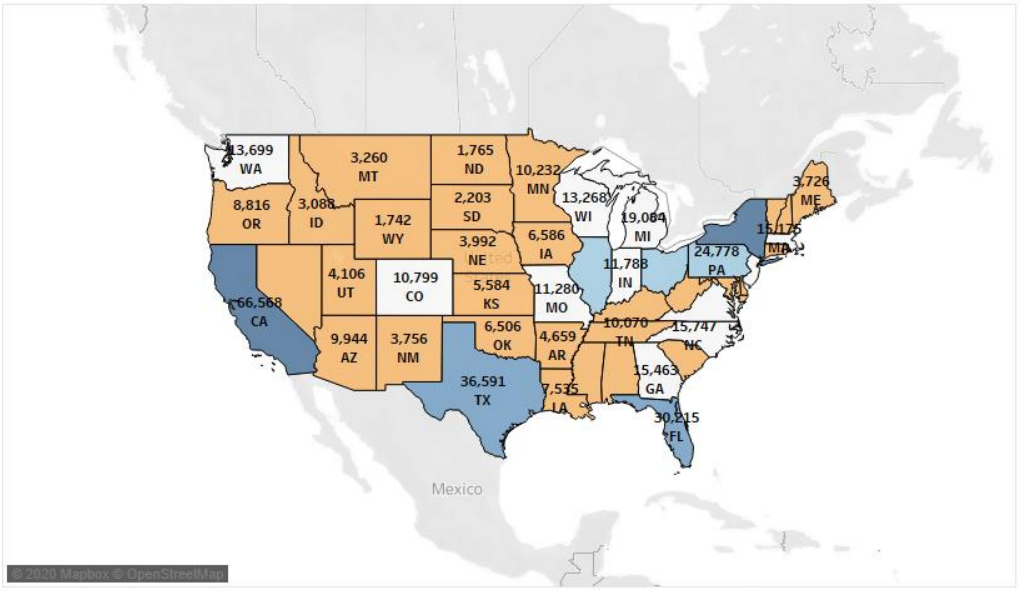
1) Download the FoodSrvcByCounty.txt file and create the following visualizations for this geographical data. The data is for the availability of food services by county in the U.S. It also has data by state (in the county field, some of them have the state names, and those rows hold the state totals,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

- **Ans:** I have used Tableau for this question first I have downloaded the file and import in the Tableau then I am doing following.

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.

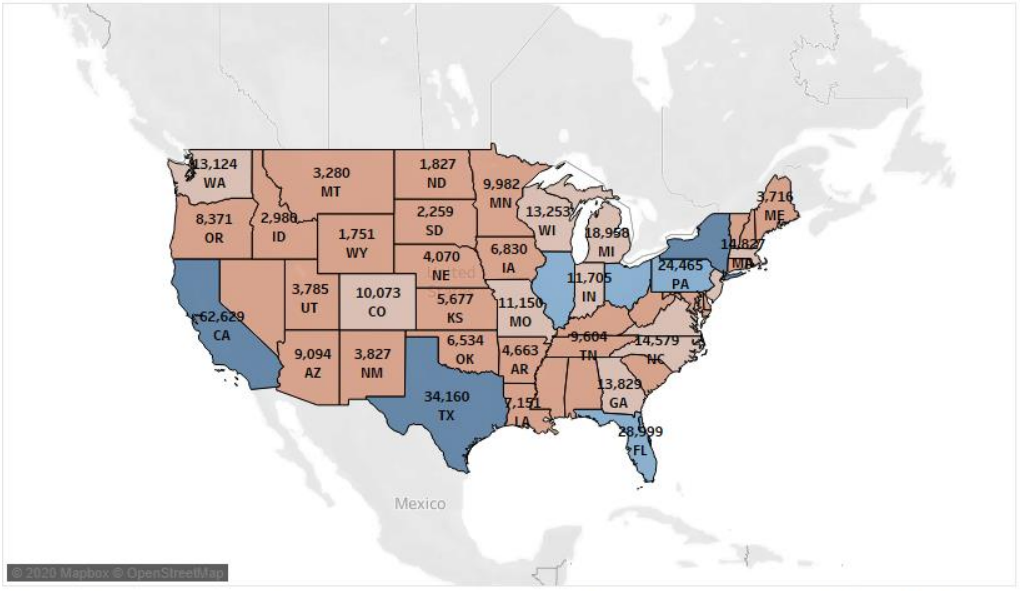
- In this graph I have discarded the null value of the United States of America and also neglect the Hawaii and Arkansas just to plot the US map more aesthetically.
- You can also find more detail of the each graph at the bottom.
- Here in the state Vs 2002 intentionally I have used little light colour than the other two graph just to show different numbers actually most of the state(not all) have less number of food service.
-

State Vs Food Services 2002

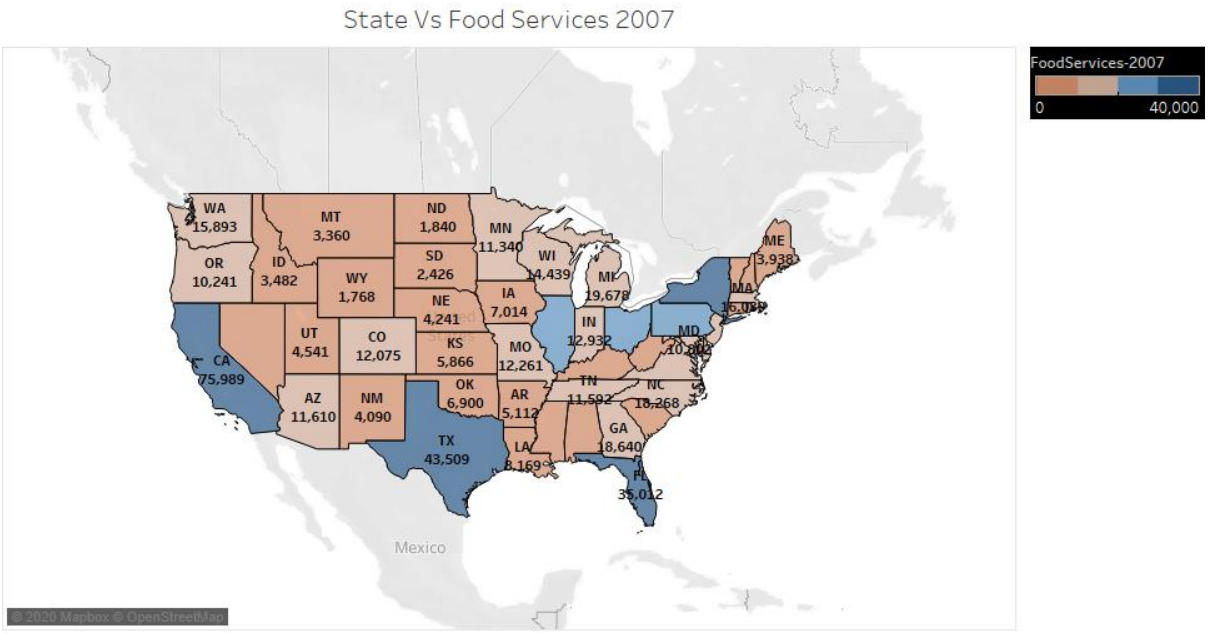


Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-2002. The marks are labelled by sum of FoodServices-2002 and State. Details are shown for State. The view is filtered on State, which excludes Null, AK and HI.

State Vs Food Services 1997



Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-97. The marks are labelled by sum of FoodServices-97 and State. Details are shown for State. The view is filtered on State, which excludes Null, AK and HI.

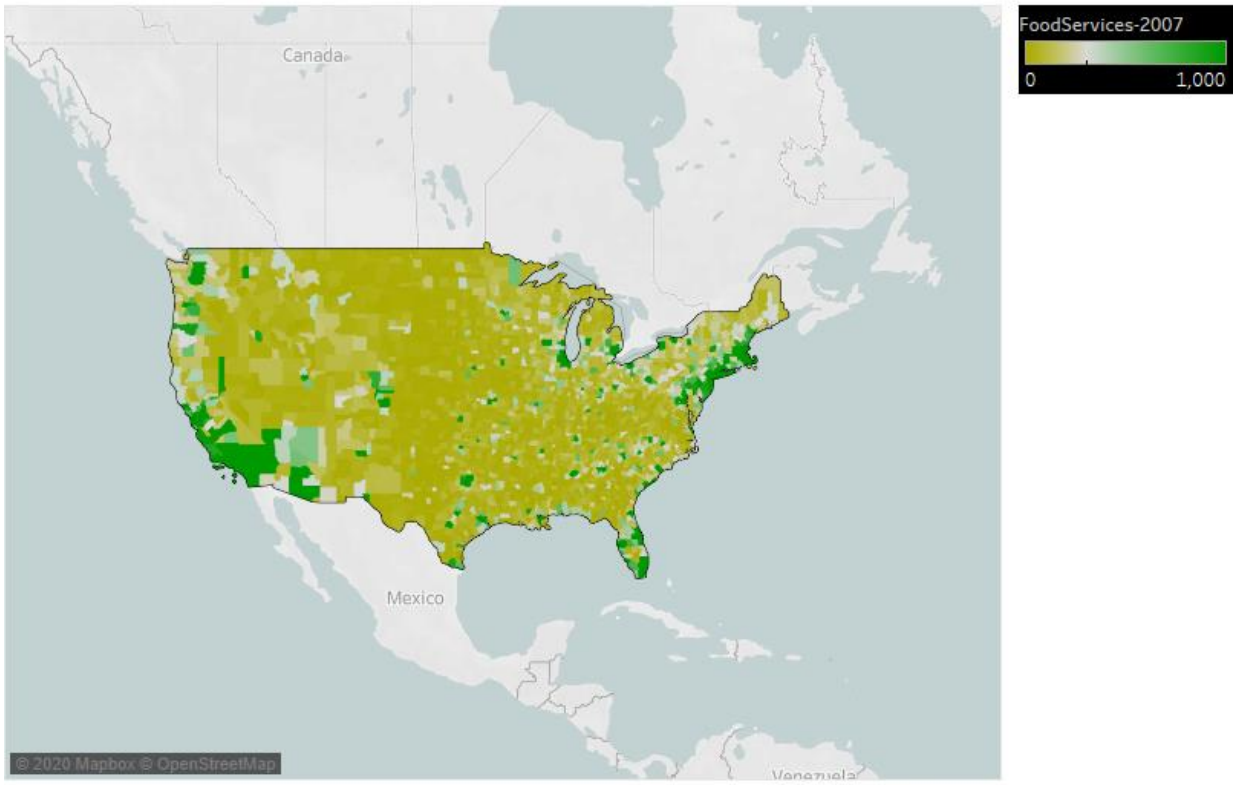


Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-2007. The marks are labelled by State and sum of FoodServices-2007. Details are shown for State. The view is filtered on State, which excludes Null, AK and HI.

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

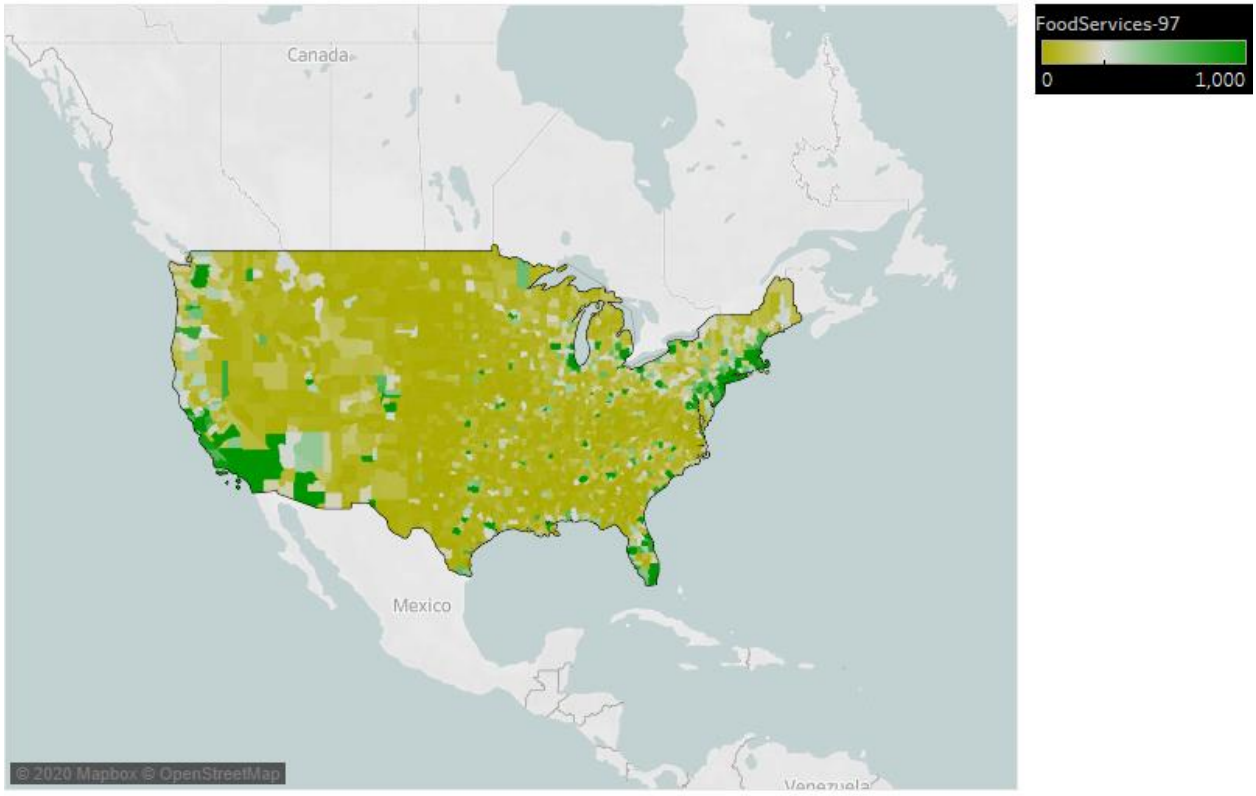
- In this graph I have discarded the null value of the United States of America and also neglect the Hawaii and Arkansas just to plot the US map more aesthetically.
- You can also find more detail of the each graph at the bottom.
- In this graph for county with higher food services I have put the Green colour and used the colour circle to show contrast for opposite low number of food services.

County Vs Food Services 2007

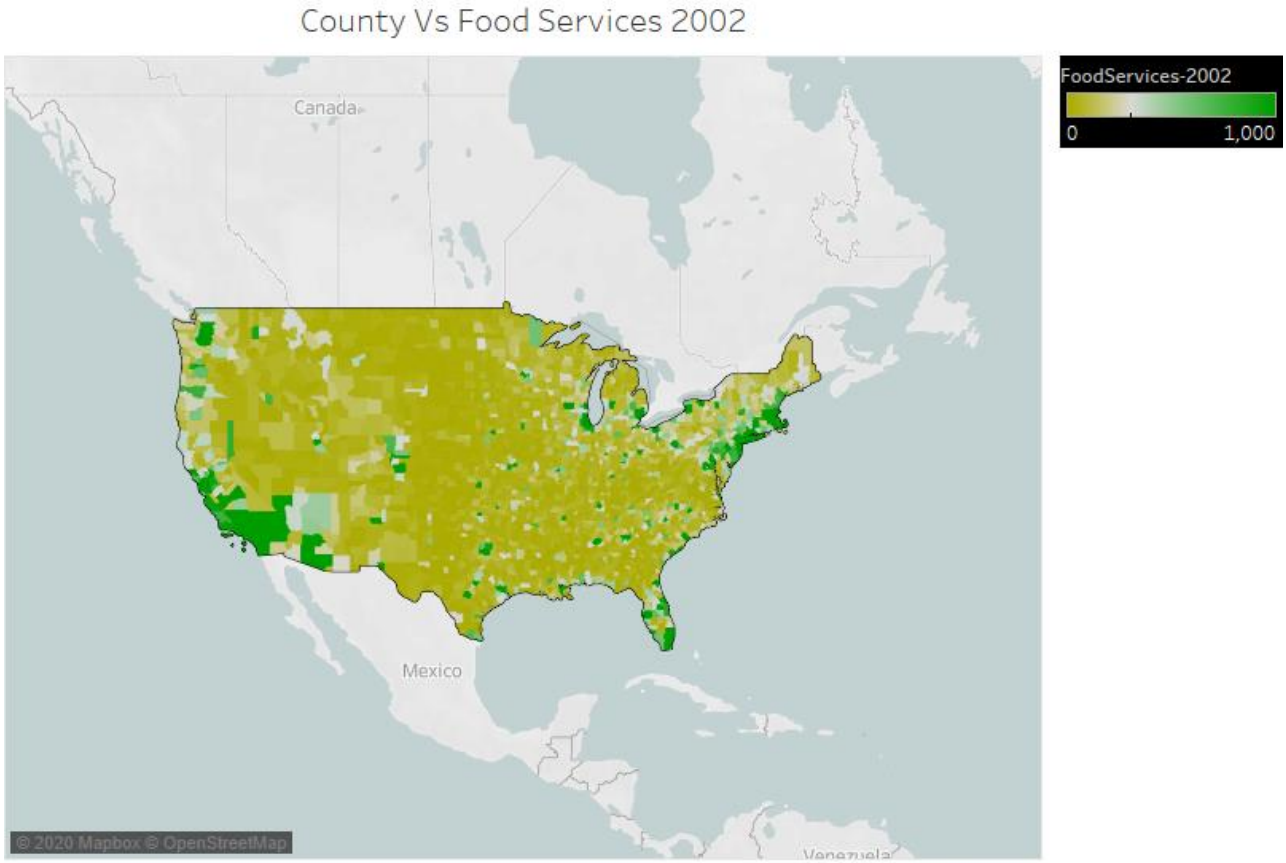


Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-2007. Details are shown for State and County. The view is filtered on State, which excludes Null, AK and HI.

County Vs Food Services 1997



Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-97. Details are shown for State and County. The view is filtered on State, which excludes Null, AK and HI.

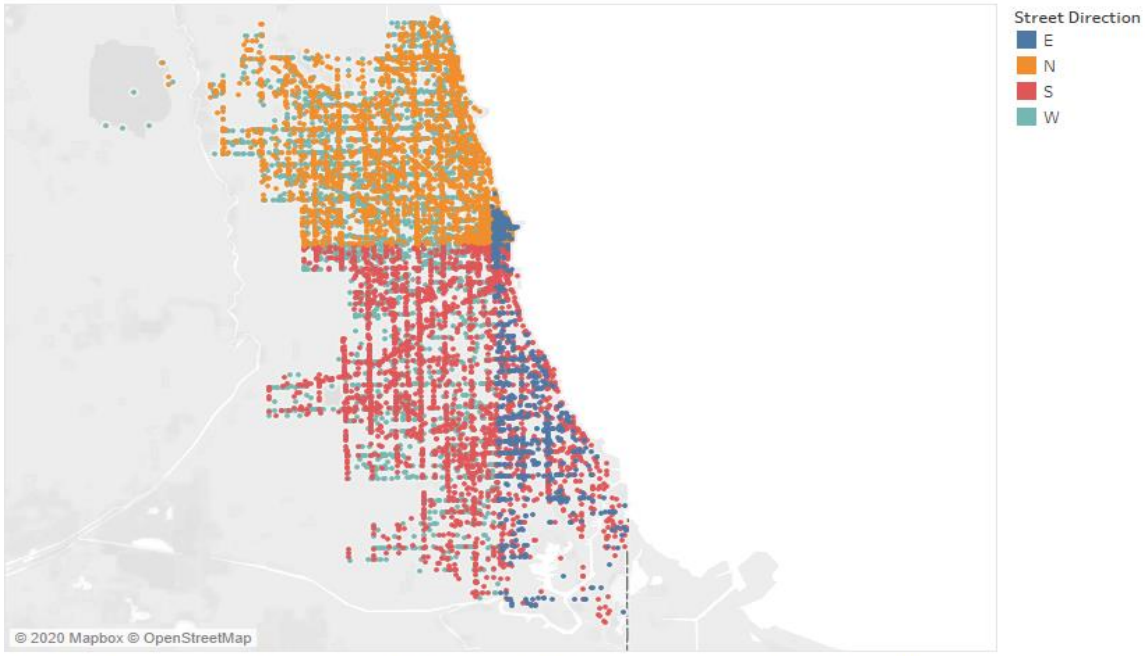


Map based on Longitude (generated) and Latitude (generated). Colour shows sum of FoodServices-2002. Details are shown for State and County. The view is filtered on State, which excludes Null, AK and HI.

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).

- From the graph you can see that are more crashes are in the north but if you see closely you can see in the west there are more crashes as geographical west is spread from south to north which covers the most of the area of the Chicago.
- Where In the East very less amount of the crashes you can found.

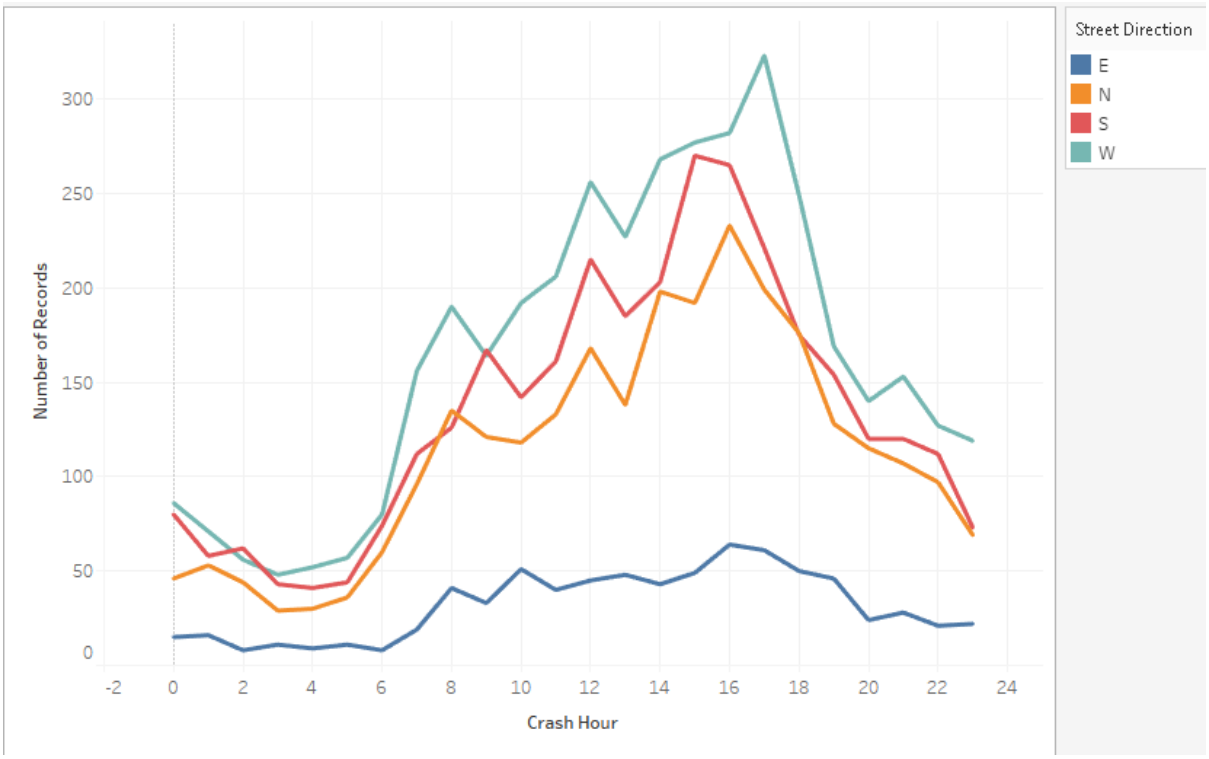
Geographical Chicago Crashes Graph:



Map based on Longitude and Latitude. Colour shows details about Street Direction. Details are shown for Street Name.

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.

- Here I have used line graph to show common crashes in the Chicago you can use bar graph as well but it looks more vibrant as you can clearly see that which part have more crashes on particular tume.
- From the graph, you can see that from 14-18 hours there are higher numbers of the crashes in all the area of the graph. So you can say that this is the peak hour for the accident to happen. Where in the west it is more than 300 crashes.
- Where in the east there are very less crash cases from that we can say that people are more mindful and drive consciously.



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.

Ans: For this question I have used R I am attaching the code and the output of the Code as below:

- After Pre-processing of water level measurement every hour I got the below graphs,

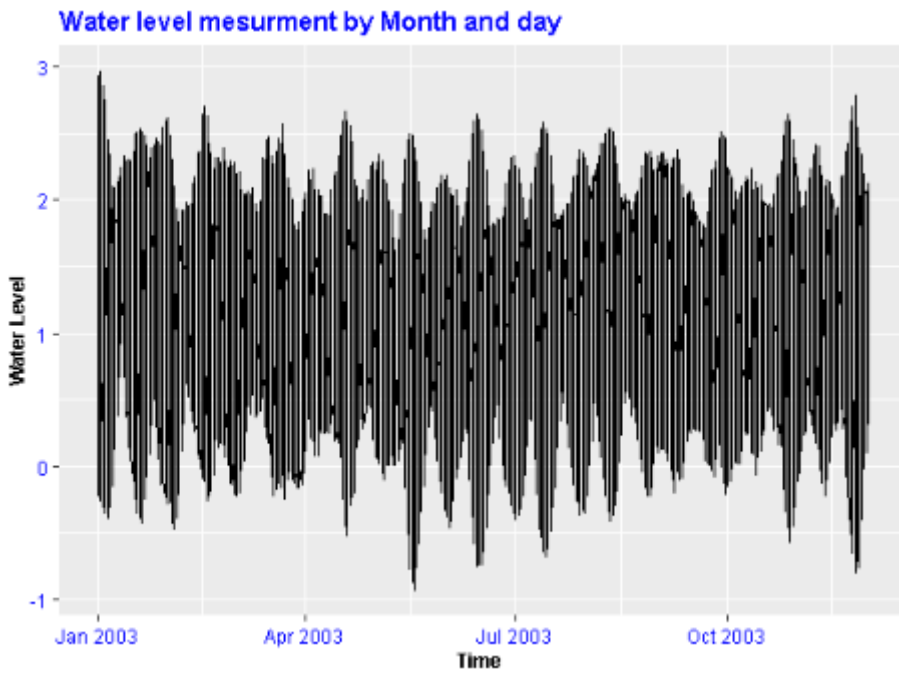
R – Code:

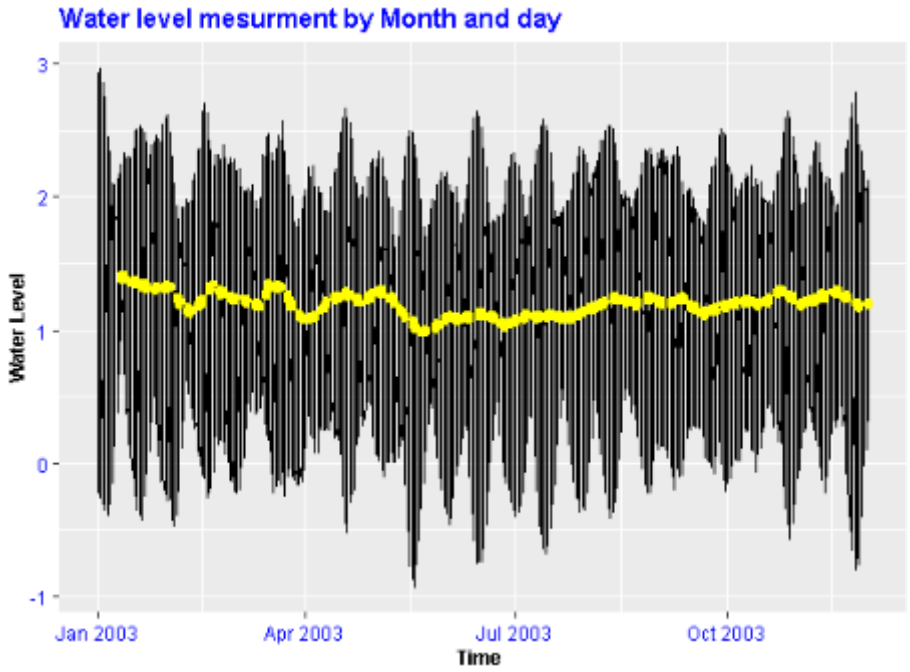
```
date = mdy(portland$Date)
plot <- ggplot(portland, aes(x=date, y=WL)) + geom_line() +
xlab("Date ")+
 ylab("Water level") +
 ggtitle("Water level Mesurment by dates and months")+
 theme(
   axis.title.x = element_text(color = "Blue", size = 8, face="bold"),
   axis.text.x = element_text(color = "Blue"),
   axis.title.y = element_text(color = "Black", size = 8, face="bold"),
```

```
  axis.text.y = element_text(color = "Blue"),
  plot.title = element_text(color = "Blue", size = 12, face="bold")
 )

plot_ma <- plot + geom_ma(ma_fun=SMA, n=360, size=1, color="red")+
xlab("Date ")+
 ylab("Water level") +
 ggtitle("Water level Mesurment by dates and months")+
 theme(
  axis.title.x = element_text(color = "Blue", size = 8, face="bold"),
  axis.text.x = element_text(color = "Blue"),
  axis.title.y = element_text(color = "Black", size = 8, face="bold"),
  axis.text.y = element_text(color = "Blue"),
  plot.title = element_text(color = "Blue", size = 12, face="bold")
 )

plot_ma
```

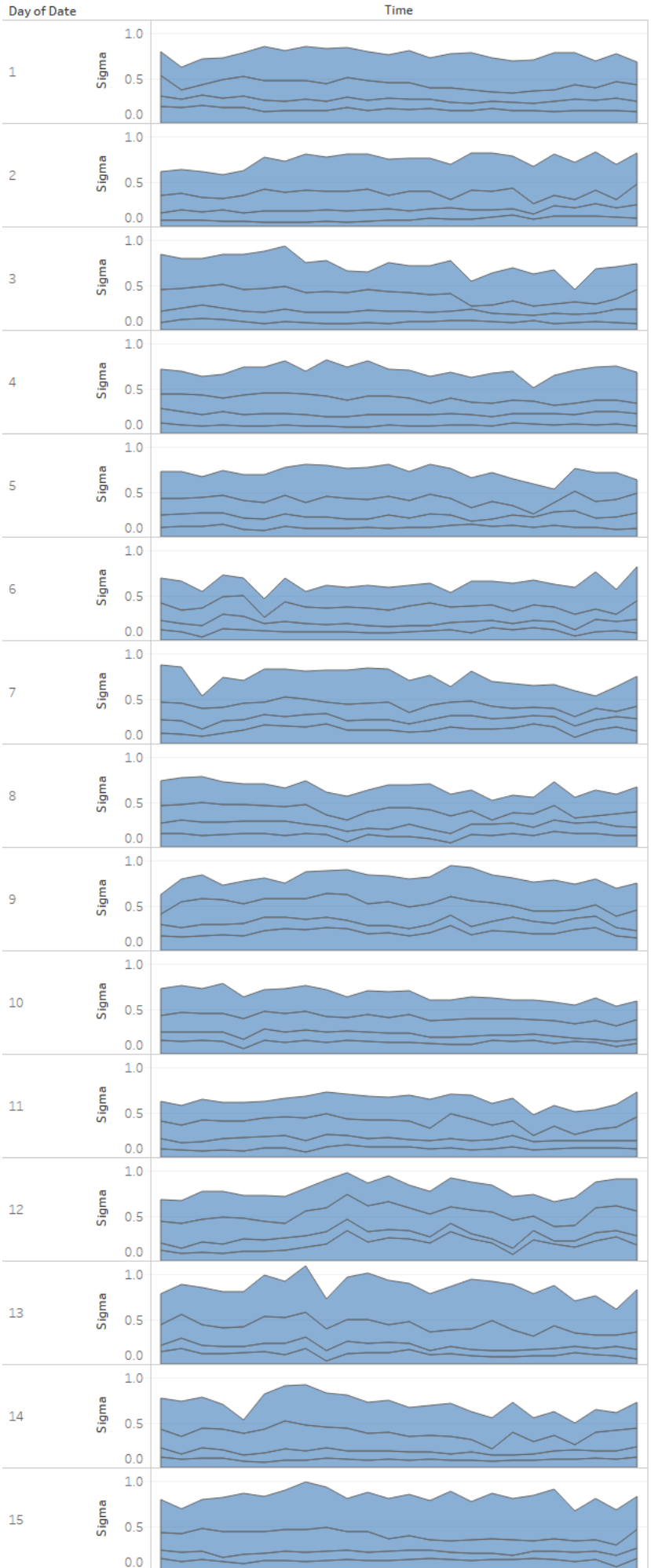




B.) Graph the cycles that happen each day (because of tides). You might try overlapping many days' data as separate overlapping time series, using a level plot, a horizon graph, etc. The point of this exercise is to try to come up with a way of showing the progression of the tides over some period of time that is rich and detailed and which shows the pattern, but which is still readable and which doesn't clutter the graph.

Ans: I have use number of days and time of the day as x and y axis and as per that you can see that tides high or low in four quarters. You can see in the graph that for particular day and particular time tide's level in the different quarters of year

Progression of tides over hours (Water Color)



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

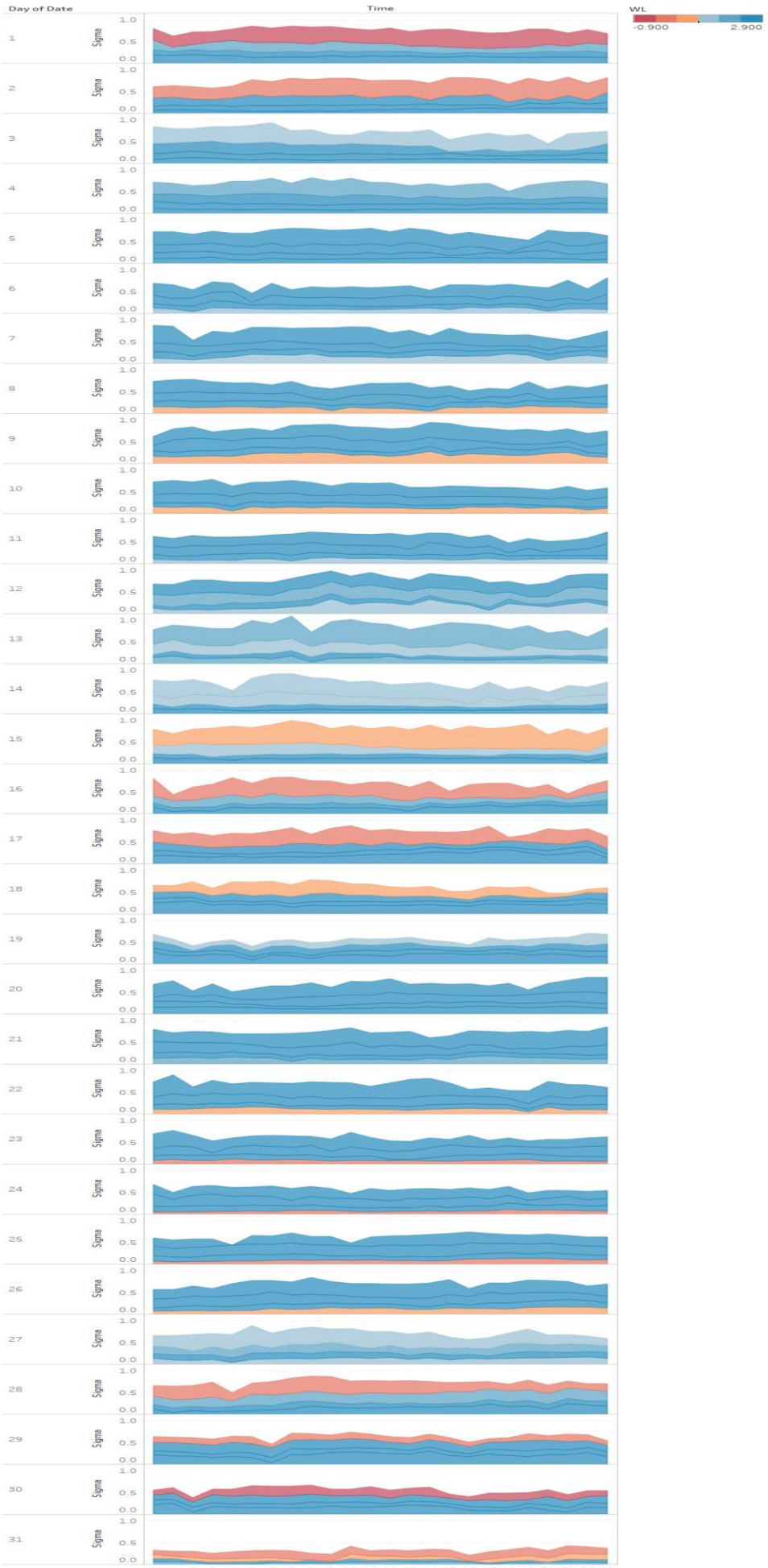
Ans: This data contains a year of data with water level (WL) measurements every hour as a function of Time. I have used the simple moving average which is very popular in the stock's closing count and then approached window method which is one of the popular and useful methods of the Image processing and data visualisation. I have use area graph for showing tides measurement as you can see and clearly distinguish the difference between the tides of each day and easily see that which day have high tides and low tides. The first graph is more subtle where the second graph is easy to interpret and you can notice the difference level of water by day wise. So last graph shows the progression of the tides over period of time that is rich and with detailed. From that you can also observe the pattern. In addition, it still readable and most common problem of cluttering of graph is not happening here.

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.

Ans:

I have use number of days and time of the day as x and y axis and as per that you can see that tides high or low in four quarters. You can see in the graph that for particular day and particular time tide's level in the different quarters of year, where if tide is low that I have showed with red color and blue for opposite.

Progression of tides over hours



Sum of Sigma for each Time Hour broken down by Date Day. Colour shows sum of WL. Details are shown for Date Quarter.