ASSIGNMENT 4

Simek, Kathryn

2020-09-20

Load the ggplot2 package

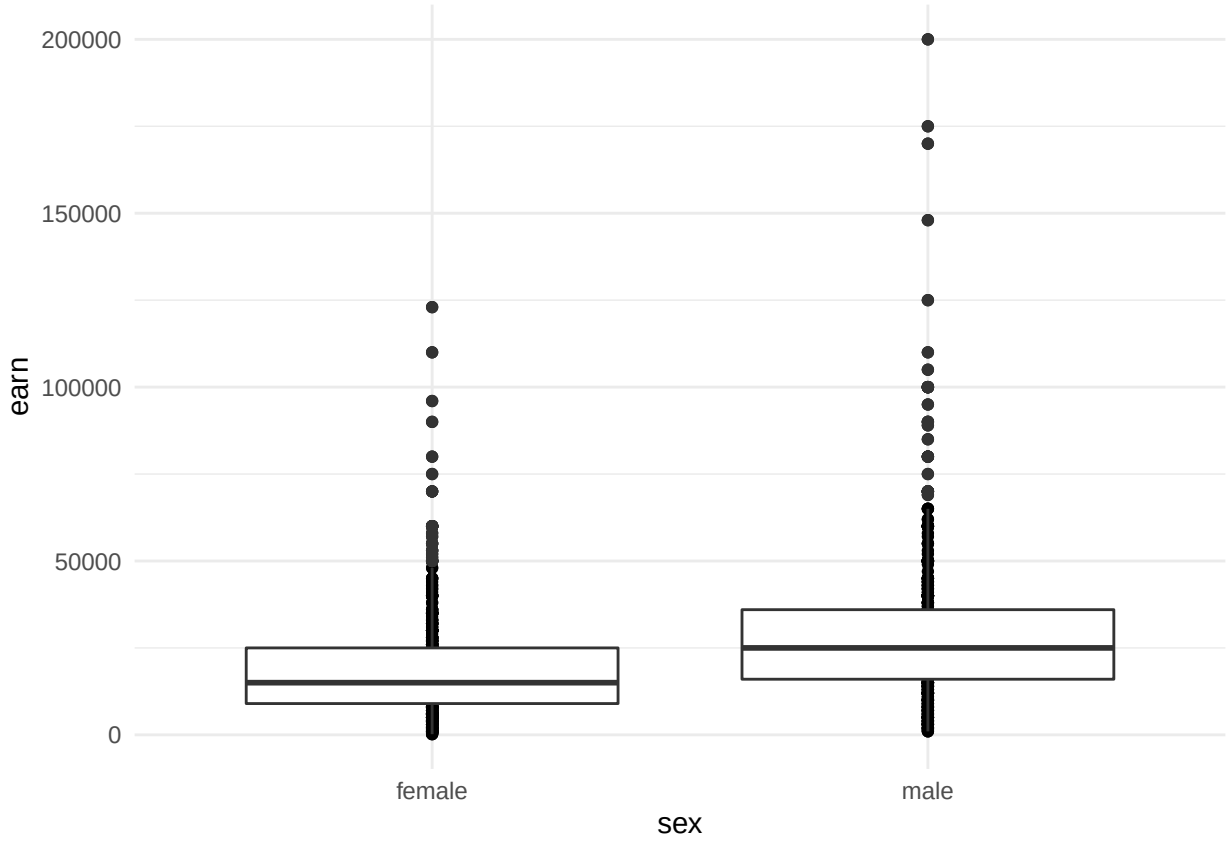
```
library(ggplot2)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
setwd("C:/Users/katie/OneDrive/Documents/GitHub/dsc520")

## Load the `data/r4ds/heights.csv` to
heights_df <- read.csv("data/r4ds/heights.csv")
```

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

and `geom_boxplot()`

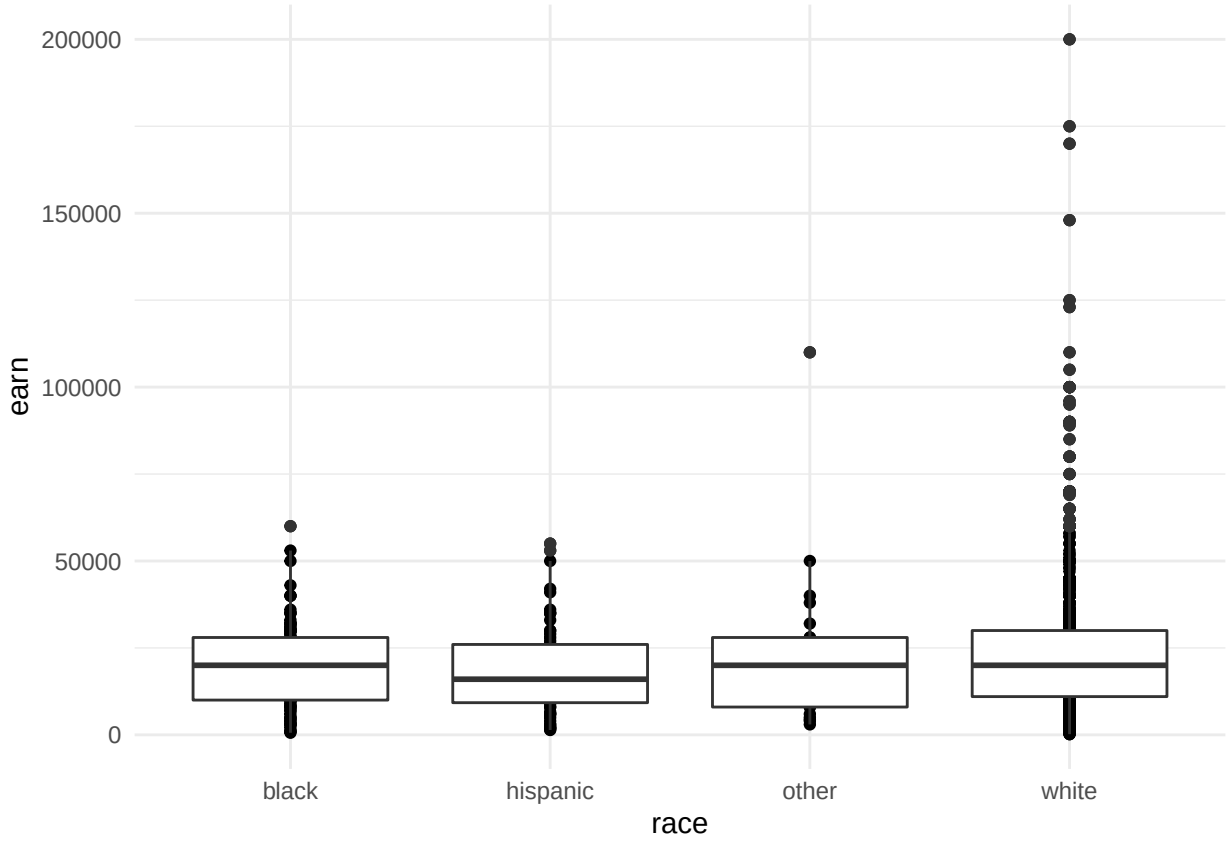
sex vs. earn

```
ggplot(heights_df, aes(x=sex, y=earn)) + geom_point()+ geom_boxplot()
```



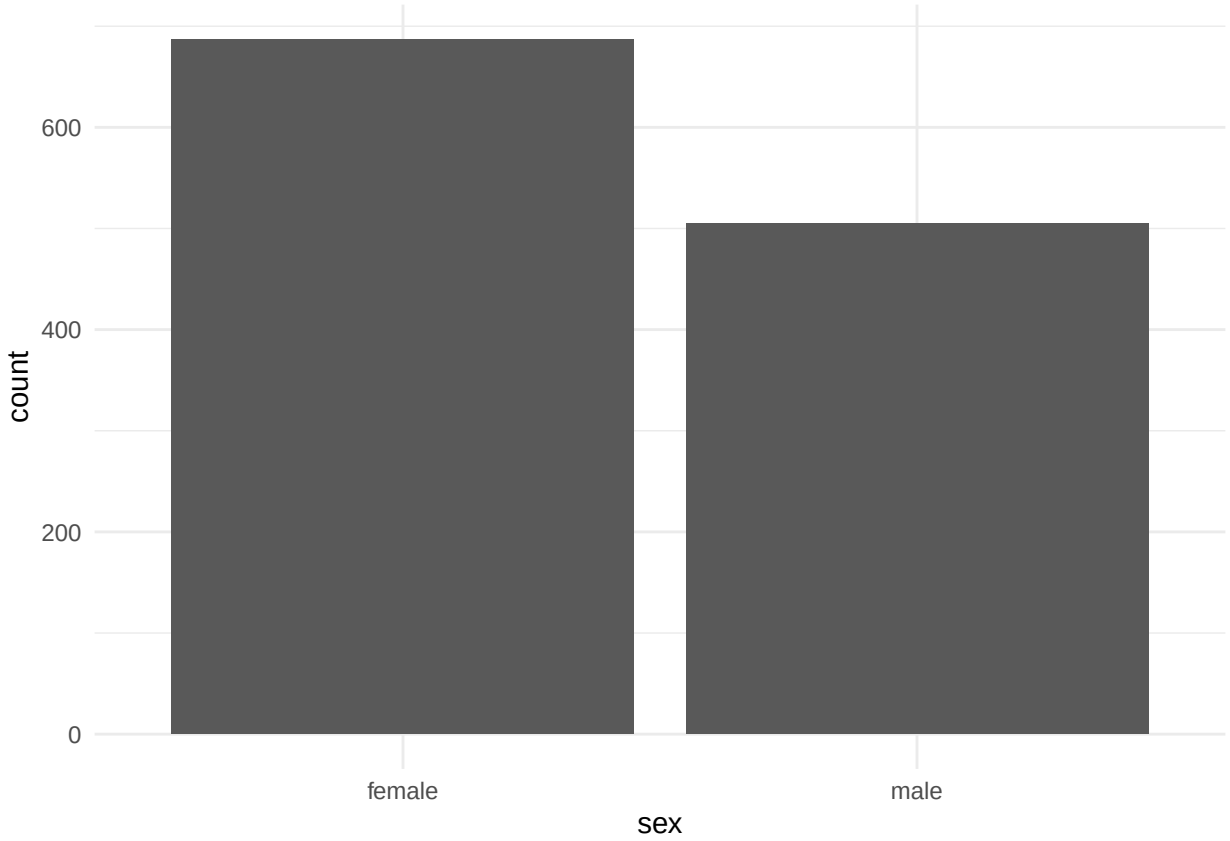
race vs. earn

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



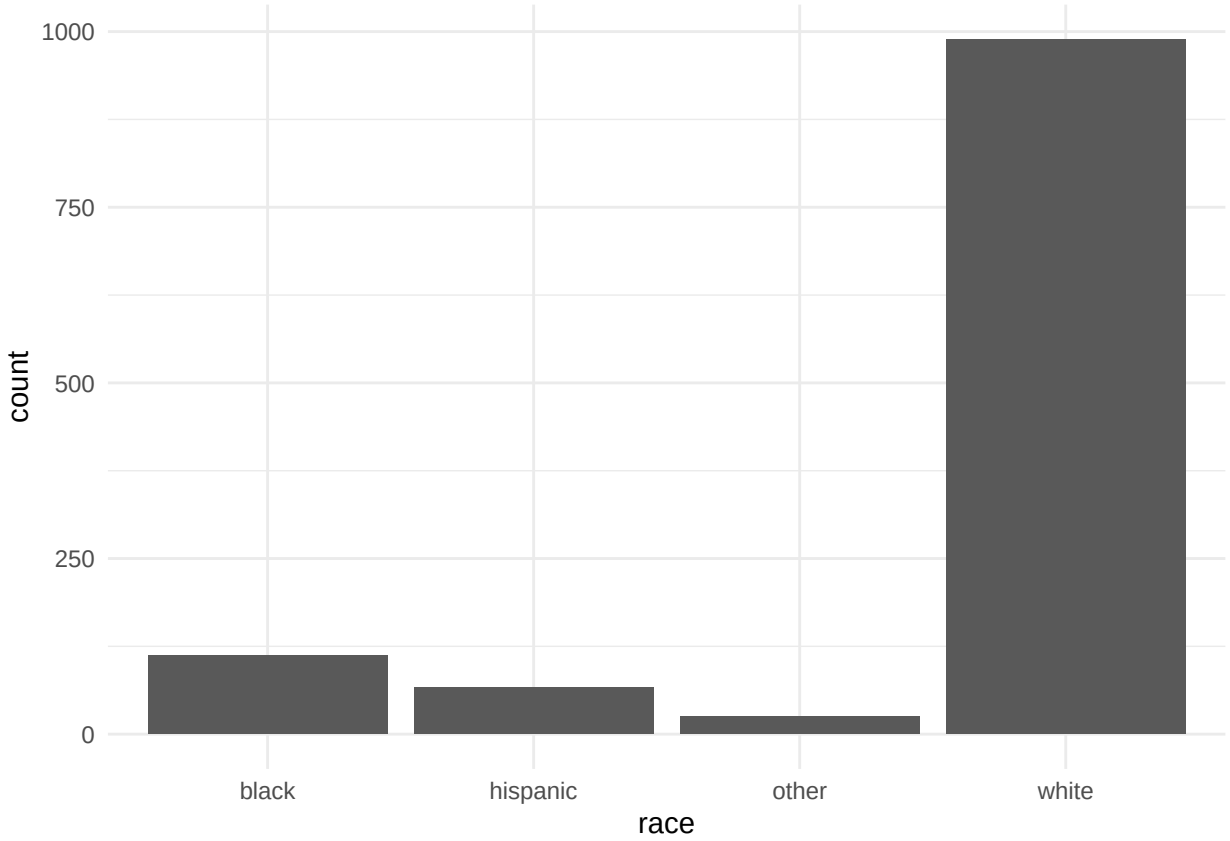
Using `geom_bar()` plot a bar chart of the number of records for each `sex`

```
ggplot(heights_df, aes(sex)) + geom_bar()
```



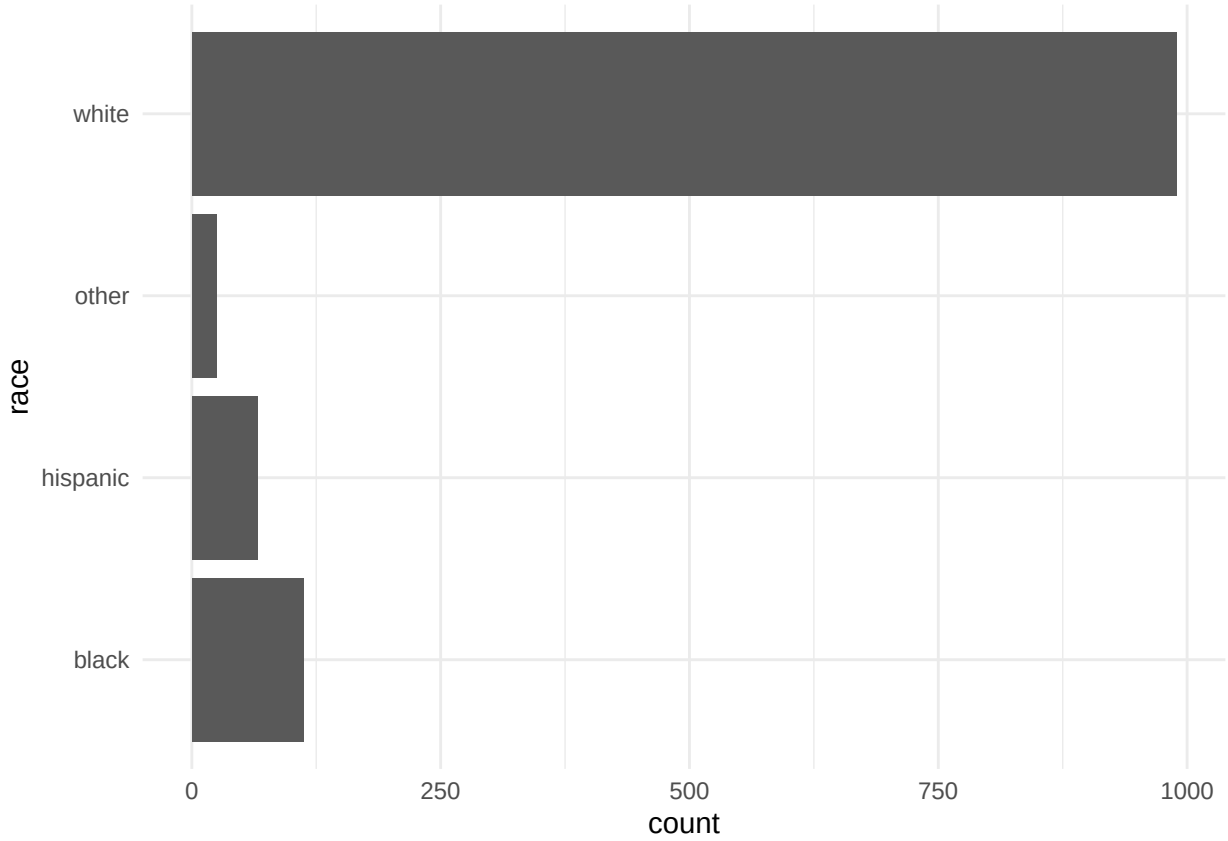
Using `geom_bar()` plot a bar chart of the number of records for each race

```
ggplot(heights_df, aes(race)) + geom_bar()
```



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

```
ggplot(heights_df, aes(race)) + geom_bar() + coord_flip()
```

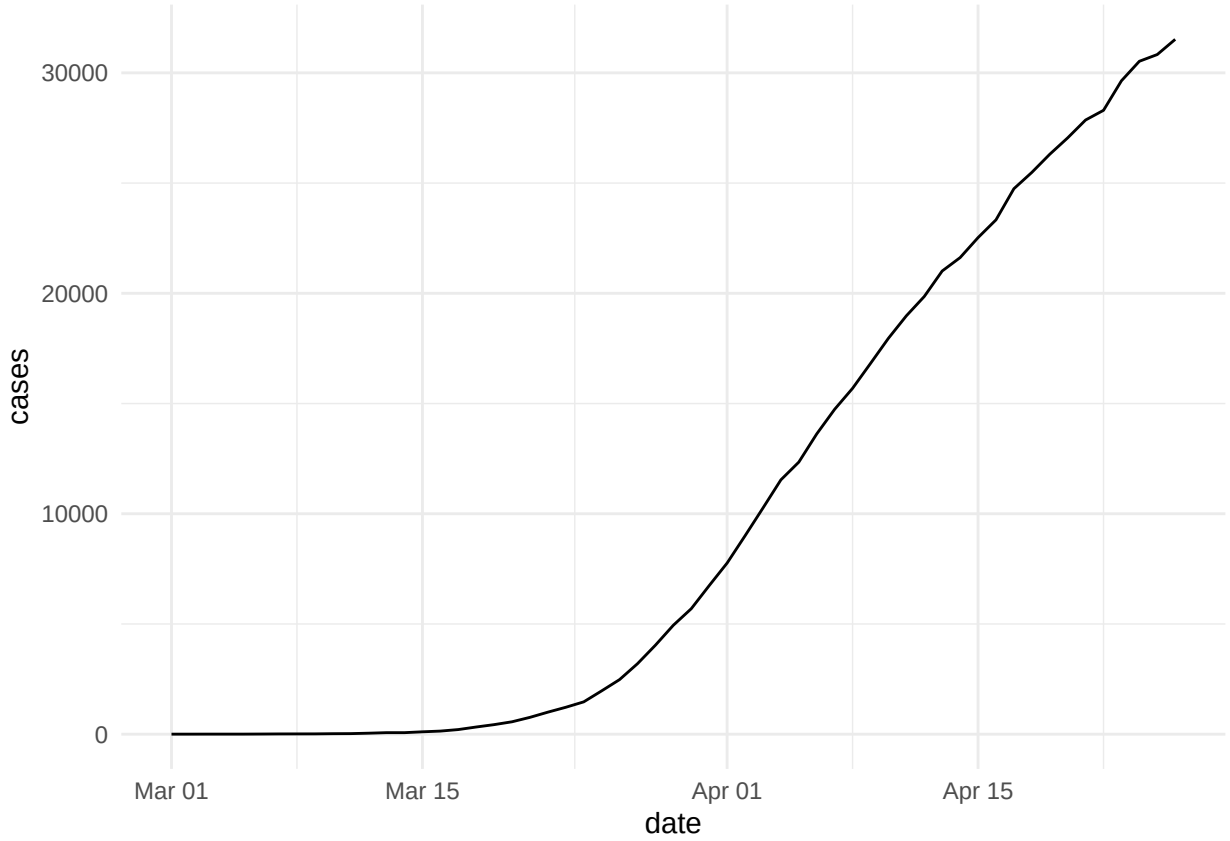


Load the file "data/nytimes/covid-19-data/us-states.csv" and assign it to the covid_df dataframe

```
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv",
                     col.names = c("date", "state", "fips", "cases", "deaths"))

## Parse the date column using `as.Date()``
covid_df$date <- as.Date(covid_df$date)

## Create three dataframes named `california_df`, `ny_df`, and `florida_df`
## containing the data from California, New York, and Florida
california_df <- covid_df[covid_df$state == "California",]
ny_df <- covid_df[covid_df$state == "New York",]
florida_df <- covid_df[covid_df$state== "Florida",]
```

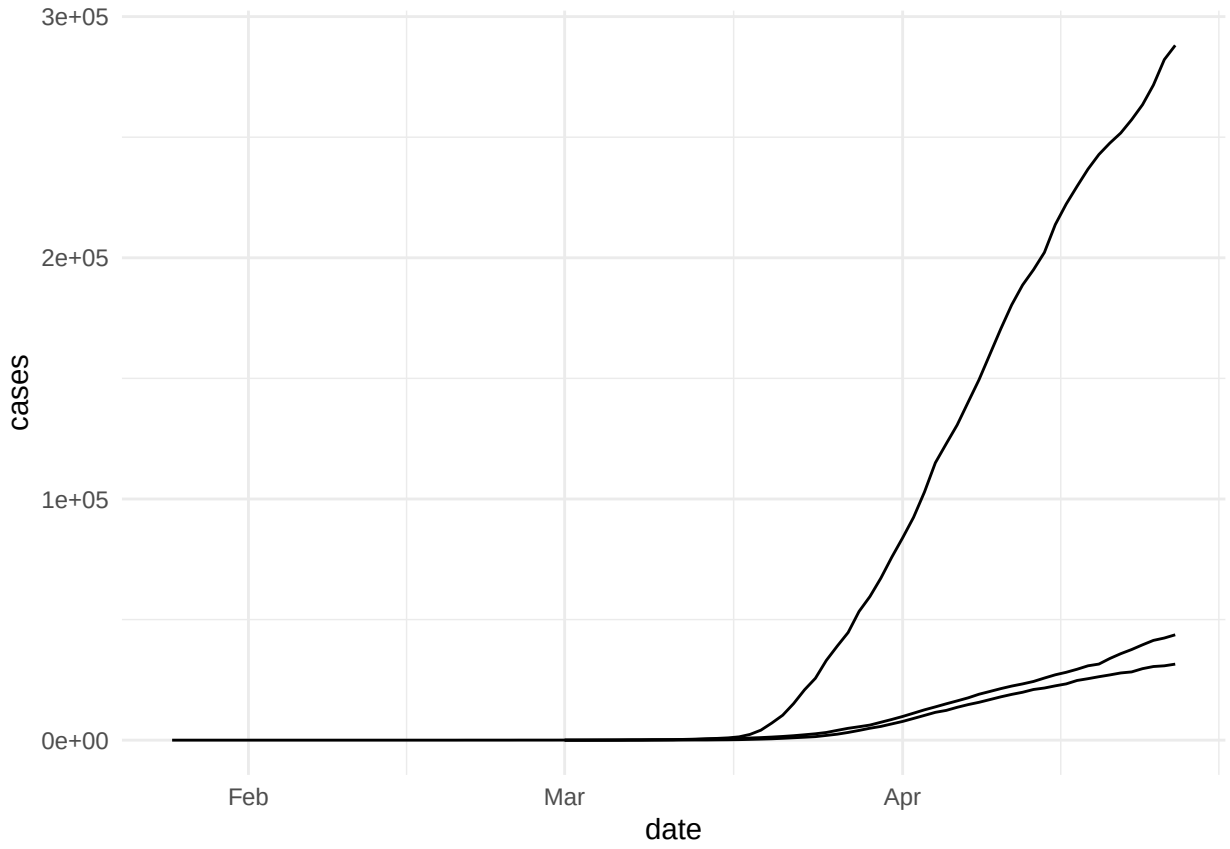
Plot the number of cases in Florida using geom_line()

```
ggplot(data=florida_df, aes(x=date, y=cases, group=1)) + geom_line()
```



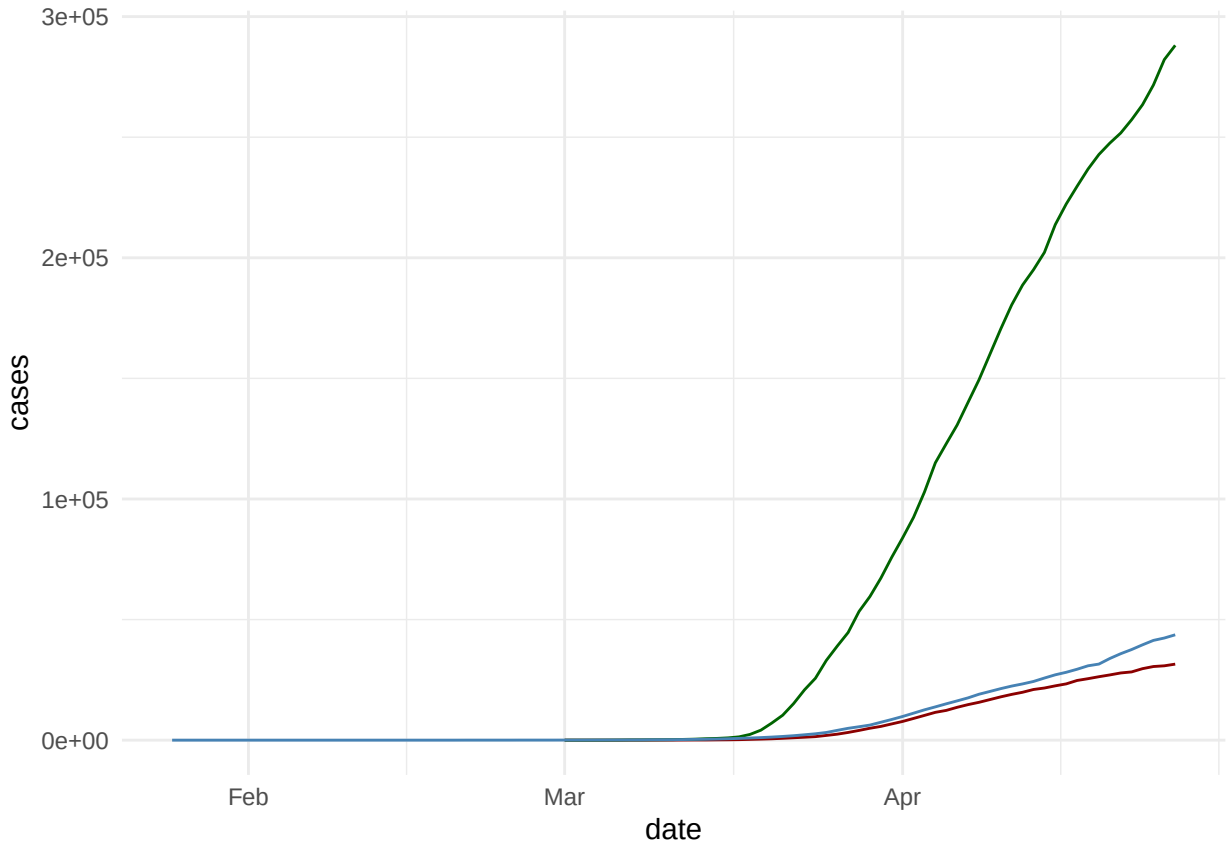
Add lines for New York and California to the plot

```
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases)) +
  geom_line(data=california_df, aes(y = cases)) +
  geom_line(data=ny_df, aes(y = cases))
```



Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California

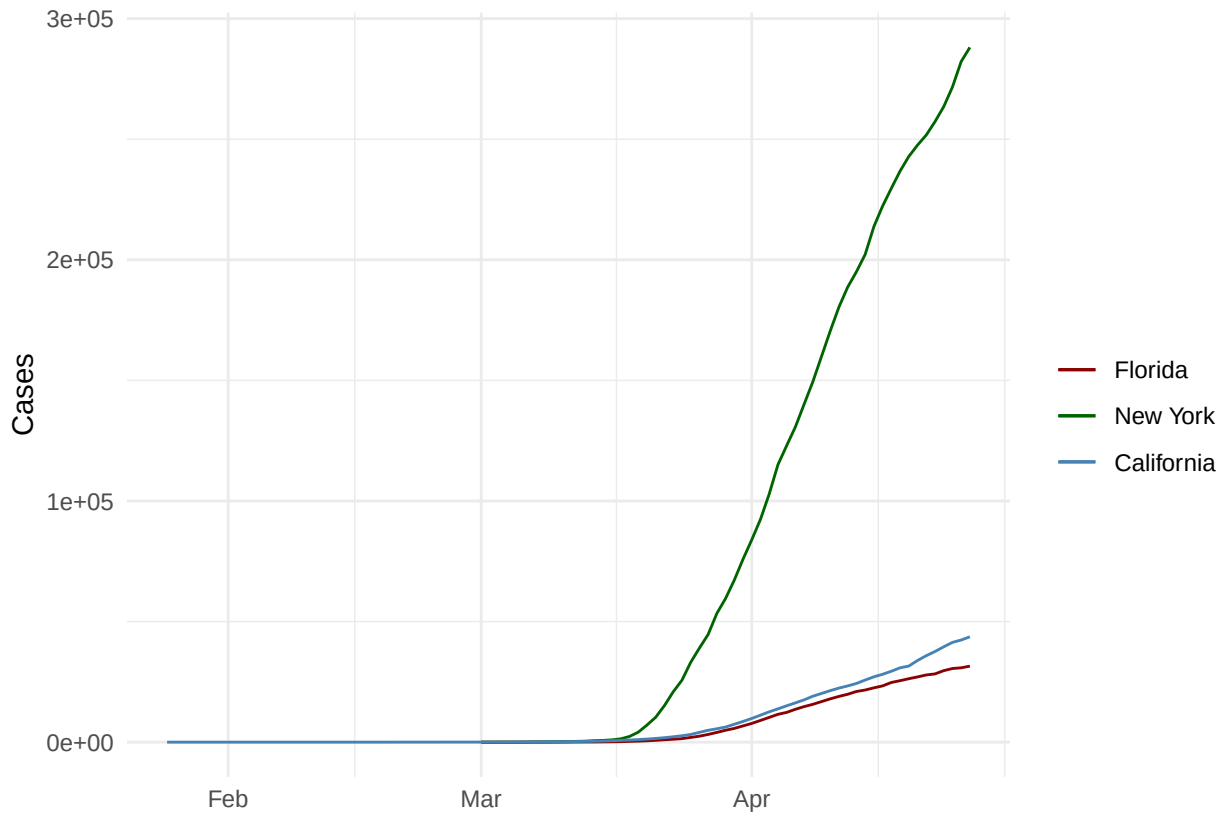
```
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = "darkred") +
  geom_line(data=ny_df, aes(y = cases), color="darkgreen") +
  geom_line(data=california_df, aes(y = cases), color = "steelblue")
```



Add a legend to the plot using `scale_colour_manual`

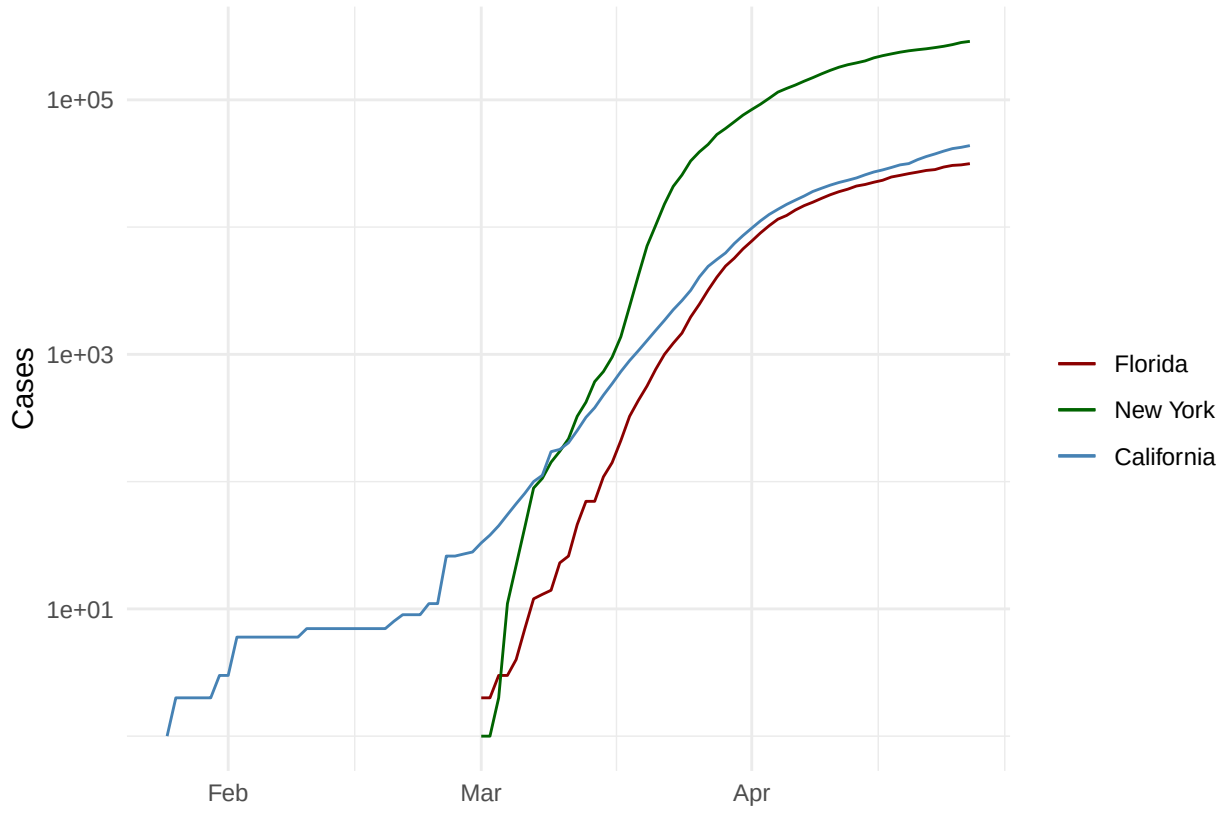
Add a blank (" “) label to the x-axis and the label”Cases" to the y axis

```
ggplot(data=florida_df, aes(x=date, 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 the y axis using `scale_y_log10()`

```
ggplot(data=florida_df, aes(x=date, 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()
```


Cases
1e+05
1e+03
1e+01
Feb
Mar
Apr
Florida
New York
California