

# Scaffolding for a community of practice

---

Tom Mock

@thomas\_mock

[github.com/jthomasmock/tidytuesday\\_presentation-user-2020](https://github.com/jthomasmock/tidytuesday_presentation-user-2020)

2020-06-20

# A weekly social data project in R

**Designed to help you practice data visualization and basic data wrangling**

- Step 1: A dataset is posted with a data dictionary and a related article
- Step 2: Participants pull data into R, then clean, analyze and plot the data
- Step 3: Participants post the results to Twitter with #TidyTuesday
- Step 4: We all learn from each other!

Check it out at: [bitly.com/tidyreadme](https://bitly.com/tidyreadme)

# Some data

6,358 Tweets

2,312 GitHub stars

2,035 Unique visits last  
week

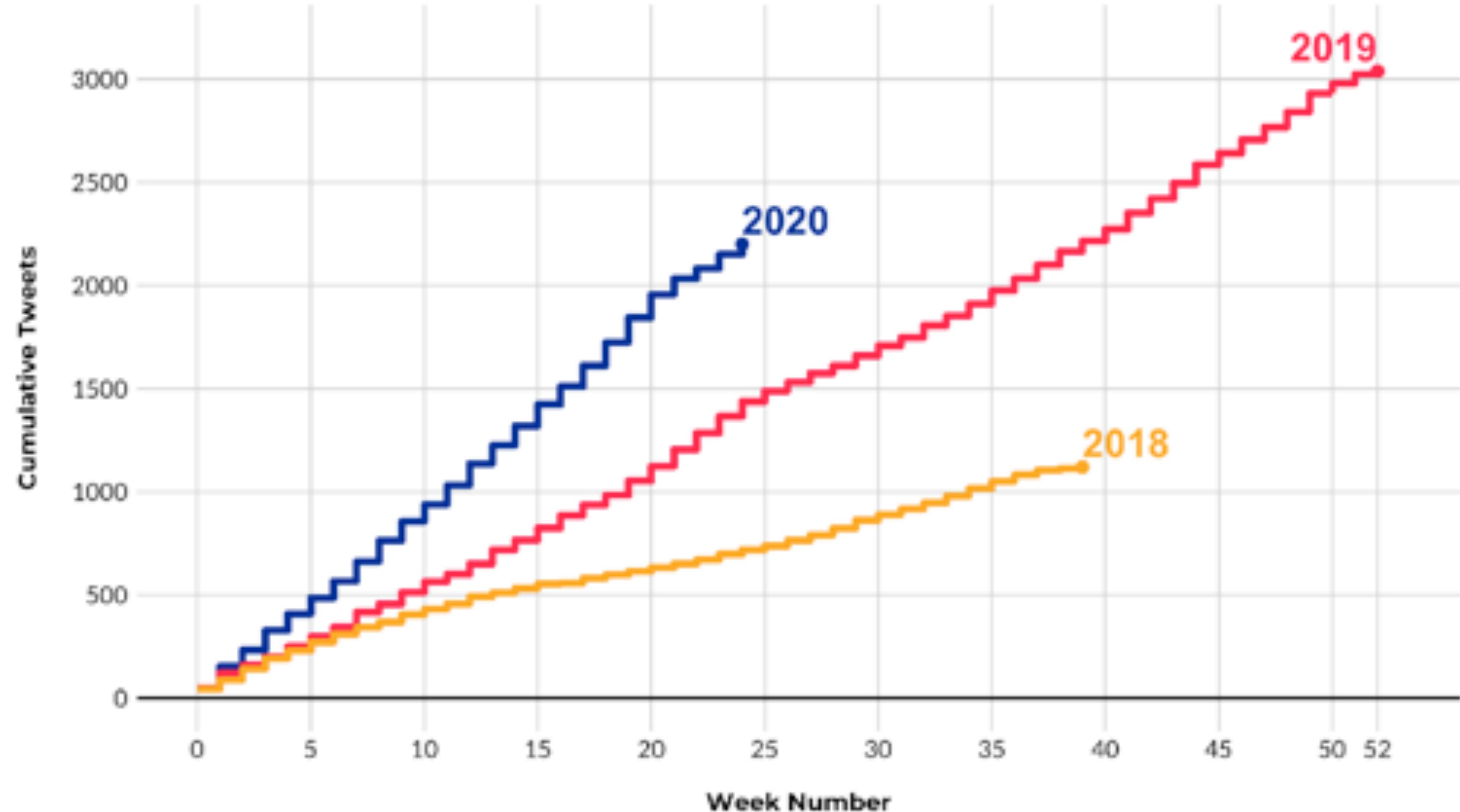
1,582 Unique Participants

116 weeks of data

Code: [bit.ly/2YndHP2](https://bit.ly/2YndHP2)

## Cumulative tweets for #TidyTuesday by year

Note that Week 1 of 2018 started in April & tweets must contain: 'rstats, code, plot, graph, viz, data or tidyverse'



Data: rtweet | Plot: @thomas\_mock

# What is TidyTuesday?

## It's just data.

- That can be read into R in a few seconds
- With an article for context
- With a data dictionary
- With dozens of example analyses **WITH** code every week
- With a community of learners and mentors

# What is TidyTuesday?

It depends on who you ask

- “I copy-paste some code and change some of it”
- “I read in a file from the internet, visualize it and post it to Twitter”
- “I do a live screencast, run some models, push the output to GitHub, and post the analysis to Twitter”
- “I start a blog post, write some code, do an analysis, and post the results to Twitter”

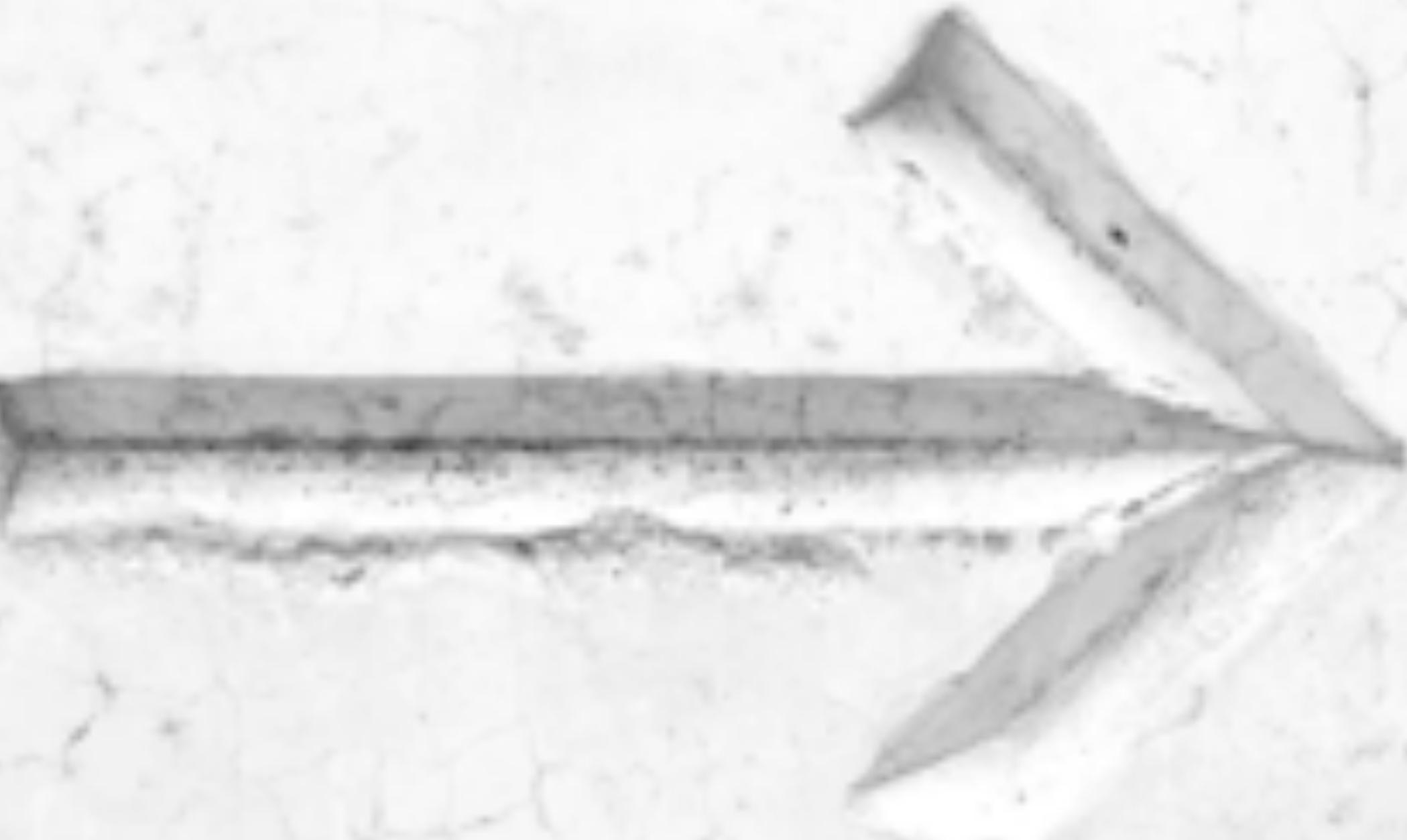
# **What is TidyTuesday?**

A scaffold  
for a self-directed  
community of practice

# Scaffolding

In education, scaffolding refers to a variety of instructional techniques used to **move** students **progressively toward** stronger **understanding** and, ultimately, greater **independence** in the **learning process**.





**Self  
Directed**

# A community of practice

- “A group of people who **share** a **passion** for something they do, and **learn** how to **do it better** as they **interact regularly**”
- **Voluntary** and driven by the community
- Successful through: “enough excitement, relevance, and value to attract and engage members”
- This presentation is not a story about me, but rather the #TidyTuesday project, the **community** that has adopted it, extended it, and made it **alive**

# Cultivating Communities of Practice

## Harvard Business School

- Create a rhythm for the community
- Design for evolution
- Open a dialogue between inside and outside perspectives
- Invite different levels of participation
- Develop both public and private community spaces
- Focus on value
- Combine familiarity and excitement

# Create a rhythm for the community



# Timeline

- Monday AM

“The @R4DScommunity welcomes you to week {n} of #TidyTuesday!

We're exploring {data}!”

- Tuesday AM

“It's #TidyTuesday y'all!”

 **Tom Mock**  
@thomas\_mock

The [@R4DScommunity](#) welcomes you to week 26 of #TidyTuesday!  
We're exploring Caribou locations!!

 [bit.ly/tidyreadme](https://bit.ly/tidyreadme)  
 [bit.ly/37XuqeT](https://bit.ly/37XuqeT) #r4ds #tidyverse #rstats #dataviz

  
Tidy Tuesday  
A weekly data project by & from the R4DS online learning community



37 9:27 AM - Jun 22, 2020

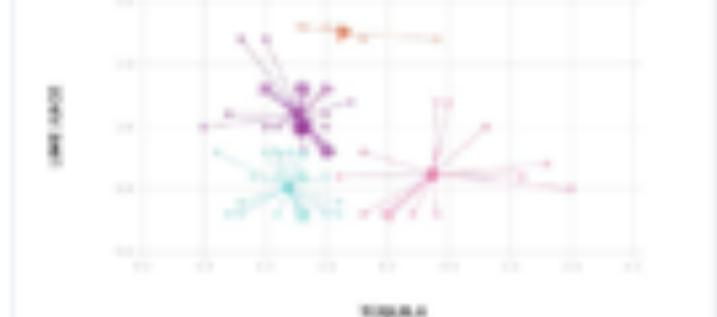
 See Tom Mock's other Tweets

 **R4DS online learning community**  
@R4DScommunity

It's #TidyTuesday y'all! Now go make something neat!  
[twitter.com/thomas\\_mock/st...](https://twitter.com/thomas_mock/st...)

 [bit.ly/tidyreadme](https://bit.ly/tidyreadme)  
 [53elg.h/3c5bfQB](https://53elg.h/3c5bfQB) #r4ds #tidyverse #rstats #dataviz

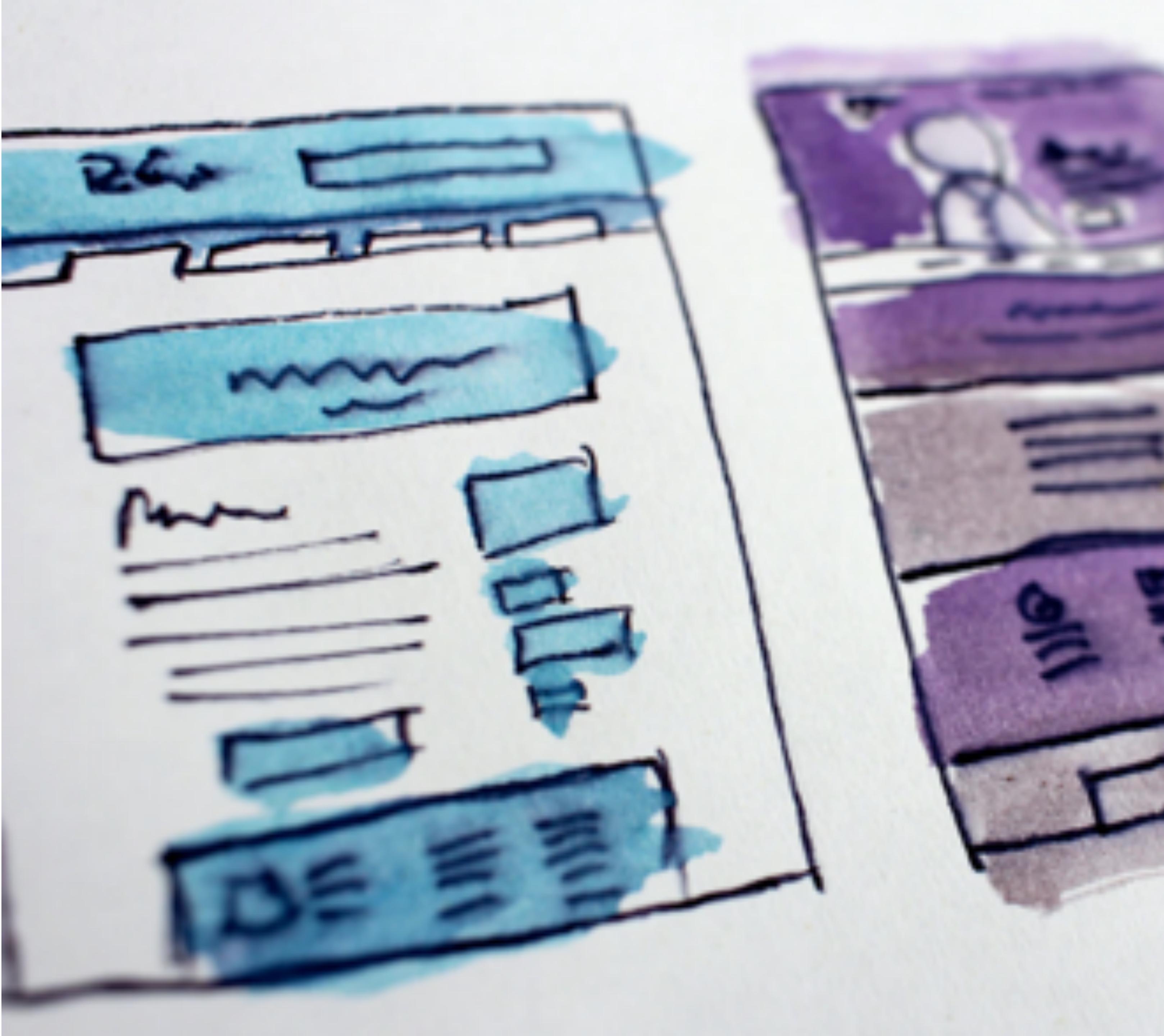
  
Tidy Tuesday  
A weekly data project by & from the R4DS online learning community



16 7:53 AM - May 26, 2020

 See R4DS online learning community's other Tweets

# Design for evolution



# The birth of an online community

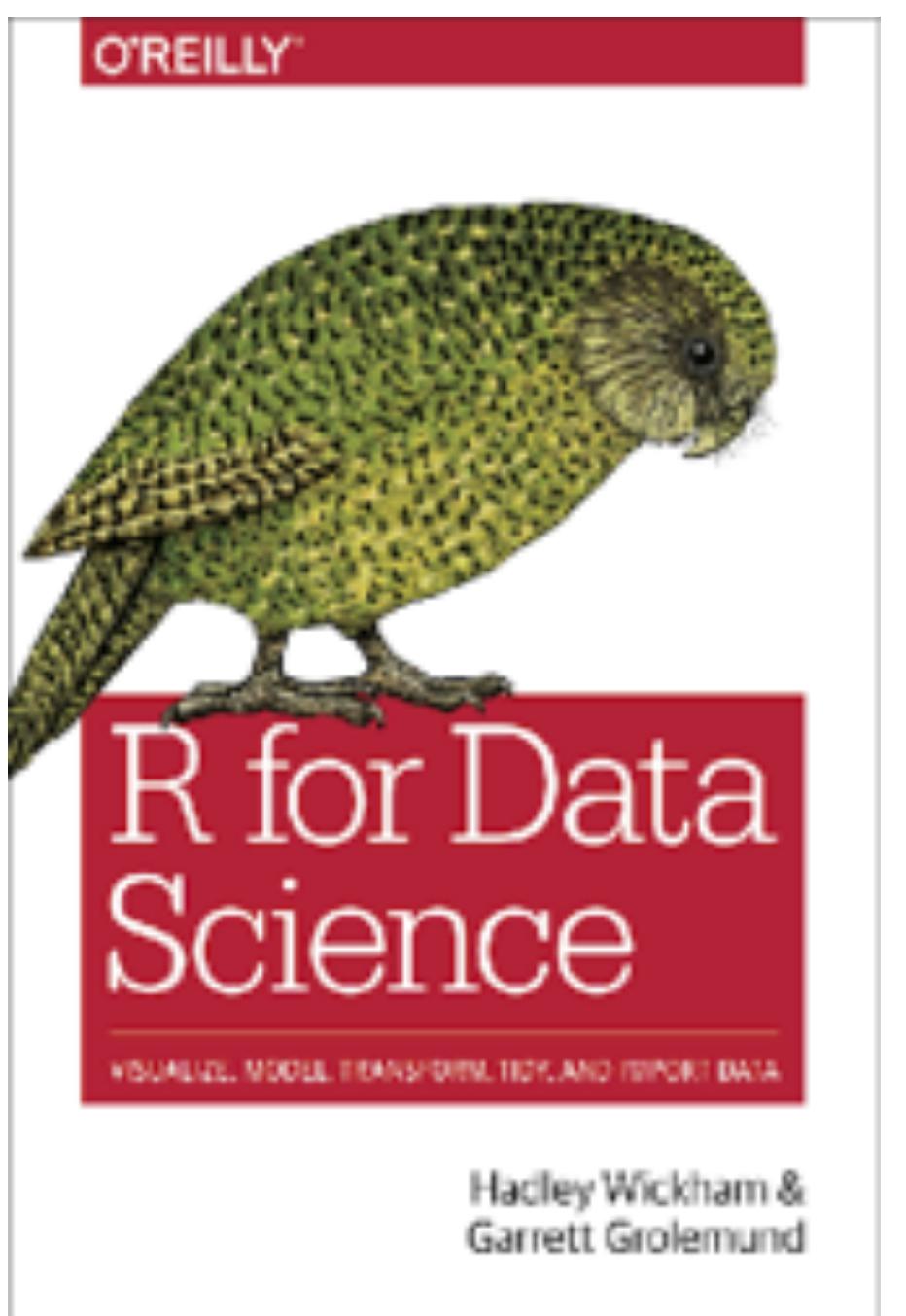
- A place to read a book
- A place to learn
- A place to mentor
- A place to build
- A **community**



Jesse Mostipak  
@kierisi



@r4dscommunity



r4ds.had.co.nz

<https://www.jessemaegan.com/post/join-the-r-for-data-science-online-learning-community/>

<https://www.jessemaegan.com/post/r4ds-v1-v2-a-retrospective/>

# A retrospective, by Jesse

“I can’t do this all by myself - what do YOU want to see in this community?”

“That’s a fantastic idea, how can I help support YOU in making that a reality?”

Communities evolve and it’s important to let others fill roles, become leaders, engage with each other.

Open a  
dialogue  
between  
inside and  
outside  
perspectives



# Make it easy

## Cleaning Script



### Caribou

This data tr  
tags of 260

h/t to [Alex C](#)

The tracking  
management  
Nature's IU

The caribou

```
# Load libraries
library(tidyverse)
library(janitor)

# Import data
individuals_raw <- read_csv("./caribou-location-tracking/raw/Mountain caribou in British Columbia-reference-data.csv")
locations_raw <- read_csv("./caribou-location-tracking/raw/Mountain caribou in British Columbia-gps.csv")

# Clean individuals
individuals <- individuals_raw %>%
  clean_names() %>%
  transmute(animal_id,
            sex = animal_sex,
            # Getting rid of whitespace to address inconsistent spacing
            # NOTE: life_stage is as of the beginning of deployment
            life_stage = str_remove_all(animal_life_stage, " "),
            reproductive_condition = animal_reproductive_condition,
            # Cause of death "cod" is embedded in a comment field
            death_cause = str_remove(animal_death_comments, ".*cod "),
            study_site,
            deploy_on_longitude,
            deploy_on_latitude,
            # Renaming to maintain consistency "deploy_on_FIELD" and "deploy_off_FIELD"
            deploy_on_comments = deployment_comments,
            deploy_off_longitude,
            deploy_off_latitude,
            deploy_off_type = deployment_end_type,
            deploy_off_comments = deployment_end_comments) %>%
  # reproductive_condition actually has two dimensions
  separate(reproductive_condition, into = c("pregnant", "with_calf"), sep = ";", fill = "left") %>%
  mutate(pregnant = str_remove(pregnant, "pregnant: ?"),
        with_calf = str_remove(with_calf, "with calf: ?")) %>%
  # TRUE and FALSE are indicated by Yes/No or Y/N
  mutate_at(vars(pregnant:with_calf), ~ case_when(str_detect(., "Y") ~ TRUE,
                                                str_detect(., "N") ~ FALSE,
                                                TRUE ~ NA))

# Description of variables
animal_id: is fitted with a location-tracking tag.
sex: individual identifier for animal
age: age class (in years) at beginning of deployment
pregnant: whether animal was pregnant at beginning of deployment
had_calf: whether animal had a calf at time of deployment
cause_of_death: cause of death
study_site: deployment site or colony, or a location-related group such as the herd or pack
time: time
deploy_on_longitude: longitude where animal was released at beginning of deployment
deploy_on_latitude: latitude where animal was released at beginning of deployment
deployment_comments: additional information about tag deployment
deploy_off_longitude: longitude where deployment ended
deploy_off_latitude: latitude where deployment ended
deployment_end_type: classification of tag deployment end (see table below for full description)
deploy_off_comments: additional information about tag deployment end
```

is fitted with a location-tracking tag.

description

individual identifier for animal

sex of animal

age class (in years) at beginning of deployment

whether animal was pregnant at beginning of deployment

whether animal had a calf at time of deployment

cause of death

deployment site or colony, or a location-related group such as the herd or pack  
time

longitude where animal was released at beginning of deployment

latitude where animal was released at beginning of deployment

additional information about tag deployment

longitude where deployment ended

latitude where deployment ended

classification of tag deployment end (see table below for full description)

additional information about tag deployment end

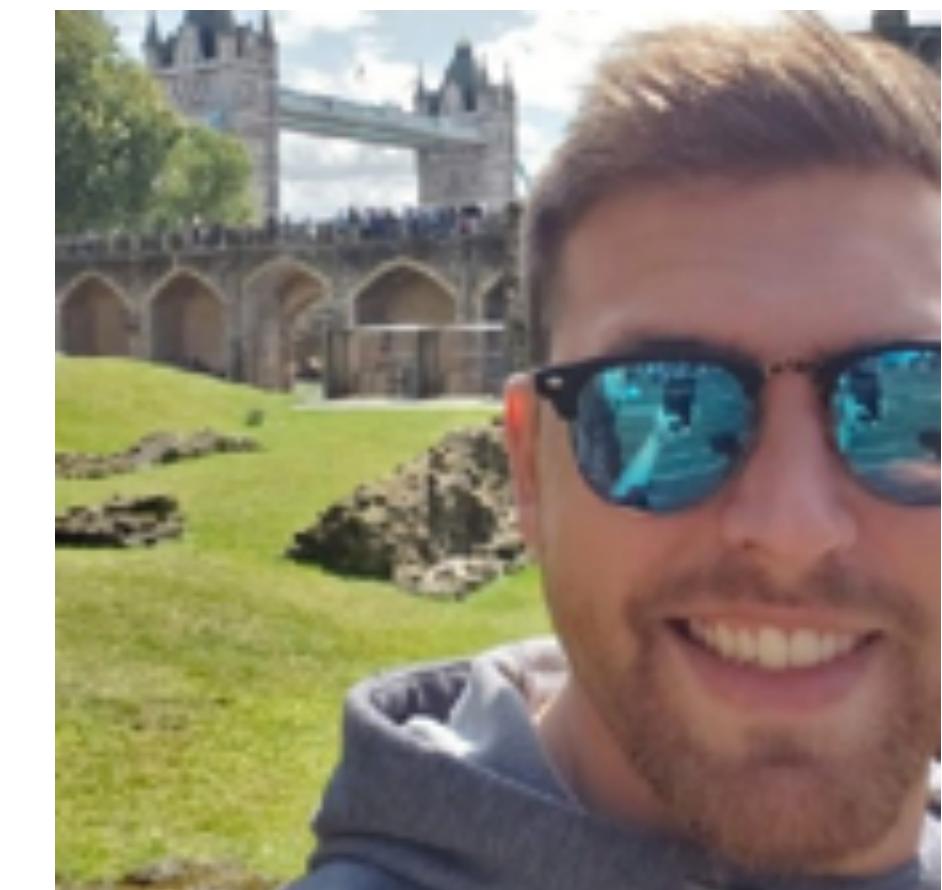
# Make it easieR

```
devtools::install_github("thebioengineer/tidytuesdayR")
```

“{tidytuesdayR} has the main goal to make it easy to participate in the weekly #TidyTuesday project. Currently this is done by assisting with the import of data posted on the R4DataScience Tidy Tuesday repository.”

tidytuesdayR lets you load:

- The readme
- The data
- All by passing just a date or week + year



Ellis Hughes  
@ellis\_hughes

# tidytuesday.rocks

- Shiny + #TidyTuesday posts



Neal Grantham  
@nsgrantham

## tidytuesday.rocks

Tidy Tuesday is a weekly social data project in R. Every week @thomas\_mock and @R4DSCommunity post a new dataset and ask R users to explore it and share their findings on Twitter with #TidyTuesday.

Since the first dataset was posted on April 2nd, 2018, there are now 111 datasets and 4,613 #TidyTuesday tweets from 1,827 users! Use the options below to filter the tweets by dataset or Twitter user and sort them by date, likes, and retweets.

tidytuesday.rocks is about 150 lines of R code and relies on your #TidyTuesday tweets, which I scrape and label every few weeks. It is built with [Shiny](#) and its source code is [on GitHub](#).

The response to tidytuesday.rocks has been amazing! It was even awarded a runner up spot in the 2019 Shiny Contest among 136 submissions. 🎉

I'd love to hear your feedback, say hi [@nsgrantham](#).

Happy plotting!

[Filter by Dataset](#)    [Filter by User](#)

Choose a dataset

Animal Crossing

- Animal Crossing
- Broadway Musicals
- GDPR Violations
- Best Rap Artists
- Tour de France
- Beer Production
- Traumatic Brain Injury
- The OJ Trial

Animal Crossing

Data Article Source

Tweets sorted by most retweets



Ted "Business Goose" Laderas  
@laderas

For this week's Animal Crossing #tidytuesday, I learned how to use the gt package to make a formatted table of the most expensive items in #ANCH. Gold Toilet FTW.

name	img_name	category	image_url
Royal Crown	1200000_Hats	Hats	
Crown	1000000_Hats	Hats	
Gold Armor	320000_Dresses	Dresses	
Golden Casket	320000_Furniture	Furniture	
Grand Piano	280000_Furniture	Furniture	
Golden Toilet	240000_Furniture	Furniture	

81 9:16 PM - May 4, 2020

22 people are talking about this

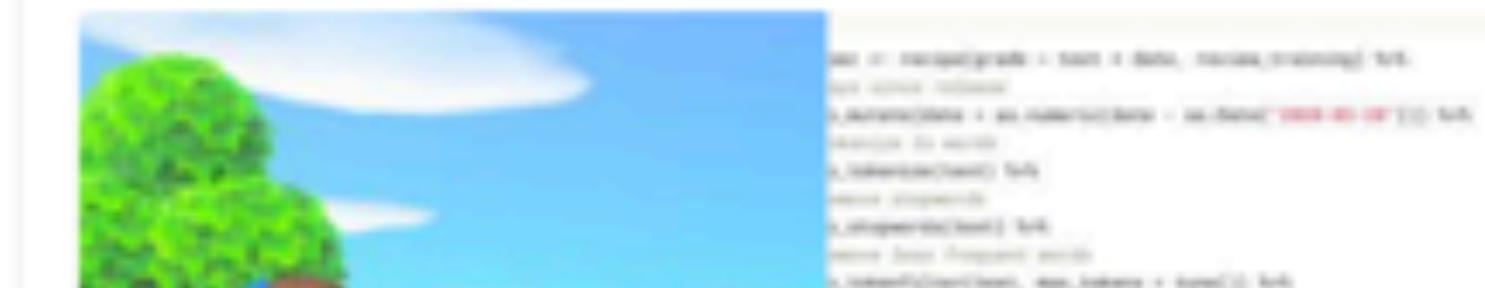


Emil Hvitfeldt  
@Emil\_Hvitfeldt

I'm starting a {textrecipes} blog post series!

The first post will be focused on Term Frequencies with step\_tf()

And we are lucky to be able to talk about Animal Crossing with this weeks #tidytuesdayhvitfeldt.me/blog/textrecip...#rstats #tidymodels #tidytext



**Invite  
different  
levels of  
participation**



learners &  
mentors &  
learners &  
mentors &  
learners &  
mentors &  
learners &  
mentors &

- “We welcome all newcomers, enthusiasts, and experts to participate”
- Regardless of your follower count, #TidyTuesday is an instant audience
- Learn new skills or expand your existing skillset
- Share your expertise
- Solve a “never-ending stream of novel problems”

# 52 weeks of TidyTuesday

Jake Kaupp  
@jakekaupp

And with that last submission marks my goal of completing the full year of [#TidyTuesday](#). The whole 52. You can find all of my submissions and code @ [bit.ly/2TMiRyW](https://bit.ly/2TMiRyW). I'm planning to write a blog post about it sometime in the New Year. But for now..



IT'S OVER

IT'S DONE

209 3:47 PM - Dec 27, 2019

33 people are talking about this

Entries for tidytuesday

r rds tidytuesday

161 commits 1 branch 0 packages 0 releases 1 contributor

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

		Latest commit dc6a2b1 yesterday
2018	finished week32	11 months ago
2019	finished 2020 week17	2 months ago
2020	cropped image	yesterday
.gitignore	added week12	3 months ago
README.md	restructured and added README	2 years ago
tidytuesdays.Rproj	restructured and added README	2 years ago
README.md		

## Whats all this then?

It's my repository for #tidytuesday entries! I'll include code and submission week by week.

## What's #TidyTuesday?

Either you have an interest in cleaning house on Tuesday or you like #rstats, #tidyverse & #R4DS!

TidyTuesday is an initiative grown out of Jesse Maegan's (@kierisi) #R4DS & the community mentor group, started by Thomas Mock (@thomas\_mock). You can read all about it and get involved [here](#)

<https://github.com/jkaupp/tidytuesdays>

# Learn by watching an expert on a weekly #TidyTuesday screencast



Julia Silge  
@juliasilge

[youtube.com/channel/UCCTTBgWyJI2HrrhQOOc710kA](https://youtube.com/channel/UCCTTBgWyJI2HrrhQOOc710kA)

[juliasilge.com/blog/](http://juliasilge.com/blog/)



Dave Robinson  
@drob

[youtube.com/user/safe4democracy](https://youtube.com/user/safe4democracy)

[github.com/dgrtwo/data-screencasts](https://github.com/dgrtwo/data-screencasts)

# Learn by watching experts walk through code



Patrick Ward  
@osppatrick



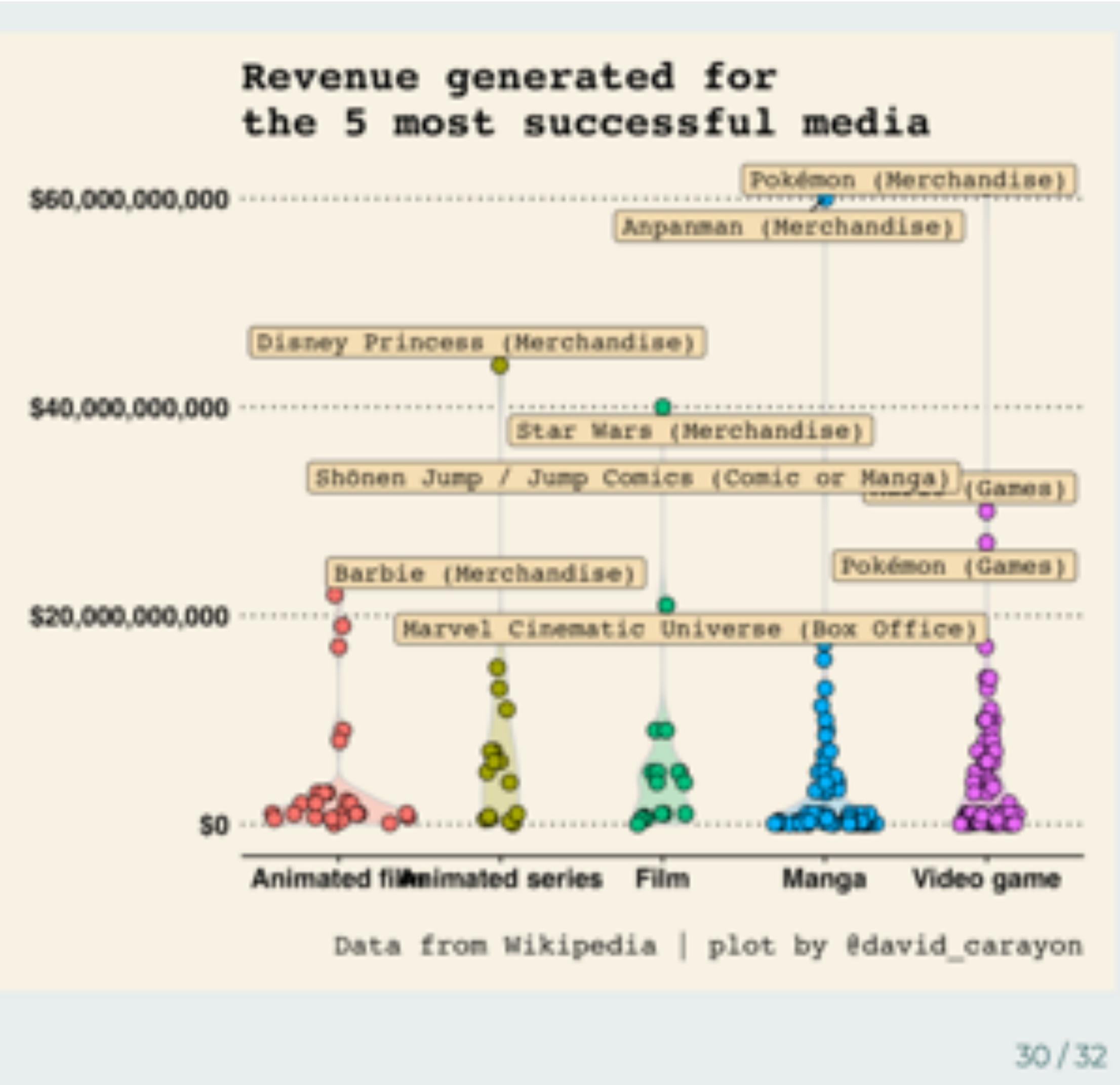
Ellis Hughes  
@ellis\_hughes



[youtube.com/channel/UCP8I94xtoemCH\\_GxByvTuFQ](https://youtube.com/channel/UCP8I94xtoemCH_GxByvTuFQ)

# Learn chunk by chunk with Flipbooks

```
ggplot(data = media_franchises_5) +  
  aes(x = original_media) +  
  aes(y = revenue) +  
  aes(fill = original_media) +  
  geom_violin(alpha = 0.3) +  
  ggforce::geom_sina(color = "black",  
                     size = 3, shape = 21) +  
  aes(label = category) +  
  ggrepel::geom_label_repel(  
    data = media_franchises_5 %>%  
      filter(revenue > 10 * 10^9),  
    color = "black", fill = "wheat",  
    family = "mono", size = 4) +  
  guides(fill = FALSE, color = FALSE) +  
  labs(x = "Original Media", y = "Total revenue") +  
  scale_y_continuous(label = scales::dollar) +  
  ggthemes::theme_minimal() +  
  labs(title = "Revenue generated for the 5 most successful media") +  
  theme(title = element_text(size = 16)) +  
  labs(caption = "Data from Wikipedia | plot by @david_carayon")
```



Gina Reynolds  
@EvaMaeRey

# Develop both public and private community spaces



# TidyTuesday Meetups



<https://www.meetup.com/Boston-useR/events/269130710/>



UCSB Bren School  
@brenucsb



Read how Bren data analysis/stats instructor [@allison\\_horst](#) and student [@GracieGW](#) launched a [#TidyTuesday](#) meetup, a no-stress, low-effort coding club to bring people together for "fun social coding." Which is definitely a thing. [openscapes.org/blog/2019/05/02/tidy-tuesday-coding-club/](https://openscapes.org/blog/2019/05/02/tidy-tuesday-coding-club/) [@juliesquid](#) [#Rstats](#) [#R4DS](#)



1 41 1:37 PM - Jul 8, 2019



[See UCSB Bren School's other Tweets](#)



<https://twitter.com/brenucsb/status/1148300034326163460?s=20>  
<https://www.openscapes.org/blog/2019/05/02/tidy-tuesday-coding-club/>  
[https://twitter.com/allison\\_horst/status/1113242674113286144](https://twitter.com/allison_horst/status/1113242674113286144)

# The R4DS Slack group



**R4DS online learning community**  
@R4DScommunity



Today we had our second [#TidyTuesday](#) Virtual Meetup! The recording starts part-way through, but join us on YouTube: [youtube.com/watch?v=kilx7R...](https://youtube.com/watch?v=kilx7R...)

And then on Slack: [r4ds.io/join!](https://r4ds.io/join)

Thank you [@CedScherer](#) for making this happen, and [@alexcookson](#) for moderating today! [#RStats](#)



[Create Account](#)

Slack is where work flows. It's where the people you need, the information you share, and the tools you use come together to get things done.

[rfordatascience.slack.com](https://rfordatascience.slack.com)

18 2:45 PM - May 8, 2020



[See R4DS online learning community's other Tweets](#)



Come join us at: [r4ds.io/join](https://r4ds.io/join)



Cédric Scherer  
@CedScherer



Jon Harmon  
@jonthegeek



Alex Cookson  
@alexcookson



@r4dscommunity

# TidyTuesday Venues

- Classroom - extend your lectures with “free exploration”
- Small Group - work together on a shared problem
- Industry - some using it as a framework for interview tests

#TIDYTUESDAY WORKSHOP

JULY 23, 2019 @ 7:00 PM - 9:00 PM UTC+0



BouldeR Meetup  
*#TidyTuesday*

Tuesday, May 19, 2020

## [ONLINE] #tidytuesday Work Group



Hosted by  
Stephanie Z. and 2 others



Stephanie Spielman, PhD  
@stephsspiel



This semester, my class [#DataScienceforBiologists](#) will be participating in [#TidyTuesday](#) for extra credit, and the first ones are in! I am \*so\* incredibly proud of their work - they started with R for the very first time literally a month ago and they are crushing it!! 1/n

62 8:07 AM - Feb 21, 2020



See Stephanie Spielman, PhD's other Tweets





Focus on value

# Ten reasons why I love Tidy Tuesday

- Getting inspired by the creativity of others
- Discovering new functions and packages
- Learning to make plots self-explanatory
- Practicing tidyverse skills
- Learning to fine-tune graphs
- Discovering new data every week
- Being part of the #rstats community
- Exploring interesting data sets
- Trying different plot types
- Getting inspired to tidy up your home



Veerle van Son  
@veerlevanson

# Build your career through projects

“Learning on your own means that you don’t have a teacher or role model....The best way to counteract the lack of a teacher is to **find a community of people where you can ask questions.**”

**“One great example is the TidyTuesday program** (<https://github.com/rfordatascience/tidytuesday>); ... every Tuesday, aspiring and practicing data scientists use R to tackle a data science problem.”

bestbook.cool (for the whole book)

<https://livebook.manning.com/book/build-your-career-in-data-science/chapter-3/109>



Jacqueline Nolis @skyetetra



Emily Robinson @robinson\_es

# “The unreasonable effectiveness of public work

**How I used to think of my goals:**



“One of the most remarkable projects that enables exploration of new data is #TidyTuesday”

**How I should have been thinking of them:**



“Share what you can learn with it”



Dave Robinson  
@drob

# Get hired

“My #rstats friends, can you believe that I got a job interview and possibly an offer because I made this treemap for #TidyTuesday?”

“You never know who is watching you till you get calls from those who would like hire you. Keep up the good work and share with us!”



# Learn by learning about learning

## Becoming Tidier Over Time: Studying #TidyTuesday as a Social Media-Based Context for Learning to Visualize Data

Joshua M. Rosenberg<sup>1</sup>, Anthony Schmidt<sup>1</sup>, Aaron Rosenberg<sup>2</sup>, Jennifer Longnecker<sup>1</sup>, Michael Mann<sup>1</sup>  
<sup>1</sup>University of Tennessee, Knoxville, <sup>2</sup>Thermo-Fisher Scientific

**Abstract:** This poster explores a new context and design to learn to visualize data, the social media-based #TidyTuesday weekly challenge. We use a novel data source for understanding learning —tweets from participating individuals over more than one year—and a combination of qualitative and computational research methods. The content individuals shared explained, qualified, or highlighted the substance of the visualizations they created, and participation over time (and longer code) was related to being recognized more by others.



Combine  
familiarity  
with  
excitement

# Same tools applied in novel ways

# Star Wars characters

# Park safety

# Online news headlines

# New York Restaurants

# Horror movies

# Number of birds colliding against airplanes

# USA Births by year

# Dog names

# NFL attendance

# Anime

# Cat and dog popularity

# Prisoner population

# Military brain injuries

# TV show ratings

# Federal R&D budgets

# Thanksgiving food popularity

# Voter turnout

# So what is TidyTuesday?

It's just data.

- That can be read into R in a few seconds
- With an article for context
- With a data dictionary
- With dozens of example analyses **WITH** code every week
- With a community of learners and mentors

**It's #TidyTuesday y'all and I hope you join us!**