# ASSIGNMENT 4

Kimberly Cable

2022-04-23

## Assignment 4 - Part 1

**Set the working directory to the root of your DSC 520 directory**

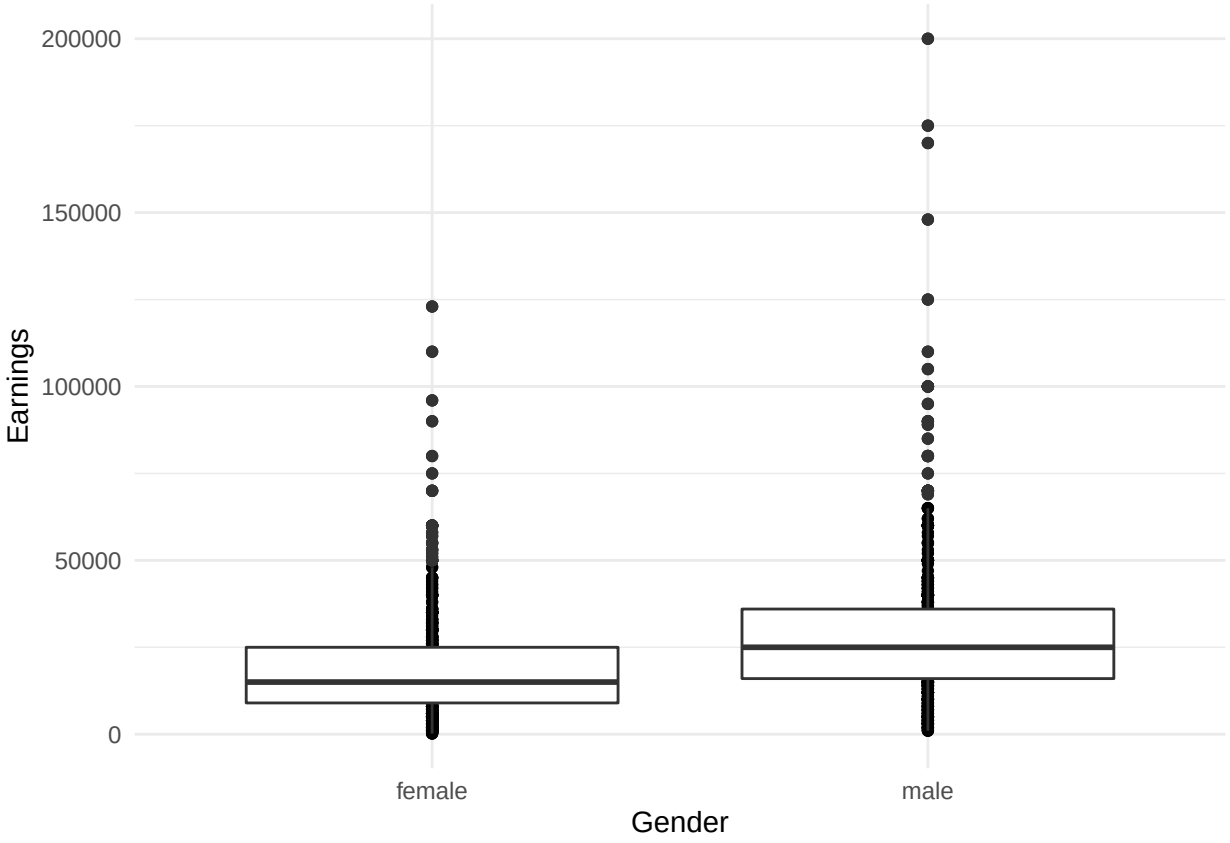
**Load the `data/r4ds/heights.csv` to**

```
##    earn   height    sex ed age  race
## 1 50000 74.42444   male 16  45 white
## 2 60000 65.53754 female 16  58 white
## 3 30000 63.62920 female 16  29 white
## 4 50000 63.10856 female 16  91 other
## 5 51000 63.40248 female 17  39 white
## 6  9000 64.39951 female 15  26 white
```

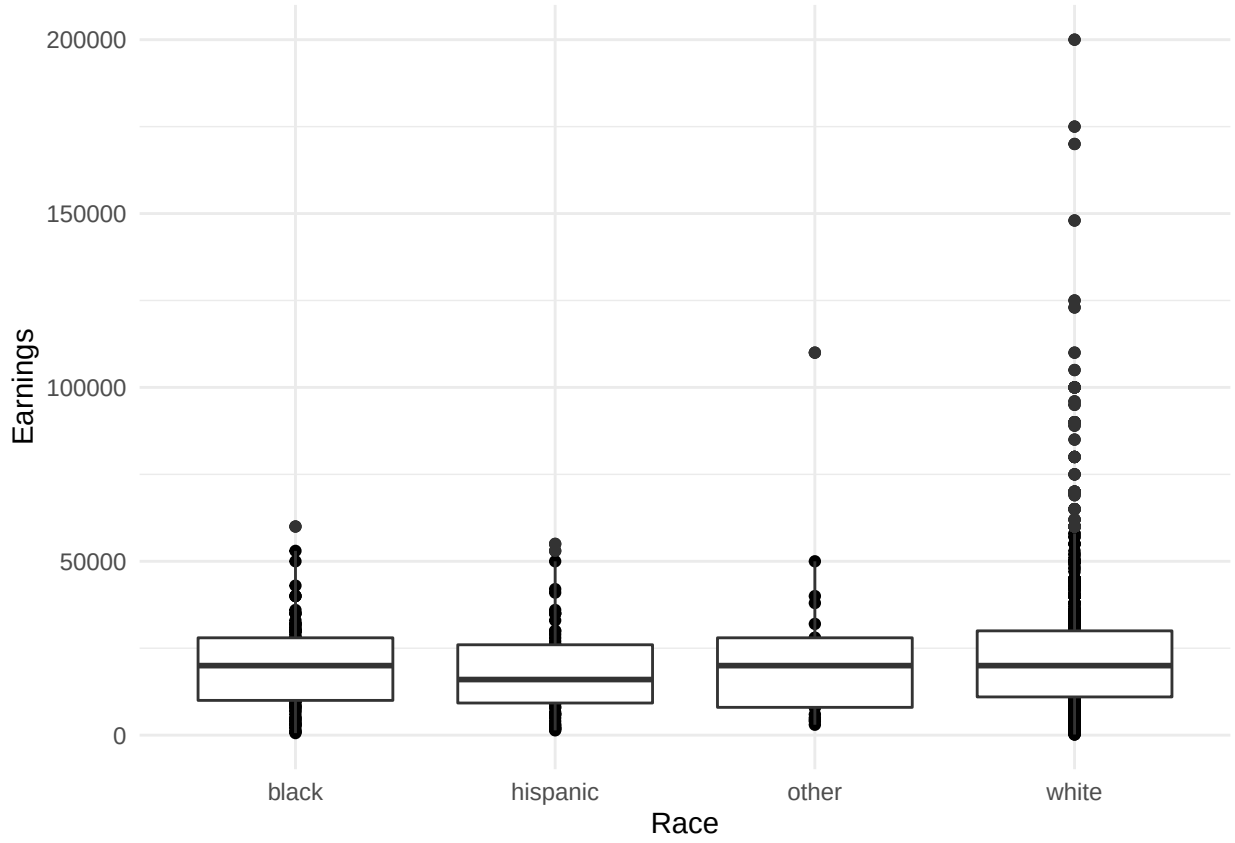
## https://ggplot2.tidyverse.org/reference/geom_boxplot.html

**Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`**

**sex vs. earn**

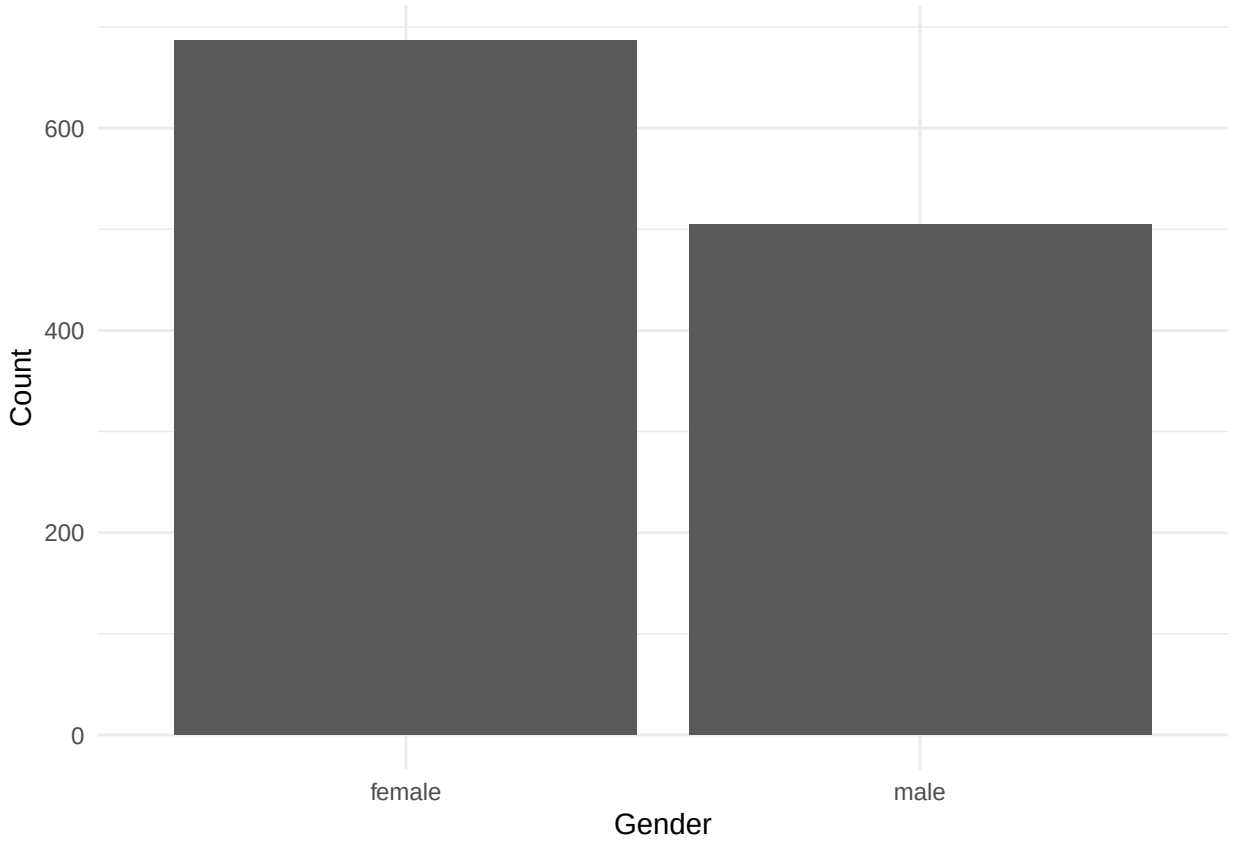


**race vs. earn**

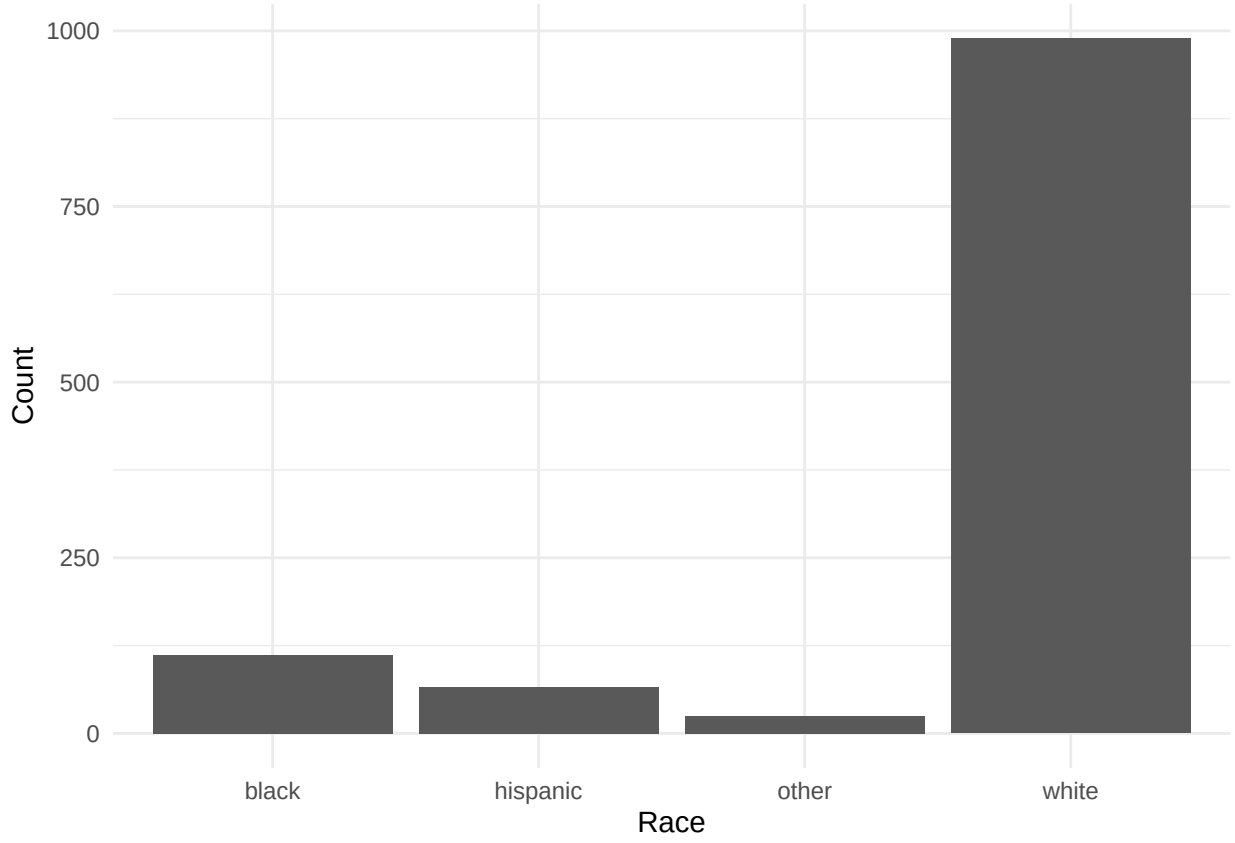

200000
150000
100000
50000
0
Earnings
black
hispanic
other
white
Race

## https://ggplot2.tidyverse.org/reference/geom_bar.html

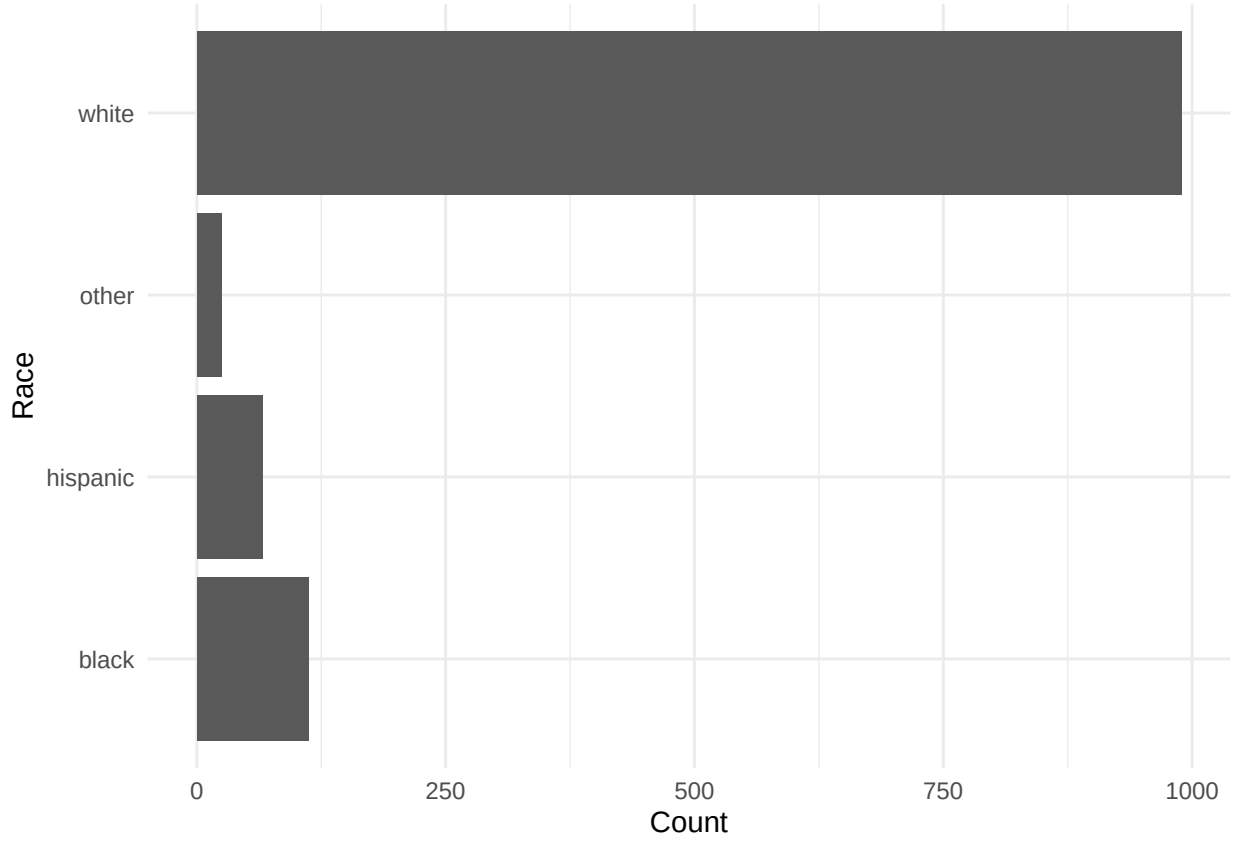
**Using `geom_bar()` plot a bar chart of the number of records for each `sex`**



**Using `geom_bar()` plot a bar chart of the number of records for each race**



**Create a horizontal bar chart by adding `coord_flip()` to the previous plot**




## https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0/topics/geom_path

**Load the file `"data/nytimes/covid-19-data/us-states.csv"` and**

**assign it to the `covid_df` dataframe**

```
##         date      state fips cases deaths
## 1 2020-01-21 Washington   53     1      0
## 2 2020-01-22 Washington   53     1      0
## 3 2020-01-23 Washington   53     1      0
## 4 2020-01-24   Illinois   17     1      0
## 5 2020-01-24 Washington   53     1      0
## 6 2020-01-25 California    6     1      0
```

**Parse the date column using `as.Date()`**

```
##         date      state fips cases deaths
## 1 2020-01-21 Washington   53     1      0
## 2 2020-01-22 Washington   53     1      0
```

```
## 3 2020-01-23 Washington   53     1      0
## 4 2020-01-24   Illinois   17     1      0
## 5 2020-01-24 Washington   53     1      0
## 6 2020-01-25 California    6     1      0
```

**Create three dataframes named `california_df`, `ny_df`, and `florida_df` containing the data from California, New York, and Florida**

## California

```
##          date      state fips cases deaths
## 6  2020-01-25 California    6     1      0
## 10 2020-01-26 California    6     2      0
## 14 2020-01-27 California    6     2      0
## 18 2020-01-28 California    6     2      0
## 22 2020-01-29 California    6     2      0
## 26 2020-01-30 California    6     2      0
```

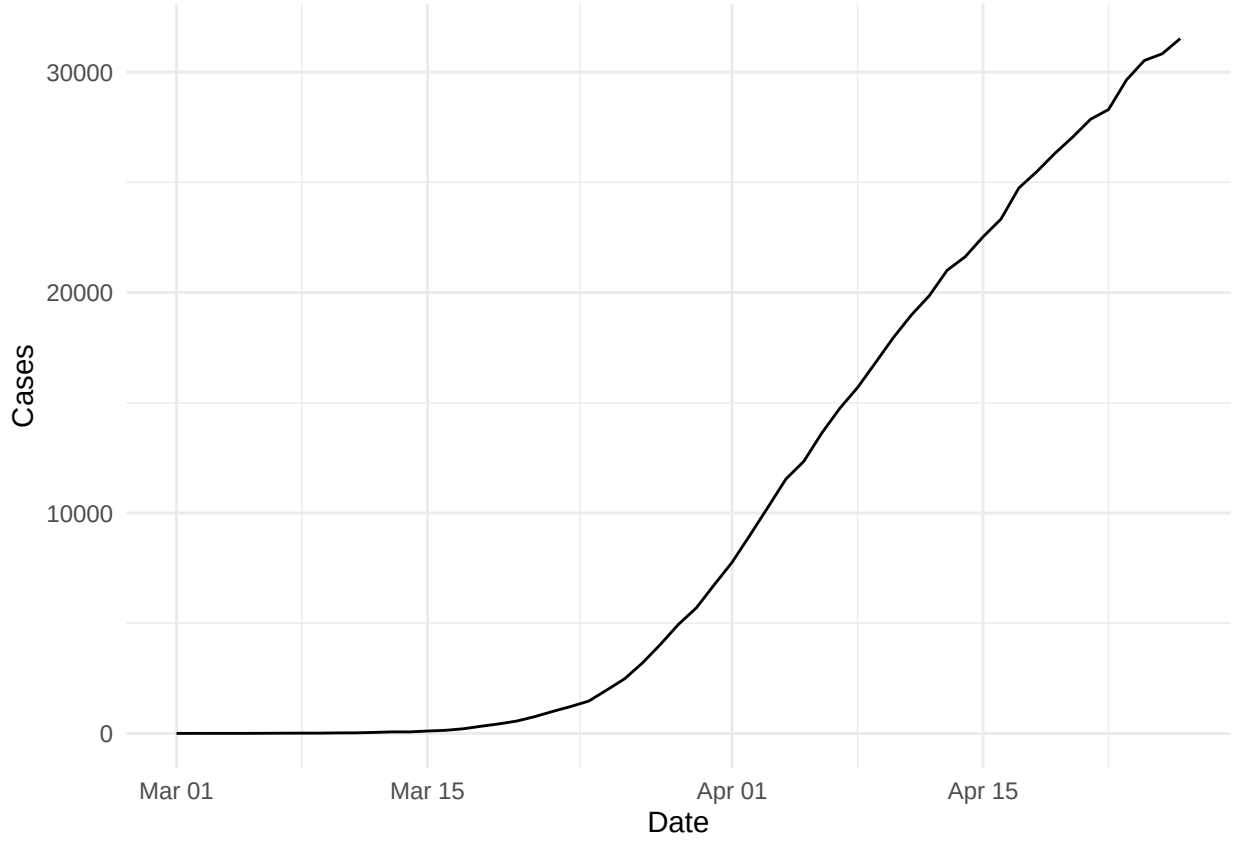
## New York

```
##           date    state fips cases deaths
## 247 2020-03-01 New York   36     1      0
## 262 2020-03-02 New York   36     1      0
## 277 2020-03-03 New York   36     2      0
## 294 2020-03-04 New York   36    11      0
## 314 2020-03-05 New York   36    22      0
## 339 2020-03-06 New York   36    44      0
```

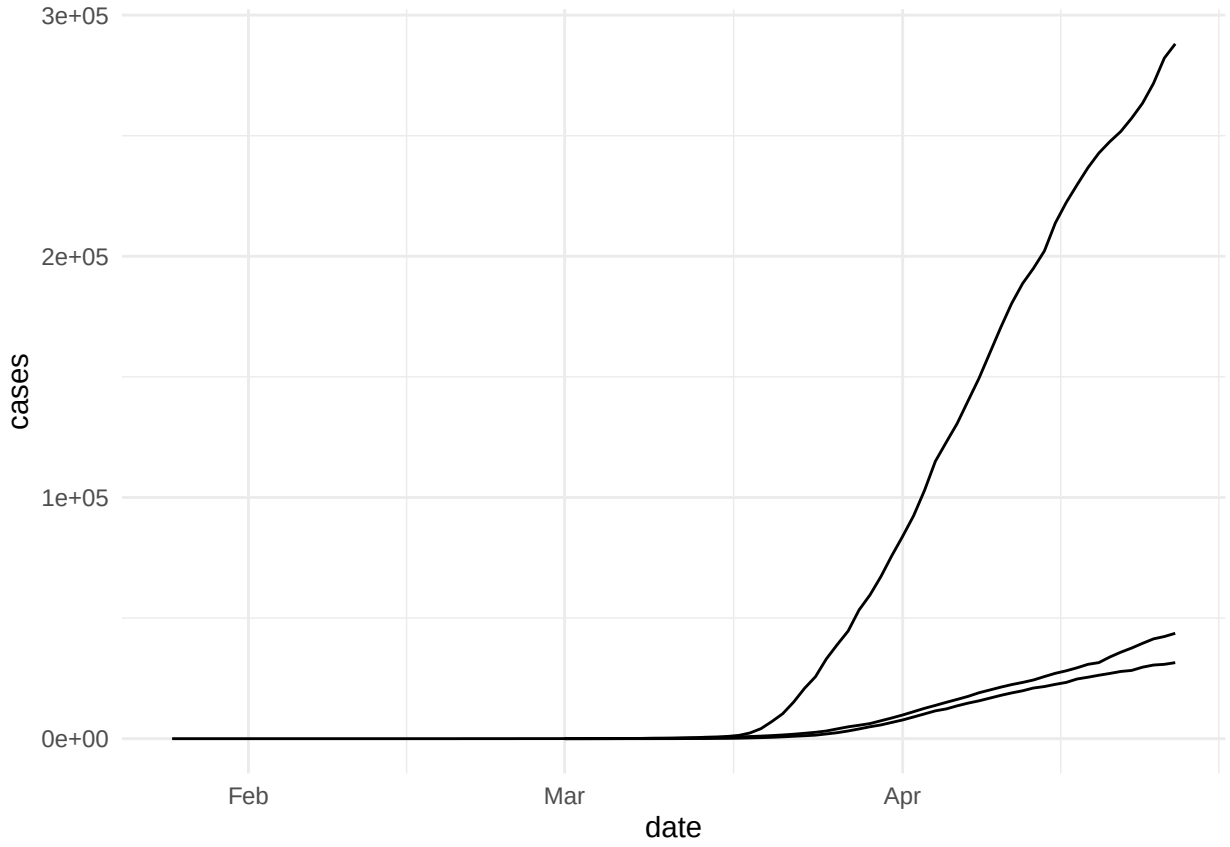
## Florida

```
##           date   state fips cases deaths
## 243 2020-03-01 Florida   12     2      0
## 256 2020-03-02 Florida   12     2      0
## 271 2020-03-03 Florida   12     3      0
## 287 2020-03-04 Florida   12     3      0
## 305 2020-03-05 Florida   12     4      0
## 326 2020-03-06 Florida   12     7      2
```

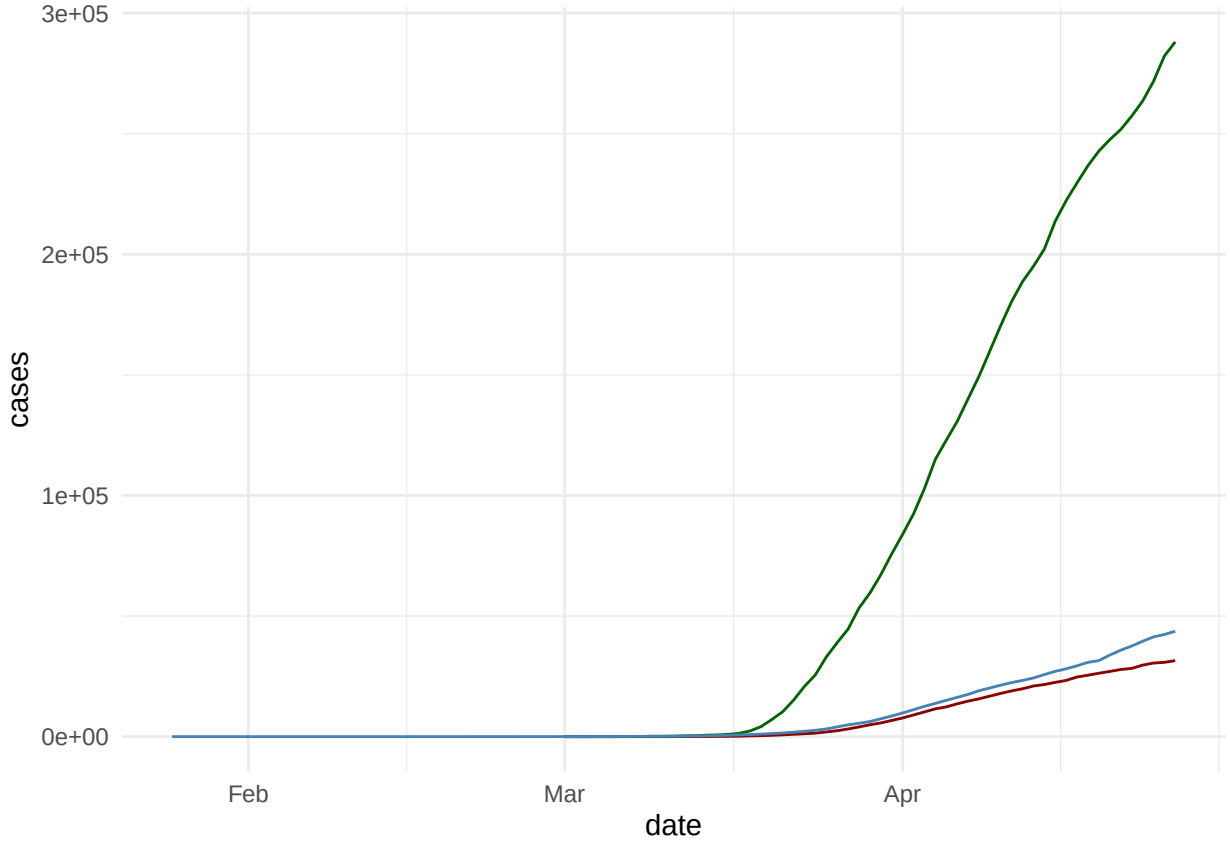
**Plot the number of cases in Florida using `geom_line()`**



**Add lines for New York and California to the plot**

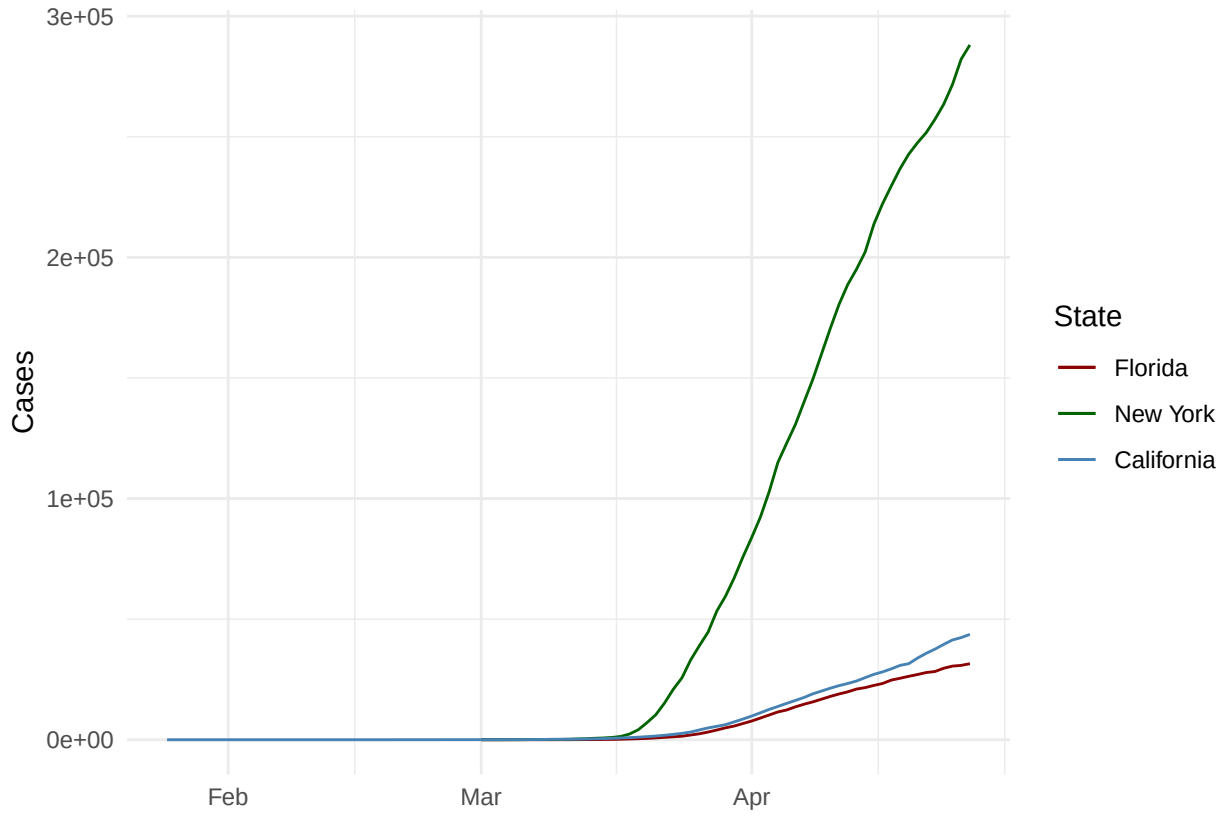


**Use the colors “darkred”, “darkgreen”, and “steelblue” for Florida, New York, and California**

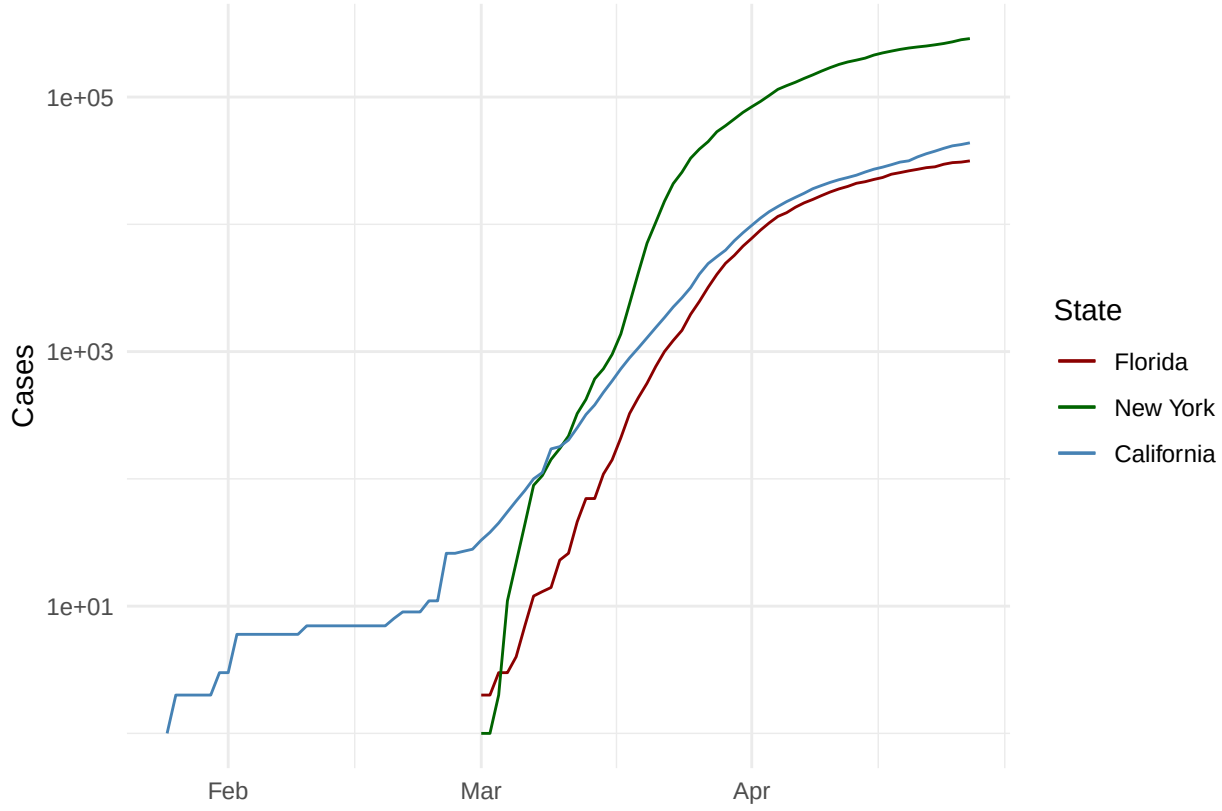


**Add a legend to the plot using `scale_colour_manual`**

**Add a blank (" ") label to the x-axis and the label"Cases" to the y axis**



### Scale the y axis using `scale_y_log10()`



# Assignment 4 - Part 2

# Markdown Basics

## Favorite Foods

1. Chocolate
2. King Crab
3. Maria's Cookies

## Images

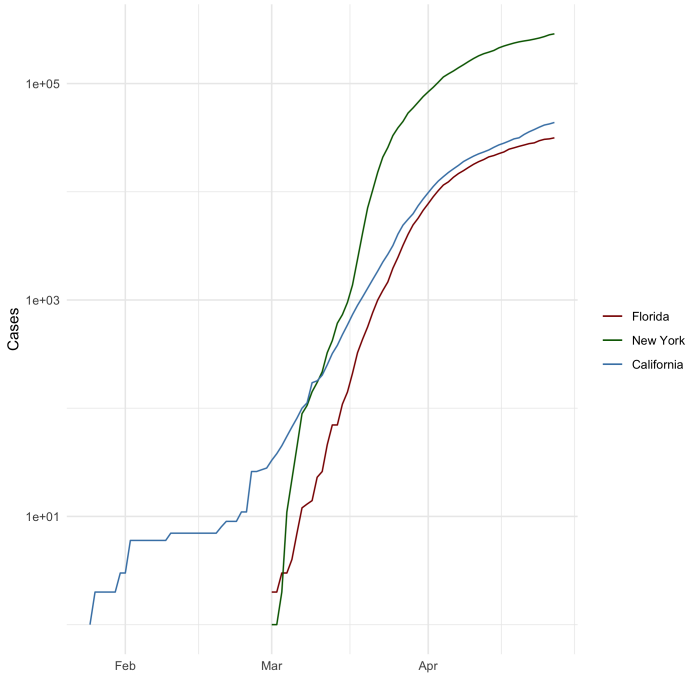



Figure 1: All Cases (Log Plot)

## Add a Quote

> *"Practically perfect in every way"*

## Add an Equation

$$CDF(x) = 1 - (\frac{x}{x_m})^{-}\alpha$$

## Add a Footnote

This is a footnote[1]

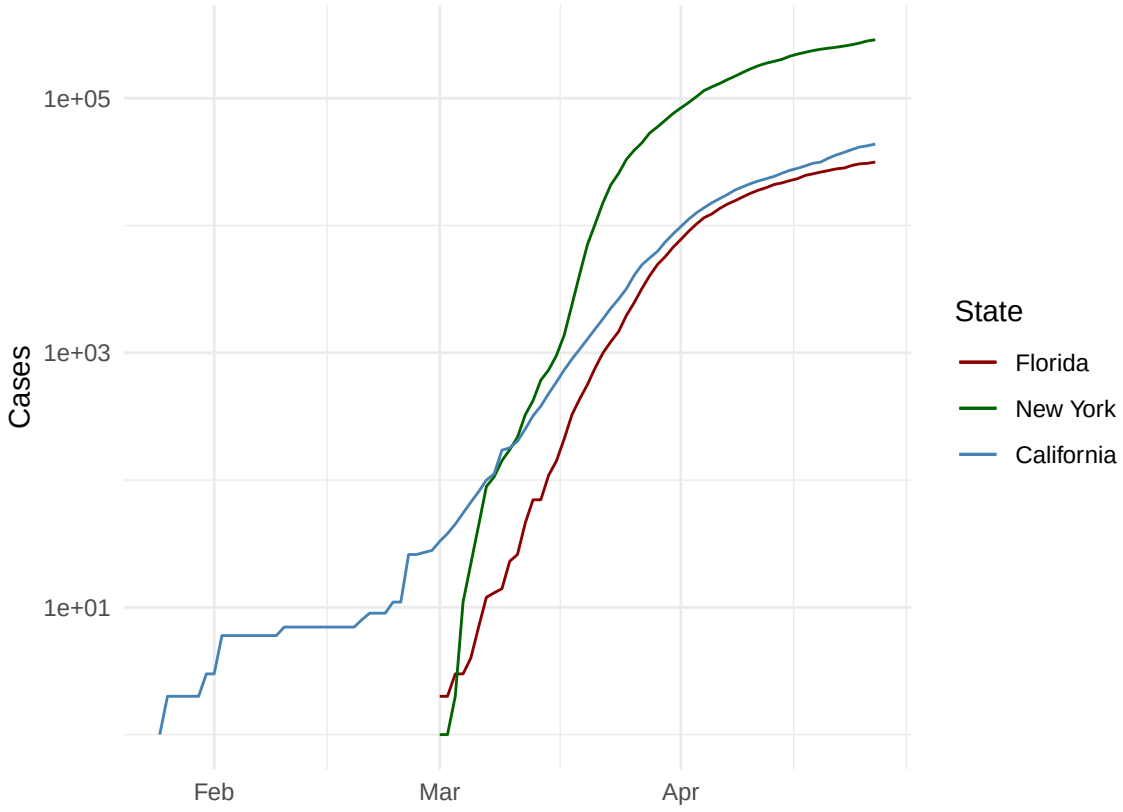
[1]This is the explanation to the footnote

### Add Citations

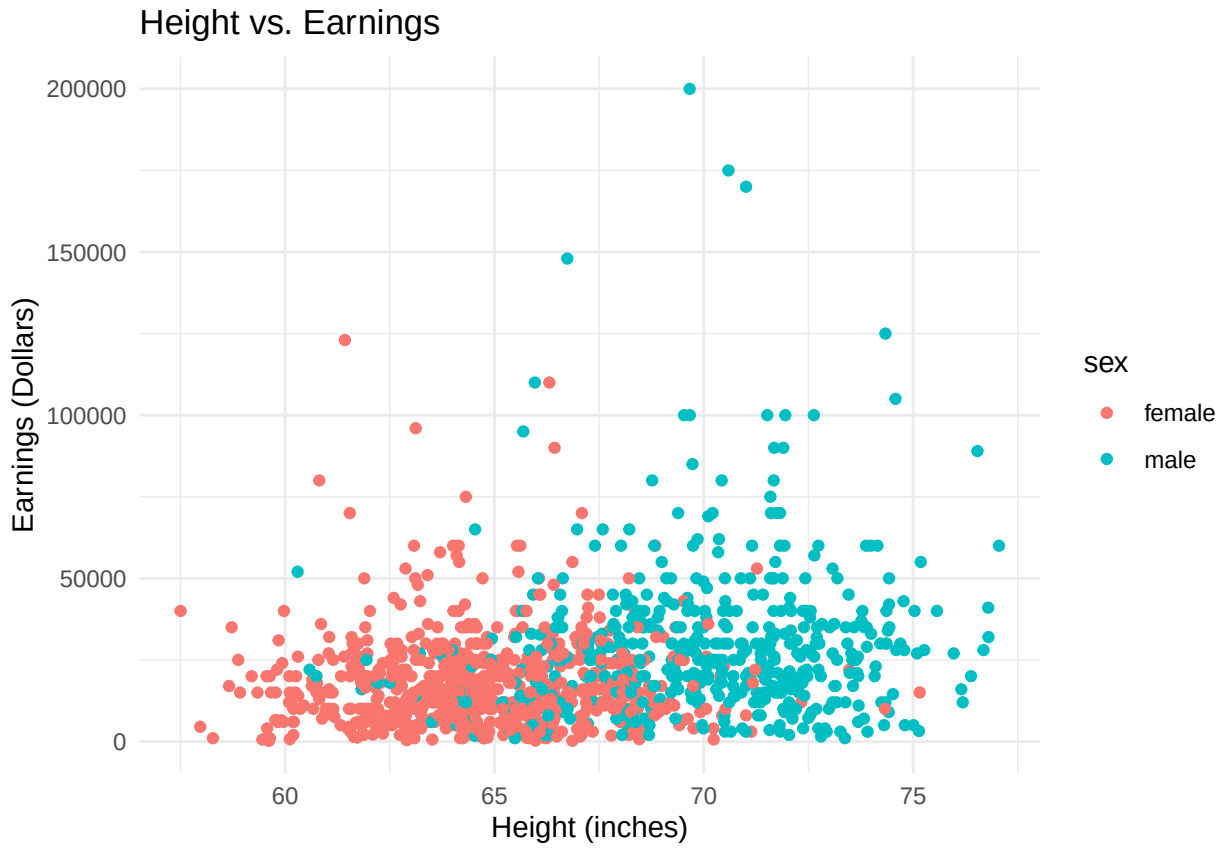
- R for Everyone (Lander 2014)
- Discovering Statistics Using R (Field, Miles, and Field 2012)

## Inline Code

### NY Times COVID-19 Data



### R4DS Height vs Earnings




# Tables

### Knitr Table with Kable

Table 1: One Ring to Rule Them All

| name | race | in_fellowship | ring_bearer | age |
|---|---|---|---|---|
| Aragon | Men | TRUE | FALSE | 88 |
| Bilbo | Hobbit | FALSE | TRUE | 129 |
| Frodo | Hobbit | TRUE | TRUE | 51 |
| Galadriel | Elf | FALSE | FALSE | 7000 |
| Sam | Hobbit | TRUE | TRUE | 36 |
| Gandalf | Maia | TRUE | TRUE | 2019 |
| Legolas | Elf | TRUE | FALSE | 2931 |
| Sauron | Maia | FALSE | TRUE | 7052 |
| Gollum | Hobbit | FALSE | TRUE | 589 |

### Pandoc Table

| name | race | in_fellowship | ring_bearer | age |
|---|---|---|---|---|
| Aragon | Men | TRUE | FALSE | 88 |
| Bilbo | Hobbit | FALSE | TRUE | 129 |
| Frodo | Hobbit | TRUE | TRUE | 51 |
| Galadriel | Elf | FALSE | FALSE | 7000 |
| Sam | Hobbit | TRUE | TRUE | 36 |
| Gandalf | Maia | TRUE | TRUE | 2019 |
| Legolas | Elf | TRUE | FALSE | 2931 |
| Sauron | Maia | FALSE | TRUE | 7052 |
| Gollum | Hobbit | FALSE | TRUE | 589 |

# References


Field, A., J. Miles, and Z. Field. 2012. *Discovering Statistics Using r*. SAGE Publications. https://books.google.com/books?id=wd2K2zC3swIC.

Lander, J. P. 2014. *R for Everyone: Advanced Analytics and Graphics*. Addison-Wesley Data and Analytics Series. Addison-Wesley. https://books.google.com/books?id=3eBVAgAAQBAJ.