BACS HW (Week 1)

108020024

due on 02/19 (Sun)

Before working on HW1, first complete the tutorial of Swirl assigned this week to get familial with r programming language.

1.What is the 5th element in the original list of ages?

For problem 1, first read the text file into a data frame named "customer".

```
customers <- read.table("customers.txt",head=TRUE) #read the target txt file
customers[5,] #print the 5th element
```

```
## [1] 45
```

Since the data frame has only one variable "age", just print the fifth row of customer,and the 5th element in the original list of ages is 45.

2. What is the fifth lowest age?

For problem 2, sort the data into order.

```
customers_sorted <- sort(customers$age) #sort the age list from smallest to largest
customers_sorted[5] #get the fifth lowest age
```

```
## [1] 19
```

So 19 is the fifth lowest age.

3. Extract the five lowest ages together.

```
customers_sorted[1:5] #get the 1 to 5 elements.
```

```
## [1] 18 19 19 19 19
```

The five lowest ages are [18 19 19 19 19].

4. Get the five highest ages by first sorting them in decreasing order first.

```r
customers_sorted2 <- sort(customers$age,decreasing = TRUE) #sort from largest to smallest
customers_sorted2[1:5] #get the 1 to 5 elements.
```

```
## [1] 85 83 82 82 81
```

The five highest ages are [85 83 82 82 81].

5. What is the average (mean) age?

```r
age_mean <- mean(customers$age) #calculate the mean of age
age_mean
```

```
## [1] 46.80702
```

The average age is 46.80702.

6. What is the standard deviation of ages?

```r
age_sd <- sd(customers$age) #calculate the standard deviation of age
age_sd
```

```
## [1] 16.3698
```

The standard deviation of ages is 16.3698.

7. Make a new variable called age_diff, with the difference between each age and the mean age.

```r
age_diff <- c(customers$age - age_mean)
```

8. What is the average "difference between each age and the mean age"?

```r
mean(age_diff) #calculate the mean of age_diff
```

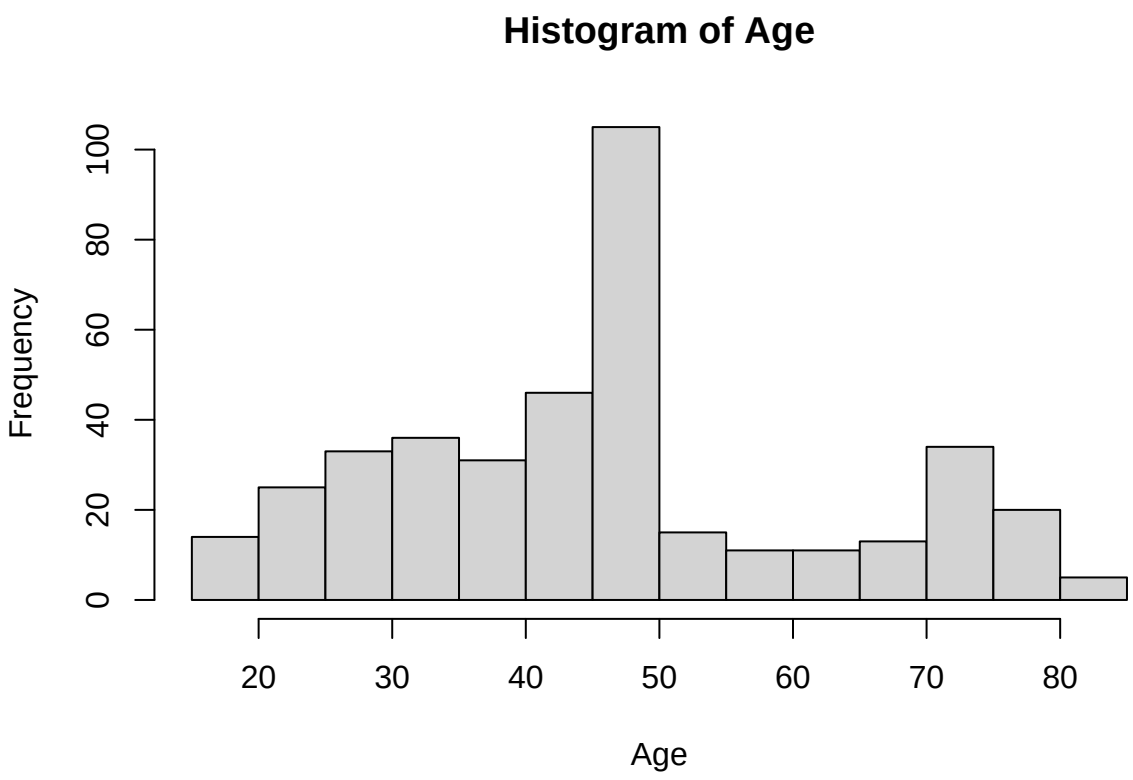
```
## [1] -1.623275e-15
```

The average "difference between each age and the mean age" is -1.623275e-15, however the computer is trying to say 0. The really small number is because of the imprecision that computers makes doing calculation.

9. Visualize the raw data as we did in class: (a) histogram, (b) density plot, (c) boxplot+stripchart

(a) histogram

```r
hist(customers$age, main = "Histogram of Age", xlab = "Age")
```

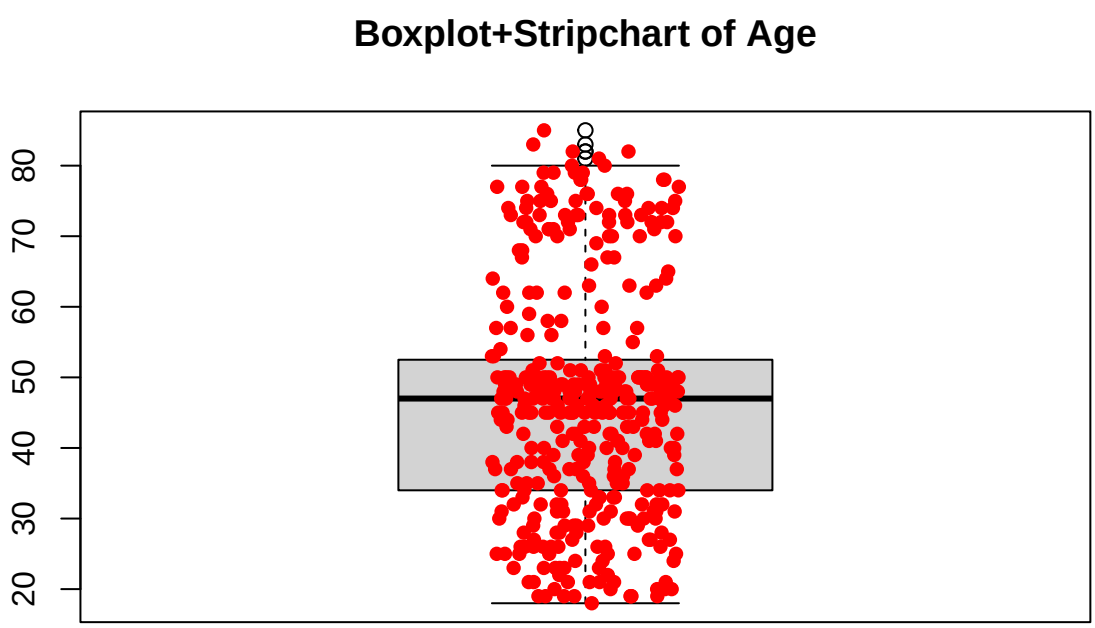


(b) density plot

```r
plot(density(customers$age), main = "Histogram of Age", xlab = "Age")
```

(c) boxplot+stripchart

```
boxplot(customers$age, main = "Boxplot+Stripchart of Age")
stripchart(customers$age, add = TRUE, vertical = TRUE,
           method = "jitter", col = "red", pch = 16)
```


Boxplot+Stripchart of Age
20
30
40
50
60
70
80