R Dataframe Functions

Create Dataframe from matrix / array

as.data.frame()

```
M1 = matrix(c(1,2,3,4),nrow=2)
#M1
df <- as.data.frame(M1)
df
```

V1 <dbl>	V2 <dbl>
1	3
2	4

2 rows

```
df$V1
```

```
## [1] 1 2
```

```
str(df)
```

```
## 'data.frame':    2 obs. of  2 variables:
##  $ V1: num  1 2
##  $ V2: num  3 4
```

```
is.data.frame(df)
```

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

```
nrow(df)
```

```
## [1] 2
```

```
ncol(df)
```

```
## [1] 2
```

```
length(df)
```

```
## [1] 2
```

```
colnames(df)
```

```
## [1] "V1" "V2"
```

```
names(df)
```

```
## [1] "V1" "V2"
```

```
M1 = matrix(c(1,2,3,4),nrow=2)
m_df <- as.data.frame(M1)
colnames(m_df) = c('c1','c2') # change column names
print(m_df)
```

```
##   c1 c2
## 1  1  3
## 2  2  4
```

```
str(m_df)
```

```
## 'data.frame':    2 obs. of  2 variables:
##  $ c1: num  1 2
##  $ c2: num  3 4
```

```
ncol(m_df)
```

```
## [1] 2
```

```
#nominal varible doesn't have any order
emp = c("John Doe", "Peter Gynn", "Jolie Hope")
#Ordinal varible has order
designation <- factor(c("Manager", "Team Lead","SME"), ordered =TRUE,          levels = c("SM
E", "Team Lead","Manager"))
salary <- c(41000, 33400, 26800)
startdate <- as.Date(c("2010-11-1", "2008-3-25", "2007-3-14"))
employee.data <- data.frame(emp, designation,salary, startdate)

str(employee.data)
```

```
## 'data.frame':    3 obs. of  4 variables:
##  $ emp        : chr  "John Doe" "Peter Gynn" "Jolie Hope"
##  $ designation: Ord.factor w/ 3 levels "SME"<"Team Lead"<..: 3 2 1
##  $ salary     : num  41000 33400 26800
##  $ startdate  : Date, format: "2010-11-01" "2008-03-25" ...
```

```
employee.data
```

emp <chr>	designation <ord>	salary <dbl>	startdate <date>
John Doe	Manager	41000	2010-11-01
Peter Gynn	Team Lead	33400	2008-03-25
Jolie Hope	SME	26800	2007-03-14

3 rows

```
employ.data <- data.frame(emp, salary, startdate, stringsAsFactors=TRUE)
str(employ.data)
```

```
## 'data.frame':    3 obs. of  3 variables:
##  $ emp      : Factor w/ 3 levels "John Doe","Jolie Hope",..: 1 3 2
##  $ salary   : num  41000 33400 26800
##  $ startdate: Date, format: "2010-11-01" "2008-03-25" ...
```

```
# variaable name can have .
employ.data <- data.frame(emp, salary, startdate, stringsAsFactors=FALSE)
# select rows based on condition
employ.data[employ.data$salary>35000,]
```

	emp <chr>	salary <dbl>	startdate <date>
1	John Doe	41000	2010-11-01

1 row

```
employ.data$salary
```

```
## [1] 41000 33400 26800
```

```
employ.data[c('emp','salary')]
```

emp <chr>	salary <dbl>

emp <chr>	salary <dbl>
John Doe	41000
Peter Gynn	33400
Jolie Hope	26800

3 rows

```
employ.data[employ.data$salary>35000,c('emp','salary')]
```

	emp <chr>	salary <dbl>
1	John Doe	41000

1 row

Reading TEXT / CSV into Data frame

read.csv(file, header = TRUE, sep = ",", quote = "\"", dec = ".", fill = TRUE, comment.char = "", …)

GET HELP –> ?read.csv

read.csv2(file, header = TRUE, sep = ";", quote = "\"", dec = ",", fill = TRUE, comment.char = "", …)

```
df = read.csv("76_attributes_heartdiseases.csv")
summary(df)
```

```
##        V1               V2          V3              V4               V5
##  Min.   :  1.00   Min.   :0   Min.   :29.00   Min.   :0.0000   Min.   :-9
##  1st Qu.: 75.25   1st Qu.:0   1st Qu.:48.00   1st Qu.:0.0000   1st Qu.:-9
##  Median :151.50   Median :0   Median :55.00   Median :1.0000   Median :-9
##  Mean   :151.52   Mean   :0   Mean   :54.41   Mean   :0.6773   Mean   :-9
##  3rd Qu.:227.75   3rd Qu.:0   3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:-9
##  Max.   :298.00   Max.   :0   Max.   :77.00   Max.   :1.0000   Max.   :-9
##        V6           V7               V8              V9
##  Min.   :-9   Min.   :-18.000   Min.   :-9.000   Min.   :  1.000
##  1st Qu.:-9   1st Qu.: -9.000   1st Qu.:-9.000   1st Qu.:  3.000
##  Median :-9   Median : -9.000   Median :-9.000   Median :  3.000
##  Mean   :-9   Mean   : -9.096   Mean   :-8.869   Mean   :  4.433
##  3rd Qu.:-9   3rd Qu.: -9.000   3rd Qu.:-9.000   3rd Qu.:  4.000
##  Max.   :-9   Max.   : -9.000   Max.   : 4.000   Max.   :130.000
##       V10             V11               V12             V13
##  Min.   :  1.0   Min.   :  0.000   Min.   : -9.0   Min.   :-9.000
##  1st Qu.:120.0   1st Qu.:  0.000   1st Qu.:212.0   1st Qu.:-9.000
##  Median :130.0   Median :  1.000   Median :244.0   Median :-9.000
##  Mean   :130.3   Mean   :  2.823   Mean   :246.8   Mean   :-8.727
##  3rd Qu.:140.0   3rd Qu.:  1.000   3rd Qu.:277.0   3rd Qu.:-9.000
##  Max.   :200.0   Max.   :253.000   Max.   :564.0   Max.   :30.000
##       V14             V15              V16                V17
##  Min.   :-9.0   Min.   :-9.00   Min.   :-9.00000   Min.   :-9.000
##  1st Qu.: 0.0   1st Qu.: 0.00   1st Qu.: 0.00000   1st Qu.:-9.000
##  Median :10.0   Median :15.00   Median : 0.00000   Median :-9.000
##  Mean   :16.5   Mean   :14.62   Mean   : 0.05319   Mean   :-8.078
##  3rd Qu.:30.0   3rd Qu.:30.00   3rd Qu.: 0.00000   3rd Qu.:-9.000
##  Max.   :99.0   Max.   :54.00   Max.   : 1.00000   Max.   : 1.000
##       V18              V19              V20             V21
##  Min.   :0.0000   Min.   : 0.000   Min.   : 1.0   Min.   : 1.00
##  1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.: 3.0   1st Qu.: 8.00
##  Median :1.0000   Median : 2.000   Median : 7.0   Median :15.00
##  Mean   :0.6099   Mean   : 1.071   Mean   : 6.5   Mean   :15.94
##  3rd Qu.:1.0000   3rd Qu.: 2.000   3rd Qu.:10.0   3rd Qu.:22.00
##  Max.   :1.0000   Max.   :11.000   Max.   :23.0   Max.   :82.00
##       V22             V23                V24               V25
##  Min.   : 0.00   Min.   :-9.00000   Min.   :-9.0000   Min.   :-9.0000
##  1st Qu.:82.00   1st Qu.: 0.00000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median :82.00   Median : 0.00000   Median : 0.0000   Median : 0.0000
##  Mean   :81.43   Mean   :-0.03191   Mean   : 0.2695   Mean   : 0.1809
##  3rd Qu.:83.00   3rd Qu.: 0.00000   3rd Qu.: 1.0000   3rd Qu.: 0.0000
##  Max.   :84.00   Max.   : 1.00000   Max.   : 1.0000   Max.   : 1.0000
##       V26                V27               V28            V29
##  Min.   :-9.00000   Min.   :-9.00000   Min.   :1.00   Min.   : 1.800
##  1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.:1.00   1st Qu.: 6.500
##  Median : 0.00000   Median : 0.00000   Median :1.00   Median : 8.500
##  Mean   : 0.03546   Mean   : 0.06028   Mean   :1.08   Mean   : 8.405
##  3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.:1.00   3rd Qu.:10.075
##  Max.   : 1.00000   Max.   : 1.00000   Max.   :9.00   Max.   :15.000
##       V30              V31             V32           V33
##  Min.   :-9.000   Min.   :  3.0   Min.   : 71   Min.   : 40.00
```

```
##  1st Qu.: 0.000   1st Qu.:  7.0   1st Qu.:132   1st Qu.: 65.00
##  Median : 3.000   Median :  9.5   Median :153   Median : 74.00
##  Mean   : 1.507   Mean   : 11.3   Mean   :149   Mean   : 75.95
##  3rd Qu.: 6.000   3rd Qu.: 12.0   3rd Qu.:165   3rd Qu.: 85.00
##  Max.   :15.000   Max.   :175.0   Max.   :202   Max.   :190.00
##       V34             V35             V36             V37
##  Min.   : 84.0   Min.   : 26.00   Min.   : 78.0   Min.   :  0.00
##  1st Qu.:152.0   1st Qu.: 70.00   1st Qu.:120.0   1st Qu.: 80.00
##  Median :168.0   Median : 80.00   Median :130.0   Median : 85.00
##  Mean   :167.3   Mean   : 79.09   Mean   :131.2   Mean   : 84.04
##  3rd Qu.:183.5   3rd Qu.: 85.00   3rd Qu.:140.0   3rd Qu.: 90.00
##  Max.   :232.0   Max.   :130.00   Max.   :200.0   Max.   :110.00
##       V38              V39              V40             V41
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.000   Min.   :-9.000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.: 1.000
##  Median :0.0000   Median :0.00000   Median :0.800   Median : 2.000
##  Mean   :0.3227   Mean   :0.03652   Mean   :1.037   Mean   : 1.433
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.600   3rd Qu.: 2.000
##  Max.   :1.0000   Max.   :1.80000   Max.   :6.200   Max.   : 3.000
##       V42              V43             V44              V45          V46
##  Min.   : -9.000   Min.   :  0.00   Min.   :-9.0000   Min.   :-9   Min.   :-9
##  1st Qu.: -9.000   1st Qu.: 90.25   1st Qu.: 0.0000   1st Qu.:-9   1st Qu.:-9
##  Median : -9.000   Median :117.50   Median : 0.0000   Median :-9   Median :-9
##  Mean   : -6.564   Mean   :121.29   Mean   : 0.4539   Mean   :-9   Mean   :-9
##  3rd Qu.: -9.000   3rd Qu.:150.00   3rd Qu.: 1.0000   3rd Qu.:-9   3rd Qu.:-9
##  Max.   :200.000   Max.   :270.00   Max.   : 3.0000   Max.   :-9   Max.   :-9
##       V47          V48          V49          V50              V51
##  Min.   :-9   Min.   :-9   Min.   :-9   Min.   :-9.000   Min.   :-9.000
##  1st Qu.:-9   1st Qu.:-9   1st Qu.:-9   1st Qu.:-9.000   1st Qu.: 3.000
##  Median :-9   Median :-9   Median :-9   Median :-9.000   Median : 3.000
##  Mean   :-9   Mean   :-9   Mean   :-9   Mean   :-8.787   Mean   : 4.369
##  3rd Qu.:-9   3rd Qu.:-9   3rd Qu.:-9   3rd Qu.:-9.000   3rd Qu.: 7.000
##  Max.   :-9   Max.   :-9   Max.   :-9   Max.   : 7.000   Max.   : 7.000
##       V52          V53          V54             V55              V56
##  Min.   :-9   Min.   :-9   Min.   :-9.00   Min.   : 1.000   Min.   : 1.00
##  1st Qu.:-9   1st Qu.:-9   1st Qu.:-9.00   1st Qu.: 3.000   1st Qu.: 8.00
##  Median :-9   Median :-9   Median :-9.00   Median : 7.000   Median :15.00
##  Mean   :-9   Mean   :-9   Mean   :-8.78   Mean   : 6.571   Mean   :16.24
##  3rd Qu.:-9   3rd Qu.:-9   3rd Qu.:-9.00   3rd Qu.:10.000   3rd Qu.:23.00
##  Max.   :-9   Max.   :-9   Max.   :11.00   Max.   :29.000   Max.   :82.00
##       V57             V58             V59             V60
##  Min.   : 0.00   Min.   :0.0000   Min.   :1.000   Min.   :1.000
##  1st Qu.:82.00   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000
##  Median :82.00   Median :0.0000   Median :1.000   Median :1.000
##  Mean   :81.15   Mean   :0.9184   Mean   :1.043   Mean   :1.145
##  3rd Qu.:83.00   3rd Qu.:2.0000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :84.00   Max.   :4.0000   Max.   :2.000   Max.   :2.000
##       V61             V62             V63                V64
##  Min.   :-9.00   Min.   :-9.000   Length:282         Min.   :-9.000
##  1st Qu.: 1.00   1st Qu.:-9.000   Class :character   1st Qu.:-9.000
##  Median : 1.00   Median :-9.000   Mode  :character   Median :-9.000
```

```
##  Mean   : 1.06   Mean   :-8.858                     Mean   :-8.784
##  3rd Qu.: 1.00   3rd Qu.:-9.000                     3rd Qu.:-9.000
##  Max.   : 2.00   Max.   : 1.000                     Max.   : 2.000
##       V65               V66              V67             V68             V69
##  Min.   :-9.0000   Min.   :-9.000   Min.   :1.000   Min.   :1.000   Min.   :1
##  1st Qu.: 1.0000   1st Qu.:-9.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1
##  Median : 1.0000   Median :-9.000   Median :1.000   Median :1.000   Median :1
##  Mean   : 0.9468   Mean   :-8.784   Mean   :1.174   Mean   :1.124   Mean   :1
##  3rd Qu.: 1.0000   3rd Qu.:-9.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1
##  Max.   : 2.0000   Max.   : 2.000   Max.   :2.000   Max.   :2.000   Max.   :1
##       V70             V71              V72             V73
##  Min.   :1.000   Min.   : 1.000   Min.   :1.000   Min.   :-9.0000
##  1st Qu.:1.000   1st Qu.: 1.000   1st Qu.:1.000   1st Qu.: 1.0000
##  Median :1.000   Median : 1.000   Median :1.000   Median : 1.0000
##  Mean   :1.007   Mean   : 1.177   Mean   :1.404   Mean   : 0.8901
##  3rd Qu.:1.000   3rd Qu.: 1.000   3rd Qu.:1.000   3rd Qu.: 1.0000
##  Max.   :3.000   Max.   :11.000   Max.   :8.000   Max.   : 4.0000
##       V74          V75                V76
##  Min.   :-9   Length:282         Length:282
##  1st Qu.:-9   Class :character   Class :character
##  Median :-9   Mode  :character   Mode  :character
##  Mean   :-9
##  3rd Qu.:-9
##  Max.   :-9
```

```
#df_sample = read.csv("sample3.txt",sep="")
#df_sample
df_sample = read.csv("sample1.txt",sep="|")
df_sample
```

c1 <int>	c2 <int>	c3 <int>
1	23	233
2	45	254
3	67	555

3 rows

```
df = read.csv("76_attributes_heartdiseases.csv")
summary(df$V73)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -9.0000  1.0000  1.0000  0.8901  1.0000  4.0000
```

```
df = read.csv("76_attributes_heartdiseases.csv",na.strings=c("-9","-18"))
summary(df$V73)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   1.000   1.142   1.000   4.000       7
```

```
summary(df)
```

```
##        V1               V2          V3              V4            V5
##  Min.   :  1.00   Min.   :0   Min.   :29.00   Min.   :0.0000   Mode:logical
##  1st Qu.: 75.25   1st Qu.:0   1st Qu.:48.00   1st Qu.:0.0000   NA's:282
##  Median :151.50   Median :0   Median :55.00   Median :1.0000
##  Mean   :151.52   Mean   :0   Mean   :54.41   Mean   :0.6773
##  3rd Qu.:227.75   3rd Qu.:0   3rd Qu.:61.00   3rd Qu.:1.0000
##  Max.   :298.00   Max.   :0   Max.   :77.00   Max.   :1.0000
##
##     V6             V7               V8              V9
##  Mode:logical   Mode:logical   Min.   :3.000   Min.   :  1.000
##  NA's:282       NA's:282       1st Qu.:3.000   1st Qu.:  3.000
##                                Median :3.000   Median :  3.000
##                                Mean   :3.333   Mean   :  4.433
##                                3rd Qu.:3.500   3rd Qu.:  4.000
##                                Max.   :4.000   Max.   :130.000
##                                NA's   :279
##       V10             V11               V12             V13
##  Min.   :  1.0   Min.   :  0.000   Min.   :126.0   Min.   : 0.00
##  1st Qu.:120.0   1st Qu.:  0.000   1st Qu.:213.0   1st Qu.:10.00
##  Median :130.0   Median :  1.000   Median :244.0   Median :20.00
##  Mean   :130.3   Mean   :  2.823   Mean   :249.5   Mean   :16.67
##  3rd Qu.:140.0   3rd Qu.:  1.000   3rd Qu.:277.5   3rd Qu.:25.00
##  Max.   :200.0   Max.   :253.000   Max.   :564.0   Max.   :30.00
##                                    NA's   :3       NA's   :279
##       V14             V15             V16             V17
##  Min.   : 0.00   Min.   : 0.00   Min.   :0.0000   Min.   :1
##  1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.:0.0000   1st Qu.:1
##  Median :10.00   Median :15.00   Median :0.0000   Median :1
##  Mean   :16.96   Mean   :15.04   Mean   :0.1505   Mean   :1
##  3rd Qu.:30.00   3rd Qu.:30.00   3rd Qu.:0.0000   3rd Qu.:1
##  Max.   :99.00   Max.   :54.00   Max.   :1.0000   Max.   :1
##  NA's   :5       NA's   :5       NA's   :3        NA's   :256
##       V18              V19             V20             V21
##  Min.   :0.0000   Min.   : 0.000   Min.   : 1.0   Min.   : 1.00
##  1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.: 3.0   1st Qu.: 8.00
##  Median :1.0000   Median : 2.000   Median : 7.0   Median :15.00
##  Mean   :0.6099   Mean   : 1.071   Mean   : 6.5   Mean   :15.94
##  3rd Qu.:1.0000   3rd Qu.: 2.000   3rd Qu.:10.0   3rd Qu.:22.00
##  Max.   :1.0000   Max.   :11.000   Max.   :23.0   Max.   :82.00
##
##       V22             V23               V24             V25
##  Min.   : 0.00   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:82.00   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :82.00   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :81.43   Mean   :0.03214   Mean   :0.3357   Mean   :0.2464
##  3rd Qu.:83.00   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :84.00   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##                  NA's   :2         NA's   :2        NA's   :2
##       V26           V27             V28            V29              V30
##  Min.   :0.0   Min.   :0.000   Min.   :1.00   Min.   : 1.800   Min.   : 0.000
##  1st Qu.:0.0   1st Qu.:0.000   1st Qu.:1.00   1st Qu.: 6.500   1st Qu.: 3.000
```

```
##  Median :0.0   Median :0.000   Median :1.00   Median : 8.500   Median : 5.500
##  Mean   :0.1   Mean   :0.125   Mean   :1.08   Mean   : 8.405   Mean   : 4.911
##  3rd Qu.:0.0   3rd Qu.:0.000   3rd Qu.:1.00   3rd Qu.:10.075   3rd Qu.: 7.500
##  Max.   :1.0   Max.   :1.000   Max.   :9.00   Max.   :15.000   Max.   :15.000
##  NA's   :2     NA's   :2                                       NA's   :69
##       V31             V32           V33              V34
##  Min.   :  3.0   Min.   : 71   Min.   : 40.00   Min.   : 84.0
##  1st Qu.:  7.0   1st Qu.:132   1st Qu.: 65.00   1st Qu.:152.0
##  Median :  9.5   Median :153   Median : 74.00   Median :168.0
##  Mean   : 11.3   Mean   :149   Mean   : 75.95   Mean   :167.3
##  3rd Qu.: 12.0   3rd Qu.:165   3rd Qu.: 85.00   3rd Qu.:183.5
##  Max.   :175.0   Max.   :202   Max.   :190.00   Max.   :232.0
##
##       V35             V36             V37              V38
##  Min.   : 26.00   Min.   : 78.0   Min.   :  0.00   Min.   :0.0000
##  1st Qu.: 70.00   1st Qu.:120.0   1st Qu.: 80.00   1st Qu.:0.0000
##  Median : 80.00   Median :130.0   Median : 85.00   Median :0.0000
##  Mean   : 79.09   Mean   :131.2   Mean   : 84.04   Mean   :0.3227
##  3rd Qu.: 85.00   3rd Qu.:140.0   3rd Qu.: 90.00   3rd Qu.:1.0000
##  Max.   :130.00   Max.   :200.0   Max.   :110.00   Max.   :1.0000
##
##       V39               V40             V41             V42
##  Min.   :0.00000   Min.   :0.000   Min.   :1.000   Min.   :105.0
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:153.8
##  Median :0.00000   Median :0.800   Median :2.000   Median :173.0
##  Mean   :0.03652   Mean   :1.037   Mean   :1.583   Mean   :162.8
##  3rd Qu.:0.00000   3rd Qu.:1.600   3rd Qu.:2.000   3rd Qu.:182.0
##  Max.   :1.80000   Max.   :6.200   Max.   :3.000   Max.   :200.0
##                                    NA's   :4       NA's   :278
##       V43              V44           V45             V46             V47
##  Min.   :  0.00   Min.   :0.0000   Mode:logical   Mode:logical   Mode:logical
##  1st Qu.: 90.25   1st Qu.:0.0000   NA's:282       NA's:282       NA's:282
##  Median :117.50   Median :0.0000
##  Mean   :121.29   Mean   :0.6594
##  3rd Qu.:150.00   3rd Qu.:1.0000
##  Max.   :270.00   Max.   :3.0000
##                   NA's   :6
##    V48             V49               V50           V51           V52
##  Mode:logical   Mode:logical   Min.   :3     Min.   :3.000   Mode:logical
##  NA's:282       NA's:282       1st Qu.:6     1st Qu.:3.000   NA's:282
##                                Median :7     Median :3.000
##                                Mean   :6     Mean   :4.659
##                                3rd Qu.:7     3rd Qu.:7.000
##                                Max.   :7     Max.   :7.000
##                                NA's   :278   NA's   :6
##    V53               V54             V55              V56
##  Mode:logical   Min.   : 2.00   Min.   : 1.000   Min.   : 1.00
##  NA's:282       1st Qu.: 2.75   1st Qu.: 3.000   1st Qu.: 8.00
##                 Median : 6.50   Median : 7.000   Median :15.00
##                 Mean   : 6.50   Mean   : 6.571   Mean   :16.24
##                 3rd Qu.:10.25   3rd Qu.:10.000   3rd Qu.:23.00
```

```
##                Max.   :11.00   Max.   :29.000   Max.   :82.00
##                NA's   :278
##       V57              V58               V59              V60
##  Min.   : 0.00   Min.   :0.0000   Min.   :1.000   Min.   :1.000
##  1st Qu.:82.00   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000
##  Median :82.00   Median :0.0000   Median :1.000   Median :1.000
##  Mean   :81.15   Mean   :0.9184   Mean   :1.043   Mean   :1.145
##  3rd Qu.:83.00   3rd Qu.:2.0000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :84.00   Max.   :4.0000   Max.   :2.000   Max.   :2.000
##
##       V61              V62           V63                 V64
##  Min.   :1.000   Min.   :1     Length:282         Min.   :1.000
##  1st Qu.:1.000   1st Qu.:1     Class :character   1st Qu.:1.000
##  Median :1.000   Median :1     Mode  :character   Median :1.000
##  Mean   :1.205   Mean   :1                        Mean   :1.167
##  3rd Qu.:1.000   3rd Qu.:1                        3rd Qu.:1.000
##  Max.   :2.000   Max.   :1                        Max.   :2.000
##  NA's   :4       NA's   :278                      NA's   :276
##       V65              V66              V67              V68              V69
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1
##  Median :1.000   Median :1.000   Median :1.000   Median :1.000   Median :1
##  Mean   :1.163   Mean   :1.167   Mean   :1.174   Mean   :1.124   Mean   :1
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1
##  Max.   :2.000   Max.   :2.000   Max.   :2.000   Max.   :2.000   Max.   :1
##  NA's   :6       NA's   :276
##       V70              V71               V72              V73
##  Min.   :1.000   Min.   : 1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.: 1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :1.000   Median : 1.000   Median :1.000   Median :1.000
##  Mean   :1.007   Mean   : 1.177   Mean   :1.404   Mean   :1.142
##  3rd Qu.:1.000   3rd Qu.: 1.000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :3.000   Max.   :11.000   Max.   :8.000   Max.   :4.000
##                                                   NA's   :7
##    V74              V75                V76
##  Mode:logical   Length:282         Length:282
##  NA's:282       Class :character   Class :character
##                 Mode  :character   Mode  :character
##
##
##
##
```

```
df_sample = read.csv("sample2.txt",skip=2,sep="|")
df_sample
```

c1	c2	c3
<int>	<int>	<int>
1	23	233

c1 <int>	c2 <int>	c3 <int>
2	45	254
3	67	555

3 rows

```
df_sample = read.csv("sample2.txt",skip=2,sep="|",blank.lines.skip=FALSE)
df_sample
```

c1 <int>	c2 <int>	c3 <int>
1	23	233
2	45	254
NA	*NA*	*NA*
NA	*NA*	*NA*
3	67	555

5 rows

```
df_sample = read.csv("sample2.txt",skip=0)
df_sample
```

This.is.a.sample.file <chr>
We need to skip first two lines
c1|c2|c3
1|23|233
2|45|254
3|67|555

5 rows

Few more IMP operations in read.csv

stringsAsFactors –> Default to TRUE. String columns will be treated as factors

```
If FALSE then Strings will be kept intact
```

encoding –> For handing unicode or other encoding problems

nrows –> integer: the maximum number of rows to read in.

```
       Negative and other invalid values are ignored.

       Useful when loading very large files
```

```
print(ncol(df))
```

```
## [1] 76
```

```
print(nrow(df))
```

```
## [1] 282
```

```
summary(df)
```

```
##        V1               V2          V3              V4            V5
##  Min.   :  1.00   Min.   :0   Min.   :29.00   Min.   :0.0000   Mode:logical
##  1st Qu.: 75.25   1st Qu.:0   1st Qu.:48.00   1st Qu.:0.0000   NA's:282
##  Median :151.50   Median :0   Median :55.00   Median :1.0000
##  Mean   :151.52   Mean   :0   Mean   :54.41   Mean   :0.6773
##  3rd Qu.:227.75   3rd Qu.:0   3rd Qu.:61.00   3rd Qu.:1.0000
##  Max.   :298.00   Max.   :0   Max.   :77.00   Max.   :1.0000
##
##     V6             V7               V8              V9
##  Mode:logical   Mode:logical   Min.   :3.000   Min.   :  1.000
##  NA's:282       NA's:282       1st Qu.:3.000   1st Qu.:  3.000
##                                Median :3.000   Median :  3.000
##                                Mean   :3.333   Mean   :  4.433
##                                3rd Qu.:3.500   3rd Qu.:  4.000
##                                Max.   :4.000   Max.   :130.000
##                                NA's   :279
##       V10             V11               V12             V13
##  Min.   :  1.0   Min.   :  0.000   Min.   :126.0   Min.   : 0.00
##  1st Qu.:120.0   1st Qu.:  0.000   1st Qu.:213.0   1st Qu.:10.00
##  Median :130.0   Median :  1.000   Median :244.0   Median :20.00
##  Mean   :130.3   Mean   :  2.823   Mean   :249.5   Mean   :16.67
##  3rd Qu.:140.0   3rd Qu.:  1.000   3rd Qu.:277.5   3rd Qu.:25.00
##  Max.   :200.0   Max.   :253.000   Max.   :564.0   Max.   :30.00
##                                    NA's   :3       NA's   :279
##       V14             V15             V16             V17
##  Min.   : 0.00   Min.   : 0.00   Min.   :0.0000   Min.   :1
##  1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.:0.0000   1st Qu.:1
##  Median :10.00   Median :15.00   Median :0.0000   Median :1
##  Mean   :16.96   Mean   :15.04   Mean   :0.1505   Mean   :1
##  3rd Qu.:30.00   3rd Qu.:30.00   3rd Qu.:0.0000   3rd Qu.:1
##  Max.   :99.00   Max.   :54.00   Max.   :1.0000   Max.   :1
##  NA's   :5       NA's   :5       NA's   :3        NA's   :256
##       V18              V19             V20             V21
##  Min.   :0.0000   Min.   : 0.000   Min.   : 1.0   Min.   : 1.00
##  1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.: 3.0   1st Qu.: 8.00
##  Median :1.0000   Median : 2.000   Median : 7.0   Median :15.00
##  Mean   :0.6099   Mean   : 1.071   Mean   : 6.5   Mean   :15.94
##  3rd Qu.:1.0000   3rd Qu.: 2.000   3rd Qu.:10.0   3rd Qu.:22.00
##  Max.   :1.0000   Max.   :11.000   Max.   :23.0   Max.   :82.00
##
##       V22             V23               V24             V25
##  Min.   : 0.00   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:82.00   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :82.00   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :81.43   Mean   :0.03214   Mean   :0.3357   Mean   :0.2464
##  3rd Qu.:83.00   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :84.00   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##                  NA's   :2         NA's   :2        NA's   :2
##       V26           V27             V28            V29              V30
##  Min.   :0.0   Min.   :0.000   Min.   :1.00   Min.   : 1.800   Min.   : 0.000
##  1st Qu.:0.0   1st Qu.:0.000   1st Qu.:1.00   1st Qu.: 6.500   1st Qu.: 3.000
```

```
##  Median :0.0   Median :0.000   Median :1.00   Median : 8.500   Median : 5.500
##  Mean   :0.1   Mean   :0.125   Mean   :1.08   Mean   : 8.405   Mean   : 4.911
##  3rd Qu.:0.0   3rd Qu.:0.000   3rd Qu.:1.00   3rd Qu.:10.075   3rd Qu.: 7.500
##  Max.   :1.0   Max.   :1.000   Max.   :9.00   Max.   :15.000   Max.   :15.000
##  NA's   :2     NA's   :2                                       NA's   :69
##       V31             V32           V33              V34
##  Min.   :  3.0   Min.   : 71   Min.   : 40.00   Min.   : 84.0
##  1st Qu.:  7.0   1st Qu.:132   1st Qu.: 65.00   1st Qu.:152.0
##  Median :  9.5   Median :153   Median : 74.00   Median :168.0
##  Mean   : 11.3   Mean   :149   Mean   : 75.95   Mean   :167.3
##  3rd Qu.: 12.0   3rd Qu.:165   3rd Qu.: 85.00   3rd Qu.:183.5
##  Max.   :175.0   Max.   :202   Max.   :190.00   Max.   :232.0
##
##       V35              V36             V37              V38
##  Min.   : 26.00   Min.   : 78.0   Min.   :  0.00   Min.   :0.0000
##  1st Qu.: 70.00   1st Qu.:120.0   1st Qu.: 80.00   1st Qu.:0.0000
##  Median : 80.00   Median :130.0   Median : 85.00   Median :0.0000
##  Mean   : 79.09   Mean   :131.2   Mean   : 84.04   Mean   :0.3227
##  3rd Qu.: 85.00   3rd Qu.:140.0   3rd Qu.: 90.00   3rd Qu.:1.0000
##  Max.   :130.00   Max.   :200.0   Max.   :110.00   Max.   :1.0000
##
##       V39               V40             V41             V42
##  Min.   :0.00000   Min.   :0.000   Min.   :1.000   Min.   :105.0
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:153.8
##  Median :0.00000   Median :0.800   Median :2.000   Median :173.0
##  Mean   :0.03652   Mean   :1.037   Mean   :1.583   Mean   :162.8
##  3rd Qu.:0.00000   3rd Qu.:1.600   3rd Qu.:2.000   3rd Qu.:182.0
##  Max.   :1.80000   Max.   :6.200   Max.   :3.000   Max.   :200.0
##                                    NA's   :4       NA's   :278
##       V43              V44           V45             V46             V47
##  Min.   :  0.00   Min.   :0.0000   Mode:logical   Mode:logical   Mode:logical
##  1st Qu.: 90.25   1st Qu.:0.0000   NA's:282       NA's:282       NA's:282
##  Median :117.50   Median :0.0000
##  Mean   :121.29   Mean   :0.6594
##  3rd Qu.:150.00   3rd Qu.:1.0000
##  Max.   :270.00   Max.   :3.0000
##                   NA's   :6
##    V48             V49               V50           V51           V52
##  Mode:logical   Mode:logical   Min.   :3     Min.   :3.000   Mode:logical
##  NA's:282       NA's:282       1st Qu.:6     1st Qu.:3.000   NA's:282
##                                Median :7     Median :3.000
##                                Mean   :6     Mean   :4.659
##                                3rd Qu.:7     3rd Qu.:7.000
##                                Max.   :7     Max.   :7.000
##                                NA's   :278   NA's   :6
##    V53               V54             V55              V56
##  Mode:logical   Min.   : 2.00   Min.   : 1.000   Min.   : 1.00
##  NA's:282       1st Qu.: 2.75   1st Qu.: 3.000   1st Qu.: 8.00
##                 Median : 6.50   Median : 7.000   Median :15.00
##                 Mean   : 6.50   Mean   : 6.571   Mean   :16.24
##                 3rd Qu.:10.25   3rd Qu.:10.000   3rd Qu.:23.00
```

```
##                Max.   :11.00   Max.   :29.000   Max.   :82.00
##                NA's   :278
##       V57             V58              V59             V60
##  Min.   : 0.00   Min.   :0.0000   Min.   :1.000   Min.   :1.000
##  1st Qu.:82.00   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:1.000
##  Median :82.00   Median :0.0000   Median :1.000   Median :1.000
##  Mean   :81.15   Mean   :0.9184   Mean   :1.043   Mean   :1.145
##  3rd Qu.:83.00   3rd Qu.:2.0000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :84.00   Max.   :4.0000   Max.   :2.000   Max.   :2.000
##
##       V61             V62          V63                V64
##  Min.   :1.000   Min.   :1     Length:282         Min.   :1.000
##  1st Qu.:1.000   1st Qu.:1     Class :character   1st Qu.:1.000
##  Median :1.000   Median :1     Mode  :character   Median :1.000
##  Mean   :1.205   Mean   :1                        Mean   :1.167
##  3rd Qu.:1.000   3rd Qu.:1                        3rd Qu.:1.000
##  Max.   :2.000   Max.   :1                        Max.   :2.000
##  NA's   :4       NA's   :278                      NA's   :276
##       V65             V66             V67             V68             V69
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1
##  Median :1.000   Median :1.000   Median :1.000   Median :1.000   Median :1
##  Mean   :1.163   Mean   :1.167   Mean   :1.174   Mean   :1.124   Mean   :1
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1
##  Max.   :2.000   Max.   :2.000   Max.   :2.000   Max.   :2.000   Max.   :1
##  NA's   :6       NA's   :276
##       V70             V71              V72             V73
##  Min.   :1.000   Min.   : 1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.: 1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :1.000   Median : 1.000   Median :1.000   Median :1.000
##  Mean   :1.007   Mean   : 1.177   Mean   :1.404   Mean   :1.142
##  3rd Qu.:1.000   3rd Qu.: 1.000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :3.000   Max.   :11.000   Max.   :8.000   Max.   :4.000
##                                                   NA's   :7
##    V74             V75                V76
##  Mode:logical   Length:282         Length:282
##  NA's:282       Class :character   Class :character
##                 Mode  :character   Mode  :character
##
##
##
##
```

Create subset of a dataframe

We can use subset() function. HELP –> ?subset

Arguments:

data / data frame

condition based on which to create the subset

select –> specify the array of columns to be selected

```
df_1 = subset(df, df$V1 > 200)
head(df_1)
```

	V1 <int>	V2 <int>	V3 <int>	V4 <int>	V5 <lgl>	V6 <lgl>	V7 <lgl>	V8 <int>	V9 <int>
185	201	0	60	0	*NA*	*NA*	*NA*	*NA*	4
186	202	0	63	0	*NA*	*NA*	*NA*	*NA*	2
187	203	0	42	1	*NA*	*NA*	*NA*	*NA*	3
188	204	0	66	1	*NA*	*NA*	*NA*	*NA*	2
189	205	0	54	1	*NA*	*NA*	*NA*	*NA*	2
190	206	0	69	1	*NA*	*NA*	*NA*	*NA*	3

6 rows | 1-10 of 77 columns

```
df_1 = subset(df, df$V1 > 200, select = c("V1","V2"))
head(df_1)
```

	V1 <int>	V2 <int>
185	201	0
186	202	0
187	203	0
188	204	0
189	205	0
190	206	0

6 rows

```
nrow(df_1)
```

```
## [1] 98
```

```
head(mtcars)
```

	mpg <dbl>	cyl <dbl>	disp <dbl>	hp <dbl>	drat <dbl>	wt <dbl>	qsec <dbl>	vs <dbl>	am <dbl>
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1

	mpg <dbl>	cyl <dbl>	disp <dbl>	hp <dbl>	drat <dbl>	wt <dbl>	qsec <dbl>	vs <dbl>	am <dbl>
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0

6 rows | 1-10 of 12 columns

```
mtcars[mtcars$mpg > 20 , ]
```

	mpg <dbl>	cyl <dbl>	disp <dbl>	hp <dbl>	drat <dbl>	wt <dbl>	qsec <dbl>	vs <dbl>	am <dbl>
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0

1-10 of 14 rows | 1-10 of 12 columns Previous **1** 2 Next

```
mtcars[mtcars$mpg > 20 , c('mpg')]
```

```
##  [1] 21.0 21.0 22.8 21.4 24.4 22.8 32.4 30.4 33.9 21.5 27.3 26.0 30.4 21.4
```

```
mtcars[(mtcars$mpg > 20) & (mtcars$cyl != 6), c('wt','cyl','mpg')]
```

	wt <dbl>	cyl <dbl>	mpg <dbl>
Datsun 710	2.320	4	22.8
Merc 240D	3.190	4	24.4

	wt <dbl>	cyl <dbl>	mpg <dbl>
Merc 230	3.150	4	22.8
Fiat 128	2.200	4	32.4
Honda Civic	1.615	4	30.4
Toyota Corolla	1.835	4	33.9
Toyota Corona	2.465	4	21.5
Fiat X1-9	1.935	4	27.3
Porsche 914-2	2.140	4	26.0
Lotus Europa	1.513	4	30.4

1-10 of 11 rows Previous **1** 2 Next

Select all Numeric Columns or All String/Character Columns

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
head(mpg)
```

manufacturer <chr>	model <chr>	displ <dbl>	year <int>	cyl <int>	trans <chr>	drv <chr>	cty <int>	hwy <int>	fl <chr>
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p
audi	a4	2.0	2008	4	auto(av)	f	21	30	p
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p

6 rows | 1-10 of 11 columns

```
Filter(is.numeric, mpg)
```

displ <dbl>	year <int>	cyl <int>	cty <int>	hwy <int>
1.8	1999	4	18	29
1.8	1999	4	21	29

displ <dbl>	year <int>	cyl <int>	cty <int>	hwy <int>
2.0	2008	4	20	31
2.0	2008	4	21	30
2.8	1999	6	16	26
2.8	1999	6	18	26
3.1	2008	6	18	27
1.8	1999	4	18	26
1.8	1999	4	16	25
2.0	2008	4	20	28

1-10 of 234 rows Previous **1** 2 3 4 5 6 ... 24 Next

```
Filter(is.character, mpg)
```

manufacturer <chr>	model <chr>	trans <chr>	drv <chr>	fl <chr>	class <chr>
audi	a4	auto(l5)	f	p	compact
audi	a4	manual(m5)	f	p	compact
audi	a4	manual(m6)	f	p	compact
audi	a4	auto(av)	f	p	compact
audi	a4	auto(l5)	f	p	compact
audi	a4	manual(m5)	f	p	compact
audi	a4	auto(av)	f	p	compact
audi	a4 quattro	manual(m5)	4	p	compact
audi	a4 quattro	auto(l5)	4	p	compact
audi	a4 quattro	manual(m6)	4	p	compact

1-10 of 234 rows Previous **1** 2 3 4 5 6 ... 24 Next

```
M1 = matrix(c(1234,2235,67,85),nrow=2)
m1_df <- as.data.frame(M1)
colnames(m1_df) = c('roll_no','marks_mod1')
m1_df
```

roll_no <dbl>	marks_mod1 <dbl>

roll_no <dbl>	marks_mod1 <dbl>
1234	67
2235	85

2 rows

```
M2 = matrix(c(1234,2235,75,68),nrow=2)
m2_df <- as.data.frame(M2)
colnames(m2_df) = c('roll_no','marks_mod2')
m2_df
```

roll_no <dbl>	marks_mod2 <dbl>
1234	75
2235	68

2 rows

```
merge(m1_df,m2_df,by = "roll_no")
```

roll_no <dbl>	marks_mod1 <dbl>	marks_mod2 <dbl>
1234	67	75
2235	85	68

2 rows

Merge Dataframes

```
M1 = matrix(c(1234,2235,67,85),nrow=2)
m1_df <- as.data.frame(M1)
colnames(m1_df) = c('roll_no','marks_mod1')
m1_df
```

roll_no <dbl>	marks_mod1 <dbl>
1234	67
2235	85

2 rows

```
M2 = matrix(c(1234,2235,75,68),nrow=2)
m2_df <- as.data.frame(M2)
colnames(m2_df) = c('r_no','marks_mod2')
m2_df
```

r_no <dbl>	marks_mod2 <dbl>
1234	75
2235	68

2 rows

```
M3 = matrix(c(1234,2235,50,45),nrow=2)
m3_df <- as.data.frame(M3)
colnames(m3_df) = c('r_no','marks_mod3')
m3_df
```

r_no <dbl>	marks_mod3 <dbl>
1234	50
2235	45

2 rows

```
merge(m1_df,m2_df,by.x = "roll_no", by.y = "r_no")
```

roll_no <dbl>	marks_mod1 <dbl>	marks_mod2 <dbl>
1234	67	75
2235	85	68

2 rows

```
merge(m2_df,m3_df,by = "r_no")
```

r_no <dbl>	marks_mod2 <dbl>	marks_mod3 <dbl>
1234	75	50
2235	68	45

2 rows

Store Datframe to a File

```
write.csv(m2_df,"new_df.csv")
```