

Exploring the Washington Post Data

Packages

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

Read in data

```
washpo <-
  read_csv(here("data_cleaning", "data", "washpo.csv")) %>%
  select(-id, -name, -longitude, -latitude, -is_geocoding_exact, -city) %>%
  mutate(year = year(date),
         month = month(date),
         day = day(date)) %>%
  select(-date)

##
## -- 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()
## )
```

Distribution of victim age

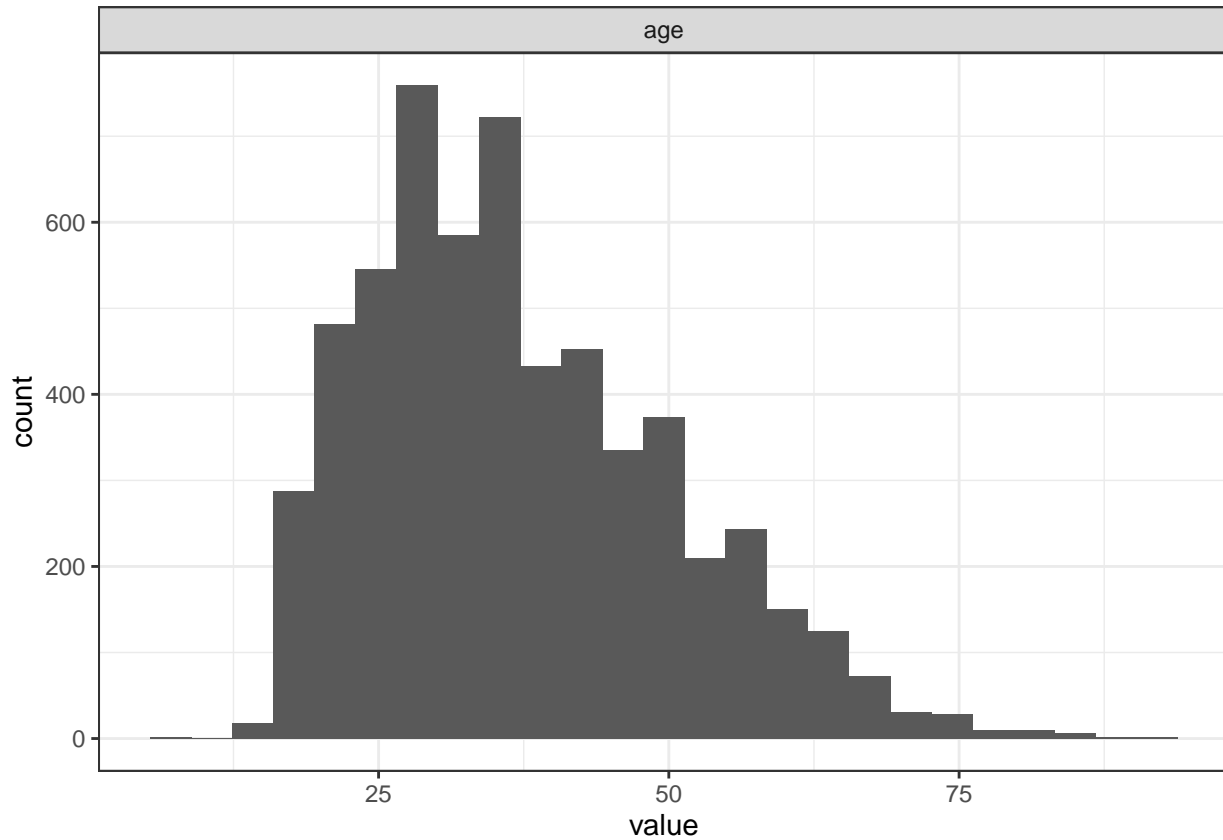
```
washpo %>%
  select(age) %>%
```

```

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

```

Warning: Removed 275 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	37.16	46.00	91.00	275

Distribution of victim age by race

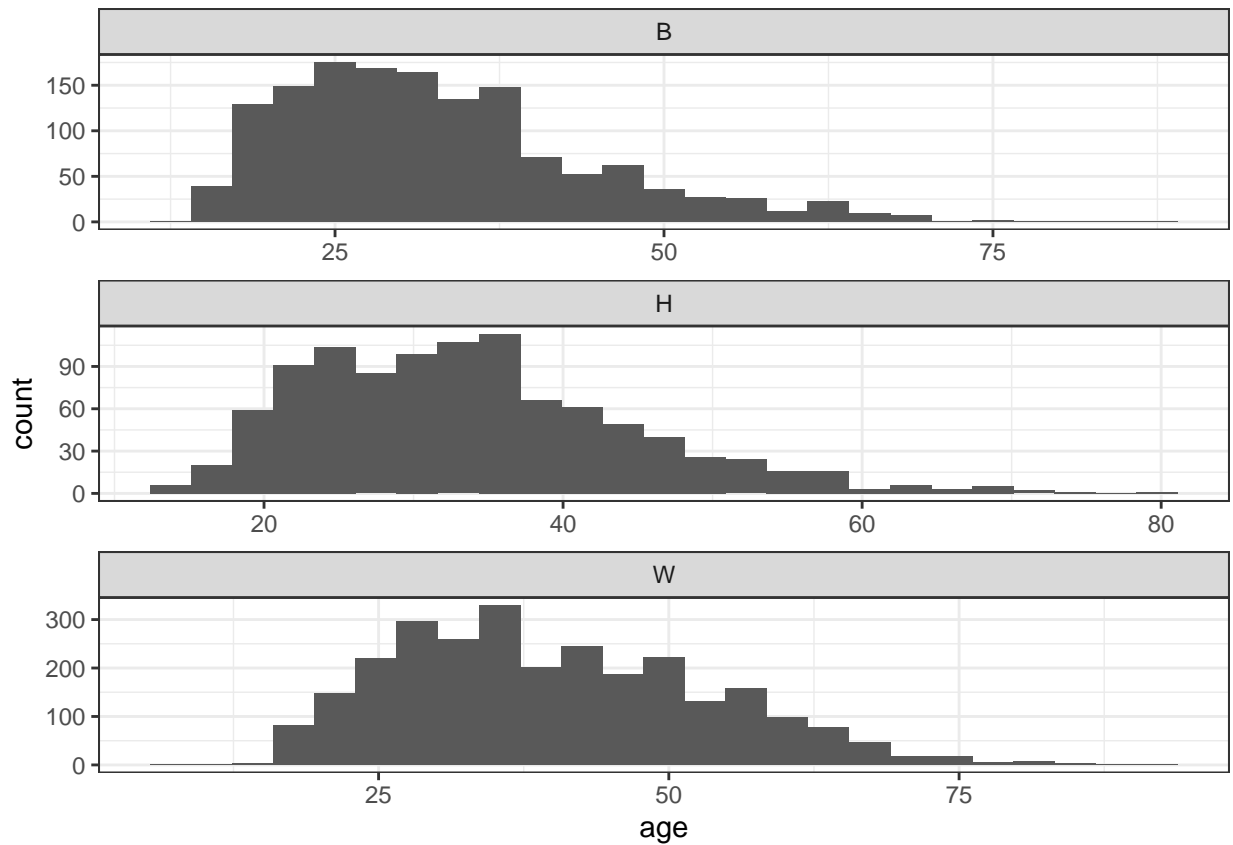
- Black American and Hispanic victims seem to be younger.

```

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 96 rows containing non-finite values (stat_bin).



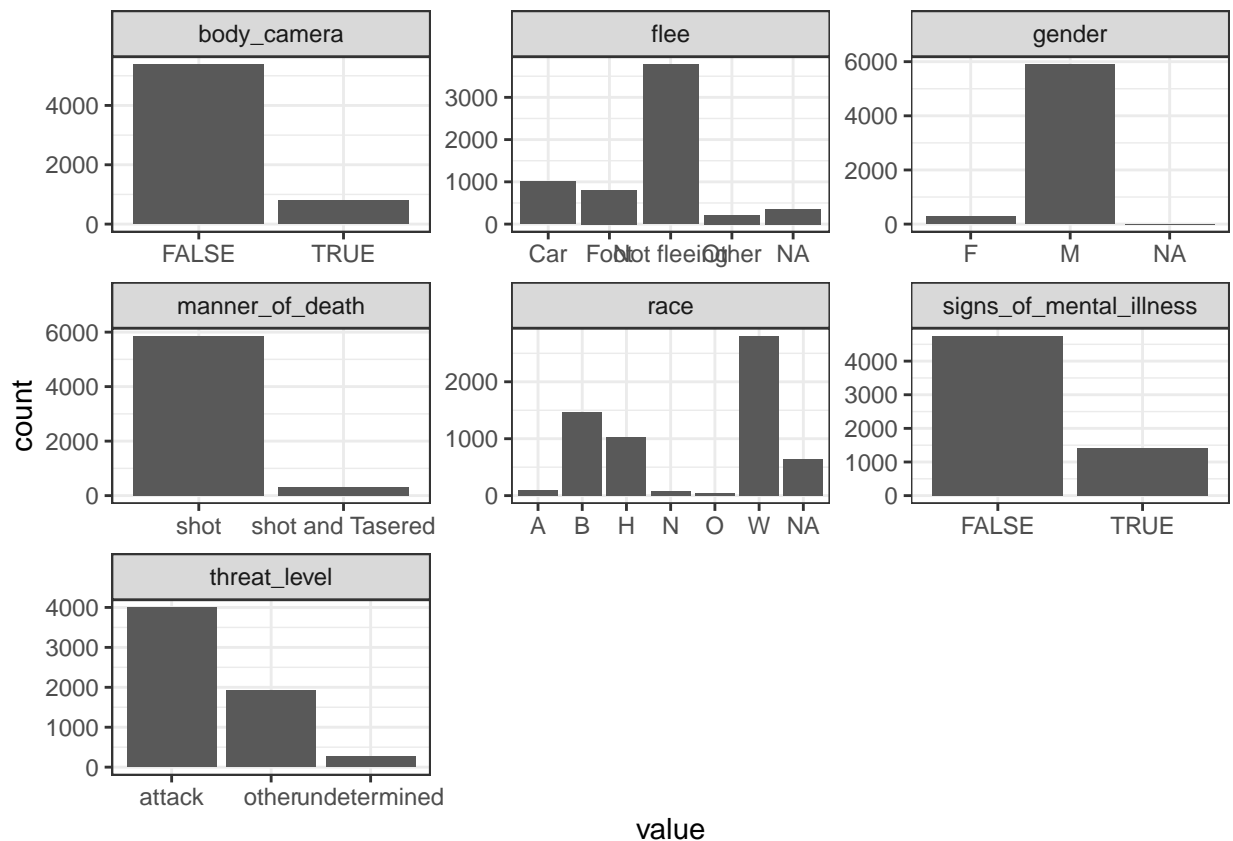
```
map(c("W", "B", "H"),
    function(df, raceCat) {
      df <- df %>% filter(race == raceCat)
      print(raceCat)
      summary(df$age)},
    df = washpo)
```

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

## [[1]]
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##     6      30      38      40      49      91     41
##
## [[2]]
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 13.00  24.00  31.00  32.69  39.00  88.00    27
##
## [[3]]
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 14.00  26.00  33.00  33.75  40.00  80.00    28
```

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

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

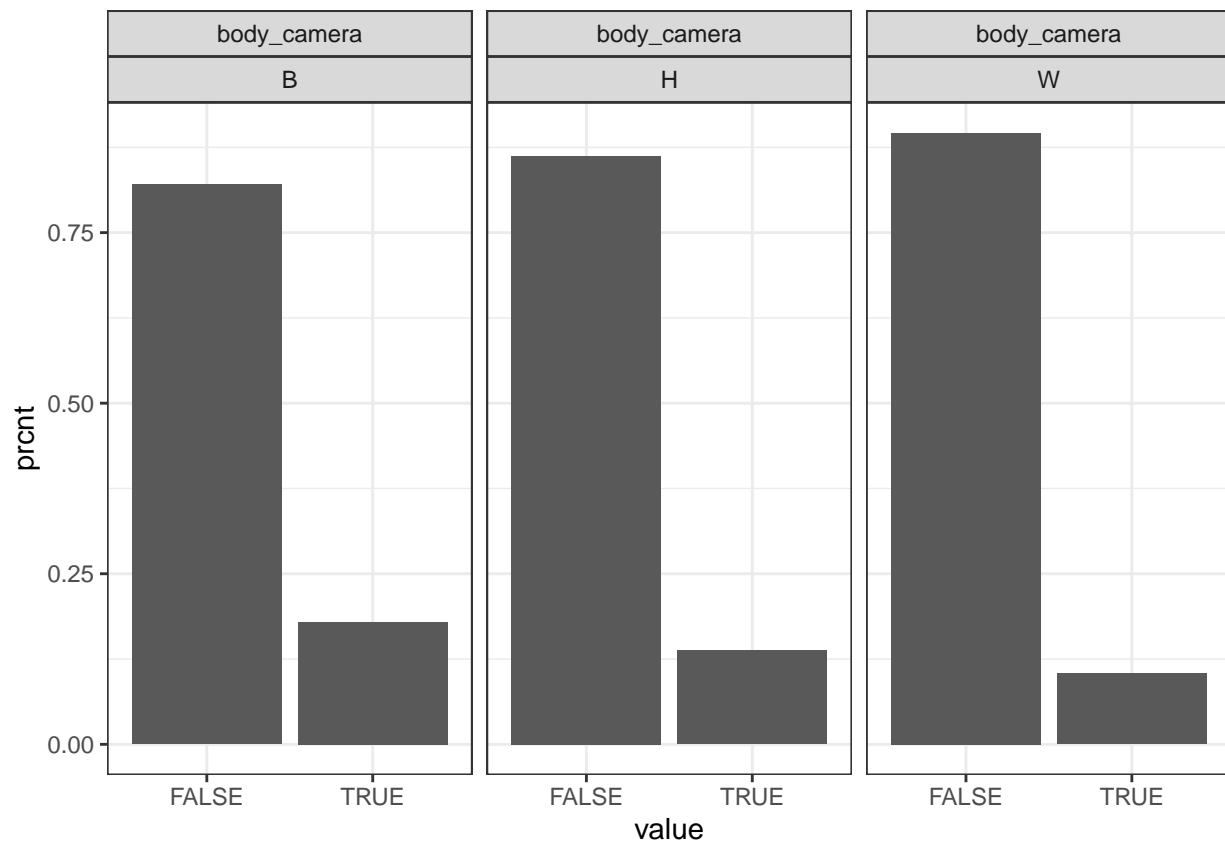
```

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

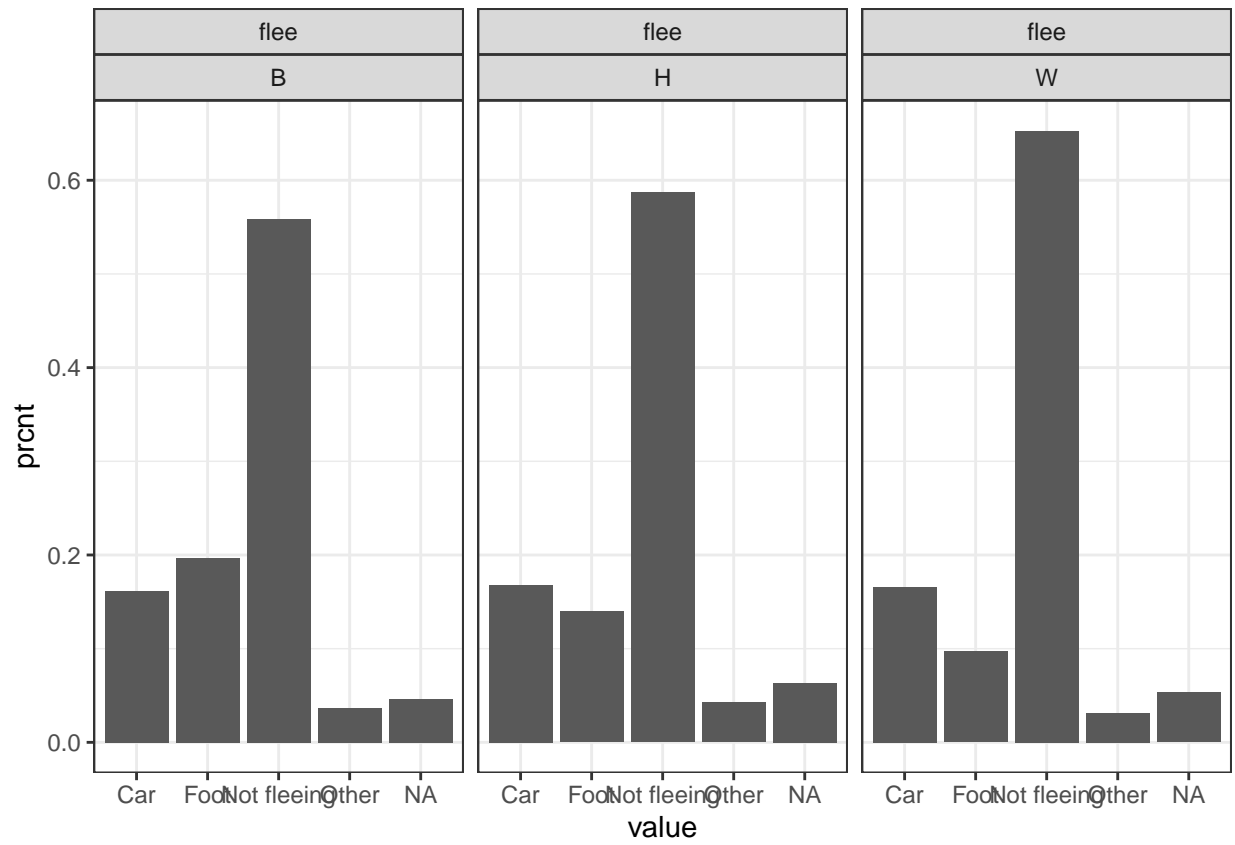
map(unique(categoriesRace$name),
  function(df, column) {
    df %>%
      filter(name == column) %>%
      ggplot(aes(x = value, y = prcnt)) +
      geom_bar(stat = "identity") +
      facet_wrap(~name+race, scales = "free_x") +
      theme_bw()},
  df = categoriesRace)

```

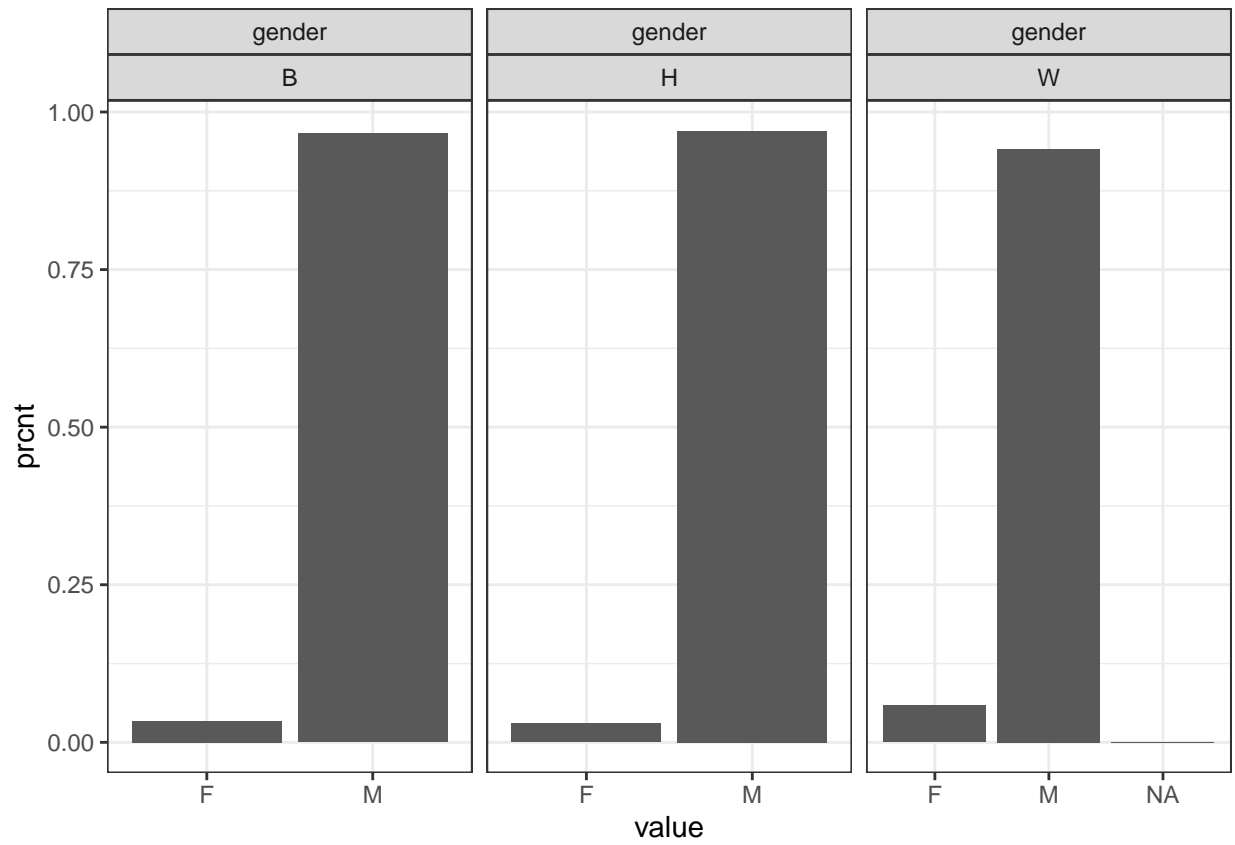
```
## [[1]]
```



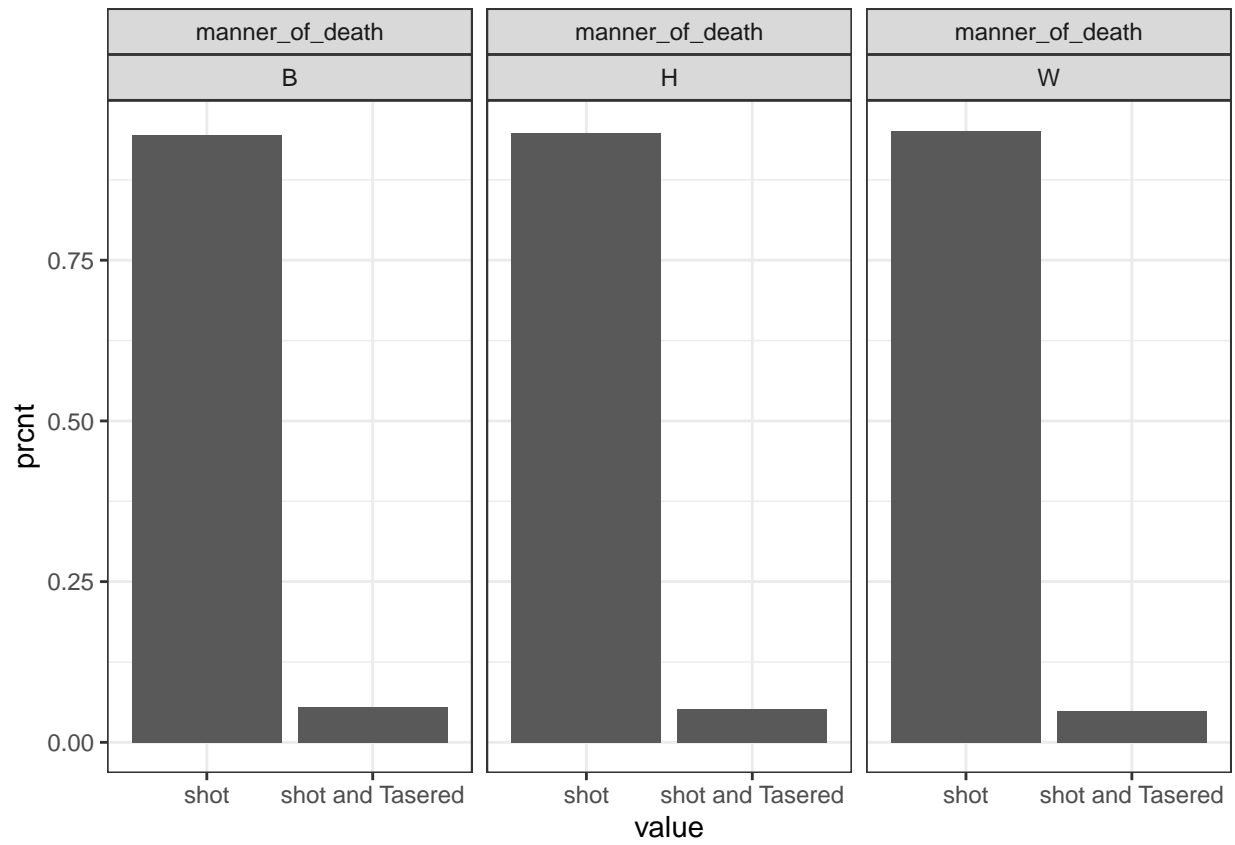
```
##
## [[2]]
```



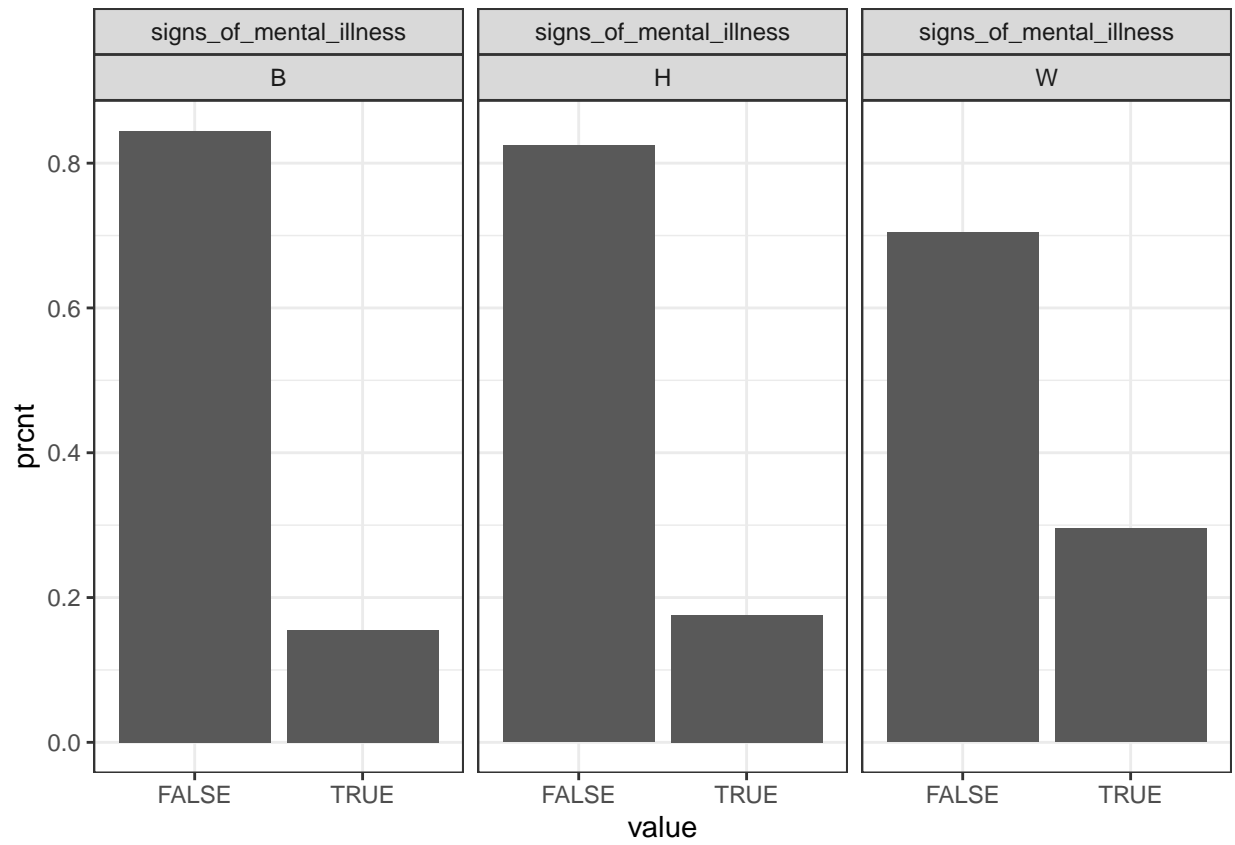
```
##
## [[3]]
```



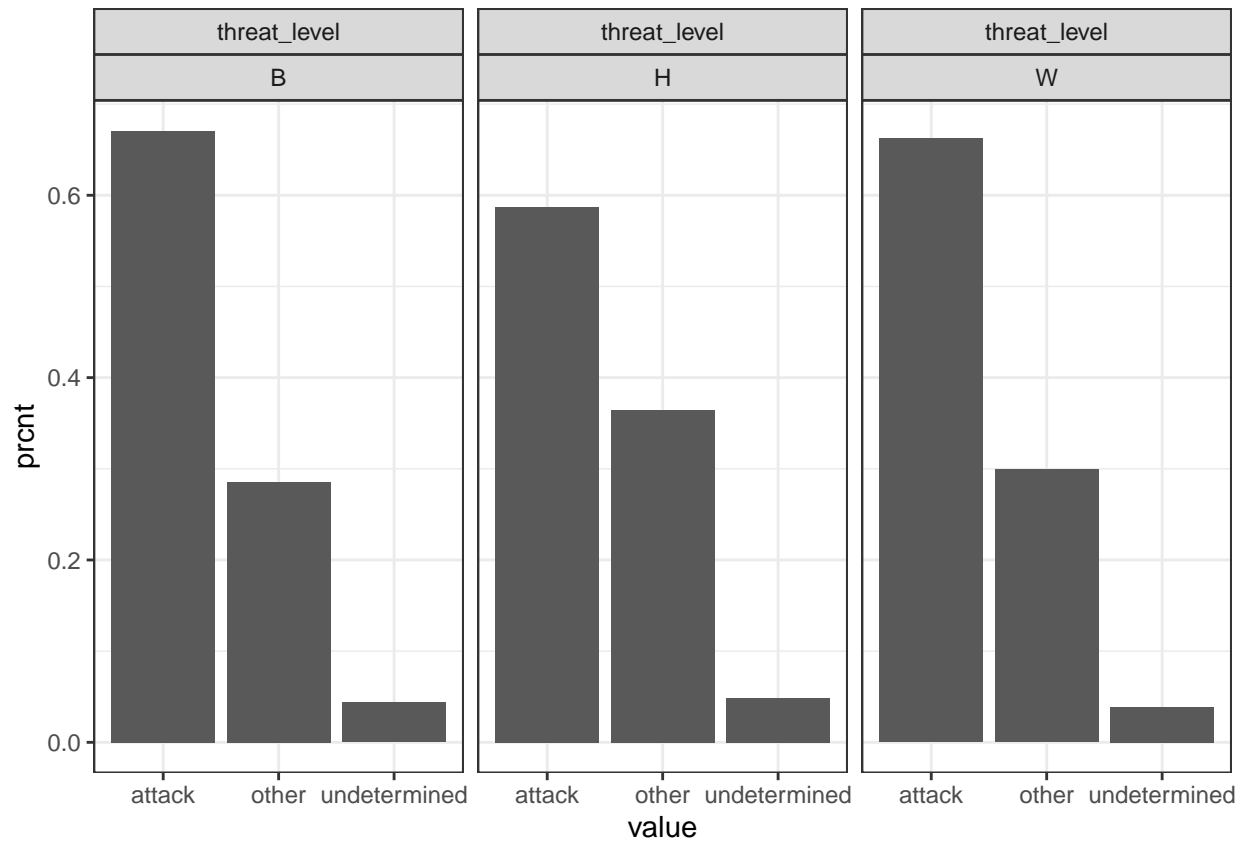
```
##  
## [[4]]
```



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

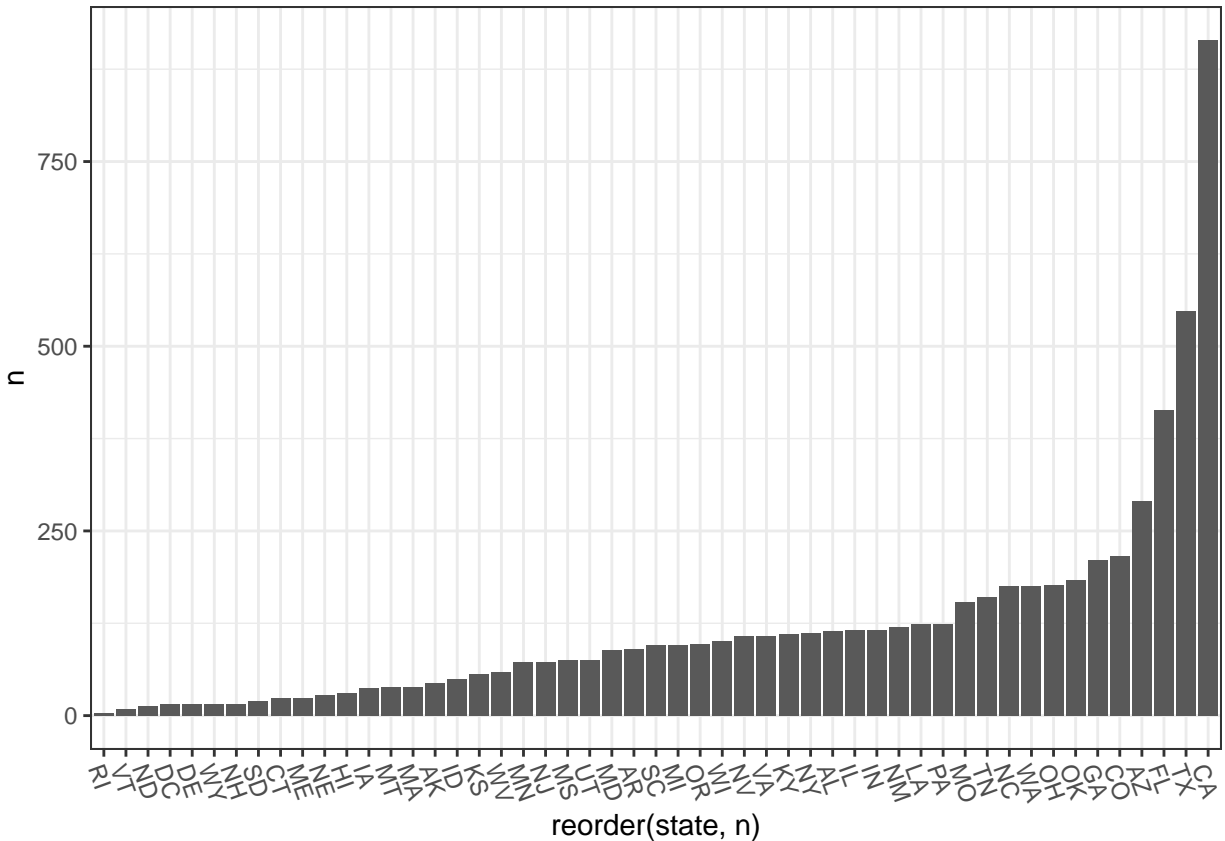



```
##  
## [[6]]
```



Distribution of victims by state

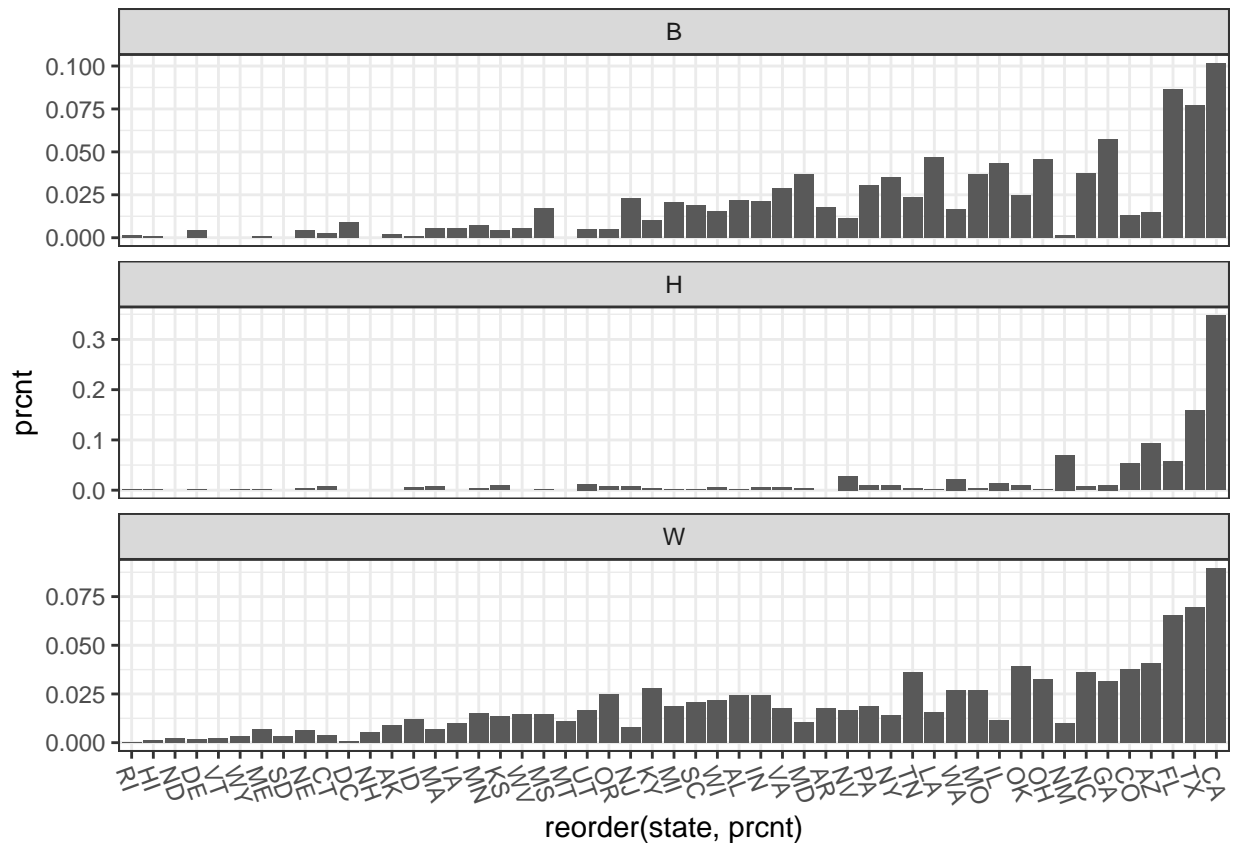
```
washpo %>%
  select(state) %>%
  group_by(state) %>%
  count(state) %>%
  ggplot(aes(x = reorder(state, n), n)) +
  geom_bar(stat = "identity") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 290, hjust = 0))
```



Distribution of victims by state by race

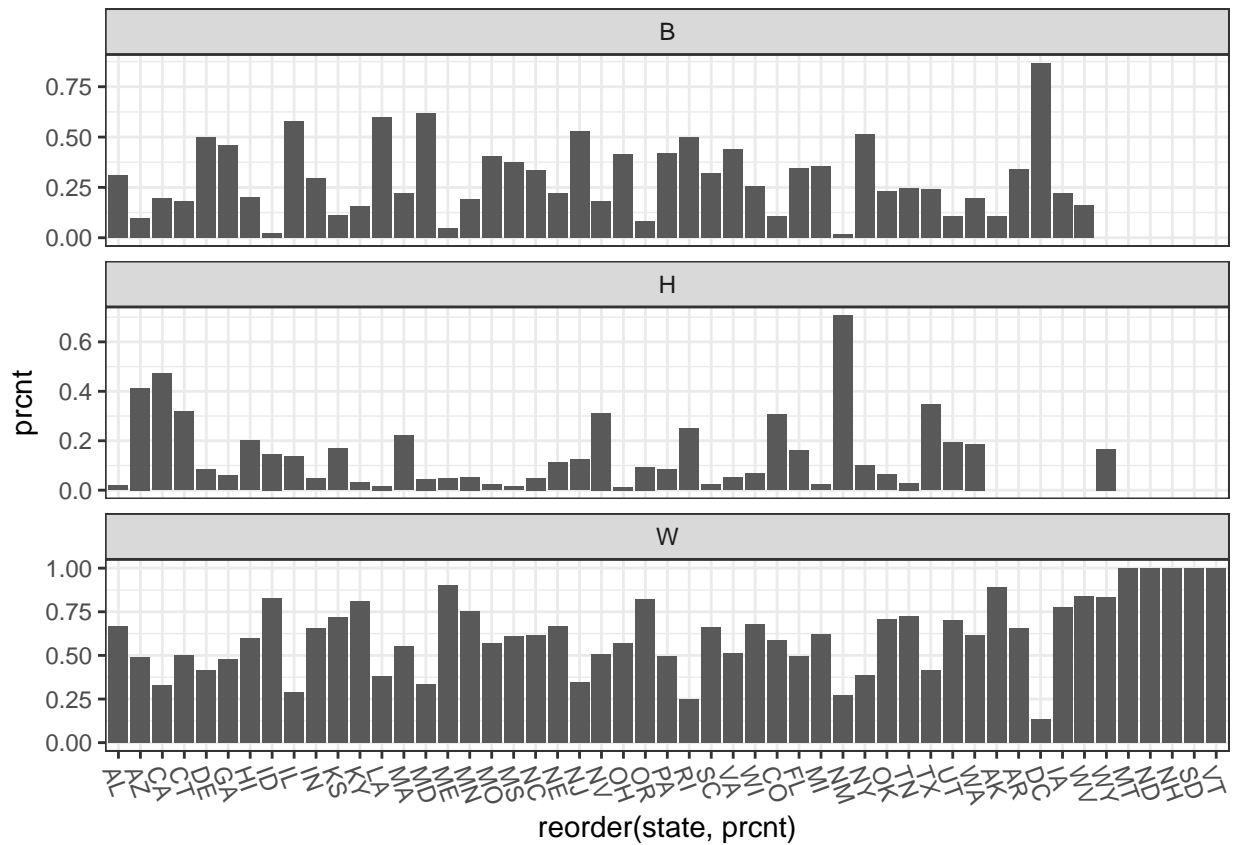
- Within each racial category, what proportion of killings took place in each state? E.g. of all Black Americans killed by the police, 10% were killed in CA. Of all White Americans killed by the police, 8% were killed in California. Of all Hispanics killed by the police, over 30% were killed in CA.
- 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 states. 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))
```



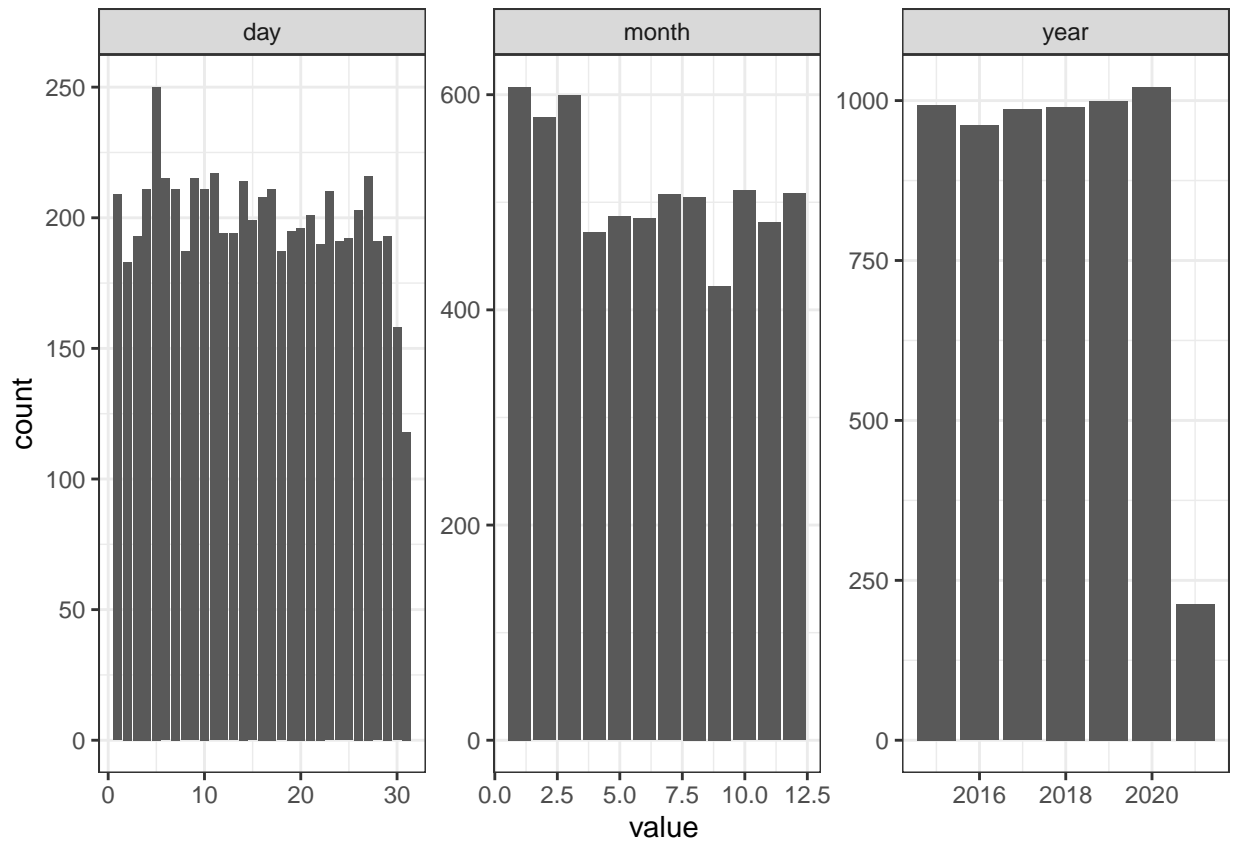
- Within each state, what proportion of killings were for each race? E.g. of all Americans killed by the police in Alabama, 25% were Black, 1% were Hispanic, and 70% were White.
- Hard to discern any immediate trends. I need population totals to get a sense of the disproportionality.

```
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") +
  facet_wrap(~race, scales = "free_y", ncol = 1) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 290, hjust = 0))
```



Number of killings by day, month, year

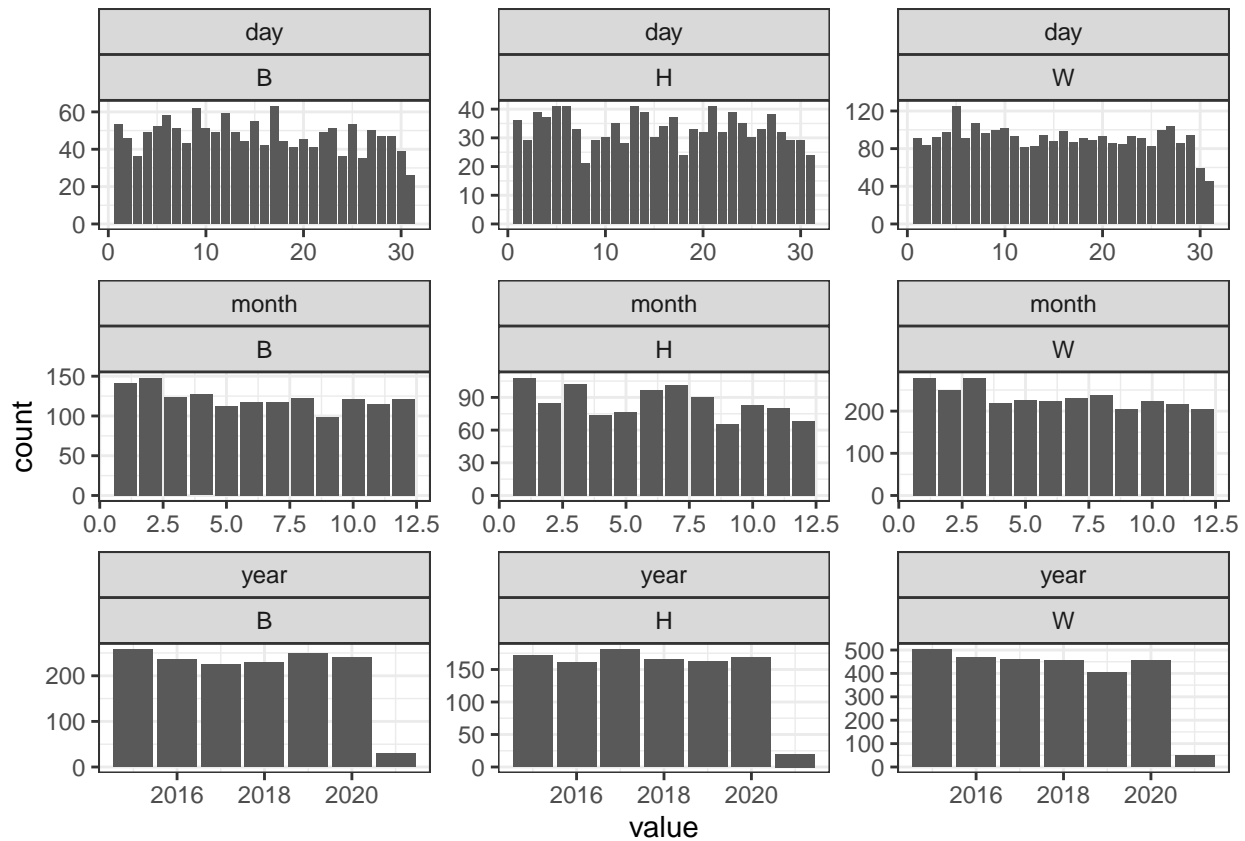
```
washpo %>%
  select(day, month, year) %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x = value)) +
  geom_bar() +
  facet_wrap(~name, scales = "free") +
  theme_bw()
```



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.

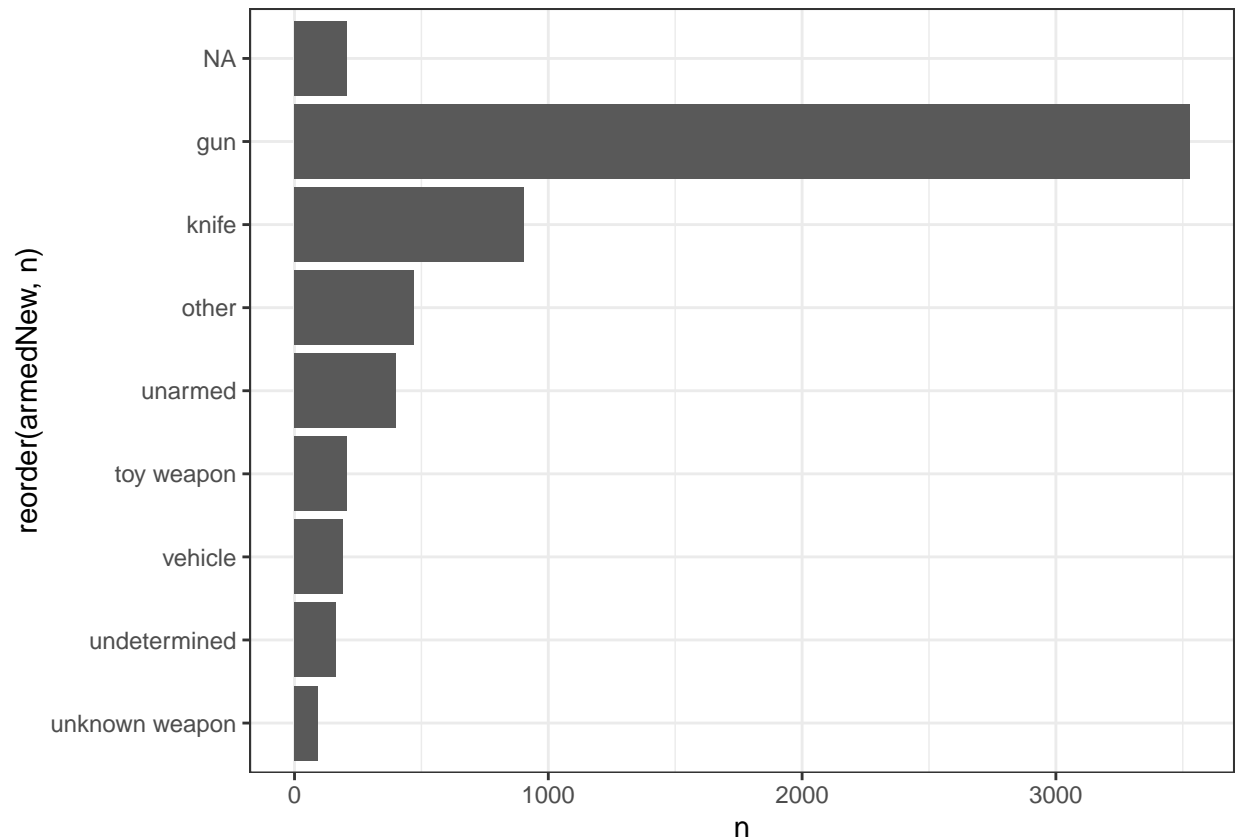
```
washpo %>%
  select(day, month, year, race) %>%
  pivot_longer(-race) %>%
  filter(race %in% c("W", "B", "H")) %>%
  ggplot(aes(x = value)) +
  geom_bar() +
  facet_wrap(~name+race, scales = "free") +
  theme_bw()
```



Distribution of victims by if they were armed or not

```
armedNewDf <-
  washpo %>%
  select(armed) %>%
  count(armed) %>%
  mutate(prcnt = n / sum(n),
         armedNew = if_else(prcnt < 0.01, "other", armed))

armedNewDf %>%
  group_by(armedNew) %>%
  summarise(n = sum(n)) %>%
  ungroup() %>%
  ggplot(aes(x = reorder(armedNew, n), y = n)) +
  geom_bar(stat = "identity") +
  theme_bw() +
  coord_flip()
```



Distribution of victims by if they were armed 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 %>%
  select(armed, race) %>%
  filter(race %in% c("W", "B", "H")) %>%
  inner_join(select(armedNewDf, armed, armedNew), by = "armed") %>%
  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()
```