

# A whirlwind tour of Rstudio, R, and Rmarkdown

## Magic for behavioral and brain data science

Jenny Rieck & Derek Beaton

Rotman Research Institute, Baycrest Health Sciences

May 16 2019

# The big outline

- ▶ Part 0: Background and Community

# The big outline

- ▶ Part 0: Background and Community
- ▶ Part 1: Rstudio

# The big outline

- ▶ Part 0: Background and Community
- ▶ Part 1: Rstudio
- ▶ Part 2: Project and environment setup

# The big outline

- ▶ Part 0: Background and Community
- ▶ Part 1: Rstudio
- ▶ Part 2: Project and environment setup
- ▶ Part 3: R (et al., eg Python)

# 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

# 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

# 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

## 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

## 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

## 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

## 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

## 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”

## 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

# What is R?

- ▶ R is for stats and general purpose programming
- ▶ R is a functional language

# 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

# 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

# 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

# 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

# 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

# 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



# 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

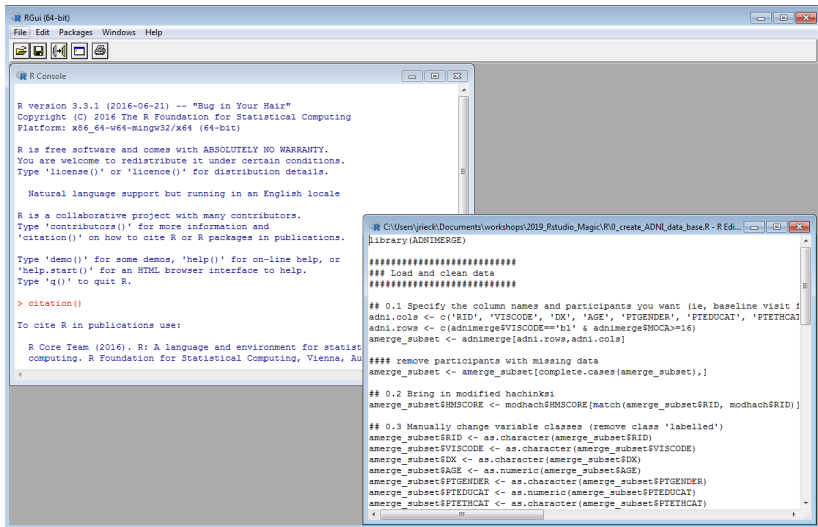
# What is R?

- ▶ 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)

# 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 toolbar 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 from the ADNI dataset.

```
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)
```

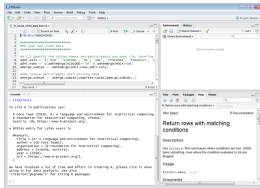
# But R has many interfaces

- ▶ Today we focus on RStudio (MatLab-like)

# But R has many interfaces

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

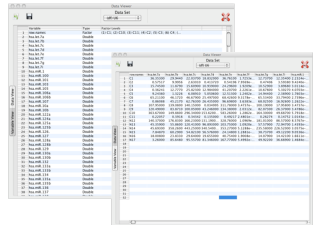
**RStudio**



**RCommander**



**Deducer**



R is a community (actually many communities!)

- ▶ Help and resources

# R is a community (actually many communities!)

- ▶ Help and resources
- ▶ Package development and distribution



# R: Help!

► <https://www.statmethods.net/>

## R: Help!

- ▶ <https://www.statmethods.net/>
- ▶ Online forums (Stack Exchange, r-lists)

## R: Help!

- ▶ <https://www.statmethods.net/>
- ▶ Online forums (Stack Exchange, r-lists)
- ▶ SpringerLink

## 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)

## 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

# 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

# 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



# 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

# 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

# Part 1: RStudio

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

# RStudio Environment

~/workshops/2019\_Rstudio\_Magic-master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Source on Save

Run

Source

Environment History Connections Git

Global Environment

Data

amerge\_subset 665 obs. of 17 variables

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 = "...")

```
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$HMScore <- modhach$HMScore[match(amerge_subset$RID, modhach$RID)]
17
18 ## 0.3 Manually change variable classes (remove class 'labelled')
19 <-
```

CONSOLE

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

Console Terminal Jobs

```
~/workshops/2019_Rstudio_Magic/ >
Mean :71.92 Mean :10.36
3rd Qu.:176.60 3rd Qu.:18.00
Max. :89.60 Max. :20.00
APOE4 FDG APOE4 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.:1.1837 1st Qu.:1.0199 1st Qu.:0.0000 1st Qu.: 8.0 1st Qu.:22.00
Median :0.0000 Median :1.2802 Median :1.1105 Median :1.0000 Median :12.0 Median :24.00
Mean :0.5248 Mean :1.2682 Mean :1.1989 Mean :1.202 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. :1.7011 Max. :2.0256 Max. :5.500 Max. :46.0 Max. :30.00
Mholatrain Hippocampus MidTemp mPACCtra1158 HMScore
Min. : 817421 Min. :12213 Min. : -38.6983 Min. :0.0000
1st Qu.: 984410 1st Qu.: 6510 1st Qu.:18535 1st Qu.: -6.4051 1st Qu.:0.0000
Median :1051621 Median : 7223 Median :20186 Median : -2.5250 Median :1.0000
Mean :1105026 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. :11486036 Max. :110602 Max. :32189 Max. : 5.3540 Max. :3.000
> view(amerge_subset)
```

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home workshops 2019\_Rstudio\_Magic

Name Size Modified

Environment 52 B May 12, 2019, 11:33 AM

2019\_Rstudio\_Magic.Rproj 218 B May 12, 2019, 6:30 PM

external

mic

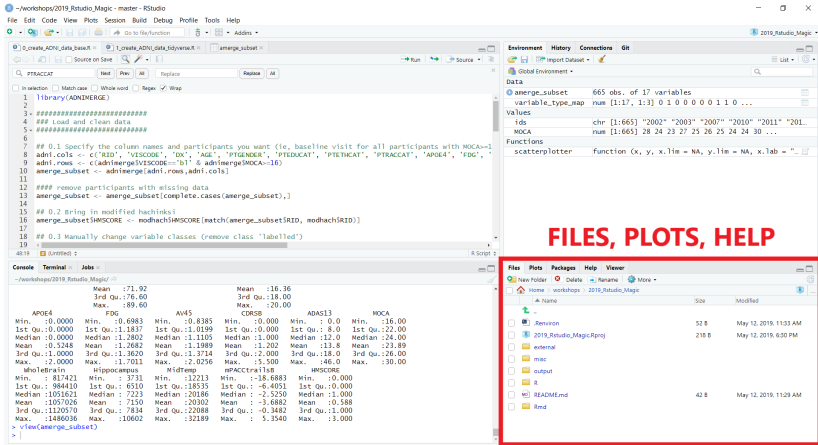
output

R

README.md 42 B May 12, 2019, 11:29 AM

Rmd

## RStudio Environment



# RStudio Environment

The screenshot displays the RStudio interface with the following components:

- Script Editor:** Contains R code for loading the `ADNI` dataset, cleaning it, and creating a subset of variables and participants. The code includes comments and function calls like `library(ADNIMERGE)`, `admi_rows <- c(adnimerge$VISCODE=="b1")`, and `amerge_subset <- adnimerge[admi_rows, admi_cols]`.
- Console:** Shows the output of the code execution, including summary statistics for various variables such as `AP0E4`, `FDG`, `AV45`, `CDR5B`, `ADAS13`, `MOCA`, `ihofetbrain`, `Hippocampus`, `MidTemp`, `hPACCtra115B`, and `IMSORE`.
- Environment Pane (Highlighted):** Displays the current environment, showing the `amerge_subset` data frame with 665 observations and 17 variables. It also lists the functions loaded in the environment, including `scatterplot`.

**VARIABLES, HISTORY, VERSION CONTROL**



## RStudio Environment

The screenshot displays an RStudio interface with the following components:

- Source Editor:** Contains R code for data cleaning. A red box highlights the first 19 lines, which include:
  - Library loading: `library(ADNIERGE)`
  - Data loading: `## load and clean data`
  - Column selection: `admi.rows <- c(adniERGE$VISCORE=="b1" | adniERGE$MCOA=="16")`
  - Data removal: `admi.rows <- admi.rows[admi.rows,admi.cols]`
  - Missing data removal: `admi.rows <- admi.rows[complete.cases(admi.rows),]`
  - Variable renaming: `admi.rows$MCOA <- modhach$MCOA[match(admi.rows$MCOA, modhach$MCOA)]`
  - Manual class change: `admi.rows$MCOA <- factor(admi.rows$MCOA, levels=c("b1", "b2", "b3", "b4", "b5", "b6", "b7", "b8", "b9", "b10", "b11", "b12", "b13", "b14", "b15", "b16", "b17", "b18", "b19", "b20", "b21", "b22", "b23", "b24", "b25", "b26", "b27", "b28", "b29", "b30", "b31", "b32", "b33", "b34", "b35", "b36", "b37", "b38", "b39", "b40", "b41", "b42", "b43", "b44", "b45", "b46", "b47", "b48", "b49", "b50", "b51", "b52", "b53", "b54", "b55", "b56", "b57", "b58", "b59", "b60", "b61", "b62", "b63", "b64", "b65", "b66", "b67", "b68", "b69", "b70", "b71", "b72", "b73", "b74", "b75", "b76", "b77", "b78", "b79", "b80", "b81", "b82", "b83", "b84", "b85", "b86", "b87", "b88", "b89", "b90", "b91", "b92", "b93", "b94", "b95", "b96", "b97", "b98", "b99", "b100", "b101", "b102", "b103", "b104", "b105", "b106", "b107", "b108", "b109", "b110", "b111", "b112", "b113", "b114", "b115", "b116", "b117", "b118", "b119", "b120", "b121", "b122", "b123", "b124", "b125", "b126", "b127", "b128", "b129", "b130", "b131", "b132", "b133", "b134", "b135", "b136", "b137", "b138", "b139", "b140", "b141", "b142", "b143", "b144", "b145", "b146", "b147", "b148", "b149", "b150", "b151", "b152", "b153", "b154", "b155", "b156", "b157", "b158", "b159", "b160", "b161", "b162", "b163", "b164", "b165", "b166", "b167", "b168", "b169", "b170", "b171", "b172", "b173", "b174", "b175", "b176", "b177", "b178", "b179", "b180", "b181", "b182", "b183", "b184", "b185", "b186", "b187", "b188", "b189", "b190", "b191", "b192", "b193", "b194", "b195", "b196", "b197", "b198", "b199", "b200", "b201", "b202", "b203", "b204", "b205", "b206", "b207", "b208", "b209", "b210", "b211", "b212", "b213", "b214", "b215", "b216", "b217", "b218", "b219", "b220", "b221", "b222", "b223", "b224", "b225", "b226", "b227", "b228", "b229", "b230", "b231", "b232", "b233", "b234", "b235", "b236", "b237", "b238", "b239", "b240", "b241", "b242", "b243", "b244", "b245", "b246", "b247", "b248", "b249", "b250", "b251", "b252", "b253", "b254", "b255", "b256", "b257", "b258", "b259", "b260", "b261", "b262", "b263", "b264", "b265", "b266", "b267", "b268", "b269", "b270", "b271", "b272", "b273", "b274", "b275", "b276", "b277", "b278", "b279", "b280", "b281", "b282", "b283", "b284", "b285", "b286", "b287", "b288", "b289", "b290", "b291", "b292", "b293", "b294", "b295", "b296", "b297", "b298", "b299", "b300", "b301", "b302", "b303", "b304", "b305", "b306", "b307", "b308", "b309", "b310", "b311", "b312", "b313", "b314", "b315", "b316", "b317", "b318", "b319", "b320", "b321", "b322", "b323", "b324", "b325", "b326", "b327", "b328", "b329", "b330", "b331", "b332", "b333", "b334", "b335", "b336", "b337", "b338", "b339", "b340", "b341", "b342", "b343", "b344", "b345", "b346", "b347", "b348", "b349", "b350", "b351", "b352", "b353", "b354", "b355", "b356", "b357", "b358", "b359", "b360", "b361", "b362", "b363", "b364", "b365", "b366", "b367", "b368", "b369", "b370", "b371", "b372", "b373", "b374", "b375", "b376", "b377", "b378", "b379", "b380", "b381", "b382", "b383", "b384", "b385", "b386", "b387", "b388", "b389", "b390", "b391", "b392", "b393", "b394", "b395", "b396", "b397", "b398", "b399", "b400", "b401", "b402", "b403", "b404", "b405", "b406", "b407", "b408", "b409", "b410", "b411", "b412", "b413", "b414", "b415", "b416", "b417", "b418", "b419", "b420", "b421", "b422", "b423", "b424", "b425", "b426", "b427", "b428", "b429", "b430", "b431", "b432", "b433", "b434", "b435", "b436", "b437", "b438", "b439", "b440", "b441", "b442", "b443", "b444", "b445", "b446", "b447", "b448", "b449", "b450", "b451", "b452", "b453", "b454", "b455", "b456", "b457", "b458", "b459", "b460", "b461", "b462", "b463", "b464", "b465", "b466", "b467", "b468", "b469", "b470", "b471", "b472", "b473", "b474", "b475", "b476", "b477", "b478", "b479", "b480", "b481", "b482", "b483", "b484", "b485", "b486", "b487", "b488", "b489", "b490", "b491", "b492", "b493", "b494", "b495", "b496", "b497", "b498", "b499", "b500", "b501", "b502", "b503", "b504", "b505", "b506", "b507", "b508", "b509", "b510", "b511", "b512", "b513", "b514", "b515", "b516", "b517", "b518", "b519", "b520", "b521", "b522", "b523", "b524", "b525", "b526", "b527", "b528", "b529", "b530", "b531", "b532", "b533", "b534", "b535", "b536", "b537", "b538", "b539", "b540", "b541", "b542", "b543", "b544", "b545", "b546", "b547", "b548", "b549", "b550", "b551", "b552", "b553", "b554", "b555", "b556", "b557", "b558", "b559", "b560", "b561", "b562", "b563", "b564", "b565", "b566", "b567", "b568", "b569", "b570", "b571", "b572", "b573", "b574", "b575", "b576", "b577", "b578", "b579", "b580", "b581", "b582", "b583", "b584", "b585", "b586", "b587", "b588", "b589", "b590", "b591", "b592", "b593", "b594", "b595", "b596", "b597", "b598", "b599", "b600", "b601", "b602", "b603", "b604", "b605", "b606", "b607", "b608", "b609", "b610", "b611", "b612", "b613", "b614", "b615", "b616", "b617", "b618", "b619", "b620", "b621", "b622", "b623", "b624", "b625", "b626", "b627", "b628", "b629", "b630", "b631", "b632", "b633", "b634", "b635", "b636", "b637", "b638", "b639", "b640", "b641", "b642", "b643",`

# RStudio Environment

~/workshops/2019\_Rstudio\_Magic-master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

## DATA VIEWER

	DX	AGE	PTGENDER	PTEDUCAT	PTETHCAT	PTRACCAT	APOE4	FDG	AV45	CDRSB	ADAS13	MOCA	WholeBrain
2002	MCI	64.8	Male	18	Not His/Latino	White	0	1.2091908	0.9794523	2.5	4	28	1123556.8
2003	MCI	65.6	Female	18	Not His/Latino	White	0	1.2899625	1.1646374	2.0	11	24	1070369.5
2007	MCI	85.4	Female	20	His/Latino	White	0	1.3058182	1.4495250	2.5	9	23	920710.1
2010	MCI	62.9	Female	20	Not His/Latino	Other	1	1.3121151	1.1472848	0.5	6	27	986402.9
2011	MCI	69.9	Female	14	Not His/Latino	White	0	1.4537199	1.0537930	1.5	7	25	967822.5
2018	MCI	76.4	Female	18	Not His/Latino	White	0	1.3148491	1.0525191	1.5	10	26	1004817.0
2022	MCI	66.0	Male	18	Not His/Latino	Other	1	1.2031270	1.3135914	1.5	6	25	1173068.2
2023	MCI	61.9	Female	14	Not His/Latino	White	0	1.4000446	1.0299761	1.0	6	24	969957.1
2031	MCI	72.5	Male	16	Not His/Latino	White	0	1.3404430	0.9939887	2.0	10	24	1059879.5
2036	MCI	66.7	Female	14	Not His/Latino	White	0	1.2892910	1.0300795	1.0	5	30	1019101.0
2037	MCI	75.8	Male	16	Not His/Latino	White	1	1.3074956	1.4389912	0.5	20	20	1104797.3
2042	MCI	69.5	Male	20	Not His/Latino	White	0	1.2083193	1.0655846	1.5	18	23	1061388.4
2043	MCI	72.2	Female	20	Not His/Latino	White	1	1.2781158	1.2040191	2.0	8	27	1032110.3

Showing 110/13 of 685 entries

Environment History Connections Git

Global Environment

Data

amerge\_subset 665 obs. of 17 variables

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
Renviron	52 B	May 12, 2019, 11:33 AM
2019_Rstudio_Magic.Rproj	218 B	May 12, 2019, 6:50 PM
external		
mic		
output		
R		
README.md	42 B	May 12, 2019, 11:29 AM
Rmd		

```
~/workshops/2019_Rstudio_Magic/ >
  Mean : 71.92
  3rd Qu.: 176.60
  Max. : 89.60

  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.:1.1837 1st Qu.:1.0199 1st Qu.:0.0000 1st Qu.: 8.0 1st Qu.:22.00
Median :0.0000 Median :1.2802 Median :1.1105 Median :1.0000 Median :12.0 Median :24.00
Mean :0.5248 Mean :1.2682 Mean :1.1989 Mean :1.2020 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. :1.7011 Max. :2.0256 Max. :5.500 Max. :46.0 Max. :30.00

  WholeBrain
Min. : 817421 Min. : 3731 Min. :12213 Min. : -38.6883 Min. :0.0000
1st Qu.: 984410 1st Qu.: 6510 1st Qu.:18535 1st Qu.: -6.4051 1st Qu.:0.0000
Median :1051621 Median : 7223 Median :20186 Median : -2.5250 Median :1.0000
Mean :1105026 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. :11486036 Max. :110602 Max. :32189 Max. : 5.3540 Max. :3.000

> view(amerge_subset)
>
```

# Benefits of RStudio

- ▶ Built-in integration with version control (git or SVN)

# Benefits of RStudio

- ▶ Built-in integration with version control (git or SVN)
- ▶ Package and documentation generation

# Benefits of RStudio

- ▶ Built-in integration with version control (git or SVN)
- ▶ Package and documentation generation
- ▶ Reproducible science!

# Benefits of RStudio

- ▶ Built-in integration with version control (git or SVN)
- ▶ Package and documentation generation
- ▶ Reproducible science!
  - ▶ R Markdown documents

# 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

# 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



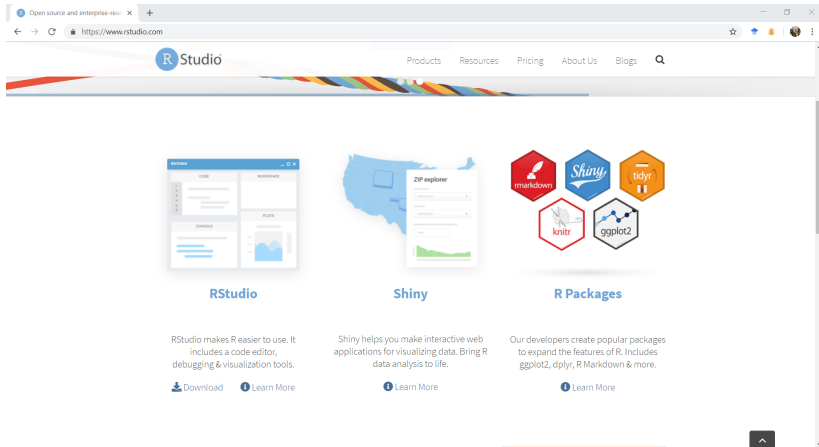
# Benefits of RStudio

- ▶ 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!)

# 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 Resources



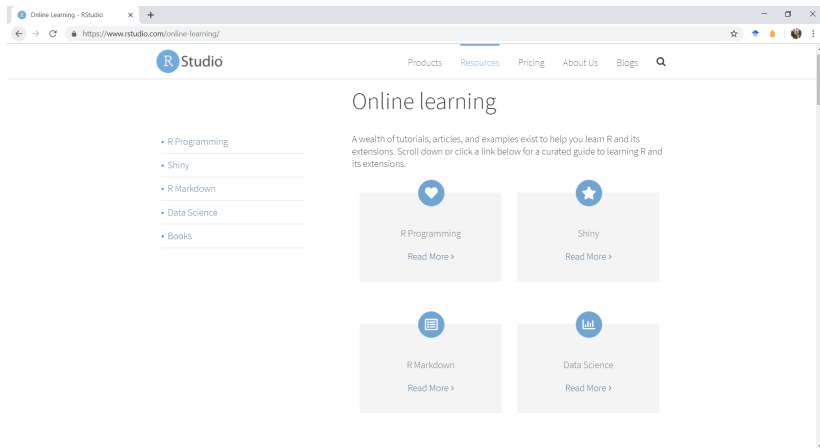
The screenshot shows the RStudio website homepage. The browser address bar displays "https://www.rstudio.com". The navigation bar includes links for "Products", "Resources", "Pricing", "About Us", and "Blogs", along with a search icon. A decorative graphic of a rainbow-colored line is positioned below the navigation bar.

The main content area features three columns of resources:

- RStudio**: Accompanied by an image of the RStudio IDE interface. The text states: "RStudio makes R easier to use. It includes a code editor, debugging & visualization tools." Below this are two links: "Download" and "Learn More".
- Shiny**: Accompanied by an image of a Shiny web application titled "ZIP explorer". The text states: "Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life." Below this is a link: "Learn More".
- R Packages**: Accompanied by an image showing logos for "markdown", "Shiny", "tidyr", "knitr", and "ggplot2". The text states: "Our developers create popular packages to expand the features of R. Includes ggplot2, dplyr, R Markdown & more." Below this is a link: "Learn More".

A small "Up" arrow icon is located in the bottom right corner of the page.

# 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 purr

- 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

Python in R Markdown

Python in R code

Object Conversion

Helpful Resources

## Part 2: Project and Environment Setup

- ▶ Hidden files & whatnot

## Part 2: Project and Environment Setup

- ▶ Hidden files & whatnot
- ▶ Have a structure ready to go on Github

## Part 2: Project and Environment Setup

- ▶ Hidden files & whatnot
- ▶ Have a structure ready to go on Github
- ▶ Explain/walk through



## Part 2: Project and Environment Setup

- ▶ Hidden files & whatnot
- ▶ Have a structure ready to go on Github
- ▶ Explain/walk through
- ▶ Discuss the helpful packages above

# RStudio Setup

## ► Download R and Rstudio

```
#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)
```

# 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)
```

# RStudio Setup

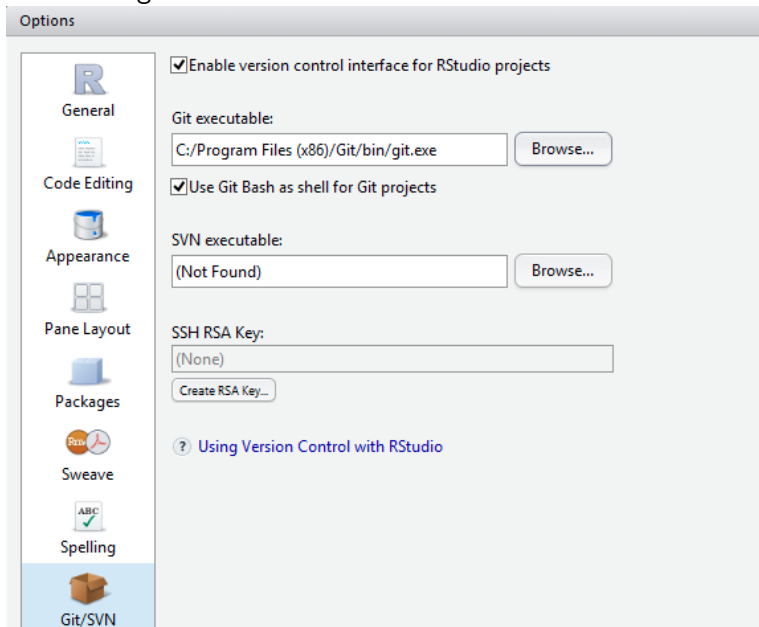
- ▶ Download R and Rstudio
- ▶ Add-on packages

```
#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> for a detailed guide

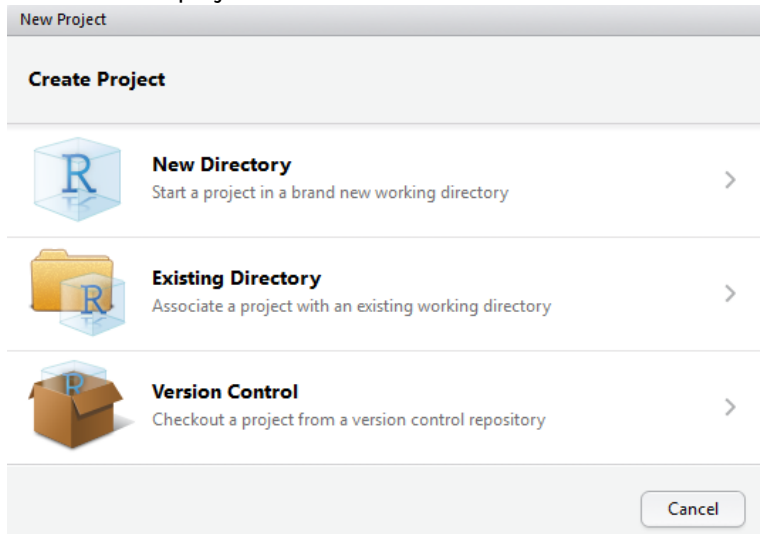
# Rstudio Setup: Projects & Git

- Download git and link to RStudio



# Rstudio Setup: Projects & Git

- Create a new project File



# Format .gitignore

- ▶ File types to ignore:

# Format .gitignore

- ▶ File types to ignore:
  - ▶ `.Rproj.user`



# Format .gitignore

- ▶ File types to ignore:
  - ▶ .Rproj.user
  - ▶ .Rhistory

# Format .gitignore

- ▶ File types to ignore:
  - ▶ .Rproj.user
  - ▶ .Rhistory
  - ▶ .Ruserdata

# Format .gitignore

- ▶ File types to ignore:
  - ▶ .Rproj.user
  - ▶ .Rhistory
  - ▶ .Ruserdata
  - ▶ .Renviron

# Format .gitignore

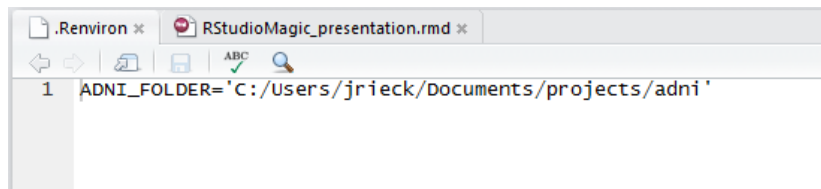
- ▶ File types to ignore:
  - ▶ .Rproj.user
  - ▶ .Rhistory
  - ▶ .Ruserdata
  - ▶ .Renvirom
  - ▶ .rda & .Rdata (to avoid pushing potentially sensitive data files to git)

# Format .gitignore

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

## Format environmental variables

- ▶ Set environmental variables (ie, directory location of data) to make code generalizable across computers

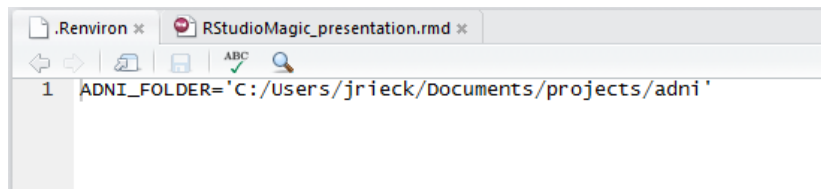


The screenshot shows a window with two tabs: ".Renviron" and "RStudioMagic\_presentation.rmd". The ".Renviron" tab is active, displaying a single line of code: `ADNI_FOLDER='C:/Users/jr.ieck/Documents/projects/adni'`. The code is on line 1. The window has a standard toolbar with icons for navigation, editing, and searching.

```
1 ADNI_FOLDER='C:/Users/jr.ieck/Documents/projects/adni'
```

# Format environmental variables

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



The screenshot shows an RStudio editor window with two tabs: `.Renvi` and `RStudioMagic_presentation.rmd`. The `.Renvi` tab is active, displaying a single line of code: `ADNI_FOLDER='C:/Users/jrleck/Documents/projects/adni'`. The code is written in a monospaced font, and the line number 1 is visible in the left margin. The RStudio interface includes standard navigation icons (back, forward, search, etc.) above the code editor.

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