Supporting Library Departments through Data Visualization sought to demonstrate the power of data visualization in supporting library departments’ potential to leverage internal data and improve communication and decision-making processes. In collaboration with the User Experience and Research and Instruction department at Davis Library – UNC Chapel Hill and the Southern Historical Collection Wilson Library – UNC Chapel Hill, I developed a series of visualizations after conducting a needs assessment of department. For Davis Library, I developed a series of visualizations analyzing user engagement with UNC Library LibGuides in effort to support their annual content review and for Wilson Library, I developed a series of visualizations to support digitization planning efforts.

Headings:

Data Visualization
Assessment
Digitization
LibGuides
User Experience
Data analysis
SUPPORTING LIBRARY DEPARTMENTS THROUGH DATA VISUALIZATION

by
Natalia Lopez

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
April 2017

Approved by

_______________________________________
Lorin Bruckner
Contents
SUPPORTING LIBRARY DEPARTMENTS THROUGH DATA VISUALIZATION... 0
1 Introduction ................................................................................................................................. 2
2 Visualizing LibGuide Data for the User Experience Department ............................... 3
  2.1.1 Needs Assessment............................................................................................................. 3
  2.1.2 Implementation Process .................................................................................................. 6
  2.1.3 Evaluation and Following Steps ....................................................................................... 9
3 Southern Historical Collection ................................................................................................. 12
  3.1 Needs Assessment ............................................................................................................... 12
  3.2 Implementation Process ...................................................................................................... 13
  3.3 Evaluation and Following Steps ......................................................................................... 15
4 Lessons Learned ....................................................................................................................... 18
1 Introduction
This project sought to support the work of two library departments in developing data visualizations that would enable them to analyze internal data and support their interdepartmental communications and/or decision making processes. Specifically, I partnered with the User Experience department and the Southern Historical Collection at the University of North Carolina Chapel Hill’s Libraries. I used Tableau Academic to create visualizations. With the support of my advisor, Lorin Bruckner, the Data Visualization Librarian at Davis Library UNC, I identified two groups that would have access to robust data sets. For each partner, I arranged a first meeting where we discussed general data collection and management practices, reviewed some of their data collection platforms, and possible data-related questions.
2 Visualizing LibGuide Data for the User Experience Department

2.1.1 Needs Assessment

The User Experience department collects data about user experience from different sources. Each of these reflects a different facet of the reference and instruction team’s work. For example, LibAnalytics is a Springshare software recently implemented within the last year or two and is used to track reference questions from patrons. Data is entered by desk staff, librarians in consultations and email reference questions. This enables administration to review the types of reference and consultation questions that the library is receiving and be able to better tailor their services and make decisions about resource allocation and staff. Likewise, chat reference questions are also captured in a database. Lastly, Google Analytics is a tool enabled on the library web pages that captures data about how users engage with the library website daily. The library website is expansive, encompassing information from all branches and departments of a Research 1 library institution. Data that is captured through Google Analytics includes page views, percentage of page scrolled, unique views, and time spent on web pages.

I met with Sarah Arnold, Instructional Technology Librarian; Arnold is one of the two key User Experience department staff. After a discussion of data management practices and needs, Arnold shared that the next major upcoming project for the User Experience department is an annual review of LibGuide data. LibGuides are web
tools librarians use to host resources and pertinent information for their patrons. They are often organized as either Course Guides or Subject Guides. This system is fairly recent, with librarians beginning to share information with their patrons online through basic HTML sites in the early 1990s. Today, LibGuides are a relatively prominent content management system utilized by many libraries, but poorly understood or assessed. At UNC libraries, both Sarah Arnold and Jacqueline Solis, Head of Research and Instruction, discussed a need for better assessment tools to design best practices and implement meaningful and effective standards. As LibGuides have grown from an organic desire to share resources, there were few centralized training or standards put in place.

As such, the department is working on distributing information to librarians about their guides, as well as collecting general data about how LibGuides are being used by patrons. In the past, the User Experience department has sent librarians individual reports with counts of unique views by page, average time spent on page, number of resources on each page, and number of broken links per page. An estimated time that it would take to update each page was also included. The department is currently exploring ways to better communicate this information to librarians, as well as supporting the promotion of best practices regarding LibGuides. While sending a personalized spreadsheet to each librarian was a positive step towards stronger assessment practices, I proposed using Tableau to better leverage the data provided by Google Analytics. The current spreadsheet system is primarily a tool that can only be used by individual librarians, rather than provide a macro overview that can then
be viewed at a more micro level. Tableau can provide some of this sophisticated functionality and may be able to guide department administrators in leading more centralized conversations.

**Fig 1.** All librarians receive a personalized LibGuide assessment document that captures data related to unique views, time on page, and number of resources on each page they own.

Specific questions that arose during our meeting include:

(a) **What is the most effective case use of LibGuides? Are they more effective as course pages or subject pages?** Librarians frequently create course pages to supplement their instruction sessions, tailoring their resources to that class. Simultaneously, they create a subject guide for general use. It is unclear which of the two is most used or most effective. Should librarians focus their efforts on creating expansive guides that can be used by a broader audience or should they expend their efforts on creating a tool for classes? Perhaps, the effectiveness of
this strategy is also impacted by the way the guide is used/shared to users. For example, if users are introduced to a subject guide at orientation and subsequent meetings with advisors, are they more likely to use the tool?

(b) **Which resources are patrons using the most? Where are they located within the page?** While the current spreadsheet-based LibGuide review process includes unique views data for main page and sub pages, it does not specifically highlight the use of each resource. This could be particularly telling of which resources are most appealing or in demand by patrons. It could also be revealing of the significance of placement/web design on the patron. For example, if patrons are primarily using the resources listed on the top third of their page, this may be indicative of web design practices.

(c) **How are LibGuides being used in conjunction with instruction or consultations?** LibGuide traffic may differ depending on the way in which it is used. Some LibGuides might have a lot of use during the actual instruction session and not be viewed at any other time. Perhaps, some are used prior to midterm and final projects. This could be useful information for librarians to be aware of in planning their instruction sessions.

### 2.1.2 Implementation Process

The project was conducted in a series of phases:

- **Data gathering:** During my meeting with Sarah Arnold, I reviewed the Google Analytics site. We set up an additional meeting to walk through tools to export the data. Google Analytics enables users to search for specific key
terms using regex and return a series of metrics. For the purposes of this project, I returned a list of sites with the following root: guides.unc.edu/ and metrics for time viewed, unique views, entrances, and exits. One significant limitation of Google Analytics is the inability to export data in bulk. Each sub metric (i.e., data on each specific resource or link) would have to be opened by selecting the parent page (i.e. homepage or a subpage of the LibGuide).

- **Sample data manipulation:** To begin exploring data visualizations that could be of use to the User Experience and Reference and Instruction departments, I began to pull a series of sample data from across the library. Included in the sample were course guides and subject guides from libraries across the University. I downloaded and cleaned data representing Kenan Science Library, Davis Library, and the Health Sciences Library. With that data, I developed visualizations based on the spreadsheet data already shared with me. In addition to looking at views of each parent page, I was able to include unique view data of each resource included on the specific page. I was also able to include data about monthly views for each LibGuide.

- **Mid-point meeting with stakeholder:** At this point of the project, I had met primarily with the User Experience department. Realizing that the Head of Reference and Instruction would be able to provide more context about the way LibGuides are currently implemented and assessed, I set up a meeting with Jacqueline Solis. In addition to discussions considerations reviewed above in the needs assessment portion of this paper, we discussed the draft visualizations, identifying the need for differentiating between course guides
and subject guides. Further, Solis confirmed that the department would benefit from looking at the use of guides throughout the semester and the need for better understanding the resources patrons are using.

- **Milestones revised:** Given the limitation of Google Analytics regarding downloading data in bulk, milestones for the project were revised. A sample of LibGuides from one department at one library branch would be used to develop a dashboard. Specifically, the Research and Instruction team at Davis Library was selected and a sample size of three LibGuides for half of the librarians in the department was deemed adequate. This would enable the creation of a sophisticated dashboard that can be reviewed by the Research and Instruction team to make general inferences and evaluate the applicability of implementing this system to their LibGuide review process.

- **Data cleaning and preparation:** Firstly, narrowing down the sample to a specific set of librarians required pulling the links of each libguide from the librarian’s homepage. I then had to search for each link on Google Analytics and I had to manually download each data set for the parent pages, and download data about the resources on each page. This often meant if one LibGuide had three or four subpages, I would have to download five to six different data sets for each one. Once all data sets were downloaded, I merged all data sets under one spreadsheet and added data about the librarian involved, the type of libguide, added the master guide title, and denoted whether each row was a Resource or Main Page. If done in bulk, this project would benefit from including a coding component that might involve
webscraping as well as pulling data from an API. For example, identifying which librarian was the creator of a libguide could be done after creating a web scraping script that went to each libguide link (provided by an API) and finding the librarian name on each page. Initial exploration into this question suggests that it would be possible using class names included in the HTML of each LibGuide.

- **Final meeting with stakeholder:** Review of final product. See below for detailed description.

2.1.3 Evaluation and Following Steps

As part of the final meeting with the department stakeholder, I reviewed the current views of the visualization. There are four primary views that are all interconnected with one another through filters:

- **Visualization 1:** illustrates what number of unique views are held if comparing course and subject guides to one another

- **Visualization 2:** illustrates number of unique views by each librarian and then continues to expand and enable us to look at unique views tied to each LibGuide

- **Visualization 3:** illustrates the list of resources and their unique views in comparison to one another

- **Visualization 4:** illustrates the unique views of each LibGuide by week and includes the number at peak use.
After meeting with the Head of Research and Instruction to show the final dashboard, she shared enthusiasm for the model and felt it could be used to communicate the use of LibGuides with her staff. Suggestions for further improvement included integration of where people are visiting the pages from. She wanted to gauge whether visitors were from outside the university system or primarily locally based. Additionally, she asked if there is a way to see whether patrons are then actually using the resource or just following the link and leaving the site. Following steps to make this project effective and useful are gaining access to the Google Analytics API to generate a
script where data can be downloaded in bulk, as well as writing a script to both consolidate the data, clean data, add tags, and associate each guide with a specific librarian.
3 Southern Historical Collection

3.1 Needs Assessment

The Reference and Instruction department of the Southern Historical Collection (SHC) uses AEON to support their reading room and instruction activities, including paging and material management. The Southern Historical Collection is home to over 5,000 distinct archival collections. These are each comprised of unique primary documents, such as diaries, correspondence, photographs, maps, and oral histories.

AEON is a complex transactional system that associates researchers and events with specific collections. Data is entered by patrons through a web form that they are redirected to from individual collections’ finding aids. While the finding aid pre-populates some of the data associated with the collection, there are some blank text fields that patrons can optionally fill in. For example, volume is an optional field that can capture which part of the collection a patron is interested in. Collections are comprised of many parts. Knowing that the University Papers are most used as a collection tells us significantly less than knowing which volumes within the University Papers are most used. Given resource limitations and staff’s limited time, decisions about what volumes within a collection are most appropriate for digitization are significantly aided by capturing this data.
Currently, the SHC has limited capacity (both time and skill level) to analyze their internal data. The Research and Instruction Librarian for the SHC initially requested support in downloading instruction data from AEON. He knew the database included functionality to export data but was unfamiliar with the technical steps involved. Specifically, he shared an interest in identifying which collections were most used or requested during instruction session. Given the sensitive nature of archival materials, if the same collection and volume are consistently pulled throughout a semester or across semesters, there is an increased chance for damaging these materials. As such, analyzing their use for instruction sessions, an event generally involving a high volume of people using these materials, is of relative importance to the function of the SHC’s mission to preserve materials in their collection.

3.2 Implementation Process

- **Data gathering**: Using the AEON querying system, I was able to download instruction data as an excel spreadsheet. AEON captures all transactions, including reading room transactions. To isolate instruction data from Fall 2016, I used the search filter “Instruction” under Activity Type and limited the date of activities to span from August 23rd to December 7th.

- **Mid-point meeting with stakeholder**: The second meeting with Matt Turi, Research and Instruction Librarian at the Southern Historical Collection was a more thorough overview of the way data is captured and used across the department. I gathered
- **Data cleaning and preparation:** The instruction data was relatively straightforward and simple. Data was consolidated in one spreadsheet. A few major changes to the data included:

- **Isolating the primary volume for each collection:** Collections are often broken up into several boxes and subfolders. When researchers or staff input this information for the purposes of paging, they often enter a string of volumes, starting from the larger container, and narrowing down to the smaller unit. For example, a patron interested in reviewing the Sam J. Ervin Senate Records, a collection comprised of hundreds of boxes of archival material, may be interested in only reviewing one box deemed relevant after looking through the collection’s finding aid. A typical request may involve a string such as Box 273, Folder 11298-11317. Using the text to columns function, I isolated the largest unit recognizing commas as the separator.

- **Using collection title as alias for item author:** The archival processes associated with manuscripts and monographs are vastly different. Monographs are generally stand-alone items with one author and one title. Manuscripts are processed as whole collections, meaning multiple unique items can have different titles but one unique author or creator. When requested from the finding aid, the title and author fields are pre-populated for monographs, while manuscripts will only populate the title and leave the author field blank. As such, I duplicated the collection title as an entry under author name for all manuscripts.
• **Cleaning entry for location of item:** Some items had duplicate entries for their location or variations of the same location. I created a conditional statement to check for the major keywords for each location and populate another cell appropriately. For example, a cell with the entries: “Rare Book Collection” and “Rare Book Collection vault” would be re-classified as “Rare Book Collection.”

• **Final meeting with stakeholder:** Review of final produce. See below for detailed description.

### 3.3 Evaluation and Following Steps

As part of the final meeting with the stakeholders at the Southern Historical Collection, Matt Turi, Research and Instruction Librarian in Manuscripts and Jason Tomberlin, Head of Research and Instructional Services, I went over the specific components of the visualization developed for the SHC. The visualization is primarily comprised of a tree map so as to showcase the collections that are used most as a whole and the components that are used within those larger wholes. Given that the SHC is primarily interested in identifying collections and parts of collections that are repeatedly used, I incorporated a filter so as to only show collections that had been used at least twice throughout the semester. Additionally, this visualization is connected to a listing of the volumes and the class name associated with it. There is also a visualization showing number of activities associated with each librarian. Both staff members from the SHC felt this visualization was a great way to assess their collections at the end of each semester. They were also interested in continuing this practice and retroactively pulling data for the past couple of years. They expressed
interest in learning how to incorporate this tool and type of assessment into their end of semester evaluation process.

Fig 3. Screenshot of treemap visualization developed for the Southern Historical Collection highlighting manuscript collections most frequently used in Instruction sessions.
Requests for additional improvements included: Incorporating a tree map that only looked at monographs. Because manuscripts are most effectively viewed through the lens of the creator, rather than the title, if we select purely by creator for monographs, it will not showcase the title because multiple monographs could have been written by one singular author. As such, I created a separate tree map examining only monographs. They also suggested having a list of requests by librarians so they could share this with their colleagues and better illustrate how much they have used a certain collection to increase buy-in. Additionally, they wanted to be able to see collection use for specific classes. As such, I developed a hierarchical graph that shows the number of resources by librarian and expands out to showcase the collection based on classes taught. I, again, incorporated a filter to showcase only records greater than two.
4 Lessons Learned

Supporting the User Experience and Research and Instruction departments at Davis Library and the Southern Historical Collection at Wilson Library through data visualization was an enlightening and rewarding experience. As anticipated, given the high volume of patron transactions, collection materials, and programming occurring at libraries, data already collected everyday by library departments offer an immense opportunity to improve services, communication, or decision making. With both departments, our first meetings indicated that each gathered data for both transactional and assessment purposes that could benefit from visualization efforts. Specific learnings from each site include:

- **Importance of identifying all stakeholder at initial meeting:** I began my project exclusively meeting with the User Experience department, the primary department who holds access to the relevant data. Once I began to explore the data, I quickly realized that I had an incomplete understanding the relevance of the data. While google analytics offers an overwhelming number of metric data, it was unclear which data points would be most useful to the actual department who it impacts on a daily basis. I reached out to the Head of Research and Instruction at Davis Library after my first two meetings to gain a greater understanding of how the department uses lib guides and their specific interests. I was quickly able to gain a better understanding of what kinds of questions they were grappling with
as a department. In the future, I will be more strategic about involving key stakeholders who support the collection of data, as well as those responsible for the actual work the data is representing. Furthermore, special collections data necessitates a general understanding of the type of data and items it represents. Given my experience as a graduate assistant in both libraries, I was relatively comfortable with the nuances of the paging process at a Special Collections library, as well as how Lib Guides are used in reference and instruction efforts at libraries. I may have needed additional time or orientation to the inner workings of a department had I partnered with a department such as the E-Resources department.

- **Visualization is as good as data:** Both data sets were populated by automatic systems, making the data relatively clean and easy to use. During discussions with stakeholders, it was clear that there are questions that they might have that could potentially be answered more quickly by more strategic data collection practices or naming processes. For example, setting standards for the directory names for lib guide webpages makes it significantly easier to search for specific guides associated with a subject or course. Final meetings might be an opportune time to walk through some suggestions for how to better capture data moving forward.

- **Explore the mechanism for gathering data early:** While the data was relatively clean for both departments, the actual mechanism for exporting data was complicated for google analytics data due to their content management system. This considerably altered the ability to export data in bulk. There is a significant difference when working with data that is consolidated in one worksheet or a few
sheets versus working with multiple sheets that have to be combined and significantly altered. In the future, I will allocate more time to exploring the mechanism for exporting data and planning strategically based on that.

- **The role of replicability:** Discussions with the Southern Historical Collection staff revealed the importance of creating visualizations that are simple enough for staff with limited technical skills to replicate. The staff shared interest in formally integrating this kind of visualization to their assessment process at the end of each semester. Many library departments have limited resources to bring in outside consultants or new staff to lead more advanced technological projects. As such, first meetings should address whether staff are interested in being able to replicate the visualization to better inform how complex the visualization should be.