Ethics, stories, and curiosity

Session 14

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

What did we just learn?

Ethics of data analytics

Ethics of storytelling

Curiosity

What did we just learn?

Course objectives

By the end of this course, you (1) will be literate in the language of causal inference, (2) will communicate evaluation outcomes clearly, and (3) will understand the ethics and limits of data analysis by designing, critiquing, coding, and running rigorous, valid, and feasible evaluations of public sector programs focused on society's most pressing problems.

Specifically, you'll be able to:

- Explain the philosophy of causation
- Identify and diagram program logic models
- Outline theories of change with directed acyclic graphs (DAGs)
- Summarize key threats to causal inference, identify these threats in evaluations, and mitigate these threats with research design
- Develop rigorous and valid statistical measures
- Run statistical models
- Explain the theory, research design, methods, and results of evaluations to all types of stakeholders, from highly trained econometricians to the general public
- Share your analyses and data with the public
- Identify ethical issues and limits in data science and program evaluation
- Become curious and confident in consuming and producing evaluations

Evaluation and causation

Program theories

Logic models Measurement

DAGs Potential outcomes

Tools and methods

Randomization Matching

Difference-in-differences

Regression discontinuity

Instrumental variables

Applied evaluation

Preregistration Ethics

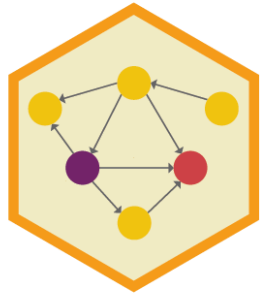
Communication

Other evaluations

R and the tidyverse

Data manipulation Modeling

R Markdown Visualization



Program Evaluation for Public Service

Main takeaways

Don't be afraid of causal language!

With careful use of DAGs and special research designs, you can make causal claims

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results.

Miguel A. Hernán, MD, DrPH

See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

> Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Many journal editors request authors to avoid causal language,[1] and many observational researchers, trained in a scientific environment that frowns upon causality claims, spontaneously refrain from mentioning the

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating "causal" as a dirty word that respectable investigators do not say in public or put in print. It is true that observational studies cannot definitely prove causation, but this statement misses the point, as discussed in this commentary.

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine

Ethics of data analytics

Powerful tools

R is an incredibly valuable skill

Causal inference is an incredibly valuable skill

These tools can be used to improve the world!

And potentially harm it

Possible pitfalls

Manipulation

Don't coerce people

Bias

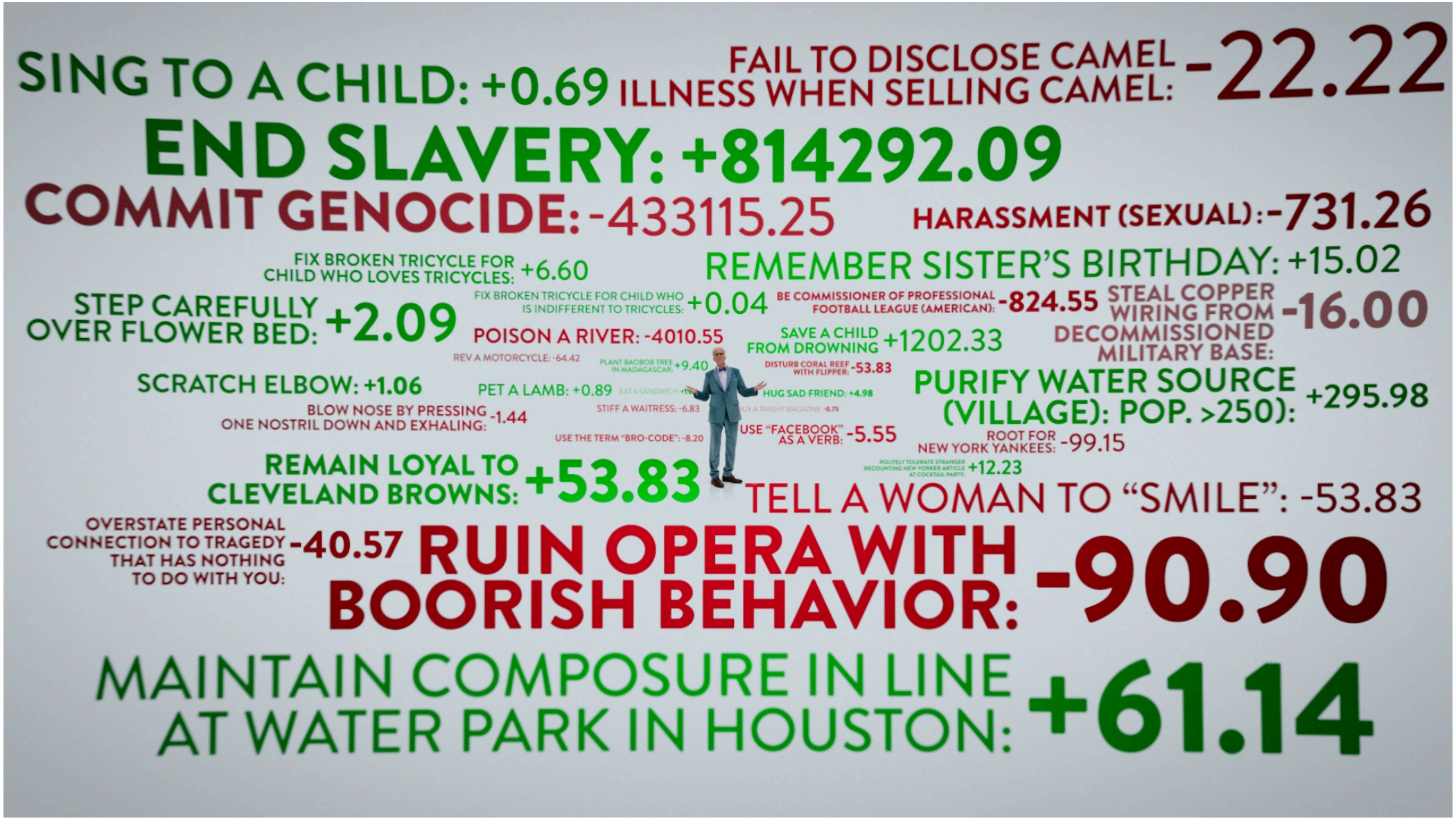
There's no such thing as objective data or models

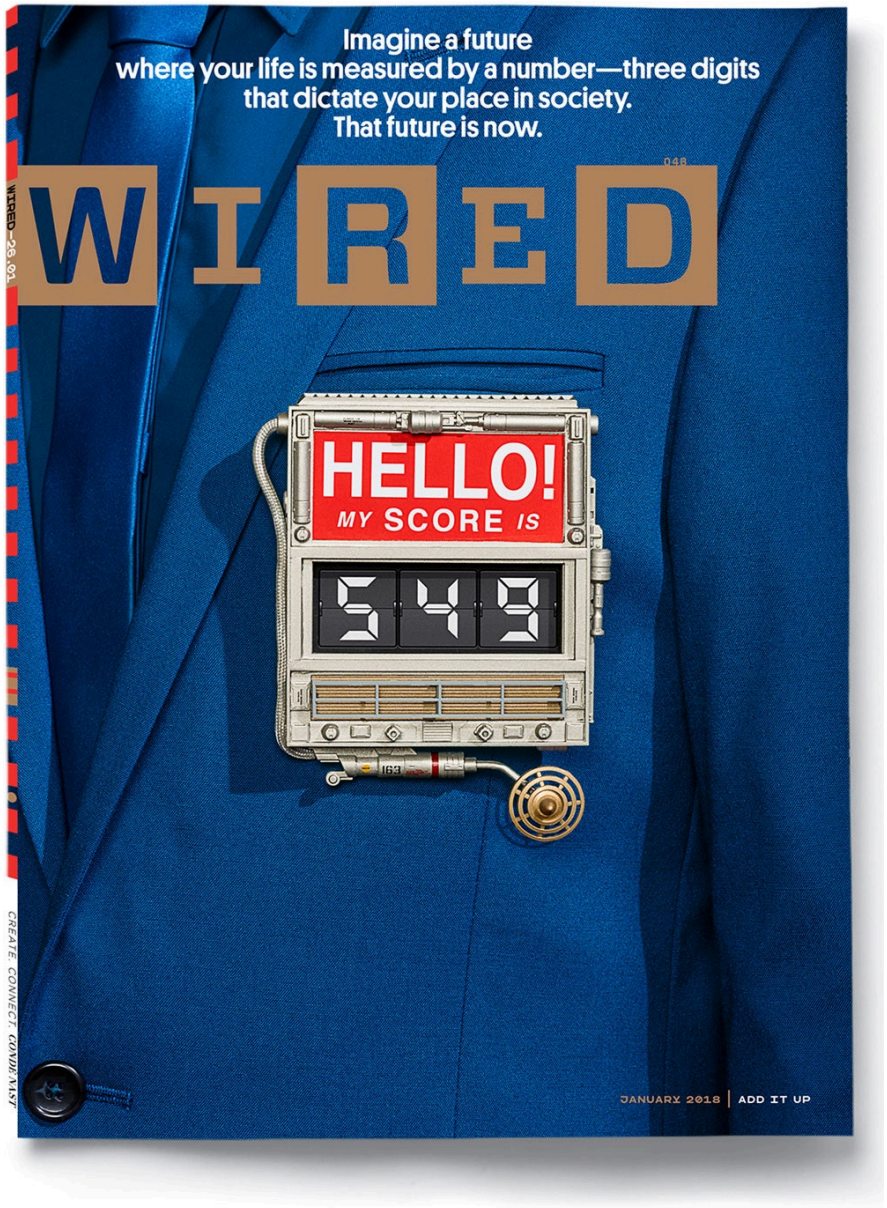
Accidental evil

Don't let stupidity transform into evil

Manipulation

Welcome!
Everything is fine.


SING TO A CHILD: +0.69
FAIL TO DISCLOSE CAMEL ILLNESS WHEN SELLING CAMEL: -22.22
END SLAVERY: +814292.09
COMMIT GENOCIDE: -433115.25
HARASSMENT (SEXUAL): -731.26
FIX BROKEN TRICYCLE FOR CHILD WHO LOVES TRICYCLES: +6.60
REMEMBER SISTER'S BIRTHDAY: +15.02
STEP CAREFULLY OVER FLOWER BED: +2.09
FIX BROKEN TRICYCLE FOR CHILD WHO IS INDIFFERENT TO TRICYCLES: +0.04
BE COMMISSIONER OF PROFESSIONAL FOOTBALL LEAGUE (AMERICAN): -824.55
STEAL COPPER WIRING FROM DECOMMISSIONED MILITARY BASE: -16.00
POISON A RIVER: -4010.55
SAVE A CHILD FROM DROWNING +1202.33
REV A MOTORCYCLE: -64.42
PLANT BAOBOB TREE IN MADAGASCAR: +9.40
DISTURB CORAL REEF WITH FLIPPER: -53.83
SCRATCH ELBOW: +1.06
PET A LAMB: +0.89
EAT A SANDWICH: +0.
HUG SAD FRIEND: +4.98
PURIFY WATER SOURCE (VILLAGE): POP. >250): +295.98
BLOW NOSE BY PRESSING ONE NOSTRIL DOWN AND EXHALING: -1.44
STIFF A WAITRESS: -6.83
BUY A TRASHY MAGAZINE: -0.75
USE THE TERM "BRO-CODE": -8.20
USE "FACEBOOK" AS A VERB: -5.55
ROOT FOR NEW YORK YANKEES: -99.15
REMAIN LOYAL TO CLEVELAND BROWNS: +53.83
POLITELY TOLERATE STRANGER RECOUNTING NEW YORKER ARTICLE AT COCKTAIL PARTY: +12.23
TELL A WOMAN TO "SMILE": -53.83
OVERSTATE PERSONAL CONNECTION TO TRAGEDY THAT HAS NOTHING TO DO WITH YOU: -40.57
RUIN OPERA WITH BOORISH BEHAVIOR: -90.90
MAINTAIN COMPOSURE IN LINE AT WATER PARK IN HOUSTON: +61.14


Imagine a future
where your life is measured by a number—three digits
that dictate your place in society.
That future is now.
WIRED
HELLO!
MY SCORE IS
549
JANUARY 2018 | ADD IT UP
CREATE. CONNECT. CONVERSE.

Instagram's feed ranking criteria

Instagram relies on machine learning based on your past behavior to create a unique feed for everyone. Even if you follow the exact same accounts as someone else, you'll get a personalized feed based on how you interact with those accounts.

Three main factors determine what you see in your Instagram feed:

1. **Interest:** How much Instagram predicts you'll care about a post, with higher ranking for what matters to you, determined by past behavior on similar content and potentially machine vision analyzing the actual content of the post.
2. **Recency:** How recently the post was shared, with prioritization for timely posts over weeks-old ones.
3. **Relationship:** How close you are to the person who shared it, with higher ranking for people you've interacted with a lot in the past on Instagram, such as by commenting on their posts or being tagged together in photos.

THE WALL STREET JOURNAL.

Blue Feed, Red Feed

See Liberal Facebook and Conservative Facebook, Side by Side

This page has been archived and will no longer be updated

Last Updated Aug. 19, 2019

To begin, click on a topic:

PRESIDENT TRUMP | HEALTH CARE | GUNS | ABORTION | ISIS | BUDGET | EXECUTIVE ORDER | IMMIGRATION

LIBERAL

SHOWING POSTS ABOUT:
"PRESIDENT TRUMP"

CONSERVATIVE

MSNBC
on Wednesday

New York Attorney Gen.-elect Letitia James says she plans to launch sweeping investigations into President Trump, his family and "anyone" in his circle who may have violated the law.

NBCNEWS.COM
Incoming New York attorney general plans wide-rang...

2.5K 334 742

LIVE: Senate Minority Leader Schumer speaks on Senate floor about his Oval Office meeting with President Trump.
https://nbcnews.to/2EezUoG

2.5K 2K 980

The Raw Story
on Wednesday

Can't fix stupid.

RAWSTORY.COM

The Federalist Papers
on Wednesday

The President just set three traps for Schumer and Pelosi on immigration yesterday and they fell for all three.

This is brilliant!

THEFEDERALISTPAPERS.ORG
Trump Just Set 3 Traps For Schumer, Pelosi Over Im...
Hook, line and sinker.

29K 1.5K 10K

The Daily Signal
on Wednesday

Michael Cohen's payments to two women during the 2016 campaign may have been unseemly—but they were not illegal.

DAILYSIGNAL.COM
Trump's Ex-Lawyer Didn't Violate Campaign Finance ...
Michael Cohen's payments to two women during the 2016 campa...

2.3K 165 1.1K

Breitbart
on Wednesday

"I'm not concerned, no. I think that the people would revolt if that happened," Trump said.

WIRED

MAT HONAN GEAR 08.11.2014 06:30 AM

I Liked Everything I Saw on Facebook for Two Days. Here's What It Did to Me

I like everything. Or at least I did, for 48 hours. Literally everything Facebook sent my way, I liked---even if I hated it.

GIMLET

GIMLET
/reply-all/
NOVEMBER 2, 2017
#109 Is Facebook Spying on You?
by REPLY ALL
LISTEN NOW
LISTEN ON Spotify
Listen on Apple Podcasts

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff
Tech
Welcome to The Not-So Private Parts where technology & privacy collide

As Pole's computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a "pregnancy prediction" score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

One Target employee I spoke to provided a hypothetical example. Take a fictional Target shopper named Jenny Ward, who is 23, lives in Atlanta and in March bought cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug. There's, say, an 87 percent chance that she's pregnant and that her delivery date is sometime in late August.

via How Companies Learn Your Secrets - NYTimes.com.

AIRLINES FACE CRACK DOWN ON USE OF 'EXPLOITATIVE' ALGORITHM THAT SPLITS UP FAMILIES ON FLIGHTS

Government ministers have condemned the practice

Helen Coffey | @LenniCoffey | Monday 19 November 2018 12:22

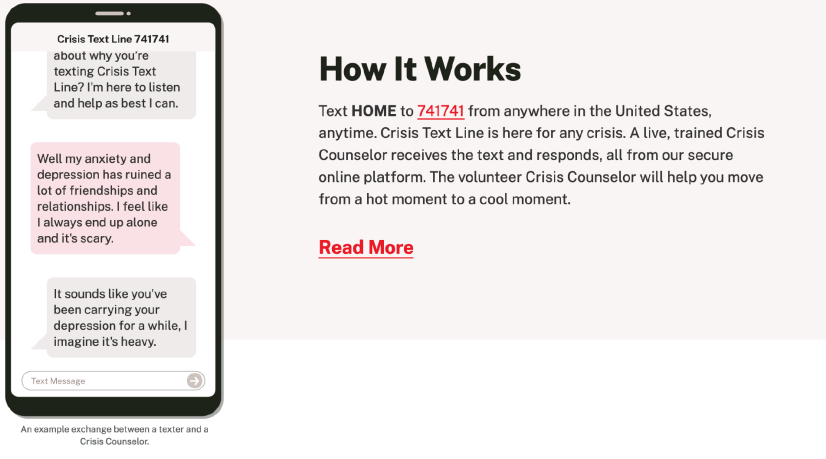
It's not all dystopian!

The White House
Office of the Press Secretary

For Immediate Release January 30, 2015

FACT SHEET: President Obama's Precision Medicine Initiative

Building on President Obama's announcement in his State of the Union Address, today the Administration is unveiling details about the Precision Medicine Initiative, a bold new research effort to revolutionize how we improve health and treat disease. Launched with a $215 million investment in the President's 2016 Budget, the Precision Medicine Initiative will pioneer a new model of patient-powered research that promises to accelerate biomedical discoveries and provide clinicians with new tools, knowledge, and therapies to select which treatments will work best for which patients.



An example exchange between a texter and a Crisis Counselor.

times. But given the growth of demand, it's critical to use data and technology to aid them. For example, the data shows the most effective conversations are between 45 and 60 messages. Or, if a texter messages in with the word "ibuprofen" they are 16 times more likely to be actively suicidal ("bridge" is 8 times and "tonight" is 3 times) and the Crisis Counselors can immediately begin a risk assessment to help de-escalate the texter.

What makes the social score and the crisis score ethically different?

Or are they the same thing?

Avoid manipulation

Think about people

Think about autonomy

Don't rely 100% on data!

Bias

ARTIFICIAL INTELLIGENCE

Predictim Claims Its AI Can Flag 'Risky' Babysitters. So I Tried It on the People Who Watch My Kids.

Brian Merchant

12/06/18 3:57PM • Filed to: AUTOMATON

109 4

At issue is the fact that I've used Predictim to scan a handful of people I very much trust with my own son. Our actual babysitter, Kianah Stover, returned a ranking of "Moderate Risk" (3 out 5) for "Disrespectfulness" for what appear to me to be innocuous Twitter jokes. She returned a worse ranking than a friend I also tested who routinely spews vulgarities, in fact. She's black, and he's white.

I tell them I am sure that they don't have a 'Do Racism' button on their program's dashboard, but wonder if systemic bias could nonetheless have entered into their datasets. Parsa says, "I absolutely agree that it's not perfect, it could be biased, it could flag things that are not really supposed to be flagged, and that's why we added the human review." But the human review let these results stand.

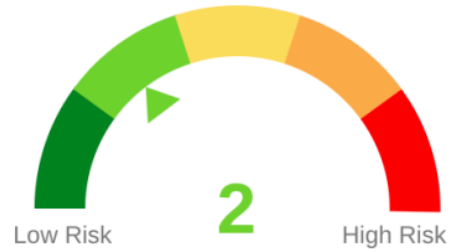
Personal Information

What does this score mean?

Kianah Jay

Scan completed on: November 27, 2018

Summary



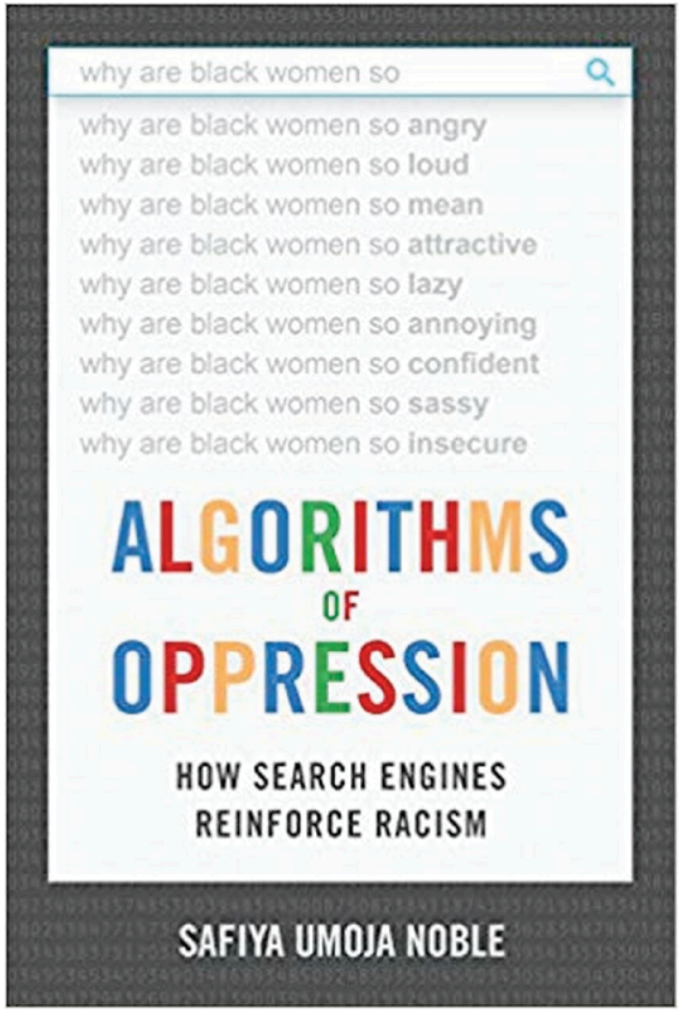
Low Risk

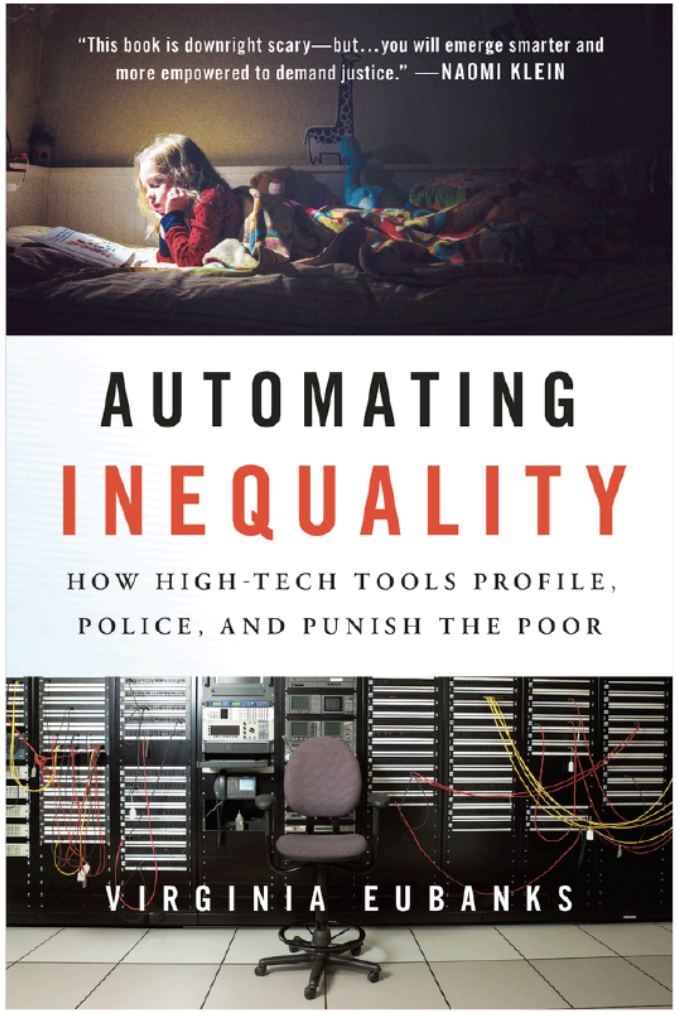
Bullying / Harassment: **2**
Disrespectful Attitude: **3**
Explicit Content: **1**
Drug Abuse: **1**

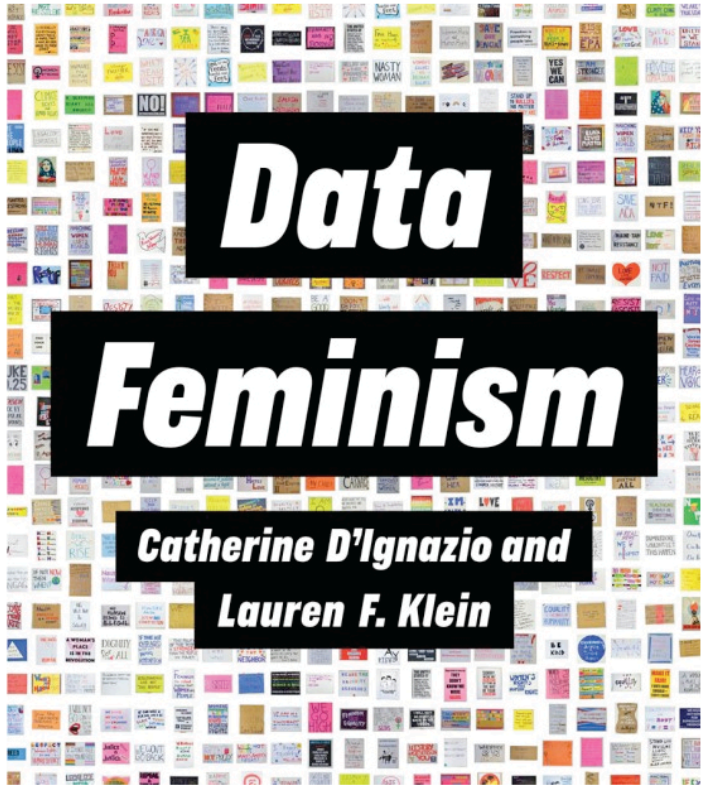
Report Summary

Initiate A New Scan

Bullying / Harassment:	Low Risk
Disrespectful Attitude:	Moderate Risk


why are black women so
why are black women so angry
why are black women so loud
why are black women so mean
why are black women so attractive
why are black women so lazy
why are black women so annoying
why are black women so confident
why are black women so sassy
why are black women so insecure
ALGORITHMS
OF
OPPRESSION
HOW SEARCH ENGINES
REINFORCE RACISM
SAFIYA UMOJA NOBLE


"This book is downright scary—but...you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN
AUTOMATING
INEQUALITY
HOW HIGH-TECH TOOLS PROFILE,
POLICE, AND PUNISH THE POOR
VIRGINIA EUBANKS


Data
Feminism
Catherine D'Ignazio and
Lauren F. Klein

After an audit of the algorithm, the resume screening company found that the algorithm found two factors to be most indicative of job performance: their name was Jared, and whether they played high school lacrosse. Girouard's client did not use the tool.

Algorithms sold to courts across the United States have been crunching those numbers since 2000. And they did so without much oversight or criticism, until *ProPublica* released an investigation showing the bias of one particular system against black defendants. The algorithm, called COMPAS, could single out those who would go on to reoffend with roughly the same accuracy for each race. But it guessed wrong about twice as often for black people. COMPAS mislabeled a person who *didn't* go on to reoffend as "high risk" almost twice as often for those individuals. And COMPAS also mistakenly assigned a higher number of "low risk" labels to white convicts who went on to commit more crimes. So the system essentially demonizes black offenders while simultaneously giving white criminals the benefit of the doubt.

MACHINE BIAS

Facebook Lets Advertisers Exclude Users by Race

Facebook's system allows advertisers to exclude black, Hispanic, and other "ethnic affinities" from seeing ads.

by Julia Angwin and Terry Parris Jr., Oct. 28, 2016, 1 p.m. EDT

UNITED STATES OF AMERICA
DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
OFFICE OF ADMINISTRATIVE LAW JUDGES

The Secretary, United States Department of Housing and Urban Development, on behalf of Complainant Assistant Secretary for Fair Housing and Equal Opportunity, Charging Party, v. Facebook, Inc., Respondent	HUD ALJ No. FHEO No. 01-18-0323-8

CHARGE OF DISCRIMINATION

I. JURISDICTION

On August 13, 2018, the Assistant Secretary for Fair Housing and Equal Opportunity ("Assistant Secretary") filed a timely complaint with the Department of Housing and Urban Development ("HUD" or the "Department") alleging that Respondent violated subsections 804(a), 804(b), 804(c) and 804(f) of the Fair Housing Act, 42 U.S.C. §§ 3601-19 ("Act"), by discriminating because of race, color, religion, sex, familial status, national origin and disability.

Marisa Robertson

Avoid bias

Make sure your sample is representative

Think about theory

Remember that NO data, models, or algorithms are neutral

Fight the algorithms

Very feebly, but still...

Incognito / private windows

adsettings.google.com

Accidental(?) evil

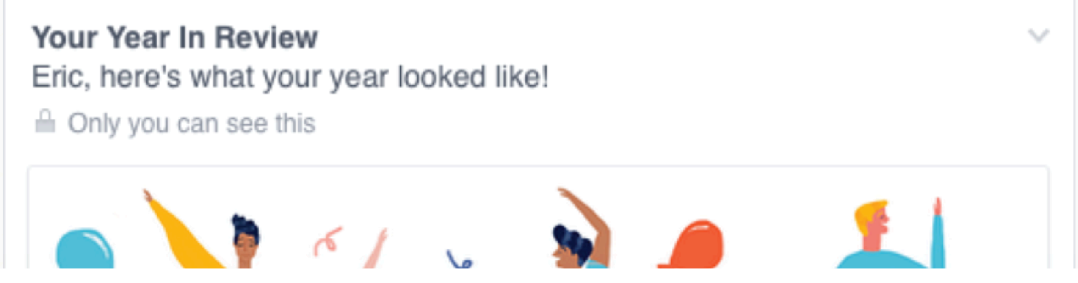
Inadvertent Algorithmic Cruelty

Published 5 years, 10 months past

I didn't go looking for grief this afternoon, but it found me anyway, and I have designers and programmers to thank for it. In this case, the designers and programmers are somewhere at Facebook.

I know they're probably pretty proud of the work that went into the "Year in Review" app they designed and developed, and deservedly so—a lot of people have used it to share the highlights of their years. Knowing what kind of year I'd had, though, I avoided making one of my own. I kept seeing them pop up in my feed, created by others, almost all of them with the default caption, "It's been a great year! Thanks for being a part of it." Which was, by itself, jarring enough, the idea that any year I was part of could be described as great.

Still, they were easy enough to pass over, and I did. Until today, when I got this in my feed, exhorting me to create one of my own. "Eric, here's what your year looked like!"



TECHNOLOGY

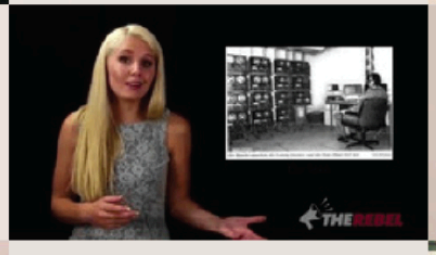
Feds Say Self-Driving Uber SUV Did Not Recognize Jaywalking Pedestrian In Fatal Crash

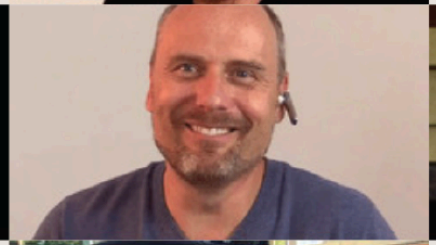
November 7, 2019 · 10:57 PM ET

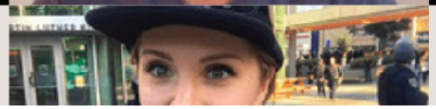
RICHARD GONZALES

The New York Times

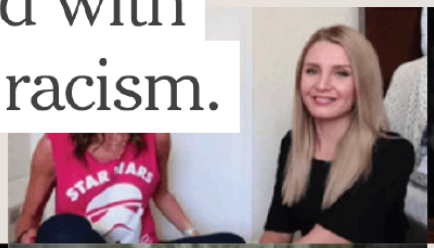
Older People Shared Fake News on Facebook More Than Others in 2016 Race, Study Says

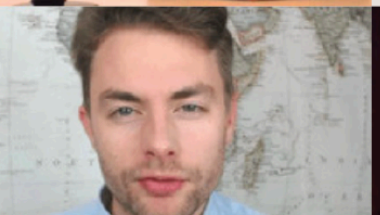






Soon, he was pulled into a far-right universe, watching thousands of videos filled with conspiracy theories, misogyny and racism.





GOAL
Recommend videos

CBS News ✓
@CBSNews

A geneticist at Harvard Medical School is working on a dating app that matches users based on their DNA. The goal: to eliminate all genetic diseases. @60Minutes reports, tonight cbsn.ws/2s5JWo8

wombot 👀
@colourmeamused_

o no no

10:44 AM · Sep 24, 2020 · Twitter Web App

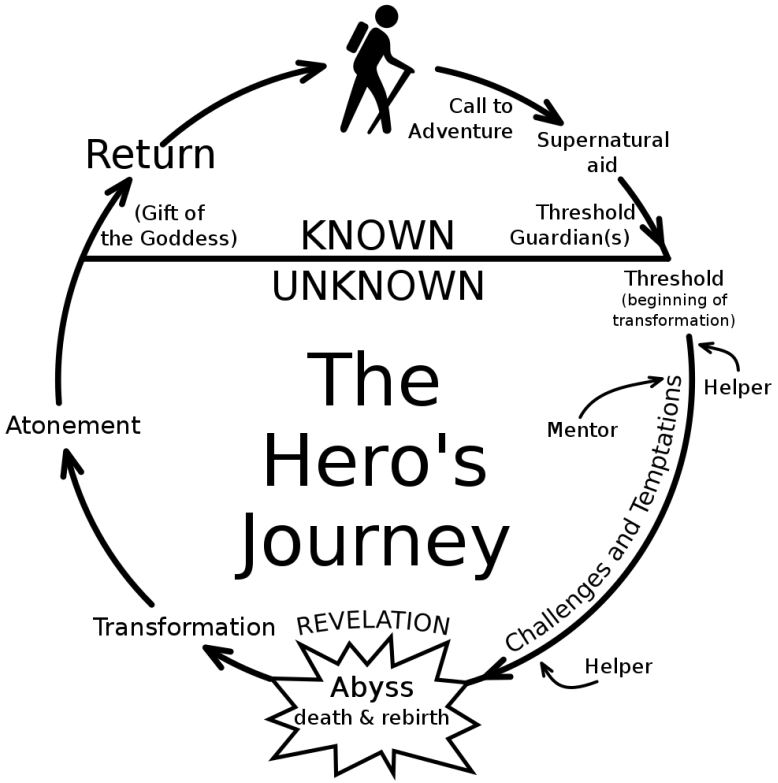
Ethics of storytelling

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

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

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.**

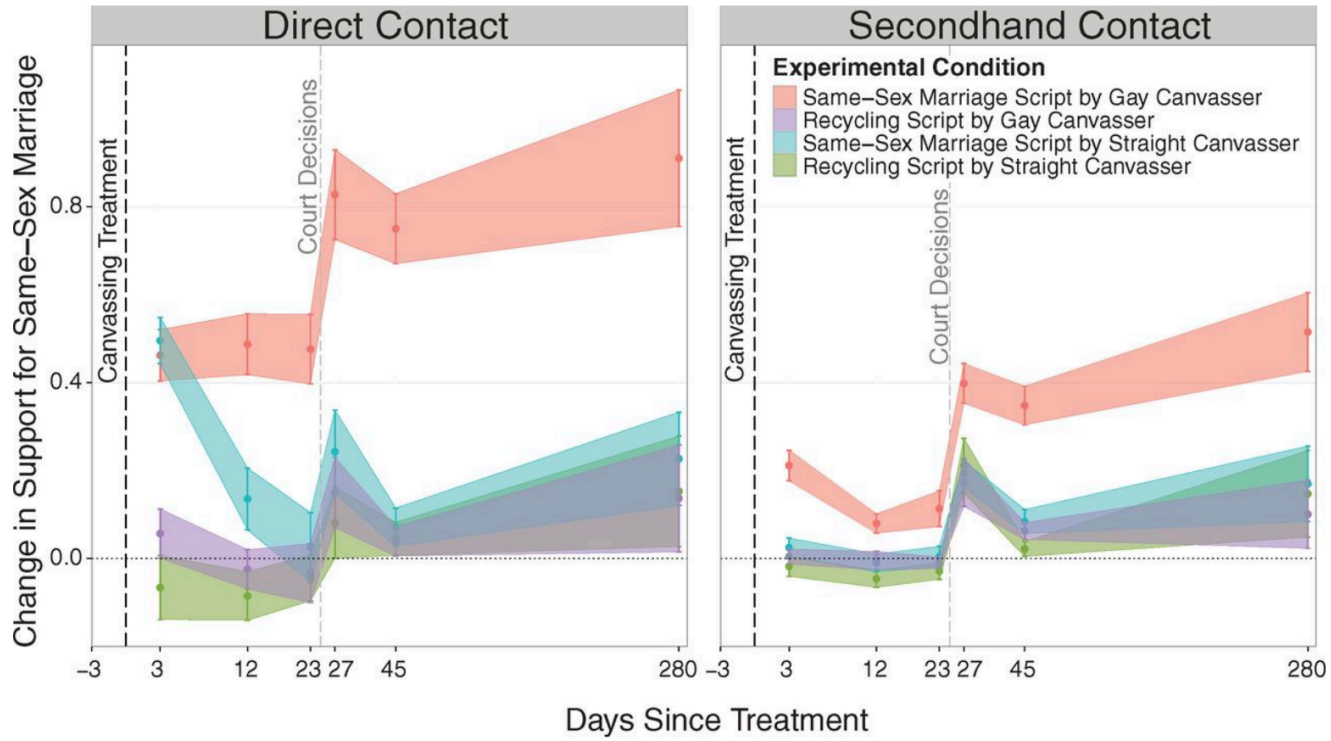
The iPad thing was LaCour's trademark. "He was sort of famous for taking his results from different studies he was working on, putting them on an iPad, and buttonholing people at the conferences and going over all of the research that he was doing, the different findings he had, and basically not letting 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

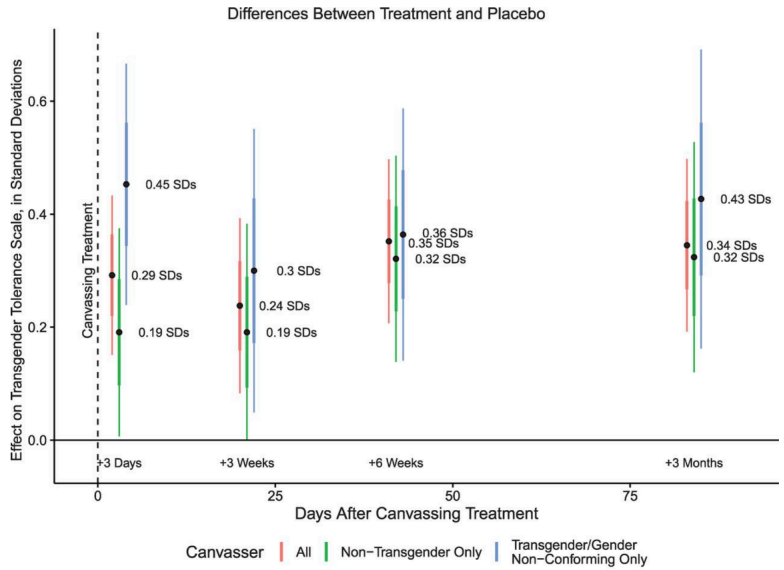
May 19, 2015

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

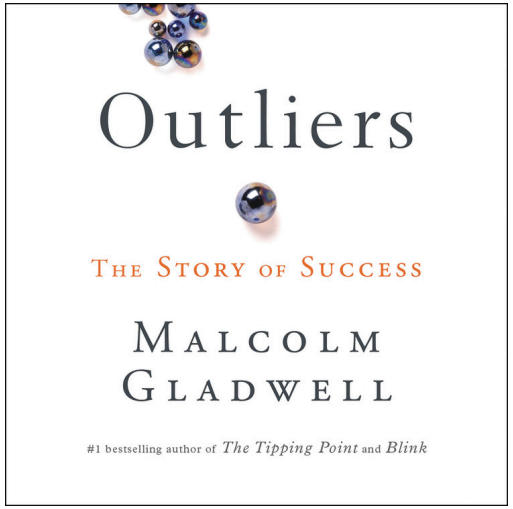


Manipulation

Don't lie

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

Misrepresentation



10,000 hours

"the magic number of greatness"

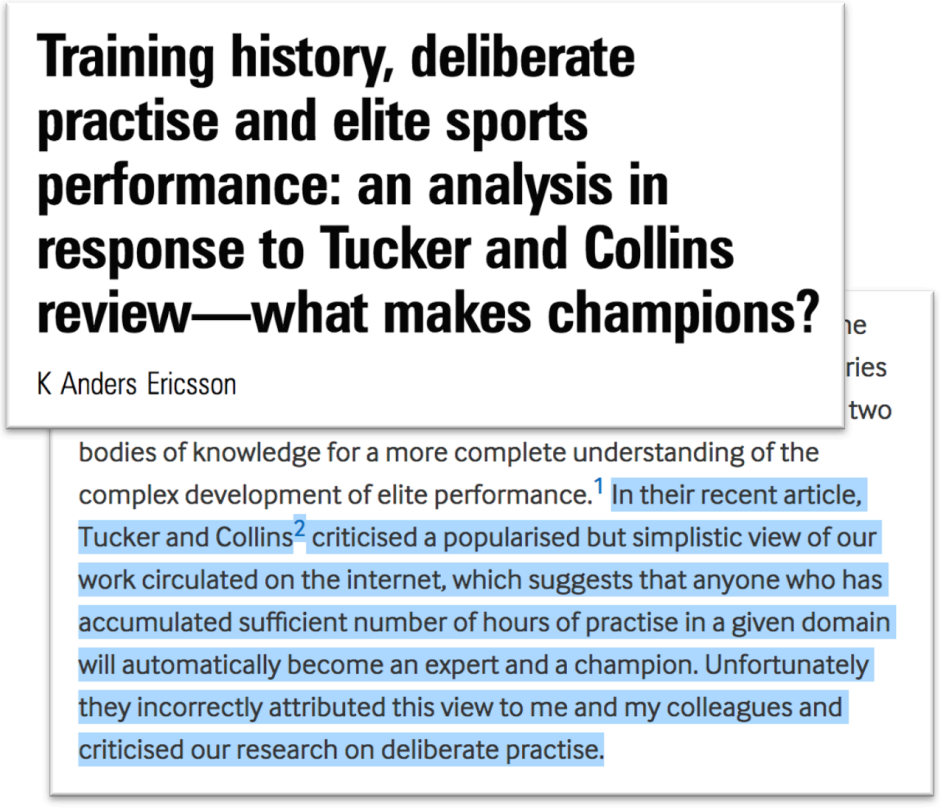
Psychological Review
1993, Vol. 100. No. 3, 363-406

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

Equity

Dumbing down vs. translation

Neuroscientist Explains One Concept in 5 Levels of Difficulty | WIRED

Translation

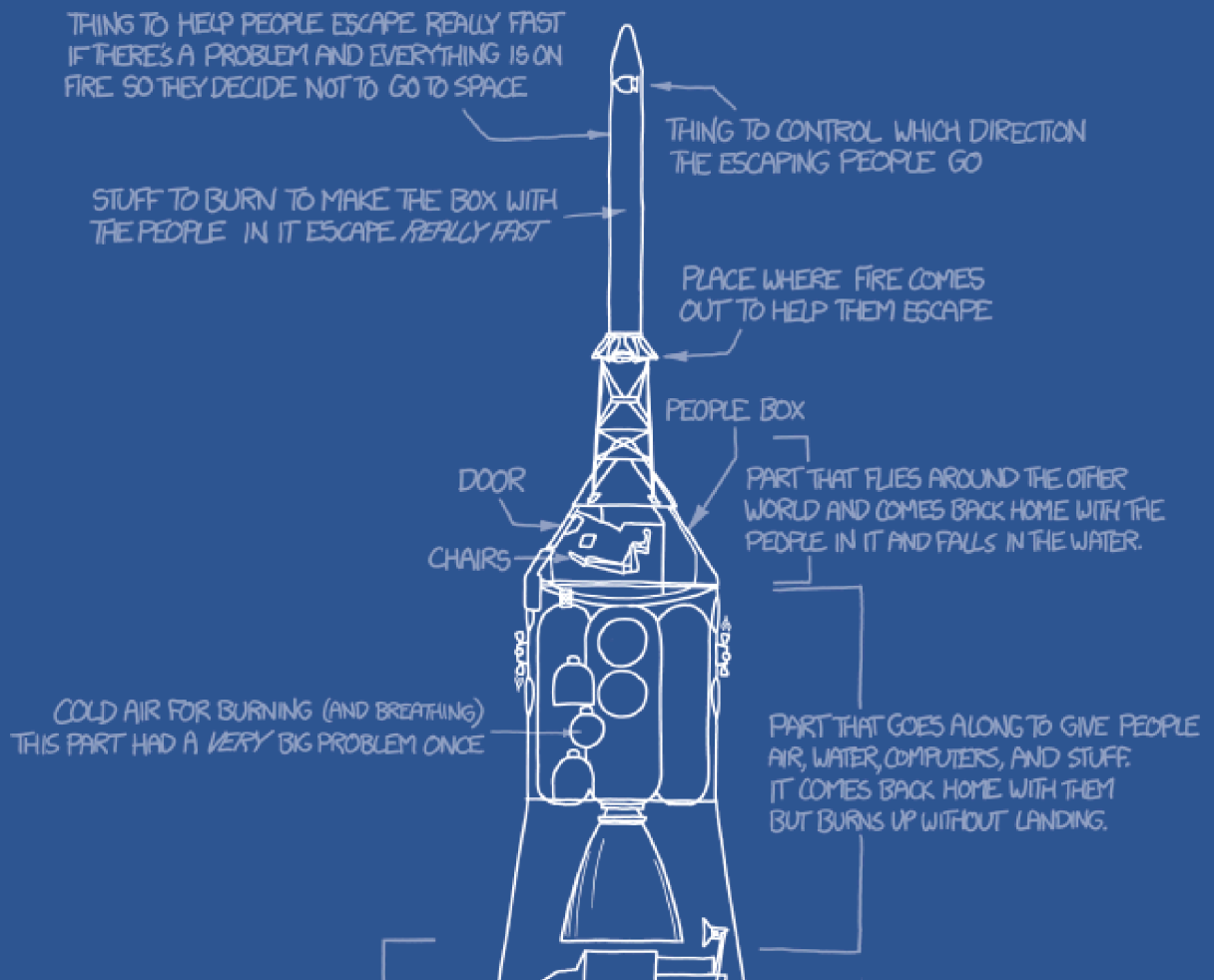
"...the task of the translator consists in finding that intended effect upon the language into which he is translating which produces in it the echo of the original"

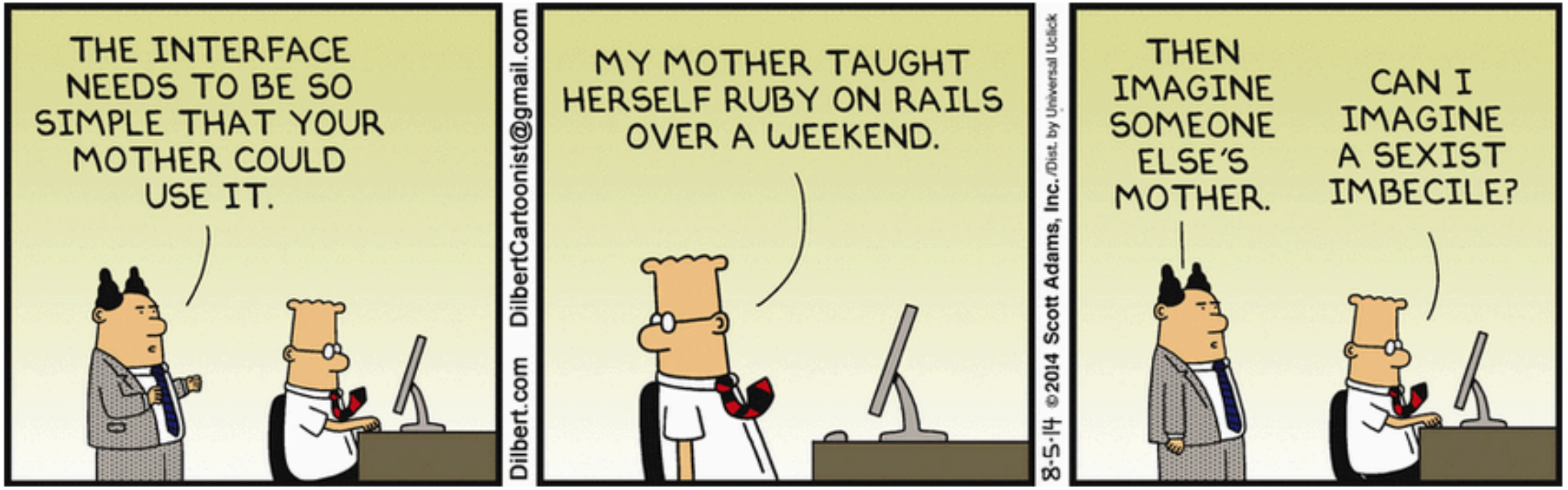
Walter Benjamin,
The Task of the Translator

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

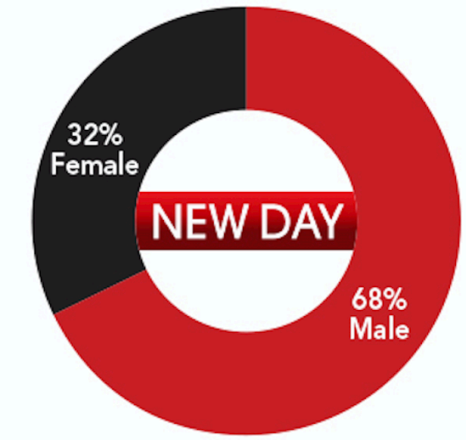
OPEN ACCESS PEER-REVIEWE

RESEARCH ARTICLE

On the Compliance of Women Engineers with a Gendered Scientific System

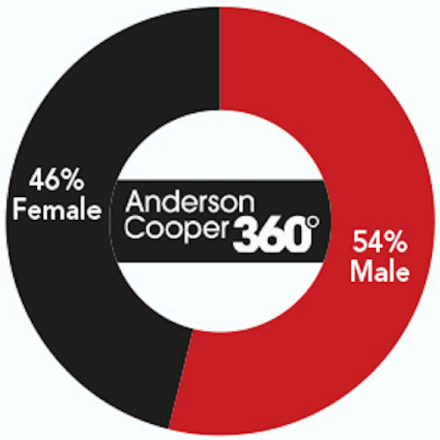
Gita Ghiasi, Vincent Larivière, Cassidy R. Sugimoto

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



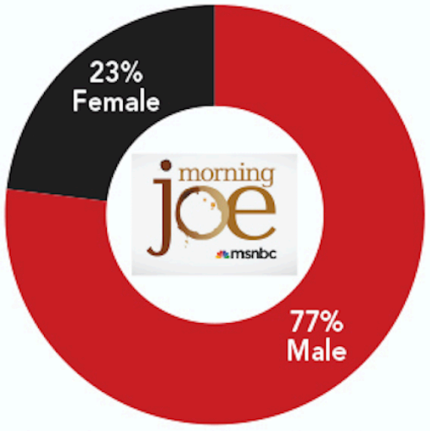
Male analysts (2486) Female analysts (1151)

CNN



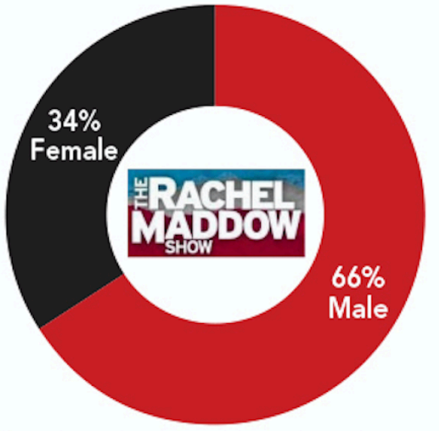
Male analysts (1216) Female analysts (1022)

CNN



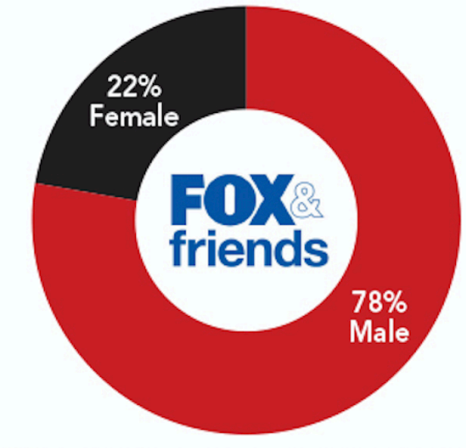
Male analysts (4121) Female analysts (1255)

msnbc



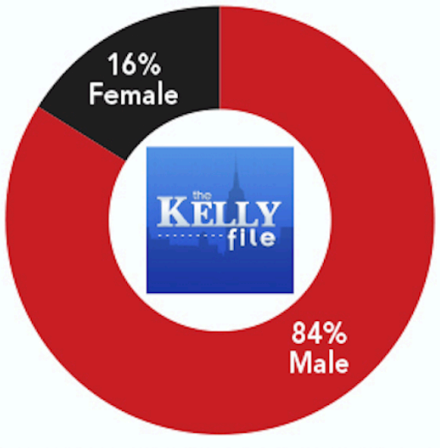
Male analysts (151) Female analysts (79)

msnbc



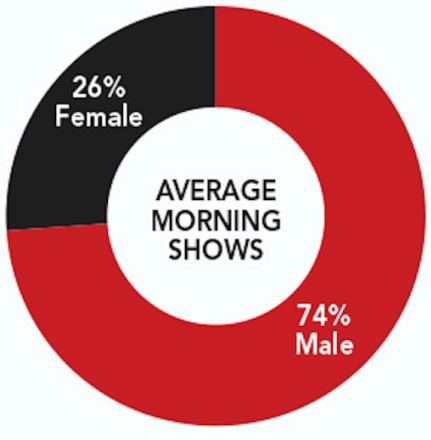
Male analysts (1313) Female analysts (373)

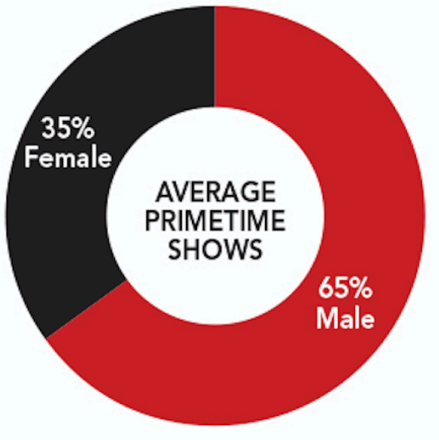
FOX



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

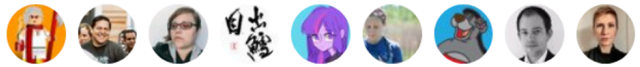
Katie Mack
@AstroKatie

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

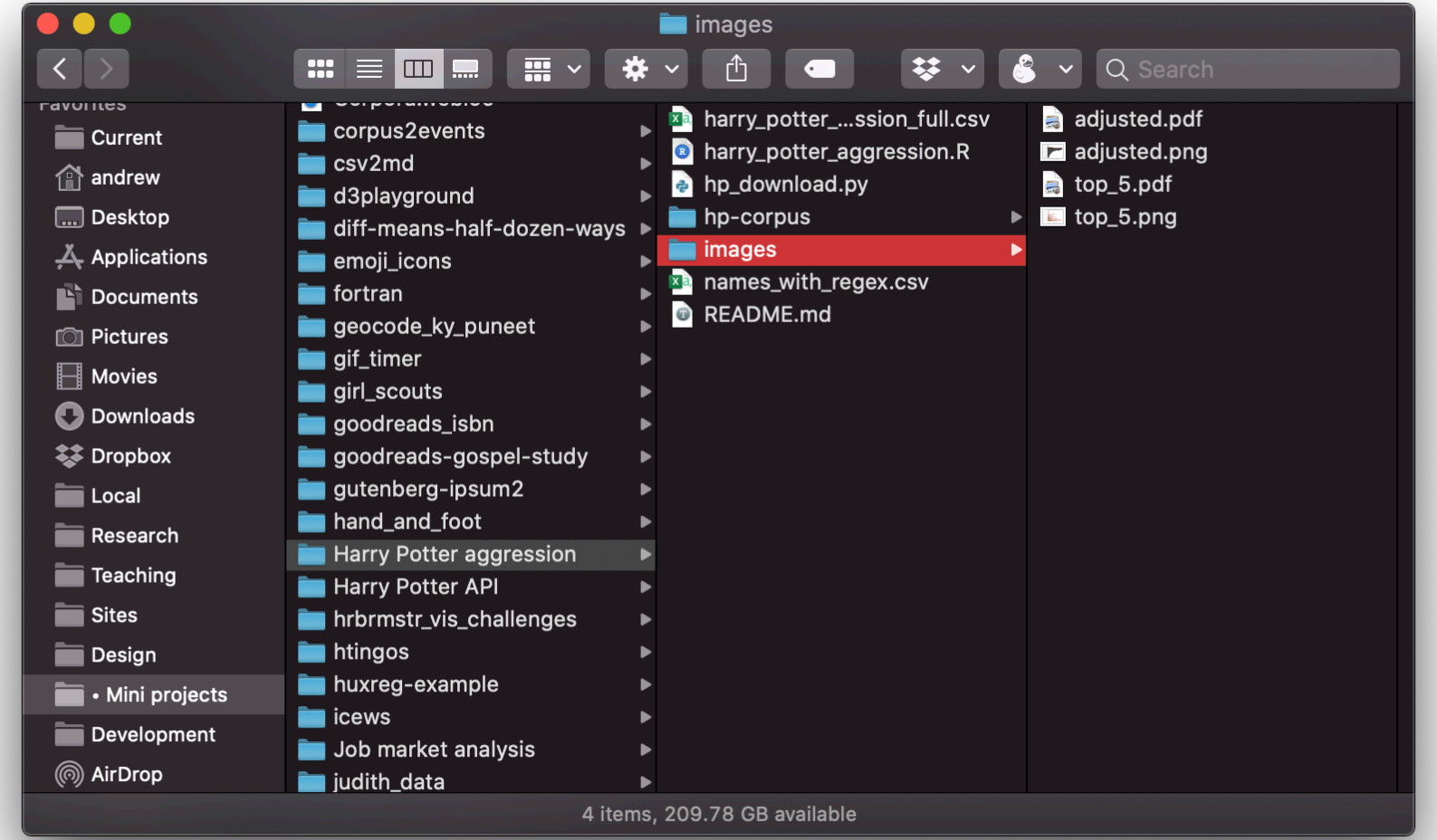
Playing with R

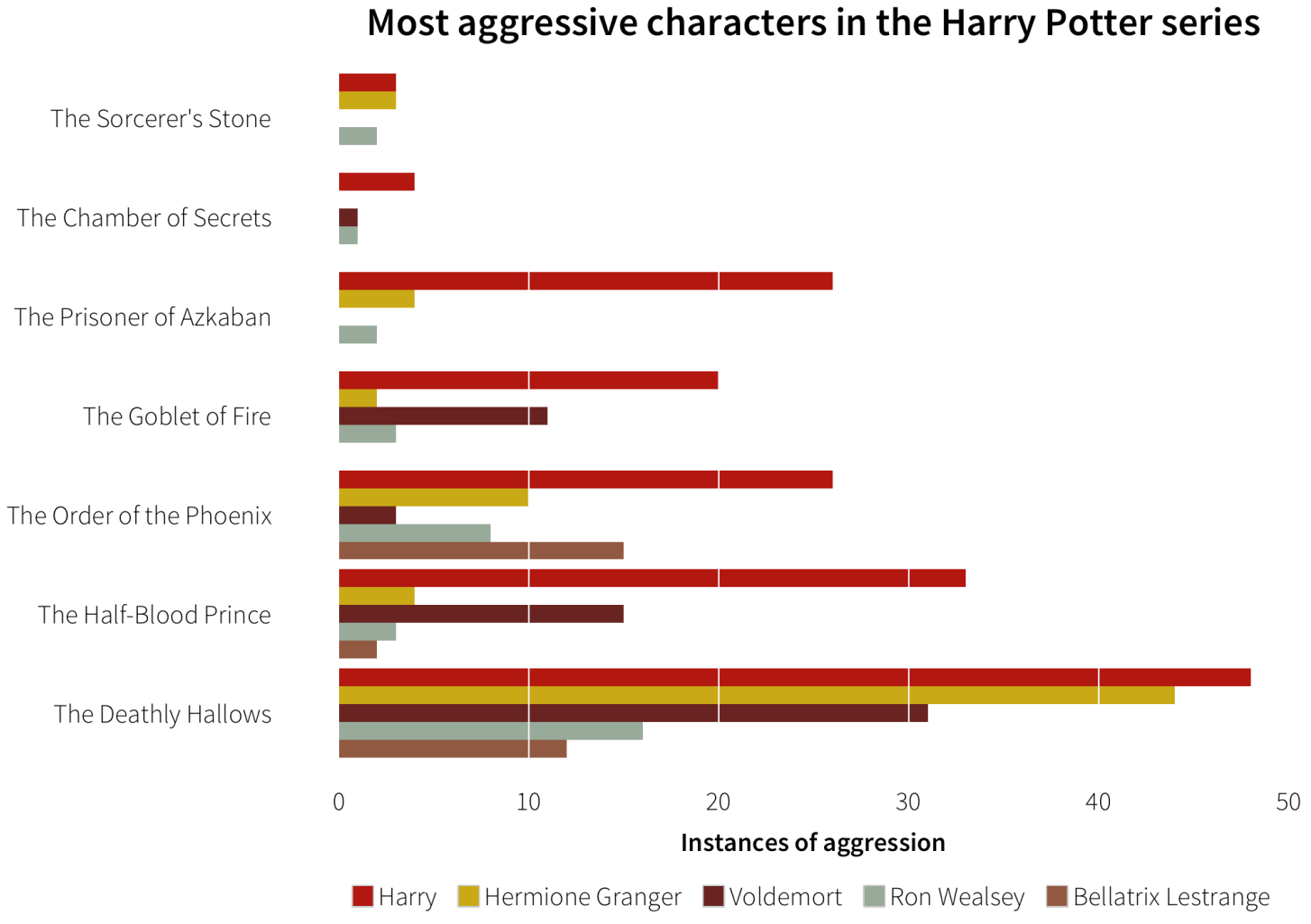
Little exploration projects

#TidyTuesday

Data play time

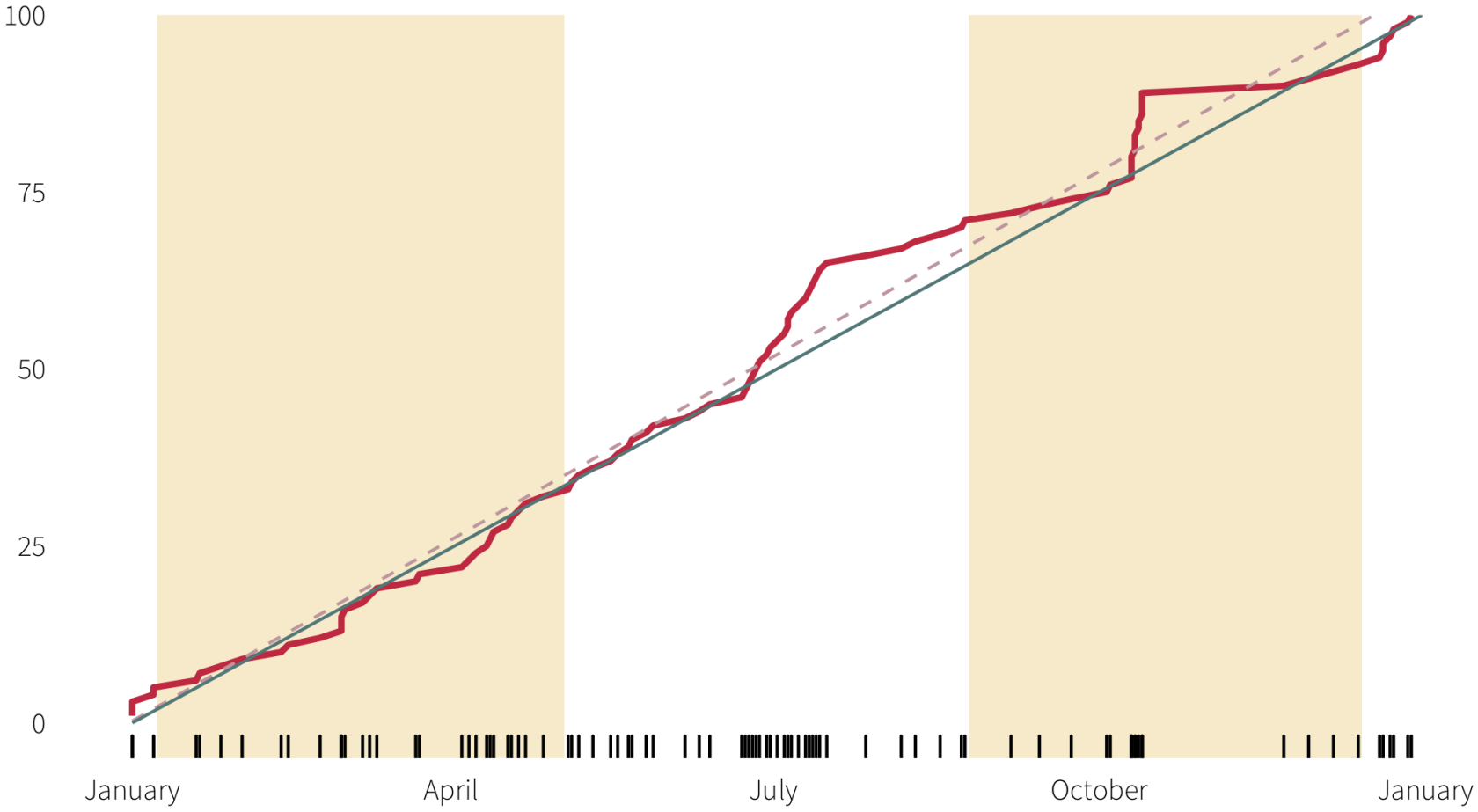
Actual projects

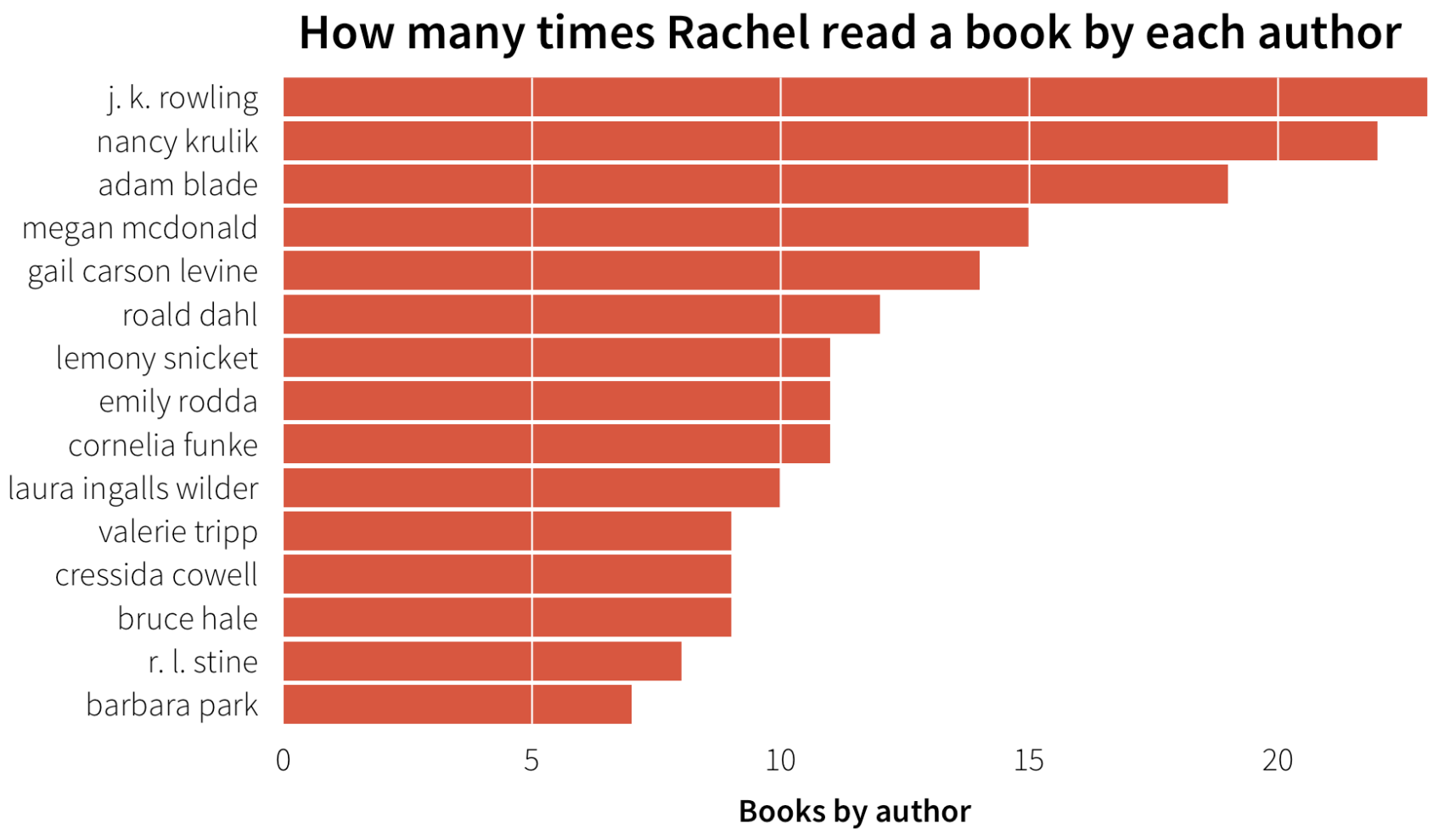

images
Search
Favorites
Current
andrew
Desktop
Applications
Documents
Pictures
Movies
Downloads
Dropbox
Local
Research
Teaching
Sites
Design
• Mini projects
Development
AirDrop
corpus2events
csv2md
d3playground
diff-means-half-dozen-ways
emoji_icons
fortran
geocode_ky_puneet
gif_timer
girl_scouts
goodreads_isbn
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

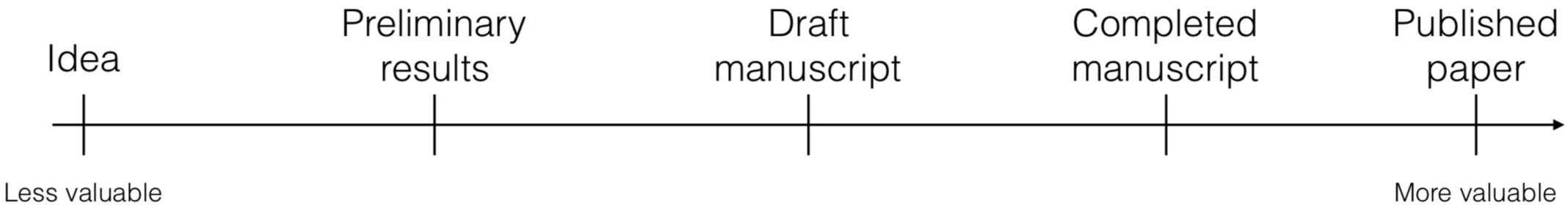
Duke semesters shaded in yellow




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 and 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)

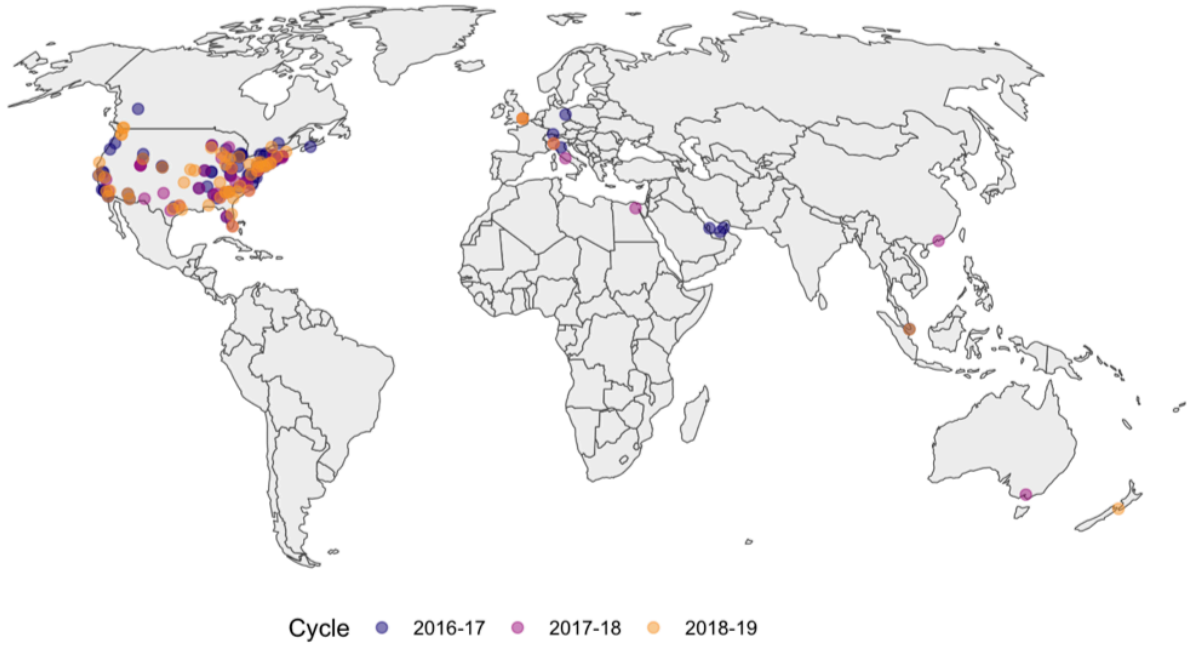
2018-19

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

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.

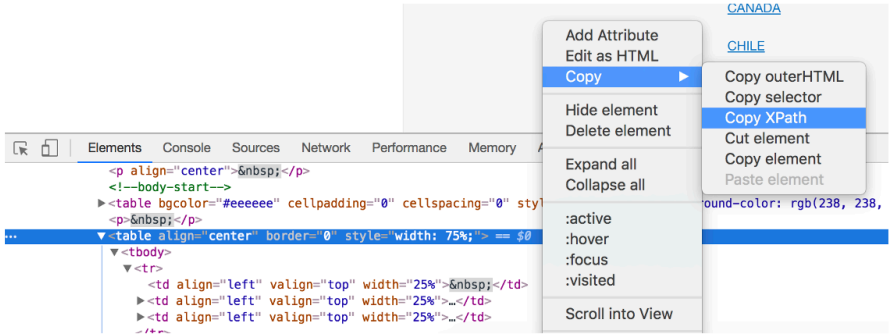
But is that bad? How does that compare to other advanced democracies, like countries in the OECD?

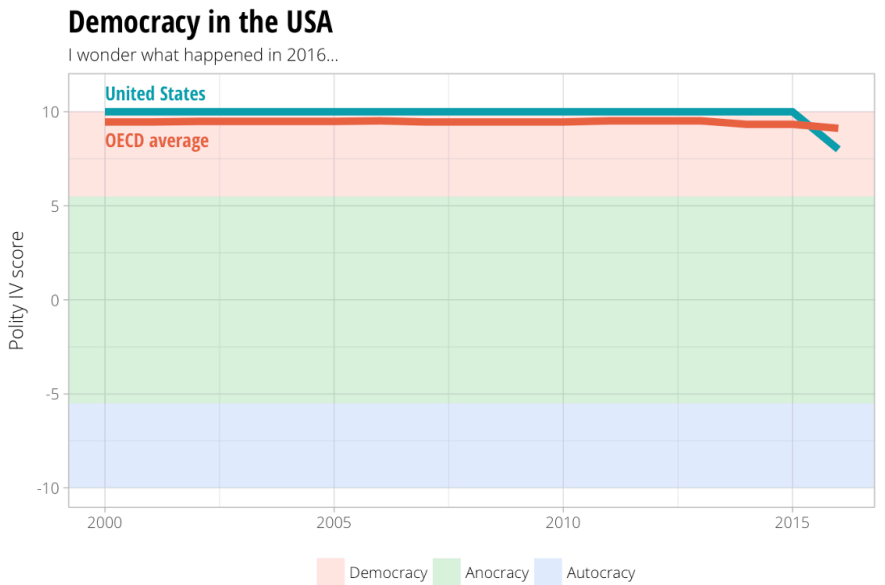
What follows below shows how relatively easy it is to quickly and reproducibly grab the new data, graph it, and compare scores across countries. (This notebook is also in a GitHub repository.)

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

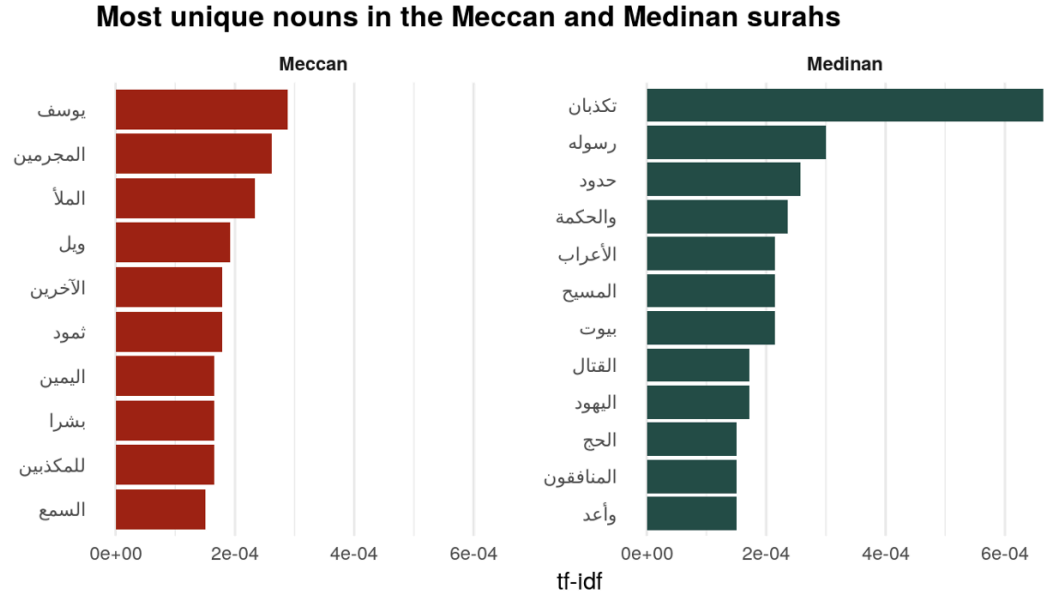
Friday, December 28, 2018

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.



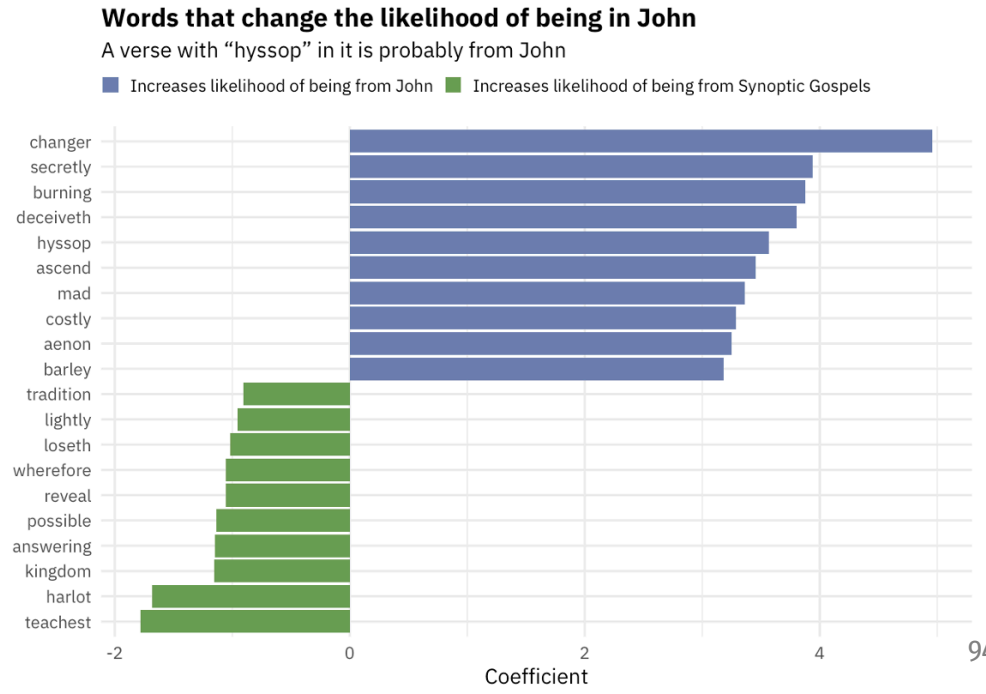
Wednesday, December 26, 2018

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

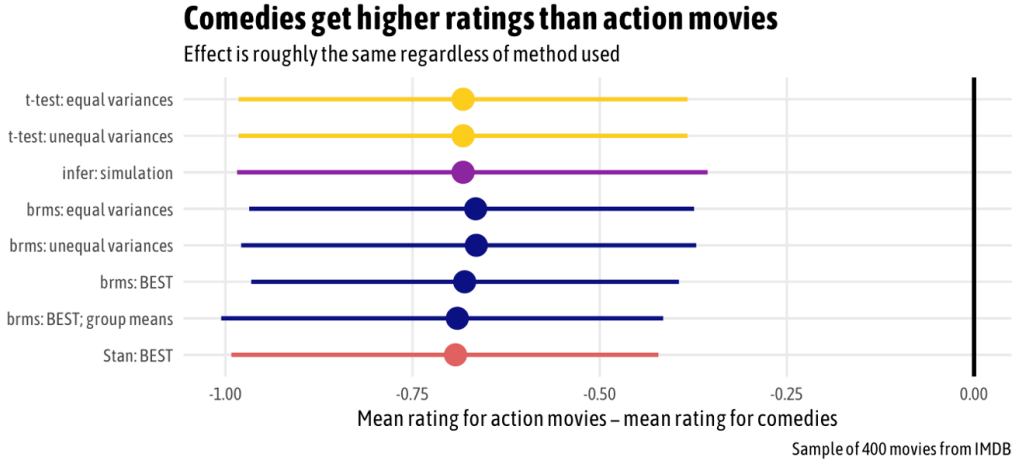
Half a dozen frequentist and Bayesian ways to measure the difference in means in two groups

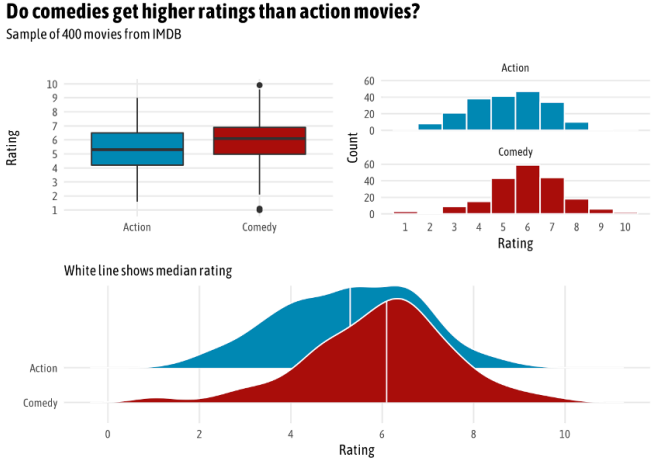
(See this notebook on GitHub)

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:)





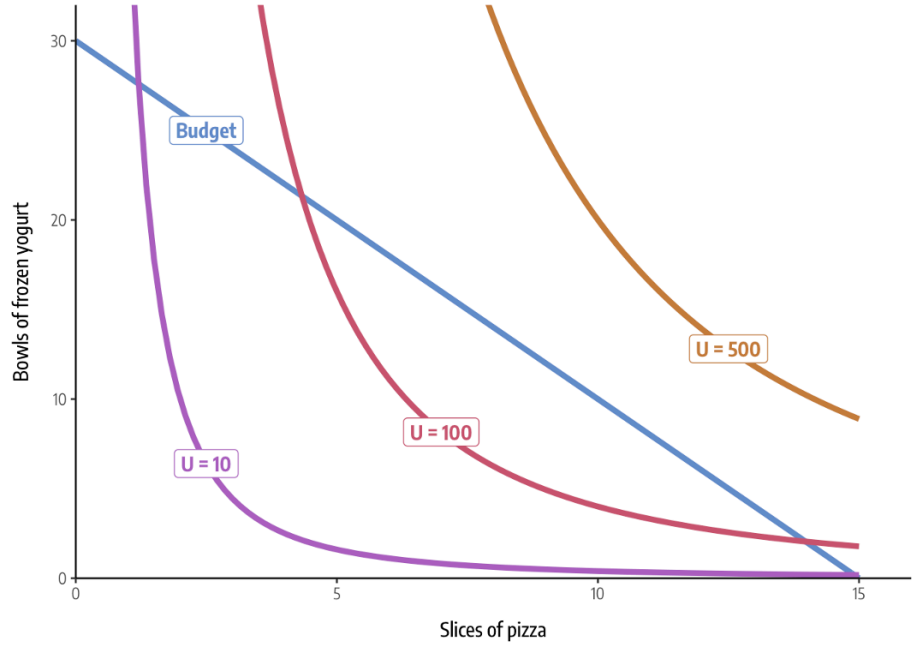
Saturday, February 16, 2019

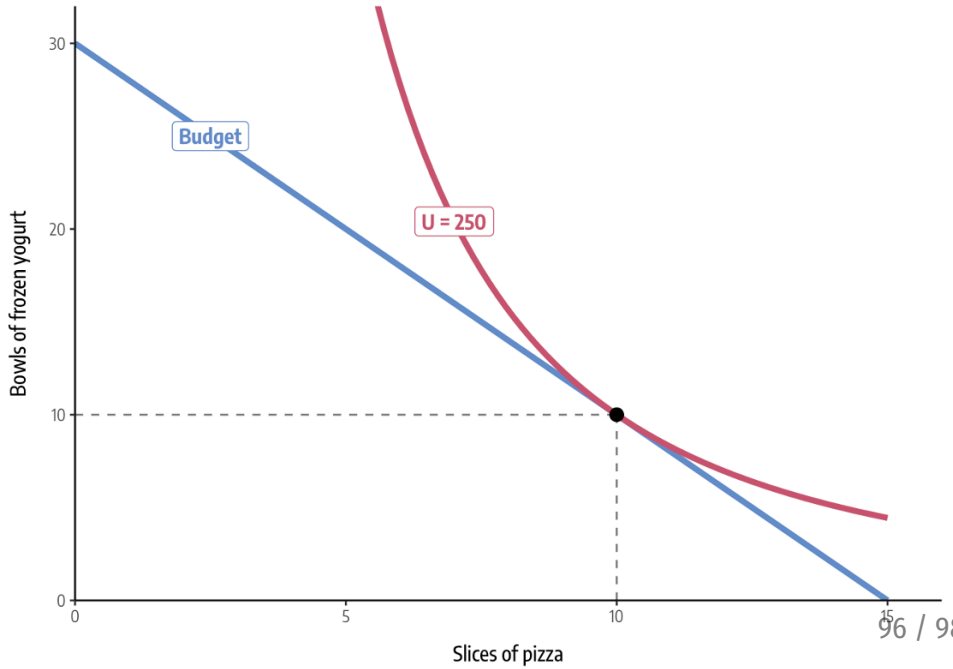
Chidi's budget and utility: doing algebra and calculus with R and yacas

(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 correctly find causal effects!