

What They Forgot to Teach You About R

rstudio::conf
from R Studio

rstudio::conf 2018 San Diego

Training Days

<https://www.rstudio.com/conference/>

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/>

Yes, all materials are available.

rstd.io/forgot

Wed Jan 31 & Thurs Feb 1, 2018

- 8am registration & light breakfast
- 9am We start!
- 10:30-11am snacks and coffee
- 12-1pm lunch
- 3-3:30pm snacks and coffee
- 5pm We stop!

Everyone is encouraged to open issues here:

 rstd.io/forgot

<https://github.com/jennybc/what-they-forgot>

Record glitches, gotchas,
good sidebar discussions, etc.
to address now or later.

Day 2, morning

Jennifer Bryan
RStudio, University of British Columbia

 @jennybc
 @JennyBryan

Guest appearance by Jim Hester

How to exploit all the R activity on GitHub

E.g., searching your code, CRAN packages, R itself

Some is here, but we will update w/ workshop goodies:

<http://happygitwithr.com/search.html>



Deep Thoughts



Trevor A. Branch

@TrevorABranch

Follow



My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. #Rstats

Reason to iterate #1:

Get it right! Keep it right!

New data?

New understanding of data?



Reason to iterate #2:

Refine and Extend

Make your code more

- Readable
- Resilient
- Efficient
- General



**beware of
monoliths**



**break logic &
output into
pieces**



smell-test.R

wrangle.R

model.R

>>>

everything.R

make-figs.R

report.Rmd

Import

smell-test.R

Tidy

wrangle.R

Transform

make-figs.R

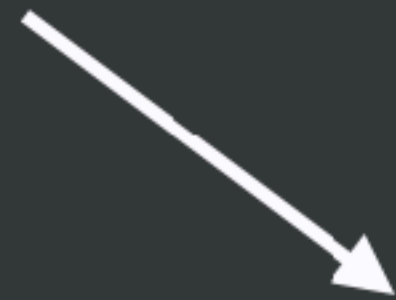
Visualise

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

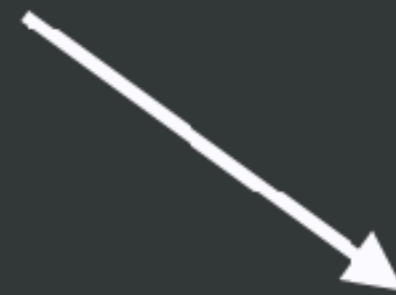
Visualise

fits.rds

Model

ests.csv

Communicate



Input	Code	Output
<i>raw data</i>	smell-test.R	<i>wisdom</i>
<i>raw data</i>	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

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.

```
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!
```

```
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

Next challenge:

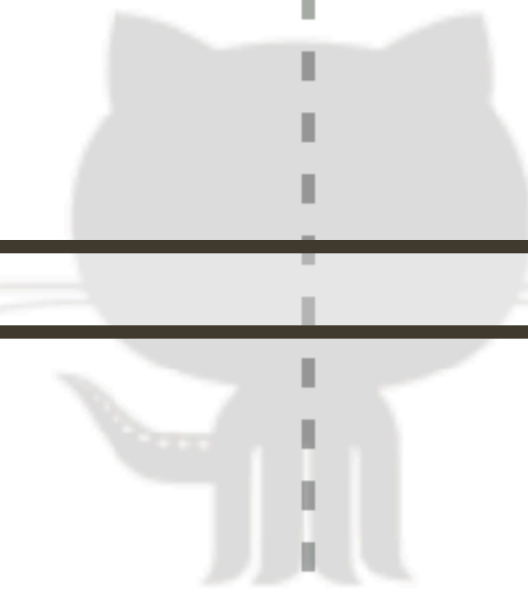
Refine and extend analysis of your libraries

Using Git and GitHub as you go

Practice safe paths

Put a humane API on this analysis

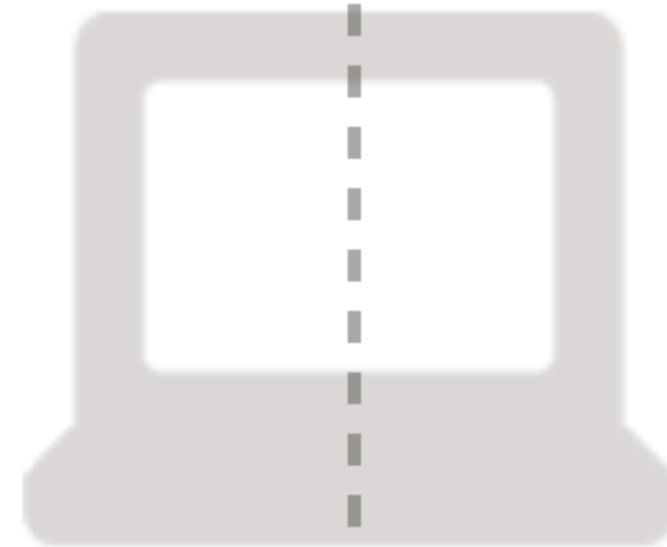
Them




You



not your
problem



"fork and clone"



Go to rstudio.io/forgot_3
<https://github.com/jennybc/packages-report>

Fork it!

Create a new RStudio project [from your fork](#)

Open README.Rmd for your instructions

See how far you can get!

work on challenge

debrief on challenge re: subdirectories,
pseudo Makefiles, and anything else that surfaces

revisit slides we skipped yesterday re: naming
things