# Fatal Force Data Visualization

Objective: Analyze the Washington Post's fatal force data set, with a focus on data visualization.

(Link to data set: https://github.com/washingtonpost/data-police-shootings (https://github.com/washingtonpost/data-police-shootings))

Load packages and download data set.

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.3     ✓ purrr   0.3.4
## ✓ tibble  3.1.1     ✓ dplyr   1.0.6
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   1.4.0     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
library(lubridate)
```

```
## 
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
## 
##     date, intersect, setdiff, union
```

```
library(knitr)
library(kableExtra)
```

```
## 
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
## 
##     group_rows
```

```
library(tinytex)
setwd("~/Desktop/RStudio Directory")
police_data <- read_csv('fatal-police-shootings-data (2).csv')
```

```
## 
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   id = col_double(),
##   name = col_character(),
##   date = col_date(format = ""),
##   manner_of_death = col_character(),
##   armed = col_character(),
##   age = col_double(),
##   gender = col_character(),
##   race = col_character(),
##   city = col_character(),
##   state = col_character(),
##   signs_of_mental_illness = col_logical(),
##   threat_level = col_character(),
##   flee = col_character(),
##   body_camera = col_logical(),
##   longitude = col_double(),
##   latitude = col_double(),
##   is_geocoding_exact = col_logical()
## )
```

Review data set.

```
view(police_data)
```

Determine the first and last dates on which shootings occurred and the total number of fatal shootings that occurred over this time frame.

```
head(police_data)
```

```
## # A tibble: 6 x 17
##      id name    date       manner_of_death  armed   age gender race  city  state
##   <dbl> <chr>   <date>     <chr>            <chr> <dbl> <chr>  <chr> <chr> <chr>
## 1     3 Tim El… 2015-01-02 shot             gun      53 M      A     Shel… WA
## 2     4 Lewis … 2015-01-02 shot             gun      47 M      W     Aloha OR
## 3     5 John P… 2015-01-03 shot and Tasered unar…    23 M      H     Wich… KS
## 4     8 Matthe… 2015-01-04 shot             toy …    32 M      W     San … CA
## 5     9 Michae… 2015-01-04 shot             nail…    39 M      H     Evans CO
## 6    11 Kennet… 2015-01-04 shot             gun      18 M      W     Guth… OK
## # … with 7 more variables: signs_of_mental_illness <lgl>, threat_level <chr>,
## #   flee <chr>, body_camera <lgl>, longitude <dbl>, latitude <dbl>,
## #   is_geocoding_exact <lgl>
```

```
tail(police_data)
```

```
## # A tibble: 6 x 17
##      id name   date       manner_of_death armed    age gender race  city   state
##   <dbl> <chr>  <date>     <chr>           <chr>  <dbl> <chr>  <chr> <chr>  <chr>
## 1  7414 Ronal… 2021-12-07 shot            gun       88 M      <NA>  River… FL
## 2  7410 <NA>   2021-12-08 shot            knife     NA M      <NA>  Color… CO
## 3  7415 <NA>   2021-12-08 shot            gun       NA <NA>   <NA>  Colum… OH
## 4  7416 <NA>   2021-12-08 shot            gun       NA M      <NA>  Stock… CA
## 5  7412 <NA>   2021-12-09 shot            undet…    NA M      <NA>  Henry… GA
## 6  7411 <NA>   2021-12-10 shot            gun       NA M      <NA>  Color… CO
## # … with 7 more variables: signs_of_mental_illness <lgl>, threat_level <chr>,
## #   flee <chr>, body_camera <lgl>, longitude <dbl>, latitude <dbl>,
## #   is_geocoding_exact <lgl>
```

```
nrow(police_data)
```

```
## [1] 6800
```

- First recorded shooting: January 2, 2015
- Last recorded shooting: December 10, 2021
- Total fatal shootings: 6,800

Determine how many fatal shootings occurred per year from 2015 - 2021.

```
police_data_tally <- police_data %>% mutate(assign_1 = 1)

police_data_year_col <- police_data_tally %>%
                        mutate(year = year(police_data_tally$date))

table_value_1 <- police_data_year_col %>% group_by(year) %>%
                                          summarize(yearly_sums = sum(assign_1))

table_value_1 %>%
  kbl(col.names = c('Year', 'Total Fatal Shootings Per Year'), align = c('c', 'c')) %>%
  kable_classic_2(full_width = F)
```

| Year | Total Fatal Shootings Per Year |
|---|---|
| 2015 | 994 |
| 2016 | 957 |
| 2017 | 984 |
| 2018 | 990 |
| 2019 | 999 |
| 2020 | 1021 |
| 2021 | 855 |

Determine the average number of fatal shootings per year from 2015 - 2021. (Note: the data set is missing shooting data for the last couple weeks of December 2021, so the average listed below is likely less than the true average.)

```
yearly_sums_data <- police_data_year_col %>%
                    group_by(year) %>%
                    summarize(yearly_sums = sum(assign_1)) %>%
                    mutate(yearly_sums = yearly_sums)

yearly_sums_data %>% summarize(mean(yearly_sums)) %>%
                     mutate_if(is.numeric, round, digits = 2)
```

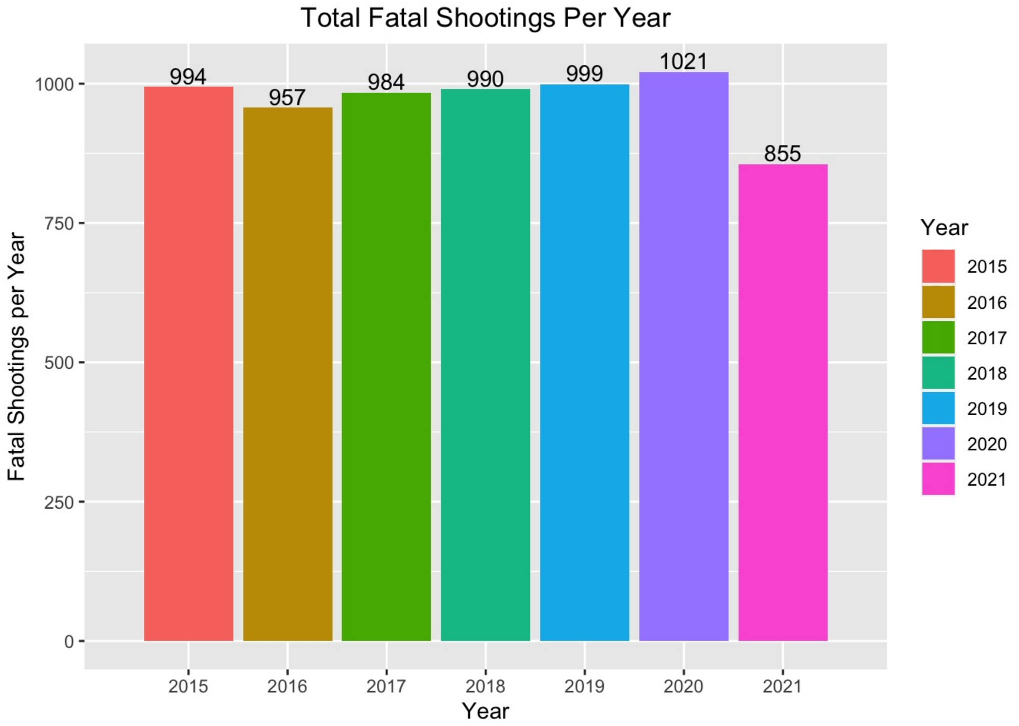
```
## # A tibble: 1 x 1
##   `mean(yearly_sums)`
##                 <dbl>
## 1                971.
```

From the code above we see that approximately 971 fatal shootings occurred per year for the years on record.

Create a bar graph that shows the total number of fatal shootings per year from 2015 - 2021.

```
ggplot() +
geom_col (yearly_sums_data, mapping = aes(x = year, y = yearly_sums, fill = as.factor(year))) +
geom_text(yearly_sums_data, mapping = aes(x = year, y = yearly_sums, label = yearly_sums, vjust = -.2)) +
scale_x_discrete(limits = c(2015, 2016, 2017, 2018, 2019, 2020, 2021),
                 labels = c("2015", "2016",
                            "2017", "2018",
                            "2019", "2020", "2021")) +
                 labs(title = "Total Fatal Shootings Per Year",
                      x = "Year",
                      y = "Fatal Shootings per Year",
                      fill = 'Year') +
                 theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




Create line graphs that show fatal shooting rates on a month-by-month basis for each year.

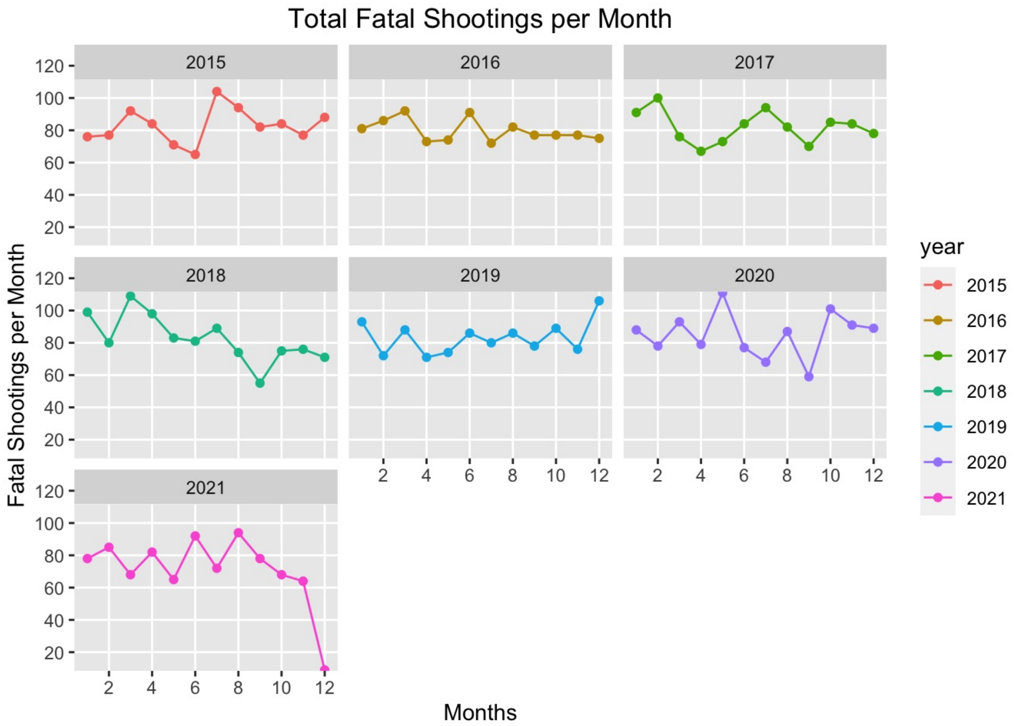
(Note: there is a steep drop off in the 2021 graph because of the missing values in December referenced above.)

```
monthly_freq <- police_data %>% mutate(year = year(date), month = month(date)) %>%
                                group_by(year, month) %>%
                                summarise(n = n())
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

```
monthly_freq %>% ggplot(aes(x = month, y = n, color = year)) +
                 geom_point(monthly_freq, mapping = aes(color = as.factor(year))) +
                 geom_line(monthly_freq, mapping = aes(color = as.factor(year))) +
                 facet_wrap(vars(year)) +
                 scale_x_discrete(limits = c(2, 4, 6, 8, 10, 12),
                 labels = c("2", "4", "6", "8", "10", "12")) +
                 scale_y_discrete(limits = c(20, 40, 60, 80, 100, 120),
                 labels = c("20", "40", "60", "80", "100", "120")) +
                 labs(title = "Total Fatal Shootings per Month",
                      x = "Months",
                      y = "Fatal Shootings per Month",
                      fill = 'Year') +
                 theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?

## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




Determine the mean, median, and range with respect to age for individuals who were fatally shot.

```
police_data %>% summarize(mean(age, na.rm = T),
                          median(age, na.rm = T),
                          range(age, na.rm = T))
```

```
## # A tibble: 2 x 3
##   `mean(age, na.rm = T)` `median(age, na.rm = T)` `range(age, na.rm = T)`
##                    <dbl>                    <dbl>                   <dbl>
## 1                   37.1                       35                       6
## 2                   37.1                       35                      92
```

```
left <- c(37.13, ' ')
middle <- c(35, ' ')
right <- c(6, 92)

table_value_2 <- data.frame (left, middle, right)


table_value_2 %>%
kbl(col.names = c('Mean', 'Median', 'Range'), align = c('c', 'c', 'c')) %>%
kable_classic_2(full_width = F)
```

| Mean | Median | Range |
|---|---|---|
| 37.13 | 35 | 6 |
| | | 92 |

Create a histogram that shows the age distribution of individuals who were fatally shot.

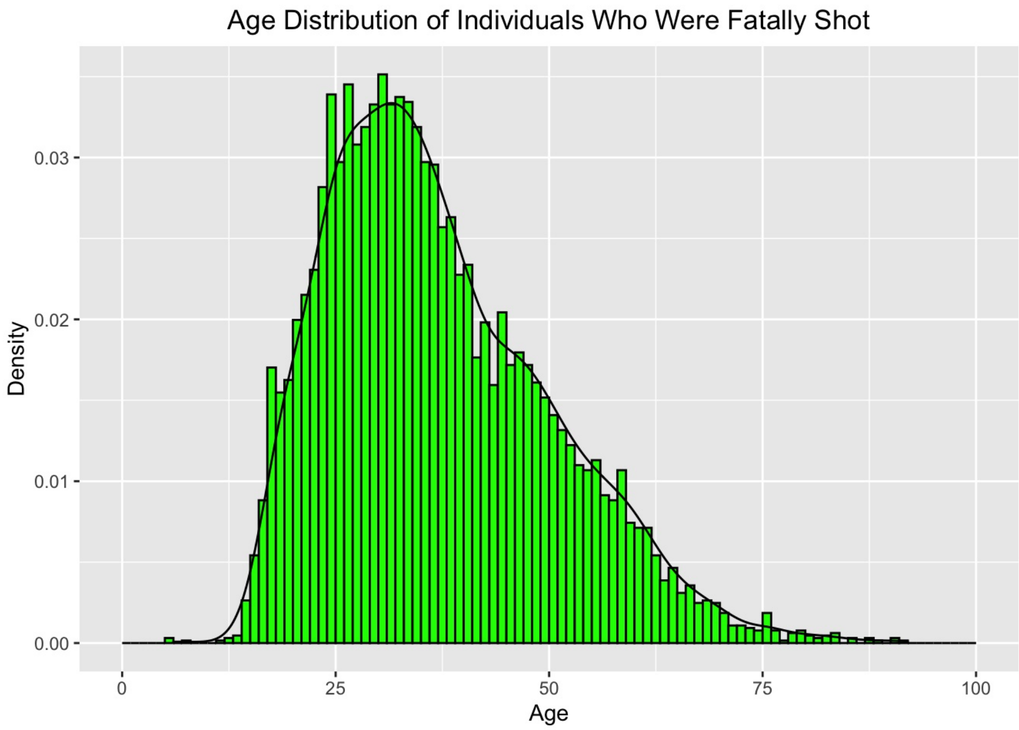
```
ggplot(police_data, aes(x = age)) +
geom_histogram (aes(y =..density..),
                breaks = seq(0, 100, by = 1),
                colour = "black",
                fill = "green") +
stat_function(fun = dnorm,
              args = list(mean = mean(police_data$age),
              sd = sd(police_data$age))) +
geom_density(bins = 100) +
labs(title = "Age Distribution of Individuals Who Were Fatally Shot",
     x = "Age",
     y = "Density") +
theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: bins
```

```
## Warning: Removed 339 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 339 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 101 row(s) containing missing values (geom_path).
```



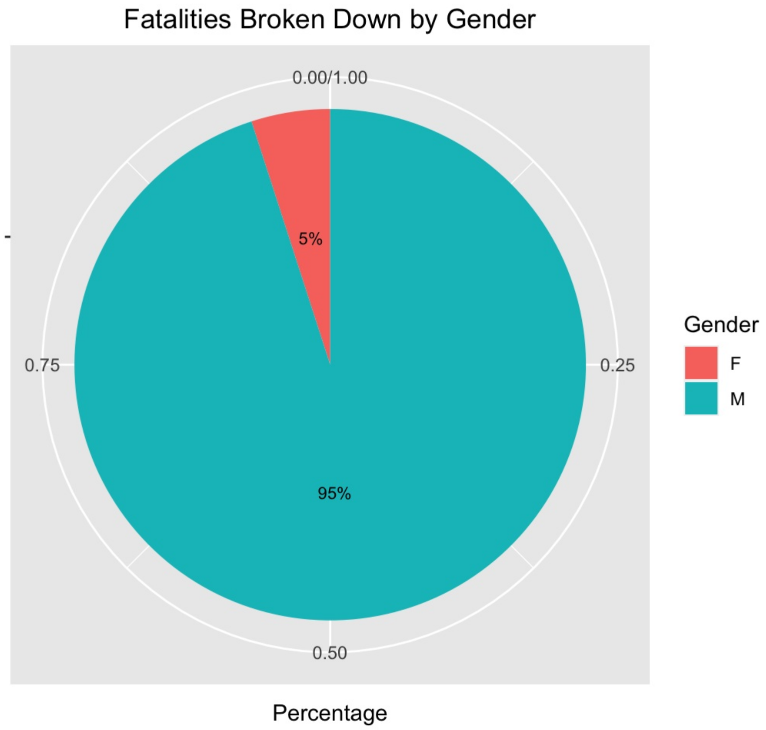

Compare the number of males who were fatally shot to the number of females who were fatally shot.

```
gender_breakdown <- police_data %>%
                    group_by(gender) %>%
                    drop_na() %>%
                    count() %>%
                    ungroup() %>%
                    mutate(perc = `n` / sum(`n`)) %>%
                    arrange(perc) %>%
                    mutate(labels = scales::percent(perc)) %>%
                    mutate_if(is.numeric, round, digits = 4)

gender_breakdown %>% kbl(col.names = c('Gender',
                                       'Total Fatal Shootings',
                                       'Decimal',
                                       'Rounded Percentage'),
                         align = c('c', 'c', 'c', 'c')) %>%
                         kable_classic_2(full_width = F)
```

| Gender | Total Fatal Shootings | Decimal | Rounded Percentage |
|---|---|---|---|
| F | 246 | 0.0498 | 5% |
| M | 4692 | 0.9502 | 95% |

Create a pie chart that compares the number of males who were fatally shot to the number of females who were fatally shot.

```
gender_breakdown %>% ggplot(aes(x= "", y = perc, fill = gender)) +
                     geom_bar(stat="identity", width=1) +
                     geom_text(aes(label = labels),
                                   position = position_stack(vjust = 0.52),
                                   cex = 2.9) +
                     coord_polar("y", start=0) +
                     labs(title = "Fatalities Broken Down by Gender",
                          x = ' ' ,
                          y = "Percentage",
                          fill = 'Gender') +
                     theme(plot.title = element_text(hjust = 0.5))
```



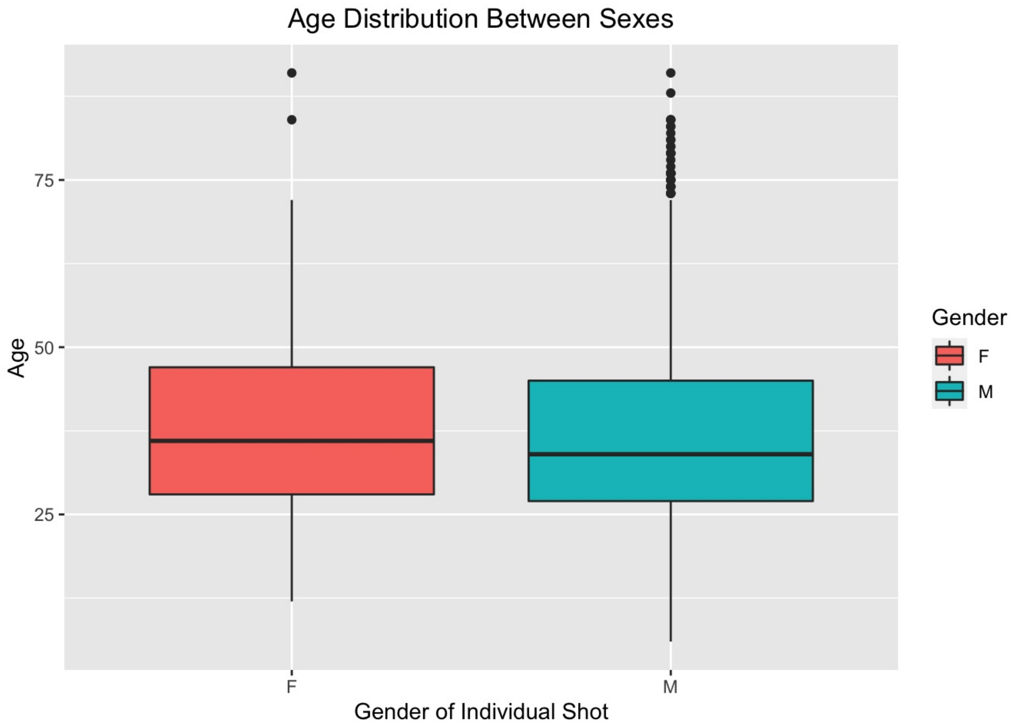

Determine the average age of men and women who were fatally shot.

```
table_value_4 <- police_data %>%
                 drop_na() %>%
                 group_by(gender) %>%
                 summarize(mean(age)) %>%
                 mutate_if(is.numeric, round, digits = 3)

table_value_4 %>%
  kbl(col.names = c('Gender', 'Average Age When Fatally Shot'),
      align = c('c', 'c')) %>%
  kable_classic_2(full_width = F)
```

| Gender | Average Age When Fatally Shot |
|---|---|
| F | 37.833 |
| M | 36.591 |

Create a box and whisker plot that incorporates data on age and gender.

```
police_data %>% drop_na() %>%
                ggplot(mapping = aes(x = gender, y = age, fill = gender)) +
                geom_boxplot() +
                labs(title = "Age Distribution Between Sexes",
                          x = 'Gender of Individual Shot' ,
                          y = 'Age',
                          fill = 'Gender') +
                theme(plot.title = element_text(hjust = 0.5))
```



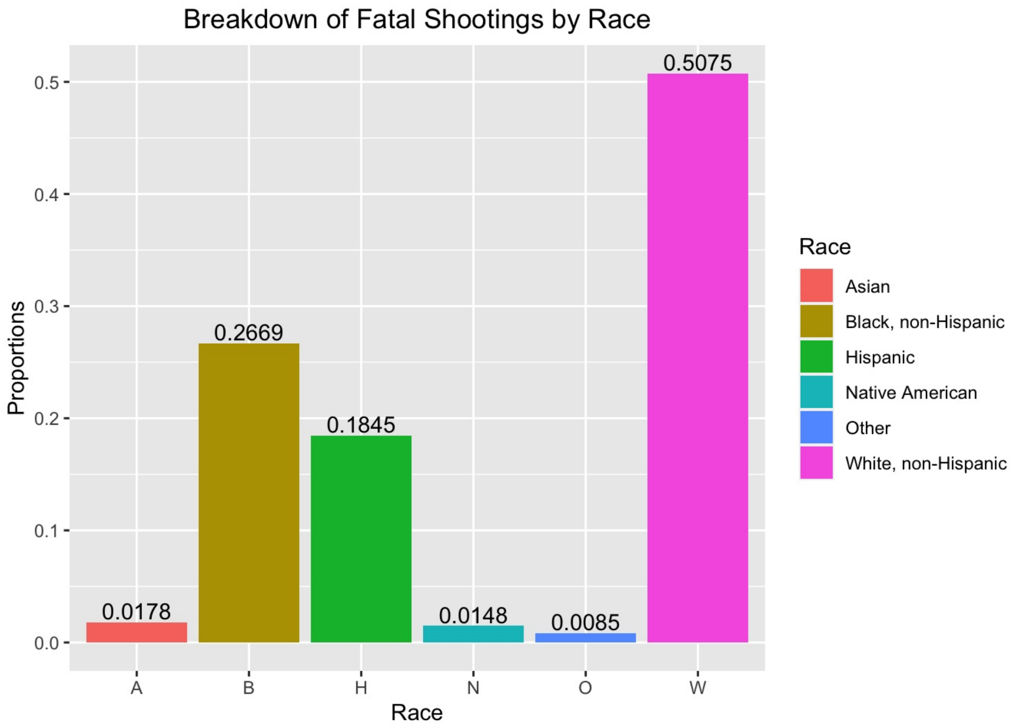

Determine the breakdown of race among individuals who were fatally shot.

```
race_breakdown <- police_data %>%
                  mutate(race_string =
                         case_when(race == 'W' ~ 'White, non-Hispanic',
                                   race == 'B' ~ 'Black, non-Hispanic',
                                   race == 'A' ~ 'Asian',
                                   race == 'N' ~ 'Native American',
                                   race == 'H' ~ 'Hispanic',
                                   race == 'O' ~ 'Other'))


table_value_5 <- race_breakdown %>% group_by(race_string) %>%
                 drop_na() %>%
                 count() %>%
                 ungroup() %>%
                 mutate(perc = `n` / sum(`n`)) %>%
                 arrange(perc) %>%
                 mutate(rounded = scales::percent(perc)) %>%
                 mutate_if(is.numeric, round, digits = 4) %>%
                 arrange(desc(n))


table_value_5 %>% kbl(col.names = c('Race',
                                    'Total Fatal Shootings',
                                    'Decimal',
                                    'Unrounded Percentage'),
                      align = c('c', 'c')) %>%
                      kable_classic_2(full_width = F)
```

| Race | Total Fatal Shootings | Decimal | Unrounded Percentage |
|---|---|---|---|
| White, non-Hispanic | 2506 | 0.5075 | 50.75% |
| Black, non-Hispanic | 1318 | 0.2669 | 26.69% |
| Hispanic | 911 | 0.1845 | 18.45% |
| Asian | 88 | 0.0178 | 1.78% |
| Native American | 73 | 0.0148 | 1.48% |
| Other | 42 | 0.0085 | 0.85% |

Create a bar graph that shows the shooting breakdown by race.

```
ggplot() +
geom_col (table_value_5, mapping = aes(x = race_string,
                                       y = perc, fill = as.factor(race_string))) +
geom_text(table_value_5, mapping = aes(x = race_string,
                                       y = perc, label = perc,
                                       vjust = -.2)) +
scale_x_discrete(limits = c("Asian", "Black, non-Hispanic",
                            "Hispanic", "Native American",
                            "Other", "White, non-Hispanic"),
                 labels = c("A", "B",
                            "H", "N",
                            "O", "W")) +
labs(title = "Breakdown of Fatal Shootings by Race",
     x = "Race",
     y = "Proportions",
     fill = 'Race') +
theme(plot.title = element_text(hjust = 0.5))
```



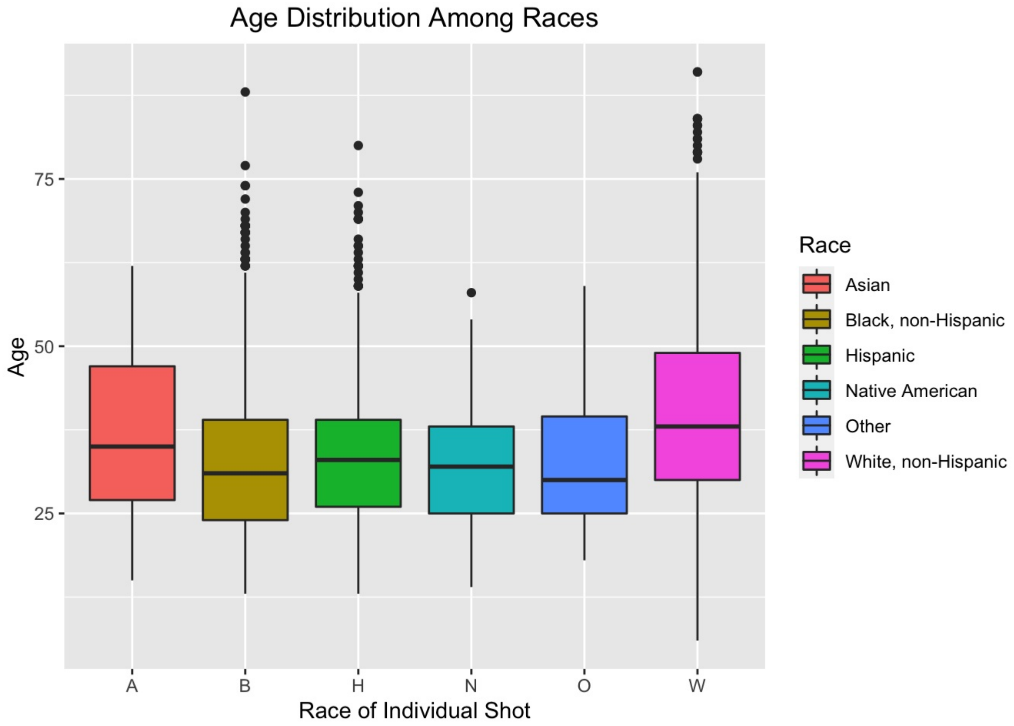

Determine the average age among individuals who were shot, according to race.

```
table_value_6 <- race_breakdown %>%
                 drop_na() %>%
                 group_by(race_string) %>%
                 summarize(mean_age = mean(age)) %>%
                 mutate_if(is.numeric, round, digits = 3) %>%
                 arrange(desc(mean_age))

table_value_6 %>% kbl(col.names = c('Race', 'Average Age When Fatally Shot'),
                      align = c('l', 'c')) %>% kable_classic_2(full_width = F)
```

| Race | Average Age When Fatally Shot |
|---|---|
| White, non-Hispanic | 40.019 |
| Asian | 36.716 |
| Hispanic | 33.704 |
| Other | 32.762 |
| Native American | 32.699 |
| Black, non-Hispanic | 32.631 |

Create a box and whisker plot that incorporates data on age and race.

```
race_breakdown %>%
drop_na() %>%
ggplot(mapping = aes(x = race_string, y = age, fill = race_string)) +
geom_boxplot() +
scale_x_discrete(limits = c("Asian","Black, non-Hispanic", "Hispanic",
                              "Native American", "Other", "White, non-Hispanic"),
                   labels = c("A", "B","H", "N", "O", "W")) +
                   labs(title = "Age Distribution Among Races",
                        x = 'Race of Individual Shot' ,
                        y = 'Age',
                        fill = 'Race') +
theme(plot.title = element_text(hjust = 0.5))
```




A few more interesting tables.

Five states with the most shootings from 2015 - 2021.

```
top_5 <- police_data %>%
                  drop_na() %>%
                  group_by(state) %>%
                  summarize(totals = n()) %>%
                  mutate(frequency = totals / sum(totals)) %>%
                  mutate_at(vars(starts_with("frequency")), funs(round(., 4))) %>%
                  arrange(desc(totals)) %>%
                  head(5)
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
```

```
top_5 %>%
kbl(col.names = c('Top States', 'Total Fatal Shootings', 'Proportions'),
    align = c('l', 'c')) %>%
kable_classic_2(full_width = F)
```

| Top States | Total Fatal Shootings | Proportions |
|---|---|---|
| CA | 730 | 0.1478 |
| TX | 422 | 0.0855 |
| FL | 343 | 0.0695 |
| AZ | 214 | 0.0433 |

| GA | 179 | 0.0362 |
|---|---|---|

Five states with the least shootings from 2015 - 2021.

```
bottom_5 <- police_data %>%
                  drop_na() %>%
                  group_by(state) %>%
                  summarize(totals = n()) %>%
                  mutate(frequency = totals / sum(totals)) %>%
                  mutate_at(vars(starts_with("frequency")), funs(round(., 4))) %>%
                  arrange(totals) %>%
                  head(5)

bottom_5 %>%
kbl(col.names = c('Bottom States', 'Total Fatal Shootings', 'Proportions'),
    align = c('l', 'c')) %>%
kable_classic_2(full_width = F)
```

| Bottom States | Total Fatal Shootings | Proportions |
|---|---|---|
| RI | 2 | 0.0004 |
| VT | 7 | 0.0014 |
| ND | 9 | 0.0018 |
| DE | 11 | 0.0022 |
| SD | 12 | 0.0024 |

Breakdown of shootings that occurred when individuals who were shot were armed vs. unarmed.

```
armed_vs_unarmed <- police_data %>% drop_na(armed) %>%
                                    mutate(armed_vs_unarmed = ifelse (
                                    armed == 'unarmed', 'Unarmed', 'Armed')) %>%
                                    group_by(armed_vs_unarmed) %>%
                                    summarize (n = n()) %>%
                                    mutate(freq = n / sum(n)) %>%
                                    mutate_at(vars(starts_with("freq")), funs(round(., 4)))


armed_vs_unarmed %>%
kbl(col.names = c('Armed vs. Unarmed', 'Total Fatal Shootings', 'Proportions'),
    align = c('l', 'c')) %>%
kable_classic_2(full_width = F)
```

| Armed vs. Unarmed | Total Fatal Shootings | Proportions |
|---|---|---|
| Armed | 6166 | 0.9351 |
| Unarmed | 428 | 0.0649 |

Breakdown of shootings that occurred with and without a police body camera.

```
body_camera <- police_data %>%
               group_by(body_camera) %>%
               summarize (n = n()) %>%
               mutate(freq = n / sum(n)) %>%
               mutate_at(vars(starts_with("freq")), funs(round(., 4)))


body_camera %>%
kbl(col.names = c('Wearing Body Camera', 'Total Fatal Shootings', 'Proportions'),
    align = c('l', 'c')) %>%
kable_classic_2(full_width = F)
```

| Wearing Body Camera | Total Fatal Shootings | Proportions |
|---|---|---|
| FALSE | 5851 | 0.8604 |
| TRUE | 949 | 0.1396 |