Andromeda in Education: 
Studies on Student Collaboration and Insight Generation with 
Interactive Dimensionality Reduction

Mia R. Taylor

Thesis submitted to the Faculty of the 
Virginia Polytechnic Institute and State University 
in partial fulfillment of the requirements for the degree of

Master of Science in 
Computer Science and Applications

Christopher L. North, Chair 
Leanna L. House 
Yalong Yang

September 20, 2022 
Blacksburg, Virginia

Keywords: data analysis, dimensionality reduction, education, natural language processing

Copyright 2022, Mia R. Taylor
Andromeda in Education:
Studies on Student Collaboration and Insight Generation with Interactive Dimensionality Reduction

Mia R. Taylor

(ABSTRACT)

Andromeda is an interactive visualization tool that projects high-dimensional data into a scatterplot-like visualization using Weighted Multidimensional Scaling (WMDS). The visualization can be explored through surface-level interaction (viewing data values), parametric interaction (altering underlying parameterizations), and observation-level interaction (directly interacting with projected points). This thesis presents analyses on the collaborative utility of Andromeda in a middle school class and the insights college-level students generate when using Andromeda. The first study discusses how a middle school class collaboratively used Andromeda to explore and compare their engineering designs. The students analyzed their designs, represented as high-dimensional data, as a class. This study shows promise for introducing collaborative data analysis to middle school students in conjunction with other technical concepts such as the engineering design process. Participants in the study on college-level students were given a version of Andromeda, with access to different interactions, and were asked to generate insights on a dataset. By applying a novel visualization evaluation methodology on students’ natural language insights, the results of this study indicate that students use different vocabulary supported by the interactions available to them, but not equally. The implications, as well as limitations, of these two studies are further discussed.
Andromeda in Education:
Studies on Student Collaboration and Insight Generation with Interactive Dimensionality Reduction

Mia R. Taylor

(GENERAL AUDIENCE ABSTRACT)

Data is often high-dimensional. A good example of this is a spreadsheet with many columns. Visualizing high-dimensional data is a difficult task because it must capture all information in 2 or 3 dimensions. Andromeda is a tool that can project high-dimensional data into a scatterplot-like visualization. Data points that are considered similar are plotted near each other and vice versa. Users can alter how important certain parts of the data are to the plotting algorithm as well as move points directly to update the display based on the user-specified layout. These interactions within Andromeda allow data analysts to explore high-dimensional data based on their personal sensemaking processes. As high-dimensional thinking and exploratory data analysis are being introduced into more classrooms, it is important to understand the ways in which students analyze high-dimensional data. To address this, this thesis presents two studies. The first study discusses how a middle school class used Andromeda for their engineering design assignments. The results indicate that using Andromeda in a collaborative way enriched the students’ learning experience. The second study analyzes how college-level students, when given access to different interaction types in Andromeda, generate insights into a dataset. Students use different vocabulary supported by the interactions available to them, but not equally. The implications, as well as limitations, of these two studies are further discussed.
Dedication

To Mom, Dad, and Kevin
Acknowledgments

Firstly, I would like to thank my advisor Dr. Christopher North for his continued support in both my academics and future career.

I would also like to thank my great collaborators who have helped me navigate this degree: Dr. Leanna House, Mr. Danny Mathieson, Dr. Lata Kodali, Dr. Yalong Yang, and more.

Lastly, I would like to thank my partner Kendall Knight.
# Contents

List of Figures \hspace{1cm} x

List of Tables \hspace{1cm} xiii

1 Introduction \hspace{1cm} 1

1.1 Research Motivations \hspace{1cm} 2

1.1.1 Collaborative Exploratory Data Analysis \hspace{1cm} 2

1.1.2 Insight-based Visualization Evaluation \hspace{1cm} 3

1.2 Research Questions \hspace{1cm} 5

1.3 Contributions \hspace{1cm} 5

1.4 Thesis Organization \hspace{1cm} 6

2 Andromeda \hspace{1cm} 7

2.1 Surface-Level Interaction \hspace{1cm} 8

2.2 Parametric Interaction \hspace{1cm} 8

2.3 Observation-Level Interaction \hspace{1cm} 12

2.4 Studies in Education \hspace{1cm} 12

3 A Study on the Collaborative Utility of Andromeda in K-12 \hspace{1cm} 15
3.1 Introduction ....................................................... 15

3.2 Related Work .................................................. 16

3.2.1 Andromeda in K-12 Education ............................... 16

3.2.2 WhiteBox Learning ......................................... 17

3.3 Methods ........................................................ 18

3.4 Results .......................................................... 20

3.4.1 Shelter ....................................................... 20

3.4.2 Dragster ...................................................... 21

3.5 Discussion ...................................................... 22

3.5.1 Collaboration: Andromeda supported student collaborations ................................. 22

3.5.2 Data Analysis: Andromeda helped students practice data analysis .......................... 23

3.5.3 Design Influence: Andromeda influenced the engineering design process ............ 28

3.5.4 Sentiment: The class had a positive experience using Andromeda ..................... 28

3.6 Limitations ...................................................... 29

3.7 Conclusions ..................................................... 30

4 A Study on College Students’ Insights using Andromeda .......................... 32

4.1 Introduction ..................................................... 32

4.2 Related Work .................................................. 34

4.3 Methods ........................................................ 37
Appendix B  Andromeda Activity Worksheet for K-12 Students

B.1  Worksheet with Solutions ........................................ 64

B.2  Worksheet without Solutions ................................. 66
List of Figures

2.1 Surface-level interaction (SLI) in Andromeda with a reduced Animals dataset imported. This is the initial projection with all variables weighted equally. Applying SLI, the Elephant point was clicked, and the cursor is hovering over Squirrel. The attribute, or feature, values of Elephant and Squirrel are shown on the right-hand side in orange and yellow, respectively. Note that applying SLI does not affect the projection nor the variable weights. 

2.2 Parametric interaction (PI) in Andromeda with a reduced Animals dataset imported. The sliders for Size and Speed were dragged to the right to increase their weight. The layout differs from the initial layout fig. 2.1 that relied on equal weights for all variables. Hovering over the variable Size changes the size of the animal circles to be proportional with the animal’s Size value. Because Size and Speed have higher weights, animals with similar Size and Speed values tend to be projected near each other such as the Siamese Cat and Squirrel.

2.3 Observation-level interaction (OLI) in Andromeda with a reduced Animals dataset imported (before layout update). The Squirrel and Siamese Cat were dragged close together in the top left. Elephant and Blue Whale in the bottom right. Clicking the update layout button will learn new variable weights that create a projection in which the groups are dissimilar, but the points in a group are similar.
2.4 Observation-level interaction (OLI) in Andromeda with a reduced *Animals* dataset imported (after layout update). After clicking the update layout button, the layout changes by learning variable weights that describe how the dragged points in fig. 2.3 are similar or dissimilar. These learned weights are applied to the entire projection. As shown by the high *Furry* variable weight, *Furry* best describes how the dragged points relate to each other.

3.1 Example screenshot of the WhiteBox Learning Shelter 2.0 assignment.

3.2 Survey statements given to the students. The students rated each statement on a scale from 0 (strongly disagree) to 10 (strongly agree). Students rated statements 7a and 7b after completing the shelter and dragster assignments, respectively.

3.3 Survey responses after the shelter assignment. The mean ($\bar{x}$) and standard deviation (s) for each survey question is provided with a visual count of responses.

3.4 Survey responses after the dragster assignment. The mean ($\bar{x}$) and standard deviation (s) for each survey question is provided with a visual count of responses.

3.5 An example of two students comparing and contrasting their shelters. Each point is assigned a unique number and represents a student’s design for the shelter assignment. The exact scaling has been altered for purposes of presentation in this paper.
3.6 An example of the class analyzing if color correlates with dragster speed. Each point is assigned a unique number and represents a student’s design for the dragster assignment. The exact scaling has been altered for purposes of presentation in this paper.

3.7 An example of the class analyzing the fastest dragsters. Each point is assigned a unique number and represents a student’s design for the dragster assignment. The exact scaling has been altered for purposes of presentation in this paper.
List of Tables

4.1 Simplified input of SIEVE that consists of two insights. Each insight has an associated group: A or B. ................................. 33

4.2 Simplified output of SIEVE. The term *Axis* results in an increase in probability that insights were generated using visualization A, while the term *Cluster* results in an increase for visualization B. This also means that *Cluster* results in a decrease in probability that insights were generated using visualization A. The details of the output (*Beta, Significance*) are discussed in ??... 34

4.3 Significant keyword differences between insights generated with versions of Andromeda. Each comparison compares two groups of insights, such as OLI and PI in the first row. *Similar* is a significant term when comparing OLI and PI insights. An insight is more likely to belong to (*Prob. ↑*) the OLI insights if it uses the term *Similar* more. A single asterisk (*) indicates that $p < 0.1$, a double asterisk (**) indicates that $p < 0.05$, and a triple asterisk (***) indicates that $p < 0.01$. See ?? for an explanation of how significance is determined. ................................................................. 41

4.4 Keywords associated with an increase in probability that an insight belongs to a version of Andromeda based on table 4.3. An insight is more likely to be generated using OLI if it uses the words *Similar, Away, Far*, and/or *Group* more. Note that *Weight* and *Variable* are associated with PI and BOTH. . . 42
4.5 Significant keyword differences between insights generated before and after recitation with a version of Andromeda. **Even** is a significant term when comparing **NONE** insights. An insight is more likely to belong to (Prob. ↑) the Pre-recitation insights if it uses the term **Even** more. A single asterisk (*) indicates that $p < 0.1$, a double asterisk (**) indicates that $p < 0.05$, and a triple asterisk (***) indicates that $p < 0.01$. See ?? for an explanation of how significance is determined.
Chapter 1

Introduction

As data analysis becomes increasingly relevant in early education, it is important to understand how students use data. Students may be intimidated by concepts such as high-dimensional data or the underlying math that supports data science. Data is considered high-dimensional when it has more than 3 dimensions. It can be difficult to think beyond two- or three-dimensions. Thus, it is important to engage students in high-dimensional thinking to encourage building data science skills throughout their entire education. This will better prepare them for future moments when they are expected to understand and draw conclusions from large datasets.

The software Andromeda is an interactive, high-dimensional data analysis tool that enables novice analysts to explore data in parallel with their personal sensemaking processes. Andromeda uses the dimensionality reduction algorithm known as Weighted Multidimensional Scaling (WMDS) [1] to project high-dimensional data into a scatterplot-like visualization. Users are not required to understand the underlying math behind dimensionality reduction to meaningfully use it. Thus, users are not required to have a strong math or data science background to use the software. With Andromeda, novice analysts can perform exploratory data analysis (EDA) by visualizing high-dimensional data in a two-dimensional space such that points follow the “near == similar” metaphor. Andromeda is used in visualization and education research to analyze how users interact with data and conduct EDA.
CHAPTER 1. INTRODUCTION

This thesis aims to further research on Andromeda in education by (1) conducting a case study on Andromeda as collaborative, supplemental lesson material in K-12 education and (2) applying a novel methodology to summarize the differences between insights generated by college students using Andromeda.

1.1 Research Motivations

1.1.1 Collaborative Exploratory Data Analysis

Exploratory data analysis (EDA) is an approach to data analysis that is mostly graphical and seeks to maximize insight into a dataset, test underlying assumptions, and more [2]. Generally, it is a way to explore a dataset and understand its basic structure. It is important for building data analysis skills and confidence in one’s own abilities. Previous studies [3, 4, 5] show promising results for introducing collaborative activities in the classroom to teach computational thinking skills. Collaboration has been shown to help students earn better scores, feel more confident in their abilities, and have greater creativity when using their abilities. These studies, however, do not focus on collaborative exploratory data analysis.

There exist a few studies on collaborative EDA. For example, CoDA is a collaborative EDA system that enables multiple students to physically interact with a single scatterplot display [6]. This system was used in a study to understand how interactive data physicalizations support collaborative data analysis among teenage students. Some students got more insight or saw different perspectives of the data because of their collaboration with others indicating that the collaborative nature of EDA was beneficial. Another study sought to understand how analysts conduct collaborative EDA around a single, shared tabletop display [7]. The findings of the study suggest that closely coupled work and communication led to successful
task completion. There is also existing research on collaborative EDA with Andromeda. “Be the Data” is a series of studies on Andromeda where students physically embody data points in a workshop room [8, 9]. A large-scale projection of Andromeda was shown above the students. Students were engaged and learned high-dimensional data concepts. This indicated that, once again, the collaborative nature of EDA was beneficial. Despite the clear benefits of collaborative EDA, the studies all required specialized hardware to support students. This is not always feasible in a typical classroom setting. There are extra time and financial costs associated with specialized hardware. To address this, the first study in this thesis analyzes collaborative EDA with publicly accessible Andromeda. This is done by conducting a case study on the collaborative use of Andromeda in an 8th grade technology education class.

1.1.2 Insight-based Visualization Evaluation

Another approach to understanding how students use Andromeda is to analyze the insights that they generate when using the software. For this work, an insight is considered any observation or understanding of data. Analyzing a visualization by the insights it helps generate is a technique within visualization research. Previous studies analyze insights by interactions through interaction logs [10] or characteristics like degree of depth or relevance [11].

Insights generated using Andromeda have been studied a few times [12, 13, 14], however these methods analyze insights by how well they exhibit characteristics such as complexity. Manually annotating individual insights by subjective characteristics is an arduous task. The language used in insights can help demonstrate the ways that analysts use Andromeda without needing to manually annotate each insight. To address this, the second study in this
thesis creates a novel method, SIEVE, to understand the difference in insight language by comparing the experimental groupings of insights generated by college students at the beginning and end of an activity session with access to different interaction types in Andromeda. The method provides a summarization of the content difference between the experimental groupings of insights. Rather than replacing existing insight analysis methods, SIEVE is meant to help evaluators understand the insights and identify whether there are distinct differences in language between groups of insights. SIEVE can be used as a supplemental insight analysis method for insight-based visualization evaluation.

This method was created to supplement the insight analysis of a previously conducted experiment by Lata Kodali. The large-scale study was conducted primarily to understand students’ change in learning from different versions of Andromeda. Participants in this study were given a series of surveys. An example survey is shown in chapter A. The surveys consisted of analytical questions which have a correct answer and interpretive questions for which students provided insights into data. Lata Kodali previously analyzed the insights for their complexity (as described in subsection 4.3.1). The results showed that insights between experimental groupings generally did not differ in complexity. This analysis did not explore whether the insights significantly different in other ways such as by their language content.

For this thesis, SIEVE was created to further analyze the insights. The results from SIEVE provides a new perspective on student learning with Andromeda by looking at the keyword differences between groupings of insights. The results from this will be used as part of the analysis of the overall larger-scale study which analyzes all of the students’ responses to understand their change in learning with Andromeda.
1.2 Research Questions

This thesis focuses on interactive high-dimensional data analysis with Andromeda in educational settings by targeting the following research questions.

1. How can Andromeda be used to facilitate collaboration in the classroom to support learning data analysis and engineering design principles?

2. Within introductory data analytics education, how do the insights that students generate relate to interaction types available within Andromeda?

3. What method can be used to identify insight differences between experimental groups using natural language processing?

1.3 Contributions

The research questions defined above are addressed by the following contributions.

1. a case study of the collaborative utility of Andromeda within an 8th grade engineering design environment

2. an analysis of student insights generated using Andromeda in a large-scale study of a college introductory data analytics class

3. the novel method SIEVE that identifies significant keyword differences between experimental groups of natural language insights
1.4 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 describes the interactive, dimensionality reduction software Andromeda. Chapter 3 presents the case study analyzing how Andromeda was used within an 8th grade engineering design environment. Chapter 4 defines the novel method SIEVE and applies the method to a large-scale, controlled study of Andromeda. Chapter 5 presents general discussions and conclusions of the work described in this thesis.
Chapter 2

Andromeda

Andromeda is an interactive visualization and data analysis tool that was originally designed to enable analysts of all skill levels to explore high-dimensional data [15, 16]. The visualization relies on a dimensionality reduction algorithm. Dimensionality reduction algorithms take in high-dimensional data as input and outputs low-dimensional data that is representative of the input data. The low-dimensional data is usually represented in 2- or 3-dimensions for visualization-purposes. Andromeda specifically uses Weighted Multidimensional Scaling (WMDS) [17]. WMDS is a dimensionality reduction algorithm that associates each dimension in the data with a weight that represents the dimension’s relative importance in the visualization. With Andromeda, users can explore the dimension, or variable, weights and the projections to better understand the high-dimensional data.

All interactions within Andromeda are designed to be simple and interpretable without requiring the user to understand dimensionality reduction or the math that supports it. Users can immediately explore and interact with the data in a way that makes sense to them. This is both useful for professional analysts [18] and students alike [13, 14].

When Andromeda first opens, the WMDS algorithm initially places equal importance on all variables when projecting the data. The projection of a dataset describing animals is shown in fig. 2.1. The left-hand side of the display shows the data projection in a scatterplot-like manner. The right-hand side of the display shows each variable and its associated weight as a slider.
CHAPTER 2. ANDROMEDA

Andromeda is often studied using a dataset describing animals because analysts do not need any domain knowledge to understand the dataset. Any future reference to the Animals dataset is specifically referring to the dataset created by Xian et al. called Animals with Attributes 2 (AWA2) [19]. The following section describe the interaction types available in Andromeda using the Animals dataset as an example. There are three types of interactions in Andromeda that enable analysts to analyze high-dimensional data: surface-level interaction, parametric interaction, and observation-level interaction.

2.1 Surface-Level Interaction

Surface-level interaction (SLI) allows users to highlight one or more data points by clicking or hovering. This interaction enables users to view the data points’ values without altering the projection or variable weights. As shown in fig. 2.1, clicking on Elephant shows, as indicated by the orange highlighting, low values for Furry and Speed, moderate values for Grazer and Smart, and high values for Size and Walks. Hovering over a data point, such as Squirrel in fig. 2.1, shows its attribute values highlighted in yellow. Figure 2.1 and subsequent Andromeda figures are from the current web-version of Andromeda. The study described in chapter 4 used an older version of Andromeda that offered identical functionality with slightly different display changes such as color. The differences are negligible.

2.2 Parametric Interaction

The second interaction type is Parametric Interaction (PI) which allows users to change variable weights. Variable weights are represented by sliders on the right-hand side of the display. The weights of variables can be increased or decreased by dragging sliders to the
2.2. Parametric Interaction

Figure 2.1: Surface-level interaction (SLI) in Andromeda with a reduced Animals dataset imported. This is the initial projection with all variables weighted equally. Applying SLI, the Elephant point was clicked, and the cursor is hovering over Squirrel. The attribute, or feature, values of Elephant and Squirrel are shown on the right-hand side in orange and yellow, respectively. Note that applying SLI does not affect the projection nor the variable weights.
Figure 2.2: Parametric interaction (PI) in Andromeda with a reduced *Animals* dataset imported. The sliders for **Size** and **Speed** were dragged to the right to increase their weight. The layout differs from the initial layout fig. 2.1 that relied on equal weights for all variables. Hovering over the variable **Size** changes the size of the animal circles to be proportional with the animal’s **Size** value. Because **Size** and **Speed** have higher weights, animals with similar **Size** and **Speed** values tend to be projected near each other such as the *Siamese Cat* and *Squirrel*. 
2.3. Observation-Level Interaction

Figure 2.3: Observation-level interaction (OLI) in Andromeda with a reduced Animals dataset imported (before layout update). The Squirrel and Siamese Cat were dragged close together in the top left. Elephant and Blue Whale in the bottom right. Clicking the update layout button will learn new variable weights that create a projection in which the groups are dissimilar, but the points in a group are similar.

right or left, respectively. This assigns different levels of importance to the variables such that a variable with a greater weight influences the layout more than a variable with a lesser weight.

After adjusting the sliders, the display updates based on the new variable weights. Also, the sliders are updated to be in decreasing order by weight. Figure 2.2 shows an example of PI.
2.3 Observation-Level Interaction

The final type of interaction is Observation-Level Interaction (OLI) [20]. OLI allows users to reposition data points in the layout. This indirectly communicates variable weight changes via inverse WMDS [21]. After dragging points to different locations on the projection and clicking the “Update Layout” button, Andromeda solves for the optimal weights that preserve the user-defined projection. Then, Andromeda updates its display with the new weights. The sliders are updated to be in decreasing order by weight. Figures 2.3 and 2.4 show an example usage of OLI.

2.4 Studies in Education

Andromeda was originally designed to enable data analysts of all skill levels to explore high-dimensional data [16]. Andromeda is available publicly as a web application [22]. It is often studied in a usability or educational context.

There are multiple Andromeda studies on the software’s usability [13, 15, 23, 24, 25, 26]. The participants in the usability studies are most often graduate students, however the research itself does not focus on education.

“Be the Data” [8, 9] is the only previous study on Andromeda that analyzes how K-12 students interact with data. Virginia Tech hosted “Be the Data” events for various student groups ranging from 3rd grade to the undergraduate level. In “Be the Data”, students embodied data points and physically moved in a room to interact with WMDS. For example, using the animals dataset, each student represented an animal and their physical location in the workshop room was projected onto a large screen with Andromeda. Then, if a student moved from one side of the room to the other, their corresponding animal data point would...
Figure 2.4: Observation-level interaction (OLI) in Andromeda with a reduced *Animals* dataset imported (after layout update). After clicking the update layout button, the layout changes by learning variable weights that describe how the dragged points in fig. 2.3 are similar or dissimilar. These learned weights are applied to the entire projection. As shown by the high *Furry* variable weight, *Furry* best describes how the dragged points relate to each other.
move in the Andromeda projection as well. Collectively, classes would answer questions such as “What characteristics differentiate good pets from bad pets?” Data from these events suggested that the students were engaged and enabled to learn about high-dimensional data and analytics.

A few education studies [14, 24] analyzed how college students performed on a series of data analysis assignments. These studies characterized insights by describing the complexity of each insight and counting the diversity of tasks as described by Amar et al. [27]. Tasks are low-level analytic activity such as sorting, clustering, or filtering data. Results showed that students tended to think in low dimensions by default. When provided more complex data analysis tools, the students began to think with higher dimensionality. As a result of using Andromeda, students generated insights that were higher in dimensionality and more complex.

This thesis consists of two studies that extend the described research on Andromeda in education. The first case study follows a middle school class that uses Andromeda to conduct collaborative EDA for their engineering design assignments. This was using the publicly available web-based version of Andromeda. This is the first analysis of K-12 students using Andromeda in a typical classroom setting. The second study in this thesis analyzes how college students generated insights when using Andromeda. Students were given access to different interaction types and generated insights throughout their session. This study uses a novel insight analysis method rather than manually characterizing insights by their complexity and diversity. Both studies provide a new perspective on how Andromeda is used by students.
Chapter 3

A Study on the Collaborative Utility of Andromeda in K-12

The contents of this chapter are adapted from the author’s publication in the American Society for Engineering Education’s Annual Conference and Exposition in 2022. The original publication is titled “Andromeda in the Classroom: Collaborative Data Analysis for 8th Grade Engineering Design” [28].

3.1 Introduction

Collaborative data analysis enables students to explore and analyze high-dimensional data together. This case study shows how collaborative data analysis can be successfully integrated into teaching the engineering design process using Andromeda.

This case study performs an observational, qualitative analysis on the collaborative use of Andromeda in an 8th grade technology education class. Students were given two engineering projects through WhiteBox Learning: Survival Shelter 2.0 and Dragster 2.0. WhiteBox Learning is a web-based STEM education software that allows students to learn STEM concepts, such as introductory physics, and practice the engineering design process. Survival Shelter 2.0 and Dragster 2.0 are two design projects that let students create an emergency survival shelter for hikers and a $CO_2$ racecar, respectively. In this case, students used
WhiteBox Learning to create, analyze, and simulate their project designs. Between design iterations, the class explored their designs in Andromeda with the teacher acting as the facilitator. That is, the teacher uploaded data describing the students’ projects to Andromeda; each point in the visualization represented a student’s design. With the teacher controlling Andromeda, students used Andromeda to visualize, analyze, and compare their designs.

Data were collected from students, along with impressions from their teacher. This data are used to assess the success of the class collaboration. The potential implications of Andromeda as a public, educational resource are discussed. An example class activity aligned with Virginia’s proposed Standards of Learning in data science is included in chapter B. Hopefully, this will encourage educators to introduce collaborative data analysis tools such as Andromeda into their classrooms.

3.2 Related Work

3.2.1 Andromeda in K-12 Education

The software Andromeda and previous educational studies on it can be found in chapter 2. The majority of the studies focused on college students. Only one study, “Be the Data”, analyzed how K-12 students used Andromeda. While “Be the Data” proved a successful collaborative use of Andromeda, “Be the Data” also required large-scale technological infrastructure such as motion detection systems, cameras, and large projectors. In this paper, the collaborative educational benefits and potential drawbacks of Andromeda in a typical K-12 classroom setting are discussed.
3.2. Related Work

3.2.2 WhiteBox Learning

WhiteBox Learning is a software learning system that offers activity-based STEM assignments for 6\textsuperscript{th} to 12\textsuperscript{th} grade science and pre-engineering classes [29]. Through WhiteBox Learning, students can research, design, analyze, simulate, and compete with classmates. Students can construct 3D renderings and print plans to physically construct their designs. This option lets students complete the engineering design process from research to end product.

The WhiteBox Learning assignments used in this case study are the Survival Shelter 2.0 (referred to as shelter) [30] and Dragster 2.0 (referred to as dragster) [31]. An example 3D rendering of a shelter can be found above in fig. 3.1. The shelter assignment enables students to design an emergency shelter for hikers who are trapped in a blizzard. After
students perform initial research on topics like energy and heat, they can design, analyze, and compete with their designs within WhiteBox Learning. The most comfortable shelter, in terms of temperature and space, wins the competition. The dragster assignment enables students to design a $CO_2$ dragster (racecar). Similar to the shelter assignment, students research relevant concepts such as Newton’s Second Law of Motion and net force. Then, they design, analyze, and compete with their designs. The fastest car wins the competition. In this case study, the students’ designs from these assignments were used as input to Andromeda. The class used Andromeda to learn about similarities and trade-offs in their designs.

### 3.3 Methods

While exploring computer science research for personal reasons, the teacher discovered Andromeda and its potential for their 8th grade technology education class. With no outside involvement from the Andromeda team, the teacher decided to use Andromeda alongside WhiteBox Learning.

To prime the students for Andromeda, the teacher introduced Andromeda by exploring the *Animals* dataset [19]. This allowed the students to understand the basic interactions available within Andromeda with an approachable high-dimensional dataset. The class explored Andromeda by answering questions such as “What makes a good pet?” and “What differentiates wild and domestic animals?”. The class grouped together animals they considered similar and increased the weight of variables they suspected would answer the questions. Once the students were familiar enough with Andromeda to understand the software’s capabilities, they then used Andromeda for their WhiteBox Learning assignments.

The class completed the two assignments, shelter and dragster, through WhiteBox Learning. They created and competed their designs using WhiteBox Learning simulations. Using the
3.3. METHODS

1. Analyzing data with Andromeda is fun.
2. I know what is meant by the term high-dimensional data.
3. Understanding math makes using Andromeda easier.
4. I know which interaction (map/observable vs. slider/parametric) to move to better understand the data.
5. Andromeda helped me understand the raw data.
6. Using Andromeda helped me understand how different variables affect my design.
7a. Using Andromeda allowed me to design a better shelter.
7b. Using Andromeda allowed me to design a better dragster.

Figure 3.2: Survey statements given to the students. The students rated each statement on a scale from 0 (strongly disagree) to 10 (strongly agree). Students rated statements 7a and 7b after completing the shelter and dragster assignments, respectively.

engineering design process, the students designed multiple iterations for each assignment. Their intermediate designs were converted to high-dimensional data where each design decision (such as material for the shelter or surface friction for the dragster) represented a single dimension, or variable, of data. The teacher consolidated the students’ designs into a single, high-dimensional dataset to be used in Andromeda.

Using this high-dimensional dataset, the students who opted-in explored and interacted with Andromeda’s visualizations. The teacher controlled the web version of Andromeda and shared their screen with these students. Of the students, about 30% were present on a videoconferencing software. The majority of discussion took place in the classroom environment, yet it was not uncommon for virtual students to add insights via chat. Due to COVID-19 protocols during 2020-2021, the students were not able to physically build their designs.

The teacher gave the students a survey after each assignment as shown in fig. 3.2. The survey
also included free response questions about how the students used the engineering design process, including application of knowledge and skills gained through the use of tools such as Andromeda and WhiteBox Learning. After the teacher collected the survey data throughout the 2020-2021 school year, the Andromeda team became involved in summer 2021 to analyze the data. The team selected not to compare the data to similar, past assignments due to the large number of unique circumstances under which the class was conducted.

3.4 Results

The following heatmaps represent students’ responses to each survey. For all questions, responses ranged from 0 (strongly disagree) to 10 (strongly agree). The surveys were offered to 36 students, but only 30 and 26 students voluntarily completed the shelter and dragster survey, respectively. There were no incomplete responses when the surveys were completed. Most students were enrolled in pre-algebra level math. While they had previous experience in technology education, they had no educational background in high-dimensional data or math related to visualizing high-dimensional data.

3.4.1 Shelter

In fig. 3.3, below, the statements with the highest and lowest average are “Using Andromeda helped me to design a better shelter.” and “Using math makes using Andromeda easier.”, with respective means of $\bar{x} = 7.93$ and $\bar{x} = 4.20$. The statement with the most variability in responses is “Using math makes using Andromeda easier.” ($s = 2.68$). The statement with the least amount of variability is “Analyzing data with Andromeda fun.” ($s = 1.74$). The statement “Using Andromeda allowed me to understand the raw data.” received relatively
3.4. Results

Figure 3.3: Survey responses after the shelter assignment. The mean ($\bar{x}$) and standard deviation ($s$) for each survey question is provided with a visual count of responses.

<table>
<thead>
<tr>
<th>Survey Response</th>
<th>$\bar{x}$ (s)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyzing data with Andromeda is fun.</td>
<td>6.87 (1.74)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>10</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>I know what is meant by the term high-dimensional data.</td>
<td>6.90 (2.16)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>13</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Understanding math makes using Andromeda easier.</td>
<td>4.20 (2.68)</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>I know which interaction (map/observable vs. slider/parametric) to move to better understand the data.</td>
<td>7.80 (1.77)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Andromeda helped me understand the raw data.</td>
<td>7.87 (1.80)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Using Andromeda helped me understand how different variables affect my design.</td>
<td>7.60 (1.83)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Using Andromeda allowed me to design a better shelter.</td>
<td>7.93 (2.03)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

high responses ($\bar{x} = 7.87$), while the statement “I know what is meant by the term high-dimensional data.” received the second-lowest average ($\bar{x} = 6.90$).

3.4.2 Dragster

The results are similar for the dragster assignment as shown in fig. 3.4 below. The statement “Using Andromeda helped me understand the raw data.” has the highest average ($\bar{x} = 8.96$) and the lowest variability ($s = 1.34$). Similar to the shelter assignment, the statement with the lowest average is “Using math makes using Andromeda easier.” ($\bar{x} = 4.15$). This statement also has the highest variability with ($s = 2.75$). The statements “Using Andromeda helped me understand how different variables affect my design.” and “I know which interaction ... to move to better understand the data.” also had high average responses with $\bar{x} = 8.38$ and $\bar{x} = 8.27$, respectively. The impact of the results represented in figs. 3.3 and 3.4 are in the next section of this paper.
CHAPTER 3. A STUDY ON THE COLLABORATIVE UTILITY OF ANDROMEDA IN K-12

Figure 3.4: Survey responses after the dragster assignment. The mean ($\bar{x}$) and standard deviation ($s$) for each survey question is provided with a visual count of responses.

3.5 Discussion

The following discussion puts the results presented in the previous section into context with impressions shared by the teacher. This is broken into four themes: collaboration, data analysis, design influence, and sentiment.

3.5.1 Collaboration: Andromeda supported student collaborations

This case study presents findings on the collaborative utility of Andromeda. The class analyzed and discussed their designs visualized within Andromeda together. According to the students’ teacher, “students didn’t start discussion with other students until Andromeda”. As the students became more comfortable exploring their data within Andromeda, they requested more interactions from their teacher on the shared screen. For example, two
3.5. Discussion

friends had different internal shelter temperatures and they wanted to explore why. They requested that the teacher drag their two points together and update the layout. This used Andromeda’s observation-level interaction as discussed in subsection 3.2.1. After seeing the updated layout and variable weights, the students discussed the most prominent difference in their designs. Example images are shown below in fig. 3.5. Interactions like this demonstrate the potential for Andromeda to serve as a conversation piece as a student group collaboratively explores a dataset.

3.5.2 Data Analysis: Andromeda helped students practice data analysis

Andromeda can be engaging without requiring an extensive math or data analysis background. The students did not fully understand dimensionality reduction before, or after, using Andromeda. Regardless, the class was able to meaningfully interact with Andromeda to improve their designs without being bogged down by Andromeda’s underlying math. These results are supported by the student responses to the statement “Understanding math makes using Andromeda easier.” ($\bar{x}_{\text{shelter}} = 4.20, \bar{x}_{\text{dragster}} = 4.15$). While not understanding the math, the students were still able to gain insights into their high-dimensional data.

Before Andromeda, “students did not know what high-dimensional data was”, according to their teacher. The students began to feel more confident in their understanding of the concept of high-dimensional data once they used Andromeda. This is shown by the positive responses to the statement “I know what is meant by the term high-dimensional data.” ($\bar{x}_{\text{shelter}} = 6.90, \bar{x}_{\text{dragster}} = 7.42$). Students also felt confident in their ability to interact with high-dimensional data by the positive responses to the statement “I know which interaction (map/observable vs. slider/parametric) to move to better understand the data.” ($\bar{x}_{\text{shelter}} =
(a) Select student design 1

(b) Drag student design 1 next to student design 2

(c) The weights and visualization are updated with new weights based on the similarities and differences of students 1 and 2. They both heavily used Newspaper in their shelter designs. Students 18, 19, 21 also had similar designs to 1 and 2.

Figure 3.5: An example of two students comparing and contrasting their shelters. Each point is assigned a unique number and represents a student’s design for the shelter assignment. The exact scaling has been altered for purposes of presentation in this paper.
While introducing concepts like high-dimensional data, Andromeda also allowed students to approach their design analytically rather than anecdotally. Andromeda helped the students better understand their designs by giving them “a new way to analyze the designs”, according to their teacher. Students tested negative hypotheses by analyzing variables they suspected would not impact their designs’ performances. Figure 3.6 shows an example interaction where the class increased the weight on the color variable for their dragsters. This used Andromeda’s parametric interaction as discussed in subsection 3.2.1. As expected, faster dragsters were not plotted near each other. Students were able to verify their negative hypotheses.

Another common interaction the class used was to cluster the top-performing designs together via Andromeda’s observation-level interaction and look at the most heavily weighted variables. These variables represented the most common values the top-performing designs had in common compared to the worse-performing designs. In the example images from fig. 3.7, the top-performing dragsters have similar values for total vehicle mass and surface friction. Interactions like these helped students understand their data and what variables affected their performance. This is supported by the high averages on the following statements: “Using Andromeda helped me understand the raw data.” ($\bar{x}_{\text{shelter}} = 7.87$, $\bar{x}_{\text{dragster}} = 8.96$) and “Using Andromeda helped me understand how different variables affect my design.” ($\bar{x}_{\text{shelter}} = 7.60$, $\bar{x}_{\text{dragster}} = 8.38$). Through using Andromeda, students were able to understand and interact with their high-dimensional data in a new and pedagogically impactful way.
(a) Increase the weight of the color variable (attribute).

(b) Hover over speed so each circle size is proportional to the design’s speed. There is no clear pattern of circle size, so dragster color does not correlate with speed, as expected.

Figure 3.6: An example of the class analyzing if color correlates with dragster speed. Each point is assigned a unique number and represents a student’s design for the dragster assignment. The exact scaling has been altered for purposes of presentation in this paper.
3.5. Discussion

(a) Hover over speed so each circle size is proportional to the design’s speed. The student designs numbered 5, 3, and 15 are fastest.

(b) Drag student designs 5, 3, and 15 together

(c) The weights and visualization are updated based on the similarities and differences of the fastest dragsters. Large rear wheels and round bottom were important variables.

Figure 3.7: An example of the class analyzing the fastest dragsters. Each point is assigned a unique number and represents a student’s design for the dragster assignment. The exact scaling has been altered for purposes of presentation in this paper.
3.5.3 Design Influence: Andromeda influenced the engineering design process

The class explored students’ overall design and individual components within their designs in Andromeda. The class did this together by identifying the top-performing designs, such as the fastest dragster based on WhiteBox Learning simulations, and comparing the designs’ projections in Andromeda. Once the class identified which variables were of importance, students focused their next design iteration on tuning those variables in WhiteBox Learning. Andromeda helped students prioritize their design work and served as an impactful tool in the analysis stage of the engineering design process.

According to the teacher, “students produced higher quality products than in previous years—a direct result of the analysis and rapid prototyping provided by the use of Andromeda and WhiteBox, respectively.” This statement is supported by the student responses to the following statements: “Using Andromeda allowed me to design a better shelter.” ($\bar{x}_{\text{shelter}} = 7.93$) and “Using Andromeda allowed me to design a better dragster.” ($\bar{x}_{\text{dragster}} = 7.77$). Students were able to meaningfully analyze designs in Andromeda and use this analysis to help create better end products.

3.5.4 Sentiment: The class had a positive experience using Andromeda

A class’ attitude toward a new concept, such as data analysis and software, is important in maintaining active engagement. The teacher found that Andromeda has a “low intimidation factor” which is important for its adoption by non-expert data analysts. As a whole, the class enjoyed using Andromeda. Students generally found Andromeda fun and useful. When asked to rate the statement “Analyzing data with Andromeda is fun.”, students responded with averages $\bar{x}_{\text{shelter}} = 6.87$ and $\bar{x}_{\text{dragster}} = 7.62$. Students also responded to the post-survey
saying Andromeda “is going to be good for upcoming projects” and is “kind of cool”. The class was able to use Andromeda meaningfully while learning new concepts and enjoying the learning experience.

3.6 Limitations

The results show that the class benefited from using Andromeda, however, this case study can not definitively show that a class without Andromeda would have performed worse on the assignments. Though the teacher reflected on class projects from previous years without Andromeda, there is no formal control group for this study. While the class agreed that using Andromeda helped them design better products, there is no data to support causation. A large, controlled experiment is necessary to further explore the causal influence of Andromeda on K-12 engineering design projects.

Though students were able to explore high-dimensional data and improve their designs, the projections visualized by Andromeda cannot be considered truly accurate. Web-based Andromeda in its current form requires all variables to be continuous. However, in this case study, the definition of continuous was conflated with categorical. Thus, nominal categorical variables, such as the dragster colors, were assigned numeric values to import data into Andromeda. This means that Andromeda considered, for example, red to have a lesser distance to blue when compared to the distance to green. In reality, these colors should be considered equidistant. This conflation may have affected the class’ visualizations, however, it does not take away from the fact that students were able to explore and understand their high-dimensional data for their engineering design projects. This error highlights potential feature improvements and the responsibility of the creators of Andromeda to communicate the software’s assumptions about its input data.
3.7 Conclusions

This case study highlights Andromeda as an educational tool within engineering education. At the time of this paper, Andromeda does not support the idea of a dependent variable, such as the final speed of the students’ dragsters. For this class, it would have been useful to conveniently display each dragster’s speed since that was what the students were seeking to improve. This could be supported in Andromeda by designating a specific variable to display on each data point. While Andromeda could still take the variable as input to its visualization algorithm, the option to explicitly display a specified variable’s value on each point may help users who are strongly interested in optimizing a single aspect of the data. This would benefit the class in this case study and Andromeda users in general.

Another future improvement of Andromeda could be to support categorical variables through using a different distance metric that supports numerical (continuous) and categorical variables. This would address the incorrect assumption in this study, as discussed in the previous section, and improve Andromeda’s flexibility as a data analysis tool. Regardless, the public Andromeda web page should be updated to specifically define what types of data it supports.

While not used in this case study, Andromeda also supports collaborative analysis that enables multiple users to interact with the same, synchronized visualization. The class would then be able to manipulate the visualization together rather than having the teacher interact with Andromeda based on the class discussion. Exploring the use of this capability in the classroom would help to further illuminate the effects of collaborative data analysis in engineering education.

This case study shows Andromeda’s potential to be used as a collaborative data analysis tool in the classroom. Though this work focused on its use in K-12 classrooms, it is expected that Andromeda would foster collaborations at the college level as well. Educators may
encourage both group and whole-class collaborations with Andromeda. As shown, the software can be used in combination with services such as WhiteBox Learning to teach students about the engineering design process. It can also be taught as a standalone lesson, using the provided sample dataset about animals, to teach introductory data science principles. To encourage educators to introduce Andromeda into their classrooms, an example class activity is provided, with and without solutions, in sections B.1 and B.2 respectively. This activity, like the presented case study, aligns with three guidelines within the proposed data science Standards of Learning for the state of Virginia [32]. These guidelines cover topics such as interpreting data in visualizations, formulating hypotheses from data, and utilizing appropriate tools for data analysis. Collaborative data analysis tools like Andromeda may enhance the K-12 data science education experience.
Chapter 4

A Study on College Students’ Insights using Andromeda

The contents of this chapter are adapted from the author’s work-in-progress publication.

4.1 Introduction

Visualizations are typically evaluated via task completion or insight generation. For task-based evaluations, researchers ask analysts to complete a task. The researchers measure metrics such as analysts’ accuracy, time to complete the task, etc. For insight-based evaluations, researchers ask analysts to generate insights about data. Then, the researchers analyze the participant-generated insights. While asking analysts to complete a task seems like a simpler method of evaluating a visualization, some argue that because visualizations are created to generate insights into data, then they should be evaluated in a similar manner [33, 34]. This chapter describes an analysis of insights from a large-scale study that uses both task-based and insight-based evaluation methods. For this work, an insight is considered any observation or understanding of data.

Within current literature, insights are typically analyzed based on how well they exhibit general characteristics such as their level of depth and relevance [34]. These insight-scoring methods have three problems. First, they require intensive labor for manual annotation.
4.1. Introduction

The data are spread evenly across the horizontal axis. A

There are two small clusters of points. B

Table 4.1: Simplified input of SIEVE that consists of two insights. Each insight has an associated group: A or B.

<table>
<thead>
<tr>
<th>ID</th>
<th>Insight</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The data are spread evenly across the horizontal axis.</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>There are two small clusters of points.</td>
<td>B</td>
</tr>
</tbody>
</table>

Second, they can be subjective. Third, they do not describe the difference in insight language between visualizations. Insights may be equally deep or relevant, but use different language to describe their ideas or observations.

The novel method, Semi-automated Insight-based Explainable Visualization Evaluation (SIEVE), takes a new approach through a semi-automated quantitative analysis. Rather than manually labeling insights, SIEVE applies basic natural language processing techniques and logistic regression modeling to compare groups of insights. Besides any added manual processing steps researchers may want to apply, the process is automated and provides valuable information into how visualizations support insight generation. It identifies the difference in the kinds of insights, based on word usage, that participants identify when using different visualizations. In essence, the method describes the summarative differences between groups of insights. SIEVE does not validate or classify insights, but rather can provide researchers and educators with a better understanding of how visualizations produce different kinds of observations.

SIEVE takes two groups of insights as input. Each insight must be associated with one of two groups. For example, say that there is a visualization study where participants are asked to generate insights given one of two visualizations. Insights belong to group A when generated using visualization A and they belong to group B when generated using visualization B. SIEVE takes these insights, written in natural language, and outputs lists of comparatively distinct terms associated with groups A and B. Two example participant insights are shown in table 4.1.
### Table 4.2: Simplified output of SIEVE.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Term</th>
<th>Beta</th>
<th>Prevalence</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &amp; B</td>
<td>Axis</td>
<td>...</td>
<td>A</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>Cluster</td>
<td>...</td>
<td>B</td>
<td>...</td>
</tr>
</tbody>
</table>

The term *Axis* results in an increase in probability that insights were generated using visualization A, while the term *Cluster* results in an increase for visualization B. This also means that *Cluster* results in a decrease in probability that insights were generated using visualization A. The details of the output (*Beta*, *Significance*) are discussed in ??.

If these insights, along with the insights from other participants, were used as input to SIEVE, the corresponding output may look as shown in table 4.2. The terms *Axis* and *Cluster* have been marked as significant. Without any context into the visualizations available to groups A and B, it may be concluded that it is easier to analyze a single dimension of data with visualization A, while visualization B supports insight generation about how data are similar to each other. With SIEVE, researchers can quickly identify the differences in insight content between visualizations. This can be useful for getting a general idea of the differences in visualizations or to supplement other visualization evaluation techniques.

The remainder of this paper describes related work in section 4.2, defines SIEVE in depth along with the experimental design in section 4.3, and applies SIEVE to a large-scale study on the software Andromeda in section 4.4. Lastly, sections 4.5 and 4.6 contain a discussion of these results and concluding remarks.

### 4.2 Related Work

The software Andromeda and previous educational studies on it can be found in chapter 2.

A common way to evaluate a visualization is through the insights it helps generate. North postulates that “the purpose of visualization is insight. The purpose of visualization evalu-
4.2. RELATED WORK

ation is to determine to what degree visualizations achieve this purpose” [34]. There exist multiple methodologies to perform insight-based visualization evaluation.

The low-level insight characterization by Amar et al. characterizes insights by creating a list of analytic activity components [27]. The researchers in this study asked students to generate questions about visualizations and processed these questions down into a list of ten tasks which describe the analytic activity necessary to answer the questions. A few tasks include retrieve value, filter, sort, cluster, etc. Using these tasks to characterize open-ended insights would prove useful when interested in the specific low-level analysis that users conduct. In this case, the words and concepts used in the insight are more interesting rather than the steps taken to generate the insight.

A popular insight characterization commonly used by visualization researchers is Saraiya et al.’s characterization that measures an insight’s degree of directness, correctness, breadth and depth [11]. North’s similar characterization measures domain value, complexity, depth, subjectivity, unexpectedness, and relevance [34]. Visualization researchers may apply these characterizations by manually assigning characteristic values to each insight and comparing these values within their data analysis. O’Brien et al. adapted Saraiya et al.’s method of insight characterization by counting metrics such as the number of total insights, insights per minute, and insight complexity level [35]. This characterization was applied by an expert with access to recorded video logs of the participants using think aloud and helped researchers to measure the effectiveness of their visualization Gremlin, against a similar system. Gomez et al. also adapted Saraiya et al.’s characterization within their method, Layered Insight- and Task-based Evaluation, that prompts users for insights about the data between predefined search tasks [36]. This method mixes insight and task-based evaluation methods to reflect on the relationship between task performance and how well a visualization promotes insight generation based on Saraiya et al.’s characterization. Lastly, He et al. adapted the same
characterization by analyzing interaction logs and insight quality to identify how analyst actions relate to insight quality [37]. While these works [35, 36, 37] showcase the effectiveness of applying Saraiya et al.’s insight characterization, these methods are dependent on manual insight characterization by an expert and do not describe the difference in the language used in insights between visualizations.

More recent work analyzed insights with different insight characterizations. Guo et al. use visualization interaction logs and insight metrics to understand how application design influences insight generation [10]. Previously existing insight-based evaluation methods did not compare how analysts arrived at insights. The researchers identified correlations between features of interaction such as select, explore, filter, etc. and insight classifications such as fact, generalization, hypothesis, etc. Zgraggen et al. use the same characterization to analyze interaction logs in combination with think-aloud protocols to study the effect of different visualization latency methods on users’ exploratory analysis [38]. Law et al. created a new insight characterization by interviewing professional visualization users. Open-coded interviews with these users uncovered the following seven characteristics of data insights: actionable, collaboratively refined, unexpected, confirmatory, spontaneous, trustworthy, and interconnecting [39]. This characterization has slight overlap with Saraiya et al and North’s [11, 34], however it does uniquely describe how insights must be actionable and refined.

The insight characterizations used within visualization research analyze insights by domain-agnostic values such as depth and complexity. Comparing visualizations by general characteristics like these does not address the idea that different visualization systems may enable different insight generation. These characterization methods were used on Andromeda studies previously [13, 14, 24]. However, it is possible that two visualizations may support equally complex insight generation, but the insights may describe completely separate topics. These general characteristics do not help researchers understand the difference in insight content.
This is addressed with novel insight analysis method SIEVE by identifying the unique language that analysts use in their insights.

4.3 Methods

The following subsections describe the experiment that was conducted to collect student insights and the method SIEVE that was used to analyze the insights.

4.3.1 Experiment Design

The following experiment was designed and conducted by a previous graduate student at Virginia Tech, Lata Kodali. The purpose of the study was to understand how students’ learning changed based on access to different interaction types in Andromeda. The large-scale classroom experiment was conducted in the STAT 2004 introductory statistics course at Virginia Tech in Spring 2017. The course had a lecture portion and an additional “recitation” section which was a 50 minute small group section per week. Recitation sections were on Mondays, Tuesdays, Wednesdays, or Thursdays. For the study, students used the web-based version of Andromeda. A total of 152 students participated.

All students were taught about Weighted Multidimensional Scaling (WMDS) and Andromeda during the lecture portion of the course. This includes a description of basic WMDS concepts. The lesson provides a high-level overview of WMDS. Specifically, the class was taught about how high-dimensional data and the variable weights were used as input to WMDS. Then, WMDS would output, in this case, two-dimensional coordinates. This was explored using an example dataset about cereals that had variables relating to nutrition. Students were able to familiarize themselves with Andromeda.
For a description of Andromeda and each interaction type, see chapter 2. Four versions (one version per recitation) of Andromeda were given to students during recitation. Each student was enrolled in a single recitation section. Data were collected from the students during recitation. The list below describes the different versions of Andromeda.

1. **NONE**: Only has access to surface-level interactions. Essentially, static WMDS.
2. **PI**: Has access to parametric and surface-level interactions.
3. **OLI**: Has access to observation-level and surface-level interactions.
4. **BOTH**: Has access to parametric, observation-level, and surface-level interactions.

These four versions of Andromeda were randomly assigned to entire recitation groups, with Monday using **PI** (42 students), Tuesday using **NONE** (40 students), Wednesday using **OLI** (40 students), and Thursday using **BOTH** (30 students).

Students completed surveys throughout recitation where they optionally consented to having their submission data collected for this study. The data were collected with approval under IRB #21-911. An example survey is shown in chapter A. Within the survey, students were asked questions that required them to interact with the animals dataset [19] using Andromeda. Some questions were graded on accuracy such as “In what ways are the **Bobcat**, **Hippopotamus**, and **Beaver** similar to each other, but also different from a **Horse**? Briefly explain why”. This work focuses solely on the responses to “Write three insights that you notice about the data”. Students were asked to write down three insights before and after using their assigned version of Andromeda. Thus, the insights can be grouped by version of Andromeda and by whether the responses were generated at the beginning or end of the recitation session.

The insights were previously analyzed by Lata Kodali by scoring the complexity of the insights. This method was adopted from previous works on Andromeda [13, 24]. The analysis
relating to insight complexity and student responses to the other survey questions are not reported in this work. The work in this chapter analyzes the insights using SIEVE. These results will be included within that overall study in the future.

The following text cleaning and processing steps are applied in context of this study, but can be altered for other datasets and experimental setups as appropriate.

1. **Combine three insights.** Concatenate each set of three insights generated by students into a single response.

2. **Remove stop words.** Remove unimportant words such as “the”, “and”, etc. These words are not important for analysis. The NLTK\(^1\) pre-made stop word list was used.

3. **Apply lemmatization.** Lemmatization is the natural language process of grouping forms of a word into a single word. For example, “changing” and “changed” are changed to their base form “change”. Lemmatization was done with the NLTK\(^1\) lemmatizer and manual lemmatization was done for any word forms that the NLTK lemmatizer missed.

4. **Collapse data point and variable names into single keywords.** For the animal dataset, convert all instances of data point names (such as Giraffe and Dog) into the keyword Animal_{Tag}. Convert all instances of variable names (such as Smelly and Size) into the keyword Variable_{Tag}. This is done because it is more important if a group of insights describes rows or columns more rather than if a specific data point or variable is mentioned more.

5. **Collapse feature, attribute, variable into a single keyword.** For the purposes of this study, these words have the same meaning and are used interchangeably by students. Similarly to the step above, these words are converted into variable.

\(^1\)Natural Language Toolkit: www.nltk.org
4.4 Results

The following describes how SIEVE is used when analyzing student insights from using different versions of the software tool Andromeda. Note that the term Animal means that an insight directly used the word “Animal”, while the term Animal_Tag means that an insight used a specific animal name such as “Giraffe”. This rule similarly applies to Variable and Variable_Tag. SIEVE is applied to compare Andromeda interactions pairwise based on word counts. Tables 4.3 to 4.5 list the models fit and notable model results. This section highlights the results.

4.4.1 Keyword Differences Between Andromeda Versions

The keyword differences reported in table 4.3 are based on student insights at the end of their recitation session. This means that the students had time to familiarize themselves with their version of Andromeda before generating these insights.

The first three comparisons in the table report the results regarding OLI compared to the other versions of Andromeda, where response $y_i=1$ when insights derive from OLI and 0 otherwise. The most keyword differences that are deemed statistically significant result when modeling OLI and PI. The probability that insights were generated using OLI increases significantly with the use of terms Similar, Far, and Away and decreases significantly with the use of terms Weight, Variable, Variable_Tag, Increase, and One. Insights generated with OLI have similar, but weaker differences with BOTH. When comparing OLI- and BOTH-supported insights, the term Away increases the probability that insights were generated using OLI, while the terms Weight, Variable, High, and Adjust increase the probability that insights were generated using BOTH. Generally, an increase in spatial-words (such as Far and Away) results in an increase in probability that insights were generated using OLI.
4.4. Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OLI vs PI</td>
<td>Similar</td>
<td>0.521</td>
<td>0.196</td>
<td>OLI</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Away</td>
<td>0.492</td>
<td>0.253</td>
<td>OLI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>0.609</td>
<td>0.318</td>
<td>OLI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>-0.725</td>
<td>0.258</td>
<td>PI</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Variable</td>
<td>-0.619</td>
<td>0.226</td>
<td>PI</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Variable_Tag</td>
<td>-0.576</td>
<td>0.209</td>
<td>PI</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Increase</td>
<td>-0.738</td>
<td>0.371</td>
<td>PI</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>One</td>
<td>-0.420</td>
<td>0.252</td>
<td>PI</td>
<td>*</td>
</tr>
<tr>
<td>OLI vs BOTH</td>
<td>Away</td>
<td>0.632</td>
<td>0.352</td>
<td>OLI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>-0.426</td>
<td>0.215</td>
<td>BOTH</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Variable</td>
<td>-0.461</td>
<td>0.212</td>
<td>BOTH</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>-0.662</td>
<td>0.362</td>
<td>BOTH</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Adjust</td>
<td>-0.458</td>
<td>0.265</td>
<td>BOTH</td>
<td>*</td>
</tr>
<tr>
<td>OLI vs NONE</td>
<td>Group</td>
<td>0.519</td>
<td>0.292</td>
<td>OLI</td>
<td>*</td>
</tr>
<tr>
<td>PI vs BOTH</td>
<td>Animal</td>
<td>0.387</td>
<td>0.216</td>
<td>PI</td>
<td>*</td>
</tr>
<tr>
<td>PI vs NONE</td>
<td>Weight</td>
<td>1.227</td>
<td>0.391</td>
<td>PI</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Variable</td>
<td>0.460</td>
<td>0.219</td>
<td>PI</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>One</td>
<td>1.179</td>
<td>0.605</td>
<td>PI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Increase</td>
<td>0.533</td>
<td>0.310</td>
<td>PI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Animal</td>
<td>0.333</td>
<td>0.200</td>
<td>PI</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Animal_Tag</td>
<td>-0.372</td>
<td>0.209</td>
<td>NONE</td>
<td>**</td>
</tr>
<tr>
<td>BOTH vs NONE</td>
<td>Weight</td>
<td>0.836</td>
<td>0.287</td>
<td>BOTH</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Animal_Tag</td>
<td>-0.382</td>
<td>0.215</td>
<td>NONE</td>
<td>**</td>
</tr>
</tbody>
</table>

Table 4.3: Significant keyword differences between insights generated with versions of Andromeda. Each comparison compares two groups of insights, such as OLI and PI in the first row. Similar is a significant term when comparing OLI and PI insights. An insight is more likely to belong to (Prob. ↑) the OLI insights if it uses the term Similar more. A single asterisk (*) indicates that $p < 0.1$, a double asterisk (**) indicates that $p < 0.05$, and a triple asterisk (***) indicates that $p < 0.01$. See ?? for an explanation of how significance is determined.
Table 4.4: Keywords associated with an increase in probability that an insight belongs to a version of Andromeda based on table 4.3. An insight is more likely to be generated using OLI if it uses the words Similar, Away, Far, and/or Group more. Note that Weight and Variable are associated with PI and BOTH.

while value-words (such as Increase and Weight) results in an increase in probability that insights were generated with PI or BOTH.

These “value-words” are also identified as keyword differences in the comparison between PI and NONE-supported insights. Specifically, the terms Weight, Variable, One, Increase, and Animal result in an increase in probability that insights were generated with PI. Only Animal_Tag results in an increase in probability that insights were generated with NONE.

The remaining pairwise comparisons generated one or two keyword difference between the groups of insights. These identified keywords are directly related to the difference in functionalities offered by the version of Andromeda. In the comparison between PI- and BOTH-supported insights, the term Animal results in an increase in probability that insights were generated with PI. The results are the same when comparing PI and NONE. Lastly, when comparing NONE- and BOTH-supported insights, Weight is associated with an increase in probability that insights were generated with BOTH while Animal_Tag is associated with an increase in probability that insights were generated with NONE. The results are further discussed in section 4.5.
### 4.4. Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NONE</td>
<td>Even</td>
<td>0.626</td>
<td>0.307</td>
<td>Pre</td>
<td>**</td>
</tr>
<tr>
<td>Pre &amp; Post</td>
<td>Much</td>
<td>-0.590</td>
<td>0.353</td>
<td>Post</td>
<td>*</td>
</tr>
<tr>
<td>PI</td>
<td>Mammal</td>
<td>0.441</td>
<td>0.228</td>
<td>Pre</td>
<td>*</td>
</tr>
<tr>
<td>Pre &amp; Post</td>
<td>Far</td>
<td>0.503</td>
<td>0.269</td>
<td>Pre</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Would</td>
<td>0.563</td>
<td>0.309</td>
<td>Pre</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Variable_Tag</td>
<td>-1.188</td>
<td>0.259</td>
<td>Post</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Variable</td>
<td>-0.919</td>
<td>0.265</td>
<td>Post</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>-0.484</td>
<td>0.209</td>
<td>Post</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Increase</td>
<td>-0.423</td>
<td>0.242</td>
<td>Post</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Lot</td>
<td>-0.430</td>
<td>0.253</td>
<td>Post</td>
<td>*</td>
</tr>
<tr>
<td>OLI</td>
<td>Variable</td>
<td>-0.710</td>
<td>0.279</td>
<td>Post</td>
<td>**</td>
</tr>
<tr>
<td>Pre &amp; Post</td>
<td>Weight</td>
<td>-1.182</td>
<td>0.546</td>
<td>Post</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Variable_Tag</td>
<td>-0.382</td>
<td>0.216</td>
<td>Post</td>
<td>*</td>
</tr>
<tr>
<td>BOTH</td>
<td>Variable_Tag</td>
<td>-0.862</td>
<td>0.294</td>
<td>Post</td>
<td>***</td>
</tr>
<tr>
<td>Pre &amp; Post</td>
<td>Variable</td>
<td>-0.639</td>
<td>0.248</td>
<td>Post</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>-0.599</td>
<td>0.253</td>
<td>Post</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>-0.723</td>
<td>0.393</td>
<td>Post</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 4.5: Significant keyword differences between insights generated before and after recitation with a version of Andromeda. *Even* is a significant term when comparing *NONE* insights. An insight is more likely to belong to (Prob. ↑) the Pre-recitation insights if it uses the term *Even* more. A single asterisk (*) indicates that $p < 0.1$, a double asterisk (**) indicates that $p < 0.05$, and a triple asterisk (***) indicates that $p < 0.01$. See ?? for an explanation of how significance is determined.
4.4.2 Keyword Differences Before & After Using Andromeda

Students were asked to generate insights at the beginning and end of their recitation session. Throughout the recitation, students completed tasks within their version of Andromeda. The keyword differences between their insights, by version, are reported in table 4.5.

Insights generated using PI, once again, had the most keywords identified. Mammal, Far, and Would increased the probability that insights were generated at the beginning while Variable_Tag, Variable, Weight, Increase, and Lot increased the probability that insights were generated at the end. For NONE-supported insights, the term Even decreases the probability that insights were generated at the end of recitation while Much increases it. Lastly, for OLI- and BOTH-Andromeda, the terms Variable_Tag, Variable, and Weight resulted in an increase in the probability that insights were generated at the end of recitation. BOTH also identified Change as having this result. Similar to the results in table 4.3, the terms Variable and Weight are often identified as significant keywords. For insights generated using PI, OLI, and BOTH, using these terms results in an increased probability that these insights were generated at the end of recitation. The implications of these results are discussed in section 4.5.

4.5 Discussion

The following discussion splits into three parts: on the results comparing insights generated between versions of Andromeda, on the results comparing insights generated before and after using a version of Andromeda, and on using SIEVE in research.
4.5. Discussion

4.5.1 Insight Differences Between Andromeda Versions

The significant keyword differences between insights generated with different versions of Andromeda are found in table 4.3. The insight differences show no significant keywords, except for Animal_Tag, when comparing NONE-supported insights to other versions of Andromeda. These data suggest that NONE insights do not include unique, clear common words relative to other interactions. This confirms the belief that Andromeda and its interactions, compared to standard, static WMDS, enable analysts to generate richer insights.

Going into this study, it was suspected that PI- and OLI-supported insights would have the most significant keyword differences. The expectation was confirmed by the data. Furthermore, PI-supported insights are consistently associated with value-words (such as Variable and Weight) and OLI-supported insights are associated with spatial-words (such as Similar and Away). This makes sense because value-words overlap with concepts of WMDS and PI requires users to focus on input parameters. Also, spatial-words pertain to the interpretation of the WMDS projection, which is essential for OLI.

It was also expected that animal names, which were converted to Animal_Tag, would be associated with OLI-supported insights. Instead, SIEVE reports that Animal_Tag is associated with NONE. This is likely because analysts with NONE Andromeda do not describe interactions and relationships as much, so they rely on simply describing animals. Similar results were expected for Variable_Tag and PI-supported insights, which was supported by Variable_Tag resulting in an increase in probability that insights were generated with PI when comparing OLI and PI.

When comparing OLI- and BOTH-supported insights, the words Weight and Variable increase the probability that insights were generated using BOTH. It is suspected that because these words are also significant to PI and PI is available in BOTH Andromeda, analysts using BOTH
were generating insights using PI. This overlap of words is shown in table 4.4. Generally, the analysts with access to PI are using WMDS-words, or value-words, and ideas to describe their insights. Analysts with access to OLI are comparatively not focusing on the mechanics of WMDS, but its interpretation. This idea is supported by the fact that OLI- and NONE-supported insights do not have any significant differences in value-words. If analysts were properly supported by OLI, WMDS-words would be expected to be associated with OLI-supported insights more than NONE-supported insights.

The keyword insight differences between Andromeda versions provide educators using Andromeda motivation to understand how and why students choose between PI and OLI when generating insights. If the initial educational goal was to enable students to interpret relationships within the data, it could be said that the analysts with access to OLI performed better. On the other hand, if the goal was to teach students to master WMDS concepts, it could be said that PI worked better in supporting students. Regardless, there is a clear difference between the insights generated by using different versions of Andromeda.

### 4.5.2 Insight Differences Before & After Using Andromeda

The students were asked to generate insights at the beginning and end of their recitation session. Table 4.5 compares the difference between the insights by Andromeda version. Generally, the words Variable_Tag, Variable, and Weight are associated with an increased in probability that insights were generated at the end of recitation when given PI, OLI, and BOTH Andromeda. This shows that even though OLI-supported insights have a weaker association with WMDS-words than PI- and BOTH-supported insights, as discussed in subsection 4.5.1, analysts with access to OLI are still increasingly using these words. The magnitude of the $\beta_1$ coefficients show that OLI-supported insights have the greatest change in keyword influence.
on insight classification. This suggests that analysts using PI and BOTH found it easier to use these words at the beginning of the session. This supports the idea that analysts rely on WMDS concepts more (as described by WMDS-words) when given access to PI compared to OLI.

For PI-supported insights, **Increase** and **Lot** result in an increase in probability that insights were generated at the end of recitation while the words **Mammal**, used to refer to a group of animals with common characteristics, and **Far** result in a decrease. This suggests that after using PI, analysts used these spatial-words less. For **NONE**-supported insights, **Much** results in an increase that insights were generated the end of recitation while **Even** results in a decrease. Unlike the other results of SIEVE, the contexts of these words are not understood immediately. Further exploration shows that most insights using **Even** are using it with the phrase “even though” as a way to contradict expectations in an insight. The word **Much** is used as a comparison word in phrases like “much more”. These findings suggest that after using **NONE**, insights are more likely to use contradictory words at the beginning of recitation and comparative words at the end. This confirms that even with just **NONE**, there is a difference in insight content at the beginning and end of recitation.

### 4.5.3 Research with SIEVE

Before creating SIEVE, explainable classifier models to classify what visualization group an insight belongs to were explored. It was found that existing classification models did not achieve the initial goal: to understand how groups of insights are different between visualizations. For example, Decision Tree-based classifiers provide explainable output that reports what words contribute most to an insight’s classification, however performance of these models were poor and inconsistent. This led to the idea of conducting statistical
significance testing based on the presence of words in insights as described in ??.

SIEVE fits logistic regression models to model the relationship between word counts ($n$ discrete counts) and group assignments (a binary value). The logistic regression model is simple, fast, and reports confidence. Fitting the models this way outputs whether a specific keyword is associated with an increase or decrease of probability that an insight belongs to a certain group. This is appropriate in the context of this study because the keywords provide the summarative difference between insight groups.

There exist potential avenues for future work with SIEVE. This is the first insight-based visualization evaluation method that analyzes insights using natural language processing. Quantitatively analyzing natural language is faster than manual annotation and, more importantly, less subjective to researcher bias. The description of insight differences provided by SIEVE can give researchers a general understanding of how different visualizations enable insight generation. This can give researchers ideas for how to appropriately apply more in-depth insight-based visualization evaluation methods.

SIEVE currently only analyzes insights on a word-by-word basis. Extending the method to look at phrases, rather than individual words, may yield interesting results. In this case, a phrase-based approach may better capture the idea of groups of points such as “aquatic animals” or “physical traits”. Within education research, SIEVE may be helpful to develop an automatic grading scheme of natural language insights, however SIEVE in its current state does not support the notion of insight accuracy. Similar to SIEVE, an automatic grading regression model could be trained to output a grade based on the counts of words present in a group of insights. Within visualization research, SIEVE could be extended to classify insights automatically. A pre-trained insight classifier may be transferable between studies. Other future work can also combine SIEVE with different insight-based evaluation methods to gain an understanding of how groups of insights are different and how that relates to the
4.6 Conclusions

SIEVE identifies keyword differences between groups of insights. This method was applied in the context of a comparative visualization study of the software Andromeda. With minimal text processing, it was found that the parametric interaction within Andromeda enabled students to use value- and WMDS-related words in their insights that describe the variable values of data points and how they change. Observation-level interaction, on the other hand, enabled students to use spatial-words more that describe the relationships between data and relative positions of data points. SIEVE describes how the interactions within Andromeda support different types of insight generation.

There are limitations to SIEVE. It does not explain whether groups of insights differ in quality, depth, accuracy, etc. While other insight-based evaluation methods measure these characteristics, the measurement is done by manually annotating insights. SIEVE does not have as in-depth output, however it does provide an analysis on keyword insight differences which is not offered by other methods. The ideal method to use in a visualization evaluation study depends on the context of the study.

Overall, the novel method SIEVE describes how visualization systems support analyst insight generation through semi-automatically analyzing natural language insights. SIEVE is simple and fast to apply. This method was applied on the software system Andromeda to understand how different interactions in the system enable analysts to generate different types of insights. SIEVE can be applied and improved to further visualization research that seeks to understand natural language insights.


Chapter 5

Conclusion

This thesis explores new research on Andromeda in education. This section contains thoughts on potential future work as well as conclusions of the work presented.

5.1 Overall Conclusions

The research questions and contributions within this thesis are listed in sections 1.2 and 1.3. The first research question asks how Andromeda can be used to facilitate collaborative data analysis and engineering design in education. This is answered by the case study in chapter 3. Middle school students used Andromeda in a collaborative setting and were able to meaningfully conduct data analysis on their engineering designs by comparing and contrasting designs, testing hypotheses, and discussing how to improve their design prototypes. The second research question asks how student-generated insights relate to the interaction types in Andromeda the students had exposure to in an introductory data analytics classroom setting. This question is addressed in chapter 4 by conducting an analysis on student insights and finding that the language used in students’ insights strongly related to the version of Andromeda used by the students. It was also found that, when given access to multiple interaction types, students may not be utilizing all interaction types equally. The final research question asks how to conduct the insight analysis just described. This was conducted by a novel insight analysis method, SIEVE, which provides a summative difference be-
5.2 Future Work

5.2.1 Andromeda as a Supplemental Lesson

Future work may identify the academic areas that Andromeda may best support data analysis. Do students find it easier to meaningfully use Andromeda when learning about data science as a primary lesson or when using Andromeda as a supplemental tool? While this thesis describes examples of Andromeda being used a primary and supplemental tool in education, the uses are not directly comparable considering the large difference in educational level, content, and setting. This interesting research question can also be generalized to teaching data science principles. It is increasingly relevant as educators, at least within the state of Virginia, begin to introduce data science principles into the classroom. Educators using Andromeda can apply Andromeda in a way that best aligns with their educational motivations and industry-defined (SOL) [32] standards. Understanding the difference in data science education when teaching data science principles as a primary versus supplemental lesson provides relevant insight into the student learning process.

5.2.2 Andromeda as a Public Tool

This thesis discusses Andromeda serving as an educational tool within K-12 and college. The public, web-based version available can be further supported. As mentioned, a disclaimer must be added to the site to support users in making accurate assumptions and conclusions about their data. It is the responsibility of researchers and developers to clearly define the
capabilities of their public tools. The disclaimer must also be written in a way that a non-technical audience can understand considering that Andromeda is designed to be used by novice and expert analysts alike.

A great benefit of publishing Andromeda as a public tool is that it spurs authentic use. As described in section 3.3, the middle school teacher used Andromeda in his classroom with no outside involvement from the Andromeda team. This is an example of data with high ecological validity. Ecological validity is defined as “a measure of how test performance predicts behaviours in real-world settings” [40]. The evaluation environment closely matched the real world context. In this case, the evaluation environment is the middle school teacher using Andromeda in the classroom with no researcher involvement. The real world context is a group of novice analysts using Andromeda to conduct data analysis. This had positive and negative implications. The tool was slightly misused, as described in section 3.5, however it provided a unique opportunity to analyze how Andromeda was used with higher ecological validity than ever before. This authentic data collection gave unique feedback on Andromeda which led multiple ideas for features and research as described next.

5.2.3 Andromeda Improvements

Collaboration is a largely unexplored area for Andromeda. The case study, as presented in chapter 3, analyzed how a class shared a single Andromeda display and meaningfully interacted with their engineering design data. The class did not use Andromeda’s built-in collaboration feature. The class was not aware of this feature. Thus, the accessibility of collaborative Andromeda should be improved and further research should be conducted to better understand how analysts collaborate. Similarly to [7], future work could compare how novice analysts collaborate when given a single, shared display versus a distributed, shared
5.2. Future Work

display by identifying collaboration patterns.

In its current state, Andromeda does not support the concept of dependent variables. For example, students in the case study were specifically focused on maximizing dragster speed. To better support these students, Andromeda may offer a functionality to display a chosen variable’s value on the data points directly. This brings up the idea of supporting causation within Andromeda. It is possible that adding this functionality may encourage analysts to think in a single-dimension which could hinder high-dimensional thinking. The analysts could become fixated on increasing or decreasing dependent variable(s). Future research could explore the concept of causation in exploratory data analysis.

As described in the discussions in section 4.5, students using Andromeda with all interaction types available are using keywords more strongly associated with parametric interaction rather than using keywords associated with observation-level interaction when generating insights. There are a few potential reasons for this which may be explored in a future study. It could either be because of the Andromeda interface itself where parametric interaction is more apparent at first glance. It could also be because students find the concept supporting parametric interaction (WMDS) is easier to understand than the concept supporting observation-level interaction (inverse WMDS). A small-scale, interview-based qualitative study may provide students an opportunity to explain the reasoning for gravitating towards certain interaction types if they prefer one over the other. This could lead to an improvement in the Andromeda user interface and/or in the instructional material given to students about using Andromeda.
5.2.4 Andromeda-supported Insights

There still exist many ways to analyze Andromeda by the insights it helps generate. For example, future research may identify how outside knowledge or continued exposure to Andromeda influences insight generation. A future study could extend the work on SIEVE by conducting an in-depth study on Andromeda interactions and insight generation via interaction logs and SIEVE results similarly to the work by He et. al [37].

As described in chapter 4, insight-based visualization evaluation by natural language processing has not been done before, thus SIEVE provides multiple opportunities for future research. Future work could extend the method by improving the natural language processing (NLP) methods used as NLP research moves forward. SIEVE focuses on keyword difference identification, but applying other NLP tasks such as insight sentiment classification or insight quality prediction may yield interesting results. The insights that analysts generate when conducting exploratory data analysis provide rich data that can be explained through NLP methods. SIEVE provides just a single exploration of this.

The insight analysis from SIEVE is part of an overall larger study as discussed in subsection 1.1.2. The next steps for this work are to use the SIEVE results in combination with the insight complexity analysis to gain a fuller understanding of how the students’ insights relate to their access to different interaction types in Andromeda. This research in progress.

5.3 Impact

As shown, there are remaining research opportunities on the use of Andromeda in education. This work is expected to become increasingly relevant as data science principles are more heavily emphasized in early education. Andromeda can introduce students of all ages to the
idea of high-dimensional, data-driven thinking in an engaging and approachable manner. We can further our understanding of how students learn with Andromeda through collaborative exploratory data analysis and visualization-supported insight generation.
Bibliography


Appendices
Appendix A

Recitation Survey Given to College Students

The following questions were given to college students as part of recitation. Each student used a different version of Andromeda. For experimental setup information, see subsection 4.3.1. Formatting and text has been changed for presentation purposes. The differences as negligible.

1. Please enter the following information.
   
   (a) Your name.
   
   (b) Recitation.
   
   (c) Today’s date.

2. In your visualization, are the weights equal across all variables?

3. In your visualization, name an animal that is most similar to a cow, relative to all other animals.

4. For question 3, why did you choose this animal?

5. In your visualization, name an animal that is very different from the cow relative to others.
6. For question 5, why did you choose this animal?

7. Write three insights that you notice about the data.

Notice the Bobcat, Hippopotamus, and Beaver are in different locations on the graph. This suggests that they are very different from each other, relative to others, when considering all of the variables equally. However, they have similarities when considering a subset of those variables.

Keep this in mind when answering the following questions. Please go on to question 8.

8. In what ways are the Bobcat, Hippopotamus, and Beaver similar to each other, but also different from a Horse? Briefly explain why.

9. In this visualization, you can assess observations AND variables. Compare animals using the smart variable.

10. Compare animals using the smart, fierce, and smelly variables.

11. If you can/could adjust variable weights, might the observation locations could change as well?

12. If you can/could adjust observation locations, might the variable weights could change as well?

13. Write three insights about the data before you leave.
Appendix B

Andromeda Activity Worksheet for K-12 Students

B.1 Worksheet with Solutions

To complete the activity, go to http://nebula.cs.vt.edu/cosmos/andromeda.html. On the top right, click the following: “Start Here”, “Load Data”, “highD/Animal_Data_square.csv”, and “Use Selected Data”. This loads a dataset describing animals where the minimum value is 0 and the maximum value is 100. Answer the following questions using Andromeda.

Solutions in blue. For questions 6-8, a few possible solutions are listed.

1. [ ] 0 100
   Approximately how fierce is a Tiger?

2. [ ] 0 100
   Approximately how timid is a Wolf?

3. [ ] 0 100
   Approximately how big is a Deer?

4. If you increase the weight on Swims, will the Dolphin and Otter move closer together or farther apart? Circle one.
   
   a. Closer Together
b. Farther Apart

5. If you increase the weight on Size, will the Squirrel and Giant Panda move closer together or farther apart? Circle one.
   a. Closer Together
   b. Farther Apart

6. What differentiates a Mouse, Raccoon, and Squirrel from a Blue Whale and a Dolphin? Fill in the 3 blanks.
   swims, furry, claws

   timid, grazer, size

8. What differentiates a good pet from a bed pet? Fill in the 3 blanks.
   good pets are highly domestic, good pets are agile, bad pets are grazers
B.2 Worksheet without Solutions

To complete the activity, go to http://nebula.cs.vt.edu/cosmos/andromeda.html. On the top right, click the following: “Start Here”, “Load Data”, “highD/Animal_Data_square.csv”, and “Use Selected Data”. This loads a dataset describing animals where the minimum value is 0 and the maximum value is 100. Answer the following questions using Andromeda.

1. 0  Approximately how fierce is a Tiger?  100

2. 0  Approximately how timid is a Wolf?  100

3. 0  Approximately how big is a Deer?  100

4. If you increase the weight on Swims, will the Dolphin and Otter move closer together or farther apart? Circle one.
   a. Closer Together
   b. Farther Apart

5. If you increase the weight on Size, will the Squirrel and Giant Panda move closer together or farther apart? Circle one.
   a. Closer Together
   b. Farther Apart

6. What differentiates a Mouse, Raccoon, and Squirrel from a Blue Whale and a Dolphin? Fill in the 3 blanks.

_________________________  _________________________  _______________________

8. What differentiates a good pet from a bed pet? Fill in the 3 blanks.

_________________________  _________________________  _______________________
