# Import the WidgeOne dataset.
Widge <- read.csv("C:\\Users\\jpriestl\\Desktop\\WidgeOne.csv")
Widge <- Widge[, -10]

# Using SQL extract only Plant, Productivity, Job Satisfaction and Years On Job.
install.packages("sqldf")
library(sqldf)
Widge2 <- sqldf('select Plant, PRDCTY, JOBSAT, YRONJOB from Widge')

# Rename "Plant" to "Location".
names(Widge2)[names(Widge2) == "Plant"] <- "Location"

# Change the Location names from N and D to "Norcross" and "Dallas".
Widge2$Location <- factor(Widge2$Location, c(" N ", " D "),
c("Norcross", "Dallas"))
table(Widge2$Location)

# Using SQL, determine the average Productivity Level by Location.
MeanProd <- sqldf('select Location, avg(PRDCTY) from Widge2 group by Location')
MeanProd

# Create a Pie Chart of Location. Use Red for Norcross and Blue for Dallas. Include the appropriate labels including percent and count.

colors <- c("Red", "Blue", "Green")
Location <- table(Widge2$Location)
pie(Location)
pie(Location, label = c("Norcross", "Dallas"), col = colors, main = "Dallas versus Norcross")

# Now let's apply the labels
prop.table(Location)
mypercents <- round(prop.table(Location) * 100, 2)
mypercents
mylabels <- paste(names(Location),", n = ", Location,", ", mypercents, ",", sep = " ")
pie(Location, labels = mylabels, col = colors, main = "Dallas versus Norcross")

# Create a cyan-colored Histogram of Productivity. Include appropriate labels/titles.
hist(Widge2$PRDCTY, freq = FALSE, col = "Cyan", main = "Distribution of Productivity", xlab = "Productivity", ylab = "Percentage")

# Create a scatterplot of Productivity and Years on Job. Include appropriate labels/titles. Would you consider these two variables to be related?
plot(Widge2$PRDCTY, Widge2$YRONJOB, main = "Years on Job relative to Productivity", xlab = "Years on Job", ylab = "Productivity", pch = 20)
# Create a new ordinal character variable from Productivity with three levels – high, medium and low. (Use the information from the histogram to determine the cut offs)

Widge2$Prodcat <- ifelse(Widge2$PRDCTY <= 75, "LOW", ifelse(Widge2$PRDCTY <= 90, "MED", "HIGH"))
table(Widge2$Prodcat)

# Create a horizontal bar chart of the new productivity variable. Include appropriate labels and titles.

barplot(table(Widge2$Prodcat), col=colors, main="Plot of Productivity Levels")