

A Gentle Introduction to Tidy Statistics in R

Thomas Mock, PhD

Who am I?





Who are you?



Jesse Mostipak

@kierisi

Following

OK #rstats Twitter, let's talk.

If you've tried learning R and stopped, what would have made learning R a better/easier/more pleasant experience for you?

If you're learning (or have learned) R, what were your stumbling blocks? What do you wish you had known when you started?

7:14 AM - 5 Feb 2018

58 Retweets 151 Likes



112



58



151



Dr. Brianna McHorse 🏳️‍🌈

@fossilosophy

When I first started learning, I didn't have many examples of code/people doing the kinds of things I needed to do (basic



David Butler

@DavidKButlerUoA

Follow

Replying to @DavidKButlerUoA @kierisi and 3 others

Instead, the R-interested I meet are doing



Thomas Mock 🧑

@thomas_mock

Replying to @kierisi

Time investment in solving many "simple" problems. 1st time I want to run an anova: read in excel file, oh need it in .csv, wait why can't it find my file? Oh setwd() fit <- aov() ERROR read as character, ok as.numeric()

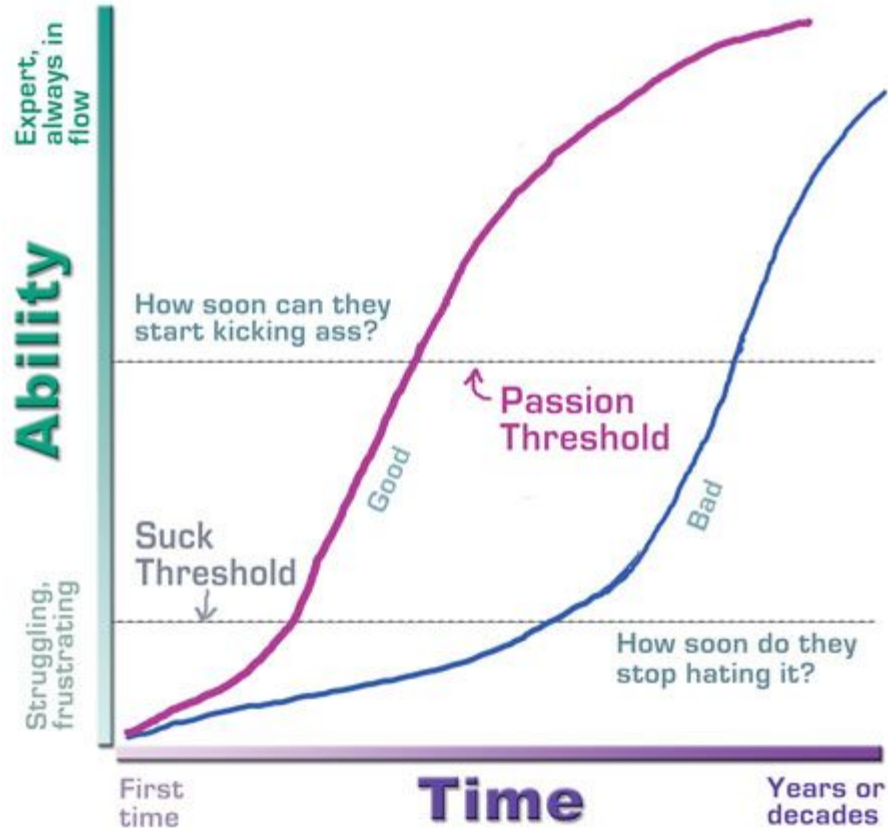
Time to make a plot, base plot 😞

8:05 AM - 5 Feb 2018

4 Likes



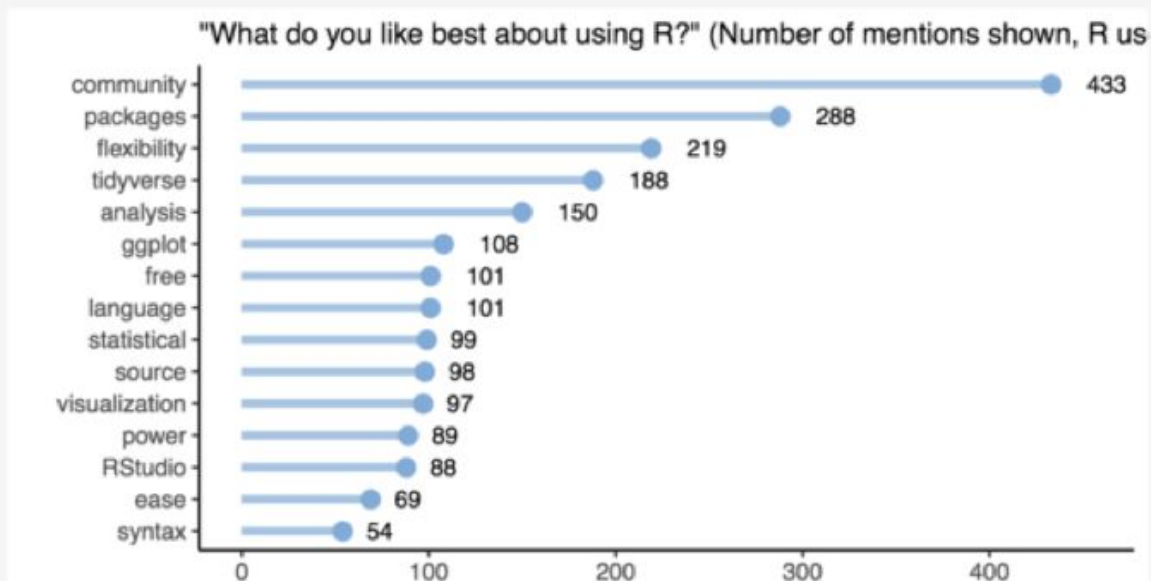
The Kick Ass Curve



Why learn R?

- Connect with an amazing community
- Programming is a superpower that everyone has access to
- Clean, analyze, plot, and communicate with your data all in one place
- Reproducibility
- Automation
- IT'S FREE!

WHAT PEOPLE LIKE BEST ABOUT R IS THE COMMUNITY





Thomas Mock @thomas_mock · Mar 21

Finalizing the slides on my #rstats presentation "Finding the YOU in the R community"

So I have a questions for the R community (at least on Twitter):

? "Why be an involved useR?"

15 7 43



Kim Cressman

@swmpkim

Following

Replying to @thomas_mock

Because it's fun and you can learn SO MUCH, without having to seek it out, just by watching what other people do. I have a lot of "OMG, you can do THAT?" moments.

I also learn a lot from what other people *ask*.

And sometimes I can help someone else, which is always nice.

8:42 PM - 21 Mar 2019

3 Likes



1 3 3



Thomas Mock @thomas_mock · Mar 21

Finalizing the slides on my #rstats presentation "Finding the YOU in the R community"

So I have a questions for the R community (at least on Twitter):

? "Why be an involved useR?"

15 7 43



Luuuda

@ludmila_janda

Following

Replying to @thomas_mock

for all these feels:

🥰 omg you can do that in R now?!

👋 oh I think I can help with this issue

👏 oh man I've so been there

😊 thank you so much for helping me with this thing i've been going crazy trying to figure out

📊 whoa awesome data viz

📚 ohhh sweet resources

6:37 PM - 21 Mar 2019

3 Retweets 22 Likes

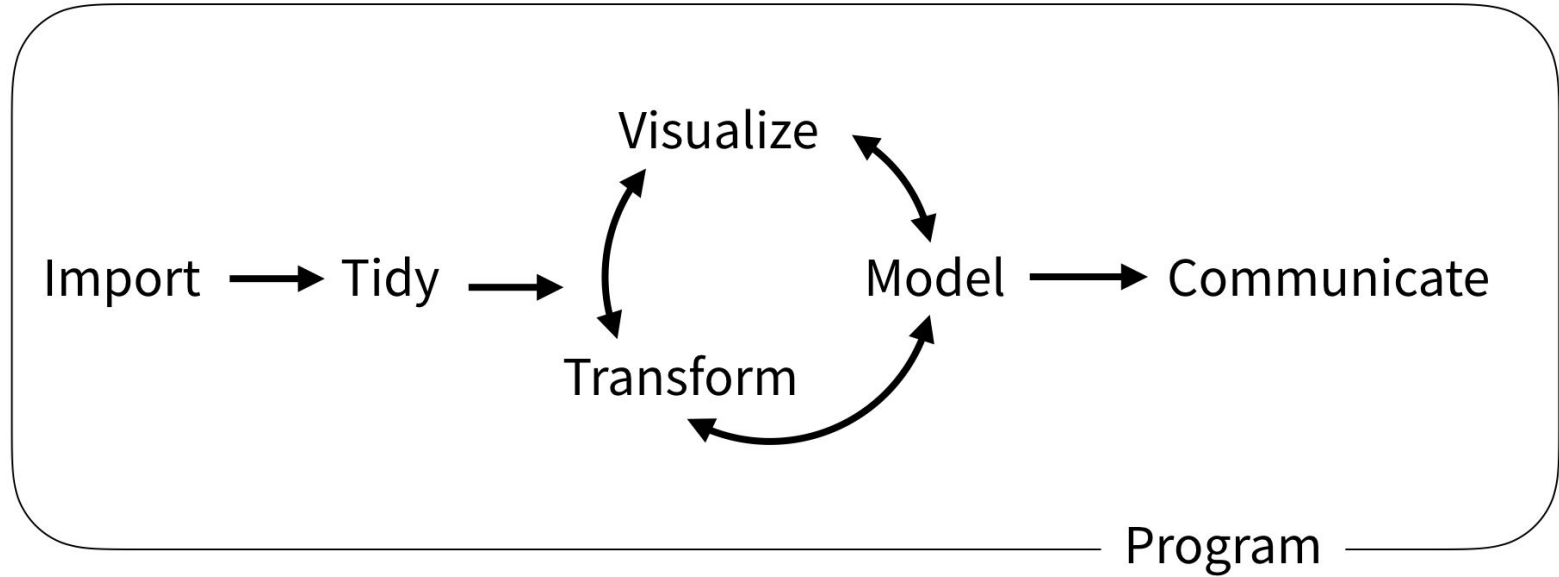


1 3 22

What we are covering today

- Basics of RStudio
- Read data into R
- Data wrangling
- Exploratory data analysis
- Basic statistics in R
- Publication plots







What we aren't doing today

- Covering what statistical test to run when
- Deep dive into statistical programming
- Expecting you to 100% get it the first time!



Ultimate Goal

Overview



R: Engine



RStudio: Interface



Studio Cloud #TidyTuesday / RStudio Thomas Mock

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins R 3.6.0

Environment History Connections

Global Environment

Environment is empty

```
~/cloud/project/
> ggplot(mtcars, aes(x = disp, y = mpg, color = as.factor(cyl))) + geom_point(size = 3)
> |
```

Files Plots Packages Help Viewer

Zoom Export Publish

mpg

disp

as.factor(cyl)

- 4
- 6
- 8

Console

Output

R: New phone



R Packages: Apps you can download



File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

code_through.rmd x

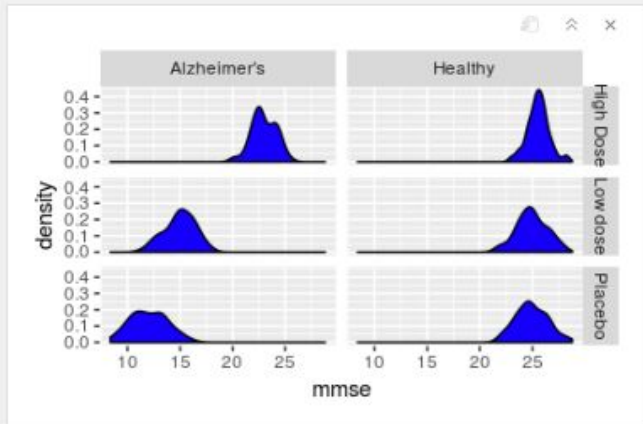
Knit Insert Run

```
31 raw_df <- readxl::read_xlsx("ad_treatment.xlsx")
32 dplyr::glimpse(raw_df)
33
34
35 ***
36
37
38 {r}
39 ggplot(data = raw_df, aes(x = mmse)) +
40   geom_density(fill = "blue") +
41   facet_grid(~health_status)
42
```

Environment History

Import Data Global Environment

Environment is empty.



Live demo

Files Plots Packages History

New Folder Upload Delete

Cloud > project

Name

..



Next steps!

Slides + RStudio Cloud instance

The screenshot displays the RStudio Cloud interface for a workspace named "intro_R". The left sidebar shows navigation options like "Spaces", "Learn", and "Info". The main editor window shows R code in a file named "code_through.rmd". The code reads an Excel file, inspects it, and creates a faceted density plot of "mmse" values for "Alzheimer's" and "Healthy" individuals across three dose levels: "High Dose", "Low dose", and "Placebo".

```
31 raw_df <- readxl::read_xlsx("ad_treatment.xlsx")
32 dplyr::glimpse(raw_df)
33 ```
34
35 ***
36
37
38 ```{r}
39 ggplot(data = raw_df, aes(x = mmse)) +
40   geom_density(fill = "blue") +
41   facet_grid(~health_status)
42 ```
```

The resulting plot is a 3x2 grid of density plots. The columns are labeled "Alzheimer's" and "Healthy". The rows are labeled "High Dose", "Low dose", and "Placebo". The x-axis is "mmse" (10-25) and the y-axis is "density" (0.0-0.4). The "Healthy" plots show a peak around 25, while the "Alzheimer's" plots show a peak around 15. The "Placebo" row shows a shift in the distribution for both groups.

On the right, the "Environment" panel shows "Global Environment" and "Environment is empty". Below it, the "Files" panel shows a "project" folder. The top right corner shows the user "Thomas Mock" and an "Access" button.

RStudio Cloud Primers

<https://rstudio.cloud/learn/primers>

- Interactive
- Ready to go
- Can't "break" anything

The sidebar navigation menu for RStudio Cloud includes the following items:

- Studio Cloud
- Spaces
 - Your Workspace
 - #TidyTuesday
 - Puzzle testing
 - + New Space
- Learn
 - Guide
 - What's New
 - Primers** (highlighted)
 - Cheat Sheets
 - Feedback and Questions
- Info
 - Terms and Conditions
 - System Status

The RStudio Primers landing page features a navigation bar with "Learn", "Guide", "What's New", "Primers", and "Cheat Sheets". The main content area is titled "R Studio Primers" and includes the text: "Learn data science basics with the interactive tutorials below." The page is organized into a grid of six primer cards:

- The Basics**: Start here to learn the skills that you will rely on in every analysis (and every primer that follows): how to inspect, visualize, subset, and transform your data, as well as how to run code.
- Work with Data**: Learn the most important data handling skills in R: how to extract values from a table, subset tables, calculate summary statistics, and derive new variables.
- Visualize Data**: Learn how to use ggplot2 to make any type of plot with your data. Then learn the best ways to visualize patterns within values and relationships between variables.
- Tidy Your Data**: Unlock the tidyverse by learning how to make and use tidy data, the data format designed for R.
- Iterate**: Master a core programming paradigm with the purrr package: for each ____ do _____.
- Write Functions**: Functions are the key to programming in R. This primer will teach you how to write and use your own reusable functions.

Book guides

- Learning statistics with R
 - <https://learningstatisticswithr.com/book/>
- R for Data Science
 - <https://r4ds.had.co.nz/>
- Data Visualization
 - <https://socviz.co/>



#TidyTuesday - R4DS Online Learning Community

A weekly data project aimed at the R ecosystem. An emphasis will be placed on understanding how to summarize and arrange data to make meaningful charts with `ggplot2`, `tidyr`, `dp1yr`, and other tools in the tidyverse ecosystem.

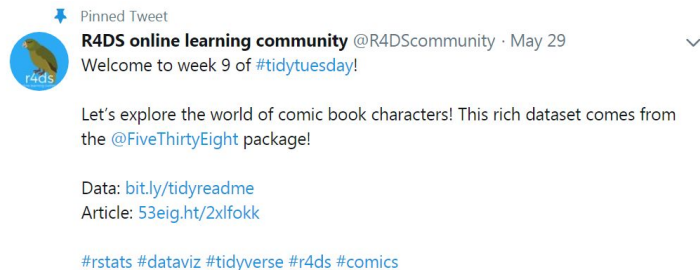
Data

- Uploaded weekly at [github/rfordatascience/tidytuesday](https://github.com/rfordatascience/tidytuesday)
- Weekly welcome from twitter.com/R4DScommunity

Basics to participate:

- Use provided data to create an image of a plot from R
- Save and share code used to create your plot and data tidying
- Submit the plot and code on **Twitter**
- Use the **#TidyTuesday** hashtag (and you can tag [@R4DScommunity](https://twitter.com/R4DScommunity) or [@thomas_mock](https://twitter.com/thomas_mock))
- Browse other submissions and like/comment on their work!

Full details, data, code of conduct: <https://github.com/rfordatascience/tidytuesday>

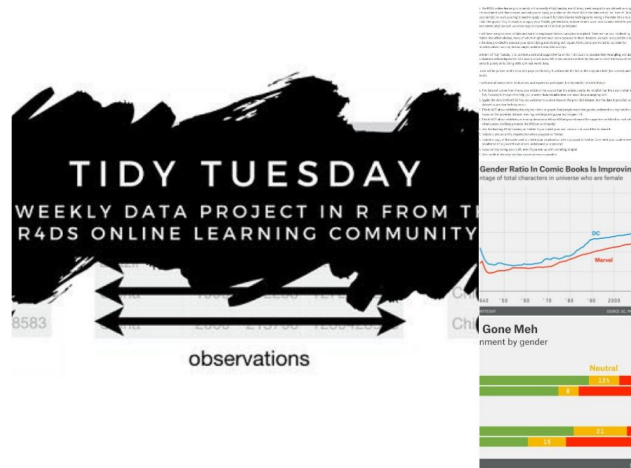


Pinned Tweet
R4DS online learning community @R4DScommunity · May 29
Welcome to week 9 of #tidytuesday!

Let's explore the world of comic book characters! This rich dataset comes from the [@FiveThirtyEight](#) package!

Data: bit.ly/tidyreadme
Article: 53eig.ht/2xlfokk

#rstats #dataviz #tidyverse #r4ds #comics



all categories ▾

all tags ▾

Latest



New (54)

Unread (2)

Categories

Top

+ New Topic

Topic	Replies	Views	Activity
<p>  Welcome to the RStudio Community!</p> <p>meta</p> <p>Welcome to community.rstudio.com — we're glad to have you! This welcome page will give you some advice on how to get the most out of the site if you're getting or giving help. We want this to be a friendly, inclusive com... read more</p>	0	4.5k	2018-07-22
<p><input type="checkbox"/> My app only loads properly some of the time; sometimes will not open •</p> <p>shinyapps.io shiny keras reticulate</p>	1	21	1m
<p><input type="checkbox"/> Preserve constant bar widths for uneven faceted plots? •</p> <p>tidyverse ggplot2 tidyverse</p>	1	14	1h
<p><input type="checkbox"/> Create columns with unique values •</p> <p>tidyverse</p>	1	16	1h
<p><input type="checkbox"/> Which test to use (animal behavior) •</p> <p>General rstudio</p>	1	10	3h
<p><input checked="" type="checkbox"/> capture streaming json over websocket</p> <p>General json websocket jsonlite</p>	7	193	5h
<p><input type="checkbox"/> Rstudio free desktop preview 1.2.1329 - can't select text in editor window</p> <p>RStudio IDE</p>	2	40	6h
<p><input type="checkbox"/> Project don't resume source panel files •</p> <p>RStudio IDE rstudio</p>	1	20	6h
<p><input type="checkbox"/> RStudio keeps crashing</p> <p>RStudio IDE</p>	3	49	6h
<p><input type="checkbox"/> tm package removing unwanted characters works in r but not knitr •</p> <p>General tm</p>	7	43	6h
<p><input type="checkbox"/> Rstudiocloud "The previous R session was abnormally terminated due to an unexpected crash" •</p> <p>RStudio Cloud crash-report 0 votes</p>	4	46	7h



R FOR DATA SCIENCE ONLINE LEARNING COMMUNITY



LEARNERS

Develop your skills in R and Data science with a friendly community



MENTORS

Help build a positive learning community for R users and expand your R debugging skills



COMMUNITY

Help develop resources for others to feel part of the #rstats community

R on Twitter!

```
> nrow(rstats_tweets)
[1] 429513
> range(rstats_tweets$created_at)
[1] "2008-09-07 19:54:26 UTC" "2018-12-19 19:32:01 UTC"
```

- #rstats
- #tidyverse
- #ggplot2
- #epitwitter
- #r4ds
- #RFeedbackFriday
- #rspatial
- Be wary of “infinite flow”

The image shows a screenshot of the #rstats Twitter page. The page is divided into several sections:

- Search filters:** A section with a 'Grow' button.
- Who to follow:** A list of accounts to follow, including Matt Miller (@mattmiller), R for Data Science (@rstats), and Rstats (@rstatstweet).
- Worldwide trends:** A section showing trending topics and their tweet counts, such as #ScPMoney (No freakin' fees + 2.25% APY), #LaRojaGHV (16.9K Tweets), Mueller (1.42M Tweets), Iowa State (12.3K Tweets), #FuerstFuode (4,533 Tweets), Harden (63.6K Tweets), Chile (219K Tweets), Ohio State (16.6K Tweets), #E3PhotoMedica (7,408 Tweets), and Backayn (5,840 Tweets).
- Tweets:** A list of tweets, including one from Hadley Wickham (@hadleywickham) asking for help with a survey, and one from Alison Hill (@alisonhill) discussing embedding a 'code download' button in HTML.
- World Map:** A world map showing the distribution of tweets, with a large cluster in North America.



Transforming science through open data and software

We build software with a community of users and developers, and educate scientists about transparent research practices.

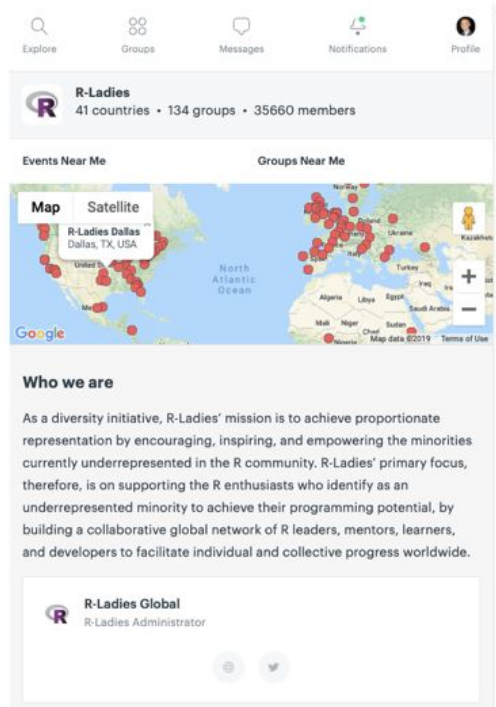
R-LADIES GLOBAL

R-LADIES IS A WORLD-WIDE ORGANIZATION TO PROMOTE GENDER DIVERSITY IN THE R COMMUNITY

HOME • ABOUT US • R-LADIES DIRECTORY • EVENTS • TWITTER • GITHUB • SLACK • BLOG • SPONSORS

MEETUP LOCATIONS

A complete list of all groups and meetups organised under R-Ladies globally may be found in the [R-Ladies organizational meetup](#), or check our founders lovely shiny dashboard!



Explore Groups Messages Notifications Profile

R-Ladies
41 countries • 134 groups • 35660 members

Events Near Me Groups Near Me

Map Satellite

R-Ladies Dallas
Dallas, TX, USA

Who we are

As a diversity initiative, R-Ladies' mission is to achieve proportionate representation by encouraging, inspiring, and empowering the minorities currently underrepresented in the R community. R-Ladies' primary focus, therefore, is on supporting the R enthusiasts who identify as an underrepresented minority to achieve their programming potential, by building a collaborative global network of R leaders, mentors, learners, and developers to facilitate individual and collective progress worldwide.

R-Ladies Global
R-Ladies Administrator



- SPONSORS -



consortium

- SOCIAL MEDIA -

SOCIAL ACCOUNTS BY CHAPTER

- COC -

CODE OF CONDUCT

R-LADIES IS A WORLDWIDE ORGANIZATION WHOSE MISSION IS TO PROMOTE GENDER DIVERSITY IN THE R COMMUNITY.

The R community suffers from an underrepresentation of minority genders (including but not limited to cis/trans women, trans men, non-binary, genderqueer, agender) in every role and area of participation, whether as leaders, package developers, conference speakers, conference participants, educators, or users ([see recent stats](#)).

As a diversity initiative, the mission of R-Ladies is to achieve proportionate representation by encouraging, inspiring, and empowering people of genders currently underrepresented in the R community. R-Ladies' primary focus, therefore, is on supporting minority gender R enthusiasts to achieve their programming potential, by building a collaborative global network of R leaders, mentors, learners, and developers to facilitate individual and collective progress worldwide.



Sharla Gelfand @sharlagelfand · Apr 19
the most imposter-syndrome inducing thing for me is fin
think is "really successful" is like, a year older than me

or worse, a year younger 🙄

🗨️ 14 🔄 4 ❤️ 157 ✉️



Jenny Bryan
@JennyBryan

Replying to @sharlagelfand

I was cured of this very early, sinc
supervisor @laan12345 is only a f
older than me. I then had 3 kids in
and was arguably working in the v
I was slow to launch. But I like wh
now! Let's hear it for the Slow Ca
movement.

10:07 AM - 19 Apr 2019

3 Retweets 171 Likes



🗨️ 5 🔄 3 ❤️ 171 ✉️



Hadley Wickham ✓
@hadleywickham

Following



I love the fact that when I don't know how to
do something in ggplot2, I can google it and
find out #rstats

7:16 AM - 28 Mar 2013



Hadley Wickham ✓
@hadleywickham

Following



I google for #rstats code all the time!

Freya Rowland @freshwaterfreya

I find it enormously comforting that several people with mad #rstats skills have
mentioned they often Google for code. #sohappyitsnotjustme

5:45 AM - 26 Jul 2016

93 Retweets 210 Likes



🗨️ 13 🔄 93 ❤️ 210 ✉️



THANK
YOU