# SCATTER PLOT

SCATTER PLOT

- This plot is used for showing relation between the two variables
- It is formed when we have two quantitative or numeric data one is arranged along x-axis and the other along y-axis
- Many of the times the data points on the graph are found to be scattered around and so this plot is called as scattered plot
- If the graph points are following certain specific pattern then this shows the correlation between the two variables.
- When the plotpoints do not follow any pattern or seems to be scattered around then we say that there is no correlation

- If the y coordinates increases with the x-coordinates or vice versa then we say that the correlation is positive

for example : hours spent in running and calories burnt
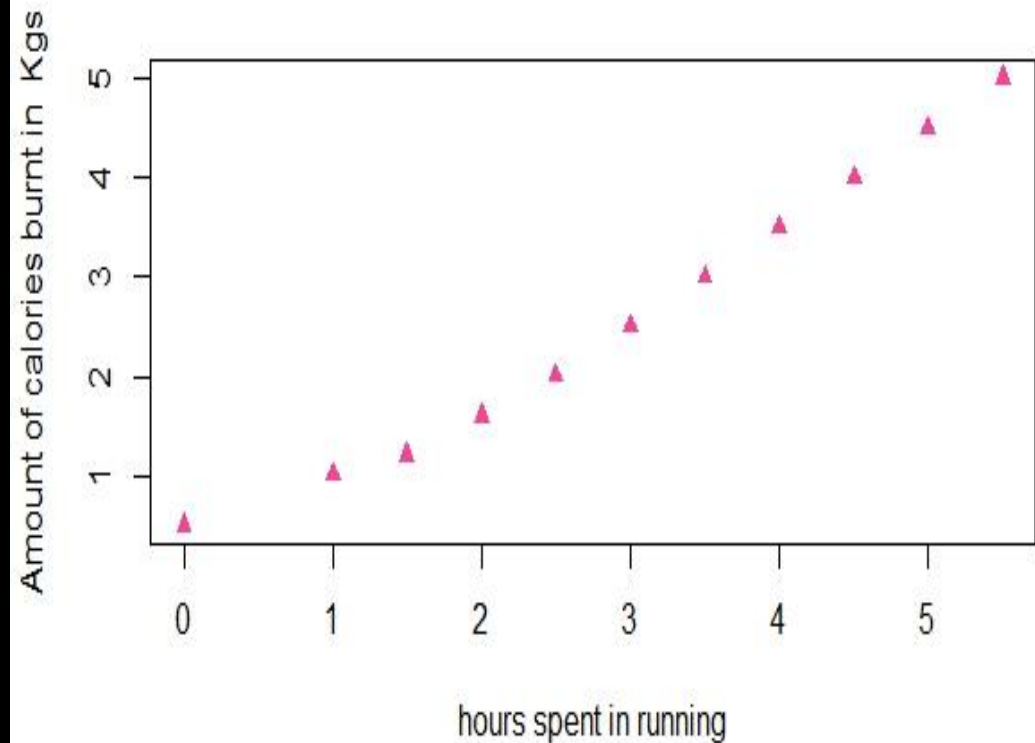
NEGATIVE CORRELATION

- If the y-coordinates are decreasing with increase in x-coordinates or vice versa then we say that the correlation is negative

- Example ; the relation between height above the sea level and temperature
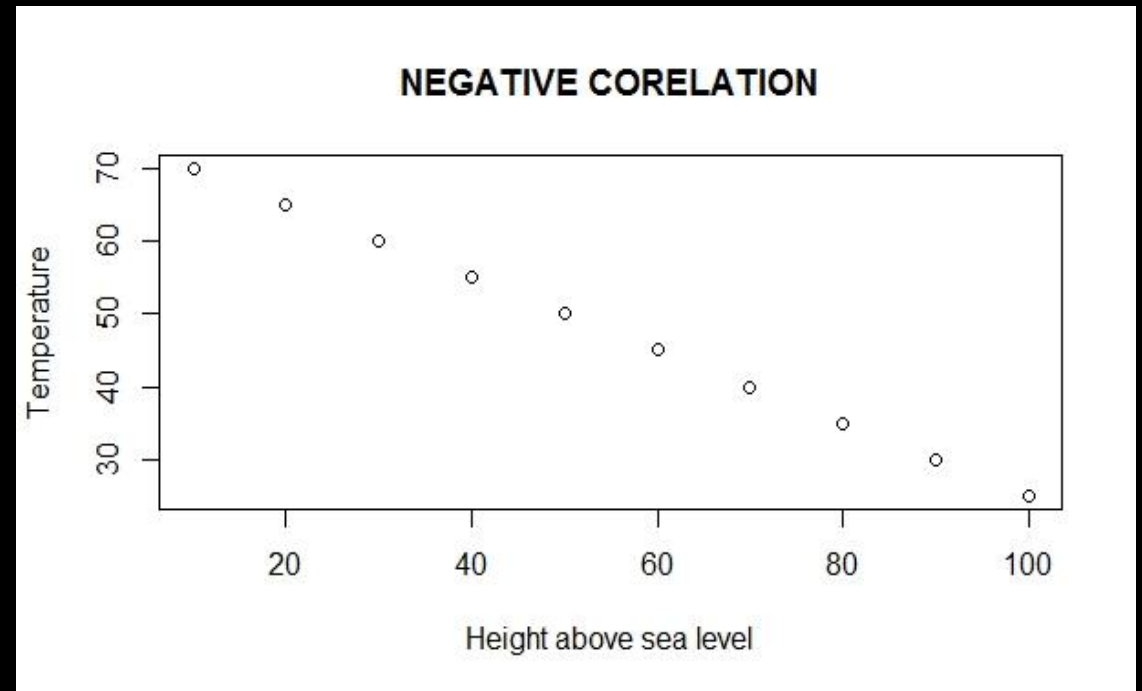
CONSTANT CORRELATION

- And if the y-coordinates remains at constant value then the correlation is said to be constant .

- Example : Relation between marks obtained and passing marks

CONSTANT CORRELATION
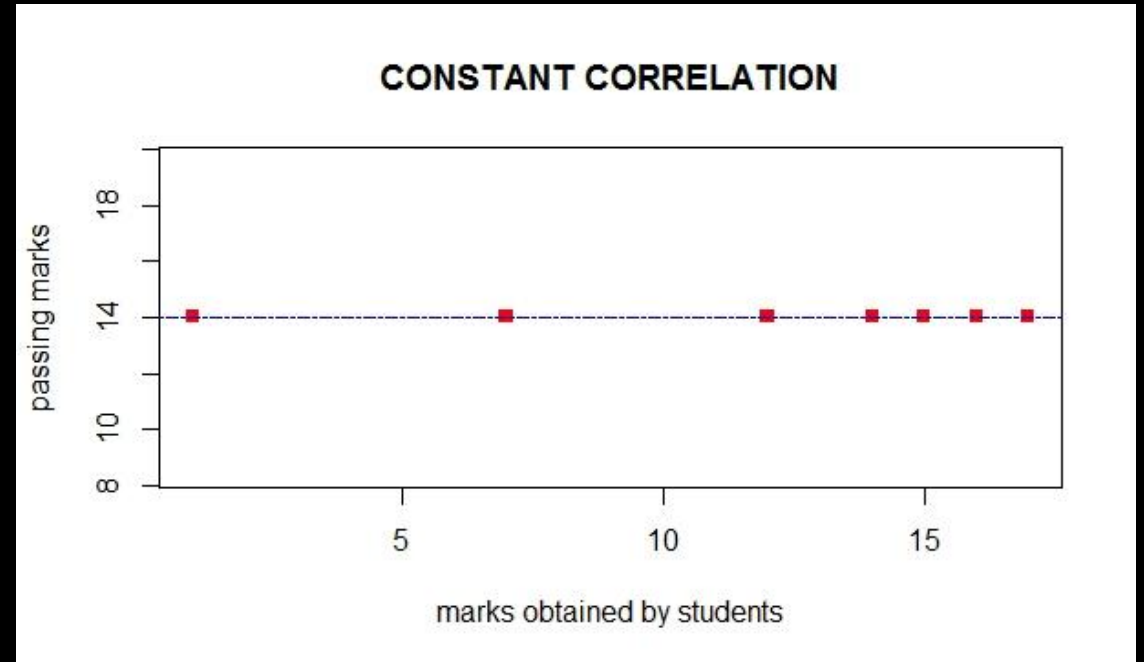
NO CORRELATION

- If the y-coordinates and x-coordinates are having their values randomly not following any pattern then we say that there is no correlation.

- For example : The shoe size of students and their IQ level
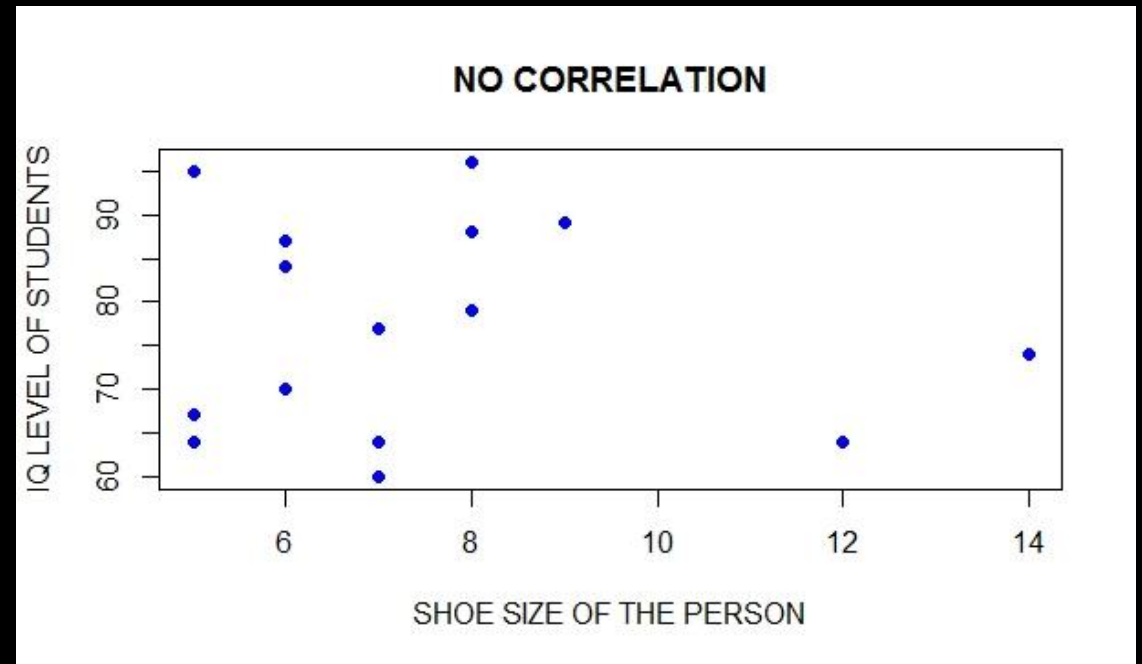
COMMANDS AND THEIR USE IN R FOR CREATING AN SCATTER PLOT

- Suppose we have a and b as two variable defined in R
- Then it is very important to check the length of the variables because the length of the two variables should be same if not we won't get any output.
- The simplest command to create an scatter plot with above two variables is plot(a,b)

  And also the first variable in the plot command always appears in horizontal axis and the second one in the vertical axis
- If we have two numeric columns (a,b) of data frame(w) then we can form scatter plot with the command
- plot(wa,wb) is the simplest scatter plot with column a and b of data frame w where a represents the horizontal and b represents the vertical column of data frame

CUSTOMIZATION COMMANDS USED IN SCATTER PLOT

- xlab='' used to label x –axis or horizontal axis
- ylab=' ' used to label y-axis or vertical axis
- xlim =c(starting point,end point)and ylim =c(starting point,end point)are used to set the range of x-axis and y-axis respectively
- pch=0 to 25, is used to choose plotting characters
- cex=0 to 25,is used to choose or alter the size of plotting characters.
- col=' ' is used to choose color of the plotting characters
- las=1 is used to rotate the numbers written on vertical axis

SCATTER PLOT WITH TWO VARIABLES

```
speed<-c(10,20,30,40,50,60,70,80,90,100)

speed

length(speed)

time<-c(10,9,8,7,6,5,4,3,2,1)

time

length(time)

plot(time,speed,pch=18,cex=2,col='springreen3',xlab='Time required to cover certain distance',ylab='speed of the car')
```
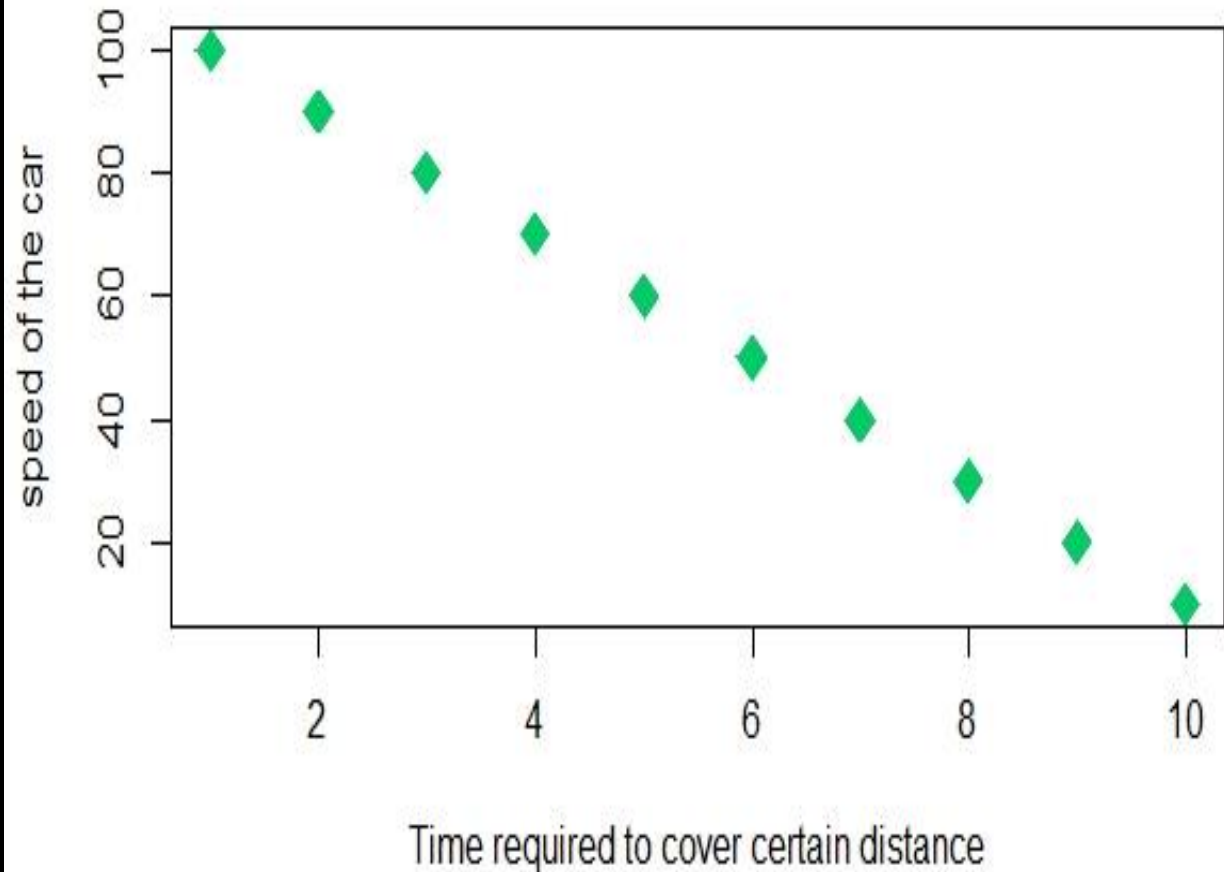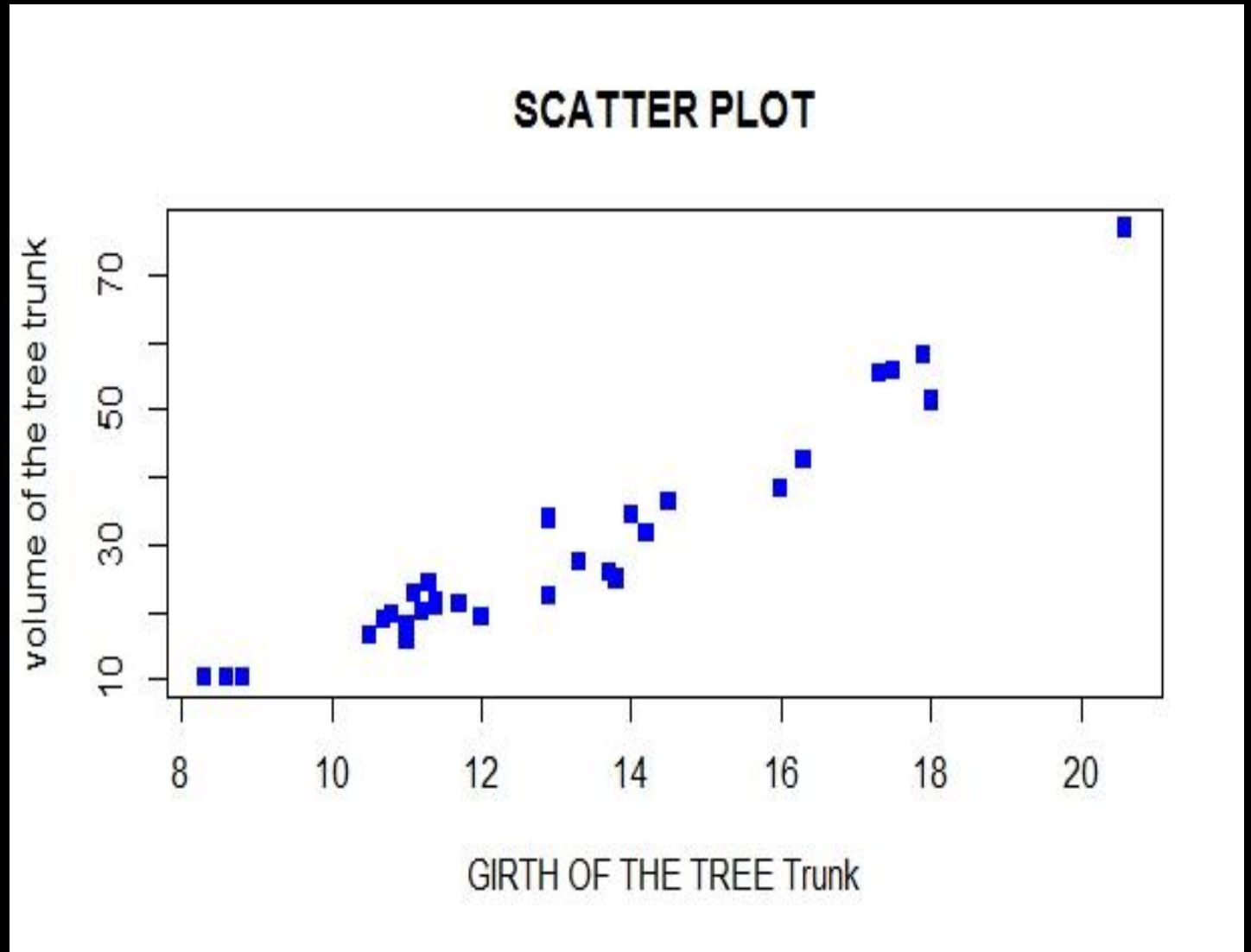
EXAMPLE OF SCATTER PLOT WITH COMMANDS

If trees are the inbuilt dataframe in R and we want to form scatter plot with the girth and volume column of trees then the commands are

trees

plot(trees$Girth,trees$Volume,col='blue',pch=15,cex=1,main='SCATTER PLOT',xlab='GIRTH OF THE TREE Trunk',ylab='volume of the tree trunk')

HOW TO ADD LINE TO SCATTER PLOT

- For adding line to scatter plot first we need to create a linear model with the two variables or the two columns of dataframe

- Linear model between the two variable is used to calculate the slope at every coordinate point

- This slope is used by R to add line in scatter plot because this line depends on the slope of the points
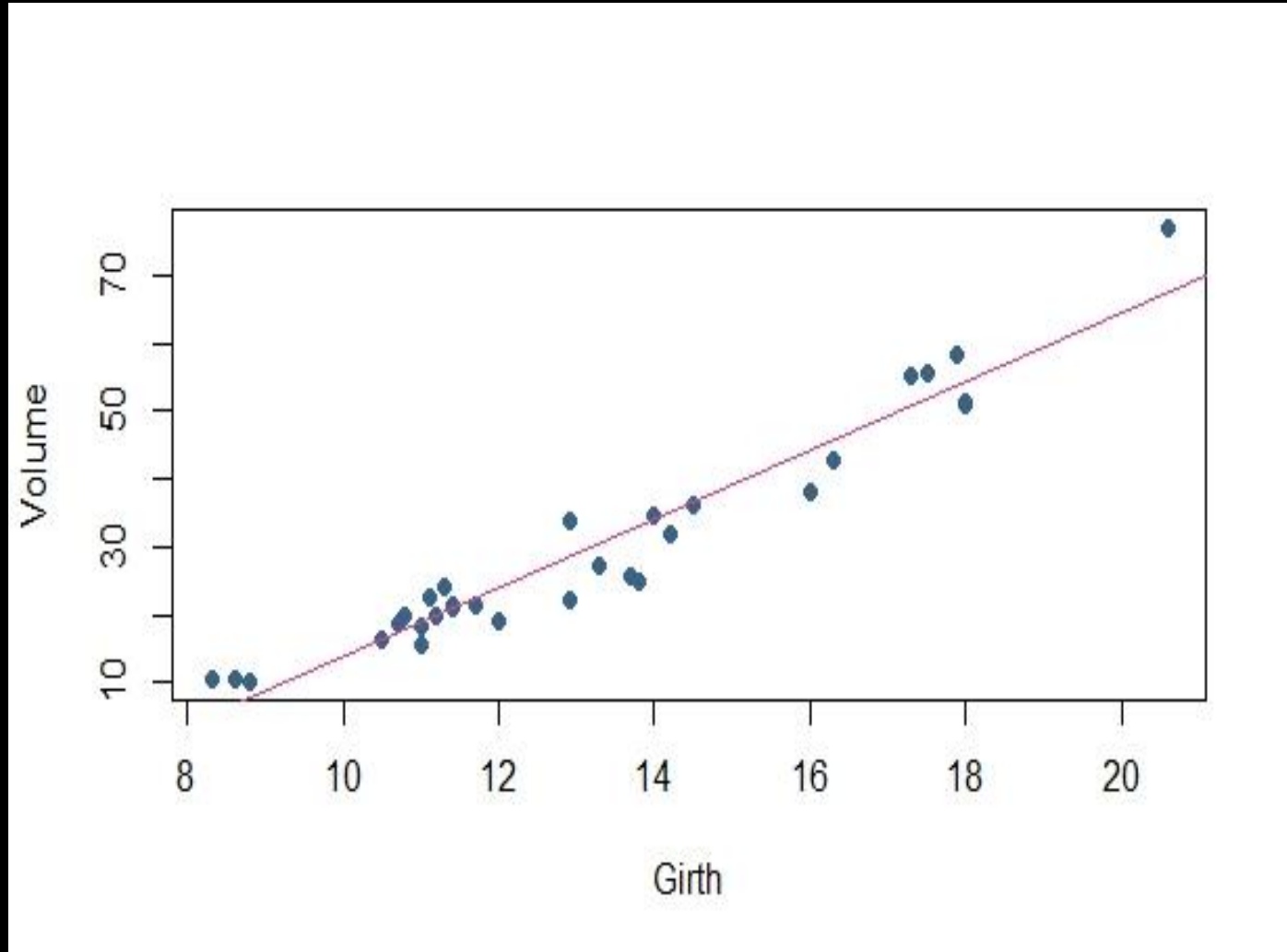
COMMANDS USED FOR ADDING A LINE TO SCATTER PLOT

plot(Volume~Girth,data=trees,col='steelblue4',pch=16)

abline(lm(Volume~Girth,data=trees,),col='violetred2')

The first variable in linear plot represents the vertical axis and the second one represents the horizontal axis.

abline helps in understanding the correlation

CUSTOMIZATION COMMANDS FOR ABLINE

- lty= is used to choose the line type in two ways either a numeric value or name in quotes

- By default line type is 1 or 'solid'

- Colour can be given to the line by using col command

- Customization commands for line should be given in the abline command

- And customization commands for plotting points should be given in the plot commands

EXAMPLE

Mtcars

#incase we just want to create scatter plot and not to add line

plot(mtcars$gear,mtcars$carb,pch=19, cex=2,col='springgreen3',main='SCATTER PLOT',xlab='GEAR',ylab='CARB')

#in case we want to make a line in the scatter plot

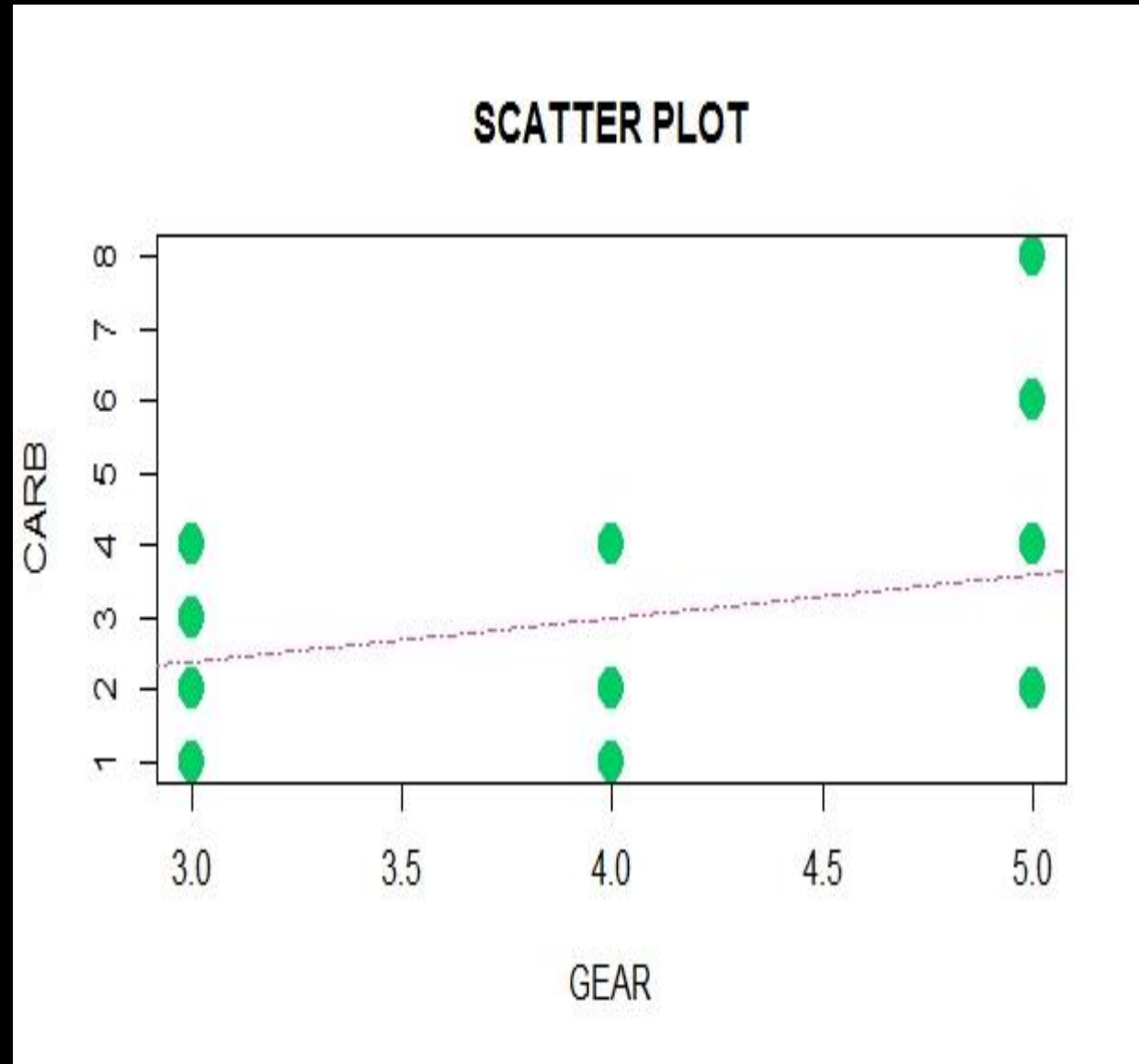plot(carb~gear,data=mtcars,pch=19,cex=2,col='springgreen3',main='SCATTER PLOT',xlab='GEAR',ylab='CARB')

)

abline(lm(carb~gear,data=mtcars),col='violetred1',lty=6)

The customization commands that are used in the plot command will make changes in coordinate points.

And the commands used in abline will make changes in line in the scatter plot
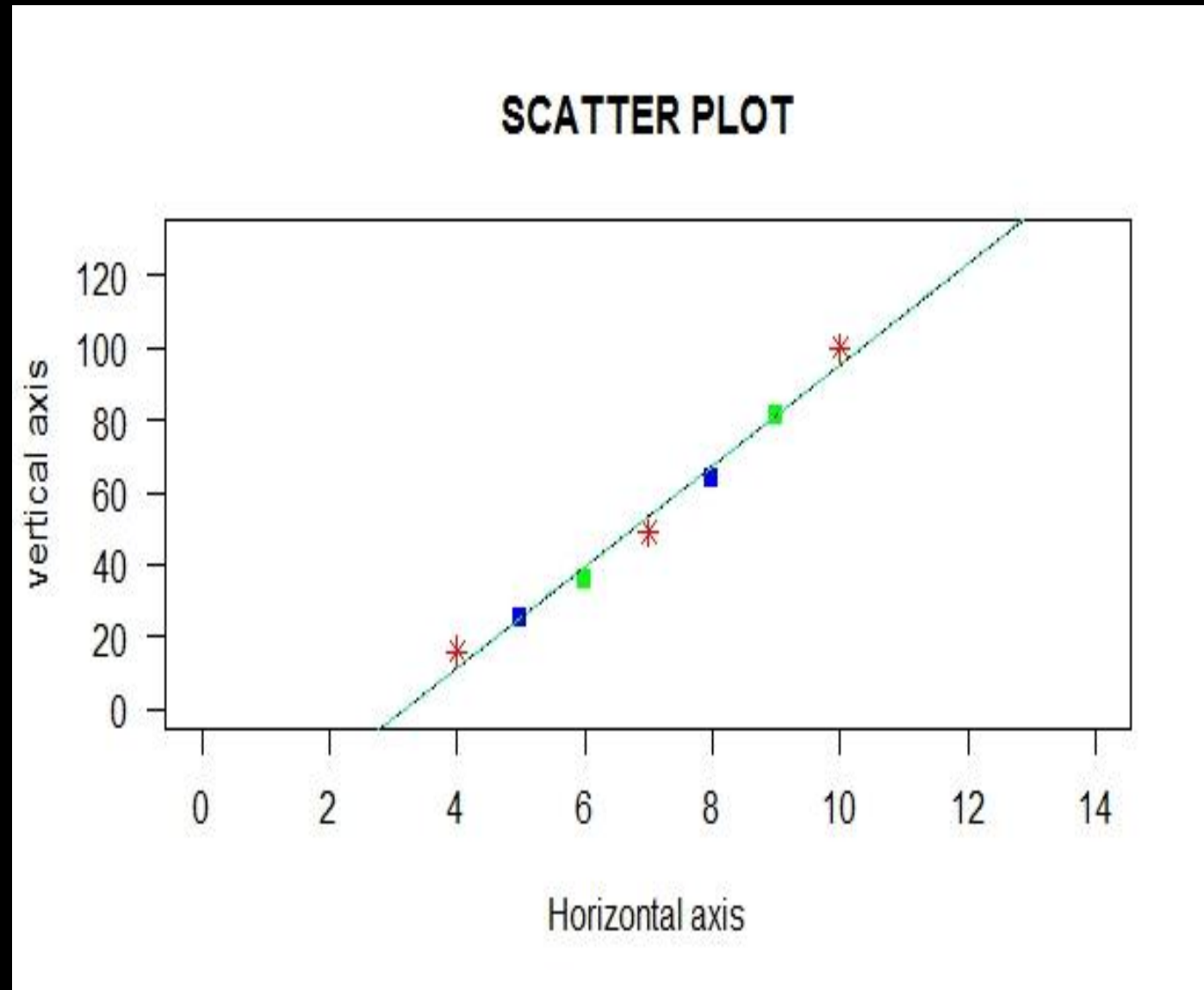
CREATING AN SCATTER PLOT

- #suppose we have two numeric datas and we want to create an scatter plot to see
- #the relation between them
- #x<-c(4,5,6,7,8,9,10)
- #y<-c(16,25,36,49,64,81,100)
- x<-c(4,5,6,7,8,9,10)
- #first we need to check the number of variable in x
- length(x)#used to find the number of elements in x
- y<-c(16,25,36,49,64,81,100)
- #similarly we need to find the length of y
- length(y)#used to find number of elements in y
- #if the length of the two variable do not come our to be equal then we won't get an
- #output for scatter plot
- plot(x,y) # this is the simplest command to form an scatter plot using two variables
- #Also the first variable in the plot command appears along horizontal axis and the second one along vertical axis
- plot(x,y,las=1)#las command is used to rotate the number written along vertical axis

PLOTTING THE SCATTER PLOT

```
plot(x,y,las=1,pch=c(8,15,15),cex=1,col=c('red','blue','green'),xlab='Horizontal axis',ylab='vertical axis',main='scatter plot',xlim=c(0,14),ylim=c(0,130))

abline(lm(y~x),lty=2,lwd=1,col='springgreen')
```

THANK YOU