NA

Neighborhood Analysis

Neighborhood Analysis

Starting with Data

Questions

- What is a data frame?
- How can I get data into and out of R?
- How can I get summary information about my data?
- How can I change the formatting of my data?

Objectives

- Describe what a data frame is
- Create a data frame manually
- Load external data into a data frame
- Summarize data frame contents
- Describe the various variable formats R recognizes

Data frame structure

Most of us are familiar with viewing data in a form like this:

	A	B	C	D	E	F	G	H	I	J
1	geoid	state	Designated	county	Type	dec_score	SE_Flag	Population	medhhincome2014_tract	PovertyRate
2	01001020200	Alabama		Autauga	Low-Income Community	4		2,196	$ 41,107	24.0%
3	01001020300	Alabama		Autauga	Non-LIC Contiguous	6		3,136	$ 51,250	10.7%
4	01001020700	Alabama	1	Autauga	Low-Income Community	9		3,047	$ 45,234	19.0%
5	01001020802	Alabama		Autauga	Non-LIC Contiguous	10		10,743	$ 61,242	15.3%
6	01001021000	Alabama		Autauga	Non-LIC Contiguous	5		2,899	$ 49,567	15.1%
7	01001021100	Alabama		Autauga	Low-Income Community	6		3,247	$ 40,801	19.4%
8	01003010100	Alabama		Baldwin	Non-LIC Contiguous	6		4,013	$ 45,667	14.0%
9	01003010200	Alabama	1	Baldwin	Low-Income Community	9		3,067	$ 33,333	27.2%
10	01003010300	Alabama		Baldwin	Non-LIC Contiguous	10		8,079	$ 47,443	6.8%
11	01003010400	Alabama	1	Baldwin	Non-LIC Contiguous	9		4,578	$ 46,696	14.8%
12	01003010500	Alabama	1	Baldwin	Low-Income Community	8		5,115	$ 45,825	16.8%
13	01003010600	Alabama	1	Baldwin	Low-Income Community	9		3,503	$ 28,219	28.2%
14	01003010904	Alabama		Baldwin	Non-LIC Contiguous	10		6,523	$ 48,521	16.3%
15	01003010906	Alabama		Baldwin	Non-LIC Contiguous	10		5,272	$ 42,120	11.5%
16	01003011000	Alabama		Baldwin	Low-Income Community	10		3,885	$ 34,883	21.8%
17	01003011401	Alabama		Baldwin	Non-LIC Contiguous	10		10,021	$ 44,886	11.9%
18	01003011406	Alabama		Baldwin	Low-Income Community	10		3,810	$ 41,867	19.0%
19	01003011407	Alabama		Baldwin	Low-Income Community	10		4,970	$ 41,840	20.8%
20	01003011501	Alabama	1	Baldwin	Non-LIC Contiguous	9		5,947	$ 48,191	17.9%
21	01003011502	Alabama	1	Baldwin	Low-Income Community	10		11,575	$ 39,563	20.3%
22	01003011601	Alabama		Baldwin	Low-Income Community	10		6,640	$ 39,586	24.3%
23	01005950100	Alabama	1	Barbour	Low-Income Community	6		3,477	$ 38,571	33.2%
24	01005950200	Alabama		Barbour	Low-Income Community	1		4,404	$ 32,742	27.2%
25	01005950300	Alabama		Barbour	Low-Income Community	1		1,657	$ 29,911	36.1%
26	01005950400	Alabama		Barbour	Non-LIC Contiguous	1		3,693	$ 33,241	19.6%
27	01005950500	Alabama		Barbour	Low-Income Community	8		3,438	$ 38,859	19.1%
28	01005950600	Alabama		Barbour	Low-Income Community	4		2,003	$ 27,708	31.0%
29	01005950700	Alabama		Barbour	Low-Income Community	6		1,959	$ 28,409	31.3%
30	01005950800	Alabama		Barbour	Non-LIC Contiguous	5		2,195	$ 40,724	14.2%
31	01005950900	Alabama		Barbour	Low-Income Community	4		3,788	$ 27,027	28.5%
32	01007010001	Alabama		Bibb Cou	Low-Income Community	7		2,783	$ 44,422	9.6%

Data frame structure

Most of us are familiar with viewing data in a form like this:

- What does each row represent?
- What does each column represent?

	A	B	C	D	E	F	G	H	I	J
1	geoid	state	Designated	county	Type	dec_score	SE_Flag	Population	medhhincome2014_tract	PovertyRate
2	01001020200	Alabama		Autauga	Low-Income Community	4		2,196	$ 41,107	24.0%
3	01001020300	Alabama		Autauga	Non-LIC Contiguous	6		3,136	$ 51,250	10.7%
4	01001020700	Alabama	1	Autauga	Low-Income Community	9		3,047	$ 45,234	19.0%
5	01001020802	Alabama		Autauga	Non-LIC Contiguous	10		10,743	$ 61,242	15.3%
6	01001021000	Alabama		Autauga	Non-LIC Contiguous	5		2,899	$ 49,567	15.1%
7	01001021100	Alabama		Autauga	Low-Income Community	6		3,247	$ 40,801	19.4%
8	01003010100	Alabama		Baldwin	Non-LIC Contiguous	6		4,013	$ 45,667	14.0%
9	01003010200	Alabama	1	Baldwin	Low-Income Community	9		3,067	$ 33,333	27.2%
10	01003010300	Alabama		Baldwin	Non-LIC Contiguous	10		8,079	$ 47,443	6.8%
11	01003010400	Alabama	1	Baldwin	Non-LIC Contiguous	9		4,578	$ 46,696	14.8%
12	01003010500	Alabama	1	Baldwin	Low-Income Community	8		5,115	$ 45,825	16.8%
13	01003010600	Alabama	1	Baldwin	Low-Income Community	9		3,503	$ 28,219	28.2%
14	01003010904	Alabama		Baldwin	Non-LIC Contiguous	10		6,523	$ 48,521	16.3%
15	01003010906	Alabama		Baldwin	Non-LIC Contiguous	10		5,272	$ 42,120	11.5%
16	01003011000	Alabama		Baldwin	Low-Income Community	10		3,885	$ 34,883	21.8%
17	01003011401	Alabama		Baldwin	Non-LIC Contiguous	10		10,021	$ 44,886	11.9%
18	01003011406	Alabama		Baldwin	Low-Income Community	10		3,810	$ 41,867	19.0%
19	01003011407	Alabama		Baldwin	Low-Income Community	10		4,970	$ 41,840	20.8%
20	01003011501	Alabama	1	Baldwin	Non-LIC Contiguous	9		5,947	$ 48,191	17.9%
21	01003011502	Alabama	1	Baldwin	Low-Income Community	10		11,575	$ 39,563	20.3%
22	01003011601	Alabama		Baldwin	Low-Income Community	10		6,640	$ 39,586	24.3%
23	01005950100	Alabama	1	Barbour	Low-Income Community	6		3,477	$ 38,571	33.2%
24	01005950200	Alabama		Barbour	Low-Income Community	1		4,404	$ 32,742	27.2%
25	01005950300	Alabama		Barbour	Low-Income Community	1		1,657	$ 29,911	36.1%
26	01005950400	Alabama		Barbour	Non-LIC Contiguous	1		3,693	$ 33,241	19.6%
27	01005950500	Alabama		Barbour	Low-Income Community	8		3,438	$ 38,859	19.1%
28	01005950600	Alabama		Barbour	Low-Income Community	4		2,003	$ 27,708	31.0%
29	01005950700	Alabama		Barbour	Low-Income Community	6		1,959	$ 28,409	31.3%
30	01005950800	Alabama		Barbour	Non-LIC Contiguous	5		2,195	$ 40,724	14.2%
31	01005950900	Alabama		Barbour	Low-Income Community	4		3,788	$ 27,027	28.5%
32	01007010001	Alabama		Bibb Cou	Low-Income Community	7		2,783	$ 44,422	9.6%

Data frame structure

A data frame is R's version of a table. Each column is a vector of the same data type.

	A	B	C	D	E	F	G	H	I	J
1	geoid	state	Designated	county	Type	dec_score	SE_Flag	Population	medhhincome2014_tract	PovertyRate
2	01001020200	Alabama		Autauga	Low-Income Community	4		2,196	$ 41,107	24.0%
3	01001020300	Alabama		Autauga	Non-LIC Contiguous	6		3,136	$ 51,250	10.7%
4	01001020700	Alabama	1	Autauga	Low-Income Community	9		3,047	$ 45,234	19.0%
5	01001020802	Alabama		Autauga	Non-LIC Contiguous	10		10,743	$ 61,242	15.3%
6	01001021000	Alabama		Autauga	Non-LIC Contiguous	5		2,899	$ 49,567	15.1%
7	01001021100	Alabama		Autauga	Low-Income Community	6		3,247	$ 40,801	19.4%
8	01003010100	Alabama		Baldwin	Non-LIC Contiguous	6		4,013	$ 45,667	14.0%
9	01003010200	Alabama	1	Baldwin	Low-Income Community	9		3,067	$ 33,333	27.2%
10	01003010300	Alabama		Baldwin	Non-LIC Contiguous	10		8,079	$ 47,443	6.8%
11	01003010400	Alabama	1	Baldwin	Non-LIC Contiguous	9		4,578	$ 46,696	14.8%
12	01003010500	Alabama	1	Baldwin	Low-Income Community	8		5,115	$ 45,825	16.8%
13	01003010600	Alabama	1	Baldwin	Low-Income Community	9		3,503	$ 28,219	28.2%
14	01003010904	Alabama		Baldwin	Non-LIC Contiguous	10		6,523	$ 48,521	16.3%
15	01003010906	Alabama		Baldwin	Non-LIC Contiguous	10		5,272	$ 42,120	11.5%
16	01003011000	Alabama		Baldwin	Low-Income Community	10		3,885	$ 34,883	21.8%
17	01003011401	Alabama		Baldwin	Non-LIC Contiguous	10		10,021	$ 44,886	11.9%
18	01003011406	Alabama		Baldwin	Low-Income Community	10		3,810	$ 41,867	19.0%
19	01003011407	Alabama		Baldwin	Low-Income Community	10		4,970	$ 41,840	20.8%
20	01003011501	Alabama	1	Baldwin	Non-LIC Contiguous	9		5,947	$ 48,191	17.9%
21	01003011502	Alabama	1	Baldwin	Low-Income Community	10		11,575	$ 39,563	20.3%
22	01003011601	Alabama		Baldwin	Low-Income Community	10		6,640	$ 39,586	24.3%
23	01005950100	Alabama	1	Barbour	Low-Income Community	6		3,477	$ 38,571	33.2%
24	01005950200	Alabama		Barbour	Low-Income Community	1		4,404	$ 32,742	27.2%
25	01005950300	Alabama		Barbour	Low-Income Community	1		1,657	$ 29,911	36.1%
26	01005950400	Alabama		Barbour	Non-LIC Contiguous	1		3,693	$ 33,241	19.6%
27	01005950500	Alabama		Barbour	Low-Income Community	8		3,438	$ 38,859	19.1%
28	01005950600	Alabama		Barbour	Low-Income Community	4		2,003	$ 27,708	31.0%
29	01005950700	Alabama		Barbour	Low-Income Community	6		1,959	$ 28,409	31.3%
30	01005950800	Alabama		Barbour	Non-LIC Contiguous	5		2,195	$ 40,724	14.2%
31	01005950900	Alabama		Barbour	Low-Income Community	4		3,788	$ 27,027	28.5%
32	01007010001	Alabama		Bibb Cou	Low-Income Community	7		2,783	$ 44,422	9.6%

What's a vector?

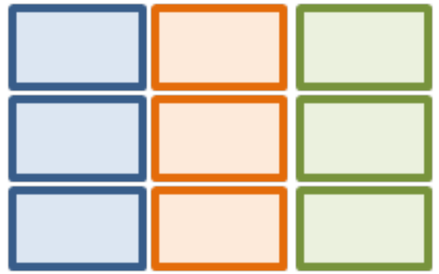
Vector

- 1 column or row of data
- 1 type (numeric or text)

Matrix

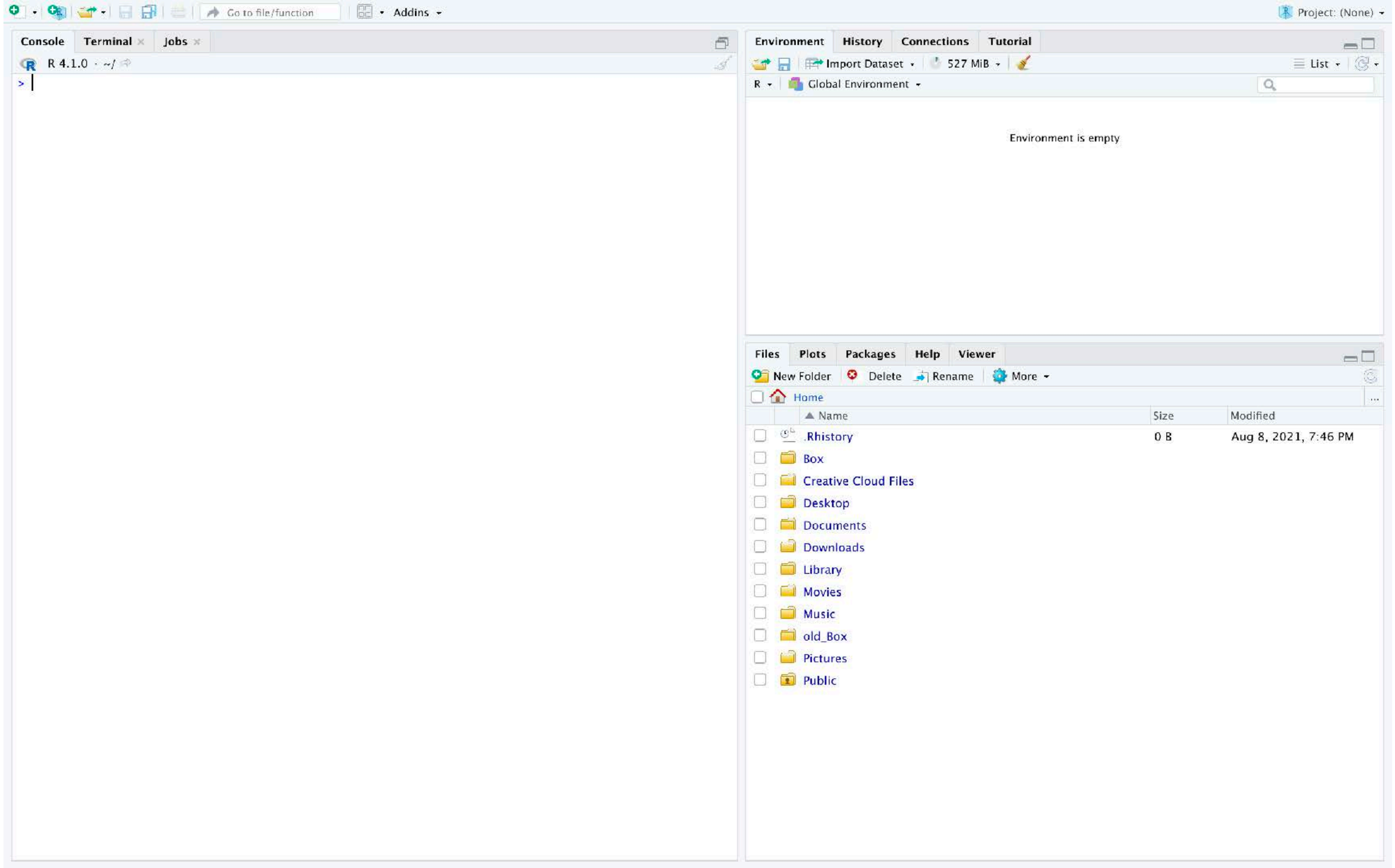
- multiple columns and/or rows of data
- 1 type (numeric or text)

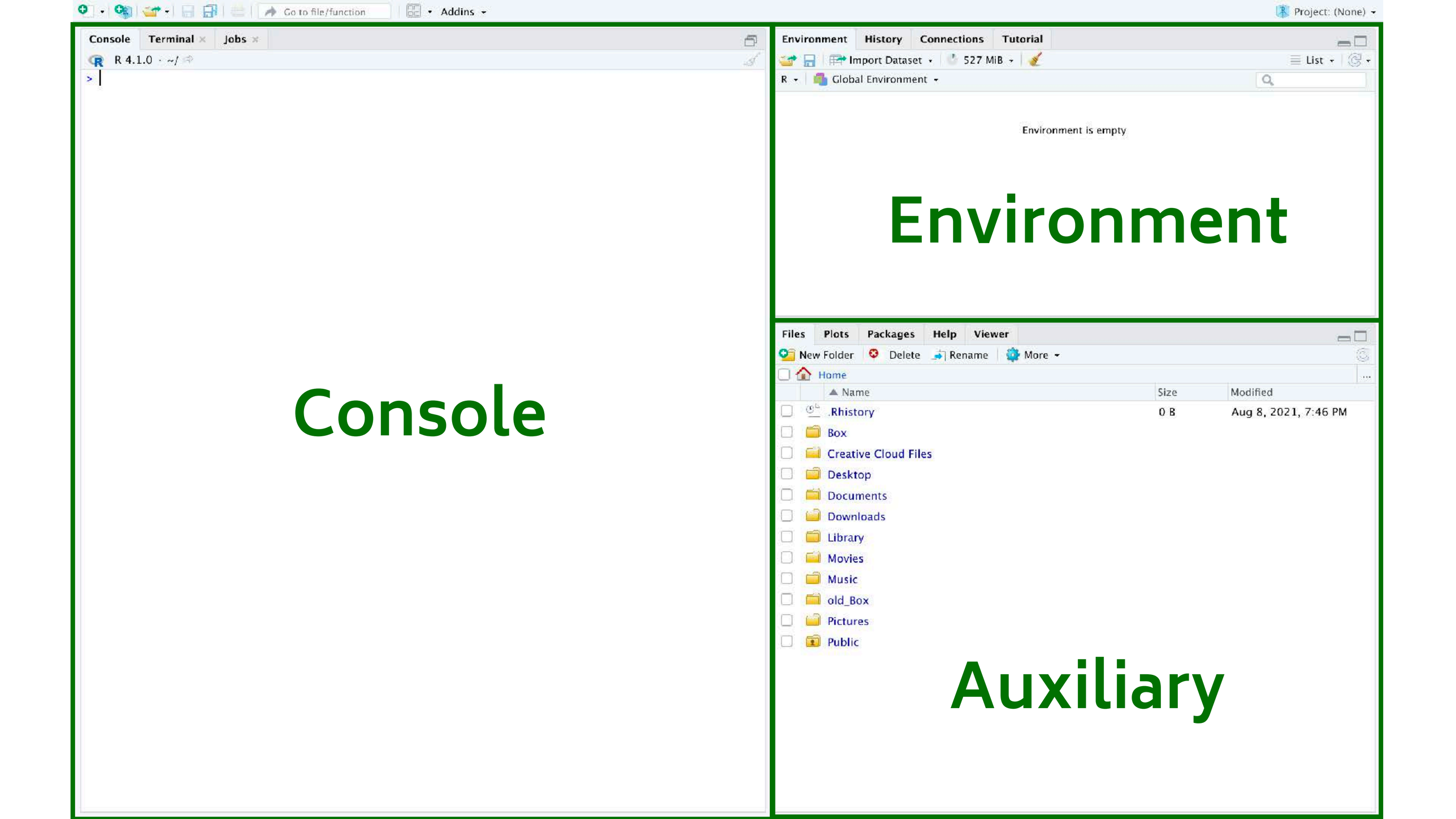
Data Frame

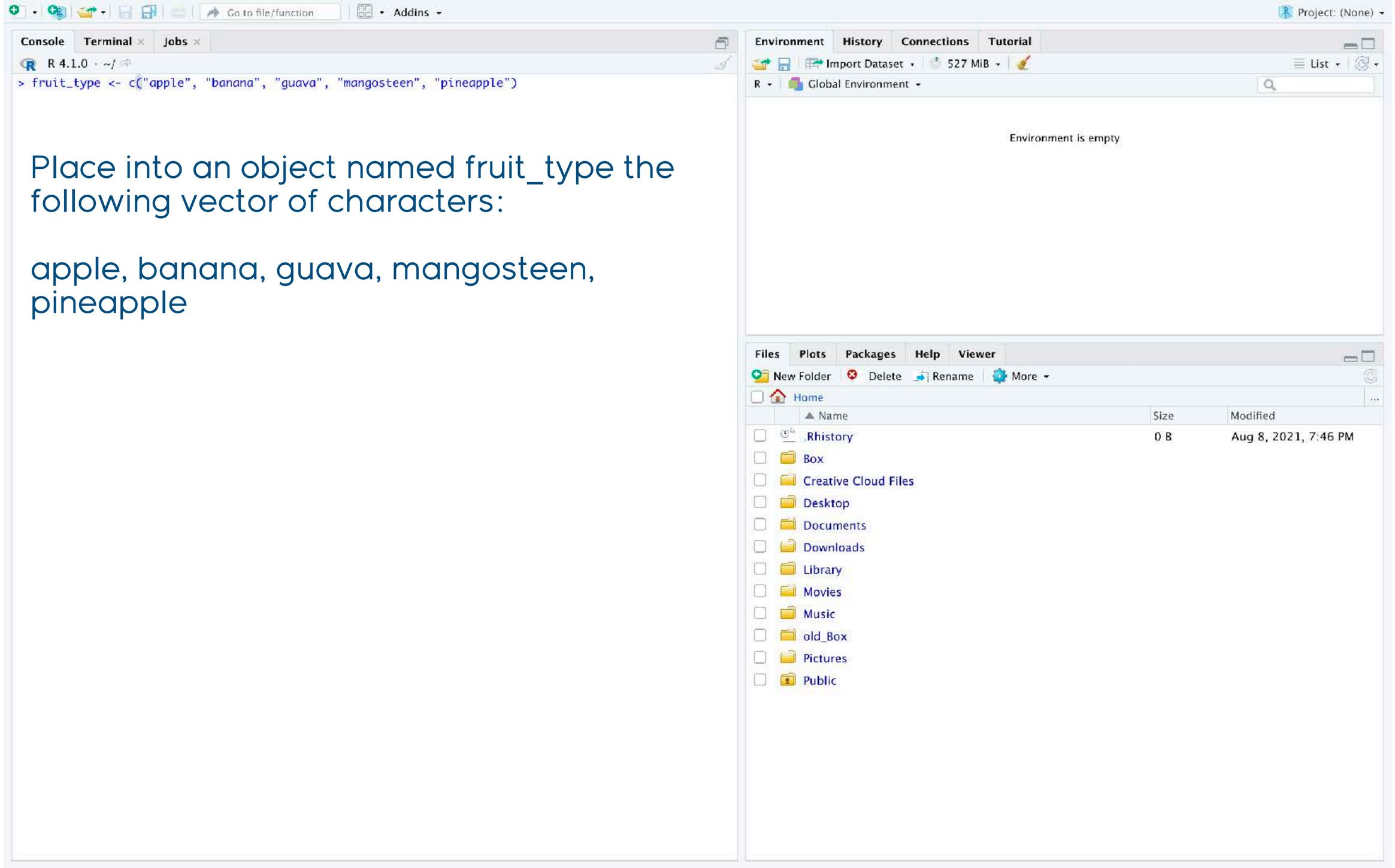


- multiple columns and/or rows of data
- multiple types

- A vector is a collection of observations that have the same format and are describing the same thing.
- A data frame is a collection of vectors.

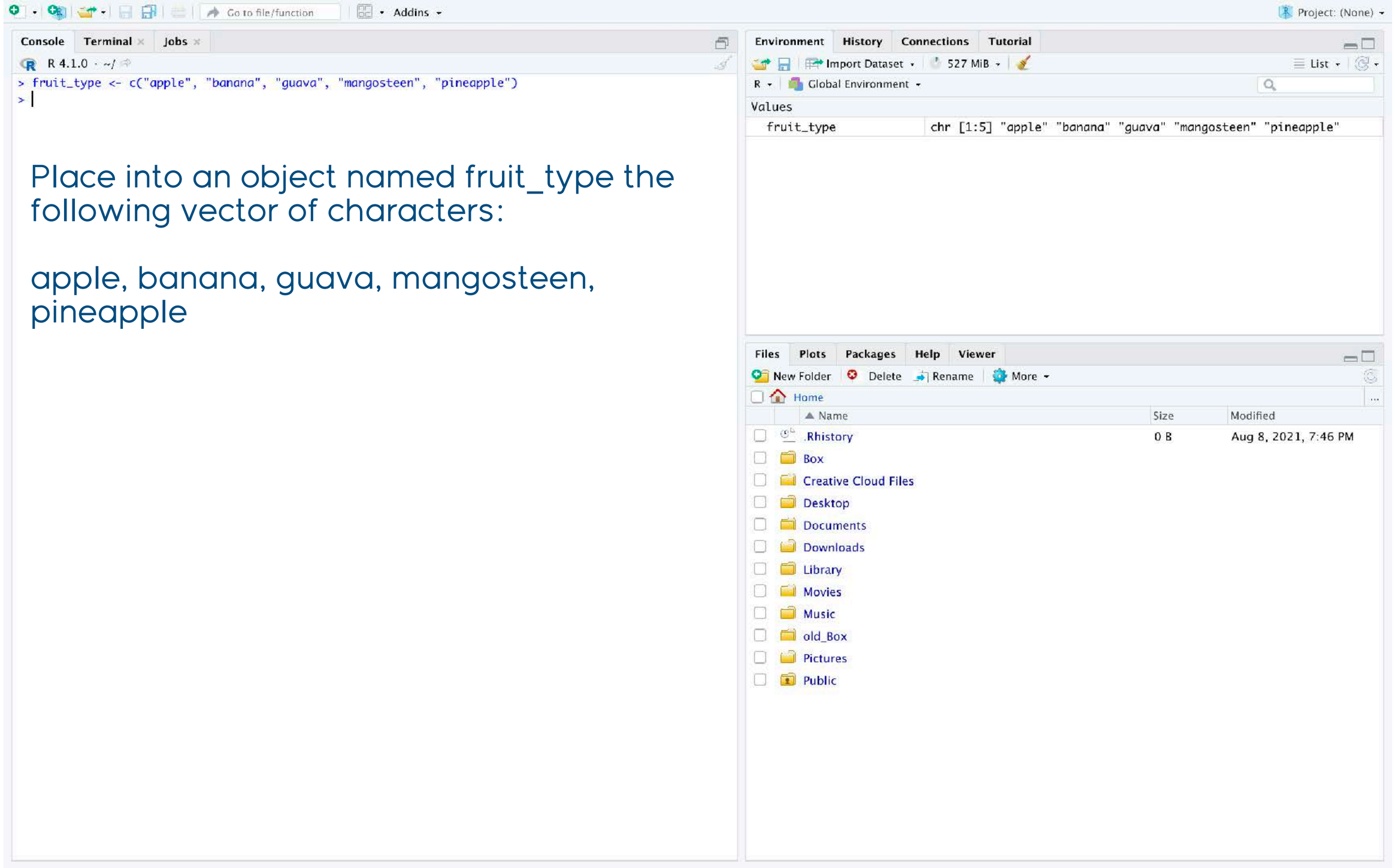

Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
R 4.1.0 · ~/
>
Environment
History
Connections
Tutorial
Import Dataset
527 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Name
Size
Modified
.Rhistory
0 B
Aug 8, 2021, 7:46 PM
Box
Creative Cloud Files
Desktop
Documents
Downloads
Library
Movies
Music
old_Box
Pictures
Public


Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
R 4.1.0 · ~/
>
Console
Environment
History
Connections
Tutorial
Import Dataset
527 MiB
List
R
Global Environment
Environment is empty
Environment
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Name
Size
Modified
.Rhistory
0 B
Aug 8, 2021, 7:46 PM
Box
Creative Cloud Files
Desktop
Documents
Downloads
Library
Movies
Music
old_Box
Pictures
Public
Auxiliary



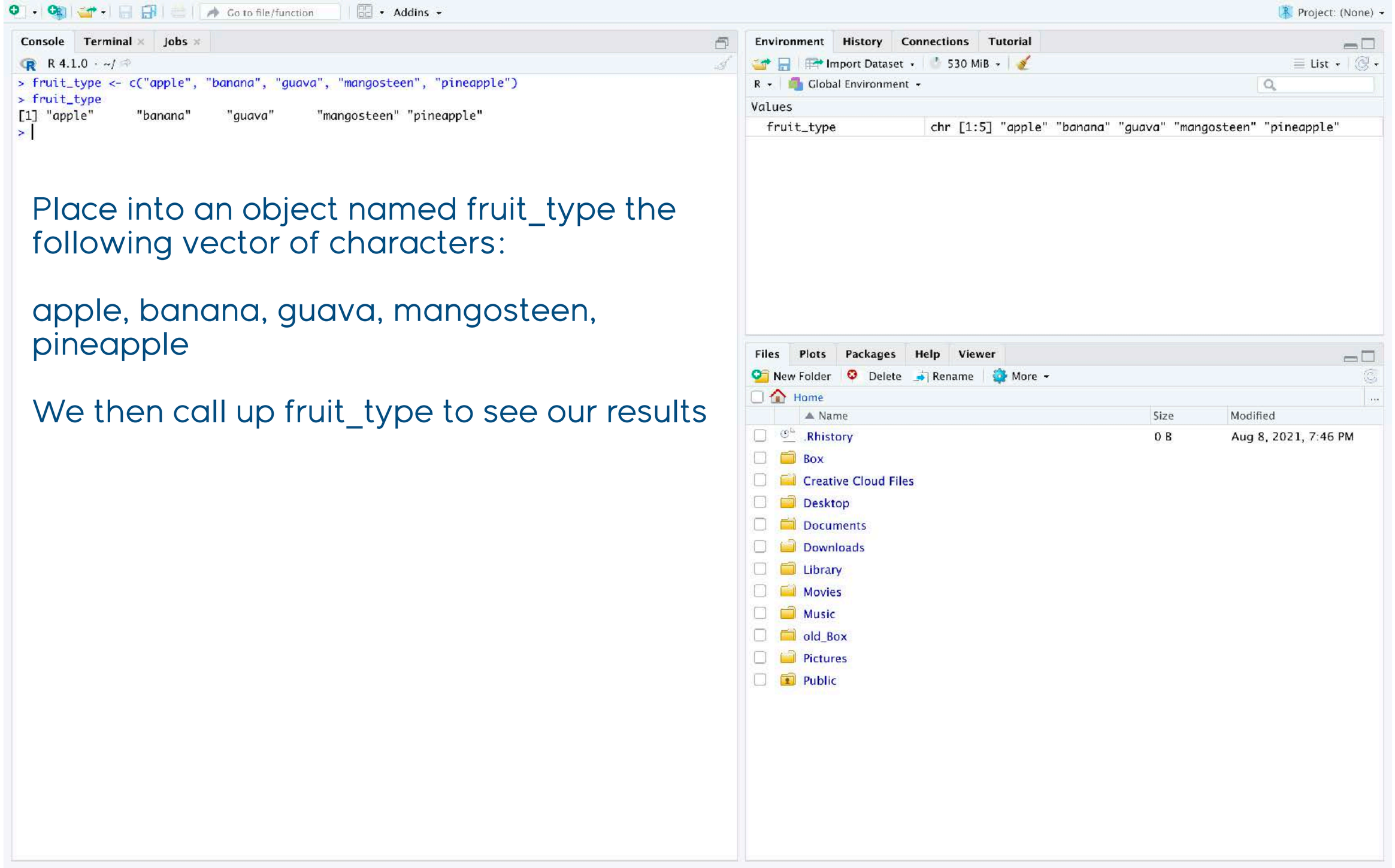
Place into an object named fruit_type the following vector of characters:

apple, banana, guava, mangosteen, pineapple



Place into an object named fruit_type the following vector of characters:

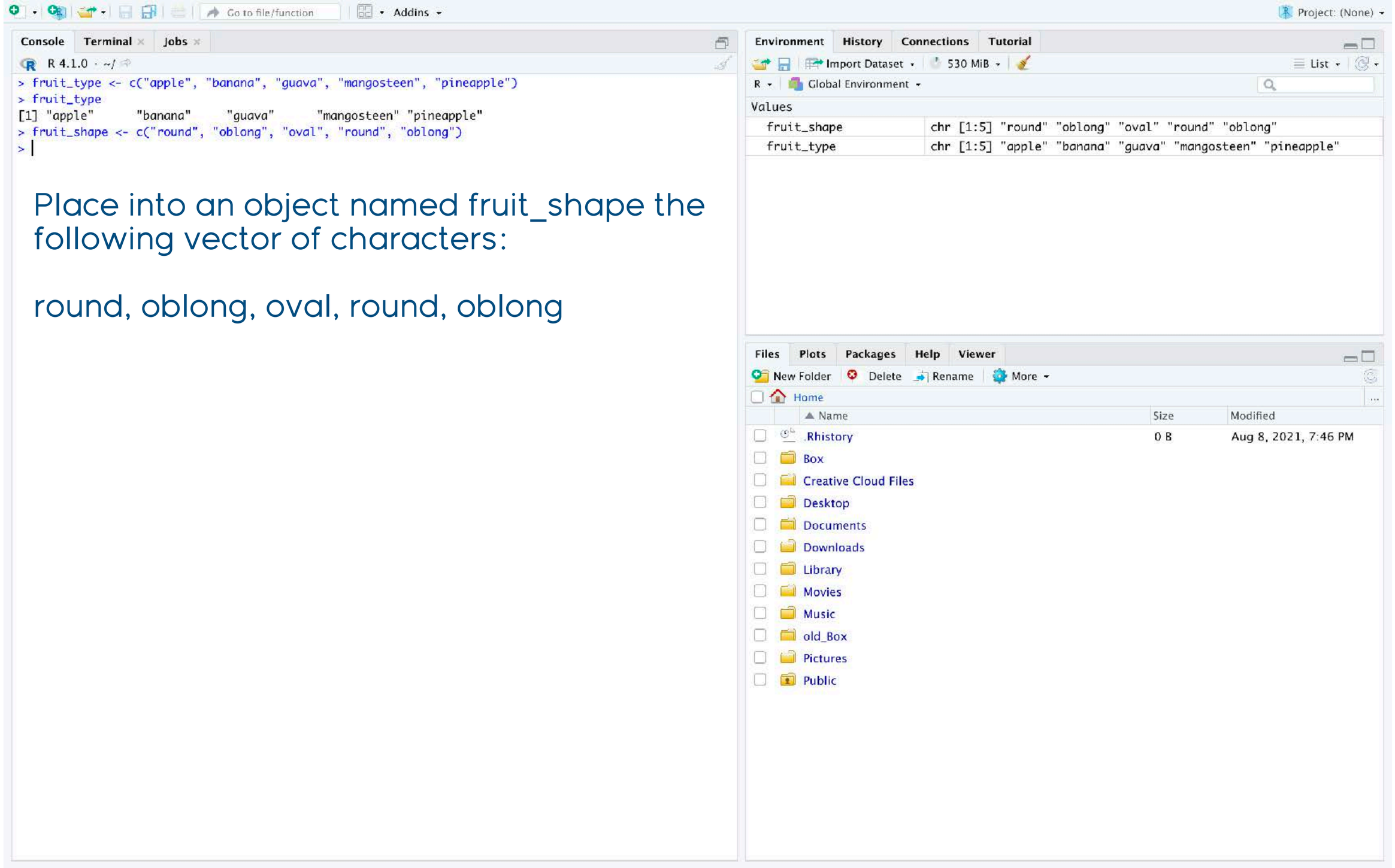
apple, banana, guava, mangosteen, pineapple



Place into an object named fruit_type the following vector of characters:

apple, banana, guava, mangosteen, pineapple

We then call up fruit_type to see our results



Place into an object named fruit_shape the following vector of characters:

round, oblong, oval, round, oblong

```
> fruit_type <- c("apple", "banana", "guava", "mangosteen", "pineapple")
> fruit_type
[1] "apple"      "banana"     "guava"      "mangosteen" "pineapple"
> fruit_shape <- c("round", "oblong", "oval", "round", "oblong")
> fruits<-data.frame(fruit_type, fruit_shape)
```

Place into an object named fruits a data frame consisting of the fruit_type and fruit_shape vectors

Values	
fruit_shape	chr [1:5] "round" "oblong" "oval" "round" "oblong"
fruit_type	chr [1:5] "apple" "banana" "guava" "mangosteen" "pineapple"

Name	Size	Modified
.Rhistory	0 B	Aug 8, 2021, 7:46 PM
Box		
Creative Cloud Files		
Desktop		
Documents		
Downloads		
Library		
Movies		
Music		
old_Box		
Pictures		
Public		

```
> fruit_type <- c("apple", "banana", "guava", "mangosteen", "pineapple")
> fruit_type
[1] "apple"      "banana"     "guava"      "mangosteen" "pineapple"
> fruit_shape <- c("round", "oblong", "oval", "round", "oblong")
> fruits<-data.frame(fruit_type, fruit_shape)
> fruits
  fruit_type fruit_shape
1      apple       round
2     banana      oblong
3      guava        oval
4 mangosteen       round
5  pineapple      oblong
>
```

What's happening here?

```
> fruit_type <- c("apple", "banana", "guava", "mangosteen", "pineapple")
> fruit_type
[1] "apple"      "banana"     "guava"      "mangosteen" "pineapple"
> fruit_shape <- c("round", "oblong", "oval", "round", "oblong")
> fruits<-data.frame(fruit_type, fruit_shape)
> fruits
  fruit_type fruit_shape
1      apple       round
2     banana      oblong
3      guava        oval
4 mangosteen       round
5  pineapple      oblong
> weight<-c(85, 120, 55, 82, 905)
> fruits<-data.frame(fruit_type, fruit_shape, weight)
> fruits
  fruit_type fruit_shape weight
1      apple       round     85
2     banana      oblong    120
3      guava        oval     55
4 mangosteen       round     82
5  pineapple      oblong    905
>
```

Place into an object named weight the following vector of numbers:

85, 120, 55, 82, 905

Place into an object named fruits a data frame consisting of the fruit_type, fruit_shape, and weight vectors.

Note that this overwrites the existing fruits object

```
> fruit_type <- c("apple", "banana", "guava", "mangosteen", "pineapple")
> fruit_type
[1] "apple"      "banana"     "guava"      "mangosteen" "pineapple"
> fruit_shape <- c("round", "oblong", "oval", "round", "oblong")
> fruits<-data.frame(fruit_type, fruit_shape)
> fruits
  fruit_type fruit_shape
1      apple       round
2     banana      oblong
3      guava        oval
4 mangosteen       round
5  pineapple      oblong
> weight<-c(85, 120, 55, 82, 905)
> fruits<-data.frame(fruit_type, fruit_shape, weight)
> fruits
  fruit_type fruit_shape weight
1      apple       round     85
2     banana      oblong    120
3      guava        oval     55
4 mangosteen       round     82
5  pineapple      oblong    905
> str(fruits)
'data.frame':   5 obs. of  3 variables:
 $ fruit_type : chr  "apple" "banana" "guava" "mangosteen" ...
 $ fruit_shape: chr  "round" "oblong" "oval" "round" ...
 $ weight     : num  85 120 55 82 905
>
```

Tell me about the structure of the object fruits.

R tells us about the number of observations (rows) and number of columns. It also tells us about the type of data in each column and shows us the first few values.

Reading in data

- Most of the time, we will not create data frames manually, but will import existing data which we can then work with.
- There are different functions that read external data with particular formats or structures.
- read_csv() for instance, is designed to read comma separated value files

```
geoid,state,county,Designated,Type,dec_score,Population,PovertyRate,medhhincome2014_tract,medvalue,medrent,pctown,vacancyrate,pctwhitealone,pctblackalone,pctHispan
ic,pctAAPIalone,HSorlower,BAorhigher,Metro,Micro,NoCBSAType
01001020200,Alabama,Autauga County,NA,Low-Income Community,4,2196,24,41107,95300,743,62.8,5.8,43.9,51.9,1.3,1.1,58.1,16.2,1,NA,NA
01001020300,Alabama,Autauga County,NA,Non-LIC Contiguous,6,3136,10.7,51250,113800,817,70.3,14,67.1,20.5,7.3,1.1,46.4,21.9,1,NA,NA
01001020700,Alabama,Autauga County,1,Low-Income Community,9,3047,19,45234,93500,695,71.1,6.2,83.3,9.2,3.4,0,54.4,11.3,1,NA,NA
01001020802,Alabama,Autauga County,NA,Non-LIC Contiguous,10,10743,15.3,61242,160400,1018,82.3,6.1,81.4,15.7,1.4,1,45,22.9,1,NA,NA
01001021000,Alabama,Autauga County,NA,Non-LIC Contiguous,5,2899,15.1,49567,102900,546,83,21.8,72.6,24.6,0.4,0,62.1,13.6,1,NA,NA
01001021100,Alabama,Autauga County,NA,Low-Income Community,6,3247,19.4,40801,71000,630,81.9,16.9,44,55.4,0.5,0,61.3,12.1,1,NA,NA
01003010100,Alabama,Baldwin County,NA,Non-LIC Contiguous,6,4013,14,45667,86300,685,91.4,35.1,79.2,16.1,0.2,0,63.8,14.5,1,NA,NA
01003010200,Alabama,Baldwin County,1,Low-Income Community,9,3067,27.2,33333,136000,640,77.1,10.5,83.9,12,0.9,0,60.7,7.9,1,NA,NA
01003010300,Alabama,Baldwin County,NA,Non-LIC Contiguous,10,8079,6.8,47443,185300,850,81.8,14.7,80.8,15.1,2.3,0,48.3,21.4,1,NA,NA
01003010400,Alabama,Baldwin County,1,Non-LIC Contiguous,9,4578,14.8,46696,135600,808,82.6,19.8,88.9,4.2,5.8,0.5,53.2,16.3,1,NA,NA
01003010500,Alabama,Baldwin County,1,Low-Income Community,8,5115,16.8,45825,129400,636,66.2,14.4,83.4,12.8,1.2,0.6,44.4,20.3,1,NA,NA
01003010600,Alabama,Baldwin County,1,Low-Income Community,9,3503,28.2,28219,88900,552,55.7,21.1,30.2,69.3,0.5,0,60.1,9.7,1,NA,NA
```

Your Lab

- Review the basic structure of vectors and data frames
- Importing external data
- Describing the structure of data
- Indexing or referencing specific data components

Questions