R for Beginners - R Code File -2

This R code book has been written by Rohit Dhankar . GitHub - https://github.com/RohitDhankar

This is the 2nd in series of R Code Files.

Refer GitHub Repository , for all Data Files –> https://github.com/RohitDhankar/R-Beginners-Online-Virtual-Learning-Session

Its a good practice from time to time to keep a track of our current Working Directory and list out all the Objects in our R ENVIRONMENT - specially so when we are committing changes to a Git Remote.

VECTOR Operations

```
getwd()
```

[1] "/home/dhankar/Desktop/R_Own/Proj_1"
```
#
ls()
```

character(0)

We could remove any object with command - rm("Object Name")

We can also use print() , to view any objects stored value.

```
# Code Section -1
a1 <- "FINANCE"
b1 <- "MARKETING"
c1 <- "SALES"
d1 <- 3.1416
char_vector <- c("x","d","c","f")
print(a1)
```

[1] "FINANCE"
```
#
print(char_vector)
```

[1] "x" "d" "c" "f"

Going further with VECTORS .

We combine two or more vectors to get another vector .

```
# Code Section -2
num_vector <- c(22,22,33,33,44)
print(num_vector)
```

[1] 22 22 33 33 44
```
num_vector1 <- c(11,12,13,14,15)
#
num_vector3 <- c(num_vector,num_vector1)
print(num_vector3)
```

[1] 22 22 33 33 44 11 12 13 14 15

```r
#
sort(num_vector3)
```

```
##  [1] 11 12 13 14 15 22 22 33 33 44
```

```r
#
order(num_vector3) # Ascending Order of  ELEMENTS without SORTING .
```

```
##  [1]  6  7  8  9 10  1  2  3  4  5
```

```r
#
# The COLON Operator is same as the seq() function seen later in this text.

seq_1<- 55:50
seq_1
```

```
## [1] 55 54 53 52 51 50
```

```r
#
seq_2<- 50:55
seq_2
```

```
## [1] 50 51 52 53 54 55
```

```r
#
# While the above - seq_1 and seq_2 are stored as Objects withing persistence storage.

50:55 ## is in Memory Only and Not Stored on any persistence storage.
```

```
## [1] 50 51 52 53 54 55
```

```
## Source -- R Manual -- https://stat.ethz.ch/R-manual/R-devel/library/base/html/Colon.html
```

Some basic Maths and Stats with VECTORS.

```r
# Code Section -3
num_vector3 + 10
```

```
##  [1] 32 32 43 43 54 21 22 23 24 25
```

```r
# Adds NUMERIC VALUE = 5 to all ELEMENTS of the Num Vector.
nmv_1<-c(20,21,211,312,413,5114)
nmv_2<-c(20,21,211,313,414,5214)
#
class(nmv_1) # "numeric" Vector
```

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

```r
#
nmv_vect <- (1/nmv_1)
class(nmv_vect)
```

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

```r
typeof(nmv_vect)
```

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

```r
typeof(nmv_2) ## CHECK --- Why not Integer or ??
```

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

```r
print(1/nmv_1)
```

```
## [1] 0.0500000000 0.0476190476 0.0047393365 0.0032051282 0.0024213075
## [6] 0.0001955417
#
min_max_nmv <- c(min(nmv_1),max(nmv_1))
min_max_nmv  ## Output - MIN == 20 , MAX ==5114
```

```
## [1]   20 5114
#

### MATHEMATICAL logical operators and Boolean calculus - present in R .
# - <, <=, >, >=, == for exact equality and != for inequality.

### Boolean calculus
# Given - nmv_1 and nmv_2 are logical expressions,
# thus nmv_1 & nmv_2 is intersection ("AND")
# nmv_1 | nmv_2 is union ("OR")
# !nmv_1 is Negation of nmv_1.

nmv_1 == nmv_2
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE FALSE
#

nmv_1 != nmv_2
```

```
## [1] FALSE FALSE FALSE  TRUE  TRUE  TRUE
#
# Lets introduce NA's - the data wranglers nightmare
#

nmv_3 <- c(11,NA,22,33,44,NA,NA)
nmv_3
```

```
## [1] 11 NA 22 33 44 NA NA
# We can use function is.na() , to find out the NOT AVAILABLE missing values
# At a letr stage we shall also look at NA management or IMPUTATION of MISSING VALUES
# Here is a prelim resource --
#
is.na(nmv_3)
```

```
## [1] FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
#
# Kaggle_Titanic [Multiple Imputation of Missing_Values] --
# http://datasciencewithrandpython.blogspot.in/2017/01/kaggle-titanic-initial-analysis-wip.html
#
# Not a NUMBER = NaN
xx <- 0/0.00
xx
```

```
## [1] NaN
```

```
#
is.nan(xx) ## TRUE
```

```
## [1] TRUE
```

```
#
is.nan(nmv_3) ## As Many FALSE - as Elements ### NA's are NOT Treated as NaN's
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
#
nmv_4 <- nmv_3 + 2 ## Any OPERATION done with a NA value , results in a NA value.
nmv_4
```

```
## [1] 13 NA 24 35 46 NA NA
```

```
#
# Creating an INTEGER Numeric Vector

int_num_vec <- c(22L,33L,44L,55L,66L)
class(int_num_vec) ## CLASS == "integer"
```

```
## [1] "integer"
```

```
typeof(int_num_vec)
```

```
## [1] "integer"
int_num_vec
```

```
## [1] 22 33 44 55 66
```

```
# Creating a DOUBLE Numeric Vector

db_num_vec <- c(22.1,33.2,44.3,55.6,66.7)
class(db_num_vec) ## CLASS == "numeric"
```

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

```
typeof(db_num_vec) ## DOUBLE
```

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

```
## [1] 22.1 33.2 44.3 55.6 66.7
```

```
# Code Section -4
num_vector1
```

```
## [1] 11 12 13 14 15
```

```
#
num_vector3
```

```
##  [1] 22 22 33 33 44 11 12 13 14 15
```

```
#
num_vector1 * num_vector3
```

```
##  [1] 242 264 429 462 660 121 144 169 196 225
```

```r
# First 5 elements of - num_vector3 multiplied by the Five Elements
# of num_vector1 and again the Next 5 elements of num_vector3
# multiplied by the Five Elements of num_vector1

# Concatenate Strings - Its ofetn required to PASTE together CHAR VARIABLES
# to create more Complex CHAR VARIABLES
# R has a handy function called PASTE -
# ?paste() # Uncomment to see help
# Concatenate vectors after converting to character.

col_names_1 <- paste(c("N","P","Q","R"), 1:20, sep="")
col_names_1
```

```
## [1] "N1"  "P2"  "Q3"  "R4"  "N5"  "P6"  "Q7"  "R8"  "N9"  "P10" "Q11"
## [12] "R12" "N13" "P14" "Q15" "R16" "N17" "P18" "Q19" "R20"
```

```r
col_name_2_1 <- paste(c("Q","Q","Q","Q"),1:4, sep="")
col_name_2_1
```

```
## [1] "Q1" "Q2" "Q3" "Q4"
```

```r
#
col_names_2 <- paste(c("X","Y","Z"), 1:3, sep="")
col_names_2
```

```
## [1] "X1" "Y2" "Z3"
```

```r
#
col_names_3 <- paste(c("M","N","P"), col_names_2, sep="")
col_names_3
```

```
## [1] "MX1" "NY2" "PZ3"
```

```r
#
#
col_names_4 <- paste( col_names_2,c("M","N","P"), sep="")
col_names_4
```

```
## [1] "X1M" "Y2N" "Z3P"
```

```r
# As seen above very handy for creating COLUMN NAMES or VARIABLE LABELS
# Kindly notice - PASTE() folllows the Order of R OBJECTS provided
# and converts them to CHAR VECTORS.
```

Check out the LENGTH of a VECTOR with length()

```r
# Code Section -5

length(num_vector1 * num_vector3)
```

```
## [1] 10
```

```r
# Code Section -6

#num_vector1 %*% num_vector3 # Error in num_vector1 %*% num_vector3 : non-conformable arguments

# Vectors are not of same Length above - below they are of same length

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

```r
nv1 <- c(6,7,8,9,10)

nv %*% nv1 # Inner Product of same Length Vectors
```

```
##      [,1]
## [1,]  130
```

```r
#  "two equal-length sequences of numbers (usually coordinate vectors) and returns a single number."
# Algeberic Dot Product as defined by WikiPedia - "https://en.wikipedia.org/wiki/Dot_product"

### CHECK --- A element wise product example to be included
```

Operate upon a ELEMENT of the Vector.

```r
# Code Section -7


log(num_vector3[2]) # Log Base 2 of 22
```

```
## [1] 3.091042
```

```r
#
log(22)
```

```
## [1] 3.091042
```

```r
#
```

Converting a CHAR Vector into a NUMERIC Vector .

```r
# Code Section -8
ch_v <- c("11","12","13","14","15")
typeof(ch_v)
```

```
## [1] "character"
```

```r
#
class(ch_v)
```

```
## [1] "character"
```

```r
#ch_v + 2 # Error in ch_v + 2 : non-numeric argument to binary operator
# Cant do a Math operation on CHAR Vector - lets Convert into NUM Vector
#
nm_v <- as.numeric(ch_v)
#
class(nm_v)
```

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

```r
nm_v + 2
```

```
## [1] 13 14 15 16 17
```

```r
#

#Summary of the Num Vector as below :-
#
summary(nm_v+2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##      13      14      15      15      16      17
#
summary(nm_v+5)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       16      17      18      18      19      20
#
nm_v
```

```
## [1] 11 12 13 14 15
sum(nm_v+5)
```

```
## [1] 90
#
sd(nm_v+5) ## CHECK  -- Standard Deviation in much detail
```

```
## [1] 1.581139
#
max(nm_v+5)
```

```
## [1] 20
#
min(nm_v+5)
```

```
## [1] 16
#
mean(nm_v+5)
```

```
## [1] 18
#
median(nm_v+5)
```

```
## [1] 18
#
#The Quantile -
#
quantile(nm_v+5)
```

```
##   0%  25%  50%  75% 100%
##   16   17   18   19   20
#
quantile(nm_v+100)
```

```
##   0%  25%  50%  75% 100%
##  111  112  113  114  115
#
#We can also specify the Quantile buckets or Percentiles as an argument to the Quantile function :-
#
nmv_q <- c(10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,100)
summary(nmv_q)
```
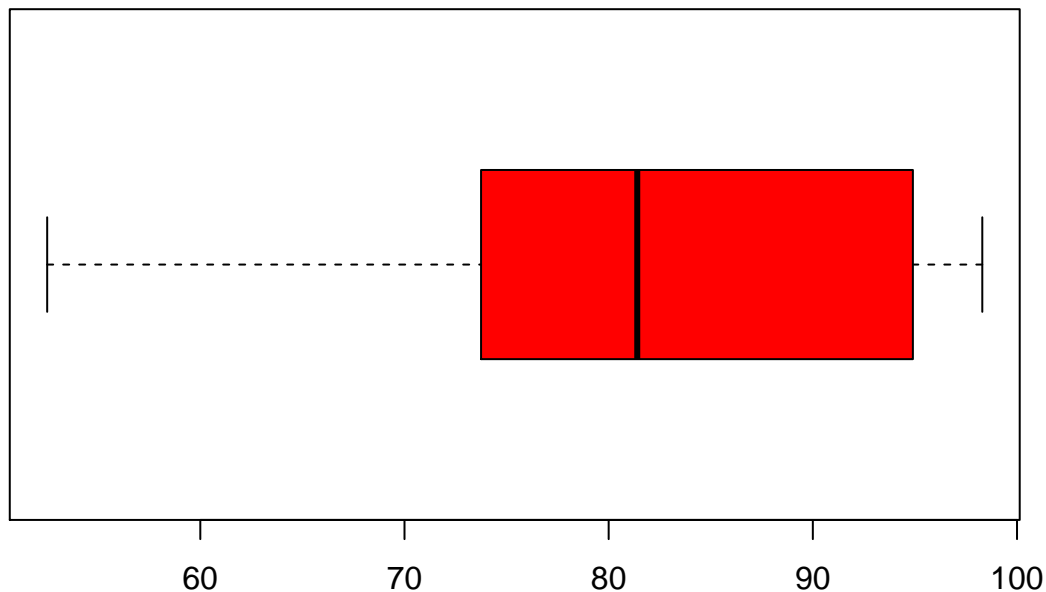
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##    10.00   31.25   52.50   52.78   73.75  100.00
percent_1 <- quantile(nmv_q, c(.50,.75,.84, .97, .99))
percent_1
```

```
##   50%   75%   84%   97%   99%
## 52.50 73.75 81.40 94.90 98.30
```

```r
boxplot(percent_1,col = "red",horizontal = TRUE,
        main = "Box and Whisker Plot of Quantiles",
        xlab = "Quantile Values")
```

Box and Whisker Plot of Quantiles


```
# Kindly note how the ARGUMENTs to boxplot()
# have been bumped to the next row - keeping in mind
# the Horizontal space of our PDF knit of the .Rmd file

# Seen above we have the MEDIAN quartile - 50% and the UPPER
# Quartile - 75% along with THREE more percentiles.


boxplot(nmv_q,col = "green",horizontal = TRUE,
        main = "Box and Whisker Plot of Quantiles",
        xlab = "Vector Values")
```
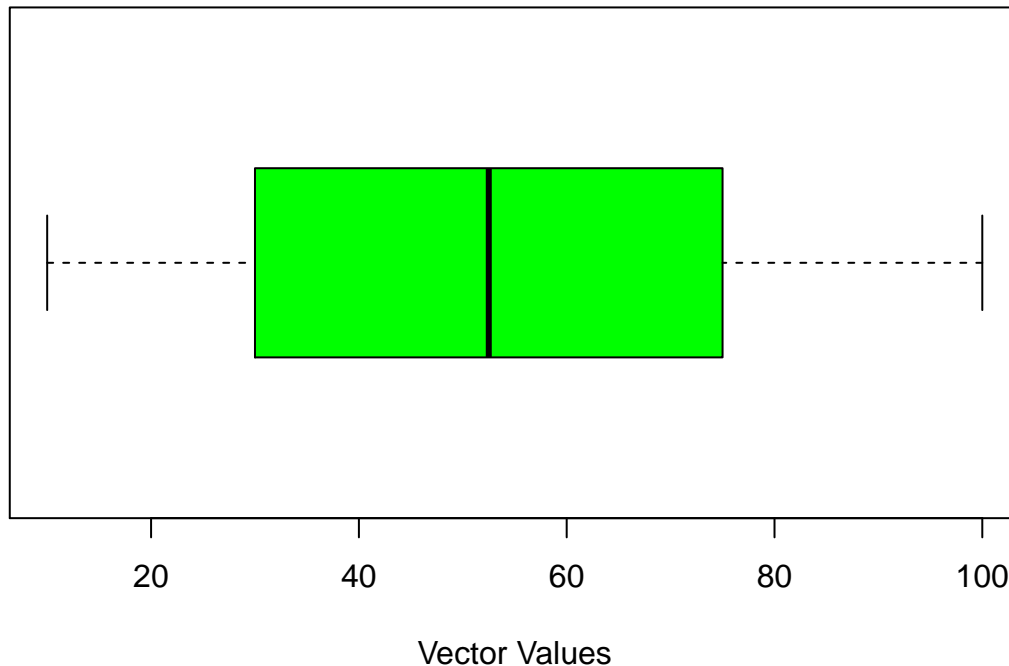
Box and Whisker Plot of Quantiles

Wiki reference – Percentile Rank - "https://en.wikipedia.org/wiki/Percentile_rank" #

Intro to ANOVA and BOXPLOTS

We also carry out ONE Way ANOVA or ANALYSIS of VARIANCE test with the BOX and WHISKERS plots as seen below :-

```
# Code Section -9
library(graphics)

nmv_q <- c(10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,100)
percent_1 <- quantile(nmv_q, c(.50,.75,.84, .97, .99))
percent_1

##   50%    75%    84%    97%    99%
## 52.50 73.75 81.40 94.90 98.30

percent_2 <- quantile(nmv_q, c(.1, .3, .16, .40, .50))
percent_2

##  10%  30%  16%  40%  50%
## 18.5 35.5 23.6 44.0 52.5

percent_3 <- quantile(nmv_q, c(.16, .40, .50,.75,.84))
percent_3

##   16%    40%    50%    75%    84%
## 23.60 44.00 52.50 73.75 81.40

col_boxes = (c("red","blue","green"))
```

```
boxplot(percent_1,percent_2,percent_3,col = col_boxes,
        names = c("perc_1","perc_2","perc_3"),horizontal = TRUE,
        main = "Box and Whisker Plot of Quantiles",
        xlab = "Quantile Values")
```

Box and Whisker Plot of Quantiles

Quantile Values
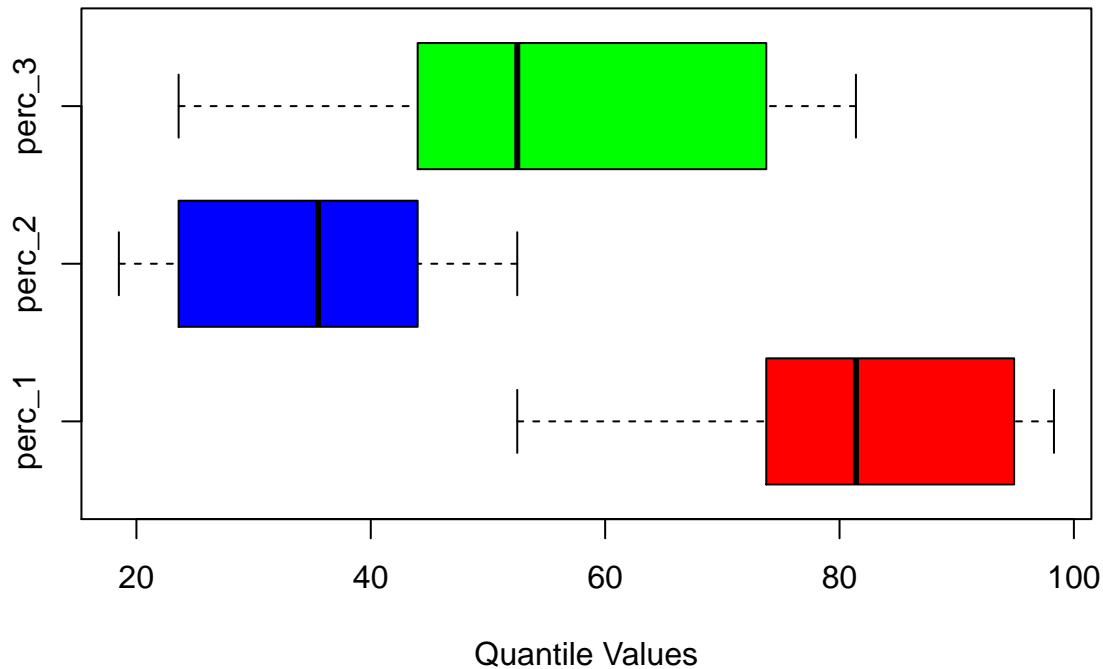
```
# Kindly note the Quantiles are randomly chosen here
# this is not the best way to choose quantiles
# we shall come back for details later in this text
```

rainbow() for Coloring Boxplots -

```
# Code Section -10

percent_4 <- quantile(nmv_q, c(.16, .40, .50,.95,.99))
percent_4

##  16%  40%  50%  95%  99%
## 23.6 44.0 52.5 91.5 98.3

percent_5 <- quantile(nmv_q, c(.16, .24,.32 ,.40,.75))
percent_5

##   16%   24%   32%   40%   75%
## 23.60 30.40 37.20 44.00 73.75

percent_6 <- quantile(nmv_q, c(.1, .5, .26, .45, .60))
percent_6

##   10%   50%   26%   45%   60%
```

```
## 18.50 52.50 32.10 48.25 61.00

percent_7 <- quantile(nmv_q, c(.3, .7, .18, .43, .70))
percent_7

##    30%    70%    18%    43%    70%
## 35.50 69.50 25.30 46.55 69.50

col_rainbow <- rainbow(7)

boxplot(percent_1,percent_2,percent_3,percent_4,percent_5,percent_6,percent_7,col = col_rainbow,
        names = c("perc_1","perc_2","perc_3","perc_4","perc_5","perc_6","perc_7"),horizontal = TRUE,
        main = "Box and Whisker Plot of Quantiles",
        xlab = "Quantile Values")
```

Box and Whisker Plot of Quantiles

Quantile Values
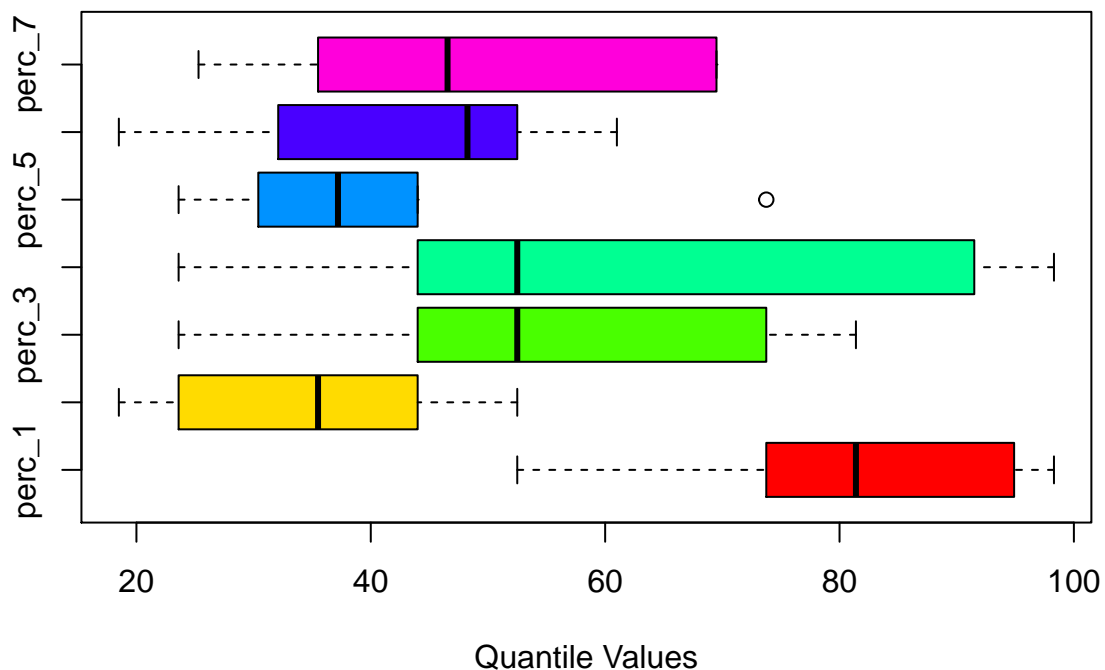
```
# Code Section -11

# Just for Fun a PIE Graph --- you always ...
# need to avoid PIE Graphs for "DESCRIPTIVE STATS"
# I personally love to use them for MARKETING
# They are usually excellent EYE CANDY :P

pie(rep(1, 7), col = rainbow(7))
```
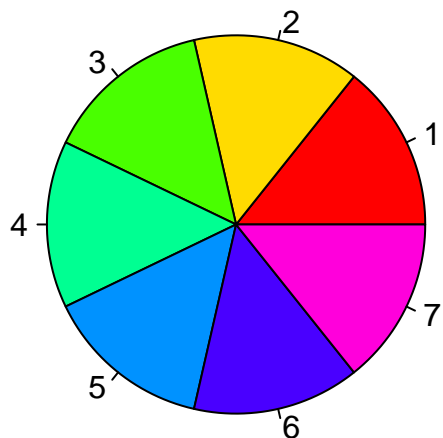
MATRICE Operations - TRANSPOSE of a MATRIX

Coming back to MATRICES lets further look at some MATRIX Operations :-

```
# Code Section -12

m1 <- matrix(data=66:69,nrow=2,ncol=2)
m1
```

```
##      [,1] [,2]
## [1,]   66   68
## [2,]   67   69
```

```
# Lets now TRANSPOSE this MATRIX - for more on TRANSPOSE of MATRICES
# kindly refer this Wiki Link -- https://en.wikipedia.org/wiki/Transpose


t(m1)
```

```
##      [,1] [,2]
## [1,]   66   67
## [2,]   68   69
```

```
# As seen below - the DIAGONAL Elements remain as -is .
# 66 and 69 do not move .
# 67 and 68 switch places , thus giving us a Transpose Matrix.

# Lets look at another example of TRANSPOSE ....

m2 <- matrix(data=10:25,nrow=4,ncol=4)
m2
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   10   14   18   22
## [2,]   11   15   19   23
## [3,]   12   16   20   24
## [4,]   13   17   21   25
```

```
class(m2)
```

```
## [1] "matrix"
```

```
## Note  in the above sequence - 10:25 - both 10 and 25 are included.
# Lets now TRANSPOSE this MATRIX - for more on TRANSPOSE of MATRICES
# kindly refer this Wiki Link -- https://en.wikipedia.org/wiki/Transpose

t(m2)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   10   11   12   13
## [2,]   14   15   16   17
## [3,]   18   19   20   21
## [4,]   22   23   24   25
```

```
# As seen below - the DIAGONAL Elements remain as-is.
# 10, 15 , 20 , 25 -- do not move .
# Non Diagonal elements are Transposed ,giving the Transpose Matrix.
```

The Semicolon Notation - RANGE or SEQUENCE

```
# Code Section -13

# Quick recap of the SEQUENCE

a_seq <- 66:69
a_seq
```

```
## [1] 66 67 68 69
```

```
# In the earlier section we create a MATRIX by using a sequence within
# the COMBINE function

# We can also use the - seq - sequence function as seen below

b_seq <- seq(from=66, to=69, by=1)
b_seq
```

```
## [1] 66 67 68 69
```

```
#

b_seq <- seq(from=66, to=69, by=2)
b_seq
```

```
## [1] 66 68
```

```
#
b_seq <- seq(from=1, to=30, by=3)
b_seq
```

```
##  [1]  1  4  7 10 13 16 19 22 25 28
```

```
#
c_seq <- seq(from=1, to=10, by=2)
c_seq
```

```
## [1] 1 3 5 7 9
```

```
#
class(c_seq)
```

```
## [1] "numeric"
typeof(c_seq)
```

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

The CBIND and RBIND Functions

We can COLUMN Bind and ROW Bind more than one data structures as seen below -

```
# Code Section -14

ma1 <- matrix(data=10:15,nrow=3,ncol=2)
ma1
```

```
##      [,1] [,2]
## [1,]   10   13
## [2,]   11   14
## [3,]   12   15
#
class(ma1)
```

```
## [1] "matrix"
#
ma2 <- matrix(data=20:25,nrow=3,ncol=2)
ma2
```

```
##      [,1] [,2]
## [1,]   20   23
## [2,]   21   24
## [3,]   22   25
#
class(ma2)
```

```
## [1] "matrix"
# ROW Bind the Matrices
ma3 <- rbind(ma1,ma2)
ma3
```

```
##      [,1] [,2]
## [1,]   10   13
## [2,]   11   14
## [3,]   12   15
## [4,]   20   23
## [5,]   21   24
## [6,]   22   25
#
# COLUMN Bind the Matrices
ma4 <- cbind(ma1,ma2)
ma4
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   10   13   20   23
## [2,]   11   14   21   24
## [3,]   12   15   22   25
```

```
# As seen below we need to have same COLUMN Numbers to do a RBIND
#m3 <- rbind(m1,m2)

# # As seen below we need to have same ROW Numbers to do a RBIND
#m3 <- cbind(m1,m2)
```

ROW BIND - Data Frames

```
# Code Section -15

df_1 <- read.csv("~/Desktop/R_Own/R_1 - Sheet1.csv",header =TRUE , sep = "," )
df_1
```

```
##     X.    Product.Name        Prod.ID Date.of.Invoice Date.of.Shipping
## 1    1 OFF-LA-10002782 MX-2014-143658      01-01-2013       02-01-2013
## 2    2 FUR-FU-10004015 MX-2012-155047      01-01-2013       02-01-2013
## 3    3 FUR-BO-10002352 MX-2012-155047      01-01-2013       02-01-2013
## 4    4 OFF-BI-10004428 MX-2012-155047      01-01-2013       02-01-2013
## 5    5 OFF-AR-10004594 MX-2012-155047      01-01-2013       02-01-2013
## 6    6 OFF-EN-10001375 MX-2012-155047      01-01-2013       02-01-2013
## 7    7 OFF-EN-10001375 MX-2013-134096      01-01-2013       02-01-2013
## 8    8 TEC-MA-10004956 MX-2013-134096      01-01-2013       02-01-2013
## 9    9 OFF-SU-10003474 MX-2013-134096      01-01-2013       02-01-2013
## 10  10 TEC-AC-10001830 MX-2013-134096      01-01-2013       02-01-2013
## 11  11 OFF-BI-10002075 MX-2013-134096      01-01-2013       02-01-2013
## 12  12 OFF-FA-10002526 MX-2013-156335      01-01-2013       02-01-2013
## 13  13 FUR-CH-10002846 MX-2013-156335      01-01-2013       02-01-2013
## 14  14 OFF-EN-10004100 MX-2014-121923      02-01-2013       04-01-2013
## 15  15 OFF-AR-10003914 MX-2014-135706      02-01-2013       03-01-2013
## 16  16 OFF-FA-10000038 MX-2014-135706      02-01-2013       03-01-2013
## 17  17 OFF-EN-10000761 US-2013-126655      02-01-2013       03-01-2013
## 18  18 FUR-FU-10003066 US-2013-126655      02-01-2013       03-01-2013
## 19  19 OFF-EN-10000075 US-2013-126655      02-01-2013       03-01-2013
## 20  20 OFF-EN-10002226 US-2013-126655      02-01-2013       03-01-2013
## 21  21 FUR-CH-10002132 MX-2013-167759      02-01-2013       04-01-2013
## 22  22 TEC-AC-10002749 MX-2013-163139      02-01-2013       02-01-2013
## 23  23 OFF-SU-10000066 MX-2013-163139      02-01-2013       02-01-2013
## 24  24 OFF-BI-10003934 US-2014-119753      02-01-2013       03-01-2013
## 25  25 OFF-BI-10003932 US-2012-133970      02-01-2013       03-01-2013
##    Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type
## 1      13.080        3      39.240              1      PRIORITY
## 2     252.160        8    2017.280              2      PRIORITY
## 3     193.280        2     386.560              3      PRIORITY
## 4      35.440        4     141.760              4      PRIORITY
## 5      71.600        2     143.200              5      PRIORITY
## 6      56.120        2     112.240              6      PRIORITY
## 7      56.120        2     112.240              7      STANDARD
## 8     344.640        3    1033.920              8      STANDARD
```

```
## 9       97.360      4     389.440          9     STANDARD
## 10     341.520      2     683.040         10     STANDARD
## 11      12.060      3      36.180         11     STANDARD
## 12      20.760      3      62.280         12     STANDARD
## 13     210.640      4     842.560         13     STANDARD
## 14      80.100      3     240.300         14     STANDARD
## 15     132.640      4     530.560         15     STANDARD
## 16      12.940      1      12.940         16     STANDARD
## 17      18.840      2      37.280         17     STANDARD
## 18     308.280      7    2157.560         18     STANDARD
## 19      40.176      2      79.952         19     STANDARD
## 20       8.784      3      25.952         20     PRIORITY
## 21     273.472      4    1093.688         21     PRIORITY
## 22      27.000      1      27.000         22     PRIORITY
## 23     207.000      9    1863.000         23     PRIORITY
## 24      60.660      3     181.580         24     PRIORITY
## 25     181.116      9    1629.644         25     PRIORITY
##               Category
## 1    Office Supplies
## 2          Furniture
## 3          Furniture
## 4    Office Supplies
## 5    Office Supplies
## 6    Office Supplies
## 7    Office Supplies
## 8         Technology
## 9    Office Supplies
## 10        Technology
## 11   Office Supplies
## 12   Office Supplies
## 13         Furniture
## 14   Office Supplies
## 15   Office Supplies
## 16   Office Supplies
## 17   Office Supplies
## 18         Furniture
## 19   Office Supplies
## 20   Office Supplies
## 21         Furniture
## 22        Technology
## 23   Office Supplies
## 24   Office Supplies
## 25   Office Supplies
```

```r
# Code Section -16

df_2 <- read.csv("~/Desktop/R_Own/R_2.csv",header =TRUE , sep = "," )
df_2
```

```
##   X.      Product.Name       Prod.ID Date.of.Invoice Date.of.Shipping
## 1  1 TEC-AC-10001830 MX-2013-134096      05-01-2013       06-01-2013
## 2  2 FUR-FU-10004015 MX-2012-155047      03-01-2013       03-01-2013
## 3  3 FUR-BO-10002352 MX-2012-155047      03-01-2013       03-01-2013
## 4  4 OFF-BI-10004428 MX-2012-155047      03-01-2013       03-01-2013
## 5  5 OFF-AR-10004594 MX-2012-155047      03-01-2013       03-01-2013
```

```
## 6   6 OFF-EN-10001375 MX-2012-155047       03-01-2013          03-01-2013
## 7   7 OFF-EN-10001375 MX-2013-134096        03-01-2013          04-01-2013
## 8   8 OFF-AR-10003914 MX-2014-135706        03-01-2013          04-01-2013
## 9   9 OFF-FA-10000038 MX-2014-135706        03-01-2013          04-01-2013
## 10 10 OFF-EN-10000761 US-2013-126655        03-01-2013          04-01-2013
## 11 11 FUR-FU-10003066 US-2013-126655        03-01-2013          04-01-2013
## 12 12 OFF-EN-10000075 US-2013-126655        03-01-2013          04-01-2013
## 13 13 OFF-EN-10002226 US-2013-126655        03-01-2013          03-01-2013
## 14 14 FUR-CH-10002132 MX-2013-167759        03-01-2013          03-01-2013
## 15 15 OFF-EN-10001375 MX-2013-134096        03-01-2013          04-01-2013
## 16 16 TEC-MA-10004956 MX-2013-134096        03-01-2013          04-01-2013
## 17 17 OFF-SU-10003474 MX-2013-134096        03-01-2013          04-01-2013
## 18 18 TEC-AC-10001830 MX-2013-134096        03-01-2013          04-01-2013
## 19 19 OFF-BI-10002075 MX-2013-134096        03-01-2013          04-01-2013
## 20 20 OFF-FA-10002526 MX-2013-156335        03-01-2013          04-01-2013
## 21 21 OFF-EN-10001375 MX-2013-134096        04-01-2013          05-01-2013
## 22 22 TEC-MA-10004956 MX-2013-134096        04-01-2013          05-01-2013
## 23 23 OFF-SU-10003474 MX-2013-134096        04-01-2013          05-01-2013
## 24 24 TEC-AC-10001830 MX-2013-134096        04-01-2013          05-01-2013
## 25 25 OFF-BI-10002075 MX-2013-134096        04-01-2013          05-01-2013
## 26 26 OFF-EN-10001375 MX-2013-134096        04-01-2013          05-01-2013
## 27 27 TEC-MA-10004956 MX-2013-134096        04-01-2013          05-01-2013
## 28 28 OFF-SU-10003474 MX-2013-134096        04-01-2013          05-01-2013
## 29 29 TEC-AC-10001830 MX-2013-134096        04-01-2013          05-01-2013
## 30 30 OFF-BI-10002075 MX-2013-134096        04-01-2013          05-01-2013
## 31 31 OFF-EN-10000761 US-2013-126655        04-01-2013          05-01-2013
## 32 32 FUR-FU-10003066 US-2013-126655        04-01-2013          05-01-2013
## 33 33 OFF-EN-10000075 US-2013-126655        04-01-2013          05-01-2013
## 34 34 OFF-EN-10002226 US-2013-126655        04-01-2013          04-01-2013
## 35 35 FUR-CH-10002132 MX-2013-167759        04-01-2013          04-01-2013
## 36 36 OFF-EN-10001375 MX-2013-134096        04-01-2013          05-01-2013
## 37 37 TEC-MA-10004956 MX-2013-134096        04-01-2013          05-01-2013
## 38 38 OFF-SU-10003474 MX-2013-134096        04-01-2013          05-01-2013
## 39 39 TEC-AC-10001830 MX-2013-134096        04-01-2013          05-01-2013
## 40 40 OFF-BI-10002075 MX-2013-134096        04-01-2013          05-01-2013
## 41 41 OFF-FA-10002526 MX-2013-156335        04-01-2013          05-01-2013
## 42 42 OFF-EN-10001375 MX-2013-134096        04-01-2013          05-01-2013
## 43 43 TEC-MA-10004956 MX-2013-134096        04-01-2013          05-01-2013
## 44 44 OFF-SU-10003474 MX-2013-134096        05-01-2013          06-01-2013
## 45 45 TEC-AC-10001830 MX-2013-134096        05-01-2013          06-01-2013
## 46 46 OFF-BI-10002075 MX-2013-134096        05-01-2013          06-01-2013
## 47 47 OFF-EN-10001375 MX-2013-134096        05-01-2013          06-01-2013
## 48 48 TEC-MA-10004956 MX-2013-134096        05-01-2013          06-01-2013
## 49 49 OFF-SU-10003474 MX-2013-134096        05-01-2013          06-01-2013
## 50 50 OFF-LA-10002782 MX-2014-143658        03-01-2013          03-01-2013
##    Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type
## 1     341.520        2     683.040             10      STANDARD
## 2     252.160        8    2017.280              2      PRIORITY
## 3     193.280        2     386.560              3      PRIORITY
## 4      35.440        4     141.760              4      PRIORITY
## 5      71.600        2     143.200              5      PRIORITY
## 6      56.120        2     112.240              6      PRIORITY
## 7      56.120        2     112.240              7      STANDARD
## 8     132.640        4     530.560             15      STANDARD
```

17

```
## 9      12.940   1     12.940    16      STANDARD
## 10     18.840   2     37.280    17      STANDARD
## 11    308.280   7   2157.560    18      STANDARD
## 12     40.176   2     79.952    19      STANDARD
## 13      8.784   3     25.952    20      PRIORITY
## 14    273.472   4   1093.688    21      PRIORITY
## 15     56.120   2    112.240     7      STANDARD
## 16    344.640   3   1033.920     8      STANDARD
## 17     97.360   4    389.440     9      STANDARD
## 18    341.520   2    683.040    10      STANDARD
## 19     12.060   3     36.180    11      STANDARD
## 20     20.760   3     62.280    12      STANDARD
## 21     56.120   2    112.240     7      STANDARD
## 22    344.640   3   1033.920     8      STANDARD
## 23     97.360   4    389.440     9      STANDARD
## 24    341.520   2    683.040    10      STANDARD
## 25     12.060   3     36.180    11      STANDARD
## 26     56.120   2    112.240     7      STANDARD
## 27    344.640   3   1033.920     8      STANDARD
## 28     97.360   4    389.440     9      STANDARD
## 29    341.520   2    683.040    10      STANDARD
## 30     12.060   3     36.180    11      STANDARD
## 31     18.840   2     37.280    17      STANDARD
## 32    308.280   7   2157.560    18      STANDARD
## 33     40.176   2     79.952    19      STANDARD
## 34      8.784   3     25.952    20      PRIORITY
## 35    273.472   4   1093.688    21      PRIORITY
## 36     56.120   2    112.240     7      STANDARD
## 37    344.640   3   1033.920     8      STANDARD
## 38     97.360   4    389.440     9      STANDARD
## 39    341.520   2    683.040    10      STANDARD
## 40     12.060   3     36.180    11      STANDARD
## 41     20.760   3     62.280    12      STANDARD
## 42     56.120   2    112.240     7      STANDARD
## 43    344.640   3   1033.920     8      STANDARD
## 44     97.360   4    389.440     9      STANDARD
## 45    341.520   2    683.040    10      STANDARD
## 46     12.060   3     36.180    11      STANDARD
## 47     56.120   2    112.240     7      STANDARD
## 48    344.640   3   1033.920     8      STANDARD
## 49     97.360   4    389.440     9      STANDARD
## 50     13.080   3     39.240     1      PRIORITY
##              Category
## 1        Technology
## 2         Furniture
## 3         Furniture
## 4   Office Supplies
## 5   Office Supplies
## 6   Office Supplies
## 7   Office Supplies
## 8   Office Supplies
## 9   Office Supplies
## 10  Office Supplies
## 11        Furniture
```

```
## 12 Office Supplies
## 13 Office Supplies
## 14          Furniture
## 15 Office Supplies
## 16         Technology
## 17 Office Supplies
## 18         Technology
## 19 Office Supplies
## 20 Office Supplies
## 21 Office Supplies
## 22         Technology
## 23 Office Supplies
## 24         Technology
## 25 Office Supplies
## 26 Office Supplies
## 27         Technology
## 28 Office Supplies
## 29         Technology
## 30 Office Supplies
## 31 Office Supplies
## 32          Furniture
## 33 Office Supplies
## 34 Office Supplies
## 35          Furniture
## 36 Office Supplies
## 37         Technology
## 38 Office Supplies
## 39         Technology
## 40 Office Supplies
## 41 Office Supplies
## 42 Office Supplies
## 43         Technology
## 44 Office Supplies
## 45         Technology
## 46 Office Supplies
## 47 Office Supplies
## 48         Technology
## 49 Office Supplies
## 50 Office Supplies
```

COLUMN BIND- Data Frame

```
# Code Section -17

summary(df_1)
```

```
##        X.              Product.Name           Prod.ID     Date.of.Invoice
##   Min.   : 1    OFF-EN-10001375: 2    MX-2012-155047:5   01-01-2013:13
##   1st Qu.: 7    FUR-BO-10002352: 1    MX-2013-134096:5   02-01-2013:12
##   Median :13    FUR-CH-10002132: 1    US-2013-126655:4
##   Mean   :13    FUR-CH-10002846: 1    MX-2013-156335:2
##   3rd Qu.:19    FUR-FU-10003066: 1    MX-2013-163139:2
```

```
## Max.     :25  FUR-FU-10004015: 1    MX-2014-135706:2
##               (Other)         :18   (Other)         :5
##    Date.of.Shipping   Cost.Price       Quantity      Sales.Price
## 02-01-2013:15      Min.   :  8.784   Min.   :1.0   Min.   :  12.94
## 03-01-2013: 8      1st Qu.: 27.000   1st Qu.:2.0   1st Qu.:  62.28
## 04-01-2013: 2      Median : 71.600   Median :3.0   Median : 181.58
##                    Mean   :122.231   Mean   :3.6   Mean   : 555.18
##                    3rd Qu.:207.000   3rd Qu.:4.0   3rd Qu.: 842.56
##                    Max.   :344.640   Max.   :9.0   Max.   :2157.56
##
##  Shipping.Index   Shipping.Type          Category
##  Min.   : 1      PRIORITY:12    Furniture       : 5
##  1st Qu.: 7      STANDARD:13    Office Supplies:17
##  Median :13                     Technology      : 3
##  Mean   :13
##  3rd Qu.:19
##  Max.   :25
##
```

```r
summary(df_2)
```

```
##        X.                Product.Name          Prod.ID
##  Min.   : 1.00   OFF-EN-10001375: 8   MX-2012-155047: 5
##  1st Qu.:13.25   OFF-SU-10003474: 6   MX-2013-134096:30
##  Median :25.50   TEC-AC-10001830: 6   MX-2013-156335: 2
##  Mean   :25.50   TEC-MA-10004956: 6   MX-2013-167759: 2
##  3rd Qu.:37.75   OFF-BI-10002075: 5   MX-2014-135706: 2
##  Max.   :50.00   FUR-CH-10002132: 2   MX-2014-143658: 1
##                  (Other)        :17   US-2013-126655: 8
##    Date.of.Invoice   Date.of.Shipping    Cost.Price       Quantity
## 03-01-2013:20      03-01-2013: 8     Min.   :  8.784   Min.   :1.00
## 04-01-2013:23      04-01-2013:14     1st Qu.: 24.430   1st Qu.:2.00
## 05-01-2013: 7      05-01-2013:21     Median : 84.480   Median :3.00
##                    06-01-2013: 7     Mean   :145.243   Mean   :3.02
##                                      3rd Qu.:308.280   3rd Qu.:3.75
##                                      Max.   :344.640   Max.   :8.00
##
##   Sales.Price     Shipping.Index   Shipping.Type          Category
##  Min.   :  12.94   Min.   : 1.00   PRIORITY:10    Furniture       : 6
##  1st Qu.:  66.70   1st Qu.: 7.25   STANDARD:40    Office Supplies:32
##  Median : 264.88   Median : 9.50                  Technology      :12
##  Mean   : 478.04   Mean   :10.64
##  3rd Qu.: 683.04   3rd Qu.:12.00
##  Max.   :2157.56   Max.   :21.00
##
```

```r
df_3_C <- cbind(df_1,df_2) ## Column Bind is not the correct operation to be done here
df_3_C
```

```
##    X.    Product.Name          Prod.ID Date.of.Invoice Date.of.Shipping
## 1   1 OFF-LA-10002782 MX-2014-143658      01-01-2013       02-01-2013
## 2   2 FUR-FU-10004015 MX-2012-155047      01-01-2013       02-01-2013
## 3   3 FUR-BO-10002352 MX-2012-155047      01-01-2013       02-01-2013
## 4   4 OFF-BI-10004428 MX-2012-155047      01-01-2013       02-01-2013
## 5   5 OFF-AR-10004594 MX-2012-155047      01-01-2013       02-01-2013
```

```
## 6   6 OFF-EN-10001375 MX-2012-155047          01-01-2013          02-01-2013
## 7   7 OFF-EN-10001375 MX-2013-134096          01-01-2013          02-01-2013
## 8   8 TEC-MA-10004956 MX-2013-134096          01-01-2013          02-01-2013
## 9   9 OFF-SU-10003474 MX-2013-134096          01-01-2013          02-01-2013
## 10 10 TEC-AC-10001830 MX-2013-134096          01-01-2013          02-01-2013
## 11 11 OFF-BI-10002075 MX-2013-134096          01-01-2013          02-01-2013
## 12 12 OFF-FA-10002526 MX-2013-156335          01-01-2013          02-01-2013
## 13 13 FUR-CH-10002846 MX-2013-156335          01-01-2013          02-01-2013
## 14 14 OFF-EN-10004100 MX-2014-121923          02-01-2013          04-01-2013
## 15 15 OFF-AR-10003914 MX-2014-135706          02-01-2013          03-01-2013
## 16 16 OFF-FA-10000038 MX-2014-135706          02-01-2013          03-01-2013
## 17 17 OFF-EN-10000761 US-2013-126655          02-01-2013          03-01-2013
## 18 18 FUR-FU-10003066 US-2013-126655          02-01-2013          03-01-2013
## 19 19 OFF-EN-10000075 US-2013-126655          02-01-2013          03-01-2013
## 20 20 OFF-EN-10002226 US-2013-126655          02-01-2013          03-01-2013
## 21 21 FUR-CH-10002132 MX-2013-167759          02-01-2013          04-01-2013
## 22 22 TEC-AC-10002749 MX-2013-163139          02-01-2013          02-01-2013
## 23 23 OFF-SU-10000066 MX-2013-163139          02-01-2013          02-01-2013
## 24 24 OFF-BI-10003934 US-2014-119753          02-01-2013          03-01-2013
## 25 25 OFF-BI-10003932 US-2012-133970          02-01-2013          03-01-2013
## 26  1 OFF-LA-10002782 MX-2014-143658          01-01-2013          02-01-2013
## 27  2 FUR-FU-10004015 MX-2012-155047          01-01-2013          02-01-2013
## 28  3 FUR-BO-10002352 MX-2012-155047          01-01-2013          02-01-2013
## 29  4 OFF-BI-10004428 MX-2012-155047          01-01-2013          02-01-2013
## 30  5 OFF-AR-10004594 MX-2012-155047          01-01-2013          02-01-2013
## 31  6 OFF-EN-10001375 MX-2012-155047          01-01-2013          02-01-2013
## 32  7 OFF-EN-10001375 MX-2013-134096          01-01-2013          02-01-2013
## 33  8 TEC-MA-10004956 MX-2013-134096          01-01-2013          02-01-2013
## 34  9 OFF-SU-10003474 MX-2013-134096          01-01-2013          02-01-2013
## 35 10 TEC-AC-10001830 MX-2013-134096          01-01-2013          02-01-2013
## 36 11 OFF-BI-10002075 MX-2013-134096          01-01-2013          02-01-2013
## 37 12 OFF-FA-10002526 MX-2013-156335          01-01-2013          02-01-2013
## 38 13 FUR-CH-10002846 MX-2013-156335          01-01-2013          02-01-2013
## 39 14 OFF-EN-10004100 MX-2014-121923          02-01-2013          04-01-2013
## 40 15 OFF-AR-10003914 MX-2014-135706          02-01-2013          03-01-2013
## 41 16 OFF-FA-10000038 MX-2014-135706          02-01-2013          03-01-2013
## 42 17 OFF-EN-10000761 US-2013-126655          02-01-2013          03-01-2013
## 43 18 FUR-FU-10003066 US-2013-126655          02-01-2013          03-01-2013
## 44 19 OFF-EN-10000075 US-2013-126655          02-01-2013          03-01-2013
## 45 20 OFF-EN-10002226 US-2013-126655          02-01-2013          03-01-2013
## 46 21 FUR-CH-10002132 MX-2013-167759          02-01-2013          04-01-2013
## 47 22 TEC-AC-10002749 MX-2013-163139          02-01-2013          02-01-2013
## 48 23 OFF-SU-10000066 MX-2013-163139          02-01-2013          02-01-2013
## 49 24 OFF-BI-10003934 US-2014-119753          02-01-2013          03-01-2013
## 50 25 OFF-BI-10003932 US-2012-133970          02-01-2013          03-01-2013
##    Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type
## 1      13.080        3      39.240              1      PRIORITY
## 2     252.160        8    2017.280              2      PRIORITY
## 3     193.280        2     386.560              3      PRIORITY
## 4      35.440        4     141.760              4      PRIORITY
## 5      71.600        2     143.200              5      PRIORITY
## 6      56.120        2     112.240              6      PRIORITY
## 7      56.120        2     112.240              7      STANDARD
## 8     344.640        3    1033.920              8      STANDARD
```

```
## 9      97.360      4      389.440          9      STANDARD
## 10    341.520      2      683.040         10      STANDARD
## 11     12.060      3       36.180         11      STANDARD
## 12     20.760      3       62.280         12      STANDARD
## 13    210.640      4      842.560         13      STANDARD
## 14     80.100      3      240.300         14      STANDARD
## 15    132.640      4      530.560         15      STANDARD
## 16     12.940      1       12.940         16      STANDARD
## 17     18.840      2       37.280         17      STANDARD
## 18    308.280      7     2157.560         18      STANDARD
## 19     40.176      2       79.952         19      STANDARD
## 20      8.784      3       25.952         20      PRIORITY
## 21    273.472      4     1093.688         21      PRIORITY
## 22     27.000      1       27.000         22      PRIORITY
## 23    207.000      9     1863.000         23      PRIORITY
## 24     60.660      3      181.580         24      PRIORITY
## 25    181.116      9     1629.644         25      PRIORITY
## 26     13.080      3       39.240          1      PRIORITY
## 27    252.160      8     2017.280          2      PRIORITY
## 28    193.280      2      386.560          3      PRIORITY
## 29     35.440      4      141.760          4      PRIORITY
## 30     71.600      2      143.200          5      PRIORITY
## 31     56.120      2      112.240          6      PRIORITY
## 32     56.120      2      112.240          7      STANDARD
## 33    344.640      3     1033.920          8      STANDARD
## 34     97.360      4      389.440          9      STANDARD
## 35    341.520      2      683.040         10      STANDARD
## 36     12.060      3       36.180         11      STANDARD
## 37     20.760      3       62.280         12      STANDARD
## 38    210.640      4      842.560         13      STANDARD
## 39     80.100      3      240.300         14      STANDARD
## 40    132.640      4      530.560         15      STANDARD
## 41     12.940      1       12.940         16      STANDARD
## 42     18.840      2       37.280         17      STANDARD
## 43    308.280      7     2157.560         18      STANDARD
## 44     40.176      2       79.952         19      STANDARD
## 45      8.784      3       25.952         20      PRIORITY
## 46    273.472      4     1093.688         21      PRIORITY
## 47     27.000      1       27.000         22      PRIORITY
## 48    207.000      9     1863.000         23      PRIORITY
## 49     60.660      3      181.580         24      PRIORITY
## 50    181.116      9     1629.644         25      PRIORITY
##              Category X.    Product.Name      Prod.ID Date.of.Invoice
## 1  Office Supplies  1 TEC-AC-10001830 MX-2013-134096      05-01-2013
## 2          Furniture  2 FUR-FU-10004015 MX-2012-155047      03-01-2013
## 3          Furniture  3 FUR-BO-10002352 MX-2012-155047      03-01-2013
## 4  Office Supplies  4 OFF-BI-10004428 MX-2012-155047      03-01-2013
## 5  Office Supplies  5 OFF-AR-10004594 MX-2012-155047      03-01-2013
## 6  Office Supplies  6 OFF-EN-10001375 MX-2012-155047      03-01-2013
## 7  Office Supplies  7 OFF-EN-10001375 MX-2013-134096      03-01-2013
## 8          Technology  8 OFF-AR-10003914 MX-2014-135706      03-01-2013
## 9  Office Supplies  9 OFF-FA-10000038 MX-2014-135706      03-01-2013
## 10         Technology 10 OFF-EN-10000761 US-2013-126655      03-01-2013
## 11 Office Supplies 11 FUR-FU-10003066 US-2013-126655      03-01-2013
```

```
## 12 Office Supplies 12 OFF-EN-10000075 US-2013-126655          03-01-2013
## 13         Furniture 13 OFF-EN-10002226 US-2013-126655          03-01-2013
## 14 Office Supplies 14 FUR-CH-10002132 MX-2013-167759          03-01-2013
## 15 Office Supplies 15 OFF-EN-10001375 MX-2013-134096          03-01-2013
## 16 Office Supplies 16 TEC-MA-10004956 MX-2013-134096          03-01-2013
## 17 Office Supplies 17 OFF-SU-10003474 MX-2013-134096          03-01-2013
## 18         Furniture 18 TEC-AC-10001830 MX-2013-134096          03-01-2013
## 19 Office Supplies 19 OFF-BI-10002075 MX-2013-134096          03-01-2013
## 20 Office Supplies 20 OFF-FA-10002526 MX-2013-156335          03-01-2013
## 21         Furniture 21 OFF-EN-10001375 MX-2013-134096          04-01-2013
## 22        Technology 22 TEC-MA-10004956 MX-2013-134096          04-01-2013
## 23 Office Supplies 23 OFF-SU-10003474 MX-2013-134096          04-01-2013
## 24 Office Supplies 24 TEC-AC-10001830 MX-2013-134096          04-01-2013
## 25 Office Supplies 25 OFF-BI-10002075 MX-2013-134096          04-01-2013
## 26 Office Supplies 26 OFF-EN-10001375 MX-2013-134096          04-01-2013
## 27         Furniture 27 TEC-MA-10004956 MX-2013-134096          04-01-2013
## 28         Furniture 28 OFF-SU-10003474 MX-2013-134096          04-01-2013
## 29 Office Supplies 29 TEC-AC-10001830 MX-2013-134096          04-01-2013
## 30 Office Supplies 30 OFF-BI-10002075 MX-2013-134096          04-01-2013
## 31 Office Supplies 31 OFF-EN-10000761 US-2013-126655          04-01-2013
## 32 Office Supplies 32 FUR-FU-10003066 US-2013-126655          04-01-2013
## 33        Technology 33 OFF-EN-10000075 US-2013-126655          04-01-2013
## 34 Office Supplies 34 OFF-EN-10002226 US-2013-126655          04-01-2013
## 35        Technology 35 FUR-CH-10002132 MX-2013-167759          04-01-2013
## 36 Office Supplies 36 OFF-EN-10001375 MX-2013-134096          04-01-2013
## 37 Office Supplies 37 TEC-MA-10004956 MX-2013-134096          04-01-2013
## 38         Furniture 38 OFF-SU-10003474 MX-2013-134096          04-01-2013
## 39 Office Supplies 39 TEC-AC-10001830 MX-2013-134096          04-01-2013
## 40 Office Supplies 40 OFF-BI-10002075 MX-2013-134096          04-01-2013
## 41 Office Supplies 41 OFF-FA-10002526 MX-2013-156335          04-01-2013
## 42 Office Supplies 42 OFF-EN-10001375 MX-2013-134096          04-01-2013
## 43         Furniture 43 TEC-MA-10004956 MX-2013-134096          04-01-2013
## 44 Office Supplies 44 OFF-SU-10003474 MX-2013-134096          05-01-2013
## 45 Office Supplies 45 TEC-AC-10001830 MX-2013-134096          05-01-2013
## 46         Furniture 46 OFF-BI-10002075 MX-2013-134096          05-01-2013
## 47        Technology 47 OFF-EN-10001375 MX-2013-134096          05-01-2013
## 48 Office Supplies 48 TEC-MA-10004956 MX-2013-134096          05-01-2013
## 49 Office Supplies 49 OFF-SU-10003474 MX-2013-134096          05-01-2013
## 50 Office Supplies 50 OFF-LA-10002782 MX-2014-143658          03-01-2013
##    Date.of.Shipping Cost.Price Quantity Sales.Price Shipping.Index
## 1        06-01-2013    341.520        2     683.040             10
## 2        03-01-2013    252.160        8    2017.280              2
## 3        03-01-2013    193.280        2     386.560              3
## 4        03-01-2013     35.440        4     141.760              4
## 5        03-01-2013     71.600        2     143.200              5
## 6        03-01-2013     56.120        2     112.240              6
## 7        04-01-2013     56.120        2     112.240              7
## 8        04-01-2013    132.640        4     530.560             15
## 9        04-01-2013     12.940        1      12.940             16
## 10       04-01-2013     18.840        2      37.280             17
## 11       04-01-2013    308.280        7    2157.560             18
## 12       04-01-2013     40.176        2      79.952             19
## 13       03-01-2013      8.784        3      25.952             20
## 14       03-01-2013    273.472        4    1093.688             21
```

```
## 15       04-01-2013      56.120    2     112.240              7
## 16       04-01-2013     344.640    3    1033.920              8
## 17       04-01-2013      97.360    4     389.440              9
## 18       04-01-2013     341.520    2     683.040             10
## 19       04-01-2013      12.060    3      36.180             11
## 20       04-01-2013      20.760    3      62.280             12
## 21       05-01-2013      56.120    2     112.240              7
## 22       05-01-2013     344.640    3    1033.920              8
## 23       05-01-2013      97.360    4     389.440              9
## 24       05-01-2013     341.520    2     683.040             10
## 25       05-01-2013      12.060    3      36.180             11
## 26       05-01-2013      56.120    2     112.240              7
## 27       05-01-2013     344.640    3    1033.920              8
## 28       05-01-2013      97.360    4     389.440              9
## 29       05-01-2013     341.520    2     683.040             10
## 30       05-01-2013      12.060    3      36.180             11
## 31       05-01-2013      18.840    2      37.280             17
## 32       05-01-2013     308.280    7    2157.560             18
## 33       05-01-2013      40.176    2      79.952             19
## 34       04-01-2013       8.784    3      25.952             20
## 35       04-01-2013     273.472    4    1093.688             21
## 36       05-01-2013      56.120    2     112.240              7
## 37       05-01-2013     344.640    3    1033.920              8
## 38       05-01-2013      97.360    4     389.440              9
## 39       05-01-2013     341.520    2     683.040             10
## 40       05-01-2013      12.060    3      36.180             11
## 41       05-01-2013      20.760    3      62.280             12
## 42       05-01-2013      56.120    2     112.240              7
## 43       05-01-2013     344.640    3    1033.920              8
## 44       06-01-2013      97.360    4     389.440              9
## 45       06-01-2013     341.520    2     683.040             10
## 46       06-01-2013      12.060    3      36.180             11
## 47       06-01-2013      56.120    2     112.240              7
## 48       06-01-2013     344.640    3    1033.920              8
## 49       06-01-2013      97.360    4     389.440              9
## 50       03-01-2013      13.080    3      39.240              1
##     Shipping.Type        Category
## 1        STANDARD      Technology
## 2        PRIORITY       Furniture
## 3        PRIORITY       Furniture
## 4        PRIORITY Office Supplies
## 5        PRIORITY Office Supplies
## 6        PRIORITY Office Supplies
## 7        STANDARD Office Supplies
## 8        STANDARD Office Supplies
## 9        STANDARD Office Supplies
## 10       STANDARD Office Supplies
## 11       STANDARD       Furniture
## 12       STANDARD Office Supplies
## 13       PRIORITY Office Supplies
## 14       PRIORITY       Furniture
## 15       STANDARD Office Supplies
## 16       STANDARD      Technology
## 17       STANDARD Office Supplies
```

```
## 18       STANDARD      Technology
## 19       STANDARD Office Supplies
## 20       STANDARD Office Supplies
## 21       STANDARD Office Supplies
## 22       STANDARD      Technology
## 23       STANDARD Office Supplies
## 24       STANDARD      Technology
## 25       STANDARD Office Supplies
## 26       STANDARD Office Supplies
## 27       STANDARD      Technology
## 28       STANDARD Office Supplies
## 29       STANDARD      Technology
## 30       STANDARD Office Supplies
## 31       STANDARD Office Supplies
## 32       STANDARD      Furniture
## 33       STANDARD Office Supplies
## 34       PRIORITY Office Supplies
## 35       PRIORITY      Furniture
## 36       STANDARD Office Supplies
## 37       STANDARD      Technology
## 38       STANDARD Office Supplies
## 39       STANDARD      Technology
## 40       STANDARD Office Supplies
## 41       STANDARD Office Supplies
## 42       STANDARD Office Supplies
## 43       STANDARD      Technology
## 44       STANDARD Office Supplies
## 45       STANDARD      Technology
## 46       STANDARD Office Supplies
## 47       STANDARD Office Supplies
## 48       STANDARD      Technology
## 49       STANDARD Office Supplies
## 50       PRIORITY Office Supplies
```

```r
summary(df_3_C)
```

```
##       X.                Product.Name         Prod.ID      Date.of.Invoice
##  Min.   : 1   OFF-EN-10001375: 4   MX-2012-155047:10   01-01-2013:26
##  1st Qu.: 7   FUR-BO-10002352: 2   MX-2013-134096:10   02-01-2013:24
##  Median :13   FUR-CH-10002132: 2   US-2013-126655: 8
##  Mean   :13   FUR-CH-10002846: 2   MX-2013-156335: 4
##  3rd Qu.:19   FUR-FU-10003066: 2   MX-2013-163139: 4
##  Max.   :25   FUR-FU-10004015: 2   MX-2014-135706: 4
##               (Other)        :36   (Other)       :10
##   Date.of.Shipping   Cost.Price        Quantity    Sales.Price
##  02-01-2013:30   Min.   :  8.784   Min.   :1.0   Min.   :  12.94
##  03-01-2013:16   1st Qu.: 27.000   1st Qu.:2.0   1st Qu.:  62.28
##  04-01-2013: 4   Median : 71.600   Median :3.0   Median : 181.58
##                  Mean   :122.231   Mean   :3.6   Mean   : 555.18
##                  3rd Qu.:207.000   3rd Qu.:4.0   3rd Qu.: 842.56
##                  Max.   :344.640   Max.   :9.0   Max.   :2157.56
##
##  Shipping.Index  Shipping.Type          Category            X.
##  Min.   : 1      PRIORITY:24   Furniture       :10   Min.   : 1.00
##  1st Qu.: 7      STANDARD:26   Office Supplies:34   1st Qu.:13.25
```

```
##  Median :13                 Technology   : 6   Median :25.50
##  Mean   :13                                    Mean   :25.50
##  3rd Qu.:19                                    3rd Qu.:37.75
##  Max.   :25                                    Max.   :50.00
##
##            Product.Name            Prod.ID     Date.of.Invoice
##  OFF-EN-10001375: 8    MX-2012-155047: 5    03-01-2013:20
##  OFF-SU-10003474: 6    MX-2013-134096:30    04-01-2013:23
##  TEC-AC-10001830: 6    MX-2013-156335: 2    05-01-2013: 7
##  TEC-MA-10004956: 6    MX-2013-167759: 2
##  OFF-BI-10002075: 5    MX-2014-135706: 2
##  FUR-CH-10002132: 2    MX-2014-143658: 1
##  (Other)        :17    US-2013-126655: 8
##    Date.of.Shipping   Cost.Price        Quantity       Sales.Price
##  03-01-2013: 8     Min.   :  8.784   Min.   :1.00   Min.   :  12.94
##  04-01-2013:14     1st Qu.: 24.430   1st Qu.:2.00   1st Qu.:  66.70
##  05-01-2013:21     Median : 84.480   Median :3.00   Median : 264.88
##  06-01-2013: 7     Mean   :145.243   Mean   :3.02   Mean   : 478.04
##                    3rd Qu.:308.280   3rd Qu.:3.75   3rd Qu.: 683.04
##                    Max.   :344.640   Max.   :8.00   Max.   :2157.56
##
##  Shipping.Index    Shipping.Type          Category
##  Min.   : 1.00   PRIORITY:10   Furniture       : 6
##  1st Qu.: 7.25   STANDARD:40   Office Supplies:32
##  Median : 9.50                 Technology     :12
##  Mean   :10.64
##  3rd Qu.:12.00
##  Max.   :21.00
##
## R does Not Complaint as such but this DF is rendered useless ...
```

ROW BIND - Data Frame

```
# Code Section -18

df_3_R <- rbind(df_2,df_1)
df_3_R
```

```
##     X.     Product.Name          Prod.ID Date.of.Invoice Date.of.Shipping
## 1   1 TEC-AC-10001830 MX-2013-134096       05-01-2013       06-01-2013
## 2   2 FUR-FU-10004015 MX-2012-155047       03-01-2013       03-01-2013
## 3   3 FUR-BO-10002352 MX-2012-155047       03-01-2013       03-01-2013
## 4   4 OFF-BI-10004428 MX-2012-155047       03-01-2013       03-01-2013
## 5   5 OFF-AR-10004594 MX-2012-155047       03-01-2013       03-01-2013
## 6   6 OFF-EN-10001375 MX-2012-155047       03-01-2013       03-01-2013
## 7   7 OFF-EN-10001375 MX-2013-134096       03-01-2013       04-01-2013
## 8   8 OFF-AR-10003914 MX-2014-135706       03-01-2013       04-01-2013
## 9   9 OFF-FA-10000038 MX-2014-135706       03-01-2013       04-01-2013
## 10 10 OFF-EN-10000761 US-2013-126655       03-01-2013       04-01-2013
## 11 11 FUR-FU-10003066 US-2013-126655       03-01-2013       04-01-2013
## 12 12 OFF-EN-10000075 US-2013-126655       03-01-2013       04-01-2013
## 13 13 OFF-EN-10002226 US-2013-126655       03-01-2013       03-01-2013
```

```
## 14 14 FUR-CH-10002132 MX-2013-167759      03-01-2013      03-01-2013
## 15 15 OFF-EN-10001375 MX-2013-134096      03-01-2013      04-01-2013
## 16 16 TEC-MA-10004956 MX-2013-134096      03-01-2013      04-01-2013
## 17 17 OFF-SU-10003474 MX-2013-134096      03-01-2013      04-01-2013
## 18 18 TEC-AC-10001830 MX-2013-134096      03-01-2013      04-01-2013
## 19 19 OFF-BI-10002075 MX-2013-134096      03-01-2013      04-01-2013
## 20 20 OFF-FA-10002526 MX-2013-156335      03-01-2013      04-01-2013
## 21 21 OFF-EN-10001375 MX-2013-134096      04-01-2013      05-01-2013
## 22 22 TEC-MA-10004956 MX-2013-134096      04-01-2013      05-01-2013
## 23 23 OFF-SU-10003474 MX-2013-134096      04-01-2013      05-01-2013
## 24 24 TEC-AC-10001830 MX-2013-134096      04-01-2013      05-01-2013
## 25 25 OFF-BI-10002075 MX-2013-134096      04-01-2013      05-01-2013
## 26 26 OFF-EN-10001375 MX-2013-134096      04-01-2013      05-01-2013
## 27 27 TEC-MA-10004956 MX-2013-134096      04-01-2013      05-01-2013
## 28 28 OFF-SU-10003474 MX-2013-134096      04-01-2013      05-01-2013
## 29 29 TEC-AC-10001830 MX-2013-134096      04-01-2013      05-01-2013
## 30 30 OFF-BI-10002075 MX-2013-134096      04-01-2013      05-01-2013
## 31 31 OFF-EN-10000761 US-2013-126655      04-01-2013      05-01-2013
## 32 32 FUR-FU-10003066 US-2013-126655      04-01-2013      05-01-2013
## 33 33 OFF-EN-10000075 US-2013-126655      04-01-2013      05-01-2013
## 34 34 OFF-EN-10002226 US-2013-126655      04-01-2013      04-01-2013
## 35 35 FUR-CH-10002132 MX-2013-167759      04-01-2013      04-01-2013
## 36 36 OFF-EN-10001375 MX-2013-134096      04-01-2013      05-01-2013
## 37 37 TEC-MA-10004956 MX-2013-134096      04-01-2013      05-01-2013
## 38 38 OFF-SU-10003474 MX-2013-134096      04-01-2013      05-01-2013
## 39 39 TEC-AC-10001830 MX-2013-134096      04-01-2013      05-01-2013
## 40 40 OFF-BI-10002075 MX-2013-134096      04-01-2013      05-01-2013
## 41 41 OFF-FA-10002526 MX-2013-156335      04-01-2013      05-01-2013
## 42 42 OFF-EN-10001375 MX-2013-134096      04-01-2013      05-01-2013
## 43 43 TEC-MA-10004956 MX-2013-134096      04-01-2013      05-01-2013
## 44 44 OFF-SU-10003474 MX-2013-134096      05-01-2013      06-01-2013
## 45 45 TEC-AC-10001830 MX-2013-134096      05-01-2013      06-01-2013
## 46 46 OFF-BI-10002075 MX-2013-134096      05-01-2013      06-01-2013
## 47 47 OFF-EN-10001375 MX-2013-134096      05-01-2013      06-01-2013
## 48 48 TEC-MA-10004956 MX-2013-134096      05-01-2013      06-01-2013
## 49 49 OFF-SU-10003474 MX-2013-134096      05-01-2013      06-01-2013
## 50 50 OFF-LA-10002782 MX-2014-143658      03-01-2013      03-01-2013
## 51  1 OFF-LA-10002782 MX-2014-143658      01-01-2013      02-01-2013
## 52  2 FUR-FU-10004015 MX-2012-155047      01-01-2013      02-01-2013
## 53  3 FUR-BO-10002352 MX-2012-155047      01-01-2013      02-01-2013
## 54  4 OFF-BI-10004428 MX-2012-155047      01-01-2013      02-01-2013
## 55  5 OFF-AR-10004594 MX-2012-155047      01-01-2013      02-01-2013
## 56  6 OFF-EN-10001375 MX-2012-155047      01-01-2013      02-01-2013
## 57  7 OFF-EN-10001375 MX-2013-134096      01-01-2013      02-01-2013
## 58  8 TEC-MA-10004956 MX-2013-134096      01-01-2013      02-01-2013
## 59  9 OFF-SU-10003474 MX-2013-134096      01-01-2013      02-01-2013
## 60 10 TEC-AC-10001830 MX-2013-134096      01-01-2013      02-01-2013
## 61 11 OFF-BI-10002075 MX-2013-134096      01-01-2013      02-01-2013
## 62 12 OFF-FA-10002526 MX-2013-156335      01-01-2013      02-01-2013
## 63 13 FUR-CH-10002846 MX-2013-156335      01-01-2013      02-01-2013
## 64 14 OFF-EN-10004100 MX-2014-121923      02-01-2013      04-01-2013
## 65 15 OFF-AR-10003914 MX-2014-135706      02-01-2013      03-01-2013
## 66 16 OFF-FA-10000038 MX-2014-135706      02-01-2013      03-01-2013
## 67 17 OFF-EN-10000761 US-2013-126655      02-01-2013      03-01-2013
```

```
## 68 18 FUR-FU-10003066 US-2013-126655        02-01-2013        03-01-2013
## 69 19 OFF-EN-10000075 US-2013-126655        02-01-2013        03-01-2013
## 70 20 OFF-EN-10002226 US-2013-126655        02-01-2013        03-01-2013
## 71 21 FUR-CH-10002132 MX-2013-167759        02-01-2013        04-01-2013
## 72 22 TEC-AC-10002749 MX-2013-163139        02-01-2013        02-01-2013
## 73 23 OFF-SU-10000066 MX-2013-163139        02-01-2013        02-01-2013
## 74 24 OFF-BI-10003934 US-2014-119753        02-01-2013        03-01-2013
## 75 25 OFF-BI-10003932 US-2012-133970        02-01-2013        03-01-2013
##    Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type
## 1     341.520        2     683.040             10      STANDARD
## 2     252.160        8    2017.280              2      PRIORITY
## 3     193.280        2     386.560              3      PRIORITY
## 4      35.440        4     141.760              4      PRIORITY
## 5      71.600        2     143.200              5      PRIORITY
## 6      56.120        2     112.240              6      PRIORITY
## 7      56.120        2     112.240              7      STANDARD
## 8     132.640        4     530.560             15      STANDARD
## 9      12.940        1      12.940             16      STANDARD
## 10     18.840        2      37.280             17      STANDARD
## 11    308.280        7    2157.560             18      STANDARD
## 12     40.176        2      79.952             19      STANDARD
## 13      8.784        3      25.952             20      PRIORITY
## 14    273.472        4    1093.688             21      PRIORITY
## 15     56.120        2     112.240              7      STANDARD
## 16    344.640        3    1033.920              8      STANDARD
## 17     97.360        4     389.440              9      STANDARD
## 18    341.520        2     683.040             10      STANDARD
## 19     12.060        3      36.180             11      STANDARD
## 20     20.760        3      62.280             12      STANDARD
## 21     56.120        2     112.240              7      STANDARD
## 22    344.640        3    1033.920              8      STANDARD
## 23     97.360        4     389.440              9      STANDARD
## 24    341.520        2     683.040             10      STANDARD
## 25     12.060        3      36.180             11      STANDARD
## 26     56.120        2     112.240              7      STANDARD
## 27    344.640        3    1033.920              8      STANDARD
## 28     97.360        4     389.440              9      STANDARD
## 29    341.520        2     683.040             10      STANDARD
## 30     12.060        3      36.180             11      STANDARD
## 31     18.840        2      37.280             17      STANDARD
## 32    308.280        7    2157.560             18      STANDARD
## 33     40.176        2      79.952             19      STANDARD
## 34      8.784        3      25.952             20      PRIORITY
## 35    273.472        4    1093.688             21      PRIORITY
## 36     56.120        2     112.240              7      STANDARD
## 37    344.640        3    1033.920              8      STANDARD
## 38     97.360        4     389.440              9      STANDARD
## 39    341.520        2     683.040             10      STANDARD
## 40     12.060        3      36.180             11      STANDARD
## 41     20.760        3      62.280             12      STANDARD
## 42     56.120        2     112.240              7      STANDARD
## 43    344.640        3    1033.920              8      STANDARD
## 44     97.360        4     389.440              9      STANDARD
## 45    341.520        2     683.040             10      STANDARD
```

```
## 46     12.060        3       36.180          11          STANDARD
## 47     56.120        2      112.240           7          STANDARD
## 48    344.640        3     1033.920           8          STANDARD
## 49     97.360        4      389.440           9          STANDARD
## 50     13.080        3       39.240           1          PRIORITY
## 51     13.080        3       39.240           1          PRIORITY
## 52    252.160        8     2017.280           2          PRIORITY
## 53    193.280        2      386.560           3          PRIORITY
## 54     35.440        4      141.760           4          PRIORITY
## 55     71.600        2      143.200           5          PRIORITY
## 56     56.120        2      112.240           6          PRIORITY
## 57     56.120        2      112.240           7          STANDARD
## 58    344.640        3     1033.920           8          STANDARD
## 59     97.360        4      389.440           9          STANDARD
## 60    341.520        2      683.040          10          STANDARD
## 61     12.060        3       36.180          11          STANDARD
## 62     20.760        3       62.280          12          STANDARD
## 63    210.640        4      842.560          13          STANDARD
## 64     80.100        3      240.300          14          STANDARD
## 65    132.640        4      530.560          15          STANDARD
## 66     12.940        1       12.940          16          STANDARD
## 67     18.840        2       37.280          17          STANDARD
## 68    308.280        7     2157.560          18          STANDARD
## 69     40.176        2       79.952          19          STANDARD
## 70      8.784        3       25.952          20          PRIORITY
## 71    273.472        4     1093.688          21          PRIORITY
## 72     27.000        1       27.000          22          PRIORITY
## 73    207.000        9     1863.000          23          PRIORITY
## 74     60.660        3      181.580          24          PRIORITY
## 75    181.116        9     1629.644          25          PRIORITY
##              Category
## 1         Technology
## 2          Furniture
## 3          Furniture
## 4    Office Supplies
## 5    Office Supplies
## 6    Office Supplies
## 7    Office Supplies
## 8    Office Supplies
## 9    Office Supplies
## 10   Office Supplies
## 11         Furniture
## 12   Office Supplies
## 13   Office Supplies
## 14         Furniture
## 15   Office Supplies
## 16        Technology
## 17   Office Supplies
## 18        Technology
## 19   Office Supplies
## 20   Office Supplies
## 21   Office Supplies
## 22        Technology
## 23   Office Supplies
```

```
## 24        Technology
## 25 Office Supplies
## 26 Office Supplies
## 27        Technology
## 28 Office Supplies
## 29        Technology
## 30 Office Supplies
## 31 Office Supplies
## 32         Furniture
## 33 Office Supplies
## 34 Office Supplies
## 35         Furniture
## 36 Office Supplies
## 37        Technology
## 38 Office Supplies
## 39        Technology
## 40 Office Supplies
## 41 Office Supplies
## 42 Office Supplies
## 43        Technology
## 44 Office Supplies
## 45        Technology
## 46 Office Supplies
## 47 Office Supplies
## 48        Technology
## 49 Office Supplies
## 50 Office Supplies
## 51 Office Supplies
## 52         Furniture
## 53         Furniture
## 54 Office Supplies
## 55 Office Supplies
## 56 Office Supplies
## 57 Office Supplies
## 58        Technology
## 59 Office Supplies
## 60        Technology
## 61 Office Supplies
## 62 Office Supplies
## 63         Furniture
## 64 Office Supplies
## 65 Office Supplies
## 66 Office Supplies
## 67 Office Supplies
## 68         Furniture
## 69 Office Supplies
## 70 Office Supplies
## 71         Furniture
## 72        Technology
## 73 Office Supplies
## 74 Office Supplies
## 75 Office Supplies
```

```
# We can SUBSET Data within R Data Structures with custom and inbuilt functions
#

summary(df_3_R$Cost.Price)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.784  23.880  71.600 137.600 273.500 344.600
#
boxplot(df_3_R$Cost.Price,col="green",horizontal = TRUE)
```
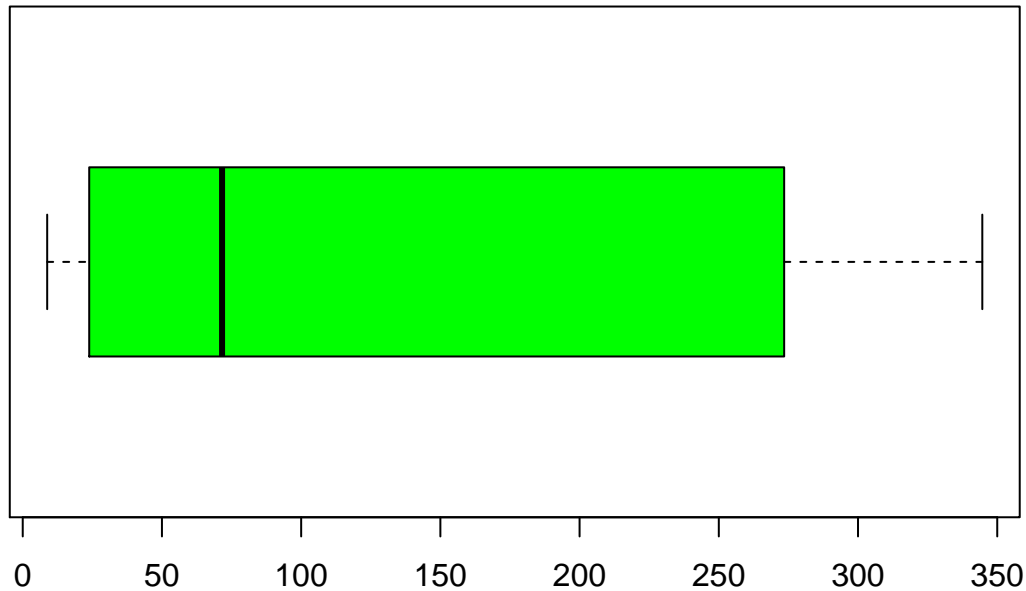


```
#
sub_df_1 <- subset(df_3_R,Cost.Price > 300)
#
summary(sub_df_1$Cost.Price) # As seen from the SUMMARY

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   308.3   341.5   341.5   336.9   344.6   344.6
# All ROWS or OBSERVATIONS with Cost.Price Values LESS THAN - 300
# have been dropped from the Sub-Set Data Frame.

# To SUBSET a Vector - we may simply-
#
summary(num_vector3)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.00   13.25   18.50   21.90   30.25   44.00
#
sub_num_vector3 <- num_vector3[-(1:3)]
# Creates a SUBSET of all ELEMENTS
# of num_vector3 - besides the FIRST THREE
#
summary(sub_num_vector3)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.00   12.50   14.00   20.29   24.00   44.00
```

31

describeBy— with Grouping variables

```
# Code Section -19


library("psych", lib.loc="~/R/x86_64-pc-linux-gnu-library/3.3")
library(psych)
#

#


#

detach("package:psych", unload=TRUE)
#?describeBy() # Seek HELP for the function - Uncomment this line.

#describeBy(df_3_C,df_3_C$Date.of.Invoice)

# In this case - Grouping Variable is - Date of Invoice .
# This Grouping variable has two Values here - 01-01-2013 and 02-01-2013

# IMPORTANT NOTE --- The DOLLAR SIGN in df_3_C$Date.of.Invoice , is used
# to access a certain variable within the DATA FRAME.

# Also note that the variables within the DATA FRAME will not be stored with the Labels
# as - is
```

Built in DATA Sets

The default R environment - comes bundled with a number of packages and data sets.

A package called DATASETS - contains a number of inbuilt data sets as seen in sections below -

```
# Code Section -20


library(help = "datasets")



#
# Seen below are the DataSets available within my installed version of R -
# yours may slightly differ :-
#
# AirPassengers          Monthly Airline Passenger Numbers 1949-1960
# BJsales                Sales Data with Leading Indicator
# BOD                    Biochemical Oxygen Demand
# CO2                    Carbon Dioxide Uptake in Grass Plants
# ChickWeight            Weight versus age of chicks on different diets
# DNase                  Elisa assay of DNase
# EuStockMarkets         Daily Closing Prices of Major European Stock
#                        Indices, 1991-1998
# Formaldehyde           Determination of Formaldehyde
# HairEyeColor           Hair and Eye Color of Statistics Students
# Harman23.cor           Harman Example 2.3
```

```
# Harman74.cor              Harman Example 7.4
# Indometh                  Pharmacokinetics of Indomethacin
# InsectSprays              Effectiveness of Insect Sprays
# JohnsonJohnson            Quarterly Earnings per Johnson & Johnson Share
# LakeHuron                 Level of Lake Huron 1875-1972
# LifeCycleSavings          Intercountry Life-Cycle Savings Data
# Loblolly                  Growth of Loblolly pine trees
# Nile                      Flow of the River Nile
# Orange                    Growth of Orange Trees
# OrchardSprays             Potency of Orchard Sprays
# PlantGrowth               Results from an Experiment on Plant Growth
# Puromycin                 Reaction Velocity of an Enzymatic Reaction
# Theoph                    Pharmacokinetics of Theophylline
# Titanic                   Survival of passengers on the Titanic
# ToothGrowth               The Effect of Vitamin C on Tooth Growth in
#                           Guinea Pigs
# UCBAdmissions             Student Admissions at UC Berkeley
# UKDriverDeaths            Road Casualties in Great Britain 1969-84
# UKLungDeaths              Monthly Deaths from Lung Diseases in the UK
# UKgas                     UK Quarterly Gas Consumption
# USAccDeaths               Accidental Deaths in the US 1973-1978
# USArrests                 Violent Crime Rates by US State
# USJudgeRatings            Lawyers' Ratings of State Judges in the US
#                           Superior Court
# USPersonalExpenditure     Personal Expenditure Data
# VADeaths                  Death Rates in Virginia (1940)
# WWWusage                  Internet Usage per Minute
# WorldPhones               The World's Telephones
# ability.cov               Ability and Intelligence Tests
# airmiles                  Passenger Miles on Commercial US Airlines,
#                           1937-1960
# airquality                New York Air Quality Measurements
# anscombe                  Anscombe's Quartet of 'Identical' Simple Linear
#                           Regressions
# attenu                    The Joyner-Boore Attenuation Data
# attitude                  The Chatterjee-Price Attitude Data
# austres                   Quarterly Time Series of the Number of
#                           Australian Residents
# beavers                   Body Temperature Series of Two Beavers
# cars                      Speed and Stopping Distances of Cars
# chickwts                  Chicken Weights by Feed Type
# co2                       Mauna Loa Atmospheric CO2 Concentration
# crimtab                   Student's 3000 Criminals Data
# datasets-package          The R Datasets Package
# discoveries               Yearly Numbers of Important Discoveries
# esoph                     Smoking, Alcohol and (O)esophageal Cancer
# euro                      Conversion Rates of Euro Currencies
# eurodist                  Distances Between European Cities and Between
#                           US Cities
# faithful                  Old Faithful Geyser Data
# freeny                    Freeny's Revenue Data
# infert                    Infertility after Spontaneous and Induced
#                           Abortion
```

```
# iris                  Edgar Anderson's Iris Data
# islands               Areas of the World's Major Landmasses
# lh                    Luteinizing Hormone in Blood Samples
# longley               Longley's Economic Regression Data
# lynx                  Annual Canadian Lynx trappings 1821-1934
# morley                Michelson Speed of Light Data
# mtcars                Motor Trend Car Road Tests
# nhtemp                Average Yearly Temperatures in New Haven
# nottem                Average Monthly Temperatures at Nottingham,
#                       1920-1939
# npk                   Classical N, P, K Factorial Experiment
# occupationalStatus    Occupational Status of Fathers and their Sons
# precip                Annual Precipitation in US Cities
# presidents            Quarterly Approval Ratings of US Presidents
# pressure              Vapor Pressure of Mercury as a Function of
#                       Temperature
# quakes                Locations of Earthquakes off Fiji
# randu                 Random Numbers from Congruential Generator
#                       RANDU
# rivers                Lengths of Major North American Rivers
# rock                  Measurements on Petroleum Rock Samples
# sleep                 Student's Sleep Data
# stackloss             Brownlee's Stack Loss Plant Data
# state                 US State Facts and Figures
# sunspot.month         Monthly Sunspot Data, from 1749 to "Present"
# sunspot.year          Yearly Sunspot Data, 1700-1988
# sunspots              Monthly Sunspot Numbers, 1749-1983
# swiss                 Swiss Fertility and Socioeconomic Indicators
#                       (1888) Data
# treering              Yearly Treering Data, -6000-1979
# trees                 Girth, Height and Volume for Black Cherry Trees
# uspop                 Populations Recorded by the US Census
# volcano               Topographic Information on Auckland's Maunga
#                       Whau Volcano
# warpbreaks            The Number of Breaks in Yarn during Weaving
# women                 Average Heights and Weights for American Women
#
```

ATTACH and DETACH Datasets

To OPERATE upon an inbuilt DATASET we need to LOAD it into R .

In lay-man terms , we may equate this to our earlier - read.csv() , but

there are subtle differences.

The ATTACH will pull in a Data Set from the DataSet Package into our R Search Path.

```
# Code Section -21

?attach()

# # Quoting below from inbuilt HELP -
# # The database is attached to the R search path. This means that the database is
```

```
# searched by R when evaluating a variable, so objects in the database can be accessed
# by simply giving their names.

?detach()

# # Quoting below from inbuilt HELP -
# # Detach a database, i.e., remove it from the search() path of available R objects.
# Usually this is either a data.frame which has been attached or a
# package which was attached by library.

# Another Function which will help automate is - zap()

# library(epicalc)
# ?zap()
# https://artax.karlin.mff.cuni.cz/r-help/library/epicalc/html/zap.html
```

```
# Code Section -22

search()
```

```
## [1] ".GlobalEnv"        "package:stats"     "package:graphics"
## [4] "package:grDevices" "package:utils"     "package:datasets"
## [7] "package:methods"   "Autoloads"         "package:base"
```

DATA_FRAME COLUMNS - The DF Column Vector

```
# Code Section -23

# Three basic methods to access a DF COLUMN - also called the
# DF Column Vector

print(df_1[[2]]) # Will Print 2nd COLUMN of the DF
```

```
##  [1] OFF-LA-10002782 FUR-FU-10004015 FUR-BO-10002352 OFF-BI-10004428
##  [5] OFF-AR-10004594 OFF-EN-10001375 OFF-EN-10001375 TEC-MA-10004956
##  [9] OFF-SU-10003474 TEC-AC-10001830 OFF-BI-10002075 OFF-FA-10002526
## [13] FUR-CH-10002846 OFF-EN-10004100 OFF-AR-10003914 OFF-FA-10000038
## [17] OFF-EN-10000761 FUR-FU-10003066 OFF-EN-10000075 OFF-EN-10002226
## [21] FUR-CH-10002132 TEC-AC-10002749 OFF-SU-10000066 OFF-BI-10003934
## [25] OFF-BI-10003932
## 24 Levels: FUR-BO-10002352 FUR-CH-10002132 ... TEC-MA-10004956
```

```
#
print(df_1[["Product.Name"]]) # Will Print the "NAMED"  COLUMN of the DF
```

```
##  [1] OFF-LA-10002782 FUR-FU-10004015 FUR-BO-10002352 OFF-BI-10004428
##  [5] OFF-AR-10004594 OFF-EN-10001375 OFF-EN-10001375 TEC-MA-10004956
##  [9] OFF-SU-10003474 TEC-AC-10001830 OFF-BI-10002075 OFF-FA-10002526
## [13] FUR-CH-10002846 OFF-EN-10004100 OFF-AR-10003914 OFF-FA-10000038
## [17] OFF-EN-10000761 FUR-FU-10003066 OFF-EN-10000075 OFF-EN-10002226
## [21] FUR-CH-10002132 TEC-AC-10002749 OFF-SU-10000066 OFF-BI-10003934
## [25] OFF-BI-10003932
## 24 Levels: FUR-BO-10002352 FUR-CH-10002132 ... TEC-MA-10004956
```

```
#
print(df_1$Product.Name) # Will Print the "NAMED" COLUMN of the DF
```

```
##  [1] OFF-LA-10002782 FUR-FU-10004015 FUR-BO-10002352 OFF-BI-10004428
##  [5] OFF-AR-10004594 OFF-EN-10001375 OFF-EN-10001375 TEC-MA-10004956
##  [9] OFF-SU-10003474 TEC-AC-10001830 OFF-BI-10002075 OFF-FA-10002526
## [13] FUR-CH-10002846 OFF-EN-10004100 OFF-AR-10003914 OFF-FA-10000038
## [17] OFF-EN-10000761 FUR-FU-10003066 OFF-EN-10000075 OFF-EN-10002226
## [21] FUR-CH-10002132 TEC-AC-10002749 OFF-SU-10000066 OFF-BI-10003934
## [25] OFF-BI-10003932
## 24 Levels: FUR-BO-10002352 FUR-CH-10002132 ... TEC-MA-10004956
```

DATA_FRAME - Row and Column Slices

```
# Code Section -24

# Slicing the DF COLUMNS and ROWS

print(df_1[2]) # Will Print 2nd COLUMN - ALL ROWS - of the DF as a SLICE.
```

```
##        Product.Name
## 1   OFF-LA-10002782
## 2   FUR-FU-10004015
## 3   FUR-BO-10002352
## 4   OFF-BI-10004428
## 5   OFF-AR-10004594
## 6   OFF-EN-10001375
## 7   OFF-EN-10001375
## 8   TEC-MA-10004956
## 9   OFF-SU-10003474
## 10 TEC-AC-10001830
## 11 OFF-BI-10002075
## 12 OFF-FA-10002526
## 13 FUR-CH-10002846
## 14 OFF-EN-10004100
## 15 OFF-AR-10003914
## 16 OFF-FA-10000038
## 17 OFF-EN-10000761
## 18 FUR-FU-10003066
## 19 OFF-EN-10000075
## 20 OFF-EN-10002226
## 21 FUR-CH-10002132
## 22 TEC-AC-10002749
## 23 OFF-SU-10000066
## 24 OFF-BI-10003934
## 25 OFF-BI-10003932
```

```
#
print(df_1[2,]) # Will Print 2nd ROW - ALL COLUMNS - of the DF as a SLICE.
```

```
##   X.    Product.Name        Prod.ID Date.of.Invoice Date.of.Shipping
## 2  2 FUR-FU-10004015 MX-2012-155047      01-01-2013       02-01-2013
##   Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type  Category
```

```
## 2      252.16       8    2017.28              2       PRIORITY Furniture
#
df_temp1 <- (df_1[2,])
df_temp1
```

```
##   X.    Product.Name        Prod.ID Date.of.Invoice Date.of.Shipping
## 2  2 FUR-FU-10004015 MX-2012-155047      01-01-2013       02-01-2013
##   Cost.Price Quantity Sales.Price Shipping.Index Shipping.Type   Category
## 2     252.16        8     2017.28              2      PRIORITY Furniture
```

```
View(df_temp1)
#
summary(df_temp1)
```

```
##       X.              Product.Name              Prod.ID    Date.of.Invoice
##  Min.   :2   FUR-FU-10004015:1   MX-2012-155047:1   01-01-2013:1
##  1st Qu.:2   FUR-BO-10002352:0   MX-2013-134096:0   02-01-2013:0
##  Median :2   FUR-CH-10002132:0   MX-2013-156335:0
##  Mean   :2   FUR-CH-10002846:0   MX-2013-163139:0
##  3rd Qu.:2   FUR-FU-10003066:0   MX-2013-167759:0
##  Max.   :2   OFF-AR-10003914:0   MX-2014-121923:0
##              (Other)        :0   (Other)        :0
##   Date.of.Shipping   Cost.Price       Quantity    Sales.Price
##  02-01-2013:1   Min.   :252.2   Min.   :8   Min.   :2017
##  03-01-2013:0   1st Qu.:252.2   1st Qu.:8   1st Qu.:2017
##  04-01-2013:0   Median :252.2   Median :8   Median :2017
##                 Mean   :252.2   Mean   :8   Mean   :2017
##                 3rd Qu.:252.2   3rd Qu.:8   3rd Qu.:2017
##                 Max.   :252.2   Max.   :8   Max.   :2017
##
##  Shipping.Index  Shipping.Type          Category
##  Min.   :2   PRIORITY:1   Furniture       :1
##  1st Qu.:2   STANDARD:0   Office Supplies:0
##  Median :2                Technology      :0
##  Mean   :2
##  3rd Qu.:2
##  Max.   :2
##
```

```
print(df_1[c(2,4,6)]) # Using an INDEX VECTOR  - created with COMBINE Function.
```

```
##        Product.Name Date.of.Invoice Cost.Price
## 1  OFF-LA-10002782      01-01-2013     13.080
## 2  FUR-FU-10004015      01-01-2013    252.160
## 3  FUR-BO-10002352      01-01-2013    193.280
## 4  OFF-BI-10004428      01-01-2013     35.440
## 5  OFF-AR-10004594      01-01-2013     71.600
## 6  OFF-EN-10001375      01-01-2013     56.120
## 7  OFF-EN-10001375      01-01-2013     56.120
## 8  TEC-MA-10004956      01-01-2013    344.640
## 9  OFF-SU-10003474      01-01-2013     97.360
## 10 TEC-AC-10001830      01-01-2013    341.520
## 11 OFF-BI-10002075      01-01-2013     12.060
## 12 OFF-FA-10002526      01-01-2013     20.760
## 13 FUR-CH-10002846      01-01-2013    210.640
```

```
## 14 OFF-EN-10004100        02-01-2013      80.100
## 15 OFF-AR-10003914        02-01-2013     132.640
## 16 OFF-FA-10000038        02-01-2013      12.940
## 17 OFF-EN-10000761        02-01-2013      18.840
## 18 FUR-FU-10003066        02-01-2013     308.280
## 19 OFF-EN-10000075        02-01-2013      40.176
## 20 OFF-EN-10002226        02-01-2013       8.784
## 21 FUR-CH-10002132        02-01-2013     273.472
## 22 TEC-AC-10002749        02-01-2013      27.000
## 23 OFF-SU-10000066        02-01-2013     207.000
## 24 OFF-BI-10003934        02-01-2013      60.660
## 25 OFF-BI-10003932        02-01-2013     181.116
```
```
# To access MULTIPLE COLUMNS - 2,4 and 6 of the DF.
```

Lets create a New Directory - Files Folder from within R and house our basic introductory graphs there - we use - dir.create()

```
# Code Section -25


#dir.create("/home/dhankar/Desktop/R_Own/Plots/Intro_2/", recursive=TRUE) # to be RUN only Once
```

Plots and Graphs - the Basics

Over the next couple of sections we shall preview the basics of Plots and Graphs.

If we were to execute the code from the following sections , in the Console- the Graphs would not appear inline but would appear within the PLOTS pane of R Studio or in certain cases as POP Up's.

HISTOGRAM an Introduction

```
# Code Section -26
# So what exactly is a HISTOGRAM -
# " estimate of the probability distribution of a continuous variable (quantitative variable)"
# Source -- Wiki -- https://en.wikipedia.org/wiki/Histogram


hist(df_1$Cost.Price,col.main="blue",col = "green")
```

Histogram of df_1$Cost.Price
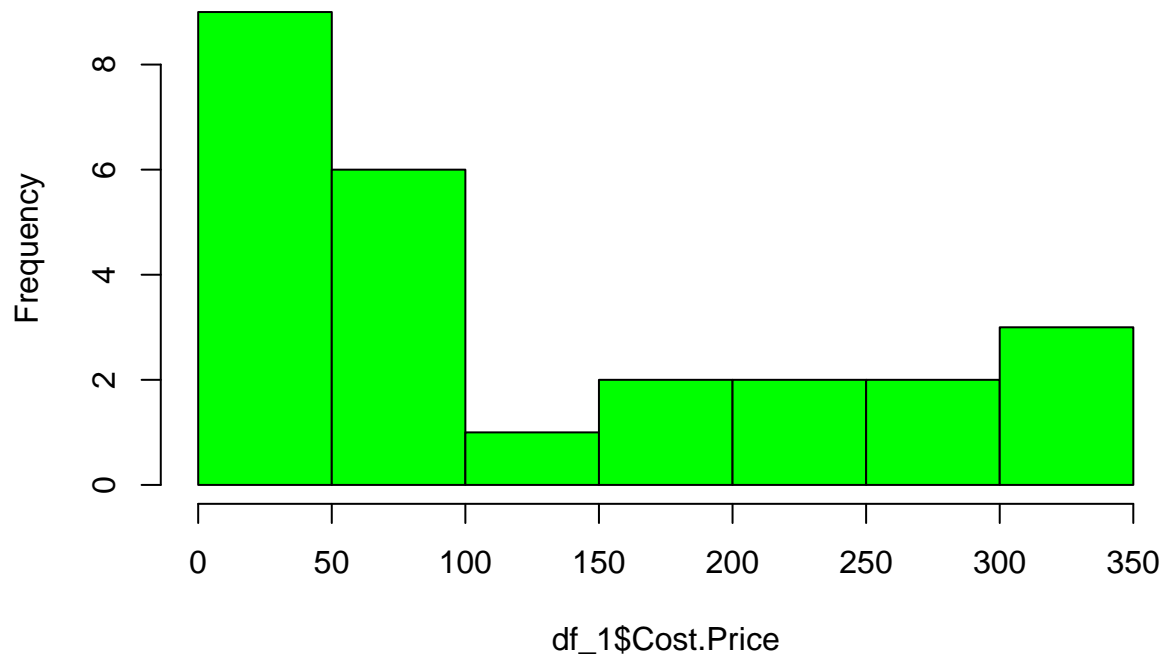

```
#
hist(df_1$Sales.Price,col.main="blue",col = "orange")
```

Histogram of df_1$Sales.Price
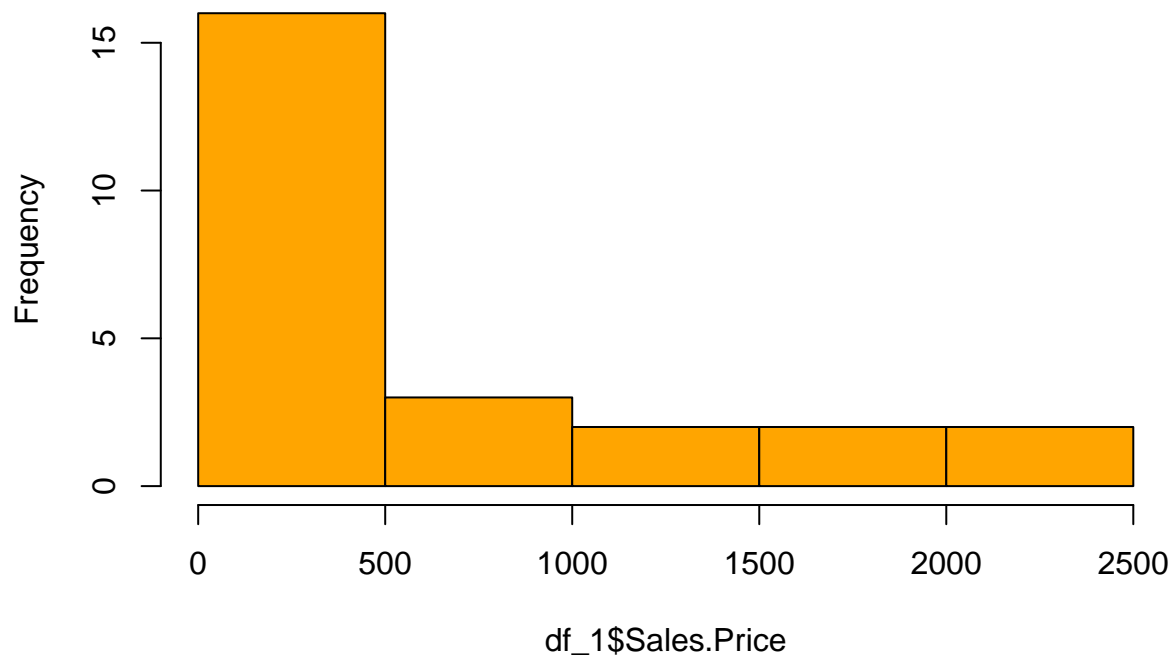
par() Function - Multiple adjacent plots

par{graphics}

Next sections we look at some basic usage of the par() function , with it we can create a matrix of nrows X ncolumns - of various plots and graphs
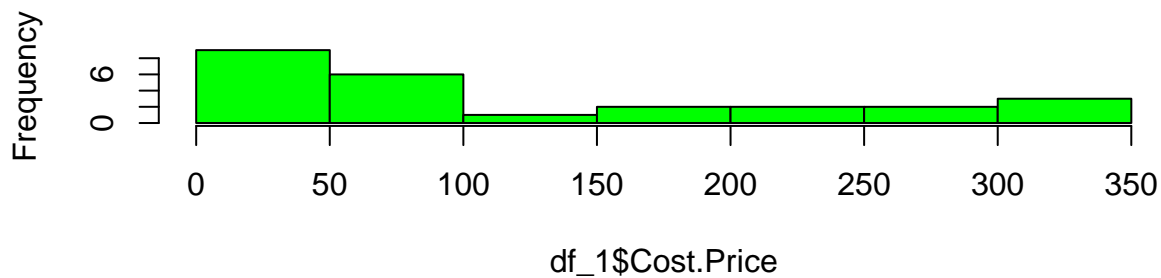
```
# Code Section -27
#?par() # Uncomment for Help

opar <- par(no.readonly=TRUE)
par(mfrow=c(2,1)) # 2 - ROWS , 1 - COLUMN - Filled in by ROWS # or we may use - mfcol=c() , to fill by
hist(df_1$Cost.Price,col.main="blue",col = "green")
hist(df_1$Sales.Price,col.main="blue",col = "orange")
```
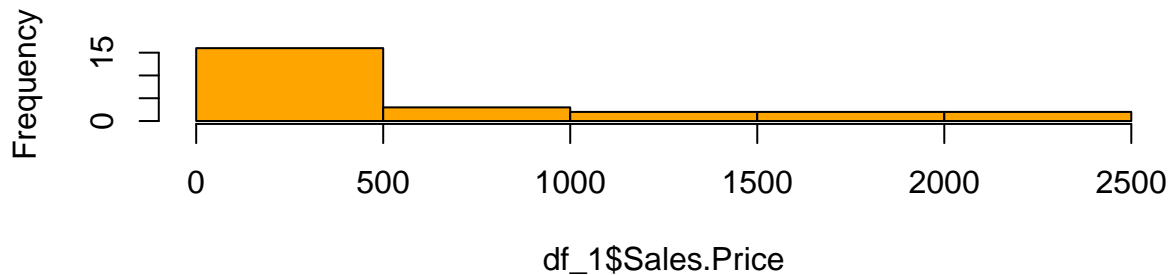


```
# Next lets try - sub-plot in 2 COLUMNS and 1 ROW

opar <- par(no.readonly=TRUE)
par(mfrow=c(1,2)) # 1 - ROW , 2 - COLUMNS
hist(df_1$Cost.Price,col.main="black",col = "yellow")
hist(df_1$Sales.Price,col.main="black",col = "red")
```
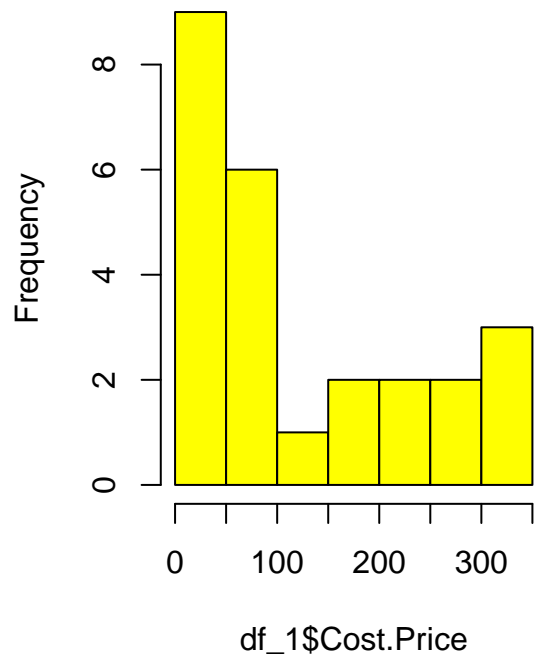
Histogram of df_1$Cost.Price　　**Histogram of df_1$Sales.Price**
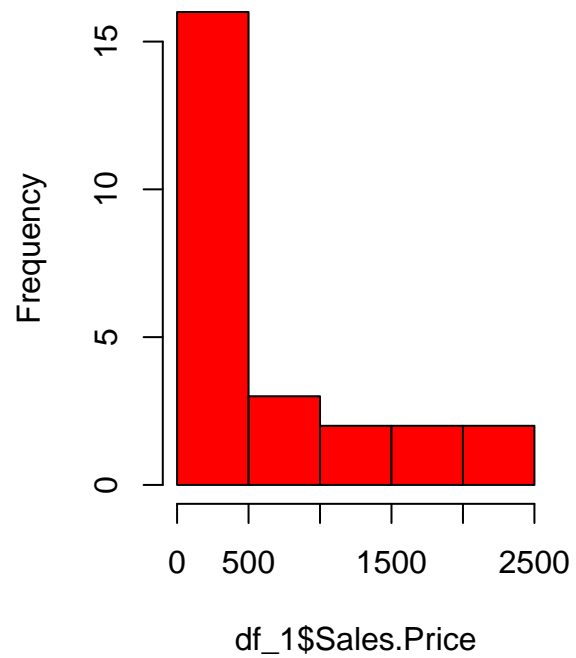

```
#
opar <- par(no.readonly=TRUE)
par(mfrow=c(2,2)) # 2 - ROWS , 2 - COLUMNS
plot(df_1$Date.of.Invoice,df_1$Sales.Price,col=c("red","yellow"))
hist(df_1$Cost.Price,col.main="black",col = "yellow")
plot(df_1$Date.of.Invoice,df_1$Cost.Price,col=c("green","yellow"))
hist(df_1$Sales.Price,col.main="black",col = "red")
```

Histogram of df_1$Cost.Price

Histogram of df_1$Sales.Price
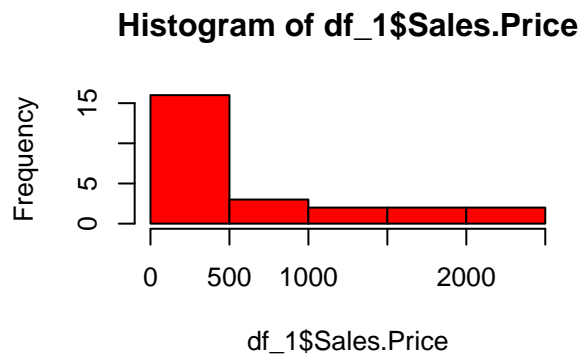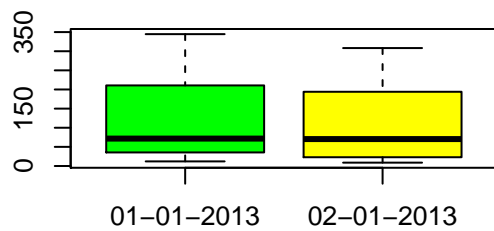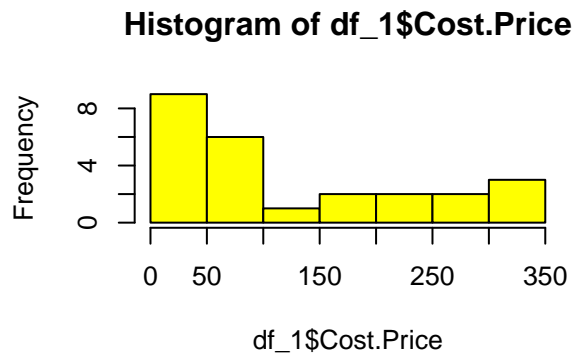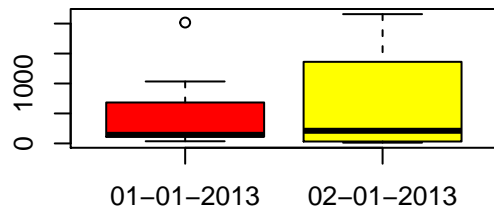
LAYOUT of Plots - beyond the par()

```
# Code Section -28
## The same set of PLOTS as earlier now formatted
# with Layout inplace of par()


#?layout()

layout(matrix(1, 2, byrow = TRUE),widths = c(4,1),heights = c(4,1))

plot(df_1$Date.of.Invoice,df_1$Sales.Price,col=c("green","yellow"))
```
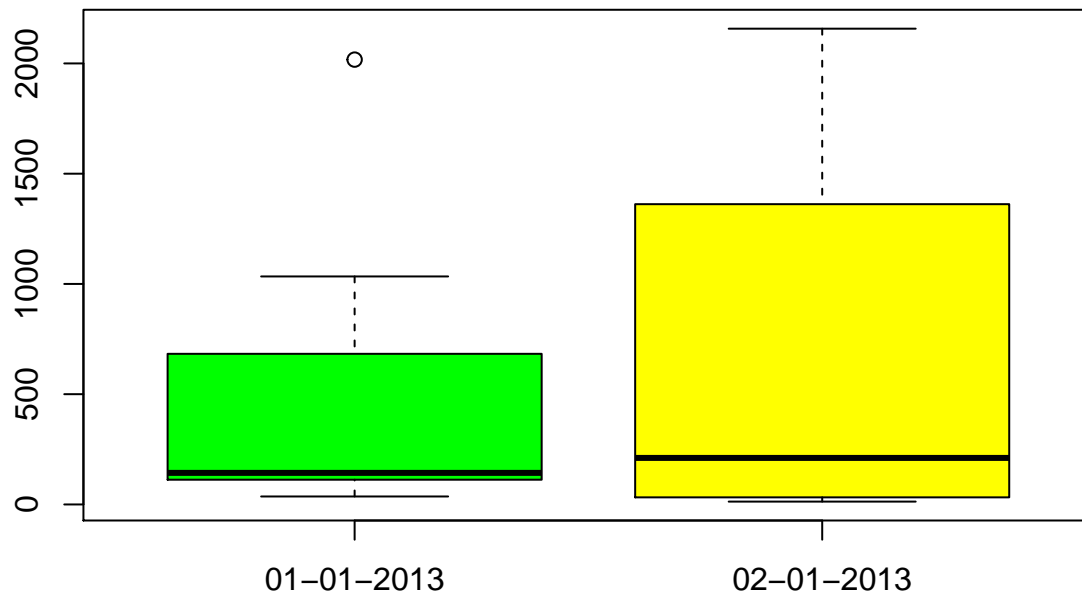
```
hist(df_1$Cost.Price,col.main="blue",col = "yellow")
```

Histogram of df_1$Cost.Price
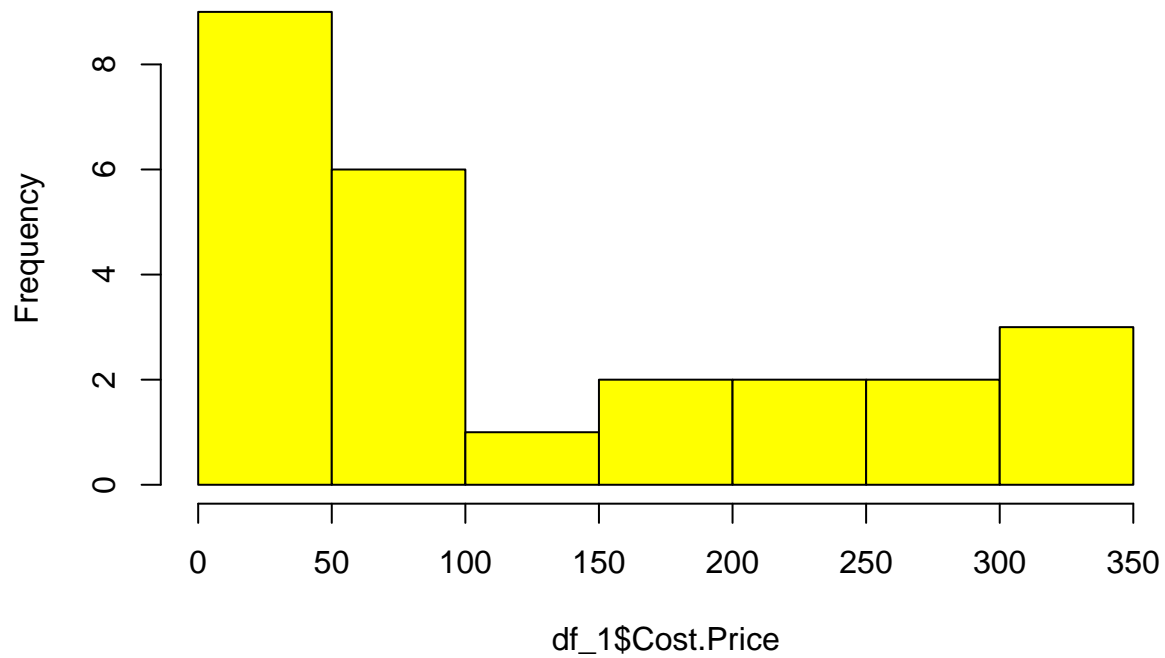

```
plot(df_1$Date.of.Invoice,df_1$Cost.Price,col=c("green","yellow"))
```
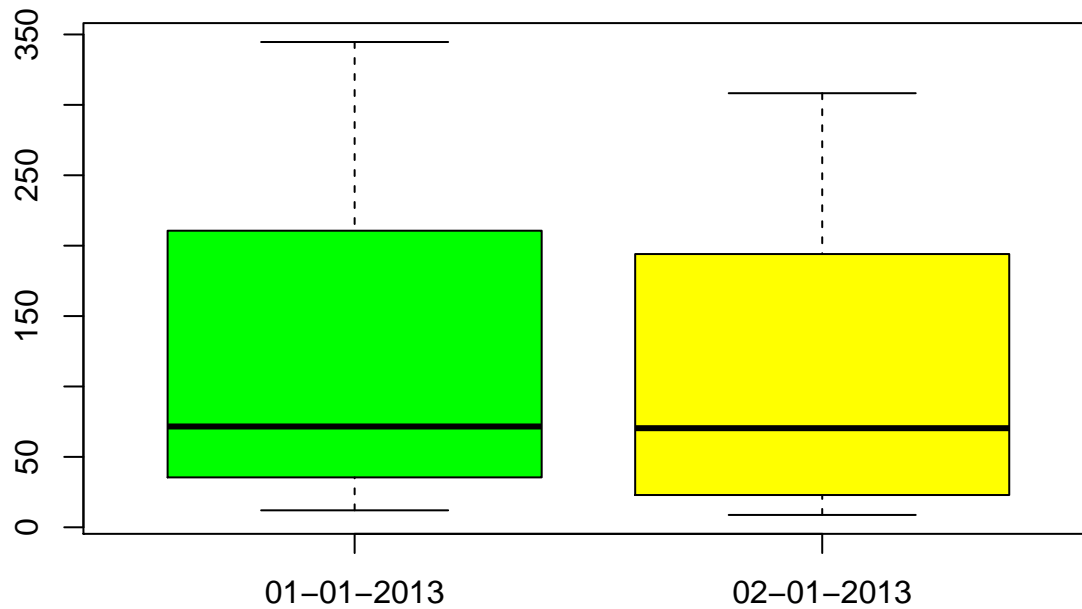
```
hist(df_1$Sales.Price,col.main="black",col = "red")
```

Histogram of df_1$Sales.Price
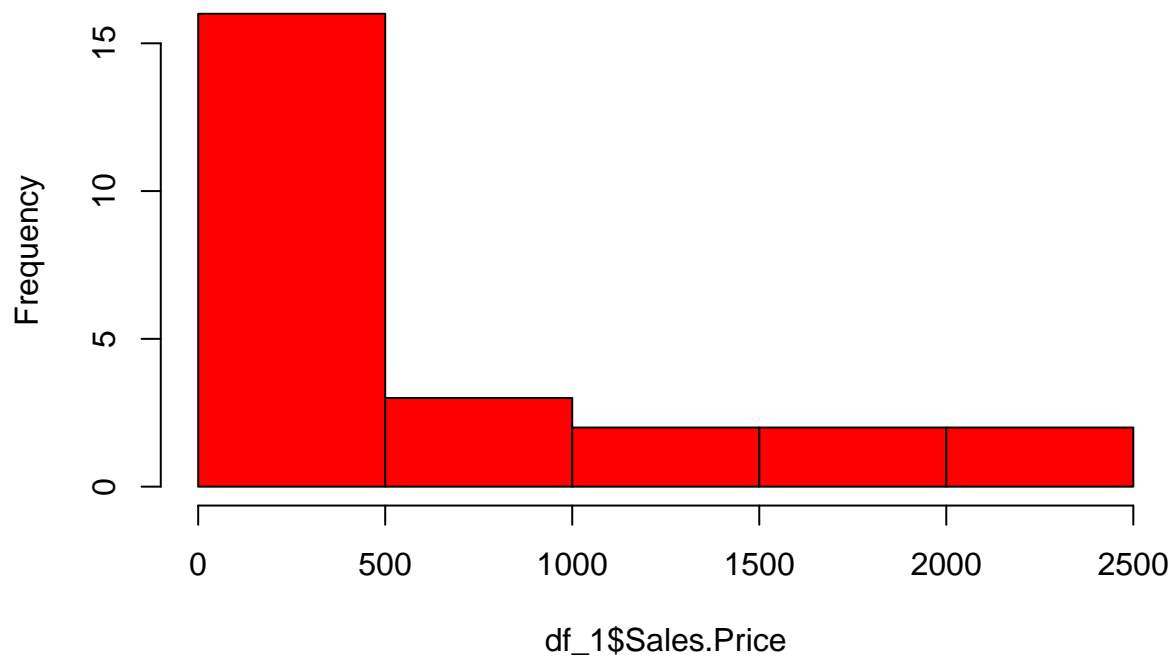
SCATTER Plot - initial Intro

```
# Code Section -29

df_a1 <- read.csv("~/Desktop/R_Own/Data_Directory/mcprod.csv")
```

```
## CHECK --- Why not Factor Variables ??
#
summary(df_a1)
```

```
##        X              a1              a2              a3
##  Prod_1 : 1   Min.   :10.4   Min.   :4.000   Min.   : 71.1
##  Prod_10: 1   1st Qu.:15.2   1st Qu.:4.000   1st Qu.:120.3
##  Prod_11: 1   Median :19.2   Median :6.000   Median :196.3
##  Prod_12: 1   Mean   :20.3   Mean   :6.158   Mean   :229.8
##  Prod_13: 1   3rd Qu.:24.0   3rd Qu.:8.000   3rd Qu.:342.0
##  Prod_14: 1   Max.   :33.9   Max.   :8.000   Max.   :472.0
##  (Other):32
##        a4              a5              a6              a7
##  Min.   : 52.0   Min.   :2.760   Min.   :1.513   Min.   :14.50
##  1st Qu.: 95.5   1st Qu.:3.098   1st Qu.:2.356   1st Qu.:16.88
##  Median :123.0   Median :3.715   Median :3.325   Median :17.41
##  Mean   :145.6   Mean   :3.614   Mean   :3.149   Mean   :17.72
##  3rd Qu.:180.0   3rd Qu.:3.920   3rd Qu.:3.690   3rd Qu.:18.83
##  Max.   :335.0   Max.   :4.930   Max.   :5.424   Max.   :22.90
##
##        a8              a9              a10             a11
##  Min.   :0.0000   Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :0.0000   Median :3.500   Median :2.000
##  Mean   :0.4211   Mean   :0.4211   Mean   :3.711   Mean   :2.684
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :5.000   Max.   :8.000
##
```

```
#
plot(df_a1$a6,df_a1$a5,col=c("red","green"))
```
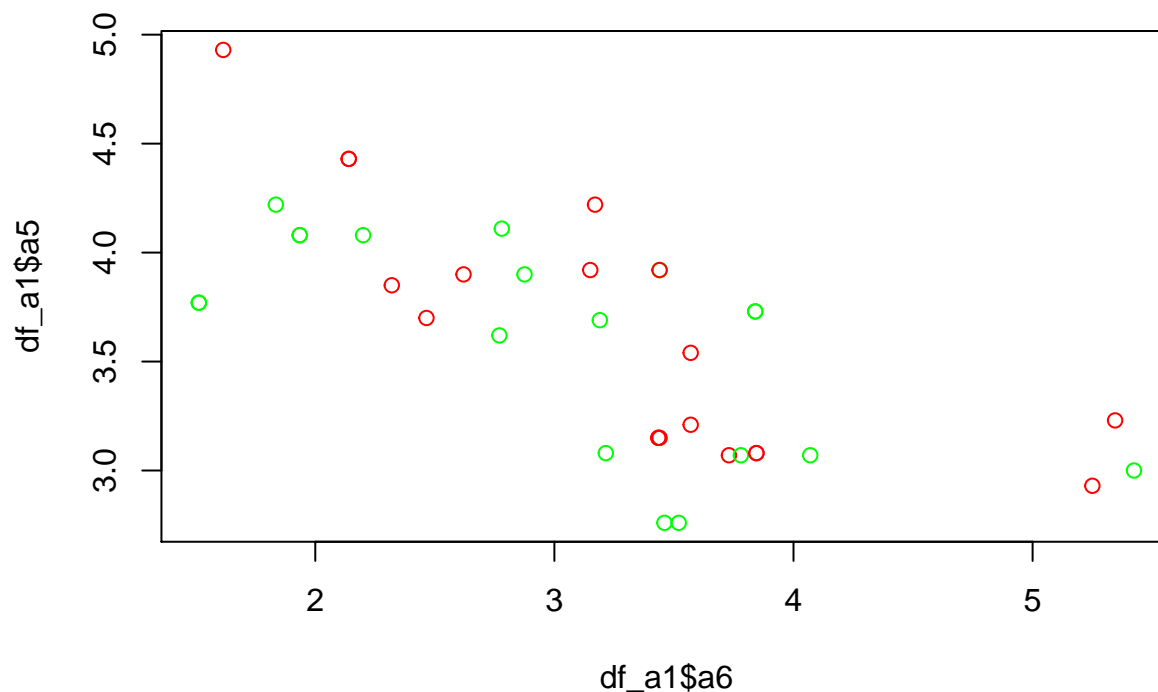
```
#
# Its a Synthetic Simulated Dataset for learning purposes ...
# As seen below - Variables a1,a3,a5-a7 == Double / Decimal
# (Double Precision Numeric Values)
# X1 == Var Char
# All others are Integers

# QUOTING Verbatim - WIKI-Scatter Diagram -- https://en.wikipedia.org/wiki/Scatter_plot
#
# If no dependent variable exists, either type of variable can be plotted
# on either axis and a scatter plot will illustrate only the
# degree of correlation (not causation) between two variables.
#
# VERY VERY VERY Important - NOT CAUSATION ... so how do we
# ascertain CAUSATION , we shall see within this text soon.

# Again we plot multiple plots within a PLOT MATRIX using par()
opar <- par(no.readonly=TRUE)
par(mfrow=c(2,2)) # 2 - ROWS , 2 - COLUMNS
#
plot(df_a1$a1,df_a1$a3,col=c("blue","red"),main="Plot_A [Negative_Corr.]")
#
plot(df_a1$a1,df_a1$a5,col=c("blue","green"),main="Plot_B [Slight Positive_Corr.]")
#
plot(df_a1$a1,df_a1$a6,col=c("blue","orange"),main="Plot_C [Negative_Corr.]")
#
plot(df_a1$a1,df_a1$a7,col=c("blue","yellow"),main="Plot_D [In-conclusive.]")
```
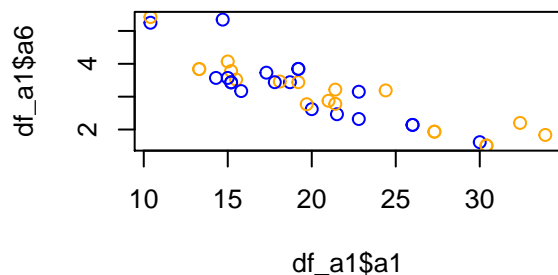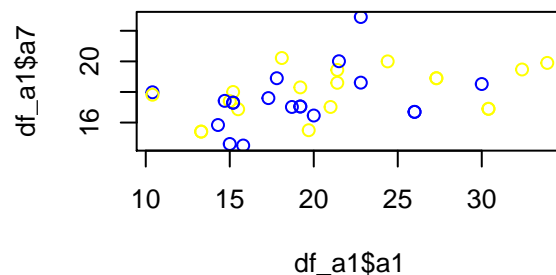

Plot_A [Negative_Corr.]
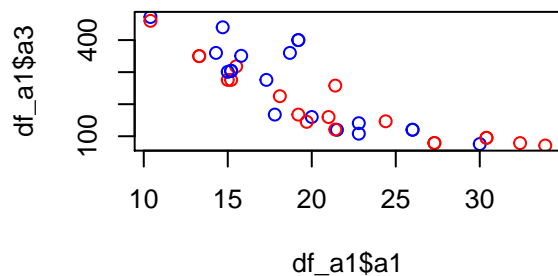
Plot_B [Slight Positive_Corr.]
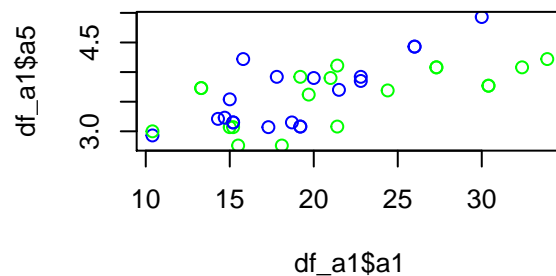
Plot_C [Negative_Corr.]

Plot_D [In−conclusive.]

```
#
```

```
##### Next Section Lets Look at CORRPLOTS --- or
# Correlation Plots for all Variables with all the Other Variables
```

Looking at these FOUR Scatter Plots we can now have a basic level discussion about - CORRELATION amongst different variables in any given Data Set.

We have chosen to have - variable "a1" on the X Axis of all Four Plots .

Plot_A - Var "a1" and "a3" seem to be Highly Correlated . As values of "a1" increase the values of "a3" decrease.

Plot_B - Var "a1" and "a5" seem to be Slightly Correlated . As values of "a1" increase , some of the values of "a5" decrease.

Plot_C - Var "a1" and "a6" seem to be Highly Correlated . As values of "a1" increase the values of "a6" decrease.

Plot_D - We are not able to draw any inference for Correlation.

```
# Code Section -30
```

```
# Code Section -32
```

```
# Code Section -33
```

```
# Code Section -34
```

```
# Code Section -35
```

```
# Code Section -36
```

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.1 LTS
##
## locale:
##  [1] LC_CTYPE=en_IN.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_IN.UTF-8        LC_COLLATE=en_IN.UTF-8
##  [5] LC_MONETARY=en_IN.UTF-8    LC_MESSAGES=en_IN.UTF-8
##  [7] LC_PAPER=en_IN.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_IN.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.8     lattice_0.20-33 digest_0.6.10   rprojroot_1.1
##  [5] grid_3.3.2      nlme_3.1-124    backports_1.0.4 magrittr_1.5
##  [9] evaluate_0.10   stringi_1.1.2   rmarkdown_1.3   tools_3.3.2
## [13] foreign_0.8-66  stringr_1.1.0   yaml_2.1.14     parallel_3.3.2
## [17] mnormt_1.5-5    htmltools_0.3.5 knitr_1.15.1
```