

# Reproducible (and collaborative) science through RStudio

A whirlwind tour with R, RMarkdown, Python, LaTeX, and more

Jenny Rieck & Derek Beaton

May 20 2019

# The big outline

- ▶ Part 0: Background
- ▶ Part 1: A bit about R
- ▶ Part 2: RStudio & Project setup
- ▶ Part 3: R, RMarkdown, & more
- ▶ Part 4: Advanced, beyond, & our favorites

## Part 0: Background

To dive right in

If you want to skip over the background & RStudio, go straight to  
**Part 2: RStudio & Project setup**

# Background

- ▶ This is a taste and to bring you into a bigger world
  - ▶ R, Python, SQL, and JavaScript are critical data science tools/languages
- ▶ R (language and community) strongly emphasizes
  - ▶ Centralization & standards
  - ▶ Rigor & reproducibility (packages, RMarkdown)
- ▶ An interesting language
  - ▶ Functional
  - ▶ With a sublanguage (or dialect?): the tidyverse

## R is a community (actually many communities!)

- ▶ Help and resources
- ▶ Package development and distribution
- ▶ An ideal example
  - ▶ Not quite always that way
  - ▶ Strong communal presence

## R: Help!

- ▶ So many websites e.g., <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
- ▶ Git
  - ▶ With open books (via bookdown)
- ▶ Twitter #rstats
- ▶ RStudio (website)
  - ▶ Videos, cheat sheets

# R Packages

- ▶ Packages are bundles of code made by someone (or many people) for everyone to use
  - ▶ There are packages for everything
  - ▶ We'll cover some of the diversity throughout
- ▶ Comprehensive & Reproducible
- ▶ Available primarily on CRAN
  - ▶ But also github (less so: r-forge)

## Part 1: A bit about 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 general purpose programming
  - ▶ Design around & for statistics
  - ▶ “for and by statisticians”
- ▶ R is a collection of tools
  - ▶ Pre-packaged software at your disposal
- ▶ R is free (as in beer and speech)
  - ▶ No cost, no restrictions
  - ▶ E.g., Microsoft (nee Revolution) R
- ▶ R is a functional language
  - ▶ Turing complete
  - ▶ Mathematical functions
  - ▶ Pass expressions and functions to and from functions

## R: Data types

- ▶ Stored as *vectors*
  - ▶ see `class()`
- ▶ numeric
  - ▶ real or decimal
  - ▶ Includes `NaN`, `Inf`, `-Inf`
- ▶ integer
- ▶ complex
- ▶ character
- ▶ logical
  - ▶ includes `NA`, `TRUE`, `FALSE`
- ▶ factor
  - ▶ factors are usually not your friends
  - ▶ generally: `stringsAsFactors = F` or convert these

## R: factor disasters

```
a_numeric_vector <- c(3, 0, 1, -2, 2, 5, 5, 2, 1)
(a_numeric_vector + 1)

## [1] 4 1 2 -1 3 6 6 3 2

(a_numeric2factor_vector <- as.factor(a_numeric_vector))

## [1] 3 0 1 -2 2 5 5 2 1
## Levels: -2 0 1 2 3 5

(as.numeric(a_numeric2factor_vector))

## [1] 5 2 3 1 4 6 6 4 3

(as.numeric(as.character(a_numeric2factor_vector)))

## [1] 3 0 1 -2 2 5 5 2 1
```

## R: Data structures

- ▶ Starts counting from 1
  - ▶ Not 0
- ▶ vector[1]
- ▶ matrix[1,1]
- ▶ array[1,1,1]
- ▶ list[[1]]
  - ▶ Can contain mixtures of types
  - ▶ or list\$name
- ▶ data.frame:
  - ▶ Is technically a list but access in three ways
  - ▶ data.frame[[1]][1]
  - ▶ data.frame[1,1]
  - ▶ data.frame\$name
  - ▶ tibbles: tidyverse data.frames

R

Some more about R here... Matlab cheat sheet Other cheat sheets

## Tidyverse

- ▶ something here about tidy
- ▶ Learn it. But don't learn *only* the tidyverse; you'll be lost in base R

## Part 2: RStudio & Project setup

# RStudio

- ▶ IDE: Integrated development environment
- ▶ RStudio: Does so much
  - ▶ We scratch the surface here
- ▶ Quick walk through
- ▶ Followed by specific set up
  - ▶ Generally, but
  - ▶ Also for this workshop

# RStudio Environment

The screenshot shows the RStudio interface with several windows open:

- Script Editor:** Displays the R script `0_create_ADNI_data.R`. The code performs data cleaning and merging, including handling missing data and matching across different datasets.
- Console:** Shows the output of the script execution.
- File Browser:** Shows the project structure with files like `0_create_ADNI_data.R`, `1_load_ADNI_data_idiverse.R`, and `2_merge_subset.R`.
- Environment:** Shows the current environment variables and objects.

# RStudio Environment

The screenshot displays the RStudio interface with several windows open:

- Code Editor:** Shows a script named `create_ADNI_data.R` containing R code for data manipulation. The code includes library imports, data loading, merging datasets, and manual variable class changes.
- Console:** Shows the output of the R code execution. It includes descriptive statistics for various variables like APOE4, FDG, AV45, CDRSB, ADAS13, and MOCA across different brain regions (Wholebrain, Hippocampus, Midtemp, nPACCtailB, and MSMScore).
- Environment:** Shows the global environment with objects like `anmerge_subset` (665 obs. of 17 variables), `ids` (chr vector), and `MOCA` (num vector).
- File Browser:** Shows the project structure with files like `README.md`, `script.R`, and `output`.

# RStudio Environment

The screenshot shows the RStudio interface with the following components:

- Script Editor:** Displays the R script `create_ADNI_data.R`. The script performs several steps:
  - Imports required packages: `tidyverse`, `lapply`, `data.table`, and `stringr`.
  - Creates a dataset `PTRACCAT` from `ADNI_data`.
  - Specifies column names and participants.
  - Loads and cleans data.
  - Removes participants with missing data.
  - Brings in modified hachimaki functions.
  - Manually changes variable classes.
- Console:** Shows statistical summaries for various variables like APOE4, FDG, AV45, CDRSB, ADAS13, MOCA, and hippocampus.
- File Browser:** Shows the directory structure of the workspace:
  - Home
  - workshops > 2019\_Rstudio\_Magic
  - |- .Renviron
  - |- 2019\_Rstudio\_Magic.Rproj
  - |- external
  - |- misc
  - |- output
  - |- R
  - |- README.md
  - |- Rmd
- Header Bar:** Shows tabs for Environment, History, Connections, and Git, along with a search bar and a status bar indicating the current project is `2019_Rstudio_Magic`.

**FILES, PLOTS, HELP**

# RStudio Environment

The screenshot shows the RStudio interface with several windows open:

- Code Editor:** Shows R code for creating an ADNI data subset. The code includes library imports, data loading, merging, and subset selection. A red box highlights the title "VARIABLES, HISTORY, VERSION CONTROL" in the center of the interface.
- Console:** Displays statistical summaries (Mean, Median, Min, Max) for various variables across different groups (APOE4, FDG, AV45, CDRSB, ADAS13, MOCA, Hippocampus, Midtemp, nPACCtrailsB, HMSCORE, wholebrain, MTR, and ROI).
- Environment:** Shows the global environment with objects like `anmerge_subset`, `ids`, `MOCA`, and a `scatterplotter` function.
- File Browser:** Shows the project structure under `workshops > 2019_Rstudio_Magic`, including files like `Renviron`, `2019_Rstudio_Magic.Rproj`, `external`, `mac`, `output`, `R`, and `README.md`.

# RStudio Environment

The screenshot displays the RStudio interface with several panes:

- Code Editor:** Shows a script named `0_create_ADNI_data.R` with code for creating an ADNI dataset. The code includes sections for loading data, merging datasets, removing participants with missing data, and modifying variable classes.
- Console:** Shows statistical summaries for variables like APOE4, FDG, AV45, CDRSB, ADAS13, and MOCA. For example, APOE4 has a mean of 71.92 and a median of 70.00. The FDG variable has a range from 0.0000 to 89.60.
- Environment:** Shows the global environment with objects like `anerge_subset`, `variable_type_map`, `values`, and `functions`.
- File Browser:** Shows the project structure with files like `README.md`, `Renviron`, and `2019_Rstudio_Magic.Rproj`.

# RStudio Environment

~\workshops\2019\_RStudio\_Magic - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

o 0\_create\_ADNI\_data.RData 1\_create\_ADNI\_data\_tidyverse.R amerge\_subset

DATA VIEWER

DX	AGE	PTGENDER	PTEDUCAT	PTRECAT	PTRACCAT	APOE4	FDG	AV45	CDRSB	ADAS13	MOCA	WholeBrain	
2002	MCI	64.0	Male	18	Not Hisp/Latino	White	0	1.2091938	0.9754323	2.5	4	28	1135568.8
2003	MCI	63.6	Female	18	Not Hisp/Latino	White	0	1.2889628	1.1645374	2.0	11	24	1070369.3
2007	MCI	85.4	Female	20	Hisp/Latino	White	0	1.305182	1.4495259	2.5	9	23	920710.1
2010	MCI	62.9	Female	20	Not Hisp/Latino	Other	1	1.3121151	1.1472844	0.5	6	27	966402.9
2011	MCI	69.9	Female	14	Not Hisp/Latino	White	0	1.4571991	1.057399	1.5	7	25	987823.5
2018	MCI	76.4	Female	18	Not Hisp/Latino	White	0	1.3148491	1.052191	1.5	10	26	1004817.0
2022	MCI	66.0	Male	18	Not Hisp/Latino	Other	1	1.2031270	1.3135914	1.5	6	25	1173068.2
2027	MCI	61.9	Female	14	Not Hisp/Latino	White	0	1.4000448	1.0297671	1.0	6	24	969957.1
2031	MCI	72.5	Male	16	Not Hisp/Latino	White	0	1.3404430	0.9939887	2.0	10	24	1059879.5
2036	MCI	66.7	Female	14	Not Hisp/Latino	White	0	1.2959310	1.0307979	1.0	5	30	1019101.0
2037	MCI	75.8	Male	16	Not Hisp/Latino	White	1	1.3074956	1.4389912	0.5	20	20	1104797.3
2042	MCI	68.5	Male	20	Not Hisp/Latino	White	0	1.2081130	1.0555841	1.5	18	23	1061388.8
2043	MCI	72.2	Female	20	Not Hisp/Latino	White	1	1.3761158	1.2040191	2.0	8	27	1039110.3

Showing 10 of 15 1665 entries

Console Terminal Jobs

~\workshops\2019\_RStudio\_Magic

```
Mean : 71.92 Mean : 16.36
3rd Qu.: 76.60 3rd Qu.: 18.00
Max. : 89.60 Max. : 20.00

APOE4 FDG AV45 CDRSB ADAS13 MOCA
Min. : 0.0000 Min. : 0.6983 Min. : 0.8385 Min. : 0.0000 Min. : 0.0 Min. : 16.00
1st Qu.: 0.0000 1st Qu.: 1.1000 1st Qu.: 1.1000 1st Qu.: 0.0000 1st Qu.: 8.0 1st Qu.: 22.00
Median : 0.0000 Median : 1.2802 Median : 1.1105 Median : 0.0000 Median : 10.0 Median : 25.00
Mean : 0.5248 Mean : 1.2682 Mean : 1.1989 Mean : 0.1200 Mean : 13.8 Mean : 23.89
3rd Qu.: 1.0000 3rd Qu.: 1.3620 3rd Qu.: 1.3714 3rd Qu.: 2.0000 3rd Qu.: 18.0 3rd Qu.: 26.00
Max. : 2.0000 Max. : 1.7012 Max. : 2.0256 Max. : 15.5000 Max. : 46.0 Max. : 30.00

WholeBrain Hippocampus MidTemp nPACCtrailsB HMSCore
Min. : 114.421 Min. : 1.011 Min. : 12213 Min. : -18.6883 Min. : 0.0000
1st Qu.: 984410 1st Qu.: 6510 1st Qu.: 2535 1st Qu.: -1.051 1st Qu.: 0.0000
Median : 1051621 Median : 7223 Median : 20186 Median : -2.5250 Median : 1.0000
Mean : 1057026 Mean : 7150 Mean : 20302 Mean : -3.6882 Mean : 0.588
3rd Qu.: 1120570 3rd Qu.: 7834 3rd Qu.: 22088 3rd Qu.: -0.3482 3rd Qu.: 1.0000
Max. : 1486036 Max. : 10602 Max. : 32189 Max. : 5.3540 Max. : 3.0000
> view(amerge_subset)
> |
```

Environment History Connections Git

Global Environment

anmerge\_subset 665 obs. of 17 variables  
variable\_type\_map num [1:17] "0 0 0 0 0 0 0 1 0 ...  
ids chr [1:665] "2002" "2003" "2007" "2010" "2011" "2012" ...  
MOCA num [1:665] 28 24 23 27 25 26 25 24 24 30 ...  
Functions scatterplotter function (x, y, x.lim = NA, y.lim = NA, x.lab = "...") {

Files Plots Packages Help Viewer

Home workshops : 2019\_RStudio\_Magic

Name	Size	Modified
Renviron	52 B	May 12, 2019, 11:33 AM
2019_RStudio_Magic.Rproj	210 B	May 12, 2019, 6:30 PM
external		
mice		
output		
R		
README.md	42 B	May 12, 2019, 11:29 AM
Rmd		

# Some 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
  - ▶ Python, D3 (JavaScript), SQL, Shiny, LaTeX, Git/SVN, HTML/CSS, and so much more.
- ▶ This workshop
  - ▶ Will walk you through some of this (and more)
  - ▶ See [https://github.com/jennyrieck/workshops/tree/master/2019\\_Rstudio\\_Magic](https://github.com/jennyrieck/workshops/tree/master/2019_Rstudio_Magic)

## RStudio is more

- ▶ Not just an IDE
- ▶ A company
- ▶ A community
- ▶ A conference
- ▶ A centralized resource

# RStudio Resources

The screenshot shows the RStudio website homepage. At the top, there's a navigation bar with links for Products, Resources, Pricing, About Us, Blogs, and a search icon. Below the navigation is a decorative banner featuring a colorful, abstract graphic of overlapping colored bands.

**RStudio**: A screenshot of the RStudio IDE interface, showing the code editor, workspace, and plots.

**Shiny**: An image of a map of the United States with a "ZIP explorer" interface overlaid.

**R Packages**: Icons for several popular R packages: `markdown`, `Shiny`, `tidyverse`, `knitr`, and `ggplot2`.

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

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

**R Packages** description: Our developers create popular packages to expand the features of R. Includes `ggplot2`, `dplyr`, `R Markdown` & more.

At the bottom, there are download and learn more buttons for each section, and a horizontal orange progress bar.

# RStudio Resources

Online Learning - RStudio

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

R Studio

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 x + - □ x

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

R Studio Products Resources Pricing About Us Blogs Q

## 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 Reticulate 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

The reticulate package makes it easy to have and use Python in R. It's a Python interface, just like R itself.

Python in R Markdown

Object Conversion

Helpers



## Project and Environment Setup

- ▶ Special & hidden files
- ▶ Having a structure

## RStudio Setup

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

## For safety & collaboration

- ▶ Project(s) files
  - ▶ SOMETHING!

# Projects through Git

- ▶ Create a new project File

New Project

Create Project

---

 **New Directory**  
Start a project in a brand new working directory >

---

 **Existing Directory**  
Associate a project with an existing working directory >

---

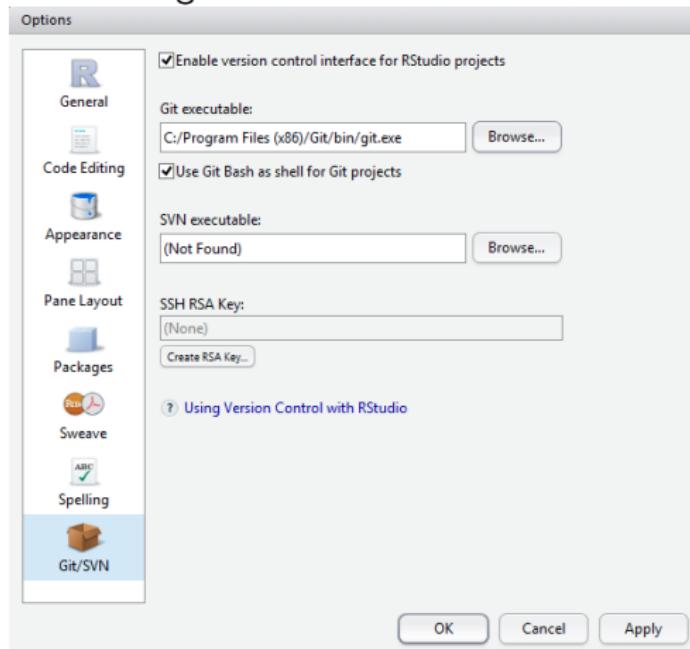
 **Version Control**  
Checkout a project from a version control repository >

Cancel

# Git & Projects

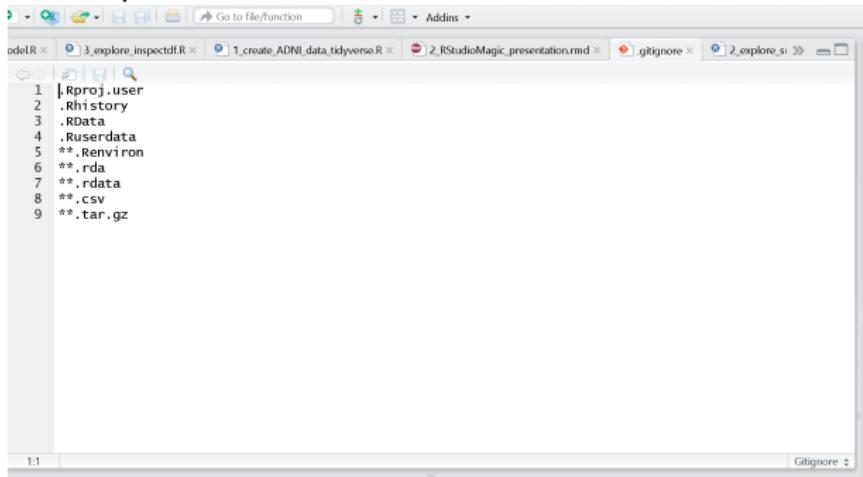
## ► Git

- Download git and link executable within RStudio



# Format .gitignore

- ▶ File types to ignore via version control
  - ▶ \*\* before each extension will match directories anywhere in the repo

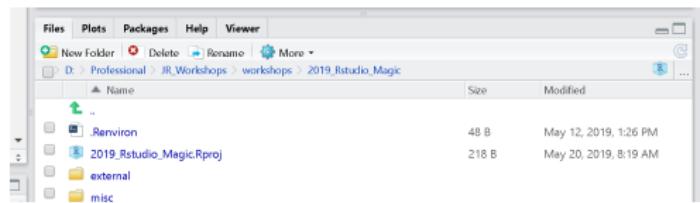


The screenshot shows the RStudio interface with the .gitignore tab selected in the top navigation bar. The main workspace displays the following content in the .gitignore file:

```
1 |Rproj.user
2 |.Rhistory
3 |.RData
4 |.Ruserdata
5 **|.Renvironment
6 **|.rda
7 **|.rdata
8 **|.CSV
9 **|.tar.gz
```

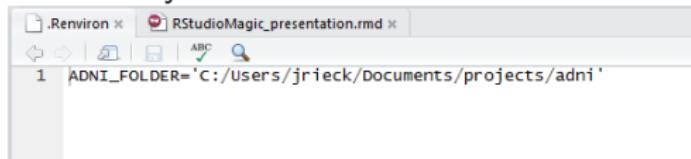
The code editor at the bottom shows the number "1:1" and the word "Gitignore" in the status bar.

# Environmental variables



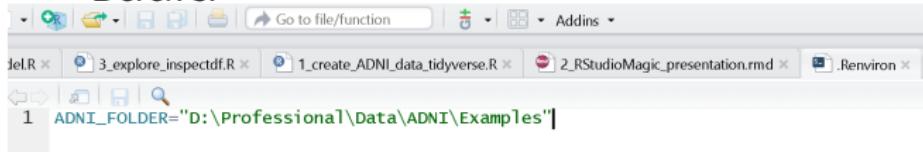
# Format environmental variables

- ▶ Set environmental variables (ie, directory location of data) to make code generalizable across computers
  - ▶ Don't commit or share these
- ▶ In **your** project folder create a `.Renvironment` file and define variables
  - ▶ Jenny's:



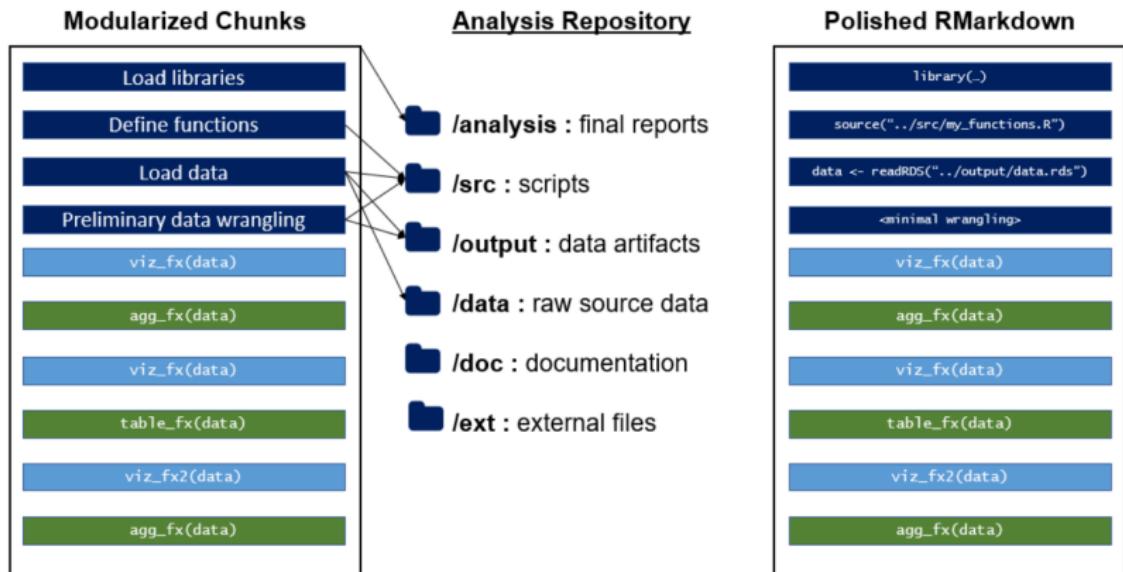
```
1 ADNI_FOLDER='C:/Users/jrieck/Documents/projects/adni'
```

- ▶ Derek's:



```
1 ADNI_FOLDER="D:\Professional\Data\ADNI\Examples"
```

# Organize your project folders and markdown



<https://emilyriederer.netlify.com/post/rmarkdown-driven-development/>

## Organize your project folders and markdown

- ▶ What works for you?
- ▶ What works for your organization or team?
- ▶ Maximize utility, minimize complexity

# This works for us

 [jennyrieck / workshops](#)

 Watch ▾ 1    Star 0    Fork 0

 Code    Issues 0    Pull requests 0    Projects 0    Wiki    Insights    Settings

Branch: master ▾ [workshops / 2019\\_Rstudio\\_Magic /](#)

 Create new file    Upload files    Find file    History

 jennyrieck added our favoRite things ...   Latest commit d818f26 6 hours ago

..

	R	more updates to manuscript example!	23 hours ago
	Rmd	added our favoRite things	6 hours ago
	external/images	reorganizing pngs	6 hours ago
	misc	reorganizing pngs	6 hours ago
	2019_Rstudio_Magic.Rproj	initial folder structure	5 days ago
	README.md	create readme	5 days ago
 README.md			
Rstudio magic for BrainHack Toronto 2019			

# This works for us

Screenshot of a GitHub repository interface showing a list of commits.

Branch: master    workshops / 2019\_Rstudio\_Magic / R /

Create new file Upload files Find file History

derekbeaton almost done now we hope    Latest commit e87b65d 16 hours ago

..

File	Message	Time Ago
0_create_ADNI_data_base.R	fixed rownames/race coding	8 days ago
1_create_ADNI_data_tidyverse.R	fixed rownames/race coding	8 days ago
2_explore_summarytools.R	almost done now we hope	16 hours ago
3_explore_inspectdf.R	almost done now we hope	16 hours ago
4_explore_DataExplorer_one_liner.R	almost done now we hope	16 hours ago
5_linear_model.R	almost done now we hope	16 hours ago
6_covstatis_example.R	please don't collide.	2 days ago

# This works for us

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Branch: master workshops / 2019\_Rstudio\_Magic / Rmd / Create new file Upload files Find file History

derekbeaton small update Latest commit 74c5384 14 hours ago

1\_a\_Simple\_RMarkdown\_PDF\_files/figure-latex more updates to manuscript example! 3 days ago

3\_RMarkdown APA Manuscript\_files updated numbers & structures 16 hours ago

1\_a\_Simple\_RMarkdown\_PDF.Rmd almost done now we hope 16 hours ago

1\_a\_Simple\_RMarkdown\_PDF.log whatever 2 days ago

1\_a\_Simple\_RMarkdown\_PDF.pdf more updates to manuscript example! 3 days ago

1\_a\_Simple\_RMarkdown\_PDF.tex tons of bells-and-whistles via the manuscript. 4 days ago

2\_RStudioMagic\_presentation.pdf small update 14 hours ago

2\_RStudioMagic\_presentation.rmd small update 14 hours ago

2\_RStudioMagic\_presentation.tex small update 14 hours ago

3\_RMarkdown APA Manuscript.Rmd updated numbers & structures 16 hours ago

3\_RMarkdown APA Manuscript.docx updated numbers & structures 16 hours ago

3\_RMarkdown APA Manuscript.pdf updated numbers & structures 16 hours ago

3\_RMarkdown APA Manuscript.tex updated numbers & structures 16 hours ago

r-references.bib updated numbers & structures 16 hours ago

# RStudio Setup

- ▶ Download R and Rstudio
- ▶ Strongly recommend Microsoft R  
(<https://mran.microsoft.com/open>)
  - ▶ Comes with Intel MKL
- ▶ Plain R is fine (<https://cran.r-project.org/>)
  - ▶ Can relink to faster libraries
- ▶ Download RStudio (<https://www.rstudio.com/>)

## Get the packages you need

```
#to install from CRAN
install.packages('devtools', dependencies = TRUE)

#to install from a git  (requires the devtools package)
dev.tools::install_github(Gibbsdavidl/CatterPlots)

#to install from a file
install.packages('/mypath/to/package/ADNIMERGE.tar.gz',
                 type='source', repos=NULL)
```

Some transition?

- ▶ A bit of background, including idiosyncrasies and unique things about R
  - ▶ Especially packages & three ways to install (somewhat covered above) CRAN, Locally, Git & others (devtools)
  - ▶ It's a functional language
  - ▶ Data types Including data frames & alts like tibbles
- ▶ Read/explore
  - ▶ explore .R scripts
- ▶ Clean/export
  - ▶ Show 0\_Create from PCA/MCA with Base, Tidyverse, Plyr (NOT dplyr), data.table
  - ▶ Reimport?
  - ▶ Analyze With MCA & covstatis

## Read in and create your dataframe

- ▶ ADNI Dataset adnimerge package
  - ▶ Reduce full dataset to only those participants (rows) and variables (columns) you're interested in
- ▶ Two methods to create your dataframe
  - ▶ using base R functions: 0\_create\_ADNI\_data\_base.R
  - ▶ Using tidyverse functions:  
`1_create_ADNI_data_tidyverse.R`

## Screenshots

Explanation

## Exploring your data

- ▶ Many packages to help explore and describe your data:
  - ▶ `summarytools`: `2_explore_summarytools.R`
  - ▶ `inspectdf`: `3_explore_inspectdf.R`
  - ▶ `DataExplorer`: `4_explore_DataExplorer_one_liner.R`

Code w/ eval=F

## Hard Break

- ▶ DataExplorer is dangerous
- ▶ Blind analyses can be *criminal*
  - ▶ de Leeuw paper quote
  - ▶ DEREK RANTS, PER USUAL.

## Analyze your data

- ▶ Linear models: 5\_linear\_model.R

## Screenshots / Code w/ eval=F

## Get experimental

- ▶ Explain motivation, not method
- ▶ covSTATIS: `6_covstatis_example.R`

## Part 3: RMarkdown

# RMarkdown

- ▶ What it is /why to use it
- ▶ A short deviation for LaTeX, and new helpers: kable & kableExtra
  - ▶ A taxonomy and how to approach this *Tying it all together through here* 1: simple RMD Plot-based visuals
    - ▶ Base, gt, ggplot, grobTable()/grid/gridExtra
    - ▶ 2: Slides (these ones here)
    - ▶ 3: Manuscripts!!
- ▶ Reporting/presentin

## RMarkdown Don(u)'ts

- ▶ Don't hardcode values
- ▶ Don't hardcode absolute file paths
- ▶ Don't do complicated database queries
- ▶ Don't litter
  - ▶ avoid eval=FALSE
  - ▶ reduce repeated code by making functions
- ▶ Don't load unnecessary libraries
- ▶ More at: <https://emilyriederer.netlify.com/post/rmarkdown-driven-development/>

## Part 4: Advanced R

## Some advanced/other things we're not covering

- ▶ package development
- ▶ Shiny
- ▶ SQL
- ▶ C/C++
- ▶ R2D3

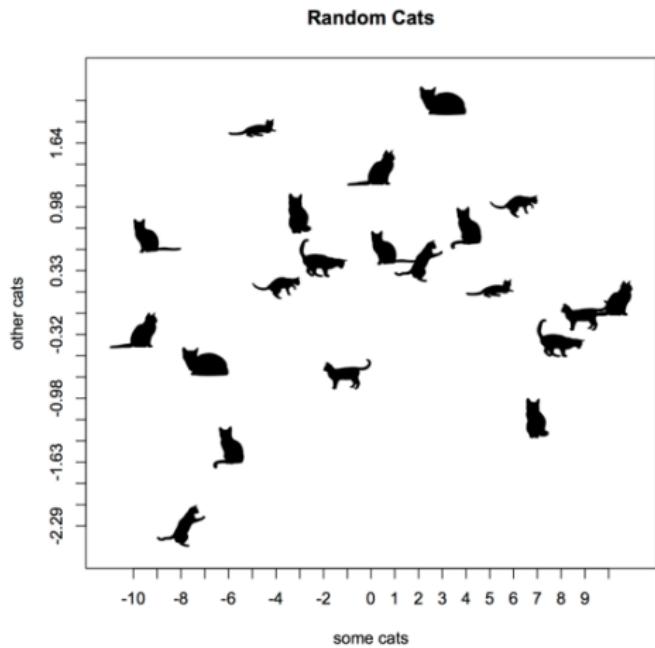
## A few of our favorite things

- ▶ Fun R do-dads

# CatterPlot for feline based graphics:

► <https://github.com/Gibbsdavidl/CatterPlots>

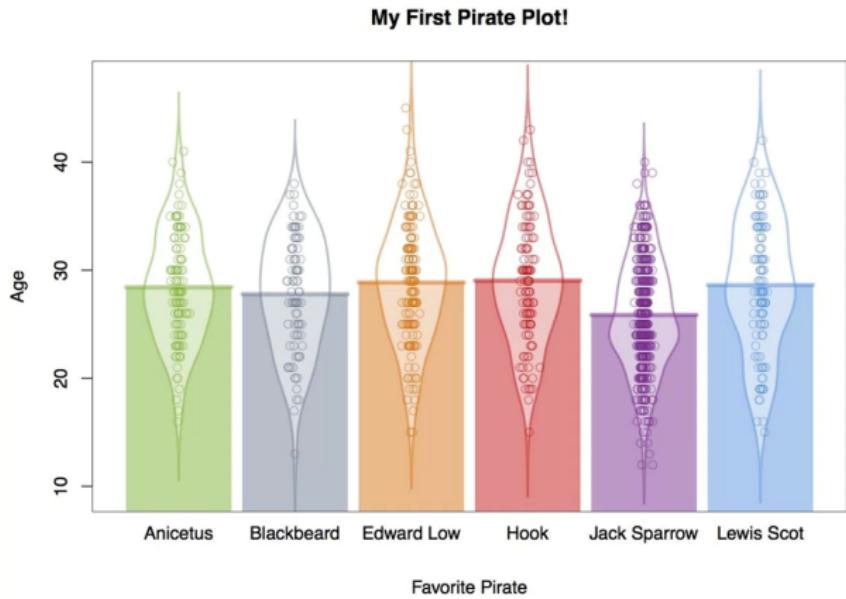
```
dev.tools::install_github(Gibbsdavidl/CatterPlots)
```



# What's a pirate's favorite programming language?

► <https://cran.r-project.org/web/packages/yarr/vignettes/pirateplot.html>

```
install.packages('yarr')
```

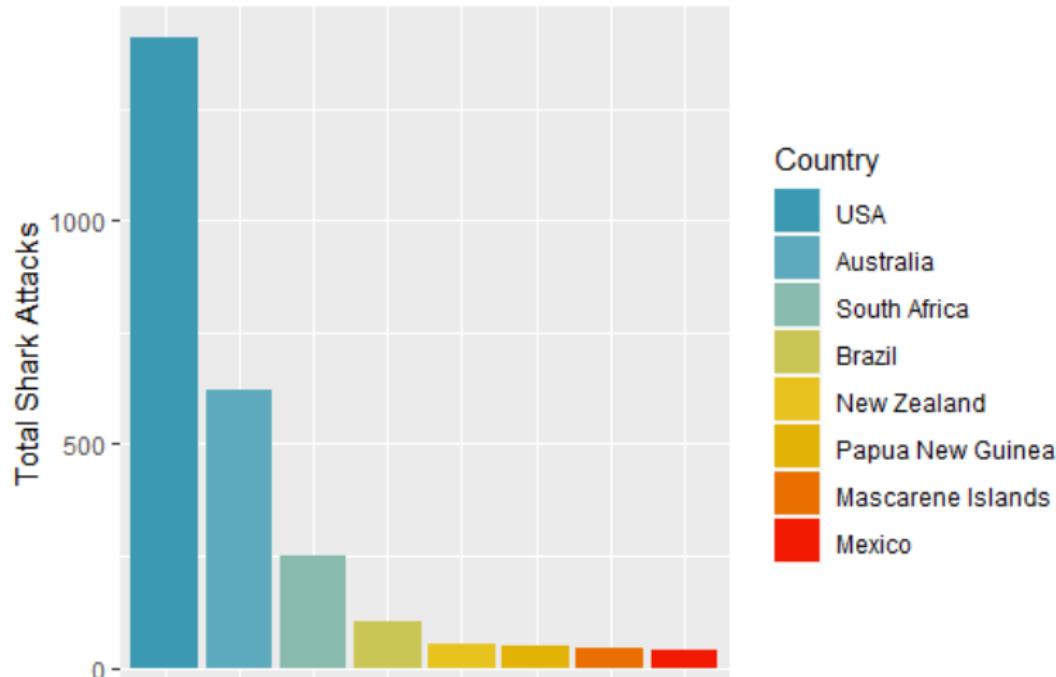


# Color palettes to fit your mood

► <https://github.com/karthik/wesanderson>

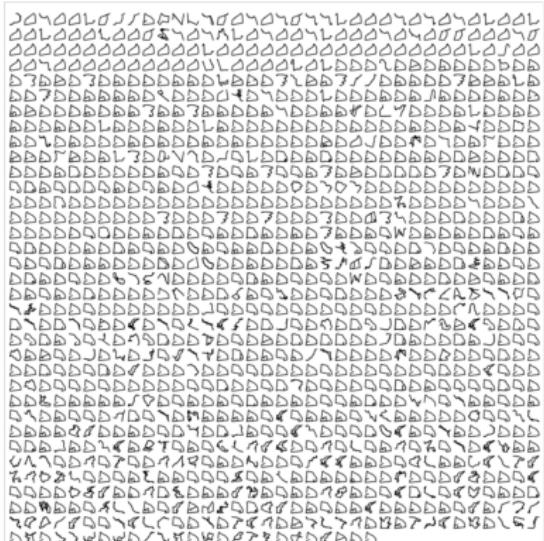
```
dev.tools::install_github(karthik/wesanderson)
```

Top countries with shark attacks  
(Esteban was eaten)



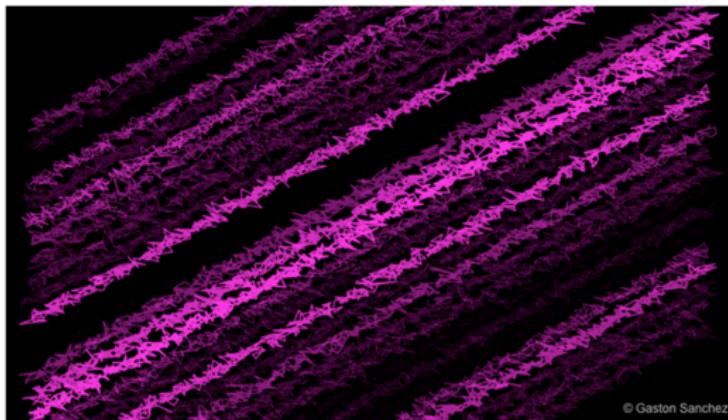
# Mapping your Strava routes

- ▶ <https://www.r-bloggers.com/strava-rides-map-in-r/>
- ▶ ALSO <https://marcusvolz.com/?p=4068>
  - ▶ `dev.tools::install_github(marcusvolz/strava)`



# Make aRt!

- ▶ R Graph Gallery
  - ▶ <http://www.r-graph-gallery.com/>
- ▶ Rtist: Gaston Sanchez
  - ▶ <http://gastonsanchez.com/Rtist/>



```
# -----
# Pink Barbs
# -----
# generate points x-y values
x <- seq(0, 100, length = 1000)
y <- x + rnorm(1000)

# -----
# Pink Barbs
# -----
# see graphical parameters
op <- par(bg = "black", mar = rep(0, 4))
# plot
plot(x, y, type = "n")
for (i in seq(-80, 70, by = 5))
{
  lines(x + rnorm(1000), x + i + rnorm(1000, 0), pch = 19,
        lwd = rnorm(0.8), lty = 1, runif(1000),
        lwd = sample(seq(0.1, 2, length = 20), 1))
}
# signature
legend("bottomright", legend = "@ Gaston Sanchez", bty = "n",
       text.col = "gray77")
# reset par
par(op)
dev.off()
```