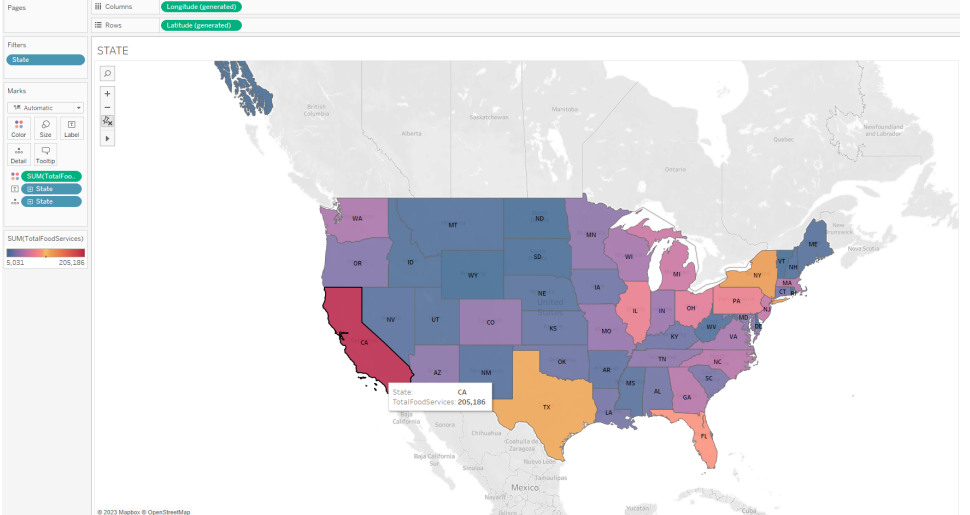
Data Visualisation Assignment

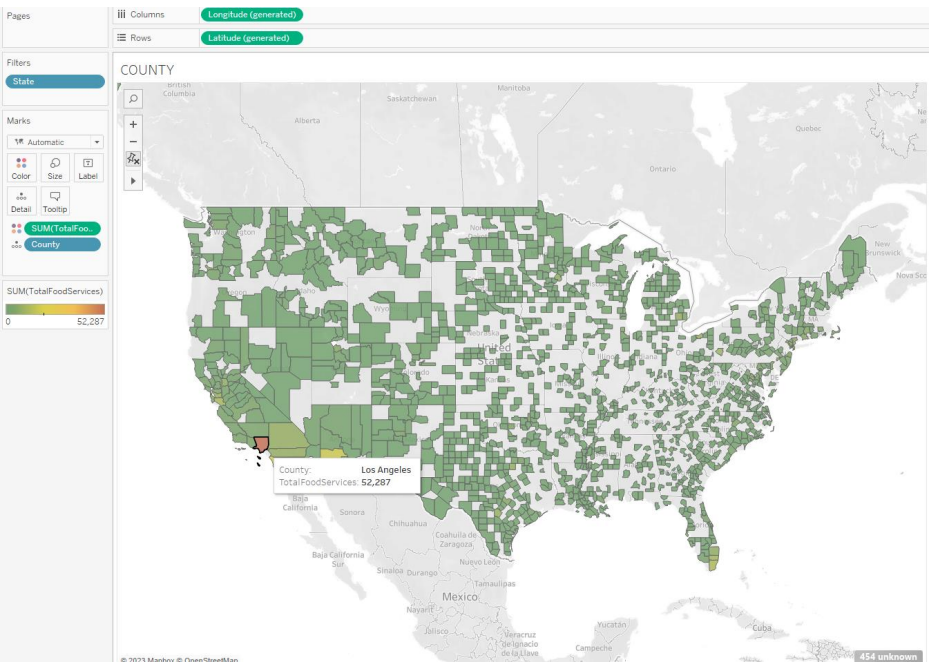
Adithya Harsha

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.



EXPLANATION: With a divergent color scheme pattern above, we can clearly see the states with maximum and minimum food services. CA and TX are clearly depicted on the top, and the darker the shade of blue lower the food services.

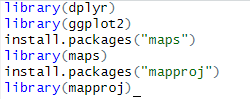
b. Graph food services by county with the same type of visualization. Again, think carefully about the color scheme



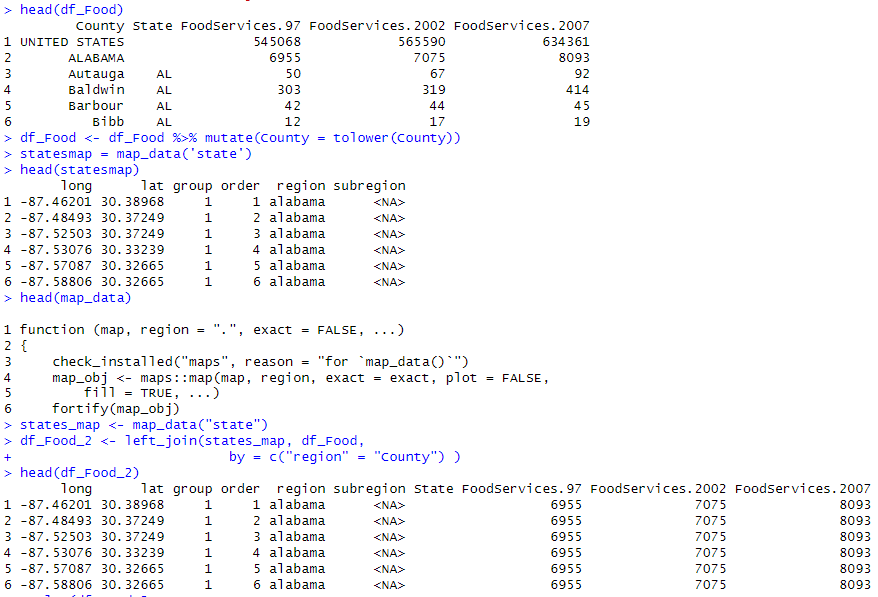
EXPLANATION:

As we can see the individual donations by each county range mostly in the lower scale, since there are a greater number of services at the lower scale, the maximum services are clearly highlighted and the maximum services goes to – Los Angeles

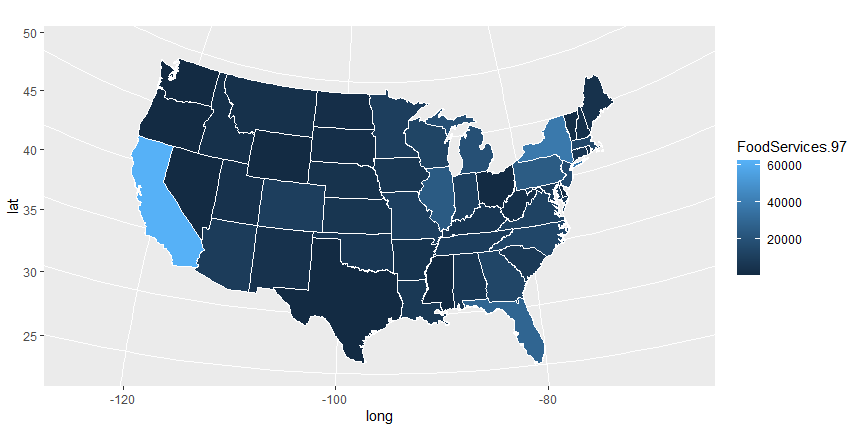
c. (Extra credit) Research how to do a diffusion or tile cartogram in R or D3 and create a cartogram of the state data from this dataset



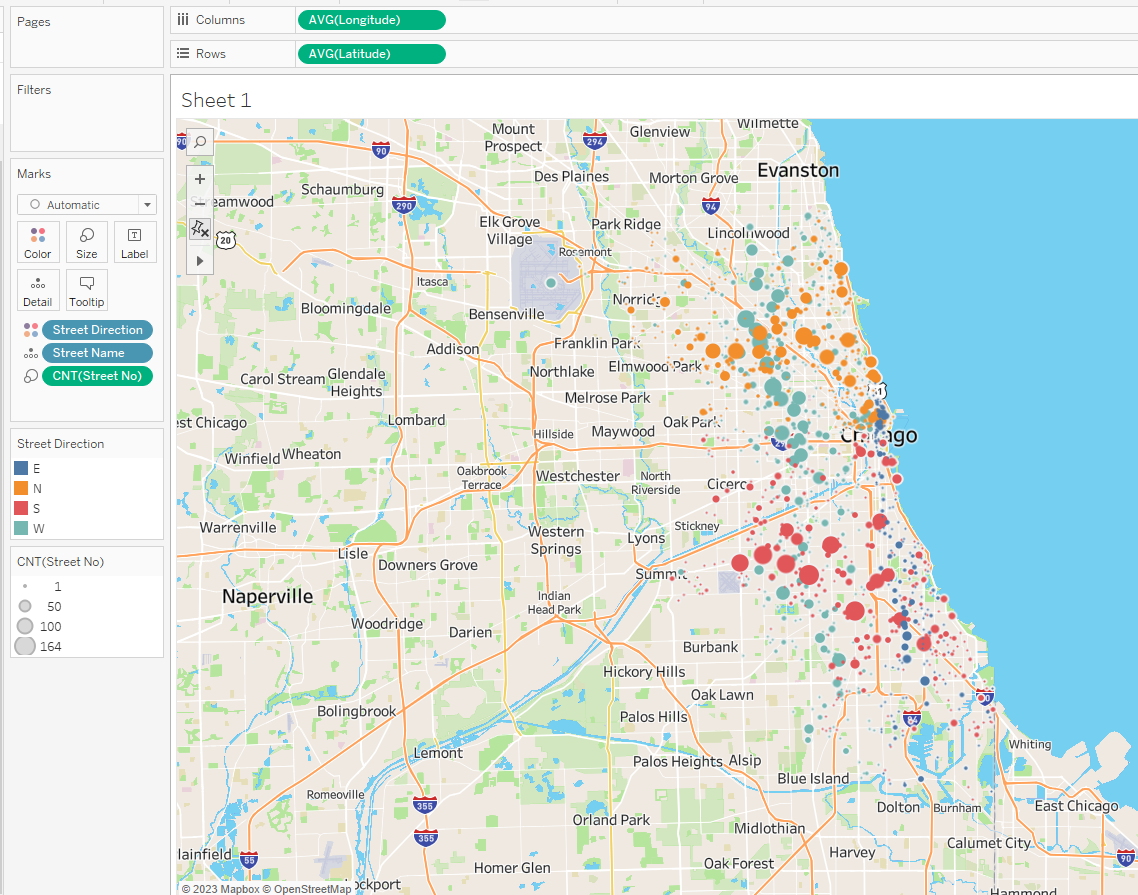








2) The Chicago\_crashes.csv file contains information on every crash recorded in Chicago in June 2019 (see Chicago’s portal at https://data.cityofchicago.org/Transportation/Traffic-CrashesCrashes/85ca-t3if for the latest data. I chose a random month because the data get dense quickly). a. Create an appropriate type of geographic plot to show where all the accidents in this data occur.

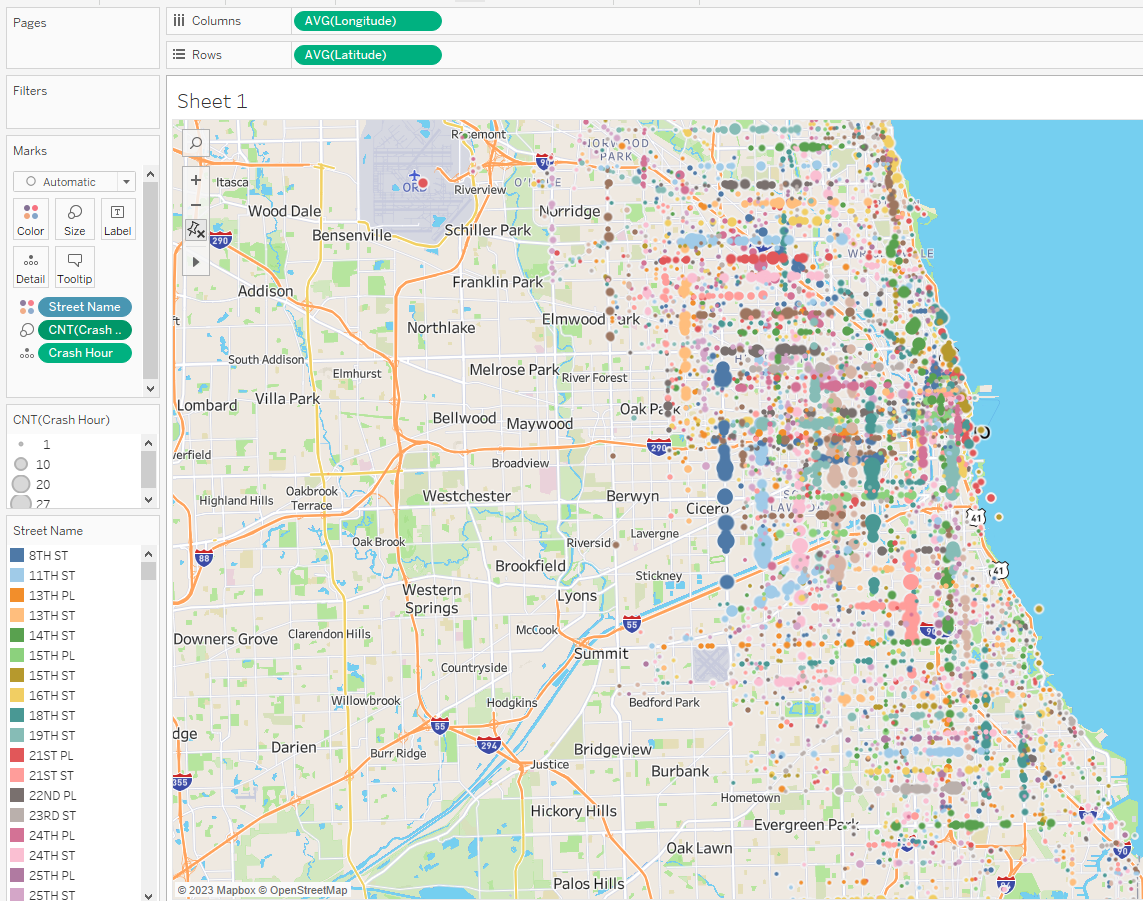


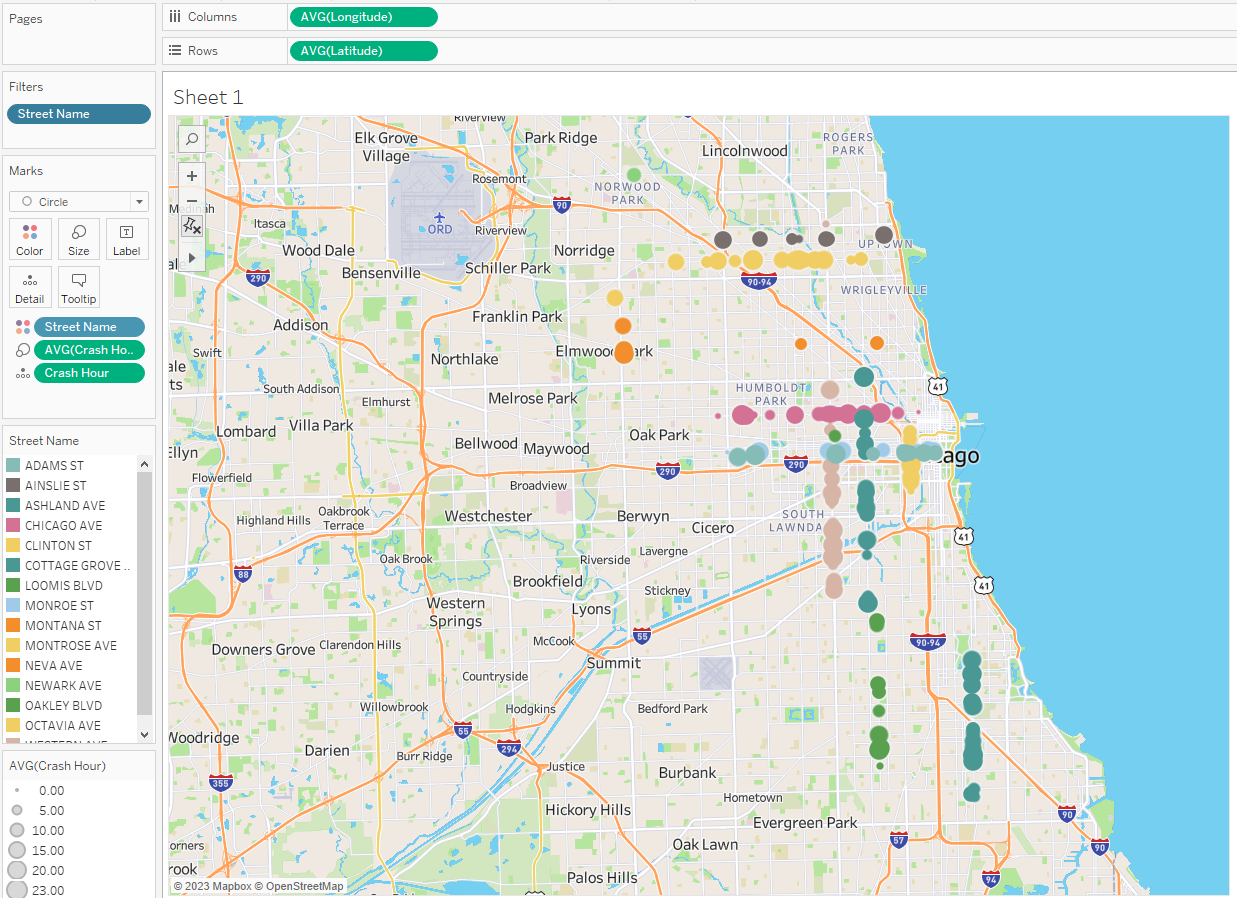
EXPLANATION:

• I have chosen to use a color scheme based on the direction E, N, S and W, to clearly see a pattern and identify the crash prone zones.

• The size of each point consists of number of crashes at that street

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.



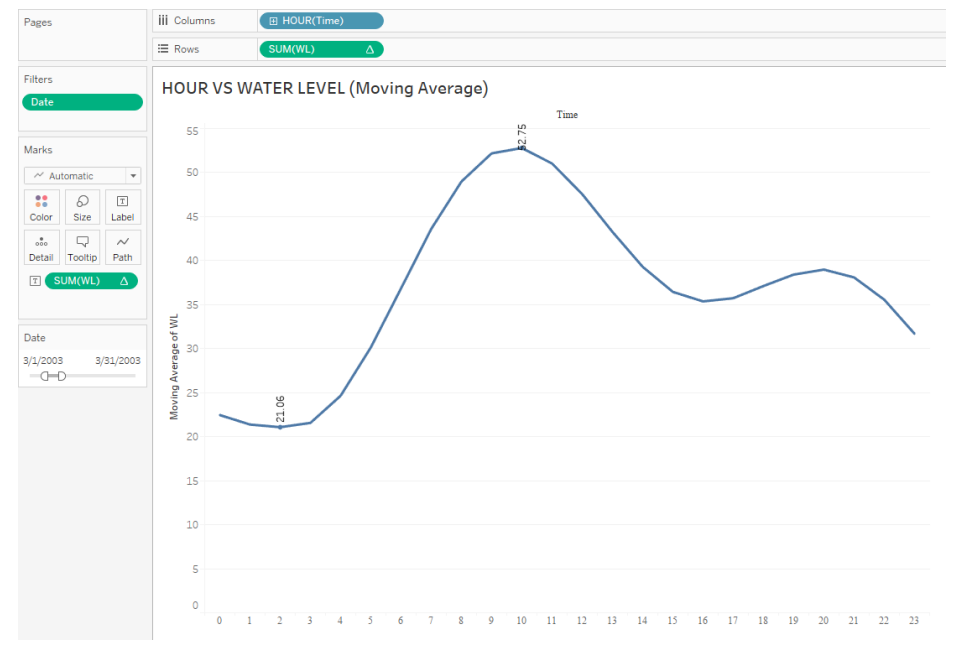


EXPLANATION: Fig 1: We can see the number of crashes that have occurred at a particular crash hour in a particular street. The size of the point depicts the number of crashes, each point depicts the crash hour and the street as well. The color scheme is based on the street.

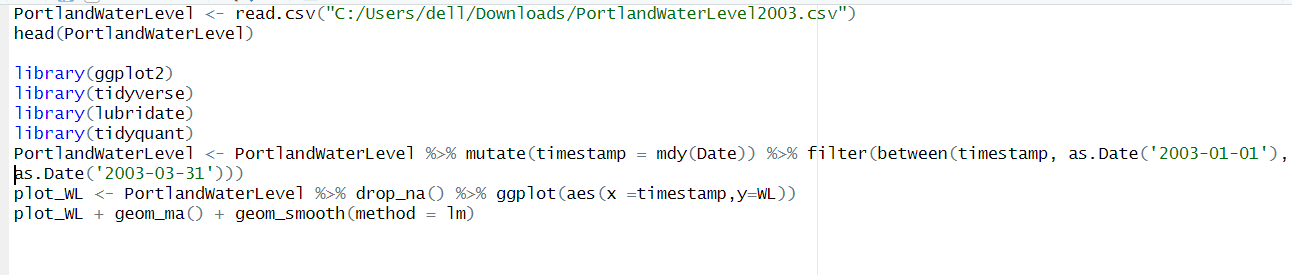
Fig 2. Is similar to figure 1 but has been filtered to show patterns, that need to be focused on to avoid crashes.

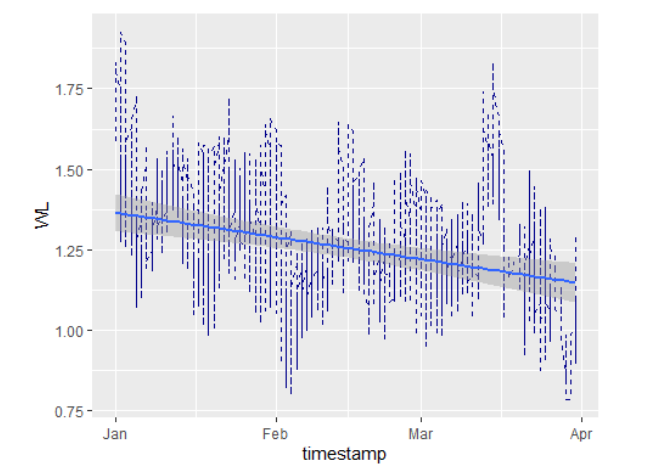
3) (20 pts) Download the Portland Water Level dataset and explore it by creating the following visualizations of the time series from the techniques described in lecture. Use both R and Tableau for at least one question part. They should, of course, adhere to the design criteria that we’ve learned, and 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.



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.





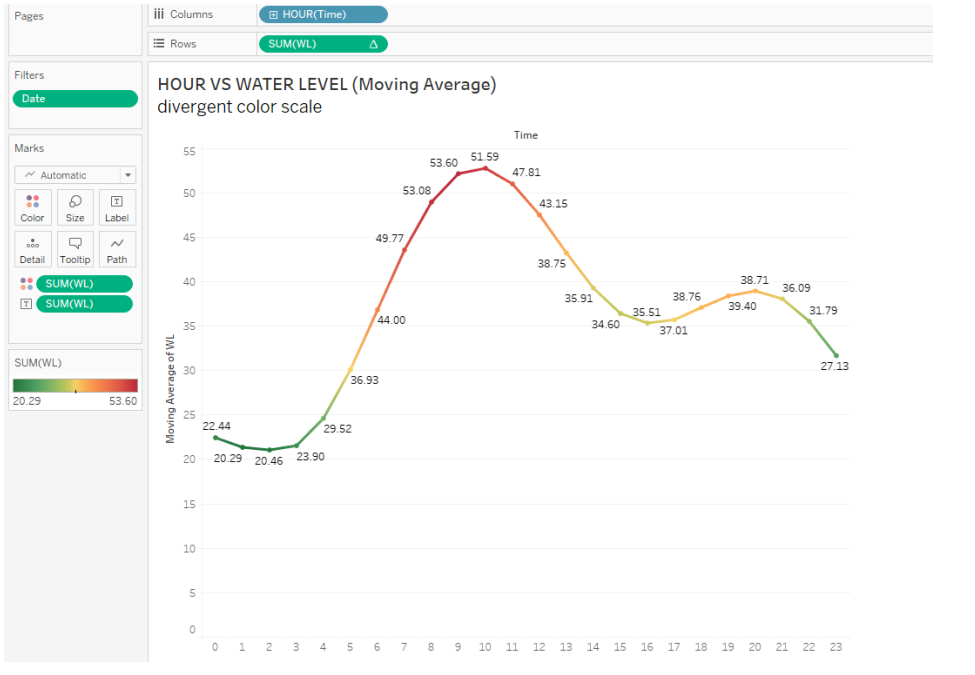
c. Then write a single paragraph outlining the differences between the information that each graph communicates.

EXPLANATION:

Fig 1: The overall trend of water level throughout the day has been averaged for 3 month (Jan – March) as “Moving Average” to notice the pattern of rise and fall in water level at different hour of the day. (the min and max have been marked to notice the peak)

Fig 2: This graph is a time series for the months Jan, Feb and Mar. The time series clearly shows two peaks at every month (occurs roughly 14 days apart) depicting - The tides build up to a maximum and fall to a minimum twice a month. One caused when the moon is closest to earth and the other when the moon is farthest. The smooth line shows a downward trend in the overall water level.

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.



EXPLANATION: The divergent color scheme has been used The peaks are red, and the lower peaks are green.

The labels consist of the moving average at each point.

I have chosen this color scheme to differentiate between the high tides and low tides easily