CSC 465 Homework 2

Name: Kavana Manvi Krishnamurthy

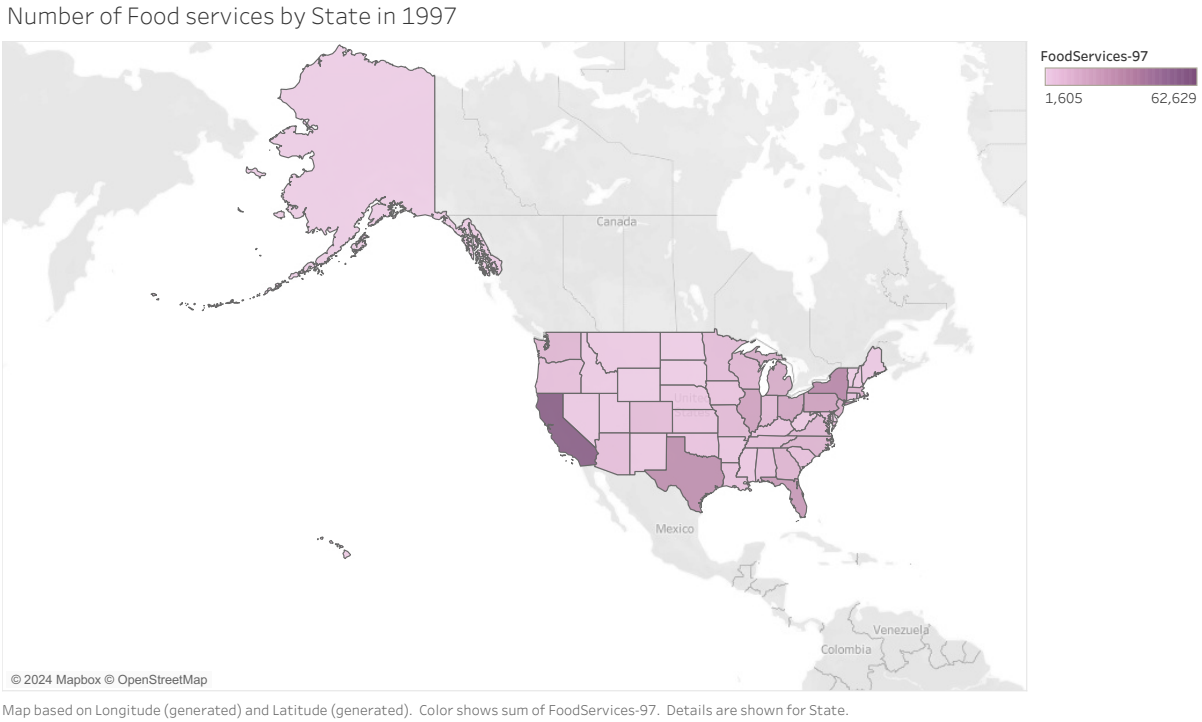
Student ID: 158984

Question 1

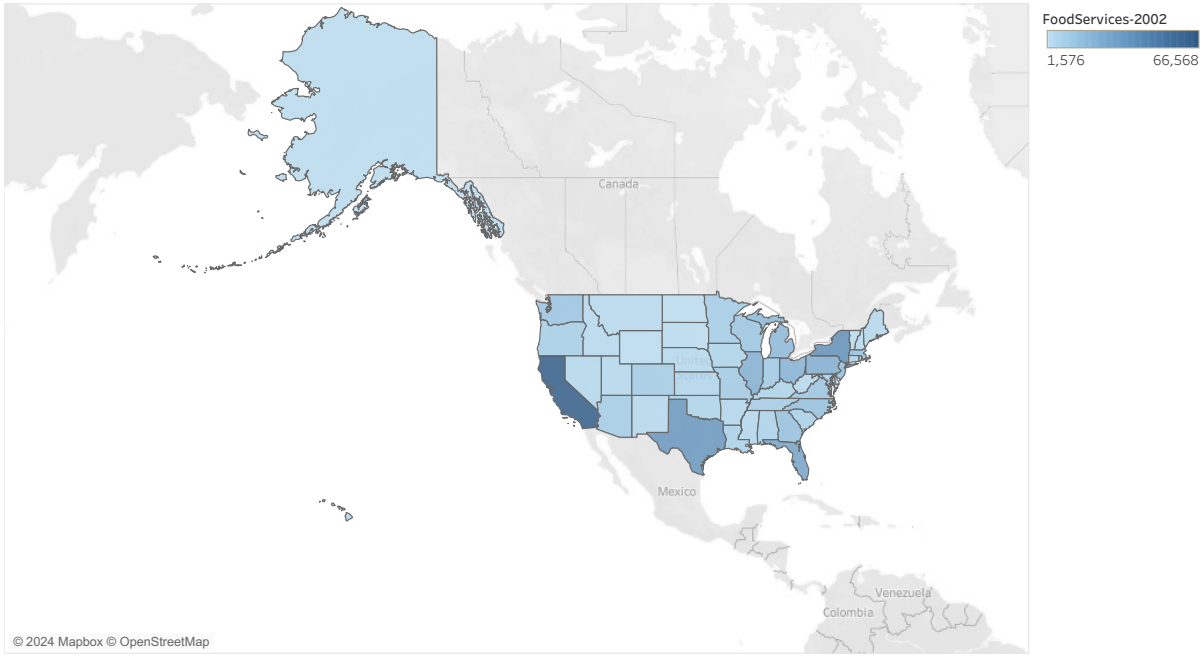
Download the FoodSrvcByCounty.txt file and create the following visualizations for this geographical data. The data is for the availability of food services by county in the U.S. It also has data by state (in the county field, some of them have the state names, and those rows hold the state totals, or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

I have used a Choropleth map to graph the food services by state. It is a great visualization to see the information by state. I have chosen a colour gradient that transitions from light to dark. The light colour indicates a low service while the darkest colour indicates a high service availability. We can see that urbane states like California, Texas, Florida, NY, Illinois etc have a darker gradient and, hence have a larger food service availability compared to other states. I have added similar maps to indicate food service counts by state in 1997, 2002 and 2007.

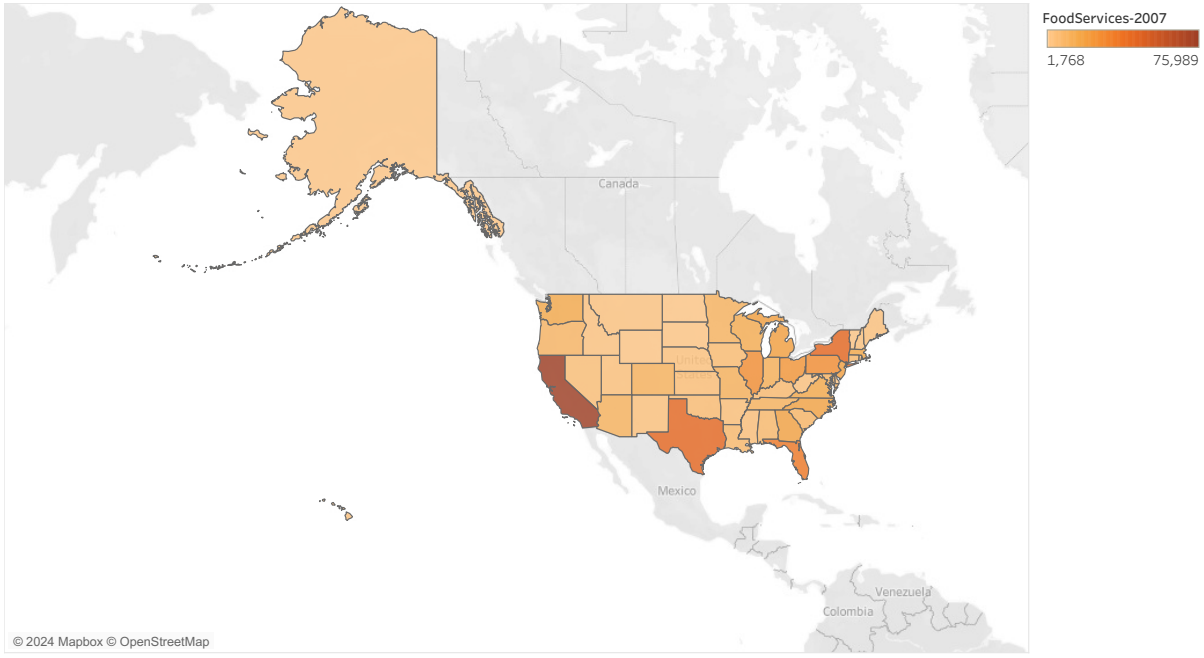


Number of Food services by State in 2002



Map based on Longitude (generated) and Latitude (generated). Color shows sum of FoodServices-2002. Details are shown for State.

Number of Food services by State in 2007

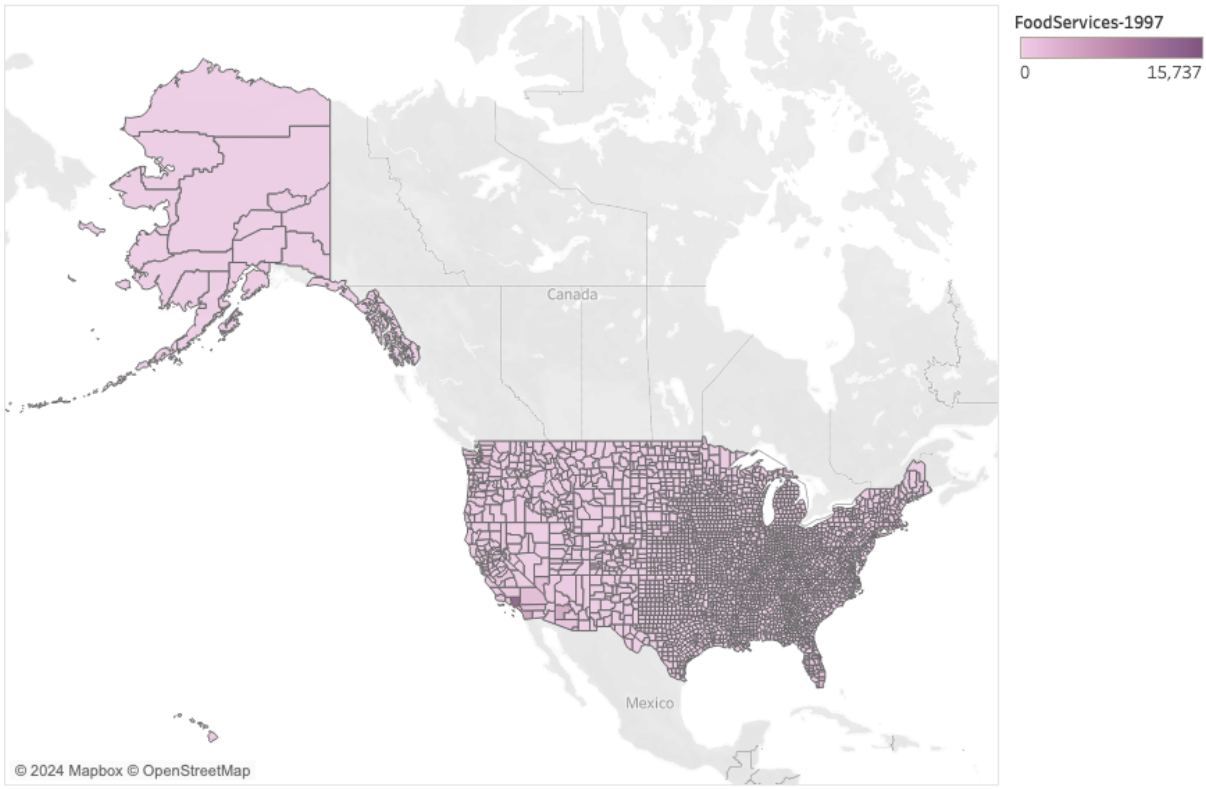


Map based on Longitude (generated) and Latitude (generated). Color shows sum of FoodServices-2007. Details are shown for State.

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

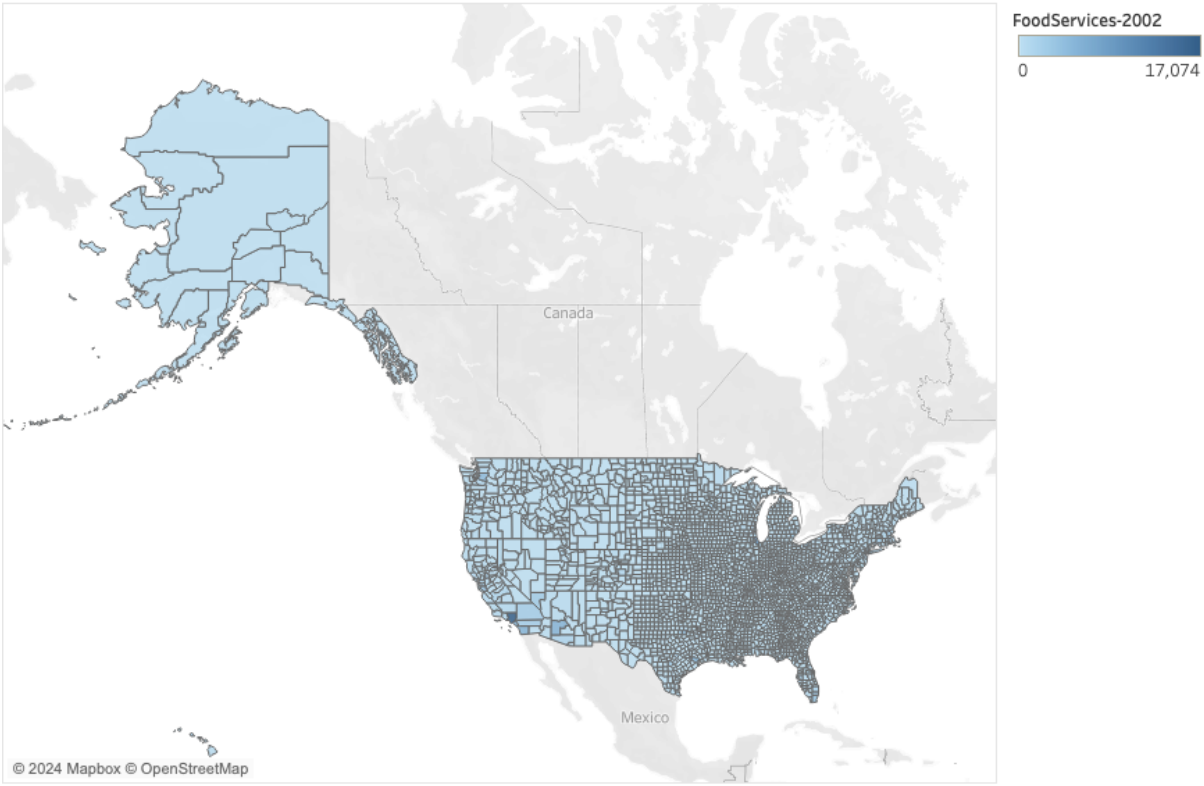
I have a similar Choropleth map but this time, I have added an additional detail which is county. I have chosen a colour gradient that transitions from light to dark. . The light colour indicates a low service while the darkest colour indicates a high service availability. We can see that large cities like Los Angeles, Dallas, and Cook County in Illinois have a darker gradient which indicates higher food service availability compared to other counties. I have added similar maps to indicate food service counts by county in 1997, 2002 and 2007.

Number of Food services by County in 1997



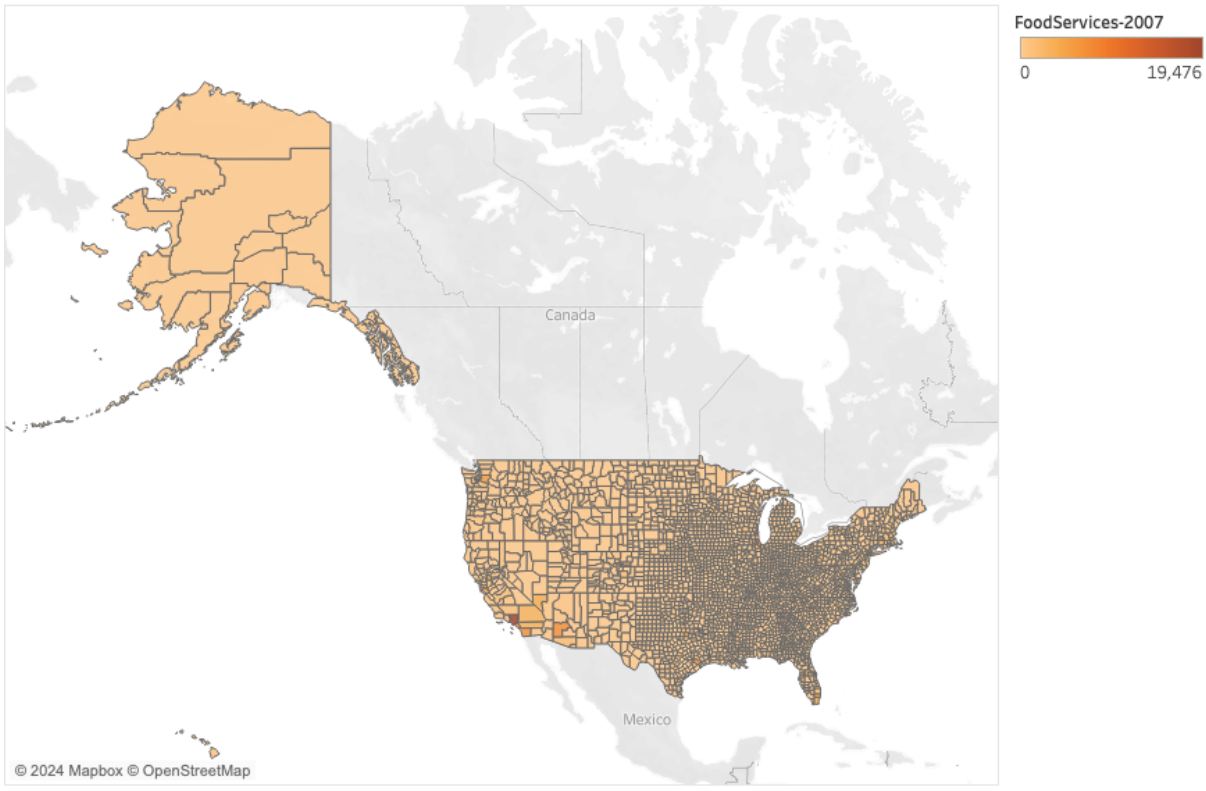
Map based on Longitude (generated) and Latitude (generated). Color shows sum of FoodServices-97. Details are shown for State and County.

Number of Food services by County in 2002



Map based on Longitude (generated) and Latitude (generated). Color shows sum of FoodServices-2002. Details are shown for State and County.

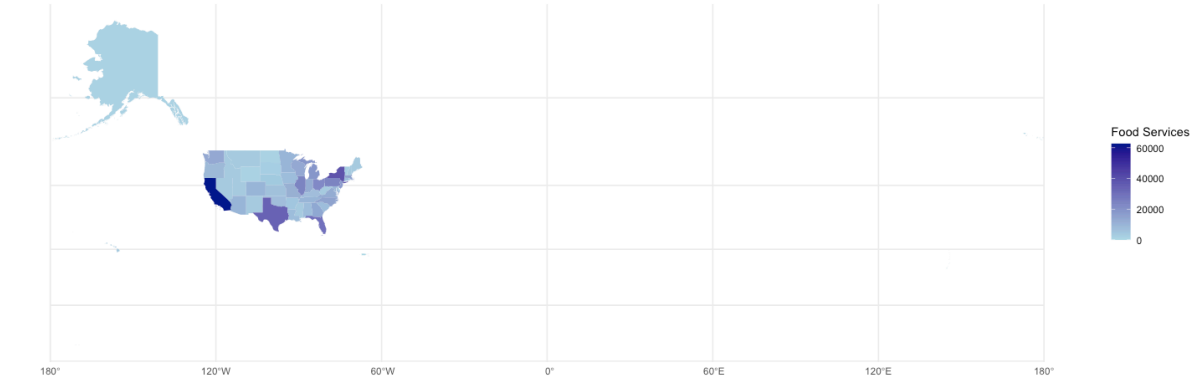
Number of Food services by County in 2007



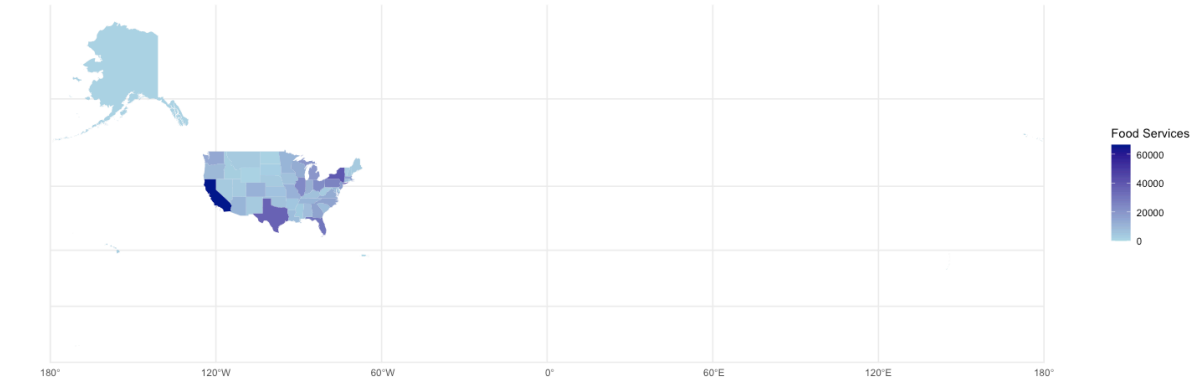
Map based on Longitude (generated) and Latitude (generated). Color shows sum of FoodServices-2007. Details are shown for State and County.

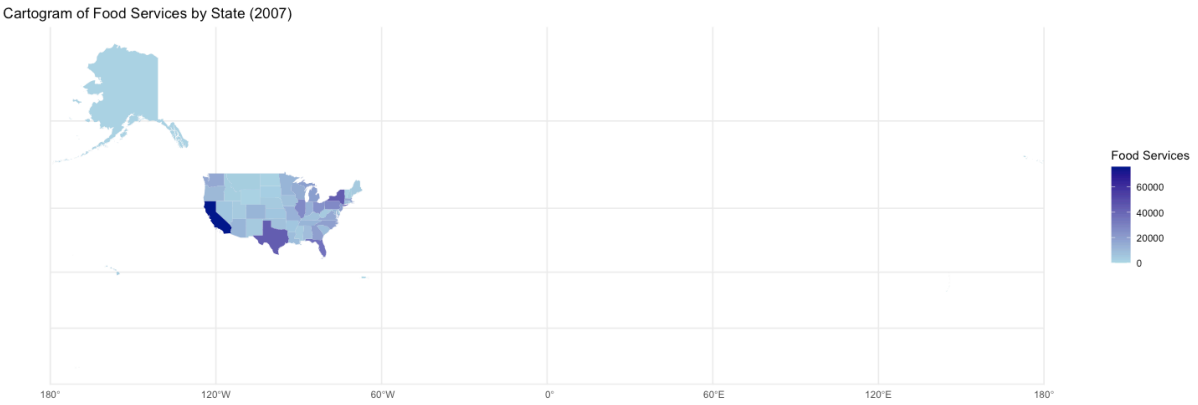
c) (Extra credit) Research how to do a diffusion or tile cartogram in R or D3 and create a cartogram of the state data from this dataset.

Cartogram of Food Services by State (1997)



Cartogram of Food Services by State (2002)





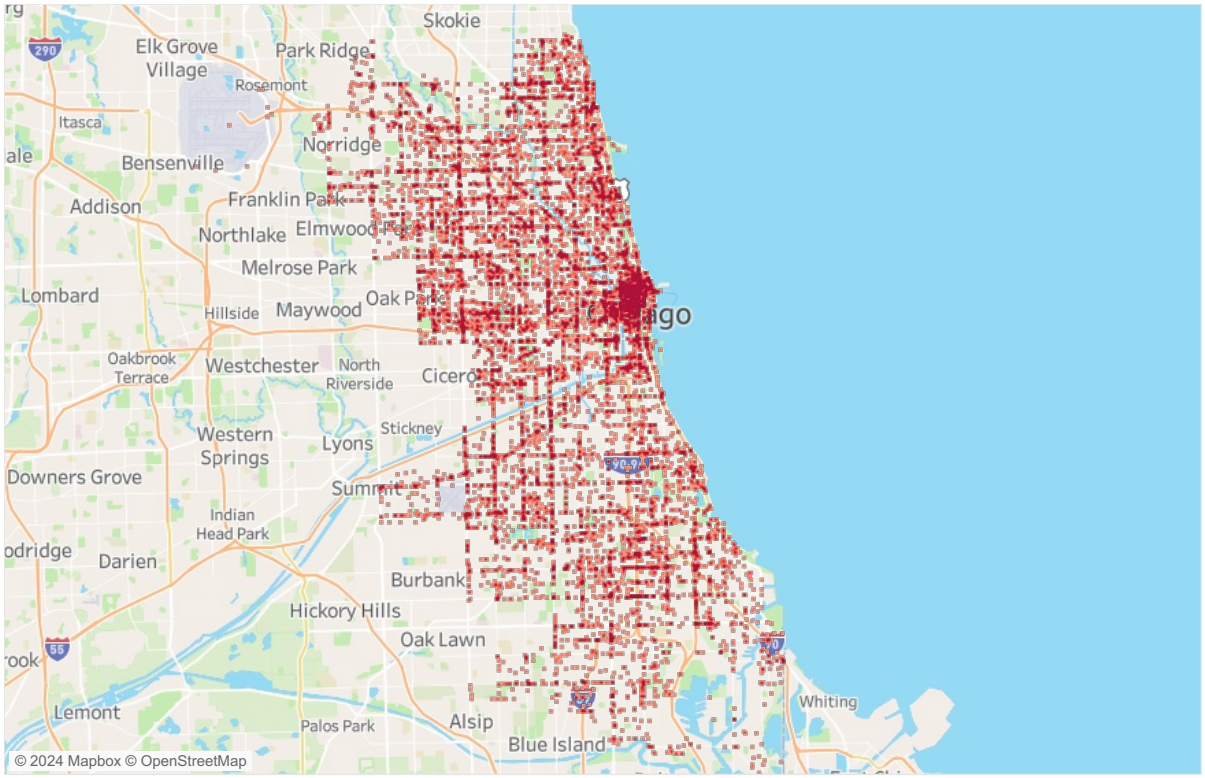
Question 2

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.

I created a density plot- map in Tableau. I chose a street-style map that indicates highways and roads along with some labels. I chose the colour red to indicate the points of crash occurrence. The density of points seems to be more intense at the centre of the city, indicating a number of crashes in that region. This can be useful information to regulate traffic and enforce stricter laws and slower speed limits.

Density plots of Accidents in Chicago

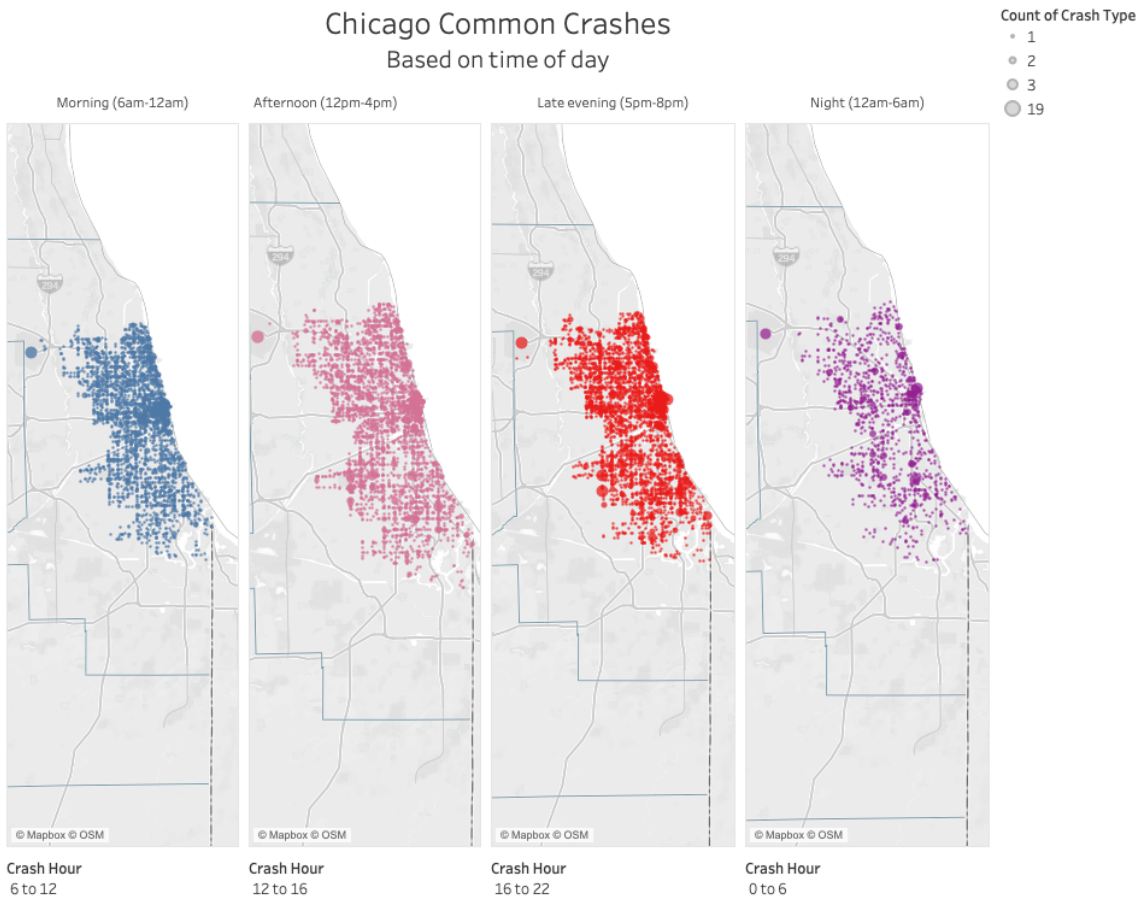


Map based on Longitude and Latitude.

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 created a time-based crash density map in Tableau, splitting the data into four time periods: Morning (6am-12pm), Afternoon (12pm-4pm), Late evening (5pm-10pm), and Night (12am-6am). The map uses different colours to represent crashes during these times—blue for morning, red for afternoon, pink for late evening, and purple for night. The density of crash points is highest in the central areas of Chicago, with the highest concentration during the late evening and night. This visualization helps understand crash trends by time of day, offering insights for targeted traffic safety measures.

The size of each point on the map represents the count of crashes at that location, with larger points indicating higher numbers of accidents. This helps visually highlight areas where crashes are more frequent. Interestingly, crashes are most concentrated during the late evening (5pm-10pm), as seen by the dense cluster of larger red points, particularly in the city's central areas. In contrast, the night period (12am-6am) has fewer crashes, reflected by the smaller and sparser purple points. This suggests that the late evening, possibly due to higher traffic volumes and rush hour, poses a greater risk for accidents compared to the quieter night hours. These patterns provide valuable insights for city planners and traffic authorities to focus their safety initiatives, especially during high-risk evening hours.

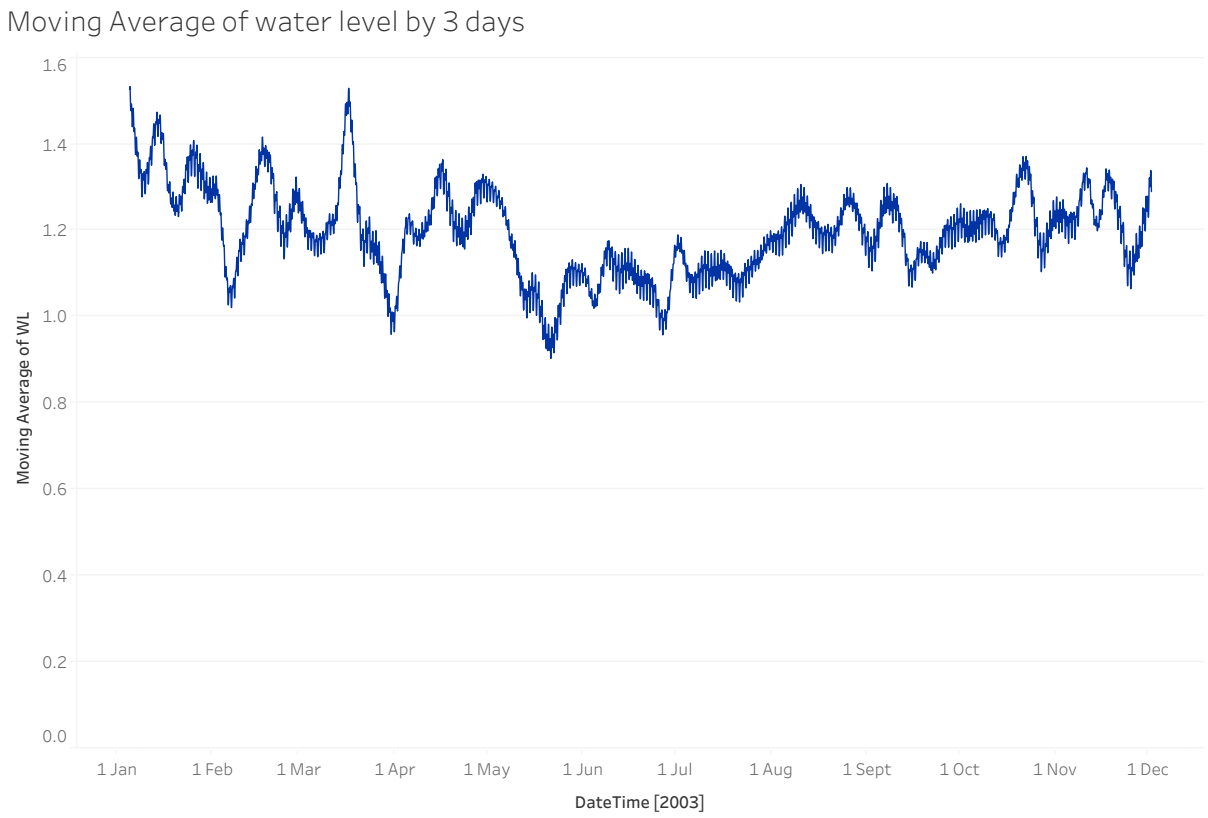


Question 3

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.

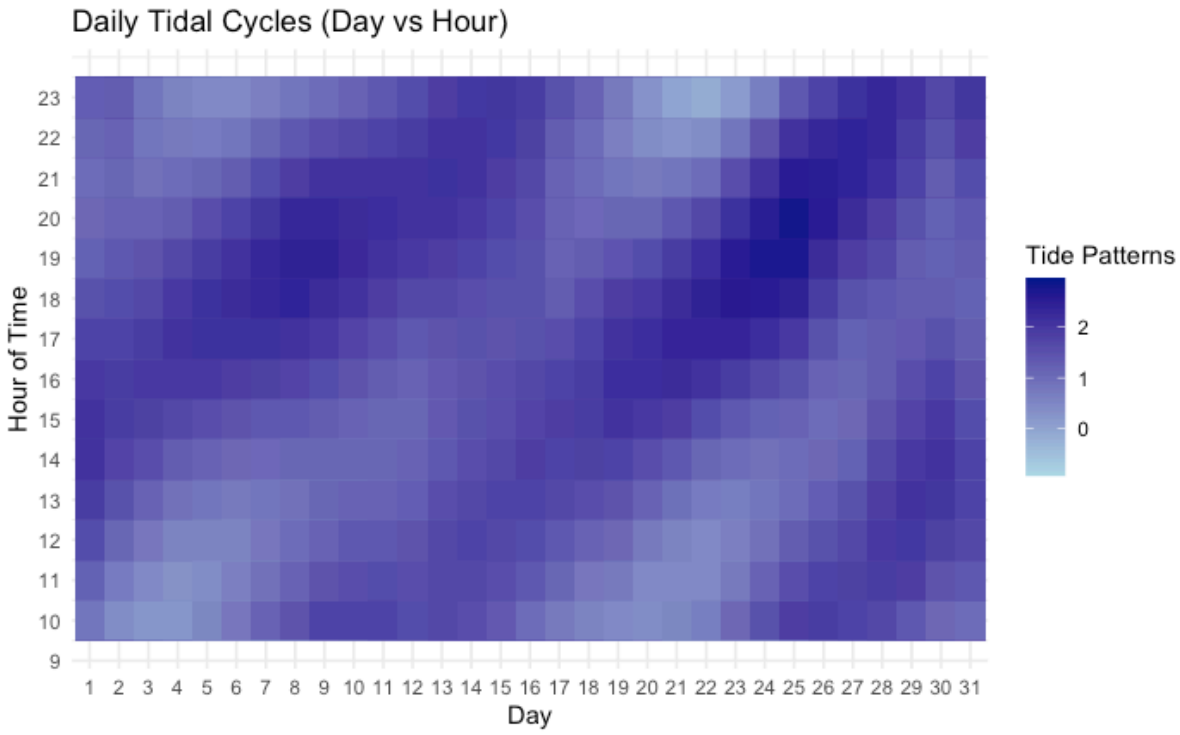
Tableau:



The trend of Moving Average of Avg. WL for DateTime.

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.

In R Studio:



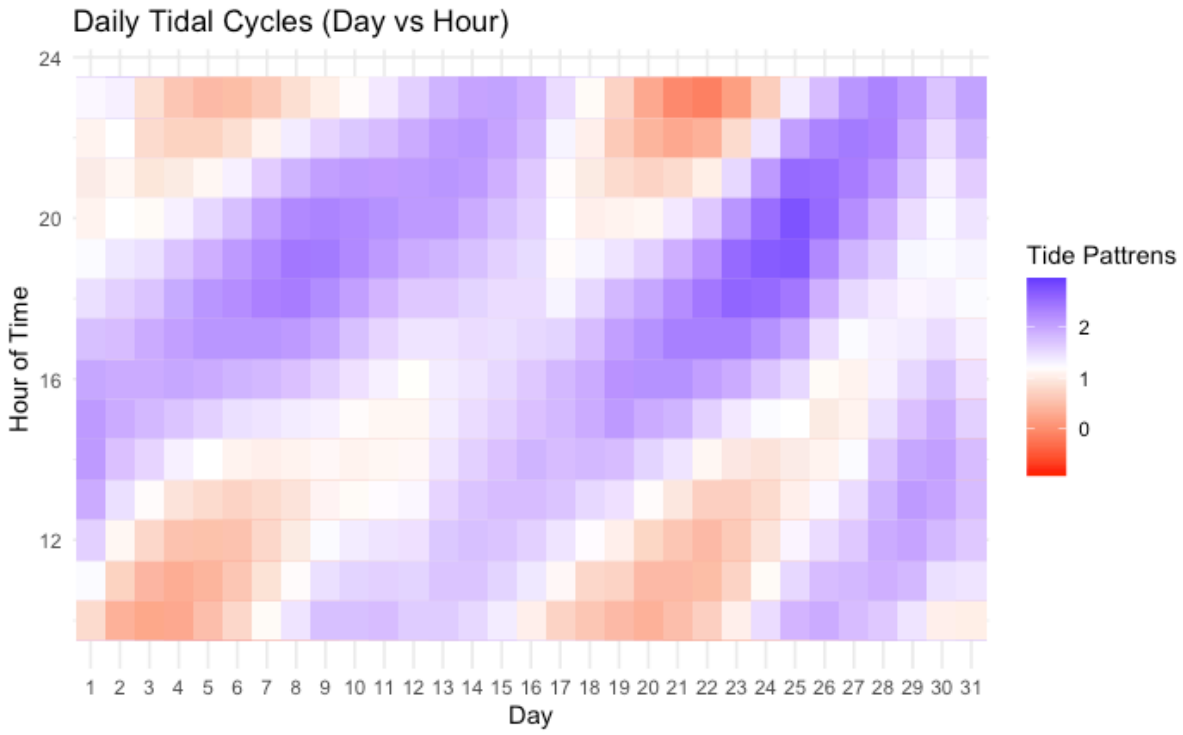
c) Then write a single paragraph outlining the differences between the information that each graph communicates.

Graph 1 focuses on the long-term water level trend over a specific period (January to October 2003). It uses a 3-day moving average to smooth out short-term fluctuations and reveal the overall pattern. The graph shows a relatively stable water level with minor variations. Graph 2 focuses on the daily and hourly variations of tidal levels. It uses a heatmap to visualize the average tidal values for each day and hour of the month. This graph reveals the patterns of tidal activity, including daily fluctuations and potential monthly variations. In summary, graph 1 provides a general overview of the water level trend, while graph 2 delves into the detailed patterns of tidal activity.

Question 4

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.**

In R Studio:


Daily Tidal Cycles (Day vs Hour)
Hour of Time
24
20
16
12
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Day
Tide Pattrens
2
1
0