The Project

My final project topic is 'Traffic Crash Analysis'. I will be using the 'Crash Data' from the Open Data Philly website. My original plan was to use crash data from Open Data Cinncinnati, however I found the Philadelphia data to be more explanatory as it had more variables for analysis. The dataset contains data from Philadelphia Police Department responses to traffic crashes in the city of Philadelphia, PA dating from 2007 through 2017. There are 90 variables related to the crash report across just under 80,000 records in the dataset. Each traffic crash response contains data relating to the weather and lighting at the time of the crash, road conditions, vehicle type, and severity of the crash among other attributes.

I will be using histograms and scatterplots to perform exploratory data analysis to learn about the data and to identify potential relationships between variables. I will be converting some of the textual and categorical data to numeric data and using regression analysis to identify explanatory variables with a statistically significant relationship to defined outcomes.

The Dataset

Website: Open Data Philly

Data Title: Crash Data 2007-2017 (CSV)

URL: https://www.opendataphilly.org/dataset/vehicular-crash-

data/resource/229497e2-ef2e-4a8d-8408-badddedb2657

Hypotheses

- There is a relationship between vehicle type and the presence of injuries in the data.

- Certain age brackets for drivers are impacted to a greater degree by weather or road conditions.

- The combination of road type and weather conditions show a relationship to traffic incidents.

- Adverse weather increases the instances of injuries among accidents.

- There is a higher incidence of accidents in areas based on latitude and longitude.

Findings

- Accident frequencies are highest on Fridays and Saturdays
- The peak for crash reports is rush hour on Fridays
- Accident frequency is lowest during the month of February, then climbs steadily to peak in May
- Rain has a significant contribution in relationship to crashes among all drivers
- Snow is a significant contributor for drivers aged 50 years or older
- Drivers aged 50 or older have a higher probability of being involved in a crash during the hours of 7AM to 2PM
- Drivers aged 16 to 20 have a higher probability of a crash between the hours of 9PM and 4AM
- Most crashes result in at least one injury (approximately 75%)
- Accidents during adverse weather conditions (rain, snow, sleet) are related to a lower rate of injuries
  - This may be due to reduced speeds during adverse weather (no supporting data)
- Accidents including heavy trucks and/or motorcycles have a significant relationship to major injuries
- Unbelted occupants has a significant relationship to major injuries and fatalities
- There is a negative significant relationship in the data between crash-related fatalities and SUVs and small trucks, indicating lower instances of fatalities for these vehicle types