P8477 EPI MODELING FOR INFECTIOUS DISEASES

Lab 1: R101

Housekeeping issues

- Add/drop class
- Waitlist

Learning Objectives

- R intro
  - Setting it up
  - Getting started
- Variables/Data types in R
- Basic math operations in R
- Basic functions in R
- Plotting (data visualization)
- Input and Output in R

Why programing?

- For ID modeler, can’t write code is like can’t drive

http://www.funnyfidos.com/just-out-for-a-sunday-drive/

R 101

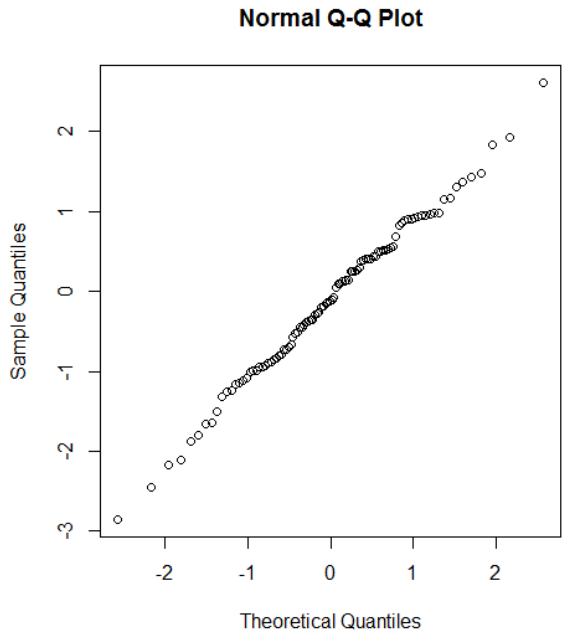
- What is R?
  - https://www.youtube.com/watch?v=ZCQHm63xc4s
- Why R?
  - Statistical programing: powerful analytical tool
    - Not just for this class
  - Open source:
    - Many developers & functionalities (packages)
    - And you don't have to pay
  - Easy to use (cp. other languages)

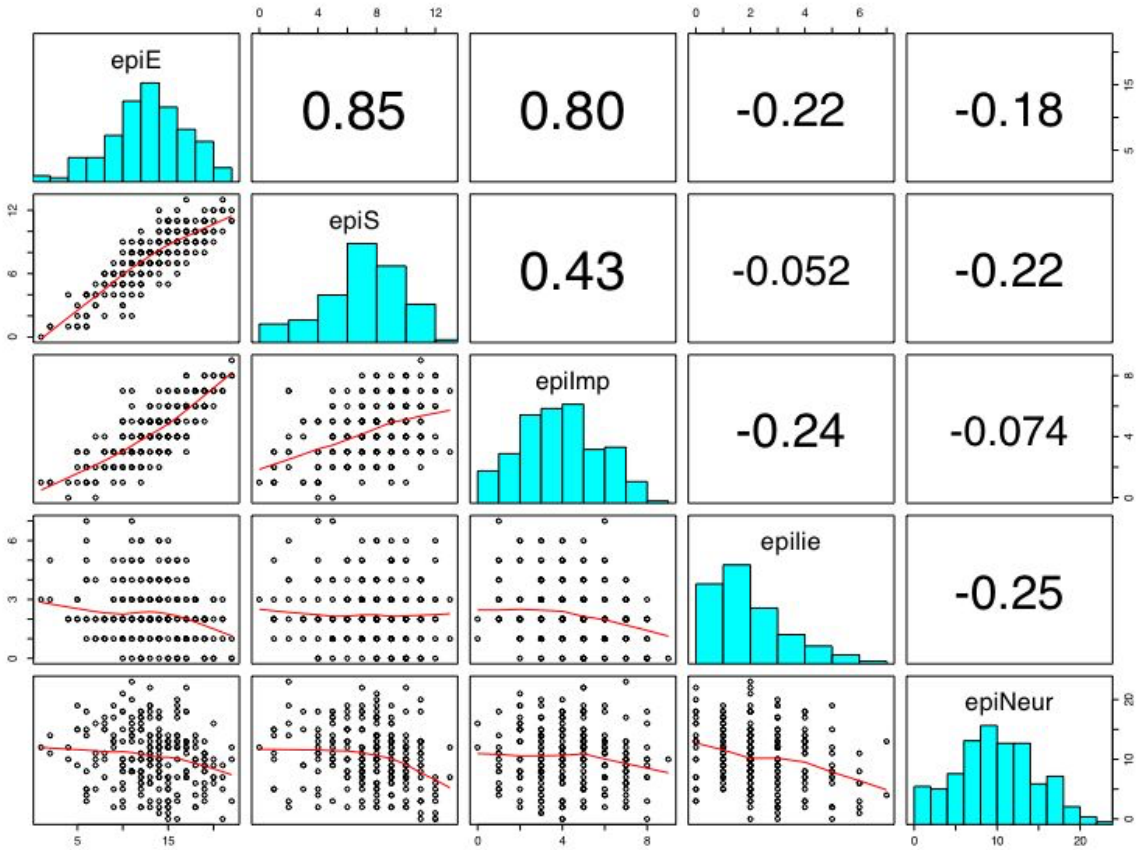
What can you do with R?

- Data analysis
- Data visualization
- Simulation (that's we will do with it mostly)
- Mapping
- …

What can you do with R?

- Data analysis
  - Distributions
  - Hypothesis tests
  - Regression models
  - …

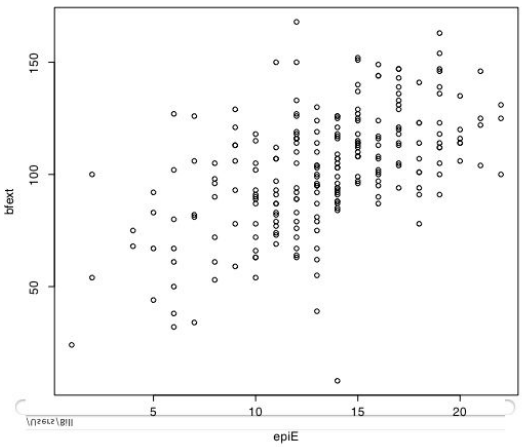


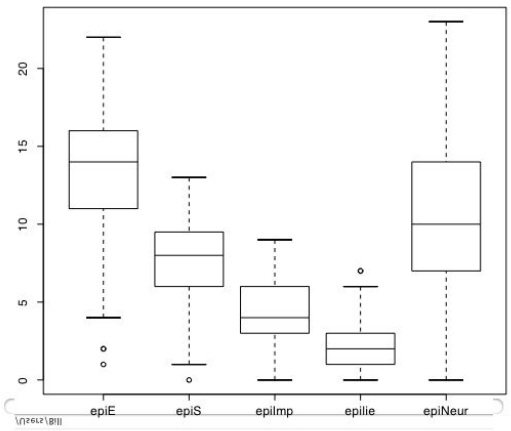


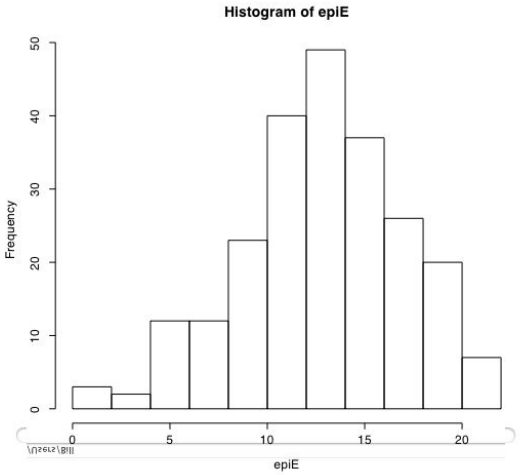
http://personality-project.org/r/r.short.html

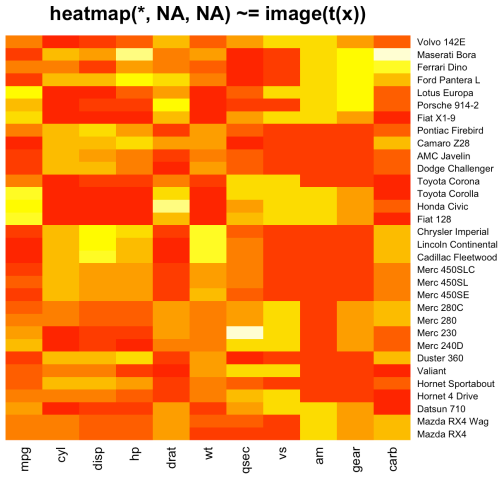
What can you do with R?

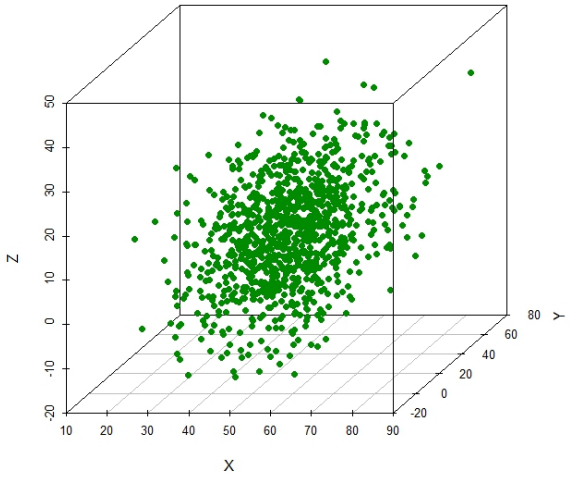
- Data visualization

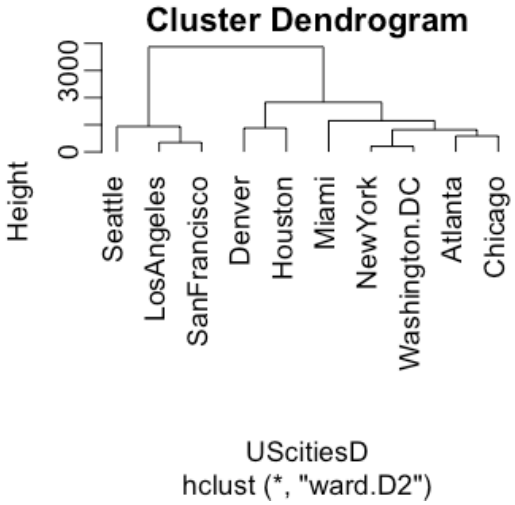












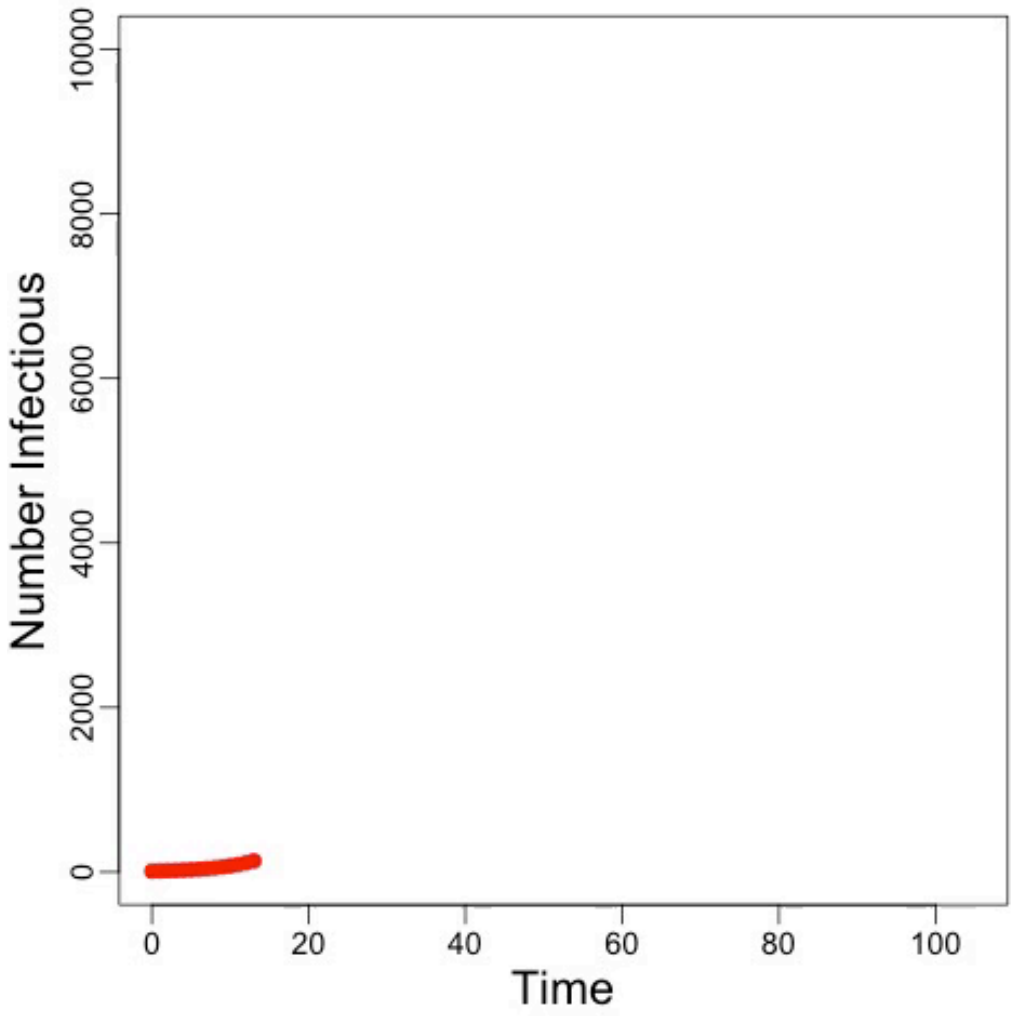
What can you do with R?

- Simulation (that's we will do with it mostly)

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \quad (1)$$

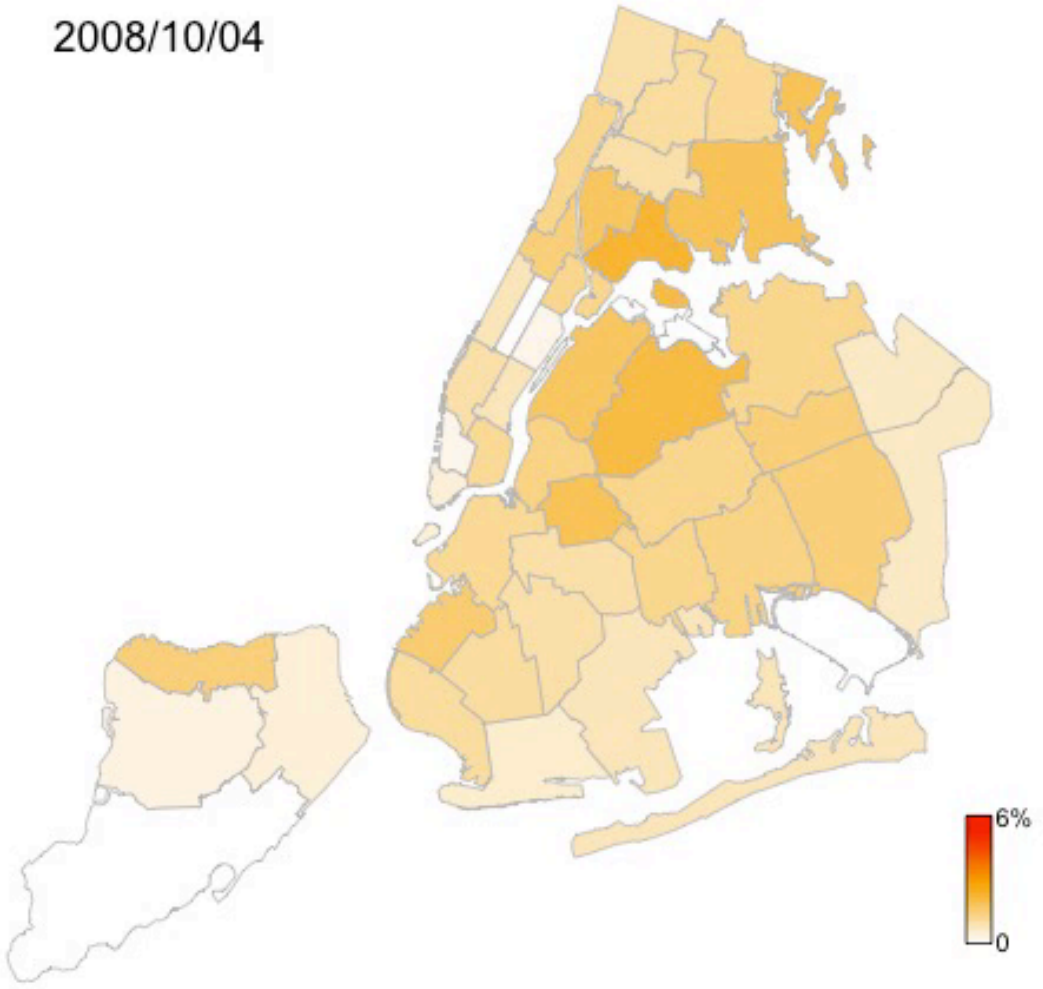
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$



What can you do with R?

- Mapping



Let's get started…

- Installation
  - R: https://www.r-project.org

[Home]

Download

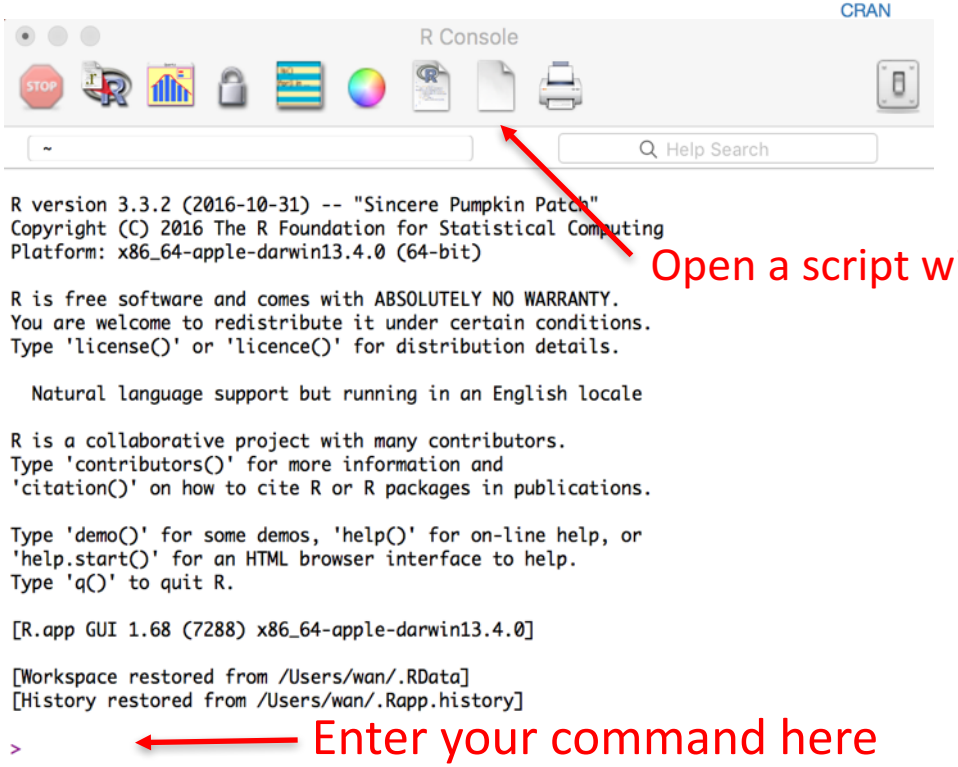
CRAN

The R Project for Statistical Computing

Getting Started

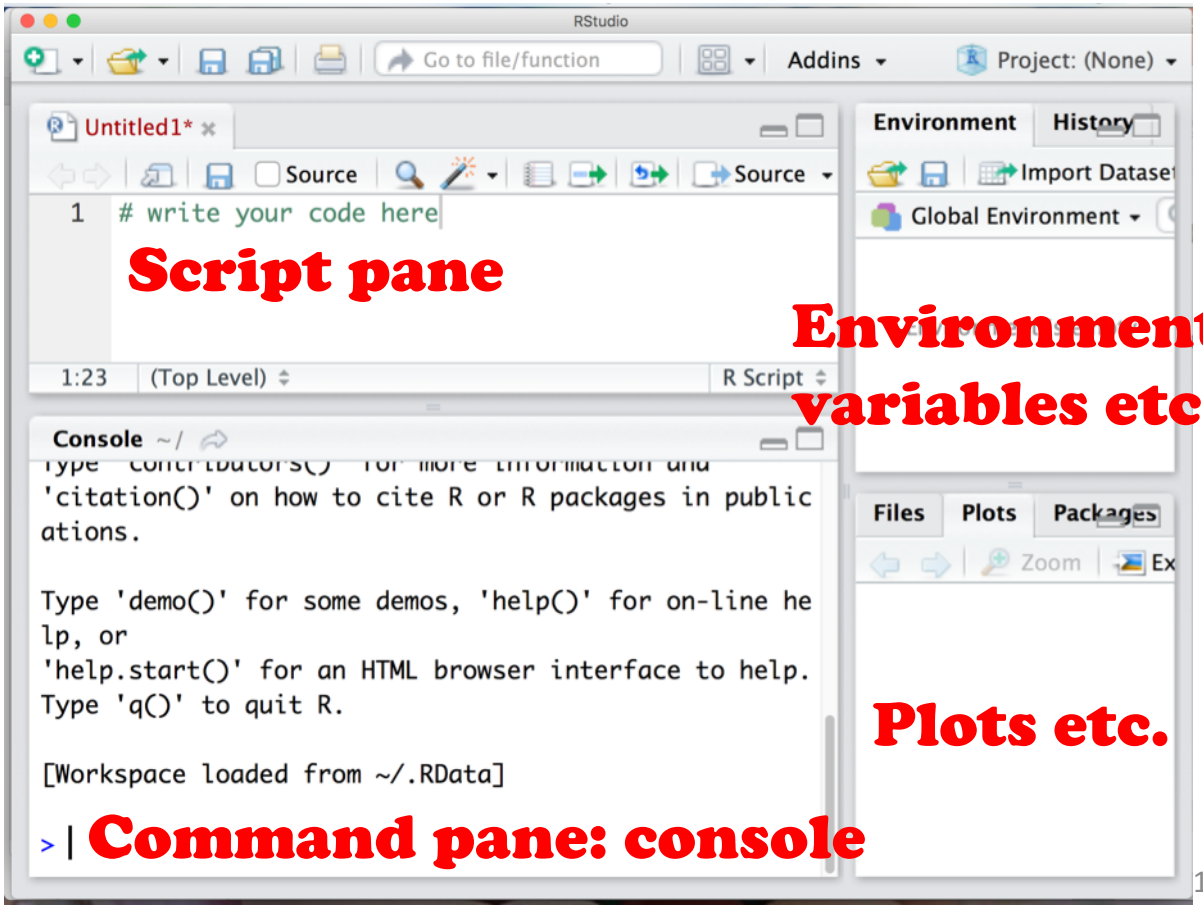
R is a free software environment for statistical computing and graphics. It con
variety of UNIX platforms, Windows and MacOS. To **download R**, please cho
mirror.

If you have questions about R like how to download and install the software,
are, please read our answers to frequently asked questions before you send a



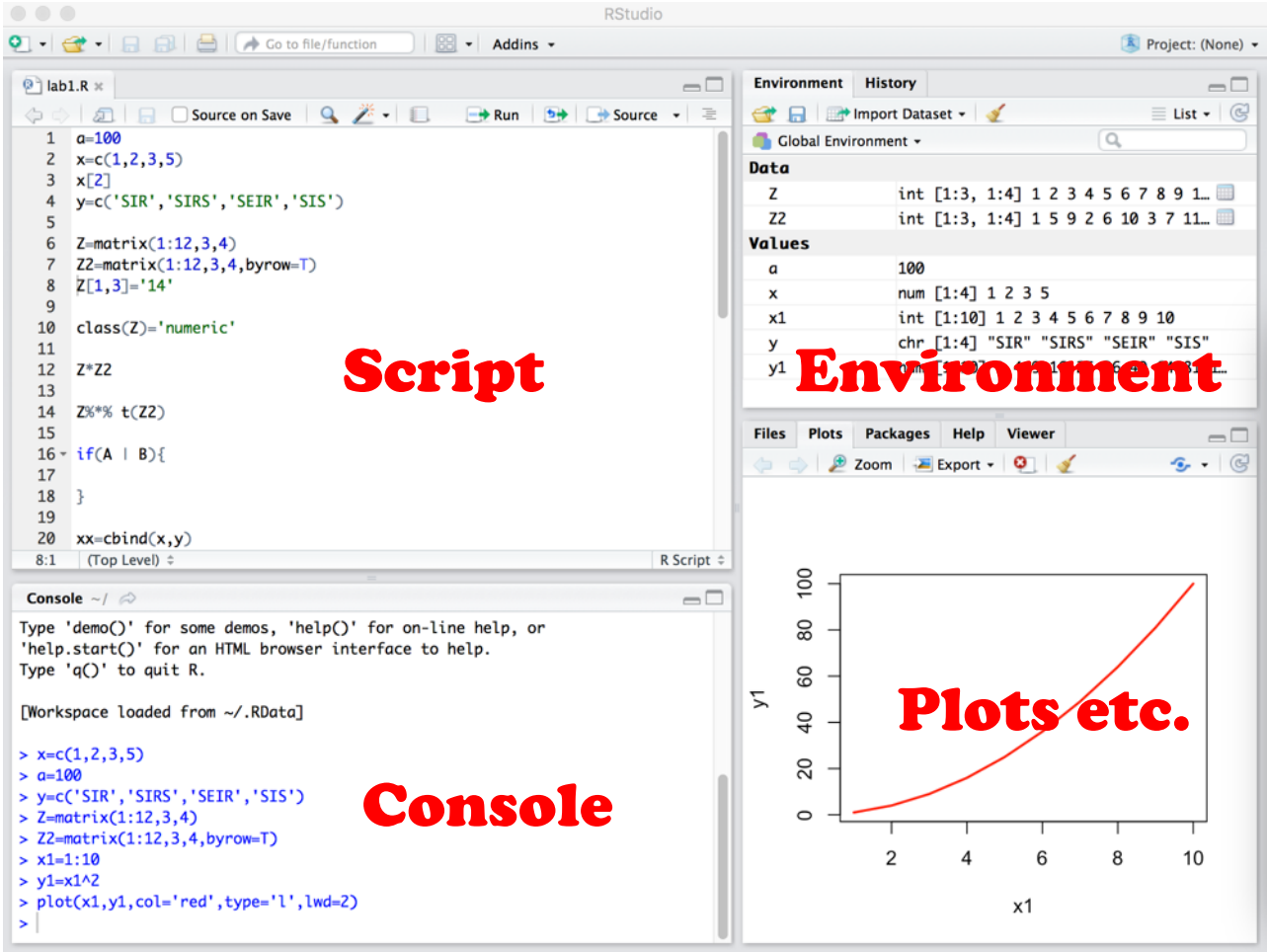
Let's get started…

- R studio: a user friendly version of R
  - https://www.rstudio.com/products/rstudio/download/
  - You'll need to install R first
- R studio interface



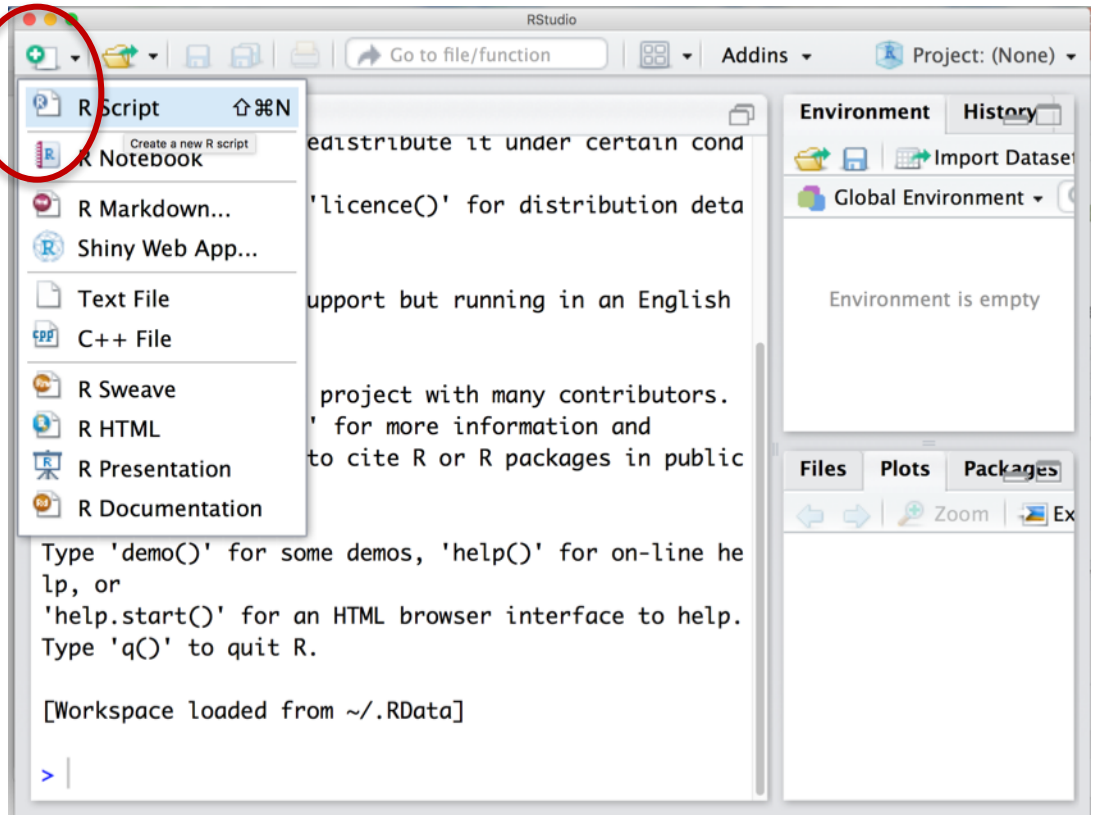
Let's get started…

- R studio: a user friendly version of R
  - https://www.rstudio.com/products/rstudio/download/
  - You'll need to install R first
- R studio interface



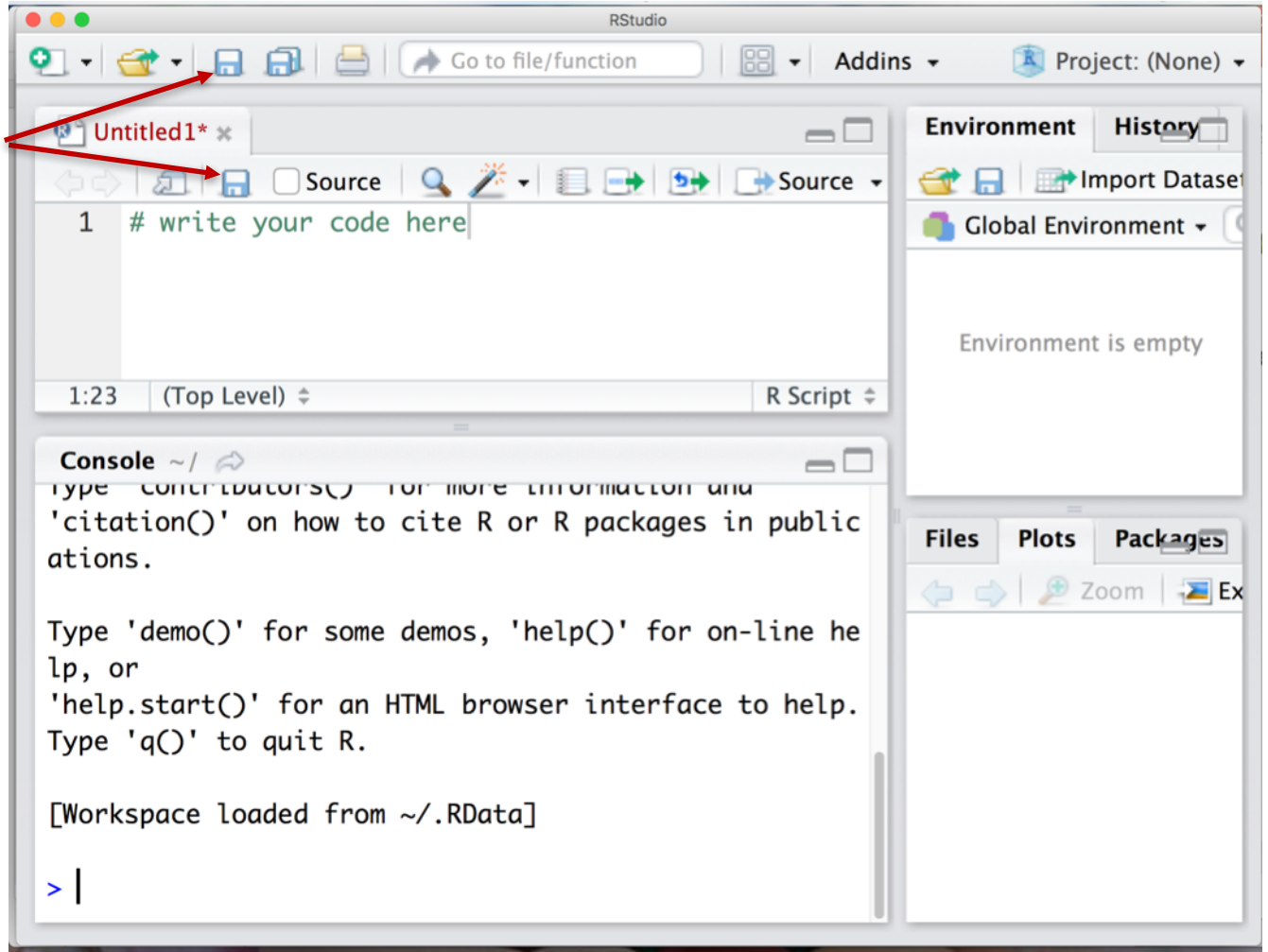
Write and Execute Commands

- Open a script pane to write your code
  - Tip #1: always save your script (so you can re-use it)



Write and Execute Commands

- Open a script pane to write your code
  - Tip #1: always save your script (so you can re-use it)

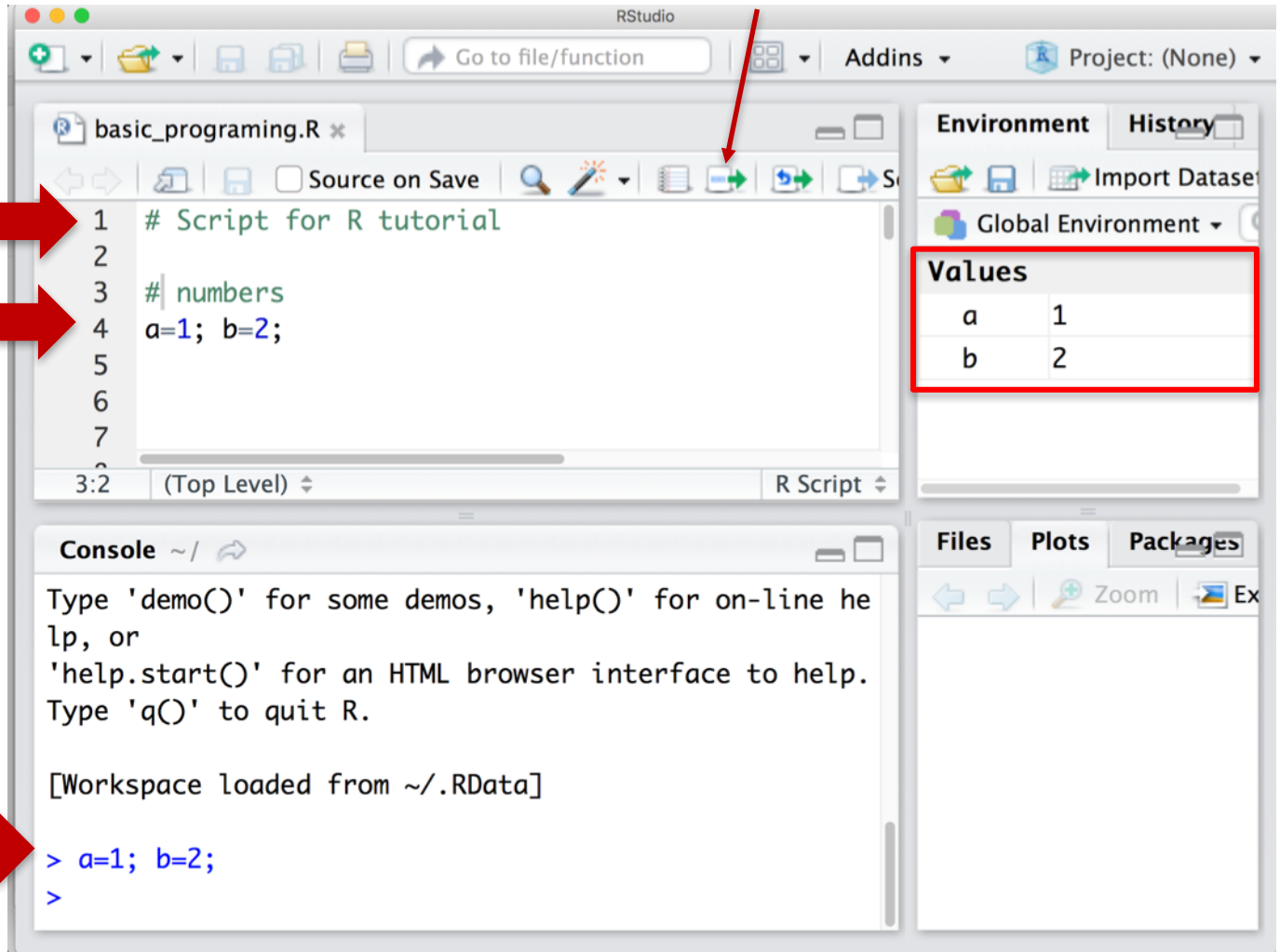


Variables/Data types

- Numbers: 1, 2, 3, …
  - Assign numbers: a=1; b=2;
  - Or a <- 1

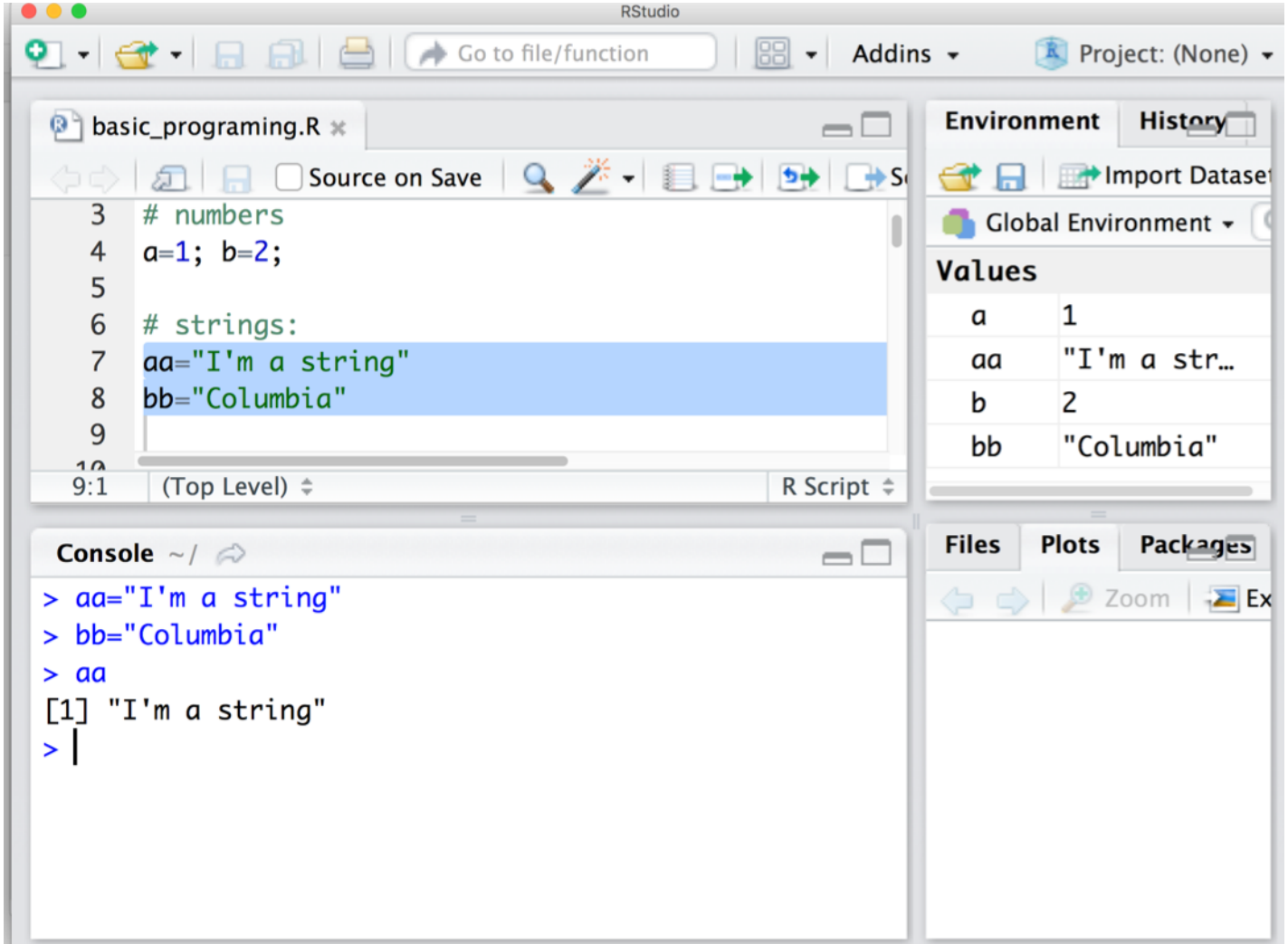
Execute commands
(or press
"Ctrl+Return"(Windows)/"Command+Return"(Mac)

Comments:
starting with "#"
(won't execute)



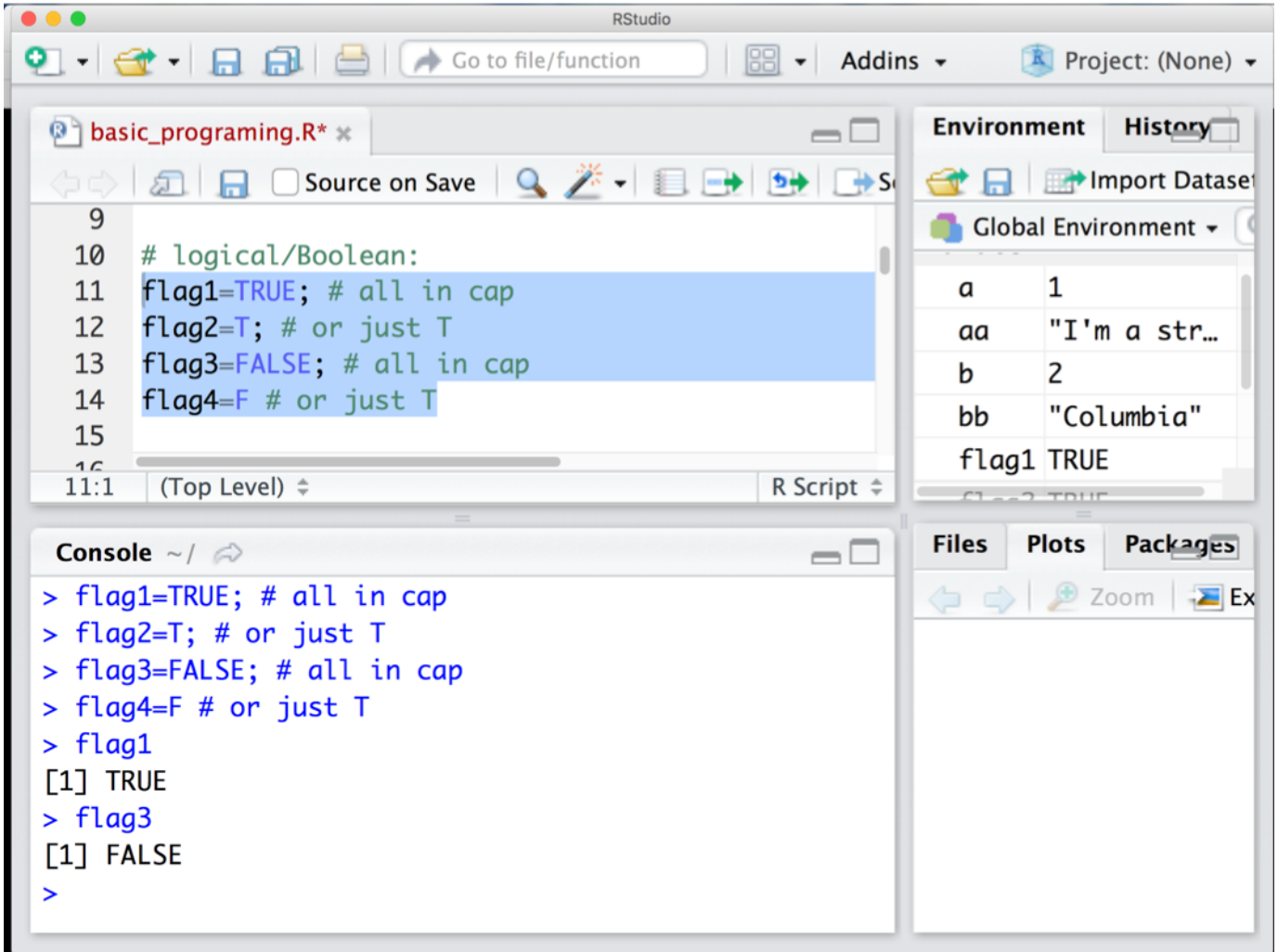
Variables/Data types

- Strings:
  - Assign strings: aa="I'm a string"; bb="Columbia"



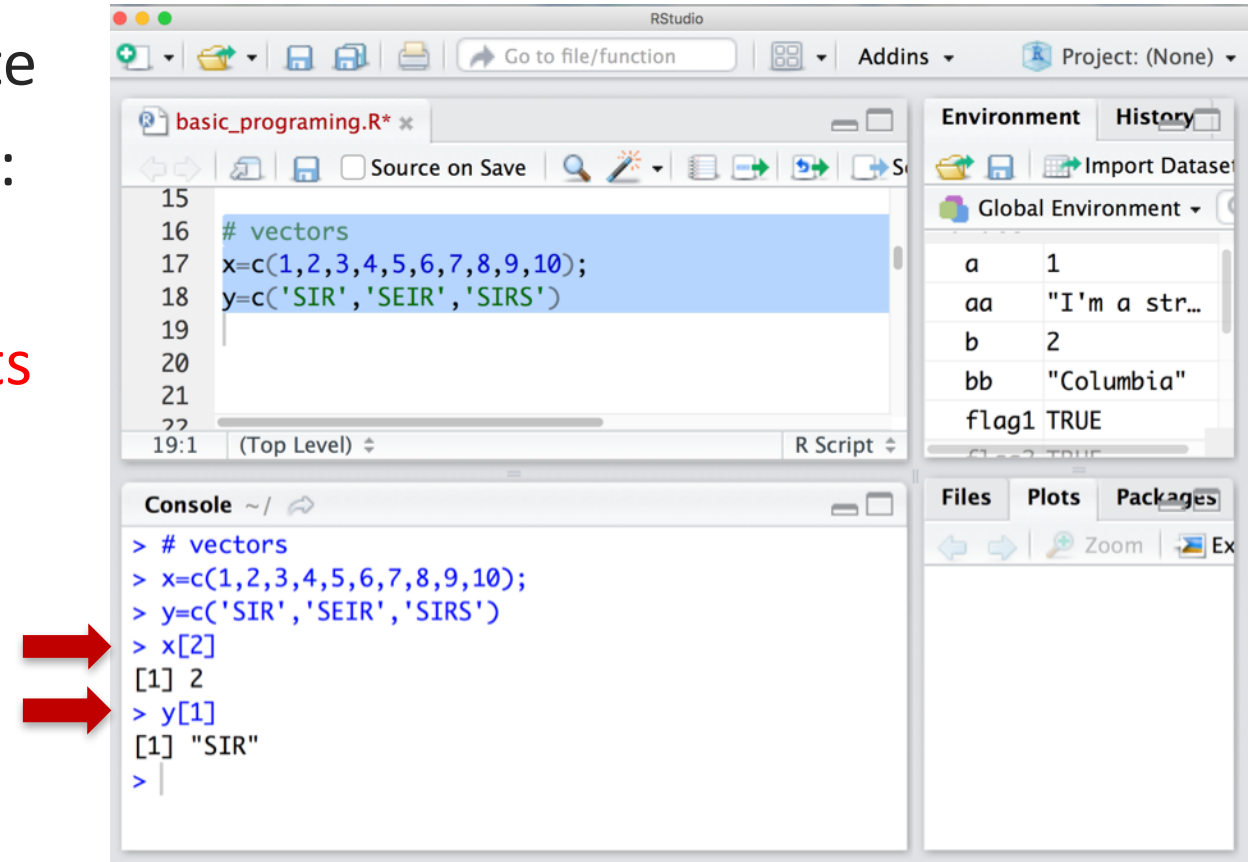
Variables/Data types

- Logical/Boolean:
  - Assign Boolean: TRUE (T), FALSE (F)



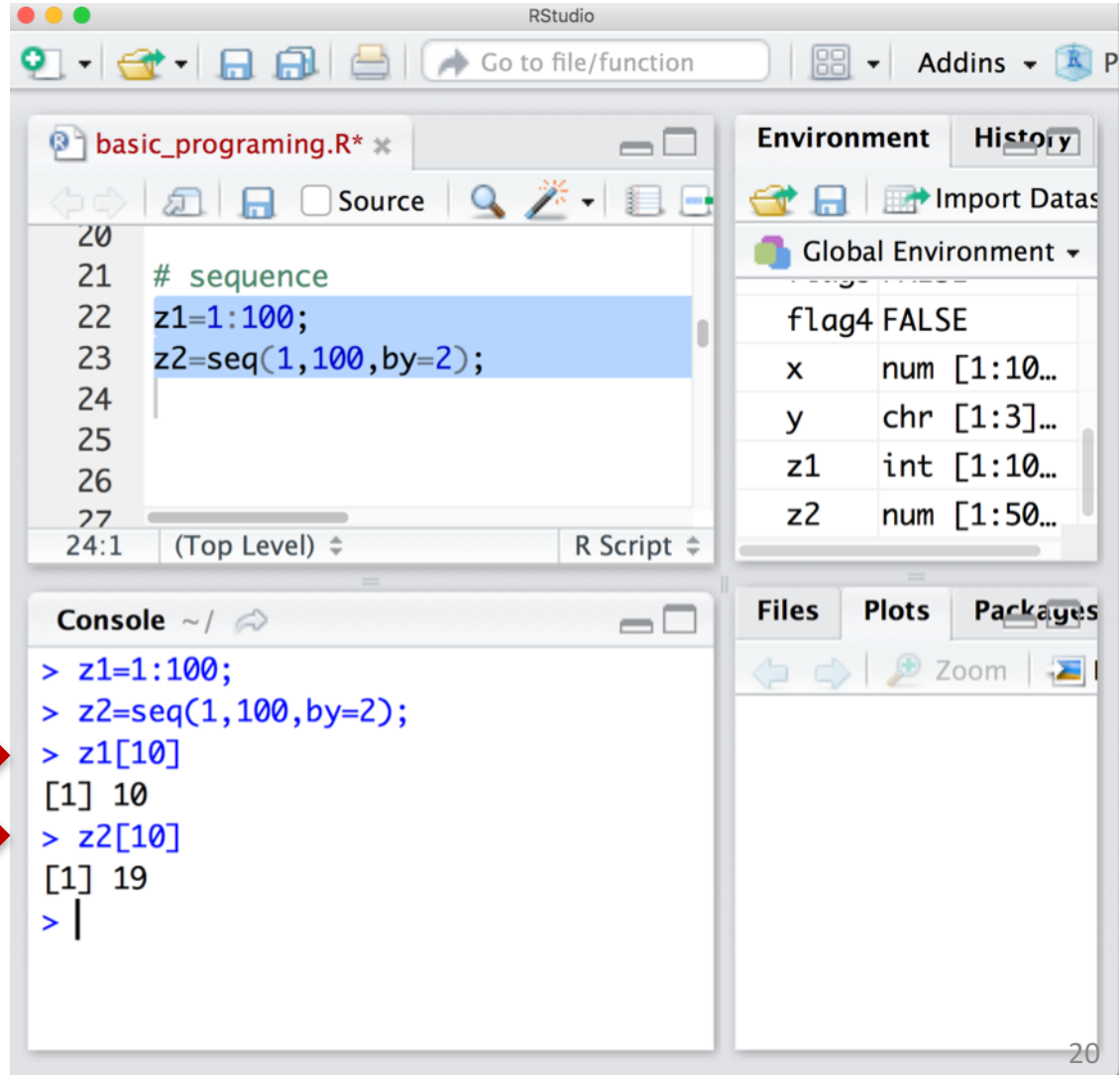
Variables/Data types

- Vectors: >=1 element
  - Assignment:
    - Numbers: x=c(1,2,3)
    - Strings: y=c('SIR', 'SEIR', 'SIRS')
    - c(): concatenate
  - Access element:
    - use [], e.g. x[1]
    - Index in R starts from 1



Variables/Data types

- Vectors: >=1 element
  - Sequence:
    - z2=seq(1,100,by=2)
      - ✓seq(): sequence
      - ✓by: increment
    - z1=1:100
      - ✓":": increase by 1

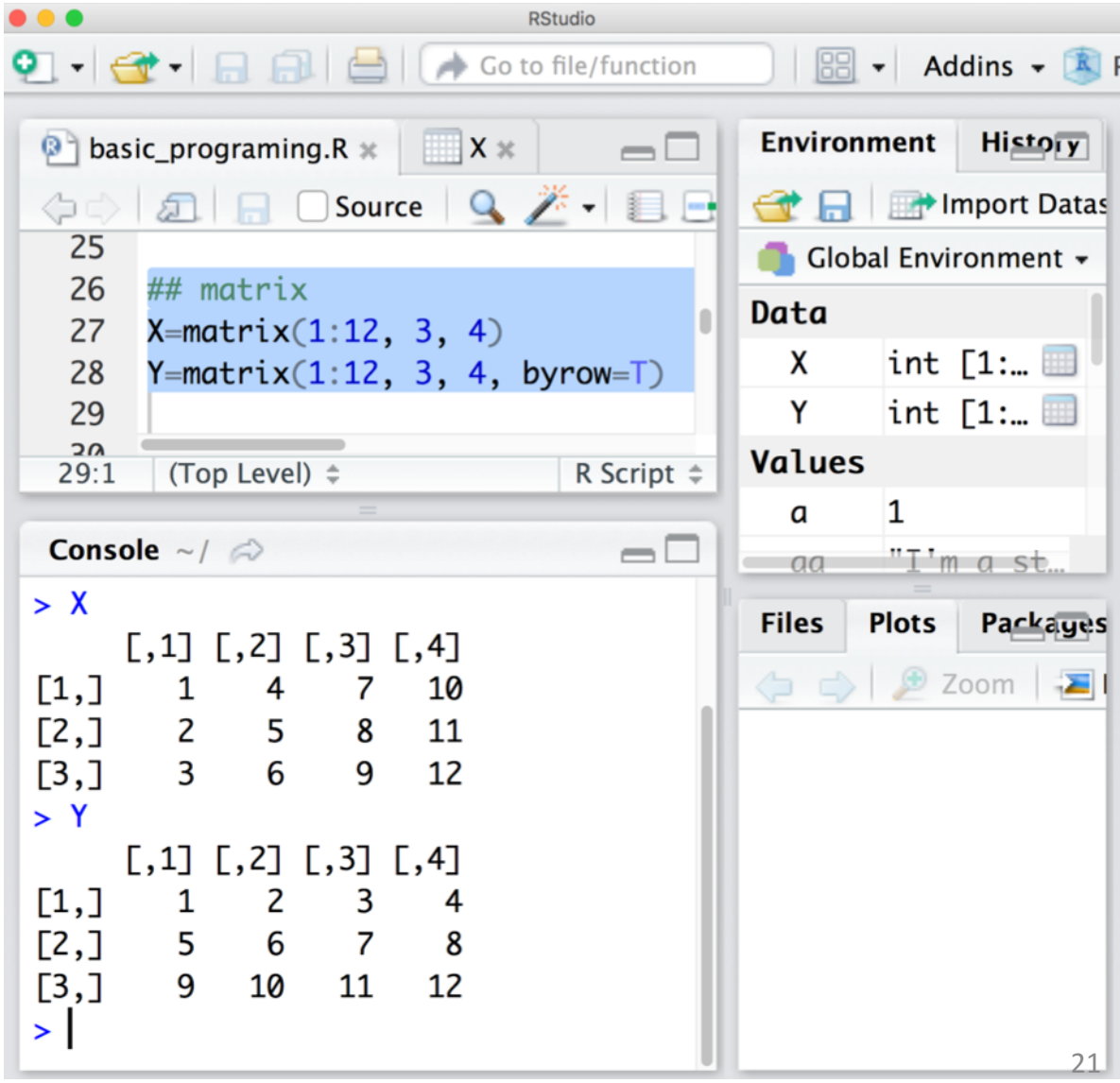


Variables/Data types

- Matrix:
  - X=matrix(1:12,3,4)

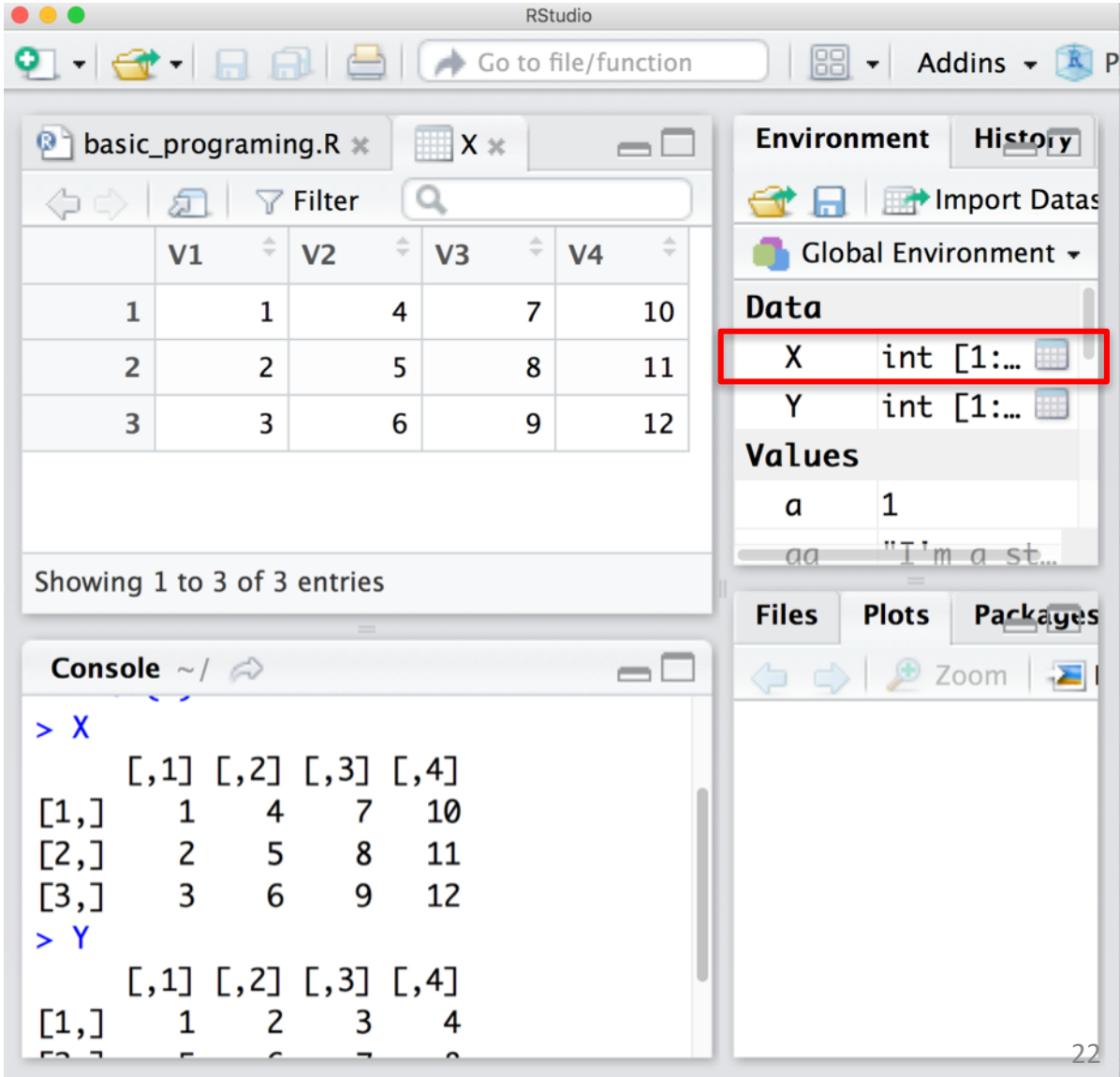
    matrix(elements to use in the matrix, #Row, #Col)
  - in R, elements are filled by column by default
  - To tell R to fill by row, use: Y=matrix(1:12,3,4, byrow=T)
  - Access element
    - e.g.: X[1,2]

      Row index, Col index



Variables/Data types

- Matrix:
  - X=matrix(1:12,3,4)
  - in R, elements are filled by column by default
  - To tell R to fill by row, use: Y=matrix(1:12,3,4, byrow=T)
  - Contains the same type of elements

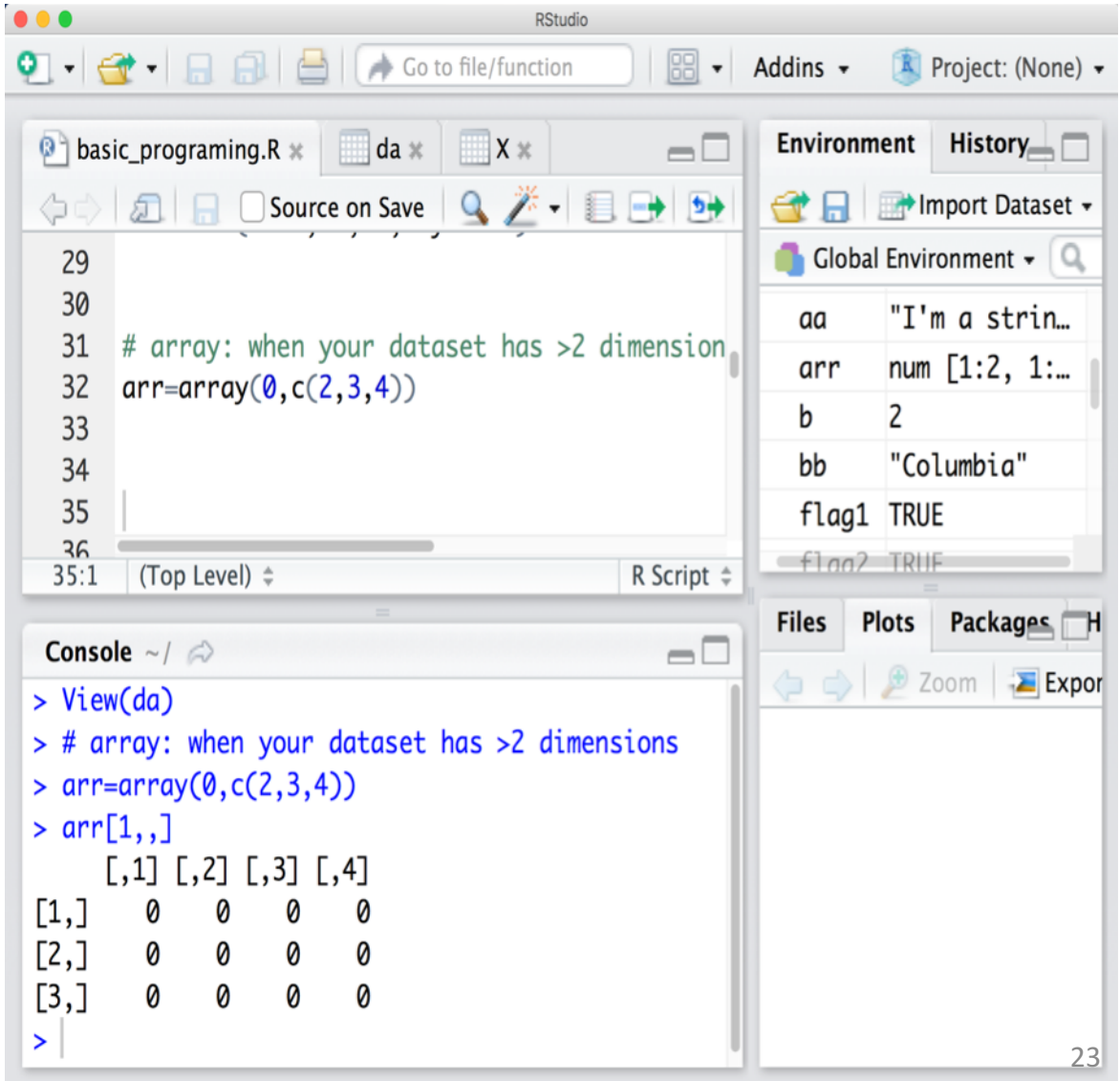


Variables/Data types

- Array: could be >2 dimensions
  - array()
  - arr=array(0,c(2,4,5))

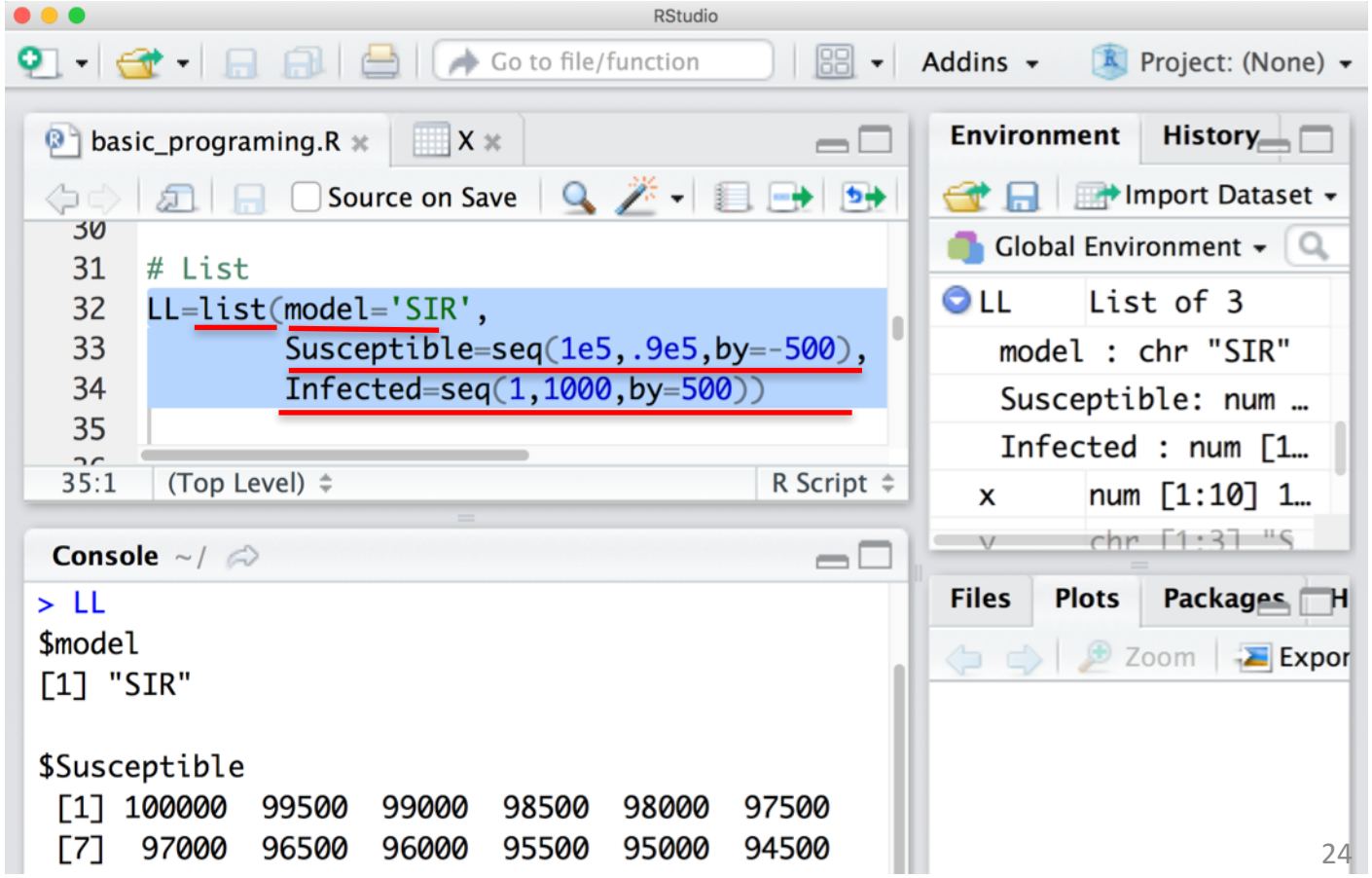
array(elements to use in the array, c(dim1, dim2,…))

  - To access elements:
    - arr[1,1,1]
    - arr[1,,]



Variables/Data types

- List: an object consisting of an ordered collection of objects known as its components.
  - list()
  - Elements can be of different data types vs. same data type in a vector



Variables/Data types

- Data Frame: an organized list
  - data.frame(...)
  - allow different data types for different columns
    - vs matrix (only 1 data type for all elements
  - e.g.: put different features together (name, weight, height, etc.); rows are records

Index	Weight	Group
1	92.9	Control
2	86.8	Control
3	81.3	Control
4	73.1	Experiment
5	94.3	Control
6	94.9	Experiment
7	67.4	Control
8	81.7	Control
9	93.7	Experiment
10	60.0	Control

Variables/Data types

- Special variables:
  - NA: missing value
  - Inf: infinity (e.g. divide by 0)
  - pi: 3.1415…

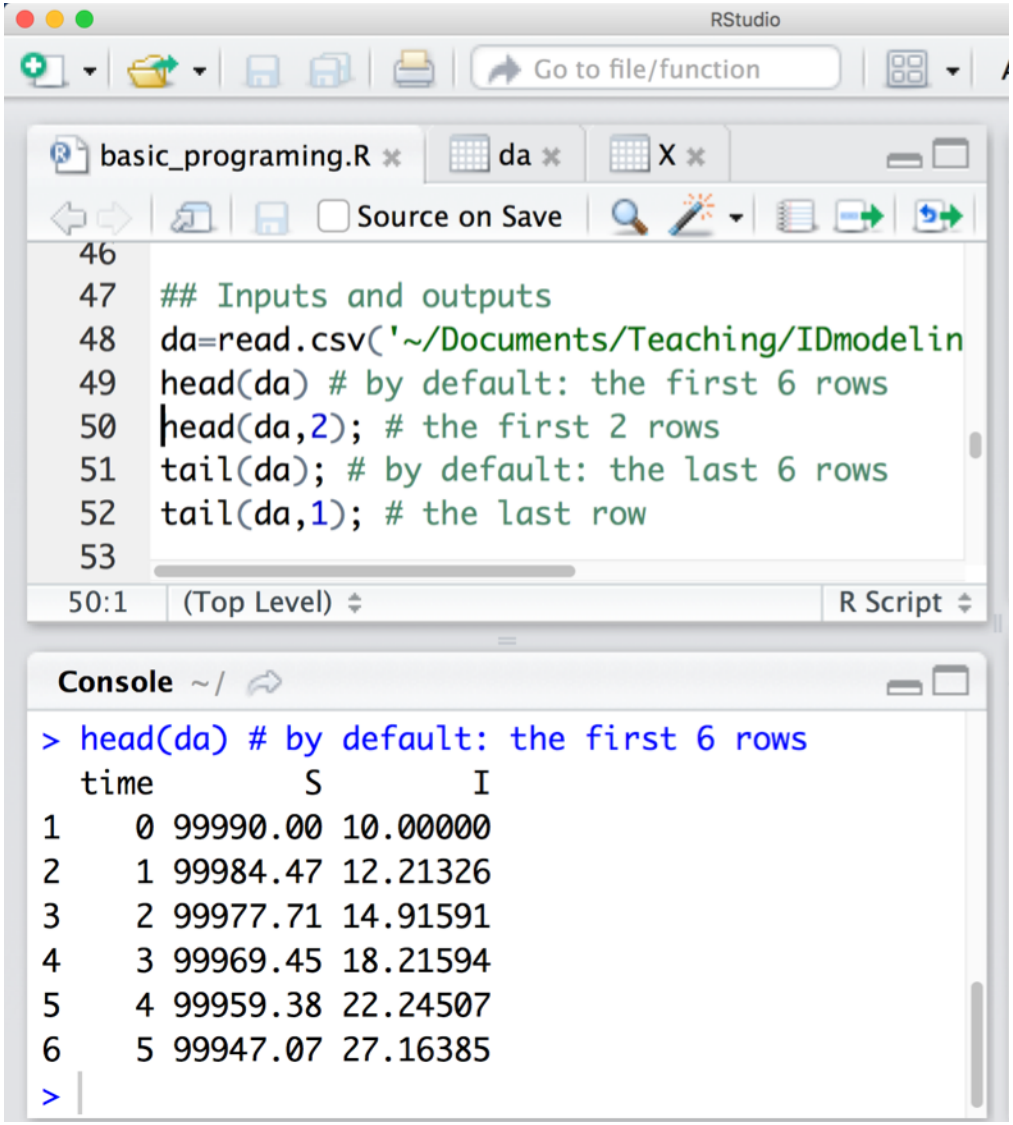
(Math) Operations

- summation: 1 + 3
- subtraction: 3 - 2
- multiplication: 3 * 2
- division: 3 / 2
- square root: sqrt()
- exponential: x^y
- logarithm:
  - ✓ natural log: log();
  - ✓ other base: log(x, base=Base)
- matrix multiplication: %*%
- mod (residual): %%
- logic: & (and); | (or)
- Use () to arrange the order of operations: (100+20)*5 v. 100+20*5

Don't forget '*': e.g., x times y should be 'x*y', not 'xy'

Built-in functions

- c(); seq(); matrix(); array()
- head(); tail()
- rep();
- sum(); mean(); sd(); median(); var();
- min(); max();
- sort(); order();
- rbind(); cbind()
- paste(); paste0()
- summary(); table()
- Learn more:
  - help()
  - ? Function name



Customized functions

- Customized functions
  - EX:

Name of the function [choose as you like, but should start with a character (not number)]

Function input(s)

The function function output: result(s)

```
194 myfunc=function(a,b){
195   result=(a+b)^2
196   return(result)
197 }
198 X=matrix(1:16,8,2)
199 y=rep(0,8)
200 for(i in 1:8){
201   y[i]=myfunc(X[i,1],X[i,2])
202 }
```

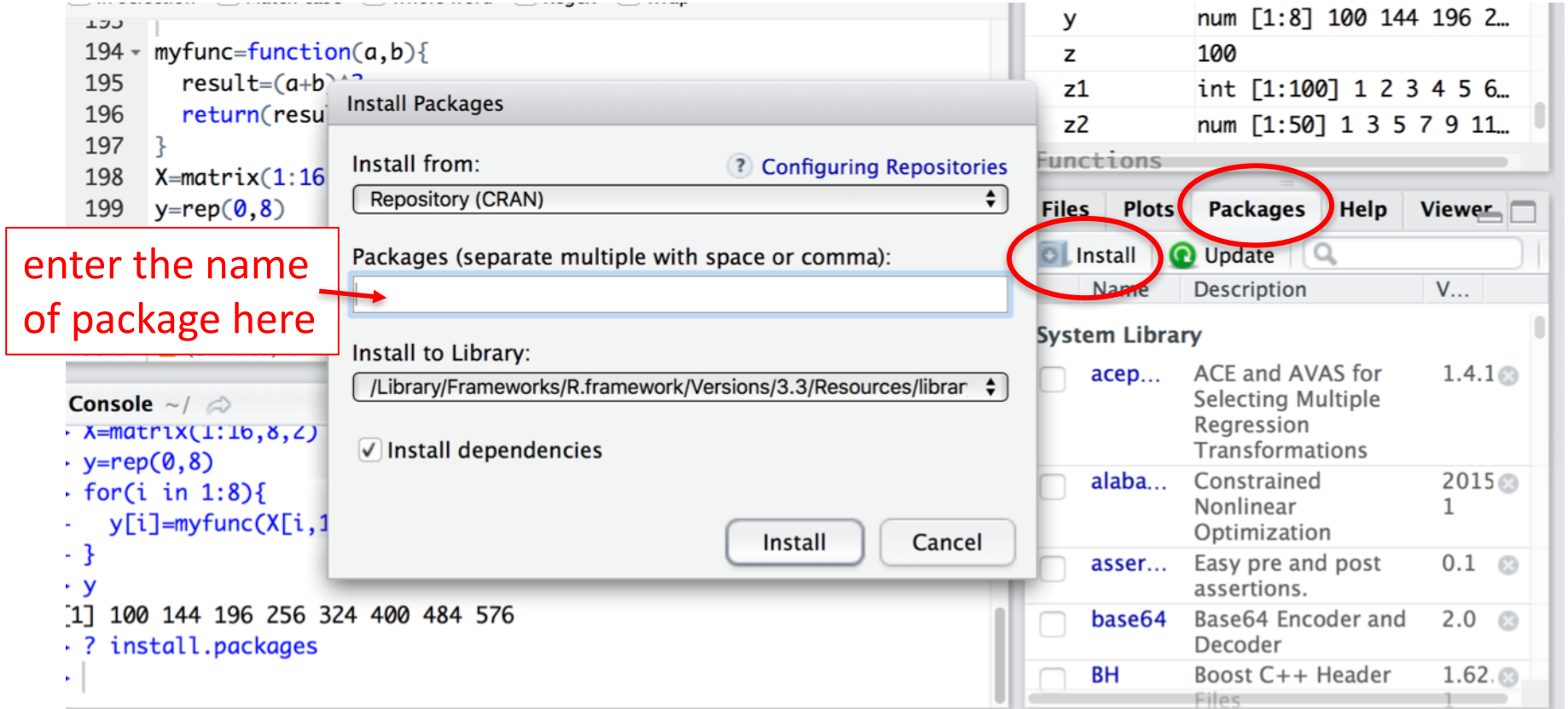
193:1 (Untitled)

Console ~/

```
+ }
> X=matrix(1:16,8,2)
> y=rep(0,8)
> for(i in 1:8){
+   y[i]=myfunc(X[i,1],X[i,2])
+ }
> y
[1] 100 144 196 256 324 400 484 576
>
```

Install a “library” (“package”)

- What is a library/package?
  - All R functions and datasets are stored in packages
- Install a library
  - command: install.packages(“NAME OF PACKAGE”)

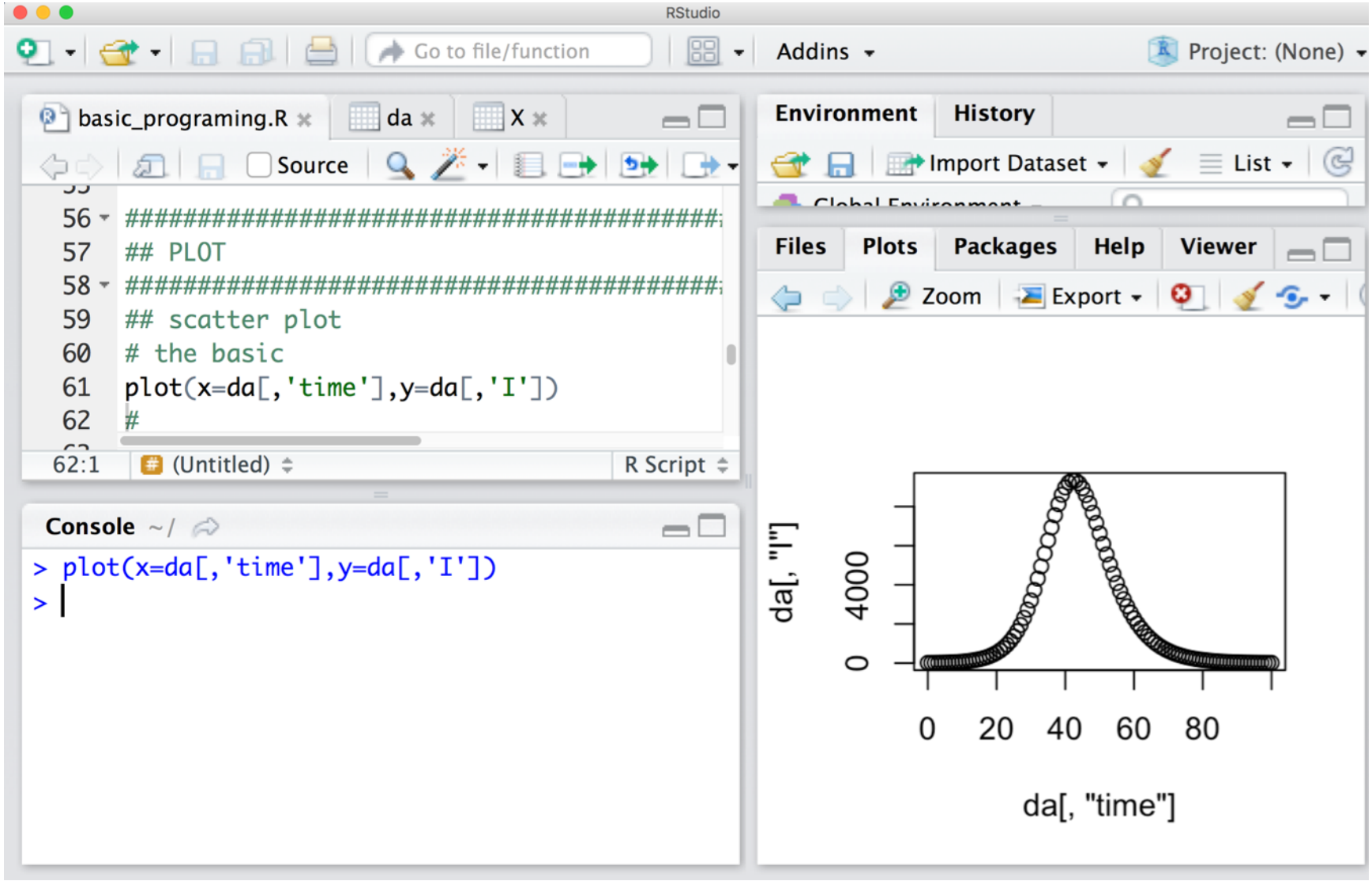


Install a “library” (“package”)

- What is a library/package?
  - All R functions and datasets are stored in packages
- Install a library
  - command: install.packages (“NAME OF PACKAGE”)
- Load a library: need to load it before use
  - command: library (“NAME OF LIBRARY”)

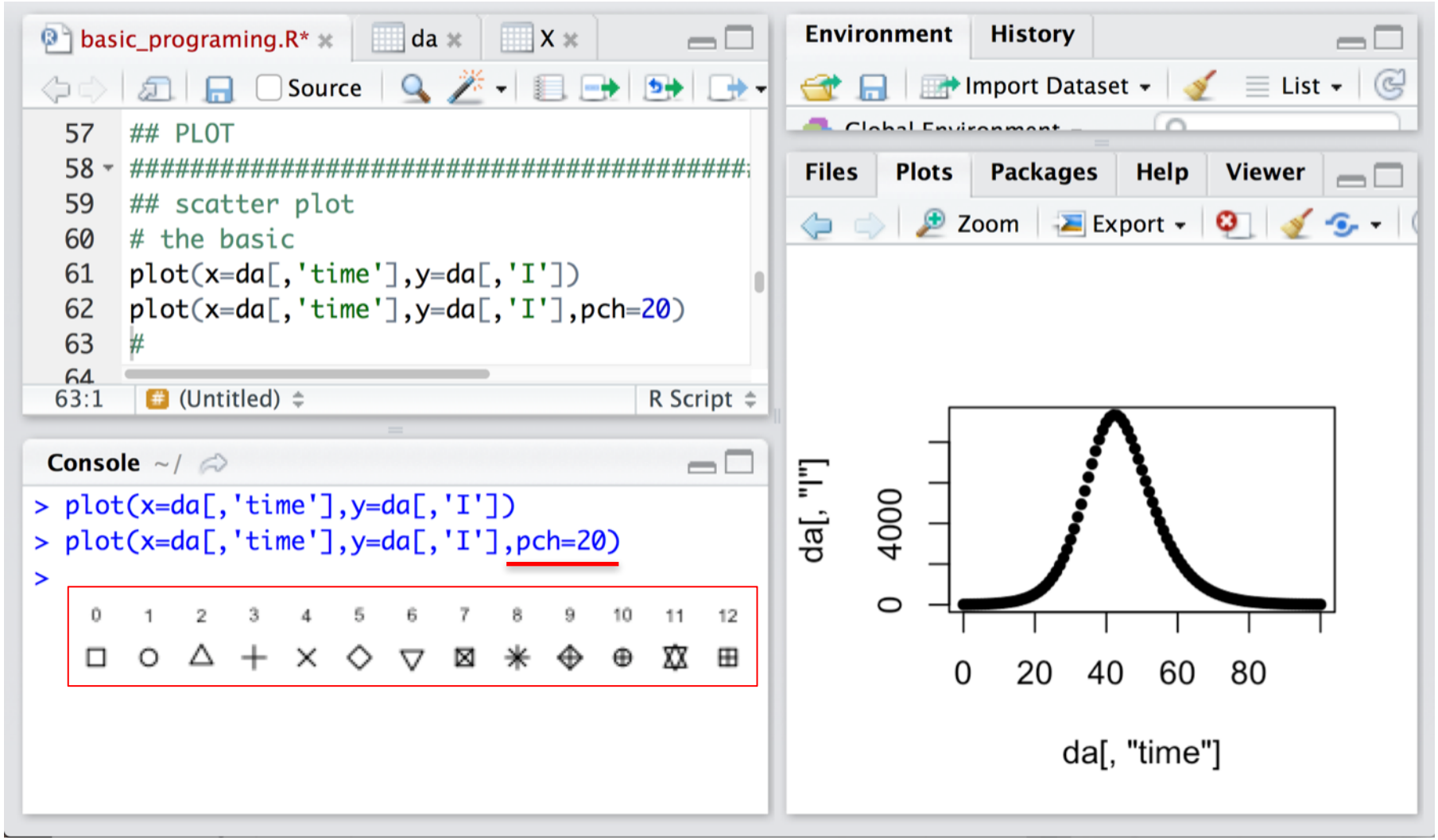
Plots

- Scatter plot: plot(x,y,…)



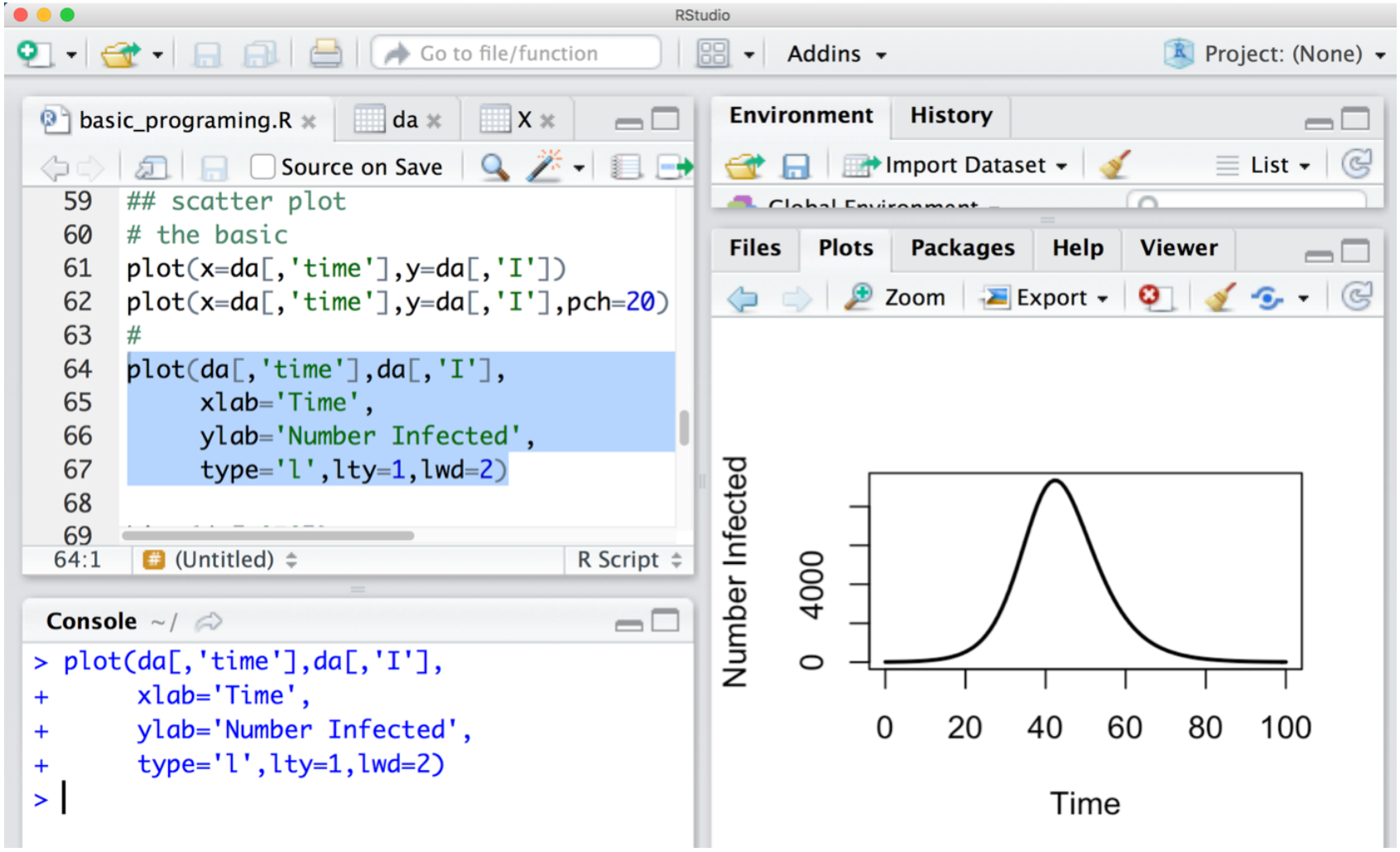
Plots

- Scatter plot: plot(x,y,…)



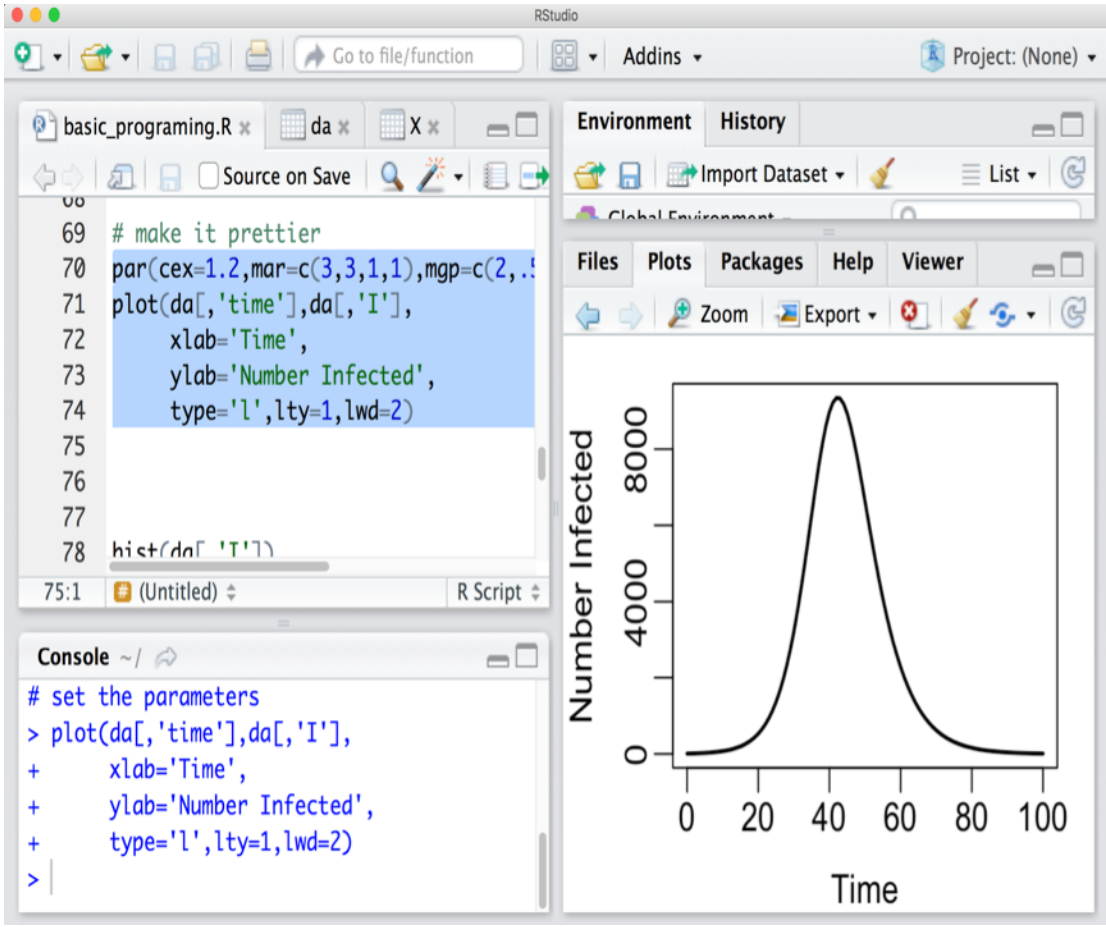
Plots

- Scatter plot: plot(x,y,…)
  - Add another line to the same plot: lines (x, y, …)



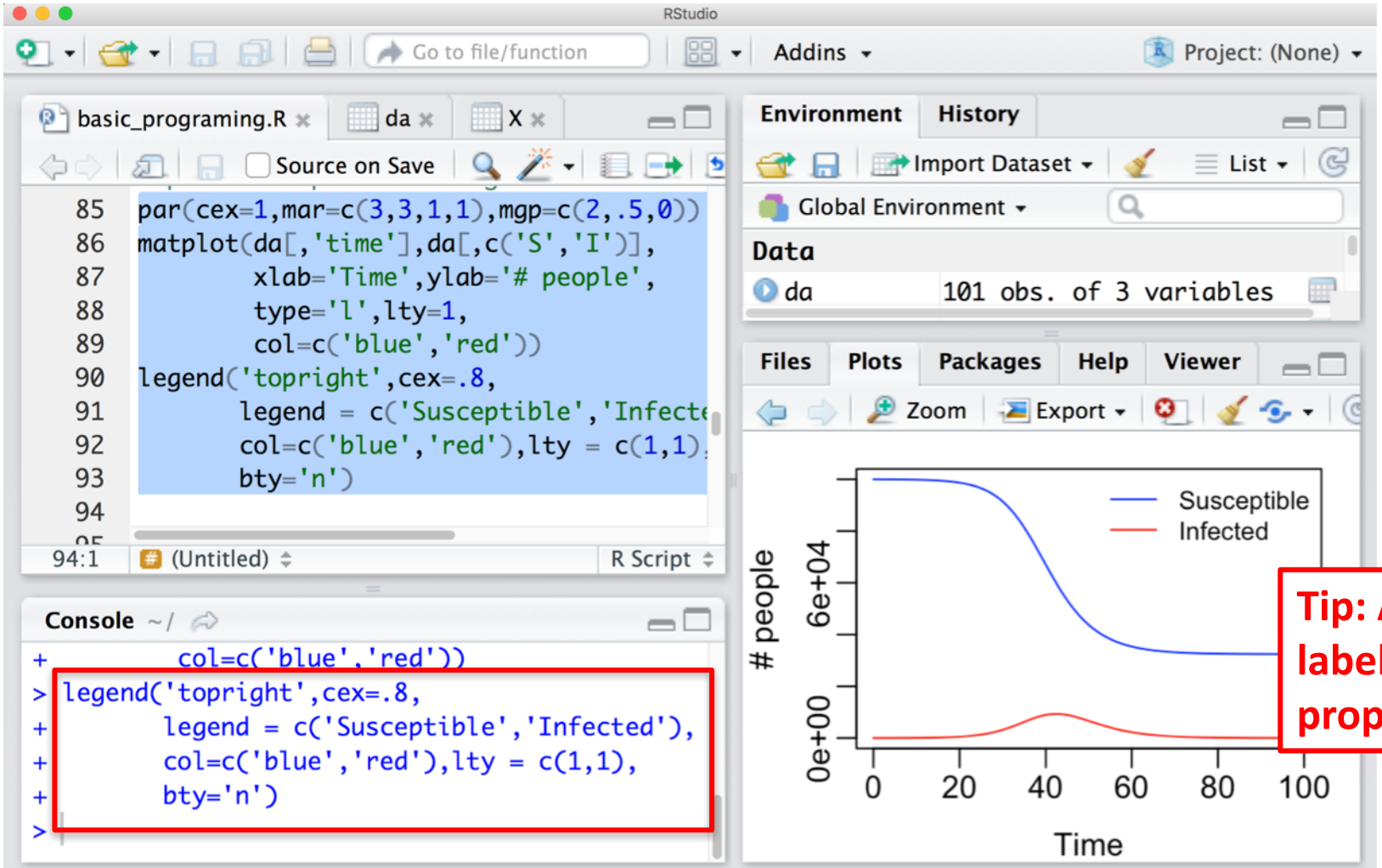
Plots

- Scatter plot: plot(x,y,…)
  - Make it prettier: par()
    - E.g.: par(mar=c(3,3,1,1), cex=1.2, mgp=c(2,.5,0))
      - ✓ Margin: mar=c(3,3,1,1)
      - ✓ Font size: cex=1.2
      - ✓ The margin to axis title, axis labels and axis line: mgp=c(2,0.5,0)



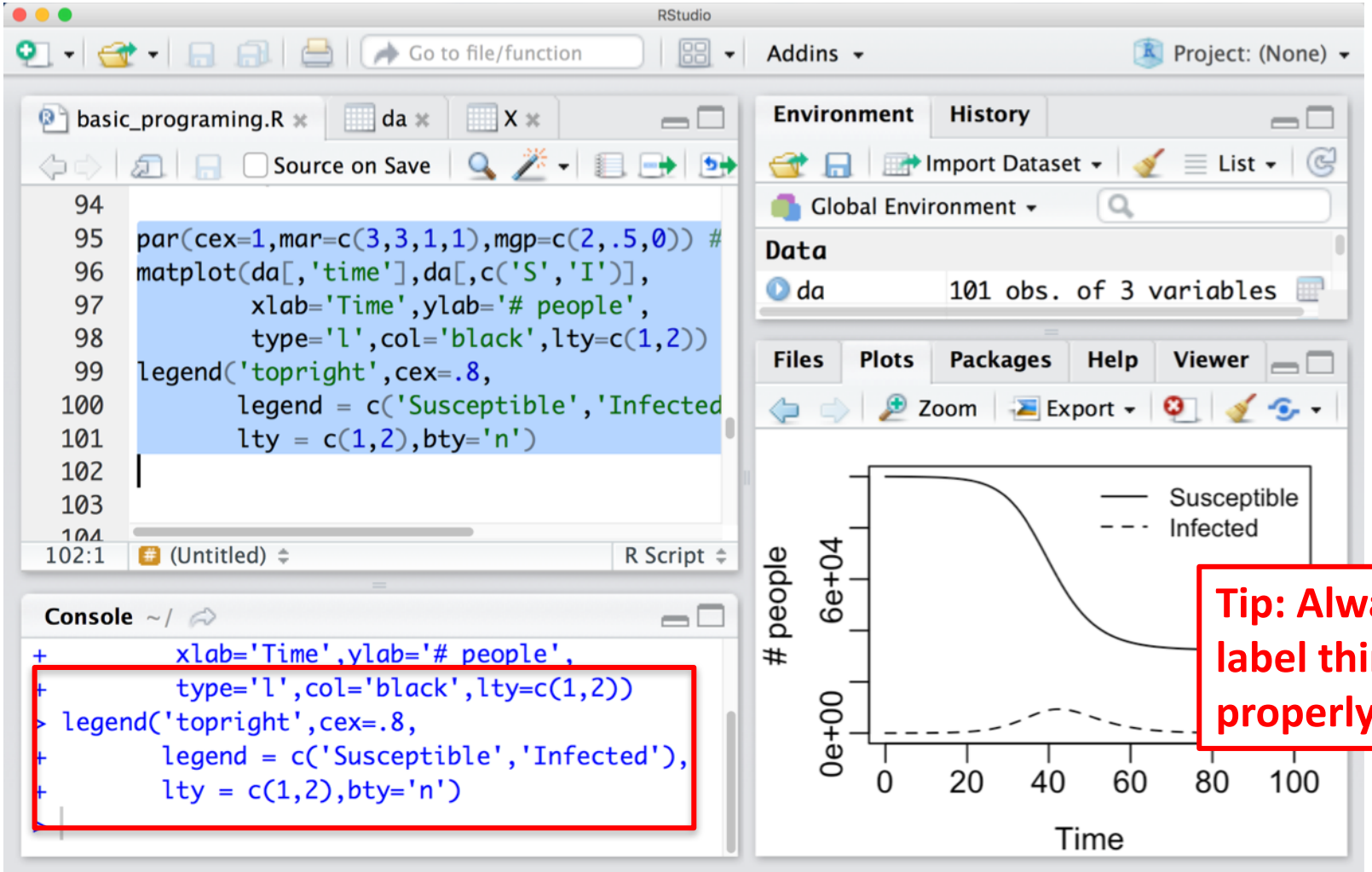
The Basic: Plots

- Scatter plot & legend: plot multiple lines together
  - matplot(x,y,…)



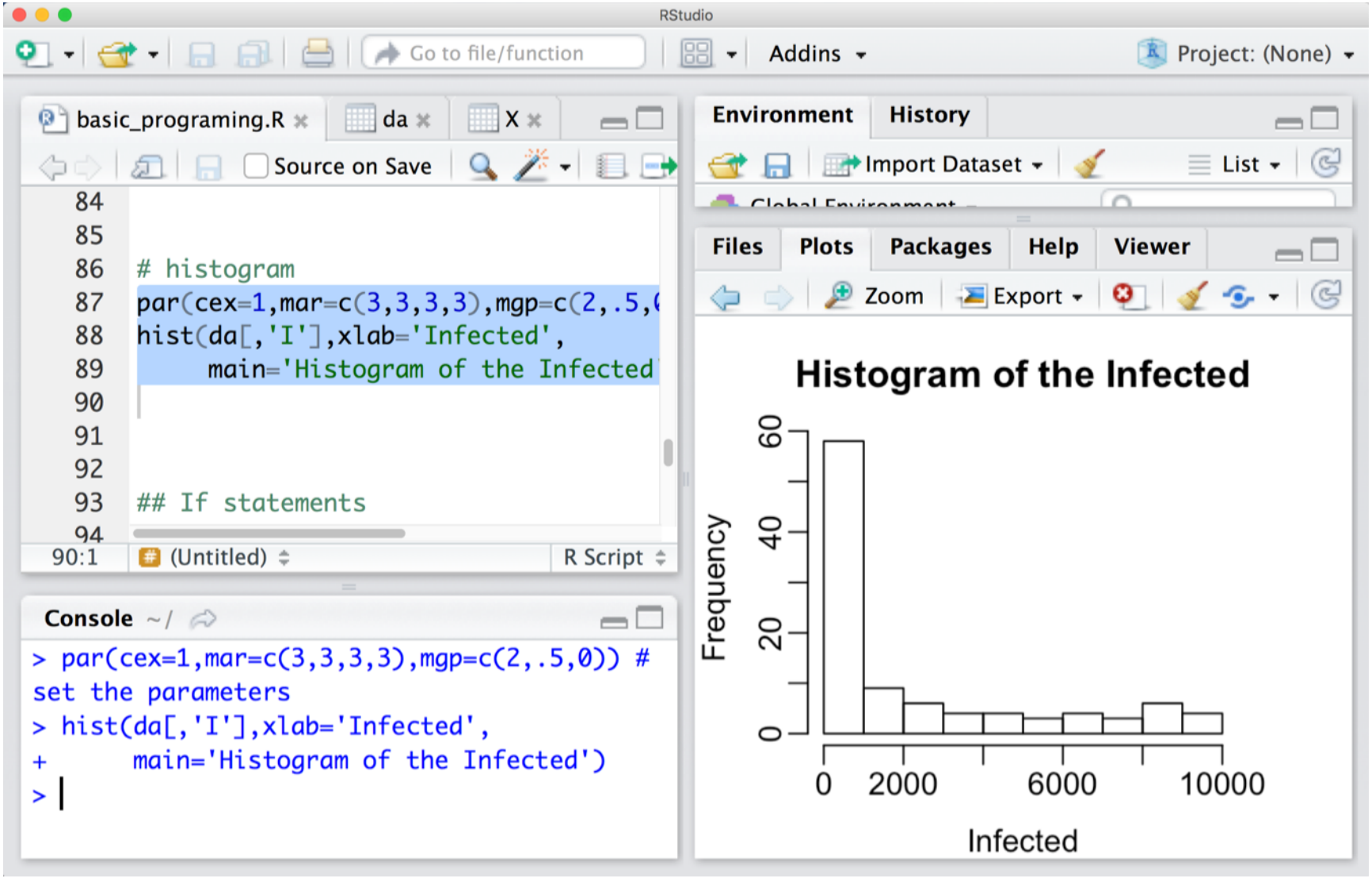
Plots

- Scatter plot & legend: plot multiples line together
  - matplot(x,y,…)



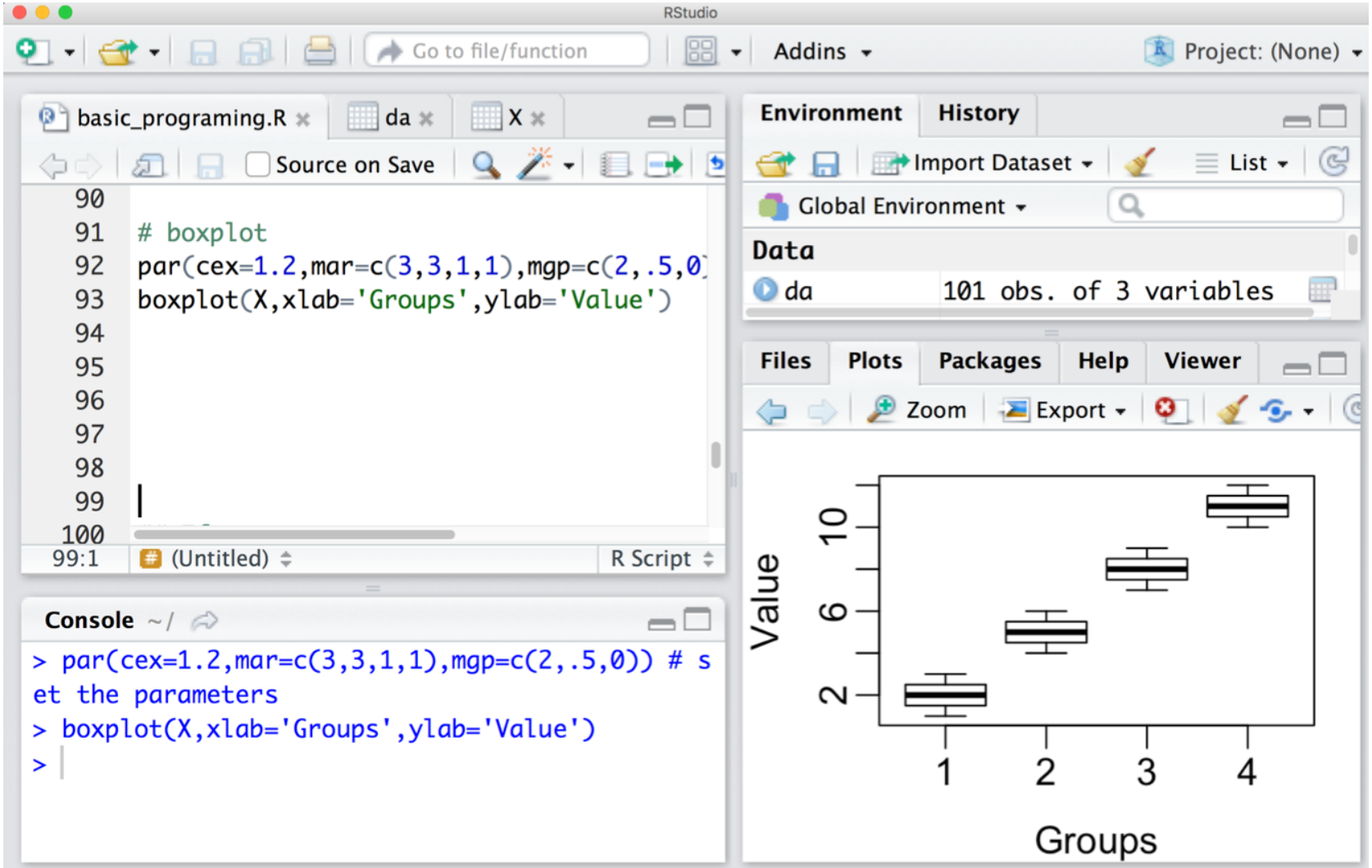
Plots

- Histogram: hist(x,…)



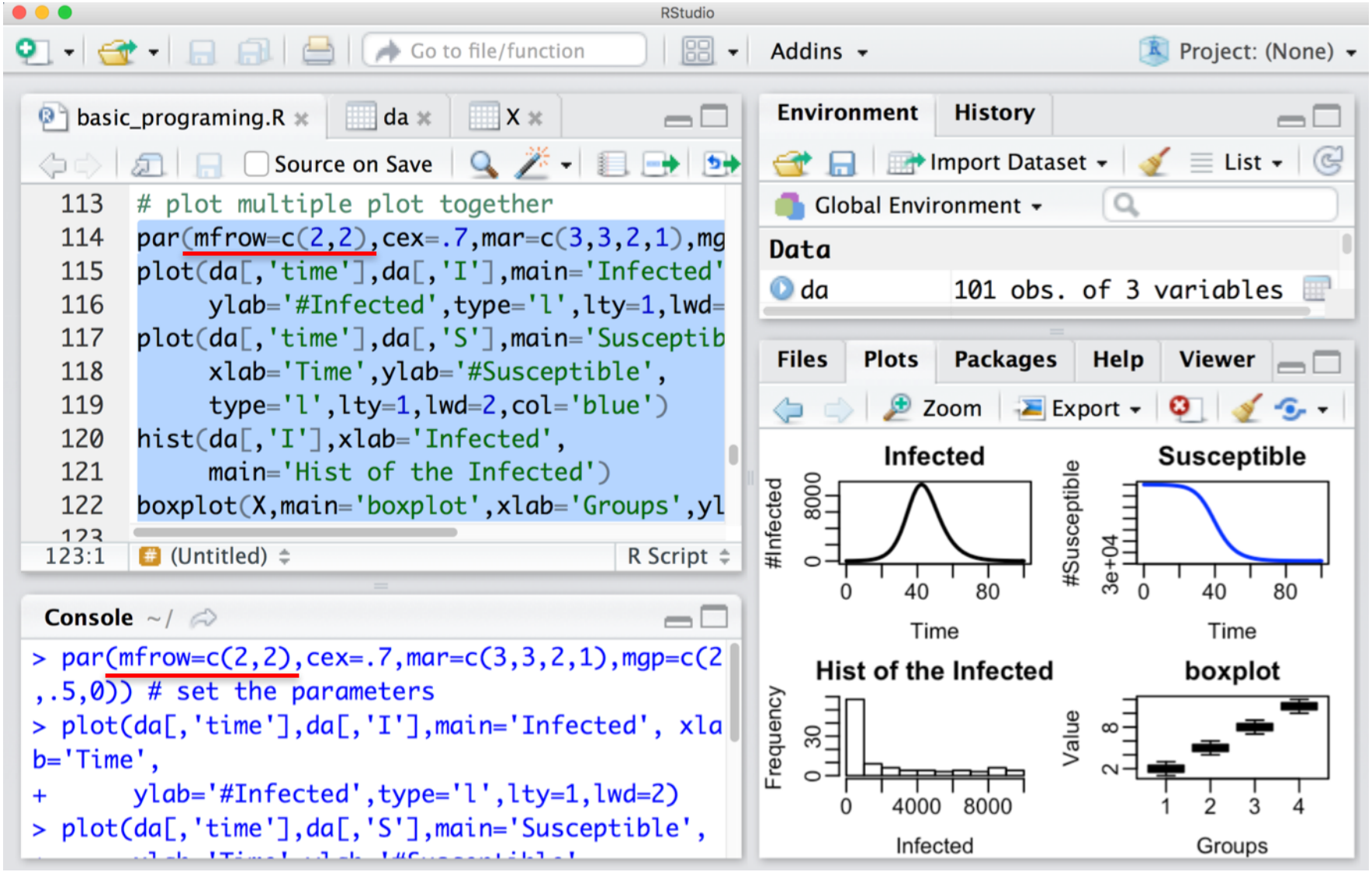
Plots

- Boxplot: boxplot (x,…)
- Barplot: barplot (…)



Plots

- Put multiple plots together



The Basic: Input and Output

- Read in a file
  - read a .csv file: read.csv('path')
  - read a table: read.table('path')
- Save files
  - save as .csv: write.table(x, 'pathToSave.csv', col.names=T, row.names=F, sep=',') # comma separated values
  - save R project:
    - save.image('path/filename.RData')

Things to Know:
General Programing Tips

- R is case sensitive
  - Aa ≠ aa
- Index in R starts from 1 (not 0)
- Save your code: write in the script pane, not the console directly
- Write comments:
  - #...SO YOU KNOW WHAT YOU ARE DOING 3 MONTHS LATER…
- Use the 'help' function to learn more about each function
  - Or Google
- Practice, practice, practice
  - You will remember the syntax after using it many times
  - To do: work through Worksheet Part I
  - Homework (Lab report #1)

More Resources on R

- More R tutorials:
  - edX (Date Science: R Basics)
    https://www.edx.org/course/data-science-r-basics-harvardx-ph125-1x-0
  - Youtube, Coursera (https://www.coursera.org)
  - http://www.cyclismo.org/tutorial/R/index.html
- Google/ R help
- CourseWorks:
  - Modules -> R Tutorial -> Short-R-Intro.pdf
    - Short-R-intro.pdf
    - Long-R-intro.pdf