Statistical Graphics

Graphic Representation of Data

- Bar Plot

- Histogram

- Scatter Plot

- Density Plot

- Box Plot

Bar Plot

- Bar chart can be drawn with the help of function **barplot**()

Syntax : barplot(height, horiz = FALSE, main = NULL, sub = NULL, xlab = NULL, ylab = NULL,...)
Where   height : A vector / matrix
                        horiz : logical; If TRUE then graph results in
                                horizontal bar graph
                        main : Title of Graph
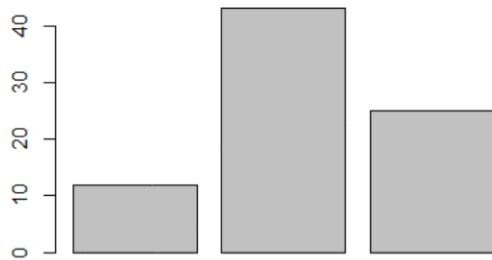   xlab : X-axis label
                        ylab : Y-axis label

Bar Plot Examples
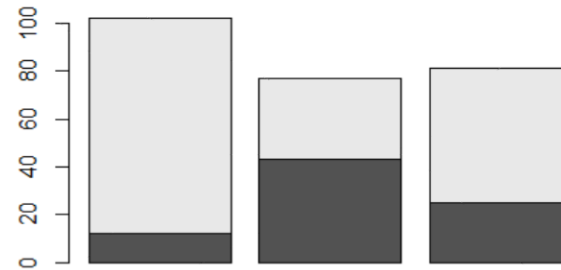
```
> ProductA <- c(12,43,25)
> barplot(ProductA)
```



```
> ProductA <- c(12,43,25)
> ProductB <- c(90,34,56)
> rb <- rbind(ProductA,ProductB)
> barplot(rb)
```

Bar Plot Examples

```
> barplot(table(Gender),col = "brown",main = "Gender-Wise Distribution")
```


Gender-Wise Distribution

Histogram

- Bar Plot on binned data can be said to be a histogram
- Difference between histogram and bar plot is that in bar plot we are free to rearrange the bars whereas in histogram we are not
- We can know the distribution of any variable with histogram

Syntax : hist(x, breaks = "", include.lowest = TRUE, main = paste("Histogram of" , xname), xlim = range(breaks), xlab = xname, ylab, ...)
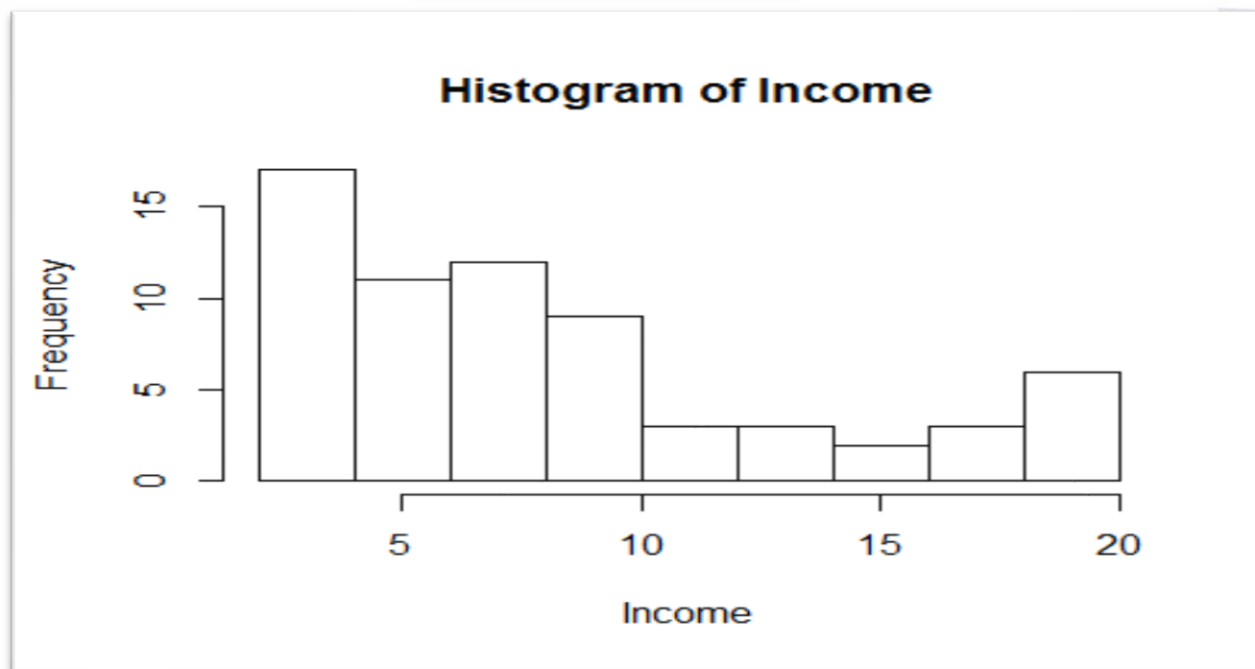
Where x : numeric vector or a function which can generate a numeric
vector
breaks : specifying the number of breaks or a numeric vector of breaks
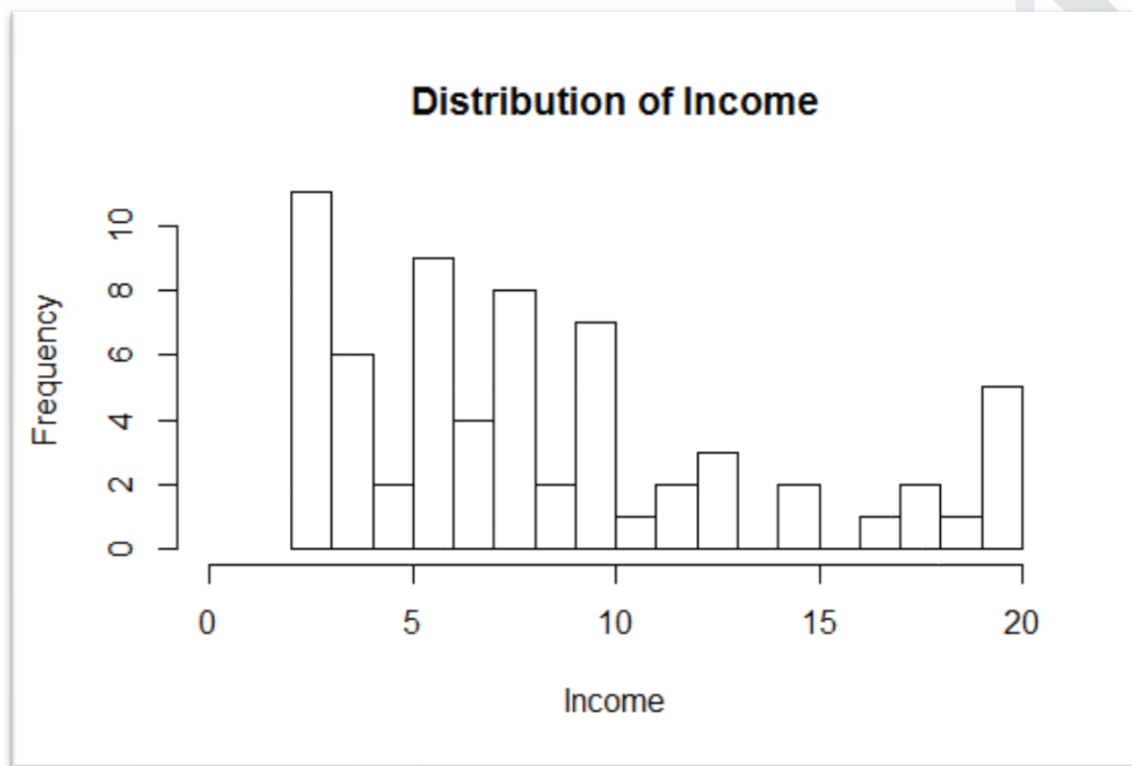xlim : Scale of limits on X axis

Histogram Examples

Histogram Examples

```
> hist(Income,breaks = 20,xlab="Income",
+        main="Distribution of Income", xlim = c(0,20))
```


Distribution of Income

Scatter Plot

- Scatter Plot can be understood as plot of points on 2-dimensional space of XY-plane
- Scatter plot gives us the information about correlation between the two variables plotted

Syntax : plot(x, y, type, col, pch...)

Where x , y : numeric vector or a function generating numerical vector

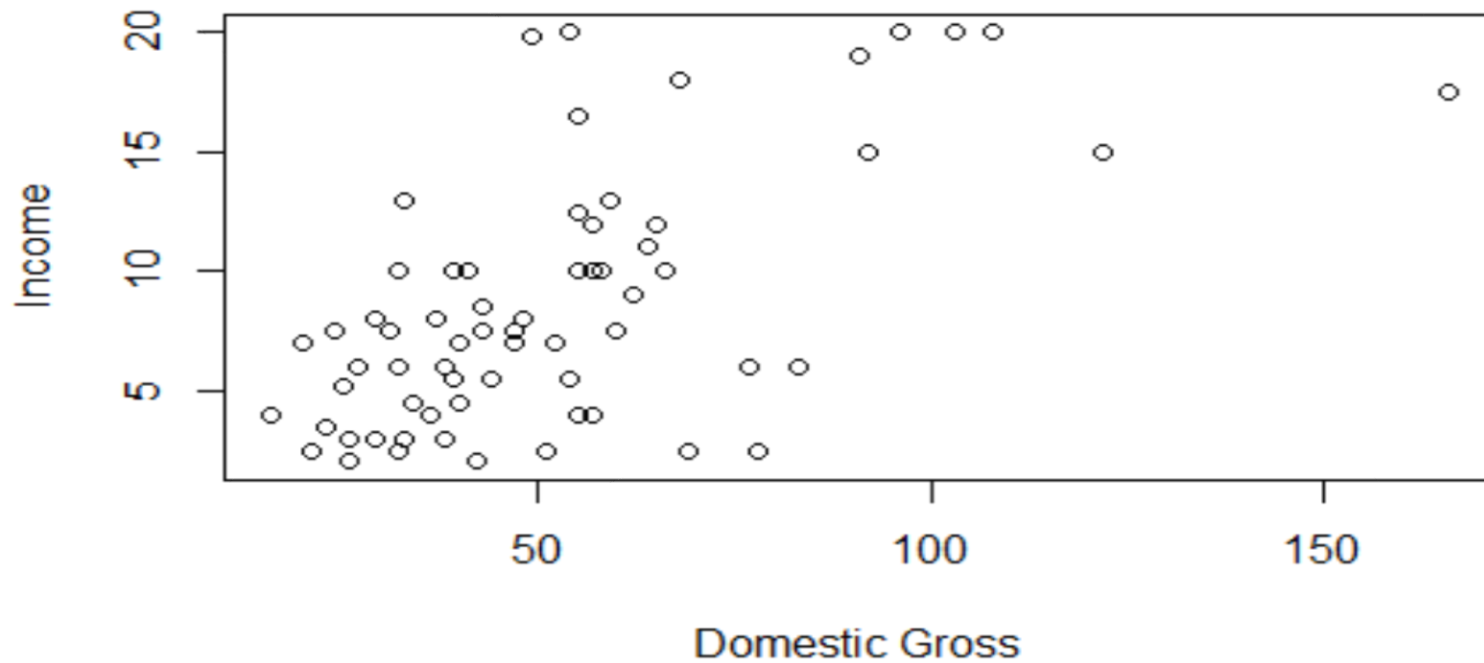   type : p for points, l for lines, b for both, s for stair   steps etc.

      col : Colour of plotting character
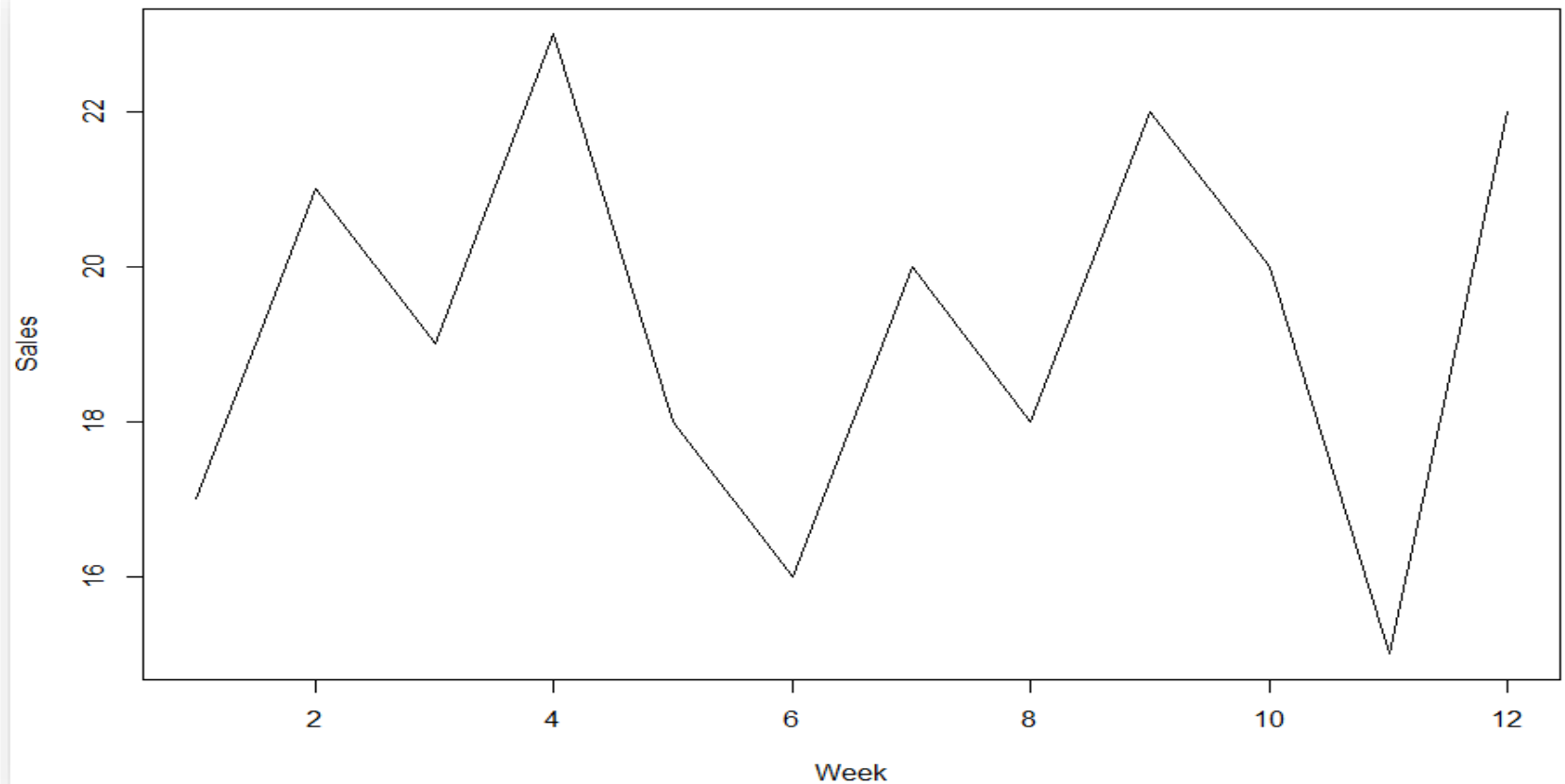
      pch : Plotting character

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
□	○	△	+	✕	◇	▽	⊠	✳	⬙	⊕	⊠	⊞	⊠	⊡	■	●	▲	◆	●	•	○	▢	◇	△	▽

Scatter Plot Examples

```
> plot(Domestic.Gross,Income,xlab="Domestic Gross",ylab="Income")
```
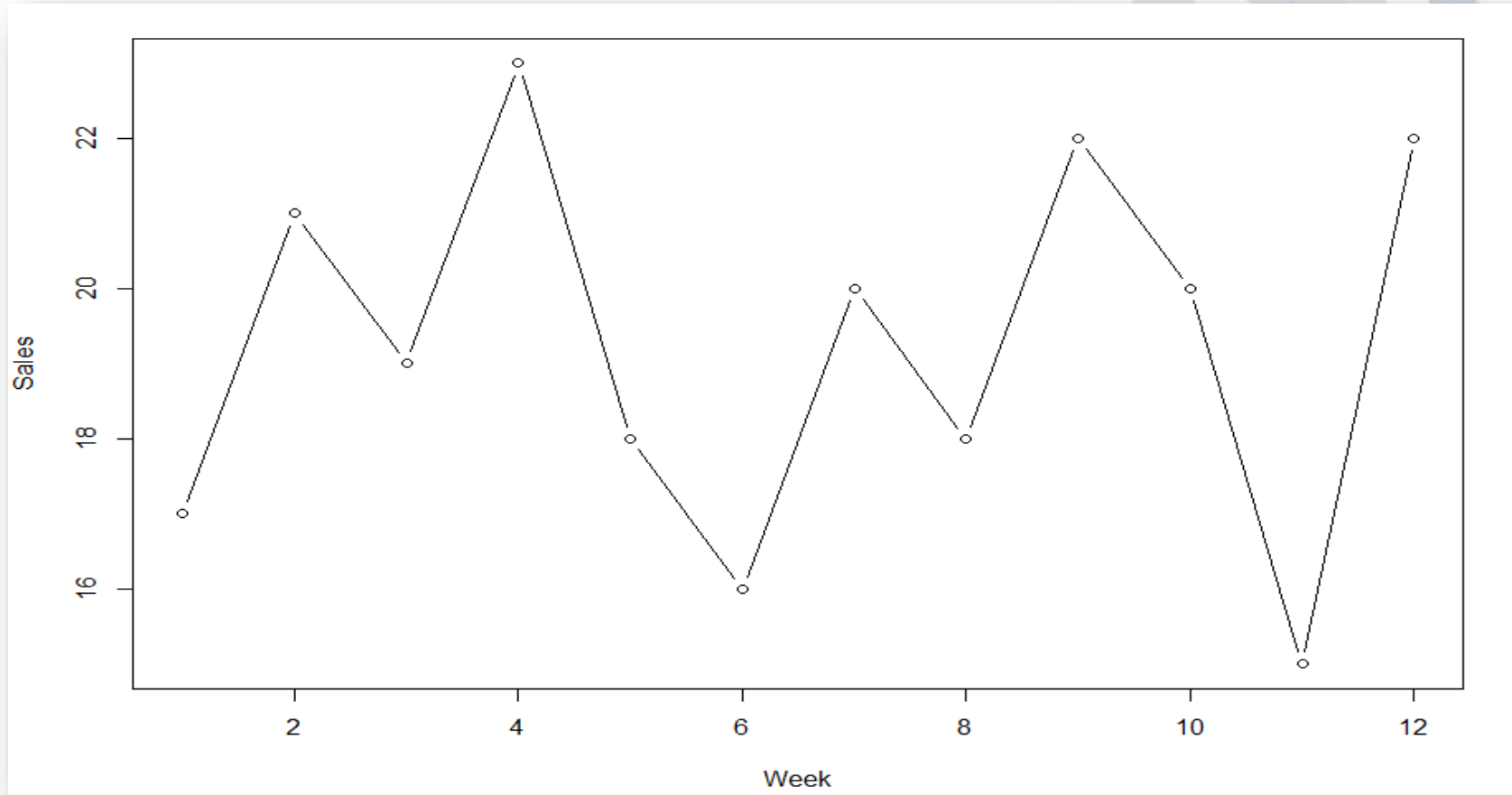
Line Plot using plot()

```
plot(gasoline$Week,gasoline$Sales,xlab = "Week",ylab = "Sales",type = 'l')
```
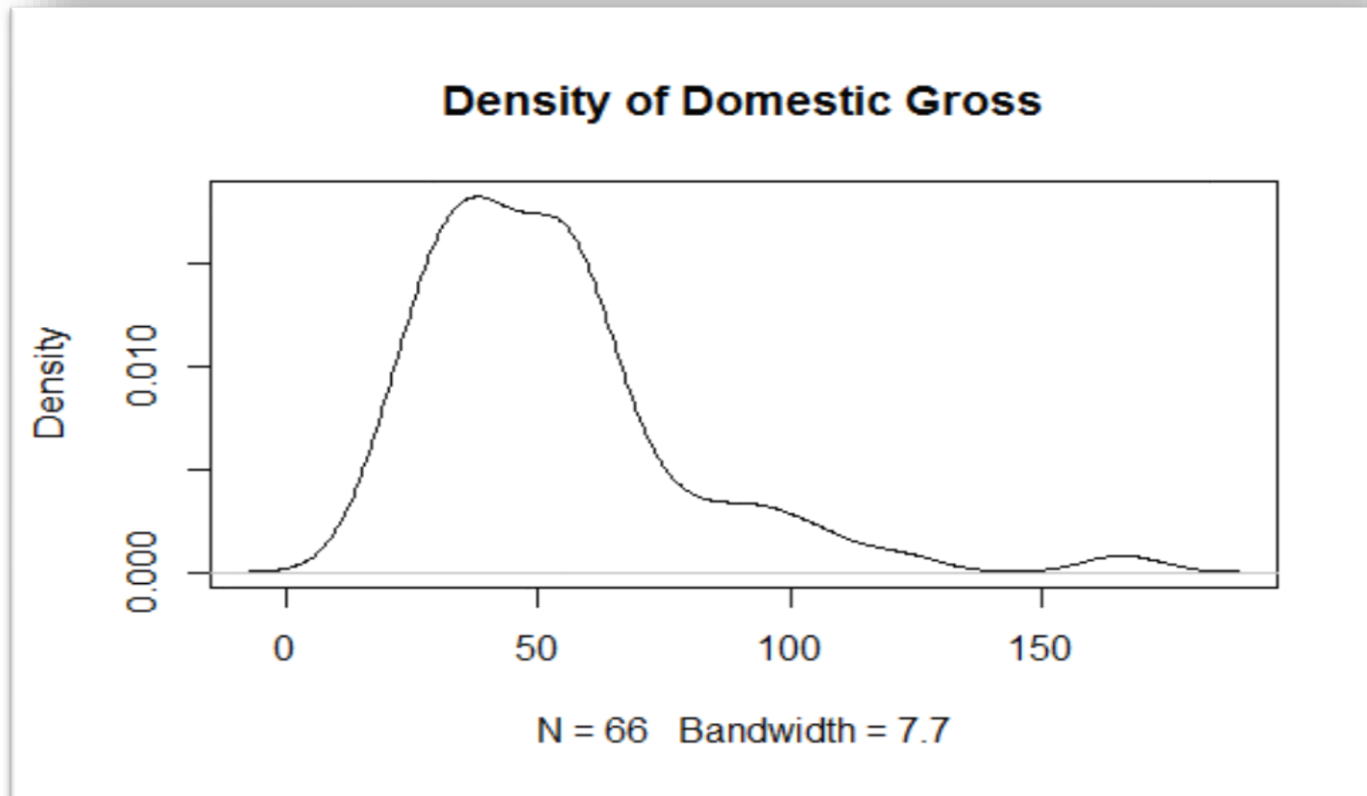
Line graph using plot()

```
plot(gasoline$Week,gasoline$Sales,xlab = "Week",ylab = "Sales",type = 'b')
```

Density Plot

- Density plot is smoothened form of histogram
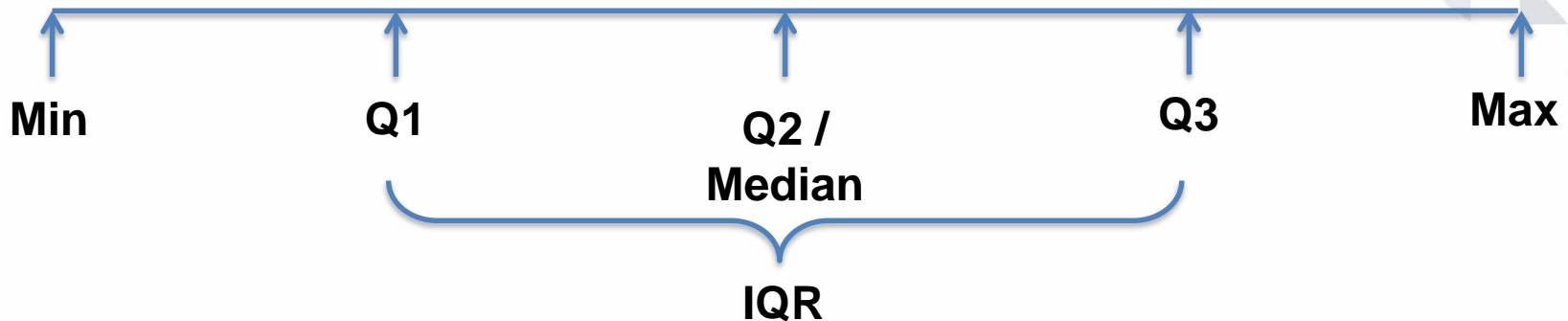- Density plot can be plotted with function plot() by calling density function

```
> plot(density(Domestic.Gross),main="Density of Domestic Gross")
```


Density of Domestic Gross

N = 66   Bandwidth = 7.7

Boxplot

- Before we understand box plot we need to understand the quartiles
- Quartiles divide the given data into four equal parts.

Data represented as a line

| **Min** | **Q1** | **Q2 /
Median** | **Q3** | **Max** |

IQR

- Inter-quartile range (IQR) is given by the formula:

$$IQR = Q3 - Q1$$

Describing with Box Plots

Here Outliers are the observations which are 1.5*IQR from the edge of the box
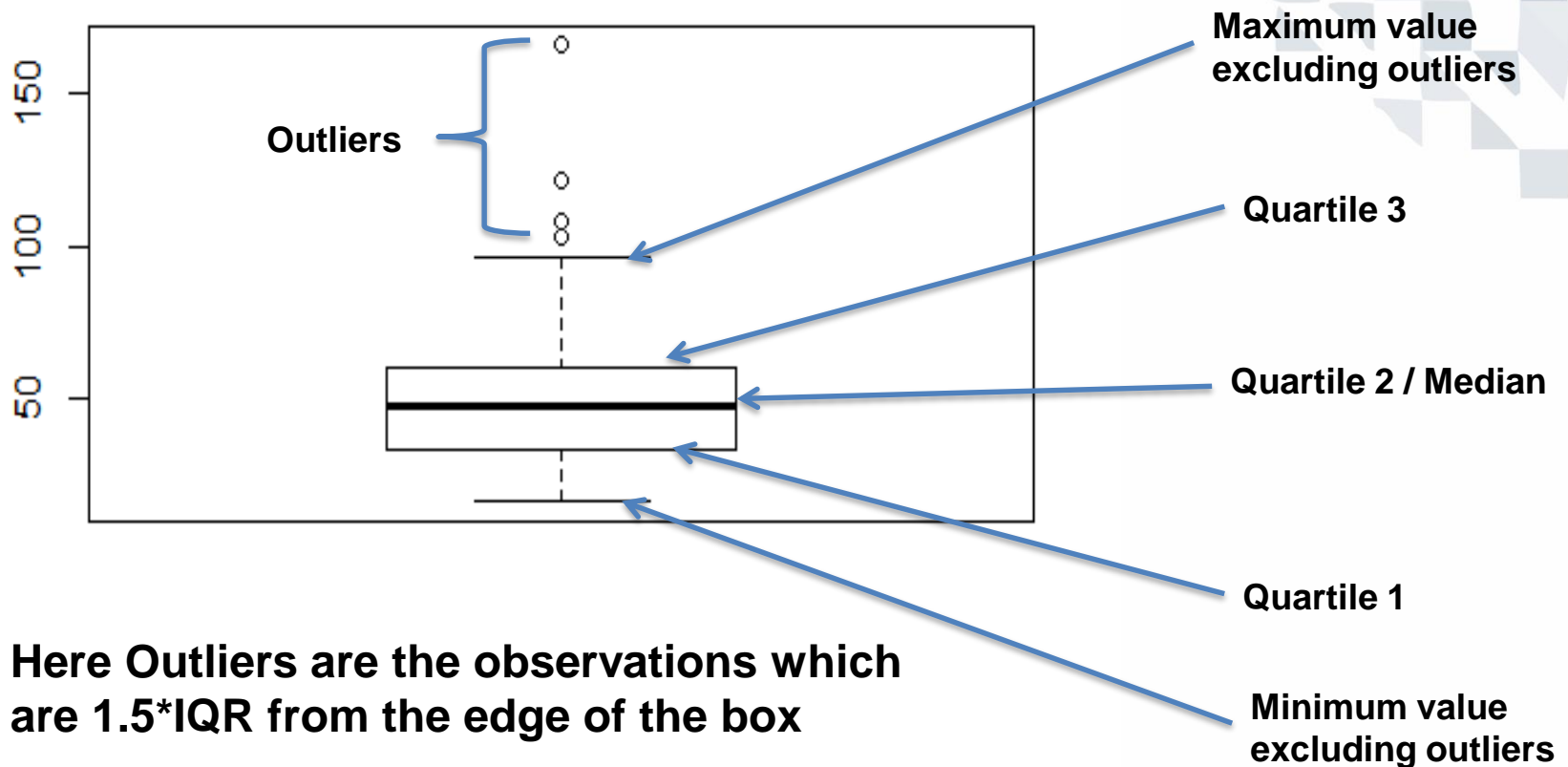
Boxplot

- Boxplot gives us the spread of the data

Syntax : boxplot(x)

Where x : numeric vector

OR

$$boxplot(y \sim x)$$

Where y : numeric vector

x : factor

Boxplot Examples

Boxplot Examples

```
> boxplot(Domestic.Gross,horizontal = TRUE,main="Domestic Gross")
```


Domestic Gross
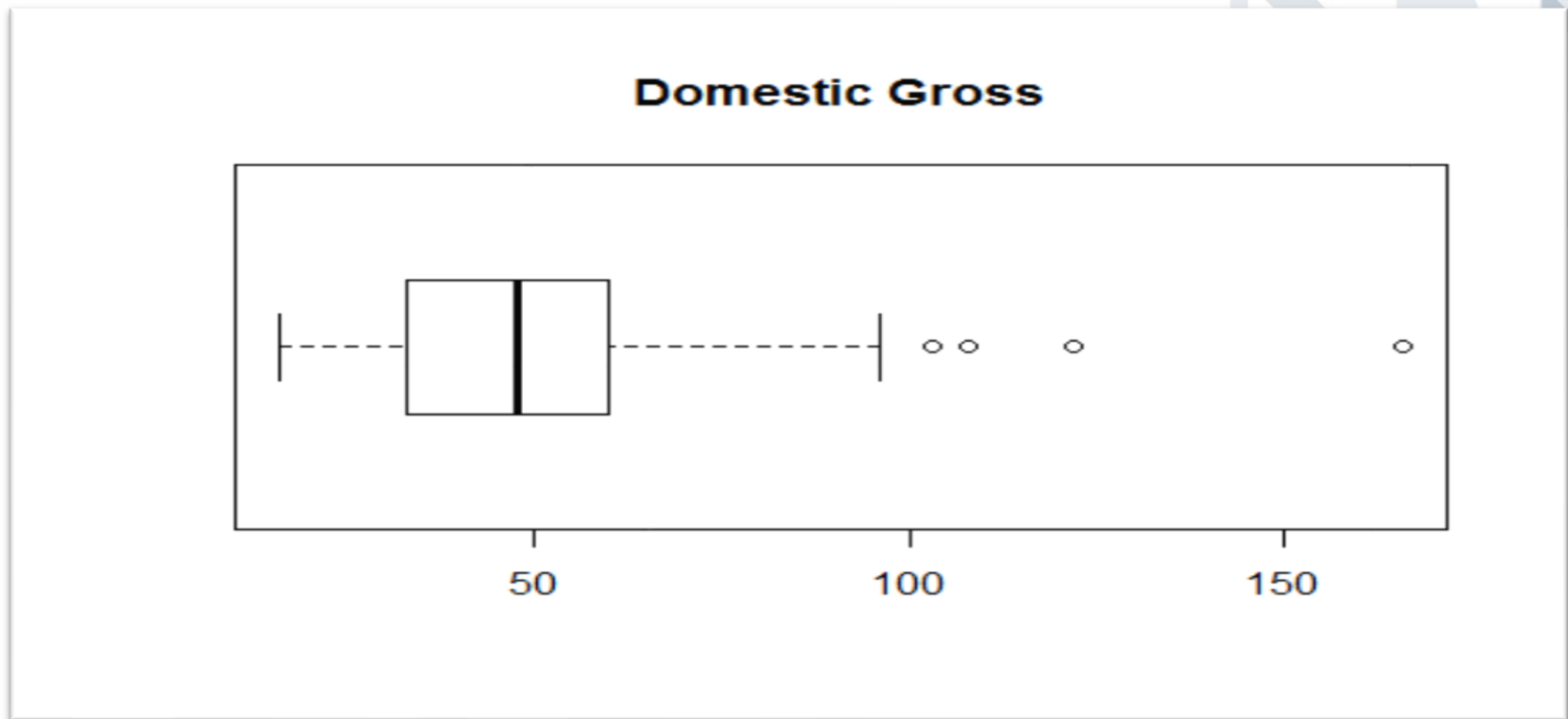
Displaying multiple graphs

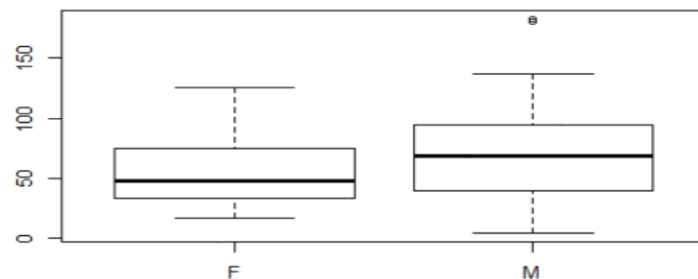- For displaying multiple graphs on the same canvas par() function can be used
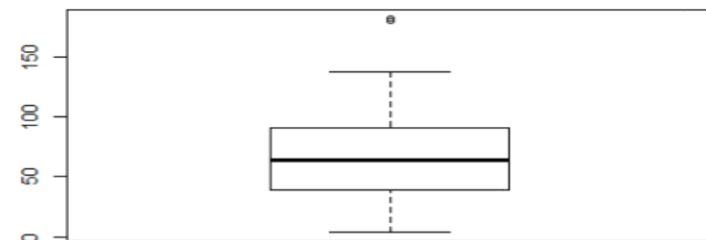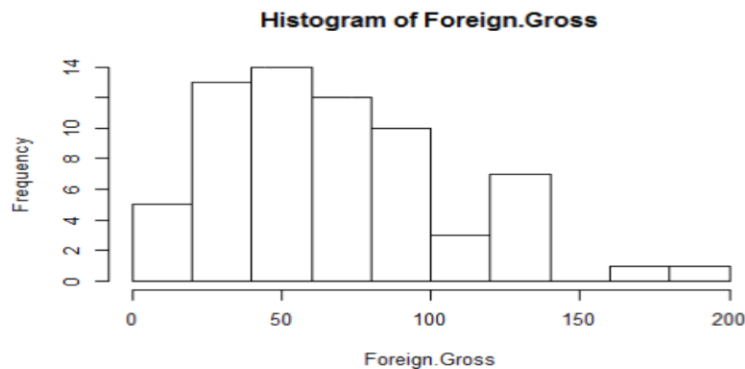
Syntax : par(mfrow=c(r,c))

Where r : no. of rows on the canvas

c : no. of columns on the canvas

par() examples

```
> par(mfrow=c(2,2))
> hist(Foreign.Gross)
> boxplot(Foreign.Gross)
> boxplot(Foreign.Gross ~ Gender)
> plot(density(Foreign.Gross))
```
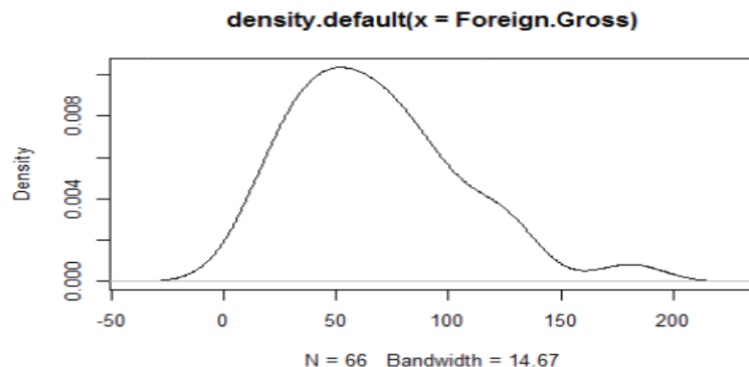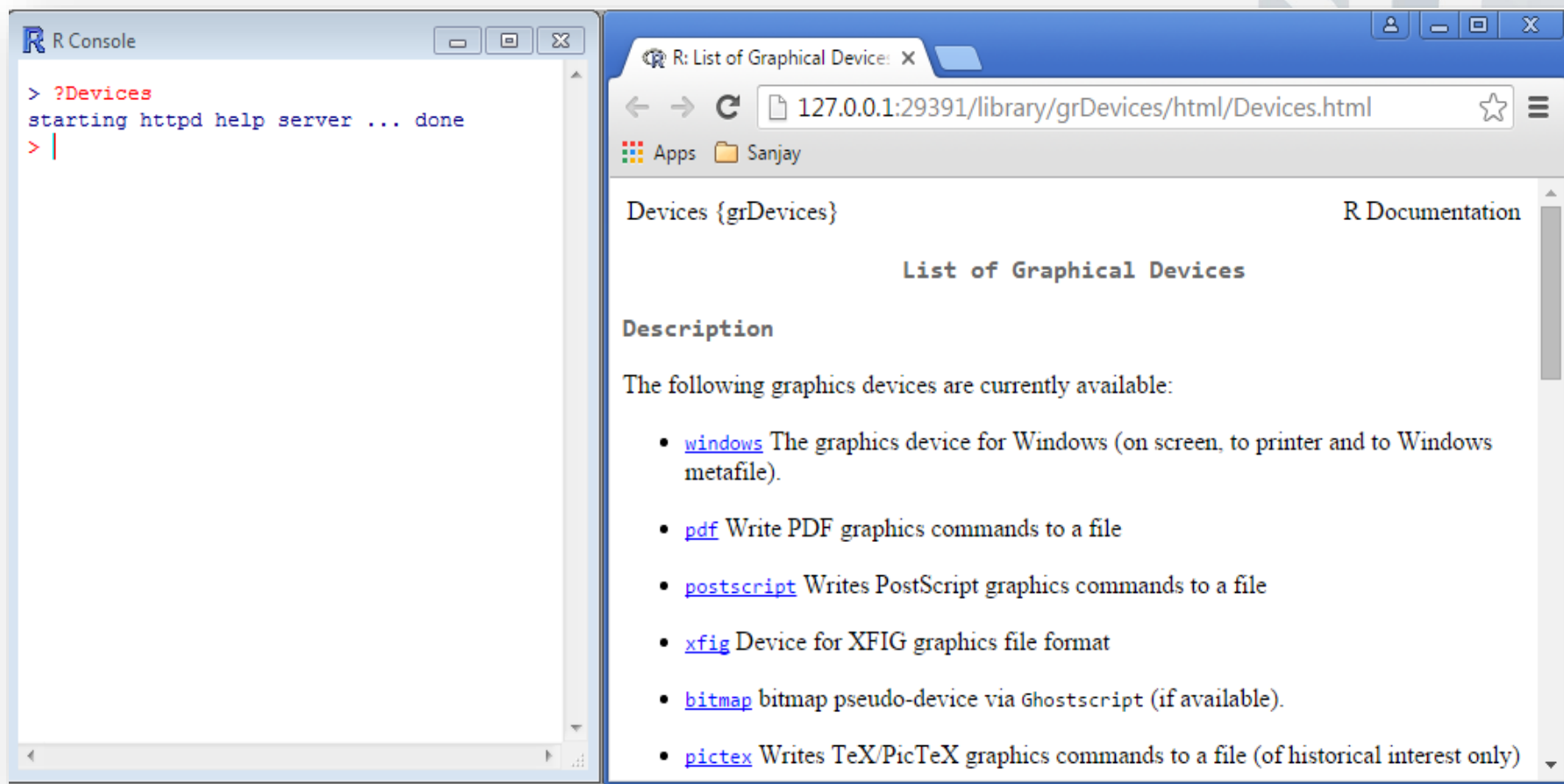
Graphic Devices

What is a Graphic Device?

- It is a destination to which our plot will be sent to as an output

- e.g.
  - On Screen (R default)
  - PDF
  - Bitmap
  - Jpeg
  - Tiff
  - etc

Graphic Devices in R

- To know which devices are available to your current workspace we can type the command **?Devices**

Creation of a Plot in R

- The following are sequence of the events happening while the plot gets shown to us in R:
  - Calling of graph function like plot(), boxplot()
  - Plot appearing on the screen
  - Annotating of the plot, if required

```
hist(Income,breaks = 20,xlab="Income",main="Distribution of Income" )
abline(v = median(Income), col="red",lwd=2) # Annotating the plot
```


Distribution of Income

Using Graphic Device

- We can introduce a specific device in the first step itself:
  - Explicitly launching a graphics devices
  - Calling of graph function like plot(), boxplot()
  - Plot appearing on the screen
  - Annotating of the plot, if required
  - Explicitly closing the graphics device using **dev.off()**

```
pdf("D:\\Data Science Training\\Outputs\\HistIncome.pdf")
hist(Income,breaks = 20,xlab="Income",main="Distribution of Income" )
abline(v = median(Income), col="red",lwd=2) # Annotating the plot
dev.off()
```