

Put it Simply: Tools and Tips for Communicating Library Collections Data

Hilary Davis

North Carolina State University, hilary_davis@ncsu.edu

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

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>.

Hilary Davis, "Put it Simply: Tools and Tips for Communicating Library Collections Data" (2011). *Proceedings of the Charleston Library Conference*.

<http://dx.doi.org/10.5703/1288284314932>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Put it Simply: Tools and Tips for Communicating Library Collections Data

Hilary Davis, Associate Head, Collection Management, North Carolina State University Libraries

Abstract:

In communicating library collections data to our stakeholders and administrators, our main goals are to be impactful, to make our points clearly and concisely, and to provide data that can move decisions forward. Accomplishing these goals requires time and creativity to experiment and refine—assets that can be hard to come by for busy librarians. This presentation will provide an introduction to a few easy-to-use data visualization tools and how they can be applied for communicating data about library collections. The main tools included in the overview will be Google Spreadsheets and Tableau Public (including the pros and cons of each) as well as guidance on how to tame Microsoft Excel's graph design biases. We will also review basic tips to take into account when graphically communicating data on the use and value of library collections to stakeholders. Real world library data such as usage statistics and collection expenditures will be included in the demonstrations of these tools and tips. No programming skills will be needed! By the end of this session, attendees will be equipped with some practical strategies and tools that make it easier to share and make sense of library collections data.

Introduction

Libraries commit significant resources to measure a very diverse set of functions including interactions at service points, online chat and in-person consultations, impromptu guidance for patrons, and website visits. Library staff work hard to expose valuable collections via catalogs, subject guides, bibliographic instruction, resource discovery systems (such as Summon), and journal and database lists. We keep track of all of these interactions by harvesting and analyzing website traffic statistics, gate counts, circulation reports, surveys, usage statistics, and feedback forums. With all of the resources that are devoted to this broad and deep data harvesting and analysis, we must be wise and impactful with the data that is collected and conveyed about patrons, services and collections.

Why do we measure all of these things year after year? Overall, library metrics help us to ensure that we are in alignment with our organizational goals. We need to demonstrate our worth and fit for our parent institutions and we need to show that people are getting value for money we get from our organizations. Collecting the data and analyzing it is only half the battle. The way that we share library collections data with our stakeholders is what ultimately matters.

When conveying data about library collections and services, we need to be able to drive home a point quickly and we need that point to have a lingering impact. Creating good data visualizations is hard work. Good examples are simple and compelling. They can be interactive or static. Most importantly,

they allow us to express value, communicate trends, and test assumptions about library services and collections. In this paper, I will provide an overview of three tools that are useful for data presentation. One you probably use on a daily basis: Microsoft Excel; one that you might already use: Google Documents Spreadsheet; and one that you might not be familiar with, but that you should know about and use: Tableau Public. I use actual library data in all of the examples. The best part is that you do not need to know any programming at all to start using the data visualization tools and techniques discussed in this paper.

- *Microsoft Excel*: I will not go into any detail about how to make graphs or manipulate data using this tool. You should already know how to make charts in Excel. Instead, I will focus on how to tame Excel so that your charts are simple, straightforward, and help you make your point.
- *Google Documents Spreadsheet*: I will briefly review how to make charts in this tool, but I will mostly focus on the kind of charts that are not already available in Excel, and how to share those charts with your stakeholders. Therefore, I would like you to think of Google Documents Spreadsheet as another tool in your portfolio, not as a replacement for Excel.
- *Tableau Public*: I will discuss how to import data into this tool and how to create a variety of different charts that can be published and shared on the open web.

Disclaimer, About Me

My job revolves around the analysis of library collections—measuring and assessing their use, composition and access. I collect a lot of usage statistics, circulation data, measurements of the scope of our collections, comparing those factors to the cost of our collections. As a result, I pretty much live and breathe in Excel spreadsheets on a daily basis. As a practitioner, I am always looking for strategies and tools that make it easier to present and make sense of library data. I am by no means an expert in data visualization, but am constantly trying to find ways to make a better case for library collections decision-making and justification.

Tame Your Graphs

Most librarians use Excel as their primary data work space. I use Excel on a daily basis in my own work because it is pervasive (it is loaded on every computer at my institution), it is a common denominator application (which means that I can share Excel files easily with my colleagues and be assured that they will be able to open and use the files), and it serves most of my daily needs for gathering, reviewing, and assessing library data. There are, however, some limitations of Excel for collaborative work and for making good graphs that make an impact. Below is a list of some elements of Excel graphs that detract from your ability to present a good data visualization.

- Grid lines & borders these are often distracting and unnecessary
- 3-D avoid more than 2 dimensions
- Legends the placement of legends or keys can take up valuable real estate on a graph that could be better used to display the data; do not display a legend if the labels are sufficient
- Titles think of the title as a news headline and use the it as an oppor-

tunity to make your point (e.g., Interlibrary loan increased 40% in 2009-2010 as subscription funds decreased)

- Labels keep labels to a minimum and only label what you need to make your point
- Units always include the units of measure (e.g., dollars (\$); time (e.g., hours, days, weeks, months, years); scale (e.g., number of duplicates, number of faculty)
- Sorting order your data in a way that makes your point, e.g., from highest to lowest expenditures rather than alphabetically
- Pie charts these are best used when trying to convey parts of a whole; avoid using 3-D pie charts since the angles and areas can be difficult to interpret when set askew in 3-D mode

Figure 1 illustrates some of the common mistakes that are made when using Excel to present library data graphically. The issues with this graphical display of library expenditures are that the pie chart is in 3-D making it more difficult to easily see which functions cost the most; the legend is duplicated by the labels; the labels are poorly positioned and make the chart look cluttered; the title does not convey any information about the point of the chart; and the data is not ordered in a logical way (from greatest to least expenditure).

A better way to present this data could be in a simple 2-D bar chart or even a table, listing the expenditures from greatest to least, as in Figure 2.

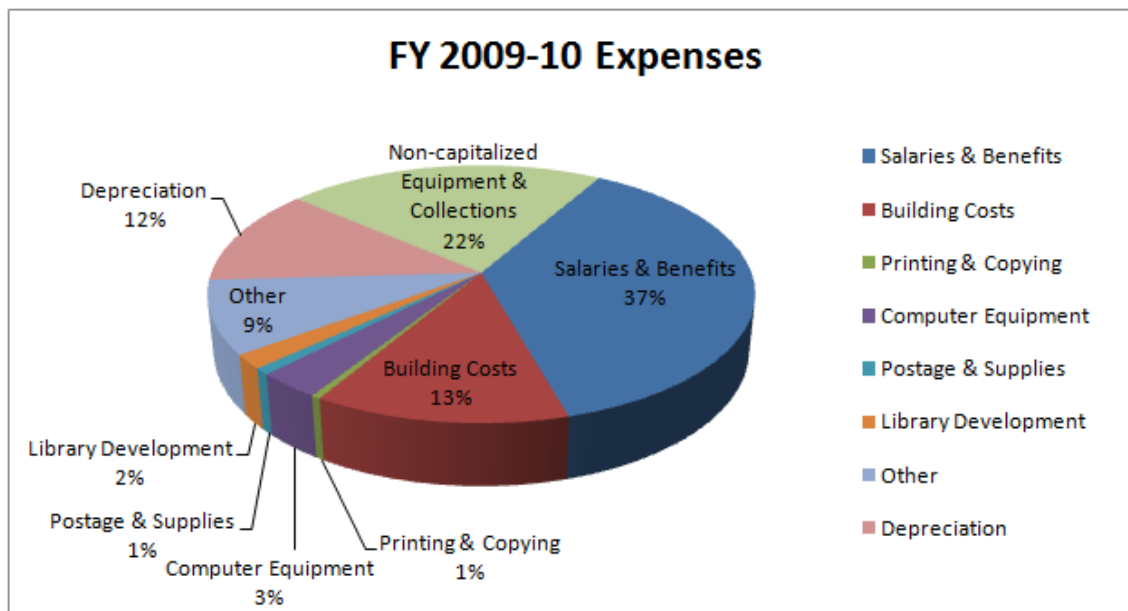


Figure 1: Illustration of the common mistakes made when presenting library data.

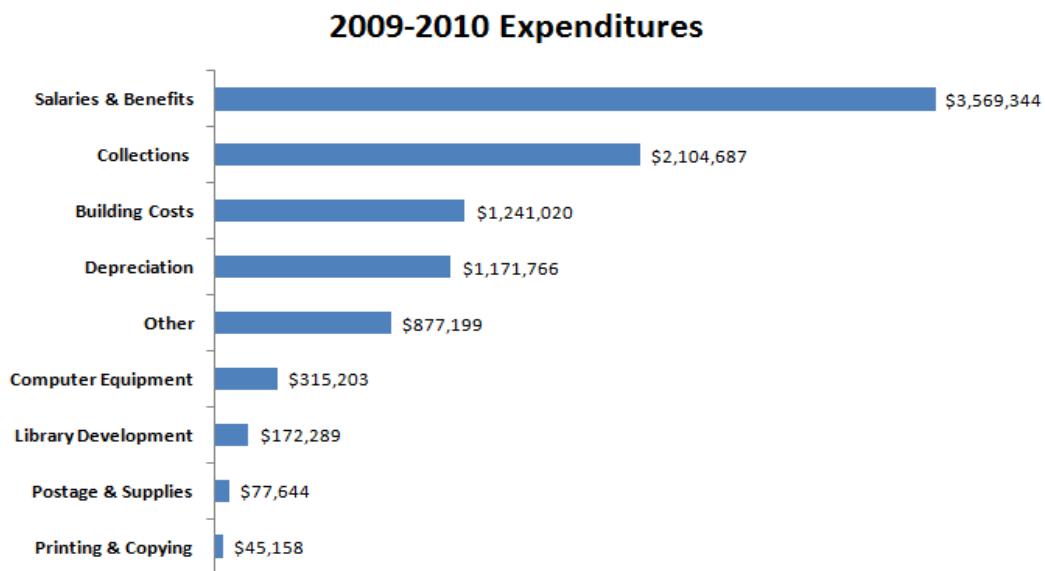


Figure 2: Representation of data presented in Figure 1 using simpler 2-D bars, shorter labels, excluding the legend, grid lines and border, ordering data from greatest to least to help make the point of which functions take up the majority of the library expenditures.

Both Figure 1 and Figure 2 were created in Excel, but they show the difference between overwhelming, unnecessary graphical design practices, and simpler, more straightforward design skills. There is

a relatively easy way to alter the default chart settings in Excel so that you can apply good graph design with the click of a button. Use Excel Chart Templates to create charts with the right design

elements. A quick summary of steps to create and repurpose chart templates in Excel 2007:

1. Create your chart, setting the design you wish to save for use with other charts.
2. Click the chart that you want to save as a template.
3. On the Design tab, in the Type group, click Save As Template.
4. In the Save in box, select the Charts folder; name your template.
5. Select another set of data that you want to display in your saved chart template.
6. On the Insert tab, in the Charts group, select All Chart Types.
7. In the dialog box, click on the Templates folder select the template that you want to use; click OK.

Collaborative work using Excel spreadsheets can be successful, but these collaborations usually take place asynchronously when collaborators are on different schedules or in different locations. As a result, managing versions quickly becomes important when multiple people are editing a spreadsheet in Excel to avoid writing over important changes or appending changes to the wrong version. Sharing graphs created in Excel with stakeholders is easy, but not ideal. Most of us copy and paste a graph into another document and share our Excel graphs as static proxies, cut off from the actual data. When it is important to be able to collaborate on working data and enable stakeholders to not only view graphs, but also interact with them, then it is time to choose a different tool such as Google Documents Spreadsheet or Tableau Public.

Google Spreadsheet:

A Complement to Microsoft Excel

Google Documents Spreadsheet tool is especially useful when you want to collaborate and share

working data with colleagues and graphs with your stakeholders. The benefits of using this tool are that it is free (all you need to do is sign up for a Google account), creating data spreadsheets is just as easy in Google as it is in Excel, and importing data from another application, such as Excel is also very straightforward. In a few simple steps, you can create a variety of different graphs:

1. Log into your Google account
2. Choose Google Documents
3. Create New Spreadsheet
4. Add your data (just as you would in Excel)
5. Once your data is ready, go to the Insert menu and choose Chart or Gadget

Google Spreadsheet includes many of the common charts you would expect such as bar, line, area, pie, etc. (see Figure 3) and even suggests a few common charts based on the kind of data you have.

One of the most interesting charts available in Google Spreadsheet is the motion chart. This type of chart is excellent for displaying data a couple of variables of data over time. They have been popularized by the Gapminder organization, a “global development and achievement” non-profit based in Sweden that focuses on trend analysis and statistical content and tools. To create a motion chart in Google Spreadsheet, your data must be set up so that the first column is the entity formatted as a string (e.g., country, state, university) and the second column is a unit of time (e.g., day, month, quarter, year). The rest of your data can be numbers or strings. Figure 4 shows an example of how to set up the data in the Google Spreadsheet tool, using data from the Association of Research Libraries (ARL) comparing library expenditures to university expenditures from 1982 through 2007.

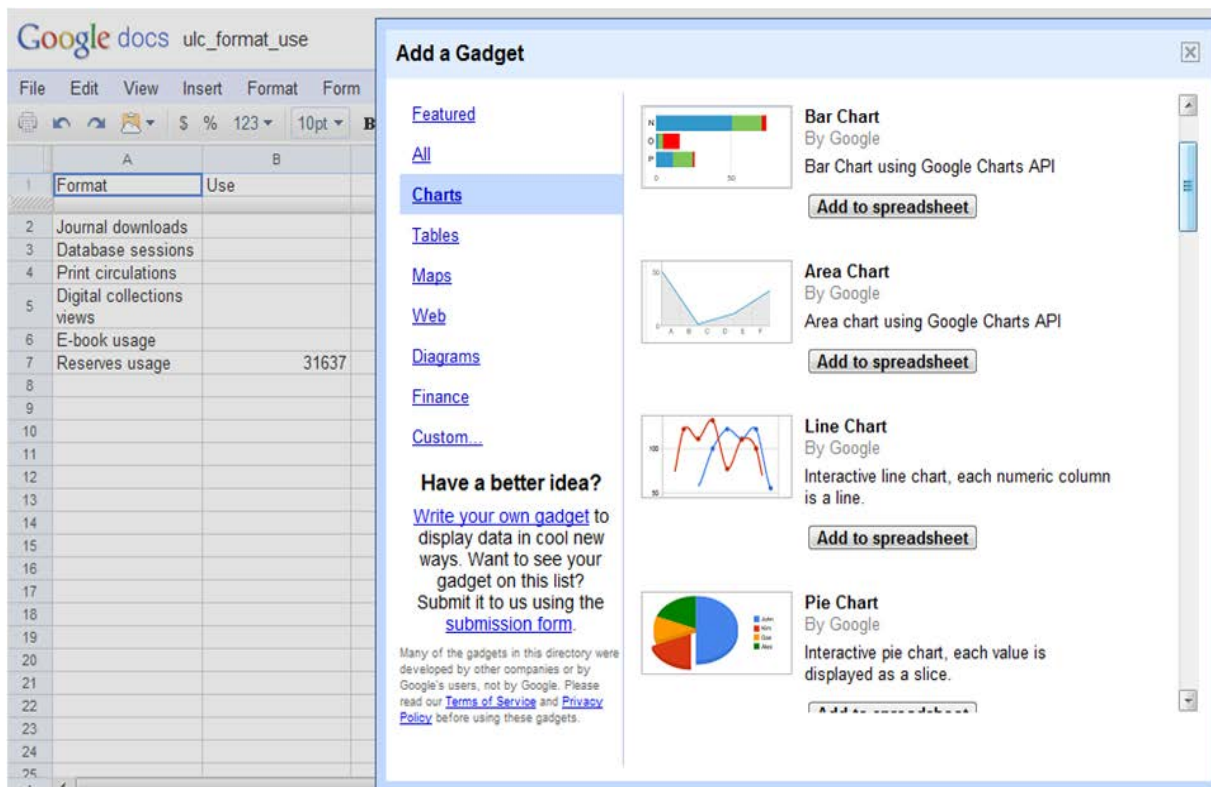


Figure 3: Google Spreadsheet includes at least 37 different chart choices, many of which are interactive such as motion charts and maps.

Institution	Year	Library Expend	University Expe	Region	Type
ALBERTA	1982	9,360,319.00	147,068,270	Canada	Canadian
BRITISH COLU	1982	11,428,727.00	197,967,553	Canada	Canadian
GUELPH	1982	3,894,000.00	78,728,000	Canada	Canadian
LAVAL	1982			Canada	Canadian
MCGILL	1982	7,787,967.00	122,870,790	Canada	Canadian
MCMASTER	1982	5,154,149.00	98,713,662	Canada	Canadian
MANITOBA	1982	5,049,186.00	93,149,179	Canada	Canadian
MONTREAL	1982			Canada	Canadian
QUEEN'S	1982	4,652,966.00	64,525,240	Canada	Canadian
SASKATCHEW	1982	3,730,665.00	69,913,714	Canada	Canadian
TORONTO	1982	16,147,563.00	228,509,429	Canada	Canadian
WATERLOO	1982			Canada	Canadian
WESTERN ON	1982	6,643,825.00	120,309,643	Canada	Canadian
YORK	1982	5,323,215.00	94,645,000	Canada	Canadian
ALBERTA	1983	10,543,759.00	168,416,066	Canada	Canadian
BRITISH COLU	1983	11,941,996.00	165,141,111	Canada	Canadian
GUELPH	1983	4,231,434.00	87,645,153	Canada	Canadian
LAVAL	1983			Canada	Canadian
MCGII I	1983	8,001,650.00	128,069,480	Canada	Canadian

Figure 4: Data from the Association of Research Libraries (ARL) on library expenditures compared to university expenditures from 1982 through 2007.

In Google Spreadsheet, select the columns and/or rows of data that you want to use to create a motion chart. From the “Insert” menu, choose “Chart” and from the Chart Editor view, select the motion chart (this should then give you a preview). Select “Insert”. The chart should appear within the

spreadsheet view; you can edit the chart to move it to its own sheet. See Figure 5 for an example of a motion chart using the data shown in Figure 4 (this is a static representation for this paper, but see the link below the figure for a live interactive version).

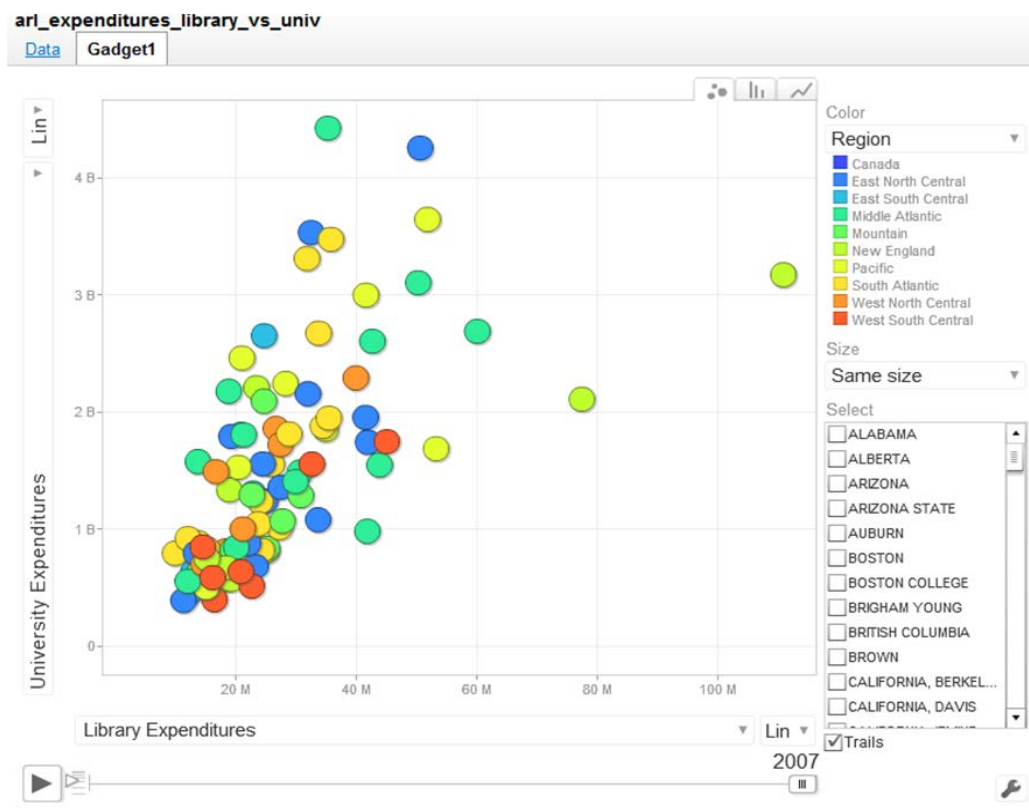


Figure 5: Google Motion chart of data from Figure 4. Live interactive version of this chart is available at: <http://spreadsheets.google.com/pub?key=tZUhX86iPl1x3PEyGPK65YA&output=html>

In Figure 5, as a function of time, as university expenditure increases, the library expenditure increases. As there are cutbacks in university expenditures, we also see cutbacks in library expenditures. A useful feature is that you can highlight the trail of one specific library and watch how its expenditure increases or decreases as compared to the rest of the ARL libraries.

Sharing Google Spreadsheets and any charts or graphs that you create is simple and straightforward. You can choose to Share or Publish the information. Choose “File”--> “Share” to add specific people that should be able to read and edit the data

and charts. Or, choose “Publish” to see a code snippet that you can paste into your website or blog that will enable you to embed your data and corresponding charts.

Tableau Public - Next Best Thing Since Sliced Bread?

Tableau Public is the free web version of the Tableau business intelligence service. The features that make Tableau Public stand out from Excel and the Google Spreadsheet tool is the pervasive drag and drop experience, the ability to publish a dashboard view to the web that combines multiple graphs, easily get graphs to interact with each other via dy-

dynamic filtering, and the richness of data visualizations and formatting options. It is relatively easy to get started with Tableau Public. First, you need to create an account and download a desktop application where you create your data visualizations.

1. Create an account at <https://public.tableausoftware.com/auth/signup>
2. Download the desktop application from <http://www.tableausoftware.com/public/download>

3. Click “Open data” from the desktop application and connect to one of several data formats including Access, Excel, and text files
4. Use the Dimensions and Measures variables to create your graphs and other chart types; use “Show me!” to get suggested graphs to fit your data
5. Create a dashboard with any of the charts/graphs that you make (see Figure 6, for example)
6. Save it all to the web (Tableau Public servers)

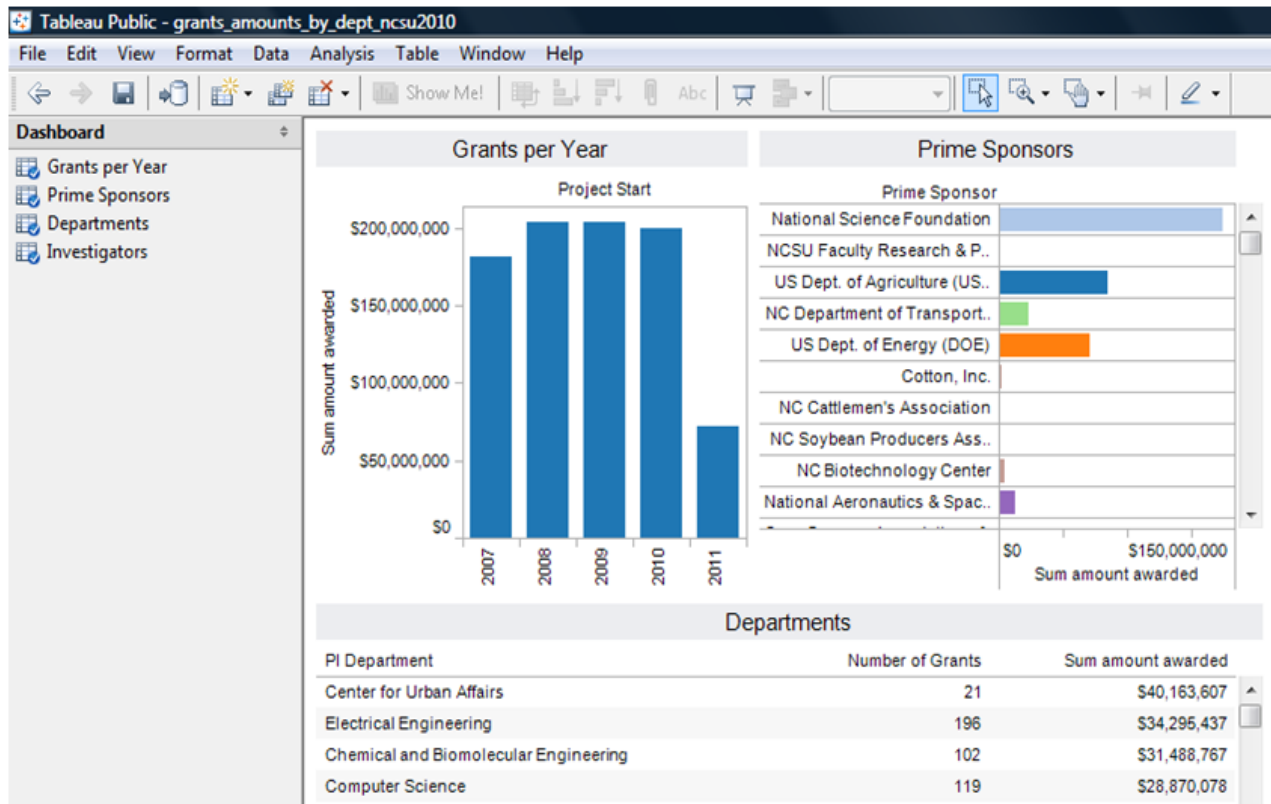


Figure 6: Dashboard view in Tableau Public showing grant data by year, by granting organization, and by department at my institution.

Tableau Public is a fantastic tool for creating panel charts (or small multiples)—that is, a series of charts that show a trends across multiple variables without having to plot everything on one chart.

Panel charts are great for comparing data across sets such as expenditure trends amongst peer libraries (Figure 7).

**Total Library Expenditures amongst NCSU peers
2004/05 - 2009/10**

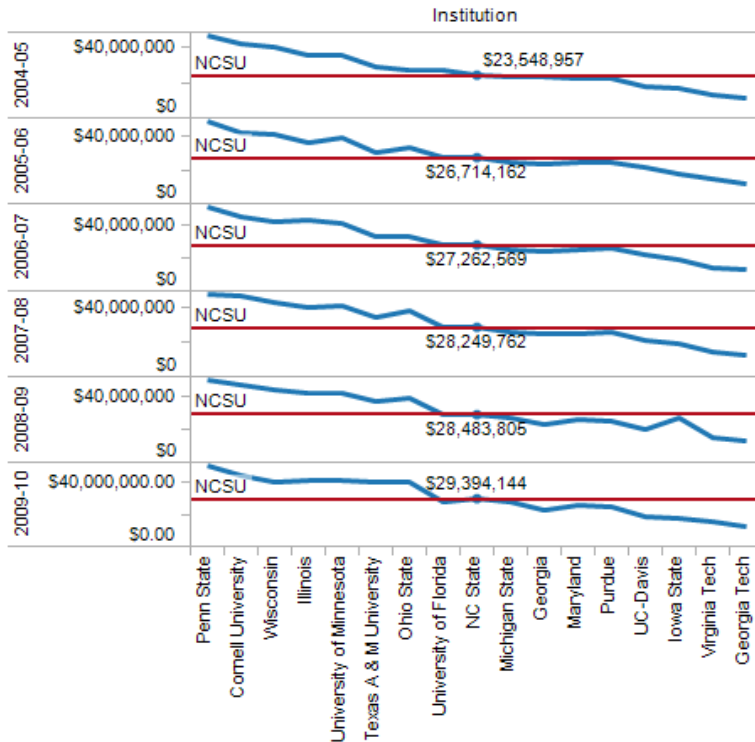


Figure 7. Example of a panel chart created using Tableau Public, comparing library expenditures amongst library peers.

While it is easy to make a simple graph in Tableau Public, but once you get into more advanced functions such as adding filters with specific actions and creating custom calculations, things can get a little tricky. There are a variety of training videos, guides, and an active user community for getting ideas and troubleshooting (<http://www.tableausoftware.com/public/training>). Beyond the “how to” of Tableau Public, there are some important limitations to be aware of before you get started.

- Limit of 100,000 rows or less per dataset
- Datasets must be configured such that the names of fields are at the top and all of the observations are in the cells below
- Graphs cannot be saved to your computer, only can be saved to the Tableau Public servers
- All data and visualizations are openly available
- 50 MB limit per account

Barring these limitations, Tableau Public is a powerful and impactful tool for analyzing your data, creating compelling visualizations, and sharing both the graphs and the underlying data with collaborators and stakeholders. It is a trivial process to share links or embed on dashboard and graphs on your website or blog.

Now, It’s Your Turn

After reading through this overview you should have enough information to tap into your favorite library data and start cleaning up your charts in Excel, reset your defaults in Excel, use Google Docs to create interactive and sharable charts, apply motion charts to data that changes over a period of time, upload data and create charts in Tableau Public, use the panel charts to show trends across related variables, and share the charts you create as well as the data for others to view and repurpose.

Inspiration

The following is a list of references and places for inspiration including links to the datasets used in this paper so that you can go and try it out for yourself if you do not have a dataset of your own ready to go.

Websites:

- <http://flowingdata.com/>
- <http://infosthetics.com/>
- <http://www.visualcomplexity.com/vc/>
- <http://www.gapminder.org/>

Books & People:

- *Visualize This* (Yau, 2011; also author of the FlowingData blog)
- *Beautiful Data* (Segaran & Hammerbacher, 2009)
- *Show me the Numbers* (Few, 2004)
- *Now you see it* (Few, 2009)
- Lisa Kurt and Will Kurt, University of Nevada Reno (<http://www.lisakurt.com/>)

- *Visual Display of Quantitative Information* (2001, and other books by Edward Tufte)
- Cory Lown, North Carolina State University (<http://www.slideshare.net/corylown/data-visualization-7522083>)

Data to play with:

- ARL library expenditures plotted against institution expenditures (Google Motion Chart)
<http://spreadsheets.google.com/pub?key=tZUhX86iPI1x3PEyGPK65YA&output=html>
- Journal usage over time (Tableau Public)
<http://public.tableausoftware.com/views/Elsevier/Dashboard1>
- Backfile cumulative cost per use (Tableau Public)
<http://public.tableausoftware.com/views/JSTOR/Sheet1>
- Funding grants dashboard (Tableau Public)
http://public.tableausoftware.com/views/grants_amounts_by_dept_ncsu2010_1/Dashboard1