Intro to Dat Science - HW 2


```
# Enter your name here:   Ryan Tervo
# Course Number:          IST 687
# Assignement Name:       Homework #2
# Due Date:               24 Oct 2022
# Submitted Date:         21 Oct 2022
```

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

```
# 1. I did this homework by myself, with help from the book and the professor.
```

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

'Summary:'
```

```
## [1] "Summary:"
```

```
summary(myQuakes)
```

```
##       lat              long           depth            mag       
##  Min.   :-38.59   Min.   :165.7   Min.   : 40.0   Min.   :4.00  
##  1st Qu.:-23.47   1st Qu.:179.6   1st Qu.: 99.0   1st Qu.:4.30  
##  Median :-20.30   Median :181.4   Median :247.0   Median :4.60  
##  Mean   :-20.64   Mean   :179.5   Mean   :311.4   Mean   :4.62  
##  3rd Qu.:-17.64   3rd Qu.:183.2   3rd Qu.:543.0   3rd Qu.:4.90  
##  Max.   :-10.72   Max.   :188.1   Max.   :680.0   Max.   :6.40  
##     stations    
##  Min.   : 10.00  
##  1st Qu.: 18.00  
##  Median : 27.00  
##  Mean   : 33.42  
##  3rd Qu.: 42.00  
##  Max.   :132.00
```

```
'Structure:'
```

```
## [1] "Structure:"
```

```
str(myQuakes)
```

```
## 'data.frame':    1000 obs. of  5 variables:
##  $ lat     : num  -20.4 -20.6 -26 -18 -20.4 ...
##  $ long    : num  182 181 184 182 182 ...
##  $ depth   : int  562 650 42 626 649 195 82 194 211 622 ...
##  $ mag     : num  4.8 4.2 5.4 4.1 4 4 4.8 4.4 4.7 4.3 ...
##  $ stations: int  41 15 43 19 11 12 43 15 35 19 ...
```

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

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

```
#   Using mean() Function:
mag <- myQuakes$mag
meanQuakes1 <- mean(mag)
meanQuakes1
```

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

```
#   Using summary() function:
meanQuakes2 <- summary(mag)[4]
meanQuakes2
```

```
##   Mean
## 4.6204
```

```
#   Verify both methods produce the same result
meanQuakes1 == meanQuakes2
```

```
## Mean
## TRUE
```

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

```
#   Using max() function:
maxQuake <- max(mag)
maxQuake
```

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

```
#   Using summary() function:
maxQuake1 <- summary(mag)[6]
maxQuake1
```

```
## Max.
##  6.4
```

```
#   Verify that both methods produce the same result
maxQuake == maxQuake1
```

```
## Max.
## TRUE
```

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

```
#   Using min() function
minQuake <- min(mag)
minQuake
```

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

```
#   Using summary() function
summary(mag)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.00    4.30    4.60    4.62    4.90    6.40
```

```
minQuake1 <- summary(mag)[1]
minQuake1
```

```
## Min.
##    4
```

```
#   Verify that both methods produce the same result
minQuake == minQuake1
```

```
## Min.
## TRUE
```

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)

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

```
## [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()

```
#  Using arrange() function
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
##  ggplot2 3.3.6      purrr   0.3.5
##  tibble  3.1.8      dplyr   1.0.10
##  tidyr   1.2.1      stringr 1.4.1
##  readr   2.1.3      forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
##  dplyr::filter() masks stats::filter()
##  dplyr::lag()    masks stats::lag()
```

```
    #    This library assumes that it's been installed already.
    #    If tidyverse is not installed then this code will not run properly

quakeSorted1 <- arrange(myQuakes, mag)
head(quakeSorted1, 10)
```

```
##       lat   long depth mag stations
## 1  -20.42 181.96   649   4       11
## 2  -19.68 184.31   195   4       12
## 3  -17.94 181.49   537   4       15
## 4  -23.55 180.80   349   4       10
## 5  -19.26 184.42   223   4       15
## 6  -22.06 180.60   584   4       11
## 7  -15.31 185.80   152   4       11
## 8  -17.70 181.70   450   4       11
## 9  -19.73 182.40   375   4       18
## 10 -19.06 182.45   477   4       16
```

```
#   Using order() function with subsetting
quakeSorted2 <- myQuakes[order(myQuakes$mag), ]
head(quakeSorted2, 10)
```

```
##       lat   long depth mag stations
```

```
## 5   -20.42 181.96   649   4       11
## 6   -19.68 184.31   195   4       12
## 26  -17.94 181.49   537   4       15
## 34  -23.55 180.80   349   4       10
## 52  -19.26 184.42   223   4       15
## 58  -22.06 180.60   584   4       11
## 71  -15.31 185.80   152   4       11
## 85  -17.70 181.70   450   4       11
## 96  -19.73 182.40   375   4       18
## 113 -19.06 182.45   477   4       16
```

```
#   Verify both sorting techniques produced the same result.
cat(sum(quakeSorted1 == quakeSorted2), 'elements of quakeSorted1 match', nrow(quakeSorted1) *
ncol(quakeSorted1), 'elements in quakeSorted2.  The data frames are the same.')
```

```
## 5000 elements of quakeSorted1 match 5000 elements in quakeSorted2.  The data frames are the
 same.
```

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

```
tempMyQuakes = myQuakes[myQuakes$stations == max(myQuakes$stations), c('lat', 'long', 'station
s') ]
'The following quake(s) were reported by the largest number of stations.'
```

```
## [1] "The following quake(s) were reported by the largest number of stations."
```

```
tempMyQuakes
```

```
##        lat   long stations
## 870 -12.23 167.02      132
```

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

```
#   Verify min
tempMyQuakes = myQuakes[myQuakes$stations == min(myQuakes$stations), c('lat', 'long', 'station
s') ]
'The following quake(s) were reported by the smallest number of stations.'
```

```
## [1] "The following quake(s) were reported by the smallest number of stations."
```

```
tempMyQuakes
```

```
##        lat   long stations
## 14  -21.00 181.66       10
## 34  -23.55 180.80       10
## 35  -16.30 186.00       10
## 146 -20.10 184.40       10
## 175 -15.03 182.29       10
```

```
## 263 -19.06 169.01       10
## 284 -17.70 185.00       10
## 327 -21.04 181.20       10
## 350 -27.21 182.43       10
## 431 -18.40 183.40       10
## 438 -20.30 182.30       10
## 482 -14.85 184.87       10
## 690 -17.60 181.50       10
## 693 -20.61 182.44       10
## 704 -25.00 180.00       10
## 763 -17.78 185.33       10
## 770 -20.70 186.30       10
## 776 -21.77 181.00       10
## 778 -21.05 180.90       10
## 995 -17.70 188.10       10
```

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:

```
#   Original code
#if  (100 < 150) "100 is less than 150" else "100 is greater than 150"

#   Modified code
if  (maxQuake > 7) "yes" else "no"
```

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

```
#   Verify maxQuake value is not greater than 7.
cat('max quake was ', maxQuake, ' which is less than 7.')
```

```
## max quake was  6.4  which is less than 7.
```

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

```
#   Original Code
#if  (100 < 150) "100 is less than 150" else "100 is greater than 150"

#   Modified code
if  (minQuake < 3) "yes" else "no"
```

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

```
#   Verify minQuake is not less than 3.
cat('min quake was ', minQuake, ' which is greater that 3.')
```

```
## min quake was  4  which is greater that 3.
```