Analysis of United States Traffic Fatalities

Alejandro Escobar, Ryan Mahtab, Amy Tsai, Eric Zhou

Carnegie Mellon University Department of Statistics and Data Science

Introduction

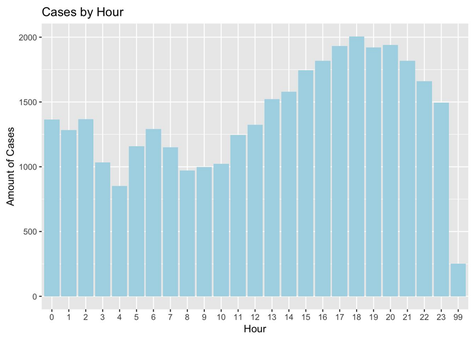
The Fatality Analysis Reporting System (FARS) contains data on a census of fatal traffic crashes within the 50 states, D.C, and Puerto Rico. There is information on over 989,451 motor vehicle fatalities and over 100 different coded data elements that characterize the crash, vehicle, and people involved. In the dataset, accident_2016, there are 52 columns, and 34,748 rows. Each observation corresponds to a different car accident. The columns include variables describing the time and location of the accident, the number of fatalities, drunk drivers, weather conditions, and other factors that describe the situation of the crash.

Time of Accidents

The cases by hour graph shows the amount of cases in the hour of day they occurred.

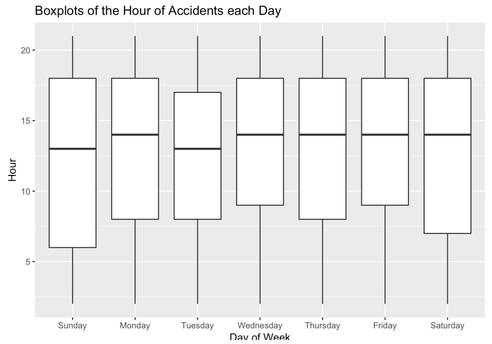
As one can see, the graph is skewed slightly right, with peak cases occurring at hours 17-20, with a noticeable uptick during rush hour.

Accidents are more likely to occur with more vehicles on the road.

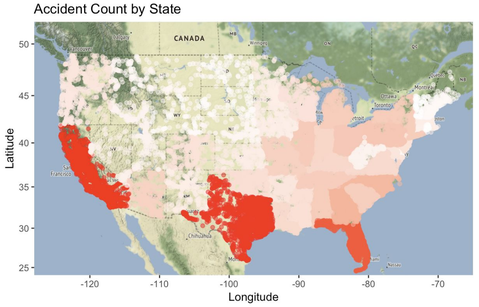


The boxplots show the distribution of accidents each hour of the day for each day of the week.

It is interesting to see that the distributions are mostly similar across the days of the week. The medians do not differ by more than 3 hours.



Accident Frequency



This map displays the frequencies of accidents in 2016 by state. Each point is a particular traffic accident. States that are more red have higher overall accident totals, while states that are more white have lower totals.

State Trends

The word cloud displays the states with the most fatalities in traffic accidents.

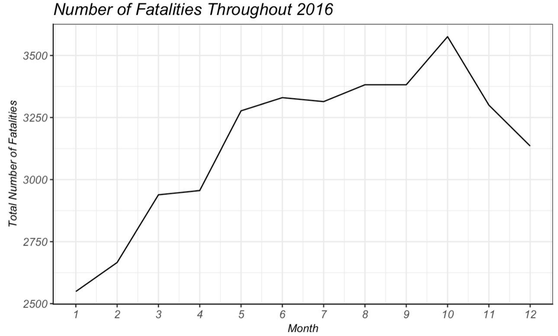
We can see that California, Florida, and Texas have the most accidents, judging by the size of the words.

This could also be attributed to the fact that these states have large populations.

Variations Over Time

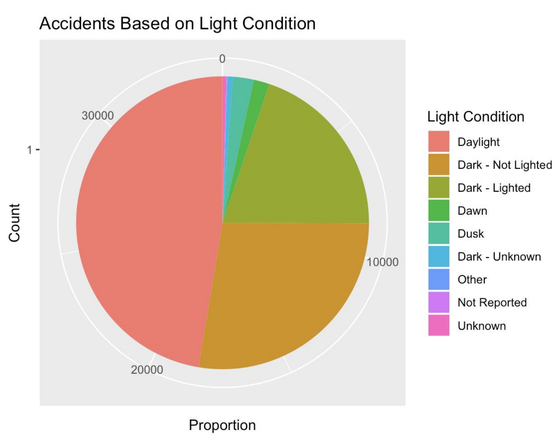
The time series graph shows the total number of fatalities by month.

The totals increase as the year progresses, peaking in October.



Light Conditions

The most common traffic accident light condition is broad daylight (roughly 45%). The next two most common conditions are Dark - Not Lighted and Dark - Lighted, making up a little under 50% of the total. The other groups are only a small fraction of the total accidents.

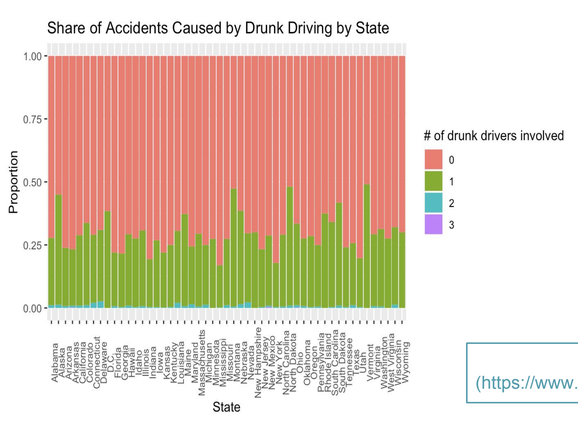


Accounting for Drunk Driving

This stacked bar chart illustrates the proportion of each state's total traffic accidents due to drunk driving, and the number of drunk drivers in the event.

These incidents make up a significant portion of the overall counts.

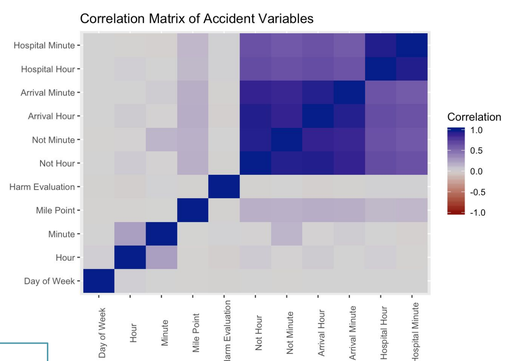
Roughly 25% of each state's totals involve 1 or more drunk driver.



Variable Correlations

DIsplayed here is a correlation matrix of the continuous variables in the accident data set.

Most variables are not correlated with each other.



Source: Kaggle Datasets (https://www.kaggle.com/usdot/nhtsa-traffic-fatalities)