

Semester Project: Targets of Unsolved Shootings

Sam C, Stephen K, Vance C, Jake G

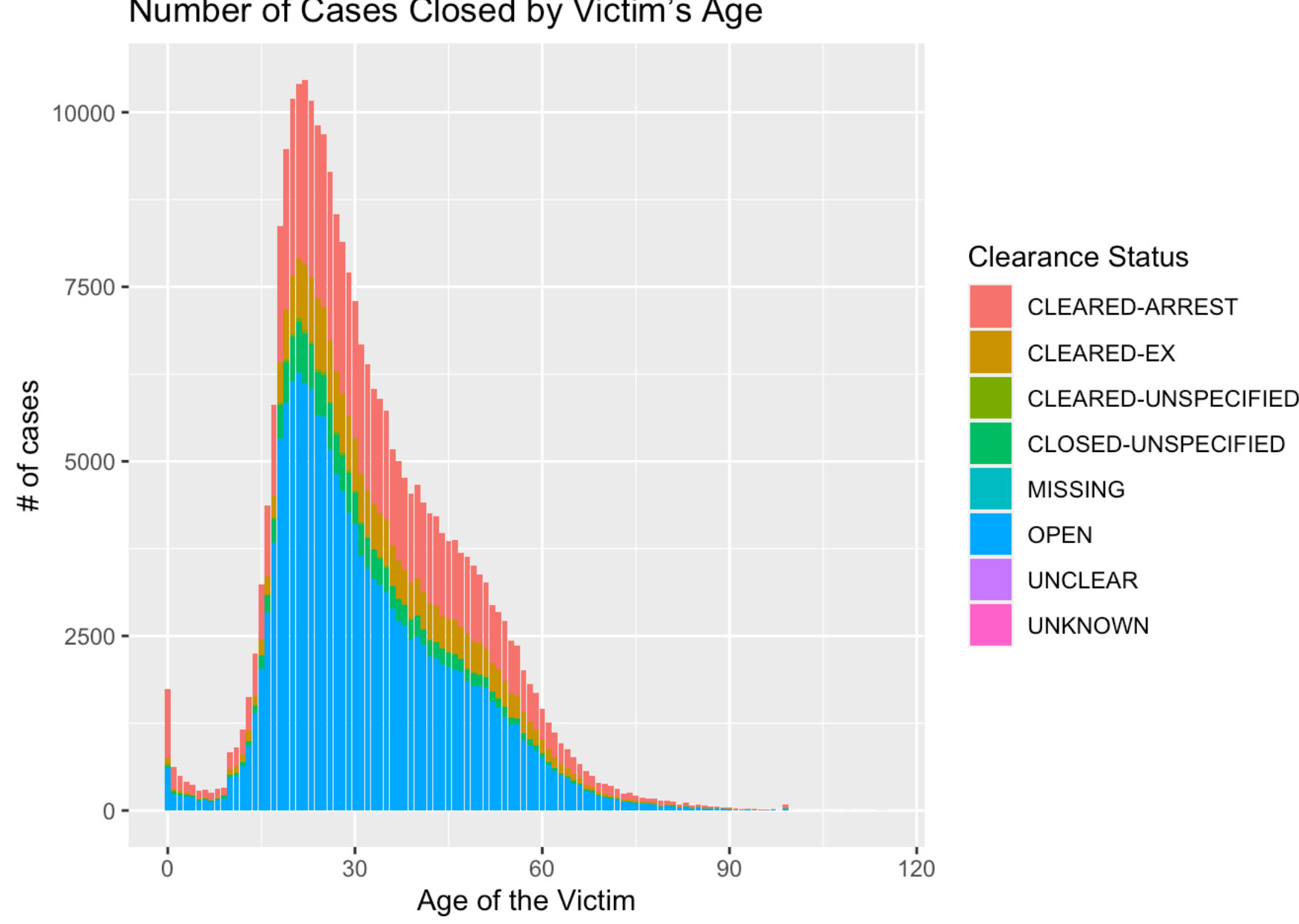
11/16/2020

Introduction

Does your age have any impact on your importance in the eyes of the Police? Does being 90 years old make you more or less likely to have a crime against you solved? As a group we wanted to see if there was a connection between age and the likelihood of your case being solved. We wanted to see if there was any sort of noticeable pattern or anything that should a strong correlation between age and the likelihood of the criminal being found. These are the types of questions that we felt like we wanted to know more about. For our data we looked at the "Unsolved Shooting" data which consists of three major sets of data. The first being the data from 22 municipal police departments directed towards discovering present day trends in arrest rates. The second set consists of data from Baltimore which has a relatively high crime rate, and looked at 3500 fatal and non-fatal shootings in the city. The last was an FBI report that helped examine the historical trends in unsolved shootings. The data comes from some of the largest cities in the US including places such as Chicago, LA, and Boston PD. The records requested from each police department were homicide reports, aggravated assault, and non-fatal shootings. Although the police departments did not choose to give over all of the data there is still enough to begin looking at trends. The data was a mix of public records and some from online sources. The data itself in an observational study, this is true because all of the data is there for anybody to look at and see because they are public record. So the people who we got the data from did not go out and experiment on this but instead began looking for trends in the already massive amount of data presented. In every observational study there is most likely bias because humans themselves are always biased. These biases can be seen in the data collection process. Although we got the data from police departments and public online records, the input of information can always be biased along with the absence of missing information. One reason could be in crimes not being recorded for certain people and the overall effectiveness of the police departments in question. Data can also be incorrectly entered into a computer which can cause bias in the data as well.

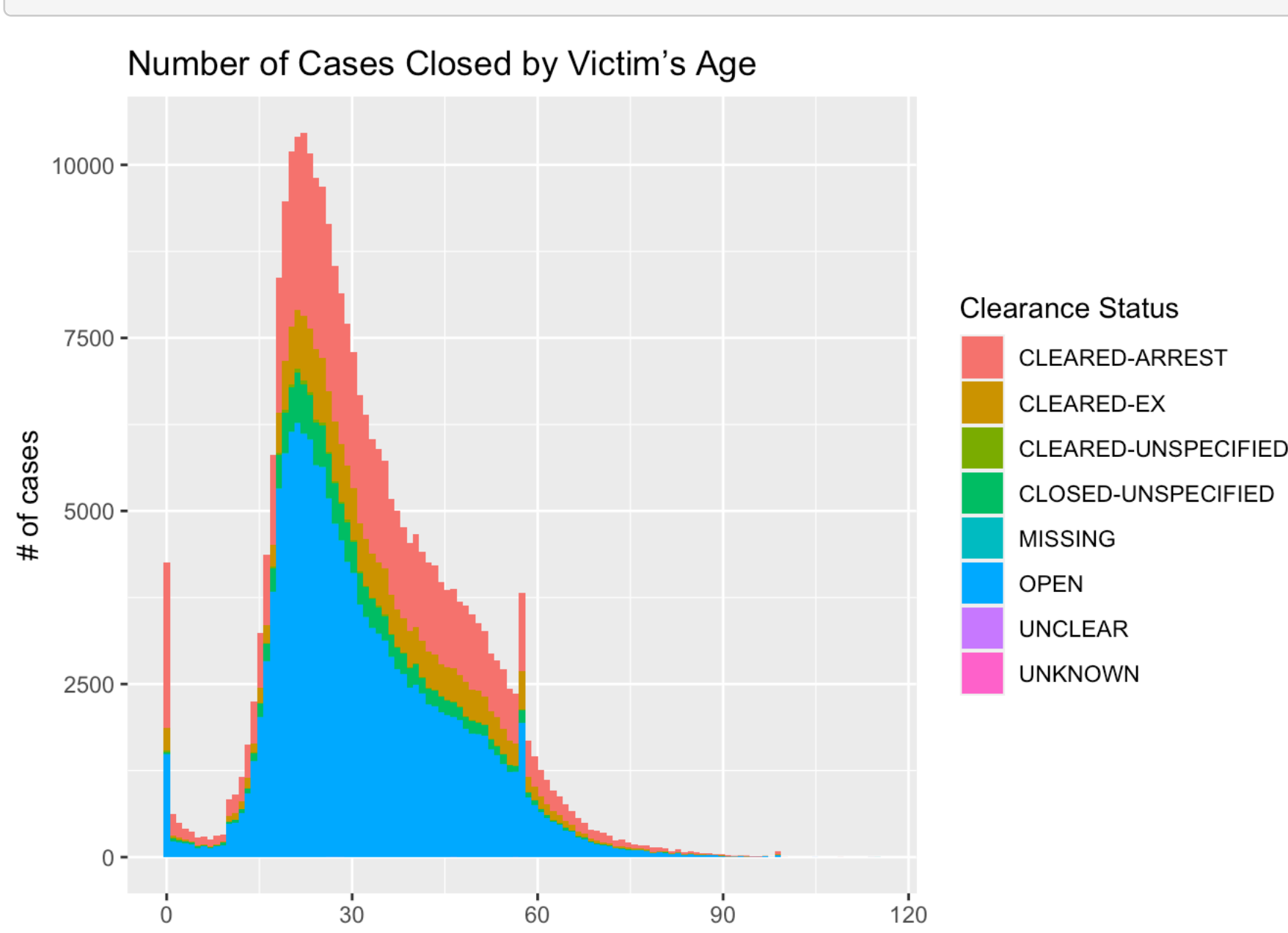
Data Analysis

```
invisible(Sclose <- S2 %>%
  filter(clearance_status != '')
)
Sclose$victim_age <- as.numeric(Sclose$victim_age)
Sclose$victim_age <- ifelse(Sclose$victim_age > 115, NA, Sclose$victim_age)
thePlot1 <- ggplot(data = Sclose,
  mapping = aes(x = victim_age, fill = clearance_status)) +
  geom_bar() +
  labs(fill = "Clearance Status", y = "# of cases", x = "Age of the Victim", title = "Number of Cases Closed by Victim's Age")
thePlot1
```



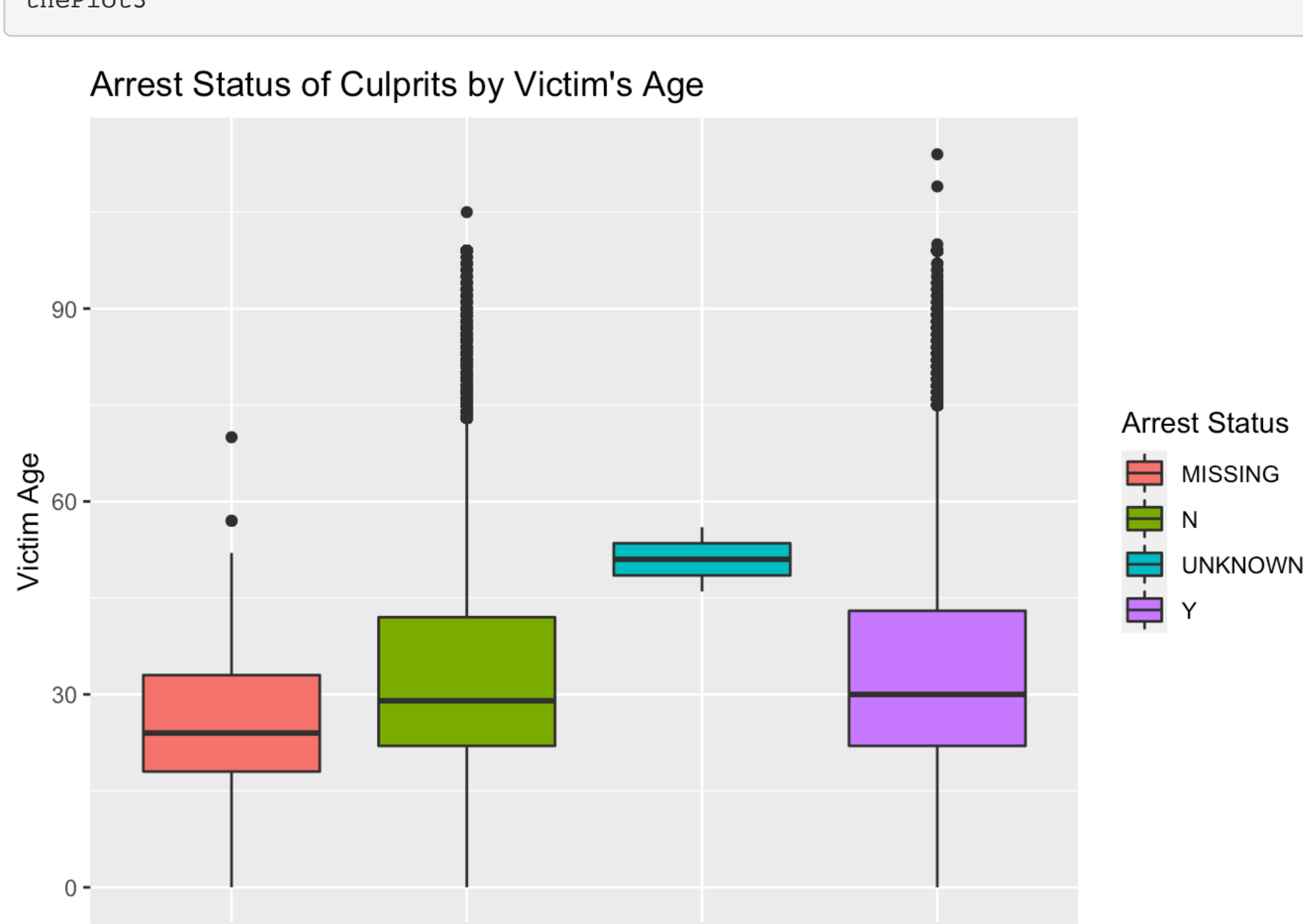
This graph shows the bell-curve relationship of the age of the victim on the clearance status of the cases, meaning whether the case against the culprit is cleared or closed or otherwise open. If a case is cleared when police make a related arrest, the case is turned over to prosecution, or formal charges are filed. A case is closed when the investigation into that case is resolved. A case's status may remain unknown or unclear when the police agency does not specify the status of the case or details on the case are clearly communicated. For further clarity, "CLEARED-ARREST" means that the case was cleared by an offender being arrested, and "CLEARED-EX" (exceptional means) means that the offender was identified but the policing agency was unable to arrest the offender, recording the culprit's offence and their location for a possible third arrest (<https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/clearances>). This curve shows the number of cases shows a few different relationships in the data, such as a steep increase in cases where the victim is 10 years old, from around 700 cases where the victim is 10-years old to around 8,200 cases when the victim is 18-years old. This is followed by a peak of cases with the victims 22-years old at a staggering number of cases around 10,625, and a gradual decrease of arrests of the culprits after age rises above 57 years old, where the number of cases suddenly peaks at around 3750 cases. The plot of the cases comes with outliers in the data at 0 and 57 years old, where there are sudden spikes of cases. The first steep increase could be because of the harsher treatment and punishments against criminals who attack children in court and in society, as well as the criminals possibly having or coming from a family that might be reminded by the age of these children. These two possibilities likely deter many people from committing heinous crimes, but not it the same for the elderly until after the victim is past 57 years old. Before then from around a victim age of 20 to 56 years old, the rate of crime declines at a slower rate possibly because of the absence of such a serious societal hatred towards acting against the particular age group. This data shows us that crime with victims of 56 years or older has less of a chance to be cleared as an arrest than cases of crime against 20-30 year old victims, along with percent of open cases to total cases by victim age increasing between the ages of 35 to 56. The sudden spike of around 4375 cases when the victim's age was less than a year old could be from mishaps or poor practices in raising an infant, while the spike in cases at 57 years of age continues to baffle me. These observations show the bias of officers towards cases with younger victims than older victims.

```
invisible(Sclose <- S %>%
  filter(clearance_status != '')
)
Sclose$victim_age <- invisible(ifelse(Sclose$victim_age == "UNDER 18", 0, Sclose$victim_age))
Sclose$victim_age <- as.numeric(Sclose$victim_age)
Sclose$victim_age <- ifelse(Sclose$victim_age > 115, NA, Sclose$victim_age)
thePlot2 <- ggplot(data = Sclose,
  mapping = aes(x = victim_age, fill = clearance_status)) +
  geom_histogram(bins = 115) +
  labs(fill = "Clearance Status", y = "# of cases", x = "Age of the Victim", title = "Number of Cases Closed by Victim's Age")
thePlot2
```



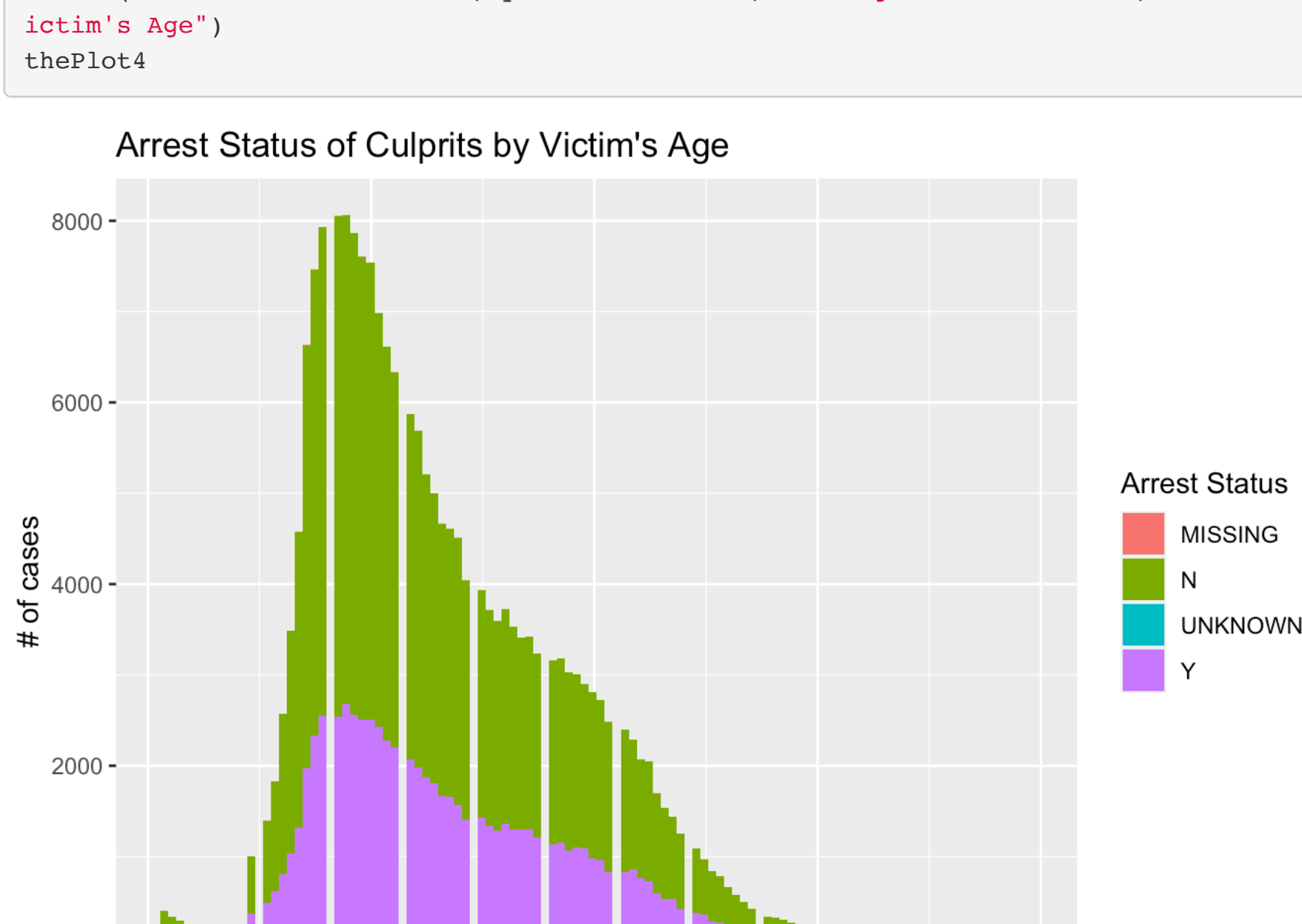
This bar graph shows the clearance status of culprits based on the age of the victim, similar to the data shown in the last graph. If a case is clear, the However, this graph shows the cases of 2,311 of the 84,750 cases labeled "under 18-years old", as shown by the reassignment of these values of "UNDER 18" of zero, and does not show the over 400,000 cases who were declared unspecified with no number entry. This graph highlights a source of error in our data which could act as a confounding variable for our argument, as the unspecified age victims could be disproportionately elderly or not. The existence of this confounding variable is vital in interpreting the meaning behind the data and the validity of the argument.

```
Sclose3 <- Sclose %>%
  filter(!is.na(victim_age), victim_age != "UNDER 18") %>%
  group_by(arrest_ind) %>%
  filter(victim_age <= 115) %>%
  filter(arrest_ind != '')
Sclose3$victim_age <- as.numeric(Sclose3$victim_age)
thePlot3 <- ggplot(data = Sclose3,
  mapping = aes(y = victim_age, x = arrest_ind, fill = arrest_ind)) +
  geom_boxplot() +
  labs(x = '', y = 'Victim Age', fill = "Arrest Status", title = "Arrest Status of Culprits by Victim's Age")
thePlot3
```



This boxplot shows the status of arrest of the culprit by the victim's age with each case. The plot shows a similar portion of cases ending in an arrest as with not an arrest for cases with victims ranging from 0 to 75 years of age, sharing the same first and third quartiles (Q1 at 22.5 years old, and Q3 at around 42 years old). However, cases with the culprit's arrest status labeled "Unknown" hold victim ages only between 45-55 years, and cases with arrest status of "Missing" have cases with victims ranging mostly from 10-50 years. These cases paint a different picture of the police possibly neglecting more cases with younger victims than cases with older victims, possibly because more victims of crimes at younger ages before 30 are kept off the record by the police. Overall, this data shows how the proportions of cases with victims between 0-25 years and cases with victims between 45-55, when they are reviewed by the police, are treated similarly, showing the cause of this misrepresentation in cases with victims of different age ranges likely coming from the authorities giving more time to younger victim age ranges than older victim age ranges.

```
Sclose2 <- S %>%
  filter(!is.na(arrest_ind), !is.na(victim_age)) %>%
  filter(victim_age != "UNDER 18") %>%
  filter(victim_age >= 115)
vicAge <- as.numeric(Sclose2$victim_age)
vicAge <- ifelse(vicAge > 115, NA, vicAge)
thePlot4 <- ggplot(data = Sclose2,
  mapping = aes(x = vicAge, fill = arrest_ind)) +
  geom_histogram(bins = 110) +
  labs(fill = "Arrest Status", y = "# of cases", x = "Age of the Victim", title = "Arrest Status of Culprits by Victim's Age")
thePlot4
```



This histogram shows the arrest status of unsolved shootings by the age of the victim. Above, we see the trend of a peak in unsolved shootings at around victim ages of 15-30 years old, with around 2,500 cases ending in arrests and 5,500 cases ending in no arrest when the age of the victim is around 20 years old. This is followed by a sudden decline in cases after 56 years old, with the number of arrests made when the victim was 50 years old numbering a little over 1,000 cases and cases ending in no arrest numbering around 1,800. However, we also see a sharper decline in cases where the culprit was arrested after victim ages of 18 years old than from the entirety of cases between victim ages of 25 to 75 years. This could possibly be because of the decline in cases from societal pressure referenced under the first graph, or this could be from the greater antagonizing of criminals who harm children than criminals who harm adults, showing an indirect bias in law enforcement against adults and the elderly. At the same time in the victim age range from 25 to 35 years old, the number of cases rapidly decreases, which could possibly undermine the weight of the indirect bias in law enforcement references before. Through the evidence given, this plot shows the reader evidence of a possibly bias that law enforcement has in judging cases based on the victim's age.

Conclusions

Most shooting cases occur to those who are young adults, with the lowest cases happening to the elderly. As seen in the histogram, the clearance status among each age appears to be mostly equally distributed, indicating that each case seems to be handled similarly. However, any difference in treatment may be caused by police giving more attention victims of a younger age than an older age. In terms of arrest indication by age, it does show that those in the young adult range may have higher rates of no arrest than other age groups. This could show some possible bias that law enforcement has based on a victim's age. The fact that the most shootings occur to the younger age group may also have an impact on the arrest status.

Limitations / Recommendations

It is hard to fully understand how many victims there were in each individual age that fall in the range of 0 to 18 because a lot of the victims' ages are labeled as "UNDER 18." This is important to factor in because a very large amount of cases is labeled under that category, which could significantly impact the data shown. The idea of who you are as a person may have a potential impact on your importance to police leads to a great variety of other research in other areas such as ethnicity or gender impacting importance, not just age. Also, understanding the ages that are impacted the most by shootings can lead to research understanding why this is and how to prevent it. There could also be more exploration in understanding fatal and non-fatal shootings among victim age groups.