Charleston Library Conference

Collection Dashboards for Selectors

Lindsay A. Cronk
University of Houston, lacronk@uh.edu

Wenli Gao
University of Houston, wgao5@uh.edu

Collection Dashboards for Selectors

Lindsay A. Cronk, Coordinator of Online Resources and Collections, University of Houston

Wenli Gao, Communication, Sociology, and Anthropology Librarian, University of Houston

Abstract

While collections dashboards are often used as an external communications tool, they have additional applications for improving internal processes and assisting subject selectors. The value of collection visualization for analytics and strategy cannot be underestimated as visualizations can help clarify complex information to improve decision-making. This paper summarizes the current efforts to deploy collection dashboards at the University of Houston libraries. Using Tableau to parse and visualize collections data, the library is embedding visualization frameworks into the acquisitions calendar to enhance selection and deselection processes.

Introduction: "Beautiful Evidence"

Of the act of visualizing data, Edward Tufte, considered the foremost leader in the movement for information design, offered (1990), "Confusion and clutter are failures of design, not attributes of information," (p. 51). This phrase becomes all the more meaningful and relevant when one turns to look at a COUNTER usage report. Tufte describes the act of visualization as one of creating "beautiful evidence" that serves to deliver to the intended audience both simple design and intense content, essentially a clear representation of the content of a dataset.

In library decision-making, we have long struggled with a surplus of evidence sources in a variety of formats. From circulation statistics to usage statistics, gate counts, budget information, inflation projections, title lists, and coverage information, we generate and receive a startling amount of quantitative data that must be factored into decision-making. Beyond these information sources, we must also work to ensure that we do not lose the qualitative data provided by subject experts, faculty advocates, students, and other users. It is, frankly, the challenge of attempting to simplify this glut of inputs that drives many collection librarians to roll their eyes at the phrase "data-driven decision-making."

At this point, there is no simple means of combining these disparate data sources for review. However, by beginning to develop frameworks that incorporate some of these data points, collections librarians may find new ways to engage selectors and other collection stakeholders in the work of assessing and managing the library collection.

With the understanding that finding ways to consolidate and standardize this data is a great undertaking, the central argument of this proceeding and its presentation is that it is an effort we must start and improve, rather than one we can put off until a vendor builds a proprietary platform or system that serves as a half-measure fix.

Ultimately, to paraphrase Nate Silver (2012), "Before we ask more of our data, we must ask more of ourselves," (p. 9). Only by critically reviewing and assessing our current collection and management practices can we ensure that we are holding ourselves to the same rigorous standards we espouse in our training and efforts in data services. If we do not begin to explore our own data with the same scholarly technique our researchers apply to their data, we may find that we are hypocritical in positioning ourselves as information experts.

Certainly, library data presents its own set of challenges, particularly vendor-provided usage data, but only by using and parsing the data will we be able to identify its failings and correct our standards. How we leverage our common need with vendors, and how we share our solutions with one another will prove the critical efforts to improving data quality. That collections librarians are all toiling separately on the same problem should be a synergy we harness for a shared solution.

This paper focuses on the University of Houston's approach to creating subject area group dashboards, breaking out the e-resource collection in three major areas: Humanities, sciences, and social sciences. Focusing on practical approaches, we seek to describe a method that has successfully integrated

visualization with a monthly database review calendar synchronized with the vendor renewals cycle.

Visualization for Collection Assessment

In 2009, Jennifer Z. McClure noted that "Collection assessment, at its best, is an art not a science, and the numbers that it generates are a means, not an end," (p. 79). The sentiment of this quote has guided the work of visualization for selectors at the University of Houston. The effort to visualize collections data is interpretive and iterative, meant to empower selector decision-making, not to dictate it.

Visualizations offer opportunity to embed nuance and context in ways spreadsheets simply cannot. Using visualization tools to better know our collections and better analyze past collection decision has near limitless potential to improve the delivery of collections as a service. This is not a hyperbolic statement. By beginning to truly analyze usage data and connect it to other data points, we may finally find means of linking collections to learning outcomes. This, many would argue, is the grail most collections librarians in academic libraries seek.

However, this paper and the presentation it describes are only speculating in this area. That said, the only way to begin to find that process is through small project steps such as the ones described here.

As has been previously discussed, collections data is a natural area where we have vast and long-standing data that would benefit from analysis and visualization. Our collections data is a place for us to begin to understand the techniques and tools for visualization our researchers will employ and the common challenges that they will encounter.

Visualization can only enhance assessment efforts because, as if in answer to McClure's exhortation, it serves as a meeting place of art and science, where information becomes simultaneously more useful and more beautiful.

Visualization at the University of Houston Libraries

In identifying data sources, the library's collection management committee has committed to reviewing three central areas: usage, cost, and content. In order to develop a sustainable practice that clearly integrates these data sources, the way in which we enter and review data has fundamentally transformed. The critical piece of this effort has been the elimination of previous formatting of spreadsheets of information.

To clarify, a spreadsheet that has been created to be presented as a visual document is often the antithesis of what is required for it to be ready for ingestion by any visualization or analysis software. Color coding, headings and subheadings, wide horizontal orientations, all of these common organizational methods within spreadsheets make the data content less usable for analysis, review, and visualization.

Additionally, when we cling to spreadsheets, and more specifically to Microsoft Excel as a visualization tool, we do a disservice to our data. We also fail in our mission as information professionals. Many libraries are moving to provide data services and data management to faculty. We will only begin to understand researcher data needs when we begin to treat our own collected data as worthy of analysis and visualization.

Other large university libraries may find, as we did at UH, that reviewing a full slate of licensed resources, which often number in the hundreds or thousands, can represent a logistical challenge. Scaling the project and considering pre-existing frameworks becomes an essential part of planning a visualization project.

Ultimately, the success of the pilot project for selector visualizations hinged on integrating the visualizations into existing workflows, mainly the acquisitions calendar (a natural review cycle) and monthly meetings of the Collection Management Committee (CMC), the body which serves to oversee collections at the University of Houston.

By beginning with an integrated process rather than an additional free standing system, the visualizations were given a better opportunity to succeed as an assessment tool. As a part of an ongoing database review system, the integration helped to inform a manageable and meaningful process that lines up usage trends, cost trends, and other important evaluation factors.

In working in conjunction with selectors, university collections stakeholders developed a sense of what drivers would influence decision-making, including not just cost or use but cost trends (ever upwards) and use trends (less obvious). By looking at cost and use over time, the selectors were able to see that some e-resources should be watched more closely, while others were of high importance to users.

Rather than creating subject-level dashboards that would create a high volume of smaller scale visualizations, the practitioners opted to develop subject area dashboards in the humanities, sciences, and social sciences. This approach was informed

both by a desire to create a sustainable practice and by the structure of the library's liaison services department, which is divided along those lines.

Ultimately, the result of focusing on e-resources by subject area was three dashboards featuring four central visualizations. The visualizations, which can be seen below in Figure 1, highlight expenditure information alongside usage information. Also highlighted are resources that are COUNTER compliant versus those that are not.

The dashboards were developed within Tableau, making ongoing updates as simple as updating the source spreadsheets with the new data lines.
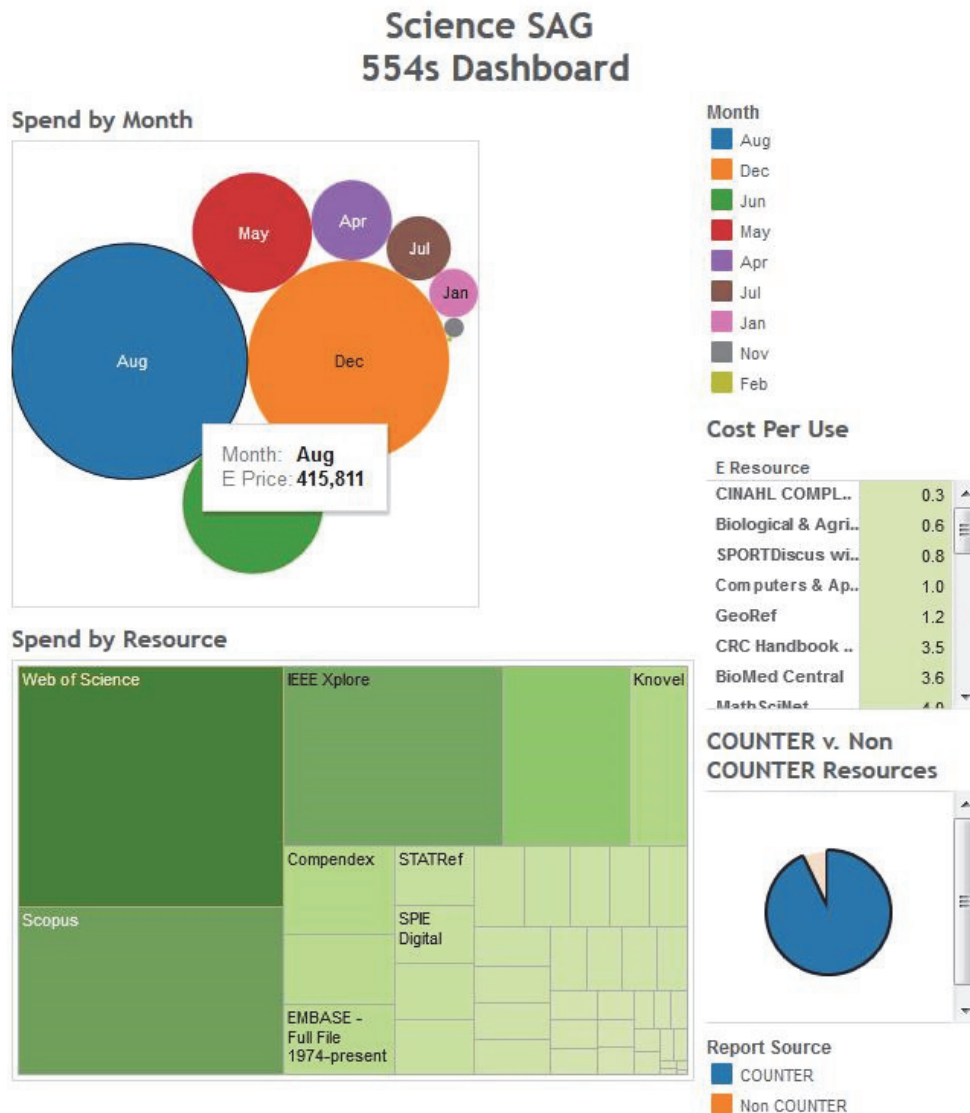

Figure 1. Science SAG 554s dashboard.

Selector Use for Visualizations

The interactive dashboard gives subject selectors a tool for a quick understanding of the expenditures by month and by key resources. It is easily identifiable that August and December are the two months when most of the renewals are paid. In the science subject area group, Web of Science, Scopus, and IEEE Xplore are the three most expensive databases. The interactive feature of the dashboard gives selectors the ability to narrow down by month. Since resources are reviewed by month according to their renewal period, this function helps selectors focus their attention to the resources they need to review for the month. Figure 2 is an example of the resources paid for the month of April. Selectors can focus their attention on the most expensive resources for that month, Petroleum Abstracts TULSA Database, SPORTDiscus with Full Text, and other medical-related databases. In addition to looking at the expenditures, the dashboard also integrates usage information and calculates cost per use. Selectors can see ranked cost per use and factor this variable into their decision-making.

However, not all resources are COUNTER-compliant. In order to recognize this, the dashboard also includes an option for selectors to distinguish resources by their COUNTER-compliant status. Overall, the dashboard frees selectors from digging into multiple spread sheets and filtering out the information they need. It is a quick and easy way for selectors to understand their resources, thus helping them focus their attentions based on the renewal cycle, and make evidence-based decisions.

Conclusions

As we continue to explore the potential for visualizations for collection assessment and strategy, the University of Houston is putting a key priority on a few areas for review and redesign moving forward. In particular, we will turn our attention to enhancing the current subject area dashboards, creating

subject level dashboards in high-needs areas, and most importantly reviewing and revising our data collection processes to find automation opportunities.

Making the process sustainable requires that we devote time and resources to a full review of our data collection for library collections. Of particular interest moving forward are opportunities to collect and quantify selector feedback and content-based information about databases and eresources.

Continuous improvement and integration will continue to inform and improve our collection visualizations. This incremental and iterative approach provides both the impetus for ongoing analysis and the framework to insure adoption. Ultimately, the fervent desire of those involved with this project is the establishment of a more robust and community-driven approach to collection visualization. Sharing our example and developing community frameworks is a next step that the authors are undertaking and facilitating.

Such a community of practice would be a powerful means of overcoming one of the central challenges for collections practitioners: Lack of resources and time for creating institution-specific solutions, which may be challenging to replicate. A community of practice, focused on developing, sharing, and ensuring the sustainability of best practices for collection visualization could inform a more balanced approach to looking at collections. Such a broad-scale collaboration would benefit all those involved and save the time of many a harried collection practitioner already in the midst of other important projects.

Sharing this small project will, the authors hope, serve as a conversation starter for the collections community. In reviewing one process of subject area dashboard creation and use, others may find a solution or tool that may prove useful.
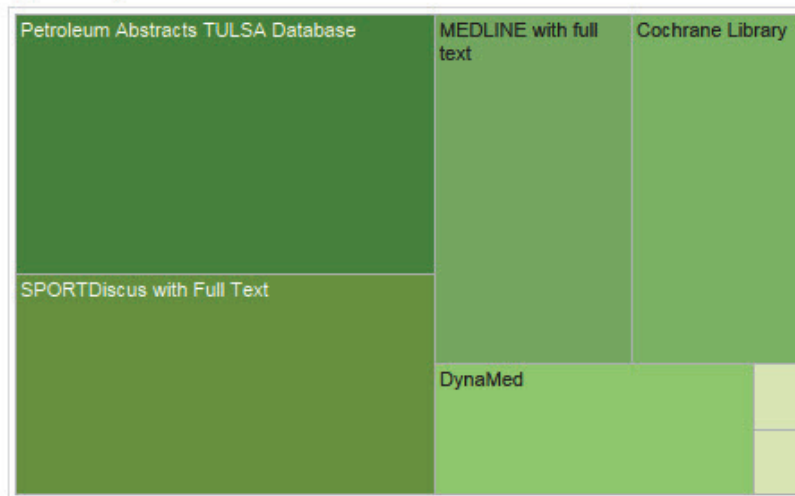
Spend by Month
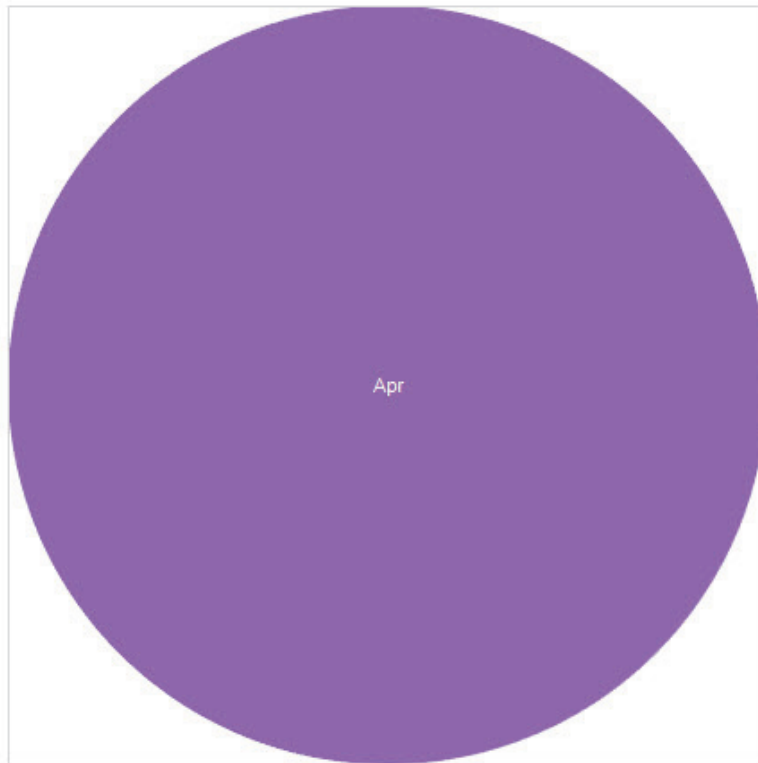
Spend by Resource

Figure 2. Using filters in science SAG dashboard.

References

McClure, J. Z. (2009). Collection assessment through WorldCat. *Collection Management, 34*(2), 79-93.

Silver, N. (2012). *The signal and the noise: Why so many predictions fail--But some don't*. New York: Penguin Press.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.