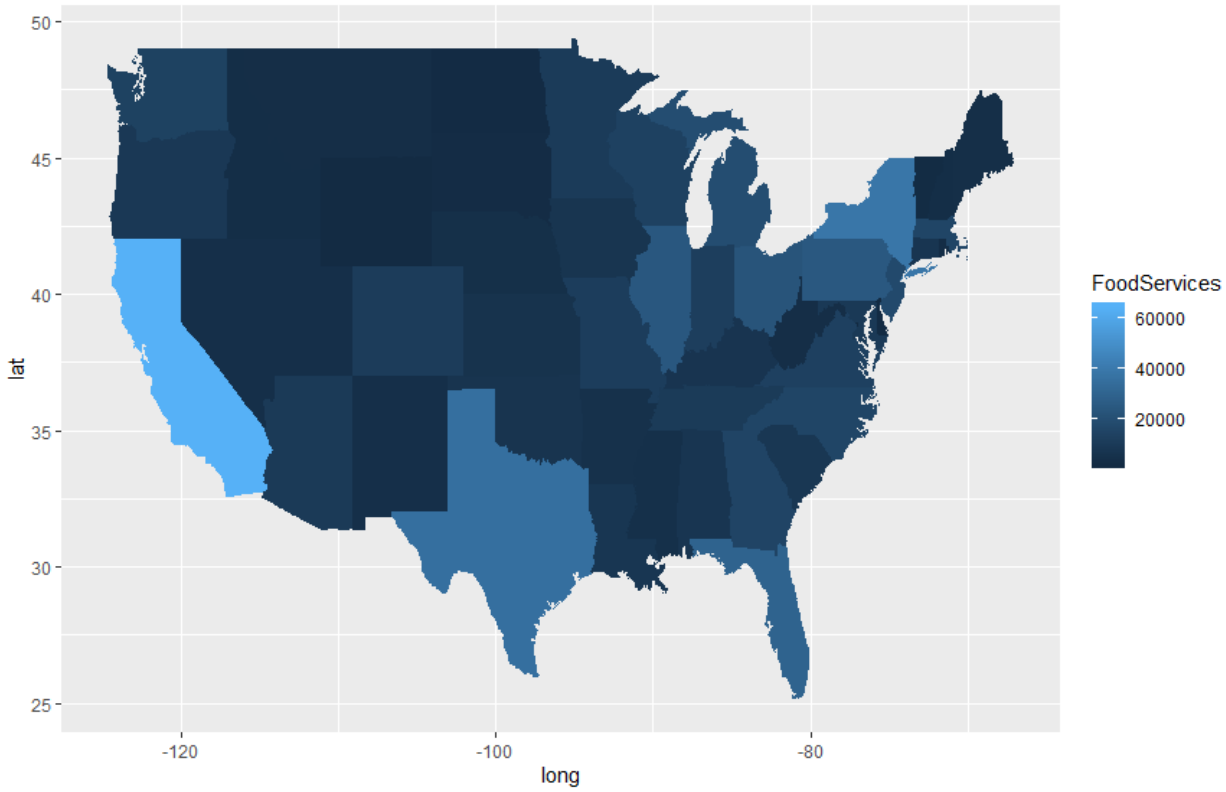
1) Download the FoodSrvcByCounty.txt file and create the following visualizations for this geographical data. The data is for the availability of food services by county in the U.S. It also has data by state (in the county field, some of them have the state names, and those rows hold the state totals, or you can aggregate by state)

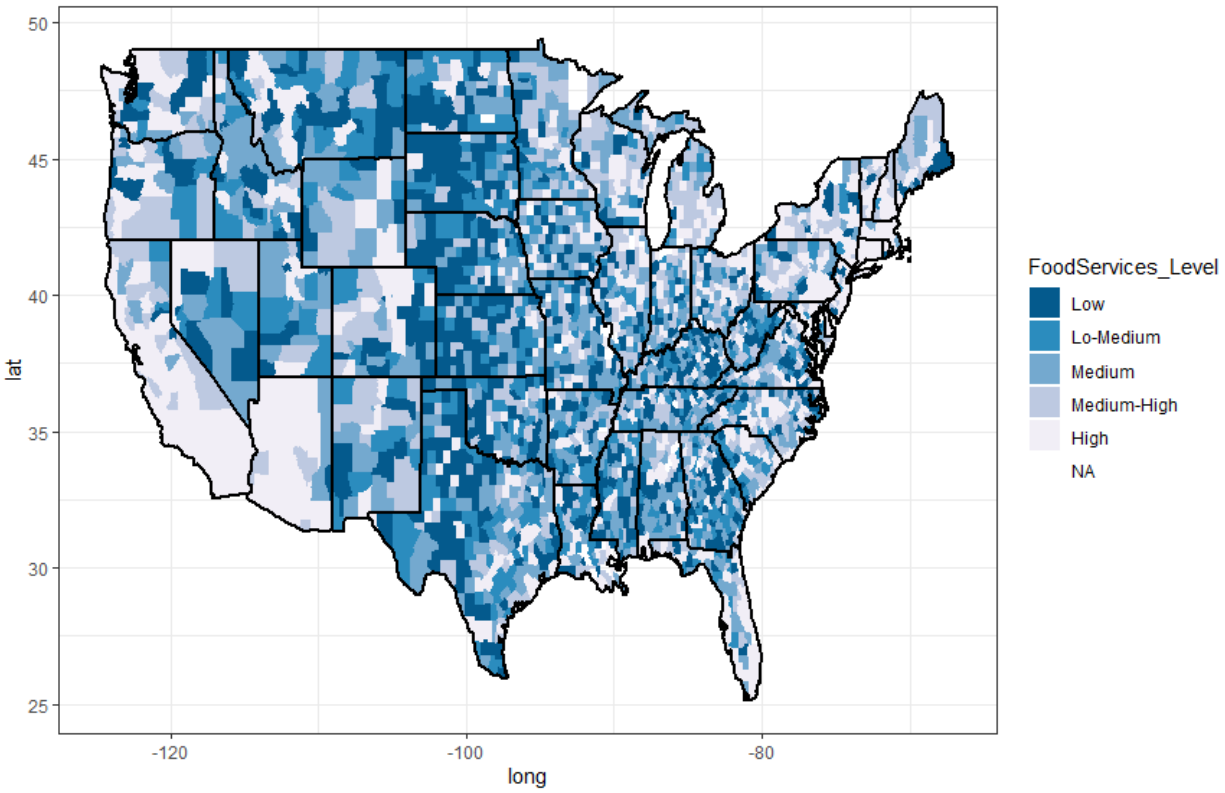
a. Graph food services by state with an appropriate geographic visualization. Note any patterns that arise. Your visualization should clearly display states that have high levels or low levels of food service availability, so think carefully about the color scheme.

To plot this graph, I first reformatted FoodSrycByCounty table by selecting State data only and change state names to lower case, so the table can be merged with states map table. After that, I added a new column called FoodServices, which is median of 3-year-FoodServices. I used medina amount to fill the geographic graph, and the darker the state shows the lower FoodService level the state has.



b. Graph food services by county with the same type of visualization. Again, think carefully about the color scheme.

As this is by counties, geographical areas are much smaller than food service level by states. Thus, I chose to use thick black line border states so it's clear to see how each county performs within the state. I grouped data into 5 difference levels, from low to high, the lighter the square is the higher level of food service that county obtains.

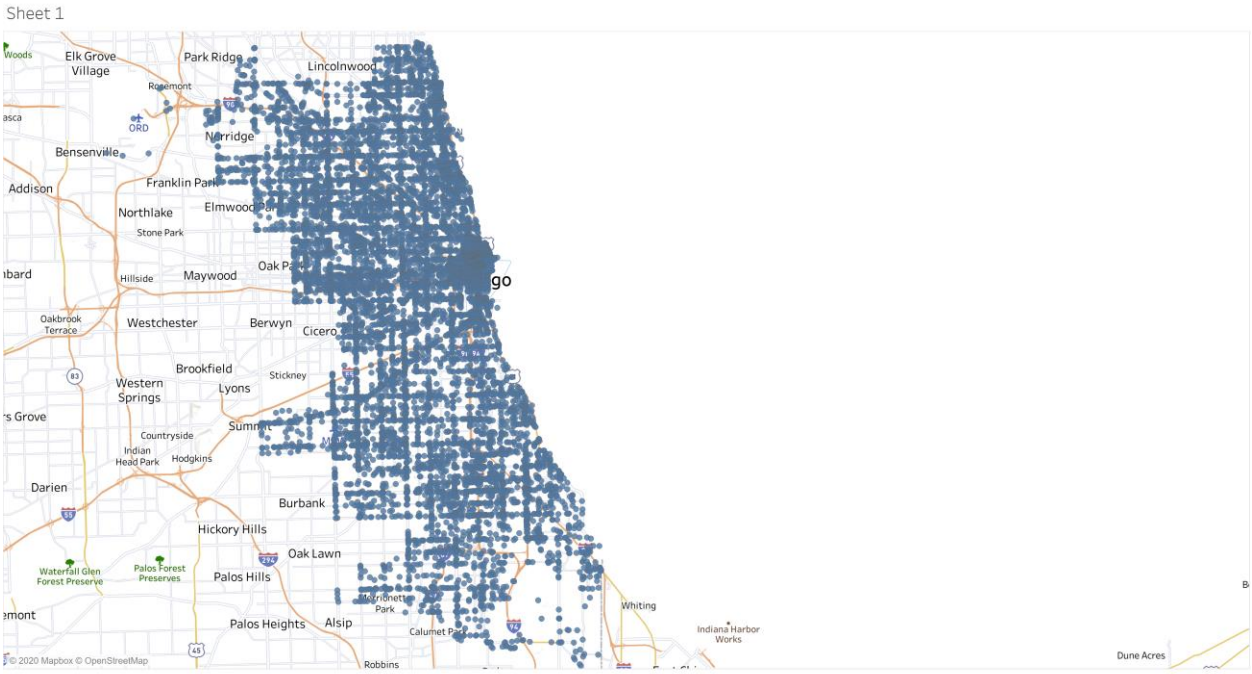


c. (Extra credit) Research how to do a diffusion or tile cartogram in R or D3 and create a cartogram of the state data from this dataset.

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-CrashesCrashes/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.

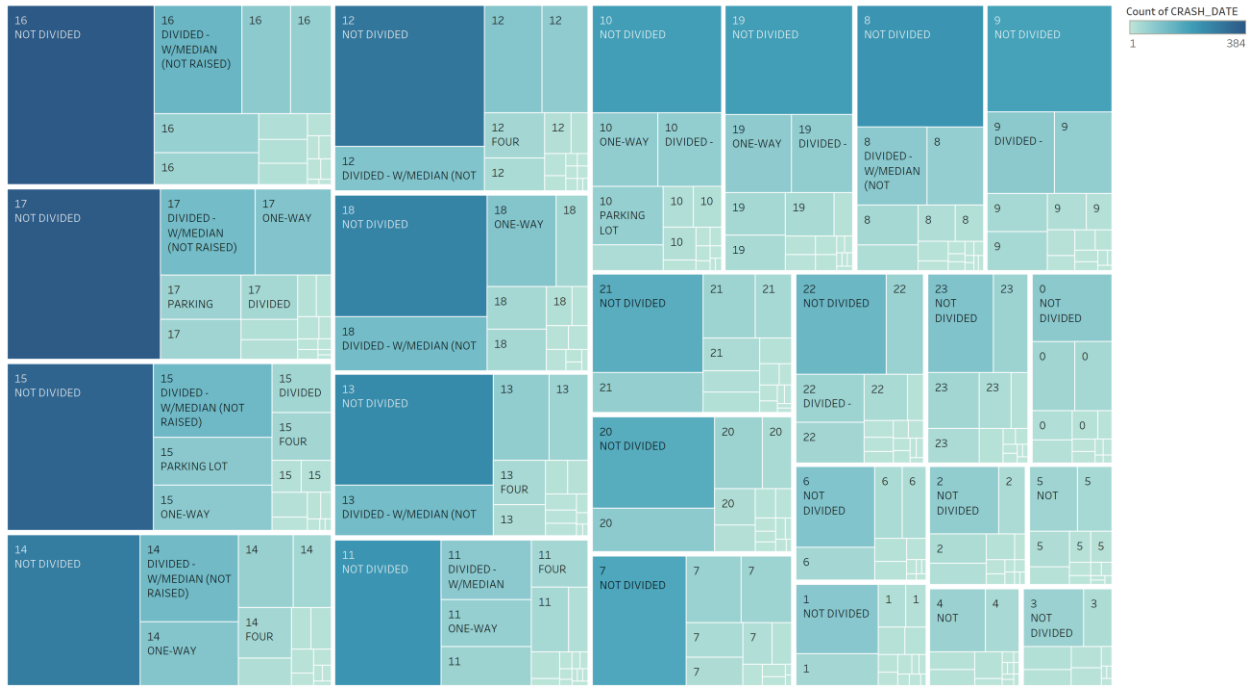
I used longitude as column and latitude as row in Tableau. For the graph type, I used symbol maps to show streets, highways and neighborhood names. Each blue point represents the location of crashes, and it's easy to see which parts of Chicago has crashed happened based on the street view of map.



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.

I used heatmap to show how common crashed are in different parts of the city. I used traffic way type to separate parts of the city. Each big square represents the hour (from 0 to 23) at when crashes happened, with the big square, each small square means different traffic type, and the area of the square determines the number of crashes happened.

2.b

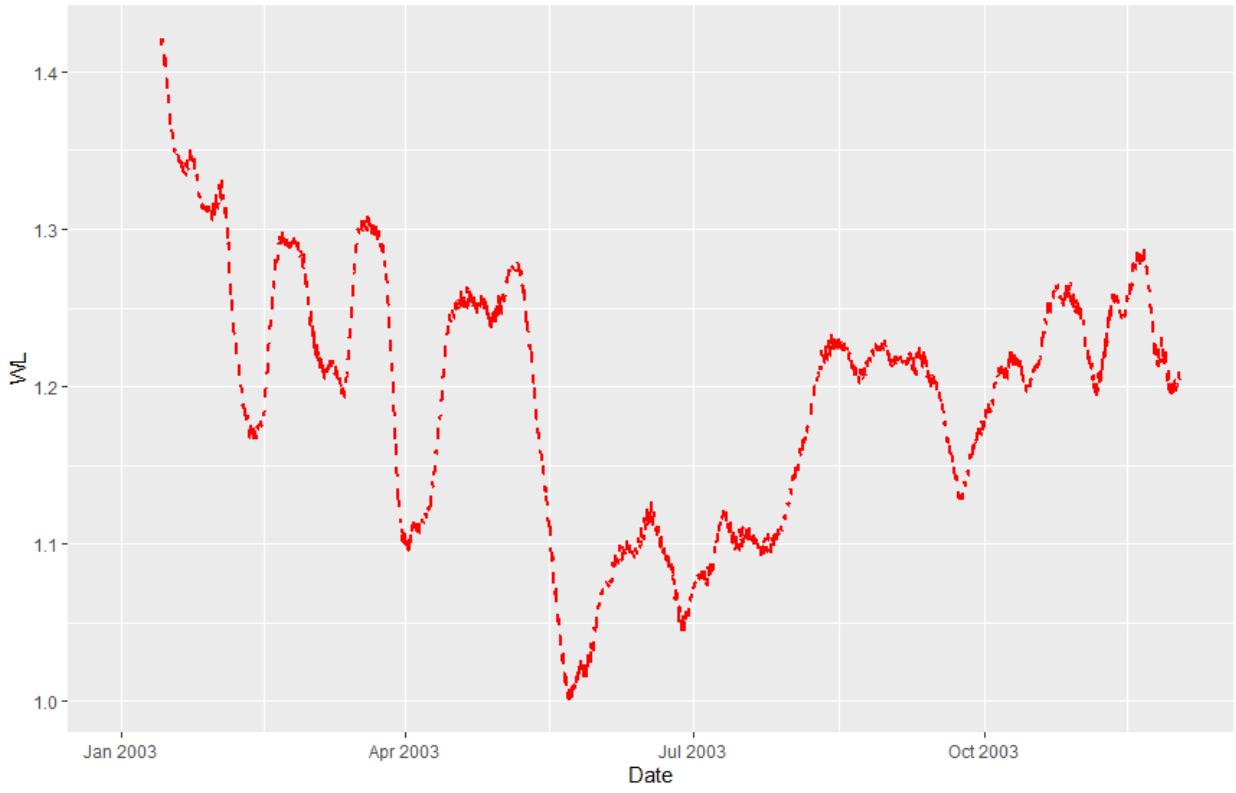


CRASH_HOUR and TRAFFICWAY_TYPE. Color shows count of CRASH_DATE. Size shows count of CRASH_DATE. The marks are labeled by CRASH_HOUR and TRAFFICWAY_TYPE.

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.

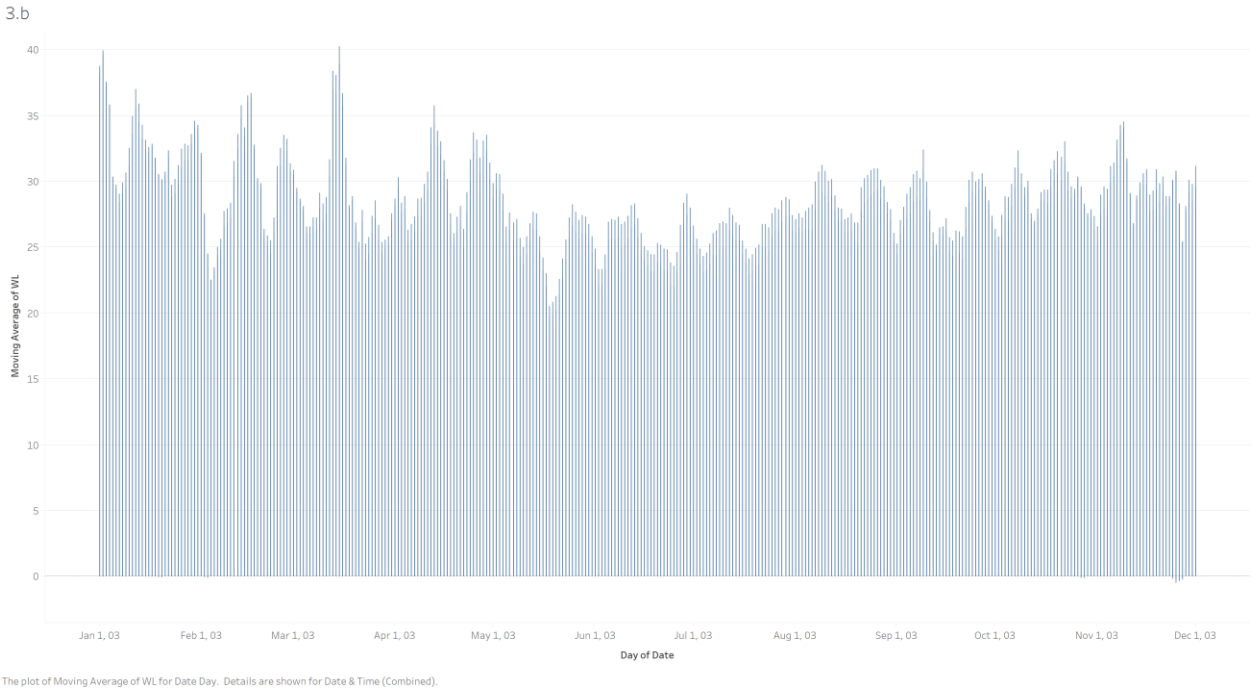
a. This data contains a year of data with water level (WL) measurements every hour as a function of Time (i.e. 365 x 24 data points!). Since there is a lot of data, clean it up by smoothing the data by calculating a moving average. Use a window approach with window size that covers a range of days (remember, 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.

I have tried different window sizes range from 100 to 900 and came up with 310 is the most appropriate size the gives me the best view of the changes and while still smoothing the noise well.



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.

I used Date-time as columns and moving average of water level as rows. Each bar represents water level movements within one day. There's apparent pattern shows that at the beginning of every month, water level is lower than the rest of the month.



c. Then write a single paragraph outlining the differences between the information that each graph communicates.

The overlapping graph shows more detail when there's small changes in water level than those with window approach. Additionally, the line graph is easier to tell big changes and lowest and highest points, while bar graph is bit harder to compare.

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.

I chose light royal blue to purple to dark blue to black as color scale. The lighter the color is the lower the water level and vice versa.