

The results below are generated from an R script.

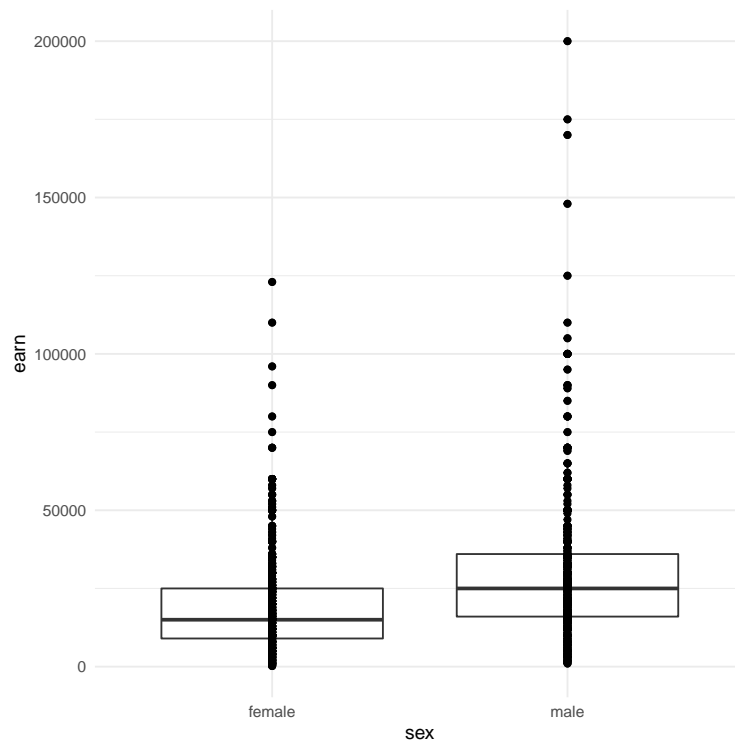
```
# Assignment: ASSIGNMENT 4
# Name: Wilson, Stewart
# Date: 2022-5-1

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

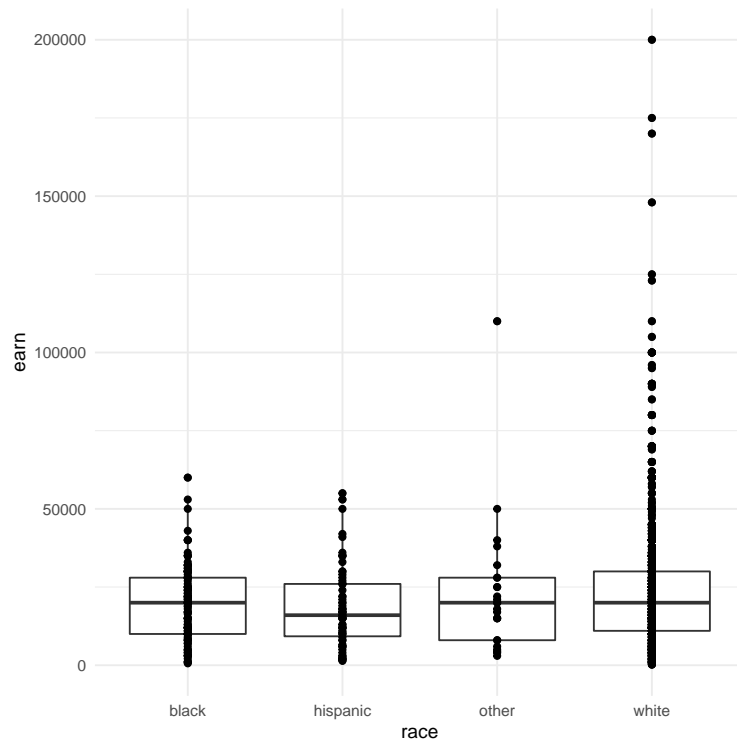
## Set the working directory to the root of your DSC 520 directory
setwd("/Users/Stewart/Documents/GitHub/dsc520")

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

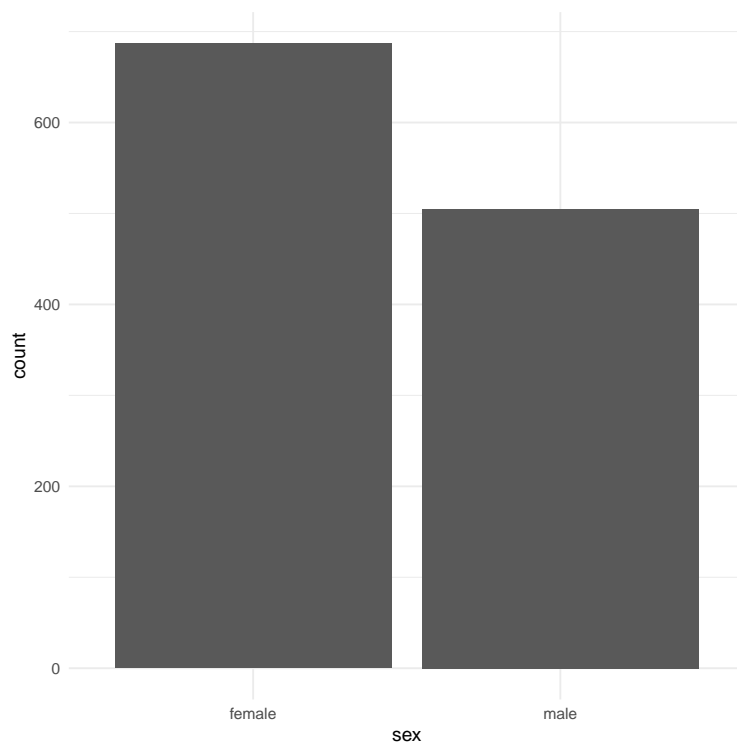
# https://ggplot2.tidyverse.org/reference/geom_boxplot.html
## Create boxplots of sex vs. earn and race vs. earn using 'geom_point()' and 'geom_boxplot()'
## sex vs. earn
ggplot(heights_df, aes(x=sex, y=earn)) + geom_boxplot() + geom_point()
```



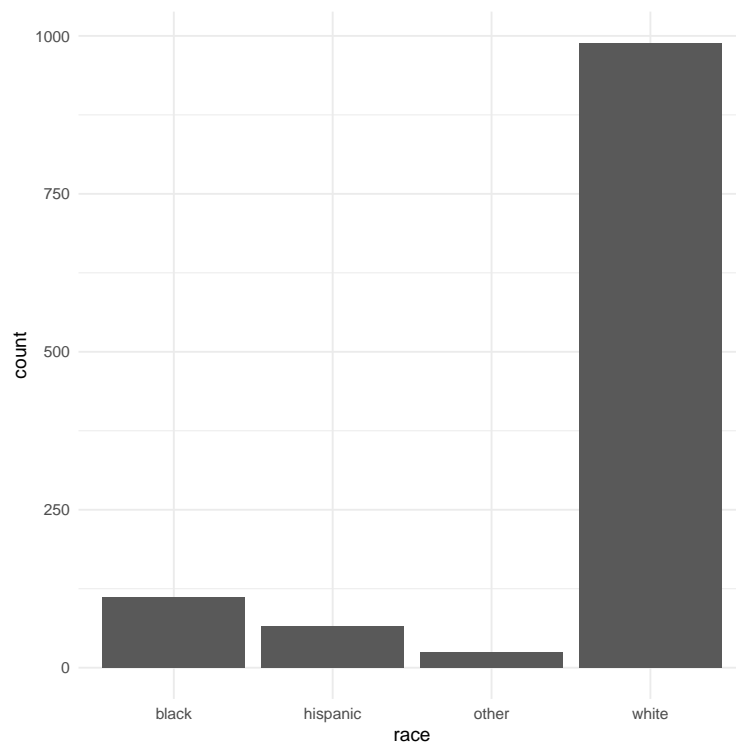
```
## race vs. earn
ggplot(heights_df, aes(x=race, y=earn)) + geom_boxplot() + geom_point()
```



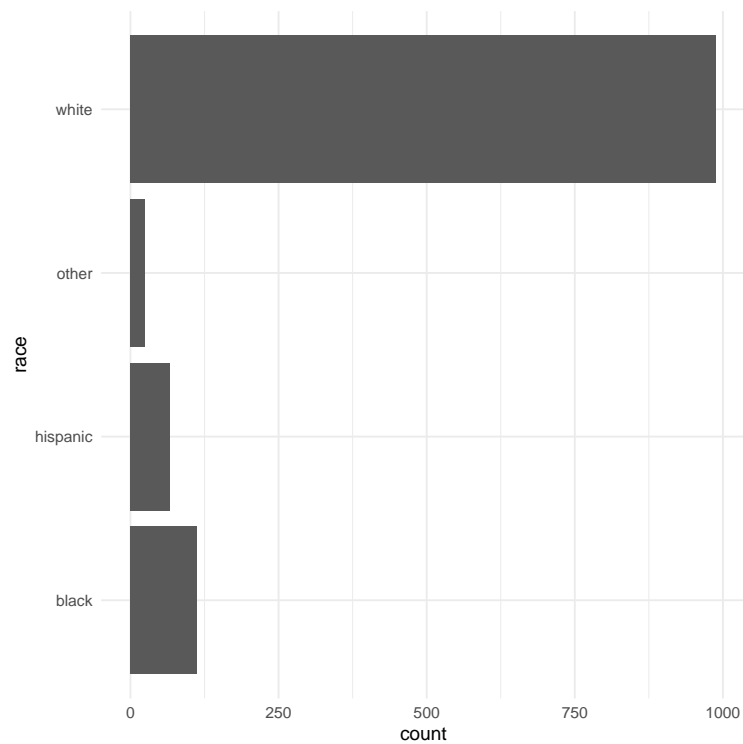
```
# https://ggplot2.tidyverse.org/reference/geom\_bar.html
## Using 'geom_bar()' plot a bar chart of the number of records for each 'sex'
ggplot(heights_df, aes(x=sex)) + geom_bar()
```



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



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

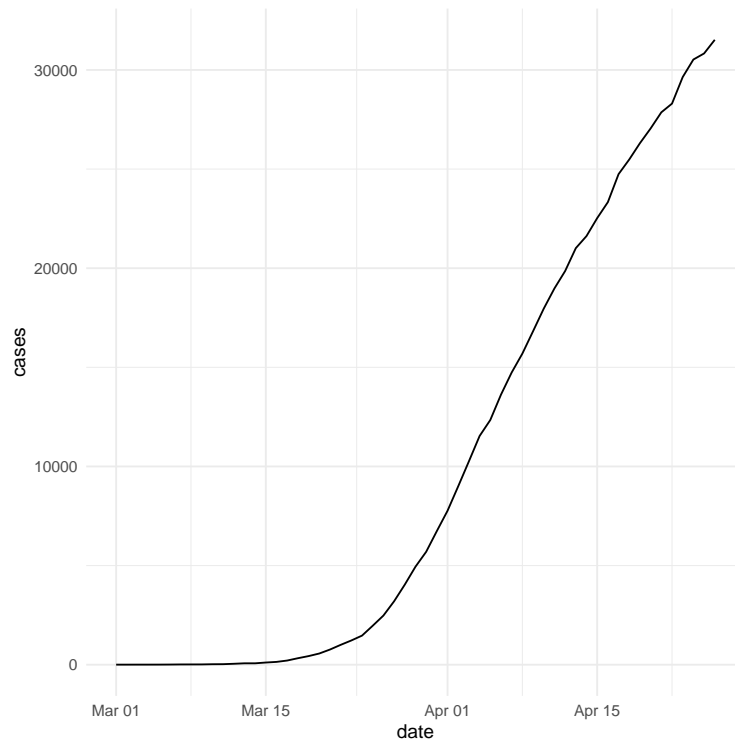


```
# https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0/topics/geom\_path
## Load the file "data/nytimes/covid-19-data/us-states.csv" and
## assign it to the 'covid_df' dataframe
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv")

## 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"), ]

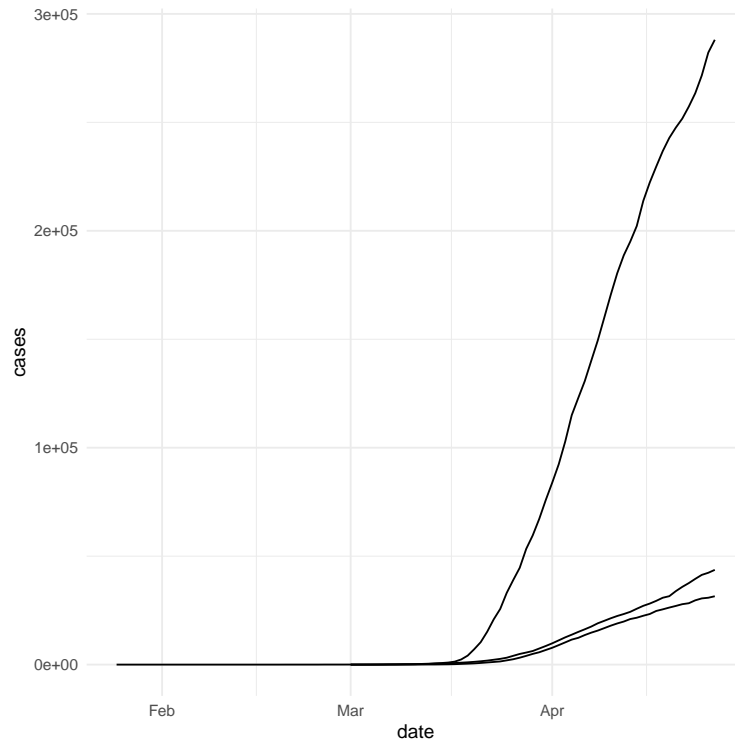
## Plot the number of cases in Florida using 'geom_line()'
ggplot(data=florida_df, aes(x=date, y=cases, group=1)) + geom_line()
```



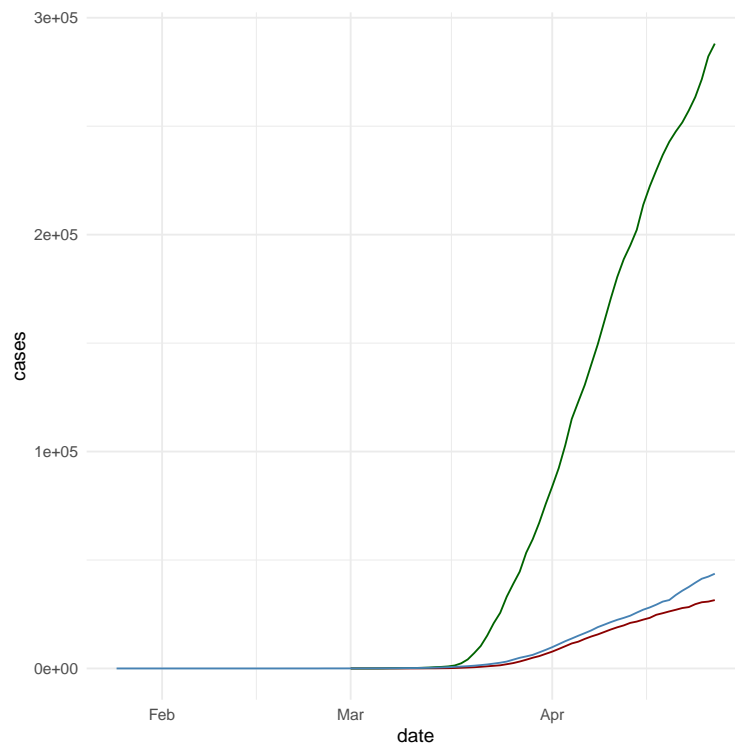
```
l## Add lines for New York and California to the plot

## Error in eval(expr, envir, enclos): object 'l' not found

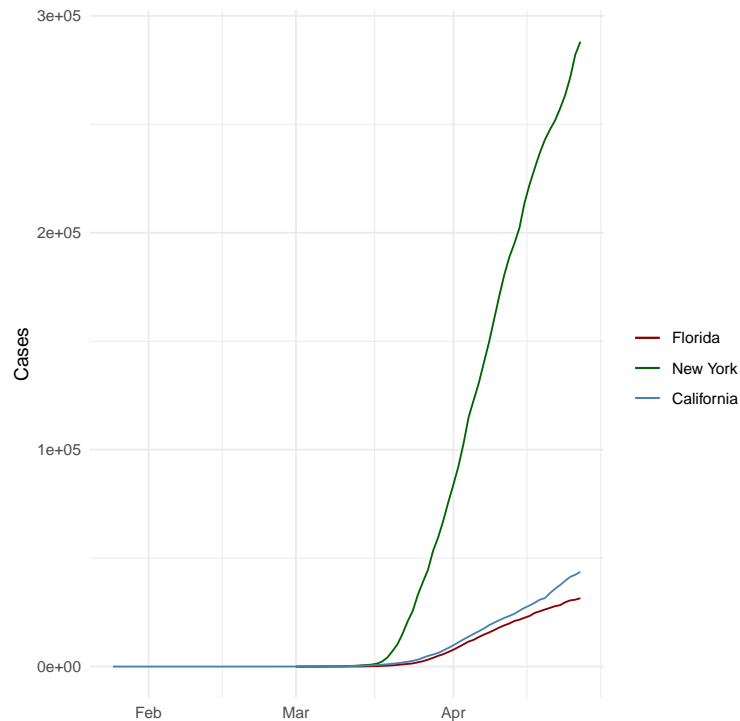
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases)) +
  geom_line(data=california_df, aes(y = cases)) +
  geom_line(data=ny_df, aes(y = cases))
```



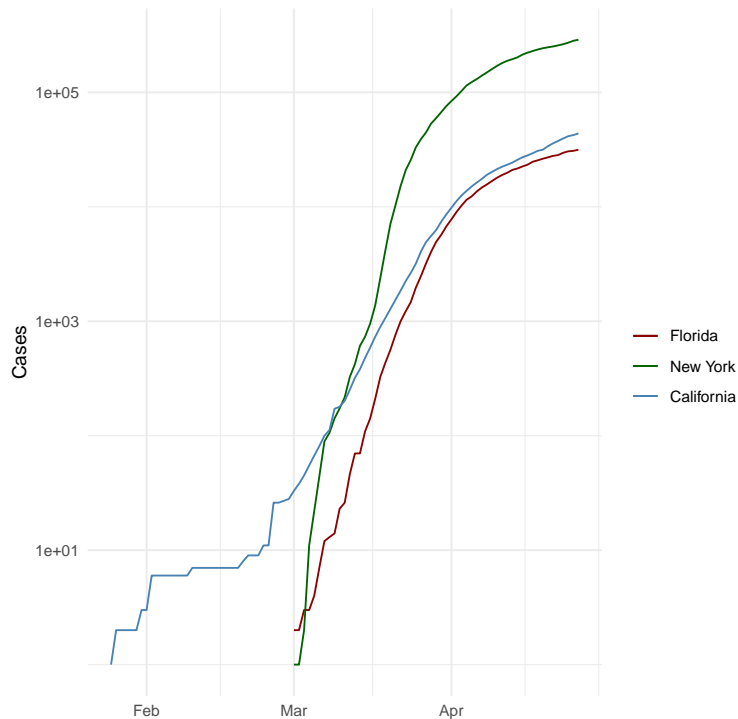
```
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = "darkred") +
  geom_line(data=ny_df, aes(y = cases), color="darkgreen") +
  geom_line(data=california_df, aes(y = cases),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



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.1.3 (2022-03-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252 LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252 LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] knitr_1.38      ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
## [1] highr_0.9      pillar_1.7.0   compiler_4.1.3 tools_4.1.3    digest_0.6.29
## [6] evaluate_0.15  lifecycle_1.0.1 tibble_3.1.6   gtable_0.3.0   pkgconfig_2.0.3
## [11] rlang_1.0.2    cli_3.2.0      rstudioapi_0.13 yaml_2.3.5     xfun_0.30
## [16] fastmap_1.1.0  withr_2.5.0    dplyr_1.0.8    stringr_1.4.0  generics_0.1.2
## [21] vctrs_0.4.1    grid_4.1.3     tidyselect_1.1.2 glue_1.6.2     R6_2.5.1
## [26] fansi_1.0.2    rmarkdown_2.13 farver_2.1.0   purrr_0.3.4    magrittr_2.0.2
## [31] scales_1.1.1   ellipsis_0.3.2 htmltools_0.5.2 colorspace_2.0-3 labeling_0.4.2
## [36] utf8_1.2.2     stringi_1.7.6  munsell_0.5.0  crayon_1.5.0
```

```
Sys.time()
```

```
## [1] "2022-05-01 07:05:23 PDT"
```