1. **sex-vs-earn**

## Load the ggplot2 package

library(ggplot2)

theme\_set(theme\_minimal())

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

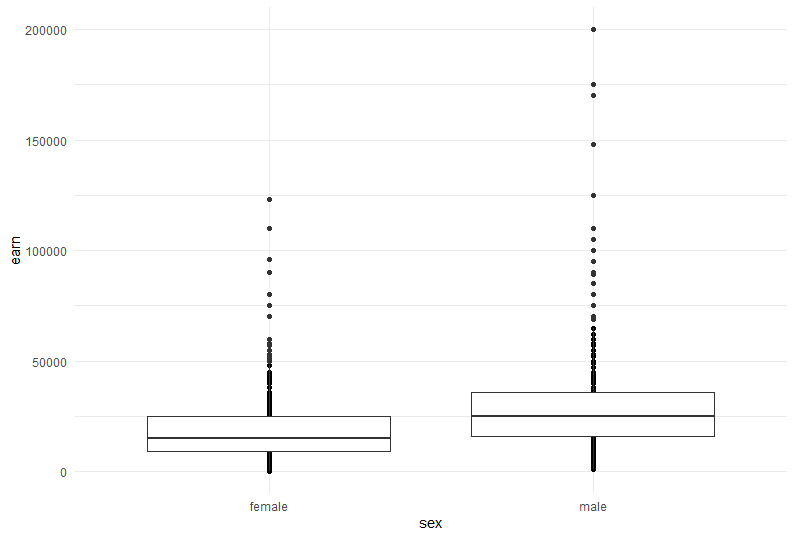
setwd("C:/Users/ragun/Documents/GitHub/dsc520-master/DSC520-new")

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

heights\_df <- read.csv("data/r4ds/heights.csv")

## sex vs. earn

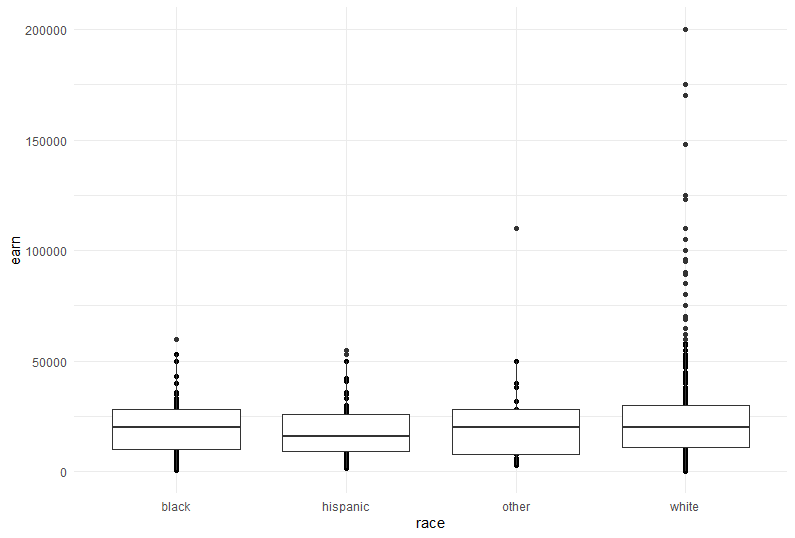
ggplot(heights\_df, aes(x=sex, y=earn)) + geom\_point()+ geom\_boxplot()



1. **Race Vs Earn**

## race vs. earn

ggplot(heights\_df, aes(x=race, y=earn)) + geom\_point()+ geom\_boxplot()

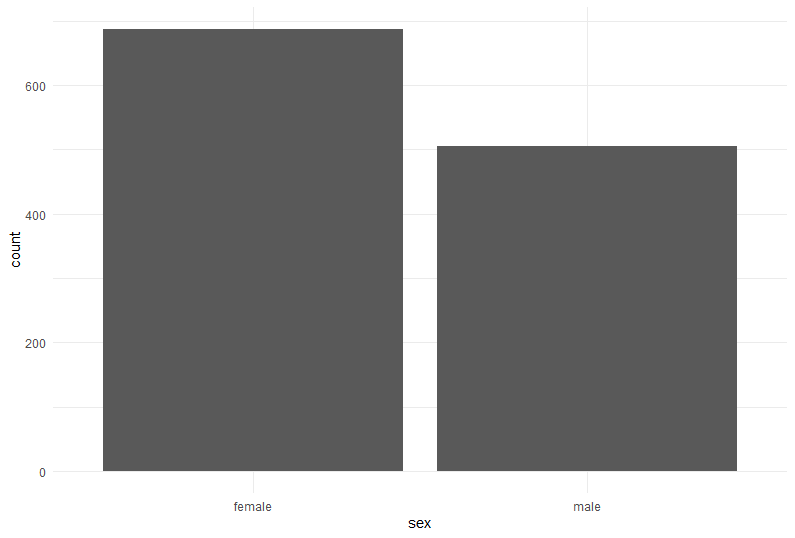


1. **sex-bar-chart**

## `ed` vs. `earn`

## Using `geom\_bar()` plot a bar chart of the number of records for each `sex`

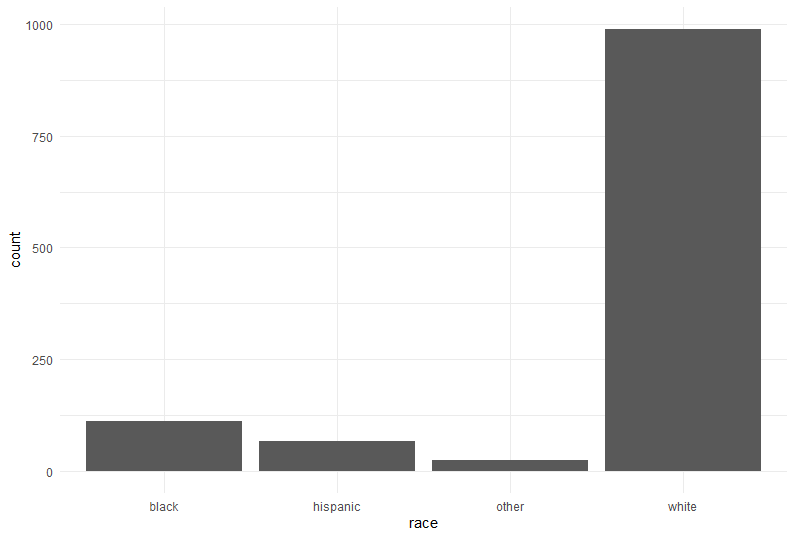
ggplot(heights\_df, aes(sex)) + geom\_bar()



1. **race-bar-chart**

## Using `geom\_bar()` plot a bar chart of the number of records for each race

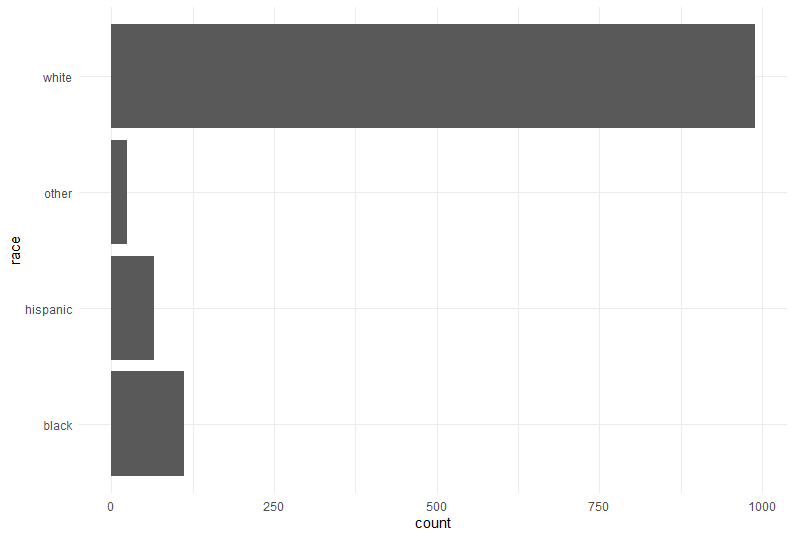
ggplot(heights\_df, aes(race)) + geom\_bar()



1. **Race-horizontal-bar-chart**

## Create a horizontal bar chart by adding `coord\_flip()` to the previous plot

ggplot(heights\_df, aes(race)) + geom\_bar()+ coord\_flip()



1. **Florida-cases**

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

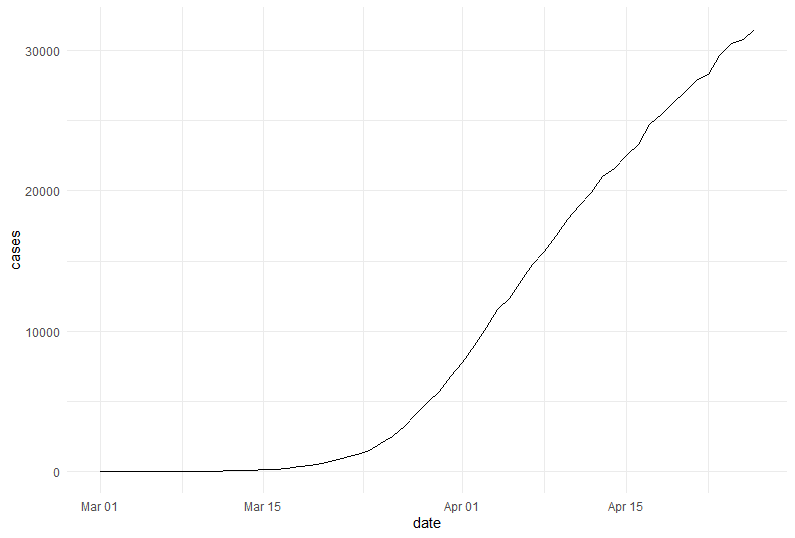
california\_df <- covid\_df[ which( covid\_df$state == "California"), ]

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()



1. **All-cases**

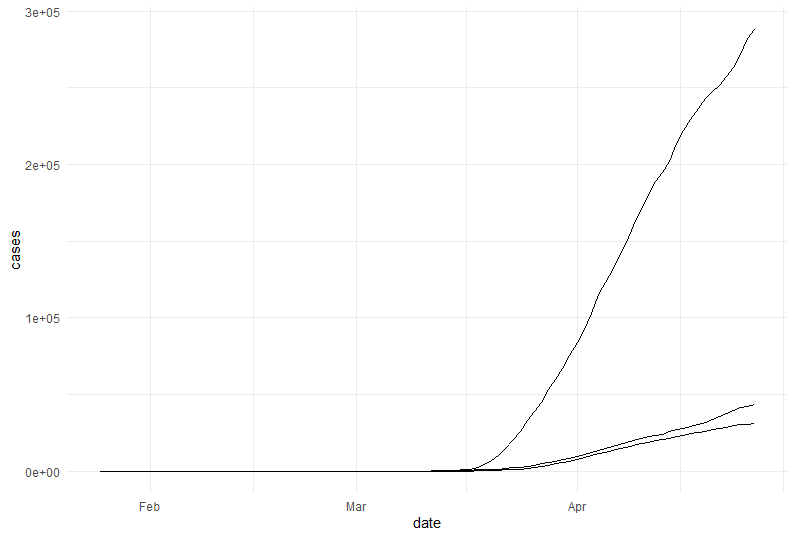
## Add lines for New York and California to the plot

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))



1. **All-cases-color**

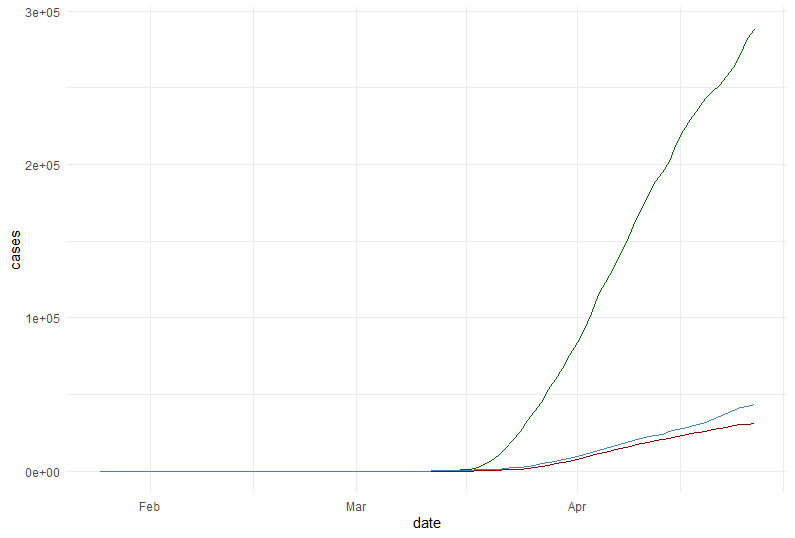
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California

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")



1. **All-cases-legend**

## 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")) +

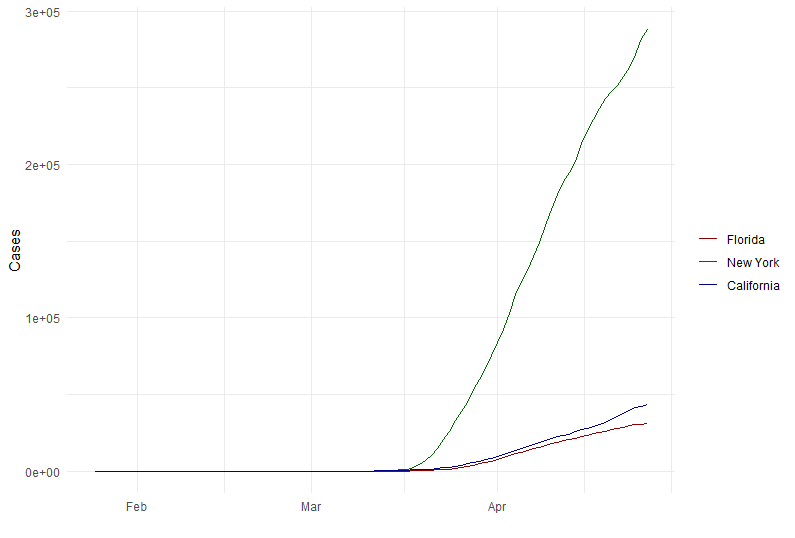
geom\_line(data=california\_df, aes(y = cases, colour="California")) +

scale\_colour\_manual("",

breaks = c("Florida", "New York", "California"),

values = c("darkred","darkgreen", "steelblue" )) +

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



1. **All-cases-log**

## 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()

