

# Introduction to R for Data Science

Lecture 2, August 27  
Fall 2014

# Outline

1. What is R
2. Why use R
3. Installing R in your own computer
4. R studio
5. Your first script
6. Data Structures and Manipulation in R

# What is R?

R is a powerful, versatile, and free statistical programming language. Scientists, statisticians, analysts, students and others who are interested in statistical analysis, data visualization, etc. are using R to do so.

Data analysis is done in R by writing or using built in scripts and functions in the R language. The R environment is not only equipped with all the standard methods, but also some of the most recent cutting-edge techniques.

R is open source. This means that you can download and use R for free, and additionally the source code is open and available for inspection and modification.

# Why use R?

- \* R is free and open.
- \* R is a language. You learn much more than just point and click.
- \* R has excellent tools for graphics and data visualization.
- \* R is flexible. You are not restricted to the built in set of functions, you can use them and extend them with your own.

# How to Obtain R for your own computer?

## Windows:

<http://cran.r-project.org/bin/windows/base/>

## MacOs X:

<http://cran.r-project.org/bin/macosx/>

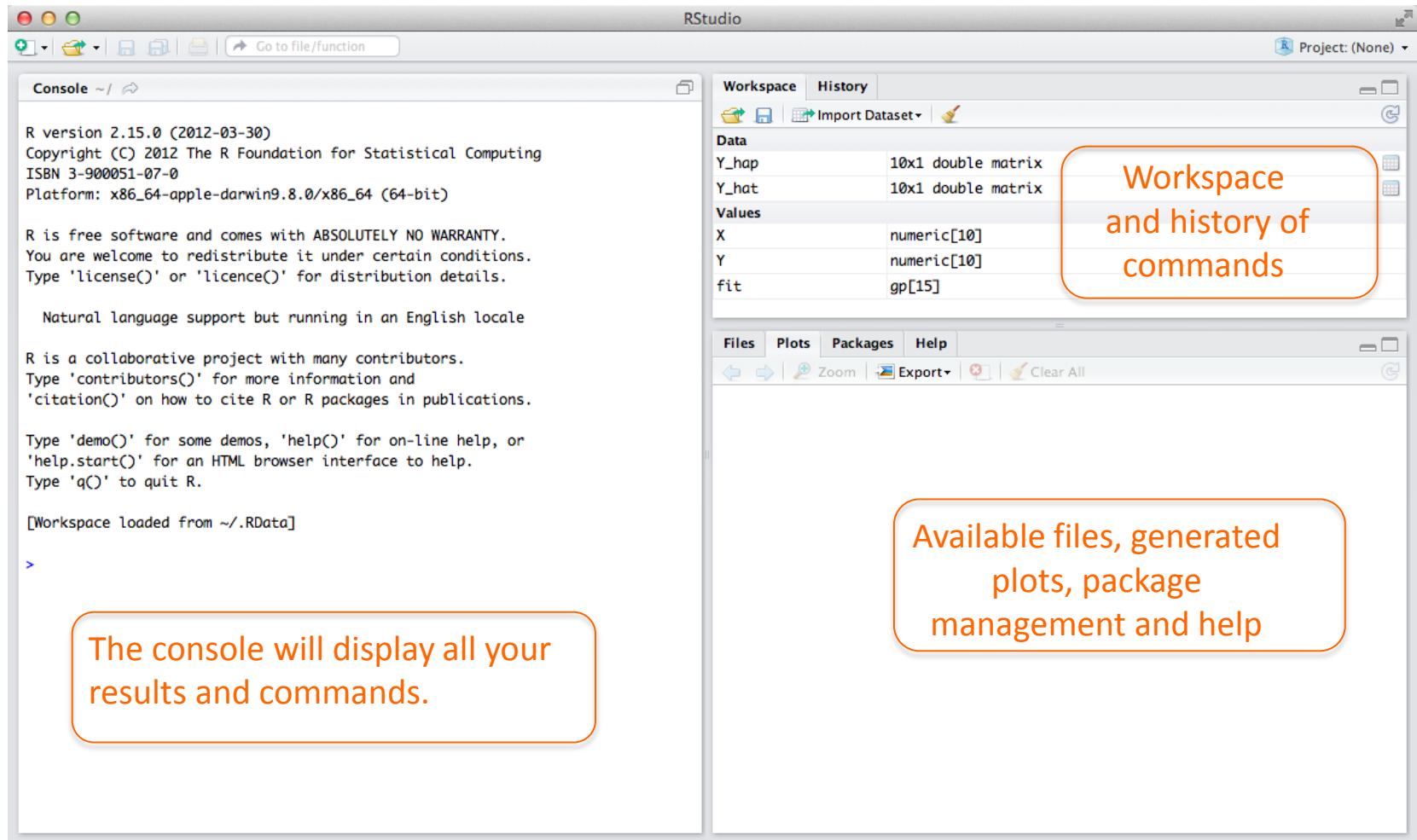
# How to Obtain R Studio for your own computer? (after you got R)

<http://www.rstudio.com/products/rstudio/download/>

We are installing RStudio Desktop.

Then choose the installer for your own system (Windows/Mac) and click on it. Follow through with the installation instructions.

# R Studio



# R Studio menu

- Open files.
- New files.
- Save files.
- Run scripts (Ctrl+R in Windows).
- "New Folder". Create a folder "Rfolder" on your computer, where you will save all your R files.



# First script

- Set working directory to the R folder you created
- Create “Hello World” object
- Create comments with “#”
- Get to help from menu or type `help.start()`
- Save R script and workspace to your R folder

# Data Structures and Manipulation

## 1. Object Creation

Expression: A command is given, evaluated and the result is printed on the screen.

Arithmetic: +, -, /, \*, ^

Logical: ==, <, >, <=, >=, !=, &; |

Assignment: Storing the results of expressions.

<- (recommended)

=

# Data Structures and Manipulation

## 2. Vectors:

The basic data structure in R. (Scalars are vectors of dimension 1).

### a. Creating sequences:

- `:` command. Creates a sequence incrementing/decrementing by 1
- `seq()` command.

### b. Vectors with no pattern. `c()` function.

### c. Vectors of characters. Also use `c()` function with the help of `""`

### d. Repeating values. `rep()` function.

### e. Arithmetic with vectors: All basic operations can be performed with vectors.

### f. Subsets: The basic syntax for subsetting vectors is: `vector[index]`

# Data Structures and Manipulation

## 3. Matrices: Objects in two dimensions.

### a. Creating Matrices

Command: `matrix(data, nrow, ncol, byrow)`.

`data`: list of elements that will fill the matrix.

`nrow, ncol`: number of elements in the rows and the columns respectively.

`byrow`: filling the matrix by row. The default is `FALSE`.

### b. Some Matrix Functions

- `dim()`: Lists the dimensions of the matrix.
- `cbind`: Creating matrix by putting columns together.
- `rbind`: Creating matrix by putting rows together.
- `diag(d)`: Creates identity matrix of dimension `d`.

# Data Structures and Manipulation

## c. Some Matrix computations

- Addition.
- Subtraction
- Inverse: function `solve()`
- Transpose: function `t()`
- Element-wise multiplication: `*`
- Matrix multiplication: `%*%`

## d. Subsets

- Referencing a cell: `matrix[r,c]`, where `r` represents the row and `c` represents the column.
- Referencing a row: `matrix[r,]`
- Referencing a column: `matrix[,c]`

# R Object Naming

Object names:

- Start with letter (recommended) or dot. IF starts with dot- second character can't be digit.
- Contain only letters, numbers, underscores, dots.
- case sensitive (lastname <> Lastname)
- make names understandable.