An Analysis into the History of Police Brutality in California

STATS 3351.001
Group 1 - Educated Nighthawks
Jeremiah Joseph, Ayushi Patel, Justin Pham
November 12, 2020

Introduction

The purpose of our study is to use R statistics to find out why police brutality exists and what significant factors contribute to its continuous existence. Through the recent events of George Floyd, Breonna Taylor, and much more within the last decade, we have been led to believe by many news outlets and social media that police brutality is a systematic method used by police officers to employ racism and abuse their power over others of minority race. To test these claims, we analyzed several datasets from 2016 to 2019 taken from the Department of Justice of California. We chose to focus our dataset on California because it has one of the highest number of cases of police brutality in the United States. Our combined dataset contains 9524 observations with 33 class attributes, which include the civilians' gender (*Gender*), race (*Race_Ethnic_Group*), age (*Age*), whether the altercation resulted in death of the civilian or resulted in injury (*Injury_Level*), whether the civilian was perceived to be armed or not (*CIVILIAN_Perceived_Armed*), whether the civilian resisted (*CIVILIAN_Resisted*), and many other attributes.

Using this dataset, we attempted to find the factors that contribute to the occurrence of police brutality. Things such as the race, gender, and age may directly contribute to the occurrences of police brutality, but are there any underlying factors that actually increase police brutality in some areas as opposed to others? Thus, we began to search for other more indirect factors such as the relative income of each county in California and whether or not it holds a correlation with the number of police brutality incidents. By tailoring a variety of graphs to display the relationships of our dataset, we were able to find conclusions regarding the correlations between our class attributes and the occurrence of police brutality as well as being able to determine which class attributes had the highest correlation.

Data Cleaning

The datasets used in this analysis required a great deal of cleaning. In the first step of our data cleaning process, we wanted to combine the separate 2016, 2017, 2018, and 2019 Use of

Force data together. To do this we noticed that there the columns *HI_Islander_Race* and *CIVILIAN_Crime_Qualifier* were not on all the sets. Because this information was not really pertinent to the goals of our analysis, we decided to remove these columns by setting them to null. We then noticed columns indicating the civilian's mental status, and whether the civilian assaulted the officer was labeled under different names, so we changed them to all having a similar name. Another thing we did was add a year column to each of the individual data sets, which allowed us to track the years in the data after combining them. This allowed us to row bind these data sets into a total set.

Initially, most of the variables were given as characters, so we had to turn them into factors. We noticed that for 2017- 2019 all values were lower case, while in 2016 all values were in all capital letters. Due to this, we decided to make the variables lowercase in the 2016 section. We then made race as a factor and gender as a factor. Another variable that we were very much interested in was injury level. In 2016 we noticed that there was slightly different syntax used to describe things. For example, in 2016 a condition was described as serious bodily injury whereas in 2017-2019 this was listed as serious_injury. We then got these terms to be the same and made that as a factor as well.

Other variables we were interested in were in what was called the incident data set. This was from the same source as the Use of Force and had the same cases (marked by a unique id) with other variables such as county, whether an arrest was made, time, etc. This was also separated by year like the Use of Force data set. We first took out columns such as *PRIMARY_AGENCY_INDICATOR* ,and *NUM_INVOLVED_AGENCIES* which were not on all the data sets. Also, in 2016, the column names were all written in lowercase, whereas they were in upper case in all other years. To deal with this we turned all 2016 columns to uppercase. When looking at the counties, we noticed that 2016 put the word county at the end of all this data. To deal with this, we took out that portion of the string. We then were able to row bind these data sets together and have a total incident data set. We then decided to right join the incident data set to the total data set by the id. This allowed us to have a big data set.

We also used other data sets which had the income of each county and the population, in order to help us create maps. We also had to manipulate variables such as age from being strings to being numeric values for different contexts of the plots we were making. We also ended up taking out values that would be hard to graph, to make the data easier to work with. This is all outlined in the code in the appendix. Ultimately, most of the data cleaning process involved ensuring different years used the same terminology and combining multiple data sets from different years.

Below is a summary of our cleaned total data:

```
> str(total)
'data.frame':	9524 obs. of  20 variables:
 $ INCIDENT_ID             : chr  "URSUS-01-0000-2019-402" "URSUS-01-0000-2019-402" "URSUS-01-0000-2019-402" "URSUS-01-0000-2019-481" ...
 $ Civilian_Officer        : chr  "Civilian" "Officer" "Officer" "Civilian" ...
 $ ORI                     : chr  "CA0010000" "CA0010000" "CA0010000" "CA0010000" ...
 $ Gender                  : Factor w/ 2 levels "male","female": 1 1 1 1 1 1 1 1 2 1 ...
 $ Race_Ethnic_Group       : Factor w/ 5 levels "asian","black",..: 2 4 4 2 4 4 4 3 4 2 ...
 $ Age                     : chr  "18_20" "26_30" "41_45" "21_25" ...
 $ Injured                 : logi  FALSE TRUE FALSE TRUE TRUE FALSE ...
 $ Injury_Level            : Factor w/ 4 levels "Injury","Serious Injury",..: 4 2 4 1 2 4 4 4 4 2 ...
 $ Injury_Type             : chr  "" "concussion" "" "abrasion_laceration" ...
 $ CIVILIAN_Perceived_Armed: logi  FALSE NA NA TRUE NA NA ...
 $ CIVILIAN_Resisted       : logi  TRUE NA NA TRUE NA NA ...
 $ CIVILIAN_Confirmed_Armed: logi  FALSE NA NA FALSE NA NA ...
 $ year                    : num  2019 2019 2019 2019 2019 ...
 $ INCIDENT_DATE_STR       : chr  "1/6/2019" "1/6/2019" "1/6/2019" "4/8/2019" ...
 $ INCIDENT_TIME_STR       : chr  "16:20" "16:20" "16:20" "10:30" ...
 $ CITY                    : chr  "DUBLIN" "DUBLIN" "DUBLIN" "DUBLIN" ...
 $ COUNTY                  : chr  "ALAMEDA" "ALAMEDA" "ALAMEDA" "ALAMEDA" ...
 $ ZIP_CODE                : int  94568 94568 94568 94568 94568 94568 94568 94568 94568 94578 ...
 $ ARREST_MADE             : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
 $ CRIME_REPORT_FILED      : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
```

Data Set Summary

Questions and Findings

In this portion of our analysis, we will break down possible factors contributing to outcomes with police force by asking questions and revealing what the data shows. In the conclusion for each section we will answer the overall question pertaining to the factor we are discussing.

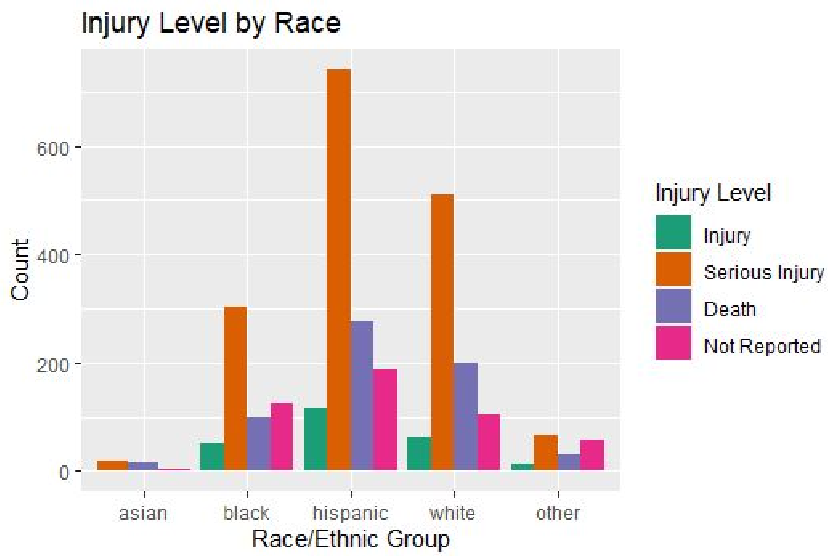
Section 1: Race

Overall Question: Does race affect the way that police officers interact with various individuals?

Sub-Questions:

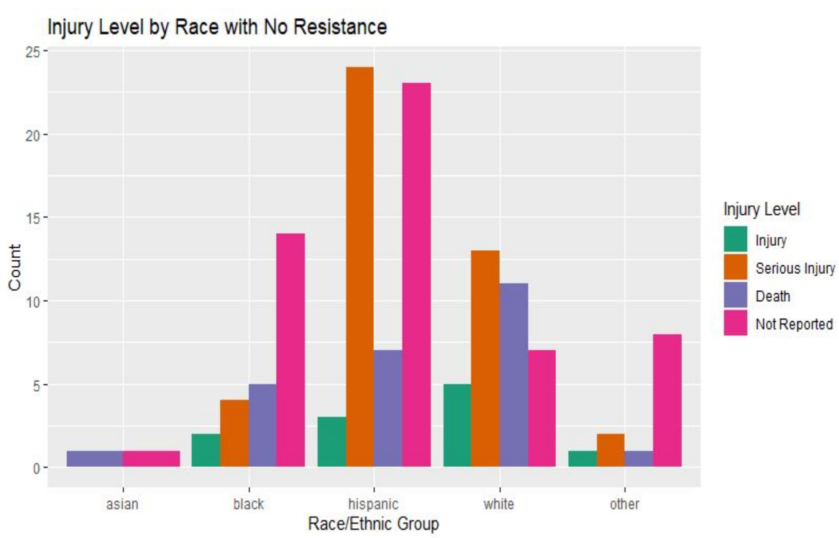
1. **How does race affect the overall injury outcomes of civilians when police officers use force?**

   When looking at the overall picture of how the races are being distributed in the Injury Level by Race plot, we see that hispanic and white individuals are the highest frequency of cases. This is not very surprising as hispanics and white individuals make up the majority of California's population. Blacks have a fairly high frequency whereas Asians represent a very small percentage. Again, these results are not very surprising as the data represents the general trend of the Californian population. Between Black, Hispanic, and White groups (the groups that make up the majority of the cases), the distribution of various severity of injury levels seem to all follow a very similar shape. Within all these cases the smallest injury frequency tends to be the lowest level of injury, followed by non-reporting. The two biggest injury levels tend to be death and serious injury for all these major races. One exception to this is Black individuals, where there are slightly more non-reported cases than deaths. Ultimately, however, the shape of the graphs tend to basically be around the same. Due to this, with the data we have, we would conclude that the overall injury outcomes are independent of race. In general, the overall frequency goes in the general order of the respective population rates, and the injury levels follow mostly the same basic shape. This, however, is based on data given by police departments, and would thus not account for non-reporting and false reporting.



2. **Does the injury outcomes change between races when there is no resistance?**

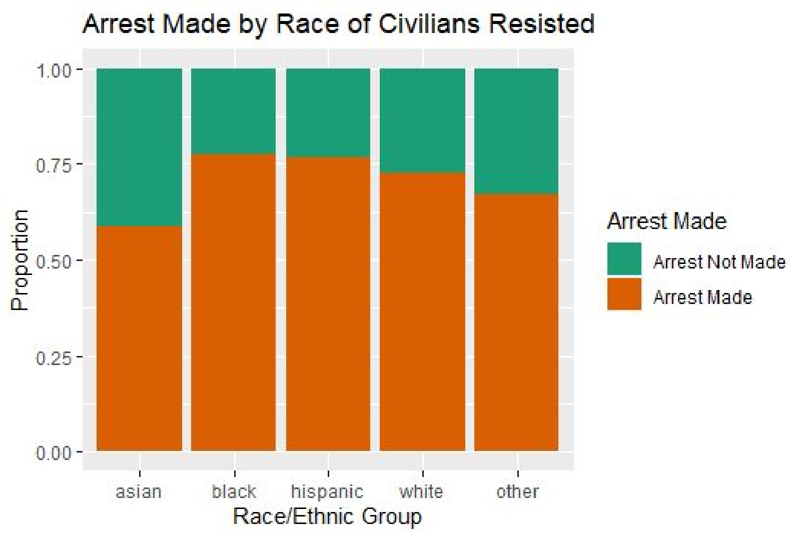
When looking at the data in the Injury Level by Race with No Resistance plot, we see that still the most minor form of injury has the lowest frequency across all races. When looking at the cases which end in death, however, African Americans are the only group where the number of cases of death is higher than the other less severe injury levels. What is also glaring in the data, is the large portion of injury levels that go unreported within the Black, Hispanic, and other groups. In fact, Caucasians are the only race that seems to have not as high of a percentage of individuals whose injury levels go unreported. Both of these may suggest that race may play a factor in how officers react to civilians who don't resist



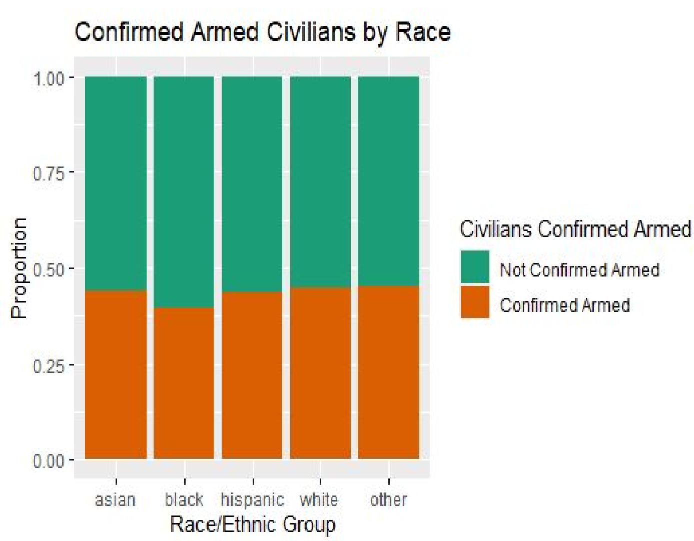
arrest and the detail in which they file reporting on these cases after the fact. When looking at the count, however, these cases are a fairly small subsection of the overall data and thus should be tested with a bigger set of data to confirm any of these observations.

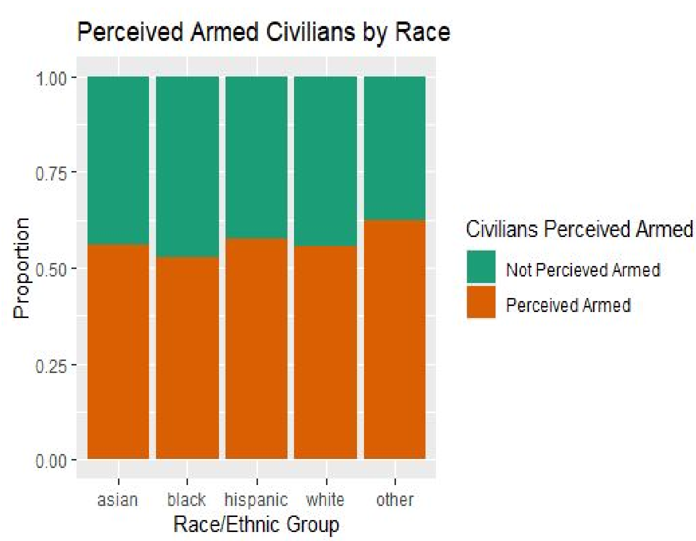
3. **How does race affect the outcomes of individuals who resist arrest?**

When considering the graph of Arrest Made by Race of Civilians Resisted, we can see that among those who resist a police's initial attempt at an arrest, Black and Hispanic groups have the highest arrest rates. White individuals have a lower arrest rate, followed by other races, and finally having the lowest arrest rate is asians. This produces a very interesting result and really raises the question as to why there is such discrepancy between the different races. According to the California Penal Code Section 148(a), resisting arrest is a misdemeanor that is punishable to up to a year in jail. With this in mind, the percentages should be equal, as there should be no inequity by race of how strictly this law is enforced. Though it is unfeasible to try to understand the specific nature as to the description of the crime that warranted the initial interaction, this data along with the graph of injury outcomes based on race when there is no resistance, points to racial prejudice as a likely contributing factor affecting these outcomes.



4. **Do officers generally view any race as more threatening than others?**





Initially when looking at this question, this may be extremely hard to quantify. A metric we decided to look at to possibly gain more insight into this is the difference, split by race, that an officer perceives that a civilian has a gun versus how many times the civilian was confirmed to be armed. Our logic was, that if there was a noticeable variance between different races, this could point to a certain race possibly being looked at as more threatening. When looking at the graphs, we can see that Black, Hispanic, and other groups, had the highest drop. Whites and Asians simultaneously had the fewest drop. This may point to police officers being more threatened by Black and Hispanic groups, and being less threatened by the White and Asian groups. This data we are looking at, however, only considers the difference between races, so with just these graphs it is impossible to be sure why this inequity exists. To truly confirm this idea, it would be ideal to test this with a larger set to see if this point holds true, and it would be important to investigate other possible variables that could explain this difference.

Conclusion: When looking at the overall picture of the data, it seems likely that racial discrimination affects the outcomes that individuals have with the police. Though there was not much evidence of racial discrimination in the overall injury outcomes, we saw that when there was no resistance Black and Hispanic individuals in particular had very high rates of unreported injury level. Also, we can see that Black and Hispanic individuals had a higher percentage of arrest made when initially resisting an officer's arrest compared to other races. Finally, we can see that Black and Hispanic groups also had some of the highest drops in people who were perceived armed versus those who were confirmed armed. It is important to keep in mind, however, that cases with no resistance were extremely low, so that reduces how conclusive this data is. Also we are unable to know all the circumstances that led to a civilian resisting arrest and why the officer suspected the suspect had a gun. For future work, we could test this data with a larger dataset and possibly look at whether other variables had a greater correlation with the arrest rate and the perceived armed rate to confirm the results of our investigations.

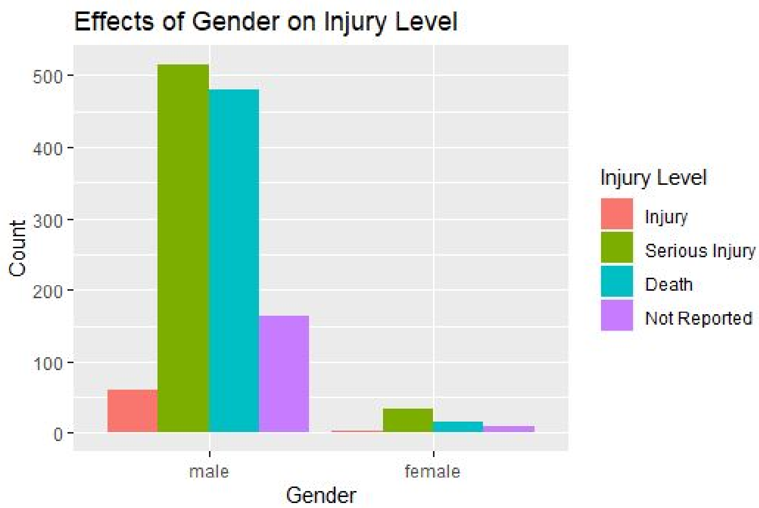
Section 2: Gender

Overall Question: How does gender play a role in predicting the likelihood of police officers using force in their encounters?

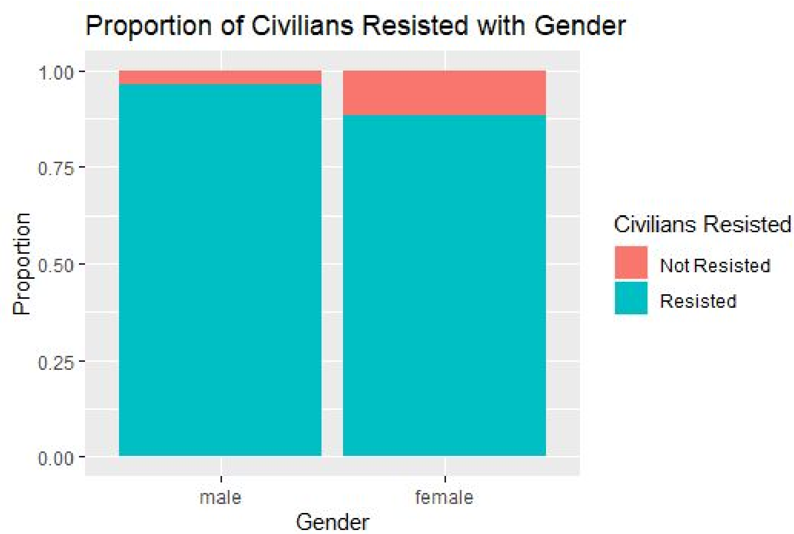
Sub-Questions:

1. **Is there a certain gender that is more likely to be involved in Use of Force incidents and how did these incidents get resolved?**

   From this graph we can tell that males do have a disproportionately higher number of incidents than females, most likely due to crime rates being predominantly male. However, we can also see that most incidents, whether it is male or female, end in serious injury. The second highest is death for both genders, but the number of deaths for males is nearly identical to the number of serious injuries, while the number of deaths for females is half of their number of serious injuries. This may lead to further speculation that police officers are more likely to use lethal force when a male assailant is involved than when a female assailant is involved. In the end, this graph was not enough to reach conclusive evidence that this may be the case, but it allowed us to explore into other related questions.



2. **How does the proportion of those who resisted arrest differ among males and females and what kind of conclusions can we draw from taking the graph above into account as well?**



Surprisingly enough, the proportion of people who resisted arrest are approximately the same between males and females with females only being slightly less than males. What we could draw from this graph is that males and females are approximately equally likely to resist arrest, which would justifiably prompt Use of Force by police officers. However, as seen from the previous graph, a larger proportion of incidents with males who resisted ended in serious injury or death than those with females who resisted, thus suggesting some sort of gender bias in the Use of Force by police officers. This, however, is not proven very conclusive from our data, so it is difficult to prove that it is a factor in whether or not the police officer will use force.

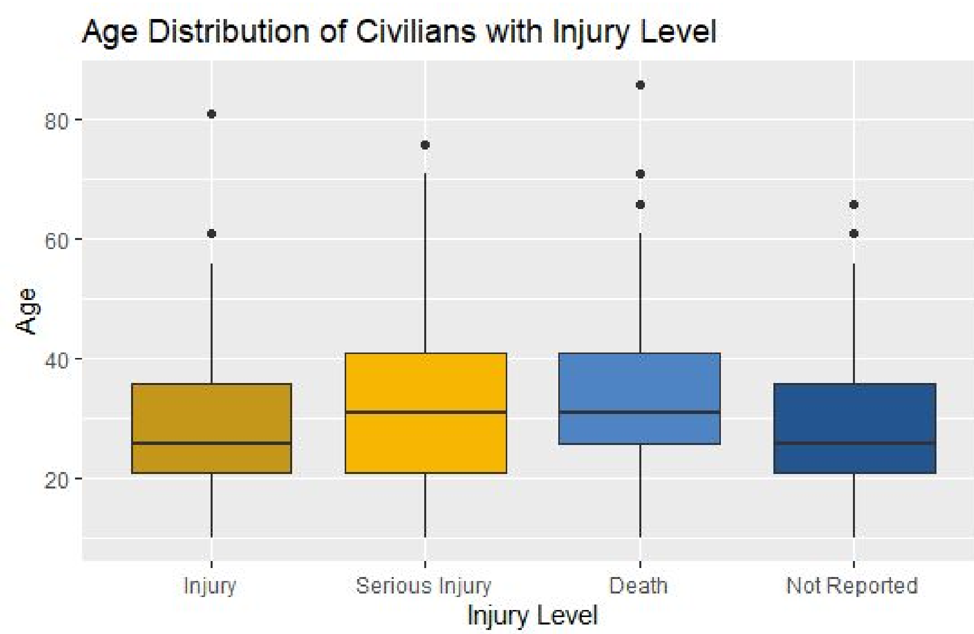
Conclusion: From our graphs, we cannot say for sure that gender plays a significant role in the likelihood of police brutality occurring due to the fact that males, on average, commit more crimes more often than females. However, we cannot dismiss it as an option because we can see that the Use of Force incidents that result in death represent a greater proportion of the total number of incidents involving males than those involving females, despite both genders being almost equally likely to resist, as seen from the second graph. This may lead to the assertion that police officers are more likely to use lethal force on males than on females, but we lack the proper evidence to prove that to be true.

Section 3: Age

<u>Overall Question</u>: What factors are affected by the age of civilians when considering law enforcement's use of force?

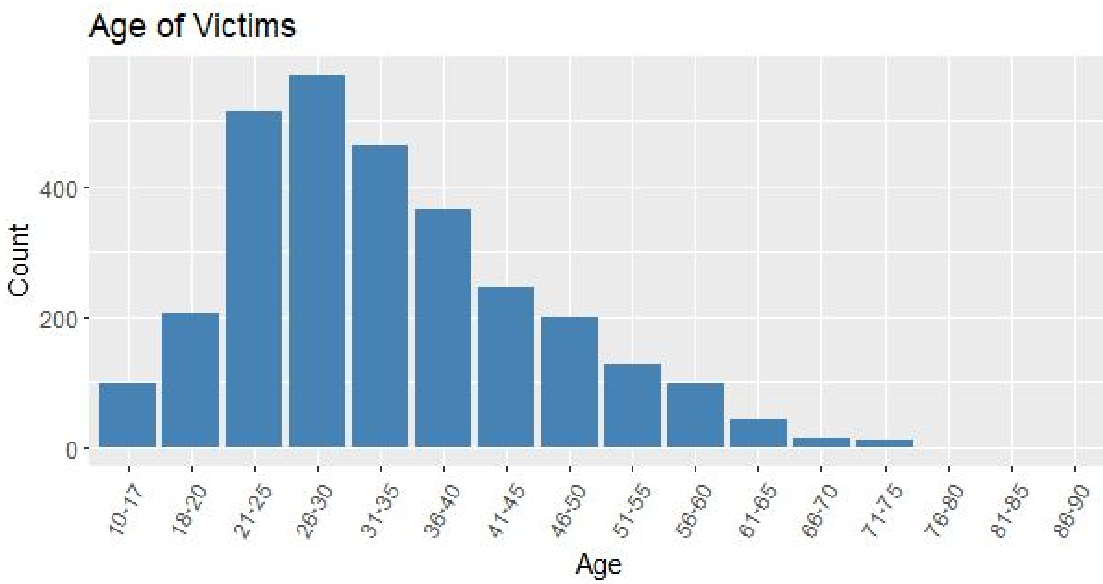
<u>Sub-Questions:</u>

1. **How does age affect the injury level of civilians after encountering the police?**



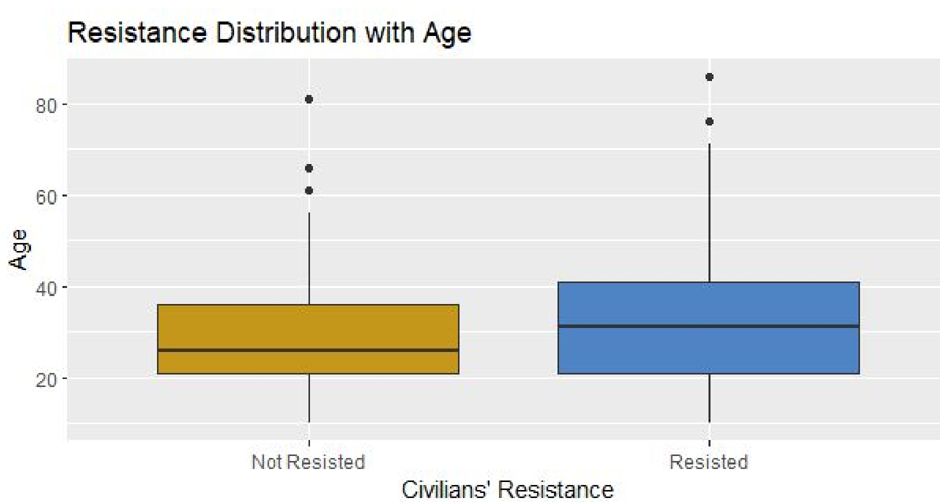
In order to observe how age affects the level of injury of civilians after encountering the police, we decided to create boxplots to visualize the age distribution for each level of injury. First, we can observe that the median age of serious injury and death is higher than the median age of civilians with the lowest level of injury, which is labeled as "Injury" in this graph. This observation supports that as age increases, the risk of injury also increases. We can also see that the interquartile range of serious injury is greater than any other injury level. It spans from the lower quartile of the lowest level of injury to the upper quartile of death. Also, the middle 50% of civilians with serious injury is distributed symmetrically, therefore, we can assume that age does not impact civilians with serious injury as greatly as another injury level.

2. **What is the most common age range for civilians who have been confronted with law enforcement?**



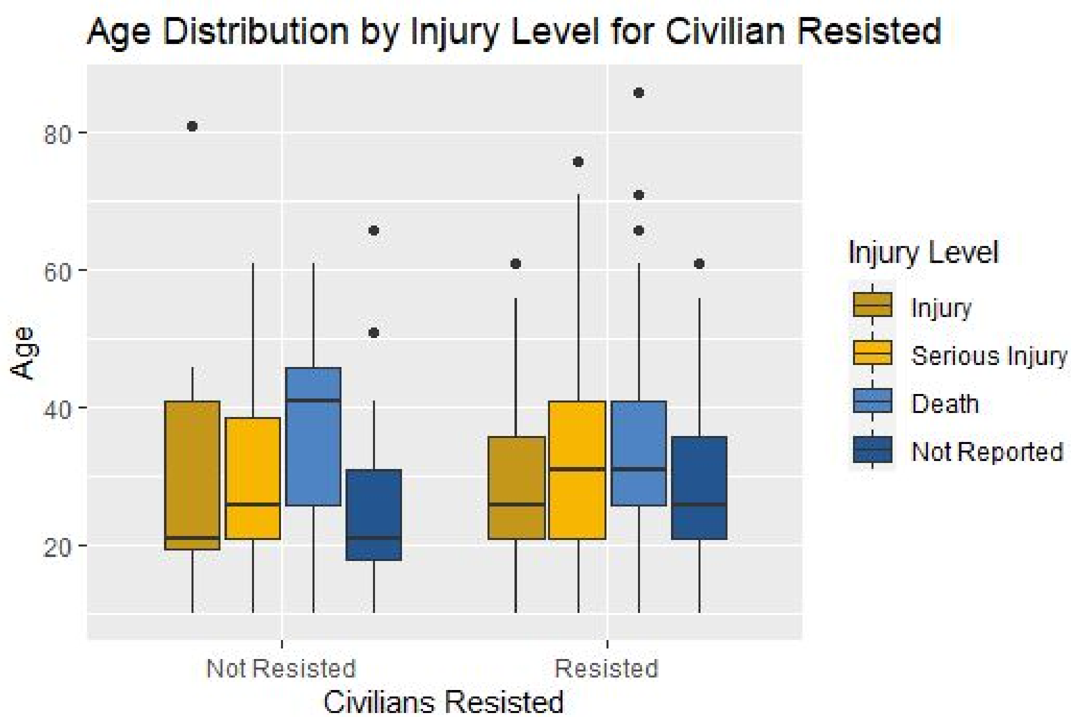
This figure depicts the age of civilians who have confronted law enforcement. The Use of Force data sets from the Department of Justice of California provided age as a range for each individual, mostly of a binwidth of 4 years, however, minors were grouped together and individuals between ages 18 and 20 were combined separately, therefore, civilians under the age of 20 were distributed differently than the other ranges of age. Nevertheless, from the barplot, we can see that the distribution of age is skewed to the right. Our data shows that the greatest number of incidents occurs with people between the ages of 26 and 30 years old. Because our graph is skewed right, we can determine that most people who have encountered the police are adults in their 20's or 30's.

3. **Is age an influence on whether a civilian would be more or less likely to resist police**



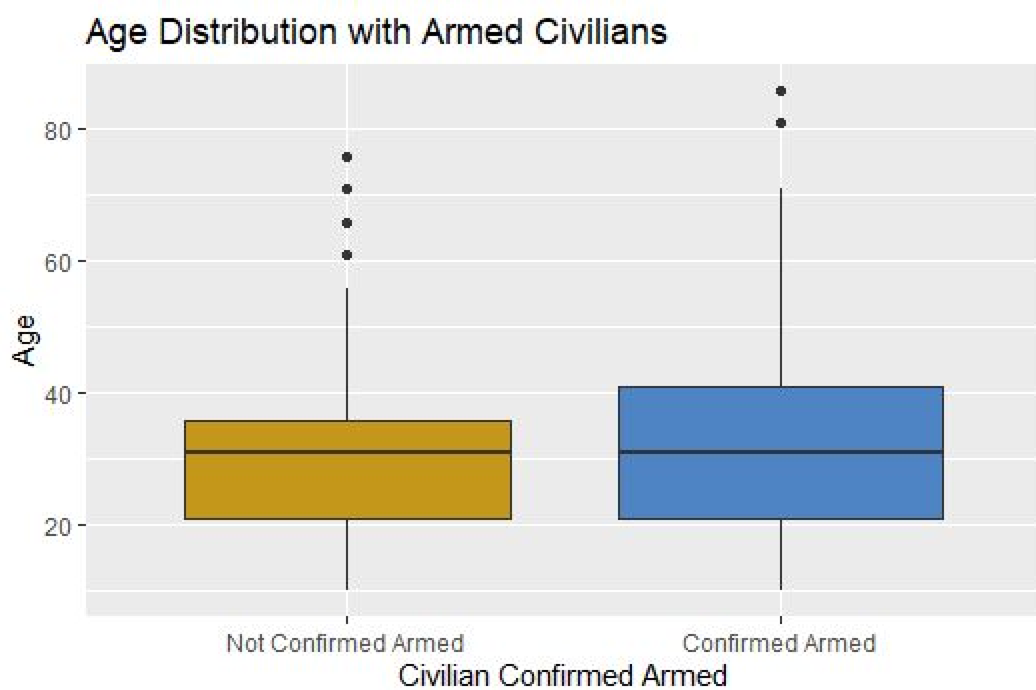
To observe if age has an affect on whether a civilian would be more or less likely to resist arrest, we graphed the distribution of age to if the individual resisted law enforcement or if the individual did not resist. We can notice that the boxplot for those who resisted has a slightly higher median, near 32 years old, while civilians who did not resist have a median age around 25 years old. Another observation is that the boxplot for civilians who did not resist is more skewed to the right than civilians who did resist, which reveals that younger adults are less likely to resist law enforcement. Overall, our graph suggests that older individuals are more likely to resist arrest from police, which is surprising because we believed that younger people are more willing to confront law enforcement. Also, noting the figure in question 2, we can see the greatest number of individuals are in their 20's, which reveals that even though the population of civilians who encounter police is largely young adults, most often these civilians do not resist arrest from law enforcement.

4. **How does age affect the degree of injury for civilians who did not resist the police?**



The figure above shows boxplots of the distribution of age for each injury level grouped together by whether the civilian resisted or not. Notice that in the median ages for each type of injury level for civilians who did not resist, firstly, we can see that the median age for individuals with the lowest level of injury, labeled "Injury" on the graph, is slightly higher than 20 years old. This observation conveys that for civilians who did not resist arrest, which in return led to limited injury by law enforcement, were of a younger age. This sustains the conclusion that age does affect the level of injury an individual may suffer. We can also see that the median age for death of civilians who did not resist arrest is a bit above 40 years old. The older an individual is, the probability of severe injury increases, regardless of police's use of force. To summarize, the level of injury and age have a positive correlation, supporting that age influences injury. Finally, when comparing civilians who have resisted with those who did not resist, we can note that the median ages for resisted levels out to 25 or 30 years old, while those who did not resist have medians skewed to either side. With this information, we can say that individuals who resisted arrest were somewhat distributed similarly across all levels of injury, while those who did not resist were affected more from their age when considering injury. However, we were unable to find any patterns with this graph that answers the question of whether age affects the degree of injury for civilians who did resist arrest.

5. **What age are most civilians who are confirmed armed versus those who were confirmed not to be armed?**



We decided to create boxplots for ages of individuals who were confirmed armed and confirmed not to be armed when confronted by law enforcement. From the figure, we can see that both median ages are the same, which tells us that age does not affect whether a civilian is likely to be actually armed or not. The interquartile range for individuals who were not confirmed armed is less than those who were confirmed armed. Another observation is that civilians confirmed not to be armed are more skewed to the right than civilians who were confirmed to be armed. Unfortunately, this graph does not provide enough support to lead to a more specific conclusion since the boxplots are not very telling. In the future, possibly comparing the ages of civilians who are confirmed armed and unarmed with the ages of civilians who are licensed to carry a gun in California would provide us more information to answer the question of whether age can be a factor to conclude if police are approaching a civilian who is armed or not armed.

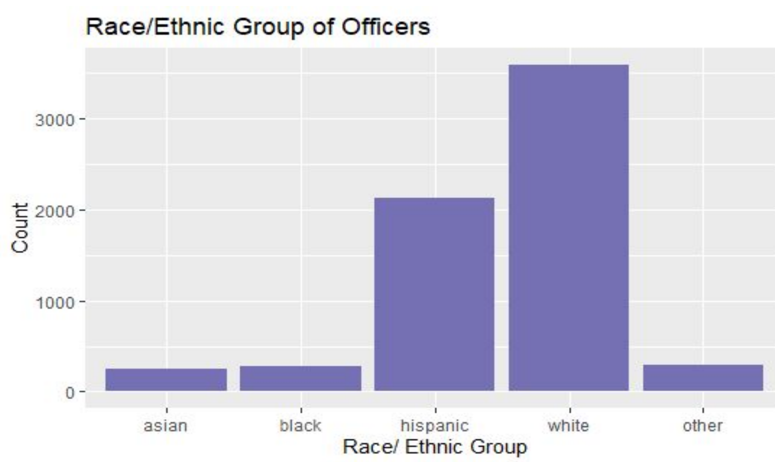
Conclusion: From analyzing our graphs, we can conclude that age affects the level of injury of a civilian after running into law enforcement. We continuously saw that older individuals suffered from a higher degree of injury, which is understandable as aging negatively impacts health, therefore risk of injury. Also, our graphs depicted that younger civilians most often had the lowest level of injury, which again supports that age influences injury. Another question we investigated was if there was a pattern with age and whether or not civilian's resisted arrest. From our graphs we can assume that age does correlate to if individuals resisted or not, we observed that younger people were less likely to resist. This is interesting, because despite this they make up the highest level of overall cases of police violence. Finally, we decided to explore civilians who were armed or unarmed and if they had a telling age distribution. Unfortunately, this factor did not provide us with a distinct conclusion as our graph showed that civilians that are armed and unarmed have similar distributions for age.

Section 4: Police Demographics

Overall Question: Is there a trend in the race of the police officers and the number of Use of Force incidents they are involved in?

Sub-Questions:

1. **What race/ethnic group is most involved in Use of Force incidents? What is the race/ethnic group that is least involved?**



Not only is the race of the victim important, but the race of the officers involved is also an extremely important factor to look at when analyzing these sensitive cases. The graph to the right shows the number of police officers of each race that were involved in incidents that were resulted in the police officer's Use of Force. From our dataset, we can see that a majority of the incidents involve white officers, with hispanic officers coming second highest. The races of officers that were least involved in Use of Force incidents in California were asian and black. While these numbers are not proportionate to the total demographic of police officers in California, we can infer from this data that white officers are more likely to use force when encountering suspects in California.

2. **How does the race/ethnic group of police officers involved in Use of Force incidents compare to the race/ethnic group of civilians involved in Use of Force incidents?**

   As seen from the Race/Ethnic Group of Officers graph, we can see that white officers are the most involved in Use of Force incidents, followed by hispanic officers. If we compare this graph to the Injury Level by Race graph, hispanic civilians have the highest number of serious injuries and deaths among all the races/ethnic groups. This is an interesting point to look at because we can see that there is a possibility of some sort of racial bias that occurs in these police encounters with white officers and hispanic civilians. This assertion is further expounded by the fact that for many of California's counties, the white population represents the vast majority of the demographic, while the hispanic population is the second greatest with a population that is half that of the white population. Despite this, the number of hispanics that were involved in Use of Force incidents were much greater than the number of white civilians.

Conclusion: According to our graph, there is a clear winner in the involvement in the most number of Use of Force incidents. White officers nearly double the hispanic officers and are vastly greater than any other race in their involvement in Use of Force incidents. Before we jump to conclusions, however, we must consider the total demographics of the entire police force of California. Since white officers are the greatest demographic in the police force, they will naturally have more encounters with potential criminals and therefore more encounters that require Use of Force. Therefore, we cannot solely identify white officers as a catalyst to the frequency of police brutality. On the other hand, we cannot completely dismiss these graphs as they give us an idea of which race is the most involved in Use of Force incidents and which side of the coin do they fall under, civilian or police.

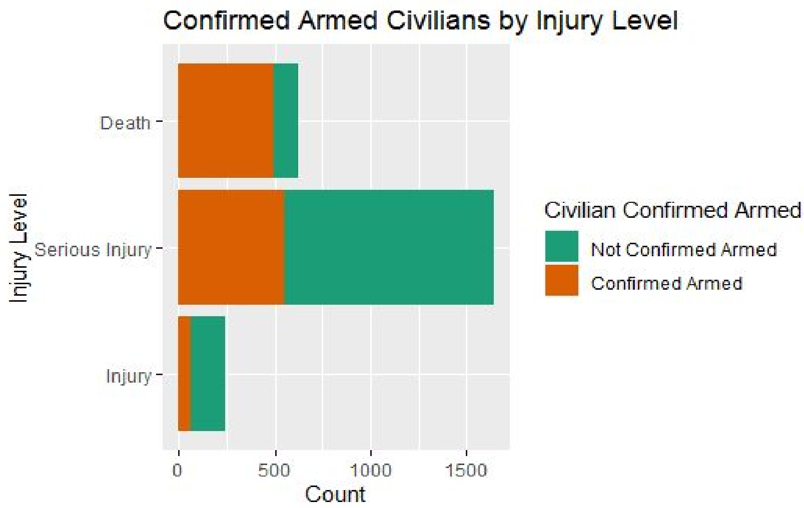
Section 5: Firearms

Overall Question: How does an addition of a firearm change the outcome of incidents of police force?

Sub-Questions:

1. **Are the more extreme injury levels linked with the presence of a firearm?**

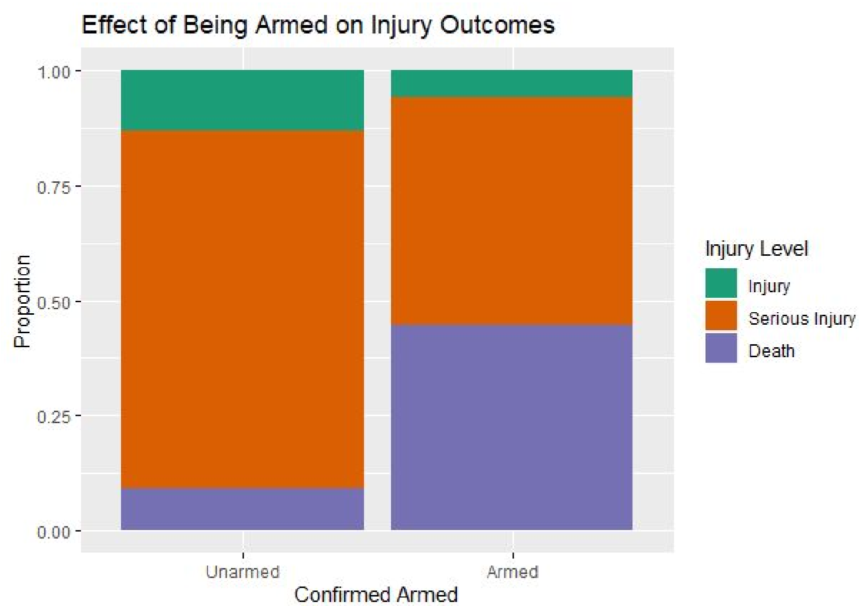
   When looking at the Confirmed Armed Civilians by Injury Level, we can see that the vast majority of cases that end in death occur when the individual was confirmed to be armed. In situations that end in a serious injury, we can see that most are unarmed but there is



still a considerable amount of individuals who are armed. In the cases where the result is simply just an injury, there are almost no cases in which the civilian was armed. Due to this, our data points to the idea that as an injury level becomes more serious, the likelihood of the civilian being armed increases as well. This is based on the data collected by police officers, and again would not account for any possible misreporting that could have occured.

2. **When being armed in a situation with police violence, what will most likely be the result?**

When looking at the Effect of Being Armed on Injury Outcomes graph we can see that there is almost an equal likelihood of a police use of force ending in a serious injury and death. This is a stark difference from the unarmed cases where death is the least likely scenario. This further points to the fact that being armed severely increases the likelihood of a very negative scenario occurring.



3. **Is there anything about the injury levels that remain consistent among both armed and unarmed cases?**

When looking again at the Effect of Being Armed on Injury Outcomes, we can see in both armed and unarmed cases, the vast majority of cases where the police use force ends in at least serious injury or death. In reality, in both cases, the most basic injury rarely ever occured. This shows that though the data points to guns being an extremely devastating factor in determining what level of force an individual receives, most cases ultimately end in at least a serious injury.

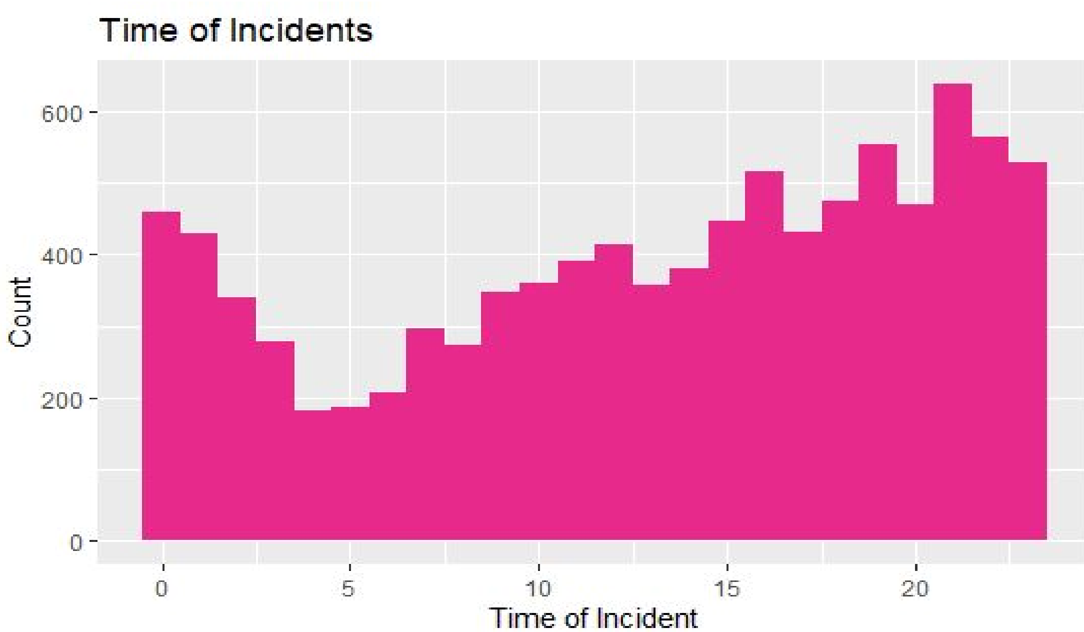
Conclusion: When looking at the data it is very likely that having a presence of a firearm contributes to a worse outcome, when a police uses force. In cases that end in death, the vast majority of individuals had a weapon. Though this is the case, it is important to note that even in cases where the civilian was unarmed, the majority of cases still end in at least a serious injury. When looking at the data, regardless of whether armed or unarmed, it is very rare that a use of force ends in only an injury. It is also important to note that this data is from the police departments, and has the possibility of being biased or misreported. With this in mind, though it is very likely that the presence of a firearm negatively affects the outcomes of police uses of force, Further testing could further prove this point.

Section 6: Time/Date

Overall Question: In what ways does when a police event occurred affect the outcome of the interaction?

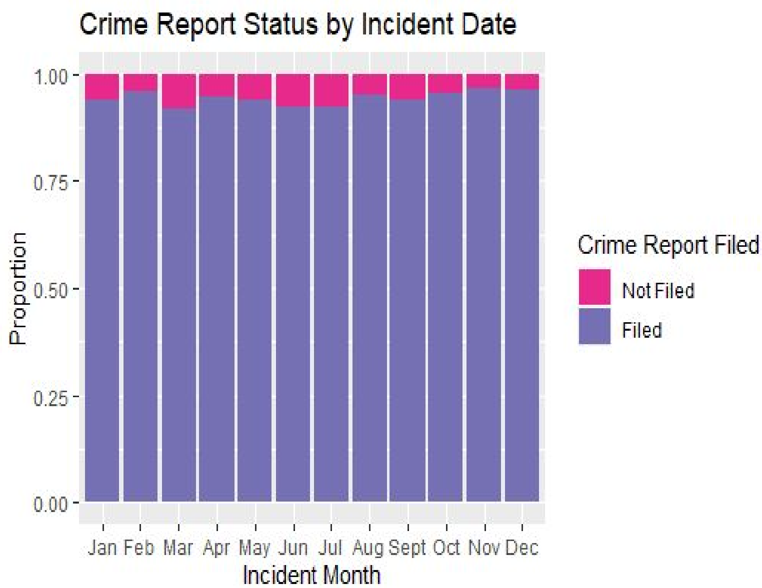
Sub-Questions:

1. **Are there any times where police are more likely to use force?**



When looking at the data of the Time of Incidents graph, we can see that the most acts of police violence occurs towards the end of the day and at the very beginning of the day. This is very interesting, because the most crime in the United States according to the department of justice occurs between 4-7pm, but that is not where we see the highest bumps in the graph. One possible conclusion is that the darkness and lack of visibility might make it more difficult for officers to see, and would thus contribute to them making more aggressive moves. This, however, cannot be tested fully as we are taking a statistic from the overall US and applying to California. Currently there is no statistic that we could find that showed what time most of the crime was committed in California. In future work, we could attempt to find this statistic in order to confirm this observation of the night having a positive correlation with police brutality.

2. **Are there any months of the year where police showed signs of doing a worse job?**



To explore this question we decided to look at the percentage of the crime reports that were filed based on the months of the year. We were wondering if there was any time that we could see a noticeable shift in the percentage of the crime reports that were filed. Our logic was that if there was any time that we saw a significant drop in the percentage crime reports being filed, such as the end of the year, we would expect that to be a time that would be worth looking at incidents that occured during that time. When looking at the Crime Report Status by Incident Date plot, we can see that no such pattern exists. For the most part crime reports are filed fairly consistently. This data comes from the police departments own reporting, however, so it could be argued that the data might be skewed to make the officers look good. In our data, we also noticed that there were holes where there were no incidents in certain months. We believe that this is likely due to failure in recording this data to the dataset. Because of this, we decided that including data on frequency of police incidents in certain months would be misleading, so we did not include any more exploration of the data with respect to month.

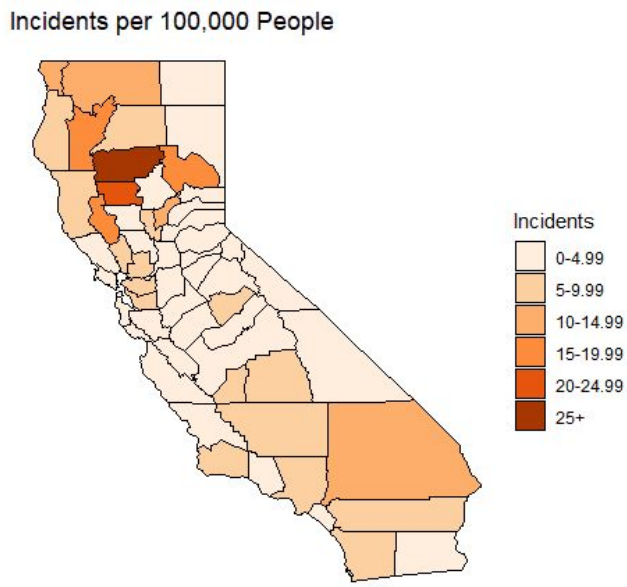
<u>Conclusion</u>: From the data we have, we can say that there is a possibility that higher rates of police brutality occurs at darker parts in a day rather than lighter parts. We cannot be very certain about this conclusion with our data, however, as we are unable to compare this with the overall amounts of crime per time of the day. Our data also leaves us unable to really make any conclusions on how the month affects police brutality as well. Overall, the small observations we saw would make a good foundation to do further research and compare results to larger data sets and with other variables.

Section 7: Incidents per County

<u>Overall Question:</u> How do the number of Use of Force incidents differ between counties? Are there any distinctive features of a specific county that may have an effect on the frequency of these incidents?

<u>Sub-Questions:</u>

1. **Which counties in California had the highest number of incidents and which counties had the lowest?**

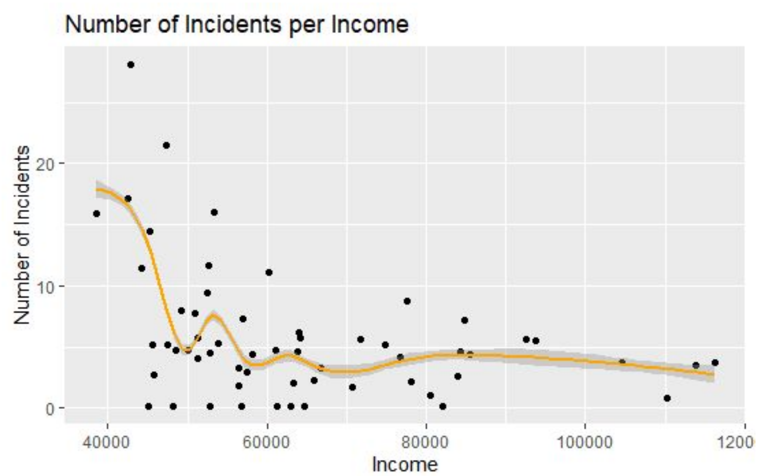


As shown in our Incident map of California, we can see the darker regions correspond with the higher number of Use of Force incidents per 100,000 people. We figured that we should scale our counties with respect to their population as the population across California may range from 1,000 in a small county, such as Alameda, to 1,000,000 in larger counties like Contra Costa. From the map, there is a pattern of higher number of incidents happening in the counties in northern California. These counties include Tehama County being the highest with 25+ incidents, Glenn County with 20-24.99 incidents, and Trinity with 15-19.99 incidents. The counties with the lowest number of incidents appear in central California, indicated by the lightest regions which represent 0-4.99 incidents.

2. **Given the previous map, what kind of conclusion can we make by comparing it to a map of the household income of each county in California? Are these 2 factors related?**

By looking at the two graphs, we can see some correlation between the household income of the county vs the number of incidents that occur in that same county. Those with a higher household income ($80,000 to $100,000+) generally have a lower number of incidents (0 to 4.99), while those with a lower household income ($0 to $59,999) have a higher number of incidents (20 to 25+). This is indicated by how the lighter regions of the household income map correspond to darker regions in the incident map, and vice versa. Although there may be other socioeconomic factors that may play a role, we can use this information to make the assertion that the household income of the environment has some influence over the likelihood of Use of Force occurring.

3. **What is the trend between household income and number of incidents? Is there enough evidence to show that there is a correlation?**



In this graph, we can more easily see a negative correlation between the number of incidents and the household income of each county. With household income data collected from IndexMundi, we were able to plot the household income of each county in California in a given year against the number of incidents that occurred in each county, which we were able to sum up from our original Use of Force dataset. As we can see from our trendline we can see an overall negative correlation, with the highest number of incidents happening in counties with an average household income on the lower side of the spectrum. From this graph, we cannot say that household income is the definitive variable in affecting the likelihood of Use of Force incidents happening, but we can say that it has some effect.

Conclusion: The number of Use of Force incidents greatly differ between each county. As for distinctive features that may cause this difference, we looked into the average household income of each county in a given year. Through our maps, we were able to highlight the correlation between low household income counties and higher rates of Use of Force incidents by police officers. Using this simple correlation, we can surmise that household income may be a probable factor in causing Use of Force incidents to happen. However, we cannot decisively state that household income is the sole factor that causes these incidents. There may be other socioeconomic factors that also play a role. For future work, we could also look at things like crime rate, number of gun owners in a given county, or number of police precincts in one county to find other underlying causes of police brutality.

Conclusion

From our efforts of identifying what factors contribute to the continuous case of police brutality, we have discovered that California has many trends that correlate with the use of force by law enforcement. Our data combined incidents from 2016 to 2019 in California which required a considerable amount of cleaning. However, after organizing our spreadsheets we were able to analyze civilians and officers involved in these cases. Because of the recent events of police brutality along with the Black Lives Matter Movement, we firstly wanted to investigate how race affects use of force by police officers. Our graphs depicted that racial discrimination does affect individuals who did not resist arrest from law enforcement, primarily those who are Black or Hispanic. We also noticed that minority races had the highest counts of being perceived armed by police. Next, we wanted to explore if gender has an influence on the use of force, our data had a greater proportion of male suspects than female, and from our figures we were unfortunately unable to confirm that gender plays a role in the likelihood of police brutality. When comparing age distribution with the level of injury after encountering the police, we were able to infer that age does influence injury level. Our graphs also support that age correlates to civilian resistance, we found that younger adults were less likely to resist. After focusing on civilian demographics, we wanted to shift gears towards common factors among police officers involved in these incidents. We identified that our data had nearly double the amount of white officers than Hispanic officers and held a greater margin between any other race in their involvement with use of force. However, considering the demographics of the total population of California, we have to recognize that white police officers cannot be the sole cause of police brutality but also acknowledge that our graphs still hold truth for racial bias throughout cases involving law enforcement. Finally, adding on data from the U.S. Census provided us the opportunity to look more in depth into incomes of each county in comparison to crime rates. There was a correlation between low household income counties and higher rates of use of force incidents.

While examining many factors that can affect the ongoing cases of police brutality, we have compiled even more questions that need answers. In the future, exploring data from all of the states and from further back than 2016 would allow us to solidify our findings and apply our conclusions to a national level since we would have a greater amount of information. We would also like to explore more into topics that were inconclusive when analyzing our findings, such as getting data that has more proportionality in gender. Additionally, comparing our results from this data, which is provided by California police departments, with findings from data that is sourced from independent records might lead to interesting discrepancies.

Appendix

Project Code

```r
library(ggplot2)
library(tidyverse)
library(dplyr)
library(maps)
library(ggpubr)

#reading data sets
URSUS_Civilian_Officer_2018 <- read.csv("URSUS_Civilian-Officer_2018.csv")
URSUS_Civilian_Officer_2019 <- read.csv("URSUS_Civilian-Officer_2019.csv")
URSUS_Civilian_Officer_2017 <- read.csv("URSUS_Civilian-Officer_2017.csv")
URSUS_Civilian_Officer_2016 <- read.csv("URSUS_Civilian-Officer_2016.csv")

incident_2019 <- read.csv("incident_2019.csv")
incident_2018 <- read.csv("incident_2018.csv")
incident_2017 <- read.csv("incident_2017.csv")
incident_2016 <- read.csv("incident_2016.csv")

#cleaning incident IDs
colnames(incident_2019)[1] <- "INCIDENT_ID"
incident_2017$PRIMARY_AGENCY_INDICATOR <- NULL
incident_2018$NUM_INVOLVED_AGENCIES <- NULL
incident_2019$NUM_INVOLVED_AGENCIES <- NULL
colnames(incident_2016) <- toupper(colnames(incident_2016))
incident_2016$COUNTY <- substr(incident_2016$COUNTY, 1,
nchar(incident_2016$COUNTY) - 6)

#combining data sets
total_incident <- rbind(incident_2019, incident_2018, incident_2017,
incident_2016)

#getting factor variable
URSUS_Civilian_Officer_2019$Gender <-
factor(URSUS_Civilian_Officer_2019$Gender, levels = c("male", "female"))

#Adding table of year
URSUS_Civilian_Officer_2019$year <- 2019
URSUS_Civilian_Officer_2018$year <- 2018
URSUS_Civilian_Officer_2017$year <- 2017
URSUS_Civilian_Officer_2016$year <- 2016


#factor variable for gender
```

```r
URSUS_Civilian_Officer_2018$Gender <-
factor(URSUS_Civilian_Officer_2018$Gender, levels = c("male", "female"))
URSUS_Civilian_Officer_2017$Gender <-
factor(URSUS_Civilian_Officer_2017$Gender, levels = c("male", "female"))
URSUS_Civilian_Officer_2016$Gender <-
factor(URSUS_Civilian_Officer_2016$Gender, levels = c("Male", "Female"))
levels(URSUS_Civilian_Officer_2016$Gender) <- c("male", "female")

#getting rid of variables that are not common
URSUS_Civilian_Officer_2018$HI_Islander_Race <- NULL
URSUS_Civilian_Officer_2019$HI_Islander_Race <- NULL
colnames(URSUS_Civilian_Officer_2019)[21] <- "CIVILIAN_Assaulted_Officer"
colnames(URSUS_Civilian_Officer_2019)[20] <- "CIVILIAN_Mental_Status"
colnames(URSUS_Civilian_Officer_2018)[20] <- "CIVILIAN_Mental_Status"
colnames(URSUS_Civilian_Officer_2019)[1] <- "Incident_ID"
URSUS_Civilian_Officer_2017$CIVILIAN_Crime_Qualifier <- NULL
URSUS_Civilian_Officer_2016$CIVILIAN_Crime_Qualifier <- NULL

#combining data sets
total <- rbind(URSUS_Civilian_Officer_2019, URSUS_Civilian_Officer_2018,
URSUS_Civilian_Officer_2017, URSUS_Civilian_Officer_2016)

#combine both total data sets:
colnames(total)[1] <- "INCIDENT_ID"
total <- right_join(total, total_incident, id = "INCIDENT_ID")

#Taking out irrelevant info
total$OFFICER_Dress <- NULL
total$Asian_Race <- NULL
total$Received_Force_Location <- NULL
total$NUM_INVOLVED_CIVILIANS <- NULL
total$NUM_INVOLVED_OFFICERS <- NULL
total$IN_CUSTODY_REASON <- NULL
total$CONTACT_REASON <- NULL
total$ON_K12_CAMPUS <- NULL
total$CIVILIAN_K12_Type <- NULL
total$Medical_Aid <- NULL
total$MULTIPLE_LOCATIONS <- NULL
total$OFFICER_Officer_Used_Force_Reason <- NULL
total$OFFICER_On_Duty <- NULL
total$OFFICER_Officer_Used_Force <- NULL
total$CIVILIAN_Custody_Status <- NULL
total$CIVILIAN_Resistance_Type <- NULL
total$CIVILIAN_Firearm_Type <- NULL
total$CIVILIAN_Highest_Charge <- NULL
total$Injury_From_Preexisting_Condition <- NULL
total$Order_Of_Force_Specified <- NULL
total$Order_Of_Force_Str <- NULL
total$Received_Force <- NULL
```

```
total$Received_Force_Type <- NULL
total$DISCHARGE_OF_FIREARM_INDIVIDUAL <- NULL
total$DISCHARGE_OF_FIREARM_INCIDENT <- NULL
total$CIVILIAN_Mental_Status <- NULL
total$CIVILIAN_Assaulted_Officer <- NULL
total$CIVILIAN_Perceived_Armed_Weapon <- NULL
total$CIVILIAN_Confirmed_Armed_Weapon <- NULL
total$STATE <- NULL

#Making lower case
total$Race_Ethnic_Group <- tolower(total$Race_Ethnic_Group)

#creating race as factor
total$Race_Ethnic_Group <- addNA(factor(total$Race_Ethnic_Group, levels =
c("asian", "black", "hispanic", "white" )))
total$Race_Ethnic_Group[which(is.na.data.frame(total$Race_Ethnic_Group))] <-
"other"
levels(total$Race_Ethnic_Group) <- c("asian", "black", "hispanic", "white",
"other")

#creating injury level as factor
total$Injury_Level <- tolower((total$Injury_Level))
total$Injury_Level[which(total$Injury_Level == "serious bodily injury")] <-
"serious_injury"
total$Injury_Level <- addNA(factor(total$Injury_Level, levels = c("injury",
"serious_injury", "death")))
levels(total$Injury_Level) <- c("Injury", "Serious Injury", "Death", "Not
Reported")

#-----------------------------------------------------------------------------
#Creating graphs

#RACE
#Comparing race injury levels when resistance is happened vs not
ggplot(data = subset(total, Civilian_Officer == "Civilian")) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill = Injury_Level), position
= "dodge") +
  labs(x = "Race/Ethnic Group", y = "Count", fill = "Injury Level", title =
"Injury Level by Race") +
  scale_fill_brewer(palette = "Dark2")

ggplot(data = subset(total, Civilian_Officer == "Civilian" &
CIVILIAN_Resisted == FALSE)) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill = Injury_Level), position
= "dodge") +
  labs (x = "Race/Ethnic Group", y = "Count", fill = "Injury Level", title =
"Injury Level by Race with No Resistance") +
  scale_fill_brewer(palette = "Dark2")
```

```
#Looking how being armed effects different races
ggplot(data = subset(total, Civilian_Officer == "Civilian" &
CIVILIAN_Perceived_Armed == TRUE)) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill = Injury_Level), position
= "dodge") +
  scale_fill_brewer(palette = "Dark2") +
  labs(title = "Injury Level by Race for Armed Civilians", x = "Race/Ethnic
Group", y = "Count", fill = "Injury Level")

#Looking at how difference between confirmed and non confirmed weapons based
on race
ggplot(data = subset(total, Civilian_Officer == "Civilian" )) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill =
CIVILIAN_Perceived_Armed), position = "fill") +
  labs(x = "Race/Ethnic Group", y = "Proportion", fill = "Civilians Perceived
Armed", title = "Perceived Armed Civilians by Race") +
  scale_fill_brewer(palette = "Dark2", labels = c("Not Perceived Armed",
"Perceived Armed"))

ggplot(data = subset(total, !is.na.data.frame(total$CIVILIAN_Confirmed_Armed)
&Civilian_Officer == "Civilian" )) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill =
CIVILIAN_Confirmed_Armed), position = "fill") +
  labs(x = "Race/Ethnic Group", y = "Proportion", fill = "Civilians Confirmed
Armed", title = "Confirmed Armed Civilians by Race") +
  scale_fill_brewer(palette = "Dark2", labels = c("Not Confirmed Armed",
"Confirmed Armed"))

#arrest made
ggplot(data = subset(total, Civilian_Officer == "Civilian" &
CIVILIAN_Resisted == TRUE)) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group,fill = ARREST_MADE), position
= "fill") +
  labs(x = "Race/Ethnic Group", y = "Proportion", title = "Arrest Made by
Race of Civilians Resisted", fill = "Arrest Made") +
  scale_fill_brewer(palette = "Dark2", labels = c("Arrest Not Made", "Arrest
Made"))



#RACE OF OFFICER
ggplot(data = subset(total, Civilian_Officer == "Officer")) +
  geom_bar(mapping = aes(x = Race_Ethnic_Group), fill = "#7470b3") +
  labs(x = "Race/ Ethnic Group", y = "Count", title = "Race/Ethnic Group of
Officers")



#FIREARM
#Seeing how the introduction of a weapon changes endings of encounters
x <- ggplot(data = subset(total, Civilian_Officer == "Civilian" &
```

```r
Injury_Level != "Not Reported")) +
  geom_bar(mapping = aes(x = Injury_Level, fill = CIVILIAN_Confirmed_Armed),
position = "stack") +
  labs(y = "Count", x = "Injury Level", fill = "Civilian Confirmed Armed",
title = "Confirmed Armed Civilians by Injury Level") +
  scale_fill_brewer(palette = "Dark2", labels = c("Not Confirmed Armed",
"Confirmed Armed"))
x + coord_flip()

#Seeing how the introduction of a weapon changes endings of encounters
ggplot(data = subset(total, Civilian_Officer == "Civilian" & Injury_Level !=
"Not Reported")) +
  geom_bar(mapping = aes(x = CIVILIAN_Confirmed_Armed, fill = Injury_Level),
position = "fill") +
  labs(x = "Confirmed Armed", y = "Proportion", title = "Effect of Being
Armed on Injury Outcomes", fill = "Injury Level") +
  scale_fill_brewer(palette = "Dark2") + scale_x_discrete(labels =
c("Unarmed", "Armed"))


#GENDER
ggplot(data = subset(total, Civilian_Officer == "Civilian" &
!is.na.data.frame(total$Gender) & CIVILIAN_Confirmed_Armed == TRUE)) +
  geom_bar(mapping = aes(x = Gender,fill = Injury_Level), position = "dodge")
+
  labs(y = "Count", fill = "Injury Level", title = "Effects of Gender on
Injury Level")

ggplot(data = subset(total, Civilian_Officer == "Civilian" &
!is.na.data.frame(total$Gender))) +
  geom_bar(mapping = aes(x = Gender,fill = CIVILIAN_Resisted), position =
"fill") +
  labs(y = "Proportion", fill = "Civilians Resisted", title = "Proportion of
Civilians Resisted with Gender") +
  scale_fill_discrete(labels = c("Not Resisted", "Resisted"))


#AGE
#Box plot of Age vs Injury Level
total$Age <- substr(total$Age, 1, 2)
total$Age <- as.numeric(total$Age)

ggplot(total) + geom_boxplot(mapping = aes(x = Injury_Level, y = Age), fill =
c("#C4961A", "#f7b602","#4e84c4", "#235591")) +
  labs(x = "Injury Level", title = "Age Distribution with Injury Level")

ggplot(data = subset(total, Civilian_Officer == "Civilian")) +
  geom_boxplot(mapping = aes(x = Injury_Level, y = Age), fill = c("#C4961A",
"#f7b602","#4e84c4", "#235591")) +
```

```r
  labs(x = "Injury Level", title = "Age Distribution of Civilians with Injury
Level")

#Box plot of Age vs Confirmed Armed
ggplot(data = subset(total, !is.na.data.frame(total$CIVILIAN_Confirmed_Armed)
& Civilian_Officer == "Civilian")) +
  geom_boxplot(mapping = aes(x = CIVILIAN_Confirmed_Armed, y = Age), fill =
c("#C4961A", "#4E84C4")) +
  labs(x = "Civilian Confirmed Armed", title = "Age Distribution with Armed
Civilians") +
  scale_x_discrete(labels = c("Not Confirmed Armed", "Confirmed Armed"))

#Box plot of age and resistance with Injury Level
ggplot(subset(total, !is.na.data.frame(total$CIVILIAN_Resisted))) +
  geom_boxplot(mapping = aes(x = CIVILIAN_Resisted, y = Age, fill =
Injury_Level)) +
  labs(x = "Civilians Resisted", fill = "Injury Level", title = "Age
Distribution by Injury Level for Civilian Resisted") +
  scale_fill_manual(values = c("#C4961A", "#f7b602","#4e84c4", "#235591")) +
  scale_x_discrete(labels = c("Not Resisted", "Resisted"))

#Box plot of age and resistance
ageresist <- ggplot(subset(total, !is.na.data.frame(total$CIVILIAN_Resisted)
& Civilian_Officer == "Civilian")) +
  geom_boxplot(mapping = aes(x = CIVILIAN_Resisted, y = Age), fill =
c("#C4961A", "#4E84C4")) +
  labs(x = "Civilians' Resistance", title = "Resistance Distribution with
Age") +
  scale_x_discrete(labels = c("Not Resisted", "Resisted"))

#Age Bar plot
total$Age <- as.character(total$age)
total$Age[which(total$Age == "0")] <- "0-9"
total$Age[which(total$Age == "10" | total$Age == "17")] <- "10-17"
total$Age[which(total$Age == "18")] <- "18-20"
total$Age[which(total$Age == "21")] <- "21-25"
total$Age[which(total$Age == "26")] <- "26-30"
total$Age[which(total$Age == "31")] <- "31-35"
total$Age[which(total$Age == "36")] <- "36-40"
total$Age[which(total$Age == "41")] <- "41-45"
total$Age[which(total$Age == "46")] <- "46-50"
total$Age[which(total$Age == "51")] <- "51-55"
total$Age[which(total$Age == "56")] <- "56-60"
total$Age[which(total$Age == "61")] <- "61-65"
total$Age[which(total$Age == "66")] <- "66-70"
total$Age[which(total$Age == "71")] <- "71-75"
total$Age[which(total$Age == "76")] <- "76-80"
total$Age[which(total$Age == "81")] <- "81-85"
total$Age[which(total$Age == "86")] <- "86-90"
```

```r
agebar <- ggplot(data = subset(total, Civilian_Officer == "Civilian" &
!is.na.data.frame(total$Age))) +
  geom_bar(mapping = aes(x=as.factor(Age)), fill = "steelblue") +
  labs(x = "Age", y = "Count", title = "Age of Victims") + theme(axis.text.x
= element_text(angle = 60, hjust = 1))

ggarrange(ageresist, agebar, nrow = 2, ncol = 1)


#COUNTY
#map for county
total$COUNTY <-toupper(total$COUNTY)

co_num <- data.frame(table((total$COUNTY[which(total$Civilian_Officer ==
"Civilian")])))

colnames(co_num)[1] <- "subregion"
co_num$subregion <- tolower(co_num$subregion)

pop <- read_excel("pop.xlsx")
colnames(pop)[1] <- "subregion"
colnames(pop)[2] <- "population"

pop$subregion <- substr(pop$subregion, 2, nchar(pop$subregion) - 19)
pop$subregion <- tolower(pop$subregion)

pop$population <- as.numeric(pop$population)
co_num$population <- as.numeric(co_num$population)
co_num <- right_join(co_num, pop, id = "subregion")

co_num$fre <- co_num$Freq / (co_num$population / 100000)
co_num$fre[which(is.na.data.frame(co_num$fre))] <- 0.1

#calculating standard deviation
sd(co_num$fre)

counties <- map_data("county")
ca_county <- counties%>%
  filter(region == "california")
head(ca_county)
ggplot(ca_county) + geom_polygon(mapping = aes(x = long, y = lat)) +
coord_quickmap()
ggplot(ca_county) + geom_polygon(mapping = aes(x = long, y = lat, group =
group), color = "white") + coord_quickmap()
borders2 <- left_join(ca_county, co_num, by = "subregion")

borders2$bin <- cut((borders2$fre) / 5,
                    breaks = 0:6, labels = c("0-4.99", "5-9.99", "10-14.99",
```

```
"15-19.99", "20-24.99", "25+"))
ggplot(borders2) + geom_polygon(mapping = aes(x = long, y = lat, group =
group, fill = bin), color = "black") +
  coord_quickmap() + scale_fill_brewer(palette = "Oranges", direction = 1) +
  labs(title = "Incidents per 100,000 People", fill = "Incidents") +
theme_void()

#map for income
total$COUNTY <-toupper(total$COUNTY)

counties <- map_data("county")
ca_county <- counties%>%
  filter(region == "california")
head(ca_county)
ggplot(ca_county) + geom_polygon(mapping = aes(x = long, y = lat)) +
coord_quickmap()
ggplot(ca_county) + geom_polygon(mapping = aes(x = long, y = lat, group =
group), color = "white") + coord_quickmap()

Counties_by_Income <- read_excel("Counties by Income.xlsx")
colnames(Counties_by_Income)[1] <- "subregion"

Counties_by_Income$subregion <- tolower(Counties_by_Income$subregion)
Counties_by_Income$Income <- as.numeric(Counties_by_Income$Income)
borders <- left_join(ca_county, Counties_by_Income, by = "subregion")
borders$bin <- cut((borders$Income) / 20000,
                   breaks = 1:6, labels = c("0-39,999", "40,000-59,999",
"60,000-79,999", "80,000-99,999", "100,000+"))
ggplot(borders) + geom_polygon(mapping = aes(x = long, y = lat, group =
group, fill = bin), color = "black") +
  coord_quickmap() + scale_fill_brewer(palette = "Oranges", direction = 1) +
  labs(title = "Income of California Counties", fill = "Income") +
theme_void()

#scatter plot for income
ggplot(data = borders2, mapping = aes(x = borders$Income, y = fre)) +
geom_point() +
  geom_smooth(color = "Orange") +
  labs(title = "Number of Incidents per Income", x = "Income", y = "Number of
Incidents")


#TIME
#Making Histogram of time of incidents
count <- 0
for(val in incident_2016$INCIDENT_TIME_STR) {
  count <- count + 1
  if(nchar(val) == 3) {
    incident_2016$INCIDENT_TIME_STR[count] <- paste0("0",
```

```
incident_2016$INCIDENT_TIME_STR[count])
  } else if(nchar(val) == 2) {
    incident_2016$INCIDENT_TIME_STR[count] <- paste0("00",
incident_2016$INCIDENT_TIME_STR[count])
  } else if(nchar(val) == 1) {
    incident_2016$INCIDENT_TIME_STR[count] <- paste0("000",
incident_2016$INCIDENT_TIME_STR[count])
  }
}

count <- 0
for(val in total$INCIDENT_TIME_STR) {
  count <- count + 1
  if(total$year[count] == 2016) {
    if(nchar(val) == 3) {
      total$INCIDENT_TIME_STR[count] <- paste0("0",
total$INCIDENT_TIME_STR[count])
    } else if(nchar(val) == 2) {
      total$INCIDENT_TIME_STR[count] <- paste0("00",
total$INCIDENT_TIME_STR[count])
    } else if(nchar(val) == 1) {
      total$INCIDENT_TIME_STR[count] <- paste0("000",
total$INCIDENT_TIME_STR[count])
    }
  }
}

count <- 0
for(val in total$INCIDENT_TIME_STR) {
  count <- count + 1
  if(total$year[count] == 2017 & nchar(val) == 4) {
    total$INCIDENT_TIME_STR[count] <- paste0("0",
total$INCIDENT_TIME_STR[count])
  }
}

total$INCIDENT_TIME_STR <- substr(total$INCIDENT_TIME_STR, 1, 2)
total$INCIDENT_TIME_STR <- as.numeric(total$INCIDENT_TIME_STR)

ggplot(total) + geom_histogram(mapping = aes(x = INCIDENT_TIME_STR), bins =
24, fill = c("#e7298b")) +
  labs(x = "Time of Incident", y = "Count", title = "Time of Incidents")

#Bar plot of Incident Time and Crime Report Filed
total$INCIDENT_DATE_STR <- substr(total$INCIDENT_DATE_STR, 1, 2)
total$INCIDENT_DATE_STR <- factor(total$INCIDENT_DATE_STR, levels = c("1/",
"2/", "3/", "4/", "5/", "6/", "7/", "8/", "9/", "10", "11", "12"))
levels(total$INCIDENT_DATE_STR) <- c("Jan", "Feb", "Mar", "Apr", "May",
"Jun", "Jul", "Aug", "Sept", "Oct", "Nov", "Dec")
```

```
ggplot(data = subset(total, Civilian_Officer == "Civilian")) +
  geom_bar(mapping = aes(x = INCIDENT_DATE_STR, fill = CRIME_REPORT_FILED),
position = "fill") +
  labs(x = "Incident Month", y = "Proportion", fill = "Crime Report Filed",
title = "Crime Report Status by Incident Date") +
  scale_fill_manual(values = c("#e7298b", "#7470b3"), labels = c("Not Filed",
"Filed"))
```

References

Use of Force (2016 - 2019) data and Incident Report (2016 - 2019)
https://openjustice.doj.ca.gov/data

California County Population:
https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html

California County Income:
https://www.indexmundi.com/facts/united-states/quick-facts/california/median-household-income#table