WRAPPING UP & GOING BEYOND THE BASICS

MPA 630: Data Science for Public Management

December 13, 2018

Fill out your reading report on Learning Suite

PLAN FOR TODAY

What the heck did we just learn?

Going beyond the basics

Ethics of data

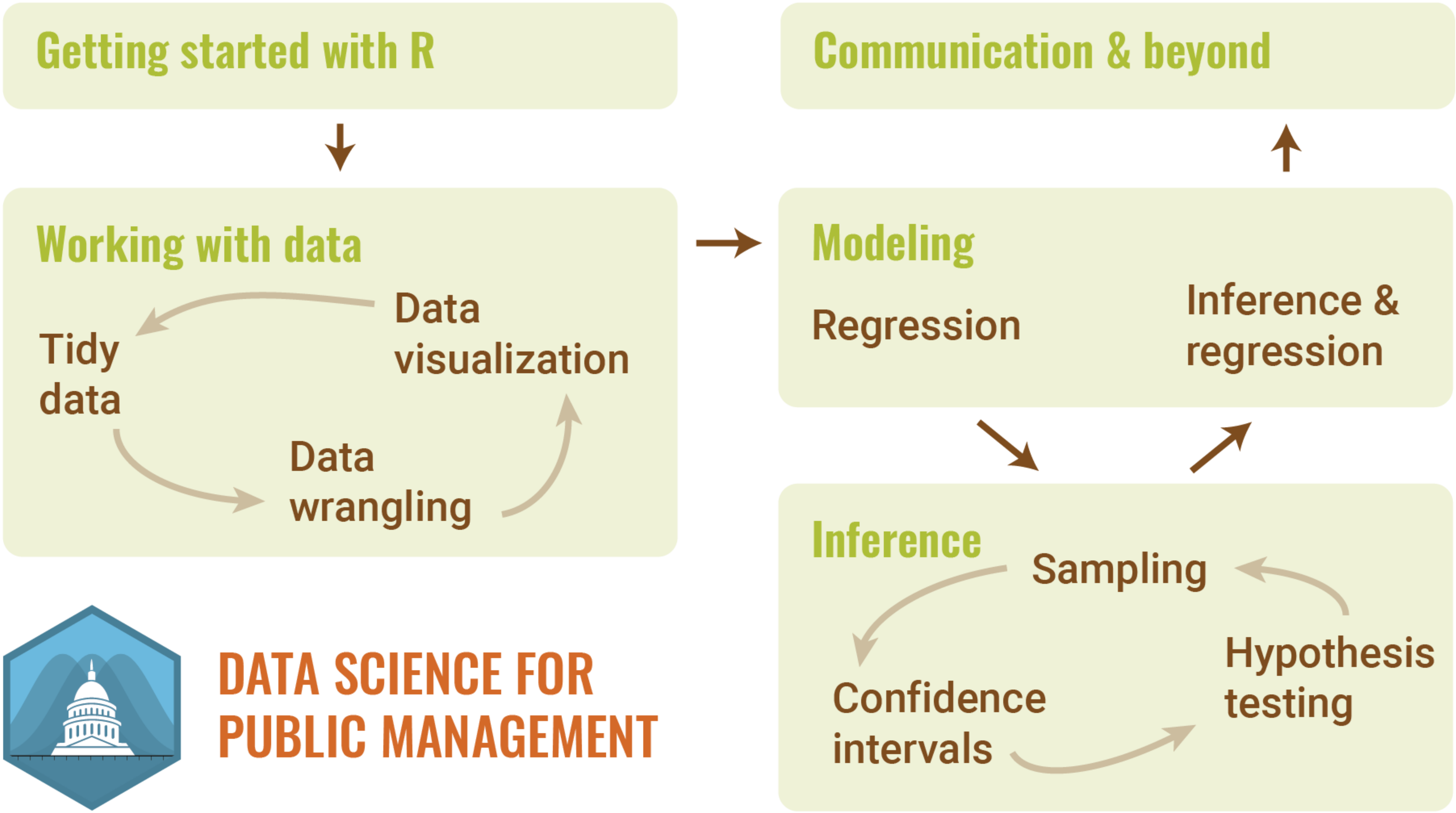
Curiosity

WHAT THE HECK DID WE JUST LEARN?

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$



Getting started with R

Working with data

Data visualization

Tidy data

Data wrangling

Modeling

Regression

Inference & regression

Inference

Sampling

Confidence intervals

Hypothesis testing

Communication & beyond

DATA SCIENCE FOR PUBLIC MANAGEMENT

WHAT IS "DATA SCIENCE"?

Turning raw data into understanding, insight, and knowledge

Collect

Analyze

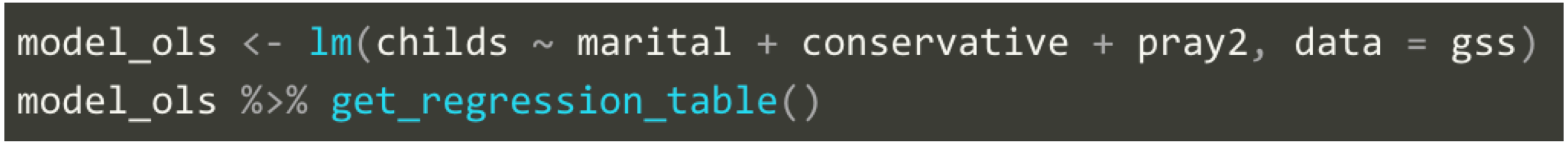
Communicate

GOING BEYOND THE BASICS

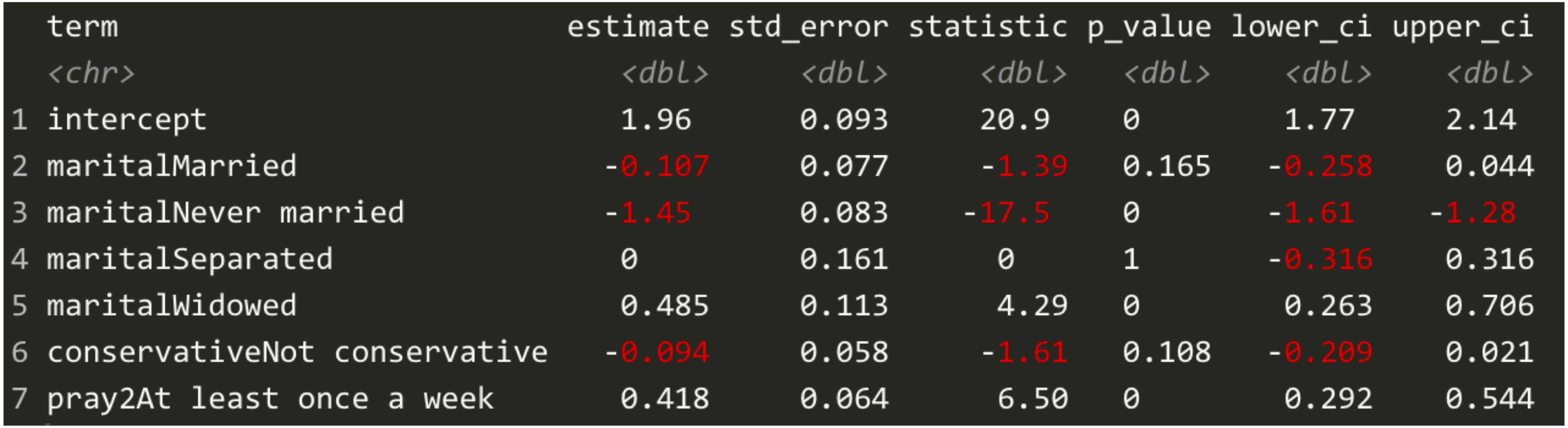
FLAVORS OF REGRESSION

Linear regression (OLS)	Y is numeric
Logistic regression	Y is 2 categories
Ordered logistic regression	Y is 3+ categories

LINEAR REGRESSION



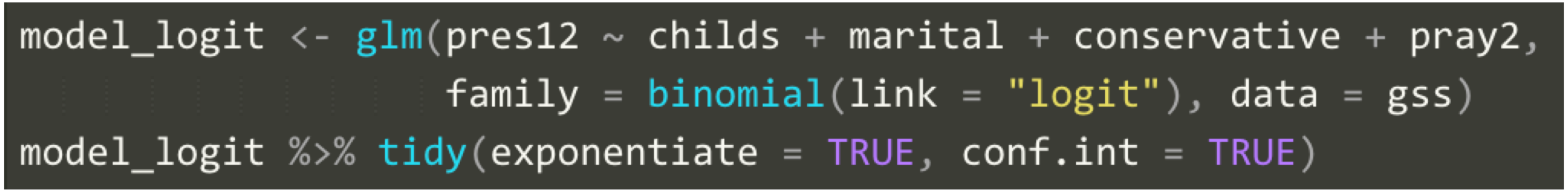
```
model_ols <- lm(childs ~ marital + conservative + pray2, data = gss)
model_ols %>% get_regression_table()
```



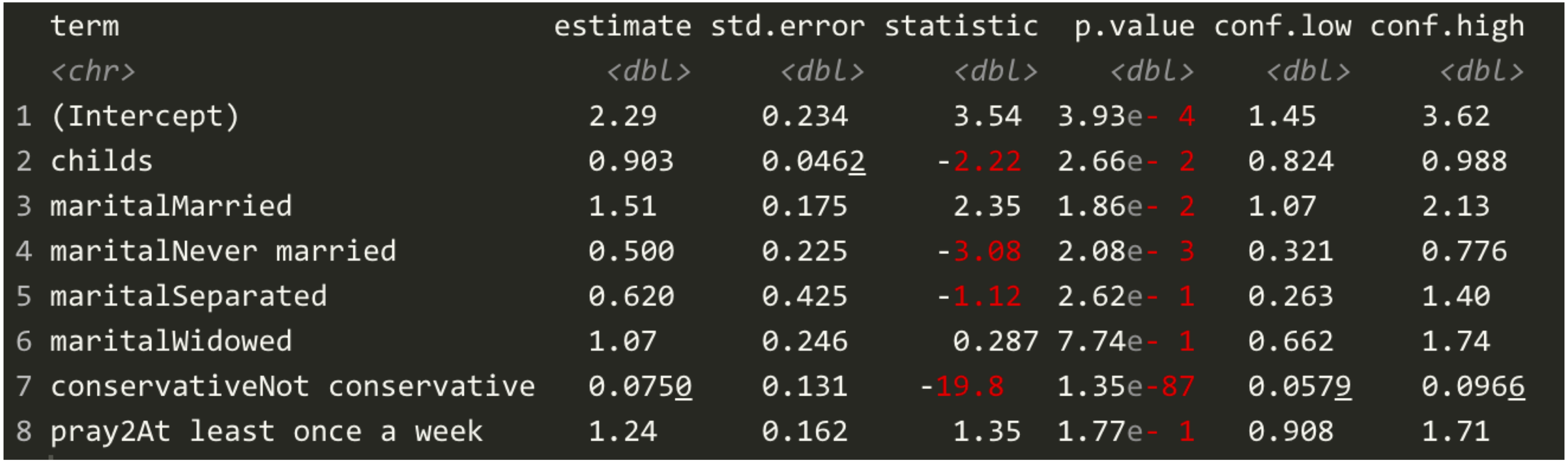
```
  term                          estimate std_error statistic p_value lower_ci upper_ci
  <chr>                            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
1 intercept                        1.96      0.093    20.9     0        1.77     2.14 
2 maritalMarried                  -0.107     0.077    -1.39    0.165   -0.258    0.044
3 maritalNever married            -1.45      0.083   -17.5     0       -1.61    -1.28 
4 maritalSeparated                 0         0.161     0       1       -0.316    0.316
5 maritalWidowed                   0.485     0.113     4.29    0        0.263    0.706
6 conservativeNot conservative    -0.094     0.058    -1.61    0.108   -0.209    0.021
7 pray2At least once a week        0.418     0.064     6.50    0        0.292    0.544
```

For every 1 unit increase in X, there’s a β change in Y

LOGISTIC REGRESSION



```r
model_logit <- glm(pres12 ~ childs + marital + conservative + pray2,
                   family = binomial(link = "logit"), data = gss)
model_logit %>% tidy(exponentiate = TRUE, conf.int = TRUE)
```



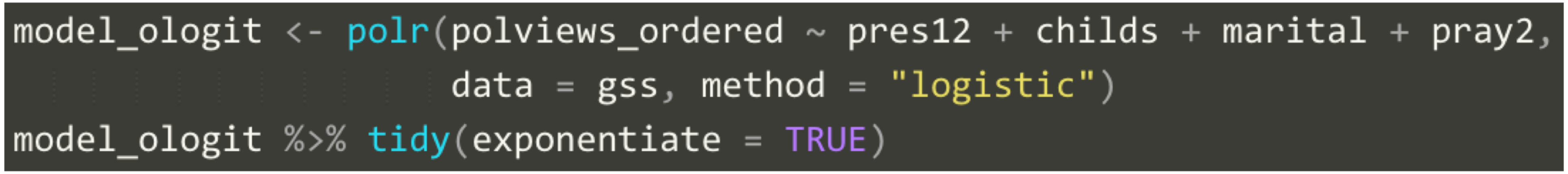
```
  term                          estimate std.error statistic  p.value conf.low conf.high
  <chr>                            <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)                     2.29      0.234      3.54  3.93e- 4   1.45      3.62
2 childs                          0.903     0.0462    -2.22  2.66e- 2   0.824     0.988
3 maritalMarried                  1.51      0.175      2.35  1.86e- 2   1.07      2.13
4 maritalNever married            0.500     0.225     -3.08  2.08e- 3   0.321     0.776
5 maritalSeparated                0.620     0.425     -1.12  2.62e- 1   0.263     1.40
6 maritalWidowed                  1.07      0.246      0.287 7.74e- 1   0.662     1.74
7 conservativeNot conservative    0.0750    0.131    -19.8   1.35e-87   0.0579    0.0966
8 pray2At least once a week       1.24      0.162      1.35  1.77e- 1   0.908     1.71
```

For every 1 unit increase in X, there's a β% change in the probability of Y happening

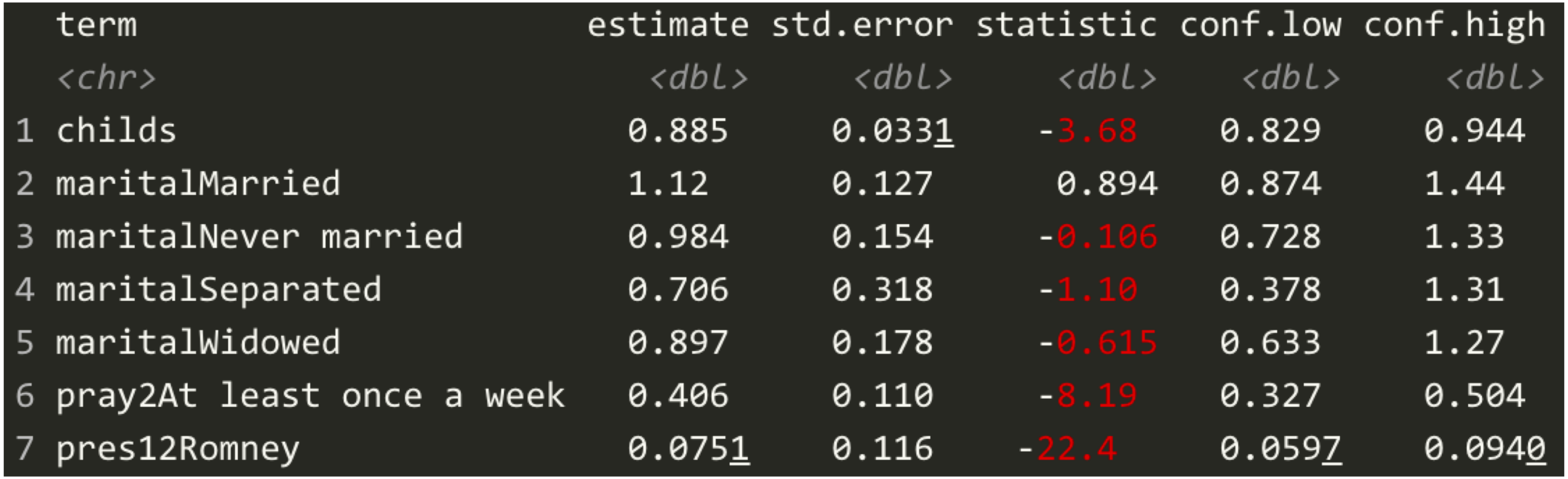
Odds ratios

Centered around 1

ORDERED LOGISTIC REGRESSION



```r
model_ologit <- polr(polviews_ordered ~ pres12 + childs + marital + pray2,
                     data = gss, method = "logistic")
model_ologit %>% tidy(exponentiate = TRUE)
```



```
  term                       estimate std.error statistic conf.low conf.high
  <chr>                         <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
1 childs                       0.885     0.0331    -3.68    0.829     0.944
2 maritalMarried               1.12      0.127      0.894   0.874     1.44
3 maritalNever married         0.984     0.154     -0.106   0.728     1.33
4 maritalSeparated             0.706     0.318     -1.10    0.378     1.31
5 maritalWidowed               0.897     0.178     -0.615   0.633     1.27
6 pray2At least once a week    0.406     0.110     -8.19    0.327     0.504
7 pres12Romney                 0.0751    0.116    -22.4     0.0597    0.0940
```

For every 1 unit increase in X, there's a β% change in the probability of moving to next level of Y

Odds ratios

Centered around 1

MORE WITH R

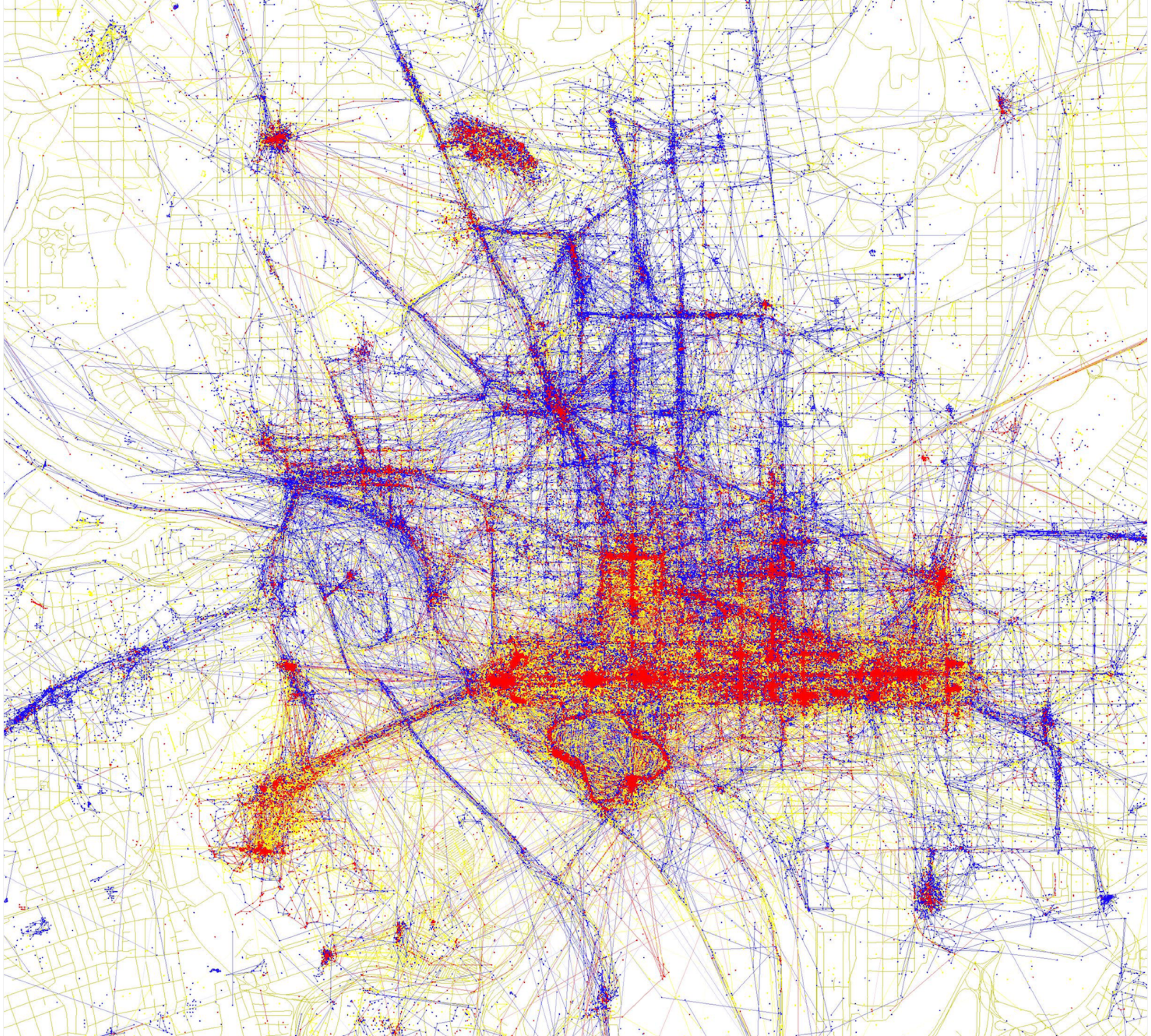
Geography

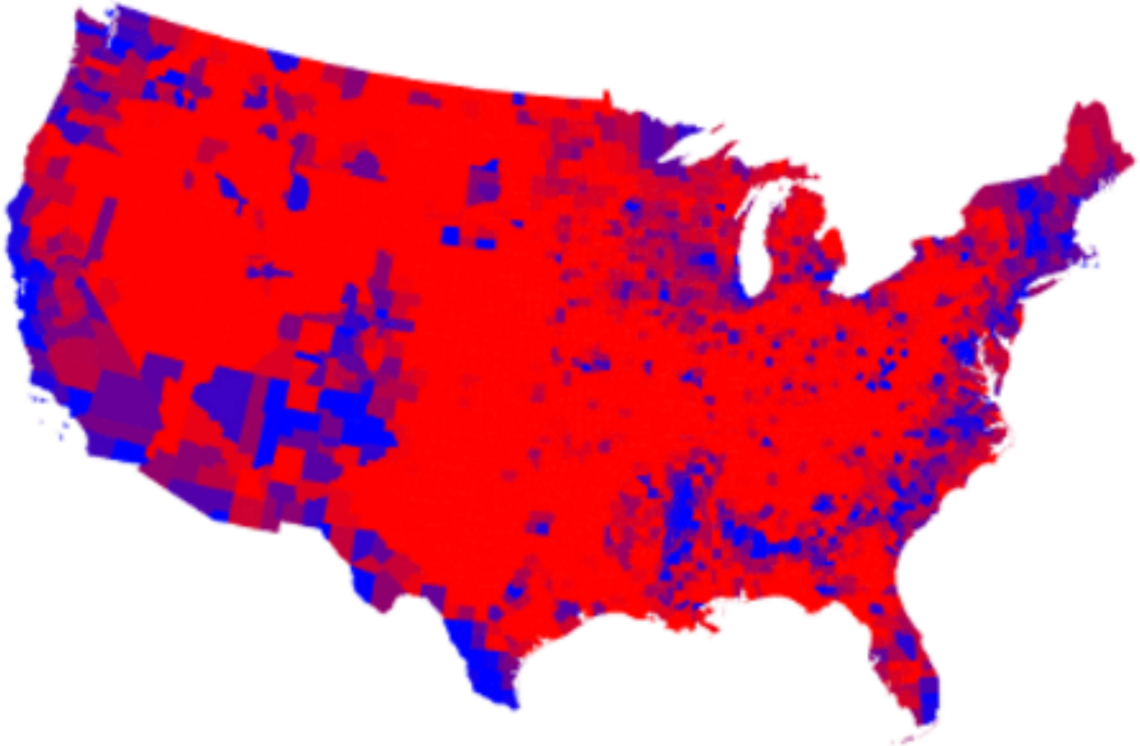
Text

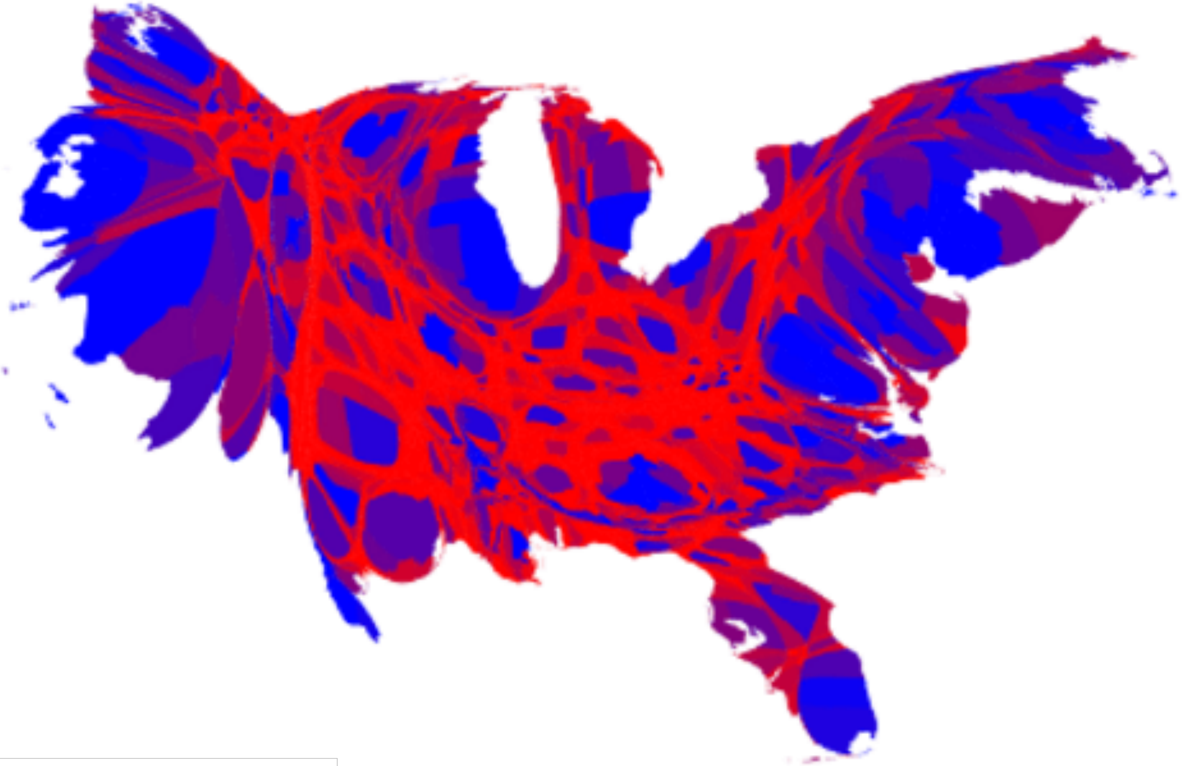
Interactivity

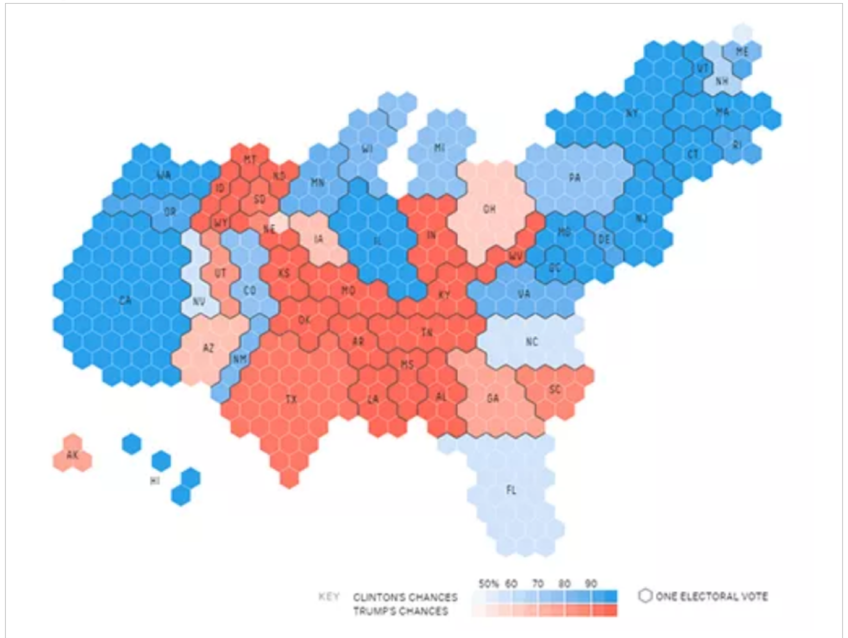
Dashboards

Sharing R Markdown



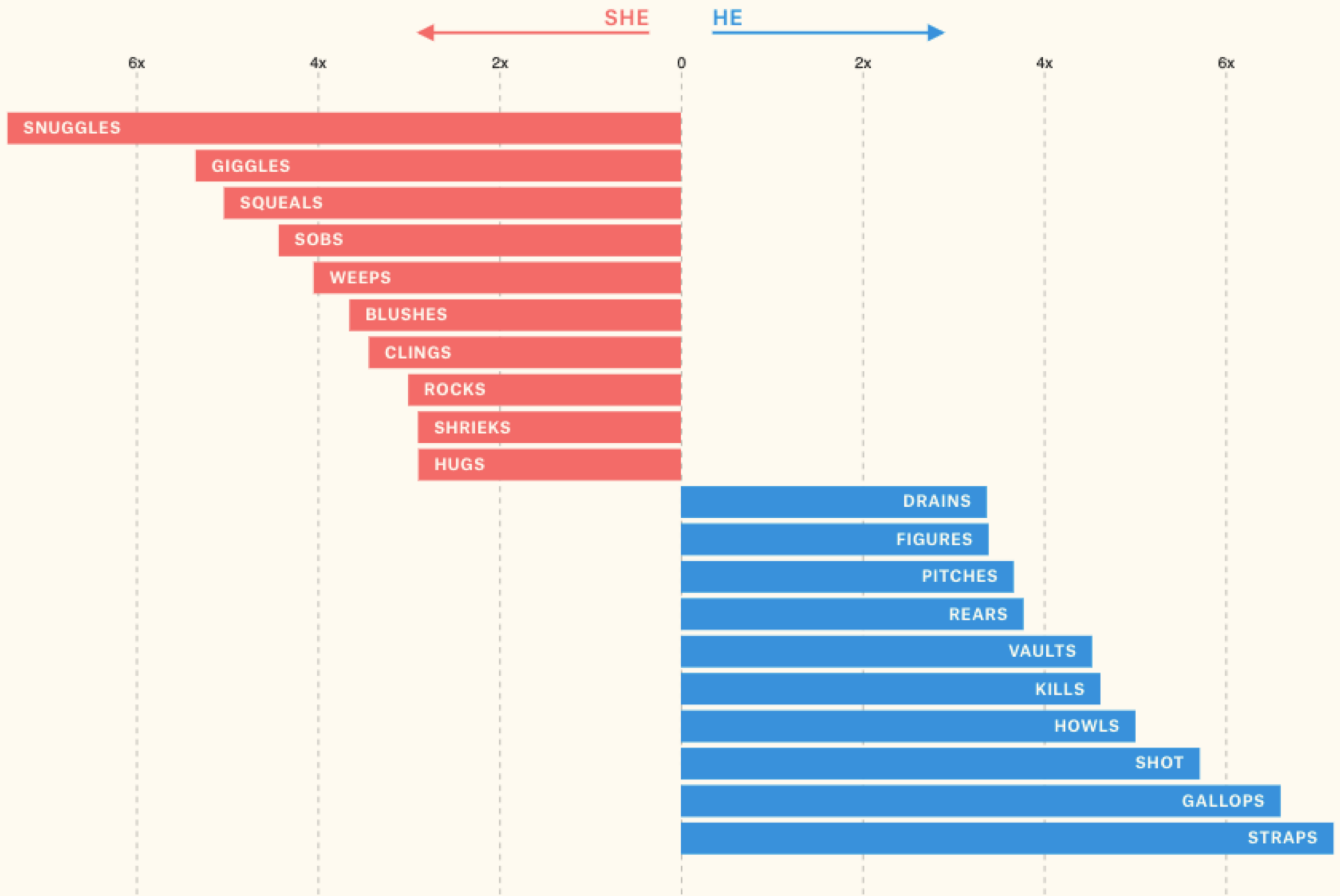


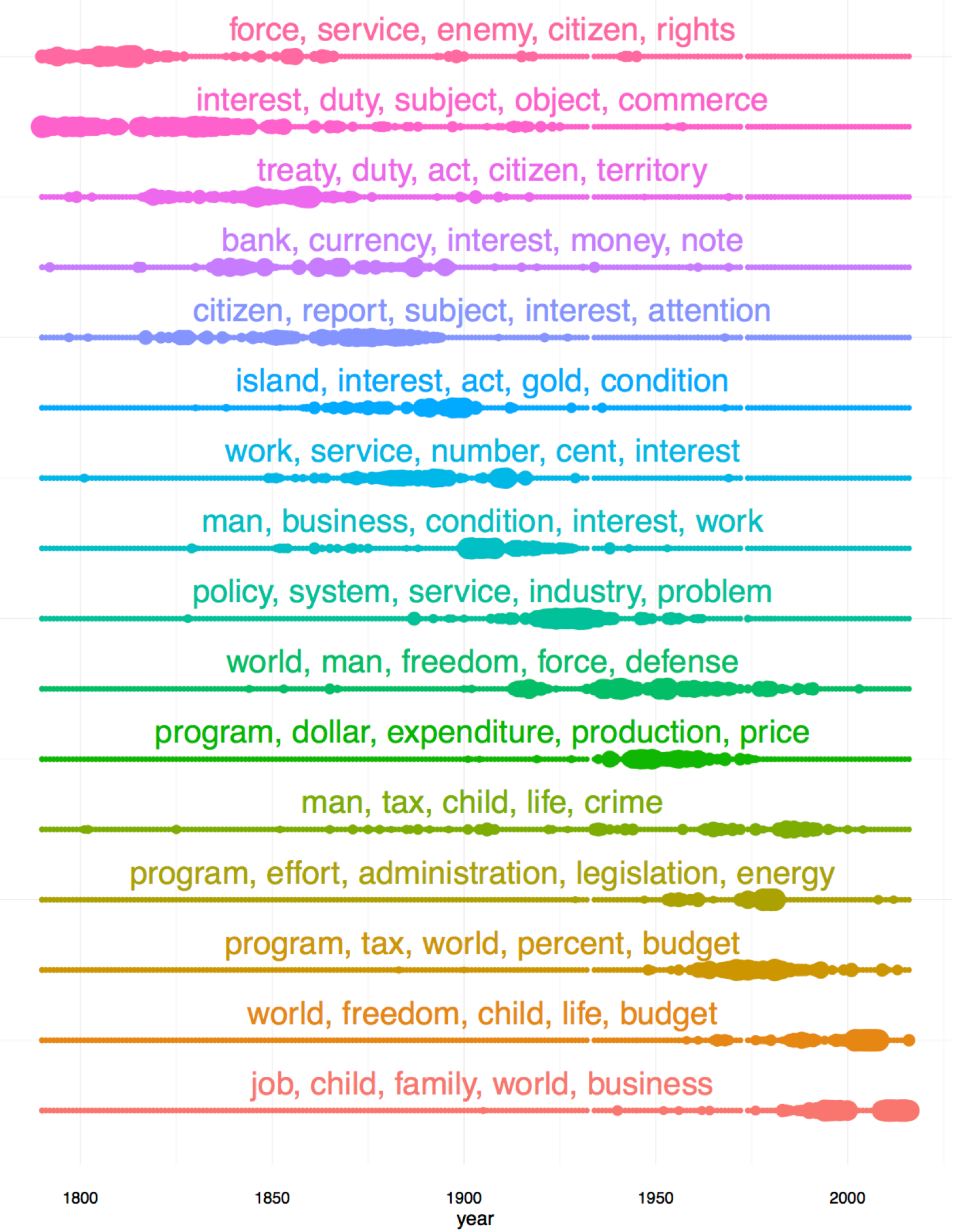


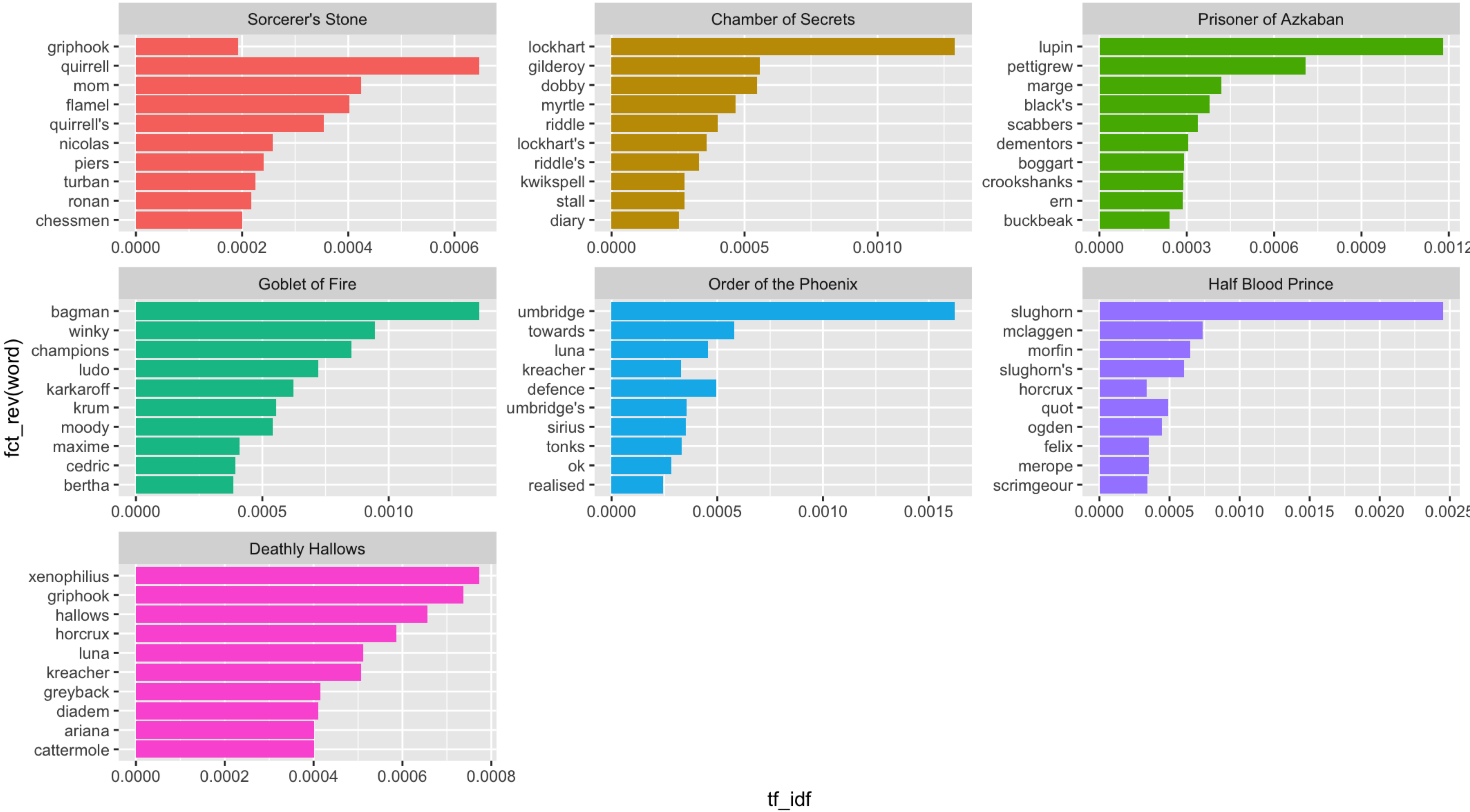

ME
VT
NH
NY
MA
CT
RI
WI
MI
MT
ND
WA
SD
MN
PA
OH
OR
ID
WY
NE
IA
IL
IN
NJ
MD
DE
WV
DC
UT
NV
CO
KS
MO
KY
VA
OK
TN
CA
NC
AR
MS
SC
AZ
NM
TX
LA
AL
GA
AK
HI
FL
KEY
CLINTON'S CHANCES
TRUMP'S CHANCES
50% 60 70 80 90
ONE ELECTORAL VOTE

The most used words for women vs. men

Likelihood that certain words appear after “she” vs. “he” in screen direction.




force, service, enemy, citizen, rights
interest, duty, subject, object, commerce
treaty, duty, act, citizen, territory
bank, currency, interest, money, note
citizen, report, subject, interest, attention
island, interest, act, gold, condition
work, service, number, cent, interest
man, business, condition, interest, work
policy, system, service, industry, problem
world, man, freedom, force, defense
program, dollar, expenditure, production, price
man, tax, child, life, crime
program, effort, administration, legislation, energy
program, tax, world, percent, budget
world, freedom, child, life, budget
job, child, family, world, business
1800
1850
1900
1950
2000
year


Sorcerer's Stone
griphook
quirrell
mom
flamel
quirrell's
nicolas
piers
turban
ronan
chessmen
0.0000
0.0002
0.0004
0.0006
Chamber of Secrets
lockhart
gilderoy
dobby
myrtle
riddle
lockhart's
riddle's
kwikspell
stall
diary
0.0000
0.0005
0.0010
Prisoner of Azkaban
lupin
pettigrew
marge
black's
scabbers
dementors
boggart
crookshanks
ern
buckbeak
0.0000
0.0003
0.0006
0.0009
0.0012
Goblet of Fire
bagman
winky
champions
ludo
karkaroff
krum
moody
maxime
cedric
bertha
0.0000
0.0005
0.0010
Order of the Phoenix
umbridge
towards
luna
kreacher
defence
umbridge's
sirius
tonks
ok
realised
0.0000
0.0005
0.0010
0.0015
Half Blood Prince
slughorn
mclaggen
morfin
slughorn's
horcrux
quot
ogden
felix
merope
scrimgeour
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
Deathly Hallows
xenophilius
griphook
hallows
horcrux
luna
kreacher
greyback
diadem
ariana
cattermole
0.0000
0.0002
0.0004
0.0006
0.0008
fct_rev(word)
tf_idf

ETHICS OF DATA

POSSIBLE PITFALLS

Manipulation

Don’t coerce people

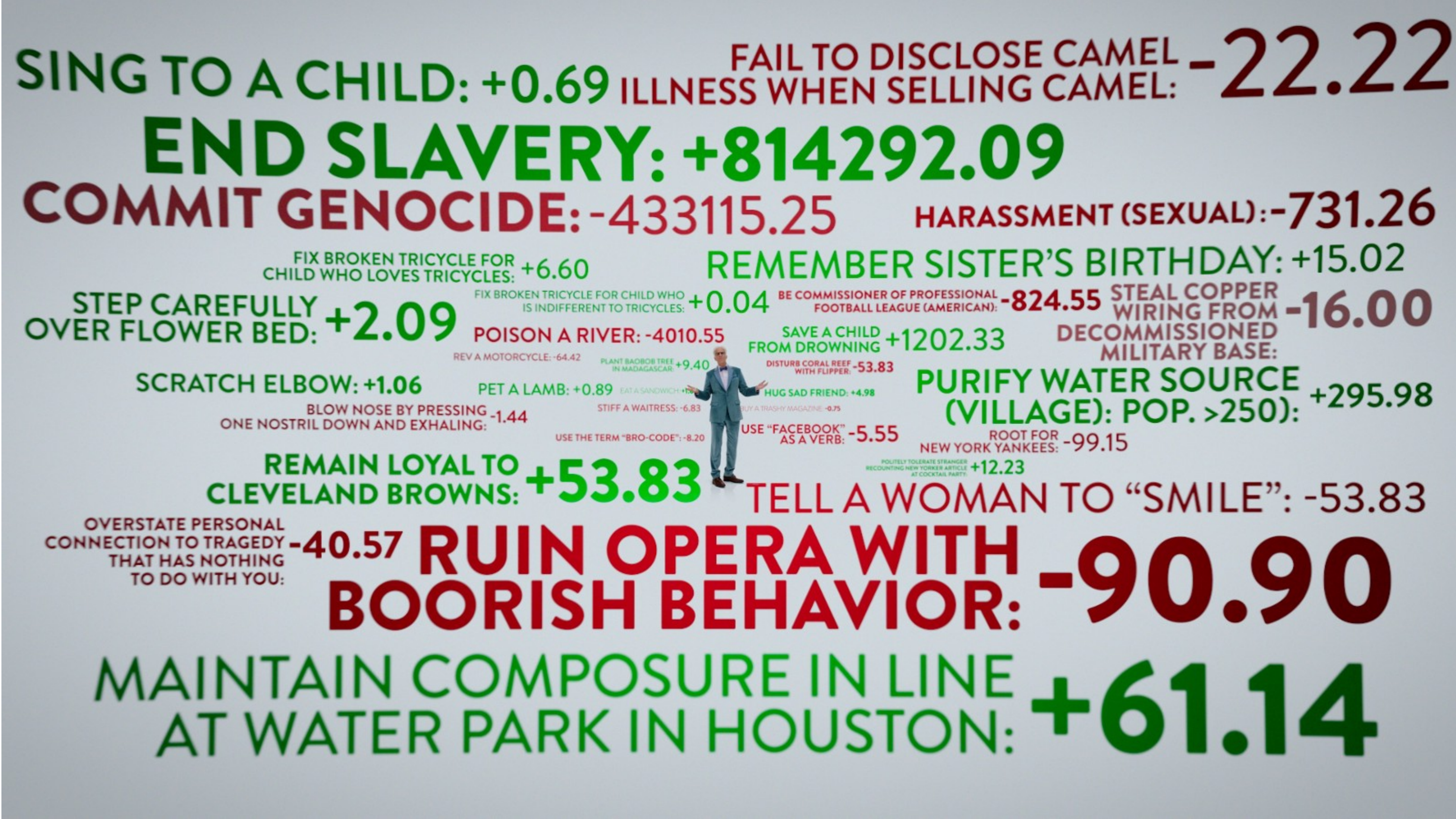
Don’t make critical decisions on data alone

Bias

There’s no such thing as an objective model

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: +1.

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

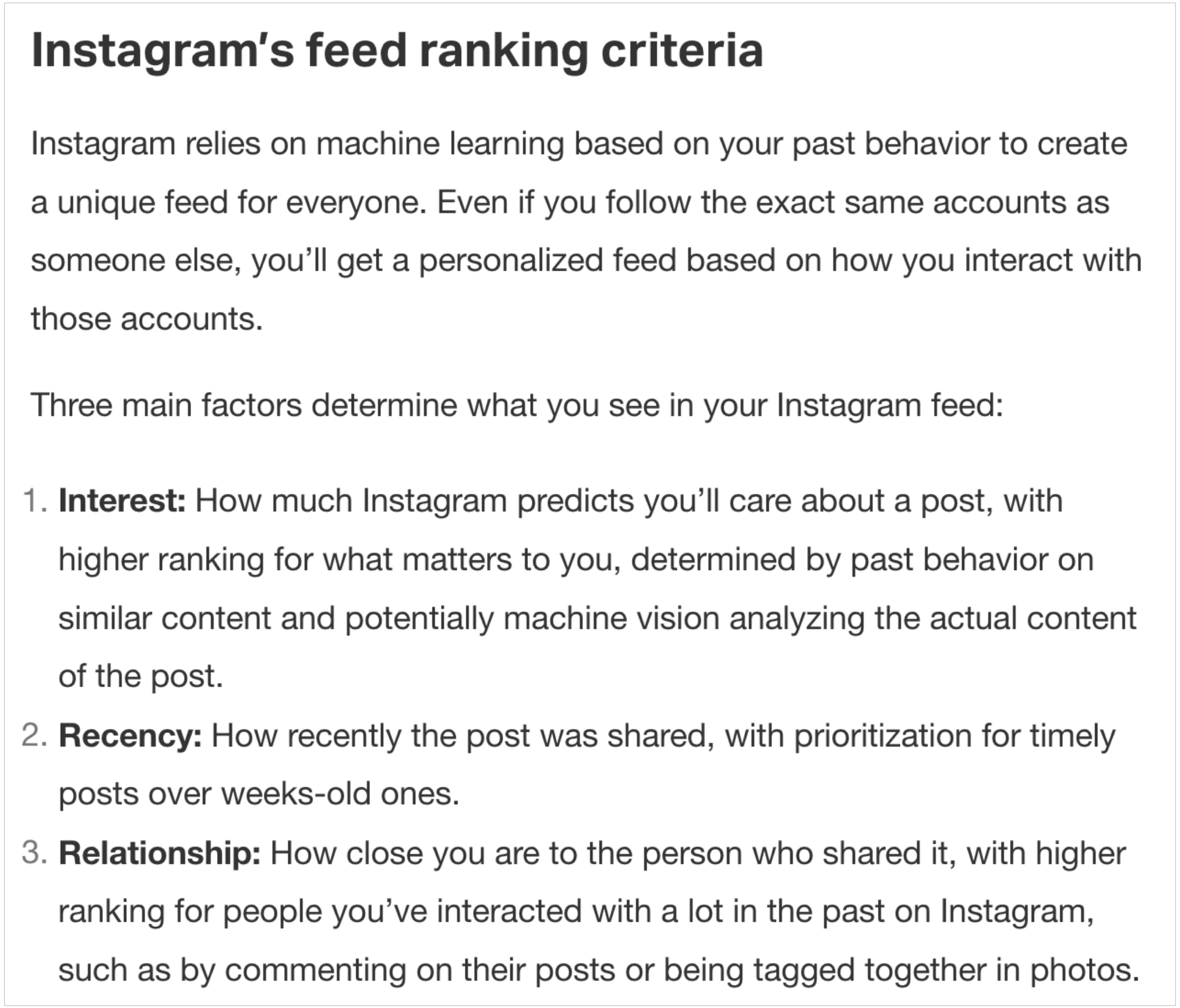
MAINTAIN COMPOSURE IN LINE AT WATER PARK IN HOUSTON: +61.14

Imagine a future
where your life is measured by a number—three digits
that dictate your place in society.
That future is now.

WIRED

048
WIRED—26.01
HELLO!
MY SCORE IS
549
163
CREATE. CONNECT. CONDÉ NAST

JANUARY 2018 | ADD IT UP

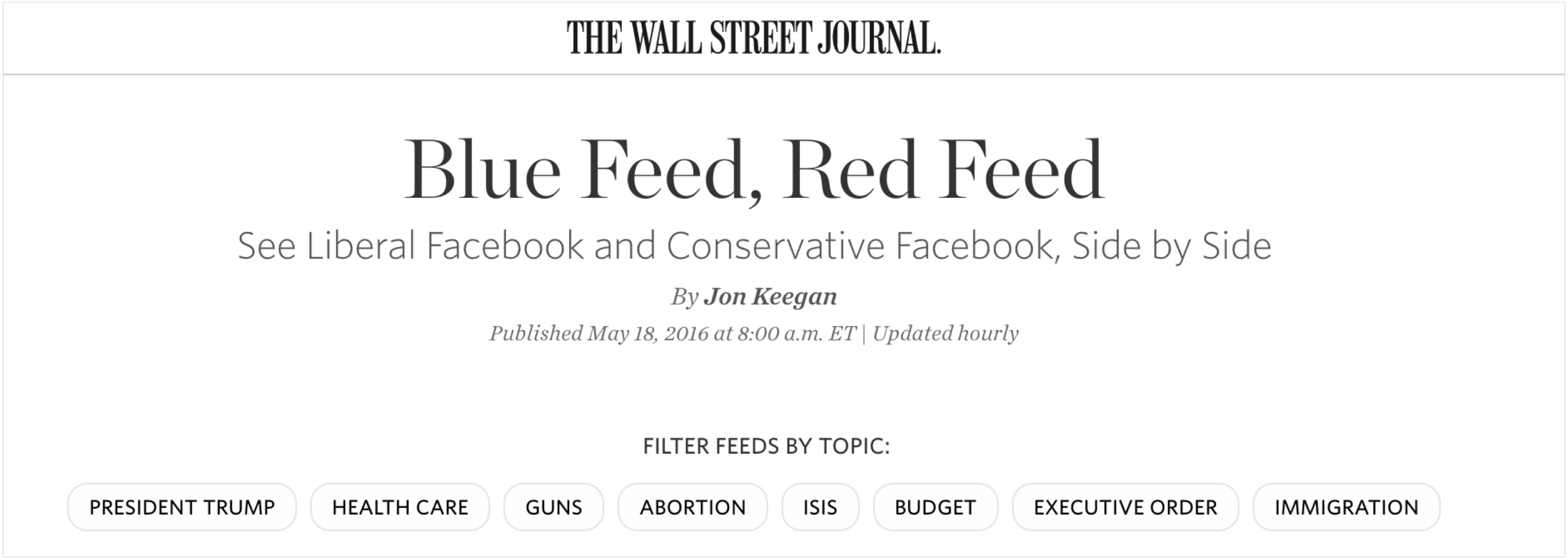


Instagram’s feed ranking criteria

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.

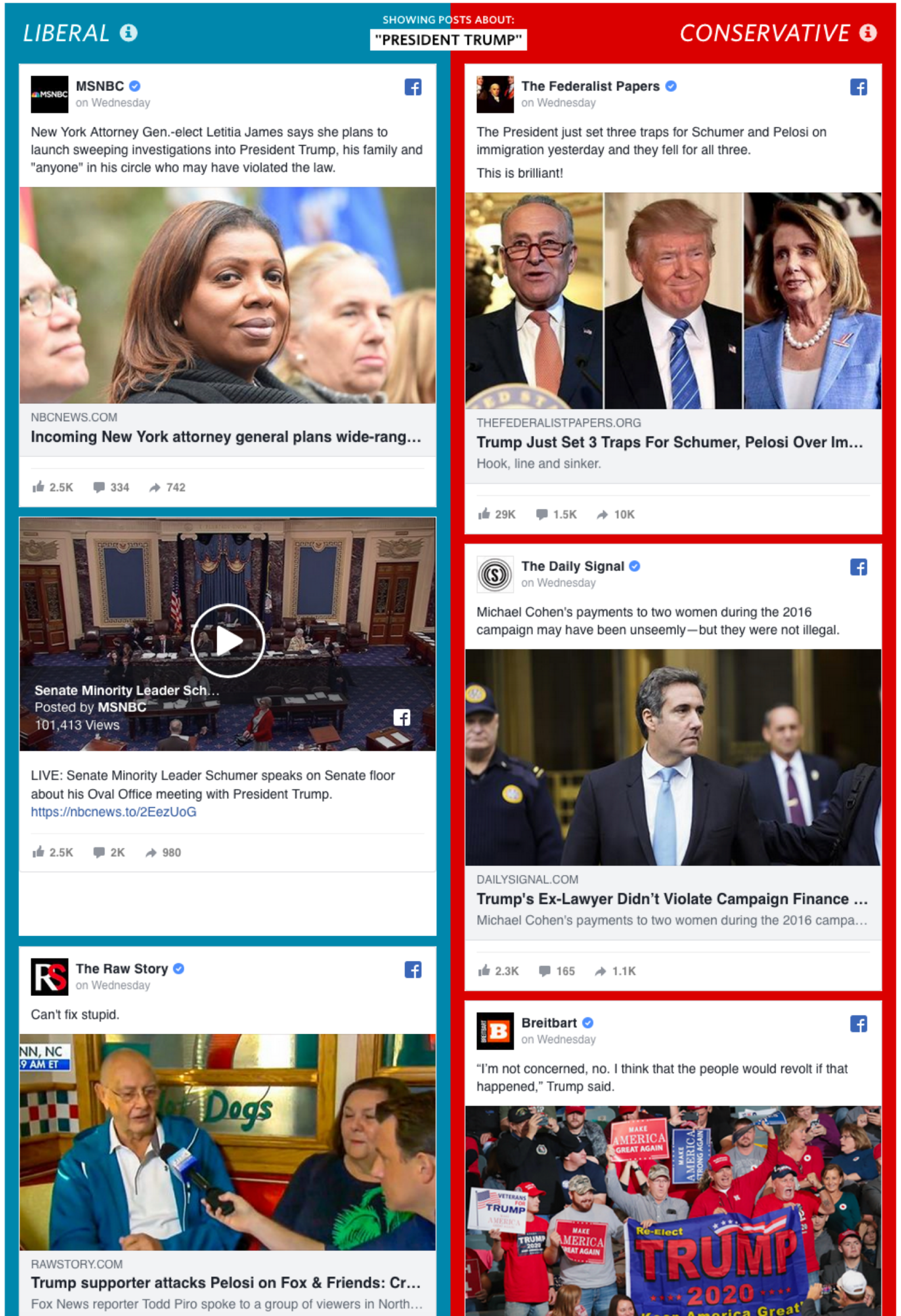
Blue Feed, Red Feed

See Liberal Facebook and Conservative Facebook, Side by Side

*By **Jon Keegan***

Published May 18, 2016 at 8:00 a.m. ET | Updated hourly

FILTER FEEDS BY TOPIC:



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

Senate Minority Leader Sch…
Posted by **MSNBC**
101,413 Views

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
Trump supporter attacks Pelosi on Fox & Friends: Cr…
Fox News reporter Todd Piro spoke to a group of viewers in North…

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

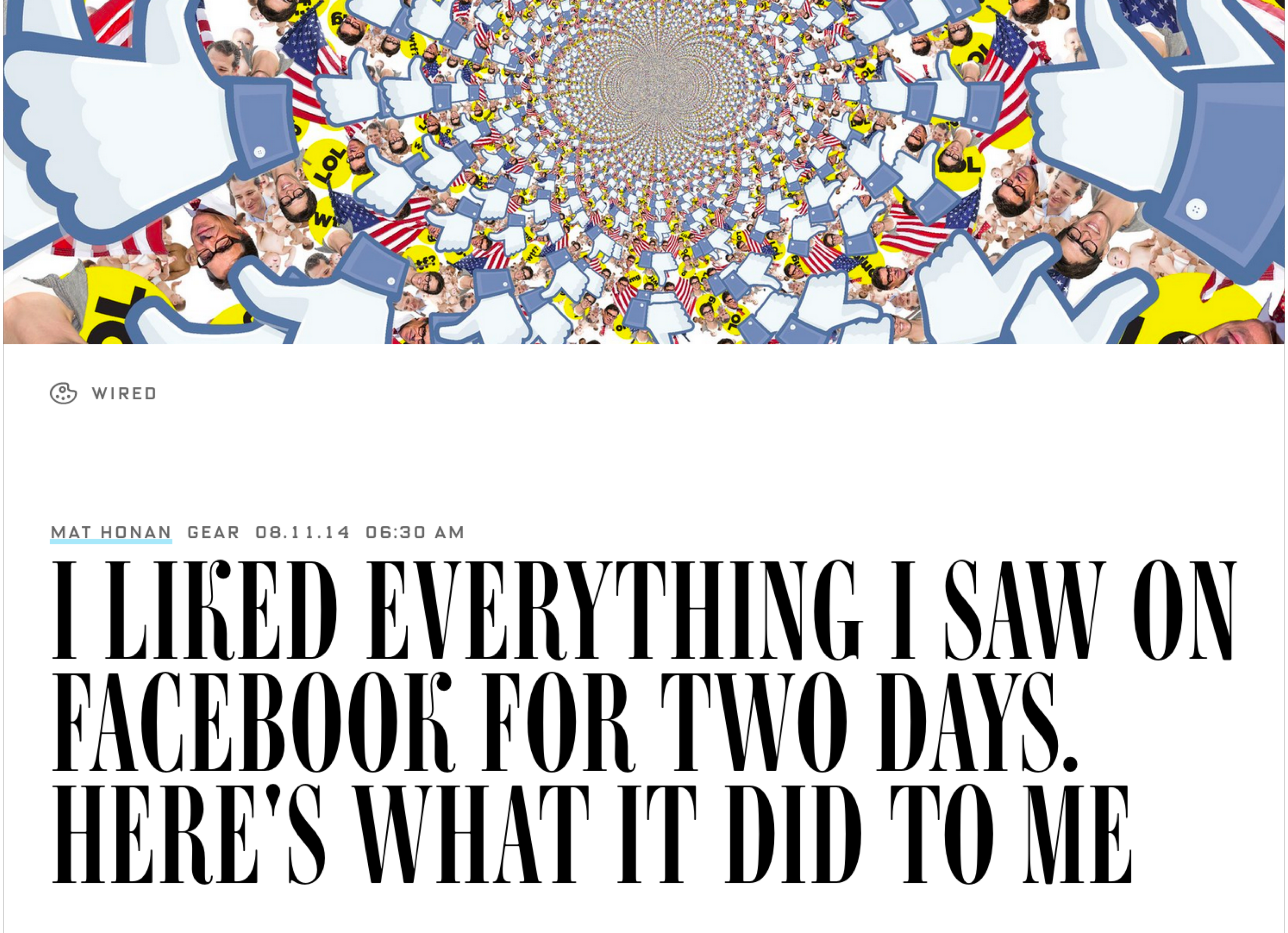
Michael Cohen's payments to two women during the 2016 campaign may have been unseemly—but they were not illegal.

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 08.11.14 06:30 AM

I LIKED EVERYTHING I SAW ON FACEBOOK FOR TWO DAYS. HERE'S WHAT IT DID TO ME

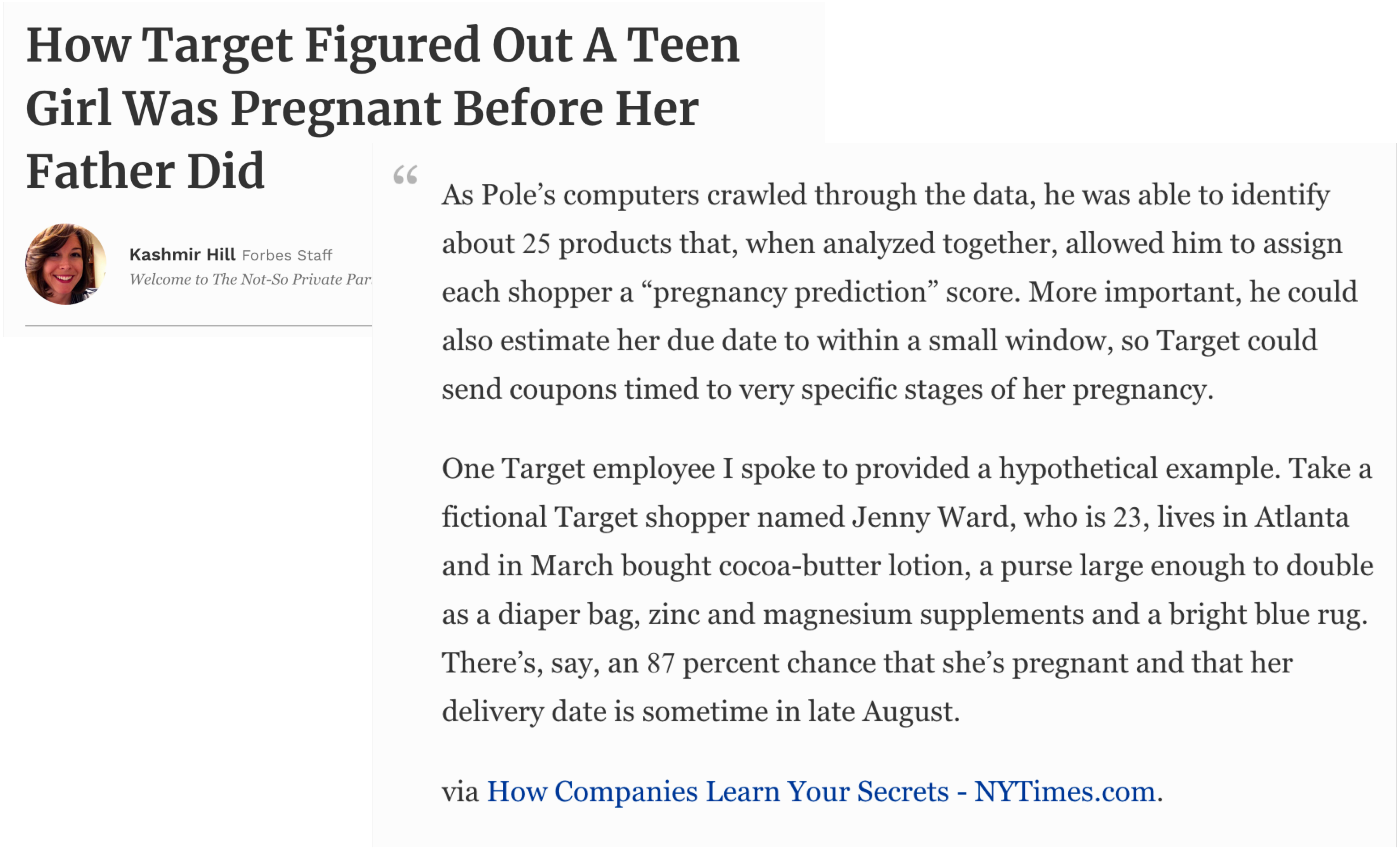
#109 Is Facebook Spying on You?

November 2, 2017

Reply All

#109 Is Facebook Spying on You?

00:32:51

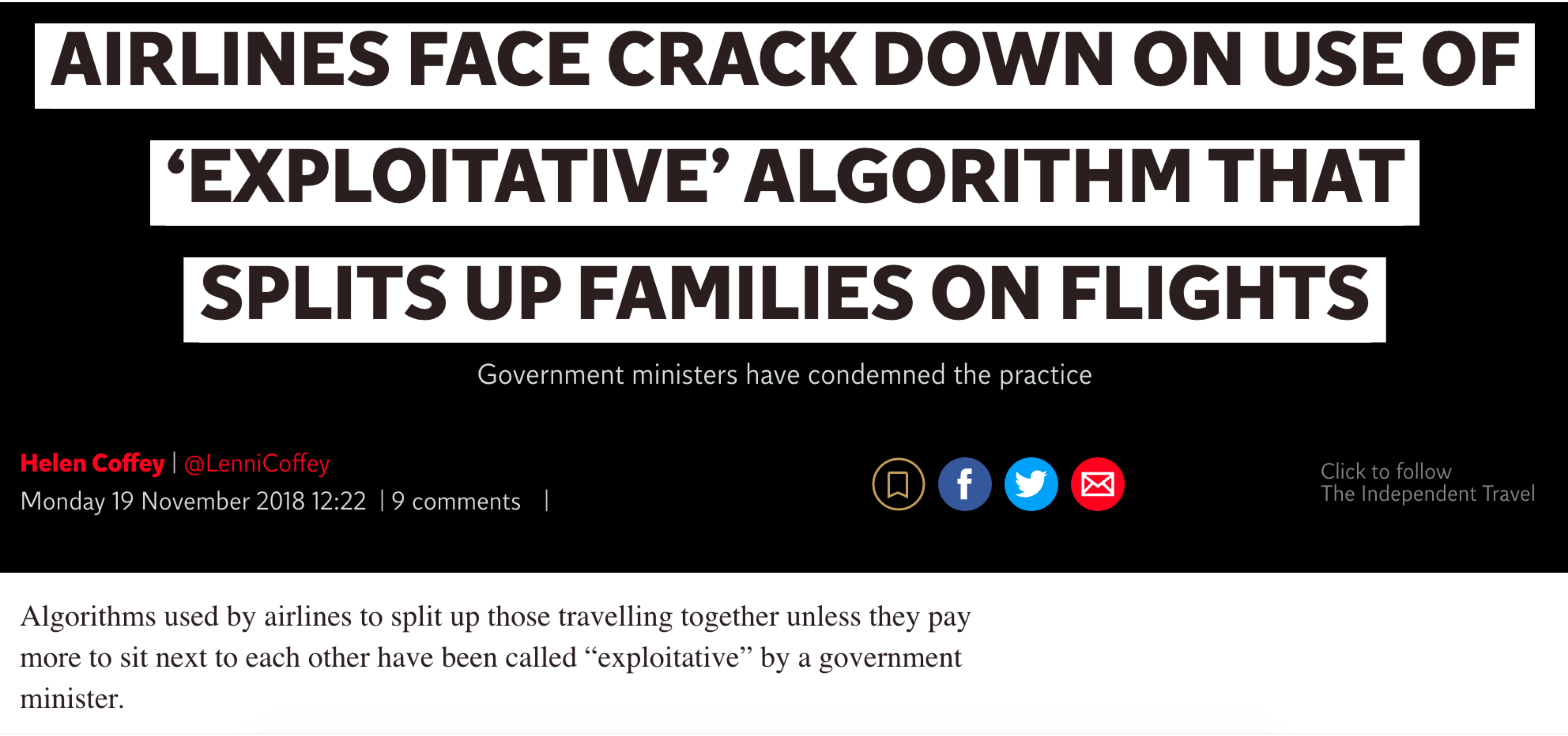


How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Forbes Staff

Welcome to The Not-So Private Par

> 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

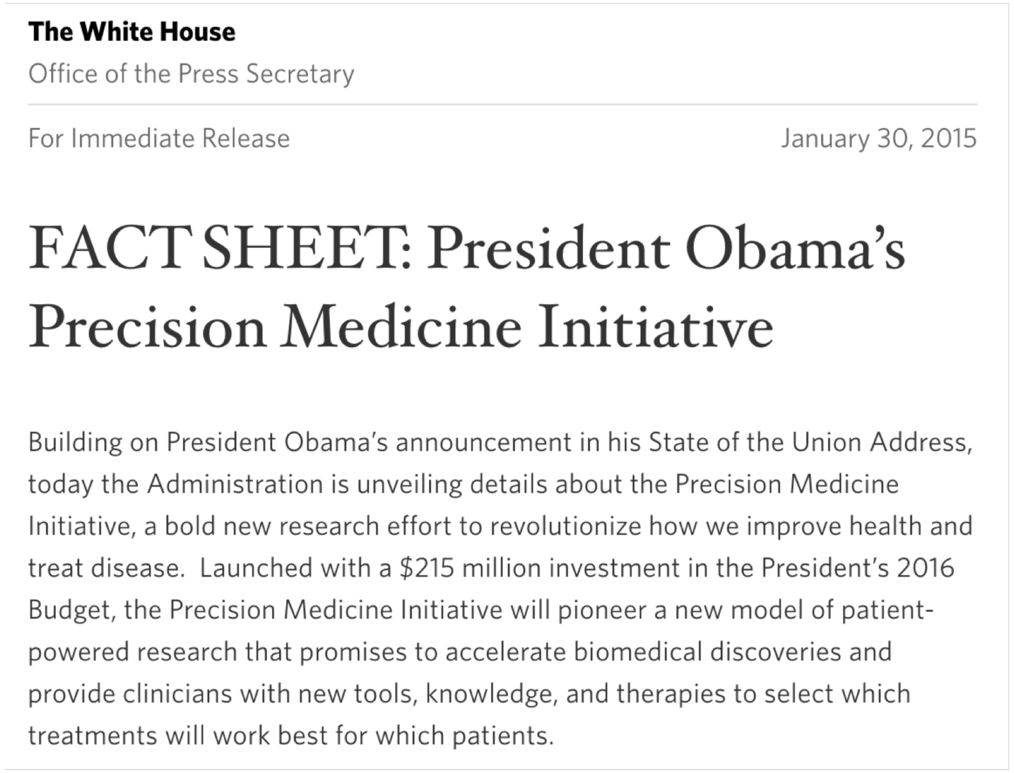
Helen Coffey | @LenniCoffey

Monday 19 November 2018 12:22 | 9 comments |

Click to follow
The Independent Travel

Algorithms used by airlines to split up those travelling together unless they pay more to sit next to each other have been called “exploitative” by a government minister.

IT'S NOT ALL DYSTOPIAN

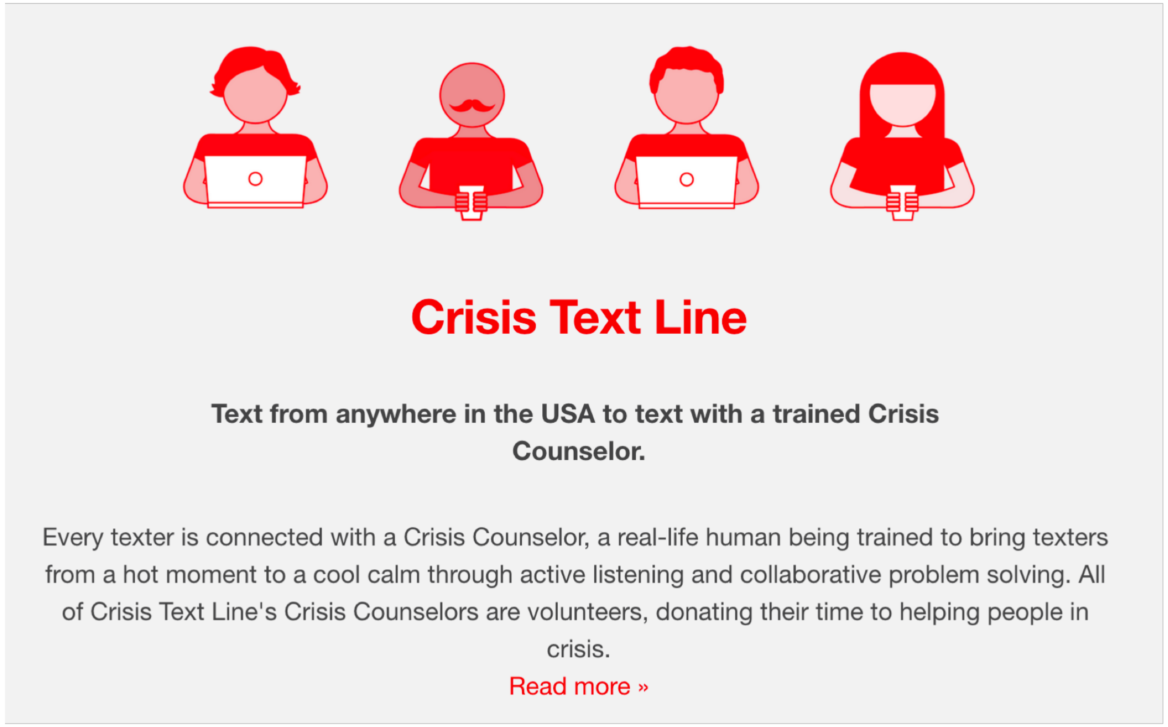


The White House
Office of the Press Secretary

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.

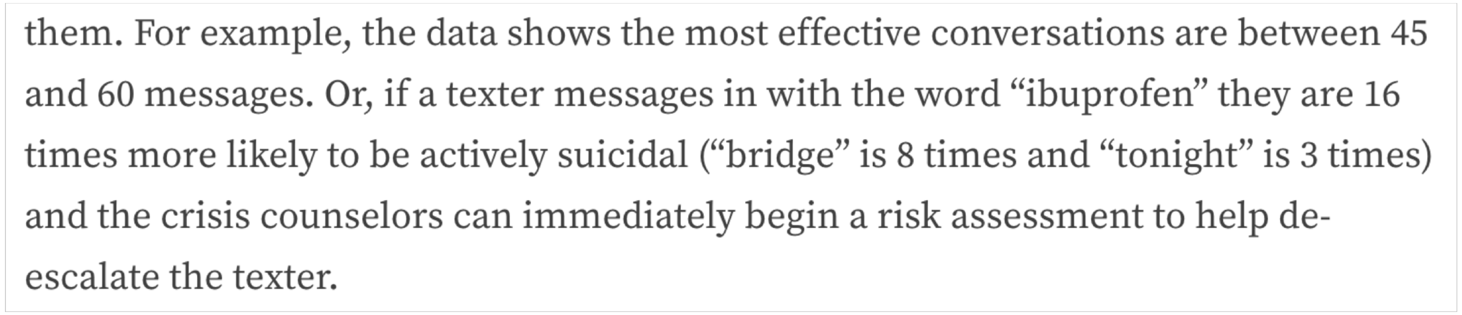


Crisis Text Line

Text from anywhere in the USA to text with a trained Crisis Counselor.

Every texter is connected with a Crisis Counselor, a real-life human being trained to bring texters from a hot moment to a cool calm through active listening and collaborative problem solving. All of Crisis Text Line's Crisis Counselors are volunteers, donating their time to helping people in crisis.

Read more »



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?

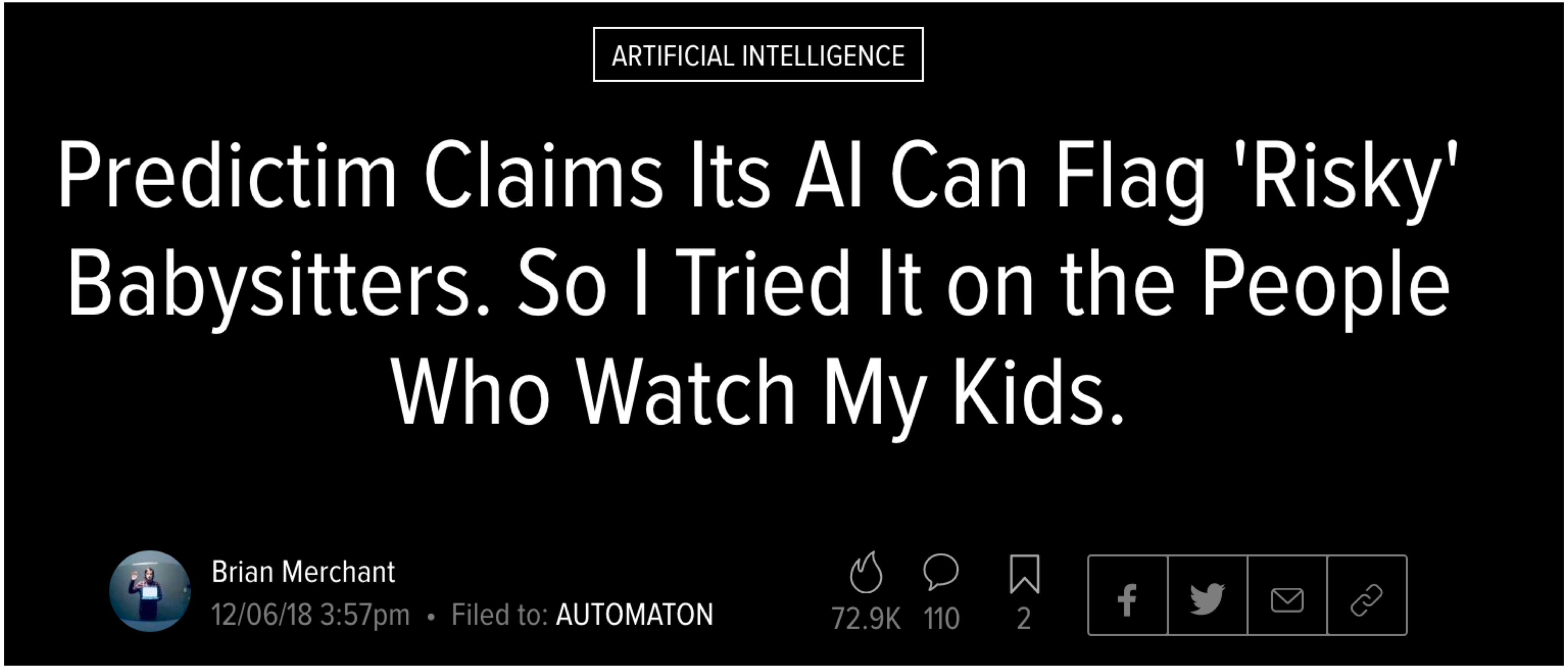
AVOID MANIPULATION

Think about people

Think about autonomy

You shall not live on data alone

Bias

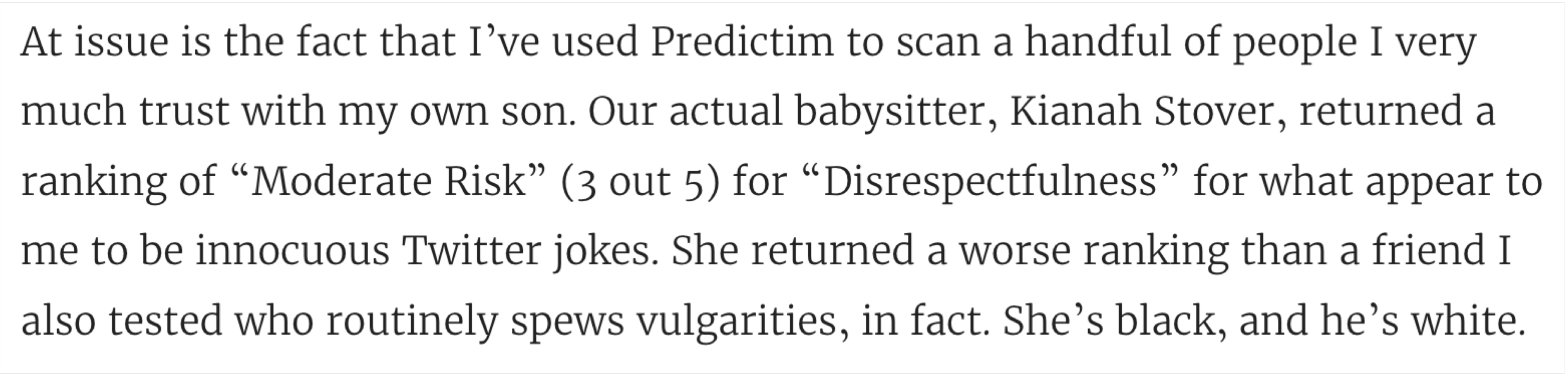


ARTIFICIAL INTELLIGENCE

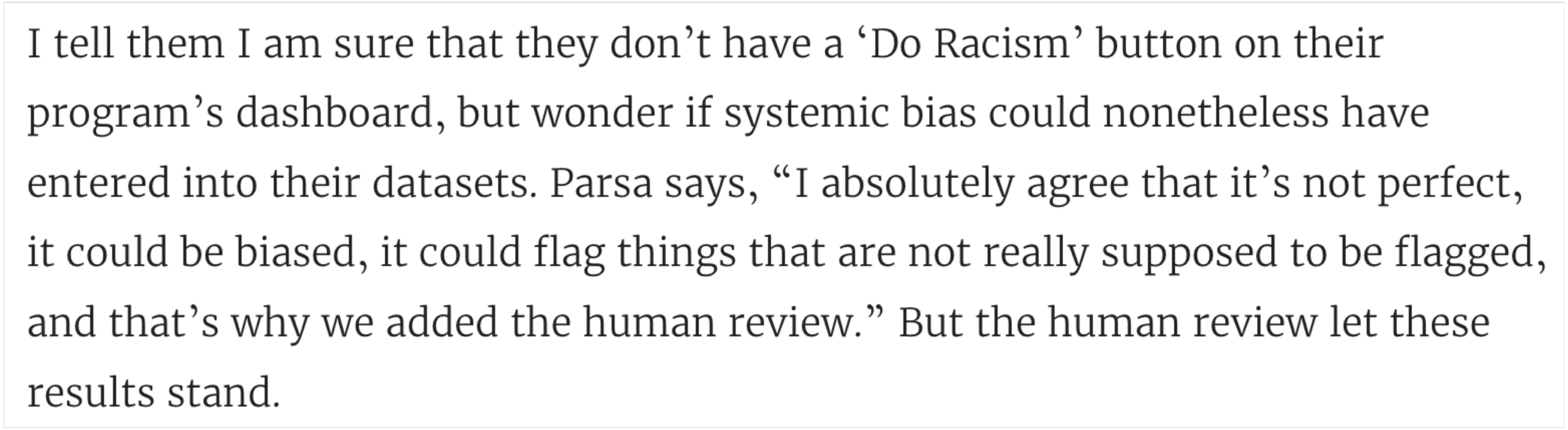
Predictim Claims Its AI Can Flag 'Risky' Babysitters. So I Tried It on the People Who Watch My Kids.

Brian Merchant

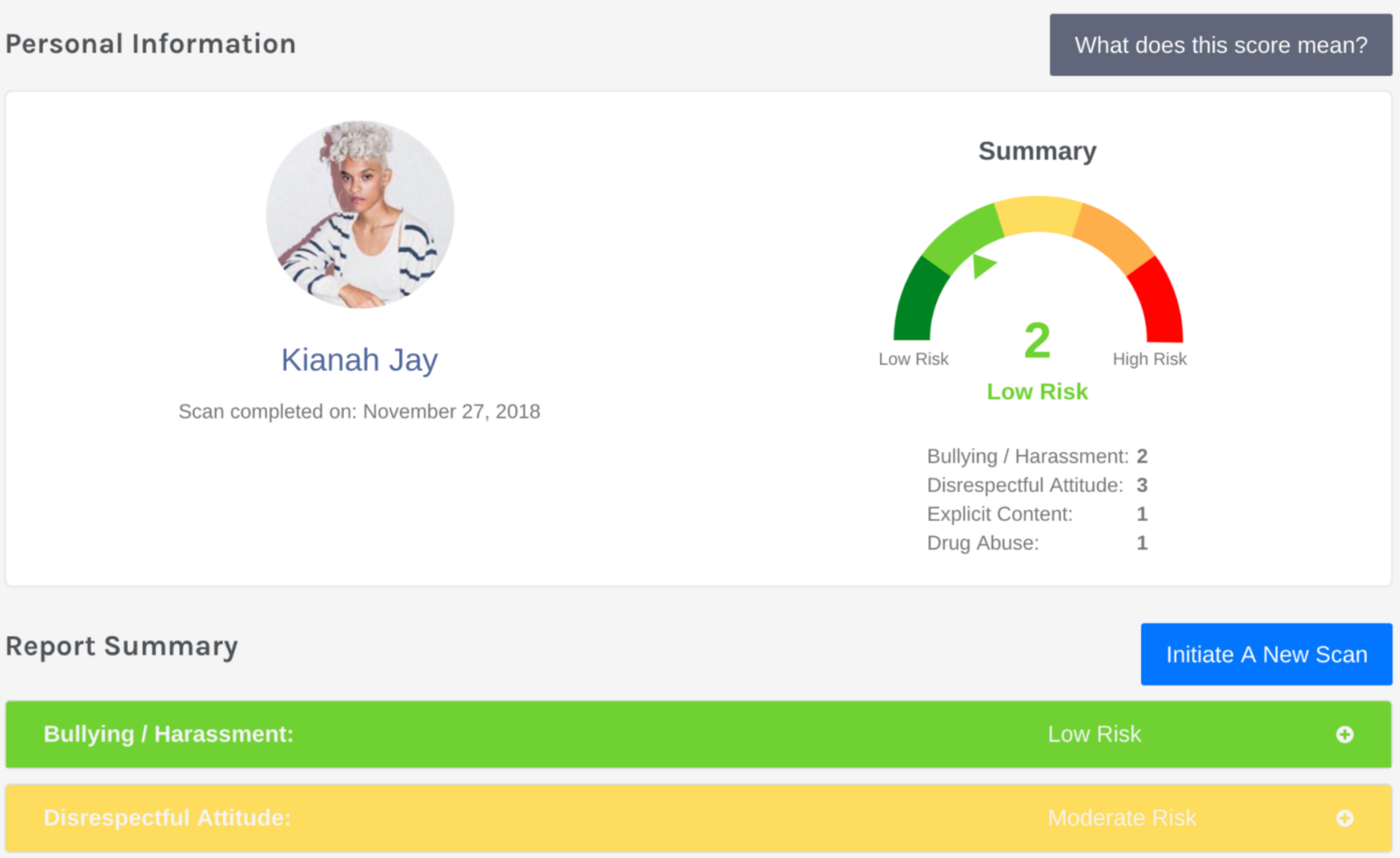
12/06/18 3:57pm • Filed to: AUTOMATON



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

Kianah Jay

Scan completed on: November 27, 2018

Summary

2

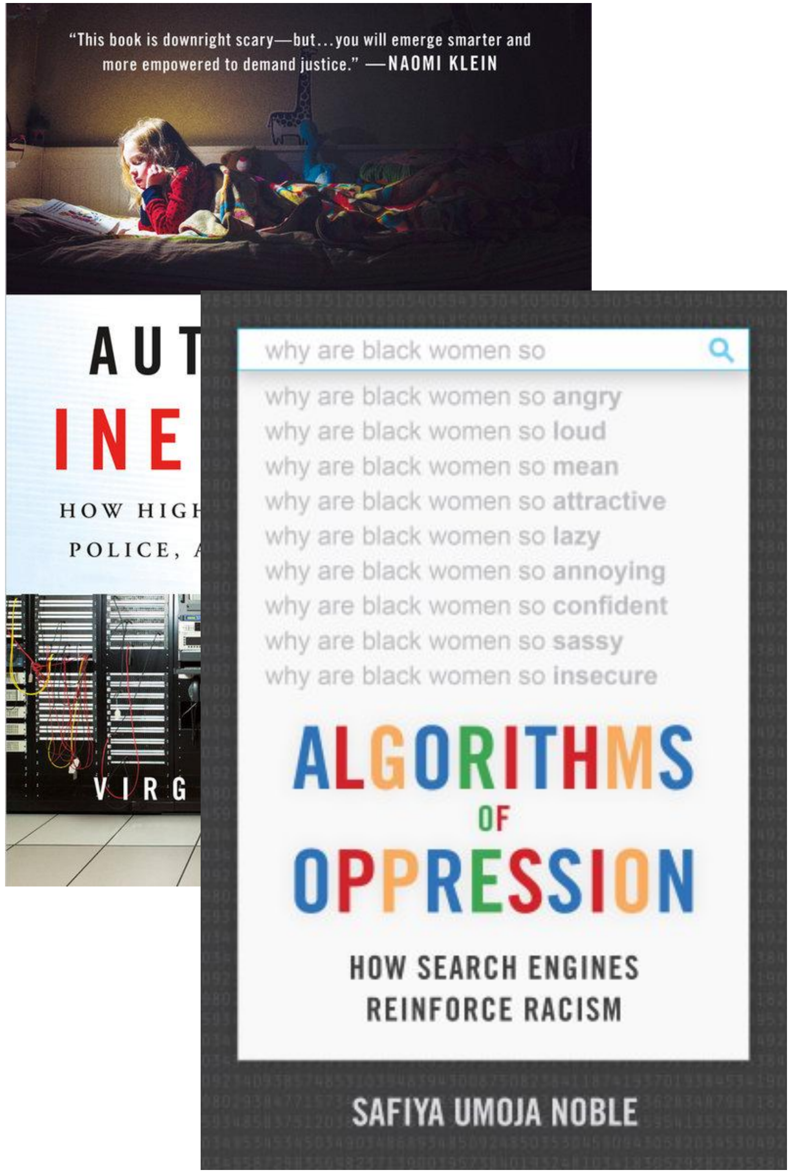
Low Risk

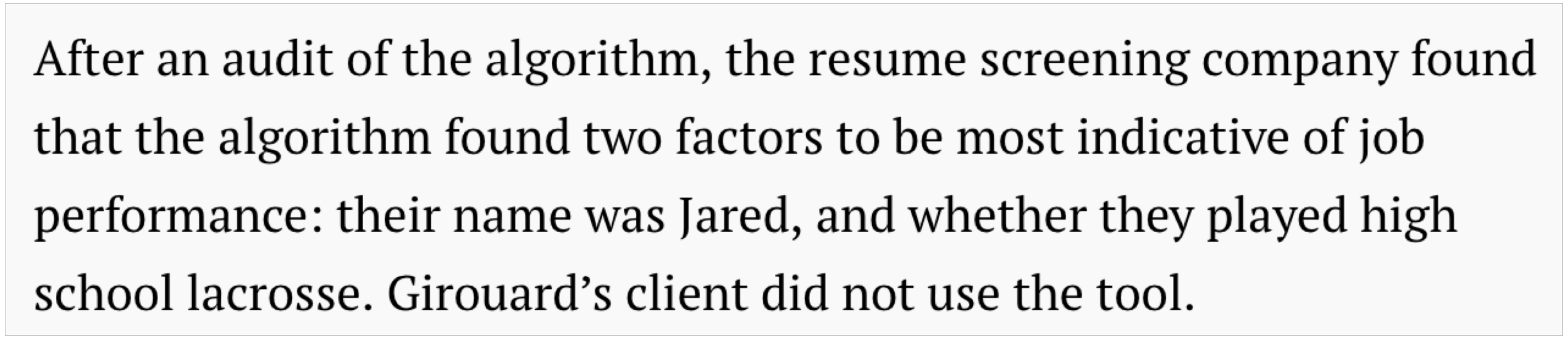
Bullying / Harassment:	2
Disrespectful Attitude:	3
Explicit Content:	1
Drug Abuse:	1

Report Summary

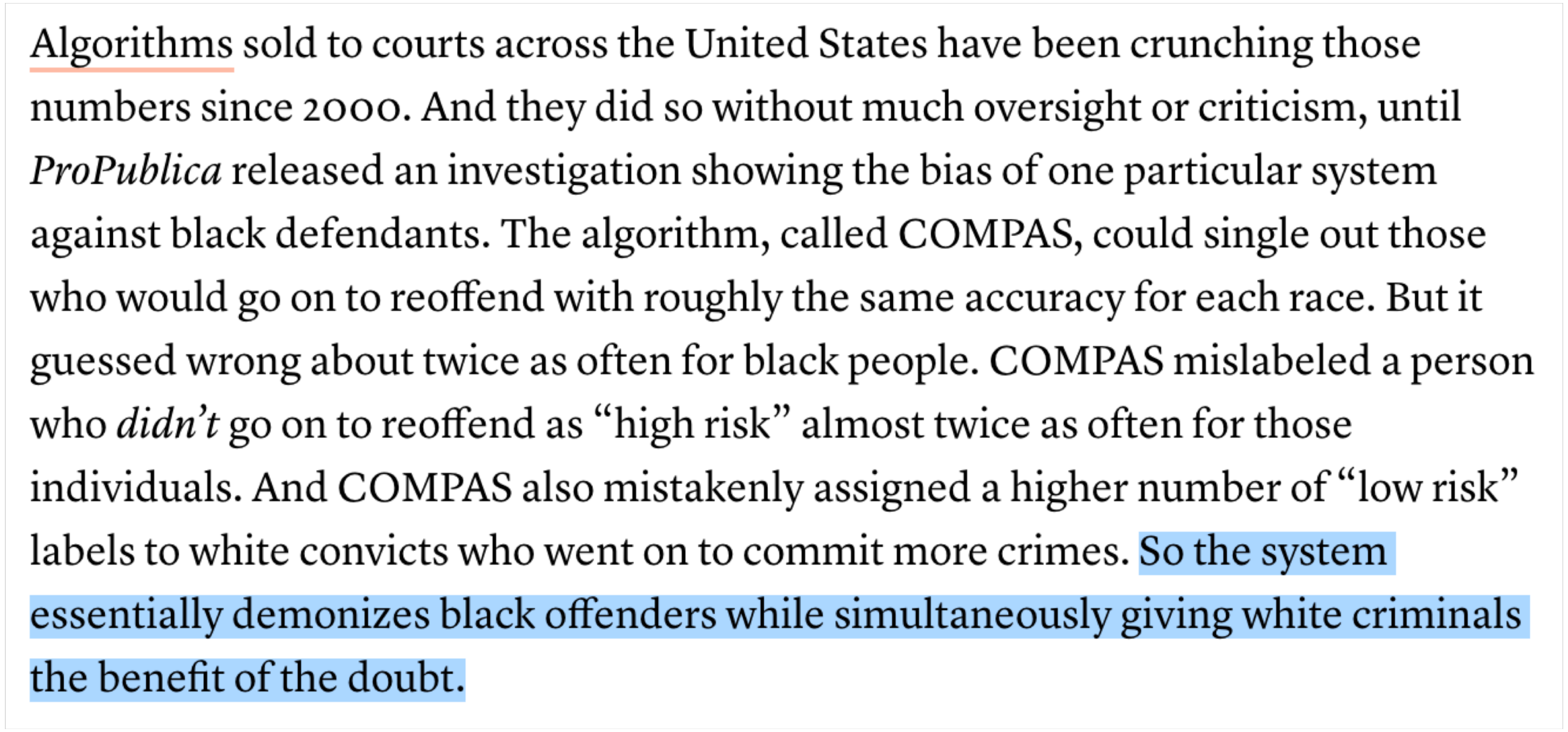
Bullying / Harassment:	Low Risk
Disrespectful Attitude:	Moderate Risk

https://www.youtube.com/watch?v=BSpAWkQLlgM


"This book is downright scary—but…you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN
AUT
INE
HOW HIGH
POLICE,
VIRG
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



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.

FIGHT THE ALGORITHMS

Incognito / private windows

adssettings.google.com

AVOID BIAS

Make sure your sample is representative

Think about theory

Remember that no data, models, or algorithms are neutral

CURIOSITY

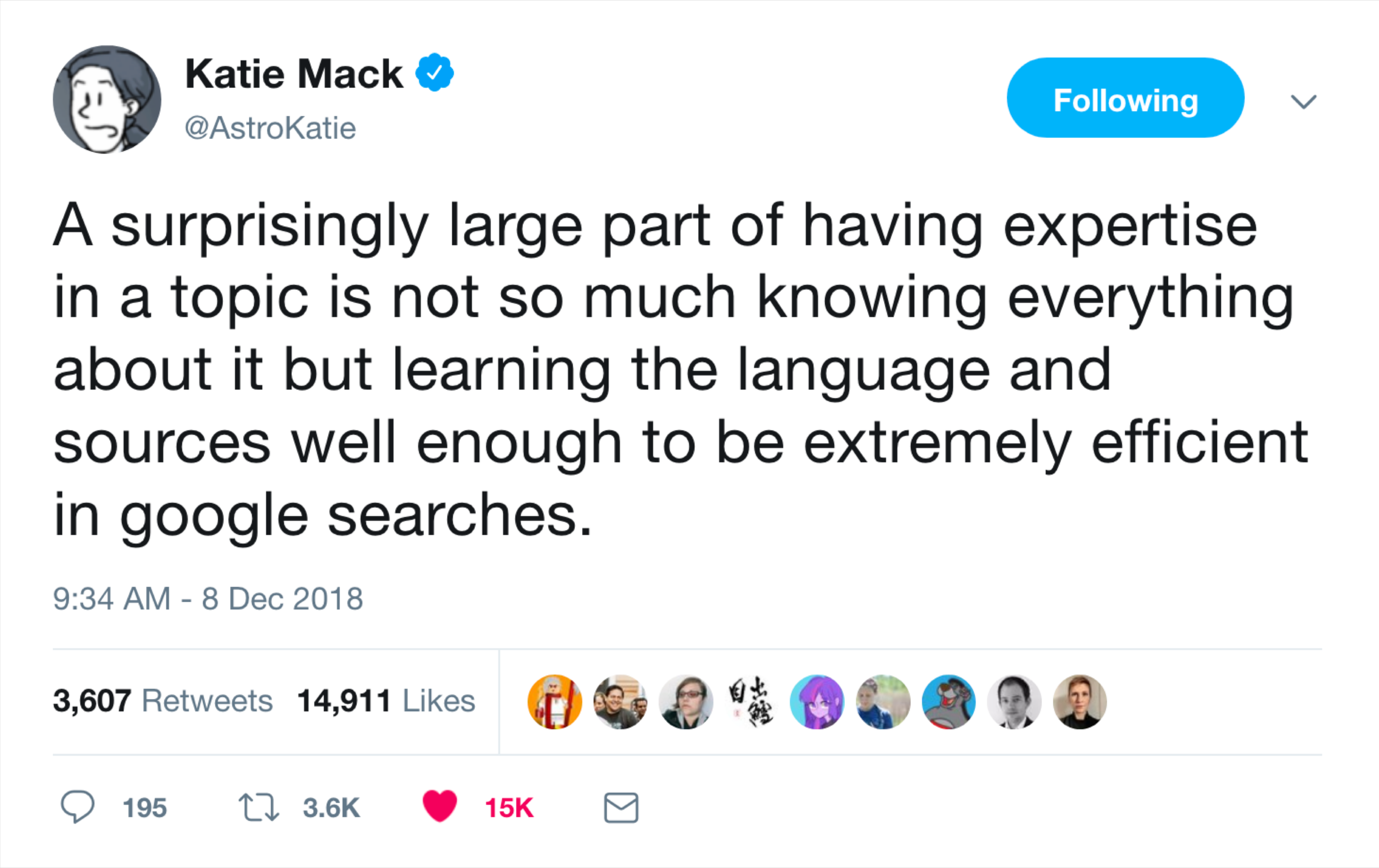
How do I keep learning R?

What class should I take next?

What book should I read next?

You don’t learn R

You learn how to do things in R

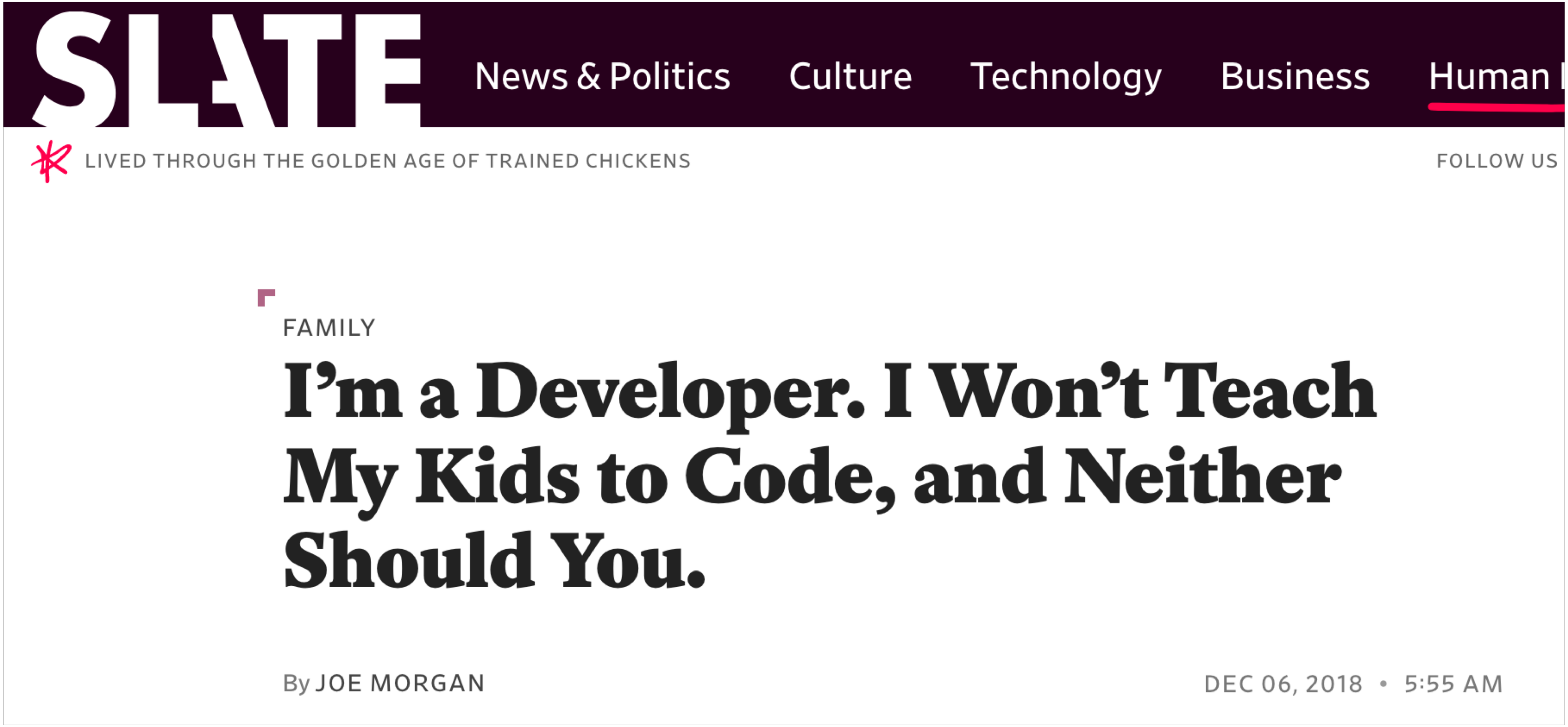


Katie Mack
@AstroKatie

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

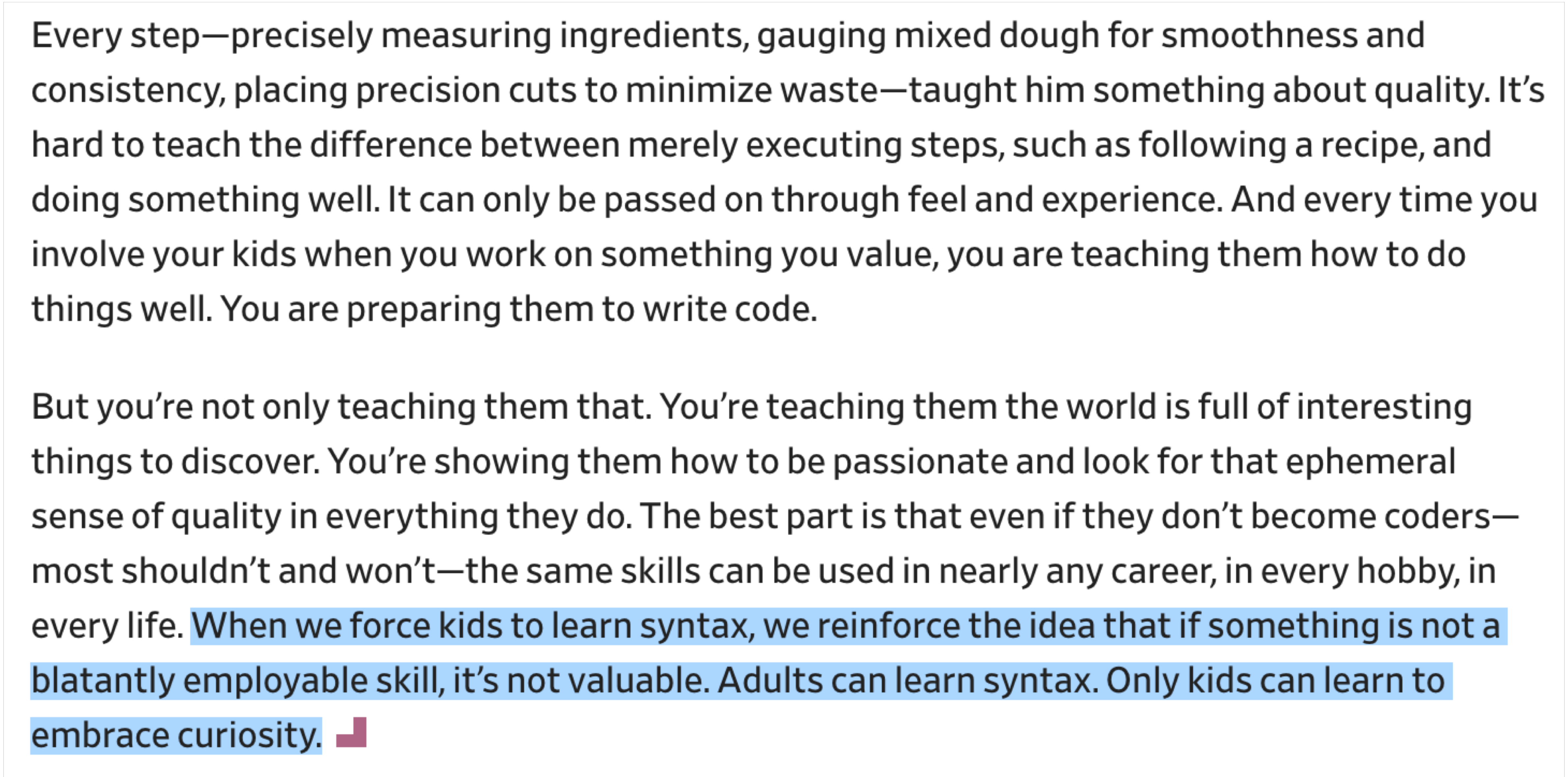


FAMILY

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.

EMBRACE CURIOSITY

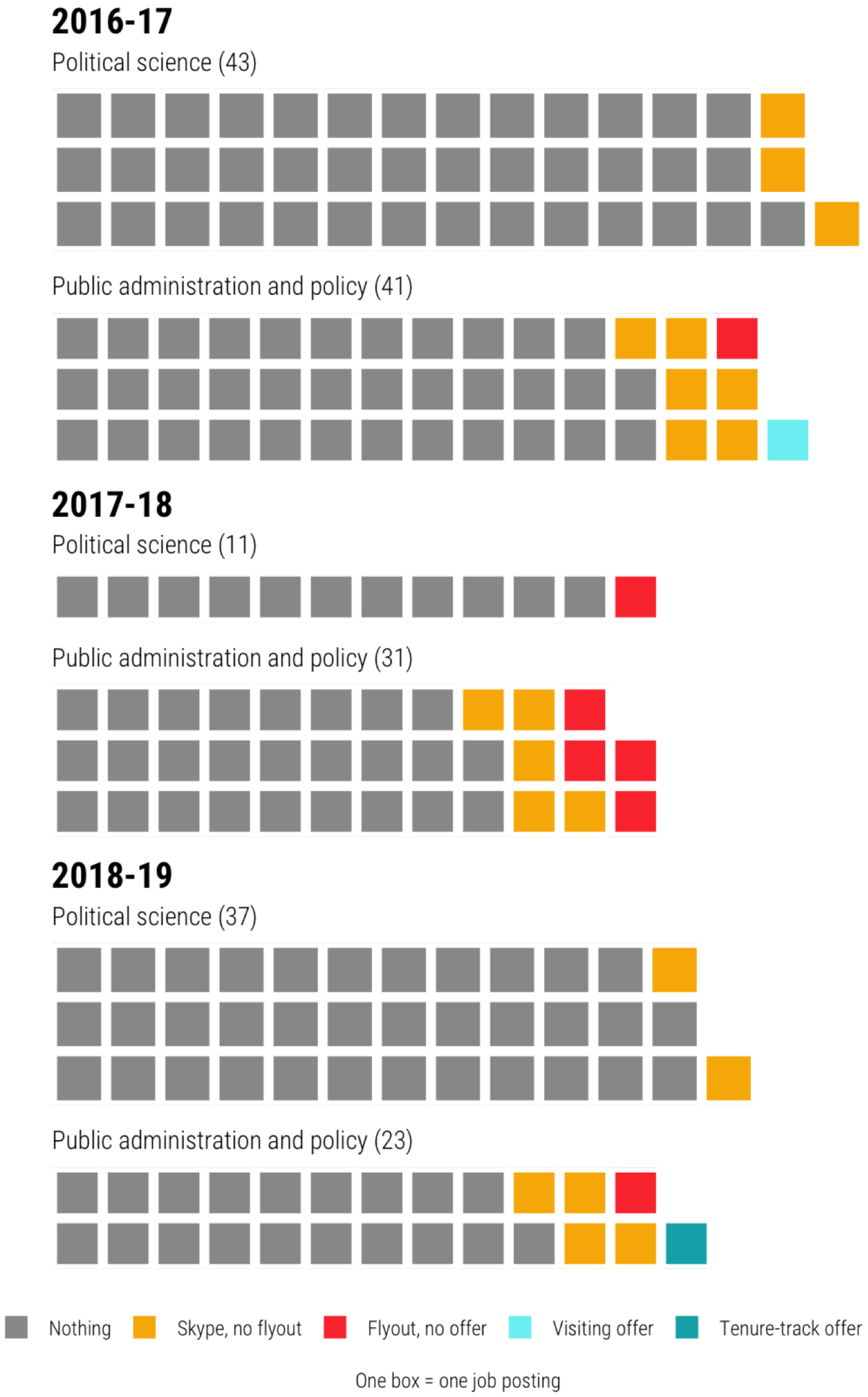
Find excuses to use R

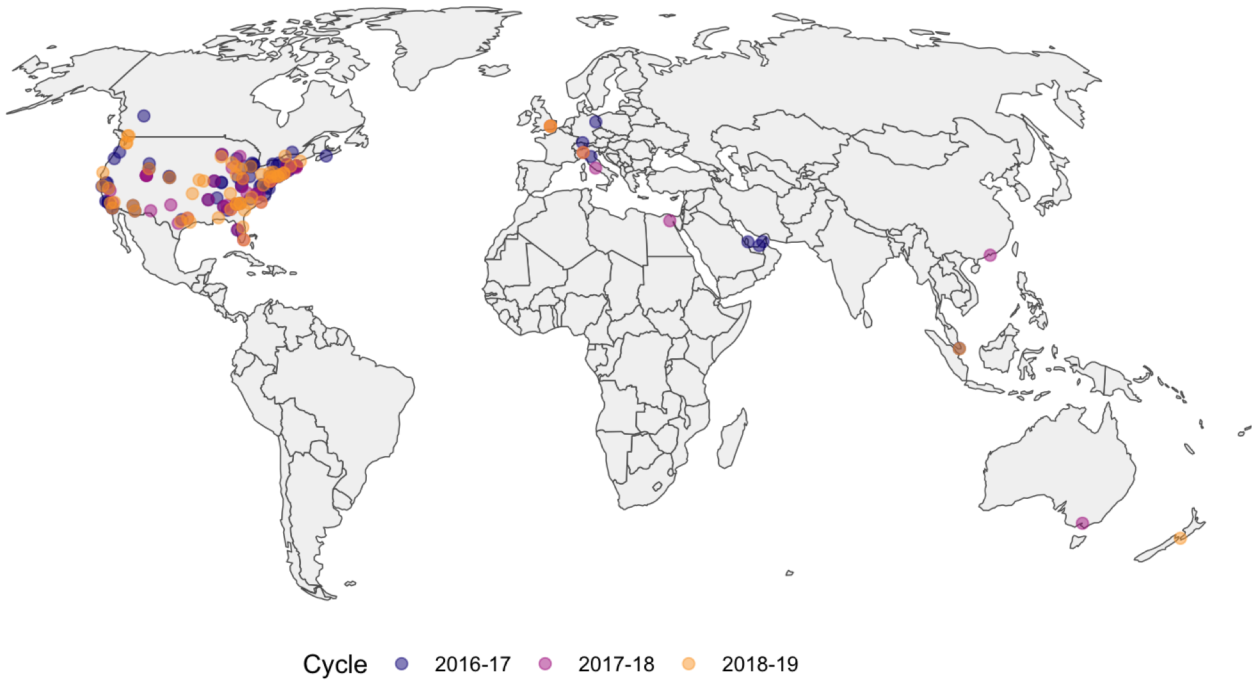
(This is why I subjected you to the code-through assignment)

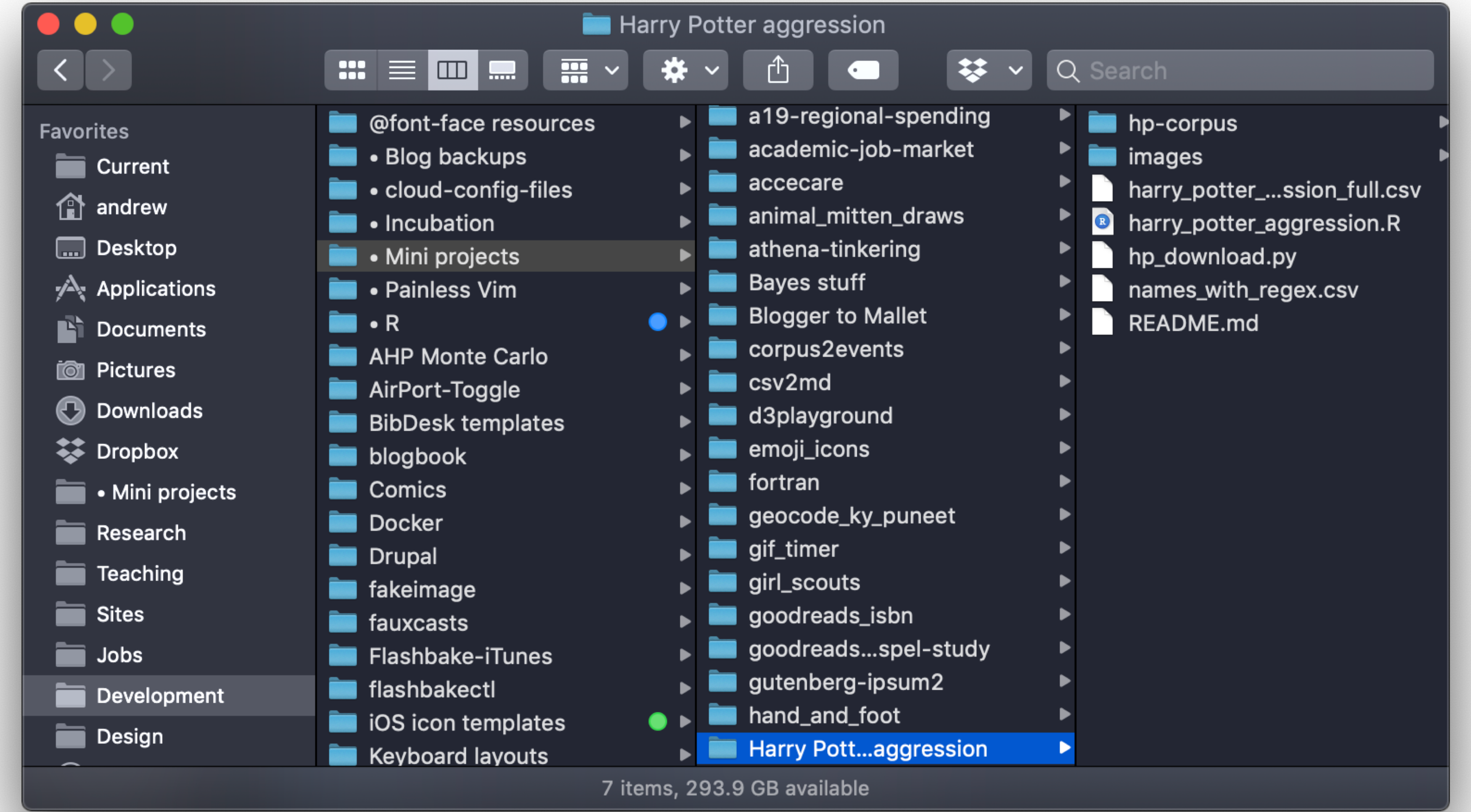
Dumb dinky projects

Data play time

Actual projects


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


Cycle
2016-17
2017-18
2018-19



Harry Potter aggression

Search

Favorites
- Current
- andrew
- Desktop
- Applications
- Documents
- Pictures
- Downloads
- Dropbox
- • Mini projects
- Research
- Teaching
- Sites
- Jobs
- Development
- Design

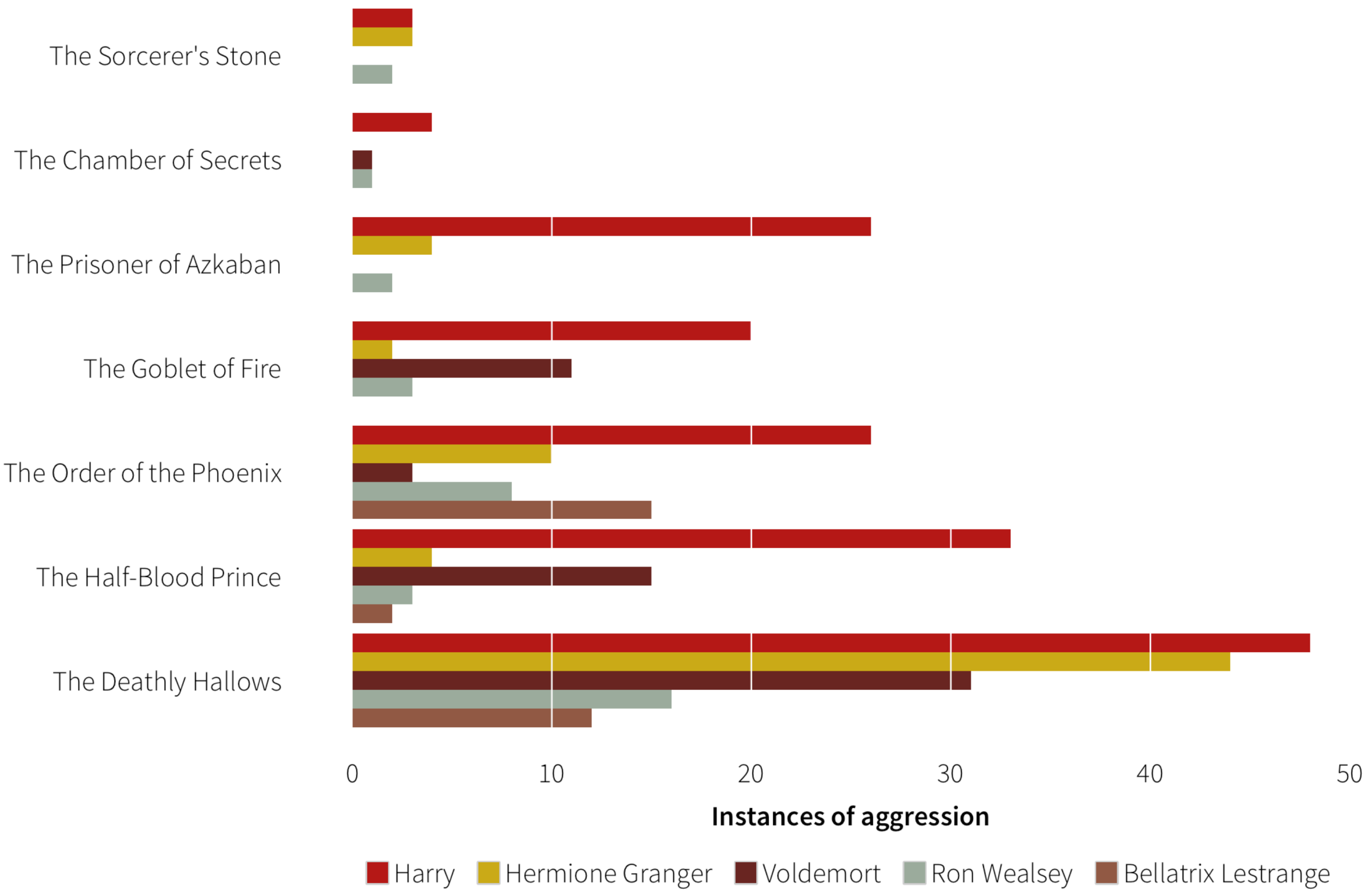
- @font-face resources
- • Blog backups
- • cloud-config-files
- • Incubation
- • Mini projects
- • Painless Vim
- • R
- AHP Monte Carlo
- AirPort-Toggle
- BibDesk templates
- blogbook
- Comics
- Docker
- Drupal
- fakeimage
- fauxcasts
- Flashbake-iTunes
- flashbakectl
- iOS icon templates
- Keyboard layouts

- a19-regional-spending
- academic-job-market
- accecare
- animal_mitten_draws
- athena-tinkering
- Bayes stuff
- Blogger to Mallet
- corpus2events
- csv2md
- d3playground
- emoji_icons
- fortran
- geocode_ky_puneet
- gif_timer
- girl_scouts
- goodreads_isbn
- goodreads...spel-study
- gutenberg-ipsum2
- hand_and_foot
- Harry Pott...aggression

- hp-corpus
- images
- harry_potter_...ssion_full.csv
- harry_potter_aggression.R
- hp_download.py
- names_with_regex.csv
- README.md

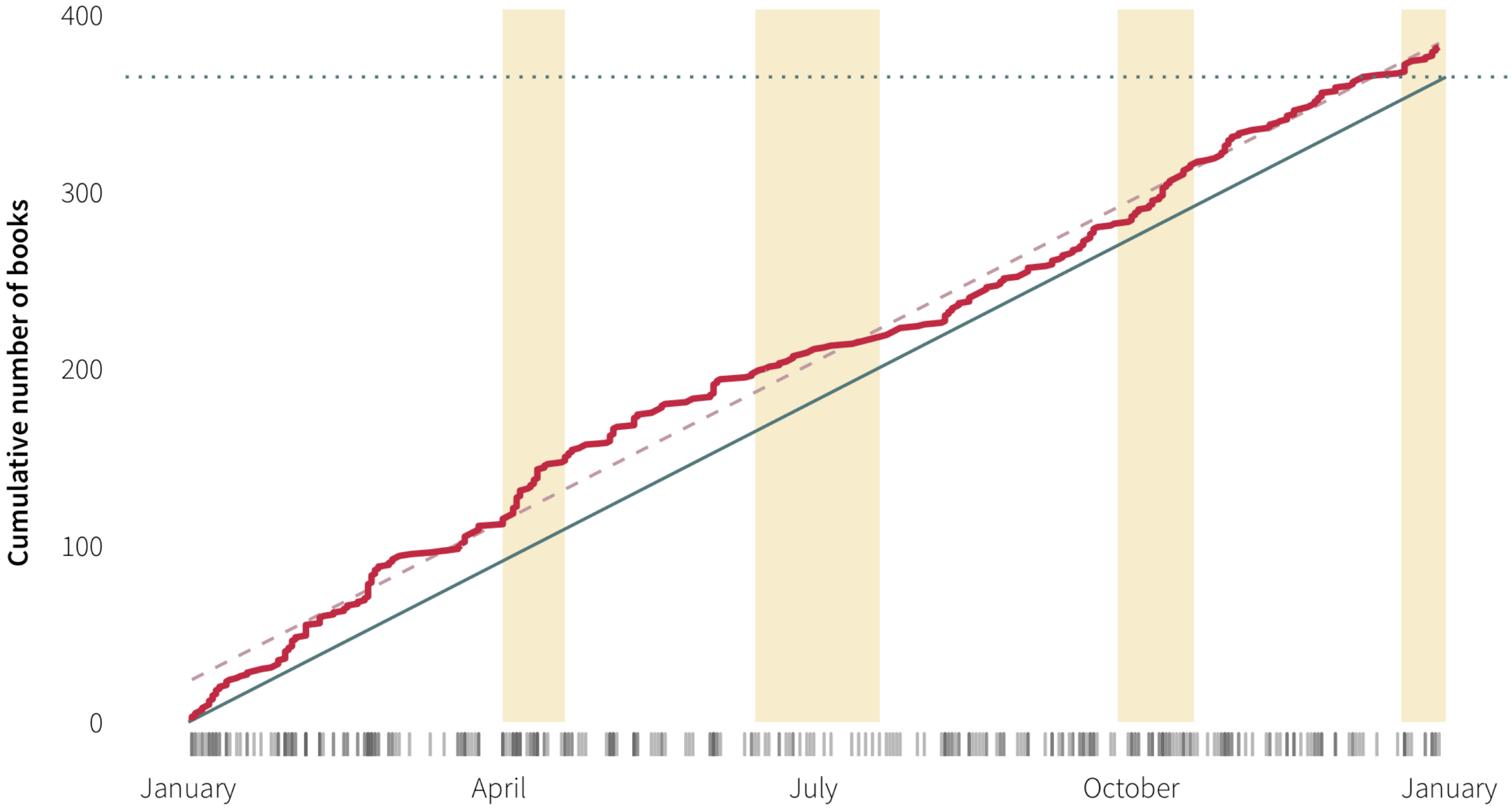
7 items, 293.9 GB available

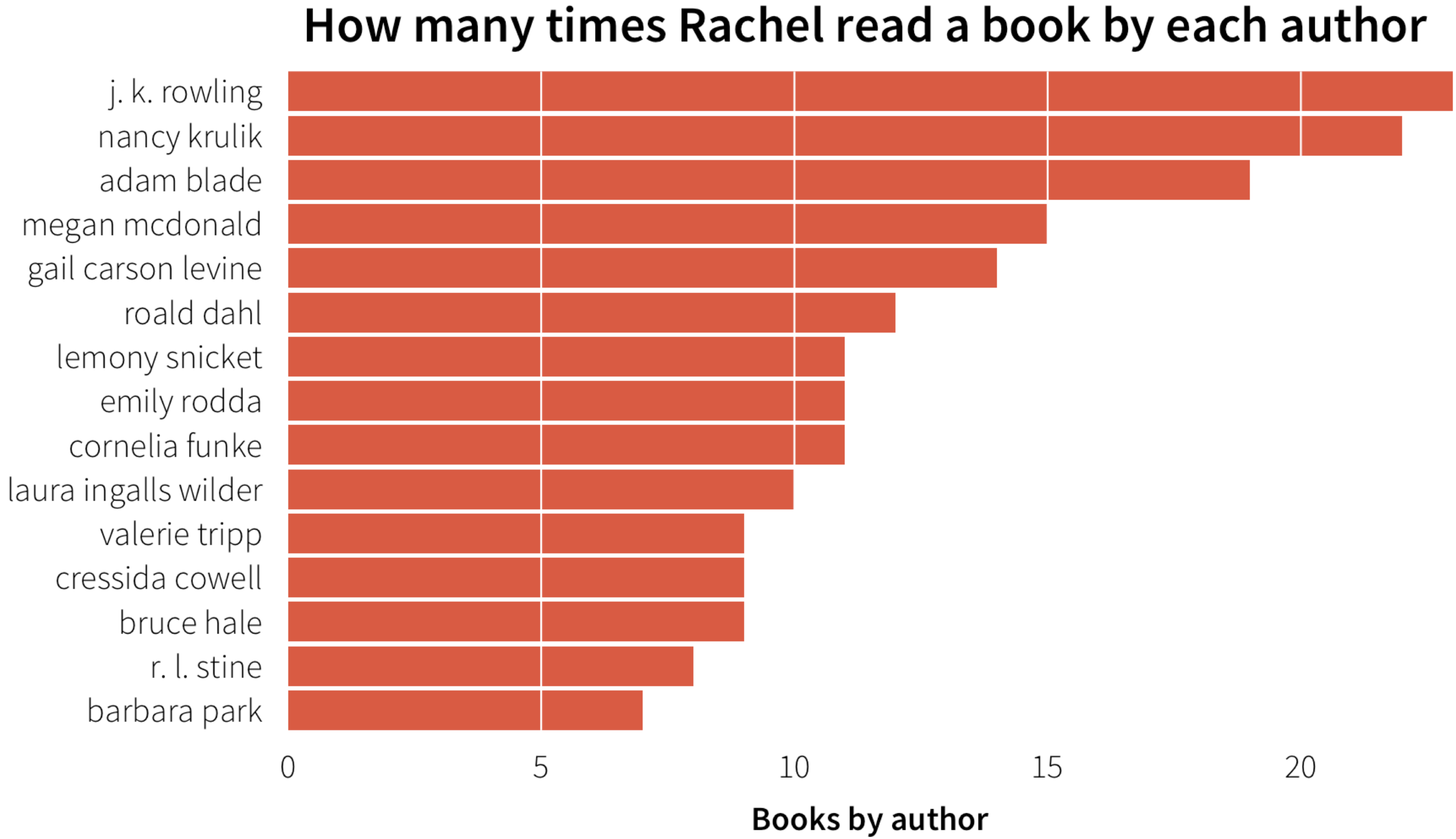
Most aggressive characters in the Harry Potter series



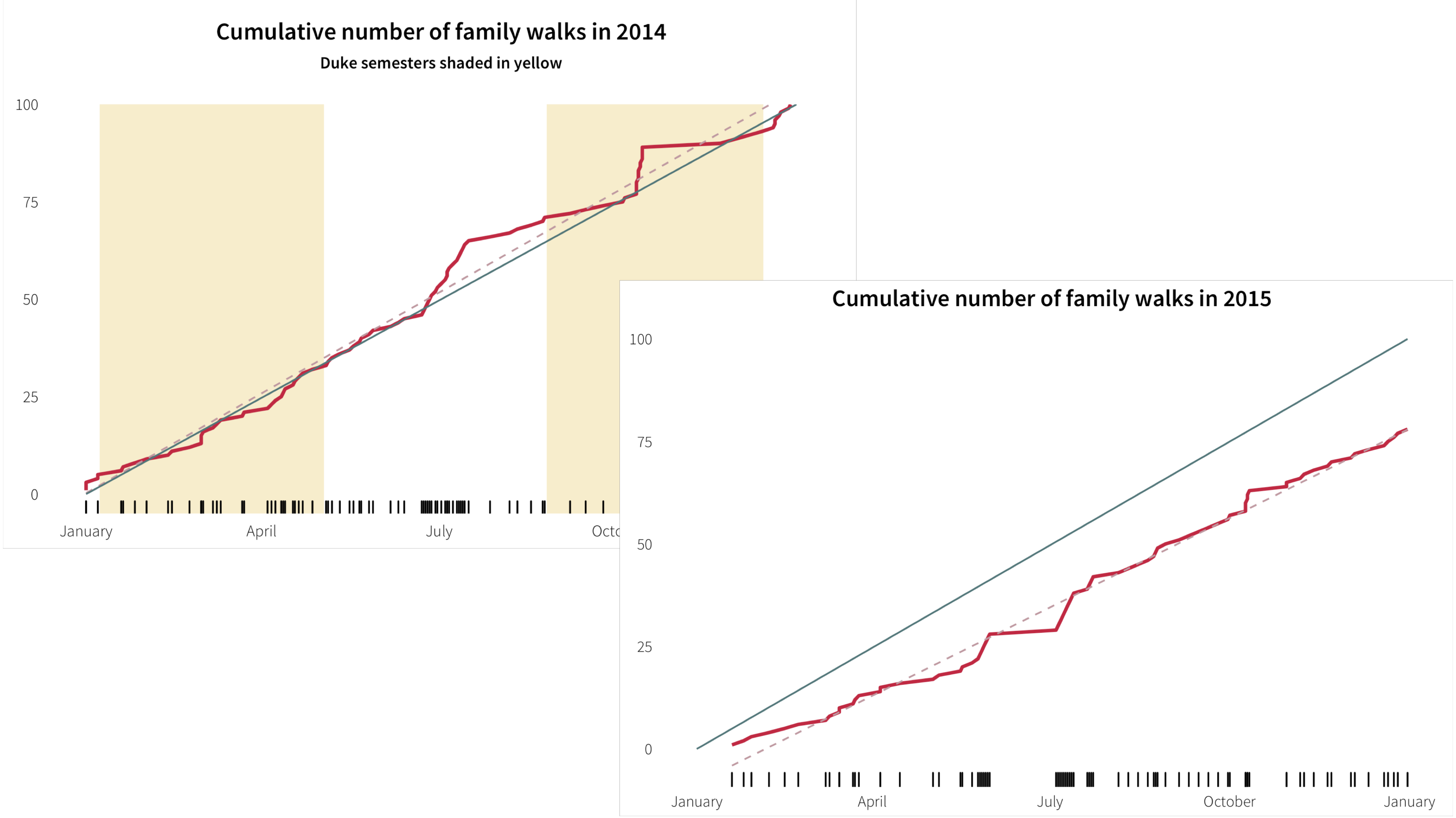
Progress toward 2014 goal

School breaks 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



Cumulative number of family walks in 2014

Duke semesters shaded in yellow

Cumulative number of family walks in 2015

GET OUT IN PUBLIC

Share everything

GitHub + Twitter + websites

#rstats

R User Groups

SLC RUG

#rladies

You are all expert enough now.

Go make stuff!