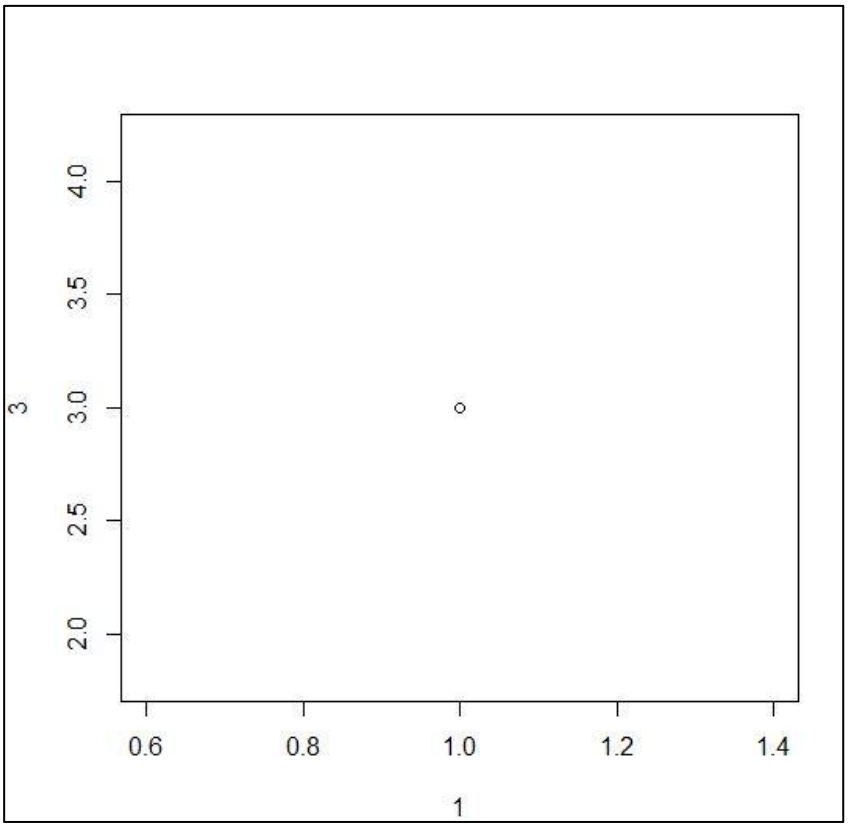
Practical No. 6

Title: - Introduction to R Graphics and Data Importing

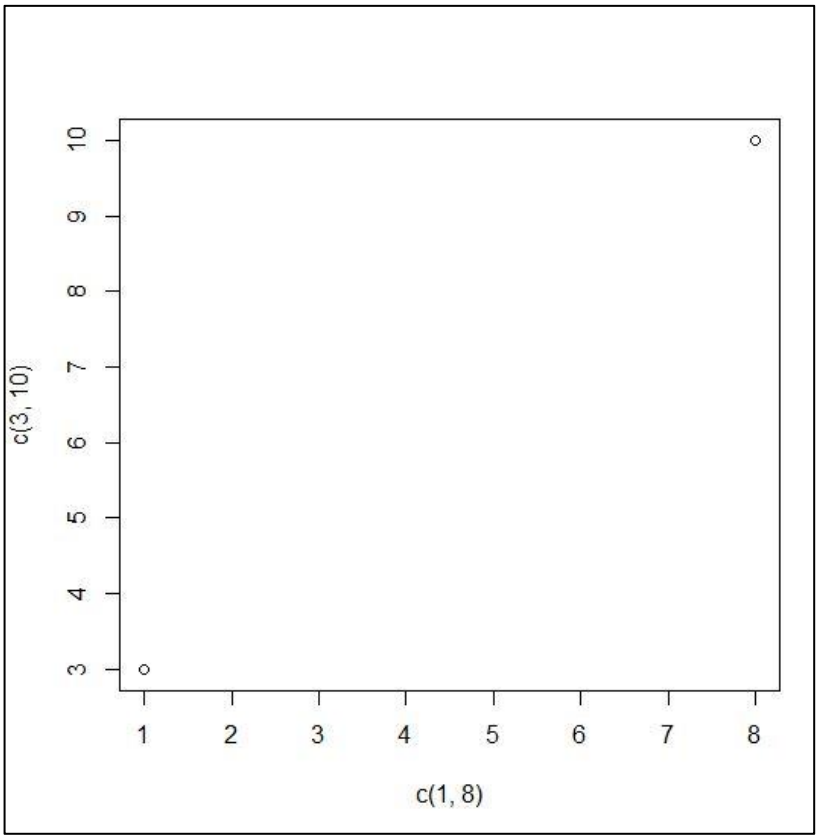
Aim: - To understand R Graphics and how to import data.

Plot() function

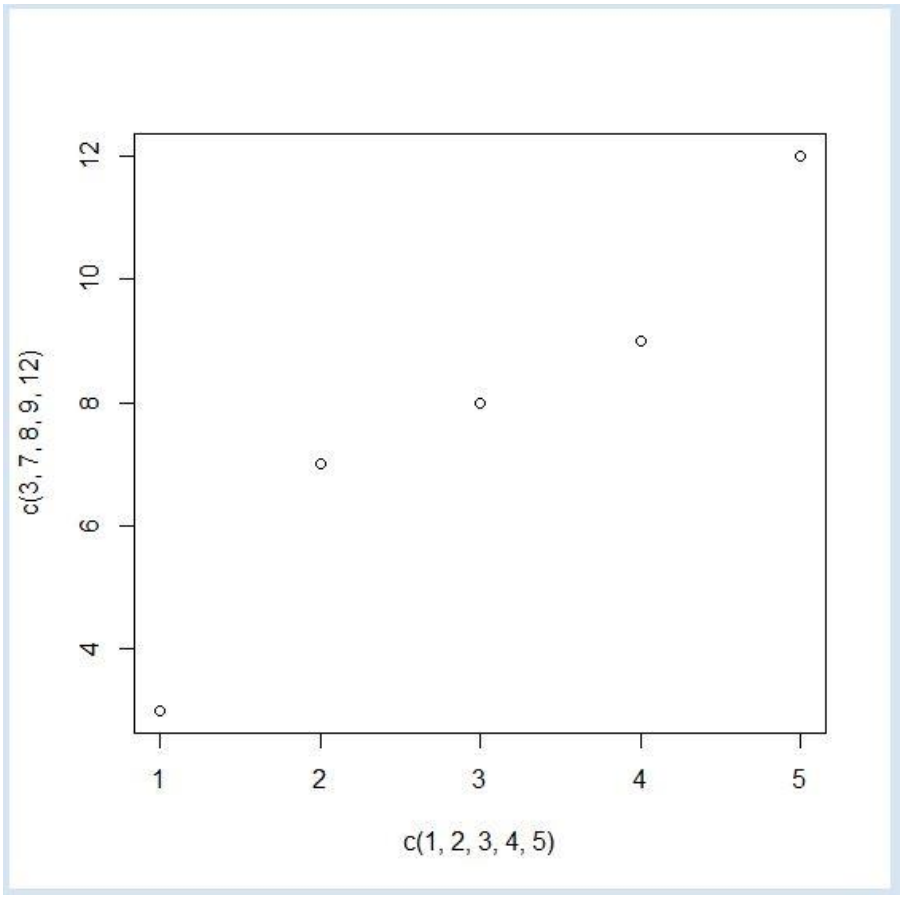
plot(1,3)



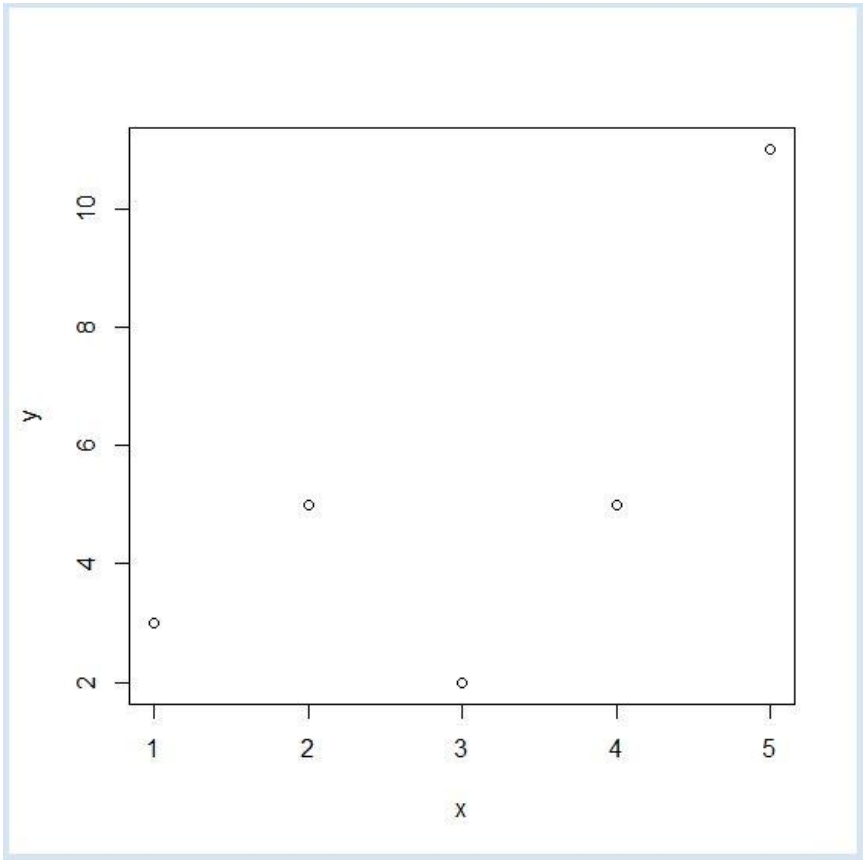
plot(c(1,8),c(3,10))



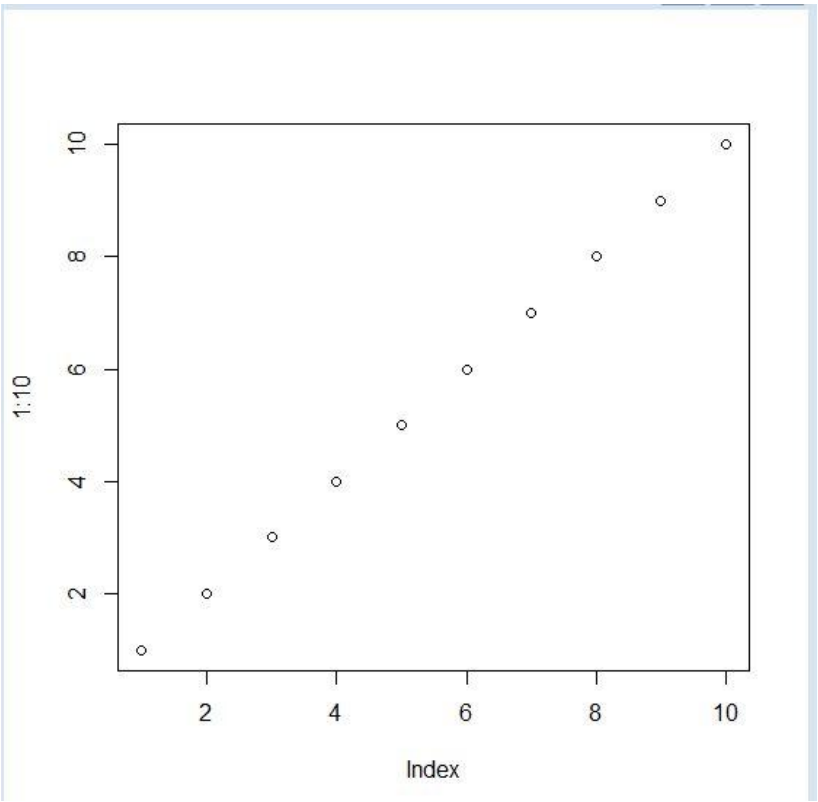
```
plot(c(1,2,3,4,5),c(3,7,8,9,12))
```



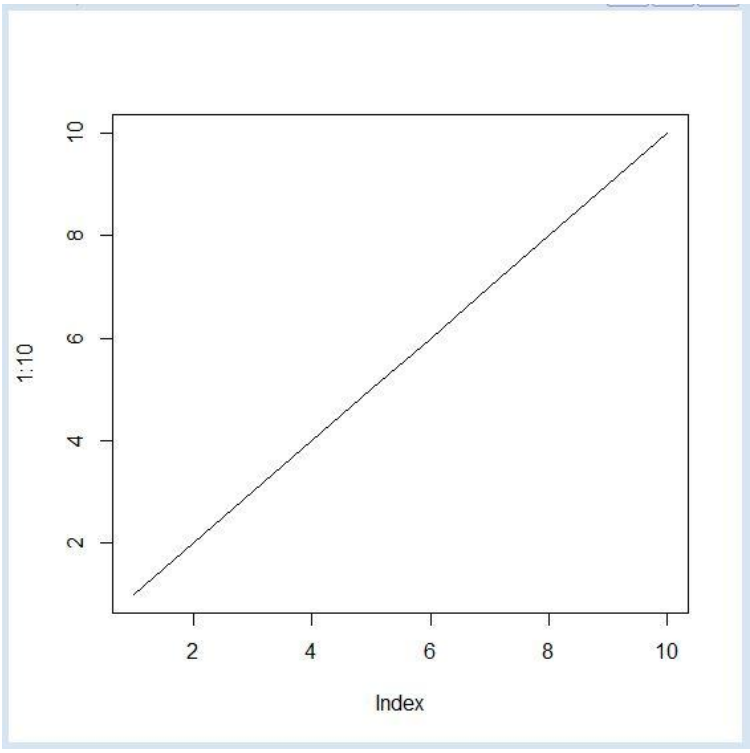

```
x <- c(1,2,3,4,5)

y <- c(3,5,2,5,11)

plot(x,y)
```



plot(1:10)

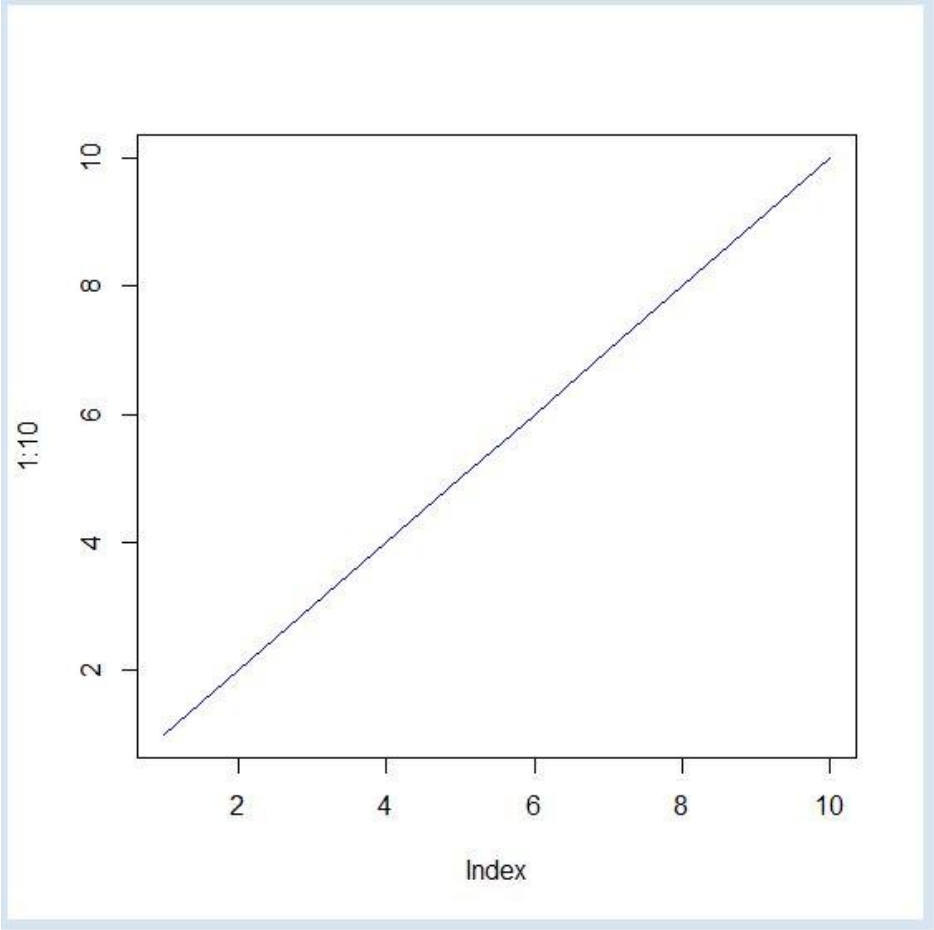


Create line chart in R
plot(v,type,col,xlab,tlab,main,lwd,lty)



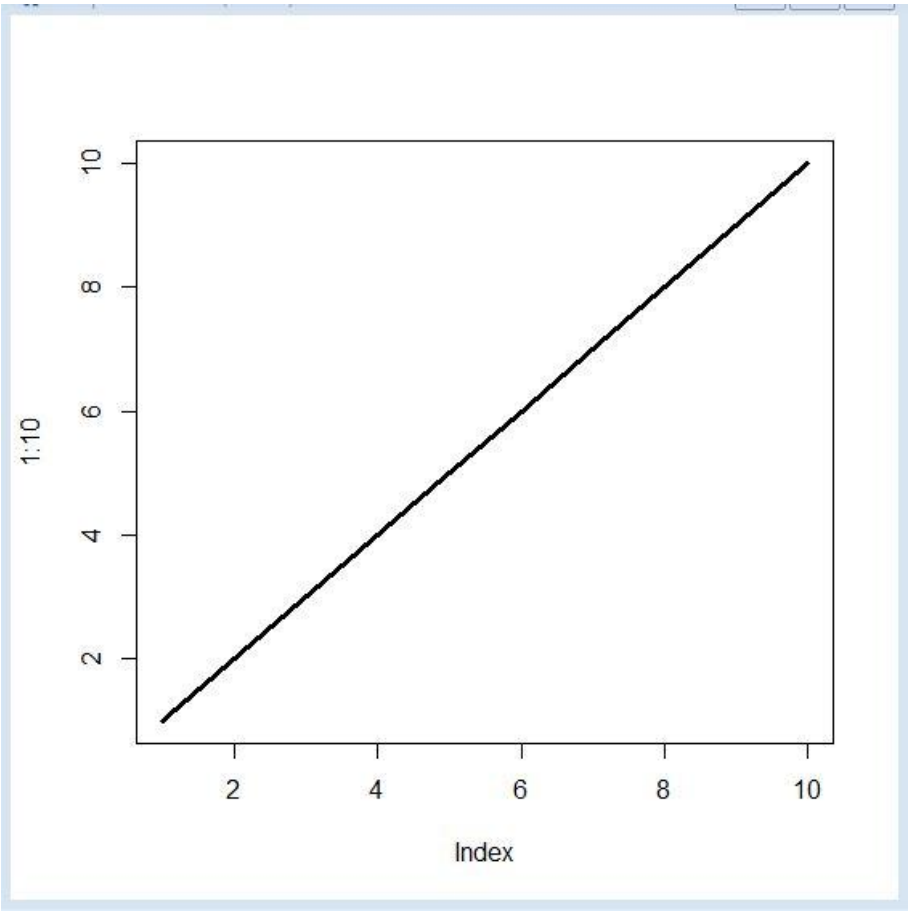
Line graph

plot(1:10,type="l")

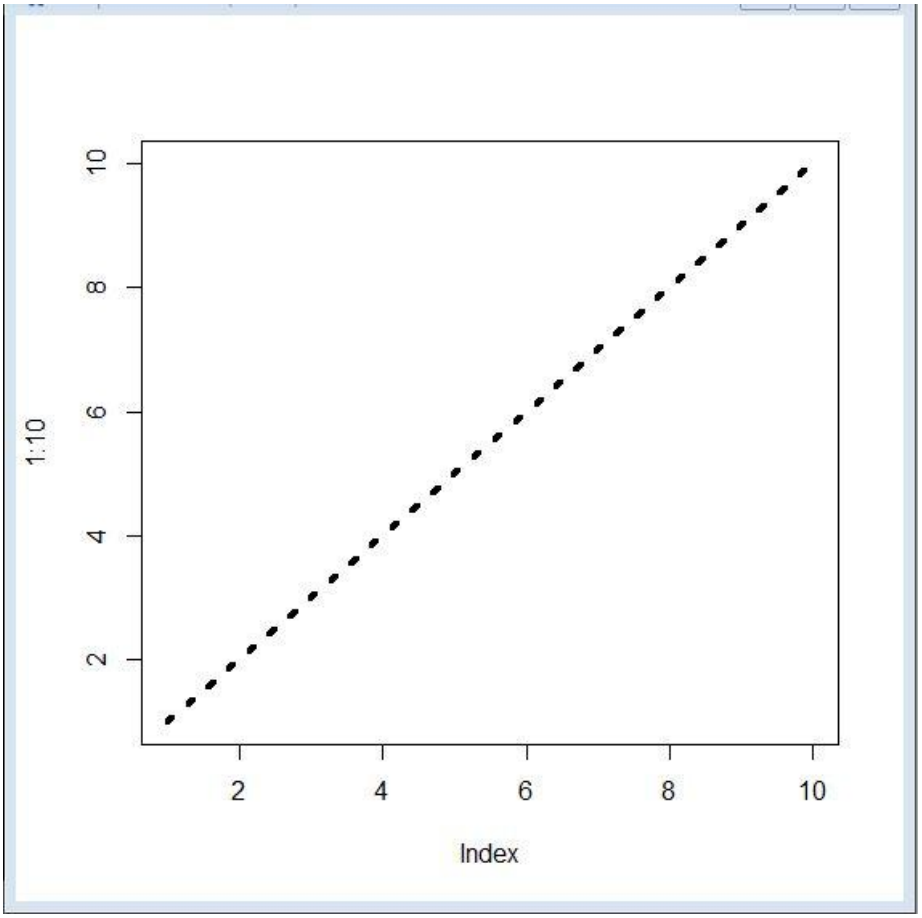


Line Graph Width

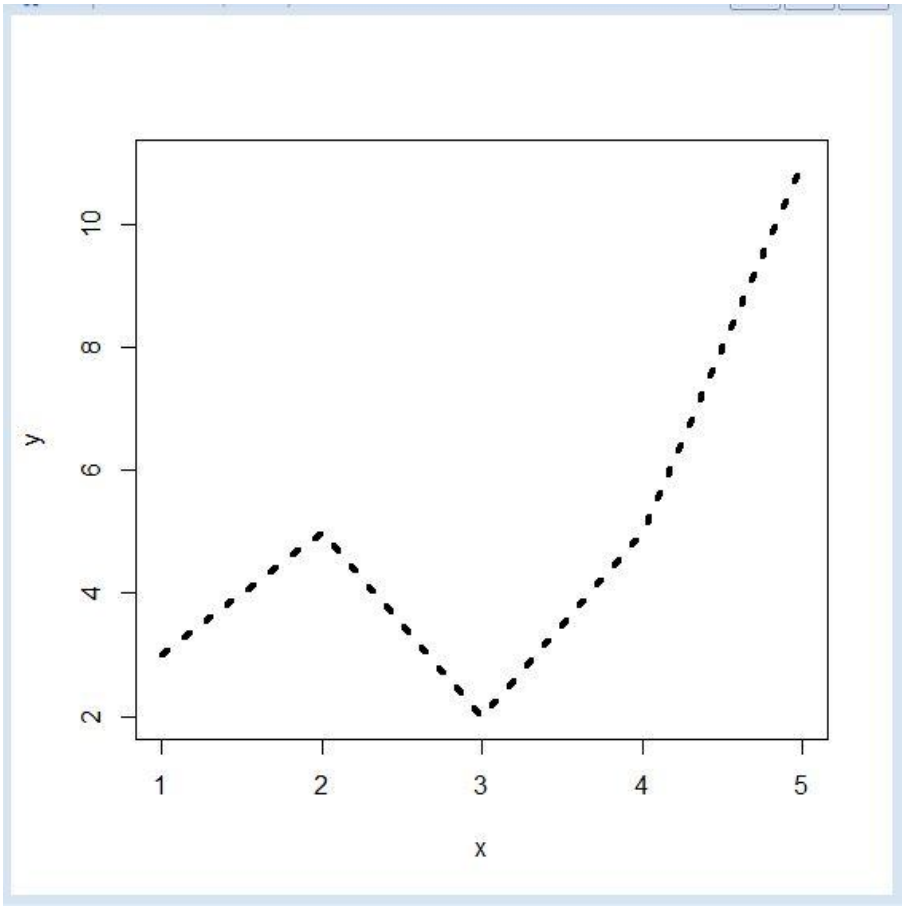
plot(1:10,type="l",lwd=4)



Line Graph Width and Color
plot(1:10,type="l",lwd=4,lty=3)

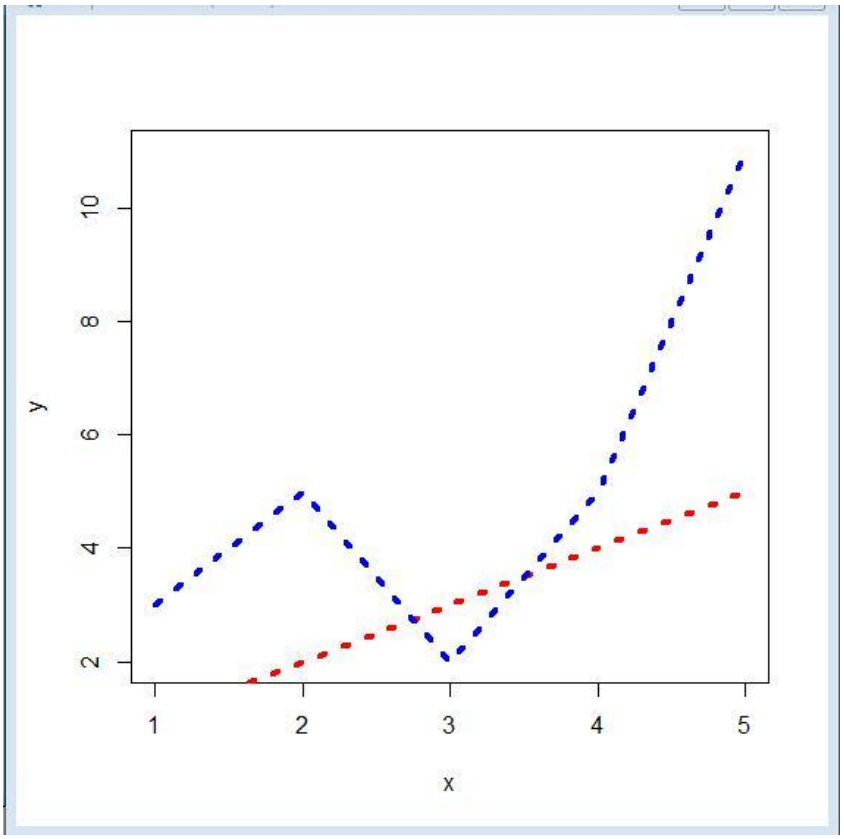


Line Graph Width and Color with data
plot(x,y,type="l",lwd=4,lty=3)



Line() function
lines(x,type="l",lwd=4,lty=3,col="Red")

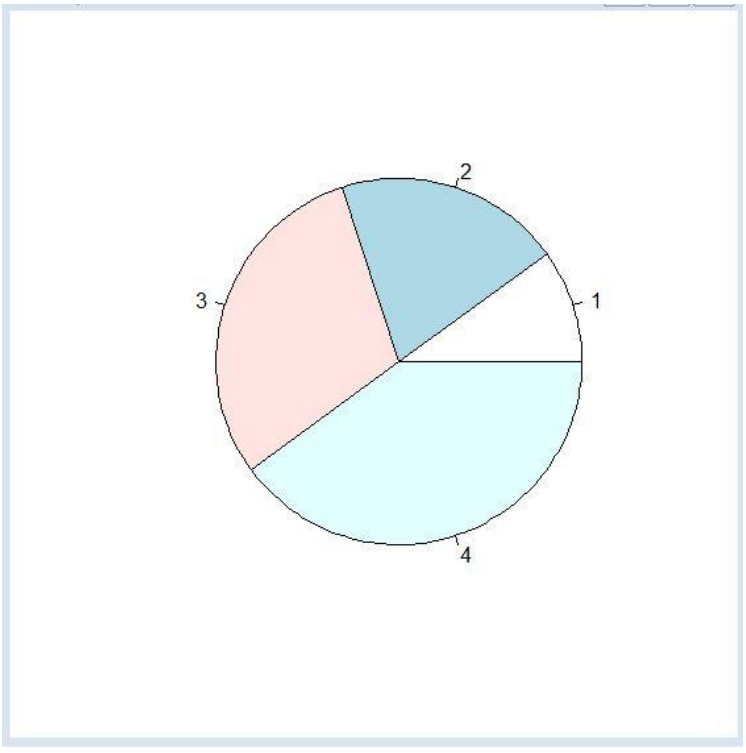
lines(y,type="l",lwd=4,lty=3,col="Red")



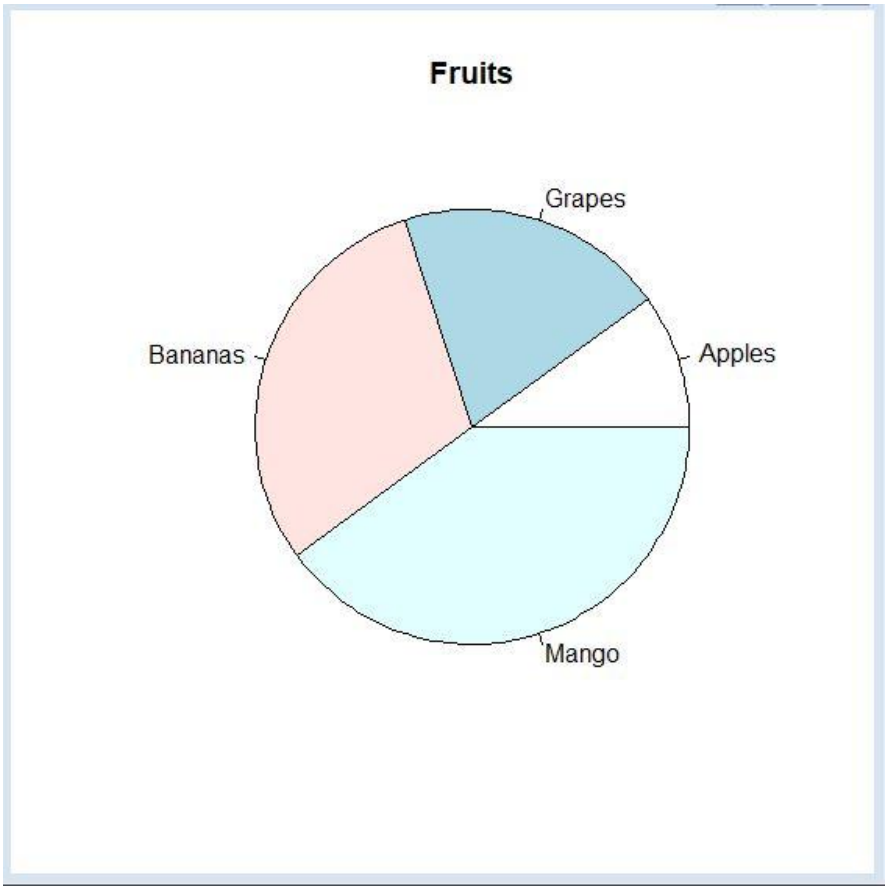
Pie Charts
pie(x,labels,radius,main,col,clockwise)

x <- c(10,20,30,40)

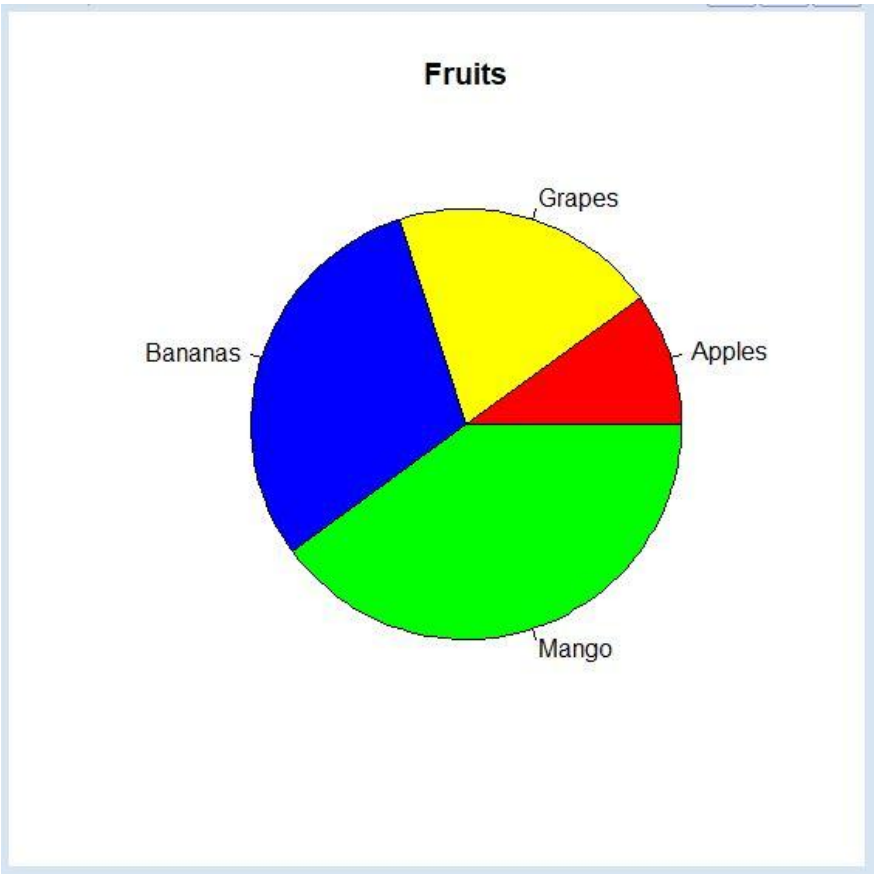
pie(x)



```
mLabel <- c("Apples","Grapes","Bananas","Mango")

pie(x,label=mLabel,main="Fruits")
```




```
colors <- c("red","yellow","blue","green")

pie(x,label=mLabel,main="Fruits",col=colors)
```



Q.1. Import employees.csv file and perform following -
1. Display the content.
2. Find the dimensions of the data in the above imported dataset.
3. Get all the people with designation “clerk”.
4. Get the people whose salary is greater than 55,000 and write the output in new excel file.
5. Summarize the above dataset

```
# Load the required library
library(readr)
install.packages("openxlsx")

# 1. Display the content
employees <- read.csv("employees.csv")
view(employees)
```

linearRegR.R | marketBasketAnalysis.R | Untitled1* | data

Filter

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
1	198	Donald	OConnell	DOCONNEL	650.507.9833	21-JUN-07	SH_CLERK	2600	-	124	50
2	199	Douglas	Grant	DGRANT	650.507.9844	13-JAN-08	SH_CLERK	2600	-	124	50
3	200	Jennifer	Whalen	JWHALEN	515.123.4444	17-SEP-03	AD_ASST	4400	-	101	10
4	201	Michael	Hartstein	MHARTSTE	515.123.5555	17-FEB-04	MK_MAN	13000	-	100	20
5	202	Pat	Fay	PFAY	603.123.6666	17-AUG-05	MK_REP	6000	-	201	20
6	203	Susan	Mavris	SMAVRIS	515.123.7777	07-JUN-02	HR_REP	6500	-	101	40
7	204	Hermann	Baer	HBAER	515.123.8888	07-JUN-02	PR_REP	10000	-	101	70
8	205	Shelley	Higgins	SHIGGINS	515.123.8080	07-JUN-02	AC_MGR	12008	-	101	110
9	206	William	Gietz	WGIETZ	515.123.8181	07-JUN-02	AC_ACCOUNT	8300	-	205	110
10	100	Steven	King	SKING	515.123.4567	17-JUN-03	AD_PRES	24000	-	-	90
11	101	Neena	Kochhar	NKOCHHAR	515.123.4568	21-SEP-05	AD_VP	17000	-	100	90
12	102	Lex	De Haan	LDEHAAN	515.123.4569	13-JAN-01	AD_VP	17000	-	100	90
13	103	Alexander	Hunold	AHUNOLD	590.423.4567	03-JAN-06	IT_PROG	9000	-	102	60
14	104	Bruce	Ernst	BERNST	590.423.4568	21-MAY-07	IT_PROG	6000	-	103	60
15	105	David	Austin	DAUSTIN	590.423.4569	25-JUN-05	IT_PROG	4800	-	103	60

Showing 1 to 16 of 50 entries, 11 total columns

```
# 2. Find the dimensions of the data
dimensions <- dim(employees)
cat("Dimensions of the dataset:", dimensions[1], "rows and", dimensions[2],
"columns\n")
```

Dimensions of the dataset: 50 rows and 11 columns

```
# 3. Get all the people with designation "clerk"
clerks <- subset(employees, JOB_ID == "PU_CLERK")
print("People with designation 'clerk':")
view(clerks)
```

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
25	115	Alexander	Khoo	AKHOO	515.127.4562	18-MAY-03	PU_CLERK	3100	-	114	30
26	116	Shelli	Baida	SBAIDA	515.127.4563	24-DEC-05	PU_CLERK	2900	-	114	30
27	117	Sigal	Tobias	STOBIAS	515.127.4564	24-JUL-05	PU_CLERK	2800	-	114	30
28	118	Guy	Himuro	GHIMURO	515.127.4565	15-NOV-06	PU_CLERK	2600	-	114	30
29	119	Karen	Colmenares	KCOLMENA	515.127.4566	10-AUG-07	PU_CLERK	2500	-	114	30

```
# 4. Get people whose salary is greater than 55,000 and write the output to a new
Excel file
high_salary_employees <- subset(employees, SALARY > 3000)
write.xlsx(high_salary_employees, "high_salary_employees.xlsx", row.names =
FALSE)
```

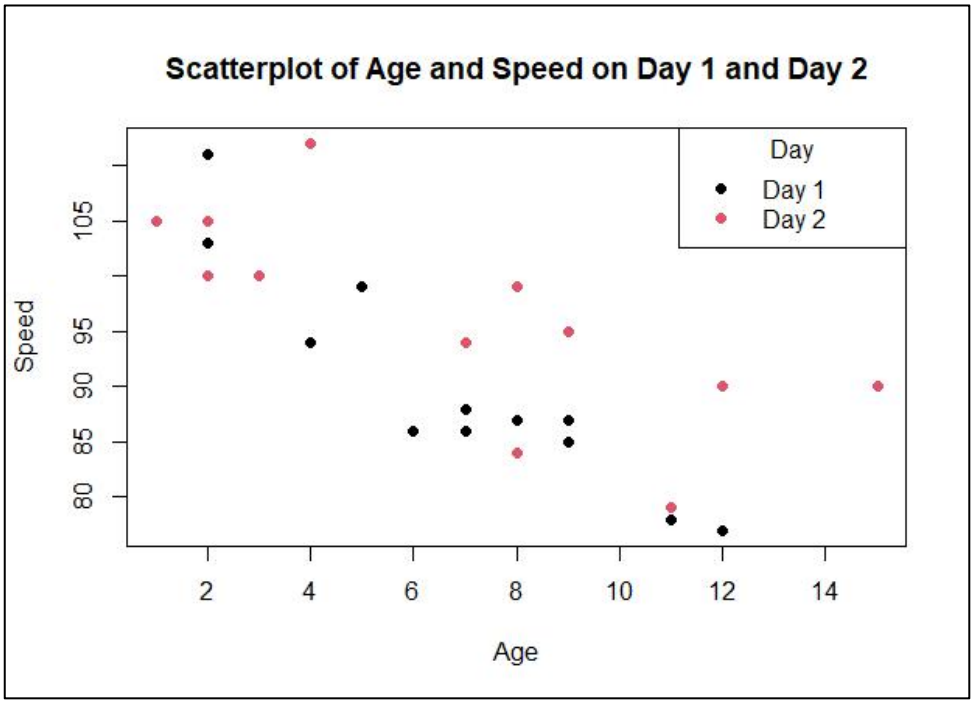
EMPLOYEE	FIRST_NAM	LAST_NAM	EMAIL	PHONE_NI	HIRE_DATI	JOB_ID	SALARY	COMMISSI	MANAGER	DEPARTME
200	Jennifer	Whalen	JWHALEN	515.123.44	17-SEP-03	AD_ASST	4400	-	101	10
201	Michael	Hartstein	MHARTSTE	515.123.55	17-FEB-04	MK_MAN	13000	-	100	20
202	Pat	Fay	PFAY	603.123.66	17-AUG-05	MK_REP	6000	-	201	20
203	Susan	Mavris	SMAVRIS	515.123.77	07-JUN-02	HR_REP	6500	-	101	40
204	Hermann	Baer	HBAER	515.123.88	07-JUN-02	PR_REP	10000	-	101	70
205	Shelley	Higgins	SHIGGINS	515.123.80	07-JUN-02	AC_MGR	12008	-	101	110
206	William	Gietz	WGIETZ	515.123.81	07-JUN-02	AC_ACCOU	8300	-	205	110
100	Steven	King	SKING	515.123.45	17-JUN-03	AD_PRES	24000	-	-	90
101	Neena	Kochhar	NKOCHHA	515.123.45	21-SEP-05	AD_VP	17000	-	100	90
102	Lex	De Haan	LDEHAAN	515.123.45	13-JAN-01	AD_VP	17000	-	100	90
103	Alexander	Hunold	AHUNOLD	590.423.45	03-JAN-06	IT_PROG	9000	-	102	60
104	Bruce	Ernst	BERNST	590.423.45	21-MAY-07	IT_PROG	6000	-	103	60
105	David	Austin	DAUSTIN	590.423.45	25-JUN-05	IT_PROG	4800	-	103	60
106	Valli	Pataballa	VPATABAL	590.423.45	05-FEB-06	IT_PROG	4800	-	103	60
107	Diana	Lorentz	DLORENTZ	590.423.55	07-FEB-07	IT_PROG	4200	-	103	60
108	Nancy	Greenberg	NGREENBE	515.124.45	17-AUG-02	FI_MGR	12008	-	101	100
109	Daniel	Faviet	DFAVIET	515.124.41	16-AUG-02	FI_ACCOU	9000	-	108	100

```
# 5. Summarize the dataset
summary_data <- summary(employees)
print("Summary of the dataset:")
print(summary_data)
```

```
 EMPLOYEE_ID    FIRST_NAME         LAST_NAME          EMAIL              PHON
E_NUMBER         HIRE_DATE           JOB_ID
 Min.   :100.0   Length:50          Length:50          Length:50          Length:50          Len
gth:50           Length:50
 1st Qu.:112.2   Class :character   Class :character   Class :character   Class :character
 Class :character   Class :character
 Median :124.5   Mode  :character   Mode  :character   Mode  :character   Mode  :char
acter   Mode  :character   Mode  :character
 Mean   :134.8

 3rd Qu.:136.8

 Max.   :206.0

    SALARY       COMMISSION_PCT     MANAGER_ID      DEPARTMENT_ID
 Min.   : 2100   Length:50          Length:50          Min.   : 10.0
 1st Qu.: 2725   Class :character   Class :character   1st Qu.: 50.0
 Median : 4600   Mode  :character   Mode  :character   Median : 50.0
 Mean   : 6182                                         Mean   : 57.6
 3rd Qu.: 8150                                         3rd Qu.: 60.0
 Max.   :24000                                         Max.   :110.0
```

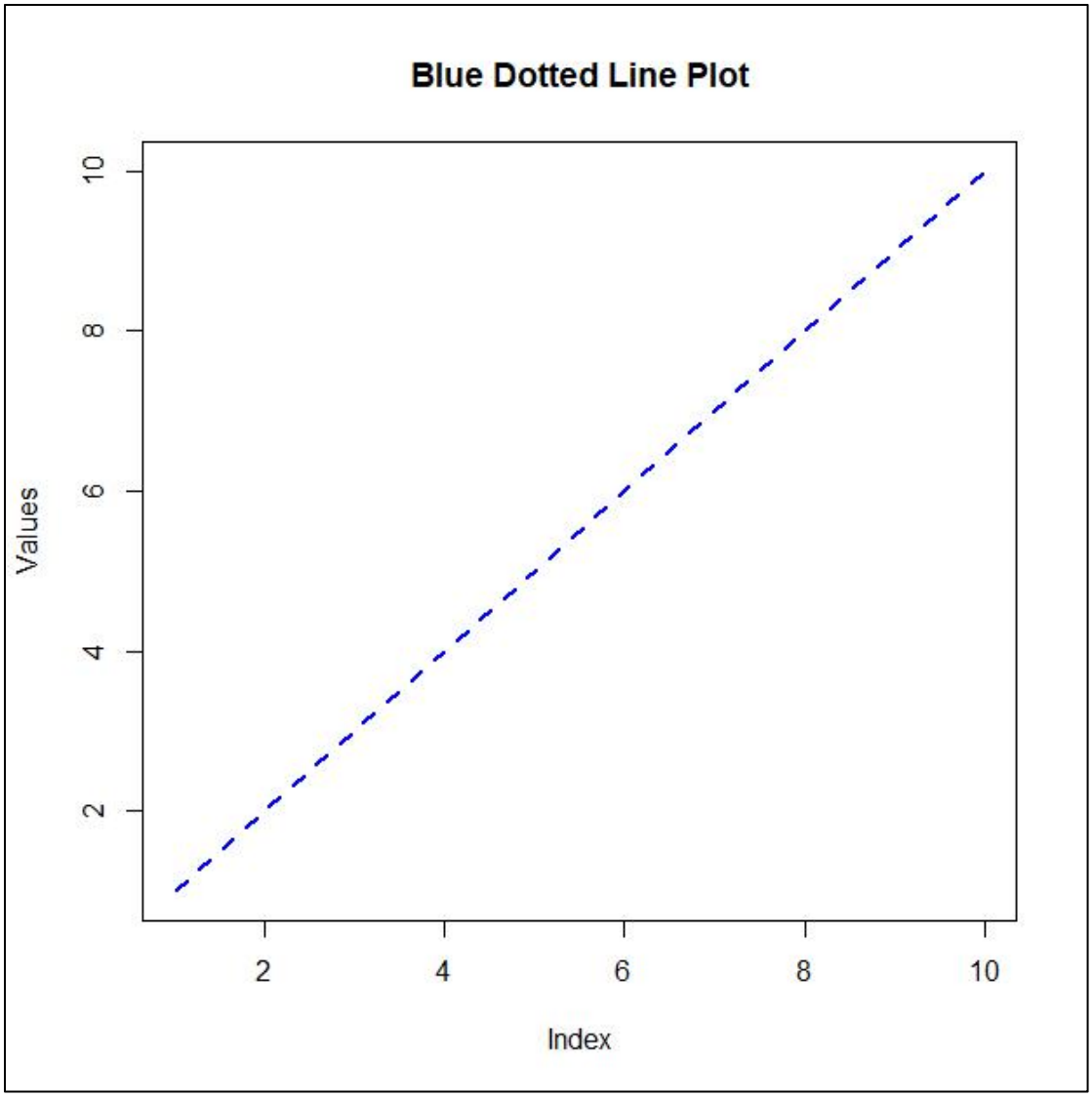
Q.2 The age and speed of 12 cars observed on day 1 are age1(5,7,8,7,2,2,9,4,11,12,9,6), speed1(99,86,87,88,111,103,87,94,78,77,85,86) and on day 2 following values are observed age2(2,2,8,1,15,8,12,9,7,3,11,4,7,14,12), speed2(100,105,84,105,90,99,90,95,94,100,79,112,91,80,85). Write a R program to draw a scatterplot that compares observations of the two days.

```
# Age and speed data for day 1
age1 <- c(5, 7, 8, 7, 2, 2, 9, 4, 11, 12, 9, 6)
speed1 <- c(99, 86, 87, 88, 111, 103, 87, 94, 78, 77, 85, 86)
# Age and speed data for day 2
age2 <- c(2, 2, 8, 1, 15, 8, 12, 9, 7, 3, 11, 4, 7, 14, 12)
speed2 <- c(100, 105, 84, 105, 90, 99, 90, 95, 94, 100, 79, 112, 91, 80, 85)
# Make sure the vectors have the same length
min_length <- min(length(age1), length(speed1), length(age2), length(speed2))
# Create a data frame
data <- data.frame(
  Day = rep(c("Day 1", "Day 2"), each = min_length),
  Age = c(age1[1:min_length], age2[1:min_length]),
  Speed = c(speed1[1:min_length], speed2[1:min_length])
)
# Draw a scatterplot
plot(
  Speed ~ Age,          # y-axis ~ x-axis
  data = data,          # Data frame
  col = as.factor(data$Day),  # Color points by Day
  pch = 16,             # Use solid circles as points
  main = "Scatterplot of Age and Speed on Day 1 and Day 2",
  xlab = "Age",
  ylab = "Speed"
)
# Add a legend
legend("topright", legend = levels(as.factor(data$Day)), col = 1:2, pch = 16, title =
"Day")
```



Q.3 Write a R program to create a vector with numerical values in a sequence from 1 to 10 and draw a blue colored dotted line of width 2 for the above vector.

```
# Create a vector with numerical values from 1 to 10
my_vector <- 1:10

# Create a plot with a blue-colored dotted line of width 2
plot(
  my_vector,              # x-axis values
  type = "l",             # "l" indicates a line plot
  col = "blue",           # Line color is blue
  lty = 2,                # Dotted line style
  lwd = 2,                # Line width is 2
  main = "Blue Dotted Line Plot",
  xlab = "Index",
  ylab = "Values"
)
```



Q.4 Write a R program to read the excel file "input.xlsx" and perform following
1. Display the content.
2. Find the dimensions of the data in the above imported dataset.
3. Get all the people working in IT department
4. Get the people who joined on or after 2014 and write the output in new excel file.
5. Summarize the above dataset

```
# 1. Display the content
input_data <- read_excel("input.xlsx")
print("Content of the dataset:")
view(input_data)
```

	id	name	salary	start_date	dept
1	1	Rick	623.30	40909	IT
2	2	Dan	515.20	41540	Operations
3	3	Michelle	611.00	41958	IT
4	4	Ryan	729.00	41770	HR
5	5	Gary	843.25	42090	Finance
6	6	Nina	578.00	41415	IT
7	7	Simon	632.80	41485	Operations
8	8	Guru	722.50	41807	Finance

```
# 2. Find the dimensions of the data
dimensions <- dim(input_data)
cat("Dimensions of the dataset:", dimensions[1], "rows and", dimensions[2],
"columns\n")
```

Dimensions of the dataset: 8 rows and 5 columns

```
# 3. Get all the people working in IT department
it_department <- subset(input_data, Department == "IT")
print("People working in IT department:")
view(it_department)
```

	id	name	salary	start_date	dept
1	1	Rick	623.3	40909	IT
3	3	Michelle	611.0	41958	IT
6	6	Nina	578.0	41415	IT

# 4. Get people who joined on or after 2014 and write the output to a new Excel file
joined_after_2014 <- subset(input_data, Joining_Year >= 2014)
write.xlsx(joined_after_2014, "joined_after_2014_output.xlsx", row.names = FALSE)

	A	B	C	D	E
1	id	name	salary	start_date	dept
2	1	Rick	623.3	40909	IT
3	2	Dan	515.2	41540	Operations
4	3	Michelle	611	41958	IT
5	4	Ryan	729	41770	HR
6	5	Gary	843.25	42090	Finance
7	6	Nina	578	41415	IT
8	7	Simon	632.8	41485	Operations
9	8	Guru	722.5	41807	Finance
10					

# 5. Summarize the dataset
summary_data <- summary(input_data)
print("Summary of the dataset:")
print(summary_data)

```
> summary_data <- summary(input_data)
> print("Summary of the dataset:")
[1] "Summary of the dataset:"
> print(summary_data)
       id          name              salary        start_date        dept
 Min.   :1.00   Length:8           Min.   :515.2   Min.   :40909   Length:8
 1st Qu.:2.75   Class :character   1st Qu.:602.8   1st Qu.:41468   Class :character
 Median :4.50   Mode  :character   Median :628.0   Median :41655   Mode  :character
 Mean   :4.50                      Mean   :656.9   Mean   :41622
 3rd Qu.:6.25                      3rd Qu.:724.1   3rd Qu.:41845
 Max.   :8.00                      Max.   :843.2   Max.   :42090
> |
```