Understanding Traffic Collisions in Los Angeles County

STA 440 Final Project

Sara Lemus

Introduction

The year 2019 in Los Angeles was a particularly troubling year in terms of car accidents: the LAPD reported that there were 236 traffic deaths and over 54,000 accidents in the year alone, with a 32% increase in fatal crashes over the last 5 years [1]. Fatal and severe traffic accidents not only devastate the lives of those who are involved, but also impact communities on both a local and national scale. A 2015 study found that the total cost of motor vehicle accidents in the United States was $242 billion, which includes costs such as hospitalization, property damage, and unemployment [4].

There have been initiatives to reduce the rising number of traffic fatalities in Los Angeles. In 2015, Los Angeles mayor Eric Garcetti launched a project known as Vision Zero, which aims to eliminate traffic deaths completely by the year 2025 [2]. The initiative uses CHP data to inform various policies, such as establishing traffic education programs and identifying and monitoring dangerous intersections in Los Angeles. Although it is a lofty goal to prevent all traffic deaths by the year 2025, further analyses of traffic data can provide a better understanding of the causes of collisions as well as improve or implement polices to make the roads safer for all drivers.

This analysis examines traffic data from motor vehicle crashes from Los Angeles county in the year 2019. There are two main goals for this analysis specifically:

- Understand if certain groups of people are overrepresented in car accidents
- Build a model to predict whether or not a crash will be fatal or severe

Previous literature has researched which factors contribute to car accidents. One comprehensive meta-analysis from 2019 found that factors such as weather, traffic congestion, and time of day may all be related to severe accidents. In terms of driver characteristics, factors such as driver sobriety, car type, age, and driving behavior also contributed to accidents [3]. Although the dataset being used for this analysis does not contain spatio-temporal traffic data, we will examine some of the other factors mentioned in the report to understand to what extent they contribute to Los Angeles car accidents.

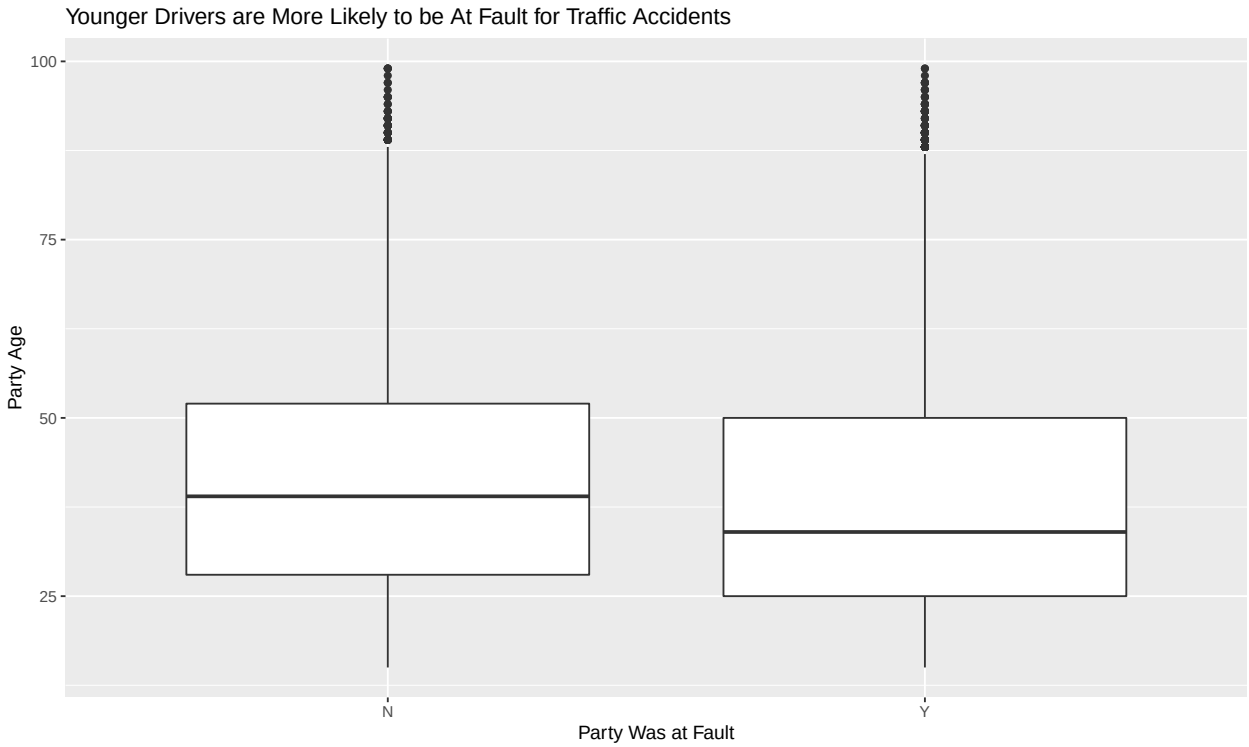
About the Data

The dataset was accessed from the Statewide Integrated Traffic Records System (SWITRS), a database which is directly managed by the CHP (California Highway Patrol). Researchers at UC Berkeley created an online tool which allows users to filter through the extensive database (which contains data on all California counties through numerous years) and retrieve certain subsets of the data. There are two datasets being used for this analysis specifically. One is the collisions dataset, which has about 40,000 rows and contains information on various factors relating to the crash, such as crash date, time, weather conditions, and reason for the crash. The second data set contains information on the parties involved in the crash. It contains about 90,000 rows (as many crashes have multiple parties involved) and specifies demographic and situational information about the parties involved in the collision, such as age, gender, and race.

Demographic EDA

The first research goal is to understand various discrepancies in demographics for involvement in all crash types.

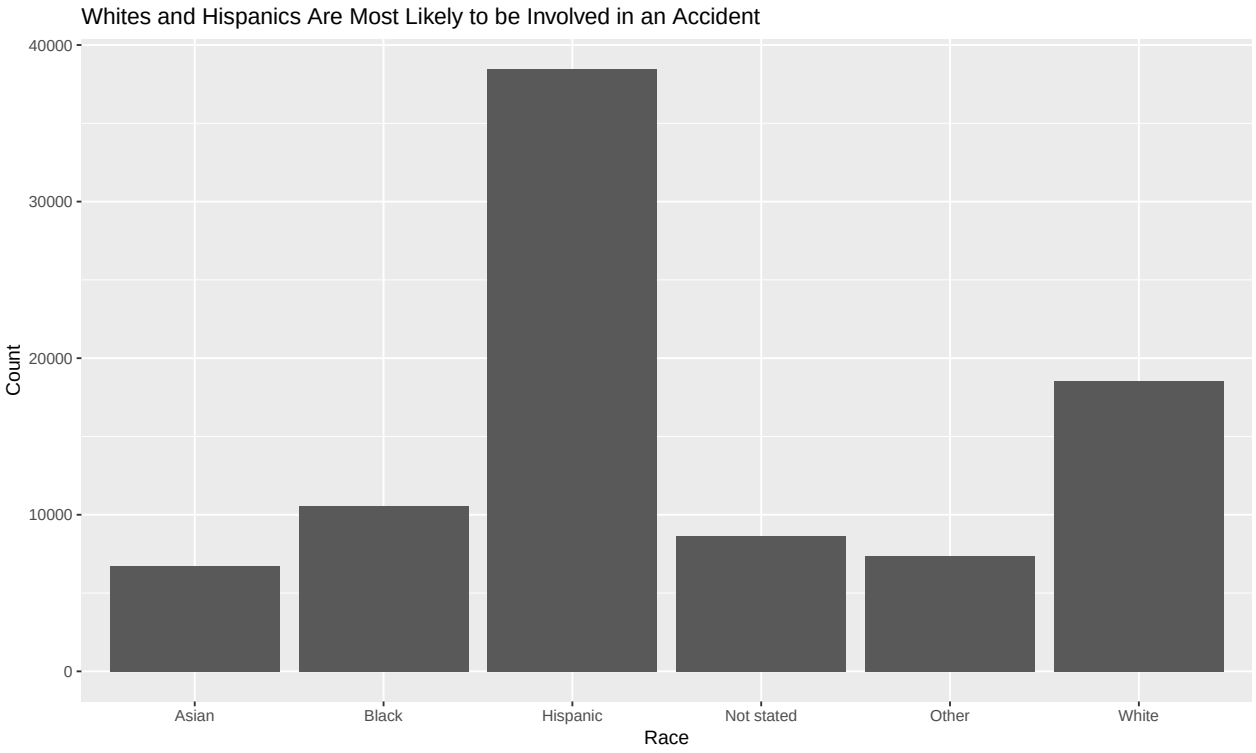
This plot suggests that younger drivers are more likely to be at fault for a crash, as the median age for drivers at fault was much lower than for those not at fault. This may be explained by the fact that younger drivers are less experienced and more likely to end up in dangerous driving situations.



We also see a difference amongst genders in traffic collision outcomes. For drivers not at fault for collisions, men were involved in about 10% more traffic accidents than women. However, when the driver was at fault, the proportion of men was about 18% higher than the proportion of women responsible for accidents. The reason for this discrepancy is not immediately clear.

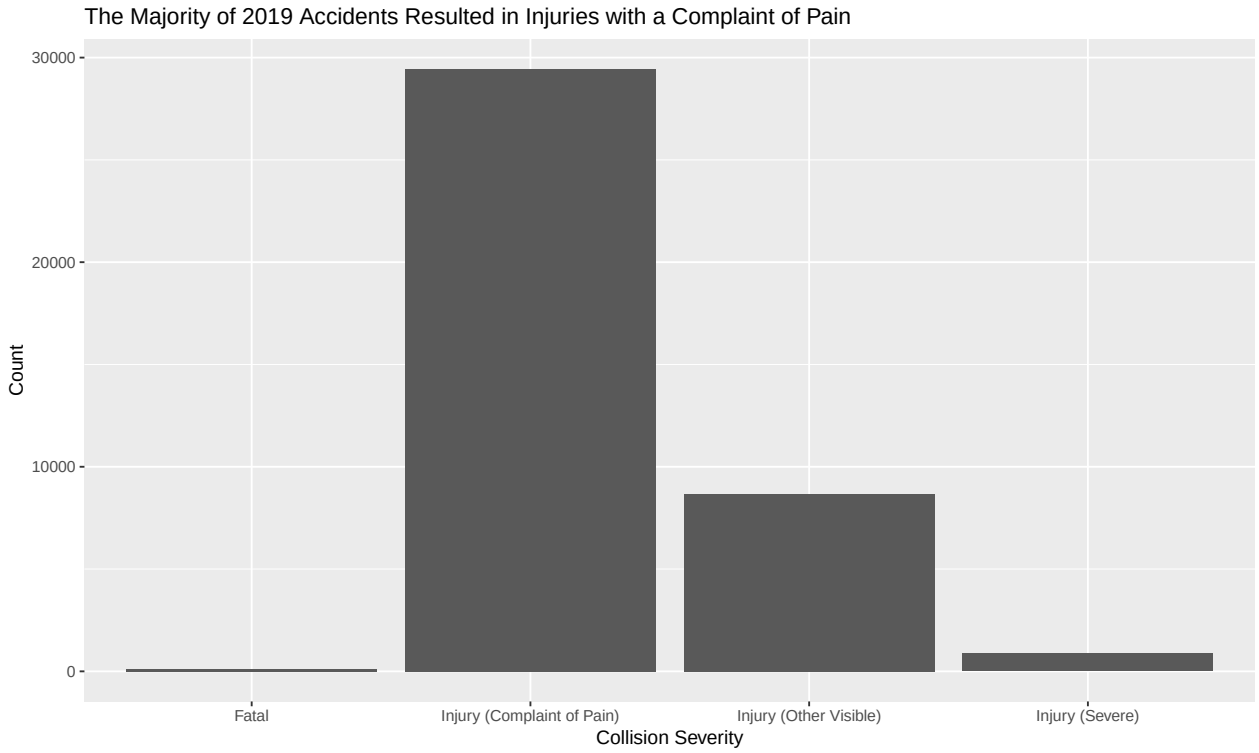
AT_FAULT	PARTY_SEX	Count	Percent
N	F	22610	44.36
N	M	27708	54.37
Y	F	13446	36.39
Y	M	20065	54.30

Finally, we examine race and collision involvement. This graph shows a histogram of what races were involved in reported traffic collisions in 2019. The plot indicates that Hispanic and White people are involved in the most crashes. However, comparing this distribution to recent LA County Census data [5], it seems that people who were identified as Asian or White may be slightly underrepresented as they make up 15.4% and 26.1% of the population demographic, respectively while being involved in 7.7% and 21.1% of accidents. Black people may conversely be overrepresented, being involved in 12.0% of accidents but only representing 9.0% of the LA population.



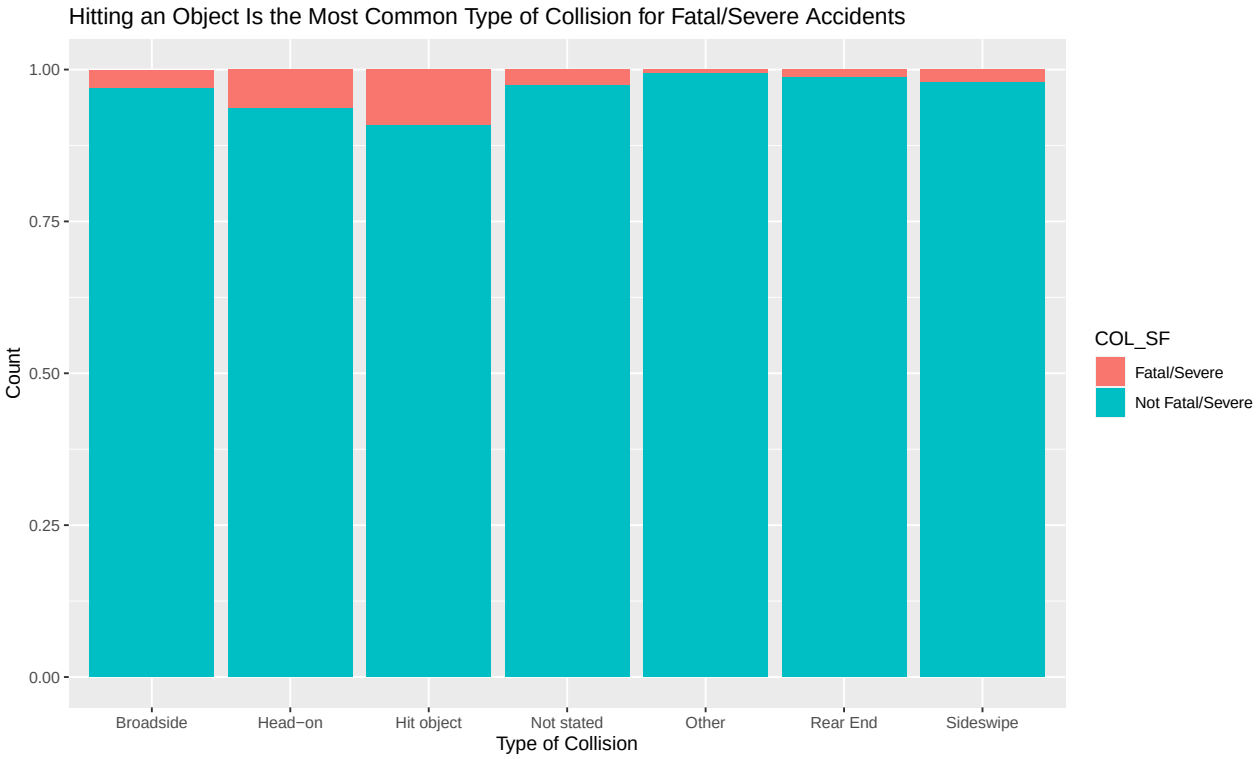
Fatal/Severe Accident EDA

The goal of this EDA section is to better understand the data set and which factors contributed specifically to fatal and severe accidents. To better address this research goal, we created a binary variable that indicates whether or not an accident was considered to be Fatal/Severe or not. We felt that it was an appropriate choice to examine both severe and fatal crashes together due to the fact that other analyses use the same definition [2], and also because the number of fatal crashes is very small compared to the number of other accidents that it may not informative to look only at fatal accidents, as evident in the figure below.

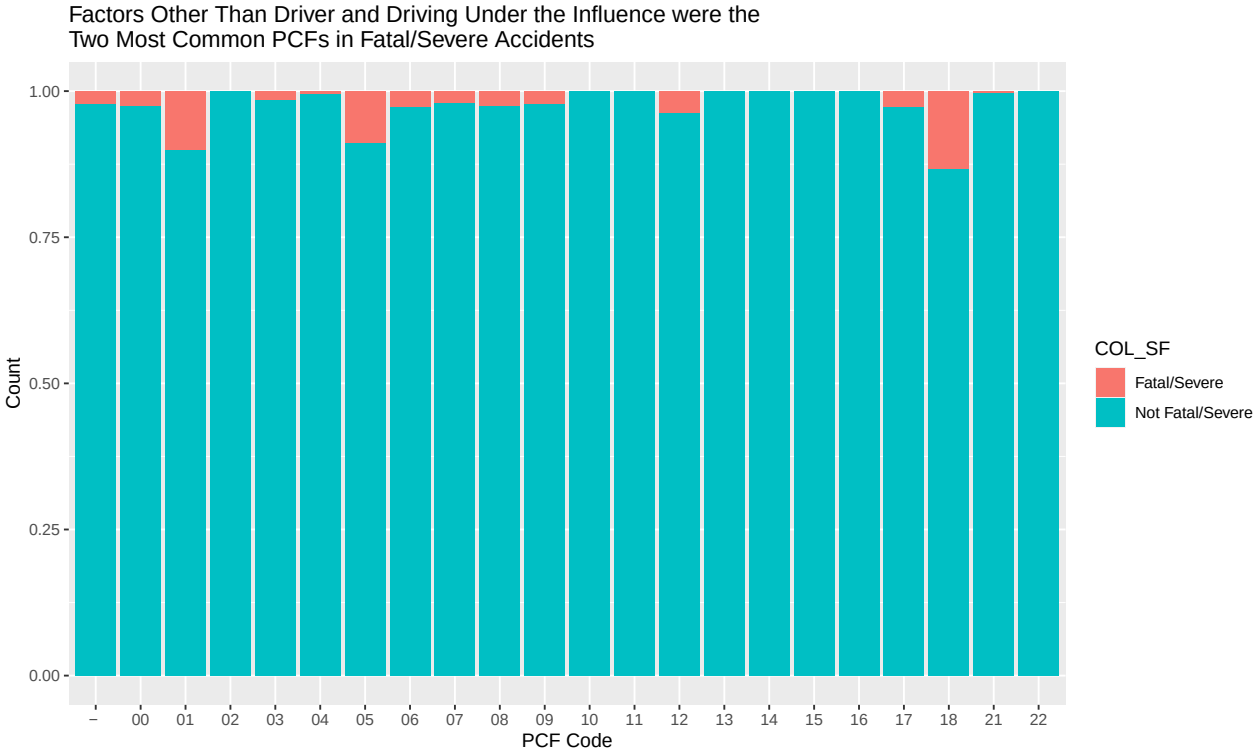


It is worth noting that crashes that resulted in property damage only (PDO) are not recorded in this dataset. That being said, it remains that only a minority of crashes are fatal or result in severe injury (about 2.54% of total accidents). The CHP defines a severe injury as "an injury other than fatal which results in a severe laceration, a broken extremity, crush injury, suspected skull, chest or abdominal injury other than bruises or minor lacerations, significant burns, unconsciousness, paralysis or any combination thereof [2]."

In terms of collision type, it seems that hitting an object is the most common collision type for accidents that end up being severe/fatal, with head-on collisions being the second most common. This is an interesting finding, as hitting an object means that another driver may not be involved and calls for further examination of driver sobriety and other characteristics that may explain this type of collision.



Collisions are often listed as PCF (primary collision factor) violations, which give information as to why the crash occurred. This plot shows that the most common PCF codes that were listed for fatal and severe collisions were 01- Driving or Bicycling Under the Influence of Alcohol or Drug, 05- Wrong Side of Road, and 18- Other Than Driver (or Pedestrian). This is in agreement with previous studies, which show that sobriety is an important factor in explaining fatal/severe accidents.



References

[1] KABC. "LAPD Urges Safe Driving after Releasing Staggering Data on Traffic Deaths across City in 2019." ABC7 Los Angeles, 30 Dec. 2019, https://abc7.com/5799520/.

[2] Vision Zero - Los Angeles County. https://pw.lacounty.gov/visionzero/. Accessed 1 Mar. 2021.

[3] Retallack, Angus Eugene, and Bertram Ostendorf. "Current Understanding of the Effects of Congestion on Traffic Accidents." International Journal of Environmental Research and Public Health, vol. 16, no. 18, Sept. 2019. PubMed Central, doi:10.3390/ijerph16183400.

[4] The Economic and Societal Impact Of Motor Vehicle Crashes, 2010 (Revised). DOT HS 812 013, National Highway Traffic Safety Administration, May 2015.

[5] U.S. Census Bureau QuickFacts: Los Angeles County, California. https://www.census.gov/quickfacts/fact/table/losangelescountycalifornia/RHI725219#RHI725219. Accessed 1 Mar. 2021.