Week 03 Assignment

1717719

2019-09-22

Let us produce numerical summaries of the variables AGE, HEIGHT and WEIGHT in the GLOW500 data. First, make sure to set your working directory.

```
setwd("D:/git/DPH101-xjtlu/Y3/week03_ass_9.23")
```

Then, read the file into memory.

```
GLOW500_WORK <- read.csv("GLOW500.csv")
```

We will calculate the mean and variance for the three variables. The formula for the mean is $\bar{x} = \frac{\sum x}{n}$.

The formula for the variance is $\sigma^2 = \frac{\sum (x-\bar{x})^2}{n-1}$.

The mean and variance of the three variables are calculated using base R functions. Let's demonstrate using AGE.

```
mean(GLOW500_WORK$AGE); var(GLOW500_WORK$AGE)
```

```
## [1] 68.562
```

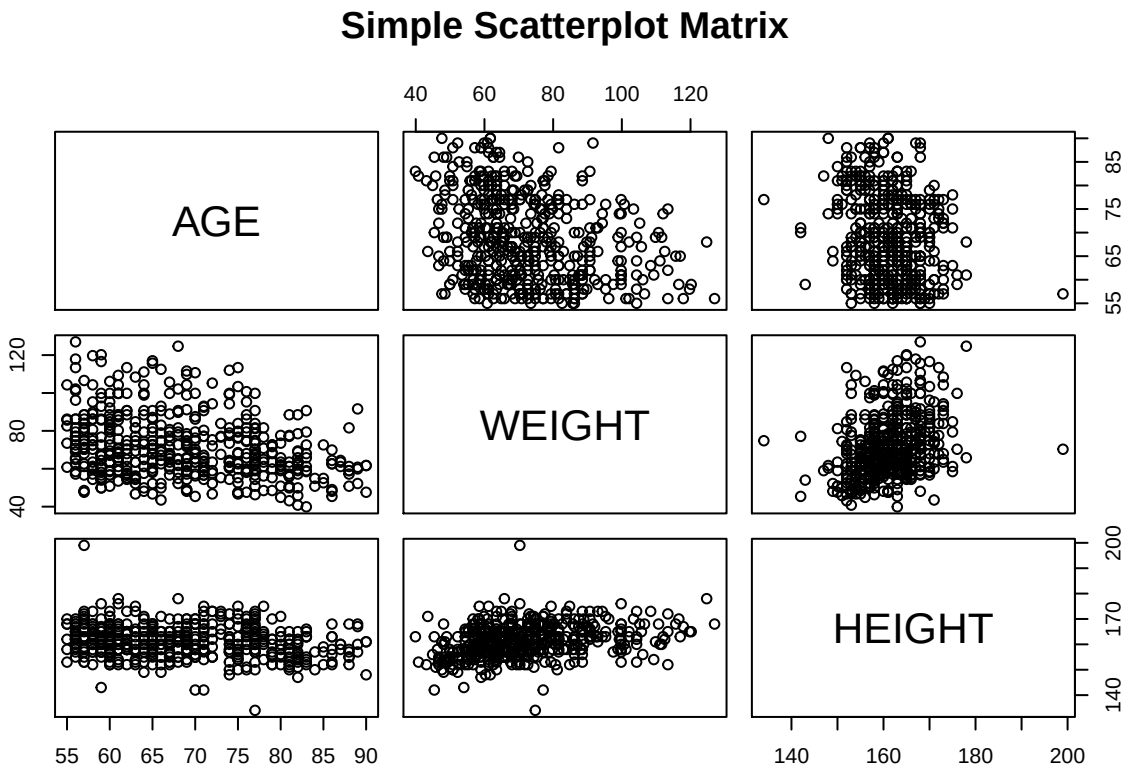
```
## [1] 80.81178
```

Table 1: Basic numerical summaries of age, weight and height in the GLOW500 study.

Variable	Mean	Variance
Age(Year)	68.6	80.81
Weight(kg)	71.82	270.142
Height(cm)	161.4	40.39

Finally, let's visualise the relationship among all three variables using a scatterplot matrix.

```
pairs(~AGE + WEIGHT + HEIGHT, data=GLOW500_WORK, main="Simple Scatterplot Matrix")
```


Simple Scatterplot Matrix
AGE
WEIGHT
HEIGHT
40
60
80
100
120
55
65
75
85
40
80
120
140
170
200
55
60
65
70
75
80
85
90
140
160
180
200

THE END