



#install.packages('psych')

iris

summary(iris)

library(psych)

describe(iris)

#assign iris dataset to an object, that can be used again later throughout code and can see in data environment as well

blue<-iris

#write the dataset out to excel as a csv

write.csv(blue, "C:/Nipa/stats stuff/DataVis/Homework Assignments/blue\_iris.csv")

#read a csv (in this case the iris dataset we just wrote out) into R

blue<-read.csv("C:/Nipa/stats stuff/DataVis/Homework Assignments/blue\_iris.csv")

#understand the structure of the dataset

str(blue)

#basic stats of dataset using summary() function

summary(blue)

#basic stats of dataset using describe() function from library psych

describe(blue)

#basic stats on each of the species after splitting the data

setosa<- subset(iris, iris$Species == "setosa")

versicolor<- subset(iris, iris$Species == "versicolor")

virginica<- subset(iris, iris$Species == "virginica")

describe(setosa)

describe(versicolor)

describe(virginica)

boxplot(setosa[,1:4], main="Setosa")

boxplot(virginica[,1:4], main="Virginica")

boxplot(versicolor[,1:4], main="Versicolor")

#OR one can use this function below and group by species to get the separate stats on each species

#blue$Species uses the dollar sign to refer to the species column in the blue dataset

describeBy(blue, group=blue$Species)

