

New Developments for Journal Package Analysis and Data Visualization

Nat Gustafson-Sundell, Collections Librarian
Pat Lienemann, ER Access & Discovery Librarian
Luwis Andradi, Collection Management Technology Graduate Assistant
Evan Rusch, Gov Docs, Reference & Instruction Librarian
Jeff Rosamond, Serials Technician

Minnesota State University, Mankato
NASIG 2021



To begin, please let me thank NASIG for this opportunity to share our work. We are a team from Minnesota State University, Mankato, a regional public access university with about 14,000 full-year-equivalent students. We are the largest university in the Minnesota State System of 37 universities and colleges. Our focus at Mankato is on resources for teaching and learning, but we additionally support many graduate programs.

SEE more information about Minnesota from the land acknowledgement at the Institute for Advanced Study, University of Minnesota:

- [Local Dakota Land Map](#) — downloadable visual and audio Dakota land maps of Minneapolis, St. Paul, and surrounding areas by local artist Marlena Myles
- [Why Treaties Matter](#) — a comprehensive and thoughtful exploration of treaties and land theft in Minnesota. For our area, we recommend you begin by reading about the [1837](#) land cession treaties with the Ojibwe and Dakota, and the [1851](#) Dakota land cession treaties

Credit: <https://ias.umn.edu/about/ias-land-acknowledgement>

3

Please let me also take this opportunity to recognize that we live, work, and learn in the homeland of the Dakota people, and whose language frames our name—Minnesota State University, Mankato.

I'll also quickly share some aspects of my positionality. I'm Nat Gustafson-Sundell, a Collections Librarian and Technical Services Coordinator, in my second career and formerly the Business Manager of a software company and then a market research firm. I'm perceived as a white, straight, able, middle-aged man. My own realities are more complicated, but I know I've definitely and substantially benefited from advantages resulting from the perception of my whiteness and my gender, as well as my socioeconomic & cultural background, education, and physical attributes.



[https://link.mnsu.edu/
mankato-nasig21](https://link.mnsu.edu/mankato-nasig21)

Download a local copy, then turn on the notes to
see the text.

There is also a link to the video
at the end of the slide deck.

4

We recognize that it's not always possible to follow along with a presentation, so we've also linked a version of this presentation here.

Collection Management Technology (CMT) Sub-Group of the Journals Review Committee (JRC)

Person (time)	Primary Role	Additional Roles
Nat (.1)	Project Manager	Design & Development, Data Wrangling, Liaison Outreach
Jeff (.25 -> .1)	Data Validation Lead	Design & Development, Report Production
Pat (.03)	Data Viz Lead	Design & Development, GA Supervisor
Evan (.02)	Finished Report Design Lead	Design & Development, Liaison Outreach
Luwis (GA)	Programming Lead	Design & Development

→ Link to slides: <https://link.mnsu.edu/mankato-nasig21>

5

I'd also like to thank the members of the Collection Management Technology (CMT) Sub-group of the Journals Review Committee (JRC) for their work. Only four of us are talking today, but that has nothing to do with the relative value of our work to this project. In particular, I want to mention Jeff Rosamond, who is a key member of our team and his work is included in the results we demonstrate today. Without his work on data validation and report production, the scope of our work would be drastically reduced.

Please let me also thank the JRC, because current and past JRC members have contributed many times to our report priorities.

Automating collection analysis data visualization in Jupyter Notebook: What's possible and why would you do it. Presented at Electronic Resources & Libraries (ER&L), Austin, TX, 2020. https://cornerstone.lib.mnsu.edu/lib_services_fac_pubs/177/

Journal collection analysis: How we developed new tools to improve collection assessment, evaluation and outreach. Presented at North American Serials Interest Group (NASIG), Atlanta, GA, 2019. https://cornerstone.lib.mnsu.edu/lib_services_fac_pubs/157/

The JCA DB: Journal collection analysis and evaluation for outreach and more! Presented at Electronic Resources & Libraries (ER&L), Austin, TX, 2018. https://cornerstone.lib.mnsu.edu/lib_services_fac_pubs/150/

It's all happening in Cornerstone!

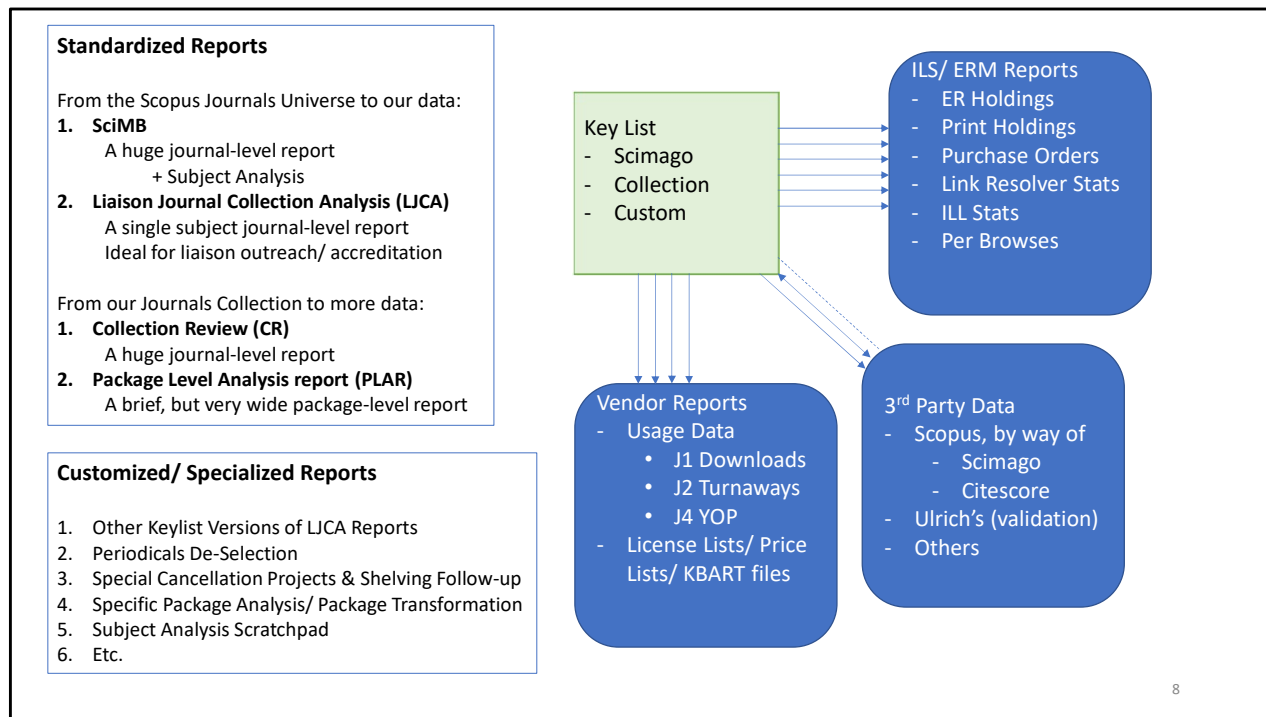
<https://cornerstone.lib.mnsu.edu/>

<https://link.mnsu.edu/Mankato-NASIG21>

6

Our report development project has now been iterating for several years. We have presented to several conferences different aspects of the project, some of which are listed on the screen as examples, in case you are interested in other applications or in learning more about how we do this work. The how will not be a focus of this presentation.

Because this is a long-running project, our results have steadily improved, growing much deeper and ever more cost-effective to produce. A single one of our standard reports takes less than 2 hours to produce. Asterisk



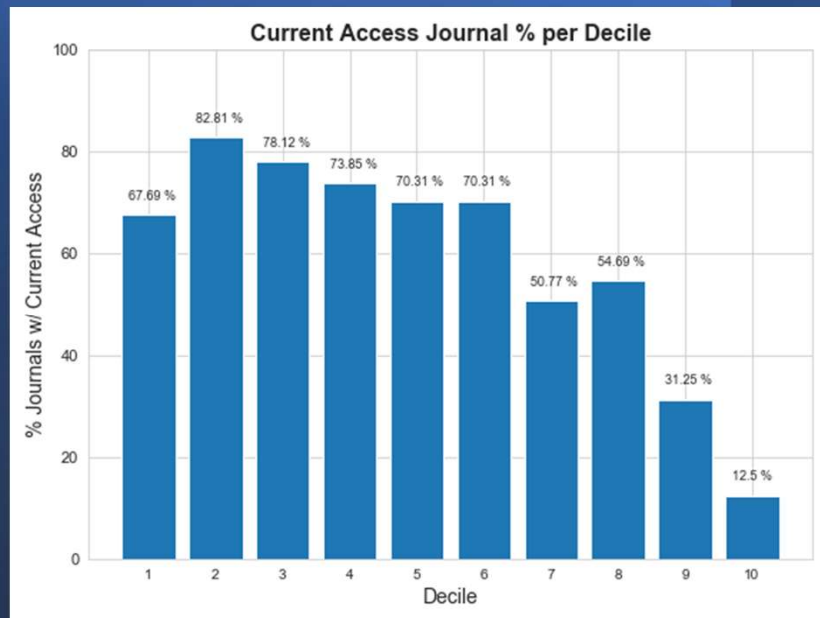
So here's the lightning fast background. Our CMT group has developed 4 standard reports, one for journal-level collection review, a second for package-level review, a third for liaison outreach and accreditation support, and a fourth which allows us to see a universe of journals and our place within it. This last report can also be sliced at the level of any number of subjects. All of these reports are standardized so that they can be produced cheaply using automated methods. We also produce customized reports for other jobs using the same underlying automation.

The basic concept behind these reports is that we match from a key list of journals to all of the relevant data available to us. Today, we are focusing on what we can do with 2 of these reports. The first is the PLAR. You'll hear us saying the word 'PLAR' often today – that stands for the Package Level Analysis Report. The PLAR is based on a different report, which we call the Collection Review report. The Collection Review is a journal-level report including all of our subscription journals, whether they are in packages or not. For the PLAR, we just roll all of the journal level data up to the package level, and then we augment that report specifically with additional package-level only data.

The other report we'll see today is what we call the SciMB, which is based on the journals listed in Scimago, a free online resource describing a universe of journals, with data derived from Scopus. When we present any of our subject analysis, we are basing that work on the SciMB.



I want to stress that these reports have vastly improved our opportunities for incisive decision-making, and these reports have *created* opportunities for outreach to each other and academic departments. We often encounter the assumption that these reports must be expensive or difficult to produce, but these reports are produced at low cost because our processes have grown increasingly automated and streamlined. In some cases, we have reduced the costs of previous work by as much as 90% or more.



In fact, this whole project, which has now run for several years was started just because Evan and I decided to explore a cost-reducing approach to some of the holdings reports he previously created for himself and his academic departments as a liaison.

I should also mention we prefer making these reports ourselves, rather than relying on more limited commercial products, because the skills and teamwork developed in this work have paid forward into other projects in many different ways. In my view, data competence is one of the essential skills of our time, and this project also builds data competence.



That said, we've come to see an essential limitation to our impact.

We have presented our reports to colleagues several times now, but only a few can take full advantage of them, and most would not feel confident presenting the reports to their liaison departments. In some ways, the reports are successful for communication purposes because they inspire confidence in our stewardship, even if the details aren't always fully grasped by our audiences. As an additional driver of our work this year and next, we've realized we need to develop more effective education and outreach strategies.



So now, let me provide a quick background and timeline for what we're presenting today.

We started the year in August expecting the unexpected. I had planned an early focus on package analysis in case we needed to make room in the budget quickly, but we didn't actually need to take any emergency steps this year. Instead, we've seen a longer-term process unfolding.

Our priorities shifted. We decided to slow down and think through what we really wanted from this process. Instead of focusing on decision-making and presenting our results for acceptance within the year, which is what we did in the past, we're more focused now on developing a process to educate our colleagues as liaisons to interact more meaningfully across campus.

Degree	AY17	AY18	AY19	AY20	AY21 >Dec
Accounting - BS	68	86	65	87	46
Accounting - MACC		14	19	15	16
Accounting - MS	29				
Aging Studies - GC			3		
Aging Studies - MS	3	1	5	4	
Alcohol and Drug Studies - BS	10	25	24	28	18
American Indigenous Studies - BA				1	
American Indigenous Studies - BS	3			1	1
American Indigenous Studies - CERT			1		
Anthropology - BA	2	4	4		
Anthropology - BS	5	7	4	10	3
Anthropology - MS	3	3	4	8	
Applied Health Science - BS	29	43	30	27	6
Applied Leadership - BS	25	32	30	38	27
Applied Statistics - MS	7		4	2	4

13

So, given a longer timeline for action than we initially expected, we developed the following overall concept. We will prioritize our packages for retention or cancellation through a process that started this year and which will continue through next year. The immediate premise for discussion right now is that I might need to change license terms for any multi-year deals, so that any package prioritized for cancellation can be actionable. Included among the cancellation priorities will be at least one big deal – we currently have big deals with all of the big 5 journal publishers, but we definitely foresee that one will have to go. The bigger picture, of course, is that our collections must be aligned as much as possible with curricular needs. We can look at the student enrollment and completion data, but we also need improve how we work with our academic programs.

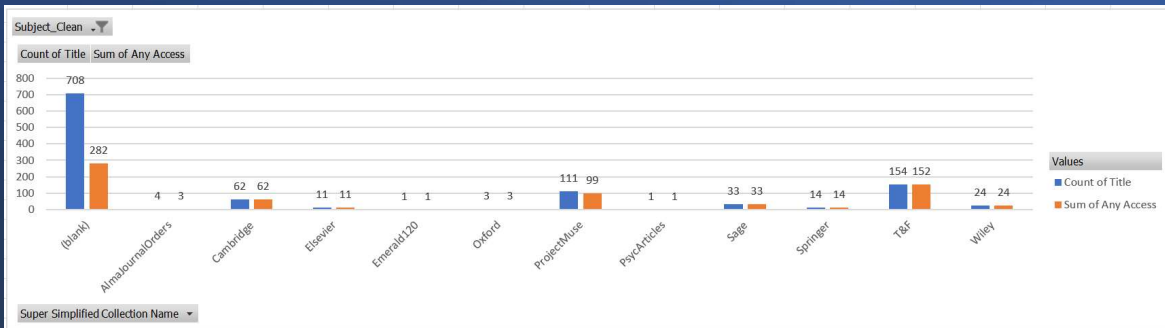
Package Level Analysis Report (PLAR)

<https://link.mnsu.edu/package>

2020 Collection	Collection Identifier	Count of Titles For All Matched Journals (FAMJ)	Count of Any NonSub Coverage Overlap	% of Any NonSub Coverage Overlap	Count of Presumed NonSub Coverage <1 Year or less moving wall> Overlap	% of Presumed NonSub Coverage <1 Year or less moving wall> Overlap	Count of Presumed Current NonSub Coverage Overlap	% of Presumed Current NonSub Coverage Overlap	Count of Best Quartile, Q1	Count of Best Quartile, Q2
Small Journal Collection	AAA	185	130	70%	111	60%	4	2%	54	55
Big Journal Collection 1 (All)	BBB_All	2119	882	42%	792	37%	58	3%	710	689
Big Journal Collection 2 (Core Collection)	CCC_Core	116	77	66%	64	55%	12	10%	70	26
Big Journal Collection 2 (Access)	CCC_Access	1187	606	51%	548	46%	102	9%	625	343
Big Journal Collection 2 (Combined for analysis)	n/a	1303	683	52%	612	47%	114	9%	695	369

I see two primary means to prioritize packages. One is to compare packages at the package level, to see how they provide value relative to one another. For this, we have the PLAR. We'll look at this a little more closely in a moment.

Subject Analysis Scratchpad



15

And the second is to consider subject and program level impacts. This is especially important for communication with campus and will also inform how we transform any cancelled packages by adding back some individual subscriptions. For subject level analysis, we have multiple means, including a new tool called the Subject Analysis Scratchpad, although Pat will present more sophisticated subject analysis.

The combination of these methods gives us a holistic view of the value and performance of our packages. Of course, we can also implement the kind of single package analysis available through UnSub, which might be familiar to many of you as a low-cost solution for package transformation, but I think it's very important to perform comparative analysis at the package and subject levels, and across wider ranging data than either UnSub or the ILS can handle, especially to support communication across campus.



In addition to working through committee for package discussions this year, I scheduled an open meeting for all librarians this semester. In this meeting, we received feedback that a workshop would be helpful, so we scheduled one for the end of the semester, and we'll schedule another for the beginning of the fall semester. We now see these workshops as a focus of our process, because it's here that we can enable our colleagues to work as more effective liaisons for journal collection development. We also hope these workshops will lead to substantial feedback for Pat, to inform his work.

Testimonials

CIVE (Heidi Southworth)

"The ASCE journals? We would be totally lost without them; it's a core component to what we teach our students, a building block of our research, and a mainstay of our mission as a public regional university. Our accreditors ABET would pop a major gasket if we didn't have access." Steve Druschel, CIVE, Nov, 2020

ECET (Heidi Southworth)

"Students in my classes have always been using the IEEE Xplore Digital Library extensively (<http://ezproxy.mnsu.edu/login?url=http://www.ieee.org/ieeexplore>). As my courses taught are related to Wireless Networks/Computer Networks/Machine Learning Applications, I always have different home/project assignments in my courses that need students to download front-end research articles from IEEE Xplore Database substantially to learn knowledge about the emerging technology development from both academia and industry. Moreover, I also have different undergraduate/graduate student helpers working with me on my research projects, for which the IEEE Xplore Digital Library has been the main source to use." Qiang Yu, ECET, Dec, 2020

17

One goal driving our work has always been to provide reports that provoke conversations. Through the workshops, we're striving ever more to put these tools into the hands of our liaison colleagues, and we will increasingly ask them to garner more substantial responses from their contacts, especially as we move through the fall. As one example, we have started soliciting testimonials to help us understand how our journal packages are being used, and these testimonials serve, in the first place, to round out and balance our reliance on numbers, and, in the second place, these testimonials can help the library to communicate with our administration.



Ok, so now let's take a super quick look at two of the reports we are supplying to liaisons. The first is the PLAR. If you are interested in this report, I've provided an example version through the link. This version does not include the data viz, but it does have a rough glossary to describe the variables.

After the PLAR, I'll also very quickly display the Subject Analysis Scratchpad. I'm not providing an example of this one because it is very large and includes data we need to think more deeply about whether we should share.



Thank you Nat.

Just a quick introduction – my name is Pat Lienemann (my pronouns are he/him). I am the Electronic Access and Discovery Librarian for MNSU. I just finished my third year at Mankato and as a part of this team.

I joined the Collection Management Team after Jeff, Evan, and Nat had been working together for quite a while. As a beginning faculty member, they invited me to work with them, I think without really considering how I could contribute towards the team. Fortunately for me, during grad school, I happened to take an elective covering Digital Humanities tools and offered to try my hand at creating Data Visualizations of our work

(I'll probably fall into calling these "dataviz" right away, so I wanted to warn you ahead of time).

Tableau for Teaching
<https://www.tableau.com/academic/teaching>

ACADEMIC PROGRAMS

Tableau for Teaching

Teach data skills across all disciplines and levels in your classroom for free.

REQUEST INDIVIDUAL LICENSE REQUEST COURSE SOFTWARE

We're on a mission to create a data literate world

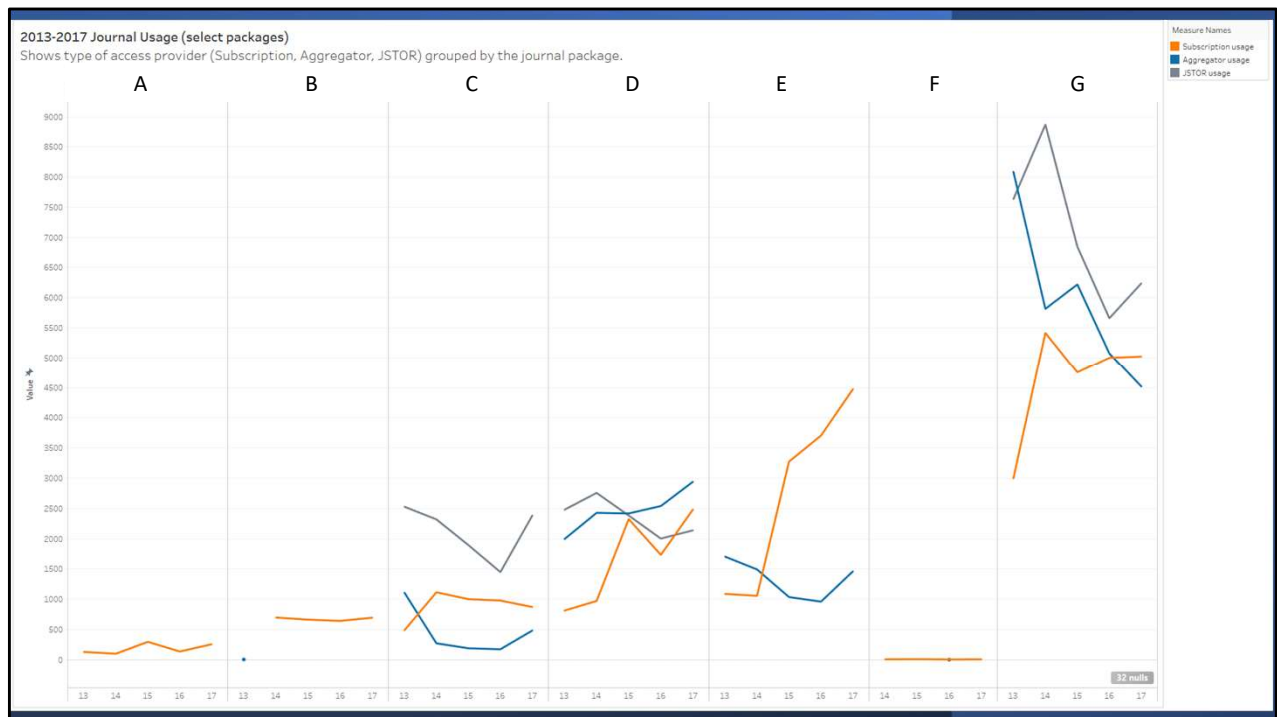
Since 2011, Tableau's Academic Programs have enabled more than 1.5 million students and teachers from accredited institutions around the world, across disciplines, with critical data skills. Tableau for Teaching (TTF) offers complimentary software, learning resources, and curricula to help instructors teach data skills that are essential in today's global economy. Get started today by requesting your [free 1 year license!](#)

20

Although there are many different tools available, when I began I focused on using Tableau* and it's still my go-to visualization software. As a team we see a lot of promise and think the future of our project is in coding our own analyses and visualizations. Luwis will talk about our thought and decision process about that and the way our development is heading.

However, although I aspire to learn python, I don't know it at this time. So in the meantime, Tableau is powerful and useful and it's good for **developing visualizations/prototyping**.

*Just a quick note: Tableau is a paid product, but they provide complimentary licenses to educators and researchers, which is how I utilize the software.



Just a brief history of what we've done/ how we utilized the viz in the past

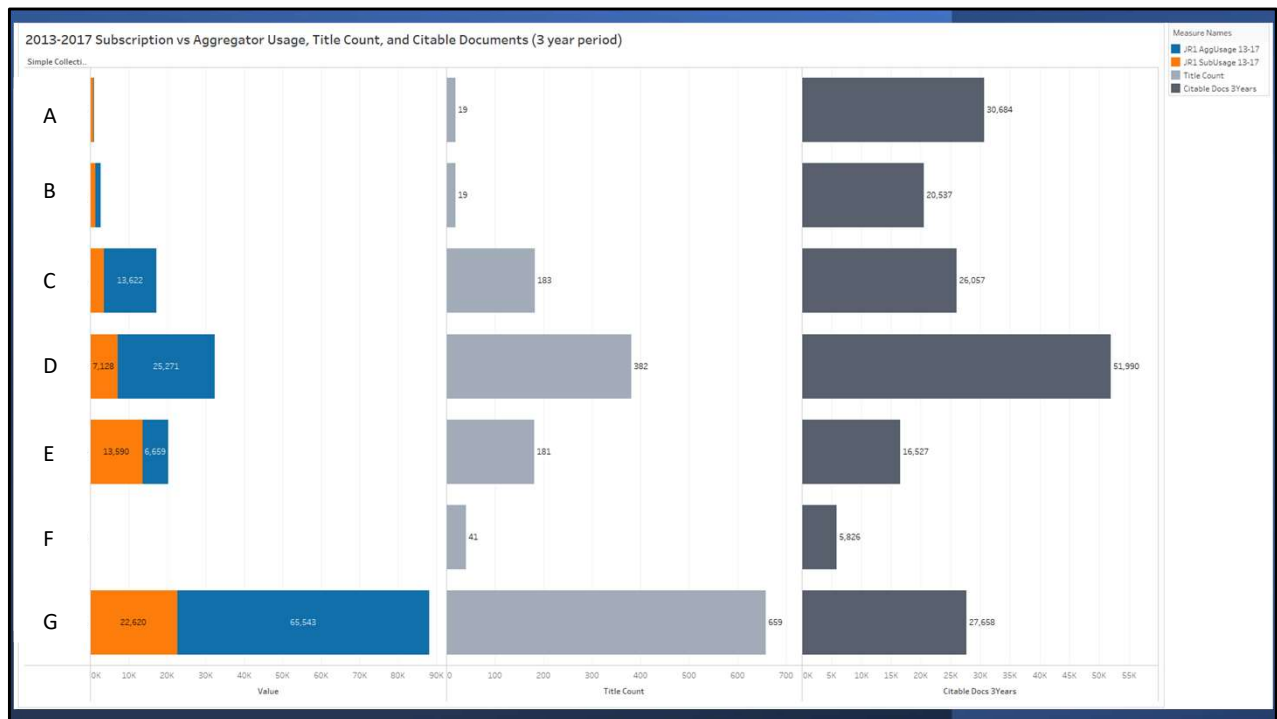
When I first started on the team, we were working with the package level data

- knew we'd have to cut
- knew we'd want to bring back some of the journal titles

So some of the things we thought were useful, like this viz:

- Shows a 5 year usage trend
 - Broken down by Usage type (Subscription - orange, Aggregator – blue, or JSTOR - grey)
 - Where are our users accessing this journal content
 - Is it the sub? [Raises questions -why or why not/does that mean we can cut it/is there a discovery issue/etc]

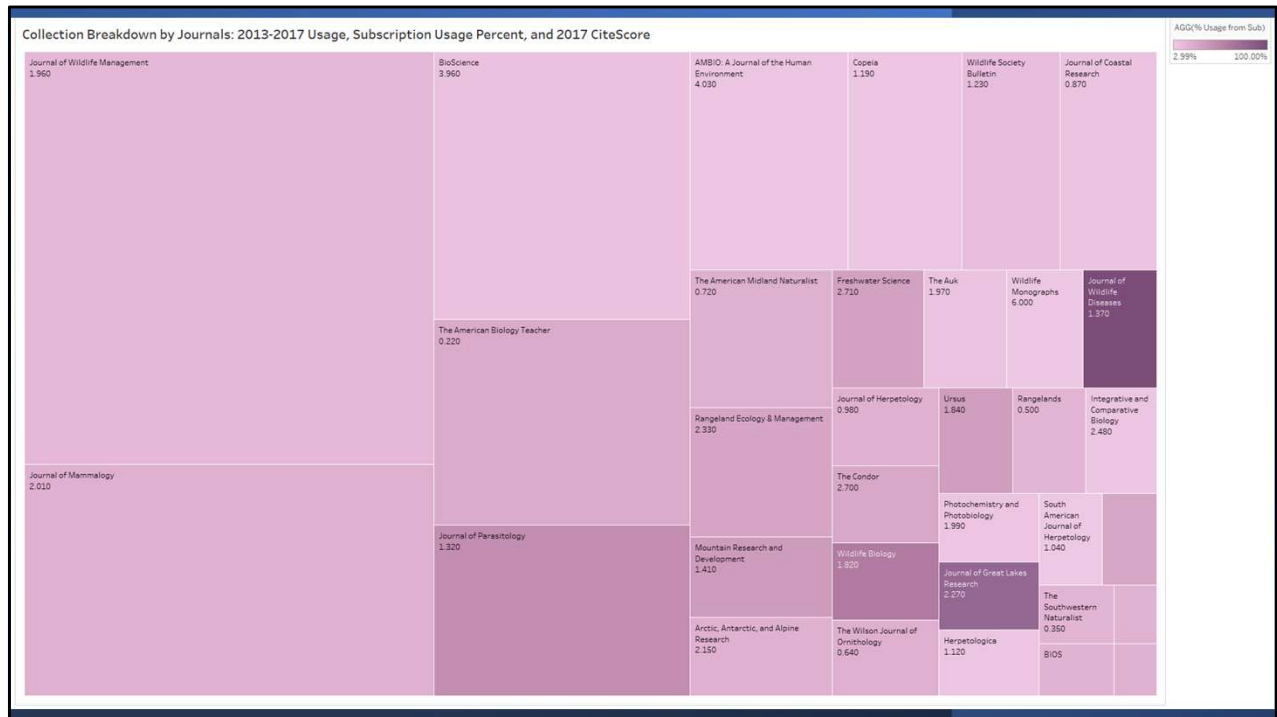
Identifies importance of Package E's subscription, Highlights historical changes in Usage of Package C



This is another visual -

- Left column: Subscription (orange) vs Aggregator Usage (blue) – [where the usage is coming from]
- Middle column: Title Count, Right column: Citable Documents in a 3-year period – [are journals fat or thin]

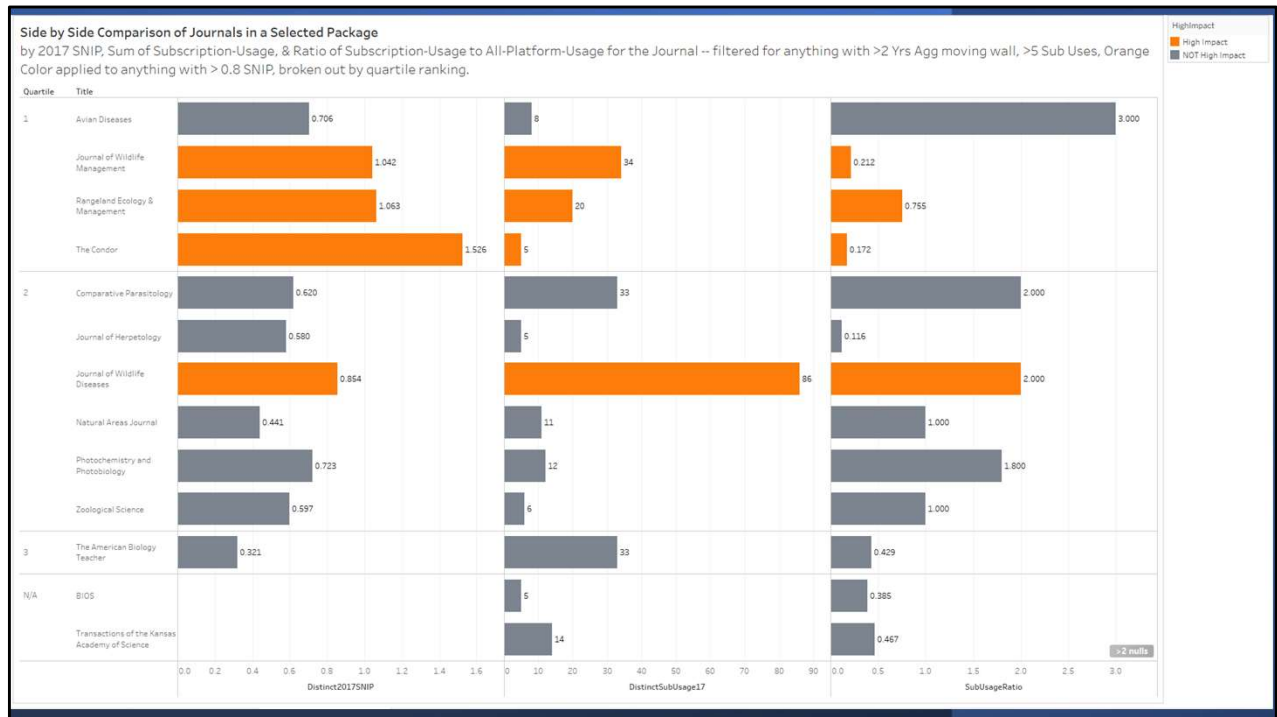
Package G for example: High Aggregator usage. Lots of Journals, not so many articles (especially compared to say A)



Once we identified a package to cut, we knew we'd want to at least consider subscribing to single titles to maintain access to the journals they wanted/needed.

So for a single package, we produced viz like these identifying 5-year usage (the size of the box), % usage from Subscription (darker shade = higher subscription usage), and a Value metric.

- Conditional shading like this makes it easy to identify the journals that rely on the subscription for access.

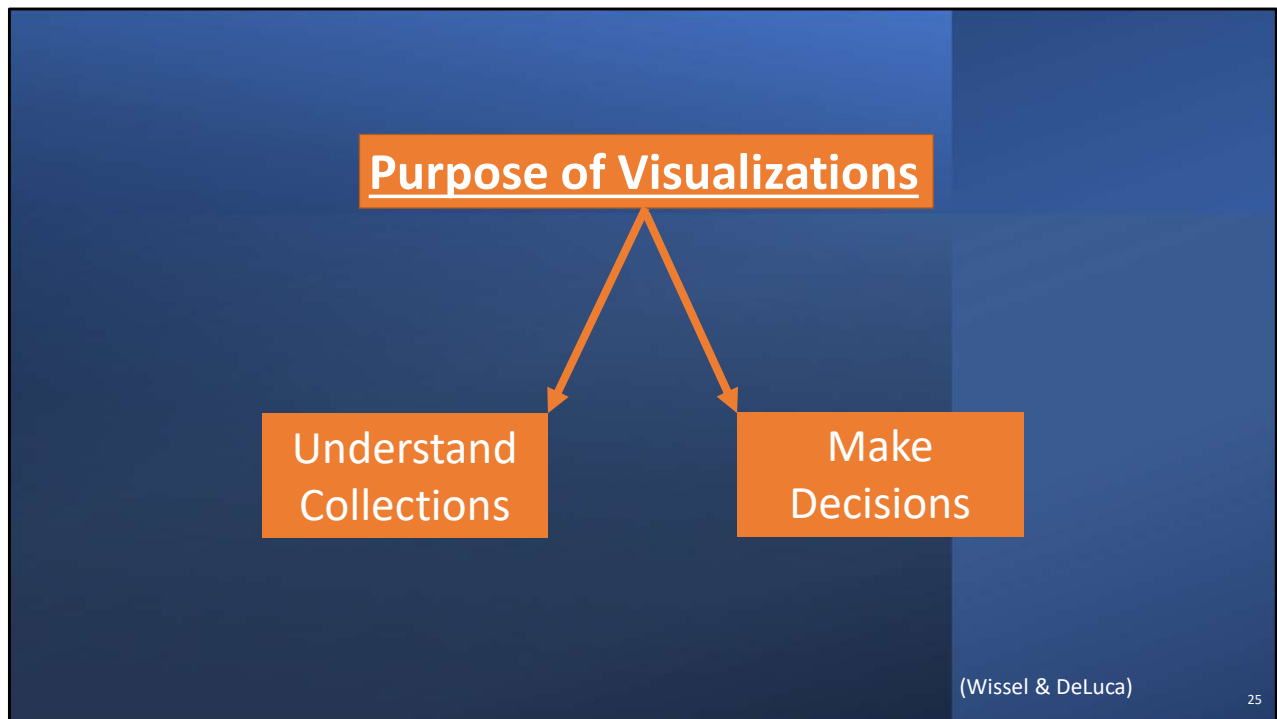


In addition to 'decision making viz' we created some information-packed visuals to illustrate our collection. This one for example combines SNIP, Sum of Sub Usage, and the ratio of Sub to All Usage (with filters and flags running in the background to highlight the info).

But this time around, I started to doubt myself:

- * Are these viz just cool looking or do they tell us what they need?
- * What metrics should we be focusing on to inform decisions?

So Nat suggested I look at some of the literature out there and see what other metrics/viz libraries are using to make these decisions.



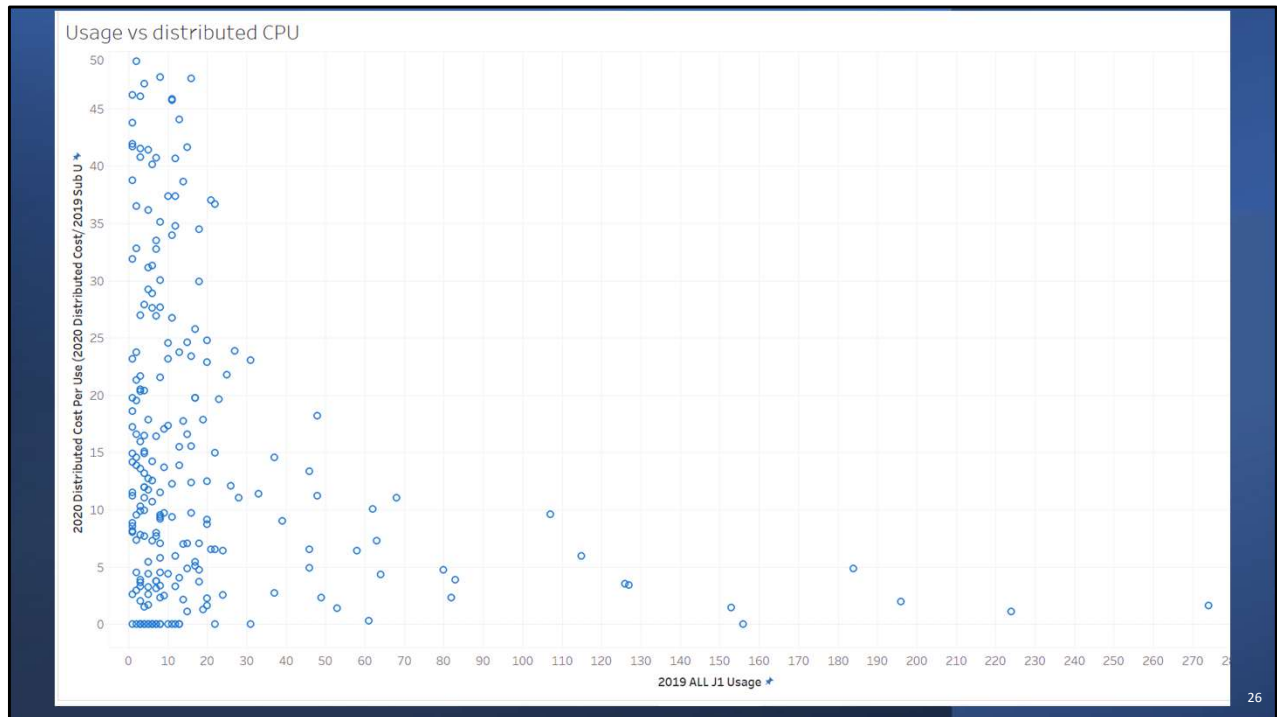
Wissel & DeLuca identify the two purposes of Visualizations

Visualizations as a way to understand collections

Visualizations as a way to make decisions

I felt that many of my viz in the past lead to understanding collection (which is useful), but I wanted to see what other people use to make decisions. I don't think any of this is groundbreaking, but I wanted to provide a summary of what I found:

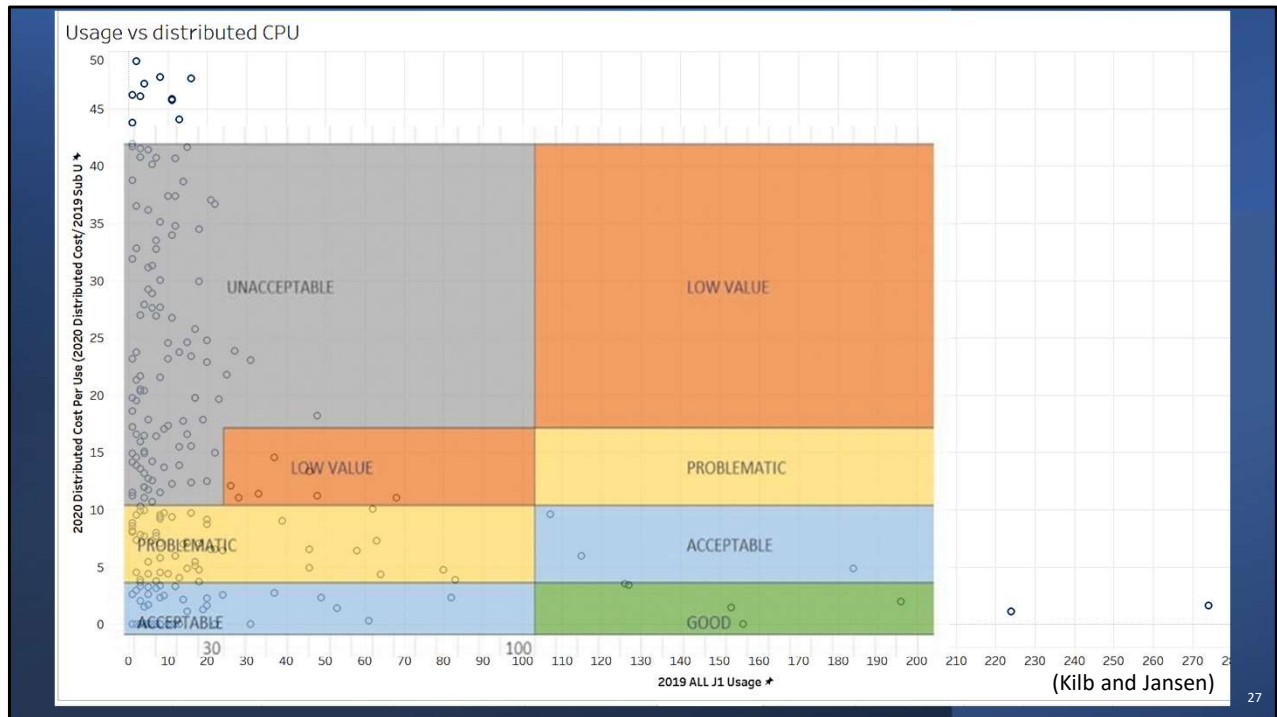
Wissel, Kathryn M., and Lisa DeLuca. "Telling the Story of a Collection with Visualizations: A Case Study." *Collection Management*, vol. 43, no. 4, Routledge, 2018, pp. 264–75, doi:10.1080/01462679.2018.1524319.



Multiple papers stated that they considered the most important metric to be CPU. There are a couple of ways to look at this:

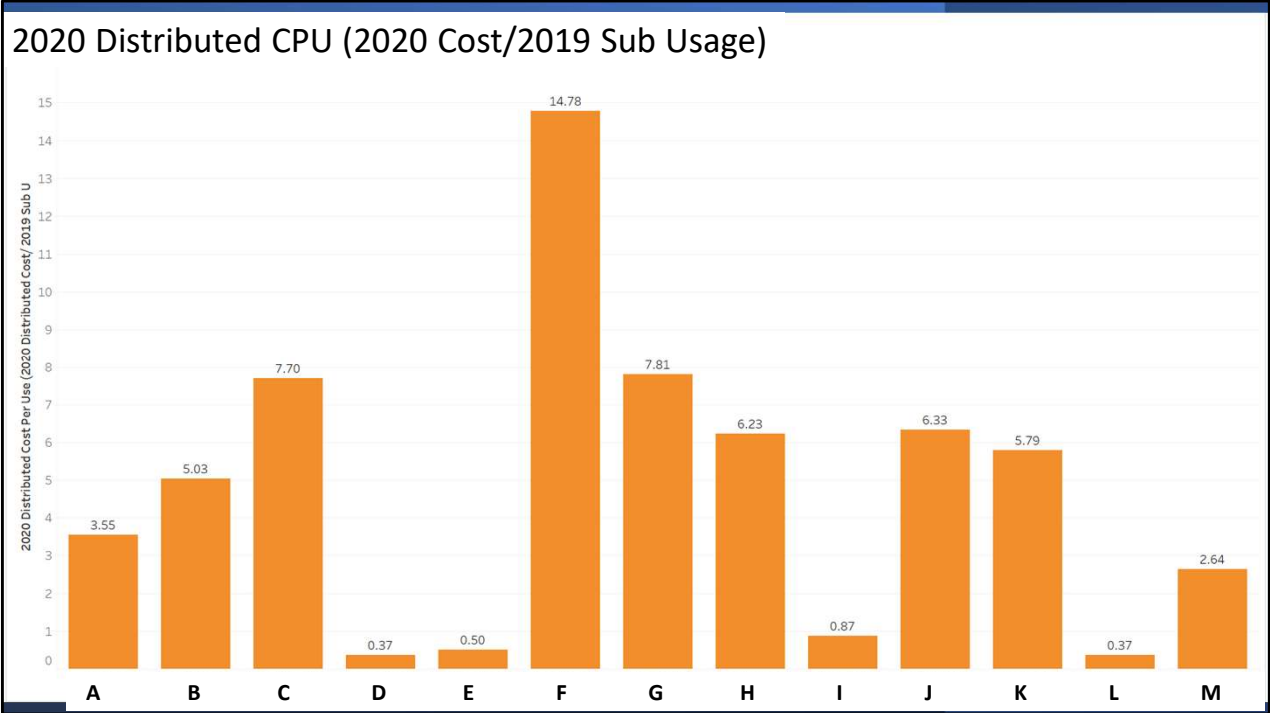
Kilb and Jansen suggest plotting Usage vs Cost Per Usage (CPU)

This is data looking at engineering journals – the usage is on the X-axis, the CPU is on the Y-axis.



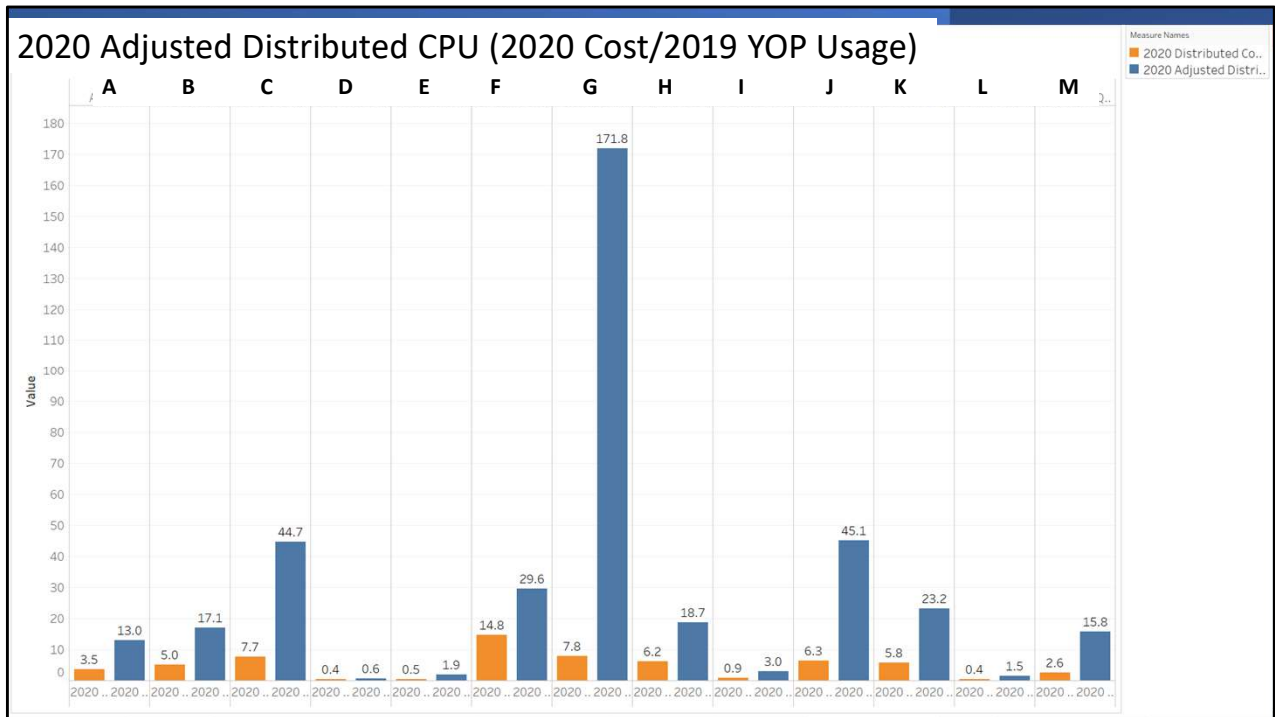
And then they created a benchmark scale to grade the value of each journal. I created a rough overlay here to show how our data would line up.

Kilb, Megan, and Matt Jansen. "Visualizing Collections Data: Why Pie Charts Aren't Always the Answer." *Serials Review*, vol. 42, no. 3, Routledge, 2016, pp. 192–200, doi:10.1080/00987913.2016.1207479.



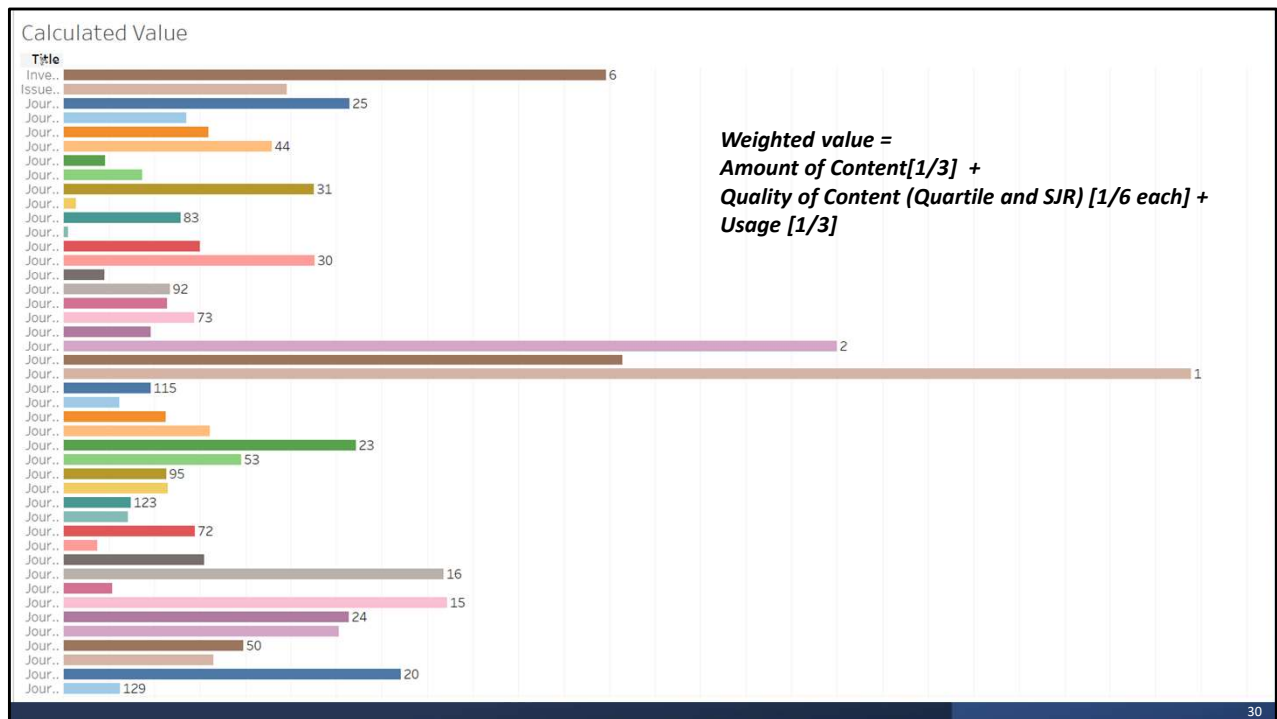
This is a different way to look at Cost per Usage. The Distributed cost for the year/a years worth of subscription usage. On face value, all 13 of these journals provide good CPU value.

However, if you consider that many licenses include PCA rights, this might not be giving a full picture.



Jabaily et al. argue that an Adjusted CPU would be better – calculating CPU for only the most recent year. This drastically changes the value of these journals.

Matthew J. Jabaily, et al. “Leveraging Use-by-Publication-Age Data in Serials Collection Decisions.” Where Do We Go From Here?, Purdue University Press, 2016, p. 292–, doi:10.2307/j.ctv15wxr5j.50.



And then finally, California Digital Library has designed a really cool Weighted Value Algorithm. When our team was brainstorming ideas, Evan thought of something similar and suggested we look at three categories:

- Amount of Content
- Quality of Content (Quartile and SJR)
- Usage

Each category is weighted at 33% and then the journal is ranked based on that value. We developed this calculation right at the end of the school year, so there is still work to be done, but it's an exciting metric to pursue.

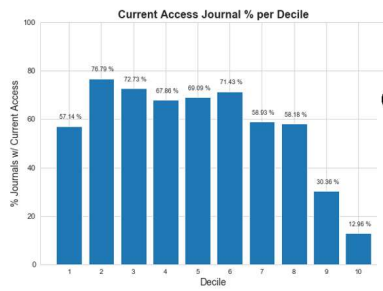
Like I said, none of this was really groundbreaking – but with the clean/validated data already – it was simple to find the model in the literature, reproduce it with our data, & compare our collection to other's standards.

And now for the actual groundbreaking work, our colleague Luwis:

<https://cdlib.org/services/collections/licensed/statistics-and-analytics/calculating-scholarly-journal-value/>

- **SciMB**

- Excel Workbook:
 - 21 Analyses
- Finished Reports-several analyses
- More than 70 subject areas.

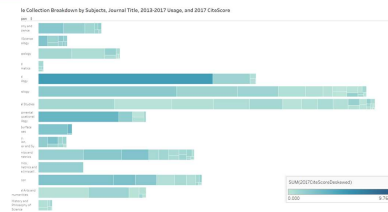
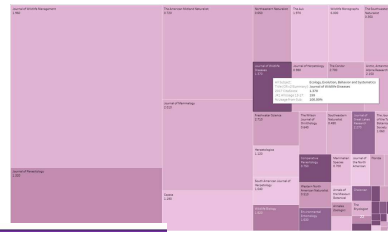


Top 15 most used journals

Journal Title	Sum of 2017 Usage/5yr usage rank out of 1103 titles
Children and Youth Services Review	849 (2)
Journal of Personality and Social Psychology	809 (4)
American Sociological Review	547 (6)
Future of Children	516 (11)
Journal of Family Violence	496 (8)
Foreign Affairs	480 (11)
Social Work	454 (9)
Violence Against Women	427 (14)
Teaching Sociology	411 (17)
Leadership Quarterly	408 (26)
Journal of Criminal Justice	389 (22)
Annals of the Am. Acad. of Political and Social Science	378 (9)
Social Forces	374 (5)
Gender and Society	365 (13)
Public Administration Review	304 (12)

- **PLAR**

- Tabular
- Several Analyses



Thanks Pat. I'll first give a little bit of summary of SciMB and PLAR analyses before these reports were automated. One key standard report is SciMB which includes more than 100 data variables. The SciMB includes 21 graphs and analyses which were manually created for each subject area. This was done using Excel. The SciMB report was designed to help program reviewers understand how the journal collection supports the curriculum of any given academic department. Data visualization included in the report can quickly provide meaningful insights, which helps also with decision-making. There is also another document called the "finished report." This is a word document based on the SciMB with a few selected graphs and analyses to share with academic departments.

Nat also gave a good introduction to the PLAR dataset. It includes about 20 rows of journal packages and more than 140 data variables. In the past, the PLAR has been used as a table only, without additional graphs or analyses.

In the past, the data visualizations for the SciMB were manually created, while the PLAR had no data visualizations. Creating the data visualizations was a tedious task. If we think about just the SciMB, there were more than 70 subject areas and a finished report had to be developed for each subject area.

Because all of these reports are standardized so that they have consistent data, we were able to automate report production.

Why We Choose Python & Jupyter Notebook

- Free and Open-Source
- Powerful, Dynamic Features
- Portability
- Free Third-Party Modules and Packages
- Productivity and Efficiency
- Easy to Read, Learn and Write
- Efficiency
- Consistency
- Reproducibility



<https://www.invensis.net/blog/benefits-of-python-over-other-programming-languages/>
<https://techvidvan.com/tutorials/python-advantages-and-disadvantages/>

We developed new versions of these standard reports using Jupyter Notebook to create the data visualizations and analyses. We replaced the previous versions of manually implemented graphs.

Why we choose Python and Jupyter Notebook?

Python is free and open source, there are tons of free third-party modules and packages. Also, Python is a high-level programming language that has English-like syntax. So it's easy to read and write.

Jupyter Notebook is a web-based interactive computational environment. It was designed to promote reproducible scientific research by providing an easily shared computational interface that mixes code, results, and text. Also, Jupyter Notebook was selected as the environment for development because it supports the achievement of the 3 major goals of efficiency, consistency, and reproducibility.

SciMB and PLAR Jupyter Notebook Codes

PLAR data set Analysis

Import below packages

- First need to importe all the packages and modules
- If there is get any error saying "There isn't any module", need to install it using pip command. It can be done using below 2 method

Command prompt

- For example if you don't have a matplotlib, you can open comand prompt and paste below code. To do some another module you have to change module name. Check pypi.org site or stackoverflow site to get reference.
- pip install matplotlib
- for reference use pypi.org site or [stackoverflow](https://stackoverflow.com)

Using jupyter notebook cell

- For example if you don't have a matplotlib, you can paste below code in to jupyter cell and run. To do some another module you have to change o module name. Check pypi.org site or stackoverflow site to get reference.
- !pip install matplotlib
- for reference use pypi.org site or [stackoverflow](https://stackoverflow.com)

```
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.ticker as mticker
from numpy.polynomial.polynomial import polyfit
import mplcursors
import squarify

from pandas import ExcelWriter
from pandas import ExcelFile
from IPython.display import FileLink, FileLinks
from datetime import datetime
import math
import xlswriter
import openpyxl
import io
```

Improt the data set as a Pandas Dataframe. Give the file path to 'filepath' variable.

```
! # Read excel file
filepath = 'Data/PLAR_AY21_2_ValuesOnly_210931.xlsx'
OriginalExcelFile = pd.read_excel(filepath)
#OriginalExcelFile =OriginalExcelFile.head()
OriginalExcelFile.head()
```

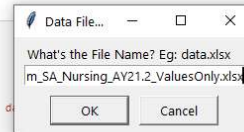
	COLLECTION	ROWLABELS	COUNT_OF_TITLE	COUNT_OF_OVERLAP	PERCENT_OF_OVERLAP
0	2020 Collection	Collection Identifier	Count of Titles For All Matched Journals (FAMJ)	Count of Any NonSub Coverage Overlap	% of Any NonSub Coverage Overlap
1	ACM Serials Only	2020ACM_SerialsOnly	136	6	0.0441176
2	ACS Core	2020ACS_Core	18	NaN	NaN
3	ACS Open Access	2020ACS_OpenAccess	2	2	1
4	ACS Token Access	2020ACS_Tokens	43	3	0.0697674

5 rows x 140 columns

In this code it collect the Base report data file name.

- Small window will open and you need to give your Excel file name along with the extension. For ex: "LJCA_Scim_EconFin_AY21.1_ValuesOnly.xlsx"

```
! ROOT = tk.TK()
ROOT.withdraw()
# the input dialog
DataFileName = simpledialog.askstring(title="Data File Name",
prompt="What's the File Name? Eg: d
#LJCA_Scim_SA_Nursing_AY21.2_ValuesOnly.xlsx
DataFileName
```



These are some of the SciMb and PLAR Jupyter Notebook Codes. Here we importing packages and modules that we need to perform the analyses. These are called cells and each cell contains a function or several functions. The cell on the screen shows the code to read a base Excel file which contains the data. We just need to give the Excel file name, then it's going to run automatically.

However, each year, it is important to update some of the code so that it continues to work properly. I'll show examples on the next slide.

Handling for Future Years

code no 6

- You need to replace current year MSU_IL_YOP field in this code "d2 = d2.rename({'2020_MSU_IL_YOP':'MSU_IL_YOP_MOSTRECENT'},axis='columns')"
- for ex: if the current year is 2020 then you need to change "2019_MSU_IL_YOP" to "2020_MSU_IL_YOP". So code should be look like "d2 = d2.rename({'2020_MSU_IL_YOP':'MSU_IL_YOP_MOSTRECENT'},axis='columns')"
- Also Change the heading years match to above code - for ex: if the current year is 2020 -> MSUIYOPHeading = "MSUIYOP 2020 (14 Usage YOP 17-19) (Citable Docs-3yrs)"

code no 7

- You need to replace current year distributed cost per use field in this code "d2 = d2.rename({'2020_DISTRIBUTED_COST_PER_USE':'DISTRIBUTED_COST_PER_USE'},axis='columns')"
- for ex: if the current year is 2020 then you need to change "2019_DISTRIBUTED_COST_PER_USE" to "2020_DISTRIBUTED_COST_PER_USE". So code should be look like "d2 = d2.rename({'2020_DISTRIBUTED_COST_PER_USE':'DISTRIBUTED_COST_PER_USE'},axis='columns')"
- Also Change the heading years match to above code - for ex: if the current year is 2020 -> DISCPUHeading = "2020 Distributed Cost Per Use (2020 Distributed Cost/ 2019 Sub Usage Only)"

code no 8

- You need to replace last year ILL_REQUESTS field in this code "d2 = d2.rename({'2019_ILL_REQUESTS':'ILL_REQUESTS_lastYear'},axis='columns')"
- for ex: if the last year is 2019 then you need to change "2018_ILL_REQUESTS" to "2019_ILL_REQUESTS". So code should be look like "d2 = d2.rename({'2019_ILL_REQUESTS':'ILL_REQUESTS_lastYear'},axis='columns')"
- Also Change the heading years match to above code - for ex: if the last year is 2019 then you need to change ILLLastYearHeading = "ILL Requests 2019"

```

# code no 4
#need to update the year
d2 = d2.rename({'2019_SUB_31_USAGE':'SUB_31_USAGE_lastYear'},axis='columns')
SUB31USAGELastYearHeading = "Sub 31 Usage 2019"

# code no 5
d2 = d2.rename({'2019_ALL_31_USAGE':'ALL_31_USAGE_lastYear'},axis='columns')
ALL31USAGELastYearHeading = "All 31 Usage 2019"
d2.ALL_31_USAGE_lastYear.fillna(0, inplace=True)

# code no 6
d2 = d2.rename({'2020_MSU_IL_YOP':'MSU_IL_YOP_MOSTRECENT'},axis='columns')
MSUIYOPHeading = "MSUIYOP 2020 (14 Usage YOP 17-19) (Citable Docs-3yrs)"

# code no 7
d2 = d2.rename({'2020_DISTRIBUTED_COST_PER_USE':'DISTRIBUTED_COST_PER_USE'},axis='columns')
DISCPUHeading = "2020 Distributed Cost Per Use (2020 Distributed Cost/ 2019 Sub Usage Only)"

# code no 8
d2 = d2.rename({'2019_ILL_REQUESTS':'ILL_REQUESTS_lastYear'},axis='columns')
ILLLastYearHeading = "ILL Requests 2019"

```

code no 4

- You need to replace most recent year. Sum of NonSub Usage field in here -- mo_re_ye_Sum_of_NonSub_Usage = "J1_NONSUB_FAMJ_2019"
- for ex: if the most recent year is 2019 then you need to change "J1_NONSUB_FAMJ_2018" to "J1_NONSUB_FAMJ_2019". So code should be look like "mo_re_ye_Sum_of_NonSub_Usage = "J1_NONSUB_FAMJ_2019"

code no 5

- You need to replace most recent year. Sum of Sub Usage field in here -- mo_re_ye_Sum_of_Sub_Usage = "J1_SUB_FAMJ_2019"
- for ex: if the most recent year is 2019 then you need to change "J1_SUB_FAMJ_2018" to "J1_SUB_FAMJ_2019". So code should be look like "mo_re_ye_Sum_of_Sub_Usage = "J1_SUB_FAMJ_2019"

code no 6

- You need to replace most recent year. Sum of All Usage field in here -- mo_re_ye_Sum_of_ALL_Usage = "J1_ALL_FAMJ_2019"
- for ex: if the most recent year is 2019 then you need to change "J1_ALL_FAMJ_2018" to "J1_ALL_FAMJ_2019". So code should be look like "mo_re_ye_Sum_of_ALL_Usage = "J1_ALL_FAMJ_2019"

```

# code no 1
mo_re_ye_Pa_Le_Sln_Sub_Usage = "J1_ALL_PACKAGE_2019"

# code no 2
mo_re_ye_Pa_Le_Sln_Sub_HTHL_Usage = "J1_HTHL_PACKAGE_2019"

# code no 3
yearHeading = "2019"

# code no 4
mo_re_ye_Sum_of_NonSub_Usage = "J1_NONSUB_FAMJ_2019"

# code no 5
mo_re_ye_Sum_of_Sub_Usage = "J1_SUB_FAMJ_2019"

# code no 6
mo_re_ye_Sum_of_ALL_Usage = "J1_ALL_FAMJ_2019"

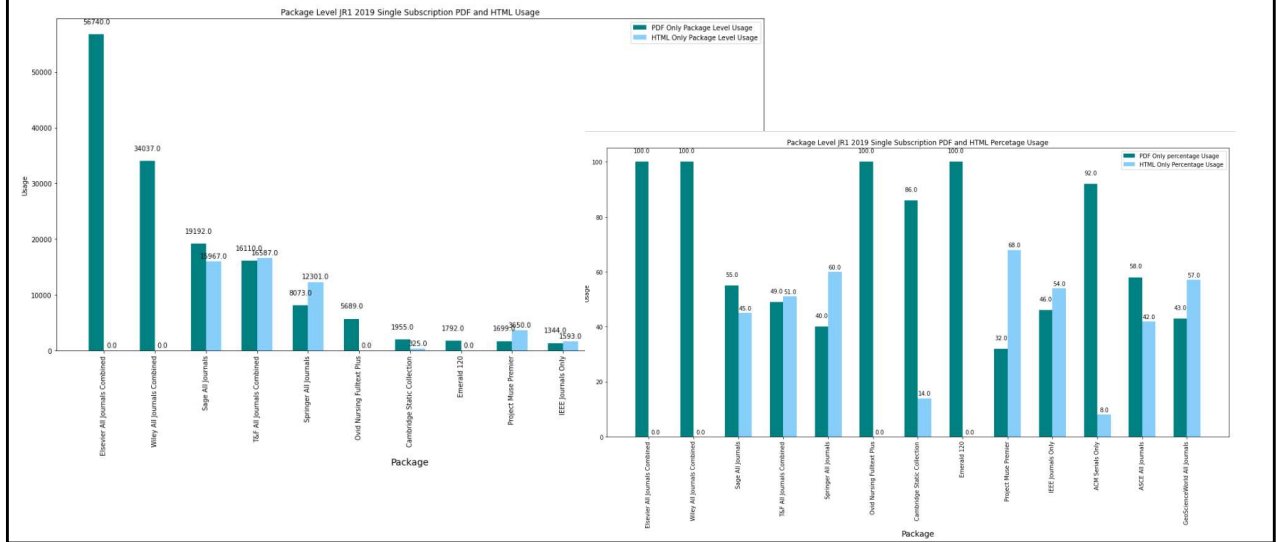
```

SciMB

PLAR

These are examples of code that need to be updated each year. I included instructions how to change the code each year. This means the code will remain useful even after my job is done.

Some examples



Here are some examples of automatically generated graphs. These graphs are different views of the same analysis. One uses actual values and the other uses percentage values



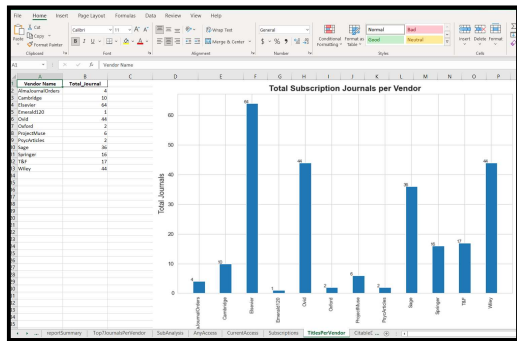
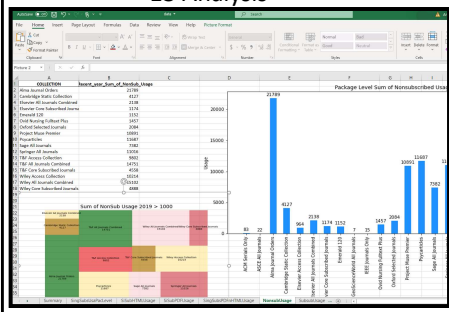
We can use packages like plotly and matplotlib to create interactive graphs, too. Here is the same analysis done by two different packages. If you use plotly, you can create nice interactive plots only using a few commands. Both packages have advantages. Plotly codes are easy. Here we give values for x and y. Then you need to give a size of the bubble. Then how the colour should need to change. I also supplied a collection= name. Then when you hover over it with your mouse, you get more information. Matplotlib is also easy, but when you create interactive plots, you have to use some other modules and packages.

Accomplishment

- SciMB
 - Excel Workbook:
 - 21 Core Analysis
 - Report Summary
 - Word Document:
 - Finish Report
- PLAR
 - Excel Workbook
 - 13 Analysis

Journal Title	ISI Impact Indicator/Scimago Impact Percentile
International Journal of Nursing Education	101.8 (54.1)
Mental Health Practice	23.9 (65.1)
Primary Health Care	13.9 (54.2)
Nurse Journal of Health, Nursing and Health Care	6.5 (64.4)
Journal of Family Nursing	3.48 (1.79)
Pediatric Nursing	1.8 (27.8)
Workforce on Evidence-Based Nursing	1.79 (27)
Nursing Economics	1.4 (27.8)
Marketing Nursing - Official Journal of the Academy of Medical Marketing Nurses	1.4 (27)
Nursing Inquiry	1.1 (15.7)
Online Journal of Issues in Nursing	1.1 (11.1)
Nursing Ethics	1.1 (11.1)
Nursing Access	1.1 (69.1)
Pacific Rim International Journal of Nursing Research	1.1 (47.8)
Critical Care Nurse	0.9 (20.9)
Journal of Community Health Nursing	0.8 (20.1)
ACORN Journal	No Access
Journal of Gerontology - Series B Psychological Sciences and Social Sciences	No Access
Journal of Research in Nursing	No Access
Journal of Transcultural Nursing	0.99 (20.1)
Journal of the American Association of Nurse Practitioners	No Access
Journal of Hospital Care	No Access
The Gerontologist	20.5 (20.9)
Journal of the American Medical Directors Association	No Access

Journal Title	Use (6 yr usage rank out of 643 titles)
Journal of Clinical Nursing	1537 (2)
Journal of Advanced Nursing	1488 (3)
British Journal of Nursing	1298 (5)
Journal of Transcultural Nursing	1164 (15)
Journal of Family Nursing	948 (6)
Medicine Nursing - Official Journal	902 (8)
ACORN Journal	616 (10)
Pediatric Nursing	528 (13)
Online Journal of Issues in Nursing	527 (7)
Journal of Nursing Management	526 (15)
Nursing Science Quarterly	473 (8)
Nursing Management	456 (24)
Clinical Journal of Oncology Nursing	438 (14)
Journal of Nursing Education	433 (12)
Nursing Ethics	403 (16)



Efficiency, consistency, and reproducibility. Why is that important to an organization? Because we need to perform the production work with the least amount of work-time or troubleshooting, to free our time to focus on communicating about the outputs.

Demonstration

Reflection and Change in Journal Collection Review

Evan Rusch (he/him/his)
Reference and Instruction Librarian
Minnesota State University, Mankato



Hi my name is Evan Rusch. I'm a reference librarian, liaison to a number of programs in the social sciences, and my role on the CMT is to help us take the amazing work of my colleagues you've just heard from and help make it more applicable to our liaisons and the departments and programs they serve. I'm going to talk about how we have learned from our past collection review process and how that has changed our thinking and goals for our current work.

2018-2019 Collection Review

- 2 Rounds
 - First round, used cost per use and overlap analysis to identify weaker packages
 - Second round, assessed weaker packages for cost of single sub add backs
- Dropped 3 subject-specific packages
- Strength was communicating and deciding on add backs
- One multisubject package slated to be cut was held because of the high number of add backs.

I want to start by recapping our last collection review. The review had two rounds. We started with an initial assessment of all of our journal packages. This relied heavily on the initial version of the PLAR that Nat described earlier. The goal was to winnow the 20 or so packages into a manageable list for a deeper analysis. The two primary variables that got packages included in the deeper review were those with high cost per use and those where there was a lot of content overlap with aggregators. The second round we dove into the individual journal titles within packages. So in this case we recalculated cost per use for individual titles with the sticker price for a journal subscription rather than the price we paid as part of our package deal. The journals with an expected low cost per use after this analysis were tagged for a single subscription in the case a package was cancelled. The cost of those journals we expected to “add back” offset the savings from dropping the package and helped to determine our final decision for the package. Initially it was determined that it was cost effective to break four packages. Later we decided against one package simply because we would be adding back enough titles that the cost of managing these subscriptions outweighed the savings from breaking the package. The 3 packages dropped were all subject specific packages rather than broader multi-disciplinary big deals.

Weaknesses of the 2018- 2019 Collection Review

Strength of analysis left little room for debate, just followed the numbers.

Lack of debate means qualitative information about programs, curriculum, specific journals, etc. can be missed.

Decisions were made with “our” analysis. Did this make it challenging for liaisons to communicate decisions with academic departments?

The focus on the “numbers” meant subject specific packages. Those cuts weren’t shared. Two departments were frustrated.

Our visualizations served the CMT well, but were they helpful to others?

As we began this year’s review of journal packages we reflected on our previous review to seek improvements. One issue that was surprising to me at the time, was that there wasn’t much debate or discussion within the JRC. I remember Nat, Pat, and I bringing charts and data to meetings, explaining them, and the reaction was that our analysis was reasonable. At the time that seemed good, and that we had come to an easy consensus. But unintentionally by conducting analysis and then sharing the results with the committee, it didn’t leave our colleagues much room for debate, as a result we didn’t hear other perspectives. We didn’t get those anecdotes about specific programs, curriculum, professors, assignments, that can help add depth to our understanding of our collections. Another concern is whether our liaison colleagues felt comfortable communicating about the collection decisions if they ultimately weren’t very involved in the analysis that drove those decisions. Would they be comfortable sharing charts or graphs we created or responding to questions from non-library faculty?

An additional problem in hindsight is that since we focused heavily on cost per use, that disproportionately targeted subject specific packages which at our institution tend to have a higher cost per use. Cuts to subject specific packages disproportionately hurt certain departments. For example, we have a very engaged Civil Engineering program and excellent librarian serving them, but a modest number of majors and student engagement with scholarly journals coming later in their time in the program causes the ASCE journal package to have a higher cost per use, even though testimonials from the department

demonstrated how essential these titles were to the program. We just simply can't make decisions on quantitative data analysis alone. Lastly, by Pat joining our group we were able to really improve visualization, and I can say that these had a huge impact on the CMT's work, but we have to ask ourselves whether these visualizations helped create understanding for others. We want to make sure visualizations help others participate.

New Approach to Collection Review

Subject Analysis

- Prioritization of packages is goal. Likely to prepare for a drop of large package.
- Need to understand impact of cut across disciplines.
- Subject impact measured by quality, usage, and supply.

Hands-on Analysis

- Helps bring broader group into analysis
- Encourages discussion (Source of qualitative data)
- Builds competence that aids in whole library communication

Visualization for Understanding

- Should support liaisons least connected to analysis
- Should contribute to non-library faculty engagement in decision making
- Can communicate at the subject/program level

For our current journal collection review our goal is to prioritize our packages so we are in a position to make a decision if a budget need arises. Because we have seen the impact of cutting subject specific packages we are likely to identify one of our multidiscipline big deals as a priority for cutting. Adding subject analysis allows us to understand how cuts will impact specific programs and departments in a multi-subject package. We can articulate the loss to a department or program if we were to cut a big deal as a loss of overall quality of holdings, a loss of used content or usage, and a loss of supply...essentially the number of articles lost.

We also have the goal of bringing more people into the discussion of analysis. We started this by holding a workshop that allowed our colleagues to play around with the different metrics available. The Scratchpad let our colleagues explore various metrics by their subject areas of interest. By working together in classroom type environment we actually have started to generate qualitative data. As colleagues explored and asked questions, liaisons talked about their perceptions of their disciplines and how they use our collections. These perspectives help balance our thinking about our quantitative data, and ensures decision making isn't simply an equation we follow thoughtlessly. Also by actively learning about metrics, we hope this builds confidence and competence in data use by our colleagues. This will improve decision making and communication. Even though we are bringing a wider group into the process of analyzing, we will not reach everyone. So we need to provide visualization that can support our colleagues who are least connected to the

analysis. We want to make it easy for them to understand and easy for them to communicate to their departments. Furthermore, we need our visualization to encourage non-library faculty to participate in prioritizing, and inspires them to provide qualitative responses that can help inform decisions and serve as testimonials of these collections' importance.


Qualitative Data Provides Balance

- Subject impact analysis might vary by department/program characteristics and attitudes.
 - Quality Metrics: Might matter more to research-oriented departments.
 - Usage Data: Do we have a sense of whether journal usage is driven to specific titles or found through discovery?
 - Supply: Are some programs more oriented towards just-in-time usage?
- Is there analysis we aren't doing that we should?
- Testimonials reaffirm value and help us build the case to our administration.

As we have already talked about qualitative information can help provide a balance to our quantitative data and give context to our numbers. We have talked about bringing together metrics of journal quality, usage, and supply to understand a package's impact for a department or discipline. Qualitative information about the idiosyncrasies of programs and their curriculum can help us adjust the emphasis on these metrics. Might a more research oriented department care about a loss of quality publications? Is there a sense that usage is driven more for a department to specific journals and articles rather than just needing "anything" on a topic? This information from liaisons or departments can enhance the quantitative data we provide. Also feedback can tell us there is analysis we aren't doing but should be. For instance in our first workshop a liaison discovered that for her discipline their usage and supply of content skewed towards aggregators. So we might need to incorporate a deeper analysis of our aggregator access for her department. Lastly testimonials reaffirm our value and help us build a case to our administration.

Conclusion

Library Strategic Planning Goals and Outcomes

- Improve liaison roles
 - Consistent relationships across disciplines
 - Systematic engagement
 - Develop a culture of evidence
 - Deeper understanding and participation in analysis
 - Enhance demonstration of fiscal responsibility
- 

In closing, the Library is finalizing a new strategic plan and I pulled these goals and outcomes from the current draft up for review. The work we are doing to improve our collection analysis fits directly in these Library wide goals. Because the plan is emphasizing these points, I hope this will further encourage wide participation in analysis by our liaisons and that participation will greatly enhance the CMT's work. Our revised collection review strategy to gear our analysis towards the impact on disciplines, to increase liaison and non-library faculty in analysis and decision making, to develop visualizations create understanding for all and inspire participation, and to increase the collection and use of qualitative data into our work, will not only help us meet these goals, but will allow our library to take full advantage of the amazing data work Nat, Jeff, Pat, and Luwis have contributed, allowing sophisticated and thoughtful analysis of our collections.

Thank You!

Here is a link to the slides:

<https://link.mnsu.edu/mankato-nasig21>

Download a local copy, then turn on the notes to see the text.

And here is a link to the first 3 video presentations:

<https://link.mnsu.edu/mankato-nasig-video>

7