

HW1

111078513

2023-02-17

Load customers data

```
data <- read.table("customers.txt", header = T)
age <- data$age
head(age)
```

```
## [1] 49 69 41 73 45 71
```

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

```
q1 <- age[5]
q1
```

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

2. What is the fifth lowest age?

remove the duplicated data from 'age' and get the fifth lowest

```
sort_age <- sort(age)
uni_sort_age <- unique(sort_age)
uni_sort_age[5]
```

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

3. Extract the five lowest ages together

```
head(uni_sort_age, 5)
```

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

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

```
tail(uni_sort_age, 5)
```

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

5. What is the average (mean) age?

```
mean(age)
```

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

6. What is the standard deviation of ages?

```
sd(age)
```

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

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

```
age_diff <- age - mean(age)
head(age_diff)
```

```
## [1] 2.192982 22.192982 -5.807018 26.192982 -1.807018 24.192982
```

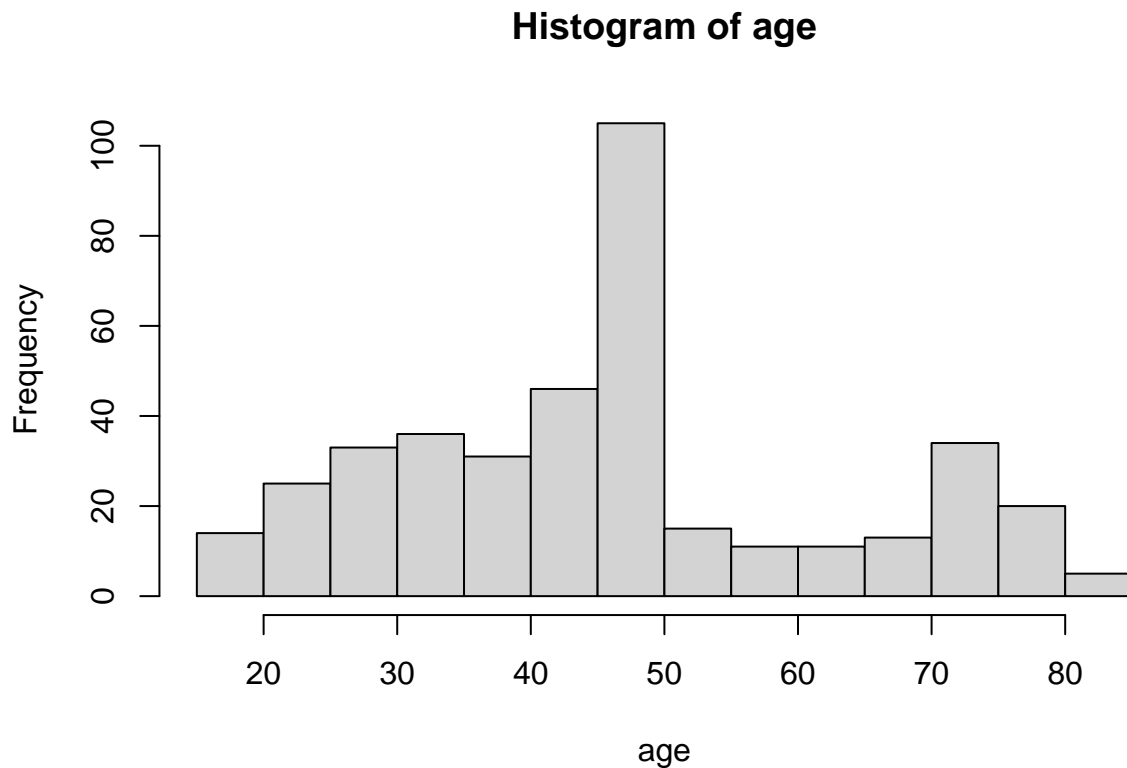
8. What is the average `age_diff`?

```
mean(age_diff)
```

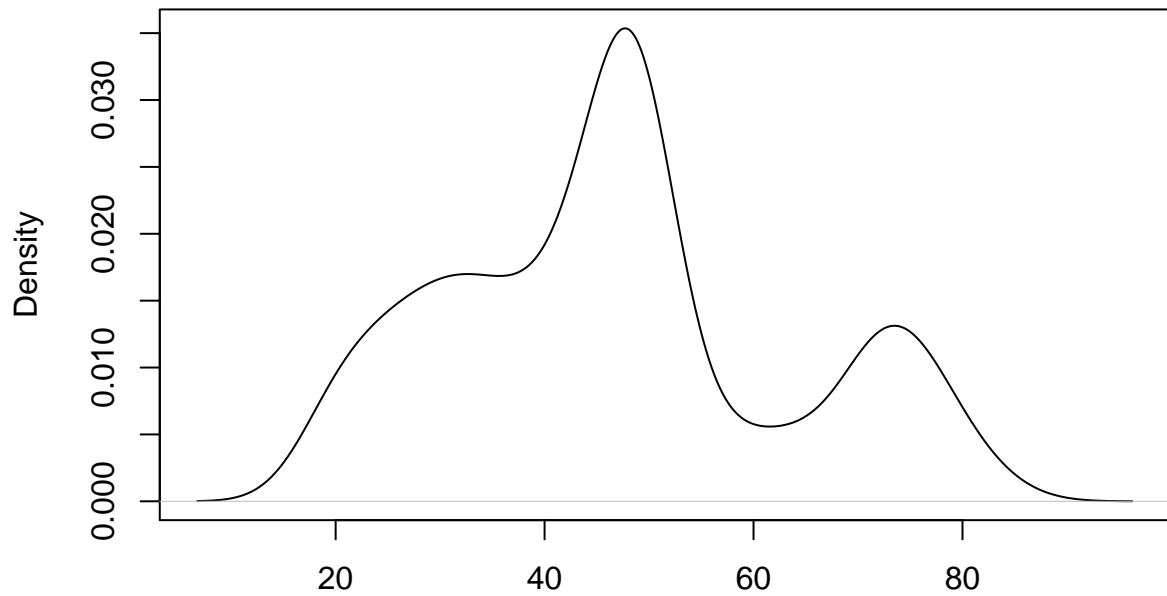
```
## [1] -1.623275e-15
```

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

```
# (a) histogram
hist(age)
```



```
# (b) density plot
plot(density(age))
```

density.default(x = age)

N = 399 Bandwidth = 3.751

```
# (c) boxplot+stripchart
```

```
boxplot(age, horizontal = TRUE)
```

```
stripchart(age, method = "stack", add = TRUE)
```

