

contributing to the tidyverse

Mara Averick
tidyverse developer advocate
@dataandme

CONTRIBUTING TO THE

tidyverse

What is the tidyverse?

The tidyverse is an opinionated collection of R packages designed for data science.

All packages share an underlying design philosophy, grammar, and data structures.

Key FOSS actors

- maintainers
- contributors
- consumers/users
- sustainers

Report team: Ben Nickolis, Pia Mancini, Justin Dorfman, Robert Gibb <<https://sustainoss.org/>>

GitHub HQ (SF) | June 19, 2017

Sustain

A ONE DAY CONVERSATION
for Open Source Software sustainers

THE REPORT

the thoughtful user

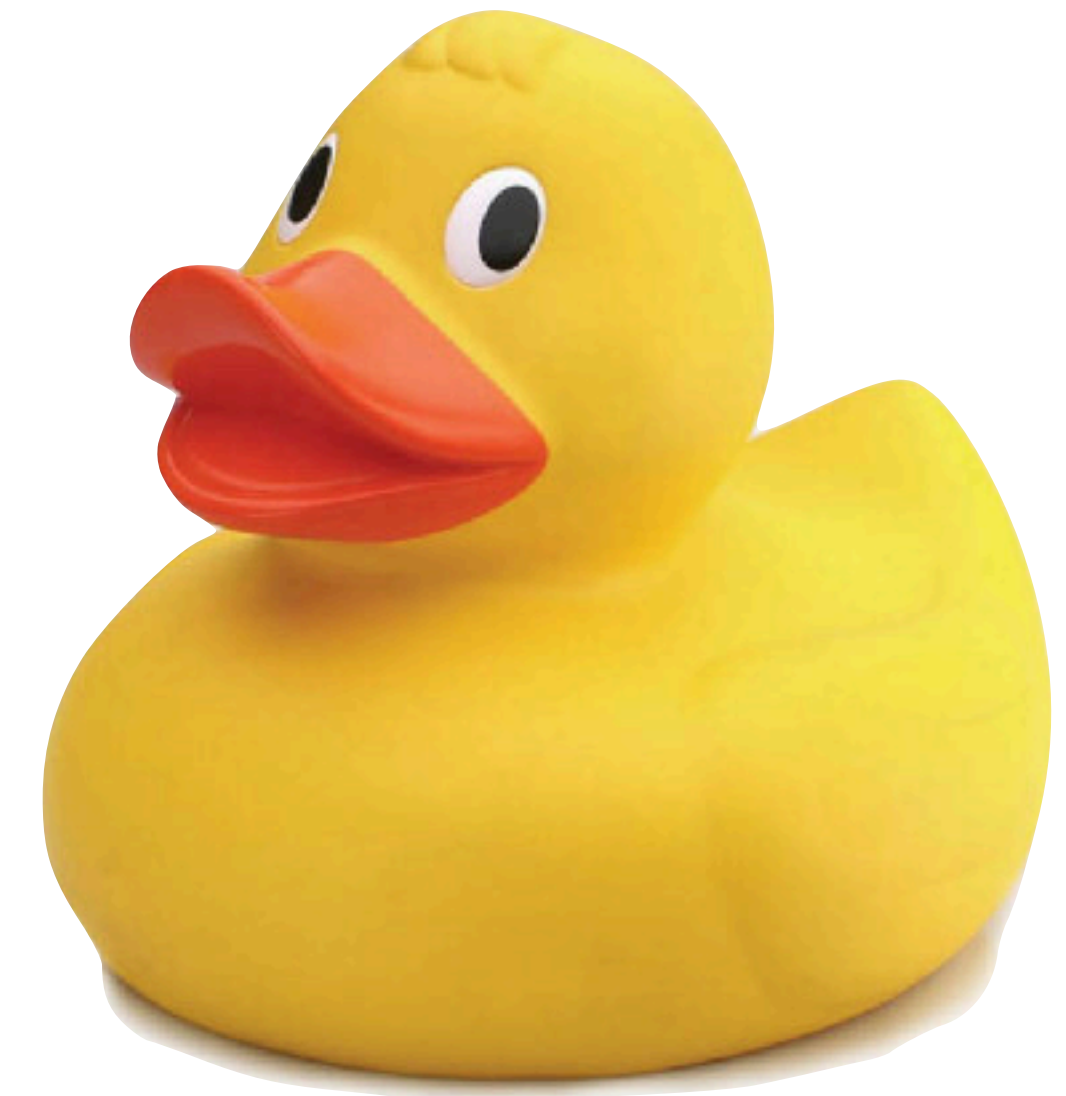
5



asking questions

The most useless problem statement that one can face is "it doesn't work", yet we seem to get it far too often.

– Thiago Maciera



where to ask



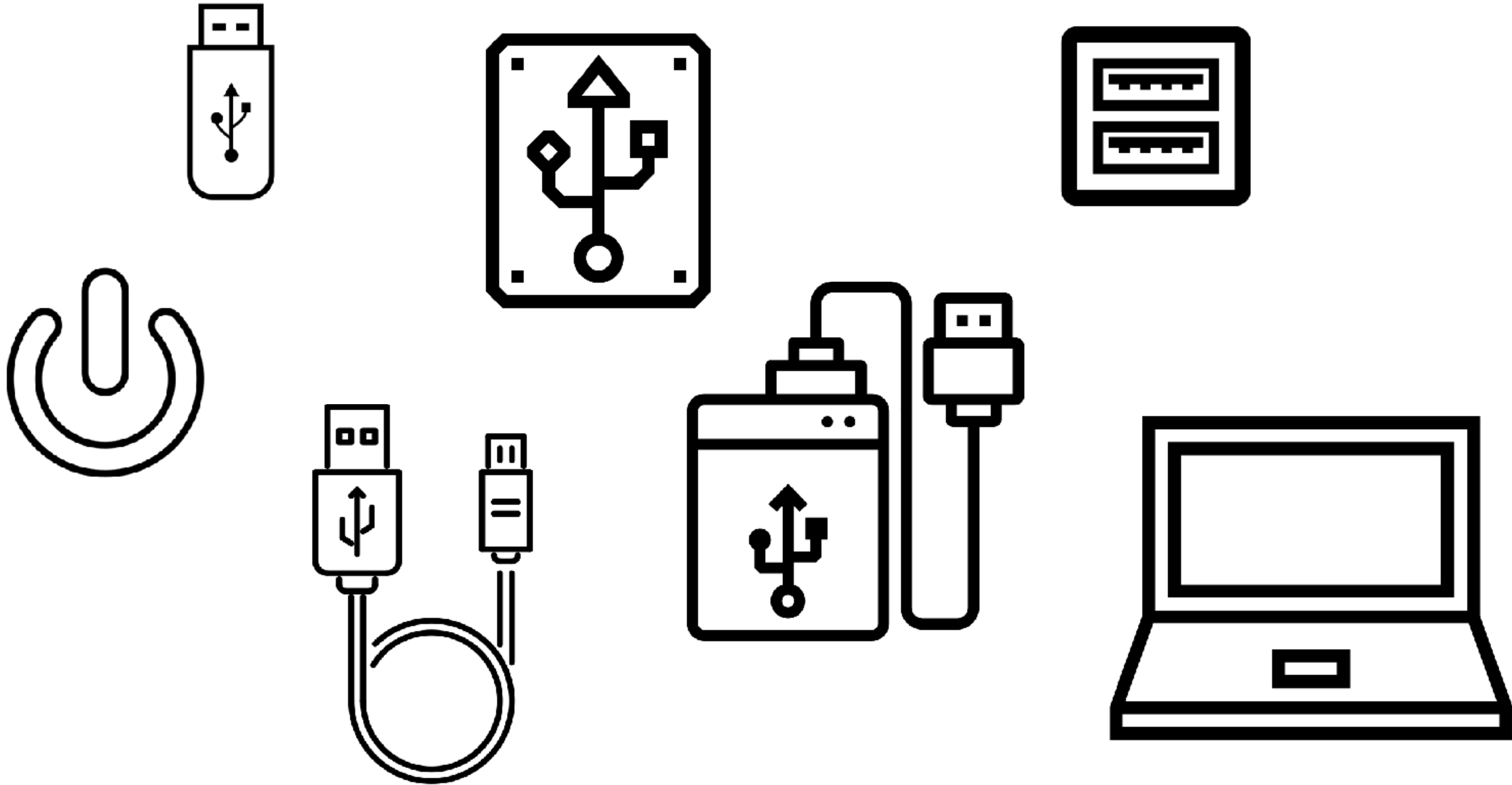
Twitter



RStudio Community



StackOverflow



The newcomer's paradox...



When you ask for help, some friendly soul will no doubt tell you that "*it's easy, just do foo, bar and baz.*" Except for you, it is not easy, there may be no documentation for foo, bar is not doing what it is supposed to be doing and what is this baz thing anyway with its eight disambiguation entries on Wikipedia?

— Leslie Hawthorne

"You'll Eventually Know Everything They've Forgotten." In *Open Advice: FOSS: What We Wish We Had Known When We Started*, edited by Lydia Pintscher, 29–32.

meta-help

10

helping others to help themselves...



meta-help



Why does head() show 6 rows by default?



hadley RStudio Employee

Nov '17

To answer a question like this, I first starting looking through the S books I have on hand (e.g. [The New S Language](#) ¹¹). I don't see `head()` mentioned in the index, so that suggests it's a function introduced by R.

Since it's an R function, I can next search [@winston](#)'s GitHub mirror of the R sources: <https://github.com/wch/r-source> ¹⁶, finding the source at <https://github.com/wch/r-source/blob/af7f52f70101960861e5d995d3a4bec010bc89e6/src/library/utils/R/head.R> ²⁰

This includes a comment which suggest we should ask Patrick Burns:

```
### placed in the public domain 2002
### Patrick Burns patrick@burns-stat.com
###
### Adapted for negative arguments by Vincent Goulet
### <vincent.goulet@act.ulaval.ca>, 2006
```

But it's worth checking just to make sure it's always used 6. I click on history, and then find the first version: <https://github.com/wch/r-source/commit/37271cdbc7e5d82c79bdb536ef305d93b644ad#diff-941bf47bf09f67538338535bd512d521> ²⁸ - so it has been six from the very beginning.

So next step, I'll email Patrick and see if he recollects...

meta-help



Why does head() show 6 rows by default?



hadley RStudio Employee

Nov '17

To answer a question like this, I first starting looking through the S books I have on hand (e.g. [The New S Language](#) ¹¹). I don't see `head()` mentioned in the index, so that suggests it's a function introduced by R.

Since it's an R function, I can next search [@winston](#)'s GitHub mirror of the R sources: <https://github.com/wch/r-source> ¹⁶, finding the source at <https://github.com/wch/r-source/blob/af7f52f70101960861e5d995d3a4bec010bc89e6/src/library/utils/R/head.R> ²⁰

This includes a comment which suggest we should ask Patrick Burns:

```
### placed in the public domain 2002
### Patrick Burns patrick@burns-stat.com
###
### Adapted for negative arguments by Vincent Goulet
### <vincent.goulet@act.ulaval.ca>, 2006
```

But it's worth checking just to make sure it's always used 6. I click on history, and then find the first version: <https://github.com/wch/r-source/commit/37271cdbc7e5d82c79bdb536ef305d93b644ad#diff-941bf47bf09f67538338535bd512d521> ²⁸ - so it has been six from the very beginning.

So next step, I'll email Patrick and see if he recollects...

Look in S books

Go to R-source mirror

Find relevant bit

See the history

Send an email

why *does* $n = 6$?



hadley RStudio Employee

Nov '17



From Pat (via email):

I came upon 'head' and 'tail' at one of my clients. That implementation had $n = 5$. I didn't think there would ever be an issue regarding ownership of the code, but I changed to 6 just to help if there were a conflict.



mara

Nov '17

$n = 6$: R :: brown M&Ms : Van Halen

Ways to contribute



PULL REQUESTS

Contributing code/making fixes.



ISSUES

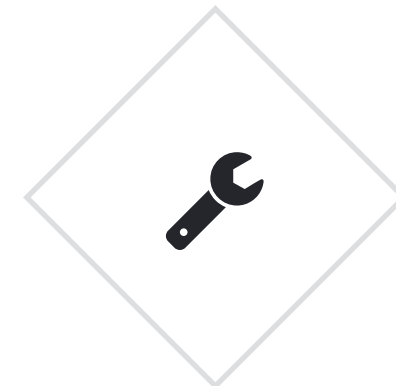
Identifying a problem, trying your best to isolate its source.



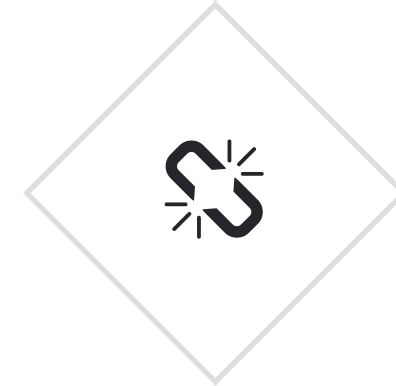
COMMENTS

Help maintainers answer questions, triage issues. Help newcomers learn how to ask better questions (e.g. the art of the reprex).

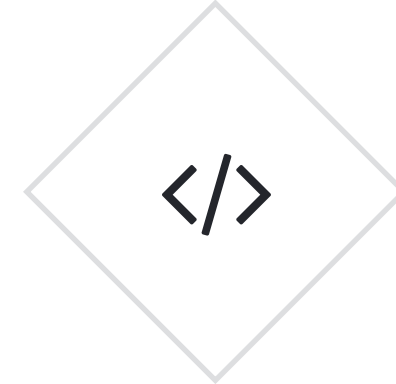
the anatomy of an issue



PROBLEM DESCRIPTION




EXPECTED BEHAVIOUR



MINIMAL REPRODUCIBLE EXAMPLE

the anatomy of an issue

Please review the [guidelines for contributing](#) to this repository.




AA ▾ B *i* “ <> 🔗 ☰ ☷ ✓ @ ★

Expected behavior

Actual behavior

Steps to reproduce the behavior

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

 Styling with Markdown is supported

My first “contribution”

15

hadley / r4ds

Watch 150 Star 1,072 Fork 1,323

Code Issues 29 Pull requests 35 Projects 0 Wiki Insights

Fixed spelling of purrr on line 121 (#208)

master (#208)

batpigandme committed with hadley on Jul 29, 2016 1 parent 8da00ed commit 6351bf51de44ef402dacc5efd159c3036a57e2e1

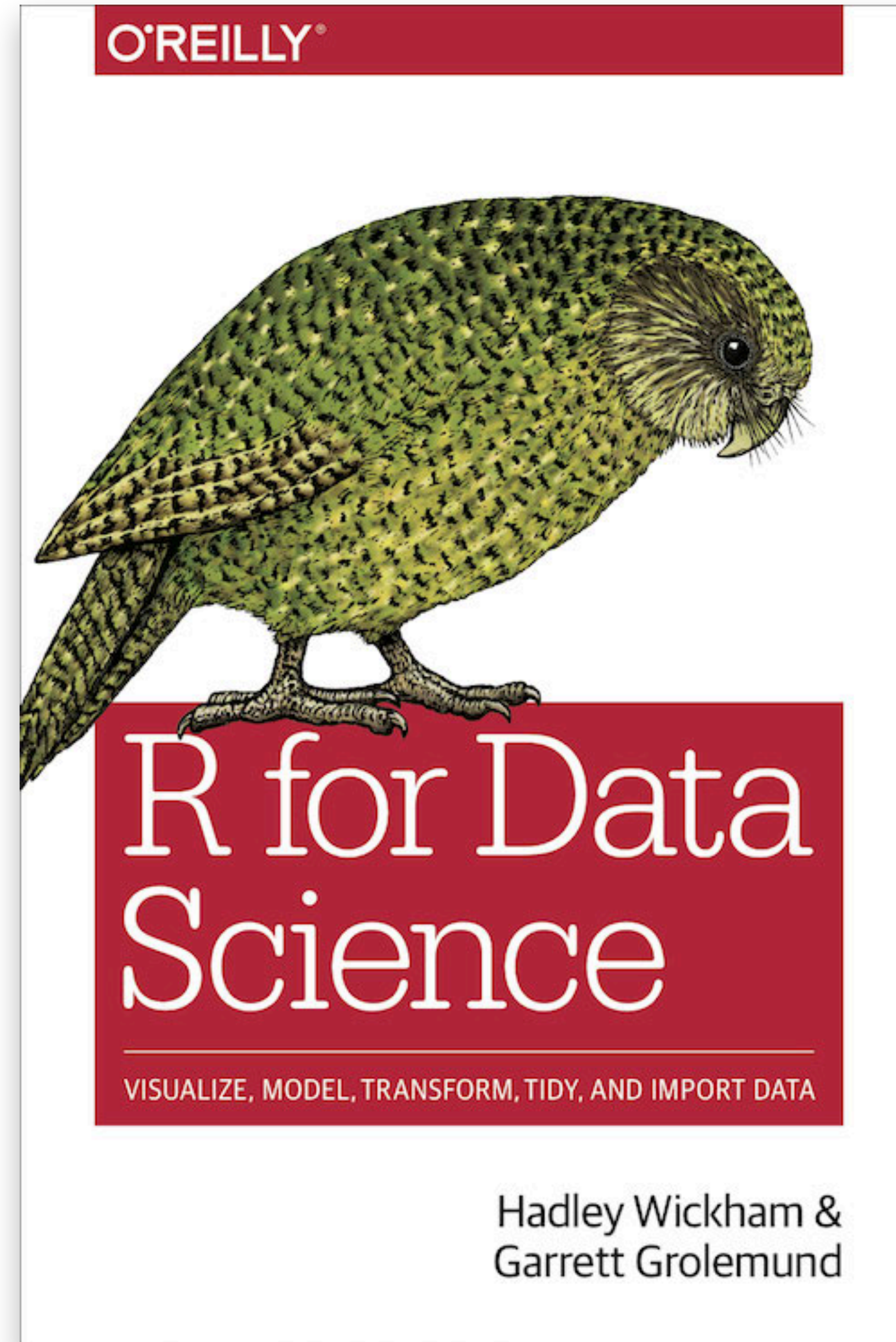
Showing 1 changed file with 1 addition and 1 deletion.

Unified Split

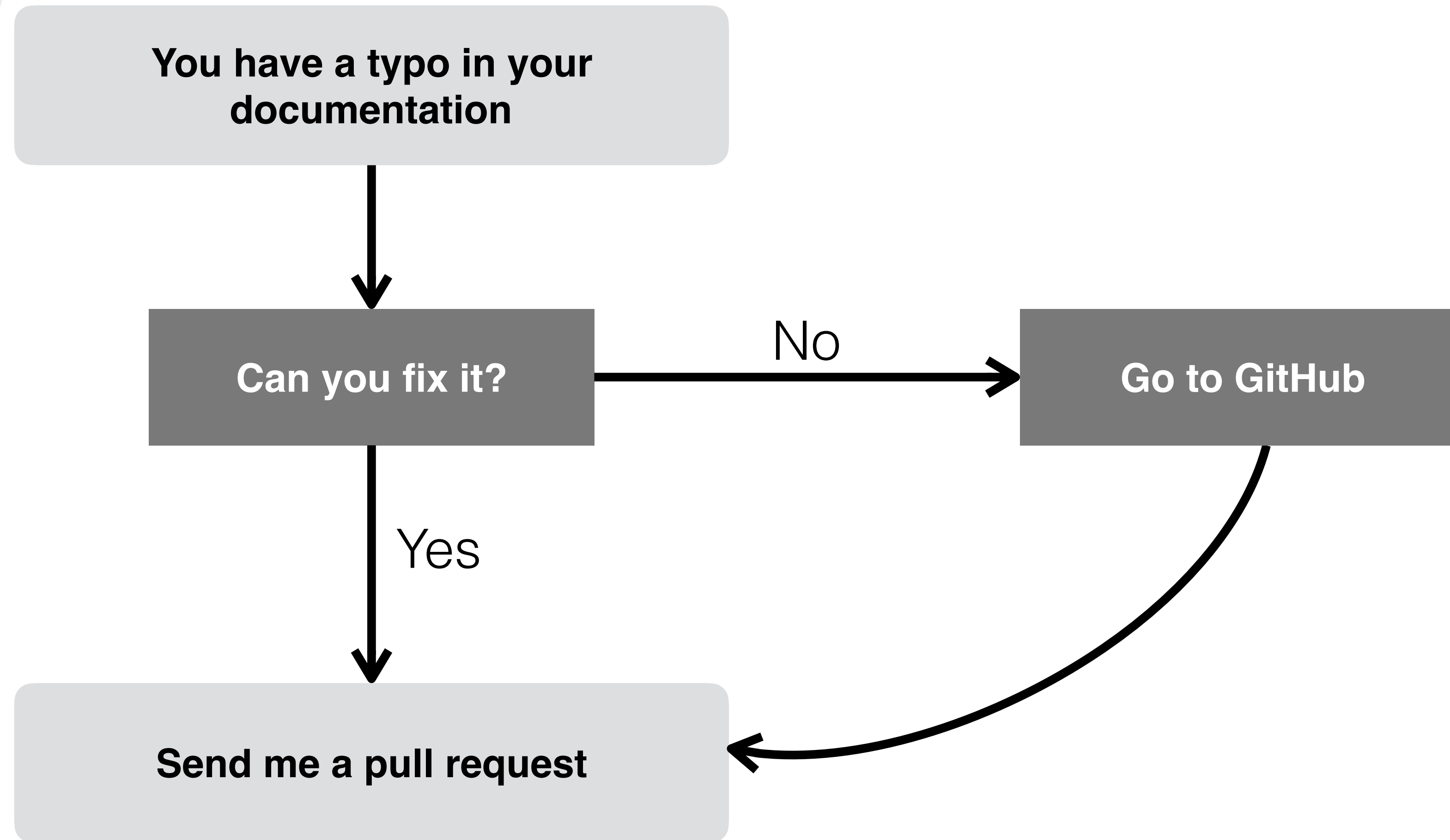
2 model-basics.Rmd

```
@@ -118,7 +118,7 @@ measure_distance <- function(mod, data) {
118 118   measure_distance(mod, 1.5), sim1)
119 119   ...
120 120   ...
121 121   -Now we use purrr to calculate the distance for all the models defined above. We need a helper function because our
    distance function expects a model as a numeric vector of length 2.
121 121   +Now we use purrr to calculate the distance for all the models defined above. We need a helper function because our
    distance function expects a model as a numeric vector of length 2.
122 122   ...{r}
123 123   ...{r}
124 124   sim1_dist <- function(a1, a2) {
```

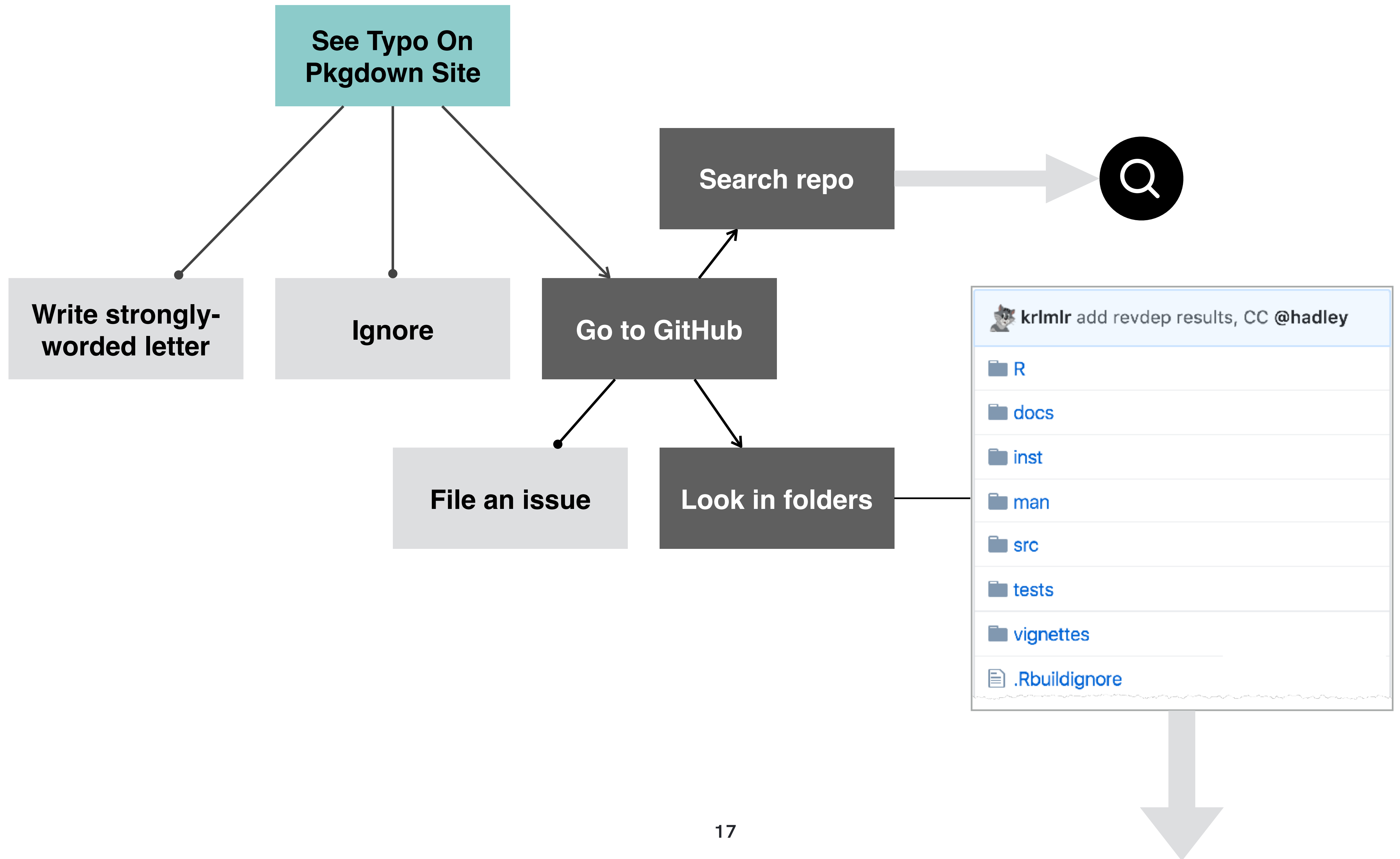
My first “contribution”



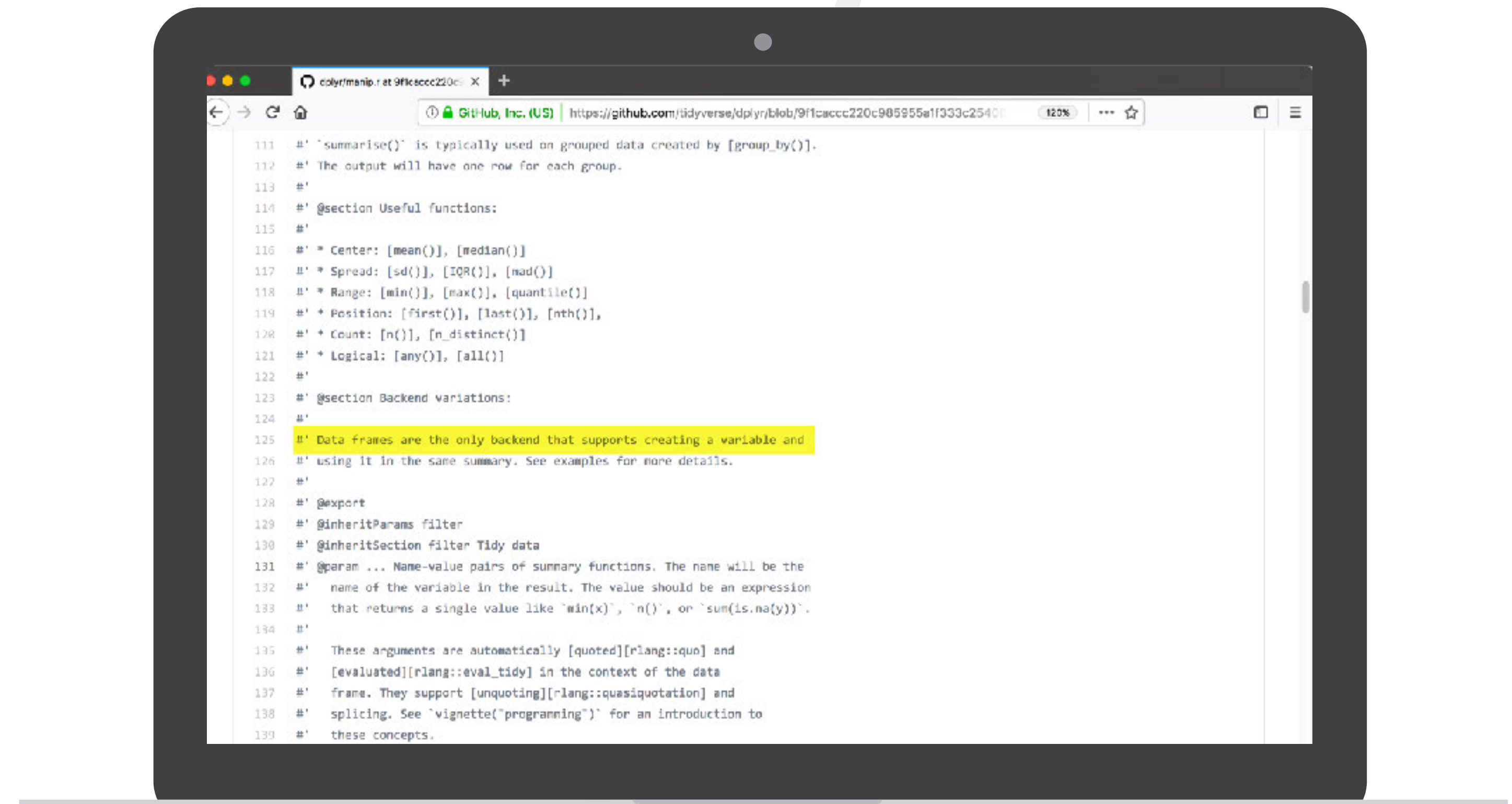
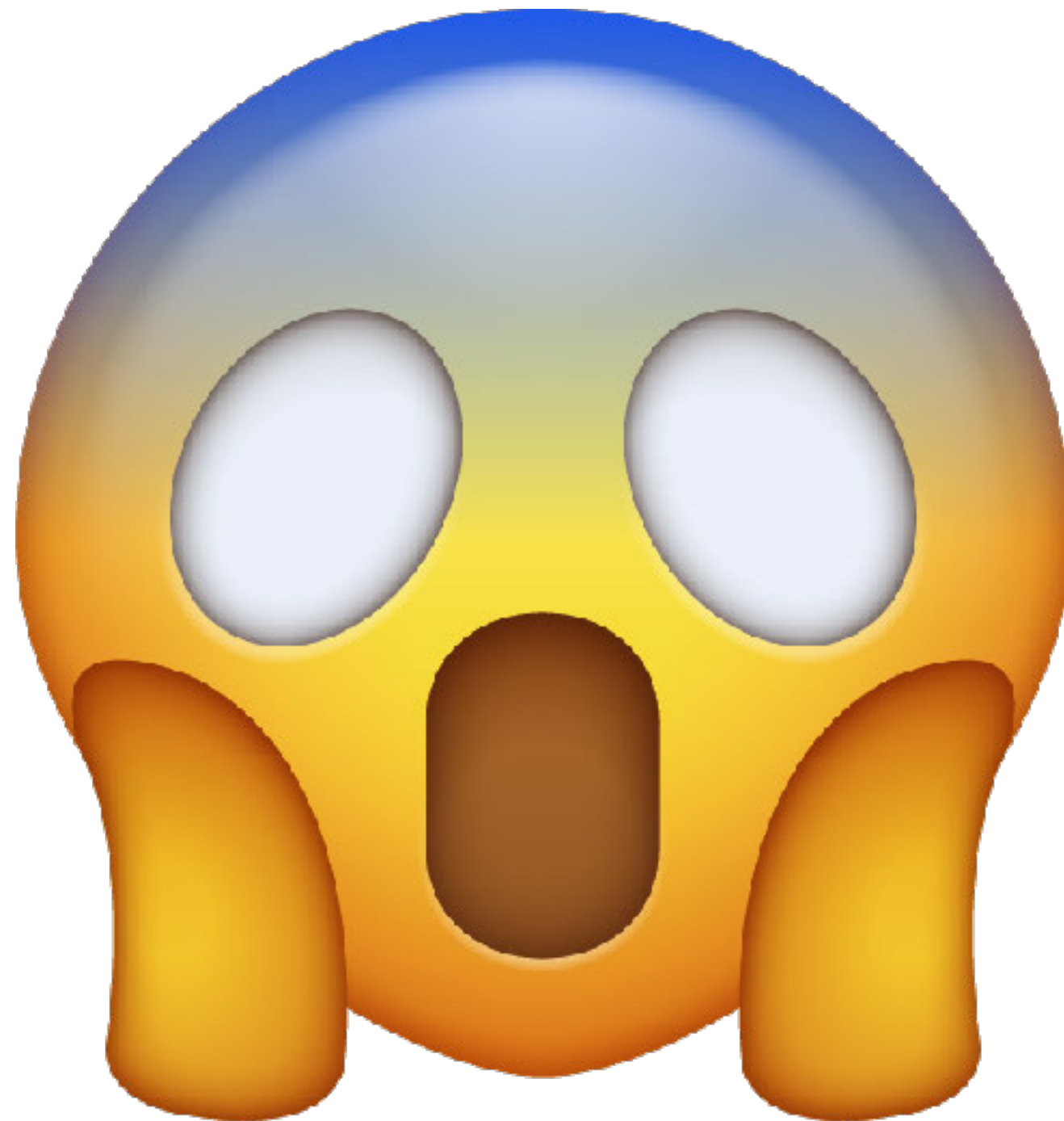
typos



Adapted from: You Do Not Need to Tell Me I Have A Typo in My Documentation by Yihui Xie



Source code?!



Source code...

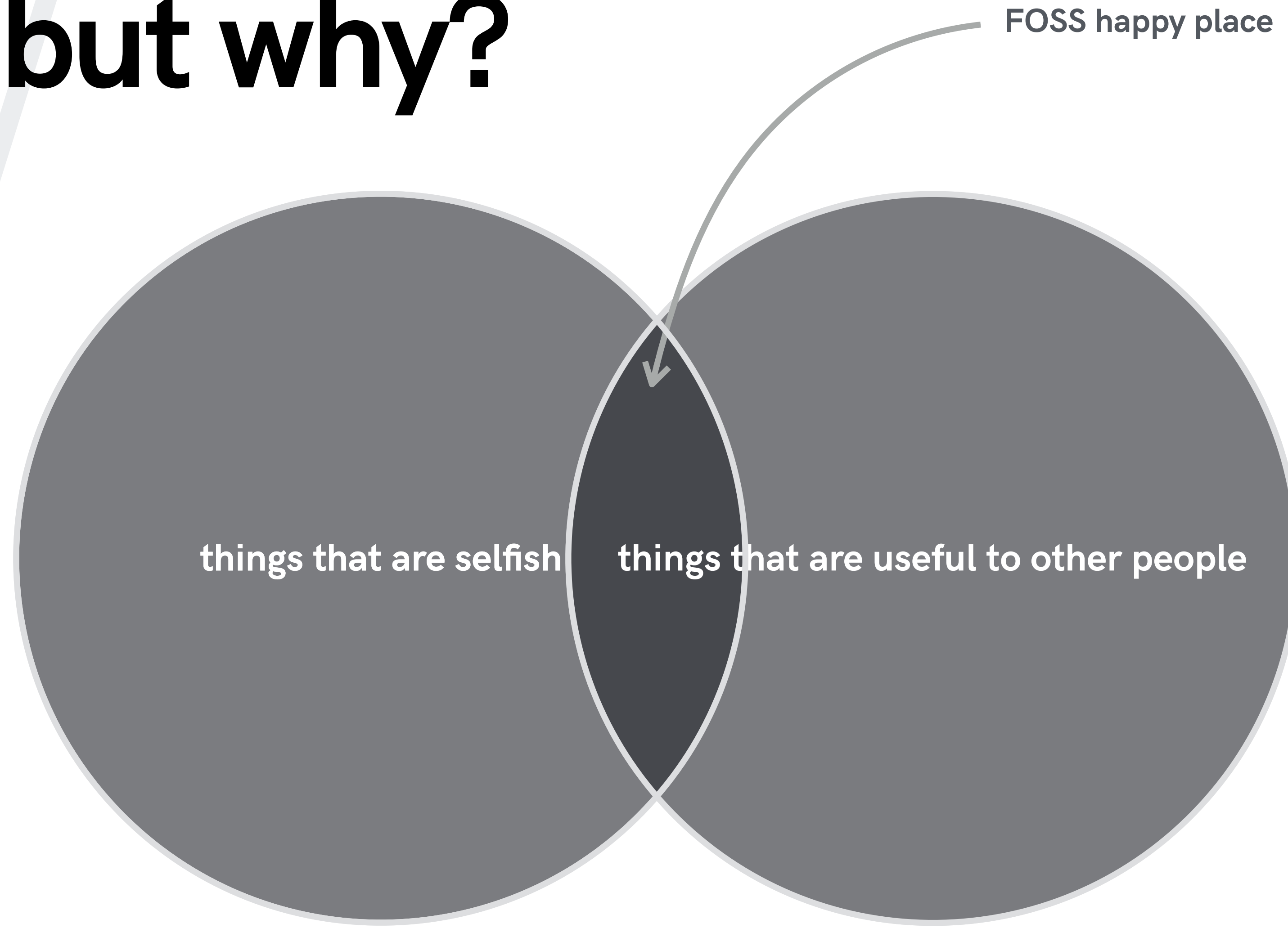


Contribute documentation

“Innocence lost is not easily regained. The designer simply cannot predict the problems people will have, the misinterpretations that will arise, and the errors that will get made.”

— Donald Norman, *The Design of Everyday Things*

wait, but why?



Hints for happy code contributions in the tidyverse



GET THE PULSE OF A PROJECT



WATCH THE REPO



READ THE CODE



DISCUSS YOUR IDEAS

CONTRIBUTING.md

↑
tidyverse & package conventions

KEY PACKAGES FOR CONTRIBUTING



ROXYGEN2

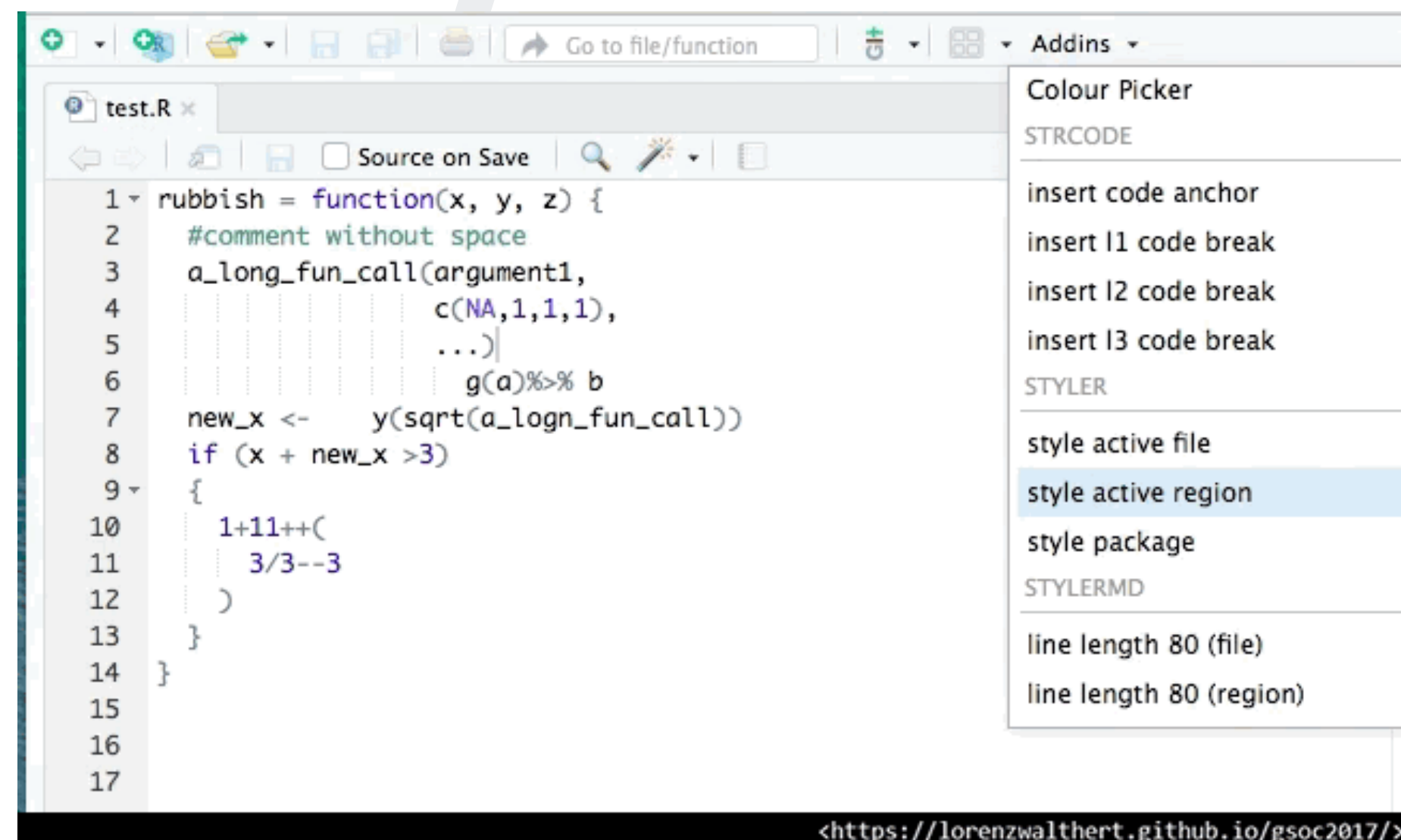


DEVTOOLS



TESTTHAT

stylish...



styler by Kirill Müller & Lorenz Walthert <http://styler.r-lib.org>

The tidyverse style guide

The tidyverse style guide

Hadley Wickham

Welcome

Good coding style is like correct punctuation: you can manage without it, but it sure makes things easier to read. This site describes the style used throughout the [tidyverse](#). It was originally derived from [Google's R style guide](#), but has evolved and expanded considerably over the years.

All style guides are fundamentally opinionated. Some decisions genuinely do make code easier to use (especially matching indenting to programming structure), but many decisions are arbitrary. The most important thing about a style guide is that it provides consistency, making code easier to write because you need to make fewer decisions.

Two R packages support this style guide:

- [styler](#) allows you to interactively restyle selected text, files, or entire projects. It includes an RStudio add-in, the easiest way to re-style existing code.

Addins

STYLER

style

Style active file

Style package

Style selection

- [lintr](#) performs automated checks to confirm that your conform to the style guide.

the magic of reprex



**reproducible
example**

Keys to reprex-cellence

- ✓ Code that **actually runs**
- ✓ Code that **doesn't have to be run**
- ✓ Code that **can be easily run**

Nailing those reprexes?

- Help others ask questions.
- Answer questions.
- Make *more* thoughtful users

**contributing is
a relationship,
not a
transaction**



baby steps build up trust

...and make life easier



every contribution counts...

The image shows a screenshot of the Tidyverse website's 'Acknowledgments' page. The page has a dark blue header with the Tidyverse logo and navigation links: Packages, Articles, Learn, Help, and Contribute. The main content area is titled 'Acknowledgments' and contains a paragraph of text listing 109 contributors. To the right of the text, there is a large black curly bracket. Three lines extend from the middle of this bracket to the right, pointing to the labels 'PULL REQUESTS', 'ISSUES', and 'COMMENTS'. The footer of the page is dark blue and contains the text 'The tidyverse is proudly supported by R Studio' and social media icons for GitHub and Twitter.

Tidyverse Packages Articles Learn Help Contribute

Acknowledgments

We received issues, pull requests, and comments from 109 people since tibble 1.2.0. Thanks to everyone: @aalexandersson, @adnbps, @AkhilNairAmey, @alexhallam, @alibat, @amjiuzi, @AndreMikulec, @andrewjpfeiffer, @ashiklom, @atribe, @bapfeld, @barnettjacob, @behman, @BillDunlap, @BruceZhaR, @cassiusoat, @cboettig, @cderv, @ckluss, @ClaytonJY, @colearendt, @csgillespie, @dalejbarr, @dan87134, @DavisVaughan, @ddiez, @dhicks, @dlpd, @drewgendreau, @drolejoel, @echasnovski, @edzer, @ElsLommelen, @etiennebr, @FabianRoger, @garrettgman, @gavinsimpson, @geotheory, @ginolhac, @hadley, @happysshows, @heavywatal, @helix123, @holstius, @huffis, @ianmcook, @manuelcostigan, @janschulz, @javierluraschi, @jennybc, @jimhester, @joelgombin, @jonathan.g, @kendonB, @kevinushey, @khughitt, @kismsu, @krlmlr, @kwstat, @LaDilettante, @lcolladotor, @lionel-, @lpmarco, @m-sostero, @MarcusWalz, @matteodefelic, @mattdfider, @mgirlich, @michaellevy, @MikeBadescu, @mkearney, @mmuurr, @Monduiz, @mubeenarasack, @mundl, @nbenn, @ncarchedi, @NikNakk, @noamross, @ntguardian, @pObs, @patperry, @pgensler, @phalexo, @pssguy, @r2evans, @rentrop, @richierocks, @Rongpeng, @s-fleck, @sainathadapa, @sebschub, @sibojan, @slonik-az, @sskim47, @t-kalinowski, @thercast, @thornend, @tjmahr, @trinker, @vnijs, @vspinu, @vvrably, @wibom, @wpetry, @yeedle, @yihui, @yutannihilation, and @Zedseayou.

PULL REQUESTS

ISSUES

COMMENTS

The tidyverse is proudly supported by R Studio

Thank You

<http://bit.ly/tverse-contrib>



resources

- Happy Git and GitHub for the useR by Jenny Bryan.
- Contribute to the tidyverse (incl. ways to contribute that *don't* involve writing code).
- Contributing Code to the Tidyverse by Jim Hester.
- R packages by Hadley Wickham.
 - Git and GitHub
 - Automated checking
 - Object documentation
 - Testing
- dplyr's NEWS.md is a good source of examples for both content and styling.
- Closing issues using keywords on GitHub.
- Autolinked references and URLs on GitHub.
- GitHub Guides: Forking Projects.