

Assignment_7

2024-02-21

Question 1

```
quest_1 = c(12,10,3,4.2,18,1,54,21,6,7)
```

```
mean(quest_1)
```

```
## [1] 13.62
```

```
median(quest_1)
```

```
## [1] 8.5
```

Question 2

```
quest_2= c(19,5,-3,41,18,11,50,25,-8,5,10,13)
```

```
quest_2[quest_2>0] # to remove any values that are less than 0 (negative values)
```

```
## [1] 19 5 41 18 11 50 25 5 10 13
```

```
mean(quest_2)
```

```
## [1] 15.5
```

```
median(quest_2)
```

```
## [1] 12
```

Question 3

```
quest_3 = c(19,5,NA,41,18,11,50,25,-8,5,10,13)
```

```
mean(quest_2, na.rm=TRUE) ## na.rm=TRUE means to remove any existing NA values
```

```
## [1] 15.5
```

```
median(quest_2, na.rm=TRUE)
```

```
## [1] 12
```

Question 4

```
#to access the csv data file, the working directory needs to be set up
```

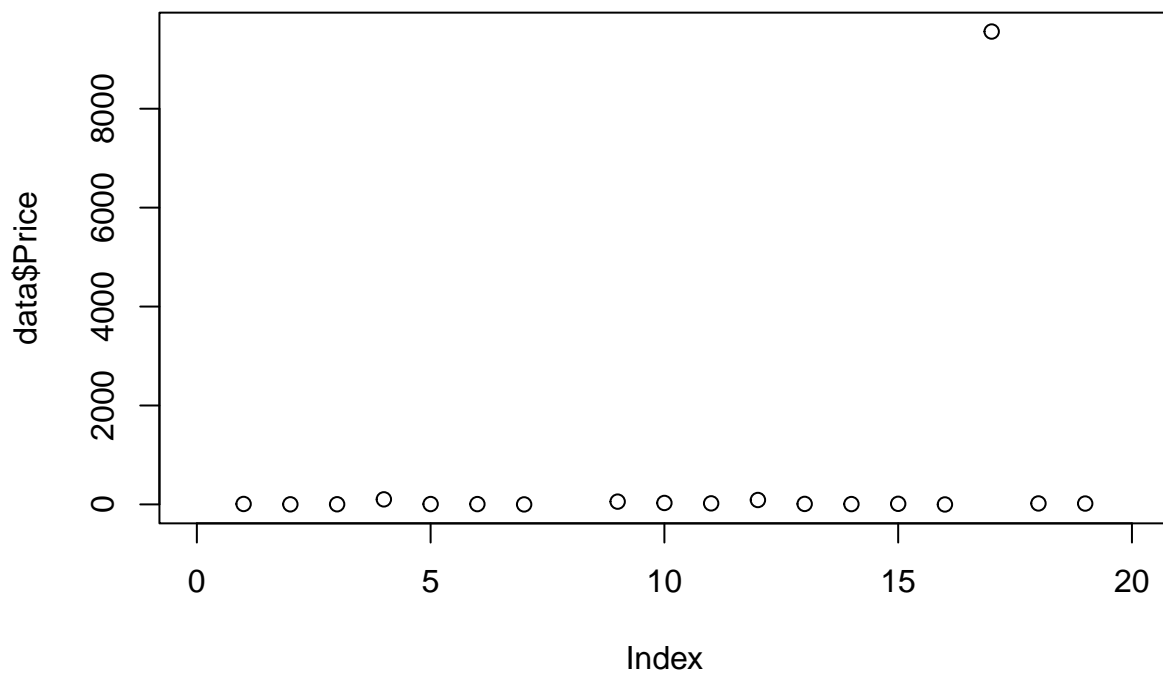
```
setwd("C://Users//amann//Desktop//Bellevue College//Winter 2024//DATA 310//Week 7")  
  
data = read.csv("BookPrice.csv")
```

General analysis on Q4:

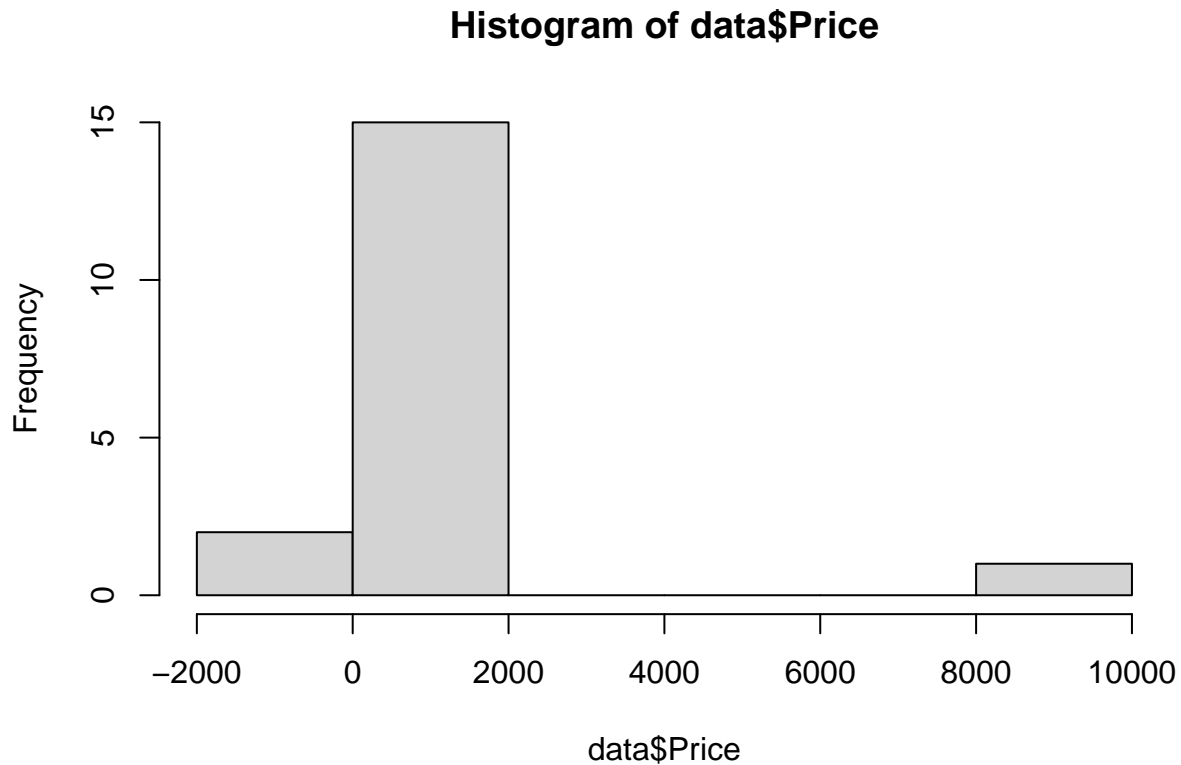
```
summary(data) #to get the summary of the data file
```

```
##      Price  
## Min.   : -0.800  
## 1st Qu.:  6.125  
## Median : 10.940  
## Mean   : 552.671  
## 3rd Qu.: 27.000  
## Max.   :9560.000  
## NA's   :1
```

```
plot(data$Price, xlim=c(0,20)) #xlim sets the limit for x axis
```



```
hist(data$Price)
```



My analysis:

From the summary function of the data, I can identify the important data points like the minimum, maximum number of the prices, what the average price of the bookstore books is and if any books have a missing price. By using the plot and summary function, I can see where most of my numbers are, which are below \$2000 dollars and even lower since they are very close to the number 0. The axis cannot be narrowed down without removing the outlier of a value greater than \$8000. The outlier of \$8000 is not normal for book prices unless it is very rare and is worth that price, which means it should be investigated. It could be a possible indication of the wrong book price. Also, by using the histogram and summary function, the data indicates that there's also a negative value that isn't correct since a book price cannot be negative. The negative value should be investigated.