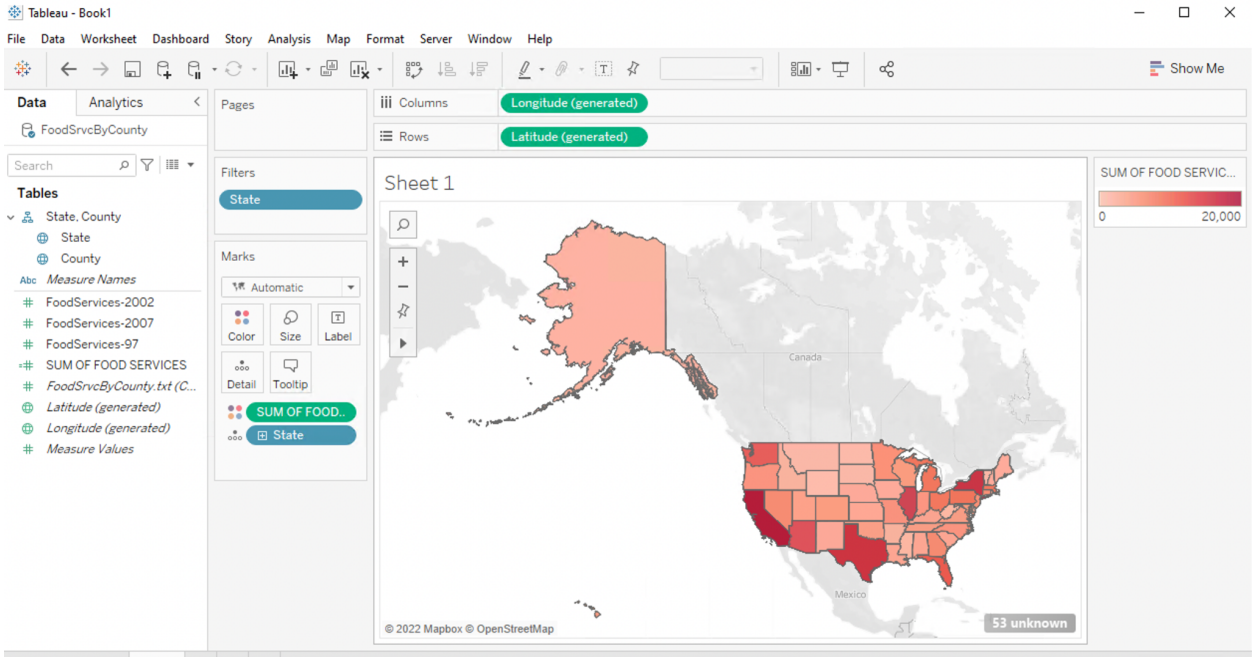
1) Download the FoodSrvcByCounty.txt file and create the following visualizations for this geographical data. The data is for the availability of food services by county in the U.S. It also has data by state (in the county field, some of them have the state names, and those rows hold the state totals,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

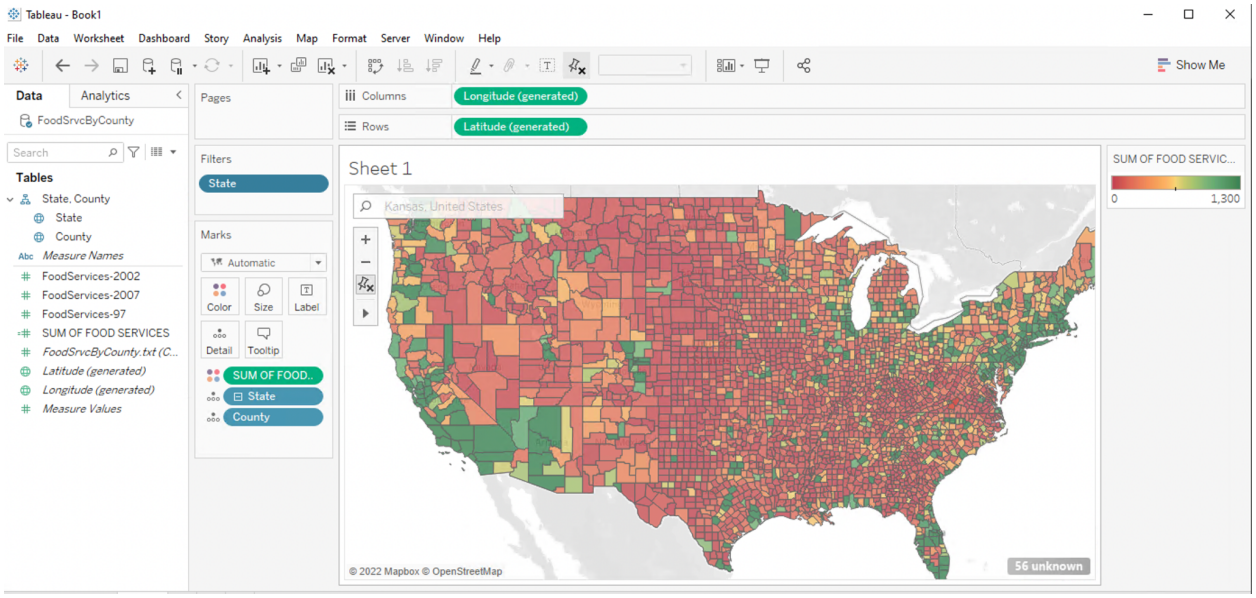
a. Graph food services by state with an appropriate geographic visualization. Note any patterns that arise. Your visualization should clearly display states that have high levels or low levels of food service availability, so think carefully about the color scheme.

It is not a reliable encoding for people to read at specific numbers by distinguishing colors.Instead, the color makes high and lows stand out, so I would use saturation and brightness with a single hue here in order to see where high and low values are.



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

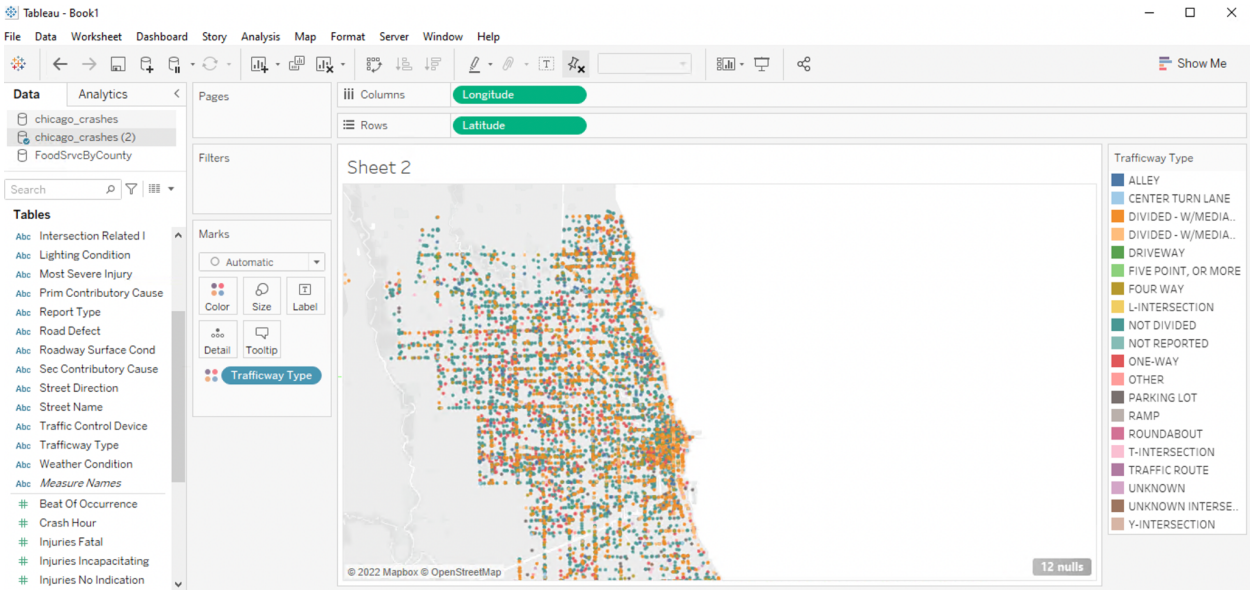
The mean of the FoodService based on country is 645 , so I would use 0 to 1300 as the range of color and 645 at the center point to see not only a changing value low to high but also a mark of the neutral area in the graph.



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 gets dense quickly).

a. Create an appropriate type of geographic plot to show where all the accidents in this data occur.

I would use trafficway type to see where all the accidents occur on the map by plotting out the accident point, so we could know where and which type of traffic road would be more dangerous.



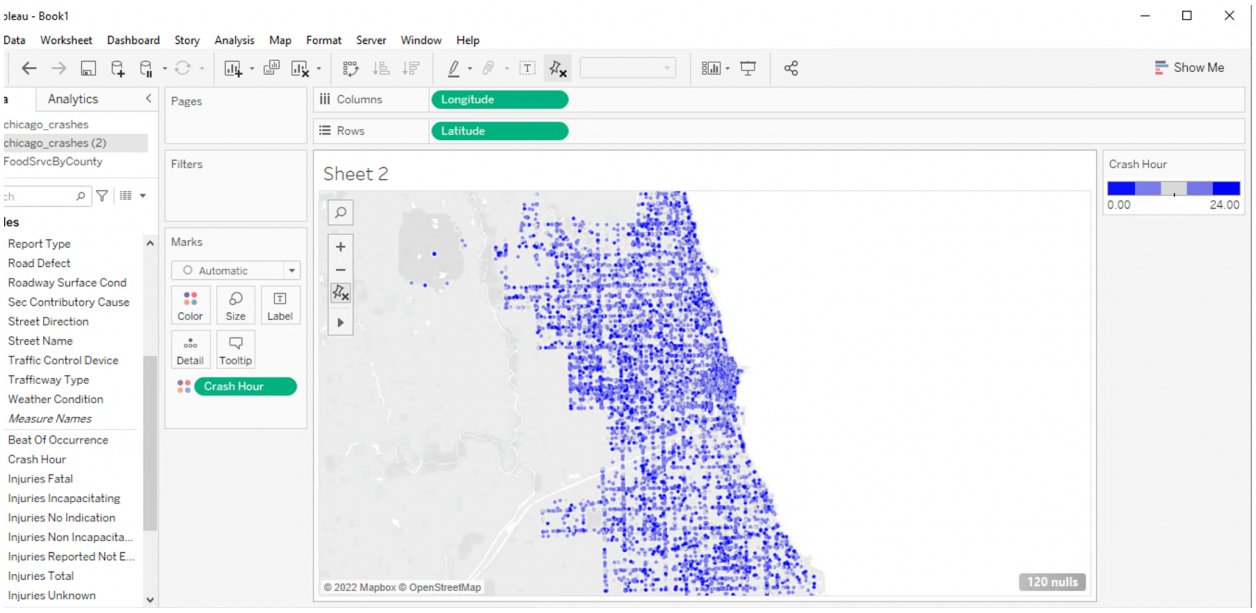
b. Create a visualization that shows how common crashes are in different parts of the city based on 24

time of day. There are multiple approaches to this. Explain your approach and what you can see in your graph.

I would use the divergent color palettes to see what time the accident could happen more frequently.

I marked the two extreme sides with the same color because I'd like to categorize them as the same time which should be at night, and I marked the neutral area with gray color which I would categorize these times as at the daytime.

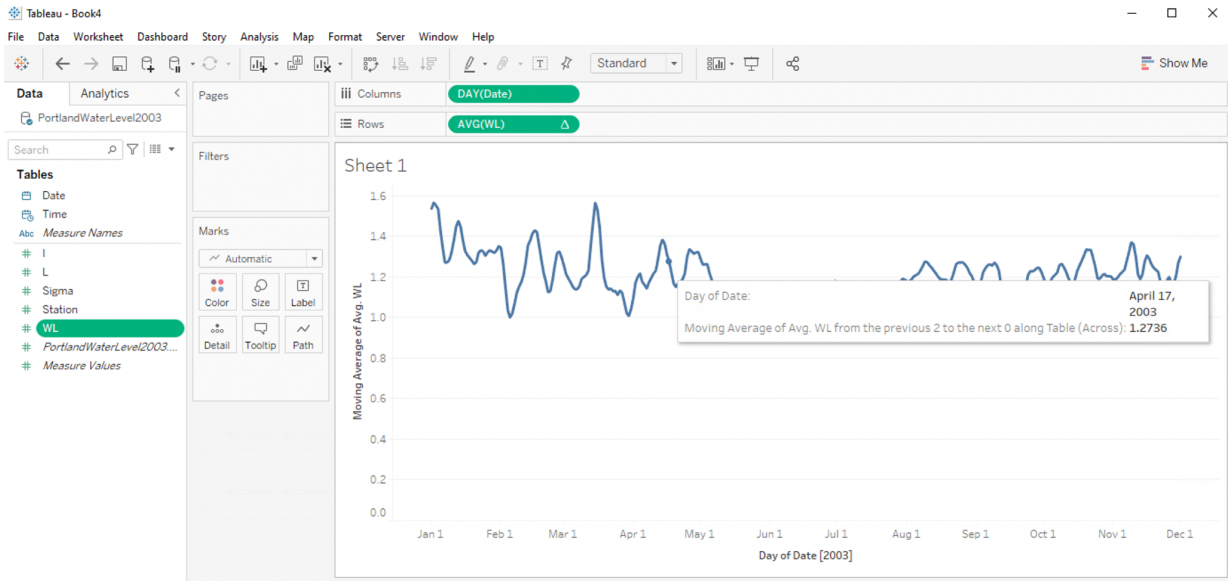
I would say there are more accidents that could happen at night by reading this graph.



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.

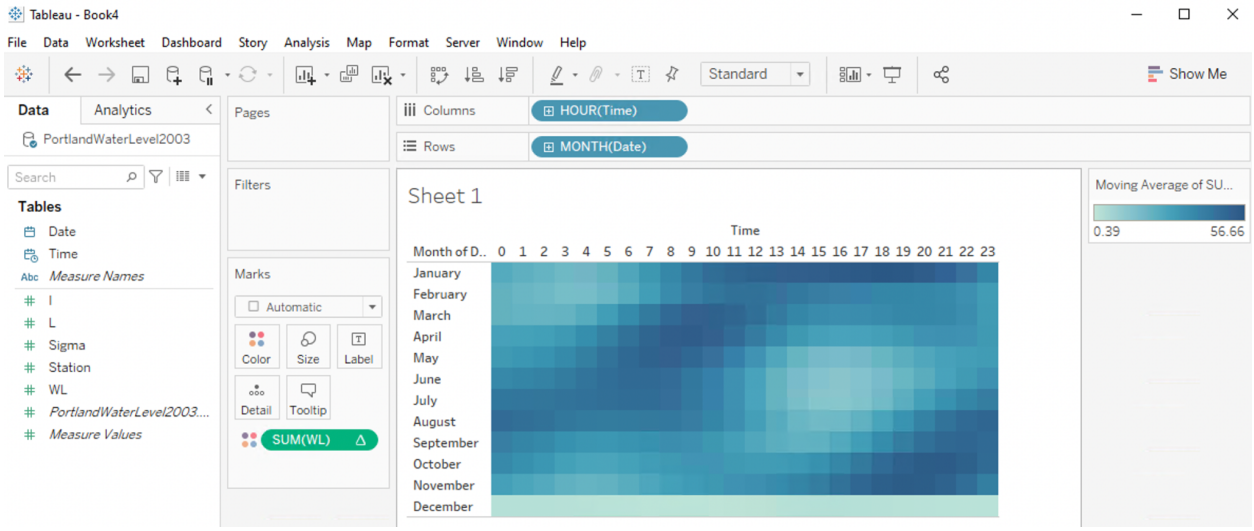
a. This data contains a year of data with water level (WL) measurements every hour as a function of Time (i.e. 365 x 24 data points!). Since there is a lot of data, clean it up by smoothing the data by calculating a moving average. Use a window approach with window size that covers a range of days (remember,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

By using the Tableau, I would use the Date in the column and WL in the row and change WL into moving average to see the change of data from the previous 2 to the next 0 along Table.



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.

By using a heat map, we could see the moving average of hours every months, we could tell that the lowest level is at 3:00, the highest level is at around 15:00 in January, but it is the opposite situation in Julyt which the lowest level is at 15:00, the highest level is at around 3:00.



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

Graph A shows that a moving average of each day in twelve months, we could see an overall water level changed in a year.
Graph B shows a moving average of hours, we could see how water level would change in a day in each month.

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 a color scale in ggplot that is interpolated in Lab space.

In Graph b I would use the red and blue which are the contrast colors to see the change from low to high and use the white color to see where the neutral points are.

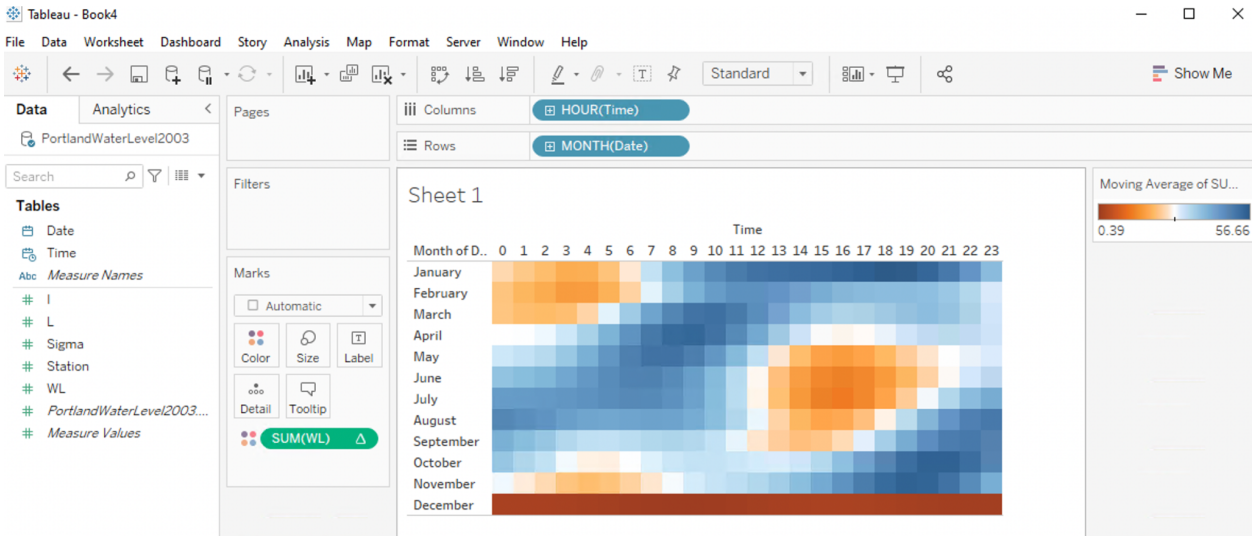


Tableau - Book4
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
PortlandWaterLevel2003
Search
Tables
Date
Time
Measure Names
I
L
Sigma
Station
WL
PortlandWaterLevel2003....
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
SUM(WL)
Columns
HOUR(Time)
Rows
MONTH(Date)
Sheet 1
Time
Month of D..
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
January
February
March
April
May
June
July
August
September
October
November
December
Moving Average of SU...
0.39
56.66