Homework 1 BACS

109006206

2023-02-19

Opening Libraries

```
require(dplyr)
```

Set Working Directories

```
setwd("/Users/olivia/Documents/Documents/Study/Semester 6/BACS")
```

Import the Data

```
data<-read.table(file="customers.txt",header=TRUE)
```

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

```
data[5,]
```

To get the 5th element we need to access the 5th row of the data so here i use data[row,column]

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

2. What is the fifth lowest age?

```
temp<-unique(arrange(data,age))
temp[5,]
```

I use unique() to sort out duplicate ages and arrange() to arrange the data based on ascending order

```
## [1] 22
```

3. Extract the five lowest ages together

```
temp[1:5,]
```

I use data[from:until,] to print out from a certain row number to a certain row number.

```
## [1] 18 19 20 21 22
```

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

```
temp2<-sort(unique(data$age),decreasing = TRUE)
temp2[1:5]
```

Here I use sort and set the decreasing into TRUE so it returns. I also unique the data to remove the duplicate ages.

```
## [1] 85 83 82 81 80
```

5. What is the average (mean) age?

```
mn<-mean(data$age)
mn
```

I use mean() to get the average age.

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

6. What is the standard deviation of ages?

```
sd(data$age)
```

I use sd() that's already provided by r to find the standard deviation.

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

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

```
data$age_diff<-abs(data$age-mn)
head(data,5)
```

Here i use abs() so all the result will be positive since it would be make more sense to find the difference. To calculate the difference I just calculate it by age-average_age

```
##   age  age_diff
## 1  49  2.192982
## 2  69 22.192982
## 3  41  5.807018
## 4  73 26.192982
## 5  45  1.807018
```

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

```
av<-mean(data$age_diff)
av
```

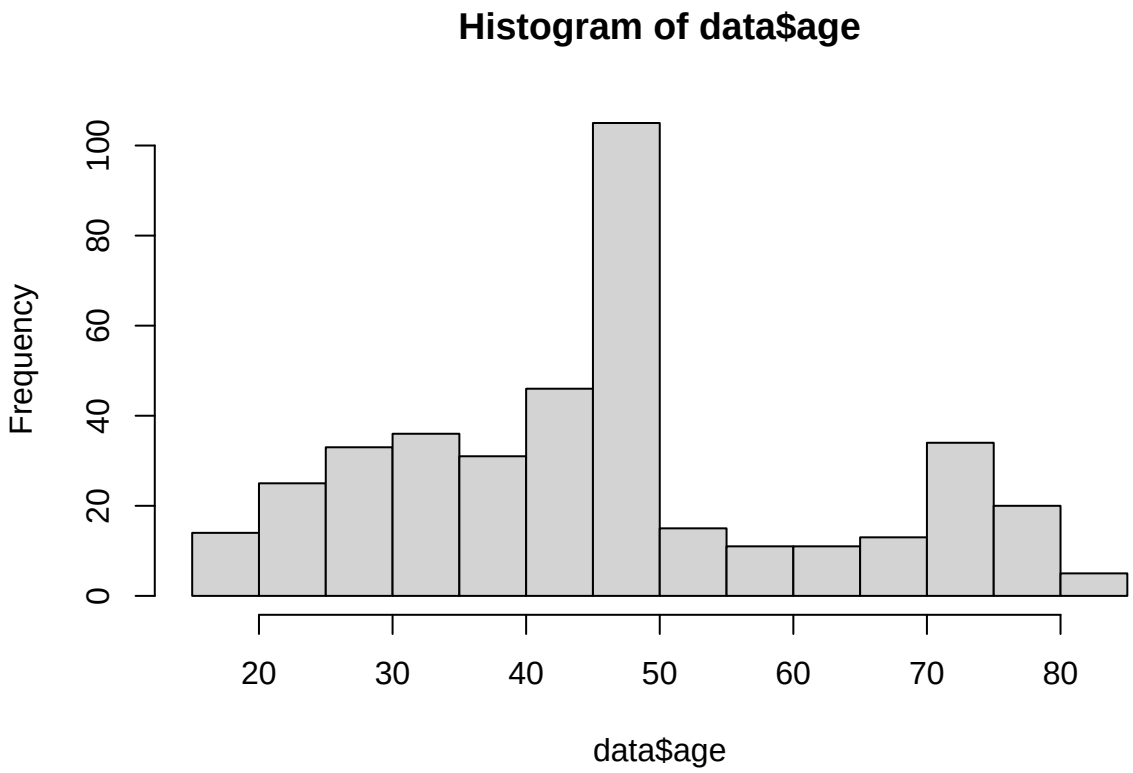
I use mean() to find the average

```
## [1] 12.66948
```

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

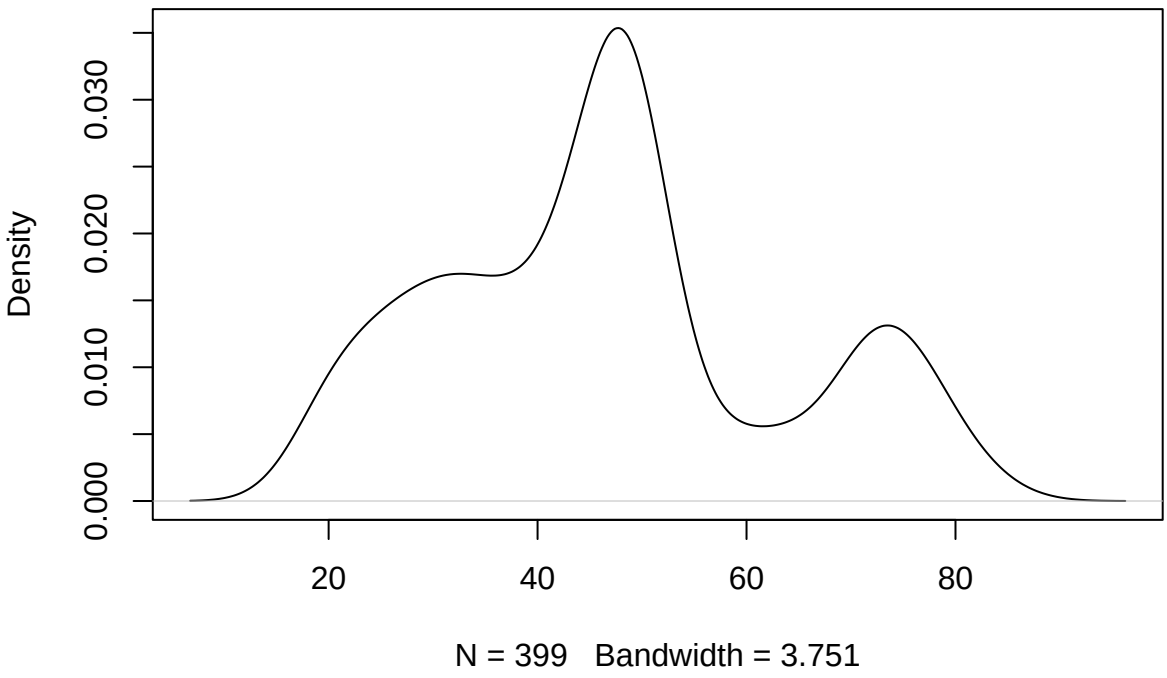
Histogram is used to summarize discrete or continuous data that are measured on an interval scale.

```
hist(data$age)
```




```
d<-density(data$age)
plot(d,main="Density of Age")
```

Density plots are used to observe the distribution of a variable in a dataset.




```
boxplot(data$age, horizontal = TRUE)
stripchart(data$age, method = "stack", add = TRUE)
```

Box plots are used to show distributions and comparison between multiple groups. A strip chart can be used to visualize dozens of time series at once.