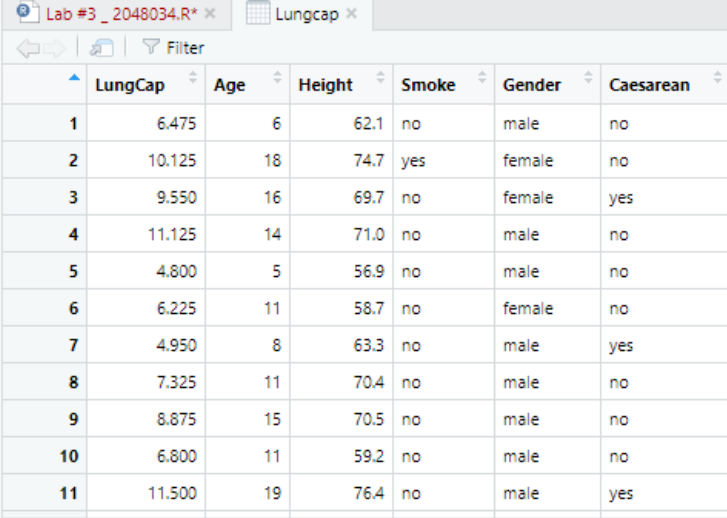


Demonstrate the concepts on Data Analysis with Summary Statistics and Scatter plots

Name: Harini G

1.Import the LungCapData into R and attach it



	LungCap	Age	Height	Smoke	Gender	Caesarean
1	6.475	6	62.1	no	male	no
2	10.125	18	74.7	yes	female	no
3	9.550	16	69.7	no	female	yes
4	11.125	14	71.0	no	male	no
5	4.800	5	56.9	no	male	no
6	6.225	11	58.7	no	female	no
7	4.950	8	63.3	no	male	yes
8	7.325	11	70.4	no	male	no
9	8.875	15	70.5	no	male	no
10	6.800	11	59.2	no	male	no
11	11.500	19	76.4	no	male	yes

Showing 1 to 12 of 725 entries, 6 total columns

2. Find the class and typeof Age and Height

```
> class(Lungcap$Age)
```

```
[1] "numeric"
```

```
> class(Lungcap$Height)
```

```
[1] "numeric"
```

```
> typeof(Lungcap$Age)
```

```
[1] "double"
```

```
> typeof(Lungcap$Height)
```

```
[1] "double"
```

3. Find the summary of the dataset

```
> summary(Lungcap)
```

LungCap	Age	Height	Smoke	Gender
Min. : 0.507	Min. : 3.00	Min. : 45.30	Length:725	Length:725
1st Qu.: 6.150	1st Qu.: 9.00	1st Qu.:59.90	Class :character	Class :character
Median : 8.000	Median :13.00	Median :65.40	Mode :character	Mode :character
Mean : 7.863	Mean :12.33	Mean :64.84		
3rd Qu.: 9.800	3rd Qu.:15.00	3rd Qu.:70.30		
Max. :14.675	Max. :19.00	Max. :81.80		

Caesarean

Length:725

Class :character

Mode :character

4.Find the strength of the relationship between Age and Height (Hint:Using Correlation)

```
> cor(Lungcap$Age,Lungcap$Height)
```

```
[1] 0.8357368
```

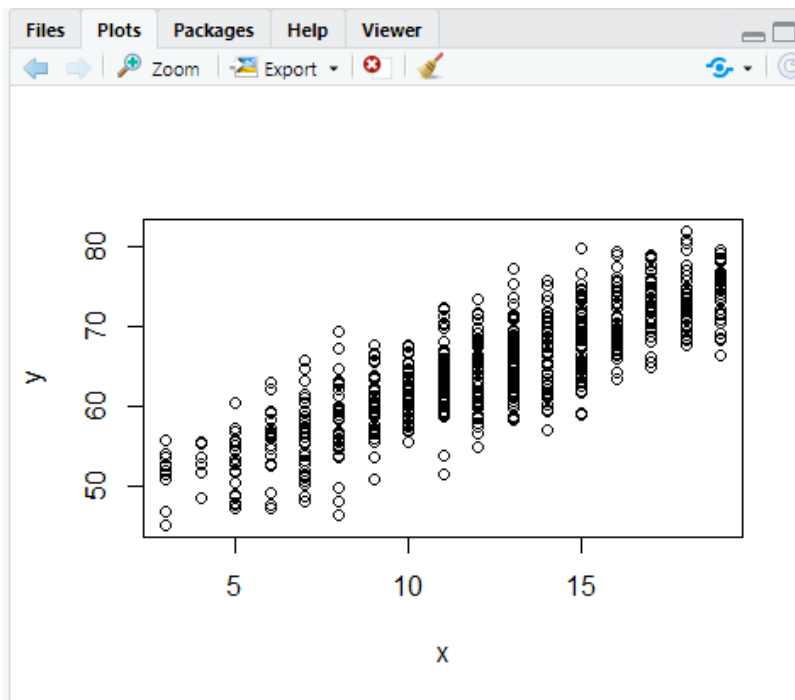
The relationship is positively strong correlated

5.Perform a scatterplot for the above (Qn.4)

```
> x=Lungcap$Age
```

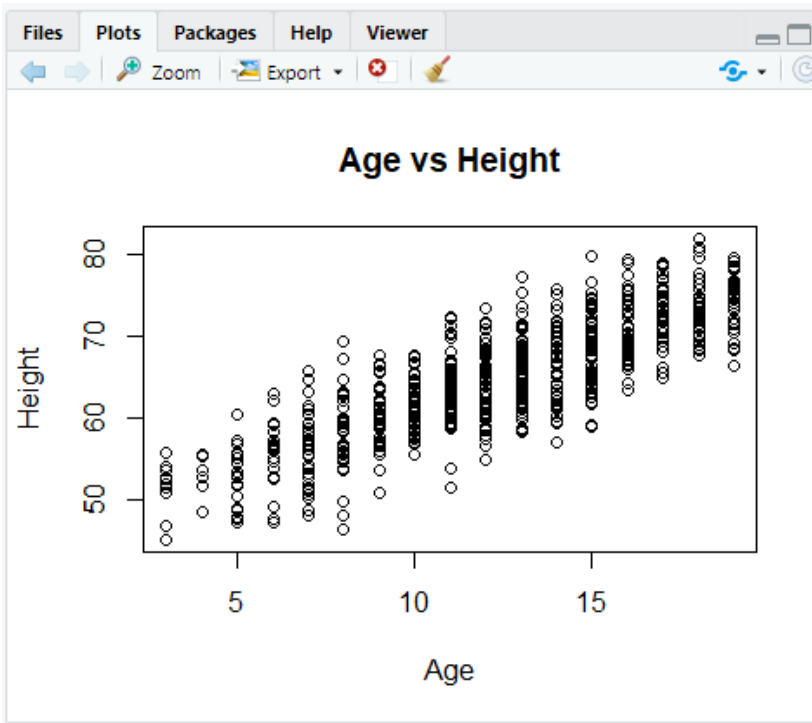
```
> y=Lungcap$Height
```

```
> plot(x,y)
```



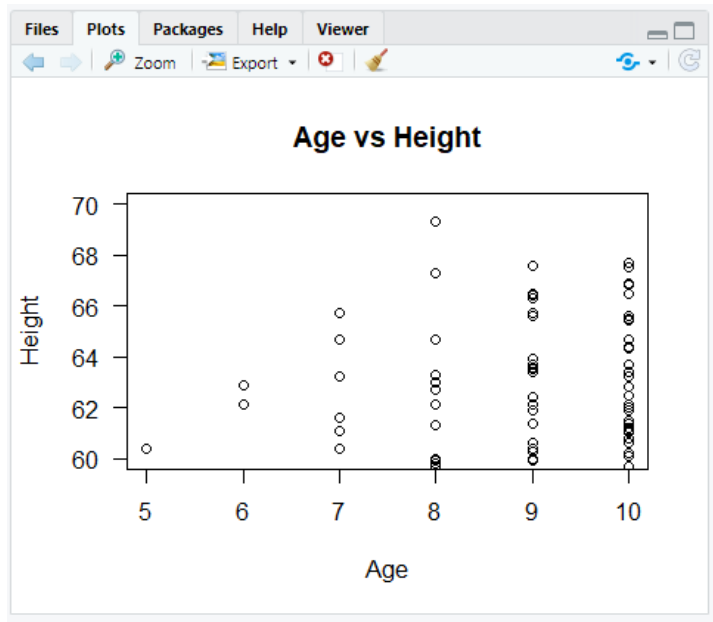
6. Add a title to the plot and Label x-axis and y-axis

```
> plot(x,y,xlab="Age",ylab="Height",main="Age vs Height")
```



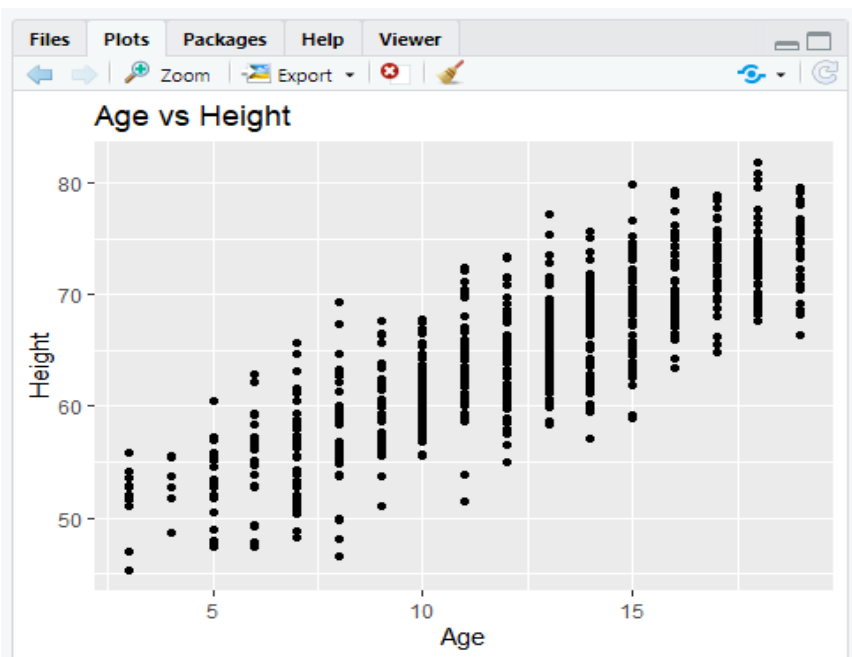
7. Rotate the values on the y-axis and also change the limits of x-axis and y-axis

```
> plot(x,y,xlab="Age",ylab="Height",main="Age vs  
Height",las=1,xlim=c(5,10),ylim=c(60,70))
```

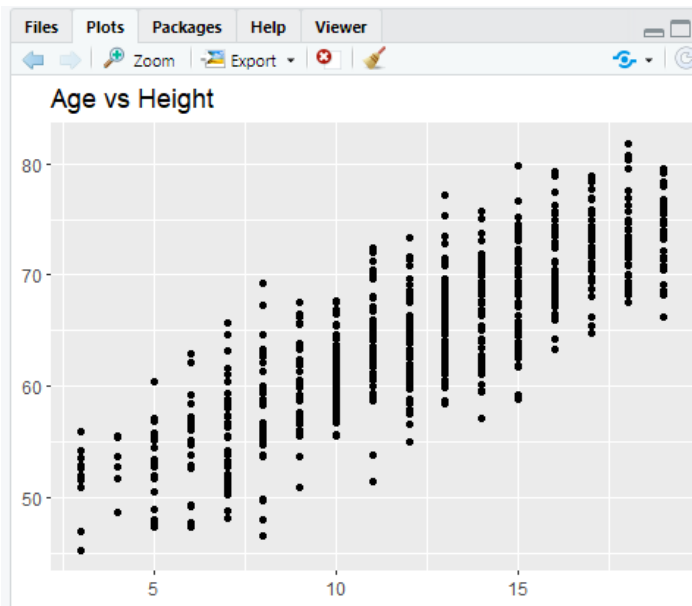


8. Remove the axes of the plot and relabel these axes

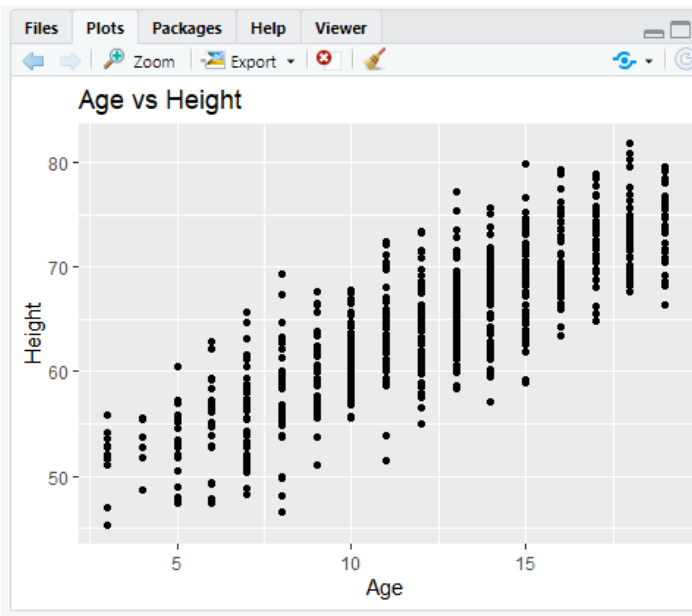
```
> p=ggplot(Lungcap,aes(x,y))+labs(x = "Age", y = "Height",title="Age vs Height")  
> p+geom_point()
```



```
p + theme(axis.title = element_blank()) + geom_point()
```

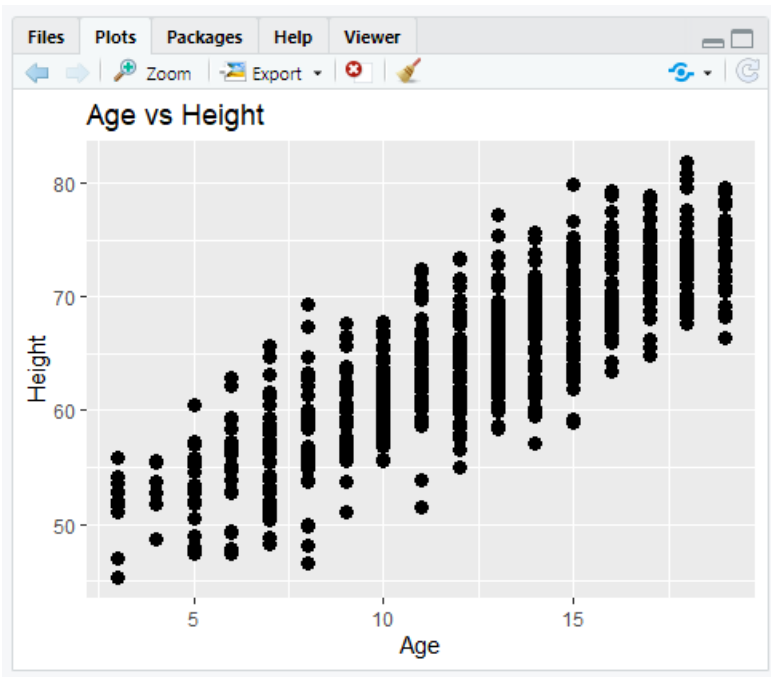


```
p + labs(x="Age",y="Height") + geom_point()
```



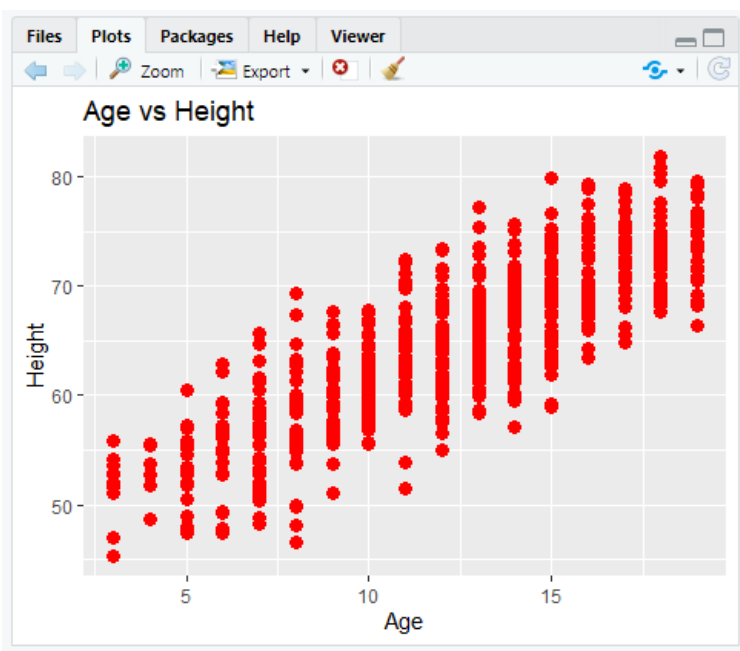
9. Change the size of the plotting characters

```
p+geom_point(size=3)
```



10. Change the color of the characters to red

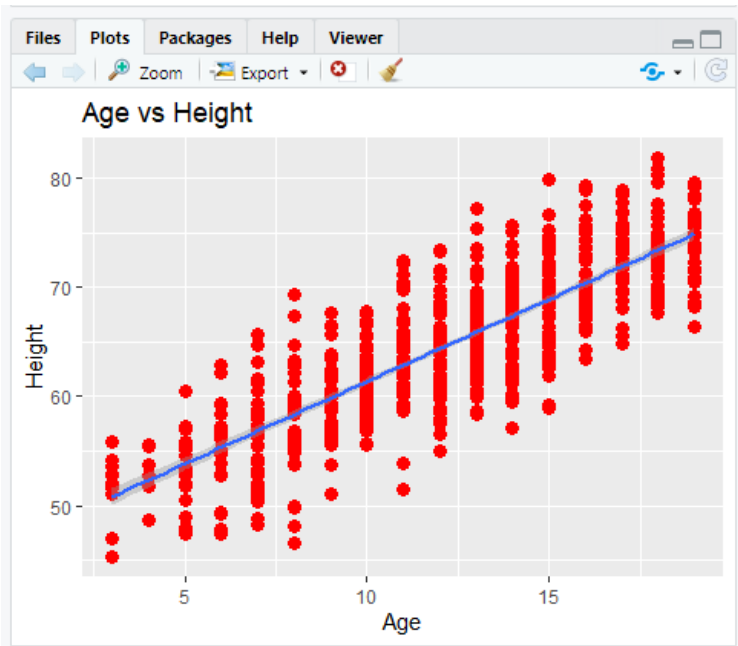
```
> p+geom_point(size=3,color="red")
```



11. Draw the regression line to the plot predicting height using age

```
> p+geom_point(size=3,color="red")+geom_smooth(method=lm)
```

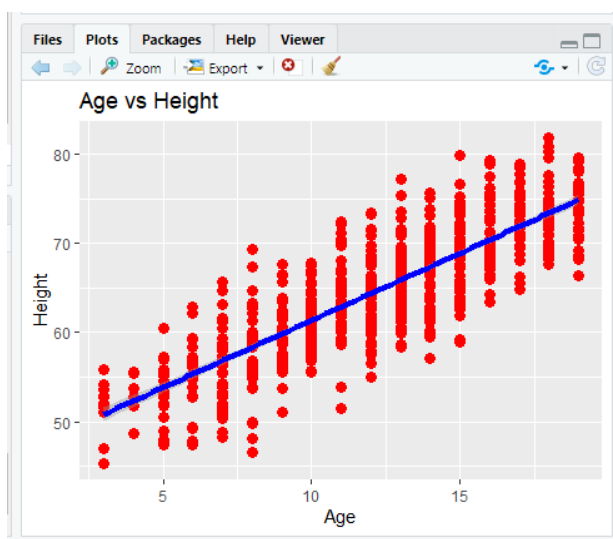
`geom_smooth()` using formula 'y ~ x'



12. Change the color of the line to blue and the width of the line

```
> p+geom_point(size=3,color="red")+geom_smooth(method=lm,size = 1.5,color="blue")
```

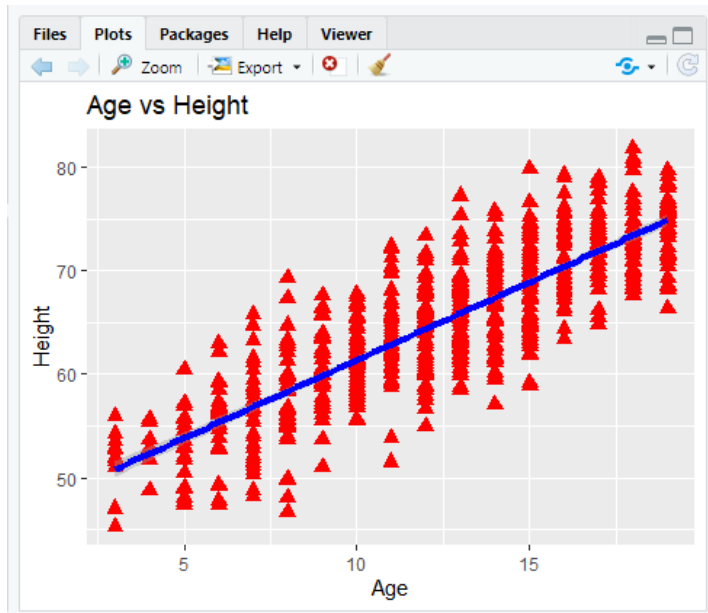
`geom_smooth()` using formula 'y ~ x'



13. Change the font of the plotting characters

```
> p+geom_point(size=3,color="red",shape =  
17)+geom_smooth(method=lm,size=1.5,color="blue")
```

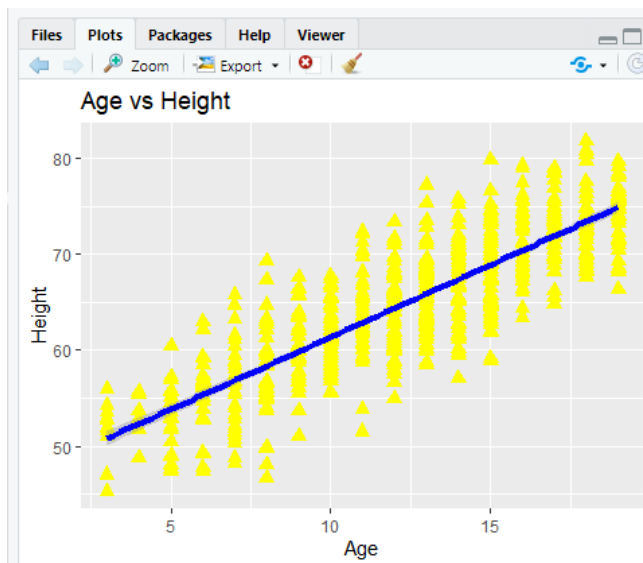
`geom_smooth()` using formula 'y ~ x'



14. Change the color of the plotting characters

```
> p+geom_point(size=3,shape =  
17,color="yellow")+geom_smooth(method=lm,size=1.5,color="blue")
```

`geom_smooth()` using formula 'y ~ x'



15. Identify gender on the plot for the age male and female

```
> ggplot(Lungcap,aes(Age,Height))+geom_point(shape =  
17,size=3,aes(colour=factor(Gender)))+geom_smooth(method=lm,size=1.5,color="blue")
```

`geom_smooth()` using formula 'y ~ x'

