```
# Assignment: ASSIGNMENT 4

# Name: Sylvest, Allison

# Date: 2021-04-25

Output:

> ## Load the ggplot2 package

> library(ggplot2)

> theme_set(theme_minimal())

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

> setwd("GitHub/dsc520/data/r4ds/")

Error in setwd("GitHub/dsc520/data/r4ds/") :

  cannot change working directory

> ## Load the `data/r4ds/heights.csv` to

> library(readr)

> heights_df <- read_csv("heights.csv")


-- Column specification --------------------------------------------------------

cols(

  earn = col_double(),

  height = col_double(),

  sex = col_character(),

  ed = col_double(),

  age = col_double(),

  race = col_character()

)


> View(heights)

Error in View : object 'heights' not found

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

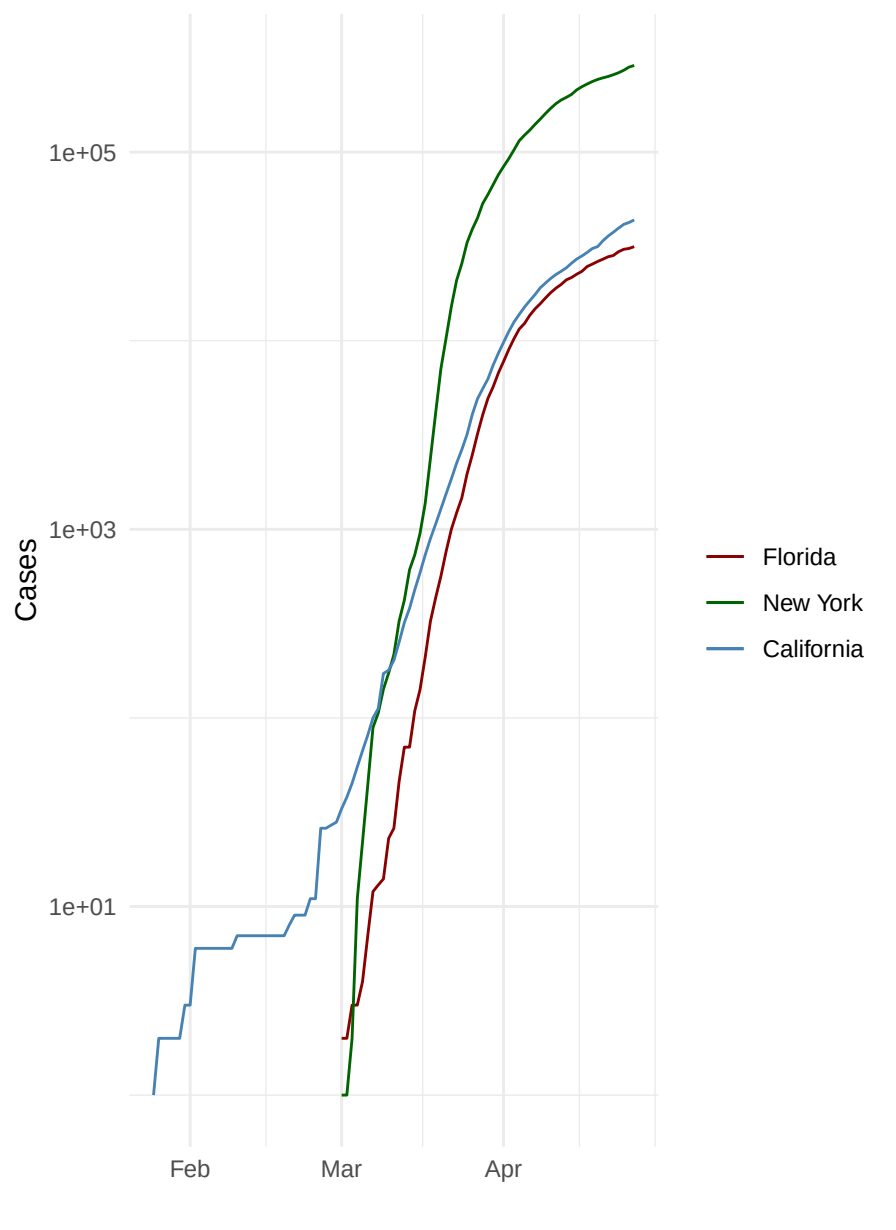
```
> ## Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`
> ## sex vs. earn
> ggplot(heights_df, aes(x=sex, y=earn)) + geom_point()+ geom_boxplot()
> ## race vs. earn
> ggplot(heights_df, aes(x=race, y=earn)) + geom_point()+ geom_boxplot()
> # https://ggplot2.tidyverse.org/reference/geom_bar.html
> ## 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
> 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("~/GitHub/dsc520/data/nytimes/covid-19-data/us-states.csv")

-- Column specification --------------------------------------------------------
cols(
  date = col_date(format = ""),
  state = col_character(),
  fips = col_character(),
  cases = col_double(),
  deaths = col_double()
)

> View(covid_df)
> ## 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[ 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= 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=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()
>
```

```
# Assignment: ASSIGNMENT 4
# Name: Sylvest, Allison
# Date: 2021-04-25

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

## Set the working directory to the root of your DSC 520 directory
setwd("GitHub/dsc520/data/r4ds/")

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

# 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
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("~/GitHub/dsc520/data/nytimes/covid-19-data/us-
states.csv")
View(covid_df)

## 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[ 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= 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=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()
```


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