

A whirlwind tour of Rstudio, R, and Rmarkdown Magic

Jenny Rieck & Derek Beaton

May 29, 2019

The big outline

- ▶ Part 0: Background and Community
- ▶ Part 1: Rstudio
- ▶ Part 2: Project and environment setup
- ▶ Part 3: R (et al., eg Python)
- ▶ Part 4: Rmarkdown
- ▶ Part 5: Advanced R and beyond
- ▶ Part 6: A few of our favorite things

Part 0: Background and Community

- ▶ What this is & isn't; a bunch of things we aren't covering but you should be aware of
 - ▶ This is a taste and to bring you into a bigger world
- ▶ Centralization, standards
- ▶ Help
- ▶ Including rigor & reproducibility of packages
- ▶ The “tidyverse”
 - ▶ Learn it. But don't learn *only* the tidyverse; you'll be lost in base R

R Background

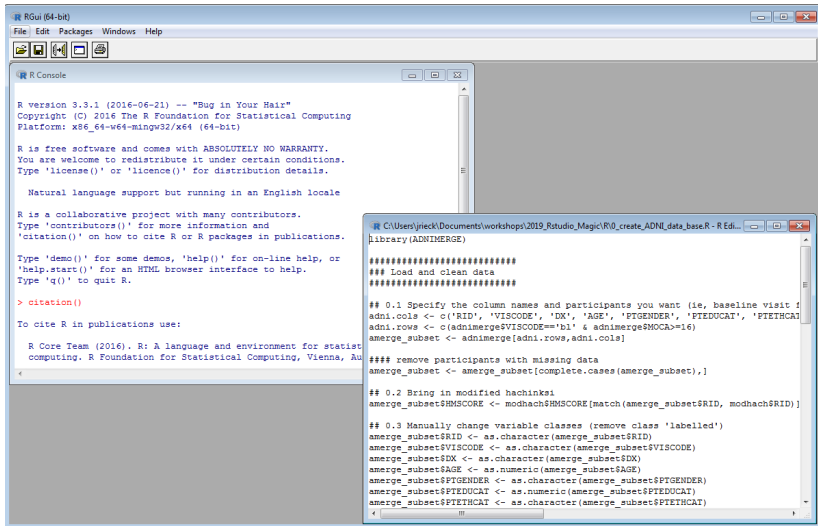
- ▶ Created in 1992 by Gentleman & Ihaka

[we] considered the problem of obtaining decent statistical software for our undergraduate Macintosh lab. After considering the options, we decided that the most satisfactory alternative was to write our own. [...] Finally we added some syntactic sugar to make it look somewhat like S. We call the result “R”.

What is R?

- ▶ R is for stats and general purpose programming
- ▶ R is a functional language
 - ▶ Turing complete – can do anything other languages can do
- ▶ R is an environment to interface with the language
 - ▶ Console based
 - ▶ Type in commands
 - ▶ no point-and-click
- ▶ R is a collection of tools
 - ▶ Pre-packaged software at your disposal
- ▶ R is free (as in beer and speech)
 - ▶ No cost, no restrictions

R is a bit ugly



The screenshot shows the R GUI (64-bit) interface. The main window has a menu bar (File, Edit, Packages, Windows, Help) and a toolbar. Below the menu bar is the R Console, which displays the R version information and welcome message. To the right of the console is a script editor window titled "C:\Users\jriek\Documents\workshops\2019_Rstudio_Magic\R0_create_ADNI_data_base.R - R Edit...". The script editor contains R code for loading and cleaning data, specifying column names, removing participants with missing data, and manually changing variable classes.

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> citation()

To cite R in publications use:

  R Core Team (2016). R: A language and environment for statisti
  computing. R Foundation for Statistical Computing, Vienna, Au
```

```
library(ADNIMERGE)

#####
### Load and clean data
#####

## 0.1 Specify the column names and participants you want (ie, baseline visit f
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT')
adni.rows <- c(adnimerge$VISCODE=="b1" & adnimerge$MOCHA==16)
amerge_subset <- adnimerge[adni.rows, adni.cols]

#### remove participants with missing data
amerge_subset <- amerge_subset[complete.cases(amerge_subset),]

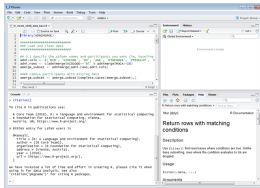
## 0.2 Bring in modified hachinks1
amerge_subset$HMScore <- modhach$HMScore[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
amerge_subset$RID <- as.character(amerge_subset$RID)
amerge_subset$VISCODE <- as.character(amerge_subset$VISCODE)
amerge_subset$DX <- as.character(amerge_subset$DX)
amerge_subset$AGE <- as.numeric(amerge_subset$AGE)
amerge_subset$PTGENDER <- as.character(amerge_subset$PTGENDER)
amerge_subset$PTEDUCAT <- as.numeric(amerge_subset$PTEDUCAT)
amerge_subset$PTETHCAT <- as.character(amerge_subset$PTETHCAT)
```

But R has many interfaces

- ▶ Today we focus on RStudio (MatLab-like)
- ▶ But see also Deducer, RCommander (SPSS-like)

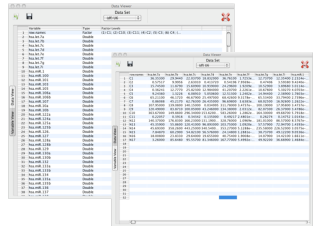
RStudio



RCommander



Deducer



R is a community (actually many communities!)

- ▶ Help and resources
- ▶ Package development and distribution

R: Help!

- ▶ <https://www.statmethods.net/>
- ▶ Online forums (Stack Exchange, r-lists)
- ▶ SpringerLink
 - ▶ All R books for free (pdf format) or for minimal cost (printed)
- ▶ Vignettes
 - ▶ step-by-step instruction guides for packages

R Packages

- ▶ Packages are bundles of code made by someone (or many people) for everyone to use
 - ▶ If you can think of a stats problem, there is a package for it
- ▶ Available primarily on CRAN
 - ▶ But also github, r-forge

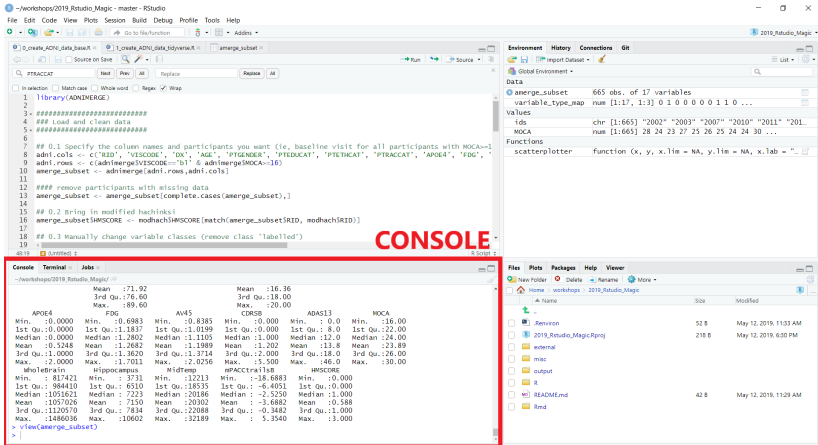
Tidyverse

- ▶ something here about tidy

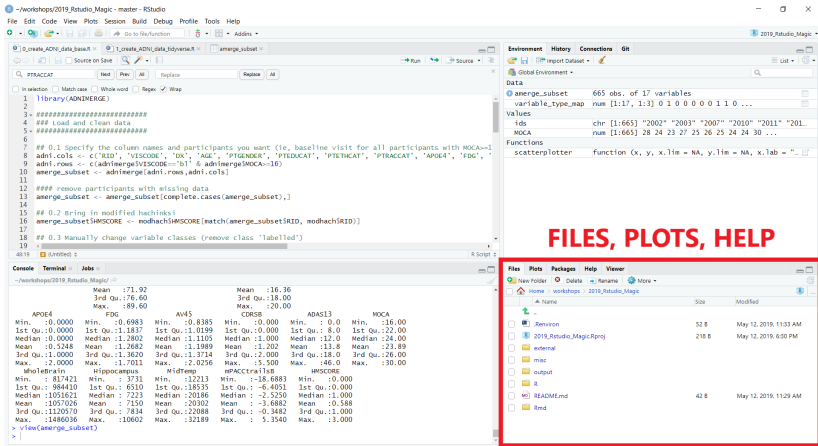
Part 1: RStudio

- ▶ Settings, a quick tour through stuff, features
- ▶ Examples on getting setup

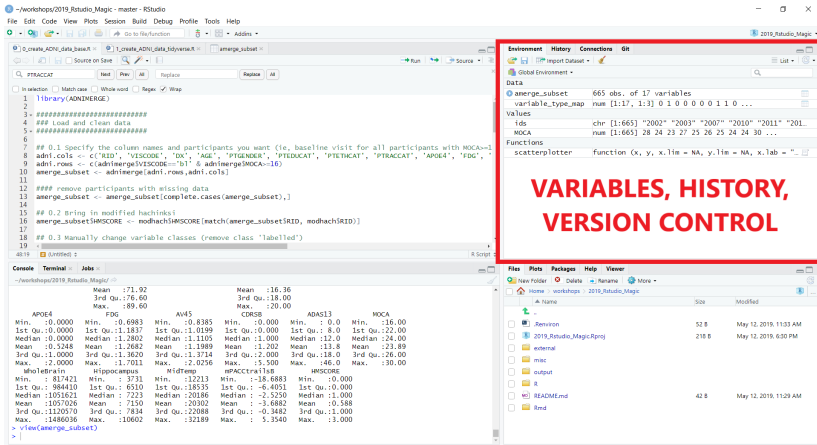
RStudio Environment



RStudio Environment



RStudio Environment



RStudio Environment

~/workshops/2019_Rstudio_Magic-master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

CODE

```
1 library(ADNIMERGE)
2
3 #####
4 ## Load and clean data
5 #####
6
7 ## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA=1
8 adni.cols <- c("RID", "VISCODE", "DX", "AGE", "PTGENDER", "PTEDUCAT", "PTETHCAT", "PTRACCAT", "APOE4", "FDG",
9 adni.rows <- c(adnimerge$VISCODE=="b1" & adnimerge$MOCA==16)
10 amerge_subset <- adnimerge[adni.rows, adni.cols]
11
12 ##### remove participants with missing data
13 amerge_subset <- amerge_subset[complete.cases(amerge_subset),]
14
15 ## 0.2 Bring in modified hachinks1
16 amerge_subset$HSCORE <- modhach1$HSCORE[match(amerge_subset$RID, modhach1$RID)]
17
18 ## 0.3 Manually change variable classes (remove class 'labelled')
19
20
```

Environment History Connections Git

Global Environment +

Data

amerge_subset 665 obs. of 17 variables

variable_type_map

variable_type_map	num	[1:17, 1:3]	0 1 0 0 0 0 1 1 0 ...
-------------------	-----	-------------	-----------------------

Values

ids	chr	[1:665]	"2002" "2003" "2007" "2010" "2011" "201...
MOCA	num	[1:665]	28 24 23 27 25 26 25 24 24 30 ...

Functions

scatterplot	function	(x, y, x.lim = NA, y.lim = NA, x.lab = "...")
-------------	----------	-----------------------------------------------

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home workshops 2019_Rstudio_Magic

Name	Size	Modified
2019_Rstudio_Magic.Rproj	218 B	May 12, 2019, 6:50 PM
2019_Rstudio_Magic.Rproj	218 B	May 12, 2019, 6:50 PM
external		
output		
R		
README.md	42 B	May 12, 2019, 11:29 AM
Rmd		

Console Terminal Jobs

```
~/workshops/2019_Rstudio_Magic/ >
> view(amerge_subset)
>
```

Mean :71.92 Min. :0.0000 1st Qu.:1.1837 Median :1.2802 Mean :0.5248 3rd Qu.:1.0000 Max. :2.0000

FDG

Mean :176.60 1st Qu.:1.6983 Median :1.2802 Mean :0.5248 3rd Qu.:1.0000 Max. :2.0000

AV45

Mean :10.36 1st Qu.:1.0199 Median :1.1105 Mean :1.1989 3rd Qu.:1.3714 Max. :2.0256

CDRSB

Mean :18.00 1st Qu.:0.0000 Median :1.0000 Mean :11.202 3rd Qu.:2.0000 Max. :5.500

ADAS13

Mean :18.00 1st Qu.:8.0 Median :12.0 Mean :13.8 3rd Qu.:18.0 Max. :46.0

MOCA

Mean :22.00 1st Qu.:22.00 Median :24.00 Mean :23.89 3rd Qu.:26.00 Max. :30.00

Hippocampus

Mean :12213 Min. :18.6883 1st Qu.:6510 Median :7223 Mean :7150 3rd Qu.:7834 Max. :110602

MidTemp

Mean :18535 1st Qu.:18535 Median :20186 Mean :20302 3rd Qu.:22088 Max. :32189

mPACCtra11sb

Mean :-6.4051 1st Qu.:-6.4051 Median :-2.5250 Mean :-3.6882 3rd Qu.:0.3482 Max. :5.3540

HSCORE

Mean :0.0000 1st Qu.:0.0000 Median :1.0000 Mean :0.588 3rd Qu.:1.0000 Max. :3.000

RStudio Environment

The screenshot displays the RStudio interface. The main window shows a data viewer for a subset of data from the '2019_RatJudo_Magic' dataset. The data is presented in a table with columns: DX, AGE, PTENDER, PTEDUCAT, PTHCAT, PTRACCAT, APOE4, FDG, AV45, CDRSB, ADAS13, MOCA, and WholeBrain. The data is filtered to show 13 of 663 entries. The R console shows the command 'view(amerge_subset)'.

DX	AGE	PTENDER	PTEDUCAT	PTHCAT	PTRACCAT	APOE4	FDG	AV45	CDRSB	ADAS13	MOCA	WholeBrain
2002	64.8	Male	18	Not Hsp/Latino	White	0	1.2061908	0.9784523	2.5	4	28	11335566
2003	63.6	Female	18	Not Hsp/Latino	White	0	1.2899626	1.1646574	2.0	11	24	1070369.5
2007	83.4	Female	20	Hsp/Latino	White	0	1.3038182	1.4492659	2.5	9	23	9207101
2010	62.9	Female	20	Not Hsp/Latino	Other	1	1.3121151	1.4712845	0.5	6	27	968402.9
2011	69.9	Female	20	Not Hsp/Latino	White	0	1.4537199	1.0537398	1.5	7	25	987622.5
2018	76.4	Female	18	Not Hsp/Latino	White	0	1.3148491	1.0525191	1.5	10	26	1004817.0
2020	66.0	Male	18	Not Hsp/Latino	Other	1	1.2093170	1.3159914	1.5	6	25	1133661.8
2027	61.9	Female	14	Not Hsp/Latino	White	0	1.4030646	1.0299761	1.0	6	24	969957.1
2031	72.5	Male	14	Not Hsp/Latino	White	0	1.3404430	0.9959887	2.0	10	24	1059676.5
2036	66.7	Female	14	Not Hsp/Latino	White	0	1.2892590	1.0500795	1.0	5	30	1019101.0
2037	75.8	Male	16	Not Hsp/Latino	White	1	1.3074956	1.4389912	0.5	20	20	1104797.1
2042	69.5	Male	20	Not Hsp/Latino	White	0	1.2083130	1.0555846	1.5	18	23	1061388.4
2043	72.2	Female	20	Not Hsp/Latino	White	1	1.3761156	1.2040191	2.0	8	27	1023110.0

Benefits of RStudio

- ▶ Built-in integration with version control (git or SVN)
- ▶ Package and documentation generation
- ▶ Reproducible science!
 - ▶ R Markdown documents
 - ▶ Save and execute code
 - ▶ Generate high quality reports that can be shared
 - ▶ Create presentations (like this one!)
 - ▶ Even write papers

RStudio Setup

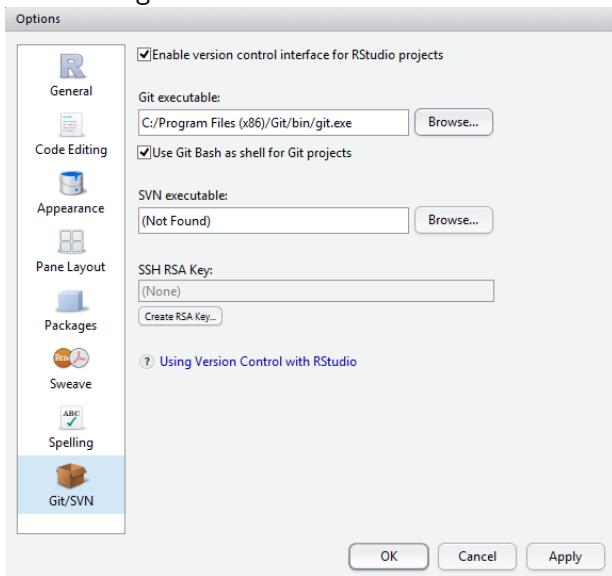
- ▶ Download R and Rstudio
- ▶ Add-on packages

```
#to install from CRAN  
install.packages('devtools', dependencies = TRUE)  
#to install from a file  
install.packages('/mypath/to/package/ADNIMERGE.tar.gz',  
                 type='source', repos=NULL)  
#to install from a git (requires the devtools package)  
dev.tools::install_github(Gibbsdavidl/CatterPlots)
```

- ▶ See <https://jennybc.github.io/2014-05-12-ubc/r-setup.html>
for a detailed guide

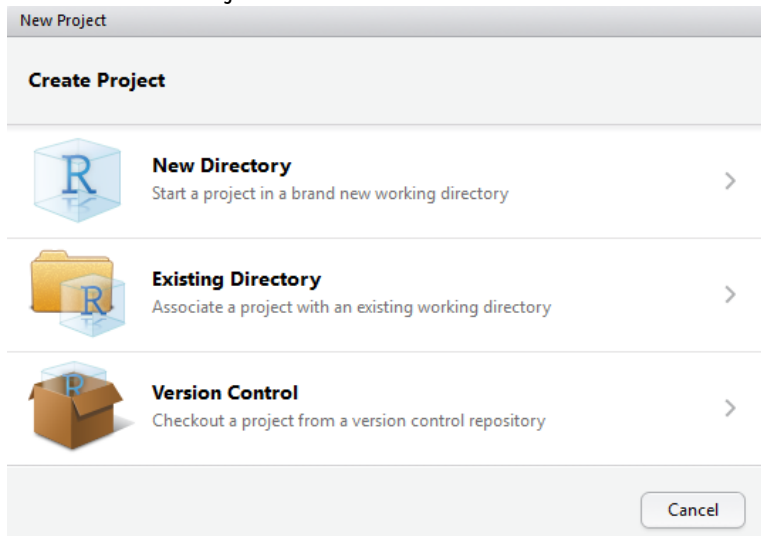
Rstudio Setup: Projects & Git

► Download git and link to RStudio



Rstudio Setup: Projects & Git

► Create a New Project File

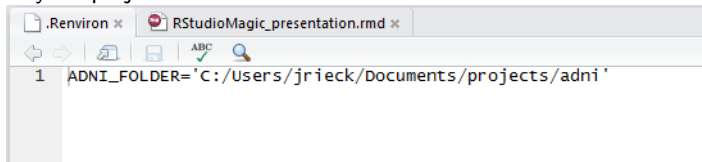


Format .gitignore

- ▶ File types to ignore:
 - ▶ .Rproj.user
 - ▶ .Rhistory
 - ▶ .Ruserdata
 - ▶ .Renvirom
 - ▶ .rda & .Rdata (to avoid pushing potentially sensitive data files to git)
 - ▶ ** before each extensions will match directories anywhere in the repo

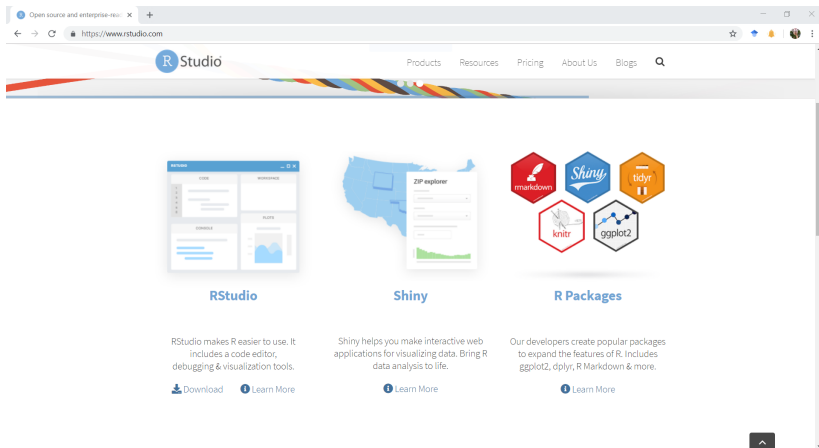
Format environmental variables

- ▶ Set environmental variables (ie, directory location of data) to make code generalizable across computers
- ▶ In your project folder create a .Renviron file and define variables



The screenshot shows an RStudio editor window with two tabs: '.Renviron *' and 'RStudioMagic_presentation.rmd *'. The '.Renviron' tab is active, displaying a single line of code: `ADNI_FOLDER='C:/Users/jriECK/Documents/projects/adni'`. The line is numbered '1' on the left margin. The editor's toolbar includes icons for navigation, saving, and searching.

RStudio Resources



The screenshot shows the RStudio website homepage. At the top is a navigation bar with the RStudio logo and links for Products, Resources, Pricing, About Us, and Blogs. Below the navigation bar is a large section with three main features highlighted: RStudio, Shiny, and R Packages. Each feature has a representative image, a title, a description, and a link to learn more or download. At the bottom right, there is a small black button with a white upward arrow.

Open source and enterprise-ready | <https://www.rstudio.com>

Products Resources Pricing About Us Blogs

RStudio

RStudio makes R easier to use. It includes a code editor, debugging & visualization tools.

[Download](#) [Learn More](#)

Shiny

Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life.

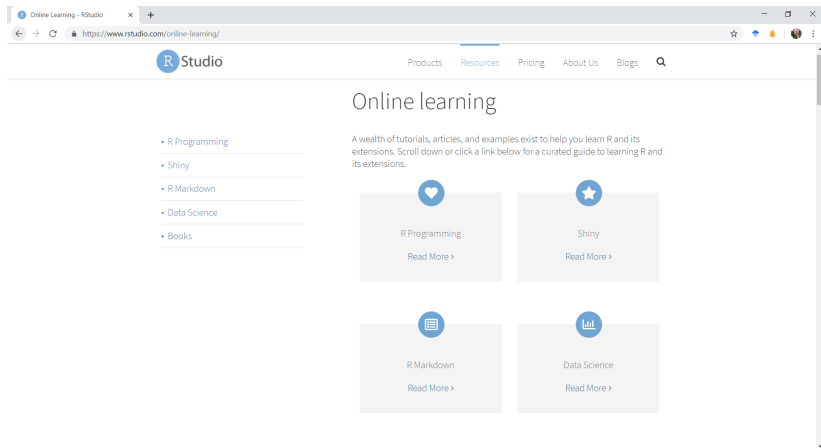
[Learn More](#)

R Packages

Our developers create popular packages to expand the features of R. Includes ggplot2, dplyr, R Markdown & more.

[Learn More](#)

RStudio Resources



The screenshot shows the RStudio website's 'Online Learning' section. The browser's address bar displays 'https://www.rstudio.com/online-learning/'. The website's navigation bar includes links for 'Products', 'Resources' (which is highlighted), 'Pricing', 'About Us', and 'Blogs', along with a search icon. On the left side, there is a vertical list of links: 'R Programming', 'Shiny', 'R Markdown', 'Data Science', and 'Books'. The main content area is titled 'Online learning' and contains a paragraph stating: 'A wealth of tutorials, articles, and examples exist to help you learn R and its extensions. Scroll down or click a link below for a curated guide to learning R and its extensions.' Below this text are four cards arranged in a 2x2 grid. Each card features a blue circular icon at the top, a title, and a 'Read More >' link. The cards are: 1) 'R Programming' with a heart icon, 2) 'Shiny' with a star icon, 3) 'R Markdown' with a document icon, and 4) 'Data Science' with a bar chart icon.

Online Learning - RStudio

https://www.rstudio.com/online-learning/


RStudio

Products Resources Pricing About Us Blogs

Online learning


A wealth of tutorials, articles, and examples exist to help you learn R and its extensions. Scroll down or click a link below for a curated guide to learning R and its extensions.

- R Programming
- Shiny
- R Markdown
- Data Science
- Books




R Programming

[Read More >](#)




Shiny

[Read More >](#)



R Markdown

[Read More >](#)



Data Science

[Read More >](#)

RStudio Resources

Cheatsheets - RStudio

https://www.rstudio.com/resources/cheatsheets/

RStudio

ProductsResourcesPricingAbout UsBlogs

RStudio Cheat Sheets

The cheat sheets below make it easy to learn about and use some of our favorite packages. From time to time, we will add new cheat sheets to the gallery. If you'd like us to drop you an email when we do, let us know by clicking the button to the right.

SUBSCRIBE TO CHEAT SHEET UPDATES HERE

- RStudio IDE
- R Markdown
- Shiny
- Package Development

- Data Import
- Data Transformation with dplyr
- Data Visualization with ggplot2
- Apply functions with purrr

- Deep Learning with Keras
- Data Science in Spark with Sparklyr
- String manipulation with stringr
- Dates and times with lubridate

Python with R and Retiulate Cheat Sheet

The reticulate package provides a comprehensive set of tools for interoperability between Python and R. With reticulate, you can call Python from R in a variety of ways including importing Python modules into R scripts, writing R Markdown Python chunks, sourcing Python scripts, and using Python interactively within the RStudio IDE. This cheatsheet will remind you how. Updated 4/19.

Use Python with R with reticulate :: CHEAT SHEET

Python in R Markdown

Python in R code

Object Conversion

Modules