Novelty Visualizations of Collections Data: Real Impact or Comic Interlude?

Nat Gustafson-Sundell
Evan Rusch

Minnesota State University Mankato
February 2, 2024

Hi, we are librarians from Minnesota State University Mankato, or MNSU, a regional, public, comprehensive university of about 14,000 students. We'd like to thank SUNYLA for this opportunity to present on this unusual topic. One other colleague, Mark McCullough, was going to join us for this presentation, but he's focusing on a different presentation. We want to thank Mark for his help as we talked through the issues involved in the development of novelty visualizations. His comments helped to inform many of our slides. Okay, so let's jump right into our presentation because we'd like to cover a lot of ground quickly.

link.mnsu.edu/noveltyviz

But first, let me provide a link to our slides. We know it can sometimes be difficult to follow along with a presentation or see all the little things on the slides, so you can download our deck at "link.mnsu.edu/noveltyviz."
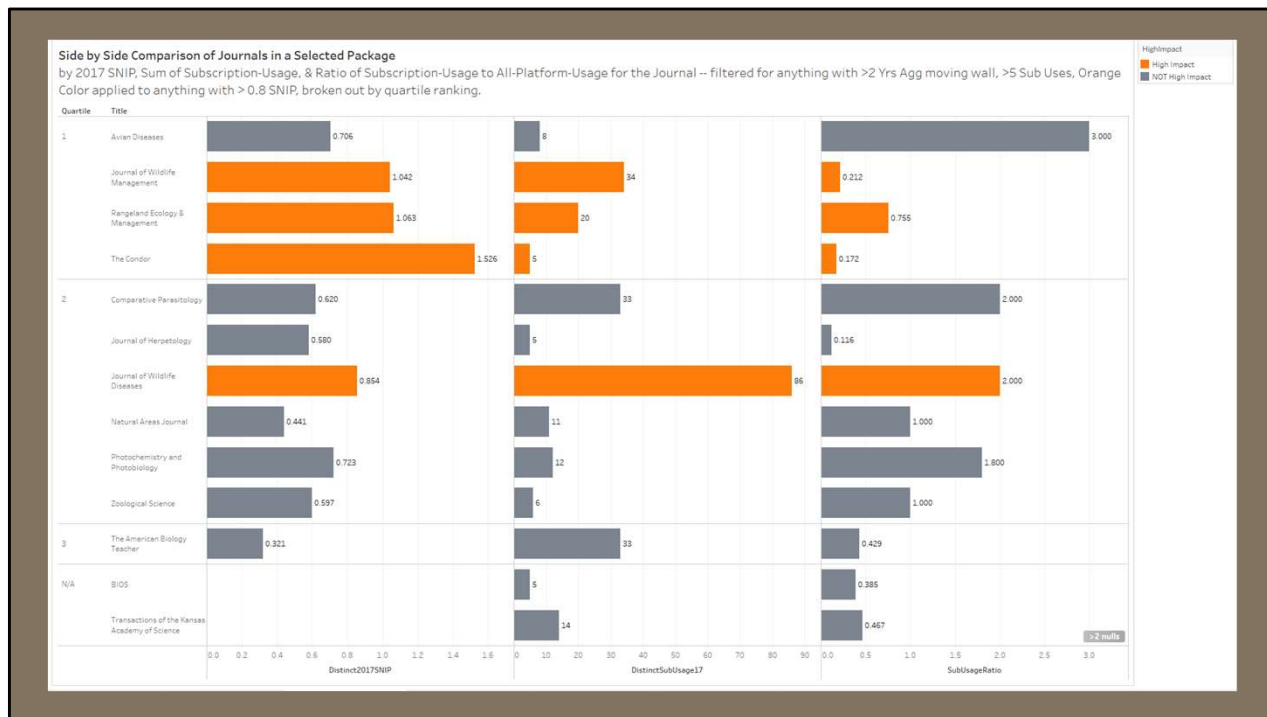
At MNSU, we've developed collection analysis and data visualization products iteratively over many years. We've previously presented our data viz implemented with Tableau, Jupyter Notebook, and Microsoft Power BI at several conferences. The image on the screen is from a presentation we gave to NASIG in 2021. I'll provide a link at the end of this presentation in case you want to see some of our other work. We've thought a lot about charts and tables, and how best to tell a story with data depending on the context. Today, we'll present an unusual approach to data viz. To be honest, we're not sure yet what to think of this new approach, thus the title of this presentation.

3
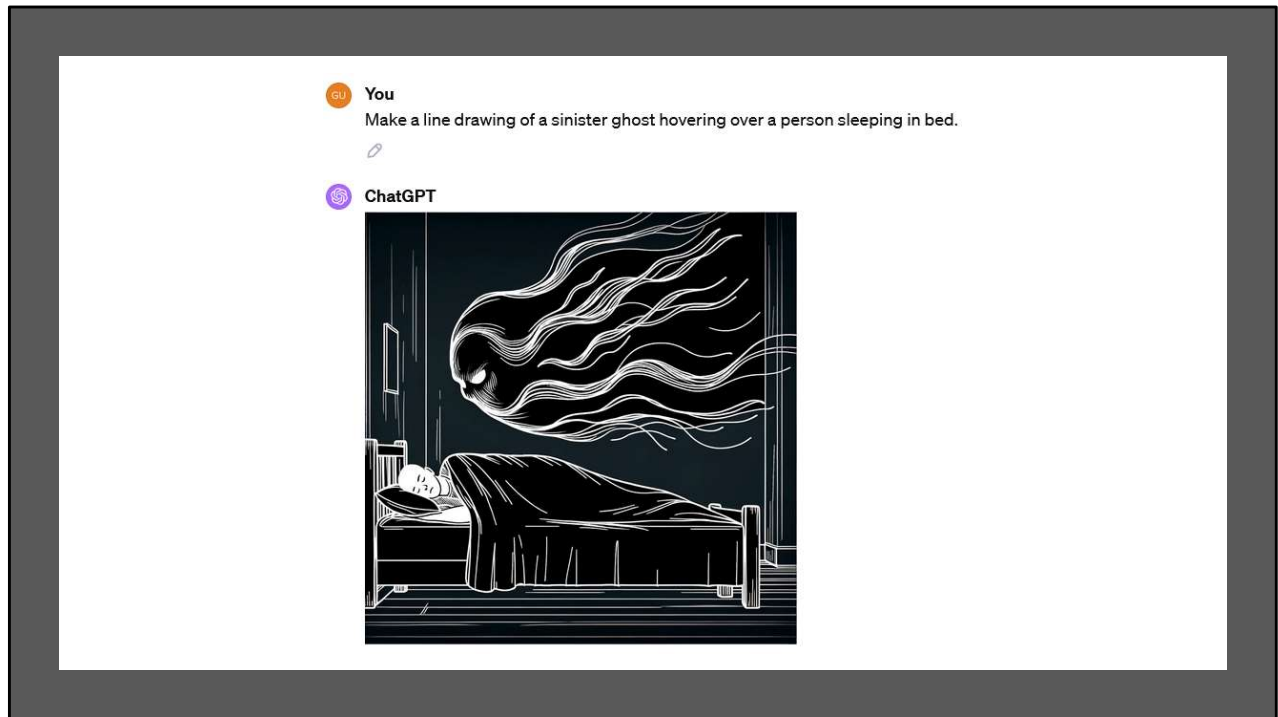
Let me provide a brief background. It was the middle of the night and I woke up, as I often do, at about 3 in the morning. That's the time when ghosts visit, so I usually try to think of something to occupy my mind until I can fall asleep again. I had recently given some presentations on Generative AI. I was thinking about possible applications for Dall-E 3 and other image generators. I thought to myself, what if I was able to create illustrations of data, not as charts or tables, but as pictures, where the objects in the picture represent data? To be clear, I didn't want to create pictures only to serve as traditional infographics, simply to illustrate the topic. I wanted the images to depict quantities accurately. And more, I wanted the pictures to depict 2-variable data relationships, so that the proportions of objects in the picture would be quantitatively accurate as a comparison.
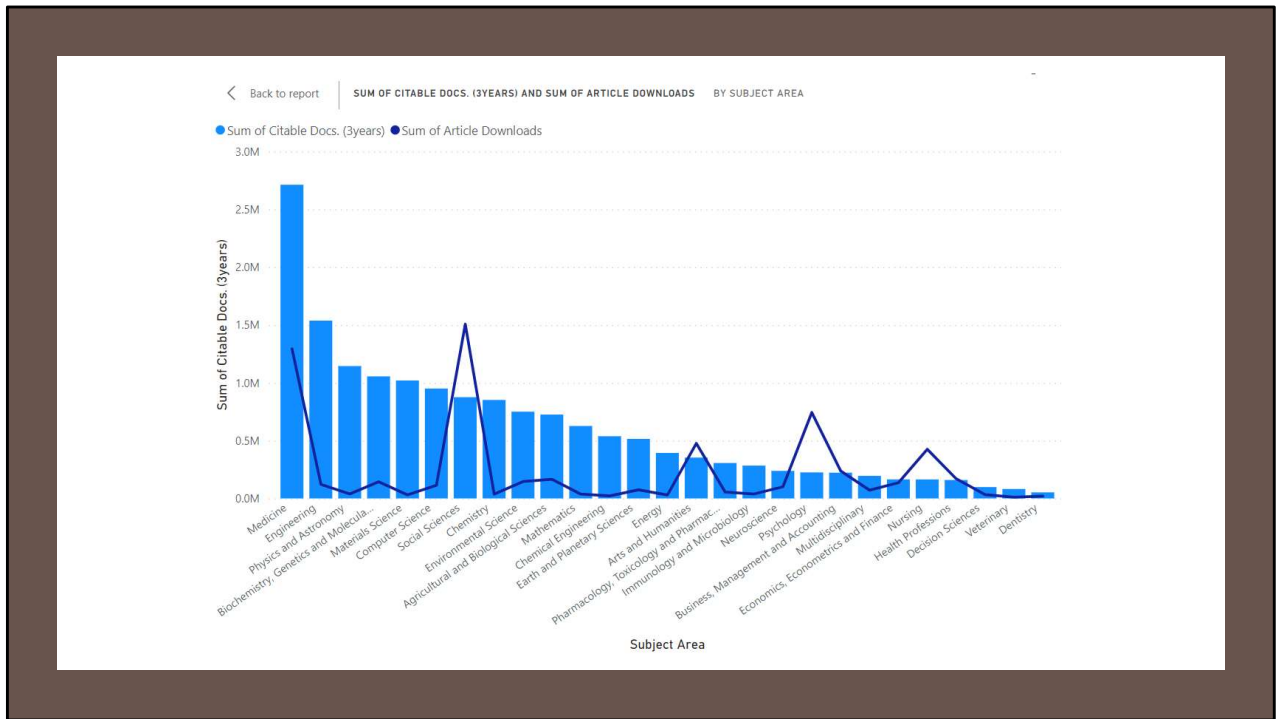
Here is an example of a traditional 2-variable comparison I slapped together in Power BI for the purposes of this presentation. The columns represent the number of citable documents published in the last 3 years, while the line shows article downloads. The point here is just that this chart compares 2 variables, basically supply and demand.
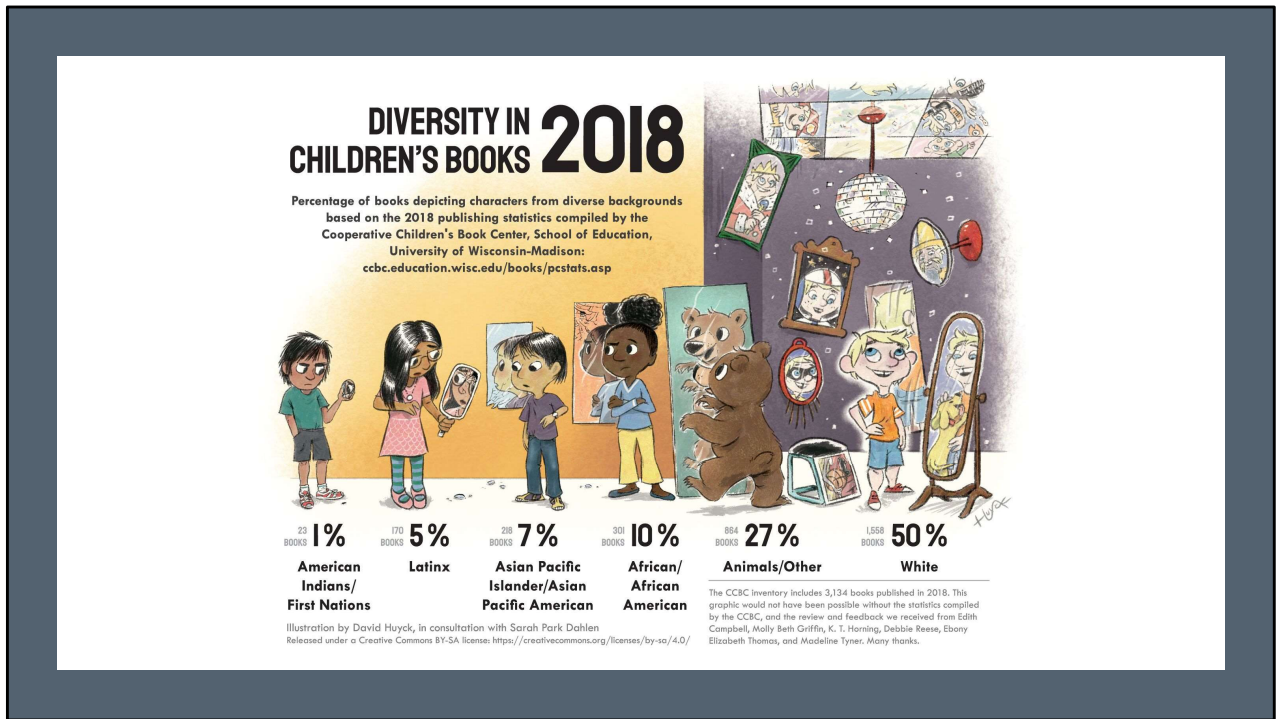
Of course, I thought it might also be interesting to create 1-variable illustrations, although I've seen these before. I think I first saw a chart like this in *Time* magazine when I was about ten. Missiles were used in bar charts to compare the nuclear armaments of the USSR and the US. We'll provide some examples of 1-variable illustrations in this presentation.

On the screen, you can see a creative example of one of these. The size of the mirror is meant to depict the supply of children's books for various populations. This picture is effective for getting its message across, but if read as a 2-variable chart, the size of the children would need to depict the sizes of the populations involved. Basically, my goal has been to create data illustrations that are accurate as charts.

Novelty Viz

- Introduction & Background
- Nat's Demonstration
- Evan's Discussion
- Conclusion

Make a picture of a medieval monk illustrating a manuscript. We can see over his shoulder. He's drawing a dragon.

Today, I'll present a few examples of my work on these data illustrations, which we're calling "Novelty Visualizations." Evan Rusch led our internal discussion about how these novelty viz might be used. He'll provide an overview of the issues we've discussed so far.

Before I demonstrate any examples, I should be perfectly clear we don't think we've been fully successful yet. The basic idea is to provide novelty visualizations that could lighten the mood of meetings with faculty or administration, or to provide a memorable insight. We would rely on our usual charts and tables for such presentations, but if we add one or two novelty viz, we might be able to create a better or a deeper impression.
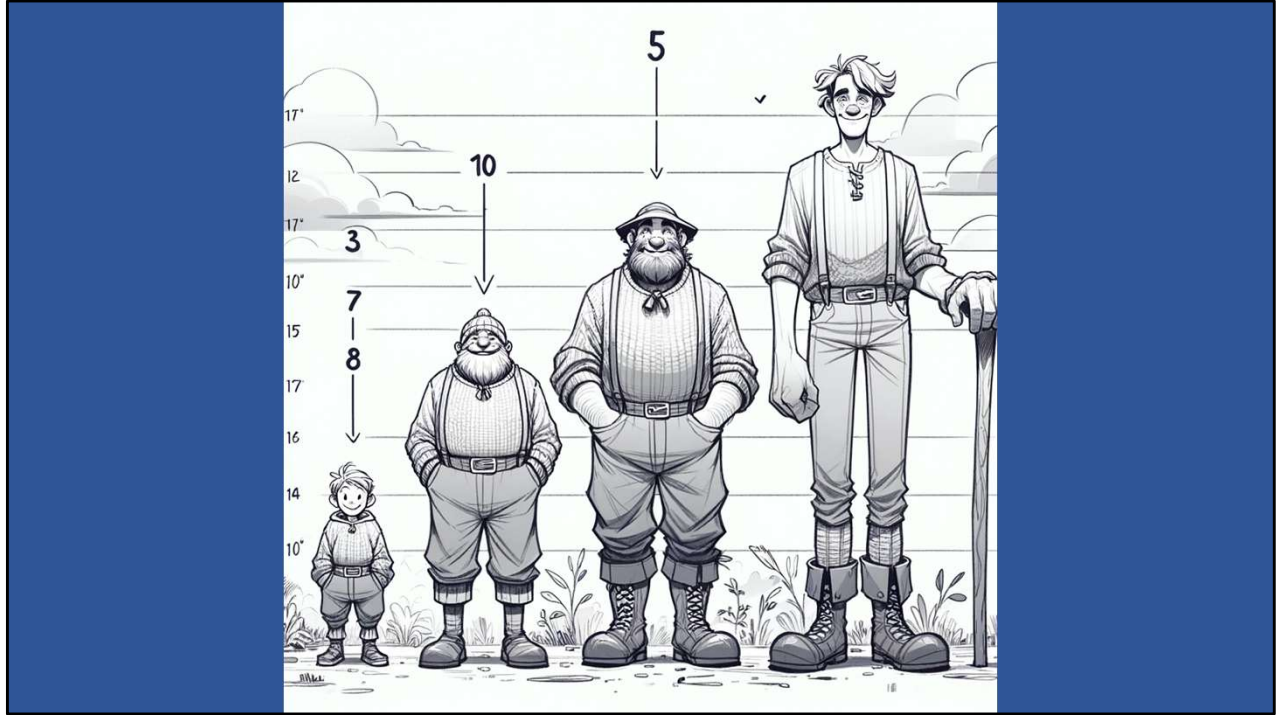
Unfortunately, though, Dall-E 3 has problems and it might not be ready for what we're trying to do. You might see bizarre stuff in some images.

And Dall-E 3 isn't good with numbers yet. I didn't know Dall-E wasn't good with numbers until I tried my first experiment. When I first asked Dall-E 3 to depict 2-variable relationships, it tried, but it couldn't do it. I'm really looking forward to Dall-E 5 or 6 when these problems are resolved, which is to say that I think Dall-E will get better at what we're trying to do now.

For Dall-E 3, I needed to dramatically simplify what I was asking.

Even with these problems, though, I was able to make a sort-of successful comparison after about 15 minutes. So, then, let's imagine we're in a meeting…

Springer Journal Supply for Subject & Quality (quartile)
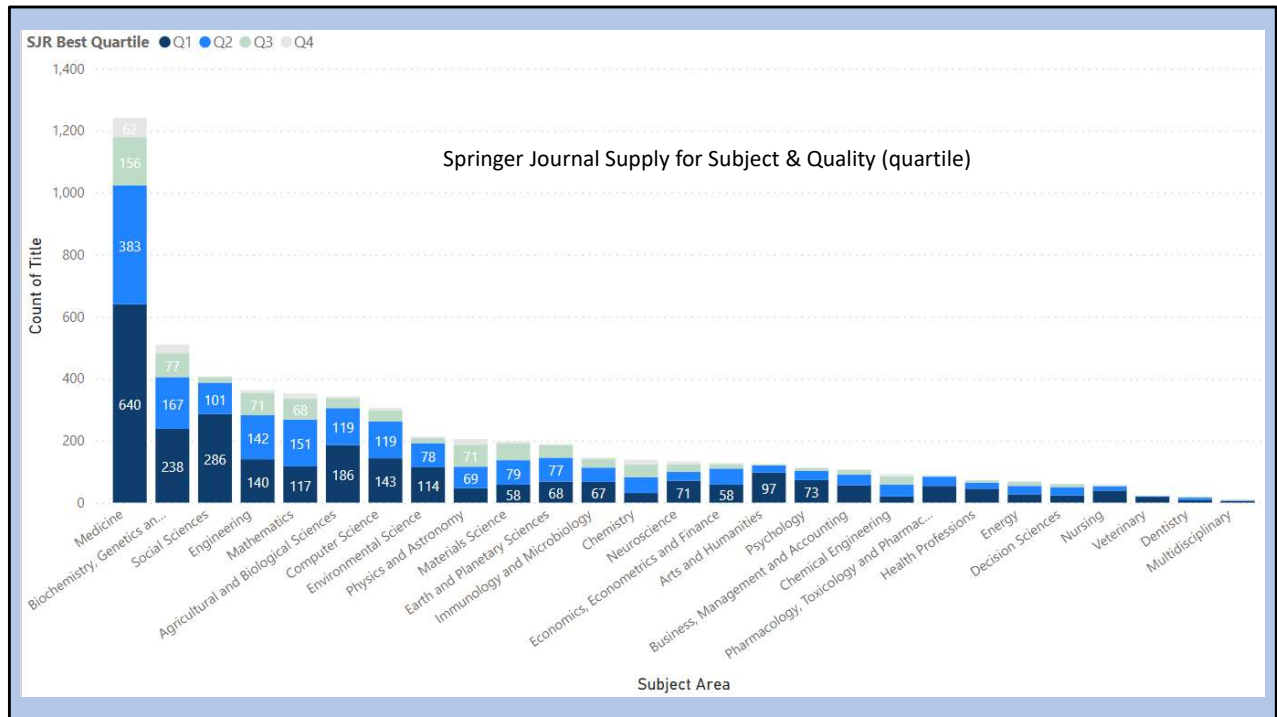
In this imaginary meeting, I'm presenting on the value provided by our Springer journal package. So there I am, "blah blah, blah blah blah…"

Is Springer lost in the crowd?
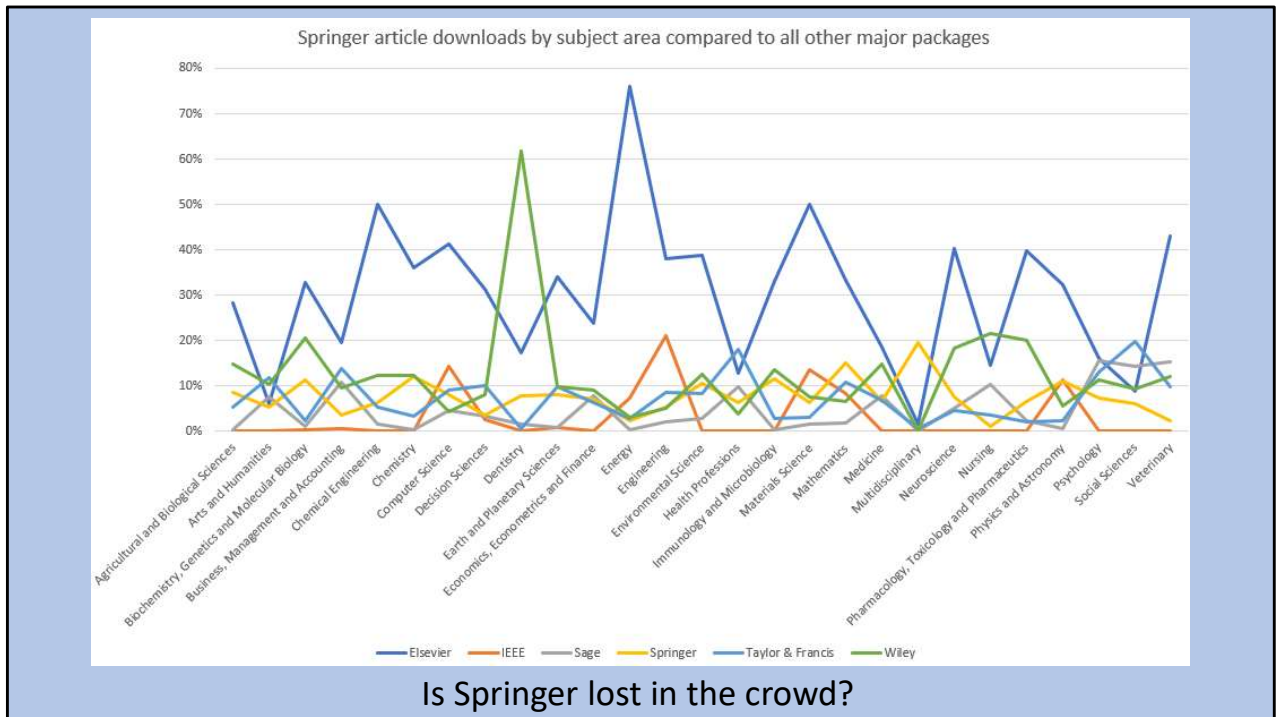
I get to a slide where I show the usage provided by all of our most expensive journal packages. Springer usage is depicted by the yellow line. I ask, "Is Springer lost in the crowd?"

Consider:

The Springer journal package costs less than one-fifth as much as the Elsevier journal package…

Proportionally, the usage of Springer is greater than Elsevier.

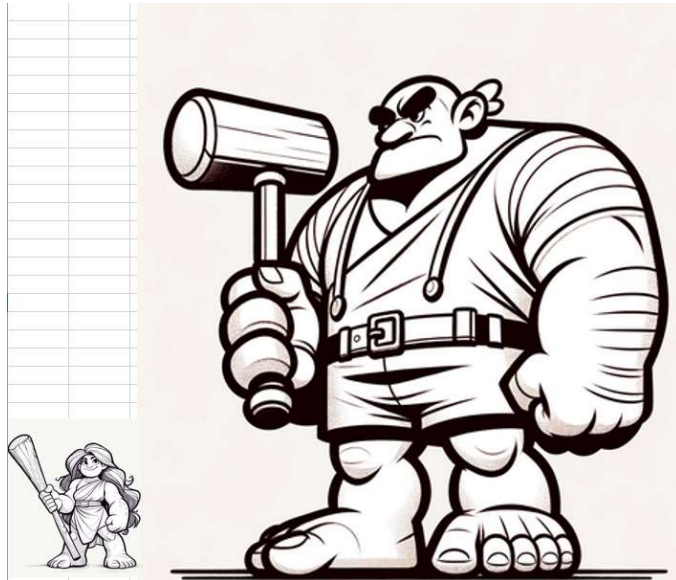(The club size indicates usage.)

I follow this with a novelty visualization depicting the overall cost and usage of the Springer package relative to Elsevier. The size of the giant equates to the cost, while the club size equates to usage. The image shows usage relative to cost.

Please remember this illustration was my very first try and does not quite accomplish what I originally envisioned. This illustration was meant to prove the concept of novelty visualizations. When Evan first saw this, he immediately pointed out how the choice of visual metaphor could be improved to better depict the variables involved.

The cost of journal big deals depends on historical subscriptions at libraries, so these costs vary from library to library. At MNSU, like at many libraries, the cost of Elsevier is far greater than the other deals. However, Elsevier also provides vastly more citable documents, according to our analysis of ScImago data. We might think of these documents as treasure for our students to discover as they adventure through their university careers.

In this picture, the size of the dragon relative to the treasure hoard is roughly the same as the cost of our Sage package relative to the number of citable documents provided.

So here is my second experiment, in which I heeded Evan's point. In this example, the size of the dragon roughly equates to the cost of our Sage journal package relative to the pile of gold, which roughly equates to the number of citable documents provided by the package in publications over the past 3 years.

The cost of journal big deals depends on historical subscriptions at libraries, so these costs vary from library to library. At MNSU, like at many libraries, the cost of Elsevier is far greater than the other deals. However, Elsevier also provides vastly more citable documents, according to our analysis of ScImago data. We might think of these documents as treasure for our students to discover as they adventure through their university careers.

In this picture, the size of the dragon relative to the treasure hoard is roughly the same as the cost of our Sage package relative to the number of citable documents provided.

Now here is the cost of Elsevier relative to the number of citable documents they provide ->

The point here is not to present the value of the Sage package, but to show the value of a better deal. So I turn the page…

… to say, "Here you can see our Elsevier journal package relative to the Sage package based on the same variables of cost and citable document supply."
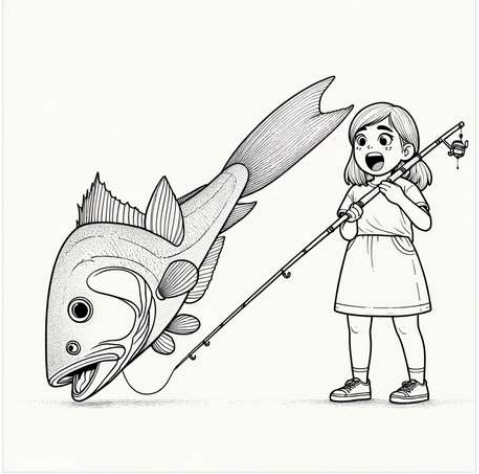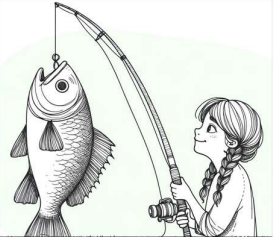
The reason I'm focusing on the value provided by these packages is because we've been developing a new concept for collection review. We're trying to flip the script a little bit. Instead of focusing on cancellations, we're focused on justifying our budget needs in a new kind of report we call the "Budget Proposal." We will be presenting on the BP at NASIG in June.

Anyway, on the screen you can see a one-variable novelty visualization. If we imagine journals in the top fiftieth percentile as living humans and journals in the bottom fiftieth percentile as zombies, we can illustrate the composition of a journal package. This novelty visualization is roughly accurate for Taylor & Francis. Actually, I started by asking for a group image, but Dall-E couldn't get the numbers right, so then I did one portrait at a time, which I copied into Excel and re-sized.

Honestly, though, the zombie metaphor might be better for a different context. Here, the zombies represent the proportion of journals providing 20 or less subscription platform downloads over 3 years in our Springer package. The proportion of zombies equates to the proportion of low use journals. This kind of novelty visualization might be best used to explain why we'd cancel a journal package.

Evan will talk about the categories of novelty visualizations we've considered so far and some of the issues involved when choosing visual metaphors. There are all kinds of metaphors to try. Evan has suggested prize pigs at a county fair, size of pig vs. size of ribbon, or number of ribbons. We're in Minnesota, so our metaphors might tend in a Minnesotan direction. How about the number of morel mushrooms on a patch of ground, or a fish on a fishing pole, where the size of the fish and the length of the pole mean something? Or we could depict a pie chart as an actual pie, or a pie chart as a cake. The possibilities seem pretty endless to me, but first, it's important to have a sense of when and how these viz might be used most effectively.

So now, Evan will dive into the issues a little more deeply.
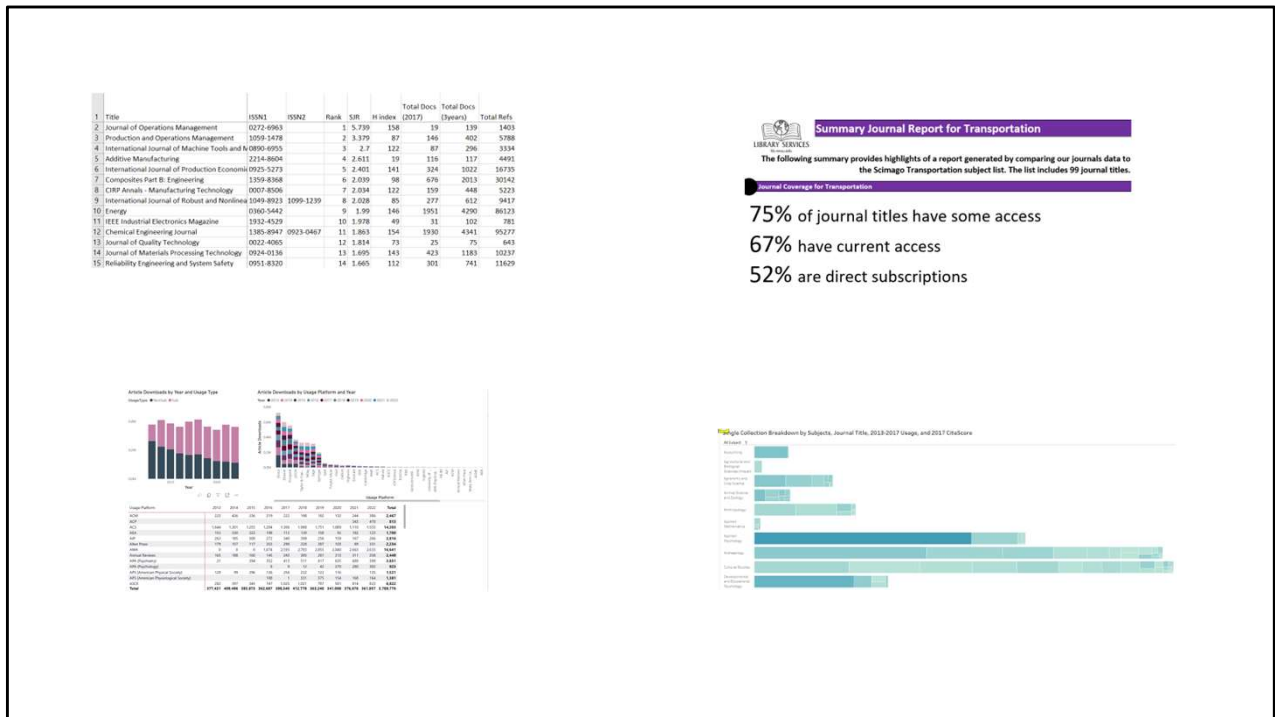
Nat and I have collaborated over the years on developing reports and tools for collection analysis and data visualization, particularly as it relates to our journal collections. As we have developed various roles in this work, my focus has often been on how to share this information, so it is useful to our librarian colleagues, our Dean, our faculty colleagues across campus, and other administrators…even accreditors.

We have used various methods to communicate and share information from sophisticated reports using Excel spreadsheets, to unique visualizations in Tableau, to glossy one-sheeters for colleagues to bring to department meetings, to interactive dashboards using Power BI.
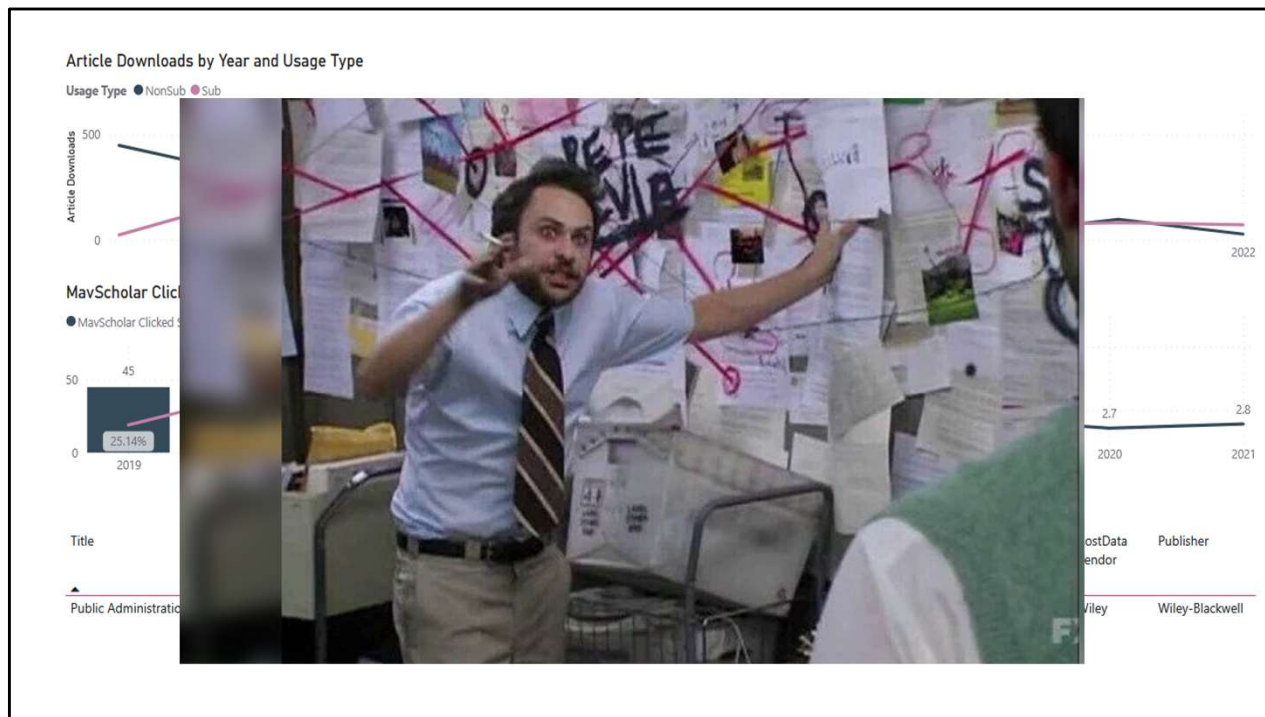
All of these reports, tools, and visualizations have strengths and weaknesses. While they have served us well, none are perfect. In any case we are trying to make, we need to adjust our message considering the context, the purpose and the audience. As mentioned before our audience can vary significantly, and therefore their knowledge, understanding, and interest can be equally varied. Even amongst our librarian colleagues there is differing levels of comfort with working with data and interpreting visualizations. To communicate about collection analysis means we need to think about various learning preferences and levels of knowledge about our collection and the data we share about them. For some, might this well intentioned chart come off like this!

In my role in communication, I am excited about novelty visualizations' potential. Can they help bridge the gap where other reports, charts or graphs have failed to connect? Or provide an alternative that really brings home the points we are trying to convey? In the past these types of visualizations would have required hiring graphic designers or artists and would need well thought out concepts. This certainly wouldn't be feasible for a quick invitation to an academic department meeting. In fact even utilizing a clip art library or other repositories of symbols takes more time than utilizing AI. We might now be able to produce an image 10 minutes before we head out the door and have a successful visit to an academic department like in this AI created image.

Why Use Images?

Draw attention or change tone of presentation

Deepen understanding and to enlighten

Create a lasting impression

Persuade

So what role might these images serve in enhancing our presentation of collection analysis data? Nat has provided a few examples of how we might use these, but here are some general ideas of how they might improve our message:

1. I think novelty visualizations could help us grab the attention of our audience or change the tone of a presentation. Any time we are presenting to a group, we are going to have people with different comfort levels and interest in data presentations. Can this help us reach the person who hears Charlie Brown's teacher voice when I am talking about journal usage data? Might this also help accommodate different learners?

2. Novelty visualizations might also help us deepen our audiences understanding or really help to bring home a point. I think initially I was thinking of AI created visualizations as potential replacements for traditional charts and graphs. I am now thinking one of the ways they would be most effective is as a supplement to the traditional tools and viz we have used in the past. Can we create that "a-ha" moment or where someone says "okay, okay, I get your point!"

3. Another potential benefit of using novelty visualizations is to create a lasting impression. Imagine a series of charts and tables in a presentation, how does one chart midway through going to stick in anyone's brain. It seems to me, novelty

visualizations can help cement a concept we are trying to get across. Even better, they can give us reference points…Rather than saying "if you remember in the chart demonstrating the composition of the big deal journal package", we could say "Do you remember when we talked about those zombie titles in the journal package?" Whether we are talking about piles of gold or pigs at the county fair, novelty visualizations could give us a shorthand for concepts that might allow for lasting impressions and reference points in future discussions.

4. Lastly, while in many cases our purpose might, as much as is possible, be to let data speak for itself. Often we are trying to be convincing. We see this as even more likely important as we trying to flip the script of our reporting and justify our budget needs rather than focus on cancellations to balance the budget. Novelty visualizations can add to our narrative and potentially be more convincing…If we can make the expensive journal package seem less like an expensive budget line, and more as an asset that provides the biggest pile of riches, we might be able to shape impressions and improve our ability to justify our collections budget. I think novelty visualizations could be especially helpful when we are trying to encourage a change in thinking or the point we are making goes against conventional wisdom. Novelty visualizations might also help soften the blow of a cancellation. Is a package full of zombie journals is easier to let go of?

In the Springer journal package, nearly **8 out of 10** of the journals in the package provided **20 or less** article downloads over **3 years.**

Bar chart

Nat has provided some of the ideas we have had for types of illustrations. Ultimately my starting point on this topic was to think through the types of viz we had created previously and how that might be translated into novelty visualization. While part of the interest in doing this work was to capture some of the complexity of previously created visualization, some of our most simple charts seemed easy candidates for illustrations. For instance, if you think back to the example of humans and zombies we can take a simple bar chart that represents journal usage in our Springer journal package. This example is sort of simplest or first level use of novelty visualization.

Count of Current Access Quartile 1 Journals
Supplied Per Interface for Accounting

| 1 | 1 | 2 | 3 | 3 | 4 |
| DOAJ | University of Chicago | Cambridge | Emerald | Sage | Springer |
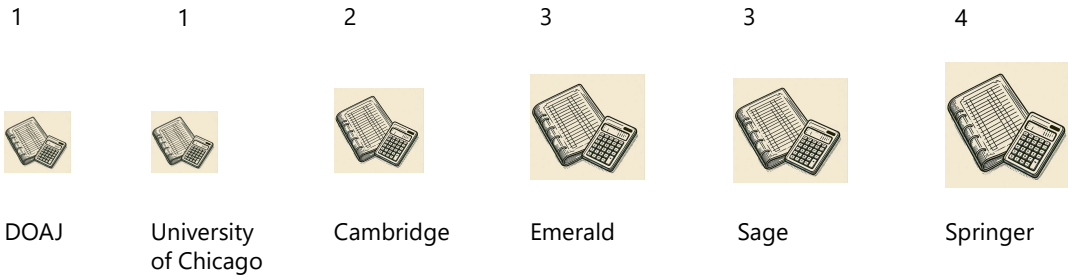
Sequential comparison, 1

A step more complex is the idea of sequential comparison. In this case we might use multiple images in order of magnitude to demonstrate proportion. Imagine looking at which journal packages provide high quality journals for a given discipline. In this example, we are looking at top quartile of accounting journals according to Scimago. Each journal package's count of journals in the top quartile in which successive images show the increasing proportion each package provides.

Count of Current Access Quartile 1 Journals
Supplied Per Interface for Accounting

6

7

8

Taylor &
Francis

EBSCO
Aggregators

Elsevier

Sequential comparison, 2

You can see here we tried a multi-level approach...grouping the packages...with a first tier of packages and then a second...

Count of Current Access Quartile 1 Journals
Supplied Per Interface for Accounting

16

Wiley

Sequential comparison, 3

With the successive increased size of the images we can emphasize just how important our Wiley package is for high quality accounting journals. This is still a pretty simple example, but successive images can help us to more easily show proportions and with unveiling the images can add a bit of drama, which could aid in creating a memorable impression.

Collection Breakdown by Journals: 2013-2017 Usage, Subscription Usage Percent, and 2017 CiteScore

In the past we have used treemaps as a way of showing the value of individual elements within a whole. You can see an example we used previously in Tableau. Treemaps have been more effective in many cases than the old pie chart…especially when we have many individual elements to show.

Treemap?
Count of Current Access Quartile 1 Journals Supplied Per Interface for Accounting

Green titles indicate 2nd variable

Could we consider extending the sequential comparison idea to allow for something resembling the treemap? This concept allows us to include that second variable as this very simple example highlights…shading sequential images representing journal quality for something like usage or cost.

We could even change the image of individual packages we wanted to pop out and be memorable within the treemap. The hope would be that we might both demonstrate a more complex concept and cause specific elements to stand out and become more memorable.
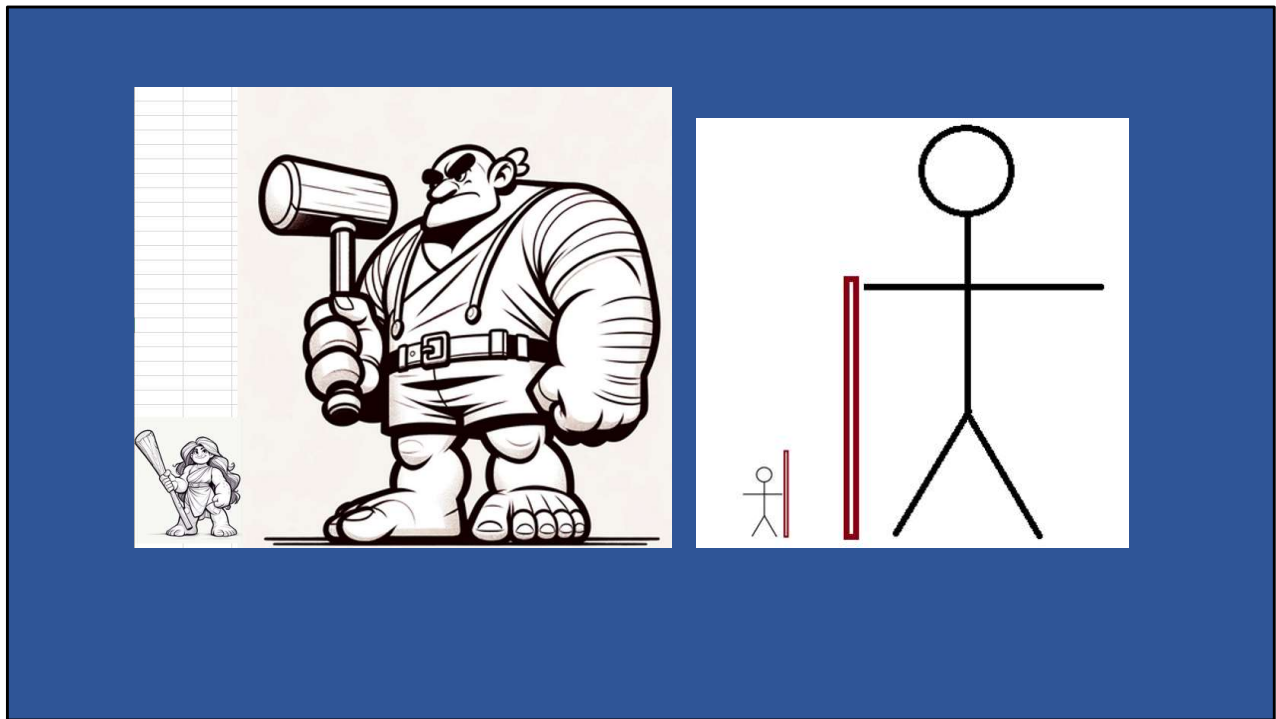
This treemap example gets us closer to that larger goal which is to create 2 variable illustrations in which the size of objects is truly proportional to the data we are trying to convey. The two giants are a great example of this. The size of the giant is the cost of the package, but the size of the club is usage. The accurate proportions and legibility of the stick figures are the goal, now can we use AI to generate the giants to meet that goal? Right now, it doesn't seem that the AI tools we have used for image generation handle numbers well enough, but we hope and expect them to improve.

As we have negotiated this challenge of not easily being able to generate proportional images, I think we do see a middle ground. For images to be effective representations of data, there needs to be contrasts in size. We can generate images where the contrast might not connect directly to the data we have, but the images can show a symbolic contrast. One example of visualizations we have created in the past are comparisons of one element against the average, the whole, or between 2 competitors. Here is an example that demonstrates that our ASME package has a high cost per use. The top factory shows an average cost per use, and the bottom shows ASME's. The steam or smoke coming from the smokestacks isn't exactly proportional to the cost per use, but does symbolize that that the higher cost per use is a problem.
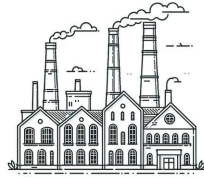
Low Cost Per Use

Such as Springer

Moderate Cost Per Use

Such as IEEE

High Cost Per Use

Such as ASME

Data Symbols
and Metaphors

In the end we can think of these examples as data symbols. They are not measurable representations of the data but show an adequate contrast to demonstrate the range within the data. And of course the use to steam or smoke helps us communicate that a greater cost per use number has a negative connotation, and hopefully would create a lasting impression and future reference point for discussing cost per use with colleagues across campus.

In this example the choice of image is a key component to creating the impression. We have tended to refer these choice of images as metaphors. In this scenario part of metaphor selection relates to whether the concept connotes a negative or positive impression. Zombies and pollution give us a negative connotation, where a pile of riches or a county fair ribbon might create a positive one.

Column chart using image library
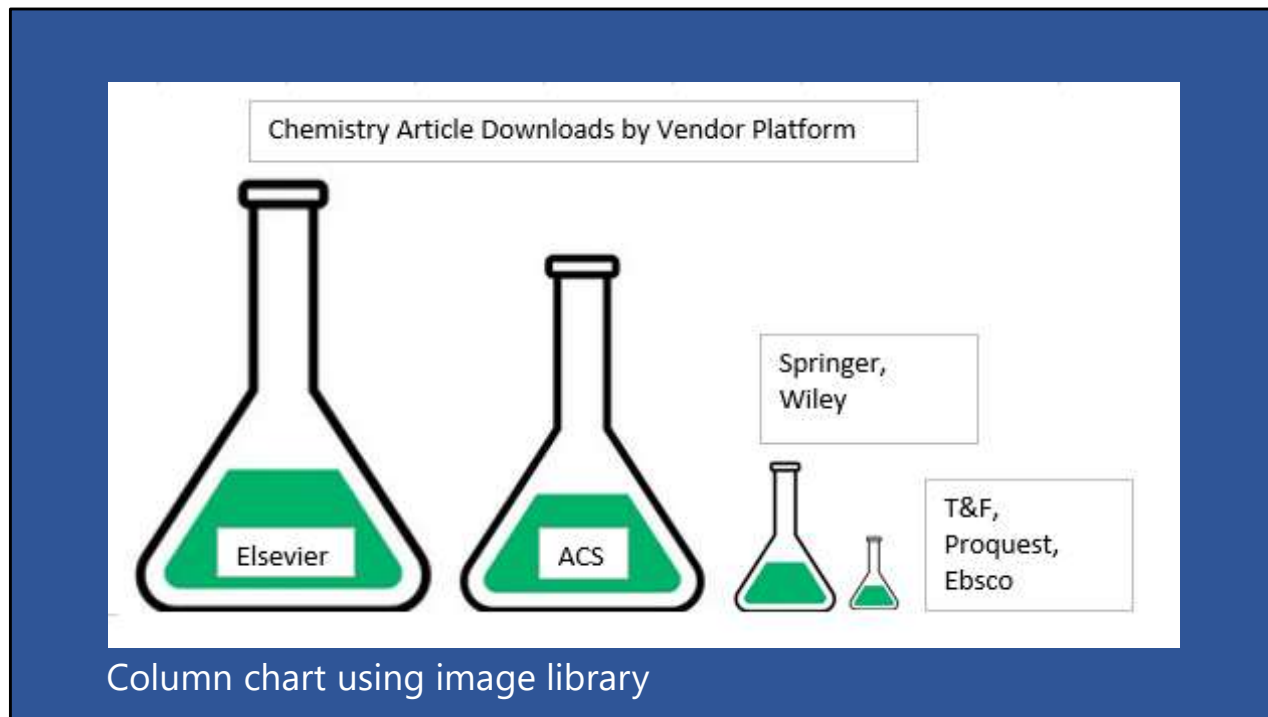
Another approach for selecting image metaphors, is to consider disciplines of our audience or collections we are analyzing. We are often presenting to liaison librarians, academic departments or even creating materials used with accreditation and program review meetings. Our choice of image can connect the presentation to the discipline and hopefully also our audience? That said this does not come without risk. If we choose imagery that plays on stereotypes, or shows a lack of awareness of the department's specific curriculum, it could hurt our credibility rather than help it.

Here is a simple example of using beakers to represent article downloads in Chemistry. The beaker is particularly useful in that it can handle 2 variable concepts because the size of beaker and how full it is can present multiple concepts. But you will also notice that these are not the fancy AI generated novelty visualizations we have been discussing…in fact these are essentially like basic clip art and something we might see in common infographics. For some of the concepts we have been talking about these are pretty effective…what might this look like with AI image generation?

Challenges and Weaknesses

- Simpler is better
- Can images speak for themselves?
- AI struggles with numbers
- Do images hurt our credibility?

This is pretty cool, but it might be that simpler is better.

I think this example of chemistry beakers suggests that in most cases less is more. If an image is too busy or complicated it loses its effect. If we are trying to clarify or emphasize a concept, we want to reduce noise, where a busy AI generated image might actually add noise and muddle our point. This is one reason that sequential comparisons seem attractive. It allows us to show each element being compared one by one.

Another challenge of complexity is how much context is needed for the audience to understand the point. The more we have to explain or add words to the image, the more I would question how effective this method of demonstrating data is for communicating about our collections. It certainly wouldn't make sense to replace a complex chart with an equally complex image that requires as much context for people to make sense of it.

While we have already mentioned this, but as of right now Dall-E 3 does not handle data well. This adds challenges to creating images that demonstrate proportionality. We need to be able to show contrasts to make points. If we exaggerate those contrasts are we moving away from these representing data?

Lastly, while we might reach some of our audience by adding novelty visualization there is a

risk that some might see the images as less serious. Walking the line to maintain credibility is important. We also might not consider connotations others might place on the images we choose. AI can create images that further stereotypes and can be downright offensive. We have to be vigilant that we are not causing harm by our choice of images.

Future Directions

| OTHER LIBRARY DATA | DEEPER STORYTELLING | VIDEO | ADVERTISING AND PROMOTION |

So where do we go from here?

1. Our focus has been on journal collections data because that is at the core of our work, but we could see this type of communication being useful with other types of library data.

2. We might also consider a deeper data storytelling in which we create broader narratives that use a succession of images to create fuller stories.

3. We have tried playing around a little bit with AI videos. There are some concepts that are better expressed with motion…think trend or longitudinal data. That idea of movement might be best demonstrated with video. A rising hot air balloon to demonstrate increasing usage of a journal makes sense to me.

4. One area I think we will definitely pursue is to use images to promote collections and advertise new acquisitions or subscriptions.

Whether or not novelty visualizations become a mainstay of our communication about collection analysis, experimenting with AI image generation has given me a new view of the power of using images to create impressions and will shape my thinking as we continue to try to effectively demonstrate the value our collections provide the university.

Thank you!

gustan2@mnsu.edu
evan.rusch@mnsu.edu

https://libguides.mnsu.edu/collection-analysis/research

Thank you for your patience. You can reach us at the email addresses on the screen if you'd like to talk. If you'd like more information about our library's work on collections, licensing, or collection analysis, you can find a reference list at "libguides.mnsu.edu/collection-analysis/research." We'd also like to thank Dall-E 3 for collaborating with us.

By the way, we recently developed a new approach to journal package analysis using subject-package measures, such as usage per subject per package or citable documents per subject per package. We'll be sharing this work at the Electronic Resources Minnesota online conference on Feb 21. The topic might be interesting to some in this crowd. ERMN is a free conference.