# Some Essentials for Data Science with R

Derek Beaton

2020 FEB 23

# Outline

- Part 0: Project set up

# Outline

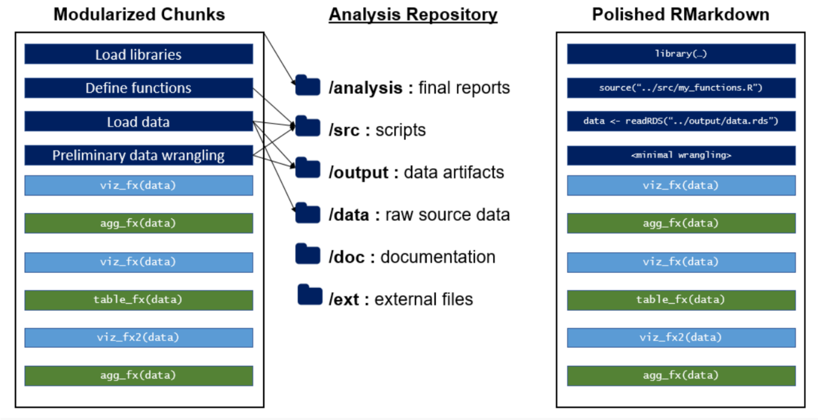
- Part 0: Project set up
- Part 1: RStudio, R, RMarkdown, Git

# Outline

- Part 0: Project set up
- Part 1: RStudio, R, RMarkdown, Git
- Part 2: Working with data

# Project set up

# Project set up



**Modularized Chunks**

Load libraries

Define functions

Load data

Preliminary data wrangling

viz_fx(data)

agg_fx(data)

viz_fx(data)

table_fx(data)

viz_fx2(data)

agg_fx(data)

**Analysis Repository**

**/analysis** : final reports

**/src** : scripts

**/output** : data artifacts

**/data** : raw source data

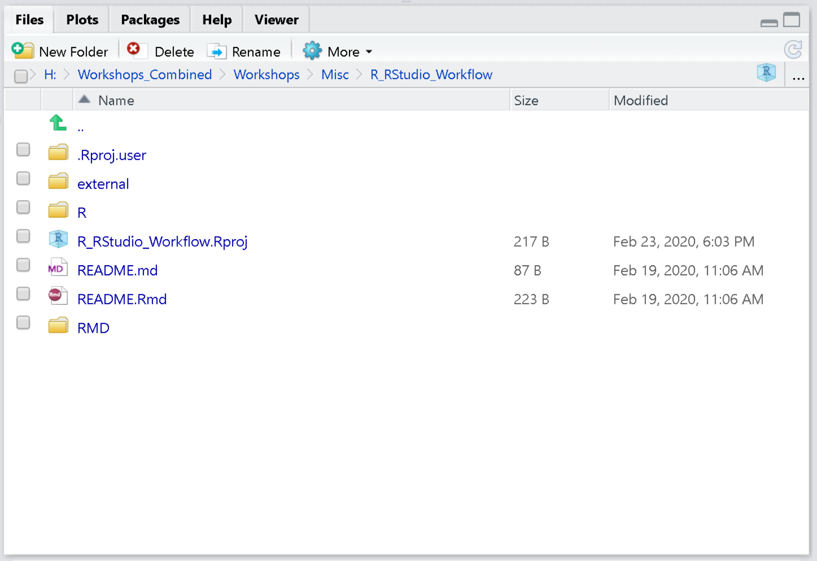
**/doc** : documentation

**/ext** : external files

**Polished RMarkdown**

```
library(…)
source("../src/my_functions.R")
data <- readRDS("../output/data.rds")
<minimal wrangling>
viz_fx(data)
agg_fx(data)
viz_fx(data)
table_fx(data)
viz_fx2(data)
agg_fx(data)
```

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



H: > Workshops_Combined > Workshops > Misc > R_RStudio_Workflow

| | Name | Size | Modified |
|---|---|---|---|
| | .. | | |
| □ | .Rproj.user | | |
| □ | external | | |
| □ | R | | |
| □ | R_RStudio_Workflow.Rproj | 217 B | Feb 23, 2020, 6:03 PM |
| □ | README.md | 87 B | Feb 19, 2020, 11:06 AM |
| □ | README.Rmd | 223 B | Feb 19, 2020, 11:06 AM |
| □ | RMD | | |

# Organize your project folders and markdown

- What works for you?

# Organize your project folders and markdown

- What works for you?
- What works for your organization or team?

# Organize your project folders and markdown

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

# Part 1

# Part 2: RStudio & Project setup

# RStudio

- IDE: Integrated development environment

# RStudio

- IDE: Integrated development environment
- RStudio: Does so much

# RStudio

- IDE: Integrated development environment
- RStudio: Does so much
  - We scratch the surface here

# RStudio

- IDE: Integrated development environment
- RStudio: Does so much
    - We scratch the surface here
- Quick walk through

# RStudio

- IDE: Integrated development environment
- RStudio: Does so much
  - We scratch the surface here
- Quick walk through
  - Followed by specific set up

# RStudio

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

# 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 Setup

- Download R and Rstudio

# RStudio Setup

- Download R and Rstudio
  - Strongly recommend Microsoft R (https://mran.microsoft.com/open)

# RStudio Setup

- Download R and Rstudio
  - Strongly recommend Microsoft R (https://mran.microsoft.com/open)
  - Comes with Intel MKL

# 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/)

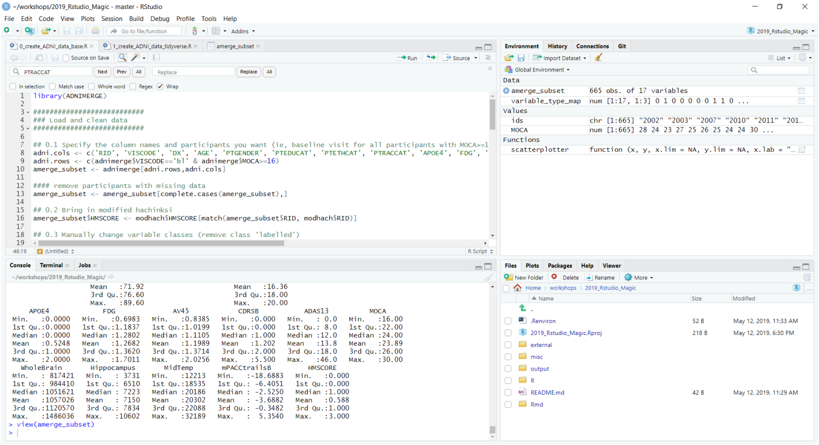
# RStudio Setup

- 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

# 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/)

# RStudio Environment



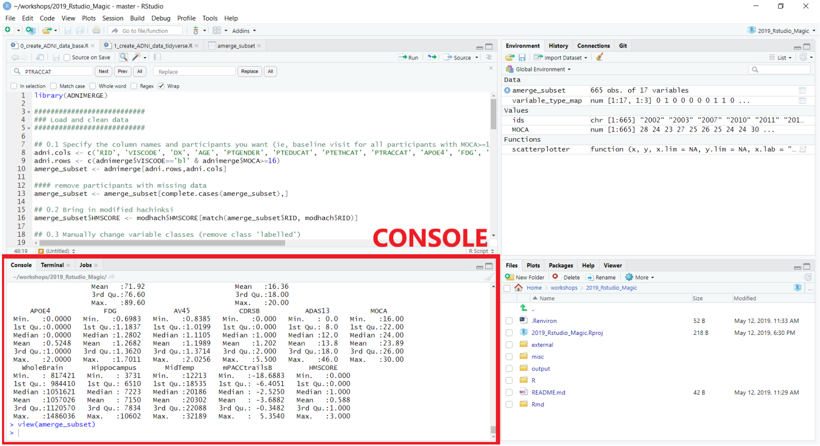
□ In selection □ Match case □ Whole word □ Regex ☑ Wrap

```r
library(ADNIMERGE)

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

## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA>=1
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT', 'PTRACCAT', 'APOE4', 'FDG', '
adni.rows <- c(adnimerge$VISCODE=='bl' & adnimerge$MOCA>=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 hachinksi
amerge_subset$HMSCORE <- modhach$HMSCORE[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
```

```
                 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.   :0.000   Min.   : 0.0   Min.   :16.00  
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0   1st Qu.:22.00  
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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       Hippocampus       MidTemp       mPACCtrailsB         HMSCORE     
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000  
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000  
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000  
 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.:1.000  
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000  
> view(amerge_subset)
> 
```

| | |
|---|---|
| Data | |
| amerge_subset | 665 obs. of 17 variables |
| variable_type_map | num [1:17, 1:3] 0 1 0 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 | |
| scatterplotter | function (x, y, x.lim = NA, y.lim = NA, x.lab = "… |

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external | | |
| misc | | |
| output | | |
| R | | |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd | | |



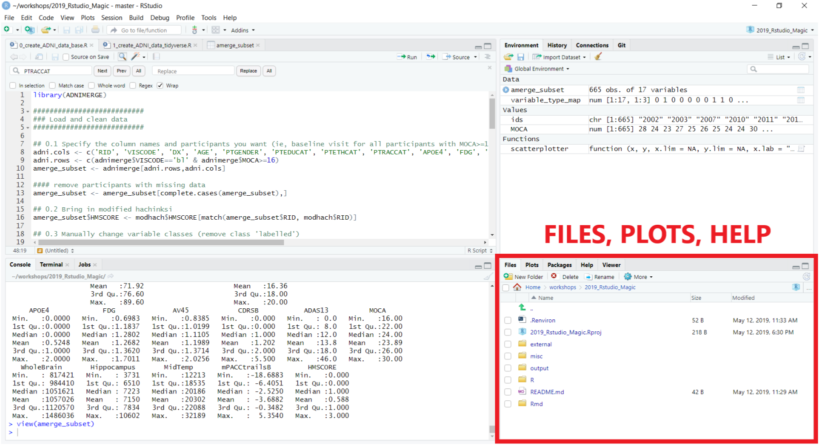
~/workshops/2019_Rstudio_Magic - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
2019_Rstudio_Magic
0_create_ADNI_data_base.R
1_create_ADNI_data_tidyverse.R
amerge_subset
Source on Save
Run
Source
PTRACCAT
Next
Prev
All
Replace
Replace
All
□ In selection □ Match case □ Whole word □ Regex ☑ Wrap
48:19
(Untitled)
R Script
CONSOLE
Console
Terminal
Jobs
~/workshops/2019_Rstudio_Magic/
Environment
History
Connections
Git
Import Dataset
List
Global Environment
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home › workshops › 2019_Rstudio_Magic

```r
library(ADNIMERGE)

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

## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA>=1
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT', 'PTRACCAT', 'APOE4', 'FDG', '
adni.rows <- c(adnimerge$VISCODE=='bl' & adnimerge$MOCA>=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 hachinksi
amerge_subset$HMSCORE <- modhach$HMSCORE[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
```

```
                   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.   :0.000   Min.   : 0.0   Min.   :16.00
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0   1st Qu.:22.00
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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       Hippocampus        MidTemp       mPACCtrailsB          HMSCORE
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000
 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.:1.000
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000
> view(amerge_subset)
> 
```

```
Data
amerge_subset          665 obs. of 17 variables
  variable_type_map    num [1:17, 1:3] 0 1 0 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
  scatterplotter       function (x, y, x.lim = NA, y.lim = NA, x.lab = "…
```

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external | | |
| misc | | |
| output | | |
| R | | |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd | | |



PTRACCAT

□ In selection □ Match case □ Whole word □ Regex ☑ Wrap

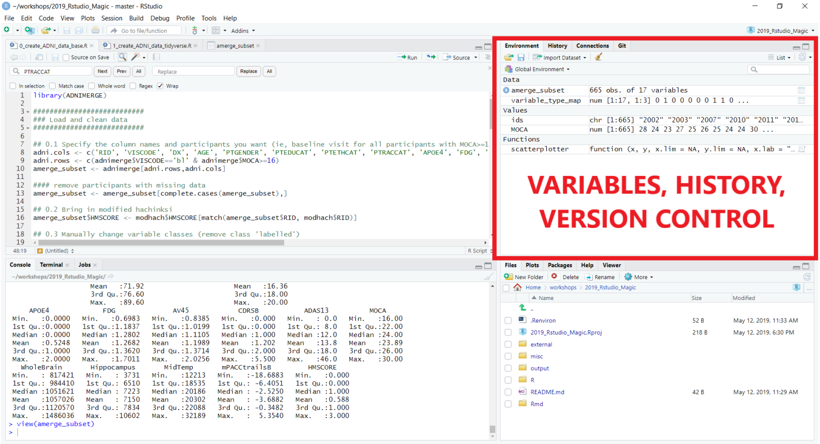
```
library(ADNIMERGE)

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

## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA>=1
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT', 'PTRACCAT', 'APOE4', 'FDG', '
adni.rows <- c(adnimerge$VISCODE=='bl' & adnimerge$MOCA>=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 hachinksi
amerge_subset$HMSCORE <- modhach$HMSCORE[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
```

```
Data
  amerge_subset        665 obs. of 17 variables
  variable_type_map    num [1:17, 1:3] 0 1 0 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
  scatterplotter       function (x, y, x.lim = NA, y.lim = NA, x.lab = "…
```

```
                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.   :0.000   Min.   : 0.0   Min.   :16.00  
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0   1st Qu.:22.00  
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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       Hippocampus       MidTemp       mPACCtrailsB        HMSCORE     
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000  
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000  
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000  
 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.:1.000  
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000  
> view(amerge_subset)
> 
```

# FILES, PLOTS, HELP

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external | | |
| misc | | |
| output | | |
| R | | |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd | | |



Find: PTRACCAT

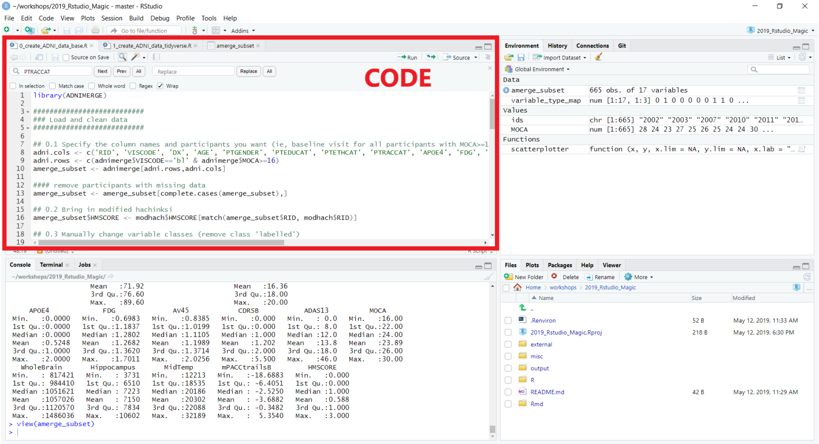
```r
library(ADNIMERGE)

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

## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA>=1
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT', 'PTRACCAT', 'APOE4', 'FDG', '
adni.rows <- c(adnimerge$VISCODE=='bl' & adnimerge$MOCA>=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 hachinksi
amerge_subset$HMSCORE <- modhach$HMSCORE[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
```

```
                 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.   :0.000   Min.   : 0.0   Min.   :16.00  
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0   1st Qu.:22.00  
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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       Hippocampus        MidTemp        mPACCtrailsB        HMSCORE     
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000  
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000  
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000  
 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.:1.000  
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000  
> view(amerge_subset)
> 
```

```
Data
  amerge_subset        665 obs. of 17 variables
  variable_type_map    num [1:17, 1:3] 0 1 0 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
  scatterplotter       function (x, y, x.lim = NA, y.lim = NA, x.lab = "…
```

**VARIABLES, HISTORY, VERSION CONTROL**

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external | | |
| misc | | |
| output | | |
| R | | |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd | | |



```
library(ADNIMERGE)

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

## 0.1 Specify the column names and participants you want (ie, baseline visit for all participants with MOCA>=1
adni.cols <- c('RID', 'VISCODE', 'DX', 'AGE', 'PTGENDER', 'PTEDUCAT', 'PTETHCAT', 'PTRACCAT', 'APOE4', 'FDG', '
adni.rows <- c(adnimerge$VISCODE=='bl' & adnimerge$MOCA>=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 hachinksi
amerge_subset$HMSCORE <- modhach$HMSCORE[match(amerge_subset$RID, modhach$RID)]

## 0.3 Manually change variable classes (remove class 'labelled')
```

Data

```
amerge_subset         665 obs. of 17 variables
  variable_type_map   num [1:17, 1:3] 0 1 0 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 ...
```

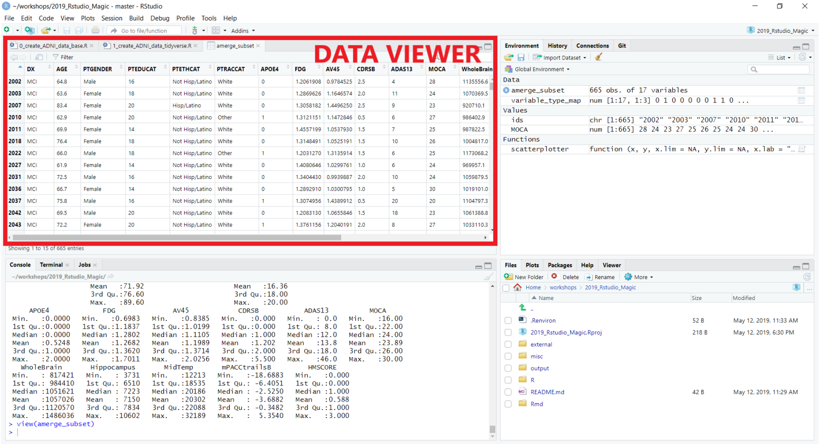
Functions

```
  scatterplotter      function (x, y, x.lim = NA, y.lim = NA, x.lab = "…
```

~/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.   :0.000   Min.   : 0.0    Min.   :16.00  
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0    1st Qu.:22.00  
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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      Hippocampus        MidTemp       mPACCtrailsB         HMSCORE     
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000  
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000  
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000  
 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.:1.000  
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000  
> view(amerge_subset)
> 
```

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external | | |
| misc | | |
| output | | |
| R | | |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd | | |


~/workshops/2019_Rstudio_Magic - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
2019_Rstudio_Magic
0_create_ADNI_data_base.R
1_create_ADNI_data_tidyverse.R
amerge_subset
Filter
DATA VIEWER
Environment History Connections Git
Import Dataset
List
Global Environment
Console Terminal Jobs
Files Plots Packages Help Viewer
New Folder Delete Rename More
Home > workshops > 2019_Rstudio_Magic

|  | DX | AGE | PTGENDER | PTEDUCAT | PTETHCAT | PTRACCAT | APOE4 | FDG | AV45 | CDRSB | ADAS13 | MOCA | WholeBrain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2002 | MCI | 64.8 | Male | 16 | Not Hisp/Latino | White | 0 | 1.2061908 | 0.9784525 | 2.5 | 4 | 28 | 1135556.6 |
| 2003 | MCI | 63.6 | Female | 18 | Not Hisp/Latino | White | 0 | 1.2869626 | 1.1646574 | 2.0 | 11 | 24 | 1070369.5 |
| 2007 | MCI | 83.4 | Female | 20 | Hisp/Latino | White | 0 | 1.3058182 | 1.4496250 | 2.5 | 9 | 23 | 920710.1 |
| 2010 | MCI | 62.9 | Female | 20 | Not Hisp/Latino | Other | 1 | 1.3121151 | 1.1472846 | 0.5 | 6 | 27 | 986402.9 |
| 2011 | MCI | 69.9 | Female | 14 | Not Hisp/Latino | White | 0 | 1.4557199 | 1.0537930 | 1.5 | 7 | 25 | 987822.5 |
| 2018 | MCI | 76.4 | Female | 18 | Not Hisp/Latino | White | 0 | 1.3148491 | 1.0525191 | 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 | 1.4080646 | 1.0299761 | 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 | 1.2892910 | 1.0300795 | 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 | 69.5 | Male | 20 | Not Hisp/Latino | White | 0 | 1.2083130 | 1.0655846 | 1.5 | 18 | 23 | 1061388.8 |
| 2043 | MCI | 72.2 | Female | 20 | Not Hisp/Latino | White | 1 | 1.3761156 | 1.2040191 | 2.0 | 8 | 27 | 1033110.3 |

Showing 1 to 15 of 665 entries

~/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.   :0.000   Min.   : 0.0   Min.   :16.00  
 1st Qu.:0.0000   1st Qu.:1.1837   1st Qu.:1.0199   1st Qu.:0.000   1st Qu.: 8.0   1st Qu.:22.00  
 Median :0.0000   Median :1.2802   Median :1.1105   Median :1.000   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.:2.000   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       Hippocampus        MidTemp       mPACCtrailsB         HMSCORE     
 Min.   : 817421   Min.   : 3731   Min.   :12213   Min.   :-18.6883   Min.   :0.000  
 1st Qu.: 984410   1st Qu.: 6510   1st Qu.:18535   1st Qu.: -6.4051   1st Qu.:0.000  
 Median :1051621   Median : 7223   Median :20186   Median : -2.5250   Median :1.000  
 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.:1.000  
 Max.   :1486036   Max.   :10602   Max.   :32189   Max.   :  5.3540   Max.   :3.000  
> view(amerge_subset)
> 
```

```
Data
amerge_subset        665 obs. of 17 variables
  variable_type_map  num [1:17, 1:3] 0 1 0 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
  scatterplotter     function (x, y, x.lim = NA, y.lim = NA, x.lab = "…
```

| Name | Size | Modified |
|---|---|---|
| .. |  |  |
| .Renviron | 52 B | May 12, 2019, 11:33 AM |
| 2019_Rstudio_Magic.Rproj | 218 B | May 12, 2019, 6:30 PM |
| external |  |  |
| misc |  |  |
| output |  |  |
| R |  |  |
| README.md | 42 B | May 12, 2019, 11:29 AM |
| Rmd |  |  |

# Benefits of RStudio

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

# Benefits of RStudio

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

# Benefits of RStudio

- Built-in integration with version control (git or SVN)
- R Markdown
  - Save and execute code

# Benefits of RStudio

- Built-in integration with version control (git or SVN)
- R Markdown
  - Save and execute code
  - Generate high quality reports that can be shared

# Benefits of RStudio

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

# Benefits of RStudio

- Built-in integration with version control (git or SVN)
- R Markdown
  - Save and execute code
  - Generate high quality reports that can be shared
  - Create presentations (like this one!)
  - Even write papers
  - This workshop

# Benefits of RStudio

- Built-in integration with version control (git or SVN)
- R Markdown
  - Save and execute code
  - Generate high quality reports that can be shared
  - Create presentations (like this one!)
  - Even write papers
  - This workshop
    - See https://github.com/derekbeaton/Workshops/tree/master/Misc/R_RStudio_Workflow

# Benefits of RStudio

- Built-in integration with version control (git or SVN)
- R Markdown
  - Save and execute code
  - Generate high quality reports that can be shared
  - Create presentations (like this one!)
  - Even write papers
  - This workshop
    - See https://github.com/derekbeaton/Workshops/tree/master/Misc/R_RStudio_Workflow
- Python, D3 (JavaScript), SQL, Shiny, LaTeX, Git/SVN, HTML/CSS, and so much more.

# RStudio is more

- Not just an IDE (integrated development environment)

# RStudio is more

- Not just an IDE (integrated development environment)
- A company

# RStudio is more

- Not just an IDE (integrated development environment)
- A company
- A community

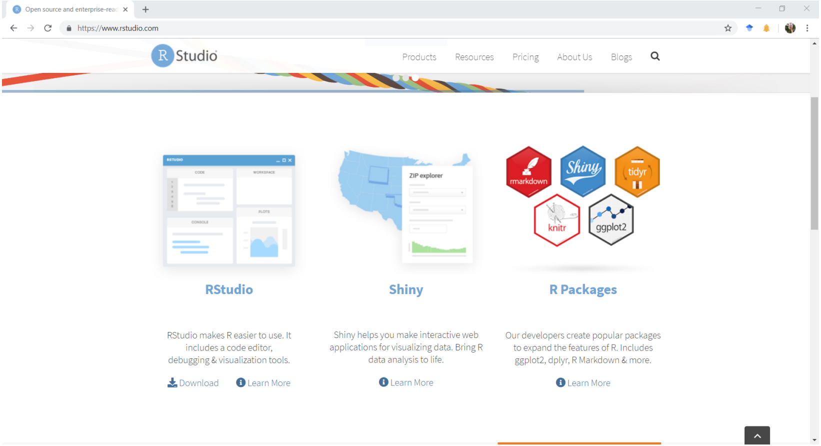
# RStudio is more

- Not just an IDE (integrated development environment)
- A company
- A community
- A conference

# RStudio is more

- Not just an IDE (integrated development environment)
- A company
- A community
- A conference
- A centralized resource

# RStudio Resources



**RStudio**

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

Download Learn More

**Shiny**

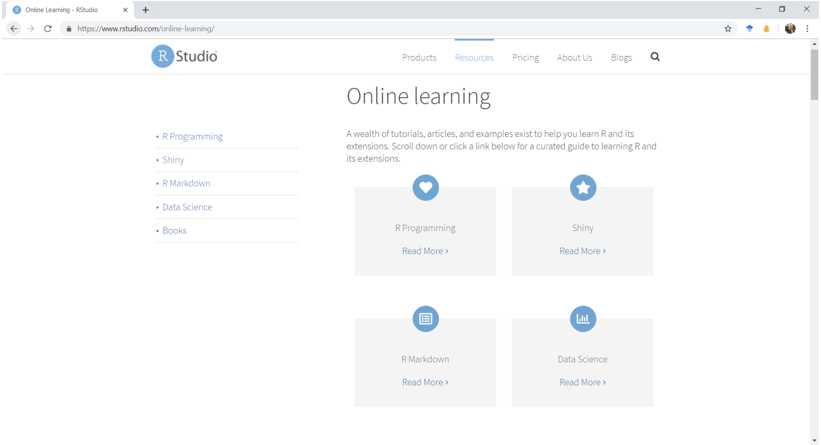
Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life.

Learn More

**R Packages**

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

Learn More



Online Learning - RStudio

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

RStudio

Products Resources Pricing About Us Blogs

# Online learning

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

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

Read More ›

Shiny

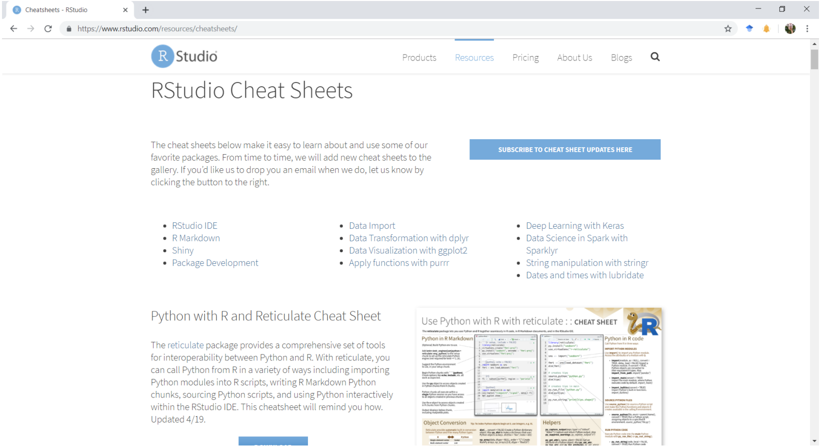
Read More ›

R Markdown

Read More ›

Data Science

Read More ›


Cheatsheets - RStudio
https://www.rstudio.com/resources/cheatsheets/
RStudio
Products
Resources
Pricing
About Us
Blogs
SUBSCRIBE TO CHEAT SHEET UPDATES HERE
Use Python with R with reticulate : : CHEAT SHEET
Python in R Markdown
Python in R code
Object Conversion
Helpers

# 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.

- 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.

# Project and Environment Setup

# RStudio Setup

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

# For safety & collaboration

- RStudio projects

# For safety & collaboration

- RStudio projects
  - “RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.”

# For safety & collaboration

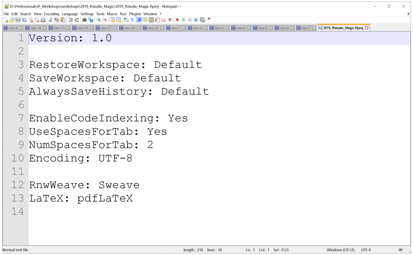
- RStudio projects
  - “RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.”
  - Allows for return to key states

# For safety & collaboration

- RStudio projects
    - “RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.”
    - Allows for return to key states
- .Rproj files

# For safety & collaboration

- RStudio projects
  - "RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents."
  - Allows for return to key states
- .Rproj files
  - Basically a text file with some parameters for start up



D:\Professional\R_Workshops\workshops\2019_Rstudio_Magic\2019_Rstudio_Magic.Rproj - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
new 8
new 16
new 11
new 23
new 27
new 24
new 28
new 1
new 2
new 3
new 4
new 5
new 6
new 7
2019_Rstudio_Magic.Rproj

```
Version: 1.0

RestoreWorkspace: Default
SaveWorkspace: Default
AlwaysSaveHistory: Default

EnableCodeIndexing: Yes
UseSpacesForTab: Yes
NumSpacesForTab: 2
Encoding: UTF-8

RnwWeave: Sweave
LaTeX: pdfLaTeX

```

Normal text file
length : 218 lines : 14
Ln : 1 Col : 1 Sel : 0 | 0
Windows (CR LF)
UTF-8
IN

# Projects

Create a new one for:

- a folder

# Projects

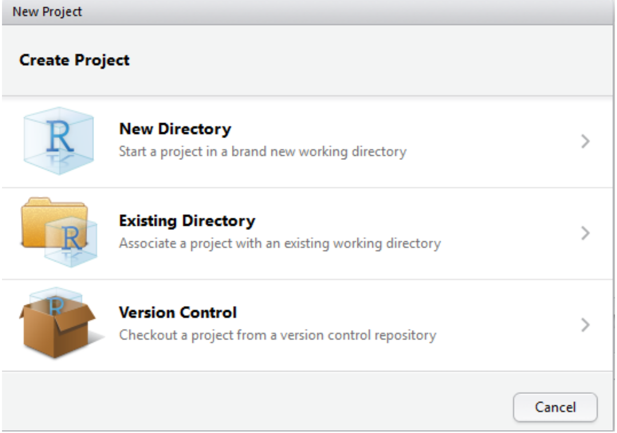
Create a new one for:

- a folder
- packages

# Projects

Create a new one for:

- a folder
- packages
- (and from) git repos:



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

# What is Git?

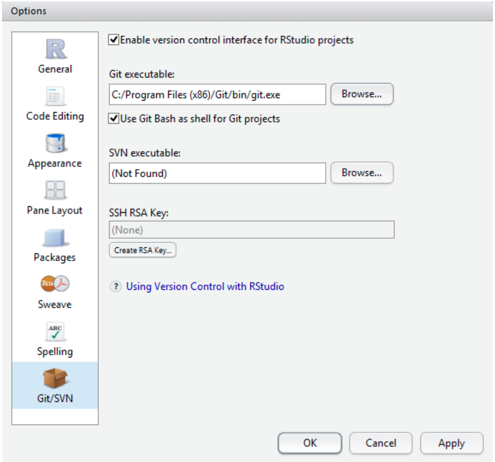
SOMETHING

# Git & Projects

- Git

# Git & Projects

- Git
- Download git and link executable within RStudio



Options

☑ Enable version control interface for RStudio projects

Git executable:

C:/Program Files (x86)/Git/bin/git.exe

☑ Use Git Bash as shell for Git projects

SVN executable:

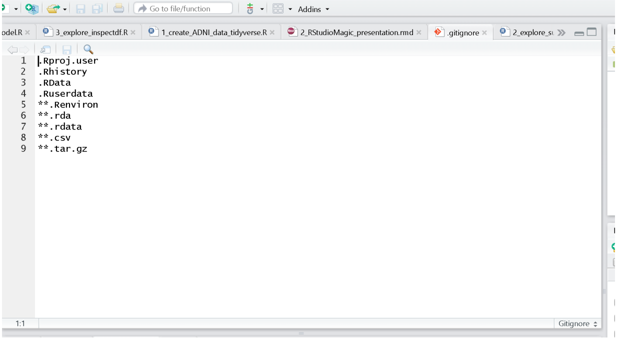
(Not Found)

SSH RSA Key:

(None)

# Format .gitignore

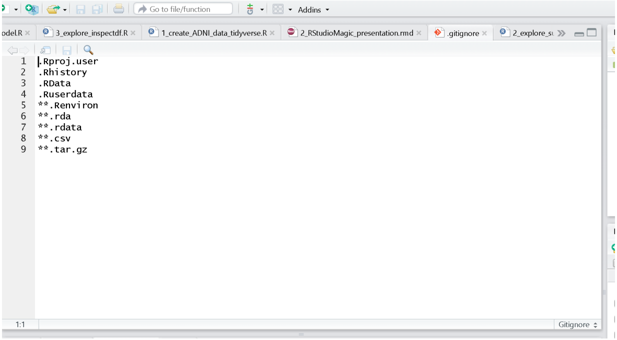
- File types to ignore via version control



```
.Rproj.user
.Rhistory
.RData
.Ruserdata
**.Renviron
**.rda
**.rdata
**.csv
**.tar.gz
```

# Format .gitignore

- File types to ignore via version control
- ∗∗ before each extentions will match directories anywhere in the repo



```
.Rproj.user
.Rhistory
.RData
.Ruserdata
**.Renviron
**.rda
**.rdata
**.csv
**.tar.gz
```