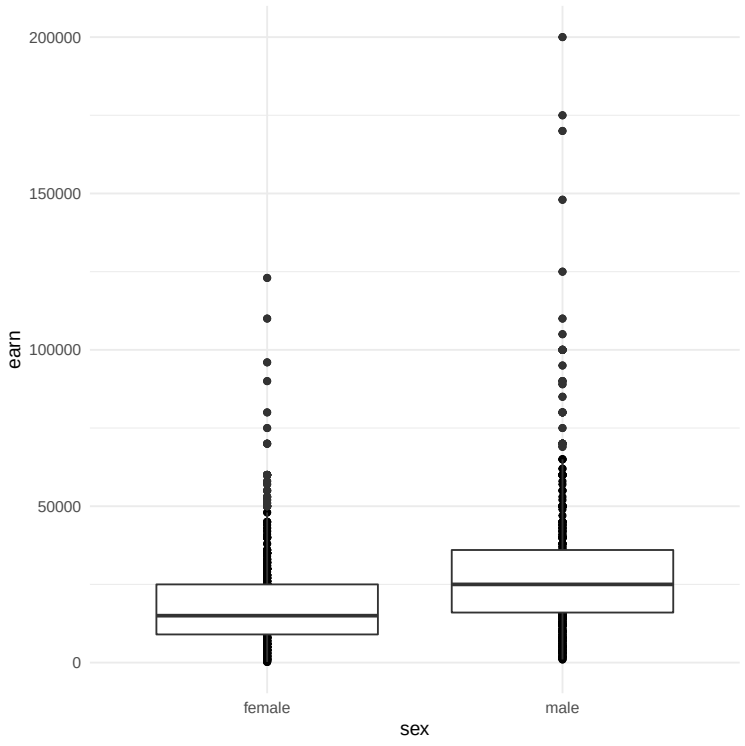
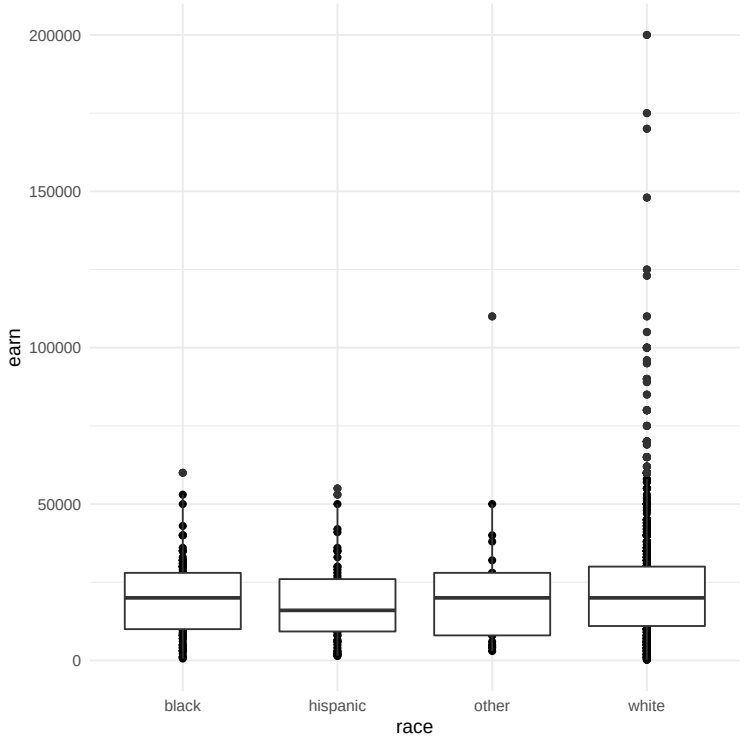
May 1, 2022

The results below are generated from an R script.

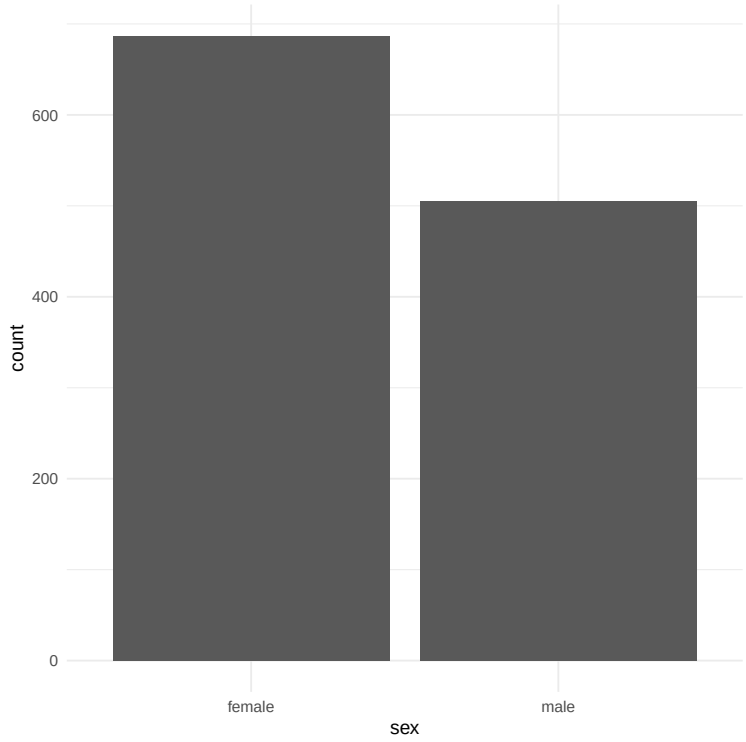
```
# Assignment: ASSIGNMENT 4
# Name: MacDonald, Mariana
# Date: 4/24/2022

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

## Set the working directory to the root of your DSC 520 directory
setwd("/Users/marianamacdonald/Documents/DATA SCIENCE/DSC 520/Statistics R/Week 2/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()
```



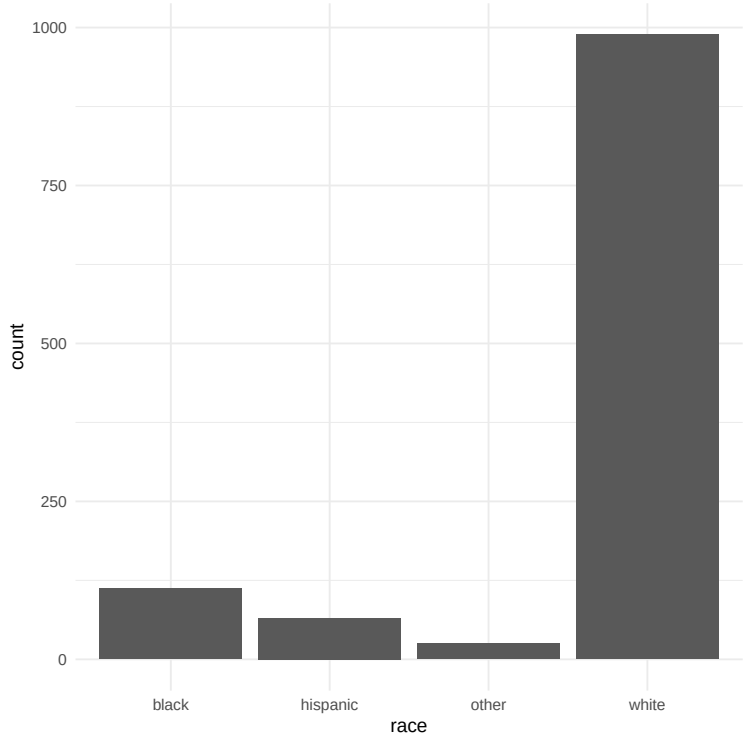
```
## race vs. earn
ggplot(heights_df, aes(x=race, y=earn)) + geom_point()+ geom_boxplot()
```



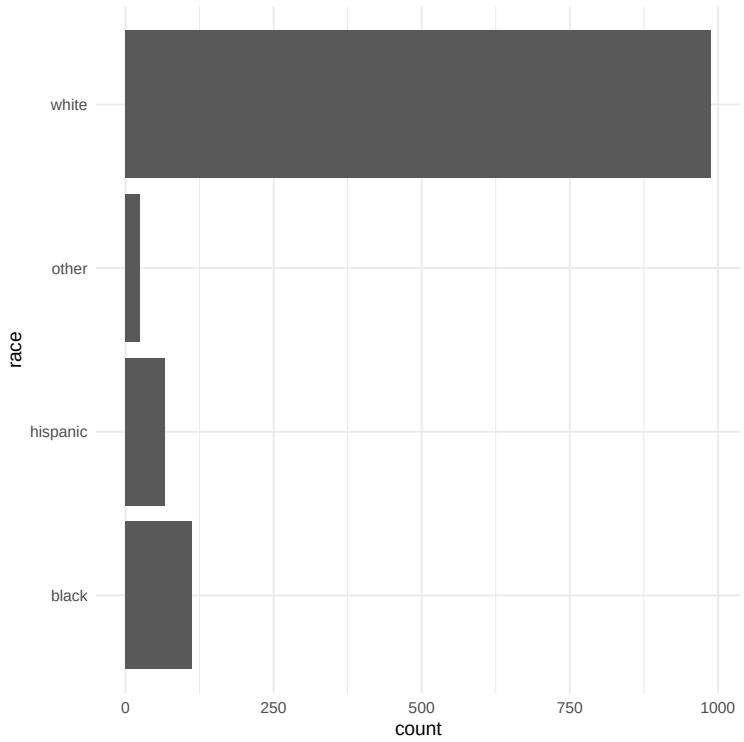

```
# https://ggplot2.tidyverse.org/reference/geom_bar.html
## Using `geom_bar()` plot a bar chart of the number of records for each `sex`
ggplot(heights_df, aes(sex)) + geom_bar()
```



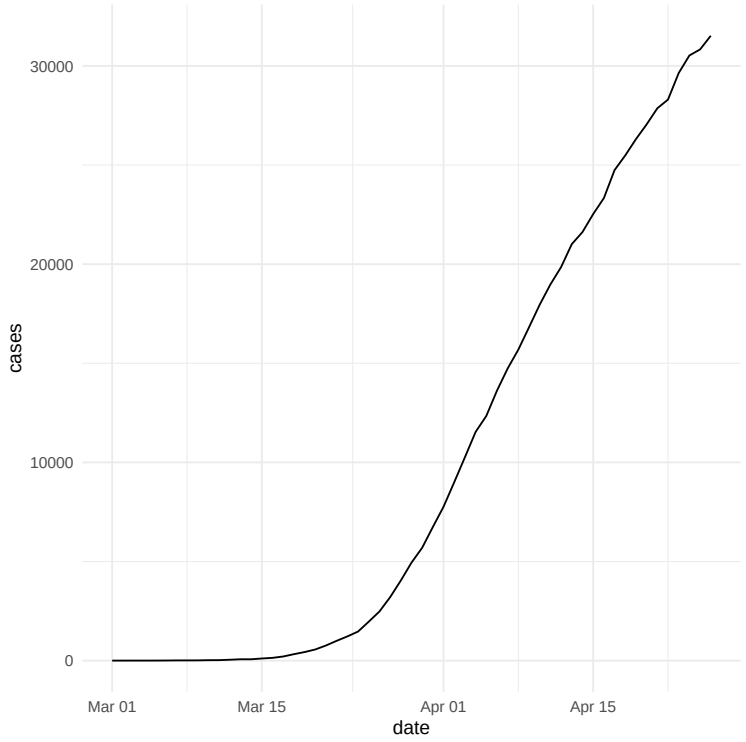

```
## Using `geom_bar()` plot a bar chart of the number of records for each race
ggplot(heights_df, aes(race)) + geom_bar()
```



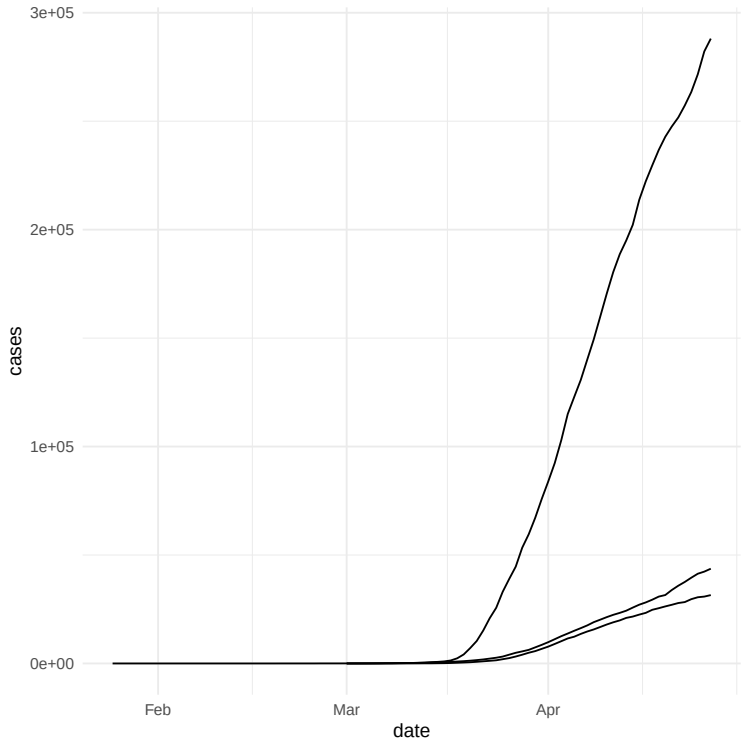
```r
## Create a horizontal bar chart by adding `coord_flip()` to the previous plot
ggplot(heights_df, aes(race)) + geom_bar() + coord_flip()
```



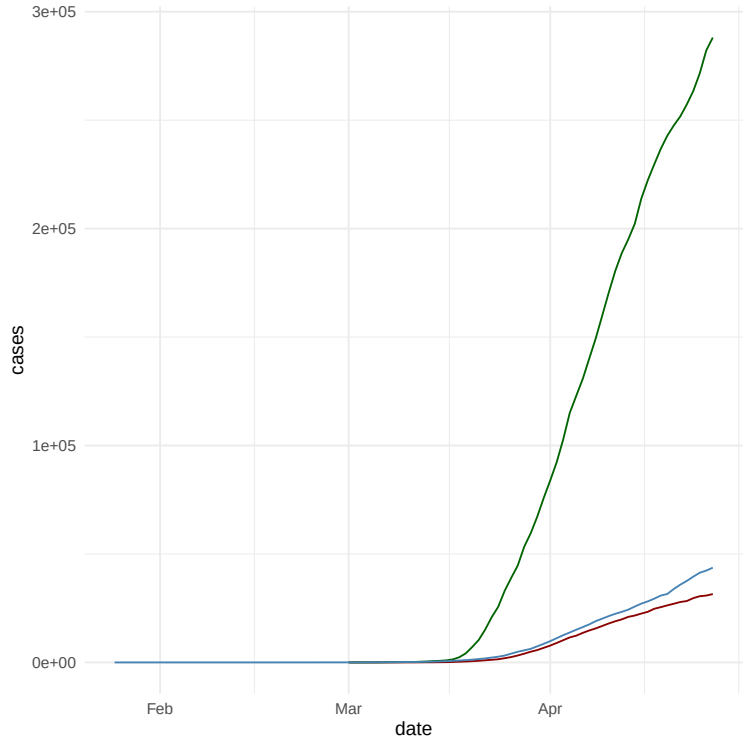

```r
# https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0/topics/geom_path
## Load the file `"data/nytimes/covid-19-data/us-states.csv"` and
## assign it to the `covid_df` dataframe
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv")

## 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()
```



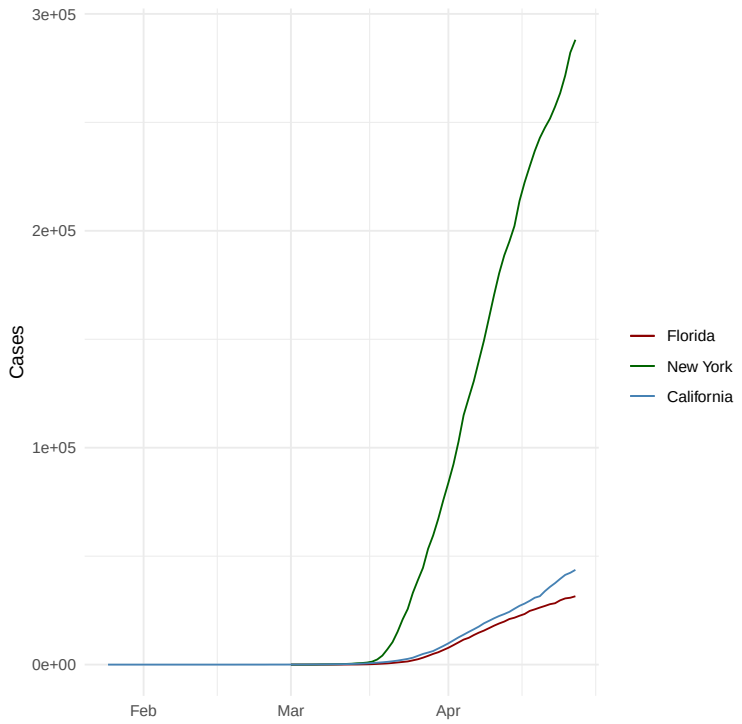

```
## Add lines for New York and California to the plot
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases)) +
  geom_line(data=ny_df, aes(y = cases)) +
  geom_line(data=california_df, aes(y = cases))
```



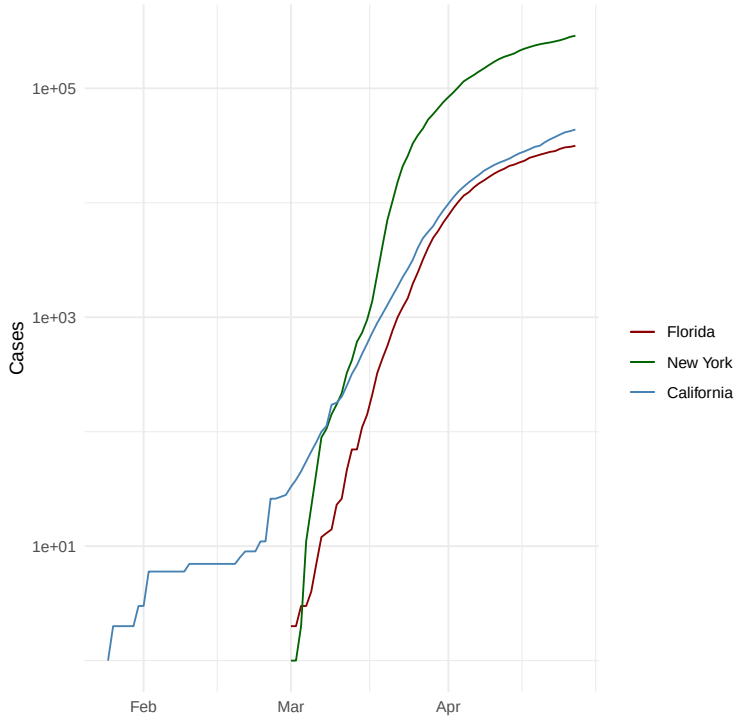
```r
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = "darkred") +
  geom_line(data=ny_df, aes(y = cases), color="darkgreen") +
  geom_line(data=california_df, aes(y = cases),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




```r
## 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-apple-darwin17.0 (64-bit)
## Running under: macOS Monterey 12.3.1
##
## Matrix products: default
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] tinytex_0.38  knitr_1.39    ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
##  [1] pillar_1.7.0     compiler_4.1.3   highr_0.9        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      DBI_1.1.2        cli_3.2.0        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_0.3.8      grid_4.1.3       tidyselect_1.1.2 glue_1.6.2       R6_2.5.1
## [26] fansi_1.0.3      rmarkdown_2.13   purrr_0.3.4      farver_2.1.0     magrittr_2.0.3
## [31] scales_1.1.1     ellipsis_0.3.2   htmltools_0.5.2  assertthat_0.2.1 colorspace_2.0-3
## [36] labeling_0.4.2   utf8_1.2.2       stringi_1.7.6    munsell_0.5.0    crayon_1.5.1

Sys.time()
```

```
## [1] "2022-05-01 15:52:30 EDT"
```