USC Marshall School of Business MBA

To: **Professor Abbass Sharif, PhD**
From: **Cheng (Xenia) Atkeisson, Justine Chao, Chris Dong, Hannah Kim, Parin Patel**
Date: **May 8, 2018**
Subject: **Statistical Computing and Data Visualization Final Report – LA Crimes**

Executive Summary

As graduate students who spend a majority of our week on USC's campus and/or in surrounding Los Angeles neighborhoods, we wanted to better understand the key trends concerning incidents of crime in Los Angeles: how have crime rates changed over time? Is there a difference between volume of violent and non-violent crimes? What is the makeup of crime victims (gender, age, race)? How have these trends played out in the areas immediately surrounding USC?

While our report will explore each of these questions/topics in greater detail, a few highlights from our analysis are below:

- Incidents of crime are rising in Los Angeles, California, with total numbers of violent crimes reaching nearly 12,500 in 2017.
- Although robberies account for the greatest percentage of violent crimes in Los Angeles year-over-year, incidents of rape have increased significantly from 2010.
- On weekdays, 8:00 - 10:00 PM are the most dangerous hours to be out in Los Angeles; however, on weekends, 12:00 - 2:00 AM is the most perilous window.
- Males between 20-25 years old are the biggest victims of violent crimes in Los Angeles, but males and females exhibit similar trends in terms of violent crimes by age group.
- African Americans remain the most frequently victimized by violent crimes in Los Angeles, which has been a constant trend since 2010.
- While Hispanic/Latino citizens account for half of the Los Angeles population, the data reveals that their exposure to violent crimes has remain relatively flat.
- Incidents of violent crime near USC's campus increased in 2011, whereas violent crimes in broader Los Angeles decreased.
- Violent crimes in the area surrounding USC appear to have decreased overall from 2010 levels, while non-violent crimes appear to be increasing. Perhaps there is a link between the increasing population and the rising number of crimes.
- Despite the deployment of a campus fence and additional security protocols, robberies at USC actually *increased* in 2017, which correlates with an increasing trend in crime throughout the city from 2013 to 2017.

To help city officials and our fellow students visualize a few of these key trends and toggle between different output views (by changing select input variables), we created a dashboard of apps (using Shiny) – the key features of each app are included below:

- Users can toggle between Males and Females to see how the breakdown of violent crimes varies by age (and of course, gender).
- Users can select various crime types to see a breakdown of violent crimes by race

The visualizations and supporting analysis of crime incidents in Los Angeles (and USC) present broad opportunities for city officials to incorporate data into their decision-making regarding the safety of the city. Leveraging the data and our heatmap, we can predict the safest times of any day to move about the city and advise citizens as to when they need to be more alert/vigilant for potential perpetrators. Perhaps more importantly, this data and its accompanying visualizations + app can be utilized by the city and the Los Angeles Police Department to better understand its staffing and deploy resources to problem areas at peak times.

If we were to take this forward in the future, we would look to see how many police officers are deployed per neighborhood and the resulting impact on incidents of crime by neighborhood. Additionally, we would overlay our findings with median household income information to analyze any resulting correlations.

Introduction

As the most densely populated city in the United States, Los Angeles is home to over four million people and stands out as the "dominant population center in California" according to the Los Angeles Times. What exactly does this mean for crime rates in the city, and how have those rates changed over time?

The answers to these two questions have profound implications for us as both city dwellers and USC graduate students, who spend countless hours each week in this metropolis. Our final report examines incidents of crime (violent and non-violent) in the city of Los Angeles and analyzes how crime rates and categories have changed over time, how crime has specifically changed over time in the areas immediately surrounding USC, and whether or not the construction of the perimeter wall was in fact effective in improving campus safety, as well as the makeup of Los Angeles crime victims.

To better understand and discern any key trends regarding incidents of crime, we sought to create visualizations (area maps and bar graphs) of the amount of crime occurring in Los Angeles, segmenting between violent and non-violent crimes; we also wanted to determine if crime in Los Angeles occurs more frequently on certain days of the week; additionally, we sought to uncover key trends regarding crime victims in Los Angeles (i.e. what is the breakdown of victims by age, gender, and race); finally, we wanted to see if we might be able to draw parallels or potential correlations between crime activity in Los Angeles and in the surrounding neighborhoods around USC's campus.

Given the increasing population, we postulate that crime in Los Angeles is also on the rise, with non-violent crimes accounting for a greater percentage of that growth. Moreover, we speculate that younger citizens (18+) account for a greater percentage of violent crimes and that crime rates in Los Angeles will certainly be correlated with those around USC. We expect that our research can help to inform the city's decisions regarding the deployment of law enforcement officials.

Data Description

Our dataset was incredibly robust, with 26 separate columns and 1,722,911 rows of data. We recognized that we would need to clean the data so that it would be in a usable format. Our original dataset had crime data from the city of Los Angeles between 2010 to Present. The data is updated weekly and our team pulled the data on April 24, 2018. The dataset can be retrieved HERE. The columns contain the following data attributes:

Column Name	Description
DR Number	Division of Records Number: Official file number made up of a 2 digit year, area ID, and 5 digits
Date Reported	MM/DD/YYYY
Date Occurred	MM/DD/YYYY
Time Occurred	In 24 hour military time. 3 or 4 digit number, with 1 or 2 digit as hour and last 2 digit as minute.
Area ID	The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. These Geographic Areas are sequentially numbered from 1-21.
Area Name	The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible for. For example 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles.
Reporting District	A four-digit code that represents a sub-area within a Geographic Area. All crime records reference the "RD" that it occurred in for statistical comparisons.
Crime Code	Indicates the crime committed. (Same as Crime Code 1)
Crime Code Description	Defines the Crime Code provided.
MO Codes	Modus Operandi: Activities associated with the suspect in commission of the crime. See attached PDF for list of MO Codes in numerical order.
Victim Age	Two character numeric. Some rows have no Victim Age information
Victim Sex	F - Female, M - Male, X - Unknown. Some rows have no Victim Sex information
Victim Descent	Descent Code: A - Other Asian B - Black C - Chinese D - Cambodian F - Filipino G - Guamanian H - Hispanic/Latin/Mexican I - American Indian/Alaskan Native J - Japanese K - Korean L - Laotian O - Other P - Pacific Islander S - Samoan U - Hawaiian V - Vietnamese W - White X - Unknown Z - Asian Indian. Some rows have no Victim Descent information
Premise Code	The type of structure, vehicle, or location where the crime took place.
Premise Description	Defines the Premise Code provided.
Weapon Used Code	The type of weapon used in the crime.
Weapon Description	Defines the Weapon Used Code provided.
Status Code	Status of the case. (IC is the default)
Status Description	Defines the Status Code provided.

Crime Code 1	Indicates the crime committed. Crime Code 1 is the primary and most serious one. Crime Code 2, 3, and 4 are respectively less serious offenses. Lower crime class numbers are more serious.
Crime Code 2	May contain a code for an additional crime, less serious than Crime Code 1.
Crime Code 3	May contain a code for an additional crime, less serious than Crime Code 1.
Crime Code 4	May contain a code for an additional crime, less serious than Crime Code 1.
Address	Street address of crime incident rounded to the nearest hundred block to maintain anonymity.
Cross Street	Cross Street of rounded Address.
Location	The location where the crime incident occurred. Actual address is omitted for confidentiality. XY coordinates reflect the nearest 100 block. The format is (lat:lon).

Data Cleaning

1. The first step we performed was to select only the useful columns in a new variable called ***clean_data***. This decreased the number of columns to 8 from 26.
2. Then, we took the *Time.Occurred* variable and formatted it via **stringr** to ensure it had the 4-digit HH:MM format.
3. Afterwards, we formatted the dates for Date.Reported and Date.Occurred to have the yyyy-mm-dd format.
4. We then mutated the dataset to include *Hour* and *Weekday* columns so that we could create a heatmap
5. Next, we removed the 2018 data because the year is currently in progress, so we don't have a full year's worth of data yet.
6. Then, we formatted the location column to create two separate columns, *Lat* and *Lon*, for latitude and longitude.
7. We created a column for crime category, *Crime.Category*, for non-violent vs. violent crimes.
8. We also removed the "NA" data from fields with no data.
9. Finally, we set up a boundary two blocks around USC's campus to see how many crimes occurred at or near the university.

Exploratory Analysis

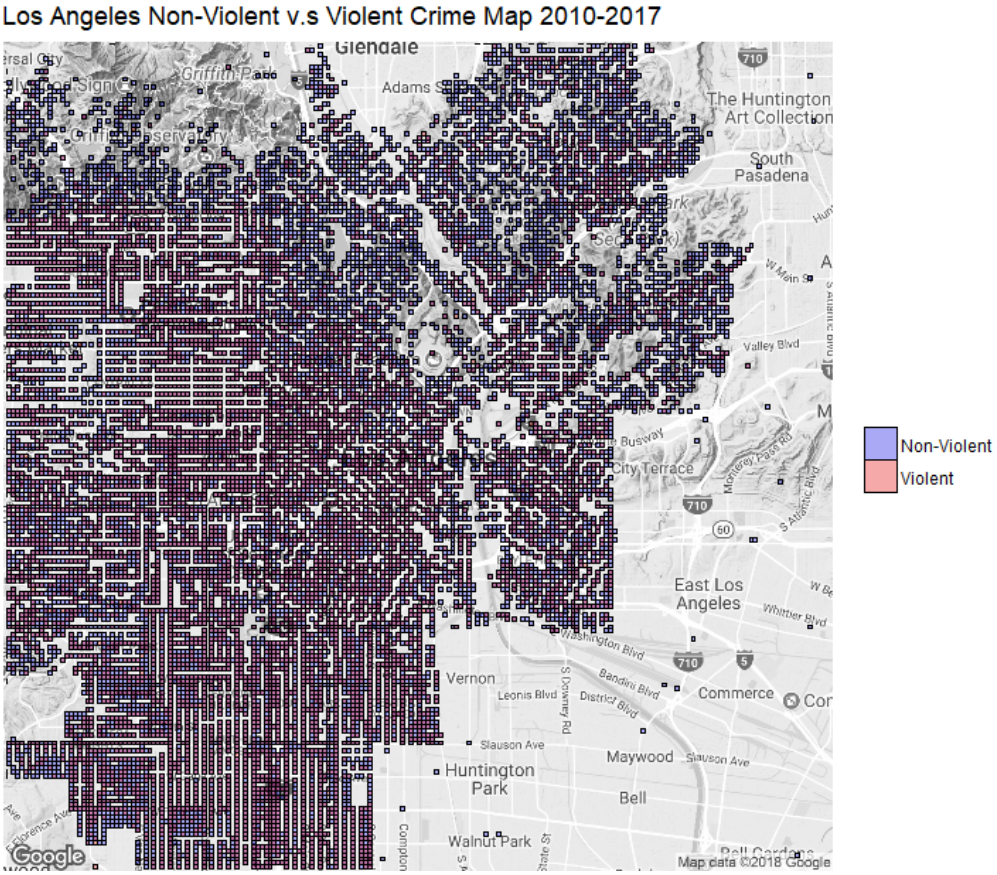
Our initial analysis aimed to see how much crime occurred both in Los Angeles and near USC, and we segmented crimes between violent and non-violent offenses. Violent crimes were listed as:

- ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER
- ASSAULT WITH DEADLY WEAPON
- AGGRAVATED ASSAULT
- ATTEMPTED ROBBERY
- CHILD ABUSE(PHYSICAL)-AGGRAVATED ASSAULT
- CHILD ABUSE(PHYSICAL)-SIMPLE ASSAULT
- CRIMINAL HOMICIDE
- MANSLAUGHTER, NEGLIGENT
- OTHER ASSAULT

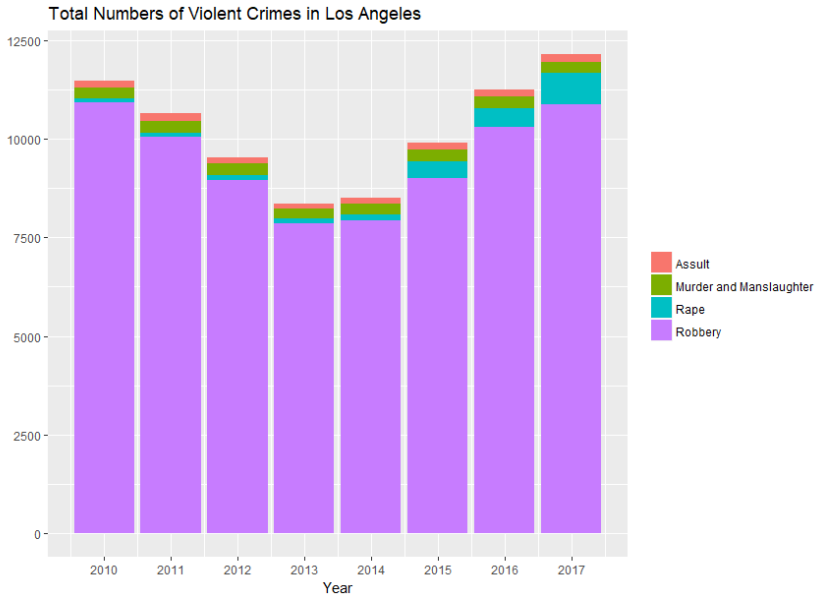
- RAPE, ATTEMPTED
- RAPE, FORCIBLE
- ROBBERY
- SEXUAL PENETRATION W/FOREIGN OBJECT
- SEXUAL PENETRATION WITH A FOREIGN OBJECT

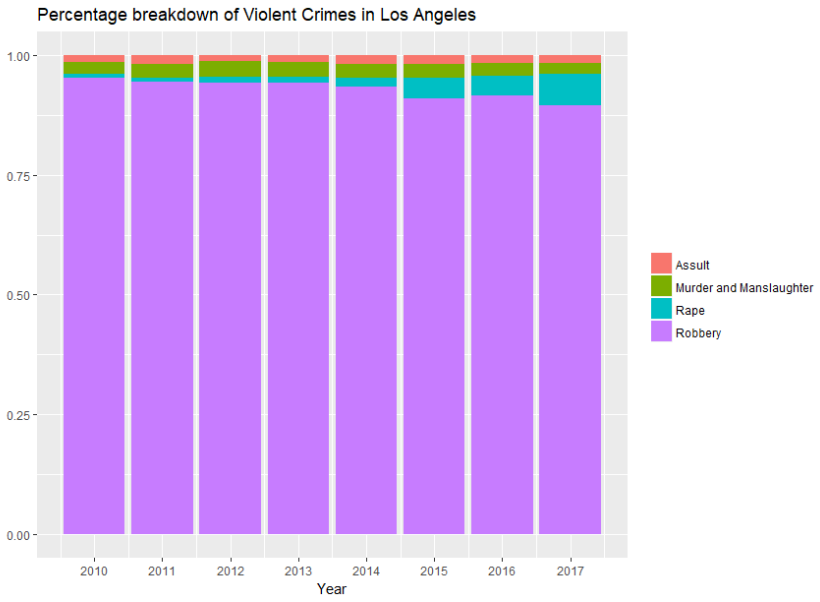
The rest of crimes were considered as non-violent crimes.

We then mapped all of the crimes in Los Angeles from 2010-2017:

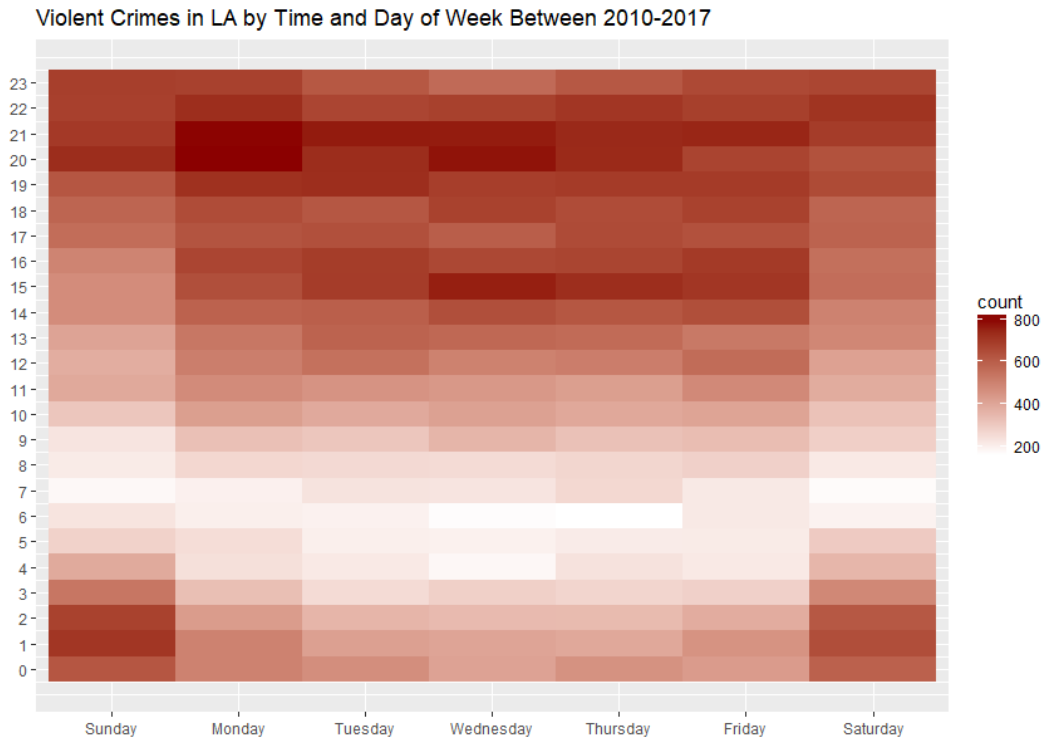


Based on the data, we determined that the map had too many data points. We decided to look at a bar plot of the data to see violent crimes in greater detail, both in terms of number and percentage of incidents:

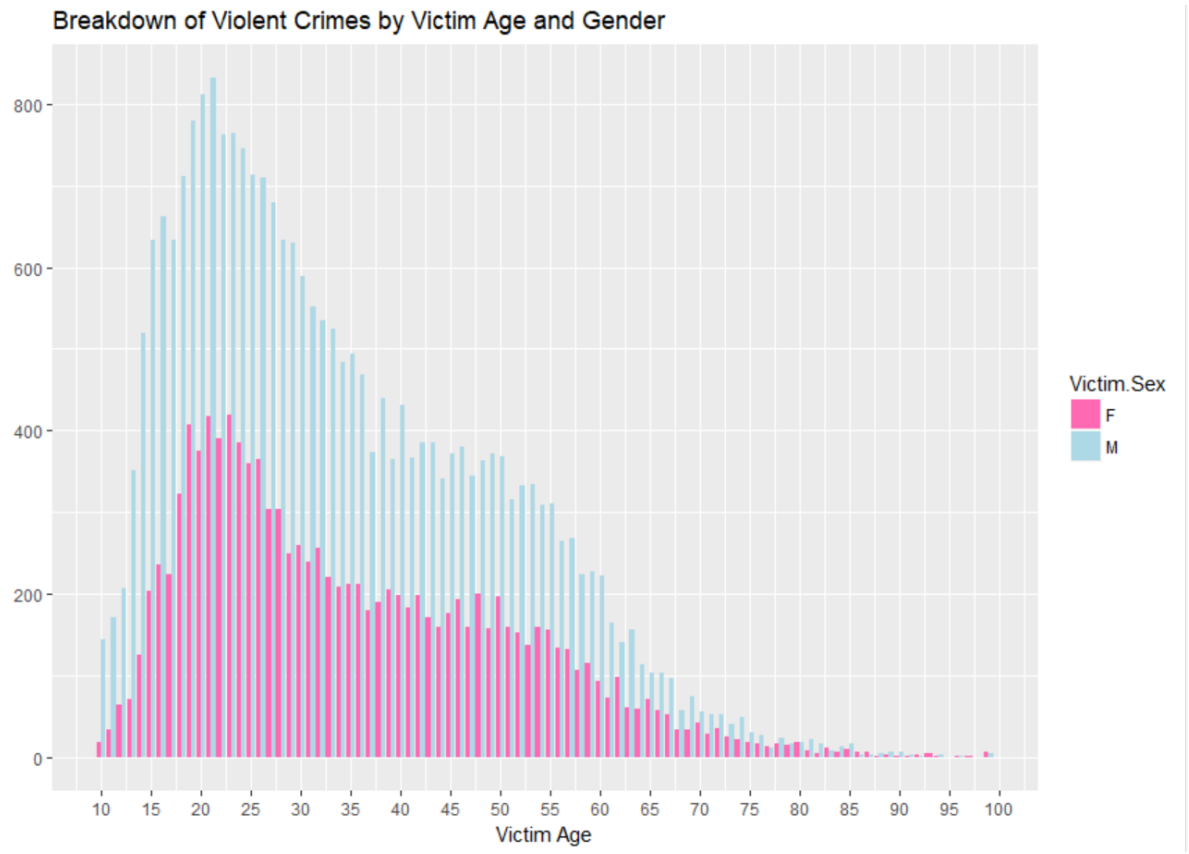




We then looked at violent crimes based on the time and day of week to see if we could find trends in safe vs unsafe times within the city. Based on our results, it appears that during weekdays, 8:00-10:00pm are the most dangerous hours to be out in Los Angeles, and 12:00-2:00am are the most dangerous hours on weekends.

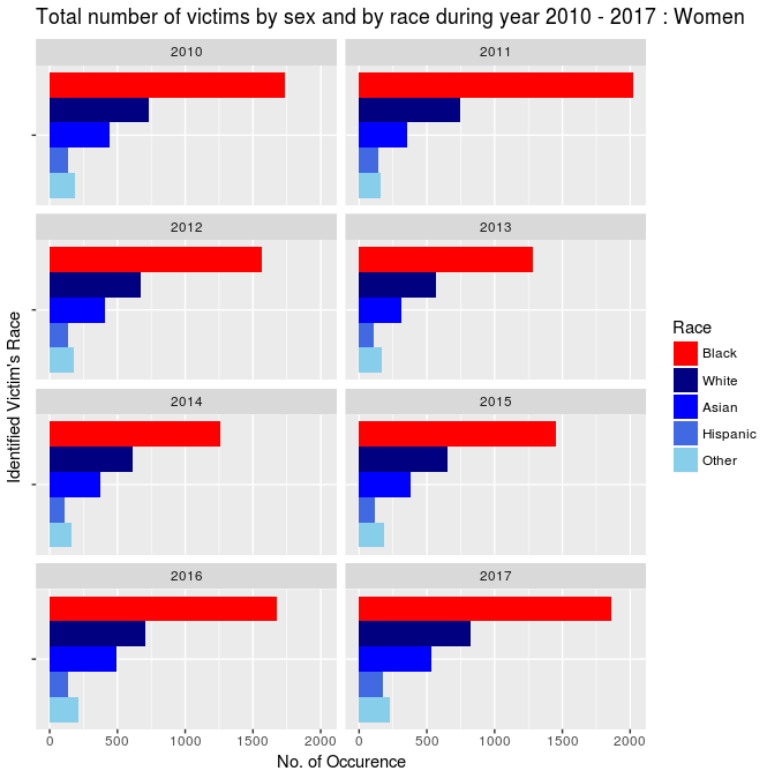


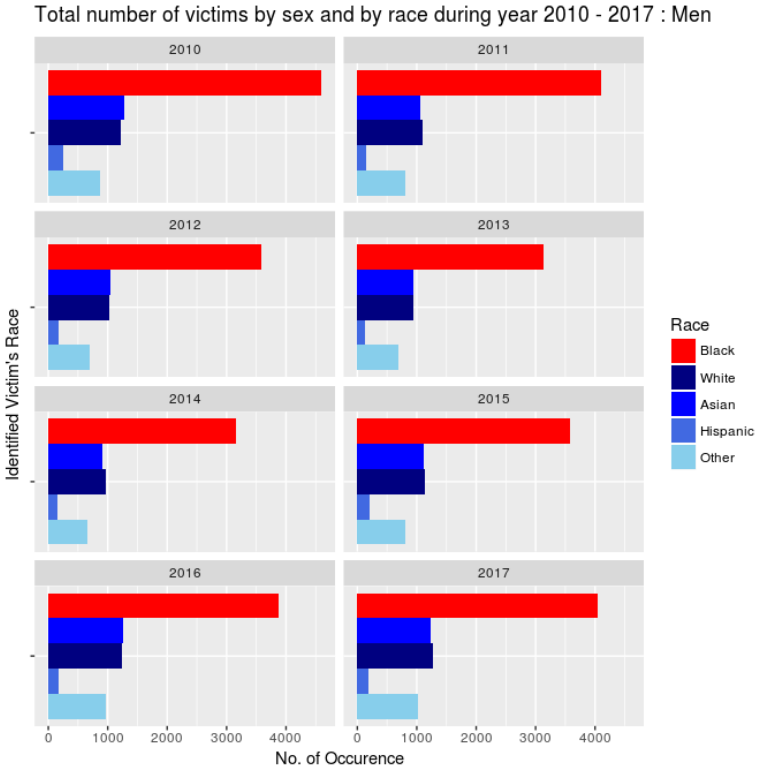
Next, we wanted to see the breakdown of violent crimes in Los Angeles by age and by gender. A dodged bar plot would allow us to create a visualization of the split (with victim ages on the X axis, dodged according to gender).



The main takeaways from our chart were that males between 20-25 are the biggest victims of violent crimes. That being said, both males and females exhibit very similar trends in terms of violent crimes by age group, which was a bit surprising to us at first. This suggests that citizens age 18-25 were the primary victims of violent crimes in Los Angeles, and these rates appear to decrease with time/increases in age.

In addition to the above, we wanted to identify which race was most vulnerable to violent crimes and track this trend from 2010-2017. A horizontal bar chart with years on the X axis and fill by race would allow us to clearly visualize this trend (see plots below):

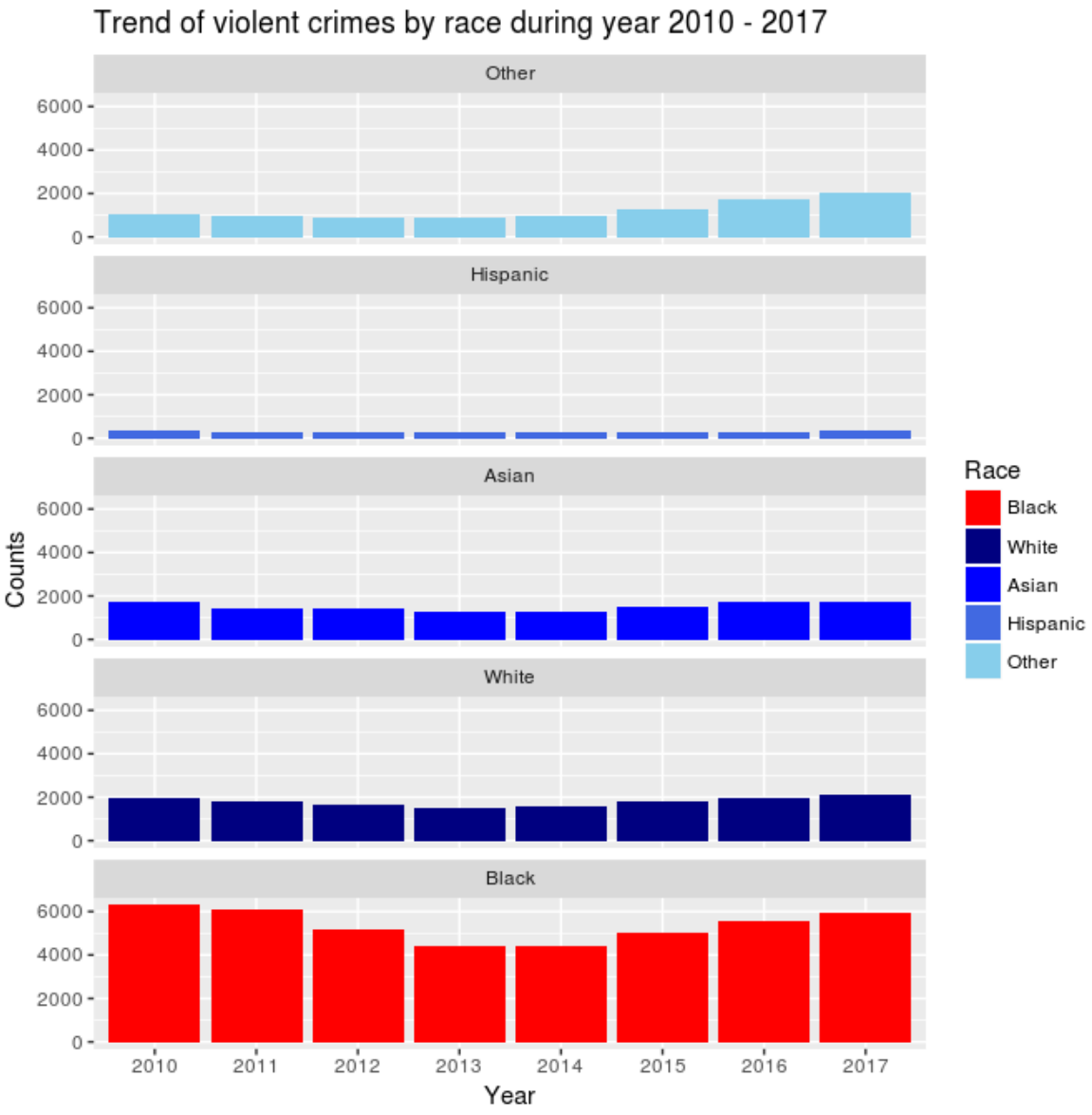




The main takeaway from our chart is that African American men and women are the biggest victims of violent crimes. Both Males and Females exhibit very similar trends in terms of violent crimes by

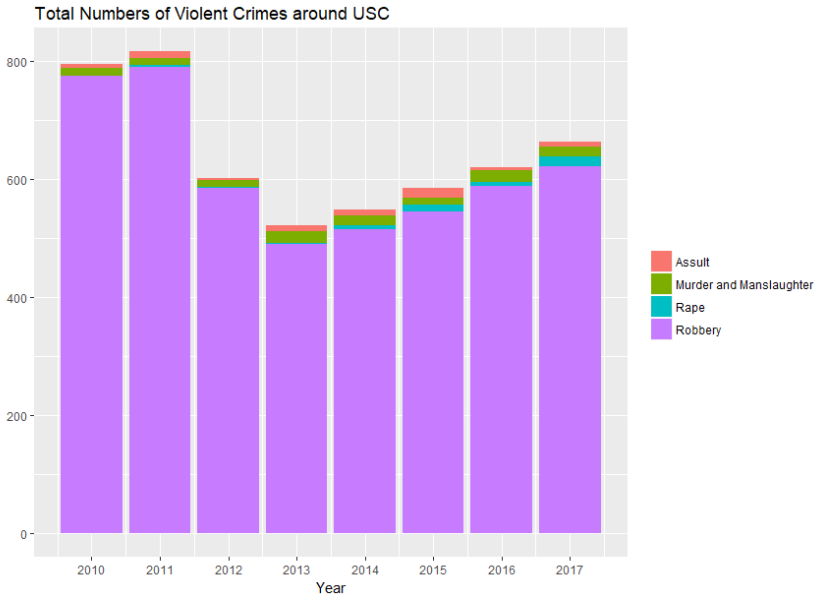
race in order; however, there does appear to be greater instances of White female victims than White Male victims, especially in comparison to the other races pictured.

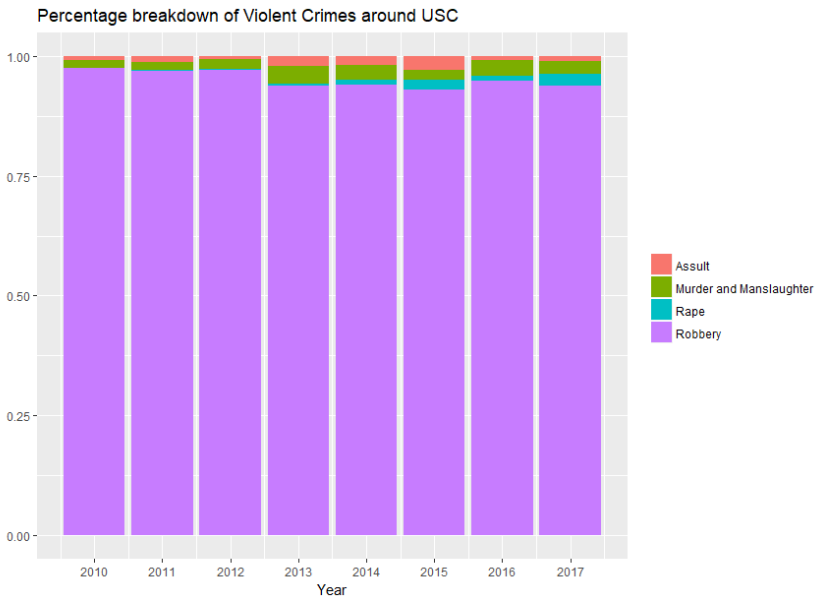
According to the 2005-2009 US Census, Los Angeles is comprised of the following races: Hispanic 47.5%, White 29.4%, Asian 10.7%, Black 9.8%, Others 25.2% (https://en.wikipedia.org/wiki/Demographics_of_Los_Angeles). Hispanic/Latino accounts for half of Los Angeles' population, but the crime data reveals that their exposure to violent crimes has remained relatively flat over the years. Expanding our research to include all crimes exhibited similar results: Hispanic/Latino victims were the lowest reported.



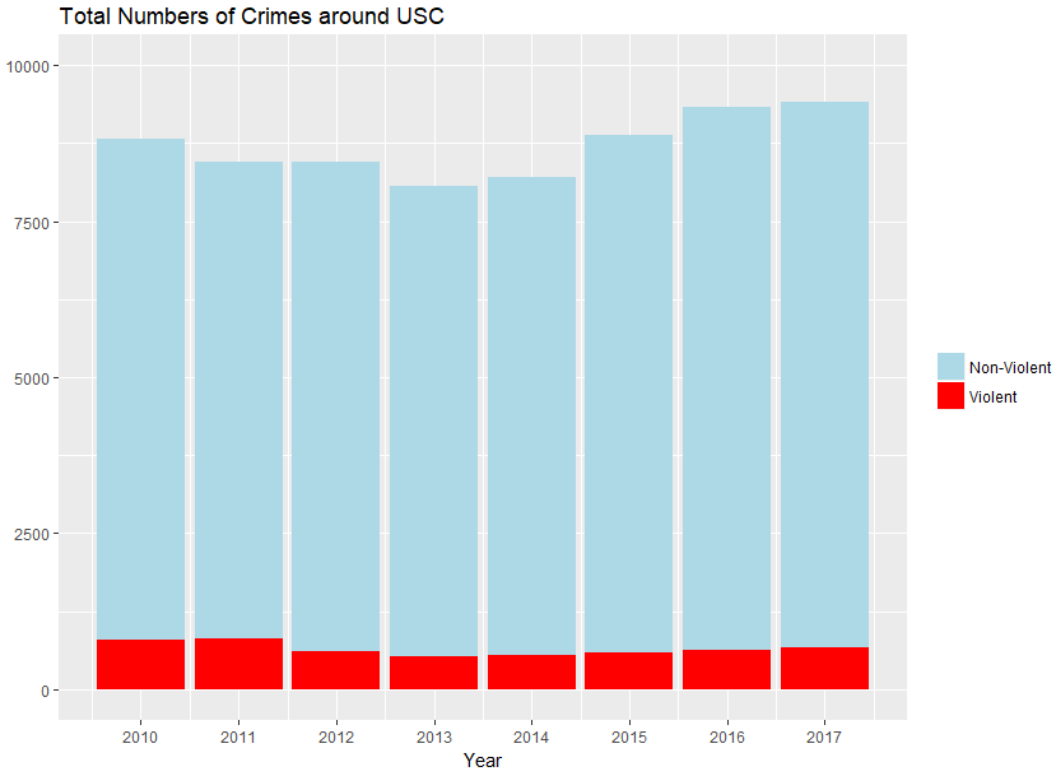
In addition to the above, we plotted the occurrence by race to see the trends over time (2010-2017). Interestingly, in 2013 and 2014, the number of violent crime decreased across all racial groups. Beginning in 2015, however, violent crimes across all races (except Hispanic/Latino) began to rise in 2015 and continued their upward trend over the last two years.

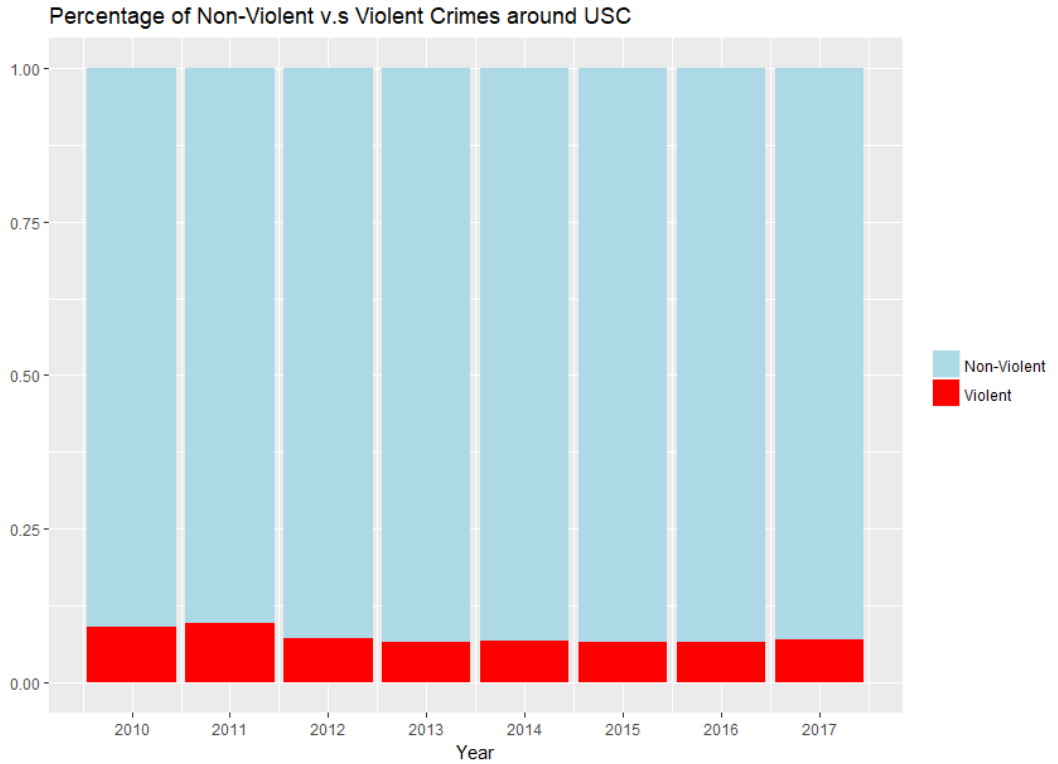
We then wanted to look at a specific region near USC to see how the data correlated with LA overall. Incidents of violent crime near USC's campus increased in 2011, whereas violent crimes in broader Los Angeles decreased.





We also looked at a breakout comparison between violent and non-violent crimes:

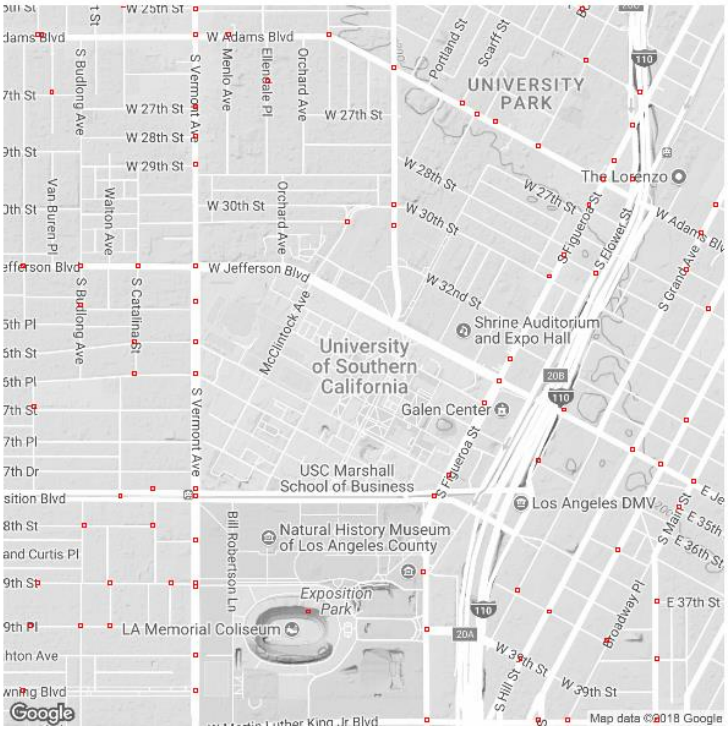

Total Numbers of Crimes around USC
10000
7500
5000
2500
0
2010
2011
2012
2013
2014
2015
2016
2017
Year
Non-Violent
Violent

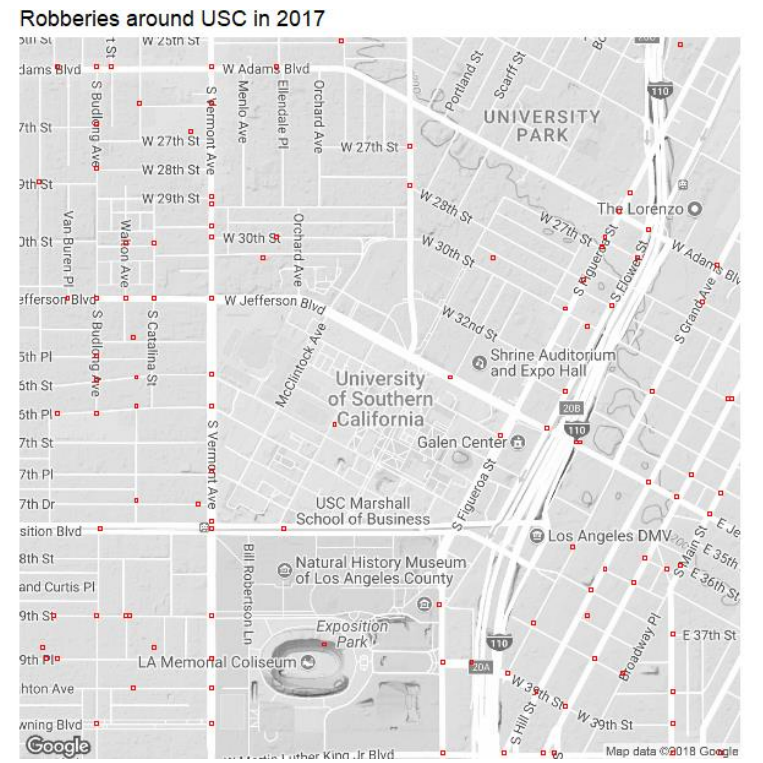

Percentage of Non-Violent v.s Violent Crimes around USC
1.00
0.75
0.50
0.25
0.00
2010
2011
2012
2013
2014
2015
2016
2017
Year
Non-Violent
Violent

Violent crimes appear to have decreased overall from 2010 levels, while non-violent crimes appear to be increasing. Perhaps this can be linked to the increase in population in Los Angeles – it could be that a correlation exists between rising instances of crime and the growing population.

We also wanted to focus on a troubling issue near campus: robberies. Given that the fence surrounding USC's campus and additional security measures were rolled out in 2013 – resulting from unforeseen crimes that affected USC students (homicide in 2012 and on-campus shooting in 2012) – we decided to look at the difference in robberies in 2013 and 2017 to examine the impact of these safety measures:

Robberies around USC in 2013



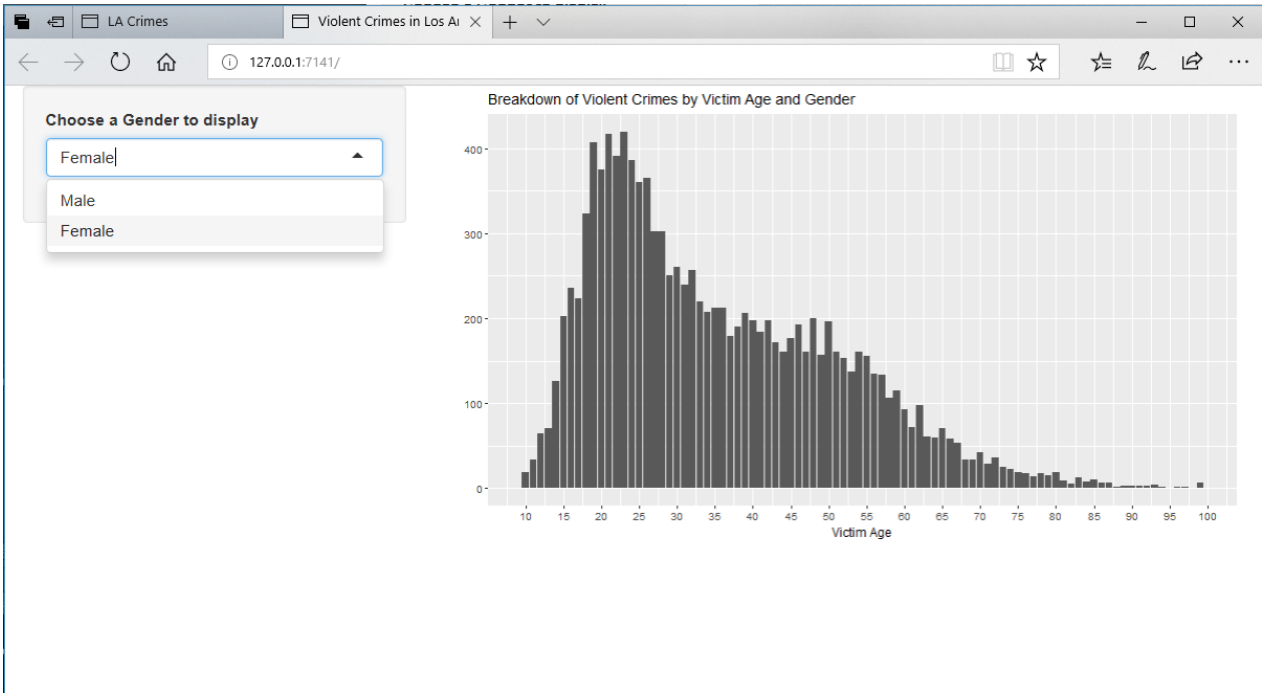


Surprisingly, robberies actually *increased* in 2017, but this also correlates with an increasing trend in crime throughout the city from 2013 to 2017.

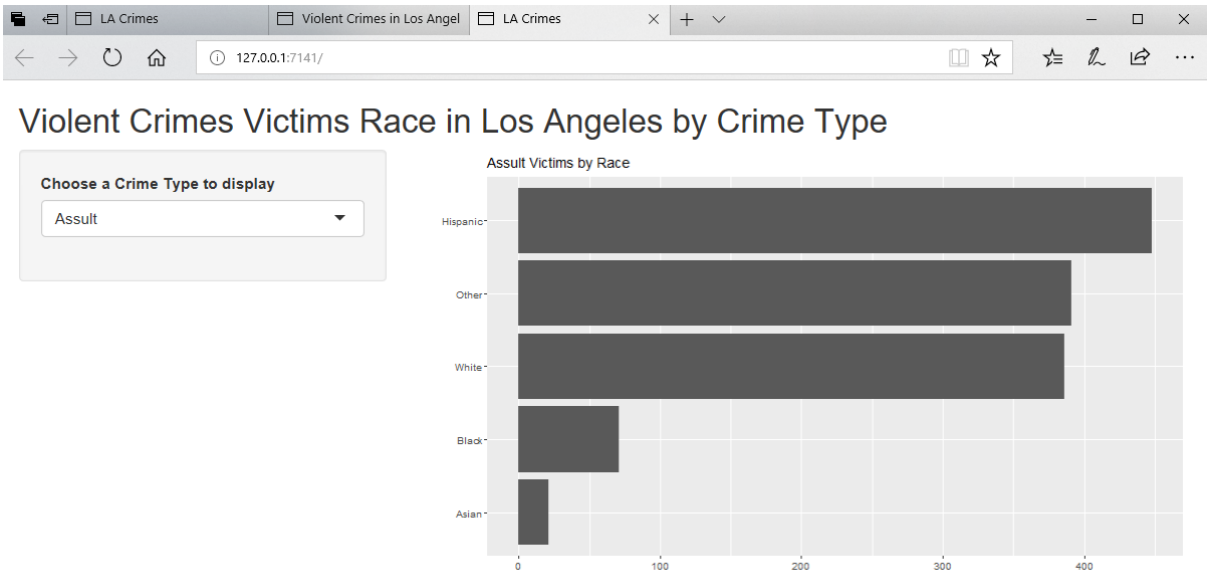
Shiny Dashboard

The dashboard we've created can be used primarily to toggle between different data output views by selecting the relevant input data/variable.

To distinguish between Male and Female victims of violent crimes more clearly, we created a Shiny App that allows users to toggle between Males and Females (on the side bar panel) to see how the breakdown of violent crimes varies by age, and of course, gender (as displayed on the main panel). A screenshot of the app appears below:



We also created a Shiny App to showcase the breakdown of violent crime victims by race and crime type. Users can select various crime types from the drop down menu of the side bar panel to see a breakdown of violent crimes by race (as displayed on the main panel). A screenshot of the app appears below:



Conclusion / Future Work

Our dataset gave us insight into Los Angeles in terms of the number of crimes, which crimes are the most common (robberies), the most common times for crimes to occur, victim classifications, and the areas of the city where crimes occur most often.

Based on the data, we can actually predict the safest times of day through the work week and weekend. This can make it easier for citizens to know when they need to be more vigilant against potential crimes and perpetrators. More importantly, the city of Los Angeles, specifically the Los Angeles Police Department, can use this data to get a better understanding of staffing for problem areas and peak time periods.

We have also noticed that crime rates across the city have gone up, so we were not surprised that crime around campus has increased, despite more neighborhood patrolling by CSC guards and the installation of the perimeter wall around USC. This may be correlated with the fact that the population in LA has risen steadily over the past 5 years.

In the future, we would like to plot the correlation between the median household income from US Census data and the crime occurrence per race and the area from Los Angeles Crime Data. We would also like to look into how many officers per neighborhood may affect crime in areas with high rates of violent and non-violent crimes.