ProjectReport

Mengdi Sun, Chien-Cheng Lai, Ning Hsu, Jiabei Li

4/17/2018

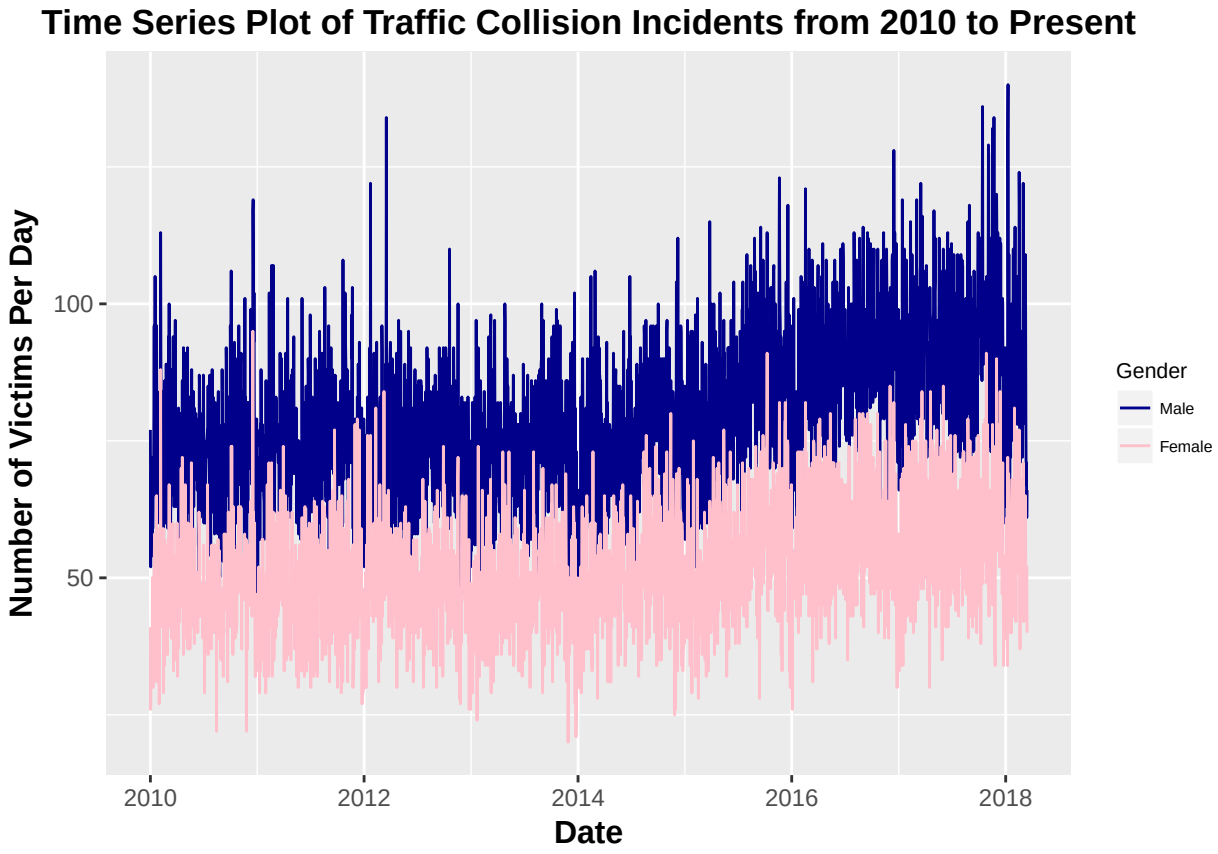
Background

This project's dataset is provided by the Los Angeles Police Department. It records all the traffic collisions from 2010 to present in the City of Los Angeles. The entire data reflects the date reported, date occured, time occurred, area name, victim age, victim sex, and victim descent. Each row in this dataset represents a traffic collision incident. There are 406,000 rows and 18 columns in this data. Our purpose is to analyze the correlations among all those factors and conclude the main factors that cause traffic collisions in the City of Los Angeles.

Objective

Our team's objective is to seek factors that correlate with the number of victims in Los Angeles. The factors we will analyze include time, age, sex, descent, unemployment rate, and personal income.

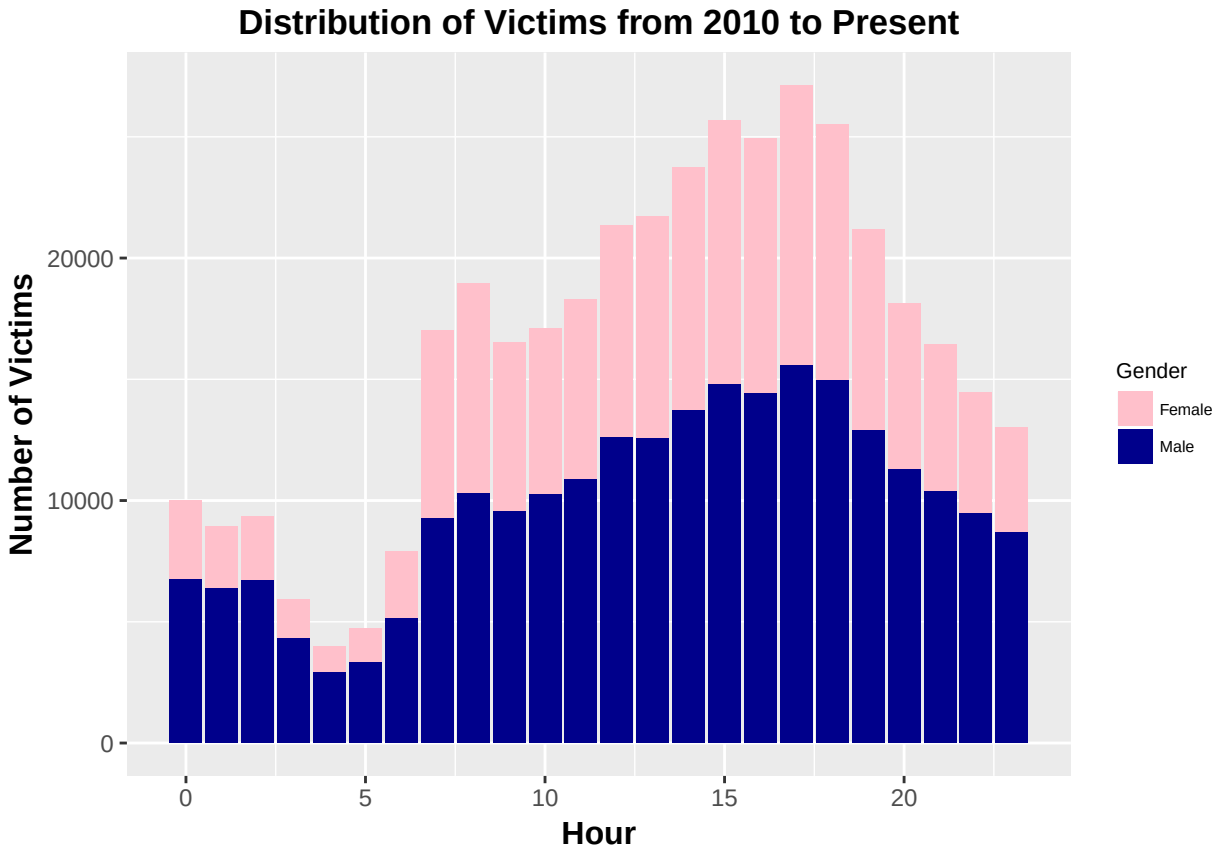
Data Summary



This time series plot summarizes the number of victims for both male and female throughout the past eight years. Clearly, male have more collisions than female in the past eight years. The trend for victims from 2010 to 2015 are consistent and stable. The highest number of victims happened during 2010 and 2015 is in year 2014. Starting from year 2016, we can see a slight increase of victims for both male and female. We hypothesize factors such as personal income and unemployment cause the increase of victims.

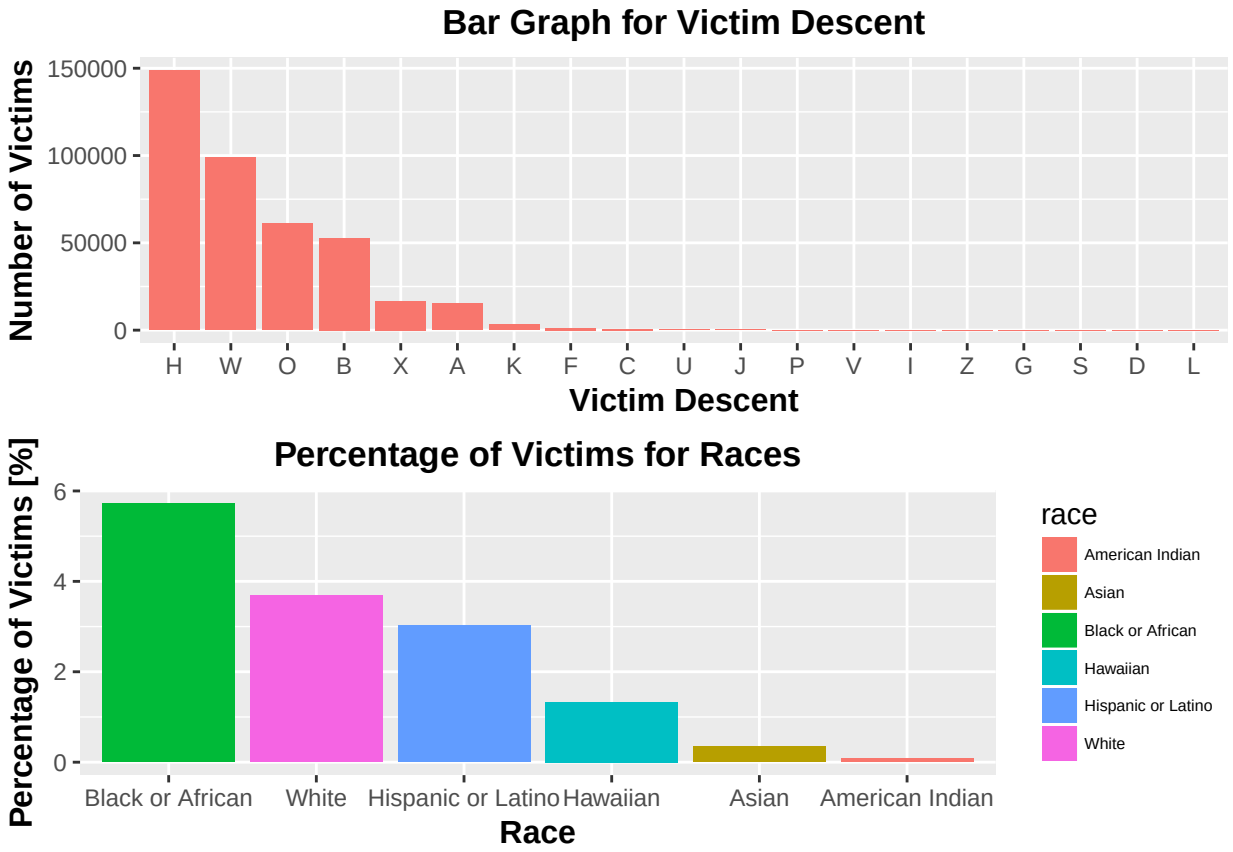
Data Analysis

1.Relationship between number of victims and time



From this bar graph, we can see that both male and female have more victims in late afternoon from 3PM to 7PM. In addition, both male and female have the least victims in early morning from 4AM to 5AM. The highest victims happens at 5PM, which is a total of 27117 victims. The lowest victims happens at 4AM, which is a total of 4716 victims. Same as the line graph above, this bar graph depicts the fact that most victims in Los Angeles happen around after work and after school time. It is very useful to compare the number of victims with both age and gender groups. With the supporting trends from both graphs, we can conclude that most victims from traffic collisions are of 25-60 years old female.

2.Relationship between number of victims and descent



(Note: H- Hispanic, W- White, O- Other, B- Black, X- Unknown, A- Other Asian, K- Korean, F- Filipino, C- Chinese, U- Hawaiian)

Most victims from traffic collisions in Los Angeles were Hispanic/ Latino/ Mexican, which is 148528. Immediately following the highest race, second highest collisions victims are White, which is a total of 99006. The third highest collisions are from victim descent of other race, which is a total of 61023. The fourth highest collisions are from victim descent of Black, which is a total of 52852. The least number of collisions are from victims of Laotian descent, which is a total of 4. Overall, number of collisions of Hispanic descent victims is 1.5001919 times as much as White descent victims. Our team found the demographics of Los Angeles for each race and used it to standardize the number of collisions. Our team divided the number of victims of each race by the total population of each race and found the percentage of victims by racial composition. Therefore, this refined bar graph depicts which race is more likely to get into a collision. From this refined bar graph, we can clearly see that people of Black or African American are most likely to be victims from traffic collisions. People of White are second most likely to be victims from traffic collisions. People of American Indian and Alaska Natives are least likely to be victims from traffic collisions. People of Asian are second least likely to be victims

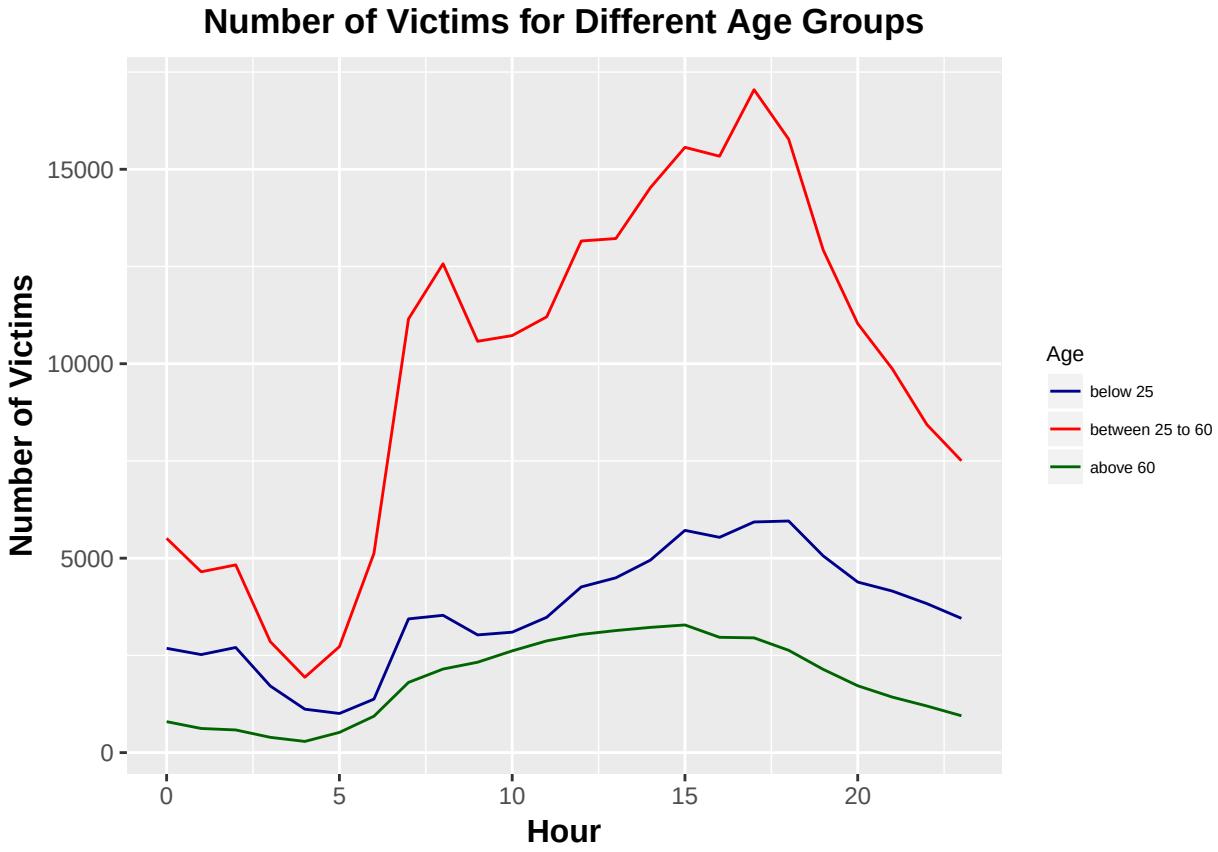
(Note: In order to clearly display all the races, our team decided to group all the victim descents into six major races: Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and other Pacific Islander, Hispanic and Latino, and White.)

3.Percentage of Victims by Race

race	percentage

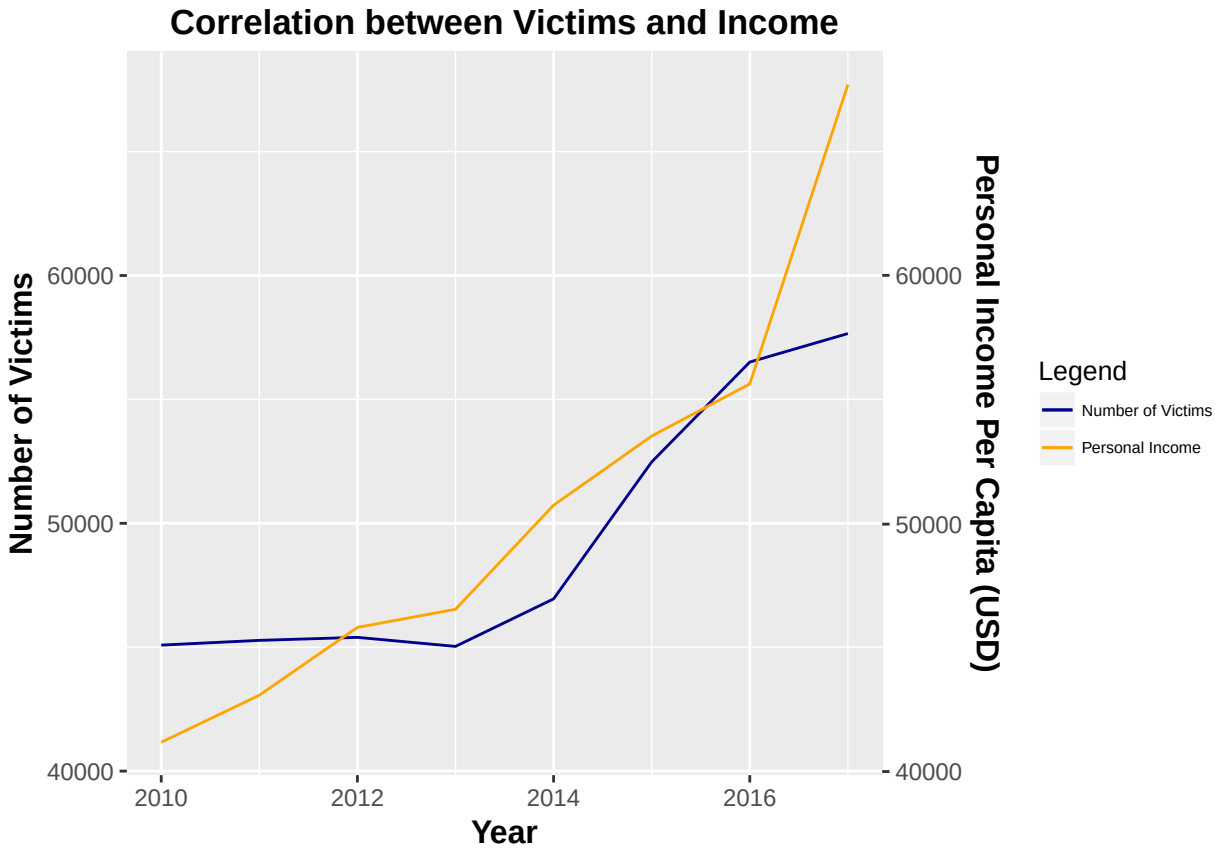
race	percentage
Black or African	5.7289019
American Indian	0.0828573
Asian	0.3519017
Hawaiian	1.3341007
Hispanic or Latino	3.0207720
White	3.6852503

4.Correlation between number of victims and age



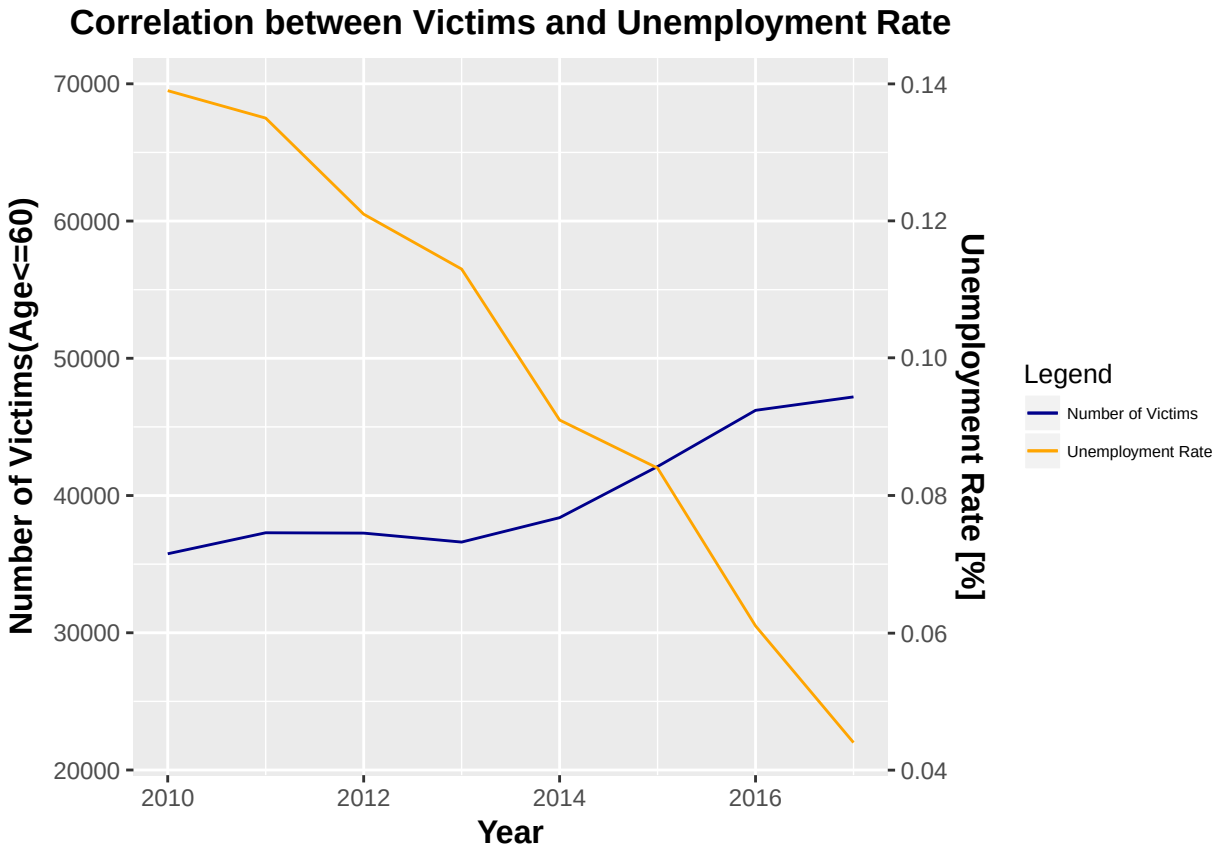
This line graph shows the correlation between victims' ages and time during the day. We concluded that most accidents happened during commuting hours, approximately at 5PM to 6PM, resulting in the highest number of victims. Least number of collisions happen around 4AM due to the streets having fewer cars. Collisions increase drastically from 4AM to 6AM due to more people on the road to work and school. Most victims are at 6PM for age groups below 25 and between 25 to 60. However, people who are below 60 years old had the most collisions around 3PM. Overall, this line graph effectively depicts the correlation between number of victims and time for different age groups, and we can clearly conclude that most victims are caused by heavy traffic and accidents during commuting time.

5.Correlation between number of victims and personal income



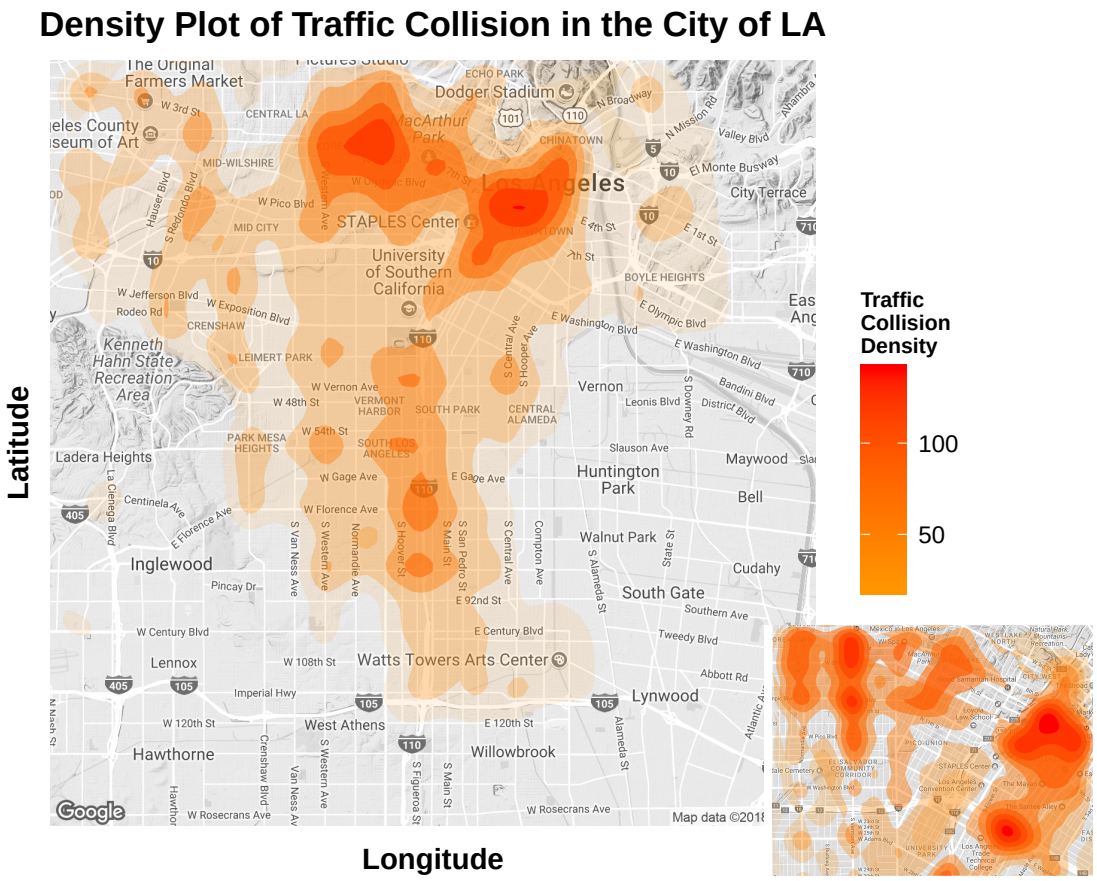
In order to get more depth and learn what other factors increase the number of victims throughout the years, our team found supporting datasets on personal income per capita of Los Angeles from year 2010 to 2017. As the number of victims increases, the amount of personal income per capita also increases. The positive correlation is very strong between 2013 and 2016. Even though number of collisions decreases during 2012 as personal income increases, the number of collisions immediately increases drastically after 2013 as personal income increases drastically too. Therefore, the graph depicts a positive correlation between both factors. We can conclude that more people are able to afford cars as personal income increases, which results in more traffic on the road and causes an increase in number of collisions and victims.

6.Correlation between number of victims and unemployment rate



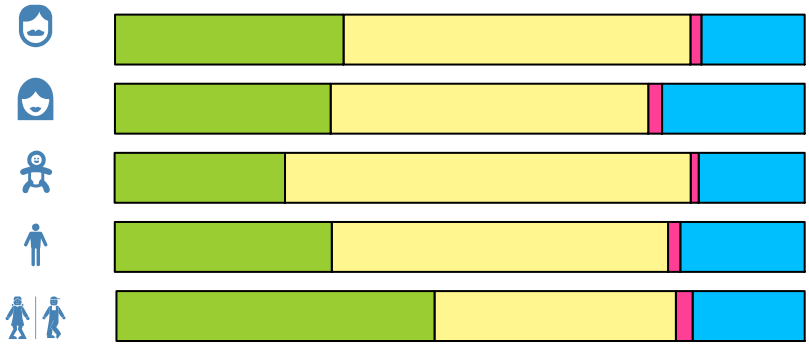
Originally, we graphed the "Number of Collisions for Different Age Groups" using hours as the x-axis. Now, we would like to use years as x-axis to graph number of collisions for age group below age of 25 and age group between 25 and 60, because number of collisions are the same as number of victims. Our team compared unemployment rate with number of collisions from 2010 to 2017. The graph depicts a strong negative correlation between number of collisions and unemployment rate. As the percentage of unemployment rate decreases, the number of collisions increases. This conclusion verifies with our team's hypothesis previously, because the higher the employment rate, the higher the number of collisions. High employment rate results in more people commuting to work, causing more collisions.

7.Density plot



Then we plot the density map of the traffic collisions happen in LA. The darker orange fields represent the areas that have the most victims whereas the lighter areas have least victims. As the map shows, most dark oranges fields are around the downtown area and along highways. This means that most collisions happen near downtown and highway, resulting in a high number of victims. This conclusion verifies the results we acquired above, which is the number of victims increase sharply at the time when people commute to work. From this density plot, we can also see a trail of dark orange patterns near highways and slowly aggregate in downtown in large, darker patches. This pattern implies that downtown is the center of traffic collisions due to its busy streets, which hold a large number of companies and workers.

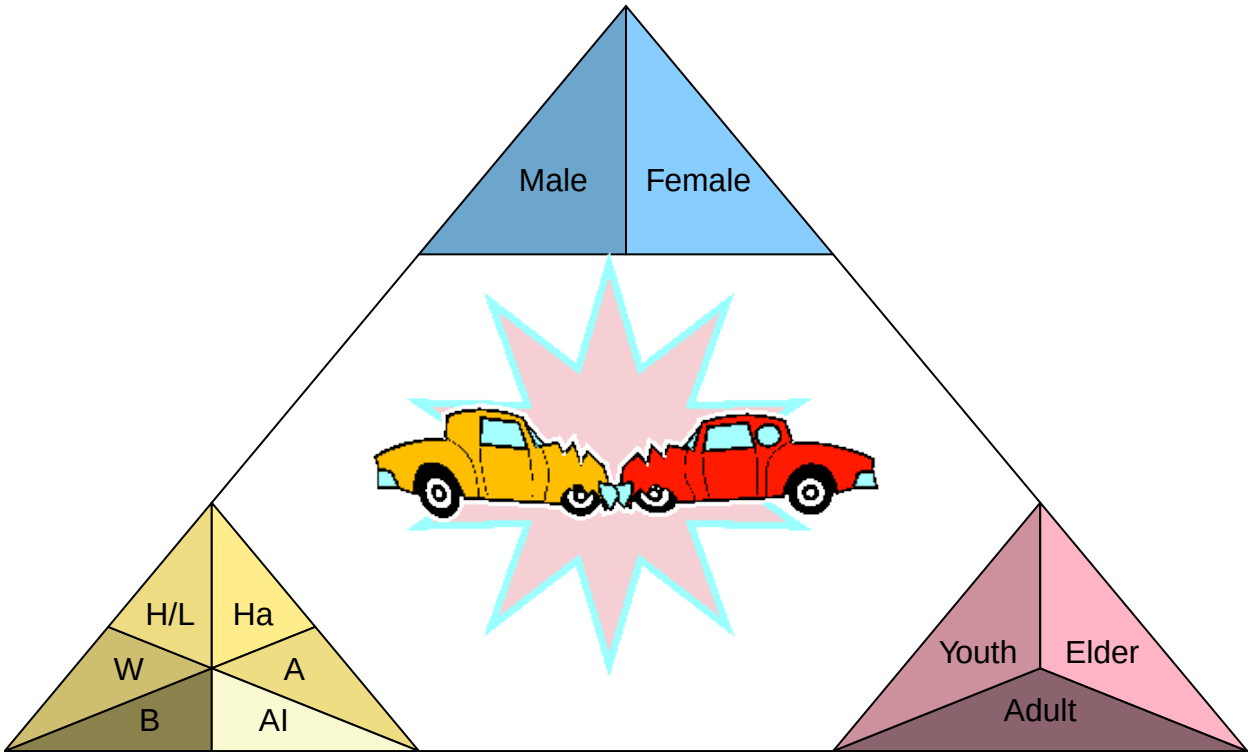
Summary Graph



Descent	Male	Female	Youth	Adult	Elder
Black or African	27270	25574	8584	36972	7296
Asian	2888	2476	647	3649	1091
Hispanic or Latino	91531	56941	32842	99974	15712
White	60303	38669	13796	64505	20705

This graph summarizes the number of victims in sex, race, and age. The first two rows represent male and female victims in different race, and the following three row represent victims in different age groups and race. The third row represents victims of age below 25 years old. The fourth row represents victims of age between 25 to 60 years old. The last row represents victims of age above 60 years old.

Summary factor graph



This factor graph summarizes all the factors we considered in order to analyze the trend and correlation for the number of traffic victims. Three three triangles represent gender, race, and age. Each triangle then divides into different categories. The darker the color, the more victims each factors has.

Conclusion

In our project, we did a broad investigation over the victims in the traffic collisions happen in Los Angeles from 2010 to 2016. The main factors that correlate with numbers of victims we look at are sex, race, and age. Furthermore, we aim to figure out other factors, such as personal income per capita, and unemployment rate, that correlate with the change of number of victims from 2010 to 2017. Sex, race, and age affect the number of victims throughout the years. We concluded that there are male victims than female victims, and most victims are mostly between 25 to 60 years old. People of Black and African American descent are most likely to be the victims from traffic collisions. Most traffic collisions occur during after work, resulting in a high number of victims. In addition, personal income positively correlate with number of victims, and unemployment rate negatively correlate with number of victims.

Sources

"Crime Data from 2010 to Present." Crime Data from 2010 to Present - Data.gov, Publisher Data.lacity.org, 10 Apr. 2018, catalog.data.gov/dataset/crime-data-from-2010-to-present.

"Traffic Collision Data from 2010 to Present | Los Angeles - Open Data Portal." Data.lacity.org, data.lacity.org/A-Safe-City/Traffic-Collision-Data-from-2010-to-Present/d5tf-ez2w.

"U.S. Census Bureau QuickFacts: Los Angeles County, California." Census Bureau QuickFacts, www.census.gov/quickfacts/fact/table/losangelescountycalifornia/PST045216.