

A Gentle Introduction to Tidy Statistics in R

Thomas Mock, PhD

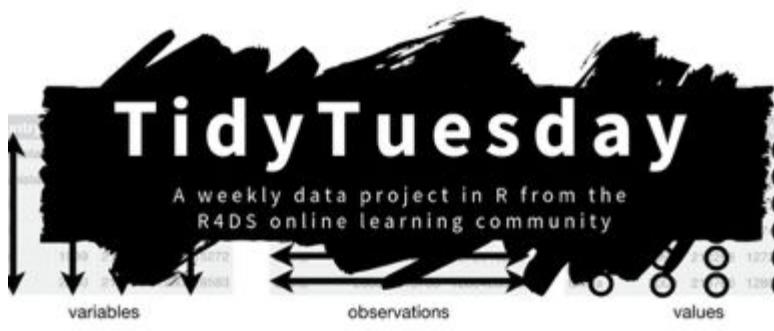
Who am I?



SPSS®

 Excel

 **ORIGIN®**
The Data Analysis and Graphing Workspace





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

Replies to @DavidKButlerUoA @kierisi and 3 others

Instead, the R-interested I meet are doing



Thomas Mock 🧑‍💻

@thomas_mock

Replies 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()

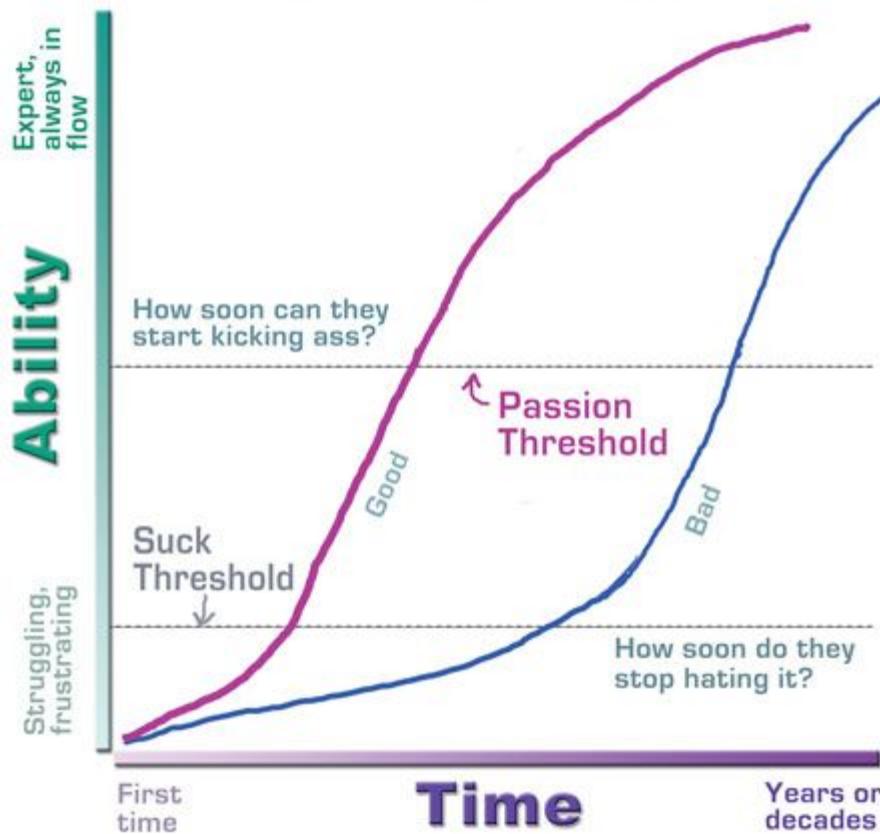
Time to make a plot, base plot 😬

8:05 AM - 5 Feb 2018

4 Likes



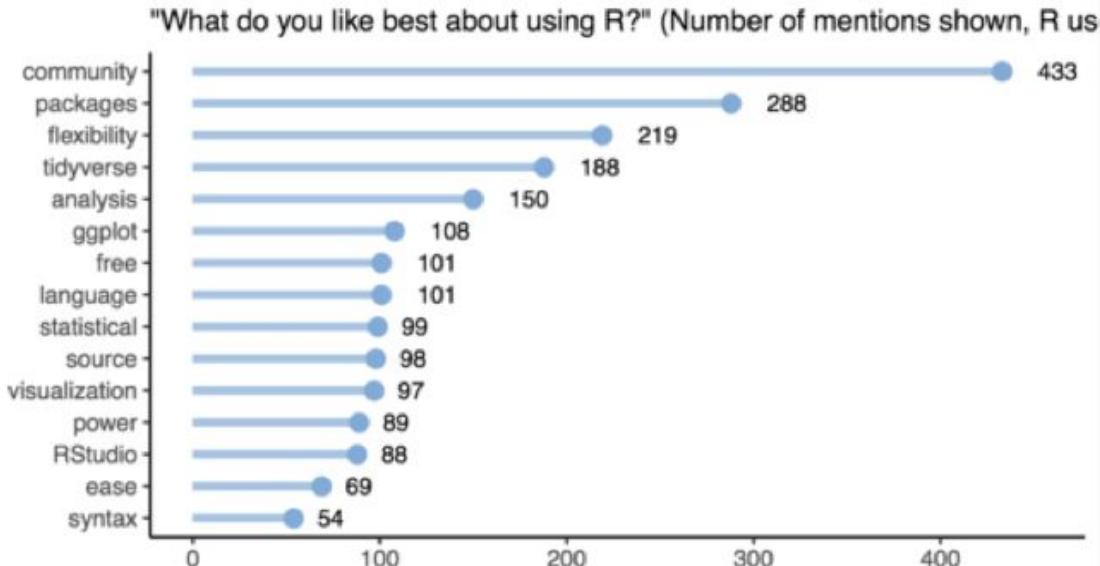
The Kick Ass Curve



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



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!?

牚牚 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

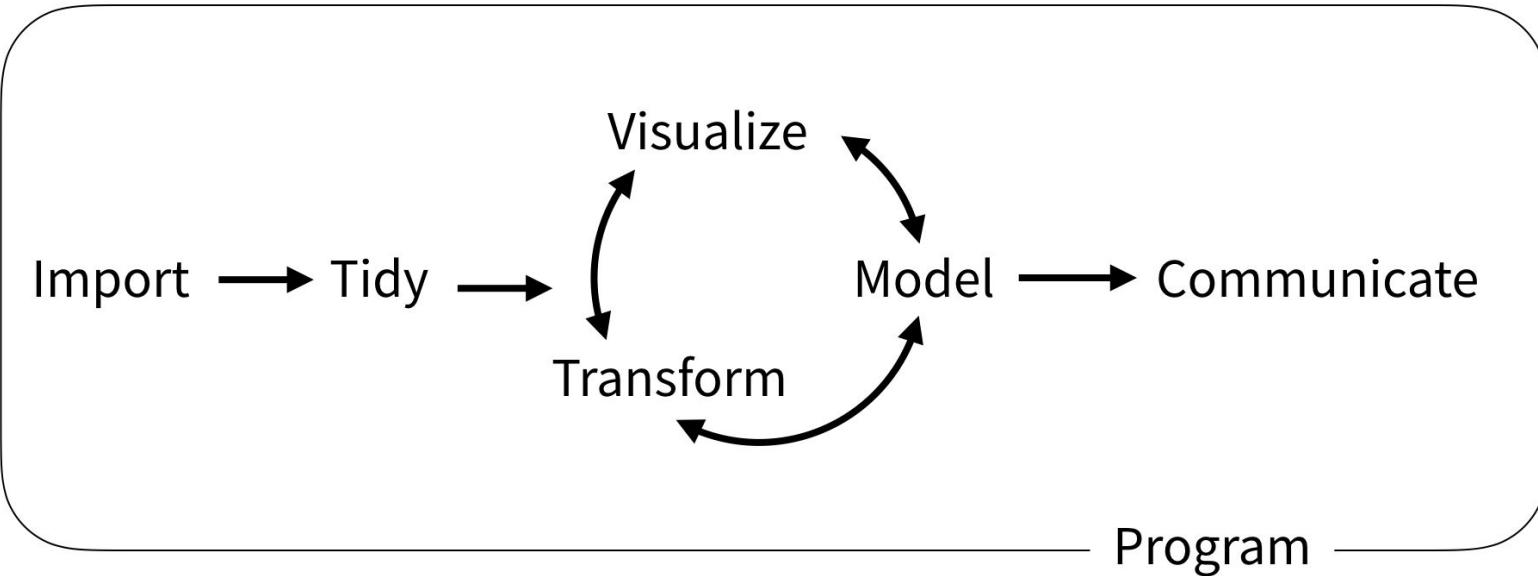
22



What we are covering today

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





A close-up photograph of a traffic light pole. The traffic light itself is in sharp focus, showing a solid red light. Behind it, a modern skyscraper with many windows is visible, though slightly out of focus, creating a sense of depth.

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

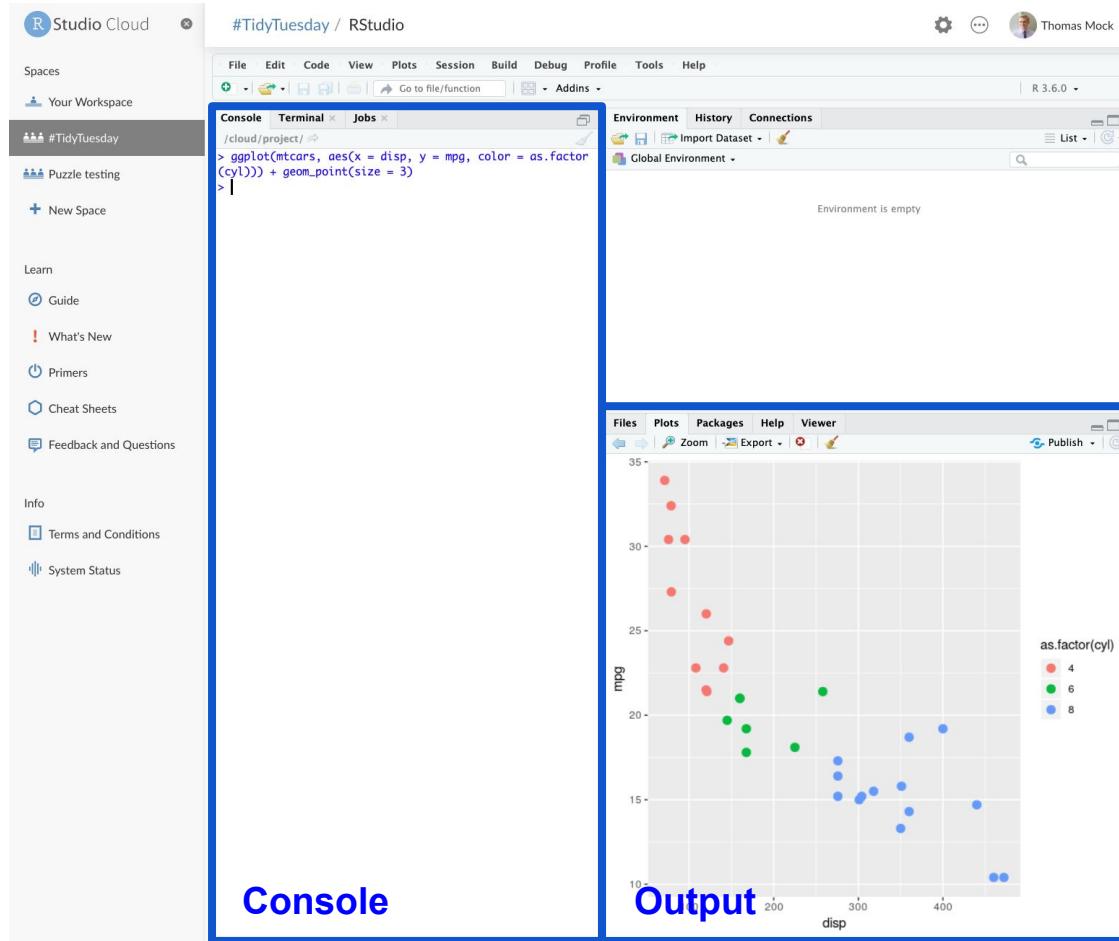


R: Engine



RStudio: Interface





R: New phone



R Packages:
Apps you can download



File Edit Code View Plots Session Build Debug Profile Tools Help

code_through.rmd x Go to file/function Addins

```
code_through.rmd
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 figure consists of a 3x2 grid of density plots. The columns represent 'Alzheimer's' and 'Healthy' status. The rows represent 'High Dose', 'Low dose', and 'Placebo' treatment groups. The x-axis for all plots is 'mmse' (ranging from 10 to 25) and the y-axis is 'density' (ranging from 0.0 to 0.4). The 'High Dose' row shows peaks around mmse 25. The 'Low dose' row shows peaks around mmse 15. The 'Placebo' row shows peaks around mmse 10.

Environment Hist

Import D.

Global Environment

Environment is empty.

Live demo

A photograph showing a person's legs and feet as they walk up a set of blue metal stairs. The person is wearing black leggings and bright orange sneakers with white soles. The stairs have a textured blue surface and are bordered by blue railings. The background shows more of the blue staircase structure.

Next steps!

Slides + RStudio Cloud instance

R Studio Cloud Your Workspace / intro_R

Spaces

Your Workspace

- #TidyTuesday
- Puzzle testing
- New Space

Learn

- Guide
- What's New
- Primers
- Cheat Sheets
- Feedback and Questions

Info

- Terms and Conditions
- System Status

File Edit Code View Plots Session Build Debug Profile Tools Help R 3.6.0

code_through.rmd

```
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)
38 ````{r}
```

The figure consists of four density plots arranged in a 2x2 grid. The top row is labeled 'Alzheimer's' and the bottom row 'Healthy'. The left column shows the 'High Dose' condition, and the right column shows the 'Low dose' condition. The x-axis for all plots is 'mmse' (ranging from 10 to 25) and the y-axis is 'density' (ranging from 0.0 to 0.4). The 'Alzheimer's' group generally has lower MMSE scores than the 'Healthy' group. The 'High Dose' and 'Low dose' plots show very similar distributions for both groups, while the 'Placebo' plot shows a distinct shift towards higher MMSE scores for the 'Healthy' group.

Environment History Con

Global Environment

Environment is empty

Info Access

Who can view this project

Everyone

Files Plots Packages H

New Folder Upload Delete

Cloud > project

Name

43

RStudio Cloud Primers

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

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

The screenshot shows the RStudio Cloud Primers interface. At the top, there's a navigation bar with links: Learn, Guide, What's New, Primers (which is highlighted in blue), and Cheat Sheets. To the right of the navigation is a user profile for Thomas Mock. On the left, a sidebar menu includes: Spaces, Your Workspace, #TidyTuesday, Puzzle testing, + New Space, Learn, Guide, What's New, Primers (highlighted in blue), Cheat Sheets, Feedback and Questions, Info, Terms and Conditions, and System Status.

R Studio Primers

Learn data science basics with the interactive tutorials below.

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`, `tidyverse`, 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 or @thomas_mock)
- Browse other submissions and like/comment on their work!

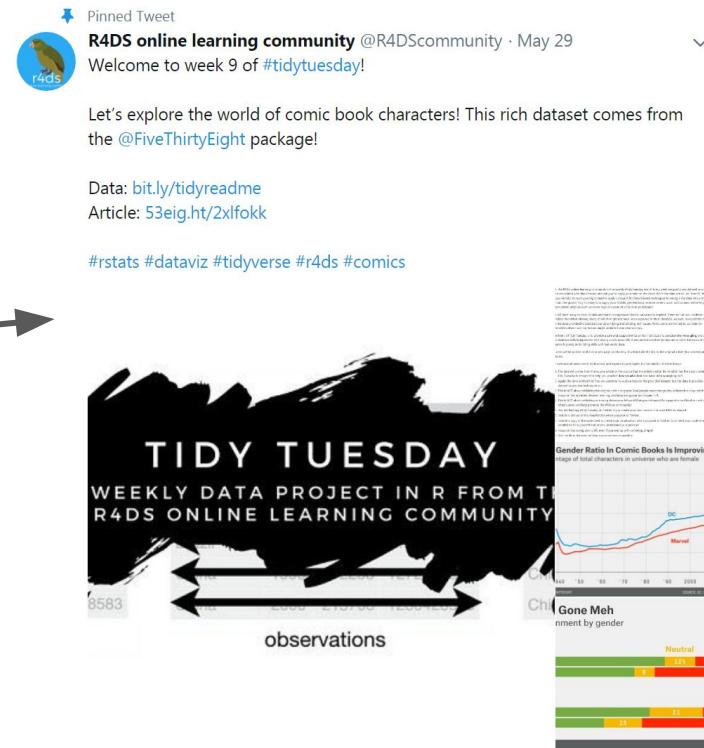
Full details, data, code of conduct: <https://github.com/rfordatascience/tidytuesday>

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

 Welcome to the RStudio Community![meta](#)

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](#)



0

4.5k

2018-07-22

 My app only loads properly some of the time; sometimes will not open •[shinyapps.io](#) [shiny](#) [keras](#) [reticulate](#)

1

21

1m

 Preserve constant bar widths for uneven faceted plots? •[tidyverse](#) [ggplot2](#) [tidyverse](#)

1

14

1h

 Create columns with unique values •[tidyverse](#)

1

16

1h

 Which test to use (animal behavior) •[General](#) [rstudio](#)

1

10

3h

 capture streaming json over websocket[General](#) [json](#) [websocket](#) [jsonlite](#)

C



J



H



P

7

193

5h

 Rstudio free desktop preview 1.2.1329 - can't select text in editor window[RStudio IDE](#)

K

2

40

6h

 Project don't resume source panel files •[RStudio IDE](#) [rstudio](#)

H



K

1

20

6h

 RStudio keeps crashing[RStudio IDE](#)

K



Q



V

3

49

6h

 tm package removing unwanted characters works in r but not knitr •[General](#) [tm](#)

Q



M



W

7

43

6h

 Rstudiocloud "The previous R session was abnormally terminated due to an unexpected crash" •[RStudio Cloud](#) [crash-report](#) 0 votes

V



M



W

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!

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

```
> 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"
```

The screenshot shows the Twitter profile for the account @rstats. The header reads '#rstats'. The top navigation bar includes 'Top', 'Latest', 'People', 'Photos', 'Videos', 'News', and 'Broadcasts'. Below the header, there's a 'Search filters' section with a dropdown set to 'Show'. The 'Who to follow' section lists three accounts: Matt Miller (@millerdatasci), R for Data Science (@rstudio), and Rstats (@rstatsweet). There's also a link to 'Find people you know'. On the right side, there are two tweets from Hadley Wickham (@hadleywickham): one about pivot verbs in tidyverse and another about code download buttons in R Markdown. Below that, there's a tweet from Alison Hill (@preashill) about embedding code download buttons. The bottom of the screenshot shows a snippet of a tweet from Thomas Eliaz Hege (@thomashege) about YAML, followed by a screenshot of a Shiny app interface.

Transforming science through open data and software

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

ERIN
LEWELL

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!

The screenshot shows the 'Meetup Locations' section of the R-Ladies Global website. At the top, there's a navigation bar with links to Home, About Us, R-Ladies Directory, Events, Twitter, GitHub, Slack, Blog, and Sponsors. Below the navigation is a search bar and a purple header bar with the R-Ladies logo and the text 'R-Ladies 41 countries • 134 groups • 35660 members'. The main content area features two maps: 'Events Near Me' (showing locations in the United States) and 'Groups Near Me' (showing locations around the world). Below the maps is a section titled 'Who we are' which contains a detailed paragraph about the organization's mission to promote gender diversity in the R community. At the bottom, there's a footer with the R-Ladies Global logo and the text 'R-Ladies Administrator'.

The screenshot shows the R-Ladies Consortium website. It features a large 'R Ladies' logo at the top left, followed by a search bar and a 'SPONSORS' section. Below this is another 'R' logo with the word 'consortium' underneath it. There are sections for 'SOCIAL MEDIA' (with links to various platforms), 'SOCIAL ACCOUNTS BY CHAPTER', and 'CODE OF CONDUCT'. At the bottom, there are social media icons for LinkedIn, GitHub, and Twitter.

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 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.



Sharla Gelfand @sharlagelfand · Apr 19

the most imposter-syndrome inducing thing for me is fine
think is "really successful" is like, a year older than me

or worse, a year younger 😱



14



4



157



Jenny Bryan

@JennyBryan

Replies to @sharlagelfand

I was cured of this very early, since my supervisor @laan12345 is only a few years older than me. I then had 3 kids in my late 20s and was arguably working in the field of data science before I was slow to launch. But I like where we are now! Let's hear it for the Slow Camp 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