Reflections on Project – Ben Wright

There was a lot I learned over the course of this project. Firstly, and most importantly, I think it made me a better researcher – not just for visualization but also in terms of implementing other CS algorithms and other academic and applied researching skills. I know I only cite one paper in the final report, but there were several times I had to look things up in medium articles, documentation, and forums to move forward with either project ideas or implementation.

Second, I learned how data transformations in high dimensions can really affect dimensionality reduction algorithms. For example, some of the datasets that came preloaded for this assignment had eight-digit identifiers for each record entry, and all other coordinates of a data point were small in absolute value. This made the data live on a line, basically, so applying T-SNE and UMAP did not make very enlightening pictures. Further, I didn’t realize before doing this project that doing data transformations can be a strict visualization trick.

One of the visualization ideas that I kept encountering over the course of this project was more of a design concept than a strictly visualization idea – it was that every choice you make should have a purpose and a justification. One thing I feel bad about, for example, is my conscious choice not to rescale the 2D data so that the same axes can be used for every dataset. But this was a choice that I have justifications for – for example. One reason I didn’t want to do this is because what if the actual values of the data under the transformation have meaning to the user? Also, doesn’t it make it more comparable across multiple algorithms allowing the axes to rescale themselves? I don’t know for sure, but this class opened my eyes to the idea that this is a choice that I made, and for that reason, I should have justifications for doing things a certain way. Another choice I made was to use interaction in changing underlying chart data rather than to display all the charts at once. Again, just an example. This is not the place to justify why I did that.

Lastly, I learned specific programming tools and algorithms like Google Collab, matplotlib, and this cluster sharpening algorithm from the paper I cited in my report. I was already a decent python programmer, but unique challenges arose trying to implement interactivity in Python notebooks…