Statistical Analysis, Data Visualization, and Business Intelligence Tools for Electronic Resources in Academic Libraries

Cheng Cheng
SUNY Oneonta, cheng.cheng@oneonta.edu

Tracy J. Gilmore
Virginia Tech, tgilmore@vt.edu

Colleen Lougen
SUNY New Paltz, lougenc@newpaltz.edu

Connie Stovall
Virginia Tech, cjstovall@ua.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

Part of the [Business Intelligence Commons](#), and the [Collection Development and Management Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Cheng Cheng, Tracy J. Gilmore, Colleen Lougen, and Connie Stovall, "Statistical Analysis, Data Visualization, and Business Intelligence Tools for Electronic Resources in Academic Libraries" (2017). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316679>

Statistical Analysis, Data Visualization, and Business Intelligence Tools for Electronic Resources in Academic Libraries

Cheng Cheng, SUNY Oneonta

Tracy J. Gilmore, Virginia Tech

Colleen Lougen, SUNY New Paltz

Connie Stovall, Virginia Tech

Abstract

Analytics, business intelligence, and data visualization have an increasingly important space in the changing assessment landscape of academic libraries. Today, most academic libraries are spending a significant amount of their total expenditure budget on electronic resources. While they are high-value assets for modern libraries, the expensive annual subscription cost and continuing price increases of e-resources also make them a substantial budgeting burden. It is therefore essential to have a clear statistical view of the trends and patterns in price changes for e-resources and vendors. This paper focuses on three major topics vital for developing such a view: first, the e-resources landscape and the need for evidence-based decision-making, analysis, and reporting; second, the methodology applied in a statistical analysis and data visualization research project, as well as business trends and patterns in the field of libraries' e-resources; and third, business intelligence tools available for this type of analysis, including low-cost options such as Excel and Redlink, and premium options such as Tableau.

Introduction

Producing evidence-based data and robust visualizations have a growing importance in the assessment landscape of academic libraries. Collecting, analyzing, and translating data into information are vital in evaluating the return on investment of expensive electronic resources. Business intelligence (BI) tools present advantages to librarians as these resources continue to proliferate and more data becomes available, and librarians are tasked with using data to convey trends and actionable insights. Redlink, Excel, and Tableau are BI tools that represent the spectrum of pricing and licensing options that are available and widely used by assessment and other academic librarians. Each is briefly summarized herein.

Redlink

Library staff are inundated with electronic resource data. Often, they spend an excessive amount of time manually downloading COUNTER and non-COUNTER usage statistics from a myriad of platforms and then consolidating the information into a single report to use in evidence-based decision making. Many libraries struggle with finding the time to collect and analyze their costly e-resources or do not possess the technical skills to conduct a homegrown assessment project.

The Redlink Library Dashboard provides an economical, subscription-based business intelligence tool that automates the harvesting and normalizing of COUNTER and non-COUNTER usage statistics for a library's electronic resources. This tool instantly merges and organizes the data into a single platform so staff can efficiently access the information on demand. Users can quickly ascertain and evaluate relative usage of a given resource. Additionally, the Library Dashboard promptly calculates the highly coveted cost per use metrics and produces robust data visualizations that make reading and understanding usage easier and more engaging.

A substantial time investment exists regarding initial setup of the Redlink Library Dashboard system. Setup requires libraries to gather and send all their SUSHI protocols, log-in credentials, and subscription/cost information to Redlink. This "onboarding" process is estimated to take 4–6 weeks.

The main advantages of the Redlink Library Dashboard system are:

- Saves critical staff time through the automated collection and standardization of usage statistics.
- Filters and engineers data on a user-friendly, Web-based, and fully hosted platform.

- Reports usage and cost data on demand for efficient sharing with staff and stakeholders.
- Provides ability to compare usage and cost data for overlapping titles from multiple sources.
- Creates dynamic data visualizations that expedite data analysis and decision making with more transparent, readily communicated information.

Some disadvantages of the Redlink Library Dashboard are:

- No trials available because of the significant onboarding process.
- No training or support documentation from Redlink by the time of the conference presentation.

The Redlink Library Dashboard is a low-cost and effective business intelligence tool that empowers librarians to extract meaningful information from the vast amount of electronic resource usage data, saving critical staff time and facilitating data-driven decisions. It is an essential tool that helps tell the story of a library's usage of electronic resources, and in turn, clearly depicts the value of these costly assets.

Excel

Excel is one of the most widely used business intelligence tools for analyzing and visualizing data. For librarians, who manage large amounts of usage and statistics data, one of Excel's many advantages is its familiarity and accessibility. The Collection Assessment Team at Virginia Tech University Libraries uses Excel extensively to communicate information about our online resources in visual, interactive reports for collection management, college librarians, and library administrators.

Data collection, analysis, and reporting of decisions based on aggregated data is a requirement in today's data-driven library environment. A library's strategic direction and its rationale for decision making are predicated on evidence-based data, and stakeholders expect libraries to generate on-demand reports that explicitly underscore the data used to support these decisions. When working with aggregated usage data and COUNTER reports, Excel spreadsheets remain our front-end tool for wrangling data. Excel features such as Power Pivot, Power Query, and Power View,

available for the 2010 and 2013 versions, provide a customized way for us to view new perspectives and insights from our data.

The Collection Assessment Team began utilizing these advanced Excel options to support a more robust program of data analysis, visualization, and transparency. Building dashboards, interactive reports, and geospatial maps is possible with PivotTables, PivotCharts, and slicers, and these tools allow us to provide an array of visualizations and data interactions for our stakeholders. The dashboards and visualizations that we create need to be shared in a variety of Web spaces, some private and some public-facing. Using Excel within the Virginia Tech, Microsoft Sharepoint server allows us to share dashboards, visual representation of the data, and the underlying data in a myriad of ways.

The Collections Assessment Team needed to become more agile in evaluating big deal packages, return-on-investment, and analysis of electronic resource cost-per-use data by platform, publisher, discipline, and college. Package analysis, cost per use, usage trends, and Paretos can all be done in Excel. Using Pivot Tables, data can be summarized and filtered, and reports can be created dynamically by changing the selected data in the PivotTable Fields list, or adding a calculated field to the table. Multiple Pivot Charts can then be created from one Pivot Table and synchronized using a slicer.

Pivot Charts create the basis for dynamic dashboard and visualizations. There are a number of chart options including pie, column, bar, line, scatter, bubble, and other selections. Slicers provide a way to filter the data in a Pivot Chart and to dynamically change the data displayed. They are basically buttons that enable the quick filtering of data without the need to open a drop-down list to find items, and Slicers can connect and filter multiple Pivot Charts.

The cost to acquire the software and train employees is minimal. At Virginia Tech, as at many universities and large institutions, Excel is provided via an enterprise license. Our team gained the necessary advanced skills training primarily through Lynda.com, an online learning platform, which was also provided via an enterprise license. However, we also utilized other free online training platforms including the Microsoft Office 365 Training Center and YouTube. Excel's powerful analytic capabilities along with its accessibility and familiarity within the business realm make it one of the most accessible and widely used BI tools.

Tableau

Beyond creating visualizations and dashboards for COUNTER usage data, the Virginia Tech team needed to work with institutional data and needed a tool that allowed for the following: sharing dynamic visualizations via the Web, connecting to a wide variety of data sources, and sharing data sources among colleagues. Product reviews and word of mouth brought us to Tableau, a well-known BI tool particularly known for its visualizations. In 2017 *PC Magazine* named Tableau Desktop Editor's Choice Winner (Baker, 2017). Users include the likes of Citibank, Toyota, Dell, and a host of universities.

Sharing of data sources works easily with Tableau Online or Tableau Server, but a strong word of caution is necessary here when it comes to most BI tools including Tableau. To build and share visualizations and dashboards, *users need a license*. To share those dynamic visualizations and dashboards created in Tableau Desktop and the underlying data sources associated with them, users must also acquire an additional license to Tableau Online or Tableau Server, both of which present an additional cost per user. Two free alternatives exist for sharing visualizations and dashboards: Tableau Reader and Tableau Public, but both are less than ideal if your library plans to *routinely* use BI tools. Tableau Reader, which functions much like Adobe Reader, is free for anyone to download and use to view workbooks created by Tableau Desktop users. Tableau Public is a good choice for small projects; it ultimately limits users who need to make routine use of a BI tool. Sharing visualizations in Tableau Public means it is truly publicly available to anyone with a Web browser (Tableau Server and Tableau Online require an account and sign-in.). Second, Tableau Public limits users to 10 gigabytes of storage space. A third limitation, and a rather large one, is that Tableau Public does not allow users to store data sources to reuse and/or share in another workbook, meaning that users may find themselves reinventing the wheel. Lastly, Tableau Public does not offer users a way to organize and search for saved work. If your library is looking only to make a handful of publicly available visualizations via the Web, Tableau Public offers users an excellent choice. Tableau Public also provides users with an embed code so that visualizations can be embedded into Web pages.

Connecting to a wide variety of data source types makes BI tools very attractive. In addition to working with data collected in Excel or CSV files, Tableau

users can connect to a number of file types and services including Excel, JSON, Microsoft Access, statistical files, Google Sheets, SAP, Snowflake, MySQL, multiple Amazon services, Dropbox, Microsoft SQL Server, and Web data connectors. The latter provides a way to connect to resources like Facebook and Twitter. Our experience at this time is limited to connecting Excel, CSV, Google Sheets, and our gate count software. Connecting to the first three were extremely simple, while the latter presented more difficulties because of problems in the gate count software.

The Oneonta Project

The following case will demonstrate what a business intelligence project looks like in real life. Today, most academic libraries are spending a significant amount of money on electronic resources. In the meantime, many vendors increase the subscription cost of their products every year. Therefore, for the purpose of budget optimization, it will be helpful if librarians have a clear statistical view about trends and patterns of those price changes. Visualized data will also help to support decision making in the library, or in the college administration. Finally, since statistics analysis and data visualization have been proven to be an efficient tool to help libraries manage their assets, including but not limited to electronic resources, it will be helpful for smaller libraries to practice methodology and techniques for such operations. Therefore, librarians from different SUNY institutions conducted a research project targeting price-changing trends of electronic resources through statistics and data visualization.

More than 80 aggregating databases with consistent payment records from 2010 to 2015 were selected for the project. In order to keep all data consistent and relative, no electronic serials, one-time purchases, or annual access fees were included. A student intern from the Department of Statistics at SUNY Oneonta was hired to process the data. Windows Excel and IBM SPSS were used as tools for statistics analysis and data visualization. After input of cleaned payment expenditure data from the ILS into Excel, graphs of average and median cost of selected titles, as well as the percent increase over the years were created. The visualized data indicate there is a consistent price increase since 2011, and the price increases in 2014 and 2015 were fierce.

IBM SPSS was used to compare prices and usages of selected resources as well as provide predictions

for future price changes, which was evaluated by their correlation to existing expenditure statistics from 2010 to 2015. The visualized data suggests resources with higher usage tend to cost more, probably because to vendors high usage indicates high demand. Meanwhile, the projections suggest that in 2017 and 2018 the price change may be relatively moderate, which is likely a sign that this price increase cycle is approaching its cap. It is important to note that the model for future price prediction is based on the expenditure data from the previous years, and that value varies based on vendor. Therefore the further into the future the prediction, the less accurate it is. It is also important to note that compared to the enormous size of the aggregating databases market with numerous products available, this project covered only a very small segment.

Through the visualized data, it is obvious that the pricing trend has been influenced by the national economy trend—the high percentage of price increase during 2014–2015 correlated to the steep value increase of the U.S. dollar at the same period, and the relatively moderate price increase of electronic resources since 2015 could also be related to the relatively moderate value change of the U.S. dollar.

However, most major vendors do not publish detailed revenue reports since they are privately owned. Hence there is a lack of information to determine causes for price changes other than the national economic picture and the product demand.

Reference

Baker, P. (2017, May 26). The best self-service business intelligence (BI) tools of 2017. *PC Magazine*. Accessed June 1, 2017. <http://www.pcmag.com/article2/0,2817,2491954,00.asp>

In conclusion, this project provided a visualized trend of electronic resources price changes, as well as a prediction of future prices. It also provided an opportunity to practice data visualization and statistics analysis in small academic libraries without any additional cost. Overall, this project was a very beneficial exercise, and provided a view into the price-changing trends of electronic resources.

Summary

Statistical analysis and data visualization is critical when communicating vision and decision data to administration and stakeholders. Increasingly, libraries will turn to a BI tool to assist with analyzing and communicating insights about their data. Each library, however, presents unique situations, especially in terms of how much an institution can invest in such a tool, but also in the demand for the number of visualizations, dashboards, and reports needed. Of the BI tools reviewed here, Tableau offers a better product if an institution desires a tool that can handle virtually any need and present sophisticated visualizations—but it comes at the highest cost. If institutions seek a tool just for electronic resources, Redlink offers an excellent choice. Excel provides a strong middle ground approach. It can handle a wide variety of data types, execute sophisticated calculations, and deliver aesthetically pleasing visualizations better than ever. Its significant drawback rests in its lack of a convenient, cloud-based sharing mechanism, yet Excel still provides strong benefits to its users.