

Collection Development and Data Visualization: How Interactive Graphic Displays Are Transforming Collection Development Decisions

Paulina Borrego
University of Massachusetts Amherst, pborrego@library.umass.edu

Rachel Lewellen
University of Massachusetts Amherst, rlewellen@library.umass.edu

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



Part of the [Library and Information Science 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>.

Paulina Borrego and Rachel Lewellen, "Collection Development and Data Visualization: How Interactive Graphic Displays Are Transforming Collection Development Decisions" (2014). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284315637>

Collection Development and Data Visualization: How Interactive Graphic Displays Are Transforming Collection Development Decisions

Paulina Borrego, Science and Engineering Librarian, University of Massachusetts Amherst

Rachel Lewellen, Assessment Librarian, University of Massachusetts Amherst

Abstract

Given the changing collection management landscape a clear tool for evaluating purchase decisions is needed to help selectors make the most of budget allocations. The UMass Amherst Libraries uses the business intelligence software Tableau to help selectors more clearly see the connection between monograph purchases and circulation data. Using dashboards, subject selectors can see the impact of monograph selections within a discipline, across the collection, and over a period of time. Graphic visualizations are easier to understand than previously used text and numerical based spreadsheets for data analysis and facilitate exploration at different levels. This paper discusses how data visualizations are used to effectively communicate monograph purchasing, circulation, and expenditures. Interactive dashboards help transform abstract ideas into a solid holistic understanding of the collections and in turn provide a common language to facilitate collection development discussions, decisions, and policies.

Introduction

6	Budget Code	Order Group	Pct	Cost	Nbr	Pct	Cost	Nbr	Pct	Cost
2202	SELECTOR PAULINA BORREGO-2009	CompSci	24%	\$534.66	8	89%	\$491.91	1	11%	\$42.75
2203	SELECTOR PAULINA BORREGO-2009	Geoscience	50%	\$83.19	1	100%	\$83.19	0	0%	\$0.00
2204	SELECTOR PAULINA BORREGO-2009	MathStat	17%	\$285.81	3	50%	\$146.60	3	50%	\$139.21
2205	SELECTOR PAULINA BORREGO-2009	Microbio	9%	\$126.65	1	100%	\$126.65	0	0%	\$0.00
2206	SELECTOR PAULINA BORREGO-2009	Physics	15%	\$334.95	2	50%	\$54.45	2	50%	\$280.50
2207	SELECTOR PAULINA BORREGO-2009	Total	24%	\$2,876.00	27	64%	\$1,773.83	15	36%	\$1,102.17
2208	SELECTOR PAULINA BORREGO-2010	Biology	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2209	SELECTOR PAULINA BORREGO-2010	Biology	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2210	SELECTOR PAULINA BORREGO-2010	ChemEng	33%	\$99.95	0	0%	\$0.00	1	100%	\$99.95
2211	SELECTOR PAULINA BORREGO-2010	Chemistry	27%	\$328.74	3	100%	\$328.74	0	0%	\$0.00
2212	SELECTOR PAULINA BORREGO-2010	CompSci	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2213	SELECTOR PAULINA BORREGO-2010	CompSci	9%	\$28.49	1	100%	\$28.49	0	0%	\$0.00
2214	SELECTOR PAULINA BORREGO-2010	Geoscience	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2215	SELECTOR PAULINA BORREGO-2010	MathStat	16%	\$145.29	1	33%	\$47.49	2	67%	\$97.80
2216	SELECTOR PAULINA BORREGO-2010	Microbio	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2217	SELECTOR PAULINA BORREGO-2010	Physics	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2218	SELECTOR PAULINA BORREGO-2010	Physics	27%	\$211.74	3	75%	\$156.64	1	25%	\$55.10
2219	SELECTOR PAULINA BORREGO-2010	Total	17%	\$814.21	8	67%	\$561.36	4	33%	\$252.85
2220	SELECTOR PAULINA BORREGO-2011	Astronomy	44%	\$674.35	6	38%	\$210.18	10	63%	\$464.17
2221	SELECTOR PAULINA BORREGO-2011	ChemEng	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2222	SELECTOR PAULINA BORREGO-2011	ChemEng	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2223	SELECTOR PAULINA BORREGO-2011	Chemistry	22%	\$121.30	2	100%	\$121.30	0	0%	\$0.00
2224	SELECTOR PAULINA BORREGO-2011	Chemistry	40%	\$3,179.88	11	35%	\$1,092.88	20	65%	\$2,087.00
2225	SELECTOR PAULINA BORREGO-2011	CompSci	18%	\$160.38	2	100%	\$160.38	0	0%	\$0.00
2226	SELECTOR PAULINA BORREGO-2011	CompSci	8%	\$296.96	0	0%	\$0.00	4	100%	\$296.96
2227	SELECTOR PAULINA BORREGO-2011	Education	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2228	SELECTOR PAULINA BORREGO-2011	ElecEng	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2229	SELECTOR PAULINA BORREGO-2011	ElecEng	50%	\$211.83	0	0%	\$0.00	2	100%	\$211.83
2230	SELECTOR PAULINA BORREGO-2011	Engin, Gen	100%	\$154.83	1	100%	\$154.83	0	0%	\$0.00
2231	SELECTOR PAULINA BORREGO-2011	MathStat	23%	\$176.29	3	100%	\$176.29	0	0%	\$0.00
2232	SELECTOR PAULINA BORREGO-2011	MathStat	11%	\$864.31	8	57%	\$397.25	6	43%	\$467.06
2233	SELECTOR PAULINA BORREGO-2011	Physics	22%	\$51.89	2	100%	\$51.89	0	0%	\$0.00
2234	SELECTOR PAULINA BORREGO-2011	Physics	24%	\$974.68	9	50%	\$528.09	9	50%	\$446.59
2235	SELECTOR PAULINA BORREGO-2011	PolymerSci	0%	\$0.00	0	0%	\$0.00	0	0%	\$0.00
2236	SELECTOR PAULINA BORREGO-2011	Total	22%	\$6,866.70	44	46%	\$2,893.09	51	54%	\$3,973.61
2237	SELECTOR PAULINA BORREGO-2012	Astronomy	17%	\$22.50	1	100%	\$22.50	0	0%	\$0.00
2238	SELECTOR PAULINA BORREGO-2012	Biology	9%	\$355.94	1	33%	\$30.88	2	67%	\$325.06
2239	SELECTOR PAULINA BORREGO-2012	Chemistry	68%	\$2,101.51	1	8%	\$169.13	12	92%	\$1,932.38

Figure 1. Sample Microsoft Excel spreadsheet of monograph purchase data provided to selectors.

Librarians
have long

sought to understand the extent to which monograph purchases meet the needs of users. Circulation is one indicator that can be captured and analyzed. This paper describes how the UMass Amherst Libraries use the business intelligence software Tableau to help librarians review monograph purchases with circulation data and expenditures to inform purchasing decisions.

Challenges of Spreadsheets

Before using dashboards, library collection data was typically shared with selectors in a Microsoft Excel spreadsheet (Figure 1). It was then up to the individual selector to manipulate the data, trying to review multiple worksheets, different data points and ultimately make sense of, and glean, trends. One of the problems with this practice is that not all library personnel are skilled with Microsoft Excel manipulation features or feel comfortable dealing with the sometimes massive spreadsheets that are created and shared. While this was very useful data, and regardless of Excel manipulation skills, it is still hard to make connections and see trends in this text and numerical based environment.

The next evolution in spreadsheet use was to provide data with subtotaling, filtering, and some of the other Excel features employed. These enhancements were greatly appreciated and made it easier for some to use the data but it still did not address the issue of discerning trends within the large amounts of data. It was still difficult to view data over time since each fiscal year was in a separate worksheet.

Advantages of Data Visualization

First introduced to Tableau at the 2011 Charleston Conference, the UMass Amherst Libraries use Tableau Desktop, Tableau Server, and Tableau Public to analyze, interact with, and communicate collection data as well as other library data. Positive aspects of Tableau products for the Libraries include an intuitive web interface that makes it easy for staff to absorb and interpret data. The ability to connect to and at times blend multiple data sources provides context and shows relationships between data sources. For example, Figure 2 shows walk-in traffic, proxy activity, and circulation counts within a single dashboard.

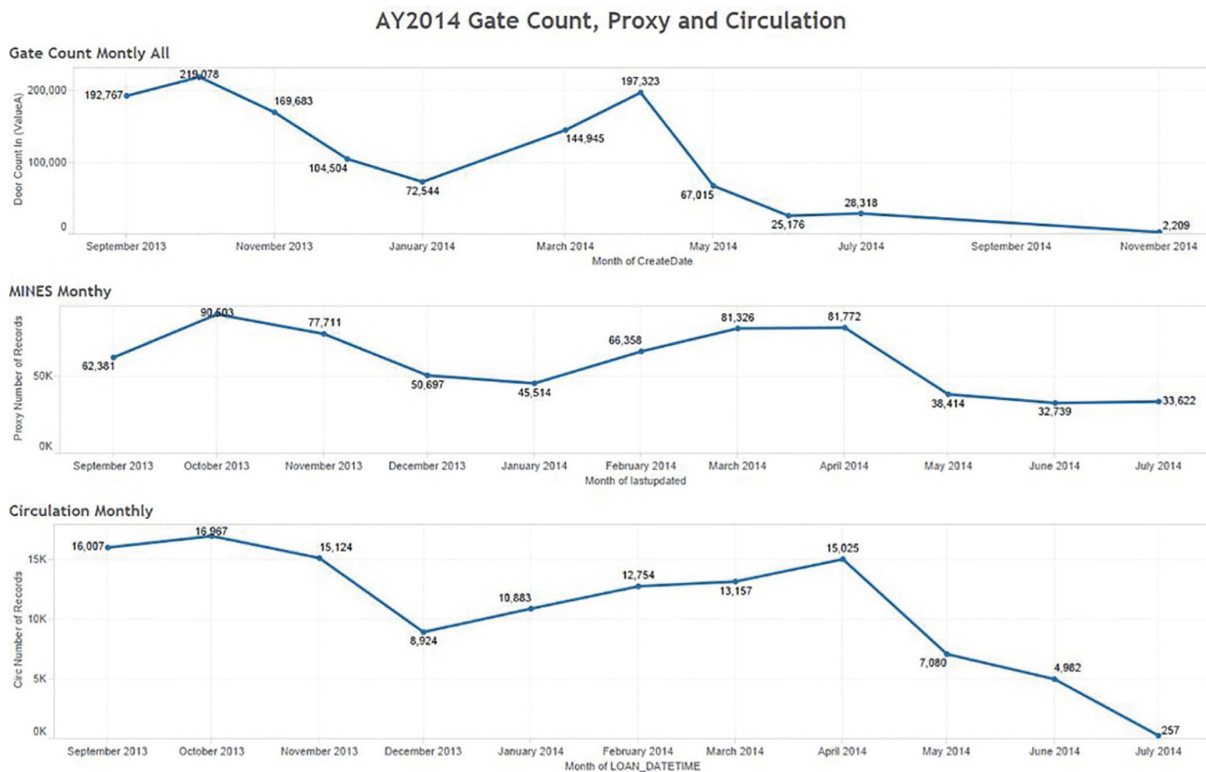


Figure 2. Gate count, proxy, and circulation dashboard.

Live interaction with real-time data and automatic updating is also beneficial. Multiple chart and graph options (bar, line, pie, map scatter plot, Gantt, Bubble, etc.) support a variety of analysis needs and learning styles. Options for private and public dissemination ensure protection of sensitive data and facilitate sharing of unrestricted data.

Using Tableau has increased the capacity for staff to use and understand data across many data sources these include: monographs, e-books, proxy traffic, gate counts, service desk interactions, instruction, space use, ARL statistics and more.

One Selectors' Collection Practice and Awakening with Tableau—Paulina Borrego

Soon after joining the staff at the UMass Amherst Science and Engineering Library Paulina Borrego was given the responsibility of selecting monographs for her various liaison departments. Being a new librarian, following a career change from teaching high school chemistry for nearly twenty years, she was unfamiliar with many library practices and had much to learn. She took her role as monograph selector seriously and tried to make the best use of the allocated budget for each of her liaison departments.

Monograph allocations were based on a number of factors including enrollment of undergraduates and graduate students as well as faculty FTE (full-time equivalent). A budget was assigned to each department and that amount was used to support the research and instruction needs of that department. The collection development policy outlines materials that are out of the scope of purchase, such as textbooks. The UMass Amherst Libraries is also part of the Five College Consortium with a shared online catalog striving to avoid unnecessary duplication of materials across the consortium.

As a selector Ms. Borrego tried to choose monographs that aligned with the faculty research area in her departments along with the overall mission of the University. She spent a great deal of time examining faculty webpages for current research interests so that she could make

monograph purchase decisions that supported those interests. This exercise of researching and noting faculty research areas was very time consuming and in the end did not guarantee that a monograph purchase decision would prove helpful or ultimately circulate. She also spent considerable time reviewing new title lists, book reviews, and current science trends to better inform her monograph purchase decisions. In hindsight, she felt she was somewhat naïve in thinking that she could accurately predict those items that would be of interest or circulate.

After her initial year of purchases she received the annual spreadsheets with circulation and duplication data. She tried to the best of her ability to deal with the massive spreadsheets to extract information dealing with each of her liaison departments, hoping to find trends and evidence to support purchasing decisions. Oftentimes she felt her Excel skills were lacking and she needed help to sort and filter the massive spreadsheets effectively. She was also frustrated with the formatting, finding that data points were hard to associate with one another and make connections. Overall, lacking the necessary computer skills to manipulate the data, along with the inability to visualize data trends given the format of the data, she felt disadvantaged in making successful selector decisions. What she wanted was indisputable evidence to make wiser collection development decisions.

Various Excel enhancements were of some help, providing an easier way to sort and filter the monograph use data, but still lacked the ability to visualize impact of seeing overall trends. Without data in a format she could reasonably use, Ms. Borrego continued to make monograph purchase decisions based on faculty research and instruction areas. Her hope was that the items purchased would satisfy the needs of the community and ultimately circulate. For some time she continued this individual collection development practice and tried to evaluate the circulation data available to make small adjustments as needed. She was mindful of trying to make informed decisions and to use her monograph allocation budget with a sense of fiscal responsibility.

Then Came Tableau, and That Changed Everything!

The same Excel spreadsheet data, previously shared with selectors, was used to create multiple dashboards and visualizations with Tableau. Selectors could now choose from multiple variables to create custom views.

When Ms. Borrego used the interactive dashboards to explore her data she found it both horrifying and freeing at the same time. She visually interacted with the data in a format that was fully understandable. Whereas before she had tried to make connections between the data points (monograph purchase decisions and circulation data) in the Excel spreadsheets, interactive dashboards presented the data in a form that was clear, simple, and undeniable.

The Order Group Dashboard (Figure 3) allowed selectors to choose the groups that correspond to their liaison departments. There is also the option to view a single year or multiple years. Horizontal bars showed monograph purchase data clearly delineated for circulated and noncirculated items. Expenditure totals along with percentages are tabulated and displayed alongside each bar. The Non Circ with Duplication Status display indicates information for monograph purchases either duplicated or unique as part of the Five College Consortium. Other dashboards allow selectors to drill down and view title level data, make comparisons to other order groups, and see trends for parts or the collection as a whole. Since the dashboards can display data for all selectors and order groups, transparency is built into the system allowing selectors to draw conclusions on many levels.

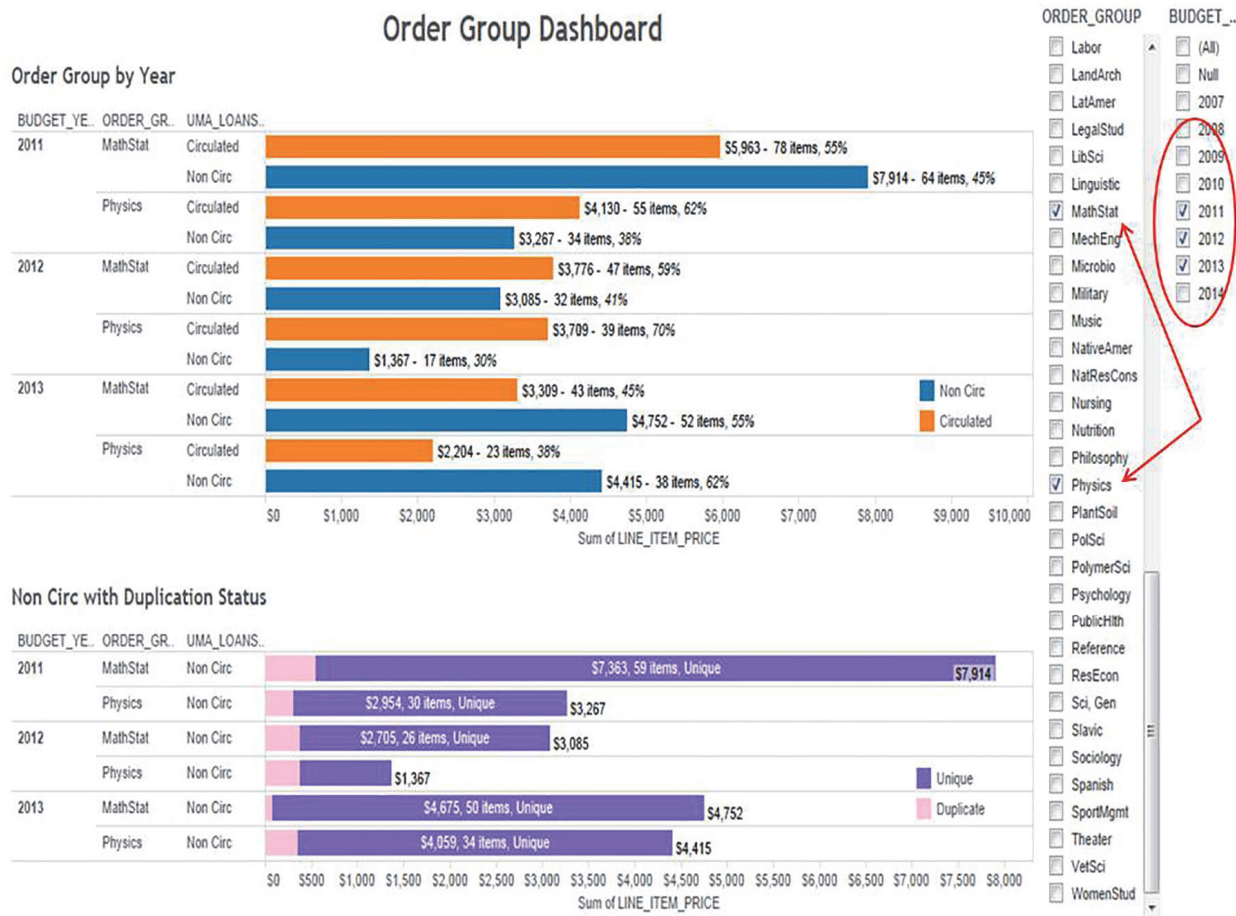


Figure 3. Sample order group dashboard of monographic purchase data in tableau format.

In dashboard form Ms. Borrego could quickly and easily see where monograph purchase decisions, no matter how informed, were not circulating. She also came to understand that in the interest of nonduplication within the Five College Consortium, she was selecting "unique" items but they were not circulating (Figure 4). Overall, she was horrified to see that the monograph titles she put so much work and effort into selecting based on research and instruction areas were not being used. Data visualization was fundamental in helping her to make this realization—clear, undeniable evidence that her monograph collection practices were not as effective as expected.

On a larger level seeing the data so clearly caused Ms. Borrego to question the practice of individually selecting titles for potential use and she immediately changed her personal collection development practice to purchasing only those monograph titles directly requested. Interacting with the data in Tableau provided the information in a format that was directly accessible and straightforward. It shaped her decision to change her approach to selection. Being able to make such a clear decision based on the data was very freeing. She now had the evidence she needed to change her course of action.

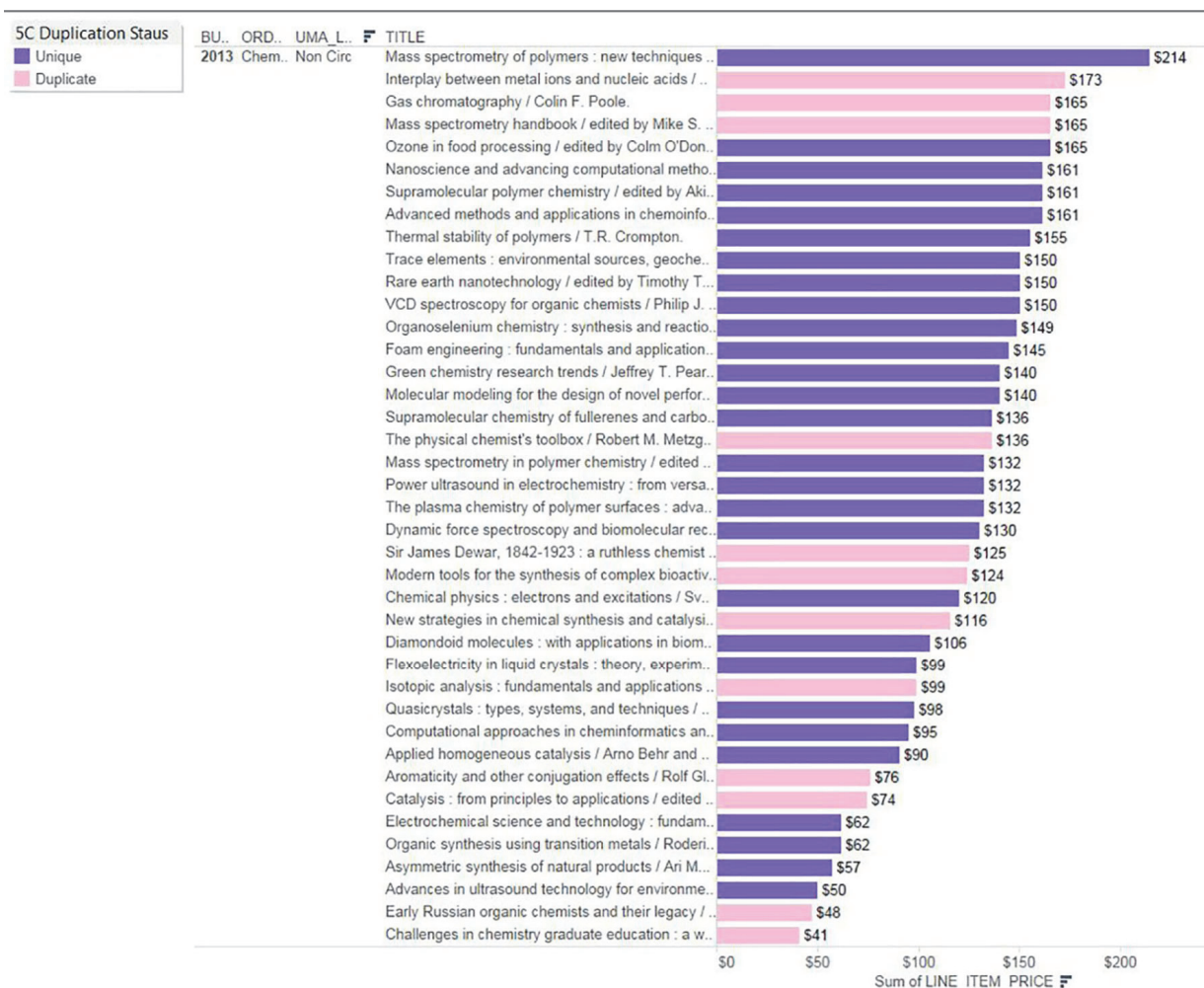


Figure 4. Titles purchased with circulation and consortial duplication status.

Visualizations and Dashboards Facilitate a Shared Perspective

Visualizations have improved how data is communicated, established a common language for discussion and contributed to decision making. Due to the transparent nature of the dashboards, selectors are free to explore their individual

monograph purchase data as well as make comparisons at other levels. Visual displays highlighted comparisons between library purchase programs such as approval plans, Books on Demand (BODEM), *New York Times* (NYT), and monographs purchased for reserves (Figure 5). These displays facilitated discussion and decision making. For example, approval plans and BODEM purchases were expanded.

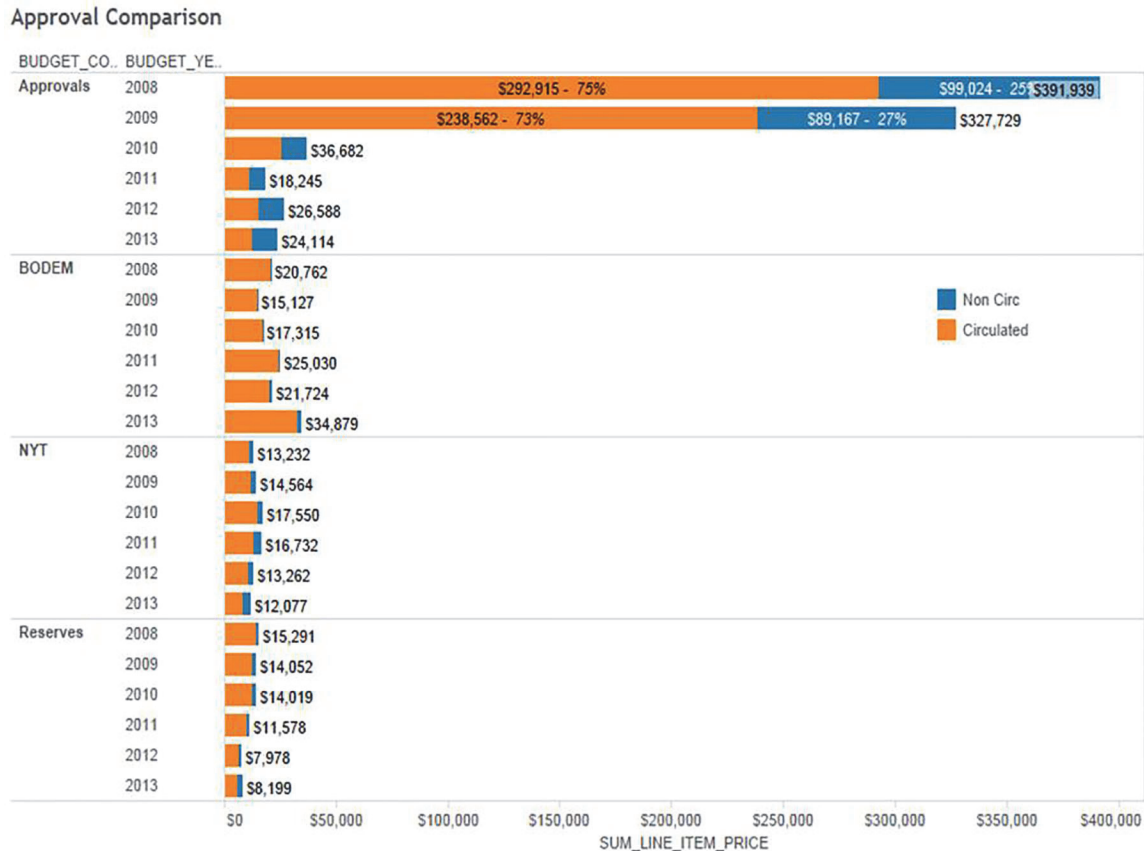


Figure 5. Tableau dashboard for various library purchase plans.

Visualizations allow selectors to drill down to view their individual order group purchase data as well as view the collection data on a larger Library of Congress (LC) scale (Figure 6). This ability to drill down as well as see the entire collection helps selectors understand their individual role in building the monograph collection as a whole. Being able to view one's individual data as part of the collective picture helps to establish a sense of

community and create a common understanding of the entire collection. Having an understanding of an individual role in the entire monograph collection is vital when budgeting and allocation methods are reviewed. Overall, these displays have helped to provide a shared assessment of the monograph collection so that discussions and decisions can be based on a common perspective and more easily interpreted data.

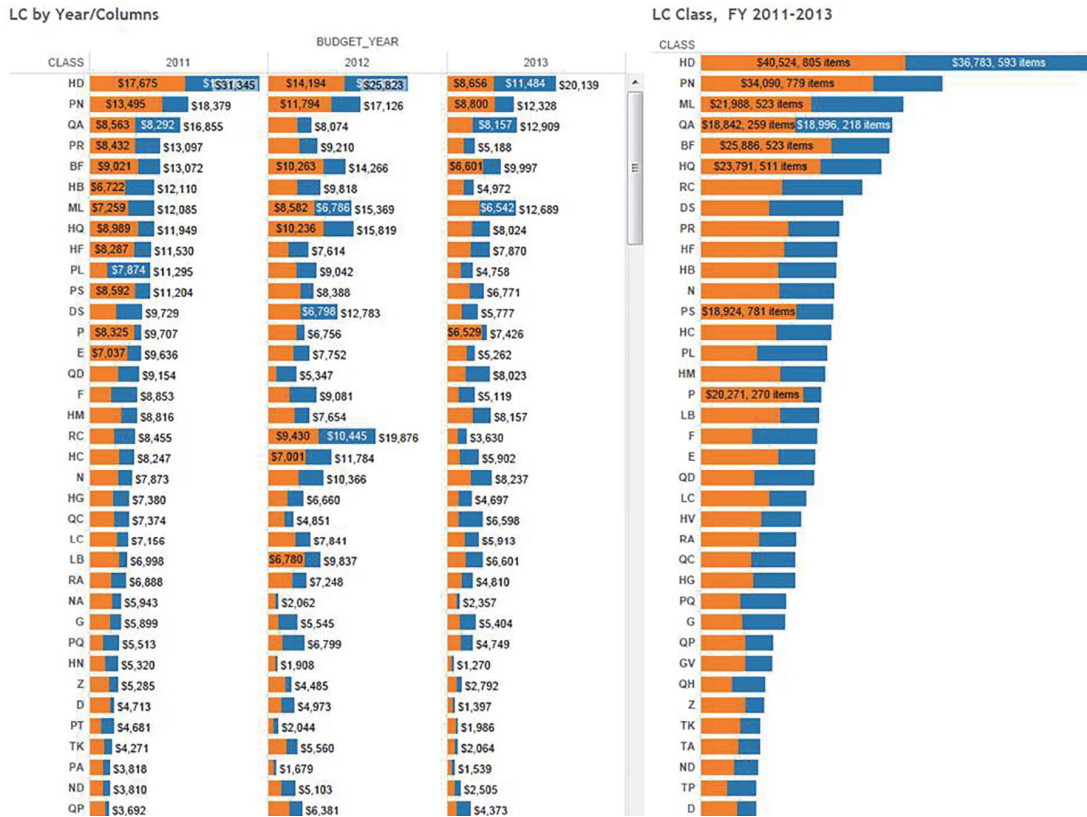


Figure 6. Tableau dashboard for monograph purchases by Library of Congress.

Data Visualizations as an Effective Communication Tool With Library Stakeholders

In the same way dashboards have helped to inform and educate library staff, they to help tell the Library story to stakeholders. Across campus, dashboards can be used on many levels, for example with faculty and deans, to effectively

communicate relevant library data. Using dashboards it is easier for faculty to see how book purchasing, spending, and use relate to each school and college (Figure 7). This mode of communication provides information that may be unfamiliar in a clear, understandable, and visually powerful format. Expenditure dashboards demonstrate transparency and fiscal responsibility.

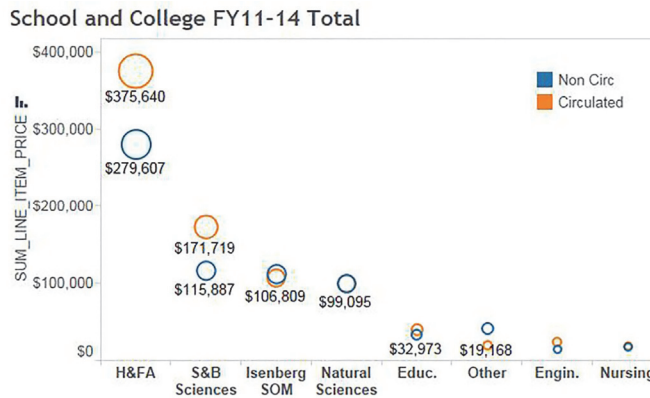


Figure 7. Spending and circulation by school and college.

Conclusion

Data visualizations enabled selectors to review monograph purchases, track and assess spending, and communicate effectively with library staff and campus stakeholders. These assessments impacted collection practices and are changing

how collections are analyzed. The UMass Amherst Libraries use Tableau as the preferred way of interaction with data and it is a major component of the Libraries' assessment program. This has increased the capacity for staff to interact with, understand, and communicate data.