Insight-Based Studies for Pathway and Microarray Visualization Tools

Purviben Bharatkumar Saraiya
Dissertation Submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science and Applications

Dr. Chris North, Chairman
Dr. Karen Duca,
Dr. T. M. Murali,
Dr. Cliff Shaffer,
Dr. Deborah Tatar,

Keywords: Empirical Studies, Pathways + Microarray data visualization tools, Insight-based method

Insight-Based Studies for Pathway and Microarray Visualization Tools

Purviben Bharatkumar Saraiya

Abstract

Pathway diagrams, similar to the graph diagrams using a node-link representation, are used by biologists to represent complex interactions at the molecular level in living cells. The recent shift towards data-intensive bioinformatics and systems-level science has created a strong need for advanced pathway visualization tools that support exploratory data analysis. User studies suggest that an important requirement for biologists is the need to associate microarray data to pathway diagrams.

A design space for visualization tools that allow analysis of microarray data in pathway context was identified for a systematic evaluation of the visualization alternatives. The design space is divided into two dimensions. Dimension 1 is based on the method used to overlay data attributes onto pathway nodes. The three possible approaches are: overlay of data on pathway nodes one data attribute at a time by manipulating a visual property (e.g. color) of the node, along with sliders or some such mechanism to animate the pathway for other timepoints. In another approach data from all the attributes in data can be overlaid simultaneously by embedding small charts (e.g., line charts or heatmap) into pathway nodes. The third approach uses miniature version of the pathways-as-glyph view for each attribute in the data. Dimension 2 decides if additional view besides pathway diagrams were used. These pathway visualizations are often linked to other type of visualization methods (e.g., parallel co-ordinates) using the concept of brushing and linking.

The visualization alternatives from pathway + microarray data design space were evaluated by conducting two independent user studies. Both the studies used timeseries datasets. The first study used visualization alternatives from both dimension 1 and dimension 2. The results suggest that the method to overlay multidimensional data on pathway nodes has a non trivial influence on accuracy of participants' responses, whereas the number of visualizations affect participants' performance time for pre-selected tasks. The second study used visualization alternatives from dimension 1 that focuses on method used to overlay data attributes on pathway nodes. The study suggests that participants using pathway visualization that display data one attribute at a time on nodes have more controlled performance for all type of tasks as compared to the participants using other alternatives. Participants using pathway visualization that display data in node-as-glyphs view have better performance for tasks that require analysis for a single node, and identifying outlier nodes. Whereas, pathway visualizations with pathways-as-glyph view provide better performance on tasks that require analysis of overall changes in the pathway, and identifying interesting timepoints in the data.

An insight-based method was designed to evaluate visualization tools for real world biologists' data analysis scenarios. The insight-based method uses different quantifiable characteristics of an 'insight' that can be measured uniformly across participants.

These characteristics were identified based on observations of the participants analyzing microarray data in a pilot study. The insight-based method provides an alternative to traditional task-based methods. This is especially helpful for evaluating visualization tools on large and complicated datasets where designing tasks can be difficult. Though, the insight-based method was developed to empirically evaluate visualization tools for short term studies, the method can also be used in real world longitudinal studies that analyzes the usage of visualization tools by the intended end-users.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem Definition

1.1.1 Pathways + Microarray Data Visualization

Biological pathways represent complex reactions at the molecular level in living cells. Pathways may be sub-typed into different categories based on the overall effect they have on functioning of an organism. Three major categories are: metabolic pathways, transcriptional and protein synthesis pathways, and signal transduction pathways. As the requirements to analyze different kinds of pathways are similar from the software development point of view, unless explicitly stated otherwise, the term pathway refers collectively to all.

Biologists use pathways to integrate results from literature, formulate hypotheses, capture empirical results, share current understanding, and even run simulations. A common goal of research in the life sciences is to develop pathway models for biological processes of many different organisms. Pathways also serve as a focal point to integrate other diverse related information, such as literature citations, experimental data, research notes, etc. To facilitate usage and exploratory analysis of complex pathways, visual representations for the pathways are necessary. Some diagrams are manually generated such as found in textbooks [1], KEGG [2] and Biocarta [3], and others are generated by interactive visualization software such as GenMapp [4] and PathwayAssist [5].

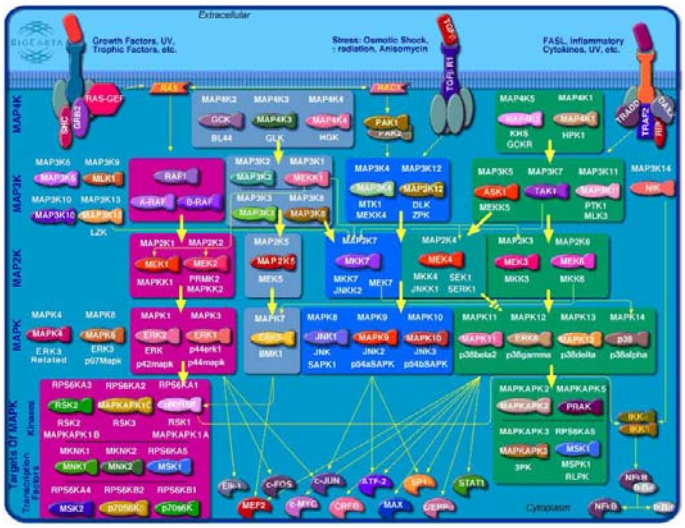
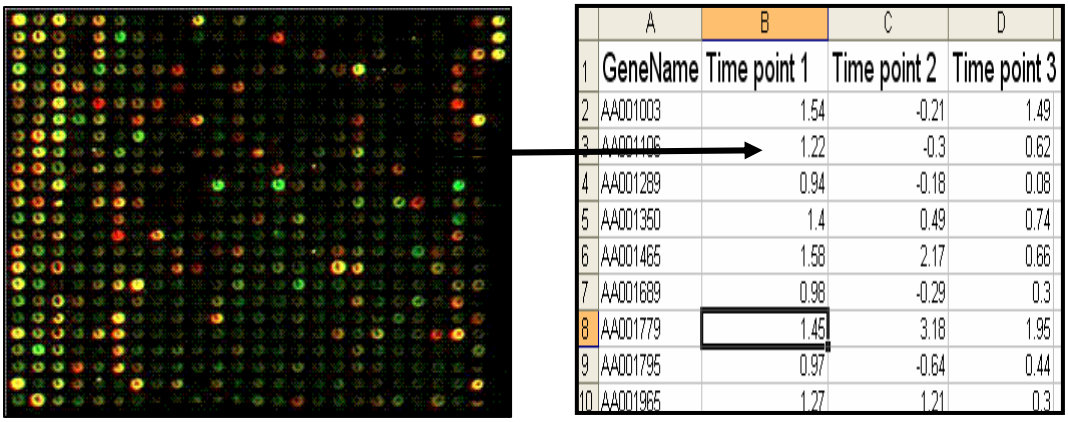


Figure 1.1 Example of a biological pathway.

In recent years, high throughput data capture technology such as microarray experiments have vastly improved life scientists' ability to detect and quantify gene, protein, and metabolite expression. A microarray experiment can simultaneously provide data about thousands of entities [6-9]. Life scientists often use microarray experiments to find cures for life threatening diseases such as cancer [10, 11]. The advent of microarray experiments is causing a shift in the way biologists do research; a shift away from simple

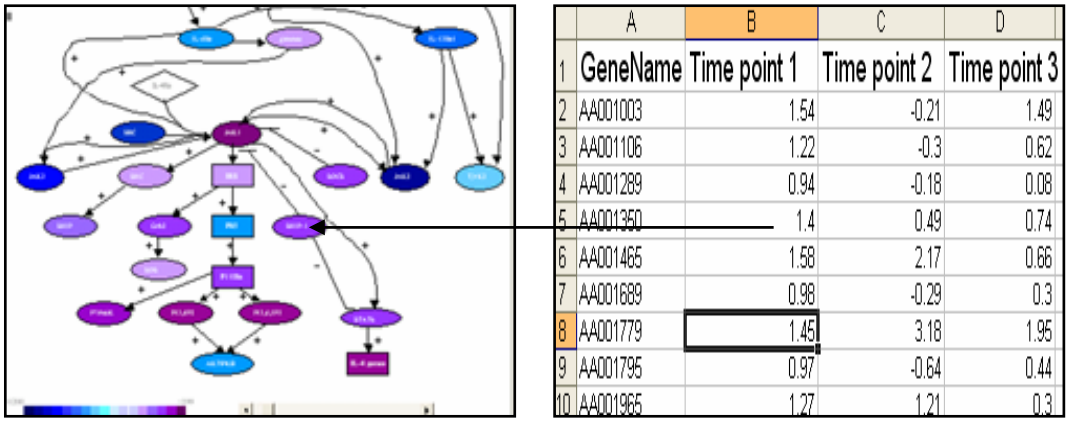
reductionist testing on a few variables towards systems-level exploratory analysis of 1000s of variables simultaneously.



	A	B	C	D
1	GeneName	Time point 1	Time point 2	Time point 3
2	AA001003	1.54	-0.21	1.49
3	AA001106	1.22	-0.3	0.62
4	AA001289	0.94	-0.18	0.08
5	AA001350	1.4	0.49	0.74
6	AA001465	1.58	2.17	0.66
7	AA001689	0.98	-0.29	0.3
8	AA001779	1.45	3.18	1.95
9	AA001795	0.97	-0.64	0.44
10	AA001965	1.27	1.21	0.3

Figure 1.2 Represents a typical microarray chip. The intensity of light at each spot corresponds to gene expression of the cDNA [9]. The microarray experiments result in large quantitative datasets.

An important requirement for the biologists is the need to associate microarray data with pathway diagrams to get the most biologically relevant insights from the data [12]. This also provides them with a biological context to otherwise plain numerical data analysis [13]. Figure 1.3 shows overlay of time series data (as an example of multidimensional microarray data) on a pathway. In response to this requirement, a large number of visualization tools that allow users to perform such analyses have been developed [14, 15]. These visualizations use different approaches to overlay data on pathways. Often the pathway visualizations are linked to other additional visualizations such as parallel co-ordinates and heat maps.



	A	B	C	D
1	GeneName	Time point 1	Time point 2	Time point 3
2	AA001003	1.54	-0.21	1.49
3	AA001106	1.22	-0.3	0.62
4	AA001289	0.94	-0.18	0.08
5	AA001350	1.4	0.49	0.74
6	AA001465	1.58	2.17	0.66
7	AA001689	0.98	-0.29	0.3
8	AA001779	1.45	3.18	1.95
9	AA001795	0.97	-0.64	0.44
10	AA001965	1.27	1.21	0.3

Figure 1.3 An example of overlaying multi-dimensional microarray data onto pathway diagrams, one time point at a time.

From information visualization point of view pathways are similar to graph visualizations using node + link diagrams. Also, microarray data are example of multidimensional datasets. Hence the main research question is: How to associate multidimensional data to the graph visualization that use node + link diagrams?

A number of studies have been performed to evaluate different graph layout algorithms e.g., [16-20]. However, these studies do not address the problem of graph + multidimensional data. A systematic approach that begins by identifying the design space for visualization tools that link graph + multidimensional data and evaluating visualization alternatives therein can help biologists decide which visualization alternative may be most suited for their data analysis.

1.1.2 Evaluation Method for Pathways + Microarray Data Visualization

Biologists analyze microarray data using pathway context in exploratory fashion. Often when analyzing their data, biologists do not have predefined hypotheses. They use visualization tools to explore data and rely extensively on the tools to provide them with new hypotheses and research questions. One of their main aims is to get most insight into their data and be able to relate insights from the quantitative data analysis to underlying biological phenomenon. Most of the insights from the tool may be observations from serendipitous data exploration rather than pre-mediated benchmark tasks.

A variety of evaluation methodologies have been used to measure effectiveness of visualizations e.g. controlled experiments, usability testing, expert inspection etc. Most of these methods fail to address the open ended and exploratory nature of the biologists' data analysis tasks. Though exploratory data analysis methods have been developed in the fields of HCI and CSCW such as interaction analysis [21], design experiments [22], situated analysis [23], and breakdown analysis [24] etc., these methods are too broad to be used directly for the purpose of evaluating visualization tools. Thus, a new evaluation method is needed that simulates the exploratory data analysis of the biologists' but is uniform enough to provide feedback about the insight capabilities of the visualization tools.

1.2 Research Questions

In order to address the problem scenarios the dissertation must address several important research questions. First, how do the pathway visualization tools fit in the overall research goals of the end users? What are the main user requirements for pathway visualization tools? How do the current pathway systems address user requirements? What are the end user reviews for these systems? Most importantly, which are the most critical requirements unmet by the current systems?

An important requirement identified in the earlier studies was the need to associate microarray data to pathways. This leads to the research questions: what is the design space for pathway + microarray data visualization tools that groups all the potential alternatives? Which is the most effective alternative from the design space in terms of the common data analysis tasks?

Most methods used so far for evaluating visualization tools do not address the real world exploratory data analysis tasks of the biologists. Hence, a new insight-based method to evaluate bioinformatics visualization was designed. This raises research questions such as: What is insight? How to identify and measure it quantifiably in experiment settings? How does the insight-based method compare to the traditional task-based method used so far to evaluate visualization tools? Do the insights generated in the evaluation study represent benchmark tasks used for evaluating visualization tools?

Finally, since the insight-based method was used for evaluating bioinformatics visualizations in a short term study, it will be interesting to know if the insight-based method can be used for a long term study. And to know if the short term Insight-based method is representative of the long term visualization tool usage.

1.3 Content

Chapter 2 describes user studies, requirements analysis, usage scenarios, literature survey for pathway visualization systems and also the approaches taken by these systems to address the end user requirements, and the heuristic evaluations on pathway visualization systems with a group of biologists. Based on these studies, a research agenda was presented concerning five critical requirements for pathway visualization systems. If addressed effectively, these requirements can prove to be most helpful in supporting exploratory pathway analysis.

An important requirement identified in Chapter 2 is the need to analyze microarray data in context of pathway diagrams. Biologists feel that the lack of pathway context can hamper their ability to derive biologically meaningful insight from microarray data [25]. Chapter 3 presents the space for microarray data + pathway visualization tools. The most common approaches to support analysis of pathways with associated microarray data include: overlay of data on pathway nodes for one treatment at a time, by manipulating color of the node. A slider is provided to animate the pathway for other treatment conditions. In another popular approach, data from all the treatments are overlaid simultaneously by using nodes-as-glyphs and alternatively miniature pathways as glyphs views. These pathway visualizations may also be linked to other visualizations (e.g., parallel co-ordinates) using brushing and linking. Systematic evaluation of visualization alternatives with these design space can help to identify the ones that will be most helpful to the biologists.

Chapter 3 also describes an initial study performed to evaluate and rank pathway + microarray visualization options based on users' performance time and accuracy of responses on predefined tasks. A timeseries data was used as an example of microarray data. The results suggest that overlaying data on pathway nodes one timepoint at a time may lead to more accurate performance for tasks involving analysis of a graph at a single timepoint, and comparisons between graph nodes for two distinct timepoints. Overlaying data simultaneously for all the timepoints on graph nodes may lead to more accurate and faster performance for tasks involving searching for outlier vertices displaying different behavior than the rest of the graph vertices for all timepoints. Single views have advantage over multiple views on tasks that require topological information. Also, the method of data overlay on the graph nodes has a non trivial influence on accuracy of responses, whereas the number of visualizations affect the participants' task performance time.

Chapter 4 presents an insight-based method to evaluate visualization tools. The method was designed as a part of the dissertation. Typically visualizations are evaluated using task-based experiments, as described in Chapter 3, that measure users' performance on predetermined tasks or using heuristic walkthroughs and expert reviews as described in Chapter 2. However, such studies fail to address the common data analysis scenario of the biologists which is exploratory and not pre-defined. Hence a more relevant evaluation method that focuses on real world data analysis scenarios is needed. The insight-based

method uses several characteristics of an ‘insight’ that allow us to recognize and quantify it in an open-ended user tests. Since the method does not use any pre-defined tasks this eliminates the need to design bench mark tasks for the user studies. This can be of help in evaluating visualization using complicated datasets for which designing tasks can be often very difficult.

Chapter 5 describes a study to evaluate visualization alternatives for pathway + microarray data using both insight-based and the task-based method. The results allow us to rank the visualization alternatives both in terms of insights reported by the participants and the performance of participants on the pre-selected tasks. Another by product of the study was the comparison between the insight and the task-based method.

Both studies reported in Chapters 4 and 5 use the insight-based method for short term studies. Chapter 6 presents a longitudinal study using the insight-based approach. Finally, Chapter 7 summarizes the lessons learnt from insight-based studies and evaluations for microarray data + pathway visualization tools.

2 Visualization for Biological Pathways

Pathway diagrams are used by biologists to represent complex interactions at the molecular level in living cells. The recent shift towards data-intensive bioinformatics and systems-level science has created a strong need for advanced pathway visualizations that support exploratory analysis. Several interviews were conducted with the biologists to understand their needs for pathway analysis. Based on these interviews, a detailed requirements analysis along with two user scenarios for pathway visualization systems are presented in this chapter. A variety of existing pathway visualization systems were also examined to list common approaches by which the contemporary systems address these requirements. A heuristic evaluation with biology domain experts for five popular pathway visualization systems was then conducted to analyze end-user perception of these systems. Based on these studies, a research agenda was presented concerning five critical requirements for pathway visualization systems. If addressed effectively, these requirements can prove to be most helpful in supporting exploratory pathway analysis. The requirements were: 1) automated construction and updating of pathways by searching literature databases, 2) overlaying information on pathways in a biologically relevant format, 3) linking pathways to multidimensional data from high throughput experiments such as microarrays, 4) overviewing multiple pathways simultaneously with inter-connections between them, 5) scaling pathways to higher levels of abstraction to analyze effects of complex molecular interactions at higher levels of biological organization.

2.1 Biological Pathways

Biological pathways represent networks of complex reactions at the molecular level in living cells. They model how biological molecules interact to accomplish a biological function and to respond to environmental stimuli. Pathways capture the current knowledge of biological processes and are derived through scientific experimentation and data analysis. Biologists use pathways to integrate results from literature, formulate hypotheses, capture empirical results, share current understanding, and even simulate processes. A common goal of research in the life sciences is to develop an ever-broadening library of pathway models for biological processes of many different organisms. Such pathways can have significant broad impacts, such as making products in biotech applications and drug discovery in the pharmaceutical industry.

Pathways also serve as a focal point to integrate other diverse related information, such as literature citations, research notes, and experimental data. In recent years, high-throughput data capture technology has vastly improved biologists' ability to detect and quantify gene, protein, and metabolite expression. Such experiments can simultaneously provide data about thousands of entities [6-9]. All this data must be analyzed in the context of the pathway diagrams to enable biologists to make inferences about the underlying biological processes and to improve the current pathway models. Hence, the increasing complexity of pathway diagrams derives not only from their size and representations, but also from the large amount of important related information.

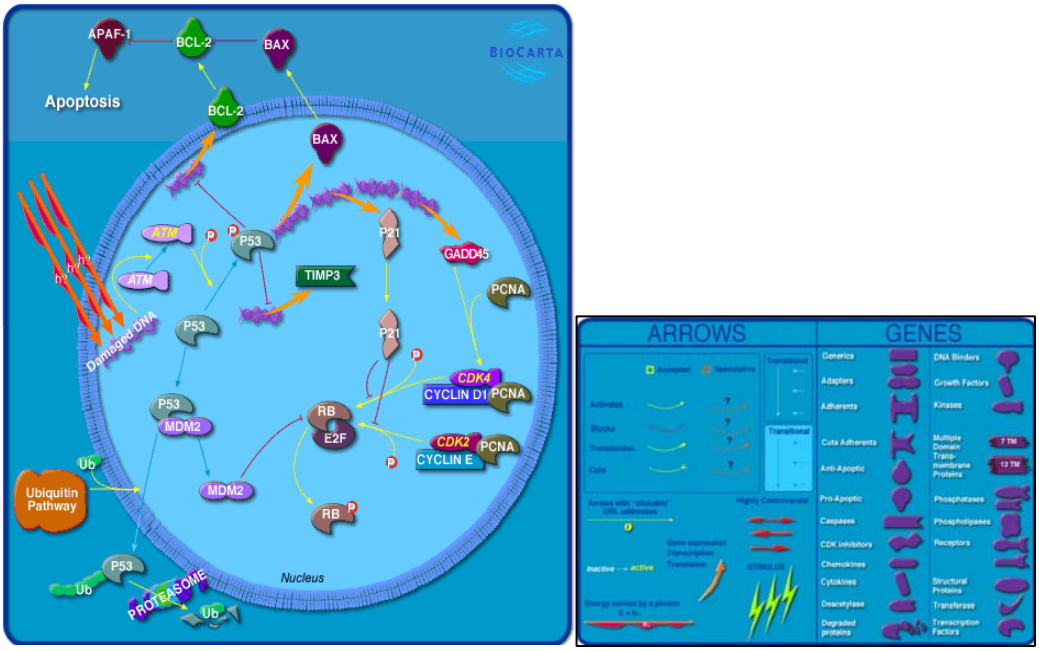


Figure 2.1 The p53 signaling pathway in a stylized diagram from BioCarta [3], including format to represent biological, spatial, and temporal properties.

The increasing importance of exploratory pathway analysis corresponds to a major shift in emphasis in biological research; a shift beyond the reductionist scientific process, which rigorously examines individual interactions of biological molecules, towards *systems-level* science, which simultaneously explores entire systems of many biological molecules. Systems-level science highlights that the whole is greater than the sum of the parts. A challenging goal for pathways is to try to convey complex global functionality, interconnections with other pathways, and their dynamic behavior.

To facilitate the exploratory analysis of complex pathways, visual representations are necessary. Pathways are typically represented as network diagrams (Figure 2.1). Some pathway diagrams are manually generated such as those found in textbooks [1] or KEGG [2], whereas others are generated by interactive visualization software such as GenMapp [4] and PathwayAssist [5]. However, although several pathway visualization systems have been developed recently, there is little guidance for the design of such tools e.g. [26, 27]. Though there have been a few studies on graph layout and aesthetics [19, 20] their utility and impact for pathway visualizations is unclear.

In discussions with biologists, we found that many are skeptical about the biological value of current pathway visualizations. When considering cost vs. benefit, the cost seems to outweigh the benefits. They are reluctant to invest time required to overcome the learning curve for many of these systems. A large amount of effort is required to gain biologically meaningful insight for specific projects from most of these systems. The tools lack many important data analysis capabilities that scientists need. Thus, to truly enable a shift towards systems-level science, more rigorous requirements analysis and evaluation of pathway visualization systems are needed.

2.2 Procedure

Generally in HCI, analysis of requirements starts with interviewing and observing current work practices of users. These observations can be contextual (users are observed as they carry out their tasks), or participatory (users are engaged in discussions). Results of these observations are scenarios and requirements that help developers understand how users will eventually use a system and its impacts [28].

We focused on biologists as the primary user class, and life science research as the primary usage scenario. To understand pathway usage, we interviewed four research professors and post-doctoral fellows having diverse research interests and several years of research experience, over a period of six months. We met with each researcher usually once or twice a week. The researchers were selected based on their availability and willingness to participate in the discussions.

We generally interviewed only one researcher at a particular time. Each interview session lasted for about one to two hours. Most of these interviews were informal and participatory. We did not ask the researchers a specific predefined set of questions. The biologists explained their research work to us and its biological significance. They also explained importance of biological pathways, different contexts in which pathways are used, different types of information needed from pathways and the current methods to obtain this. The biologists also discussed their research work, experiments, data analysis tasks, and how pathway diagrams fit into their overall research goals. We also attended presentations and seminars conducted by these biologists to understand their work in a broader context.

In addition to the interviews, we conducted two focus group meetings, with about ten biologists (two of these were researchers we interviewed extensively). In the group meetings, we discussed the requirements derived from earlier interviews. In addition, we attended the journal club meetings of a life science research group, where we discussed published research about high-throughput data experiments. Based on these studies and group meetings, we derived a final list of requirements for pathway analysis. To get feedback from additional biologists, a short questionnaire was sent via email listservs. The scientists were requested to rate the degree to which they agree or disagree with the requirements.

To analyze the end-user perception of existing pathway visualization systems, we conducted a heuristic evaluation with six biologists on five pathway analysis systems. Participation in the evaluation was voluntary. This heuristic evaluation was a form of user study in which biology domain experts reviewed systems to suggest advantages and disadvantages against the list of requirements [29]. This approach helps to further elucidate the requirements and how the systems meet biologists' needs. The results provide useful guidance for developing pathway visualization software.

2.3 Pathway User Scenarios

The use of pathways depends on the progress of the research project. Two scenarios are described here. The first scenario is for a project in its initial stage. The second scenario describes microarray data analysis in context of pathways, and is more likely to occur towards the middle or completion of a project.

2.3.1 User Scenario -1

Consider Jim, PhD in microbiology, working at a research institute. One of his primary research interests is to understand the effects of cigarette smoke on lung cells. Besides this, he also wants to know how flu's effect on lung cells is altered by the presence of cigarette smoke. He is working with several different pathways. He started with the cell apoptosis, i.e., programmed cell death (an example of a signaling pathway) pathway. He wants to know which components of this pathway are affected by cigarette smoke and flu.

Due to his domain knowledge, Jim has some intuition and ideas about different entities and how they interact. But for his project, he needs an accurate, update and detailed knowledge. He needs to collect as much available information as possible about the apoptosis pathway. He started by looking at various sites on the Internet and scientific journals. Over time, a large of research articles, hyperlinks and papers relevant to his work were collected.

As he was collecting information, Jim wants to link these materials to understand how different part of pathway function collectively to produce an overall effect. He needs a way to organize information so that he can refer later part of the pathway and paper it belongs to. Some parts and interactions are well established where as some can be hypothesis proposed by researchers but need more proof to be validated. He agrees with some researchers whereas some he is skeptical about. Before he conducts an experiment, he may have hypothesis about parts of the pathway that can be affected by smoke and/or flu. He needs to represent this in his pathway diagrams so that he can later analyze his predictions in terms of final experimental results.

It is possible that as Jim is working, new results about the pathway are published in the scientific literature. So the pathway has to be constantly updated. He needs to keep track of how other scientists are dealing with similar pathways and their hypotheses and results.

Jim saw a demonstration, in a research seminar he attended, of Cytoscape, a pathway analysis tool. He felt that the tool is easy to use and will not require too much learning time. An important selection factor was that the tool is free. This was ideal for his early analysis and will help him to decide if a software tool can actually help. His experience will also guide him if he later decided to purchase a commercial product.

Figure 2.2 shows the apoptosis pathway diagram he constructed in Cytoscape. The color of the node reflects its type, i.e., gene, kinase, etc. Similarly the color of the edge reflects the type of interaction, i.e., inhibitory, stimulatory, etc. Jim also attached different annotations to the nodes, e.g., which paper was used to obtain information, the behavior of node in a particular experiment, etc. Jim wanted to include representation of cellular location (e.g., cytoplasm, nucleus, cell membrane, etc.). He found this difficult as no direct features are provided by Cytoscape to do this. He used shape and size of node to impart this information. He was using color to focus on type of node. He was considering use of arrow thickness and type to overlay information about reaction types.

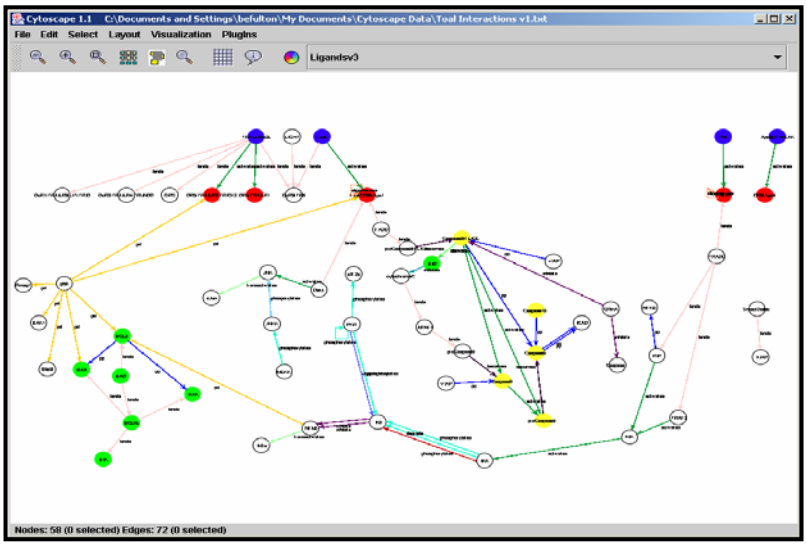


Figure 2.2 Shows the apoptosis pathway that Jim is building. He still needs to add more proteins and genes to make it more complete. He used Cytoscape [30] to build his pathway.

A problem he faced with Cytoscape was to save and update pathways. He found this too time consuming. He needed to list names of nodes and edges in text files and then load this into the pathway displayer. The .gml files used by Cytoscape let him save the look of a pathway, but he could not directly add nodes to this files. Cytoscape lets users link gene ontology information to the pathways. Jim focused on human pathways, and so needed to link humane gene ontology information to the apoptosis but could not figure out how to do this in Cytoscape.

Later, when Jim conducts an experiment he will like to see how pathway components are affected by different experiment conditions. He will also like to compare the result of his experiments with other researchers who have performed experiments, on the apoptosis pathway.

If the end results of Jim's experiments are not as hypothesized, he will like to determine how the pathway or his hypotheses can be modified to explain the results. For this he will need to have deeper understanding into the functioning of the apoptosis pathway. He will need to understand the effects of different pathways on the cell apoptosis pathway to explain the overall observed results. This can lead to more hypothesis generation and experiments.

2.3.2 User Scenario -2

Drought is one of the several causes for decreased crop yields. To genetically engineer plants more resistant to drought stress and thereby increase crop production, biologists at present are trying to discover genes that enable plants to cope with reduced water conditions. For our second scenario we focus on Jonathan, a life scientist working in the area of plant physiology. He has a Ph.D. in horticulture. For his research he focuses on pine trees.

To identify genes associated with stress acclimation, he needs to understand changes in the gene expression patterns in response to the stress. He analyzes the genetic expression profile of trees that were subjected to stress compared with those that were not

stressed. For his experiment, pine trees were subjected to three cycles of either mild or severe drought stress and their physiologic response to the stress measured.

The gene expression profiles allowed him to analyze changes in gene expression related to acclimation under mild stress and to identify differences in the gene expression between mild and severe stress levels. He found that trees grown under mild stress acclimated to the stress such that they were able to maintain growth. This was only noticeable in trees in the second and third cycles, meaning that after the initial cycle of stress, trees had acclimated and were able to withstand subsequent stress cycles. No such acclimation was seen for severe stress.

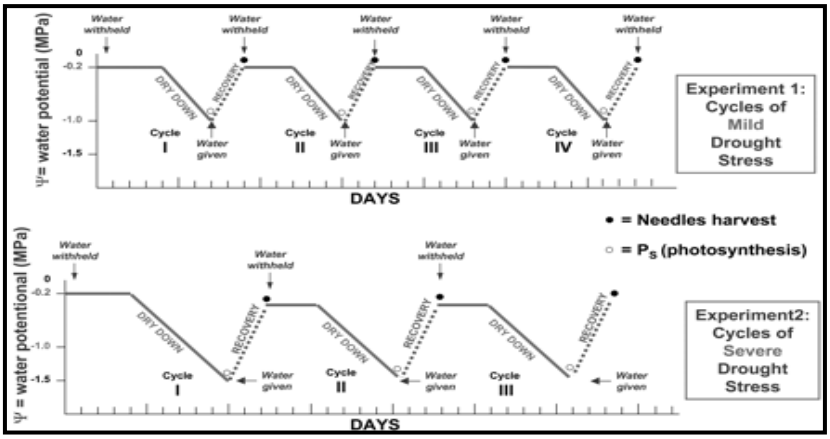


Figure 2.3 Represents water levels and days. The measurements were taken for three mild and severe drought stress cycles [31].

About two thousand different pine genes were represented on the microarray. These were selected from 5 pine EST libraries and were selected to represent all 15 MIPS [32] functional categories. Data for the six experimental conditions were recorded. While analyzing data, biologists generally look for genes that show change in expressions. Jonathan reclassified genes into functional categories designed to reflect drought stress specific responses. Some of the most important inferences are made by computer algorithms that associate gene expression changes with functional categorization

It is difficult and very time consuming to analyze microarray data manually, due to the quantity of data from a single experiment. To help him, Johnathan works with different computer scientists. He depends on different data mining algorithms and Inductive Logic Programming (ILP) [33] for identifying patterns of gene expression. An example of the ILP rules is: if a gene (X) is in the category of (Y), then it is positively expressed under (Z) stress. The functional categories with significant rules for both mild and severe stress cycles are shown in the figure 2.4.

As shown in the figure 2.4, Genes belonging to Carbon metabolism were found to show most change in all the three mild cycle conditions. However, they did not show significant changes in severe cycle. Jonathan wanted to analyze which genes belonging to this category changed and how they affect overall metabolism. He referred to different reference books and papers to determine how changes in carbon metabolism might affect or relate to other pathways understand what other pathways entities belonging to carbon metabolism pathway affect and if the changes in this pathway were reflected in other pathways.

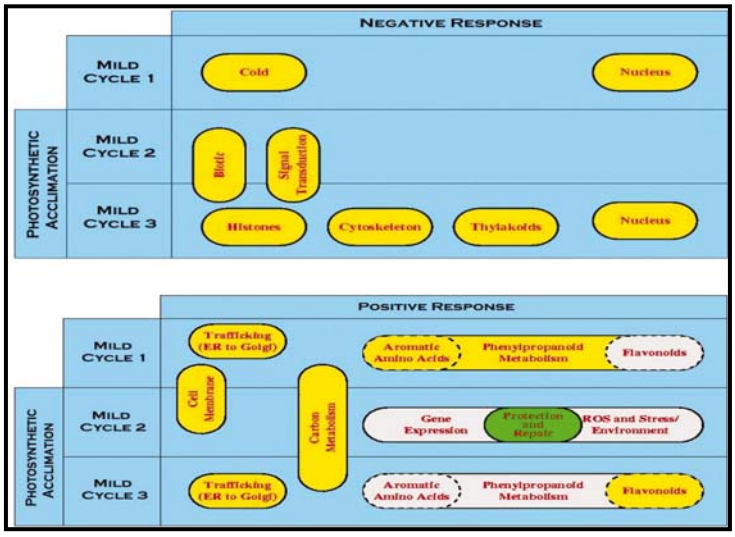


Figure 2.4 Different functional categories identified by ILP that showed a significant change for both severe and mild stress cycles [31].

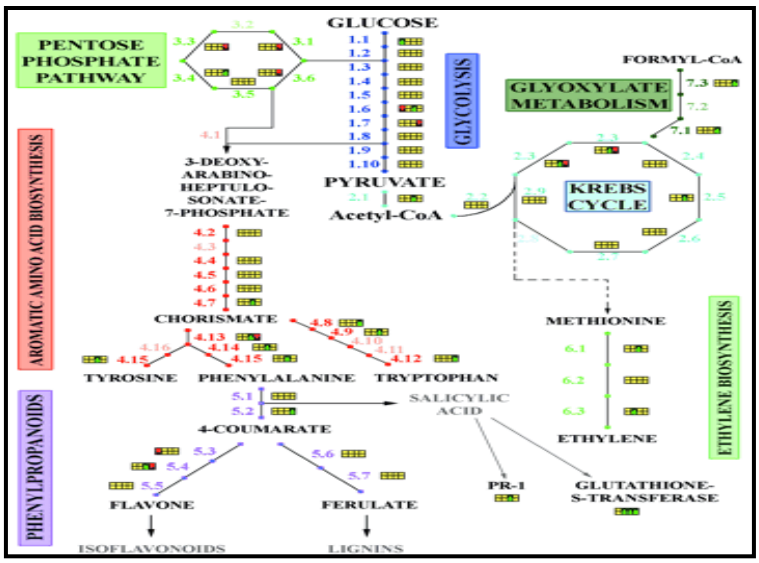


Figure 2.5 represents change in gene expressions of seven selected metabolic pathways in a time series dataset [34].

He initially wanted to depict changes in these functional category on diagrams that also showed the inter relationships between these genes and pathways. He came across the following diagrams used in a research article and found them interesting. He tried to use similar visualizations to present his results. The initial pathway diagram, he constructed, are shown in Figure 2.7. He used Microsoft – Powerpoint to create the diagram. (We were not able to obtain the original diagram, the below diagram is very similar to the original. It was created using XFig).

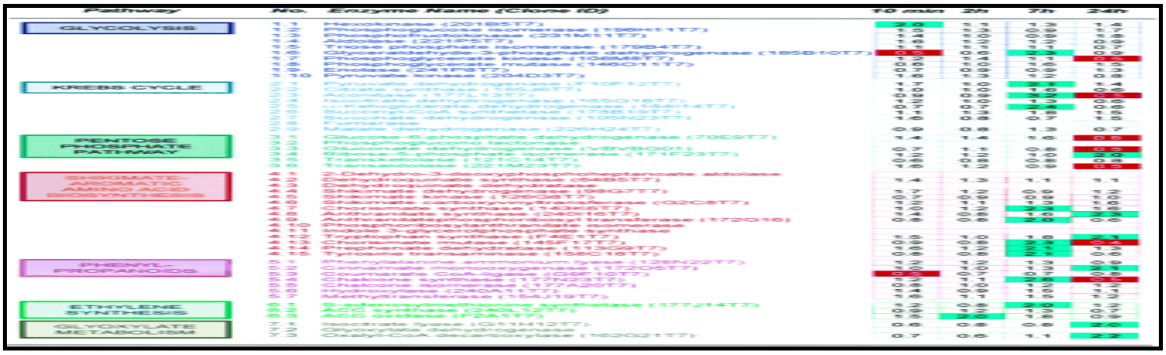


Figure 2.6 shows detailed view of the genes belonging to pathways shown in the figure 2.5 along with their expression values [34].

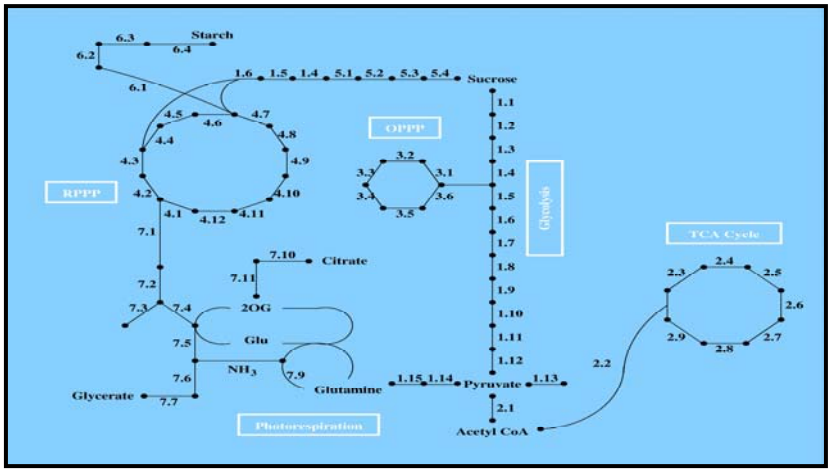


Figure 2.7 Diagram representing genes involved in various carbon metabolisms and their interrelations.

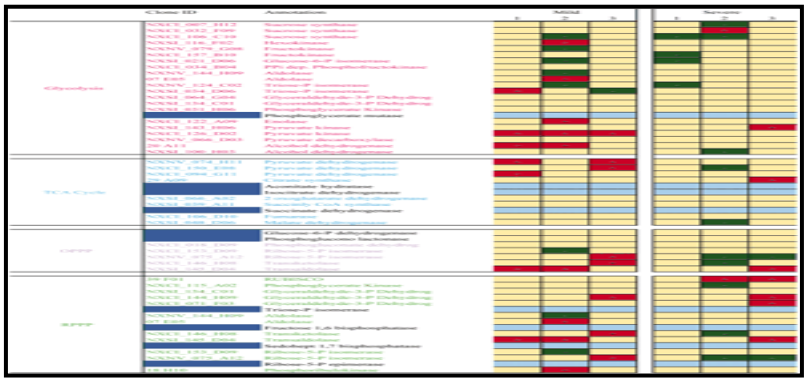


Figure 2.8 Diagram used by Jonathan to show gene expression change in different experimental conditions [31].

Jonathan wanted to overlay expression values for the entities in all the six experimental conditions. This would enable him to compare functionally how the pathways changed in different conditions. Some enzymes are encoded by multiple gene families, and for some enzymes several members of the gene family were present on the

array. This means, that for some (nodes/edges) in the pathway, there were multiple genes with different expression profiles. He decided that it was too time consuming and difficult to show all this information using Microsoft powerpoint. So he decided to use the diagrams shown in Figure 2.8 to represent the changes detected in the microarray experiments. Figure 2.8 shows genes that show significant changes across different experimental conditions. Red color implies that the gene was up regulated and green means it was down regulated. The genes for this diagram were all in the same functional category. Jonathan divided the diagram to represent different branches of the main carbon metabolism pathways. The sequence of interactions in the pathway is implied through linear ordering of genes in the table. The replicate genes have same names and can be uniquely identified by their clone Ids. From figure 2.8 he reached to conclusions that the genes showed more changes in their expression in mild cycle as compared to severe cycle. There were more positively expressed genes in the mild stress cycle. He had to refer to the scientific literature and also use his domain knowledge and intuitions to speculate what other pathways were affected by these genes.

Jonathan's main compounds of interest are flavanoids. He knows that this group of molecules is involved with stress responses in plants. It would help him to see how the flavonoid biosynthetic pathways are changed and what entities or pathways might also affect flavonoid synthesis. This would help him make better conclusions about how the plant is able to control aspects of metabolism to mobilize a response to the stress.

Ideally, Jonathan would like to overlay the experimental data on an entire metabolic pathway and see how the pathway is changing in different experimental conditions. He would like to see how changes in some genes affect the overall metabolism, and if he could relate this change to the observed physical changes in an organism. Jonathan also wanted to know what experiments besides his showed similar results. He would like to refer this and present his work relating to the previous work done in this area.

Most biologists know the pathways shown in the above diagrams due to their domain knowledge. However, certain pathways such as Lignins and Flavonoids require more specialized knowledge. To readers unfamiliar with these pathways it would be more helpful to represent these pathways using a graphical map showing interactions between entities involved and connections with other pathways.

Subsequent to his analysis, he became aware of a tool called MapMan [35] in one of the presentations he attended. He found that the visualizations created by this tool were very simplistic, which made them easy to follow. However, it lacked details on members of gene families. The visualizations were very good for the experiments that were being explained at the conference. He felt that this tool may be useful for his future work.

2.4 Biological Pathways

2.4.1 Pathway Description

There is not yet a standardized language for pathway components, as it is highly dependent on the domain and the particular need that motivates the construction of any given pathway. In many cases, a "pathway" is the user-defined network of the biological interactions under study in a particular research group. Pathways in life science research are extremely diverse. Some capture higher level abstractions, while others are very specific. Some are sketchy, while others are rigorous. Figure 2.1 shows an example of a

pathway. Overall, pathways provide an approximate model or explanation of the underlying biological process.

Typically pathways are represented as a graph, consisting of nodes and edges. A node in a pathway usually represents a biological molecule, but could also be used to summarize another entire pathway that interconnects with the one under study, or to represent any other relevant phenomena such as an environmental stimulus (e.g., heat or light). A node representing a biological molecule in a pathway diagram may be either metabolite, nucleic acid, or protein. Nucleic acids can be DNA, mRNA, tRNA, and structural RNA, etc. Proteins can be enzymes, structural proteins, chemical effectors, etc. Enzymes are further divided into ligases, phosphatases, kinases, etc. Structural protein can be microtubules, actin filaments, etc. Chemical effectors can be hormones, cytokines, chemokines, growth factors, etc. An edge in a pathway usually represents a relationship or some form of interaction between the nodes. The interaction could of many types: gene expression, inhibition, catalysis, chemical modification, etc.

Pathway graphs can be complex multi-modal or hyper-graphs. While simple graphs can capture the very basic events represented in the pathway, complex biochemical dynamics do not lend themselves well to basic graph representations. An edge could connect three nodes or might connect a node to another edge. For example, an inhibitory interaction (edge) actually indicates a deeper process by which one molecule (node) might prevent some other interaction (edge) from occurring.

Based on the overall effect they have on the functioning of an organism, pathways may be divided into several different categories. Three example categories are: metabolic pathways, gene regulation/transcription pathways, and signal transduction pathways. In this dissertation, we emphasize this fairly broad notion of pathways. We do not focus on one type of pathway or specific set of pathway elements because (a) the requirements to analyze different kinds of pathways are similar, and (b) it is a long term goal to produce software that can integrate a broad variety of pathways to support the grand vision of combined systems-level analysis. Unless explicitly stated otherwise, a pathway in this discussion refers collectively to all types.

2.4.2 User Classes

The primary users of pathway visualization tools are advanced academic, industrial and government researchers in the life sciences (i.e. biologists, biochemists, chemists, biomedical researchers, etc.). Their goals are to construct pathway diagrams that model biological phenomena as closely as possible, based on literature and experimental results. This is somewhat analogous to a computer scientist attempting to reverse engineer an algorithm by running the compiled code on a variety of inputs and examining the outputs. Each researcher is generally focused on contributing to a small set of pathways representing their area of interest and expertise. They are very knowledgeable about the details of these pathways. However, they must make use of other pathways for which they may have only general knowledge or know little about.

The biologists interviewed in this study work in small teams of about 5-10 people. A team includes undergraduate and graduate students, lab technicians, post-docs and senior researchers. Data to construct pathways is generally provided by more senior investigators. Multiple research scientists in the same or different research institutes may collaborate on identical problems. At the highest levels, there are internationally

renowned scientists who curate newly made discoveries and resolve discrepancies in research findings, e.g., The Alliance for Cellular Signaling [32].

2.4.3 Pathway Research Overview

Pathway research is strongly iterative and evolving. A critical component of the research process that enables biologists to continue the experimental feedback loop [36] is inference. Inference enables them to turn experimental data results into refined hypotheses. Some common pathway inference tasks that biologists perform include: 1) recognition of changes between experiment and control or between time points; 2) detection of changes in relationship between components of a pathway or between entire pathways; 3) identification of global patterns across a pathway; and 4) mapping pathway state to phenotype (observable effects at the physical level in living organisms) or other biological information [13]. Sometimes, the new discoveries fail to support past assumptions, leading to further experimentation and research, culminating in modified pathways. Pathway modification is a continuous, evolutionary process.

Some hypotheses and research questions are relatively simple, and can be answered through scientific reduction methods. However, with the advent of systems-level analysis, it is becoming more common to examine hypotheses that are significantly more complex. Researchers are typically interested in pathways that contain approximately 50 to 500 nodes. But when inputs to these nodes from other pathways (that in turn may be affected by several other pathways) need to be taken into account, things quickly get more complicated. Inferences that must be made in these cases are equally complex, requiring the recognition of subtle effects at various levels of scale involving multiple pathway networks. These inferences are well beyond the capabilities of current pathway visualization techniques.

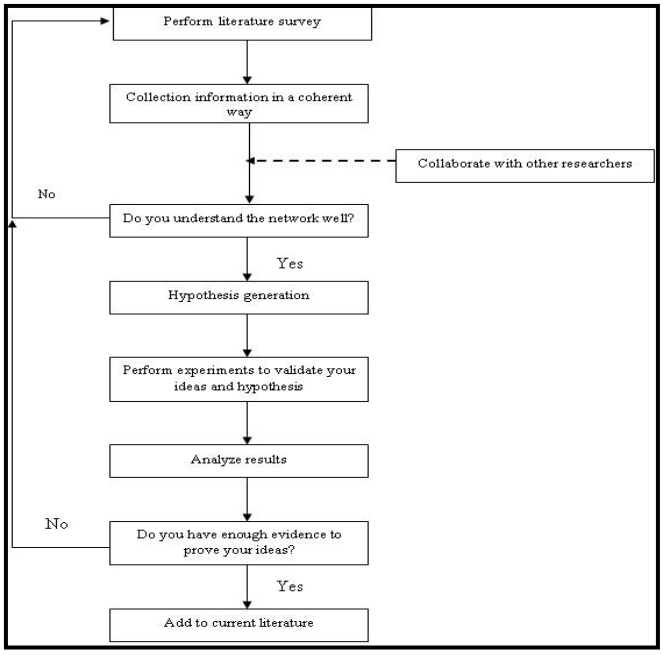


Figure 2.9 represents a sequence of tasks carried out by a life scientist to perform their research.

2.5 Requirements Analysis

Based on the interviews and focus group meetings with biologists, a list of requirements for pathway visualizations were developed as shown in Table 2.1. The requirements are grouped into three main categories: pathway assembly, information overlay, and pathway analysis. These categories are described in the following subsections.

Accomplishing these requirements will require interactive dynamic visualizations. Static, textbook-like pathway representations will not be adequate in the long term. While these functional requirements provide guidance, they do not directly dictate visualization design. It might not be possible to adequately satisfy all requirements with a single design, and tradeoffs will likely need to be carefully balanced.

Categories	Requirements	Tasks
Pathway assembly	1. Construct & Update	Collect and link pathways from multiple resources
	2. Context	Provide information about pathways
	3. Uncertainty	Maintain alternate hypotheses and information reliability
	4. Collaboration	Enable group work
Information overlay	5. Node & edge representation	Details about pathway entities and interactions
	6. Source	Details about source resources
	7. Spatial information	Physical locations of pathway entities in the cell
	8. Temporal information	Time related properties
	9. High throughput data	Expression data from high-throughput experiments
Pathway analysis	10. Overview	Comprehend large or multiple pathways
	11. Interconnectivity	Intra- and inter-pathway effects of entities on each other
	12. Multi-scale	Relate pathways at different levels of abstraction
	13. Notebook	Track accumulated research information

Table 2.1 Summarizes requirements for pathway visualization systems. The requirements are grouped into three main categories: pathway assembly, information overlay, and pathway analysis.

2.5.1 Category: Pathway Assembly

The requirements in the pathway assembly category support the assembly and maintenance process for pathways.

R1. Construct & Update: A complete pathway is generally not available from a single source. Biologists often must combine different parts of a pathway from various sources, including reference archives such as KEGG[2], research articles, etc. It is also important to continually capture updates of source information in order to keep a pathway in sync with the latest knowledge.

R2. Context: A pathway may be clear to the author because of deeper understanding of the components (nodes and edges) involved. But the same diagram may be difficult to understand by someone not familiar with the underlying biological process. It is therefore advisable to include information such as pathway significance, specific conditions for it to function, collective effects of the pathway components, history of updates, etc., in some form when creating a pathway. If a pathway from a community resource is modified, then the rationale for doing so should be stated explicitly.

R3. Uncertainty: Pathways are constantly evolving. Some relationships between pathway components may be uncertain, and may require more research to be accepted.

Known facts should be distinguished from hypotheses. Representations for alternate, potentially conflicting, hypothesis should be supported.
R4. Collaboration: More than one life scientist can be working together on the same pathways. They need ways to communicate effectively with each other.

2.5.2 Category: Information Overlay

Pathways are tightly linked to many other types of biological information, and it is critical that pathway visualizations depict this richness of information in order to be biologically relevant. Pathway visualizations that look like simple ball-and-stick graph drawings are likely to be considered information-poor, and not biologically meaningful.

R5. Node and Edge Representation: Pathway nodes and edges have information attributes that visualizations should reveal through their visual representations. Quick interactive access to further details should also be provided. Pathway nodes can represent many different types of entities (e.g., genes, enzymes, etc.), which may have different chemical properties that visualizations should depict. Nodes labels for the entity names must be clearly visible. Biologists need to attach notes to pathway nodes for future reference, and be able to link them to databases such as GenBank and Gene Ontology for up-to-date information. An edge between two nodes usually implies a certain type of relationship (e.g., expression, catalysis, etc.), perhaps with properties such as rates, that visualizations should depict.
R6. Source: To evaluate a pathway, it is important to have access to the source information for its components, such as literature citations, experimental data, etc.
R7. Spatial Information: Visualizations should represent the physical, spatial attributes of the biology of the pathway, such as location within the cell, relative distance, containment, nodes bound to each other, etc. Sometimes the entity represented by the node can be present in different parts of the cell in different states.
R8. Temporal Information: Pathways often have time lag information associated with edges. Events can occur strictly in a particular sequence, simultaneously, cyclic, or mutually exclusive. Many pathways have a primary linear structure, with supporting secondary branches.
R9. High-throughput Data: A crucial requirement is to examine changes in pathway components based in high throughput data experiments such as microarrays. Microarrays allow biologists to measure expression of several thousand genes simultaneously. The raw dataset needs to be preprocessed before it can be used for analysis. Typically, for each experiment, data can be captured for each gene over multiple time points as well as multiple conditions. Hence, pathway nodes contain multi-dimensional quantitative data. This data could also be generated through simulation.

2.5.3 Category: Pathway Analysis

Pathway visualizations must enable analysis of complex pathways and hypotheses, beyond simple small effects to very large systems-level interactions.

R10. Overview: Pathways can be large, containing hundreds or even thousands of nodes, with complex interactions throughout. Furthermore, since each pathway provides a specialized focused ‘view’ on a certain biological function within the larger biological

system, pathways are neither independent nor isolated. Biologists need to overview multiple pathways collectively, with layouts that reveal global patterns and effects in context. Figure 2.10 from KEGG, provides a comprehensive overview for metabolic processes.

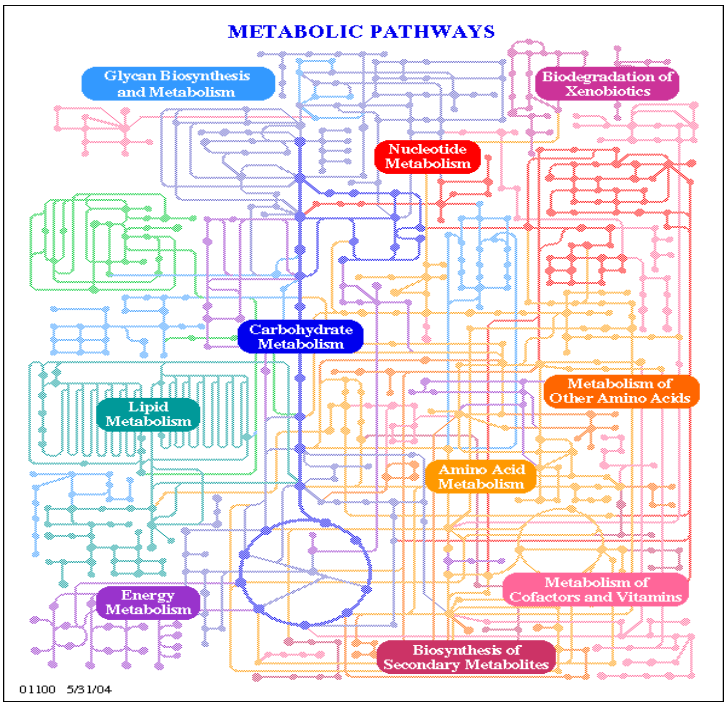


Figure 2.10 Provides an overview diagram that shows interconnectivity between metabolic pathways, taken from KEGG.

R11. Interconnectivity: Pathways are highly interconnected. Components can affect each other directly or indirectly. A single node could be involved in multiple pathways. As complexity increases, it becomes more difficult to understand connections between distant components. Biologists need to see both upstream and downstream effects from a local region of interest, including other pathways that might be affecting the focal pathway.

R12. Multi-scale: Higher level pathways can be composites of more basic pathways. In the extreme, a small change in a molecular interaction can have substantial effects at physiological levels. In such cases it is necessary to create multiple levels of abstractions to relate molecular components to higher level abstractions, and to be able to relate effects across these levels of scale.

R13. Notebook: A research group might work for several years on a set of pathways. During this time, they might obtain many results about the pathway entities. They need a logical way to keep track of collected information, along with textual notes.

2.5.4 Questionnaire

To validate and prioritize requirements and get feedback from more biologists, we sent a questionnaire to about 100 biologists using email listservs. We asked the scientists to rate each requirement according how much they agreed or disagreed with the requirement. Ten scientists responded to the questionnaire. Requirements that are highly rated

(strongly agree) by more scientists provides a basis for priority over lower rated (strongly disagree) requirements. Table 2.2 describes the questionnaire and the number of responses.

Most of the biologists agreed with the requirements list we compiled. A few of the requirements received many high ratings. The need to assemble pathways from different resources, to link source information, to infer the change in pathway components over several different experiment treatments, and to analyze the influence of pathways on one another were considered very important requirements. Most biologists commented that they were not satisfied with diagrams provided by current pathway visualization software. The visualizations should provide information about the biological properties and about the spatial and temporal relationships between the pathway components.

	Pathway Questions	**Strongly Agree**	**Agree**	**Neutral**	**Dis-agree**
Category: Pathway Assembly					
R1: Construct & Update					
1	In my work, the entire pathway(s) is generally not available from a single source.	4	6		
2	It would be valuable to have tools that allow pathway import from multiple sources.	7	3		
3	Assembling the pathway manually is one of the most time consuming processes in the whole endeavor.	4	4	1	
4	Tools that can partially build the pathway from literature or other sources would be of great value to me.	6	2	2	
R2: Context					
5	For my work, even if the pathway is fairly well known, I need to be able to modify it if I got it from a published source.	2	5	3	
R3: Uncertainty					
6	I want to represent hypothetical connections and/or nodes that have not yet been validated.	2	4	4	
R4: Collaboration					
7	I collaborate with others and need my tool to allow them to enter changes from remote sites.	1	4	4	1
Category: Information Overlay					
R5: Node and Edge Representation					
8	I am satisfied if just the name of the bio-molecules is displayed on the pathway diagram.		2	4	4
9	I need to have more information displayed on the pathway diagram than just names and connectivity.		8	2	
10	If two molecules interact, a line drawn between them is adequate for my needs.	1	1	5	3
11	I want the edge between the interacting components to have information about the nature of the interaction attached.	3	6	1	
12	I need the edges to provide more information about the nature of the interaction.	4	5	1	
13	I need the line to indicate in some manner how certain it is that the interaction actually exists.	3	5	2	
14	I want the lines to indicate in some manner alternate options/theories in pathway connectivity.	1	6	3	
R6: Source					
15	I need to link the molecule to a database or other sources of additional information.	6	3	1	
16	I need to have a lot of annotation and references for my diagram.	2	7	1	

R7: Spatial Information					
17	Representing the cellular compartment where the components are located is important for my work.	3	3	4	
R8: Temporal Information					
18	I need to view time series data and want to see how the pathways change with time.	2	4	3	
19	I need to view how components move between cell compartments over time.	1	6	3	
R9: High-throughput data overlay					
20	Adding results from multiple experiments to the pathway diagram would be of value to me.	2	7	1	
21	I need my pathway tool to link to statistical programs for further analysis.	3	4	2	1
Category: Pathway Analysis ***R10: Overview***					
22	I need information about how the pathway I am viewing links to other pathways not displayed.	7	3		
R11: Interconnectivity					
23	I need a large amount of interactivity with the pathway diagram.	2	3	5	
R13: Notebook					
24	I need to have a history function to record all the changes I've made to the diagram with reasons for them	4	4	2	
25	I perform repetitive steps for pathway analysis session to session.	1	5	4	

Table 2.2 The questionnaire used to rate each individual requirement. The biologists were requested to rate each requirement according to how much they agree or disagree with it. The table shows the number of scientists (out of 10) that agree or disagree with each individual requirement. There were no 'strongly disagree' ratings.

2.6 Survey of Pathway Visualization Systems

A large number of systems are available for pathway visualization [14, 15, 37]. It would be very difficult to review all the pathway systems. Here, we focus on systems that were selected based on availability, popularity in the bioinformatics community, and visualization and data analysis capabilities. Though the list is not exhaustive, it provides a general overview of capabilities provided and approaches used by the current pathway visualization systems. Due to the wide range of requirements, it would be difficult for any one system to address all. We group the systems based on the category of requirements they address and the approach that they use.

2.6.1 Category: Pathway Assembly

A large number of systems have been developed to facilitate pathway construction, using different approaches. Table 2.3 groups the systems based on the pathway assembly requirements they address and the approaches used by these systems to meet the requirements. Reference archives such as KEGG provide a comprehensive list of pathways for different cellular processes. Biologists frequently use these databases for accurate and up-to-date information on pathway components. A comprehensive list of such reference databases is provided by Pathway Databases [38]. The visualizations provided by these databases are typically static and textbook-like.

Requirements	Approaches	Systems
R1: Construct & Update	Reference	KEGG, BIND [39], STKE [40], BioCarta [3], EcoCyc [41]
	Pathway editor tools	Pathway Editor [42], Knowledge Editor [43], Unipath [44]
	Construct pathways using NLP algorithms on literature databases	PathwayAssist [5], Pathway Finder [45], PubGene [46], GENIES [47], VectorPathBlazer[48], Omniviz [49]
	Construct pathways from microarray data	GenePath [50], GeneSys [51], GENEW [52]
	NLP algorithms to update local database	PathwayAssist
	Update database manually	Patika [53]
	Update pathways manually	GenMapp [4], Cytoscape [30]
R2: Context	Attach notes	GenMapp, PathwayAssist, Cytoscape
R3: Uncertainty	manipulate node and edge properties (e.g., shape, size and color)	GenMapp, Cytoscape
R4: Collaboration	Facilitate sharing across group members	Omniviz, Biological Story Editor [54]

Table 2.3 Groups systems by the Pathway Assembly requirements addressed and approaches used.

Editor tools, such as Pathway Editor [42] and Knowledge Editor [43], allow users to create pathway visualizations manually. A large number of systems, such as PathwayAssist, Pathway Finder [45], and PubGene [46], use Natural Language Processing (NLP) algorithms to generate pathways automatically from research articles retrieved from search engines. Systems such as GenePath [50] infer pathways from microarray data. VectorPathBlazer [48] can create pathways by combining information from different reference databases such as KEGG and BIND [39].

2.6.2 Category: Information Overlay

Table 2.4 presents pathway systems grouped by the information overlay requirements they address and the approaches they use. Different systems provide different ways to visually represent biological properties of pathway elements. Biological properties of pathway elements are represented in Cytoscape by manipulating visual node properties such as shape, size, and color. Systems such as Patika [53], PathwayAssist, and GenMapp provide pre-defined shapes to represent different types of pathway nodes. The Patika visualization is spatially divided into fixed areas to represent different cellular locations, such as nucleus or cytoplasm. Temporal information can be shown through animation, and is often partially revealed with top-to-bottom or left-to-right ordering of primary pathway flows. Because the amount of information to overlay on nodes is large, visualizations can easily become confusing if too many node properties are visually represented.

Requirements	Approaches	Systems
R5: Node and edge Representation	Manipulate node and edge visual properties (shape, size, color, etc.)	GenMapp, Cytoscape, GScope [55]
	Provide shapes for different types of nodes	Unipath[44] , Patika, PathwayAssist
R6: Source	Attach source information on nodes and edges	GenMapp, Cytoscape, PathwayAssist
R7: Spatial Information	Provide different shapes to show different cellular locations	GenMapp

	Manipulate node properties or use fixed layout	Cytoscape, GenMapp, STKE [40], PathwayAssist
	Divide visualization into different areas	Patika
R8: Temporal Information	Manipulate edge length, or layout pathway elements in the order in which they react	Cytoscape , GenMapp, PathwayAssist, VectorPathBlazer [48]
	Animations	STKE
R9: High-throughput Data	Overlay data on nodes (using color), one condition at a time	Cytoscape, Pathway Assist, GenMapp
	Embedded views, for multiple conditions (data visualizations such as heatmaps or line charts embedded on or near nodes)	GScope [55]
	Multiple linked views, for multiple conditions (pathways linked to other data visualizations)	GeneSpring [56]
	Visualizations for a functional group	MapMan [35]
	Automatically infer relationships between entities from data	GenePath [50]
	Overlaying replicates	GenMapp

Table 2.4 Groups systems by the Information Overlay requirements addressed and approaches used.

MapMan [35] enables users to analyze microarray data for genes grouped by their functional relationships. Users can zoom into pathways to focus on areas of interest. GenMapp (Figure 2.11), Cytoscape (Figure 2.12), and PathwayAssist (Figure 2.14) allow users to overlay data from microarray experiments on pathways. Usually, the color of a node is used to encode its expression value in an experiment, using a standard color ramp from green (down expressed) to yellow (no change) to red (up expressed). Most tools limit users to overlay microarray data for one experiment condition at a time. Then, users can animate the colors to infer changes across conditions. GScope (Figure 2.13) [55] allows users to overlay expression data for several experiment conditions at once, by embedding small charts onto each node within the pathway visualization. GeneSpring [56] uses multiple views to display separate data visualizations (such as parallel-coordinate plots or heatmaps) of multiple experiment conditions, which are interactively linked to the pathway visualization. Users can then relate the information by interactively selecting nodes in the pathway to highlight the corresponding nodes' data in the data visualizations, and vice versa.

2.6.3 Category: Pathway Analysis

Table 2.5 groups systems by the analysis requirements they address and approaches used. As shown in Figure 2.10, KEGG provides an overview representing all the interconnections between the metabolic pathways. GScope uses fisheye techniques to provide an overview for pathways, with a magnified focus region for details. Gscope also allows users to dynamically simulate the effects of a change in a relationship between two nodes on all pathways of interest. Patika and PathwayAssist let users query pathway interconnections, such as finding all nodes between two nodes of interest, or finding relationships between pathways of interest. As one form of multi-scale view, GeneSpring links pathways to separate visualizations of gene locations on the chromosome. Biological Story Editor [54] uses a novel metaphor of story telling to organize and share research information and arguments about a pathway among collaborators.

Requirements	Approaches	Systems
R10: Overview	Functional groups	KEGG, MapMan
	Zooming	Cytoscape
	Fish eye views	GScope
R11: Inter-connectivity	Up-down cascades	GScope
	Query pathways	PathwayAssist, Patika
R12: Multi-scale	Chromosome location + pathways	GeneSpring
R13: Notebook	Attach notes to nodes and edges	GenMapp, Cytoscape
	Build stories about pathway elements	Biological Story Editor

Table 2.5 Groups systems by the Pathway Analysis requirements addressed and approaches used.

2.7 Heuristic Evaluation

Based on the systems survey (Section 2.5), we selected six systems for evaluation against the requirements with users. These were selected based on their availability. Some users had favorable experiences with GenMapp and PathwayAssist and requested their inclusion in our analysis. The systems were evaluated with six biologists divided into two groups. Although most users were not familiar with all systems, their reviews are important as end-user perception, and valuable to visualization designers. The systems are listed in the order in which they were evaluated.

2.7.1 User Reviews

GenMapp: GenMapp (Figure 2.11) provides drafting tools for biologists to create pathways. Though the scientists felt that the tool was easy to use, they said that they would be interested in using GenMapp only if pre-made pathways for their interests were available. Creating pathways from scratch would be too time consuming.

GenMapp does not allow users to link pathways and analyze interconnectivity between them. The biologists felt that it would be difficult to show concurrent, dependent and mutually exclusive events. Unless arrows representing relationships were labeled it was not easy to tell their type (e.g., stimulatory or inhibitory). Ability to overlay information from microarray experiments was considered helpful. GenMapp allows users to overlay information from one experimental treatment at a time. GenMapp also recognizes and highlights replicates in a microarray experiment. The scientists were skeptical of the statistical algorithms used by MappFinder [57], but said it can provide a good start to suggest pathways of interest from a long list.

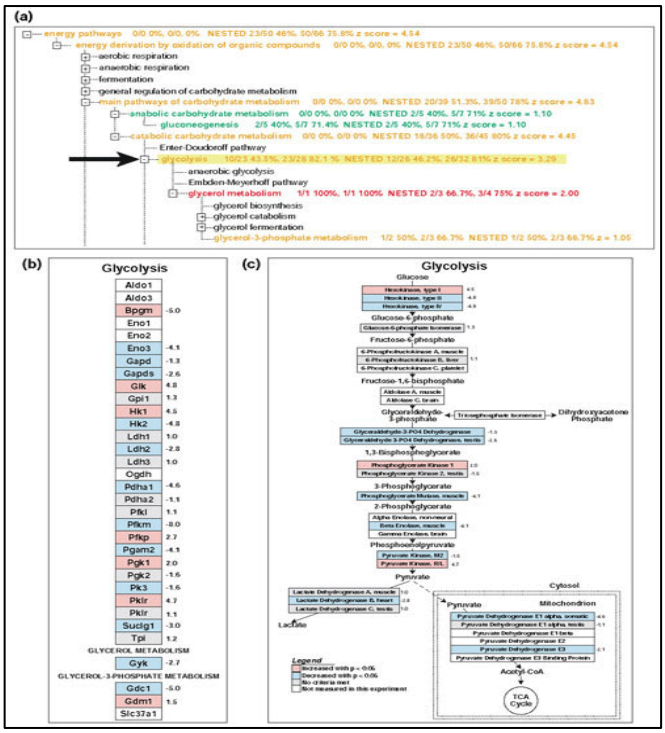


Figure 2.11 GenMapp [58]. A visualization of Glycolysis pathway in GenMapp linked to MappFinder [57]. Mappfinder, along with GenMapp, lets users perform statistical analysis on pathways to identify the most changed for a treatment. Results are displayed using the GO hierarchy as shown in (a). Users can click a pathway of interest in the hierarchy (a) for more detailed information. Pathway nodes are listed in (b). The relationships between nodes are shown in (c). The nodes are color coded based on their expression in a microarray treatment (b, c).

Cytoscape: The biologists commented it would be very difficult to understand maps created by someone else in Cytoscape (Figure 2.12). Some commented that the tool represents computer scientists' conceptions of pathways. In the overview mode, it was difficult to see the labels of genes and their properties. Without this information, a pathway is not helpful to them. They felt it would be difficult to include spatial and temporal information in Cytoscape. While information about connectivity of a node to other nodes in a pathway can be analyzed, it is difficult to comprehend overall pathway connectivity. Because of these fundamental problems, they were not impressed by the zooming capability to overview pathways. Cytoscape is created for analyzing microarray data in pathway context and provides various analytical plug-ins. Our users were mainly focused on the visualization aspects.

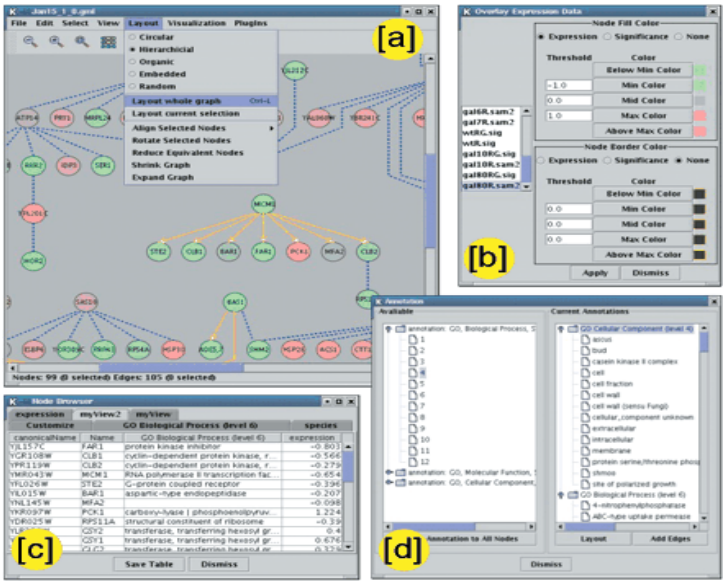


Figure 2.12 Cytoscape [59]. The color of nodes corresponds to expression data for a microarray experiment as shown in (a). Users are provided with various menus to manipulate node and edge properties (b). It is also possible to overlay annotation (c) and gene ontology information (d) on pathway nodes.

GScope: For biologists not familiar with them, fish-eye views were confusing. The distorted view and the re-orientation of the nodes when moving the fish-eye caused disorientation. Visualizations either showed too much information in the overview, or too few nodes in the case of the 'clipped view' option. It was difficult to see how a single node is related to the overall pathway. GScope (Figure 2.13) lets users simultaneously overlay gene expression data for multiple experimental treatments on the nodes. However, the pathway nodes are divided to show values for different conditions using heat map visualizations. The division of nodes, combined with fish-eye distortion, made it difficult to see overall changes in the pathway for different conditions. The scientists preferred animating the pathway node colors, showing one experiment condition at a time as done in GenMapp, over the GScope approach. There were mixed comments about the 'cascade' functionality that simulates the effect of a node manipulation on the overall pathway. One group said that this could be helpful when combined with a better means to overview the pathway. The other group, more familiar with pathway simulation tools that use differential equations (e.g, Copasi [60]), was skeptical of this implementation.

PathwayAssist: All the scientists were impressed with PathwayAssist's (Figure 2.14) pathway assembly capabilities. Some wanted to analyze the software to check if the tool really fulfills its claims of creating pathways automatically by searching the literature. They liked the ability to create pathways directly from the ResNet database [5] and from PubMed using NLP algorithms. They were excited to learn that its database has information about more than 140,000 entities, and that more can be added as required. They said that the ability to automatically link scientific references with node interactions was very helpful. The visualization also depicts the interaction type. One of the scientists was concerned about the possibility for misuse and failure to appreciate the shortcomings of NLP. Proper indication of the reliability of NLP derived information should be indicated.

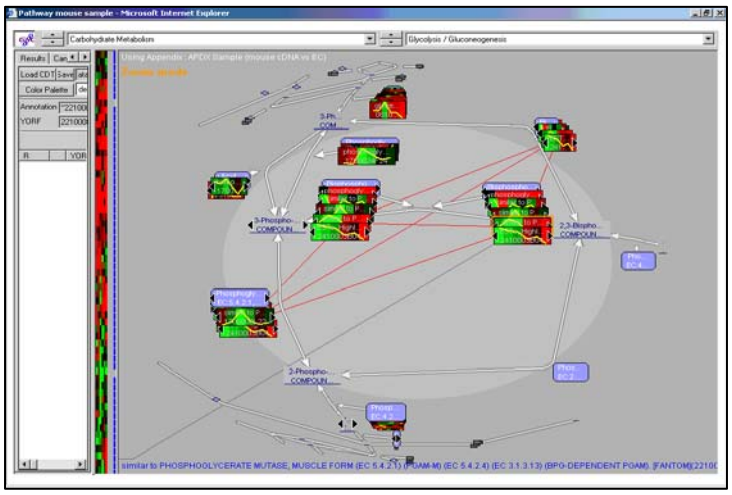


Figure 2.13 GScope [55]. Fish-eye view is used to reveal details within global context. Multiple treatments of microarray time-series data are overlaid on pathways, using colored heatmaps and line charts.

Patika: Currently, Patika (Figure 2.15) is a niche product for use in cancer research. A serious limitation is that its database is limited, and has information for just 4,000 different entities. The scientists stated that visualizations provided by Patika were more informative than other tools, because it shows multiple states of a molecule in a pathway and shows the cell compartments where the reactions take place. If information is available from the database, they found it easy to create a pathway in Patika by formulating simple queries to search for connecting entities.

BioCarta: Though we had not originally planned to include it, several scientists commented during the analysis that pathway diagrams provided by BioCarta (Figure 2.1) are among the best they have seen for providing biological context to pathways. Different types of pathway entities, the sequence of reactions between them, and the spatial relationships are all shown clearly. The symbols, shapes, and organization of the diagrams are familiar, and similar to those found in textbooks. Simply clicking on a node name reveals more information about a pathway entity. They said it is easy to comprehend the information-richness of biological pathways from these cartoon-like visualizations. They felt that none of the other pathway analysis tools provided as much information in such a helpful and biologically-meaningful visual format. It should be noted that BioCarta, unlike the other tools discussed, is simply a repository of pathway diagrams. The diagrams are manually constructed. It does not provide features like the other tools to automate pathway analysis or overlay gene expression data, but can serve as a reference library for users to construct their pathways. Hence, it serves as an excellent educational resource.

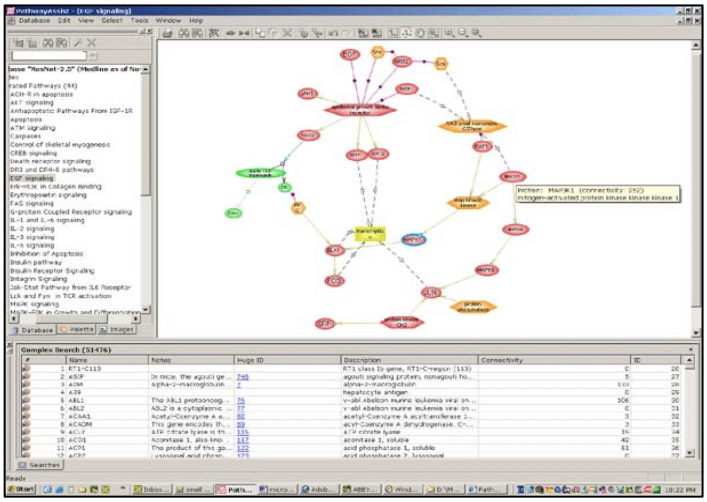


Figure 2.14 PathwayAssist [5]. EGF signaling pathway visualized in PathwayAssist. The pathway is constructed automatically using NLP algorithms, and needs to be curated by a researcher. The color and shape of the nodes denote different types of biological molecules. Also, the edges indicate if the relationship between two biological molecules is inhibitory or stimulatory. The research papers from which the information is obtained are linked to the edges. Note PathwayAssist is now called PathwayStudio.

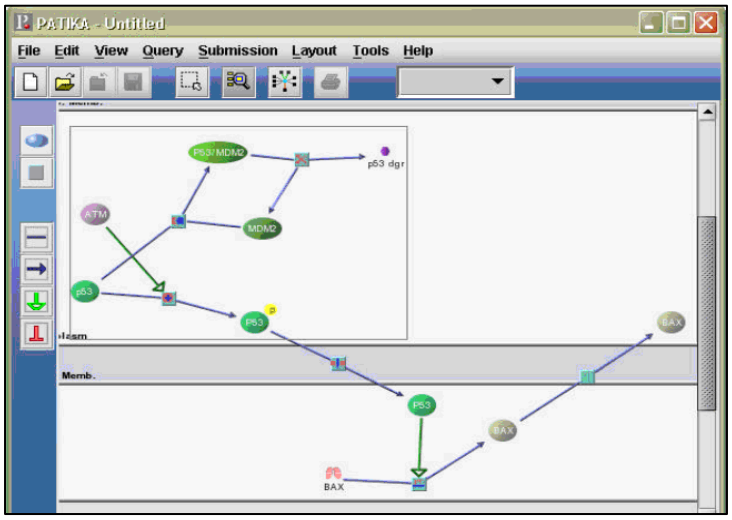


Figure 2.15 Patika [53]. The pathway diagram is divided into different regions to represent different cellular regions, such as nucleus, cytoplasm, etc. The visual properties of nodes indicate their biological properties.

2.8 Research Agenda

Chapter 2 provides a comprehensive list of requirements for pathway visualizations. We also conducted a software survey and heuristic evaluation to analyze how existing pathway visualization tools address user needs. We found that most tools allow users to perform broader data analysis tasks. A serious shortcoming of these tools at present is that they do not provide adequate domain-specific biological meaning, and users must perform many tedious operations to search for and extract relevant information. Unless the tools provide users with rapid biologically-relevant insight that relates the data to the underlying biological meanings (for example, to phenotype), most biologists will be

reluctant to use them. The sections (2.8.1 – 2.8.5) discuss the most important unmet requirements, and a research agenda to address these shortcomings.

2.8.1 Pathway Construction & Update

Biologists use many references to construct the pathways they need. Hence, creating pathways requires a significant time investment. Most biologists pointed out that no matter how valuable the other visualization capabilities, they will not be interested in tools that require them to create large pathways (approximately greater than 100 nodes) from scratch; it is simply too large a time investment, and requires a huge amount of background work to make it meaningful. The tools must be able to construct pathways by retrieving and building on previous relevant pathways. All the biologists in this study showed particular interest in PathwayAssist, because this tool allows users to automatically search for relevant pathway information and periodically update local databases. The biologists felt that this capability could save them a significant amount of time and effort. At the same time, users were very wary about a completely automated pathway builder and wanted some degree of human curation.

2.8.2 Information Overlay

Much information needs to be overlaid on pathway entities. Most tools let users impart various entity attributes by manipulating simple visual properties of nodes and edges. Different graph layouts can help reveal spatial and temporal relationships. Patika visualizations were appreciated by the biologists due to the representation of different states of molecules, along with their spatial cellular locations. BioCarta diagrams were considered most biologically meaningful, and were preferred by biologists over ball-and-stick graphs. None of the visualizations capture the actual complexity of pathway dynamics. For example, STKE provides some animated visualizations to explicitly show sequences of events in a signaling pathway, including movement of biological molecules within the cellular structure. One potential approach for more meaningful visualizations is to represent pathways based on central dogma. Pathway entities can be presented based on their categories such as genes, RNA message, proteins, metabolites, etc.

Defining consistent representations for pathways and entities is needed. Though a large number of pathway visualization systems exist, there is no standardized vocabulary. The green-yellow-red color encoding for gene expression data is one of the few standardized features among these tools (a side effect of microarray imaging technology). This is also true for reference databases and other reference sources. Biologists need to constantly learn new representation styles for visualizations created in different systems. An important research area is to define a consistent language for pathways and their visual representations.

2.8.3 Overlay Data from High-throughput Experiments

The goal of high-throughput data analysis is to infer biological meaning. Biologists must observe high-throughput data within the context of information-rich pathways. In a separate evaluation study of microarray data visualization tools, it was found that the lack of pathway context severely hampered scientists' ability to derive biologically meaningful insight from the microarray data [13]. Further work is needed to effectively combine pathway and microarray visualization tools.

It will be difficult to design visualizations that relate pathway diagrams to quantitative multi-dimensional microarray data, consisting of expression values for potentially multiple treatments and multiple time points. In general, there are several possible design alternatives that must be comparatively evaluated to determine effectiveness:

- Nodes as data glyphs: Most pathway tools will color nodes according to a single microarray treatment (usually the green/red color scale for down/up-regulated).
- Pathway animation: Cycling through several nodes-as-glyphs views over time enables the visualization of a time series. Sliders or other controls can be used to directly navigate the animation loop.
- Small multiples [61] of pathways: Layout several nodes-as-glyphs pathway views in miniature form, likely in a grid of treatments vs. time series.
- Complex node glyphs, or data visualizations embedded within nodes: While nodes-as-glyphs supports only one value per node, embedding small visualizations of microarray data within each node enables the simultaneous display of values for multiple treatments or time points. For example, GScope embeds heatmaps and line charts. Cytoscape has explored the use of radial bars of different lengths around a node [62]. A disadvantage is that these visualizations can become complex and difficult to read.
- Linked pathway and microarray visualizations: Pathway and microarray visualizations can be separated, enabling advanced microarray data visualization methods such as parallel coordinates and clustering (e.g. GeneSpring). The visualizations are interactively linked to enable users to relate nodes to their corresponding microarray data values.

2.8.4 Pathway Overview and Interconnectivity

Most systems list pathway names (as Windows Explorer lists directory names) to let users select a particular pathway of interest. Biologists prefer visualizations that provide an overview of pathways displaying interconnections between them, as in Figure 2.10. Incoming and outgoing visual links could enable users to view how other pathways can potentially affect or be affected by the focus pathway at each node. In a densely populated pathway, it is important to be able to analyze connectivity between components. Simple interactive queries for pathway analysis, such as upstream and downstream components from a node at predefined depths or steps, are considered more useful than having to do this manually. This all suggests highly interactive pathway visualizations [63].

2.8.5 Multi-scale Pathways

As pathways become large and complex, methods such as semantic zooming [64] or hierarchical decomposition [65] are needed to aggregate and abstract entire pathways or pathway portions into small units that can be displayed within larger pathway systems. These aggregates should be simple visual representations that reveal enough information of its contents to enable analysis of the high-level effects. For most applications, pathway visualizations must provide sophisticated multi-scaling to view lower level molecular interactions in the context of higher level physiological changes.

2.9 Conclusions

Thus, though a large number of pathway tools have been developed, the tools that allow researchers to effectively explore large complex biological systems of many integrated pathways are still needed. We believe that pursuit of this research agenda to develop tools that address the requirements listed here will lead to significant improvements in biologists' ability to utilize pathway representations, and facilitate the transition to systems-level science in bioinformatics. The dissertation addresses the requirement: ***Overlay of data from high throughput experiments*** on pathway diagrams. The later chapters describe design space based on the approaches taken by the current tools to address this requirement, and also the user studies conducted to evaluate visualization alternatives from the design space.

A critical component of the pathway research is ***'insight'*****.** Insight enables the biologists to turn experimental data into refined hypothesis. Some data analysis tasks that biologists typically perform include: 1) recognition of changes between experiment and control or between time points; 2) detection of changes in relationship between components of a pathway or between entire pathways; 3) identification of global patterns across a pathway; and 4) mapping pathway state to phenotype (observable effects at the physical level in living organisms) or other biological information. Chapter 3 describes a pilot study conducted to evaluate pathway visualization tools using these common data analysis tasks biologists perform for insight into the data.

3 Visualization for Pathways + Microarray Data

This chapter presents a design space to describe possible alternatives for pathways + microarray data visualization tools. We are focused on pathway diagrams that use a node-link representation. The design space groups visualization tools on two dimensions. Such grouping allows for a more systematic selection of the visualization alternatives to evaluate and relate results to the range of possible options.

A user study to evaluate and rank pathway + microarray data visualization options based on users' performance time and accuracy of responses on predefined tasks is also described. The tasks were selected based on the common biologists' data analysis tasks for pathway + microarray data analysis listed in Chapter 2. Timeseries data was used as an example of multidimensional data. The results suggest that overlaying data on pathway nodes one timepoint at a time may lead to more accurate performance for tasks involving analysis of a graph at a single timepoint, and comparisons between graph nodes for two distinct timepoints. Overlaying data simultaneously for all the timepoints on graph nodes may lead to more accurate and faster performance for tasks involving searching for outlier nodes displaying different behavior as compared to the rest of the graph nodes for all timepoints. Single views have advantage over multiple views on tasks that require topological information.

3.1 Introduction

In bioinformatics, pathways are often used to show how bio-molecules (genes and proteins) interact with each other. Data from high throughput experiments such as gene expression microarrays [7] measure quantity levels of the molecules, and are often analyzed in context of biological graphs. Usually, data is collected for several experimental treatments. An example dataset could be expression values for a viral infection over time. The pathways represent complex biological phenomenon and provide a biological context to otherwise numerical data analysis [12]. In a separate evaluation study, it was found that the lack of graph context severely hampered scientists' ability to derive biologically meaningful insight from microarray data [13]. Figure 3.1 shows overlay of time series data (as an example of multidimensional data) on a pathway. Each node in the pathway corresponds to a tuple row in the dataset, and each experiment treatment is an attribute column.

Some common tasks for analyzing multidimensional data in pathway context for bioinformatics are: What are the values of a specific pathway node in a particular experimental treatment? How do different pathway nodes change over different conditions? Which node displays a particular pattern of behavior across different experimental treatments? How does the behavior of a particular pathway node affect other nodes connected directly or indirectly to it?

A wide variety of pathway visualizations have been created to support analysis of multidimensional data in pathway context [14], and [15]. These visualizations use different approaches to overlay data on pathways. Often the pathway visualizations are linked to other additional visualizations such as parallel co-ordinates and heat maps.

The goal of this chapter is to present a design space for overlaying multidimensional data on pathways, and to comparatively evaluate instances of visualizations within the design space on the common data analysis tasks. This study was performed as a pilot study to investigate if there were differences in the participants' performance based on the different visualization alternatives used. The results from this study will guide further research towards evaluating pathway + microarray data visualization.

3.2 Literature Survey

3.2.1 Visualization for Graph + Multidimensional Data

A large variety of tools that allow analysis of multidimensional data in context of graphs have been created. A survey of different graph visualization tools is presented in [63]. In bioinformatics, a variety of tools use different visualizations to support graph data analysis. GenMapp [4] and PathwayAssist [5] allow overlay of data on graphs using one attribute at a time. The nodes are colored on a user defined scale to represent their values in a particular attribute. Though data is overlaid one attribute at a time in GeneSpring [56], users can link graph visualization to other visualizations such as heat maps, parallel co-ordinate, etc., using brushing and linking. The tools that lay data one attribute at a time on graph vertices usually provide sliders or similar mechanisms to let users iterate over other attributes.

In another approach, more complex glyphs or miniature charts can be embedded in graph vertices. This enables the simultaneous display of values for multiple conditions on the node. For example, GScope [55] embeds heatmaps and line charts on graph vertices. The graph visualizations are linked to a parallel co-ordinate display in GScope. Cytoscape has explored the use of radial bars of different lengths around a node [62] to represent multiple attribute values simultaneously. Visual elements such as images or renderable geometry is used in MoireGraphs [66] to represent various physical entities (e.g., Protein structure, web page, etc). A new focus+context radial layout algorithm along with other interaction techniques assist in exploration of the graphs.

Besides bioinformatics, graph visualizations have been created for other domains too. SeeNet [67] uses static display for spatial information, animation and manipulates different visual properties of vertices and links to represent network data. GraphViz [68] allows users to represent structural information in large number of domains. A few visual properties of nodes can be manipulated to represent different attributes of the nodes. [69] use arc height, grouping and thresholding to visualize topology and properties of Internet's Multicasting Backbone (MBone).

3.2.2 Evaluation for Graph Visualization

A number of studies have been performed to evaluate different graph layout algorithms. E.g., a study to measure cognitive cost of graph aesthetics for the task of finding shortest paths in spring layout algorithm is described in [20]. An evaluation to access readability of two graph representations: matrix based and node-link based is described in [19]. The evaluation was based on seven generic tasks and provides recommendations regarding graph representation based on their size and density. A framework for defining and validating metrics to measure difference between two drawings of the same graph is presented in [17]. The paper also presents experimental analysis on several simple

metrics. Several ideas to define similarity for comparisons between two graph drawings are presented in [16] and evaluated in a user study. A formal metrics based on seven common aesthetics criteria, applicable to any graph drawing of any size are presented in [70]. An analysis of graph drawings produced by some common layout algorithms (e.g., spring layout algorithm, DAG, etc.) based on the seven metric formulae is also presented to demonstrate the application of the metrics. A comparison of hyperbolic tree browser and conventional browser is described in [18]. The users finished their tasks faster with the hyperbolic tree browser in presence of strong information scent.

Thus, though a wide range of studies have been performed to analyze graph drawings and layouts, little work has been conducted to evaluate visualization of multidimensional data associated with graph vertices. The rise of bioinformatics pathways and gene expression analysis has brought this need to the forefront.

3.3 Design Space

The design space is based on identifying design dimensions that will allow us to group visualization tools. All the systems that analyzed for pathways + microarray data in Chapter 2 can be grouped along this space.

3.3.1 Dimension 1: Data Overlay Method

This dimension defines the method to overlay multidimensional data on pathways. The pathway maintains it node-link structure. The three possible alternatives are:

Single Attribute (using single glyphs): In this approach a visual property of nodes is manipulated (usually color) to overlay a single data attribute (Figure 3.1). Cycling through several views for other attributes enables visualization of multidimensional data. Sliders or other controls are often used to directly navigate the animation loop. This design strategy focuses on the display of 1 data attribute at a time, using simple node glyphs, with interactive access to other attributes.

Small Multiples (graph as glyphs): For this visualization design, multiple repeated views of the graph in miniature form are presented, one view for each attribute [61] (Figure 3.2). This design strategy focuses on separating each data attribute into multiple views of the pathway, still using simple node glyphs.

Multiple Attributes (nodes as glyphs): While colored pathway supports only one value per node, embedding small visualizations of multidimensional data attributes within each node enables the simultaneous display of values for all the attributes. E.g., Gscope uses a heatmap and line graphs (Figure 3.3) to display attribute values of nodes. This design strategy focuses on simultaneously combining all data attributes into a single pathway view, using complex node glyphs.

Figure 3.1 An example of overlaying data one condition at a time using color encoding.

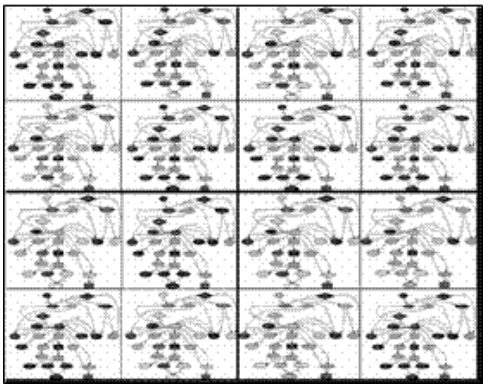


Figure 3.2 An example of laying out multiple graph views in a grid of conditions or treatments in data.

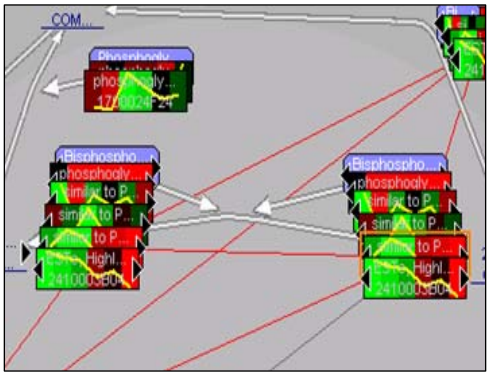


Figure 3.3 An example of embedding multiple data attributes simultaneously within each node.

3.3.2 Dimension 2: Number of Views

Dimension 2 determines if other linked multidimensional data views are used in addition to the graph visualization for data analysis. Each of the pathway visualizations mentioned in Dimension 1 can be linked with other multidimensional visualizations of the data. Thus, for three options in dimension 1, there are three possible options in dimension 2. Using brushing-and-linking approach users can select nodes in the graph to highlight the corresponding data in the multidimensional view, and vice versa.

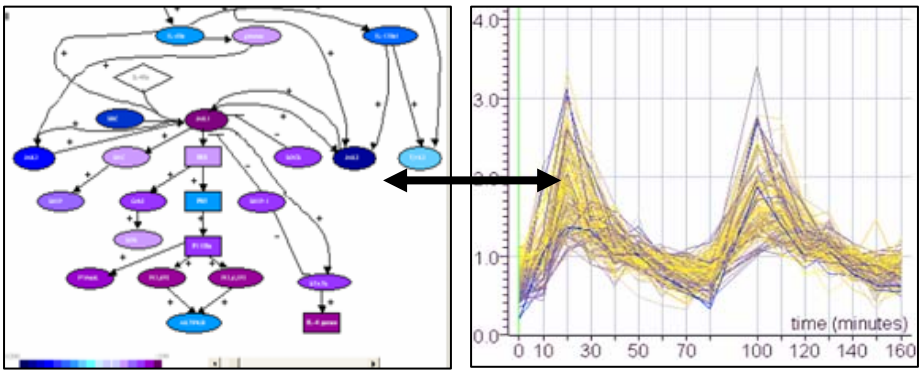


Figure 3.4 Pathway diagrams that overlay data one condition at a time is linked to parallel co-ordinate visualization.

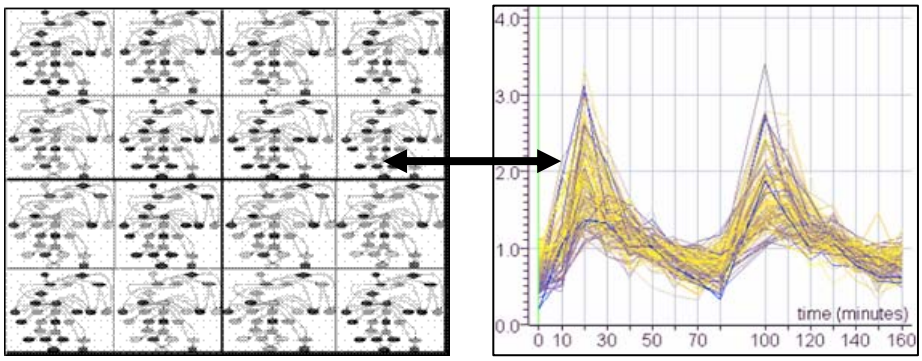


Figure 3.5 Pathway diagrams that use graph as glyph view is linked to parallel co-ordinate visualization.

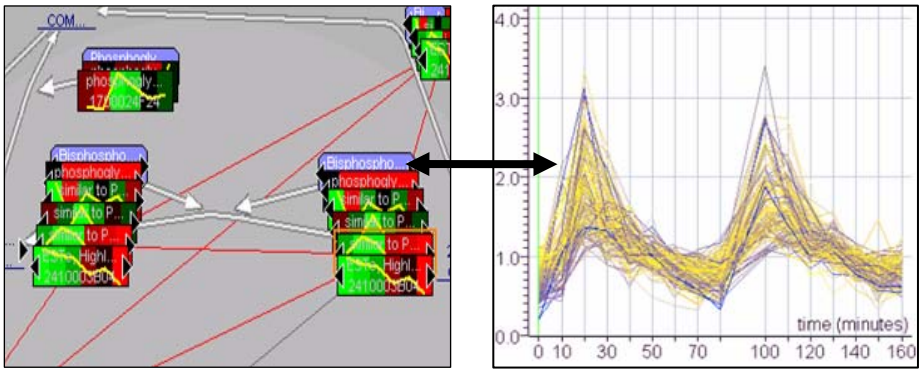


Figure 3.6 Pathway diagrams that use node as glyph view is linked to parallel co-ordinate visualization.

3.4 Pilot Study

Common options were developed for overlaying all timeseries data attributes simultaneously on graph nodes (Figure 3.7). These were evaluated in a pilot study, to select the final version for the main experiment. The alternatives used line graph (A), color (B), and both color + line graph (C) to display values of node in different conditions. We used different intensities of green color to display negative values, yellow for values around zero, and different intensities of orange color for positive values.

Time series data for 10 time points was overlaid on a 50 node directed graph for the study. The visualizations were evaluated between subjects. We had a total of 15 participants, five for each representation. The participants performed predefined tasks described in table 3.4. The tasks were in the form of multiple choice questions. Participants' answers to each task, and the response times were measured. We ranked the visualizations based on the number of correct user responses and shortest time taken to answer.

We observed that participants using color and color + line graphs had more correct responses to the tasks. On an average, participants using just the line graphs had 5.8/11, color had 6.8/11 and color + line graph had 7.2/11 correct answers. On performing ANOVA analysis on performance times we found that participants using color + line graph displays performed faster ($p<0.05$) than participants using line graphs and color only. The average times for all the 11 tasks for the participants were, for line graph: 64.51 sec, color: 54.95 sec, and color + line graph 47.6 sec. Based on these results, we selected color + line graph for the main study.

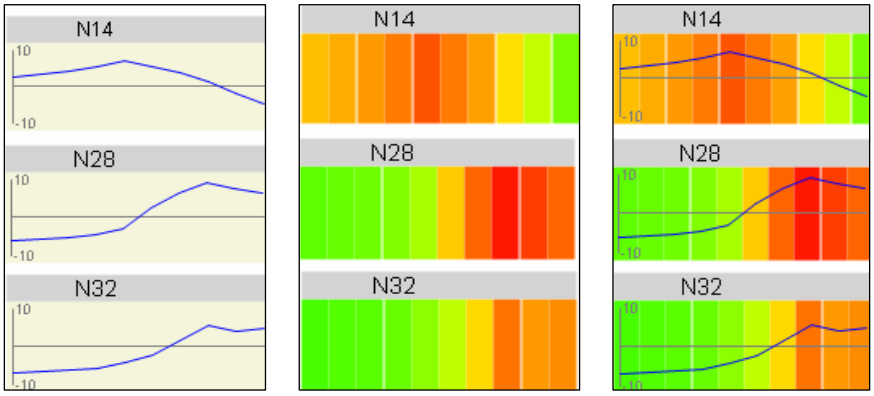


Figure 3.7 Nested visualization alternatives to overlay multidimensional timeseries data simultaneously on graph nodes.

3.5 Experiment Design

The aim of this study is to evaluate alternate visualizations in the design space that support analysis of multidimensional data in context of a pathway. A 2x2 between-subjects design examines the following two independent variables.

1. Data overlay method: Two methods were used, single attribute (simple glyphs with animation), and multiple attributes (complex glyphs in nested visualization).
2. Two choices for use of additional multidimensional view: single view (pathway visualization only), vs. multiple views (pathway visualization + linked parallel coordinates visualization).

The design space identified six different possible design alternatives for pathway + microarray data visualization tools. Since our main focus is on visualizations used in the bioinformatics domain, we selected the option to overlay data using the simple glyph with animation and the nested visualization approach, as these are the two most widely used methods. We used these in both single and multiple view conditions. This would allow us to evaluate the alternatives along both dimensions in the design space. Table 3.1 highlights the portion of the design space selected for evaluation.

Number of Views **Data overlay method**	**Single View**	**Multiple View**
Single Attribute	Yes	Yes
Small Multiples	No	No
Multiple Attributes	Yes	Yes

Table 3.1 Lists the alternatives from design space selected for evaluation.

Most often in bioinformatics, green color is used to show down regulation or negative values, yellow to display values around zero and red for positive values. We preserved this standard color scale for the visualizations in the study. Since we worked with a timeseries data, we linked the graph visualizations to parallel co-ordinate displays for multiple view visualizations.

3.5.1 Visualization Tools

We used four visualizations in the study. Table 3.2 lists visualization alternatives and their interaction features for the experiment. Confirming to the general trend in bioinformatics, we used a color scale from yellow to green for displaying negative values, and yellow to red for displaying positive values. The tools were custom developed for this study to ensure consistency between conditions. For all the visualizations, moving the mouse over a node displayed numerical values corresponding to the color. For both the single attribute visualizations a slider was provided to let users iterate over all the attributes in the data.

Single Attribute + Single View (SS): This visualization overlaid values for one attribute on a node at a time. It was same as in Figure 3.8, but did not have parallel co-ordinates view linked to it.

Single Attribute + Multiple Views (SM): This visualization is shown in Figure 3.9. It was similar to Single Attribute + Single View but was linked to a parallel co-ordinate view using brushing and linking.

Multiple Attribute + Single View (MS): This visualization overlays data from all the attributes on a node using both a heat map and a line graph. It was similar to visualization in Figure 3.10, but did not have a parallel co-ordinate view linked to it.

Multiple Attribute + Multiple Views (MM): Figure 3.11 shows this visualization. It was similar to Multiple Attributes + Single View but was linked to a parallel co-ordinate visualization using brushing and linking

Number of Views **Data Overlay Methods:**	**Single View**	**Multiple Views**
Single Attribute	Slider Mouse over (Figure 3.8)	Slider Brushing Mouse over (Figure 3.9)
Multiple Attributes	Mouse Over (Figure 3.10)	Brushing Mouse over (Figure 3.11)

Table 3.2 Design space and interaction features for visualization tools in the experiment.

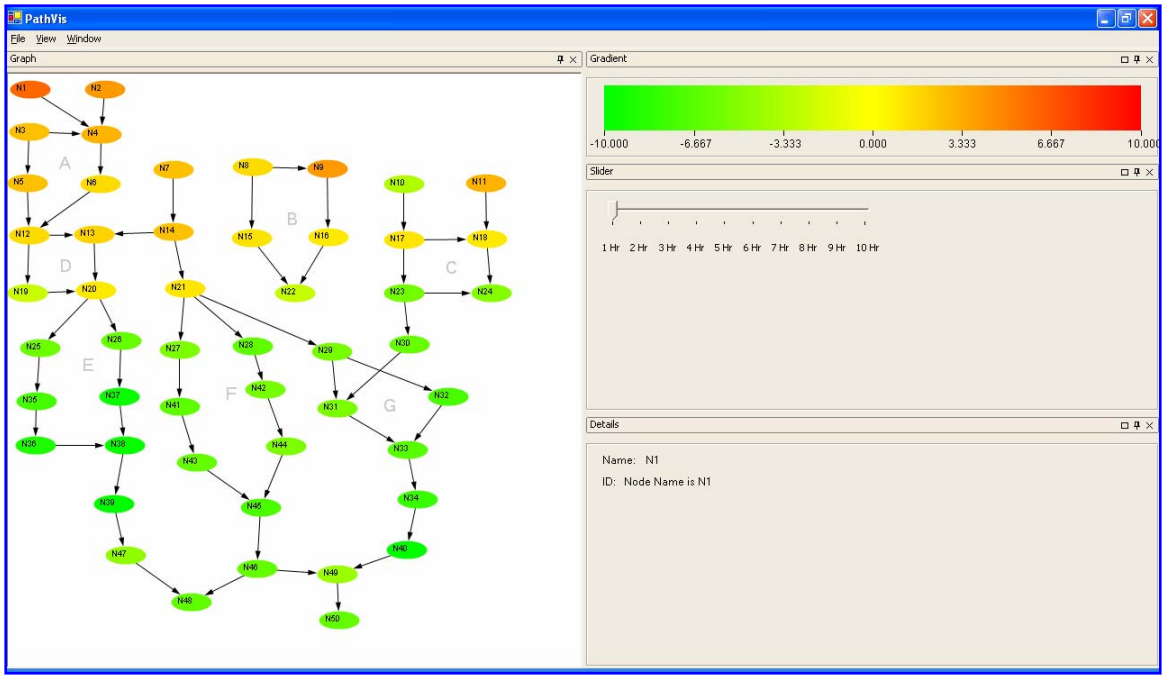


Figure 3.8 Overlay of a single attribute on graph nodes by color, and using single view (SS). A slider enables user to select which attribute to visually overlay on the graph.

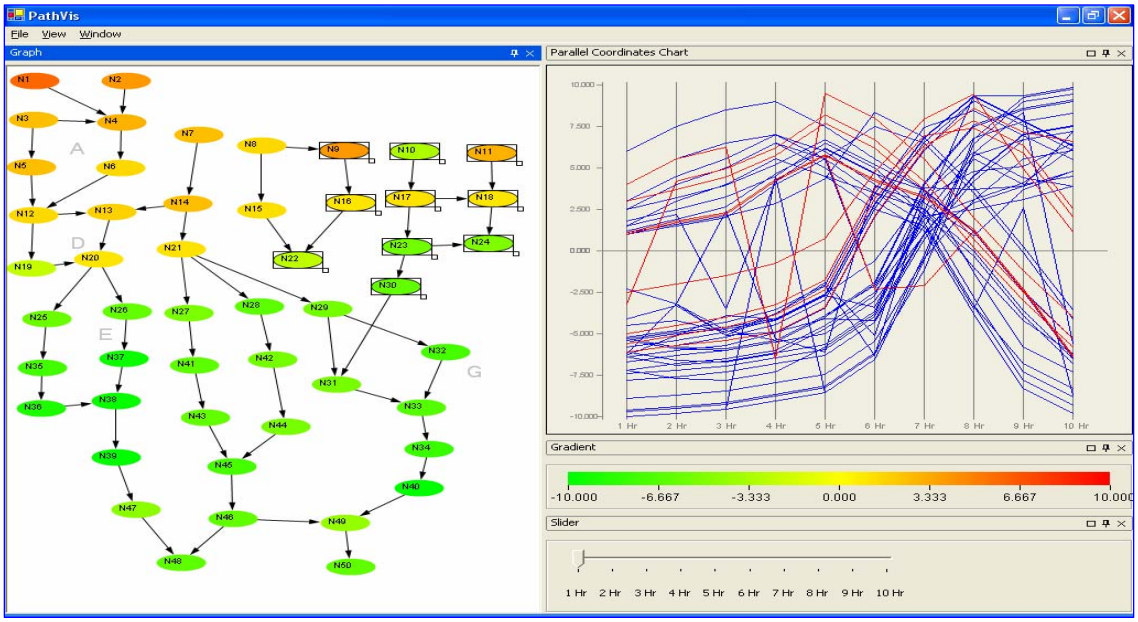


Figure 3.9 Overlay of a single attribute on graph nodes by color, and using multiple views (SM). The graph visualization is linked with parallel co-ordinate visualization using brushing and linking. A slider enables user to select which attribute to visually overlay on the graph.

Figure 3.10 Overlay of multiple attributes on graph nodes by heat maps and line charts, and using single view (MS).

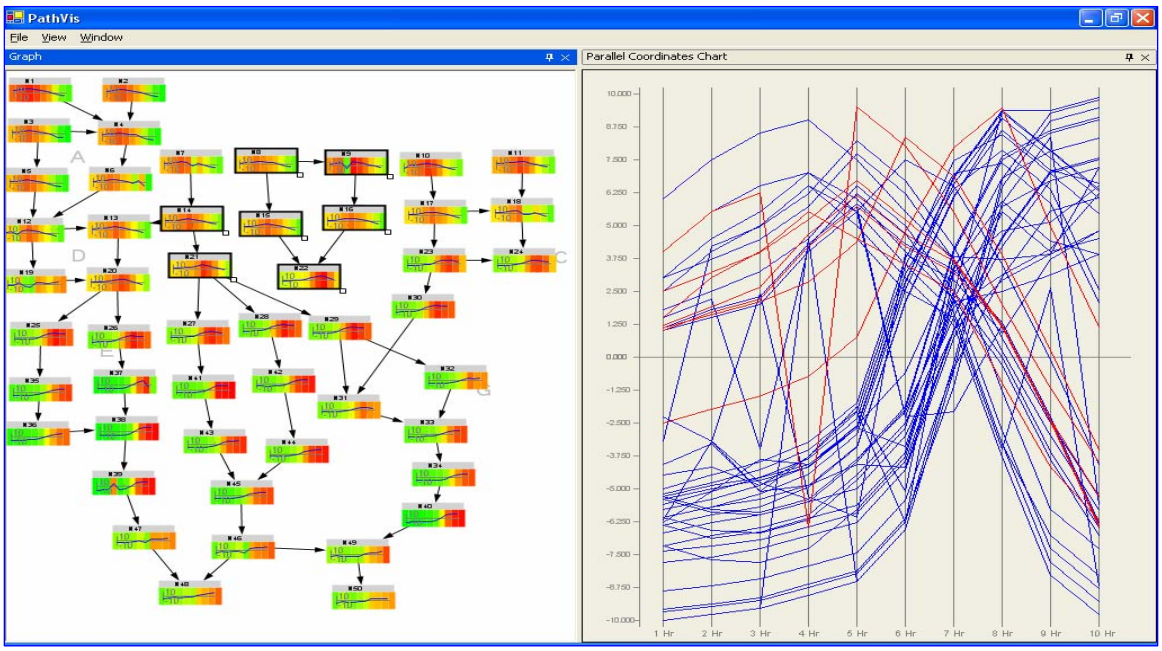


Figure 3.11 Overlay of multiple attributes on graph nodes by heat maps and line charts, and using multiple views (MM). The graph visualization is linked with parallel co-ordinate visualization using brushing and linking.

3.5.2 Data

A pathway having 50 nodes and time series data with 10 time point attributes was used. Some of the nodes in the pathway were grouped together and named by displaying textual information next to them, as shown in Figures 3.8 – 3.11, as is common in bioinformatics pathways. The sizes of the graph and dataset are based on typical needs in bioinformatics. The average size of graphs in the STKE library (www.stke.org) is under 50 nodes. Table 3.3 summarizes the data used for the experiment.

Data Type	Description
Graph	A directed graph having 50 nodes and 56 edges. Each node had an out degree of 0 to 3.
Multi-Dimensional	Time series data, having values for 10 time points for each node.

Table 3.3 Data used for the experiment.

3.5.3 Task List

Participants performed 11 tasks listed in Table 3.4. Tasks are based on common needs in bioinformatics pathway analysis, but abstracted to general graph tasks. Since a time-series data was used for the study, the tasks are more relevant to such type of data. Participants performed four practice tasks to get familiar with the user interface and the visualization after which they were given the actual tasks to perform. All the tasks were described as multiple choice questions, with five possible choices. We recorded responses to the tasks, and time taken by the participants' to perform each task.

T pts.	Nodes	Goal	Task #, Task
1	4	Read value	**1.** What are the values of Nodes C 17-18-23-24 at time point 6?
1	4	Search node	**2.** Which node of the group G 29-31-32-33 is most positive at time point 7?
1	50	Search nodes	**3.** Find a group of 4 nodes, out of which three are positive and one is negative at time point 7?
2	1	Differences	**4.** What is change in N7 from time point 5 to time point 8?
2	4	Differences	**5.** What is change in value of nodes C 17-18-23-24 from time point 6 to time point 8?
10	1	Trend	**6.** How does N8 change over time?
10	3	Topology	**7.** How many time points does it take for N29 and N30 to trigger N40?
10	5	Outlier - node	**8.** Which node is an outlier in the group B 8-9-15-16-22 that displays most different behavior than the others?
10	4	Search –T pt	**9.** At what time point is the value of nodes D 12-13-19-20 most negative?
10	50	Trend Search – nodes	**10.** Find a node that shows a continuous increase up to timepoint 9 and then a sharp decrease.
10	50	Outlier group	**11.** Find a group of nodes that display most different behavior than the rest of the graph over all the time points?

Table 3.4 Lists the tasks used in the study, T pts. = number of timepoints, Nodes = number of nodes required for the task, goal = task type, Task = the task participants performed.

3.5.4 Experiment Protocol

40 participants, 10 for each visualization participated in the experiment. All the participants in the study were freshman or sophomore undergraduate students and business majors. None of the data analysis tasks required specific biological knowledge.

So we did not require participants to have biological background. The participants were given a brief introduction to the visualization and explanation of some basic terminology used to describe tasks in the study. Table 3.5 lists the independent and dependent variables for the study.

Independent Variables	• Tool: data overlay method • Tool: Single vs. multiple views • Task
Dependent Variable	• Time to answer each question • Number of correct responses

Table 3.5 Lists independent and dependent variables for the study.

3.6 Hypothesis

3.6.1 Overall Performance

Single Attribute vs. Multiple Attributes: Since the data for all the timepoints is displayed simultaneously on the graphs nodes for the multiple attribute visualization, we thought that this would make it easier for the participants to analyze changes in the graph over time. Hence, participants using multiple attribute visualization alternatives may complete the tasks faster than the participants using single attribute visualization.

Single View vs. Multiple Views: We believed that participants using multiple view visualization may perform faster, as they had an advantage of an additional view over the participants using single views especially for tasks that required analysis for all the 10 timepoints.

Accuracy: We also believed that since all the visualizations were simple and easy to understand there may be no significant overall differences in the participants' performance on accuracy.

Based on above assumptions participants using multiple attributes + multiple views should display faster performance on the pre-selected tasks as compared to the other participants using other visualization alternatives. Most of the hypotheses were based on the general guidelines for single view vs. multiple view visualization tools [71], [72] and previously conducted experiments [73].

3.6.2 Task-based

Tasks involving 1 Timepoint

Tasks 1 – 3 (Table 3.4) required analyzing graphs at a single timepoint. We believed that participants using single attribute visualization may perform faster than participants using multiple attribute visualizations, as the single attribute visualization method displays data on the graph just for a single timepoint. This allows the participants to focus on the timepoint of interest. Since the tasks did not require much usage of the additional non-graph visualization, there should be no performance differences between participants using single view vs. multiple views. We did not hypothesize that the visualization alternatives will have effects on the accuracy of participants' responses.

Tasks involving 2 Timepoints

Tasks 4 and 5 (Table 3.4) involved comparing values of node(s) at two different timepoints. Since the multiple attribute visualization method overlays all the values for

the nodes at different timepoints simultaneously we believed that participants using multiple attribute may perform faster than single attribute visualization.

Since task 5 required participants to analyze changes in values for several nodes simultaneously, we believed that participants using multiple view visualization may have an advantage over single view visualization alternatives. We did not have any hypothesis of effect of different visualization types on accuracy of participants' responses.

Tasks involving all 10 Timepoints

Since the tasks 6 – 11 (Table 3.4) required participants to analyze changes for node(s) related to all the 10 timepoints. We believed that the participants using multiple attribute visualization may perform faster than the participants using single attribute visualization. Also, participants using multiple view visualization alternatives may have an advantage of having an additional view over participants using single view visualization alternatives.

3.7 Results

3.7.1 Overall Performance

On performing ANOVA analysis over all tasks, we found that there were significant differences in accuracy of participants' responses based on the overlay method used. Participants using single attribute graph visualizations were significantly more accurate [$F(1, 360) = 1.94$, $p = 0.001$] than participants using multiple attribute graph visualization. There was no interaction between views and attributes on accuracy [$F(1, 360) = 0.24$, $p = 0.625$].

Participants using single view visualizations performed significantly faster [$F(1,360) = 7.10$, $p = 0.011$]) as compared to multiple view visualizations. There was no interaction between views and attributes on performance time [$F(1, 360) = 0.00$, $p = 0.991$] The ANOVA model and analysis are presented in Appendix A. The results are summarized in table 3.6.

Overall Performance	Single Views	Multiple Views
Single Attribute	More accurate Faster	More accurate Slower
Multiple Attributes	Less accurate Faster	Less accurate Slower

Table 3.6 Summary from ANOVA analysis for overall performance and accuracy between participants for all the tasks.

The results were further strengthened on performing ANOVA analysis between visualization options. We found that participants using single attribute + single view visualization were more accurate than participants using multiple attribute + single view visualization [$F(1, 218) = 6.96$, $p = 0.008$], and participants using single attribute + multiple view visualization performed significantly more accurate than the participants using multiple attribute + multiple view visualization [$F(1, 218) = 12.18$, $p = 0.000584$]. There were no differences in the accuracy of responses between participants using single attribute + single view and single attribute + multiple views ($p=0.87$), and participants

using multiple attribute + single views and participants using multiple attribute + multiple views (p=0.64).

Similarly for performance times: participants using single attribute + single view were faster than participants using single attribute + multiple views [F(1, 218) = 7.73, p =0.00589], and participants using multiple attribute + single view were faster than participants using multiple attribute + multiple views [F(1, 218) = 8.36, p =0.004]. There was not much performance difference between participants using single attribute + single views, and multiple attribute + single views and participants using single attribute + multiple views and participants using multiple attribute + multiple views. The results are summarized in the table 3.7.

Overall Performance	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	68*	69*	50	46
Average Time per task (in sec)	51*	66	47*	62

Table 3.7 Lists average time in seconds for each task, and percentage of correct responses for all the four visualization options. * indicates better performance.

3.7.2 Tasks Involving 1 Timepoint

On performing ANOVAs we found that participants using Single attribute visualizations were somewhat more accurate than participants using multiple attribute visualization tools [F(1, 116) = 3.45, p = 0.065]. There was no interaction between views and node attributes for the accuracy [F(1, 116) = 1.24, p = 0.267]. Since the overall analysis combines multiple task types, a deeper analysis broken down by task type is warranted. There were no performance differences between participants on accuracy for tasks 1 and 3, however, participants using single attribute were significantly more accurate [F(1, 36) = 10.56, p=0.0025] than participants using multiple attributes on task 2. The data analysis is presented in Appendix A.1.3.1.

On performing ANOVAs we found that participants using single view visualization were significantly faster [F(1, 116) = 12.17, p=0.006] than participants using multiple view visualization tools. There was no interaction between views and attributes on participants' task performance time [F(1,116) = 1.74, p =0.189]. There were no significant time performance differences on task 1 and task 2. However, participants using single view visualizations were significantly faster than participants using multiple view visualization on task 3 [F(1, 36) = 5.64, p = 0.022]. The results are summarized in tables 3.8 and 3.9. Table 3.10 summarizes the results on individual task analysis.

T1 – T3	Single View	Multiple Views
Single Attribute	Faster	Slower
Multiple Attribute	Faster	Slower

Table 3.8 Summary of ANOVA analyses for overall accuracy and performance time differences between the participants on tasks tasks 1 – 3.

Overall Performance	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	73*	60*	46	53

Average Time per task (in sec)	45*	81	42*	69

Table 3.9 Lists average time in seconds for the tasks, and percentage of correct responses for all the visualization options. * indicates better performance.

Task	Description	Factor
T2	Which node of the group G 29-31-32-33 is most positive at time point 7?	Accuracy: Single attributes
T3	Find a group of 4 nodes, out of which three are positive and one is negative at time point 7?	Performance Time: Single view

Table 3.10 Summarizing results from Individual task analysis.

3.7.3 Tasks Involving 2 Timepoints

For both the tasks T4 and T5 (Table 3.12), ANOVA analysis, participants using single attribute performed were significantly more accurate than participants using multiple attribute visualizations [$F(1,76) = 5.21$, $p=0.025$] whereas on both the tasks, participants using multiple attribute visualizations were faster[$F(1,76) = 4.00$, ($p=0.008$)] than single attribute displays. There was no interaction between attributes and visualization [$F(1,76) = 0.379$, $p = 0.539$]. Table 3.11 summarizes these results.

T4 – T5	Single View	Multiple Views
Single Attribute	More accurate Slower	More accurate Slower
Multiple Attribute	Less accurate Faster	Less accurate Faster

Table 3.11 Summary of ANOVA analyses for overall accuracy and performance differences between the participants on tasks 4 and 5.

Overall Performance	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	85*	90*	60	50
Average Time per task (in sec)	56	64	45	50

Table 3.12 Lists average time in seconds for the tasks, and percentage of correct responses for all the visualization options. * indicates better performance.

3.7.4 Tasks Involving all 10 Timepoints

For tasks (T6 – T11) involving all the 10 time points, on performing ANOVA analysis, participants using single attribute graph visualizations were more accurate than participants using multiple attribute visualizations [$F(1, 236) = 10.82$, $p = 0.0011$]. There was no interaction between attributes and views for accuracy [$F(1, 236) = 0.43$, $p = 0.511$]. There were no significant differences on the performance time.

Also, participants using single view visualization were faster than participants using multiple view visualization. Also, participants using multiple attribute visualization were faster than participants using single attribute visualization ($p < 0.05$) Table 3.13 summarizes these results.

T6 – T11	Single View	Multiple View
Single Attribute	More accurate	More accurate
Multiple Attribute	Less accurate	Less accurate

Table 3.13 Summary of ANOVA analyses for overall accuracy and performance differences between the participants on tasks 6 -11.

Both T6 and T10 (Table 3.14) required analyzing a node behavior over 10 time points. For both the tasks there were no differences for the accuracy of results. Also, participants using multiple view visualization were faster than participants using single attribute visualization. Also, participants using single attribute displays were somewhat more accurate than participants using multiple attribute displays. Table 3.14 summarizes these results. On both the tasks T6 and T10 participants using multiple views performed faster than participants using single views. For T6, the performance difference was [$F(1, 36) = 3.124$, $p=0.08$].

T6	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	80	100	80	70
Average Time per task (in sec)	43	36	59	48
T10				
% Accurate Responses	80	80	40	50
Average Time per task (in sec)	38	32	64	53

Table 3.14 Lists average time in seconds for the tasks T6 and T11, and percentage of correct responses for all the visualization options. * indicates better performance.

On T7, that required searching for the number of time points involving topological information, we found that single attribute displays were better than multiple attribute displays in terms of accuracy [$F(1, 36) = 4.891$, $(p < 0.05)$]. Tables 3.15 and 3.16 summarize these results.

T7	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	90*	80*	50	60
Average Time per task (in sec)	32	56	45	55

Table 3.15 Lists average time in seconds for the tasks T7, and percentage of correct responses for all the visualization options. * indicates better performance.

Thus for task 7, the following table summarizes the results:

Task	Description	Factor
T7	How many time points does it take for N29 and N30 to trigger N40?	Accuracy: Single attributes

Table 3.16 Summarizing results from Individual task analysis for T7.

On the most complex tasks, T8 and T11 (Table 3.17), that required searching for a node showing different behavior than the rest of the graph, participants using multiple attribute views were faster [$F(1, 36) = 4.98$, $p = 0.035$] than the participants using single attribute views. Also participants using multiple attribute visualization were more accurate than participants using single attribute visualizations on task 8 [$F(1,36) = 5.56$, $p = 0.039$], and somewhat more accurate on task 11 [$F(1,36) = 4.0186$, $p = 0.07$].

T8	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	40	40	75*	70*
Average Time per task (in sec)	67	81	54*	67*
T11				
% Accurate Responses	35	40	60	65
Average Time per task (in sec)	47	65	38	46

Table 3.17 Lists average time in seconds for the tasks T8 and T11, and percentage of correct responses for all the visualization options. * indicates better performance.

Task	Description	Factor
T8	Which node is an outlier in the group B 8-9-15-16-22 that displays most different behavior than the others?	Accuracy: Multiple attributes Performance time: Multiple attributes

Table 3.18 Summarizing results from Individual task analysis for tasks T8 and T11.

For task 9 (Table 3.19), though participants using single attribute display were more accurate than participants using multiple attribute displays [$F(1,36)$, $p=0.04$], participants using multiple attribute displays were faster than the participants using single attribute display ($p=0.1$). These results are summarized in tables 3.19 and 3.20.

T9	Single Attribute Single View	Single Attribute Multiple Views	Multiple Attribute Single View	Multiple Attribute Multiple Views
% Accurate Responses	70*	90*	60	40
Average Time per task (in sec)	82	71	49	48

Table 3.19 Lists average time in seconds for the task T9, and percentage of correct responses for all the visualization options. * indicates better performance.

Task	Description	Factor
T9	At what time point is the value of nodes D 12-13-19-20 most negative?	Accuracy: Single attributes

Table 3.20 Summarizing results from Individual task analysis for task T9.

3.8 Hypothesis vs. Results

3.8.1 Overall Performance

We had assumed that participants using Multiple Attribute + Multiple View visualization may turn out to be fastest. However, from results we found that participants using Single Attribute + Single View visualization outperformed the other visualization methods in terms of both performance time and accuracy. The comparison of the hypothesis vs. results for the participants' overall performance for 11 tasks is listed in Table 3.21.

Visualization Method	Hypothesis	Results
Single vs. Multiple Attributes	Multiple attribute will be faster than the single attribute	No differences
Single vs. Multiple Views	Multiple views will be faster than the single views	Single views were found to be faster than the multiple views
Accuracy	There will be no differences in the participants' performance on accuracy	Single attributes were found to be more accurate than the multiple attribute.

Table 3.21 Comparing results with hypothesis for the overall performance on all the 11 predefined tasks.

3.8.2 Tasks Involving 1 Timepoint

We believed that participants using single attribute + single view visualization would perform faster than the other participants. However, we found that participants using single view visualization performed faster than the participants using multiple view visualization alternatives. Also, we had assumed that there may not be any differences on accuracy of participants' responses for the visualization alternatives. We found that participants using single attribute visualization method were more accurate on task 2. Table 3.22 summarizes the comparisons between hypothesis and results for tasks 1 -3.

Tasks	Hypothesis	Results
T1 – T3	For tasks 1 – 3 we had believed that participants using single attribute visualization may perform faster than participants using multiple attribute visualization methods. Also, since the tasks required analysis for 1 timepoint only we believed that there will be no benefits of having additional views. Thus, there may be no differences between participants using single view vs. multiple views.	We found that number of visualization alternatives rather than the number of attributes displayed on the nodes had effect on participants' performance time. We found that participants using single view visualization were faster than the participants using multiple view visualization.
T2	We believed that there will be no significant differences between the participants performance on accuracy.	We found for task 2 that participants using single attribute were more accurate than participants using multiple attribute visualization methods.

Table 3.22 Comparing results with hypothesis for participants' performance on tasks 1 – 3.

3.8.3 Tasks Involving 2 Timepoints

We hypothesized that participants using multiple attribute visualization may be faster than the participants using single attribute visualization alternatives. This was indeed found on performing data analysis. We also found that participants using single attribute

visualization alternatives were more accurate than the participants using multiple attribute visualization alternatives. The comparisons between hypothesis and results for tasks 4 – 5 are listed in Table 3.23.

Tasks	Hypothesis	Results
T4 – T5	We assumed that participants using multiple attribute visualization alternatives may be faster than the participants using single attribute visualization methods. Also, we assumed that participants using Multiple attribute + multiple view may perform faster than the participants using other visualization alternatives. We had assumed that there will be no differences in participants' accuracy of responses between visualization types.	We indeed found that participants using multiple attribute visualization alternatives were faster than the participants using single attribute visualization alternative. However, we did not find any significant differences in performance times between multiple view vs. single view visualization alternatives. Participants using multiple attribute + multiple view visualization alternative were slower than the participants using multiple attribute + single view visualization alternative. However, we found that participants using single attribute visualization alternatives were more accurate than the participants using multiple attribute visualization alternatives.

Table 3.23 Comparing results with hypothesis for participants' performance on tasks 4 & 5.

3.8.4 Tasks Involving all 10 Timepoints

Tasks	Hypothesis	Results
T6 – T11	Since these tasks required participants to perform analyses for all the 10 timepoints, we believed that participants using multiple attribute visualization method may complete these tasks faster than the participants using single attribute visualization method. We did not have any hypothesis regarding accuracy of participants' responses.	We found that participants using single attribute visualization were more accurate than the participants using multiple attribute visualization alternatives. We did not find any significant differences in overall performance time between participants using single attribute visualization vs. multiple attribute visualization.

Table 3.24 Comparing results with hypothesis for participants' performance on tasks 6 – 11.

Thus, we found that participants using single attribute visualization were more accurate than the participants using multiple attribute visualization. Also the results did not confirm our hypothesis that the participants using multiple attribute visualization method would be faster than the participants using single attribute visualization. However, there were some tasks (T8 and T11) on which participants using multiple attribute visualization had faster performance than the participants using single attribute visualization.

3.9 Summary of Results

We conducted a study to measure performance of participants on predefined tasks for graph visualizations that used different options to overlay data on the nodes. We had earlier hypothesized that participants using multiple views may complete the tasks faster as they had an added advantage of using an additional visualization. However, the results failed to support this. For most tasks single views were faster than the multiple views.

We had hypothesized that the participants using multiple attribute visualization may perform tasks faster as compared to single attribute visualization. The results suggested that the multiple attribute visualizations are indeed faster than single attribute visualizations for tasks that require comparisons of node values between two timepoints, searching for outlier nodes that display different behavior than most other nodes in the graph, and searching for a timepoint at which a node displays particular behavior.

Finally, we believed that there will be no differences in terms of accuracy between participants using different visualization alternatives. However, the results suggested that for most tasks single attribute visualization alternative proved to be more accurate than the multiple attribute alternative. Tables 3.25 and 3.26 summarize the data analysis.

Single Attribute	Multiple Attribute
+ More accurate for single time point analysis. + More accurate for comparisons between two time points. + More accurate for analyzing behavior of a single node for all the time points. + More accurate for searching graph requiring topological information + More accurate for searching a timepoint for which a node shows a particular behavior.	+ Faster results for comparisons between two timepoints. + More accurate and faster performance for searching graph for outlier nodes i.e. nodes or group of nodes that display different behavior than the other nodes + Faster performance for searching a timepoint at which a node shows a particular behavior.

Table 3.25 Results for Single vs. Multiple Attribute visualization based on the data analysis.

Single View	Multiple Views
+ Faster graph analysis at a single time point + Faster for searching a node requiring topological information. + Faster performance for searching graph for outlier nodes i.e. nodes or group of nodes that display different behavior than the other nodes	+ Faster performance for analyzing behavior of a single node for all the time points. + Faster performance on searching for a node/group of nodes that displays a particular behavior + Faster performance for analyzing values for multiple nodes at one or more time points.

Table 3.26 Results for Single vs. Multiple view visualization based on the data analysis.

3.10 Discussion of Results

The most interesting finding of the study is that the number of attributes displayed on the nodes has more influence on accuracy of user responses, whereas the number of visualizations affects the performance time. However, as can be inferred from the results, visualizations should be designed based on which data analysis tasks need to be supported.

Most participants in the study were non-technical (freshman or sophomore business majors) and unfamiliar with graph terminology. Though, once given an explanation they understood the visualizations and graph terms used in the tasks. The participants were given a more thorough explanation of the graph than the parallel co-ordinate visualization. Since the participants were novice users, they were also not experienced with performing data analysis on multiple views simultaneously. Also, the data used for analysis in the study was fairly straightforward, wherein almost all the nodes in the graph followed a regular pattern except a few. Many of the tasks for the study could be performed using just the graph visualization, eliminating the necessity for using parallel co-ordinates. Due to these reasons, it is likely that the experimental design biased the overall results towards single views.

We also noticed that the participants using multiple views performed most of the tasks in the graph visualization, and used the additional parallel coordinate view for confirming their results. Perhaps having more noisy data where graph nodes did not follow a regular pattern would have required participants to utilize both the visualizations. Also, giving participants a longer training period on brushing and linking might have been helpful for them to better utilize the reverse brushing direction in which the parallel coordinate view is used to query the graph view. Despite these concerns, we noticed that multiple views were utilized by participants to analyze behavior of nodes over all the time points, mainly as a read-only view. It also helped participants to compare behavior of a group of nodes simultaneously.

Graph visualizations that overlaid data by a single attribute at a time were most helpful to analyze graphs at a particular time point. The reason being this visualization technique lets users focus just on a particular timepoint of interest. These views are also helpful on search tasks that require topological information. The graph visualization using multiple attributes can get cluttered due to the amount of information being visualized simultaneously. This may make interpretation of topology of a graph more difficult. We found that the graph visualization with multiple attributes needs an interaction mechanism to select and highlight a single timepoint across all the nodes, somewhat analogous to the slider's behavior in the single-attribute version.

Displaying multiple attributes on nodes leads to better performance for tasks that requires searching graphs for outlier nodes, i.e., nodes that display most different behavior than most other nodes in the graph. This option lets users visualize behavior of nodes at all the time points simultaneously, making it easier to pick the nodes that are outliers.

For tasks that involved comparing graph nodes between two time points we found that graph visualization that overlaid data for multiple attributes simultaneously on nodes were faster than visualizations that overlaid data just one time point at a time. However single time point displays were more accurate. This may be due to the fact that though mousing-over nodes in both the graph visualizations displayed values, they didn't display the time point's label (attribute name). More accurate results may have been possible if the mouse-over tooltip in multiple attributes displayed both the value and the timepoint label.

The study under discussion was influenced for the data analysis needs in the bioinformatics domain. The choice of color scale (green – yellow – red), number of graph nodes, visual representations were based on the data representation typically used by the life scientists. But the need to associate time series data with graph representations is common in other domains (computer networks, communications, etc). The data analysis tasks though influenced by pathway analysis requirements in Chapter 2, were generalized enough to be applicable for other types of graph analysis too. However, more niche visualization representations (the color scale, number of nodes used) for a particular domain may cause different results. The users were not tested for green – red color blindness.

The data used for this study was time series. The data analysis requirements for time series data are different than for categorical data or multi-categorical data. Hence, though we can use the results as an initial guide to design visualizations for other datasets for similar tasks to time series data, unless a study is conducted for tasks with respect to a

particular dataset we cannot accurately generalize these results to other datasets. Also, the participants for the study were non experienced data analysts. It is possible that a different trend of results is observed with more experienced users.

3.11 Limitations of the Task-Based Method

The study reported in this chapter evaluated the selected pathway visualization alternatives using pre-selected benchmark tasks. The use of such controlled studies with benchmark tasks is one of the primary methods for evaluating visualization tools. Since the benchmark-task method relies on the pre-selected tasks to evaluate visualization tools, the usefulness of results from the method is dependent primarily on the tasks used in the study. It is important that the pre-selected tasks have definite completion times and responses to reliably measure performance times and accuracy of participants' responses. To uniformly evaluate accuracy of participants' responses, the selected tasks are often specified as multiple choice questions. The multiple responses have to be specified as clearly as possible so that the participants can understand them unambiguously. Because of these the experimenters are often forced to use over simplify the tasks. Such tasks often are not representative of the real world visualization usage.

The task-based study may allow determining if the users can quickly and accurately complete the tasks. However, the method is too limited to provide a broader indication of the different kinds of insight a visualization tool can provide. Conclusions can be made about the visualization tools only about the selected tasks. The results may not be indicative for other tasks that were not selected. Also, providing tasks to the participants may force them into line of thoughts that they may not take otherwise. Even the choice and phrasing of the tasks can bias the participants towards a particular visualization tool. Thus, though benchmark task method can provide a rigorous method to evaluate visualization tools they do not provide reliable methods to evaluate effectiveness of the visualization tools for insight [74]. The task-based method fails to address the open ended nature of biologists' data analysis process. To address these issues, a new evaluation method is presented in Chapter 4 that addresses the limitations of the benchmark task method and is better representative of biologists' real world visualization usage scenarios.

4 Insight-Based Evaluation Method

This chapter presents an insight-based evaluation method. The method provides a way to evaluate and rank bioinformatics visualization tools based on real world data analysis scenarios and addresses the limitations of the task-based method discussed in Chapter 3. The method was used to evaluate microarray data visualization tools rather than pathway + microarray data visualization tools.

The dissertation started by identifying critical requirement for pathway visualization tools. An important requirement identified was the need to associate microarray data to pathway diagrams. Design space to group all the visualization alternatives that allow overlay of microarray data over pathway diagrams was presented in Chapter 3. Though the study reported in Chapter 3 presented interesting results about the pathway + microarray data visualization alternatives it had several limitations. Primarily, it failed to address the open-ended exploratory nature of biologists' data analysis tasks. To evaluate the visualization alternatives based on biologists' real world analysis scenarios a new evaluation method is needed. Since the focus was on defining a new method we worked with microarray data visualization tools. Later chapters use the method in evaluating visualization alternatives that overlay microarray data on pathways and extend the study and results reported in Chapter 3. But for the purpose of this chapter we are more focused primarily on the insight-based evaluation method.

4.1 Introduction

The advent of microarray experiments is causing a shift in the way biologists do research; a shift away from simple reductionist testing on a few variables towards systems-level exploratory analysis of 1000s of variables simultaneously [75]. These experiments result in datasets that are very large. Biologists use these data to infer complex interactions between genes and proteins. Due to its magnitude, it is prohibitively difficult to analyze microarray data without the help of computational methods. Hence, the biologists use various data visualizations to derive domain-relevant insights. The main purpose in using these visualizations is to gain insight into the extremely complex and dynamic functioning of living cells.

In response to these needs, a large number of visualization tools targeted at this domain have been developed [14, 15]. However, in collaborations with biologists, we received mixed feedback and reviews about these tools. With such a wide variety of available options, we need an evaluation method that allows biologists to choose the right tool for their needs. The method should address the open-ended and exploratory nature of the biologists' tasks, and allow us to determine if the tools provide insights valuable to their end users.

A primary purpose of visualization is to generate ***insight*** [76], [77]. The main consideration for any life science researcher is discovery. Arriving at an insight often sparks the critical breakthrough that leads to discovery: suddenly seeing something that previously passed unnoticed, or seeing something familiar in a new light. The primary function of any visualization and analysis tool is to make it easier for an investigator to glean insight, whether from their own data or from external databanks. A measure of an effective visualization can also be its ability to generate unpredicted new insights, beyond

predefined data analysis tasks. After all, visualization should not only enable biologists to find answers but to also find questions that identify new hypotheses. We sought to evaluate a few popular microarray data visualization tools, such as Spotfire® [78]. Some research questions we addressed are: How successful are these tools in assisting the biologists in arriving at domain-relevant insights? How do various visualization techniques affect users' perception of data? How does user's background affect the tool usage? How do visualizations support hypothesis generation and suggest directions for future investigation? Most importantly, can insight be measured in a controlled experimental setting, uniformly across a group of participants? Our primary focus here is on insight**.**

Typically, visualization evaluations have previously focused on controlled measurements of user performance and accuracy on predetermined tasks [79, 80]. However, to answer these research questions requires an evaluation methodology that better addresses the needs of the bioinformatics data analysis scenario. Hence, we developed an evaluation protocol that focuses on recognition and quantification of insights gained from actual exploratory use of visualizations. This chapter presents a detailed explanation and discussion of the methodology, as well as detailed results of applying the method to bioinformatics visualizations.

4.2 Survey of Evaluation Research

4.2.1 Methods to Evaluate Visualization Tools

A variety of evaluation methodologies have been used to measure effectiveness of visualizations. Many studies have evaluated visualization effectiveness through rigorously controlled experiments[79, 80] for summative or scientific hypothesis testing. In these studies, typical independent variables control aspects of the tools, tasks, data, and participant classes. Dependent variables include accuracy and performance measures. Accuracy measures include precision, error rates, number of correct and incorrect responses, whereas performance includes measures of time to complete predefined benchmark tasks. Such studies compare effectiveness of two or more tools (e.g. [81] compares three different visualization systems), or examine human visual perception (e.g. [82] compares graphical mappings of information).

Formative usability tests typically evaluate visualizations to identify and solve user interface problems. A typical method for usability studies involves observing participants as they perform designated tasks, using a 'think aloud' protocol. Evaluators note the usability incidents that may suggest incorrect use of the interface, and compare results against a predefined usability specification [83]. Refer to [84] for an example of a professional formative usability study of a visualization.

Analytic evaluations include inspections of user interfaces by experts, such as with heuristics [29]. Examples of specific metrics for visualizations include expressiveness and effectiveness criteria [85], data density and data/ink [61], criteria for representation and interaction [86], high-level design guidelines [87], principles based on pre-attentive processing and perceptual independence [88], and rules for effectiveness of various visual properties [89]. Cognitive models, such as CAEVA [90], can be used to simulate visualization usage and thereby examine the low-level effects of various visualization techniques.

A longitudinal study of information visualization adoption by data analysts [91] suggests advantages when visualizations are used as complementary products rather than stand alone products. Rieman [92] examines users' long-term exploratory learning of new user interfaces, with 'eureka reports' to record learning events. An insight-based study to evaluate microarray data visualization using more realistic exploratory data analysis is reported in [93]. Three case studies, and a user survey to evaluate effectiveness of Hierarchical Clustering Explorer (HCE), a visualization tool, are reported in [94]. The authors also compare both the evaluation methods used to measure tool effectiveness based on results they provided about the tool usage.

4.2.2 User Studies in Bioinformatics

Biologists use microarray experiments to answer complex biological research questions. As these experiments result in very large datasets, biologists need computational methods to derive domain-relevant insights. A detailed description of the microarray data analysis process is in [95] and [96]. Since this process is very complicated, considerable research is currently being conducted to search for new and improved methods [97]. Extensive evaluations for raw data normalization and statistical algorithms for data analysis have been conducted. For example, different normalization methods based on data variance and bias are compared in [98] lists a review of statistical methods to discover differentially expressed genes. Case studies describing data analysis procedures using clustering algorithms and suggestions for new and improved methods have been published [99]. A comprehensive list of publications for this area can be obtained from [15].

A large number of information visualization tools targeting this domain have been developed [14, 15], and a number of user studies have also been conducted. A case study using GeneSifter [100] to analyze microarray data is reported in [101]. A survey of biologists' tasks for a general query system is reported in [27]. [102] reports observations from user studies with molecular biologists to identify information needs unmet by the current tools. End user participatory design process is used in [103] to create prototype electronic laboratory notebooks. A combination of end-user interviews, heuristic evaluations and surveys was used to elicit the end user requirements for pathway visualization software [12].

Thus, though there has been significant emphasis placed on improving data analysis techniques for bioinformatics, very few studies have actually been conducted to investigate the analytic process and the use of visualization tools from the end-user's perspective.

4.2.3 Visual Analytics

Visual analytics deals with the capabilities of visualization tools that help users make judgments about the data. It is important to create visualization tools that maximize human capabilities to perceive and understand complex and dynamic data. The research agenda in [104] provides a comprehensive list of key aspects that influence visual analytics, the process by which users gain insight into complex data. For discussion here, the most relevant aspects are: science of analytical reasoning, and visual representations and interaction techniques. A detailed description about each aspect of visual analytics,

along with an extensive literature survey, and suggestions for future research work is presented in [104].

4.3 Pilot Study

The main challenge we faced was precisely defining insight and how to measure it. The word ‘insight’ in ordinary usage is vague and can mean different things to different people. However, for the purpose of our study we needed this term to be quantifiable and reproducible. To examine this, we undertook an initial pilot study to observe how users recognized and categorized information obtained from microarray data using visualization tools with limited training. We used both GeneSpring® [56] and Spotfire® [78] to ascertain that these commercial tools were not too difficult to learn and could be used by novice as well as expert users.

As the pilot experiment was exploratory in nature, we presented no strict protocol as to how users ought to proceed. We recruited five subjects at our institute to participate. As our recruits had no prior experience using these particular tools, we reduced their initial learning time by offering a brief introduction to the tool they would use along with a summary of the different visualization techniques provided by the tool. Users were encouraged to think aloud and report any findings they had about the dataset. Pilot participants were supplied two datasets to work with, a table containing fake data that contained information about just ten genes, and the Lupus dataset used in the final experiment. We selected the smaller dataset to help users become familiar with the visualization techniques. Once comfortable with using the visualization tool, users were instructed to move onto the Lupus data.

Due to the volume and rapidity of observations reported, we concluded that we needed to record any future sessions on videotape. We also discovered that the users grew weary analyzing the practice dataset, despite being told that it was just a learning aid. They tended to spend too much time on it and, by the time they began looking at actual data, they were already fatigued. We found that our test subjects could learn a visualization technique just as quickly from real data, hence, we decided to use only the real data for final experiments. From the users’ comments we recognized various quantifiable characteristics of ‘insight’.

4.3.1 Insight Characteristics

To measure insights gained from visualization, a rigorous definition and coding scheme is required. We recognized in the pilot that we could capture and characterize specific individual insights as they occurred in participants’ open-ended data analysis process. This provided more detailed information about the insight capabilities of the tools than subjective measures from post-experiment surveys.

We define an *insight* as an individual observation about the data by the participant, a unit of discovery. It is straightforward to recognize insight occurrences in a think-aloud protocol as any data observation that the user mentions. The following quantifiable characteristics of each insight can then be encoded for analysis. We applied this scheme in the main experiment. Although we present them here in the context of biological and microarray data, we believe that this can be applied to other data domains as well. The characteristics of each insight are:

Observation: The actual finding about the data. We counted distinct data observations by each participant.
Time: The amount of time taken to reach the insight. Initial training time is not included.
Domain Value: The value, importance, or significance of the insight. Simple observations such as "Gene A is high in experiment B" are fairly trivial; whereas, more global observations of a biological pattern such as "deletion of the viral NS1 gene causes a major change in genes relating to cytokine expression" are more valuable. The domain value is coded on a scale of 1 to 5 by a biology expert familiar with the results of the data. In general, trivial observations earn 1-2 points, insights about a particular process earn an intermediate value of 3, and insights that confirm, deny, or create a hypothesis earn 4 or 5 points.
Hypotheses: Some insights lead users to identify a new biologically-relevant hypothesis and direction of research. These are most critical because they suggest an in-depth data understanding, relationship to biology, and inference. They lead biologists toward 'continuing the feedback loop' of the experimental process, in which data analysis feeds back into design of the next experimental iteration [36].
Directed vs. Unexpected: Directed insights answer specific questions that users want to answer. Unexpected insights are additional exploratory or serendipitous discoveries that were not specifically being searched for. This distinction is recognized by asking participants to identify specific questions they want to explore about the dataset at the beginning of the trial.
Correctness: Some insights are incorrect observations that result from misinterpreting the visualization. This is coded by an expert biologist and visualization expert together.
Breadth vs. Depth: Breadth insights present an overview of biological processes, but not much detail; e.g., "there is a general trend of increasing variation in the gene expression patterns". Depth insights are more focused and detailed; e.g., "gene A mirrors the up-down pattern of gene B, but is shifted in time". This also is coded by a domain expert.
Category: Insights are grouped into four main categories: overview (overall distributions of gene expression), patterns (identification or comparison across data attributes), groups (identification or comparison of groups of genes), and details (focused information about specific genes). These common categories were identified from the pilot experiment results after insights were collected.

4.4 Experiment Design

The aim of the main study was to evaluate five popular bioinformatics visualization tools in terms of the *insight* that they provide to the users. A 3x5 between-subjects design examined these two independent variables:

1. Microarray datasets, 3 treatments:
    - Timeseries dataset – 5 time-points
    - Virus dataset (Categorical) – 3 viral strains
    - Lupus dataset (Multi-categorical) - 42 healthy, 48 patients
2. Microarray visualization tool, 5 treatments:
    - Clusterview
    - TimeSearcher
    - HCE
    - Spotfire®

- GeneSpring®

4.4.1 Microarray Datasets

To examine a range of data scenarios, we used data from three common types of microarray experiments. The datasets are all quantitative, multi-dimensional data. Values represent a gene's measured activity level (or gene expression) with respect to a control condition. Hence, higher (lower) values indicate an increased (decreased) gene activity level. Since our study is focused on the interactive visualization portion of data analysis, the datasets were preprocessed, normalized, pre-filtered, and converted to the required formats (as discussed in [6, 8]) in advance. In general, the biologists' goal is to identify and understand the complex interactions among the genes and conditions, essentially to reverse engineer the genetic code. The following three datasets were used.

Time-series Dataset: Users were given an unpublished dataset from Karen Duca's lab [105]. HEK293 cells, a human embryonic kidney cell line, were infected with the A/WSN/33 strain of influenza virus *in vitro* at an MOI of 5. At defined time points across the entire viral replication cycle *in vitro*, mRNA was extracted from infected and mock-infected cultures. The values in the columns were the $\log_2$ of the normalized ratios of experimental signal to control signal. The dataset used for analysis had 1060 rows (genes) over 5 time points. Two additional columns represent the gene name and standard ID.

GeneName	**GenBank**	**1.5 Hr**	**4 hr**	**6 Hr**	**8 Hr**	**12 Hr**
aquaporin 4	AA00100	1.54	-0.21	1.49	-0.12	0.96
…	…	…	…	…	…	…

Table 4.1 Time-series dataset used in the experiment.

Viral Dataset: Part of a published dataset from Michael Katze's lab [106] was given to users. A549 cells, a human lung epithelial cell line, were infected with one of three influenza viruses in vitro (wild type A/PR/8/34, recombinant strain of PR8 with the NS1 partially deleted, called NS1 (1-126), recombinant strain derived from PR8 with the NS1 gene completely deleted, called delNS). Other than in the NS1 gene, all three viruses are identical. At 8 hours post infection, mRNA was extracted from infected and mock-infected cultures. The dataset used for analysis had 3 columns (representing the 3 viral conditions) and 861 rows (genes). Two additional columns represent the gene name and standard ID.

Name	**Description**	**wt PR8**	**NS1 (1-126)**	**delNS1**
ADCY9	adenylate-cyclase-	0.54	0.91	5.8
…	…	…	…	…

Table 4.2 Viral dataset used in the experiment.

Lupus dataset: Participants were presented a subset of published data from Timothy Behren's lab [107]. In this study, after blood draw, peripheral blood mononuclear cells (PBMCs), comprising monocytes/macrophages, B and T lymphocytes, and NK cells, were isolated from control and Systemic Lupus Erythematosus (SLE) samples. mRNA was harvested for expression profiling using Affymetrix technology [108]. The column

values represented expression values (average difference or AD) for each gene. Scaling was performed to allow comparison between chips. The dataset had 90 columns (consisting of gene expression from 48 SLE samples and 42 healthy control samples) and 170 rows (genes). Two additional columns represent the gene name and standard ID.

Accession #	**Gene**	**Ctrl 1**	**…**	**Ctrl 42**	**SLE 1**	**…**	**SLE 48**
AB008775	Aquaporin 9	-63.7	…	100.1	4418.1	…	3433.2
…	…	…	…	…	…	…	…

Table 4.3 Lupus dataset used in the experiment.

4.4.2 Visualization Tools

For practical reasons, we limited this study to five microarray visualization tools. We chose the tools based on their popularity and availability. We attempted to select a set of tools that would span a broad range of analytical and visual capabilities and techniques. Cluster/Treeview (Clusterview) [99], TimeSearcher [109], and Hierarchical Clustering Explorer (HCE) [110] are free tools, while Spotfire® and GeneSpring® are commercial. Table 4 summarizes the visualization and interaction techniques supported by each tool.

Clusterview (Figure 4.1) uses a heat-map visualization for both data overview and details. A compressed heat-map provides an overview of all values in the dataset, in row-column format. Users can select a part of the overview to study in more detail. It is standard practice in bioinformatics to visually encode increased gene-expression values with a red brightness scale, decreased gene-expression values with a green brightness scale, and no-change as black. As a slight variation, some tools use a continuous red-yellow-green scale with yellow in the no-change region.

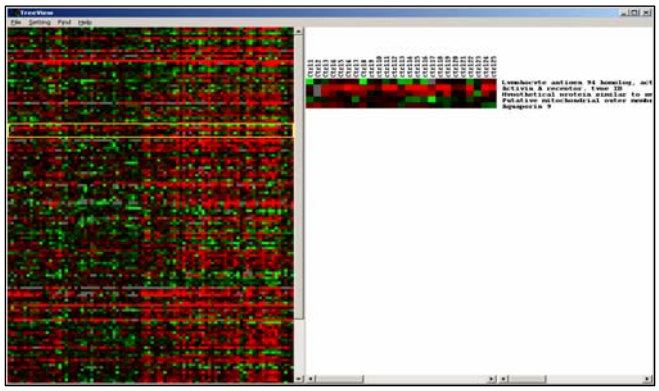


Figure 4.1 Clusterview [99] visualization of the Lupus dataset.

TimeSearcher (Figure 4.2) introduces a new concept of time-boxes to query a set of entities with temporal attributes. The visualization used for data overview is a time series display of all the data attributes. Line graphs and detailed information are also provided for each individual data entity. The views are tightly coupled using the concept of interactive ‘brushing and linking’; selecting a gene in one view highlights it in all views.

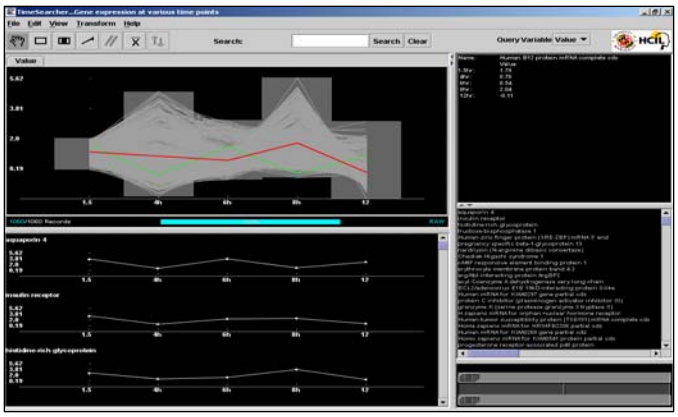


Figure 4.2 TimeSearcher [109] visualization of Time-series dataset.

HCE (Figure 4.3) provides several different visualizations: scatter plots, histograms, heat maps, and parallel-coordinates. HCE's primary display uses dendrogram visualizations to present hierarchical clustering results. This clusters similar data items near each other in the tree display. HCE also provides histograms and scatter plots for data analysis. In a multidimensional dataset, the number of scatterplots possible is large. HCE introduces a new concept of 'rank by feature' [111] to allow users to quickly find interesting histograms and scatterplots, although this feature was not available for this study. The visualizations are tightly coupled for interactive brushing users can manipulate various properties of the visualizations and also zoom into areas of interest.

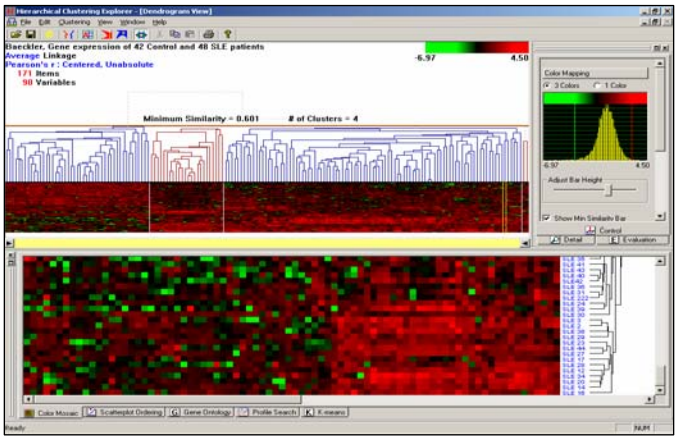


Figure 4.3 HCE [110] visualization of the Viral dataset.

Spotfire® (Figure 4.4) offers a wide range of visualizations: scatter plots, bar graphs, histograms, line charts, pie charts, parallel coordinates, heat maps, and spreadsheet views. Spotfire® presents clustering results in multiple views, placing each cluster in a separate parallel coordinate view. The visualizations are linked for brushing. Selecting data items in any view, shows feedback in a common detail window. Users can zoom, pan, define data ranges, and customize visualizations. The fundamental interaction technique in Spotfire® is the dynamic query sliders, which interactively filter data in all views.

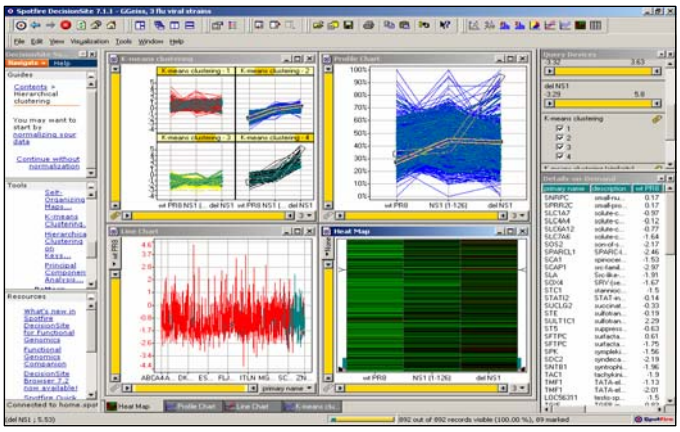


Figure 4.4 Spotfire® [78] visualization of the Viral dataset.

GeneSpring® (Figure 4.5) provides the largest variety of visualizations for microarray data analysis: parallel coordinates, heat-maps, scatter plots, histograms, bar charts, block views, physical position on genomes, array layouts, pathways, ontologies, spreadsheet views, and gene-to-gene comparison. As we did not have information such as position of genes on chromosome, and organization of gene clones on microarray chip for all the experiments, we could not use some of the visualizations, such as physical position and array layout views provided by GeneSpring®. The visualizations are linked for brushing. Users can manipulate the visualizations in several ways e.g., zooming, customizing visualizations by changing the color, range, etc. GeneSpring® also includes data clustering capabilities.

Figure 4.5 GeneSpring® [56] visualization.

Tool	**Visual Representations**	**Interactions**
Cluster/ Treeview	Heat-map, Clustered heat-map	O+D
Time-Searcher	Parallel coordinates, line graph	Brushing, O+D, DQ
HCE	Cluster dendrogram, parallel coordinates, heat-map, scatterplot, histogram	Brushing, Zooming, O+D, DQ

Spotfire® 7.2 Functional Genomics	Parallel coordinates, heat-map, scatterplots (2D/3D), histogram, bar/pie chart, tree view, spreadsheet view, Clustered parallel coordinates	Brushing, Zooming, O+D, DQ
GeneSpring® 5.0	Parallel coordinate, heat-map, scatterplots (2D/3D), histogram, bar chart, block view, physical position view, array layout view, pathway view, spreadsheet view, compare gene to gene, Clusterested parallel coordinates	Brushing, Zooming

Table 4.4 Summarizes the visualization and interaction techniques supported by each visualization tool. O+D = overview+details; DQ= dynamic queries.

4.4.3 Participants

Thirty test subjects volunteered from the university community. We allotted six users per tool, with two per dataset per tool. We required all users to have earned at least a Bachelor's degree in a biological field and be familiar with microarray concepts. To prevent undue advantage and to measure learning time, we assigned users to a tool that they had never used before. Users were randomized within this constraint. Based on their profiles, the users fit into one of three categories summarized in Table 4.5.

Category	**Participant Background**	**N**
Domain Expert	Senior researchers with extensive experience in microarray experiments and microarray data analysis. Possess a Ph.D. in a biological field.	10
Domain Novice	Lab technicians or graduate student research assistants, having an M.S. or B.S. in a biological field. Some experience with microarray data analysis.	11
Software Developers	Professionals who implement microarray software tools. Have an M.S. in a biological field and also M.S. in computer science.	9

Table 4.5 Summarizes the number of participants (N) and their backgrounds.

4.4.4 Protocol and Measures

To evaluate the visualization tools in terms of their ability to generate insight, a new protocol and set of measures is used that combines elements of the controlled experiment and usability testing methodologies. This approach seeks to identify individual insight occurrences as well as overall amount of learning while participants analyze data in an open-ended think-aloud format. No benchmark tasks were assigned. Also, we decided to focus on new users of the tools with only minimal tool training. We have found that success in the initial usage period of a tool is critical for tool adoption by biologists.

Each user was assigned one dataset and one tool. Before starting their analysis, users were given a background description about the dataset. To reduce initial learning time, the users were given a brief 15-minute tutorial about the primary visualization and interaction techniques of the tool. Users then listed some analysis questions they would typically ask about such a dataset. Then, they were instructed to continue to examine the data with the tool until they felt that they would not gain any additional insight. The entire session was videotaped for later analysis. Users were allowed to ask the administrator about using the tool if they could not understand a feature. The training in this protocol was intended to simulate how biologists often learn to use new tools from their colleagues.

While they were working, users were asked to comment on their observations, inferences and conclusions. Approximately every 10-15 minutes, users were asked to estimate how much of the total potential insight they felt they had obtained so far about the data, on a scale of 0–100%. When they felt they were finished, users were asked to

assess their overall experience with the tool, including any difficulties or benefits. Later, we analyzed the videotapes to identify and codify all individual occurrences of insights. Table 4.6 summarizes the dependent variables.

1	User's initial questions about the dataset
2	Total time spent with the tool
3	Amount learned (as a percentage), periodic and final
4	List of insights and characteristics
5	Visualization techniques used
6	Usability issues
7	Participant demographics

Table 4.6 Dependent variables for the study.

4.5 Results

Results are presented in terms of users' data questions, insights, visualization usage, and user background.

4.5.1 Initial Questions

At the start of each session, users were requested to formulate questions about the data that they expected the visualization tool to answer (Table 4.7). Almost all the users wanted to know how the gene expression changed and its statistical significance with each experimental condition, different expression patterns, and obtain pathway information and known literature for the genes of interest. More biologically specific questions focused on location of genes of interest on chromosomes and pathways. They said that it would be valuable to know what pathways show correlations.

There were collectively 31 distinct questions for all the datasets. It was not possible to answer some of the questions during the experiment, due to insufficient data. e.g. the Lupus dataset did not have information about disease severity or patient demographics as would be required for questions 23 and 26 in Table 4.7. Nor did the datasets include pathway information for questions 4, 7, 15, 18, and 30 listed in Table 4.7. However, GeneSpring® (31/31) and Spotfire® (27/31) can potentially address most of the questions posed by the participants, if adequate data were provided. Clusterview (11/31), TimeSearcher (14/31), and HCE (15/31) can answer more specific subsets of the questions.

	Information participants wanted from the data	**Num.**
Questions for Time series dataset		
1	Change in overall expression with time	10/10
2	Different patterns of expression	10/10
3	Genes that responded early to a treatment and were later followed by other genes	5/10
4	Functional details of genes showing high change	2/10
5	Genes showing similar expression pattern to a specific gene of interest	1/10
6	Relate change in gene expressions to physiological changes in the cells	1/10
7	Pathway information for genes having similar expression patterns	2/10
8	Relate gene expression to chromosome position	1/10
9	Retrieve known information for selected genes	10/10
Questions for Viral dataset		
10	Difference in overall expression for three viruses	10/10
11	Genes that show similar/different behavior to the experimental hypothesis	3/10

12	Expression patterns different from the hypothesis	3/10
13	Genes with high or low expression for each virus	10/10
14	Different patterns of gene expression	10/10
15	Pathway information for genes of interest	3/10
16	Correlations between different pathways	3/10
17	Chromosom location of genes with similar change	3/10
18	Functional information of selected genes	1/10
19	Statistical significance in changes between different viral strains	1/10
Questions for Lupus dataset		
20	Difference in expression between 2 groups	10/10
21	Statistical significance of difference between groups	3/10
22	Different patterns of gene expression	10/10
23	Relate expressions to severity of disease	1/10
24	The range of gene expression for each group	1/10
25	Statistical significance of variability of expression for genes in each group	4/10
26	In case of variability, if this is based on patients' age, sex, race, etc.	1/10
27	Analyses such as genes that show more than 50% increase from control to lupus patients	1/10
28	A list of housekeeping genes to evaluate experiment results	1/10
29	Patient characteristics such as those who used some drug vs. those who did not use any drug, males vs. females etc.	1/10
30	Behavior of Immune pathway genes	2/10
31	Calculate average expression for each group	6/10

Table 4.7 List of data questions asked by the participants.

4.5.2 Insight Characteristics

Listed here are the measured results for each insight characteristic described earlier, aggregated by visualization tool. Since this evaluation method is more qualitative and subjective than quantitative, and the number of participants is limited, general comparison of tendencies in the results is most appropriate (Figure 4.6 and Table 4.8). However, we include some statistical analysis that provides useful indicators.

Appendix B lists overall (count of insights, domain value, time to first insight, final amount learned, and total time spent with the tool) for each participant along with the type of dataset they worked with. The characteristics hypothesis, directed vs. unexpected, incorrect and breadth vs. depth were determined based on analysis of each insight by the domain expert.

Insights: We counted the total number of insights, i.e. distinct observations about the data by each participant. Participants who analyzed the same dataset with a particular tool reported very similar insights about the data. Thus, the reported insights were repetitive across participants. As shown in Figure 6, the count of insights was highest for Spotfire® and lowest for HCE.

Total Domain Value: The sum of the domain value of all the insight occurrences. Insight value was highest for Spotfire®. Participants using Spotfire® gained significantly more insight value than with GeneSpring® [$F(1, 10) = 6.92$, $p<0.05$]. Though, numeric value was lowest for HCE, there were no significant differences between Spotfire® or other tools and HCE due to high variance in the performance of HCE users on different datasets as explained in 4.5.4.

Time: The following two temporal characteristics (average time to first insight and average total time) summarize the time to acquire insights:

Average Time to First Insight: The average time into the session, in minutes, of the first insight occurrence of each participant. Lower times suggest that users are able to get immersed in the data more quickly, and thus may indicate a faster tool learning time. The participants using Clusterview took a very short time to reach first insight. TimeSearcher and Spotfire® were also fairly quick to first insight, while HCE and GeneSpring® took twice as long on average. Clusterview users took significantly less time [$F(4, 25) = 4.87$, $p<0.01$) to reach the first insight than the other users, while GeneSpring® took the longest.

Average Total Time: The average total time users spent using the tool until they felt they could gain no more insight. Lower times indicate a more efficient tool, or possibly that users gave up on the tool due to lack of further insight. In general, Clusterview users finished quickly while GeneSpring® users took twice as long. The participants using Cluterview took significantly less time as compared to the other users [$F(4 , 25) = 9.3$, $p <0.01$].

Average Final Amount Learned: The average of users' final stated estimate of their amount learned. The amount learned is a percentage of total potential insight, as perceived by users. In contrast to other insight characteristics reported, this metric gauges users' belief about insight gained, and about how much the tool is or is not enabling them to discover. Spotfire® users were most confident in their perceived insight. The similarity between this metric and total domain value might indicate that users are fairly accurate in their assessment.

Hypotheses: Only a few insights led users to new biological hypotheses (Table 4.8). These insights are most vital because they suggest future areas of research and result in real scientific contributions. For example, one user commented that parts of the time series data showed a regular cyclic behavior. He searched for genes that showed similar behavior at earlier time points, but could not find any. He offered several alternative explanations for this behavior related to immune system regulation, and said that it would compel him to perform follow-up experiments to attempt to isolate this interesting periodicity in the data. For the viral dataset, two users commented that there were two patterns of gene expression that showed negative correlation. They inquired whether this means that the transcription factors of these genes have inhibitory or stimulatory effects on each other. They said that they wanted more information about the functions and pathways to which these genes belong, to better relate the data to biological meaning. Spotfire® resulted in one hypothesis for each dataset, thus a total of three. Clusterview also led users to a hypothesis for the Viral and Lupus datasets.

Directed vs. Unexpected Insights: The participants using HCE with the Viral dataset noticed several facts about the data that were completely unrelated to their initial list of questions. Clusterview provided a few unexpected insights from the Lupus dataset, and TimeSearcher provided unexpected insights about the time series data. Spotfire® had one each for time series and Lupus.

Incorrect Insights (Correctness): HCE proved helpful to users working with the viral dataset. However, users working with the time series or Lupus datasets did not gain much insight from the data. When prompted to report their data findings, they stated some observations about the data that were incorrect. The two users that reported incorrect insights were in the domain expert and software developer categories. The errors may have been due to inferring the color scale backwards, or due to misinterpreting the way

that HCE reorders the rows and columns of the heat map by hierarchical clustering. None of the other tools resulted in incorrect findings.

Breadth vs. Depth: Though we had initially thought this to be an interesting criterion, on data analysis we found that most user comments were of the type 'breadth'. For this experiment, all the users worked with a visualization tool they were not familiar with. It will be difficult for the first time users to learn all the features of both Spotfire® and GeneSpring® within the time span of the experiment. Also, many users were not familiar with the specific genes in the datasets used for the study. We discovered that to get deeper insights into the data, the participants need to be more familiar with the data background. Hence, for the purpose of this study, we did not pursue this characteristic in detail.

	Clusterview	**TimeSearcher**	**HCE**	**Spotfire®**	**GeneSpring®**
Hypotheses	2	1	1	3	0
Unexpected Insights	3	3	5	2	0
Incorrect Insights	0	0	2	0	0

Table 4.8 Summarizes total number of hypotheses generated, unexpected insights, and incorrect insights for each tool.

Together, higher total value and count indicate a more effective tool for providing useful insight. Lower time to first insight indicates a faster learning curve for a tool. Ideally a visualization tool should provide maximum amount of information in shortest possible time.

Overall, Spotfire® resulted in the best general performance, with higher insight levels and rapid insight pace. Clusterview and TimeSearcher appear to specialize in rapid insight generation, but to a limit. With GeneSpring®, users could infer the overall behavior of the data and the patterns of gene expressions. However because the users found the tool complicated to use, most of them were overly consumed with learning the tool rather than analyzing the data. They had difficulty getting beyond simple insights. HCE's strengths will become clear in the next two sections.

4.5.3 Insight per Dataset

This section compares the tools within each dataset.

Time series data: In general, Spotfire® and TimeSearcher performed the best of the 5 tools in this dataset. Participants using Spotfire® and TimeSearcher had insights with more domain value [$F(4, 5) = 8.38$, $p<0.05$] from time series data than the other tools. However, participants using Spotfire® felt they learned more from the data (73%) compared to TimeSearcher (53%). Both Spotfire® and TimeSearcher had nearly equivalent performance in terms of value and number of insights. Time to first insight was slightly lower for TimeSearcher (4 min) as compared to Spotfire® (6 min). At the bottom, participants using HCE took significantly longer [$F(4, 5) = 12.13$, $p<0.01$] to reach the first insight than the other tools. Participants using GeneSpring® took significantly longer ($F(2, 3) = 15.44$, $p<0.05$) than TimeSearcher and Clusterview.

Virus data: HCE proved to be the best tool for this dataset. Participants using HCE had better performance in terms of insight value as compared to other users. However, there were no significant differences between the other users. HCE provided 5 unexpected

insights that were different than the initial information users were searching for in this dataset.

Lupus data: Participants using Clusterview and Spotfire® had more insight value, whereas participants using HCE [$F(4, 5) = 7.26$, $p<0.05$] had the least value as compared to the other tools.

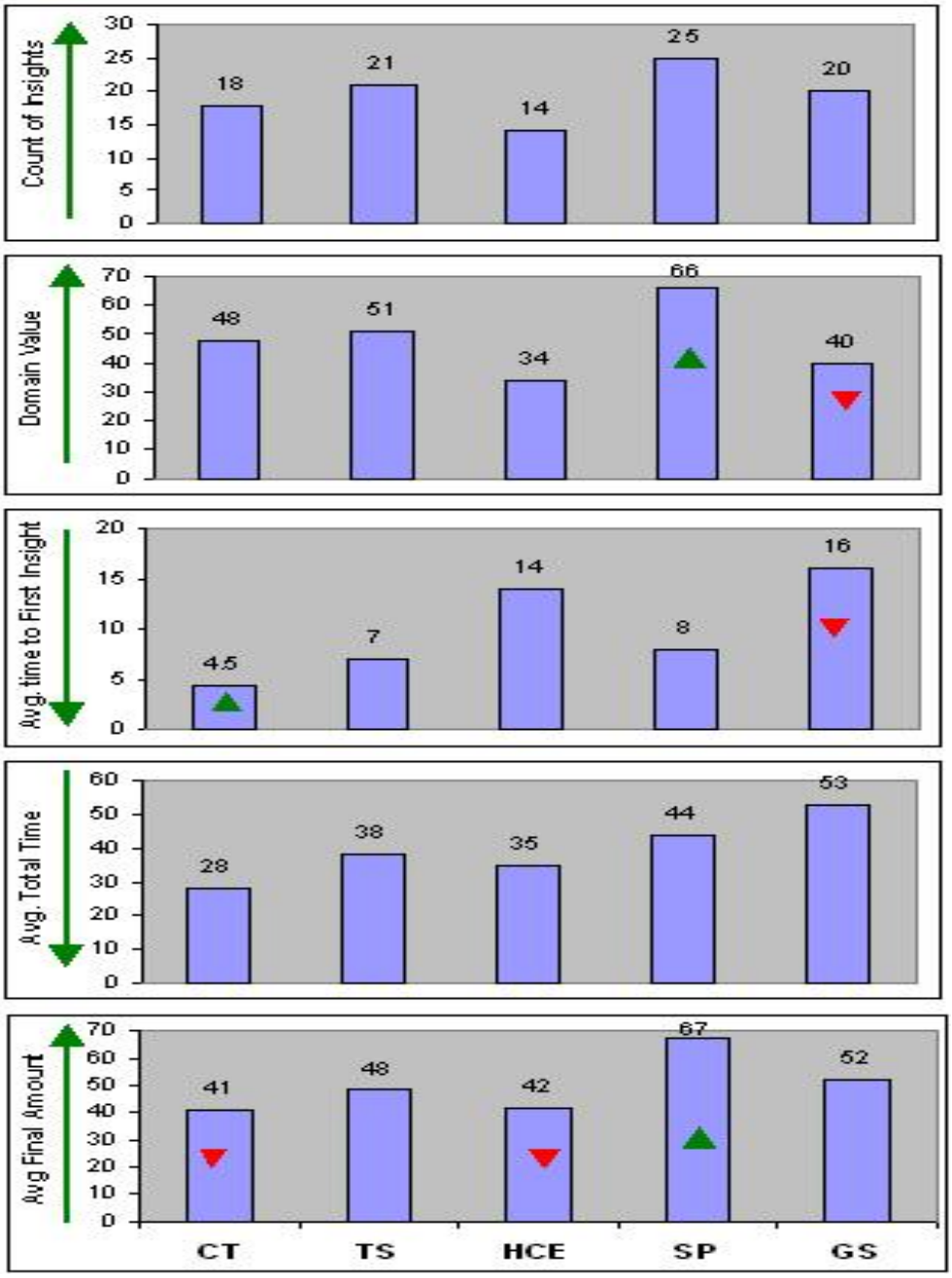


Figure 4.6 Count of insights, total insight domain value, average time to first insight, average total time, and average final amount learned for each tool. ▲/▼ indicates significantly better/worse performance differences. Y-axis arrows indicate direction of better performance.

4.5.4 Tools vs. Datasets

This section examines individual tools across the three datasets. TimeSearcher and HCE had interesting differences among the datasets (Figure 4.7), while the other tools were well rounded.

TimeSearcher: Participants using TimeSearcher performed comparatively best with the time series data as compared to the other two datasets. With time series data, they had over double the value and number of insights than with the Viral and Lupus datasets.

HCE: In contrast, participants using HCE did best on the Viral dataset. On Viral dataset, they had a significant better performance advantage on insight value [$F(2, 3) = 24.8$, $p<0.01$], number of insights [$F(2, 3) = 21.5$, $p<0.05$] and time to first insight [$F(2, 3) = 30.65$, $p<0.05$] as compared to the other datasets. They also felt they learned much more from the data. Participants using Lupus data spent less overall time with the tool [$F(2, 3) = 9.5$, $p \sim 0.05$] as they felt they could not learn much from the data using HCE.

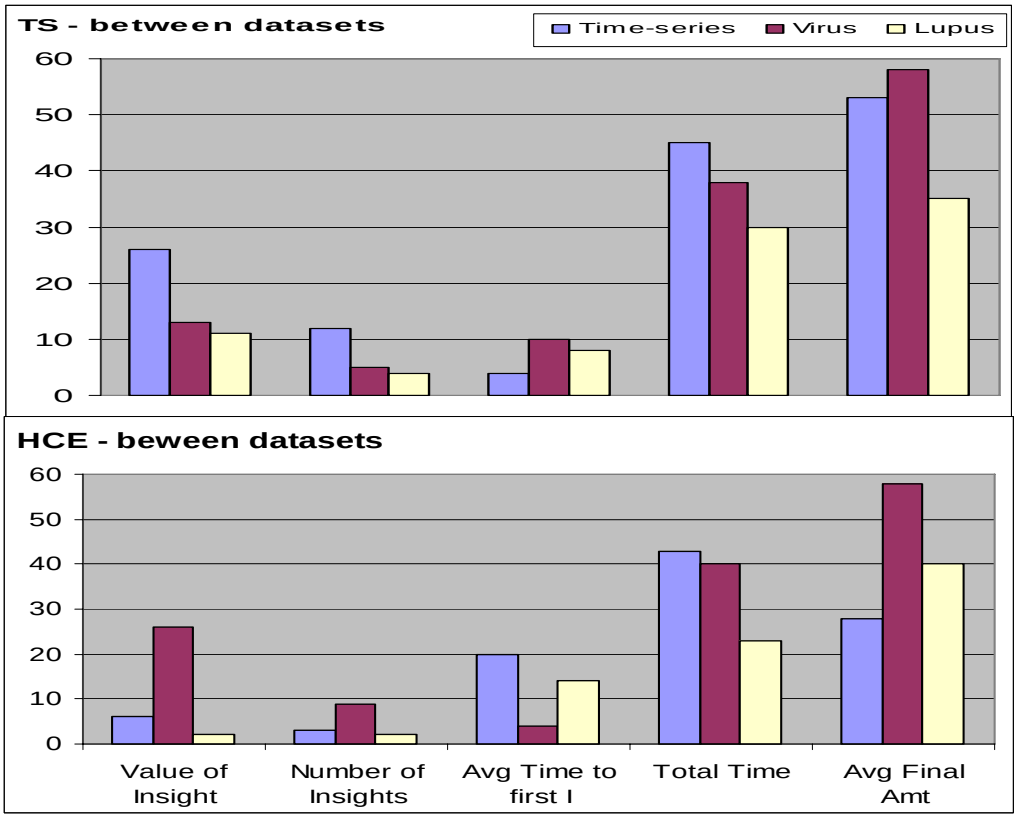


Figure 4.7 TimeSearcher and HCE specialize in the Time-series and Viral data, respectively.

4.5.5 Insight Categories

Though a wide variety of insights were made, most could be categorized into a few basic groups.

Overview: These described and compared overall expression distributions for a particular experimental condition. For example, a user analyzing time series data reported that “at time points 4 and 8 a lot of genes are up regulated, but at time point 6 a lot are down regulated”. Several users analyzing the virus dataset commented that more genes showed a higher expression level for delNS1 virus as compared to wt virus, and the gene

expression seems to be increasing with the deletion. Most users working with the Lupus dataset reported that gene expression for SLE patients appeared higher than the control group.

Expression Patterns: Most users considered the ability to search for patterns of gene expressions very valuable. Most started by using different clustering algorithms (e.g., K-Means, SOMS, Hierarchical Clustering) provided by the tools to extract the primary patterns of expression. They compared genes showing different patterns. For example, some users noted that while most genes showed higher expression value for Lupus group as compared to Control group, there were other genes that were less expressed for the Lupus group. They thought it would be interesting to obtain more information about these genes in terms of their functions and the pathways they belong to.

Groups: Some users, mainly those working with Spotfire® and GeneSpring®, grouped genes based on some criteria. For example, a user working with Spotfire® wanted to know all genes expressed similarly to the gene HSP70. Users working with GeneSpring® used gene ontology categories to group genes. GeneSpring® provides several ways in which users can group their data. They found this functionality very helpful. Also most of the users were very pleased to learn that they could link the biological information, such as gene functions, with the groups.

Detail Information: A few users wanted detailed information about particular genes that were familiar to them. For Time series data, a user noticed about 5% of genes high at 1.5 hr were also high at 12 hr and followed a regular cycle. He looked up the annotations for a few of these genes and tried to obtain more information about them to see if they could be responsible for the cyclic nature of the data.

Category	Clusterview	TimeSearcher	HCE	Spotfire®	Gene-Spring®
Overview	9	10	6	13	5
Patterns	10	8	5	10	8
Groups	0	0	0	1	4
Details	2	3	1	1	1

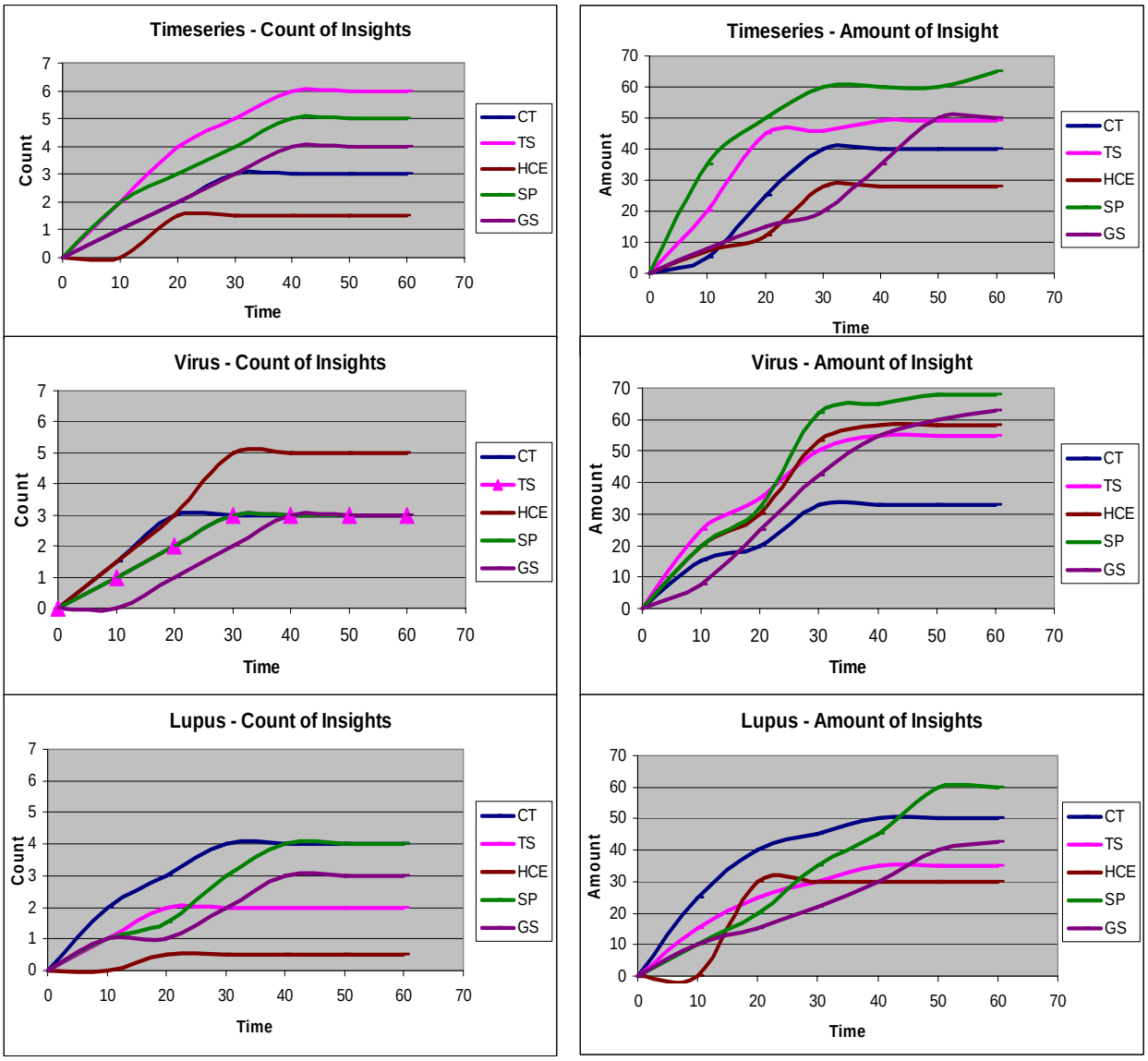
Table 4.9 summarizes the number of each type of insight by tool.

4.5.6 Insight Curves

This approach to measuring insight also enables the examination of how insight accumulates over time. This section shows the insight curves for actual insight counts as well as users' perceived insight amount. These graphs show the rate of insight generation for the tools. Figure 4.8 (A) represents the average accumulation of insight occurrences over time for each tool and dataset. Figure 4.8 (B) shows the users' average estimated percentage of total insight acquired over time. During the course of the experiment, users were asked every 10-15 minutes to report how much they felt they had learned about the data as a percentage of total potential insight.

Some of the tools stand out on certain datasets as providing faster or slower rate of insight, and strengthen finding reported earlier. TimeSearcher and Clusterview provide an early jump in insight on time series and Lupus datasets respectively. While Spotfire® eventually catches up, other tools plateau sooner. HCE rises above other tools in the viral dataset in actual insight count. However, in the other datasets, HCE shows a step-like curve perhaps indicating an initial period of learning the tool, followed by a small number of insights, followed by a plateau and termination by the users.

There is some similarity between Figures 4.8 (A) and (B) for the time series and lupus datasets, in terms of the general shape of curves and order of the tools. This could indicate some relative accuracy of participants' insight estimates. An interesting difference is that, for Spotfire® and GeneSpring®, the users' estimated insight curves continue to rise even after their corresponding curves in actual insight counts plateau. That is, even after they make no new insights, they still felt they were gaining more insight. This may be due to the fact that, after continuing to explore the data in the many different visual representations within these tools, participants became more confident in their findings and felt that they had not missed much after all.



(A) (B)

Figure 4.8 (A): Average number of Insights, over time, for each dataset and tool, (B): Average percentage of total insight gained as periodically estimated by participants, over time, for each dataset and tool.

4.5.7 Visual Representations and Interaction

Spotfire® users preferred the heat-map visual representation, whereas GeneSpring® users preferred the parallel coordinate view. This is despite the fact that both of these tools offer both representations. Most of these users performed the same analyses, but using different views.

Though there were no particular preferences of visualizations for particular the datasets, we noticed that for the Lupus dataset Spotfire® and Clusterview users preferred the heat-map visualization. The heat-map allowed them to group Control and Lupus data neatly into two distinct groups and they could easily infer patterns within and across both groups. Participants using these tools showed a higher performance on these datasets using these visualizations. This finding is strengthened by the fact that both TimeSearcher and GeneSpring® users showed average performance on this dataset. Users of these tools used parallel coordinate visualizations to analyze the datasets.

We noticed that even though tools like Spotfire® and GeneSpring® provide a wide range of visualizations to users, only a few of these were used significantly during the study. Most users preferred visualizations showing outputs of clustering algorithms, such as provided by Clusterview, Spotfire®, and GeneSpring®. These enabled the users to easily see different patterns in the data. However, many said that it would be more helpful to them if the interaction capabilities of this representation were increased, e.g. to better enable comparison of the groups, subdividing, etc.

HCE's primary overview presents the data in a dendogram heat-map that is re-ordered based on the results of hierarchical clustering algorithms. Columns and samples with the most similar expression values are placed near each other. Thus, for both the time series and Lupus datasets, where a particular column arrangement is useful to recognize changes across the experimental conditions, HCE showed poorer performance. Users were not aware of the fact that they could turn off that feature (such customization capabilities of views were not demonstrated in the initial short training session). Also, none of the four users who would have benefited the most from turning off this feature considered the possibility of turning it off, and they did not inquire about it. This turned out to be a critical feature that should be made more prominent in the tool, or in hindsight should be included in the training.

4.5.8 Participants Comments about the Visualization Tools

At the end of each experiment, users were requested to comment on their experience with the tool they used. The following sections summarize users' comments.

Clusterview: Users felt that the tool was extremely simple to use. Some users (3/6) required a brief explanation of the heat-map view of the data. The users felt that the information provided by Clusterview is very basic, and they will need to perform additional analysis with other methods to get further information from the data. The users who worked with time series data commented that the heat map was not a very efficient way to represent data and they preferred visualizations similar to parallel-coordinates.

TimeSearcher: Feedback on TimeSearcher varied for different datasets. The users found the parallel-coordinate overview provided by TimeSearcher was easy to understand. Users working with the time series data found the tool very helpful. They were able to easily identify trends and patterns in the data. Users working with Lupus dataset said that it was very difficult for them to see all the 90 data points clearly. Some participants

found a few features of TimeSearcher such as ‘Angular Queries’ and ‘Variable Time-Boxes’ difficult to interpret. As TimeSearcher does not provide any clustering capabilities, users have to manually search for every pattern in the data using ‘time boxes’, which can prove tedious in a large dataset.

HCE: Most users were impressed with HCE. The tool provides a wide variety of features for data analysis. HCE was more helpful to participants working with the viral dataset. Users working with the Lupus dataset gave up data analysis within 20 minutes, complaining that it was very difficult for them to analyze data using HCE.

Spotfire®: Users working with Spotfire® were impressed with it. They did not require any special assistance to understand the tool. They said that most of the visualizations were easy to understand. Most users preferred the heat-map visualization of the Spotfire® over its parallel coordinate or Profile chart display (Figure 4.9). Though the user found the visualization displaying different clusters in the data helpful, they said that it should be easier to interact with. They found it annoying that they could not select and focus on a particular cluster of interest.

GeneSpring®: Users felt that they would have to spend a long time learning GeneSpring®. A few users (2/6), spent an initial 45 minutes just trying to get famil iar with GeneSpring®, after which they gave up the data analysis saying that it will take them too long to comprehend what the tool does. A few users commented that it will be great to have some sort of automation that would show them which visualization to begin the data analysis and how to change the visualization properties. One user said that the basic things should be easy, and visualizing an already normalized dataset should not be so difficult. None of the users could change different properties of visualization such as color, scale, or amount of data to be visualized without help. Users were pleased to know that GeneSpring® provided features to make lists of genes based on different criteria. The users commented that such features could prove to be very helpful. Also, features that allow users to add pathway information to gene lists were considered very useful.

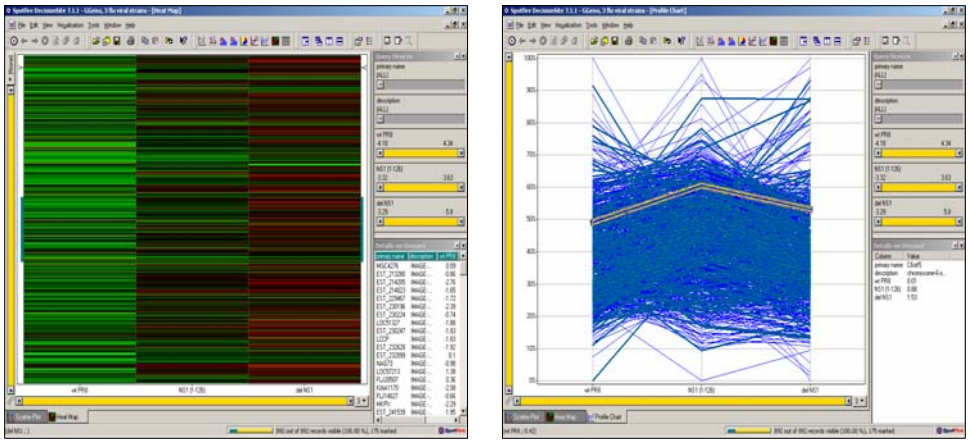


Figure 4.9 The heat map and parallel coordinate views in Spotfire®.

4.5.9 Participants’ Background

One might conjecture that users with more domain experience or software development experience would gain more insight from the data visualizations. Yet, we found that the insight domain value and total number of insights did not appear to depend on participant

background. Averages were similar, and no significant difference between user categories was detected. Due to limited number of subjects, full factorial analysis within tool or dataset groups is not feasible. Trends within user categories followed the same general trends for tools and datasets identified previously. We did not find any differences in the number of insights, value of insight, and hypothesis generation based on participants' background. Rather, we found that these factors were more dependent on the user motivation.

Software developers on average felt that they learned less from the data as compared to others, whereas domain novices felt they learned more from the data. Novices also spent comparatively more time in the study as compared to others. A noticeable difference was in the users' behavior during the experiment. Novice users needed more prompting to make comments about the datasets. They were less confident to report their findings. Software developers almost always made the first insight faster than the novice users.

4.6 Discussion

Commercial vs. Free: Both Spotfire® and Clusterview users resulted in equivalent insight from the Lupus dataset. However, participants using Spotfire® felt they learned much more from the data as compared to Clusterview. Analyzing data in multiple visual representations gave Spotfire® users more confidence that they did not miss any information. Whereas, Clusterview users were more skeptical about their progress, believing that they must be missing something. A simple visualization tool used on an appropriate dataset can have performance comparable to more comprehensive software containing many different visualizations and features.

Free research software like TimeSearcher and HCE tend to address a smaller set of closely related tasks. Hence, they provide excellent insight on certain datasets. Also, since they are focused on specific tasks, they have simpler user interfaces that emphasize a certain interaction model. This reduces the learning time and enables users to generate insights quickly. Spotfire®, despite having a large feature set, has a learning time almost equivalent to the simple tools, which is commendable. This is likely due to Spotfire®'s unified interaction model. The brushing and dynamic query concepts were quickly learned by users, and resulted in early rapid insight generation.

Domain Relevance: A serious shortcoming of the tools is that they did do not adequately link the data to biological meaning. The fact that domain experts performed on par with domain novices, and the small numbers of hypotheses generated, indicates that the tools did not leverage the domain expertise well. Before we conducted the study, we believed that users with more expertise in biology would gain more from visualizations than a novice. We were also curious about whether software development experience would lead to better usage of the tools. However, these background differences did not reveal themselves in the actual insights generated. The difference was only in the users' believed insight, in which novices were overconfident and developers were skeptical.

If the tools could provide a more information-rich environment, such as linking data directly to public gene databases or literature sources, expert biologists could better exploit their domain knowledge to construct higher level, biologically relevant hypotheses. In this experiment, the tools helped users identify patterns in the data, but did not enable them to connect these numerical patterns to the underlying biological

phenomena. A critical need is for highly integrated visualization environments that excel at domain relevance and inference. In this case, understanding gene expression patterns must lead to inference of underlying pathways that model the interactions of the genes (Figure 4.10). Visualization must support this level of inference.

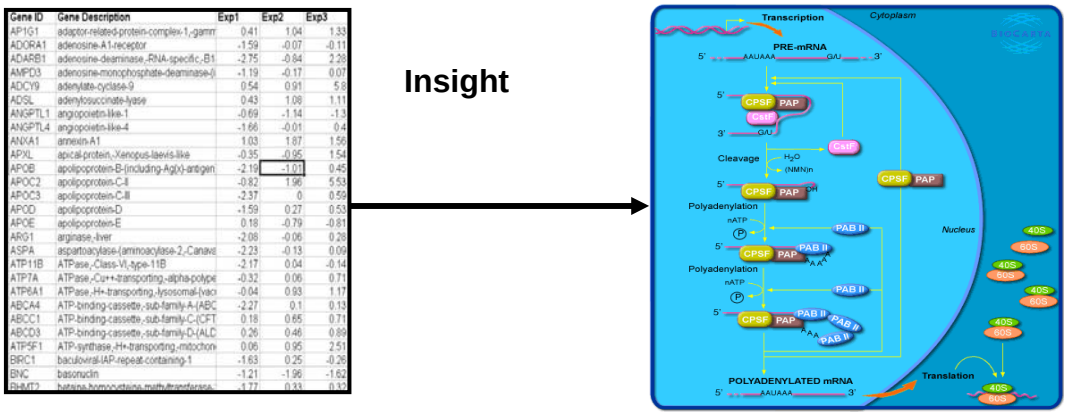


Figure 4.10 Visualizations must support domain-relevant inference, from microarray dataset to pathway models describing interactions within a cell [3].

Interaction Design: The design of interaction mechanisms in visualization is critically important. Usability of interactions can outweigh the choice of visual representation. Spotfire® users mainly focused on the heat-map representation, while GeneSpring® users focused on the parallel coordinates, even though both tools support both representations. The primary reason for this, based on comments from users, was that users preferred parallel coordinates but Spotfire®'s parallel coordinates view employs a poorly designed selection mechanism. Selecting lines in its parallel coordinates view results in unusual and occluding visual highlight feedback that made it very difficult for users to determine which genes were selected and what other genes were nearby (Figure 4.9).

The ability to select and group genes was the most common interaction that users performed. The grouping of genes into semantic groups is a fundamental need in bioinformatics visualization. GeneSpring® provided useful grouping features that enabled more insights in the 'groups' category. More tools need better support for grouping items, based on interactive selections as well as computational clustering, and managing groups. GeneSpring® is the most feature-rich tool of the five, and therefore perhaps the most difficult to learn. However, even though users tended to focus on a small number of basic visualization features, usability issues (such as the large quantity of clicks required to accomplish tasks) reduced their overall insight performance.

Clustering: Certain visualizations, such as the clustering visualizations for both Spotfire® and GeneSpring®, were the most popular in the study. Users commented that it would be very helpful if the interaction techniques for these clustered views were improved, so that they were better integrated into the overall interaction model.

Clustering (Figure 4.11) was a very useful feature throughout, but care should be taken to provide non-clustered overviews first. As in HCE, clustering can potentially bias users into a particular line of thought too quickly. In comparing Spotfire® and Clusterview, users were also more confident when they could confirm their findings between clustered and non-clustered views of Spotfire®.

User Motivation: We noticed that an important factor in gaining insight is user motivation. Clearly, participants in our study did not analyze the data with as much care as they would if the data were from their own experiments. They mainly focused on discovering the overall effects in the data, but were not sufficiently motivated for detailed analysis. Most of the insights generated were classified as breadth rather than depth. However, the visualizations were able to provide a sizeable number of breadth insights in spite of low motivation levels.

4.7 Conclusions

Empirical research methods that use exploratory protocols are prevalent in the fields of HCI and CSCW. Some examples include the critical incident theory [112], Design experiments [22], Breakdown analysis [24], Situated analysis [23], Interaction analysis [21], etc. Most of these methods provide a general protocol to collect data. The data analysis is more contextual. For e.g. the critical incident method defines an incident as "Any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act". However, it is left to experimenters' discretion as to which incidents should be collected for the systems of interest and how these are analyzed. These methods have much broader range of application and are defined to be used in variety of situations. Most of these methods need to be specialized to be applied to visualization evaluation.

The insight-based method provides an alternative method to the traditional task-based method. The method uses different characteristics of an 'insight' to rank visualization tools. Though the insight-based method may not be novel in its experimental protocol, the identified insight characteristics are unique to the method. The main contribution of the method is in the field of bioinformatics as though a large variety of visualization tools have been developed not many evaluation studies have been performed. The insight characteristics provide a way for the biologists' to select the most appropriate tool for their analysis.

5 Task-Based vs. Insight-Based Method on Pathway + Microarray Data Visualization

This chapter presents a study conducted to evaluate pathway + microarray data visualization alternatives using the new insight-based evaluation method identified in Chapter 4. Besides the insight-based method, the study also used the task-based method. This was to allow comparisons between the two methods and to investigate their advantages vs. disadvantages over each other for evaluating visualization tools.

5.1 Introduction

Chapter 3 describes the design space to identify all the solutions for pathway + microarray data visualization alternatives. The design space is divided into two main dimensions: the method to overlay data on pathway diagrams and the number of visualization alternatives used. The three possible alternatives to overlay data on pathway diagrams include: (a) Overlaying data on graph vertices for one timepoint at a time by manipulating a visual property (e.g. color) of the node, and using sliders or similar interaction to animate the graph to other timepoints; (b) Data from all the timepoints can be overlaid simultaneously by using complex node glyphs; Or, (c) small multiples can be used to simultaneously display a miniature graph for each timepoint. All of these alternatives can be used either by themselves or linked to other visualization alternatives using the concept of brushing and linking.

An initial study to evaluate some of the visualization alternatives along both the dimensions was conducted in Chapter 3. The main result from the study was that the method to overlay data on pathway nodes has non-trivial effect on accuracy of participants' responses, whereas the use of single vs. multiple view visualization alternatives has non-trivial effect on the performance times. Since we rated accuracy more important over the performance time, we decided to focus on dimension 1 i.e. method to overlay data on pathway diagrams for the second study presented here.

One of the main limitations of the study reported in Chapter 3 was that it failed to address the real world data analysis scenarios of the biologists. Though, the task-based evaluation method provides controlled means to evaluate visualization tools it has several limitations identified in Chapter 3. To address these, an insight-based evaluation method was reported in Chapter 4. The insight-based method identified several quantifiable characteristics of an insight that can be used to uniformly evaluate visualization alternatives. Thus, to evaluate pathway + microarray data on more realistic scenarios, a study was conducted using the insight-based method.

Also, though the insight-based method appeared useful, there are open questions about how the method compares to the traditional benchmark task method and whether the method can be used instead of the benchmark task method to provide meaningful statistical analyses between visualization tools. Thus, secondary goal of the study reported here was to compare both methods: the task-based and insight-based methods. Such studies to compare empirical research methods are more common to the field of usability engineering, but less frequent in the information visualization domain.

5.2 Literature survey for experiments to compare evaluation methods

The literature for visualization of pathways and graphs with multidimensional data is summarized in Chapters 2 and 3. While the study in Chapter 3 examined the use of multiple views including parallel-coordinates plots, this paper focuses on the primary graph representation itself.

5.2.1 Comparison of Studies for Information Visualization

Different types of studies have been used to evaluate visualization tools, as summarized in [113], [13]. The shortcomings of these studies and the need to develop new evaluation methods for visualization tools that better represent real world data analysis scenarios and also provide better feedback about the usability of the data representation method have been suggested [113], [74].

The literature for comparisons of empirical research methods used to evaluate information visualization tools is sparse, and mostly anecdotal. General guidelines for better tasks and methods to evaluate visualizations can be found in [114]. Recommendations for more consistent and comparable user studies based on a meta analysis is presented in [80]. Authors' comments about the user studies for information visualization, and the lessons learned from these studies and how these were used to design more effective visualization tools and evaluation studies is presented in [115]. A panel discussion summarizing research for visualization evaluation using human subjects, and suggestions and guidelines for conducting such studies by several visualization experts based on their experiences is presented in [116]. Expert reviews as an alternative in certain contexts where designing and conducting user studies can be difficult is suggested in [117].

5.2.2 Comparison of Studies in Usability Engineering

Several studies have been conducted to analyze and compare methods typically used for evaluating user interfaces. A comparison of usage based evaluation techniques and inspection method for groupware systems is provided in [118]. A study to compare the effectiveness of local vs. remote usability studies is reported in [119]. Two methods for children's computer games are compared in [120]. Usability testing methods with multiple participants is compared to heuristic evaluation in [121]. A list of criteria that can be used to compare usability evaluation methods is presented in [122]. Detailed case study of 6 usability methods that evaluates each method's usability error predictive power to actual user tests is reported in [123]. A comparison of different usability testing methods for information retrieval tasks is provided in [124].

Though studies have been conducted to evaluate usability methods that analyze user interfaces with respect to each other, studies to evaluate empirical research methods for evaluating visualization tools are rare. Most of the usability methods are compared based on the number of usability errors found, severity of these errors and participants' and facilitators' experience in the study. Since the dependent variables for the usability methods are usually the same (usability errors) such direct comparisons between the evaluation methods are possible. However, the dependent variables for the benchmark task-based method (performance time, accuracy), and the insight-based method (data insights) are different. Also, the evaluation for visualization tools investigates a wider

range of options (e.g., data representation method, interaction mechanisms used, etc) as compared to the user interface evaluation. Hence higher level measures such as the conclusions about the visualization tools, time spent by the participants in the study, effort spent to analyze the resulting empirical data, etc. need to be used for meaningful comparisons between these two evaluation methods.

5.3 Experiment Design

The aim of this study is to analyze and compare two empirical evaluation methods, using three different visualization alternatives that support analysis of timeseries data in context of graphs. A 2X3 between subjects design examines the following two independent variables:

1. Two empirical evaluation methods: benchmark tasks method, and the insight-based method.
2. Three graph visualization alternatives.

5.3.1 Data

The biologists we were collaborating with conducted a gene expression microarray experiment to analyze impacts of tobacco smoking on flu infection immune response. The actual data was 45,001 rows (genes) X 72 columns (timepoints and conditions). The biological significance of the data and the actual analysis process for this data by bioinformaticians are presented in [125].

A directed graph, having 46 vertices (or genes) and 36 edges (representing gene interactions) representing an actual immune response pathway, was linked to a timeseries dataset representing gene expression for 12 timepoints (Table 5.1). Thus, the participants in the experiment were working with a small subset of the actual data. However, the graph size was based on the typical size of the biological pathways used by the biologists, and corroborated in general by STKE [40].

Data Type	Description
Graph	A directed graph having 46 vertices and 36 edges. Each node had an out degree of 0 to 3.
Timeseries data	Gene expression values for 12 timepoints for each node. Of these, 6 timepoints measured expression values for flu infection for non-smokers, and the remaining 6 timepoints corresponded to flu infection for smokers.

Table 5.1 Data used for the study.

5.3.2 Visualization Alternatives

Three graph visualization alternatives were used in the study. The visual encoding of the data was based on the general color scheme used in bioinformatics, i.e. the color scale from yellow to green was used to display negative data values, and yellow to red was used to display positive data values. These alternatives represent dimension 1 (method to overlay microarray data on pathways) from the design space identified in Chapter 3.

The results from the study in chapter 3 indicated that the method to overlay data on pathways affect accuracy of participants' response whereas, the number of visualization alternatives affect their performance time. As we ranked accuracy more important as compared to performance time we decided to evaluate methods to overlay data on pathway diagrams more rigorously for the second study.

Single Timepoint (1 Tpt): This visualization overlays values for one timepoint on a node at a time (Figure 5.2). A slider lets users iterate over all the timepoints in the data. Mousing over nodes displays the numerical value corresponding to the color.
Multiple Timepoint (M Tpts): This visualization overlays data from all the timepoints on a node using a heat map (Figure 5.3). Mousing over the heatmap cells displays the corresponding numerical value and the timepoint.
Multiple Graphs (M Graphs): This visualization displays a miniature graph for all the timepoints in the data (Figure 5.4). Mousing over a node displays its numerical value, the name of the node (because nodes are too small to clearly show name labels), and also the time point corresponding to it

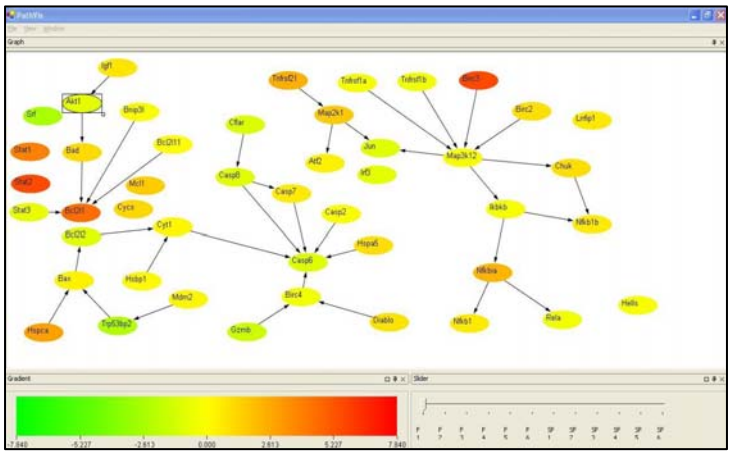


Figure 5.1 Overlay a single timepoint on graph vertices. A slider is used to navigate between different timepoints.

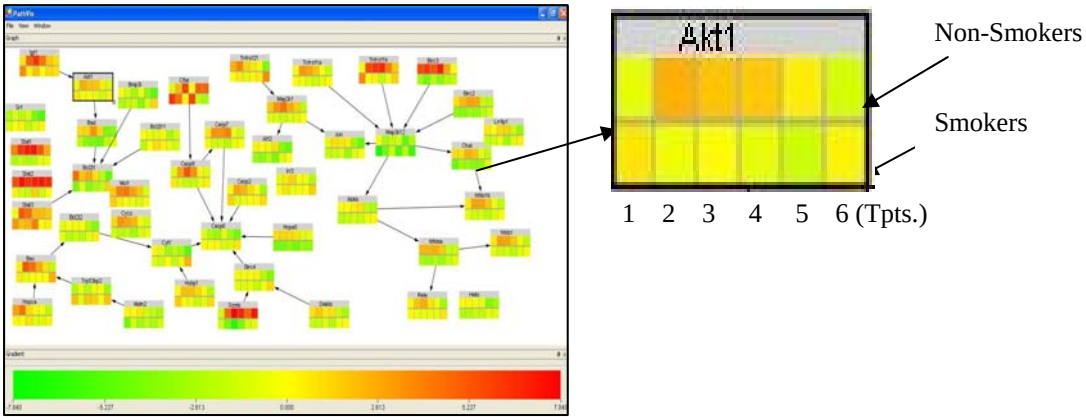


Figure 5.2 Overlay all the timepoints on graph vertices.

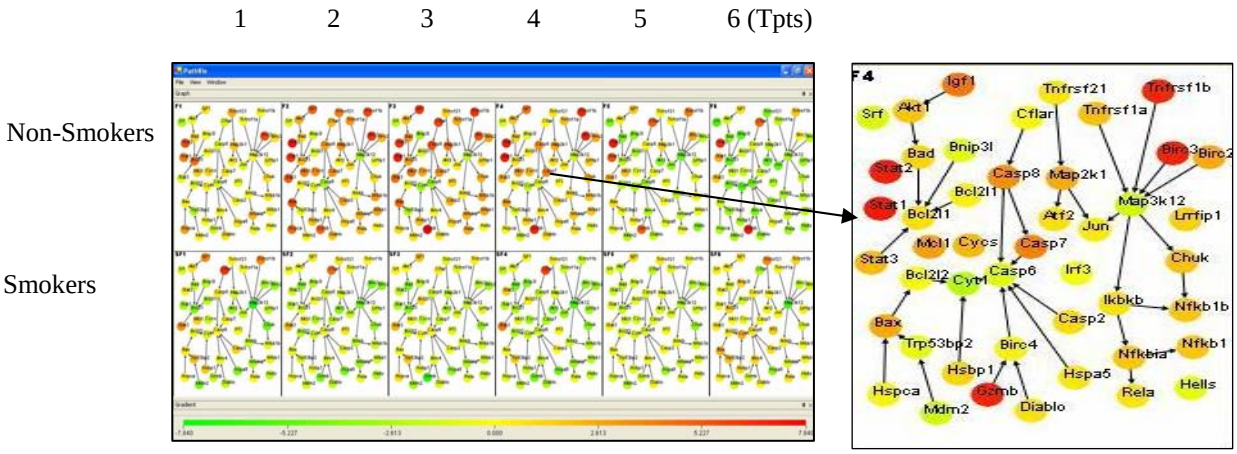


Figure 5.3 Multiple small graphs to display all data timepoints.

5.3.3 Participants

60 participants, 10 for each visualization alternative for each evaluation method, participated in the study. Since the data had a biological background, all the participants in the study were sophomore or junior biology students. Hence, the participants were all familiar with the basic concepts of the data, although were not familiar with this specific dataset.

5.3.4 Experiment Protocol

Before beginning, the participants were given a brief introduction to the visualization alternative that they were assigned and the data background used in the study. Then, the protocols were different depending on the assigned evaluation method.

Task-Based Method Protocol: Participants were required to perform 7 tasks listed in Table 5.2. All the tasks were multiple choice questions, with five possible choices. The tasks were based on the observed analysis tasks of the bioinformaticians who designed the biology experiment and analyzed the actual data [125]. Time and correctness were measured for each task (Table 5.2).

No.	Task
1	Which of the following nodes shows a positive value for all Flu timepoints but negative value for all Smoking+Flu timepoints?
2	What is the overall expression pattern for Flu timepoints vs. Smoking+Flu timepoints?
3	Which of the following nodes is negative for all 12 timepoints?
4	Which of the following timepoints has the maximum number of positive nodes?
5	Which of the following timepoints has the maximum number of negative nodes?
6	At which of the following timepoints, for both conditions, do most nodes change their expression values from previous timepoints?
7	How many nodes are between Map3k12 and Rela?

Table 5.2 Lists tasks for the Task-based method.

Dependent variables	• Time to answer each question • Number of correct answers • Overall time spent in the study

	• Feedback about the visualization alternative

Table 5.3 Lists dependent variables for the task-based method.

Insight-Based Method Protocol:
We requested the participants to analyze the data in a think aloud fashion, until they felt that they had learned all they could from the data. The experimenter sat next to the participants during the study, silently observing the participants' data analysis process and also recording (on a laptop) the data insights and the times at which these were made since beginning the study (Table 5.4).

Dependent variables	• Data insights • Time at which each insight was reported • Overall time spent in the study • Feedback about the visualization alternative

Table 5.4 Lists dependent variables for the Insight-based method

5.4 Experiment Hypothesis

5.4.1 Task-Based Method

The tasks were selected to balance out the performance across the visualization alternatives. We expected none of the visualization alternative to show overall better performance than any other alternative either in terms of accuracy or performance time. Since the graph visualization and data were not very complicated (values for 12 timepoints for 46 nodes) we had no hypothesis related to participants' accuracy of responses.

Task	Hypothesis
T1	Since task 1 requires analysis for a single node over all the timepoints, we believed that the participants using Multiple Timepoint visualization would perform faster than the participants using other methods. Also since Multiple Graph visualization method focuses on the overall changes in the graph expression, we believed that the participants using this visualization may be slower than the participants using other visualization alternatives.
T2	Task 2 requires analysis for overall gene expression Hence, participants using Multiple Graph visualization should perform faster than the participants using other methods.
T3	Task 3 required analysis for a single node over all the timepoints. Since Multiple timepoint visualization facilitates such analysis we assumed that participants using this alternative would perform faster.
T4	Since task 4 requires analysis for a particular timepoint, we believed that the participants using multiple timepoint may take longer to perform this task.
T5	-same hypothesis as task 4-
T6	- same hypothesis as task 4-
T7	This task was to test if there were differences in performances related to a topological task using the visualization alternatives.

Table 5.5 Lists hypothesis for each task for the task-based protocol.

5.4.2 Insight-Based Method

The insight-based method evaluates the visualization alternatives based on the insights participants report. Since the method uses an unguided and an exploratory protocol, it

will be difficult to predict data analysis before hand. However, initial hypothesis for each visualization alternative is listed in Table 5.6

Visualization Alternative	Hypothesis
1 Timepoint Visualization	Since the 1 timepoint visualization displays graph behavior for a single timepoint, we expected the participants to report mostly analysis related to a particular timepoint.
Multiple Timepoint Visualization	The multiple timepoint visualization displays values for all timepoints on a node simultaneously, hence we expected the participants to report insights related to behavior of a particular node.
Multiple Graph Visualization	Here, the participants should report insights related to the overall changes in the graph with respect to timepoints and between conditions.

Table 5.6 Lists hypothesis for each visualization alternative.

5.5 Results – Task-based Method

5.5.1 Overall Analysis

On performing ANOVA analysis, we found that there were no significant differences between the participants on the total time spent in the study or overall differences on the accuracy for the tasks, for all the three visualization alternatives. However, the participants using single timepoint visualization were somewhat [F(1, 117) = 1.52, (p=0.06)] more accurate than multi timepoint visualization.

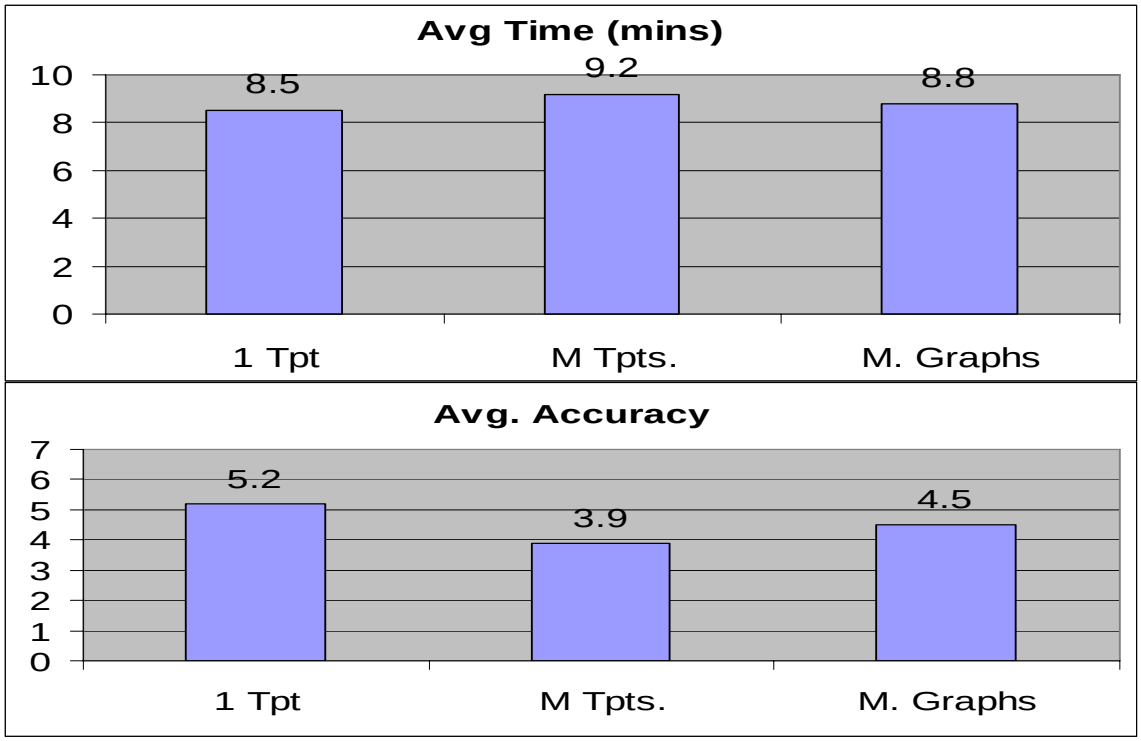


Figure 5.4 Displays the average time (in mins) that participants spent in the study and the average number of correct responses (out of 7 tasks) for the three visualization types for the task-based protocol.

5.5.2 Individual Task Analysis

Significant results from paired ANOVA analysis on tasks between the three visualization alternatives for time and accuracy are summarized in Table 5.7. There were no differences in participants' performance, in terms of accuracy and task completion time, for tasks 2 and 3. Though tasks 4 and 5 were equivalent, task 5 required more careful analysis as compared to task 4, as the timepoint at which most nodes were positive was more obvious as compared to the timepoint at which there were most negative nodes.

Tasks	1 Tpt	M Tpts	M Graphs
T1		Faster [F(2, 27) = 3.68, p=0.038]	
T2	-	-	-
T3	-	-	-
T4		Less accurate [F(2, 27) = 3.85, p=0.033]	Faster [F(1, 18) = 8.47, p=0.01] than M Tpts
T5	Somewhat more accurate than M Tpts [F(1, 18) = 3.42, p=0.08]	Slower [F(2, 27) = 3.49, p=0.044]	-
T6	-	Somewhat slower than 1 Tpt [F(1, 18) = 3.03, p=0.098]	Somewhat more accurate than M Tpts [F(1, 18) = 3.42, p=0. 08]
T7	-	-	Faster than M Tpts. [F(1, 18) = 4.98, p=0.038]

Table 5.7 Summary of Anova analysis for the task-based protocol

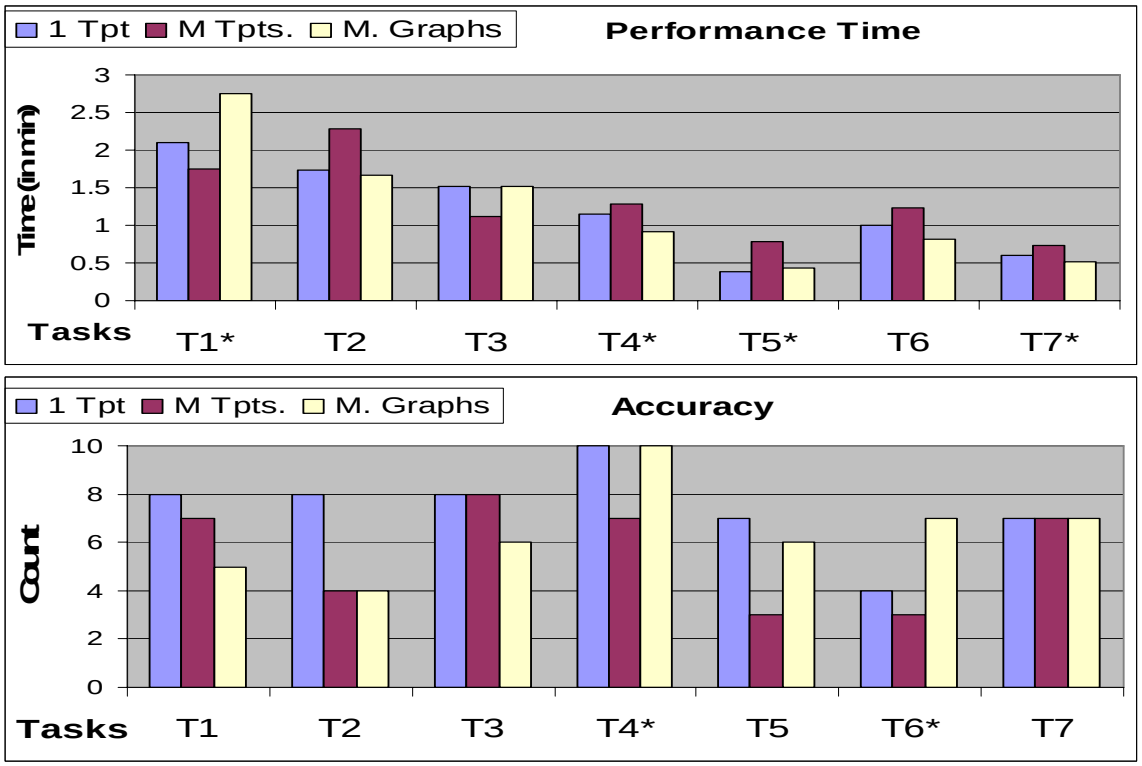


Figure 5.5 Displays average time in minutes, and total count of correct responses (out of 10 participants) for each task, for the three visualization alternatives. * on a task indicates significant performance differences.

5.5.3 Hypothesis vs. Results

The comparisons between the hypothesis and the actual results for the tasks are listed in Table 5.8

Task	Hypothesis	Results
T1	Participants using Multiple Timepoint visualization alternative should display the fastest performance. We had also assumed that the participants using Multiple Graph visualization alternative would have the slowest performance	The results confirmed the hypothesis
T2	We had assumed that the participants using the Multiple Graph visualization would display the fastest performance.	We did not observe any differences between participants' performance either on task completion time or accuracy.
T3	We had assumed that the participants using Multiple Timepoint visualization would display the fastest performance.	We did not observe any differences between participants' performance either on task completion time or accuracy.
T4	Since the task required analysis of graph at a Single Timepoint, we assumed that the participants using Multiple Timepoint visualization would be slower than the participants using other visualization alternatives	Participants using Multiple Graph visualization were faster than the participants using Multiple Timepoint visualization. Also, the participants using Multiple Timepoint visualization were less accurate than the other visualization alternative.
T5	- same as task 4-	Participants using Multiple Timepoint visualization were slower than the participants using other visualization alternatives. Also, the participants using Single Timepoint were more accurate than the participants using Multiple Timepoint visualization.
T6	- same as task 4-	Participants using Multiple Timepoint were slower than the participants using other visualization type. Also, the participants using Multiple Graphs were more accurate than the participants using Single Timepoint visualization.
T7	- none -	We found that the participants using Multiple Timepoint visualization were faster than the participants using other visualization alternatives.

Table 5.8 Lists comparison for hypothesis vs. results for the task-based protocol.

As shown in the table, the visualization alternatives had most influence on the performance time. Though for tasks T4, T5 and T6 it was observed that the participants using Multiple Timepoint visualization were somewhat less accurate as compared to the other visualization types.

5.5.4 Conclusions

The analysis for the individual task performance times and accuracy leads to the conclusions about the visualization alternatives summarized in Table 5.9. We found that the single timepoint visualization had more controlled performance as compared to the other visualization alternatives. Both multiple timepoints and multiple graph visualization

showed different performances depending on the tasks. The multiple timepoint visualization showed best performances for analyses related to a single node expression, whereas multiple graphs showed best performance for analysis related to overall gene expression. The multiple graphs provided a good data overview that allowed users to easily select the timepoints they wanted to focus on.

1 Tpt	M Tpts	M Graphs
+ More consistent performance for all the tasks.	+ Faster performance for single node analysis. – Slower and less accurate for overall graph expression at a particular timepoint.	+ Faster performance for overall expression. + More accurate and faster for finding interesting timepoints. + Faster than M Tpts for graph topology tasks.

Table 5.9 Summarizes conclusions about visualization alternatives from the task-based study.

Most conclusions about the single and multiple timepoint visualization alternatives in the second study are similar to the first study reported in Chapter 3. As in the first study, the participants using multiple timepoint visualization were less accurate and slower for tasks involving analysis of the graph at a single timepoint. For the tasks that required searching for timepoints of interest we found that similar to study 1 the participants using single timepoint visualization showed better performance in terms of accuracy and time as compared to the participants using multiple timepoint visualization. However, the participants using multiple graph visualization showed the fastest performance.

In the previous study it was also found that the multiple timepoints were faster and more accurate for tasks that required searching for outlier nodes, i.e. nodes that display different behavior than most other nodes. Since we did not have a task to represent this information, it was not possible to make such conclusions from the second study.

5.6 Results - Insight-Based Method

5.6.1 Overall Performance Analysis

On performing ANOVA analysis we found that the participants using single timepoint visualization spent significantly more amount of time in the study as compared to other participants [$F(2, 27) = 4.33$, $p = 0.02$]. The participants using single timepoint visualization had more number of distinct data insights [$F(2, 27) = 3.02$, $p = 0.065$] than both the multiple timepoint [$F(1, 18) = 3.63$, $p = 0.07$] and the participants using multiple graphs [$F(1, 18) = 4.79$, $p = 0.04$]. Note that the data insights are distinct for a participant. However, the data insights may be repeated across participants when more than one participant reported the same data insights.

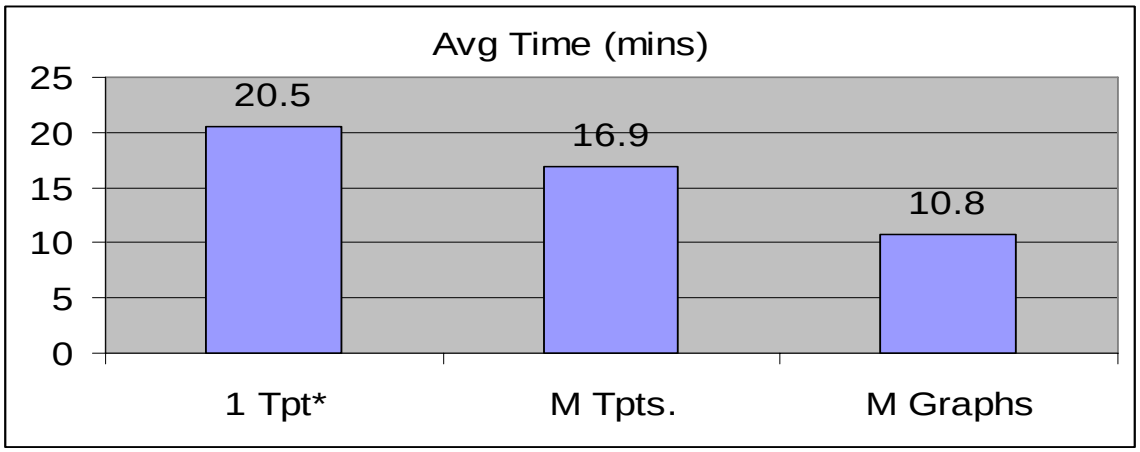


Figure 5.6 Average time in minutes participants spent in the study for each visualization type, * indicates significantly performance differences. Participants using single timepoint visualization spent significantly more time (p=0.02) in the study.

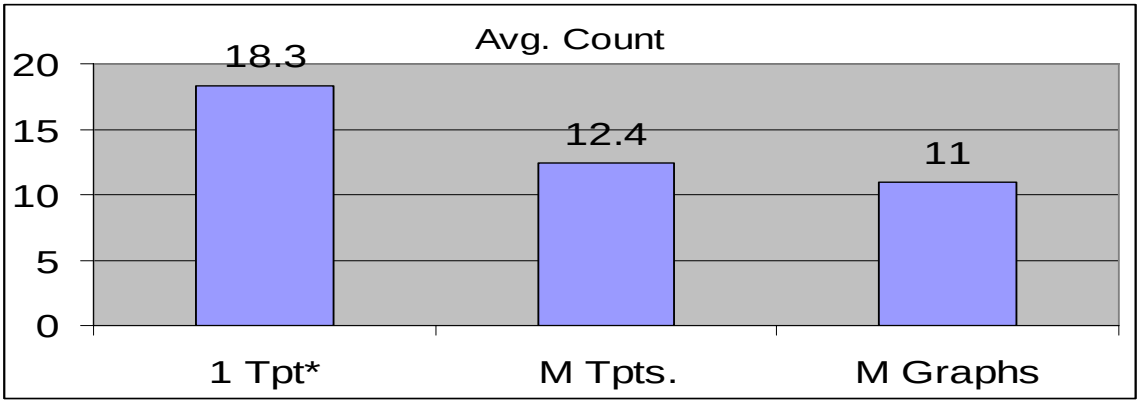


Figure 5.7 Average Count of Insights for each visualization type, * indicates somewhat better performance. Participants using single timepoint visualization had significantly more number of distinct insights as compared to multiple graph visualization (p=0.04) and somewhat more number of insights as compared to the multiple timepoint visualization (p=0.07).

5.6.2 Performance based on Insight Categories

On analyzing the participants' data insights, we found that all of these could be grouped into 7 distinct categories. A data insight belongs to only one of the seven categories.

Gene Expression: Most frequent data insights reported expression pattern for just one gene. E.g.: "Gene Gzmb displays positive values for all the flu timepoints except the first timepoint, but is negative for all the smoking timepoints". "Gene Irf3 displays similar expression patterns for flu infection for both the non-smokers and smokers".

Topology: Some of the insights reported used only the graph topology. This did not include any information about the associated timeseries data. E.g.: "The map3k12, casp6, and the bcl2ll genes seem to be major focal points in the graphs as they have a lot of arrows pointing towards them". "There are a few breaks in the graphs e.g., a few nodes that are not connected to anything". None of the participants using M. Graphs reported such insights.

Topology + Expression: Some of the insights reported by the participants investigated gene expression based on graph topology or effects of genes on each other connected directly or indirectly through other genes. E.g. “All the genes towards outside, i.e. Trnf2, birc3, etc. are positive for almost all the timepoints as compared to the inside ones that they are supposed to affect”. “The genes on right side of the graph have more regular pattern i.e. either they become more or less expressed with timepoints, whereas genes on the left side have a more random pattern”.
Timepoint analysis: Some participants reported insights that investigated overall graph expression at a particular timepoint. E.g.: “A lot of genes are negatively expression at timepoint 5 for smokers as compared to all other timepoints”. “Almost all the genes get more positive for non-smokers between timepoints 3 and 4 except Cyt1 and Trnsf21”. None of the participants using Multiple Timepoints reported such insights.
Experiment Conditions: All the participants in the study tried to evaluate the differences in the gene expression between non-smokers and smokers. E.g.: “Overall non-smokers have more positively expressed genes as compared to smokers”. “Genes for non-smokers seem to get more negative with time, whereas for smokers an opposite effect is seen”.
Outliers: Some participants identified a few genes that displayed different expression values than other genes in the graph. E.g.: “Stat1 gene is different than most other genes as it gets more positive with time for non-smokers, whereas most other genes get more negative”. “Trp53bp2 is unique as it is more expressed in smokers vs. non-smokers”.
Summary: Some participants tried to summarize their findings about the data or suggested future research based on their data analysis. These insights are most similar to the insight characteristic hypothesis that was ranked very high in the study reported in [Saraiya1 et al., 2005]. E.g.: “There is no correlation between expression values for genes that have direct influence on each other, this suggests that the information presented here is incomplete, or there may be several other biological factors influencing the genes and is not shown by the graph visualization”. “Smokers don’t have that many highly expressed genes, seems like a lot of them may reduce the gene expression of the subsequent genes. This may eventually lead to less expression for the overall immune system against the flu for smokers”.

The results from ANOVA analysis between participants, on number of distinct insights for each category reported by each participant using a particular visualization alternative are listed. As shown in table 5.10, there were no differences on the number of insights for the “Experiment Condition” category across the visualization alternatives.

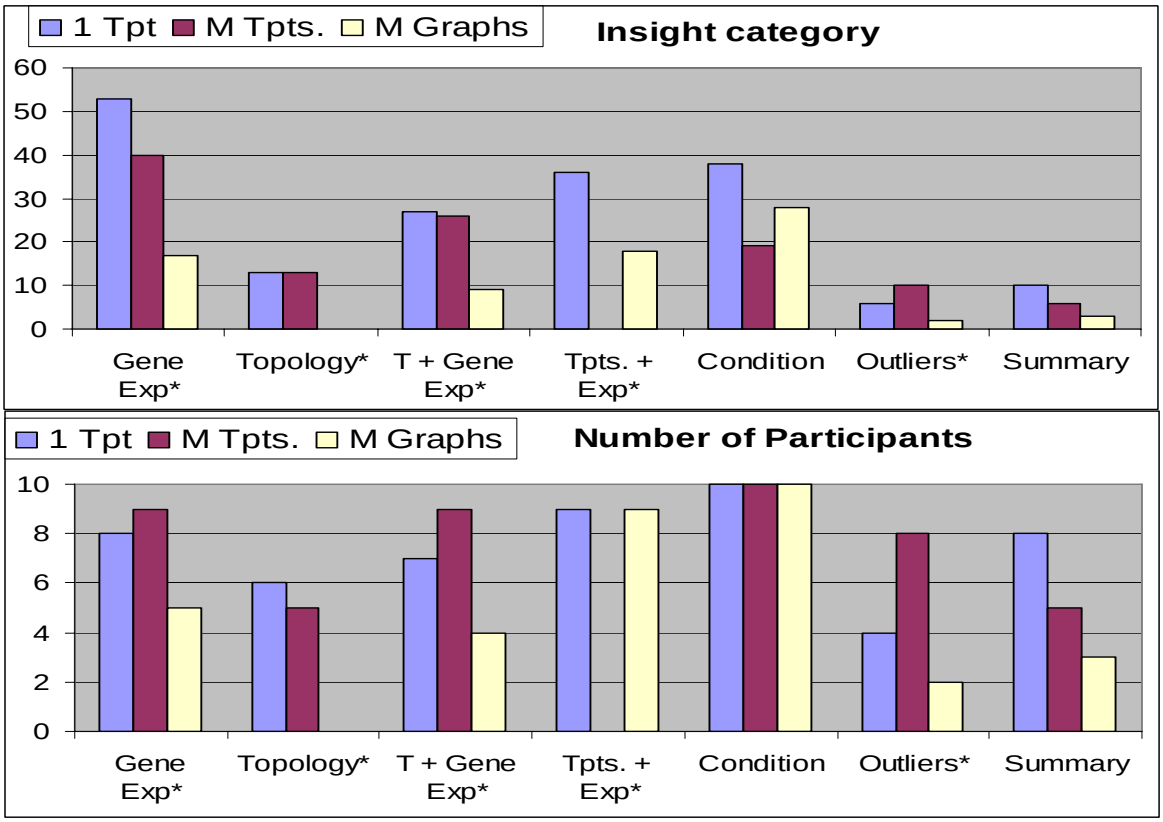


Figure 5.8 Displays the total number of insights, and the number of participants who reported these for each insight category. * indicates significant performance differences.

Category	1 Tpt	M Tpts	M Graphs
Gene Expression			Less than 1 Tpt [F(1, 18) = 3.73, p=0.069]
Topology			Significantly less than the other two visualization alternatives [F(2, 27) = 3.604, p=0.04)]
Topology + Expression			Somewhat less [F(2, 27) =2.525, p=0.09] Significantly less than M Tpts [F(1, 18) = 4.7, p=0.04] Somewhat less than 1 Tpt [F(1, 18) = 4.22, p=0.054]
Timepoint Analysis		Significantly less than others [F(2, 27) =16.2, p=2.38E-05]	Significantly less than 1 Tpt [F(1, 18) = 5.4, p=0.03]
Condition	-	-	-
Outliers		Somewhat more than others [F(2, 27) =3.08, p=0.06]	Significantly less than M Tpts [F(1, 18) = 10.28, p=0.004]
Summary			Significantly less than 1 Tpt [F(1, 18) = 7.22, p=0.015]

Table 5.10 Summarizes the results for insight categories.

5.6.3 Hypothesis vs. Results

Table 5.11 lists the differences between the hypotheses for each visualization alternative and the results from the insight-based study. Since for the insight-based method the

participants perform data analysis in an unguided and exploratory fashion, it will be difficult to predefine all the insight categories without performing the experiments. However, we had hypothesized the general influence of the visualization alternatives. There were some unanticipated results about the visualization alternatives after analyzing the reported insights.

Visualization Alternative	Hypothesis	Results
1 Timepoint Visualization	Since the 1 timepoint visualization displays graph behavior for a single timepoint at a time, we expected that the participants using this alternative will report most analysis related to a particular timepoint.	The participants using this visualization alternative displayed most controlled performance for all the insight categories. The participants using this visualization alternative reported most insights belonging to the category 'summary'.
Multiple Timepoint Visualization	The multiple timepoint visualization displays values for all timepoints on a node, we expected the participants to report most insights related to behavior of a particular node.	Participants using this visualization alternative missed analysis of the graph for a particular timepoint. However, participants using this visualization alternative reported most insights for the category 'outliers'.
Multiple Graph Visualization	Here, the participants should report insights related to the overall changes in the graph with respect to timepoints and between conditions.	The participants using this visualization alternative reported fewer insights than the other visualization alternatives for most insight categories. This was most noticeable for the categories: topology, topology + expression, and outliers. The participants using this visualization had fewer insights than the Single Timepoint visualization for the categories: Summary and Timepoint analysis.

Table 5.11 Comparison between the hypothesis and results for different visualization alternatives from the insight-based study.

5.6.4 Conclusions

Participants' performance on the insight categories lead to the conclusions about the visualization alternatives listed in table 5.12.

1 Tpt	**M Tpts**	**M Graphs**
+ Somewhat better at summarizing findings. + Best for single timepoint analysis. + More consistent performance for all insight categories.	+ Best for identifying outlier nodes. – Difficult to analyze a single timepoint.	– Difficult to focus on expression values for a single node. – Difficult to analyze graph topology.

Table 5.12 Conclusions about the visualization alternatives from the insight-based method.

5.7 Comparisons between Methods

5.7.1 Total Time Spent

On performing ANOVA analysis, overall participants in the insight-based method spent significantly more total time in the study as compared to the task-based method [$F(1,58) = 21.27$, $p<0.01$]. Participants using Single Timepoint [$F(1,18) = 28.35$, $p<0.01$] and Multiple Timepoint [$F(1, 18) = 9.006$, $p<0.01$] visualizations spent significantly more time in the insight-based method as compared to the task-based method.

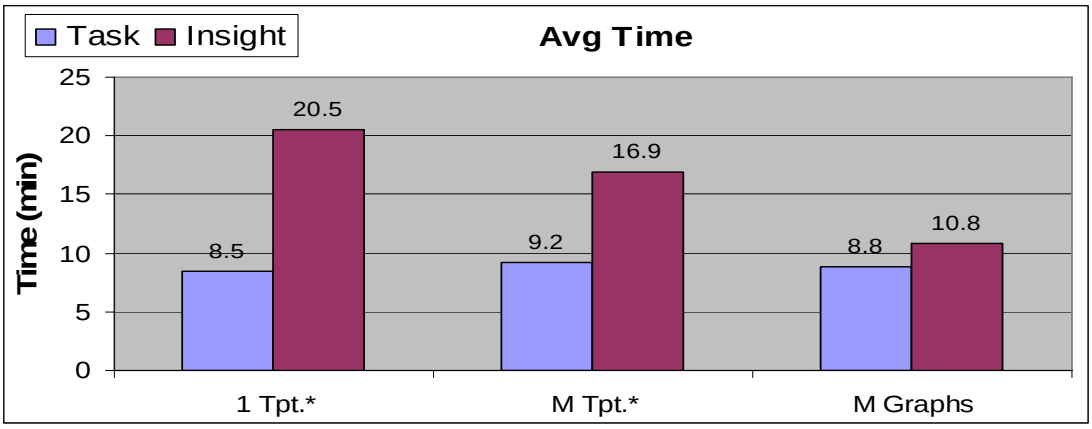


Figure 5.9 Average time participants spent in the study for each visualization method, for task-based and Insight-based methods. * indicates significant performance differences.

5.7.2 Comparisons on Pre-selected Tasks

Number of participants who reported equivalent insights: Some participants in the insight-based method reported insights exactly similar to the tasks that were used in the task-based method. E.g.: Similar to task 4 in the task method, the participants in the insight-based method reported that timepoint 4 for non-smokers has maximum number of positively expressed genes. Figure 5.11 summarizes the number of participants that made insights comparable to a particular task. Since task 7 was very specific, we omitted it from the analysis. As displayed in Figure 5.11, none of the participants using Multiple Timepoints visualization reported insights requiring data analysis for a single timepoint (tasks 4, 5, and 6). The participants using Multiple Timepoints visualization were significantly slower or less accurate on these tasks in the task-based method. The effect was confirmed and found even more significant in the insight-based method when analyzing the insight category: timepoint analysis.

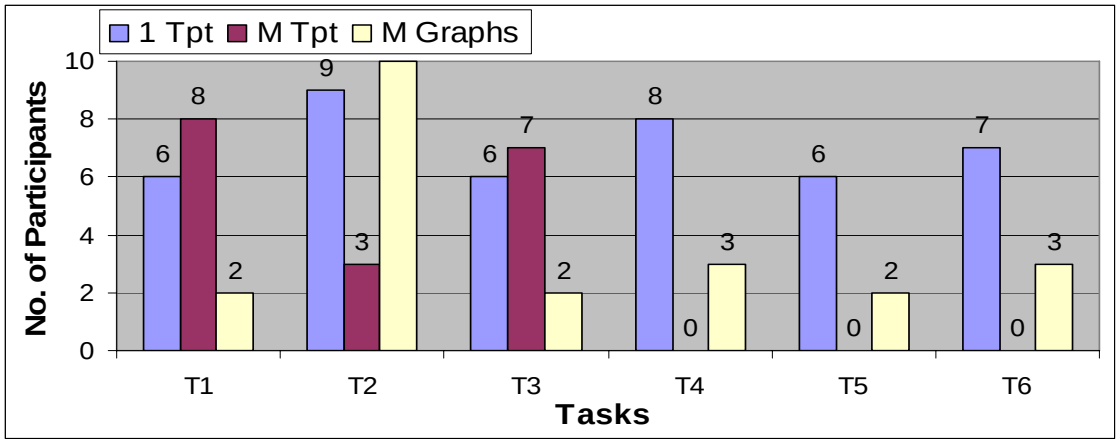


Figure 5.10 Number of participants for each visualization who reported insights equivalent to the tasks in the task-based method.

Tasks vs. insight categories: Table 5.13, lists the tasks and the corresponding insight categories. Conclusions about the visualization alternatives based on the method type are also listed. For most tasks the differences found in task-based method were made more significant in the insight-based method. tasks 1 and 3 required the participants to analyze expression values for a single node. From the task-based method it was found that M. Tpts were significantly faster at one of these tasks. However, there were no differences between M Tpts and 1 Tpt visualization alternatives for the insight-based method. The participants using M Graphs had the least number of such insights. There were no performance differences in either method for task 2. The most interesting results are for tasks T4 – T7. These tasks required analysis of the graphs at a single timepoint. In the task-based method it was found that M Timepoint visualization were either slower or less accurate for these tasks. However, none of the participants using M Timepoint visualization in the insight-based method reported any insights relative to these tasks. For task 7 we found that though the participants using M Graphs performed this task faster, however in the insight-based method none of the participants reported any data insight belonging to this category. Thus, providing a set of predefined tasks may have forced the participants to perform this analysis that the visualization would otherwise not encourage.

Furthermore, the insight-based method found additional categories that we had found difficult to create tasks for using the task-based method. These were primarily the Topology + Expression and Outlier Categories. Hence, the insight-based method revealed further advantages of some of the visualization alternatives for those categories.

Tasks	Insight Category	Task-based study	Insight-based study
T1	Gene expression	M Tpts significantly faster.	M Graphs least number of such insights
T2	Condition	No differences	No differences
T3	Gene expression	No differences	M Graphs least number of such insights
T4	Timepoint analysis	M Tpts least accurate. M Graphs faster.	1 Tpt most. M Tpts least.
T5	Timepoint analysis	M Tpts slower. 1 Tpt more accurate than M Tpts.	1 Tpt most. M Tpts least.
T6	Timepoint analysis	M Tpts slower.	1 Tpt most. M Tpts least.

*T7	Topology	M Graphs faster than M Tpts.	M Graphs least.

Table 5.13 Compares participants' performance on the selected tasks for both the methods. * in the table indicates different conclusions about the visualizations between both methods.

5.7.3 Empirical Data Analysis

The data analysis process for the task-based method was more straightforward as compared to the insight-based method. It required the use of standard statistical analysis methods like ANOVA and paired t-tests. It took about 6-7 hours to finish the entire process, as the investigators had previous experience analyzing such data.

The data analysis process for the insight-based method is more complex. The amount of empirical data collected for the insight-based method can support more rich analysis options. The participants' insights were analyzed first to find suitable categories to group the insights. The choice of categories can be dependent on the investigators' preferences and data understanding. A discussion was required between the investigators to finally agree to a list. With meetings involved it took about 3-4 days to finish the data analysis. Thus in contrast to the task-based method, data analysis for insight-based method is more complicated and subjective. It is possible that other data analysts may have grouped the insights differently. For future work, a more generalized insight categorization as in the task categorization in [126] can be tried.

5.7.4 Feedback about the Visualization Interface

Usability errors: Though both methods were conducted to evaluate visualization alternatives, the insight-based method required more interaction with the participants. The experiment protocol for the insight-based method required a closer observation of the participants' data analysis procedure and more one-to-one interaction. This made it easier to notice if the participants were having any difficulties with the user interface. Also while performing data analysis participants commented about the visualization interfaces such as "the choice of color is weird", "the timepoint labels are difficult to understand", etc. Such information is easier to miss in the task-based method. We also noticed in the insight-based method that participants using Single Timepoint visualization enjoyed the study because the visualization was more interactive in comparison to the other visualization alternatives. This may have prompted these participants to spend more time in the study as compared to the other participants.

Data representation: Participants in the insight-based method provided more feedback about the visual representation used to visualize the graph. While analyzing the data, the participants would comment on the difficulties and suggest other data representation methods that they thought would support some of their data analysis tasks in a better way. E.g., the participants using Multiple Graph visualization commented that it was difficult for them to focus on a single gene only. The participants using Multiple Timepoints visualization commented that they were having trouble focusing on a single timepoint. They said that somehow the visualization was prompting them to focus on the overall node expressions, and that interactions could be used to drill down.

5.7.5 Time Analysis

The task-based method allows comparison of the visualization alternatives based on the time participants take to complete specific tasks. Usually, the visualization tool that allows participants to complete a task faster is considered better for that particular task.

For the insight-based method, the data insights and the time at which they were made are recorded. This allows the analysts to investigate what types of insights were reported by the participants earlier in the study, and as the participants spent more time in the study how these initial insights were modified or how were new insights added to the previous insights. This may allow an extraction of the participants' data analysis patterns across the visualization tools.

5.7.6 Effect of Individual Differences

The task-based method provides all the participants with an equivalent set of tasks. The list of tasks provides very specific direction to the participants throughout the study. This prevents the participants from getting confused about what to do next. Also, it makes the experience similar for most of the participants.

The insight-based method is completely unguided. It is important for the study that the participants think aloud. It is possible that some participants are more communicative than others, and may report more insights as compared to other participants who may have actually had similar data insights but choose not to verbalize them. Sometimes participants, depending on the type of visualization alternative they were using, felt that some insights were so noticeable that they may be too trivial and not worth reporting. Thus, findings from the insight-based study are more likely to be affected by the individual differences between the participants.

The participants in the insight-based method were suspicious of our intentions, and some asked if the data insights they were reporting made sense, or if they can be provided with more idea as to what they should be reporting so that they can be more helpful. When the participants in the insight-based method became confused, sometimes they needed to be encouraged to report insights. We would just say "yes, that makes sense". Some users required more prompting than the others. It may be helpful in the future to decide if the participants should be provided with such encouragement to make the study more uniform. A few participants reported that the entire study felt as if there was some catch involved to it. They thought there was either something that they were supposed to definitely notice, or that we wanted them to completely miss. At the end of the study, when participants were ready to leave, they wanted to know if they behaved as we expected them to or what was the point of the entire study.

5.7.7 Participant Motivation for Data Analysis

All the participants in the study were undergraduate biology students. To encourage participation in the study, they received some course credit. It is possible that some participants came just for the credit and were not motivated to perform data analysis. For the task-based method, it could be either lack of motivation or difficultly in understanding the visual representation that can affect participants' performance. For the insight-based method it was easier to notice such unmotivated participants because there was more communication with the investigator. The participants would often comment that they were tired or say "I just came from class, my mind is blank, please give me a

minute to rest". We also noticed that participants who came during the weekend were more relaxed and interactive in the study, whereas the participants who came during the weekdays were less inclined to spend as much time in the study. Potentially, in the future, since such unmotivated subjects can be recognized in the insight-based method, they could be filtered from the study so as to focus on a more realistic scenario.

5.8 Conclusions about the Visualization Alternatives

Table 5.14 summarizes conclusions for the visualization alternatives using both methods. Since the dependent variables for both methods are different they provide different conclusions about the visualizations. The task-based method provides feedback in terms of accuracy and performance time. The insight-based method provides feedback based on the types of data insights the visualization generated. Since the tasks are pre-selected, they provide a more reliable feedback for the visualization in terms of the tasks. This allows designers to judge accurately if a visualization design *supports* a particular task or not. An unguided method provides feedback at a higher abstraction level, suggesting what kinds of data analysis a particular visualization method *motivates*. The fact that users may not perform certain tasks with it may not mean that the task is not supported, but that the visualization encourages the users to focus on other data analysis aspects.

Vis.	Task-based method	Insight-based method
1 Tpts	+ More consistent performance for all the tasks.	+ Somewhat better for insight category Summary. + Best for single timepoint analysis. + Most consistent performance for all insight categories.
M Tpts	+ Faster performance for single node analysis. – Slower and less accurate for timepoint analysis.	– Didn't result in any insights related to a single timepoint + Best for identifying outlier nodes.
M Graphs	+ Faster performance for overall expression. + More accurate and faster for finding interesting timepoints. + Faster than M Tpts for graph topology tasks.	– Few insights related to expression values for a single node. – Didn't result in insights involving graph topology.

Table 5.14 Comparison of the conclusions about the visualization alternatives from both methods.

5.9 Conclusions about the Experiment Protocols

The study reported here was conducted to compare two empirical research methods for evaluating visualization alternatives. Since the dependent variables for both the methods are different, the studies were compared on higher level criteria most relevant to evaluating visualization tools. A difference between the insight method and the task-based method is that the task-based method is more uniform across the participants both in terms of the user experience and the data collected from the experiment. The insight method, on the other hand, is somewhat subjective. It is possible that given a dataset two participants may analyze it in different ways and report different insights. Hence, a higher level analysis such as grouping them in categories or assigning domain value is needed, making the data analysis partly subjective. Thus, though the insight-based method

provides a way to capture a real world data analysis scenario and a wider range of comparison factors for the visualization, two experimenters may analyze the data differently and present slightly different conclusions about the visualization tools.

There are several key findings between the methods in the comparison of the visualization alternatives. In general, many of the findings in the task method were confirmed, or even exaggerated, in the insight method. This may provide some validation of the insight method to detect effects found by the task method. However, some findings were counter, indicating that users behave differently when not in the forced direction of a task-based method. Overall, the task method tended to favor the Multi Graphs visualization, while the insight method emphasizes advantages of the Single Timepoints visualization.

Though the task-based method is more uniform, it provides feedback only on the tasks selected. Designing proper tasks is non trivial [117]. This can be even more difficult for complicated and deeper datasets. Also, selecting pre-defined tasks requires better understanding as to how a particular dataset may be analyzed by an actual data analyst. Such information is not always available. The insight study found that Multiple Timepoints visualization performed well for finding the outlier nodes. We did not get this information from the task-based method because we did not have benchmark tasks to reflect that evaluation. However, because of its unguided protocol, the insight method may allow participants to miss certain type of insights. None of the participants using Multiple Graph visualization made insights about the graph topology, even though they performed this task with a fast performance time in the task-based method. Thus, the fact that the participants did not perform a task does not mean that the task is not supported by the visualization tool, but indicates that the visualization prompts the participants to focus on other tasks.

Due to the amount of interaction required for the insight-based method between participants and the experimenter, a closer observation about how the data representation and interaction mechanisms are used by the participants is possible. This may allow visual designers to conclude if the data representations and user interaction features were used as planned or not. Also, different participants may perceive a representation or use the interaction mechanism in different ways, suggesting to visualization designers a wide range of possible combinations in which the visualization may be used. Table 5.15 summarizes the comparisons between the two experiment protocols for evaluating visualizations.

Comparison Factor	**Task-based Method**	**Insight-based Method**
Data & Visualization Tools	• Best with simple data. • Better with simple visualization.	• Better with complex data. • Better with richer visualization tools.
Participants	• Can be applied with any users.	• Best with expert users.
Limitations	• Feedback only on selected tasks. Difficult for deeper data analysis tasks.	• Motivating participants without biasing them for data analysis.
Empirical Data Analysis	• More uniform. • Faster.	• Richer analysis options but higher variance, and more subjective. • Longer analysis.
Primary Outputs	• Indicates whether tasks are	• Indicates what tasks a

	supported by a visualization.	visualization motivates.

Table 5.15 Summarizing comparisons between the task-based and the insight-based methods.

5.10 Discussion

The insight-based method presented in Chapter 4 recognized several characteristics of an *insight* such as hypothesis generation, breadth vs. depth, directed vs. undirected, and domain value. For the data analysis discussed here, we decided to focus just on the category of data insights. Grouping insights by categories provided us with sufficient basis to evaluate the visualization tools for the present discussion.

Most of the participants in this study were undergraduate biology students. For the insight-based method, at the end of the data analysis some participants were confident about the data analysis and could summarize the data or make hypothesis about the biological phenomenon suggested by the data. Such comments were ranked very high in the earlier insight-based study reported in Chapter 4. However, though the participants had biological background, they did not have enough familiarity with the specific immunity phenomena examined by this dataset. Any such hypotheses were just speculations. They would not be able to judge the actual value of such findings. The data insights from the visualization tools ranked and evaluated by the actual data analysts will be different than those by the actual user. Also, the data and tools used in both the studies (Chapters 3 and 5) were more simplistic to reduce the learning time and allow users to complete the analysis in limited time.

For real world data analysis scenarios, a data analyst spends much more time analyzing the data. Also, the type of visualization tools and the data analysis procedures of an actual data analyst will be different compared to the participants in the short term study. Chapter 6 presents a longitudinal study of visualization tools that investigates long term visualization usage by actual data analysts.

6 Insight-Based Longitudinal Study

The insight-based studies conducted to evaluate visualization tools in the earlier chapters were short term. We performed a longitudinal study to analyze if the insight-based method can be used for long term studies. This chapter describes a longitudinal study of a bioinformatics dataset analysis. The main focus of this work was to capture the entire analysis process that an analyst goes through from a raw dataset to the insights sought from the data. The study provided interesting observations about the use of visual representations and interaction mechanisms provided by the tools, and also about the process of insight generation in general.

6.1 Introduction

An initial attempt to capture the real world exploratory data analysis scenario in a short-term controlled study using an insight-based methodology is reported in Chapters 4 and 5. Though the studies provided interesting observations about the visualization tools, they had limitations. The studies measured the insight process for short term data analysis by the participants, and thus failed at capturing the long term insight gained by users who spend more time analyzing the data. The amount, time and type of insight generated may change as one becomes more familiar with the visualization tool as compared to using it for the first time.

Most importantly, the participants in the insight-based studies reported earlier were unfamiliar with the experimental context of the data used in the study. Hence, the data did not mean as much to them because, simply put, it was not their data. Since the participants were not self-motivated to perform data analysis, they had to be prompted during the study to report insights. Thus, the study failed to address the most important factor i.e. ***motivation*** that drives a data analyst to spend days and often months analyzing a particular dataset. Also the study did not capture the ability of a data analyst to judge the ***significance*** of reported insights, which is usually based on users' domain knowledge and also familiarity with the data background and the experimental context.

To address these issues we performed a longitudinal study by working closely with bioinformaticians who were ready to start analyzing data from a microarray experiment using visualization tools. We wanted to analyze if the short term studies reported in earlier chapters are representative of the real-world data analysis process. Other goal of the study was to gain basic understanding into the visual analytics process. The primary research questions addressed by the longitudinal study were: How are different visualization tools used to gain insight into the data? How much effort and time are required to derive the most interesting insights (e.g. hypothesis generation Chapter 4)? What process is followed by users to get needed insights? How is insight synthesized over time? Is it by constantly discovering the unexpected trends in the data or is it a gradual process that builds newer and deeper insights in the context of previously generated ones?

A primary use of the visualization tool is to gain insight into the data [76], [77]. For this, a visualization tool not only provides data representations but also supports interaction mechanisms. We were also interested to learn: Which visualization techniques and interaction mechanism combinations were most effective in providing insights? And more importantly, how do users overcome the shortcomings of a visualization tool?

6.2 Experiment Design

In this longitudinal study, we observed bioinformaticians over a long period of time as they analyzed their data from a microarray experiment.

6.2.1 User Background

Two bioinformaticians worked closely together to analyze the data and interpret the results. A post-doc was mainly in charge of performing the bioinformatics data analysis using software visualization tools. A bioinformatics faculty member supervised the overall analysis. Though not new to microarray technology, the bioinformaticians had little previous experience with the specific software tools used for this data analysis. Later in their analysis, they collaborated with a larger group of biologists to examine broader impacts.

6.2.2 Visualization Tools

The following visualization tools were chosen by the bioinformaticians for data analysis and reporting. Microsoft Excel was also used extensively for data formatting.

- Spotfire® [78] (Figure 6.1)
- PathwayAssist® 3.0 [5] (Figure 6.2)
- GenMapp [4] (Figure 6.3)
- Q-Value Software [127] (Figure 6.4)
- KaleidaGraph [128] (Figure 6.5)

The bioinformaticians started with Spotfire and PathwayAssist 3.0 because software licenses for them were already purchased by their lab. They also tried to use other tools like Affymetrix GCOS [108], and R [129], and different versions of PathwayAssist (2.0, 3.0, and 4.0). They found that they preferred PathwayAssist 3.0. However, they did not search for other software tools rigorously, as on performing some data analysis they felt that both Spotfire and PathwayAssist supported their tasks very well. GenMapp was used because the bioinformaticians liked the mouse signaling pathways provided by that tool. They also used a Q Value software package to minimize false discoveries, and finally, KaleidaGraph to create readable static graphs to present their results.

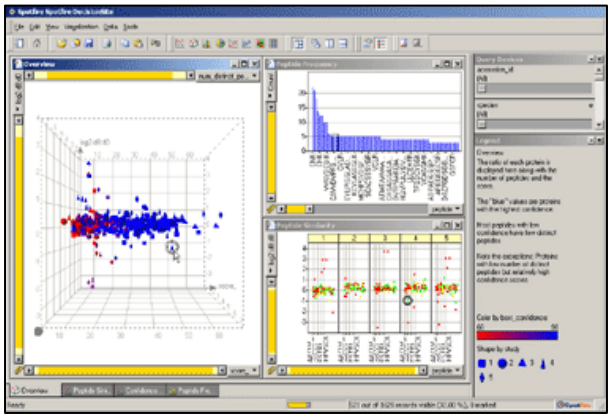


Figure 6.1 Spotfire®.

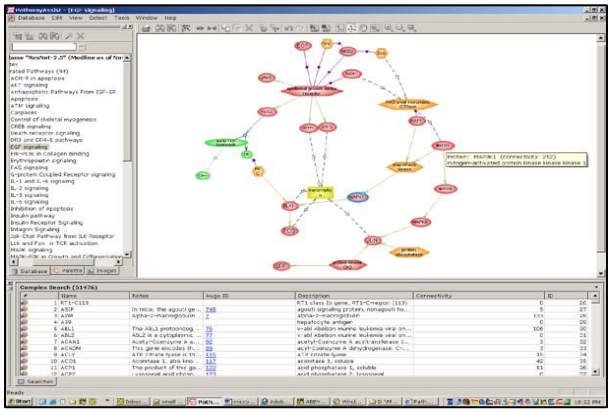


Figure 6.2 PathwayAssist®. PathwayAssist is now known as Pathway Studio.

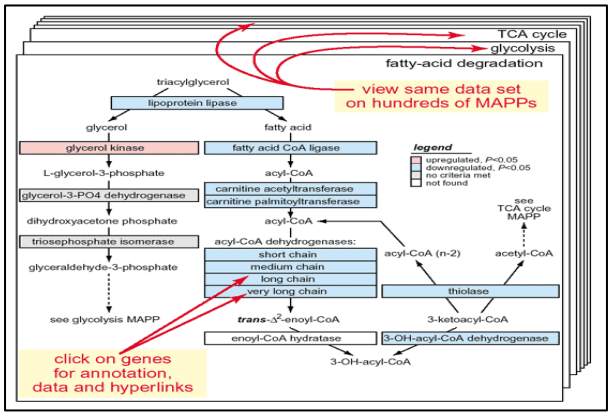


Figure 6.3 GenMapp.

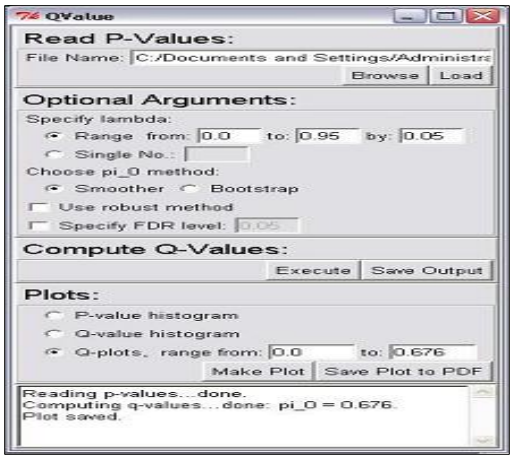


Figure 6.4 Q-Value analysis.

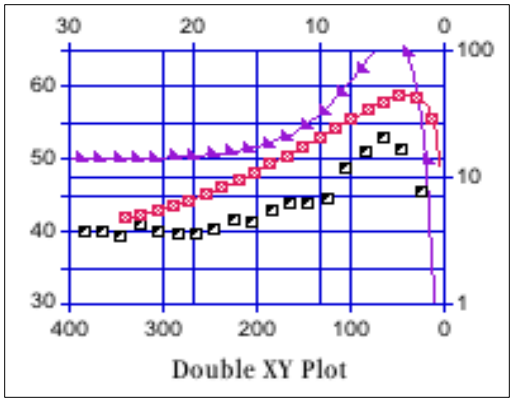


Figure 6.5 KaleidaGraph.

6.2.3 Dataset

The dataset measured mRNA expression levels from mouse lung tissue under 4 different conditions (control group, flu infected, tobacco smoke exposed, both flu infected and tobacco smoke exposed). The measurements were taken for 6 timepoints (6, 20, 30, 48, 72, and 96 hrs) with three replicates for each timepoint, resulting in: 4 conditions X 6 Timepoints X 3 Replicates = 72 data conditions for 45001 probe sets (genes). Thus, the dataset was 45,001 rows X 72 columns.

In general, these bioinformaticians' scientific goal is to understand the pathogenesis of flu infection and the impact of tobacco smoking on that process. Their analysis is exploratory, and is not limited to simply verifying a specific hypothesis.

6.2.4 Protocol

To keep the experiment as close to real world data analysis as possible, we did not require the bioinformaticians to follow an unusual protocol. They were requested to keep a diary of the process they undertook, the insights gained from the data, the visualization and interaction techniques that led to the insights, and the successes and frustrations they experienced with the software tools. We also met regularly, once every 2-3 weeks over a 3 month period, to discuss the data insights and their experience with the tools. The bioinformaticians did not perform data analysis every day, but rather based on how it fit with their normal job activities. However, when analyzing the data they usually spent about 3-4 hours at a time. To judge the significance of insights, at the end of data analysis, we requested the bioinformaticians to rank the data insights on a scale of 1-5, with 5 being the most significant.

An important requirement of the study was that we did not impact their normal data analysis process in any way, except for the diary keeping and debriefing meetings. We ***did not*** provide any help with the software tools or guide their data analysis in any way. The analytic process, the selection of tools, and the data were all determined by their own normal procedures that they had planned regardless of our observation.

6.3 Data Analysis Procedure and Insights

The bioinformaticians started from a raw Affymetrix microarray dataset. They used Microsoft Excel to convert the data into the format they needed for further analysis. Description about different file formats used by Affymetrix and their meaning and

significance can be obtained from [108]. This process was non-trivial and required about 15 hours of extensive data manipulation.

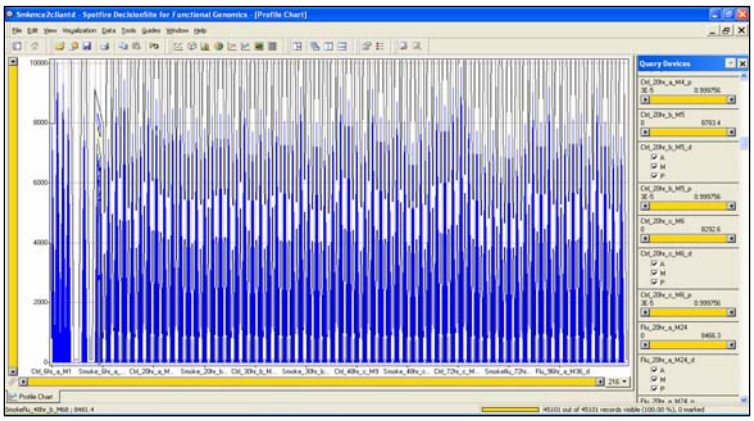


Figure 6.6 The entire dataset (45,001 rows X 72 columns) visualized using Profile chart in Spotfire.

Once the data was in the required format, they loaded it into Spotfire to get an initial overview. Figure 6.6 displays the visualization of the entire dataset (45,001 X 72) using the profile chart (similar to parallel coordinate visualization) provided by Spotfire. To make data analysis more manageable, they decided to filter some genes. They first removed the genes that had absent or null values in the data, by using sorting and column reordering features of Excel. For further filtering, they decided to remove genes that did not show much change from one condition to another, using dynamic queries provided by Spotfire. The final dataset had 30,000 rows.

They began the data analysis by using the scatterplot visualization in Spotfire to plot expression (data values) for each control timepoint with respect to timepoints of the other conditions (Figure 6.7). Each point in Figure 7 corresponds to a probe value (or a row) in the dataset. This was an extremely time consuming process due to combinatorial explosion. They initially wanted to use the profile chart to get an overview, however due to the sheer volume of data they found it confusing due to visual clutter (Figure 6.8 shows the profile chart for Control vs. Flu timepoints). Thus, they had to manually check individual time points to make data size manageable.

One of their data analysis aims was to search for probes, from the entire dataset, that displayed different expression values for selected conditions. Hence, they used scatter plots, since that view made it easier for them to identify outliers that displayed distinct behavior for the selected time points. They also tried to increase the dimensions visualized by coloring the plot using a third dimension. E.g., in Figure 7, though the plot visualizes Control 6 Hrs. vs. Flu 6 Hrs, the color for each dot is based on its expression values for Smoke Exposed at 6 Hrs. However, they found this confusing and focused more on the layout without taking color into account.

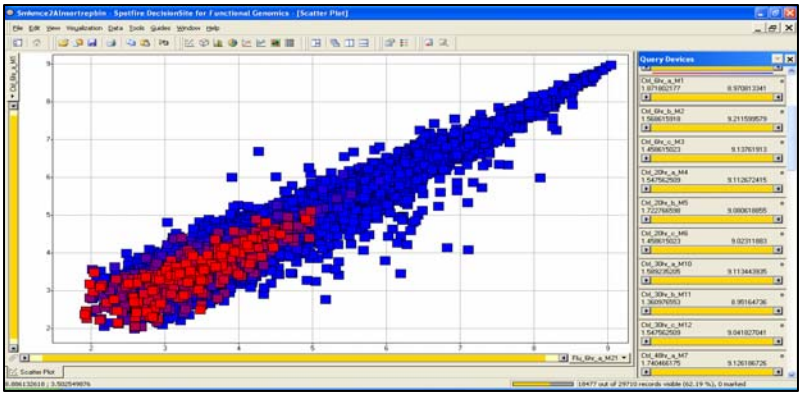


Figure 6.7 Visualization of Control 6 Hrs vs. Flu 6 Hrs using the scatter plot in Spotfire. Each dot in the figure corresponds to a probe (or a row) in the dataset. The color of each dot corresponds to expression value in Smoke Exposed 6 Hrs.

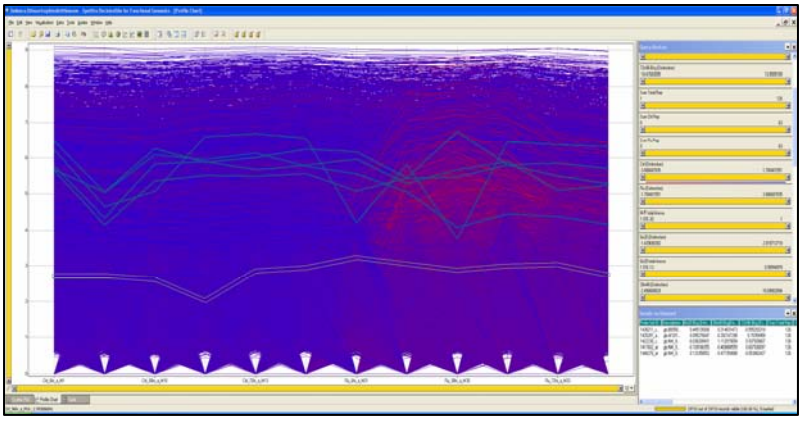


Figure 6.8 Visualization of all the 6 timepoints for Control and Flu conditions in profile chart in Spotfire for 30,000 probes.

They tried to use 3-D scatterplot visualizations (Figure 6.9) to have an overview of more timepoints simultaneously and save some data analysis time. However, they immediately gave up the idea as they had difficulty interpreting the visualization and found it actually took longer for them to think through the meaning this way.

They also tried K-means and SOMS clustering algorithms, and treatment comparison feature provided by Spotfire to get an overview of the common gene expression trends in the data. Figure 6.10 shows the visualization resulting from grouping the data by 3X3 SOMS clustering. They also checked the clusters to verify if various familiar genes displayed the behavior they expected, and if biologically functionally related genes were appropriately grouped together. Table 6.1 lists the insights they obtained using these views.

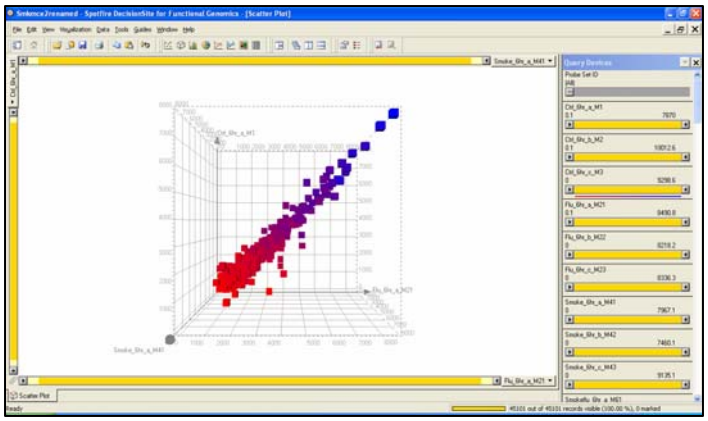


Figure 6.9 Visualization of Control 6 Hrs vs. Flu 6 Hrs vs. Smoke Exposed 6 Hrs using the 3-D scatter plot in Spotfire. The dots are colored based on the expression value in Smoke Exposed + Flu 6 hrs.

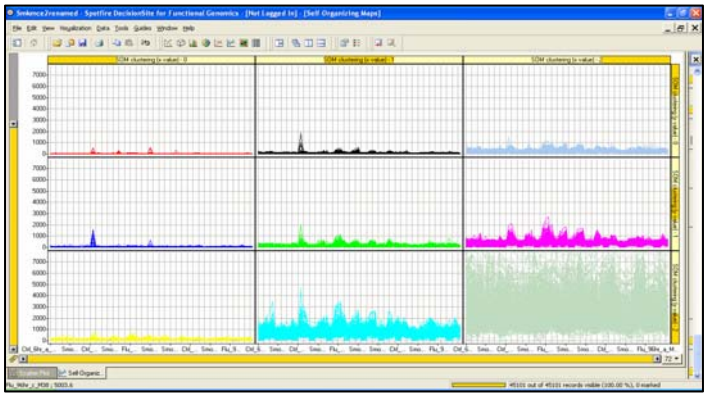


Figure 6.10 A visualization of 3X3 SOMS clustering of the data.

Date	Visualization	Insight	Value
8/12	Scatter Plots	• Noticed very up-regulated genes in flu 96 vs. ctrl 96 on scatter plot. Same effect is seen in the time series.	1
8/12	SOMS Clustering	• Self-organizing-maps at 3x3 grid show some interesting profiles e.g., one where genes are only up-regulated in flu 20 hr and others only up regulated at smoke exposed+flu 30 hr. If these are same genes, then smoking delays flu induction. • Certain matrix metalloproteases are up regulated along with interferon activated genes.	4
8/12	K-Means Clustering	• K mean clustering appear to be much better than SOM in sorting out different dynamics of gene regulation, especially on IFN genes.	3
8/12	Treatment comparisons	• An important proteolytic enzyme of relevance to the group appears to be down regulated in flu, smoke exposed, and smoke exposed +flu. • Several immune system activating genes are all up regulated by smoke exposure	3

Table 6.1 Lists insights gained at the start of data analysis.

The clustering algorithms group genes based on the similarity in their expression profiles. The bioinformaticians were worried to discover that genes with distinct time

profiles were also grouped together. Also, the algorithms do not take into account the biological functionality of the genes. However, they liked the dynamic query interaction method provided by Spotfire as a way to quickly explore many criteria. Hence, they decided to focus on the profile chart and scatter plot visualizations for more detailed analysis. E.g., Figure 6.11 displays a profile chart visualization for all genes that are up regulated for Flu as compared to Control condition. Table 6.2 lists the insights obtained by this process. The bioinformaticians also made a list of interesting genes from the queries and saved them for further investigation.

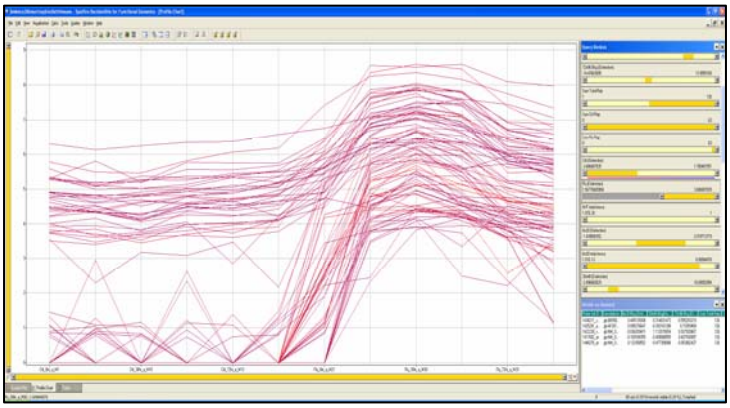


Figure 6.11 Visualizes Control vs. Flu for all 6 time points for one replicate. The display is manipulated to show genes that were up-regulated for flu as compared to control condition.

Date	Visualization	Insight	Value
8/15	Scatter plots	• Heat shock proteins, Retnla, cathepsins, serum amyloid A3, interferon induced proteins, certain matrix based proteins are up-regulated in flu infected mice. Also slower in up-regulated items include MHC molecules by 10 hours (in smoking).	3
8/29	Profile Chart	• Noticed that some heat shock proteins are up-regulated only at 30 hours in control mice. • Cathepsins are upregulated by flu.	4

Table 6.2 Lists insights by using scatter plot and profile chart visualizations along with dynamic queries.

Now that the bioinformaticians were more familiar with the data, they needed different visualizations to get more biologically relevant insights. They decided to use PathwayAssist for further data analysis involving biological pathways. Pathways are network-based models of complex biological processes [12]. They had already made lists of genes they needed to investigate further. They wanted to build pathways involving these genes, using search capabilities provided by PathwayAssist. This would show other genes that have a direct influence on these genes of interest. PathwayAssist uses NLP algorithms to extract information about relationships between genes from various search engines such as PubMed. Figure 6.12 shows an initial pathway created for genes they selected. Since the visualization had more information than they could handle, they abandoned the idea of depending on pathways created automatically. Also, they cannot completely trust the automatic pathways created by the tool. They would have to

manually curate the pathways, because the NLP algorithms usually provide some level of information that is irrelevant to their data analysis or is incorrect.

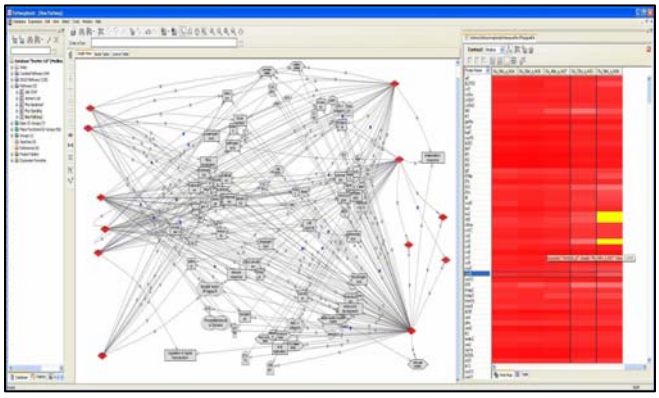


Figure 6.12 Automatic pathway created by PathwayAssist for a selected list of genes.

The bioinformaticians decided to focus on the apoptosis signaling pathway because their research group is most interested in that topic. They knew that GenMapp provides pre-built pathways. Although they preferred the pathway provided by GenMapp, they had problems overlaying the expression data onto it. The expression data manager in GenMapp (Figure 6.13) required them to define color scales for each individual column. Since they had 72 columns, they thought this would be time consuming. They decided to transport the GenMapp pathway to PathwayAssist and then link it to the microarray data. This involved importing genes and reconstructing the pathway.

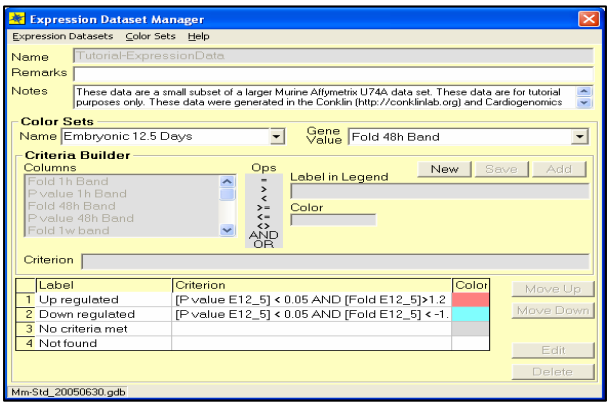


Figure 6.13 Expression Dataset Manager in GenMapp.

They utilized the heatmap visualization provided by PathwayAssist to investigate time-dependent regulation of the pathway. They found it easier to click on the column name to display data related to a particular condition on pathways in PathwayAssist (Figure 16.4). Using this, they found genes that were suppressed by smoke exposure but up regulated in flu. Table 6.3 lists insights resulting from pathway visualization.

Figure 6.14 Data from Flu 20 Hrs is overlaid on cell apoptosis pathway and linked to heatmap visualization in PathwayAssist. The color is used to denote expression value, red denotes up-regulated genes whereas green implies down regulation.

Date	Visualization	Insight	Value
9/01	Heatmap	• A list of pathway genes that are suppressed by smoking but up-regulated by flu.	4
9/12 9/13 9/21	Pathway visualization	• The up-regulation of Mx by flu is suppressed by smoking even though smoking itself did not have an effect on basal Mx activity. • Genes involved in apoptosis are regulated, particularly DAXX which is up-regulated in flu infections. • Flipping through time points on PA, noticed that CHUK and IRAK1 of the NFKB signaling is only up-regulated in flu vs. control.	3

Table 6.3 Lists insights resulting from using the heatmap and pathway visualization

Along with the data analysis, the bioinformaticians also became more familiar with additional features and functionalities of the visualization tools by reading the help documentation and calling technical support. They used profile search (Figure 6.15) for genes that display expression similar to a specified pattern, and statistical analysis methods such as t-tests and Anovas. Table 6.4 lists insights from this process.

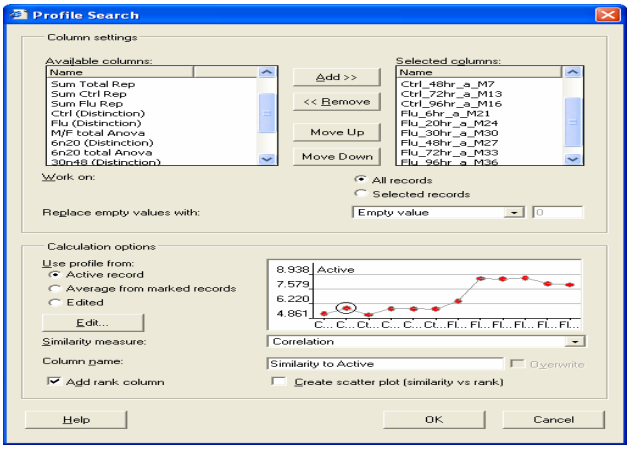


Figure 6.15 Profile search feature in Spotfire.

Date	Visualization	Insight	Value
9/28	Profile Search	• Used profile search to find genes that are regulated (over all conditions) similar to Mx. • Smoke exposure REALLY suppresses some heat shock accessory proteins. The NKB system is responding similarly. Maybe through Tol receptors?	3
9/28	Data-Pattern-Distinction	• Few distinctive genes that may be indicative of smoking. There are several candidates. • Influenza infection is typified by the up-regulation of certain genes that are activated very early by interferon. • Retnla has a very interesting profile, up-regulated in flu and EARLY in smoking recovery! VERY interesting.	5
9/30	Anovas, t-tests	• T-test/ANOVA shows that three genes are the optimal indicators of smoking, including flu infected individuals.	4

Table 6.4 Lists insights obtained from profile search and statistical analysis.

Table 6.5 lists the process towards the end of their data analysis. The bioinformaticians were trying new methods to get more insights from the data to ensure that they do not unintentionally neglect any unexpected results. The complexity of the procedure indicates more familiarity with detailed features of the tools. They were also refining their findings to ensure the most accurate insights.

Towards the end of the study, the bioinformaticians were evaluating the best statistical tests to apply to the data. In addition to using t-tests and p-values to minimize the number of false-positive tests, they were also using q-value analysis to minimize the number of false discoveries [130]. The q-value offers a less conservative approach to measuring the statistical significance of genomic data than the traditional Bonferonii-corrected p-value. Although Microsoft Excel supports this analysis, they felt that the Q-value software was more suited for bioinformatics data analysis.

Date	Visualization	Data Analysis Procedure
10/31	Data formatting + Pattern Distinction + Biology Database search	• Removed absent call data points and used the discovered binary sorting to count replicate present calls. Using distinction factor (correlated to t-test) to find flu indicators in Spotfire and then export to Pathway Assist to find biological significance.
11/01	Profile search	• Examining profile search and treatment distinctions in Spotfire. Trying to find the best way to differentiate different time profiles of expression. Particularly, should absent calls be considered 0 or null?

Table 6.5 Lists the later data analysis procedures.

6.4 Insight Presentation

The bioinformaticians work in close collaboration with another large international biological research group. They recently presented their data analysis results to the other group. Most of their presentation was related to immune system genes and used Microsoft PowerPoint slides. Since the international group is less conversant with microarray data analysis, the bioinformaticians shared their time-series data analysis experiences including data filtering and normalization methods.

The international group is primarily interested in chronic respiratory diseases, not flu infections per se. Hence, they have a different set of genes of interest than the bioinformaticians. However, the bioinformaticians were able to easily provide

information about the other genes during their presentation and later meetings, by using the data filtering capabilities of Spotfire. Spotfire allowed them to easily narrow down the genes of interest using text filters. For a given text string, Spotfire can list all the genes containing that text. Many genes having similar functionalities have similar names like Casp1, Casp2, etc. The search capabilities worked well to find such groups. Spotfire also let them analyze the time profiles for selected genes (Figure 6.16). The audience found the dynamic query mechanism provided by Spotfire to be helpful because it allowed them to search for genes based on the expression values. They also performed t-testing on the fly to check significance of the results.

Based on this interactive collaboration, they discovered that smoking suppresses expression of Slfn genes. This finding was considered very exciting, since not much data is currently available for genes belonging to that family. They also concluded that smoking suppresses expression of genes involved in DNA repair and those that facilitate cell cycle (insight value = 5).

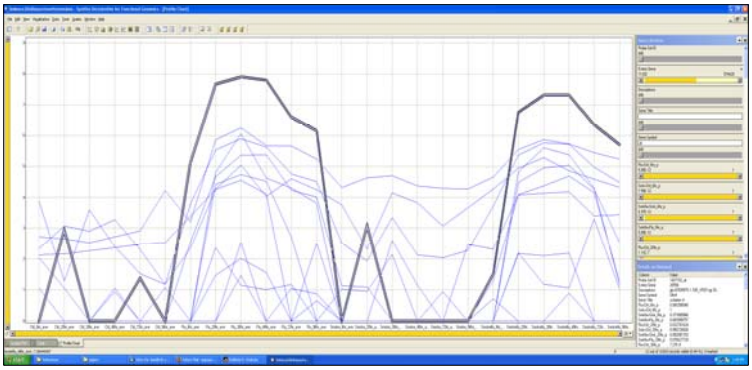


Figure 6.16 Time profiles for a group of selected genes. The profile indicates suppressed values for Smoking + Flu condition as compared to the Flu condition.

The biologists are currently working towards publishing their data analysis results. Their main conclusions are that smoke exposure suppresses overall gene expression under conditions of flu infection. They will also report a list of genes that were found to be significantly affected, the biological functions of these genes and the overall significance of these effects on biological processes. There was one major new insight involving the DNA repair mechanisms that will be explored in future collaborative work. It should be emphasized that, despite use of software tools, a significant amount of manual exploration and the input of several biological domain experts was necessary to derive useful biological understanding from the experiment data.

They will use KaleidaGraph to graph the results and gene expression time profiles. Though other software tools allow them to easily transfer screenshots to Microsoft Word, they are accustomed to KaleidaGraph and find it better suited for simple static data presentation. They also prefer the quality of images in terms of print resolution. KaleidaGraph also provides better capabilities to manipulate graph display details, such as labeling, for presenting information. An example of a graph presenting expression profiles for a selected gene using KaleidaGraph is shown in Figure 6.17. The profiles are color coded based on the four main conditions in the experimental paradigm.

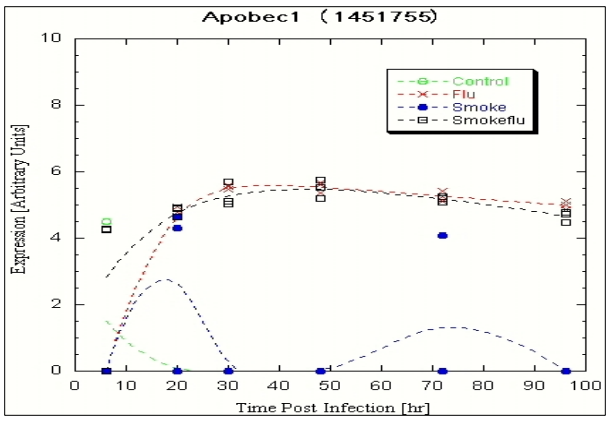


Figure 6.17 Time profiles for a selected gene in KaleidaGraph. The profile is divided into four different colored lines to represent the 4 conditions in the data.

6.5 Discussion

6.5.1 Data Analysis Procedure

Microarray experiments result in very large datasets that require extensive preprocessing before they can be analyzed for insight. The bioinformaticians spent about 15 hours formatting the data in Microsoft Excel. Excel was used because it provides an extensive and efficient functionality for data manipulation. Another reason for selecting Excel was the bioinformaticians' familiarity with it. They commented that even though a visualization tool may provide ways to manipulate data, they preferred Excel to save time learning new software.

Once they had formatted data they needed help to load the data into Spotfire. They had separate files for information relate to the genes and their expression values. They called Spotfire technical support to figure out a way to efficiently combine both the datasets into one so that they could proceed with data analysis. Although Spotfire allowed them to combine more than one column by taking averages, they did not find a mechanism for row arithmetic. To get around this limitation, the bioinformaticians had to format the data in Excel and re-import it to Spotfire.

The first step in data analysis described here was getting familiar with the data. The bioinformaticians used multiple visualization representations in Spotfire for this. They started initially by using scatter plots, profile charts, and cluster visualizations provided by Spotfire, and eventually used features such as statistical algorithms and data pattern distinction for more complex analysis. They seemed to follow the general HCI approach of "overview, zoom, and filter" in their process. For biologically relevant insights they found Spotfire alone inadequate and needed to rely on other tools, as well as domain experts. They used the list of genes selected in Spotfire to examine their biological functionalities and relationships with other genes in PathwayAssist. Towards the end of their process, they were trying different statistical analysis methods to ensure more accurate and statistically sound results. Moreover, they needed to ascertain that their results were robust to various choices that are commonly made in the community. There is no single accepted method as yet for microarray data analysis and their process reflected their professional judgment.

Once, the data analysis is completed, it is equally important to have an efficient mechanism to present information. Though the bioinformaticians worked with Spotfire and PathwayAssist for analyzing the data, they needed to use yet another tool, KaleidaGraph for creating readable graphs to present their results. They found both Spotfire and Microsoft Excel inadequate to create all data representations needed for publication, although they will use exported plots from both SpotFire and Pathway Assist in their published research report.

Thus, the bioinformaticians used a combination of different software tools during the course of their data analysis process. They picked out key features from different tools for different purposes. The use of multiple visualization tools required the bioinformaticians to export and import information, several times, from one tool to another. This required additional data formatting that was time consuming. Hence, it is important to provide better interfaces for the software tools to facilitate data exchange between them.

6.5.2 Effect of Interaction Mechanism

The bioinformaticians used multiple visual representations in Spotfire to get an initial overview of the data. One of the main reasons that they spent time exploring the data in Spotfire, using scatter plots and profile charts, is the dynamic queries provided by the tool. They said that this might have even motivated them to spend more time with the visualization then they wanted initially. Dynamic queries provided an efficient way for them to manage a large amount of data. Rather than worrying about 45,001 probe-set values they could easily focus on the genes of their interest.

For pathway analysis, the bioinformaticians preferred PathwayAssist because the tool allowed them to easily overlay data values for a selected condition on the pathway of interest using color coding. The tool also automatically filtered out all the genes that did not belong to the pathways. Though seemingly trivial, this was one of the main reasons that encouraged the analysts to continue working with these tools. Filtering is critical for making the dataset tractable for human exploration, yet they worried that they may have been missing important information in this process. They did not rigorously search for other software visualization tools, as they felt both these tools supported their tasks well.

Spotfire served well for providing dynamic queries. Even during their short term presentations to the international collaborators, the tool provided them with an efficient way to highlight genes of interest for other researchers. Using text search mechanisms the bioinformaticians could easily create lists of genes of interest to the other researchers and also display their time profiles.

It is important to maintain a history of user actions, and provide replication capability. The bioinformaticians spent a lot of time rearranging the data columns in Spotfire to visualize timepoints of interest next to each other. However, each time they restarted Spotfire, these rearrangements were lost and they had to redo them again. They had the same experience with zooming on areas of interest within visualizations. For example, when they changed the data columns in scatter plots, the zoom position was lost. Thus, these seemingly minor usability problems that developers might not have considered important had a major effect on these bioinformaticians when they had to repeat arduous operations 72 times.

Thus, methods to efficiently interact with and selectively filter the data to focus on points of interest were considered equally or even more important than the visual representations. In fact, they tended to prefer the more simple visual representations. The bioinformaticians had no trouble restarting data analysis with the selected tools even after a gap of a few days. An efficient interaction method can make the entire experience with the visualization tool, and thus the insight generation process, more rich and enjoyable.

6.5.3 The Process of Insight Generation

In the reported scenario, the subjects started the data analysis process by searching for potential insights. They did not have a prior list of specific hypotheses to validate, although they used the past 20 years of interferon research as a benchmark for validating their experimental results. They wanted to find as many interesting facts as possible in the data for more detailed later exploration. Though they wanted to use profile charts, they found the scatterplot visualization more informative due to the data size (Figures 6.7 and 6.8). However, gaining an overview of the data by examining only 2 columns at a time was time consuming.

They used clustering algorithms and treatment comparison features to get an initial idea about various patterns in the data. Most of these visualization features were not considered difficult to learn. But there were many steps to execute and many combinations to explore. It is also important to interpret results from each combination in terms of biological domain knowledge to ensure that the results make sense. In fact, the most novel insights were not revealed directly by tools, but by experienced investigators who connected the patterns of changes in two particular pathways to their prior knowledge about the underlying biological processes.

An important process is not just analyzing data using different combinations, but also interpreting which combination is best suited to analyze a particular dataset. For instance, Spotfire provides several different clustering algorithms including SOMS, k-means, and hierarchical clustering. Interpreting how each method groups the genes, and resolving conflicting results from these methods can take time. Similarly, different normalization methods yield different results from the data. Hence, selecting the appropriate method depends on understanding how each method affects the data in terms of experimental context. This clearly suggests the influence of domain knowledge on data analysis.

The bioinformaticians decided not to rely on clustering algorithms for data analysis because the algorithms grouped genes with non-distinct time profiles in similar groups, and the algorithms did not take biological functions of the genes into account. Later, the bioinformaticians used profile chart visualizations to explore the data in more details. The visual representation along with the dynamic query interaction mechanism provided a valuable combination to explore the data. They found they could easily combine many different queries to filter data, resulting in a high user satisfaction. They used this technique to find a list of interesting genes specific to a particular biological function to focus on. They were especially interested in finding genes that were differentially expressed in smoke exposure + flu condition as compared to the flu condition. This would indicate infection-related genes that were affected by smoking. They spent about 1-2 weeks exploring the data in profile charts. They needed the experience of exploring many possible combinations to simply observe all facets of the data. This gave them confidence in their coverage, and resulted in some serendipitous findings as well.

For domain specific information, they needed more biologically relevant visualizations. Though Spotfire ontology gave them some clues about patterns of expression for functionally related genes, it was not sufficient. The bioinformaticians needed to see the interactions of the genes that they selected with respect to other genes that have a direct influence on them. They decided to use pathway visualizations in PathwayAssist for this. They initially decided to use the gene list to create pathways automatically. However, since the queries resulted in too much information that was difficult to comprehend and interact with in the visualization (Figure 6.12), they decided to manually curate pathways. The process of pathway analysis was more complex and required about 2-3 weeks of data analysis and interaction with the tools. In general, it seemed a constant struggle for the bioinformaticians to continually reduce the complexity of the data to a comprehendible amount. Even with the use of visualization tools, they were forced to focus on smaller pieces so that they could wrap their minds around the observed biological behaviors.

The bioinformaticians found the most exciting insights after almost 1.5 months of data analysis and several months of "learning" time with the software. However, from the values of insights reported earlier, it is clear that later analysis is influenced by findings from the earlier analysis. Also, the bioinformaticians used more complex queries and features in the tools to reach them. This suggests more familiarity and confidence with the tools. Moreover, they feel that despite a state-of-the-art analysis, there is much untapped information waiting for mining by different domain experts.

Once they were done with pathway analysis, they then used other visualizations in Spotfire to ensure that they did not unintentionally miss any unexpected insight from the data. The later data analysis process dealt with analyzing their insights and to ensure correct statistical interpretation. They also tried another data formatting method to check if this resulted in any other insights or conflicts with earlier observations. Their most recent data analysis involves capabilities of more than one visualization tool simultaneously, requiring a lot of back-and-forth processing.

From the discussion it is clear that the choice of visualization methods used to analyze the data is based on the subjects' domain knowledge. Discovering an appropriate visual representation and procedure to interpret the data could be considered procedural insight. This is usually a non-trivial task, and requires trial-and-error attempts with many combinations. The subjects reported that in the future they will be able to analyze a similar dataset in a relatively shorter time. Such use of learned domain knowledge is very difficult to reproduce in short-term experiments.

6.6 Short Term vs. Long Term Insight-Based Studies

Three user-studies (Chapters 4 – 6) were performed using the insight-based method. Two studies investigated the use of visualization tools in a controlled laboratory-type setting, whereas one study investigated the actual real world long time visualization usage. The short term insight studies had the following three limitations

1. The study measured insight from short term usage. In real world scenarios, biologists spend days, weeks and even months analyzing data. Long-term insight may be very different than short term insight. Long term insight can provide broader understanding that guides biologists through multiple cycles of microarray experiments.

2. The participants in the study were unfamiliar with the data, and not personally invested in its creation. The only background knowledge they had was what we provided during the course of study. It was very difficult to appreciate the biological relevance of the microarray data they were analyzing. Hence, the hypotheses they reported were more speculative. Yet, the insights were not trivial, which suggests that the visualizations are provoking users to think deeply about the data and to apply the insight in their domain.
3. Each participant was unfamiliar with the visualization tool that they used. Gaining expertise with a visualization tool may change the method in which it is used and the insight it provides.

To address these issues we performed a longitudinal study to analyze how visualizations are used to get insight into the data. For this we worked with bioinformaticians who were beginning data analysis. Since they had an undeniable motivation in performing the data analysis, it was possible for us to observe the process for an extended period of time. It would have been impossible to perform this study if the subjects were not intrinsically interested in the data. Also, they were able to provide us with more meaningful feedback about insights and their utility.

To keep the data analysis as natural as possible, we worked primarily through a research diary maintained by the subjects. This saved us from having to continuously observe the user. It also indicates the viability of a self-reporting approach to longitudinal insight studies. The bioinformaticians did not have to do anything difficult beyond maintaining the research diary, in which they noted insights and captured screen shots. Most of these notes are things they would want to capture anyway. The data analysis process proceeded according to their normal job activity. Thus, the longitudinal study requirements were very light in extra effort for subjects and straightforwardly manageable for the evaluators. Moreover, the HCI investigators have a long term relationship with the bioinformaticians and have observed many aspects of their process in other contexts over approximately three years.

As the study was over an extended period of time, it was possible for us to study the long term insight generation process. We were also able to observe the use of different visual representations and interaction techniques over a long period of time. Due to their familiarity, the subjects could provide more relevant feedback about their insights and about the visualization tools and their limitations, including long-term usability problems. Thus, the longitudinal study enabled observations that would not have been possible in a short term study. Table 6.7 lists general comparisons between the short term vs. long term insight-based studies.

Comparison Factor	Short term study	Longitudinal study
No of Users	Can be large. The study reported in Chapters 4 and 5 used 30 participants each.	Limited. The study reported here used about 2 participants. It is much more difficult to find participants willing to commit for longer period of time for the study.
Experiment Protocol	The experiment protocol can be controlled as the experiment uses lab type of setting. The investigator is responsible to identify data insights. This can make insight collection	The participants for the study are responsible for keeping track of data insights. This increases work load for the participants. Since, the insights are discussed once in a week or two weeks,

	process more uniform.	most of the participants' comments are in retrospection.
Visualization Tools	It is possible to evaluate more number of tools with this type of study.	The number of tools used will be limited and also based on participants preferences.
Datasets	Visualization tools can be evaluated on different types of datasets. This allows the evaluators to check if the visualization tool is biased towards a particular dataset.	The number of datasets that will be evaluated using this approach is limited. Most often the participants have just one dataset to work with. Also, the comparisons of the tools and insights will be over different datasets since all the participants are working with their own data.
Statistical Analysis	It is possible to perform statistical analysis across the insight categories.	Since the number of participants is limited, it is difficult to perform any statistical analysis.
User Motivation	Participants work with pre-selected datasets. Also since most often the data selected is not theirs, they have limited motivation for the experimental study.	Most often the participants will be working with their own datasets that they want to analyze. This increases their motivation to work with the visualization tools.
Insight Value	Since the participants are not working with their datasets, it will be difficult for them to judge the value of their reported insights. The insights need to be coded by an independent domain expert.	The participants mostly work with their own datasets, hence they can determine the value of their insights.

Table 6.6 Comparison of the short term insight study to the longitudinal insight study.

7 Conclusions & Future Work

7.1 Insight-Based Approach

The main purpose of a visualization tool is to provide insight. This can be difficult to measure quantifiably. Although the definition and the quantifiable characteristics of an 'insight' identified in Chapter 4 are not comprehensive, they provide an approximation of users' learning from the data using visualization tools. This, in turn, enabled us as evaluators to gain insight into the effectiveness of the visualization tools. The technique evaluates users' findings from the data. More, valuable, faster, and deeper data findings correspond to more effective visualizations as it suggests users can gain more *insight* into the data with the tool.

Unlike the task-based method, the insight-based method uses an unguided data analysis protocol. The visualization evaluators record open-ended insight generation by not restricting users to a set of preplanned benchmark tasks. The recorded insights provide a way to analyze the kinds of data insights are motivated by the visualization tool. The insight-based method cannot indicate what kinds of insights are not supported by visualization tools. If the participants fail to report certain kinds of insights that would usually mean that the participants failed to notice such insights on their own. Providing tasks in such instances can force the participants to perform the corresponding data analyses with the tools that they may miss on their own. The insight method thus, addresses the research question: "What kind of insights does a visualization tool motivate?" Whereas the task-based method allows determining for a particular type of task(s), the visualization tool that lets participants complete task(s) most accurately. The accuracy may be more reliable means of determining which kind of tasks that the visualization tools supports vs. does not support.

The task-based method requires pre-selection of the tasks to evaluate the visualization tool. It is possible that the evaluators may select the tasks that are most interesting to them or select tasks to bias results towards a particular visualization tool. Also, though the task-based method provides feedback for the visualization tools on pre-selected tasks fails to report reliably for unselected tasks. Due to unguided protocol, it is possible that given a dataset two participants may analyze the dataset in different ways and report different insights. With sufficiently large number of participants this can cover several different types of possible insights for the visualization tools. Thus, the study may often provide feedback for the kinds of analysis that the evaluators may have not thought about earlier. Although, since the participants report just the insights for the insight-based method a higher level analysis of the reported insights such as grouping them into different categories or assigning domain value is needed. It is possible that two different evaluators may analyze the reported insights differently resulting in different conclusions about the visualization tools. Therein lies the strength and the weakness of the insight method. The subjective group makes the insight method less uniform, however different insights specified by different participants suggest different possible insights and the tool usage by the visualization tool. Also, it helps to determine if participants' with different backgrounds may use the visualization tool in different ways.

Besides the short term study, a longitudinal study was also performed with the insight method. The long term insight-based method provides a more detailed feedback as to

how different data representations along with the interaction mechanisms and features provided by a visualization tool were used by an actual end user for data analysis. It provides feedback on different kinds of insights that were found using the tools and the value of these insights. Though, it will be difficult to make statistical comparisons about the tools with the long term study, such a detailed report can be used by developers and tool designers to prioritize end user data analysis tasks and requirements based on the kinds of insights that are needed. The diary recordings display viability of self-reporting insights.

7.2 Pathway + Microarray Data Visualization

The design space presented in Chapter 3 was divided in two dimensions. Dimension 1 was based on the methods to overlay data on pathway nodes. Dimension 2 was based on the number of visualization alternatives used besides the pathway diagrams. There are six possible alternatives based on Dimension 1 and Dimension 2. Two studies (Chapters 3 and 5) were performed to evaluate the visualization alternatives.

7.2.1 Data Overlay Method

The design space in Chapter 3 presented 3 visualization options for the data overlay method for pathway + microarray data visualization tools.

Single Attribute Visualization: The single attribute visualization displayed the multidimensional data one attribute at a time on the pathway nodes. The participants were provided with sliders to iterate through the graph to analyze data for other timepoints. For both the experiments, participants using single attribute visualization provided more controlled performance as compared to the participants using other visualization types. The participants using this visualization tool were neither slower nor less accurate for any of the selected tasks. The participants using this alternative had better performance for tasks involving single timepoint analysis for both the experiments. Our hypothesis prior to conducting the experiment was that the participants using single attribute visualization method may perform slower for tasks involving overall graph analysis as they needed to iterate through all the data attributes one at a time. However, from the results we found that there were no differences in the performance time between the participants using single attribute and other types of visualization alternatives.

Both the studies described in this dissertation were conducted with pathway visualization alternatives with only about 50 nodes. However, with new visualization methods the number of nodes that the biologists work with is increasing. Also, often the number of data attributes is larger than 10-15 used for the studies here. For instance, in the longitudinal study (Chapter 6), the participants worked with data having 72 attributes. With large number of pathway nodes and data attributes, tools using single attribute visualization may be simpler to understand as the other visualization alternatives may become too complicated to understand and more difficult to scale.

Multiple Attribute Visualization: The multiple attribute visualization alternatives displayed all the multidimensional data attributes on the pathway nodes using nodes-as-glyphs method. The mouse over displayed both the numerical values and the corresponding timepoint for the nodes. This allowed faster comparisons between node

values for different data attributes. The multiple attribute visualization method provided faster analyses for identifying outlier nodes i.e. nodes that display different behavior than the other nodes in the graph. However, displaying all the attributes simultaneously on the graph nodes made it difficult for the participants to focus on the overall graph expression at a single data attribute. Also, the participants had difficultly identifying most interesting attributes on the data to focus on.

Multiple Graph Visualization: Multiple graph visualization alternative displayed all the attributes of the visualization using graph-as-glyph approach. The mouse over and the visualization displayed the timepoint corresponding to the attribute and the numerical value of the node. This allowed faster searching for the interesting timepoints. This also made analysis of the overall graph expression easier. Though, it made easier for participants to select the attribute of interest we found that participants had difficulty focusing on the graph layout and performing analysis for expression values of a single node.

7.2.2 Single View vs. Multiple Views

The study described in Chapter 3 evaluated single vs. multiple views for pathway diagrams. The multiple views used other visualization methods besides the pathway diagrams. We found that participants could perform single attribute analysis using just the single pathway diagram view. This allowed them to not only analyze the expression values for the graph nodes but also check this in relation to the other nodes. The multiple views were found to be helpful for analyzing exact values of the nodes for different data attributes. Overall, the participant in the study didn't make much use of the additional visualization alternative. We found that this may be as the data analysis was simplistic.

The use of multiple views was more visible for the longitudinal study. The participants used the non-pathway diagrams to perform overall data analysis with a larger set of the nodes. They used the other views to find more interesting timepoints and genes in the data. These were then short listed to analyze in the pathway diagrams to see their influence on the other nodes that are directly related to them. Thus, though the participants worked with more number of visualization alternatives in the longitudinal study this was more contextual based. The non-pathway visualization alternatives such as heatmap and parallel-co-ordinate were to get familiarity with the data, whereas the pathway diagrams were for more biologically relevant analysis.

7.3 Future Work

7.3.1 Evaluation Research

Recently a new field 'visual analytics' has been defined. Visual analytics focuses on different factors of the visualization tools that make these tools effective for the end user data analysis [104]. The field also calls for a better understanding of the data representation methods and interaction mechanisms. Such an understanding is critical to create insightful visualization tools that best supports users' data analysis tasks. Clearly, the insight-based method reported here is just a start. Though, we identified several characteristics of an insight, this definition is by no means comprehensive. The protocol identified here, provides an alternate mean to traditional task-based approach for

evaluating visualization tools. More methods that rely on real world data analysis scenarios are needed that provide better feedback about the characteristics of visualization tools that lead to better insights. This also requires a more in-depth understanding of the end user needs for visualization usage, better definition for the term 'insight' and the process of insight generation. Along with the short term studies, the longitudinal study reported in Chapter 6th stresses the need for long term visualization evaluation. Such long term studies allow detailed observations about the visualization tools that are not possible in the short term studies. They are also better representative of the actual real world visualization usage. More studies need to be conducted to investigate how reliable are the short term study results for long term real world visualization usage.

Besides evaluating visualization alternatives, the study in Chapter 5 also compared two different protocols for short-term visualization usage. Such comparison between the experimental protocols is not frequent for the field of Information Visualization. With the development of new protocols it will be important to decide factors to consider for comparing different evaluation protocols.

7.3.2 Visualization for Biological Pathways

The research agenda presented in Chapter 2 identified several requirements for pathway visualization tools. A research space was identified for overlaying microarray data on pathway diagrams. Evaluation studies were then conducted on different alternatives. Several visualization problems still need to be addressed to create more biologically relevant pathway visualization tools. The studies here just provide a start, with better computation methods available analysts now work with larger visualization and more data attributes. This requires more sophisticated means to overview and analyze inter-connectivity between pathway diagrams.

In addition to the microarray data, other biologically relevant information to be overlaid on the pathway diagrams to make them more meaningful to the biologists. Besides the nodes, the links too in the pathway diagrams have large amount of information associated with them. With all these information, better means of pathway visualization methods may be needed. Most of the studies have taken spatial properties of the pathways in nature. Besides spatial, the pathway components have temporal properties associated to them. Depicting temporal states of the pathway components is complicated and challenging research problem.

Reference

[1] H. Lodish, A. Berk, S. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular Cell Biology*: W. H. Freeman, New York, and Houndsmills, 2000.

[2] "Kanehisha Laboratories, KEGG: Kyoto Encyclopedia of Genes and Genomes," *http://www.genome.jp/kegg/*, 2006.

[3] "BIOCARTA, Charting Pathways of life," *www.biocarta.com* 2006.

[4] "GenMapp," *www.genmapp.org*, 2006.

[5] "PathwayAssist™, Ariadne Genomics," *http://www.ariadnegenomics.com/products/pathway.html*, 2006.

[6] G. Churchill, "Fundamentals of experimental design for cDNA microarrays," *Nature Genetics*, vol. 32, pp. 490-495, 2002.

[7] D. Duggan, B. Bittner, Y. Chen, P. Meltzer, and J. Trent, "Expression profiling using cDNA microarrays," *Nature Genetics*, vol. 21, pp. 11-19 1999.

[8] J. Quackenbush, "Microarray data normalization and Transformation," *Nature Genetics*, vol. 32, pp. 496-501, 2002.

[9] L. Shi, " DNA Microarray – Genome Chip," http://www.gene-chips.com/GeneChips.html#What, 2006.

[10] H. Brown, "The real value of Microarray technology," *The lancet Oncology*, vol. 4, 2003.

[11] S. E. Dobrin and D. Stephan, "Integrating microarrays into disease-gene identification strategies," *Expert Review of Molecular Diagnostics*, vol. 3, pp. 375-385, 2003.

[12] P. Saraiya, C. North, and K. Duca, "Visualization of Biological Pathways: Requirements Analysis, Systems Evaluation, and Research Agenda," *Information Visualization*, vol. 4, 2005.

[13] P. Saraiya, C. North, and K. Duca, "An Insight-based Methodology for Evaluating Bioinformatics Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, 2005.

[14] N. Bolshakova, "Microarray Software Catalogue," *http://www.cs.tcd.ie/Nadia.Bolshakova/softwaretotal.html*, 2006.

[15] Y. Leung, "Functional Genomics," *http://genomicshome.com*, 2006.

[16] S. Bridgeman and R. Tamassia, "Difference Metrics for Interactive Orthogonal Graph Drawing Algorithms," *Journal of Graph Algorithms and Applications*, vol. 4, pp. 47-74, 2000.

[17] M. Ghoniem, J.-D. Fekete, and P. Castagliola, "A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations," presented at The IEEE Symposium on Information Visualization, pp. 17-24, 2004.

[18] P. Pirolli, S. Card, and M. M. Van Der Wege, "Visual information foraging in a focus + context visualization," Proceedings of The ACM Conference on Human Factors in Computing Systems (CHI '01), pp. 506-513, 2001.

[19] H. C. Purchase, "Metrics for Graph Drawing Aesthetics," *Journal of Visual Languages and Computing*, vol. 13, pp. 501-516, 2002.

[20] C. Ware, H. Purchase, L. Colpoys, and M. McGill, "Cognitive Measurements of Graph Aesthetics," *Information Visualization*, vol. 1, pp. 103-110, 2002.

[21] B. Jordan and A. Henderson, "Interaction Analysis: Foundations and Practise," *The Journal of the Learning Sciences*, vol. 4, pp. 39-03, 1995.
[22] A. Collins, D. Joseph, and K. Bielaczyc, "Design Research: Theoretical and Methodological Issues," *Journal of the Learning Sciences*, vol. 13, pp. 15-42, 2004.
[23] C. Chen and R. Rada, "Modelling situated actions in collaborative hypertext databases," *Journal of Computer-Mediated Communication*, vol. 2, 1996.
[24] S. A. R. Scrivener, S. Pongut Urquijo, and H. K. Palmn, "The Use of Breakdown Analysis in Synchronous CSCW System Design," presented at The Thomas (Ed.)CSCW; Requirements and Evaluation, pp. 281-293, 1996.
[25] P. Saraiya, C. North, and K. Duca, "Evaluation of microarray visualization tools for biological insight," presented at The IEEE Symposium on Information Visualization, pp. 1-8, 2004.
[26] M. Gerhard, "On Representation of Pathways," *BioSystems*, vol. 47, pp. 1-7, 1998.
[27] R. Stevens, C. Goble, P. Baker, and A. Brass, "A classification of tasks in Bioinformatics," *Bioinformatics*, vol. 17, pp. 180-188, 2001.
[28] M. Rosson and J. Carroll, *Usability Engineering: Scenario-Based Development of Human Computer Interaction*: Morgan Kauffman, 2001.
[29] J. Nielsen, "Finding usability problems through heuristic evaluation," Proceedings of The ACM Conference on Human Factors in Computing Systems (CHI' 92), pp. 373-380, 1992.
[30] "Cytoscape," *www.cytoscape.org*, 2006.
[31] J. Watkinson, A. Sioson, C. Vasquez-Robinet, M. Shukla, D. Kumar, M. Ellis, L. Heath, N. Ramakrishnan, B. Chevone, L. Watson, L. Van Zyl, U. Egertsdotter, R. Sederoff, and R. Grene, "Photosynthesis Acclimation Is Reflected in Specific patterns of gene expression in drought-stressed loblolly pine," *Plant Physiology*, vol. 133, pp. 1702-1716, 2003.
[32] "The Alliance for Cellular Signaling (AfCS)," in *Nature*, vol. 420, 2002.
[33] N. Lavrac and S. Dzeroski, *Inductive Logic Programming: Theory and Applications*: Ellis Horwood, 1994.
[34] M. Scheideler, N. Schlaich, K. Fellenberg, T. Beissbarth, N. Hauser, M. Vingron, A. Slusarenko, and J. Hoheisel, "Monitoring the Switch from Housekeeping to Pathogen Defense Metabolism in Arabidopsis thaliana Using cDNA Arrays," *J. Biol. Chem.*, vol. 277, pp. 10555-10561, 2002.
[35] O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L. Müller, S. Rhee, and M. Stitt, "MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes," *The Plant Journal*, vol. 37, pp. 914-939, 2004.
[36] L. Heath and N. Ramakrishnan, "The Emerging Landscape of Bioinformatics Software System," *IEEE Computing*, vol. 35, pp. 41-45, 2002.
[37] "Bioinformatics Links Directory," *http://bioinformatics.ubc.ca/resources/links_directory/*, 2006.
[38] "Pathway database," *http://www.bioinf.mdc-berlin.de/~schober/AnnotationDTBs.htm*, 2006.

[39] G. Baderr, I. Donaldson, C. Wolting, B. Ouellette, T. Pawson, and C. Hogue, "BIND – the Biomolecular Interaction Network Database," in *http://bind.ca/*, 2006.
[40] "STKE, Signal Transduction Knowledge Environment," *http://stke.sciencemag.org/*, 2006.
[41] P. Karp, J. Collado-Vides, J. Ingraham, I. Paulsen, and M. Saier, "Ecocyc: Encyclopedia of Escherichia coli K12 Genes and Metabolism," in *http://www.ecocyc.org/*, 2006.
[42] B. Zhang, "Pathway Editor: A Tool for Creating and Editing Biological Pathways Data," Masters Project, EECS Dept CWRU, 2003.
[43] T. Toyoda, K. Hirosawa, and A. Konagaya, "KnowledgeEditor: a new tool for interactive modeling and analyzing biological pathways based on microarray data," *Bioinformatics*, vol. 19, pp. 433-434, 2003.
[44] M. Lee, S. Hyun, and S. Park, "UniPath: A Knowledge Representation System for Biological Pathways," *Genome Informatics*, vol. 14, pp. 681-682, 2003.
[45] D. Yao, K. Qu, J. Wang, Y. Lu, N. Noble, H. Sun, X. Zhu, N. Lin, D. Payan, and M. Li, "PathwayFinder: paving the way towards automatic pathway extraction," presented at The second conference on Asia-Pacific bioinformatics, pp. 53-62, 2004.
[46] "PubGene™," *http://www.pubgene.com/*, 2006.
[47] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky, "GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles," *Bioinformatics*, vol. 17, pp. 74-82, 2001.
[48] "Vector PathBlazer™, Informax Inc. Solutions," *http://register.informaxinc.com/solutions/pathblazer/*, 2006.
[49] "Omniviz™," *http://www.omniviz.com/applications/pathways.htm*, 2006.
[50] B. Zupan, J. Demsar, I. Bratko, P. Juvan, J. Halter, A. Kuspa, and G. Shaulsky, "GenePath: a system for automated construction of genetic networks from mutant data," *Bioinformatics*, vol. 19, pp. 383 -389, 2003.
[51] Ä. Glass and l. Gierl, "A bioinformatic approach for generating genetic networks," *Biosystems and Medical Technology*, vol. 2, 2000.
[52] K. Kim, K. Lee, S. Park, M. Shin, and H. Cho, "GENAW: Genetic Network Analysis Workbench," presented at The 11th International Conference on Intelligent Systems for Molecular Biology, poster, 2003.
[53] E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk, "PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways," *Bioinformatics*, vol. 18., pp. 996-1003, 2003.
[54] A. Kuchinsky, K. Graham, D. Moh, and M. Creech, "Biological Storytelling: A Software Tool for Biological Information Organization Based upon Narrative Structure," presented at The Advanced Visual Interfaces (AVI), pp. 4-5, 2002.
[55] T. Toyoda, Y. Mochizuki, and A. Konagaya, "GSCOPE: A Clipped Fisheye Viewer for Biomolecular Network Graphs," *Bioinformatics*, vol. 19, pp. 437-438, 2003.
[56] "Genespring™," *Silicon Genetics*, *www.silicongenetics.com*, 2006.

[57] S. Doniger, N. Salomonis, K. Dahlquist, K. Vranizan, S. Lawlor, and B. Conklin, "MAPPFinder: using Gene Ontology and GenMAPP to create a global gene expression profile from microarray data," *Genome Biology*, vol. 4, 2003.

[58] K. Dahlquist, N. Salomonis, K. Vranizan, S. Lawlor, and B. Conklin, "GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways," *Nature Genetics*, vol. 31, pp. 19-20, 2002.

[59] P. Shannon, A. Markiel, O. Ozier, S. Baliga, T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, pp. 2498-2504, 2003.

[60] "COPASI, Complex Pathway Simulator," *http://www.copasi.org/tiki-index.php*, 2006.

[61] E. Tufte, *The Visual Display of Quantitative Information*: Graphics Press, 1983.

[62] A. Markiel, "Cytoscape: A Network Modeling Environment with Applications to Biomolecular Interaction Networks," presented at The IEEE Symposium on Information Visualization, Interactive Demos, 2003.

[63] I. Herman, G. Melancon, and M. Marshall, "Graph visualization and navigation in information visualization: A survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, pp. 24-43, 2002.

[64] B. Bederson and J. Hollan, "Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics," in Proceedings of The User Interface Software and Technology (UIST), pp. 17-26, 1994.

[65] S. Feiner, "Seeing the forest for the trees: Hierarchical display of hypertext structure," presented at The Proceedings of the ACM Conference on Office Information Systems, pp. 205-212, 1988.

[66] T. J. Jankun-Kelly and K. Ma, "MoireGraphs: Radial focus+context visualization and interaction for graphs with visual nodes," presented at The IEEE Symposium on Information Visualization, pp. 59-66, 2003.

[67] R. A. Becker, Eick, S. G., and Wilks, A. R., "Visualizing network data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, pp. 16-28, 1995.

[68] S. North, "GraphViz," *http://www.graphviz.org/*, 2006.

[69] T. Munzner, E. Hoffman, K. Claffy, and B. Fenner, "Visualizing the Global Topology of the MBone," presented at The IEEE Symposium on Information Visualization, pp.85-92, 1996.

[70] S. Bridgeman and R. Tamassia, "A User Study in Similarity Measures for Graph Drawing," *Journal of Graph Algorithms and Applications*, vol. 6, pp. 225-254, 2002.

[71] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for Using Multiple Views in Information Visualization," presented at The working conference on Advanced visual interfaces (AVI), pp. 110-119, 2000.

[72] J. C. Roberts, "On encouraging multiple views for visualization," presented at The IEEE Symposium on Information Visualization, pp. 8-14, 1998.

[73] B. Yost and C. North, "Single Complex Glyphs Versus Multiple Simple Glyphs," Proceedings of The ACM Conference on Human Factors in Computing Systems (CHI '05), pp. 1889-1892, 2005.

[74] C. North, "Toward Measuring Visualization Insight," *IEEE Computer Graphics & Applications, Visualization Viewpoints*, 2006.
[75] D. Bassett, M. Eisen, and M. Boguski, "Gene Expression Informatics - its all in your mine," *Nature Genetics Supplement*, vol. 21, pp. 51-55, 1999.
[76] R. Spence, *Information Visualization*: Addison-Wesley, 2001.
[77] S. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization – Using Visualization to Think*. San Francisco: Morgan Kaufmann, 1999.
[78] "SPOTFIRE® Decisionsite™ for functional Genomics," *http://www.Spotfire.com*, 2006.
[79] C. Chen and M. Czerwinski, "Empirical evaluation of information visualizations: an introduction," *Int. J. Human- Computer Studies*, vol. 53, pp. 631-635, 2000.
[80] C. Chen and Y. Yu, "Empirical studies of information visualization: a meta-analysis," *Int. J. Human-Computer Studies*, vol. 53, pp. 851-866, 2000.
[81] A. Kobsa, "An Empirical Comparison of Three Commercial Information Visualization Systems," presented at The IEEE Symposium on Information Visualization, pp. 123-130, 2001.
[82] P. Irani and C. Ware, "Diagramming information structures using 3D perceptual primitives," *ACM Transactions on Computer-Human Interaction*, vol. 10, pp. 1-19, 2003.
[83] H. Hartson and D. Hix, *Developing User Interfaces: Ensuring Usability Through Product and Process*: John Wiley, 1993.
[84] G. Rao and D. Mingay, "Report on usability testing of census bureau's dynamaps CD-ROM product," http://infovis.cs.vt.edu/cs5764/papers/dynamapsUsability.pdf, 2001.
[85] J. D. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Transactions on Graphics*, vol. 5, pp. 110 – 141, 1986.
[86] C. Freitas, P. Luzzardi, R. Cava, M. Pimenta, W. A., and L. Nedel, "Evaluating Usability of Information Visualization Techniques," presented at The Advanced Visual Interfaces (AVI), poster, 2002.
[87] B. Shneiderman, "The eyes have it: a task by data type taxonomy," presented at The IEEE Symposium on Visual Languages, pp. 336-343, 1996.
[88] C. Ware, *Information Visualization: Perception for Design*: Morgan Kaufmann, 2004.
[89] W. S. Cleveland, *The Elements of Graphing Data*. Monterey, California: Wadsworth Advanced Books and Software, 1980.
[90] O. Juarez, "CAEVA: Cognitive Architecture to Evaluate Visualization Applications," presented at The IEEE Symposium on Information Visualization, pp. 589-595, 2003.
[91] V. Gonzales and A. Kobsa, "A workplace study of the adoption of information visualization systems," presented at The I-KNOW"03: 3rd International Conference on KnowledgeManagement, pp. 92-102, 2003.
[92] J. Rieman, "A field study of exploratory learning Strategies," *ACM Transactions on Computer-Human Interaction*, vol. 3, pp. 189-218, 1996.
[93] P. Saraiya, P. Lee, and C. North, "Visualization of Graphs with Associated Time-series data," *The IEEE Symposium on Information Visualization*, pp. 30-37, 2005.

[94] J. Seo and B. Shneiderman, "Knowledge Discovery in High Dimensional Data: Case Studies and a User Survey for the Rank-by-Feature Framework," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, 2006.

[95] H. Causton, J. Quackenbush, and A. Brazma, *A beginners guide to microarray data analysis*: Blackwell Publishing, 2003.

[96] D. Berrar, W. Dubitzky, and M. Granzow, *Practical approach to microarray data analysis*. Kluwer academic publishers, 2004.

[97] "Cambridge Healthtech Institute's Data visualization and Interpretation, Making Breakthroughs possible in the Omics Research," *http://www.healthtech.com/2002/dvs/*, 2006.

[98] "Cambridge Healthtech Institute's Data visualization and Interpretation, Deciphering the Data Deluge," *http://www.healthtech.com/2003/mde/index.asp*, 2006.

[99] M. Eisen, P. Shellman, P. Brown, and D. Bostein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 95, pp. 14963-14968, 1998.

[100] "GeneSifter," *http://www.genesifter.net/web/*, 2006.

[101] N. E. Olson, "Identification of Cell Cycle Genes Regulated During Erythroid Differentiation," http://www.microarraysuccess.org/web/, 2006.

[102] V. L. O'Day, A. Adler, A. Kuchinsky, and A. Bouch, "When worlds collide: Molecular biology as interdisciplinary collaboration," *ECSCW*, 2001.

[103] W. E. Mackay, G. Pothier, C. Letondal, K. Boegh, and H. E. Sorensen, "The missing link: augmenting biology laboratory notebooks," in Proceedings of The User Interface Software and Technology (UIST), pp. 41-50, 2002.

[104] J. Thomas and K. Cook, *Illuminating the path: The research and development agenda for visual analytics*: IEEE Press, 2005.

[105] K. A. Duca, H. Goto, Y. Kawaoka, and J. Yin, "Time-Resolved mRNA Profiling During Influenza Infection: Extracting Information from a Challenging Experimental System," presented at The American Society for Virology, 20th Annual Meeting, Madison, WI, 2001.

[106] G. Geiss, M. Salvatore, T. Tumpey, V. Carter, X. Wang, C. Basler, J. Taubenberger, R. Bumbarner, P. Palese, M. Katze, and A. Garcia-Sastre, "Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: The role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 99, pp. 10736–10741, 2002.

[107] E. Baechler, F. Batliwala, G. Karvpis, P. Gaffney, W. Ortmann, K. Espe, K. Shark, W. Grande, K. Hughes, K. Kapur, P. Gregersen, and T. Behrens, "Interferon-inducible gene expression signature in peripheral blood cells of patients with severe SLE," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 100, pp. 2610-2615, 2003.

[108] "Affymetrix™," in *http://www.affymetrix.com/index.affx*.

[109] H. Hochheriser, E. H. Baehrecke, S. M. Mount, and B. Shneiderman, "Dynamic Querying for Pattern Identification in Microarray and Genomic Data," presented at The IEEE International Conference on Multimedia and Expo, 2003.

[110] J. Seo and B. Shneiderman, "Interactively Exploring Hierarchical Clustering Results," *IEEE Computer*, vol. 35, pp. 80-86, 2002.
[111] J. Seo and B. Shneiderman, "A Rank-by-Feature Framework for Unsupervised Multidimensional Data Exploration Using Low Dimensional Projections," presented at The IEEE Symposium on Information Visualization, pp. 65-72, 2004.
[112] J. C. Flanagan, "The Critical Incident Technique," *Psychological Bulletin*, vol. 51, pp. 327-358, 1954.
[113] C. Plaisant, "The Challenge of Information Visualization Evaluation," presented at The Advanced Visual Interfaces (AVI), pp. 109-116, 2004.
[114] G. Golovchinsky and N. J. Belkin, "Innovation and Evaluation of Information Exploration Interfaces: A CHI98 Workshop," The SIGCHI Bulletin, pp. 22-25, 1999.
[115] R. Kosara, C. G. Healey, V. Interrante, D. Laidlaw, and C. Ware, "Thoughts on User Studies: Why, How, and When," *IEEE Computer Graphics & Applications, Visualization Viewpoints*, vol. 23, pp. 20-25, 2003.
[116] D. House, V. Interrante, D. Laidlaw, R. Taylor, and C. Ware, "Design and Evaluation in Visualization Research," presented at The IEEE Symposium on Information Visualization, Panel discussion, 2005.
[117] M. Tory and T. Möller, "Evaluating Visualizations: Do Expert Reviews Work?," *IEEE Computer Graphics and Applications*, vol. 25, pp. 8-11, 2005.
[118] M. P. Steves, E. Morse, C. Gutwin, and S. Greenberg, "A Comparison of Usage Evaluation and Inspection Methods for Assessing Groupware Usability," presented at The International ACM SIGGROUP Conference on Supporting Group Work, pp. 125-134, 2001.
[119] A. J. Brush, M. Ames, and J. Davis, "A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface," presented at The ACM Conference on Human Factors in Computing Systems, pp. 1179-1183, 2004.
[120] E. Baauw and M. Bekker, "A comparison of two analytical evaluation methods for children's computer games," presented at The Interact 2005 Workshop - WS7, 2005.
[121] R. Jeffries and H. Desurvire, "Usability Testing vs. Heuristic Evaluation: Was there a Contest," The SIGCHI Bulletin, pp. 39-41, 1992.
[122] H. R. Hartson and T. S. Andre, "Criteria for evaluating usability methods," *International Journal of Human-Computer Interaction*, vol. 13, pp. 373-410, 2001.
[123] B. John and S. Marks, "Tracking the effectiveness of usability evaluation methods," *Behaviour and Information Technology*, vol. 16, pp. 188-202, 1997.
[124] A. Doubleday, M. Ryan, M. V. Springett, and A. G. Sutcliffe, "A Comparison of Usability Techniques for Evaluating Design," presented at The Symposium on Designing Interactive Systems, pp. 101-110, 1997.
[125] P. Saraiya, C. North, V. Lam, and K. Duca, "An Insight-based Longitudinal Study of Visual Analytics," *The IEEE Transactions on Visualization and Computer Graphics*, 2006.
[126] R. Amar, J. Eagan, and J. Stasko, "Low-Level Components of Analytic Activity in information Visualization," presented at The IEEE Symposium on Information Visualization, pp.111-117, 2005.

[127] "Q Value Software," *http://faculty.washington.edu/jstorey/qvalue/*, 2006.
[128] "KalaidaGraph," *http://www.synergy.com/*, 2006.
[129] "R project for statistical computing," *http://www.r-project.org/*, 2006.
[130] J. Storey and R. Tibshirani, "Statistical significance for genomewide studies," *Proceedings of the National Academy of Science (PNAS)*, vol. 100, pp. 9440-9445, 2003.

Appendix A: Data Analysis for Chapter 3

A.1 Overall Performance

A.1.1 ANOVA: Task Completion Time versus Views, Node Attributes, Task, *Participants*

```
Factor                                Type    Levels
Participants(Views Node Attributes)  random      10
Views                                fixed        2
Node Attributes                      fixed        2
Task                                 fixed       11

Factor                                Values
Participants(Views Node Attributes)  p1, p10, p2, p3, p4, p5, p6, p7, p8, p9
Views                                Multiple View, Single View
Node Attributes                      Multiple Attribute, Single Attribute
Task                                T1, T10, T11, T2, T3, T4, T5, T6, T7, T8, T9

Analysis of Variance for Performance

Source                                DF        SS       MS     F      P
Participants(Views Node Attributes)   36  116505.8   3236.3  3.35  0.000
Views                                  1   22982.7  22982.7  7.10  0.011
Node Attributes                        1    1776.0   1776.0  0.55  0.464
Views*Node Attributes                  1       0.4      0.4  0.00  0.991
Task                                  10   81325.9   8132.6  8.42  0.000
Node Attributes*Task                  10   31085.0   3108.5  3.22  0.001
Views*Task                            10   36188.9   3618.9  3.75  0.000
Views*Node Attributes*Task            10   10935.3   1093.5  1.13  0.337
Error                                360  347695.6    965.8
Total                                439  648495.6

S = 31.0777   R-Sq = 46.38%   R-Sq(adj) = 34.62%
```

Figure A.1 ANOVA analysis for task completion time between views, node attributes on 11 tasks.

A.1.2 ANOVA: Accuracy versus Views, Node Attributes, Task, Participants

```
Factor                                Type    Levels
Participants(Views Node Attributes)  random      10
Views                                fixed        2
Node Attributes                      fixed        2
Task                                 fixed       11

Factor                                Values
Participants(Views Node Attributes)  p1, p10, p2, p3, p4, p5, p6, p7, p8, p9
Views                                Multiple View, Single View
Node Attributes                      Multiple Attribute, Single Attribute
Task                                  T1, T10, T11, T2, T3, T4, T5, T6, T7, T8,
T9

Analysis of Variance for Accuracy

Source                                DF        SS      MS      F      P
Participants(Views Node Attributes)   36   12.2000  0.3389   1.94  0.001
Views                                  1    0.0364  0.0364   0.11  0.745
Node Attributes                        1    4.4000  4.4000  12.98  0.001
Views*Node Attributes                  1    0.0818  0.0818   0.24  0.626
Task                                  10    9.9682 1.9968   11.45  0.000
```

```
Node Attributes*Task                  10    2.1500  0.2150   1.23  0.269
Views*Task                            10    1.6136  0.1614   0.93  0.510
Views*Node Attributes*Task            10    3.4682  0.3468   1.99  0.034
Error                                360   62.8000  0.1744
Total                                439  106.7182

S = 0.417665   R-Sq = 41.15%   R-Sq(adj) = 28.24%
```

Figure A.2 ANOVA analysis for accuracy between views, node attributes on 11 tasks.

A.2 1-way ANOVA analysis between visualization options

A.2.1 Anova analysis on accuracy

Single Attribute + Single View vs. Single Attribute + Multiple views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SS	110	75	0.681818	0.218932		
SM	110	76	0.690909	0.215513		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.004545	1	0.004545	0.020925	0.885116	3.884469
Within Groups	47.35455	218	0.217223			
Total	47.35909	219				

Table A.1 Results from 1 – way anova between single attribute visualization on accuracy.

Multiple Attributes + Single View vs. Multiple Attributes + Multiple views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
MS	110	56	0.509091	0.25221		
MM	110	51	0.463636	0.250959		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.113636	1	0.113636	0.451682	0.502249	3.884469
Within Groups	54.84545	218	0.251585			
Total	54.95909	219				

Table A.2 Results from 1 – way anova between multiple attribute visualization on accuracy.

Single Attribute + Single View vs. Multiple Attribute + Single view

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SS	110	75	0.681818	0.218932		
MS	110	56	0.509091	0.25221		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.640909	1	1.640909	6.965658	0.008909	3.884469
Within Groups	51.35455	218	0.235571			
Total	52.99545	219				

Table A.3 Results from 1 – way anova between single view visualization on accuracy.

Single Attribute + Multiple Views vs. Multiple Attribute + Multiple Views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SM	110	76	0.690909	0.215513		
MM	110	51	0.463636	0.250959		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.840909	1	2.840909	12.1804	0.000584	3.884469
Within Groups	50.84545	218	0.233236			
Total	53.68636	219				

Table A.4 Results from 1 – way anova between multiple view visualization on accuracy.

A.2.2 ANOVA analysis on performance times

Single Attribute + Single View vs. Single Attribute + Multiple Views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SS	110	5670	51.54545	758.452		
SM	110	7267	66.06364	2239.877		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11592.77	1	11592.77	7.73282	0.005897	3.884469
Within Groups	326817.8	218	1499.164			
Total	338410.6	219				

Table A.5 Results from 1 – way anova between single attribute visualization on performance time.

Multiple Attributes + Single View vs. Multiple Attributes + Multiple Views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
MS	110	5235	47.59091	871.932		
MM	110	6818	61.98182	1852.091		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11390.4	1	11390.4	8.362927	0.004217	3.884469
Within Groups	296918.6	218	1362.012			
Total	308309	219				

Table A.6 Results from 1 – way anova between multiple attribute visualization on performance time.

Single Attribute + Single View vs. Multiple Attribute + Single Views

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SS	110	5670	51.54545	758.452		
MS	110	5235	47.59091	871.932		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	860.1136	1	860.1136	1.055106	0.305473	3.884469
Within Groups	177711.9	218	815.192			
Total	178572	219				

Table A.7 Results from 1 – way anova between single view visualization on performance time.

Single Attribute + Multiple Views vs. Multiple Attribute + Multiple Views

Anova: Single Factor						

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
SM	110	7267	66.06364	2239.877		
MM	110	6818	61.98182	1852.091		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	916.3682	1	916.3682	0.447886	0.504049	3.884469
Within Groups	446024.5	218	2045.984			
Total	446940.9	219				

Table A.8 Results from 1 – way anova between multiple view visualization on performance time.

A.3 Analysis for tasks involving 1 Timepoint

A.3.1 ANOVA analysis on Accuracy

Overall Analysis for task1 – task3

Anova: Two-Factor With Replication						
SUMMARY	Single Views	Multiple View	Total			
Single Attribute						
Count	30	30	60			
Sum	22	18	40			
Average	0.733333	0.6	0.666667			
Variance	0.202299	0.248276	0.225989			
Multiple Attributes						
Count	30	30	60			
Sum	14	16	30			
Average	0.466667	0.533333	0.5			
Variance	0.257471	0.257471	0.254237			
Total						
Count	60	60				
Sum	36	34				
Average	0.6	0.566667				
Variance	0.244068	0.249718				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Sample	0.833333	1	0.833333	3.452381	0.065696	3.922879
Columns	0.033333	1	0.033333	0.138095	0.71086	3.922879
Interaction	0.3	1	0.3	1.242857	0.267226	3.922879
Within	28	116	0.241379			
Total	29.16667	119				

Table A.9 Anova analysis for tasks 1- 3 for accuracy.

Analysis for task 1

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	10	8	18			
Average	1	0.8	0.9			
Variance	0	0.177778	0.094737			
Multiple Attribute						
Count	10	10	20			
Sum	7	8	15			
Average	0.7	0.8	0.75			
Variance	0.233333	0.177778	0.197368			
Total						
Count	20	20				
Sum	17	16				
Average	0.85	0.8				
Variance	0.134211	0.168421				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.225	1	0.225	1.528302	0.224373	4.113165
Columns	0.025	1	0.025	0.169811	0.682723	4.113165
Interaction	0.225	1	0.225	1.528302	0.224373	4.113165
Within	5.3	36	0.147222			
Total	5.775	39				

Table A.10 Anova analysis for task 1 for accuracy.

Analysis for task 2

Anova: Two-Factor With Replication						
SUMMARY	Single	Multiple	Total			

	View	View				
Single Attribute						
Count	10	10	20			
Sum	7	6	13			
Average	0.7	0.6	0.65			
Variance	0.233333	0.266667	0.239474			
Multiple Attribute						
Count	10	10	20			
Sum	4	0	4			
Average	0.4	0	0.2			
Variance	0.266667	0	0.168421			
Total						
Count	20	20				
Sum	11	6				
Average	0.55	0.3				
Variance	0.260526	0.221053				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2.025	1	2.025	10.56522	0.002502	4.113165
Columns	0.625	1	0.625	3.26087	0.079319	4.113165
Interaction	0.225	1	0.225	1.173913	0.285802	4.113165
Within	6.9	36	0.191667			
Total	9.775	39				

Table A.11 Anova analysis for task 2 for accuracy.

Analysis for task 3

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	5	4	9			
Average	0.5	0.4	0.45			
Variance	0.277778	0.266667	0.260526			
Multiple Attribute						
Count	10	10	20			
Sum	3	8	11			
Average	0.3	0.8	0.55			
Variance	0.233333	0.177778	0.260526			

Total						
Count	20	20				
Sum	8	12				
Average	0.4	0.6				
Variance	0.252632	0.252632				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.1	1	0.1	0.418605	0.521738	4.113165
Columns	0.4	1	0.4	1.674419	0.203908	4.113165
Interaction	0.9	1	0.9	3.767442	0.06012	4.113165
Within	8.6	36	0.238889			
Total	10	39				

Table A.12 Anova analysis for task 3 for accuracy.

A.3.2 Analysis for Performance Time

Overall Analysis task1 – task3

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	30	30	60			
Sum	1342	2421	3763			
Average	44.73333	80.7	62.71667			
Variance	626.754	486.2862	551.2234			
Multiple Attribute						
Count	30	30	60			
Sum	1275	2041	3315			
Average	42.53333	68.86667	55.6			
Variance	1149.085	1290.395	1226.214			
Total						
Count	60	60				
Sum	2617	4062				
Average	43.63333	80.78333				
Variance	908.2701	1047.359				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Sample	10811.01	1	10811.01	12.17277	0.006862	3.922879
Columns	297.675	1	297.675	0.33517	0.563753	3.922879
Interaction	1548.008	1	1548.008	1.742997	0.189359	3.922879
Within	103023.1	116	888.1302			
Total	115679.8	119				

Table A.13 Anova analysis for Tasks 1-3 for performance time.

Analysis for Task 1

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	431	623	1054			
Average	43.1	62.3	57.7			
Variance	367.6556	358.6778	356.2211			
Multiple Attribute						
Count	10	10	20			
Sum	428	596	1024			
Average	42.8	59.6	51.2			
Variance	587.2889	1070.489	815.9579			
Total						
Count	20	20				
Sum	919	959				
Average	42.95	61.95				
Variance	460.8921	819.8395				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2102.5	1	2102.5	3.52752	0.068476	4.113165
Columns	40	1	40	0.067111	0.797066	4.113165
Interaction	774.4	1	774.4	1.299268	0.261874	4.113165
Within	21457	36	596.0278			
Total	24373.9	39				

Table A.14 Anova analysis for task 1 for performance time.

Analysis for task 2

Anova: Two-Factor With Replication						
SUMMARY	Single	Multiple	Total			

	View	View				
Single Attribute						
Count	10	10	20			
Sum	530	887	1417			
Average	53	88.7	71.35			
Variance	814.2222	879.1222	802.2395			
Multiple Attribute						
Count	10	10	20			
Sum	537	770	1307			
Average	53.7	77	65.35			
Variance	967.3444	1631.778	1255.924			
Total						
Count	20	20				
Sum	1067	1657				
Average	53.35	82.35				
Variance	897.7132	1332.239				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	3534.4	1	3534.4	3.293584	0.07789	4.113165
Columns	270.4	1	270.4	0.251976	0.618743	4.113165
Interaction	202.5	1	202.5	0.188703	0.666591	4.113165
Within	38632.2	36	1073.117			
Total	42639.5	39				

Table A.15 Anova analysis for task 2 for performance time.

Analysis for task 3

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	381	831	1212			
Average	38.1	83.1	60.5			
Variance	709.8778	37.65556	363.5684			
Multiple Attribute						
Count	10	10	20			
Sum	325	640	965			
Average	32.5	64	48.25			
Variance	1942.5	1273.778	1552.513			

Total						
Count	20	20				
Sum	916	961				
Average	30.3	73.55				
Variance	1318.8	888.9974				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	5593.225	1	5593.225	5.64429	0.022956	4.113165
Columns	50.625	1	50.625	0.051087	0.822461	4.113165
Interaction	680.625	1	680.625	0.686839	0.412703	4.113165
Within	35674.3	36	990.9528			
Total	41998.78	39				

Table A.16 Anova analysis for task 3 for performance time.

A.4 Analysis for Tasks involving 2 Timepoints

A.4.1 ANOVA analysis for Accuracy

Overall accuracy for task4-task5

Anova: Two-Factor With Replication						
SUMMARY	Single View	MultipleView	Total			
Single A						
Count	20	20	40			
Sum	17	18	35			
Average	0.85	0.9	0.875			
Variance	0.197368	0.094737	0.148077			
Multiple A						
Count	20	20	40			
Sum	12	10	22			
Average	0.6	0.5	0.55			
Variance	0.221053	0.263158	0.246154			
Total						
Count	40	40				
Sum	29	28				
Average	0.725	0.7				
Variance	0.204487	0.215385				
ANOVA						

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	1.0125	1	1.0125	5.216949	0.025161	3.96676
Columns	0.0125	1	0.0125	0.064407	0.800347	3.96676
Interaction	0.6125	1	0.6125	3.155932	0.079653	3.96676
Within	14.75	76	0.194079			
Total	16.3875	79				

Table A.17 Anova analysis for tasks 4-5 for accuracy.

Analysis for task 4

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	9	8	17			
Average	0.9	0.8	0.85			
Variance	0.1	0.177778	0.134211			
Multiple Attribute						
Count	10	10	20			
Sum	5	7	12			
Average	0.5	0.7	0.6			
Variance	0.277778	0.233333	0.252632			
Total						
Count	20	20				
Sum	14	15				
Average	0.7	0.75				
Variance	0.221053	0.197368				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.625	1	0.625	3.169014	0.08349	4.113165
Columns	0.025	1	0.025	0.126761	0.723893	4.113165
Interaction	0.225	1	0.225	1.140845	0.292584	4.113165
Within	7.1	36	0.197222			
Total	7.975	39				

Table A.18 Anova analysis for task4 for accuracy.

Analysis for task 5

Anova: Two-Factor With Replication			

SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	2	3	5			
Average	0.2	0.3	0.25			
Variance	0.177778	0.233333	0.197368			
Multiple Attribute						
Count	10	10	20			
Sum	2	2	4			
Average	0.2	0.2	0.2			
Variance	0.177778	0.177778	0.168421			
Total						
Count	20	20				
Sum	4	5				
Average	0.2	0.25				
Variance	0.168421	0.197368				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.025	1	0.025	0.130435	0.720093	4.113165
Columns	0.025	1	0.025	0.130435	0.720093	4.113165
Interaction	0.025	1	0.025	0.130435	0.720093	4.113165
Within	6.9	36	0.191667			
Total	6.975	39				

Table A.19 Anova analysis for task5 for accuracy.

A.4.2 Analysis for Performance Time

Overall Analysis task4-task5

Anova: Two-Factor With Replication						
SUMMARY	Single V	Multiple V	Total			
Single A						
Count	20	20	40			
Sum	1292	1380	2672			
Average	56.6	64	60.3			
Variance	673.4105	1431.368	1030.369			
Multiple A						
Count	20	20	40			

Sum	991	1237	2228			
Average	45.11	50.12	47.5			
Variance	884.6816	295.3974	613.7026			
Total						
Count	40	40				
Sum	2283	2617				
Average	57.075	65.425				
Variance	817.1481	854.3532				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2464.2	1	2464.2	4.000678	0.0087286	3.96676
Columns	1394.45	1	1394.45	1.698034	0.19648	3.96676
Interaction	312.05	1	312.05	0.379986	0.539453	3.96676
Within	62412.3	76	821.2145			
Total	66583	79				

Table A. 20 Anova analysis for tasks 4-5 for performance time.

Analysis for task 4

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	619	567	1186			
Average	61.9	56.7	59.3			
Variance	930.7667	1312.678	1069.8			
Multiple Attribute						
Count	10	10	20			
Sum	450	558	1008			
Average	45	55.8	50.4			
Variance	629.7778	391.2889	514.3579			
Total						
Count	20	20				
Sum	1069	1125				
Average	53.45	56.25				
Variance	814.3658	807.3553				
ANOVA						

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	792.1	1	792.1	0.970559	0.331116	4.113165
Columns	78.4	1	78.4	0.096063	0.758393	4.113165
Interaction	640	1	640	0.784191	0.38174	4.113165
Within	29380.6	36	816.1278			
Total	30891.1	39				

Table A.21 Anova analysis for task 4 for performance time.

Analysis for task 5

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	673	813	1486			
Average	67.3	81.3	74.3			
Variance	474.6778	1372.9	926.7474			
Multiple Attribute						
Count	10	10	20			
Sum	541	679	1220			
Average	54.1	67.9	61			
Variance	1191.878	150.9889	686.2105			
Total						
Count	20	20				
Sum	1214	1492				
Average	60.7	74.6				
Variance	835.2737	769.0947				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	1768.9	1	1768.9	2.217747	0.145141	4.113165
Columns	1932.1	1	1932.1	2.422358	0.128363	4.113165
Interaction	0.1	1	0.1	0.000125	0.991128	4.113165
Within	28714	36	797.6111			
Total	32415.1	39				

Table A.22 Anova analysis for tasks 5 for performance time.

A.5 Analysis for Tasks involving all 10 Timepoints

A.5.1 ANOVA analysis for Accuracy

Overall Performance for tasks 6 – task 10

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single A						
Count	60	60	120			
Sum	38	40	78			
Average	0.633333	0.666667	0.65			
Variance	0.236158	0.225989	0.229412			
Multiple A						
Count	60	60	120			
Sum	28	25	53			
Average	0.466667	0.416667	0.441667			
Variance	0.253107	0.247175	0.248669			
Total						
Count	120	120				
Sum	66	65				
Average	0.55	0.541667				
Variance	0.24958	0.25035				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2.604167	1	2.604167	10.8233	0.001156	3.881163
Columns	0.004167	1	0.004167	0.017317	0.895417	3.881163
Interaction	0.104167	1	0.104167	0.432932	0.511195	3.881163
Within	56.78333	236	0.240607			
Total	59.49583	239				

Table A.23 Anova analysis for tasks 6-10 for accuracy.

Analysis for task6

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	80	100	18			
Average	0.8	1	0.9			

Variance	0	0.177778	0.094737			
Multiple Attribute						
Count	10	10	20			
Sum	8	7	15			
Average	0.8	0.7	0.75			
Variance	0.233333	0.177778	0.197368			
Total						
Count	20	20				
Sum	16	17				
Average	0.8	0.85				
Variance	0.134211	0.168421				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.225	1	0.225	1.528302	0.224373	4.113165
Columns	0.025	1	0.025	0.169811	0.682723	4.113165
Interaction	0.225	1	0.225	1.528302	0.224373	4.113165
Within	5.3	36	0.147222			
Total	5.775	39				

Table A.24 Anova analysis for task 6 for accuracy.

Analysis for task7

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	9	8	17			
Average	0.9	0.8	0.85			
Variance	0.266667	0.233333	0.239474			
Multiple Attribute						
Count	10	10	20			
Sum	5	6	11			
Average	0.5	0.6	0.05			
Variance	0.1	0	0.05			
Total						
Count	20	20				
Sum	14	14				
Average	0.7	0.7				

Variance	0.197368	0.134211				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.9	1	0.9	4.891	0.031303	4.113165
Columns	0.1	1	0.1	0.666667	0.419587	4.113165
Interaction	-4.4E-15	1	-4.4E-15	-3E-14	#NUM!	4.113165
Within	5.4	36	0.15			
Total	6.4	39				

Table A. 25 Anova analysis for task 7 for accuracy.

Analysis for task8

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	4	4	8			
Average	0.4	0.4	0.8			
Variance	0.233333	0.266667	0.239474			
Multiple Attribute						
Count	10	10	20			
Sum	7	7	14			
Average	0.7	0.7	0.7			
Variance	0.266667	0	0.168421			
Total						
Count	20	20				
Sum	11	11				
Average	0.55	0.55				
Variance	0.260526	0.221053				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2.025	1	2.025	5.56522	0.039502	4.113165
Columns	0.625	1	0.625	3.26087	0.079319	4.113165
Interaction	0.225	1	0.225	1.173913	0.285802	4.113165
Within	6.9	36	0.191667			
Total	9.775	39				

Table A.26 Anova analysis for task 8 for accuracy.

Analysis for task9

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	7	9	16			
Average	0.7	0.9	0.8			
Variance	0.233333	0.1	0.168421			
Multiple Attribute						
Count	10	10	20			
Sum	5	5	10			
Average	0.5	0.5	0.5			
Variance	0.277778	0.277778	0.263158			
Total						
Count	20	20				
Sum	12	14				
Average	0.6	0.7				
Variance	0.252632	0.221053				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.9	1	0.9	4.35	0.041699	4.113165
Columns	0.1	1	0.1	0.45	0.506616	4.113165
Interaction	0.1	1	0.1	0.45	0.506616	4.113165
Within	8	36	0.222222			
Total	9.1	39				

Table A.27 Anova analysis for task 9 for accuracy.

Analysis for task10

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	8	8	16			
Average	0.8	0.8	0.8			
Variance	0.233333	0.1	0.168421			
Multiple Attribute						

Count	10	10	20			
Sum	4	5	9			
Average	0.4	0.5	0.45			
Variance	0.177778	0.266667	0.252632			
Total						
Count	20	20				
Sum	12	13				
Average	0.6	0.65				
Variance	0.197368	0.239474				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.4	1	0.4	2.057143	0.160124	4.113165
Columns	0.1	1	0.1	0.514286	0.477917	4.113165
Interaction	0.9	1	0.9	4.628571	0.038228	4.113165
Within	7	36	0.194444			
Total	8.4	39				

Table A.28 Anova analysis for task 10 for accuracy.

Analysis for task11

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	3	4	7			
Average	0.3	0.4	0.35			
Variance	0.277778	0.266667	0.260526			
Multiple Attribute						
Count	10	10	20			
Sum	6	6	12			
Average	0.6	0.6	0.6			
Variance	0.233333	0.177778	0.260526			
Total						
Count	20	20				
Sum	9	10				
Average	0.45	0.5				
Variance	0.252632	0.252632				

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.1	1	0.1	4.018605	0.071738	4.113165
Columns	0.4	1	0.4	1.674419	0.203908	4.113165
Interaction	0.9	1	0.9	3.767442	0.660122	4.113165
Within	8.6	36	0.238889			
Total	10	39				

Table A.29 Anova analysis for task 11 for accuracy.

A.5.2 Analysis for Performance Time

Overall Performance for tasks 6 – task 10

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single A						
Count	60	60	120			
Sum	3036	4666	7702			
Average	50.6	77.76667	64.18333			
Variance	794.0068	2968.792	2051.647			
Multiple A						
Count	60	60	120			
Sum	2548	3575	6123			
Average	42.46667	59.58333	51.025			
Variance	692.4904	2674.281	1743.1			
Total						
Count	120	120				
Sum	5584	8241				
Average	46.53333	68.675				
Variance	753.6796	2881.179				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	10388.5	1	10388.5	3.228404	0.096532	3.881163
Columns	29415.2	1	29415.2	3.503210	0.072342	3.881163
Interaction	1515.038	1	1515.038	0.850002	0.357493	3.881163
Within	420644.7	236	1782.393			
Total	461963.4	239				

Table A.30 Anova analysis for tasks 6 – 10 for performance time.

Analysis for task 6

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	430	360	790			
Average	43	36	39.5			
Variance	258.2667	1177.789	784.45			
Multiple Attribute						
Count	10	10	20			
Sum	591	482	1073			
Average	59.1	48.2	53.65			
Variance	63.06667	8217.333	4248.358			
Total						
Count	20	20				
Sum	1021	1173				
Average	51.1	42.1				
Variance	181.7789	4456.134				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	87.025	1	87.025	0.035826	0.850939	4.113165
Columns	7590.025	1	7590.025	3.124606	0.085594	4.113165
Interaction	585.225	1	585.225	0.240921	0.62652	4.113165
Within	87448.1	36	2429.114			
Total	95710.38	39				

Table A.31 Anova analysis for task 6 for performance time.

Analysis for task 7

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	473	657	1130			
Average	47.3	65.7	56.5			
Variance	126.0111	853.7889	553.2105			
Multiple Attribute						

Count	10	10	20			
Sum	385	465	850			
Average	38.5	46.5	42.5			
Variance	585.8333	1019.833	777.4211			
Total						
Count	20	20				
Sum	858	1122				
Average	42.9	56.1				
Variance	357.5684	984.5158				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	1960	1	1960	3.032335	0.090161	4.113165
Columns	1742.4	1	1742.4	2.695684	0.109329	4.113165
Interaction	270.4	1	270.4	0.418338	0.52187	4.113165
Within	23269.2	36	646.3667			
Total	27242	39				

Table A.32 Anova analysis for task 7 for performance time.

Analysis for task8

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	666	810	1476			
Average	66.6	81	53.8			
Variance	737.1556	1264	1191.116			
Multiple Attribute						
Count	10	10	20			
Sum	540	671	1211			
Average	54	67.1	60.55			
Variance	252.3222	219.2111	232.8316			
Total						
Count	20	20				
Sum	1206	1481				
Average	60.3	74.05				
Variance	477.25	940.9974				

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	3204.1	1	3204.1	4.983183	0.035853	4.113165
Columns	3312.4	1	3312.4	5.358377	0.026438	4.113165
Interaction	1488.4	1	1488.4	2.407743	0.129484	4.113165
Within	22254.2	36	618.1722			
Total	30259.1	39				

Table A.33 Anova analysis for task 8 for performance time.

Analysis for task9

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	381	543	924			
Average	70	90	46.2			
Variance	176.5444	1509.344	867.6421			
Multiple Attribute						
Count	10	10	20			
Sum	458	612	1070			
Average	60	40	53.5			
Variance	442.6222	706.1778	606.5789			
Total						
Count	20	20				
Sum	839	1155				
Average	41.95	57.75				
Variance	308.8921	1061.987				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	532.9	1	532.9	0.75197	0.391597	4.113165
Columns	2496.4	1	2496.4	3.522644	0.068659	4.113165
Interaction	1.6	1	1.6	0.002258	0.962365	4.113165
Within	25512.2	36	708.6722			
Total	28543.1	39				

Table A.34 Anova analysis for task 9 for performance time.

Analysis for task10

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	380	320	700			
Average	38	32	35			
Variance	1884.622	504.7111	1162.484			
Multiple Attribute						
Count	10	10	20			
Sum	640	530	1170			
Average	64	53	58.5			
Variance	1330.322	320.9333	782.7658			
Total						
Count	20	20				
Sum	1321	850				
Average	51.1	42.5				
Variance	1797.418	530.3053				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	7645.225	1	7645.225	3.568426	0.192393	4.113165
Columns	378.225	1	378.225	0.374426	0.544447	4.113165
Interaction	216.225	1	216.225	0.214053	0.646394	4.113165
Within	36365.3	36	1010.147			
Total	44604.98	39				

Table A.35 Anova analysis for task 10 for performance time.

Analysis for task11

Anova: Two-Factor With Replication						
SUMMARY	Single View	Multiple View	Total			
Single Attribute						
Count	10	10	20			
Sum	470	650	1120			
Average	47	65	56			
Variance	270.6667	7132.544	5586.576			
Multiple Attribute						
Count	10	10	20			

Sum	379	461	840			
Average	37.9	46.1	42			
Variance	1006.544	4422.944	2984.537			
Total						
Count	20	20				
Sum	849	1111				
Average	42.45	55.55				
Variance	605.2079	6090.116				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	5640.625	1	5640.625	1.958204	0.113201	4.113165
Columns	41280.63	1	41280.63	1.267321	0.398346	4.113165
Interaction	6076.225	1	6076.225	1.893982	0.177254	4.113165
Within	115494.3	36	3208.175			
Total	168491.8	39				

Table A. 36 Anova analysis for task 11 for performance time.

Appendix B: Data Analysis for Chapter 4

B.1 Data for Insight Characteristics

B.1.1 Clusterview

Tool	Data Type	Domain Values	Count of Insights	Final Amt.	Time To First Insight (mins)	Total Time (mins)
C/T	Timeseries	5	3	30	5	30
C/T	Timeseries	6	3	35	5	25
C/T	Virus	5	2	30	5	25
C/T	Virus	9	3	50	10	30
C/T	Lupus	10	3	50	2	23
C/T	Lupus	13	4	50	3	35

Table B.1 Insight data for Clusterview for different insight characteristics.

B.1.2 Timesearcher

Vis. Tool	Data Type	Domain Values	Count of Insights	Final Amt	Time To First Insight (min)	Total Time
TS	Timeseries	10	5	15	5	45
TS	Timeseries	16	7	90	2	45
TS	Virus	9	3	70	10	30
TS	Virus	5	2	45	10	45
TS	Lupus	7	3	30	5	35
TS	Lupus	4	1	40	10	25

Table B.2 Insight data for Timesearcher for different insight characteristics.

B.1.3 HCE

Vis. Tool	Data Type	Domain Values	Count of Insights	Final Amt	Time To First Insight	Total Time
HCE	Timeseries	2	1	25	20	40
HCE	Timeseries	4	2	30	20	45
HCE	Virus	15	5	60	5	45
HCE	Virus	11	4	55	3	35
HCE	Lupus	1	1	30	20	20
HCE	Lupus	1	1	50	15	25

Table B.3 Insight data for HCE for different insight characteristics.

B.1.4 Spotfire

Vis. Tool	Data Type	Domain Values	Count of Insights	Final Amt	Time To First Insight	Total Time
SP	Timeseries	14	8	50	10	25

SP	Timeseries	13	5	95	2	55
SP	Virus	10	3	60	10	45
SP	Virus	5	3	75	10	35
SP	Lupus	14	4	70	10	45
SP	Lupus	10	3	50	3	45

Table B.4 Insight data for Spotfire for different insight characteristics.

B.1.5 GeneSpring

Vis. Tool	Data Type	Domain Values	Count of Insights	Final Amt	Time To First Insight	Total Time
GS	Timeseries	9	5	80	10	50
GS	Timeseries	6	3	20	10	45
GS	Virus	7	3	65	20	55
GS	Virus	7	3	60	15	55
GS	Lupus	3	2	25	25	60
GS	Lupus	8	4	60	15	50

Table B.5 Data for GeneSpring for different insight characteristics.

B.2 Analysis on Insight Characteristics

B.2.1 Count of Insights

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	6	18	3	0.4		
Timesearcher	6	21	3.5	4.7		
HCE	6	14	2.333333	3.066667		
Spotfire	6	25	4.166667	4.566667		
GeneSpring	6	20	3.333333	1.066667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	10.86667	4	2.716667	0.9843	0.434039	2.75871
Within Groups	69	25	2.76			
Total	79.86667	29				

Table B.6 Analysis for count of insights between visualization tools.

B.2.2 Domain Value

Overall Analysis

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	6	48	8	10.4		
Timesearcher	6	51	8.5	18.7		
HCE	6	34	5.666667	35.06667		
Spotfire	6	66	11	12		
GeneSpring	6	40	6.666667	4.266667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	98.8	4	24.7	1.535433	0.222432	2.75871
Within Groups	402.1667	25	16.08667			
Total	500.9667	29				

Table B.7 Analysis for domain value between visualization tools.

Spotfire vs. GeneSpring

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Spotfire	6	66	11	12		
GeneSpring	6	40	6.666667	4.266667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	56.33333	1	56.33333	6.92623	0.025086	4.964591
Within Groups	81.33333	10	8.133333			
Total	137.6667	11				

Table B.8 Analysis for domain value between Spotfire and GeneSpring.

B.2.3 Time to First Insight

Overall

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	6	28	4.666667	4.266667		
Timesearcher	6	42	7	12		
HCE	6	93	15.5	103.5		
Spotfire	6	45	7.5	15.1		
GeneSpring	6	95	15.83333	34.16667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	647.5333	4	161.8833	4.788503	0.005246	2.75871
Within Groups	845.1667	25	33.80667			
Total	1492.7	29				

Table B.9 Analysis for Time to first insight for the visualization tools.

B.2.4 Total Time

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	6	168	28	20		
Timesearcher	6	225	37.5	77.5		
HCE	6	210	35	110		
Spotfire	6	265	44.16667	44.1666667		
GeneSpring	6	315	52.5	27.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2083.533	4	520.8833	9.32925373	9.34E-05	2.75871
Within Groups	1395.833	25	55.83333			
Total	3479.367	29				

Table B.10 Analysis for Total time for the visualization tools.

B.2.5 Final Amount Learned

Anova: Single Factor					

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	6	245	40.83333	104.1667		
Timesearcher	6	290	48.33333	746.6667		
HCE	6	250	41.66667	226.6667		
Spotfire	6	400	66.66667	296.6667		
GeneSpring	6	310	51.66667	566.6667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2620	4	655	1.687419	0.184365	2.75871
Within Groups	9704.167	25	388.1667			
Total	12324.17	29				

Table B.11 Analysis for final amount learnt for the visualization tools.

B.3 Timeseries Data Analysis

B.3.1 Domain Value

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	2	11	5.5	0.5		
Timesearcher	2	26	13	18		
HCE	2	6	3	2		
Spotfire	2	27	13.5	0.5		
GeneSpring	2	15	7.5	4.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	171	4	42.75	8.382353	0.019265	5.192168
Within Groups	25.5	5	5.1			
Total	196.5	9				

Table B. 12 Analysis for domain value for the visualization tools.

B.3.2 Time to First Insight

Overall

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	2	10	5	0		
Timesearcher	2	7	3.5	4.5		
HCE	2	40	20	0		
Spotfire	2	12	6	32		
GeneSpring	2	20	10	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	354.4	4	88.6	12.13699	0.008703	5.192168
Within Groups	36.5	5	7.3			
Total	390.9	9				

Table B.13 Analysis for Time to first insight for the visualization tools.

GeneSpring, Timesearcher and Clusterview

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	2	10	5	0		
Timesearcher	2	7	3.5	4.5		
GeneSpring	2	20	10	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	46.33333333	2	23.16666667	15.44444444	0.026339	9.552094
Within Groups	4.5	3	1.5			
Total	50.83333333	5				

Table B.14 Analysis for Time to first insight for Clusterview, Timesearcher, and GeneSpring.

B.4 Lupus Data Analysis

B.4.1 Domain Value

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Clusterview	2	23	11.5	4.5		
Timesearcher	2	11	5.5	4.5		
HCE	2	2	1	0		
Spotfire	2	24	12	8		
GeneSpring	2	11	5.5	12.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	171.4	4	42.85	7.262712	0.025885	5.192168
Within Groups	29.5	5	5.9			
Total	200.9	9				

Table B.15 Analysis for domain value for visualization tools on Lupus data.

B.5 Tools vs. Datasets

B.5.1 HCE vs Datasets

Domain Value

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Timeseries	2	6	3	2		
Virus	2	26	13	8		
Lupus	2	2	1	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	165.3333	2	82.66667	24.8	0.013621	9.552094
Within Groups	10	3	3.333333			

Total	175.3333	5				

Table B.16 Analysis for domain value for HCE by datasets.

Number of Insights

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Timeseries	2	3	1.5	0.5		
Virus	2	9	4.5	0.5		
Lupus	2	2	1	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	14.33333	2	7.166667	21.5	0.016655	9.552094
Within Groups	1	3	0.333333			
Total	15.33333	5				

Table B.17 Analysis for number of insights for HCE by datasets.

Time to First Insight

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Timeseries	2	40	20	0		
Virus	2	8	4	2		
Lupus	2	35	17.5	12.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	296.3333	2	148.1667	30.65517	0.010075	9.552094
Within Groups	14.5	3	4.833333			
Total	310.8333	5				

Table B.18 Analysis for time to first insight for HCE by datasets.

Total Time

Anova: Single Factor						
SUMMARY						

Groups	*Count*	*Sum*	*Average*	*Variance*		
Timeseries	2	85	42.5	12.5		
Virus	2	80	40	50		
Lupus	2	45	22.5	12.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	475	2	237.5	9.5	0.050356	9.552094
Within Groups	75	3	25			
Total	550	5				

Table B.19 Analysis for total time participants spent with HCE in the study by datasets.

Appendix C: Data analysis for Chapter 5

C.1 Task-based Analysis

C.1.1 Overall Analysis

ANOVA for Accuracy

```
Factor         Type    Levels  Values
Visualization  fixed        3  1 Timepoint, M Graphs, M Timepoints
Participants   random      10  p1, p10, p2, p3, p4, p5, p6, p7, p8, p9
Tasks          fixed        7  T1, T2, T3, T4, T5, T6, T7
Analysis of Variance for Accuracy
Source          DF       SS      MS     F      P
Visualization    2   0.7524  0.3762  1.80  0.168
Participants     9   2.6143  0.2905  1.39  0.194
Tasks            6   3.5619  0.5937  2.84  0.011
Error          192  40.0667  0.2087
Total          209  46.9952

S = 0.456816   R-Sq = 14.74%   R-Sq(adj) = 7.19%
```

Figure C.1 ANOVA analysis for accuracy between the visualization types for the 10 participants on the pre-selected 7 tasks.

ANOVA for Performance Time

```
Factor         Type    Levels  Values
Visualization  fixed        3  1 Timepoint, M Graphs, M Timepoints
Participants   random      10  p1, p10, p2, p3, p4, p5, p6, p7, p8, p9
Tasks          fixed        7  T1, T2, T3, T4, T5, T6, T7
Analysis of Variance for performance time
Source          DF      SS     MS      F      P
Visualization    2    1761    880   0.53  0.588
Participants     9   37540   4171   2.52  0.009
Tasks            6  249131  41522  25.10  0.000
Error          192  317606   1654
Total          209  606038

S = 40.6718   R-Sq = 47.59%   R-Sq(adj) = 42.95%
```

Figure C.2 ANOVA analysis for performance time between the visualization types for the 10 participants on the pre-selected 7 tasks.

ANOVA for accuracy between 1 Timepoint vs. M Timepoints visualization

```
Factor         Type    Levels  Values
Participats    random      10  p1, p10, p2, p3, p4, p5, p6, p7, p8, p9
Visualization  fixed        2  1 Timepoint, M Timepoints
Tasks          fixed        7  T1, T2, T3, T4, T5, T6, T7
Analysis of Variance for Accuracy
Source                DF       SS      MS     F      P
Participats            9   2.7429  0.3048  1.52  0.149
Visualization          1   0.7143  0.7143  3.56  0.062
Tasks                  6   3.2857  0.5476  2.73  0.016
Visualization*Tasks    6   0.6857  0.1143  0.57  0.753
Error                117  23.4571  0.2005
Total                139  30.8857

S = 0.447759   R-Sq = 24.05%   R-Sq(adj) = 9.77%
```

Figure C.3 ANOVA analysis for Accuracy between 1 Tpt and M Tpts visualization for the 10 participants on the pre-selected 7 tasks.

C.1.2 Individual Task Analysis for Accuracy

Analysis for task 1

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	8	0.8	0.177778		
M Timepoints	10	7	0.7	0.233333		
M Graphs	10	5	0.5	0.277778		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.466667	2	0.233333	1.016129	0.375422	3.354131
Within Groups	6.2	27	0.22963			
Total	6.666667	29				

Table C.1Anova analysis for task 1 on accuracy.

Analysis for task 2

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	8	0.8	0.177778		
M Timepoints	10	4	0.4	0.266667		
M Graphs	10	4	0.4	0.266667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.066667	2	0.533333	2.25	0.124801	3.354131
Within Groups	6.4	27	0.237037			
Total	7.466667	29				

Table C.2 Anova analysis for task 2 on accuracy.

Analysis for task 3

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	8	0.8	0.177778		
M Timepoints	10	8	0.8	0.177778		

M Graphs	10	6	0.6	0.266667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.266667	2	0.133333	0.642857	0.533647	3.354131
Within Groups	5.6	27	0.207407			
Total	5.866667	29				

Table C.3 Anova analysis for task 3 on accuracy.

Analysis for task 4

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	10	1	0		
M Timepoints	10	7	0.7	0.233333		
M Graphs	10	10	1	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.6	2	0.3	3.857143	0.033616	3.354131
Within Groups	2.1	27	0.077778			
Total	2.7	29				

Table C.4 Anova analysis for task 4 on accuracy.

Analysis for task 5

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	7	0.7	0.233333		
M Timepoints	10	3	0.3	0.233333		
M Graphs	10	6	0.6	0.266667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.866667	2	0.433333	1.772727	0.189074	3.354131
Within Groups	6.6	27	0.244444			

Total	7.466667	29				

Table C.5Anova analysis for task 5 on accuracy.

1 Timepoint vs. M Timepoint – task 5

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	7	0.7	0.233333		
M Timepoints	10	3	0.3	0.233333		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.8	1	0.8	3.428571	0.080554	4.413873
Within Groups	4.2	18	0.233333			
Total	5	19				

Table C. 6 Anova analysis for task 5 between.

Analysis for task 6

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	4	0.4	0.266667		
M Timepoints	10	3	0.3	0.233333		
M Graphs	10	7	0.7	0.233333		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.866667	2	0.433333	1.772727	0.189074	3.354131
Within Groups	6.6	27	0.244444			
Total	7.466667	29				

Table C.7 Anova analysis for task 6 on accuracy.

M Timepoints vs. M Graphs – task 6

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoints	10	3	0.3	0.233333		
M Graphs	10	7	0.7	0.233333		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.8	1	0.8	3.428571	0.080554	4.413873
Within Groups	4.2	18	0.233333			
Total	5	19				

Table C.8 Anova analysis between M Timepoints and M Graphs for accuracy on task 6.

Analysis for task 7

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	7	0.7	0.233333		
M Timepoints	10	7	0.7	0.233333		
M Graphs	10	7	0.7	0.233333		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	-2.7E-15	2	-1.3E-15	-5.7E-15	#NUM!	3.354131
Within Groups	6.3	27	0.233333			
Total	6.3	29				

Table C.9 Anova analysis for task 7 on accuracy.

C.1.3 Individual Task Analysis for Performance Time

Analysis for task 1

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	1243	124.3	1046.011		
M Timepoints	10	1057	105.7	2149.344		
M Graphs	10	1646	164.6	4184.711		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	18130.87	2	9065.433	3.685102	0.038455	3.354131
Within Groups	66420.6	27	2460.022			

Total	84551.47	29				

Table C.10 Anova analysis for task 1 on performance time.

Analysis for task 2

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	1043	104.3	3414.233		
M Timepoints	10	1373	137.3	9381.122		
M Graphs	10	1008	100.8	3127.289		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8111.667	2	4055.833	0.764163	0.475533	3.354131
Within Groups	143303.8	27	5307.548			
Total	151415.5	29				

Table C.11 Anova analysis for task 2 on performance time.

Analysis for task 3

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	906	90.6	2223.6		
M Timepoints	10	670	67	1145.333		
M Graphs	10	909	90.9	2278.989		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3760.867	2	1880.433	0.998827	0.381515	3.354131
Within Groups	50831.3	27	1882.641			
Total	54592.17	29				

Table C.12 Anova analysis for task 3 on performance time.

Analysis for task 4

Anova: Single Factor					

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	235	23.5	92.94444		
M Timepoints	10	469	46.9	1018.767		
M Graphs	10	258	25.8	317.9556		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3326.867	2	1663.433	3.490534	0.044845	3.354131
Within Groups	12867	27	476.5556			
Total	16193.87	29				

Table C.13Anova analysis for task 4 on performance time.

M timepoints vs. M Graphs – task 4

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoints	10	788	78.8	265.9556		
M Graphs	10	538	53.8	471.2889		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3125	1	3125	8.477514	0.009309	4.413873
Within Groups	6635.2	18	368.6222			
Total	9760.2	19				

Table C. 14 Anova analysis on performance time between M Timepoints vs. M Graphs on task 4.

Analysis for task 5

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	235	23.5	92.94444		
M Timepoints	10	469	46.9	1018.767		
M Graphs	10	258	25.8	317.9556		
ANOVA						
Source of	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Variation						
Between Groups	3326.867	2	1663.433	3.490534	0.044845	3.354131
Within Groups	12867	27	476.5556			
Total	16193.87	29				

Table C.15 Anova analysis for task 5 on performance time.

Analysis for task 6

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	601	60.1	245.6556		
M Timepoints	10	740	74	1694		
M Graphs	10	488	48.8	396.6222		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3186.467	2	1593.233	2.045861	0.148835	3.354131
Within Groups	21026.5	27	778.7593			
Total	24212.97	29				

Table C.16 Anova analysis for task 6 on performance time.

1 Timepoint vs. M Timepoints – task 6

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Multiple Condition	10	740	74	1694		
Small Multiples	10	488	48.8	396.6222		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3175.2	1	3175.2	3.037565	0.098415	4.413873
Within Groups	18815.6	18	1045.311			
Total	21990.8	19				

Table C.17 Anova analysis on performance time between 1 Timepoint and M Timepoints on task 6.

Analysis for task 7

Anova: Single Factor					

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	360	36	340.6667		
M Timepoints	10	448	44.8	314.4		
M Graphs	10	313	31.3	51.12222		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	939.2667	2	469.6333	1.995075	0.155558	3.354131
Within Groups	6355.7	27	235.3963			
Total	7294.967	29				

Table C.18 Anova analysis for task 7 on performance time.

M Timepoints vs. M Graphs – task 7

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Multiple Condition	10	448	44.8	314.4		
Small Multiples	10	313	31.3	51.12222		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	911.25	1	911.25	4.986017	0.03849	4.413873
Within Groups	3289.7	18	182.7611			
Total	4200.95	19				

Table C.19 Anova analysis between M Timepoints and M Graphs on performance time for task 7.

C.2 Insight-Based Analysis

C.2.1 Overall Performance

Time spent in the study

Participants	**1 Timepoint**	**M Timepoints**	**M Graphs**
p1	29	12	10
p2	16	18	26
p3	21	35	23
p4	35	18	6
p5	17	16	5
p6	25	20	5

p9	18	11	12
p10	12	10	10
Avg Time	**20.5**	**16.9**	**10.8**

Table C.20 Average time participants spent in the study for all three visualization alternatives.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	205	20.5	47.38889		
M Timepoints	10	169	16.9	60.32222		
M Graphs	10	108	10.8	58.84444		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	480.8667	2	240.4333	4.330687	0.023374	3.354131
Within Groups	1499	27	55.51852			
Total	1979.867	29				

Table C.21: Anova analysis for total time participants spent in the study for all three visualization alternatives.

Count of Insights

Participants	**1 Timepoint**	**M Timepoints**	**M Graphs**
p1	26	13	10
p2	18	16	26
p3	18	24	23
p4	33	8	6
p5	22	4	5
p6	14	15	5
p7	15	12	6
p8	20	18	5
p9	11	8	12
p10	6	6	10
Avg. Count	**18.3**	**12.4**	**10.8**

Table C.22 Count of Insights for all three visualization alternatives for the participants.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	183	18.3	58.45556		
M Timepoints	10	124	12.4	37.37778		
M Graphs	10	108	10.8	58.84444		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	312.0667	2	156.0333	3.026291	0.065184	3.354131
Within Groups	1392.1	27	51.55926			
Total	1704.167	29				

Table C.23 Anova analysis for differences between the three visualization alternatives on count of insights.

1 Timepoint vs. M Timepoints – Count of Insights

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	183	18.3	58.45556		
M Timepoints	10	124	12.4	37.37778		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	174.05	1	174.05	3.632348	0.072764	4.413873
Within Groups	862.5	18	47.91667			
Total	1036.55	19				

Table C.24 Anova analysis for differences in count of insights between 1 tpt and MTpts. visualization alternatives.

1 Timepoint vs. M Graphs – Count of Insights

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	183	18.3	58.45556		
M Graphs	10	108	10.8	58.84444		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	281.25	1	281.25	4.795396	0.041949	4.413873
Within Groups	1055.7	18	58.65			
Total	1336.95	19				

Table C.25Anova analysis for differences in count of insights between 1 tpt and M graph visualization alternatives.

C.2.2 Performance based on Insight Categories

Total count of insights for each category

Participants	Gene expression	Topology	Topology + expression	Timepoint analysis	Condition	Summary	Outliers
p1	10	3	6	5	1	1	0
p2	1	2	6	3	5	1	0
p3	3	0	5	4	3	0	3
p4	16	0	0	7	7	2	1
p5	8	3	0	7	3	1	0
p6	4	3	2	0	4	0	1
p7	2	1	3	3	4	1	1
p8	9	1	4	2	3	1	0
p9	0	0	1	3	5	2	0
p10	0	0	0	2	3	1	0
Total	**53**	**13**	**27**	**36**	**38**	**10**	**6**

Table C.26 Number of insights, grouped by participants for the selected insight categories for 1timepoint visualization alternative.

Participants	Gene expression	Topology	Topology + expression	Timepoint analysis	Condition	Summary	Outliers
p1	4	3	2	0	1	1	2
p2	5	1	3	0	5	2	0
p3	10	0	8	0	4	1	1
p4	1	1	3	0	1	1	1
p5	2	0	0	0	2	0	0
p6	4	4	1	0	4	0	2
p7	5	0	3	0	3	0	1
p8	8	4	2	0	3	0	1
p9	1	0	2	0	4	0	1
p10	0	0	2	0	2	1	1
Total	**40**	**13**	**26**	**0**	**29**	**6**	**10**

Table C.27 Number of insights, grouped by participants for the selected insight categories for M timepoints visualization alternative.

Participants	Gene expression	Topology	Topology + expression	Timepoint analysis	Condition	Summary	Outliers
p1	1	0	3	1	2	0	0
p2	4	0	1	1	4	1	0
p3	8	0	2	2	3	0	1
p4	0	0	0	0	4	0	0
p5	0	0	0	2	2	0	0
p6	0	0	0	3	2	0	0
p7	0	0	0	3	3	0	0
p8	0	0	0	3	1	1	0
p9	3	0	3	2	4	0	1

p10	1	0	0	1	3	1	0
Total	**17**	**0**	**9**	**18**	**28**	**3**	**2**

Table C.28 Number of insights, grouped by participants for the selected insight categories for multiple graphs (M Graphs) visualization alternative.

Insight Category: Gene Expression

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	53	5.3	27.78889		
M Graphs	10	17	1.7	6.9		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	64.8	1	64.8	3.736067	0.069139	4.413873
Within Groups	312.2	18	17.34444			
Total	377	19				

Table C. 29 Anova analysis for insight category: gene expression between 1 tpt and M graph visualization alternative.

Insight category: Topology

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	13	1.3	1.788889		
M Timepoints	10	13	1.3	2.9		
M Graphs	10	0	0	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11.26667	2	5.633333	3.604265	0.040982	3.354131
Within Groups	42.2	27	1.562963			
Total	53.46667	29				

Table C.30 Anova analysis for insight category: Topology between all the three visualization alternatives.

Insight Category: Topology + Expression

Anova: Single Factor						
SUMMARY						

Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	27	2.7	6.011111		
M Timepoints	10	26	2.6	4.488889		
M Graphs	10	9	0.9	1.655556		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	20.46667	2	10.23333	2.525594	0.098745	3.354131
Within Groups	109.4	27	4.051852			
Total	129.8667	29				

Table C.31 Anova analysis for insight category: Topology + Expression between all three visualization alternatives.

M Timepoints vs. M Graphs – Topology + Expression

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoints	10	26	2.6	4.488889		
M Graphs	10	9	0.9	1.655556		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	14.45	1	14.45	4.703436	0.043744	4.413873
Within Groups	55.3	18	3.072222			
Total	69.75	19				

Table C.32 Anova analysis for insight category: Topology + Expression between M Tpt and M Graph visualization alternatives.

1 Timepoint vs. M Graphs – Topology + Expression

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	27	2.7	6.011111		
M Graphs	10	9	0.9	1.655556		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Between Groups	16.2	1	16.2	4.226087	0.054608	4.413873
Within Groups	69	18	3.833333			
Total	85.2	19				

Table C. 33 Anova analysis for insight category: Topology + Expression between 1 Tpt and M Graph visualization alternatives.

Insight Category: Timepoint Analysis

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	36	3.6	4.933333		
M Timepoints	10	0	0	0		
M Graphs	10	18	1.8	1.066667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	64.8	2	32.4	16.2	2.38393E-05	3.354131
Within Groups	54	27	2			
Total	118.8	29				

Table C.34 Anova analysis for insight category: Timepoint analysis between all three visualization alternatives.

1 Timepoint vs. M Timepoints – Timepoint Analysis

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	36	3.6	4.933333		
M Timepoints	10	0	0	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	64.8	1	64.8	26.27027	7.08216E-05	4.413873
Within Groups	44.4	18	2.466667			
Total	109.2	19				

Table C.35 Anova analysis for insight category: Timepoint analysis between 1 Tpt and M Tpt visualization alternatives.

M Timepoints vs. M Graphs – Timepoint Analysis

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoints	10	0	0	0		
M Graphs	10	18	1.8	1.066667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	16.2	1	16.2	30.375	3.11232E-05	4.413873
Within Groups	9.6	18	0.533333			
Total	25.8	19				

Table C.36 Anova analysis for insight category: Timepoint analysis between M Tpt and M Graph visualization alternatives.

1 Timepoint vs. M Graphs – Timepoint Analysis

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	36	3.6	4.933333		
M Graphs	10	18	1.8	1.066667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	16.2	1	16.2	5.4	0.032044676	4.413873
Within Groups	54	18	3			
Total	70.2	19				

Table C.37 Anova analysis for insight category: Timepoint + Expression between 1 Tpt and M Graph visualization alternatives

Insight Category: Condition

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	38	3.8	2.622222		
M Timepoints	10	29	2.9	1.877778		
M Graphs	10	28	2.8	1.066667		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6.066667	2	3.033333	1.634731	0.213721	3.354131
Within Groups	50.1	27	1.855556			
Total	56.16667	29				

Table C.38 Anova analysis for insight category: condition between all three visualization alternatives.

Insight Category: Outliers

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	6	0.6	0.933333		
M Timepoints	10	10	1	0.444444		
M Graphs	10	2	0.2	0.177778		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.2	2	1.6	3.085714	0.0621	3.354131
Within Groups	14	27	0.518519			
Total	17.2	29				

Table C.39 Anova analysis for insight category: outliers between all three visualization alternatives.

M Timepoint vs. M Graphs – Outliers

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoint	10	10	1	0.444444		
M Graphs	10	2	0.2	0.177778		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.2	1	3.2	10.28571	0.004885	4.413873
Within Groups	5.6	18	0.311111			
Total	8.8	19				

Table C.40 Anova analysis for insight category: outliers between M Tpt and M graph visualization alternatives.

Insight Category: Summary

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	9	9	1	0.5		
M Timepoints	9	5	0.555556	0.527778		
M Graphs	9	3	0.333333	0.25		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.074074	2	1.037037	2.434783	0.108956	3.402826
Within Groups	10.22222	24	0.425926			
Total	12.2963	26				

Table C.41 Anova analysis for insight category: summary between all three visualization alternatives.

1 Timepoint vs. M Graphs – Summary

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint	10	10	1	0.444444		
M Graphs	10	3	0.3	0.233333		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.45	1	2.45	7.229508	0.015005	4.413873
Within Groups	6.1	18	0.338889			
Total	8.55	19				

Table C.42 Anova analysis for insight category: summary between 1 Tpt and M graph visualization alternatives.

C.3 Comparison between Methods

C.3.1 Total Time Spent

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Task-based	30	265	8.833333	5.522989		
Insight-based	30	482	16.06667	68.27126		

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	784.8167	1	784.8167	21.2704	2.25E-05	4.006873
Within Groups	2140.033	58	36.89713			
Total	2924.85	59				

Table C.43 Anova analysis for total time participants spent in the study for Task-based vs. Insight-based methods.

1 Timepoint Visualization – Time

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1 Timepoint - TB	10	85	8.5	3.388889		
1 Timepoint - Insight	10	205	20.5	47.38889		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	720	1	720	28.35886	4.62E-05	4.413873
Within Groups	457	18	25.38889			
Total	1177	19				

Table C.44 Anova analysis for total time participants spent in the study for 1 timepoint visualization for Task-based vs. Insight-based methods.

M Timepoints Visualization – Time

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
M Timepoints-TB	10	92	9.2	5.511111		
M Timepoints-Insight	10	169	16.9	60.32222		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	296.45	1	296.45	9.006076	0.007669	4.413873

Within Groups	592.5	18	32.91667			
Total	888.95	19				

Table C.45 Anova analysis for total time participants spent in the study for M timepoints visualization for Task-based vs. Insight-based method.