

FiveThirtyEight's *Data Journalism Workflow*

With 

Andrew Flowers, Quantitative Editor
FiveThirtyEight



@andrewflowers



FiveThirtyEight



• **TheUpshot**

NICAR

the guardian

PRICEONOMICS

So, what is this “data journalism” you speak of?

One answer: Empirical social science **on deadline**.

What do we mean by that?

- Story first, data **follows** (data **rarely** tells a story itself)
- Use rigorous but **interpretable** methods

Mandate:

- Be **accurate**
- Be **fast**
- Be **transparent**

Why do we use R?

Preaching to the choir



Hallelujah!

No, question is: why *should* journalists use R?

While we use other tools, R is the winner

But five big reasons we're fans of R:

1. Open source (transparency, GitHub)
2. ggplot2 (custom theme, weird charts)
3. Data wrangling (speed, messy data)
4. Collaboration (RStudio/Git/GitHub integration)
5. Interactives (Shiny prototypes, data processing)

That said...some staffers use Stata and Excel



(#1) Open source: github.com/fivethirtyeight/data

[fivethirtyeight / data](#)

Unwatch 622 Unstar 3,666 Fork 1,167

Code Issues 7 Pull requests 2 Wiki Pulse Graphs Settings

Data and code behind the stories and interactives at FiveThirtyEight — Edit

598 commits 3 branches 0 releases 24 contributors

Branch: master ▾ New pull request Create new file Upload files Find file Clone or download ▾

andrewflowers committed on GitHub Merge pull request #68 from karasinski/master ...	Latest commit 86f4403 42 minutes ago	
airline-safety	Update README.md	a year ago
alcohol-consumption	Revert "Update drinks.csv"	2 years ago
avengers	add avengers data	a year ago
bad-drivers	add bad drivers data	2 years ago
bechdel	format email address	2 years ago
biopics	for race_known as unknown, make subject_race blank, not White	11 months ago
births	fix README	28 days ago
bob-ross	cleaned up bob ross clustering script	2 years ago
buster-posey-mvp	add buster posey mvp scripts	11 months ago
classic-rock	fixed entries with #REF! excel errors on two rows of classic-rock-raw...	2 years ago

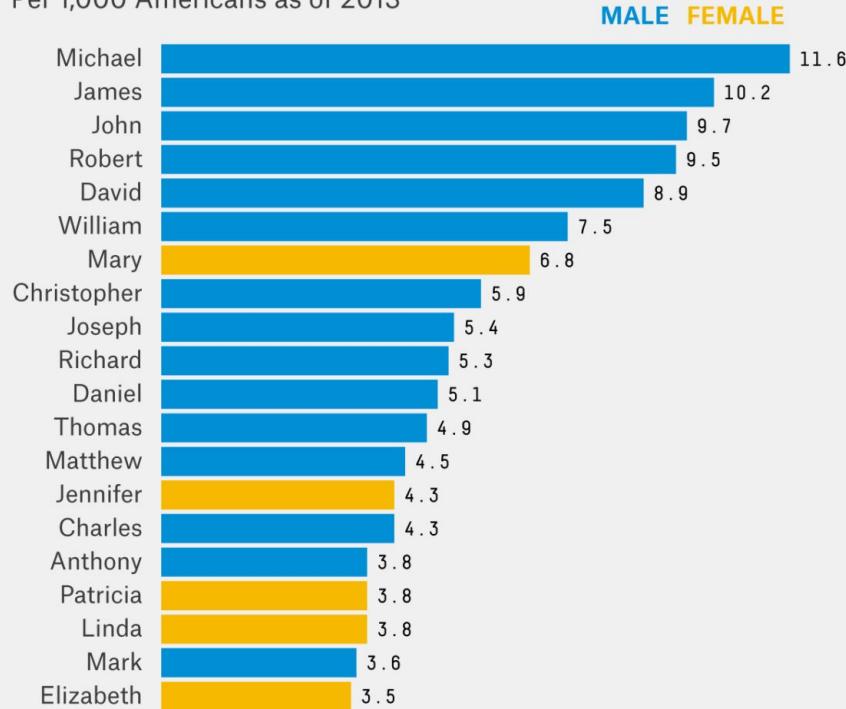
What's the most common name in America?



Michael Smith...or James Smith?

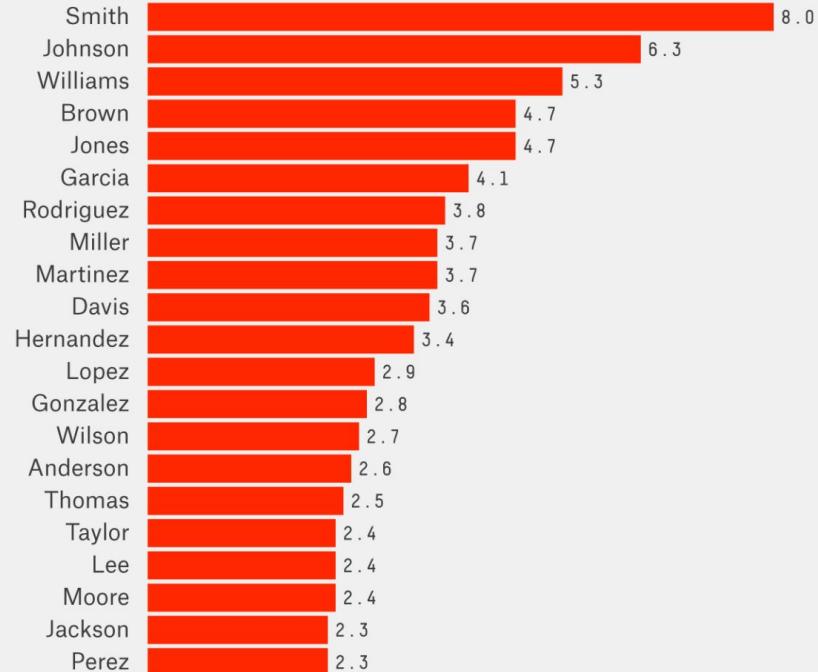
Most Common First Names

Per 1,000 Americans as of 2013



Most Common Surnames

Per 1,000 Americans as of 2013



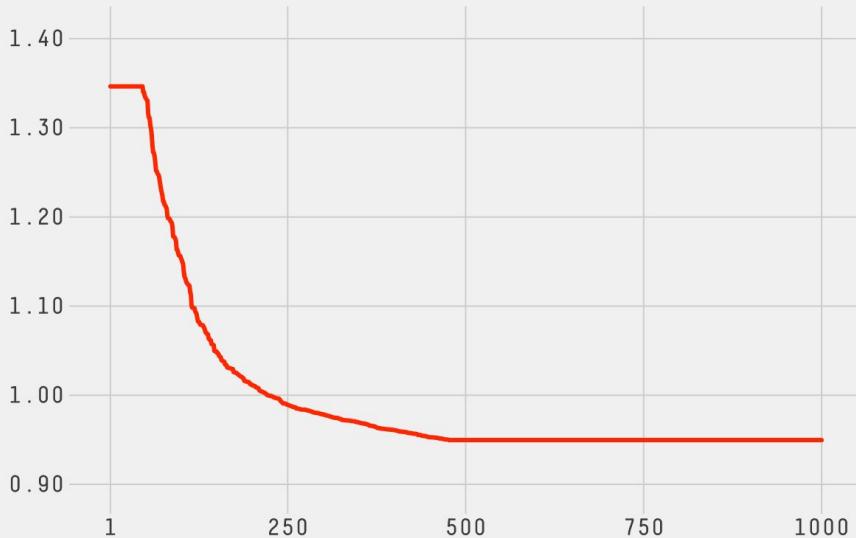
But names aren't consistently correlated

	Miller	Anderson	Martin	Smith	Thompson	Wilson	Moore	White	Taylor	Davis	Johnson	Brown	Jones	Thomas	Williams	Jackson	Lee	Garcia	Martinez	Rodriguez
John	8.4	6.9	13.4	-23.5	3.6	2.5	3.6	1.7	0.4	-3.8	-35.2	-34.3	-26.5	5.1	-11.3	-27.7	-8.3	-66.4	-63.6	-69.5
Michael	18.1	2.4	13.0	5.6	6.7	2.6	15.7	-1.7	0.6	2.5	-4.8	4.2	-2.4	2.9	-8.1	-16.8	-9.8	-54.7	-54.4	-63.2
James	20.3	24.0	35.9	21.1	33.6	40.9	43.0	41.0	37.4	32.8	9.6	25.4	8.9	24.8	20.7	19.8	12.2	-84.4	-83.9	-87.4
Robert	29.6	27.0	16.8	8.7	18.0	22.0	17.2	18.1	39.2	7.2	5.9	14.2	12.6	4.8	-1.3	-1.8	12.1	-57.4	-55.1	-61.0
David	27.9	28.1*	15.8	10.8	11.0	12.6	8.8	16.2	-1.5	-44.7	3.5	10.3	5.0	0.9	-7.3	-15.4	10.1	-30.5	-28.5	-31.7
Mary	21.5	18.2	26.2	14.7	18.5	18.9	22.8	22.5	16.0	14.8	11.7	13.1	12.2	18.9	13.1	8.7	-16.4	-42.4	-41.3	-54.0
William	22.5	7.0	33.3	16.6	29.3	21.5	31.9	38.2	29.2	23.1	4.3	22.3	15.4	19.8	-59.1	9.6	-11.4	-82.3	-80.2	-74.2
Richard	24.1	22.9	12.6	6.2	8.4	1.7	7.3	12.2	5.0	5.2	-2.0	3.9	-4.6	-2.3	-17.0	-21.6	-6.6	-43.2	-43.0	-47.7
Thomas	12.9	3.0	22.1	-2.6	-3.5	9.7	32.2	13.2	4.3	-5.9	-15.0	1.8	-3.5	-81.1	-11.5	-11.0	-8.9	-77.8	-75.6	-84.2
Jennifer	24.9	17.3	22.5	16.7	15.2	14.8	11.5	14.8	14.9	8.7	10.3	6.5	24.0	-1.6	-7.4	-6.3	22.5	-32.9	-32.2	-33.2
Patricia	20.9	10.7	25.3	17.1	16.7	6.7	19.0	17.2	19.4	11.8	7.1	15.2	11.2	15.9	10.6	8.8	-22.9	-16.0	-16.9	-23.3
Joseph	-8.9	-34.8	0.1	-27.6	-24.8	-26.1	-20.4	-10.6	-23.9	-26.7	-33.9	-23.9	-39.0	-7.5	-26.4	-26.0	-26.4	-45.8	-40.3	-52.8
Linda	34.0	25.8	29.6	27.5	24.8	24.4	29.3	26.1	28.6	24.4	22.9	19.3	21.1	17.4	15.3	16.1	6.5	-49.2	-50.0	-56.5
Maria	-77.6	-77.3	-52.1	-78.8	-76.4	-77.4	-79.3	-77.7	-80.3	-79.7	-78.3	-78.5	-79.4	-75.6	-79.4	-80.9	-77.5	663.9	614.1	639.8
Charles	36.3	25.7	29.3	38.4	38.1	45.7	49.0	43.4	41.2	43.6	24.8	39.9	30.5	37.0	33.3	29.0	4.6	-83.7	-84.5	-88.5
Barbara	35.5	25.3	25.1	24.0	20.6	25.8	24.6	25.2	24.7	24.2	20.3	26.7	16.2	16.6	14.8	18.5	-18.7	-66.9	-67.5	-67.9
Mark	42.1	45.9	-38.5	4.1	32.4	22.0	-23.1	-16.8	11.0	7.3	14.7	-14.0	-16.7	2.5	-7.3	-23.4	-33.6	-67.0	-63.0	-71.3
Daniel	19.9	-19.1	10.3	-5.2	-6.4	-13.9	-8.5	-5.0	21.3	-22.2	-17.8	-15.5	-25.1	-25.3	-28.7	-32.3	-1.3	35.1	33.2	24.0
Susan	32.1	28.4	15.8	3.9	5.4	3.5	3.5	-0.9	3.0	-5.8	-6.5	-4.6	-17.1	-1.5	-24.2	-31.9	3.7	-68.2	-68.0	-71.7
Elizabeth	13.3	7.4	13.5	1.7	9.1	7.9	8.4	4.0	1.7	-2.4	-3.7	-3.4	-5.5	-3.3	-13.9	-18.2	-20.2	16.3	23.4	22.5

It's James Smith, but Maria Garcia is rising

Hispanic Correction Factors

For top 1,000 most common first names



FIVETHIRTYEIGHT

Branch: master data / most-common-name / most-common-name.R

andrewflowers most common name data and script

bef7964 on Nov 20, 2014

1 contributor

218 lines (157 sloc) | 10.5 KB

Raw Blame History

```
1 ##### Story: "Dear Mona, What's The Most Common Name In America?"  
2 ##### Url: http://fivethirtyeight.com/features/whats-the-most-common-name-in-america/  
3 ##### Authors: Mona Chalabi (Mona.Chalabi@fivethirtyeight.com) and Andrew Flowers (andrew.flowers@fivethirtyeight.com)  
4  
5 require(babynames)  
6 require(dplyr)  
7 require(reshape2)  
8 require(zoo)  
9 require(datasets)  
10  
11 # Census population parameters  
12 pop2000 <- 276059000 # year 2000 population  
13 pop2013 <- 316128839 # year 2013 population  
14  
15 hispPopShare <- .171 # Hispanic share of overall population  
16 foreignPopShare <- .127 # Foreign-born share of overall population  
17  
18 # Census growth rates by racial categories, 2000-2013  
19 whiteGrowth <- 1.01155164  
20 blackGrowth <- 1.138799977  
21 asianGrowth <- 1.110695106  
22 asianGrowth <- 1.553975166  
23 twoRaceGrowth <- 1.817182595  
24 hispGrowth <- 1.531490233
```



A Walmart employee puts Lionsgate's "The Hunger Games: Catching Fire" Blu-ray Combo Pack and DVD on the rack prior to the midnight release at Walmart on March 6, 2014 in Orange, California. JEROD HARRIS / GETTY IMAGES

APR 1, 2014 AT 1:52 PM

The Dollar-And-Cents Case Against Hollywood's Exclusion of Women

By [Walt Hickey](#)

Filed under [Movies](#)



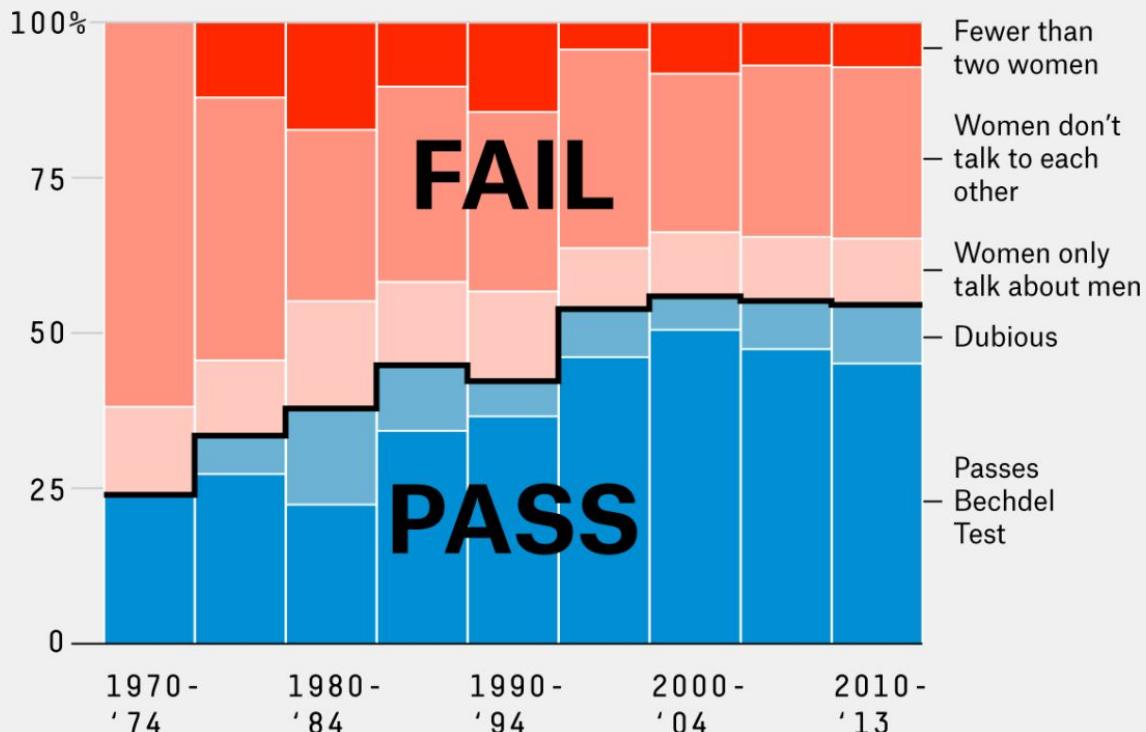
RECOMMENDED

[Trump's Scoring Of Data May Not Hurt Him, But It'll Hurt The GOP](#)

[Senate 2016: The Democrats Strike Back](#)

The Bechdel Test Over Time

How women are represented in movies



Branch: master ▾

data / bechdel / analyze-bechdel.R

[Find file](#) [Copy path](#)



ascheink format email address

3e71708 on Apr 8, 2014

2 contributors

79 lines (57 sloc) | 2.7 KB

[Raw](#)

[Blame](#)

[History](#)



```
1 # Calculates summary statistics and conducts basic regression analysis to determine
2 # whether movies which pass the Bechdel test do better or worse at the box office,
3 # using data from www.bechdeltest.com and www.the-numbers.com
4
5 # By Andrew Flowers <andrew.flowers@fivethirtyeight.com>
6 # See also http://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/
7
8 # Install and load required packages
9 # install.packages(c("gdata", "cwhmisc"))
10 library(gdata)
11 library(cwhmisc)
12
13 # Load data
14 rawData<-read.csv("movies.csv", na.strings="#N/A")
```



Students walk across the campus of UCLA in Los Angeles. KEVORK DJANSEZIAN / GETTY IMAGES

SEP 12, 2014 AT 7:37 AM

The Economic Guide To Picking A College Major

By [Ben Casselman](#)



RECOMMENDED

Senate 2016: The Democrats Strike Back

Why Pennsylvania Could Decide The 2016 Election

MAJOR	# OF MAJORS	EARNINGS (x1,000)			% WORKING IN JOBS		
		MED.	25TH	75TH	PART-TIME	NON-COLLEGE	LOW-PAYING
Petroleum Eng.	2,339	\$110	\$95	\$125	13	19	10
Mining & Mineral Eng.	756	\$75	\$55	\$90	23	42	8
Metallurgical Eng.	856	\$73	\$50	\$105	19	28	0
Naval Architecture & Marine Eng.	1,258	\$70	\$43	\$80	12	16	0
Chemical Eng.	32,260	\$65	\$50	\$75	18	20	4
Nuclear Eng.	2,573	\$65	\$50	\$102	11	37	13

Branch: master ▾

data / college-majors / college-majors-rscript.R

[Find file](#)[Copy path](#)**BenCasselman** Create college-majors-rscript.R

613c858 on Sep 11, 2014

1 contributor

166 lines (148 sloc) | 9.17 KB

[Raw](#)[Blame](#)[History](#)

```
1 #####  
2 #  
3 # COLLEGE MAJORS AND EARNINGS  
4 # This is the code used to generate data for FiveThirtyEight's  
5 # story on earnings by college major.  
6 # Analysis is based off the 2010-2012 American Community Survey  
7 # microdata.  
8 # Download data here: http://www.census.gov/acs/www/data_documentation/pums_data/  
9 # Documentation here: http://www.census.gov/acs/www/data_documentation/pums_documentation/  
10  
11 # First download data and select records for which college major  
12 # (variable FOD1P) is present. Save into data frame as MAJORS1012.  
13 # Also download github_majorslist.csv  
14  
15 require(dplyr)  
16  
17 load("MAJORS1012")  
18 MajorsList <- read.csv("github_majorslist.csv",header=TRUE,stringsAsFactors=FALSE)  
19  
20 working <- merge(MAJORS1012,MajorsList,by="FOD1P") # Check merge properly  
21 MAJORS1012 <- working  
22 rm(working)
```

NOV 14, 2015 AT 6:26 PM

The Rise Of Terrorism Inspired By Religion In France

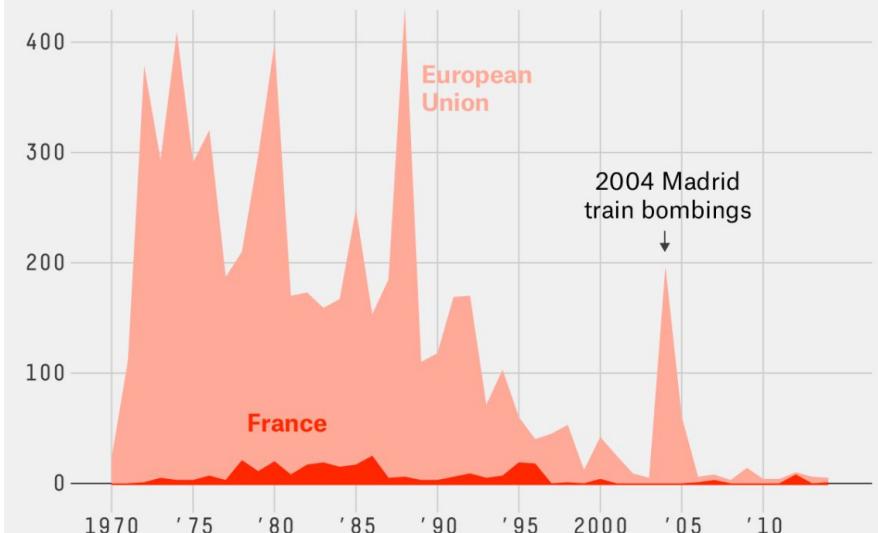
By [Carl Bialik](#)

Filed under [Terrorism](#)



Previous Terrorism Deaths In France And The EU

European Union member countries as of 1986



FIVETHIRTYEIGHT

SOURCE: GLOBAL TERRORISM DATABASE

```
30 ##### France analysis #####
31
32 # Look at incidents in France
33 france <- rawData %>% filter(country_txt=="France")
34 dim(france)
35
36 # Incidents by year -- 1993 is missing from this data
37 table(france$iyear, useNA="ifany")
38 france %>% group_by(iyear) %>%
39   summarize(incidents=n()) %>%
40   arrange(desc(iyear))
41
42 # Fatalities by year
43 fraFatByYear <- france %>% group_by(iyear) %>%
44   summarize(fatalities=sum(nkill, na.rm=T)) %>%
45   arrange(desc(fatalities))
46
47 # Add in 1993 data -- 5 fatalities in France
48 fraFatByYear <- rbind(fraFatByYear,
49                       data.frame(iyear=1993,
50                                   fatalities=stats1993[match("France", stats1993$Country),]$`Number Killed`))
51
52 # Analysis: France had 274 fatalities from terrorism incidents between 1972 and 2014.
53 fraFatByYear
54 sum(fraFatByYear$fatalities, na.rm=T)
```

JAN 6, 2016 AT 6:00 AM

You Can't Trust What You Read About Nutrition

We found a link between cabbage and innie bellybuttons, but that doesn't mean it's real.

By [Christie Aschwanden](#)

Nutrition Studies

This directory contains data and code behind the story [You Can't Trust What You Read About Nutrition](#).

Many studies of diet and nutrition include multiple variables with vast amounts of data, making it easy to p-hack your way to sexy (and false) results. We learned this firsthand when we invited readers to take a survey about their eating habits known as the food frequency questionnaire and answer a few other questions about themselves. We ended up with 54 complete responses and looked for associations much as researchers look for links between foods and dreaded diseases. It was easy to find them.

Warning: This is evil (statistical) work. Do not go to the dark side. Do not try this at home.

SPURIOUS CORRELATION A

Cabbage

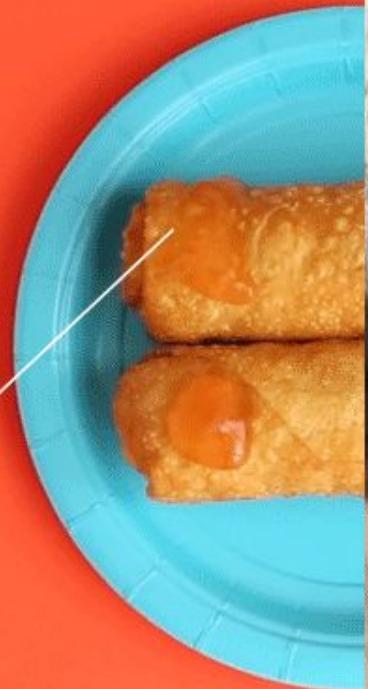


Innies



SPURIOUS CORRELATION B

Egg rolls

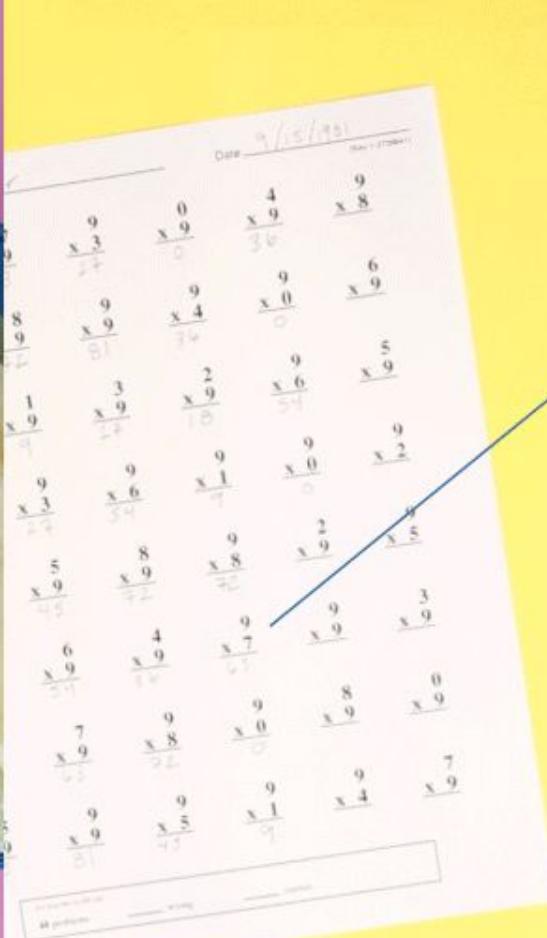


Dog ownership



SPURIOUS CORRELATION C

Potato chips



Higher scores
in math

(#2) ggplot2

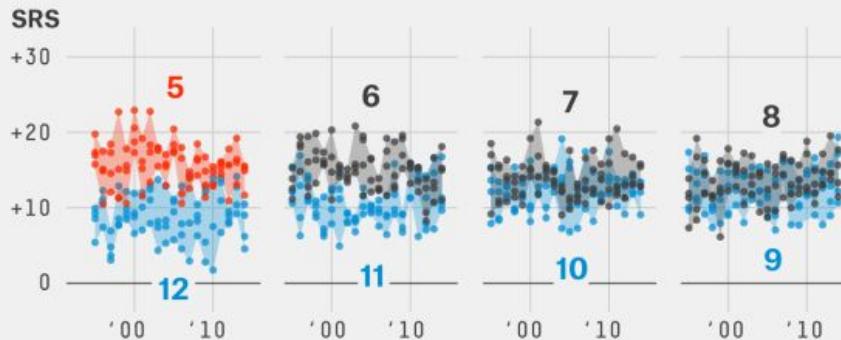
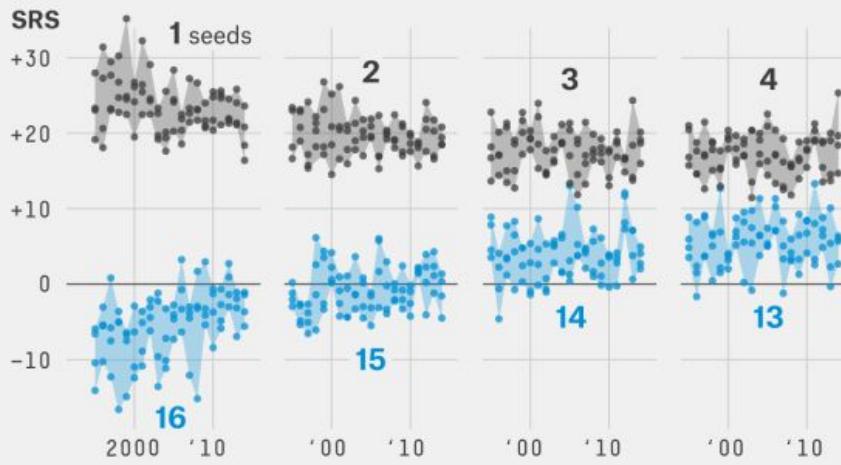
ggplot is awesome! (duh!)

Why it works so well for us?

- It's within R
- It allows us to make **weird** charts
- Custom theme
- Pass the SVG/PDF files to Adobe Illustrator

The Closeness Of Five And 12

The SRS of every team playing in the NCAA tournament since 1995 at the start of its round-of-64 matchup



The Oldest Persons In The World

KEY

Each line represents
somebody who was the
oldest person in the world

BECOMES
OLDEST
PERSON

PASSES
AWAY

115
yrs. old

110

1960

'70

'80

'90

2000

'10

Jeanne Calment

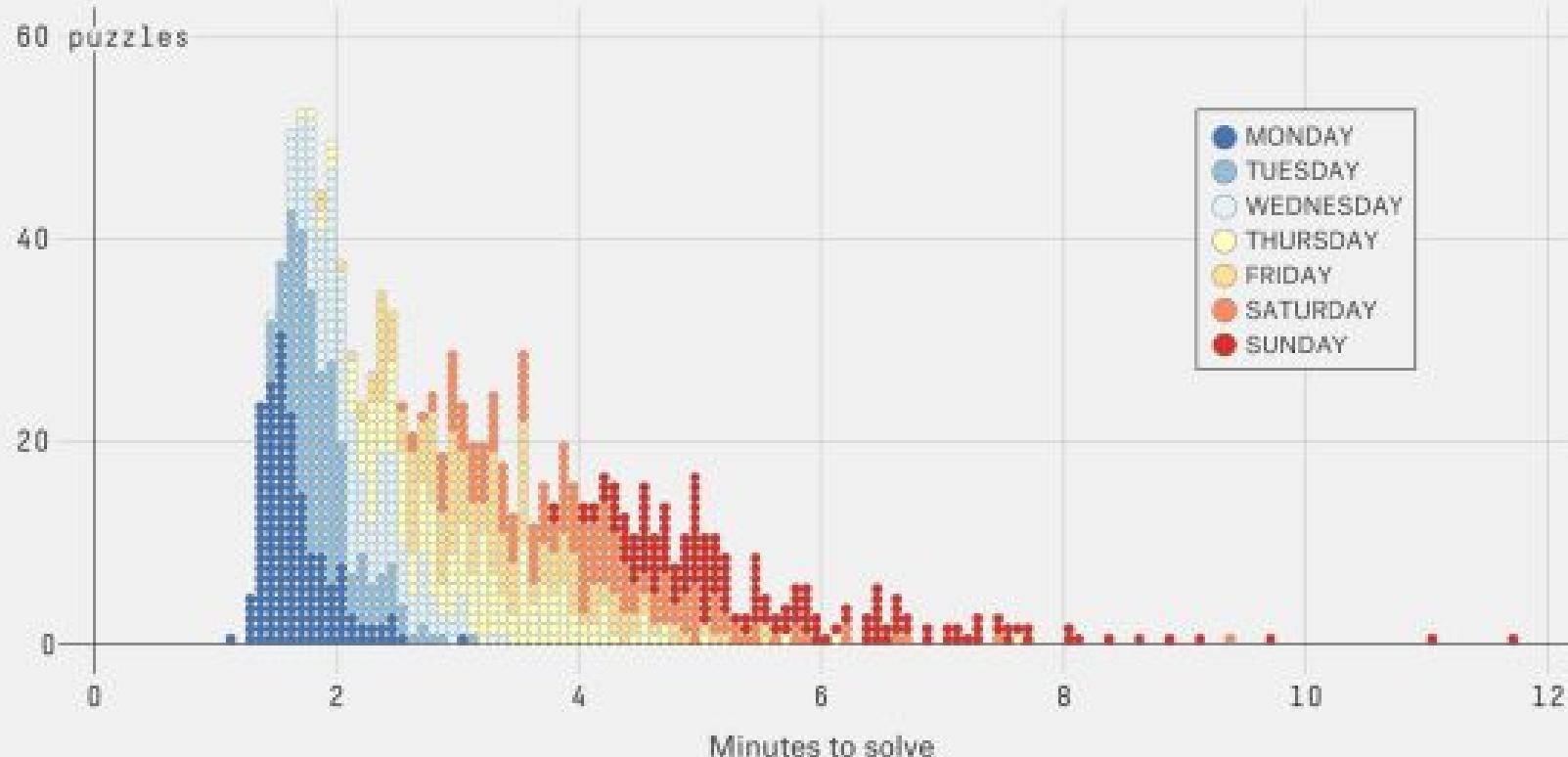
FRANCE

Lived to 122 years
and 164 days

Since 2000, the
average tenure of the
oldest living person has
shortened, as has the
age gap between her
and her successor.

The Puzzling Speed Of Dan Feyer

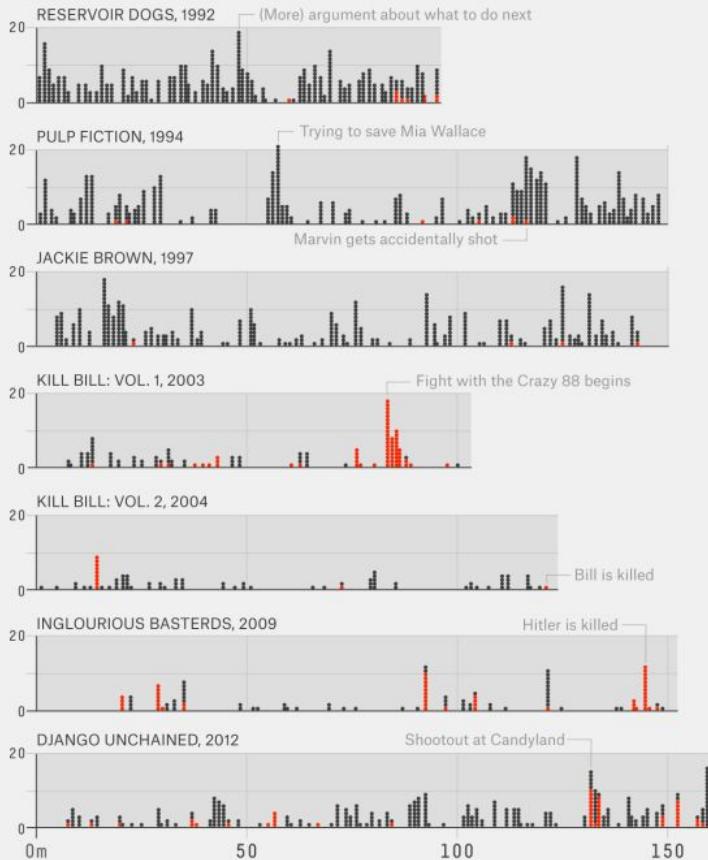
Solve times for the past 1,208 New York Times crossword puzzles, by day of the week



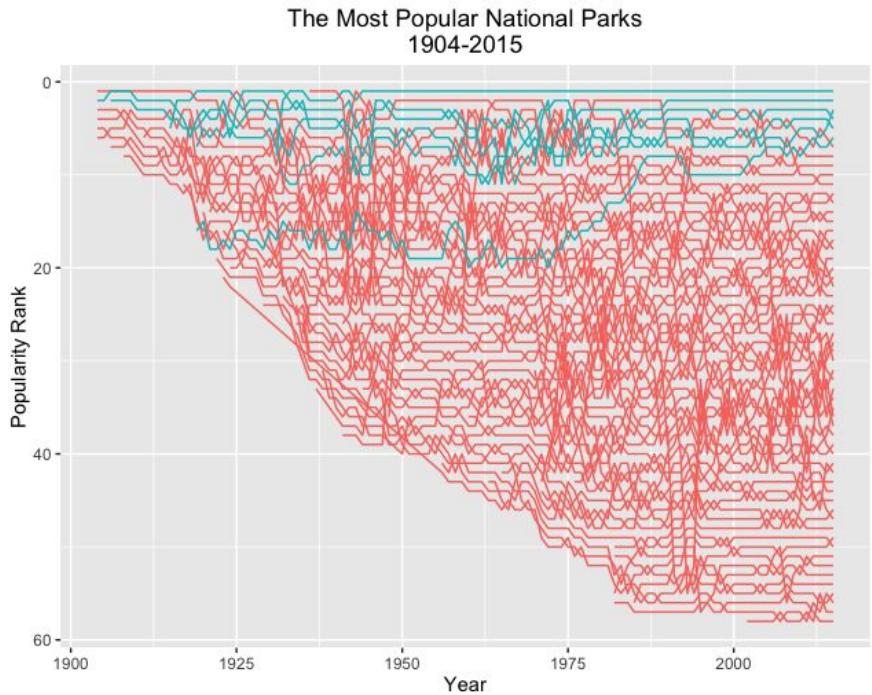
The complete obscene guide to Tarantino

Time stamp of every instance of profanity and each death in feature films directed by Quentin Tarantino

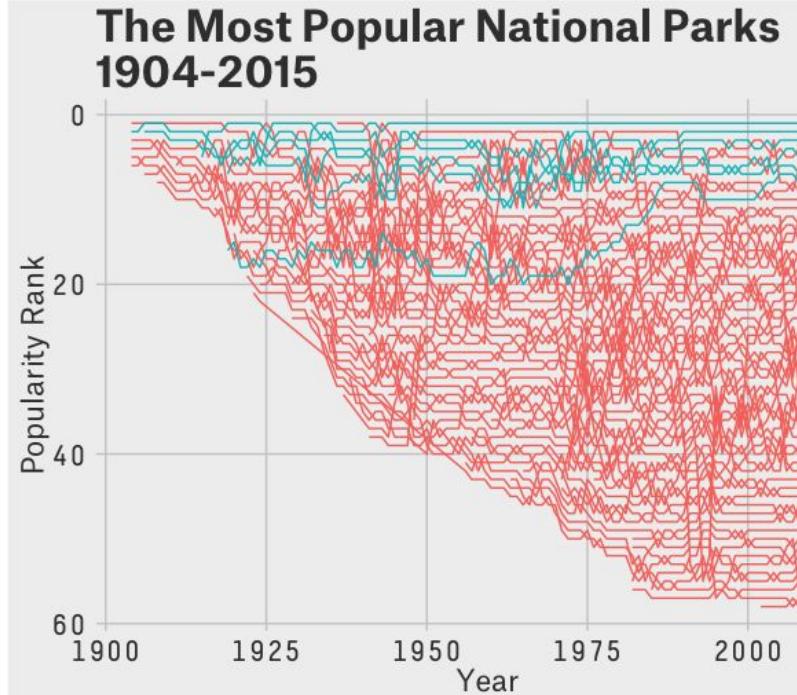
- PROFANITY
- DEATH



Custom ggplot theme

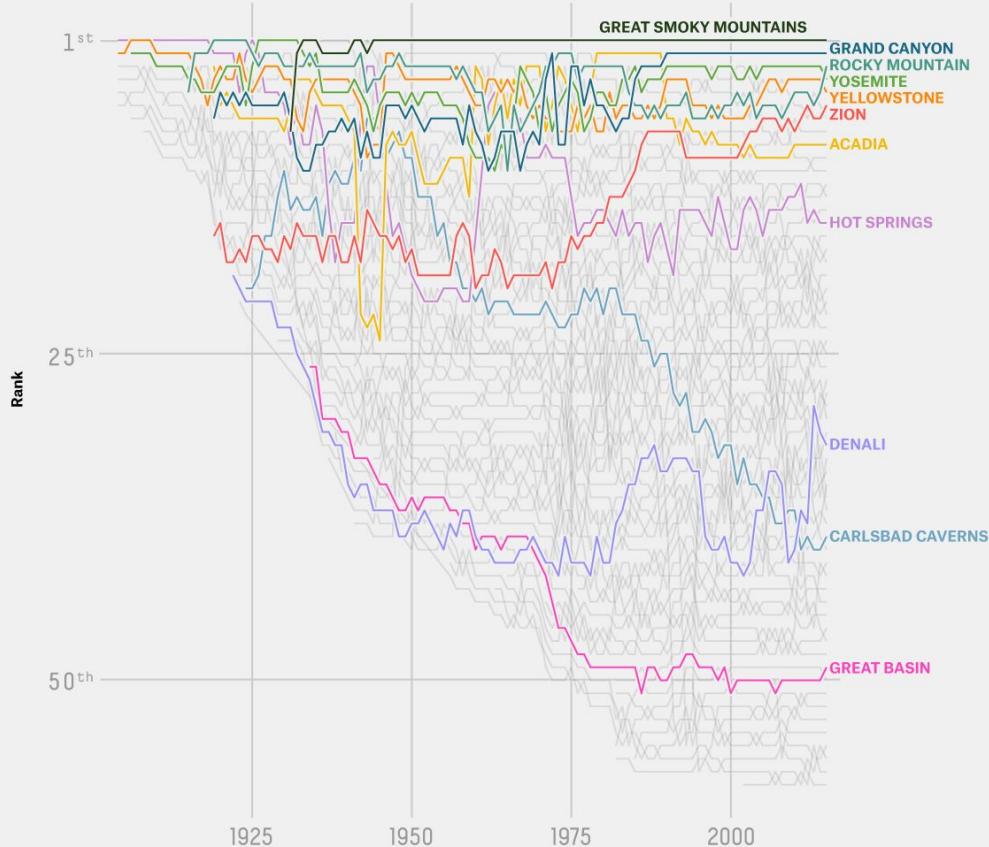


theme538 PRIVATE



The most popular national parks

National parks ranked by number of visitors in a given year



(#3) Data wrangling



Data journalism means exposure to LOTS of different (and messy) data

To be fast, we use:

- dplyr (should win a Nobel peace/economics/literature prize)
- tidyr, lubridate, stringr, readr, etc.

FiveThirtyEight has an internal R learning group

The screenshot shows an RStudio interface with the following components:

- Environment View:** Displays a table titled "all_elos_data" with columns: game_id, lg_id, year_id, date_game, is_playoffs. Rows 1 through 7 are shown, all corresponding to NBA games from November 1, 1946.
- Code Editor:** Shows two examples of filtering data:
 - # Without the pipe operator
filter(all_elos_data, year_id==2015)
 - # With pipe operator
all_elos_data %>% filter(year_id==2015)
- Console:** Shows the command "all_elos_data %>% filter(year_id==2015) %>% head()" and its output:
 - Source: local data frame [6 x 11]
 - Variables not shown: elo_i (dbl), elo_n (dbl), opp_id (chr), opp_pts (int)
 - Data (6 rows):

game_id	lg_id	year_id	date_game	is_playoffs	fran_id	pts
1	201410280LAL	NBA	2015 10/28/2014	0	Rockets	108
2	201410280LAL	NBA	2015 10/28/2014	0	Lakers	90
3	201410280NOP	NBA	2015 10/28/2014	0	Pelicans	101
4	201410280NOP	NBA	2015 10/28/2014	0	Magic	84
5	201410280SAS	NBA	2015 10/28/2014	0	Spurs	101
6	201410280SAS	NBA	2015 10/28/2014	0	Mavericks	100
- Tools View:** Shows a vertical toolbar with icons for various RStudio features like file operations, help, and help.
- Bottom Status Bar:** Shows the time as 17:41 / 1:03:47 and a message from Google Hangouts.

Example: Uber vs. Taxis in NYC

Data obtained via FOIA request:
<https://github.com/fivethirtyeight/uber-tlc-foil-response>

Uber TLC FOIL Response

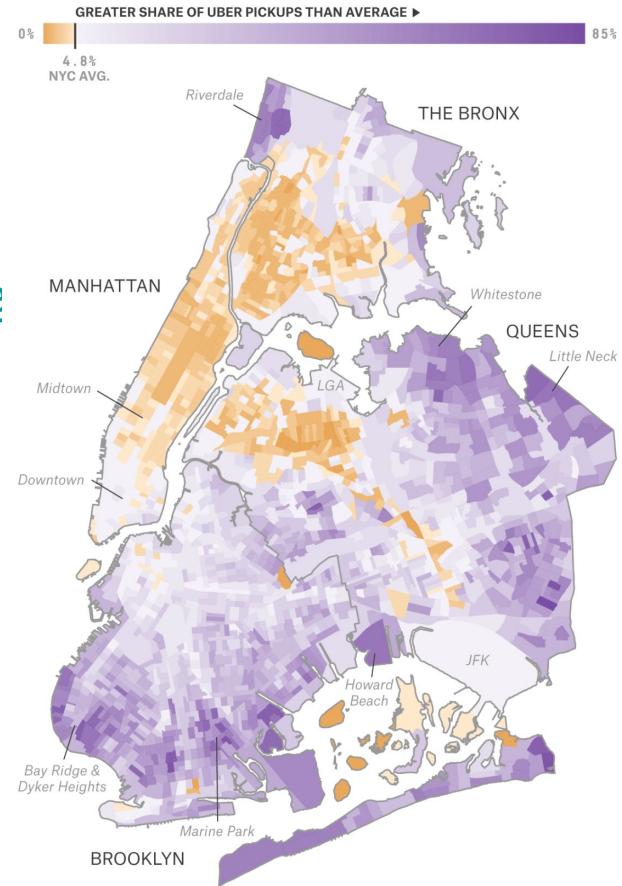
This directory contains data on over 4.5 million Uber pickups in New York City from April to September 2014, and 14.3 million more Uber pickups from January to June 2015. Trip-level data on 10 other for-hire vehicle (FHV) companies, as well as aggregated data for 329 FHV companies, is also included. All the files are as they were received on August 3, Sept. 15 and Sept. 22, 2015.

FiveThirtyEight obtained the data from the [NYC Taxi & Limousine Commission \(TLC\)](#) by submitting a Freedom of Information Law request on July 20, 2015. The TLC has sent us the data in batches as it continues to review trip data Uber and other FHV companies have submitted to it. The TLC's correspondence with FiveThirtyEight is included in the files

[TLC_letter.pdf](#), [TLC_letter2.pdf](#) and [TLC_letter3.pdf](#). TLC records requests can be made [here](#).

This data was used for four FiveThirtyEight stories: [Uber Is Serving New York's Outer Boroughs More Than Taxis Are](#), [Public Transit Should Be Uber's New Best Friend](#), [Uber Is Taking Millions Of Manhattan Rides Away From Taxis](#), and [Is Uber Making NYC Rush-Hour Traffic Worse?](#).

New York City's Edges Are Uber-Heavy
Share of all Uber, yellow cab and green cab pickups that were by Ubers from April through September 2014, by census tract



(#4) Collaboration

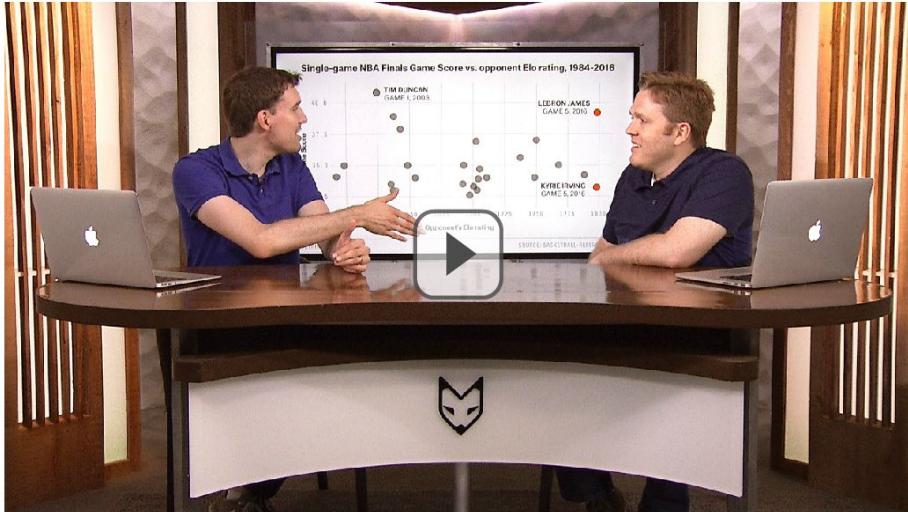
- RStudio's nifty Git/GitHub integration is great for collaboration
- But it's also a life-saver during the journalist's editing process

JUN 14, 2016 AT 2:21 PM

LeBron, Kyrie And Draymond All Made History In The Cavs' Win

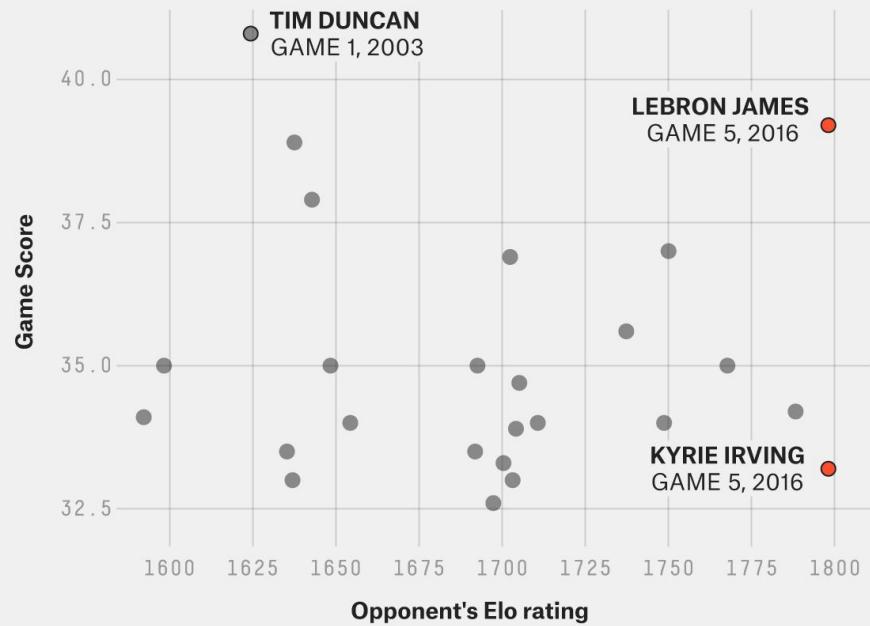
By [Neil Paine](#) and [Andrew Flowers](#)

Filed under [NBA Playoffs](#)



LeBron's historic night (Kyrie wasn't bad, either)

Single-game NBA Finals Game Score vs. opponent Elo rating, 1984-2016



MAY 26, 2016 AT 2:04 PM

The Thunder Are On A Historic Playoff Run

By [Andrew Flowers](#) and [Neil Paine](#)

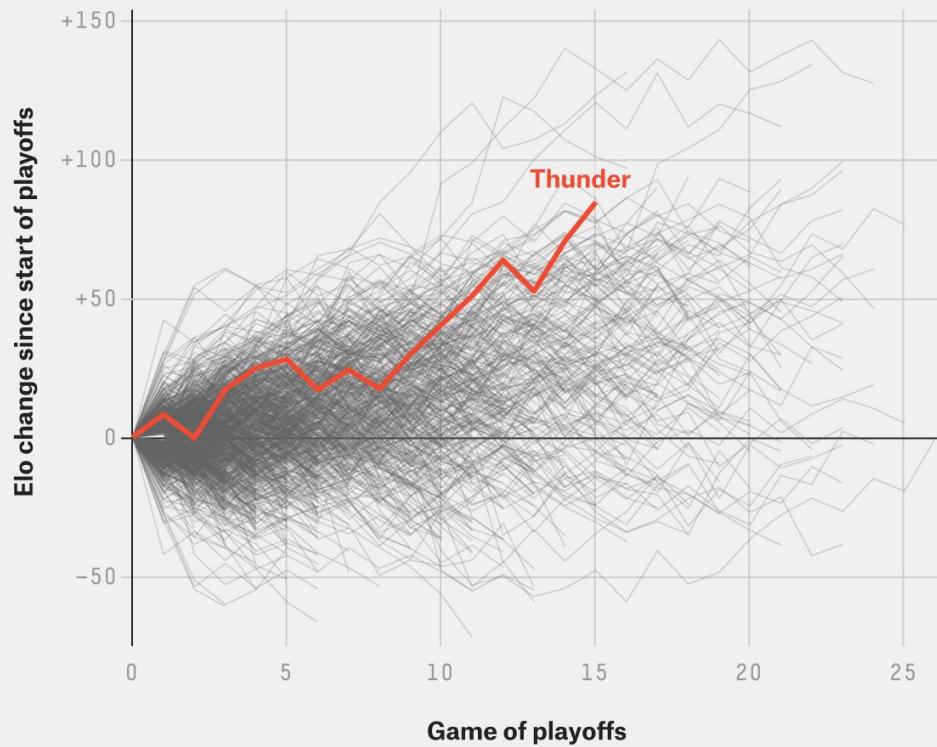
Filed under [NBA Playoffs](#)



The Thunder have run through the playoffs like few others. ANDREW BERNSTEIN / NBAE VIA GETTY IMAGES

The Thunder's historic playoff run

Elo change since start of playoffs for all NBA playoff runs from 1947 through May 24, 2016



Durant And Westbrook Have Never Faced A Defense Like The Spurs'

By [Andrew Flowers](#) and [Neil Paine](#)

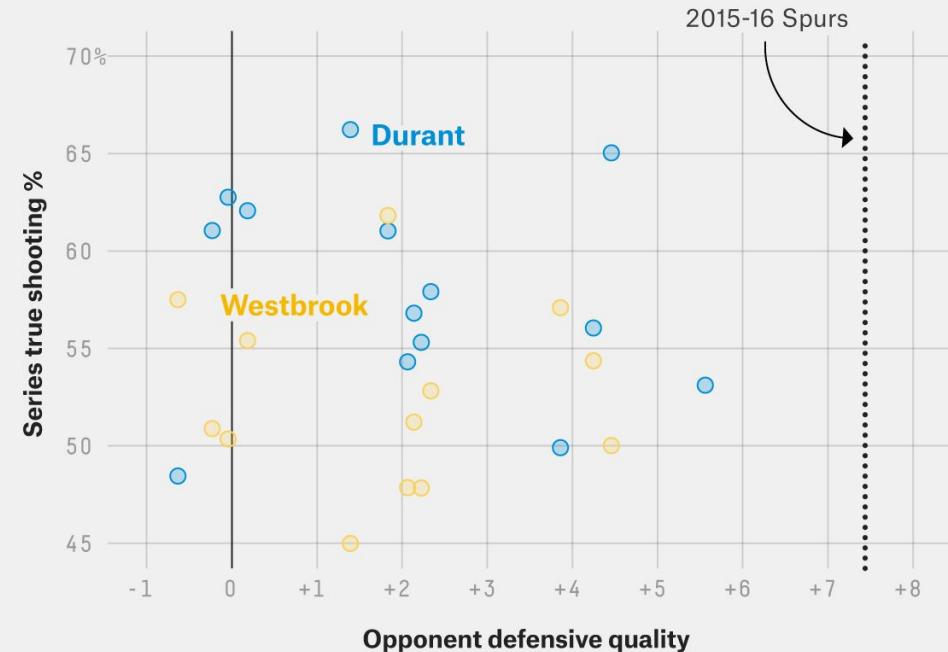
Filed under [NBA Playoffs](#)



Kawhi can't check both. EZRA SHAW / GETTY IMAGES

KD and Russ haven't faced a D this dominant

True shooting percentage vs. opponent defensive rating (relative to NBA average) for Kevin Durant and Russell Westbrook in every playoff series of their careers, 2010-16



 NeilPaine538 More commits.

Latest commit 27d3e20 13 days ago

 .gitignore	Initial commit.	14 days ago
 2016_NBA_games.csv	Add Kimono scrape, which somehow worked after all.	13 days ago
 NBA_BPM_talent_2016.csv	Added player talent file.	14 days ago
 NBA_finals_G6_preview.Rproj	Initial commit.	13 days ago
 README.md	Update README.md	13 days ago
 all_player_game_logs_2015-16.csv	combined csv of all player game logs	13 days ago
 kyrie-chart.pdf	More commits.	13 days ago
 kyrie_plot.pdf	More commits.	13 days ago
 nba-elo-2016-06-13.csv	Initial commit.	14 days ago
 nba-finals-boxscore-data.csv	Add Finals boxscore data.	14 days ago

Branch: master

NBA_finals_G6_preview / scrape_bbr_player_logs.R

Find file Copy path

Unlike Excel, editors can check a writer's data work if they use R.

Fewer mistakes, fewer corrections.

 andrewflowers fix read.csv() and date column of gamelogs

78e1af9 13 days ago

1 contributor

37 lines (27 sloc) | 1.51 KB

Raw Blame History

```
1 # Scrape player logs for 2015-2016 seasons
2
3 setwd("~/repos/NBA_finals_G6_preview/")
4
5 require(readr)
6 require(dplyr)
7 require(lubridate)
8 require(ggplot2)
9 require(rvest)
10 require(stringr)
11
12 # Set url
13 base_url <- 'http://www.basketball-reference.com/play-index/pg1_finder.cgi?request=1&player_id=&match=game&year_min=2016&year_max=2016&age='
14 offsets <- seq(17500, 27900, 100)
15 scraped_data <- data.frame()
16
17 for (offset in offsets){
18
19   #offset <- 27900
20   scrape_url <- paste0(base_url, as.character(offset))
21 }
```

(#5) Interactives

- Our amazing interactive team creates award-winning D3.js graphics
- They sometimes use R for prototyping (using Shiny)
- Also, R scripts do key data processing behind our interactives

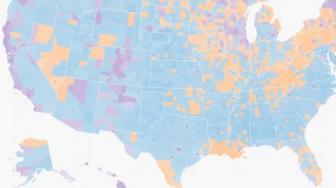
The Facebook Primary

Where 2016 presidential candidates are winning the battle for likes.

By MATTHEW CONLEN and REUBEN FISCHER-BAUM

UPDATED APRIL 18, 2016

THE FACEBOOK LEADERS ARE LAPING THE FIELD



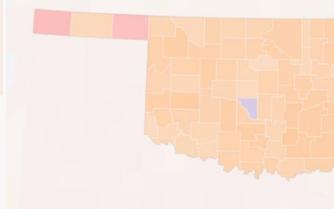
HOW THE 'ESTABLISHMENT' GOP CONTENDERS STACK UP



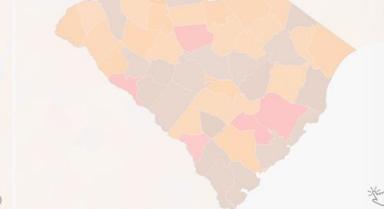
TRUMP VS. CLINTON IN NYC



COLLEGE TOWNS ARE CRAZY FOR SANDERS



WHICH TOP REPUBLICANS ARE STRONG IN S.C.?



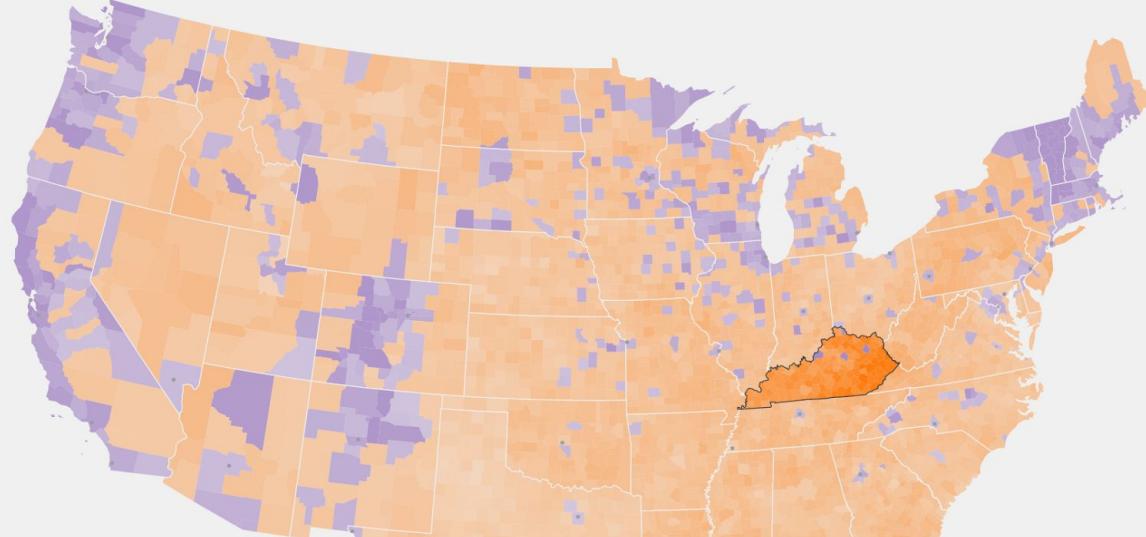
Kentucky

CANDIDATE	LIKES	VS. U.S.
✓ Trump	30%	+6
Carson	23%	-5
✓ Sanders	17%	-8
Cruz	11%	-
Rubio	7%	-
✓ Clinton	6%	-3
Kasich	1%	-
Bush	1%	-

ⓘ Click candidate names to select

Share of likes Share vs. U.S.

Map shows which candidate has the greatest share of Facebook likes by county



A History Of Sumo

By Matthew Conlen



ROUND-BY-ROUND PROBABILITIES

Bracket Table



What Would It Take To Turn Red States Blue ?

By AARON BYCOFFE and DAVID WASSERMAN

Change the settings below to see how shifts in party preference and turnout by different demographic groups would affect the 2016 presidential election. [Read more »](#)

PUBLISHED 2:00 PM EST | DEC 3, 2015

DEMOCRATS

ELECTORAL VOTES: 332 ✓
POPULAR VOTE: 51 . 7%



REPUBLICANS

ELECTORAL VOTES: 206
POPULAR VOTE: 46 . 6%



COLLEGE-EDUCATED
WHITE

NON-COLLEGE-EDUCATED
WHITE

BLACK

HISPANIC/LATINO

ASIAN/OTHER



Team-By-Team Forecast

Forecast from Oct. 27 (preseason) ▾

ELO	CARM-ELO	1-WEEK CHANGE	TEAM	CONF.	AVG. SIMULATED SEASON		PLAYOFF CHANCES		
					RECORD	POINT DIFF/G	MAKE PLAYOFFS	TOP SEED	WIN TITLE
1645	1732	+	Cavaliers	East	63-19	+9.3	>99%	77%	28%
1743	1731	-	Warriors	West	60-22	+8.1	98%	32%	18%
1564	1690	+	Thunder	West	58-24	+7.0	97%	21%	11%
1667	1685	+	Spurs	West	57-25	+6.6	97%	20%	11%
1647	1672	+	Clippers	West	56-26	+6.0	95%	13%	9%
1617	1636	+	Rockets	West	52-30	+4.5	88%	7%	5%
1583	1609	+	Grizzlies	West	50-32	+3.5	81%	4%	3%
1520	1573	+	Celtics	East	49-33	+3.0	90%	7%	2%
1570	1564	-	Bulls	East	48-34	+2.6	88%	5%	2%
1521	1555	+	Pelicans	West	45-37	+1.5	62%	<1%	1%
1562	1543	+	Hawks	East	46-36	+2.0	85%	4%	2%
1543	1542	-	Jazz	West	45-37	+1.7	65%	1%	1%
1502	1526	+	Raptors	East	44-38	+1.1	78%	3%	1%
1536	1498	+	Wizards	East	41-41	-0.1	66%	1%	<1%
1544	1489	+	Mavericks	West	38-44	-1.2	29%	<1%	<1%

```

if (nrow(nbaallello) > nrow(old_games)) {
  nbaallello$date_game <- as.Date(nbaallello$date_game, '%m/%d/%Y')

  # estimate when the 2016 season will end
  start <- as.Date('10/28/2015', '%m/%d/%Y');
  end <- as.Date('6/21/2016', '%m/%d/%Y');
  estimate <- as.numeric(end - start) + 1;

  nba_elo <-
    nbaallello %>%
    subset(lg_id == 'NBA') %>%
    group_by(year_id) %>%
    mutate(min_date = min(date_game)) %>%
    mutate(max_date = max(date_game)) %>%
    mutate(length = as.numeric(max_date - min_date) + 1) %>%
    #mutate(length = as.numeric(max_date - min_date) + 1) %>%
    mutate(timeline = year_id + (as.numeric(date_game - min(date_game)) + 1)/length) %>%
    transform(timeline = round(timeline,3)) %>%
    transform(elo_n = round(elo_n,2)) %>%
    ungroup()

  aba_elo <-
    nbaallello %>%
    subset(lg_id == 'ABA') %>%
    group_by(year_id) %>%
    mutate(min_date = min(date_game)) %>%
    mutate(max_date = max(date_game)) %>%
    mutate(length = as.numeric(max_date - min_date) + 1) %>%
    #mutate(length = ifelse(year_id != 2015, as.numeric(max_date - min_date) + 1, estimate)) %>%
    mutate(timeline = year_id + (as.numeric(date_game - min(date_game)) + 1)/length) %>%
    transform(timeline = round(timeline,3)) %>%
    transform(elo_n = round(elo_n,2)) %>%
    ungroup()
}

```

Conclusion

- R is great for data journalism because:
 - Free and open source = more transparency
 - Great for graphics (ggplot!)
 - Best data wrangling tools on the planet
 - Improves accuracy
 - One environment for reporters to work a story A-to-Z
- More FiveThirtyEight R news to come:
 - Our first package?
 - Stay tuned...