

Final Project

Project Goal: In this project, I want to analyze possible correlation among features of car accidents that happened in the US. I aim to deliver visualizations that provide information about accidents in the US among states, and show what factors may cause severe accidents.

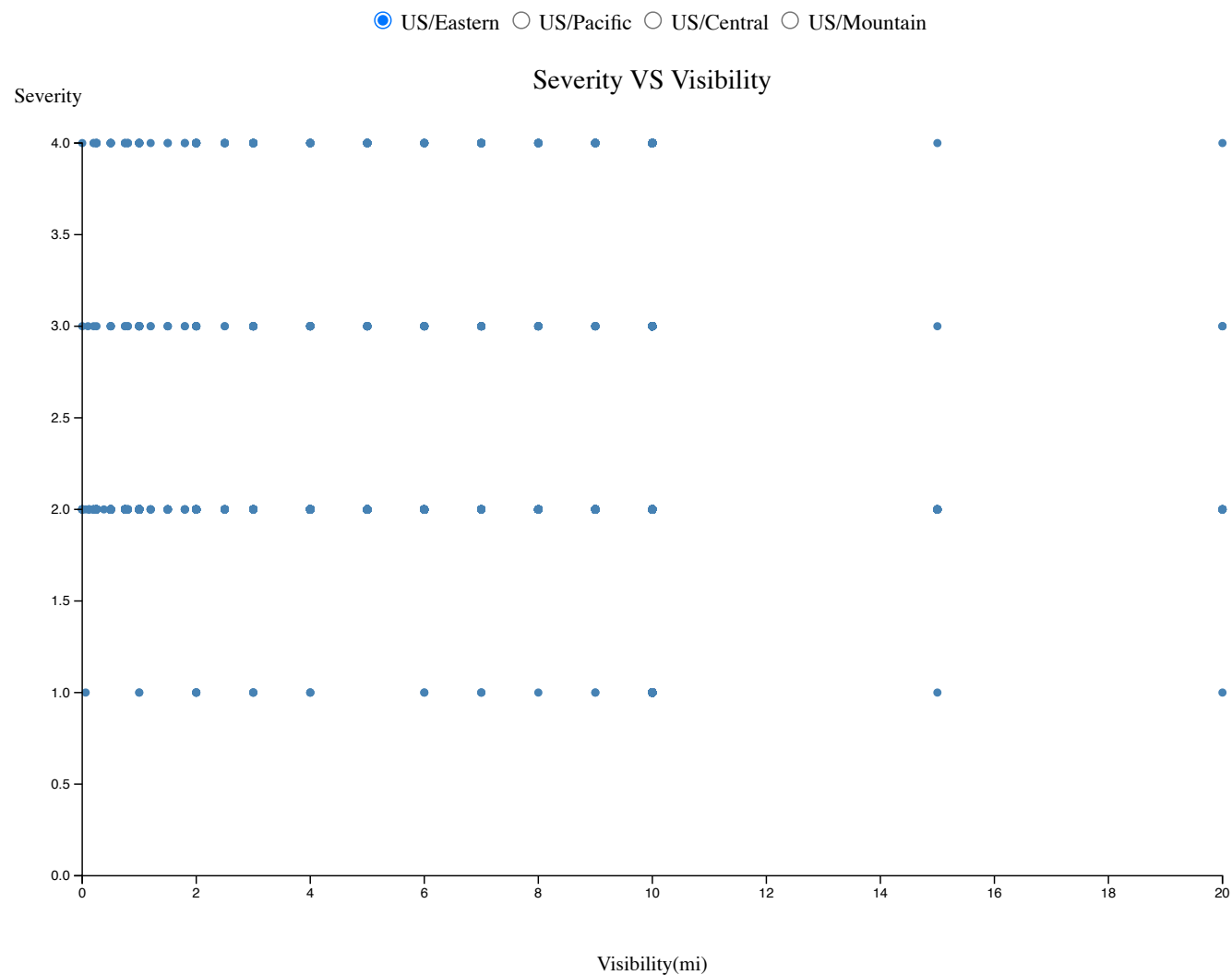
Plot1: Scatterplot

Choice of color scheme: I chose steelblue, plum, mediumaquamarine, and salmon as the color of the dots in my four time zones shown by the radio buttons below.

Marks and channels: The mark of this plot is points, and the channels are horizontal and vertical positions.

Question to answer: Is there a significant relationship between the visibility and severity of the car accident that happened in the US?

Plot:



Story Telling: According to the plot above, I believe that visibility plays some role in the the severity of the accidents. The four plots that you can see by clicking on the radio buttons above the plot shows that accidents marked as severity level 1 are unlikely to happen when the visibility is short. It is interesting to notice that in each time zone, there are more accidents marked as severity level 2 and 3 when the visibility is short. This happens probably because that most of the accidents are marked as level 2 and 3 in my dataset. It is also quite suprising to see that the distribution of accidents marked as severity level 1 and 4 in respect to visibility are quite same.

Plot 2: Barplot

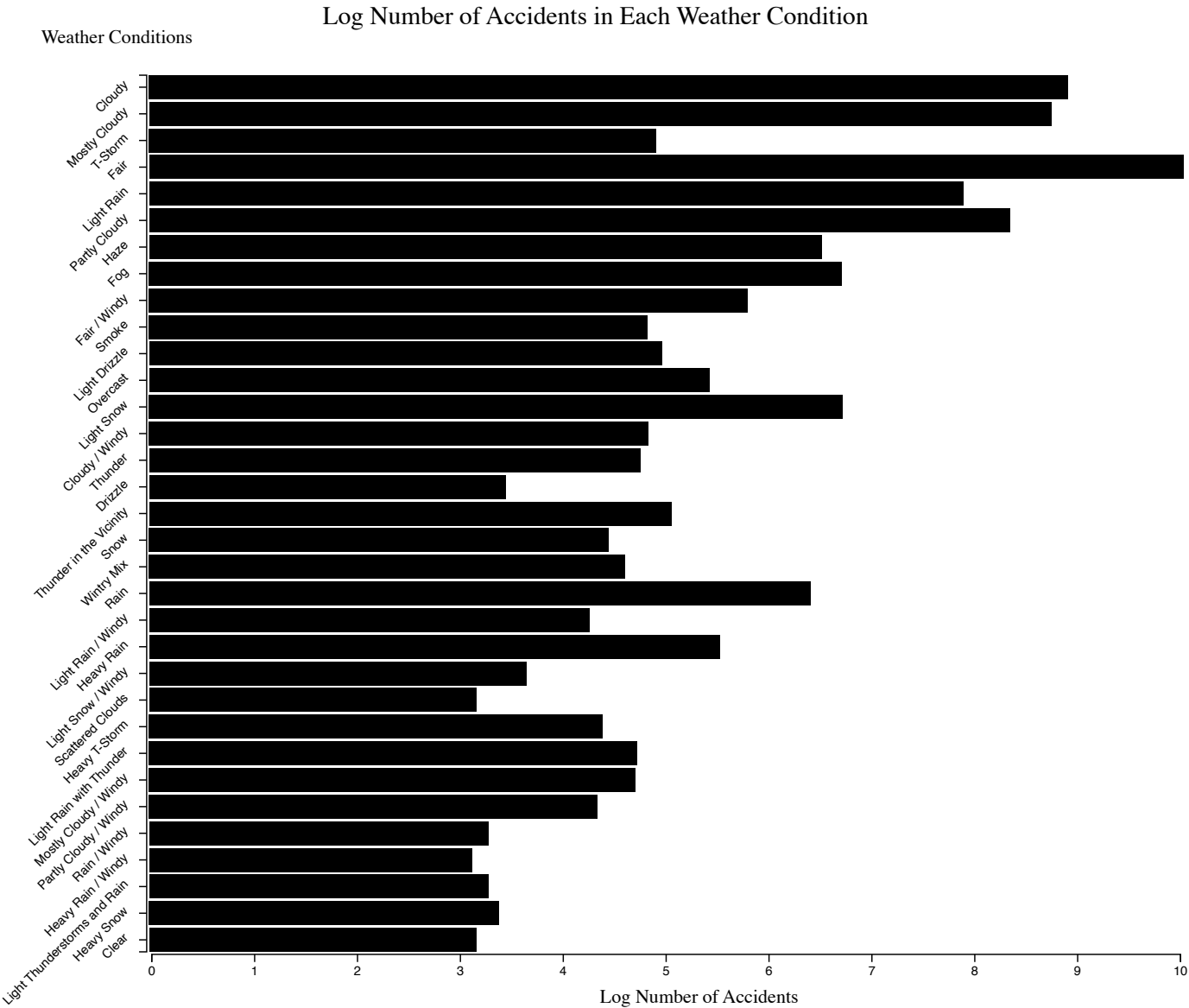
Choice of color scheme: I chose cornflower blue as the color of the bar when the user's mouse hovers above a bar.

Marks and channels: The mark of this plot is lines, and the channels are horizontal and vertical positions.

Question to answer: Does weather conditions significantly affect the number of car accidents that happen in the US? If so, what weather conditions are most likely to have more car accidents?

Plot:

Sort Bars!



Story Telling: According to the plot above, I believe that there is a relationship between weather conditions and the number of car accidents that happen in the US. After clicking the sorting bar on the top left, we can see that in this dataset, the weather that has most accidents happen is fair weather. There are also many accidents happened in cloudy and slight raining days. Suprisingly, the weather that has least number of accidents happen is heavy rain/winding. A possible explanation for it may be that people are more cautious when driving in a heavy raining/winding day.

Plot 3: Streamgraph

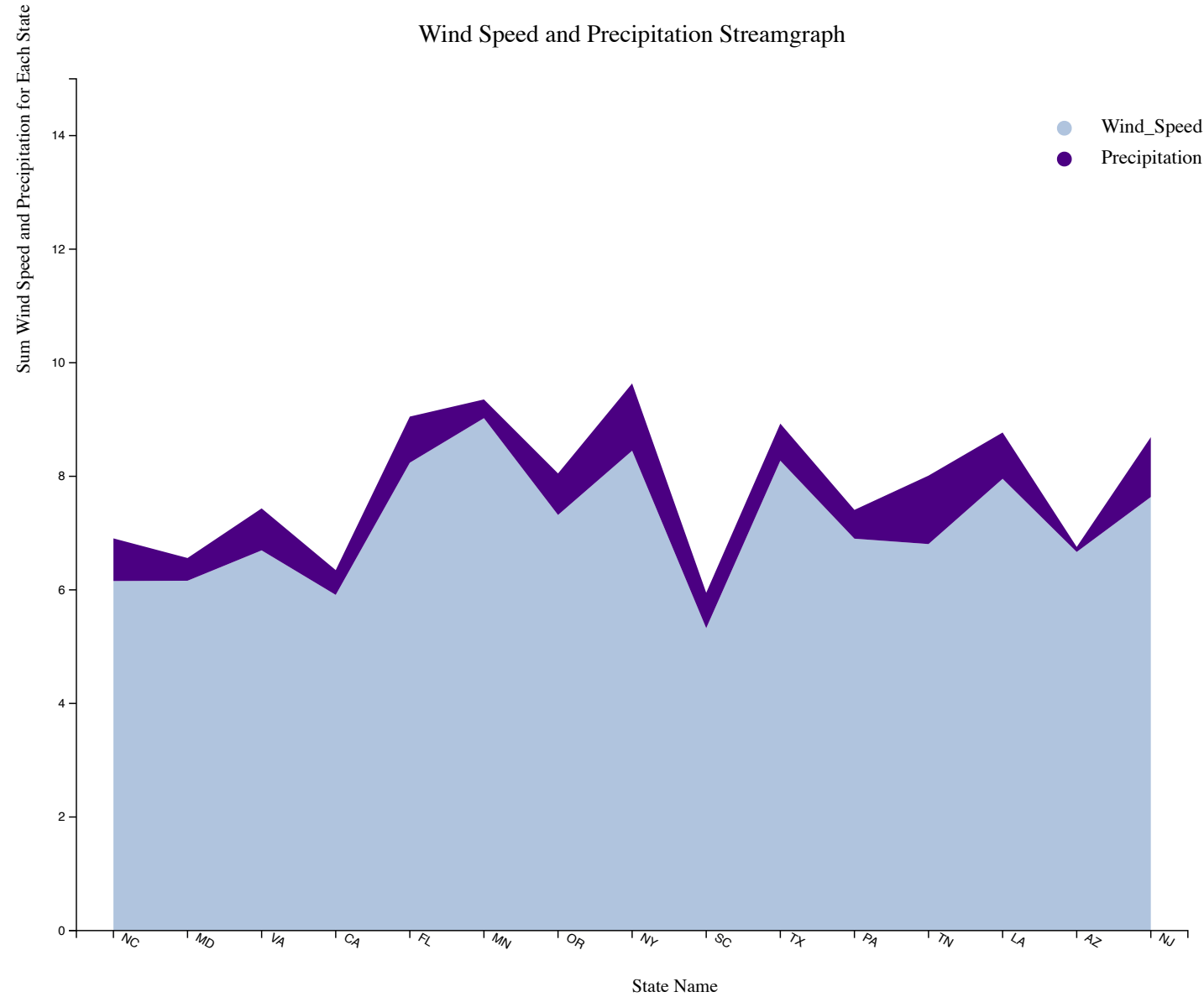
Choice of color scheme: I chose lightsteelblue as the color of the wind speed area, and I chose indigo as the color of the precipitation area.

Marks and channels: The mark of this plot is areas, and the channels are Color hue, horizontal position, and vertical position.

Question to answer: For the top 15 states that have the most accidents, what is the average wind speed and precipitation for each state when accidents happen?

Plot:

Wind Speed and Precipitation Streamgraph



Story Telling: According to the plot above, we can see that among the top 15 states that have the most accidents, South Carolina and California has the lowest average wind speed when accidents happen. Also, Arizona has the lowest average precipitation when accidents happen. Other than that, the top states have pretty similar average wind speed and precipitation when accidents happens. Therefore, I believe that there is no significant relationship among average wind speed and precipitation of a state and the probability that an accidnets will happen in that state.

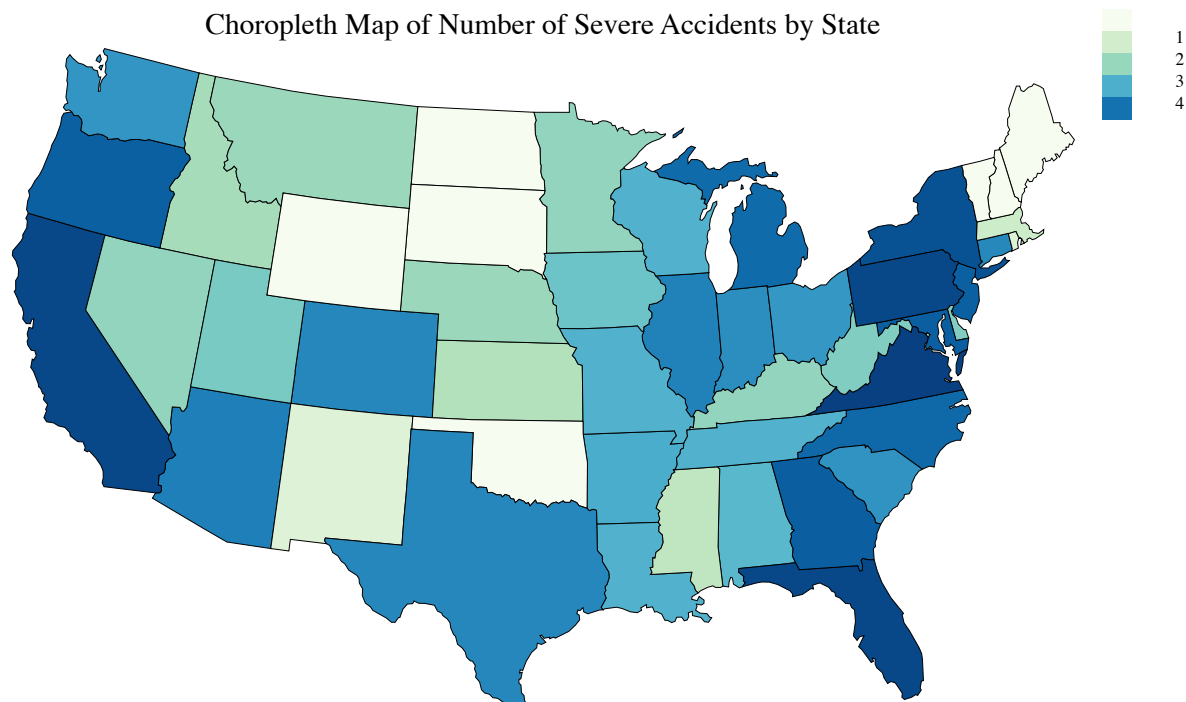
Plot 4: US Geomap

Choice of color scheme: I chose d3.interpolateGnBu as the color scheme for my choropleth map.
Marks and channels: The mark of this plot is areas, and the channels are color, horizontal position, and vertical position.
Question to answer: How many severe accidents happened in each state? Which state has the largest number of severe accidents?

Plot:

Zoom In

Zoom Out



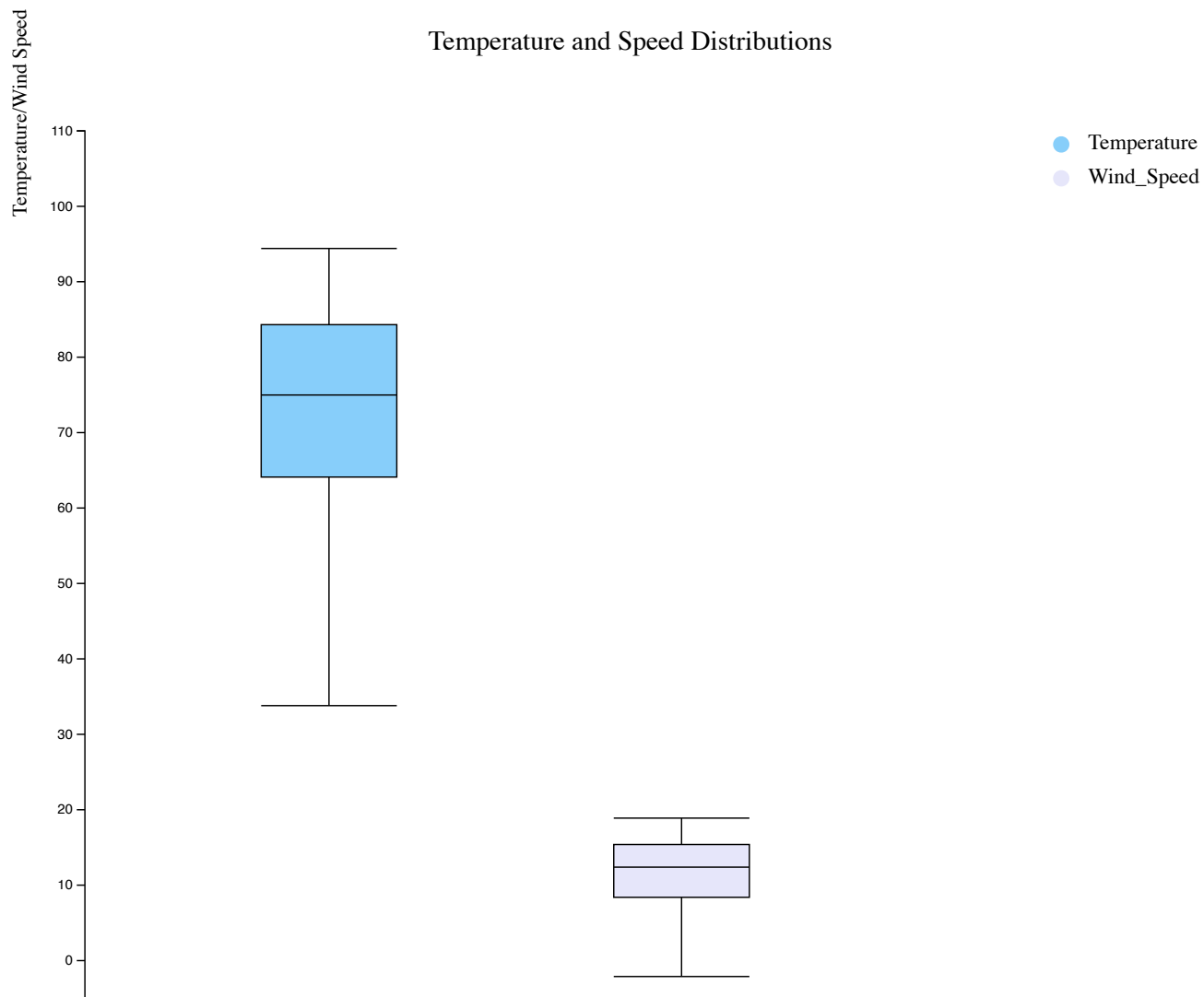
Story Telling: The choropleth map shows the number of severe accidents(have severity level 4) for each state. The darker the color fill on the state, the higher number of accidents the state has. From this map, I found that the top states that have high number of severe accidents are California, Florida, and Virginia as they are colored in dark blue. The states that have low number of severe accidents are Oklahoma, Wyoming, North Dakota, and South Dakota. A possible explanations may be the difference in populations among different states. People's different driving habits in each state may also be a possible reason for the difference in number of severe accidents.

Plot 5: Box Plots

Choice of color scheme: I chose lightskyblue and lavender to represent temperature and wind speed.

Marks and channels: The mark of this plot is lines, and the channels are colors, horizontal position, and vertical position.
Question to answer: What are the distributions of temperature and wind speed for all the accidents recorded in the dataset?

Plot:



Story Telling: From the combined box plot shown above, the middle 50 percent of the accidents happened when temperature ranges from 50 Fahrenheit to 76 Fahrenheit, and when wind speed ranges from 3 miles per hour to 10 miles per hour. According to the schema that I found from Internet, temperature between 50 Fahrenheit and 76 Fahrenheit are considered as warm and cool. Wind speed between 3 miles per hour and 10 miles per hour. This shows that most accidents happen when temperature is warm and cool, and wind is light and gentle breeze. A possible explanation for that is people may be more likely to go out in warm/cool and breezing days.

Summary

In this project, I drew five visualizations to analyze the possible correlation among features of car accidents that happened in the US. From my visualizations, I found that California, Florida, and Virginia have the highest number of severe accidents happened, while Oklahoma, Wyoming, North Dakota, and South Dakota have the lowest number of severe accidents happened. Among the accident recorded in my dataset, when the visibility of the driver is short, most of the accidents are marked as severity level 2 and 3, some of the accidents are marked as severity level 4, and only a few are marked as severity level 1. This happens probably because that most of the accidents are marked as level 2 and 3 in my dataset.

I also found that there are some relationships between weather conditions and the number of car accidents that happened in the US. From my bar chart, the weather that has most accidents happen is fair weather. There are also many accidents happened in cloudy and slight raining days. Moreover, the combined box plot shown in plot shows that most accidents happen when temperature is warm and cool, and wind is light and gentle breeze. To my surprise, the weather that has least number of accidents happen is heavy rain/winding. The number of accidents happen light thunderstorms and raining days is also lower than what I expected. A possible reason for that may be that people prefer to go out when weather is nice and warm, and choose to stay at home in bad weathers.