Here's What (and Whom) to Avoid When Driving

Desmond Cole, Teerth Patel, Yunbin Peng

April 14, 2018

Introduction

We analyze traffic fatality data provided by the National Highway Traffic Safety Administration (NHTSA) to assess various predictors of traffic fatalities and develop a limited profile of the circumstances associated with traffic fatalities.

Exploratory Analysis and Visualization

Geographic Patterns

Time Trends

Daily Cycle

At the national and state level, the cycle of fatal accidents throughout the day is fairly consistent. There is a local maximum in the early morning, correlating with morning rush hour. Beginning just before noon, the level of fatal accidents rises consistently to peak at between 7 and 8PM, before declining steadily through to about 3 or 4 in the morning.

(Insert National graphic)

With State-by-State plots, the daily fatal accident cycle is roughly similar, without significant change. The below plots show results from shape-based time series clustering of different states, according to daily fatal accident patterns.

(Insert side-by-side state-level centroid plots)

Weekly Cycle

Selected Predictors of Traffic Fatalities

We used a mix of classifiers to assess factors relevant to incident fatalities, with a particular focus on driver behavior, environmental conditions, and vehicle manufacturer. In many cases, we had to grapple with significant class imbalance. For example, in predicting whether an accident results in multiple fatalities

Driver Behavior

Drugs/Alcohol Distraction

Car Type

This section explores in some detail different fatality rates for different car types (by automaker, vehicle body, etc.).[1] For this assessment, we focused on the 15 largest automakers that together produce more than 95% of the cars and light trucks on American roads. In addition, we subsetted the data to focus specifically on smaller vehicles, excluding commercial vehicles, semi-trucks, etc.

The plot below shows a basic ranking of car manufacturers according to the number of fatalities per million vehicles.

(Insert graph of fatalities/fatal accidents by car manufacturer)

The plot above, although suggestive of meaningful difference across manufacturers, fails to fully account for the various contextual differences which may be unobservable. To further explore the relevance of car manufacturer, we consider the relationships between car model and various fatality-related predictors, to understand if a given manufacturers' products tend to be associated with high-risk behaviors or other crash factors. The table below shows the results from various multi-class classifications of car manufacturer ran using a set of crash-relevant predictors .... The results do not suggest any substantial mechanism(s) by which a car's make determines the risk of it being involved in a fatal incident.

(Insert table of classifications and results)

Environmental Conditions

Conclusion

References

Batterman, Stuart, Richard Cook, and Thomas Justin. "Temporal variation of traffic on highways

[1] Note that the numbers in this assessment focus on car brands *involved* in fatal incidents. Thus, they do not imply specifically that the vehicle types considered here actually caused death(s), or that the driver(s) of the vehicle(s) themselves died.

and the development of accurate temporal allocation factors for air pollution analyses." *Atmos Environ.* Apr. 2015.

Zador, P.L, S.A. Krawchuk, and R.B. Voas. "Relative Risk of Fatal Crash Involvement by BAC, Age, and Gender." *U.S. Department of Transportation - National Highway Traffic Safety Administration.* Apr. 2000.