Final Report: LA Crime

LeGrand Dudley, Angel Miranda-Acost, Rui, Xu, Rajpal Bal, Jacob Eguia

December 1, 2022

1 Introduction

As city populations have expanded in the 21st century, a new-found problem in the United States is the prevalence of crime. Despite efforts in the 20th century to combat crime rates through the "War-on-Drugs" program and prior administrations' "Tough on Crime" approach, a significant amount of crime is present. While many look at specific cities suffering from higher-than-usual amounts of crime (e.g., Chicago), other cities have their fair share of problems that may be consequentially related to the governance of the state and city.

1.1 What's in the Dataset?

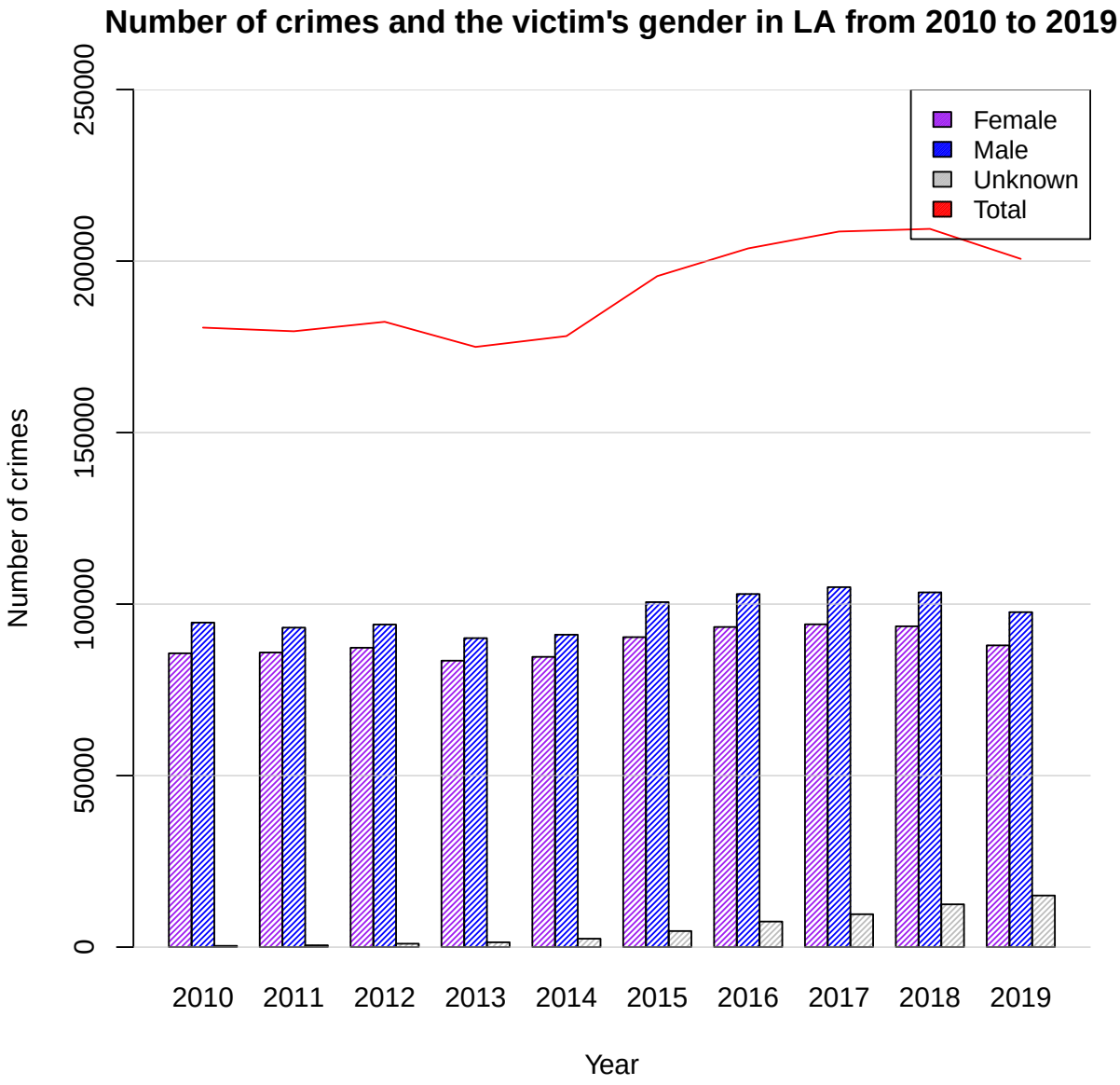
This project will attempt to separate the variables found in our dataset, 2010-2019 LA Crime Data (LAPD) in order to gain a better understanding of the distribution of reported crime in LA County. As per the data ranging from a time period of 9 years, the LAPD has sorted the data into a numerous amount of categories ranging from:

Time of Occurrence, Date of occurrence, Type of Crime, Crime Code (following LAPD code guidelines), Area, Victim Ethnicity, Victim Sex, Victim Age, Description of the Premise, Weapon Description, Status of the Case File, Location, Latitude, and Longitude.

Each of these variables have about 2 million responses collected, barring cases in which a response is missing (e.g., cases where a weapon aren't used wouldn't have a weapon description). As a result of the numerous amount of variables, we can create a few visualizations to better understand the data such as: Who are commonly targeted as victims of crime? Which region is frequently targeted? Do specific regions have differing frequencies over the years? Are specific ethnicities being targeted in specific regions? When are crimes frequently reported?

2 Breakdowns by Gender

2.1 Gender and time

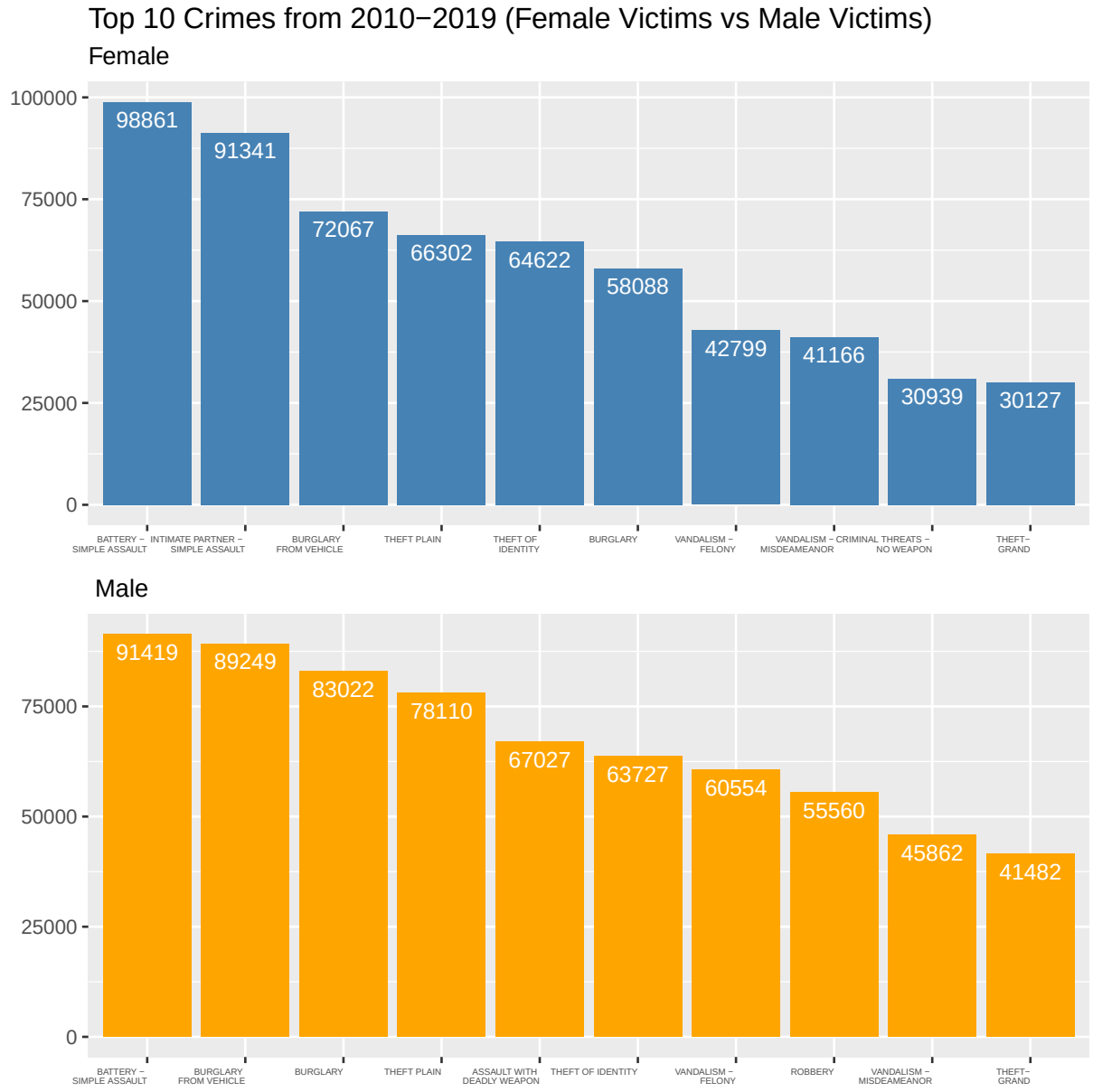


As seen by this graph, it is noticeable that men comprise of the majority of victims between 2010 and 2020 in LA and are followed by women. Reported crime tends to have decreased as a whole following the COVID-19 Pandemic in 2020, which is to be expected given the circumstances, but even with the aforementioned decrease in crime, men are still the majority of the victims. Non-binary and privacy-maintained individuals are subjected to the least amount of crimes by gender in LA county, but this can be attributed to their slim population in the U.S. as a whole. Some further analysis that we could do is breakdown gender and ethnicity by year and view political campaign support in LA county across 2010 to 2020. A recent focus within politics is the "tough on crime" campaign and if we can see the amount of times a political campaign focuses on it, it may describe the interests of the citizens of LA.

When we look at the grouped bar chart, we can see that men are more likely to be victims than women

from 2010 to 2019. This creates an interesting question pertaining to why? It could be that men are more comfortable engaging w/the outside world, placing them at greater risk to items like homicide. It can also be explained if we were to look at the population markup of LA County - if there are significantly more men than women, that may explain a bit of the difference between instances of perpetrations against women vs. men.

2.2 Difference of prepetrated crime by gender

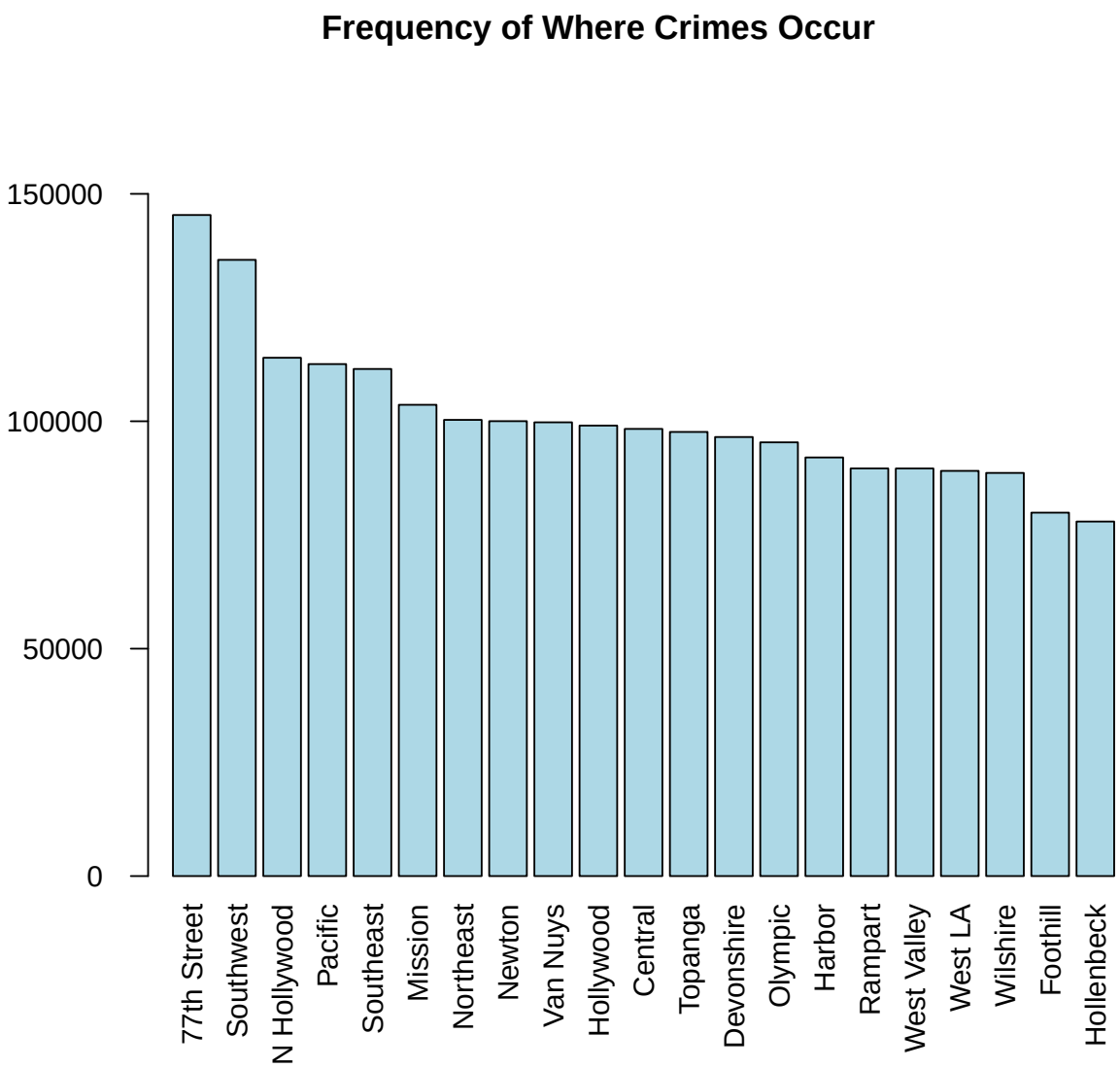


From the previous exploration, we found that men are more likely to be victims than women from 2010 to 2019. To discover some of the reasoning behind said finding, we listed the 10 most prevalent instances of crime against both females and males.

From the above bar plot, it's noticeable that the majority of crimes suffered by women are simple assault, burglary, theft, or threats with no weapon. However, for men, the number of crimes is more evenly distributed among different types. Even though the amount of the 2 most-occurring crimes suffered by men is slightly less than women, men are more likely to be involved in crimes such as assault with deadly weapons and robbery. In addition, the overall frequency of every subsequent crime (i.e., the 3rd most frequent, the 4th most frequent, etc.) is higher for men. Again, this may be due to how comfortable men are with engaging in the outside world, as women encounter threats of stalking, assault, misogyny, etc.

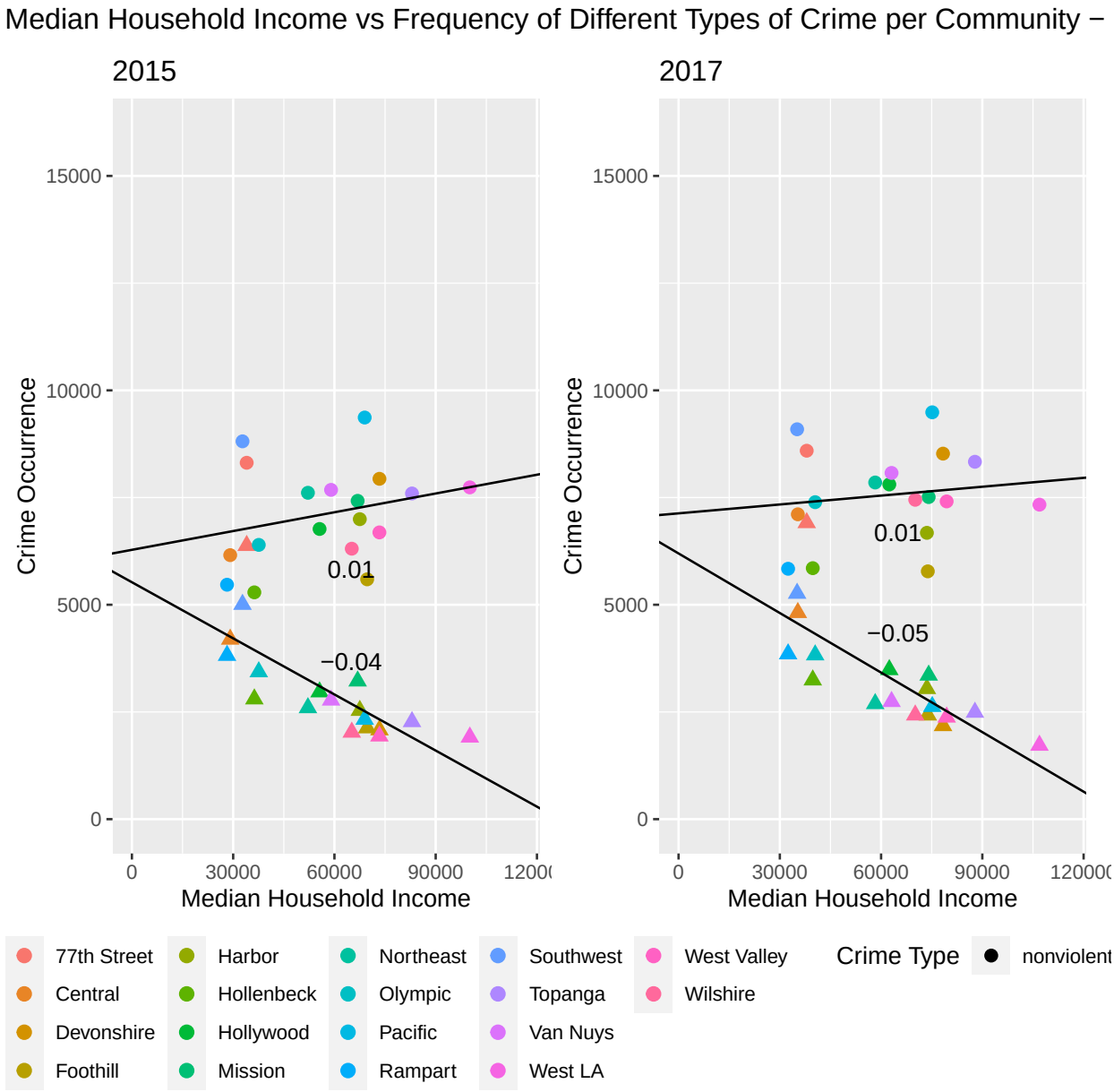
3 Breakdowns by LA Location

3.1 Frequency by Community



Above is a barchart that displays the frequencies of how many reported instances of crimes are within each area code in LA County. In short, this displays some areas of interest that can be described as "high areas of crime" in comparison to other areas. For example, 77th street could be further analyzed by measures of socioeconomic status to further analyze some potential reasonings as to why it's a relatively higher area of crime. In fact, using a dataset from the LA Almanac, we can create a graph that can showcase if there is a relationship between 1 measure of socioeconomic status - median household income per patrol district. If an inverse correlation exists between frequency of crime and income, it can imply that it is a significant predictor of troubled areas and can help governments attempt to combat said inequalities to quell crime.

3.2 Median community income vs. crime occurrence



Following the focus on viewing communities, we wanted to visualize if there was a relationship between criminal behavior in certain communities based on how wealthy a community is. Since the types of crime prevalent in communities often differ, we created a simple linear model to predict the amount of violent and nonviolent crime for a community based on the community's median income.

Focusing on the 2015 data, we can see that while there is a very weak correlation
(Adj Rsquared of 0.011
) between income and nonviolent crime, there is a much stronger negative correlation
(Adj Rsquared of 0.54
) between income and violent crime. The weak correlation would suggest that the income of a community is not associated with any variations in the frequency of nonviolent crimes. On the other hand, the negative correlation would suggest that the amount of violent crimes decreases as median income increases. Specifically for every additional dollar increase in median income of a community, there are 0.4 less occurrences of violent crimes.

For perspective, the average adult works about 1,801 hours a year. If they received a wage raise of 1/hr, they now make an additional 1,801 a year. For that increase, the relationship says that their community will experience 72 less instances of violent criminal activity in their community that year. Why would this be? Perhaps communities with greater wealth have access to better security, in the form of gated communities, neighborhood watches, and distance from lower-income communities.

3.3 Businesses in relative low-crime vs high-crime communities

From the city of Los Angeles, we obtained a dataset containing the names and locations of all operating businesses in LA County. With this dataset, we decided to see the distribution of specific business practices in the lowest-crime community and the highest-crime community community. In particular, we wanted to see if a high-crime community would contain more "predatory" practices that facilitate poverty traps - a cycle in which a low-income individual becomes unable to improve their position. In addition, we want to see ifhigh-income communities have greater access to recreational services for their personal leisure. Because they have a higher income, it'd enable them to take more days off, work less, and enjoy their leisure, and thus, have greater access to nearby leisure activities than their low-income counterparts.

Table 1: Frequency of types of businesses present

	77th Street	Hollenbeck
Health	128	201
Predatory	13	6
Education	99	57
Recreation	186	259

Based on the above table, we can see a few remarkable differences. For starters, 77th Street has lower access to Health and Recreational services and greater access to Predatory services, as compared to their high-income counterpart, Hollenbeck. Surprisingly, they also have greater access to local Educational services, something we did not expect. While a simple comparison, it does provide a bit of additional insight as to how these communities are treated.

For example, the table below showcases the first 5 results of Health for both 77th Street and Hollenbeck.

Table 2: First five health results

77th Street	Hollenbeck
THE PUBLIC HEALTH FOUNDATION OF LOS ANGELES COUNTY INC	MAKE YOUR HEALTH A PRIORITY LLC
BROWNE AND MURRAY LICENSED CLINICAL SOCIAL WORKER	RISE OF THE PHOENIX WHOLE HEALTH CENTER, INC.
DAMIAN HEALTH FOOD CATERING SERVICES LLC	NORTH FIGUEROA ANIMAL HOSPITAL, INC.
BARBARA CARE HEALTH SOLUTIONS INC	SIERRA HOME HEALTHCARE SERVICES INC
BEHAVIORAL HEALTH BILLER INC	SAINT BISHOY MEDICAL GROUP

As can be seen in the above table, 77th Street has health services, but they're not actual clinics or personal health related services; they're foundations, medical supply companies or a health billing group. However, the first five pulled from Hollenback has services towards improving physical health, two health services (one being for animals), a health center, and a medical insurance group. This dataset isn't organized in a particular fashion (e.g., alphabetically or by zip code), but rather by the company's registered business code. This may exemplify the availability of true health-related services in these areas.

Lastly, the table below showcases the first 8 results of Predatory practices for both 77th Street and Hollenbeck.

Table 3: First eight predatory-business results

77th Street	Hollenbeck
CALIFORNIA CHECK CASHING STORES LLC	PETTY CASH LLC
CASH CLOUD INC	CASH CLOUD INC
CASH CLOUD INC	MUNDO CHECK CASHING INC
MUNDO CHECK CASHING INC	KLM CASH LLC
PACIFIC CASH ADVANCE LLC	CASHBAK LLC
CASH B CONSULTANTS INC	PLS CHECK CASHERS OF CALIFORNIA INC
CASH CLOUD INC	NA
EMMETT CASH III	NA

Through the use of a text filter by searching "cash", "pawn", "check", "loan", and similar like terms, the results indicate that 77th Street has a greater availability and proportion of predatory services than Hollenbeck. In addition, Hollenbeck's results showcase businesses that facilitate cash flow/management rather than actual predatory services. 77th Street has cash advance, checking cashing, ez loans, ez cash, and similar poverty trap style business regimes, unlike Hollenbeck. While small and reliant on text filters, this supports the idea that low-income communities are exposed to poverty-traps that greatly expands the inequality of their citizens as compared to Hollenbeck - a higher income community

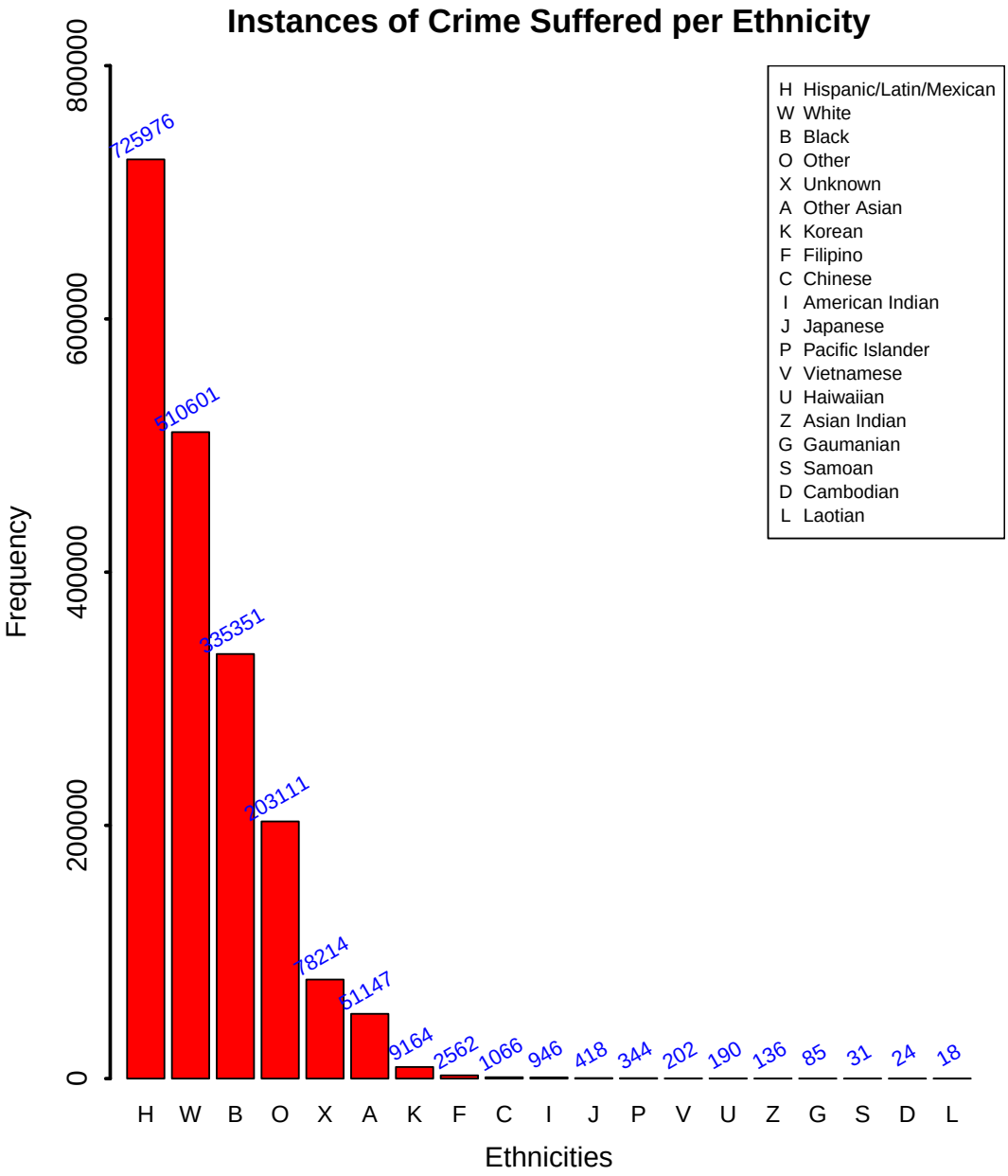
Table 4: Types of businesses present in low, median, and high-income communities

	Low Income Communities	Mid Income Communities	High Income Communities
Health	934	2201	2381
Predatory	44	68	94
Education	386	476	630
Recreation	1375	5323	6088

Lastly, we wanted to determine the frequency of these types of businesses across multiple low, median, and high income communities. Given that there are 21 communities, 7 of the lowest income communities are in low, and the same logic applies to the remaining categories. Viewing it as a whole, it seems the the lower income communities aren't disparately exposed to the aforementioned poverty traps, but they have significantly lower access to recreational and health related services compared to median and high income communities. This same relationship also seems to apply for education, but maybe not to a significant margin.

4 Breakdowns by Victim Ethnicity

4.1 Frequency committed against ethnicity



This graph conducts a different view to analyze the victim's ethnicities from the reported crimes. As has been seen in recent times, specific ethnicities are likelier to be victims of crime compared to other ethnicities. For example, minority ethnic groups, like Hispanics, can be targeted at greater rates compared to non-minority ethnic groups, such as Anglo-Americans, as shown in the above graph. Sorted in decreasing order, the graph shows which groups face the most targeted attacks among all citizens in LA county.

Given that the U.S. isn't known to be kind to ethnic minorities, it is interesting to specifically see which ethnic groups are targeted more often than others, with Hispanics/Latinos being the most targeted. While this graph only shows targeted ethnicities, a more in depth graph could show which ethnic groups faces specific criminal deeds. For example, are ethnic minorities more likely to be a victim of violent crime (assault, murder, manslaughter, etc.) than other ethnic groups? In addition, are ethnic minorities more

likely to be a victim in communities they're a majority in than other communities?

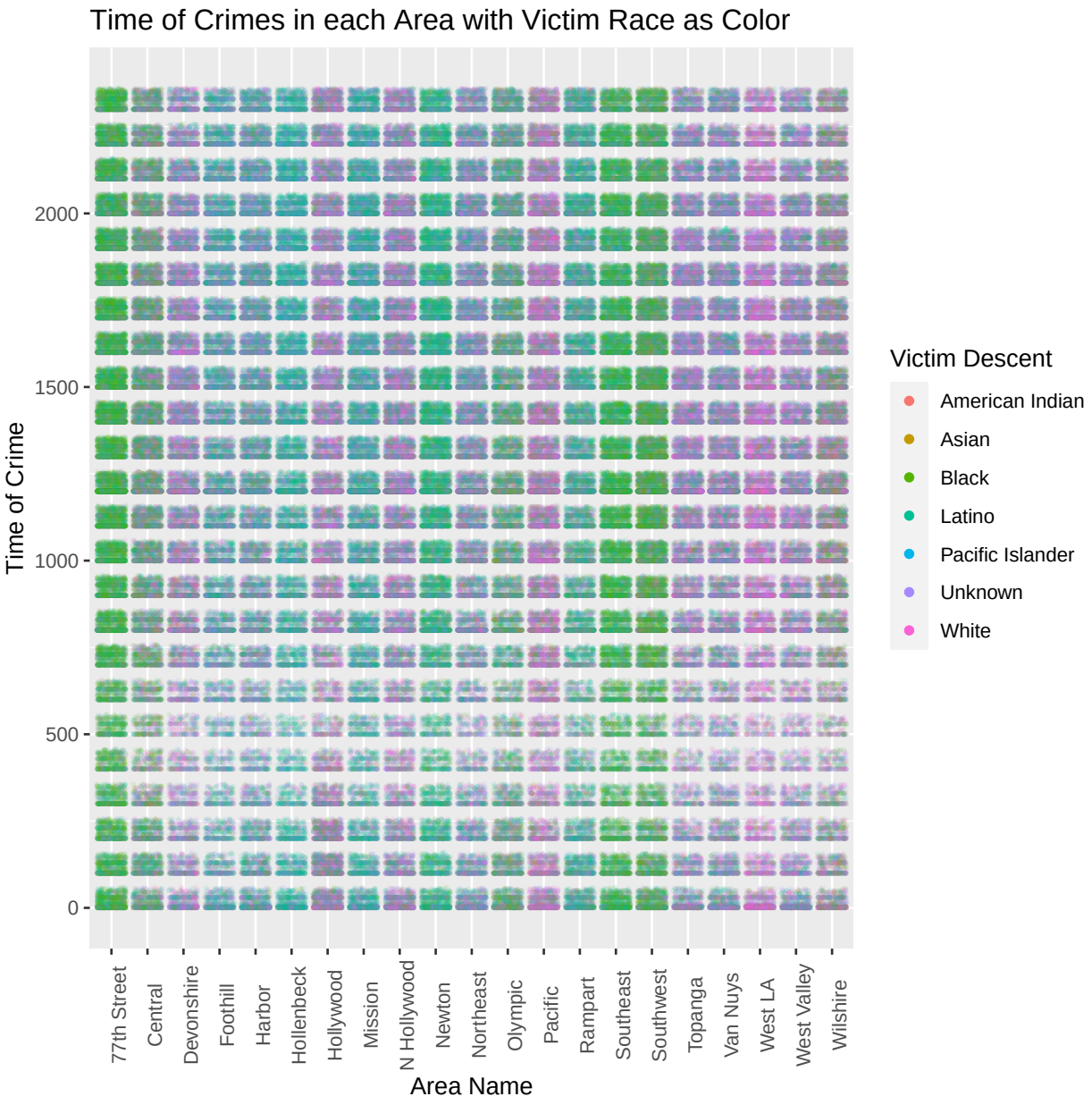
4.2 Victim's Ethnicity Evaluated by Time and Area

Below is a plot of the times each of the crimes occurred based on the area they were committed in. The color represents the descent of the victim and the columns represent crimes in an area, and the height of a point of crime is when it occurred in local military time. Any variation in the x direction within a column does not represent anything and is simply for clarity in the density of crimes.

Around the 5:00 am mark, the graph gets lighter, indicating that across all areas, there is less crime in the early morning. It is clear that each area has a race that has the most victims. For example, 77th Street and Southeast both are majority green, meaning that the victims in those areas are predominately African-American. The density of the bars represents how often crime occurs in the areas relative to other areas.

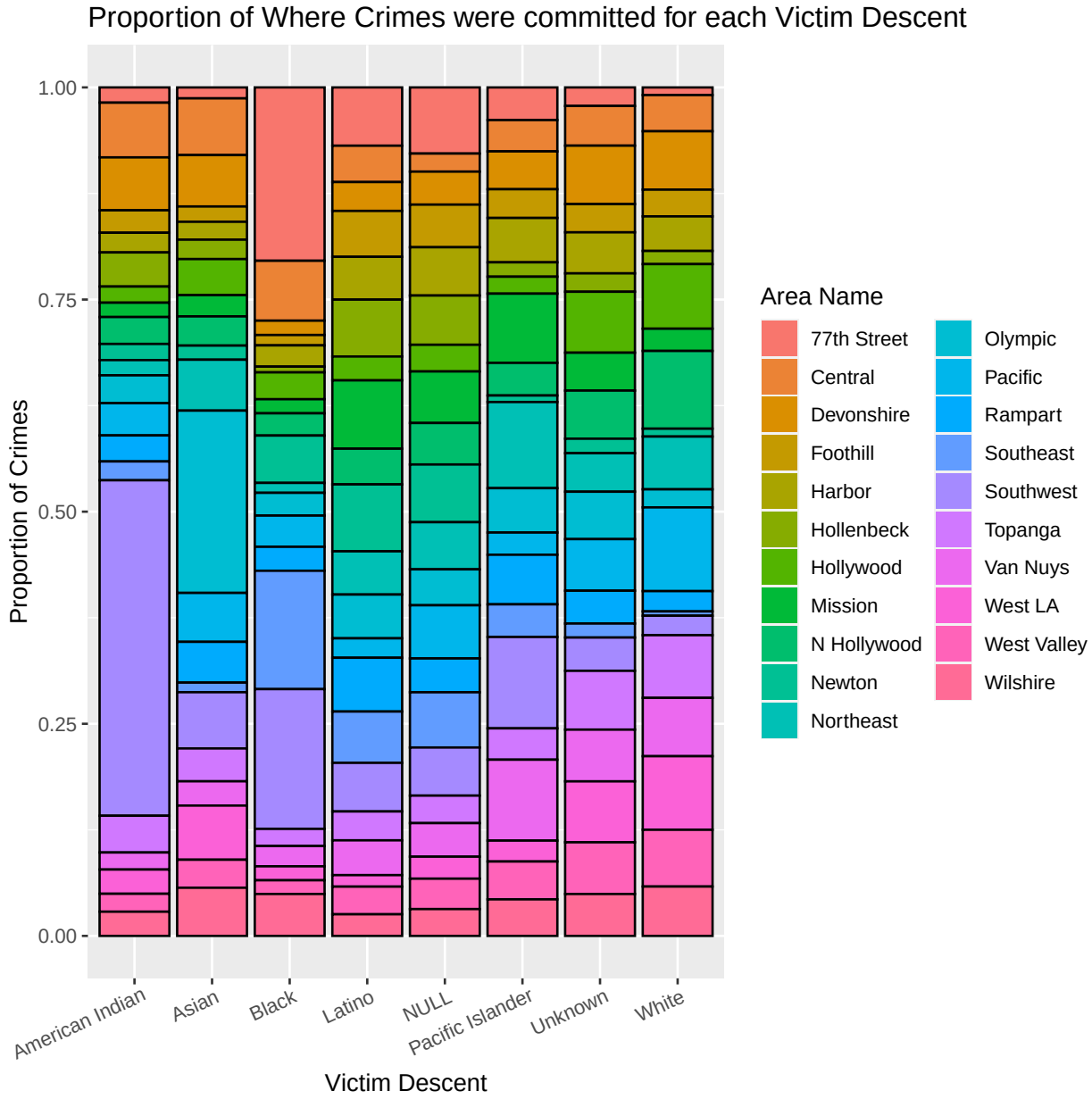
What is striking about this graph is that many of the areas have a single descent that has the most crime committed against. This can be seen in the columns that have a single color dominating. The areas that appear more brown, like Wilshire, have a mix of races that crimes are being committed against. In each of the areas, time seems to have no affect on what the victim descent is for each crime.

Another feature that can be seen is when the crimes are reported. Each hour has a large band at the start of the hour, meaning many crimes are reported to have occurred at the hour. There is also a lighter band at the 30 minute mark and, in addition, there's also a dark band visible at 12:00 pm, perhaps as a generic time of a reported crime occurring.


Time of Crimes in each Area with Victim Race as Color
Time of Crime
2000
1500
1000
500
0
77th Street
Central
Devonshire
Foothill
Harbor
Hollenbeck
Hollywood
Mission
N Hollywood
Newton
Northeast
Olympic
Pacific
Rampart
Southeast
Southwest
Topanga
Van Nuys
West LA
West Valley
Wilshire
Area Name
Victim Descent
American Indian
Asian
Black
Latino
Pacific Islander
Unknown
White

4.3 Proportion of area of reported crime by ethnicity

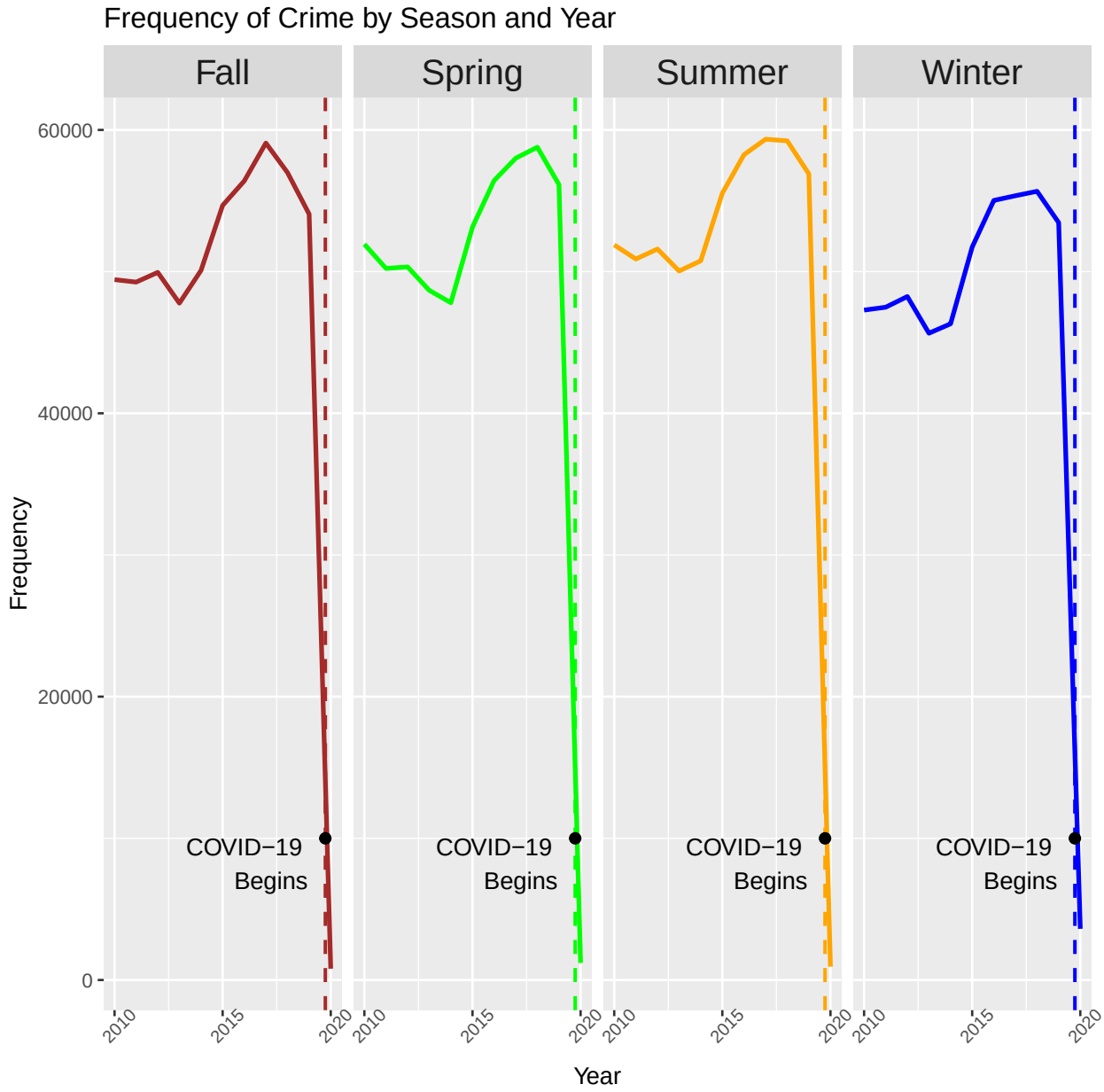
Below is a bar chart based on the proportion of where crimes against victim descents occurred. The proportion is based on the number of crimes committed in an area to a particular victim descent, and the total number of crimes against the same descent. The larger the box in the column the greater the proportion of crimes committed against that race was in that L.A. area. Different colors represent different L.A. areas.



Many of the victim descents have boxes that take up a larger proportion. This means that given a specific descent, more crimes against that group occur in these areas. For example, American Indians experience more crime in the Southwest area. Asians have their highest area as Olympic. African-Americans have 3 areas that make up a large portion: 77th Street, Southwest, and Southeast. The rest of the descents have no area that holds a higher proportion of crimes, despite some having more overall crimes committed against them.

5 Breakdown by time

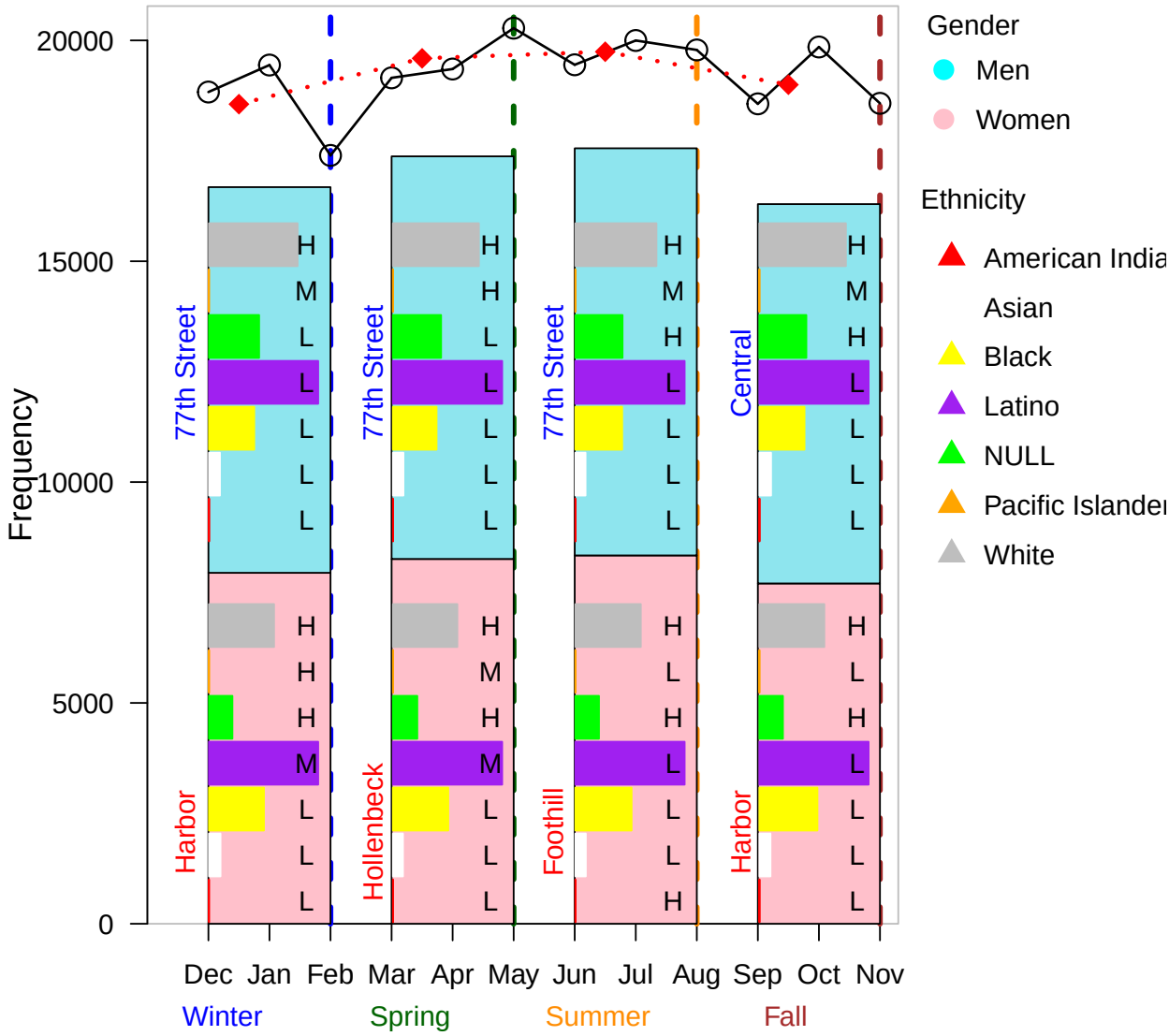
5.1 Logitudinal breakdown by year and season



As seen by this graph, there is a noticeable decline in frequency between "Summer" and "Winter" across all year, barring 2020+ due to a world-wide anomaly - The COVID-19 Pandemic. Compared to "Fall" and "Spring", "Summer" still seems to hold a greater margin in terms of reported crime frequency, but not to the same extent as it does when compared to "Winter". But why? Isn't "Summer" hotter than all other seasons?

Yes, it is, but this may be attributed to the fact that LA County is in South California, a notably hot region, as compared to a midwestern or nothern area (e.g., Salt Lake City). Given this, it may cause a reduced effect if the area is essentially hot year-long, but data showing daily temperatures during this time could help explain this.

6 Killer Plot



The above killer plot showcases multiple variables that we think are crucial to understanding the relationship and overall trends of crime within LA county. For starters, the circular points represent the frequency of crime per month in a set year. The red triangles represent the average crime frequency per the season, in order to show the overall trends over the years. Using this information, all years showcase the lowest crime frequency during the winter season.

In addition, the bar graphs showcase the stacked proportion of crime-victims' gender. Moreover, the horizontal bars highlight the impacted proportion of each gender's ethnicity. What this mean can be clearly show by viewing the gender difference for African Americans (Black). Of all women, black women were victims of a crime at a significantly higher proportion than black men from all men. This relationship is consistent for every year, but for simplicity, this report will only include the year of 2018.

Lastly, we decided to add the communities with the lowest and highest frequency (where lowest frequency communities are at the bottom and vice versa at the top). We also added the income level of specific ethnicites, but none of the of the visual information seems to be consistent.

7 Conclusion

Overall, we visualized specific variables in hopes to emphasize specific relationships between them and the frequency of crime. Although specific portions of our analysis were cut short due to the unavailability of specific data (such as ethnic distribution without each LA community, the number of active police officers per community, and median household income for all years), we were able to find information that can help answer these concerns of crime.

The availability of resources in these communities is incredibly significant with low-income communities having 5x less the number of recreational services. By restricting leisure time, lower income individuals pent up stress and may retaliate. Additionally, the relationship between average median community income and violent crime frequency is quite descriptive. Following the previous logic for resource availability, a low income in a high-cost-of-living area severely limits any kind stress relief, and can also lead to retaliation.

Of course, our visual analysis needs further support with a more detailed sample. Additionally, while these issues can be solved by increasing the number of resources and financial assistance to these low-income areas, California has one of the best socio-economic safety nets in the nation. Future work should analyze the impact of the suggested solutions on crime frequency.

Lastly, future work should focus on the impact of correctional services on criminal offenders. While our dataset maintains the confidentiality of the victims and perpetrators, once the offenders are caught, it's unknown to us if they offend again in future or not. Because of the recent focus on jailing reform in the U.S., an analysis focusing on the repeat offenders or offenders who reintegrate into upstanding citizens should be conducted to determine specific factors that contribute to people committing crimes.