Grocery Analysis

2024-11-22

## R Markdown

<https://github.com/fivethirtyeight/data/tree/master/police-killings>

library(readr)  
police\_data <- read\_csv("police\_killings.csv")

## Rows: 467 Columns: 34  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (17): name, age, gender, raceethnicity, month, streetaddress, city, stat...  
## dbl (17): day, year, latitude, longitude, state\_fp, county\_fp, tract\_ce, geo...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Check the structure of the dataset  
str(police\_data)

## spc\_tbl\_ [467 × 34] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ name : chr [1:467] "A'donte Washington" "Aaron Rutledge" "Aaron Siler" "Aaron Valdez" ...  
## $ age : chr [1:467] "16" "27" "26" "25" ...  
## $ gender : chr [1:467] "Male" "Male" "Male" "Male" ...  
## $ raceethnicity : chr [1:467] "Black" "White" "White" "Hispanic/Latino" ...  
## $ month : chr [1:467] "February" "April" "March" "March" ...  
## $ day : num [1:467] 23 2 14 11 19 7 27 26 28 7 ...  
## $ year : num [1:467] 2015 2015 2015 2015 2015 ...  
## $ streetaddress : chr [1:467] "Clearview Ln" "300 block Iris Park Dr" "22nd Ave and 56th St" "3000 Seminole Ave" ...  
## $ city : chr [1:467] "Millbrook" "Pineville" "Kenosha" "South Gate" ...  
## $ state : chr [1:467] "AL" "LA" "WI" "CA" ...  
## $ latitude : num [1:467] 32.5 31.3 42.6 33.9 41.1 ...  
## $ longitude : num [1:467] -86.4 -92.4 -87.8 -118.2 -81.4 ...  
## $ state\_fp : num [1:467] 1 22 55 6 39 4 6 6 48 26 ...  
## $ county\_fp : num [1:467] 51 79 59 37 153 13 29 37 41 81 ...  
## $ tract\_ce : num [1:467] 30902 11700 1200 535607 530800 ...  
## $ geo\_id : num [1:467] 1.05e+09 2.21e+10 5.51e+10 6.04e+09 3.92e+10 ...  
## $ county\_id : num [1:467] 1051 22079 55059 6037 39153 ...  
## $ namelsad : chr [1:467] "Census Tract 309.02" "Census Tract 117" "Census Tract 12" "Census Tract 5356.07" ...  
## $ lawenforcementagency: chr [1:467] "Millbrook Police Department" "Rapides Parish Sheriff's Office" "Kenosha Police Department" "South Gate Police Department" ...  
## $ cause : chr [1:467] "Gunshot" "Gunshot" "Gunshot" "Gunshot" ...  
## $ armed : chr [1:467] "No" "No" "No" "Firearm" ...  
## $ pop : num [1:467] 3779 2769 4079 4343 6809 ...  
## $ share\_white : chr [1:467] "60.5" "53.8" "73.8" "1.2" ...  
## $ share\_black : chr [1:467] "30.5" "36.2" "7.7" "0.6" ...  
## $ share\_hispanic : chr [1:467] "5.6" "0.5" "16.8" "98.8" ...  
## $ p\_income : chr [1:467] "28375" "14678" "25286" "17194" ...  
## $ h\_income : num [1:467] 51367 27972 45365 48295 68785 ...  
## $ county\_income : num [1:467] 54766 40930 54930 55909 49669 ...  
## $ comp\_income : num [1:467] 0.938 0.683 0.826 0.864 1.385 ...  
## $ county\_bucket : num [1:467] 3 2 2 3 5 1 4 4 2 3 ...  
## $ nat\_bucket : num [1:467] 3 1 3 3 4 1 4 4 1 2 ...  
## $ pov : chr [1:467] "14.1" "28.8" "14.6" "11.7" ...  
## $ urate : num [1:467] 0.0977 0.0657 0.1663 0.1248 0.0635 ...  
## $ college : num [1:467] 0.1685 0.1114 0.1473 0.0501 0.404 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. name = col\_character(),  
## .. age = col\_character(),  
## .. gender = col\_character(),  
## .. raceethnicity = col\_character(),  
## .. month = col\_character(),  
## .. day = col\_double(),  
## .. year = col\_double(),  
## .. streetaddress = col\_character(),  
## .. city = col\_character(),  
## .. state = col\_character(),  
## .. latitude = col\_double(),  
## .. longitude = col\_double(),  
## .. state\_fp = col\_double(),  
## .. county\_fp = col\_double(),  
## .. tract\_ce = col\_double(),  
## .. geo\_id = col\_double(),  
## .. county\_id = col\_double(),  
## .. namelsad = col\_character(),  
## .. lawenforcementagency = col\_character(),  
## .. cause = col\_character(),  
## .. armed = col\_character(),  
## .. pop = col\_double(),  
## .. share\_white = col\_character(),  
## .. share\_black = col\_character(),  
## .. share\_hispanic = col\_character(),  
## .. p\_income = col\_character(),  
## .. h\_income = col\_double(),  
## .. county\_income = col\_double(),  
## .. comp\_income = col\_double(),  
## .. county\_bucket = col\_double(),  
## .. nat\_bucket = col\_double(),  
## .. pov = col\_character(),  
## .. urate = col\_double(),  
## .. college = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

# Check for missing values  
summary(police\_data)

## name age gender raceethnicity   
## Length:467 Length:467 Length:467 Length:467   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## month day year streetaddress   
## Length:467 Min. : 1.00 Min. :2015 Length:467   
## Class :character 1st Qu.: 8.00 1st Qu.:2015 Class :character   
## Mode :character Median :16.00 Median :2015 Mode :character   
## Mean :15.83 Mean :2015   
## 3rd Qu.:23.00 3rd Qu.:2015   
## Max. :31.00 Max. :2015   
##   
## city state latitude longitude   
## Length:467 Length:467 Min. :19.92 Min. :-159.64   
## Class :character Class :character 1st Qu.:33.34 1st Qu.:-111.95   
## Mode :character Mode :character Median :35.77 Median : -94.76   
## Mean :36.40 Mean : -96.97   
## 3rd Qu.:39.94 3rd Qu.: -82.96   
## Max. :61.22 Max. : -68.10   
##   
## state\_fp county\_fp tract\_ce geo\_id   
## Min. : 1.00 Min. : 1.00 Min. : 100 Min. :1.003e+09   
## 1st Qu.: 8.00 1st Qu.: 29.00 1st Qu.: 5202 1st Qu.:8.022e+09   
## Median :24.00 Median : 63.00 Median : 40200 Median :2.403e+10   
## Mean :25.34 Mean : 91.58 Mean :236937 Mean :2.543e+10   
## 3rd Qu.:40.00 3rd Qu.:111.00 3rd Qu.:378450 3rd Qu.:4.011e+10   
## Max. :56.00 Max. :740.00 Max. :980000 Max. :5.601e+10   
##   
## county\_id namelsad lawenforcementagency cause   
## Min. : 1003 Length:467 Length:467 Length:467   
## 1st Qu.: 8022 Class :character Class :character Class :character   
## Median :24033 Mode :character Mode :character Mode :character   
## Mean :25434   
## 3rd Qu.:40112   
## Max. :56005   
##   
## armed pop share\_white share\_black   
## Length:467 Min. : 0 Length:467 Length:467   
## Class :character 1st Qu.: 3358 Class :character Class :character   
## Mode :character Median : 4447 Mode :character Mode :character   
## Mean : 4784   
## 3rd Qu.: 5816   
## Max. :26826   
##   
## share\_hispanic p\_income h\_income county\_income   
## Length:467 Length:467 Min. : 10290 Min. : 22545   
## Class :character Class :character 1st Qu.: 32625 1st Qu.: 43804   
## Mode :character Mode :character Median : 42759 Median : 50856   
## Mean : 46627 Mean : 52527   
## 3rd Qu.: 56190 3rd Qu.: 56832   
## Max. :142500 Max. :110292   
## NA's :2   
## comp\_income county\_bucket nat\_bucket pov   
## Min. :0.1840 Min. :1.000 Min. :1.000 Length:467   
## 1st Qu.:0.6454 1st Qu.:1.000 1st Qu.:1.000 Class :character   
## Median :0.8696 Median :2.000 Median :2.000 Mode :character   
## Mean :0.8959 Mean :2.498 Mean :2.497   
## 3rd Qu.:1.0815 3rd Qu.:4.000 3rd Qu.:3.000   
## Max. :2.8652 Max. :5.000 Max. :5.000   
## NA's :2 NA's :27 NA's :2   
## urate college   
## Min. :0.01133 Min. :0.01355   
## 1st Qu.:0.06859 1st Qu.:0.10617   
## Median :0.10518 Median :0.16954   
## Mean :0.11740 Mean :0.22022   
## 3rd Qu.:0.14083 3rd Qu.:0.28454   
## Max. :0.50761 Max. :0.82807   
## NA's :2 NA's :2

# Check the column names  
colnames(police\_data)

## [1] "name" "age" "gender"   
## [4] "raceethnicity" "month" "day"   
## [7] "year" "streetaddress" "city"   
## [10] "state" "latitude" "longitude"   
## [13] "state\_fp" "county\_fp" "tract\_ce"   
## [16] "geo\_id" "county\_id" "namelsad"   
## [19] "lawenforcementagency" "cause" "armed"   
## [22] "pop" "share\_white" "share\_black"   
## [25] "share\_hispanic" "p\_income" "h\_income"   
## [28] "county\_income" "comp\_income" "county\_bucket"   
## [31] "nat\_bucket" "pov" "urate"   
## [34] "college"

## Cleaning

# Convert columns to numeric where necessary  
police\_data$age <- as.numeric(police\_data$age)

## Warning: NAs introduced by coercion

police\_data$share\_white <- as.numeric(police\_data$share\_white)

## Warning: NAs introduced by coercion

police\_data$share\_black <- as.numeric(police\_data$share\_black)

## Warning: NAs introduced by coercion

police\_data$share\_hispanic <- as.numeric(police\_data$share\_hispanic)

## Warning: NAs introduced by coercion

police\_data$p\_income <- as.numeric(police\_data$p\_income)

## Warning: NAs introduced by coercion

police\_data$h\_income <- as.numeric(police\_data$h\_income)  
  
# Check for missing values again after conversion  
summary(police\_data)

## name age gender raceethnicity   
## Length:467 Min. :16.00 Length:467 Length:467   
## Class :character 1st Qu.:28.00 Class :character Class :character   
## Mode :character Median :35.00 Mode :character Mode :character   
## Mean :37.37   
## 3rd Qu.:45.00   
## Max. :87.00   
## NA's :4   
## month day year streetaddress   
## Length:467 Min. : 1.00 Min. :2015 Length:467   
## Class :character 1st Qu.: 8.00 1st Qu.:2015 Class :character   
## Mode :character Median :16.00 Median :2015 Mode :character   
## Mean :15.83 Mean :2015   
## 3rd Qu.:23.00 3rd Qu.:2015   
## Max. :31.00 Max. :2015   
##   
## city state latitude longitude   
## Length:467 Length:467 Min. :19.92 Min. :-159.64   
## Class :character Class :character 1st Qu.:33.34 1st Qu.:-111.95   
## Mode :character Mode :character Median :35.77 Median : -94.76   
## Mean :36.40 Mean : -96.97   
## 3rd Qu.:39.94 3rd Qu.: -82.96   
## Max. :61.22 Max. : -68.10   
##   
## state\_fp county\_fp tract\_ce geo\_id   
## Min. : 1.00 Min. : 1.00 Min. : 100 Min. :1.003e+09   
## 1st Qu.: 8.00 1st Qu.: 29.00 1st Qu.: 5202 1st Qu.:8.022e+09   
## Median :24.00 Median : 63.00 Median : 40200 Median :2.403e+10   
## Mean :25.34 Mean : 91.58 Mean :236937 Mean :2.543e+10   
## 3rd Qu.:40.00 3rd Qu.:111.00 3rd Qu.:378450 3rd Qu.:4.011e+10   
## Max. :56.00 Max. :740.00 Max. :980000 Max. :5.601e+10   
##   
## county\_id namelsad lawenforcementagency cause   
## Min. : 1003 Length:467 Length:467 Length:467   
## 1st Qu.: 8022 Class :character Class :character Class :character   
## Median :24033 Mode :character Mode :character Mode :character   
## Mean :25434   
## 3rd Qu.:40112   
## Max. :56005   
##   
## armed pop share\_white share\_black   
## Length:467 Min. : 0 Min. : 0.00 Min. : 0.00   
## Class :character 1st Qu.: 3358 1st Qu.:26.20 1st Qu.: 1.40   
## Mode :character Median : 4447 Median :56.50 Median : 7.40   
## Mean : 4784 Mean :51.92 Mean :17.94   
## 3rd Qu.: 5816 3rd Qu.:77.50 3rd Qu.:23.70   
## Max. :26826 Max. :99.60 Max. :99.80   
## NA's :2 NA's :2   
## share\_hispanic p\_income h\_income county\_income   
## Min. : 0.0 Min. : 5457 Min. : 10290 Min. : 22545   
## 1st Qu.: 3.5 1st Qu.:18257 1st Qu.: 32625 1st Qu.: 43804   
## Median :10.9 Median :22348 Median : 42759 Median : 50856   
## Mean :22.0 Mean :24309 Mean : 46627 Mean : 52527   
## 3rd Qu.:32.9 3rd Qu.:28556 3rd Qu.: 56190 3rd Qu.: 56832   
## Max. :98.8 Max. :86023 Max. :142500 Max. :110292   
## NA's :2 NA's :2 NA's :2   
## comp\_income county\_bucket nat\_bucket pov   
## Min. :0.1840 Min. :1.000 Min. :1.000 Length:467   
## 1st Qu.:0.6454 1st Qu.:1.000 1st Qu.:1.000 Class :character   
## Median :0.8696 Median :2.000 Median :2.000 Mode :character   
## Mean :0.8959 Mean :2.498 Mean :2.497   
## 3rd Qu.:1.0815 3rd Qu.:4.000 3rd Qu.:3.000   
## Max. :2.8652 Max. :5.000 Max. :5.000   
## NA's :2 NA's :27 NA's :2   
## urate college   
## Min. :0.01133 Min. :0.01355   
## 1st Qu.:0.06859 1st Qu.:0.10617   
## Median :0.10518 Median :0.16954   
## Mean :0.11740 Mean :0.22022   
## 3rd Qu.:0.14083 3rd Qu.:0.28454   
## Max. :0.50761 Max. :0.82807   
## NA's :2 NA's :2

# Handle missing values - For example, remove rows with missing income data  
police\_data\_clean <- police\_data[!is.na(police\_data$h\_income), ]

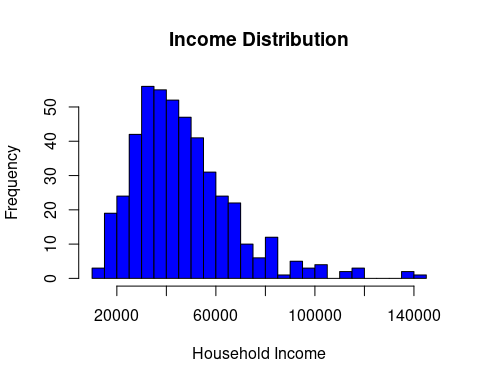
# Basic EDA: Check the distribution of raceethnicity  
table(police\_data\_clean$raceethnicity)

##   
## Asian/Pacific Islander Black Hispanic/Latino   
## 10 133 67   
## Native American Unknown White   
## 4 15 236

# Investigate cause of death  
table(police\_data\_clean$cause)

##   
## Death in custody Gunshot Struck by vehicle Taser   
## 14 409 12 27   
## Unknown   
## 3

# Explore the income distribution  
hist(police\_data\_clean$h\_income, main = "Income Distribution", xlab = "Household Income", col = "blue", breaks = 20)



# Research Questions

### 1. **Geographic Patterns and Police Killings** (Exploratory Analysis)

* **RQ4**: *Is there a regional variation in the number of police killings across different U.S. states?*
  + **Analysis Approach**:
    - **Descriptive Statistics**: Calculate the total number of police killings by state and plot the distribution.
    - **Visualizations**: Create bar charts or maps to visualize the frequency of police killings across states.
    - **Chi-Square Test**: To check if the distribution of police killings is independent of regions or states.
    - **Key Variables**: state, year, name (for counts), gender, raceethnicity.

### 2. **Socio-Economic Factors and Police Killings** (Regression Analysis)

* **RQ9**: *Is there a relationship between income levels (e.g., personal, household, or county income) and the rate of police killings in different areas?*
  + **Analysis Approach**:
    - **Regression Analysis**: Perform a linear regression to analyze the relationship between income and the rate of police killings in a region.
    - **Variables**:
      * **Independent Variables**: p\_income (personal income), h\_income (household income), county\_income (county income).
      * **Dependent Variable**: pop (population of the area or the number of police killings).
      * **Control Variables**: raceethnicity, state, armed.
    - **Model**: Use a multiple linear regression model to estimate how income levels (individual, household, or county) influence the frequency of police killings, controlling for other factors like race and whether the individual was armed.
    - **Hypothesis**: Higher income areas will have fewer police killings due to different levels of law enforcement resources, socioeconomic stability, or community-police relationships.

## Summary statistics

# Example: Assuming 'data' is a dataframe with a column 'state'  
# If you haven't defined 'data' yet, create a dummy dataframe for testing  
  
  
# List of left-wing (Democratic-leaning) and right-wing (Republican-leaning) states  
left\_wing\_states <- c("CA", "NY", "IL", "WA", "OR", "CO", "MA", "MI", "PA", "MN", "VA", "NJ", "MD", "CT", "RI", "VT", "HI")  
right\_wing\_states <- c("TX", "FL", "GA", "OH", "AZ", "NC", "SC", "WI", "MO", "IN", "KY", "LA", "AL", "OK", "TN", "MS", "AR")  
  
# Create a new variable 'state\_category' based on the political leaning  
police\_data\_clean$state\_category <- ifelse(police\_data\_clean$state %in% left\_wing\_states, "Left-Wing",   
 ifelse(police\_data\_clean$state %in% right\_wing\_states, "Right-Wing", "Unaffiliated"))

library(gtsummary)  
  
library(gtsummary)  
  
# Example: Summarizing police killings dataset with key variables (age, gender, raceethnicity, income)  
table2 <- police\_data\_clean |>   
 tbl\_summary(  
 include = c(age,gender, raceethnicity, p\_income, h\_income, county\_income), # Variables to summarize  
 by = state\_category, # Split the table by state  
 missing = "no" # Do not list missing data separately  
 ) |>   
 add\_n() |> # Add column with total number of non-missing observations  
 add\_p() |> # Test for a difference between groups (state in this case)  
 modify\_header(label = "\*\*Variable\*\*") |> # Update column header  
 bold\_labels() # Bold labels for clarity

## The following errors were returned during `modify\_header()`:  
## ✖ For variable `raceethnicity` (`state\_category`) and "estimate", "p.value",  
## "conf.low", and "conf.high" statistics: FEXACT error 7(location). LDSTP=18570  
## is too small for this problem, (pastp=12.6125, ipn\_0:=ipoin[itp=88]=439,  
## stp[ipn\_0]=16.1961). Increase workspace or consider using  
## 'simulate.p.value=TRUE'

# View the summary table  
table2

| **Variable** | **N** | **Left-Wing** N = 192*1* | **Right-Wing** N = 230*1* | **Unaffiliated** N = 43*1* | **p-value***2* |
| --- | --- | --- | --- | --- | --- |
| **age** | 461 | 35 (27, 45) | 35 (27, 46) | 35 (31, 44) | 0.6 |
| **gender** | 465 |  |  |  | 0.3 |
| Female |  | 9 (4.7%) | 9 (3.9%) | 4 (9.3%) |  |
| Male |  | 183 (95%) | 221 (96%) | 39 (91%) |  |
| **raceethnicity** | 465 |  |  |  |  |
| Asian/Pacific Islander |  | 8 (4.2%) | 2 (0.9%) | 0 (0%) |  |
| Black |  | 57 (30%) | 73 (32%) | 3 (7.0%) |  |
| Hispanic/Latino |  | 34 (18%) | 29 (13%) | 4 (9.3%) |  |
| Native American |  | 1 (0.5%) | 2 (0.9%) | 1 (2.3%) |  |
| Unknown |  | 9 (4.7%) | 6 (2.6%) | 0 (0%) |  |
| White |  | 83 (43%) | 118 (51%) | 35 (81%) |  |
| **p\_income** | 465 | 23,077 (18,471, 30,481) | 21,680 (17,696, 26,047) | 26,643 (20,296, 31,080) | 0.004 |
| **h\_income** | 465 | 48,123 (34,879, 62,312) | 39,111 (31,169, 51,367) | 41,875 (37,026, 58,600) | <0.001 |
| **county\_income** | 465 | 55,909 (49,563, 67,177) | 46,310 (41,835, 51,444) | 52,873 (46,566, 60,196) | <0.001 |
| *1*Median (Q1, Q3); n (%) | | | | | |
| *2*Kruskal-Wallis rank sum test; Fisher's exact test | | | | | |

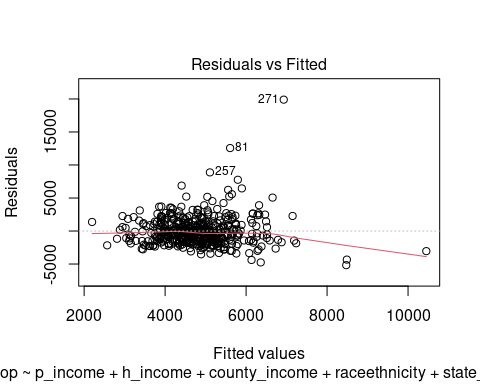
# Fit a full model with all relevant predictors  
full\_model <- lm(pop ~ p\_income + h\_income + county\_income + raceethnicity + armed + state\_category, data = police\_data\_clean)  
  
# Perform stepwise regression (both forward and backward)  
stepwise\_model <- step(full\_model, direction = "both")

## Start: AIC=7180.74  
## pop ~ p\_income + h\_income + county\_income + raceethnicity + armed +   
## state\_category  
##   
## Df Sum of Sq RSS AIC  
## - armed 7 25944783 2215748083 7172.2  
## <none> 2189803300 7180.7  
## - raceethnicity 5 48074912 2237878212 7180.8  
## - county\_income 1 13231470 2203034770 7181.5  
## - p\_income 1 26953103 2216756403 7184.4  
## - state\_category 2 56602274 2246405574 7188.6  
## - h\_income 1 134517284 2324320584 7206.5  
##   
## Step: AIC=7172.22  
## pop ~ p\_income + h\_income + county\_income + raceethnicity + state\_category  
##   
## Df Sum of Sq RSS AIC  
## <none> 2215748083 7172.2  
## - raceethnicity 5 52164502 2267912585 7173.0  
## - county\_income 1 13703118 2229451201 7173.1  
## - p\_income 1 26190355 2241938438 7175.7  
## - state\_category 2 51064026 2266812109 7178.8  
## + armed 7 25944783 2189803300 7180.7  
## - h\_income 1 128674964 2344423047 7196.5

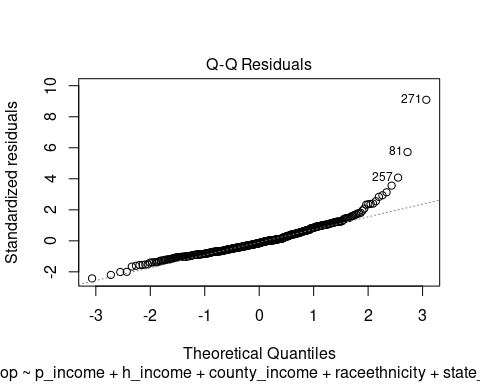
# Summary of the final model after stepwise selection  
summary(stepwise\_model)

##   
## Call:  
## lm(formula = pop ~ p\_income + h\_income + county\_income + raceethnicity +   
## state\_category, data = police\_data\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5166.8 -1421.3 -345.4 1005.0 19896.1   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.081e+03 9.344e+02 1.157 0.24772   
## p\_income -4.838e-02 2.089e-02 -2.317 0.02097 \*   
## h\_income 5.014e-02 9.765e-03 5.135 4.2e-07 \*\*\*  
## county\_income 1.730e-02 1.032e-02 1.676 0.09450 .   
## raceethnicityBlack 7.803e+02 7.302e+02 1.069 0.28579   
## raceethnicityHispanic/Latino 1.500e+03 7.541e+02 1.989 0.04733 \*   
## raceethnicityNative American 1.927e+03 1.318e+03 1.461 0.14459   
## raceethnicityUnknown 1.714e+03 9.091e+02 1.886 0.05994 .   
## raceethnicityWhite 1.352e+03 7.264e+02 1.861 0.06336 .   
## state\_categoryRight-Wing 7.886e+02 2.504e+02 3.149 0.00175 \*\*   
## state\_categoryUnaffiliated 7.029e+02 3.910e+02 1.798 0.07290 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2209 on 454 degrees of freedom  
## Multiple R-squared: 0.1417, Adjusted R-squared: 0.1228   
## F-statistic: 7.496 on 10 and 454 DF, p-value: 4.467e-11

# Residuals vs Fitted values plot  
plot(stepwise\_model, which = 1)



# Normal Q-Q plot for residuals  
plot(stepwise\_model, which = 2)



# Check for multicollinearity using VIF (Variance Inflation Factor)  
library(car)

## Loading required package: carData

vif(stepwise\_model)

## GVIF Df GVIF^(1/(2\*Df))  
## p\_income 3.375672 1 1.837300  
## h\_income 3.813772 1 1.952888  
## county\_income 1.705389 1 1.305905  
## raceethnicity 1.192872 5 1.017793  
## state\_category 1.441965 2 1.095819

# Check residuals for normality  
shapiro.test(resid(stepwise\_model))

##   
## Shapiro-Wilk normality test  
##   
## data: resid(stepwise\_model)  
## W = 0.8508, p-value < 2.2e-16