STATISTICAL ANALYSIS, DATA VISUALIZATION AND BUSINESS INTELLIGENCE TOOLS FOR ELECTRONIC RESOURCES IN ACADEMIC LIBRARIES

CHENG CHENG, ACQUISITIONS & ELECTRONIC RESOURCES LIBRARIAN, SUNY ONEONTA
TRACY GILMORE, COLLECTIONS ASSESSMENT LIBRARIAN, VIRGINIA TECH
COLLEEN LOUGEN, ELECTRONIC RESOURCES LIBRARIAN, SUNY NEW PALTZ
CONNIE STOVALL, ASSISTANT DIRECTOR FOR STRATEGIC RESEARCH & ANALYSIS, VIRGINIA TECH

Introduction

Why ???

Statistical analysis is vital for efficient assets management, and user needs satisfaction

- It help us understand user demand, predict future trends, and optimize resources

Statistical analysis is one of the most solid ways to support decision making

- Data doesn't lie

INTRODUCTION

Business Intelligence Tools

- How stats and data are analyzed and visualized

We are going to talk about:

- RedLink
- Excel
- Tableau
- Oneonta Project

INTRODUCTION

- Data visualization is the method to make stats understandable for human
- Administration and stakeholder need something understandable

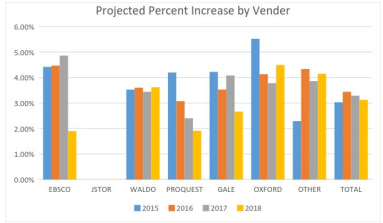
Introduction

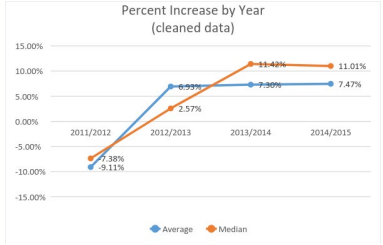
Before:

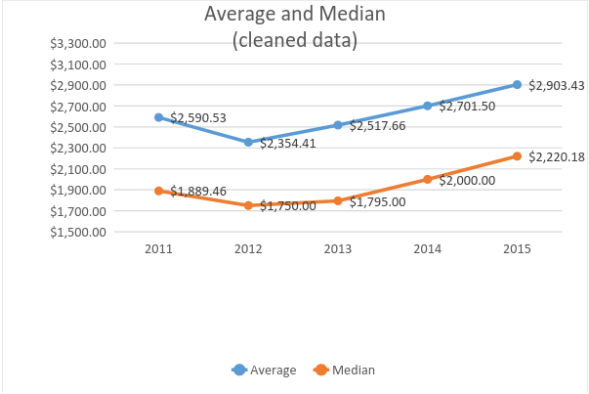
	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	2011	2012	2013	2014	2015	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	Overall Increase	Vender	Per Year Increase	2014 Usgae	2015 Usage	% Total Usage 2014	% Total Usage 2015	% Of Collection
2	EBSCO																	
3	$ 1,300.00	$ 1,365.00	$ 1,433.00	$ 1,505.00	$ 1,580.00	0.00%	5.00%	4.98%	5.02%	4.98%	21.54%	EBSCO	4.31%	1,670	5,286	0.73%	1.98%	1.20%
4	2,084.00	2,188.00	2,297.00	2,412.00	2,659.00	4.99%	4.99%	4.98%	5.01%	10.24%	21.51%	EBSCO	4.30%	1,475	1,717	0.64%	0.64%	1.20%
5	4,800.00	4,922.00	5,242.00	5,504.00	5,779.00	49.42%	2.54%	6.50%	5.00%	5.00%	79.90%	EBSCO	15.98%	2,120	1,810	0.92%	0.68%	1.20%
6	2,968.00	3,116.40	3,272.22	3,435.46	3,607.23		5.00%	5.00%	4.99%	5.00%	21.54%	EBSCO	5.38%	1,521	1,992	0.66%	0.75%	1.20%
7	2,099.20	2,001.18	2,100.45	2,217.49	2,330.35	16.30%	-4.67%	4.96%	5.57%	5.09%	29.11%	EBSCO	5.82%	3,393	2,529	1.48%	0.95%	1.20%
8	2,895.00	2,895.00	3,040.00	3,192.00	3,352.00	0.00%	0.00%	5.01%	5.00%	5.01%	15.79%	EBSCO	3.16%		3,135	0.00%	1.18%	1.20%
9	5,900.00	6,136.00	6,443.00	6,765.00	7,103.00	300.00%	4.00%	5.00%	5.00%	5.00%	381.56%	EBSCO	76.31%	3,393	3,410	1.48%	1.28%	1.20%
10	473.00	473.00	473.00	473.00	473.00	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	EBSCO	0.00%	1,019	5,286	0.44%	1.98%	1.20%
11	995.00	995.00	995.00	995.00	995.00		0.00%	0.00%	0.00%	0.00%	0.00%	EBSCO	0.00%	100,266	111,299	43.69%	41.73%	1.20%
12	**9**	**9**	**9**	**9**	**9**							**TOTAL**		**114,857**	**136,464**	**50.05%**	**51.16%**	**9.64%**
13	**$ 2,612.69**	**$ 2,676.84**	**$ 2,810.63**	**$ 2,944.33**	**$ 3,097.62**		**2.46%**	**5.00%**	**4.76%**	**5.21%**	**64.95%**	**COUNT**	16.24%	**8**	**9**			
14	**$ 2,099.20**	**$ 2,188.00**	**$ 2,297.00**	**$ 2,412.00**	**$ 2,659.00**	**16.30%**	**4.23%**	**4.98%**	**5.01%**	**10.24%**	**47.31%**	**AVERAGE**		**14,357.13**	**15,162.67**			
15	**$ 23,514.20**	**$ 24,091.58**	**$ 25,295.67**	**$ 26,498.95**	**$ 27,878.58**							**MEDIAN**		**1,895**	**3,135**			
16																		
17	GALE																	
18		$ 299.92	$ 313.41	$ 329.08	$ 342.00			4.50%	5.00%	3.93%	14.03%	GALE	4.68%	68	334	0.03%	0.13%	1.20%
19		313.50	327.66	327.66	357.20			4.52%	0.00%	9.02%	13.94%	GALE	4.65%	329	248	0.14%	0.09%	1.20%
20		2,798.32	2,924.74	3,071.04	3,071.04			4.52%	5.00%	0.00%	9.75%	GALE	3.25%	322	243	0.14%	0.09%	1.20%
21	875.00	901.25	946.31	993.63			3.00%	5.00%	5.00%		13.56%	GALE	4.52%	776	559	0.34%	0.21%	1.20%
22		299.92	313.41	329.08	342.00			4.50%	5.00%	3.93%	14.03%	GALE	4.68%	364	330	0.16%	0.12%	1.20%
23	**1**	**5**	**5**	**5**	**4**							**TOTAL**		**1,859**	**1,714**	**0.81%**	**0.64%**	**6.02%**
24	**$ 875.00**	**$ 922.58**	**$ 965.11**	**$ 1,010.10**	**$ 1,028.06**		**5.44%**	**4.61%**	**4.66%**	**1.78%**	**11.43%**	**COUNT**	2.86%	**5**	**5**			
25	**$ 875.00**	**$ 313.50**	**$ 327.66**	**$ 329.08**	**$ 349.60**		**-64.17%**	**4.52%**	**0.43%**	**6.24%**	**11.52%**	**AVERAGE**		**371.80**	**342.80**			
26	**$ 875.00**	**$ 4,612.91**	**$ 4,825.53**	**$ 5,050.49**	**$ 4,112.24**							**MEDIAN**		**329.00**	**330.00**			
27																		

INTRODUCTION

- After:







DISCUSSION OF REDLINK

COLLEEN LOUGEN

SUNY NEW PALTZ

RedLink

- Analytics company founded in 2013
- Provides products to libraries and publishers
- Includes the subscription-based Library Dashboard and the free Redlink network service

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

- Library Dashboard aggregates and standardizes usage data from COUNTER reports and non-COUNTER
- Instantly view and share usage and cost data among staff
- Organizes usage and cost data by title of publication, publisher, discipline, product
- Provides a dashboard-style landing screen with shortcuts to data visualizations

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

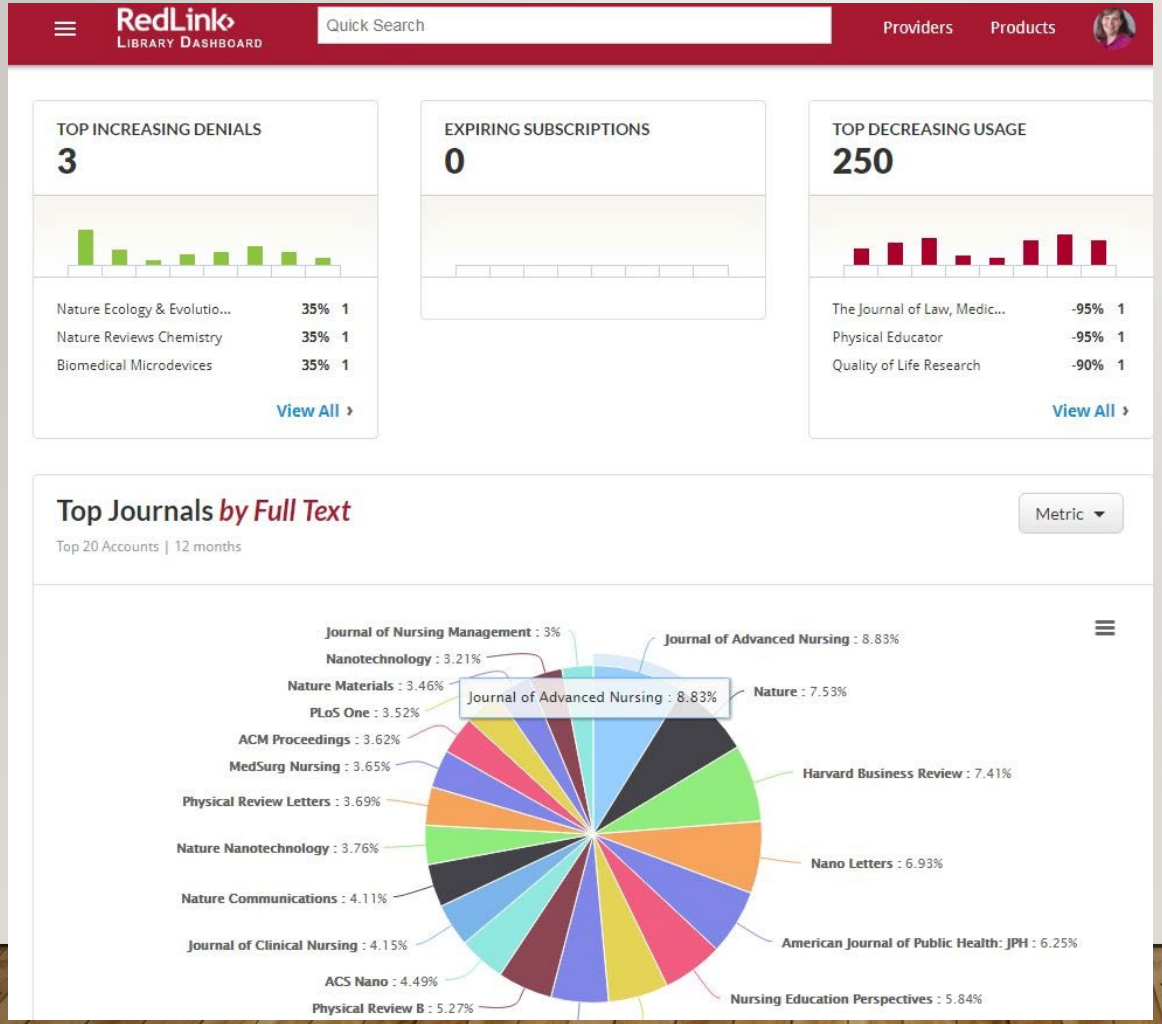
- Web-based and fully-hosted requiring no software installation
- Includes unlimited accounts for library staff and unlimited number of data sources that one can track
- Configured to utilize SUSHI protocol or COUNTER standard

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

- Initial setup is estimated at 4-6 weeks
- Option to send historical reports to upload to dashboard
- Fetches most of the required COUNTER reports (JR1, JR2, JR5, DB1, DB2, BR1-3)

New Paltz
STATE UNIVERSITY OF NEW YORK

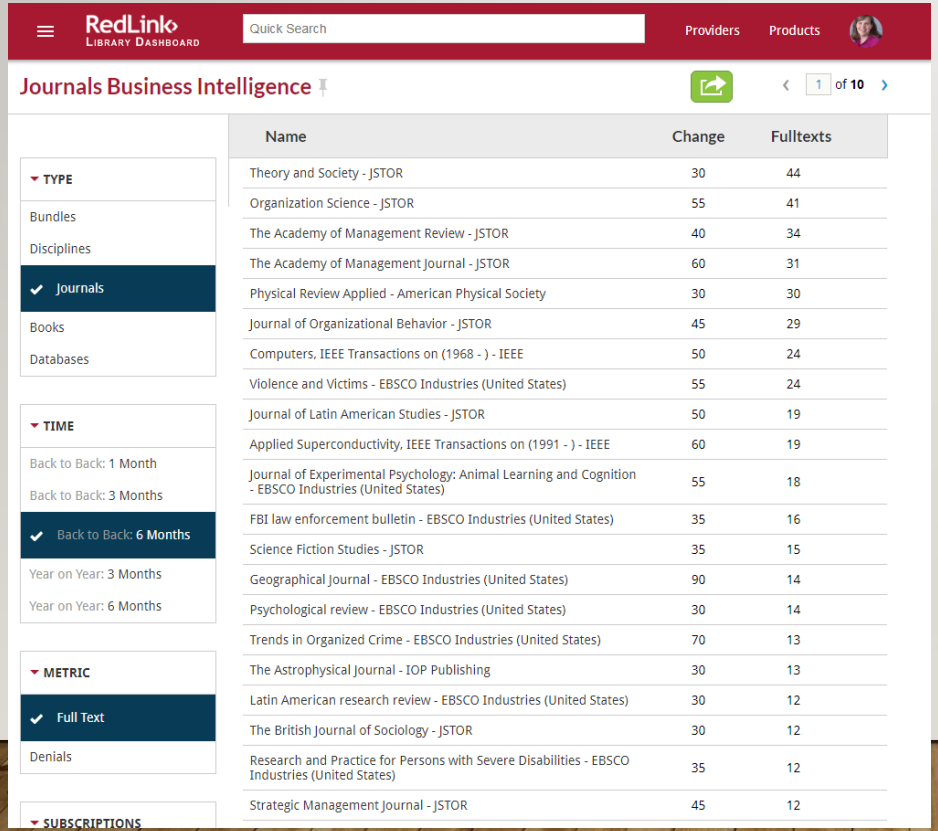


Dashboard shortcuts:

- *Top increasing denials*
- *Expiring subscriptions*
- *Top decreasing usage*
- *Top journals by full text usage or denials*

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink



Drill down further with search filters:

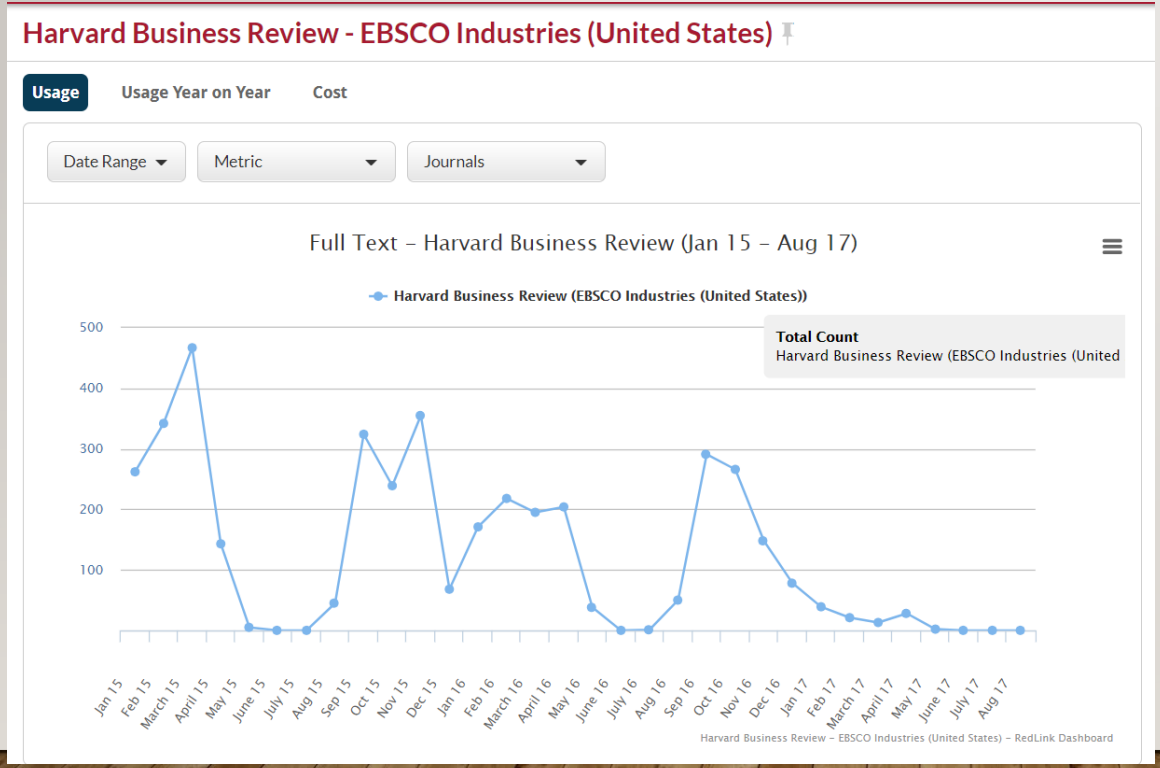
- *Content type (book, journal, bundle, database)*
- *Time period*
- *Discipline*
- *Metric (full text or denials)*
- *Subscription status*
- *Degree of change*

- *Bookmark or save as a pin*
- *Export excel or data visualization*

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

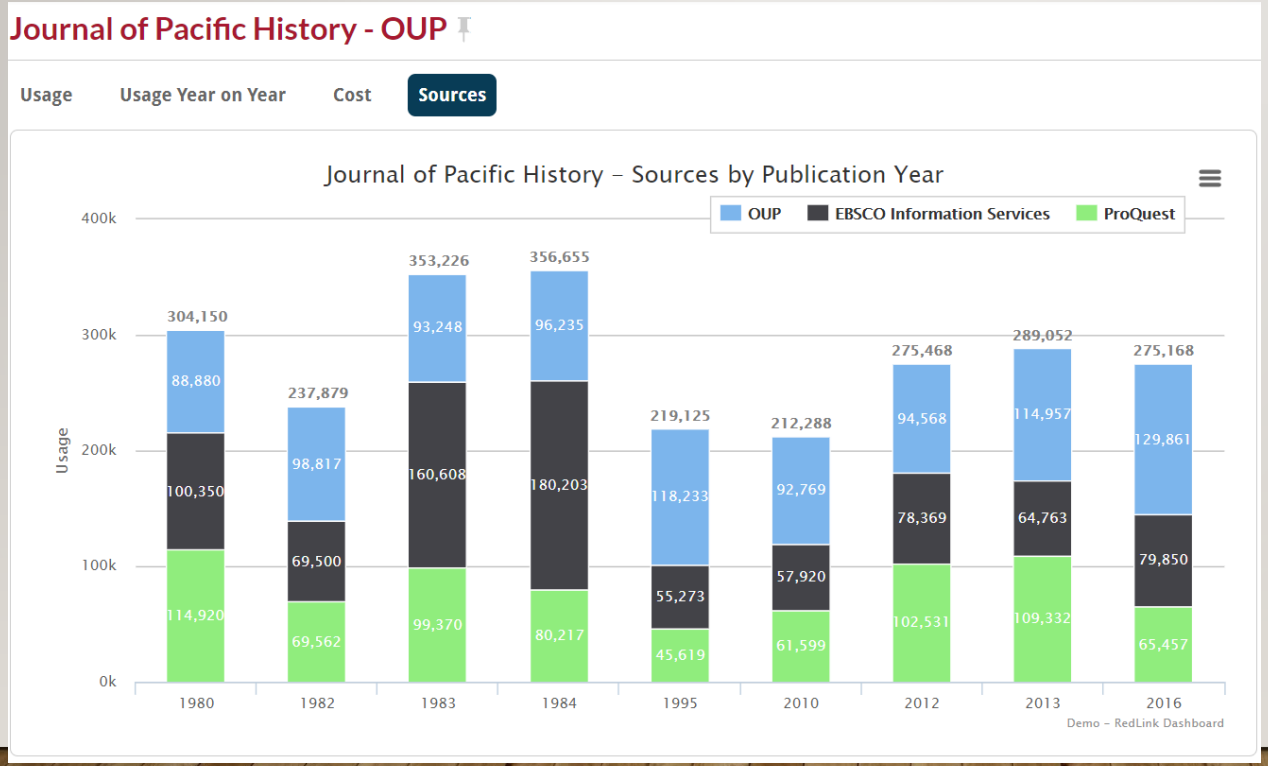
Publication-level analytics



New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

Overlap analytics



New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

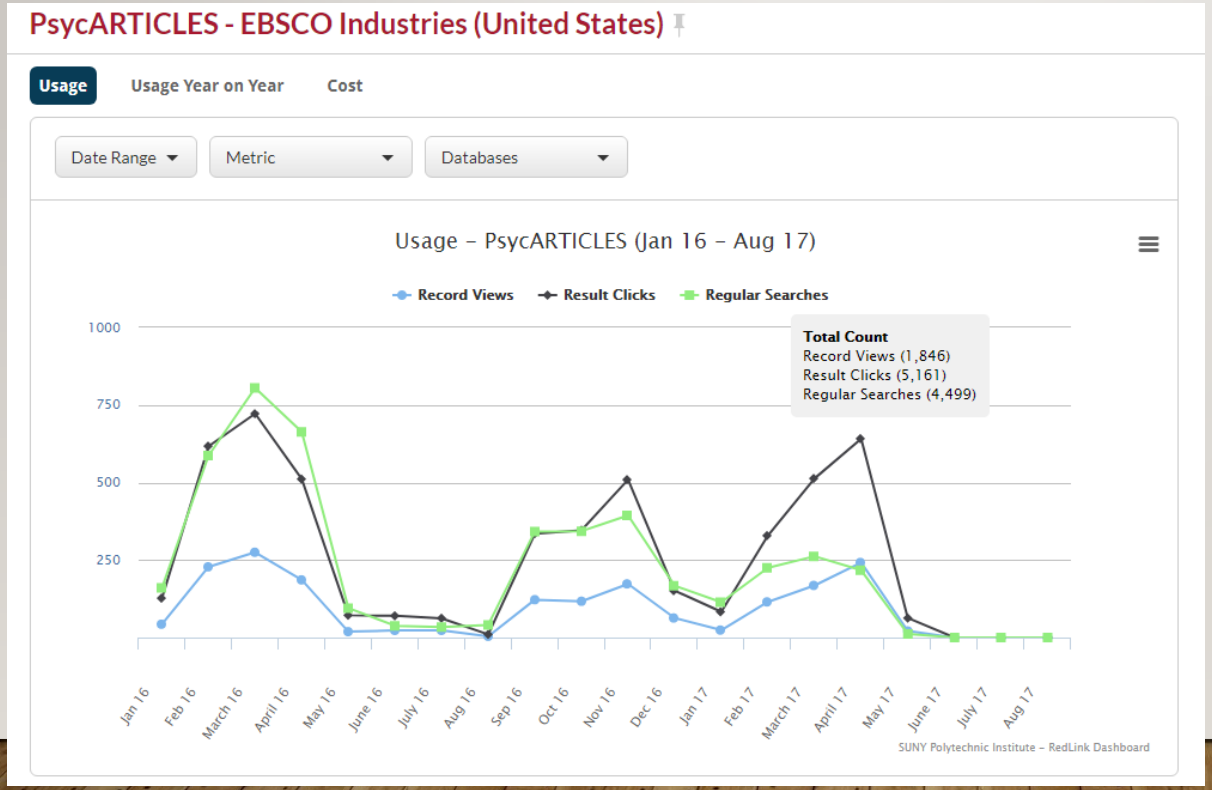
Content slice

- *Pulls data from JR5 about usage by publication year*

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink

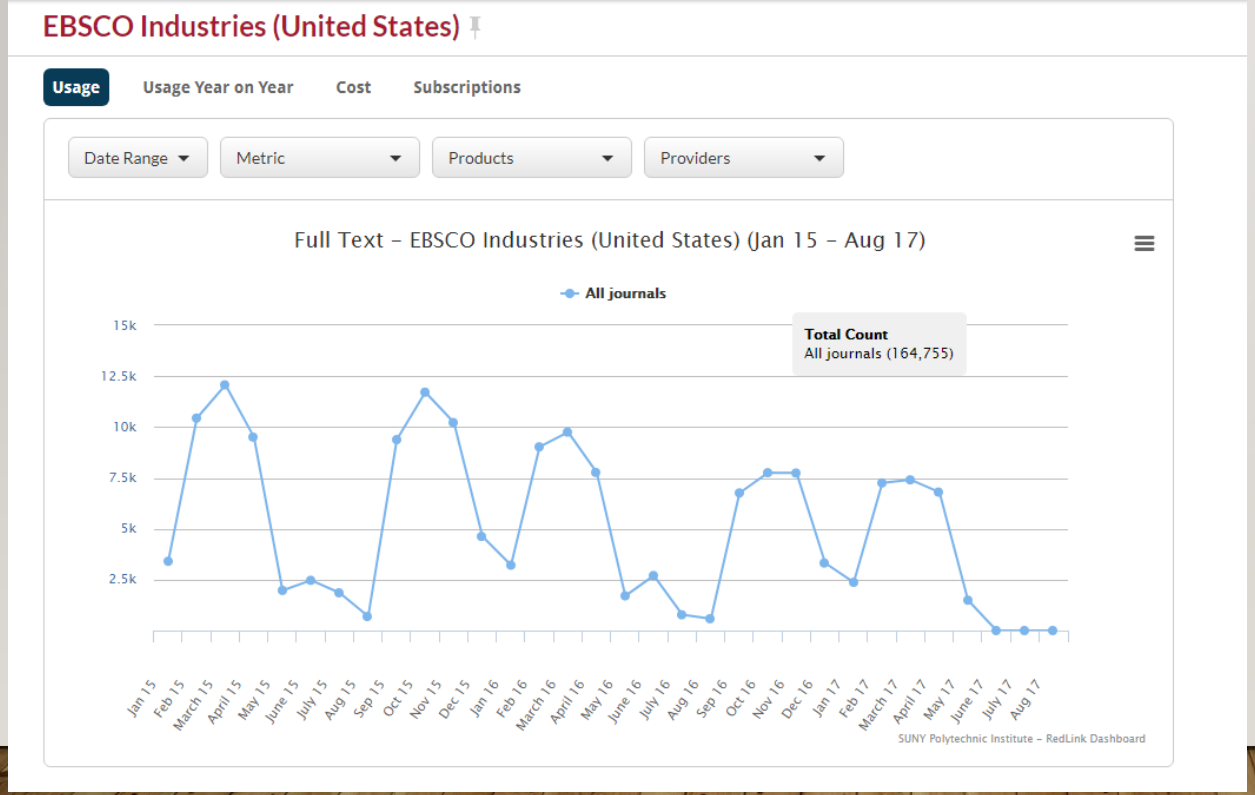
Product-level analytics



New Paltz
STATE UNIVERSITY OF NEW YORK

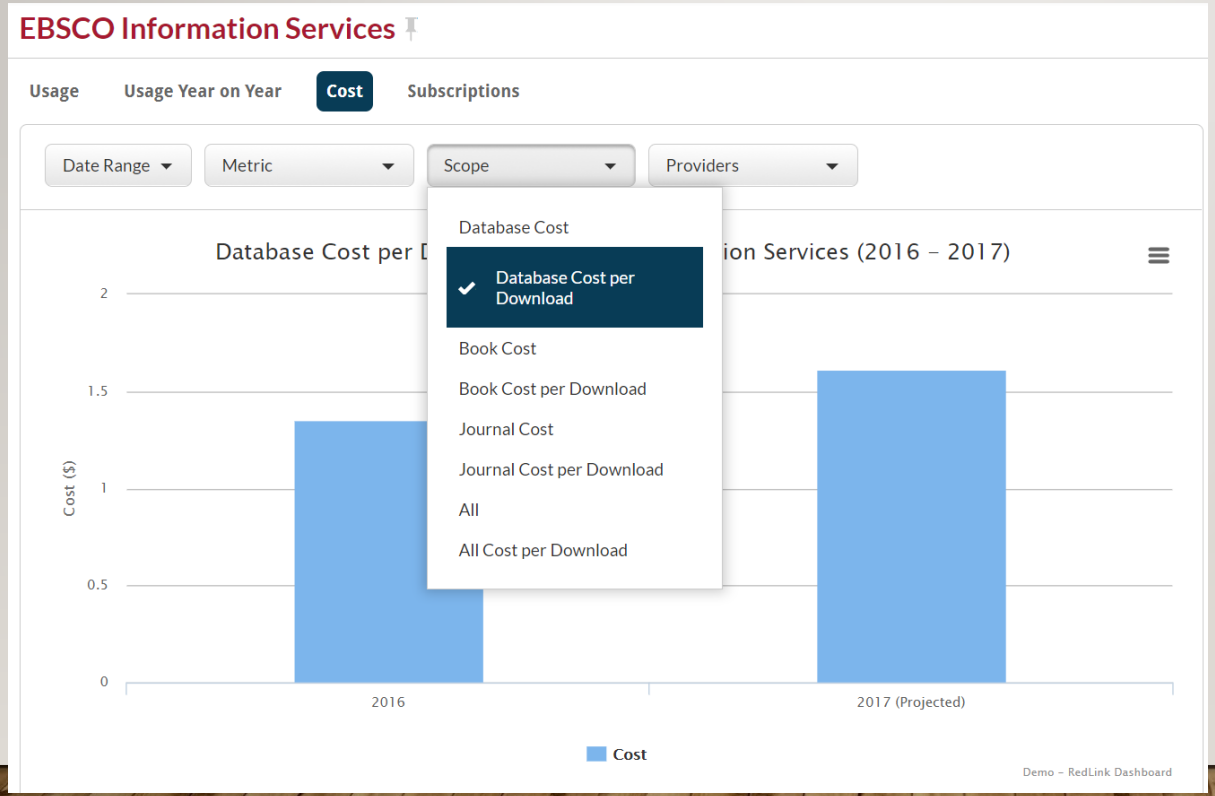
RedLink

Provider-level analytics



New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink



Provider-level cost analytics

Total and per download costs by

- ***Database***
- ***Book***
- ***Journal***

New Paltz
STATE UNIVERSITY OF NEW YORK

RedLink®

- Facilitates evidence-based collection assessment
- Initial onboarding requires time investment
- Beneficial for smaller libraries with limited staffing

New Paltz
STATE UNIVERSITY OF NEW YORK

DISCUSSION OF EXCEL

TRACY GILMORE

VIRGINIA TECH

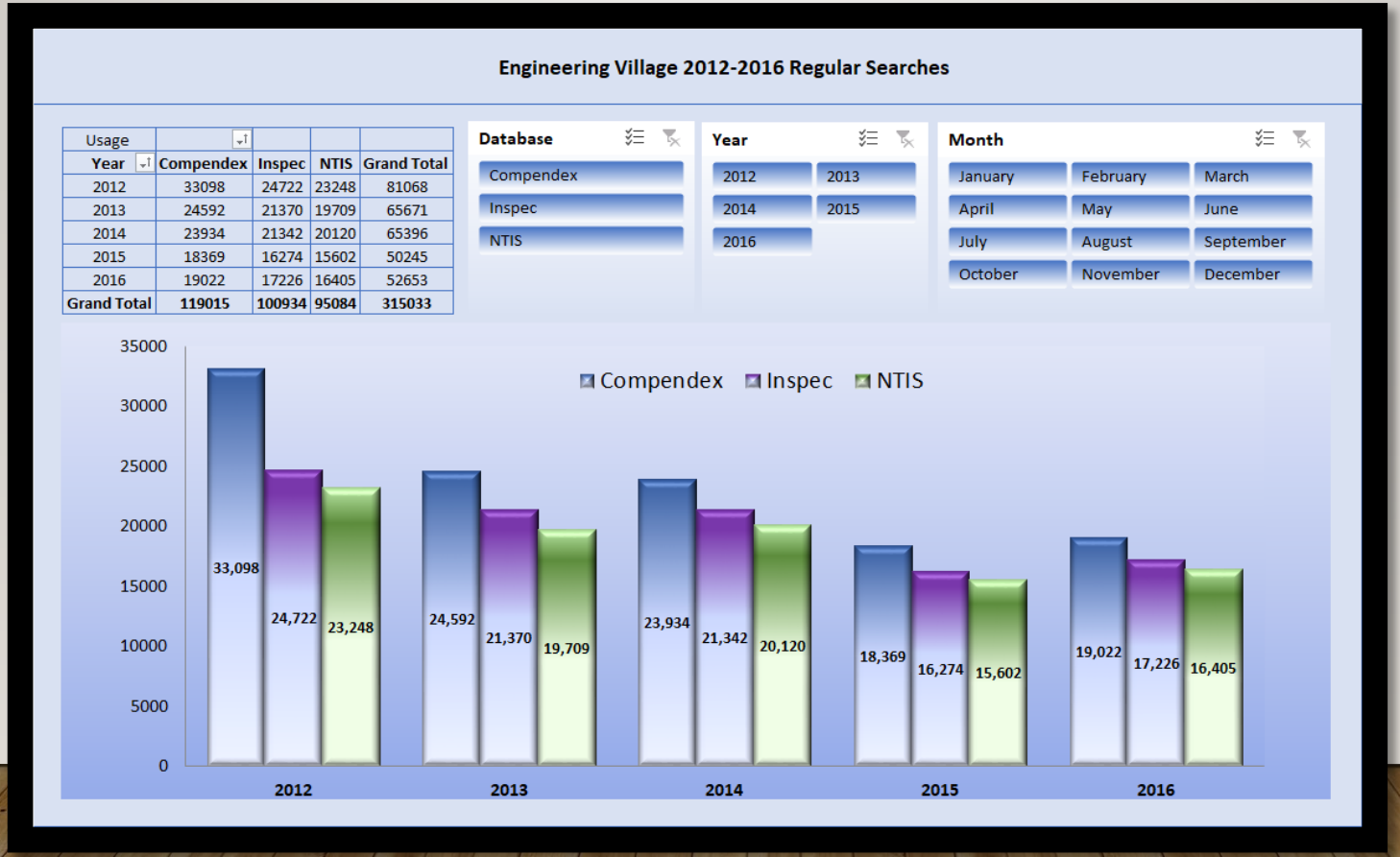
Microsoft® Excel

Creating Interactive Dashboards with Excel using:

- Power Pivot
- Pivot Tables
- Pivot Charts
- Sliders
- Some Design

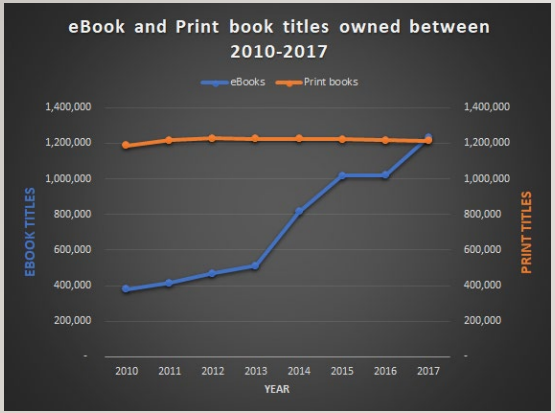
VirginiaTech Invent the Future®

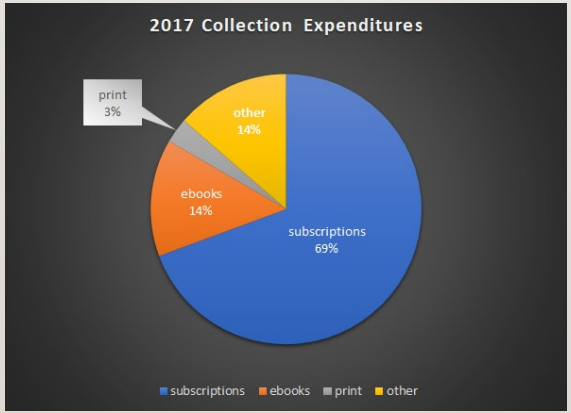
INTERACTIVE DASHBOARDS

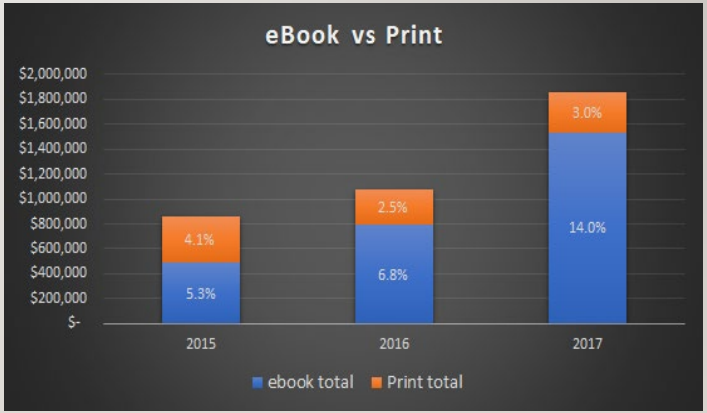


Crash Course in Dashboards with Excel

EXCEL - INTERNAL REPORTS







Microsoft® Excel

Training Resources

- Free and online

Excel Versions

- 2010
- 2013
- 2016

lynda.com

YouTube

Microsoft

Google

DISCUSSION OF TABLEAU

CONNIE STOVALL

VIRGINIA TECH

- Connects to many data types (Excel, csv, databases, web data)
- Provides tools for shaping data into table format
- Allows users to build calculations (like Excel) but also equipped with quick analysis tools
- Allows for custom design for beautiful, dynamic dashboards, vizzes, and stories

VirginiaTech
1872
Invent the Future®

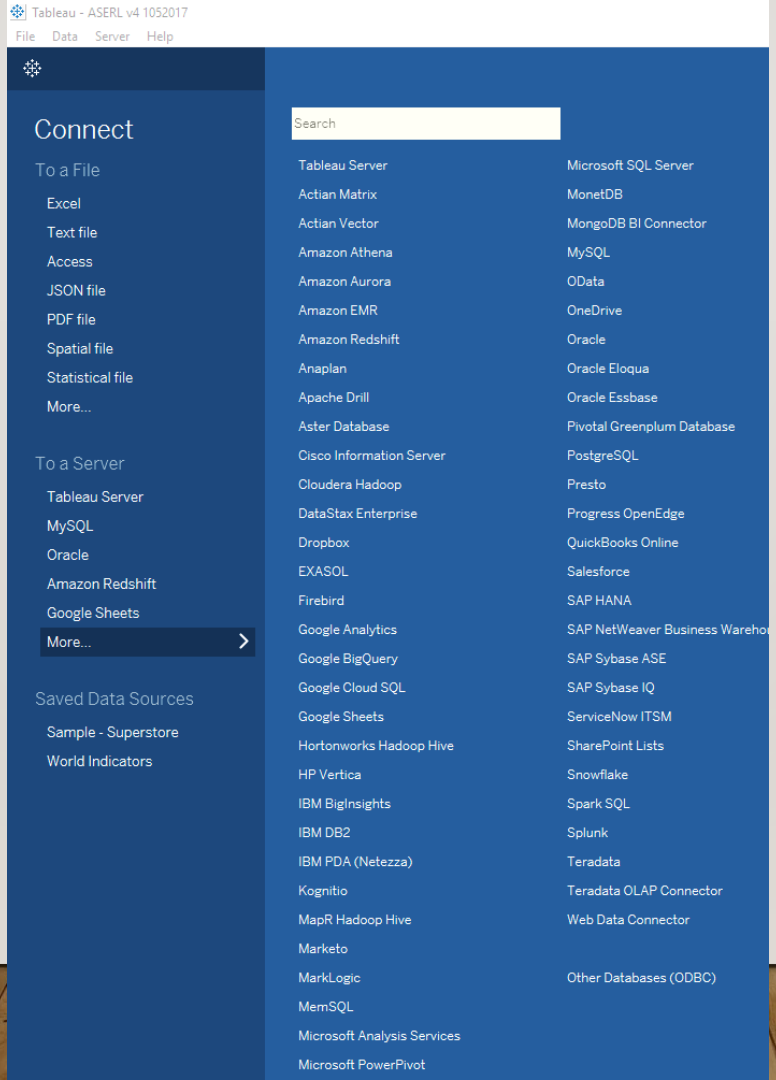
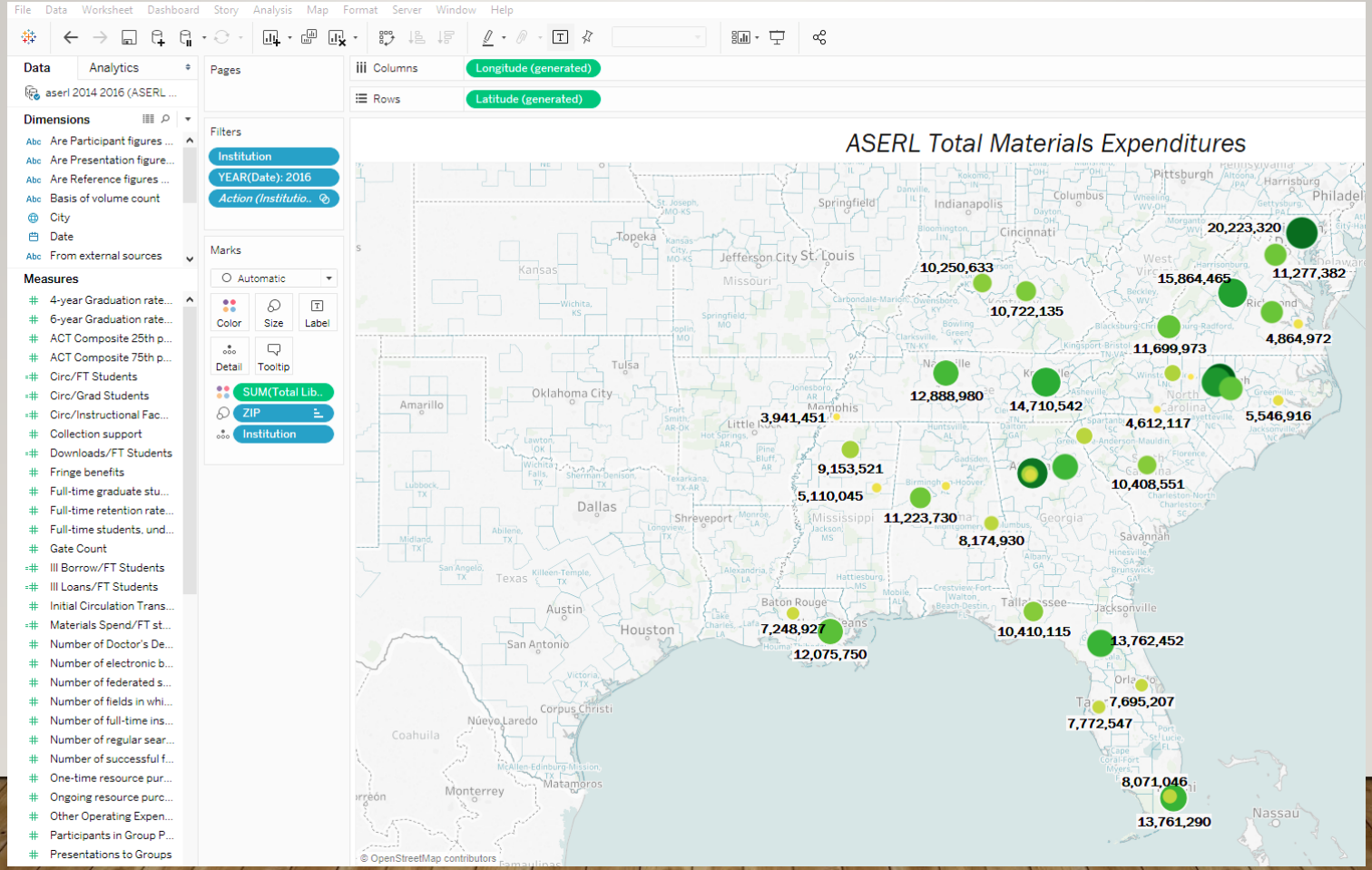
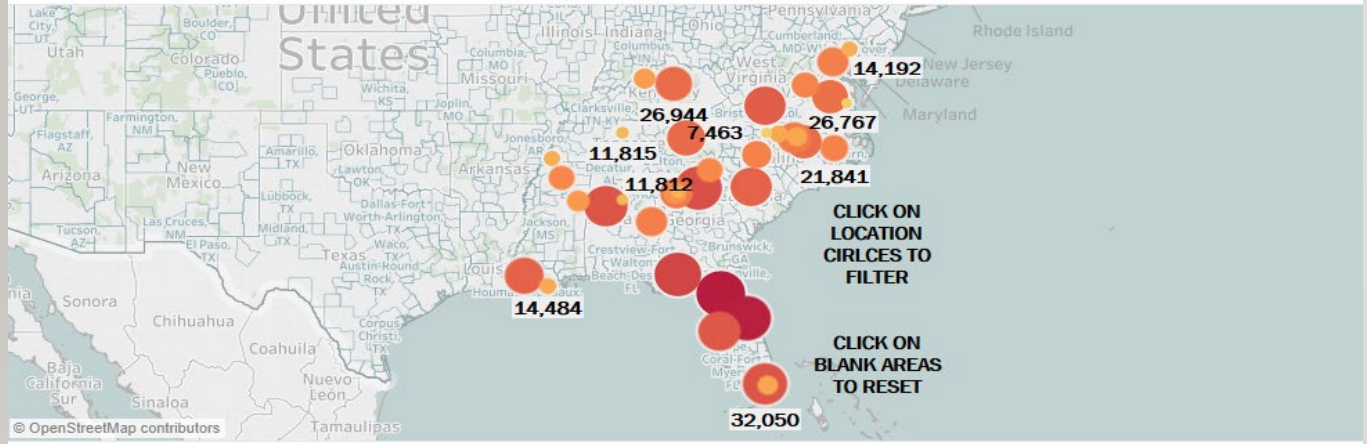


Tableau - ASERL v4 1052017
File Data Server Help
Connect
To a File
Excel
Text file
Access
JSON file
PDF file
Spatial file
Statistical file
More...
To a Server
Tableau Server
MySQL
Oracle
Amazon Redshift
Google Sheets
More...
Saved Data Sources
Sample - Superstore
World Indicators
Search
Tableau Server
Actian Matrix
Actian Vector
Amazon Athena
Amazon Aurora
Amazon EMR
Amazon Redshift
Anaplan
Apache Drill
Aster Database
Cisco Information Server
Cloudera Hadoop
DataStax Enterprise
Dropbox
EXASOL
Firebird
Google Analytics
Google BigQuery
Google Cloud SQL
Google Sheets
Hortonworks Hadoop Hive
HP Vertica
IBM BigInsights
IBM DB2
IBM PDA (Netezza)
Kognitio
MapR Hadoop Hive
Marketo
MarkLogic
MemSQL
Microsoft Analysis Services
Microsoft PowerPivot
Microsoft SQL Server
MonetDB
MongoDB BI Connector
MySQL
OData
OneDrive
Oracle
Oracle Eloqua
Oracle Essbase
Pivotal Greenplum Database
PostgreSQL
Presto
Progress OpenEdge
QuickBooks Online
Salesforce
SAP HANA
SAP Sybase ASE
SAP Sybase IQ
ServiceNow ITSM
SharePoint Lists
Snowflake
Spark SQL
Splunk
Teradata
Teradata OLAP Connector
Web Data Connector
Other Databases (ODBC)

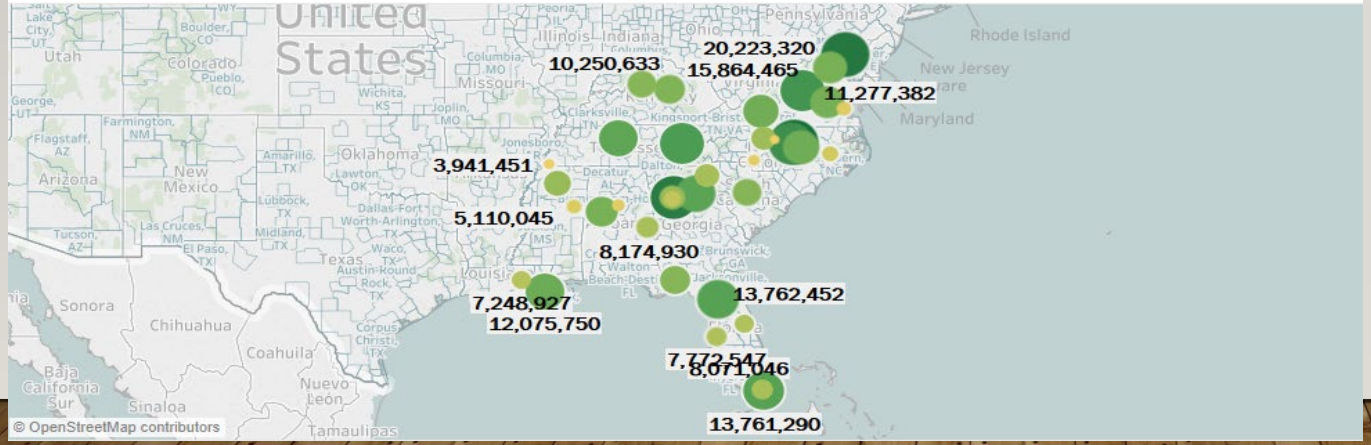
Inside Tableau



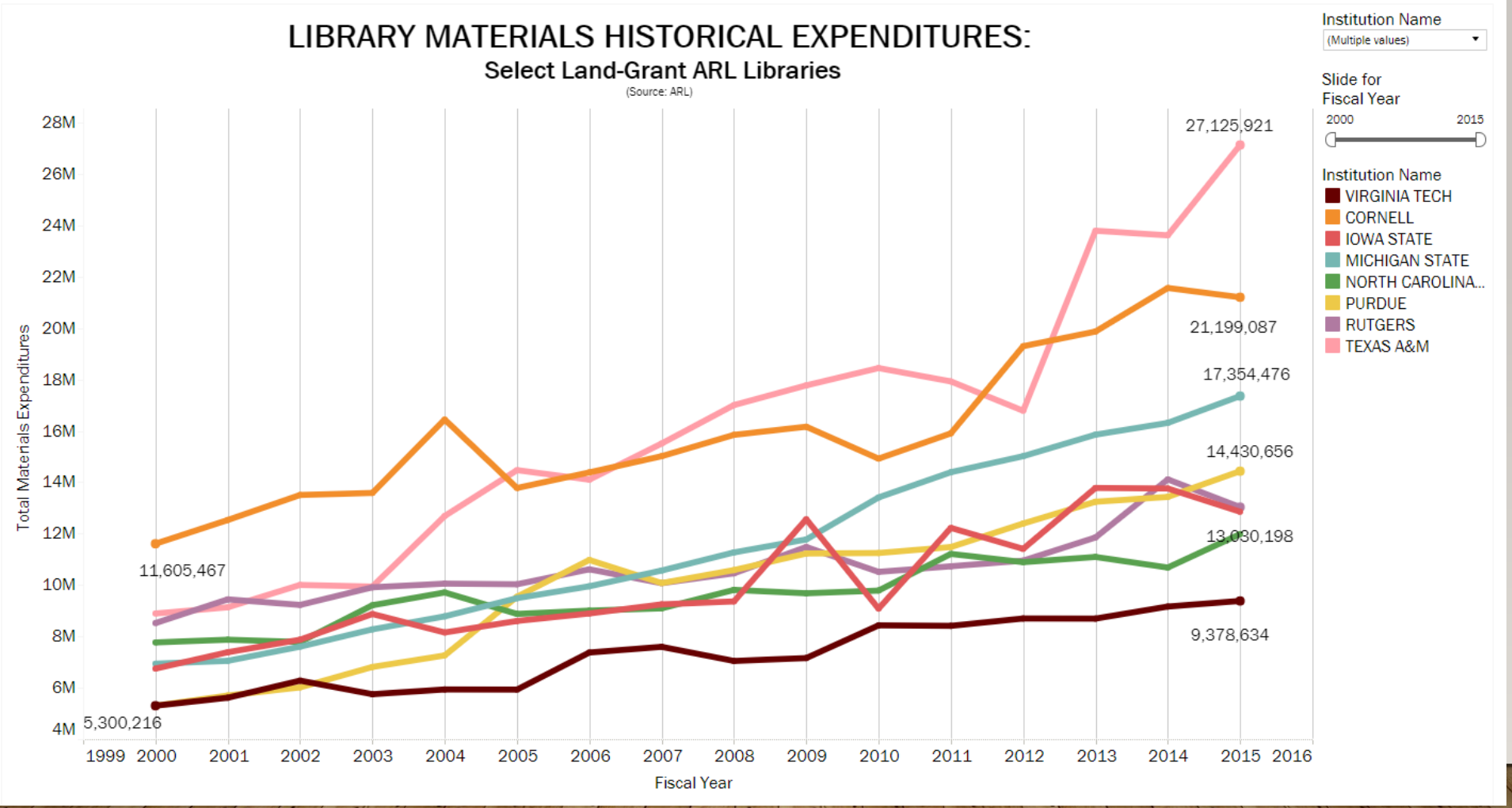
ASERL Full Time Students



ASERL Total Materials Expenditures



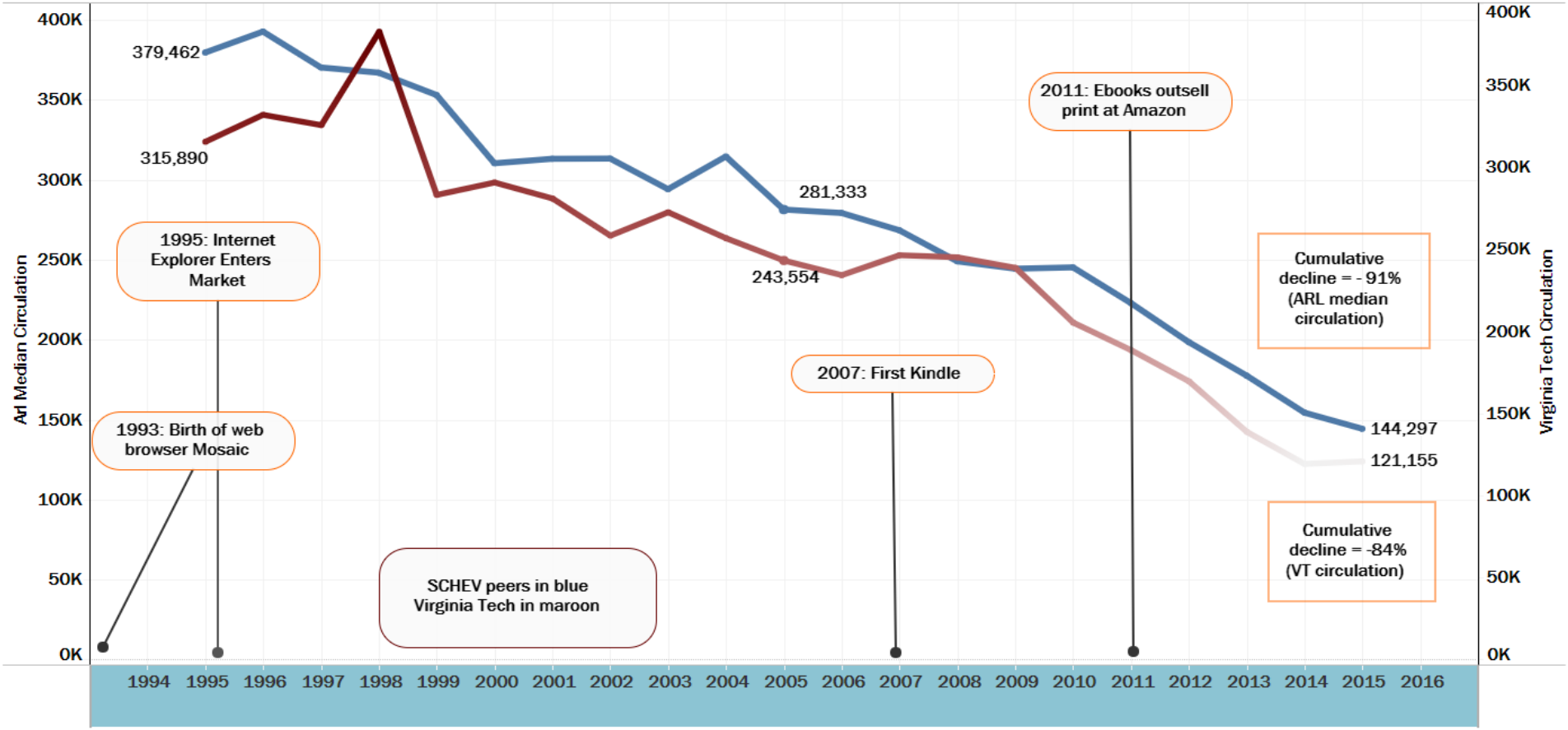
Map Vizzes



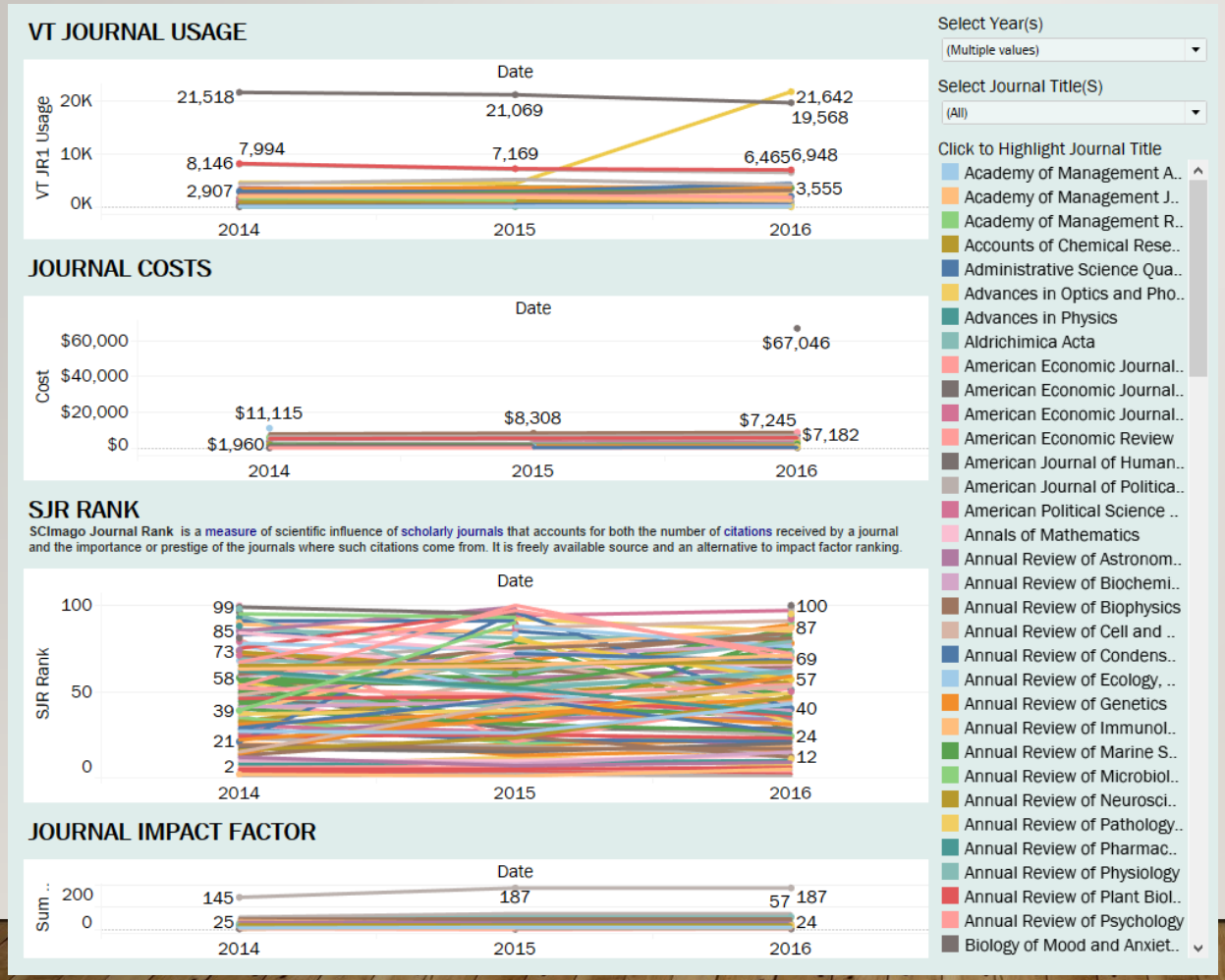
Historical Materials Viz

CIRCULATION 1995 - 2015

Source: ARL



Historical Circulation Viz



Multiple Dashboards

LOOKING AT ASERL DATA IN NEW WAYS

(JUST A DRAFT. DON'T QUOTE, PLEASE.)

ABOUT ASERL | ANSWERING NEW QUESTIONS WITH EXISTING DATA | QUESTIONS TO ANSWER | FIRST THINGS FIRST | FULL TIME STUDENT MAP | LIBRARY MATERIALS EXPENDITURES | QUE

aserl

ASSOCIATION OF SOUTHEASTERN RESEARCH LIBRARIES

About ASERL

For 60 years, the Association of Southeastern Research Libraries has brought together leaders from research, federal, and state libraries in the region to foster a high standard of library excellence through inter-institutional resource sharing and other collaborative efforts. Today, ASERL is the largest regional research library consortium in the United States and consists of 37 unique institutions. By working together, ASERL members continue to provide and maintain top quality resources and services for the students, faculty, and citizens of their respective communities.

New Metrics for Research Libraries

In 2016, ASERL began a pilot project that will help libraries develop new ways to tell their story beyond traditional metrics such as circulation and door counts. We propose to examine our existing metrics in new ways that emphasize how research libraries support teaching and learning. **What you see within is just an initial step and a very rough draft to begin the process.** The three-year pilot project plans to establish small groups of variables in three areas: collection use, mediation services, and facility usage. By the end of the project we plan to deliver a model for other libraries to use in telling their stories – a model that uses data already available to libraries.

tableau

Creating Stories

VirginiaTech 1872 Invent the Future®

- Learning Curve
- Tech Support and Implementation
- Licensing Options and Sharing Work
- Take a look at the Tableau Public Gallery

Product	Tableau Desktop	Tableau Online	Tableau Server	Tableau Reader	Tableau Public
Functions	create workbooks, vizzes, dashboards, stories, store locally	store, share, discover workbooks, vizzes, dashboards, stories	store, share, discover workbooks, vizzes, dashboards, stories	view workbooks, vizzes, dashboards, stories	create and store workbooks, vizzes, dashboards, stories
Licensing	required	required	required	free	free/10 GB storage limit

DISCUSSION OF ONEONTA PROJECT

CHENG CHENG

SUNY ONEONTA

ONEONTA PROJECT--BACKGROUND

- Need vision for what's going on in the market
- Administration and stakeholder need something understandable
- A practice for future operation

OBJECTIVES

- Analyze and visualize the price changing pattern of common electronic resources
- Provide predictions for future price changes at vendor level
- Discover any potential cause of such price changes
- A practice of skills and techniques for statistical analysis and data visualization

Criteria and data cleaning

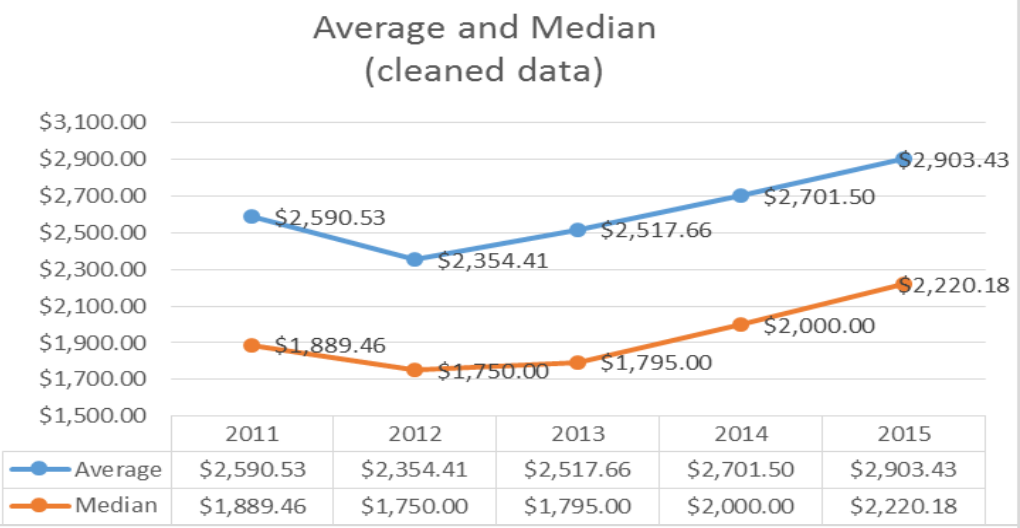
- More than 80 aggregating databases with consistent annual subscriptions from 2010-2015
- No electronic serials, ebooks, one-time purchases and annual access fees were included
- Titles without continuing payment records, or have value far away from the mean, were also eliminated

Tools:

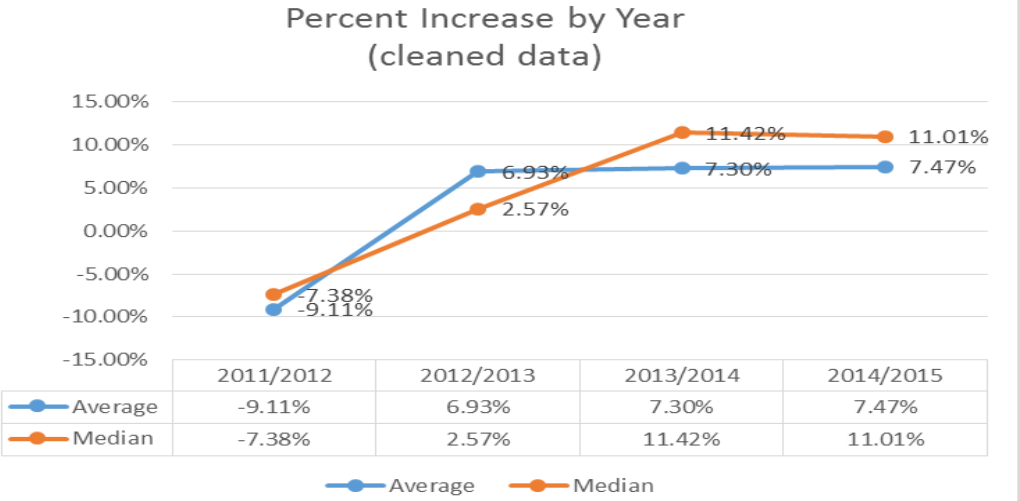
- Office Excel
- IBM SPSS
- And most importantly……
- ……HUMAN

On the right, we have both price and percent increase of selected resources over years.

As we can see, graphics show there is a consistent price increase since 2011; The price increases in 2014 and 2015 were fierce

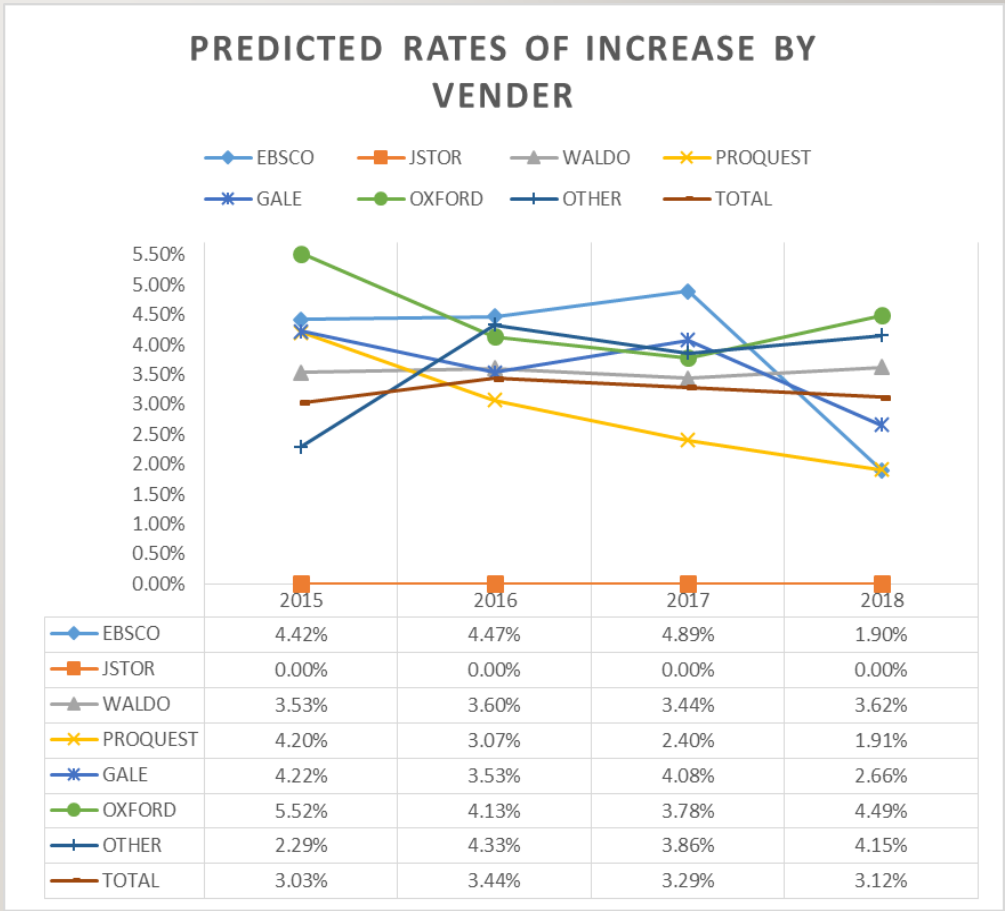


	2011	2012	2013	2014	2015
Average	$2,590.53	$2,354.41	$2,517.66	$2,701.50	$2,903.43
Median	$1,889.46	$1,750.00	$1,795.00	$2,000.00	$2,220.18



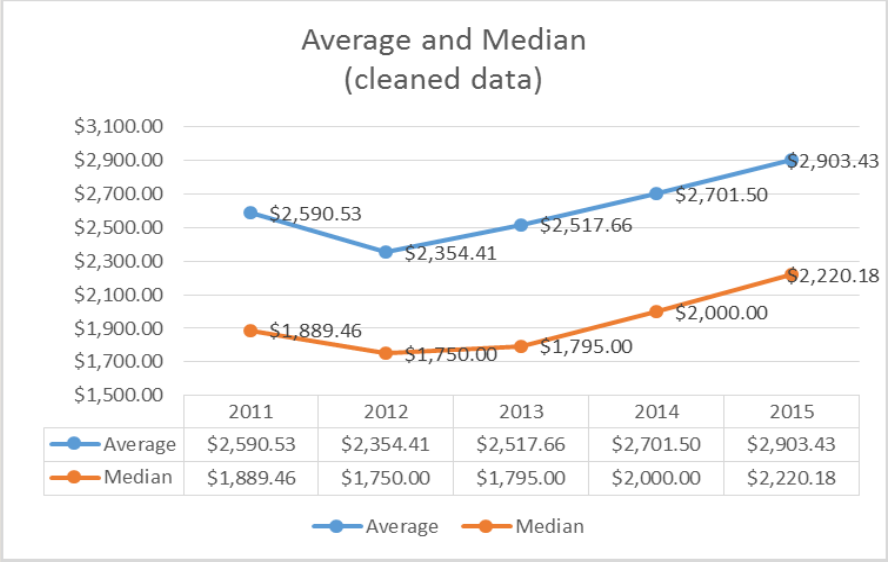
	2011/2012	2012/2013	2013/2014	2014/2015
Average	-9.11%	6.93%	7.30%	7.47%
Median	-7.38%	2.57%	11.42%	11.01%

- The projectile (on the right) based on past stats suggest in 2017 and 2018 it may be relatively moderate, which is likely a sign that this price increase cycle is approaching its cap.
- The model for future price prediction is based on the expenditure data from the previous years, and that value varies based on vendor.
- It is important to note that the further into the future the prediction, the less accurate it is.
- It is also important to note that compare to the enormous size of the aggregating databases market with numerous products available, this project only covered a very small facet of it.

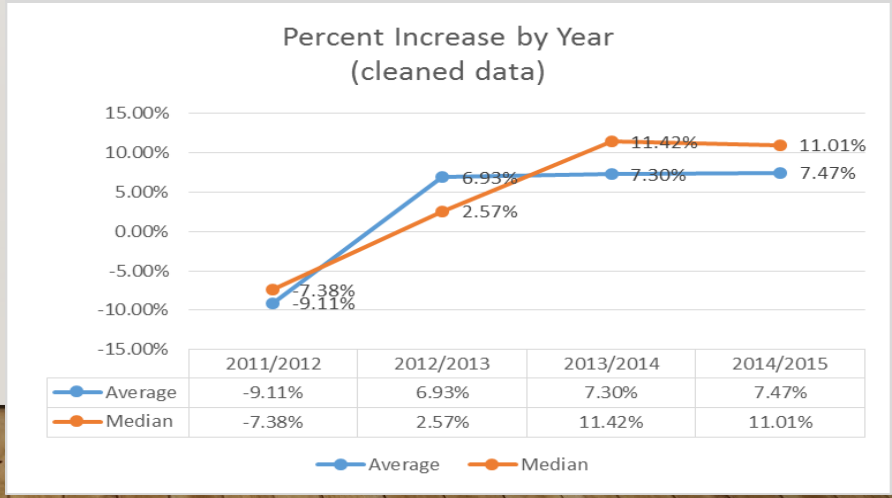


	2015	2016	2017	2018
EBSCO	4.42%	4.47%	4.89%	1.90%
JSTOR	0.00%	0.00%	0.00%	0.00%
WALDO	3.53%	3.60%	3.44%	3.62%
PROQUEST	4.20%	3.07%	2.40%	1.91%
GALE	4.22%	3.53%	4.08%	2.66%
OXFORD	5.52%	4.13%	3.78%	4.49%
OTHER	2.29%	4.33%	3.86%	4.15%
TOTAL	3.03%	3.44%	3.29%	3.12%

Compared the price changing trends and the picture of general economy (up) at the same period, it is very obvious that the pricing trend has been influenced by the national economy trend

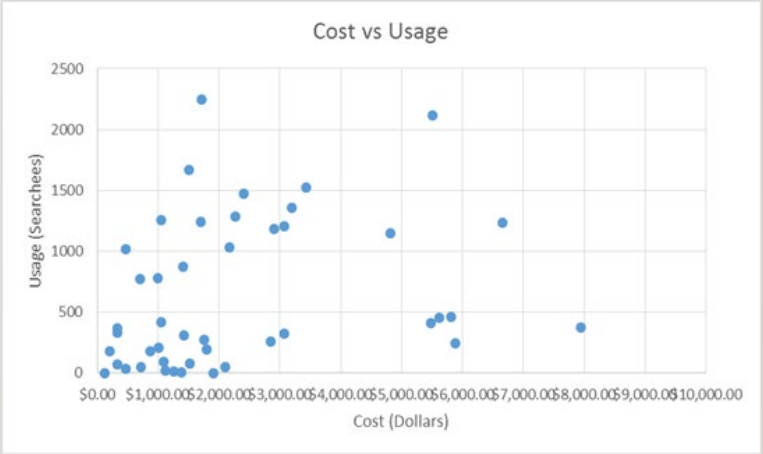


	2011	2012	2013	2014	2015
Average	$2,590.53	$2,354.41	$2,517.66	$2,701.50	$2,903.43
Median	$1,889.46	$1,750.00	$1,795.00	$2,000.00	$2,220.18



	2011/2012	2012/2013	2013/2014	2014/2015
Average	-9.11%	6.93%	7.30%	7.47%
Median	-7.38%	2.57%	11.42%	11.01%

The connection between usage and cost is less obvious, although resources with higher usage tend to cost more, probably because to vendors, high usage indicates high-demand.



PRO & CON

- Provide visualized trend of Electronic Resources price changes
- Prediction for future price change
- It proved that a statistical analysis and data visualization project could be done in small academic libraries with minimum cost.

CONCLUSION

- This project is conducted in a very small scale, based on the limited resources and data
- Need further research and better understanding of current business model to determine the true causes
- A similar research project at much larger scale will be necessary