Project Summary

Steven

Mohammed

Hubeyb

Eli

How old are they?

To categorize ages, we created new columns for each age group and used conditionals to dictate a value of either 1 or 0 for each row. Next, we grouped the data frame by year and took the sum to get final numbers to plot in a bar graph. Our observation here is that the data did not support our hypothesis as the leading age group of drivers was between 25 and 50. More so, out box plot of driver age revealed that the mean age is between 35 and 37 which is outside of our defined University Student demographic age of 18-21.

Where are they?

What Streets had the highest volume of accidents?

Our heat map revealed some hot spot areas near the University but also areas that were further away, so we declared this as inconclusive to our hypothesis. To confirm, we also created a bar graph of the top 20 street names by accident volume and observed that the results were not all strictly near ASU.

What time?

At what time of day are most accidents happening?

To confirm the most common time of day when accidents are occurring we first converted our time data into a simple hour format and used binning to plot using a bar graph. Our observation here is that most accidents are happening during late night hours which does support our defined demographic from our hypothesis. The most common hours were 11pm,10pm and 1 am.

Additional plots

In addition, we were curious to see if looking into Gender, Drug/Alcohol Influence or Injury Severity would lead us to any new discoveries.

We used the same approach as we did with age to gather values from Driver 1 and Driver 2 gender. The results presented a slight increase year by year but did not indicate that Gender played a significant factor.

We looked at Injury severity by first grouping by Injury Severity then taking the sum of the values and creating a new data frame. Our new data frame showed injury severity by total injury count and total Male/Female count within each level.

Finally, we looked into drug/alcohol influence by using the same approach as with age and gender previously to combine drug influence and alcohol influence. Our results were that most accidents involved no substance influence of any kind. Our conclusion was that the resulting data did not support our hypothesis.