

# final\_project

Eleanor Jiang 305002785

2024-11-16

## Read

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)

data <- read.csv("police_killings.csv")
set.seed(123)
sampled_data <- data[sample(nrow(data), 503, replace = TRUE), ]
shuffled_data <- sampled_data[sample(nrow(sampled_data)), ]

# Save the shuffled data to a new CSV file
write.csv(shuffled_data, "police_killings_processed.csv", row.names = FALSE)
```

## EDA

```
##           name           age           gender
##           0             0             0
##   raceethnicity         month           day
##           0             0             0
##           year   streetaddress           city
##           0             1             0
##           state       latitude       longitude
##           0             0             0
##   state_fp   county_fp   tract_ce
##           0             0             0
##   geo_id     county_id   namelsad
##           0             0             0
## lawenforcementagency   cause       armed
##           0             0             0
##           pop   share_white   share_black
##           0             0             0
##   share_hispanic   p_income   h_income
```

```

##           0           0           0
## county_income comp_income county_bucket
##           0           0           30
## nat_bucket pov urate
##           0           0           0
## college
##           0

## Warning: NAs introduced by coercion
## [1] "numeric"
## [1] 4

## age age_num
## 154 Unknown NA
## 193 Unknown NA
## 244 Unknown NA
## 355 Unknown NA

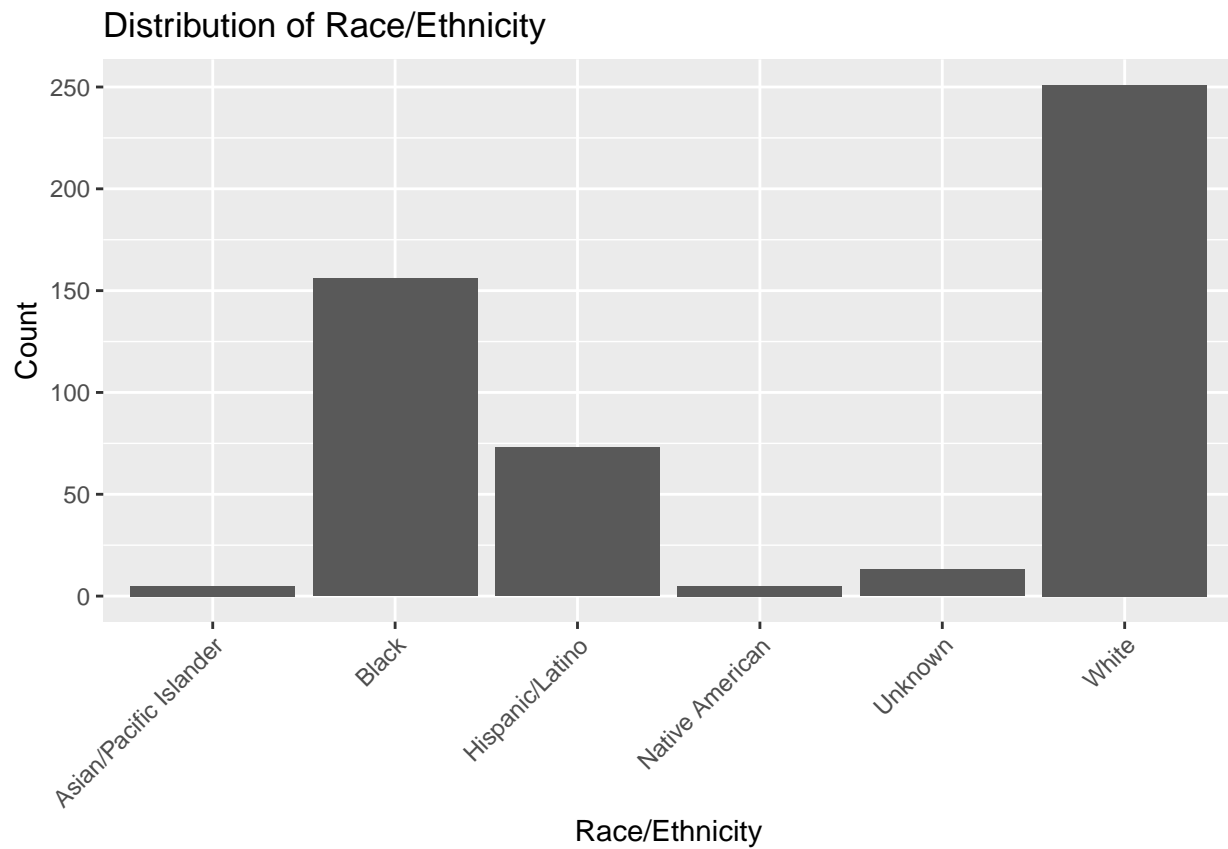
## Warning: NAs introduced by coercion
## Race/Ethnicity Distribution:

##
## Asian/Pacific Islander Black Hispanic/Latino
##           5           156           73
## Native American Unknown White
##           5           13           251

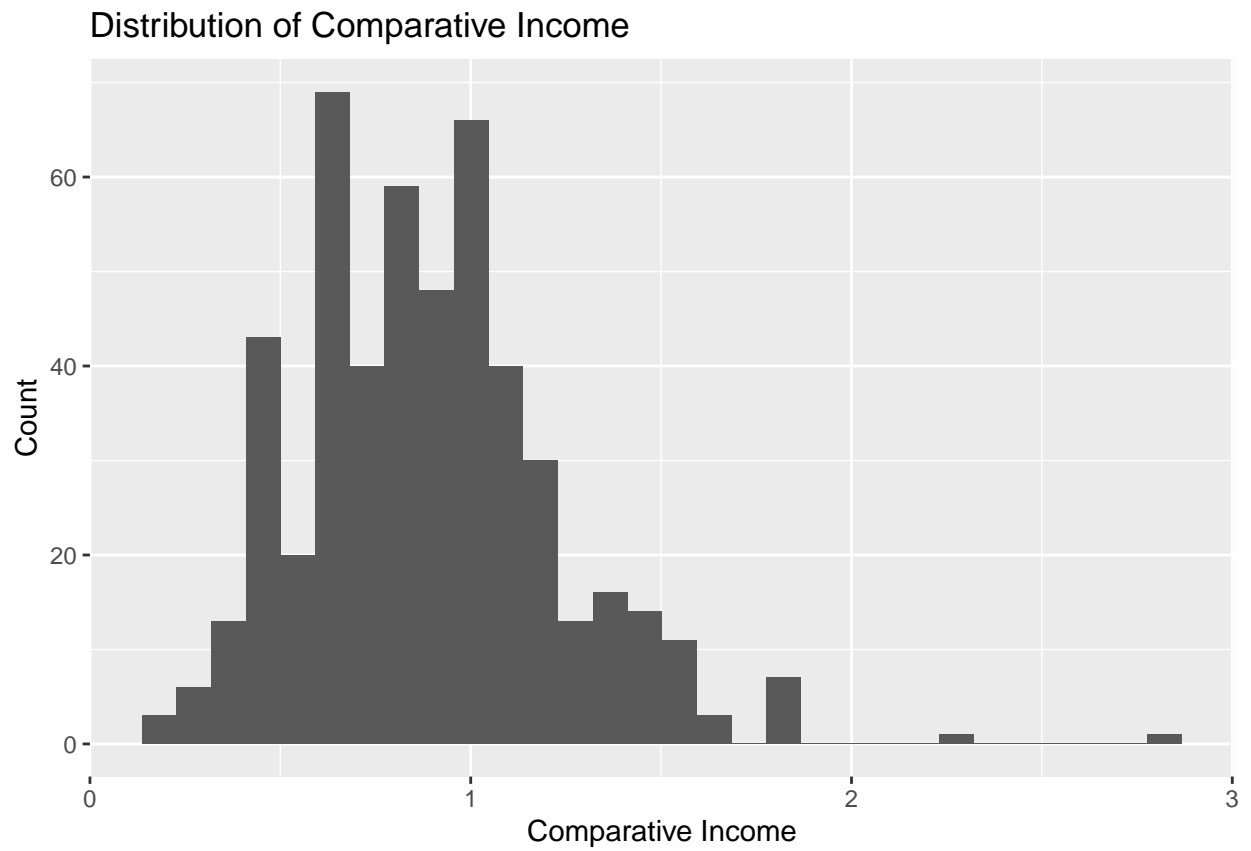
##
## Percentages:

##
## Asian/Pacific Islander Black Hispanic/Latino
## 0.9940358 31.0139165 14.5129225
## Native American Unknown White
## 0.9940358 2.5844930 49.9005964

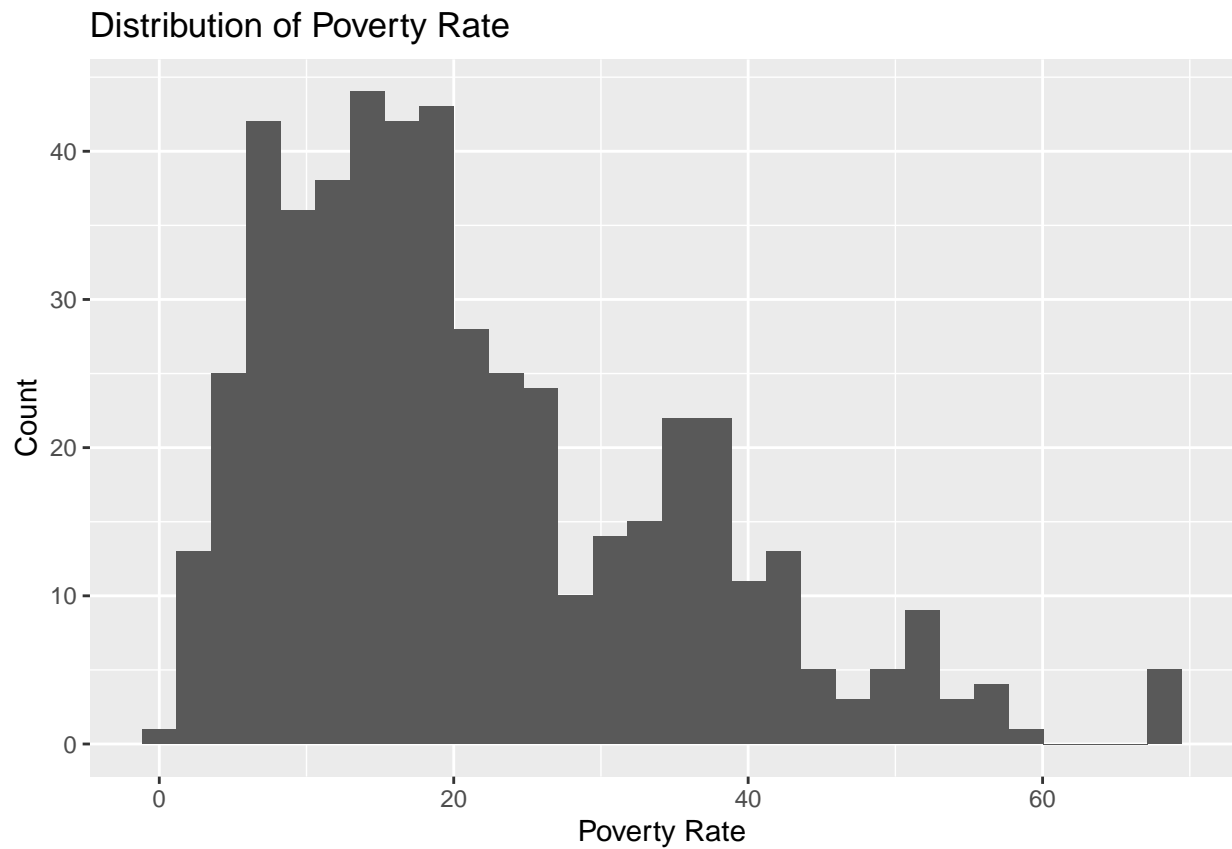
```



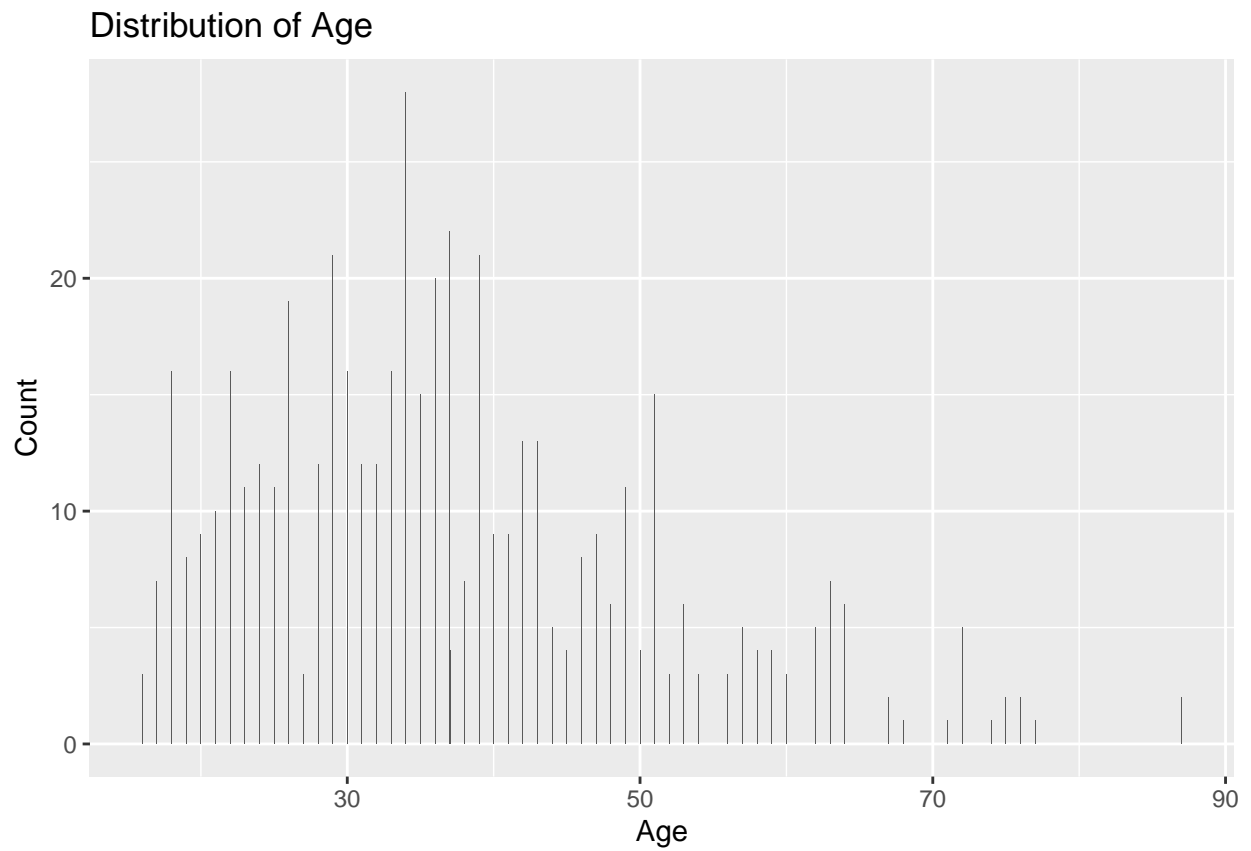
```
##
## Summary Statistics for Comparative Income:
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.2248  0.6374  0.8638  0.8870  1.0771  2.8652
```



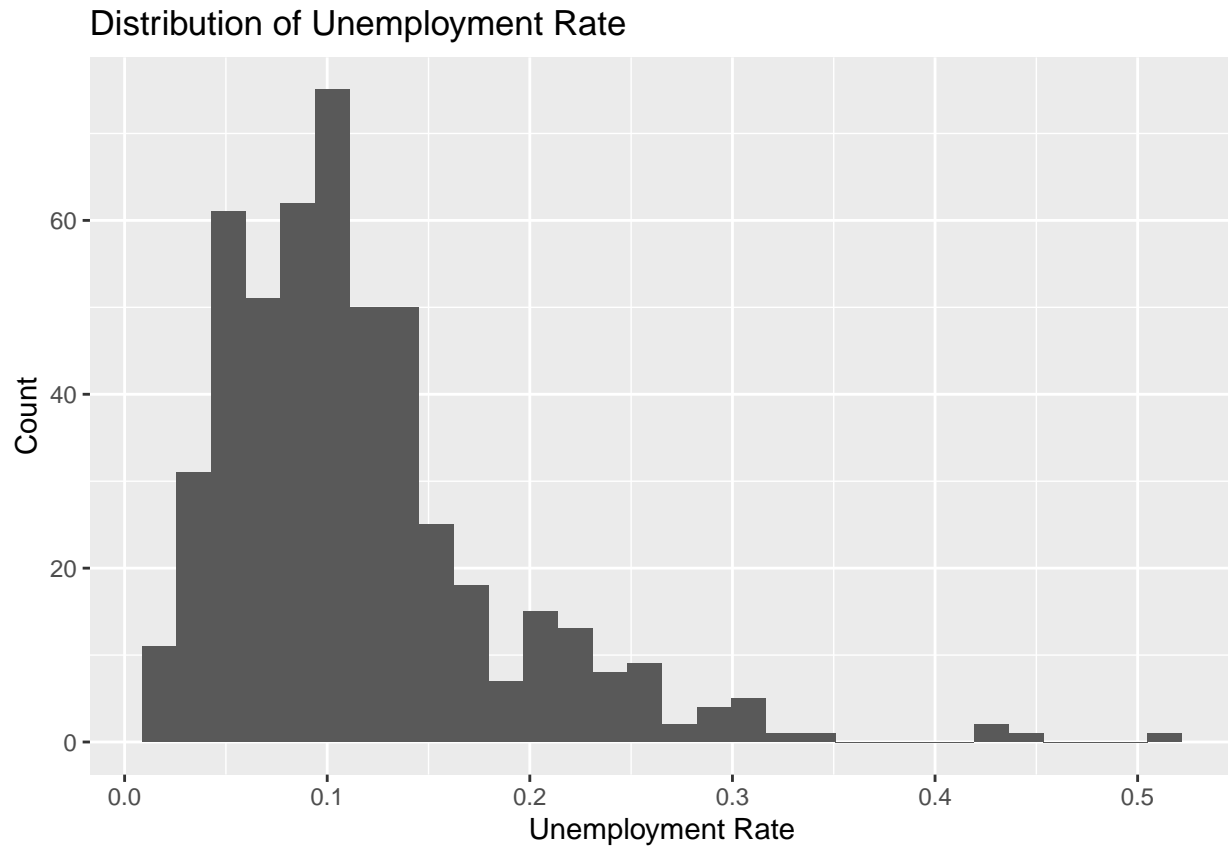
```
##  
## Summary Statistics for Poverty Rate:  
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##   1.10  10.90   18.00   21.49  31.10   69.40
```



```
##
## Summary Statistics for Age:
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  16.00  28.00   35.00   37.06  44.00   87.00
## Warning in geom_bar(bins = 30): Ignoring unknown parameters: `bins`
```



```
##
## Summary Statistics for Unemployment Rate:
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01133 0.06808 0.10205 0.11542 0.13860 0.50761
```

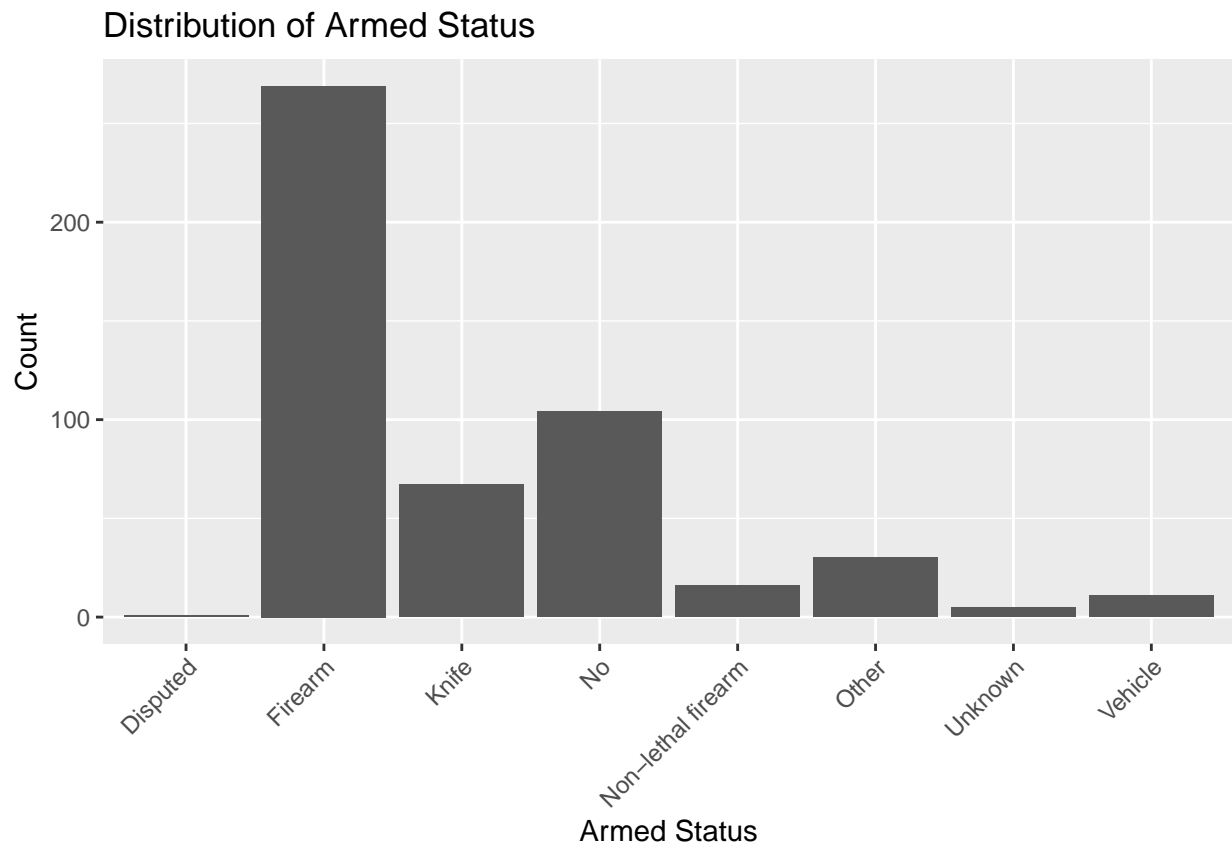


```
##
## Armed Status Distribution:
```

```
##
##      Disputed      Firearm      Knife      No
##           1         269         67       104
## Non-lethal firearm      Other      Unknown      Vehicle
##           16          30          5         11
```

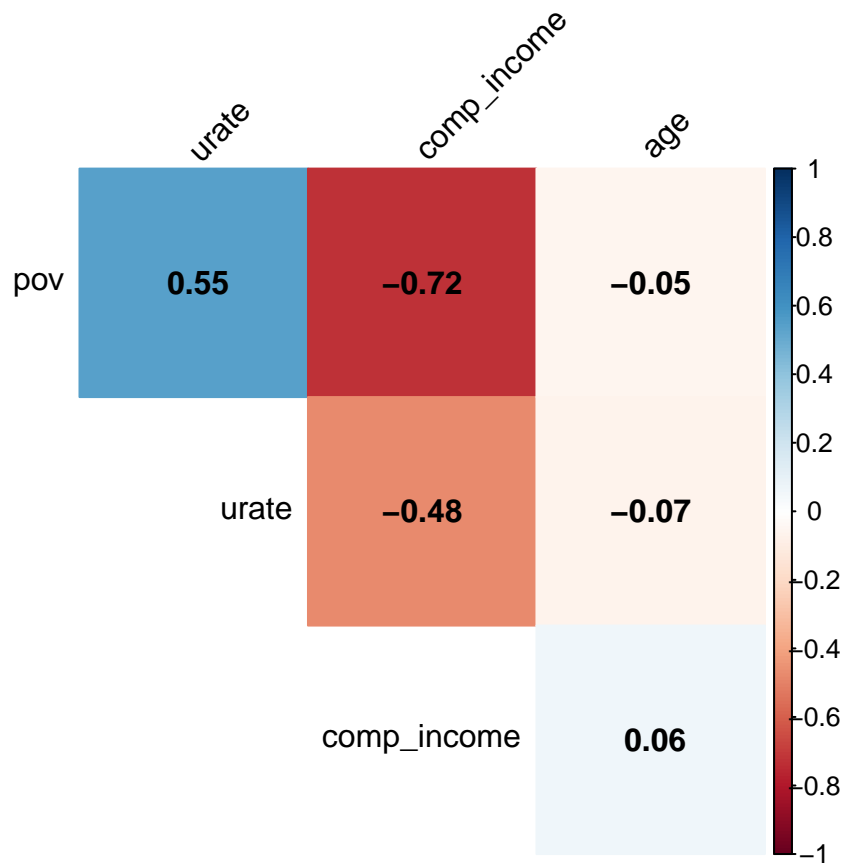
```
##
## Percentages:
```

```
##
##      Disputed      Firearm      Knife      No
##      0.1988072    53.4791252    13.3200795    20.6759443
## Non-lethal firearm      Other      Unknown      Vehicle
##      3.1809145     5.9642147     0.9940358     2.1868787
```



## corrplot 0.95 loaded





```
library(corrplot)
corrplot(cor_matrix,
  method = "color",
  type = "upper",
  order = "hclust",
  addCoef.col = "black",
  tl.col = "black",
  tl.srt = 45,
  diag = FALSE)
```

