

CSC 465

Homework 2

Submit a PDF file with your answers to D2L. Clearly label which answer and visualization goes with which question. If it is not easy to find your answers, you may lose credit.

Include text answering questions and images of your visualizations (from screenshots or copying and pasting right from Tableau or RStudio into your document). Explain very briefly how you created the visualization and include R code files in your Dropbox submission.

- 1) Download the FoodSrvByCounty.txt file and create the following visualizations for this geographical data. The data is for the availability of food services by county in the U.S. It also has data by state (in the county field, some of them have the state names, and those rows hold the state totals, or you can aggregate by state)
 - a. Graph food services by state with an appropriate geographic visualization. Note any patterns that arise. Your visualization should clearly display states that have high levels or low levels of food service availability, so think carefully about the color scheme.
 - b. Graph food services by county with the same type of visualization. Again, think carefully about the color scheme.
 - c. **(Extra credit)** Research how to do a diffusion or tile cartogram in R or D3 and create a cartogram of the state data from this dataset.
- 2) The Chicago_crashes.csv file contains information on every crash recorded in Chicago in June 2019 (see Chicago's portal at <https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if> for the latest data. I chose a random month because the data get dense quickly).
 - a. Create an appropriate type of geographic plot to show where all the accidents in this data occur.
 - b. Create a visualization that shows how common crashes are in different parts of the city based on time of day. There are multiple approaches to this. Explain your approach and what you can see in your graph.
- 3) **(20 pts)** Download the Portland Water Level dataset and explore it by creating the following visualizations of the time series from the techniques described in lecture. Use both R and Tableau for at least one question part. They should, of course, adhere to the design criteria that we've learned, and should clearly display the information described in each part.
 - a. This data contains a year of data with water level (*WL*) measurements every hour as a function of *Time* (i.e. 365 x 24 data points!). Since there is a lot of data, clean it up by smoothing the data by calculating a moving average. Use a window approach with

window size that covers a range of days (remember, the data is hourly) and graph the smoothed result. Work with the window to see what size window gives you the best view of the changes in the data while still smoothing the noise well.

Remember that the moving average is in the Quick-Table calculations inside of the right click menu on the data item in Tableau, and we can compute it in R quite easily as shown in the tutorial.

- b. Graph the cycles that happen each day (because of tides). You might try overlapping many days' data as separate overlapping time series, using a level plot, a horizon graph, etc. The point of this exercise is to try to come up with a way of showing the progression of the tides over some period of time that is rich and detailed and which shows the pattern, but which is still readable and which doesn't clutter the graph.
 - c. Then write a single paragraph outlining the differences between the information that each graph communicates.
- 4) Return to the Portland Water Level dataset. Recreate one of your plots from Question 3 with a custom color scale. Specifically, create a divergent color scale with the average water level at the midpoint and two separate colors used to show when the water is getting very high and very low. The point of this exercise is to experiment with creating a color scale, so choose your own distinctive colors to use for the endpoints and center. Make sure that they are reasonable choices given what you know about color scales. Use HSV space to choose the colors and explain how you made your decision. In Tutorial 4, you can see how to create color scale in ggplot that is interpolated in Lab space.