Analysis of Deer Related Crashes in Pennsylvania

Introduction:

An analysis of 87, 976 deer-related crashes in Pennsylvania between 1997 and 2017. The goal of this analysis is to garner a better understanding of deer related crashes in hopes of offering mitigation solutions.

Motivation:

According to data from State Farm, between July 1, 2017 through June 2018, Pennsylvanians reported 141,000 deer- related insurance claims. This was nearly 10% of all deer-related car accident claims in the United States in that time frame. The average cost of these claims is around $4,341 for property damage alone. The cost can rise steeply if injuries are incurred.

While Pennsylvania has the highest deer-related car accidents of any other state, it lags behind all other states in mitigation solutions such as wildlife crossings.

Data:

The data used for this analysis is sourced from the opendata PA. I utilized opendata's own query tool to narrow down the data to only crash incidents that were deer-related. I then further manipulated the data utilizing OpenRefine.

Note that this data only consists of incidents that were reported to the police.

OpenRefine Manipulation:

1) Removed columns that were redundant – there were many instances where there was both a yes/no and a count column for the same attribute (i.e., a column for motorcycle – yes/no as well as a column with a count of how many motorcycles that were involved). In these cases, I removed the yes/no columns.
2) Any remaining yes/no columns, I changed to 1's(yes) and 0's(no) to aid in later analysis.
3) Ensured all numeric data had a numeric data type rather than character
4) I removed any special characters from the column heads.

Analysis:

Basic Stats:

Total number of injuries between 1997 – 2017: 29, 173
Total number of fatalities between 1997 – 2017: 233
Total number of vehicles towed between 1997 – 2017: 84,233 (almost 96% of crashes reported to police)

Chi-Square Tests:

H_o for all below tests: Both variables are independent of each other and therefore it is "random chance" that they occurred in the same incident.

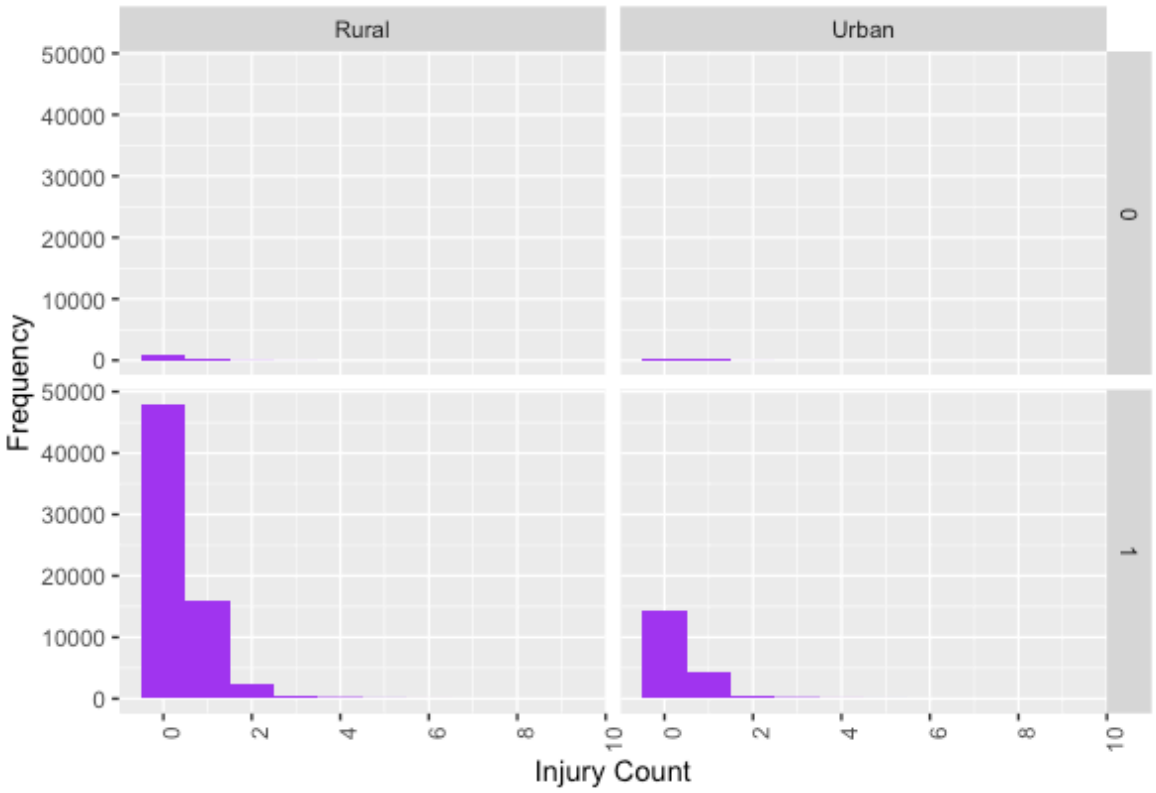
alpha = 0.05

For a chi-square test run for the variables "Sudden Deer" and "Illumination Dark," the p-value was 6.228x10^-6. Therefore, we reject the null hypothesis. "Sudden Deer" is not independent of a lack of illumination.

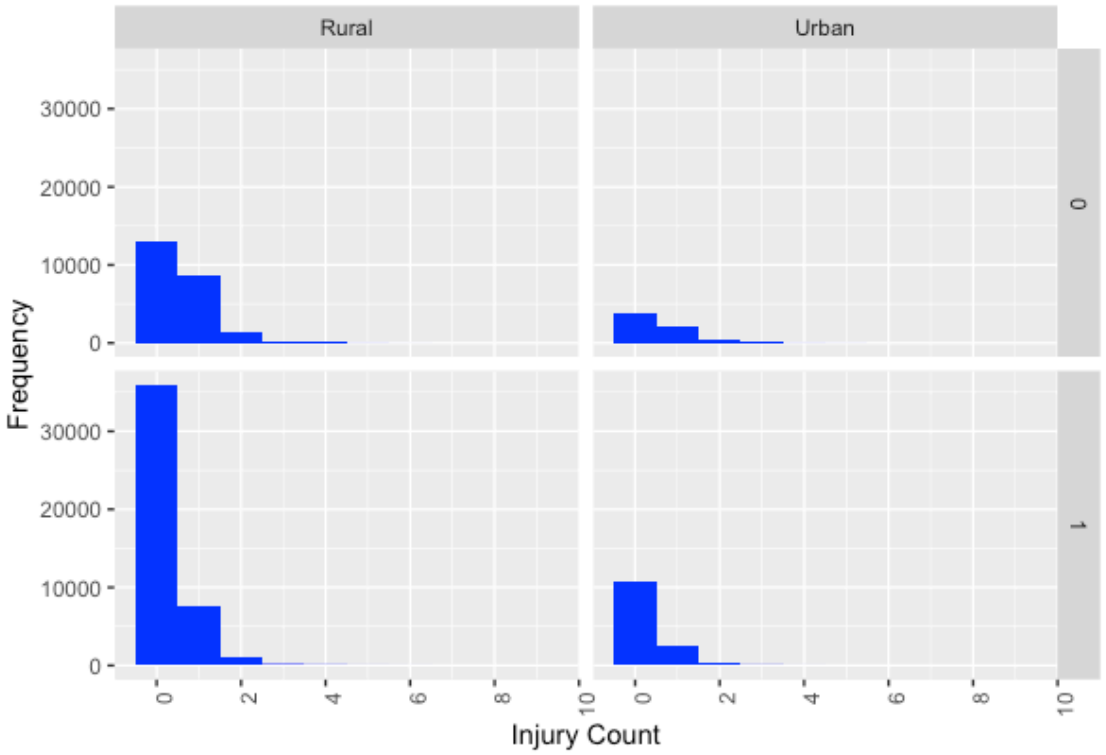
For a chi-square test run for the variables "Sudden Deer" and "Fatigue_Asleep," the p-value is 0.607 so we fail to reject the null hypothesis that these variables are independent of each other.

For a chi-square test run for the variables "Sudden Deer" and "Cell Phone," the p-value is 0.02561. So we reject the null hypothesis. "Sudden Deer" is not independent of the driver's use of a cell phone at the time of the inicident.

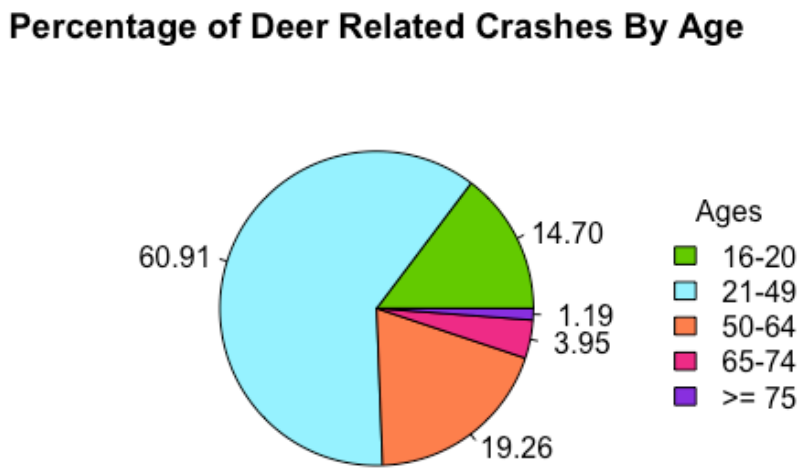
For a chi-square test run for the variables "Hit Deer" and "Cell Phone," the p-value is 1.031x10-7. Therefore we reject the null hypothesis. Instances of the driver hitting the deer is not independent of whether they were using their cell phone.



The above chart is a faceted histogram depicting the injury counts for “sudden deer” related incidents in urban and rural settings.



The above is a faceted histogram depicting the injury counts for incidents in which the vehicle hit a deer in urban and rural settings.



The above is a pie chart depicting the percentage of incidents involving each age range of the driver.

Conclusion:

As of now, this project is incomplete. There is more analyses I would like to do with other variables – driver impairment, types of vehicles, etc. I also wish to do a more thorough cost-benefit analysis for mitigation strategies and techniques such as wildlife crossings.