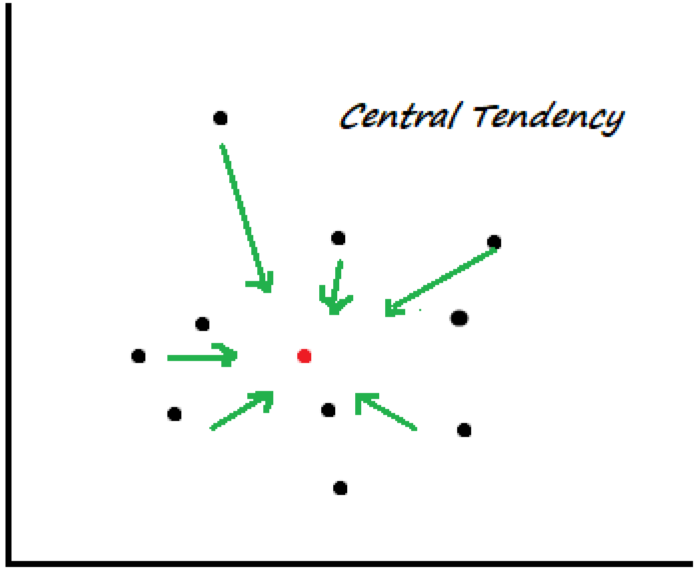
Measures of Central Tendency



There are three measures of central tendencies:

mean

median

mode

Mean

The **Arithmetic Mean** or simply **Mean** is the average of the numbers:
a calculated "central" value of a set of numbers.

To calculate it:

- add up all the numbers,
- then divide by how many numbers there are.

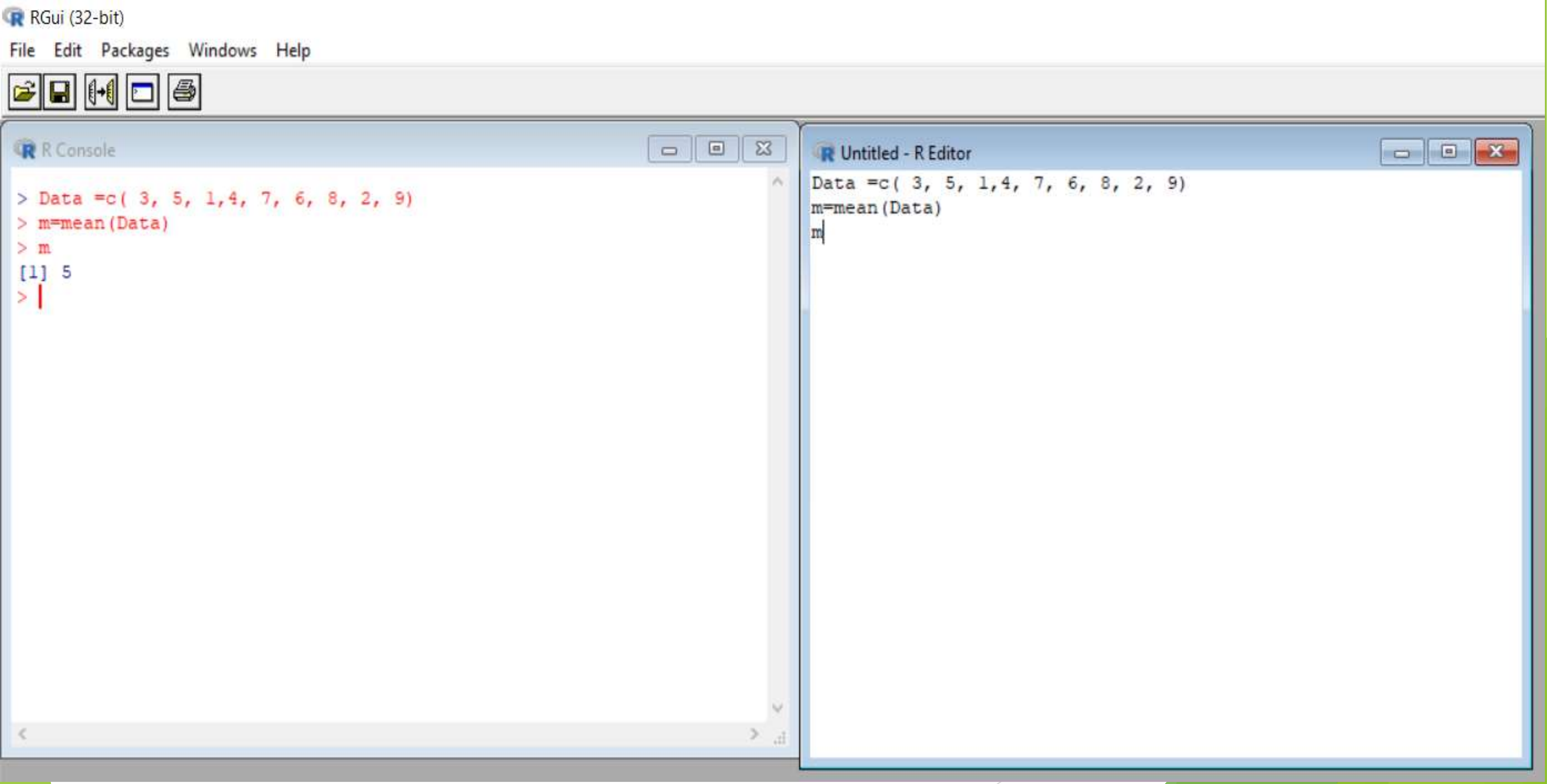
Example :

Data Set = 3, 5, 1,4, 7, 6, 8, 2, 9

Number of Elements in Data Set = 9

Mean

=(3+5+1+4+7+6+8+2+9)/ 9 =45/9=5


RGui (32-bit)
File Edit Packages Windows Help
R Console
> Data =c(3, 5, 1,4, 7, 6, 8, 2, 9)
> m=mean(Data)
> m
[1] 5
>
Untitled - R Editor
Data =c(3, 5, 1,4, 7, 6, 8, 2, 9)
m=mean(Data)
m

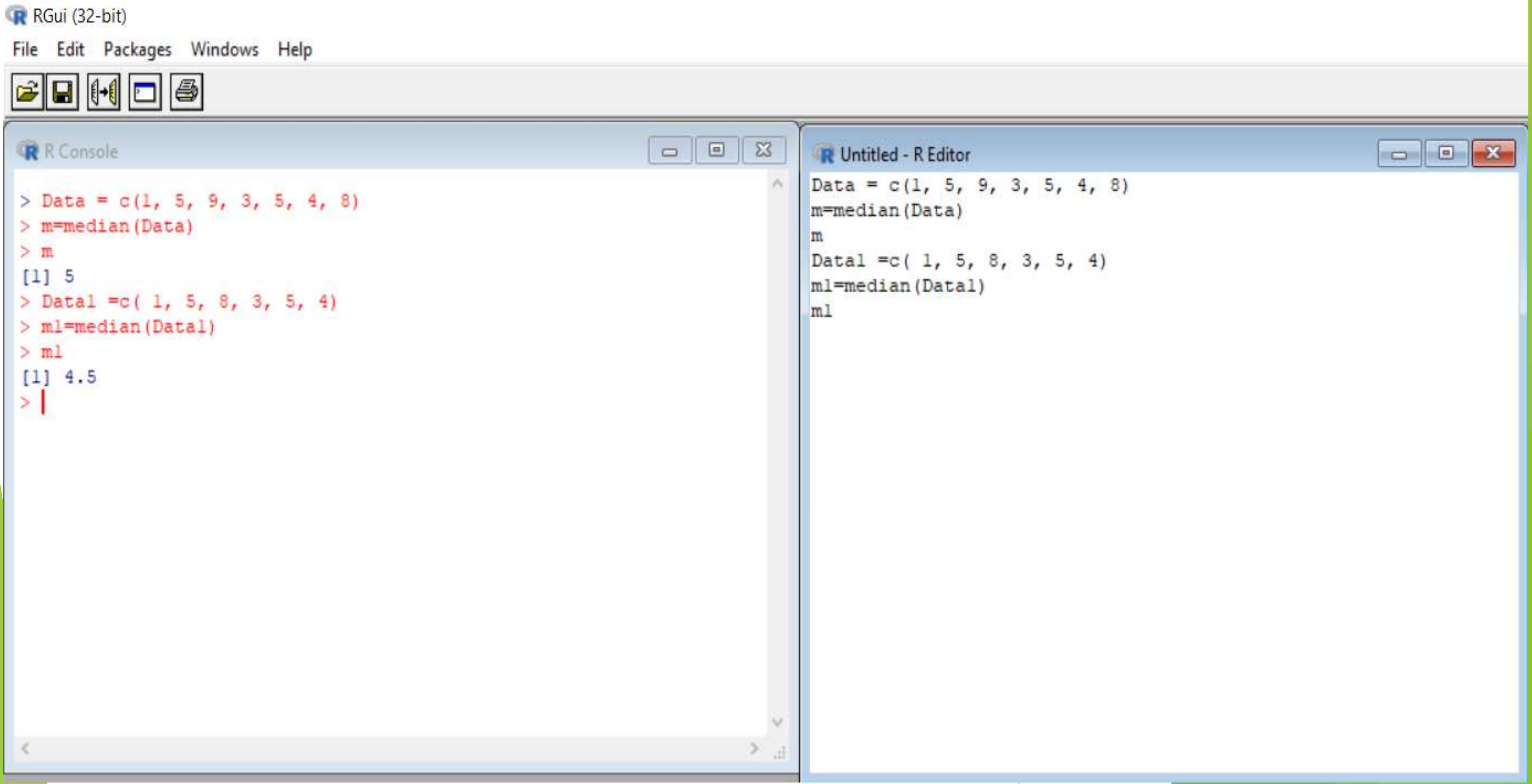
Median

The median is the middle number in a sorted, ascending or descending, list of numbers and can be more descriptive of that data set than the average.

Examples : Odd Number of Elements

Data Set = 1, 5, 9, 3, 5, 4, 8
Reordered = 1, 3, 4, 5, 5, 8, 9
Median = 5

Examples : Even Number of Elements
Data Set = 1, 5, 8, 3, 5, 4
Reordered = 1, 3, 4, 5, 5, 8
Median = (4 + 5) / 2 = 4.5


RGui (32-bit)
File Edit Packages Windows Help
R Console
> Data = c(1, 5, 9, 3, 5, 4, 8)
> m=median(Data)
> m
[1] 5
> Data1 =c(1, 5, 8, 3, 5, 4)
> m1=median(Data1)
> m1
[1] 4.5
>
Untitled - R Editor
Data = c(1, 5, 9, 3, 5, 4, 8)
m=median(Data)
m
Data1 =c(1, 5, 8, 3, 5, 4)
m1=median(Data1)
m1

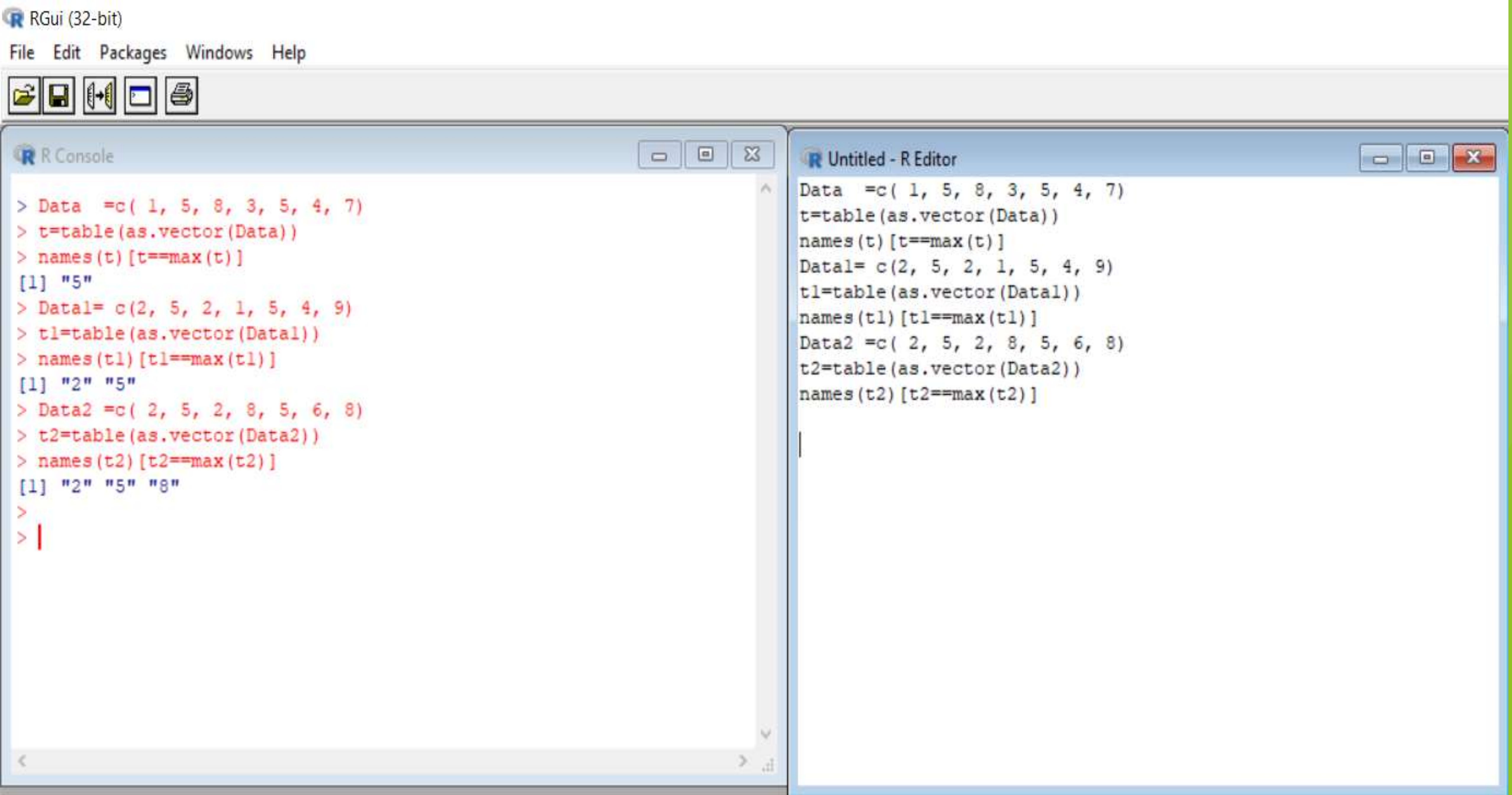
Mode

The mode is the value that appears most frequently in a data set. A set of data may have one mode, more than one mode, or no mode at all.

Examples:
Single Mode Data Set = 1, 5, 8, 3, 5, 4, 7
Mode = 5

Examples:
Bimodal Data Set = 2, 5, 2, 1, 5, 4, 9
Modes = 2 and 5

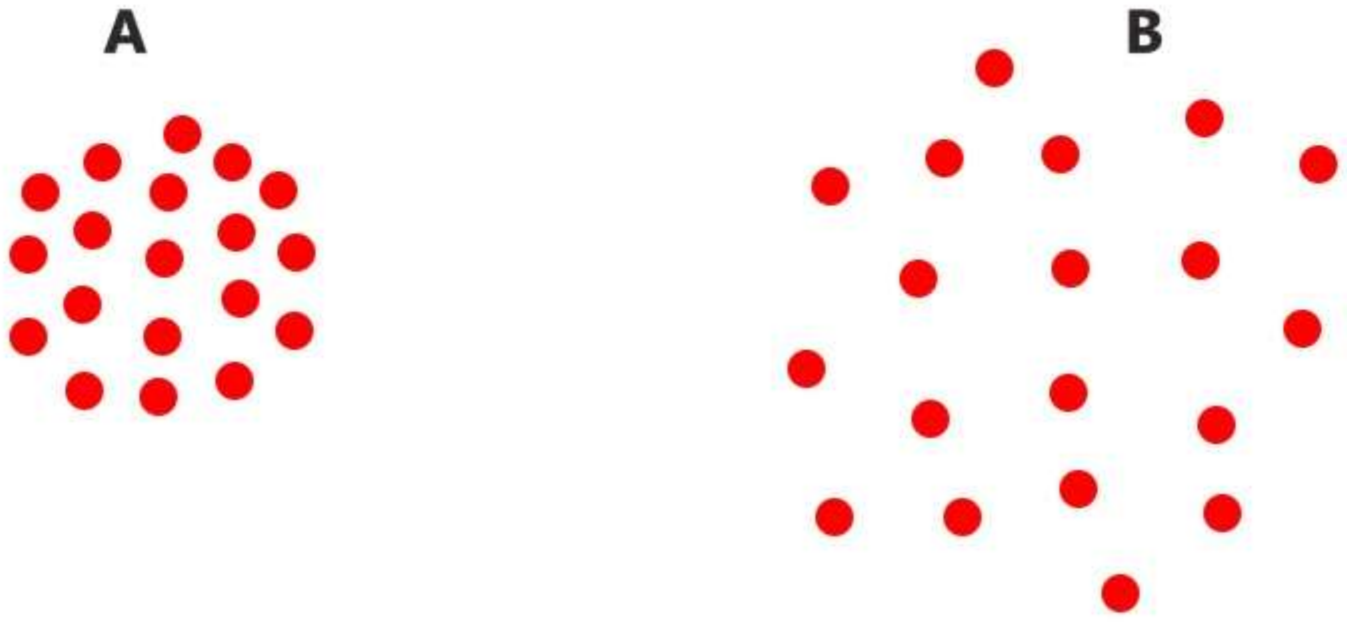
Examples:
Trimodal Data Set = 2, 5, 2, 8, 5, 6, 8
Modes = 2, 5, and 8



RGui (32-bit)
File Edit Packages Windows Help
R Console
> Data =c(1, 5, 8, 3, 5, 4, 7)
> t=table(as.vector(Data))
> names(t)[t==max(t)]
[1] "5"
> Data1= c(2, 5, 2, 1, 5, 4, 9)
> t1=table(as.vector(Data1))
> names(t1)[t1==max(t1)]
[1] "2" "5"
> Data2 =c(2, 5, 2, 8, 5, 6, 8)
> t2=table(as.vector(Data2))
> names(t2)[t2==max(t2)]
[1] "2" "5" "8"
>
>
Untitled - R Editor
Data =c(1, 5, 8, 3, 5, 4, 7)
t=table(as.vector(Data))
names(t)[t==max(t)]
Data1= c(2, 5, 2, 1, 5, 4, 9)
t1=table(as.vector(Data1))
names(t1)[t1==max(t1)]
Data2 =c(2, 5, 2, 8, 5, 6, 8)
t2=table(as.vector(Data2))
names(t2)[t2==max(t2)]

Mean- Average
Median - Middle of the data set
Mode - Most often

Measures of Dispersion

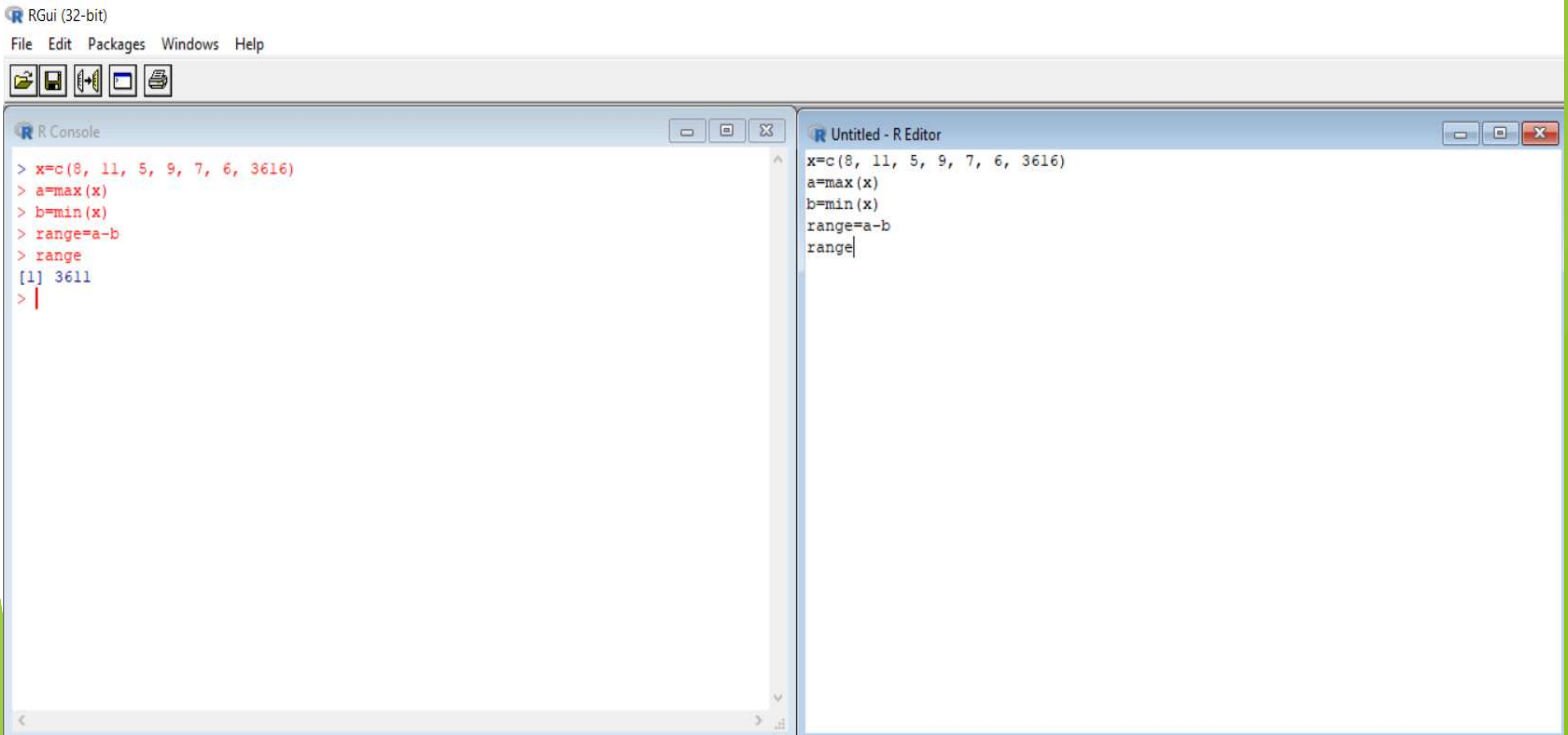


The important measures of dispersion is given by:

- Range
- Semi Inter quartile range
- Standard deviation
- Coefficient of variation

RANGE

The Range is the difference between the lowest and highest values in a data set



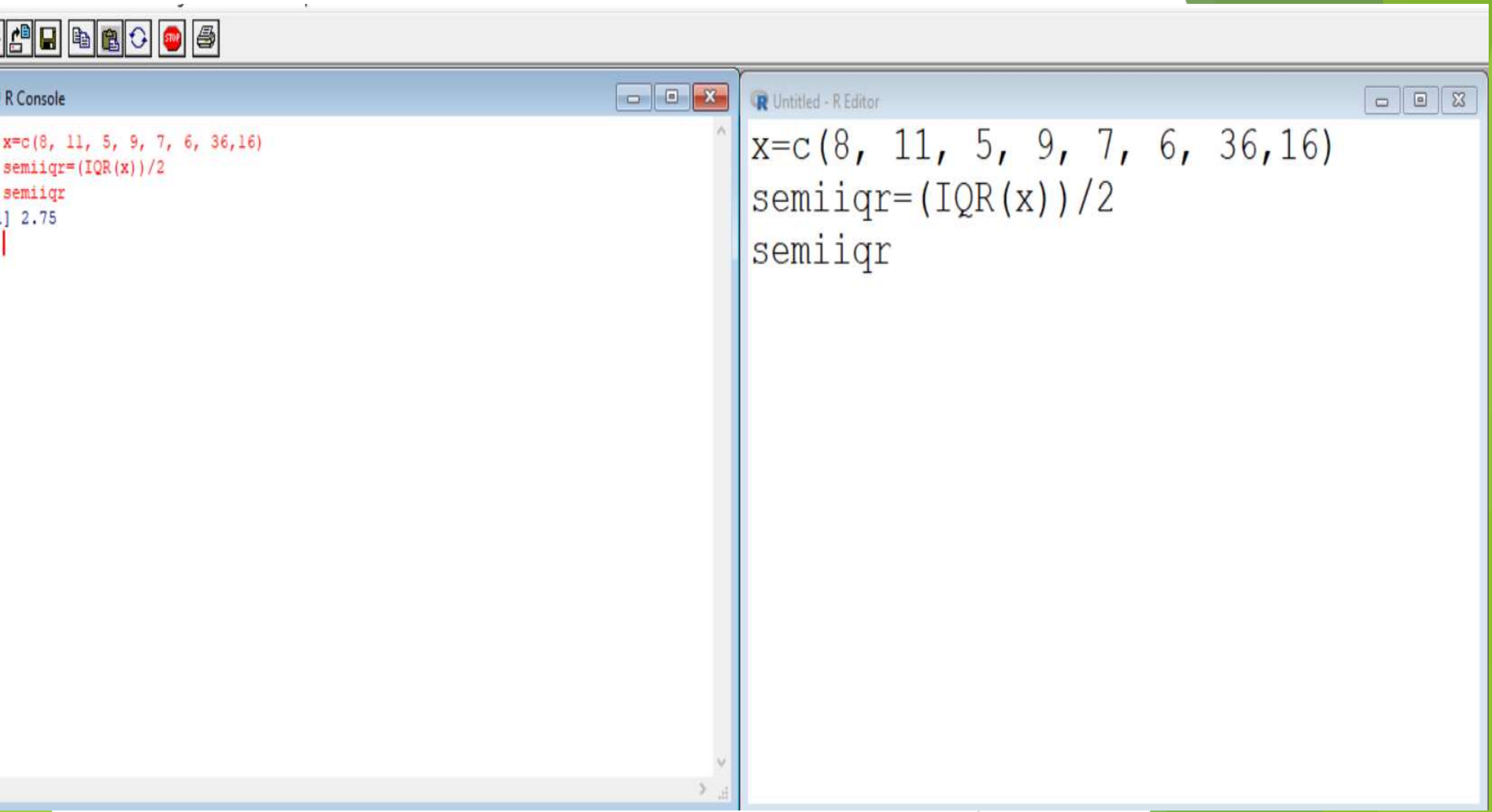
Semi Inter Quartile Deviation

It depends on the lower quartile Q1 and the upper quartile Q3.

The difference Q3−Q1 is called the inter quartile range.

The difference Q3−Q1 divided by 2 is called semi-inter quartile range

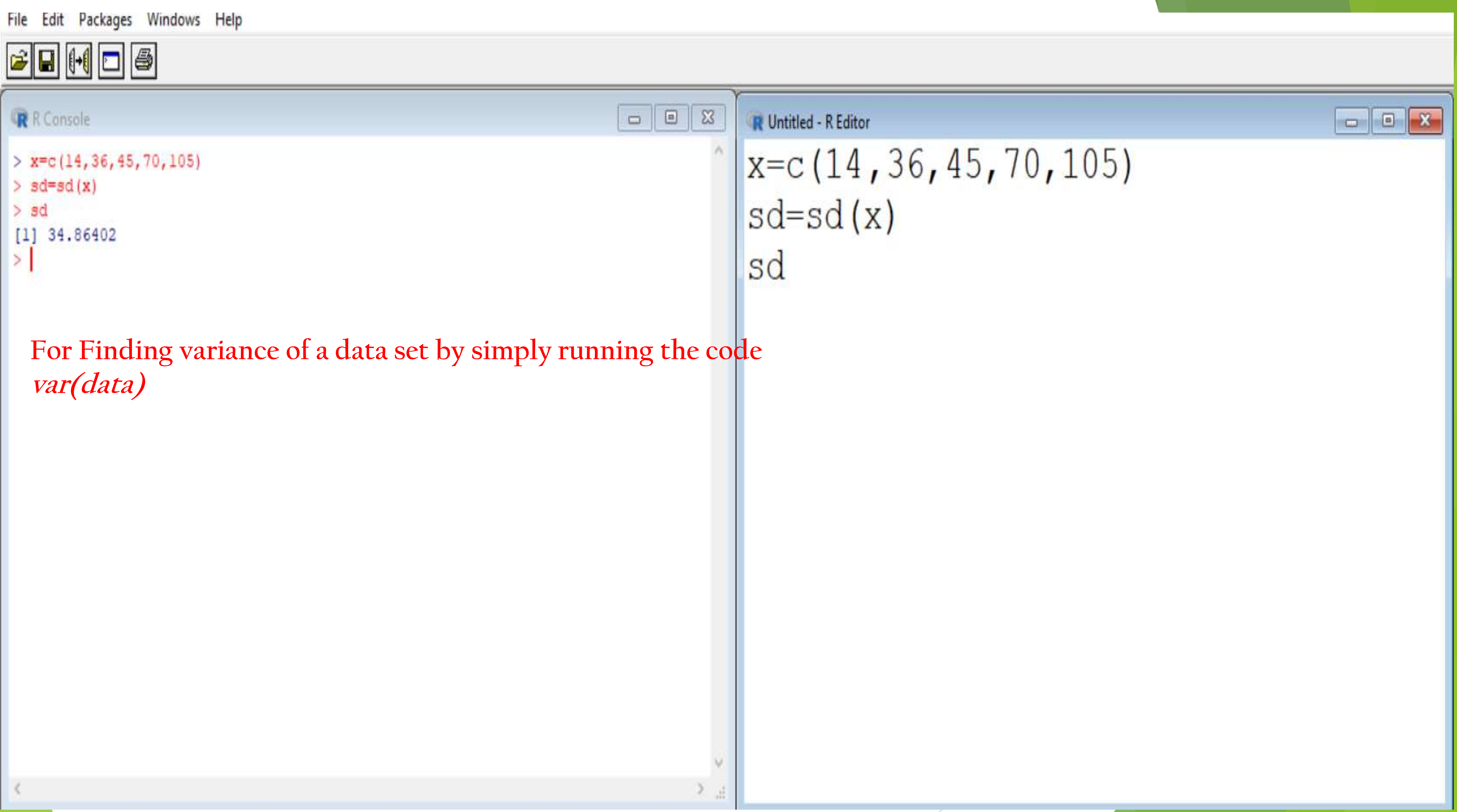
or the quartile deviation.


R Console
x=c(8, 11, 5, 9, 7, 6, 36,16)
semiiqr=(IQR(x))/2
semiiqr
] 2.75
Untitled - R Editor
x=c(8, 11, 5, 9, 7, 6, 36,16)
semiiqr=(IQR(x))/2
semiiqr

Standard Deviation

Standard deviation is the square root of the average of squared deviations of the items from their mean.

Symbolically it is represented by σ



For Finding variance of a data set by simply running the code *var(data)*

Coefficient of Variation

When comparison has to be made between two series then the relative measure of dispersion, known as coefficient of variation is used.

$$cv = \frac{\sigma}{\overline{X}} \times 100$$

Example

Problem Statement:

From the following data. Identify the risky project, is more risky:

Year	1	2	3	4	5
Project X (Cash profit in Rs. lakh)	10	15	25	30	55
Project Y (Cash profit in Rs. lakh)	5	20	40	40	30

R Console

```
> x=c(10,15,25,30,55)
> mx=mean(x)
> m
[1] 27
> sdx=sd(x)
> sdx
[1] 17.53568
> cvx=(sdx/m)*100
> cvx
[1] 64.94696
> y=c(5,20,40,40,30)
> my=mean(y)
> sdy=sd(y)
> cvy=(sdy/my)*100
> cvy
[1] 54.9348
> #Since coeff.of variation is higher for project X than for project Y, hence d$
>
```

Untitled - R Editor

```
x=c(10,15,25,30,55)
mx=mean(x)
m
sdx=sd(x)
sdx
cvx=(sdx/m)*100
cvx
y=c(5,20,40,40,30)
my=mean(y)
sdy=sd(y)
cvy=(sdy/my)*100
```

Since coeff.of variation is higher for project X than for project Y, hence project X is more risky.