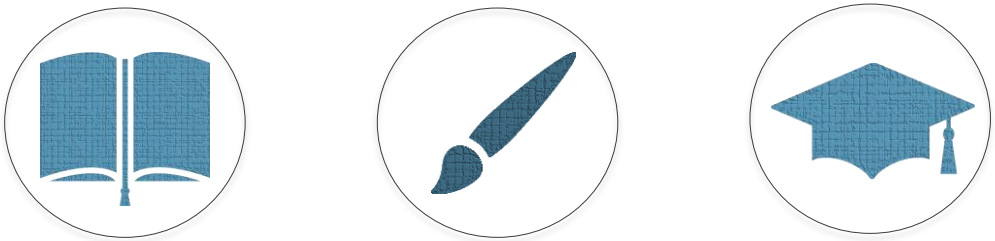


Data Science Using R

Lesson01–Overview of R

Objective

After completing this lesson you will be able to:

- Describe the genesis of R
- Locate and install R in the system
- Explain R Studio interface
- Install packages from the repositories

Overview and History of R

R is a dialect of the S language. S is a language that was developed by John Chambers and others at Bell Labs.

Features:

- Runs on almost any standard computing platform/OS
- Frequent releases (annual + bug fix releases); active development
- Quite lean, as far as software goes; functionality is divided into modular packages
- Very sophisticated graphics capabilities; better than most stat packages
- Useful for interactive work, but contains a powerful programming language for developing new tools
- Very active and vibrant user community; contains R-help and R-developer mailing lists and Stack Overflow

Design of the R System

- The R system is divided into two conceptual parts:

Base R

- Base R can be downloaded from CRAN (http://cran.r-project.org).
- The base R system contains, among other things, the base package, which is required to run R and contains the most fundamental functions.

R Extension

- There are about 4000 packages on CRAN that have been developed by users and programmers around the world.
- There are also many packages associated with the Bioconductor project (http://bioconductor.org).

R Resources

- Number of references and reading material can be found on R.

R resources available from CRAN

- An Introduction to R
- Writing R Extensions
- R Data Import/Export
- R Installation and Administration (mostly for building R from sources)
- R Internals (not for the faint of heart)

Standard books

- Chambers (2008). Software for Data Analysis, Springer (your textbook)
- Chambers (1998). Programming with Data, Springer
- Venables & Ripley (2002). Modern Applied Statistics with S, Springer
- Venables & Ripley (2000). S Programming, Springer
- Pinheiro & Bates (2000). Mixed-Effects Models in S and S-PLUS, Springer
- Murrell (2005). R Graphics, Chapman & Hall/CRC Press

R Dataset and Tools

- R comes with a number of sample datasets that you can experiment with.

On R Studio Console:

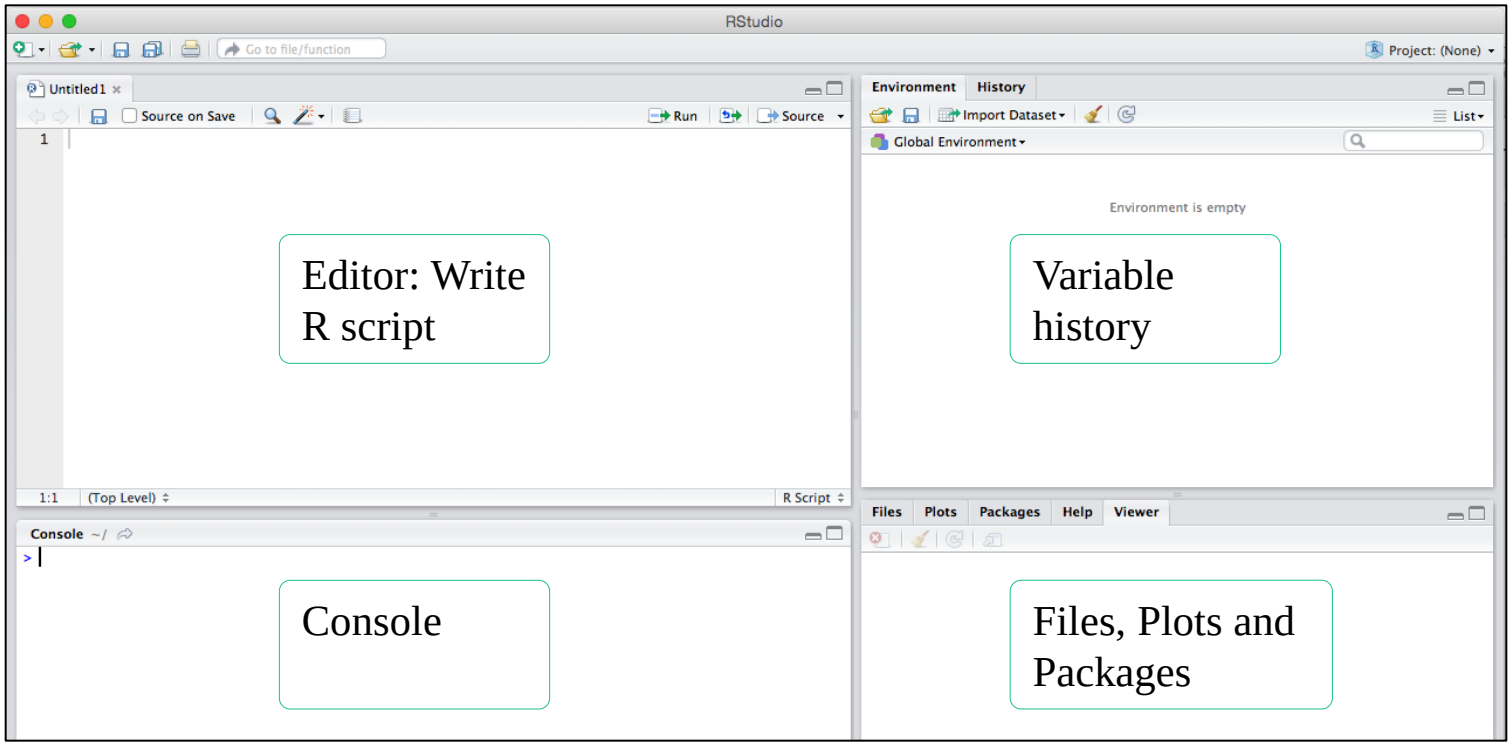
```
data( ) #to see the available datasets in the installed packages
help(datasetname)  #for details on a sample dataset
```

Help on the error: "Failed to set default locale". Open terminal and type the following:

```
defaults write org.R-project.R force.LANG en_US.UTF-8
```

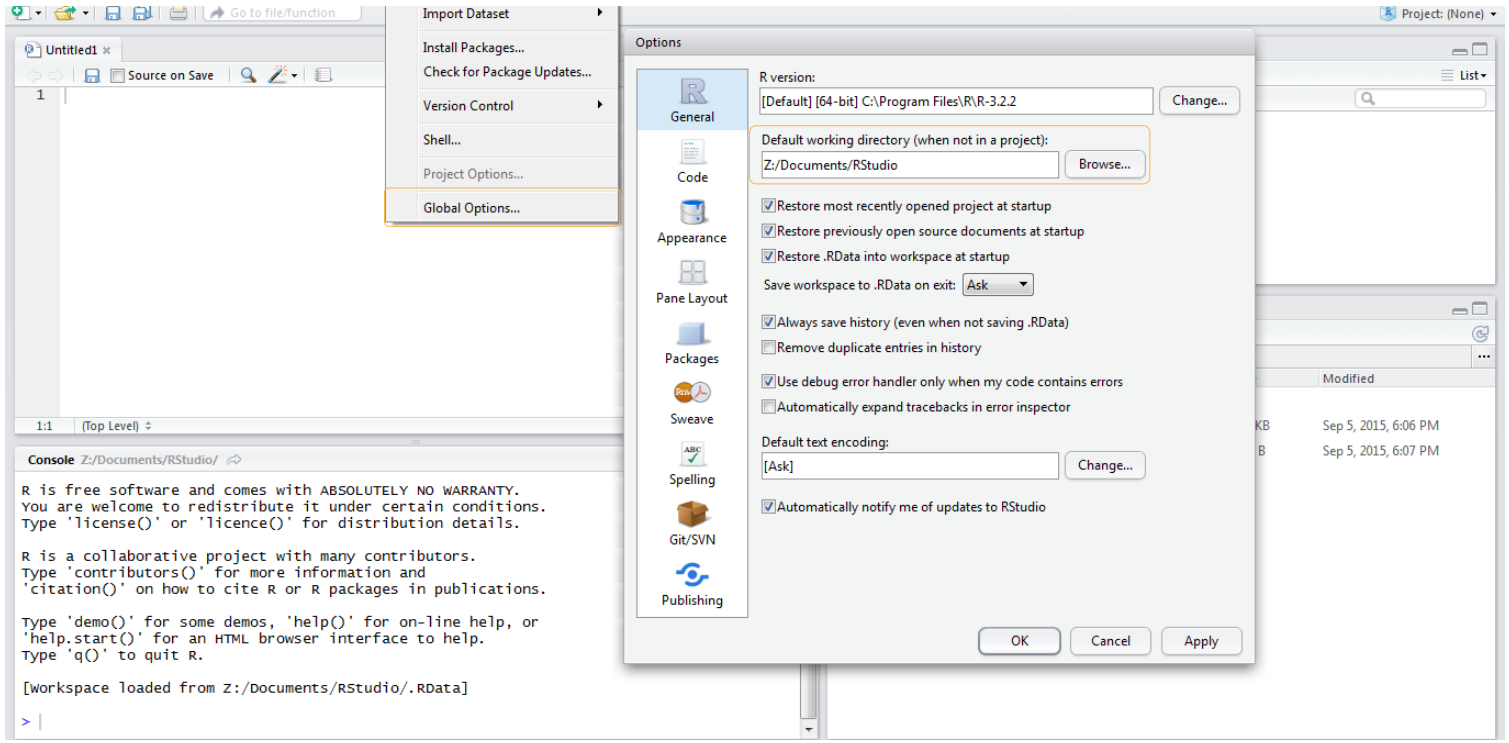
Introducing R Studio

- R Studio (version 0.99.442) console has four primary blocks:



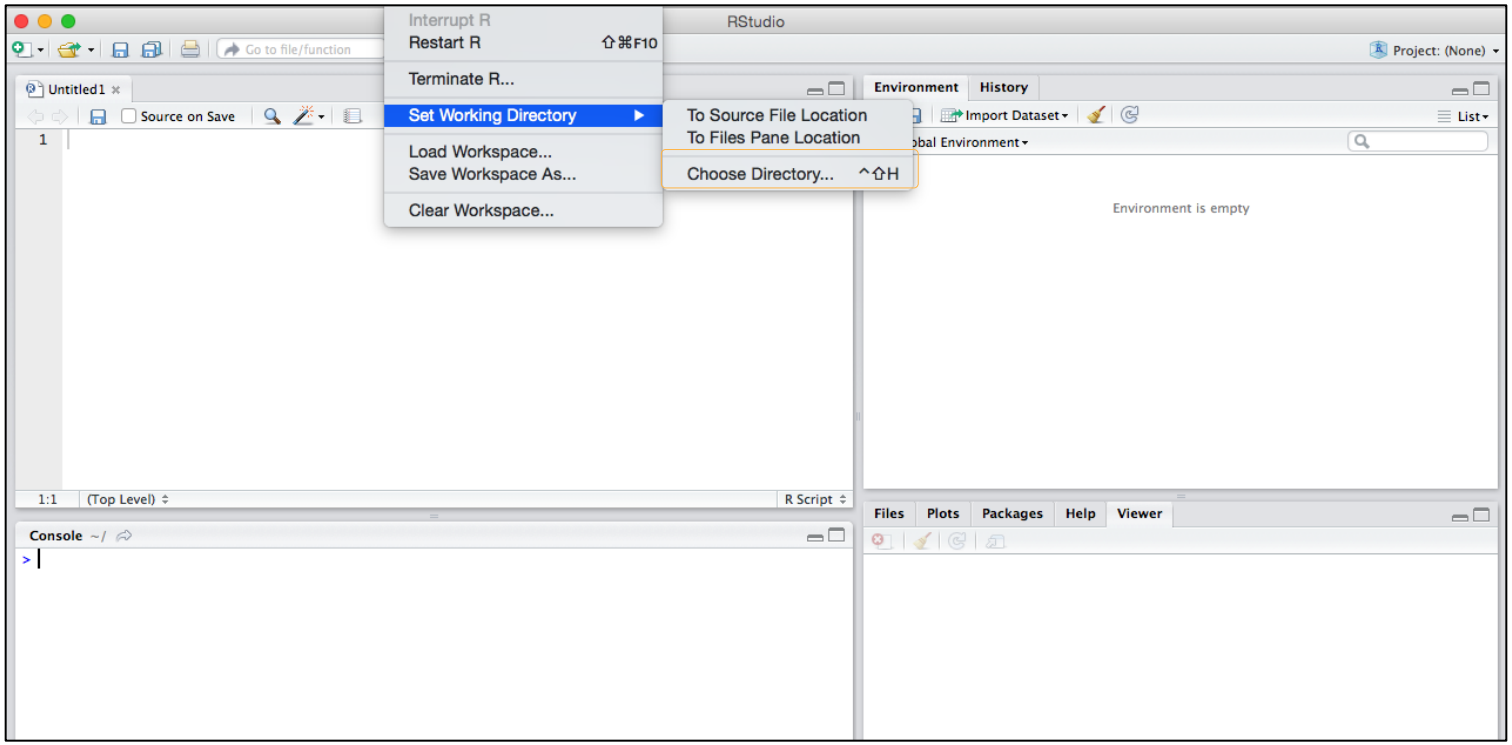
R Studio—Setting Global Working Directory

- Set the global working directory through the option below:



R Studio—Setting Local Working Directory

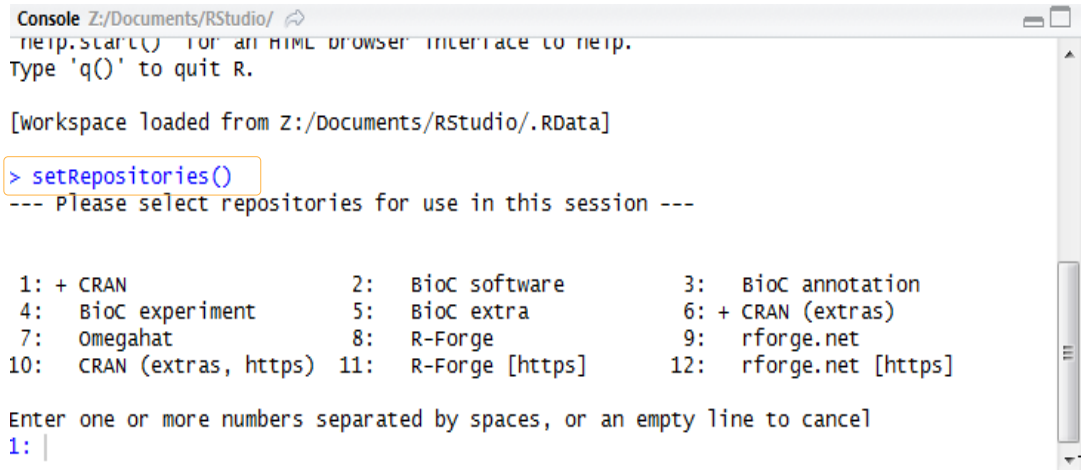
- Local working directory can be set as shown below. Two useful commands: getwd() and setwd()



Choosing a CRAN Mirror and Setting Repositories

- CRAN mirrors contain R packages that can extend the functionality of R.
- Choose a mirror located close to you as that will most likely give you the fastest downloads
- Repositories host the packages. Some of the examples of repositories are:

CRAN, CRANextra, BioCsoft, BioCann, BioCexp, BioCext, Omegahat, R-Forge and rforge.net

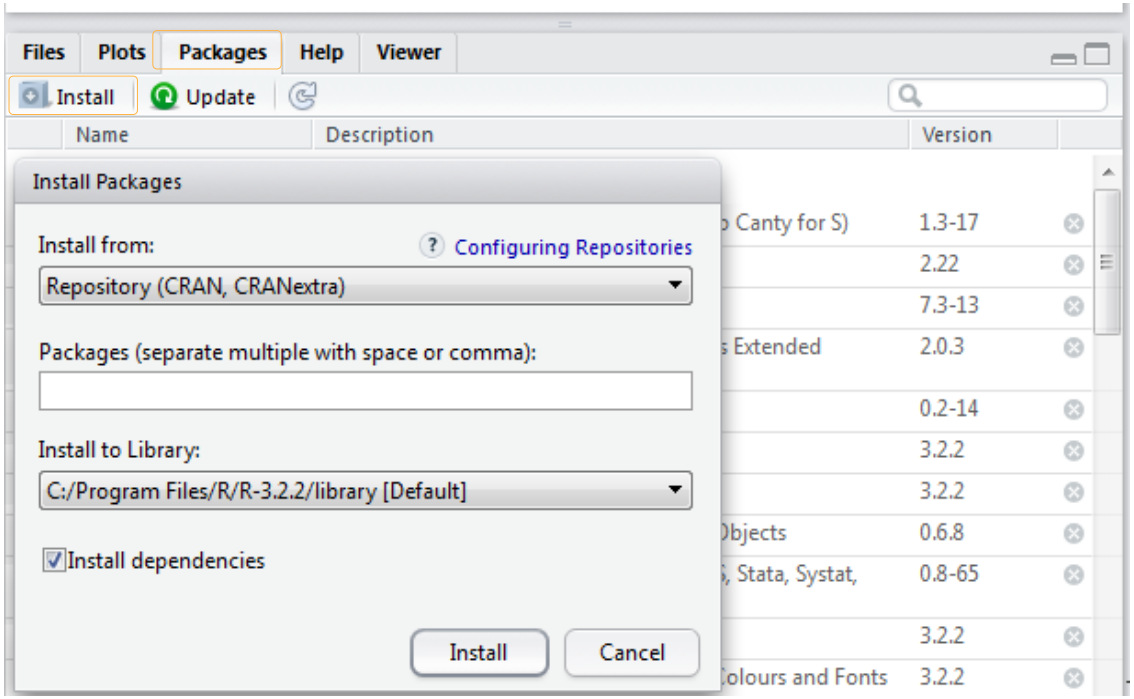


- Use this code to set your repositories: `setRepositories()`

Package TAB

Packages can be installed as and when required.

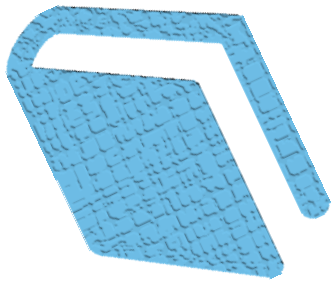
- The Packages tab lets you see what packages are installed
- A package must be loaded before you can use it
  - In Rstudio, this is accomplished by clicking the checkbox next to the package name in the package tab



This demo will show the R Studio features.

Summary

Summary of the topics covered in this lesson:



- R is a dialect of the S language. S is a language that was developed by John Chambers and others at Bell Labs.
- R comes with many built in dataset which can be used to practice the analytical concepts.

QUIZ TIME

© Copyright 2015 All rights reserved.

Quiz Question 1

Quiz 1

Which of the following is true about R?

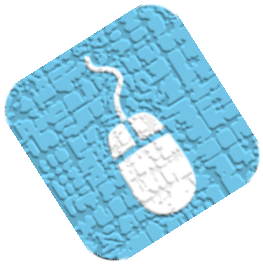
Select all that apply.

a. *Runs on almost any standard computing platform/OS*

b. *R has a very active and vibrant user community*

c. *R has very sophisticated graphics capabilities; better than most stat packages*

d. *User community does not provide help on R*



Quiz Question 1

Quiz 1

Which of the following is true about R?

Select all that apply.

a. *Runs on almost any standard computing platform/OS*

b. *R has a very active and vibrant user community*

c. *R has very sophisticated graphics capabilities; better than most stat packages*

d. *User community does not provide help on R*

Correct answer is: ***a , b & c***

All the options are correct except d. User community is helpful.

Quiz Question 2

Quiz 2

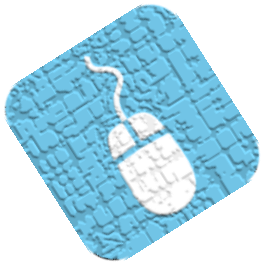
Which of the following is a name of the repository from where packages can be downloaded for use in R? *Select all that apply.*

a. *CRAN*

b. *Cranberry*

c. *R-Forge*

d. *Cran soft*



Quiz Question 2

Quiz 2

Which of the following is a name of the repository from where packages can be downloaded for use in R? *Select all that apply.*

a. *CRAN*

b. *Cranberry*

c. *R-Forge*

d. *Cran soft*

Correct answer is: b and d are not the name of repositories in R.

a & c

End of Lesson01–Overview of R



R is an official part of the Free Software Foundation's GNU project.
RStudio and Shiny are affiliated projects of the Foundation for Open Access Statistics.