R Day

In RMarkdown files (\*.Rmd), you can add headers using the pound symbol (#). For example:

# This is my header.

## And a smaller header.

Underneath headers, I can describe other things that I am doing.... In R script files (.R), however, # serves to comment out R code that you do not want to execute. Similarly, if you embed your R code in your RMarkdown file, the # will work in the same way as it would in an R script file. For example, I can insert a block of R code and show this here:

# I use the pound symbol in R code to indicate that I don't want this line of code to run  
# 4 + 4  
  
# If I don't use the pound symbol, RMarkdown will run/execute the code.  
8 + 8

## [1] 16

If I need a certain library to execute code, I can load them.

library("psych")

## Warning: package 'psych' was built under R version 3.2.3

library("moments")

I can check to see what my working directory is to make sure it's correct.

getwd() # This should be Psyc109 on your desktop

## [1] "G:/Dropbox/Progs/\_install/R/Exercises"

Because my data file is in my working directory, I can read in the csv file and assign the file to an R data frame objec with a certain name. I can read the data from the Center for Disease Control and assign that data frame to an object named cdc and make sure that it's a data frame.

cdc <- read.csv("cdc.csv")  
is.data.frame(cdc) # Determine if the data is a data frame object. Should return TRUE

## [1] TRUE

I can look a the structure of the data frame so that I can see how many vector/variables there are and what type they are. And I can also take a look at what the data frame looks like by looking at the first or last several cases.

str(cdc) # Let's you know the type of the data (e.g., numeric, integer, Factor, etc.)

## 'data.frame': 20000 obs. of 9 variables:  
## $ genhlth : Factor w/ 5 levels "excellent","fair",..: 3 3 3 3 5 5 5 5 3 3 ...  
## $ exerany : int 0 0 1 1 0 1 1 0 0 1 ...  
## $ hlthplan: int 1 1 1 1 1 1 1 1 1 1 ...  
## $ smoke100: int 0 1 1 0 0 0 0 0 1 0 ...  
## $ height : int 70 64 60 66 61 64 71 67 65 70 ...  
## $ weight : int 175 125 105 132 150 114 194 170 150 180 ...  
## $ wtdesire: int 175 115 105 124 130 114 185 160 130 170 ...  
## $ age : int 77 33 49 42 55 55 31 45 27 44 ...  
## $ gender : Factor w/ 2 levels "f","m": 2 1 1 1 1 1 2 2 1 2 ...

head(cdc) # Shows the first few cases of your data set

## genhlth exerany hlthplan smoke100 height weight wtdesire age gender  
## 1 good 0 1 0 70 175 175 77 m  
## 2 good 0 1 1 64 125 115 33 f  
## 3 good 1 1 1 60 105 105 49 f  
## 4 good 1 1 0 66 132 124 42 f  
## 5 very good 0 1 0 61 150 130 55 f  
## 6 very good 1 1 0 64 114 114 55 f

tail(cdc) # Shows the last few cases of your data set

## genhlth exerany hlthplan smoke100 height weight wtdesire age  
## 19995 good 0 1 1 69 224 224 73  
## 19996 good 1 1 0 66 215 140 23  
## 19997 excellent 0 1 0 73 200 185 35  
## 19998 poor 0 1 0 65 216 150 57  
## 19999 good 1 1 0 67 165 165 81  
## 20000 good 1 1 1 69 170 165 83  
## gender  
## 19995 m  
## 19996 f  
## 19997 m  
## 19998 f  
## 19999 f  
## 20000 m

dim(cdc) # Shows the dimensions of the data frame; 20,000 observations; 9 columns

## [1] 20000 9

names(cdc) # What are the column/variable names in the data frame?

## [1] "genhlth" "exerany" "hlthplan" "smoke100" "height" "weight"   
## [7] "wtdesire" "age" "gender"

I can also View the data frame in a more user-friendly viewer as well as examing the descriptive statistics of certain variables by specifying the data frame object (e.g., cdc) and the vector/variable I wish to examine.

View(cdc) # View the data file  
mean(cdc$weight) #Shows the Mean of the variable weight

## [1] 169.683

median(cdc$weight) #Shows the Median of the variable weight

## [1] 165

var(cdc$weight) #Shows the variance of the variable weight

## [1] 1606.484

sd(cdc$weight) #Shows the Standard Deviation of the variable weight

## [1] 40.08097

summary(cdc$weight) #Provides a summary of the variable weight

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 68.0 140.0 165.0 169.7 190.0 500.0

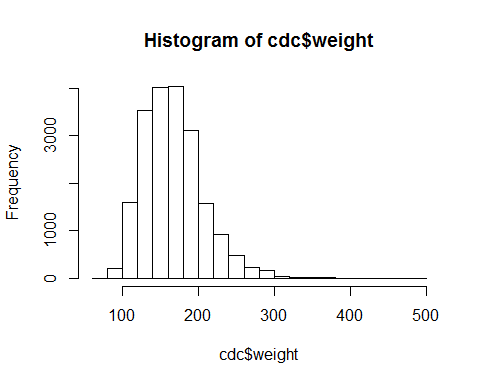
skewness(cdc$weight) #Shows the Skewness of the variable weight

## [1] 0.9556563

kurtosis(cdc$weight) #Shows the Kurtosis of the variable weight

## [1] 4.995356

hist(cdc$weight) #Produces a simple histogram of the variable weight



Once I am done with my coding, I can save my RMarkdown file (*.Rmd) and click the Knit HTML button to produce a pretty HTML file. This file will also be saved in my working directory (Psyc109) and will have the same name as my* .Rmd file.