

The results below are generated from an R script.

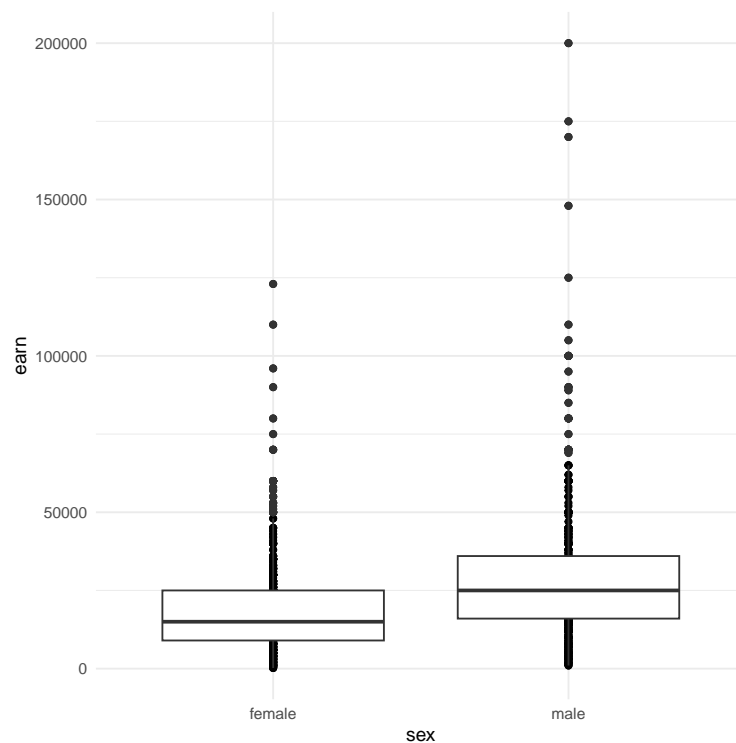
```
# Assignment: ASSIGNMENT 6
# Name: Reppeto, Reppeto
# Date: 2023-07-14

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

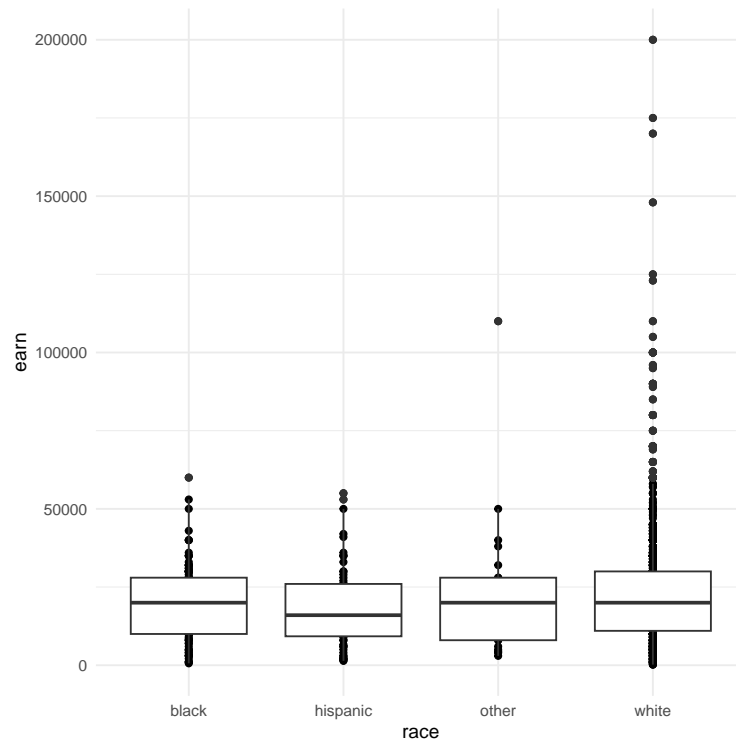
## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/dsc520")

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

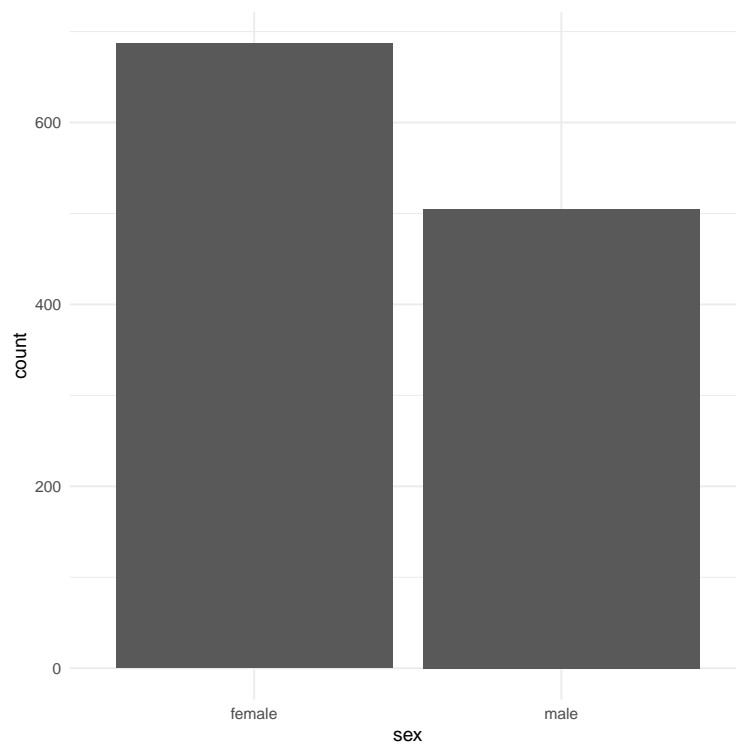
# https://ggplot2.tidyverse.org/reference/geom_boxplot.html
## Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`
## sex vs. earn
ggplot(heights_df, aes(x=sex, y=earn)) + geom_point()+ geom_boxplot()
```



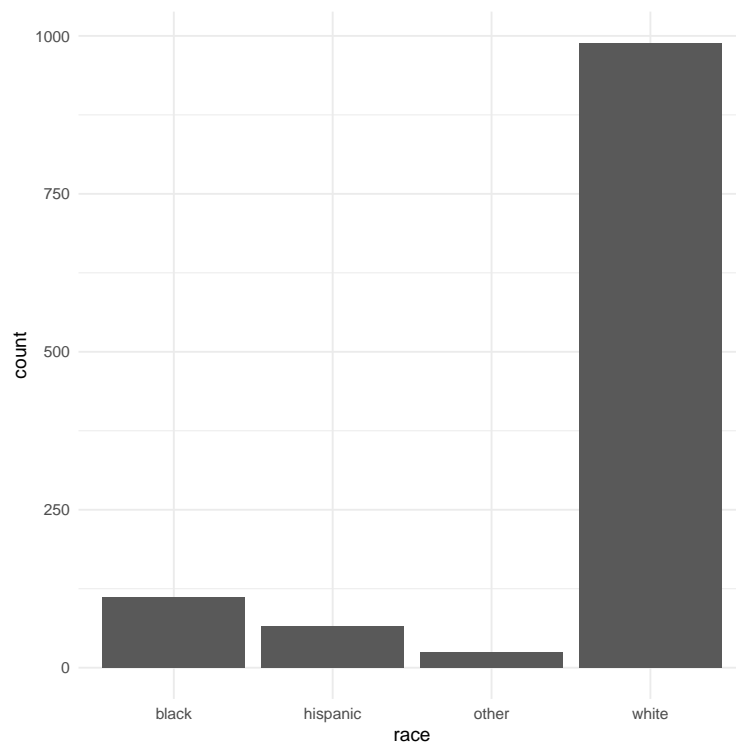
```
## race vs. earn
ggplot(heights_df, aes(x=race, y=earn)) + geom_point()+ geom_boxplot()
```



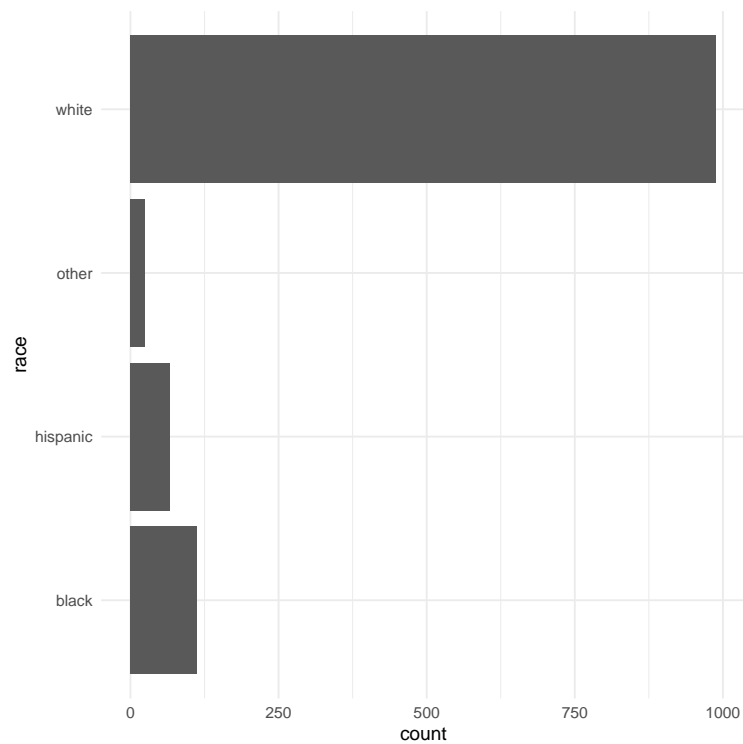
```
# https://ggplot2.tidyverse.org/reference/geom\_bar.html
## Using `geom_bar()` plot a bar chart of the number of records for each `sex`
ggplot(heights_df, aes(sex)) + geom_bar()
```



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

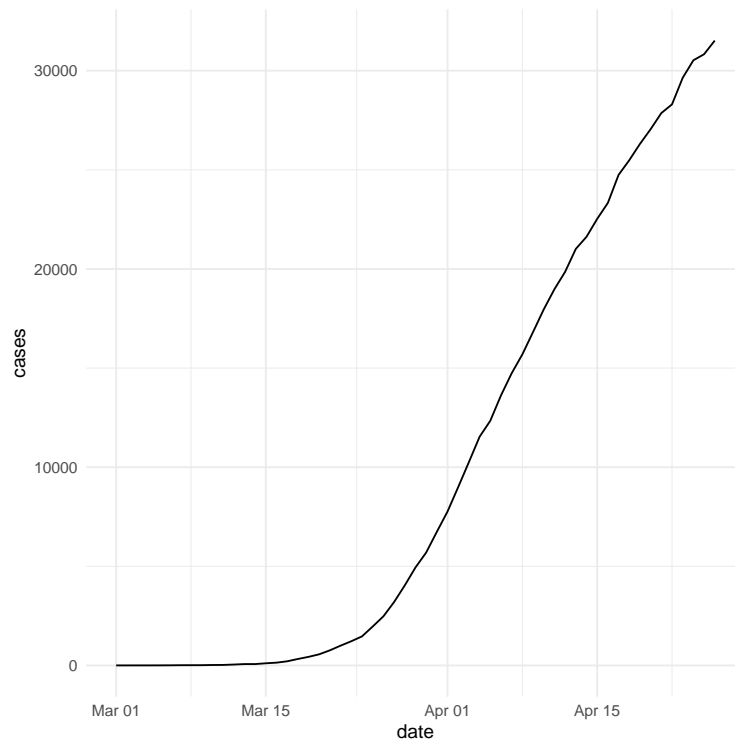
## Parse the date column using `as.Date()`
covid_df$date <- as.Date(covid_df$date)

## 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$state == "California"), ]
ny_df <- covid_df[covid_df$state == "New York", ]
florida_df <- covid_df[covid_df$state == "Florida", ]

#View(ny_df)

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

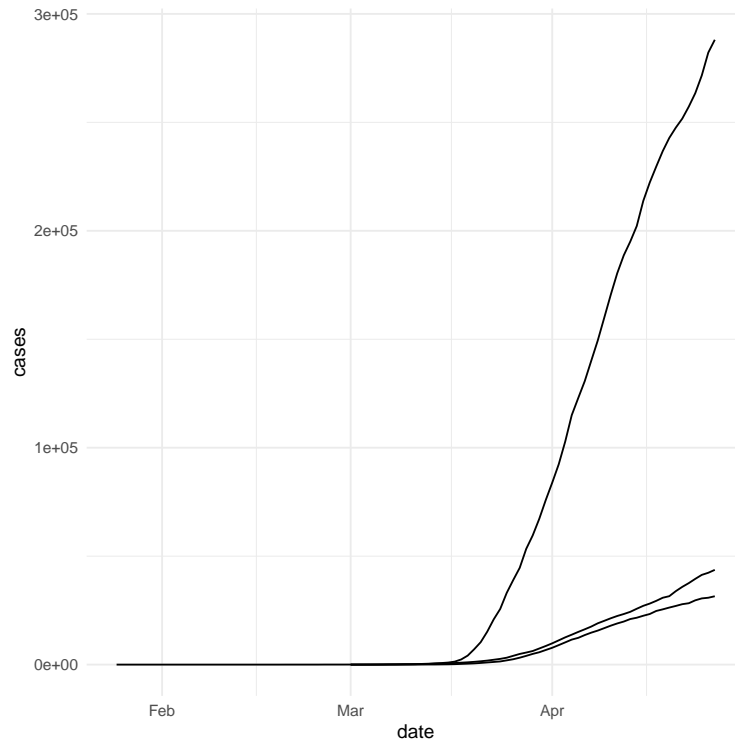
```



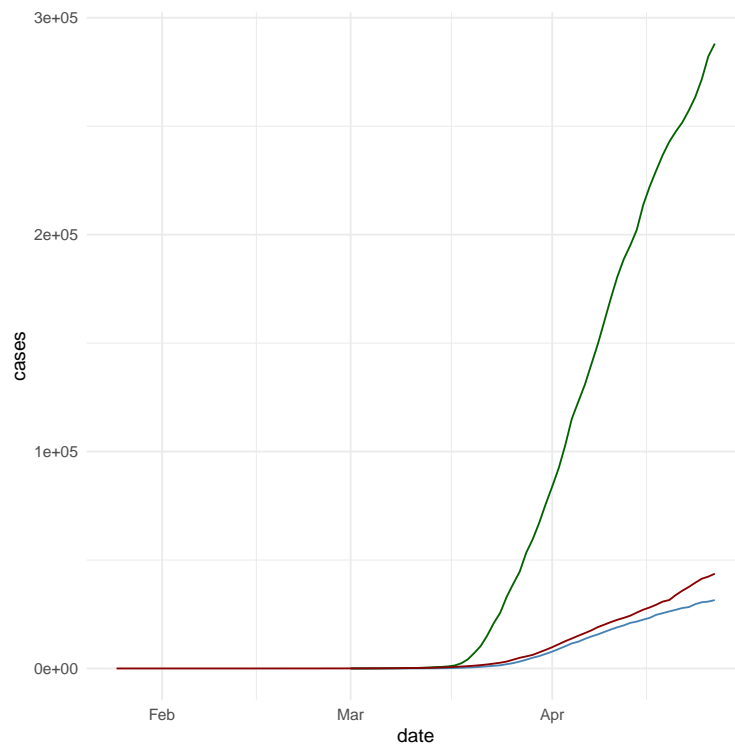
```

## 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))

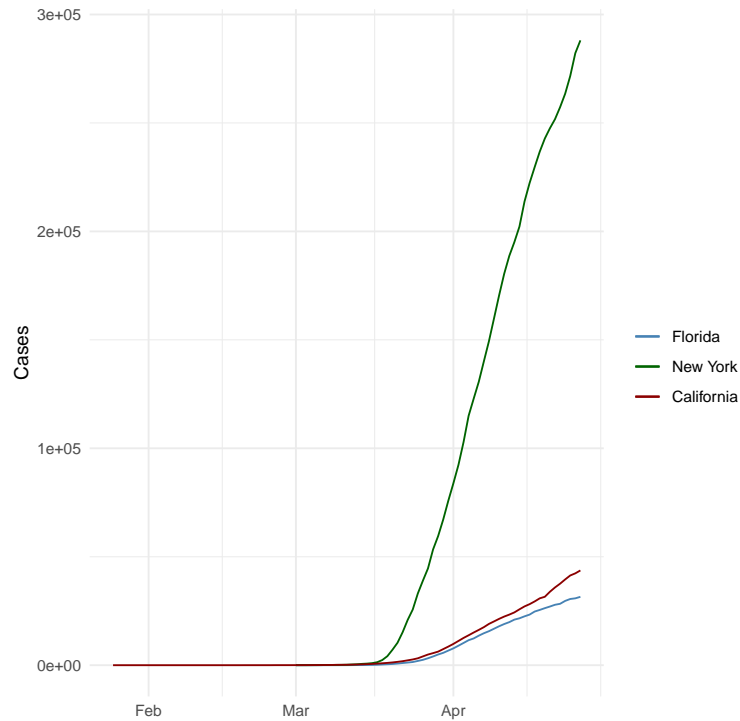
```



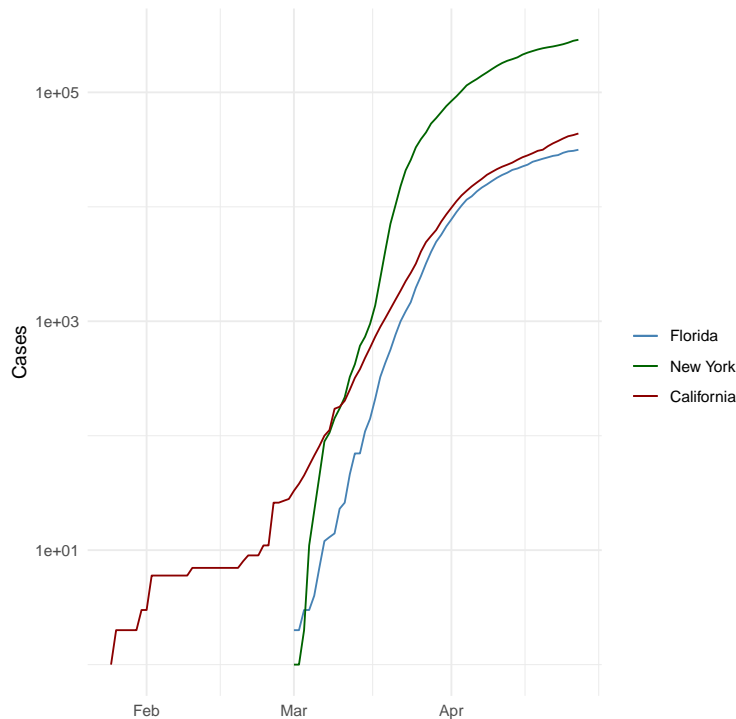
```
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = "steelblue") +
  geom_line(data=ny_df, aes(y = cases), color= "darkgreen") +
  geom_line(data=california_df, aes(y =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



```
## 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("steelblue", "darkgreen", "darkred")) +
  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("steelblue", "darkgreen", "darkred")) +
  xlab(" ") + ylab("Cases") + scale_y_log10()
```



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

```
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/Libraries/libBLAS.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK version 3.11.0
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] pandoc_0.1.0  knitr_1.43    tinytex_0.45  readr_2.1.4  ggplot2_3.4.2
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.3    dplyr_1.1.2    compiler_4.3.0  highr_0.10     tidyselect_1.2.0
## [6] scales_1.2.1    yaml_2.3.7     fastmap_1.1.1   R6_2.5.1       labeling_0.4.2
## [11] generics_0.1.3  tibble_3.2.1   munsell_0.5.0   pillar_1.9.0   tzdb_0.4.0
## [16] rlang_1.1.1     utf8_1.2.3     xfun_0.39       fs_1.6.2       cli_3.6.1
## [21] withr_2.5.0     magrittr_2.0.3 digest_0.6.31   grid_4.3.0     rstudioapi_0.14
## [26] rappdirs_0.3.3 hms_1.1.3      lifecycle_1.0.3 vctrs_0.6.3    evaluate_0.21
```

```
## [31] glue_1.6.2      farver_2.1.1     fansi_1.0.4      colorspace_2.1-0 rmarkdown_2.23
## [36] tools_4.3.0     pkgconfig_2.0.3  htmltools_0.5.5

Sys.time()

## [1] "2023-07-16 18:22:55 EDT"
```