

Joel-Cedric Beugre
Homework 0

Description of the dataset

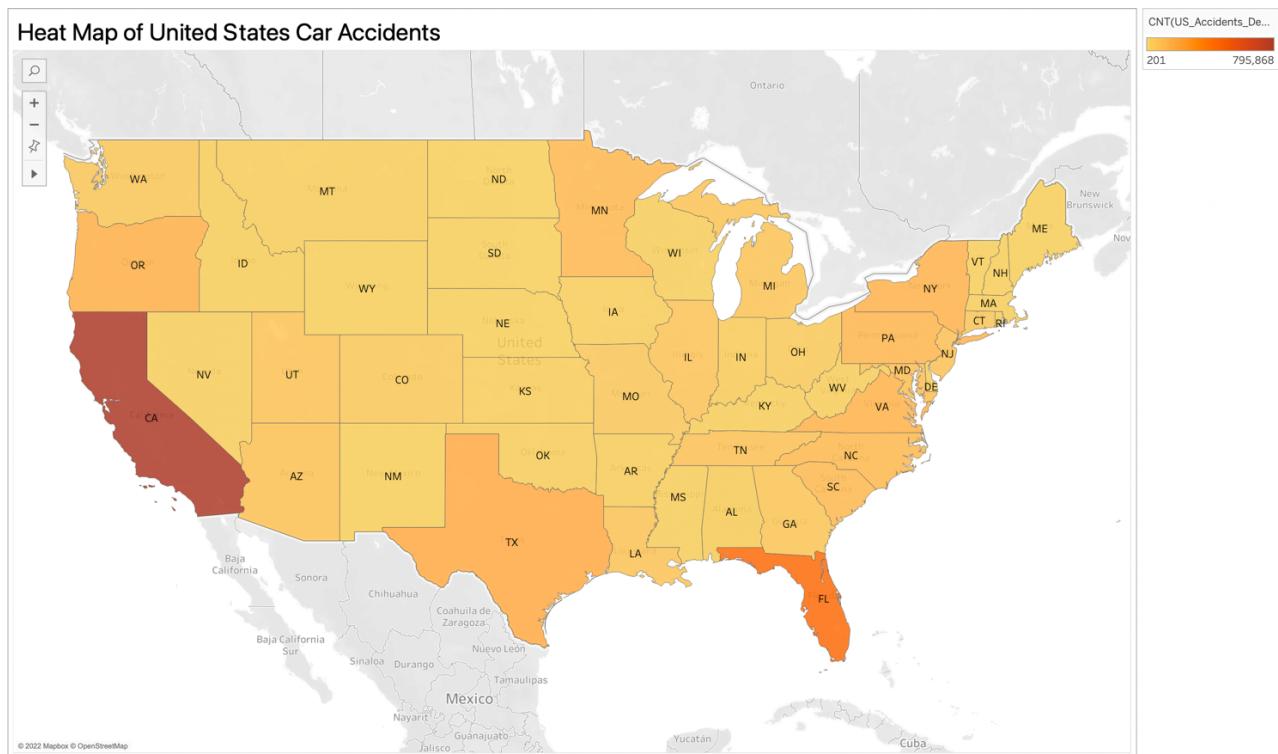
The dataset chose for this homework is a countrywide car accident dataset generated by multiples API. This dataset covers 49 states and has been collected in real-time from February 2016 to December 2021. This dataset contains 47 columns and 2 million of rows.

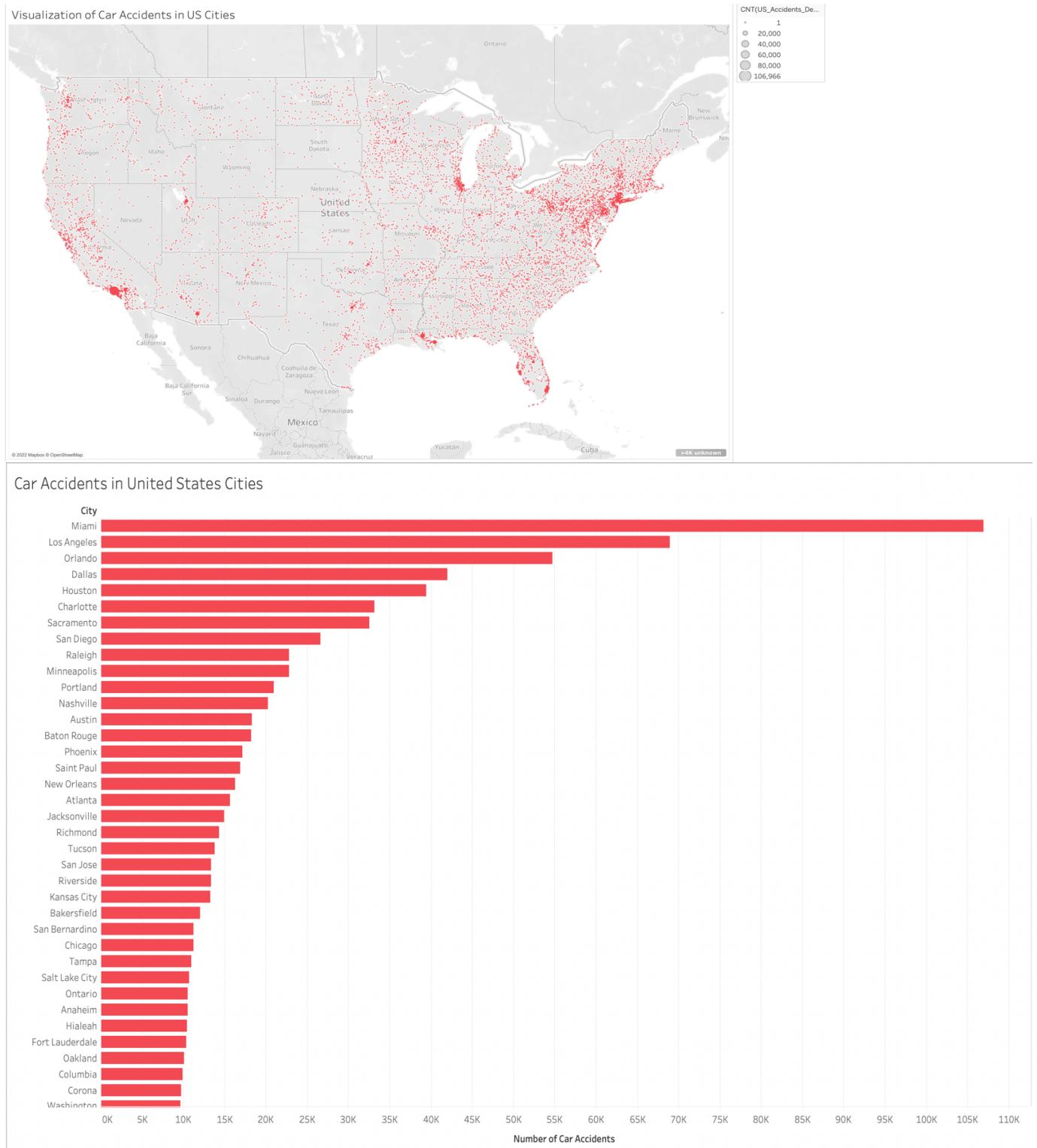
Source

<https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents>

Analysis Notebook

- Which geographic zone registers more accidents ?

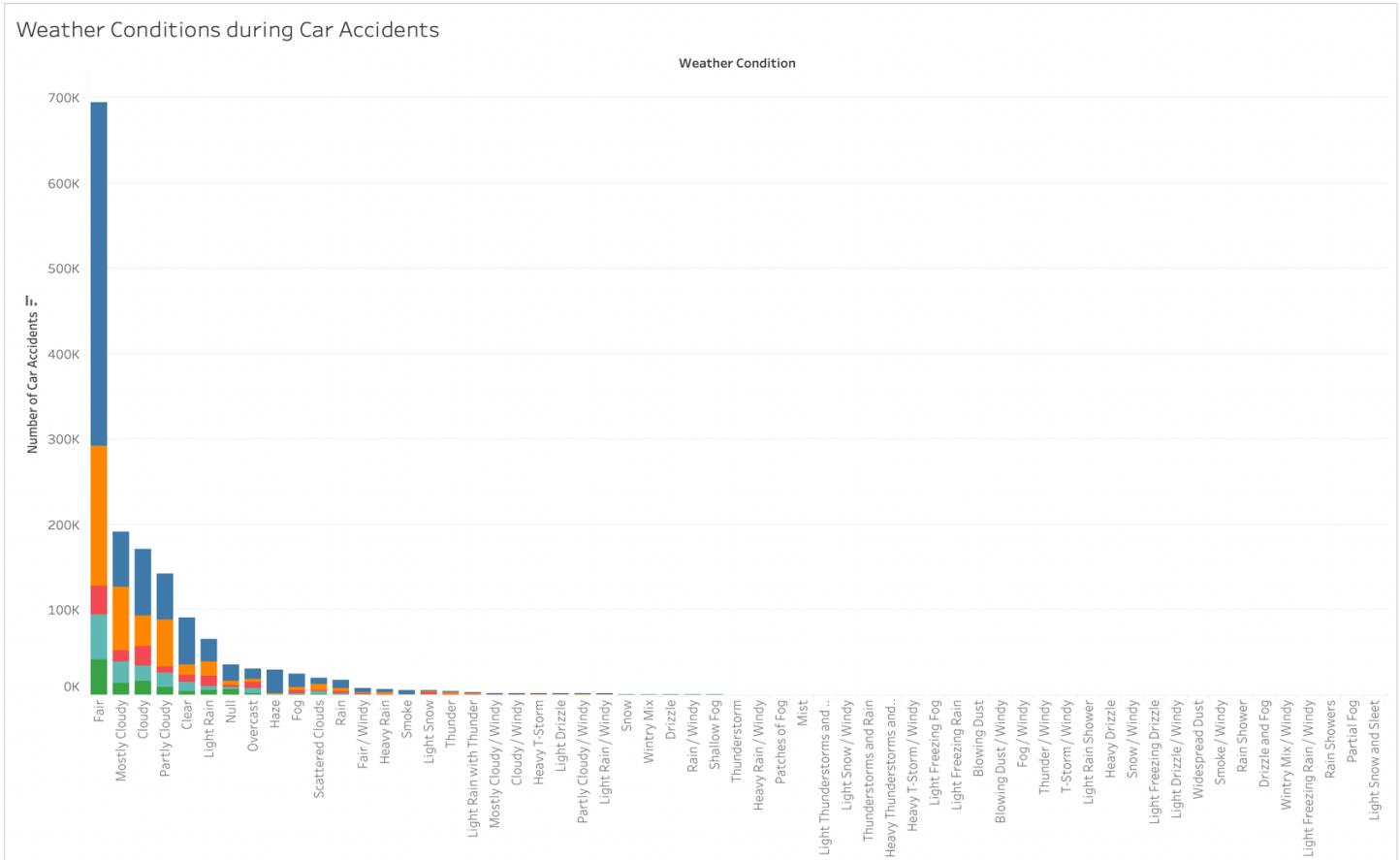




The heatmap of United States and the visualization of Car Accidents in US cities are showing us that most of the accident that occur in the United States happen around the coast and borders. From those data, we could suppose that this is probably because of the

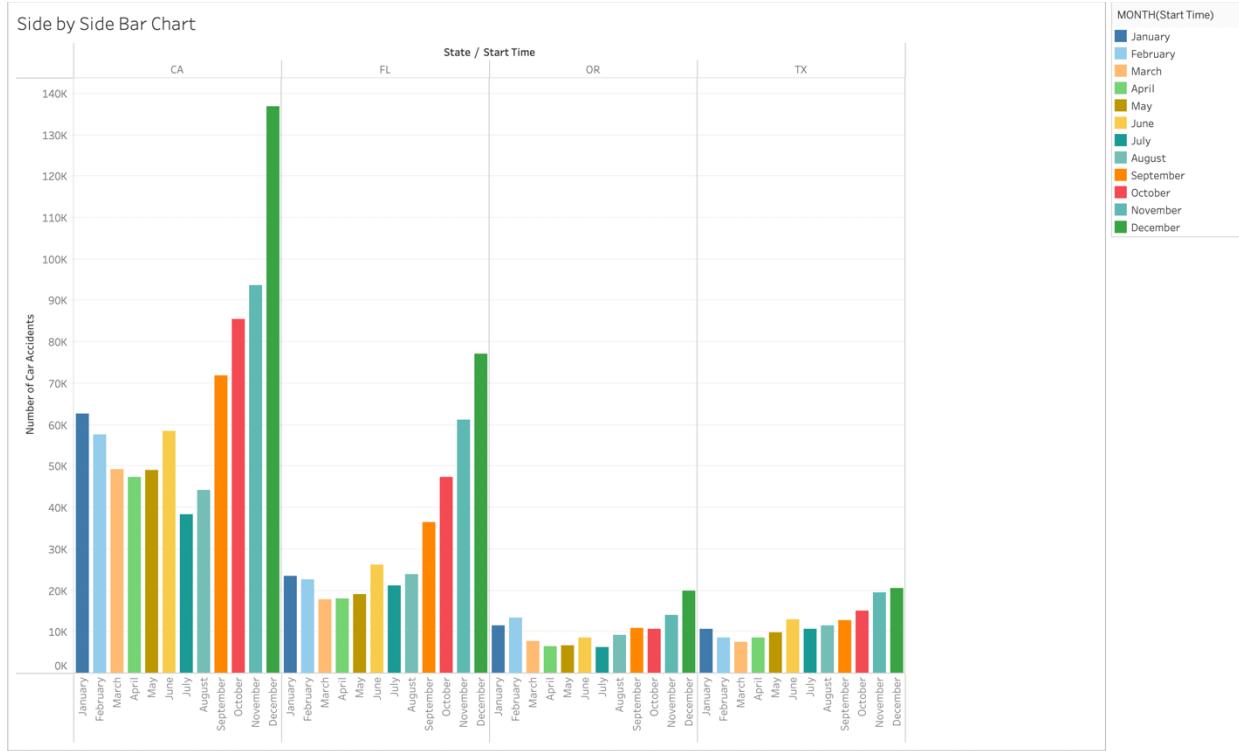
weather, but the bar chart of Car accidents in the United States cities also shows that most of the accident happen in big cities.

- *Is the fact that the states with the most accidents are on the coast related to the weather?*



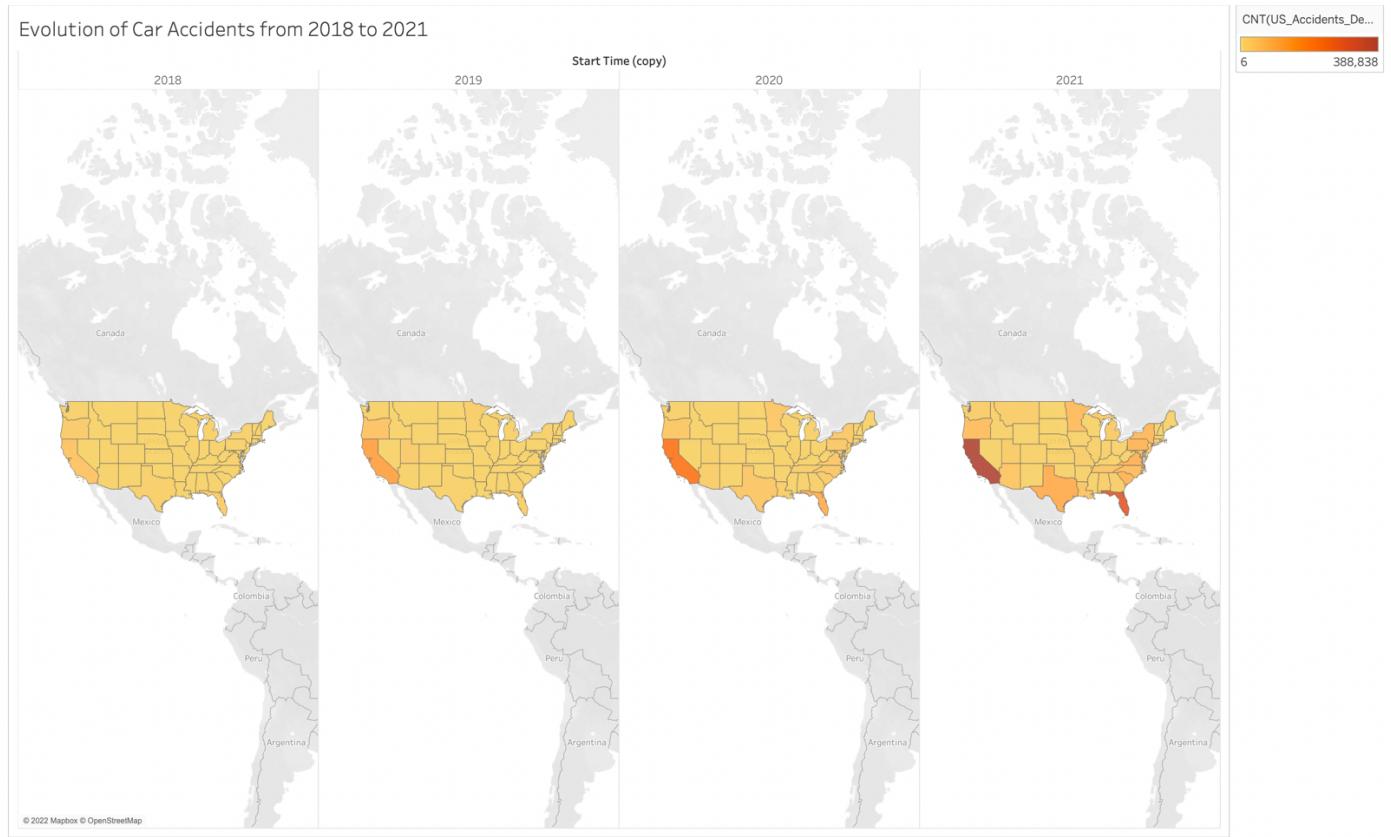
From this stacked bar chart, we can see that more than half of the Car accidents happen when the weather was fair. Aside from fair weather, cloudy weather appears to be the weather condition that caused the most accidents, likely due to visibility.

- *Is there a period in time when the accident occurs most on those zones?*



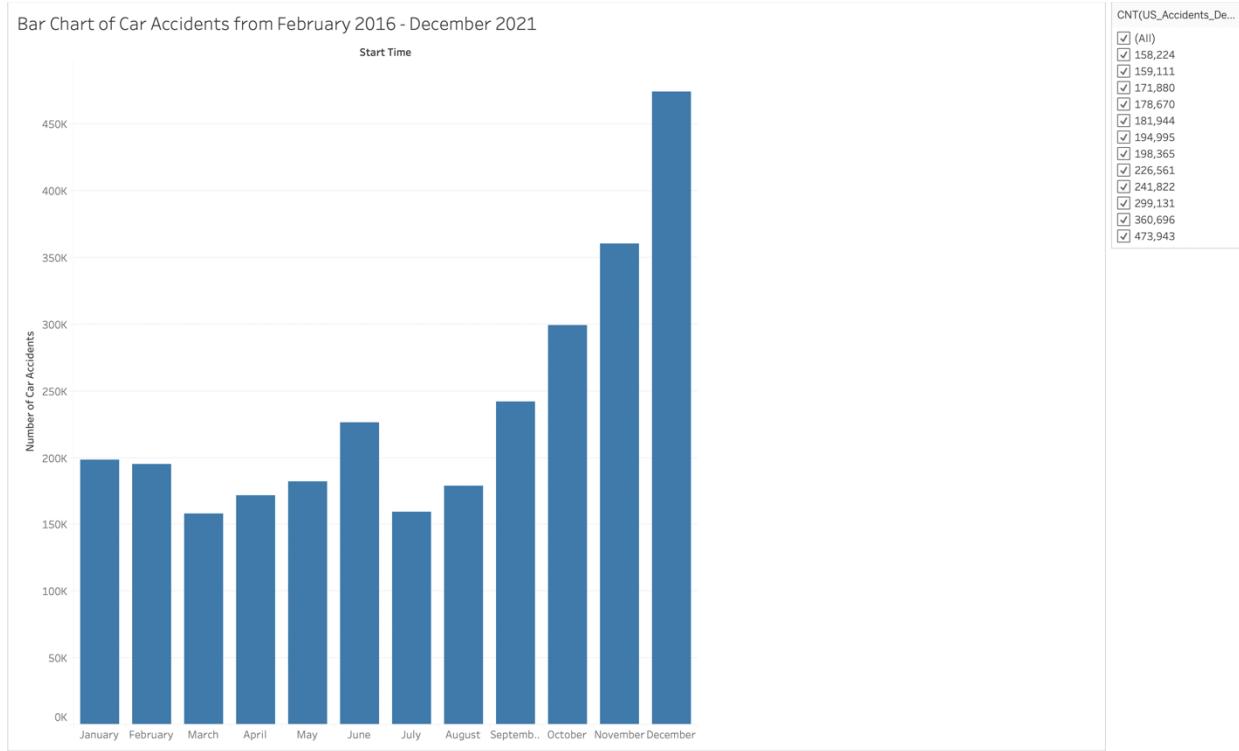
We can observe from side-by-side bar chart that most accidents occur during the winter season in all the states with the most accidents, although the previous graph shows us that most accidents occur during fair weather and not snow. We can then suppose that it's not about the weather but probably because of the number of cars in those cities since they are known for their number of inhabitants.

- *How has the number of car accidents on the coast changed over time?*



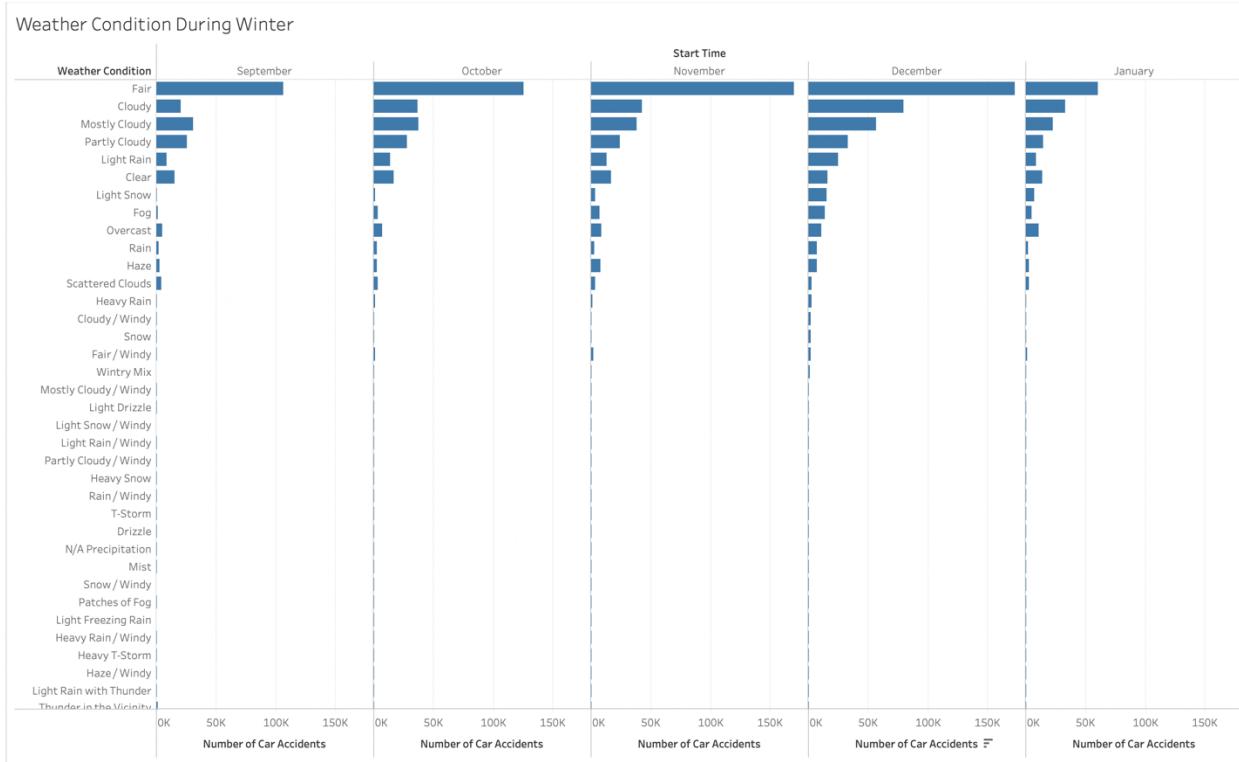
This heatmap showing the evolution of Car Accidents showed that the amount of car accidents increases each year, which can be explained by the population growth and the amount of car, but it can also be the consequence of global warming.

- What is the period of the year when we have the highest probability of car accident in the US?



This bar chart reveals that most of the accidents globally happen at the end of the year, but also have a sudden increase in June which can make us think of a correlation between season and the number of car accidents.

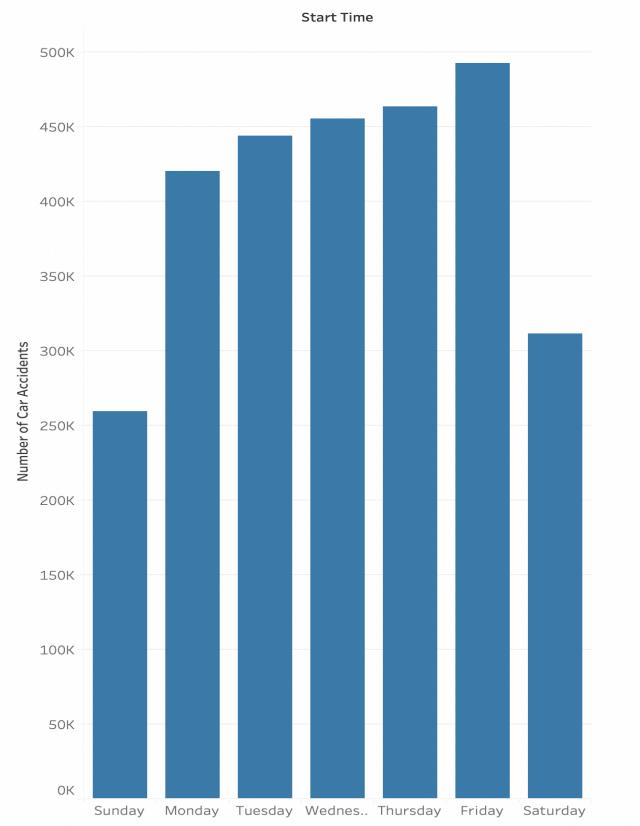
- *Is the high number of accidents at the end of the years related to the season?*



Although most accidents happen during winter, we can observe on this graph that the weather is fair in the majority of accidents. But we can also notice that car accidents during cloudy gradually increase when it approaches the end of the year, meaning that weather slightly influences car accident. The snow is low on this graph probably because most car accidents observed are in zone with low snow precipitation during winter

- Which weekday has the most accident during the dataset time period?

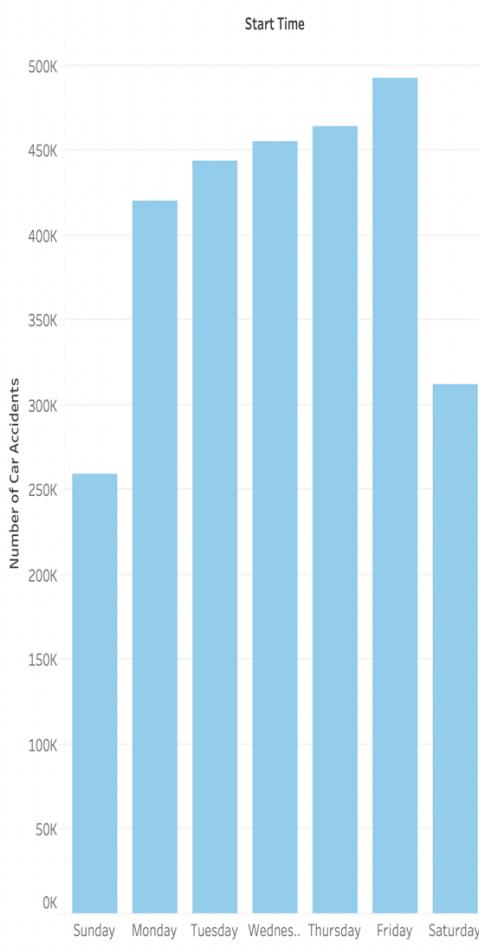
Bar Chart of Car Accidents



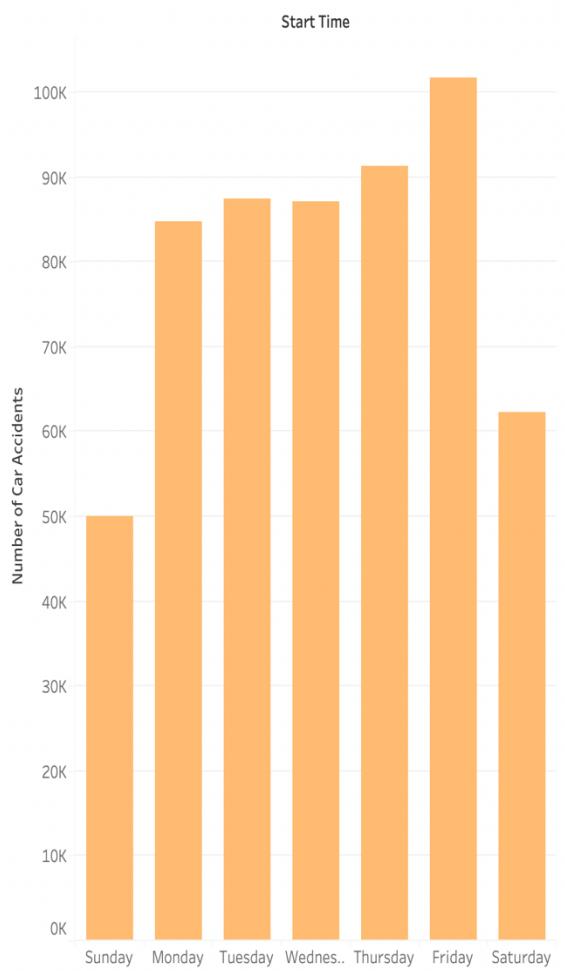
This bar chart displays the number of car accidents that happen during each weekday. We can see that the majority of accidents occur on work day what can be explained by the number cars on the street increase during the business days. A reasonable explanation of why most accident occur during fair weather.

- *Does the weather season affect the weekdays with the most accidents?*

Bar Chart of Car Accidents by Week Date during Winter



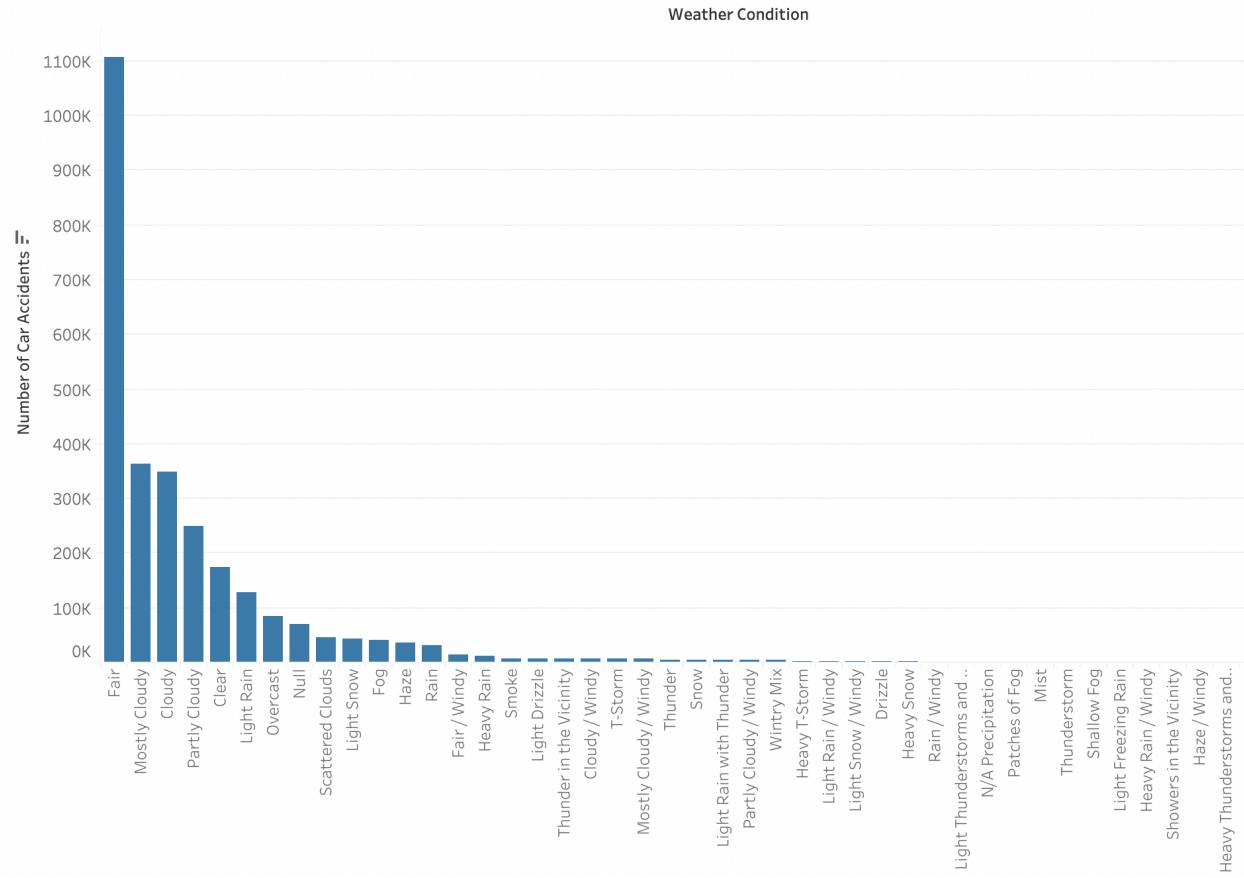
Bar Chart of Car Accidents by Week Date during Summer



The number of car accidents is greater in the winter, but we can see that both bar charts have approximately the same shape except the fact that there's not an important change on Monday, Tuesday and Wednesday during the summer, probably because traffics are less busy on summer.

- Which natural conditions cause the most accidents in the US ?

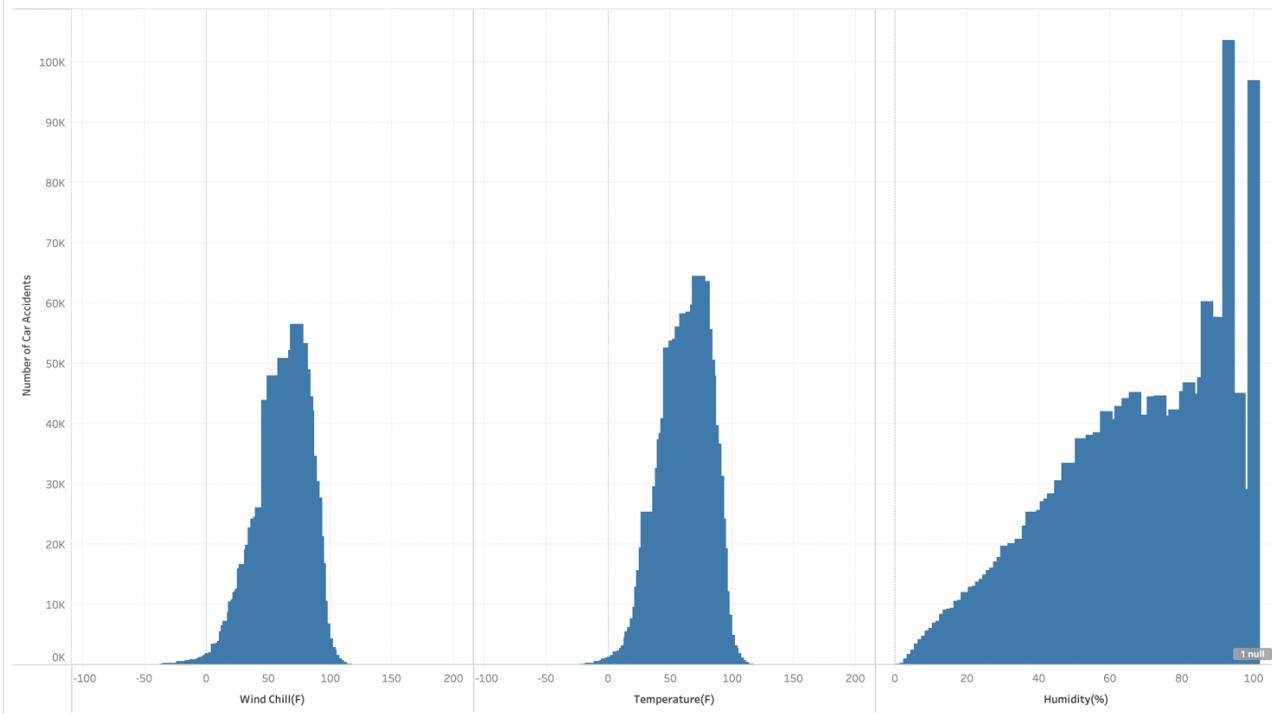
Weather Condition During Car accident (2106-2021)



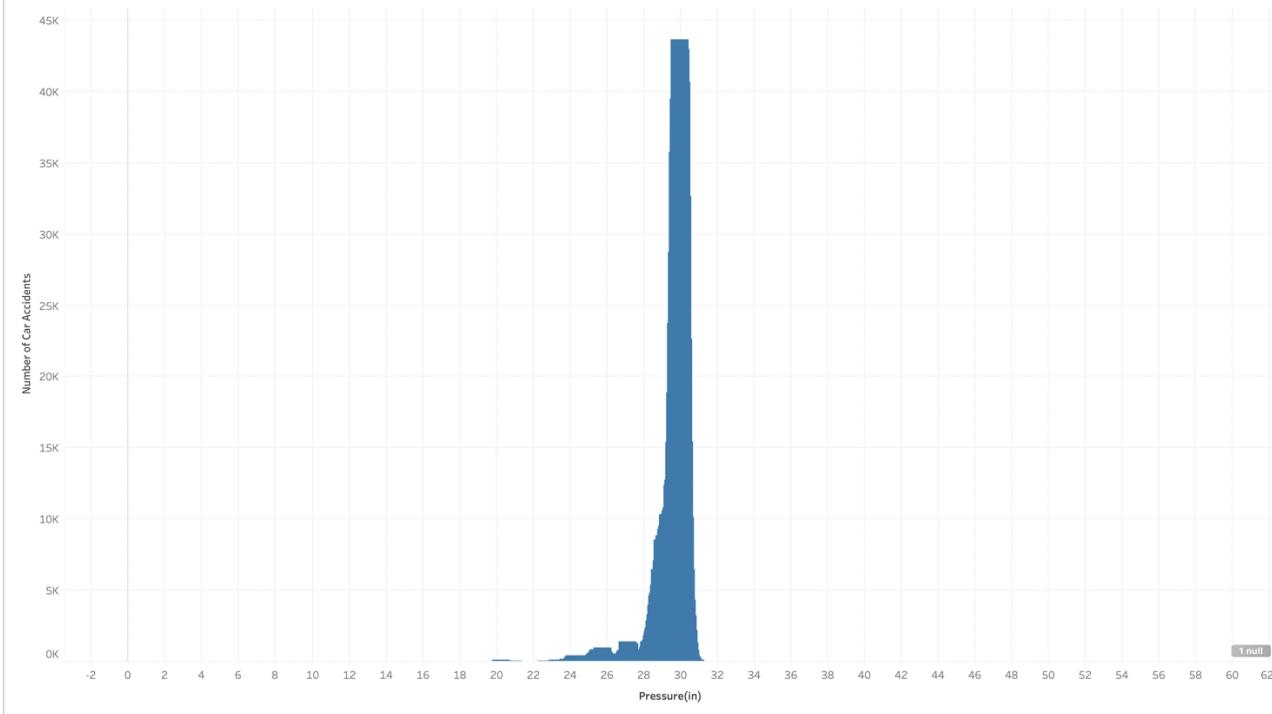
Except from fair Weather Condition when most of the car accident happen when can observe that Cloudy weather came. This can make us wonder if it is because of the visibility or other weather condition.

- *What other natural cause other than the climate cause the most accident during fair climate?*

Histogram of Natural Condition during Car Accidents

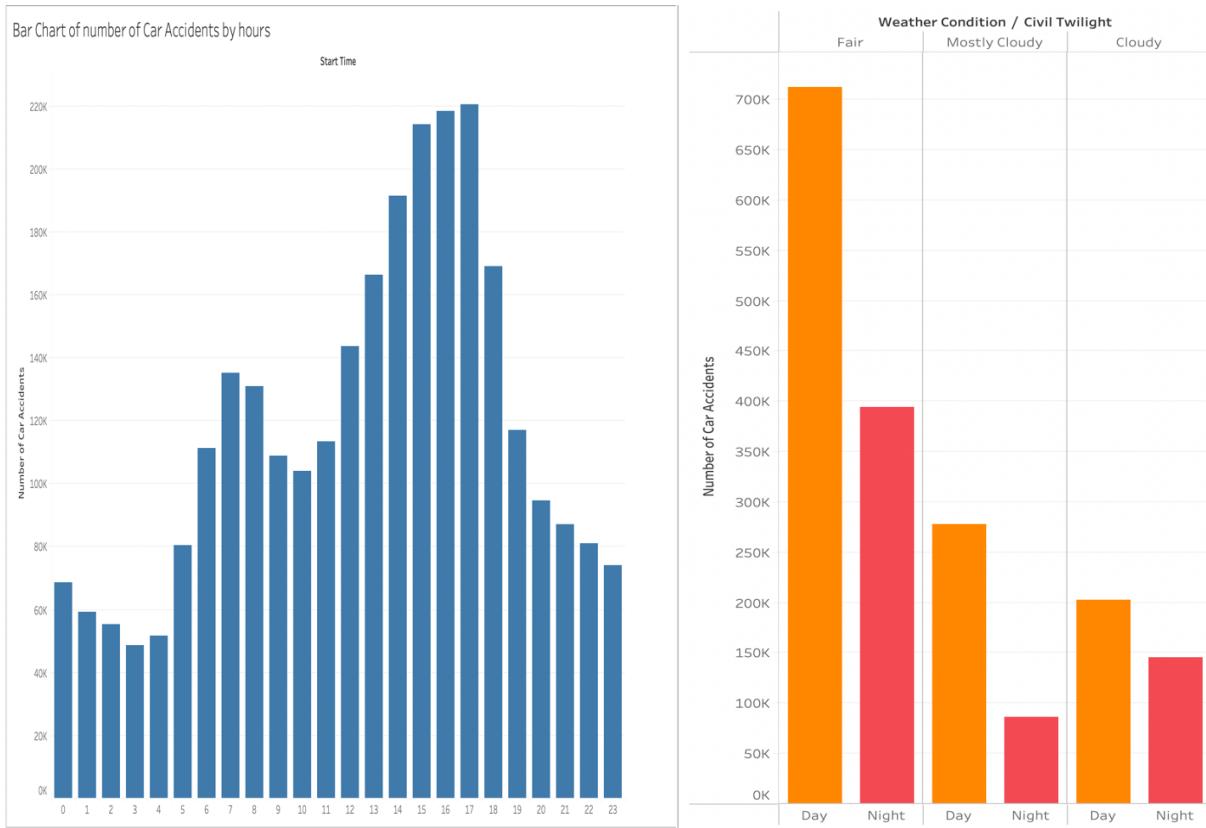


Histogram of Pressure during Car Accidents



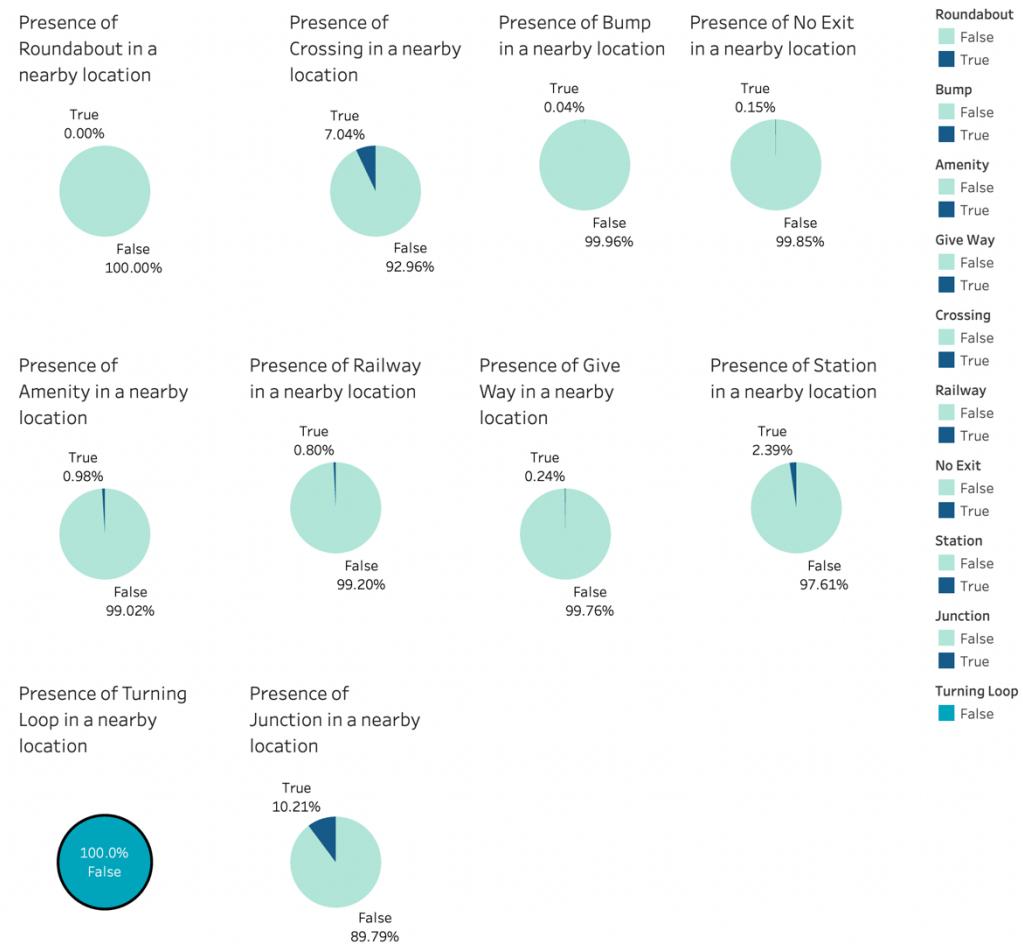
On those histograms, we can see that the majority of Accident occurs when Wind chills and Temperature are around 50-degree Fahrenheit. We also see that humidity is really important because the more humidity the percentage increases, the more accidents occur.

- *Is the time of the day have an impact on the number of accidents?*



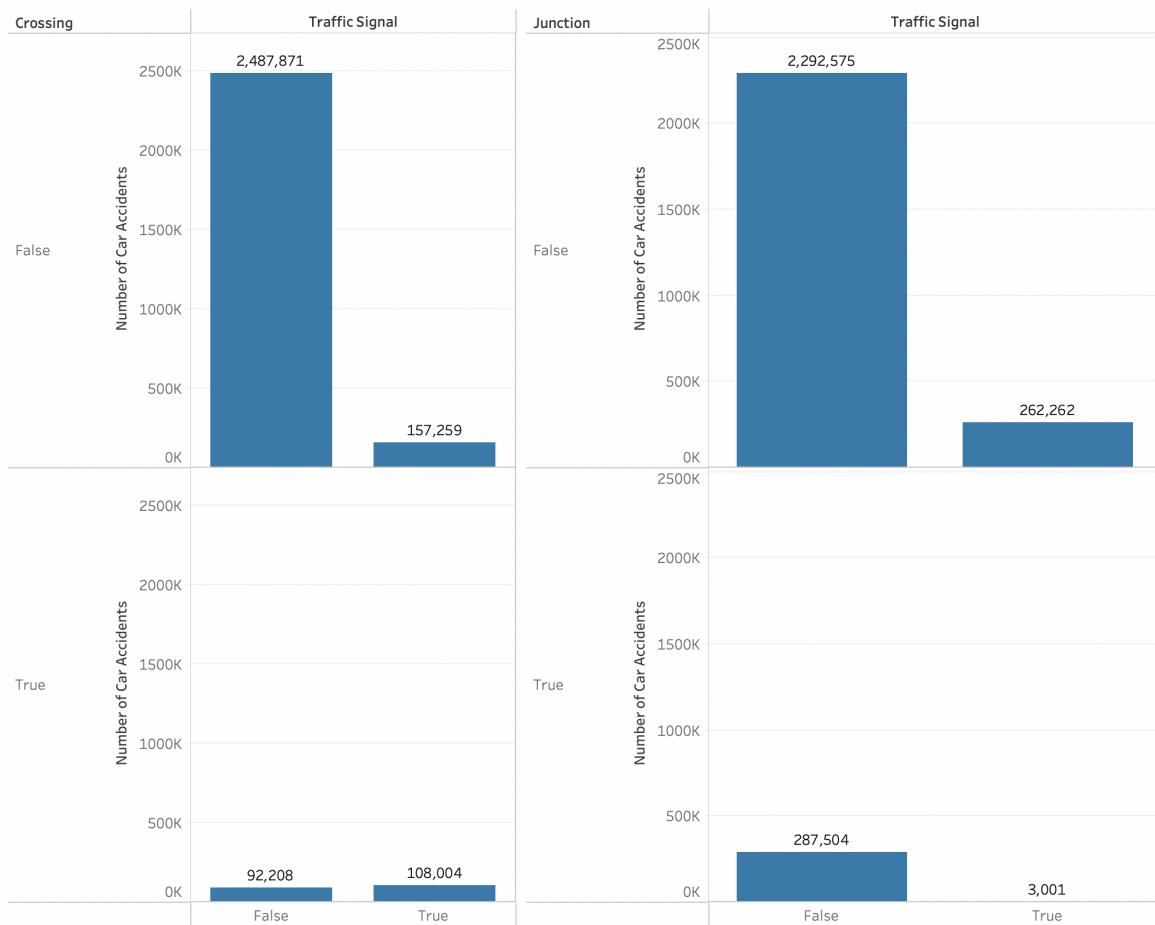
On the first graph, we can notice that both peaks are the time when sunset and sunrise happen and both are also natural condition that can cause difficulty for drivers to see obstacles. The second graph shows that most of the accident happen during the daylight regardless of the weather Condition. Another proof showing that the number of cars on the streets influences the number of car accidents.

- Is there a relationship between the location and accidents ?



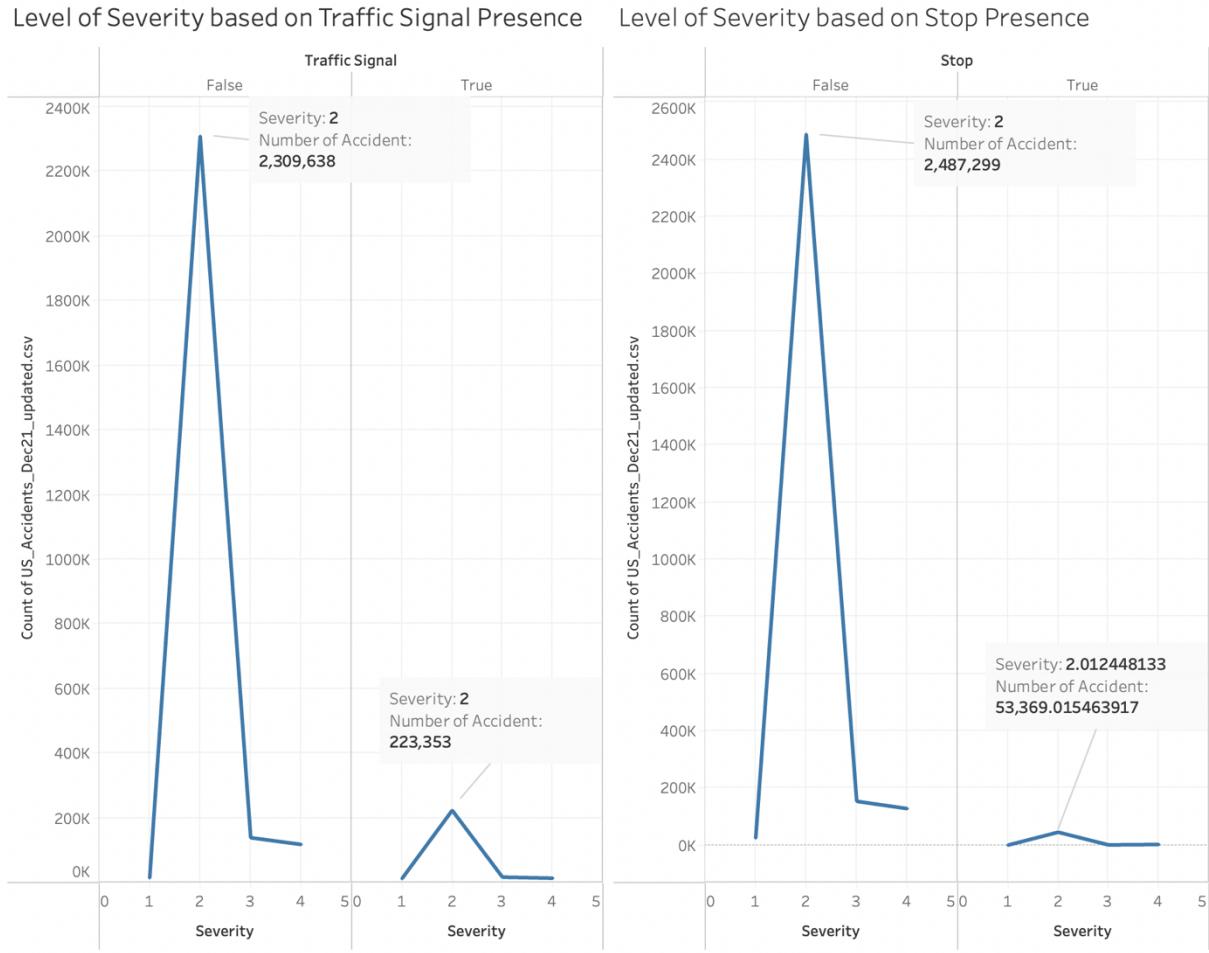
Those pie charts show that some road locations don't seem to affect the number of accidents. Except Junction and Crossing, which are above 5%, the others are very close to zero showing that the infrastructure around the place where accidents happen do not really matter.

- *Do traffic rules help significantly with preventing accidents on Junction and Crossing?*



Crossings without Traffic Signal have almost the same amount of car accidents with crossings with Traffic signal, probably because it's easier to see the cars coming to the intersection. In the junction case it seemed like there is a great relationship with the traffic signal. Junction without traffic signal nearly double the number of car accidents at Crossing with traffic signal.

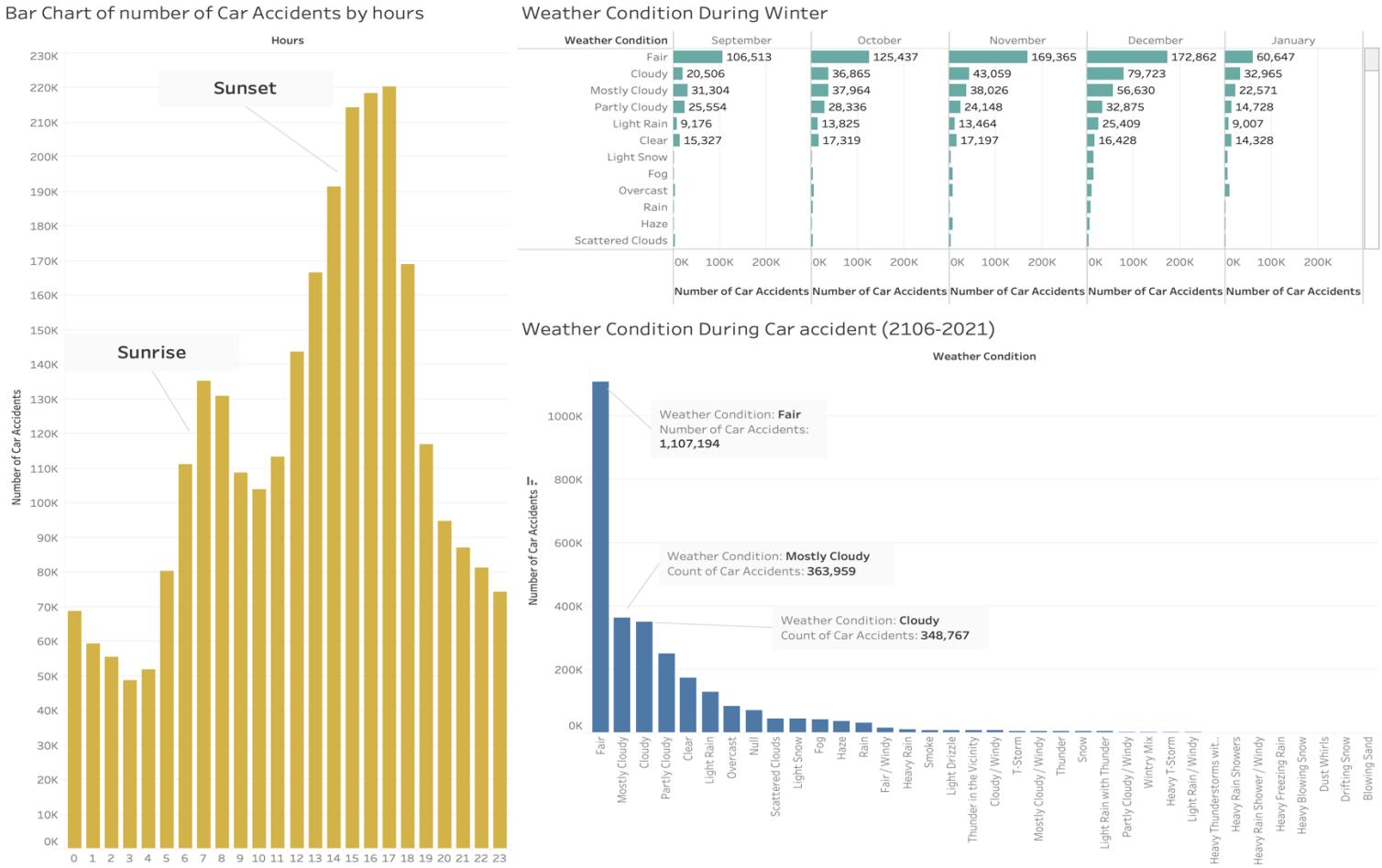
- *Do accidents near traffic signs decrease the severity of accidents?*



The severity of accidents are less important with the presence of traffic light and Stop because they slow the traffic and then prevent other driver to get caught in the accident

A final visualization

- Which natural conditions cause the most accidents in the US?



This is the final visualization to answer to the question above. During the analysis of this dataset, I was surprised by the fact that the majority of accidents occur when the weather was fair and as you can see on the Weather condition graphs above, even during the winter season, car accidents seem to occur most of the time during days when the weather was fair. So come the graph about the number of Car Accident by hours.

Since the position of the sun is still a natural event, I thought I would be wise to include it in my analysis. Indeed, the bar chart of number of Car Accident by hours shows us that most accidents occur during Sunrise and Sunset. Probably because of the visibility, but I believe is more likely because it's also the moment when people go to work and come from work. Those are also the moment when people are tired.

From these analyses, other questions can be asked to understand how the time of the day influences the probability of getting into an accident.

Citation

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "[A Countrywide Traffic Accident Dataset](#).", 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "[Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights](#)." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.