## Creating a csv file:

Using files create() function, a new file can be created from console, or trancates if already exists.

### Syntax:
file.create(" ").

### Example:
* Create a File
* The file created can be seen
* In your working directory.

## Reading a csv file:

Using read table() function R, files can be read and output is shown as as data frame.

### Syntax:
read.title
    read.table (Cfile).

### Ex:
# Reading txt file.
    data ← read.table (file = "GFG.txt").

# Print
    print (data)

## Writing into a csv file:

R can create csv file form existing data frame. The write.csv() function is used.

---

To create the csv file. This file gets created in the working directory.

```
# Create a data frame
    data ← read.csv ("input.csv").

retual ← subset (data, as.Data (start-date)
        >as.Date (2004-10-01°)).

# Write filtered data into a new file.
    write.csv (retual, "output.csv")
    newdata ← read.csv ("output.csv")
    print (newdata).
```

## Data Manipulation:

round (x,n) # round the values of x to n decimal places

ceiling (x) # vector x of small integers >x

floor (x) # vector x of largest integer <x

as.integer # Truncates real x to integers (compare to round (x,0)).

## Statistics:

min() → lowest value from given data

mean() → Average value

median() → Middle value $Q_1$, $Q_2$, $Q_3$.

sum() → Total.

var() → # produces the variance covariance matrix.

---

sd() → # standard deviation.

## Transformation:

five num() → # Tukey five numbers min, lower hinge, median, upper hinge, max.

Table () → # frequency counts of entries, ideally the entries are factors (although it works with integers or even reals).

scale (data, scale=T); # centers around the mean and scales by the sd.

## Input and Display:

read table (file name, header = True) →
    # read files with tables in first row.
    # read a tab or space delimited file

read.table (file name. header = True, sep= "1")
    # read csv files.

x = c(1:10) → # create a data vector with elements 1-10.

vect = c(x,y) → # combine them into a vector matrix.

Suppose that the data for the analysis includes the attribute age, the age value for the data tuples are 13, 15, 16, 16, 19, 20, 20, 21, 22, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70. Find the first quartile and the 3rd quartile of the data.

+ The first quartile ($Q_1$) is the 25th percentile and the third quartile ($Q_3$) is the 75th percentile in a data set.

* To find $Q_1$ and $Q_3$ we first need to order the data set and find the median ($Q_2$). The formula to find the median is,

for odd number of elements in the dataset.

Median = $(N+1)/2$ the element of the sorted dataset, where N is the numbers of the elements in the dataset.

for even number of elements in the dataset

Median = ($(N/2)$th element + $(N/2+1)$th element)/2 of the sorted dataset, when N is the number of elements in the dataset.

Have we have 26 elements in the dataset, so the median is the average of the 13th & 14th elements, which are 19 and 20 respectively.

Therefore $Q_2 = (19+20)/2 = 19.5$

Now that we have $Q_2$, we can find $Q_1$.

and $Q_3$ by finding the median of the lower and upper halves of the data set respectively.

For the lower half of the data set, we have the following values.

13, 15, 16, 16, 19.

The median of this set is 16.

For the upper half of the data set have the following values:

20, 20, 21, 22, 22, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

The median of this set is 35.

$Q_3 = 35$.

Therefore the first quartile $(Q_1) = 16$

and the third quartile $(Q_3) = 35$