THE EFFECTS OF WEATHER AND TIME ON TRAFFIC ACCIDENTS

Joel Meinhardt, Mike Terkhorn,

Brad Vawter and Cindy Ford

Motivation and Summary

Hypothesis and Questions

HYPOTHESES:

- Inclement weather increases traffic accidents
- Accident rates are higher at certain times of the day (i.e. Rush Hour)

QUESTIONS:

1. What is the correlation between weather conditions and traffic accidents in Indiana?
2. What time of day do most accidents occur?
3. What weather conditions affect severity of accidents?
4. Where do most accidents occur within a geographical area?

Why:

To demonstrate the most dangerous areas, what time of day is the most likely to have an accident and if weather conditions contributed significantly to the number of accidents or the severity.

Summary

Summary:

We were able to answer the majority of the questions, however we did not find 2020 data.

1. What is the correlation between weather conditions and traffic accidents in Indiana?
   - There were fewer accidents recorded during inclement weather conditions vs. the fair weather conditions (cloudy being the highest rate of accidents)
2. What time of day do most accidents occur?
   - The highest incidents of traffic accidents in IN was during the morning rush hour (7am-9am) and then there was a significant decrease after 9am.
3. Do weather conditions have an effect on the number or severity of accidents?
   - Weather did not have an impact on the number of accidents, but did have an impact on the severity of the accidents, with snow having the highest average severity
4. Where do most accidents occur within a geographical area?
   - Within IN the accidents occurred most frequently along major highways and interstates
   - Nationally, 8 of the top twenty counties with the highest incidents of accidents per population were in South Carolina

Questions and Data

- We were able to find a dataset that contained accidents, severity, weather, latitude, longitude, city, state and multiple other data points for accidents across the country from 2016 to 2019.
  - https://www.kaggle.com/sobhanmoosavi/us-accidents
- Utilized census data to normalize the accidents to geographic areas by population
  - https://www.census.gov/data/datasets/2010/dec/rdo/section-203-determination-pums.html

Exploration and Cleanup

Data Cleanup and Exploration

Exploration

- Searched for a dataset that had sufficient data, various measurable factors, and was recent. Ideally, we would have wanted to see 2020 data; however it wasn't available
  - The dataset used had ~3M lines and ~50 columns
- Expecting to see a larger impact of weather on accidents, but discovered that the most incidents occurred during fairer weather
- Expected to be able to locate more free APIs surrounding historical traffic data and weather
- Originally, we would have liked to see how the stay-at-home order was effecting current traffic data; however 2020 data was not available, so we had to modify our questions

```
file.columns
```

```
Index(['ID', 'Source', 'TMC', 'Severity', 'Start_Time', 'End_Time',
       'Start_Lat', 'Start_Lng', 'End_Lat', 'End_Lng', 'Distance(mi)',
       'Description', 'Number', 'Street', 'Side', 'City', 'County', 'State',
       'Zipcode', 'Country', 'Timezone', 'Airport_Code', 'Weather_Timestamp',
       'Temperature(F)', 'Wind_Chill(F)', 'Humidity(%)', 'Pressure(in)',
       'Visibility(mi)', 'Wind_Direction', 'Wind_Speed(mph)',
       'Precipitation(in)', 'Weather_Condition', 'Amenity', 'Bump', 'Crossing',
       'Give_Way', 'Junction', 'No_Exit', 'Railway', 'Roundabout', 'Station',
       'Stop', 'Traffic_Calming', 'Traffic_Signal', 'Turning_Loop',
       'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight',
       'Astronomical_Twilight'],
      dtype='object')
```

Data Cleanup and Exploration (cont.)

Data Cleanup

- Many of the columns in the dataset were unrelated to the questions, so those columns had to be removed to better view the data
- There was approximately 3M lines of data. Initially we narrowed it down to just the data from IN.
- We encountered a problem with the weather descriptions not being standardized.
  - For example, there were conditions such as “Thunderstorms” and “T-Storms” and “Thunder” that were all combined under one condition.
  - We had around 50 types and consolidated to 8 types by using .map
- Census data changed from state name to state abbreviation using a dictionary in order to make the datasets consistent

Cleanup Code

Weather Condition Mapping

```
#Combine weather conditions by similar labels
file_small["Weather"] = file_small["Weather_Condition"].map({"Clear":"Clear",
                                                            "Mostly Cloudy": "Cloudy",
                                                            "Partly Cloudy": "Cloudy",
                                                            "Scattered Clouds": "Cloudy",
                                                            "Cloudy / Windy": "Cloudy",
                                                            "Overcast" : "Cloudy",
                                                            "Fair" : "Fair",
                                                            "Light Rain": "Rainy",
                                                            "Cloudy" : "Cloudy",
                                                            "Light Snow" : "Snow",
                                                            "Rain" : "Rainy",
                                                            "Light Drizzle" : "Rainy",
                                                            "Fog" : "Foggy",
                                                            "Heavy Rain" : "Rainy",
                                                            "Haze" : "Foggy",
                                                            "Snow" : "Snow",
                                                            "Light Thunderstorm and Rain" : "Storms",
                                                            "Mist" : "Rainy",
                                                            "Mostly Cloudy / Windy" : "Cloudy",
                                                            "T-Storm": "Storms",
                                                            "Light Freezing Rain" : "Icy",
                                                            "Patches of Fog" : "Foggy",
                                                            "Thunderstorm" : "Storms",
                                                            "Thunderstorms and Rain": "Storms",
                                                            "Heavy Thunderstorms and Rain": "Storms",
                                                            "Drizzle" : "Rainy",
                                                            "Light Freezing Drizzle" : "Icy",
                                                            "Fair / Windy" : "Fair",
                                                            "Thunder in the Vicinity" : "Storms",
                                                            "Light Snow / Windy" : "Snow",
                                                            "Light Freezing Fog" : "Icy",
                                                            "Partly Cloudy / Windy": "Cloudy",
                                                            "Light Rain / Windy" : "Rainy",
                                                            "Heavy T-Storm" : "Storms",
                                                            "Light Rain with Thunder" : "Storms",
                                                            "Shallow Fog" : "Foggy",
                                                            "Heavy Snow" : "Snow",
                                                            "Wintery Mix" : "Snow",
                                                            "Light Ice Pellets": "Icy",
                                                            "Blowing Snow" : "Snow",
                                                            "Heavy Drizzle": "Rainy",
                                                            "Ice Pellets" : "Icy",
                                                            "Thunder": "Storms"
                                                           })
```

Renaming a column to be consistent across data sets

```
cen_file_df["County, State"] =  cen_file_df["County"] + ", " + cen_file_df["a"]
cen_df = cen_file_df.drop(columns=["County","State","a"])
cen_df["County, State"] = cen_df["County, State"].str.replace("Parish","County")
```

Dropping columns to make the data more readable

```
file_small = file.drop(columns = ["End_Lat", "End_Lng",
                                  "Country", "Number",
                                  "TMC", "Distance(mi)",
                                  "Street", "Side",
                                  'Timezone', 'Airport_Code',
                                  'Weather_Timestamp', 'Wind_Direction',
                                  'Wind_Speed(mph)', 'Amenity',
                                  'Bump', 'Crossing',
                                  'Give_Way','Junction',
                                  'No_Exit', 'Railway',
                                  'Roundabout', 'Station',
                                  'Stop', 'Traffic_Calming',
                                  'Traffic_Signal', 'Turning_Loop',
                                  'Sunrise_Sunset', 'Civil_Twilight',
                                  'Nautical_Twilight', 'Astronomical_Twilight',
                                  'Wind_Chill(F)','Humidity(%)',
                                  'Pressure(in)','Visibility(mi)', 'Precipitation(in)'])
```

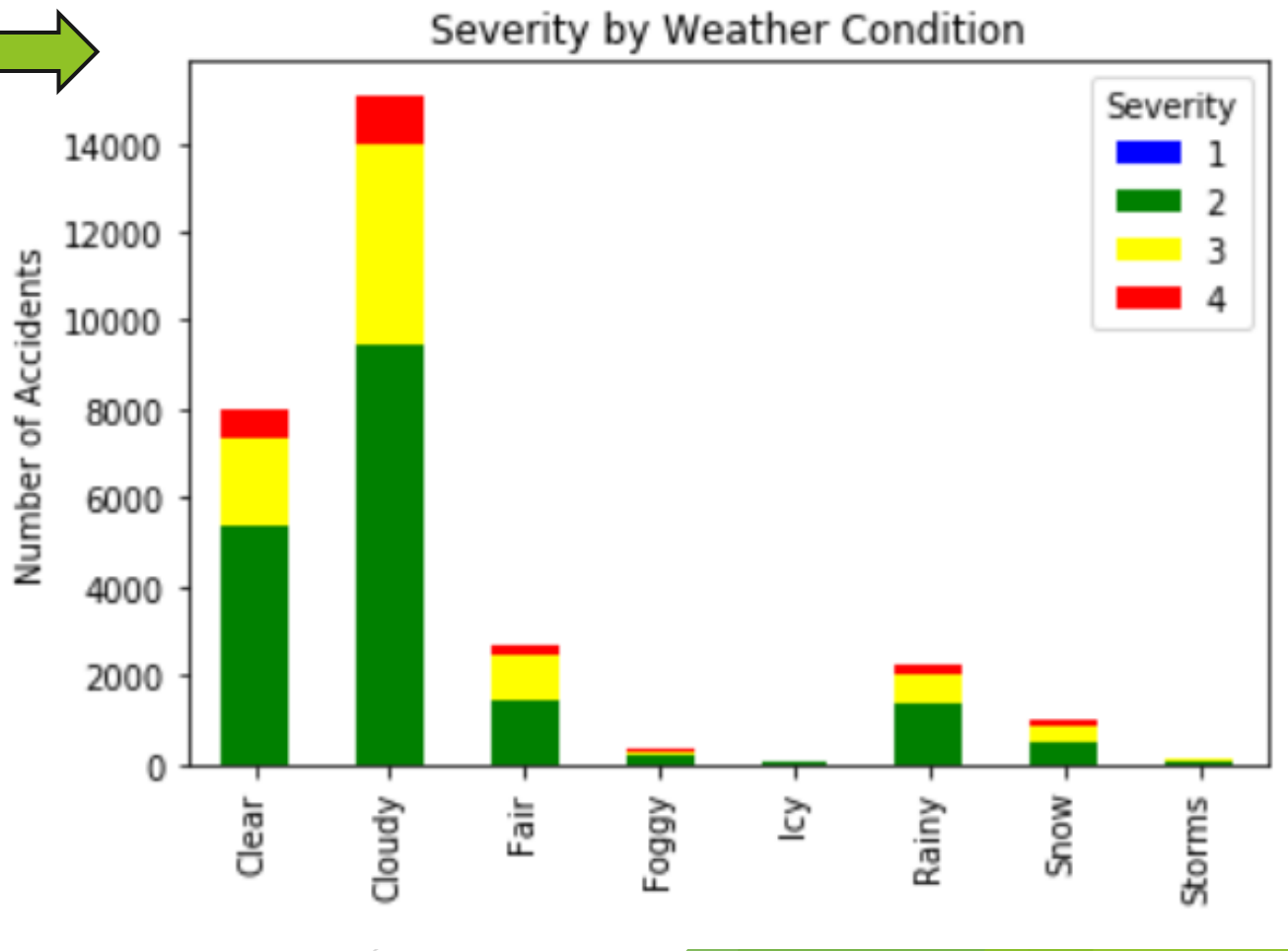
Data Analysis

Severity by Weather Condition

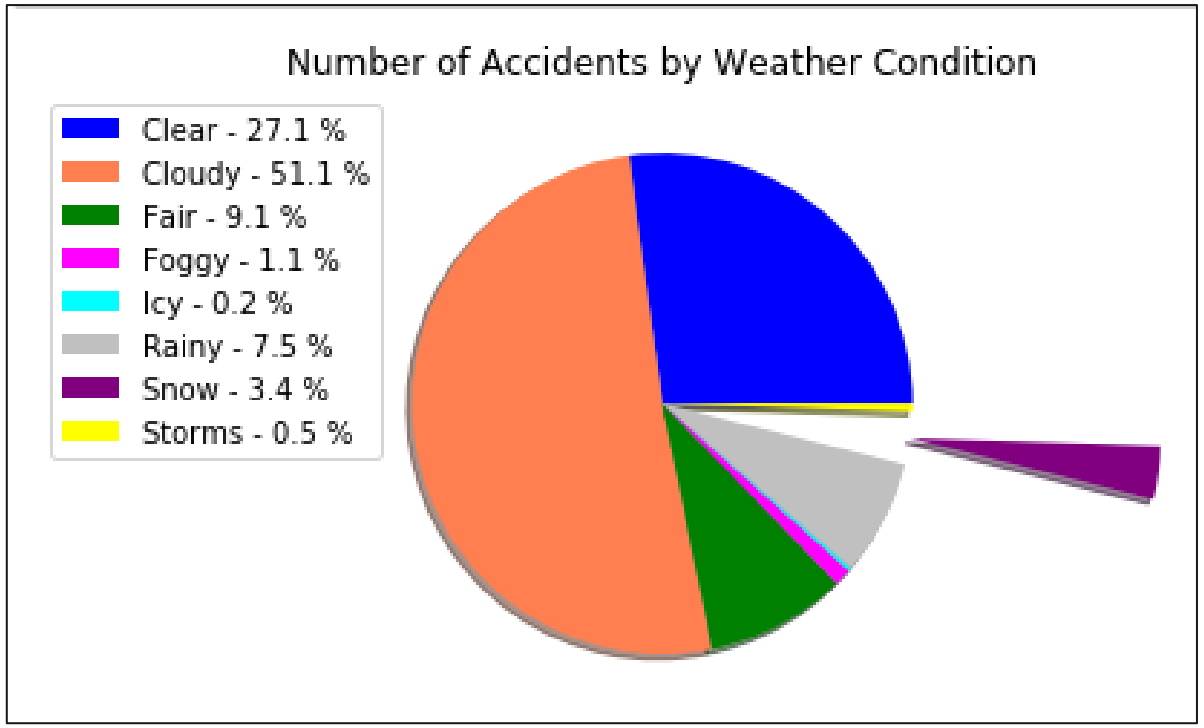
	Weather	Severity	ID
0	Clear	1	5
1	Clear	2	5347
2	Clear	3	1977
3	Clear	4	689
4	Cloudy	1	6
5	Cloudy	2	9474
6	Cloudy	3	4483
7	Cloudy	4	1129
8	Fair	1	2
9	Fair	2	1438

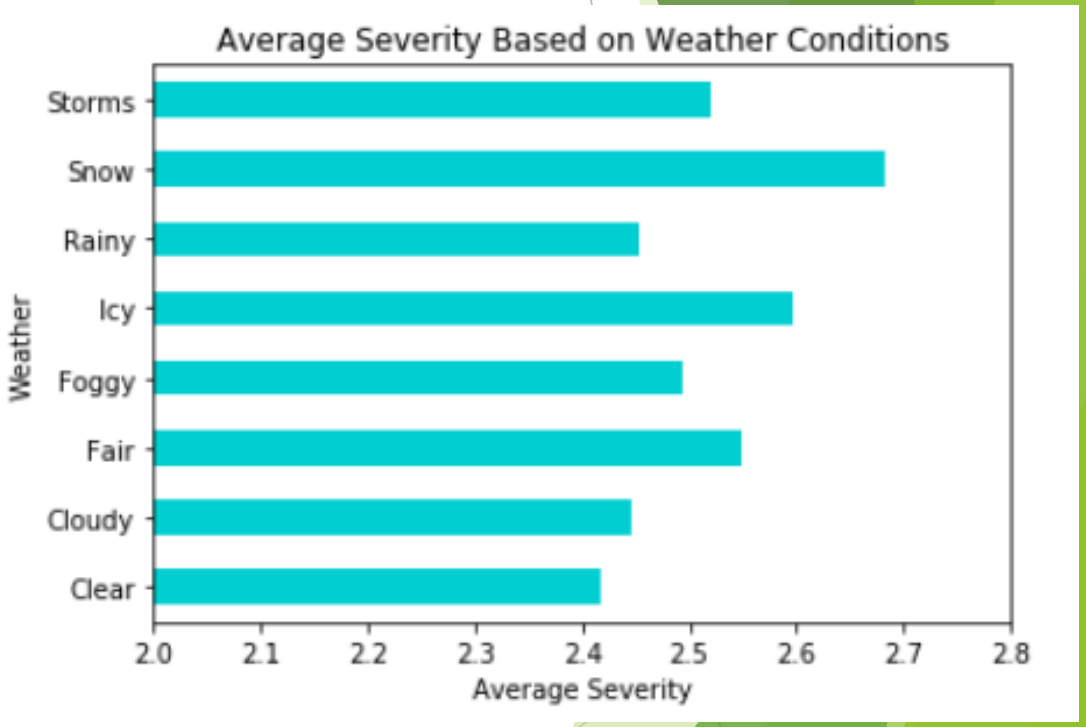
Severity	1	2	3	4
Weather				
Clear	5.0	5347.0	1977.0	689.0
Cloudy	6.0	9474.0	4483.0	1129.0
Fair	2.0	1438.0	1015.0	229.0
Foggy	1.0	202.0	81.0	40.0
Icy	1.0	35.0	21.0	10.0
Rainy	2.0	1393.0	634.0	186.0
Snow	0.0	477.0	375.0	157.0
Storms	0.0	79.0	58.0	9.0

```
pivot2_df = weather_gb4.pivot(index = "Weather", columns ="Severity", values = "ID")
pivot2_df = pivot2_df.fillna(0)
pivot2_df
```

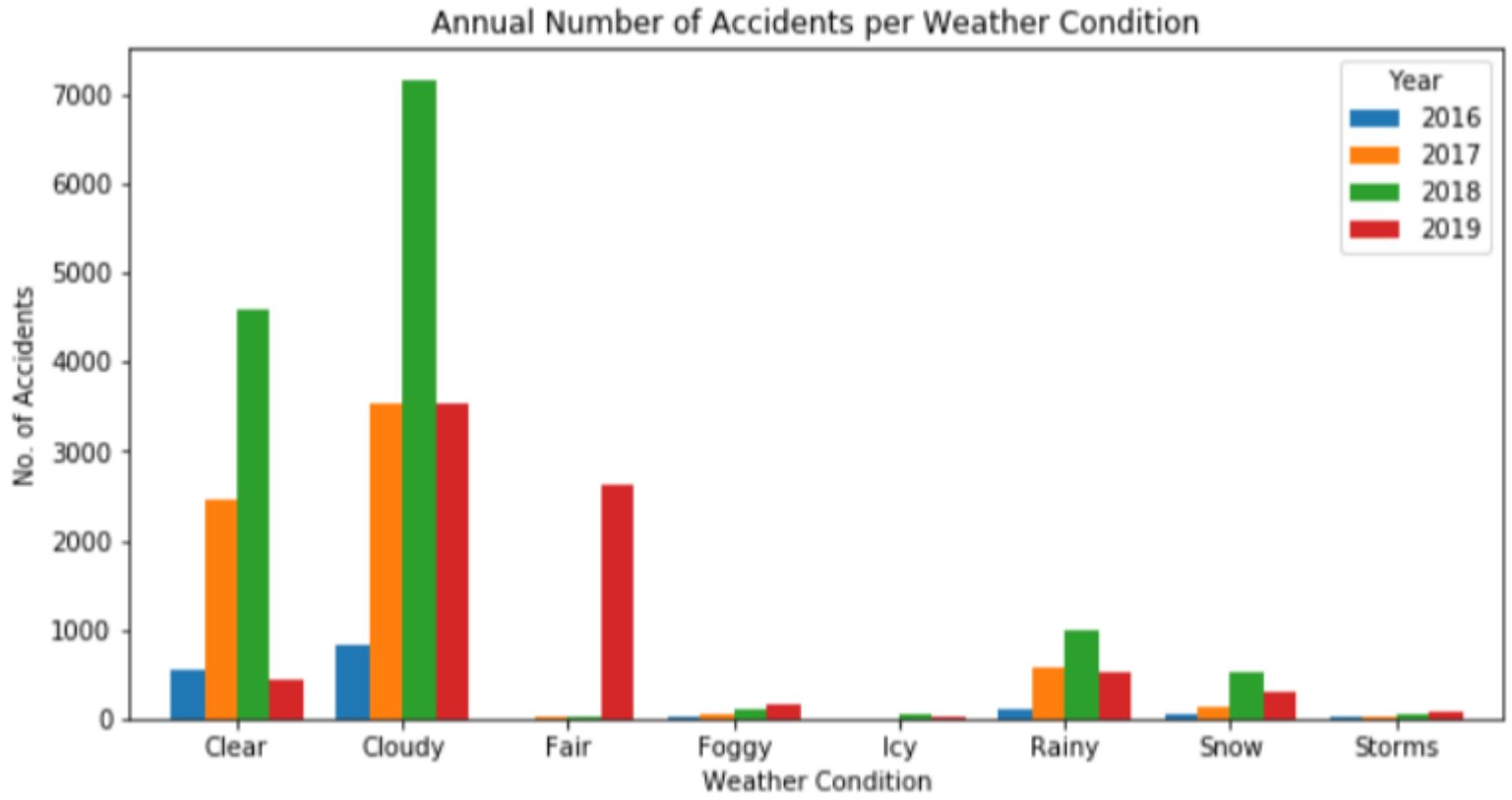


Correlation between weather conditions and traffic accidents in Indiana



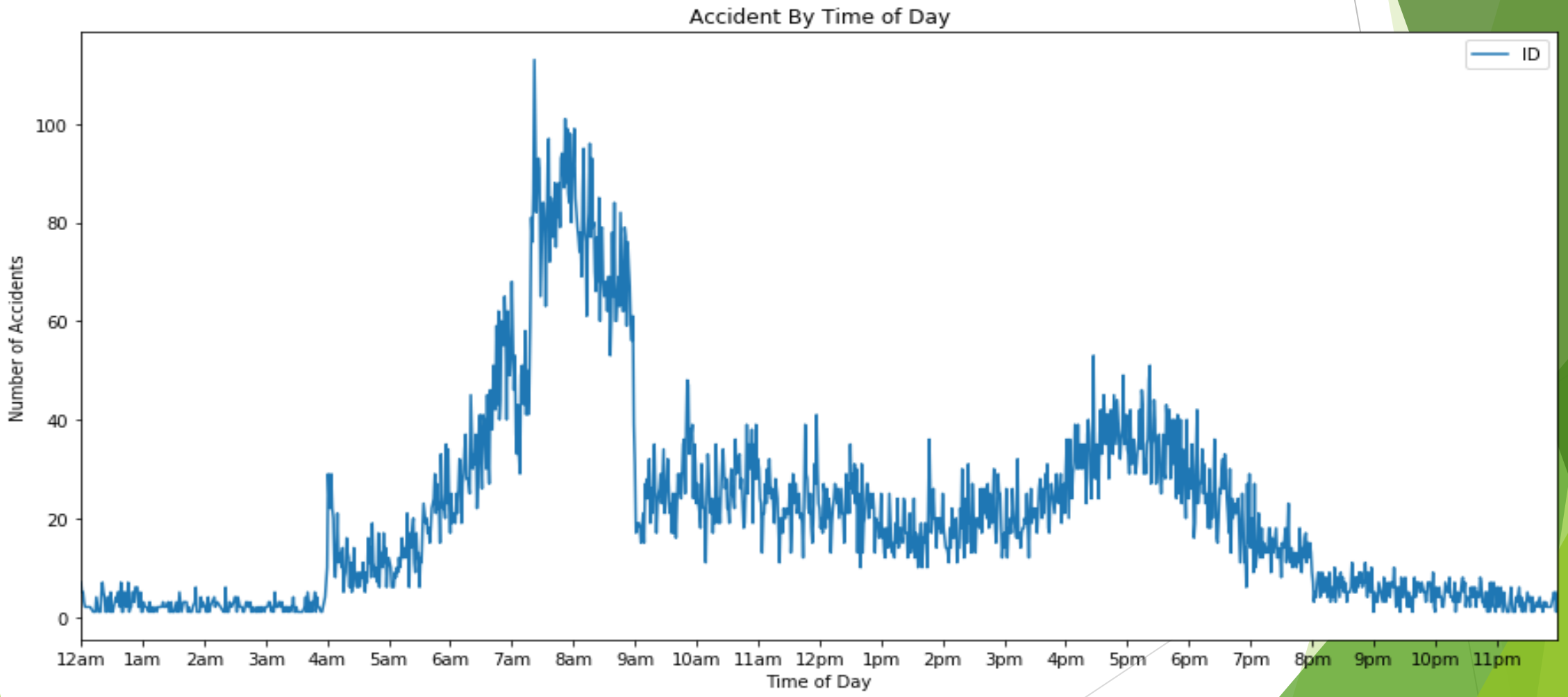


Data Analysis

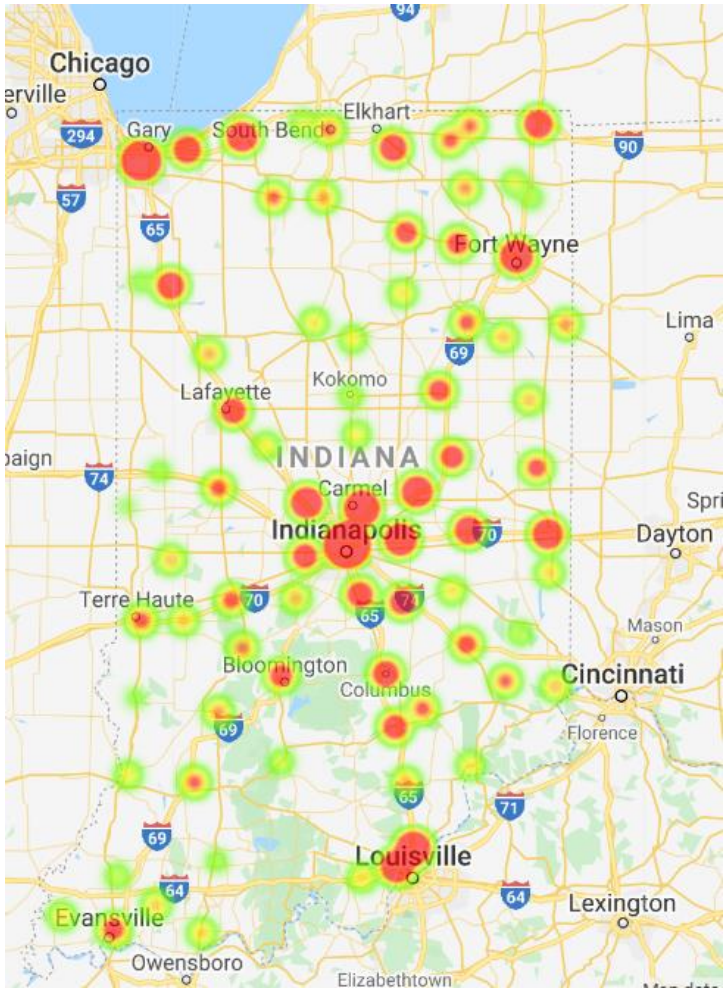


Data Analysis

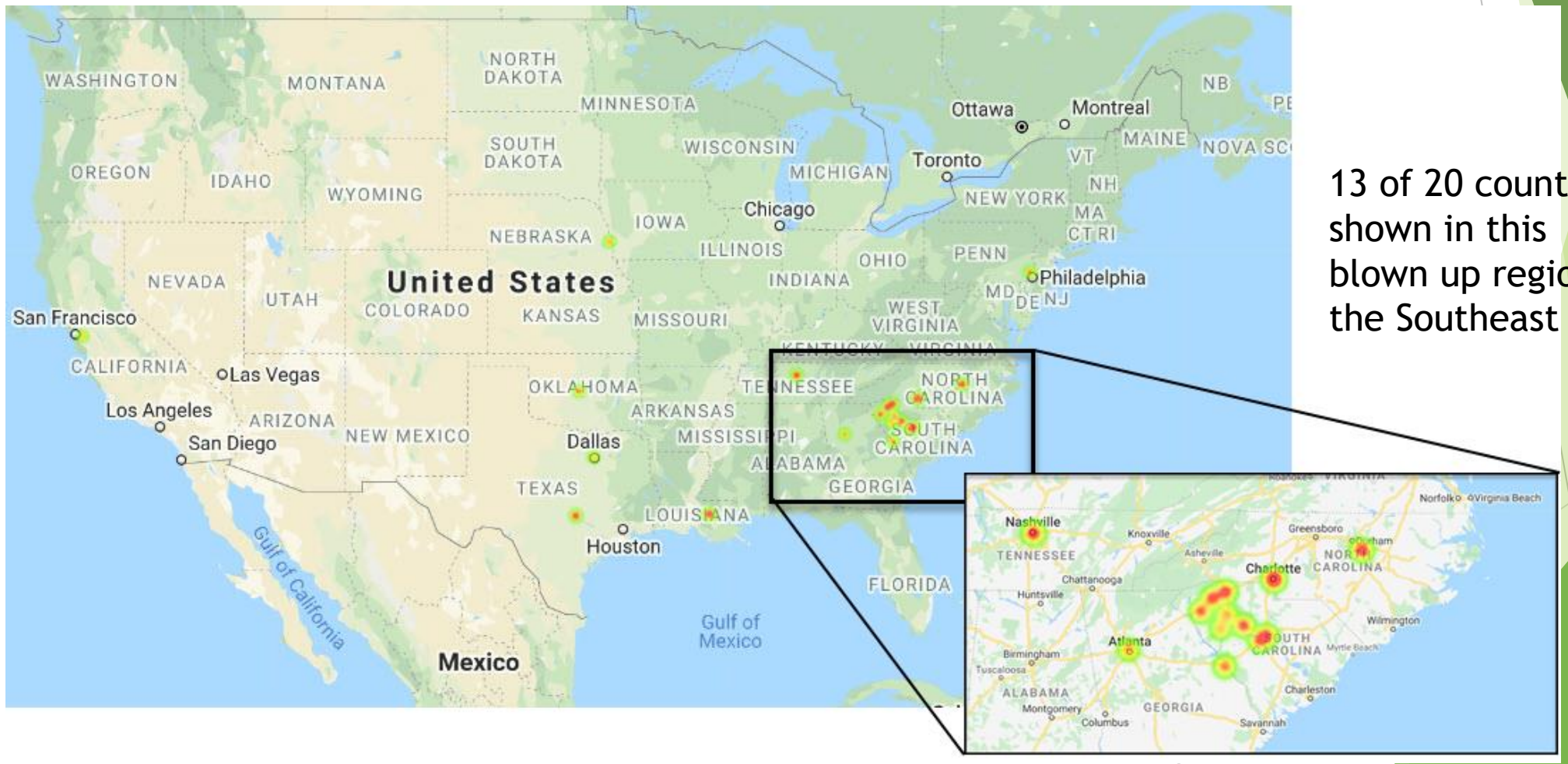
What time of day do most accidents occur?



Heat map of Accident magnitude for each county



Top 20 counties nationally for accidents per population



13 of 20 counties shown in this blown up region in the Southeast

Discussion and Postmortem

Discussion

- Originally wanted more recent data especially in an API, but we couldn't find one that had current and historical data
- In Indiana, the majority of the accidents occurred along major interstates and highways
- The weather did not significantly effect the number of accidents that occurred, but it could contribute to increased severity
- Most accidents occurred during cloudy weather in IN
- The highest incidents of accidents occurred during the morning and evening rush hours with a greater spike during the morning

Postmortem

Things we were not expecting:

- Normalized accident rates by population were not mostly in higher population density areas or traffic corridors
- 2018 had significantly more accidents than any of the other years and it would be interesting to discover why that might have been

Things we struggled with:

- Getting the state and county data aligned between the two data sets
- Extracting time from the data/time format utilized since it was all in one cell

If more time and current data were available:

- We would want to be able to do a state by state look to compare the weather conditions and highest incident times across all states
- Does the time of year effect the number of morning rush hour accidents due to sunrise time
- We would have liked to been able to see the effects of the travel bans and stay-at-home orders

Questions?