

```
library(dplyr)

rladies_global %>%
  filter(city == 'London')
```



10 Steps to Becoming a Tidyverse Contributor

Nic Crane



Hello!

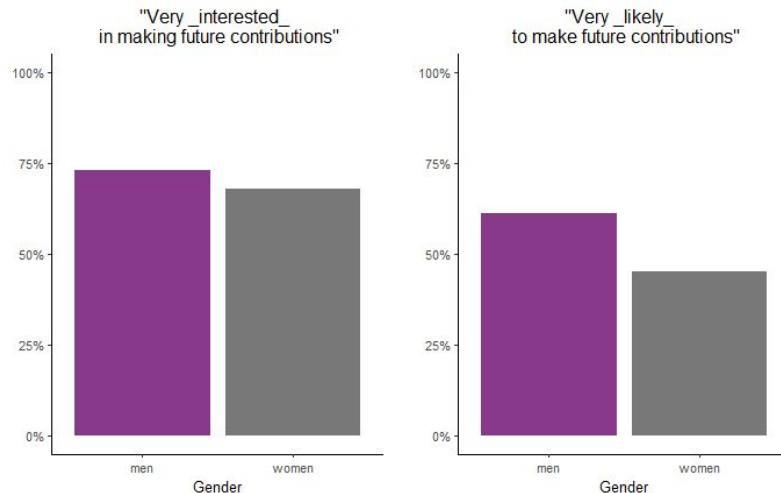
Nic Crane

@nic_crane 

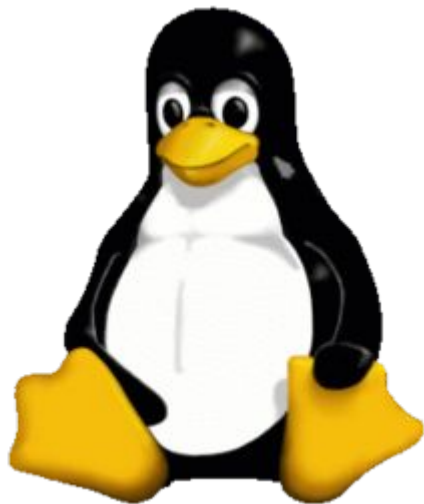
“

The gender imbalance in open source remains profound...Open source contributors don't yet reflect its broad audience of users

—GitHub Survey 2017



Culture is Changing in Open Source





How Does R Compare?

2012:
RLadies
Founded

2015:
R Consortium
established

2015:
R Forwards
set up

2016: "The R
community is
awesome
(and fast)"

2017: "The R
community is one
of R's best
features"

2017: RStudio
Community
Forums goes
live



Why Should You Contribute?

GITHUB SURVEY 2017

- “Half of contributors say that their open source work was somewhat or very important in getting their current role.”
- “Open source work helps people build their professional reputation.”
- “Improving contributor representation can help create a more representative tech sector overall.”



Why Should You Contribute?

MY PERSONAL REASONS

- Improve code knowledge
- Interact with the community
- ~~Overcome~~ Lessen imposter syndrome

“

When I started my transition into data science, I said yes to pretty much every opportunity that came my way, even if it felt slightly beyond my skill set or experience level...I said yes to many of these things when it felt like I wasn't sure if I was ready.



—Julia Silge

Interview with And Comfort (October 2018)

“

Not always easy for beginners. Helps to have a team that will onboard. Give beginner level work. Mentorship. Pick projects that are welcoming.



—Gabriela de Queiroz [summarised tweet]

Grace Hopper Celebration of Women in Computing (2018)



Step 1

Decide How to Contribute

- Identifying issues
- Documentation
- Helping people out on Twitter and forums
- Blogging
- **Code**



Step 2

Learn!

- R functions
- git (clone, commit, push)
- Building R packages (basics)
- GitHub – making a pull request



Step 3

Find an Issue

beginner-friendly

help wanted

good first issue

Step 3

Find an Issue



CRAN - Package broom

https://cran.r-project.org/web/packages/broom/index.html

broom: Convert Statistical Analysis Objects into Tidy Tibbles

Summarizes key information about statistical objects in tidy tibbles. This makes it easy to report results, create plots and consistently work with large numbers of models at once. Broom provides three verbs that each provide different types of information about a model. `tidy()` summarizes information about model components such as coefficients of a regression. `glance()` reports information about an entire model, such as goodness of fit measures like AIC and BIC. `augment()` adds information about individual observations to a dataset, such as fitted values or influence measures.

Version: 0.5.0

Depends: R (≥ 3.1)

Imports: [backports](#), [dplyr](#), [methods](#), [nlme](#), [purrr](#), [reshape2](#), [stringr](#), [tibble](#), [tidyr](#)

Suggests: [AER](#), [akima](#), [AUC](#), [bbmle](#), [betareg](#), [biglm](#), [binGroup](#), [boot](#), [brms](#), [btergm](#), [car](#), [caret](#), [coda](#), [covr](#), [e1071](#), [emmeans](#), [ergm](#), [gam](#) (≥ 1.15), [gamlss](#), [gamlss.data](#), [gamlss.dist](#), [geepack](#), [ggplot2](#), [glmnet](#), [gmm](#), [Hmisc](#), [irlba](#), [joineRML](#), [Kendall](#), [knitr](#), [ks](#), [Lahman](#), [lavaan](#), [lfe](#), [lme4](#), [lmodel2](#), [lmtree](#), [lsmeans](#), [maps](#), [maptools](#), [MASS](#), [Matrix](#), [mclust](#), [mgcv](#), [muhaz](#), [multcomp](#), [network](#), [nnet](#), [oreut](#) (≥ 2.2), [ordinal](#), [plm](#), [plyr](#), [poLCA](#), [psych](#), [quantreg](#), [rgeos](#), [rmarkdown](#), [robust](#), [rsample](#), [rstan](#), [rstanarm](#), [sp](#), [speedglm](#), [statnet.common](#), [survey](#), [survival](#), [testthat](#), [tseries](#), [xergm](#), [zoo](#)

Published: 2018-07-17

Author: David Robinson [aut, cre], Alex Hayes [aut], Matthieu Gomez [ctb], Boris Demeshev [ctb], Dieter Menne [ctb], Benjamin Nutter [ctb], Luke Johnston [ctb], Ben Bolker [ctb], Francois Briatte [ctb], Jeffrey Arnold [ctb], Jonah Gabry [ctb], Luciano Selzer [ctb], Gavin Simpson [ctb], Jens Preussner [ctb], Jay Hesselberth [ctb], Hadley Wickham [ctb], Matthew Lincoln [ctb], Alessandro Gasparini [ctb], Lukasz Komsta [ctb], Frederick Novometsky [ctb], Wilson Freitas [ctb], Michelle Evans [ctb], Jason Cory Brunson [ctb], Simon Jackson [ctb], Ben Whalley [ctb], Michael Kuehn [ctb], Jorge Cimentada [ctb], Erle Holgersen [ctb], Karl Dunkle Werner [ctb]

Maintainer: David Robinson <admiral.david@gmail.com>

BugReports: <http://github.com/tidyverse/broom/issues>

License: [MIT](#) + file [LICENSE](#)

URL: <http://github.com/tidyverse/broom>

NeedsCompilation: no

Materials: [README](#) [NEWS](#)



Step 4

Ask if You Can Help



thisisnic commented on 26 Jul • edited ▼

Contributor



I'd love to have a go at fixing one or two of these as my first contribution!



1



ellessenne commented on 27 Jul

Contributor



Hi @alexpghayes, I can deal with the methods for the `mjoint` class 😊



1



alexpghayes commented on 27 Jul

Collaborator



@thisisnic I'm in the process of adding a couple more tests, and will follow up with you once I've got them written!

@ellessenne Awesome, much appreciated. There isn't much doc for the new test system but if you run `devtools::install_github("tidymodels/broom")` I believe you should get everything you need. Let me know if you run into any problems!

Thanks to both of you!



Step 5

Fork the Repo

tidymodels / broom

Watch 57 Star 769 Fork 201

Code Issues 106 Pull requests 9 Projects 0 Wiki Insights

Convert statistical analysis objects from R into tidy format <https://broom.tidyverse.org>

r tidy-data modeling

871 commits 2 branches 10 releases 73 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download

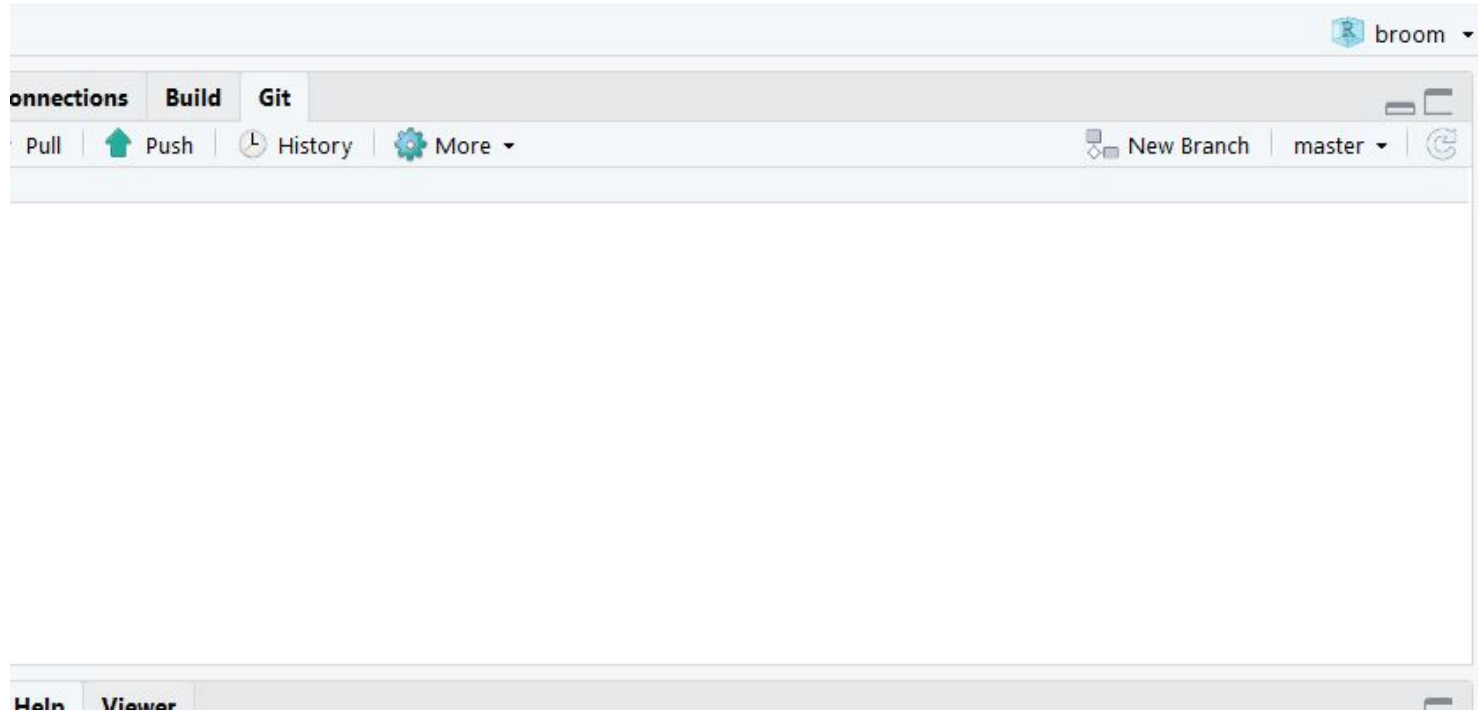
alexphayes Update mediate tidiers (#519) Latest commit bea70cd 17 days ago

.github	New README and gam/mgcv doc linking	4 months ago
R	save object as tibble and increase coverage	18 days ago
data-raw	Move standardized tests and glossaries to modeltests package (#449)	4 months ago
docs	Auament broaress and aeneral cleanup	3 months ago



Step 6

Open in RStudio & Make a Branch





Step 7

Familiarise Yourself with the Code

```
test-stats-kmeans.R | MD NEWS.md | stats-kmeans-tidiers.R
Source on Save
Run Source

1 context("stats-kmeans")
2
3 skip_if_not_installed("modeltests")
4 library(modeltests)
5
6 set.seed(2)
7 x <- rbind(
8   matrix(rnorm(100, sd = 0.3), ncol = 2),
9   matrix(rnorm(100, mean = 1, sd = 0.3), ncol = 2)
10 )
11
12 fit <- kmeans(x, 2)
13
14 test_that("kmeans tidier arguments", {
15   check_arguments(tidy.kmeans)
16   check_arguments(glance.kmeans)
17   check_arguments(augment.kmeans, strict = FALSE)
18 })
19
20 # tidy.kmeans uses the original column names to name columns in output.
21 # Therefore, strict must be set to FALSE for this test to pass.
22 test_that("tidy.kmeans", {
23   td <- tidy(fit)
24   check_tidy_output(td, strict = FALSE)
25 })
26
27 test_that("tidy.kmeans", {
28   gl <- glance(fit)
29   check_glance_outputs(gl)
30 })
31
```

Step 8

Write Code



8 R/stats-kmeans-tidiers.R

View



@@ -2,7 +2,7 @@

```
2 #' @template title_desc_tidy
3 #'
4 #' @param x A `kmeans` object created by [stats::kmeans()].
5 - #' @param col.names Dimension names. Defaults to `x1, x2, ...`
6
7 #' @template param_unused_dots
8 #'
9 #' @evalRd return_tidy("size", "withinss", "cluster")
```

@@ -13,7 +13,11 @@

```
13 #' @export
14 #' @seealso [tidy()], [stats::kmeans()]
15 #' @family kmeans tidiers
16 - tidy.kmeans <- function(x, col.names = paste0("x", 1:ncol(x$centers)), ...) {
17
18   ret <- as.data.frame(x$centers)
19   colnames(ret) <- col.names
20   ret$size <- x$size
21 }
```

```
2 #' @template title_desc_tidy
3 #'
4 #' @param x A `kmeans` object created by [stats::kmeans()].
5 + #' @param col.names Dimension names. Defaults to the names of the variables in x. Set to NULL to get
6   names `x1, x2, ...`.
7
8 #' @template param_unused_dots
9 #'
10 #' @evalRd return_tidy("size", "withinss", "cluster")
11
12
13 #' @export
14 #' @seealso [tidy()], [stats::kmeans()]
15 #' @family kmeans tidiers
16 + tidy.kmeans <- function(x, col.names = colnames(x$centers), ...) {
17 +
18 +   if(is.null(col.names)){
19 +     col.names <- paste0("x", 1:ncol(x$centers))
20 +   }
21
22   ret <- as.data.frame(x$centers)
23   colnames(ret) <- col.names
24   ret$size <- x$size
25 }
```



Step 9

Submit PR and Wait

#450 - update tidy.kmeans and document why strict = FALSE is OK for its tests #486

Merged alexpghayes merged 5 commits into tidymodels:master from thisisnic:kmeans-tidier on 1 Oct

Conversation 4 Commits 5 Checks 0 Files changed 4

thisisnic commented on 13 Aug Contributor + 👤 ...

tidy.kmeans uses column names from the input to name columns in the output, and so will not pass tests if strict is set to TRUE. I have added a comment above the test explaining this.

In addition, the tidier was naming the columns x1...xN (where N is the number of columns in the input). I have updated it to use the original column names if present, for clarity in the output. I wasn't sure whether this change would be acceptable given it is beyond the scope of the original issue.

Is this change OK, and if so, do I need to document this elsewhere? If not, shall I submit a new PR containing just the comments on setting strict to FALSE?

#450 - update tidy.kmeans and document why strict = FALSE is OK for i... b92b173

alexpghayes commented on 7 Sep Collaborator + 👤 ...

I think this is a reasonable change. I would add some documentation that `col.names` should be set to `NULL` to the get variable names `x1`, `x2`, ...

I would also add this change to `NEWS` so that it doesn't surprise anyone.



Step 10

Celebrate! Then Encourage Others to Get Involved and Share Your Experience

Nic Crane @nic_crane · Sep 8
Just had my first #tidyverse pull request accepted! 🥳 Thanks @dataandme for your awesome talk at rstudio::conf that prompted me to get involved, and @alexpghayes for taking the time to guide me through the process! ❤️ #rstats

date tidier for survival::survdiff (#485) Add label

alex hayes @alexpghayes pushed 1 commit. - d818d94 Merge branch 'master' into 7 Sep

alex hayes 7 Sep
Awesome, thanks for documenting this! — You are receiving this because you

alex hayes to tidymodels/broom, me, Auth... Yesterday View details

6 11 81

The diagram illustrates the tidyverse ecosystem. At the top, it shows 'contributing to the tidyverse' and 'tidyverse'. Below, it asks 'What is the tidyverse?' and lists 'Key FOSS actors'. It also includes a section for 'the thoughtful user' and 'ways of contributing', with a small illustration of a person and a rubber duck.

```
library(dplyr)
```

```
rladies_global %>%  
  filter(city == 'London')
```



Thanks! 🙄

Nic Crane
@nic_crane 



Resources

Talk by Mara Averick on contributing:

<https://www.rstudio.com/resources/videos/contributing-to-tidyverse-packages/>

rOpenSci contributing guide:

https://ropensci.github.io/dev_guide/contributingguide.html

Repo walking through making a pull request:

<https://github.com/firstcontributions/first-contributions>

Guide to contributing code to the tidyverse:

<https://www.tidyverse.org/articles/2017/08/contributing/>