Truth, beauty, and data revisited

Session 15

PMAP 8921: Data Visualization with R
Andrew Young School of Policy Studies
Spring 2025

Plan for today

Telling stories with data

Curiosity

Telling stories with data

Truth is beautiful

Truth ≠ facts

Truth comes from aesthetic combination of content and form

Facts require beauty to be true

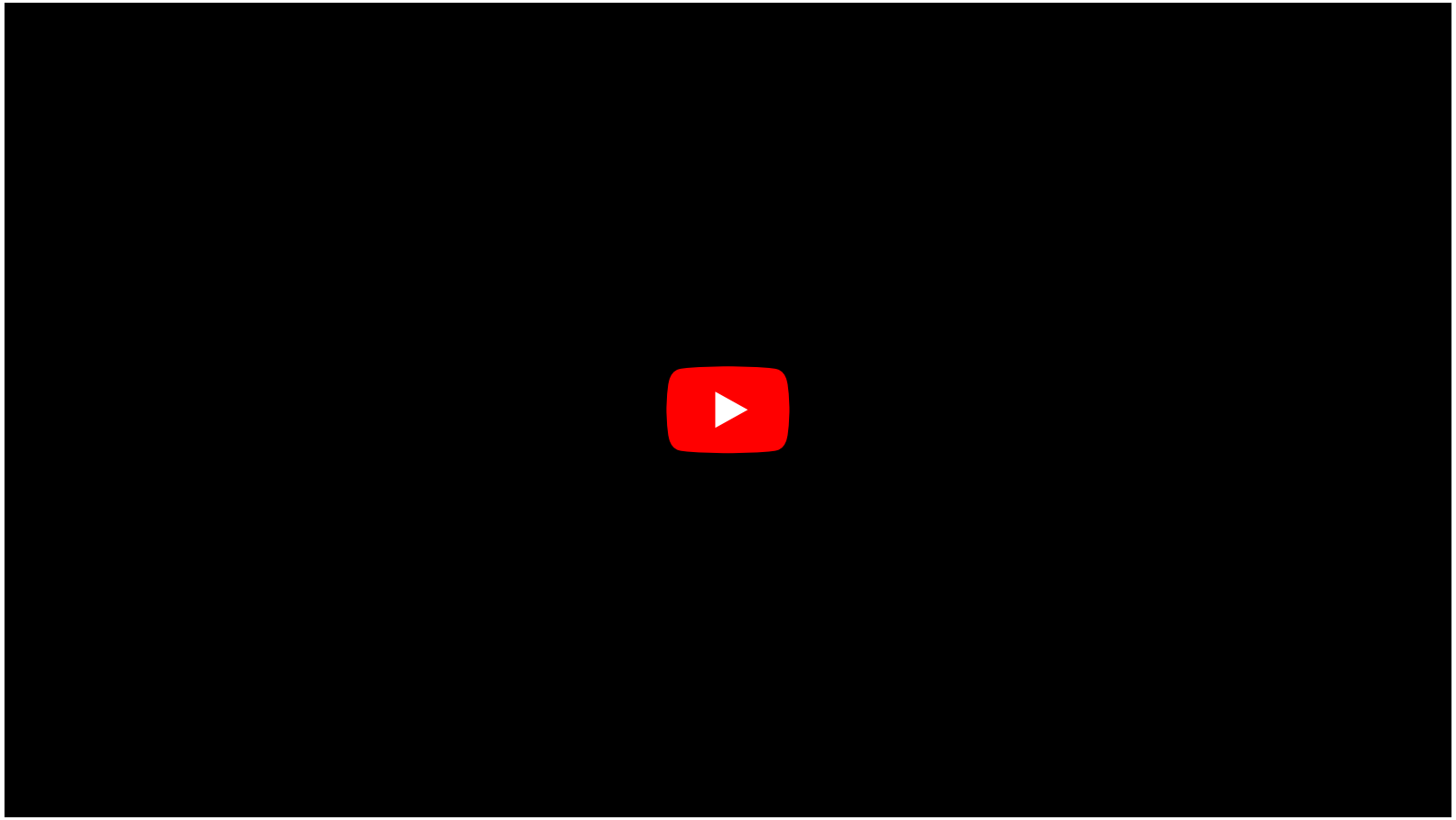
Art

Art is how we translate core, essential content (or truth!) to different forms for specific audiences.

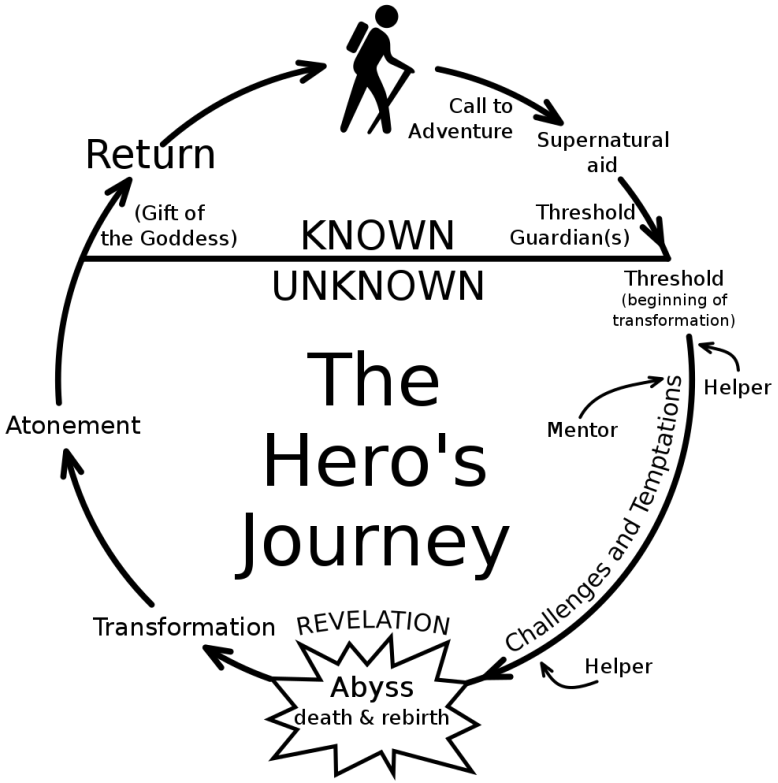
Stories as art

Stories are an *art form* for translating core, essential **content** to different **forms** for specific **audiences**.

Every story is the same



Heroes


Call to Adventure
Supernatural aid
Threshold Guardian(s)
Threshold
(beginning of transformation)
Helper
Mentor
Challenges and Temptations
Helper
REVELATION
Abyss
death & rebirth
Transformation
Atonement
Return
(Gift of the Goddess)
KNOWN
UNKNOWN
The Hero's Journey

1. YOU
2. NEED
3. GO
4. SEARCH
5. FIND
6. TAKE
7. RETURN
8. CHANGE

You are not the hero

- About us
  - Company history
  - Market cap
  - # employees and # locations
- About our product and service
  - What it is
  - How it works
  - Why it's better than the alternative
- Call to action (ideally)

XYZ Co. Equity Partners, LLC

- Founded in 1988 in Anchorage, Alaska
- Invest in companies who:
  - Provide professional IT services
  - Offer exceptional technical and project management expertise
  - Deliver complex data and information management solutions as systems and/or applications integrators
- Average annual revenue: $51.5M

XYZ Co. Software

- Established in 1984
- Headquarters: San Francisco, CA
- Integrated P&C Insurance software and services
- Focused on Alternative Risk & Self-Insured markets
- Recognized leader in risk management solutions
- Over 100 customers in U.S. and Canada

From Cole Nussbaumer Knaflic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*

Should you tell stories though?

Published: 30 July 2013

Points of view

Storytelling

Martin Krzywinski & Alberto Cairo

Nature Methods **10**, 687(2013) | Cite this article

1529 Accesses | **100** Altmetric | Metrics

Familiar elements underpin most stories: introduction, question, conflict, buildup and resolution. These can also be applied to data graphics. For example, use the idea of a story arc and make your presentation episodic—unfold it, don't dump it. In each part, make not only its content clear but its purpose easily discernible. This is particularly relevant when communicating to the general public, who may lack sufficient background knowledge to identify what is relevant or why it matters. At the same time, do not underestimate your colleagues' desire to be presented with a cogent exposition of your findings.

Published: 30 October 2013

Against storytelling of scientific results

Yarden Katz

Nature Methods **10**, 1045(2013) | Cite this article

862 Accesses | **147** Altmetric | Metrics

To the Editor:

Krzywinski and Cairo[1] beautifully illustrate the widespread view that scientific writing should follow a journalistic 'storytelling', wherein the choice of what data to plot, and how, is tailored to the message the authors want to deliver. However, they do not discuss the pitfalls of the approach, which often result in a distorted and unrepresentative display of data—one that does not do justice to experimental complexities and their myriad of interpretations.

Possible pitfalls

Manipulation

Don't lie or manipulate data

Ethos

Credentials ≠ ethos

Misinterpretation

Temper expectations

Equity

Don't dumb down

Amplify underrepresented voices

Manipulation

THIS AMERICAN LIFE
FROM WBEZ

555: The Incredible Rarity of Changing Your Mind

APR 24, 2015

It's rare for people to change what they believe, and if they do it, it's usually a long process. This week, stories of those very infrequent instances where people's opinions flip on fundamental things that they believe. Why does it happen in these particular and unusual circumstances? We explain. **NOTE: One of the authors of a study covered in this episode has asked that the study be retracted.**

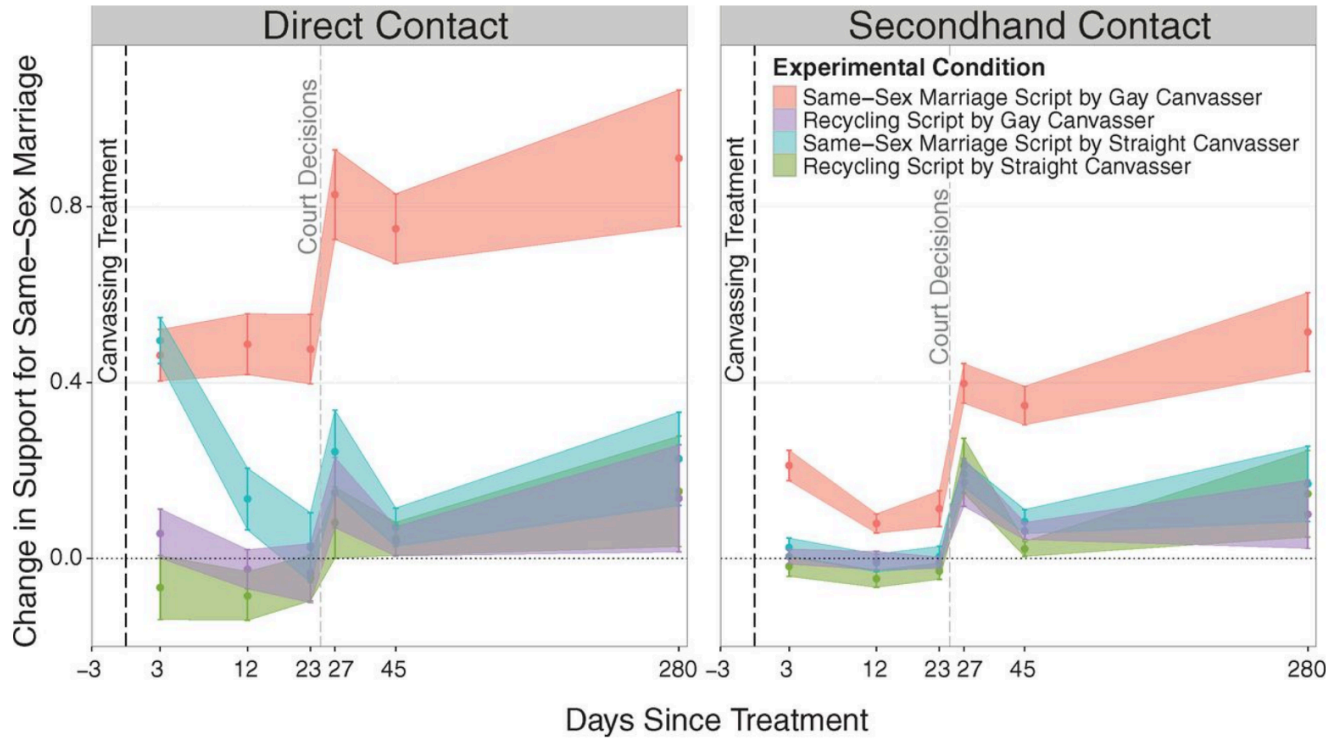
The iPad thing was LaCour's trademark. "He was sort of famous for taking his results from different studies he was working on, putting them on an iPad, and buttonholing people at the conferences and going over all of the research that he was doing, the different findings he had, and basically not letting the people go until they had an idea of what he was working on," says Tim Groeling, a communications professor at UCLA, who is listed as one of LaCour's references on his curriculum vitae. "It was infectious," continues Groeling. "Really cool stuff was on that iPad."

When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour[1], **Donald P. Green**[2]

+ See all authors and affiliations

Science 12 Dec 2014:
Vol. 346, Issue 6215, pp. 1366-1369
DOI: 10.1126/science.1256151



Irregularities in LaCour (2014)

David Broockman, Assistant Professor, Stanford GSB (as of July 1), dbroockman@stanford.edu
Joshua Kalla, Graduate Student, UC Berkeley, kalla@berkeley.edu
Peter Aronow, Assistant Professor, Yale University, peter.aronow@yale.edu

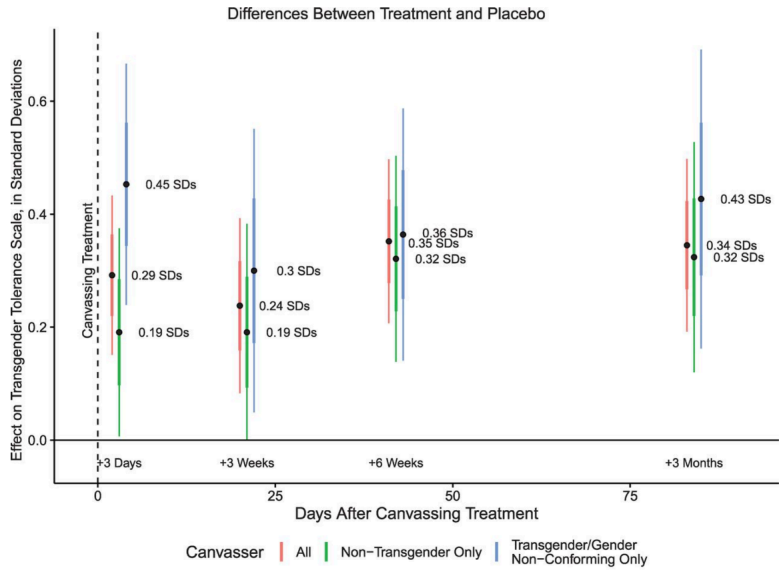
May 19, 2015

REPORT

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1,*], Joshua Kalla[2]
+ See all authors and affiliations

Science 08 Apr 2016:
Vol. 352, Issue 6282, pp. 220-224
DOI: 10.1126/science.aad9713

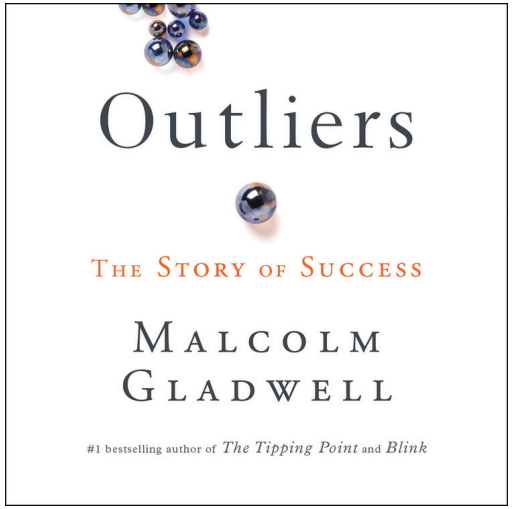


Manipulation

Don't lie

**Emphasize the story,
but make full data available**

Misrepresentation



10,000 hours

"the magic number of greatness"

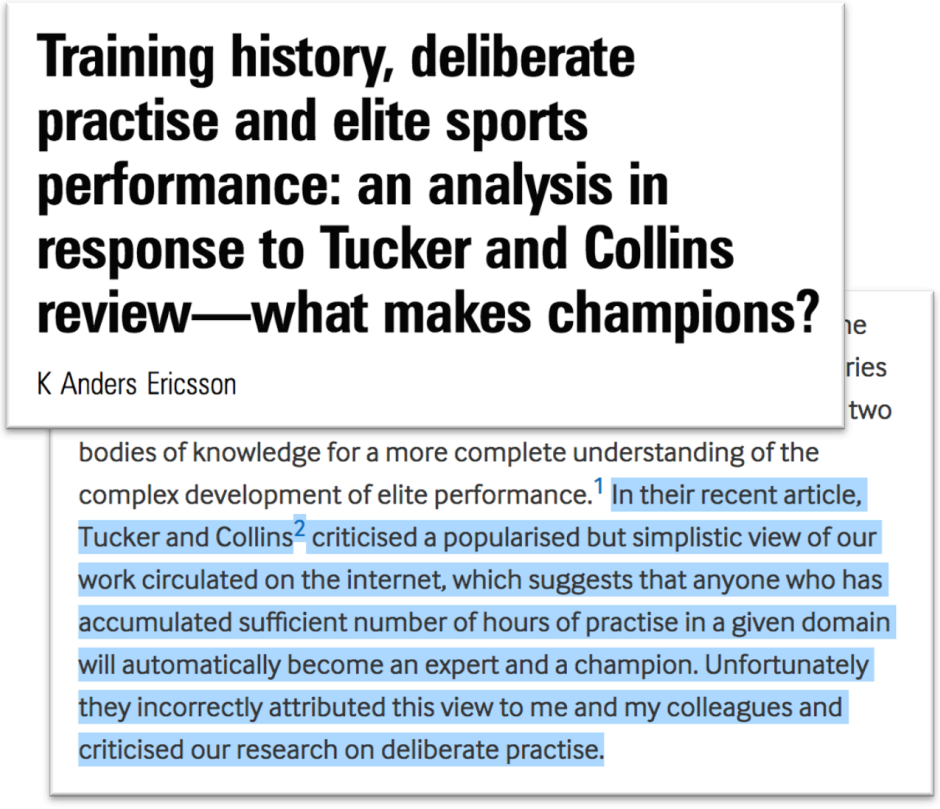
Psychological Review
1993, Vol. 100. No. 3, 363-406

Copyright 1993 by the American Psychological Association, Inc.
OO33-295X/93/S3.OO

The Role of Deliberate Practice in the Acquisition of Expert Performance

K. Anders Ericsson, Ralf Th. Krampe, and Clemens Tesch-Romer

The theoretical framework presented in this article explains expert performance as the end result of individuals' prolonged efforts to improve performance while negotiating motivational and external constraints. In most domains of expertise, individuals begin in their childhood a regimen of effortful activities (deliberate practice) designed to optimize improvement. Individual differences, even among elite performers, are closely related to assessed amounts of deliberate practice. Many characteristics once believed to reflect innate talent are actually the result of intense practice extended for a minimum of 10 years. Analysis of expert performance provides unique evidence on the potential and limits of extreme environmental adaptation and learning.



Training history, deliberate practise and elite sports performance: an analysis in response to Tucker and Collins review—what makes champions?

K Anders Ericsson

bodies of knowledge for a more complete understanding of the complex development of elite performance.[1] In their recent article, Tucker and Collins[2] criticised a popularised but simplistic view of our work circulated on the internet, which suggests that anyone who has accumulated sufficient number of hours of practise in a given domain will automatically become an expert and a champion. Unfortunately they incorrectly attributed this view to me and my colleagues and criticised our research on deliberate practise.

"[A] popularized but simplistic view of our work, which suggests that anyone who has accumulated sufficient number of hours of practice in a given domain will automatically become an expert and a champion."

10,000 is average • Quality matters • There are other factors

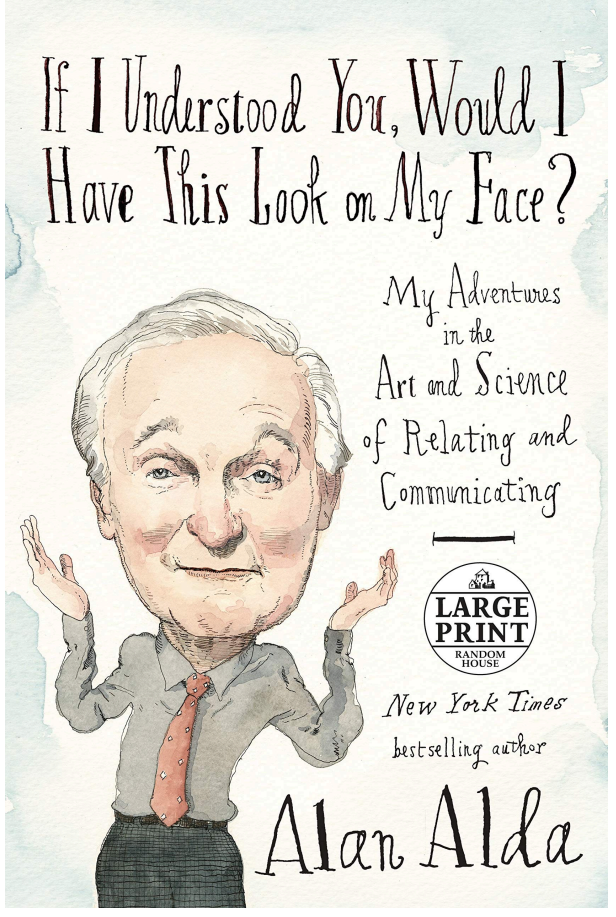
Misinterpretation

Be narrative, but not too narrative

Temper expectations

Ethos

Who can talk about science?



Science communication

RADIOLAB

Gimlet
SCIENCE VS

SCI FRI

Radiolab presents
MORE PERFECT

Ethos

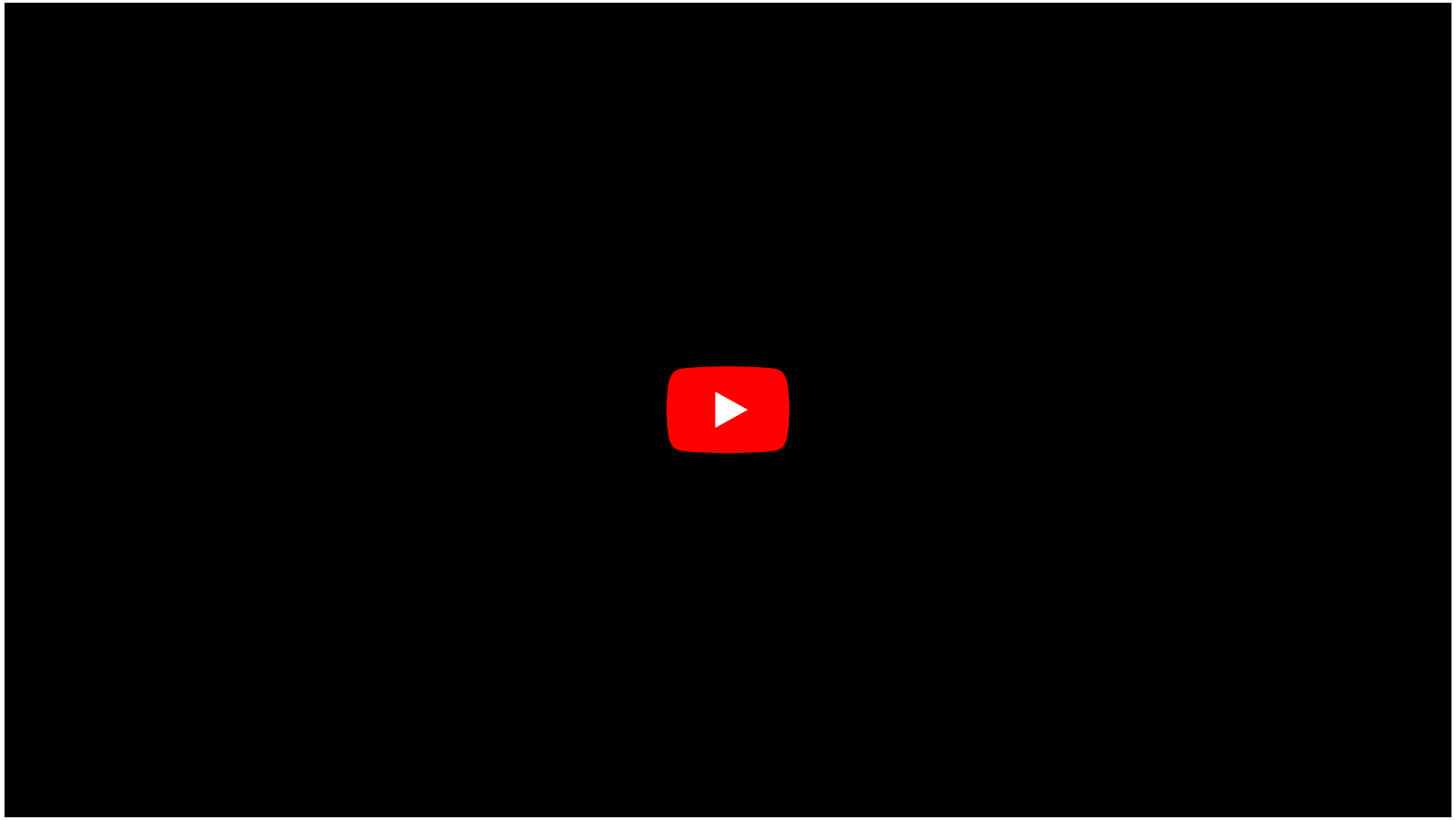
Credentials do not make an expert

Credentials do not *not* make an expert

Credentials ≠ ethos

Equity

Dumbing down vs. translation



Translation

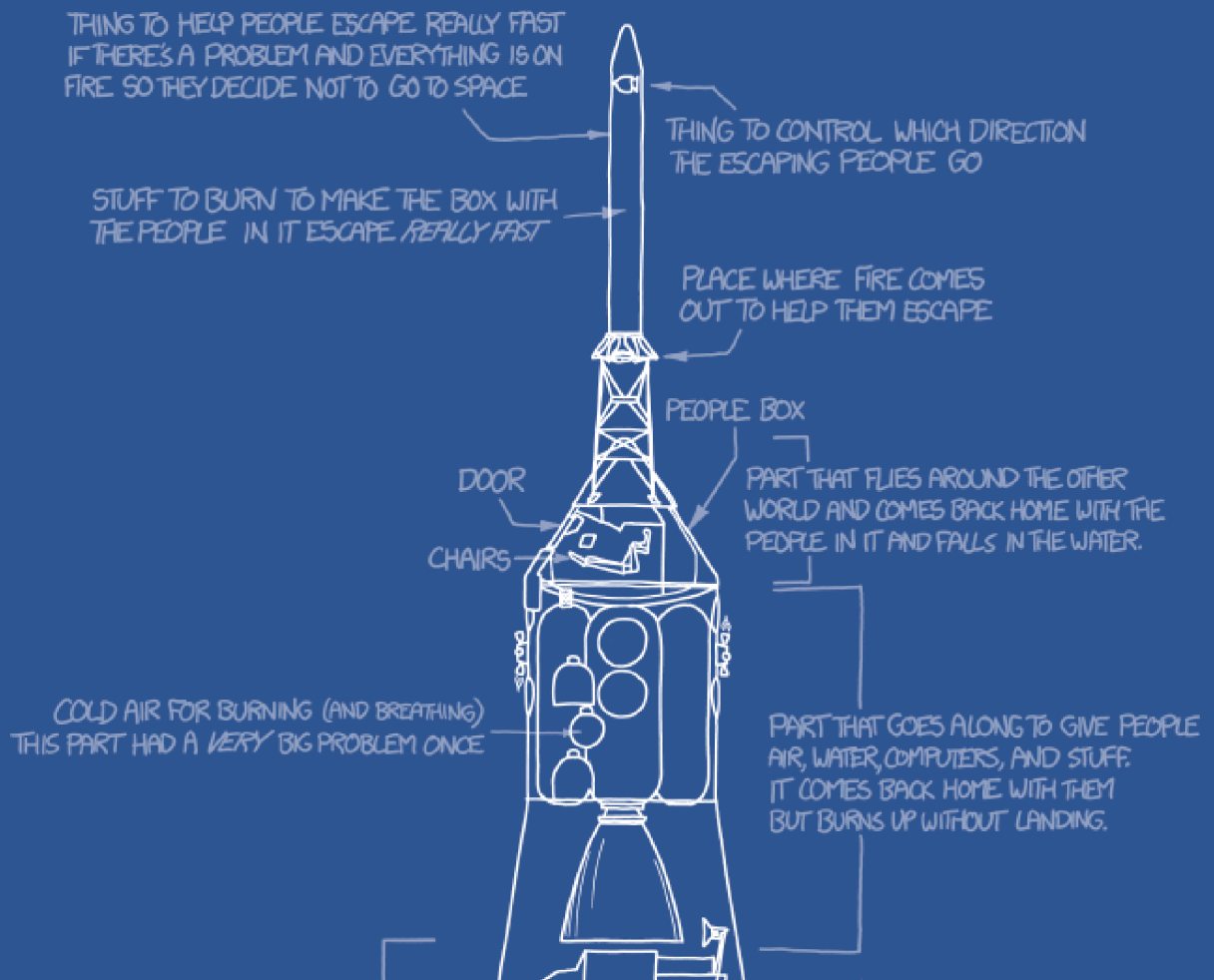
"...the task of the translator consists in finding that intended effect upon the language into which he is translating which produces in it the echo of the original"

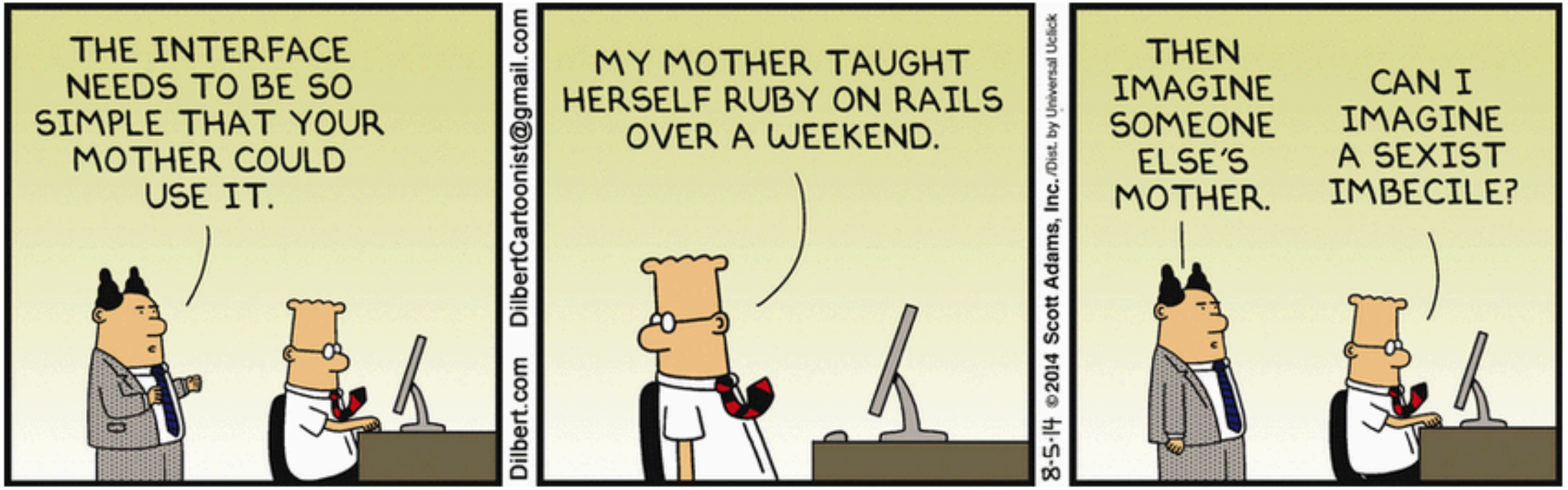
Walter Benjamin,
The Task of the Translator

US SPACE TEAM'S UP GOER FIVE

THE ONLY FLYING SPACE CAR THAT'S TAKEN ANYONE TO ANOTHER WORLD

(EXPLAINED USING ONLY THE TEN HUNDRED WORDS PEOPLE USE THE MOST OFTEN)




THE INTERFACE NEEDS TO BE SO SIMPLE THAT YOUR MOTHER COULD USE IT.
Dilbert.com DilbertCartoonist@gmail.com
MY MOTHER TAUGHT HERSELF RUBY ON RAILS OVER A WEEKEND.
8-5-14 ©2014 Scott Adams, Inc./Dist. by Universal Uclick
THEN IMAGINE SOMEONE ELSE'S MOTHER.
CAN I IMAGINE A SEXIST IMBECILE?

Casey Johnston @caseyjohnston · 4h
So many “solutions” to the lack of women in tech don’t get at the actual problems arstechnica.com/business/2014/…
8
16
View summary
Tomas Sancio @tsancio · 2h
@caseyjohnston read the full article. There's a chicken and egg problem w/ female tech role models. Men want to be the next Jobs/Gates/etc.
View conversation
Casey Johnston @caseyjohnston · 45m
@tsancio I wrote the article
3
52
View conversation

Quantitative evaluation of gender bias in astronomical publications from citation counts

Neven Caplar, Sandro Tacchella & Simon Birrer

Nature Astronomy **1**, Article number: 0141
(2017)
doi:10.1038/s41550-017-0141

The Gender Citation Gap in International Relations

Daniel Maliniak, Ryan Powers and Barbara F. Walter

International Organization / *FirstView* Article / August 2013, pp 1 - 34
DOI: 10.1017/S0020818313000209, Published online: 28 August 2013

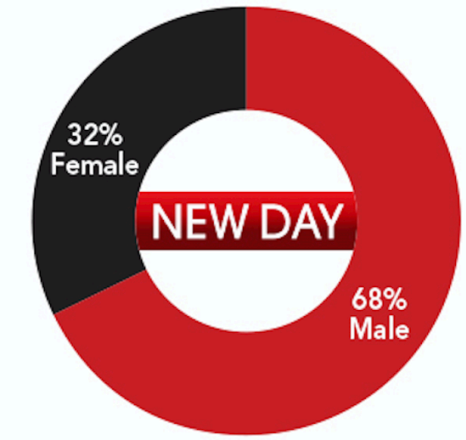
OPEN ACCESS PEER-REVIEWE

RESEARCH ARTICLE

On the Compliance of Women Engineers with a Gendered Scientific System

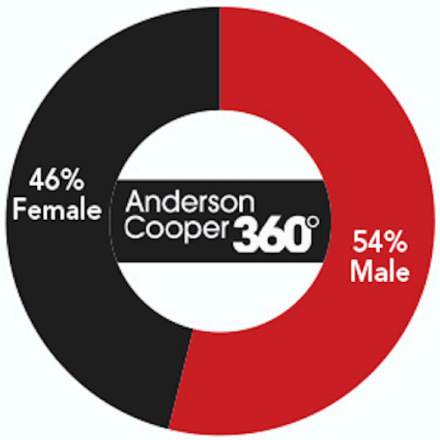
Gita Ghiasi, Vincent Larivière, Cassidy R. Sugimoto

Published: December 30, 2015 • https://doi.org/10.1371/journal.pone.0145931



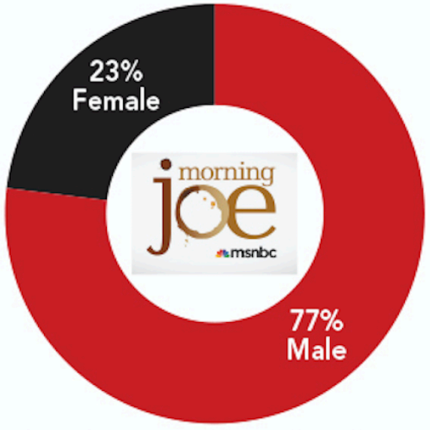
Male analysts (2486) Female analysts (1151)

CNN



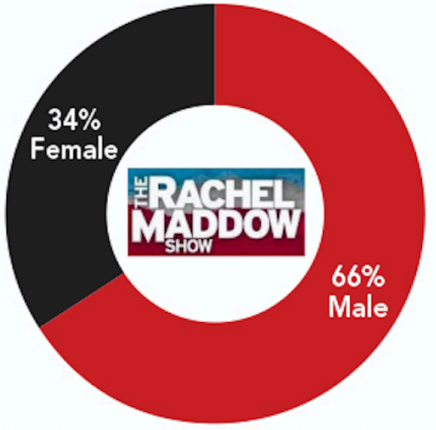
Male analysts (1216) Female analysts (1022)

CNN



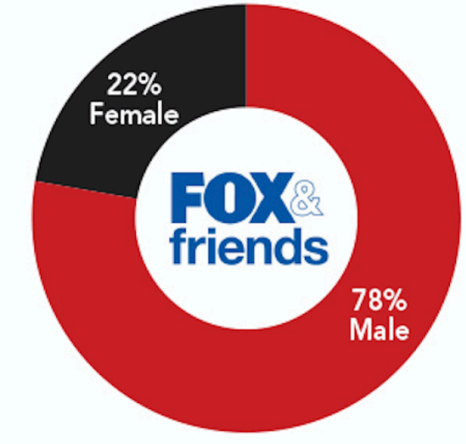
Male analysts (4121) Female analysts (1255)

msnbc



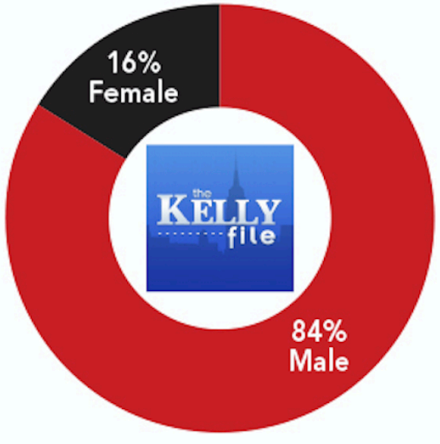
Male analysts (151) Female analysts (79)

msnbc



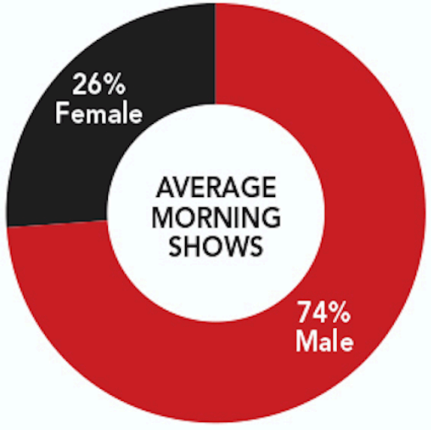
Male analysts (1313) Female analysts (373)

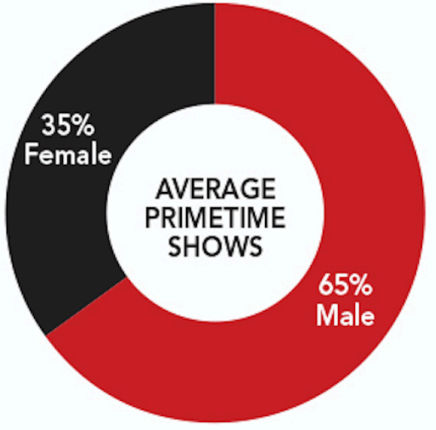
FOX



Male analysts (1019) Female analysts (201)

FOX





LGBT Scholar Network
@LGBTscholars Follows you

POCAlsoKnowStuff
@POCalsoknow

People of color also know stuff! (inspired by @womenalsoknow)

Academic Women in PA
@AWPARocks

A network of women who are (or seeking to be) faculty in public administration seeking to address gender issues in the field.

Gender Balance Assessment Tool (GBAT)

Women are cited less often than men, and are also underrepresented in syllabi. Yet even well-meaning scholars may find that they have difficulty assessing how gender-balanced their bibliographies and syllabi really are. Counting is tedious and prone to human error, and scholars may not know the gender identities of all t
help with that, by automating the process of evalua
name and then providing an estimate of what perc
women.

Your assigned readings are approximately

47.43

percent woman-authored.

Race breakdown (probabilistic)

6.48% Asian, 14.39% Black, 2.74% Hispanic,

2.68% Other, 73.71% White

https://jlsumner.shinyapps.io/syllabustool/

Equity

Don't dumb down your findings

You are a translator

Treat audience with respect

Amplify underrepresented voices

Curiosity

How do I keep learning R?

What class should I take next?

What book should I read next?

How do I keep learning R?

~~What class should I take next?~~

~~What book should I read next?~~

Be curious!

Teaching yourself

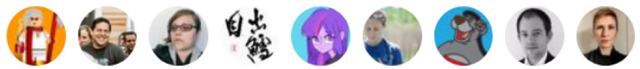
Katie Mack
@AstroKatie

Following

A surprisingly large part of having expertise in a topic is not so much knowing everything about it but learning the language and sources well enough to be extremely efficient in google searches.

9:34 AM - 8 Dec 2018

3,607 Retweets **14,911** Likes



195 3.6K 15K

FAMILY

I'm a Developer. I Won't Teach My Kids to Code, and Neither Should You.

By JOE MORGAN

DEC 06, 2018 • 5:55 AM

Every step—precisely measuring ingredients, gauging mixed dough for smoothness and consistency, placing precision cuts to minimize waste—taught him something about quality. It's hard to teach the difference between merely executing steps, such as following a recipe, and doing something well. It can only be passed on through feel and experience. And every time you involve your kids when you work on something you value, you are teaching them how to do things well. You are preparing them to write code.

But you're not only teaching them that. You're teaching them the world is full of interesting things to discover. You're showing them how to be passionate and look for that ephemeral sense of quality in everything they do. The best part is that even if they don't become coders—most shouldn't and won't—the same skills can be used in nearly any career, in every hobby, in every life. When we force kids to learn syntax, we reinforce the idea that if something is not a blatantly employable skill, it's not valuable. Adults can learn syntax. Only kids can learn to embrace curiosity.

Two secrets to master R

1: Find excuses to use it

2: Share and work in public

Find excuses to use R

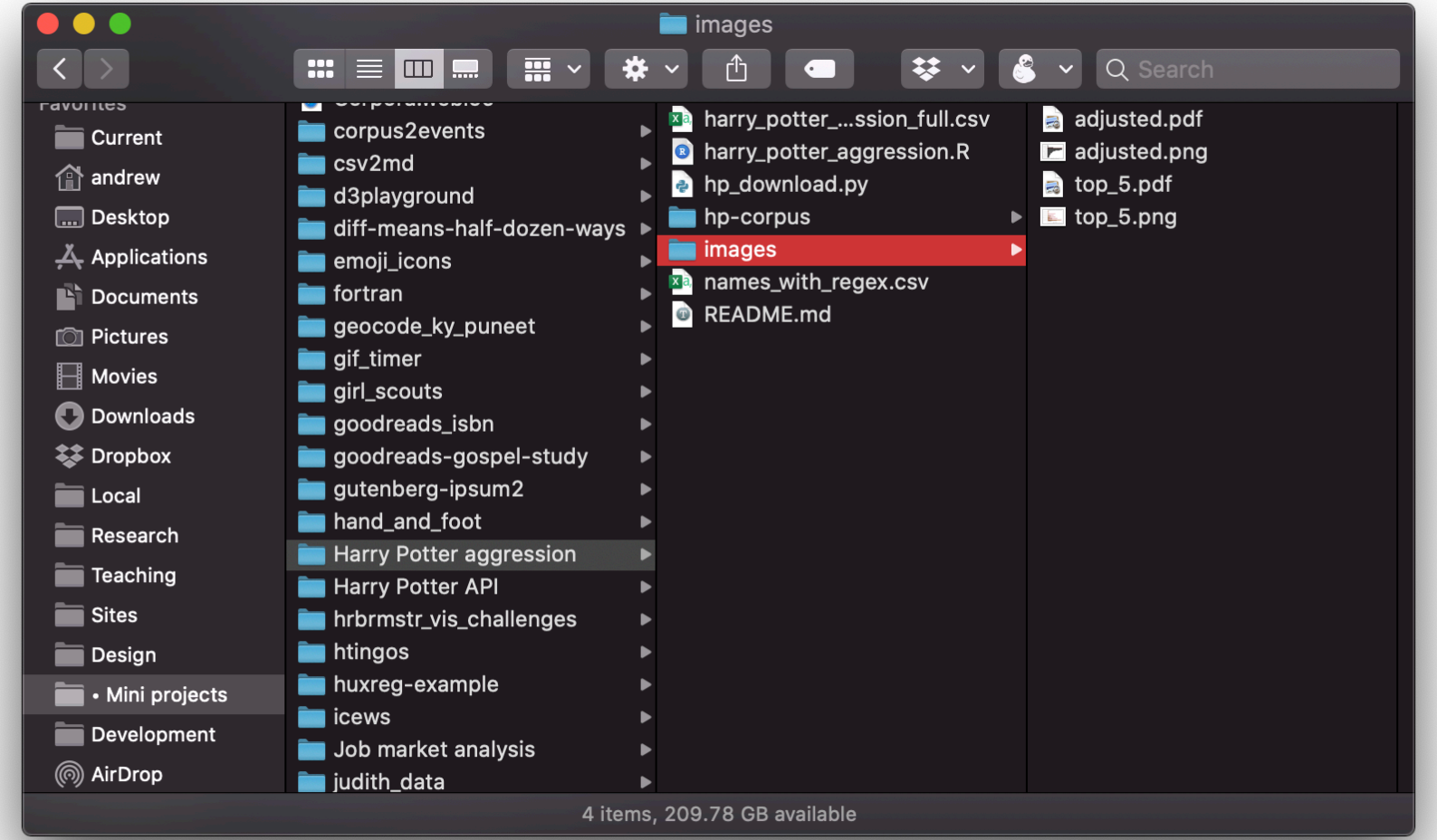
Playing with R

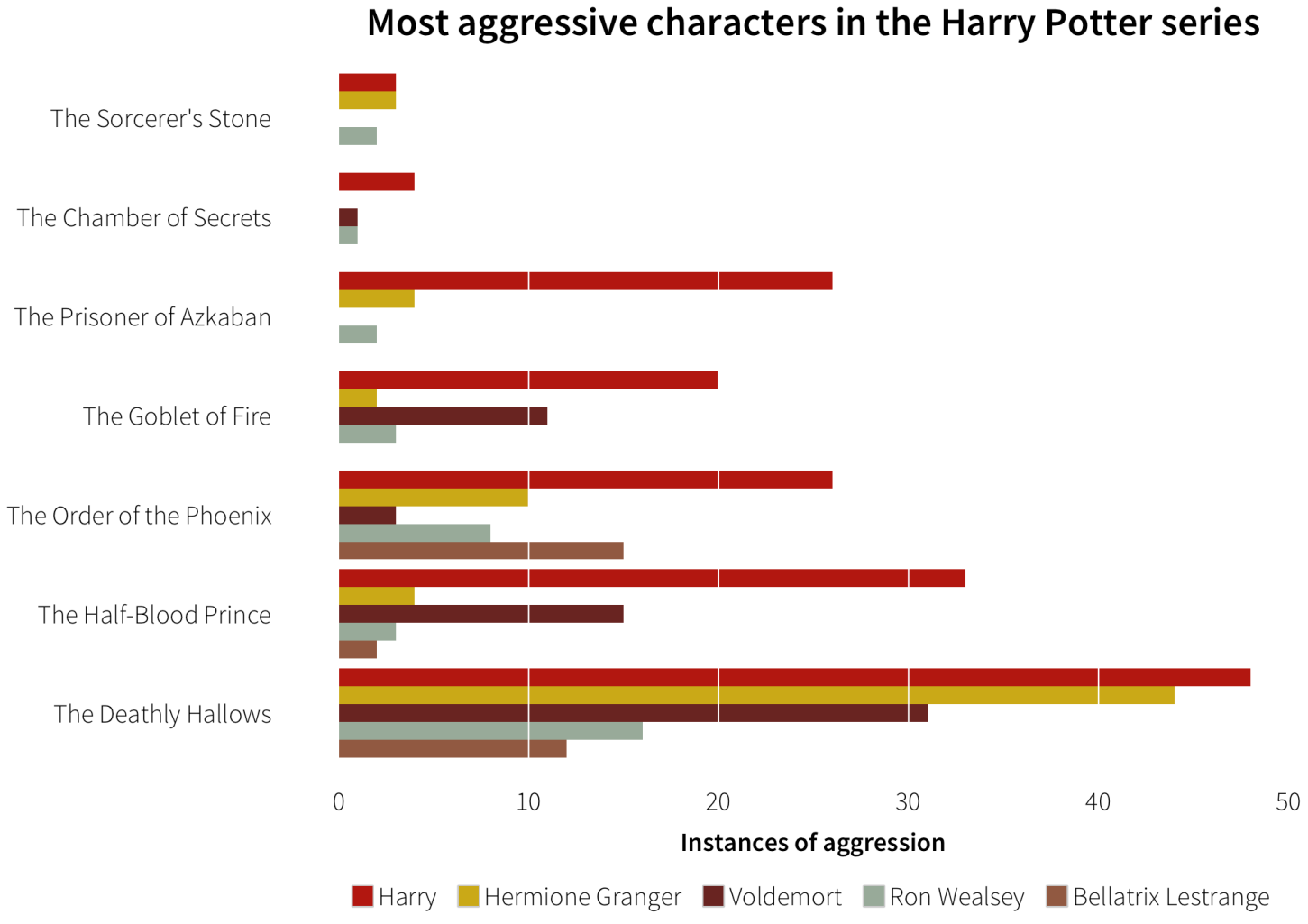
Little exploration projects

#TidyTuesday

Data play time

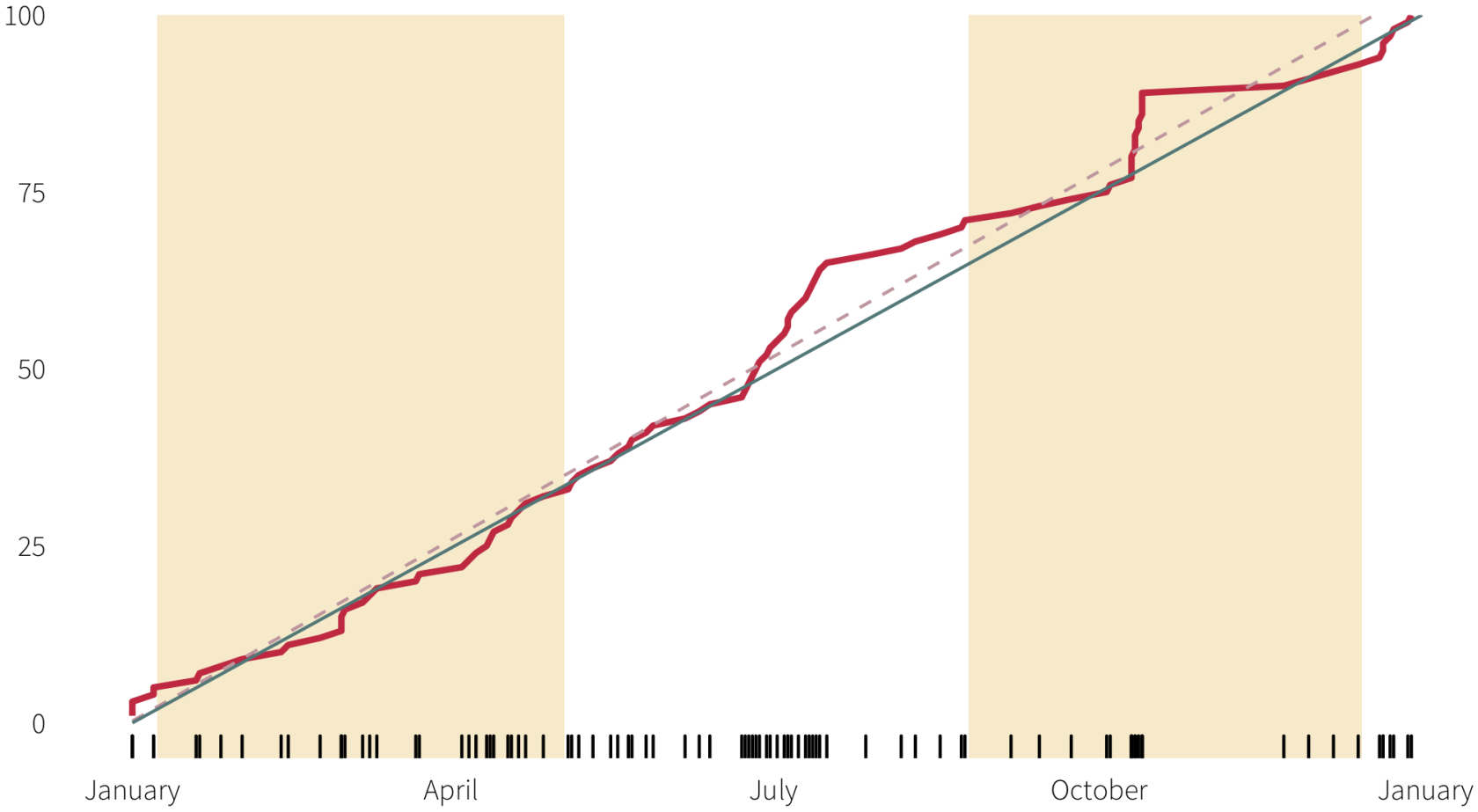
Actual projects

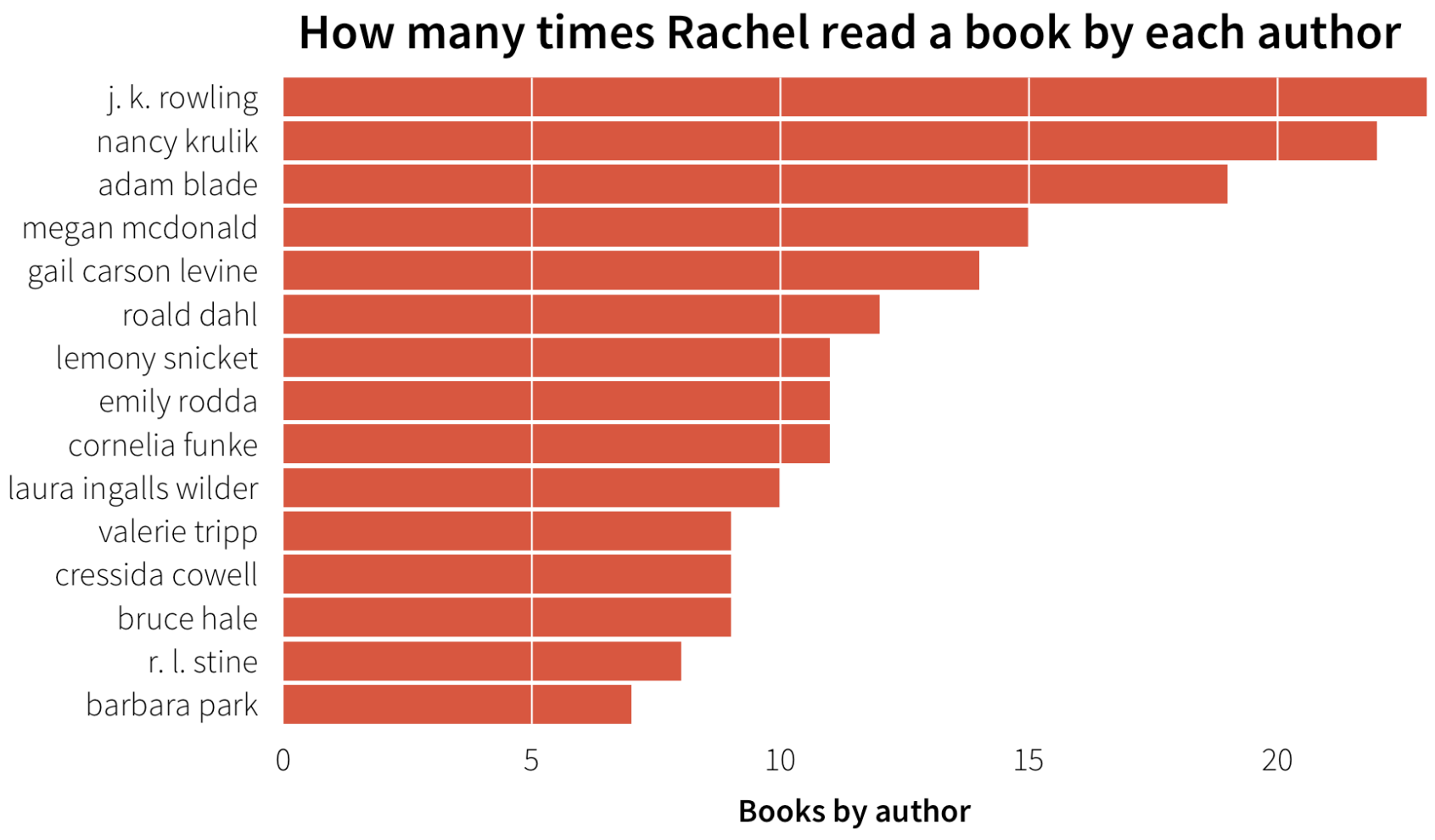

images
Search
Favorites
Current
andrew
Desktop
Applications
Documents
Pictures
Movies
Downloads
Dropbox
Local
Research
Teaching
Sites
Design
• Mini projects
Development
AirDrop
corpus2events
csv2md
d3playground
diff-means-half-dozen-ways
emoji_icons
fortran
geocode_ky_puneet
gif_timer
girl_scouts
goodreads_isbn
goodreads-gospel-study
gutenberg-ipsum2
hand_and_foot
Harry Potter aggression
Harry Potter API
hrbrmstr_vis_challenges
htingos
huxreg-example
icews
Job market analysis
judith_data
harry_potter_...ssion_full.csv
harry_potter_aggression.R
hp_download.py
hp-corpus
images
names_with_regex.csv
README.md
adjusted.pdf
adjusted.png
top_5.pdf
top_5.png
4 items, 209.78 GB available


Most aggressive characters in the Harry Potter series
The Sorcerer's Stone
The Chamber of Secrets
The Prisoner of Azkaban
The Goblet of Fire
The Order of the Phoenix
The Half-Blood Prince
The Deathly Hallows
0
10
20
30
40
50
Instances of aggression
Harry
Hermione Granger
Voldemort
Ron Wealsey
Bellatrix Lestrange

Cumulative number of family walks in 2014

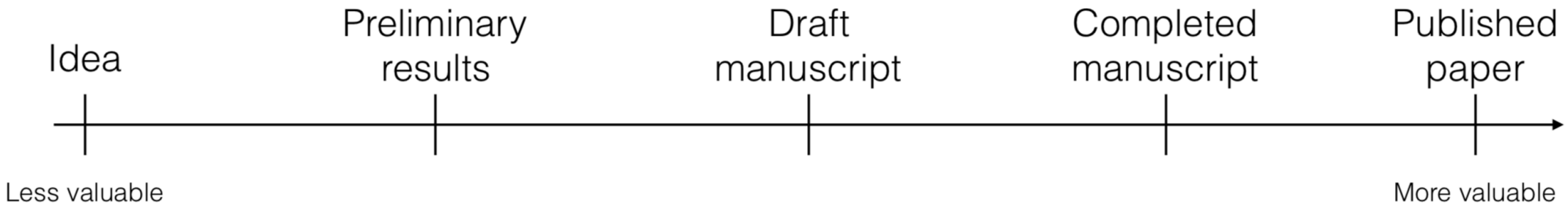
Duke semesters shaded in yellow




How many times Rachel read a book by each author
j. k. rowling
nancy krulik
adam blade
megan mcdonald
gail carson levine
roald dahl
lemony snicket
emily rodda
cornelia funke
laura ingalls wilder
valerie tripp
cressida cowell
bruce hale
r. l. stine
barbara park
0
5
10
15
20
Books by author

Radical transparency & public work

How we normally think of our work and goals



How we should think of our work and goals

David Robinson, The unreasonable effectiveness of public work"

Benefits of working in public

Build reputation

Learn more

Grow the community

Early feedback on ideas

Validation

2016-17

Political science (43)

Public administration and policy (41)

2017-18

Political science (11)

Public administration and policy (31)

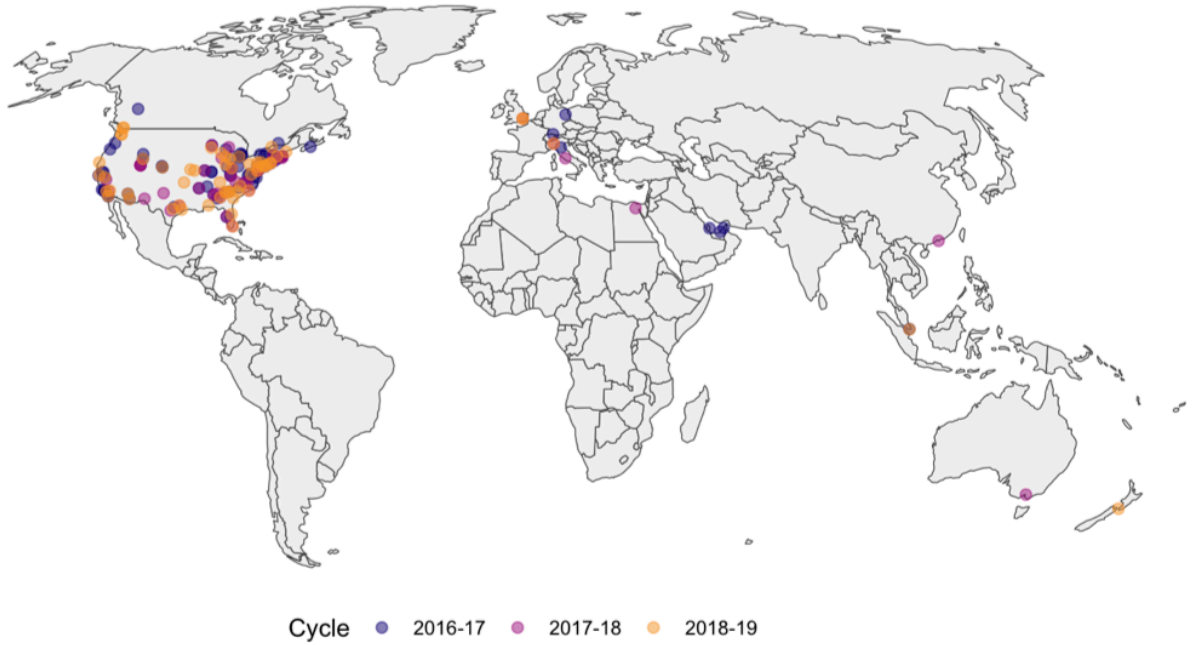
2018-19

Political science (37)

Public administration and policy (23)

Nothing Skype, no flyout Flyout, no offer Visiting offer Tenure-track offer

One box = one job posting



Andrew Heiss

International NGOs, nonprofit management, authoritarianism, data science, and R

• 2007–2019

ORCID iD: 0000-0002-3948-3914

PGP public • PGP fingerprint:
4AA2 FA83 A8B2 05A4 E30F
610D 1382 6216 9178 36AB

Code for site

Monday, December 17, 2018

The academic job search finally comes to an end

I am *so beyond thrilled* to announce that I'll be joining the Andrew Young School of Policy Studies at Georgia State University in Fall 2019 as an assistant professor in the Department of Public Management and Policy. I'll be teaching classes in statistics/data science, economics, and nonprofit management in beautiful downtown Atlanta, and we'll be moving back to the South. I am so so excited about this! The Andrew Young School does amazing work in public policy, administration, and nonprofit management, and I'll be working with phenomenal colleagues and students. I still can't believe this is real.

Part of the reason I'm in shock is that for the past 2.5 years, I've been ripped apart and destroyed by the academic job market. This job market is a horrendous beast of a thing. It is soul-crushing and dream-shattering and a constant stream of rejection. While facing rejection is good and builds grit etc., etc., in reality it's awful.

In an effort to stay On Brand™, here are a bunch of fancy graphs and numbers showing what it's been like to apply for nearly 200 jobs since August 2016. Unlike many of my other blog posts, I haven't included any of the code to generate these. That code is all available in a GitHub repository (see `README.Rmd`), along with the raw data that I've collected over the past few years (for the morbidly curious).

Application count and outcomes

Between August 31, 2016 and November 18, 2018, I applied for 186 tenure-track and non-tenure-track academic jobs at R1 schools, liberal arts colleges, and teaching-focused public universities. I was offered one two-year visiting assistant professorship at the Romney

GitHub, Inc. (US) https://github.com/andrewheiss/academic-job-market/blob/master/README.Rmd 90%

523 lines (430 sloc) | 25.8 KB Raw Blame History

```
---
title: "The academic job search finally comes to an end"
output: github_document
editor_options:
  chunk_output_type: console
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, fig.retina = 2)
```

> See the [actual blog post](https://www.andrewheiss.com/blog/2018/12/17/academic-job-market-visualized/).

I am *so beyond thrilled* to announce that I'll be joining the [Andrew Young School of Policy Studies](https://aysps.gsu.edu/) a

Part of the reason I'm in shock is that for the past 2.5 years, I've been ripped apart and destroyed by the academic job market

In an effort to stay On Brand™, here are a bunch of fancy graphs and numbers showing what it's been like to apply for nearly 20

```{r load-libraries-data, warning=FALSE, message=FALSE}
library(tidyverse)
library(lubridate)
library(here)
library(sf)
library(waffle)
library(ggstance)
library(scales)
library(countrycode)
# library(mapview)  # For interactive maps!
library(units)
library(patchwork)

# Load jobs data
jobs_clean <- read_csv(here("data", "jobs_clean.csv")) %>%
  mutate_at(vars(`Skype interview`, `Flyout`, contains("ffer")),
            funs(bin = !is.na(.)))
```
```

How to work in public

Tweet, blog, and meet people

Play with data in public

Teach concepts (for yourself too!)

Communities

#rstats

R User Groups

#rladies

Rmd websites, blogdown, bookdown

Play with data in public

Saturday, August 26, 2017

Quickly play with Polity IV and OECD data (and see the danger of US democracy)

The Polity IV Project released new data yesterday, with democratization scores for 169 countries up to 2016. I wanted to check if the ongoing erosion of US democratic institutions since the 2016 elections registered in the US's Polity score, and, lo and behold, it did! We dropped from our solid, historically consistent 10 to an 8.

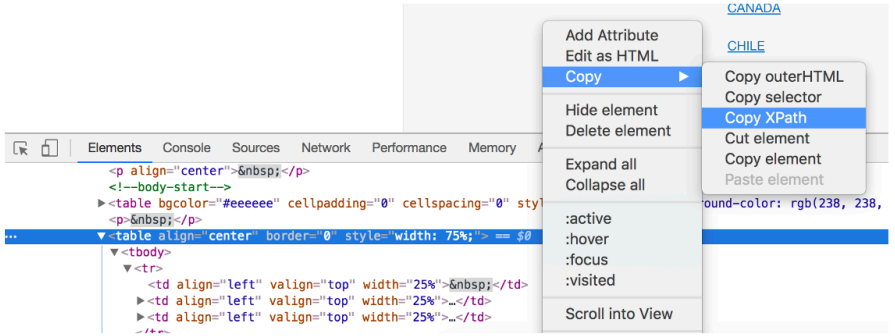
But is that bad? How does that compare to other advanced democracies, like countries in the OECD?

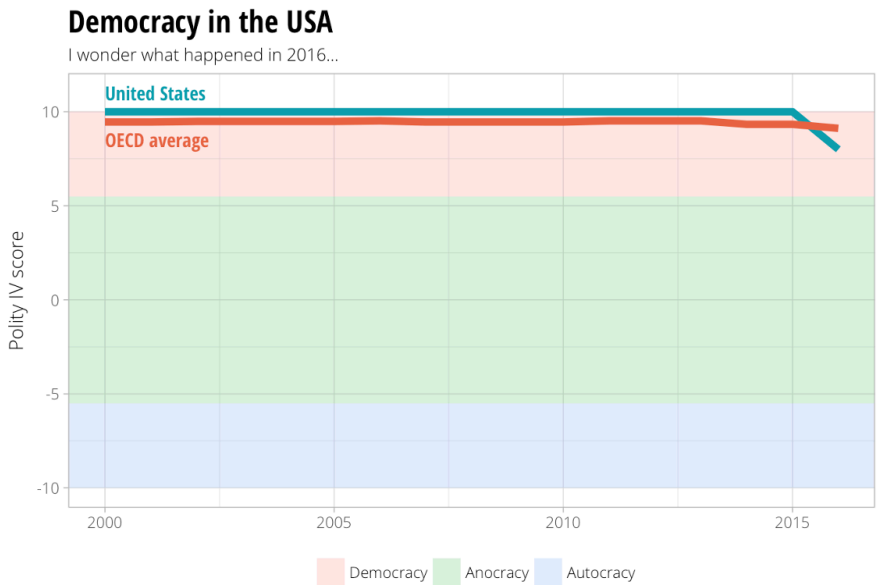
What follows below shows how relatively easy it is to quickly and reproducibly grab the new data, graph it, and compare scores across countries. (This notebook is also in a GitHub repository.)

Before we start, we'll load all the libraries we'll need:

```
library(tidyverse)      # dplyr, ggplot, etc.
library(readxl)         # Read Excel files
library(forcats)        # Deal with factors
library(countrycode)    # Deal with country codes and names
library(rvest)          # Scrape websites
library(httr)           # Download stuff
library(ggrepel)        # Place non-overlapping labels on plots
```

First, we have to download the new Polity data. We could navigate to the Polity IV data page and download the data manually, but that's not scriptable. Instead, we can use `GET()` from





Source: Polity IV Project

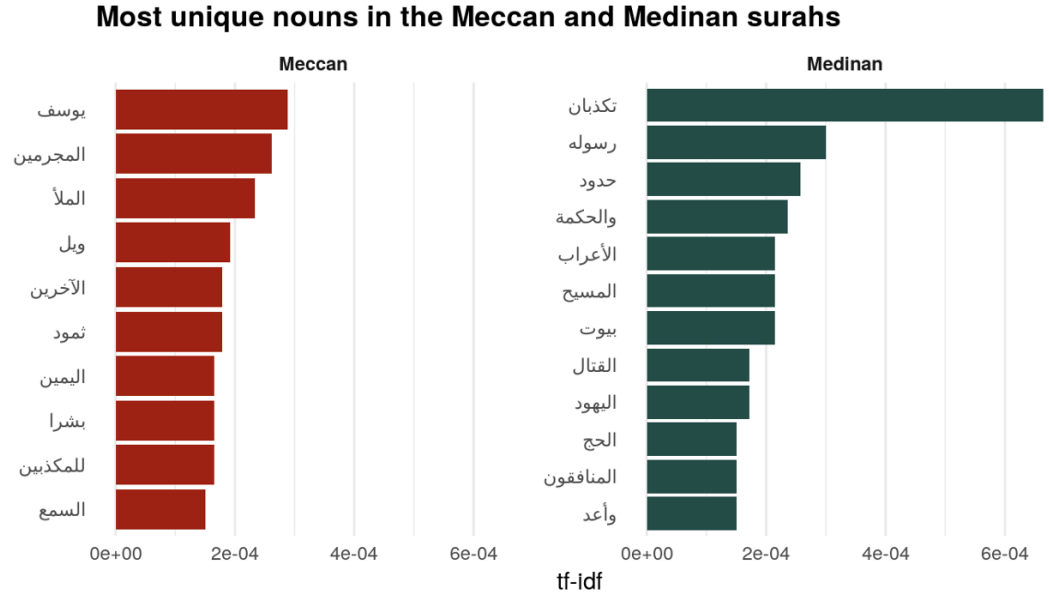
Friday, December 28, 2018

Tidy text, parts of speech, and unique words in the Qur'an

(See this notebook on GitHub)

As I showed in a previous blog post, the **cleanNLP** package is a phenomenal frontend for natural language processing in R. Rather than learn the exact syntax for NLP packages like spaCy or CoreNLP, you can use a consistent set of functions and let **cleanNLP** handle the API translation behind the scenes for you.

Previously, I used spaCy to tag the parts of speech in the Four Gospels to find the most distinctive nouns and verbs in the Gospel of John. Here, I'll show a quick example of how to use CoreNLP to tag parts of speech in Arabic. CoreNLP is far far far slower than spaCy, but it can handle languages like Arabic and Chinese, which is pretty magical.



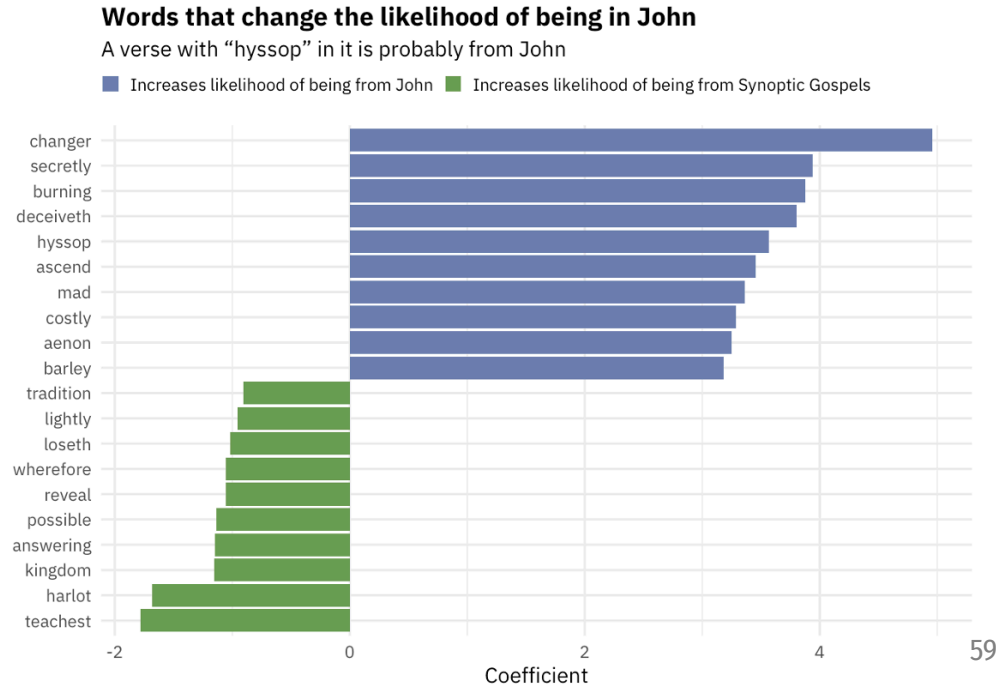
Wednesday, December 26, 2018

Tidy text, parts of speech, and unique words in the Bible

(See this notebook on GitHub)

As part of my goal to read some sort of religiously themed book every day (what I've read so far), I've been reading Eric Huntsman's new *Becoming the Beloved Disciple*, a close reading of the Gospel of John from an LDS perspective.

Near the beginning, Huntsman discusses several word frequencies that make John unique compared to the synoptic gospels of Matthew, Mark, and Luke (which all draw on the same Q source). For instance, Huntsman states that John focuses more on themes of discipleship



Teach a concept

Tuesday, January 29, 2019

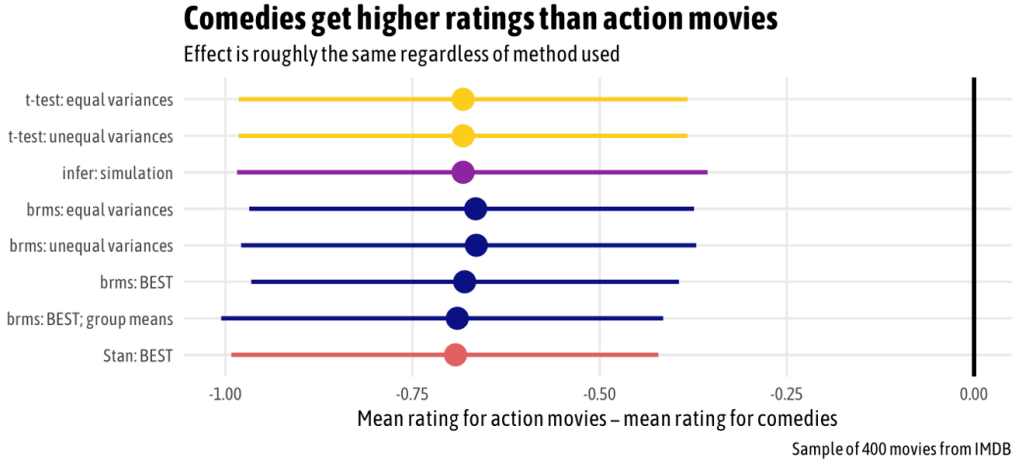
Half a dozen frequentist and Bayesian ways to measure the difference in means in two groups

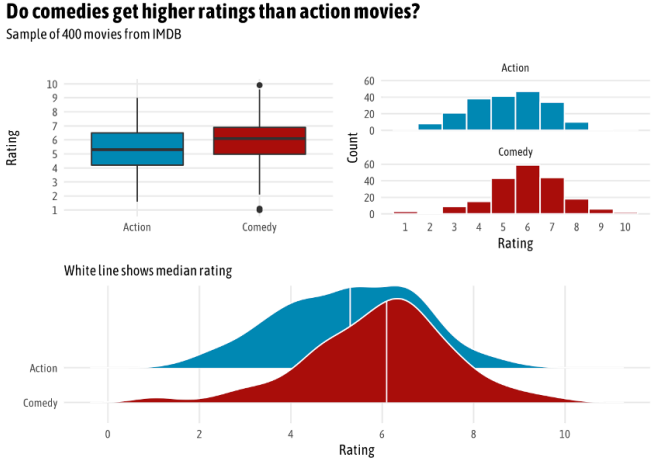
(See this notebook on GitHub)

Taking a sample from two groups from a population and seeing if there's a significant or substantial difference between them is a standard task in statistics. Measuring performance on a test before and after some sort of intervention, measuring average GDP in two different continents, measuring average height in two groups of flowers, etc.—we like to know if any group differences we see are attributable to chance / measurement error, or if they're real.

Classical frequentist statistics typically measures the difference between groups with a t-test, but t-tests are 100+ years old and statistical methods have advanced a lot since 1908. Nowadays, we can use simulation and/or Bayesian methods to get richer information about the differences between two groups without worrying so much about the assumptions and preconditions for classical t-tests.

Mostly as a resource to future me, here are a bunch of different ways to measure the difference in means in two groups. I've done them all in real life projects, but I'm tired of constantly searching my computer for the code to do them:)





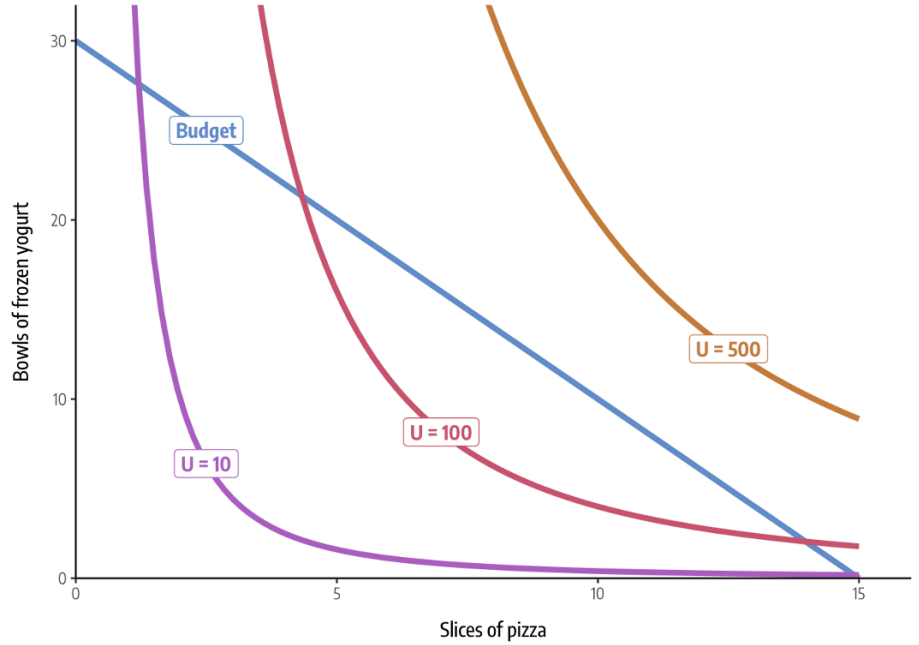
Saturday, February 16, 2019

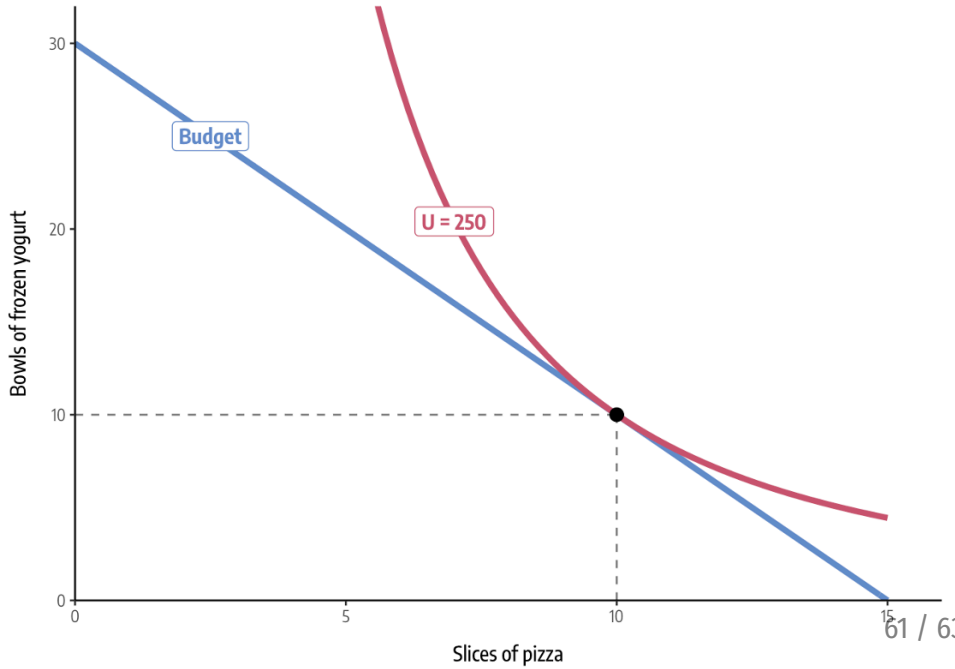
Chidi's budget and utility: doing algebra and calculus with R and yacas

(See this notebook on GitHub)

A year ago, I wrote about how to use R to solve a typical microeconomics problem: finding the optimal price and quantity of some product given its demand and cost. Doing this involves setting the first derivatives of two functions equal to each other and using algebra to find where they cross. I showed how to use neat functions like `Deriv::Deriv()` and `splinefun()` and make fancy plots showing supply and demand and it's pretty cool. I wrote it mostly because I was teaching an introductory microeconomics course and wanted an easy, generalizable, and manual math-less way to make these plots for my students' exercises and problem sets, and it works great.

I'm teaching microeconomics again this year and decided to tackle a trickier problem that involves curvier curves, more variables, and more math. And the results are even cooler and open the door for more doing math and symbolic algebra directly with R.





You are all expert enough now.

Go make beautiful things!