Traffic Accidents in the United States

UNC Bootcamp - Jan 2020
Erika Berry, Tyler Carey, Matthew Fahys, Kate Minyard, Sree Yallapragada

Research Questions

- What times of the day are traffic accidents most prevalent?
- What locations nationally and locally have a greater frequency of traffic accidents?
- What locations nationally and locally have a greater severity of traffic accidents?
- Which, if any, weather events lead to an increase in traffic accidents?
- Does the presence of precipitation indicate a greater severity of traffic accidents?

Our Hypotheses

- Rush hour in the afternoon is the time of day that a traffic accident is most likely.
- Major US cities have a greater frequency of traffic accidents. Locally, the arteries into downtown Raleigh have the greatest frequency of traffic accidents.
- Where major US highways in major US cities intersect, those are the locations where we'll find the greatest severity of traffic accidents. Locally, the arteries into downtown Raleigh have the greatest severity of traffic accidents.
- Heavy rain and sleet/ice are the weather events that will most likely lead to an increase in traffic accidents.
- Severity will be greater in inclimate weather

Our Data

- Kaggle.com - US Accident Data
  - Easy accessibility, and format of their datasets
  - Expressed interest in traffic accidents as they directly relate to our everyday lives.
  - Variety of different points

Data Exploration/Cleaning

In [16]:

```
# Data cleanup

# Filters out dates containing 2016, 2017, & 2019 -- too much data in dataset to handle with lapto
ps
accident_data_revised = accident_data[~accident_data["Start_Time"].str.contains("2016")]
accident_data_revised = accident_data_revised[~accident_data_revised["Start_Time"].str.contains("2
017")]
accident_data_revised = accident_data_revised[~accident_data_revised["Start_Time"].str.contains("2
019")]

# Splits the Start_Time column into individual columns
accident_data_revised[["Date", "Time"]] = accident_data_revised["Start_Time"].str.split(expand=Tru
e)

# Renames columns to be more readable
accident_data_revised = accident_data_revised.rename(columns={"Start_Lat":"Lat", "Start_Lng":"Ln
g", "Weather_Condition":"Weather"})

# Filters and rearranges dataset to display most useful columns
accident_data_revised = accident_data_revised[["Date", "Time", "Lat", "Lng", "City", "State",
         "County", "Weather", "Temperature(F)", "Severity", "Nautical_Twilight"]]

# Iterates through all column data to find NaN values and adds them to a list
accident_data_revised.columns[accident_data_revised.isna().any()].tolist()

accident_data_revised = accident_data_revised.reset_index()

accident_data_revised.head(20)
```

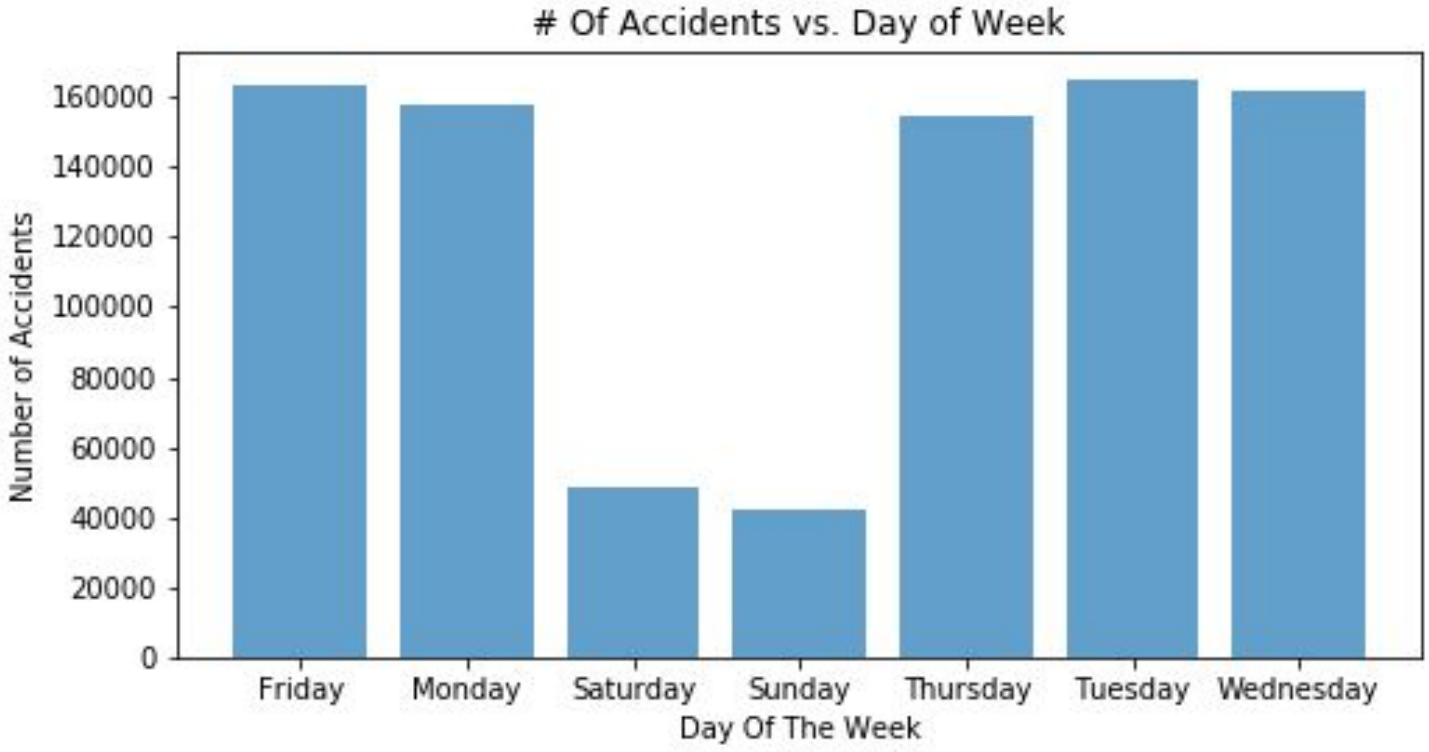
What day of the week are traffic accidents most prevalent?

Date
2018-12-31

Day of Week
Monday

```
#Creating a day of the week column with current time stamps
import datetime
weekday = []
for d in range(0, len(df["Date"])):
    dt = pd.Timestamp(df["Date"][d])
    day_ofwk = dt.weekday_name
    weekday.append(day_ofwk)
new_df = pd.DataFrame({"Day of Week": weekday})
df.insert(1,"Day of Week",new_df["Day of Week"],True)
df.head()
```

What day of the week are traffic accidents most prevalent?

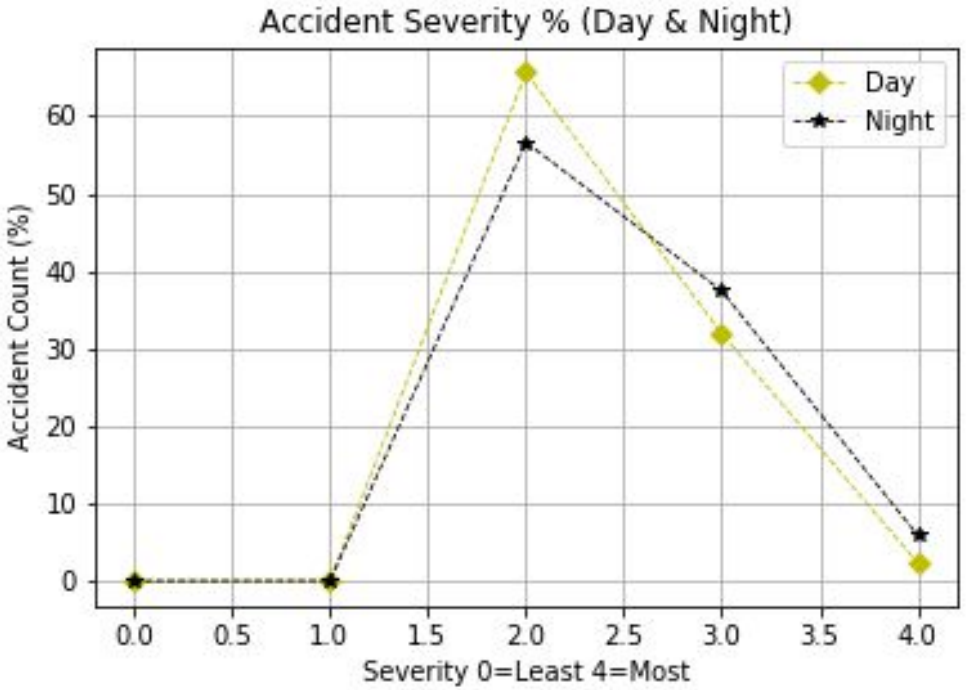


What time of day is traffic most severe?

Total Accidents 2018

Day: 758,753

Night: 133,861



What locations nationally and locally have a greater frequency of traffic accidents?

In creating a heatmap to answer this question, we needed a weight. For this question, the “weight” should be frequency of accidents, so we added a column to the dataframe showing frequency of accidents at specific coordinates.

In [18]:
```
# Add column to dataframe showing frequency of accidents at specific coordinates
accident_data_revised['frequency'] = accident_data_revised['Coordinates'].map(accident_data_revise
d['Coordinates'].value_counts())
accident_data_revised.head()
```

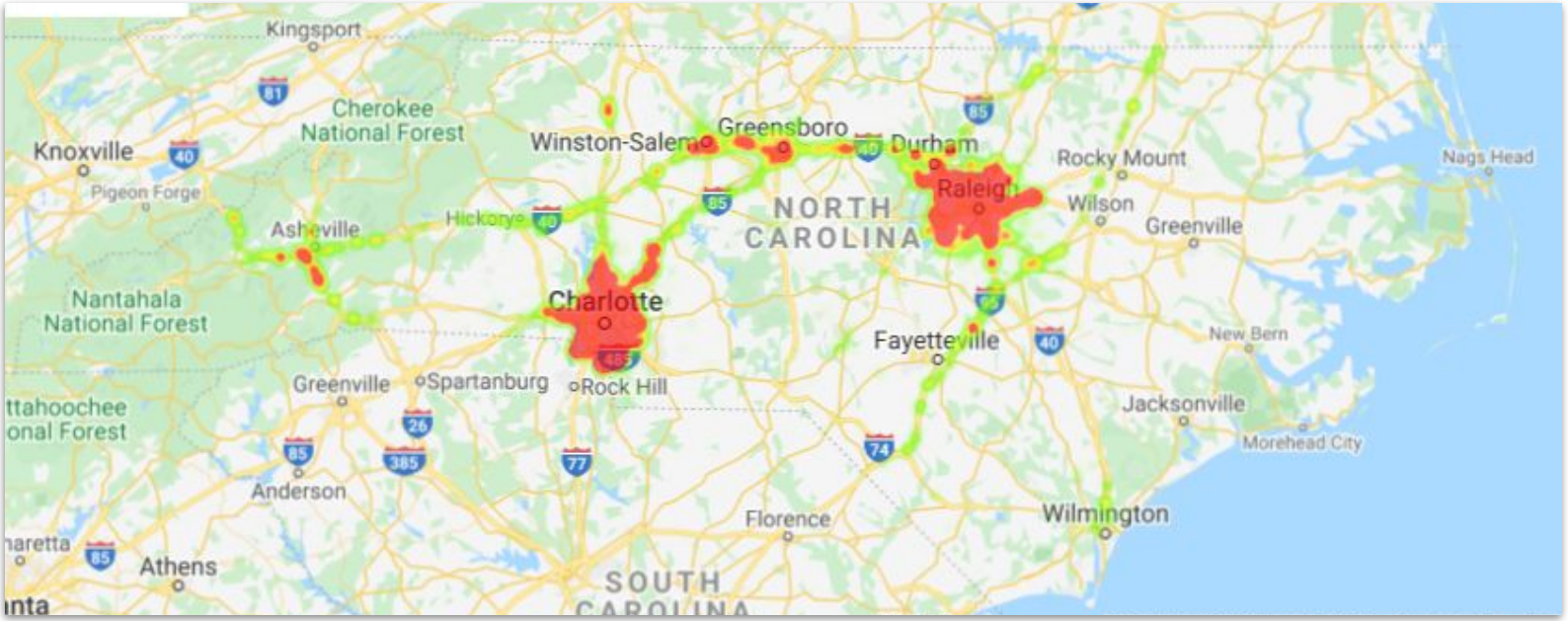
Then we grouped coordinates by accident frequency and narrowed the DataFrame to show unique coordinates and include the frequency of their traffic accidents.

In [19]:
```
# Group coordinates by accident frequency
accident_data_sorted = accident_data_revised.sort_values(by ='frequency' , ascending=False)
accident_data_sorted.head()
```

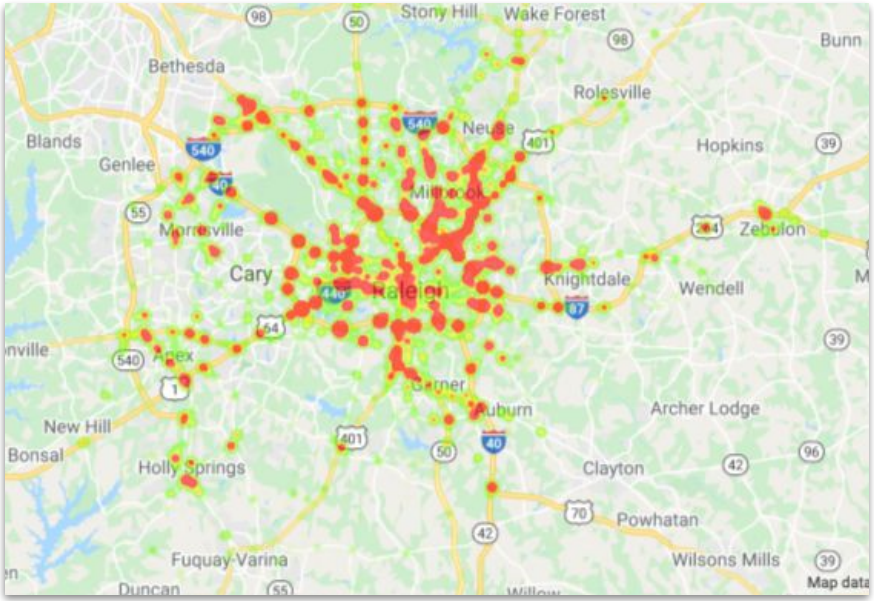
In [20]:
```
# Find unique coordinates per frequency (this will delete some rows of data)
coordinates_by_frequency = accident_data_sorted.drop_duplicates(subset=['Coordinates'])
coordinates_by_frequency.head()
```

What locations nationally have a greater frequency of traffic accidents?

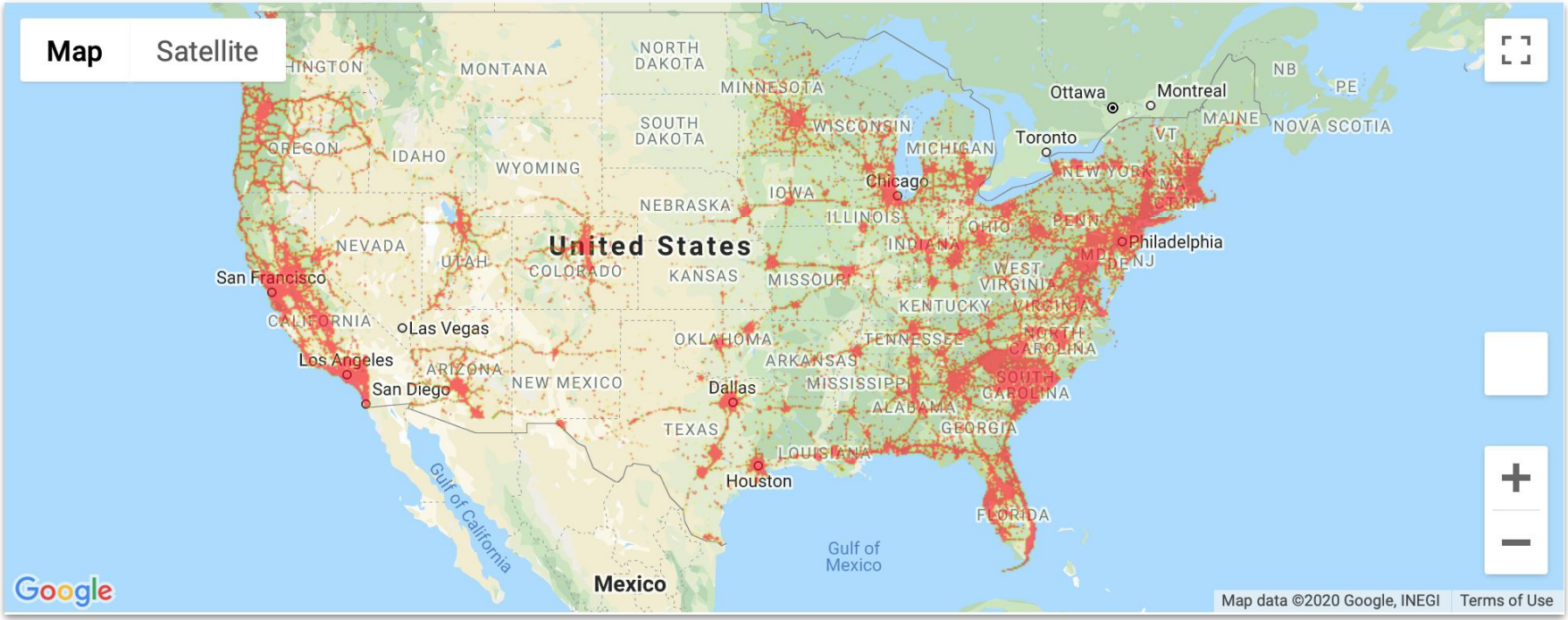
What locations locally have a greater frequency of traffic accidents?



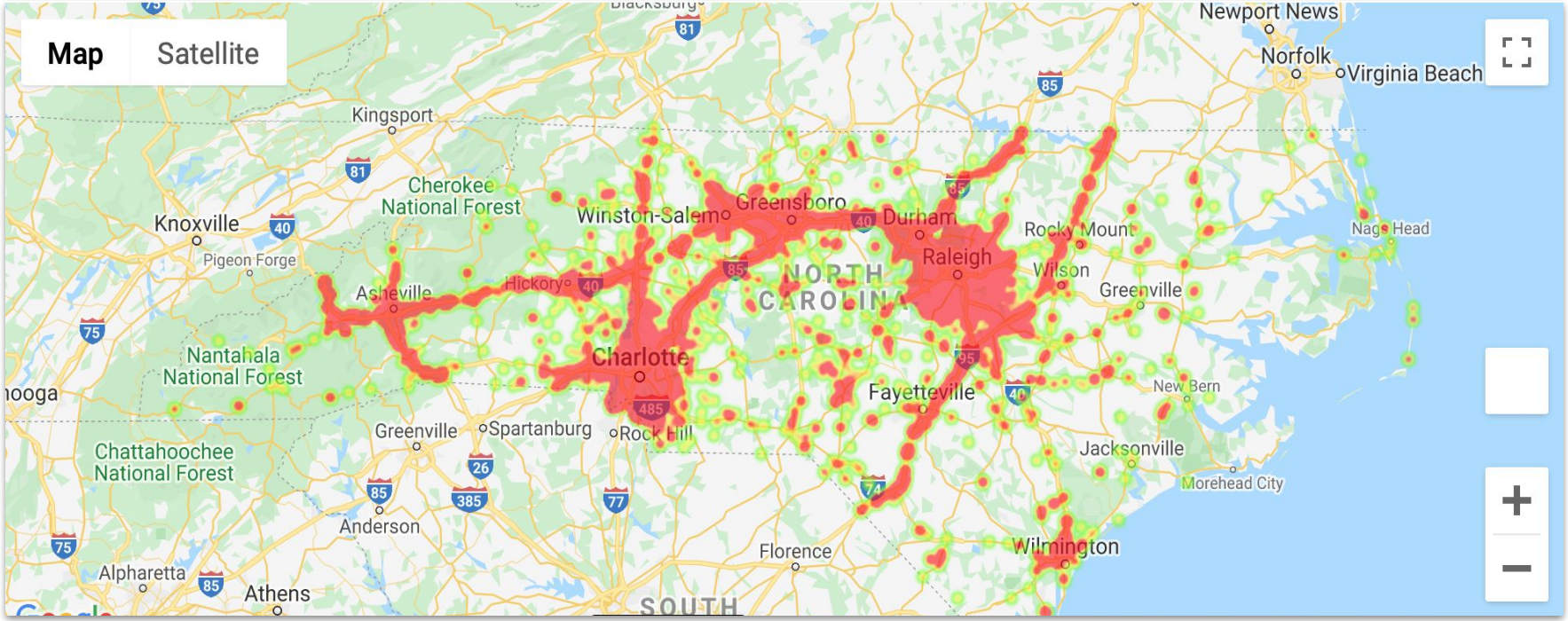
What locations locally have a greater frequency of traffic accidents?



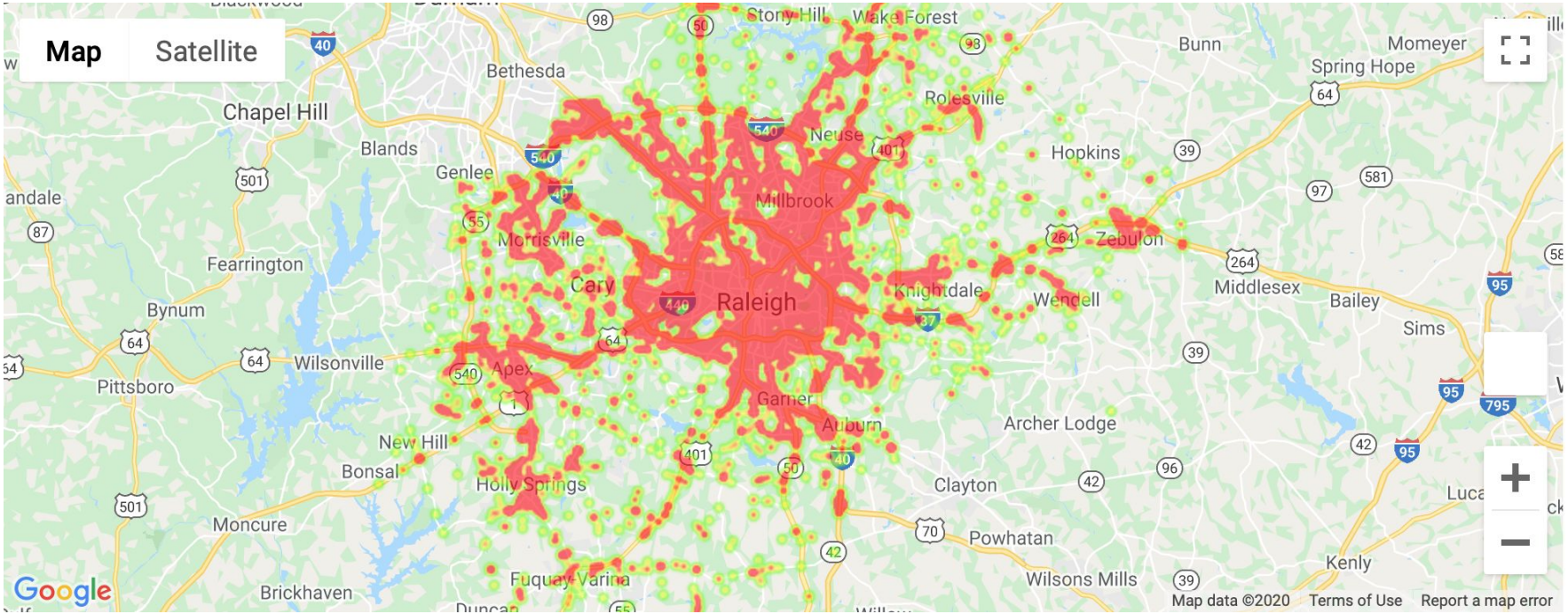
What locations nationally have a greater severity of traffic accidents?



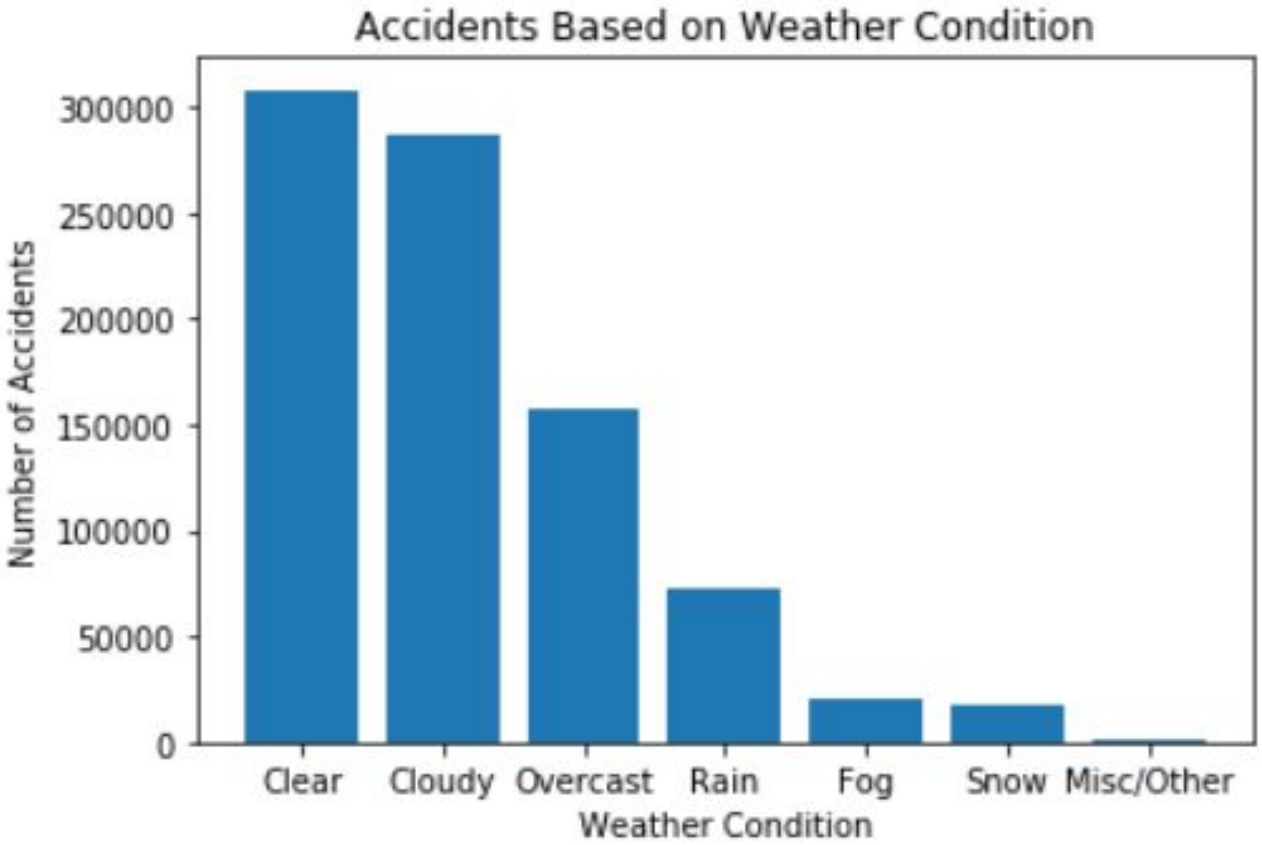
What locations at the state level have a greater severity of traffic accidents?

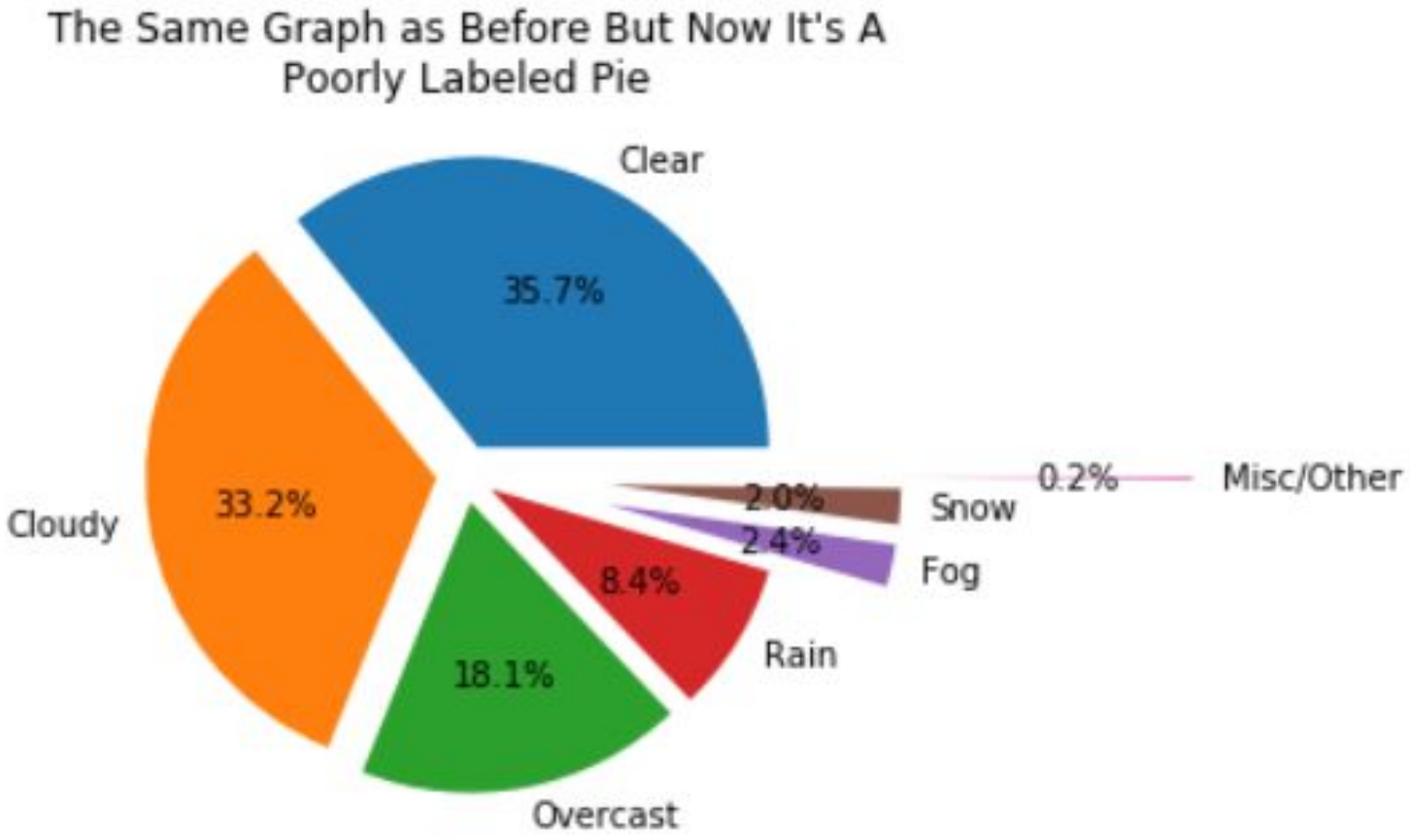


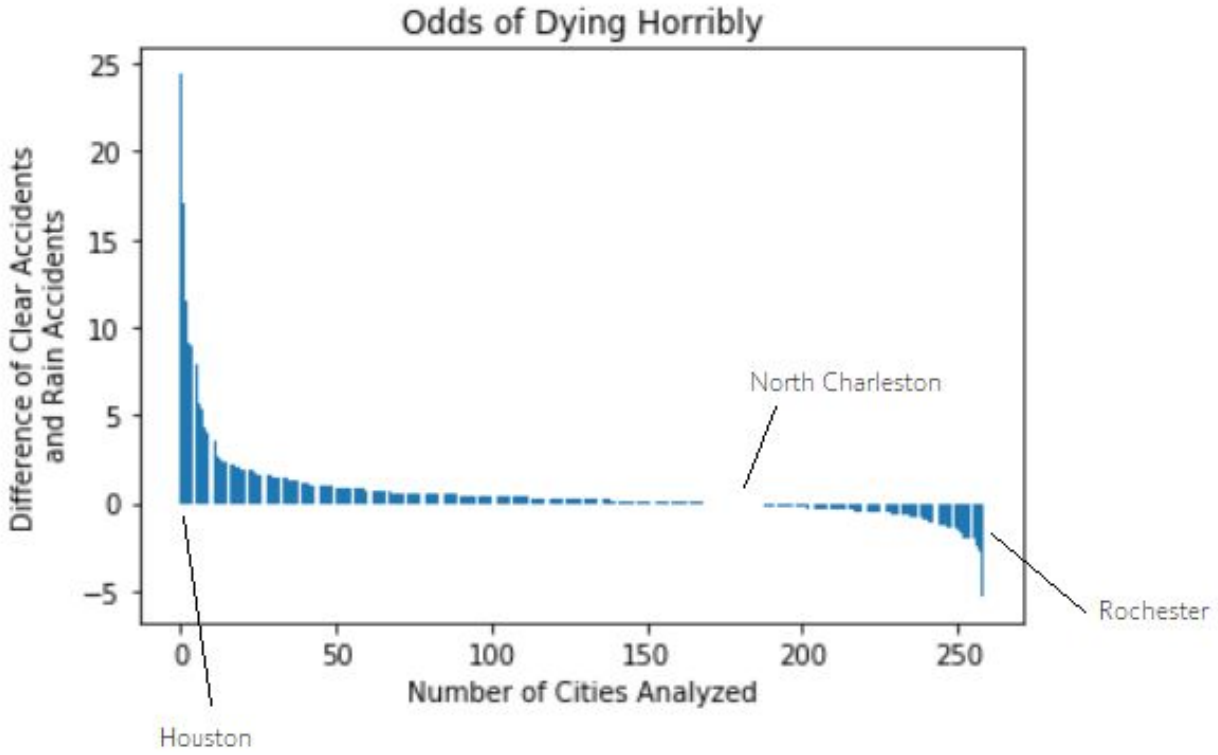
What locations locally have a greater severity of traffic accidents?



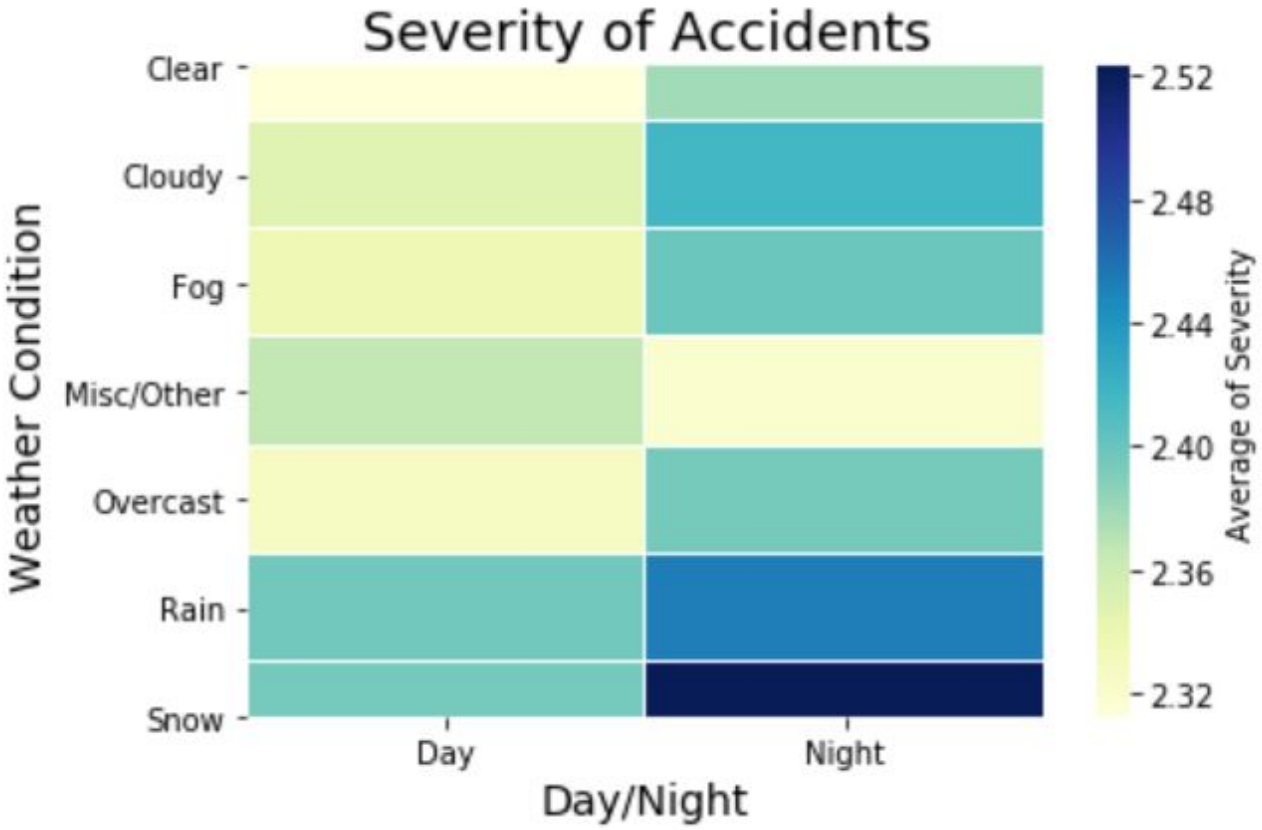
Do weather conditions impact the number of accidents reported?


Accidents Based on Weather Condition
Number of Accidents
300000
250000
200000
150000
100000
50000
0
Clear
Cloudy
Overcast
Rain
Fog
Snow
Misc/Other
Weather Condition


The Same Graph as Before But Now It's A Poorly Labeled Pie
Clear
35.7%
Cloudy
33.2%
Overcast
18.1%
Rain
8.4%
Fog
2.4%
Snow
2.0%
Misc/Other
0.2%


Odds of Dying Horribly
Difference of Clear Accidents and Rain Accidents
25
20
15
10
5
0
−5
0
50
100
150
200
250
Number of Cities Analyzed
North Charleston
Houston
Rochester

Severity of Accidents Grouped By Weather Condition



Questions?