Visualizing a Small Academic Collection Using Only Excel and Tableau Public:
A Case Study of an Anthropology Librarian and Collection

Kim Looby, Instruction and Information Literacy Librarian, University of North Carolina Charlotte

Getting Started

1. Download a report of the call number range for your subject area, Anthropology is GN. This should come from your collection services librarians.
2. You will need a basic understanding of a spreadsheet program and a basic understanding of Tableau. Tableau is quite expensive, but has a free version called Tableau Public, that does all of the basic things as regular Tableau.
3. Include: call number, title, author, publication date, check out totals, and anything else your ILS system collects
4. Clean your data: this will vary by column. Make sure everything in the column is in the exact same format – spelling, punctuation, and capitalization. In my sheet, dates had a lot of variation.
5. There are three thousand books in this example collection.

Simple: Library of Congress Classification

1. Make new columns for each LOC classification. Anthropology has four subsections.
2. By hand, fill in each box with the corresponding classification for the call number. Make you type the classification exactly the same each time.
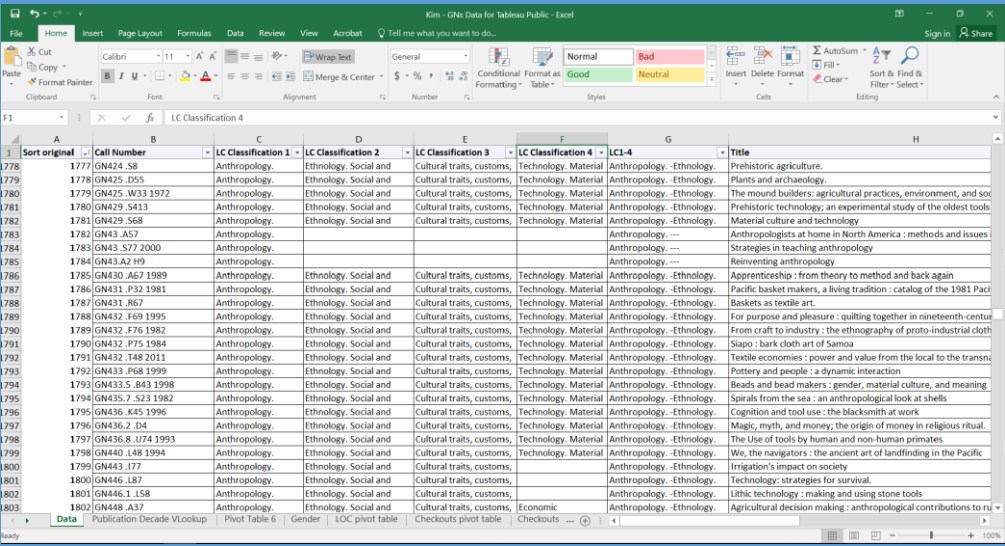3. Combine each classification subsection: concatenate!
4. Create a pivot table.

Image: Original Data Tab

Concatenate

1. Create a second column and name it
2. In the first cell of the new column, type: =concatenate(A1,B1,C1,D1). You can also just click in the first cell of each column you want to combine.
3. If you want to add a space between the combined fields, type: =concatenate(A1," ",B2)
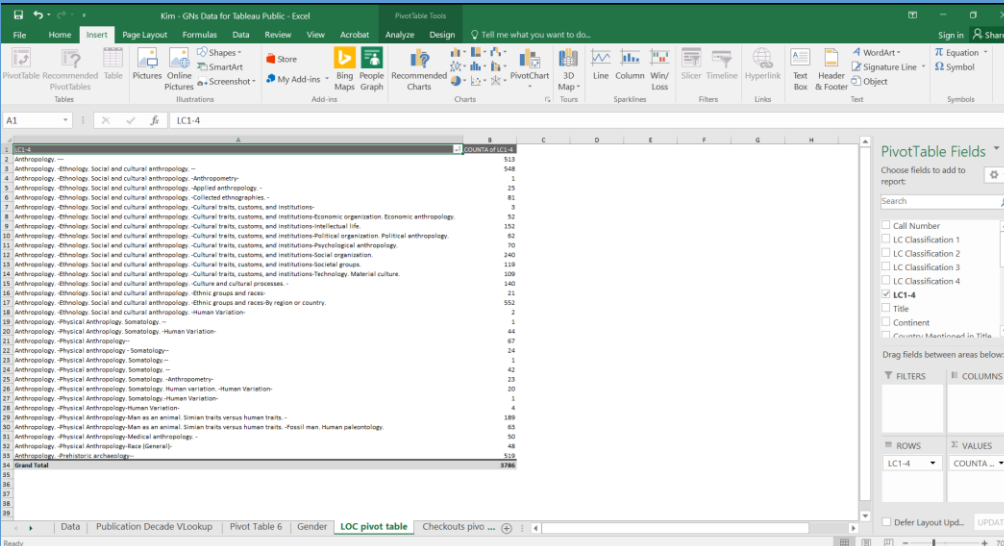4. Now my four separate columns for each subsection are combined into one column.

Image: Pivot Table for Library of Congress Classification

Pivot Table

1. Now you need to take your long list and get a count of how many books you have in each subsection.
2. Select Insert > Pivot Table. This will put it in a new tab.
3. You will see here any discrepancies you had in typing out the subsections. See if you can spot mine!

Subject Area: Countries

In every book collection, there will be recurrent themes in the titles that quickly point to the subject area. In Anthropology, an easily identifiable one was a country and continent. Let's look at countries.
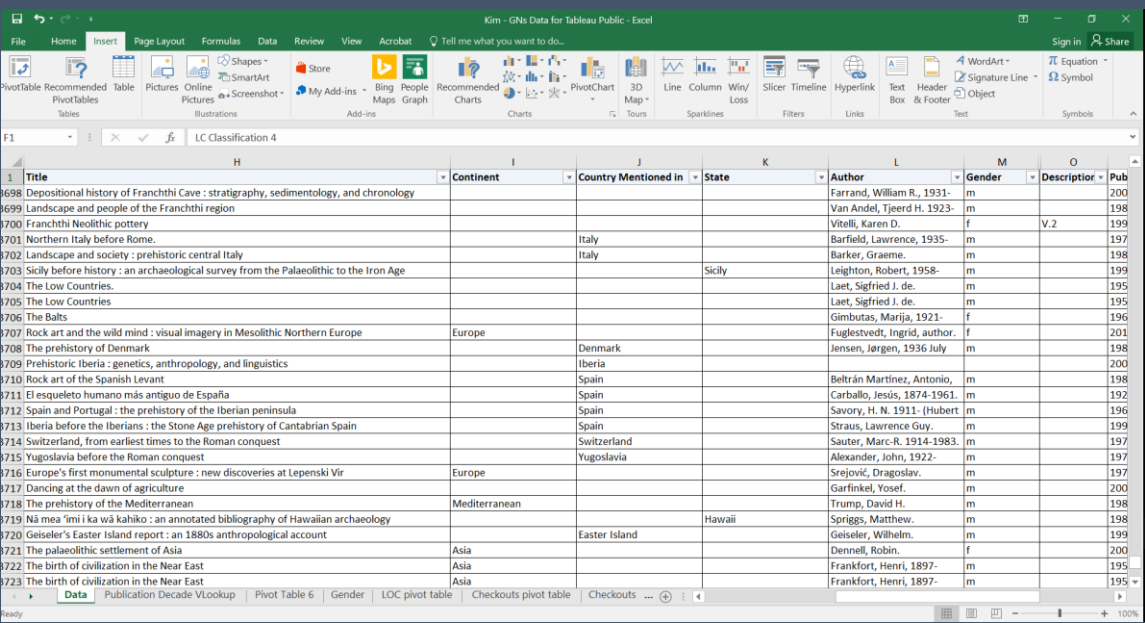

Image: Original Data tab showing columns for adding continent, country and state codes

By Hand

1. Create new columns for however many categories of code you want. I made columns for "continent", "country", and "state".
2. By hand: enter the word or "code" you want to represent that title. If a title I had mentioned "Asia", I put that in the Continent Column, etc.
3. These spreadsheets are great to have when you want something mindless to do.
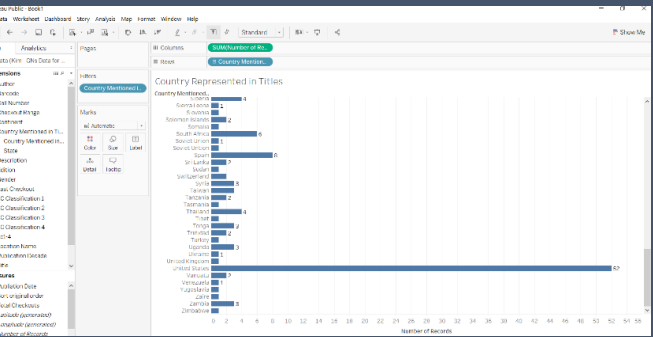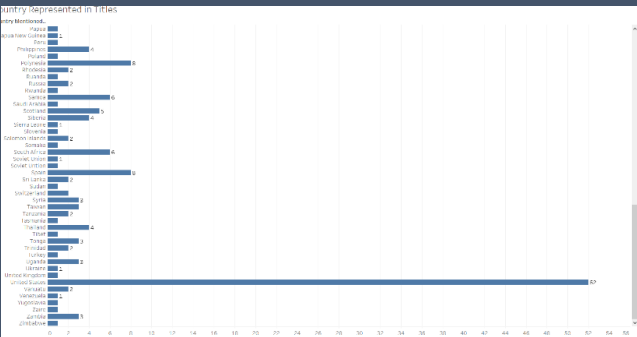
Tableau has many different options for data visualization. Showing an image of how I set up the data in the row and column is easier than describing steps.
Above: The countries are in a bar graph, selecting "mark label" will give you the number at the end.
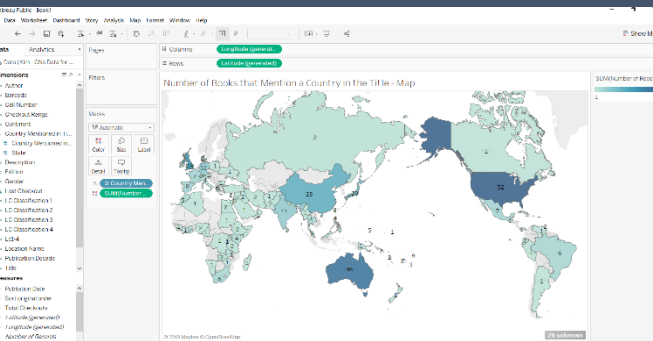Below: Tableau automatically used my data to make a map.

Image: Countries in Titles in Tableau Workbook

Image: Tableau Data Visualization of Countries in Titles

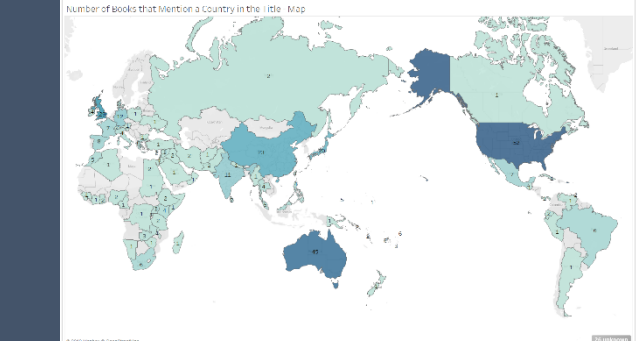Image: Countries in Titles in map form Tableau Workbook

Image: Tableau Data Visualization of Countries in Titles Visualized in map form

How can I use this?

This gives you a snapshot of the representation of your collection in terms of smaller subject areas. I chose countries for Anthropology because many of the books had a country in the title. I noticed in my final lists that we had an abundance of titles about Australia.

To better line up your collection, look at the classes your professors teach and their areas of research. If I have a professor who heavily studies and teaches on Japan, it would be good to have a better representation of books on Japan.

Since the Anthropology department is relatively small, we have had a relatively small budget. To make the budget go as far as possible, I look for books that cover a wide variety of geographic areas or for a geographic area that covers a wide variety of topics.

Don't forget, many of your titles may not mention the subjects you decided to code for. Remember, you're trying to get a quick visualization, this doesn't need to be perfect.

Year of Publication

Image: Original Data Tab focused on publication date
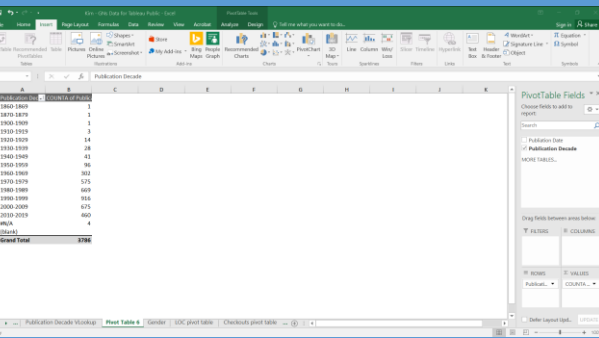Image: Publication Date Range VLOOKUP tab with corresponding year and decade columns
Image: Pivot table of Publication Decades and number of books checked out

For data to be more user friendly, sometimes it helps to break it down into ranges such as turning single years into decades, or number of checkouts into checkout ranges.

Prepare Your Data

1. Create a new tab/sheet within your original spreadsheet and name it "Publication Range"
2. On your original Data tab, sort your publication years from first to last to establish your earliest and most recent year. Mine was 1864-2019
3. In the Publication Range tab, name Column A "Publication Date"
   1. In A2, enter the first year (1864) then populate the column to your final year (2019).
   2. You now have a column that has each year listed from 1864-2019
4. In the Publication Range tab, name Column B "Decade"
   1. In B2, enter your decade range, and be consistent
   2. I chose 1860-1869, 1870-1879, etc
5. Enter the corresponding date range in Column B, that matches the year in Column A.
   1. The year 1864 in Column A would have the corresponding date range 1860-1869 in Column B
   2. Manually do this for all of the years in Column A
6. Return to your original Data tab and make a new column called "Publication Range"

VLOOKUP – Turn Dates into Date Ranges

1. On your original Data tab, in the new Publication Range column, in the first cell underneath the heading, type in: =VLOOKUP
   1. You are now entering a formula, so pay close attention to punctuation and cells because one tiny mistake can give an error
   2. Your cell now says: =VLOOKUP(
2. Click into the cell immediately to the left of the cell with your equation, this cell should be the first cell in your publication date column. We are still on the Data tab
   3. The equation now reads: =VLOOKUP(P2,
      1. Make sure you add the comma after clicking on the cell to the left
      2. Mine says P2 because my Publication Date in my Data tab was in Column P
4. Now you want to go to the dataset, where you want to associate the new information, which is the Publication Range tab that you created which has the list of years and corresponding decade ranges
5. Highlight Column A and Column B
6. The equation now reads: =VLOOKUP(P2,PublicationDecade!A:B,
   1. Make sure to add the comma

7. Put in the number: 2 for the second column which is Column B in your Publication Decade tab. I cannot explain why it is a 2, and not a B, but it is. Add another comma.
8. The equation now reads: =VLOOKUP(P2,PublicationDecade!A:B,2,
9. Now type: FALSE into the equation. In code language, FLASE means "exact match".
10. The equation now reads: =VLOOKUP(P2,PublicationDecade!A:B,2,FALSE
11. Close your parentheses and hit enter
12. Final equation: =VLOOKUP(P2,PublicationDecade!A:B,2,FALSE)
13. Go back to your original Data tab.
14. You should now see the decade range next to the year in the first two cells of the columns Publication Date and Publication Range
15. Double click on the small square at the bottom right of the cell to auto populate down the whole column. If this doesn't work, you can drag it down.
16. Now every year should match the correct decade range.

How can I use this?

You can now see how many books that were purchased during a decade. You can put it into Tableau and find other interesting data visualizations.
Remember: these are totals and not percentages. Maybe your institution experienced budget changes or a switch to more electronic books. Keep in mind other things that may affect your data.

Checkouts

To prepare your checkout data, follow the same directions for creating ranges. Your checkouts will have too many individual numbers and it is more useful to put them into individual ranges. This will depend on the range of numbers you have.

I chose the following ranges for checkouts:
0
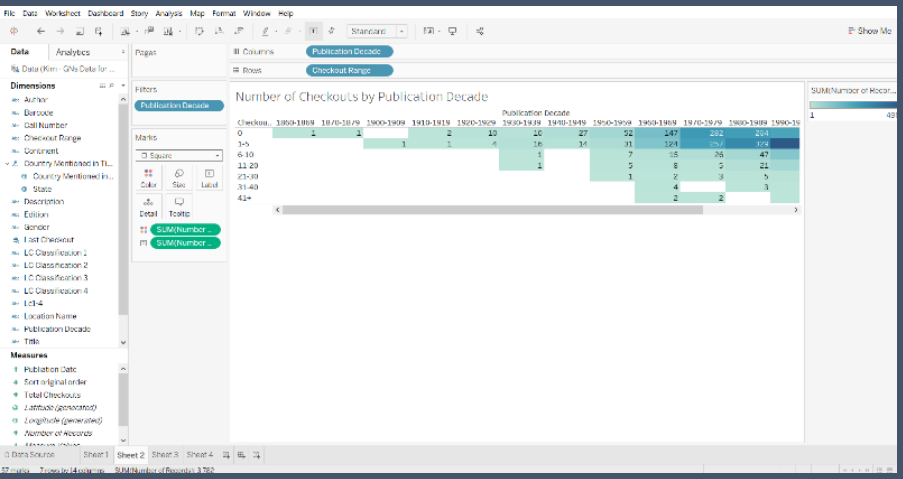1+5
6-10
11-20
21-30
31-40
41+

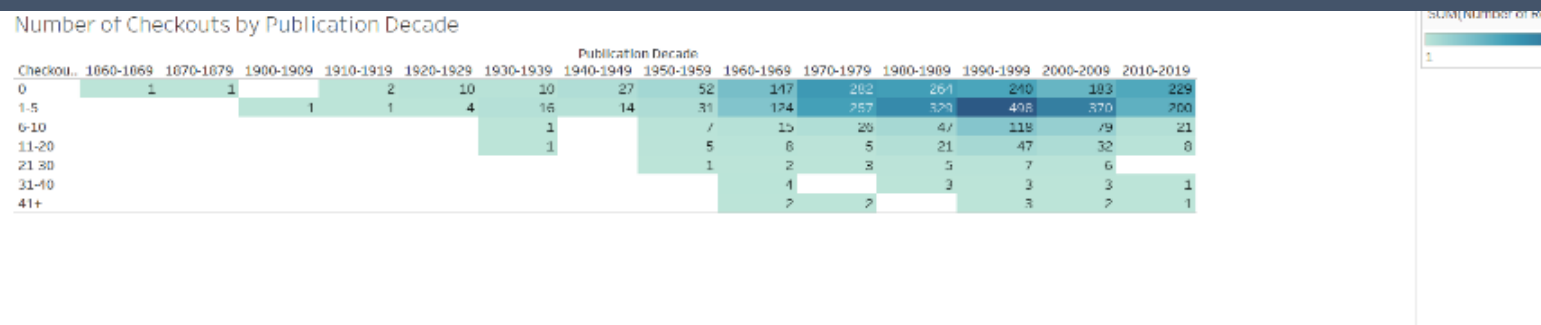Image: Number of Checkouts by Publication Decade Tableau Workbook

Image: Tableau Data Visualization of Number of Checkouts by Publication Decade

Using checkout/usage statistics to evaluate your print collection is daunting, but it can tell you how much of the collection is being used. It can also lead to interesting discussions of the overall value of what we purchase for the library.

Trapped in an era of budget restrictions, libraries everywhere have to make choices of where to spend money. Libraries are also running out of space from years of pressure to acquire large collections to meet national standards.

What determines success? - To check out books? To meet quotas? To have available just in case?

How do you justify purchasing books that are never touched?

What do you prioritize in your library? Collections? Technology? Librarians?

According to this collection, 40% of the books, 1,451 have never been checked out. What is a library without books?

What are books if they aren't being used? What is a library without print books?

How can I use this?

If the average Anthropology book costs around $70.00, this is a loss of around $100,000 just for one small collection over about 60 years which is about $1500 a year. Depending on your collection development philosophy, here are some other options.

Other Options
1. Purchase on Demand
2. Join or create a consortium
3. Join an organization that identifies libraries to hold on to certain titles
4. Check metadata to make sure books are findable in your collection
5. Promote usage of books to students and faculty
6. Only purchase e-books that are DRM free with unlimited users

Be realistic, find out what students, staff and faculty need. If it isn't books, consider purchasing fewer books and move funds toward other types of collections.

Final Data Considerations

1. Totals and percentages tell different stories. Make sure to look at both when assessing change over time.
2. The numbers don't always tell the whole story. Consider outside influences that may affect purchasing books over time
   1. Budgets
   2. National economy
   3. Technological changes
   4. Collection policies
   5. Individual librarian preference
3. If errors arise – ask the internet. Someone has had your problem before and has made a page or video about it.

Tell More Stories

Try out different combinations to see interesting visualizations

- Gender of author x publication date
- Country x checkout
- Checkout x LOC classification
- Top 10 titles
- Country x zero check outs
- Location of book in the library x checkout

What ideas do you have or what data from your collection have you used to make decisions?

Further Reading

Borin, Jacqueline, Yi, Hua, and Borin, Jacqueline. "Assessing an Academic Library Collection through Capacity and Usage Indicators: Testing a Multi-Dimensional Model." Collection Building 30, no. 3 (January 1, 2011): 120–125. http://search.proquest.com/docview/1671334491/.

Gillies, Scott, and Stephenson, Carol. "Three Libraries, Three Weeding Projects: Collaborative Weeding Projects Within a Shared Print Repository." Collection Management 37, no. 3-4 (July 1, 2012): 205–222.

Holderman, Sharon. "The Decline of Circulation: Reversing the Trend through Technology Checkout." The Journal of Academic Librarianship 43, no. 5 (September 2017): 451–453.

Kelly, Madeline. "Applying the Tiers of Assessment: A Holistic and Systematic Approach to Assessing Library Collections." The Journal of Academic Librarianship 40, no. 6 (November 2014): 585–591.

Kessler, Jane. "Use It or Lose It! Results of a Use Study of the Print Sources in an Academic Library Reference Collection." The Reference Librarian 54, no. 1 (January 1, 2013): 61–72.

Knowlton, Steven, and Hackert, Lauren. "Value Added: Book Covers Provide Additional Impetus for Academic Library Patrons to Check Out Books." Library Resources & Technical Services 59, no. 3 (July 1, 2015): 112–119. http://search.proquest.com/docview/1700035768/.

Ladwig, J. Miller, Thurston, and Ladwig, J. "Are First Circulation Patterns for Monographs in the Humanities Different from the Sciences?" Library Collections, Acquisitions, & Technical Services 37, no. 3-4 (January 1, 2013): 77–84. http://www.sciencedirect.com/science/article/pii/S1464904913.

Rose-Wiles, Lisa M. "Are Print Books Dead? An Investigation of Book Circulation at a Mid-Sized Academic Library." Technical Services Quarterly 30, no. 2 (April 1, 2013): 129–152.

Tyler, David C., Falci, Christina, Melvin, Joyce C., Epp, MaryJou, and Kreps, Antaim. "Patron-Driven Acquisition and Circulation at an Academic Library: Interaction Effects and Circulation Performance of Print Books Acquired via Librarians' Orders, Approval Plans, and Patrons' Interlibrary Loan Requests." Collection Management 38, no. 1 (January 1, 2013): 3–32.

Ward, Suzanne M. Rightsizing the Academic Library Collection Chicago, [Illinois: ALA Editions, 2015.