R Tutorial for Lab 1

Author: Leonore Findsen, Min Ren

0. Downloading Data

You will have to download all of the data used from the internet before R can access the data.

If the file accessed via a link, then right click on the file name and save it to a directory on your hard drive. It can be difficult to find the files, so if you are unfamiliar with the PC, I suggest that you download it directly to the W drive (main campus drive).

R can not read a zipped file. Therefore, if any file is zipped, remember to extract it before you try to access it in R.

Please remember where your files are stored so that you can access them for the labs.

1. General Information for R

a) R uses functions to perform operations. To run a function with two inputs called funcname, we type funcname(input1, input2).

b) There is online help that is available. All you need to do is type "? *command*" at the '>'/command prompt and the help window in R studio will contain the appropriate help file. For example, "?setwd" at the '>' prompt will open the available help documentation on utilizing the "setwd()" function. I also find that google is very effective if you are confused about a command.

c) The '#' symbol/sign indicates a comment in R, i.e. everything to the right of the '#' sign will NOT be run as code.

d) Capitalization is important.

e) **RStudio:**

R Tutorial for Lab 1

Author: Leonore Findsen, Min Ren

Once you open RStudio, you will see a window similar to what is provided below:

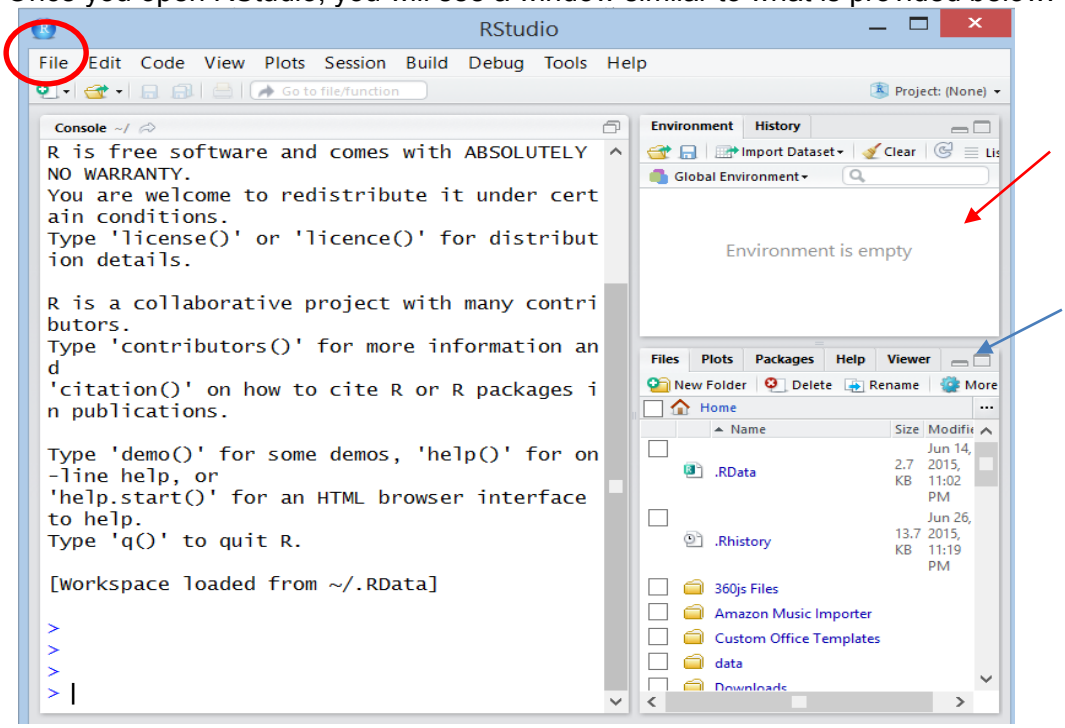
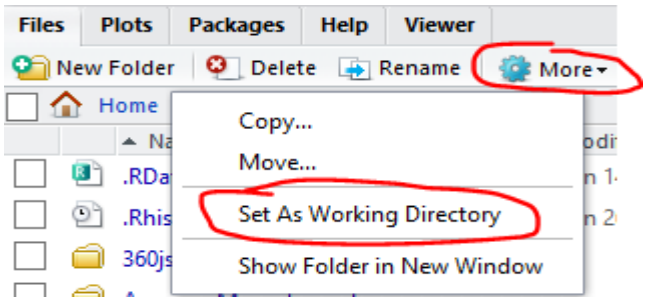


Fig. 1

Set working Directory

I strongly suggest you set your default directory to where your data files are located. To accomplish this, navigate in the "Files" window (the lower right hand panel indicated by the blue arrow in Fig. 1) to the directory that you want. Then, Files → More → Set as Working Directory.



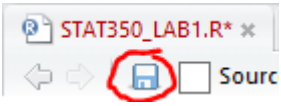
Create R Script file

For your convenience, I recommend that you create a R script for each lab. To accomplish go to the red circled File (Fig. 1) → New File → R Script, then an “Untitled1” file will be created.

R Tutorial for Lab 1

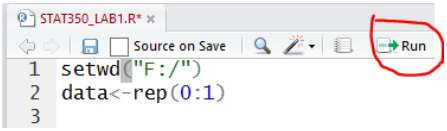
Author: Leonore Findsen, Min Ren

You can use the small “save” button to save and change the name of “Untitled1” File.



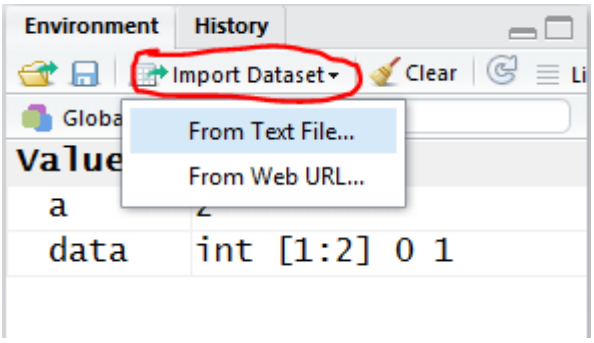
Execution:

Move your cursor to the beginning of the first line of code that you wish to run. Then click the Run button circled below. The cursor will automatically move down one line. If you want to run multiple lines, you may repeat this process. Alternatively, you may select all of the lines that you wish to run and then click on the Run button.

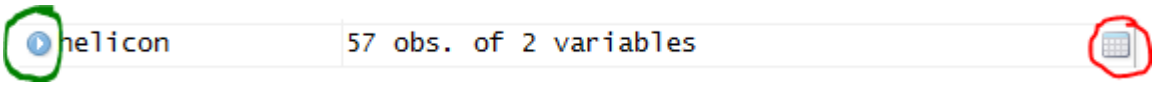


Import data

In the right hand side of RStudio (red arrow in Fig. 1), you will find the Environment tab. Click import Dataset and this wizard will help you intuitively import the dataset and automatically display the data.



You can use “View” command to check your data manually. Note this is a capital 'V'. For example, “View(data)” This can also be accomplished by clicking on the 'View icon' which is circled in red in the diagram below. Please note all the variables will be displayed in the “Environment” window when you click the icon circled in green below.

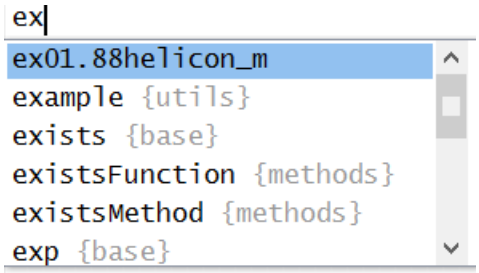


Autocomplete

In the RStudio, you can enjoy the “autocomplete” feature by hitting “Tab”. The prompt window will be like as below:

R Tutorial for Lab 1

Author: Leonore Findsen, Min Ren



2. Importing Data sets, cleaning, manipulating, and printing them.

Hummingbirds and flowers. (Data Set: ex01-88helicon_m.txt) Different varieties of the tropical flower *Heliconia* are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the form of the hummingbirds' beaks have evolved to match each other. Here are data on the lengths in millimeters of three varieties of these flowers on the island of Dominica:

H. bihai							
47.12	46.75	46.81	47.12	46.67	47.43	46.44	46.64
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36
***H. caribaea* red**							
41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07
38.10	37.97	38.79	38.23	38.87	37.78	38.01	
***H. caribaea* yellow**							
36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.1
35.17	36.82	36.66	35.68	36.03	34.57	34.63	

When you import your data set, the name will be the name of the file. I strongly recommend that you change that name to something easier to type.

```
helicon<-ex01.88helicon_m
```

Please name your table in context. In this case, we are interested in *Heliconia* flowers, so I named the table helicon. Since capitalization is important, I often only use small letters. If you want to put a space in the name, use an underscore, _.

```
helicon
```

This line will print out the entire data set. DO NOT DO THIS FOR LARGE DATA SETS. This is equivalent to using the View command which was mentioned earlier. Remember to always look at your data after you read it to be sure that there are no problems in the data set. Unless explicitly stated, never provide your data in your lab reports. In very large data sets where you don't want to look at the whole data set, there are two commands called "head(tablename") and" "tail(tablename)" which will print out the top and bottom rows of the data set.

R Tutorial for Lab 1

Author: Leonore Findsen, Min Ren

	Variety	Length
1	bihai	47.12
2	bihai	46.75
3	bihai	46.81
4	bihai	47.12
5	bihai	46.67
6	bihai	47.43
7	bihai	46.44
8	bihai	46.64

I have highlighted the first five data points from the results from the View command.

If you want to copy tables (or parts of tables) or graphs from the R output, I suggest that you use the Snipping Tool. You can also use that tool to highlight your answer. This is the procedure that I used in the above table. If you are just copying information from the console the Snipping Tool is not required.

(OPTIONAL) Assuming that helicon is in your current R session, run the following commands below and observe their outputs. Do not worry if you do not understand the output from each line as we are only currently attempting to build familiarity to R at this point in time. I would **strongly recommend** that you look at the 'head' command below because it will be useful in Lab 1.

```
head(helicon)
tail(helicon)
helicon[20:30, ]
names(helicon)
dim(helicon)
attach(helicon)
length(helicon[ , 2])
tapply(helicon[ , 2], helicon[ , 1], summary)
sd(helicon[ , 2])
sort(tapply(helicon[ , 2], helicon[ , 1], sd))
subset(helicon, subset = helicon[ , 1] != "red")
class(helicon)
class(helicon[1, ])
class(helicon[ , 2])
class(helicon[ , 1])
levels(helicon[ , 1])
sum(helicon[ , 2])
summary(helicon[ , 2])
```

Cleaning and saving datasets

Since there is often missing values in datasets, it is important to be able to remove that data before we start the analysis. The following command will remove the specific data that are not complete.

```
helicon_cleaned <-helicon[complete.cases(helicon),]
View(helicon)
```

Remember, the 'View' command has a capital V.

R Tutorial for Lab 1

Author: Leonore Findsen, Min Ren

Once you have cleaned the data set, you will want to save it so that you don't have to clean the data each time you use the data set. The procedure to save the new data set is as follows:

```
write.table(helicon_cleaned,file="helicon_cleaned.txt",quote=F,
    row.names=F, sep="\t")
```

Manipulating data

For readability, you might want to change the shortened name or abbreviation to the full version. This is done by the following commands:

```
helicon_new <- helicon_cleaned
helicon_new$Variety <- as.character(helicon_new$Variety)
helicon_new$Variety[helicon_new$Variety=="red"]<-"Caribaea_Red"
helicon_new$Variety[helicon_new$Variety=="yellow"]<-"Caribaea_yellow"
View(helicon_new)
```

Some of the output is shown below:

36	Caribaea_Red	38.23
37	Caribaea_Red	38.87
38	Caribaea_Red	37.78
39	Caribaea_Red	38.01
40	Caribaea_yellow	36.78
41	Caribaea_yellow	37.02
42	Caribaea_yellow	36.52
43	Caribaea_yellow	36.11

In addition, often you want to create a new variable based on mathematical operations from old variable(s). You can use the sample code below to convert the lengths of the beaks from millimeters to inches. The conversion factor is 1/25.4.

```
helicon_new$length_inches = helicon_new$Length/25.4
head(helicon_new)
```

```
  Variety Length length_inches
1   bihai  47.12      1.855118
2   bihai  46.75      1.840551
3   bihai  46.81      1.842913
4   bihai  47.12      1.855118
5   bihai  46.67      1.837402
6   bihai  47.43      1.867323
```

You will see a new column in the data set called length_inches:

I strongly recommend that if you are making changes to a data set that you change the name of the table. When performing calculations, keep in mind which table that you are using.