Visualizing Value of Library Collections Relative to the University Teaching and Research Enterprise: An Application of the CDL Journal Weighted Value Algorithm by Three BTAA Libraries

Sarah Murphy
The Ohio State University Libraries, USA

Mary Schoenborn and Jim Stemper
University of Minnesota Libraries

Lee Konrad and Stephen Meyer
University of Wisconsin-Madison Libraries

Abstract

Purpose: Academic libraries license many e-resources through state or regional consortia. Differences in school demographics, disciplinary emphases, budgets, priorities, and licensing restrictions can make the analysis of use and cost patterns in shared collections challenging. Studies have shown that a minority of e-journals in publisher packages get a majority of the downloads. However, not all articles downloaded are later used in teaching or research, leading to the question: What other metrics of use should be considered and how can they be presented to support decision-making?

Design & Methodology: The Big Ten Academic Alliance (BTAA) recently purchased a subset of the Clarivate Web of Science (WOS) dataset, which includes the entire WOS database. The authors, at three out of the fourteen schools in the Alliance, adapted for analysis and visualization the California Digital Library (CDL) Journal Weighted Value Algorithm, which generates a value score for individual journals using authorship, citation, and usage data. This project shows how to model in Tableau COUNTER download data with bibliographic, authorship, and citation data from Web of Science, plus average cost data from Library Journal's annual periodicals price survey, and then effectively display this data through a series of dashboard visualizations. Specifically, the authors created a series of four prototype dashboards to visually answer several questions by subject discipline and publisher package at both broad and more granular levels. The questions included:

- How many BTAA schools subscribe to each title?
- What journals do BTAA faculty and researchers publish their research in? (articles authored)
- What journals do BTAA faculty and researchers cite? (citations)
- What journals do BTAA faculty and researchers download? (downloads)

- Do authorship/citation rates vary widely by subject discipline? (articles authored vs. citations)
- Do 20% of the titles represent 80% of total downloads? (downloads)
- How does a publisher perform in relation to all titles in a given subject area?

Each dashboard shows how many BTAA schools subscribe to each title. Dashboard users can also limit the view to a particular school.

Practical Implications & Value: Tableau offers an opportunity to automate the packaging and display of large datasets, allowing librarians to design useful visualizations and then set a schedule to refresh this data, on a daily, weekly, or monthly basis or on a customized schedule. This saves library analysts significant time when information is needed to inform decision-making. This project shows how to model, analyze, and assess the value of publisher journal collections held locally or by a consortia by visualizing trends for downloads, citations, and authorships. It also provides a proof of concept for the long-range potential for automating the visual analysis of big data to enhance academic library collection development.

Academic libraries license many e-resources through state or regional consortia. Differences in school demographics, disciplinary emphases, budgets, priorities, and licensing restrictions all challenge analyses of use and cost patterns, especially since not all articles downloaded are later cited or used in teaching or research, but do inform practice. To further complicate matters, "although there may be broad agreement between institutions on the most highly used journals in a particular discipline, the actual journals most often used may differ between institutions, depending on local interests and priorities" (Belter and Kaske 2016, 420).

State and regional consortia invest significant resources to purchase large packages of commercial publishers titles. An examination of download data from thirteen OhioLINK universities, however, found "80% of the total articles downloaded [came] from approximately 30% of the titles" (Gatten and Sanville 2004). A similar study at the University of Minnesota-Twin Cities echoed these results (Stemper and Jaguszewski 2003).

To better inform collection planning, and inform licensing and purchasing decisions, the California Digital Library (CDL) introduced the Journal Weighted Value Algorithm in 2010. Comprised of download data, citation data, journal rankings and costs, the algorithm produces a score for each journal licensed by ten University of California campuses. More specifically, the CDL algorithm facilitates objective comparisons between journal titles using a decile rank approach for assigning scores to individual titles. Such scores may then be used to meaningfully group titles by subject and by publisher to inform licensing and purchasing decisions (Li and Eggleston 2016; Anderson 2018). Other schools and organizations, including the University of Minnesota, the University of Memphis, and the Canadian Research Knowledge Network, have adapted this framework for their own decision-making processes (Chew et al. 2012; Knowlton, Sales, and Merriman 2014; Jurczyk and Jacobs 2014).

This study investigates an application of Tableau, a data visualization software, to assess the value of a publisher's collection of titles at three Big Ten Academic Alliance (BTAA) universities. As libraries increasingly use data visualization to effectively analyze collection usage, Tableau has become a popular tool to analyze and share metrics (Finch 2016; Murphy 2015; Wiersma 2016; Wissel and DeLuca 2018). The authors created several prototype dashboards to present data at three levels: by individual title; by subject; and by publisher. The dashboards both visualize trends for downloads, citations, and articles authored and assign a CDL score to individual titles and an aggregate CDL score to subjects, and publisher packages. The dashboards answer several questions including:

- How many BTAA schools subscribe to each title?
- What journals do BTAA faculty and researchers publish their research in? (articles authored)

- What journals do BTAA faculty and researchers cite? (citations)
- What journals do BTAA faculty and researchers use? (downloads)
- Do authorship/citation rates vary widely by subject discipline? (articles authored vs. citations)
- Do 20% of the titles represent 80% of total usage? (downloads)
- How does a publisher perform in relation to all titles in a given subject area?

Tableau facilitates real-time, data informed decision-making, allowing practitioners to design useful visualizations and then set a schedule to refresh data, either daily, weekly, monthly, or on a customized schedule. Tableau also provides a no-code solution for combining disparate data sources to enhance analysis. These and several other features all save library analysts significant time and resources.

Methods

Appendix A provides a rough outline of the steps required and the calculations needed to harvest, model and prepare the data visualized in this project. The team gathered and assembled citation and articles authored data from Web of Science (WOS) and usage data from each institution's COUNTER 4.0 reports for 2015–2017. The team also identified journal holdings data for each institution, and added LC classification numbers to each journal title to assign one of 76 broad subjects. The team then used the subject field to add the 2018 average cost per title located in table 3 of Library Journal's annual periodical price survey (Bosch, Albee, and Henderson 2018).

To calculate the weighted value for each journal title, the team next calculated the decile rank for each metric. Metrics included total downloads, total articles authored, total citations, and the average cost per title in a journal's given subject category. To allow users to weight each component of the final CDL score in a meaningful way, the authors created float parameters with a minimum value of 0 and a maximum value of 1 for each metric, and a check weights calculation to ensure that total weights assigned cannot equal more than 1. Parameters function as variables in Tableau, enhancing interactivity by replacing a constant value in a calculation. The final weighted CDL score was constructed by multiplying each metric by its weight and then adding the result together and multiplying by 100 to display a whole number.

The authors then created empty sets for the subject, journal title, and WOS collection fields. Empty sets enhance the interactivity of visualizations or dashboards by enabling the use of Tableau's set action feature. Unlike filters and highlights, which can only show the values the user has selected, or values that are IN a set, sets differ by placing dimensions either IN or OUT of a set. Set actions allow the user to dynamically identify what values are IN a set and what values are OUT of a set. When subject and title sets are initially set to empty, the visualization or dashboard creator allows the user to identify what values should be IN the set. The set can then be used as a filter to either only show the values that are IN the set, or as a highlighter to show the values that are

both IN the set and OUT of the set. This approach gives the creator the flexibility to create dashboards consisting of trend charts that report data for one title in a collection in tandem with a scatterplot that shows that one title in relation to all of the titles in a collection on the same dashboard. A non-set filter on the title field alone would only show the value for the title selected.

Results

The authors created three interactive, one filter, and two summary dashboards after modeling and preparing the data for analysis.

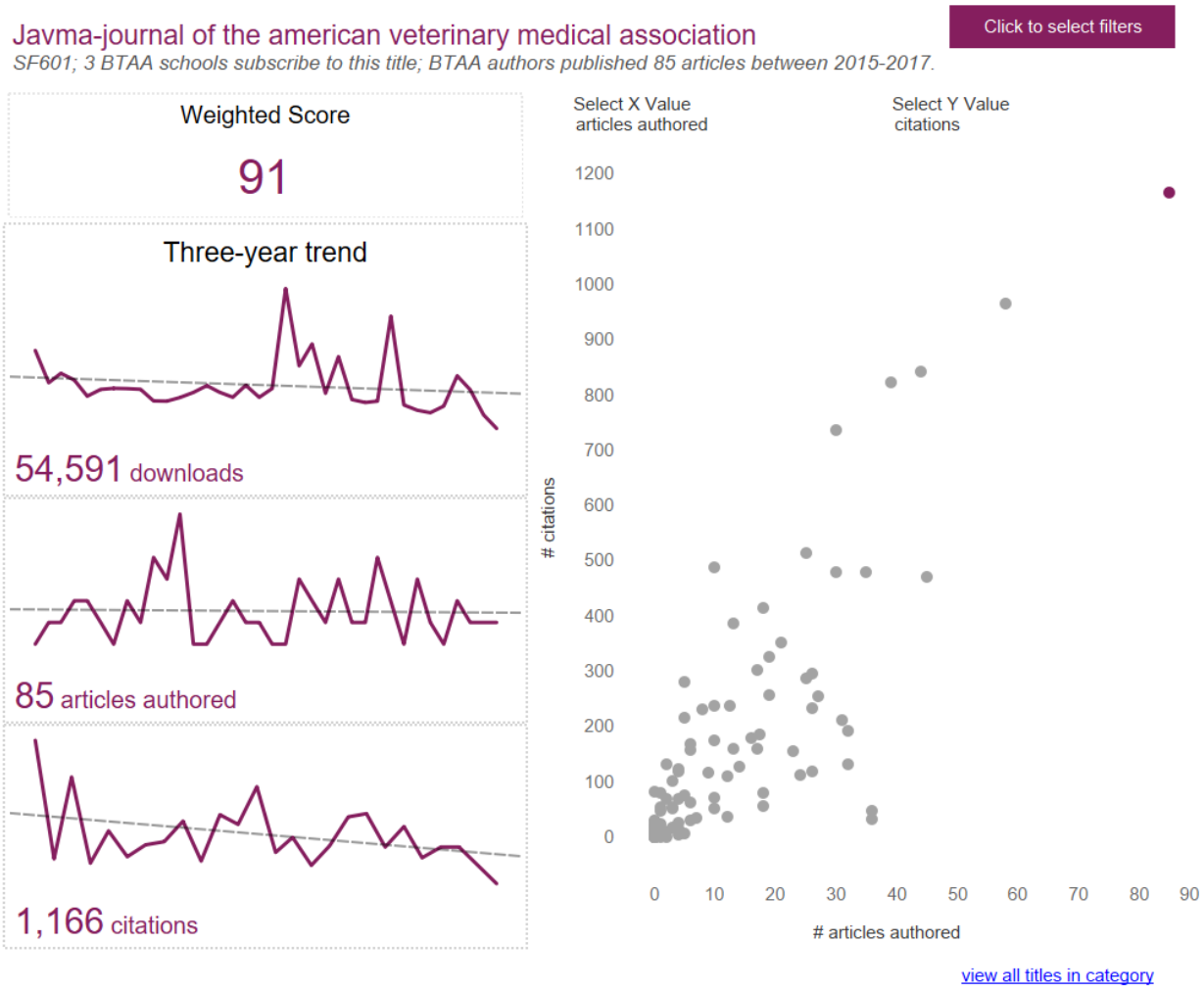
The Subject/Title Dashboard

The Subject/Title Dashboard provides information about a title in relation to other titles in a given subject area (Figure 1). The weighted score appears in the upper left, under descriptive information about the title, the number of BTAA schools that subscribe to the title, and the number of articles BTAA authors published in the title between 2015 and 2017. A series of three trend-charts showing the number of times BTAA authors downloaded the title, the number of articles BTAA authored in the title, and the number of times BTAA authors cited the title appear beneath the weighted score. Together the trend charts highlight the data used to calculate the weighted score, further demonstrating whether scholars' demand for a title is increasing, decreasing, or remaining relatively stable. Researchers at BTAA schools, for instance, might be publishing fewer articles in a particular journal title, for instance, while article downloads are increasing.

The scatterplot allows the user to control what measure is displayed on the x or y axis and then view how a title performs in relation to other titles in a discipline based on the indicators selected. Axis options include the number of articles authored, citations, and downloads. In Figure 1, *Javma-Journal of the American Veterinary Medical Association* appears at the top right in the berry color, while the gray circles represent other titles included in the veterinary medicine subject category.

Tooltips are embedded throughout the dashboard, allowing users to hover over individual data points to view additional context or information. Tooltips open for each point on the trend charts, for example, allowing users to see the specific number of downloads, articles authored, and citations for every month from 2015 to 2017. Hovering over any point on the scatterplot displays a title, the weighted score for that title, and the exact values for the axis measures selected.

Figure 1. The Subject/Title Dashboard

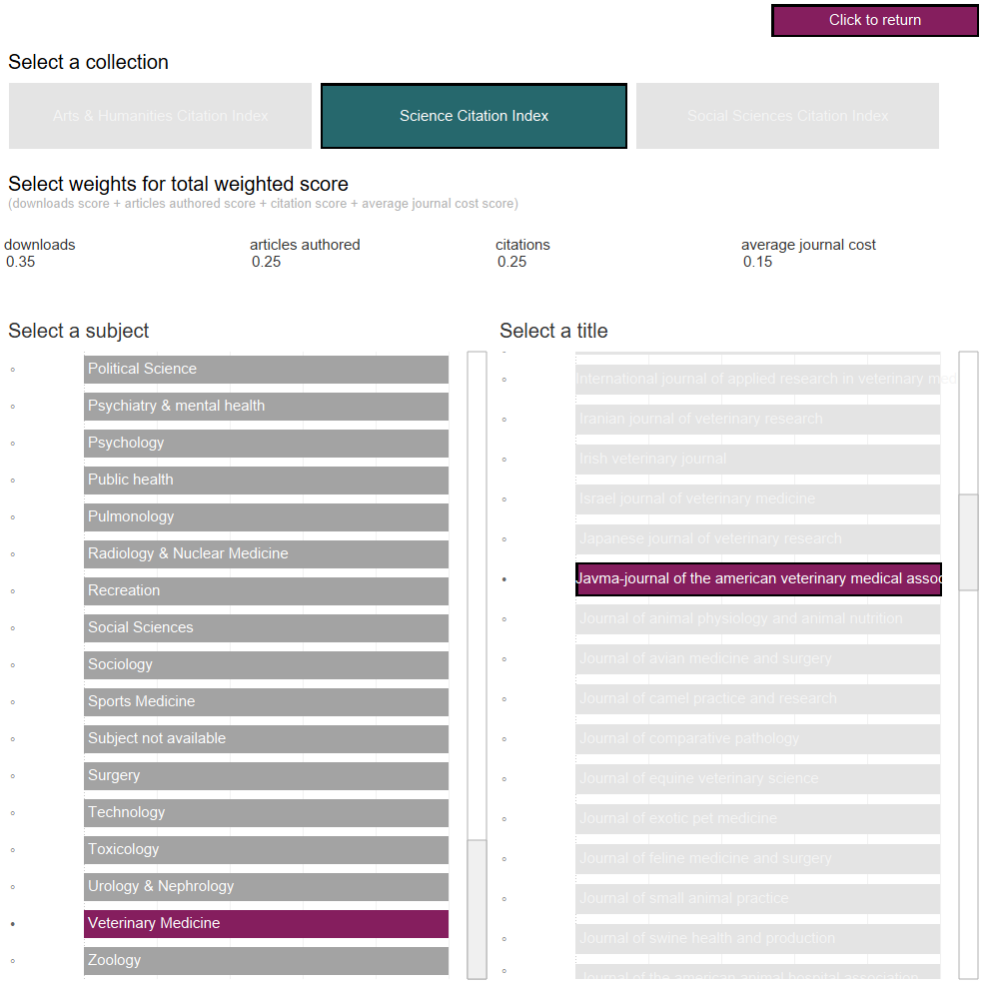


Using the view all titles in category link on the bottom right, users may download summary data for all titles in a subject category to analyze independently. A summary dashboard opens with a text table showing the title, total weighted score, downloads, articles authored, and citations. To export, users need to use the shift key to highlight all titles in the table, and then open the tooltip on the title to select the data icon in Tableau.

To change titles on the Subject/Title Dashboard, users may click the select filters button on the upper right. This opens the Filters Dashboard, where users may select a WOS collection, identify how they want to weight the final CDL score, select a subject, and then select a title within that subject category (Figure 2). Set actions populate the subject and title sections of this dashboard. For instance, when a user selects the Science Citation Index collection, Tableau populates the empty set for WOS collection. Tableau then fills the empty set for subjects included in the Science Citation Index collection and these subjects appear on the bottom left section of the dashboard. When

the user selects a subject, the color for that subject changes from gray to berry, and Tableau then populates the empty set for journal titles included in the subject area. These titles appear on the bottom right. The user then selects a title from the list on the right, and clicks to return to the Subject/Title dashboard using the button on the top right. The Subject/Title dashboard then recalculates and displays the data for the newly selected title.

Figure 2. The Filters Dashboard



The Publisher Dashboard

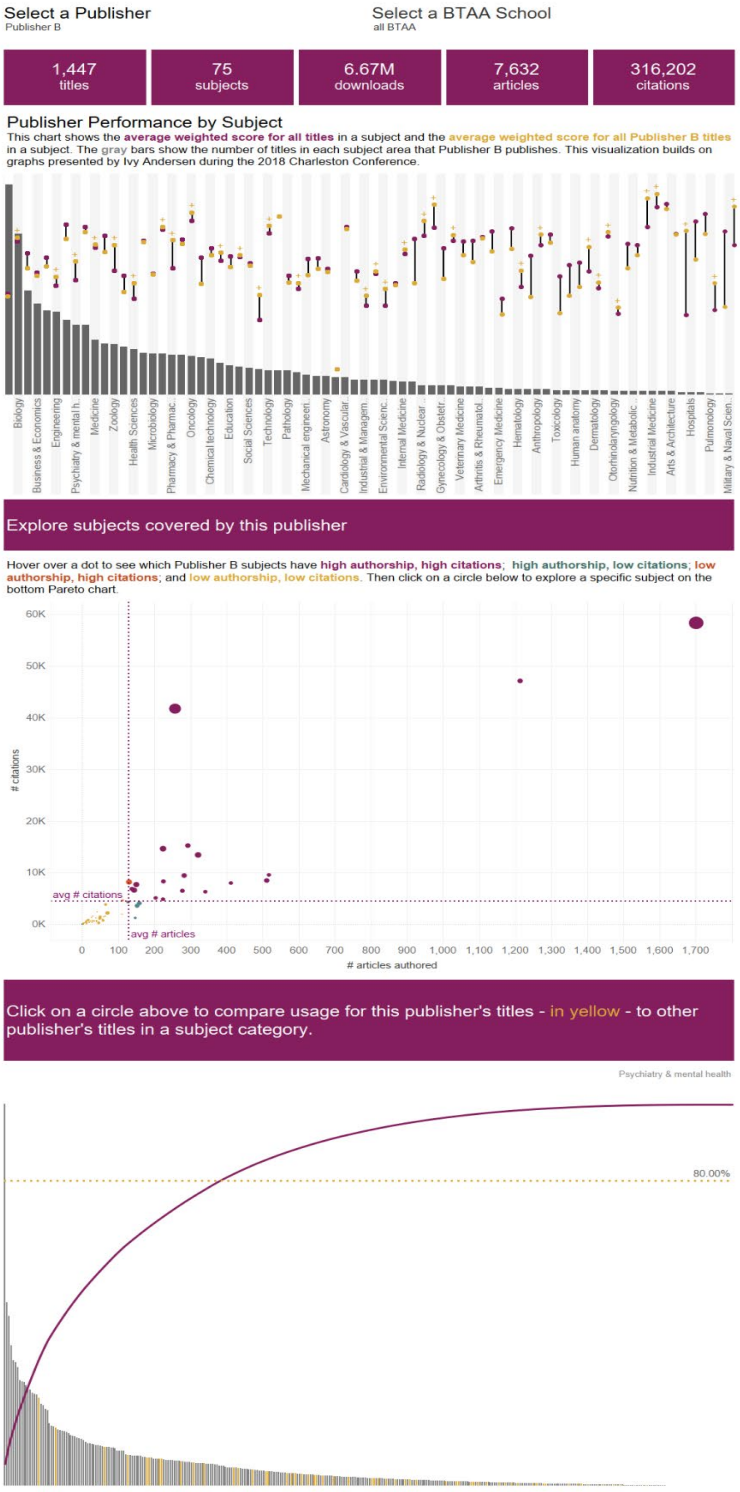
The Publisher Dashboard shows the performance of a publisher's package of titles in each subject area in relation to the performance of all titles in a subject area (Figure 3). Filters allow the user to select a publisher and view the data either collectively, for all BTAA schools, or for one individual school. Callout number boxes display the total number of titles, subjects covered by these titles, downloads, articles authored, and citations for the selected publisher. A dumbbell chart is then provided to convey how a publisher's average weighted score compares to the average weighted score for all publishers in a subject grouping. This chart was inspired by a visualization compiled by

Ivy Anderson for the 2018 Charleston conference (Anderson 2018). Anderson's "Publisher A Journals by Subject" chart used lines to display the two weighted scores for a single subject and bars to show the number of titles in a publisher's collection dedicated to that subject. The dumbbell chart builds on Anderson's combination chart by visually aligning the bars showing the number of titles published with a dumbbell line that clearly shows the gap between the publisher's average weighted score and the average weighted score for all titles. The yellow '+' sign further enhances the visualization by denoting where the publisher's average weighted score outpaces the average weighted score for all titles in a field, indicating the titles included in the publisher's package are significant for that academic discipline.

A quadrant chart sits immediately below the dumbbell chart, allowing users to explore all subjects covered by the selected publisher. Organized using an Eisenhower-Box decision-making matrix, the chart helps organize decision priorities by arranging subject categories into four quadrants: high authorship, high citations; high authorship, low citations; low authorship high citations, and low authorship, low citations. Journals in a subject category with low article authorship and a low number of citations in the bottom left quadrant of the graph might be expendable, while journals with a high number of downloads and high citations in the upper right quadrant of the graph might be necessary to keep or license. Circle size denotes the total number of downloads for the titles in each subject category for the selected publisher. When the user hovers over a circle on this chart, a narrative tooltip appears, sharing the number of downloads, as well as the number of articles authored by and cited by BTAA authors in the publisher's collection of journals in the subject area.

A Pareto chart rests on the bottom of the Publisher Dashboard. This chart is filtered by subject when the user clicks on a subject on the quadrant chart immediately above. The chart allows users to compare usage for the selected publisher's titles to titles of other publishers in the same subject area. The subject is identified in the top right corner of the graph and the graph is sorted in descending order by the number of downloads for each title in the subject area, regardless of publisher. The bars for the selected publisher's titles are colored yellow to show where they appear in relation to the other titles. The user can hover over the intersection of the berry colored cumulative percent of total line and the dotted yellow constant line to see how many titles in a subject area constitute 80% of total use. In the psychiatry & mental health subject, for example, BTAA schools downloaded 80% of articles from 102 out of 344 (29.6%) titles.

Figure 3. The Publisher Dashboard[1]

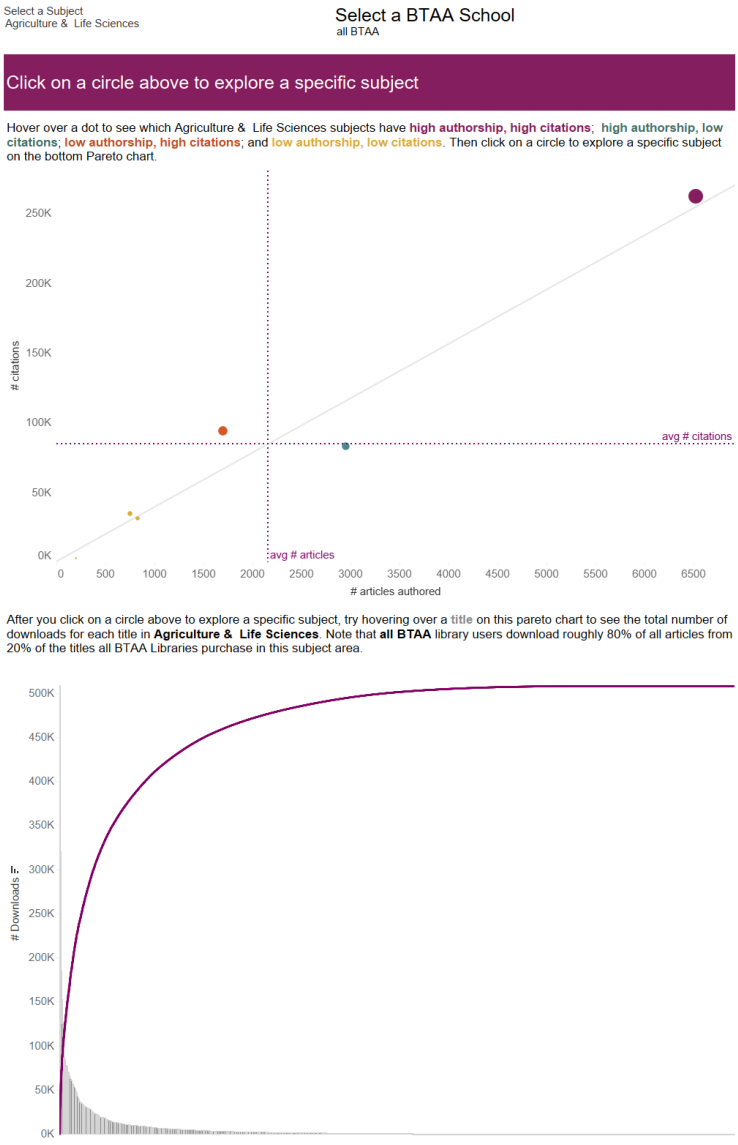


[1] Callout numbers on the publisher dashboard are masked with random numbers.

The Subject Dashboard

The Subject Dashboard mirrors the Publisher Dashboard above, but is designed to filter the quadrant and Pareto charts using broader subject categories. The authors consolidated the original 81 modified LC subject categories to a list of ten broad subjects to allow users to explore subject categories in more detail. This dashboard was created after realizing the modified LC subjects may be too specific for meaningful comparisons, especially for the health sciences. The trend line on the quadrant chart provides a linear model for the data. Tooltips on the trend line indicate how well the data fits the model by providing a R-squared measure. A p-value is also included, indicating significance.

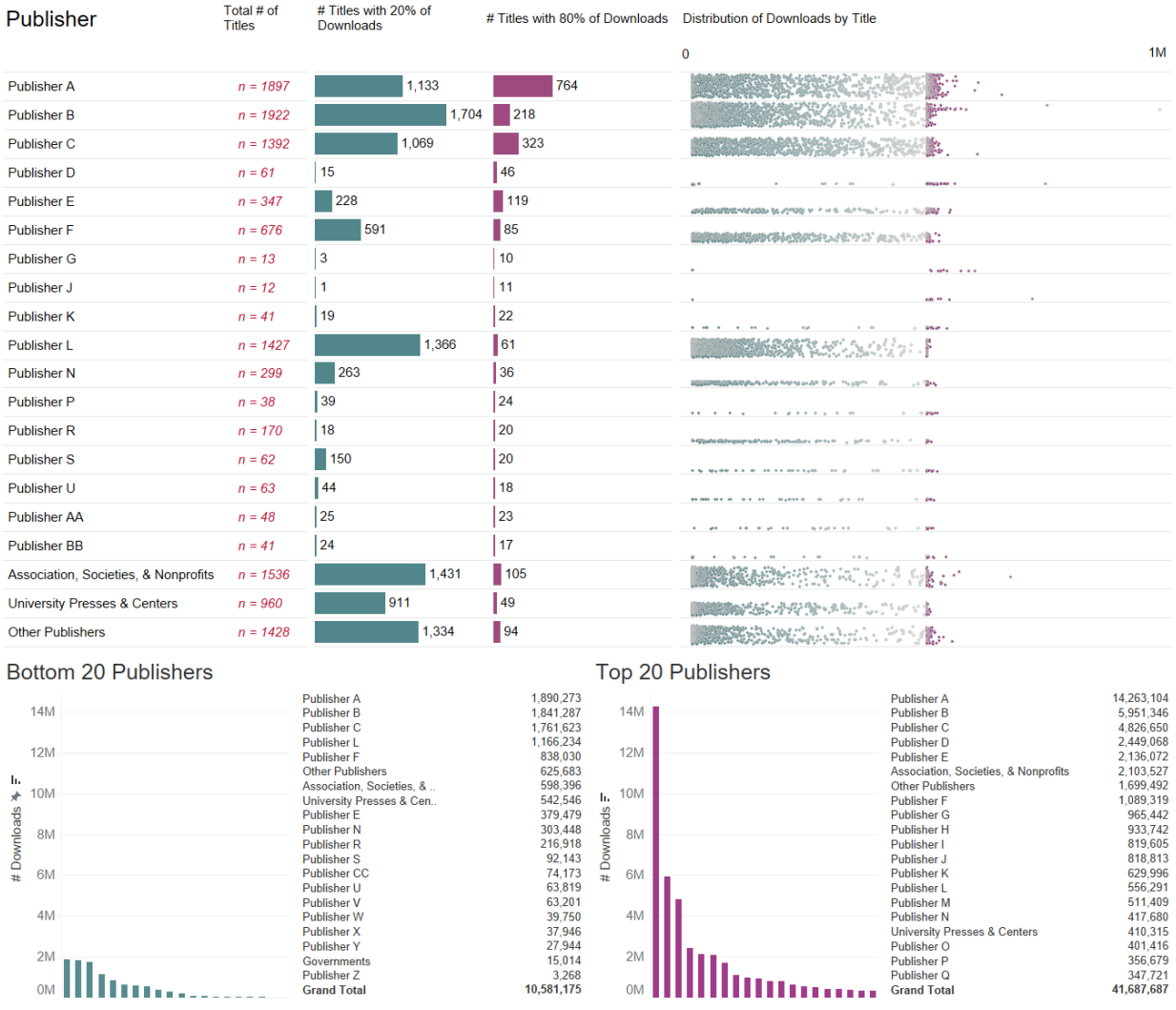
Figure 4. The Subject Dashboard



The Summary Dashboard

The Summary Dashboard broadly answers the authors' questions by publisher package. The top left of the dashboard displays the total number of titles for each publisher. Two horizontal bar charts in the middle of this section display the total number of titles in a publisher's collection with 20% of downloads, and the total number of titles with 80% of downloads. Each dot on the jitter plot to the right of the horizontal bar charts represents one title, and shows the number of downloads for that title. Throughout the dashboard, titles colored teal are in the bottom 20% of downloads. Titles assigned the berry color are in the top 80% of downloads. This chart visually shows that a few publisher package titles, including several outliers, account for 80% of all use.

Figure 5. The Summary Dashboard, with Publishers Sorted by the Number of Titles with 80% of Downloads



Analysis

While the 80% threshold of total use is reached at 44,576,655 downloads, the vertical bar chart on the bottom right of the Summary Dashboard shows that 93.5% of titles generating 80% of downloads are published by just 20 publishers. Across all publishers, titles with 80% of all use represent only 15.8%, or 2,169 of the 13,770 journal titles in the entire dataset. This percentage is smaller than the 80/20 distribution found in previous studies. The top three publishers with titles generating 80% of downloads—publishers A, B, and C—were all large commercial entities and are represented in the berry colored horizontal bar chart located in the top middle section. Publisher E, the highest ranking university publisher followed, along with a broad grouping of association, society, and nonprofit publishers, and a broad grouping of other publishers. Publishers F and L, also large commercial publishers, albeit with far fewer titles also ranked high on the list, followed by a broad grouping of university presses and centers, and publishers D and N, two large society publishers with vastly different download profiles. The remaining top 20 publishers with titles generating 80% of all titles represent various independent association, society, and non-profit publishers, as well as university presses and centers with large publishing programs.

Alternatively, examining the total number of downloads for titles generating 80% of use by publishers using the berry colored vertical bar chart on the bottom right of the summary dashboard reveals a slightly different profile. The big three commercial publishers remain at the top of this list, with Publishers B and C merely trading places, but two publishers with less than 10 titles—Publishers G and I—make the list of top 20 publishers, ranking 9th and 11th respectively. University Publishers E and N remain on this chart, along with most of the top association, society, and non-profit publishers, confirming the publishers with a small, focused collection of titles often significantly impact an academic library's collection.

Across all publishers, 11,601 or 84.2% of the 13,770 journals included in the dataset generated 20% of total use (n = 11,143,988). A descending sort of the teal horizontal bar chart located in the top middle section shows the number of titles in the bottom 20% of use for the top 20 publishers with titles generating 80% of all use. Of interest, publishers A, B, C, F, and L—all large commercial publishing houses—collectively produced just over 50% of the 10,368 titles visualized on this chart (n=5863). The broad groupings of miscellaneous association, society, and nonprofit publishers ranked 2nd on this list with 1431 titles, university presses and centers were 7th with 911 titles, and the two large university presses—Publishers E and N—were ranked 10th and 9th respectively with a combined 491 titles.

Analyzing the total number of downloads for titles generating 20% of use by publishers using the teal colored vertical bar chart on the bottom left of the summary dashboard reveals the three large commercial publishers A, B, and C again remain at the top of list,

with publishers F and L immediately behind. The two large university press publishers E and N rank 9th and 10th respectively, yet there is much more variance among societies. Several of the societies with titles generating 80% of downloads are not represented on this list—including publishers D, G, I, J, K, O, and Q. Instead, Publishers R, U, AA, and BB make the list of the top 20 publishers with titles classified as generating 20% of use, sorted by total number of downloads.

It is revealing to compare the percentage of a publisher's titles that are in the top versus the low tier of use, as represented by the jitter plot on the top right of the summary dashboard. Each dot on a jitter plot represents one title and the number of downloads for each title is plotted across the horizontal axis. Berry colored dots represent titles with 80% of downloads and teal colored dots represent titles with 20% of downloads. A quick look at the underlying data across the three Big 10 schools reveals that 60% of publisher A's titles, 89% of publisher B's titles, 77% of publisher C's titles; 87% of publisher F's titles, and 96% of publisher L's titles were represented by the teal colored dots. For the top three society publishers producing more than 13 titles, only 25% of publisher D's titles, 8% of publisher J's titles, and 23% of publisher G's titles were represented by the teal colored dots. For the two largest university publishers, 66% of publisher E's titles and 88% of publisher N's titles were represented by the teal colored dots.

Of particular note is that the rate of low use for big university press titles is similar to that of big commercial publishers. The percentage of low use for the big association, society and nonprofit titles, by comparison however, is significantly smaller, even with these publisher's shorter title lists. This echoes Ted Bergstrom's finding decades ago that society publications are often of greater value on a per-title basis than commercial publications (Bergstrom 2001).

Publisher Dashboard

The publisher dashboard's dumbbell chart shows the gap between the average weighted CDL score for all titles in a subject and the average weighted CDL score for a publisher's titles in the same subject. Looking across each chart by publisher, we can explore how well a publisher's titles score, by subject, compared to other titles (Table 1).

Table 1. Exploration of Publisher Dashboard's Dumbbell Chart, Comparing Average Weighted CDL Score for All Titles in a Subject to the Average Weighted CDL Score for a Publisher Included on Summary Dashboard

Type of publisher	Publisher	CDL score (for most subjects)
Commercial	Publisher A Publisher B Publisher C Publisher V Publisher BB	Higher than average
Commercial	Publisher F Publisher L Publisher T Publisher X Publisher W Publisher Y Publisher Z	Lower than average
Society	Publisher D Publisher G Publisher H Publisher I Publisher J Publisher M Publisher O Publisher P Publisher R Publisher U Publisher AA	Higher than average
Society	Publisher Q	Mixed
University Press	Publisher E Publisher S	Higher than average
University Press	Publisher N	Mixed

Subject Dashboard

The differences in teaching and scholarship among disciplines makes constructive comparison among disparate subjects difficult. Authorship and citation patterns in education versus physiology journals differ significantly, for example. The quadrant chart on the subject dashboard, however, visually shows that over a broad discipline's composite subjects, some basic observations may inform collection management. When the total number of citations and the total number of articles are plotted for all 81

modified LC subject categories and all three BTAA schools are included in the view, the trend model shows a significant linear relationship (R-squared = 0.851001; p-value < 0.0001) (Table 2). Selecting a specific broad subject category for each of the BTAA schools, however, reveals three broad subjects showing a significant linear relationship between the number of articles cited and the number of articles authored: health sciences (R-squared = 0.808377; p-value < 0.0001), physical sciences & engineering (R-squared = 0.904315; p-value < 0.0001), and social sciences (R-squared = 0.944588; p-value = < 0.0001).

Table 2. Overview of Subject Dashboard's Quadrant Chart Model by Broad Discipline

Broad Discipline	Total # of articles authored	Total # of citations	High authorship and high citation	Low authorship and low citation	R-Squared	p-value
Agriculture & Life Sciences	12955	508243	Biology	Environmental Sciences Botany Zoology	.955927	.0007394
Arts & Humanities	2455	20285	Language & Literature History	Arts & Architecture Philosophy & Religion	0.941625	0.0296266
Health Sciences	38318	1079324	Physiology Medicine Psychiatry & Mental Health Surgery Oncology	Forensic Medicine	0.808377	<0.0001*
Physical Sciences & Engineering	21250	768721	Physics Chemistry	Military & Naval Science Aero. Eng. & Astronautics	.904315	<0.0001*
Social Sciences	8206	268906	Business & Economics Psychology Sociology	Library Science Recreation Political Science Anthropology Geography	0.944588	<0.0001*

*p-value <0.0001 = significant

Discussion

Robust data visualization software greatly aids the analysis of large datasets. The ability to not only relate and refresh data sources, but to create dynamic, interactive dashboards is a huge advantage of Tableau. Specifically, for this project Tableau allowed the authors to view trends over time for multiple subjects, understand patterns of use, and show library contributions to interdisciplinary research both at their individual and collective schools. A previous iteration of a subject dashboard assembled by two of the authors prior to this project showed how important psychology journals were to business researchers (Schoenborn and Stemper, 2018). The dashboards created for this project reaffirmed the importance of psychology journals within the social sciences. These titles rose to the top of the social sciences subject dashboard when comparing authorship and citation patterns.

It is important for libraries to demonstrate their value proposition to college deans and university administrators. In addition to the usual download statistics and ARL holdings figures, having the title, publisher, subject, and summary dashboards is a helpful addition to our assessment toolbox. Each dashboard uniquely acknowledges differences in disciplinary research practices and gives subject coordinators the means to analyze their collections in depth. The CDL score, by providing a single number that combines a title's download, articles authored, and citation metrics, greatly simplifies the comparison of titles, publishers, and subjects.

Since the summary dashboard confirms that a minority of titles generates the majority of the downloads across most publisher packages, it makes sense to focus on adding and removing titles from the collection by subject rather than by publisher. Further focusing at the consortial level, rather than at the institutional level, would give schools their best leverage in the marketplace. It may be advantageous to consider more limited, subject-oriented title packages for big commercial publishers. Not all of the titles academic libraries subscribe to are indexed in Web of Science, a limitation of this study. The phrase "less frequently used" is also relative; for example, the Oxford title Journal of Neuropathology and Experimental Neurology is classified in the lowest 20% of use category, yet still amassed 5276 downloads over three years. Niche subject strengths for each university must be considered and the proper mode of access–subscription or interlibrary loan–adopted.

To maximize return on consortial investment, as informed by downloads, articles authored, and citation data for the three BTAA schools, titles in the following ten subjects may be advantageous for collective purchase.

1. Biology
2. Physics
3. Physiology
4. Chemistry

5. Surgery
6. Medicine
7. Psychiatry & Mental Health
8. Oncology
9. Business & Economics
10. Psychology

Select publishers' title lists in these categories outpace all titles within the category, suggesting the highest return on investment. Conversely, collections of publisher's titles scoring lower than average in relation to other titles within a subject area represent an opportunity to shed titles and redirect resources.

The Tableau dashboards created for this demonstration project are available upon request to BTAA research libraries. The authors welcome adding other BTAA schools metrics to the data model. While the data analyzed was limited to the 2015–2017 calendar years, the dashboards can be quickly updated and refreshed with more recent data. A data-free template of the Tableau workbook may also be available for adaptation by other universities in the near future, and rough instructions outlining how to gather, model, and prepare data for analysis are provided in Appendix A. Lastly, usability testing with a broader contingent of liaisons, subject coordinators, and collection strategist is a natural next step for continued development of this project.

References

Anderson, Ivy. 2018. "Data Expeditions—Mining Data for Effective Decision-Making" (Presentation at the Charleston Conference)." YouTube. https://www.youtube.com/watch?v=PR4nf7zjEi0&list=PLIGLt62pr1M7qeeElNTWhQLdrCFYA08ow&index=4.

Belter, Christopher W., and Neal K. Kaske. 2016. "Using Bibliometrics to Demonstrate the Value of Library Journal Collections." *College & Research Libraries* 77 (4): 410–22. 10.5860/crl.77.4.410.

Bergstrom, Theodore. 2001. "Free Labour for Costly Journals?" *Journal of Economic Perspectives* 15 (4): 183–198. 10.1257/jep.15.4.183.

Bosch, Stephen, Barbara Albee, and Kittie Henderson. 2018. "Death By 1000 Cuts | Periodicals Price Survey 2018." *Library Journal*. https://www.libraryjournal.com/story/death-1000-cuts-periodicals-price-survey-2018.

Chew, Katherine, Jim Stemper, Caroline Lilyard, and Mary Schoenborn. 2012. "User-defined valued metrics for electronic journals." Library Assessment Conference. http://old.libraryassessment.org/bm~doc/Chew_Katherine_2012.pdf.

Finch, Jannette L. 2016. "Using Data Visualization to Examine an Academic Library Collection." *College & Research Libraries* 77 (6): 765–768. https://doi.org/10.5860/crl.77.6.765.

Gatten, Jeffrey N., and Tom Sanville. 2004. "An Orderly Retreat from the Big Deal: Is it Possible for Consortia?" *D-Lib Magazine* 10 (10). http://www.dlib.org/dlib/october04/gatten/10gatten.html.

Jurczyk, Eva, and Pamela Jacobs. 2014. "What's the Big Deal? Collection Evaluation at the National Level." *portal: Libraries and the Academy* 14 (4): 617–631. https://doi.org/10.1353/pla.2014.0029.

Knowlton, Steven A., Adam C. Sales, and Kevin W. Merriman. 2014. "A Comparison of Faculty and Bibliometric Valuation of Serials Subscriptions at an Academic Research Library." *Serials Review* 40 (1): 28–39. https://doi.org/10.1080/00987913.2014.897174.

Li, Chan, and Holly Eggleston. 2016. "CDL Journal Weighted Value Algorithm Assessment & Revisions." California Digital Library. https://cdlib.org/wp-content/uploads/2019/04/CDL_Journal_Weighted_Value_Algorithm_Assessment_And_Update_2016.pdf.

Murphy, Sarah A. 2015. "How Data Visualization Supports Academic Library Assessment: Three Examples from The Ohio State University Libraries Using Tableau." *College & Research Libraries News* 76 (9): 482–6. https://crln.acrl.org/index.php/crlnews/article/view/9379/10546.

Schoenborn, Mary, and Jim Stemper. 2018. "Benchmarking the Value of Business Research: The University of Minnesota Libraries' Model." Charleston Conference. https://doi.org/10.26226/morressier.5b8d1f8c21c0500063b98475.

Stemper, James A., and Janice M. Jaguszewski. 2003. "Usage Statistics for Electronic Journals: An Analysis of Local and Vendor Counts." *Collection Management* 28 (4): 3–22. https://doi.org/10.1300/J105v28n04_02.

Wiersma, Gabrielle. n.d. "Collection Analysis With Tableau." Presentation at the Colorado Alliance of Research Libraries Library Assessment Workshop, November 17, 2016. Accessed June 29, 2022. https://coalliance.org/sites/default/files/WiersmaG.pdf.

Wissel, Kathryn M., and Lisa DeLuca. 2018. "Telling the Story of a Collection with Visualizations: A Case Study." *Collection Management* 43 (4): 264–75. https://doi.org/10.1080/01462679.2018.1524319.

Appendix A. Data Preparation

Step 1. Gather and Model Data

Using raw data from the Web of Science XML dataset, the team identified articles authored by BTAA researchers between 2015 and 2017 by searching for variations of their three schools names in the organization and full-address fields (articles authored). Results were filtered to only include articles listed in the Arts & Humanities, Science Citation Index Expanded, and the Social Sciences Citation Index (articles authored = 86,021). To determine which WOS indexed journals were cited by BTAA authors, the team then harvested the citation list associated with each BTAA authored article (citations = 2.82M). Download data from the three BTAA schools was added to each title using COUNTER reports and holdings data to identify which titles were acquired by each school, either individually, or as part of a publisher package (downloads = 55.72M). All data was structured to allow users to view data on the Publisher Dashboard and the Subject Dashboard by individual school, or the three BTAA schools collectively.

To filter the dashboards meaningfully, the data was then enhanced by adding a LC classification number to each journal title. This number was then used as a guide to assign one of 76 broad subject categories to each title. Subjects included:

Aeronautics
Engineering & Astronautics
Agriculture
Allergy & Immunology
Anthropology
Arthritis & Rheumatology
Arts & Architecture
Astronomy
Biology
Botany
Business & Economics
Cardiology & Vascular Medicine
Chemical technology
Chemistry
Civil Engineering
Dentistry
Dermatology
Education
Electrical engineering Electronics
Nuclear engineering
Emergency Medicine
Endocrinology
Engineering
Environmental Sciences
Forensic medicine
Gastroenterology
General Science
General Works
Geography
Geology
Geriatrics
Gynecology & Obstetrics
Health Sciences
Hematology
History
Hospitals
Human anatomy
Industrial & Management Engineering
Industrial Medicine
Infectious & Parasitic Diseases
Internal Medicine
Language & Literature
Law
Library Science
Math & Computer Science
Mechanical engineering and machinery
Medicine
Microbiology
Military & Naval Science
Mining engineering.
Metallurgy
Nursing

Nutrition & Metabolic Diseases
Oncology
Ophthalmology
Orthopedics
Otorhinolaryngology
Pathology
Pediatrics
Pharmacy & Pharmacology
Philosophy & Religion
Physics
Physiology
Political Science
Psychiatry & mental health
Psychology
Public health
Pulmonology
Radiology & Nuclear Medicine
Recreation
Social Sciences
Sociology
Sports Medicine
Surgery
Technology
Toxicology
Urology & Nephrology
Veterinary Medicine
Zoology

The subject field was next used to assign the 2018 average cost per title found in table 3 of Library Journal's annual periodicals price survey. Publisher names were normalized to simplify dashboard filters. Lastly, all sourced data was assembled using a series of joins and relationships in Tableau's Data Source window.

Figure 6. The Data Source Window

Step 2. Prepare Data

The team modified the CDL algorithm slightly to produce the total weighted CDL score. Decile ranks for the articles authored, citations, downloads, and average cost per title metrics were calculated using a combination of level of detail expressions with aggregate, and logical calculations in Tableau.

Example of Decile Rank Calculation to Score Downloads

```
FLOAT(IF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED
[Title] :zn(SUM([Downloads])) }, .1) } THEN ".1"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .2) } THEN ".2"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .3) } THEN ".3"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads])) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .4) } THEN ".4"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .5) } THEN ".5"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .6) } THEN ".6"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .7) } THEN ".7"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .8) } THEN ".8"
ELSEIF{ FIXED [Title] : zn(SUM( [Downloads] )) } <= {PERCENTILE({ FIXED [Title]
:zn(SUM([Downloads])) }, .9) } THEN ".9"
ELSE "1.0"
END)
```

Float parameters with a minimum value of 0 and a maximum value of 1 were then created to allow users to weight each component of the final CDL score by each individual metric. This required a check be added to the final CDL score calculation, to ensure that the total weights assigned by the user cannot equal more than 1. If the weights are more than one, the user receives an error message stating "Try again. Weights must =1."

Check Weights Calculation

```
if ([downloads parameter]+[articles authored parameter]+[citation
parameter]+[average journal cost parameter])=1 then True else False end
```

Check Weights Error Message

```
if [check weights=1]=True then  ''
else 'Try again. Weights must = 1'
end
```

The final CDL score for each title was constructed by first confirming the weights are less than or equal to one. Each metric was multiplied by its weights and then added together and multiplied by 100 to display a whole number.

Final CDL Score Calculation

```
//user chooses weights
if [check weights=1]=True then
(([Score Downloads]*[downloads parameter]) +
([Score Articles Authoredt]*[articles authored parameter])  +
([Score Citations]*[citations parameter])+
([Score Average Cost Per Title]*[average journal cost parameter]))*100
end
```

The dynamic reference lines on the Publisher Dashboard and Subject Dashboard's quadrant charts adjust when the user clicks on a subject category to show the average number of citations and the average number of articles for the subject. To calculate these lines correctly, a level of detail expression is needed for both the citations and the articles authored metrics. Both calculations then need to be set to average when setting the reference line.

Level of Detail Expression for Articles Authored

```
{fixed [Subject modified],[Publisher]:sum(zn([Articles authored]))}
```

A calculation is also required to dynamically color the four quadrants on the Publisher Dashboard and Subject Dashboard's quadrant charts based on the filters selected by the user.

This means that when a user selects a new publisher on the Publisher Dashboard, the reference lines and quadrant colors will reset to reflect the new average number of citations and average number of articles authored by BTAA authors for that publisher.

Calculation to Color Quadrants

```
Results are computed along Table (across).
if sum(zn([Articles authored]))>=WINDOW_AVG(sum(zn([Articles authored]))) and
sum(zn([Citations]))>=WINDOW_AVG(zn(sum([Citations])))
then 'High Authorship, High Citation'
elseif sum(zn([Articles authored]))<=WINDOW_AVG(zn(sum([Articles authored])))
and
sum(zn([Citations]))<=WINDOW_AVG(zn(sum([Citations])))
then 'Low Authorship, Low Citation'
elseif sum(zn([Articles authored]))<=WINDOW_AVG(sum(zn([Articles authored])))
and
sum(zn([Citations]))>=WINDOW_AVG(zn(sum([Citations])))
then 'Low Authorship, High Citation'
```

```
elseif sum(zn([Articles authored]))>=WINDOW_AVG(zn(sum([Articles authored])))
and

sum(zn([Citations]))<=WINDOW_AVG(zn(sum([Citations])))
then 'High Authorship, Low Citation'
end
```