R Notebook

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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

```
library(magrittr)
```

```
##
## Vedhæfter pakke: 'magrittr'

## Det følgende objekt er maskeret fra 'package:purrr':
##
##     set_names

## Det følgende objekt er maskeret fra 'package:tidyr':
##
##     extract
```

```
data = readRDS(url("https://stacks.stanford.edu/file/druid:yg821jf8611/yg821jf8611_ca_san_francisco_202
```

data is now loaded!

```
glimpse(data)
```

```
## Rows: 905,070
## Columns: 22
## $ raw_row_number                  <chr> "869921", "869922", "869923", "86992~
## $ date                            <date> 2014-08-01, 2014-08-01, 2014-08-01,~
## $ time                            <time> 00:01:00, 00:01:00, 00:15:00, 00:18~
## $ location                        <chr> "MASONIC AV & FELL ST", "GEARY&10TH ~
## $ lat                             <dbl> 37.77300, 37.78090, 37.78692, 37.746~
## $ lng                             <dbl> -122.4459, -122.4686, -122.4267, -12~
## $ district                        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ subject_age                     <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
## $ subject_race                     <fct> asian/pacific islander, black, hispa~
## $ subject_sex                      <fct> female, male, male, male, male, male~
## $ type                             <fct> vehicular, vehicular, vehicular, veh~
## $ arrest_made                      <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, F~
## $ citation_issued                  <lgl> FALSE, TRUE, TRUE, FALSE, TRUE, TRUE~
## $ warning_issued                   <lgl> TRUE, FALSE, FALSE, TRUE, FALSE, FAL~
## $ outcome                          <fct> warning, citation, citation, warning~
## $ contraband_found                 <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ search_conducted                 <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, F~
## $ search_vehicle                   <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, F~
## $ search_basis                     <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ reason_for_stop                  <chr> "Mechanical or Non-Moving Violation ~
## $ raw_search_vehicle_description   <chr> "No Search", "No Search", "No Search~
## $ raw_result_of_contact_description <chr> "Warning", "Citation", "Citation", "~
```

Tidying the data

```
data %>%
  drop_na(subject_race)
```

```
## # A tibble: 905,070 x 22
##    raw_row_number date       time   location      lat   lng district subject_age
##    <chr>          <date>     <time> <chr>       <dbl> <dbl> <chr>          <int>
##  1 869921         2014-08-01 00:01  MASONIC AV~  37.8 -122. <NA>              NA
##  2 869922         2014-08-01 00:01  GEARY&10TH~  37.8 -122. <NA>              NA
##  3 869923         2014-08-01 00:15  SUTTER N O~  37.8 -122. <NA>              NA
##  4 869924         2014-08-01 00:18  3RD ST & D~  37.7 -122. <NA>              NA
##  5 869925         2014-08-01 00:19  DIVISADERO~  37.8 -122. <NA>              NA
##  6 869926         2014-08-01 00:30  3RD ST AND~  37.7 -122. <NA>              NA
##  7 869927         2014-08-01 00:30  COLUMBUS A~  37.8 -122. <NA>              NA
##  8 869928         2014-08-01 00:35  INGALLS & ~  37.7 -122. <NA>              NA
##  9 869929         2014-08-01 01:00  17TH ST & ~  37.8 -122. <NA>              NA
## 10 869930         2014-08-01 01:00  FULTON ST.~  37.8 -122. <NA>              NA
## # ... with 905,060 more rows, and 14 more variables: subject_race <fct>,
## #   subject_sex <fct>, type <fct>, arrest_made <lgl>, citation_issued <lgl>,
## #   warning_issued <lgl>, outcome <fct>, contraband_found <lgl>,
## #   search_conducted <lgl>, search_vehicle <lgl>, search_basis <fct>,
## #   reason_for_stop <chr>, raw_search_vehicle_description <chr>,
## #   raw_result_of_contact_description <chr>
```

```
data_plot = data %>%
  group_by(subject_race, search_conducted) %>%
  summarise(n = n())
```

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

```
data_plot
```

```
## # A tibble: 10 x 3
## # Groups:   subject_race [5]
##    subject_race           search_conducted      n
##    <fct>                  <lgl>             <int>
```

```
##  1 asian/pacific islander FALSE            154823
##  2 asian/pacific islander TRUE               2861
##  3 black                  FALSE            128574
##  4 black                  TRUE              23622
##  5 hispanic               FALSE            104569
##  6 hispanic               TRUE              11445
##  7 white                  FALSE            360611
##  8 white                  TRUE              11707
##  9 other                  FALSE            103112
## 10 other                  TRUE               3746
```

```
data_plot %>%
  ggplot(aes(x = subject_race, y = n, fill = search_conducted)) + geom_col(position = "fill")
```