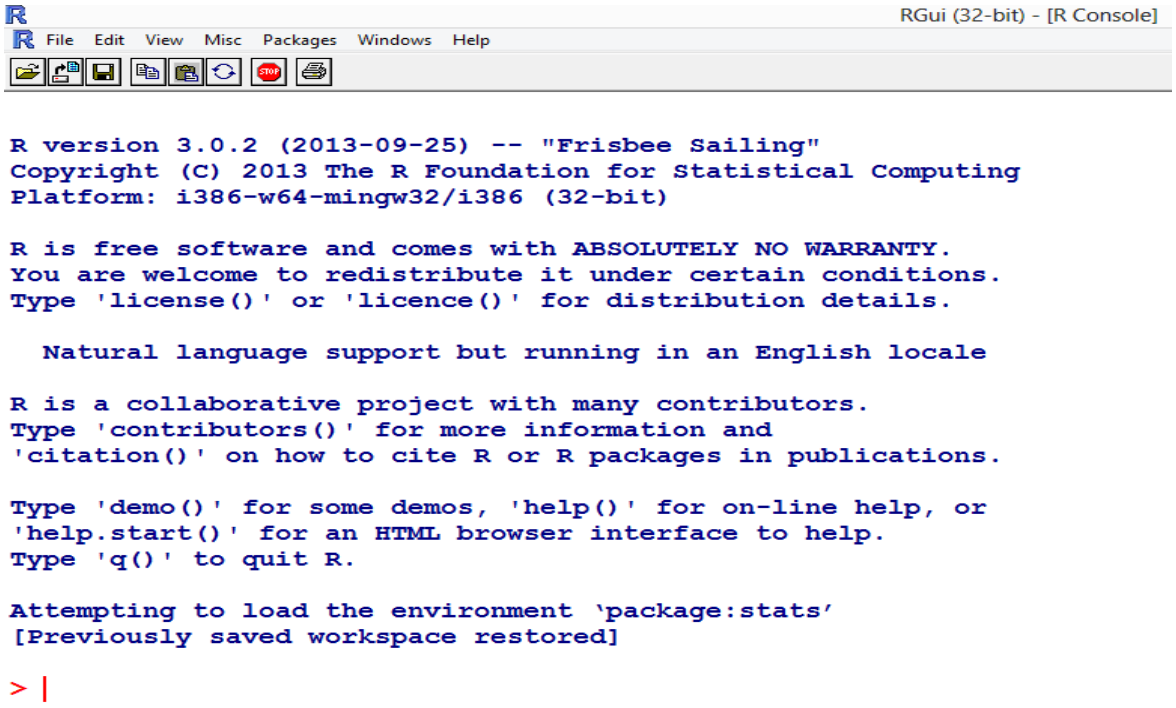
Getting Started With R

Data Science: Getting Started with R



R is an open-source software program, developed by volunteers as a service to the community of scientists, researchers, and data analysts who use it. R is free to download and use. Lots of advice is available online to help users learn R, which is good because it is a powerful and complex program, in reality a full-featured programming language dedicated to data.

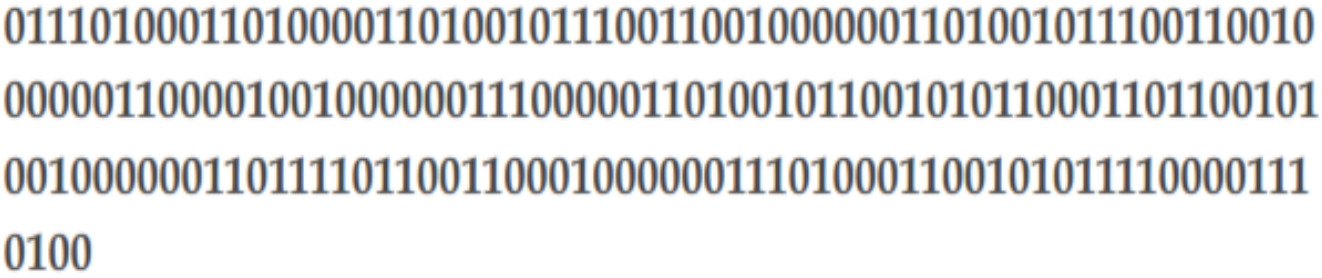
Some Basic Information on R

- Command line oriented
- Not especially good at giving feedback or error messages
- One Needs to know the data

An Introduction to Data: Strings

Character string

“this is a piece of text”



R code:

myText <- “this is a piece of text”

An Introduction to Data: Type / Mode

List of integers **43, 42, 12, 8, 5**

Each integer represents the age of a family member.

Integer list is all the same "type/mode."

R refers to a list as a "vector."

R code for this vector looks like

```
c (43, 42, 12, 8, 5)
```

```
myFamilyAges <- c (43, 42, 12, 8, 5)
```

Data set

Getting Started in R

- Start up R
- Create/key in at the command prompt '>'

  > MyText<- "this is a piece of text"

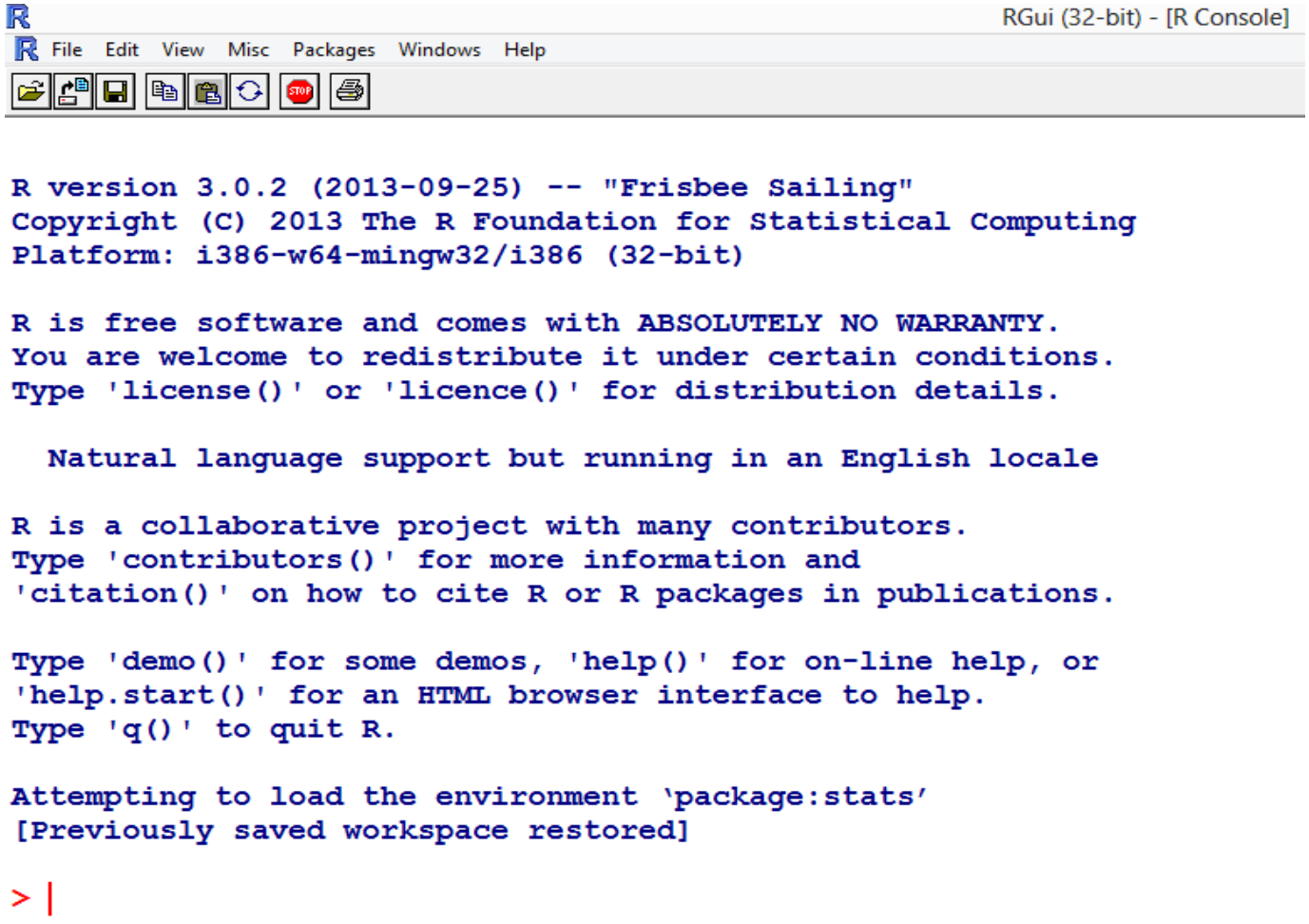
  Retrieve MyText

  Retrieve mytext
- Create/key in at the command prompt '>'

  > myFamilyAges <-c(**43, 42, 12, 8, 5**)

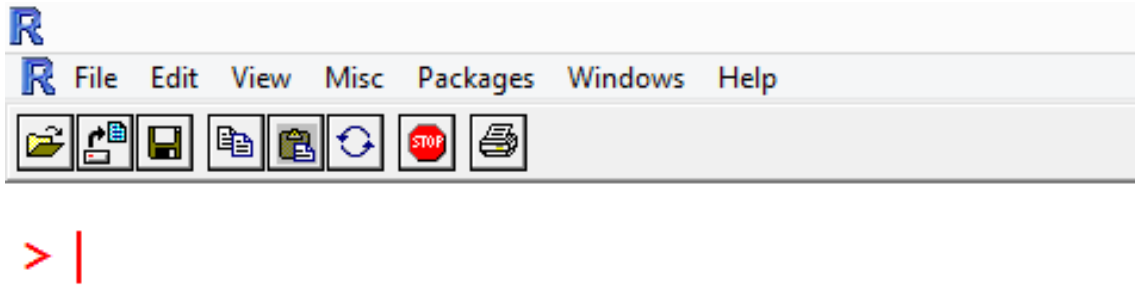
  Retrieve myFamilyAges
- End R …. q() at the command prompt '>'
- Respond 'yes' to save your work space

The Initial Console View

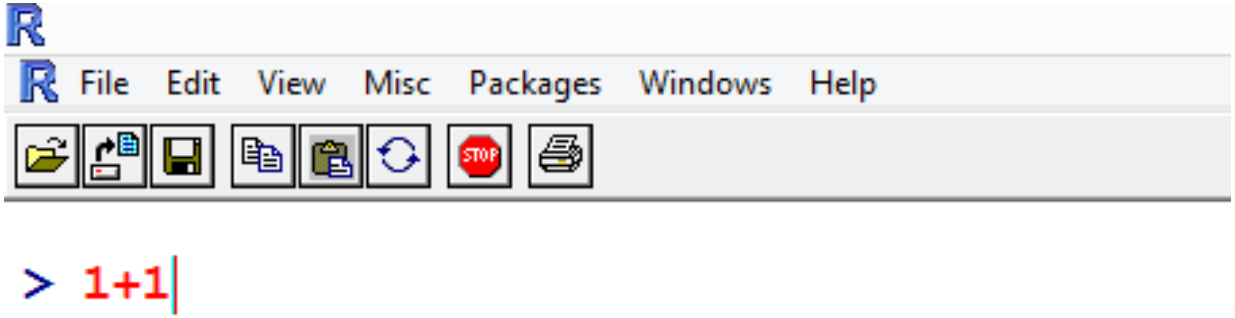


Clear the Console

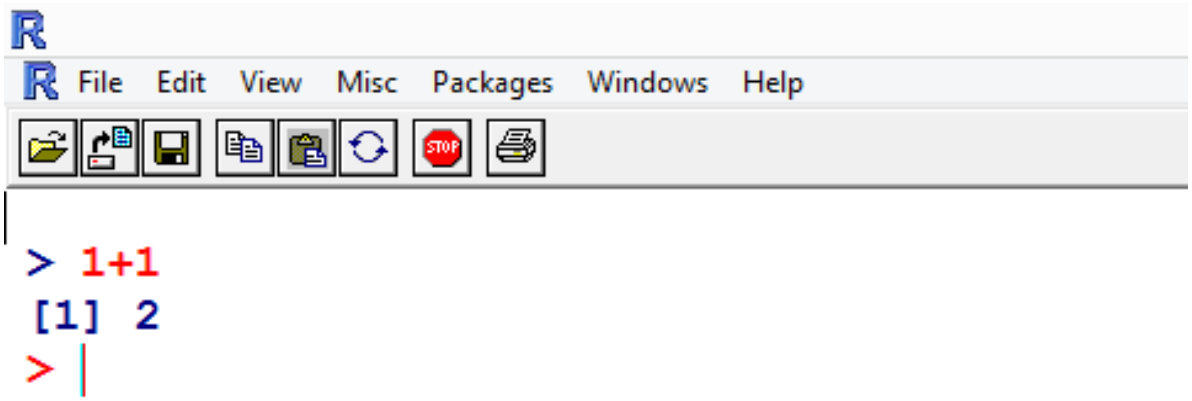
Edit → Clear Console or Control L



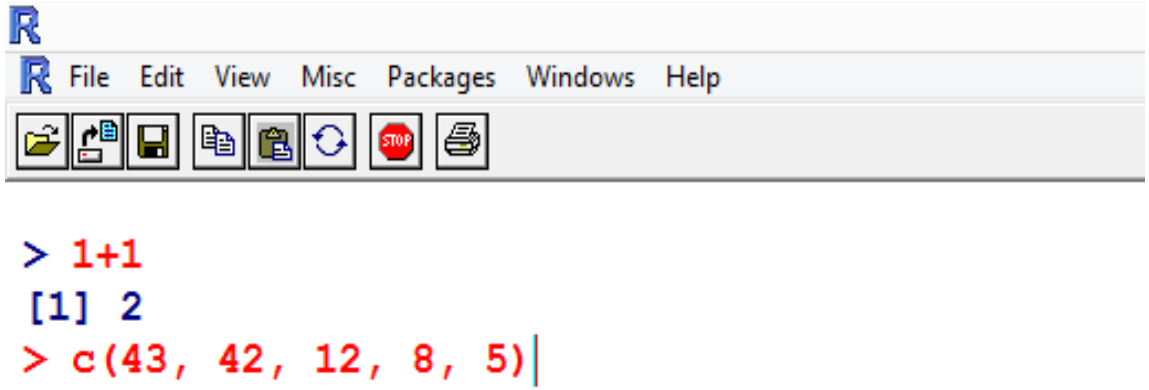
Adding Two Numbers



Adding Two Numbers

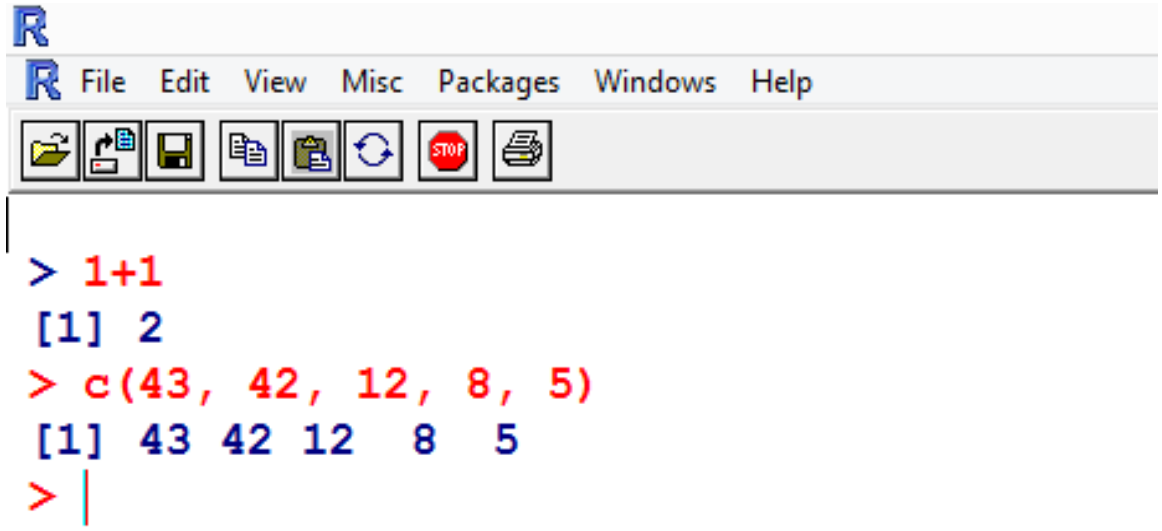


Creating a Vector

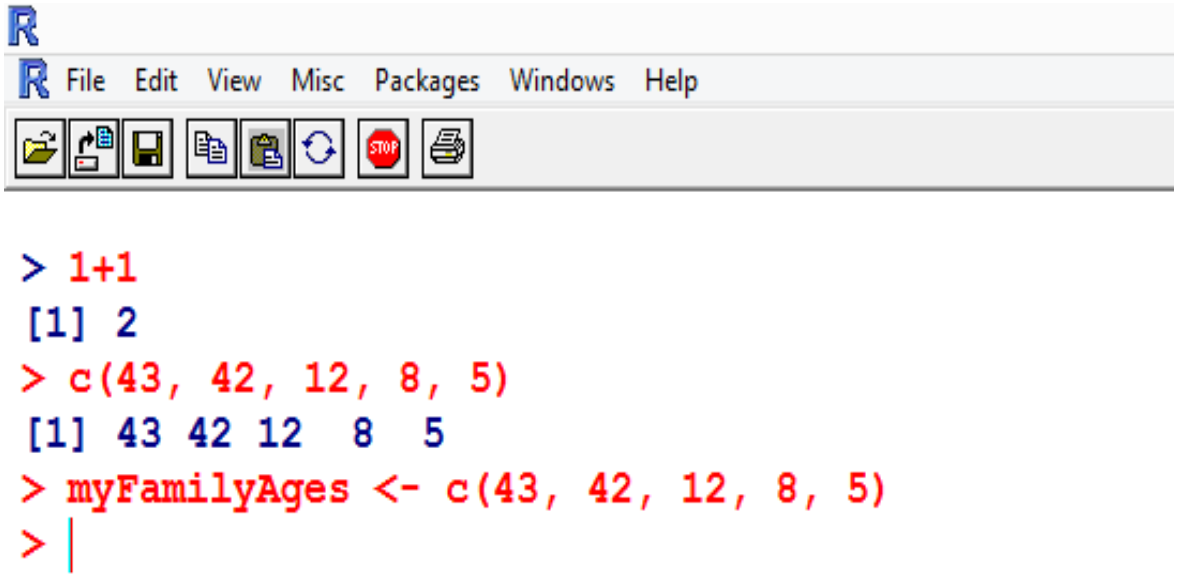


Note that the ‘c’ command stands for concatenate of combine

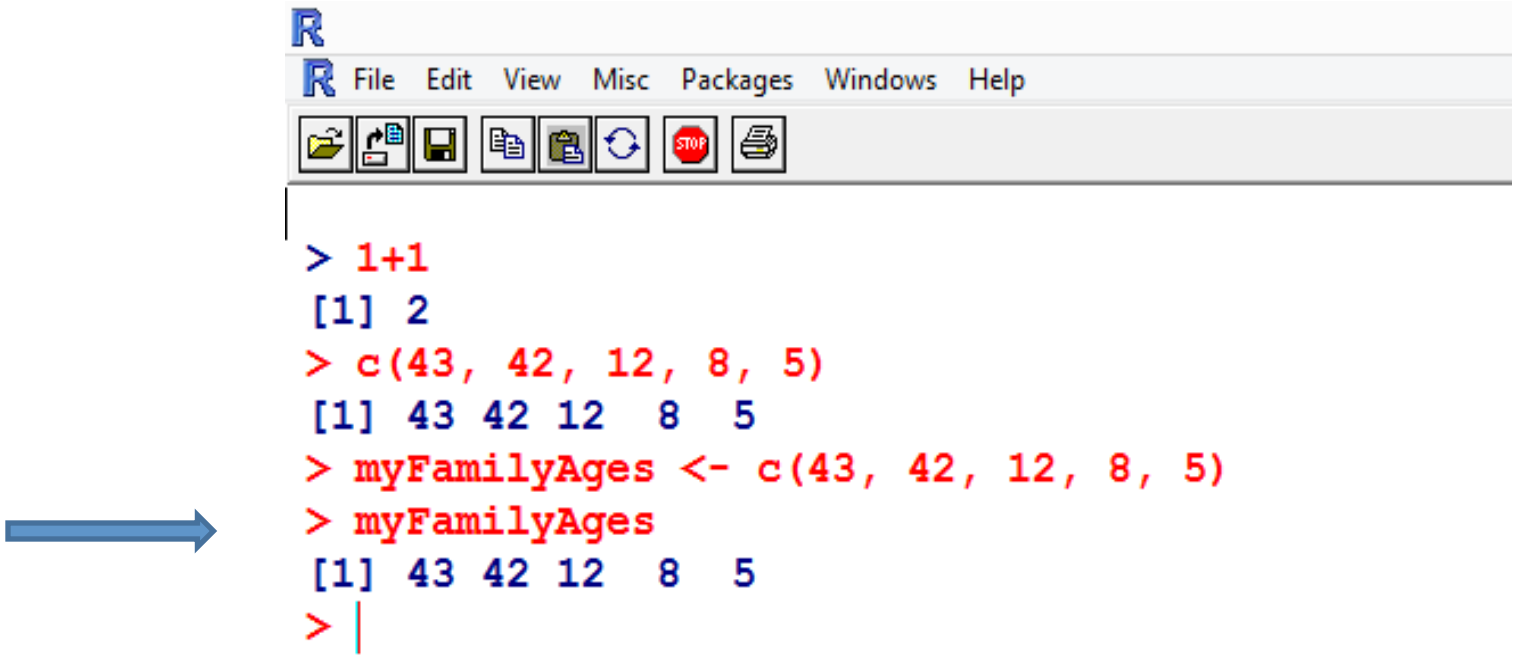
Creating a Vector



Storing a Vector



Storing a Vector

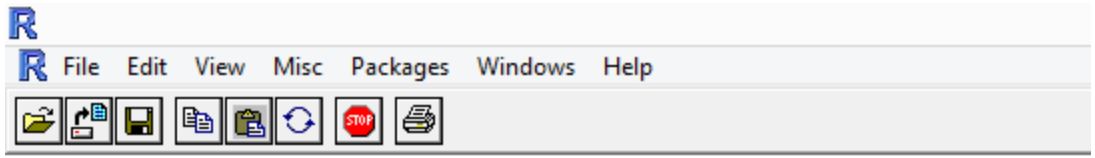


Some Simple Functions on a Vector

sum

mean

range




```
> 1+1
[1] 2
> c(43, 42, 12, 8, 5)
[1] 43 42 12  8  5
> myFamilyAges <- c(43, 42, 12, 8, 5)
> myFamilyAges
[1] 43 42 12  8  5
> sum(myFamilyAges)
[1] 110
> mean(myFamilyAges)
[1] 22
> range(myFamilyAges)
[1]  5 43
>
```

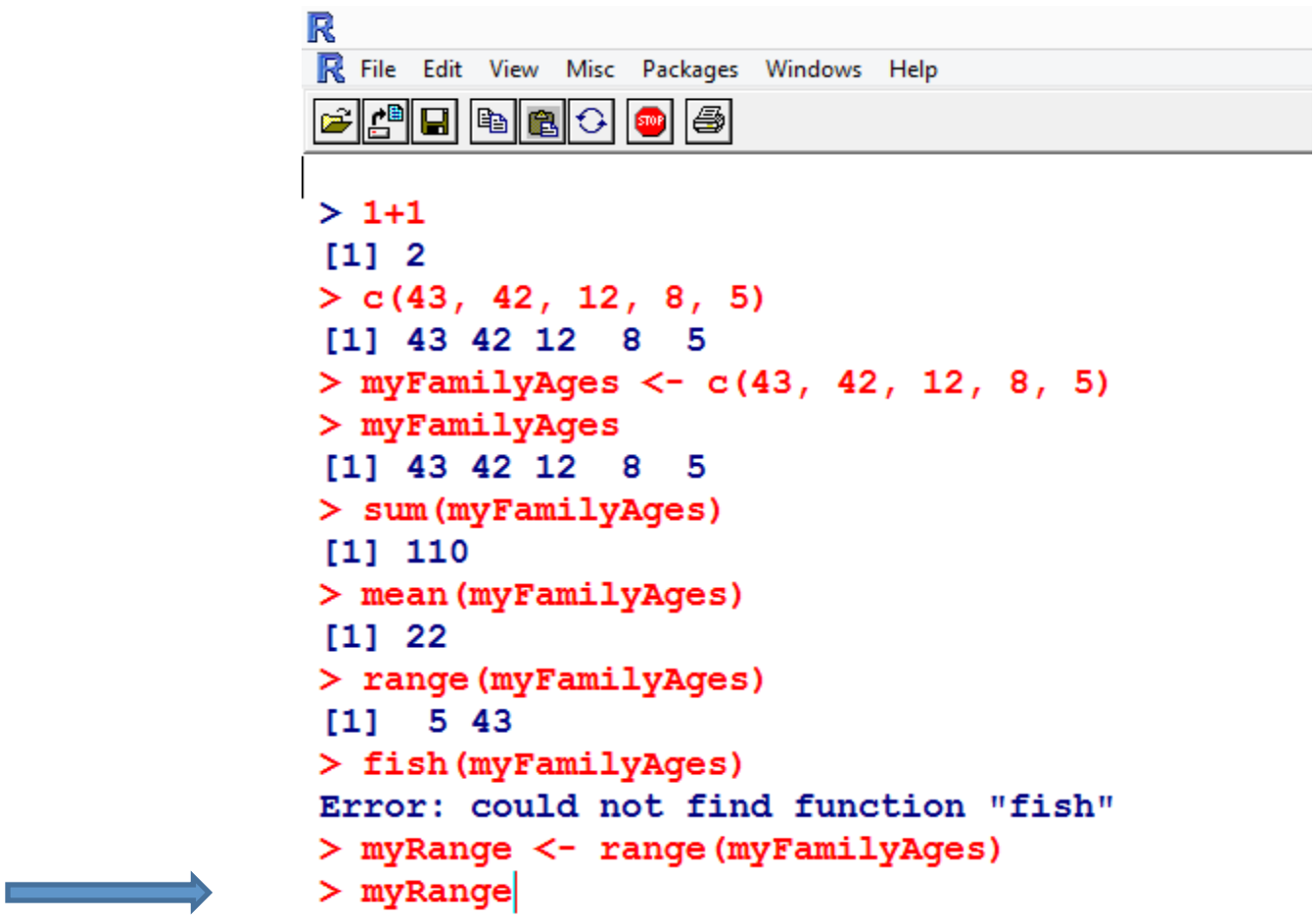
Errors in R

```
> 1+1
[1] 2
> c(43, 42, 12, 8, 5)
[1] 43 42 12  8  5
> myFamilyAges <- c(43, 42, 12, 8, 5)
> myFamilyAges
[1] 43 42 12  8  5
> sum(myFamilyAges)
[1] 110
> mean(myFamilyAges)
[1] 22
> range(myFamilyAges)
[1]  5 43
> fish(myFamilyAges)
Error: could not find function "fish"
>
```

Storing Function Results

```
> 1+1
[1] 2
> c(43, 42, 12, 8, 5)
[1] 43 42 12  8  5
> myFamilyAges <- c(43, 42, 12, 8, 5)
> myFamilyAges
[1] 43 42 12  8  5
> sum(myFamilyAges)
[1] 110
> mean(myFamilyAges)
[1] 22
> range(myFamilyAges)
[1]  5 43
> fish(myFamilyAges)
Error: could not find function "fish"
> myRange <- range(myFamilyAges)
>
```

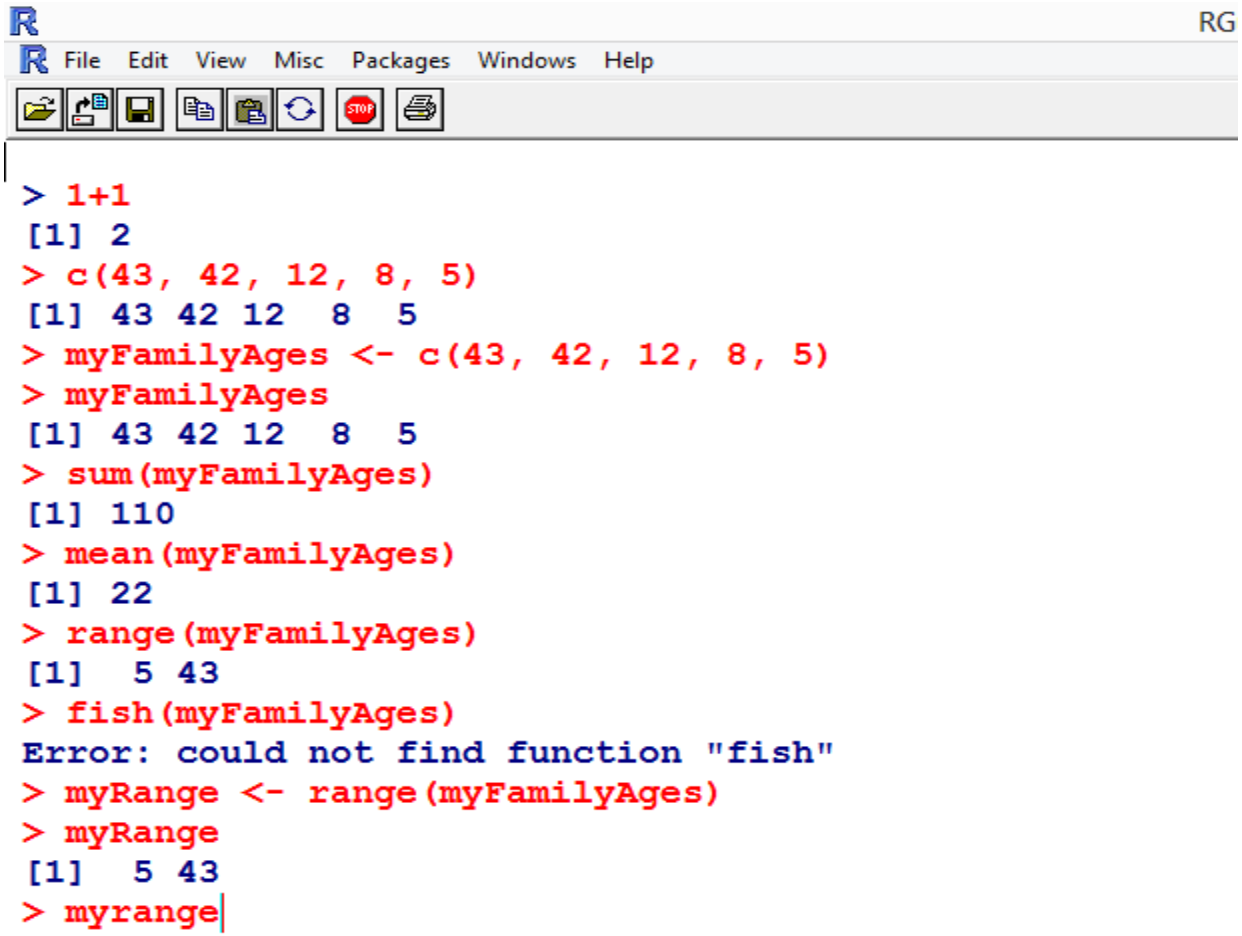
Storing Function Results



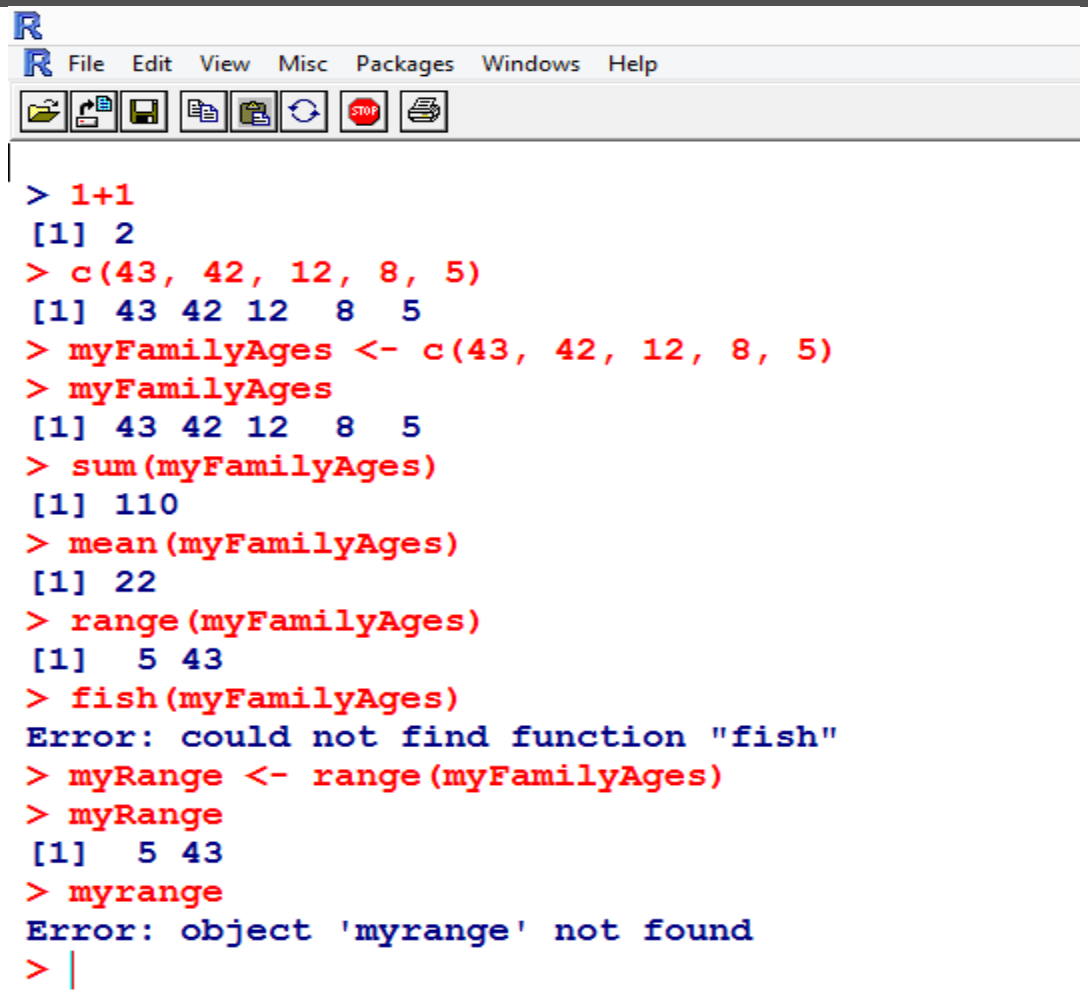
Storing Function Results

```
> 1+1
[1] 2
> c(43, 42, 12, 8, 5)
[1] 43 42 12  8  5
> myFamilyAges <- c(43, 42, 12, 8, 5)
> myFamilyAges
[1] 43 42 12  8  5
> sum(myFamilyAges)
[1] 110
> mean(myFamilyAges)
[1] 22
> range(myFamilyAges)
[1]  5 43
> fish(myFamilyAges)
Error: could not find function "fish"
> myRange <- range(myFamilyAges)
> myRange
[1]  5 43
>
```

Caps are NOT Ignored



Caps are NOT Ignored



Some Takeaways

- The use of the “c()” function.
- The vector is a basic form of data storage.
- Numbers & text can be collected in lists (Vector)
- A vector can be stored in a named location using the arrow “<-” (e.g. myFamilyAges)
- You can get the data object a named location by typing the name.

Some Takeaways

- Vector has a length (number of items in it).
- Vector has a mode (type of data).
- Capitalization matters.
- Legal functions include sum, mean, range, and fish is not a legal function.
- Typically more than one variable or data set is used for analysis.

Data Science

Data Science Fun Meter


SYRACUSE UNIVERSITY
SUOS
CULTORES
SCIENTIA
CORONAT