What They Forgot to Teach You About R

RStudio

2018 October Seattle

What They Forgot to Teach You About R

RStudio

2018 October Seattle

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

To view a copy of this license, visit
http://creativecommons.org/licenses/by-sa/4.0/

Thurs Oct 4 & Fri Oct 5, 2018

- 8am setup & breakfast
- 9-10:30am work session 1
- 10:30-11am break
- 11-12:30pm work session 2
- 12:30-1:30pm lunch
- 1:30-3pm work session 3
- 3-3:30pm break
- 3:30-5pm work session 4
- 5:20-6ish office "hours" (Thurs only)

Yes, materials are available.

All will be revealed and downloaded.

You have my word.

Jennifer (Jenny) Bryan
RStudio (& UBC)

@jennybc
@JennyBryan

Jim Hester
RStudio

@jimhester
@jimhester_

Alison Hill
OHSU (& special friend of RStudio)

apreshill
apreshill

rstd.io/wtf-seattle

office "hours" sign up

http://bit.ly/wtf-seattle-2018

Everyone is encouraged to open issues here:

rstd.io/wtf-seattle

https://github.com/jennybc/whattheyforgot-seattle

Record glitches, gotchas,

good sidebar discussions, etc.

to address now or later.

Go here now:

rstd.io/wtf-seattle

https://github.com/jennybc/whattheyforgot-seattle

Go to the Issues.

What's your OS? Put an emoji on your OS.

What if I'm in over my head?

What if I could teach this workshop?

Meet your helpers

Meet your helpers*

* your neighbours

Go here now:

rstd.io/wtf-seattle

https://github.com/jennybc/whattheyforgot-seattle

Click on https://gitter.im/what-they-forgot/seattle-2018

Live chat -- totally optional!

Day 1, morning

Jennifer (Jenny) Bryan
RStudio (& UBC)

@jennybc
@JennyBryan

What *Did* They Forget to Teach You?

Everything else

Statistical analysis

Deep Thoughts

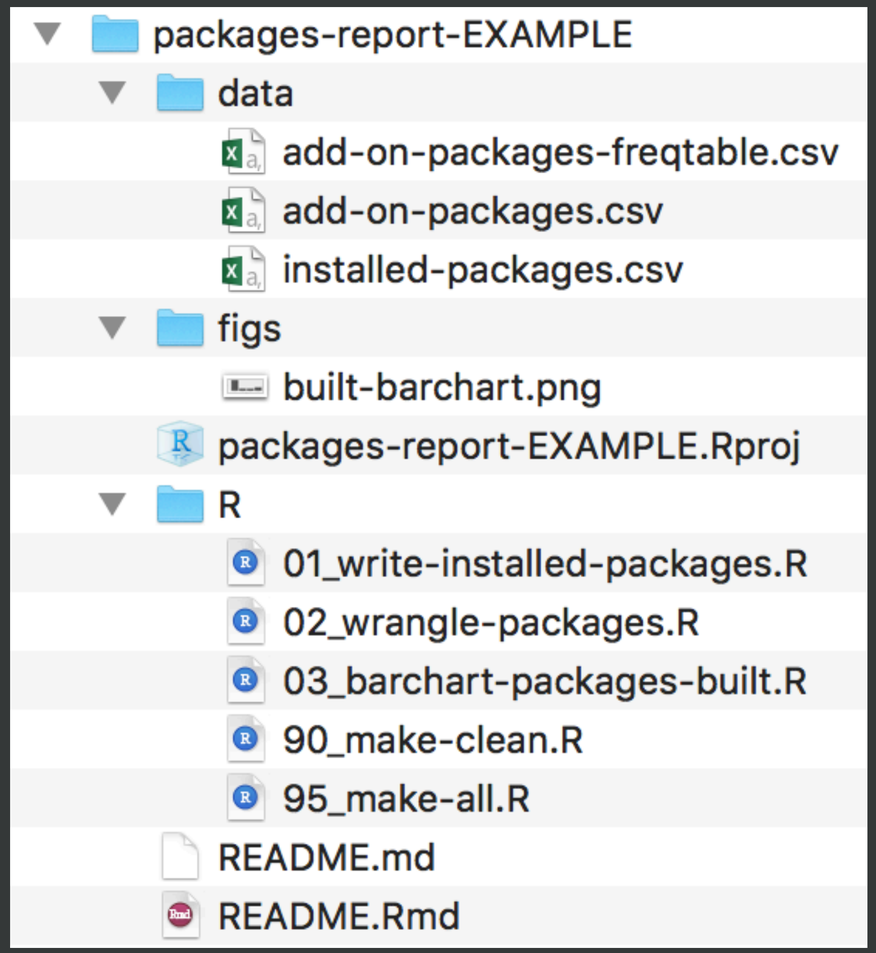
Be organized

do this as you go, not "tomorrow"

but also don't fret over past mistakes
raise the bar for *new* work

Be organized

self-explaining >>> wordy, needy explainers



>>>

file salad

+ an out-of-date README

Good enough practices in scientific computing

Wilson, Bryan, Cranston, Kitzes, Nederbragt, Teal

https://doi.org/10.1371/journal.pcbi.1005510

http://bit.ly/good-enuff

Day 1 Practical Example: Explore your R installation

R package = the natural unit for distributing R code

base R ≈ 14 base + 15 recommended packages ship with all binary distributions of R

can use right out of the box:

```
library(lattice)
```

CRAN has > 13K additional packages

install, then attach:

```
install.packages("devtools")
library(devtools)
```

And then there's GitHub ...

install via devtools, then attach:

```
devtools::install_github("tidyverse/dplyr")
library(dplyr)
```

Where do packages live locally?

By default, in the default library

`.Library`

All libraries for current session:

```
.libPaths()
```

All installed packages:

```
installed.packages()
```

```
install.packages("usethis")
library(usethis)
use_course("rstd.io/wtf-1")
```

```
install.packages("usethis")
library(usethis)
use_course("rstd.io/wtf-1")
```

Pick one to open and flesh out:

```
01_explore-libraries_spartan.R
01_explore-libraries_comfy.R*
```

* worst case, there's always `jenny`

```
install.packages("usethis")
library(usethis)
use_course("rstd.io/wtf-1")
```

Stay relaxed.

We will refine this code and where it lives soon enough.

You **<u>want</u>** to leave rough edges and gaps to address later.

work on challenge

Deep Thoughts

Adopt a project-oriented workflow

Why?

- work on more than 1 thing at a time
- collaborate, communicate, distribute
- start and stop

Adopt a project-oriented workflow

How?

- dedicated directory
- RStudio **P**roject
- Git repo, probably syncing to a remote

If you do this at the top of your scripts, I might set your computer on 🔥:

```
setwd("C:\Users\jenny\path\that\only\I\have")
rm(list = ls())
```

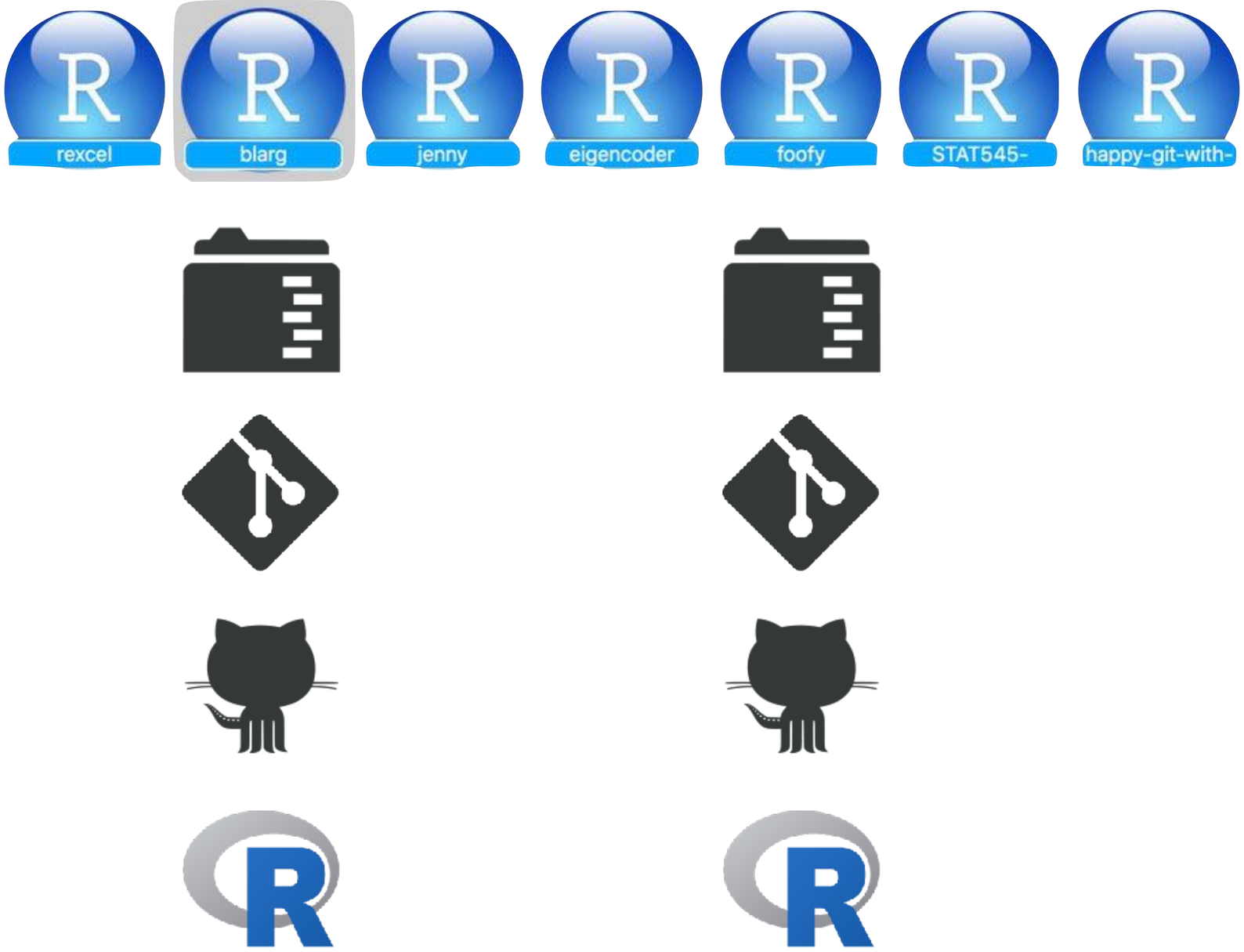
Project-oriented workflow designs this away:

https://www.tidyverse.org/articles/2017/12/workflow-vs-script/

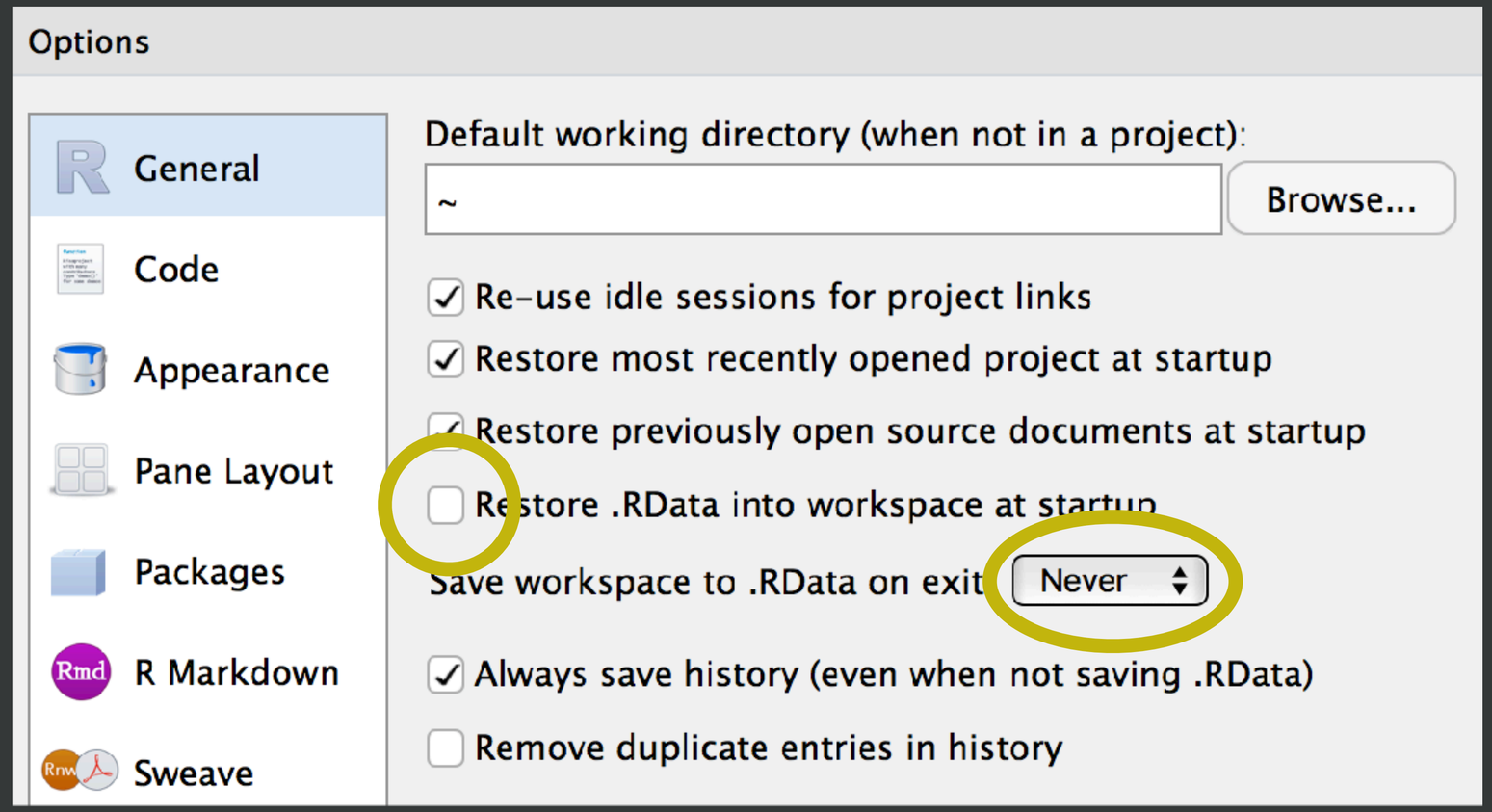
https://whattheyforgot.org ← see "A holistic workflow"

What does it mean to be an RStudio **P**roject?

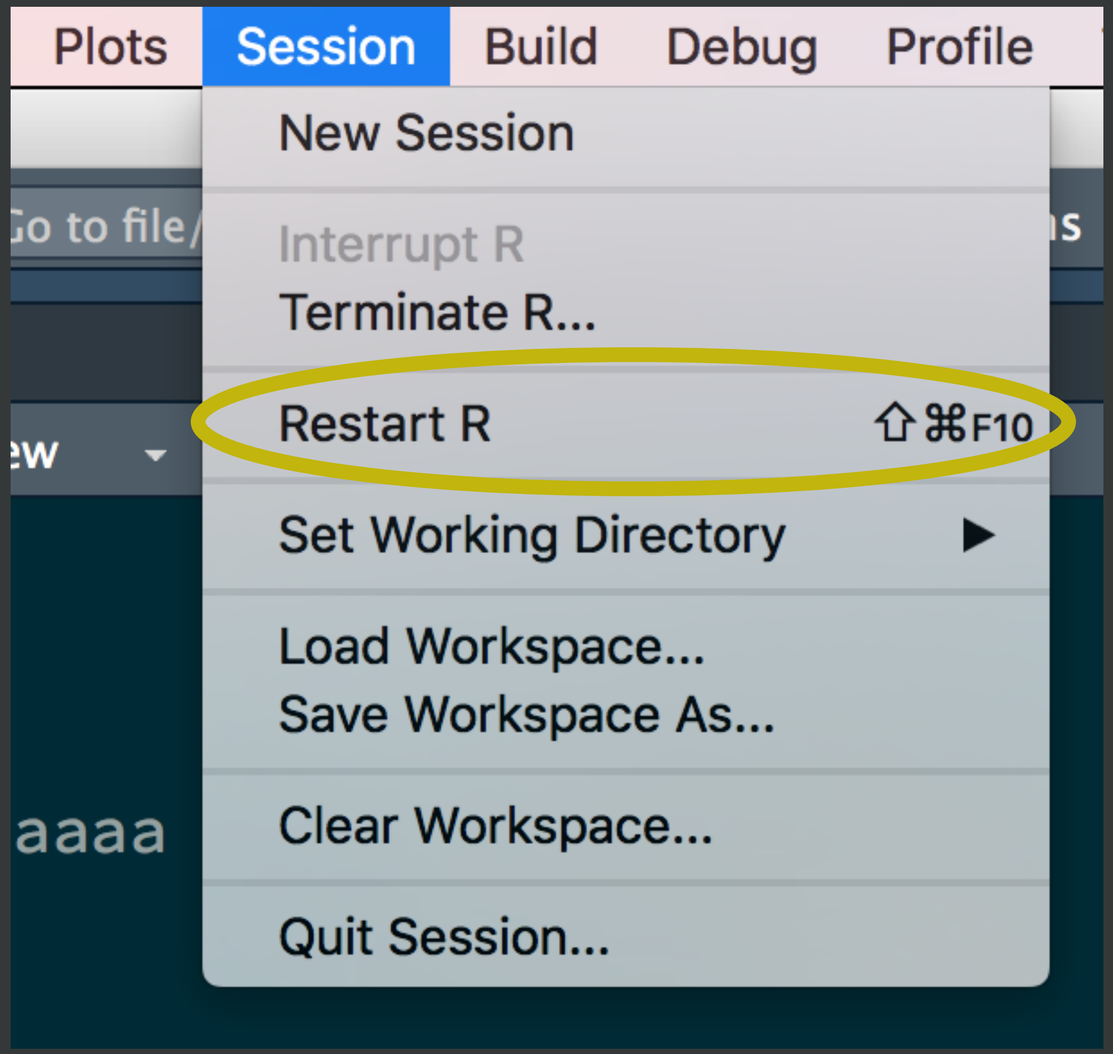
- RStudio leaves notes to itself in `foo.Rproj`
- Open Project = dedicated instance of RStudio
  - dedicated R process
  - file browser pointed at Project directory
  - working directory set to Project directory


rexcel
blarg
jenny
eigencoder
foofy
STAT545-
happy-git-with-

Use a "blank slate"



Restart R often



Project initiation: the local case

Pick one and do this now!

New folder + make it an RStudio Project

- `usethis::create_project("~/i_am_new")`
- RStudio > New Project... > New Directory > New Project

Make existing folder into an RStudio Project

- `usethis::create_project("~/i_exist")`
- RStudio > New Project... > Existing Directory

break here?

Deep Thoughts

Practice "safe paths"

Practice "safe paths"

relative to a stable base

use file system functions,

not `paste()`, `strsplit()`, etc.

Do you know where your files are?

What is working directory?

https://devs.mailchimp.com/blog/the_freddie_mercury_project/

Our second launch was supposed to be The One, but we ran into a critical issue with some of our code.... The code worked fine as we were testing on command line since the location of PHP was known by our shell, but once the code was added to cron for automation, the **location of PHP wasn't known, and the scripts continuously failed to send.**

```
install.packages("fs")
install.packages("here")
```

fs = filepath handling

here = project-relative paths

Examples of a stable base

- Project directory
  - `here::here("data", "raw-data.csv")`
- User's home directory
  - `file.path("~", ...)`
  - `fs::path_home(...)`
- Official location for installed s/w
  - `library(thingy)`
  - `system.file(..., package = "thingy")`

I have nothing against absolute paths.
Some of my best friends are absolute paths!
But don't hard-wire them into your scripts.
Instead, form at runtime relative to a stable base.

```
> (GOOD <- fs::path_home("tmp/test.csv")
[1] "/Users/jenny/tmp/test.csv"

> (BAD <- "/Users/jenny/tmp/test.csv")
[1] "/Users/jenny/tmp/test.csv"
```

Practice "safe paths"

Use the here package to build paths inside a project.

Leave working directory at top-level at all times, during development.

Absolute paths are formed at runtime.

```r
library(here)
#> here() starts at <snip, snip>/here-demo
system("tree")
#> .
#>  └── one
#>      └── two
#>          └── awesome.txt

here("one", "two", "awesome.txt")
#> [1] "<snip, snip>/here-demo/one/two/awesome.txt"
cat(readLines(here("one", "two", "awesome.txt")))
#> OMG this is so awesome!

setwd(here("one"))
getwd()
#> [1] "<snip, snip>/here-demo/one"
here("one", "two", "awesome.txt")
#> [1] "<snip, snip>/here-demo/one/two/awesome.txt”
cat(readLines(here("one", "two", "awesome.txt")))
#> OMG this is so awesome!
```

```r
ggsave(here("figs", "built-barchart.png"))
```

Works on my machine, works on yours!

Works even if working directory is in a sub-folder

Works for RStudio projects, Git repos, R packages, …

Works with knitr / rmarkdown

break here?

Deep Thoughts

Names matter

machine readable

human readable

sort nicely

😔

myabstract.docx
Joe’s Filenames Use Spaces and Punctuation.xlsx
figure 1.png
homework1.R
JW7d^(2sl@deletethisandyourcareerisoverWx2*.txt

😍

2018-01_bryan-abstract-rstudio-conf.docx
joes-filenames-are-getting-better.xlsx
fig01_scatterplot-talk-length-vs-interest.png
bryan_hw01.R
1986-01-28_raw-data-from-challenger-o-rings.txt

"machine readable"

regular expression and globbing friendly

- avoid spaces, punctuation, accented characters, case sensitivity

easy to compute on

- deliberate use of delimiters

"human readable"

name contains info on content
name anticipates context
concept of a slug from user-friendly URLs

1986-01-28_raw-data-from-challenger-o-rings.txt

"sort nicely"

put something numeric in there
left pad with zeros for constant width
use the ISO 8601 standard for dates
order = chronological or ... consider common sense

```
> ft <- tibble(files = dir_ls(glob = "*.R"))
> ft
# A tibble: 6 x 1
  files
  <fs::path>
1 00_filesystem-practice_comfy.R
2 00_filesystem-practice_jenny.R
3 00_filesystem-practice_spartan.R
4 01_explore-libraries_comfy.R
5 01_explore-libraries_jenny.R
6 01_explore-libraries_spartan.R
```

👍 file names

```
> ft <- tibble(files = dir_ls(glob = "*.R"))
> ft
# A tibble: 6 x 1
  files
  <fs::path>
1 00_filesystem-practice_comfy.R
2 00_filesystem-practice_jenny.R
3 00_filesystem-practice_spartan.R
4 01_explore-libraries_comfy.R
5 01_explore-libraries_jenny.R
6 01_explore-libraries_spartan.R
```

Anyone can guess at file's purpose

```
> ft %>%
+   filter(str_detect(files, "explore"))
# A tibble: 3 x 1
  files
  <fs::path>
1 01_explore-libraries_comfy.R
2 01_explore-libraries_jenny.R
3 01_explore-libraries_spartan.R
```

Easy to filter in R (or the shell or whatever)

```
> ft %>%
+   mutate(files = path_ext_remove(files)) %>%
+   separate(files, into = c("i", "topic", "flavor"), sep = "_")
# A tibble: 6 x 3
  i     topic               flavor
* <chr> <chr>               <chr>
1 00    filesystem-practice comfy
2 00    filesystem-practice jenny
3 00    filesystem-practice spartan
4 01    explore-libraries   comfy
5 01    explore-libraries   jenny
6 01    explore-libraries   spartan
```

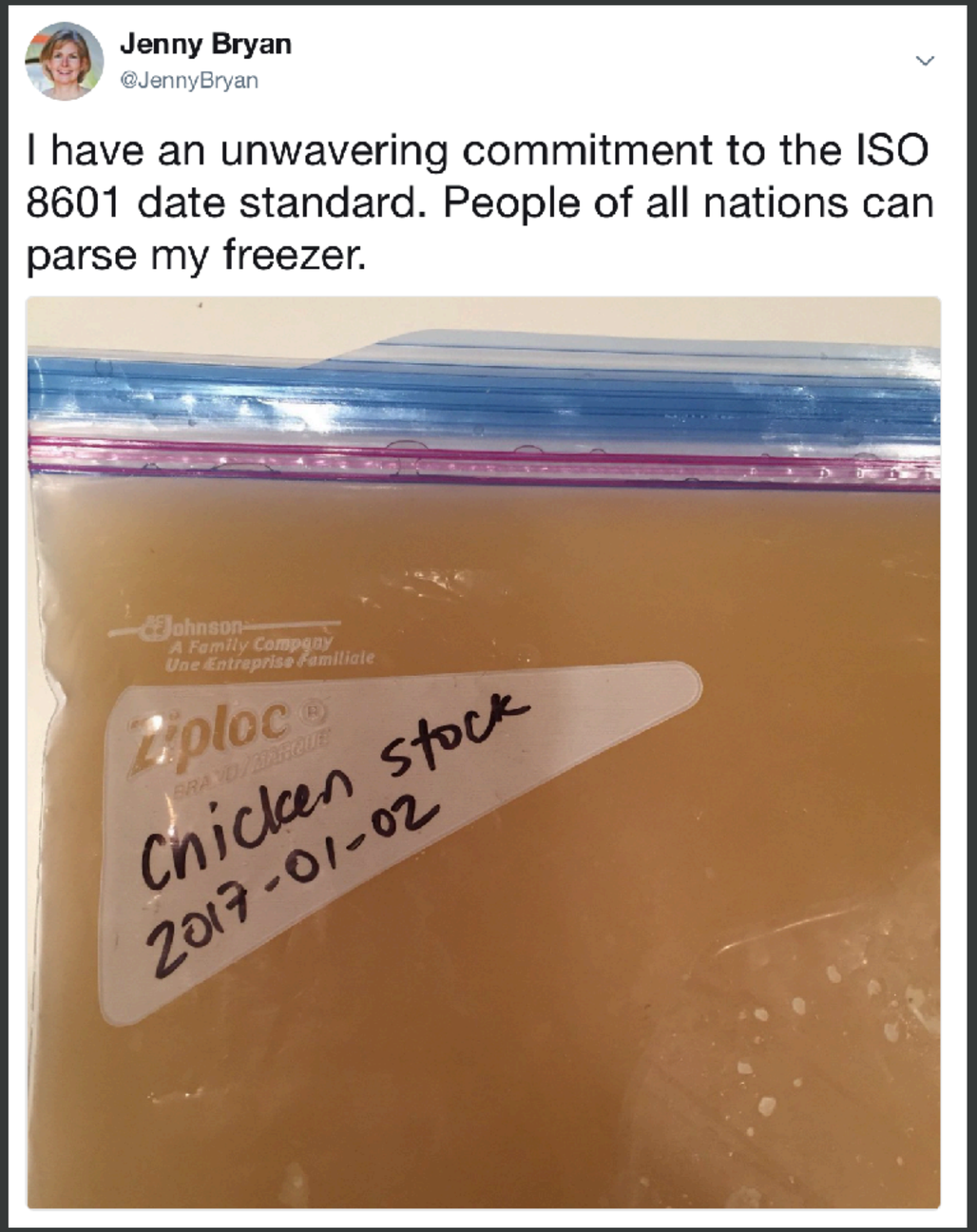
Intentional use of delimiters = meta-data easy to recover

"_" delimits fields

"-"delimits words so my eyes don’t bleed

```
> dirs <- dir_ls(path_home("Desktop"), type = "directory")
> (dt <- tibble(dirs = path_file(dirs)))
# A tibble: 2 x 1
  dirs
  <fs::path>
1 day1_s1_explore-libraries
2 day1_s2_copy-files
```

Sorts in the same order as you experience in real life



YYYY-MM-DD

ISO 8601

Names matter

machine readable

human readable

sort nicely

Names matter

easy to implement NOW

payoffs accumulate as your skills evolve and projects get more complex

Deep Thoughts

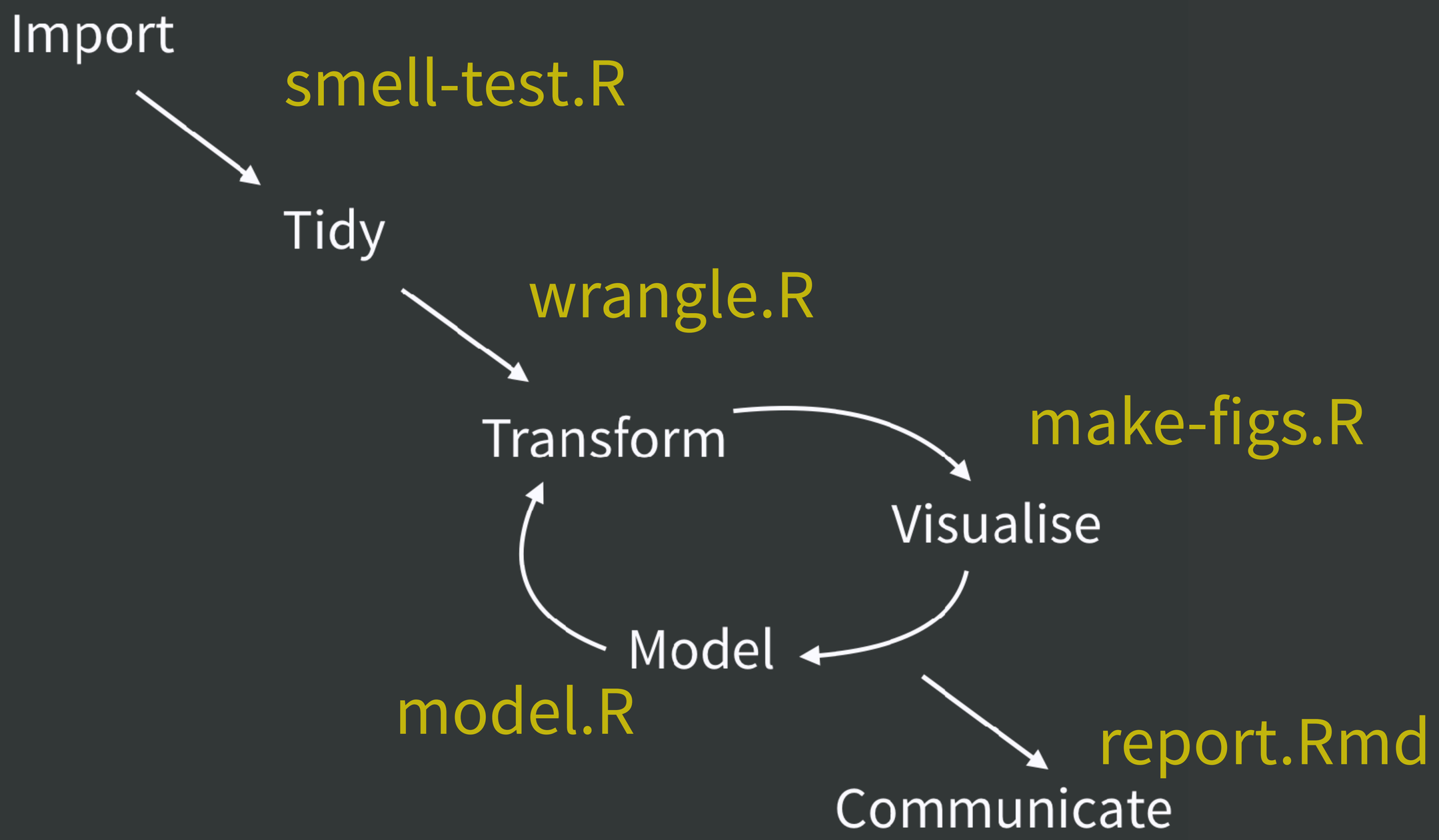
beware of monoliths

break logic & output into pieces

smell-test.R
wrangle.R
model.R
make-figs.R
report.Rmd

>>>

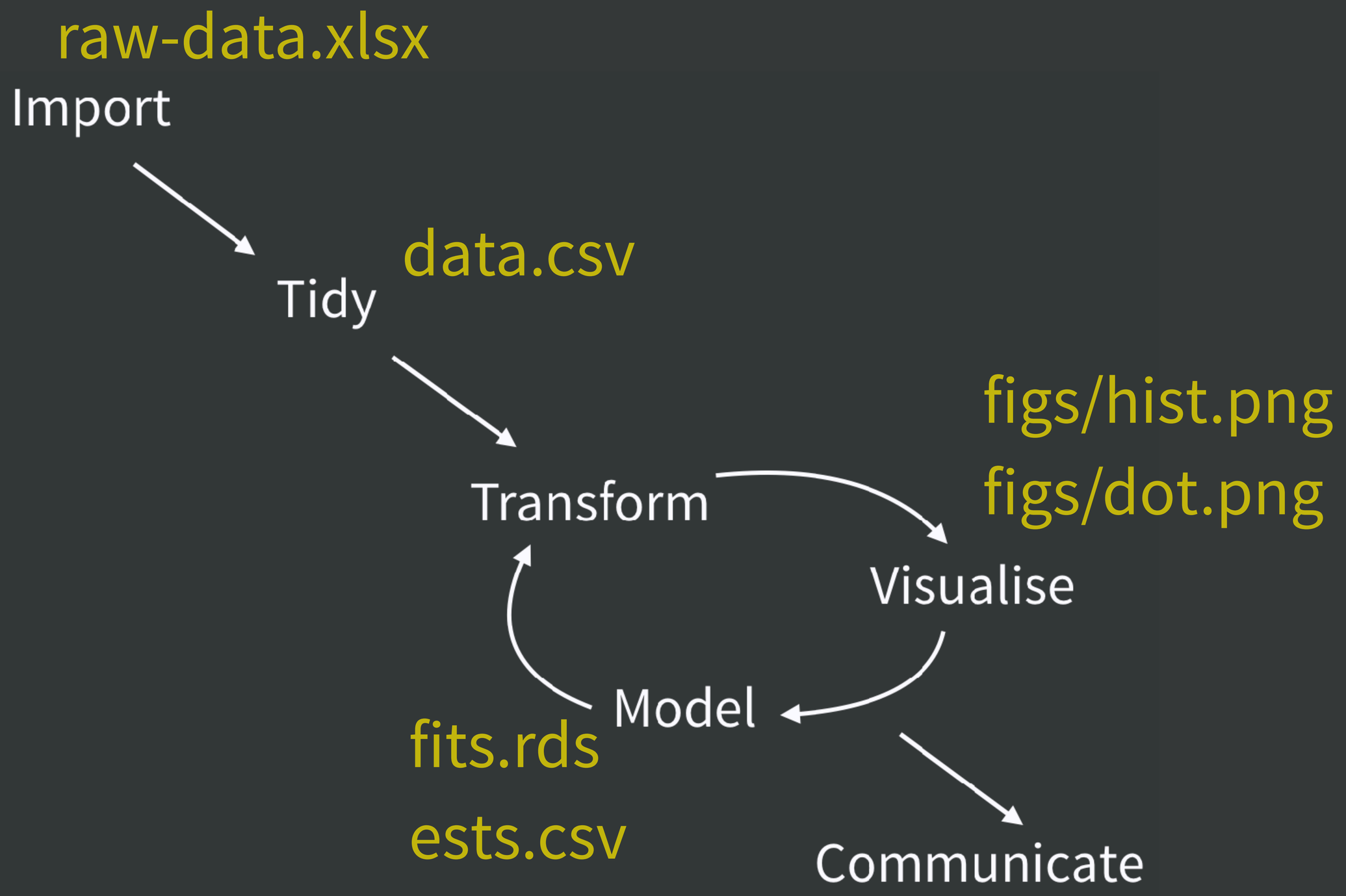
everything.R


Import
smell-test.R
Tidy
wrangle.R
Transform
make-figs.R
Visualise
Model
model.R
report.Rmd
Communicate

raw-data.xlsx
data.csv
fits.rds
ests.csv

>>>

.Rdata


raw-data.xlsx
Import
Tidy
data.csv
Transform
figs/hist.png
figs/dot.png
Visualise
Model
fits.rds
ests.csv
Communicate

Input	Code	Output
raw data	`smell-test.R`	*wisdom*
raw data	`wrangle.R`	`data.csv`
`data.csv`	`model.R`	`fits.rds` `ests.csv`

`data.csv`
`fits.rds`
`ests.csv`
	`make-figs.R`	`figs/*`

`figs/*`
`ests.csv`
	`report.Rmd`	`report.html`
		`report.docx`
		`report.pdf`

a humane API
for your analysis

```
library(usethis)
use_course("rstd.io/wtf-2")
```

- you know the drill
- download to same location as previous
- should open as RStudio project
- if not, you can make it an RStudio project

Project initiation strategies, the local case

New folder + make it an RStudio Project

- `usethis::create_project("~/i_am_new")`
- RStudio > New Project... > New Directory > New Project

Make existing folder into an RStudio Project

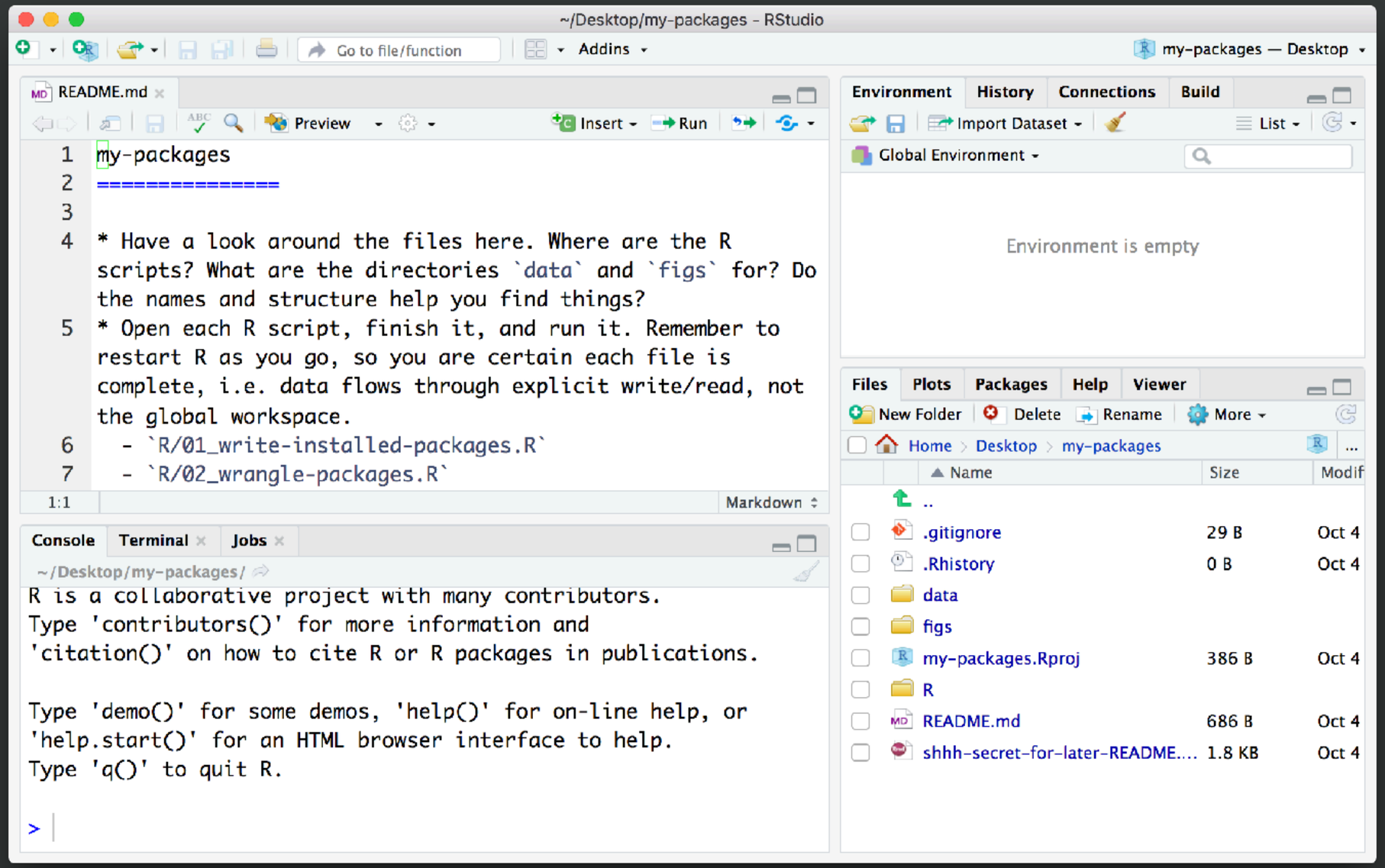
- `usethis::create_project("~/i_exist")`
- RStudio > New Project... > Existing Directory

You should now be in RStudio,
in the new RStudio Project named
my-packages.

How to launch an RStudio Project?

- double-click on .Rproj file
- use File > Open Project and friends
- use Project drop down in upper right corner
- Alfred trick 😉

We are here!


~/Desktop/my-packages - RStudio
Go to file/function
Addins
my-packages — Desktop
README.md
Preview
Insert
Run
my-packages
==========
* Have a look around the files here. Where are the R scripts? What are the directories `data` and `figs` for? Do the names and structure help you find things?
* Open each R script, finish it, and run it. Remember to restart R as you go, so you are certain each file is complete, i.e. data flows through explicit write/read, not the global workspace.
- `R/01_write-installed-packages.R`
- `R/02_wrangle-packages.R`
1:1
Markdown
Console
Terminal
Jobs
~/Desktop/my-packages/
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.
>
Environment
History
Connections
Build
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Desktop
my-packages
Name
Size
Modif
..
.gitignore
29 B
Oct 4
.Rhistory
0 B
Oct 4
data
figs
my-packages.Rproj
386 B
Oct 4
R
README.md
686 B
Oct 4
shhh-secret-for-later-README....
1.8 KB
Oct 4

work on challenge

README.md has instructions