DATA FRAMES

(With inbuilt data sets and vectors)

Subject: Computer Algebra Systems and Related Software

Name: Harneet Kaur Matharoo

Roll no.: MAT/19/23

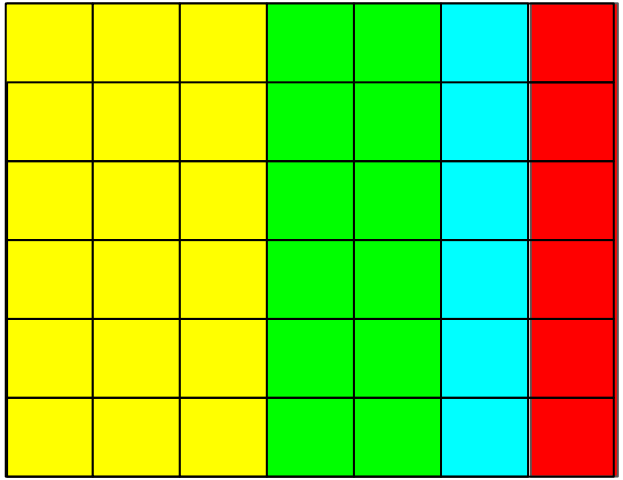
What is a Data Frame?

A Data Frame is a two-dimensional array-like structure which contains different types of objects. It is a collection of columns that can be of different objects.
There are three types of data frames:

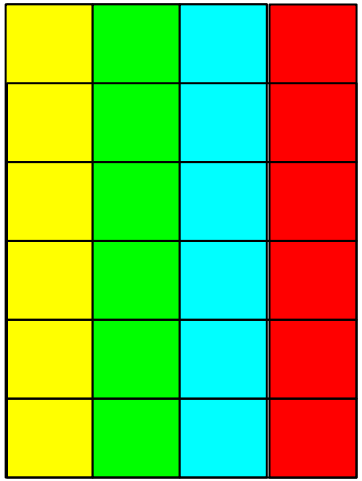
- Using inbuilt sets
- Using vectors
- Using excel sheets

Data Frames can have different types of data inside it. While the first column can be character, the second and third can be numeric or logical. However, each column should have the same type of data.

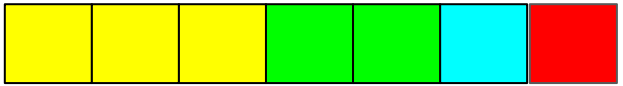
Data Frame



Represent columns of equal length having same type of elements



Represents rows of the data frame



Operations Performed on Data Frame in R

- Creating a DataFrame
- Accessing rows and columns
- Selecting the subset of the data frame
- Editing dataframes
- Adding extra rows and columns to the data frame
- Add new variables to dataframe based on existing ones

```
> day<-c("Monday","Tuesday","Wednesday","Thursda
y","Friday","Saturday","Sunday")
> date<-c(19,20,21,22,23,24,25)
> week<-data.frame(day,date)
> print(week)
        day date
1    Monday   19
2   Tuesday   20
3 Wednesday   21
4  Thursday   22
5    Friday   23
6  Saturday   24
7    Sunday   25
> class(week)
[1] "data.frame"
>
```

```
> data_frame<-data.frame(x=1:7,y=letters[1:7])
> print(data_frame)
  x y
1 1 a
2 2 b
3 3 c
4 4 d
5 5 e
6 6 f
7 7 g
>
```

```
> data<-data.frame(id=c(1:5),name=c("Rick","Dan","Michell
e","Ryan","Gary"),salary=c(623.3,515.2,611.0,729.0,843.25),
start_date=c("1st January, 2012","23rd September, 2013","15
th November,2014","11th May,2014","27th March, 2015"))
> print(data)
  id      name salary           start_date
1  1      Rick 623.30    1st January, 2012
2  2       Dan 515.20 23rd September, 2013
3  3 Michelle 611.00   15th November,2014
4  4      Ryan 729.00        11th May,2014
5  5      Gary 843.25     27th March, 2015
>
```

Inbuilt data set : **USArrests**

I. Creating a Data Frame

Data Frame using vectors

The data.frame() function is used to create a data frame and then vectors created are passed as arguments to the function.

For creating a data frame using vectors:

- Vectors should be of same length
- Type of vectors may differ
- Length of vectors should be equal

```
> data<-data.frame(id=c(1:5),name=c("Rick","Dan","Michell
e","Ryan","Gary"),salary=c(623.3,515.2,611.0,729.0,843.25),
start_date=c("1st January, 2012","23rd September, 2013","15
th November,2014","11th May,2014","27th March, 2015"))
> print(data)
  id      name salary           start_date
1  1      Rick 623.30    1st January, 2012
2  2       Dan 515.20 23rd September, 2013
3  3 Michelle 611.00   15th November,2014
4  4      Ryan 729.00        11th May,2014
5  5      Gary 843.25     27th March, 2015
>
```

- For verifying the class of the function, we use class() command.
- length() command displays the total number of columns present in the data set.

```
> length(week)
[1] 2
> data_frame<-data.frame(x=1:7,y=letters[1:7])
> length(data_frame)
[1] 2
> length(USArrests)
[1] 4
>
> length(USArrests$Murder)
[1] 50
>
```

```
> day<-c("Monday","Tuesday","wednesday","Thursda
y","Friday","Saturday","Sunday")
> date<-c(19,20,21,22,23,24,25)
> week<-data.frame(day,date)
> print(week)
        day date
1    Monday   19
2   Tuesday   20
3 Wednesday   21
4  Thursday   22
5    Friday   23
6  Saturday   24
7    Sunday   25
> class(week)
[1] "data.frame"
>
```

Data Frames using inbuilt sets

- data() command displays all the data sets available in R.
- head() command displays the first 6 rows of the data set
- tail() command displays the last 6 rows of the data set.

```
> head(USArrests)
           Murder Assault UrbanPop Rape
Alabama      13.2     236       58 21.2
Alaska       10.0     263       48 44.5
Arizona       8.1     294       80 31.0
Arkansas      8.8     190       50 19.5
California    9.0     276       91 40.6
Colorado      7.9     204       78 38.7
>
> tail(USArrests)
             Murder Assault UrbanPop Rape
Vermont         2.2      48       32 11.2
Virginia        8.5     156       63 20.7
Washington      4.0     145       73 26.2
West Virginia   5.7      81       39  9.3
Wisconsin       2.6      53       66 10.8
Wyoming         6.8     161       60 15.6
>
```

II. Accessing rows and columns

Row names and column names of any data frame can be altered using rownames() and colnames() commands.

rownames(): Displays the names of all the rows present in the data frame.

colnames(): Displays the names of all the columns present in the data frame.

```
> rownames(week)
[1] "1" "2" "3" "4" "5" "6" "7"
> colnames(week)
[1] "day"  "date"
>
```

```
> rownames(data_frame)<-c("First","Second","Thir
d","Fourth","Fifth","Sixth","Seventh")
> colnames(data_frame)<-c("Numbers","Letters")
> data_frame
        Numbers Letters
First         1       a
Second        2       b
Third         3       c
Fourth        4       d
Fifth         5       e
Sixth         6       f
Seventh       7       g
>
```

dimnames() displays the row names and column names in a single command.

```
> dimnames(USArrests)
[[1]]
 [1] "Alabama"        "Alaska"         "Arizona"        "Arkansas"       "California"
 [6] "Colorado"       "Connecticut"    "Delaware"       "Florida"        "Georgia"
[11] "Hawaii"         "Idaho"          "Illinois"       "Indiana"        "Iowa"
[16] "Kansas"         "Kentucky"       "Louisiana"      "Maine"          "Maryland"
[21] "Massachusetts"  "Michigan"       "Minnesota"      "Mississippi"    "Missouri"
[26] "Montana"        "Nebraska"       "Nevada"         "New Hampshire"  "New Jersey"
[31] "New Mexico"     "New York"       "North Carolina" "North Dakota"   "Ohio"
[36] "Oklahoma"       "Oregon"         "Pennsylvania"   "Rhode Island"   "South Carolina"
[41] "South Dakota"   "Tennessee"      "Texas"          "Utah"           "Vermont"
[46] "Virginia"       "Washington"     "West Virginia"  "Wisconsin"      "Wyoming"

[[2]]
[1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

- max() gives maximum value in the complete data set
- min() gives the minimum value of the complete data set
- sum() gives the sum of all entries in the data set.
- length() displays the number of columns in the data set

```
> max(USArrests)
[1] 337
> min(USArrests)
[1] 0.8
> sum(USArrests)
[1] 13266
> length(USArrests)
[1] 4
>
```

```
> mean(USArrests$Murder)
[1] 7.788
> median(USArrests$Murder)
[1] 7.25
> var(USArrests$Murder)
[1] 18.97047
> sd(USArrests$Murder)
[1] 4.35551
>
```

- mean() displays the mean of the specific column in the data set
- median() display the median of the specific column in the data set
- var() displays the variance of the specific column in the data set
- sd() displays the standard deviation of the specific column in the data set.

summary() command is used to to summarize the data from a Data Frame, which includes mean, median, quantile, maximum and minimum values of each column present in the data frame.

```
> summary(USArrests)
     Murder          Assault         UrbanPop          Rape       
 Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30  
 1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07  
 Median : 7.250   Median :159.0   Median :66.00   Median :20.10  
 Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23  
 3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18  
 Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00  
> 
```

- `rowSums()`: displays the sum of individual rows in the data set
- `rowMeans()`: displays mean of the individual rows in the data set.
- `colSums()`: displays the sum of individual columns in the data set.
- `colMeans()`: displays mean of the individual columns in the data set

```
> rowSums(USArrests)
       Alabama         Alaska        Arizona       Arkansas     California       Colorado
         328.4          365.5          413.1          268.3          416.6          328.6
   Connecticut       Delaware        Florida        Georgia         Hawaii          Idaho
         201.4          331.7          462.3          314.2          154.5          190.8
      Illinois        Indiana           Iowa         Kansas       Kentucky      Louisiana
         366.4          206.2          126.5          205.0          187.0          352.6
         Maine       Maryland  Massachusetts       Michigan      Minnesota   Mississippi
         143.9          406.1          254.7          376.2          155.6          336.2
      Missouri        Montana       Nebraska         Nevada  New Hampshire     New Jersey
         285.2          184.4          184.8          391.2          124.6          274.2
    New Mexico       New York North Carolina   North Dakota           Ohio       Oklahoma
         398.5          377.2          411.1           97.1          223.7          245.6
        Oregon   Pennsylvania   Rhode Island South Carolina   South Dakota      Tennessee
         260.2          199.2          272.7          363.9          147.6          287.1
         Texas           Utah        Vermont       Virginia     Washington  West Virginia
         319.2          226.1           93.4          248.2          248.2          135.0
     Wisconsin        Wyoming
         132.4          243.4
> colSums(USArrests)
  Murder  Assault UrbanPop     Rape
   389.4   8538.0   3277.0   1061.6
>
```

```
> rowMeans((USArrests))
     Alabama       Alaska      Arizona     Arkansas   California     Colorado
      82.100       91.375      103.275       67.075      104.150       82.150
  Connecticut     Delaware      Florida      Georgia       Hawaii        Idaho
      50.350       82.925      115.575       78.550       38.625       47.700
     Illinois      Indiana         Iowa       Kansas     Kentucky    Louisiana
      91.600       51.550       31.625       51.250       46.750       88.150
        Maine      Maryland Massachusetts     Michigan    Minnesota  Mississippi
      35.975      101.525       63.675       94.050       38.900       84.050
     Missouri      Montana     Nebraska       Nevada New Hampshire   New Jersey
      71.300       46.100       46.200       97.800       31.150       68.550
   New Mexico     New York North Carolina North Dakota         Ohio     Oklahoma
      99.625       94.300      102.775       24.275       55.925       61.400
      Oregon Pennsylvania Rhode Island South Carolina South Dakota    Tennessee
      65.050       49.800       68.175       90.975       36.900       71.775
       Texas         Utah      Vermont     Virginia   Washington West Virginia
      79.800       56.525       23.350       62.050       62.050       33.750
   Wisconsin      Wyoming
      33.100       60.850

> colMeans((USArrests))
  Murder  Assault UrbanPop     Rape
   7.788  170.760   65.540   21.232
>
```

sort(): used for sorting the data in ascending order whereas rev(sort()) arranges data in descending order.

order(): tells the positioning of elements when arranged in ascending order whereas rev(order()) tells the positioning of elements in descending order.

```
> sort(data_frame$Numbers)
[1] 1 2 3 4 5 6 7
> rev(sort(data_frame$Numbers))
[1] 7 6 5 4 3 2 1
>

> order(data_frame$Letters)
[1] 1 2 3 4 5 6 7
> rev(order(data_frame$Letters))
[1] 7 6 5 4 3 2 1
>
```

III. Selecting a subset of a Data Frame (Subsetting)

```
> data_frame[3,]
      Numbers Letters
Third       3       c
> data_frame[c(1,3,5),]
      Numbers Letters
First       1       a
Third       3       c
Fifth       5       e
> data_frame[-2,]
        Numbers Letters
First         1       a
Third         3       c
Fourth        4       d
Fifth         5       e
Sixth         6       f
Seventh       7       g
>
```

For rows:

- [3,] displays the 3rd row.
- [c(2,4,6),] displays just the rows 2, 4 and 6
- [−2,] displays all the rows except the second row

For columns:

- [,2] displays just the second column
- [, c(2,4,6)] displays the 2nd, 4th , and 6th columns
- [, −2] displays every column except the second column.

```
> data_frame[,2]
[1] "a" "b" "c" "d" "e" "f" "g"
> data_frame[,-2]
[1] 1 2 3 4 5 6 7
>
```

Bibliography

- Class notes
- Beginning R The Statistical Programming Language (Book)

Examples from:

- https://www.w3schools.com/r/r_data_frames.asp
- https://www.tutorialspoint.com/r/r_data_frames.htm
- https://www.geeksforgeeks.org/dataframe-operations-in-r/

THANK YOU!