Introduction to R / R Studio

MATH 6312

Department of mathematics, UTA

What is R?

- Programming Language S = developed at Bell Labs for statistics, simulation, graphics (Becker and Chambers, 1984)
  - S-PLUS: commercial implementation
  - R: implementation under GPL (GNU General Public License), open source.
- For R related tutorials and/or resources see the following links:
  - https://cran.r-project.org/doc/manuals/R-intro.pdf
  - https://github.com/berkeley-scf

Why R?

- R is widely used and has the widest statistical functionality. Wide usage helps to improve quality and reduce bugs.
- R is free and available on all major platforms
- Users add functionality via packages all the time, and access to fast growing number of packages
  - CRAN (>3500 packages) general data analysis
  - Bioconductor (>600 packages) bioscience data analysis
- As a scripting language, R is very powerful, flexible, and easy to use
- Efficient functions and data structures for data analysis; Powerful graphics.
- Standard for data mining and biostatistical analysis
- R is built on C and can call user-written and external C code and packages

Why not R?

- Other software is better than R at various tasks.
  - Python is very good for interacting with the operating system, and as a glue for tying together various applications/software in a workflow.
  - R can be much slower than compiled languages.
  - R's packages are only as good as the person who wrote them.
- Other programming languages
  - Python
  - Matlab/Octave
  - Julia
  - C/C++
  - SAS

R Studio

- R Studio allows the user to run R in a more user-friendly environment.
  - Open-source and available at https://www.rstudio.com
  - R Studio desktop can be installed in your local machine. R Studio server can be accessed using your web browser.
  - Nice GUI for
    - File management/navigation, code highlighting, auto completion, R package management/development, version control, debugging, etc.

- R Markdown is an extension to the Markdown markup language that makes it easy to write HTML in a simple plain text format.
  - http://rmarkdown.rstudio.com/
- R Sweave enables integration of R code into LaTeX documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change.
  - http://yihui.name/knitr/
- Version Control with Git and SVN
  - https://support.rstudio.com/hc/en-us/articles/200532077-Version-Control-with-Git-and-SVN/
- R Studio cheat sheet
  - https://www.rstudio.com/resources/cheatsheets/
- R for Matlab/Octave users
  - http://mathesaurus.sourceforge.net/octave-r.html

RStudio

Source: where you create/open/write your scripts
To run one line: C-RET
To run everything in source: C-S-RET

Files: all the files and folders in your default workspace
Plots: where you see all graphs.
Packages: a list of packages or add-ons.

Environment: where you see active objects.
History: a list of commands you used so far.

Console: where you input commands and see outputs

RStudio

Go to file/function

Project: (None)

MyScript.R

Source on Save

Run

Source

```
1  A = diag(3)
2  B = matrix(1:9,3,3)
3  y = array(5:7,c(3,1))
```

3:22 (Top Level)

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > Dropbox > STAT5474 > lab

Name	Size	Modified
..		
.Rhistory	304 B	Jan 26, 2016, 7:09 PM
	53 B	Jan 26, 2016, 2:14 AM

Environment | History

Import Dataset

Global Environment

Data

A	num [1:3, 1:3] 1 0 0 0
B	int [1:3, 1:3] 1 2 3 4
y	int [1:3, 1] 5 6 7

Options

General
Code
Appearance
Pane Layout
Packages
Sweave
Spelling
Git/SVN
Publishing

Choose the layout of the panes in RStudio by selecting from the controls in each quadrant.

Source

Files, Plots, Packages, Help, Vi

- [ ] Environment
- [ ] History
- [x] Files
- [x] Plots
- [x] Packages
- [x] Help
- [ ] Build
- [ ] VCS
- [x] Viewer

Environment, History, Build, VC

- [x] Environment
- [x] History
- [ ] Files
- [ ] Plots
- [ ] Packages
- [ ] Help
- [x] Build
- [x] VCS
- [ ] Viewer

Console

OK | Cancel | Apply

Layout can be changed from tools -> global options -> pane layout

RStudio

Go to file/function

Project: (None)

MyScript.R × A ×

Filter

	V1	V2	V3
1	1	0	0
2	0	1	0
3	0	0	1

Showing 1 to 3 of 3 entries

Environment History

Import Dataset List

Global Environment

Data

A	num [1:3, 1:3] 1 0 0 0 1 0 0 0 1
B	int [1:3, 1:3] 1 2 3 4 5 6 7 8 9
y	int [1:3, 1] 5 6 7

Files Plots Packages Help Viewer

New Folder Delete Rename More

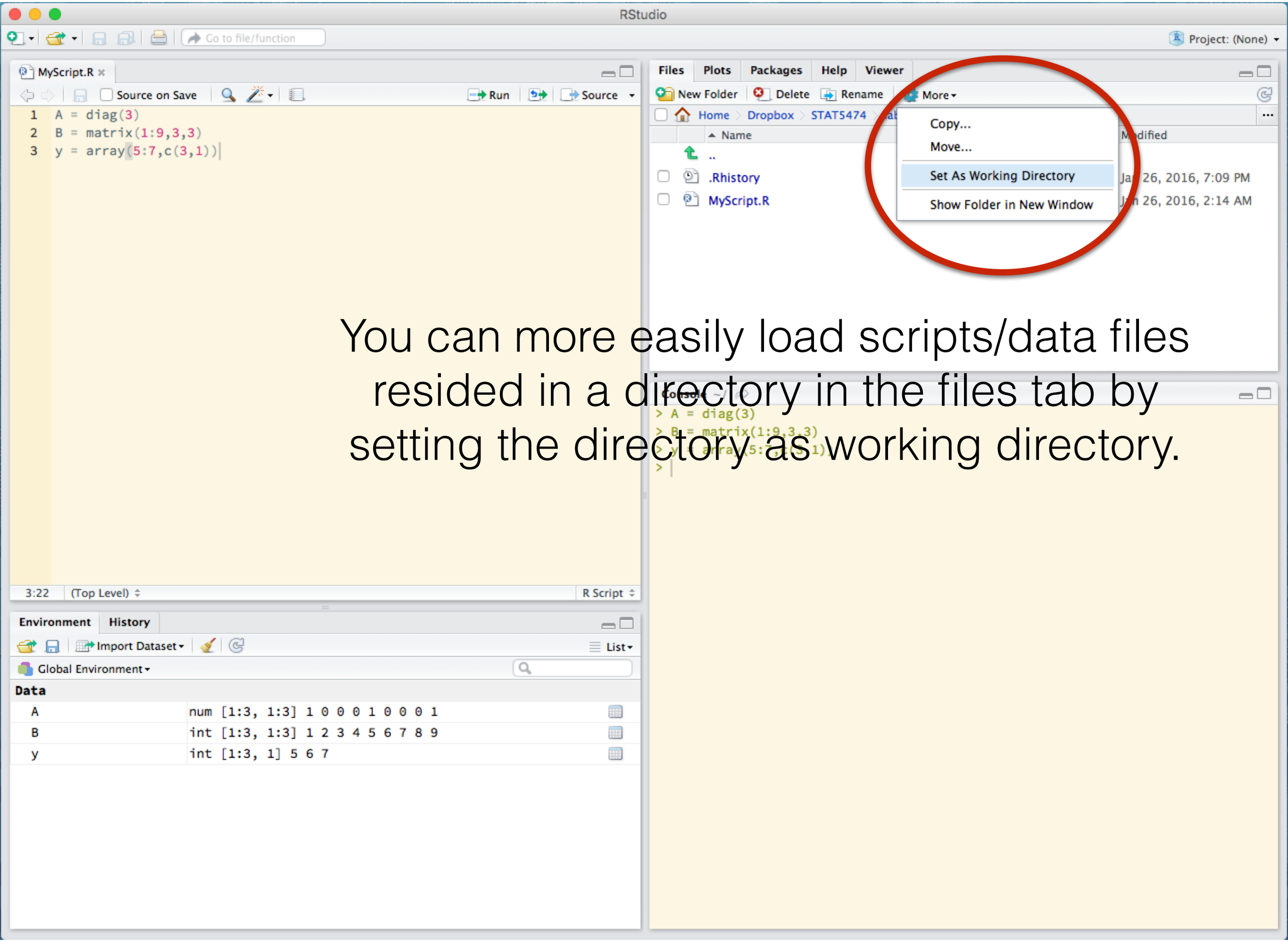
Home › Dropbox › STAT5474 › lab

Name	Size	Modified
..		
.Rhistory	304 B	Jan 26, 2016, 7:09 PM
MyScript.R	53 B	Jan 26, 2016, 2:14 AM

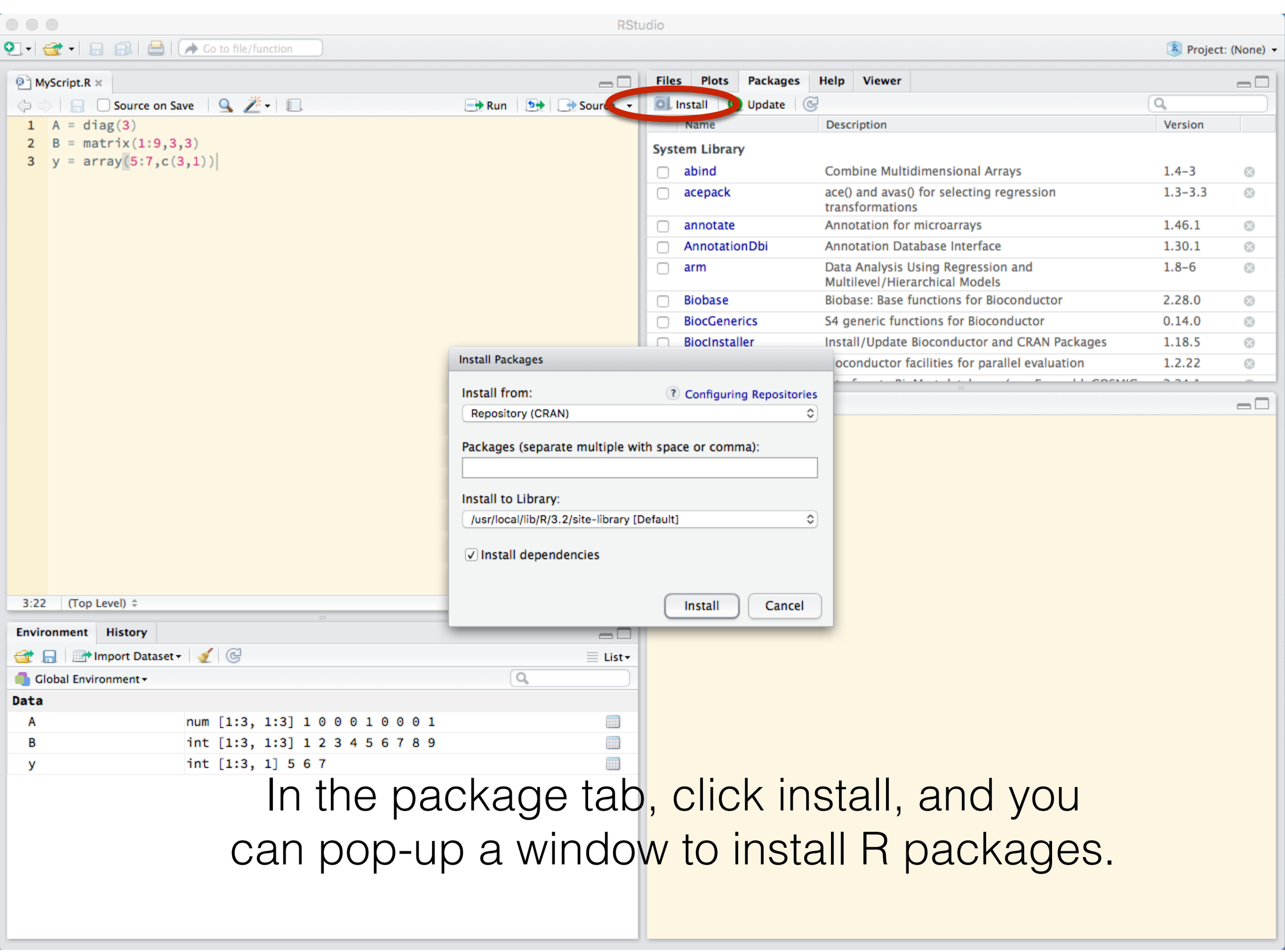
Console ~/

```
> A = diag(3)
> B = matrix(1:9,3,3)
> y = array(5:7,c(3,1))
> View(A)
>
```

click one of these, and you can see objects in the source window



You can more easily load scripts/data files resided in a directory in the files tab by setting the directory as working directory.



In the package tab, click install, and you can pop-up a window to install R packages.

RStudio

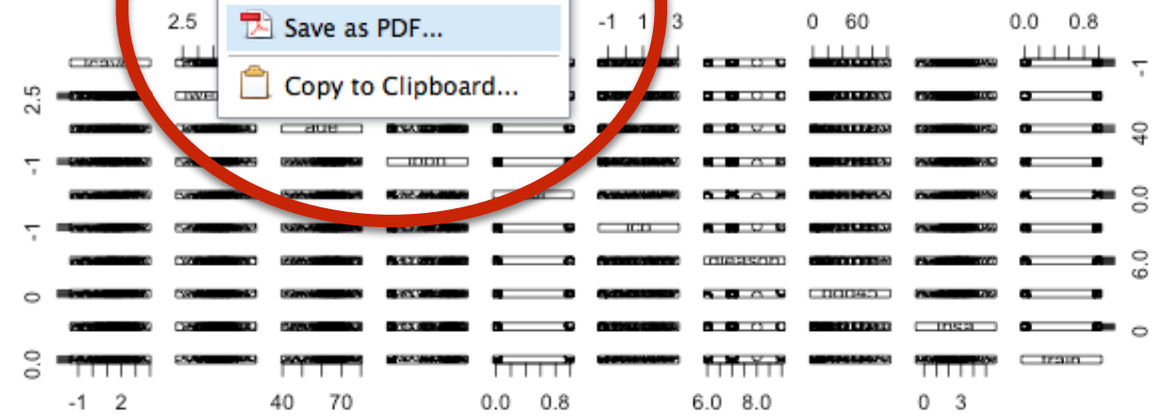
MyScript.R

```
1  A = diag(3)
2  B = matrix(1:9,3,3)
3  y = array(5:7,c(3,1))
```

Files | Plots | Packages | Help | Viewer

Export | Publish

Save as Image...
Save as PDF...
Copy to Clipboard...



Console ~/Dropbox/STAT5474/lab/

```
> prostate=read.table("prostate.txt")
> plot(prostate)
>
```

Environment | History

Import Dataset | List

Global Environment

Data	
A	num [1:3, 1:3] 1 0 0 0 1 0 0 0 1
B	int [1:3, 1:3] 1 2 3 4 5 6 7 8 9
prostate	97 obs. of 10 variables
y	int [1:3, 1] 5 6 7

From the plots tab, export -> save as image/ PDF lets you save a plot as a file

RStudio

Go to file/function

Project: (None)

MyScript.R

Source on Save

Run

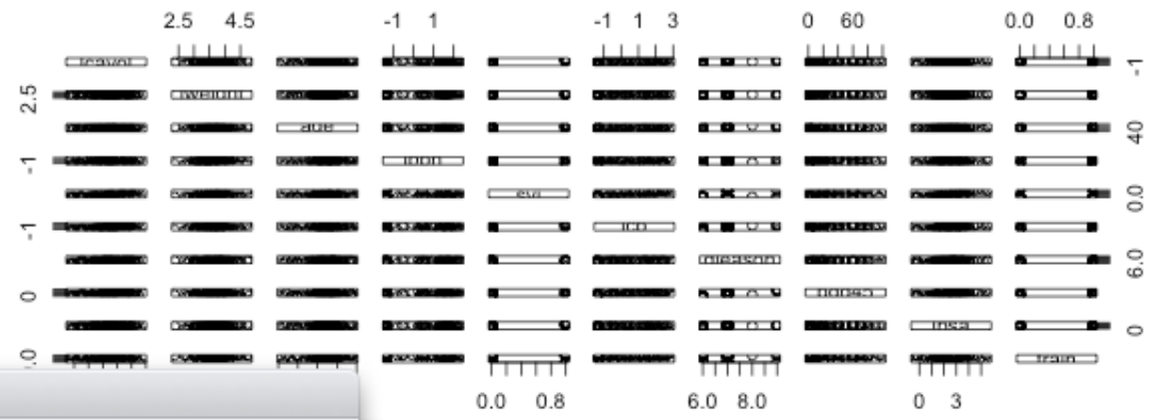
Source

```
1  A = diag(3)
2  B = matrix(1:9,3,3)
3  y = array(5:7,c(3,1))
```

3:22 (Top Level)

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



Save Plot as PDF

PDF Size: US Letter 8.50 x 11.00 inches

Orientation: Portrait / Landscape

Options: Use cairo_pdf device (requires X11)

Directory... ~/Dropbox/STAT5474/lab

File name: prostate

View plot after saving

Preview Save Cancel

Environment | History

Import Dataset List

Global Environment

Data

A	num [1:3, 1:3] 1 0 0 0 1 0 0 0 1
B	int [1:3, 1:3] 1 2 3 4 5 6 7 8 9
prostate	97 obs. of 10 variables
y	int [1:3, 1] 5 6 7

ab/

```
ostate.txt")
```

R Markdown

- R Markdown is a format that enables easy authoring of reproducible reports from R.

- It combines the core syntax of Markdown with embedded R code chunks that are run so their output can be included in the final document. i.e. you do not need to copy and paste plots or text from R to the final document

- R Markdown can produce HTML/PDF/WORD document, but it is optimized for HTML.

RStudio

MyScript.R | Rbasics.Rmd

Knit HTML | Run | Chunks

Knit HTML
Knit PDF
Knit Word
View in Pane
View in Window

```
1  ---
2  output: ht
3  ---
4  % STAT5474
5
6  # R as a calculator
7
8  ```{r computing,echo=TRUE,results='markup'}
9  2 + 2 # add numbers
10 2 * pi # multiply by a constant
11 7 + runif(1) # add a random number
12 3^4 # powers
13 sqrt(4^4) # functions
14 log(10)
15 log(100, base = 10)
16 ```
17
18 # Assigning values to R objects
19
20 A key action in R is to store values in the form of R objects, and to
   examine the value of R objects.
21
22 ```{r}
```

4:26 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > Dropbox > STAT5474 > lab

Name	Size	Modified
..		
.Rhistory	304 B	Jan 26, 2016, 7:09 PM
MyScript.R	105 B	Jan 27, 2016, 2:51 AM
prostate.txt	7 KB	Jan 27, 2016, 2:39 AM
Rbasics.Rmd	2 KB	Jan 27, 2016, 3:58 AM

Rmd is the extension of R Markdown files. You can open/compile Rmd files to generate the final document in R Studio.

Console | R Markdown

~/Dropbox/STAT5474/lab/

>

Environment | History

Import Dataset | List

Global Environment

Environment is empty