Exploring the Vice data

Packages

```
library(here)
```

```
## here() starts at /home/joe/Documents/police_shooting
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(readr)
library(purrr)
library(tidyr)
library(ggplot2)
library(stringr)
library(lubridate)
```

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

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

Read in the data

Columns in common:

- Date
- Number of Subjects (It's always 1 in vice subject dataset).
- Fatal
- Subject Armed (with a gun)
- Notes
- Subject Race
- Subject Gender
- Subject Age
- Nature of Stop
- Number of Shots
- Number of Officers
- Officer Race
- Officer Gender

- Department
- Full Narrative
- City

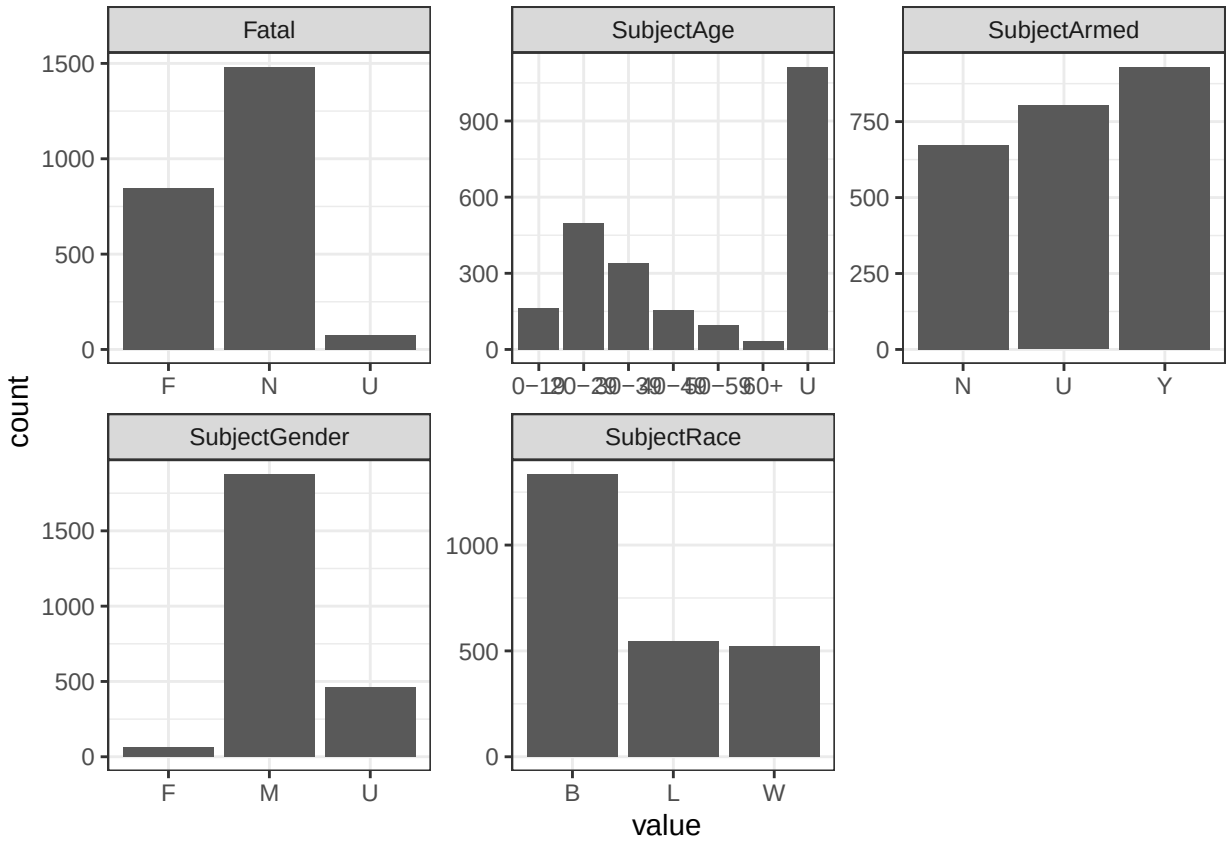
Columns unique to vice incidents:

- year, month, day: turns the date into 3 columns.
- fdate: turns dates which are missing month and/or day into a fixed date.
- b, l, w, and a: Total number of subjects in each racial category.
- tr: Total number of racial categories identified. Mostly agrees with Number of Subjects, but it will differ if racial categories are missing.
- AvgAge: The average age of the subjects.
- VA: Binary variable indicating if any subject was armed or not.
- F: Binary variable indicating if any subject died.
- nshots: Total number of shots. Turns NumberOfShots into a continuous, numeric variable.
- sit: Categorizes Nature of Stop into a more manageable set of categories.
- other: Was a weapon other than a knife/gun/replica weapon used? Taken from notes and narrative.
- knife: Was a knife used? Taken from notes and narrative.
- replica: Was a replica weapon used? Taken from notes and narrative.
- weapon: What weapon was used? Taken from notes and narrative.

```
load(here("create_test_train", "output", "raceTrain.RData"))
load(here("create_test_train", "output", "fatalTrain.RData"))
viceRace <- raceTrain[["viceDgHM"]]
viceFatal <- fatalTrain[["fatalHM"]]
```

Bar charts for demographics and fatal vs. non-fatal shootings

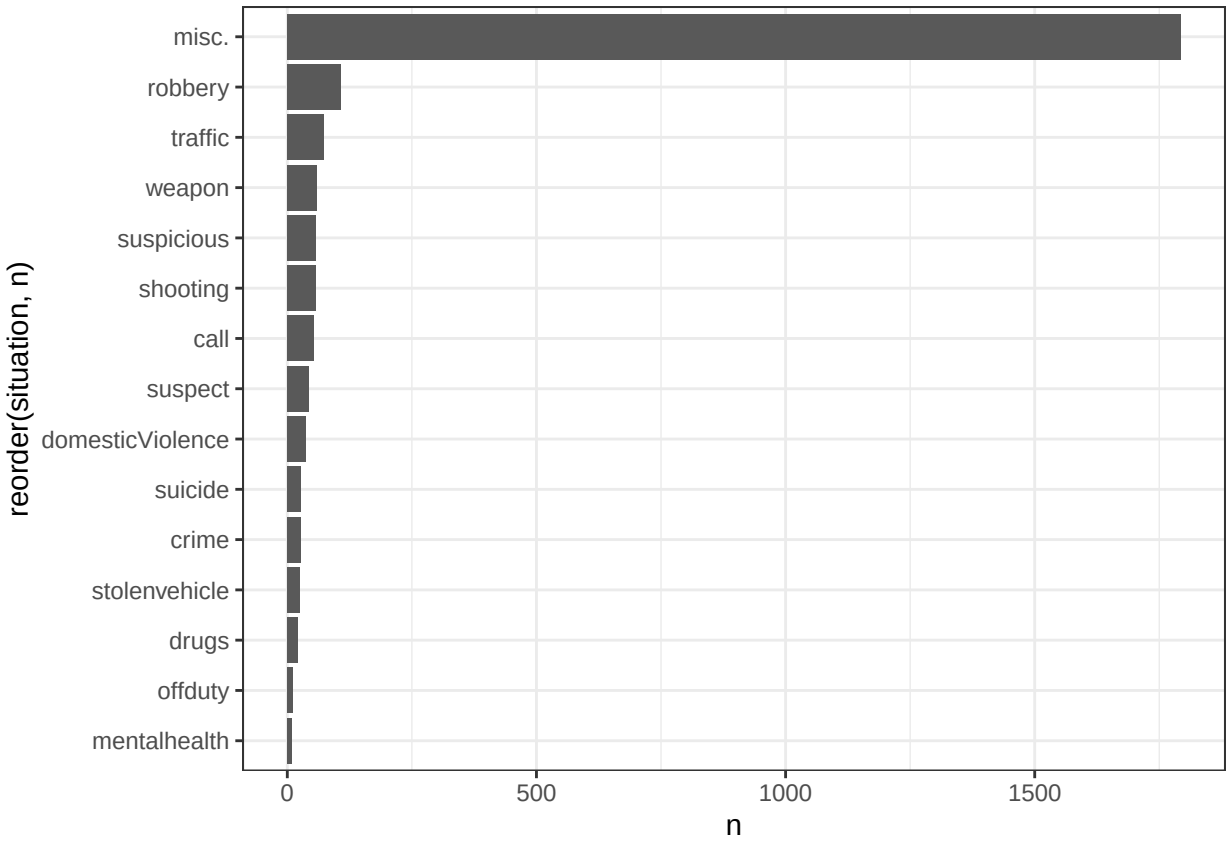
```
viceRace %>%
    select(c(!where(is.numeric)), -situation, -departmentFixed) %>%
    pivot_longer(everything(),
                 values_transform = list(value = as.character)) %>%
    ggplot(aes(x = value)) +
    geom_bar() +
    facet_wrap(~name, scales = "free") +
    theme_bw()
```



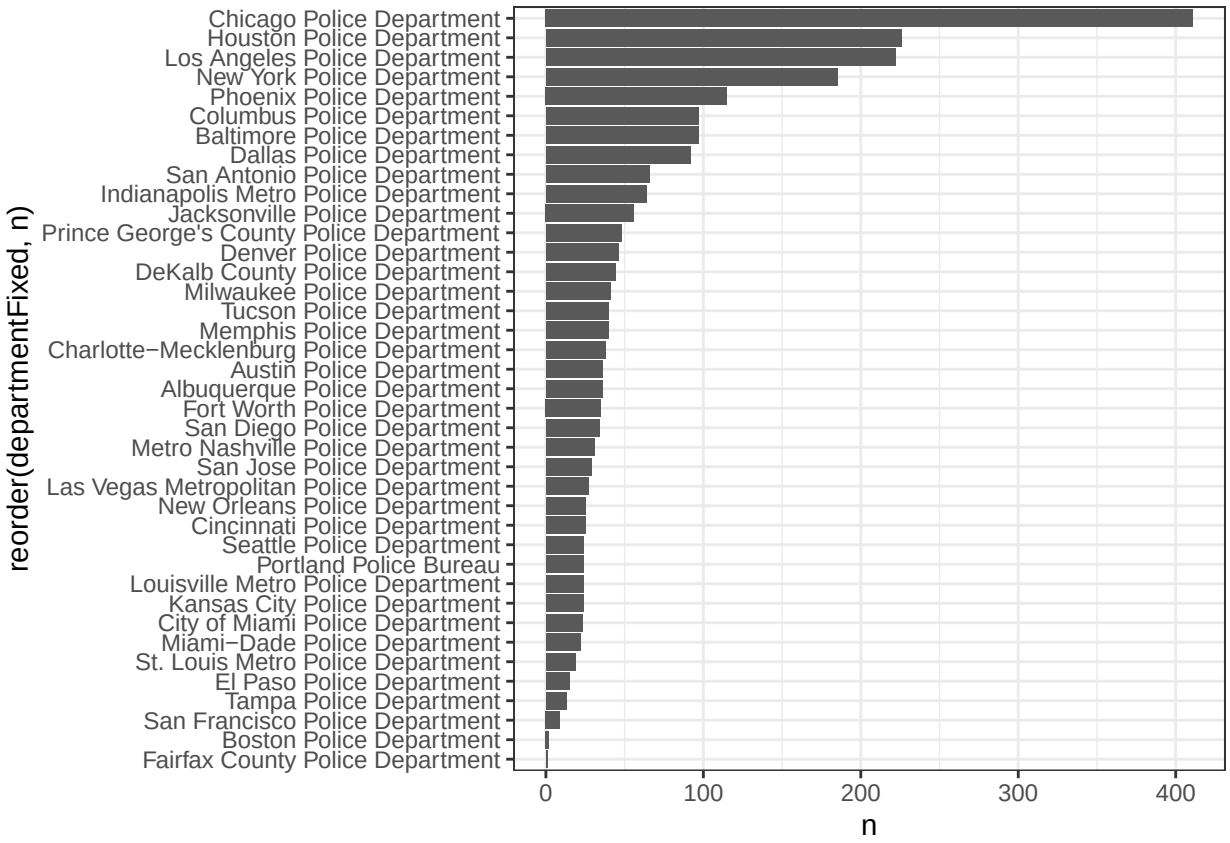
Bar charts for situation and department

```
map(list("situation", "departmentFixed"), function(df, column){

    df %>%
        count(.data[[column]]) %>%
        ggplot(aes(x = reorder(.data[[column]], n), y = n)) +
        geom_bar(stat = "identity") +
        theme_bw() +
        coord_flip()

}, df = viceRace)
```

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

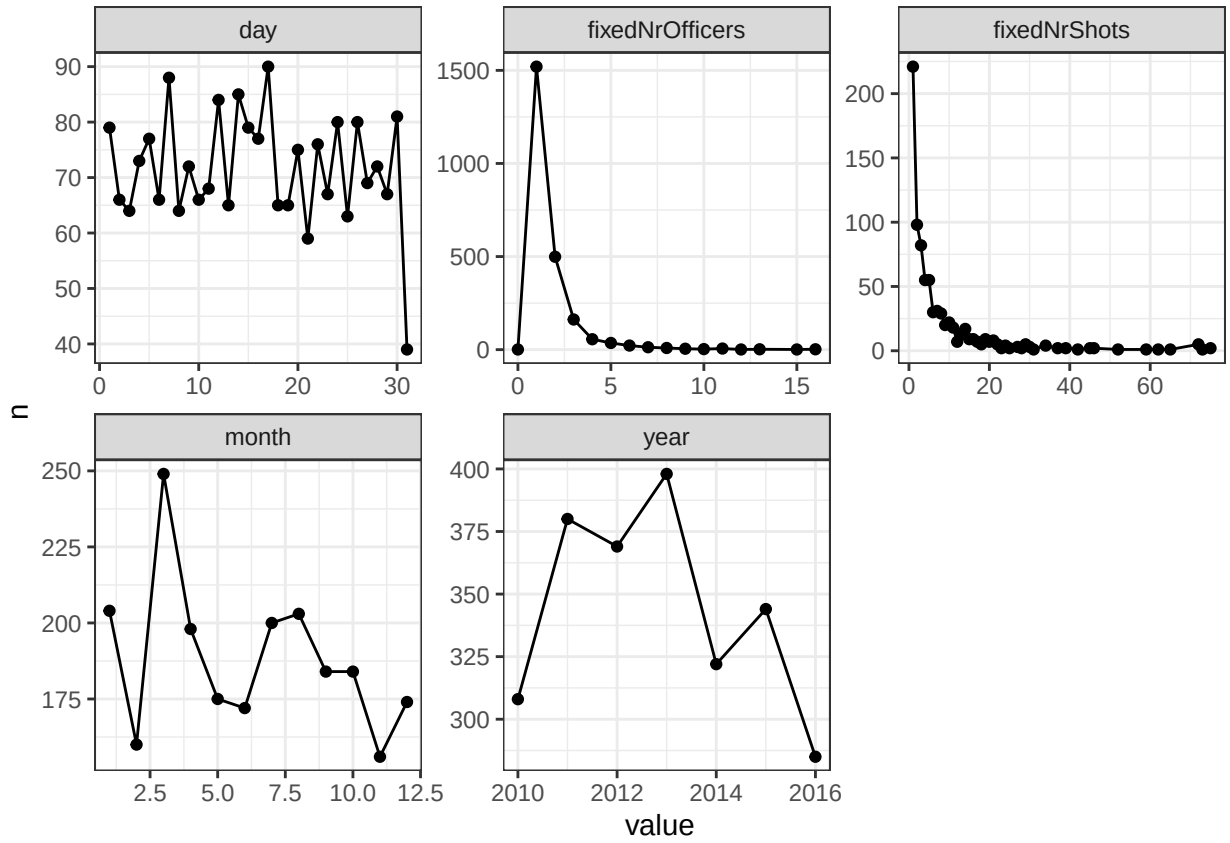

misc.
robbery
traffic
weapon
suspicious
shooting
call
suspect
domesticViolence
suicide
crime
stolenvehicle
drugs
offduty
mentalhealth
reorder(situation, n)
0
500
1000
1500
n

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



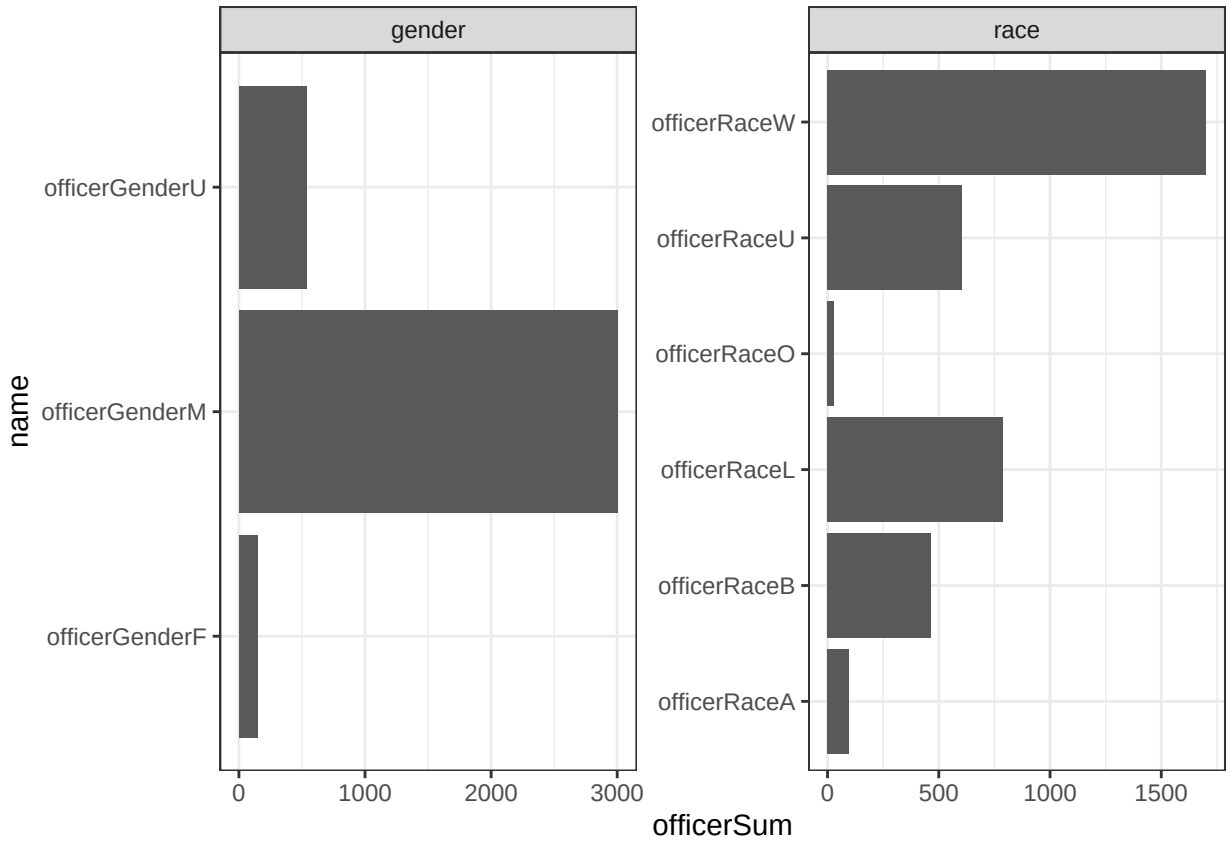
Line charts for numerical variables

```
viceRace %>%
    select(c(where(is.numeric), -matches("^officer"))) %>%
    pivot_longer(everything()) %>%
    count(value, name) %>%
    filter(!is.na(value)) %>%
    ggplot(aes(x = value, y = n)) +
    geom_line() +
    geom_point() +
    facet_wrap(~name, scales = "free") +
    theme_bw()
```



Bar charts for officer gender and race

```
viceRace %>%
    select(matches("^officer")) %>%
    pivot_longer(everything()) %>%
    group_by(name) %>%
    summarise(officerSum = sum(value)) %>%
    ungroup() %>%
    mutate(group = if_else(str_detect(name, "Gender"), "gender", "race")) %>%
    ggplot(aes(x = name, y = officerSum)) +
    geom_bar(stat = "identity") +
    facet_wrap(~group, scales = "free") +
    theme_bw() +
    coord_flip()
```



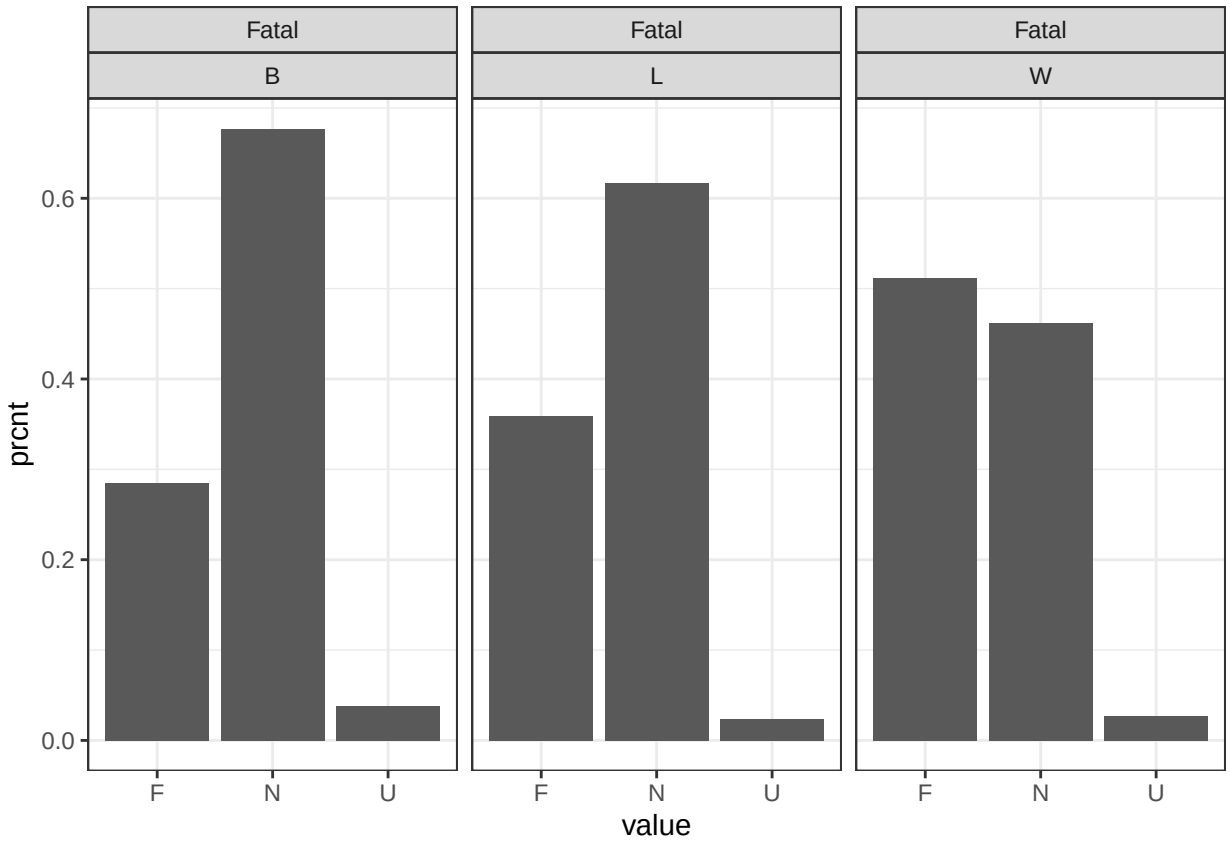
Bar charts for demographics and fatal vs. non-fatal shootings by race

- In terms of fatal shootings, Black-Americans, percentage-wise, suffer the least amount of fatal shootings while White Americans suffer the most fatal shootings, percentage-wise. This isn't what I would have expected beforehand.
- Age distributions looks pretty similar.
- White Americans seem the most likely to be armed followed by Hispanics than by Black Americans. However uncertainty is lowest for White Americans increasing for Hispanics and increasing even more for Black Americans.
- The gender breakdowns are roughly the same across racial groups. Uncertainty is lowest for White Americans and higher for Hispanics and is the highest for Black Americans.

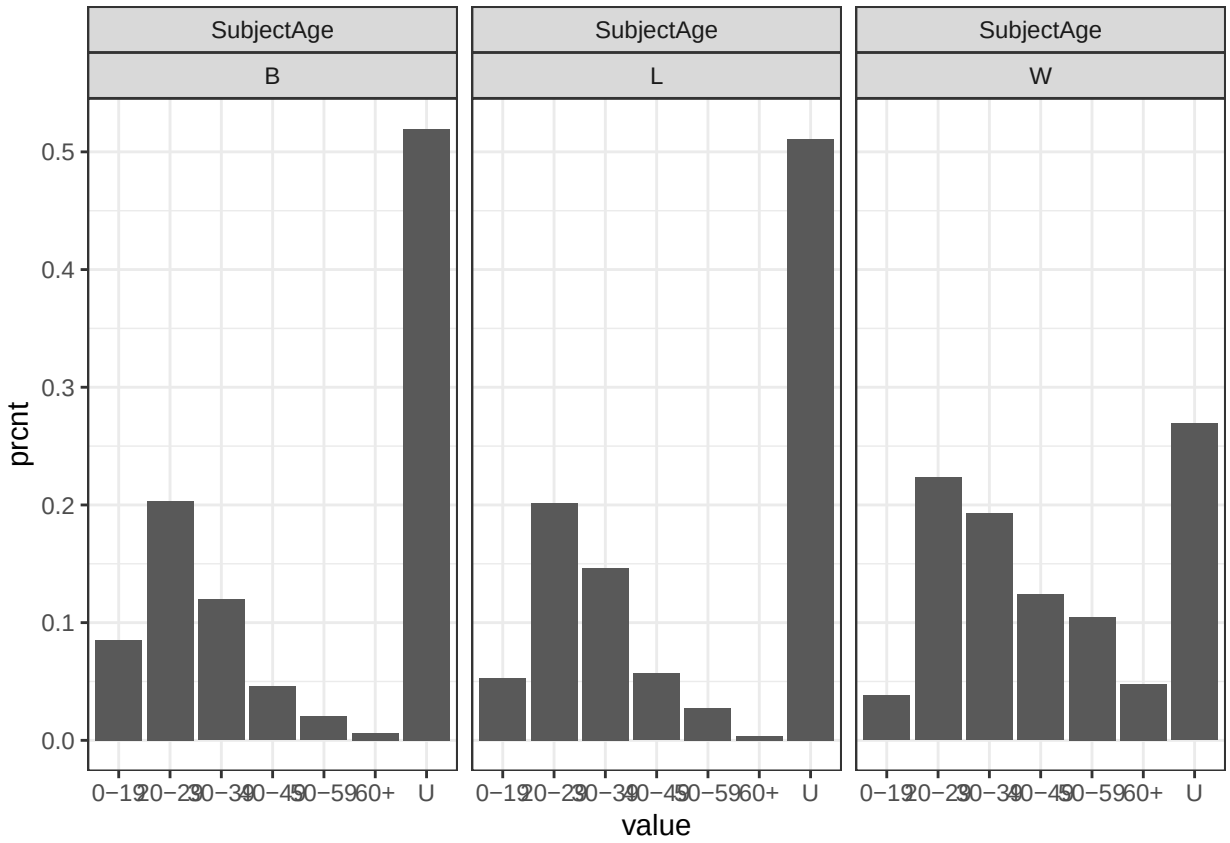
```
categoriesRace <-
    viceRace %>%
    select(c(!where(is.numeric)), -situation, -departmentFixed) %>%
    pivot_longer(-SubjectRace) %>%
    filter(SubjectRace %in% c("W", "L", "B")) %>%
    count(SubjectRace, name, value) %>%
    group_by(SubjectRace, name) %>%
    mutate(prcnt = n / sum(n)) %>%
    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+SubjectRace, scales = "free_x") +
            theme_bw()},
    df = categoriesRace)
```

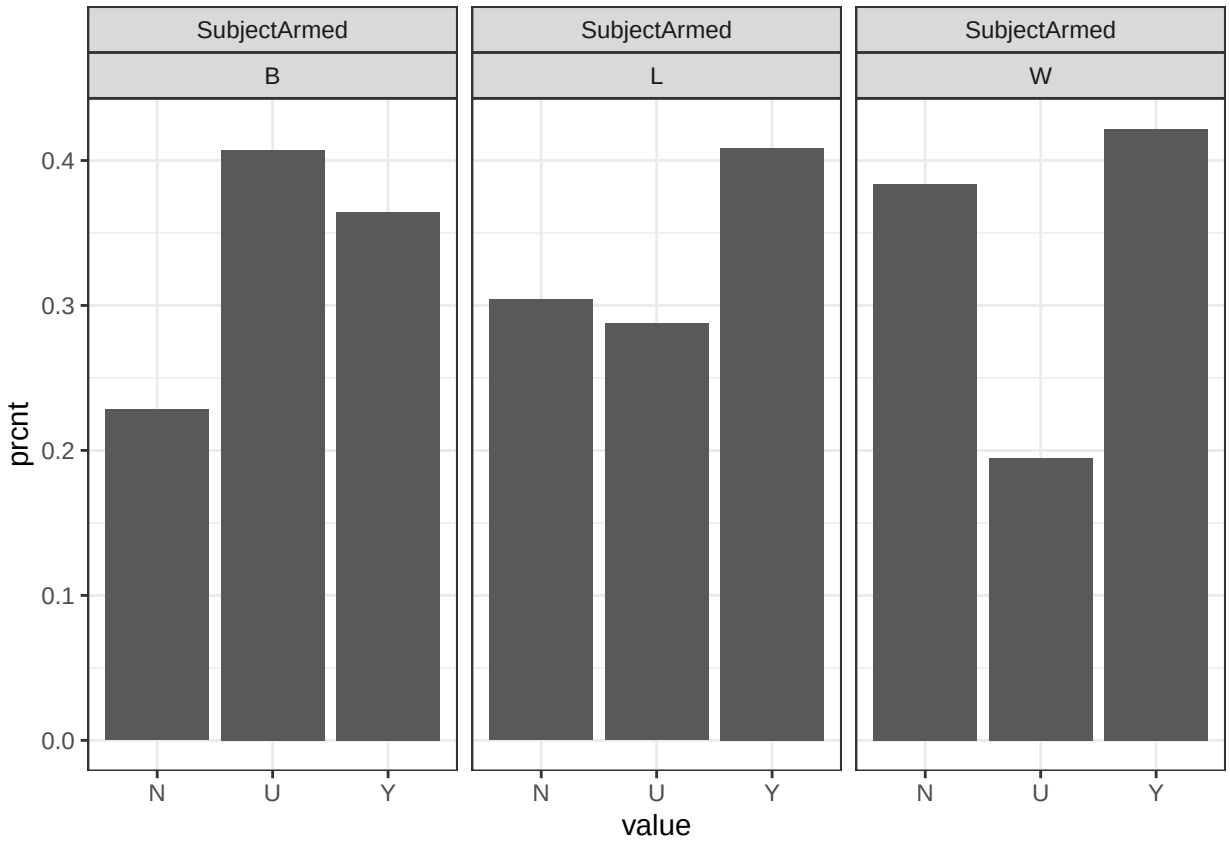
```
## [[1]]
```



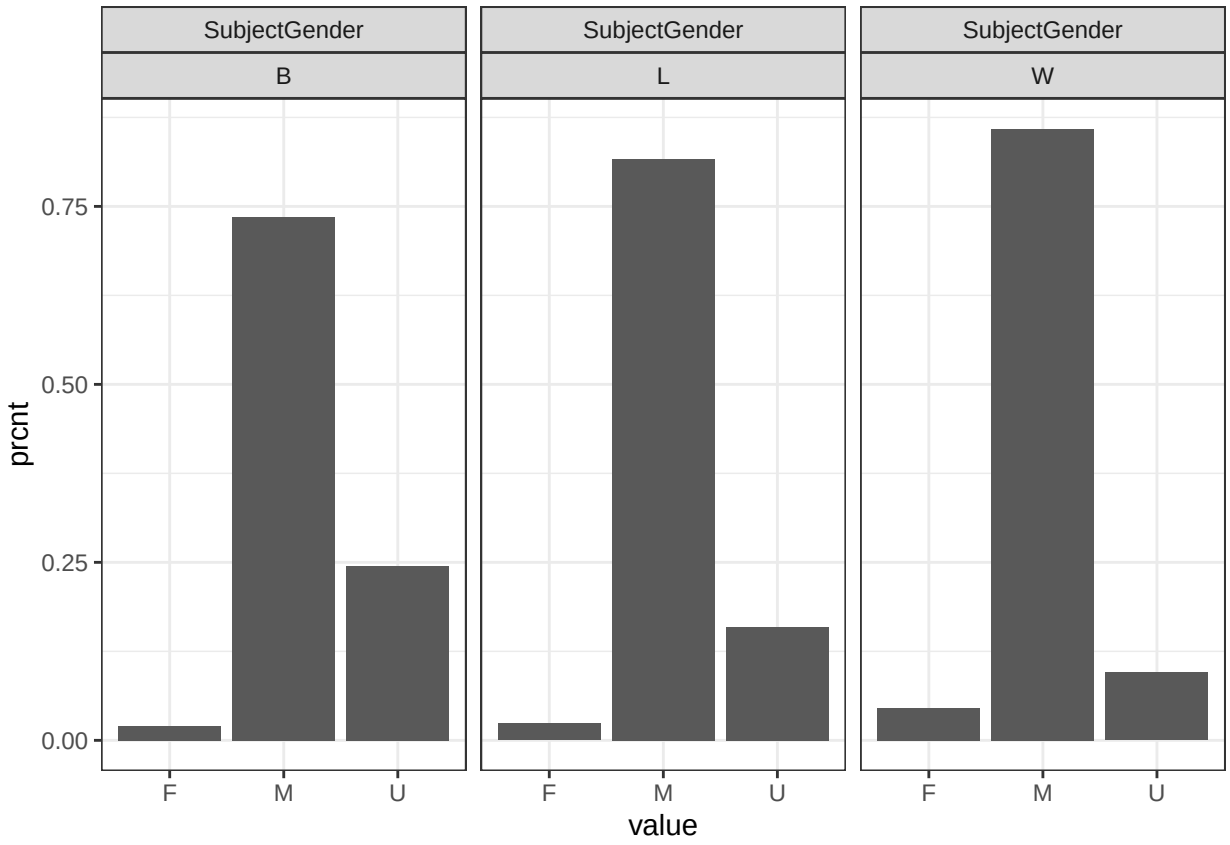

```
##
## [[2]]
```


SubjectAge
B
SubjectAge
L
SubjectAge
W
prcnt
0.0
0.1
0.2
0.3
0.4
0.5
0–19
20–29
30–39
40–49
50–59
60+
U
value

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


SubjectArmed
B
SubjectArmed
L
SubjectArmed
W
prcnt
0.0
0.1
0.2
0.3
0.4
N
U
Y
value

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

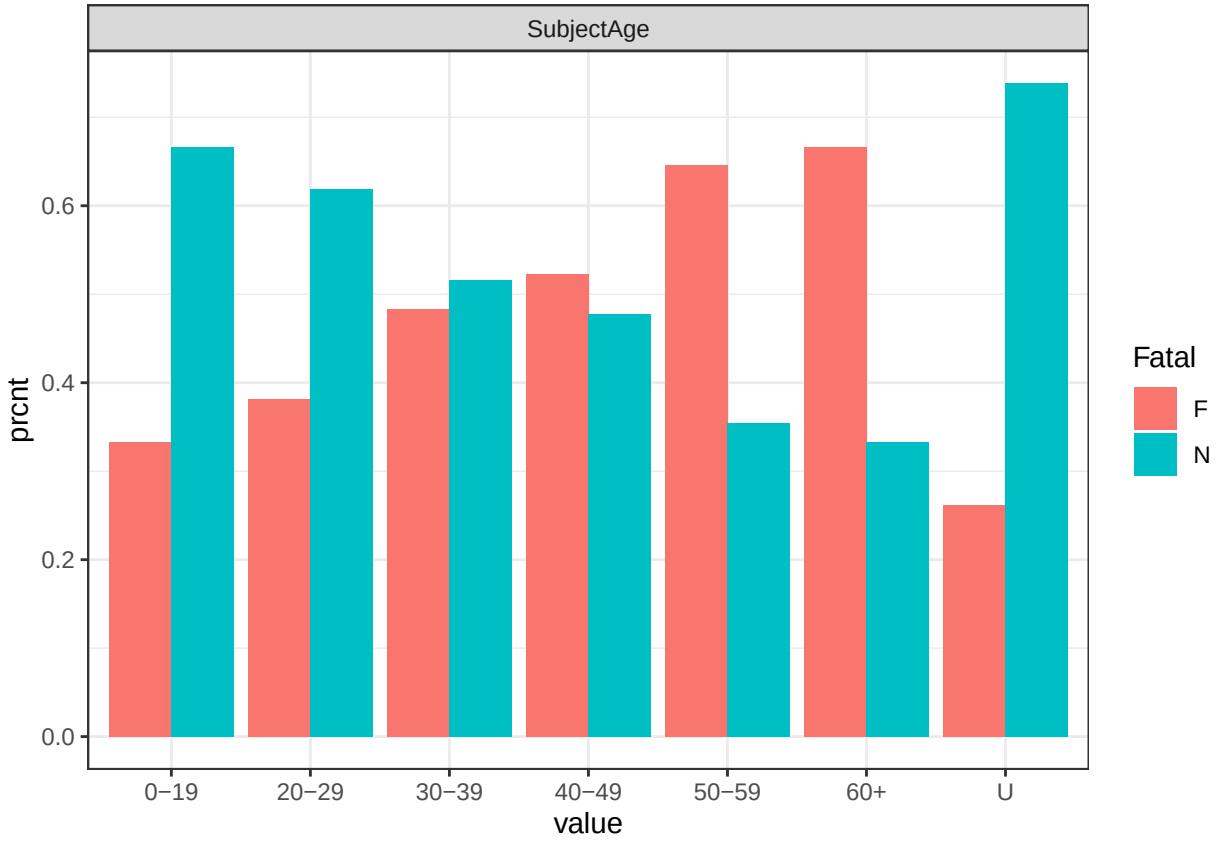


Bar charts for demographics by fatal vs. non-fatal shootings

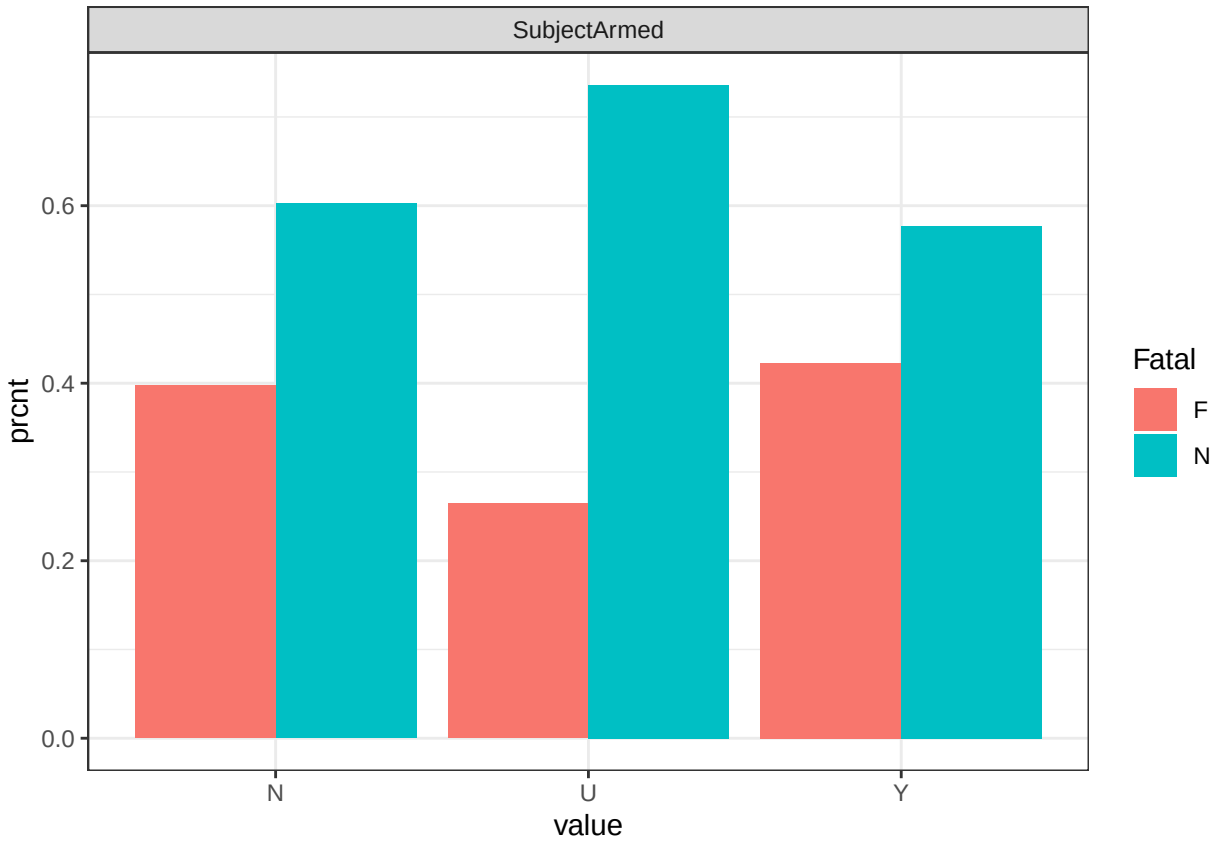
- Interestingly shootings become more and more fatal as one gets older.
- Surprisingly roughly the same proportion of armed vs. non-armed subjects are fatally shot.
- Roughly equal proportions of men and women are fatally shot.
- Whites are fatally shot at moderately higher proportions vs. Black-Americans and Hispanics.

```
categoriesFatal <-
    viceRace %>%
    select(c(!where(is.numeric)), -situation, -departmentFixed) %>%
    pivot_longer(-Fatal) %>%
    filter(Fatal %in% c("F", "N")) %>%
    count(Fatal, name, value) %>%
    group_by(name, value) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup()

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

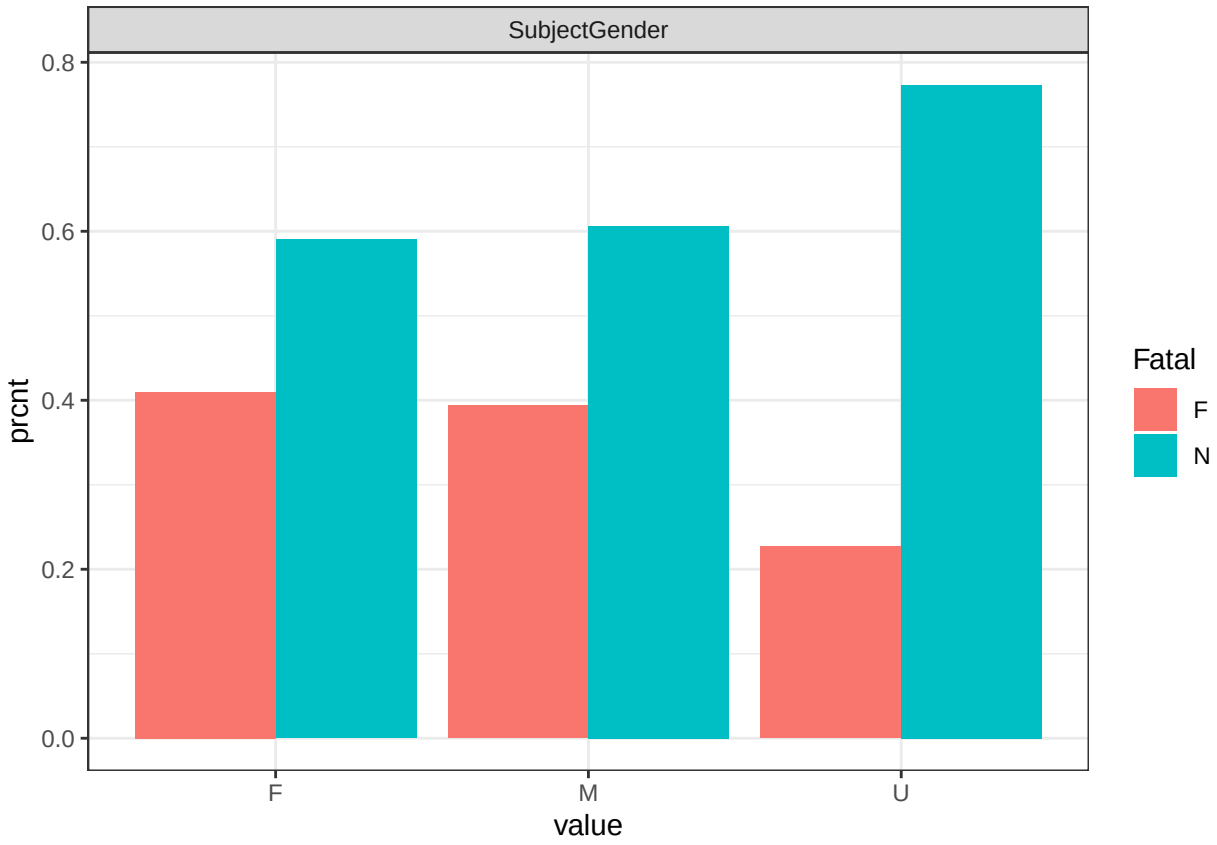
```
## [[1]]
```


SubjectAge
prcnt
0.0
0.2
0.4
0.6
0–19
20–29
30–39
40–49
50–59
60+
U
value
Fatal
F
N

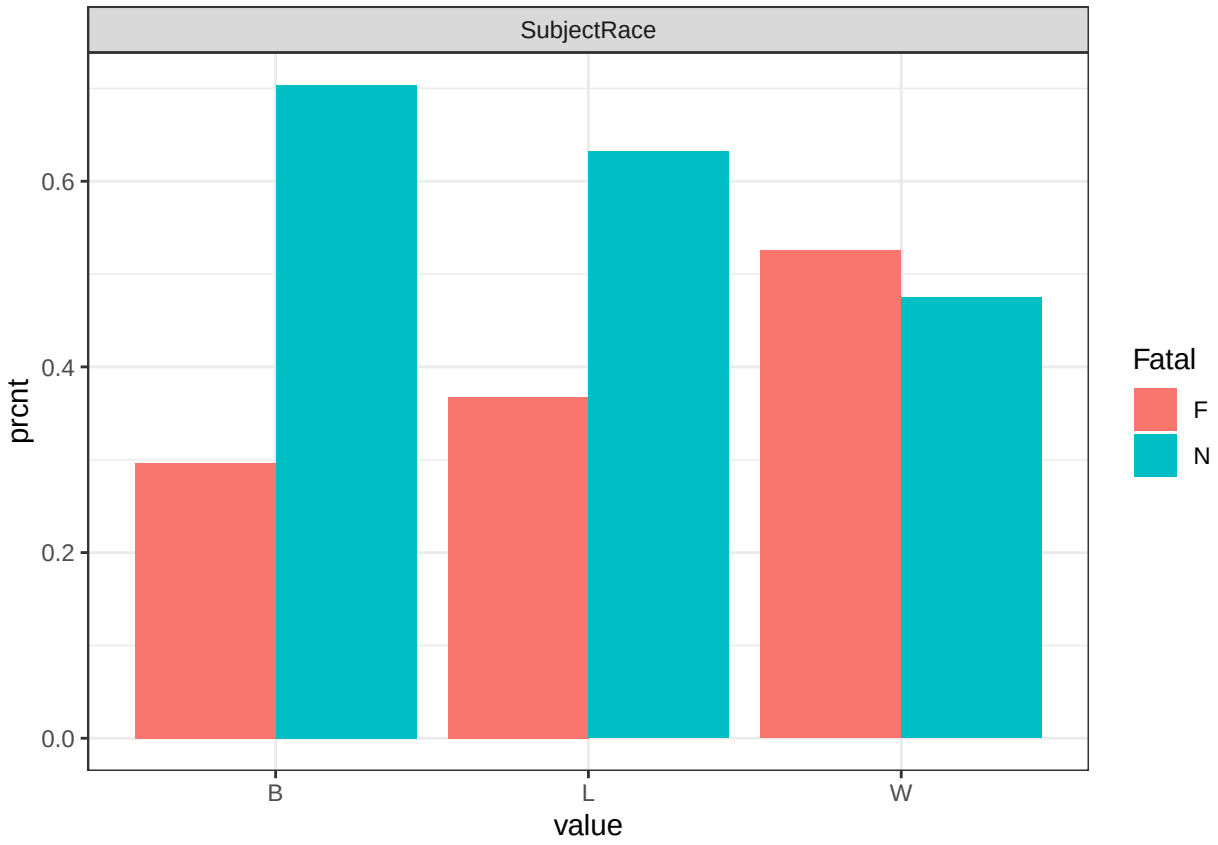
```
##
## [[2]]
```




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



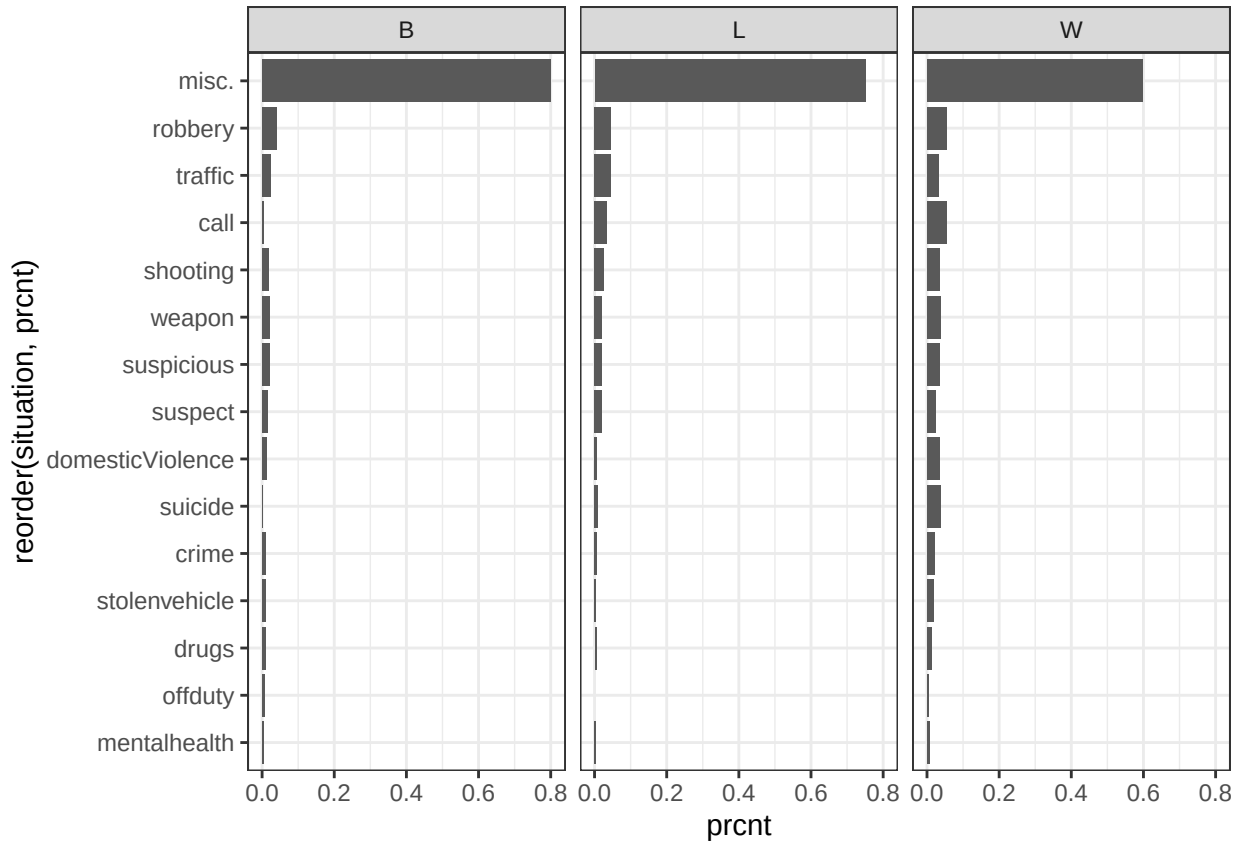

```
##
## [[4]]
```



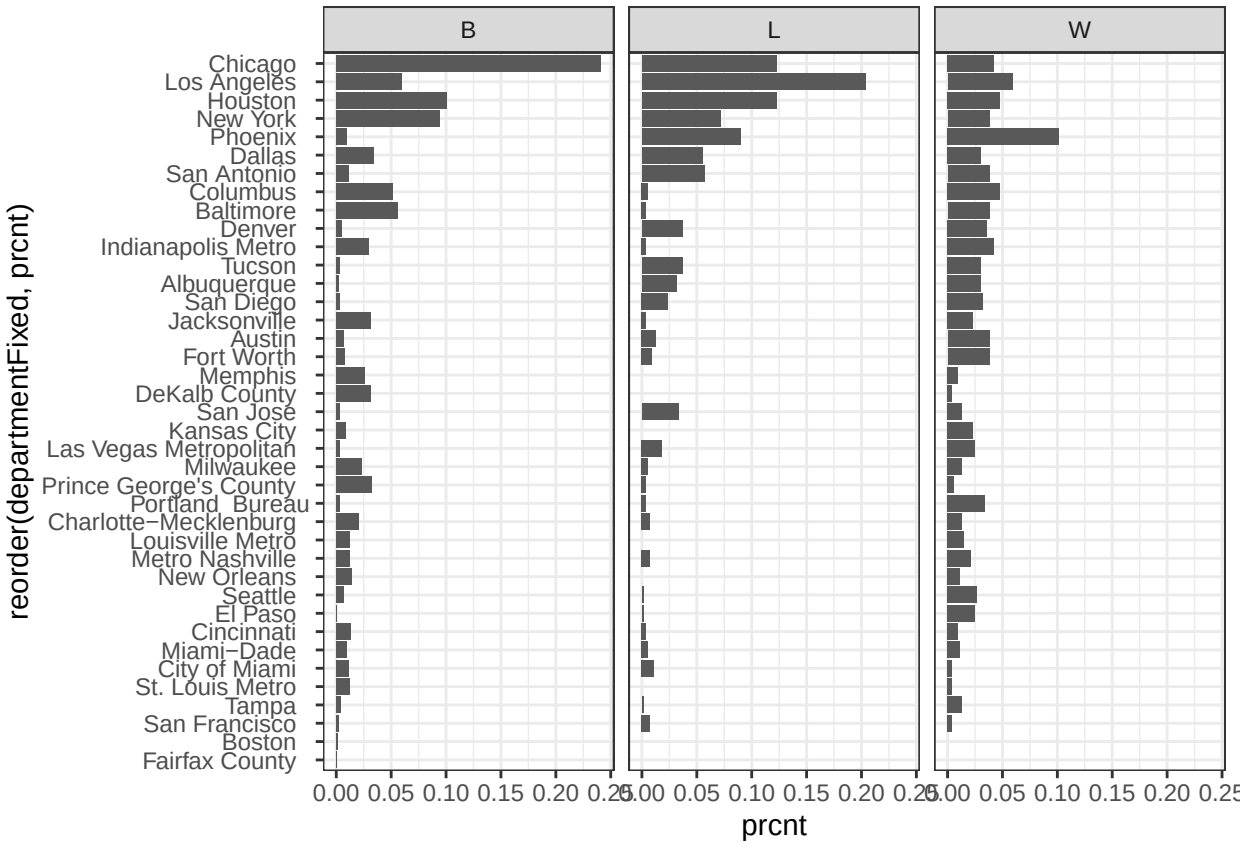
Bar charts for situation and department by race

```
temp <-
    viceRace %>%
    mutate(departmentFixed =
               str_replace_all(departmentFixed, "Police|Department", ""))

map(list("situation", "departmentFixed"), function(df, column){

    df %>%
        count(SubjectRace, .data[[column]]) %>%
        group_by(SubjectRace) %>%
        mutate(prcnt = n / sum(n)) %>%
        ungroup() %>%
        ggplot(aes(x = reorder(.data[[column]], prcnt), y = prcnt)) +
        geom_bar(stat = "identity") +
        facet_wrap(~SubjectRace) +
        theme_bw() +
        coord_flip()

}, df = temp)
```

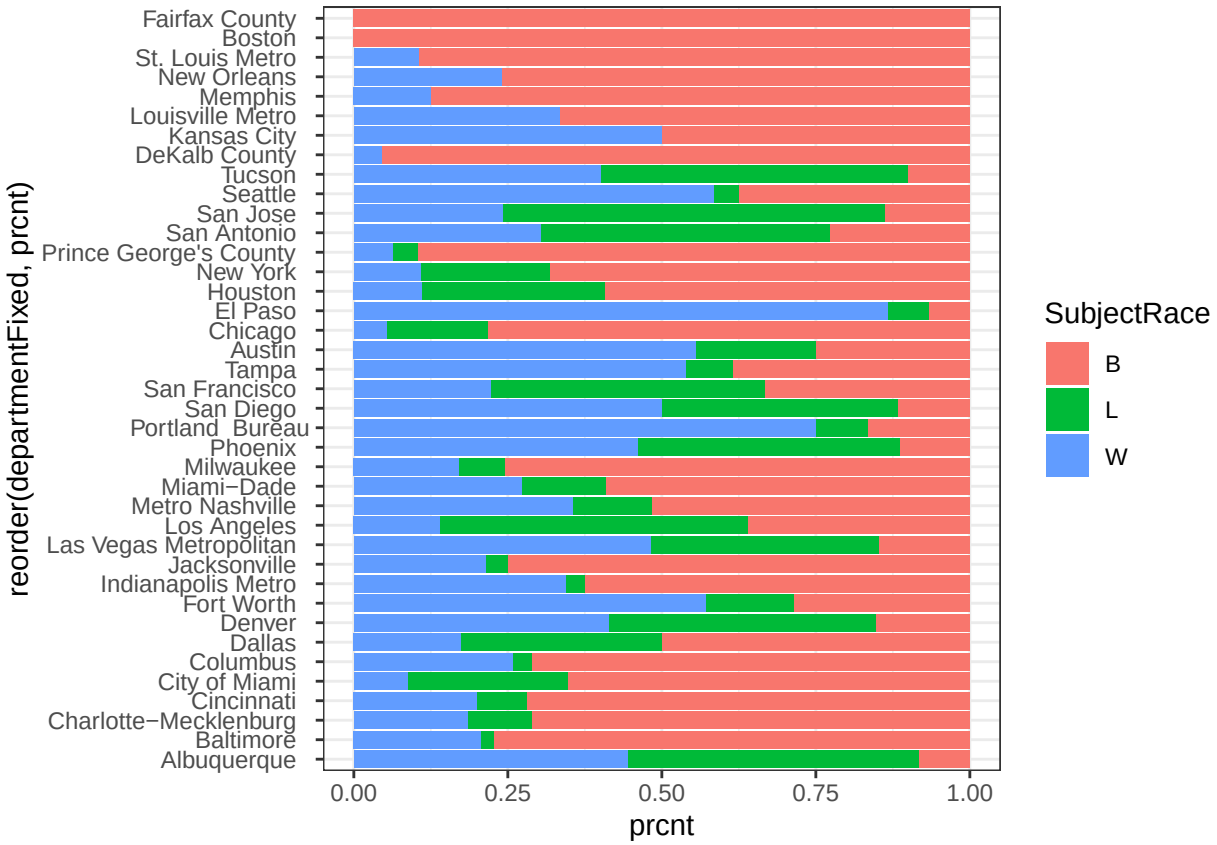
```
## [[1]]
```


B
L
W
misc.
robbery
traffic
call
shooting
weapon
suspicious
suspect
domesticViolence
suicide
crime
stolenvehicle
drugs
offduty
mentalhealth
reorder(situation, prcnt)
0.0
0.2
0.4
0.6
0.8
prcnt

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



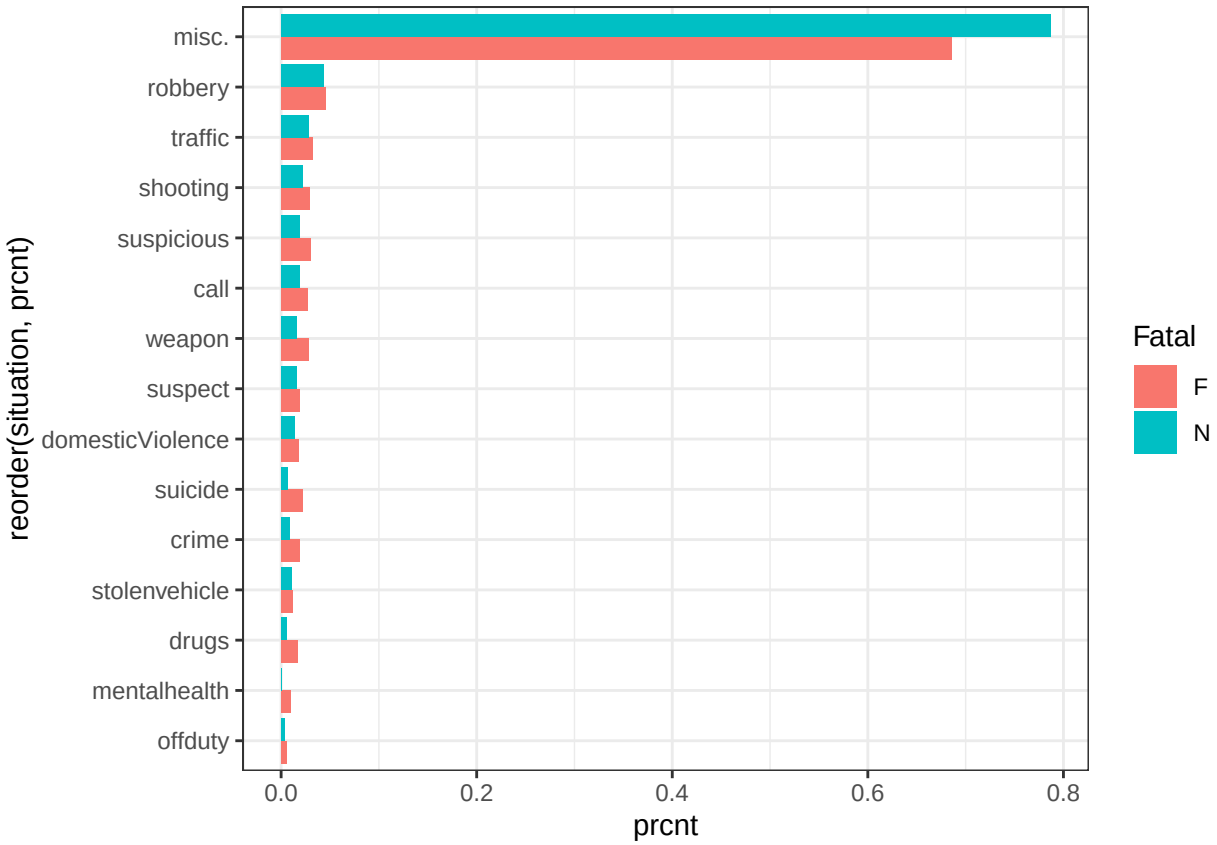

```
temp %>%
    count(SubjectRace, departmentFixed) %>%
    group_by(departmentFixed) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(departmentFixed, prcnt), y = prcnt)) +
    geom_bar(stat = "identity", aes(fill = SubjectRace)) +
    theme_bw() +
    coord_flip()
```



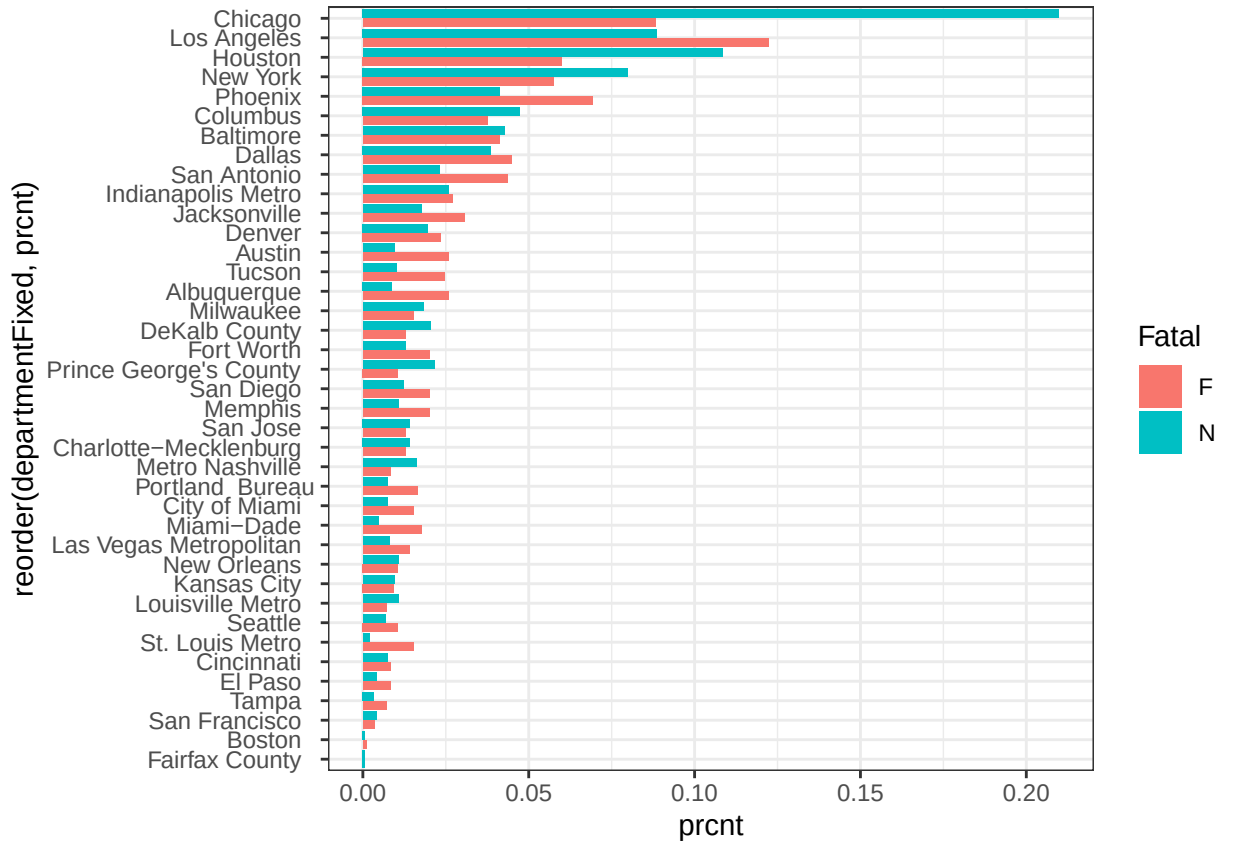
Bar charts for situation and department by non-fatal vs. fatal shooting

```
temp <-
    viceFatal %>%
    mutate(departmentFixed =
               str_replace_all(departmentFixed, "Police|Department", ""))

map(list("situation", "departmentFixed"), function(df, column){

    df %>%
        count(Fatal, .data[[column]]) %>%
        group_by(Fatal) %>%
        mutate(prcnt = n / sum(n)) %>%
        ungroup() %>%
        ggplot(aes(x = reorder(.data[[column]], prcnt), y = prcnt)) +
        geom_bar(stat = "identity", aes(fill = Fatal), position = "dodge") +
        theme_bw() +
        coord_flip()

}, df = temp)
```

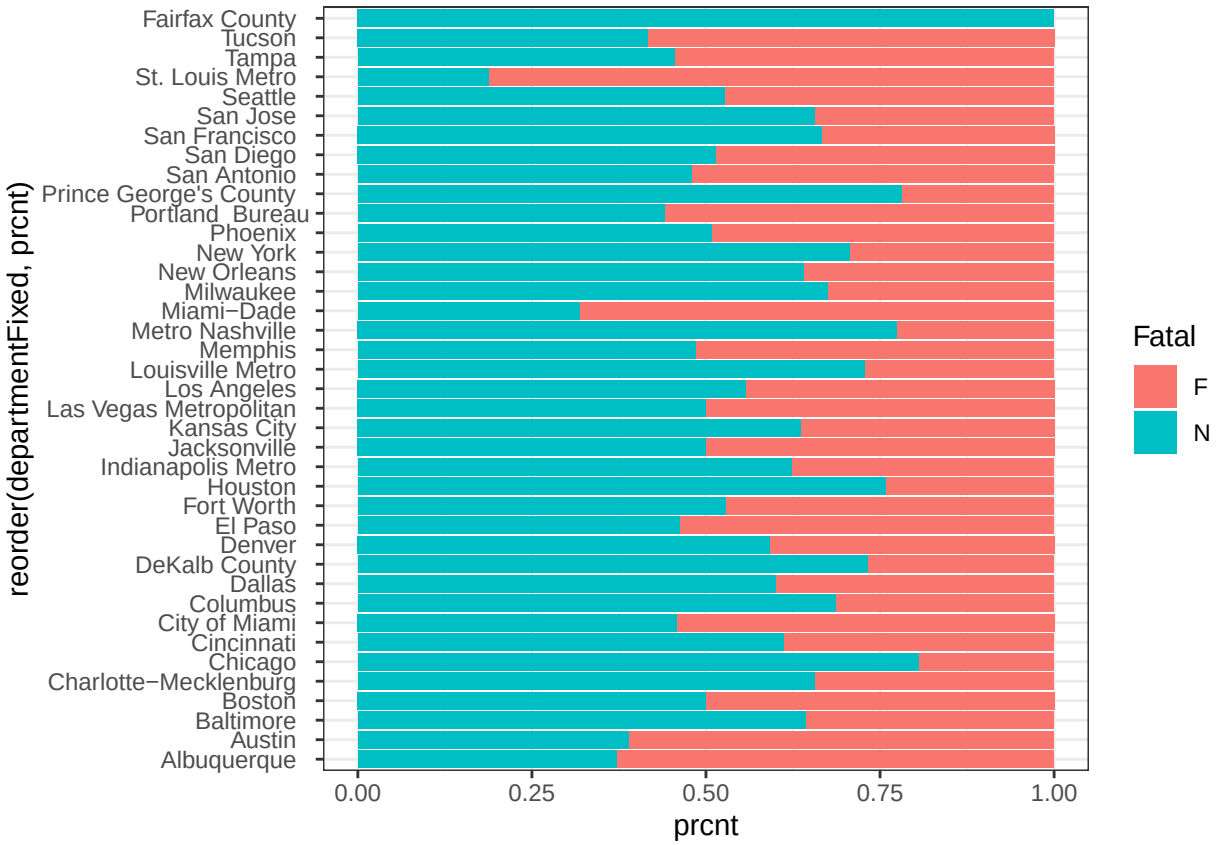
```
## [[1]]
```


misc.
robbery
traffic
shooting
suspicious
call
weapon
suspect
domesticViolence
suicide
crime
stolenvehicle
drugs
mentalhealth
offduty
reorder(situation, prcnt)
0.0
0.2
0.4
0.6
0.8
prcnt
Fatal
F
N

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



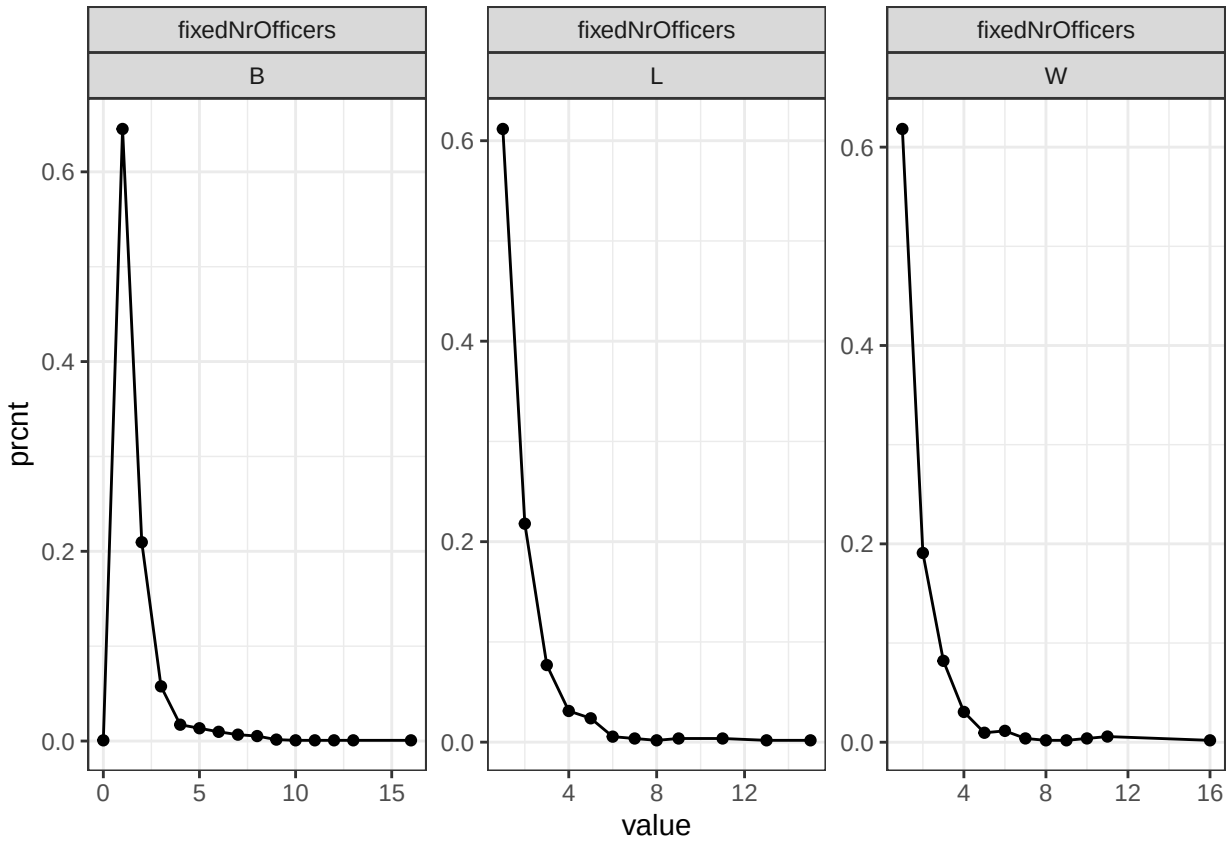

```
temp %>%
    count(Fatal, departmentFixed) %>%
    group_by(departmentFixed) %>%
    mutate(prcnt = n / sum(n)) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(departmentFixed, prcnt), y = prcnt)) +
    geom_bar(stat = "identity", aes(fill = Fatal)) +
    theme_bw() +
    coord_flip()
```



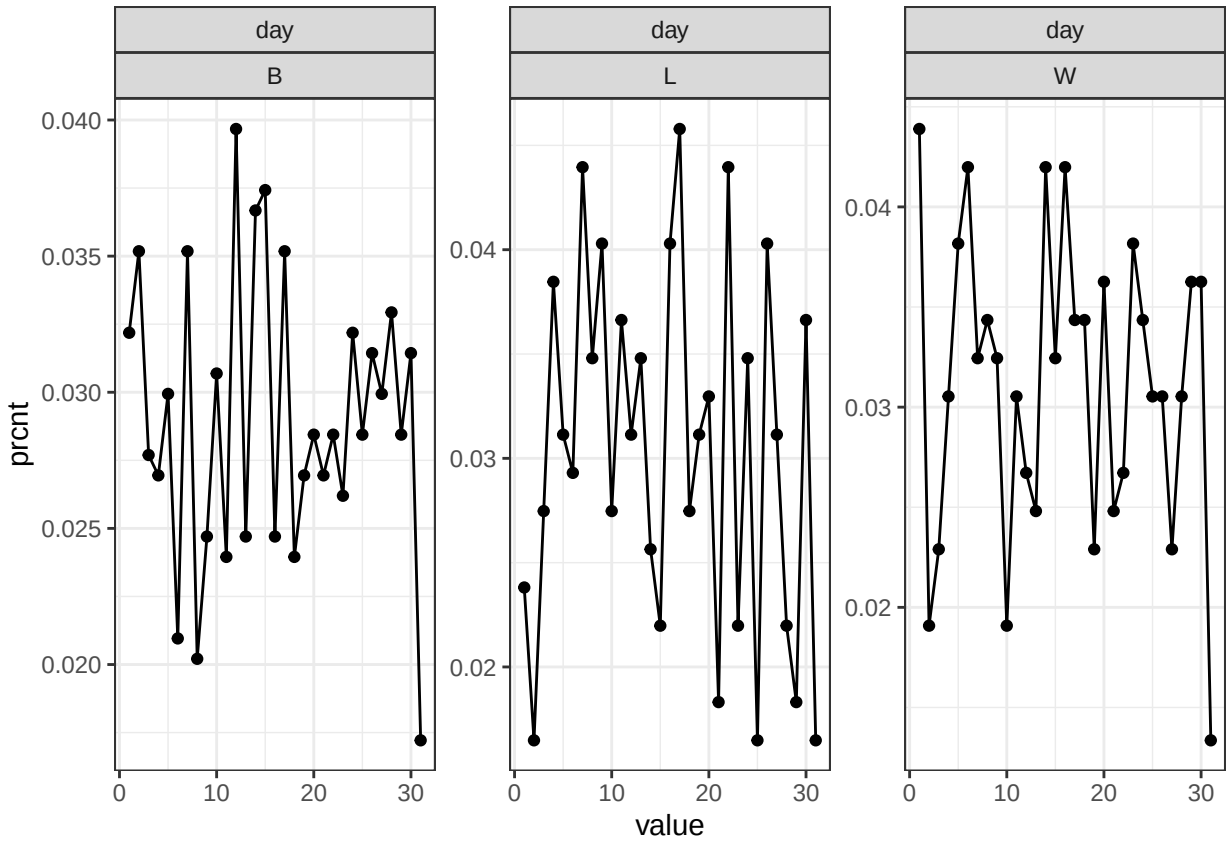
Line charts for numerical variables by race

```
categoriesRace <-
    viceRace %>%
    select(c(where(is.numeric), -matches("^officer")), SubjectRace) %>%
    pivot_longer(-SubjectRace) %>%
    count(SubjectRace, value, name) %>%
    group_by(SubjectRace, name) %>%
    mutate(prcnt = n / sum(n)) %>%
    filter(!is.na(value))

map(unique(categoriesRace$name),
    function(df, column) {
        df %>%
            filter(name == column) %>%
            ggplot(aes(x = value, y = prcnt)) +
            geom_line() +
            geom_point() +
            facet_wrap(~name+SubjectRace, scales = "free") +
            theme_bw()},
    df = categoriesRace)
```

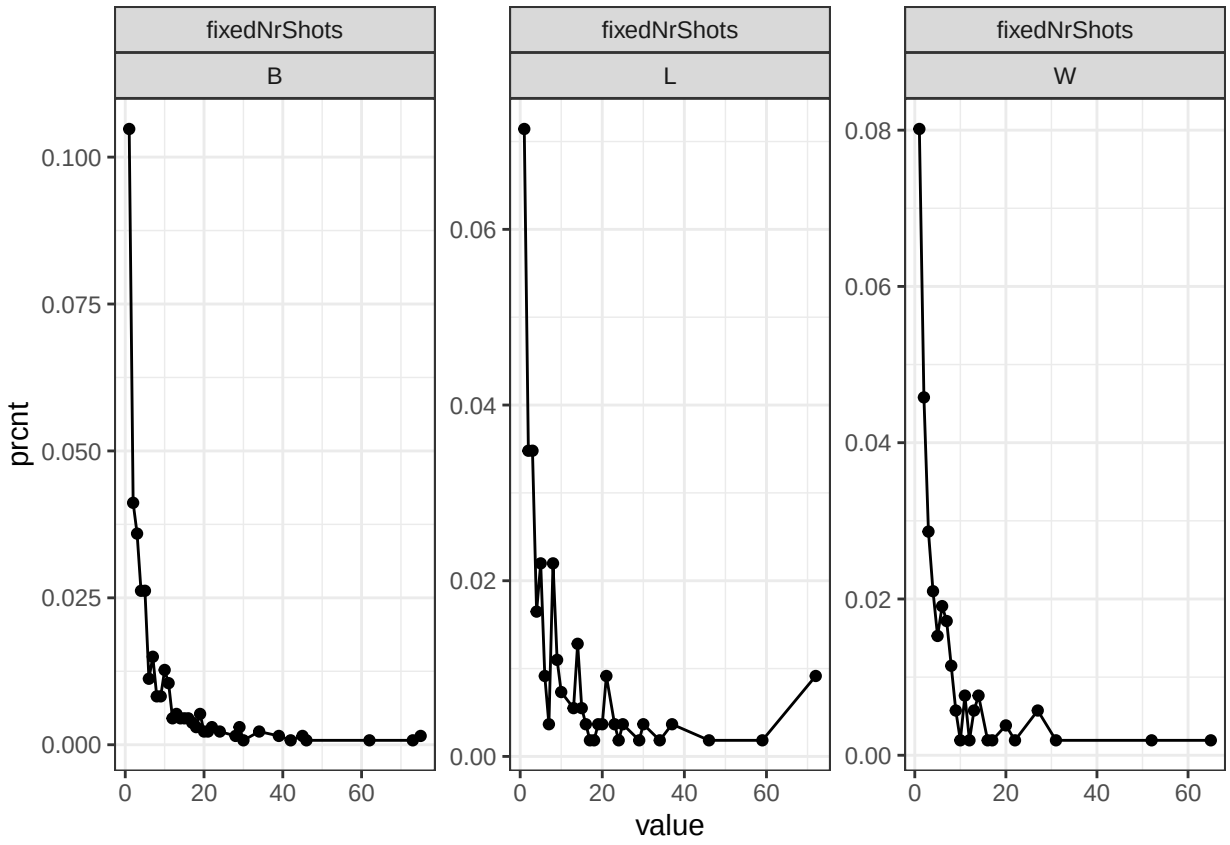
```
## [[1]]
```


fixedNrOfficers
B
fixedNrOfficers
L
fixedNrOfficers
W
prcnt
value
0.0
0.2
0.4
0.6
0
5
10
15
4
8
12
16

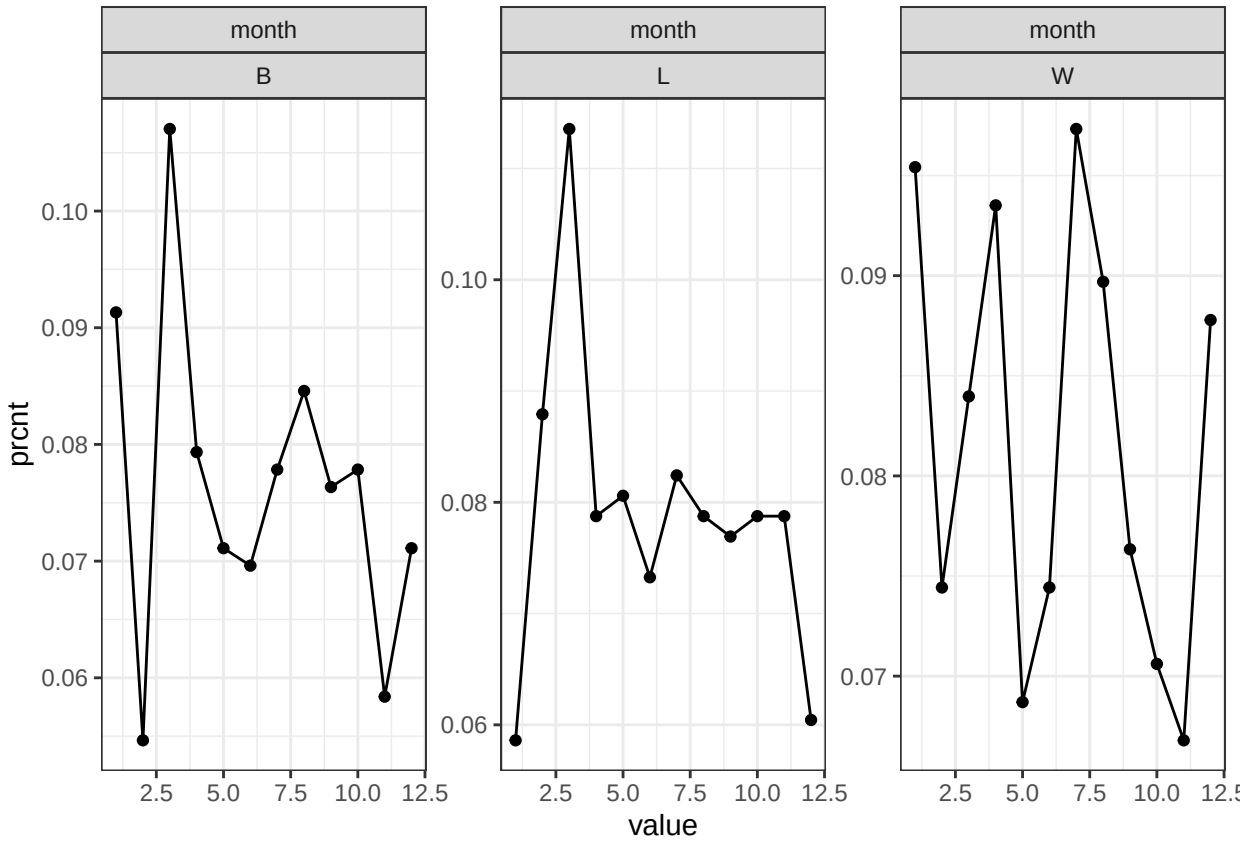
```
##
## [[2]]
```


day
B
L
W
prcnt
value
0.040
0.035
0.030
0.025
0.020
0.04
0.03
0.02
0
10
20
30

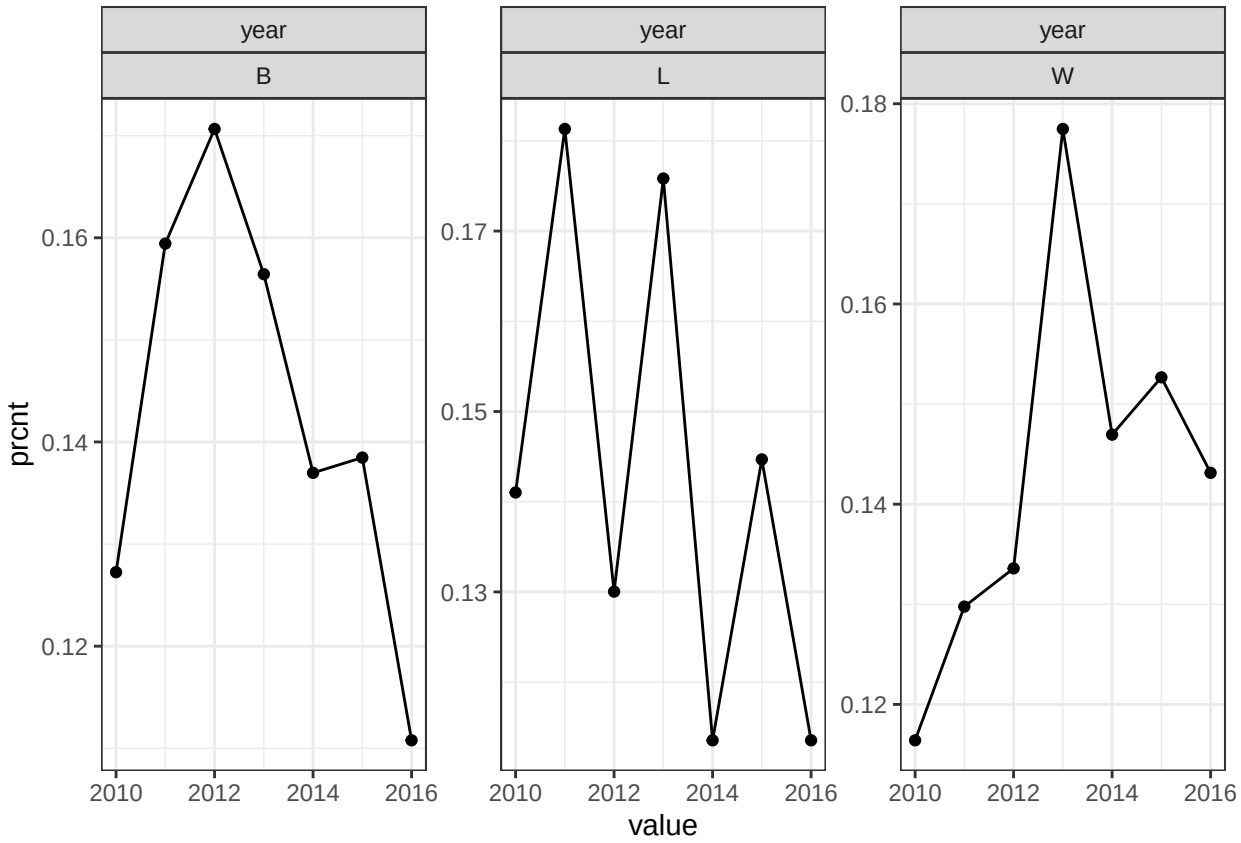
```
##
## [[3]]
```


fixedNrShots
B
fixedNrShots
L
fixedNrShots
W
prcnt
value
0.100
0.075
0.050
0.025
0.000
0.06
0.04
0.02
0.00
0.08
0
20
40
60

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



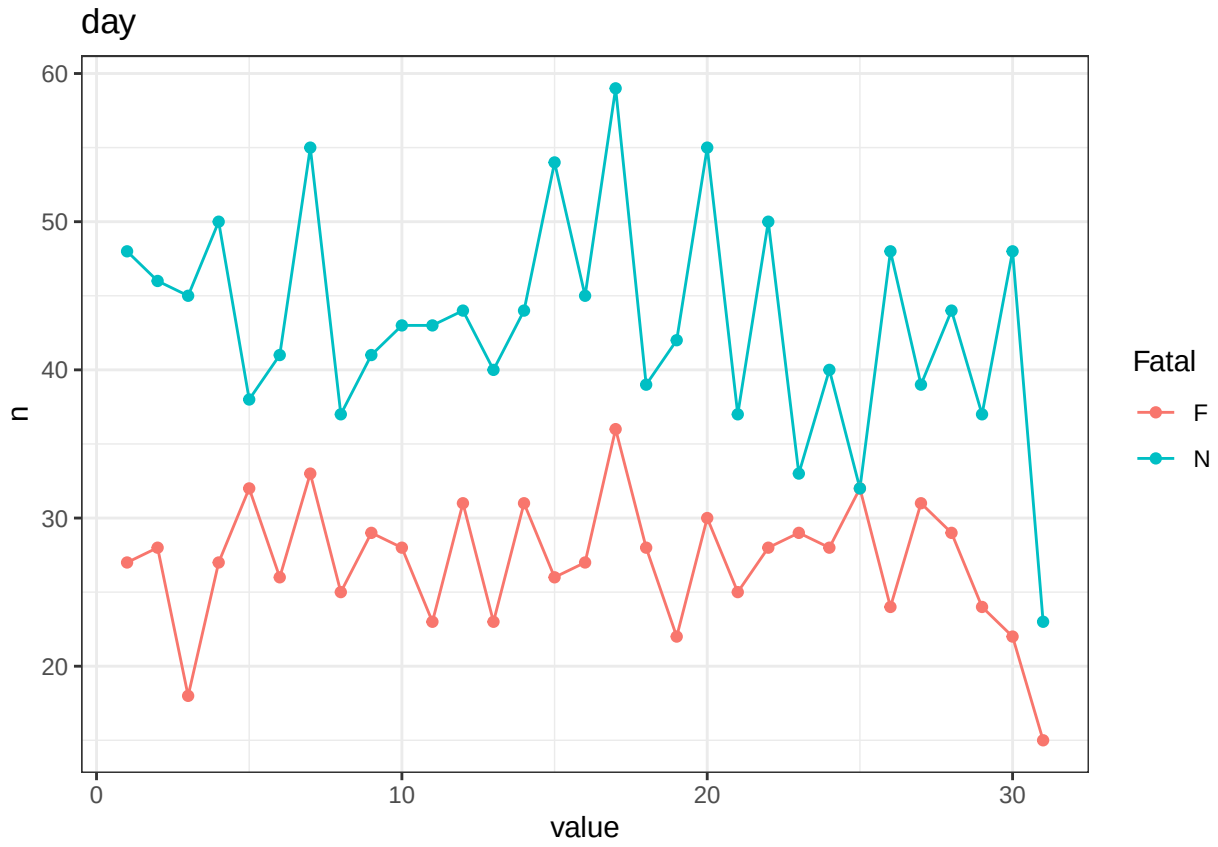

```
##
## [[5]]
```



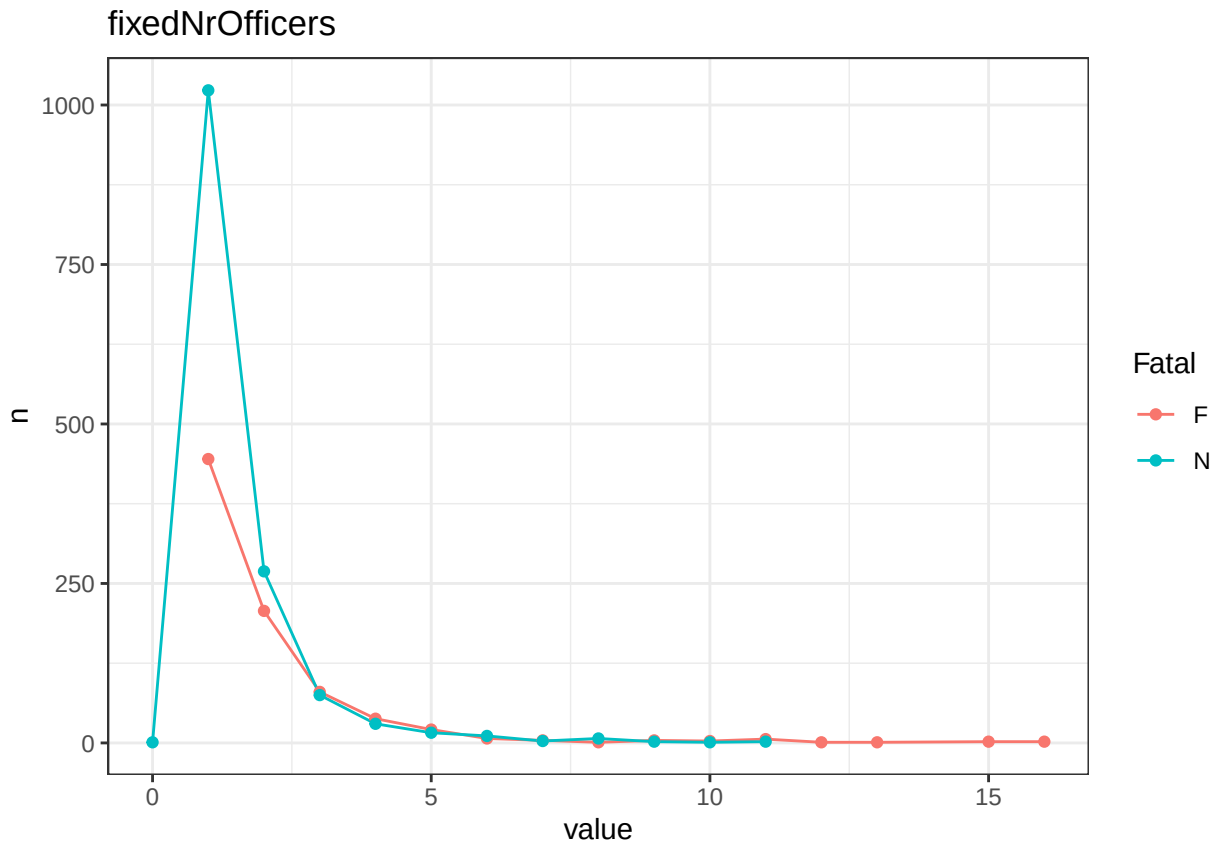
Line charts for numerical variables by fatal vs. non-fatal shooting

```
categoriesFatal <-
    viceFatal %>%
    select(c(where(is.numeric), -matches("^officer")), Fatal) %>%
    pivot_longer(-Fatal) %>%
    count(Fatal, value, name) %>%
    group_by(Fatal, name) %>%
    mutate(prcnt = n / sum(n)) %>%
    filter(!is.na(value))

map(unique(categoriesFatal$name),
    function(df, column) {
        df %>%
            filter(name == column) %>%
            ggplot(aes(x = value, y = n)) +
            geom_line(aes(color = Fatal, group = Fatal))+
            geom_point(aes(color = Fatal)) +
            #facet_wrap(~name+Fatal, scales = "free") +
            theme_bw() +
            labs(title = column)},
    df = categoriesFatal)
```

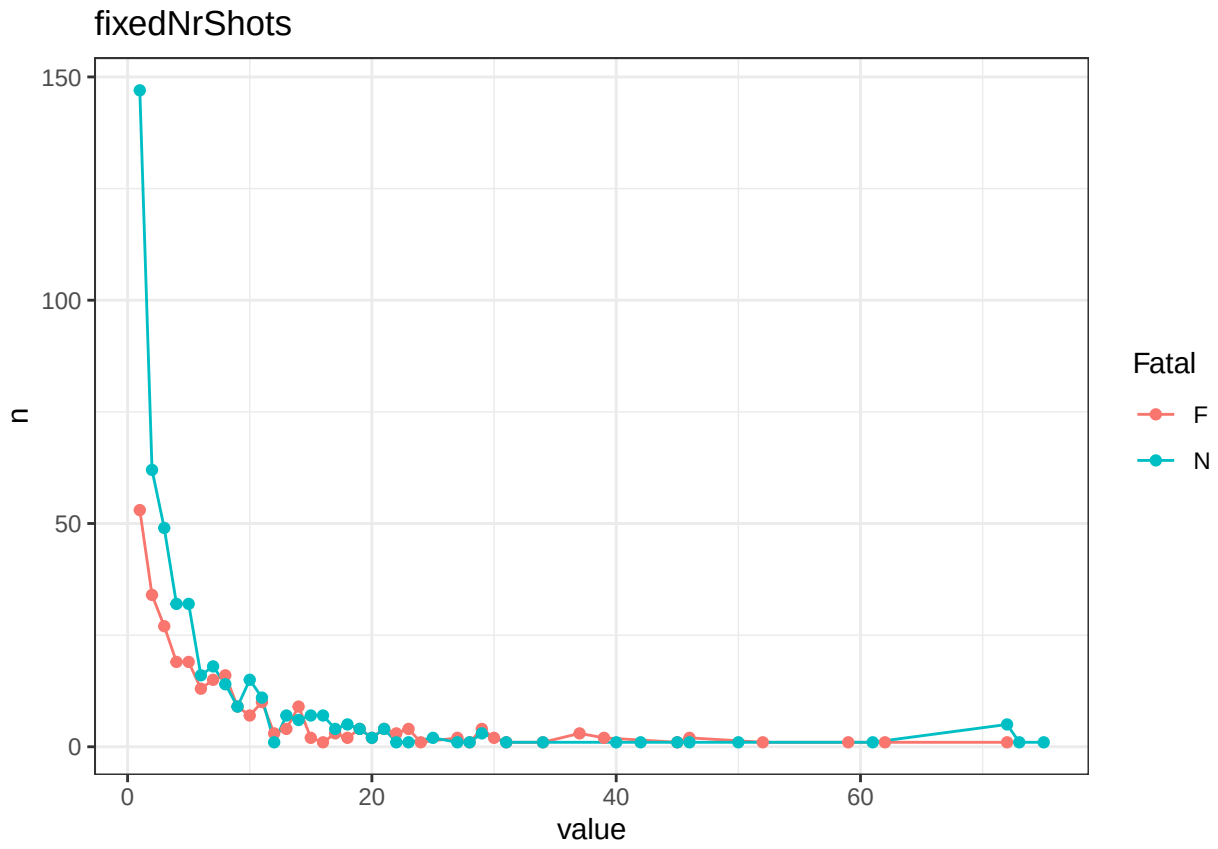
```
## [[1]]
```



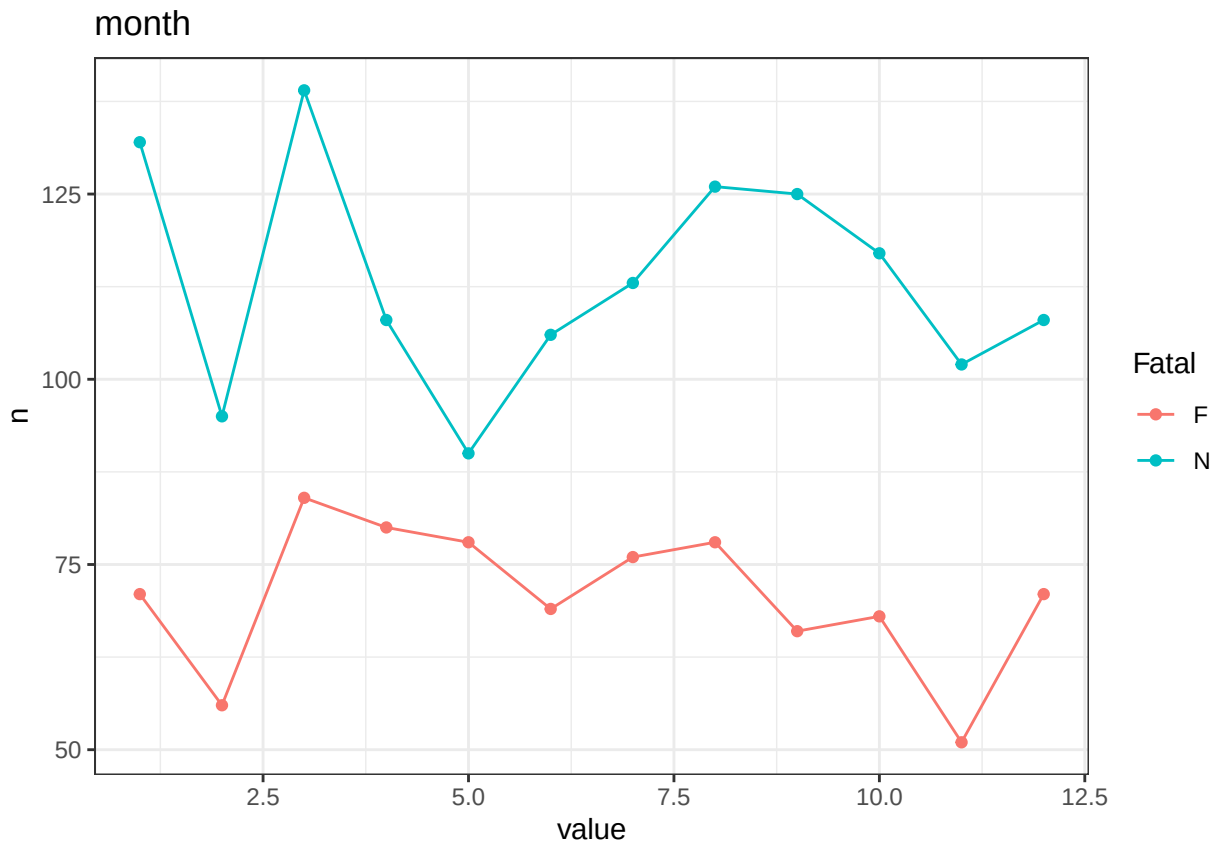

```
##
## [[2]]
```




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




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



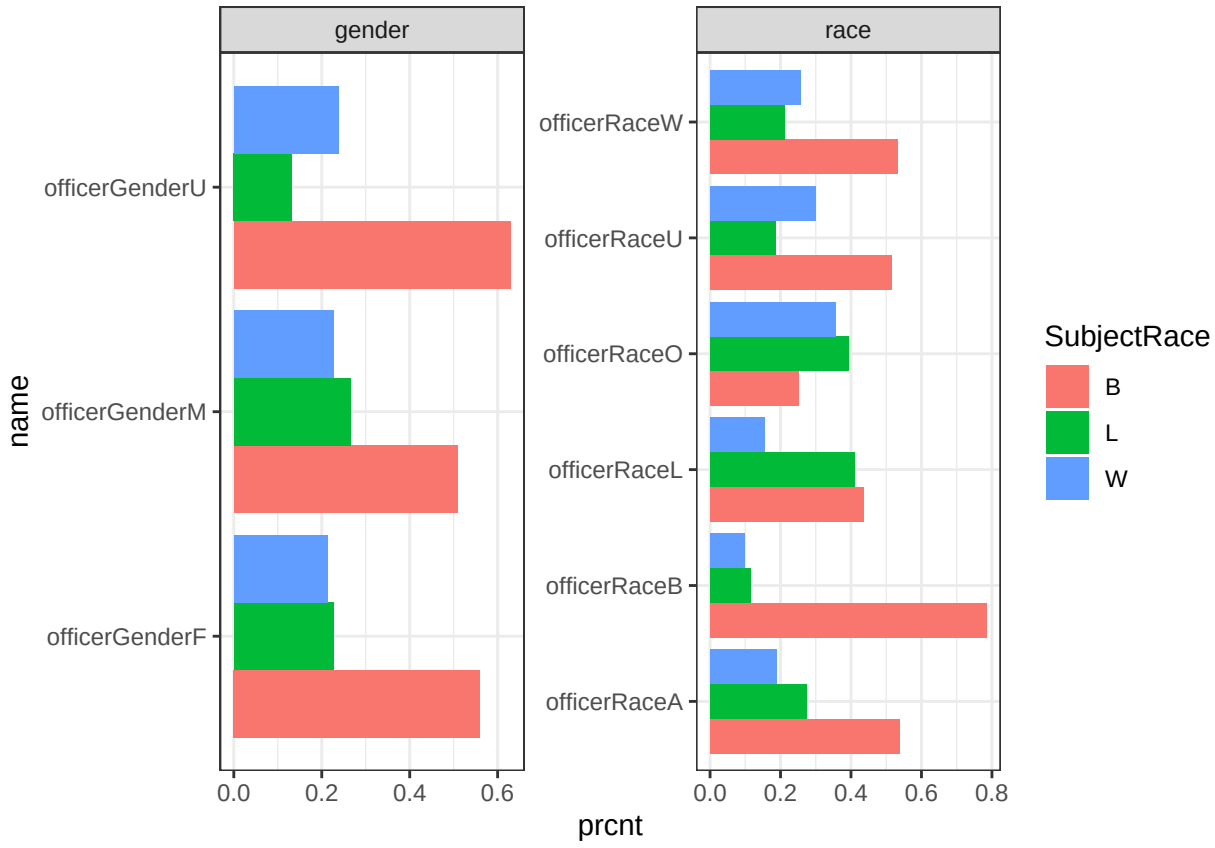

```
##
## [[5]]
```

Bar charts for officer gender and race BY subject race

- Female officers, proportionally, shoot slightly more Black subjects vs. male officers.
- All racial officer groups proportionally shoot Black offenders the most (except for the racial other category).
  - Interestingly though, for Black American officers, nearly 80% of all suspects shot at were Black.
  - A similar phenomenon occurs for Latino offficers where the proportion of supsects shot at dramatically increases for Latino suspects.

```
viceRace %>%
    select(c(matches("^officer")), SubjectRace) %>%
    pivot_longer(-SubjectRace) %>%
    group_by(SubjectRace, name) %>%
    summarise(officerSum = sum(value)) %>%
    group_by(name) %>%
    mutate(prcnt = officerSum / sum(officerSum)) %>%
    ungroup() %>%
    mutate(group = if_else(str_detect(name, "Gender"), "gender", "race")) %>%
    ggplot(aes(x = name, y = prcnt)) +
    geom_bar(stat = "identity", position = "dodge", aes(fill = SubjectRace)) +
    facet_wrap(~group, scales = "free") +
    theme_bw() +
    coord_flip()
```

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

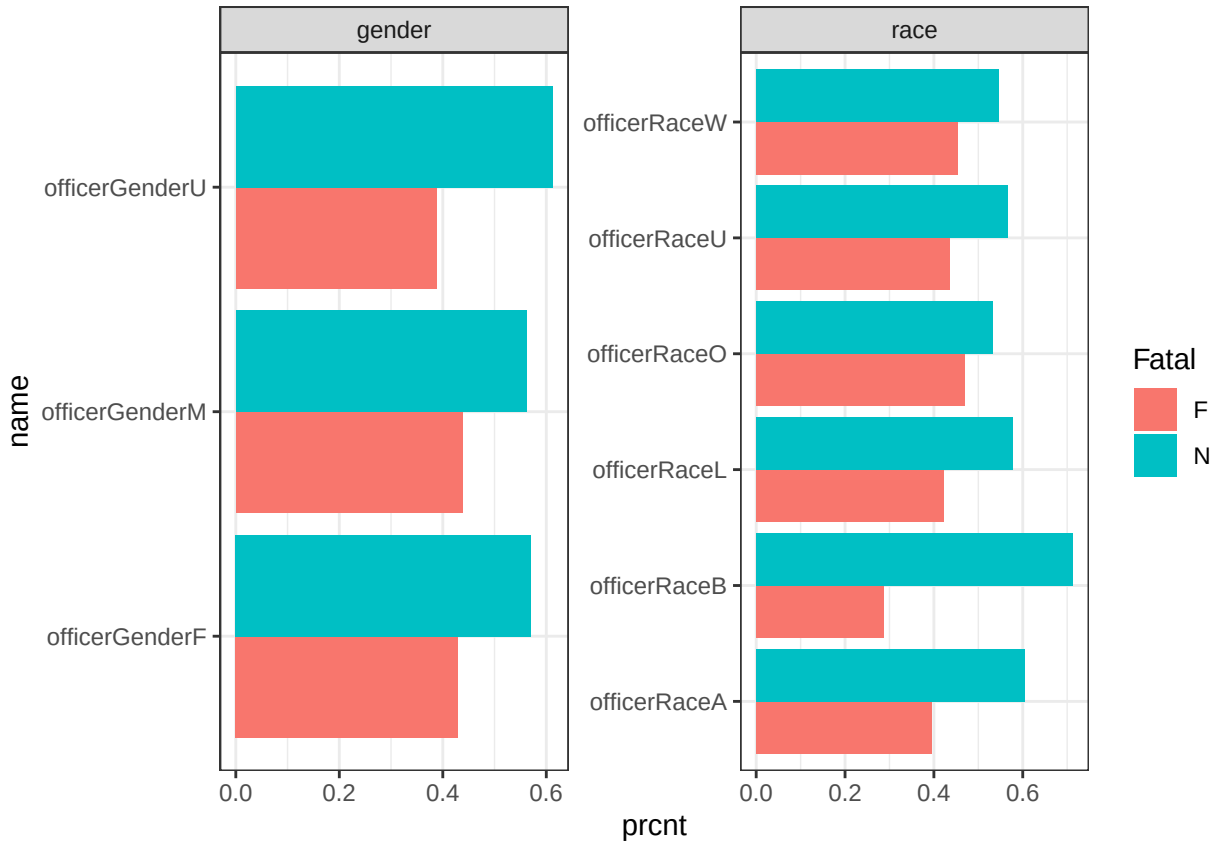


Bar charts for officer gender and race BY fatal vs. non-fatal shooting

- Female officers have a slighlty, slightly lower proportion of shootings being fatal.
- Minority police officers (Asian, Latino, Black-American) have smaller proportions of fatal shootings relative to White police officers.

```
viceFatal %>%
    select(c(matches("^officer")), Fatal) %>%
    pivot_longer(-Fatal) %>%
    group_by(Fatal, name) %>%
    summarise(officerSum = sum(value)) %>%
    group_by(name) %>%
    mutate(prcnt = officerSum / sum(officerSum)) %>%
    ungroup() %>%
    mutate(group = if_else(str_detect(name, "Gender"), "gender", "race")) %>%
    ggplot(aes(x = name, y = prcnt)) +
    geom_bar(stat = "identity", position = "dodge", aes(fill = Fatal)) +
    facet_wrap(~group, scales = "free") +
    theme_bw() +
    coord_flip()
```

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


gender
race
officerGenderU
officerGenderM
officerGenderF
officerRaceW
officerRaceU
officerRaceO
officerRaceL
officerRaceB
officerRaceA
name
0.0
0.2
0.4
0.6
prcnt
Fatal
F
N