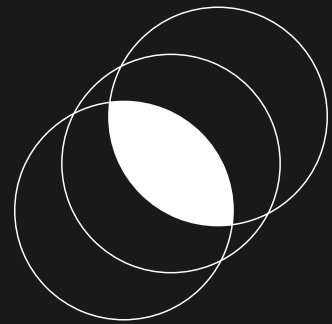


High Profile Police Killings in 2015

05/09/2023 —————> Presented by Camryn C. Way



Intro

The year 2015 was marked by several high-profile incidents of police violence, including the shooting of Walter Scott in South Carolina and the death of Freddie Gray in Baltimore. These incidents sparked a national conversation about police use of force, racial bias, and accountability.



Walter Scott during his service in the U.S. Coast Guard

Research Question

What were the patterns and trends of police killings in the United States in 2015, and how did they vary by demographic factors such as location, income, ethnicity, and age?

Introduction of Data	Page 4-5
Noteworthy Exploratory Analysis	Page 6-12
Logistic Regression	Page 13-21
Final Predictions	Page 22
Conclusion	Page 23

Data Sources

01

The Guardian Media Group

Launched an interactive database with info on Americans killed by police in the year 2015 named, "The Counted".

The Database is a combination of media coverage, reader submissions, police and coroners' reports, voter registration data, witness testimony, court records, photographs, and reliable open-sourced efforts like "Fatal Encounters" and "Killed by Police".

02

American Community Survey

Census data was calculated at the tract level from the 2015 5-year American Community Survey using the tables S0601 (demographics), S1901 (tract-level income and poverty), S1701 (employment and education) and DP03 (county-level income).

Census tracts were determined by geocoding addresses to latitude/longitude using the Bing Maps and Google Maps APIs and then overlaying points onto 2014 census tracts.

03

Colleges.csv

The data set `colleges.csv` contains information about 1991 graduation rates for the top 200 American universities and colleges determined by "Money" magazine.

Note: This dataset was only used to pull the column `Region`.



Brief Overview of Key Variables

Each observation represents an individual who was killed in America by police during the year 2015. The police_killings data set features 467 observations and 34 Columns.

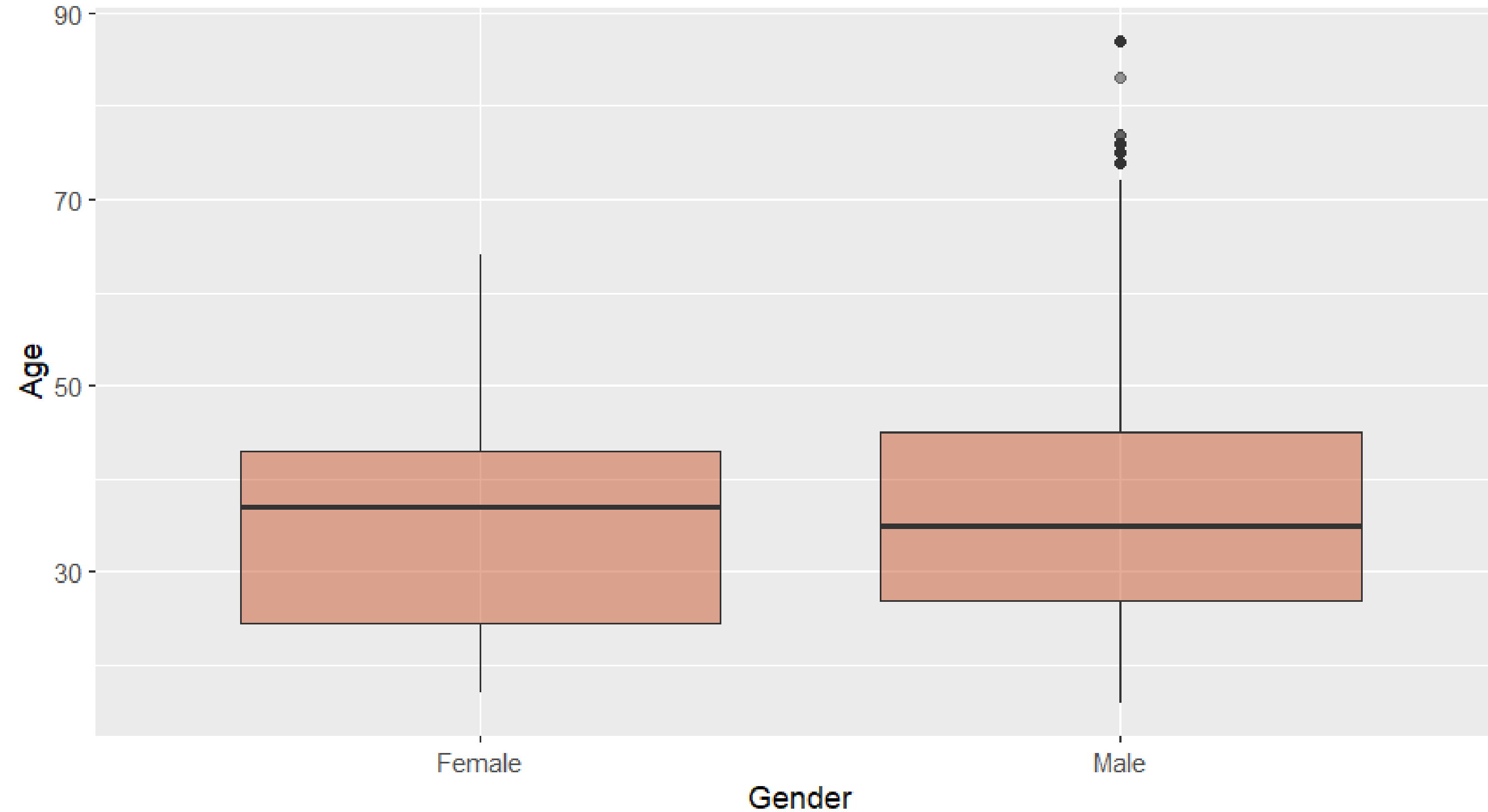
The dependent variable for the predictive model is Race/Ethnicity split into the categories of minority and non-minority.

The independent variables comprise a wide range of factors, including the victim's age, household income, county income, gender, incident location, cause of death, armament, as well as the college and unemployment rate of the area.



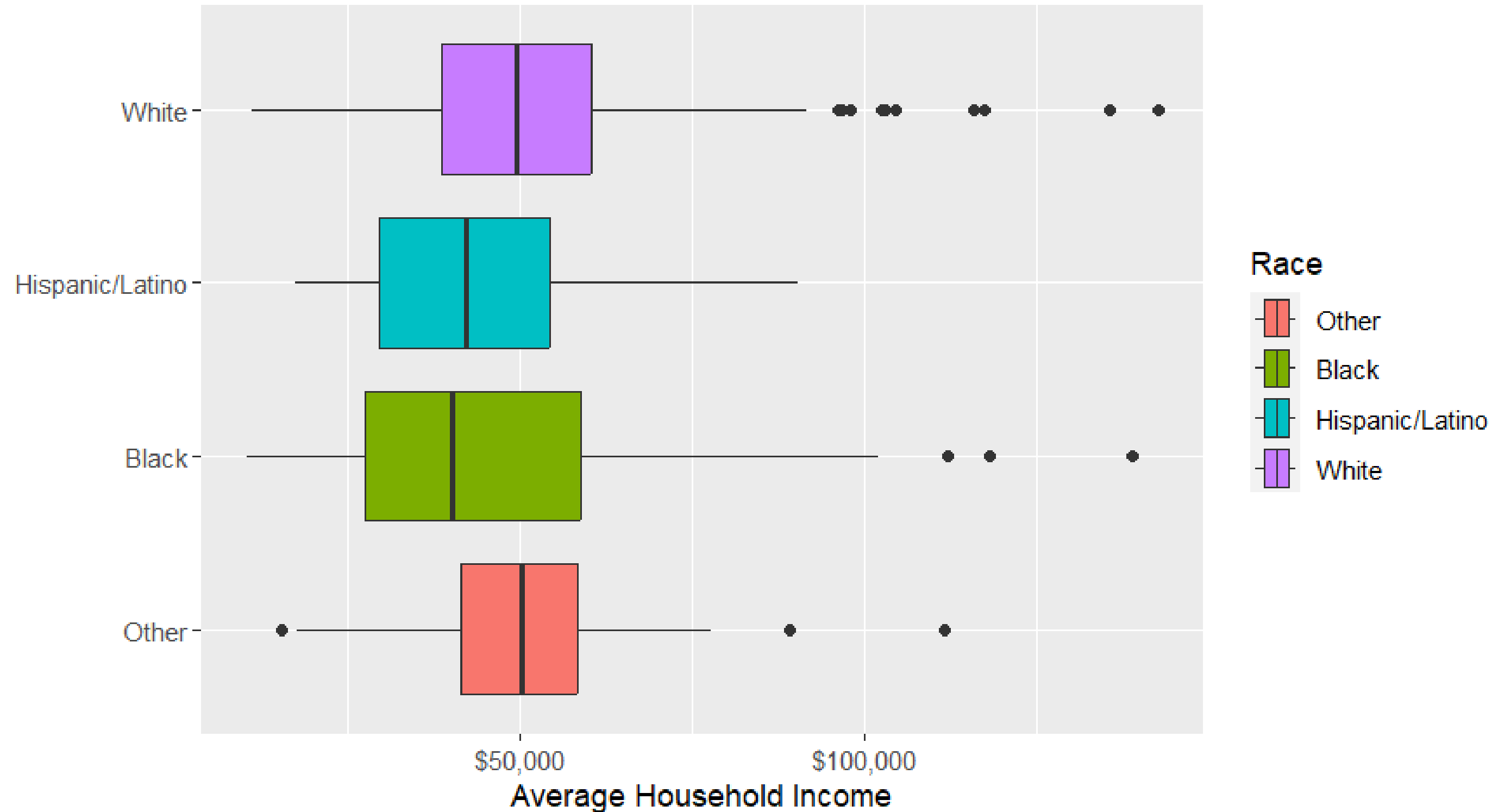
Age Distribution by Gender

Among Individuals Killed by Police in 2015



- Males have a distribution with a larger range and a greater number of outliers compared to females.
- This pattern is particularly evident when examining the age of the individuals in the dataset. The oldest male (Louis Becker) was 87 years old at the time of his death, while the oldest female (Hue Dang) was only 63 years old.

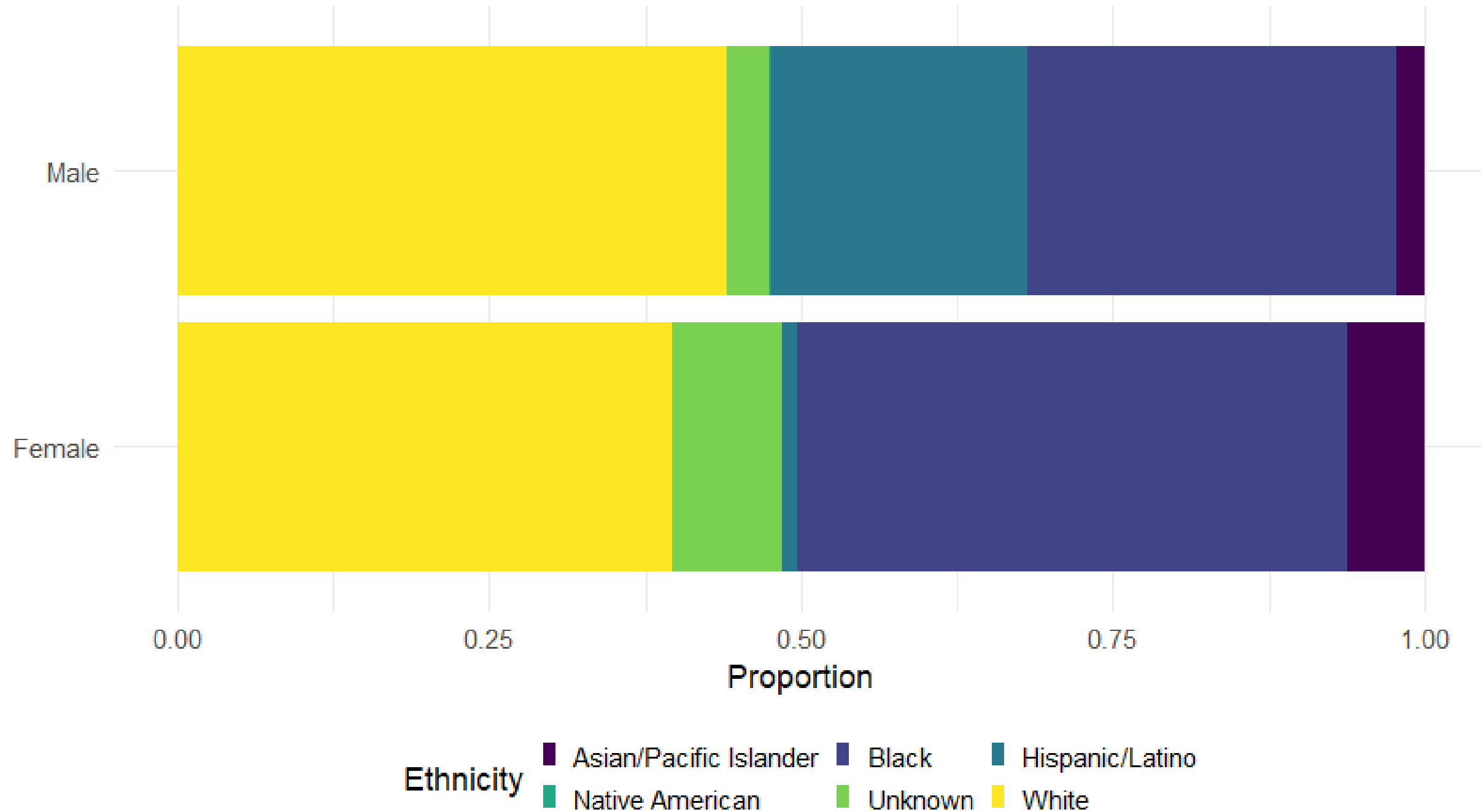
Average Household Income by the Three Most Common Races Among Individuals Killed by Police in 2015



- Hispanic/Latino households have the lowest income range among the demographic groups analyzed.
- White households have the largest income range, followed by Black households.

The United States Census Bureau estimates the real median household income in the United States was \$55,775. Both the Mean and median household income for all people who have been killed in America by police is below that average at 46,627.18 and 42,759 respectively. Similarly, the median for all of our racial groups in the above distribution is at or below 50,000.

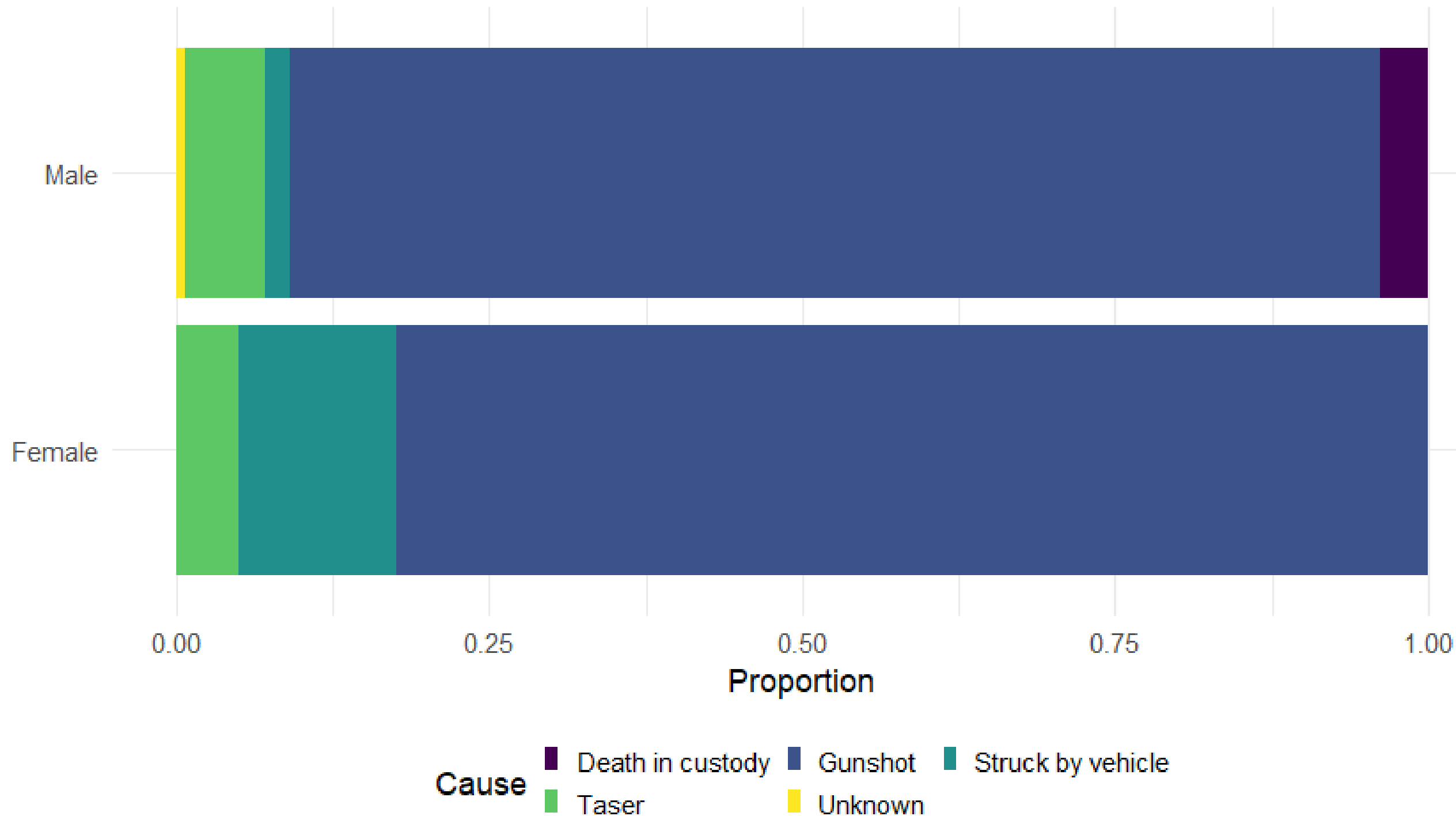
Distribution of Ethnicity by Gender
Among Individuals Killed by Police in 2015



- White men accounted for the highest proportion of police killings among all other male racial groups. at 44%
- Black women accounted for the highest proportion of police killings among all female racial groups at 44%.
- There is a higher proportion of male Hispanic/Latino Men(20.4%) killed than Hispanic/Latino Women(1.3%).
- There is also a higher proportion of black women(44%) who were killed in comparison to black men(30%).

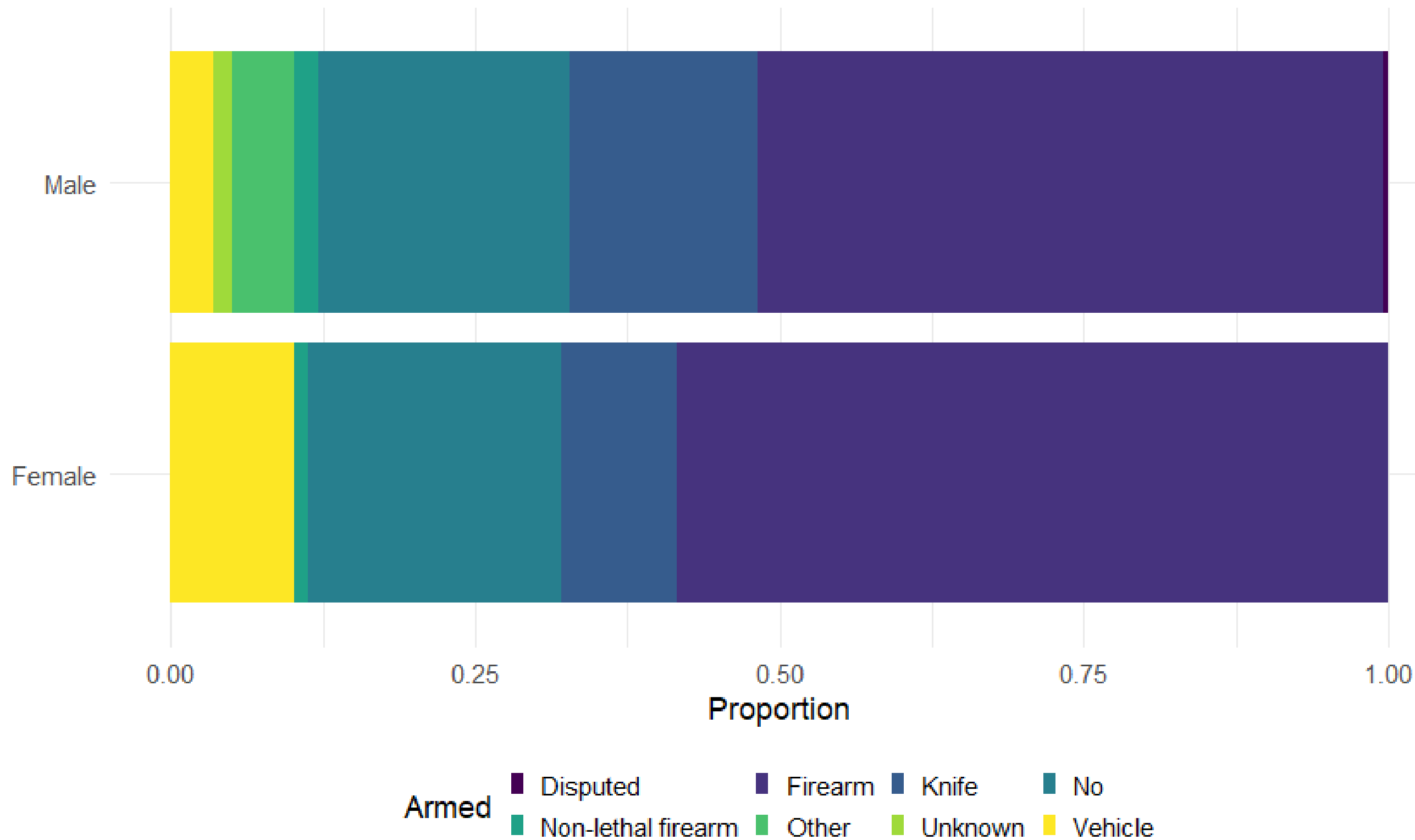
Distribution of Cause by Gender

Among Individuals Killed by Police in 2015



- 82% of women and 87% of men were killed by gunshots, which is a common method used by law enforcement.
- 5% of women were killed by tasers versus 6.4% of men.
- It is also very important to note that only male victims had a proportion (3.8%) of individuals who died in custody.
- 12.5% of women were killed by being struck by a vehicle versus only 2% of men.

Proportion of Police Killings by Gender and Armed Status in 2015



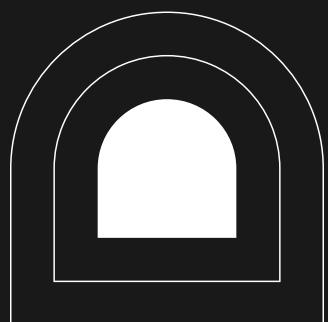
- The data shows that 58.5% of women were carrying a firearm at the time of their killing versus only 51.4% of men
- A higher proportion of men were killed while carrying a knife(15.4%) than women(9.4%).
- A relatively equal proportion of both men and women were killed while having no weapon in hand.
- And a higher proportion of women(10%) than men(1.6%)were killed who were considered to be using a vehicle as a weapon

Police Killings

Future research could benefit from examining specific states with high numbers of police killings, such as California and Texas, in order to gain a more detailed understanding of the patterns and trends in police violence in those areas.

```
## # A tibble: 377 × 2
##   lawenforcementagency      n
##   <chr>                  <int>
## 1 Los Angeles Police Department    130
## 2 New York Police Department      104
## 3 Pennsylvania State Police        66
## 4 Los Angeles County Sheriff's Department    65
## 5 Fort Worth Police Department      48
## 6 San Francisco Police Department    39
## 7 Austin Police Department         36
## 8 El Paso Police Department        36
## 9 San Antonio Police Department      36
## 10 New Jersey State Police         30
## # ... with 367 more rows
```

The Los Angeles Police Department has the highest number of police killings in the dataset, followed by the Los Angeles County Sheriff's Department and the Oklahoma City Police Department. These findings reinforce the observation from the US Proportions map that California is a hotspot for police killings. It is important to investigate why certain law enforcement agencies and geographic regions have higher rates of police violence in order to identify potential contributing factors to develop targeted interventions.



Logistic Regression

Creating Minority Groupings

[Back to Intro Page](#)

Data Wrangling

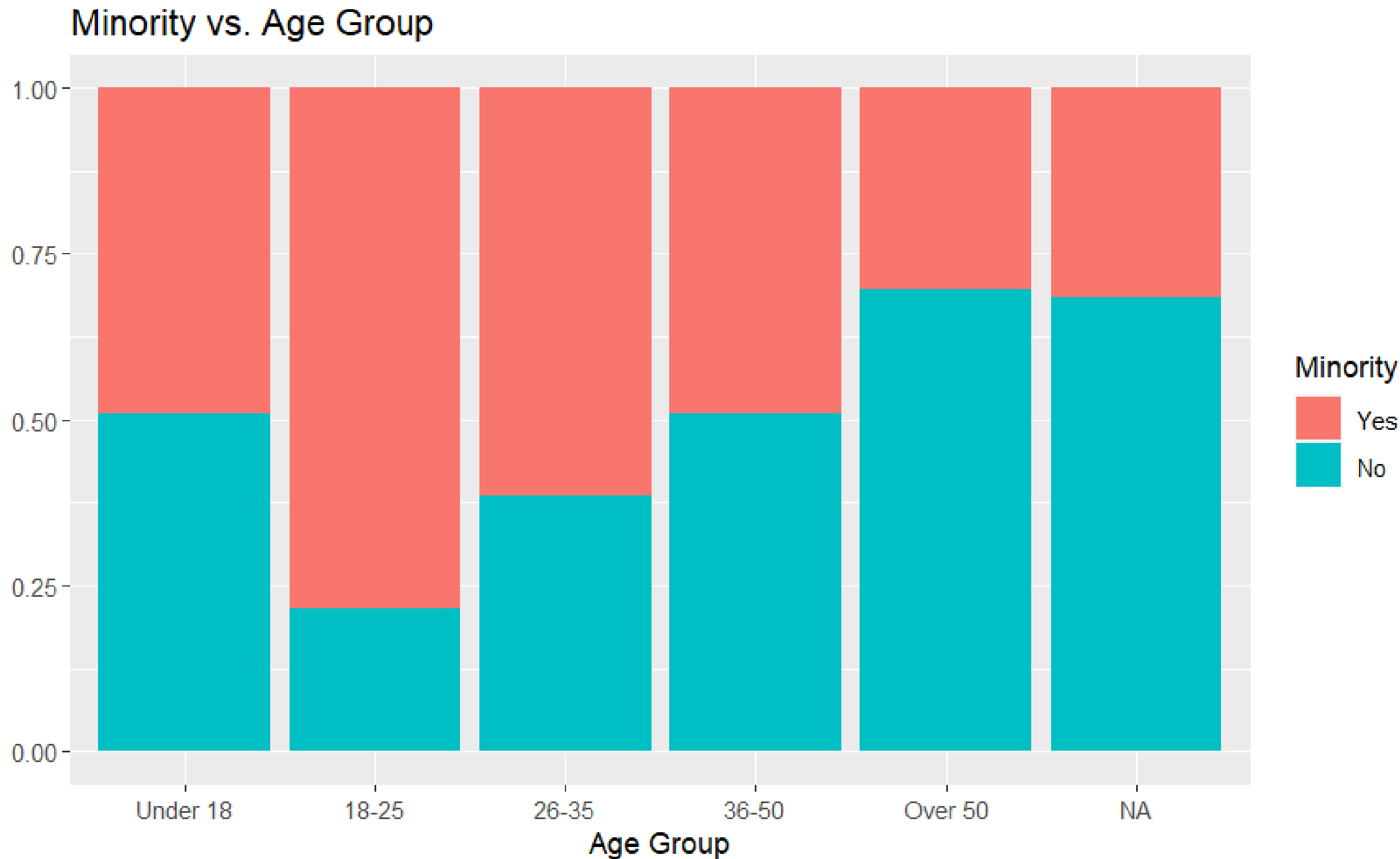
```
```{r lump_race, echo = FALSE, warning=FALSE}
pkfilt <- pk_age_grp %>%
 filter(raceethnicity != "Unknown")
minority_categories <- c("Asian/Pacific Islander",
 "Black",
 "Hispanic/Latino",
 "Native American")
...

```

```
```{r, echo = FALSE, warning=FALSE}
pk_race_cat <- pkfilt %>%
  mutate(minority = factor(ifelse(raceethnicity %in% minority_categories, 0, 1)))
...

```

The mean age of minority racial groups, including Asian/Pacific Islander, Black, Hispanic/Latino, and Native American, is 33.5, which is lower than the mean age of non-minority groups, predominantly white, at 40.5.



- It is apparent that a larger number of minorities who were killed by the police in 2015 fell within the 18-25 and 26-35 age groups, whereas more non-minorities were killed when they were over 50 or their age was unknown.
- It's also important to note that the under 18 and 36-50 age groups are relatively even in being a minority or non-minority.

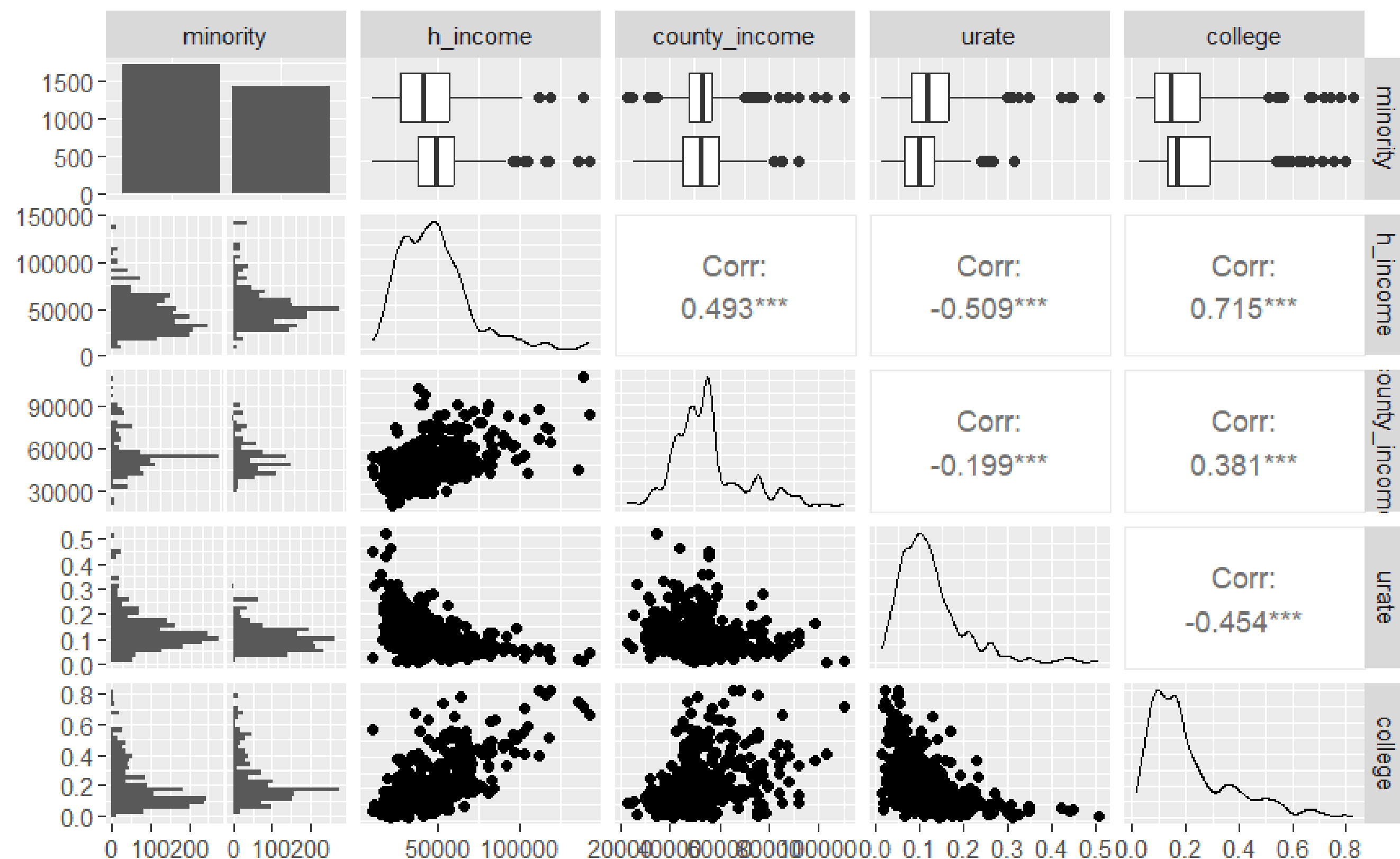
Selecting a Model

Model 1: Contains all potential useful interactions for the analysis.

```
glm(formula = minority ~ age + armed + region + urate + college +  
    cause + h_income + county_income, family = "binomial", data = pk_race_cat)  
  
Null deviance: 4238.9  on 3078  degrees of freedom  
Residual deviance: 3474.8  on 3057  degrees of freedom  
(71 observations deleted due to missingness)  
AIC: 3518.8
```

Model 2: Contains significant p-values

```
glm(formula = minority ~ age + h_income + region + cause + urate +  
    county_income, family = "binomial", data = pk_race_cat)  
  
Null deviance: 4238.9  on 3078  degrees of freedom  
Residual deviance: 3551.1  on 3065  degrees of freedom  
(71 observations deleted due to missingness)  
AIC: 3579.1
```



- There is a moderate positive correlation (0.461) between Household Income and County Income, which is expected. However, I do not believe it is significant enough to warrant its removal from the model.
- Overall the correlations I was worried about, such as college/urate and household income/county income, were not high enough to make any changes to the model.

Predictions Based On Just Age

By far, age seems to be the most significant interaction based on p-value. Breaking down the specific connection between age and minority may be worthwhile.

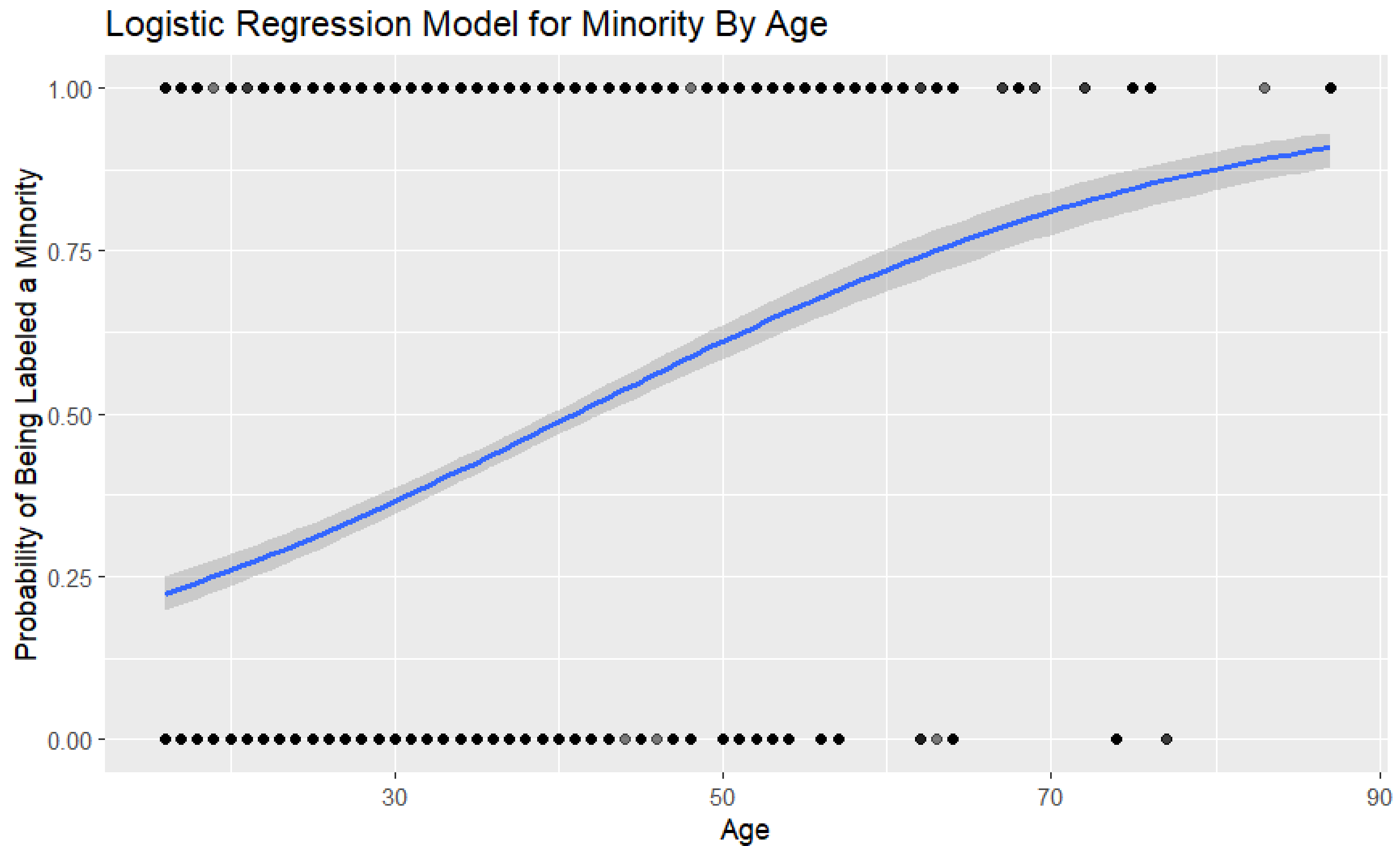
$$\widehat{minority} = -1.6630 + 0.04769age$$

```
newdata <- tibble(age = c(18, 37, 52))  
minority_fit %>% predict(newdata, type = "response" )
```

```
##           1           2           3  
## 0.2401395 0.4500058 0.6342176
```

Based on the analysis, the following probabilities were observed:

- The probability that a person who is 18 years old is a minority is 0.309.
- The probability that a person who is 37 years old is a minority is 0.525.
- The probability that a person who is 52 years old is a minority is 0.694.



Based on the logistic regression model, it appears that age has a significant impact on the likelihood of being a minority. As age increases, the probability of being a minority also tends to increase. This trend highlights the importance of age as a predictor variable in the model.

Confusion Matrix

Even though Model 2 only included the more significant interactions, it was outperformed in terms of prediction accuracy by Model 1. The larger model predicted at a rate of 72.71% accuracy, while the smaller model predicted at a rate of 71.08% accuracy. Additionally, removing the least significant interactions from the larger model one at a time and testing the accuracy also resulted in decreased accuracy estimates. These findings suggest that including more interactions in the model can improve its predictive performance.

Model 1: Useful interactions

Confusion Matrix and Statistics

		Reference	
Prediction		0	1
	0	266	94
	1	73	179

Accuracy : 0.7271

95% CI : (0.69, 0.7621)

No Information Rate : 0.5539

P-Value [Acc > NIR] : <0.000000000000000002

Model 2: Significant p-values

Confusion Matrix and Statistics

		Reference	
Prediction		0	1
	0	256	94
	1	83	179

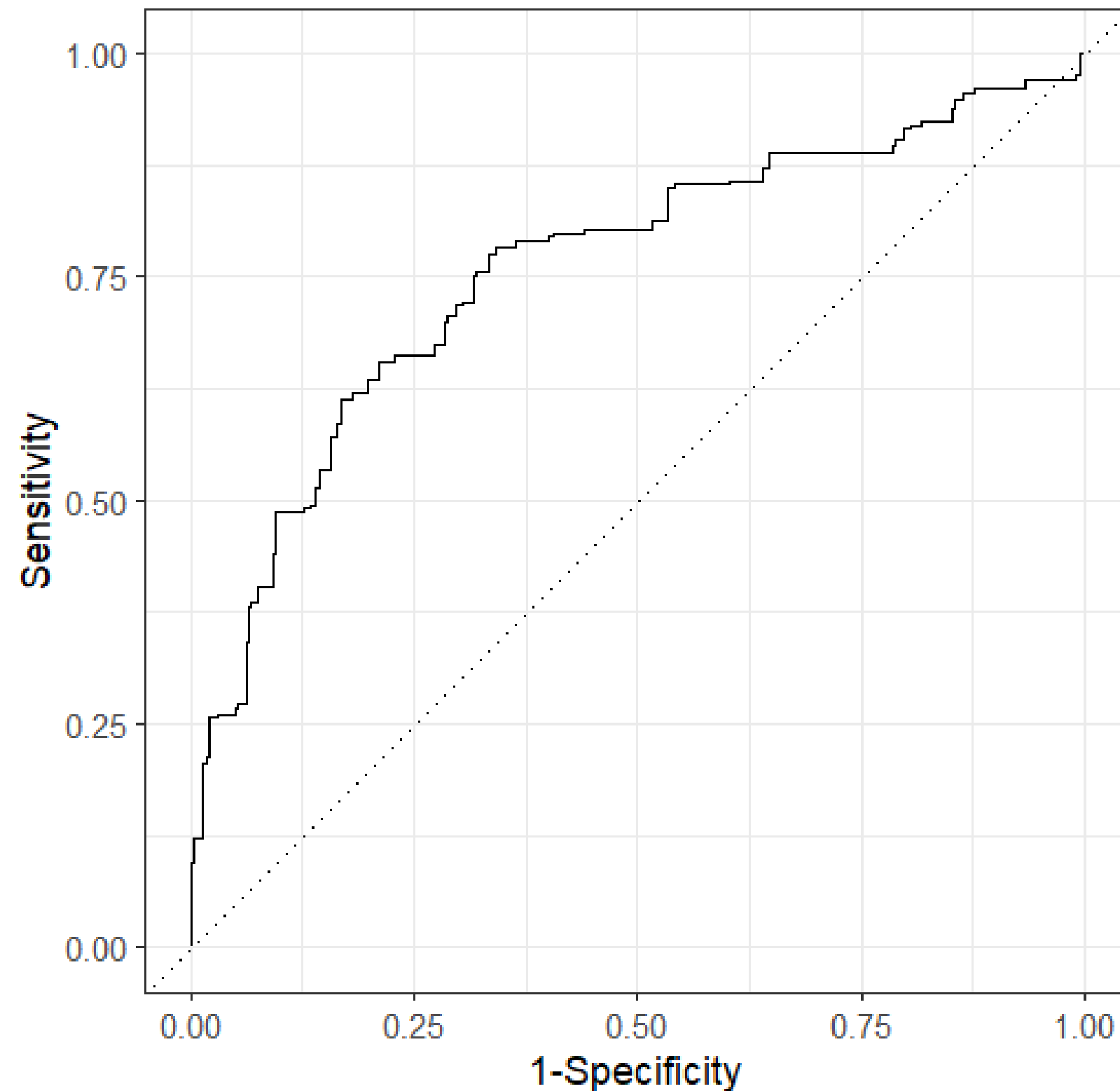
Accuracy : 0.7108

95% CI : (0.6731, 0.7464)

No Information Rate : 0.5539

P-Value [Acc > NIR] : 0.000000000000000001137

True Positive Rate vs. False Positive Rate for Minority



The model's AUC is 75.9, indicating that the model performs reasonably well in distinguishing between positive and negative cases.

The higher the AUC, the better the model's ability to differentiate between the two classes. In general, an AUC above 70% is considered good, while an AUC above 90% is considered excellent.

Final Predictions

```
newdata <- tibble(age=23,region="South",urate=0.03859900,college=0.5266234,cause="Gunshot",armed="No",h_income=57174,county_income=66006)
m_fit %>% predict(newdata, type = "response")
```

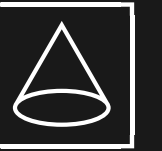
```
##           1
## 0.306456
```

Based on the logistic regression model built, we can make a prediction for whether or not a person with the given parameters is a minority or non-minority.

Let's say, for example, That the person in question is 23 years old, unarmed, from the south region, has a low unemployment rate of 0.03859900, a college rate of 0.5266234, a household income of 57174, a county income of 66006, and the cause of death is listed as a gunshot.

The model's computed probability that this person is a minority is 30.64%.

Conclusion



The analysis of police killings in the US in 2015 showed that demographic factors like age, income, ethnicity, and location had a significant influence on the patterns and trends of police violence.

- The data indicated that young people, especially minorities within the 18-25 and 26-36 age range, were more likely to be killed by police, while non-minorities were more likely to be killed in the over-50 or unknown age bracket.
- White men accounted for the highest proportion of police killings among male racial groups, while black women had the highest proportion of deaths among female racial groups and black men.
- The median household income for all people who have been killed in America by police is below the 2015 national average of \$55,775
- All of The above factors, as well as the Unemployment Rate(urate), College Rate, County Income, cause, and armament, resulted in fitting a model with a 72.7% accuracy rate for predicting whether an individual killed by police in 2015 was a minority or not.

Future research

- Consider additional variables such as the ethnicity of the police officer involved and the impact of police training and community policing initiatives on police killings. The findings have important implications for policy and training aimed at reducing police violence in the US.
- Examining specific states with high numbers of police killings, such as California and Texas, in order to gain a more detailed understanding of the patterns and trends in police violence in those areas.
- Compare 2015 data with future and past years.