

Assignment: ASSIGNMENT 4
Name: Koby-Hercsky, Theodore
Date: 2021-04-20

1. Load the ggplot2 package
 - a. `library(ggplot2)`
 - b. `theme_set(theme_minimal())`
2. Set the working directory to the root of your DSC 520 directory
 - a. `setwd("Documents/Bellevue University Classes/DSC520")`
3. Load the `data/r4ds/heights.csv` to
 - a. `heights_df <- read_csv("data/r4ds/heights.csv")`

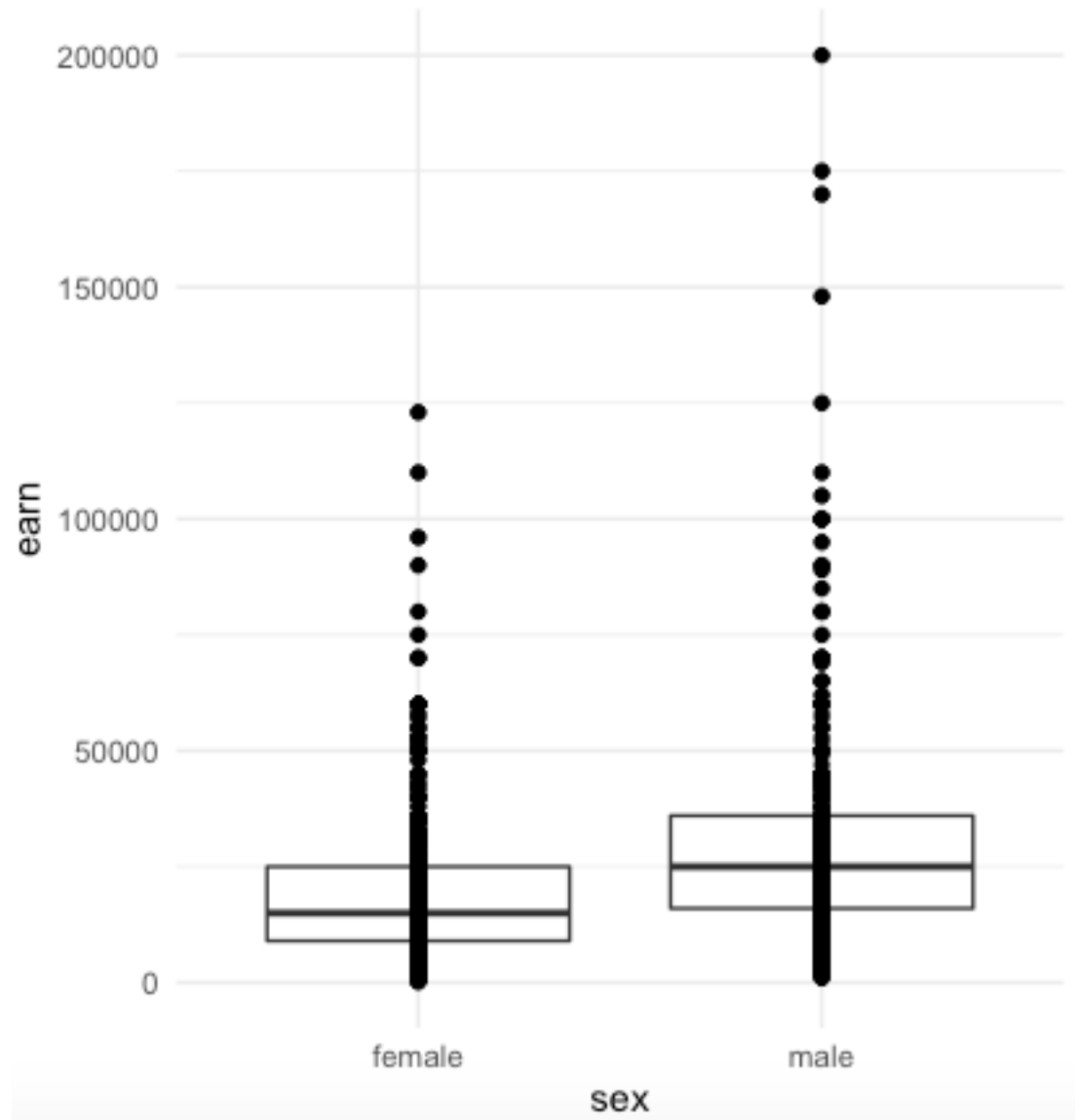
— Column specification —

```
cols(  
  earn = col_double(),  
  height = col_double(),  
  sex = col_character(),  
  ed = col_double(),  
  age = col_double(),  
  race = col_character()  
)
```

	↑	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
7		29000	61.65633	female	12	49	white
8		32000	72.69854	male	17	46	white
9		2000	72.03947	male	15	21	hispanic
10		27000	72.23493	male	12	26	white
11		6530	69.51215	male	16	65	white
12		30000	68.03161	male	11	34	white
13		12000	67.55693	male	12	27	white
14		12000	65.43059	female	12	51	white
15		22000	65.66285	female	16	35	white
Showing 1 to 16 of 1,192 entries, 6 total columns							

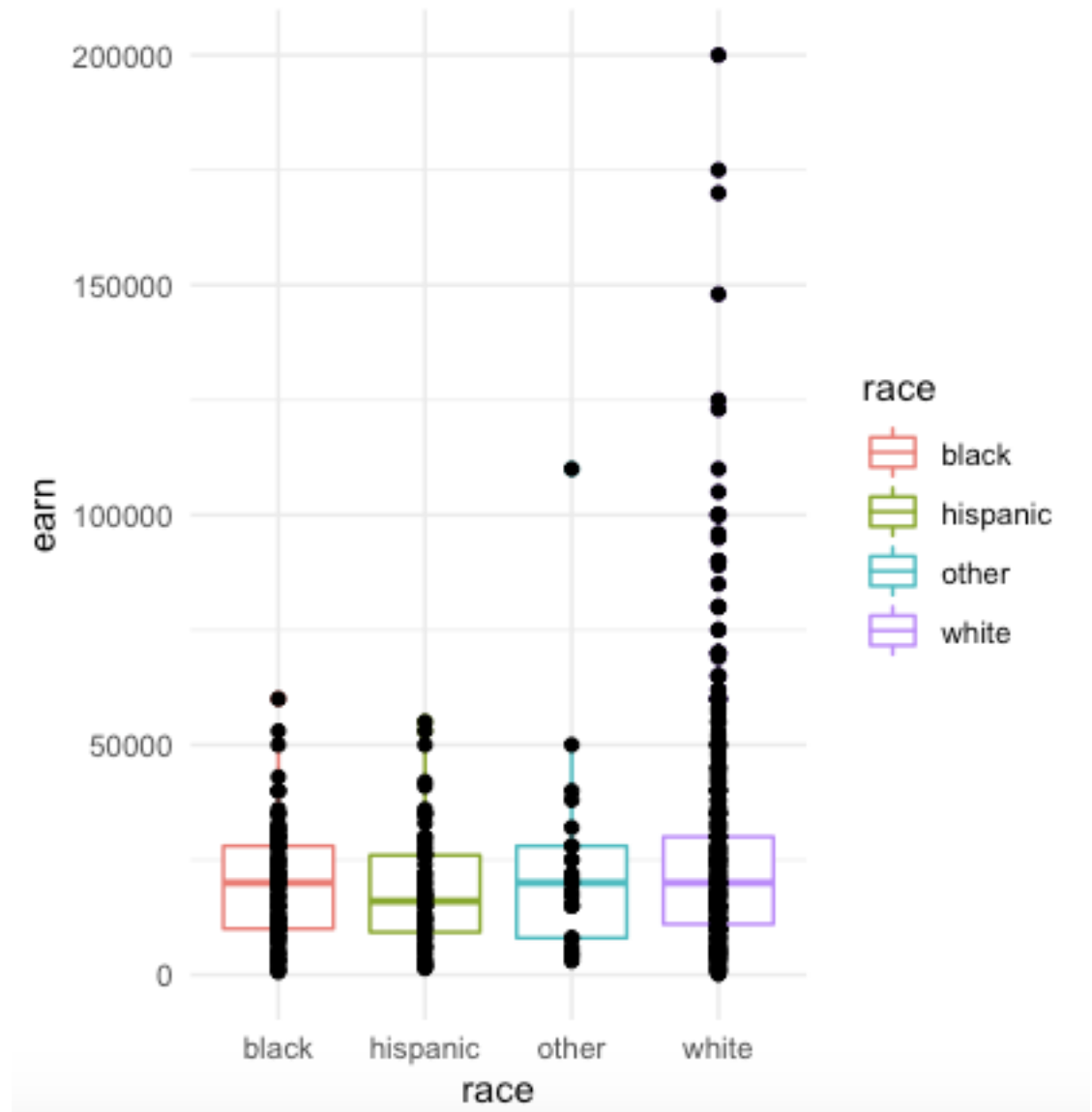
- b.
4. Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`
 - a. sex vs. earn

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



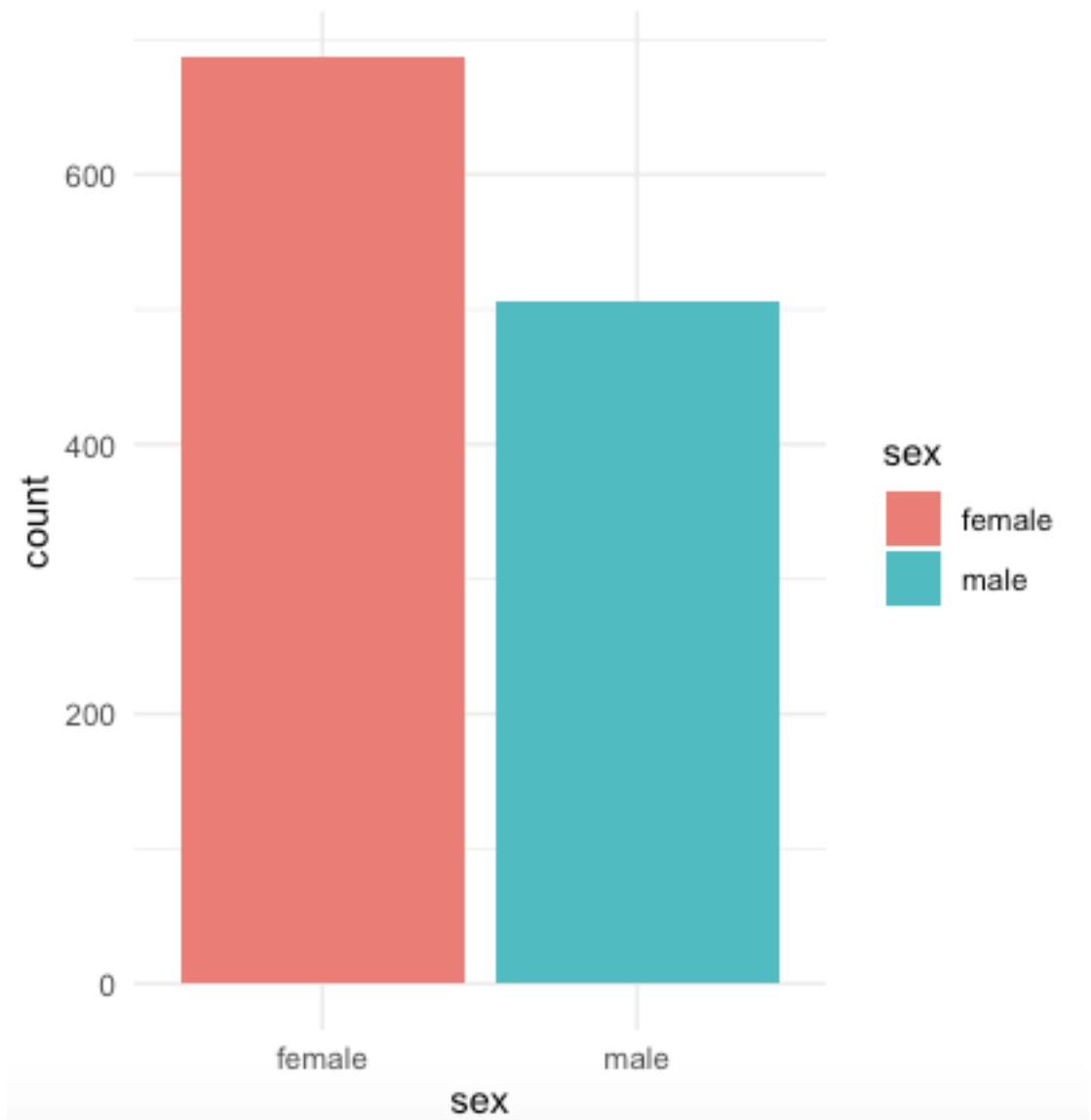
b. race vs. earn

- i. `ggplot(heights_df, aes(x=race, y=earn))+geom_boxplot(aes(colour = race))+geom_point()`



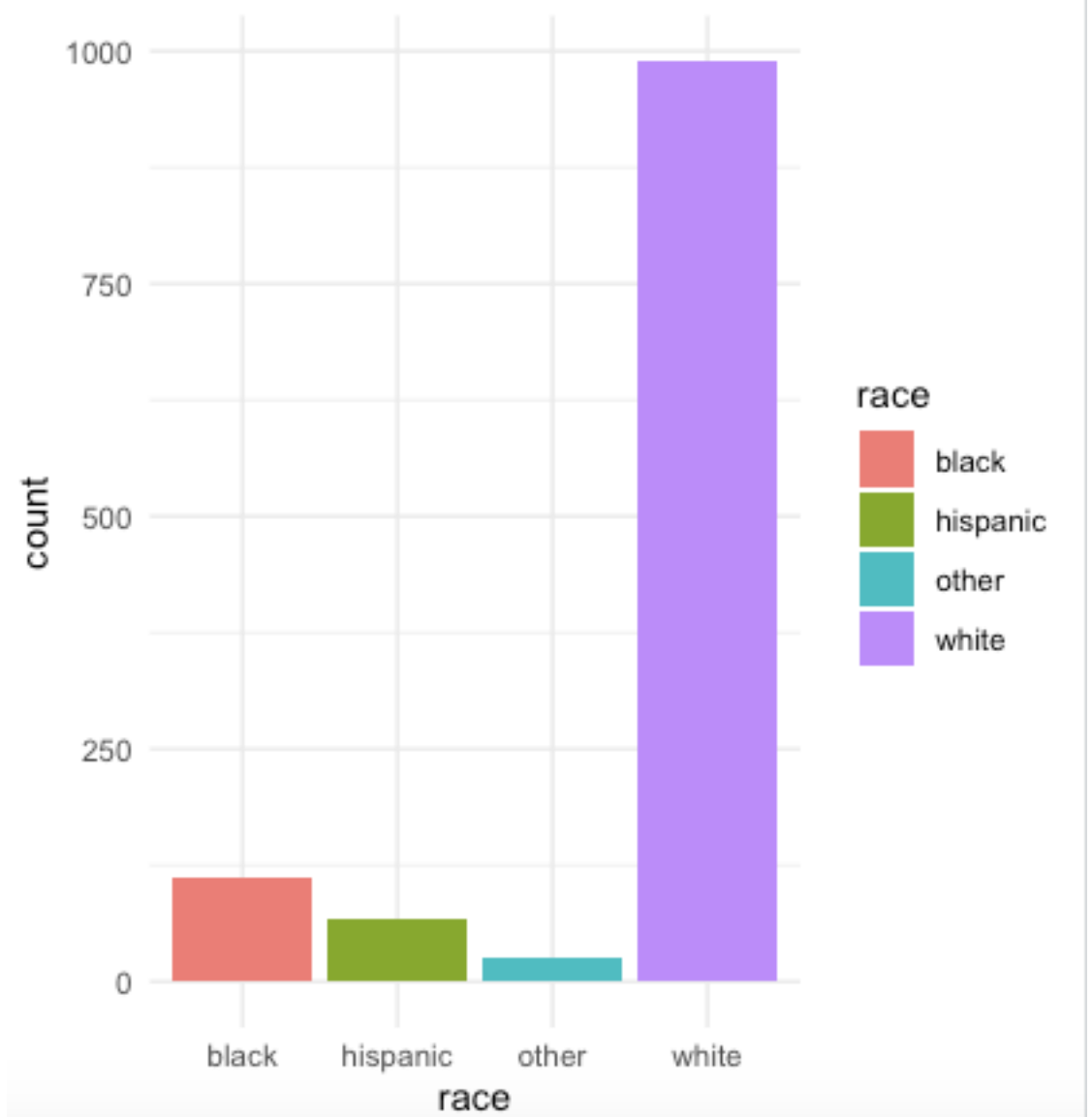
5. Using `geom_bar()` plot a bar chart of the number of records for each 'sex'

a. `ggplot(heights_df, aes(sex))+geom_bar(aes(fill = sex))`



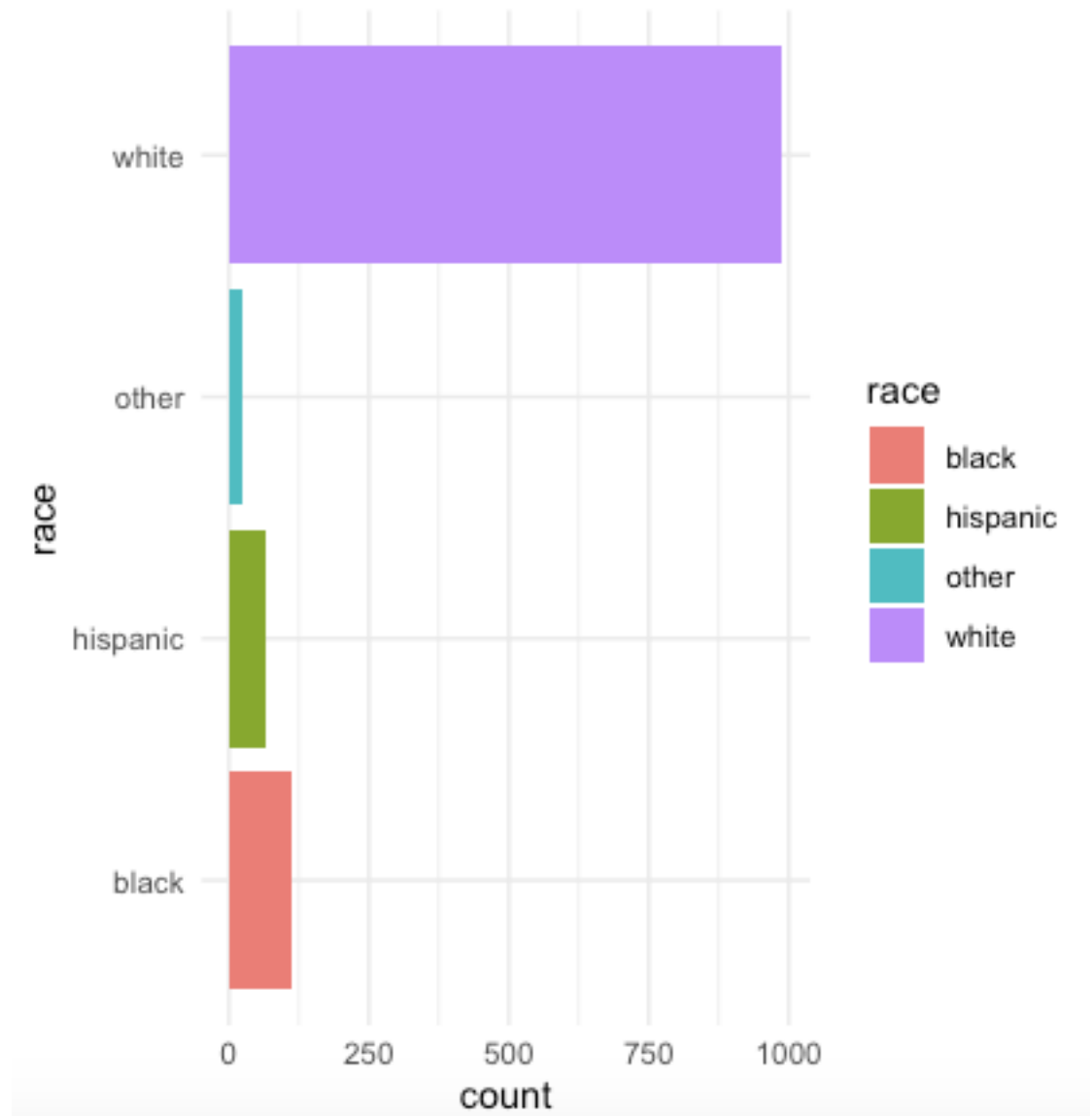
6. Using `geom_bar()` plot a bar chart of the number of records for each race

a. `ggplot(heights_df, aes(race))+geom_bar(aes(fill = race))`



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

a. `ggplot(heights_df, aes(race))+geom_bar(aes(fill = race))+coord_flip()`



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

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

	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	06	1	0
7	2020-01-25	Illinois	17	1	0
8	2020-01-25	Washington	53	1	0
9	2020-01-26	Arizona	04	1	0
10	2020-01-26	California	06	2	0
11	2020-01-26	Illinois	17	1	0
12	2020-01-26	Washington	53	1	0
13	2020-01-27	Arizona	04	1	0
14	2020-01-27	California	06	2	0
15	2020-01-27	Illinois	17	1	0

Showing 1 to 16 of 3,039 entries, 5 total columns

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

a. `covid_df$date <- as.Date(covid_df$date, "%m/%d/%y")`

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


```
a. california_df <- covid_df[ which( covid_df$state == "California"), ]
```

	date	state	fips	cases	deaths
1	2020-01-25	California	06	1	0
2	2020-01-26	California	06	2	0
3	2020-01-27	California	06	2	0
4	2020-01-28	California	06	2	0
5	2020-01-29	California	06	2	0
6	2020-01-30	California	06	2	0
7	2020-01-31	California	06	3	0
8	2020-02-01	California	06	3	0
9	2020-02-02	California	06	6	0
10	2020-02-03	California	06	6	0
11	2020-02-04	California	06	6	0
12	2020-02-05	California	06	6	0
13	2020-02-06	California	06	6	0
14	2020-02-07	California	06	6	0
15	2020-02-08	California	06	6	0

Showing 1 to 16 of 93 entries, 5 total columns

b. `ny_df <- covid_df[which(covid_df$state == "New York"),]`

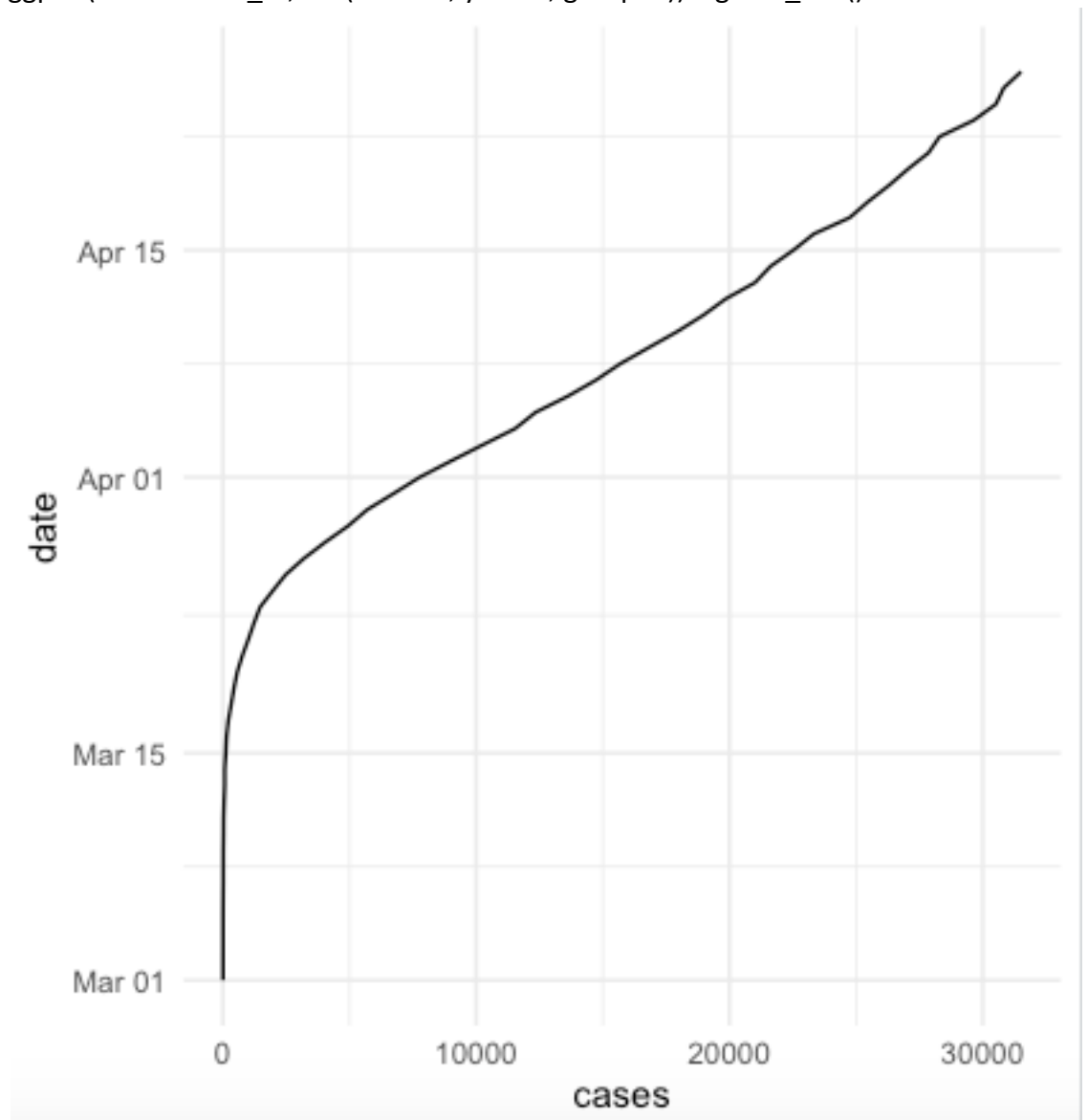
	date	state	fips	cases	deaths
1	2020-03-01	New York	36	1	0
2	2020-03-02	New York	36	1	0
3	2020-03-03	New York	36	2	0
4	2020-03-04	New York	36	11	0
5	2020-03-05	New York	36	22	0
6	2020-03-06	New York	36	44	0
7	2020-03-07	New York	36	89	0
8	2020-03-08	New York	36	106	0
9	2020-03-09	New York	36	142	0
10	2020-03-10	New York	36	173	0
11	2020-03-11	New York	36	217	0
12	2020-03-12	New York	36	326	0
13	2020-03-13	New York	36	421	0
14	2020-03-14	New York	36	610	2
15	2020-03-15	New York	36	732	6
Showing 1 to 16 of 57 entries, 5 total columns					

c. `florida_df <- covid_df[which(covid_df$state == "Florida"),]`

	▲	date	▼	state	▼	fips	▼	cases	▼	deaths	▼
1		2020-03-01		Florida		12		2		0	
2		2020-03-02		Florida		12		2		0	
3		2020-03-03		Florida		12		3		0	
4		2020-03-04		Florida		12		3		0	
5		2020-03-05		Florida		12		4		0	
6		2020-03-06		Florida		12		7		2	
7		2020-03-07		Florida		12		12		2	
8		2020-03-08		Florida		12		13		2	
9		2020-03-09		Florida		12		14		2	
10		2020-03-10		Florida		12		23		2	
11		2020-03-11		Florida		12		26		2	
12		2020-03-12		Florida		12		46		2	
13		2020-03-13		Florida		12		70		2	
14		2020-03-14		Florida		12		70		3	
15		2020-03-15		Florida		12		109		3	
Showing 1 to 16 of 57 entries, 5 total columns											

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

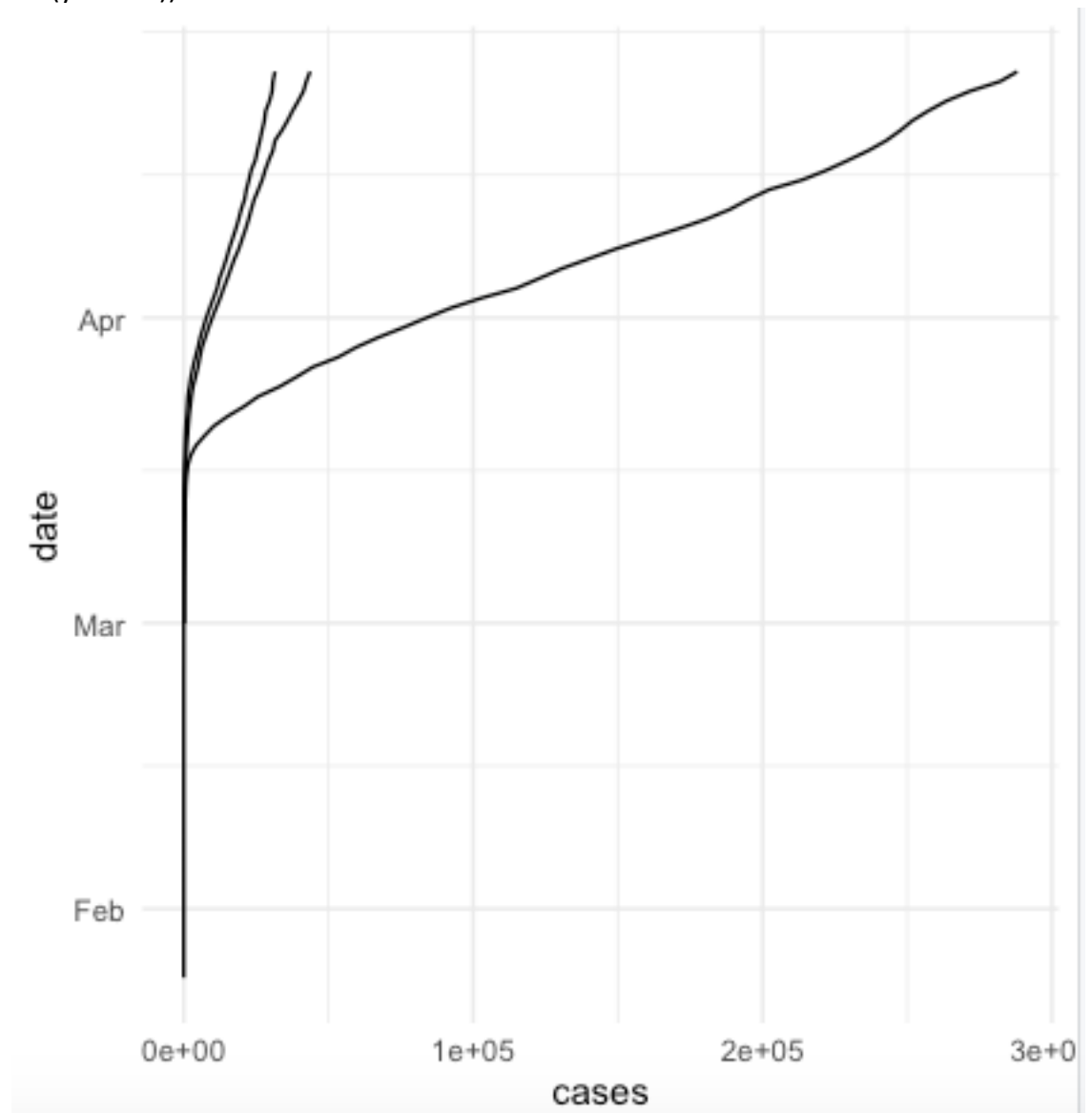
a. `ggplot(data=florida_df, aes(x=cases, y=date, group=1)) + geom_line()`



12. Add lines for New York and California to the plot

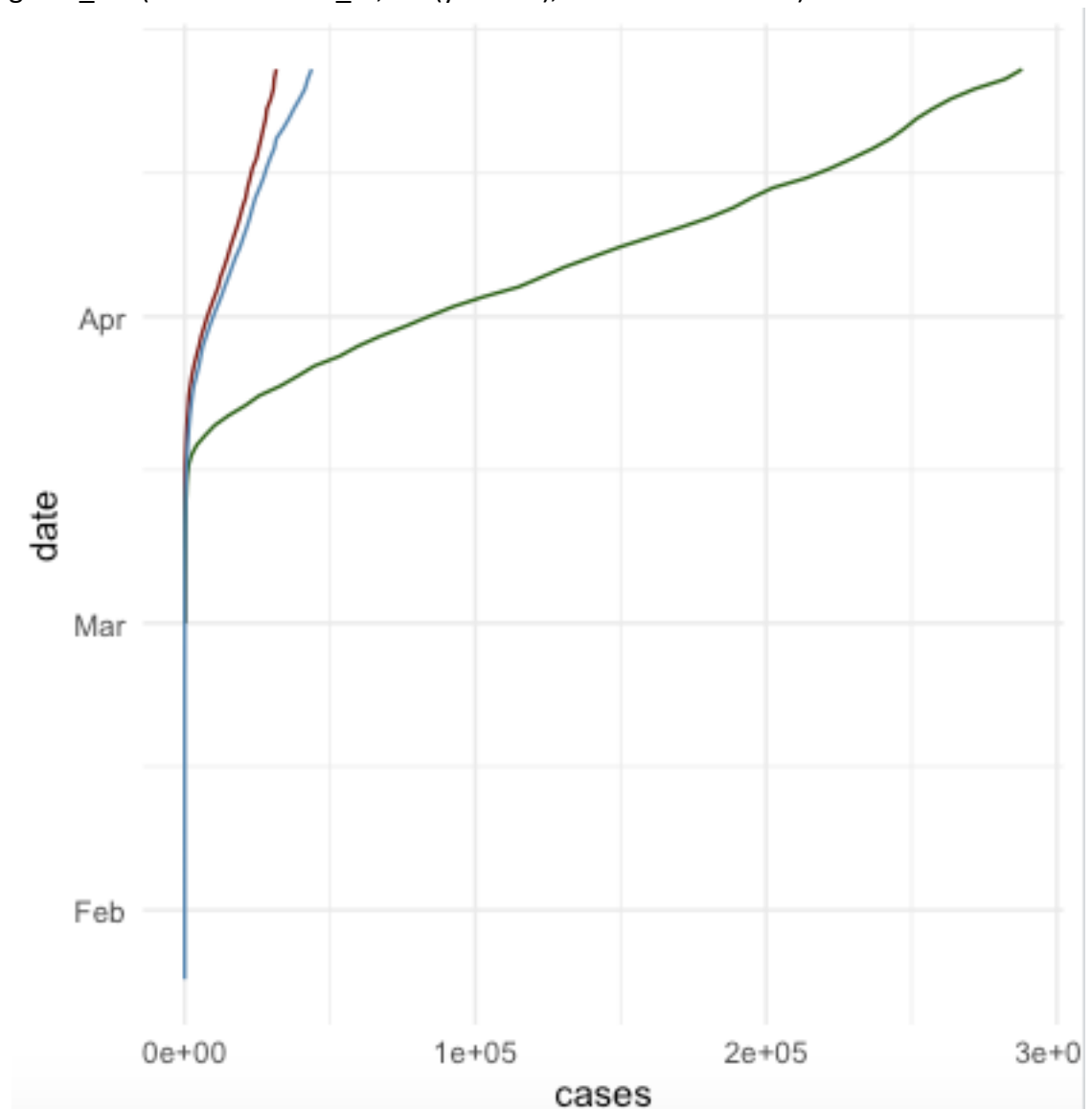
a. `ggplot(data=florida_df, aes(x=cases, y=date, group=1)) + geom_line(aes(y = date)) + geom_line(data=ny_df, aes(y = date)) + geom_line(data=california_df,`

```
aes(y = date))
```



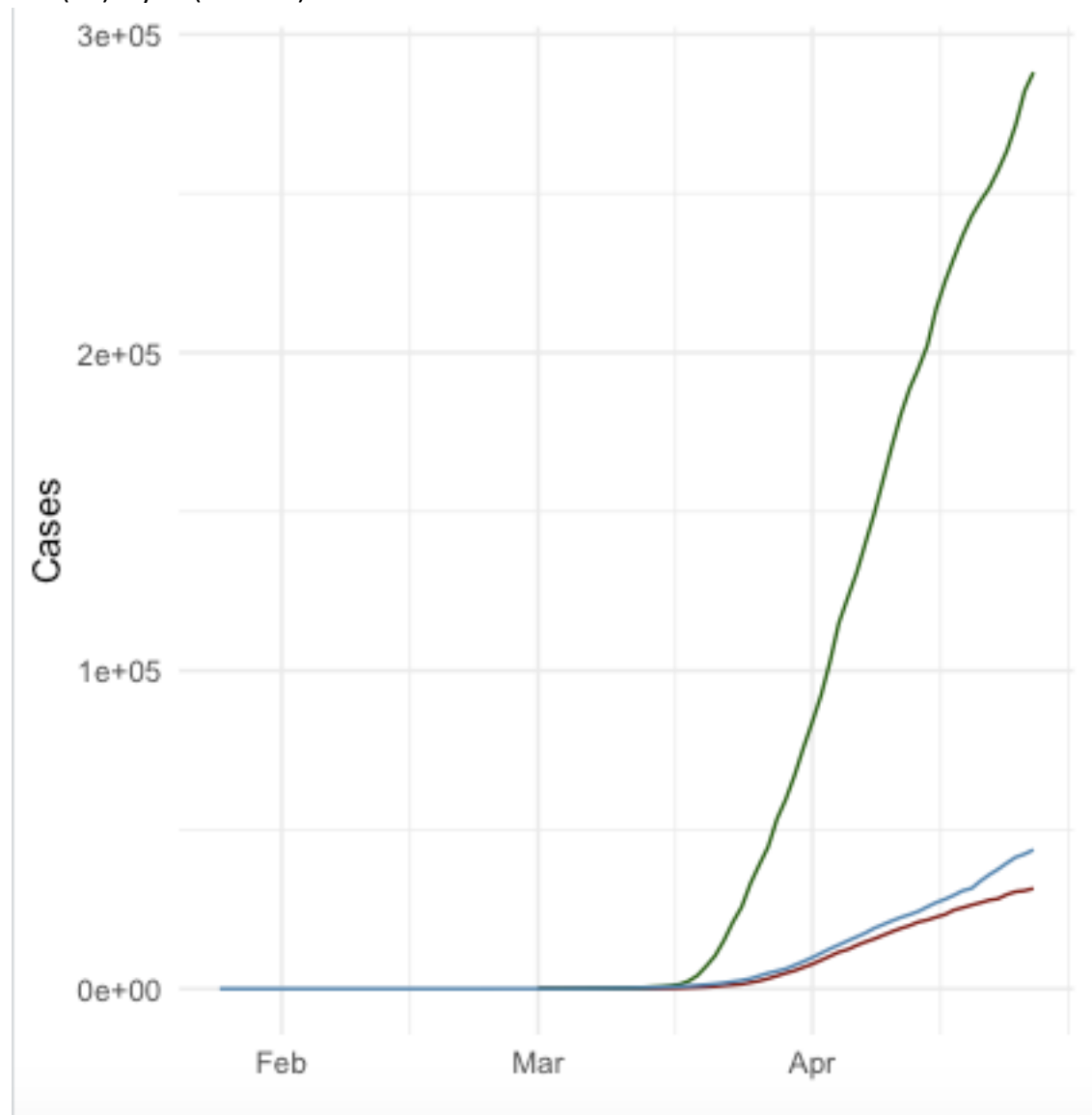
13. Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
- `ggplot(data=florida_df, aes(x=cases, group=1)) + geom_line(aes(y = date), color = "darkred") + geom_line(data=ny_df, aes(y = date), color = "darkgreen") +`

```
geom_line(data=california_df, aes(y = date), color = "steelblue")
```



14. 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
 - a. `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("10,000", "20,000", "30,000"), values = c("Florida"="darkred", "New York"="darkgreen", "California"="steelblue")) +`

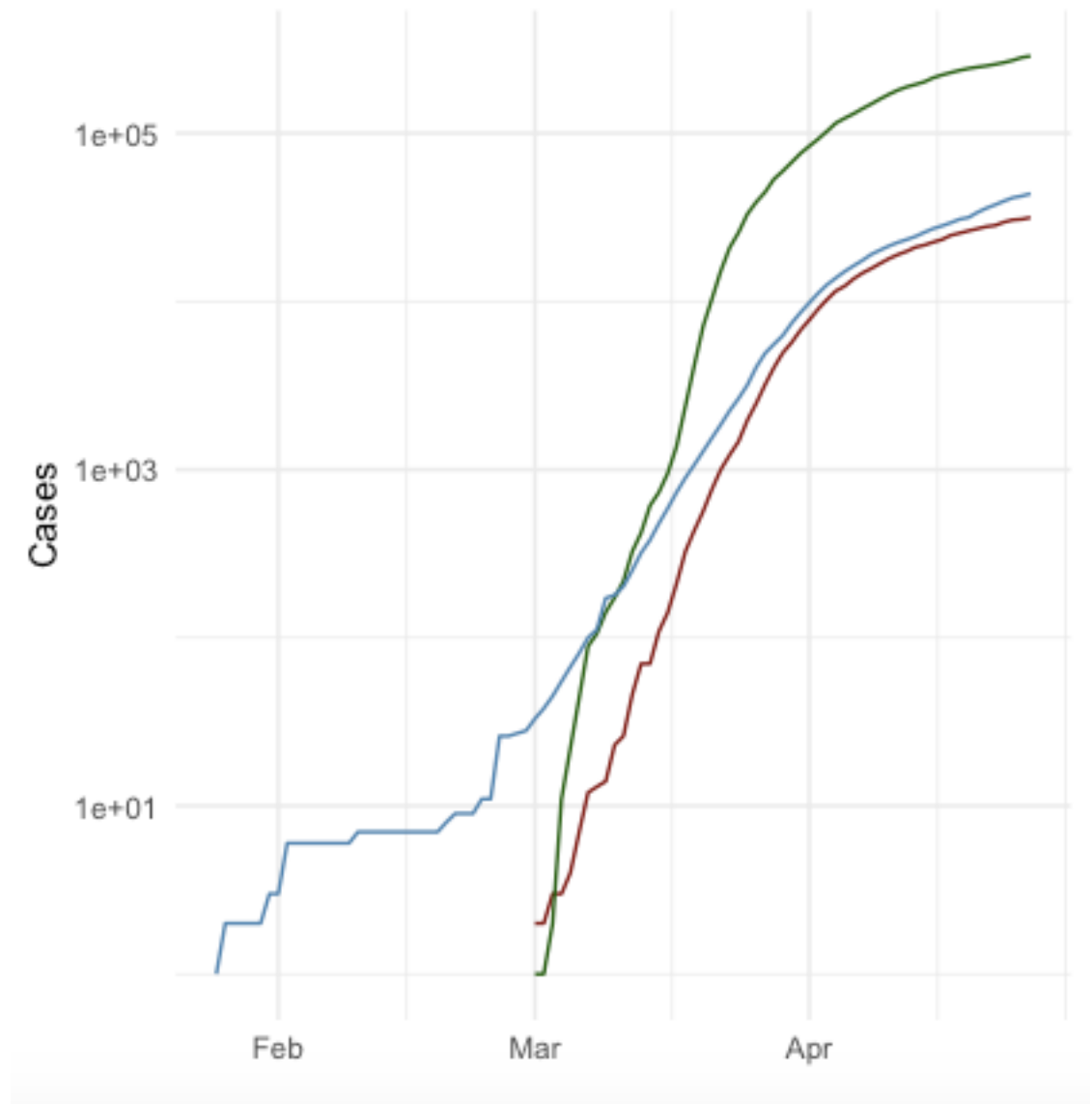
```
xlab(" ") + ylab("Cases")
```



15. Scale the y axis using ``scale_y_log10()``

- a. `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("10,000", "20,000", "30,000"), values = c("Florida"="darkred", "New York"="darkgreen", "California"="steelblue")) +`

```
xlab(" ") + ylab("Cases") + scale_y_log10()
```



Raw Code for Assignment 04

```
# Assignment: ASSIGNMENT 4  
# Name: Koby-Hercsky, Theodore  
# Date: 2021-04-20
```

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

```
## Set the working directory to the root of your DSC 520 directory
```



```

setwd("Documents/Bellevue University Classes/DSC520")

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

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

# 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(aes(fill = sex))

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

## Create a horizontal bar chart by adding `coord_flip()` to the previous plot
ggplot(heights_df, aes(race))+geom_bar(aes(fill = race))+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")

## Parse the date column using `as.Date()`
covid_df$date <- as.Date(covid_df$date,"%m/%d/%y")

## Create three dataframes named `california_df`, `ny_df`, and `florida_df`
## containing the data from California, New York, and Florida
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()`
ggplot(data=florida_df, aes(x=cases, y=date, group=1)) + geom_line()

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

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

```

```
ggplot(data=florida_df, aes(x=cases, group=1)) + geom_line(aes(y = date), color = "darkred") +  
geom_line(data=ny_df, aes(y = date), color = "darkgreen") + geom_line(data=california_df,  
aes(y = date), 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("10,000", "20,000",  
"30,000"), values = c("Florida"="darkred", "New York"="darkgreen", "California"="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("10,000", "20,000",  
"30,000"), values = c("Florida"="darkred", "New York"="darkgreen", "California"="steelblue")) +  
xlab(" ") + ylab("Cases") + scale_y_log10()
```