## This section is on descriptive statistics and will use the supermarket.xlsx dataset

## Any line with the double hashtag symbol is considered a note inside of R. You can copy this line right into R and it will recognize it as a note and skip to the next line. Great for documenting your work

## This will create a frequency table

table(supermarket$Gender)

## If we wanted to do a two-way table. First Variable will be on the rows and second on the columns

table(supermarket$Gender,supermarket$`Marital Status`)

## This will make a three way table. Var 1 is rows, var2 is columns, and var 3 is separating into different tables

table(supermarket$Gender,supermarket$`Marital Status`,supermarket$`State or Province`)

## This is a three-way table but it is formatted a little nicer

ftable(table(supermarket$Gender,supermarket$`Marital Status`,supermarket$`State or Province`))

## This is a column chart using categorical data. You need to nest a “table()” function inside in order for R to know how many there are in each category. The “main” option is the title of the graph and the “xlab” is the title for the x axis.

barplot(table(supermarket$`State or Province`), main="State Distribution",xlab="State")

## This function gives some basic information about the data. Min, max, median, mean, and quartiles

summary(supermarket$Revenue)

## This is how to install a package. The psych package has some nice descriptive stats

install.packages("psych")

## This is how to load a package once you have installed it

library("psych")

## the describe function will only work if the psych package is loaded. It gives us some good stats about the shape of the distribution

describe(supermarket$Revenue)

## This function will break down the descriptive stats by a category. In this one, by state and province

describeBy(supermarket$Revenue,supermarket$`State or Province`)

## Boxplots give us a visual representation of the shape and spread of the data

boxplot(supermarket$Revenue)

## We can separate boxplots into categories using the tilde ~ and then the category

boxplot(supermarket$Revenue~supermarket$`State or Province`)

## Histograms are really the way to go for visualizing data

hist(supermarket$Revenue)

## Scatter plots are great for comparing two variables

plot(supermarket$`Units Sold`,supermarket$Revenue)

## The correlation gives us the strength of linear relationship between two variables

cor(supermarket$`Units Sold`,supermarket$Revenue)