**VISUALLY INSPECTING DATA**

1. **Histograms**

A histogram is a descriptive statistical method that shows data by dividing the range of values into intervals and plotting the frequency/density per interval as a bar.

R creates histogram with hist() function. The function takes vector as input and uses additional parameters to plot histograms. Eg. of Histogram is as follows

wages = c(400,500,200,700,500,200,400,900,500,600,300,500,400)

h =hist(wages, main = “Daily wages”, xlab =”wages in Rs”, ylab = “freq”, xlim = c(200,900), col=”pink”, freq = TRUE)

1. **Density Plots**

density plots are usually a much more effective way to view the distribution of a variable. Create the plot using plot(density(*x*)) where *x* is a numeric vector.

d <- density(mtcars$mpg)

plot(d, main="Kernel Density of Miles Per Gallon")

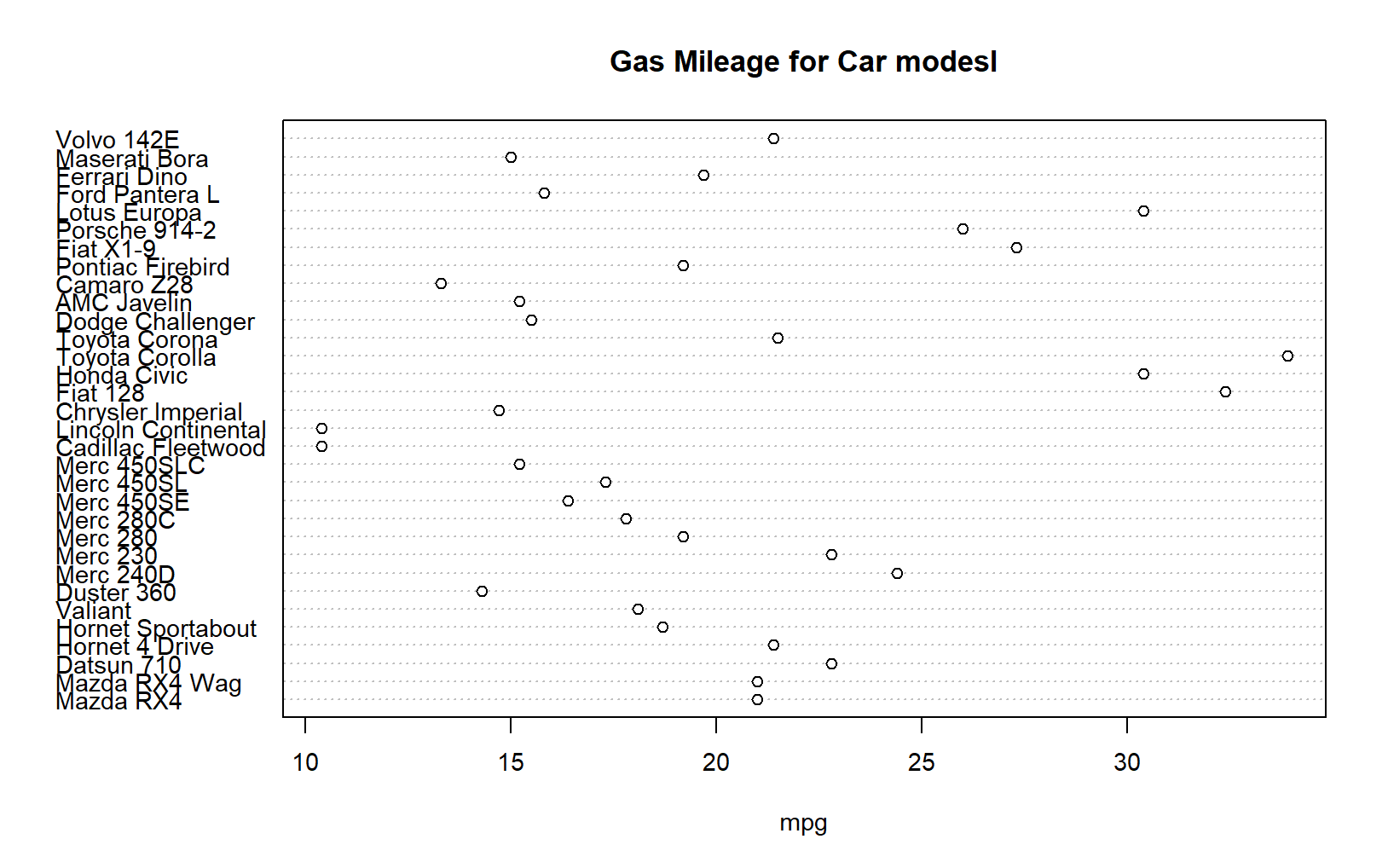
polygon(d, col="red", border="blue")

1. **Dot Plots**

A dot plot, also known as a dot diagram, is a statistical chart consisting of data points on a relatively simple scale. Dot plots are considered as one of the easiest statistical plots, used for small [data sets](https://www.vedantu.com/maths/data-sets). Dot plots are used for highlighting clusters, gaps, skews in distribution, and outliers. When working with datasets larger than 20 or 30, it is better to use another statistical chart such as a [histogram](https://www.vedantu.com/maths/histogram), as the dot plot will be less manageable.

**dotchart(mtcars$mpg, labels = row.names(mtcars),**

**cex = 0.9, , xlab = "mpg", main = "Gas Mileage for Car modesl")**

****

1. **Line Charts**

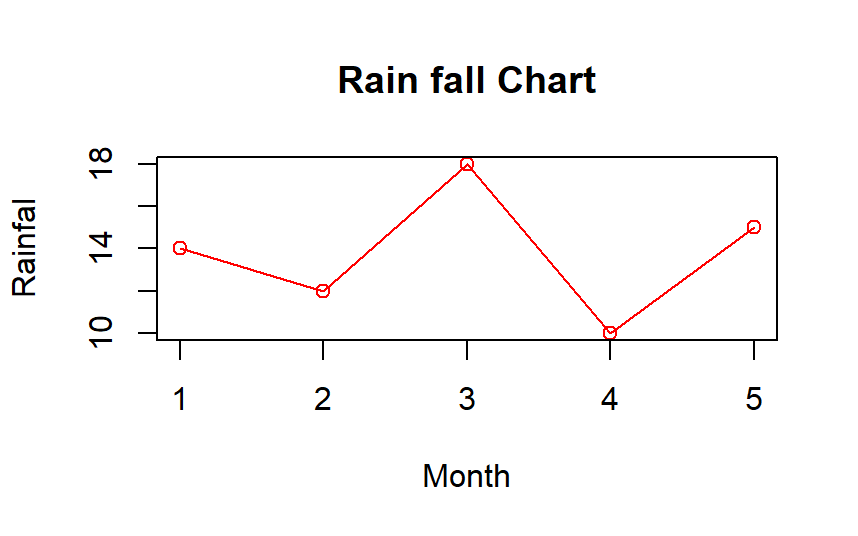
A Line chart is a graph that connects a series of points by drawing line segments between them. Plot function in R is used to create line graphs.

Eg

***y=c(14,12,18,10,15)***

***x=c(1,2,3,4,5)***

***plot(x,y,type="o", col="red", xlab="Month", ylab="Rainfal", main= "Rain fall Chart")***

****

1. **Pie Charts**

Pie chart is a representation of values as slices of a circle with different colours. The slices are labelled and the numbers corresponding to each slice is also represented in the chart. pie() function is used, which takes positive numbers as vector input.

Consider the below data that represents the monthly expenditure of an individual.

Housing = 600

Food = 500

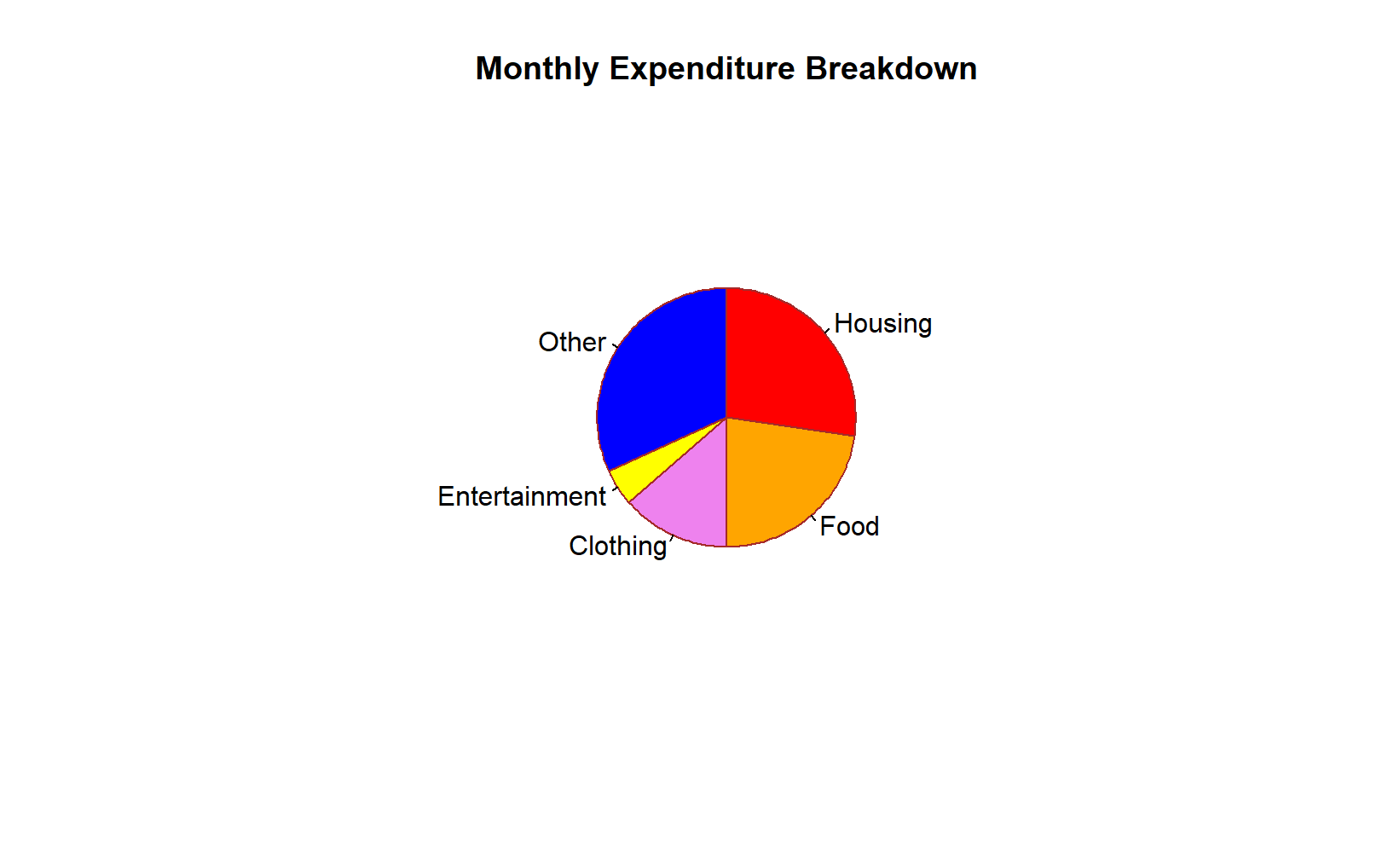
Clothing = 300

Entertainment = 100

Other = 700

***expenditure = c(600,500,300,100,700)***

***pie(expenditure,labels=c("Housing","Food","Clothing","Entertainment","Other"), main="Monthly Expenditure Breakdown", col=c("red","orange","violet","yellow","blue"), border="brown",clockwise=TRUE)***



1. **Box Plots**

Boxplots are a measure of how well distributed is the data in the dataset,

boxplot(x, data, notch, varwidth, names, main)

x is a vector or a formula.

data is the data frame.

notch is a logical value. Set as TRUE to draw a notch.

varwidth is a logical value. Set as true to draw width of the box proportionate to the sample size.

names are the group labels which will be printed under each boxplot.

main is used to give a title to the graph.

***input <- mtcars[,c('mpg','cyl')]***

***print(head(input))***

***boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders",***

***ylab = "Miles Per Gallon", main = "Mileage Data")***

***With notch***

***boxplot(mpg ~ cyl, data = mtcars,***

***xlab = "Number of Cylinders",***

***ylab = "Miles Per Gallon",***

***main = "Mileage Data",***

***notch = TRUE,***

***varwidth = TRUE,***

***col = c("green","yellow","purple"),***

***names = c("High","Medium","Low")***

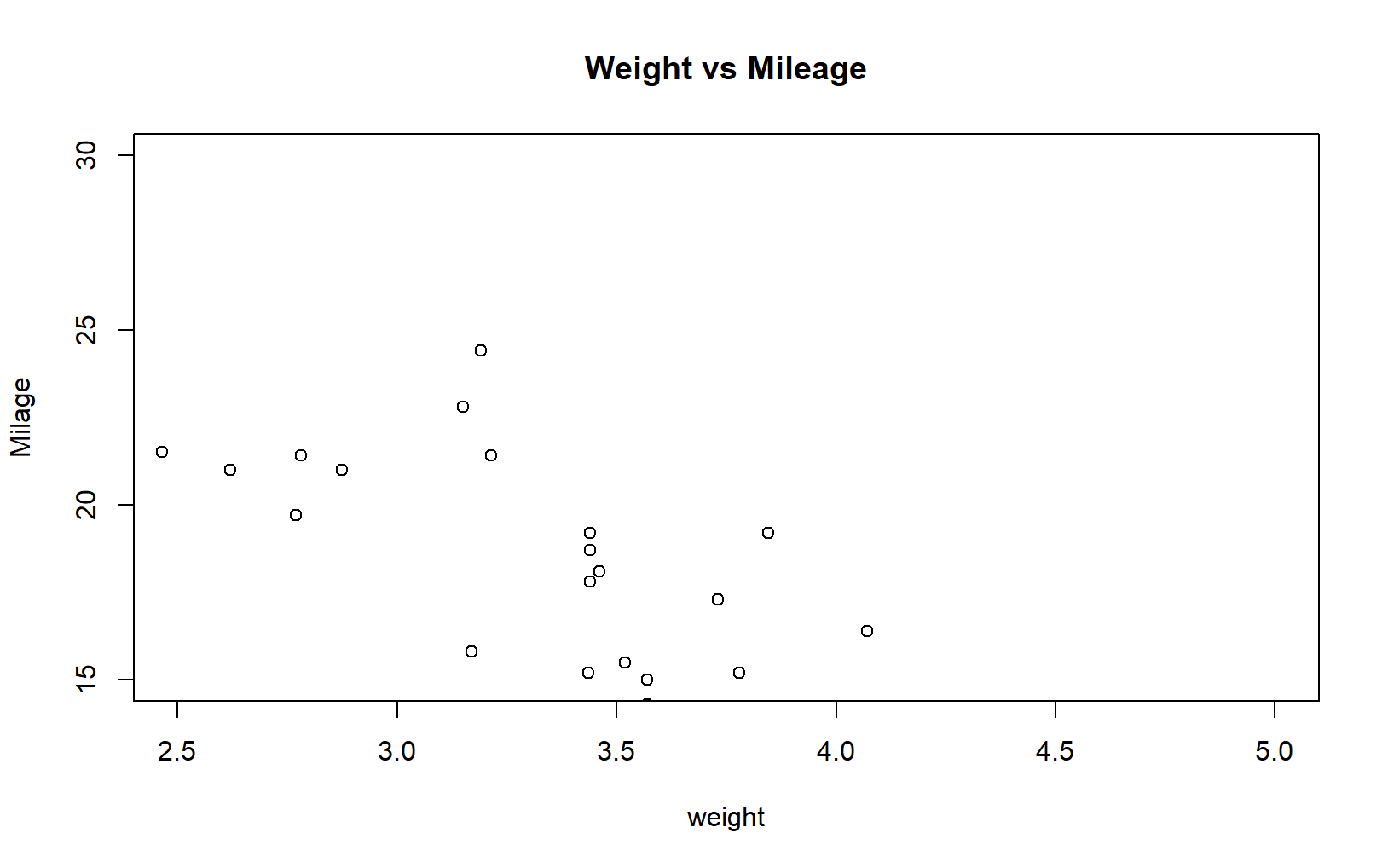
***)***

1. **Scatterplots**

A scatterplot is a mathematical diagram using cartesian coordinates to display values for typically two variables for a set of data. A scatterplor can be used either when one continuous variable that is under control of the experiment and the other depends on it or when both are independent.

***input = mtcars[,c('wt','mpg')]***

***plot(x=input$wt,y=input$mpg,xlab="weight",ylab="Milage",xlim=c(2.5,5),ylim=c(15,30),main="Weight vs Mileage")***

****