Loren Lee
Prof. Eric Field
SARC 5400: Data Visualization
28 February 2023

Assignment 3: Visual Data Analysis

All of these data visualization tools and methods are completely new to me, and although I've tried my hand at different digital skills (ArcGIS, R, Python, etc.) here and there, I'm still a complete novice, especially since I'm trying to work so far outside of my normal area of study: French medieval literature. Because of this, I've found myself alternating between being so excited about what is possible, feeling surprised by what I'm able to do so far, and growing frustrated with what I still can't manage to create. I used this assignment to make my first attempts at creating data visualizations of some of my own research data: the metadata of forty-three late medieval French manuscripts that make up the corpus of my dissertation research. Here is my dataset: ME French MSS Metadata

1. What is your question? What specific problem or question are you asking of this dataset? What are you trying to know or resolve?

My questions going into this assignment were pretty basic. I, first, wanted to see how I could make my metadata more readable by both computers and people because I know that my spreadsheet as it is now is not particularly tidy. For instance, I do not currently have a standard way of describing the origin dates for all of these manuscripts, and that column of data is filled with ambiguous text like "~end of 13th cent. or start of 14th cent." Then once I could see more clearly what needed cleaning up, I wanted to look for anything meaningful about this collection of manuscripts. Few specialists working with medieval manuscripts attempt to draw conclusions on a macro-level for many reasons: we do not currently have universal standards for cataloging manuscripts in such a way that renders particularly tidy data; there are so many gaps in the archive of what survives into the present day and what doesn't that it's difficult to be conclusive, etc. But because my dataset is of a smaller, more manageable size, I think I can at least draw some conclusions about these specific manuscripts. Do the different versions of my core text, *La Vie de sainte Marie l'Égyptienne*, tend to be copied into codices of similar sizes? Having a better understanding of codex size tells us about what the codex may have been used primarily for (personal or communal prayer, travel reading, etc.). Where are all of these manuscripts? In what

collections should I be spending the most of my time when I have the opportunity to do research in person? How have these manuscripts traveled over time? And then, of course, I was hoping to notice something new from my dataset. What have I been overlooking? What do I need to see more clearly in this data?

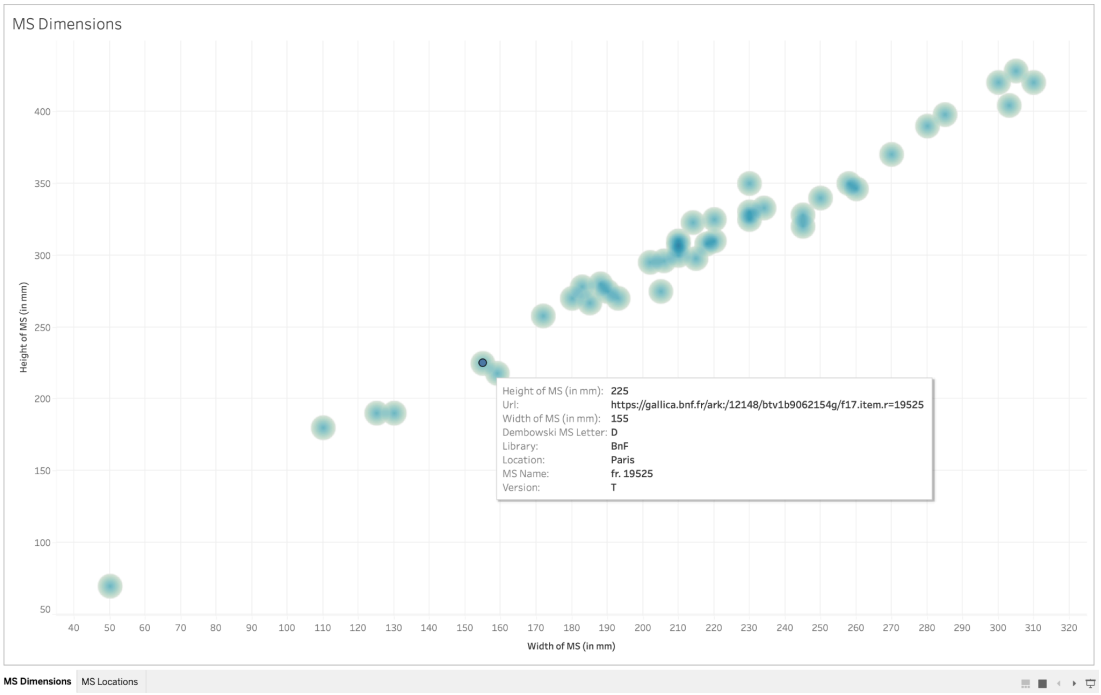
2. What visualization approaches and techniques are you using, and how do they work? Structurally, how are you going about resolving the question? Visually, how are you organizing and encoding the data to see into it? What specific techniques are you trying, and why? Consider the types and strategies that we have been discussing. How are you using color, shape, weight, location? Use our discussions of cognitive load theory, chunking, schema development, split attention, ... Refer back to Tufte's `Escaping Flatland` and `Grand Principles`. Now that you've tried them, how appropriate are these strategies to your question and to this data? How effective are these strategies at organizing, comparing, revealing, and increasing your understanding of the issue(s)?

3. What insight did you gain? Specifically, did you discover something about your question? Why? How exactly did your visual / organizational technique help (or hurt) to do so?

4. Compare your techniques critically and suggest improvements. Did one technique or approach show something uniquely different or better than another? What did each reveal that the others did not? What did each hide? What could be done better to improve your understanding of this data or question? What might be the next steps? Critique your strategies!

Initially, I thought I would like to try working on Observable to create my visualizations, but then, I was feeling lost when it came to the minor coding aspects of Observable, even though in general it’s really user-friendly. I think because I know so little about what can possibly be done with these digital tools, I find myself not knowing where to begin or even what to look for to meet my ideas. I turned to Tableau, which I found I could learn to work with much more quickly since it has that menu “Show Me” at the right that guides you toward how to build different visualizations. With that guide, I could quickly see what elements I needed to pull in in order to make different visualizations.

Prototype Visualization #1: Manuscript Dimensions (Tableau Density Plot)



For my first visualization, I wanted to tackle the question of size. What is the spread of height and width I'm seeing in this set of manuscripts? Do certain versions of the text tend to appear in manuscripts of particular size? Is there no discernible correlation between text and codex size? Manuscripts are typically measured in millimeters, so these two measurements make up my x and y axes. (Thankfully, I had already made this information tidy in my dataset by having height and width in two separate columns in my data source.) Then, I wanted to be able to quickly see which manuscript copy is which, so I next added attributes to label each point. When you mouse over each point on the plot, you have this small pop-up window that gives you basic information about the manuscript, and if you click on the point, it will take you directly to the digitized manuscript (if it has been digitized) or the digital catalog record of that manuscript. I like how the density bubbles allow you to see how the sizes of these manuscripts overlap quite a bit without obscuring the individual data points, and you can see that there is also a great range of sizes. (Though I should note that the smallest one is only so small because it is a fragment of a torn manuscript and not a full codex.) I think the plot would be better if the points were all labeled in the chart with their version letter (T, X, V, W, N, O, etc.) so that you could instantly see where all the versions might be clustering, but I couldn't quite work out how to add these

labels to the bubbles. Another annoying thing I couldn't change was the order in which attributes appear in the pop-up window. Ideally, I would like the order to be Version, Dembowski MS Letter, Location, Library, MS Name, URL, but I couldn't figure out how to make this small change for the time being.

I know that this bubble plot is really simple, but I found it really helpful in getting a better sense of my manuscript sizes. I have visited some of these manuscripts in person, and each time that I saw one in person, I was taken aback by their dimensions (either by how large or small they were!) because I have such a hard time thinking in millimeters. Perhaps a better (or additional) visualization for this same data could be some sort of visualization with stacked rectangles representing the manuscripts themselves. That way, you could more easily compare individual codices to one another as though you were physically stacking them on top of each other — something you're definitely not allowed to do in special collections libraries unless you're the great Will Noel, Eric White, and Paul Needham at Princeton →

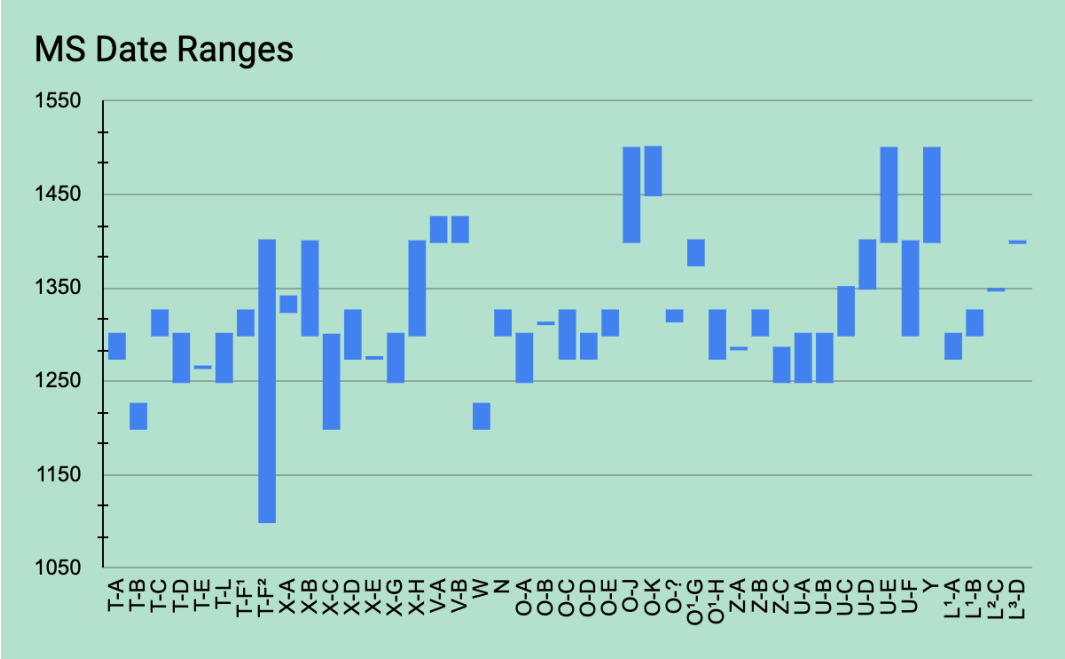
Prototype Visualization #2: Manuscript Locations (Tableau Density Map)

My second visualization was also developed in Tableau. For this one, I wanted to address questions of location: where are my manuscripts currently located? Where are they clustered together? Does this tell me something about their use and how widespread the text was in the later medieval period? I had a lot of trouble at first putting this map together because my location information in my dataset was not very neat. Originally, I had the column "Location" in my spreadsheet, which included both city and library location together, so first, I separated these out into two separate columns. When I pulled this information into Tableau, I was hoping Tableau would automatically generate longitude and latitude for me from the names of the cities in my dataset, and Tableau did try to do this, but I kept getting error messages that prevented the whole thing from working. So then, I just found all of the longitude and latitude information about each location myself using Google Maps. This took a little bit of time, but since I only have forty-three manuscripts (and many of these are located in the same set of libraries), it didn't take me too long to manually generate this data. Then, I was finally able to pull it all into Tableau and successfully generate this map. As with my first visualization, if you mouse over each point, you will have a pop-up window telling you details about each manuscript. I again decided to use the density feature to show where the majority of my manuscripts are located (many are in Paris as you can see on the map), but I'm a bit dissatisfied with how I can't easily identify individual

manuscripts located in the same place. Yes, I can see that many manuscripts are in Paris, but which ones and how many exactly? I think this map could be improved by some sort of layering that would allow me to see which manuscripts are where more quickly. Maybe this could be done with some sort of list or legend off to the side inserted using Adobe Illustrator?

Something that would make this visualization much more meaningful and dynamic too would be if I could add a layer of locations by *provenance* of each manuscript rather than just their current locations in special collections libraries. Sure, this map tells me about where all my specific manuscripts are, but really, that just means I've mapped the locations of significant French manuscript collections in Europe, and that is not very enlightening information. Instead, I could have a layer for each location type (original location and current location), and then viewers could toggle between the layers to see how manuscripts move over time. If I had a lot more information about the provenance of these manuscripts, I could even make a dynamic visualization that follows the paths of these manuscripts traveling over time. It could also be interesting to color code the dozen different versions of the text I'm examining so that I could easily see where different manuscript versions are clustered together. I'm not currently able to do that though because I only have provenance information for about half of these manuscripts, and the other half are all rather ambiguous geographic regions like "northern France". Most of these manuscripts lack detailed information about their movements from collection to collection over the centuries, but I think that trying to visualize this trajectory over time would be useful for helping viewers feel the sense of awe that I feel when I think about how long these artifacts have survived and how far they have traveled in the hands of countless people. You might notice that there is one outlier on my map: there is one manuscript fragment of my text that was discovered in Damascus, Syria. Unfortunately, this fragment has now been lost, and all that survives of it is a grainy turn-of-the-century, black-and-white photograph. Despite the fact that the physical manuscript is gone, I included the Damascus fragment on the map to demonstrate how spatial information may challenge our assumptions about the medieval period. Namely, the assumption that medieval Western Europeans did not travel or trade with people far away from home. Someone carried a twelfth-century Anglo-Norman manuscript about an Egyptian saint hundreds of miles to Damascus. Why? Without this spatial visualization, someone looking at my dataset could easily gloss over this fragment without considering its implications.

Prototype Visualization #3: Manuscript Date Ranges (Google Sheets Candlestick Chart)



I struggled the most with this last visualization, and I wasn't able to get Tableau to do what I was envisioning, so I turned to Google Sheets to generate a more basic, ugly chart. The final prototype is not exactly what I would have wanted, but it does get at the kind of idea I was hoping to visualize. What I was hoping to do was consider the data I have available concerning the date ranges for the approximate origins for each of my manuscripts. In my mind, I wanted to create a sort of timeline in which the spans of years for each manuscript appeared adjacent to one another, side-by-side on the same chart. I wanted to do this with the axes flipped from how they are now (with years spanning horizontally and the labels of manuscripts on the vertical axis) because I think, in general, we tend to imagine time in a horizontal orientation. But for some incredibly annoying reason, the Google Sheets chart function demanded that the y-axis have a numerical value. However, I do think this visualization does somewhat address my questions about time. For one thing, you can see how precisely (or not) we are able to estimate the dates for the different manuscripts. Some have dates as precise as "August 14th 1399" while others simply have descriptors as broad as "late medieval," and so their bars on the chart are much longer. You can also see on this chart how some versions, like the verse redaction Version T for example, appear and circulate somewhat earlier than the later prose versions, which to me suggests Version T's proximity to the first (now certainly lost) vernacular translation of *La Vie de*

sainte Marie l'Égyptienne in French. This visualization also allows us to see the overlap of these manuscripts that were copied and circulated among readers at the same time, and it gives us some sense of when the majority of these manuscripts were produced (about 1250-1350 – Maybe inserting a color band to highlight this range of years would have been helpful?). These dates are for the most part only estimates though, so I do not think I can safely say much more of anything meaningful about these ranges until I can get my hands on more precise data.

In terms of improving the visualization itself, I think the current iteration of this prototype could use more detail. Maybe annotations of the more precise dates would make this visualization more clear? I worry about making visualizations too cluttered, but I suppose it can't hurt to be precise. Maybe color coding could also be applied to make it clearer which bars are connected to which versions? At the moment, I see that it's difficult to tell which bar refers to which manuscript since they're all so smushed together.