

# Lab3.R

rstudio-user

2021-01-29

## *#1. Reading file*

```
install.packages("readxl")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'  
## (as 'lib' is unspecified)
```

```
library("readxl")  
fl <- read_excel("lungcap.xlsx")  
print(fl)
```

```
## # A tibble: 725 x 6  
##   LungCap Age Height Smoke Gender Caesarean  
##   <dbl> <dbl> <dbl> <chr> <chr> <chr>  
## 1 6.48 6 62.1 no male no  
## 2 10.1 18 74.7 yes female no  
## 3 9.55 16 69.7 no female yes  
## 4 11.1 14 71 no male no  
## 5 4.8 5 56.9 no male no  
## 6 6.22 11 58.7 no female no  
## 7 4.95 8 63.3 no male yes  
## 8 7.32 11 70.4 no male no  
## 9 8.88 15 70.5 no male no  
## 10 6.8 11 59.2 no male no  
## # ... with 715 more rows
```

## *#2. Find the class and typeof Age and Height*

```
class(fl$Age)
```

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

```
typeof(fl$Age)
```

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

```
class(fl$Height)
```

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

```
typeof(fl$Height)
```

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

## *#3. Find the summary of the dataset*

```
summary(fl)
```

```
##   LungCap           Age           Height           Smoke  
## Min.      : 0.507   Min.      : 3.00   Min.      :45.30   Length:725
```

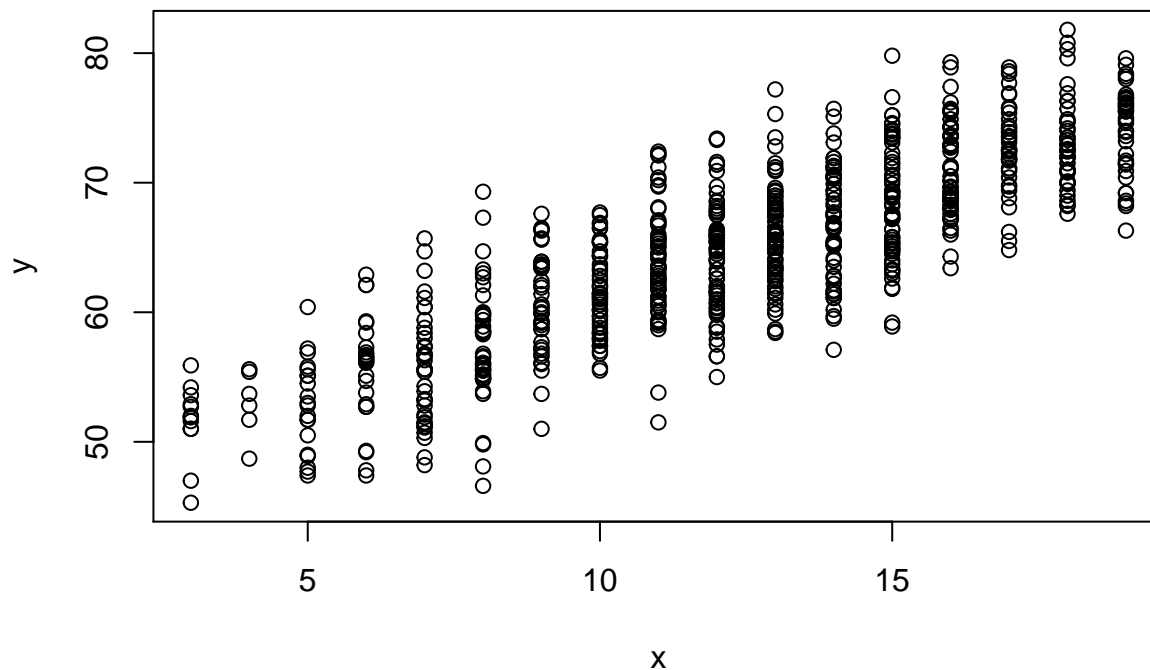
```
## 1st Qu.: 6.150 1st Qu.: 9.00 1st Qu.:59.90 Class :character
## Median : 8.000 Median :13.00 Median :65.40 Mode :character
## Mean : 7.863 Mean :12.33 Mean :64.84
## 3rd Qu.: 9.800 3rd Qu.:15.00 3rd Qu.:70.30
## Max. :14.675 Max. :19.00 Max. :81.80
## Gender Caesarean
## Length:725 Length:725
## Class :character Class :character
## Mode :character Mode :character
##
##
##
```

```
#4. Find the strength of the relationship between Age and Height
cor(fl$Age, fl$Height)
```

```
## [1] 0.8357368
```

```
#5. Perform a scatter plot for the above (Qn.4)
```

```
x <- fl$Age
y <- fl$Height
plot(x, y)
```



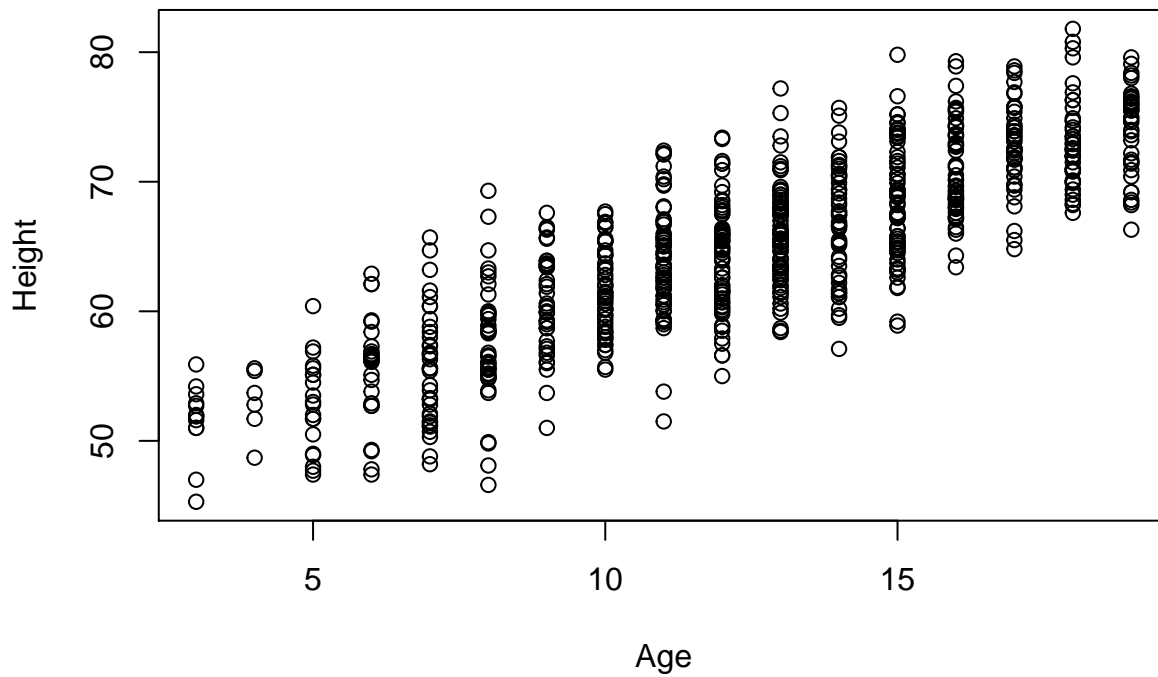
```
#6. Add a title to the plot and Label x-axis and y-axis
plot(x, y, xlab="Age", ylab="Height", main="Age vs Height Correlation")
```

```
#7. Rotate the values on the y-axis and also change the limits of x-axis and y-axis
install.packages("ggplot2")
```

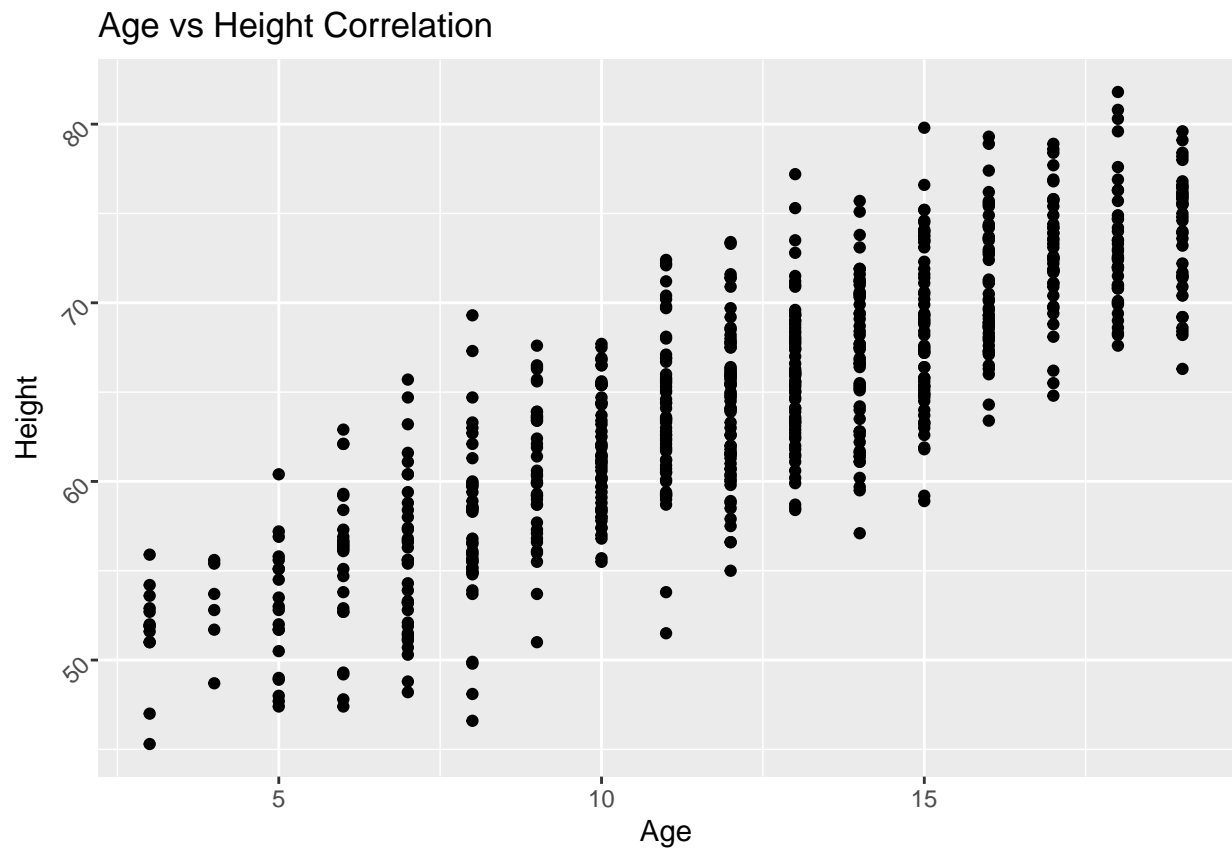
```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
```

## Age vs Height Correlation

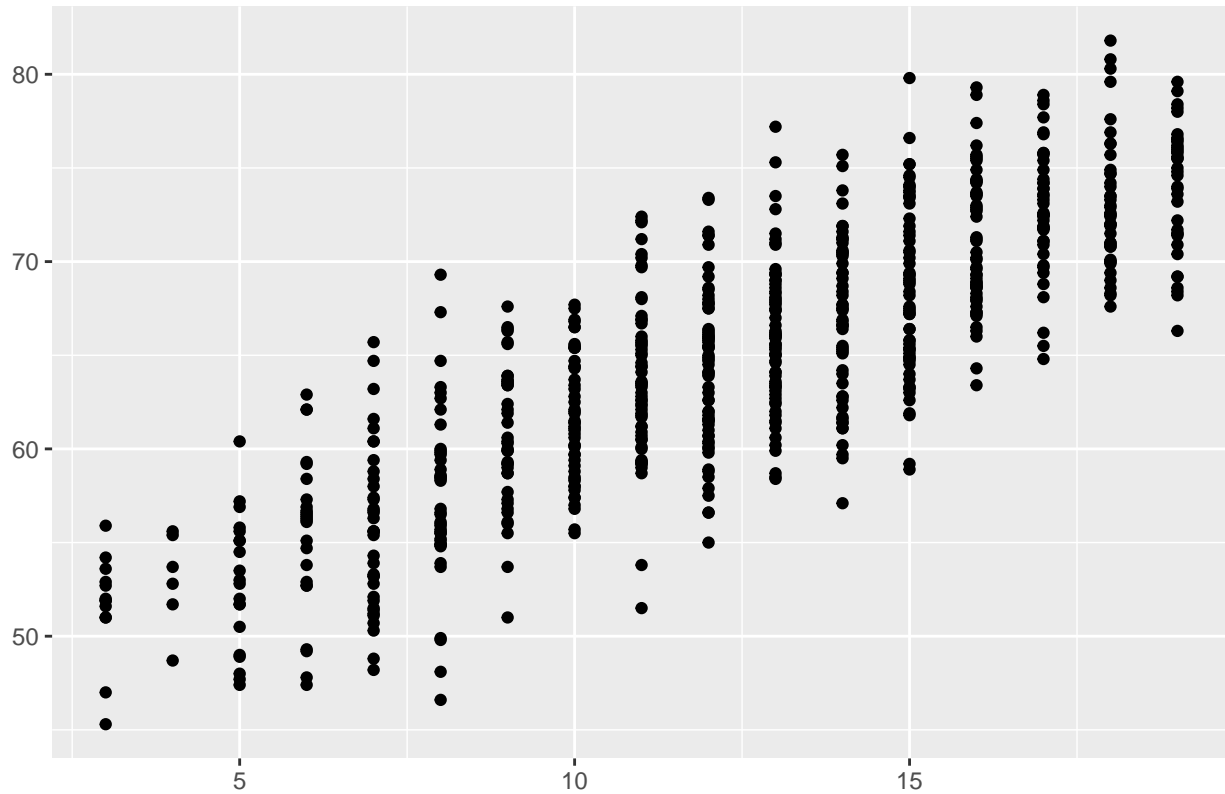


```
fig = ggplot(fl, aes(Age, y=Height))  
fig+ggtitle("Age vs Height Correlation")+theme(axis.text.y=element_text(angle=45))+geom_point()
```



```
#8. Remove the axes of the plot and relabel these axes
fig+ggtitle("Age vs Height Correlation")+theme(axis.title.x=element_blank(), axis.title.y=element_blank()
```

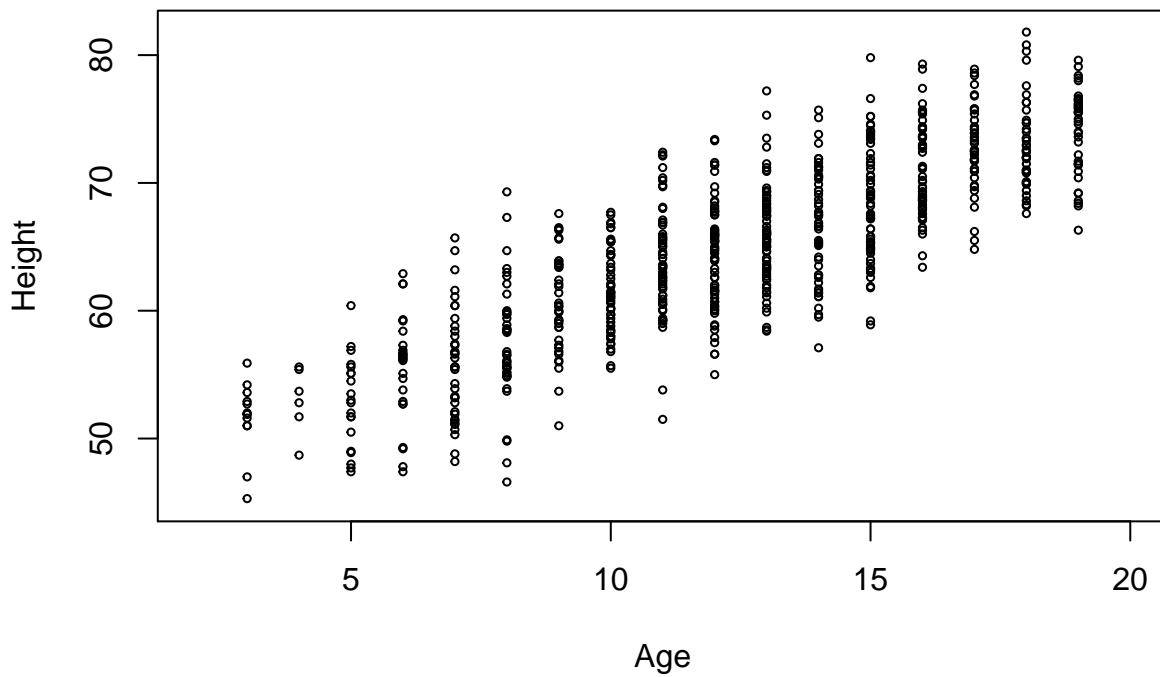
Age vs Height Correlation



*#9. Change the size of the plotting characters*

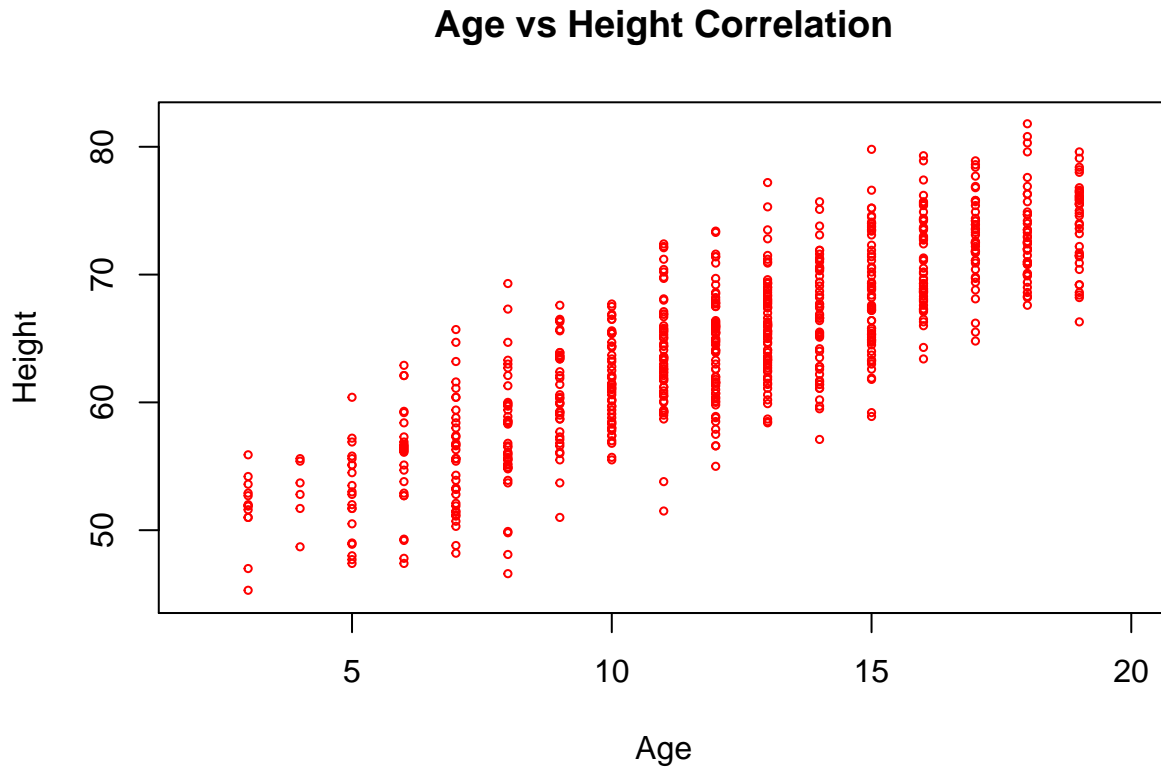
```
plot(x, y, xlab="Age", ylab="Height", main="Age vs Height Correlation", xlim=c(2,20), ylim=c(45, 82), cex=1.5)
```

Age vs Height Correlation



#10. Change the color of the characters to red

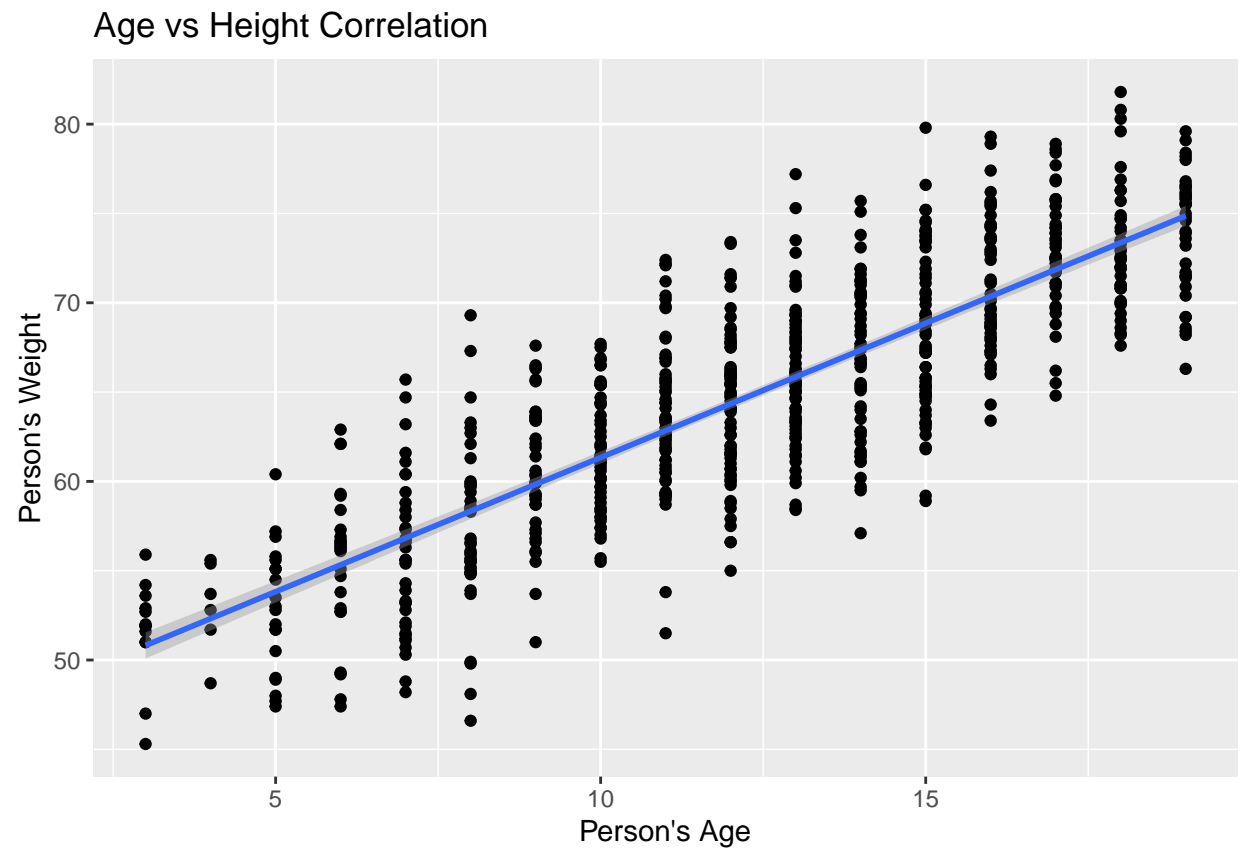
```
plot(x, y, xlab="Age", ylab="Height", main="Age vs Height Correlation", xlim=c(2,20), ylim=c(45, 82), col="red")
```



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

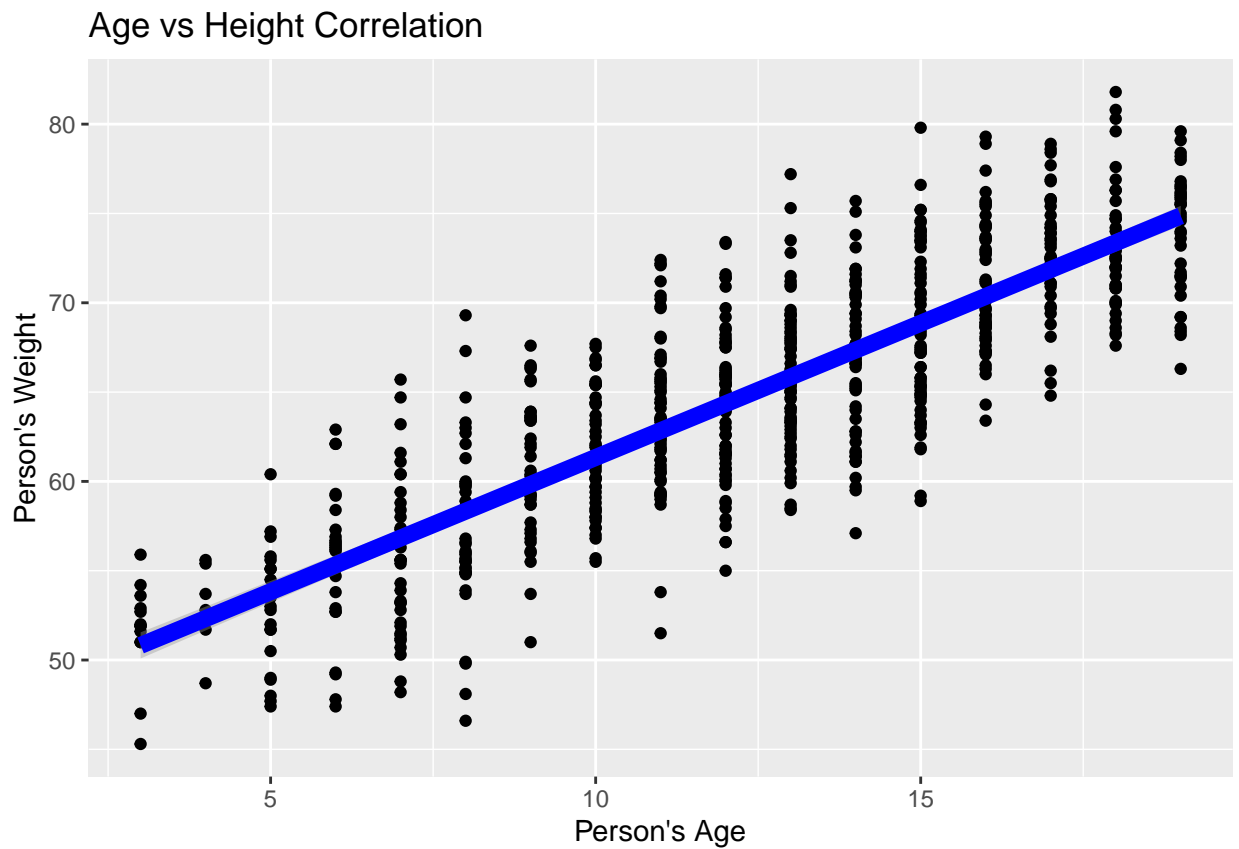
```
fig+ggtitle("Age vs Height Correlation")+labs(x="Person's Age", y="Person's Weight")+geom_point()+geom_smooth()
```

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



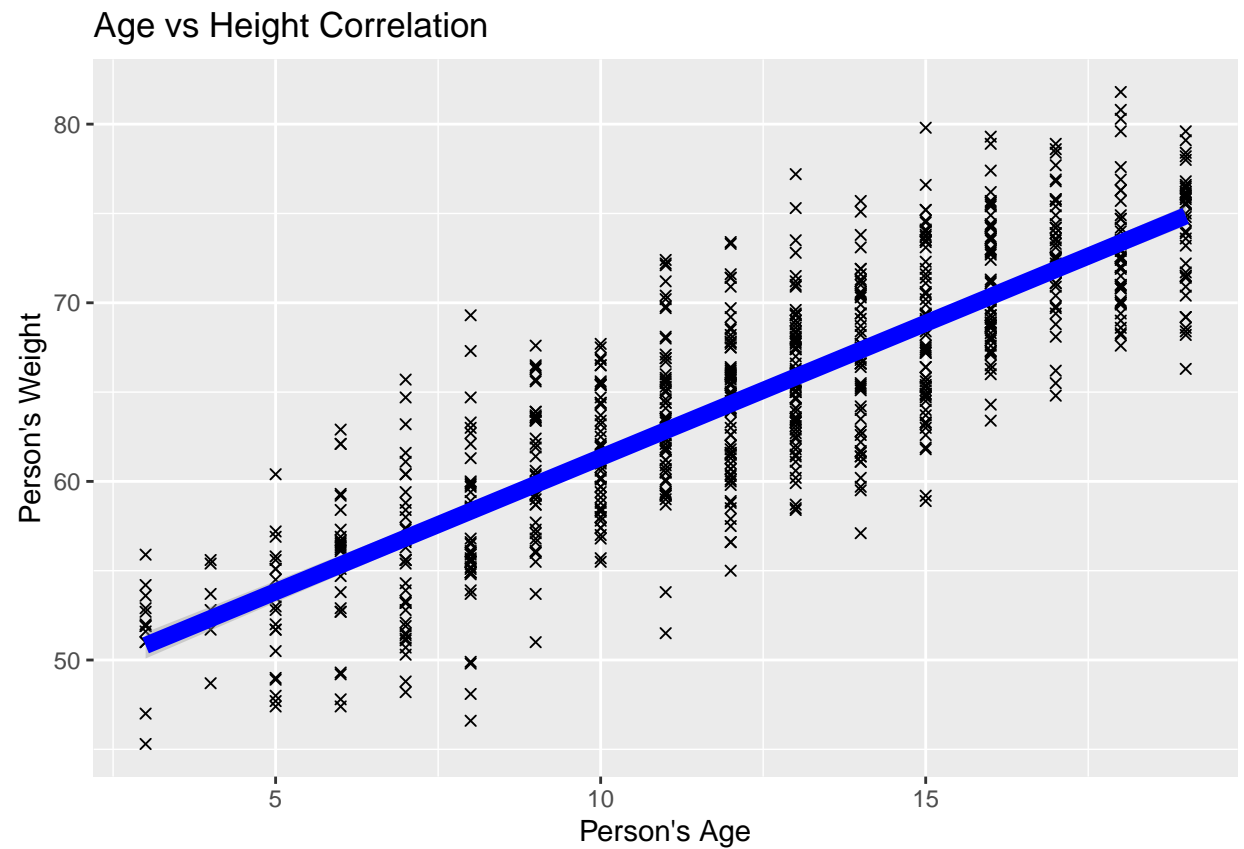
```
#12. Change the color of the line to blue and the width of the line
fig+ggtitle("Age vs Height Correlation")+labs(x="Person's Age", y="Person's Weight")+geom_point()+geom_smooth(linetype="solid", color="blue", size=2)

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



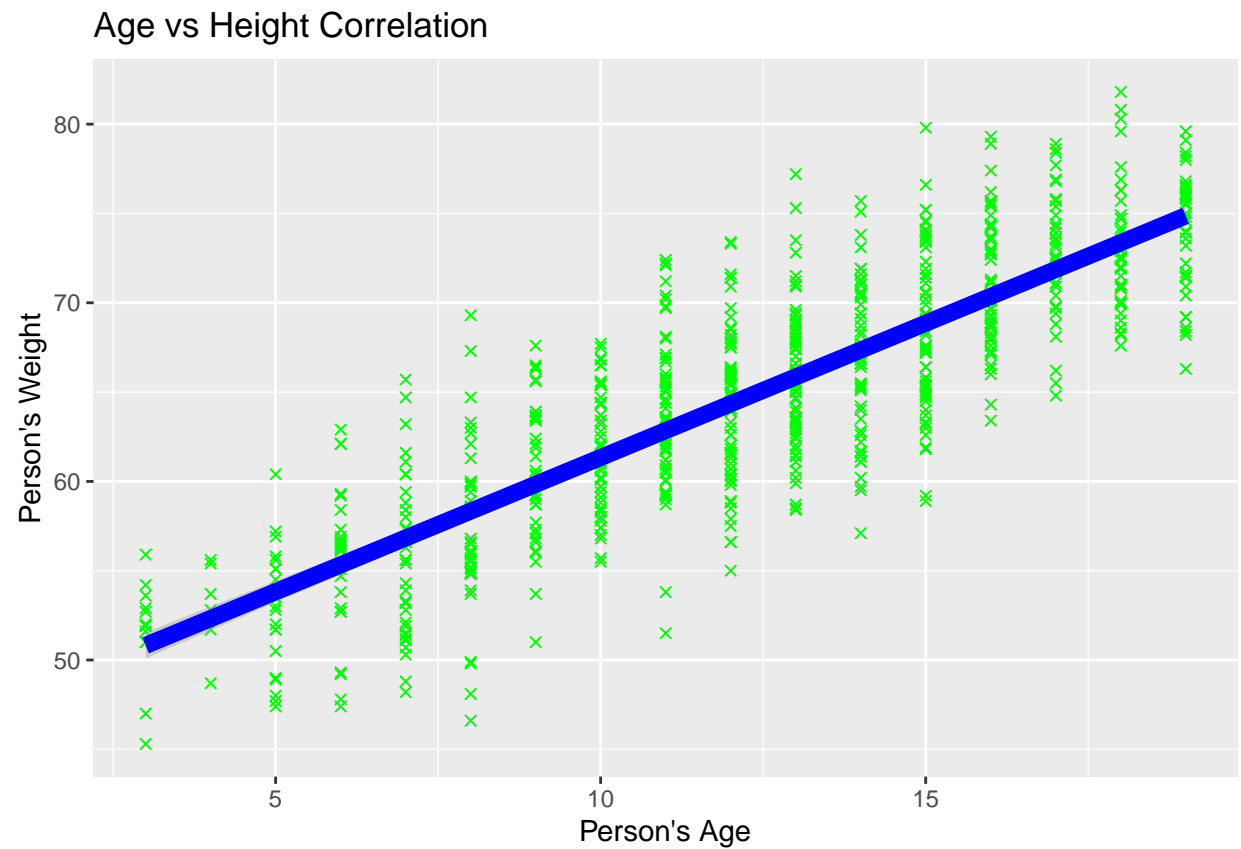
```
#13. Change the font of the plotting characters  
fig+ggtitle("Age vs Height Correlation")+labs(x="Person's Age", y="Person's Weight")+geom_point(shape=4,  
## `geom_smooth()` using formula 'y ~ x'
```





```
#14. Change the color of the plotting characters
fig+ggtitle("Age vs Height Correlation")+labs(x="Person's Age", y="Person's Weight")+geom_point(shape=4)

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



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

```
ggplot(fl, aes(Age, y=Height, colour=factor(Gender)))+ggtitle("Age vs Height Correlation")+labs(x="Person's Age", y="Person's Weight")
```

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

