

Exercise 2

Tobias Famos

Reading the Data

```
educationDataSet <- read.table("Education.txt", header=T)
```

Exercise 1: Plotting

Yes you can change all the Labels on a graph with plot by adding main, ylab and xlab parameters.

```
plot(educationDataSet$Education, educationDataSet$Wage, xlab="Education in Years", ylab="Wage per Year"
```



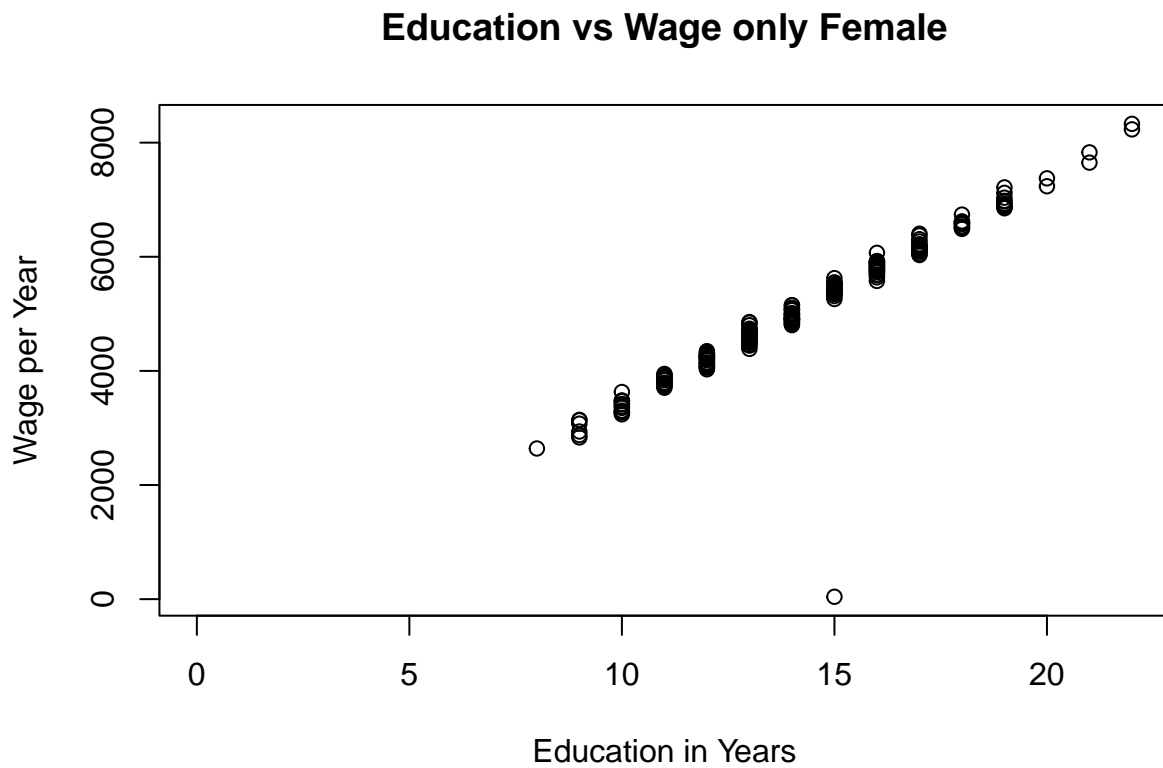
##

Filtering by Gender First create the subsets

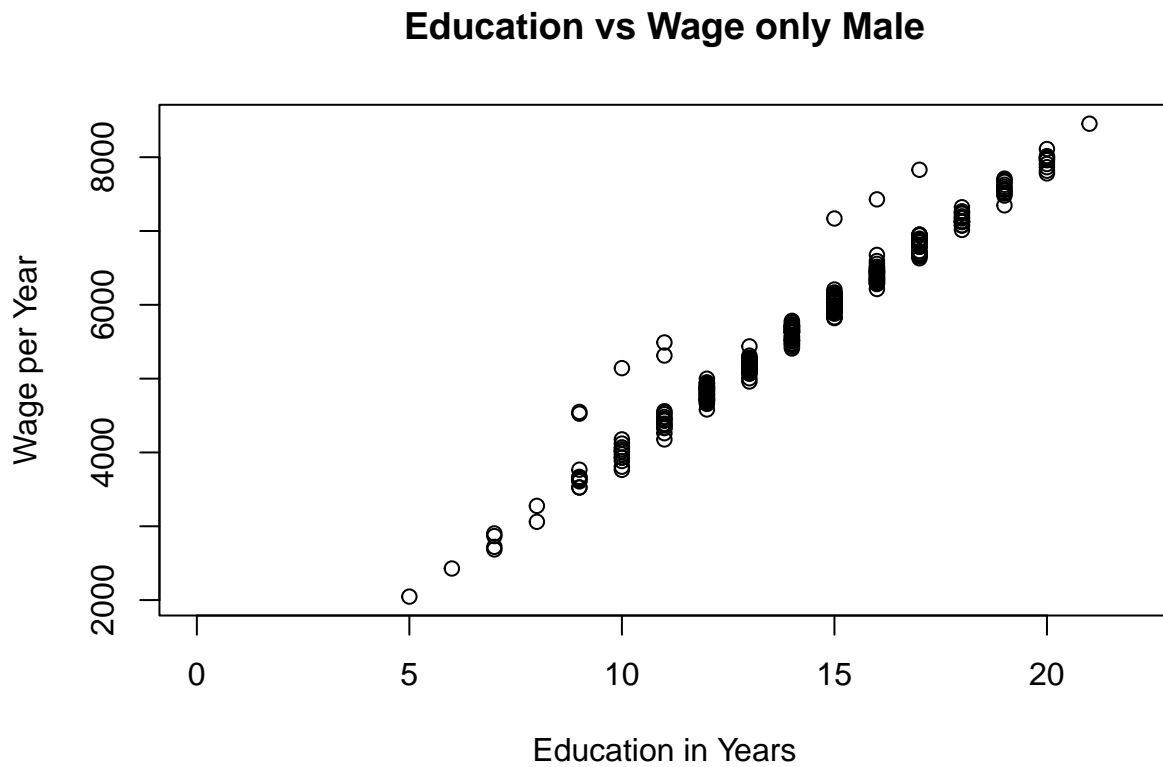
```
males <- subset(educationDataSet, Gender == 1)  
females <- subset(educationDataSet, Gender == 2)
```

Now plot the females and the males

```
plot(females$Education, females$Wage, xlab="Education in Years", ylab="Wage per Year", main="Education vs Wage only Female")
```



```
plot(males$Education, males$Wage, xlab="Education in Years", ylab="Wage per Year", main="Education vs Wage only Male")
```



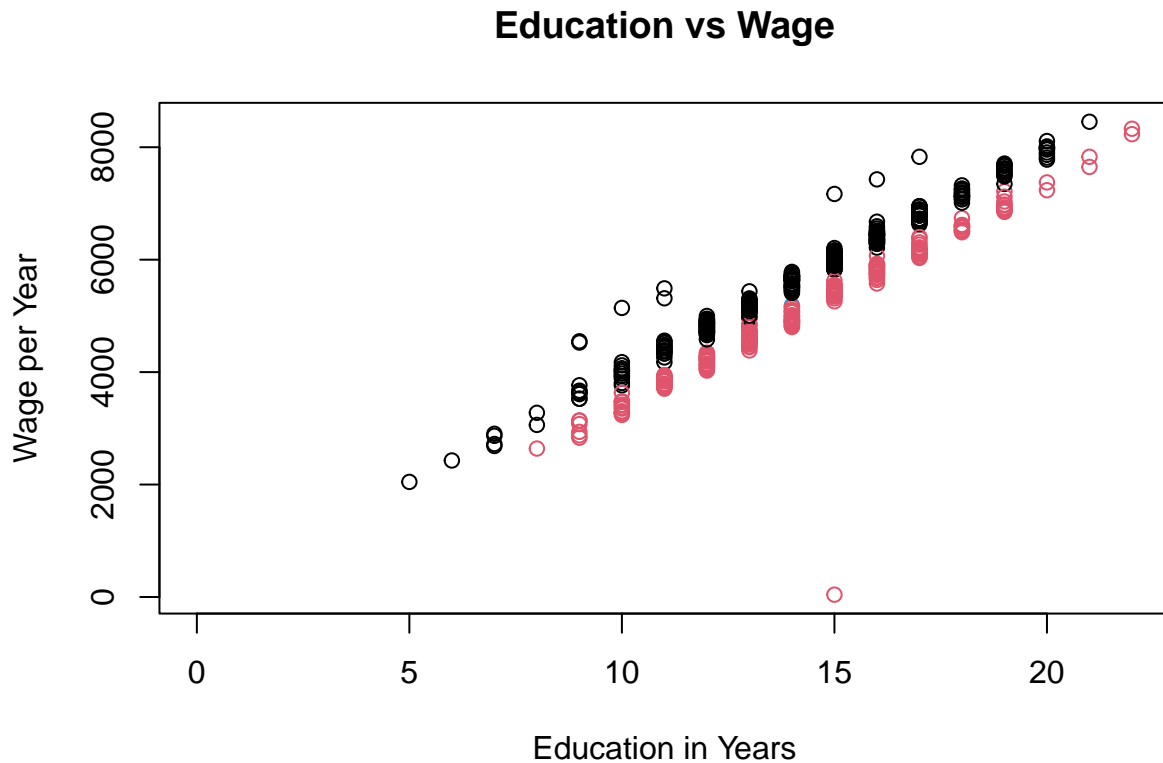
Plot them both with different colors.

Also

```

colors <- educationDataSet$Gender
plot(educationDataSet$Education, educationDataSet$Wage, xlab="Education in Years", ylab="Wage per Year"

```



##

Simpson's Hospital

```

hospitalDataSet <- read.table("Simpson.txt", header=T)
hospitalDataSet$Treatment <- factor(hospitalDataSet$Treatment)
drugs <- subset(hospitalDataSet, as.integer(Treatment)==1)
surgery <- subset(hospitalDataSet, as.integer(Treatment)==2)
summary(drugs)

```

##	Size	Treatment	Result
##	Min. :1.000	drugs :1000	Min. :1.000
##	1st Qu.:1.000	surgery: 0	1st Qu.:1.000
##	Median :1.000		Median :1.000
##	Mean :1.182		Mean :1.239
##	3rd Qu.:1.000		3rd Qu.:1.000
##	Max. :2.000		Max. :2.000

```
summary(surgery)
```

##	Size	Treatment	Result
##	Min. :1.000	drugs : 0	Min. :1.000
##	1st Qu.:2.000	surgery:1000	1st Qu.:1.000
##	Median :2.000		Median :1.000
##	Mean :1.895		Mean :1.342
##	3rd Qu.:2.000		3rd Qu.:2.000
##	Max. :2.000		Max. :2.000

Just from looking at the mean of the Result, the Drugs seem to be more effective, as the drug mean is 1.239 where the Treatment mean is 1342

```

drugsSmall <- subset(drugs, Size == 1)
surgerySmall <- subset(surgery, Size == 1)
summary(drugsSmall)

```

```

##      Size      Treatment      Result
## Min.   :1  drugs :818  Min.   :1.00
## 1st Qu.:1  surgery: 0  1st Qu.:1.00
## Median :1                      Median :1.00
## Mean   :1                      Mean   :1.18
## 3rd Qu.:1                      3rd Qu.:1.00
## Max.   :1                      Max.   :2.00

```

```
summary(surgerySmall)
```

```

##      Size      Treatment      Result
## Min.   :1  drugs : 0  Min.   :1.000
## 1st Qu.:1  surgery:105 1st Qu.:1.000
## Median :1                      Median :1.000
## Mean   :1                      Mean   :1.105
## 3rd Qu.:1                      3rd Qu.:1.000
## Max.   :1                      Max.   :2.000

```

Yes I change my recommendation as for the small tumors, the surgery has a mean of 1.105 and the drugs have a mean of the result of 1.18.

If the tumor is large:

```

drugLarge <- subset(drugs, Size == 2)
surgeryLarge <- subset(surgery, Size == 2)
summary(drugLarge)

```

```

##      Size      Treatment      Result
## Min.   :2  drugs :182  Min.   :1.000
## 1st Qu.:2  surgery: 0  1st Qu.:1.000
## Median :2                      Median :2.000
## Mean   :2                      Mean   :1.505
## 3rd Qu.:2                      3rd Qu.:2.000
## Max.   :2                      Max.   :2.000

```

```
summary(surgeryLarge)
```

```

##      Size      Treatment      Result
## Min.   :2  drugs : 0  Min.   :1.00
## 1st Qu.:2  surgery:895 1st Qu.:1.00
## Median :2                      Median :1.00
## Mean   :2                      Mean   :1.37
## 3rd Qu.:2                      3rd Qu.:2.00
## Max.   :2                      Max.   :2.00

```

I would recommend the Surgery, as it has a lower mean on the Result. Thus is more effective.

Am I Consistent?

I am not consistent in the cases above. The treatments have different effectivenesses for the size of the tumors. Thus the treatment decision must be done by taking into account the size of the tumor