Final Write-Up

Azka Javaid & Caleb Ki

December 19, 2016

Introduction

In 2015, 35092 people died in motor vehicle accidents according to the National Highway Traffic Safety Association (NHTSA). This is increase of 2348 fatalities from 2014. Part of the reason that there was an increase in the number of fatal traffic accidents overall is that the number of fatal traffic accidents involving alcohol impaired drivers increased by 322. The CDC reports that around 31% of traffic fatalities include a drunk driver and that alcohol impaired vehicle accidents costs around $44 billion.

Identifying the risk factors and the behavior of drunk drivers could go a long way in reducing drunk driving. Knowing who is more likely to drunk drive gives information about who to target with intervention methods. In addition, if we are able to identify the times that driver are more likely to be drunk, the general public could be informed (e.g. if there are more alcohol immpaired drivers during weekends, the general public should be aware and would then be able to be more viligant while driving during the weekend).

Furthermore, although the dangers of and facts concerning drunk driving are well documented, the known risk factors are generally only known at the individual level. What we mean is that there is relatively little literature exploring why there are large discrepancies in rate of drunk driving between different U.S. counties. For example, in 2015, Fremont County, Wyoming, had 27 fatal traffic accidents involving alcohol per 100,000 people whereas the same statistic for Morris County, New Jersey is only 0.2. The difference between the two counties is massive. But what makes Fremont County so different from Wyoming? Why is the rate of traffic fatalities including drunk drivers much higher in county than in another? What economic and demographic differences of the two counties explain why?

Determining the factors that could help explain the discrepancy in the rates of fatal traffic accidents involving a drunk driver could pinpoint the risk factors of drunk driving at the county level. This in turn would indicate which geographic areas would benefit the most from different types of intervention.

For our project we are looking to determine what predictors/factors at the individual and county level are good predictors for whether a driver involved in a fatal traffic accident was alcohol impaired or not. In addition we would also like to see what behaviors are good predictors as well. Specifically, if a fatal accident occurs during the night or the weekend, is it more likely that the drivers were drinking?

Data

In order to answer our questions we used data from two different sources. The first and primary source of data came from the Fatality Accident Report Service (FARS) managed by the NHTSA. FARS provides information concerning every vehicle and person involved in a fatal traffic accident. The data is provided at an accident, vehicle, and person level. Each accident is given a unique identifier, and each row of the vehicle and person datasets includes the unique ID indicating which accident that the vehicle or person was a part of. This allowed us to seamlessly join the data together. The manual provided in this link contains information about each variable within the datasets.

The second and supplementary source of data came from the American Community Survey (ACS) conducted by the U.S. Census Bureau. The ACS provides social, economic, and demographic data at the county, state, and national level. The FARS data contains geographic information like latitude, longitude, county, and state. By calculating FIPS codes, we were able to match the county level data we were looking for to each accident, vehicle, and person. Information about the ACS can be found here.