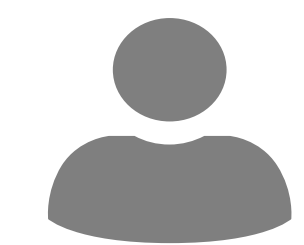


Becoming “TidyR” over Time:
Data Visualization Development
in an Online Community of
Practice



PRESENTER:
Anthony Schmidt
Evaluation, Statistics, and Measurement
aschmi11@utk.edu / anthonyschmidt.co



AUTHORS
Schmidt, Anthony (ESM)
Rosenberg, Joshua (TPTE)

Communities of Practice

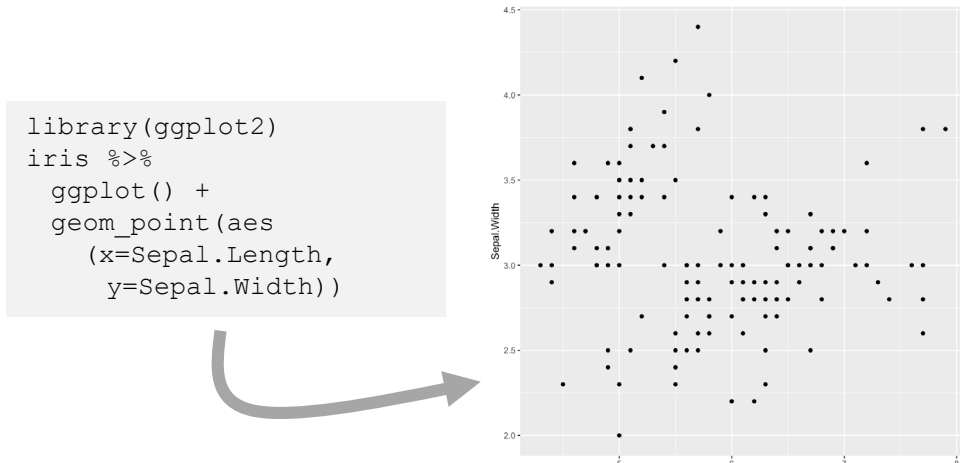
- Groups of people who share a common interest and participate in activities that promote domain-specific learning.

#TidyTuesday

- The #TidyTuesday hashtag is a community of practice for those interested in data visualization and data science.
- Every Tuesday, a unique dataset is posted on Twitter using the hashtag #TidyTuesday.
- Participants use the dataset to create visualizations using the R programing language
- Participants then share their visualizations on Twitter and link to their code (on Github).

R

- R is a statistical programming language that is used for data analysis, modeling, machine learning, and visualization.



Research Purpose

- Understand how #TidyTuesday serves as an online community of practice.
- Determine if participation in #TidyTuesday can foster data visualization skill development.

Data / Methods

- Data: One year of #TidyTuesday tweets from tidyuesday.rock.s app
- Descriptive analytics (frequencies, means, correlations)
- Qualitative thematic coding
- Content analysis of code contributions (using TidyCode)

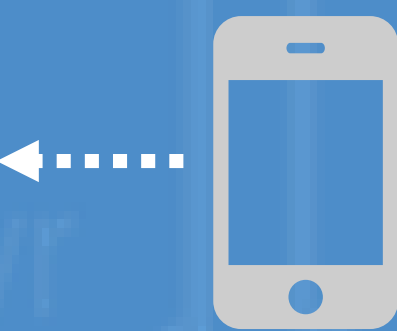
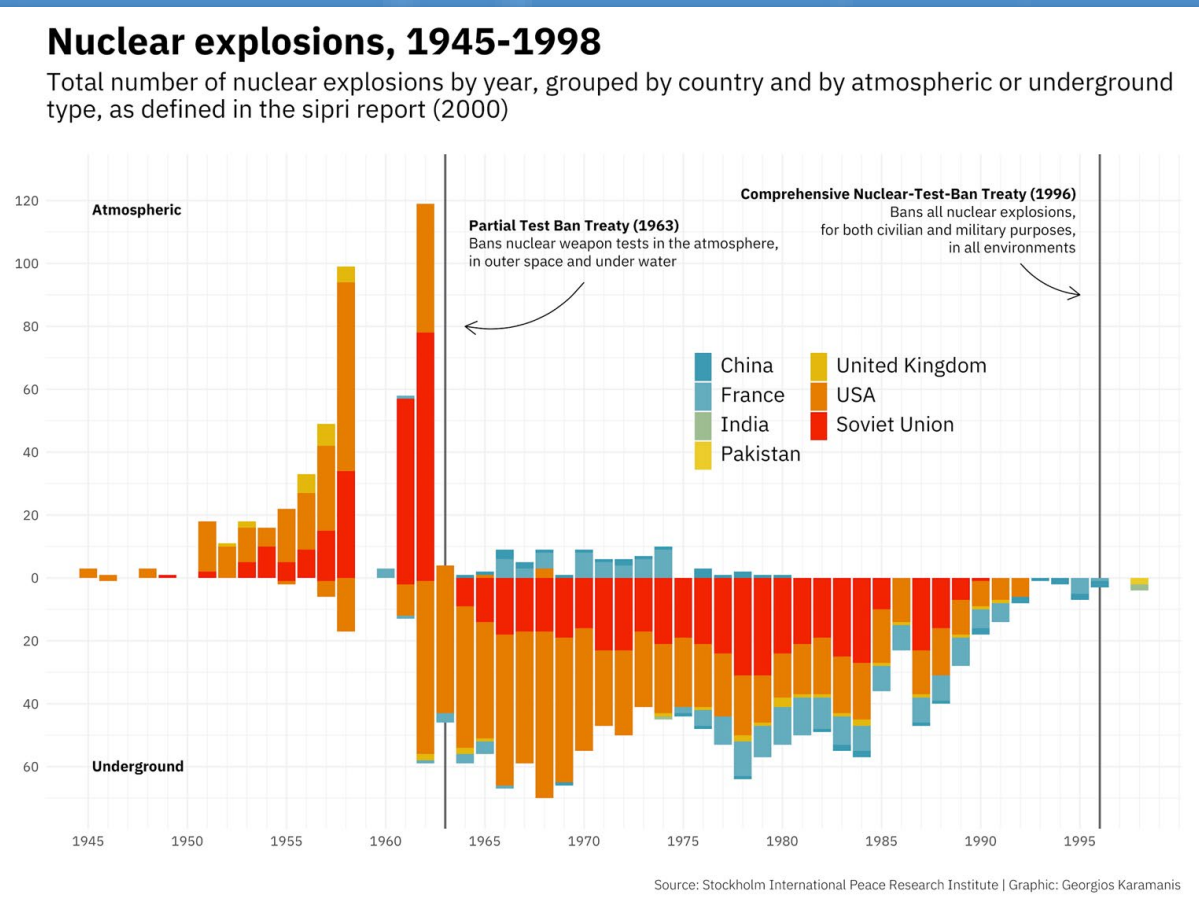
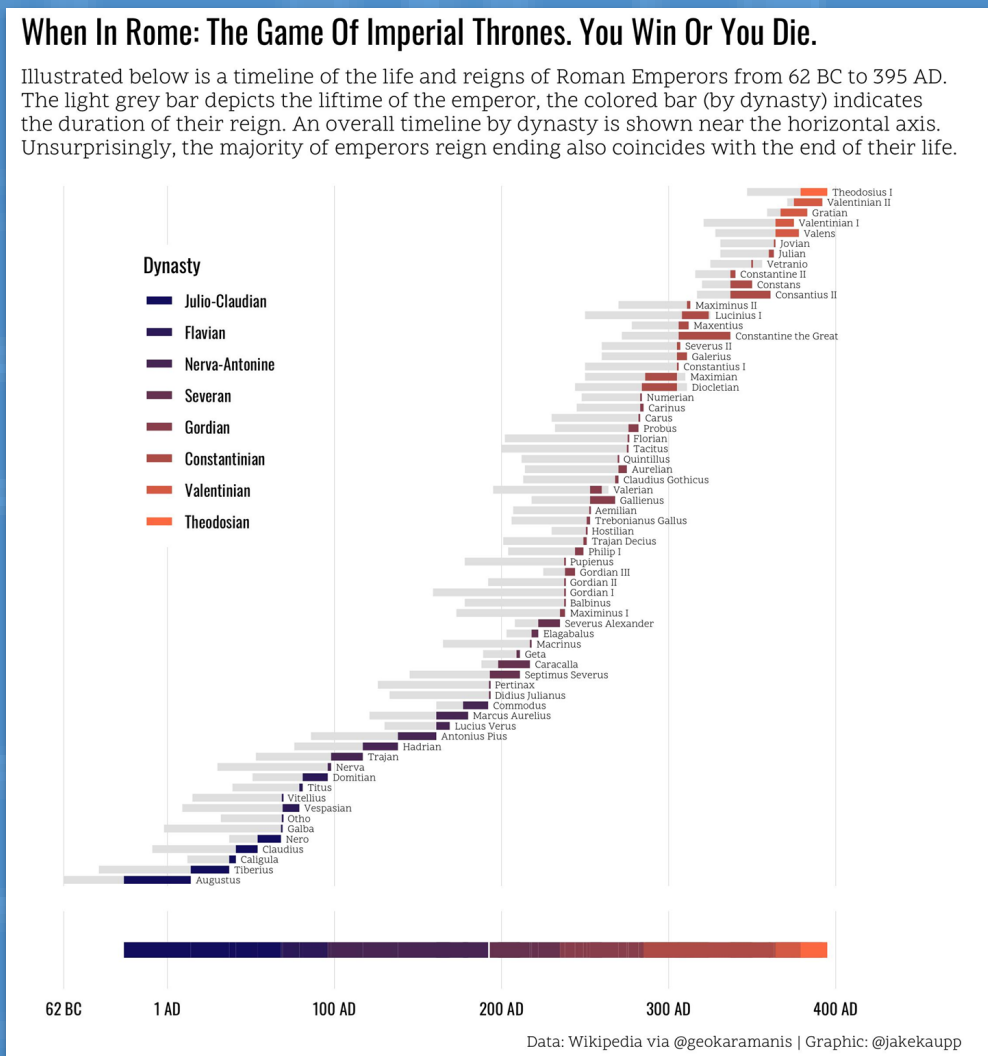
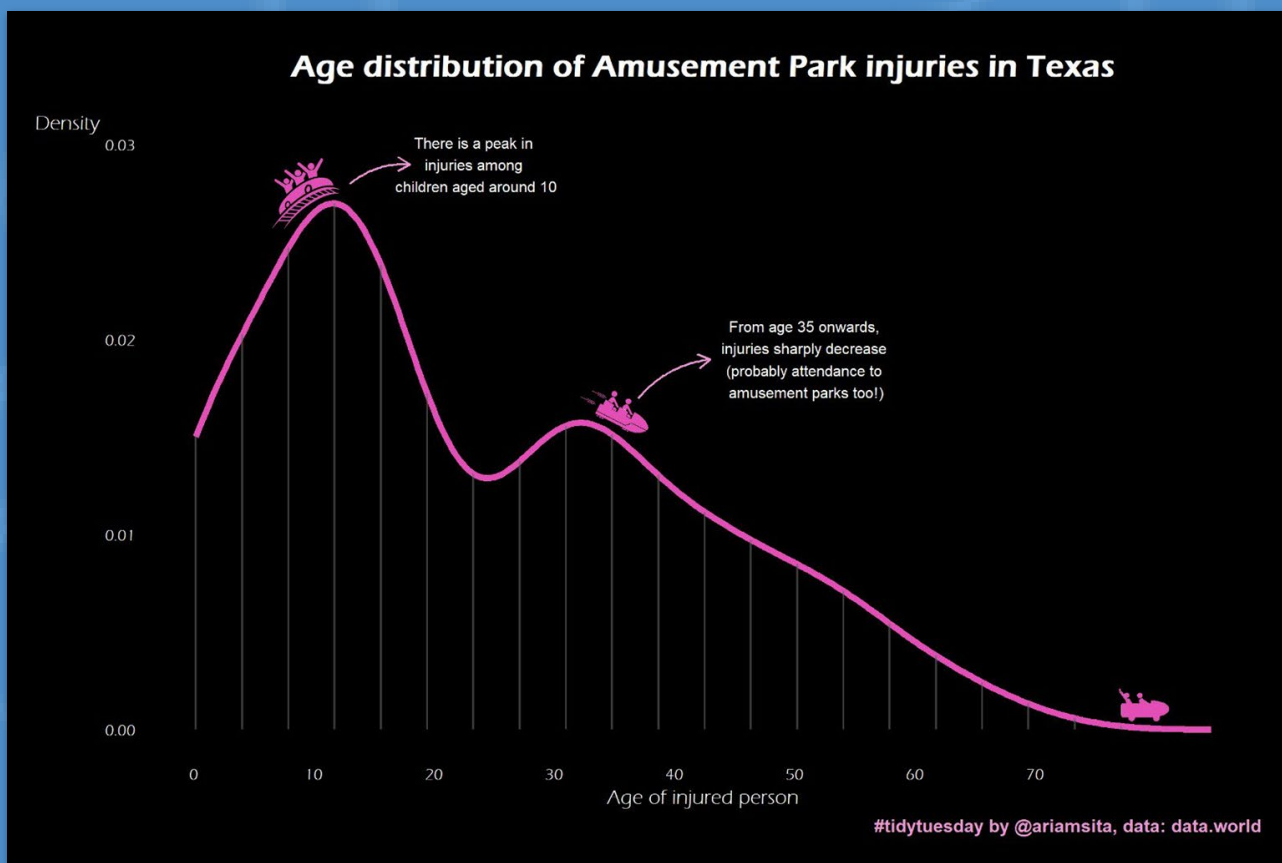
Descriptives

Findings

- 4,418** Total number of Tweets
- 2,428** Unique contributions
- 800** Unique contributors
- 46.7** Mean contributions per week
- 5.93** Mean contributions per user
- .95** Correlation between likes and retweets
- Code becomes **longer**, using **more functions**, over time.
- Visualization functions remain constant.
- Data wrangling functions **decrease**.
- Communication functions have **increased**.
- Initial tweets **explain** visualizations, describe findings, or engage the community.
- Retweets **celebrate** others' work, discuss learning, or share others' visualizations and resources.
- Conclusion:** Participation can help develop and improve visualization skills.

Data visualization
skills can be developed
by participating in the
#TidyTuesday
Twitter-based
community.

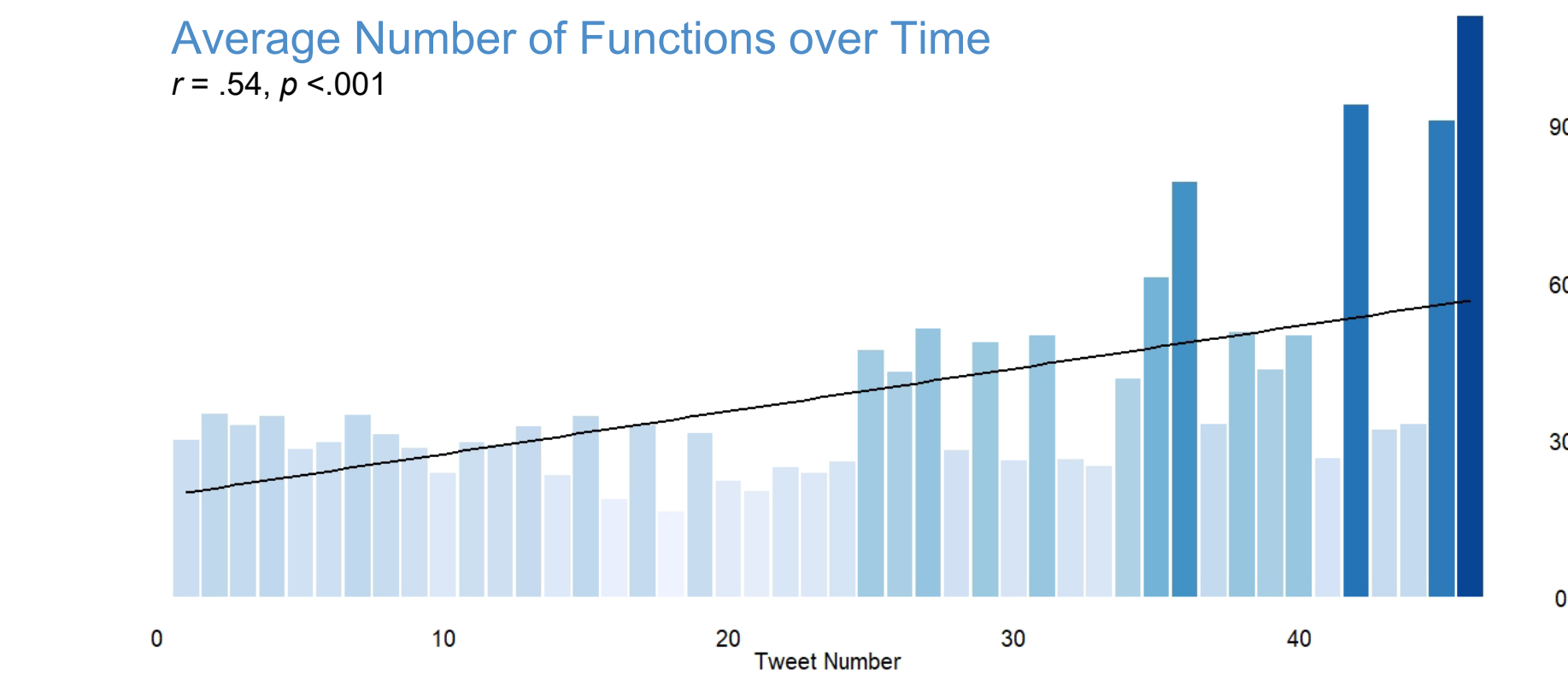
#TidyTuesday Examples



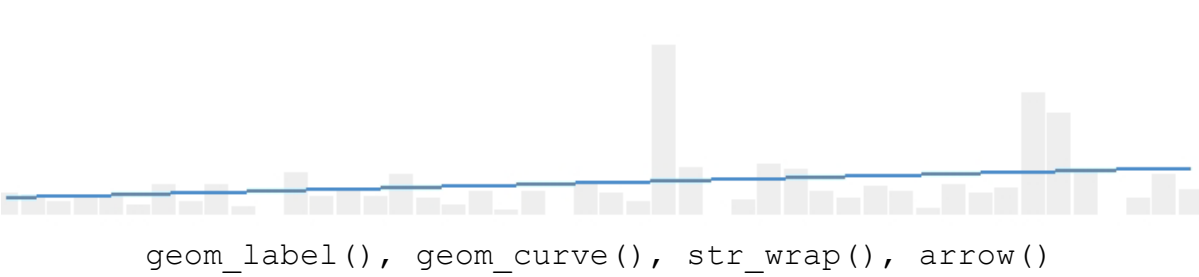
Take a picture to
download the poster and short paper

Content Analysis

Over time, contributors write **longer code** and **use more functions**.
This suggests their code may become more sophisticated over time.



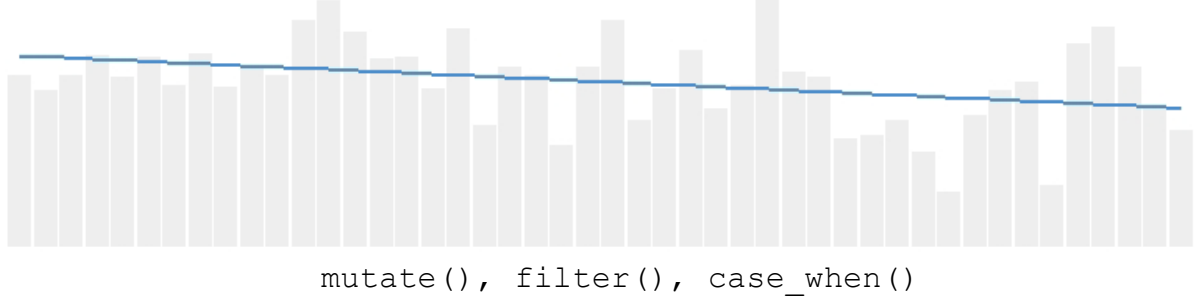
Communication



Functions related to **communication** are **increasing**.

Annotations and a focus on textual elements increasingly become important.

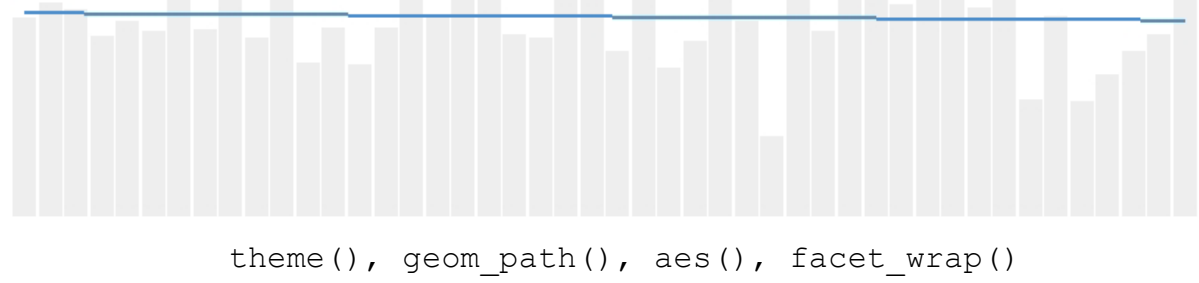
Data Cleaning/Wrangling



Data cleaning, **wrangling**, and munging functions **decrease**.

This suggest less processing of data or using more efficient code.

Data Visualization

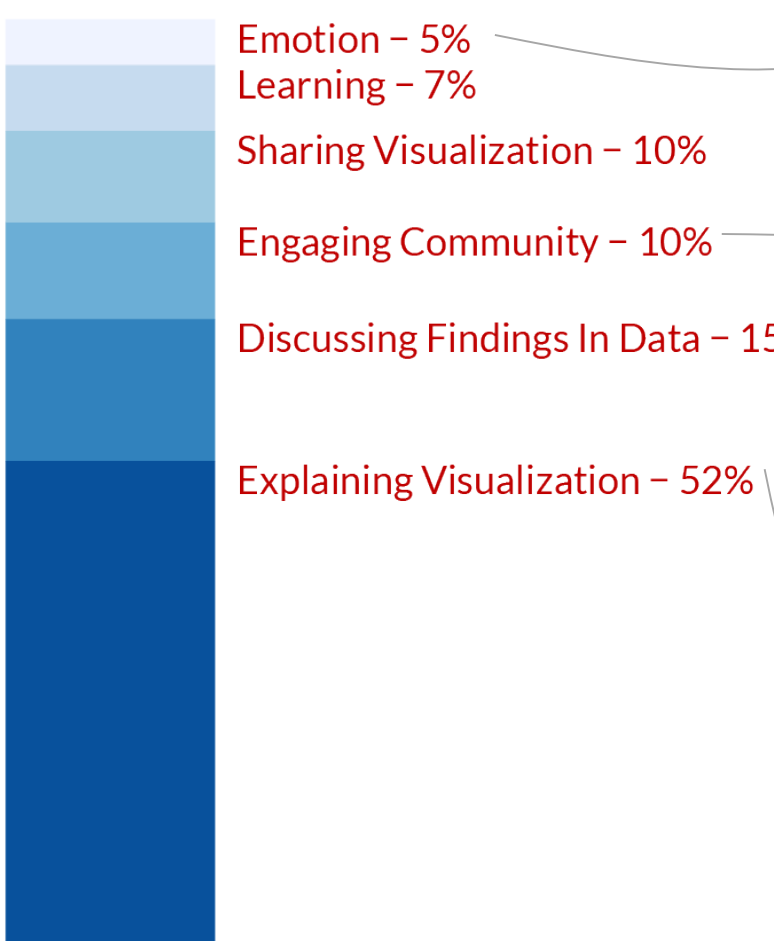


Functions related to **visualization** remain mostly **constant**.

This suggests visualizations remain the primary focus of coding.

Qualitative Analyses

Major Themes in Tweets



First ever #TidyTuesday submission! I've always been reluctant to share, because I see so many great submissions. But today I've done some new cool things (first time I made a map in R!!) and I'm quite happy with the result.

My #tidytuesday viz this week is inspired by awesome people who make scarves from data. Can't knit, but I can geom_linerange

#tidytuesday I used to think all nobel peace price winners are grey old guys. Turns out, they are getting younger compared to the other folks.

#TidyTuesday turned into tidy Saturday this week! Looking at plastic pollution around the globe. I made a waffle chart of the five countries with most mismanaged plastic waste.

Major Themes in Retweets

This thread highlights one of my favorite parts of #TidyTuesday, spontaneous and friendly collaboration!!

Sara continues to do awesome deep dives on the #tidytuesday datasets! Lots of great content here and on her GitHub.

Thanks to #TidyTuesday for making it so easy to learn/use #R and to get to know others! #grateful

Jake's dataviz work as part of #TidyTuesday is always beautiful. This one is especially great. Check out the legend he made using ggplot.

