Exercise #6

Due 10/12/18

**1. Write R code that replicates the functionality of the head function we used in Unix. Your code should define a variable with the file to return lines from and a variable representing the number of lines to be returned from the top of the indicated file. The selected file content should be printed to the terminal $**

**- use the 'iris.csv' file to complete this**

iris = read.csv("iris.csv") #read in iris.csv data as a variable called iris

iris #view the data in the variable iris

head (iris,10) #take the top 10 rows of the variable iris

**2. Load the data contained in the provided ‘iris.csv’ file and write R code to do the following things**

iris = read.csv("iris.csv") #read in iris.csv data as a variable called iris

iris #view the data in the variable iris

**- print the last 2 rows in the last 2 columns to the R terminal**

last2columns = data.frame(iris$Petal.Width,iris$Species) #create a dataframe from the last two columns species and petal.width from the variable

last2columns #unnecessary, but to check that the data.frame last2columns the last two columns was created

tail(last2columns,2) #select the last 2 rows of the dataframe last2columns

**- get the number of observations for each species included in the data set**

unique(iris$Species) #to see how many different species groupings there are

#for species group setosa

setosa=iris$Species=="setosa" #define variable where species is setosa

setosa #check that the previous function worked

sum(setosa,na.rm = TRUE) #determine how many TRUE values there are that match setosa aka how many setosa there are

#for species group versicolor

versicolor=iris$Species=="versicolor" #define variable where species is versicolor

versicolor #check that previous function worked

sum(versicolor,na.rm = TRUE) #count the number of TRUE variables where species matches versicolor as TRUE

#for species group virginica

virginica=iris$Species=="virginica" #define variable where species is virginica

virginica#check that previous function worked

sum(virginica,na.rm=TRUE) #count number of TRUE for virginica aka the number of virginica

**- get rows with Sepal.Width > 3.5**

##get rows with Sepal.width > 3.5##

sepalover3.5 = iris[which(iris[,2]>3.5),] #create a new object sepalover3.5 that subsets the data from the second column where values are over 3.5

sepalover3.5 #print what rows subsetted from the sepal.width over 3.5

**- write the data for the species setosa to a comma-delimited file names ‘setosa.csv’**

##write the data for the species setosa to a comma-delimited file named 'setosa.csv'##

iris #print the data of the object iris to see which rows contain setosa

setosadata = head(iris,50) #make object with the selected rows 1-50, which contain setosa as the species

setosadata #check that only rows 1-50 with setosa as species are in the object setosadata

write.csv(setosadata, file="setosa.csv") #write a csv of only setosadata called setosa.csv

**- calculate the mean, minimum, and maximum of Petal.Length for observations of virginica**

##calcuate the mean, minimum, and maximum of Petal.Length for observations of virginica##

iris #print the data of the object iris to see which rows contain virginica

virginicadata = tail(iris,50) #create an object where the selected last 50 rows which contain species as virginica

virginicadata #print the object for virginicadata

summary(virginicadata$Petal.Length) #summary statistics for petal length of object virginica data including minimum and maximum and average values