Introduction to Spatial Data Manipulation and Visualization

Final Project - Catalina Burch

I chose to investigate the WDFW culvert prioritization index (PI) in comparison to other PIs within the Washington Injunction Case Area. My research questions were:

(1) How many culverts are scored by both the WDFW PI and the local PI?
(2) How do the two scores for the same culvert compare?

This project required working with a variety of data-frames and geographies. I first downloaded culvert inventories from all of the jurisdictions within the case area including, WDFW, Cold Water Connection Campaign, Chehalis Basin, Thurston County, King County, and Bellingham. Second, I downloaded boundary geographies for the Case Area, Water Resource Inventory Areas (WRIA), counties, and Bellingham city limits. I converted the PI inventories to sf objects and matched IDs from local PIs with the WDFW inventory. This method identified the number of culverts that were scored by both WDFW and each local PI. I chose to compare the PI scores by making a bivariate choropleth map for each jurisdiction.

In the process of making my final maps, I encountered many tricky design decisions. For my basemap, I tried several different options before settling on the toner-lite Stamen map. I liked this map because it was simple, and the black and white color scheme does not overpower or distract from the colored data points. Kent and Klosterman (2000) note that beginner map-makers tend to design overly complicated and busy maps, so I tried to avoid that pitfall when choosing my basemap. The next design decision was to choose how large to make my data points. I chose a large point size for Bellingham and King, which have lower n values, and a smaller point size for Chehalis and CWCC, which have higher n values. This decision was to try to reduce overlapping points. I assume that the viewer (Sunny) will be looking at the maps on a computer, and has the ability to zoom in for more detail. I also added a black outline for each point so that the lightest color would not fade into the background. I omitted the outline for the high n value maps because it obscured the color of the points. I added opacity so that the viewer can see where the points overlap. I debated the necessity of adding a scale bar and north arrow, but ultimately decided it would be wise based on the 5 basic map necessities stated in Kent and Klosterman (2000). Although, I did choose to omit several of the 5 necessities from the map itself because they are included in this report, such as the map maker, data sources, and inset map (included in presentation).

In addition to design decisions, I also faced data visualization and interpretation choices. Both Monmonier (2005) and Kent and Klosterman (2000) discuss the dangers of choosing data breaks for choropleth maps. They give multiple examples of how binning can significantly change the interpretation of the data. Thus, it was important for me to consider how breaking the data changes the interpretation of my figure. I normalized all of the PI scores to a 0-100 scale so that they could be directly comparable. I then chose to break the data by 3 quantiles, which sorts the observations evenly into three categories, low, medium, and high scores, where each category represents 33% of scored barriers. From a resource managers perspective, this representation is useful if the manager wants to remove a target amount of barriers, but is less useful if they want to parse out a smaller number of outlying high scoring barriers. From the perspective of my research question, sorting by quantiles allows me to compare if the top 33% of WDFW barriers are similar to the top 33% of local PI barriers.

I found that my method worked well for the smaller jurisdictions, but not as well for larger jurisdictions. CWCC and Chehalis both had a lot of barriers that overlap due to the spatial scale of the map. To try to reduce overlap I made the points very small, but when you zoom in to see the points you no longer have the legend visible for reference. A visual interpretation of all the maps shows that the WDFW score varies often from the local PI score, which is apparent because of the number of barriers colored pink or teal. I had a loose hypothesis that WDFW would more often prioritize upstream barriers, because they do not consider upstream passability in their scoring methods. I didn't see any obvious proof of this theory, because there are pink and blue dots scattered throughout the watershed, not just upstream. CWCC is a potential exception, because it appears that there are more teal dots on the coast, and more pink/purple dots upstream. However, statistical analysis is needed in order to draw any definitive conclusions on this point.

Throughout my analysis I encountered several road blocks. The inventory for Thurston County did not have coordinates or an ID to match with WDFW, so I omitted this PI from my analysis. Four of the Bellingham culverts IDs did not match WDFW, which was my only means of attaching a geography to the Bellingham scores, so omitted these four data points. My original plan was to use scatterpie to map each jurisdiction showing the proportion of barriers scored by WDFW, local jurisdiction, and both PIs. I decided to pivot from this idea after creating a preliminary plot. I found the scatterpie package to be a bit buggy, and that the information was better displayed in a table or bar graph. Although, choropleth maps also have inherent limitations. For example, Monmonier (2005) states that bivariate choropleth maps force the user to look back and forth between map and key, which is not as intuitive as a single hued scale. This issue is why I chose to only break into 3 quantiles, so that there would be fewer colors, which is less confusing for the viewer. The final issue is the most significant, which is that the dots overlap and many are concealed. I've thought a lot about how to fix this issue and didn't reach any particular solution. Adding opacity makes it harder to match the color to the legend, and making the dots smaller forces the viewer to zoom way in to identify the color of the point. I looked into adding jitter, but I couldn't figure out how to add jitter to an sf object. Ultimately, I decided that if I had more time the best option would be to display the larger jurisdictions on leaflet, that way the viewer can zoom in to the clusters and uncover hidden points with the legend scaling down to the zoom. Overall, I learned a lot in the process of making these maps and I'm satisfied with the final product.