Using COUNTIF and COUNTIFs Formulas to Help Analyze & Graphically Represent your Instruction Data

Jennifer Sharkey, Illinois State University

Libraries have a deep tradition of being a communal repository for information. This means libraries gather and retain significant, voluminous amounts of data on a daily basis. Analysis of the data, particularly non-numerical data, on the other hand may not necessarily be part of normal practice. Reasons for lack of analysis can vary, from technical to budgetary to philosophical. However, a library's ability to effectively communicate via their data is becoming increasingly important (Tenopir, 2013). Additionally, telling the story visually provides a much stronger impact and creating those representations are much easier with numbers then text (Elguindi and Mayer, 2010). While most libraries gather instruction data, what to collect can be a bit dubious as well as determining how to analyze what has been collected (Brooks-Kieffer, 2010). Compile too much data and analysis becomes overwhelming without access to the right tools; gather too little and a library is prevented from effectively telling their story.

At Illinois State University's Milner Library, the instruction department is responsible for gathering and analyzing the library's teaching and learning data. However, until recently, we were undertaking the first part of our responsibility, but not so much the second. The reasons why this was happening were due to the data being virtually locked away on a server or it was paper-based and there was no access to a robust analytics tool. Consequently, our data analysis was extremely lacking and analyzing non-numerical data suffered the most, as that was a time intensive process. This lack of analysis prevented us from gaining an in-depth understanding of what, why, and how we were teaching, how we were demonstrating student acquisition of specific skills and competencies, and communicating our efforts to the appropriate audiences.

Finding a Solution

In trying to find a solution, quite a bit of thought went into how we wanted to analyze the data. Some of it was simple summing (e.g., total number of sessions). However, much of it was a need to count the occurrence of specific data points, sometimes in isolation and at other times in combination with each other. For instance, we wanted to be able to determine how many graduate students were taught in a fall semester for the history department. Also, we wanted to know how many times students used a specific word or phrase in their answers to open-

ended questions on some of the assessments we gave. Even though this data was gathered via various web forms and in a digital format, it was comprised of over 5000 individual instruction records and several thousand answers to open-ended assessment questions. The volume alone made easy or quick manual analysis an impossibility.

The hard part was finding a better way to analyze all of the data we had gathered without requiring the purchase of an entirely new platform, since the library administration was unwilling to fund such a purchase. Our homegrown instruction database had zero analysis functionality and the University's license to Select Survey did not include the needed analysis functions. Our investigation centered primarily on Microsoft Access® and Excel®, which turned up the spreadsheet COUNTIF and COUNTIFS formulas. Using these formulas resolved the time barrier of manual data analysis and reduced required technical skills for examining non-numerical data since knowledge of creating macros or VBA programming was no longer needed. For a librarian or library that does not have access to fee-based analytics tools, the COUNTIF and COUNTIFS formulas can save a tremendous amount of time and provide the opportunity to use gathered data in different ways.

How These Formulas Work

For some of our reporting, just basic charts or tables needed to be generated. The COUNTIF and COUNTIFS formulas allowed us to easily generate rows and columns of numerical data for easy conversion into charts, graphs, and tables. For our non-numerical data, such as students describing how they search for information, these formulas allow us to do a count for specific words, like "Google" or phrases, like "read the article abstract."

Both the COUNTIF and COUNTIFS formulas have specific purposes and are useful in unique ways. The COUNTIF function counts the number of cells within the designated range that contains one specific criteria (Microsoft, 2015a). This function works best when counting a large range of cells and the counting parameters are fairly straightforward such as counting the number of cells that contain the word or a version of the word like "Google" (Figure 1).

Figure 1: Example of a COUNTIF formula

| Formula Name | Worksheet Name | Cell Range | Criteria |

```
=COUNTIF('pretest'!$M$1:$M$5000,"googl*")
```

Table 1

Element	Important Parameters	Required
Comma (,)	Distinguishes the cell range from the criteria	**X**
Double quotes (")	Encases criteria regardless if it is text or numerical	**X**
Worksheet Name	Use when the data being counted is in a different worksheet Place the worksheet name, in single quotes, before the cell range Place an exclamation point (!) between the worksheet name's final single quote and the cell range	
Dollar sign ($)	Locks the cell range (making it an absolute cell reference); prevents the automatic shifting of a cell range when copying and pasting a formula.	
Wildcards (?, *)	Works just like wildcards in search engines Can truncate at the beginning, end and middle (search*; *library catalog*; w*m; wom?n) ? represents a single character * denotes one or more characters	
Variable limits	Must be enclosed in quotes such as ">=4" Can be used with text and numbers <> -- not equal to <= -- less than or equal to >= -- greater than or equal to = -- equal to (although often not necessary with text as text in quotes -- e.g., "journal" -- is interpreted as equals to)	

The COUNTIFS function counts the number of cells that meet multiple parameters (Microsoft, 2015b). This type of count would be similar to a complex Boolean Operator search where AND and NOT are used in multiple search boxes. This is the function to use when the counting parameters are more complex and counting different criteria in different columns is needed. For instance, counting the number of ENG101 sections that covered the topic plagiarism on or after January 1, 2014 (Figure 2).

Figure 2: Example of the COUNTIFS formula

```
=COUNTIFS
($L$1:$L$5000,"plagiarism",
$C$1:$C$5000,"eng101",
$A$1:$A$5000,">=1/1/2014")
```

In deciding what to count, we realized we needed to identify what data should be included in the same chart. For instance, did we want a chart just focusing on plagiarism instruction across all sessions or did we want to a chart that showed the top five skills taught? Once this was determined, we grouped the relevant COUNTIFS formulas into appropriately labeled rows and columns in a worksheet separate from the actual data.

Dissecting the Formulas

For both formulas, the two key components that must be present are designating the cell range(s) and the specific criteria to be counted. Other aspects of the formula may or may not be required. What is required depends on several different factors as it relates to what specifically is being counted and the specific needs for that count. Table 1 identifies different elements and important parameters that make the formulas work.

Moving Beyond Charts and Graphs

Data visualization is an area that is currently receiving an incredible amount of attention within higher education due to importance of communicating key statistics and information in a way that makes an impact and is easy to understand. With non-numerical data, this can be particularly challenging because most types of visualizations require numerical data. Utilizing the counting formulas is a good way to cross that barrier.

With the availability of free infographic software, creating interesting visual representations of data is rela-
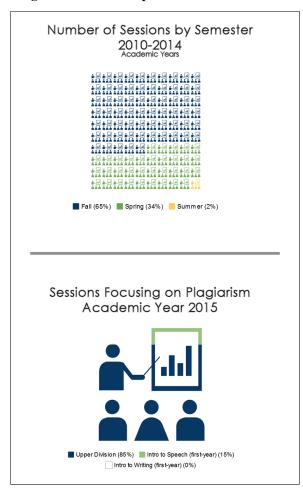
tively easy, particularly if that data is numeric (or has been converted from being non-numeric). In my department, we use the free version of Piktochart for our visualizations (Ching, 2015). With this application, the unique charting feature converts shapes to represent specific data. You can even upload your own icons if the standard Piktochart library does not have exactly what you want to use.

It is important to remember that the point of visualizing one's data is not to make prettier pictures. The strength of this method is to correctly represent data and identify trends (Phetteplace, 2012). Looking at the top image in Figure 3, it gives a strong visual impact, more so then a bar or pie chart, that over a 5-year period only 2% of instructional sessions occurred during the summer. A crucial next step, for instance, would be to use this in-

formation to discuss and determine how library instruction might happen during this time period. This is particularly important when people resources need to be utilized for other types of projects during this time period. The bottom image shows the percentage of sessions that focused on plagiarism. The visual impact here is the icon used and a clear indication where plagiarism is being taught. We can then make any necessary adjustments in our instructional programming to fill gaps and shift what competencies are being taught at what levels. Overall, when creating visual representations like these we found it important to consider where and how the information was being communicated as well as the types of discussion needing to be initiated.

Figure 3: Two Examples of Data Visualizations

Conclusion

While counting is not always the best method of analysis, it is an important component and by generating a numeric representation of specific criteria it is easier to represent the data in various ways — particularly visually via charts, graphs, and even infographics. For different types of analysis, the COUNTIF and COUNTIFS formulas are just one option. Excel® provides a variety of formulas that utilize counting, summing, and if/than logic. As online, fee-based tools become increasingly sophisticated, the need for analysis using separate tools will decrease. Until that time and for libraries not able to afford subscription analytics, utilizing an application such as Microsoft Excel® can offer a wide variety of options for analyzing data that go beyond counting the number of sessions conducted.

References

Brooks-Kieffer, J. (2010). Yielding to persuasion: Library data's hazardous surfaces. In D. Orcutt (Ed.), *Library data: Empowering practice and persuasion* (pp. 3-16). Santa Barbara, CA: Libraries Unlimited.

Ching, A. (2015). How can I make an infographic in five steps? Retrieved from http://support.piktochart.com/hc/en-us/articles/202363710-How-can-I-make-an-Infographic-in-Five-Steps

Elguindi, A. C., & Mayer, B. (2010). Telling your library's story: How to make the most of your data in a presentation. In D. Orcutt (Ed.), *Library data: Empowering practice and persuasion* (pp. 25-33). Santa Barbara, CA: Libraries Unlimited.

Microsoft Corporation. (2015a). COUNTIF function. Retrieved from http://office.microsoft.com/en-us/excel-help/countif-function-HP010342346.aspx

Microsoft Corporation. (2015b). COUNTIFS function. Retrieved from http://office.microsoft.com/en-us/excel-help/countifs-function-HA010342347.aspx

Phetteplace, E. (2012). Effectively visualizing library data. *Reference & User Services Quarterly, 52*(2), 93-97.

Tenopir, C (2013). Building evidence of the value and impact of library and information services: Methods, metrics and ROI. *Evidence Based Library and Information Practice, 8*(2): 270–274.

For a list of Additional Resources regarding Excel, and formulas, see here: http://bitly.com/421_Shark_Apdx