Black Communities, Crime, and Police

Aser Atawya

5/8/2022

Introduction

The BLM movement and the demand for equal justice for Black people have been growing exponentially since the tragic incident of George Floyd. Inspired by the BLM, I believe the scientific study and analysis of the data sets recording police interactions with the public and the crime rates in Black communities can support Black activism by highlighting the racial discrimination against Black people. This project aims to answer three important questions: What is the racial distribution of the victims and the perpetrators of homicide crimes in the U.S? What is the influence of race on the victims of police shootings? Is there a racial discrimination against Black people when it comes to the drivers stopped by police and the duration of their stop? To answer the aforementioned questions, I plan to analyze four data sets. The first data set has a record for all police shootings since 2000, including the race and gender of the victim. The second data set has a detailed record for every homicide case in the US from 1980 to 2014, including gender and race of both the victim and perpetrator. The third data set has a record of all police stops in Rhode Island from 2005 to 2015. The police stops are limited to Rhode Island only due to the size of the data; there are 200 million police stops recorded across the U.S, so analyzing such data sets is beyond the scope of this project. The fourth data set is a census of every county and state in the U.S by race and ethnicity.

The project found that there is a disproportionate ratio between the percentage of Black perpetrators of homicide crimes in the U.S and the percentage of Black people in the U.S population. Similarly, there is an even more skewed ratio between the percentage of Black victims of homicide crimes in the U.S and the proportion of Black people in the U.S population. Thus, the answer to the first question shows that Black communities are considerably vulnerable to crimes. The findings, also, showed that over one fifth of police shootings victims were Black even though the proportion of Black people in the U.S population is less than 15%. Furthermore, not only were the proportion of Black drivers stopped by the police in Rhode Island far more than the percentage of Black people in the state but also the Black drivers were stopped for longer duration compared to White drivers. The answers to the second and third questions underlines the unjust treatment of Black people by the police across the U.S.

Data Source

US Police Shootings: https://fatalencounters.org/ (https://fatalencounters.org/) The data was first crowdsourced and validated by comparing the data to public media reports. Then, the organization hired paid researchers to collect data from reliable sources, such as the LA Times Homicide report. Furthermore, the organization collected data directly from law enforcement agencies, requesting over 2300 public records of state, federal and local law enforcement agencies.

Homicide Reports, 1980-2014: https://www.kaggle.com/datasets/murderaccountability/homicide-reports (https://www.kaggle.com/datasets/murderaccountability/homicide-reports) The data was collected directly from the FBI master files of the Uniform Crime Report's "Return A", which has a summary of the homicide crimes in the U.S that ended with the arrest of the perpetrator, and the FBI's original Supplementary Homicide Report, which records the details of every homicide crime in the U.S since 1976.

Stanford Open Policing Project: https://www.kaggle.com/datasets/faressayah/stanford-open-policing-project (https://www.kaggle.com/datasets/faressayah/stanford-open-policing-project) The data was collected by Stanford Computational Policy Lab in collaboration with Big Local News by making public record requests of state and local police departments.

Population by Race and Ethnicity: https://www.statsamerica.org/downloads/default.aspx (https://www.statsamerica.org/downloads/default.aspx) The source of the data is the U.S Census Bureau, which collects data from federal, state, and local governments; in addition, some of the data is collected directly from citizens through surveys. Data sources include federal agencies, such as the IRS, and commercial businesses working on data collection. The data on race and ethnicity are based on self-identification.

Data Ethics

The analysis of the data brings light to the racial discrimination against Black communities in multiple areas, police interactions and homicide rates. The provision of accurate data regarding this matter is the basis of the intervention to achieve social efficiency, which is an ethically important gain. The analysis of the racial distribution of police shootings and police stops should impel police departments across the U.S to adopt better institutional policies to combat this racial bias, which is an ethically important gain. Highlighting the skewed ratio between the percentage of Black people in the population and the percentage of Black homicide victims and perpetrators develops our human understanding of the issue and lets us act more wisely to provide more safety to Black communities, which is another ethically important gain.

The data collected by the U.S Census Bureau, Stanford Open Policing Project, and the Murder Accountability Project avoids any harm to privacy and security because the data is anonymous in contrast to the data of police shootings, which has the full name of the victims. The data of the police shootings can be used to link the victims of police shootings to their family members using the full name of the victim; if the data was used in such a manner, the family members may be treated unfairly because they are linked to potentially dangerous criminals who were shot by the police, which is an ethically significant harm. Furthermore, this can lead the relatives of the victims to experience reputational or emotional harm, which are ethically significant harms.

Data Import, Cleaning and Tidying, and Exploration

```
#calls all the needed packages
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
## Warning: package 'tidyr' was built under R version 4.1.3
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## Warning: package 'stringr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.3
```

```
library(ggplot2)
library(modelr)
```

```
## Warning: package 'modelr' was built under R version 4.1.3
```

```
library("scatterplot3d")
options(na.action = na.warn)
```

- Data Import

```
#imports the four databases
police_shootings <- read_csv("Police_Shootings.csv")
```

```
## New names:
## * `` -> `...33`
## * `` -> `...34`
```

```
## Warning: One or more parsing issues, see `problems()` for details
```

```
## Rows: 31498 Columns: 36
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (27): Name, Gender, Race, Race with imputations, Imputation probability,...
## dbl  (7): Unique ID, Age, Latitude, Longitude, UID Temporary, Unique ID form...
## lgl  (2): ...33, ...34
##
## 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.
```

```
police_stops <- read_csv("Police_Stops.csv")
```

```
## Rows: 91741 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): driver_gender, driver_race, violation_raw, violation, search_type,...
## dbl  (2): driver_age_raw, driver_age
## lgl  (4): county_name, search_conducted, is_arrested, drugs_related_stop
## date (1): stop_date
## time (1): stop_time
##
## 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.
```

```
homicide <- read_csv("Homicide.csv")
```

```
## Rows: 638454 Columns: 24
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (18): Record ID, Agency Code, Agency Name, Agency Type, City, State, Mon...
## dbl  (6): Year, Incident, Victim Age, Perpetrator Age, Victim Count, Perpetr...
##
## 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.
```

```
US_race <- read_csv("Population by Race.csv")
```

```
## Rows: 95800 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): Statefips, Countyfips, Description
## dbl (11): IBRC_Geo_ID, Year, Total Population, White Alone, Black Alone, Ame...
##
## 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.
```

```
police_shootings
```

```
## # A tibble: 31,498 x 36
##    `Unique ID` Name           Age Gender Race  `Race with imp~` `Imputation pr~`
##          <dbl> <chr>        <dbl> <chr>  <chr> <chr>            <chr>
##  1       25746 Samuel H. K~    17 Male   Euro~ European-Americ~ Not imputed
##  2       25747 Mark A. Hor~    21 Male   Afri~ African-America~ Not imputed
##  3       25748 Phillip A. ~    19 Male   Afri~ African-America~ Not imputed
##  4       25749 Mark Ortiz      23 Male   Hisp~ Hispanic/Latino  Not imputed
##  5           1 LaTanya Jan~    24 Female Afri~ African-America~ Not imputed
##  6           2 Lester Mill~    53 Male   Race~ African-America~ 0.947676492
##  7       25750 Billy Jones     30 Male   Afri~ African-America~ Not imputed
##  8       25751 Terry Rouse     29 Male   Afri~ African-America~ Not imputed
##  9       28711 Jimmy Todd ~    25 Male   Race~ <NA>             <NA>
## 10       28712 Carl Gallis~    35 Male   Race~ <NA>             <NA>
## # ... with 31,488 more rows, and 29 more variables:
## #   `URL of image (PLS NO HOTLINKS)` <chr>,
## #   `Date of injury resulting in death (month/day/year)` <chr>,
## #   `Location of injury (address)` <chr>, `Location of death (city)` <chr>,
## #   State <chr>, `Location of death (zip code)` <chr>,
## #   `Location of death (county)` <chr>, `Full Address` <chr>, Latitude <dbl>,
## #   Longitude <dbl>, `Agency or agencies involved` <chr>, ...
```

```
police_stops
```

```
## # A tibble: 91,741 x 15
##    stop_date  stop_time county_name driver_gender driver_age_raw driver_age
##    <date>     <time>    <lgl>       <chr>                 <dbl>      <dbl>
##  1 2005-01-02 01:55     NA          M                      1985         20
##  2 2005-01-18 08:15     NA          M                      1965         40
##  3 2005-01-23 23:15     NA          M                      1972         33
##  4 2005-02-20 17:15     NA          M                      1986         19
##  5 2005-03-14 10:00     NA          F                      1984         21
##  6 2005-03-23 09:45     NA          M                      1982         23
##  7 2005-04-01 17:30     NA          M                      1969         36
##  8 2005-06-06 13:20     NA          F                      1986         19
##  9 2005-07-13 10:15     NA          M                      1970         35
## 10 2005-07-13 15:45     NA          M                      1970         35
## # ... with 91,731 more rows, and 9 more variables: driver_race <chr>,
## #   violation_raw <chr>, violation <chr>, search_conducted <lgl>,
## #   search_type <chr>, stop_outcome <chr>, is_arrested <lgl>,
## #   stop_duration <chr>, drugs_related_stop <lgl>
```

```
homicide
```

```
## # A tibble: 638,454 x 24
##    `Record ID` `Agency Code` `Agency Name` `Agency Type` City  State  Year Month
##    <chr>       <chr>         <chr>         <chr>         <chr> <chr> <dbl> <chr>
##  1 000001      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 Janu~
##  2 000002      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 March
##  3 000003      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 March
##  4 000004      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 April
##  5 000005      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 April
##  6 000006      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 May
##  7 000007      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 May
##  8 000008      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 June
##  9 000009      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 June
## 10 000010      AK00101       Anchorage     Municipal Po~ Anch~ Alas~  1980 June
## # ... with 638,444 more rows, and 16 more variables: Incident <dbl>,
## #   `Crime Type` <chr>, `Crime Solved` <chr>, `Victim Sex` <chr>,
## #   `Victim Age` <dbl>, `Victim Race` <chr>, `Victim Ethnicity` <chr>,
## #   `Perpetrator Sex` <chr>, `Perpetrator Age` <dbl>, `Perpetrator Race` <chr>,
## #   `Perpetrator Ethnicity` <chr>, Relationship <chr>, Weapon <chr>,
## #   `Victim Count` <dbl>, `Perpetrator Count` <dbl>, `Record Source` <chr>
```

```
US_race
```

```
## # A tibble: 95,800 x 14
##    IBRC_Geo_ID Statefips Countyfips Description  Year `Total Population`
##          <dbl> <chr>     <chr>      <chr>       <dbl>              <dbl>
##  1           0 0         000        U.S.         1990          249622814
##  2           0 0         000        U.S.         1991          252980941
##  3           0 0         000        U.S.         1992          256514224
##  4           0 0         000        U.S.         1993          259918588
##  5           0 0         000        U.S.         1994          263125821
##  6           0 0         000        U.S.         1995          266278393
##  7           0 0         000        U.S.         1996          269394284
##  8           0 0         000        U.S.         1997          272646925
##  9           0 0         000        U.S.         1998          275854104
## 10           0 0         000        U.S.         1999          279040168
## # ... with 95,790 more rows, and 8 more variables: `White Alone` <dbl>,
## #   `Black Alone` <dbl>, `American Indian or Alaskan Native` <dbl>,
## #   `Asian Alone` <dbl>, `Hawaiian or Pacific Islander Alone` <dbl>,
## #   `Two or More Races` <dbl>, `Not Hispanic` <dbl>, Hispanic <dbl>
```

- Data Tidying, Cleaning, and Transformation

```
#calculates the percentage of each race in the U.S population
USA_Race <- US_race %>% filter(Description == "U.S.") %>% mutate(
  Black_Proportion  = `Black Alone`*100/ `Total Population`, White_Proportion = `White Alone`*100/ `Tota
l Population`,
  Asian_Proportion = `Asian Alone`*100/ `Total Population`, Hispanic_Proportion = Hispanic*100/ `Total P
opulation`,
  Native_Proportion = `American Indian or Alaskan Native`  *100/ `Total Population`,
  Hawaiian_Proportion = `Hawaiian or Pacific Islander Alone` *100/ `Total Population`
)
USA_Race
```

```
## # A tibble: 30 x 20
##    IBRC_Geo_ID Statefips Countyfips Description  Year `Total Population`
##          <dbl> <chr>     <chr>      <chr>       <dbl>              <dbl>
##  1           0 0         000        U.S.         1990          249622814
##  2           0 0         000        U.S.         1991          252980941
##  3           0 0         000        U.S.         1992          256514224
##  4           0 0         000        U.S.         1993          259918588
##  5           0 0         000        U.S.         1994          263125821
##  6           0 0         000        U.S.         1995          266278393
##  7           0 0         000        U.S.         1996          269394284
##  8           0 0         000        U.S.         1997          272646925
##  9           0 0         000        U.S.         1998          275854104
## 10           0 0         000        U.S.         1999          279040168
## # ... with 20 more rows, and 14 more variables: `White Alone` <dbl>,
## #   `Black Alone` <dbl>, `American Indian or Alaskan Native` <dbl>,
## #   `Asian Alone` <dbl>, `Hawaiian or Pacific Islander Alone` <dbl>,
## #   `Two or More Races` <dbl>, `Not Hispanic` <dbl>, Hispanic <dbl>,
## #   Black_Proportion <dbl>, White_Proportion <dbl>, Asian_Proportion <dbl>,
## #   Hispanic_Proportion <dbl>, Native_Proportion <dbl>,
## #   Hawaiian_Proportion <dbl>
```

```
#checks the values in the Race and Race with imputations columns
police_shootings %>% distinct(Race)
```

```
## # A tibble: 12 x 1
##    Race
##    <chr>
##  1 European-American/White
##  2 African-American/Black
##  3 Hispanic/Latino
##  4 Race unspecified
##  5 Native American/Alaskan
##  6 Asian/Pacific Islander
##  7 Middle Eastern
##  8 european-American/White
##  9 African-American/Black African-American/Black Not imputed
## 10 European-American/European-American/White
## 11 Christopher Anthony Alexander
## 12 <NA>
```

```
police_shootings %>% distinct(`Race with imputations`)
```

```
## # A tibble: 10 x 1
##    `Race with imputations`
##    <chr>
##  1 European-American/White
##  2 African-American/Black
##  3 Hispanic/Latino
##  4 <NA>
##  5 Race unspecified
##  6 Asian/Pacific Islander
##  7 Native American/Alaskan
##  8 Middle Eastern
##  9 HIspanic/Latino
## 10 european-American/White
```

```
#removes the different spellings of the same race
police_shootings_clean <- mutate(police_shootings,
Race = ifelse(Race %in% c("European-American/White", "european-American/White", "European-American/Europ
ean-American/White"), "White", Race),
Race = ifelse(Race %in% c("African-American/Black", "African-American/Black African-American/Black Not i
mputed"), "Black", Race),

#if imputation probability is more than 50% and the race is unspecified, assign the imputed race
Race = ifelse( Race == "Race unspecified" & `Race with imputations` %in% c("African-American/Black", "Af
rican-American/Black African-American/Black Not imputed") & `Imputation probability` > 0.5, "Black", Rac
e),
Race = ifelse( Race == "Race unspecified" & `Race with imputations` %in% c("European-American/White", "e
uropean-American/White", "European-American/European-American/White") & `Imputation probability` > 0.5,
"White", Race),
) %>%

#removes one row with invalid Race Value
filter( Race != "Christopher Anthony Alexander", !is.na(Race))

police_shootings_clean
```

```
## # A tibble: 31,496 x 36
##    `Unique ID` Name           Age Gender Race  `Race with imp~` `Imputation pr~`
##          <dbl> <chr>        <dbl> <chr>  <chr> <chr>            <chr>
##  1       25746 Samuel H. K~    17 Male   White European-Americ~ Not imputed
##  2       25747 Mark A. Hor~    21 Male   Black African-America~ Not imputed
##  3       25748 Phillip A. ~    19 Male   Black African-America~ Not imputed
##  4       25749 Mark Ortiz      23 Male   Hisp~ Hispanic/Latino  Not imputed
##  5           1 LaTanya Jan~    24 Female Black African-America~ Not imputed
##  6           2 Lester Mill~    53 Male   Black African-America~ 0.947676492
##  7       25750 Billy Jones     30 Male   Black African-America~ Not imputed
##  8       25751 Terry Rouse     29 Male   Black African-America~ Not imputed
##  9       28711 Jimmy Todd ~    25 Male   Race~ <NA>             <NA>
## 10       28712 Carl Gallis~    35 Male   Race~ <NA>             <NA>
## # ... with 31,486 more rows, and 29 more variables:
## #   `URL of image (PLS NO HOTLINKS)` <chr>,
## #   `Date of injury resulting in death (month/day/year)` <chr>,
## #   `Location of injury (address)` <chr>, `Location of death (city)` <chr>,
## #   State <chr>, `Location of death (zip code)` <chr>,
## #   `Location of death (county)` <chr>, `Full Address` <chr>, Latitude <dbl>,
## #   Longitude <dbl>, `Agency or agencies involved` <chr>, ...
```

```
#parses date and chooses only the relevant columns
police_shootings_clean_2 <- police_shootings_clean %>% mutate(Date = parse_date(`Date of injury resultin
g in death (month/day/year)`,"%m/%d/%Y"), Year = str_sub(Date, 1, 4)) %>% select(Year, Gender, Race, `Ar
med/Unarmed`)

#calculates the total police shootings each year
police_shootings_by_year <- police_shootings_clean_2 %>% group_by(Year)  %>% summarize(n())
#calculates the police shootings by race each year
police_shootings_by_race <- police_shootings_clean_2 %>% group_by(Year)  %>%  count(Race) %>% pivot_wide
r(names_from = Race, values_from = n)

#joins the two tibbles and calculates the racial distribution of the people shot by police
police_shootings_by_race_and_year <-  full_join(police_shootings_by_year, police_shootings_by_race, by =
"Year") %>% rename( Total = `n()`   ) %>% mutate(Year = parse_double(Year), Asian_Percentage = `Asian/Pa
cific Islander`*100/ Total, Black_Percentage = Black*100/ Total, Hispanic_Percentage = `Hispanic/Latino`
*100/ Total, Native_Percentage =  `Native American/Alaskan`*100/ Total, Unspecified_Percentage =`Race un
specified`*100/ Total, White_Percentage = White*100/ Total, Middle_Eastern_Percentage =  `Middle Eastern
`*100/ Total)

#joins the tibble with the racial distribution of the people shot by police and the tibble with the raci
al distribution of the U.S population
police_shootings_And_US_Race <- inner_join(police_shootings_by_race_and_year, USA_Race, by = "Year")
police_shootings_And_US_Race
```

```
## # A tibble: 20 x 35
##     Year Total `Asian/Pacific Islander` Black `Hispanic/Latino` `Native Americ~`
##    <dbl> <int>                    <int> <int>             <int>            <int>
##  1  2000   865                       12   241                57               13
##  2  2001   923                       14   216                75                3
##  3  2002   987                       12   242                87                6
##  4  2003  1062                       17   318               115               11
##  5  2004  1038                       16   276               141                5
##  6  2005  1170                       21   307               146               13
##  7  2006  1270                       18   370               163               13
##  8  2007  1259                       13   357               165                7
##  9  2008  1216                       13   360               163                9
## 10  2009  1262                       21   353               170                9
## 11  2010  1298                       18   363               156               13
## 12  2011  1408                       21   420               194                9
## 13  2012  1483                       22   446               213               17
## 14  2013  1778                       32   515               313               11
## 15  2014  1711                       30   473               302               19
## 16  2015  1611                       31   433               249               24
## 17  2016  1609                       26   437               216               30
## 18  2017  1772                       20   473               273               33
## 19  2018  1864                       33   449               278               29
## 20  2019  1828                       40   494               249               17
## # ... with 29 more variables: `Race unspecified` <int>, White <int>,
## #   `Middle Eastern` <int>, Asian_Percentage <dbl>, Black_Percentage <dbl>,
## #   Hispanic_Percentage <dbl>, Native_Percentage <dbl>,
## #   Unspecified_Percentage <dbl>, White_Percentage <dbl>,
## #   Middle_Eastern_Percentage <dbl>, IBRC_Geo_ID <dbl>, Statefips <chr>,
## #   Countyfips <chr>, Description <chr>, `Total Population` <dbl>,
## #   `White Alone` <dbl>, `Black Alone` <dbl>, ...
```

```
#calculates the total homicide crimes each year
Homicide_by_year <- homicide %>% group_by(Year)  %>% summarize(n())

#calculates the total homicide perpetrators each year by race
Homicide_by_perpetrator_race <- homicide%>% group_by(Year)  %>%  count(`Perpetrator Race`) %>% pivot_wid
er(names_from = `Perpetrator Race`, values_from = n)

#calculates the total homicide victims each year by race and rename the columns to avoid overlapping wit
h the tibble Homicide_by_perpetrator_race
Homicide_by_victim_race <- homicide%>% group_by(Year)  %>%  count(`Victim Race`) %>% pivot_wider(names_f
rom = `Victim Race`, values_from = n) %>% rename(  `Asian/Pacific Islander_victim` = `Asian/Pacific Isla
nder`, Black_victim = Black  , `Native American/Alaska Native_victim` = `Native American/Alaska Native`
 ,  Unknown_victim = Unknown, White_victim = White)

#joins all the three tibbles and calculates the racial distribution of homicide victims and perpetrator
Homicide_by_race_and_year <-  full_join(Homicide_by_year, Homicide_by_perpetrator_race , by = "Year") %
>%  full_join(Homicide_by_victim_race , by = "Year") %>% rename( Total = `n()`   ) %>% mutate( Asian_Per
centage = `Asian/Pacific Islander`*100/ Total, Black_Percentage = Black*100/ Total,  Native_Percentage =
`Native American/Alaska Native`*100/ Total, Unspecified_Percentage =`Unknown`*100/ Total, White_Percenta
ge = White*100/ Total, Asian_Percentage_victim = `Asian/Pacific Islander_victim`*100/ Total, Black_Perce
ntage_victim = Black_victim*100/ Total,  Native_Percentage_victim =  `Native American/Alaska Native_vict
im`*100/ Total, Unspecified_Percentage_victim =`Unknown_victim`*100/ Total, White_Percentage_victim = Wh
ite_victim *100/ Total)

#joins the tibble with the racial distribution of homicide and the tibble with the racial distribution o
f the U.S population
Homicide_And_US_Race <- inner_join(Homicide_by_race_and_year, USA_Race, by = "Year")
Homicide_And_US_Race
```

```
## # A tibble: 25 x 41
##     Year Total `Asian/Pacific Islander` Black `Native American/A~` Unknown White
##    <dbl> <int>                    <int> <int>                <int>   <int> <int>
##  1  1990 21246                      138  7383                  105    6464  7156
##  2  1991 22657                      209  7906                  106    7485  6951
##  3  1992 23793                      244  8178                  110    8135  7126
##  4  1993 24335                      243  8618                  111    8079  7284
##  5  1994 23246                      239  8221                  114    7549  7123
##  6  1995 21179                      251  6997                  127    7028  6776
##  7  1996 18969                      214  6410                   94    6017  6234
##  8  1997 17794                      208  5960                   99    5811  5716
##  9  1998 16061                      148  5213                   93    4913  5694
## 10  1999 14621                      188  4714                  119    4522  5078
## # ... with 15 more rows, and 34 more variables:
## #   `Asian/Pacific Islander_victim` <int>, Black_victim <int>,
## #   `Native American/Alaska Native_victim` <int>, Unknown_victim <int>,
## #   White_victim <int>, Asian_Percentage <dbl>, Black_Percentage <dbl>,
## #   Native_Percentage <dbl>, Unspecified_Percentage <dbl>,
## #   White_Percentage <dbl>, Asian_Percentage_victim <dbl>,
## #   Black_Percentage_victim <dbl>, Native_Percentage_victim <dbl>, ...
```

```
#calculates the percentage of each race in the population of Rhode Island
Rhode_Island_Race <- US_race %>% filter (Description == "Rhode Island") %>% mutate(
  Black_Proportion = `Black Alone`*100/ `Total Population`, White_Proportion = `White Alone`*100/ `Total
Population`,
  Asian_Proportion = `Asian Alone`*100/ `Total Population`, Hispanic_Proportion = Hispanic*100/ `Total P
opulation`,
  Native_Proportion = `American Indian or Alaskan Native`  *100/ `Total Population`,
  Hawaiian_Proportion = `Hawaiian or Pacific Islander Alone` *100/ `Total Population`
) %>%
  # rename Hispanic to avoid the automatic renaming when joining this tibble in the next code chunk
  rename(Hispanic_race = Hispanic)
Rhode_Island_Race
```

```
## # A tibble: 30 x 20
##    IBRC_Geo_ID Statefips Countyfips Description   Year `Total Population`
##          <dbl> <chr>     <chr>      <chr>        <dbl>              <dbl>
##  1       44000 44        000        Rhode Island  1990            1005995
##  2       44000 44        000        Rhode Island  1991            1010649
##  3       44000 44        000        Rhode Island  1992            1012581
##  4       44000 44        000        Rhode Island  1993            1015113
##  5       44000 44        000        Rhode Island  1994            1015960
##  6       44000 44        000        Rhode Island  1995            1017002
##  7       44000 44        000        Rhode Island  1996            1020893
##  8       44000 44        000        Rhode Island  1997            1025353
##  9       44000 44        000        Rhode Island  1998            1031155
## 10       44000 44        000        Rhode Island  1999            1040402
## # ... with 20 more rows, and 14 more variables: `White Alone` <dbl>,
## #   `Black Alone` <dbl>, `American Indian or Alaskan Native` <dbl>,
## #   `Asian Alone` <dbl>, `Hawaiian or Pacific Islander Alone` <dbl>,
## #   `Two or More Races` <dbl>, `Not Hispanic` <dbl>, Hispanic_race <dbl>,
## #   Black_Proportion <dbl>, White_Proportion <dbl>, Asian_Proportion <dbl>,
## #   Hispanic_Proportion <dbl>, Native_Proportion <dbl>,
## #   Hawaiian_Proportion <dbl>
```

```
#creates a Year column
police_stops_2 <- police_stops %>% mutate(Year = str_sub(stop_date, 1, 4))

#calculates the total police stops each year in Rhode Island
police_stops_2_by_year <-police_stops_2 %>% group_by(Year)  %>% summarize(n())

#calculates the police stops by race each year in Rhode Island
police_stops_2_by_race <- police_stops_2 %>% group_by(Year)  %>%  count(driver_race) %>% pivot_wider(nam
es_from = driver_race, values_from = n)

#joins the two tibbles and calculates the racial distribution of the people stopped by police in Rhode I
sland
police_stops_by_race_and_year <-  full_join(police_stops_2_by_race, police_stops_2_by_year, by = "Year")
%>% rename( Total = `n()`   ) %>% mutate(Year = parse_double(Year), Asian_Percentage = Asian*100/ Total,
Black_Percentage = Black*100/ Total, Hispanic_Percentage = Hispanic *100/ Total,  White_Percentage = Whi
te*100/ Total)

#joins the tibble with the racial distribution of the people stopped by the police in Rhode Island and t
he tibble with the racial distribution of Rhode Island population
police_stops_And_Rhode_Island_Race <- inner_join(police_stops_by_race_and_year, Rhode_Island_Race, by =
"Year")
police_stops_And_Rhode_Island_Race
```

```
## # A tibble: 11 x 31
## # Groups:   Year [11]
##     Year Asian Black Hispanic Other White  `NA` Total Asian_Percentage
##    <dbl> <int> <int>    <int> <int> <int> <int> <int>            <dbl>
##  1  2005    85   298       56     9  2057    53  2558             3.32
##  2  2006   278  1212      199    24  8429   497 10639             2.61
##  3  2007   241  1014      682    25  6943   571  9476             2.54
##  4  2008   227  1099      853    23  5949   601  8752             2.59
##  5  2009   215  1126      858    26  5012   671  7908             2.72
##  6  2010   162  1034      812    29  4958   566  7561             2.14
##  7  2011   207  1101      856    22  5389   551  8126             2.55
##  8  2012   237  1448     1343    54  7314   574 10970             2.16
##  9  2013   163  1175     1091     9  4983   503  7924             2.06
## 10  2014   241  1380     1322    12  5894   379  9228             2.61
## 11  2015   203  1357     1435     7  5230   367  8599             2.36
## # ... with 22 more variables: Black_Percentage <dbl>,
## #   Hispanic_Percentage <dbl>, White_Percentage <dbl>, IBRC_Geo_ID <dbl>,
## #   Statefips <chr>, Countyfips <chr>, Description <chr>,
## #   `Total Population` <dbl>, `White Alone` <dbl>, `Black Alone` <dbl>,
## #   `American Indian or Alaskan Native` <dbl>, `Asian Alone` <dbl>,
## #   `Hawaiian or Pacific Islander Alone` <dbl>, `Two or More Races` <dbl>,
## #   `Not Hispanic` <dbl>, Hispanic_race <dbl>, Black_Proportion <dbl>, ...
```

```
#checks all the values of the stop_duration column
police_stops_2 %>% distinct(stop_duration)
```

```
## # A tibble: 6 x 1
##   stop_duration
##   <chr>
## 1 0-15 Min
## 2 16-30 Min
## 3 30+ Min
## 4 <NA>
## 5 2
## 6 1
```

```
#groups the data by stop_duration and counts the people stopped by their race and avoids the irrelevant
 values of stop_duration
police_stops_duration_by_race <- police_stops_2 %>% filter(stop_duration != 2 & stop_duration != 1) %>%
 group_by(stop_duration)  %>%  count(driver_race) %>% pivot_wider(names_from = driver_race, values_from
 = n)
police_stops_duration_by_race
```

```
## # A tibble: 3 x 6
## # Groups:   stop_duration [3]
##   stop_duration Asian Black Hispanic Other White
##   <chr>         <int> <int>    <int> <int> <int>
## 1 0-15 Min       1899  9500     7441   201 50502
## 2 16-30 Min       291  2102     1569    34  9639
## 3 30+ Min          69   641      497     5  2016
```

```
#groups the data by race and counts the people stopped by their stop_duration and avoids the irrelevant
 values of stop_duration
police_stops_duration_by_race2 <- police_stops_2 %>% filter(stop_duration != 2 & stop_duration != 1) %>%
group_by(driver_race)  %>%  count(stop_duration) %>% pivot_wider(names_from = stop_duration, values_from
= n)
police_stops_duration_by_race2
```

```
## # A tibble: 5 x 4
## # Groups:   driver_race [5]
##   driver_race `0-15 Min` `16-30 Min` `30+ Min`
##   <chr>            <int>       <int>     <int>
## 1 Asian             1899         291        69
## 2 Black             9500        2102       641
## 3 Hispanic          7441        1569       497
## 4 Other              201          34         5
## 5 White            50502        9639      2016
```

```
#removes the irrelevant values in the stop_duration column
police_stops_filtered <- police_stops_2 %>% filter(stop_duration =="0-15 Min" | stop_duration =="16-30 M
in" | stop_duration =="30+ Min")
police_stops_filtered
```

```
## # A tibble: 86,406 x 16
##    stop_date  stop_time county_name driver_gender driver_age_raw driver_age
##    <date>     <time>    <lgl>       <chr>                 <dbl>      <dbl>
##  1 2005-01-02 01:55     NA          M                      1985         20
##  2 2005-01-18 08:15     NA          M                      1965         40
##  3 2005-01-23 23:15     NA          M                      1972         33
##  4 2005-02-20 17:15     NA          M                      1986         19
##  5 2005-03-14 10:00     NA          F                      1984         21
##  6 2005-03-23 09:45     NA          M                      1982         23
##  7 2005-04-01 17:30     NA          M                      1969         36
##  8 2005-06-06 13:20     NA          F                      1986         19
##  9 2005-07-13 10:15     NA          M                      1970         35
## 10 2005-07-13 15:45     NA          M                      1970         35
## # ... with 86,396 more rows, and 10 more variables: driver_race <chr>,
## #   violation_raw <chr>, violation <chr>, search_conducted <lgl>,
## #   search_type <chr>, stop_outcome <chr>, is_arrested <lgl>,
## #   stop_duration <chr>, drugs_related_stop <lgl>, Year <chr>
```

```
#checks the values of the Description column of US_race
US_race %>% distinct(Description)
```

```
## # A tibble: 3,199 x 1
##    Description
##    <chr>
##  1 U.S.
##  2 Alabama
##  3 Autauga County, AL
##  4 Baldwin County, AL
##  5 Barbour County, AL
##  6 Bibb County, AL
##  7 Blount County, AL
##  8 Bullock County, AL
##  9 Butler County, AL
## 10 Calhoun County, AL
## # ... with 3,189 more rows
```

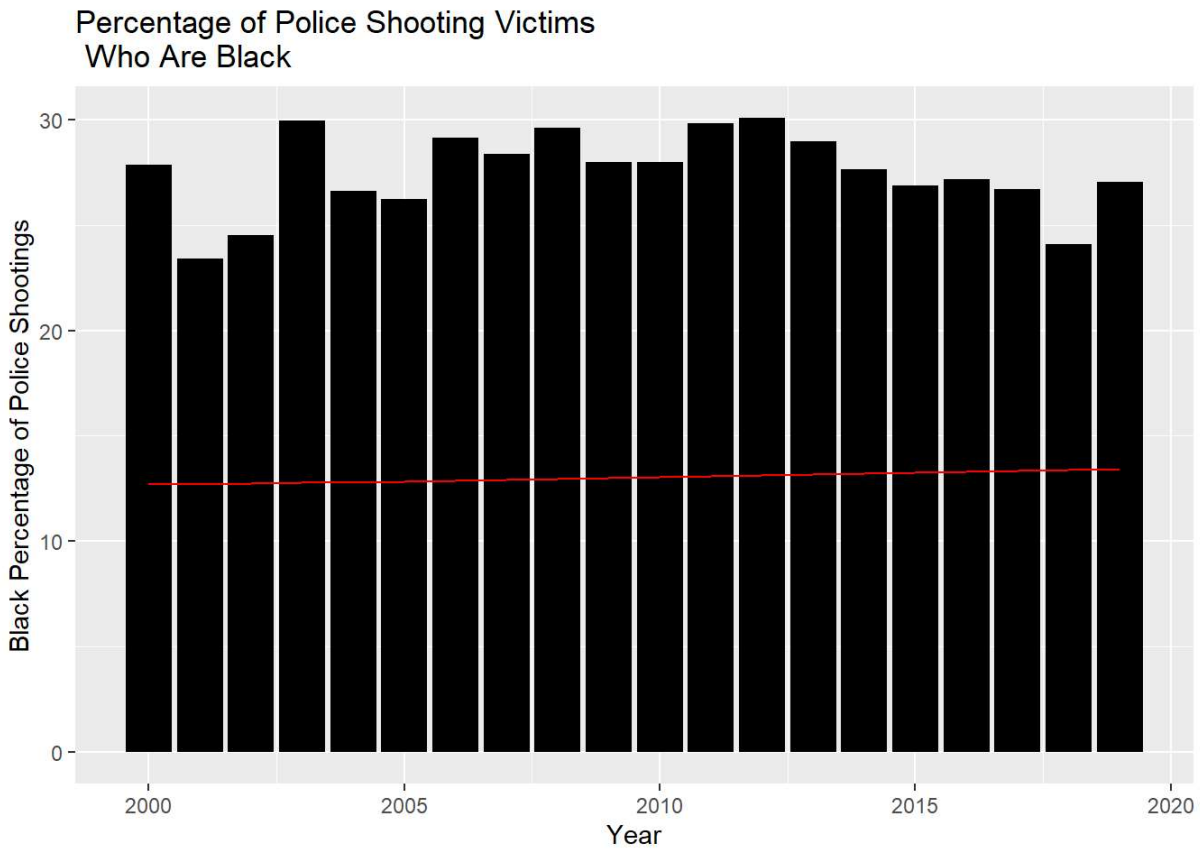
```
#finds the U.S states where Black people are a majority
#the data was filtered to keep only the 50 states and D.C
US_race_filtered <- US_race %>% filter(!str_detect(Description, ","), Description != "U.S.", Description
!= "Puerto Rico")
US_race_filtered %>% mutate(majority = pmax(US_race_filtered $`White Alone`, US_race_filtered $`Black Al
one`, US_race_filtered $`American Indian or Alaskan Native`, US_race_filtered $`Asian Alone`, US_race_fi
ltered $`Hawaiian or Pacific Islander Alone`, US_race_filtered $`Two or More Races`, na.rm = TRUE), blac
k_majority = ifelse(majority == US_race_filtered$`Black Alone`, 1, 0)) %>% select(majority, black_majori
ty, everything()) %>% filter(black_majority == 1) %>% distinct(Description)
```

```
## # A tibble: 1 x 1
##   Description
##   <chr>
## 1 District of Columbia
```

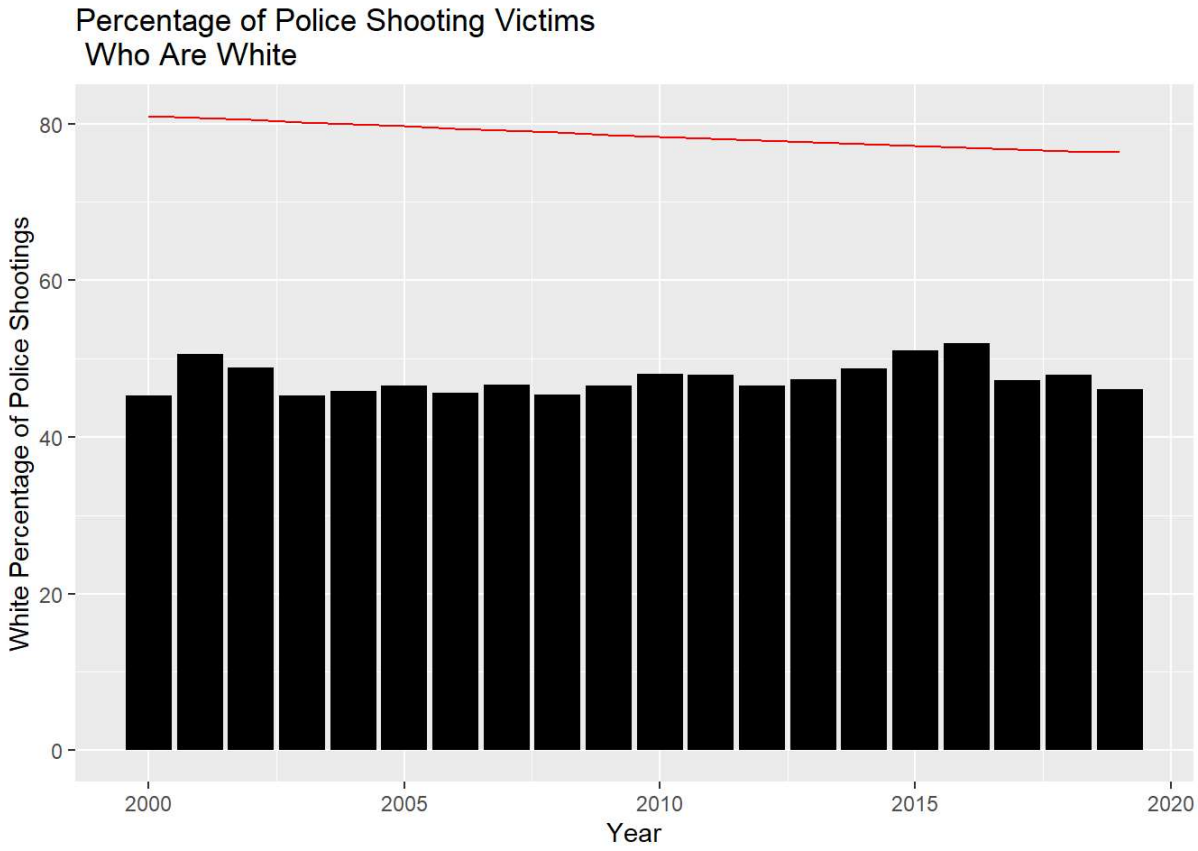
None of the 50 U.S states has a Black majority. Only the District of Columbia has a black majority.

- Data Visualization

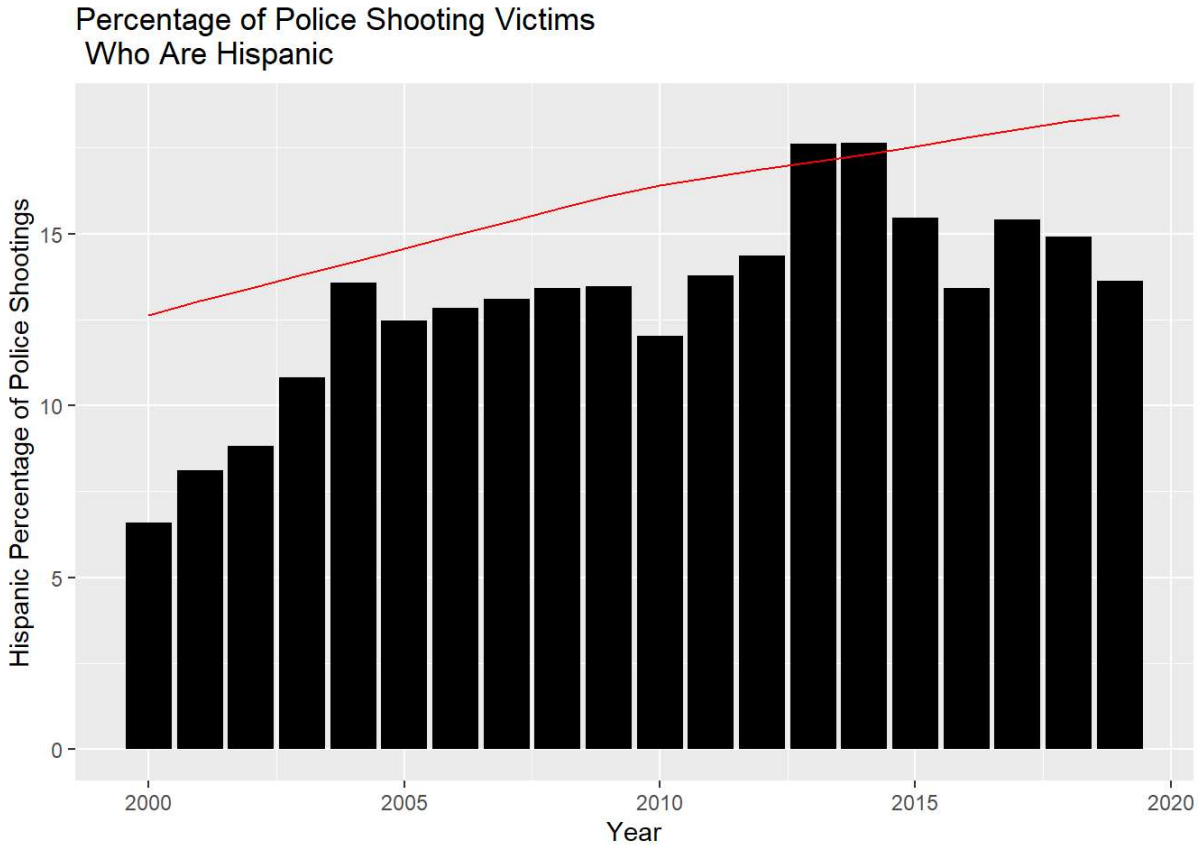
```
#column plot of the percentage of police shooting victims who are Black with a red line showing the perc
entage of Black people in the population

ggplot(data = police_shootings_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Black_Percentage), f
ill = "Black") + labs(title="Percentage of Police Shooting Victims \n Who Are Black", x = "Year", y = "B
lack Percentage of Police Shootings") + geom_line(aes(x = Year, y = Black_Proportion), color = "Red")
```



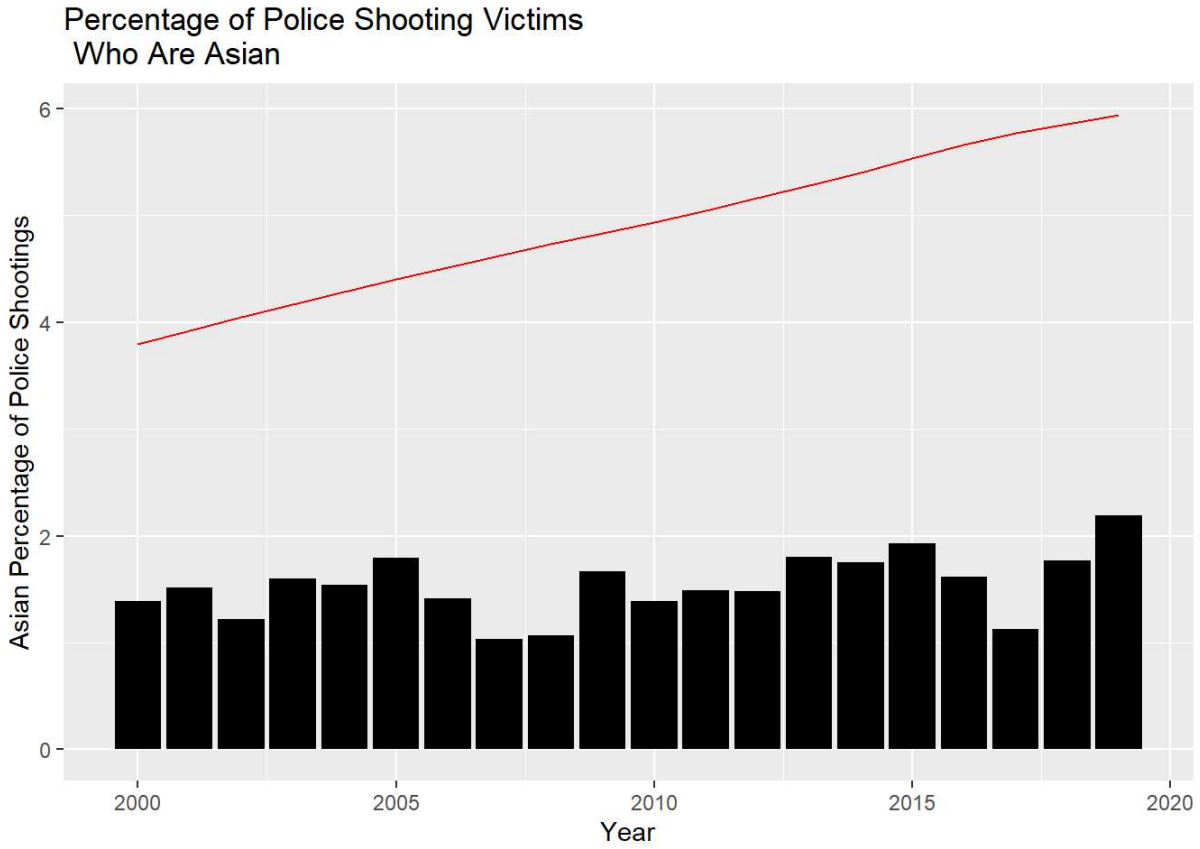

```
#column plot of the percentage of police shooting victims who are White with a red line showing the perc
entage of White people in the population

ggplot(data = police_shootings_And_US_Race ) + geom_col(mapping = aes(x = Year, y = White_Percentage), f
ill = "Black") + labs(title="Percentage of Police Shooting Victims \n Who Are White", x = "Year", y = "W
hite Percentage of Police Shootings") + geom_line(aes(x = Year, y = White_Proportion), color = "Red")
```



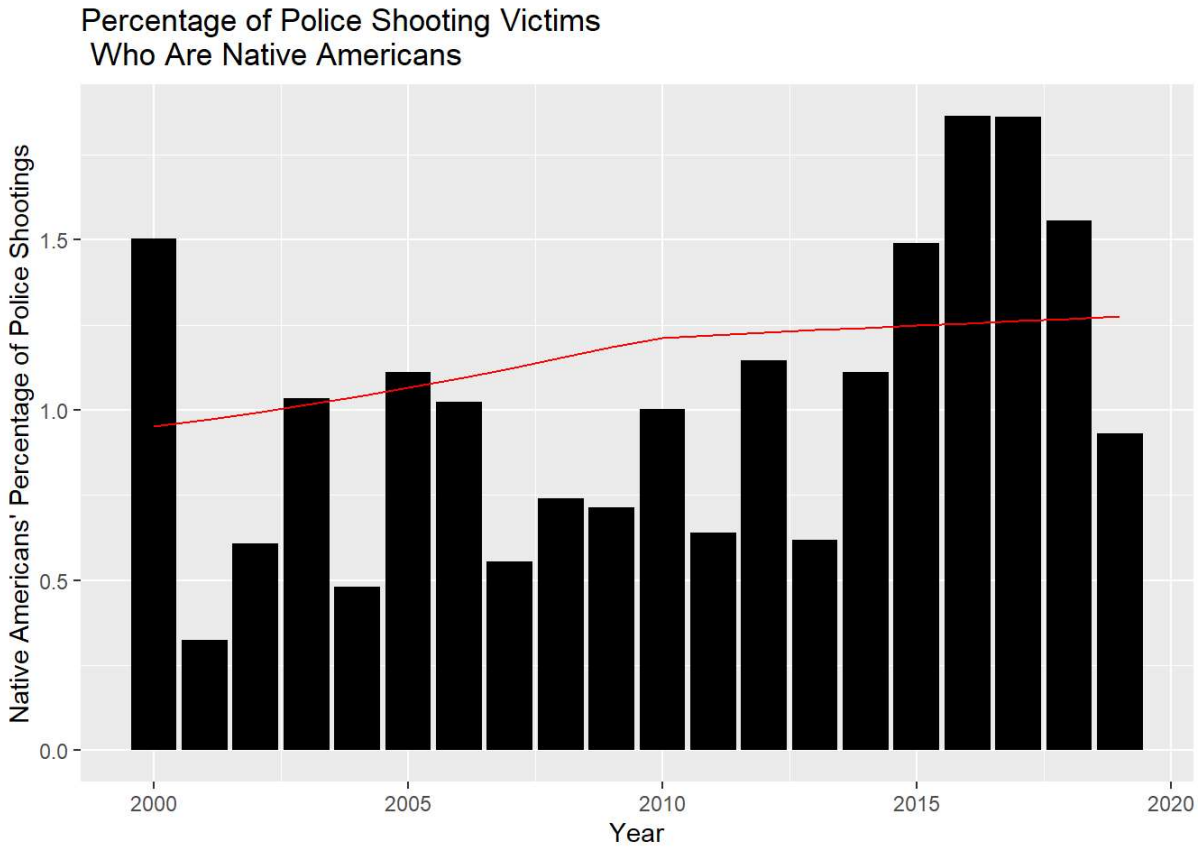

```
#column plot of the percentage of police shooting victims who are Hispanic with a red line showing the p
ercentage of Hispanic people in the population

ggplot(data = police_shootings_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Hispanic_Percentag
e), fill = "Black") + labs(title="Percentage of Police Shooting Victims \n Who Are Hispanic", x = "Year"
, y = "Hispanic Percentage of Police Shootings") + geom_line(aes(x = Year, y = Hispanic_Proportion), col
or = "Red")
```



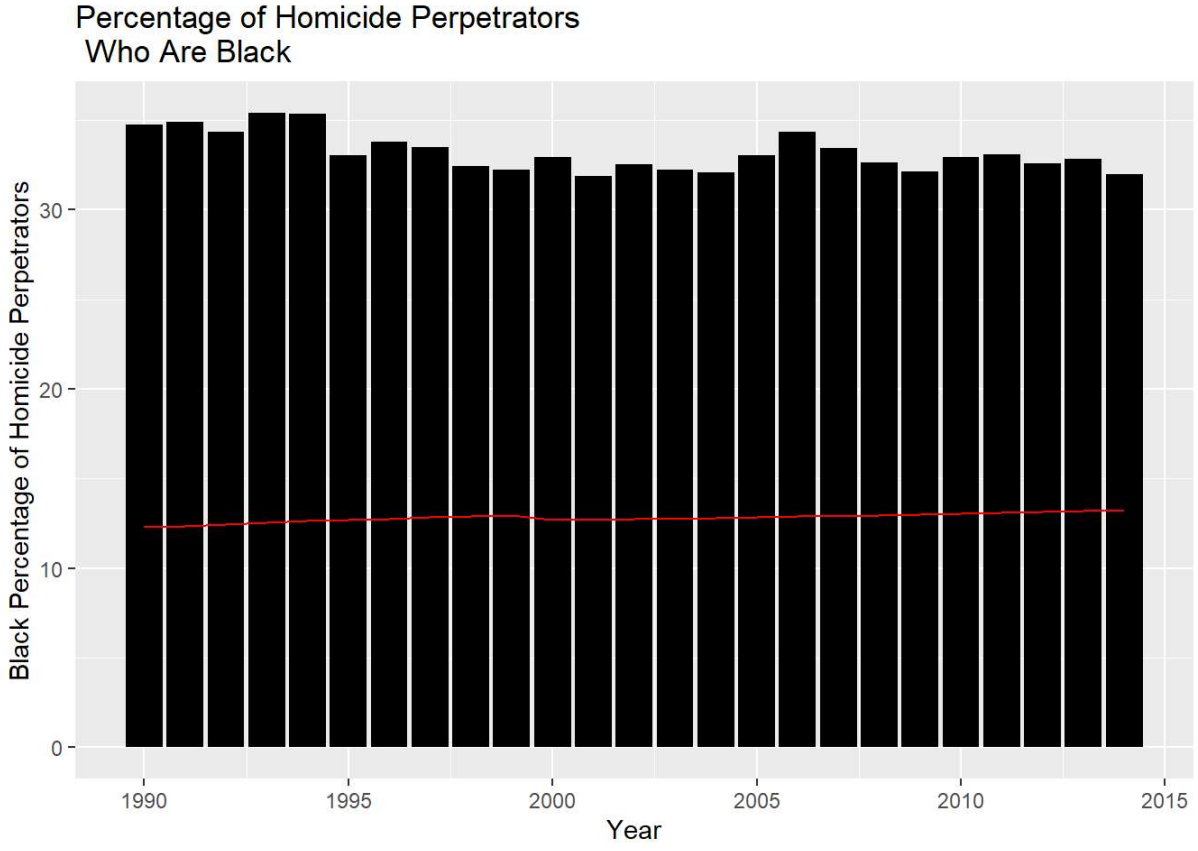
```
#column plot of the percentage of police shooting victims who are Asian with a red line showing the perc
entage of Asian people in the population

ggplot(data = police_shootings_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Asian_Percentage), f
ill = "Black") + labs(title="Percentage of Police Shooting Victims \n Who Are Asian", x = "Year", y = "A
sian Percentage of Police Shootings") + geom_line(aes(x = Year, y = Asian_Proportion), color = "Red")
```



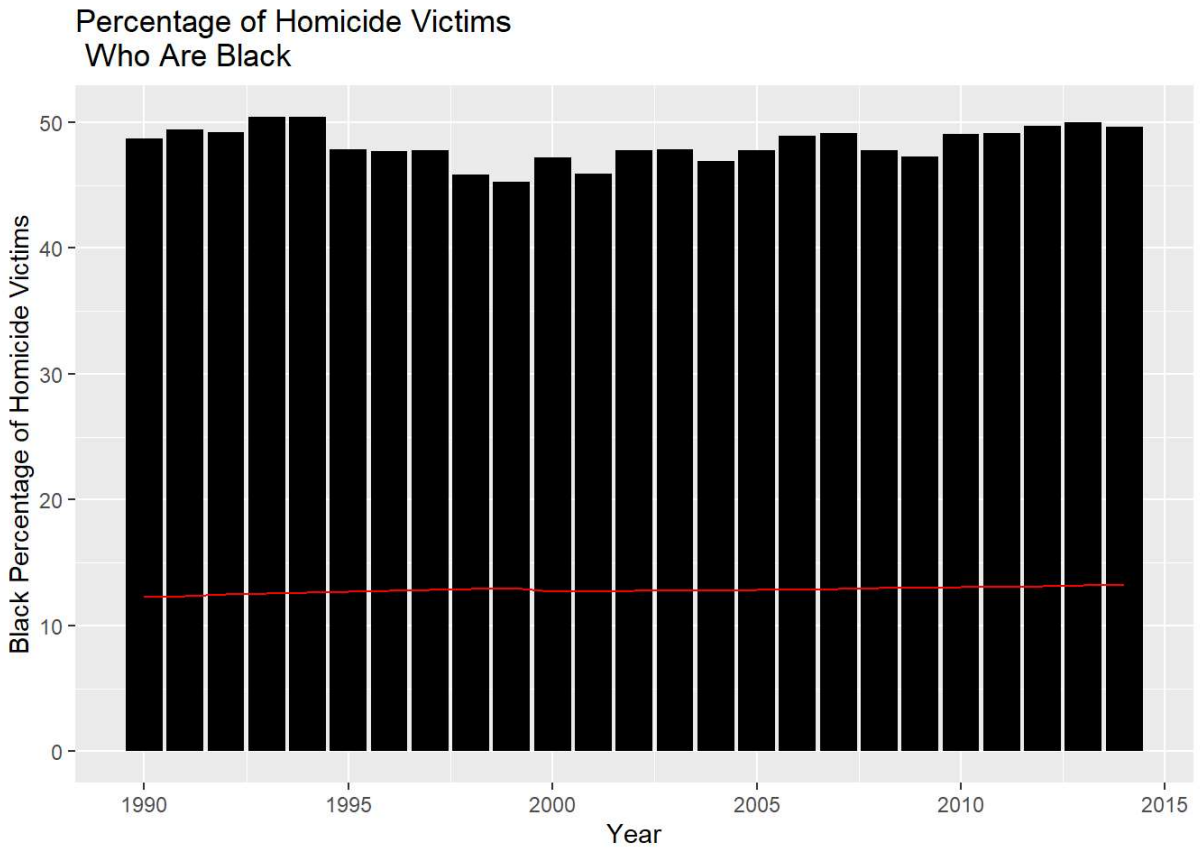

```
#column plot of the percentage of police shooting victims who are Native Americans with a red line showi
ng the percentage of Native Americans people in the population

ggplot(data = police_shootings_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Native_Percentage),
 fill = "Black") + labs(title="Percentage of Police Shooting Victims \n Who Are Native Americans", x =
"Year", y = "Native Americans' Percentage of Police Shootings") + geom_line(aes(x = Year, y = Native_Pro
portion), color = "Red")
```



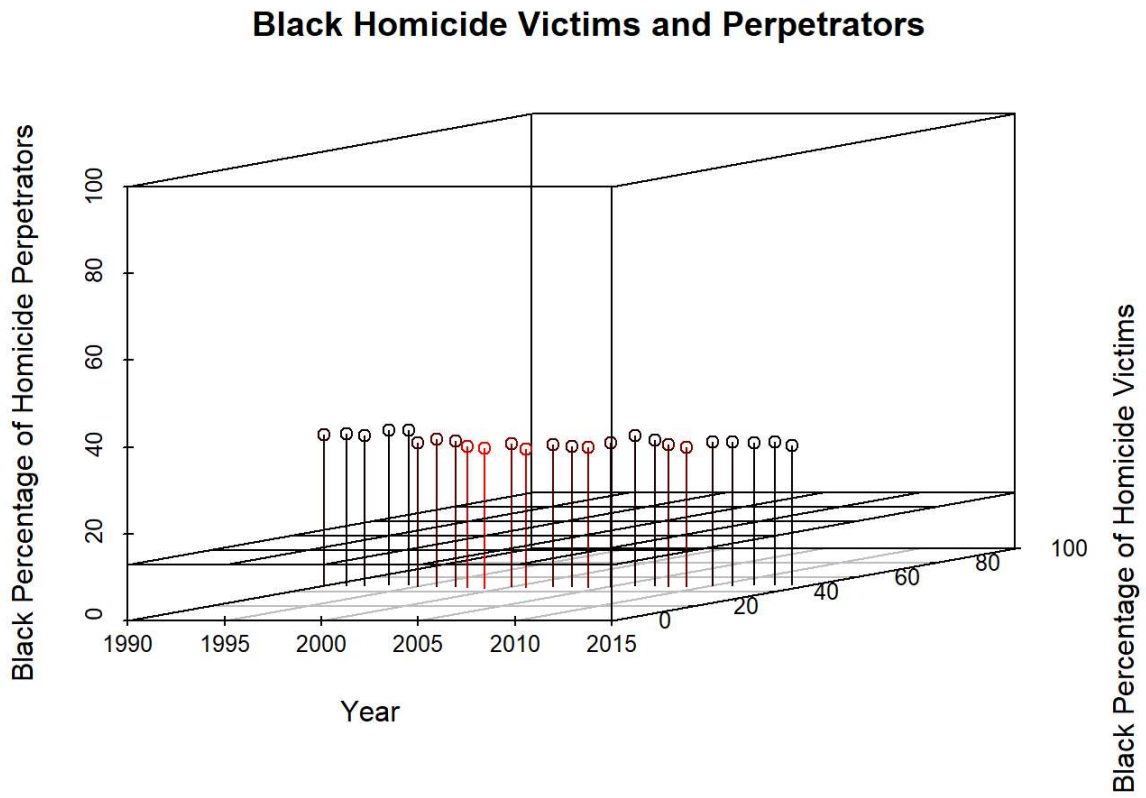

```
#column plot of the percentage of homicide perpetrators who are Black with a red line showing the percen
tage of Black people in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Black_Percentage), fill = "B
lack") + labs(title="Percentage of Homicide Perpetrators \n Who Are Black", x = "Year", y = "Black Perce
ntage of Homicide Perpetrators") + geom_line(aes(x = Year, y = Black_Proportion), color = "Red")
```



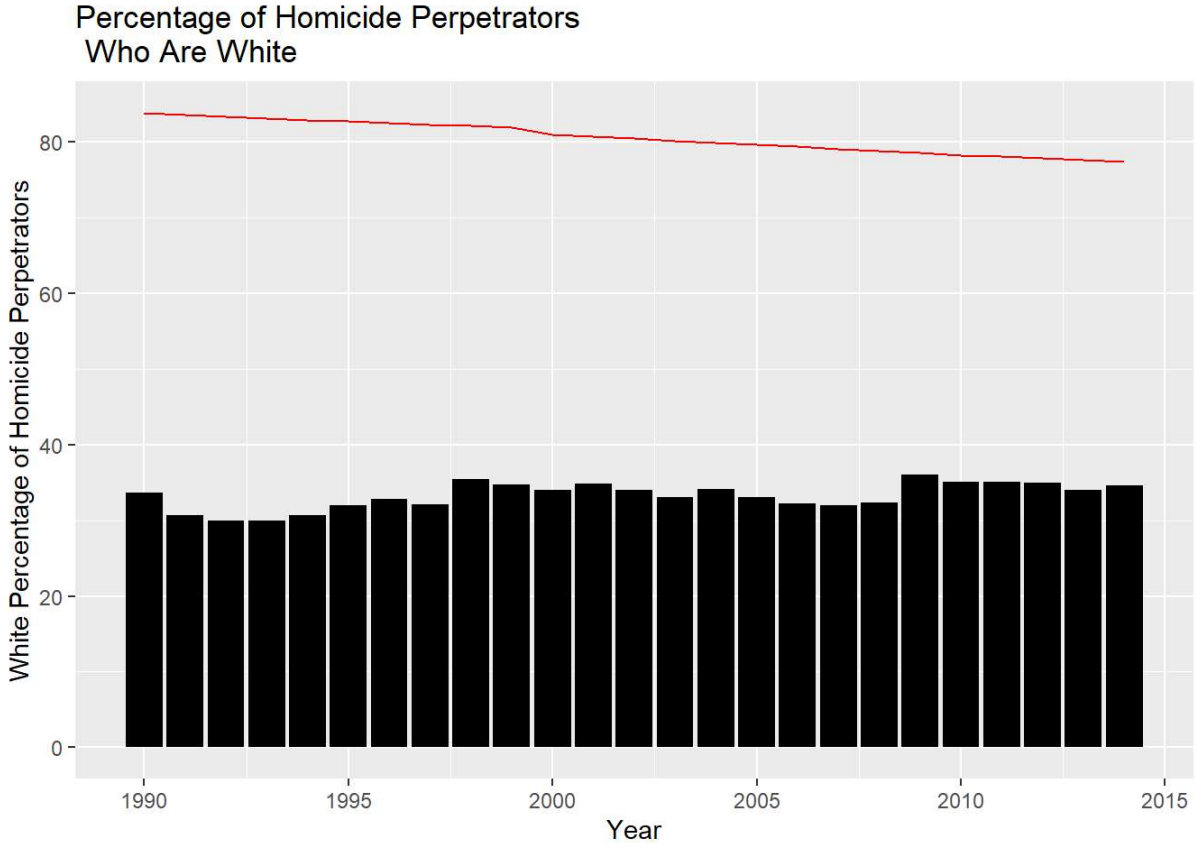
```
#column plot of the percentage of homicide victims who are Black with a red line showing the percentage
 of Black people in the population
ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Black_Percentage_victim), fi
ll = "Black") + labs(title="Percentage of Homicide Victims \n Who Are Black", x = "Year", y = "Black Per
centage of Homicide Victims") + geom_line(aes(x = Year, y = Black_Proportion), color = "Red")
```



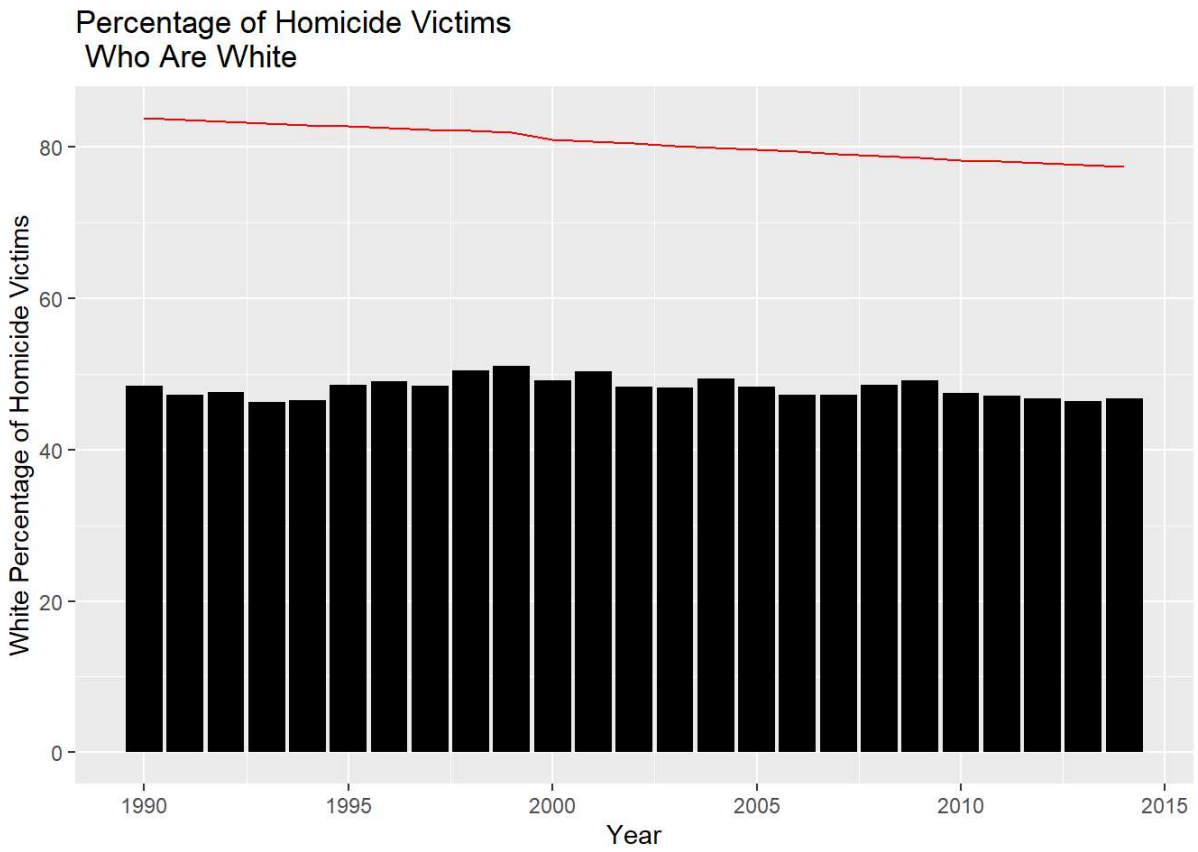

```
#3dplot of the percentage of homicide perpetrators who are Black and the percentage of homicide victims
 who are Black with a plane showing the average percentage of Black people in the population

s <- scatterplot3d(x = Homicide_And_US_Race$Year, y = Homicide_And_US_Race$Black_Percentage_victim, z =
 Homicide_And_US_Race$Black_Percentage, ylim = c(0,100), zlim = c(0,100),type = "h", angle = 15, highlig
ht.3d = TRUE, xlab = "Year", ylab = "Black Percentage of Homicide Victims", zlab = "Black Percentage of
 Homicide Perpetrators", main="Black Homicide Victims and Perpetrators")
s$plane3d(mean(Homicide_And_US_Race$Black_Proportion),0,0,lty="solid")
```



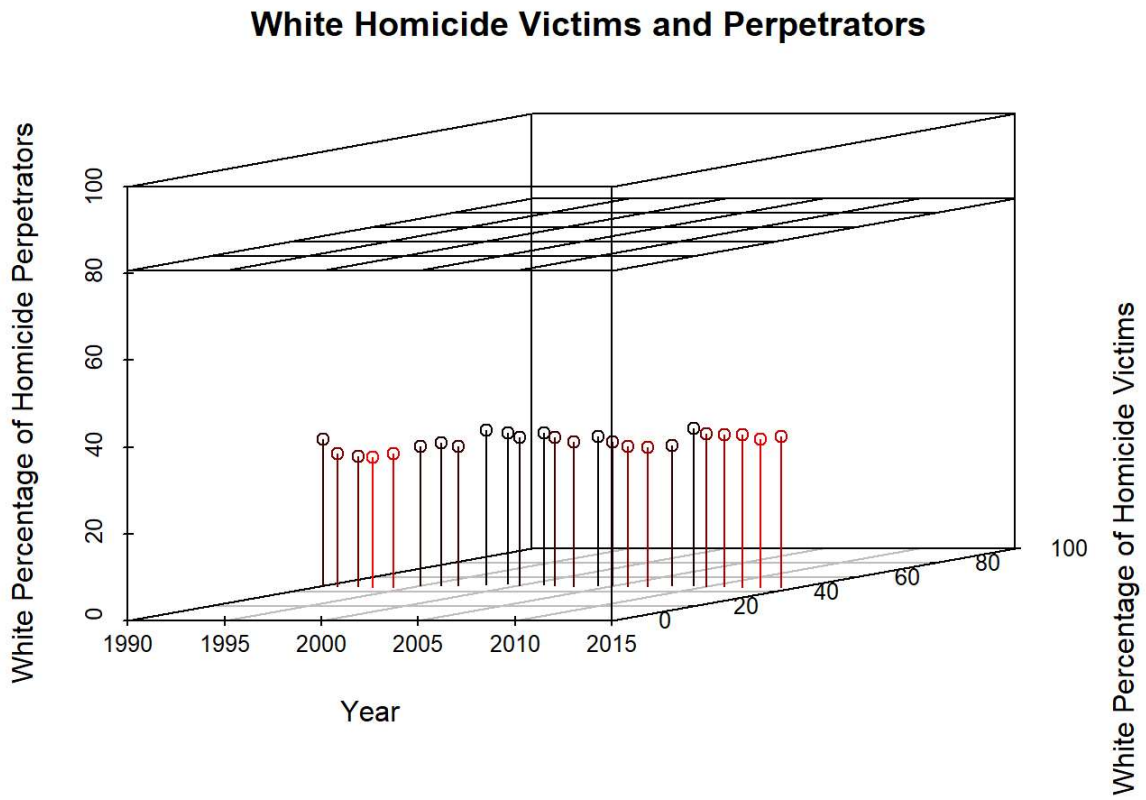

```
#column plot of the percentage of homicide perpetrators who are White with a red line showing the percen
tage of White people in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = White_Percentage), fill = "B
lack") + labs(title="Percentage of Homicide Perpetrators \n Who Are White", x = "Year", y = "White Perce
ntage of Homicide Perpetrators") + geom_line(aes(x = Year, y = White_Proportion), color = "Red")
```



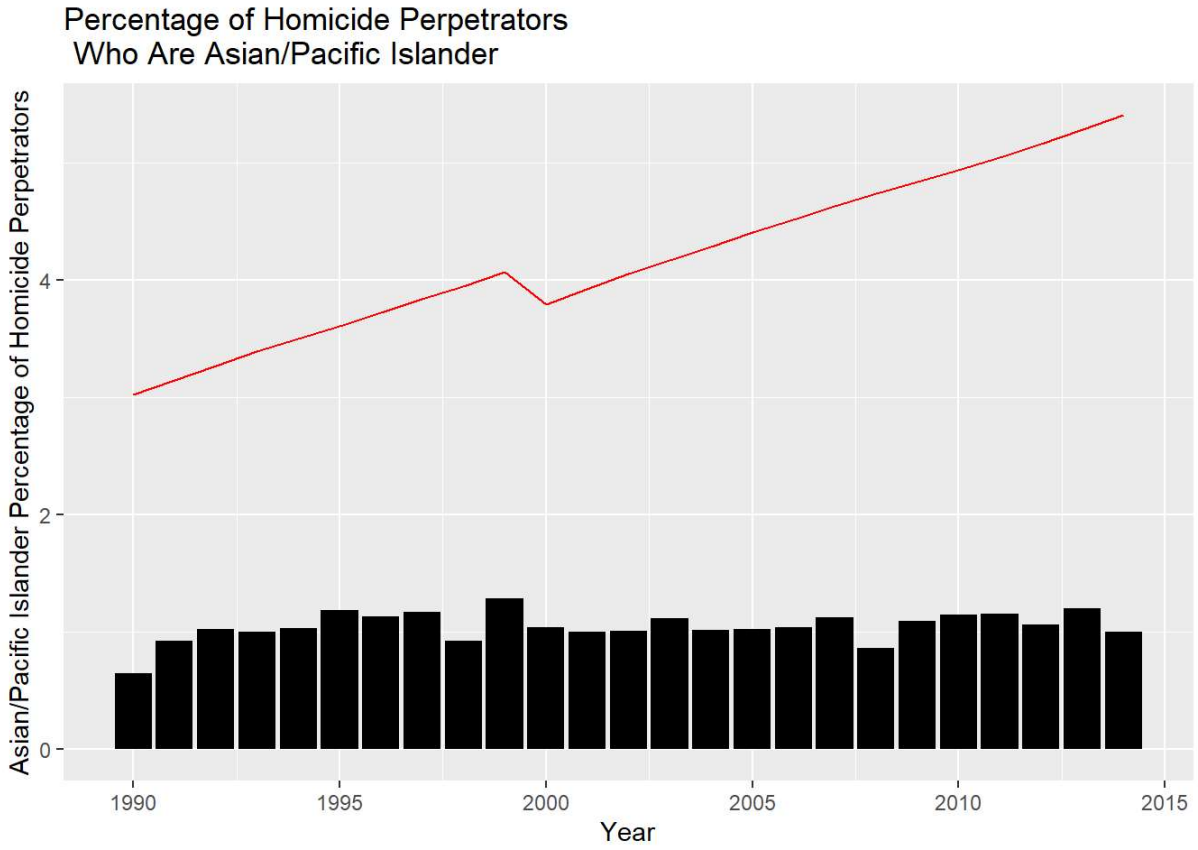
```
#column plot of the percentage of homicide victims who are White with a red line showing the percentage
 of White people in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = White_Percentage_victim), fi
ll = "Black") + labs(title="Percentage of Homicide Victims \n Who Are White", x = "Year", y = "White Per
centage of Homicide Victims") + geom_line(aes(x = Year, y = White_Proportion), color = "Red")
```



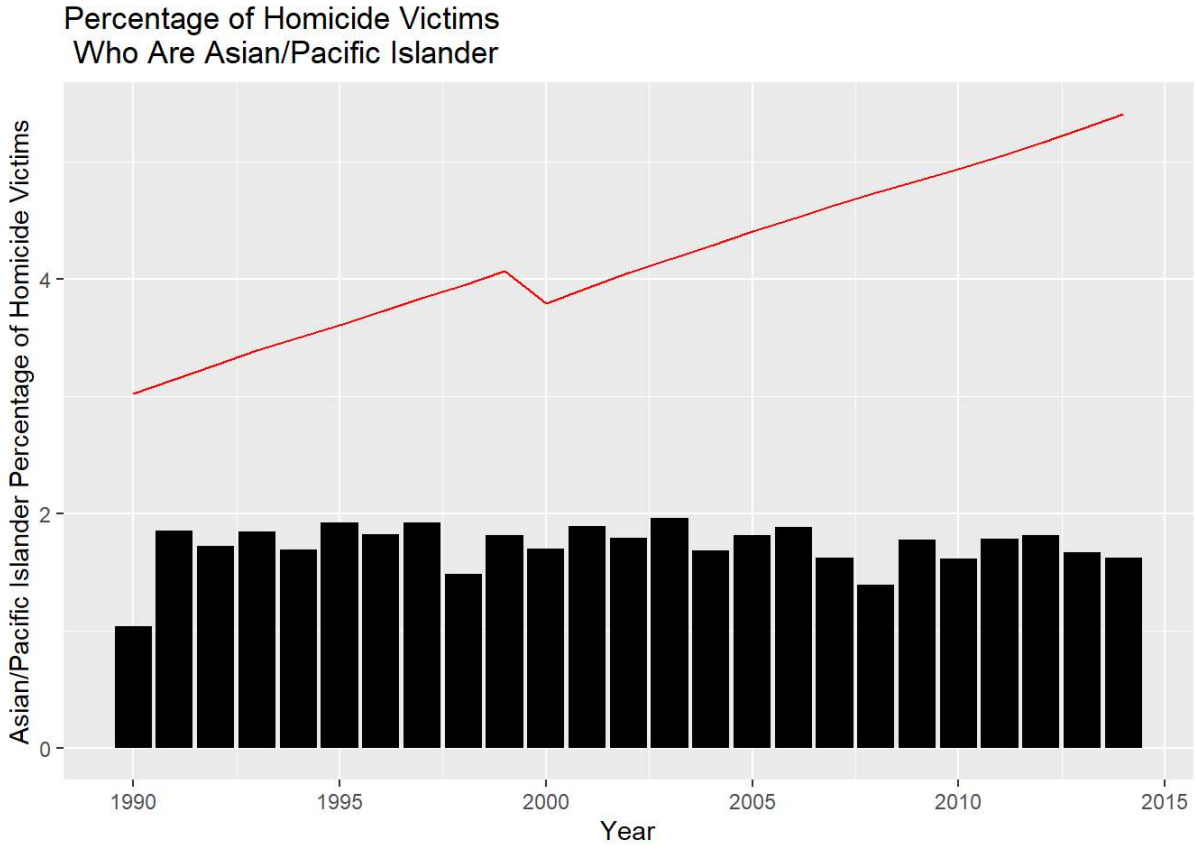

```
#3dplot of the percentage of homicide perpetrators who are White and the percentage of homicide victims
 who are White with a plane showing the average percentage of White people in the population

s <- scatterplot3d(x = Homicide_And_US_Race$Year, y = Homicide_And_US_Race$White_Percentage_victim, z =
 Homicide_And_US_Race$White_Percentage, ylim = c(0,100), zlim = c(0,100),type = "h", angle = 15, highlig
ht.3d = TRUE, xlab = "Year", ylab = "White Percentage of Homicide Victims", zlab = "White Percentage of
 Homicide Perpetrators", main="White Homicide Victims and Perpetrators")
s$plane3d(mean(Homicide_And_US_Race$White_Proportion),0,0,lty="solid")
```



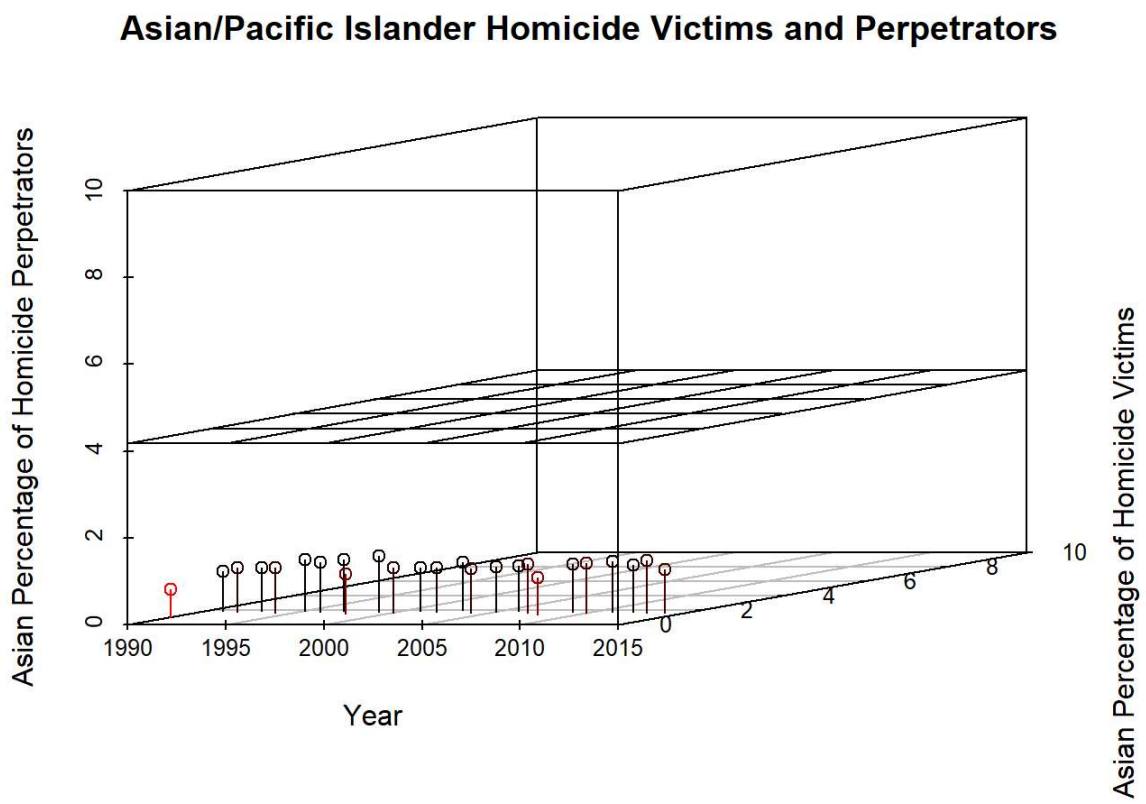

```
#column plot of the percentage of homicide perpetrators who are Asian with a red line showing the percen
tage of Asians in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Asian_Percentage), fill = "B
lack") + labs(title="Percentage of Homicide Perpetrators \n Who Are Asian/Pacific Islander", x = "Year",
y = "Asian/Pacific Islander Percentage of Homicide Perpetrators") + geom_line(aes(x = Year, y = Asian_Pr
oportion), color = "Red")
```



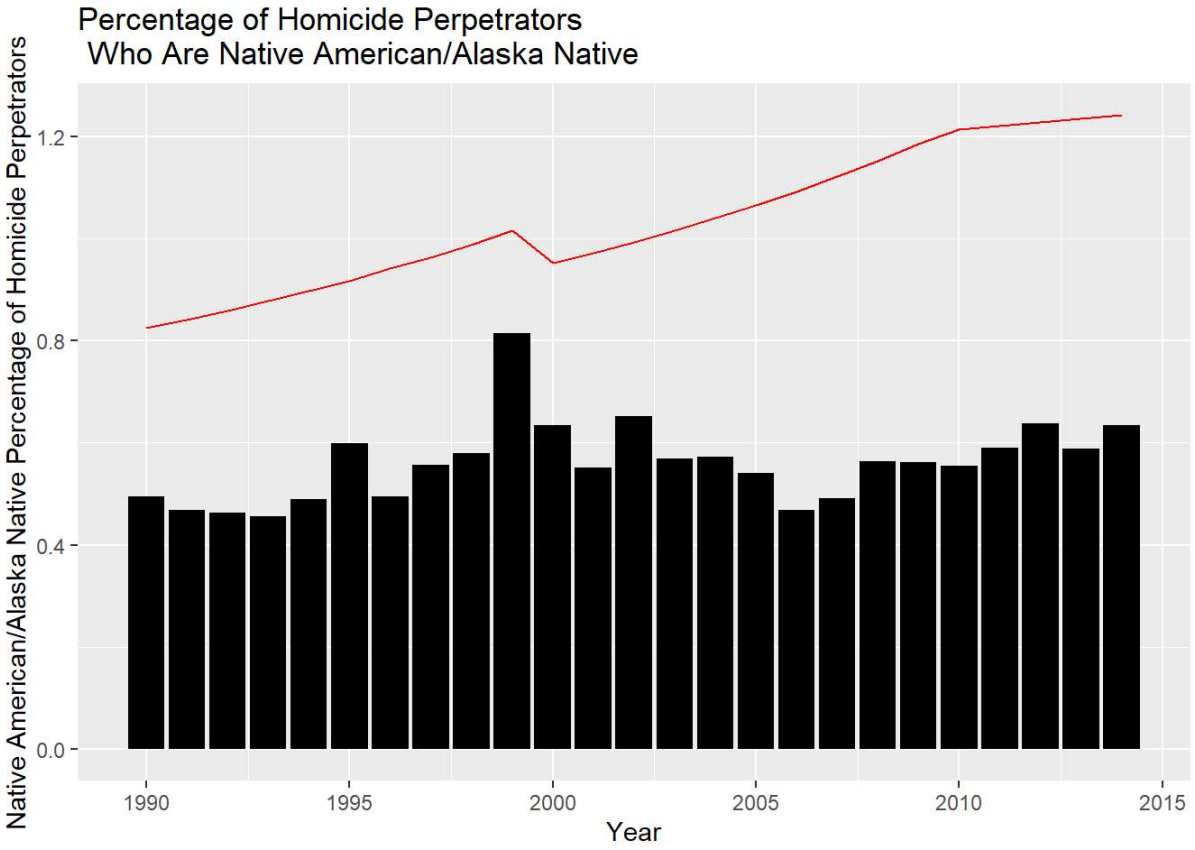
```
#column plot of the percentage of homicide victims who are Asian with a red line showing the percentage
 of Asians in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Asian_Percentage_victim), fi
ll = "Black") + labs(title="Percentage of Homicide Victims \n Who Are Asian/Pacific Islander", x = "Yea
r", y = "Asian/Pacific Islander Percentage of Homicide Victims") + geom_line(aes(x = Year, y = Asian_Pro
portion), color = "Red")
```



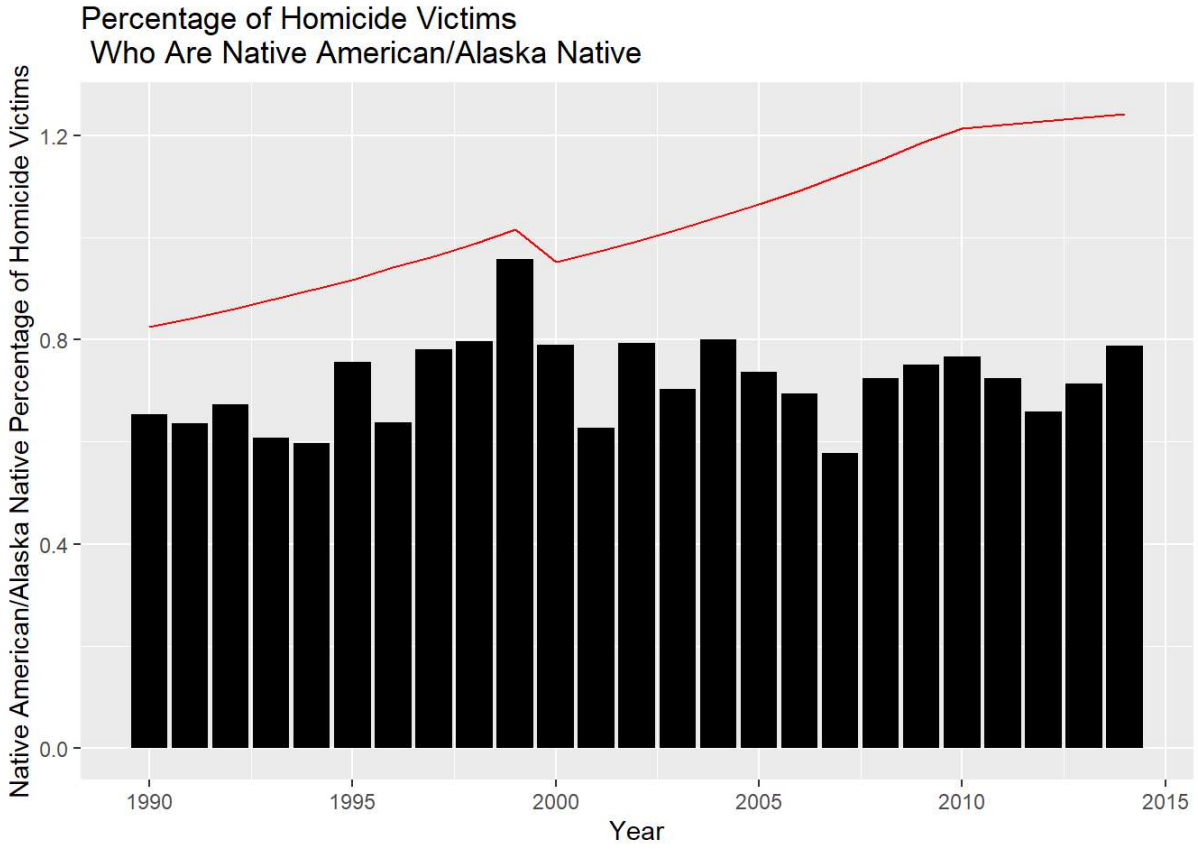

```
#3dplot of the percentage of homicide perpetrators who are Asian and the percentage of homicide victims
 who are Asian with a plane showing the average percentage of Asians in the population

s <- scatterplot3d(x = Homicide_And_US_Race$Year, y = Homicide_And_US_Race$Asian_Percentage_victim, z =
 Homicide_And_US_Race$Asian_Percentage, ylim = c(0,10), zlim = c(0,10),type = "h", angle = 15, highligh
t.3d = TRUE, xlab = "Year", ylab = "Asian Percentage of Homicide Victims", zlab = "Asian Percentage of H
omicide Perpetrators", main="Asian/Pacific Islander Homicide Victims and Perpetrators")
s$plane3d(mean(Homicide_And_US_Race$Asian_Proportion),0,0,lty="solid")
```



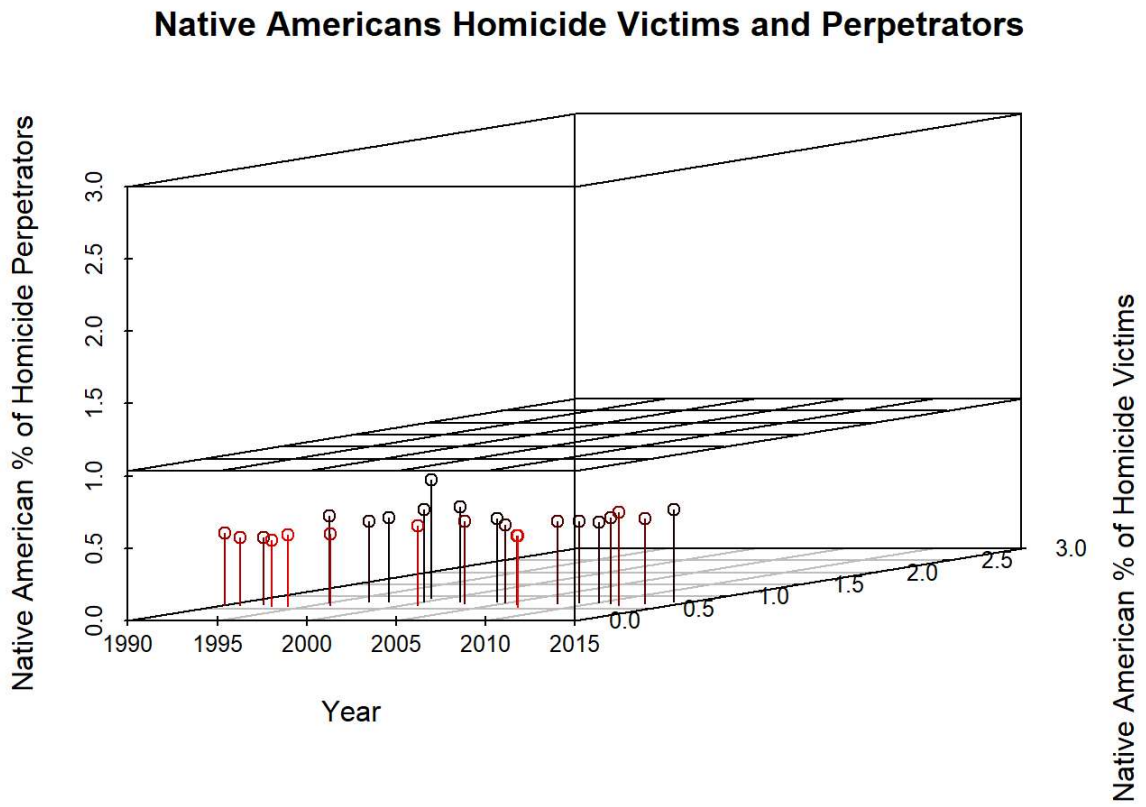

```
#column plot of the percentage of homicide perpetrators who are Native Americans with a red line showing
the percentage of Native Americans in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Native_Percentage), fill =
"Black") + labs(title="Percentage of Homicide Perpetrators \n Who Are Native American/Alaska Native", x
 = "Year", y = "Native American/Alaska Native Percentage of Homicide Perpetrators") + geom_line(aes(x =
 Year, y = Native_Proportion), color = "Red")
```



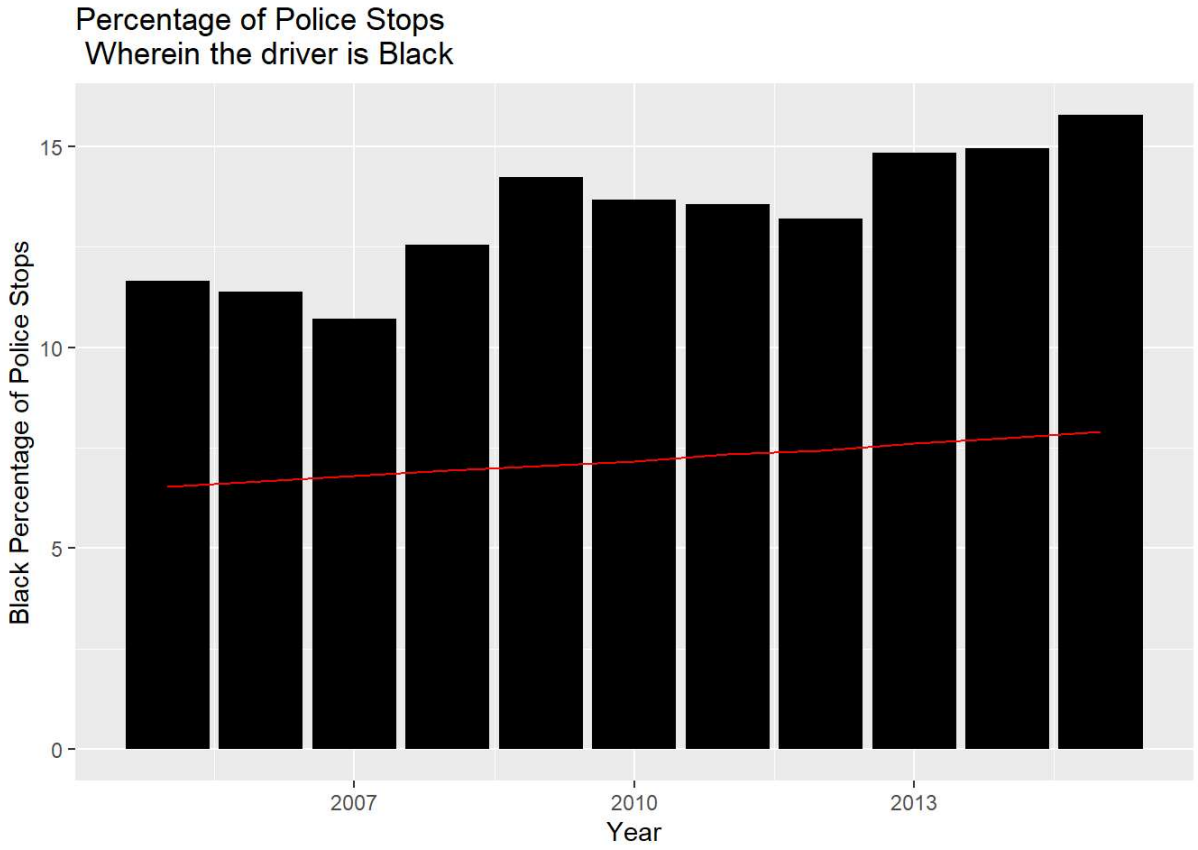
```
#column plot of the percentage of homicide victims who are Native Americans with a red line showing the
 percentage of Native Americans in the population

ggplot(data = Homicide_And_US_Race ) + geom_col(mapping = aes(x = Year, y = Native_Percentage_victim), f
ill = "Black") + labs(title="Percentage of Homicide Victims \n Who Are Native American/Alaska Native", x
= "Year", y = "Native American/Alaska Native Percentage of Homicide Victims") + geom_line(aes(x = Year,
 y = Native_Proportion), color = "Red")
```



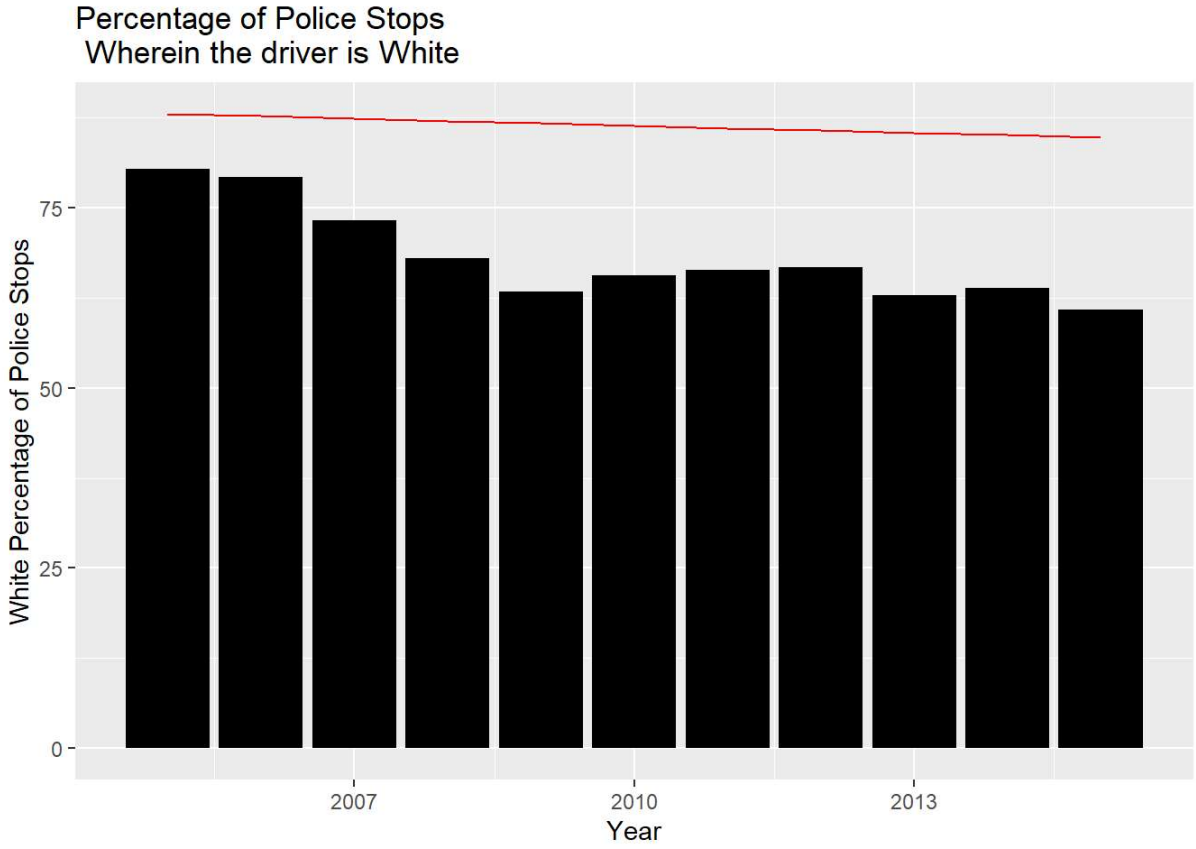

```
#3dplot of the percentage of homicide perpetrators who are Native Americans and the percentage of homici
de victims who are Native Americans with a plane showing the average percentage of Native Americans in t
he population

s <- scatterplot3d(x = Homicide_And_US_Race$Year, y = Homicide_And_US_Race$Native_Percentage_victim, z =
Homicide_And_US_Race$Native_Percentage, ylim = c(0,3), zlim = c(0,3),type = "h", angle = 15, highlight.3
d = TRUE, xlab = "Year", ylab = "Native American % of Homicide Victims", zlab = "Native American % of Ho
micide Perpetrators", main="Native Americans Homicide Victims and Perpetrators")
s$plane3d(mean(Homicide_And_US_Race$Native_Proportion),0,0,lty="solid")
```



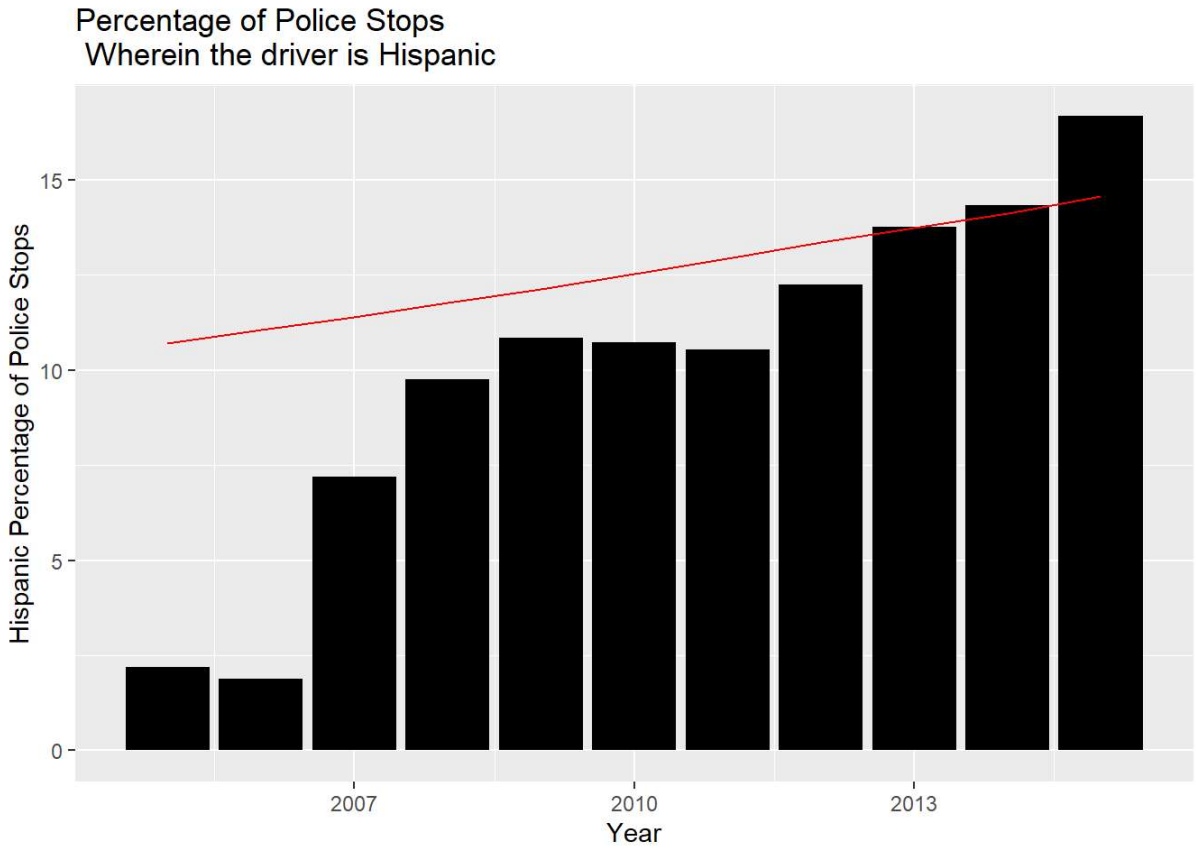

```
#column plot of the percentage of police stops who are Black with a red line showing the percentage of B
lack people in the population of Rhode Island

ggplot(data = police_stops_And_Rhode_Island_Race) + geom_col(mapping = aes(x = Year, y = Black_Percentag
e), fill = "Black") + labs(title="Percentage of Police Stops \n Wherein the driver is Black", x = "Year"
, y = "Black Percentage of Police Stops") + geom_line(aes(x = Year, y = Black_Proportion), color = "Red"
)
```



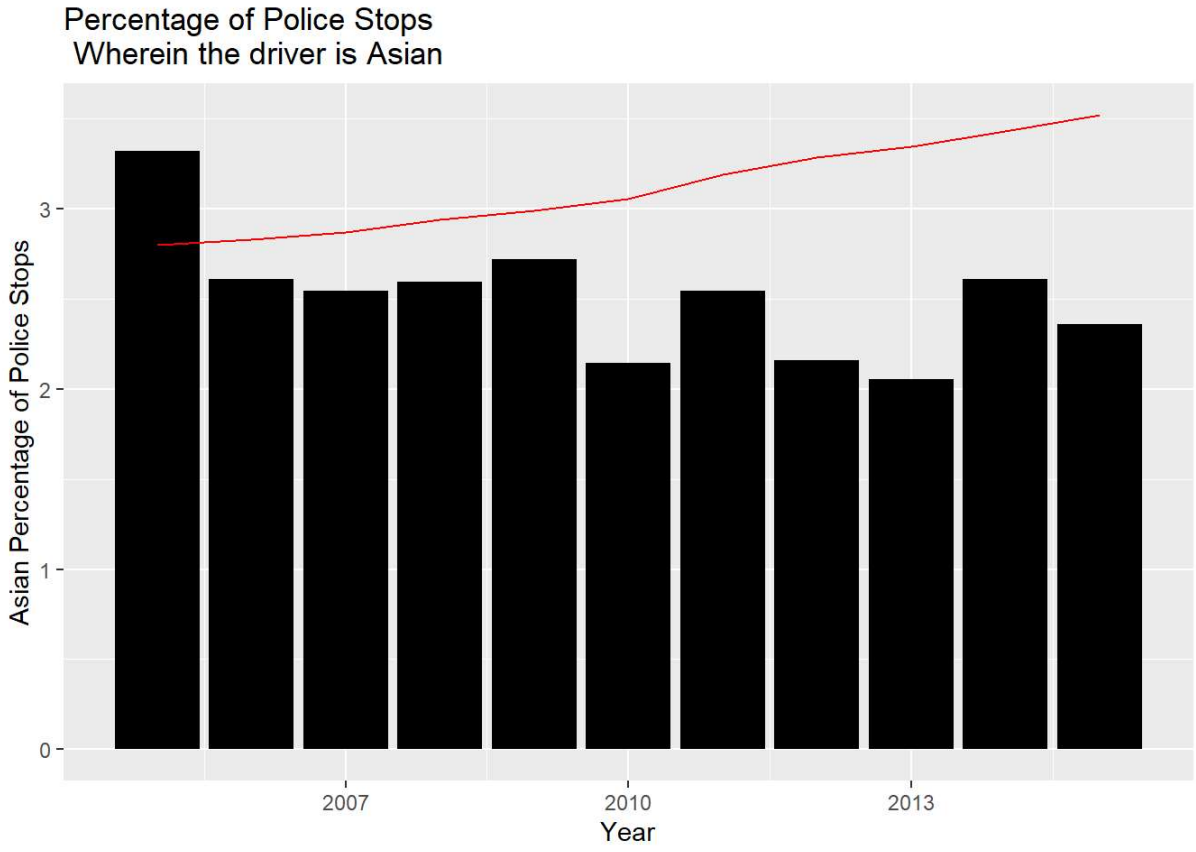
```
#column plot of the percentage of police stops who are White with a red line showing the percentage of W
hite people in the population of Rhode Island

ggplot(data = police_stops_And_Rhode_Island_Race) + geom_col(mapping = aes(x = Year, y = White_Percentag
e), fill = "Black") + labs(title="Percentage of Police Stops \n Wherein the driver is White", x = "Year"
, y = "White Percentage of Police Stops") + geom_line(aes(x = Year, y = White_Proportion), color = "Red"
)
```



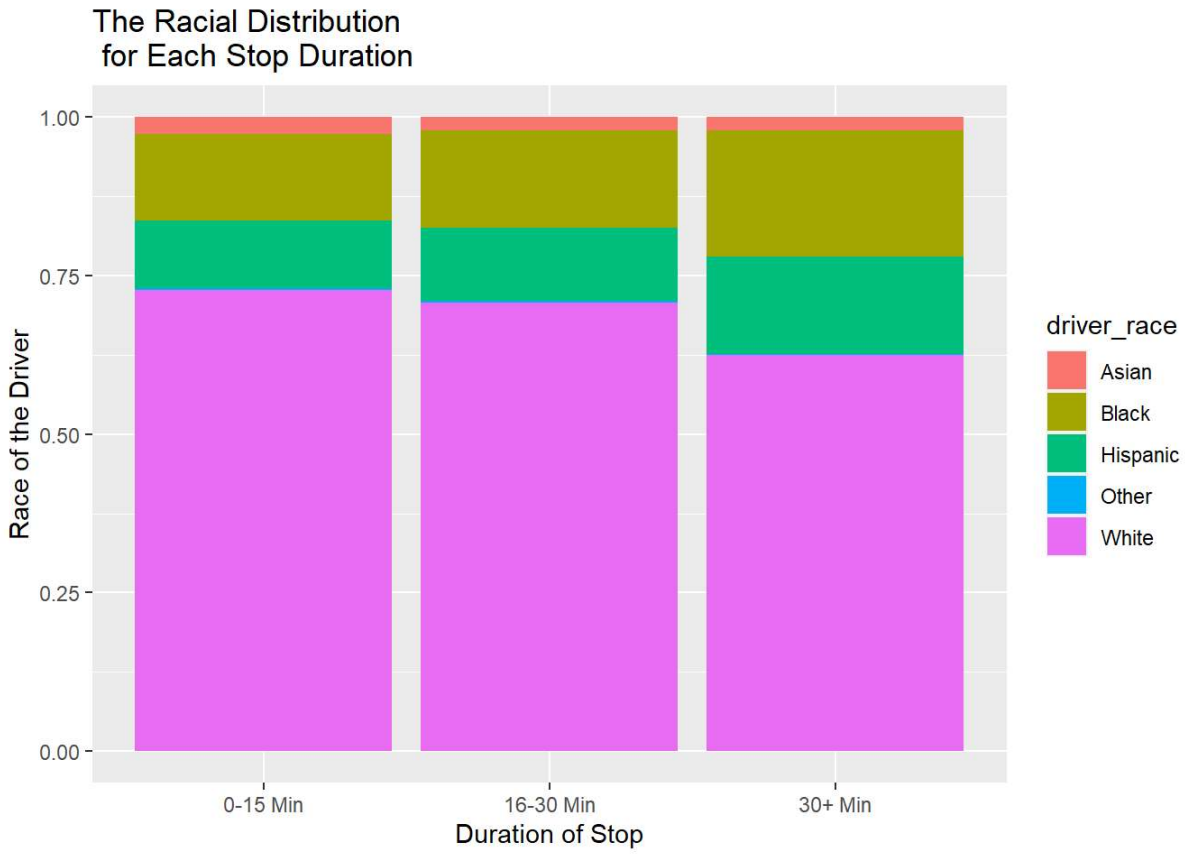

```
#column plot of the percentage of police stops who are Hispanic with a red line showing the percentage o
f Hispanics in the population of Rhode Island

ggplot(data = police_stops_And_Rhode_Island_Race) + geom_col(mapping = aes(x = Year, y = Hispanic_Percen
tage), fill = "Black") + labs(title="Percentage of Police Stops \n Wherein the driver is Hispanic", x =
"Year", y = "Hispanic Percentage of Police Stops") + geom_line(aes(x = Year, y = Hispanic_Proportion), c
olor = "Red")
```



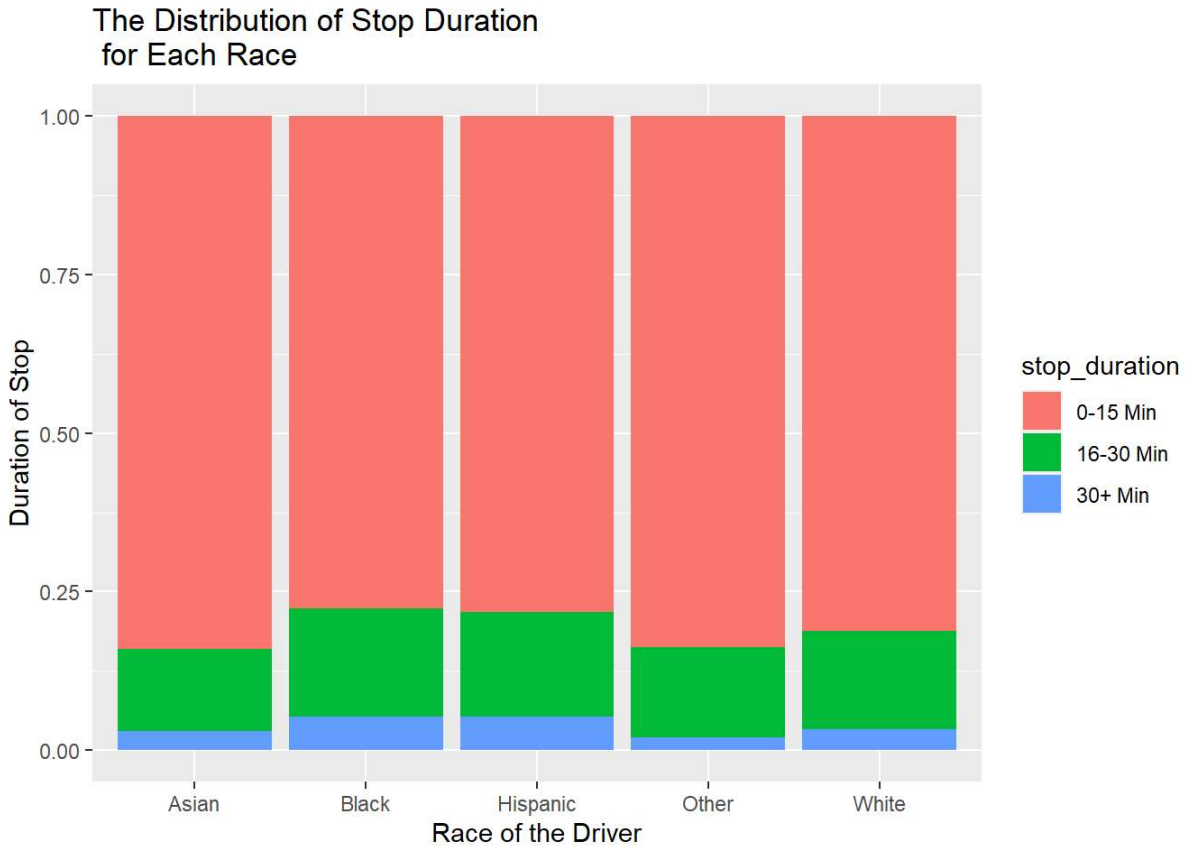

```
#column plot of the percentage of police stops who are Asian with a red line showing the percentage of A
sian in the population of Rhode Island

ggplot(data = police_stops_And_Rhode_Island_Race) + geom_col(mapping = aes(x = Year, y = Asian_Percentag
e), fill = "Black") + labs(title="Percentage of Police Stops \n Wherein the driver is Asian", x = "Year"
, y = "Asian Percentage of Police Stops") + geom_line(aes(x = Year, y = Asian_Proportion), color = "Red"
)
```



```
#bar plot shows the racial distribution for each stop_duration
ggplot(police_stops_filtered ) + geom_bar(mapping = aes(x= stop_duration, fill = driver_race), position
 = "fill")  + labs(title="The Racial Distribution \n for Each Stop Duration", x = "Duration of Stop", y
 = "Race of the Driver")
```




```
#bar plot shows the distribution of stop_duration for each race
ggplot(police_stops_filtered ) + geom_bar(mapping = aes(x= driver_race, fill = stop_duration), position
 = "fill")  + labs(title="The Distribution of Stop Duration\n for Each Race", x = "Race of the Driver",
 y = "Duration of Stop")
```



- Data Modelling

```
#creates training data out of the original data after selecting the even years
Homicide_And_US_Race_training <- Homicide_And_US_Race %>% filter(Year %% 2 == 0)

#finds whether Black or White is the highest possibility for the race of a homicide victim
mean(Homicide_And_US_Race_training$Black_Percentage_victim)
```

```
## [1] 48.36427
```

```
mean(Homicide_And_US_Race_training$White_Percentage_victim)
```

```
## [1] 48.14077
```

The average percentage of a homicide victim being black is higher than any other race. Thus, the baseline model is that for every state, the majority of homicide victims are black.

```
#creates test data by selecting the odd years and grouping the homicide victims by states
homicide_test <- homicide %>% filter(Year %% 2 == 1) %>% group_by(State)  %>%  count(`Victim Race`) %>%
 pivot_wider(names_from = `Victim Race`, values_from = n) %>% rename(  `Asian/Pacific Islander_victim` =
`Asian/Pacific Islander`, Black_victim = Black  , `Native American/Alaska Native_victim` = `Native Ameri
can/Alaska Native` ,  Unknown_victim = Unknown, White_victim = White) %>% mutate( Total = sum(`Asian/Pac
ific Islander_victim`, Black_victim, `Native American/Alaska Native_victim` ,Unknown_victim, White_victi
m, na.rm = TRUE), Asian_Percentage_victim = `Asian/Pacific Islander_victim`*100/ Total, Black_Percentage
_victim = Black_victim*100/ Total,  Native_Percentage_victim =  `Native American/Alaska Native_victim`*1
00/ Total, Unspecified_Percentage_victim =`Unknown_victim`*100/ Total, White_Percentage_victim = White_v
ictim *100/ Total)

#adds the result of the test to the tibble
homicide_test1 <- homicide_test %>% mutate(Black_majority = ifelse(max(Asian_Percentage_victim, Black_Pe
rcentage_victim, Native_Percentage_victim, Unspecified_Percentage_victim, White_Percentage_victim, na.rm
= TRUE) == Black_Percentage_victim, 1, 0))

#checks for who many states was the model correct
sum(homicide_test1$Black_majority, na.rm = TRUE)
```

```
## [1] 21
```

The base model is true for 21 states. Black is the most likely race of a homicide victim is 21 states even though only D.C has a black majority.

Conclusion

The findings show that from 1990 to 2014, over 45% of the victims and 30% of the perpetrators of homicide crime in the U.S are Black even though the percentage of Black people in the U.S didn't exceed 15%, which demonstrates that Black communities are rife with crimes in contrast to White communities. Furthermore, in 21 states, the most likely race for the victim of Homicide is Black even though only D.C had a Black majority during the years of the study. In addition, over 20% of police shootings victims each year from 2000 to 2019 were Black, and the percentage was as high as 30% in 2012. In Rhode Island, the percentage of drivers stopped by the police who are Black are as high as double the proportion of Black people in the population of Rhode Island; furthermore, Black drivers were stopped for longer duration than non-Black drivers. For instance, 5% of the Black drivers were stopped by the police for more than 30 minutes compared to 3% of White drivers. These findings illustrate that Black people are targeted by the police at unjustified rates and underscores how Black communities are prone to the most dangerous crime, homicide.

The analysis of the police stops in Rhode Island is limited in its span and can't be generalized to all the United States. Also, one of my aims was to visualize the trend in racial bias in homicide, police shootings, and police stops across the time frame of each dataset. However, the years 1980 to 1989 in the homicide data set were omitted after joining the homicide and the Census Bureau data sets because the census data set didn't have data for the U.S population from 1980 to 1990. A next step might be working with more comprehensive data, such as a U.S census data set that includes all the years since 1980 and a data set of police stops across all the United States.