

assignment_06_01_PothineniKalyan

PothineniKalyan

2023-04-22

R Markdown

```
# Assignment: ASSIGNMENT 6.2  
# Name: Pothineni, Kalyan  
# Date: 2023-04-20
```

```
## Load the ggplot2 package  
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

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

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

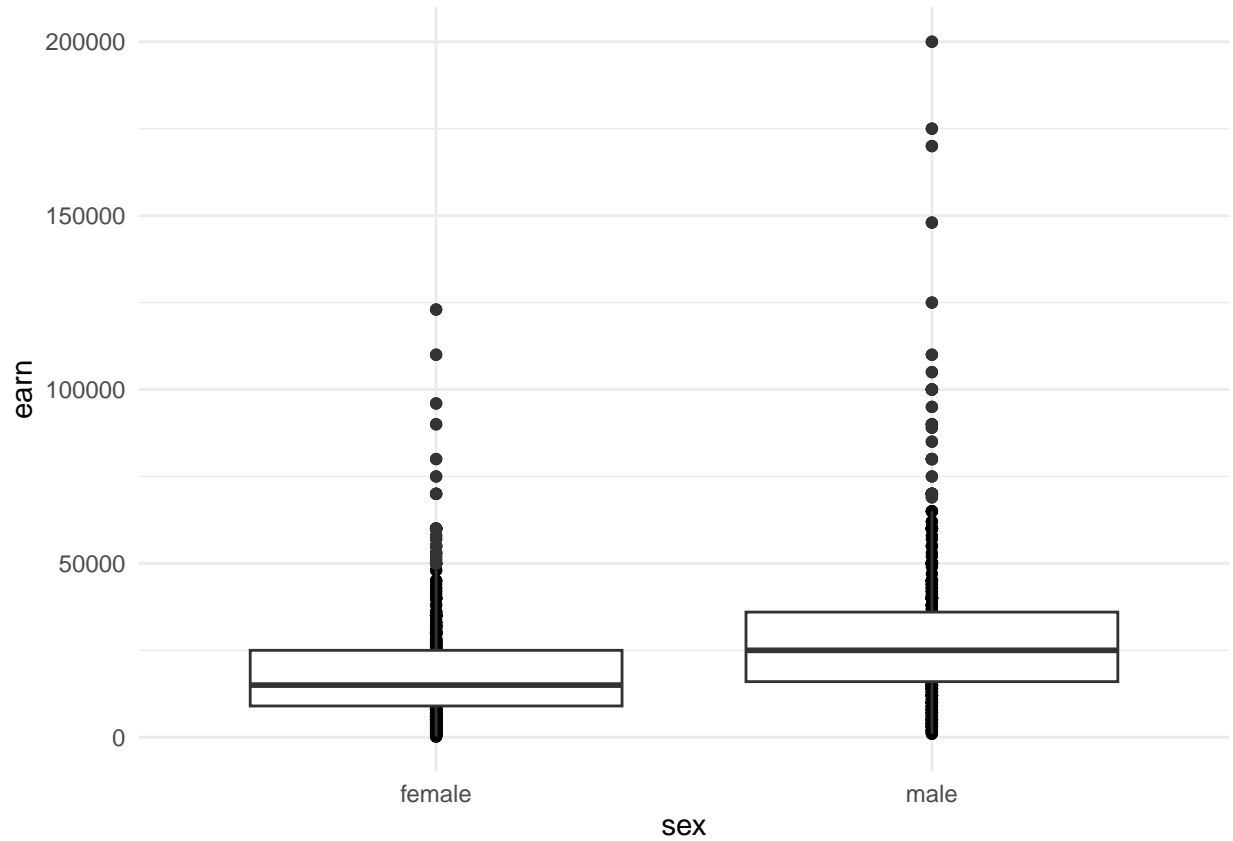
```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
theme_set(theme_minimal())
```

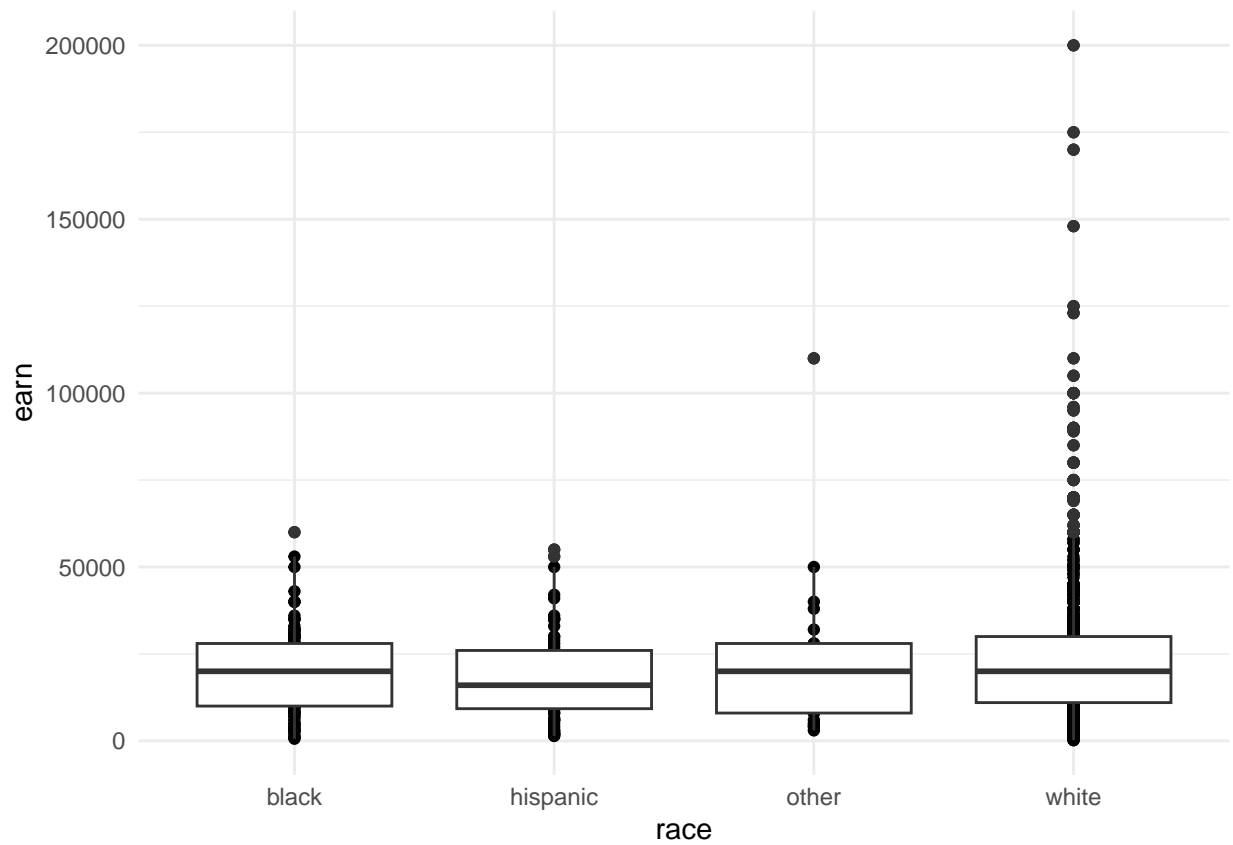
```
## Load the `data/r4ds/heights.csv` to  
heights_df <- read.csv("data/r4ds/heights.csv")  
head(heights_df, n=5)
```

```
##   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
```

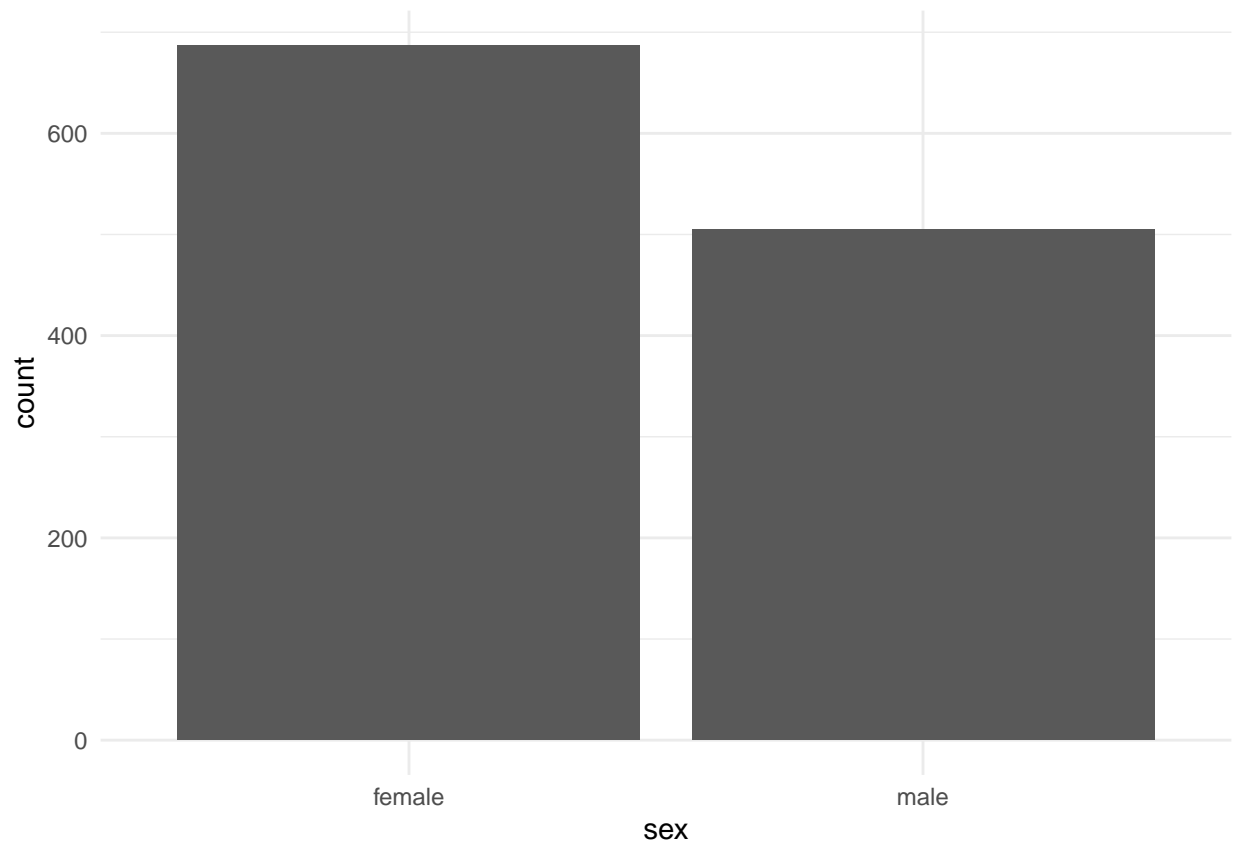
```
# https://ggplot2.tidyverse.org/reference/geom\_boxplot.html
## Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`
## sex vs. earn
ggplot(heights_df, aes(x=sex, y=earn)) + geom_point() + geom_boxplot()
```



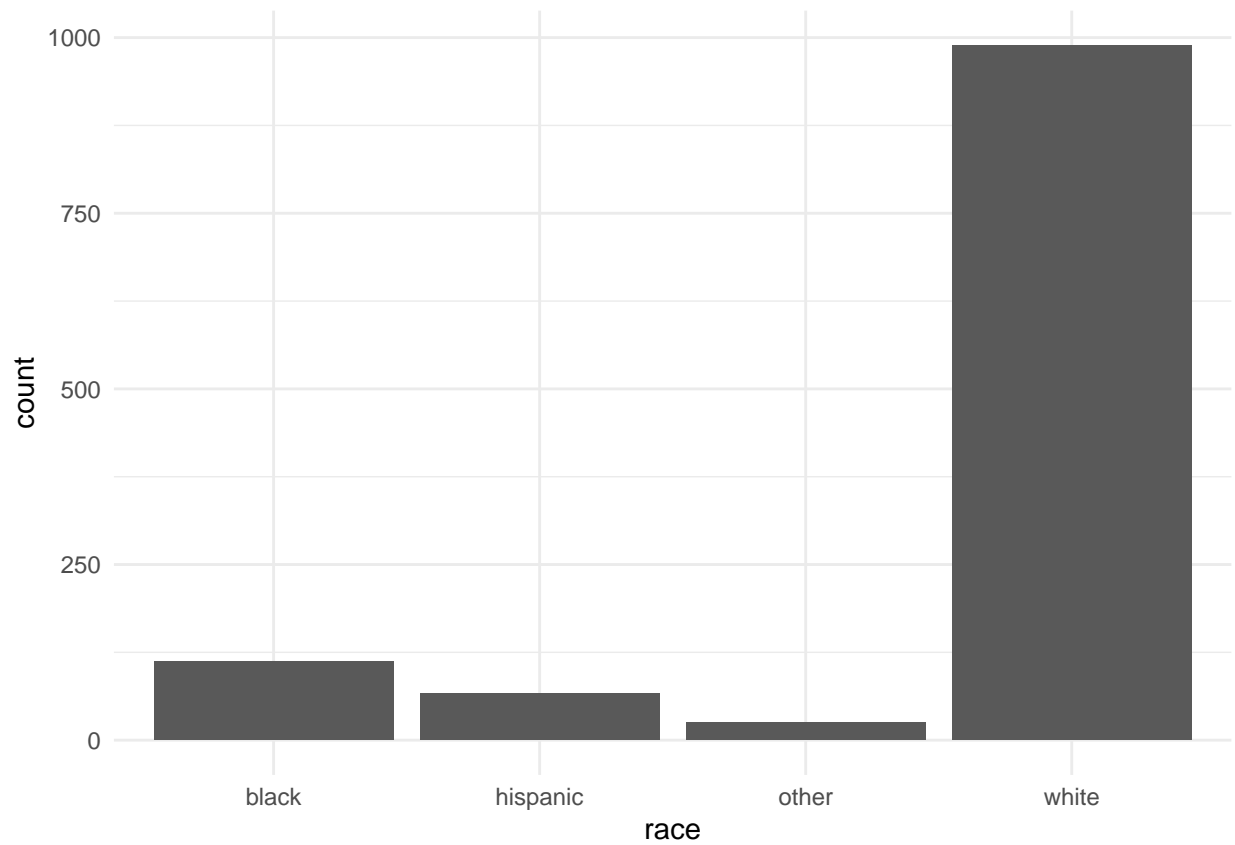
```
## race vs. earn
ggplot(heights_df, aes(x=race, y=earn)) + geom_point() + geom_boxplot()
```



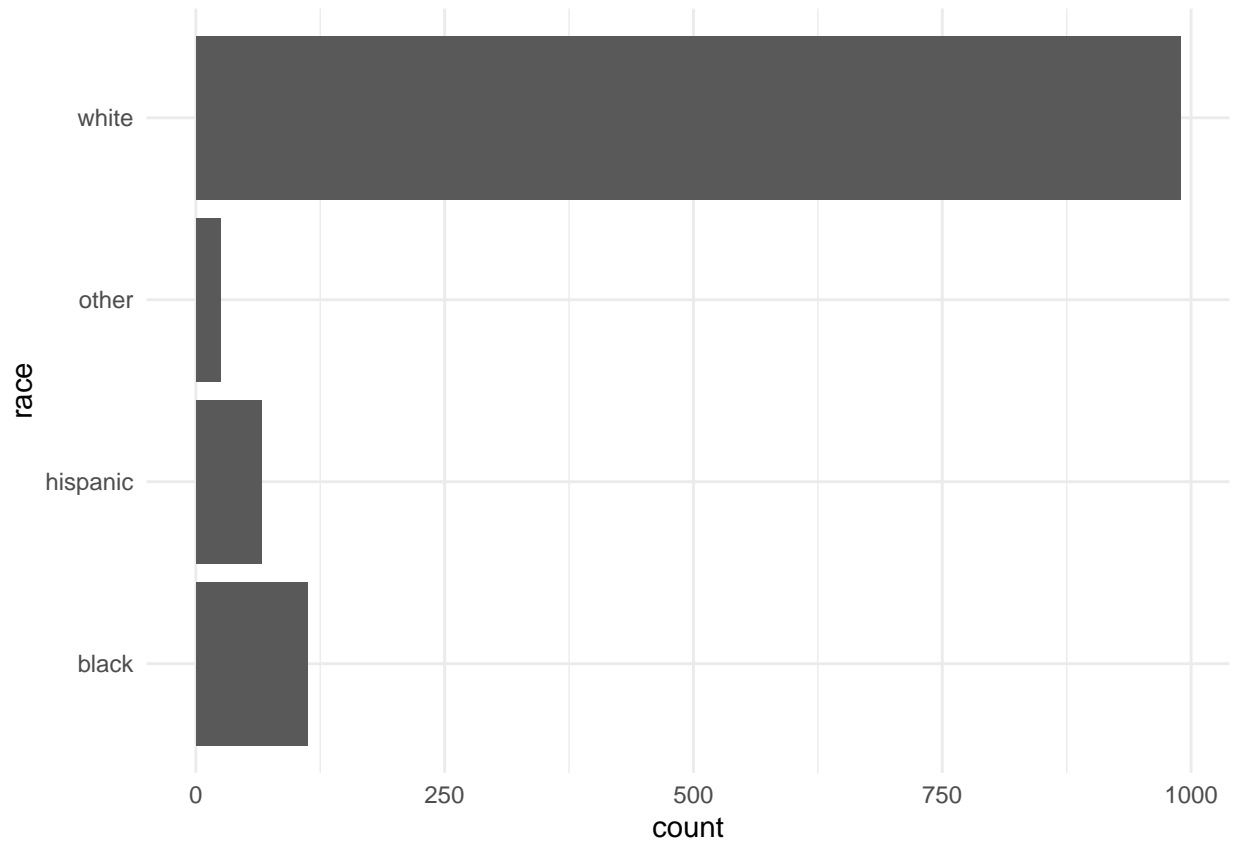
```
## https://ggplot2.tidyverse.org/reference/geom\_bar.html  
## Using `geom_bar()` plot a bar chart of the number of records for each `sex`  
ggplot(heights_df, aes(sex)) + geom_bar()
```



```
## Using `geom_bar()` plot a bar chart of the number of records for each race  
ggplot(heights_df, aes(race)) + geom_bar()
```



```
## Create a horizontal bar chart by adding `coord_flip()` to the previous plot  
## Flipping horizontal to vertical and vise versa  
ggplot(heights_df, aes(race)) + geom_bar() + coord_flip()
```



```
# 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
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv")

head(covid_df, n=5)
```

```
##      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
```

```
## Parse the date column using `as.Date()`
covid_df$date <- as.Date(covid_df$date)
head(covid_df$date, n=5)
```

```
## [1] "2020-01-21" "2020-01-22" "2020-01-23" "2020-01-24" "2020-01-24"
```

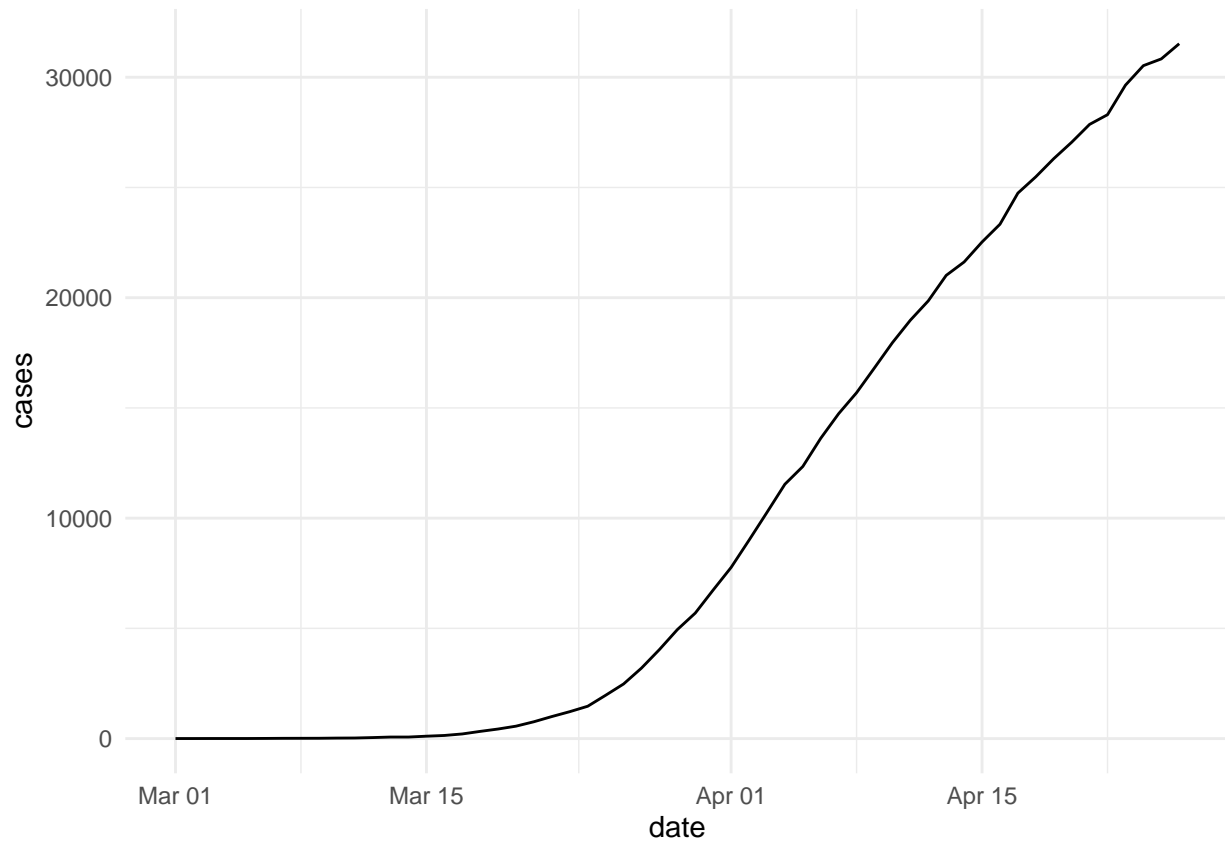
```
## Create three dataframes named `california_df`, `ny_df`, and `florida_df`
## containing the data from California, New York, and Florida
distinct(covid_df, state)
```

##	state
## 1	Washington
## 2	Illinois
## 3	California
## 4	Arizona
## 5	Massachusetts
## 6	Wisconsin
## 7	Texas
## 8	Nebraska
## 9	Utah
## 10	Oregon
## 11	Florida
## 12	New York
## 13	Rhode Island
## 14	Georgia
## 15	New Hampshire
## 16	North Carolina
## 17	New Jersey
## 18	Colorado
## 19	Maryland
## 20	Nevada
## 21	Tennessee
## 22	Hawaii
## 23	Indiana
## 24	Kentucky
## 25	Minnesota
## 26	Oklahoma
## 27	Pennsylvania
## 28	South Carolina
## 29	District of Columbia
## 30	Kansas
## 31	Missouri
## 32	Vermont
## 33	Virginia
## 34	Connecticut
## 35	Iowa
## 36	Louisiana
## 37	Ohio
## 38	Michigan
## 39	South Dakota
## 40	Arkansas
## 41	Delaware
## 42	Mississippi
## 43	New Mexico
## 44	North Dakota
## 45	Wyoming
## 46	Alaska
## 47	Maine
## 48	Alabama
## 49	Idaho
## 50	Montana
## 51	Puerto Rico
## 52	Virgin Islands
## 53	Guam

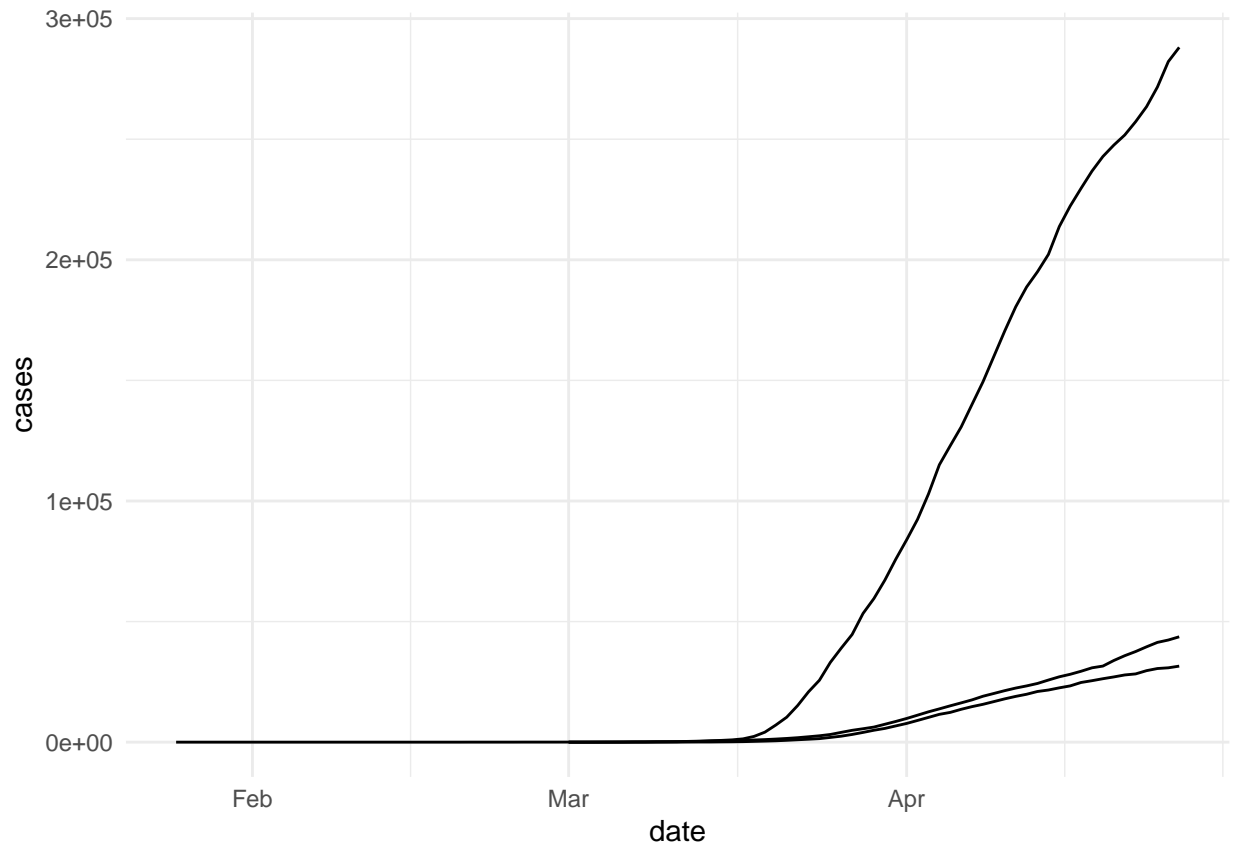
```
## 54          West Virginia
## 55 Northern Mariana Islands
```

```
california_df <- covid_df[ which( covid_df$state == "California"), ]
ny_df <- covid_df[ which( covid_df$state == "New York"), ]
florida_df <- covid_df[ which( covid_df$state == "Florida"), ]
```

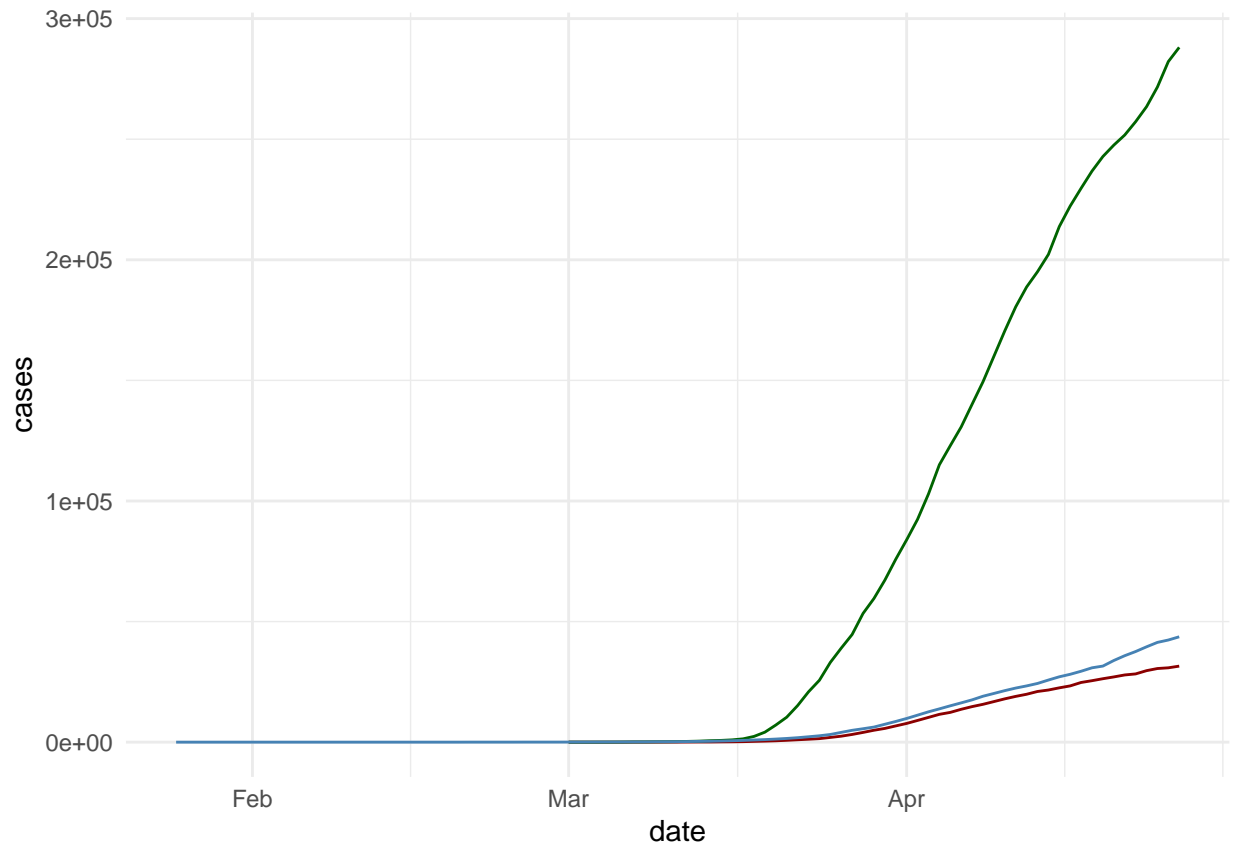
```
## Plot the number of cases in Florida using `geom_line()`
##group=1 to plot all the data points as a single group
ggplot(data=florida_df, aes(x=date, y=cases, group=1)) + geom_line()
```



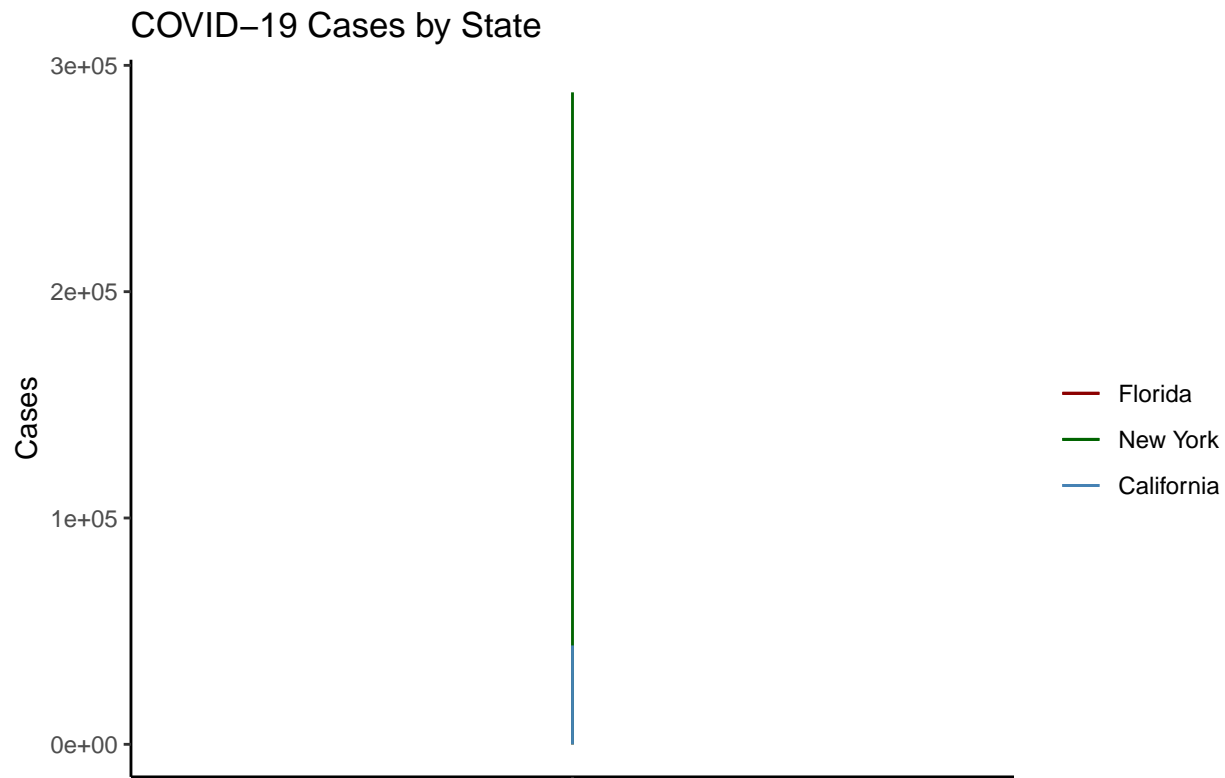
```
## Add lines for New York and California to the plot
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases)) +
  geom_line(data=ny_df, aes(y = cases)) +
  geom_line(data=california_df, aes(y = cases))
```

```
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = "darkred") +
  geom_line(data=ny_df, aes(y = cases), color="darkgreen") +
  geom_line(data=california_df, aes(y = cases), color= "steelblue")
```



```
## 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
ggplot(data=florida_df, aes(x=" ", group=1)) +
  geom_line(aes(y = cases, colour = "Florida")) +
  geom_line(data=ny_df, aes(y = cases, colour="New York")) +
  geom_line(data=california_df, aes(y = cases, colour="California")) +
  scale_colour_manual("",
                      breaks = c("Florida", "New York", "California"),
                      values = c("darkred", "darkgreen", "steelblue")) +
  xlab(" ") + ylab("Cases") +
  ggtitle("COVID-19 Cases by State") +
  theme_classic()
```



```
## Scale the y axis using `scale_y_log10()`
ggplot(data=florida_df, aes(x=" ", group=1)) +
  geom_line(aes(y = cases, colour = "Florida")) +
  geom_line(data=ny_df, aes(y = cases, colour="New York")) +
  geom_line(data=california_df, aes(y = cases, colour="California")) +
  scale_colour_manual("",
                      breaks = c("Florida", "New York", "California"),
                      values = c("darkred", "darkgreen", "steelblue")) +
  xlab(" ") + ylab("Cases") + scale_y_log10() +
  ggtitle("COVID-19 Cases by State") +
  theme_classic()
```

