

Storytelling with R— featuring several digressions

Mara Averick ([@dataandme](#))

Tidyverse Developer Advocate, RStudio





Mara Averick

TIDYVERSE DEV ADVOCATE, RSTUDIO

Not a Real Data Scientist™





Mara Averick

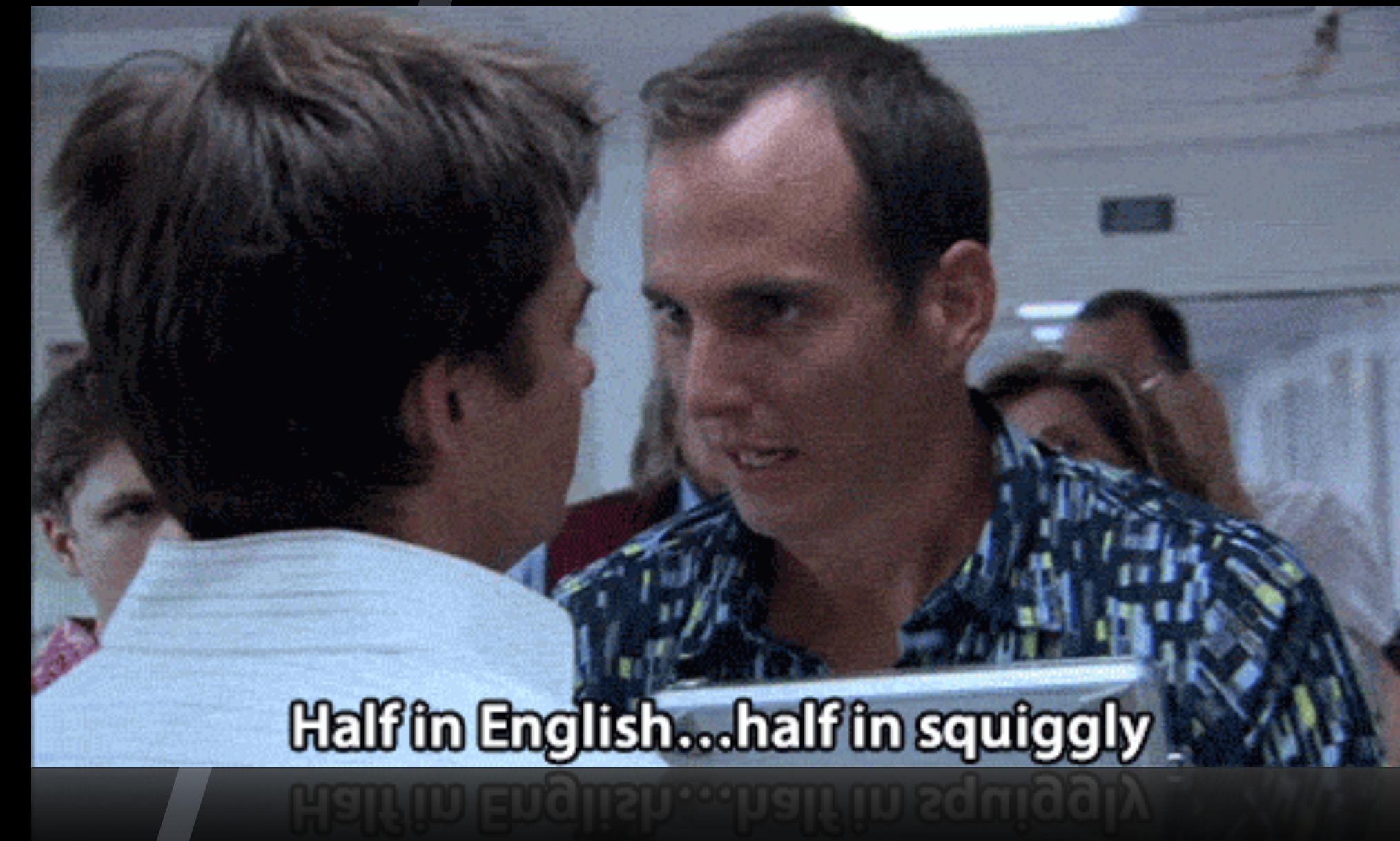
TIDYVERSE DEV ADVOCATE, RSTUDIO



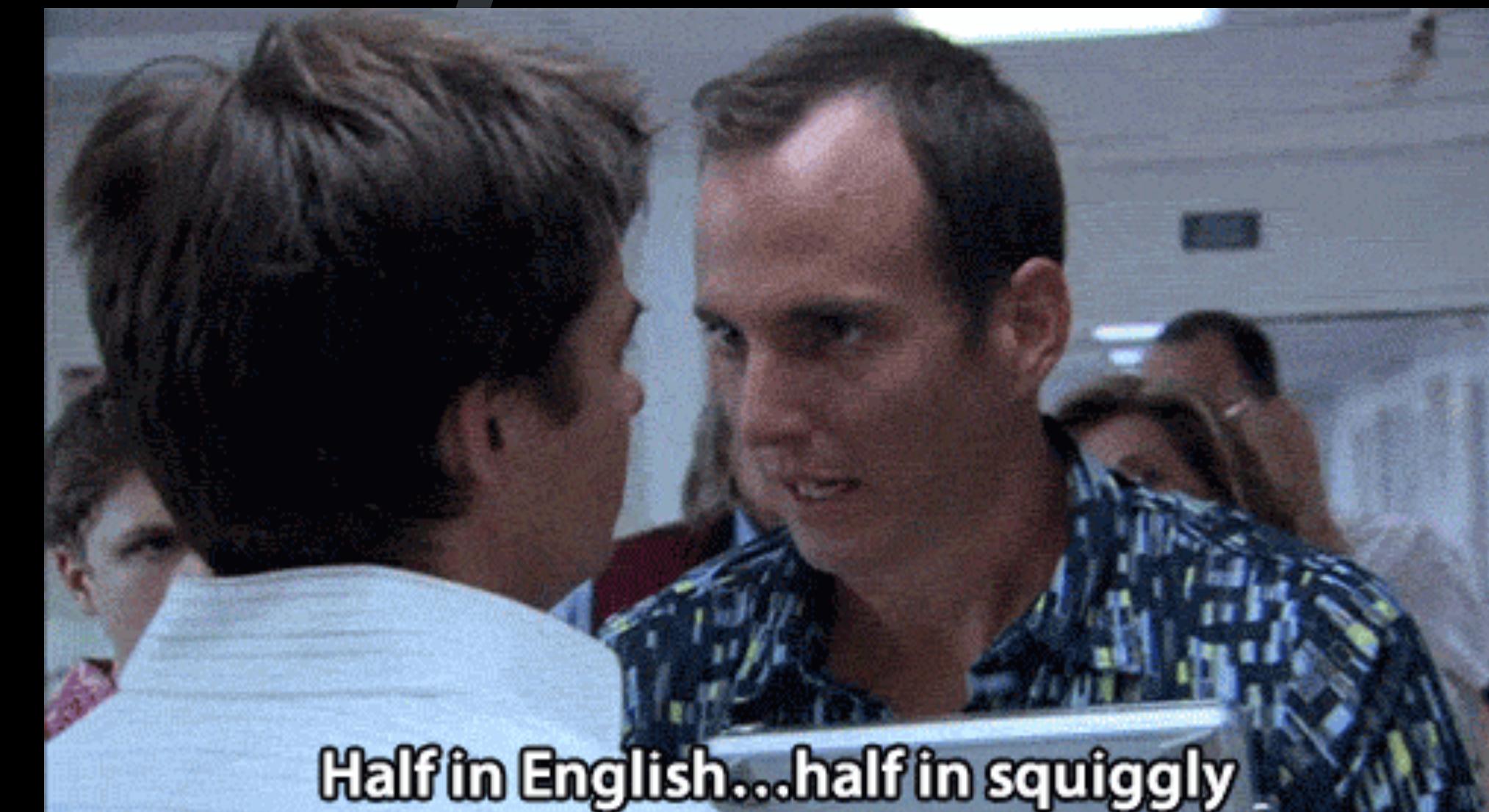
ex•o•ter•ic

adj. *understandable by outsiders*

or the general public



Communicating across the English-squiggly divide



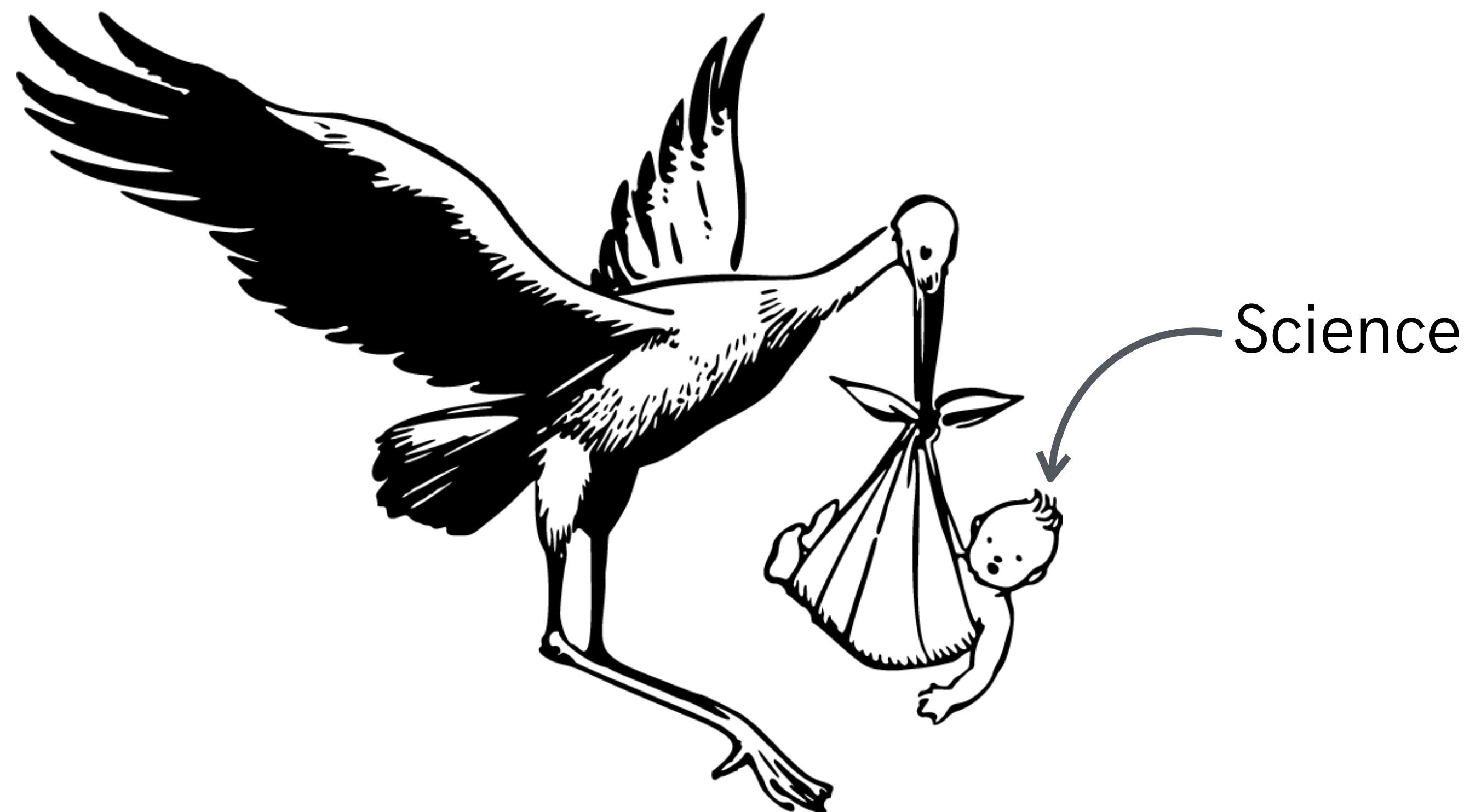
Half in English...half in squiggly

data science-ing

```
library(tidyverse)  
  
data <- read_csv(data.csv)  
  
model <- fancy_algo(data)  
  
model
```



Mom, where do data come from?



**“Is not science a
dispassionate recording of
facts, uncontaminated by
value judgements?”**

(Ross 1962)

**science
in context**



CONTEXT



What does
doing science
look like?



What does
doing science
look like?

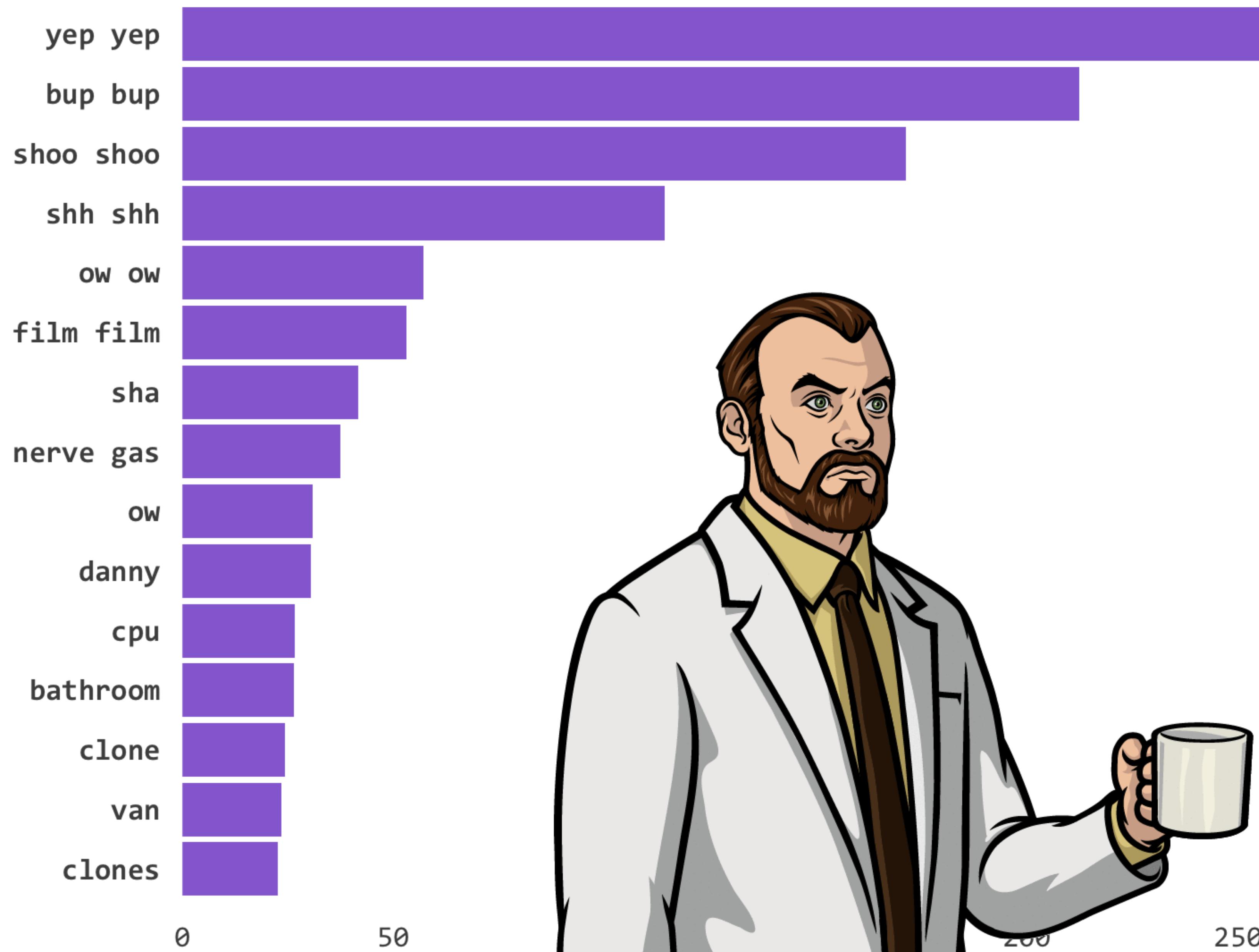


OH, THAT'S WHAT YOU mean BY context

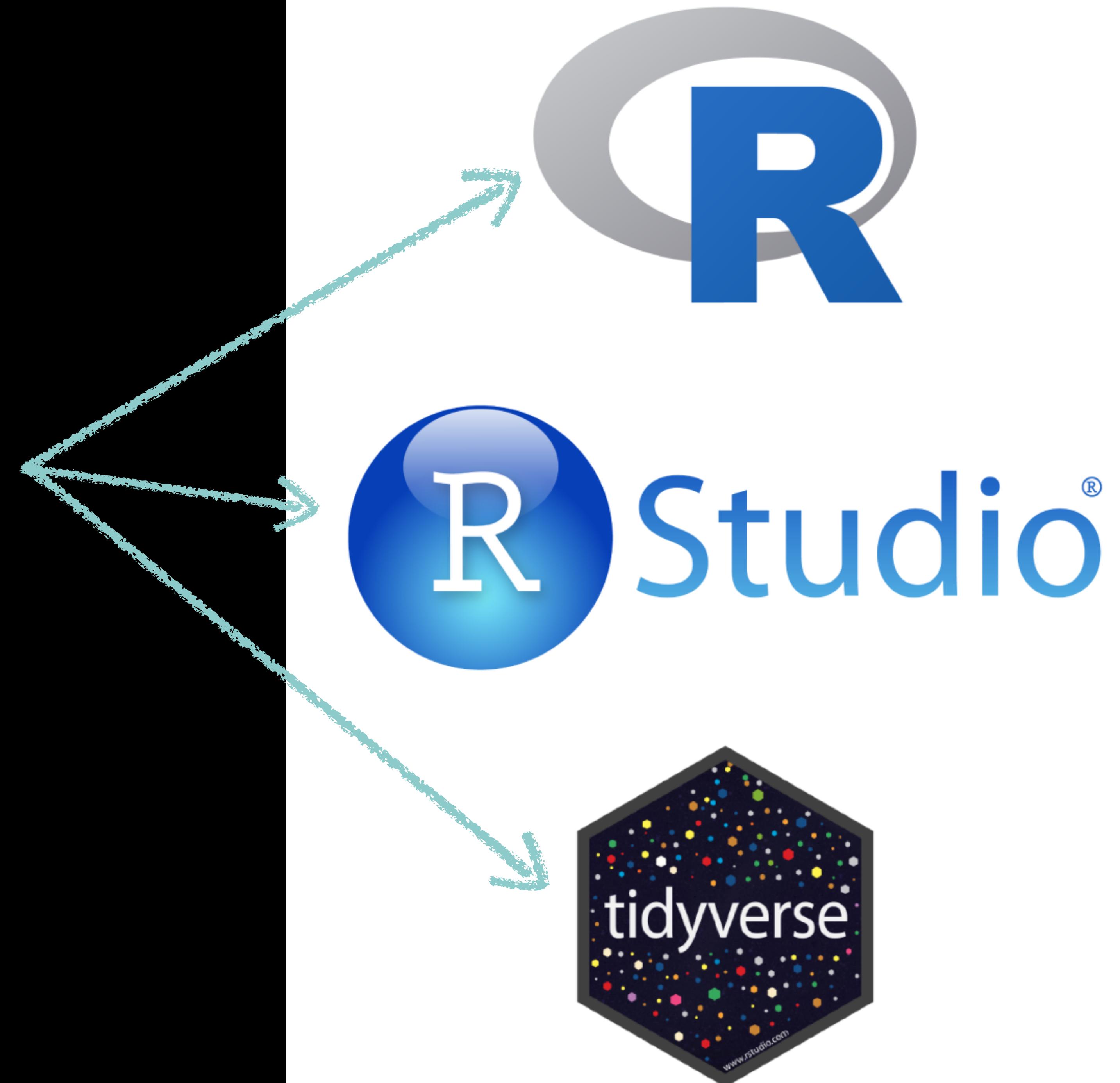
What About R?

Oh, I'm super into it!

KRIEGER'S MOST DISTINCTIVE TERMS



my favourite
tools



What is the tidyverse?

The tidyverse is an opinionated collection of R packages designed for data science.

All packages share an underlying design philosophy, grammar, and data structures.

TIDY TOOLS

Functions
should be...

SIMPLE

Do one thing and do it well.

COMPOSABLE

Combine with other functions for multi-step operations.

DESIGNED FOR HUMANS

Use evocative verb names, making them easy to remember.

Import

readr
readxl
haven
xml2

Tidy → Transform

tibble
tidyverse
dplyr
forcats
hms
lubridate
stringr

purrr
magrittr

Program

Visualise

ggplot2

recipes
rsample
tidybayesian
yardstick

Model

broom
modelr

shiny
rmarkdown

Communicate

Import



Tidy →

Store data
consistently

Transform

Create new variables & new summaries



Visualise

Surprises, but doesn't scale



Model

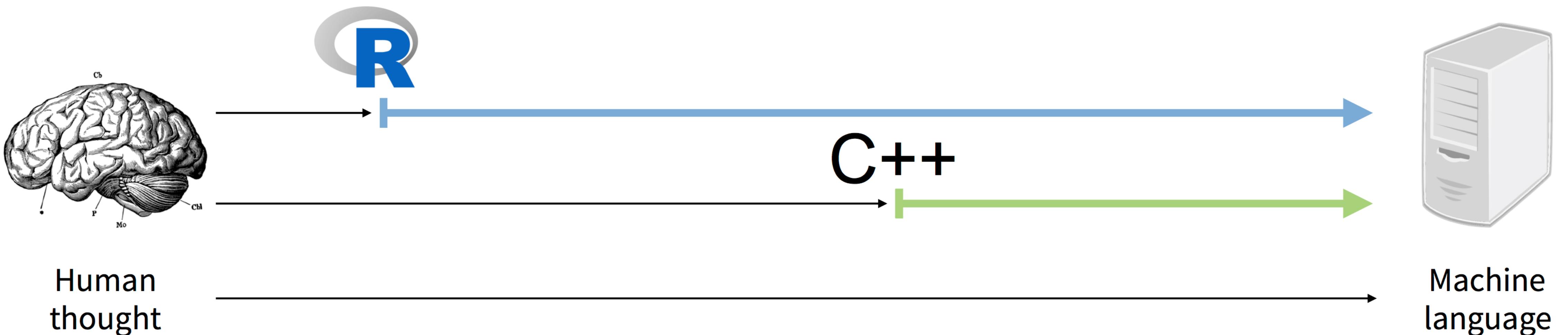
Scales, but doesn't (fundamentally) surprise

Communicate

Automate

Understand

A computer language for scientists



data wrangling



Rectangling

data wrangling



Idea

Rectangling

~~data wrangling~~



We gotta find a way to make this fit into the hole for this



with visualization we want to be sure that we communicate the right thing, and that we don't inadvertently communicate the wrong thing. In order to do so, we need to take into account the audience(s) and contexts in which these visualizations are used — we also need to account for the application/use of the visualization, esp. when being used as a "decision-support" mechanism. for example, what our adverse outcome is might inform the way in which we depict uncertainty. we have to ask ourselves what information is being removed in our rendering — are those conversations being eliminated? information fidelity is crucial, and/but we use visualization to help maximize our ability to communicate quite a lot of information. the user/viewer inherently brings his or her own experience to bear on the interpretation of the knowledge being conveyed (if it is in fact knowledge at that point in time). so, what is to be done? do we attempt to measure efficacy, and, if so, how? visualization, in fact, can also be a medium through which we consider new information; we're able to visualize information about our data before we actually know its contents sometimes this is a matter of good, efficient procedures. so, what tools do we even trust to help us w/ such an endeavor. how do we bring more of the tools in a way that helps us augment our strengths as humans?

comment your life

what do you do when you have an idea but don't, at that moment, know how to "make" that idea happen?

a thing I actually have done/will likely continue to do in order to avoid things taking forever in life always

and also b/c it allows a different kind of thinking

note: below I started to write as though someone else would actually read that (note the spacing)

at times we don't calibrate correctly for the audience, ourselves included

or this are "off" I for in later found by their

author (me) could not be properly parsed/interpreted. However, in trying to get back to what it was that I meant to convey, I am able to make use of clues/signals which, if asked to do so, I would likely be unable to "formalize" or make explicit. even when I do "get" it, I, as reader, bring things to the table that weren't

with visualization we want to be sure that we communicate the right thing, and that we don't inadvertently communicate the wrong thing. In order to do so, we need to take into account the audience(s) and contexts in which these visualizations are used — we also need to account for the application/use of the visualization, esp. when being used as a "decision-support" mechanism. for example, what our adverse outcome is might inform the way in which we depict uncertainty. we have to ask ourselves what information is being removed in our rendering — are those conversations being eliminated?

information fidelity is crucial, and/but we use visualization to help maximize our ability to communicate quite a lot of information. the user/viewer inherently brings his or her own experience to bear on the interpretation of the knowledge being conveyed (if it is in fact knowledge at that point in time). so, what is to be done? do we attempt to measure efficacy, and, if so, how? visualization, in fact, can also be a medium through which we consider new information; we're able to visualize information about our data before we actually know its contents sometimes this is a matter of good, efficient procedures. so, what tools do we even trust to help us w/ such an endeavor. how do we bring more of the tools in a way that helps us augment our strengths as humans?

comment your life

what do you do when you have an idea but don't, at that moment, know how to "make"

you consider your audience. if it's you, you capture the idea in a way that's different from how you would do so for a second-party or an audience

↗ a thing I actually have done/will likely continue to do in order to avoid things taking forever in life always

and also b/c it allows a different kind of thinking

note: below I started to write as though someone else would actually read that (note the spacing)

↙ at times we don't calibrate correctly for the audience, ourselves included

sometimes, your mental model/calculations for this are "off." I, for one have left myself a bunch of notes which, when later found by their author (me) could not be properly parsed/interpreted. However, in trying to get back to what it was that I meant to convey, I am able to make use of clues/signals which, if asked to do so, I would likely be unable to "formalize" or make explicit. even when I do "get" it, I, as reader, bring things to the table that weren't there at the time of their writing. to find the common ground

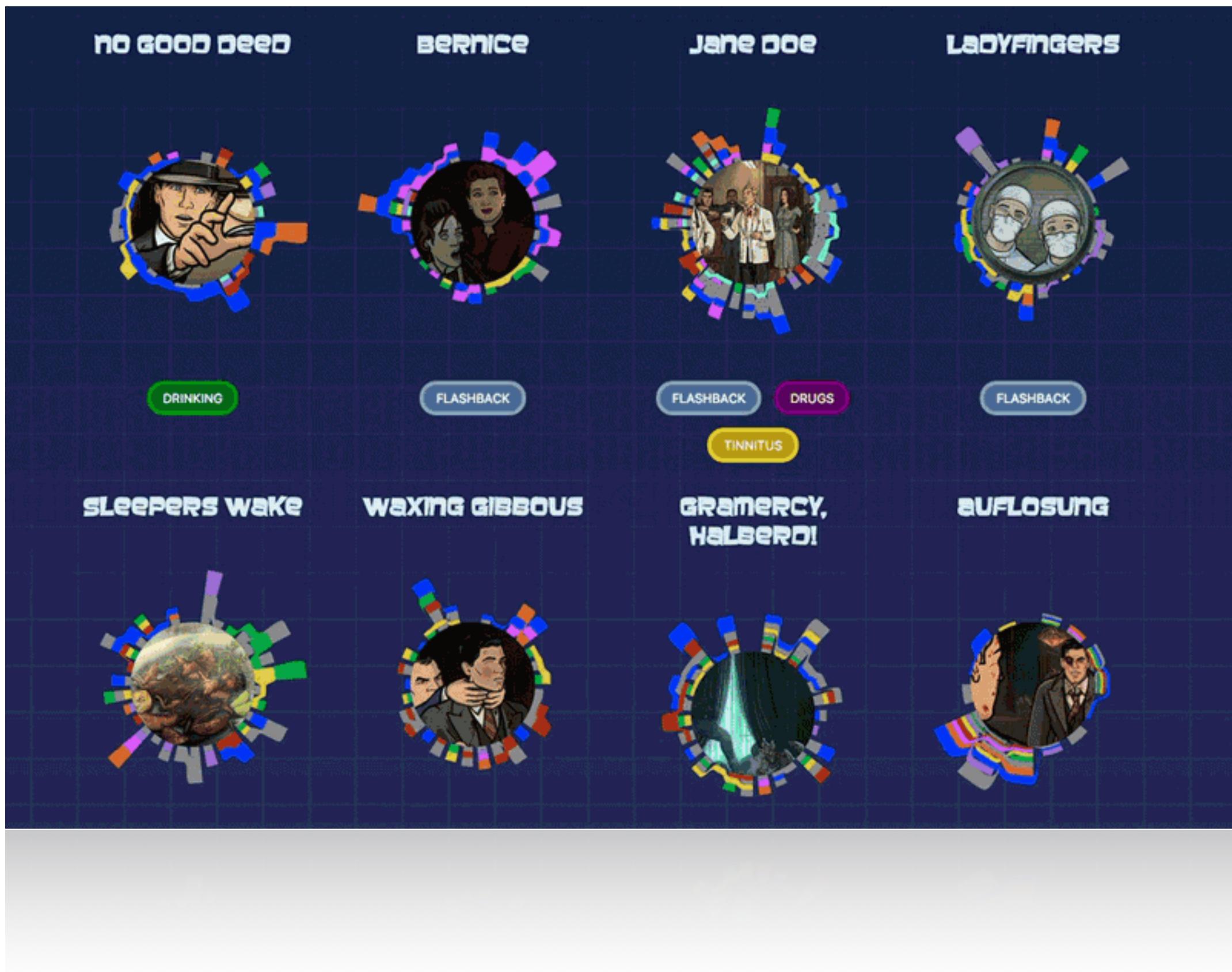
Upping your storytelling game...

But, first, a tragic tale...

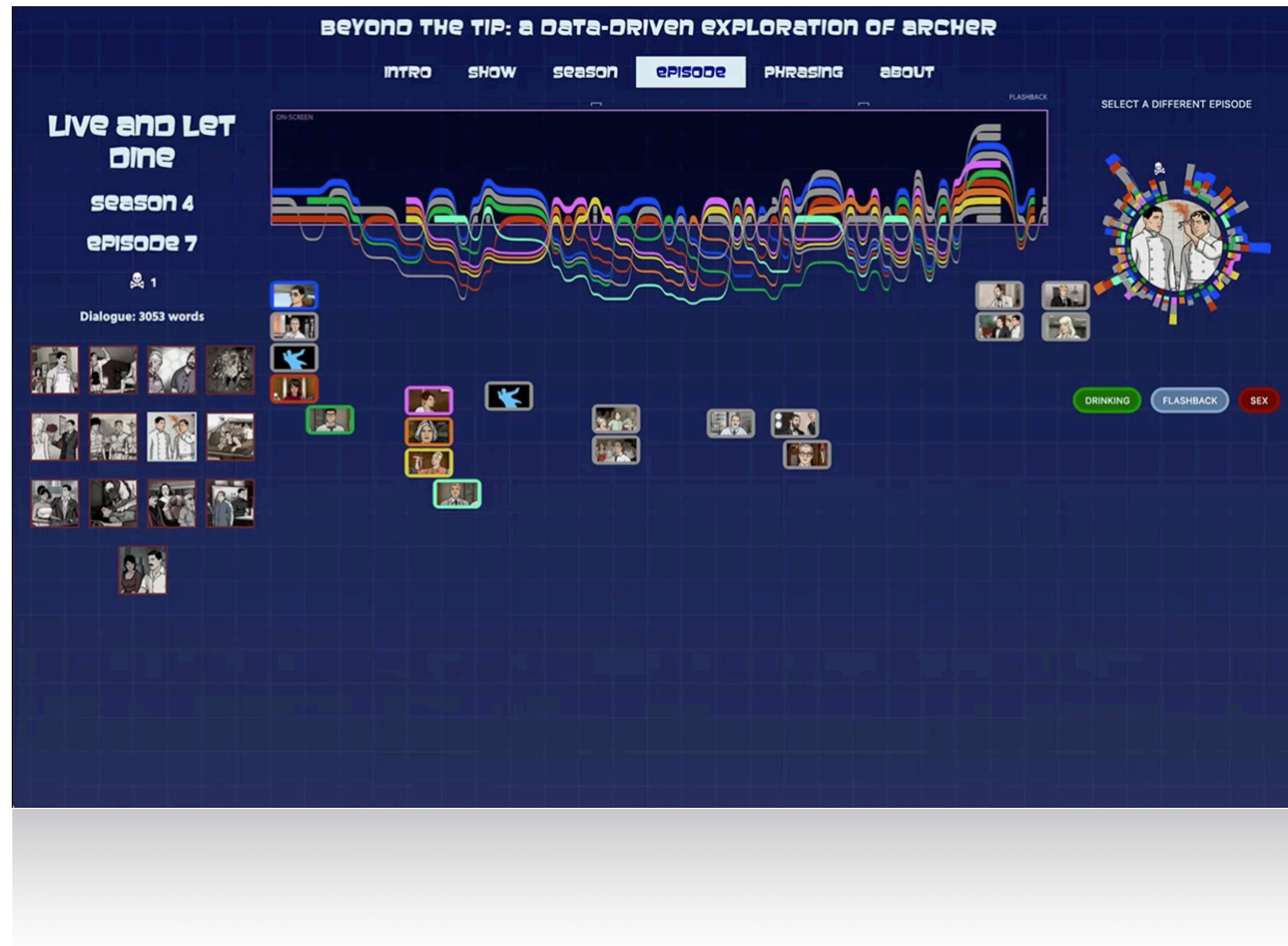


<https://archervisualization.herokuapp.com>

But, first, a tragic tale...



But, first, a tragic tale...

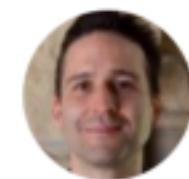


<https://archervisualization.herokuapp.com>

Making-of deets by Elijah Meeks

Towards
Data Science

Whenever you tell someone you
painstakingly annotated something by hand
they grimace and get uncomfortable like
you told them you enjoy thrash metal.



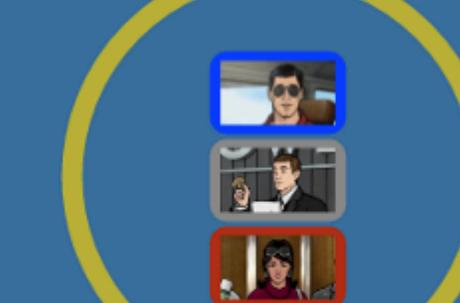
Visualizing Archer
Elijah Meeks

M

Half-drawn First/Final

First and final appearances on-screen of characters begin after or end before the drawing of a continuing character ribbon.

ON-SCREEN



Scene Tag

The active scene tag will be highlighted for any scenes with that tag.

Scene Length

Scenes length is based on the percentage of dialogue of the entire episode represented in that scene. This means that long action-heavy but dialogue light scenes will appear shorter than scenes with a lot of talking.

Scene Minimum

All scenes have a minimum length, even scenes with no dialogue at all, so that you can see them.

Entry GIF

A GIF for each character, bordered by the ribbon color of that character, is shown in a position corresponding to the first appearance of that character. GIFs are stacked and offset dynamically and the gif area is calculated and adjusted based on the width of the overall visualization.



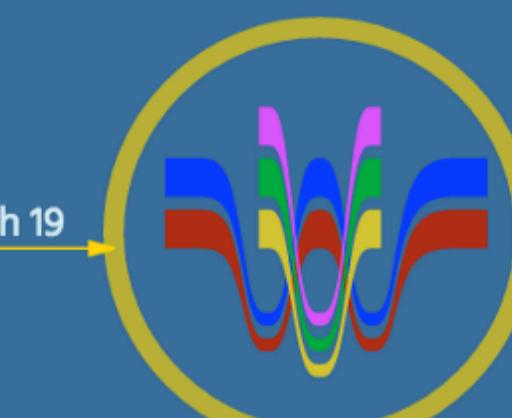
Character Order

Characters are ordered alphabetically rather than by scene participation to reduce the amount of visual shifting. Since the layout is sorting from the baseline, on-screen characters are sorted in reverse-alphabetical order from the baseline (so that Archer tends to appear on top) and in alphabetical order below the baseline (so that character order is maintained when groups of characters shift off-screen).

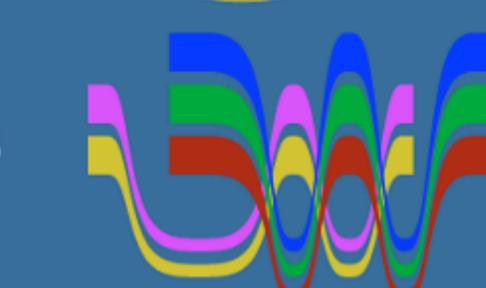
Beats

Simple beats are calculated during the layout process. These are defined as repetitions of groups of characters appearing on-screen concurrently. This emphasizes a pattern in how Archer presents episodes, with two groups of characters appear in scenes one after another several times. The beats are shown and labeled separately below the visualization.

Scenes 15 through 19



Scenes 30 through 35



Scene Spokes

Each scene is a spoke as thick as the combination of the dialogue contribution of each participant of the scene. Spokes start and end at 12 o'clock, progressing sequentially clockwise.

Kill Count

Total count of how many people die during this scene.

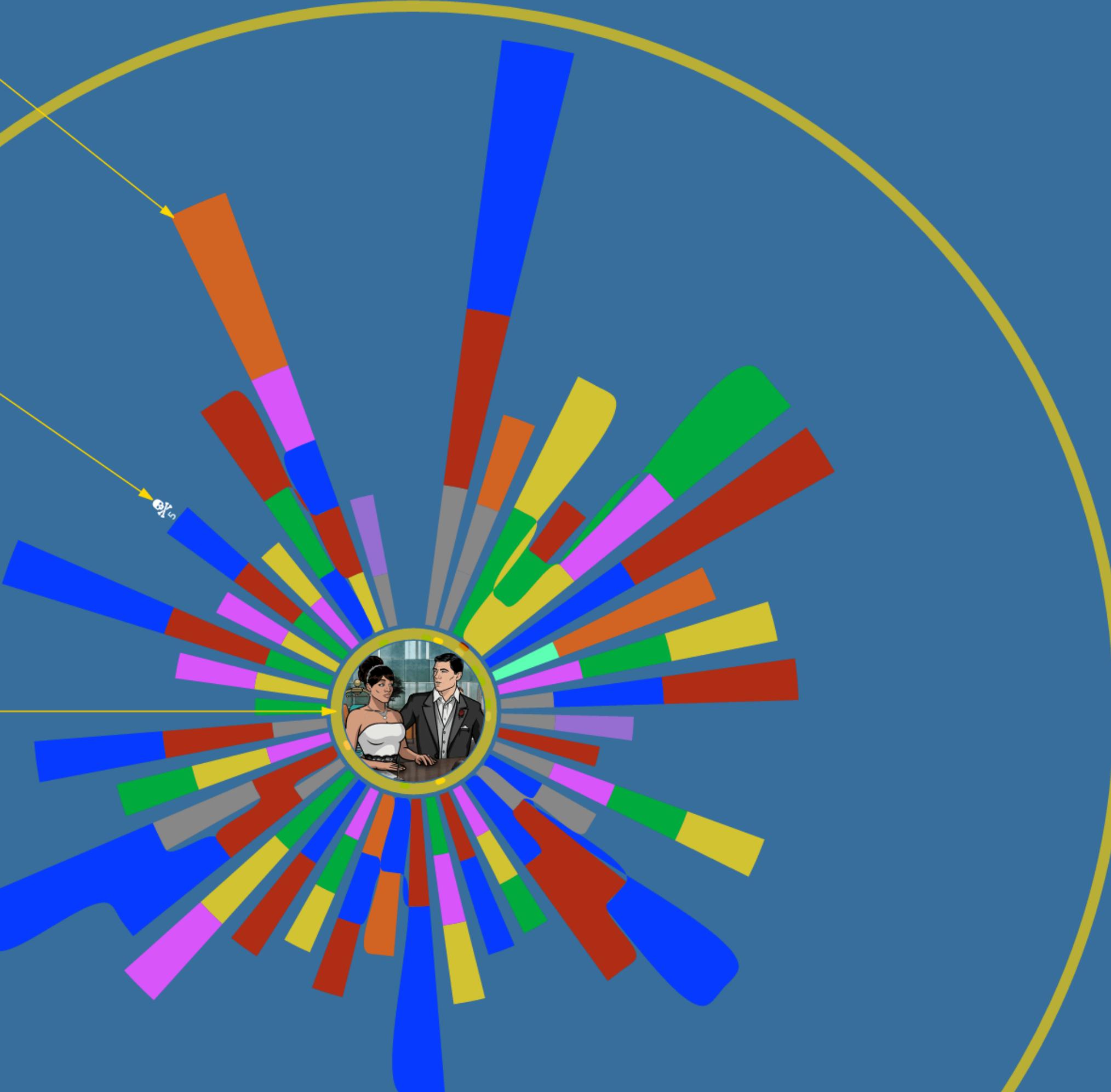
Layout

An interrupted radial bump area chart with an internal radius to hold the gif and scene information. This is a combinatorial layout, meaning it's not a novel algorithm, just a combination of various other methods (radial charts, bump area charts, interrupted charts). Radials charts are glyphy and compact, but have all the problems of pie charts along with strange interpolation behavior.

Episode GIF

Each episode has an associated GIF of the most memorable moment of the episode, as imperially decided by Mara Averick.

THE HONEYMOONERS



Lessons learned?

Towards
Data Science

That said, when I look at the complexity of the shapes and the spartan explanations for how to read them, I think it could have been a bit more user friendly. I think we could all learn a few things from Algernop Krieger and focus less on whether we can make something and more on what someone might think about the thing we made.



Visualizing Archer
Elijah Meeks



Lessons learned?



Infinite exploration
puts the onus of
storytelling on the
user

Upping your storytelling game...

- ✓ Steal (borrow) from the best
- ✓ Annotate like a boss
- ✓ Create cranky code
- ✓ Get good at getting help

glue



distill

Allaire, et al. (2018, Sept. 10). Distill for R Markdown.
Retrieved from <https://rstudio.github.io/distill>

RStudio Distill for R Markdown

Authoring ▾ Websites ▾ Publishing ▾ News 🔍

Distill for R Markdown

Scientific and technical writing, native to the web

AUTHORS	AFFILIATIONS	PUBLISHED	CITATION
JJ Allaire	RStudio	Sept. 10, 2018	Allaire, et al., 2018
Rich Iannone	RStudio		
Yihui Xie	RStudio		

Distill for R Markdown is a web publishing format optimized for scientific and technical communication. Distill articles include:

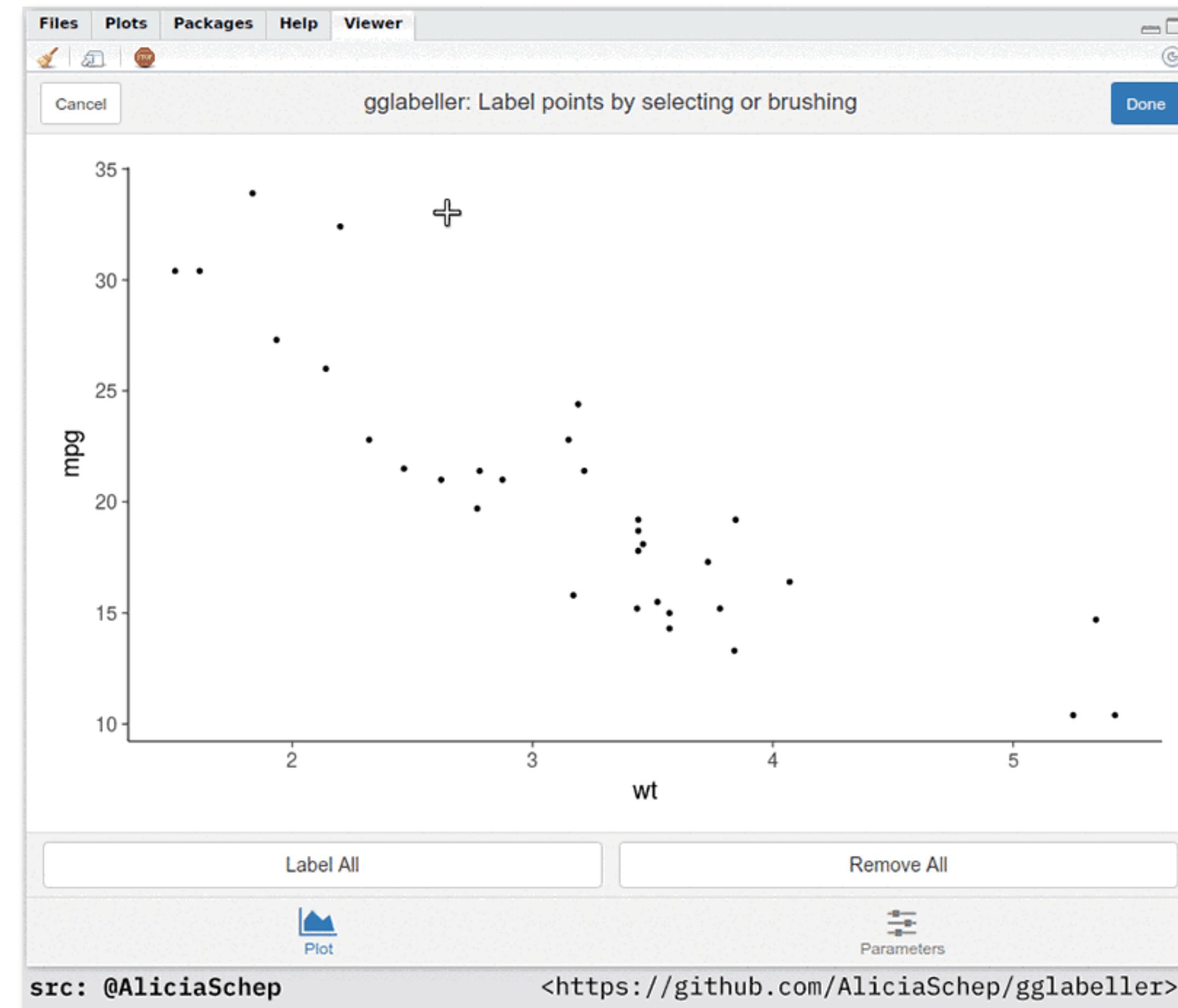
- Reader-friendly typography that adapts well to mobile devices.
- Features essential to technical writing like [LaTeX math](#), [citations](#), and [footnotes](#).
- Flexible [figure layout](#) options (e.g. displaying figures at a larger width than the article text).
- Attractively rendered [tables](#) with optional support for pagination.
- Support for a wide variety of [diagramming tools](#) for illustrating concepts.
- The ability to incorporate JavaScript and D3-based [interactive visualizations](#).
- A variety of ways to [publish articles](#), including support for publishing sets of articles as a [Distill website](#).
- The ability to [create a blog](#) composed of a collection of Distill articles.

Distill for R Markdown is based on the [Distill web framework](#), which was originally created for use in the Distill Machine Learning Journal [\[1\]](#). Distill for R Markdown combines the technical authoring features of Distill with [R Markdown](#), enabling a fully reproducible workflow based on literate programming [\[2\]](#).

Installation

To create an [R Markdown](#) document that uses the Distill format, first install the [distill](#) R package from GitHub:

gglabeller



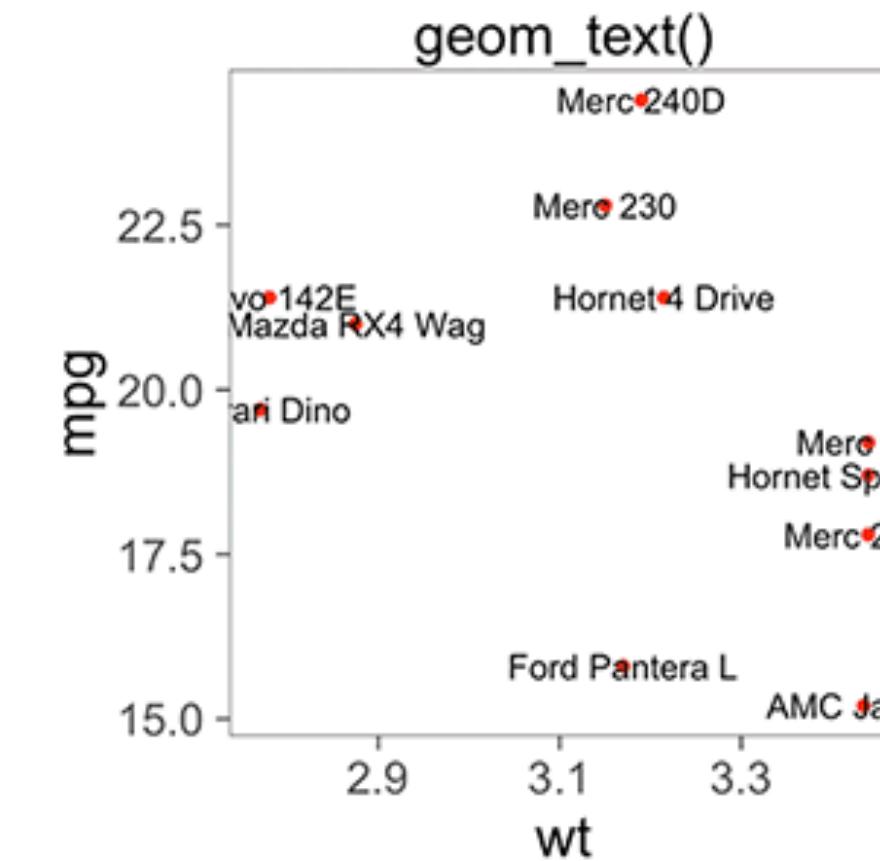
by Alicia Schep <<https://github.com/AliciaSchep/gglabeller>>

ggrepel



Problem: text placement

```
library(ggplot2)
ggplot(mtcars) +
  aes(
    x = wt, y = mpg,
    label = rownames(mtcars)
  ) +
  geom_point(color = "red") -
  geom_text()
```

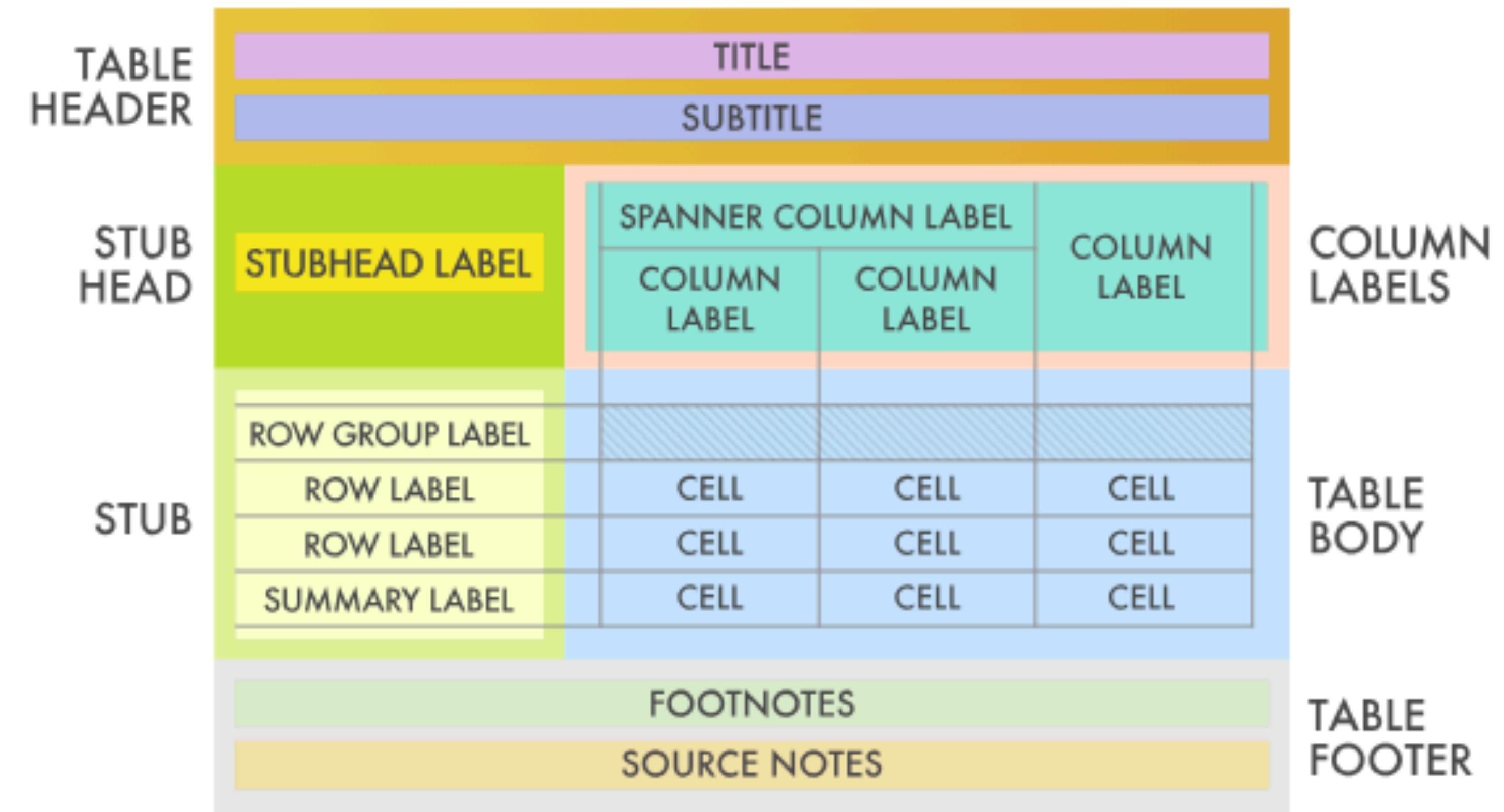


Intro to **{ggrepel}** by @slowkow

<http://slowkow.com/ggrepel>

gt

Parts of a gt Table



highcharter

HIGHCHARTER

Sections ▾ Github

WELCOME
Let's start

SHOWCASE
Some fancy tries

API
What can we do

hchart FUNCTION
The Magic

SHORTCUTS
Add data series from R objects

THEMES
Changing the look

SHINY
Output & Render functions

HIGHCHARTS
Some fun and miscellaneous

HIGHSTOCK
Using the quantmod package

HIGHMAPS
Give me the geojson data

PLUGINS
Some extensions

Docs
Functions and data

HIGHCHARTER

Highcharts.com © Natural Earth

Source: SWAPI via rstars package

by Joshua Kunst<<http://jkunst.com/highcharter/>>



The screenshot shows a web browser window for the Discourse website at <https://www.discourse.org>. The header includes the Discourse logo, navigation links for About, Features, Community, Demo, and Pricing, and standard browser controls.

The main heading on the homepage is "Civilized discussion for your community". Below it is a blue button labeled "Try it FREE".

A modal window titled "Your Community" is displayed, showing a list of posts from the "Net Neutrality Day of Action" thread. The posts include:

- Net Neutrality Day of Action (4 reactions, Jul '17)
- Sharing our experience changing the Discourse UI (6 reactions, Jul '17)
- A measure of success (0 reactions, 763 views)
- Discourse app: just GREAT (0 reactions, 830 views)
- The community platform of champions (0 reactions, 890 views)
- Loving Discourse so far (0 reactions, 807 views)
- Discourse for discussions: superiority to e.g. GitHub (0 reactions, 1.2k views)
- Discourse running great on forum with a few hundred users (0 reactions, 823 views)

Below the list, there is a note: "Moving along? If you're ready to request features or start a thread, [Get started now!](#)"

To the right of the main content, a sidebar displays a list of recent posts from other threads:

- Discourse is Best Forum Solution in the 2017 CMS Critic Awards (7 reactions, Nov '17)
- Converting, combining, migrating - Yes! (4 reactions, Nov '17)
- Successful Migration (0 reactions, Oct '17)
- Thanks to discourse (6 reactions, Oct '17)
- Post your Discourse T-shirt pics! (45 reactions, Oct '17)
- What an amazing platform - (0 reactions, Oct '17)

<https://www.discourse.org/>

RStudio Community

https://community.rstudio.com

Sign Up Log In

all categories ► all tags ► Latest Categories Top

Topic	Category	Users	Replies	Views	Activity
<input checked="" type="checkbox"/> Welcome to the RStudio Community! 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	meta		0	2.8k	Jul 22
<input type="checkbox"/> Integrating shiny apps shiny	shiny		1	34	16m
<input type="checkbox"/> Create pdf from Rnw: missing just one font character from eastern European language	R Markdown		4	39	30m
<input type="checkbox"/> UNC pathway when downloading packages ide ide-issue	RStudio IDE		0	6	39m
<input type="checkbox"/> Are packrat bundles really portable across different platforms?	R Admins		12	592	1h
<input type="checkbox"/> /usr/bin/env no such file or directory terminal rstudioserver	R Admins		1	10	1h
<input type="checkbox"/> Deploying APP on shinyapps.io which keeps asking for ggplot2 ggplot2 shinyappsi0 package-installation	shiny		13	92	1h
<input type="checkbox"/> when I installed R Studio, the packages were visible, but not seeing packages not seeing anymore, how to get back packages in rstudio? ide ide-issue	RStudio IDE		0	5	2h
<input type="checkbox"/> How to read Netcdf files	General		8	58	4h

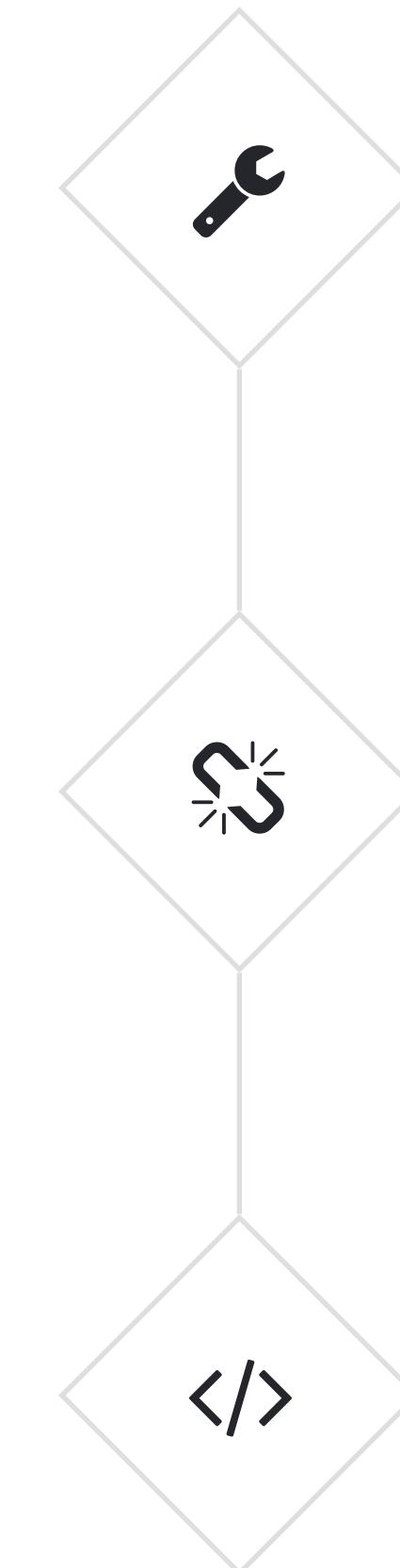
The art of the question

The most useless problem statement that one can face is "it doesn't work", yet we seem to get it far too often.

- Thiago Maciera



the anatomy of an issue



PROBLEM DESCRIPTION

EXPECTED BEHAVIOUR

MINIMAL REPRODUCIBLE EXAMPLE

the anatomy of an issue

"Creating an issue template for your repository" – GitHub Help <<https://help.github.com/articles/creating-an-issue-template-for-your-repository/>>

Error Information: Packages missing

Description of issue - [REDACTED]

Steps ta'

System

- R
- R
- C
- R Version:

Steps taken so far - i tried and tried..

Also:

- RStudio diagnostics report:
- Your sessionInfo():
- RStudio crash report:
- RStudio application log files:

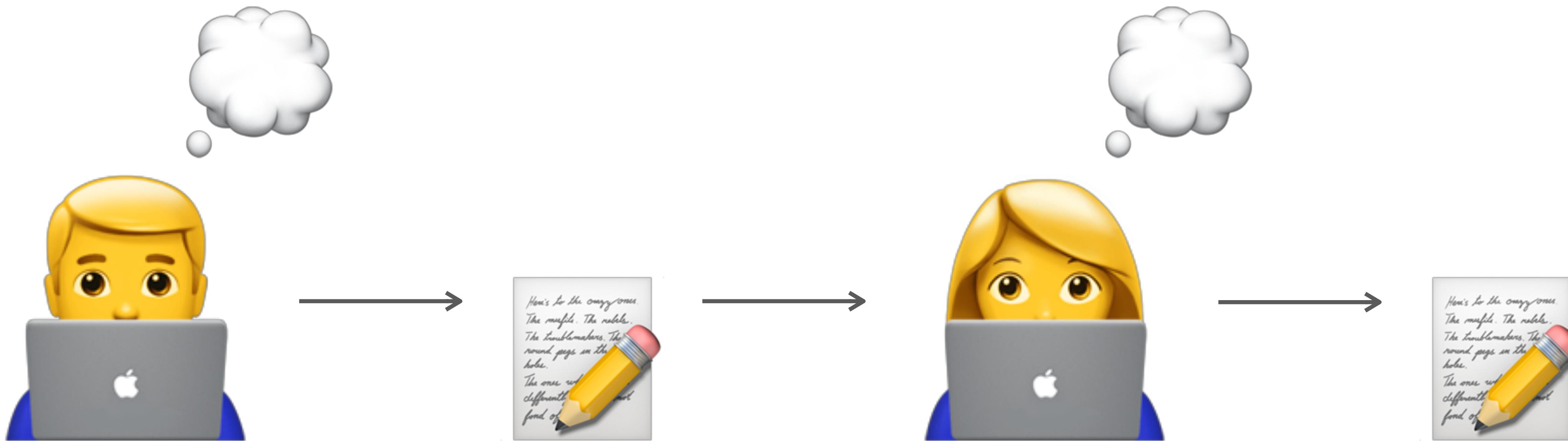
From [Troubleshooting Guide: Using RStudio](#)

They're not wrong

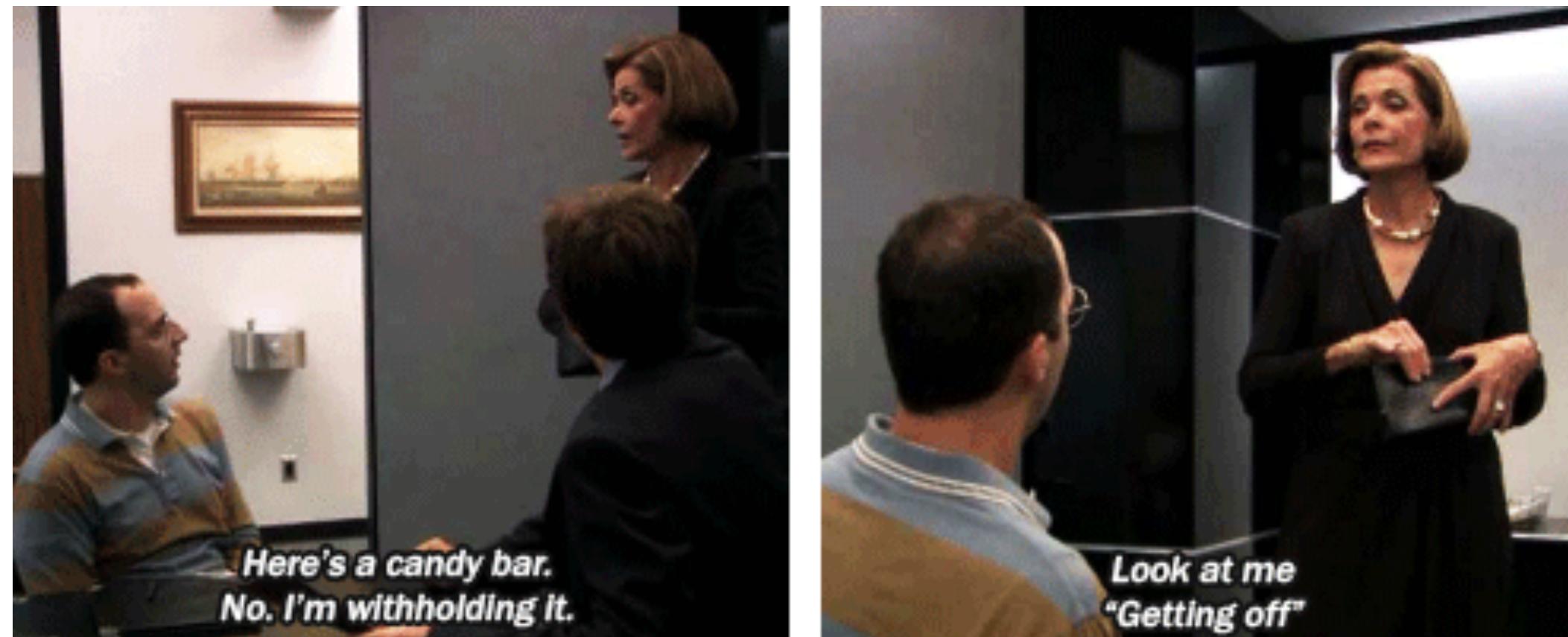
“It is impossible to speak in such a way that you cannot be misunderstood.”

– Karl Popper

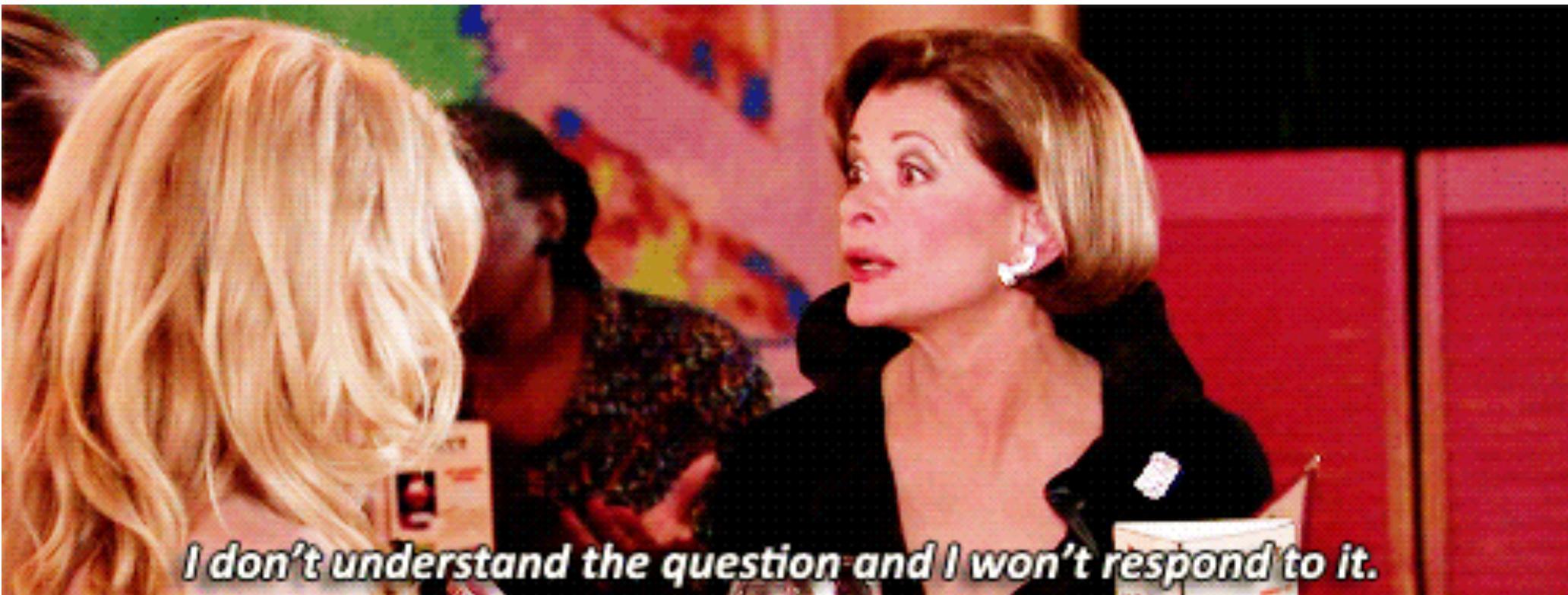
Writing prose about code...



Writing prose about code...



Writing prose about code...



I don't understand the question and I won't respond to it.

the magic of reprex



reprex raison d'être



Keys to reprex-cellence

- ✓ Code that **actually runs**
- ✓ Code that **doesn't have to be run**
- ✓ Code that **can be easily run**

The screenshot shows the RStudio interface. On the left is the 'Script Editor' window titled 'Untitled1*', containing the following R code:

```
1 library(ggplot2)
2
3 df <- data.frame(x = 1)
4
5 ggplot(df, aes(x, x)) + geom_point() +
6   ggtitle("gjpjQ") +
7   theme(plot.title = element_text(size = 10))
8
9 ggplot(df, aes(x, x)) + geom_point() +
10  ggtitle("gjpjQ") +
11  theme(plot.title = element_text(size = 100))
```

The right side of the interface is the 'Console' window, which is currently empty. Below the main windows are the 'Files', 'Plots', 'Packages', 'Help', and 'Viewer' tabs.

<https://maraaverick.rbind.io/2018/06/reprexcellence/>

Thank You

