Homework 8

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

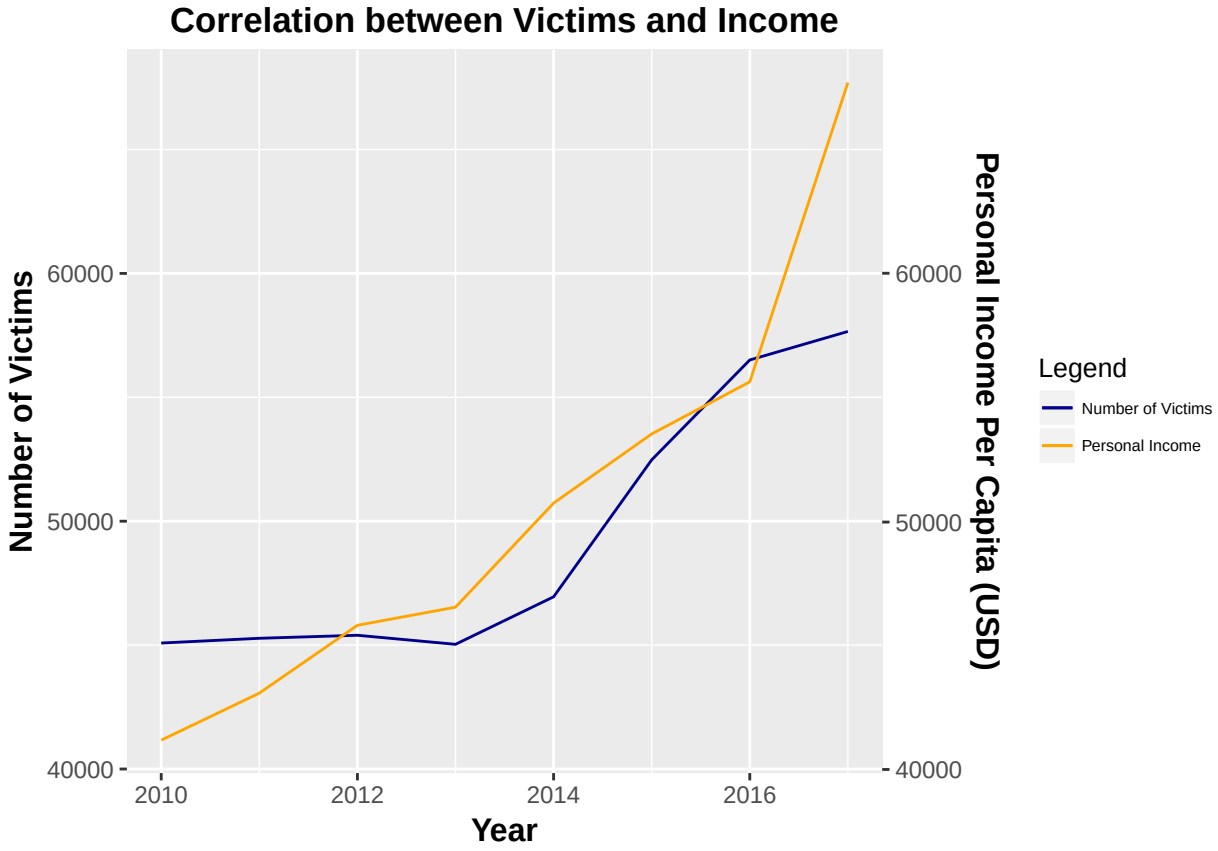
4/3/2018

Introduction

Last time, our team concluded that the largest group of traffic collisions are Hispanic/ Latino males between age 25 to 60. We would like to further our research by using two other auxiliary data and analyze which race is most likely to be traffic collision victim, and other factors that can also increase traffic collisions in Los Angeles.

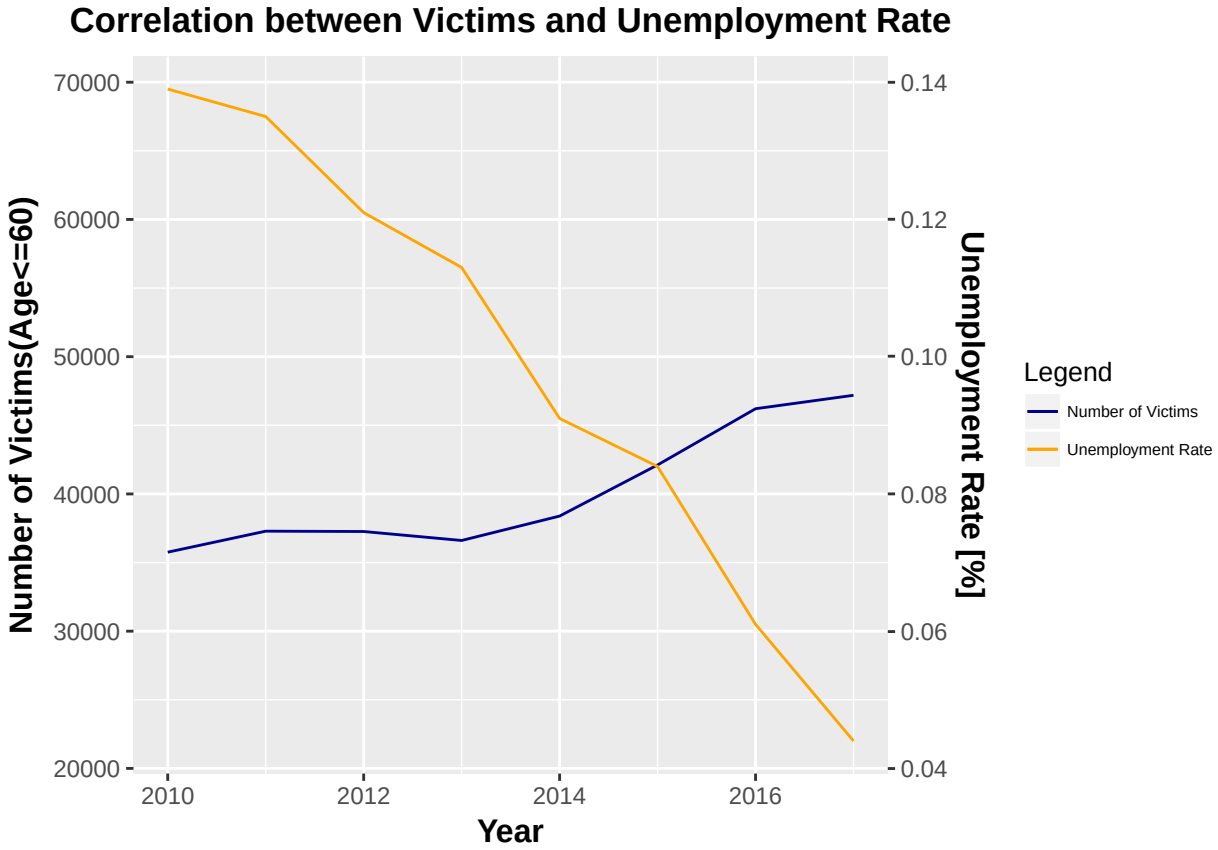
Question

What is the correlation between sex, age, and race to the total number of collisions occurred in Los Angeles from year 2010 to present? To be more specific, which race is most likely to be the victim and least likely to be the victim? In addition, are there any other factors that can increase collisions, such as personal income and unemployment rate?



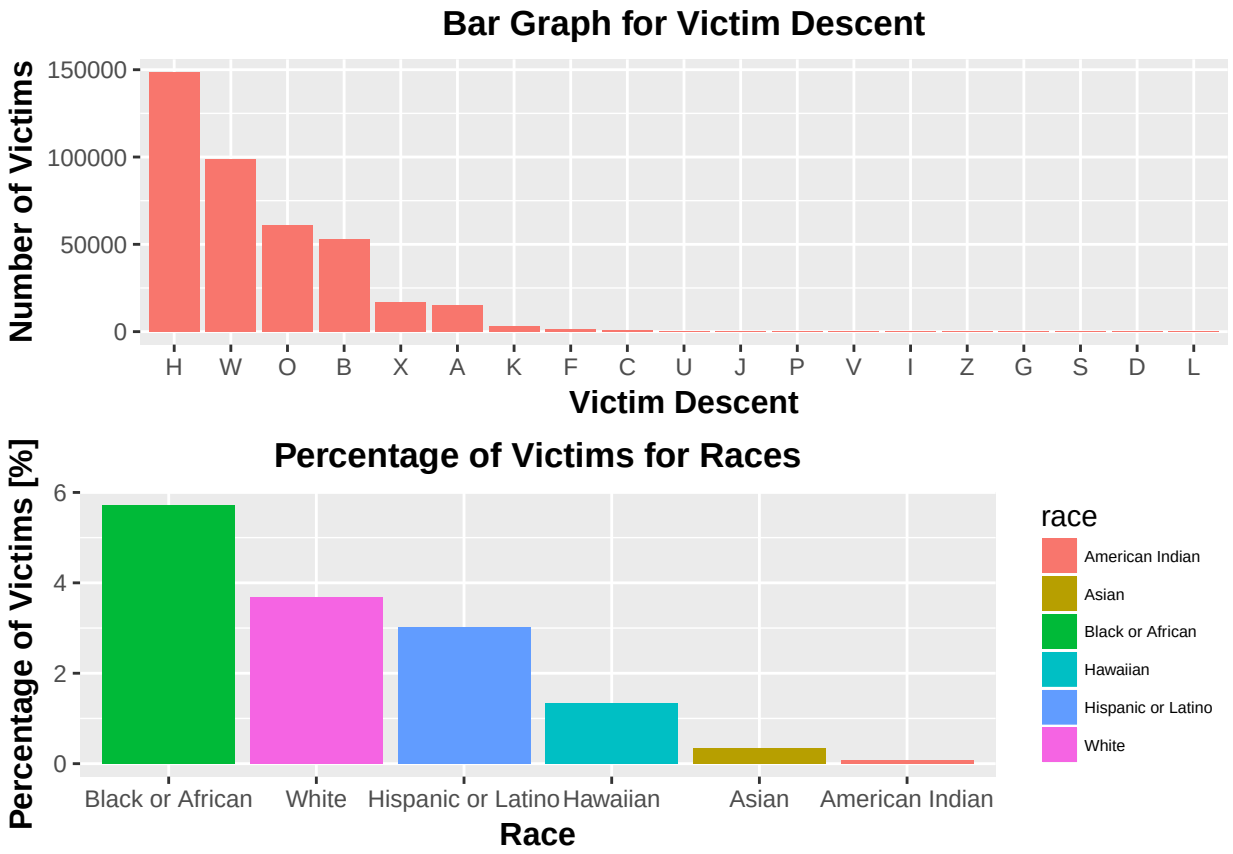
In order to get more depth and learn what other factors increase the number of collisions throughout the years, our team found supporting datasets on personal income per capita of Los Angeles from year 2010 to 2017. As the number of collisions increases, the amount of personal income per capita also 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.



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.

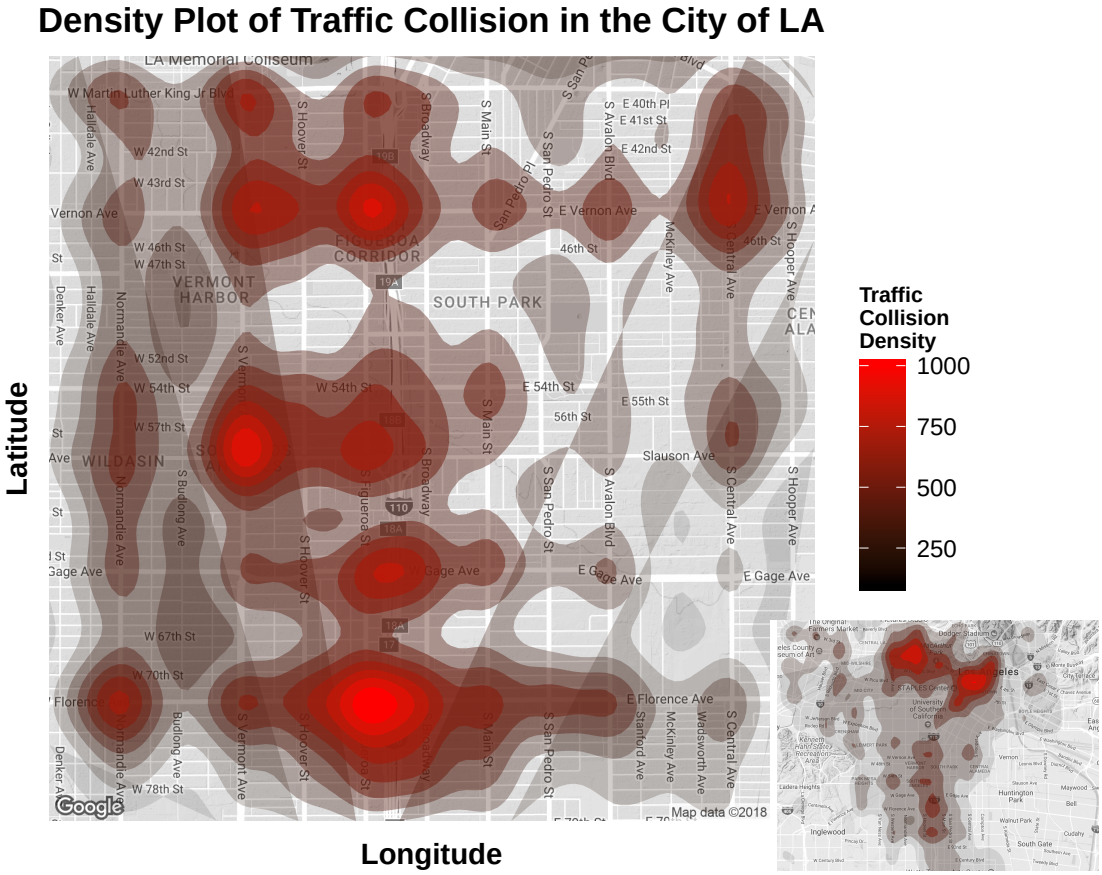


Previously, our team used bar graph to display total number of collisions by type of victim descent. We concluded that victim descent of Hispanic/ Latin/ Mexican has the highest number of collisions compare with other races. However, this conclusion can be misleading because higher number of collisions doesn't relate to higher likelihood of collision. The high number of collisions from Hispanic descent can be caused from high population of Hispanic descent in Los Angeles.

In order to improve the graph, 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.)



The main graph is a zoomed in density plot, and the small plot on the right corner is an overall view of the density plot. From the map density plot, we can see that traffic collisions happen near the downtown area and on highways. Specifically, the highest density is on HWY 110. All the density spots locate on this highway, meaning that most victims.

Conclusion

Our team concluded that people of Black and African American descent are most likely to be victims from traffic collisions. Sex, descent, and age are not only the main factors that cause collisions, because personal income per capita and unemployment rate are also strongly correlated with collisions. More personal income and employment rate increase the number of collisions.