Exploring the Washington Post Data

Packages

```
library(here)
library(readr)
library(dplyr)
library(purrr)
library(tidyr)
library(ggplot2)
library(lubridate)
```

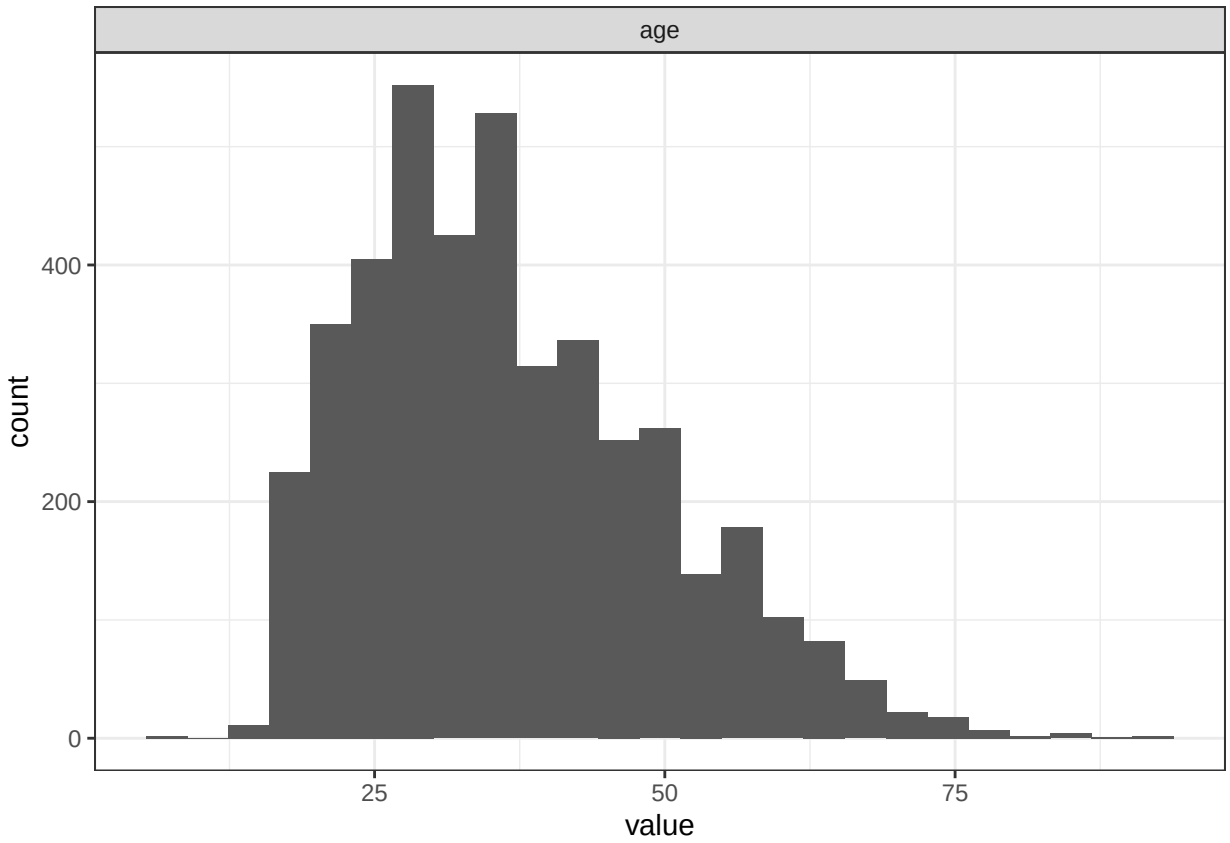
Read in data

```
load(here("create_test_train", "output", "raceTrain.RData"))
washpo <- raceTrain[["washpoDgHM"]]
```

Distribution of victim age

```
washpo %>%
    select(age) %>%
    pivot_longer(everything()) %>%
    ggplot(aes(x = value)) +
    geom_histogram(bins = 25) +
    facet_wrap(~name, scales = "free_x") +
    theme_bw()
```

```
## Warning: Removed 77 rows containing non-finite values (stat_bin).
```




```
summary(washpo$age)
```

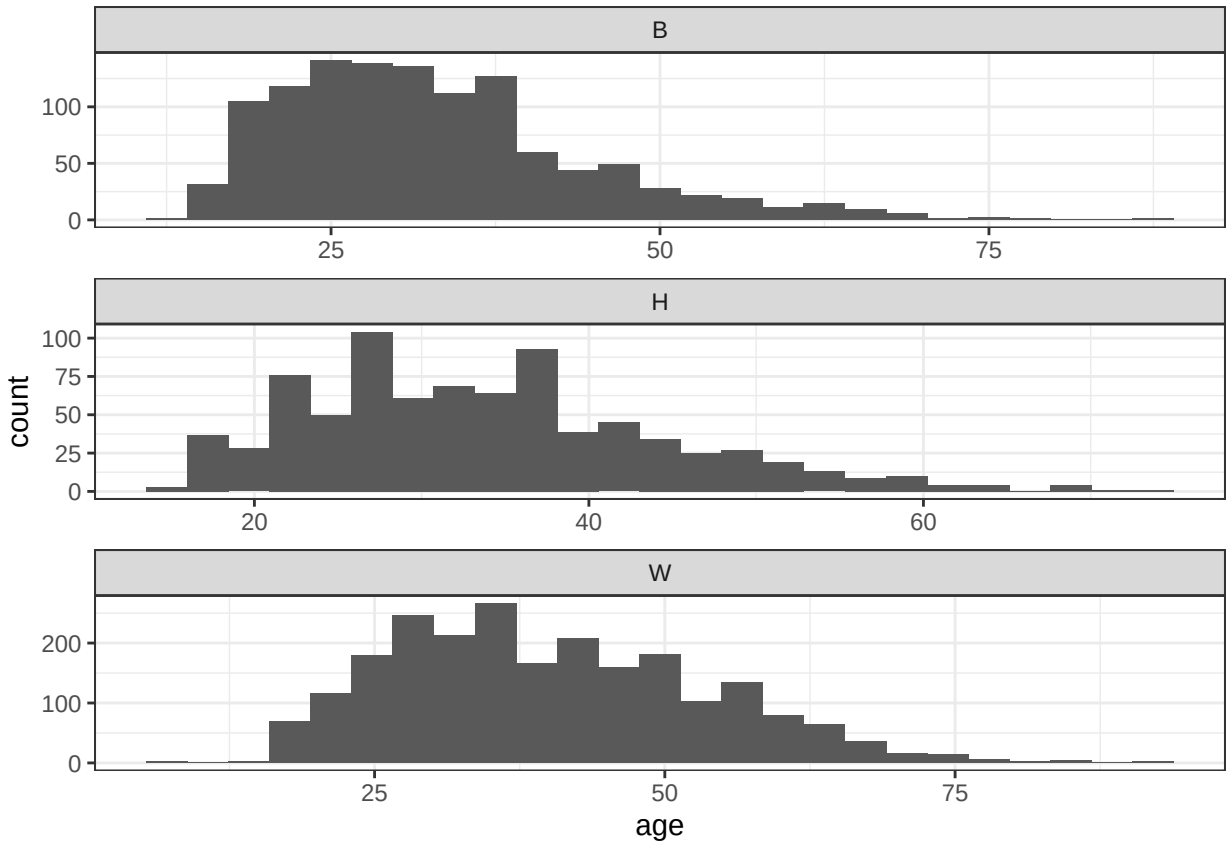
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   27.00   35.00   36.83   45.00   91.00      77
```

Distribution of victim age by race

- Black American and Hispanic victims seem to be younger.

```
washpo %>%
    select(age, race) %>%
    filter(race %in% c("W", "B", "H")) %>%
    ggplot(aes(x = age)) +
    geom_histogram(bins = 25) +
    facet_wrap(~race, scales = "free", ncol = 1) +
    theme_bw()
```

```
## Warning: Removed 77 rows containing non-finite values (stat_bin).
```



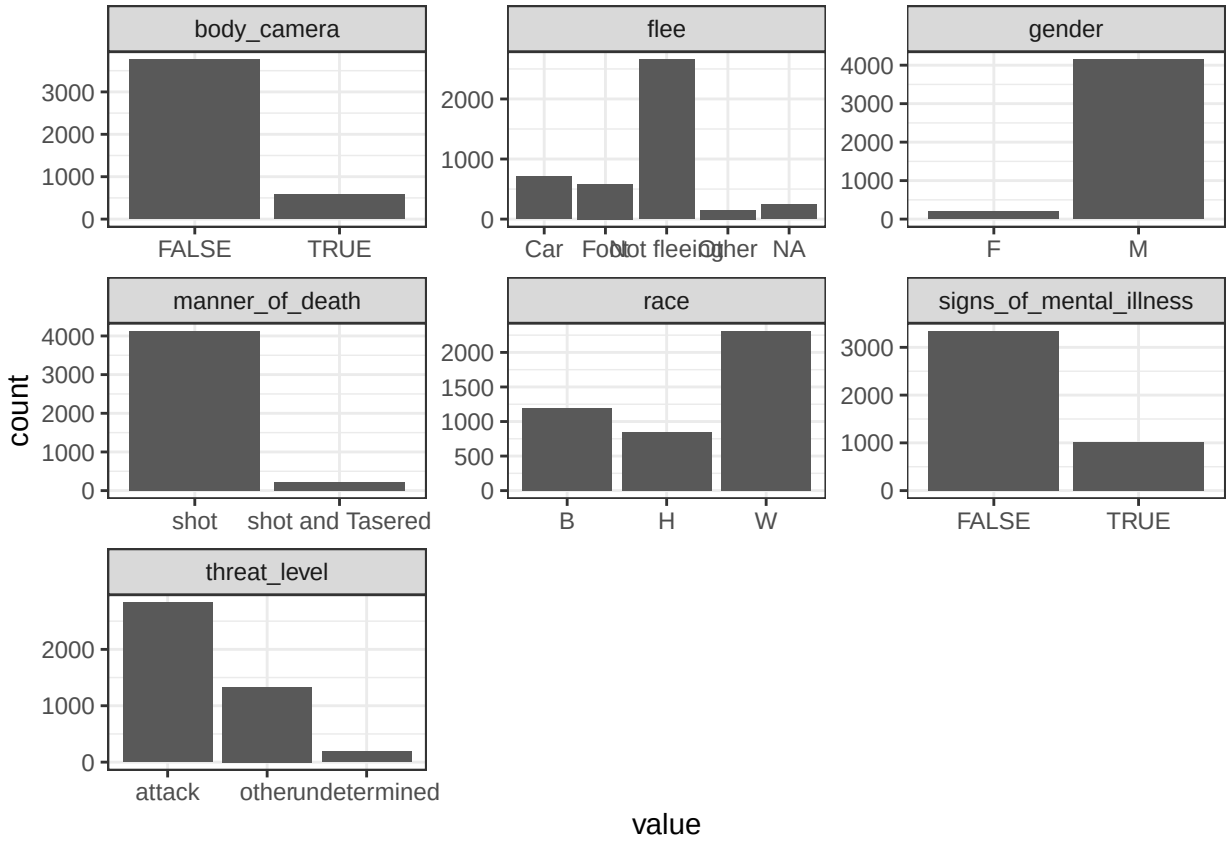

```
map(c("W", "B", "H"),
    function(df, raceCat) {
        df <- df %>% filter(race == raceCat)
        print(raceCat)
        summary(df$age)},
    df = washpo)
```

```
## [1] "W"
## [1] "B"
## [1] "H"
```

```
## [[1]]
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   30.00   38.00   40.03   49.00   91.00      36
##
## [[2]]
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   13.00   24.00   31.00   32.74   39.00   88.00      20
##
## [[3]]
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   14.00   26.00   33.00   33.83   40.00   73.00      21
```

Bar charts for categorical outcomes (not state, armedNew, or date)

```
washpo %>%
    select(c(!where(is.numeric), -armedNew, -state, -day, -month, -year)) %>%
    pivot_longer(everything(),
                 values_transform = list(value = as.character)) %>%
    ggplot(aes(x = value)) +
    geom_bar() +
    facet_wrap(~name, scales = "free") +
    theme_bw()
```



Bar charts for categorical outcomes by race (not state, armed, or date)

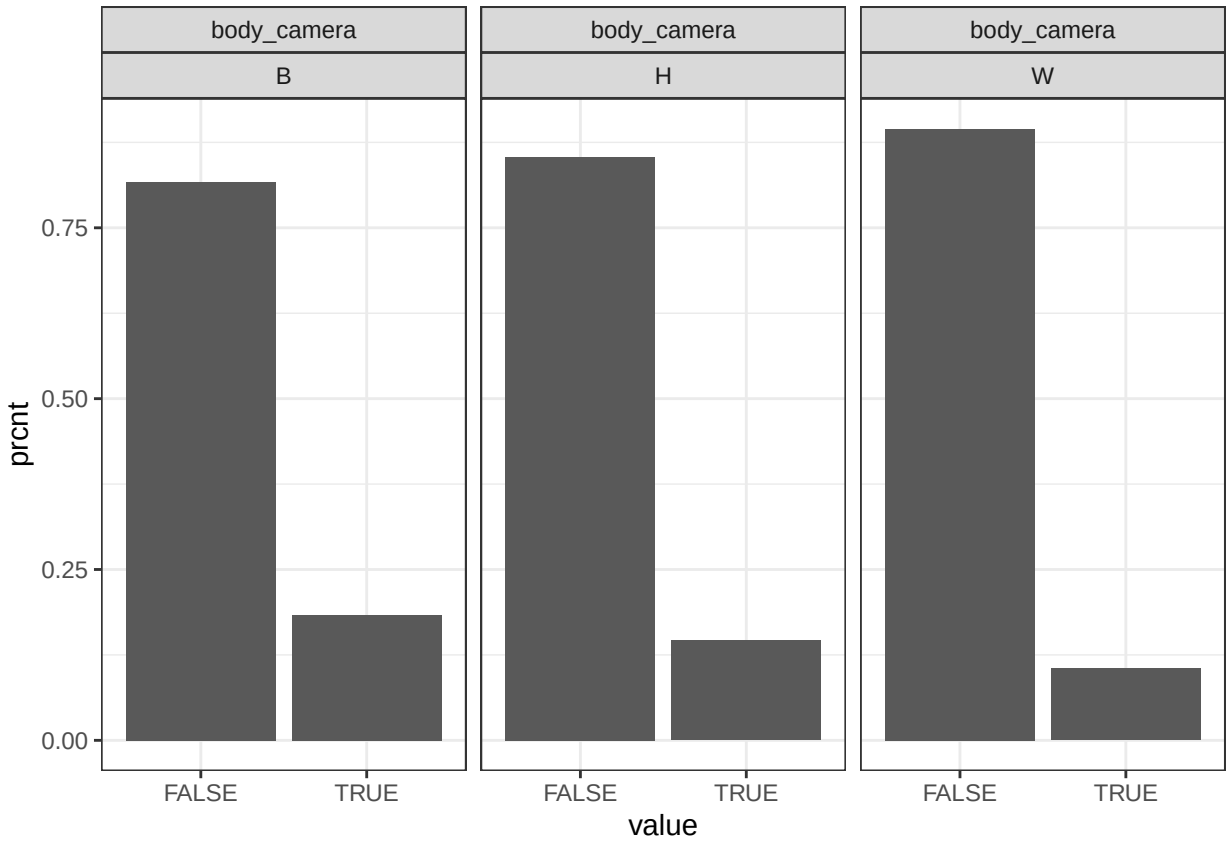
- Slightly more instances in which body cameras were used for Black and Hispanic victims vs. White victims.
- White victims slightly more likely to not be fleeing vs. Black and Hispanic victims.
- White victims slightly, slightly, less likely to be male vs. Black and Hispanic victims.
- Manner of death is universally the same.
- White victims moderately more likely to have a mental health history or signs of mental illness vs. Black and Hispanic victims.
- Similar levels of threat presented by race however the types of threat vary slightly across race.

```
categoriesRace <-
    washpo %>%
    select(c(!where(is.numeric), -armedNew, -state, -day, -month, -year)) %>%
    pivot_longer(-race,
                 values_transform = list(value = as.character)) %>%
    filter(race %in% c("W", "H", "B")) %>%
```

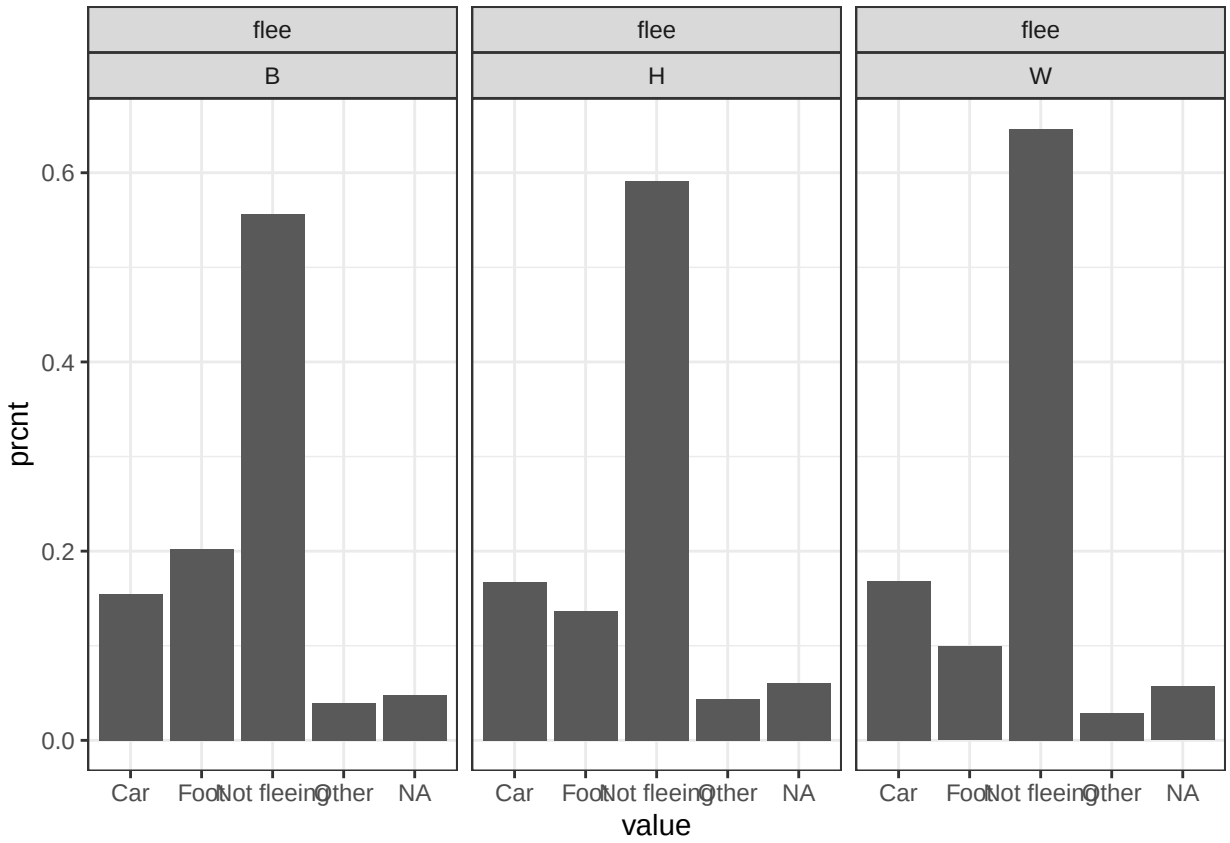
```
    count(race, name, value) %>%
    group_by(race, name) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup()

map(unique(categoriesRace$name),
    function(df, column) {
        df %>%
            filter(name == column) %>%
            ggplot(aes(x = value, y = prcnt)) +
            geom_bar(stat = "identity") +
            facet_wrap(~name+race, scales = "free_x") +
            theme_bw()},
    df = categoriesRace)
```

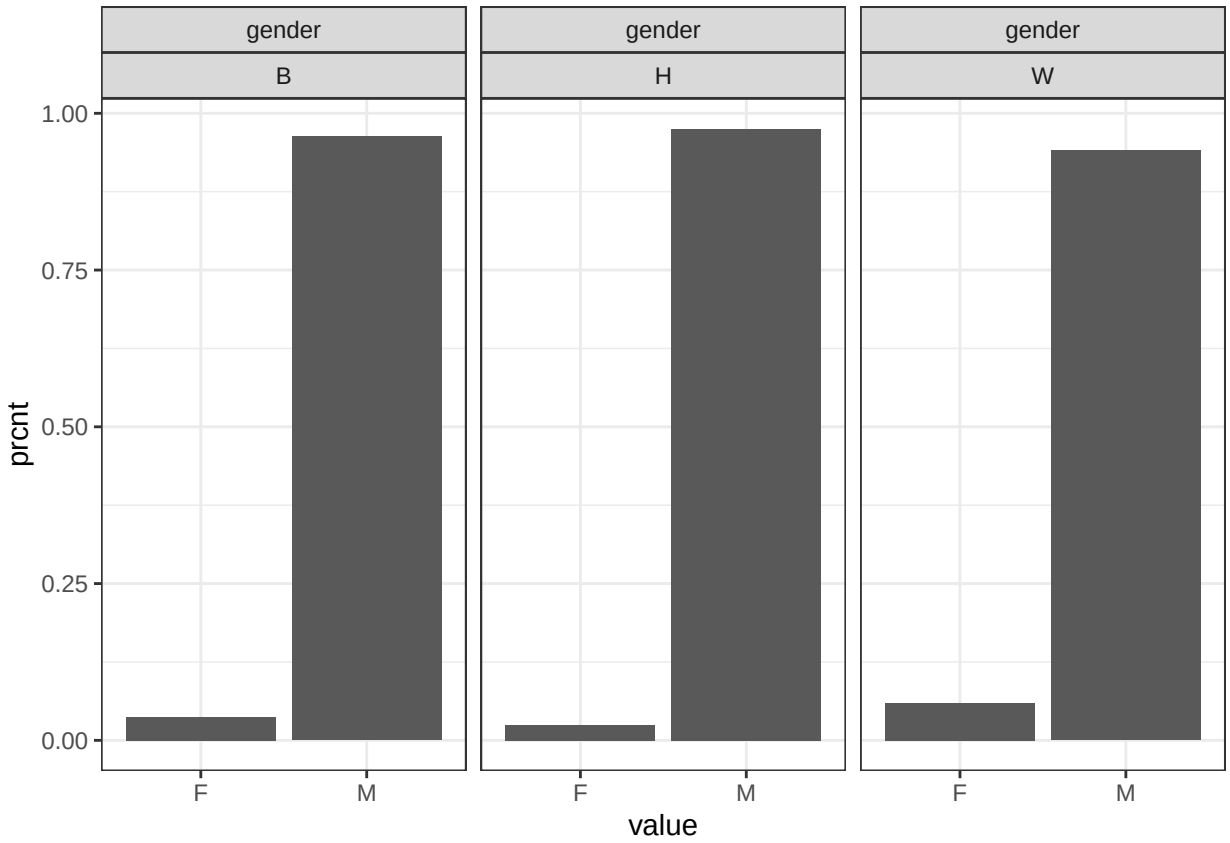
```
## [[1]]
```



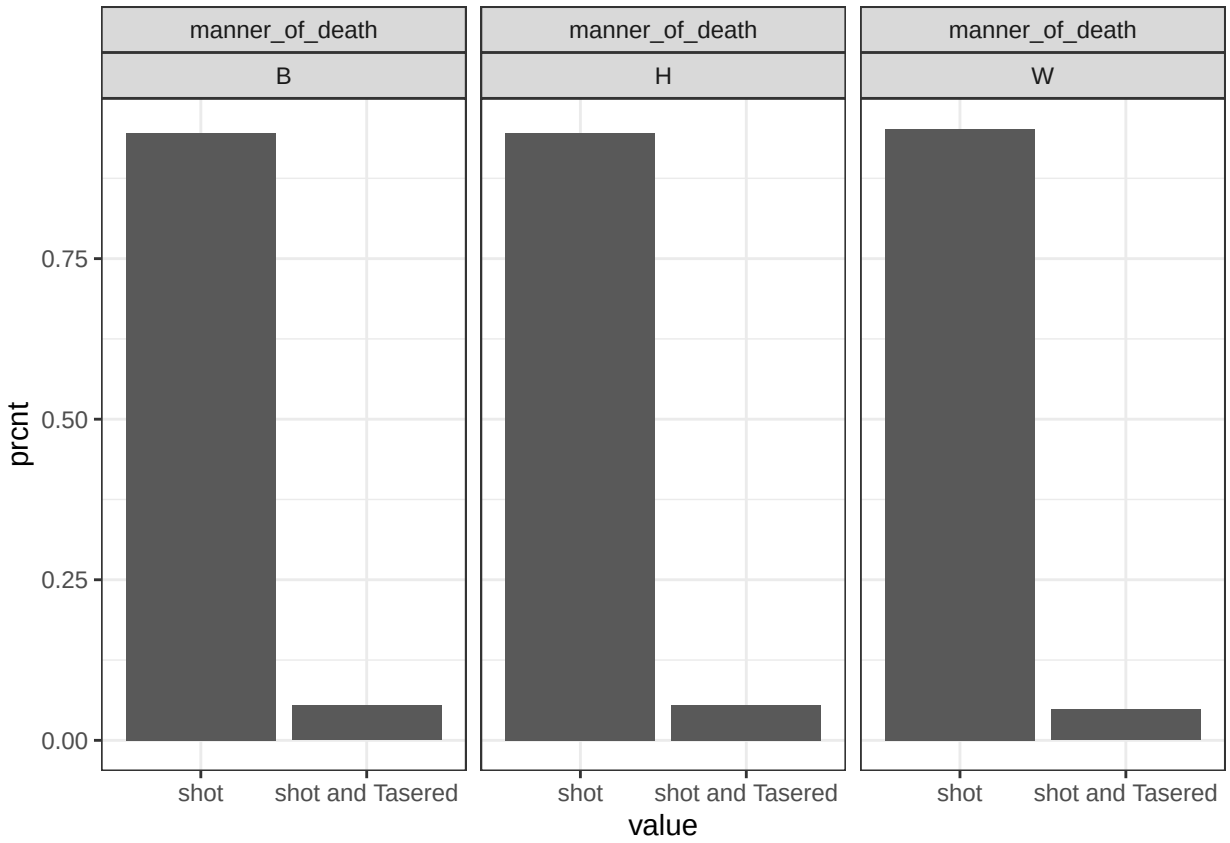

```
##
## [[2]]
```



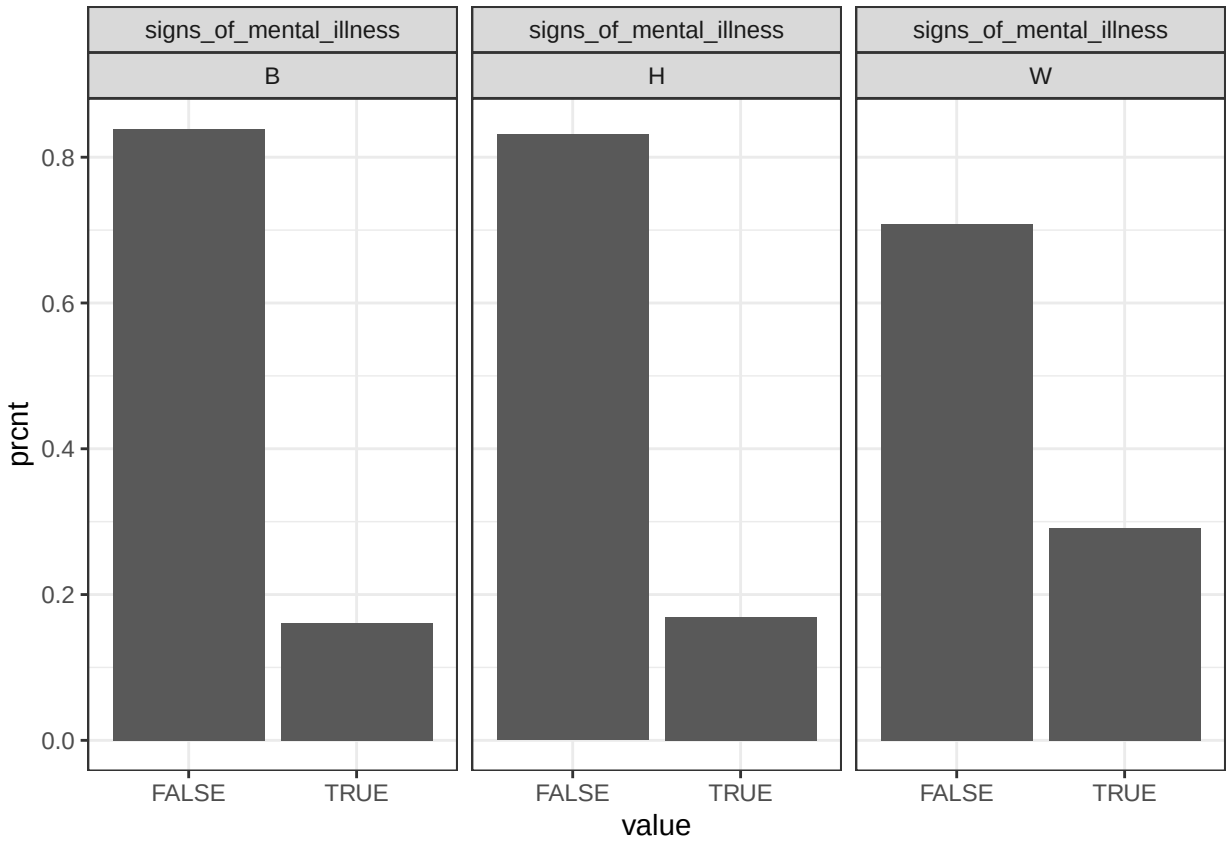

```
##
## [[3]]
```


gender
B
gender
H
gender
W
1.00
0.75
0.50
0.25
0.00
prcnt
F
M
F
M
F
M
value

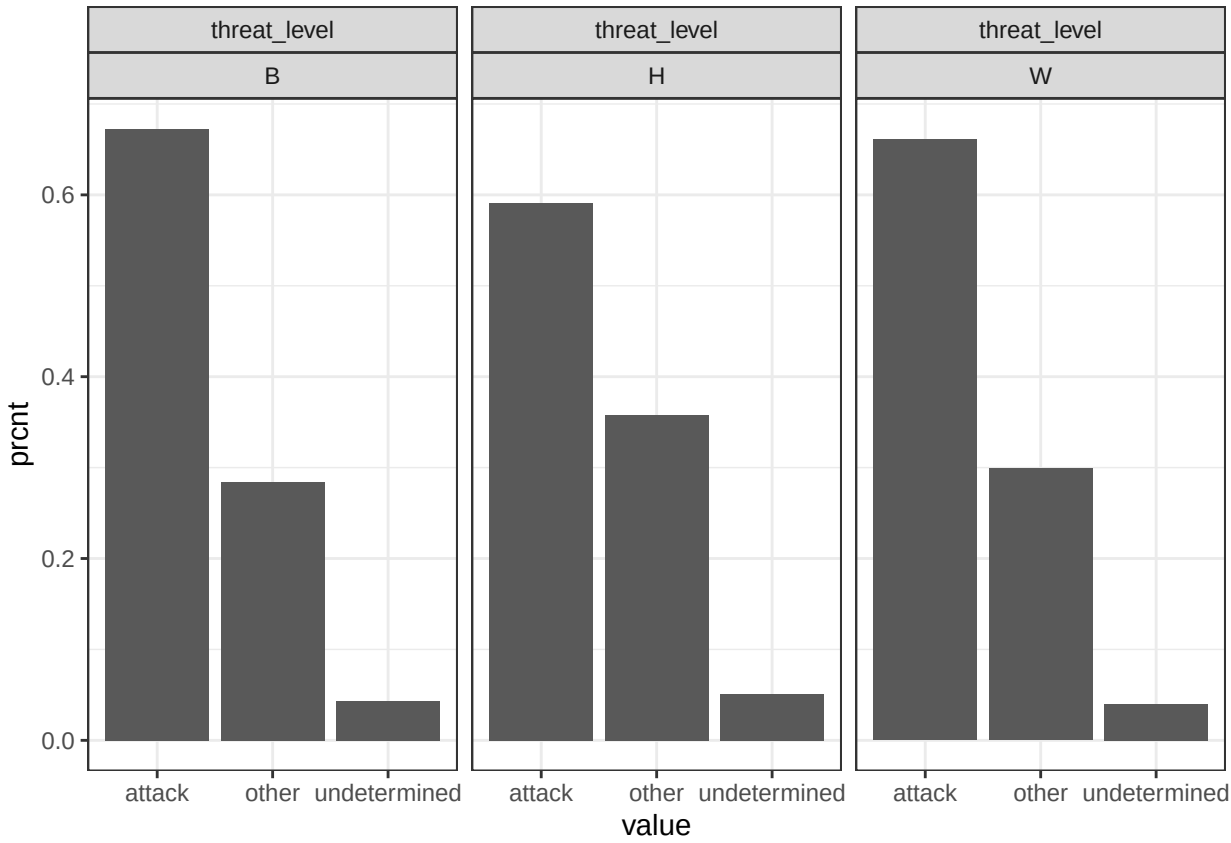
```
##
## [[4]]
```


manner_of_death
B
manner_of_death
H
manner_of_death
W
prcnt
0.75
0.50
0.25
0.00
shot
shot and Tasered
shot
shot and Tasered
shot
shot and Tasered
value

```
##
## [[5]]
```

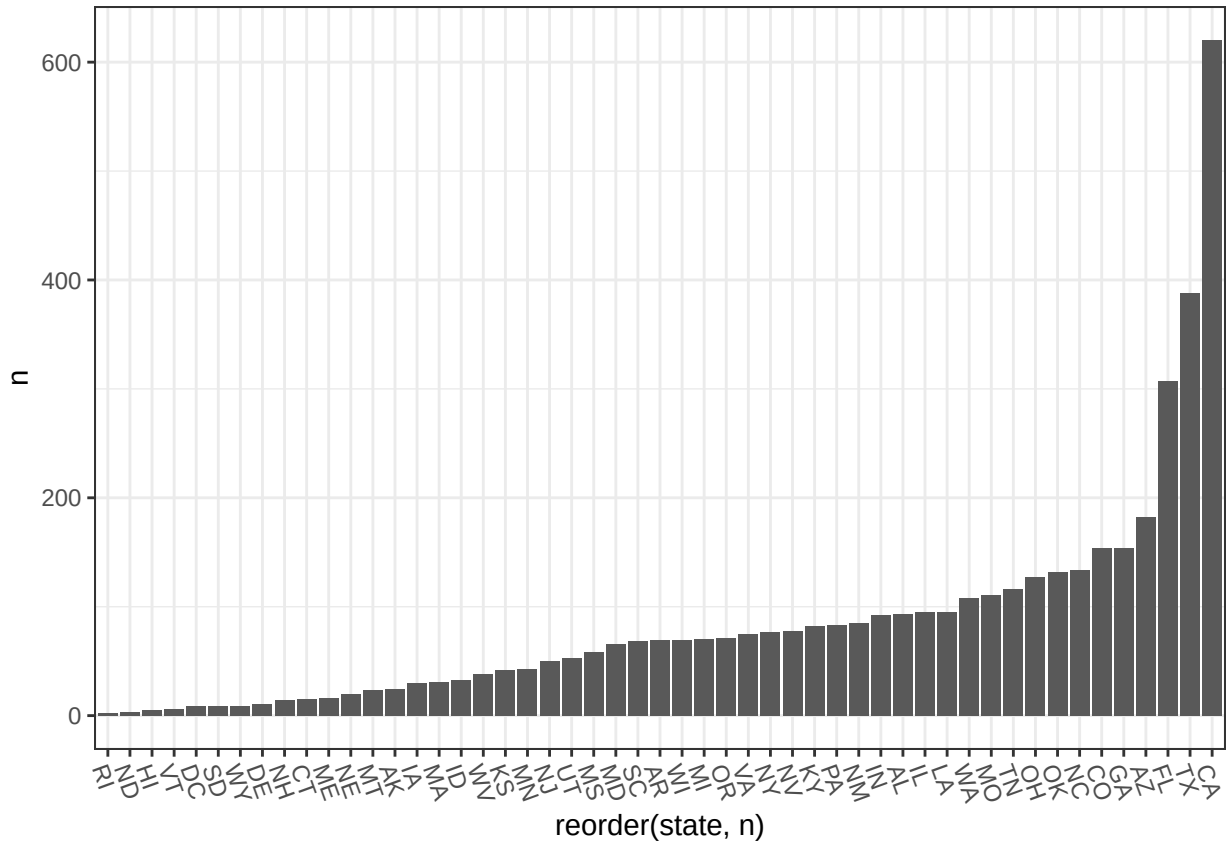

signs_of_mental_illness
B
signs_of_mental_illness
H
signs_of_mental_illness
W
prcnt
0.0
0.2
0.4
0.6
0.8
FALSE
TRUE
FALSE
TRUE
FALSE
TRUE
value

```
##
## [[6]]
```



Distribution of victims by state

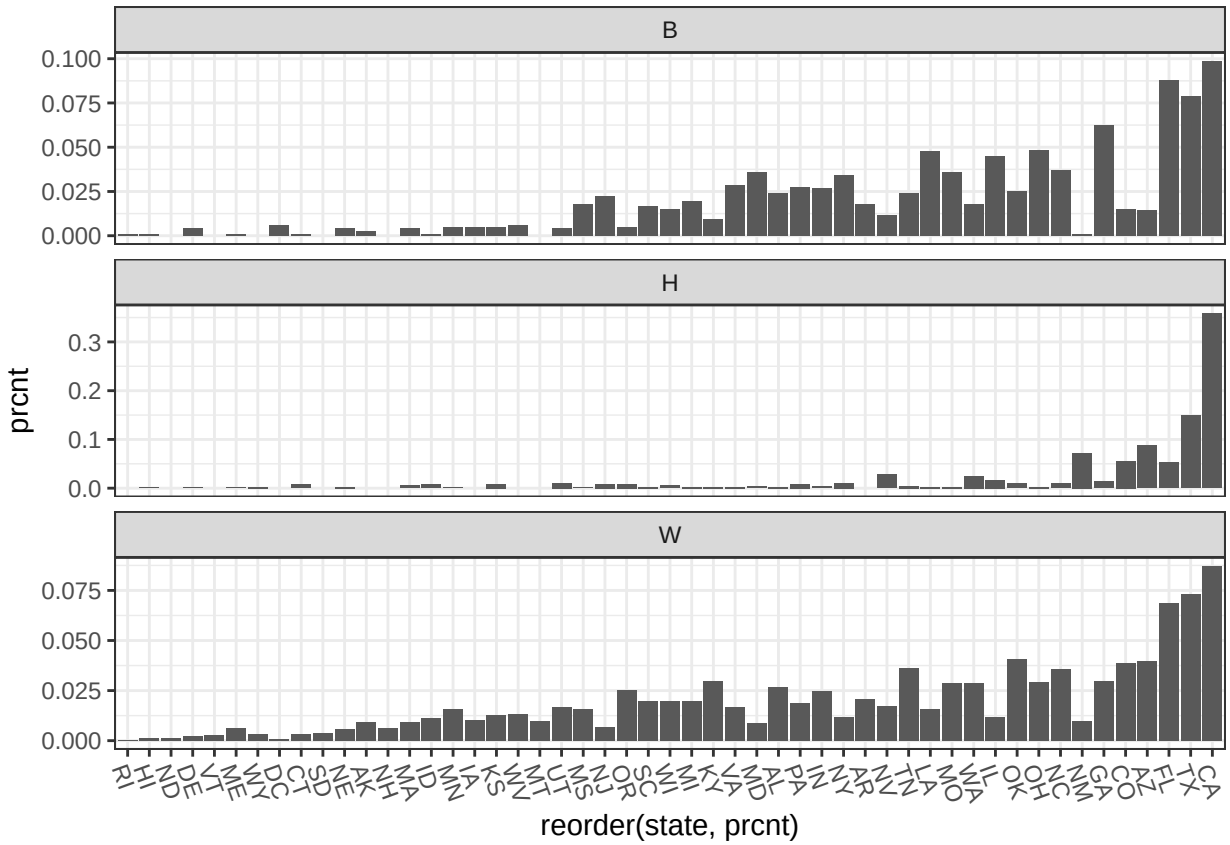
```
washpo %>%
    select(state) %>%
    group_by(state) %>%
    count(state) %>%
    ggplot(aes(x = reorder(state, n), n)) +
    geom_bar(stat = "identity") +
    theme_bw() +
    theme(axis.text.x = element_text(angle = 290, hjust = 0))
```



Distribution of victims by state by race

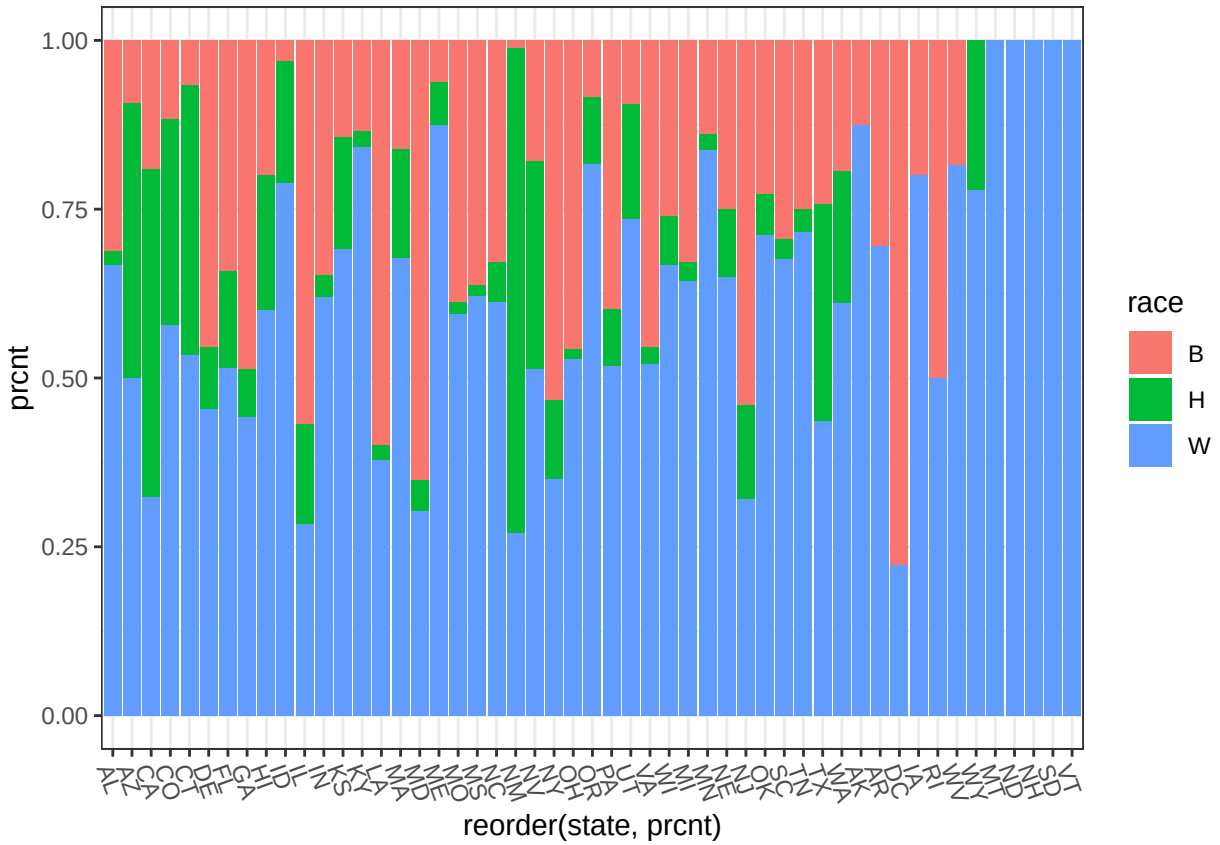
- Within each racial category, what proportion of killings took place in each state? E.g. of all Black Americans killed by the police, 10% were killed in CA. Of all White Americans killed by the police, 8& were killed in California. Of all Hispanics killed by the police, over 30% were killed in CA.
- Hard to discern any immediate trends. The biggest thing which stands out is the fact that Hispanics are overwhelmingly killed in only a small proportion of stats. Possibly tied to where Hispanic populations are largest.

```
washpo %>%
    select(state, race) %>%
    filter(race %in% c("W", "B", "H")) %>%
    count(state, race) %>%
    group_by(race) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(state, prcnt), y = prcnt)) +
    geom_bar(stat = "identity") +
    facet_wrap(~race, scales = "free_y", ncol = 1) +
    theme_bw() +
    theme(axis.text.x = element_text(angle = 290, hjust = 0))
```



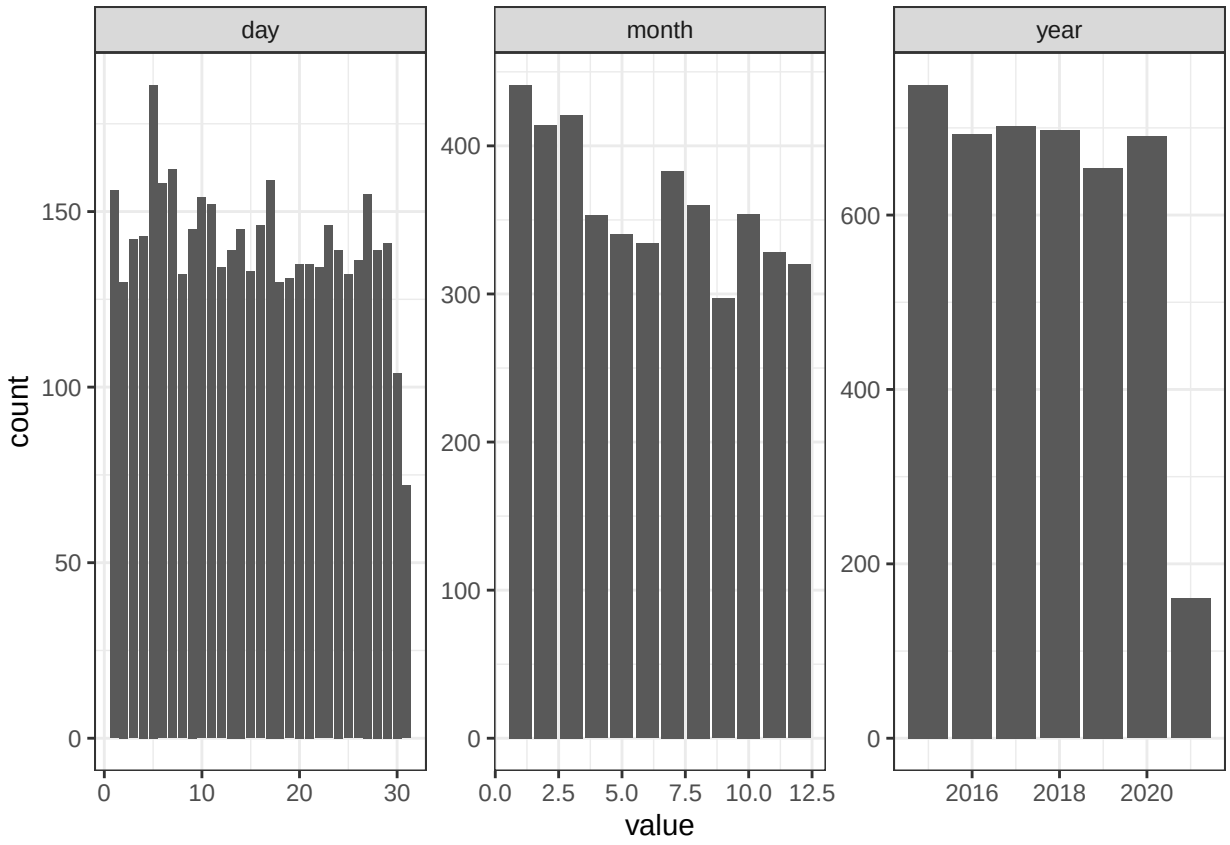
- Within each state, what proportion of killings were for each race? E.g. of all Americans killed by the police in Alabama, 25% were Black, 1% were Hispanic, and 70% were White.
- Hard to discern any immediate trends. I need population totals to get a sense of the disproportionality.

```
washpo %>%
    select(state, race) %>%
    filter(race %in% c("W", "B", "H")) %>%
    count(state, race) %>%
    group_by(state) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(state, prcnt), y = prcnt)) +
    geom_bar(stat = "identity", aes(fill = race)) +
    theme_bw() +
    theme(axis.text.x = element_text(angle = 290, hjust = 0))
```



Number of killings by day, month, year

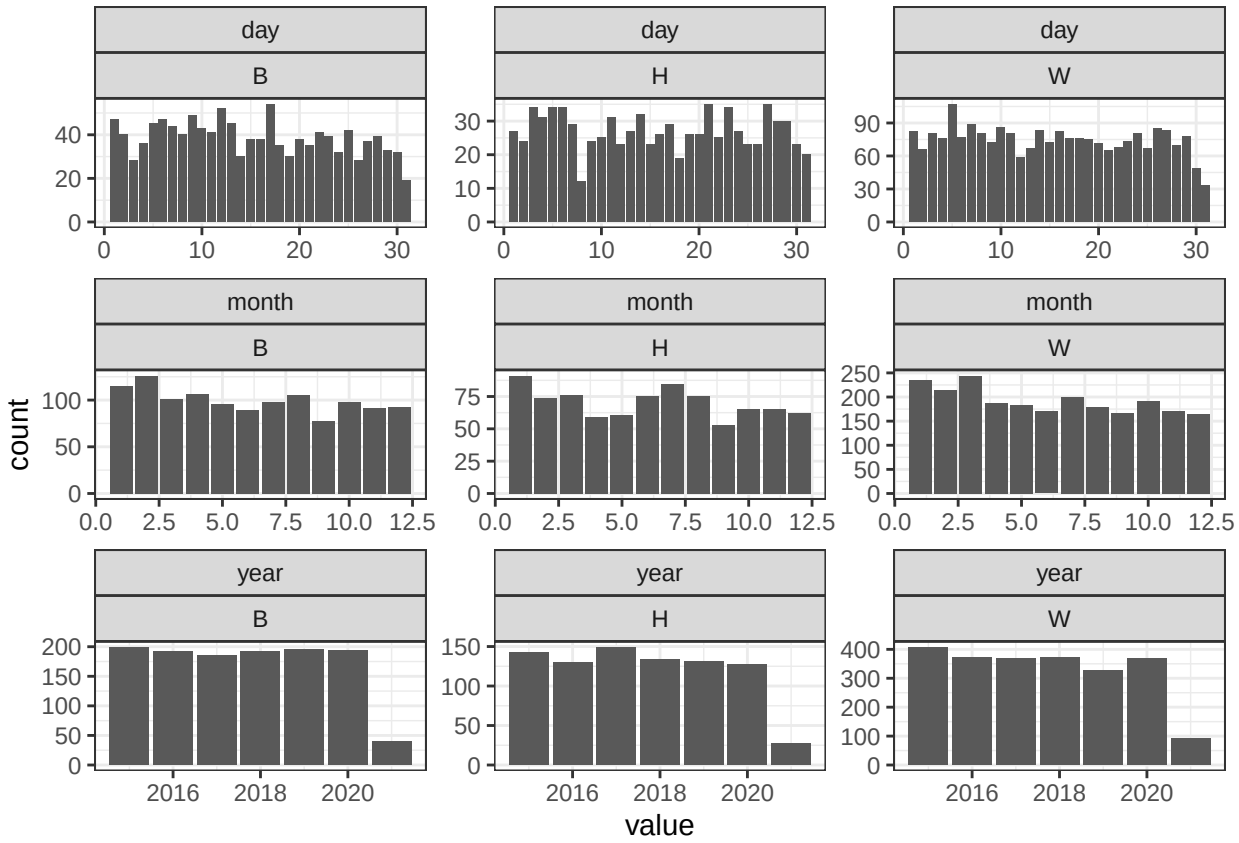
```
washpo %>%
    select(day, month, year) %>%
    pivot_longer(everything()) %>%
    ggplot(aes(x = value)) +
    geom_bar() +
    facet_wrap(~name, scales = "free") +
    theme_bw()
```



Number of killings by day, month, year and by race

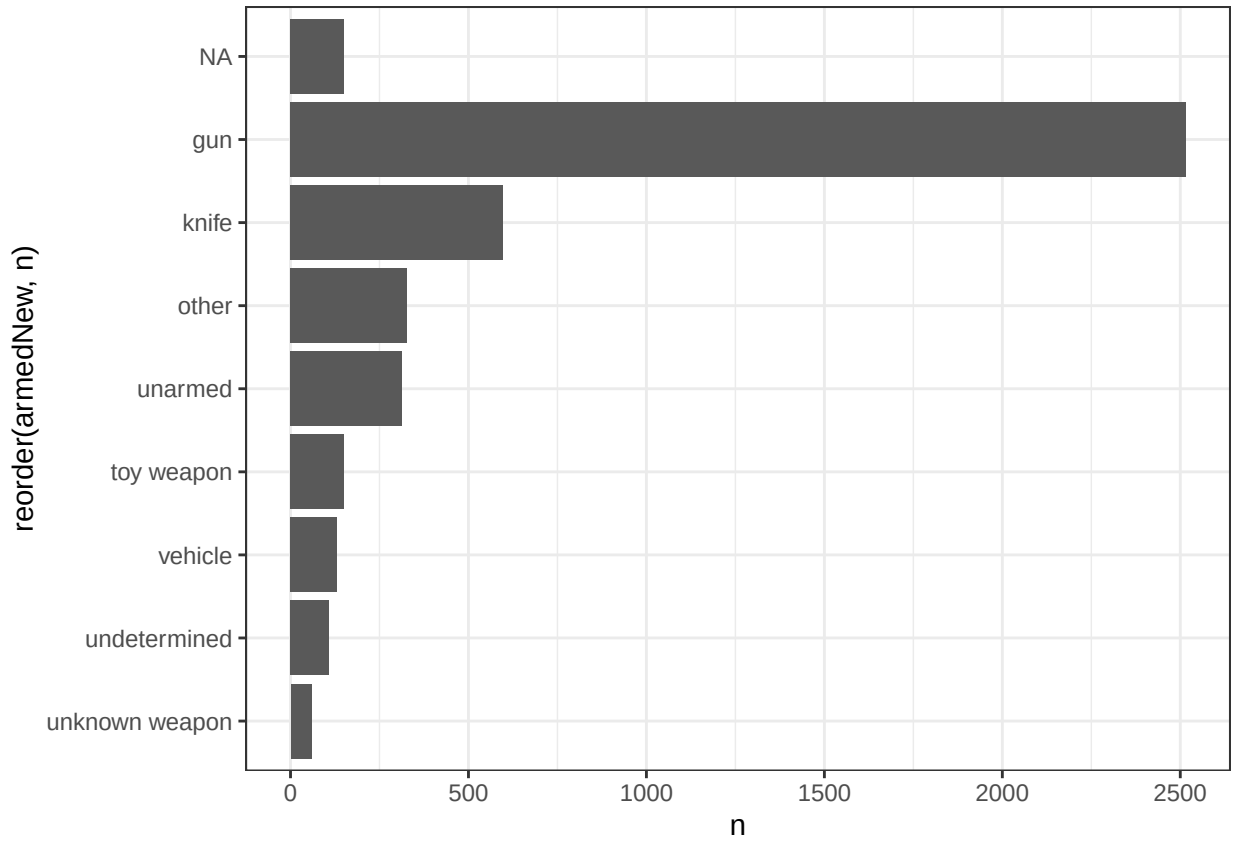
I wasn't expecting there to be any disparity by date and race, but it sure would've been interesting if there was disparity.

```
washpo %>%
    select(day, month, year, race) %>%
    pivot_longer(-race) %>%
    filter(race %in% c("W", "B", "H")) %>%
    ggplot(aes(x = value)) +
    geom_bar() +
    facet_wrap(~name+race, scales = "free") +
    theme_bw()
```



Distribution of victims by if they were aremed or not

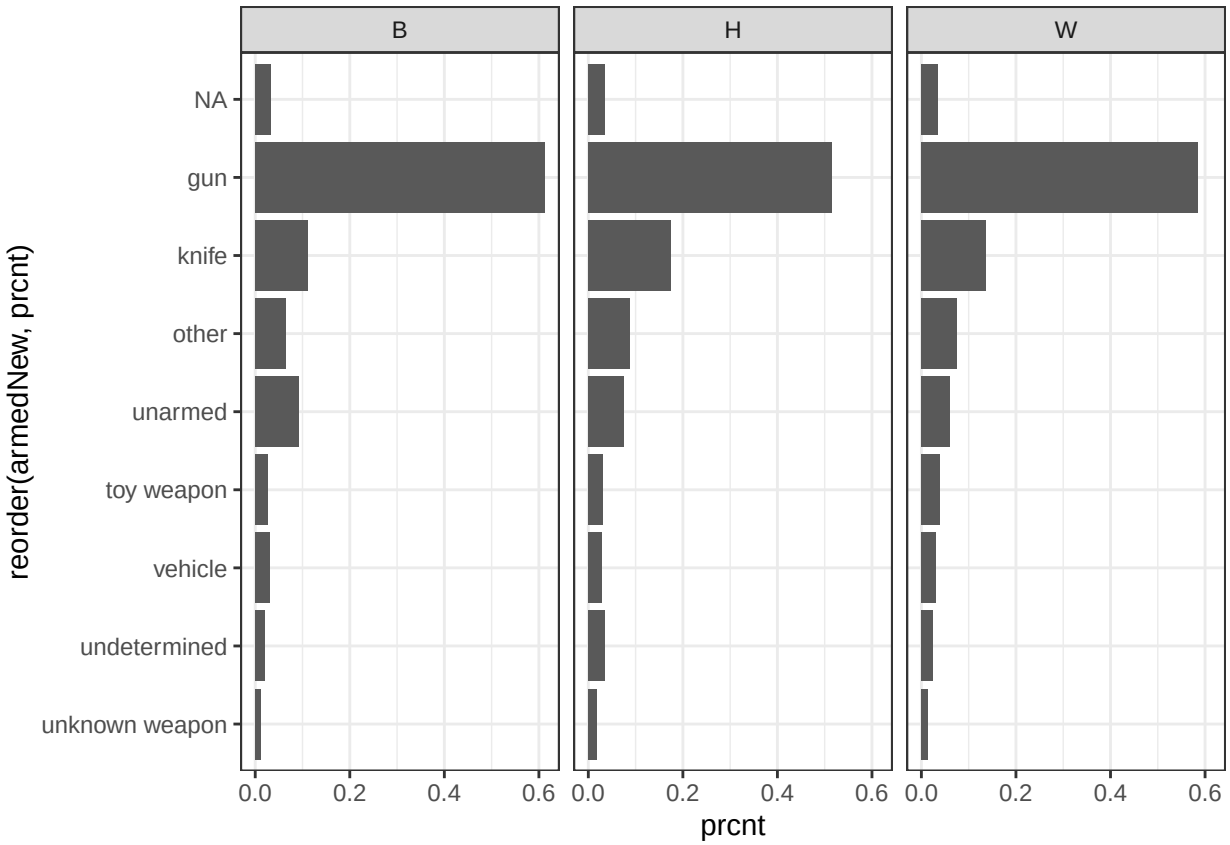
```
washpo %>%
    count(armedNew) %>%
    ggplot(aes(x = reorder(armedNew, n), y = n)) +
    geom_bar(stat = "identity") +
    theme_bw() +
    coord_flip()
```



Distribution of victims by if they were aremed or not and by race

No big differences in types of weapons used it seems. Although Black-Americans are slightly, slightly more likely to be unarmed.

```
washpo %>%
    filter(race %in% c("W", "B", "H")) %>%
    count(armedNew, race) %>%
    group_by(race) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(armedNew, prcnt), y = prcnt)) +
    geom_bar(stat = "identity") +
    facet_wrap(~race) +
    theme_bw() +
    coord_flip()
```


B
H
W
NA
gun
knife
other
unarmed
toy weapon
vehicle
undetermined
unknown weapon
reorder(armedNew, prcnt)
0.0
0.2
0.4
0.6
prcnt