Intro to Dat Science - HW 2

```
# Enter your name here: Victoria Haley
```


Attribution statement: (choose only one and delete the rest)

```
# 2. I did this homework with help from the book and the professor and these Internet sources: statisti
```

Reminders of things to practice from last week:

Assignment arrow <- The combine command c( ) Descriptive statistics mean( ) sum( ) max( ) Arithmetic operators + - * / Boolean operators > < >= <= == !=

This Week: Explore the **quakes** dataset (which is included in R). Copy the **quakes** dataset into a new dataframe (call it **myQuakes**), so that if you need to start over, you can do so easily (by copying quakes into myQuakes again). Summarize the variables in **myQuakes**. Also explore the structure of the dataframe

```
myQuakes <- quakes
```

Step 1: Explore the earthquake magnitude variable called **mag**

A. What is the average magnitude? Use mean() or summary():

```
mean(myQuakes$mag)
```

```
## [1] 4.6204
```

B. What is the magnitude of the largest earthquake? Use max() or summary() and save the result in a variable called **maxQuake**:

```
max(myQuakes$mag)
```

```
## [1] 6.4
```

```
maxQuake <- max(myQuakes$mag)
```

C. What is the magnitude of the smallest earthquake? Use min() or summary() and save the result in a variable called **minQuake**:

```
min(myQuakes$mag)
```

```
## [1] 4
```

```
minQuake <- min(myQuakes$mag)
```

D. Output the **third row** of the dataframe

```
myQuakes[3, ]
```

```
##   lat  long depth mag stations
## 3 -26 184.1    42 5.4       43
```

E. Create a new dataframe, with only the rows where the **magnitude is greater than 4**. How many rows are in that dataframe (use code, do not count by looking at the output)

```
bigQuakes <- myQuakes[myQuakes$mag > 4,]
nrow(bigQuakes)
```

```
## [1] 954
```

F. Create a **sorted dataframe** based on magnitude and store it in **quakeSorted1**. Do the sort two different ways, once with arrange() and then with order()

Rstudio would not let me knit to PDF without including the "library("dplyr") string

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
quakeSorted1 <- arrange(myQuakes, mag)
quakeSorted1 <- myQuakes[ order(myQuakes$mag),]
```

G. What are the latitude and longitude of the quake reported by the largest number of stations?

```
which.max(myQuakes$stations)
```

```
## [1] 870
```

```
myQuakes[870,c(1,2)]
```

```
##        lat   long
## 870 -12.23 167.02
```

H. What are the latitude and longitude of the quake reported by the smallest number of stations?

```
which.min(myQuakes$stations)
```

```
## [1] 14
```

```
myQuakes[14,c(1,2)]
```

```
##    lat   long
## 14 -21 181.66
```

Step 3: Using conditional if statements

I. Test if **maxQuake** is greater than 7 (output "yes" or "no") **Hint:** Try modifying the following code in R:

```
if  (maxQuake > 7) "yes" else "no"
```

```
## [1] "no"
```

J. Following the same logic, test if **minQuake** is less than 3 (output "yes" or "no"):

```
if(minQuake < 3) "yes" else "no"
```

```
## [1] "no"
```