Car Accidents

Michael Seidenberg
Kelby Wilson
Niama Bagaga
Jipeng Yan

Our Data and Approach

- Data set from Kaggle.com
  https://www.kaggle.com/sobhanmoosavi/us-accidents

- Huge data set of about 3 million rows and 49 columns, 1GB
  - Very fun to load ;)

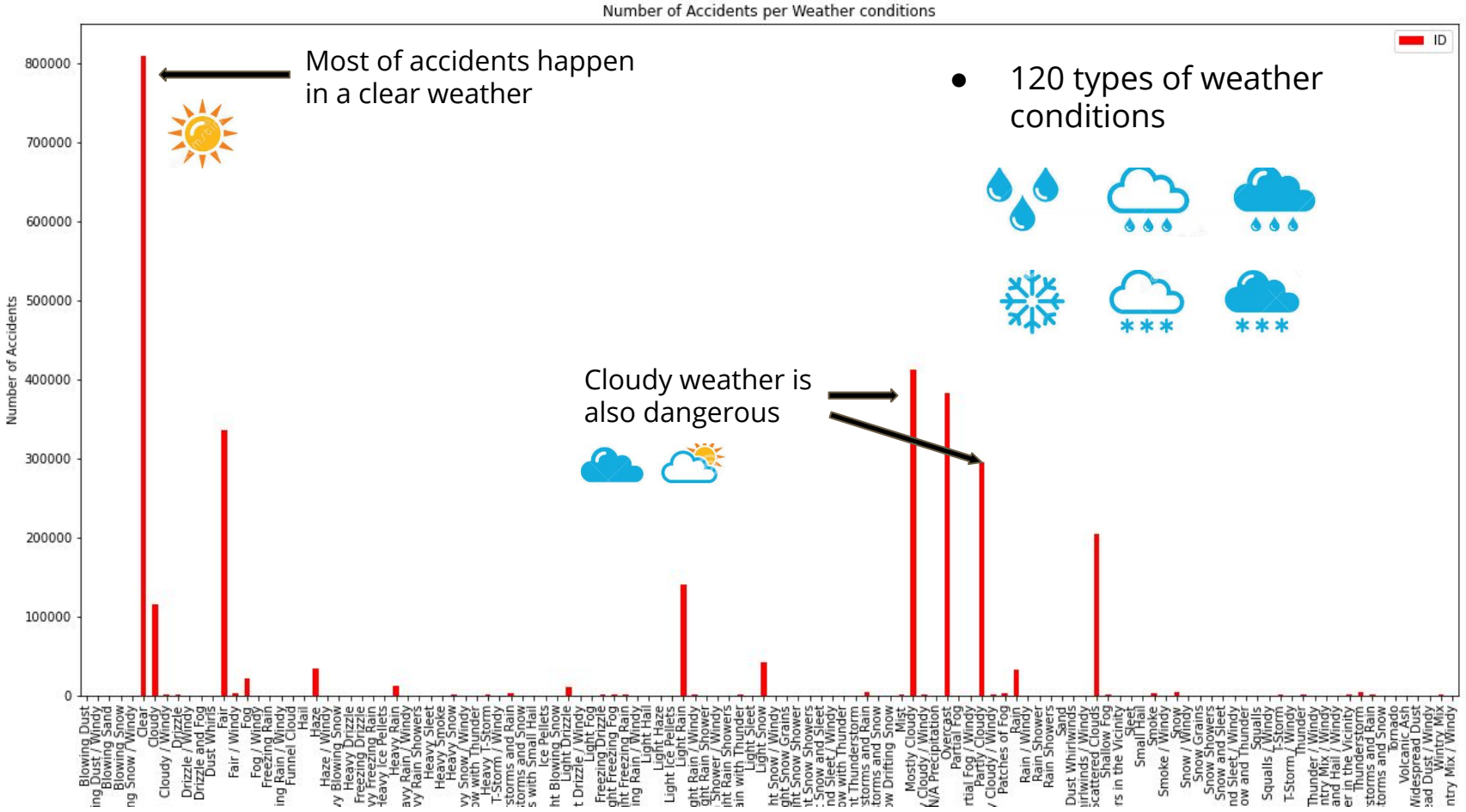
- Divide and conquer
  - Jipeng and Kelby- Weather; Niama- Time; Michael- Location

Our Overall Hypothesis and Motivation

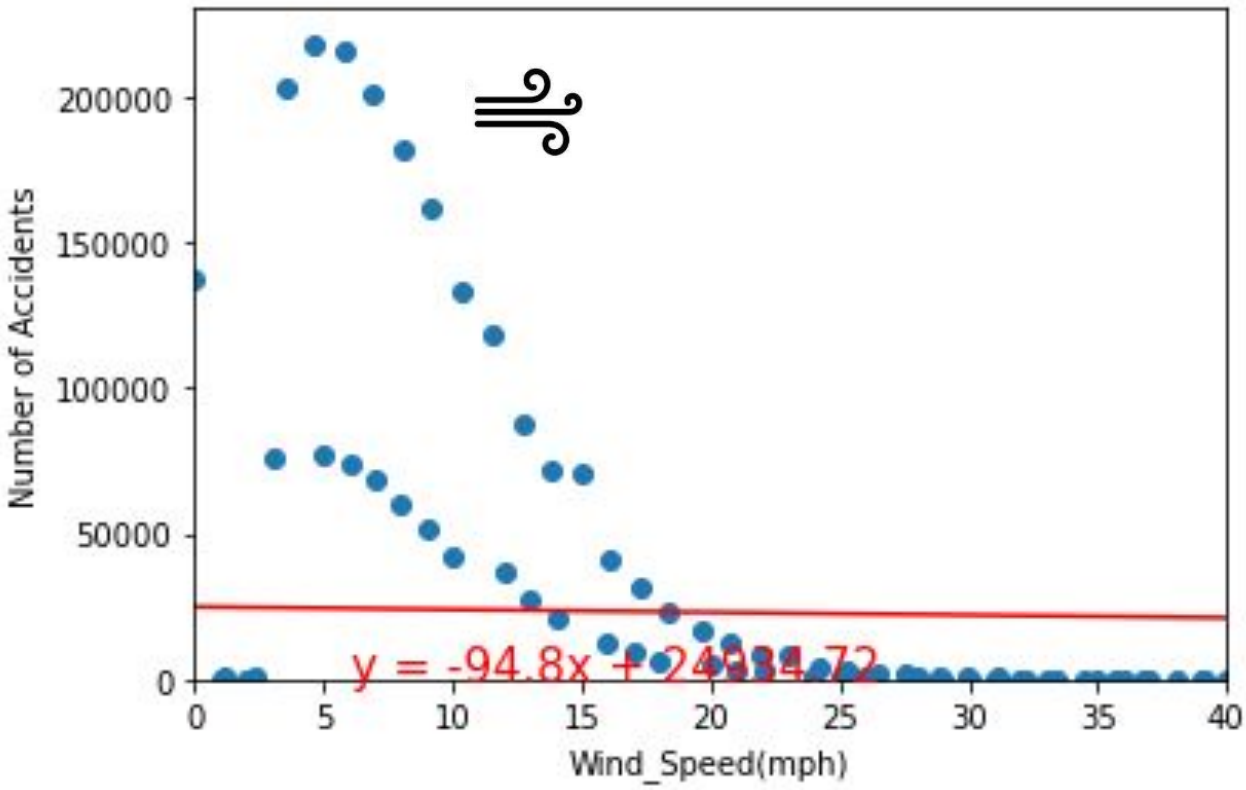
Weather, time and location have an impact on accident likelihood and severity.

Motivated to learn if we should be more careful driving in certain areas, weather conditions, or times of day

Weather conditions vs. Number of Accidents

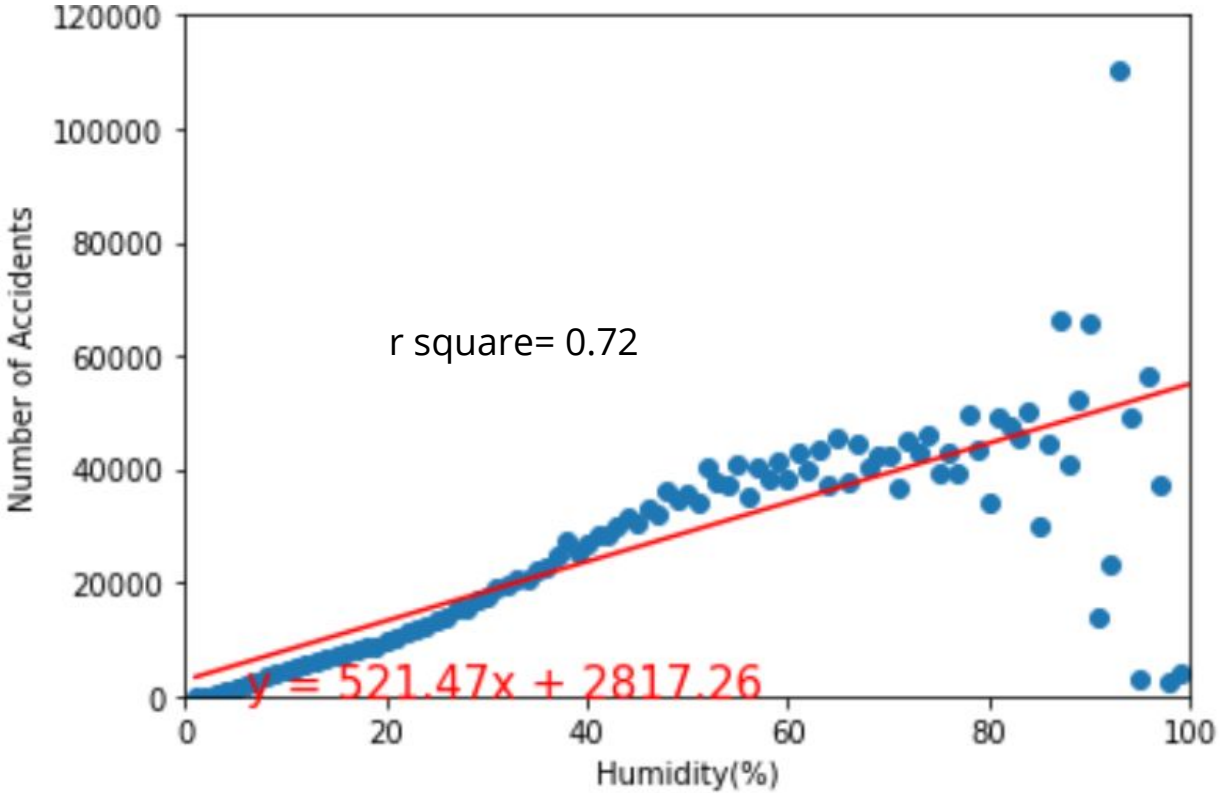


Wind Speed vs. Number of Accidents



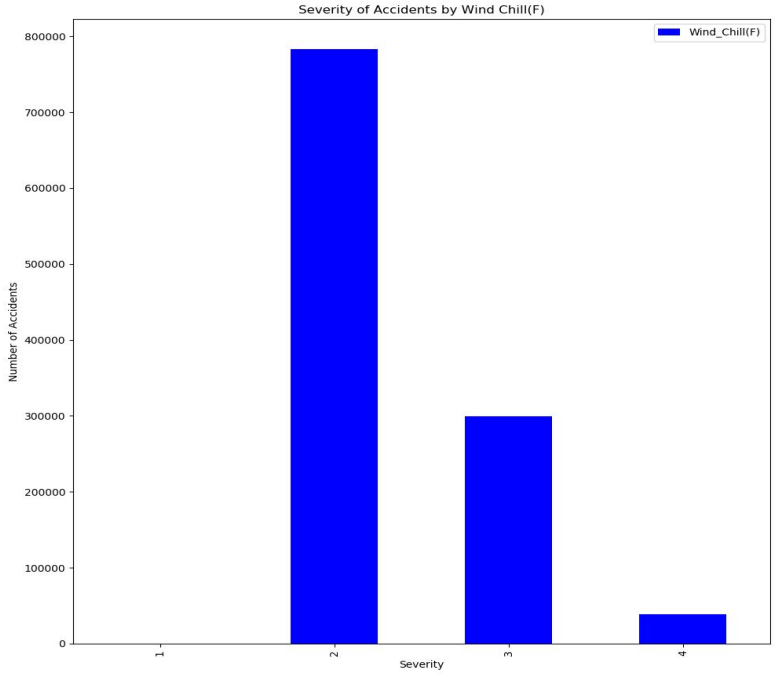
- No clear relationship between wind speed and number of accidents

Humidity vs. Number of Accidents



- Strong positive correlation between humidity and number of accidents

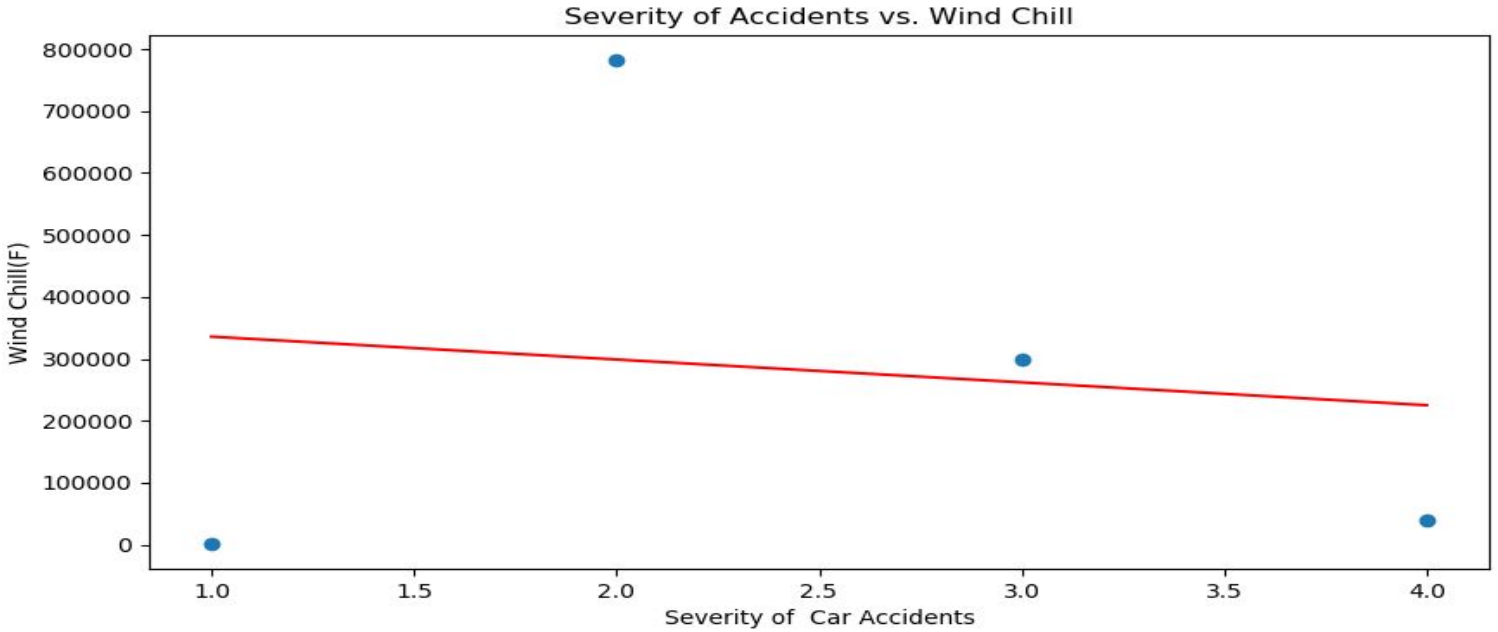
Severity of Accidents by Wind Chill(F)



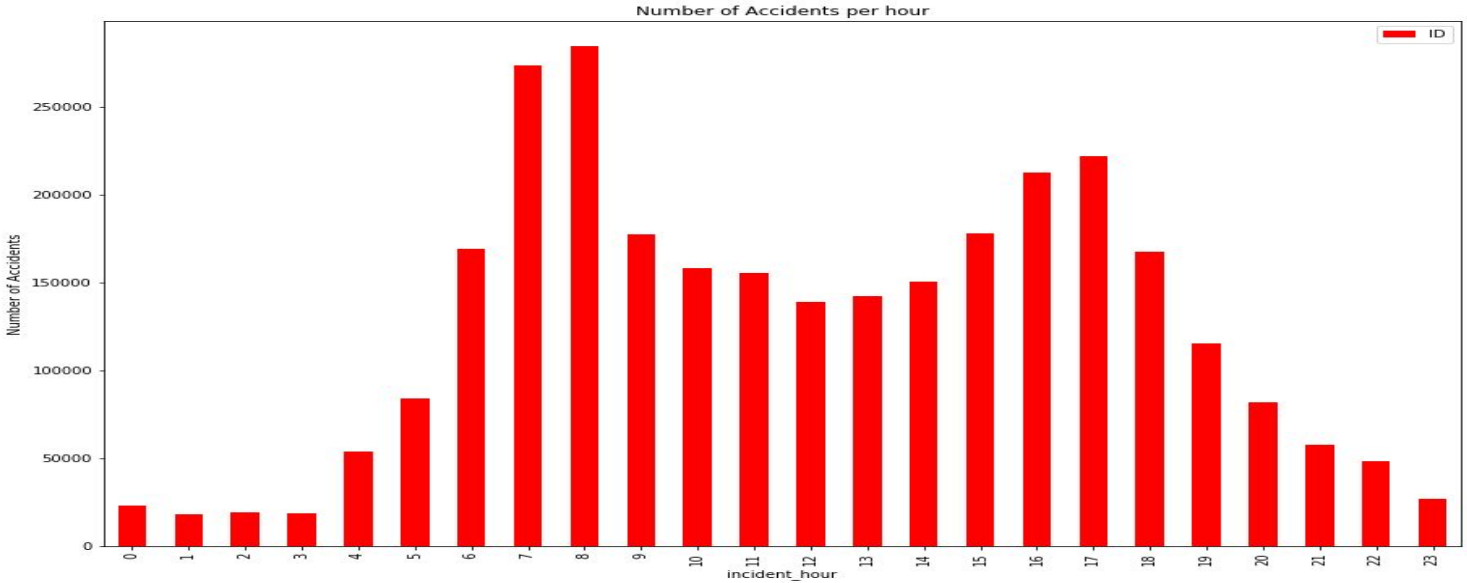
Most accidents were of severity 2, the second to lowest measure.

Linear Regression of Severity vs. Wind Chill

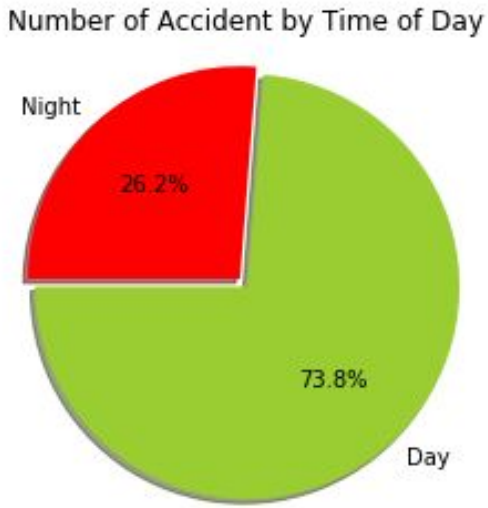
The r-squared is: -0.13207633815224823. This is a very weak negative correlation.

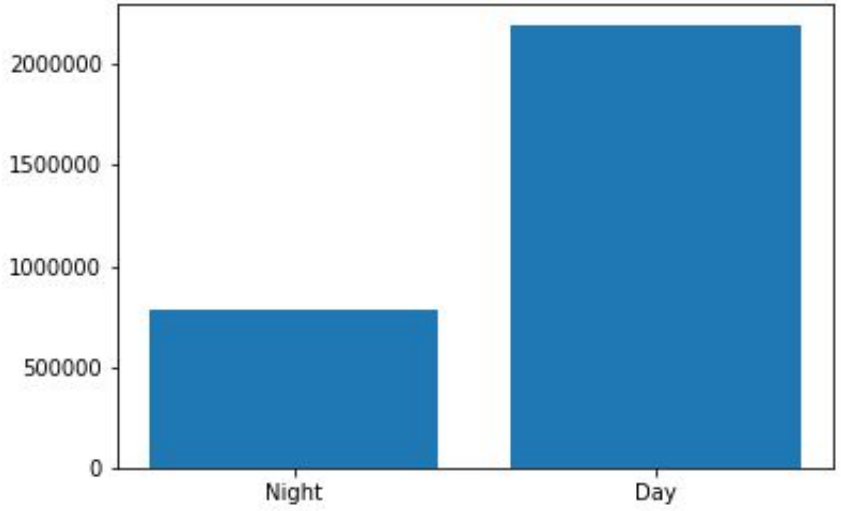


Number of accidents per hour throughout the day

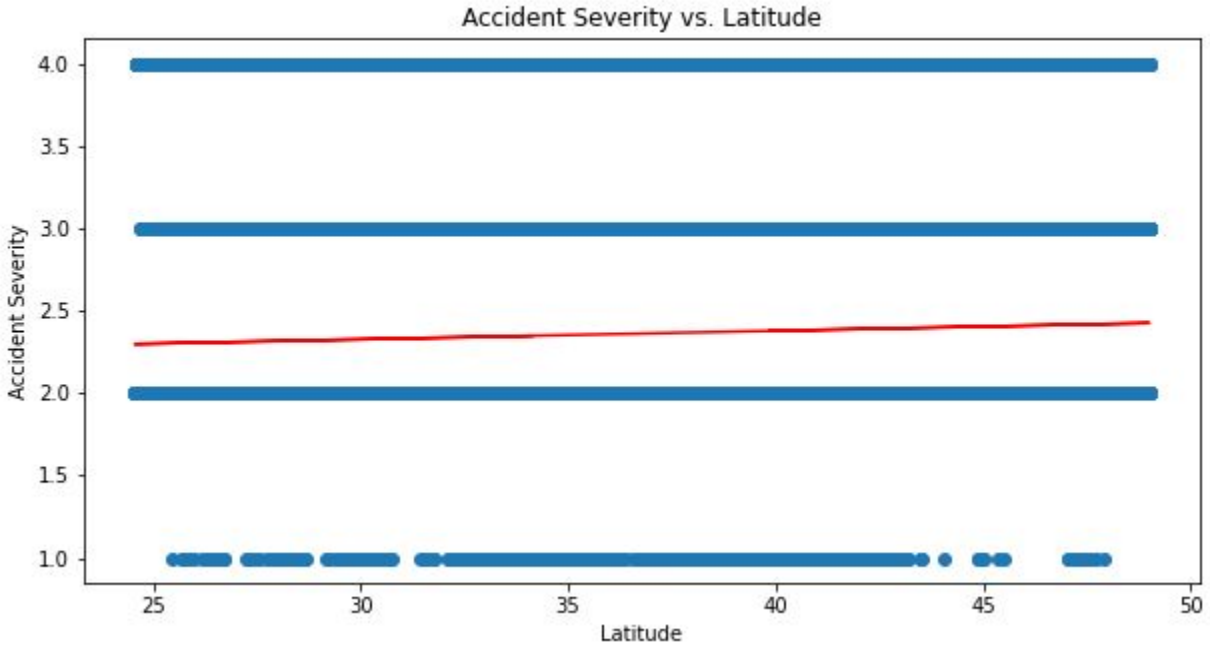


The different level of accident between Night and Day

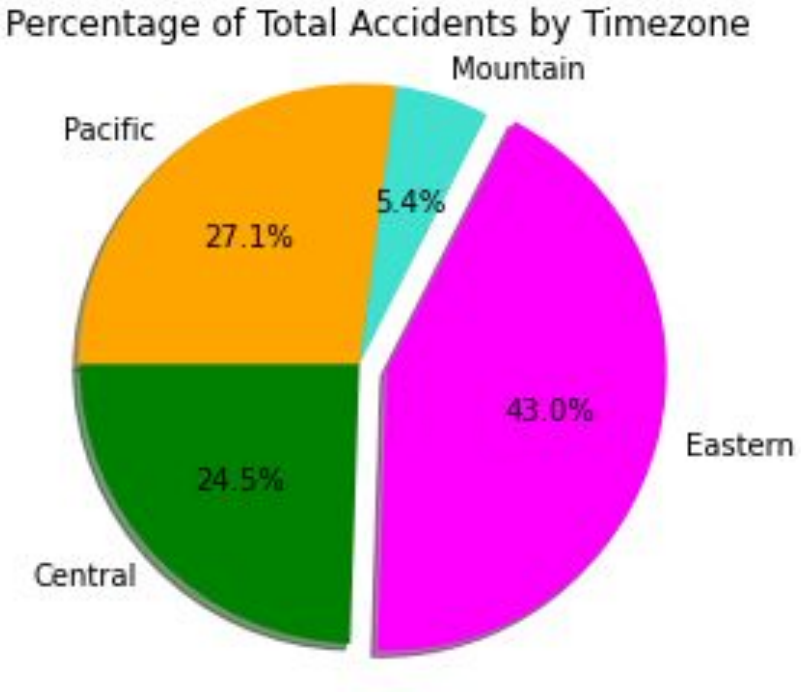




Accidents are more severe in certain Latitudes...NOT!
r-squared = 0.04, No correlation at all

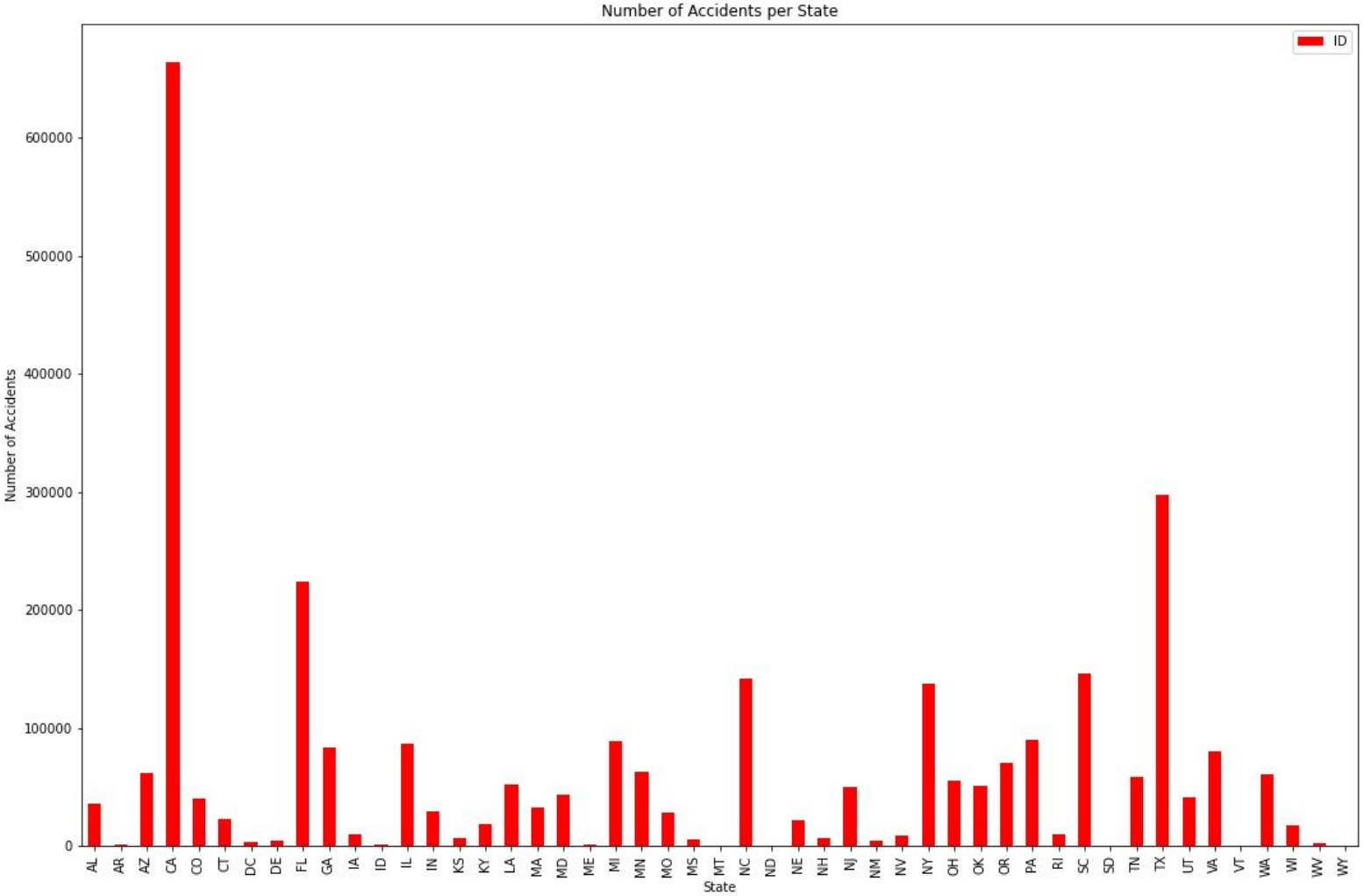


More accidents happen in more populated areas of the US

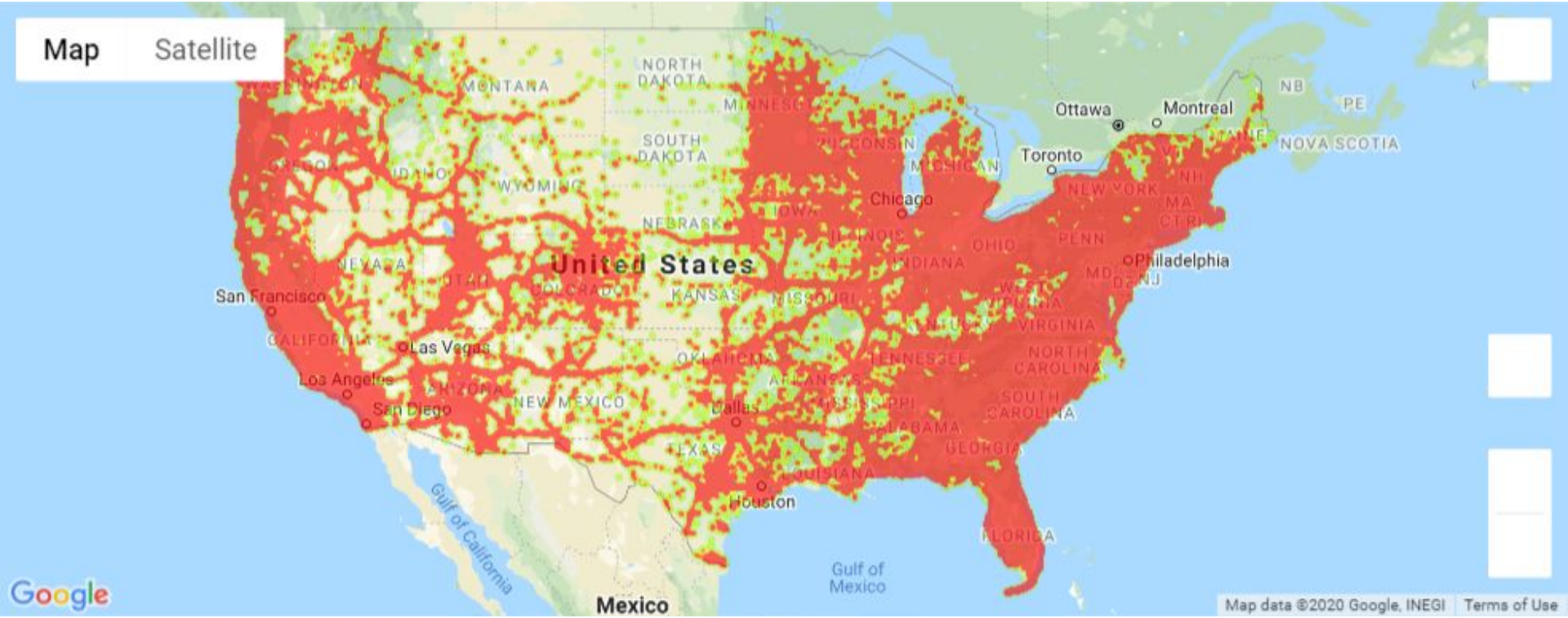


More accidents happen in more populated areas of the US.

SC is the exception.



It’s getting hot in here…



More cars on the road = more accidents

- Most common weather conditions
- Times of day with the most drivers
- Areas with the highest population