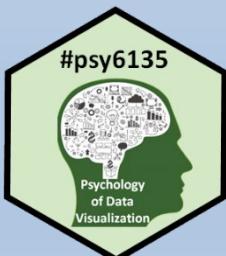
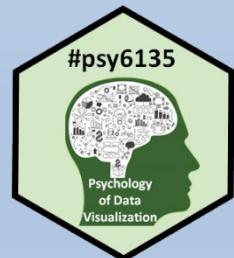


Data Journalism

Michael Friendly Psych 6135

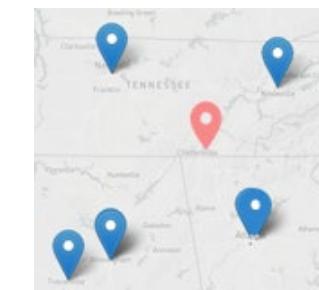
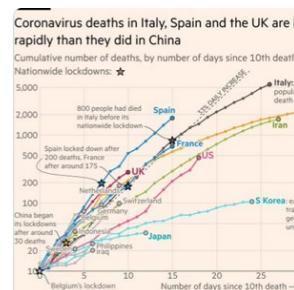
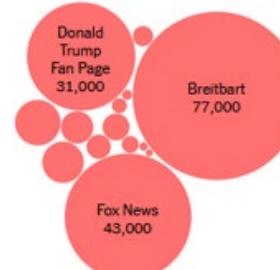
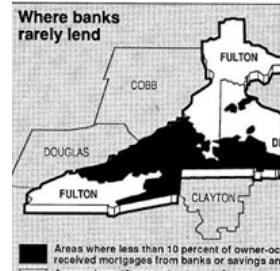


<https://friendly.github.io/6135/>



Topics

- What is DDJ?
- Some early examples
- Modern examples
 - election reporting
 - health statistics
 - social media
- Best examples??
 - Panama Papers
- Satellite journalism



What is data journalism?

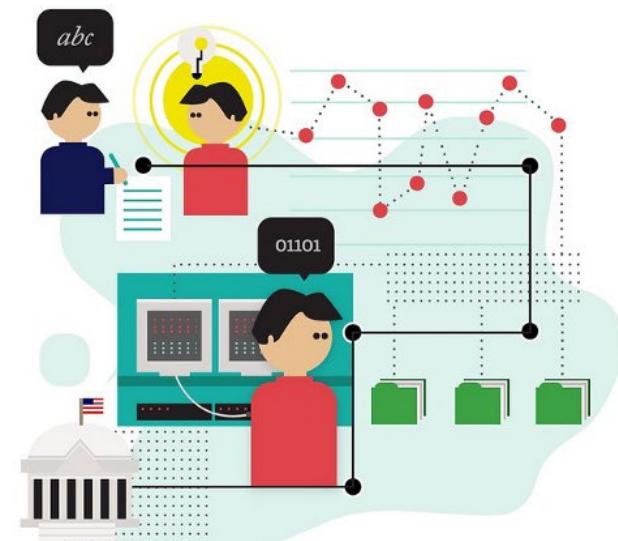
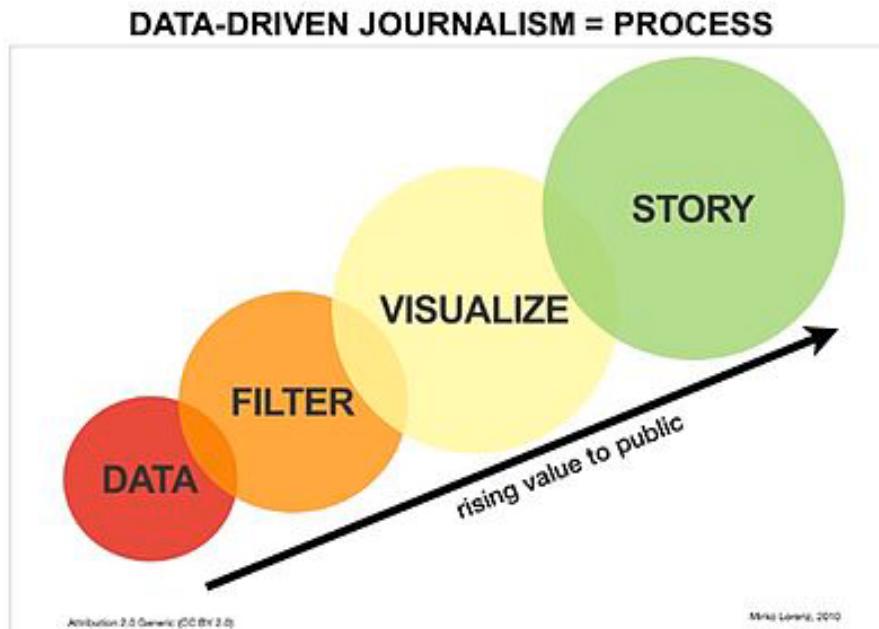
- **Data journalism** reflects the increased role of numerical data for reporting in the digital era. **Data driven** or **DDJ** more useful
 - Reflects increased interaction between journalists and {visualization design, computer science and statistics}.
- Aspects:
 - **Database analysis**: FOI requests on topics like crime, racial discrimination, judicial sentencing, toxic pollution, ...
 - **Visualization**: infographic←→data vis., emphasis on the **story**
 - Increasing use & development of data vis. **software tools**
 - analysis: R, python, ...
 - visualization libraries: R, D3, javascript, ...
 - Increasing use of **dynamic updating** of data sources
 - election polls / prediction
 - disease tracking

What is data journalism?

To an outsider, it is often hard to tell concretely what #DDJ is, or is not

Two motivational infographics of #DDJ as a process

- ❖ How does it actually work?
- ❖ What makes it successful?



Bill Dedman: *The Color of Money*

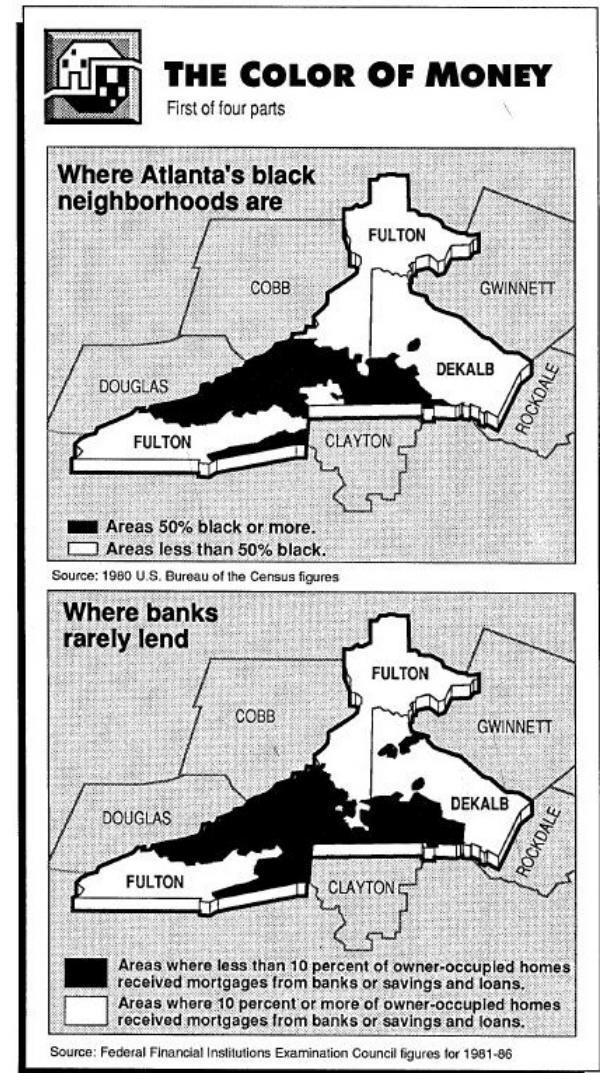
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

In a series of articles, May 1-4, 1988, *The Color of Money*, he showed that although banks made loans in even the poorest white neighborhoods, they did not lend in middle-class or more affluent black neighborhoods.

This work won the Pulitzer prize in investigative reporting in 1989.

It helped establish the idea of computer-assisted reporting and the importance of compelling visualizations in telling the story.

See: <http://powerreporting.com/color/> for all the articles and images



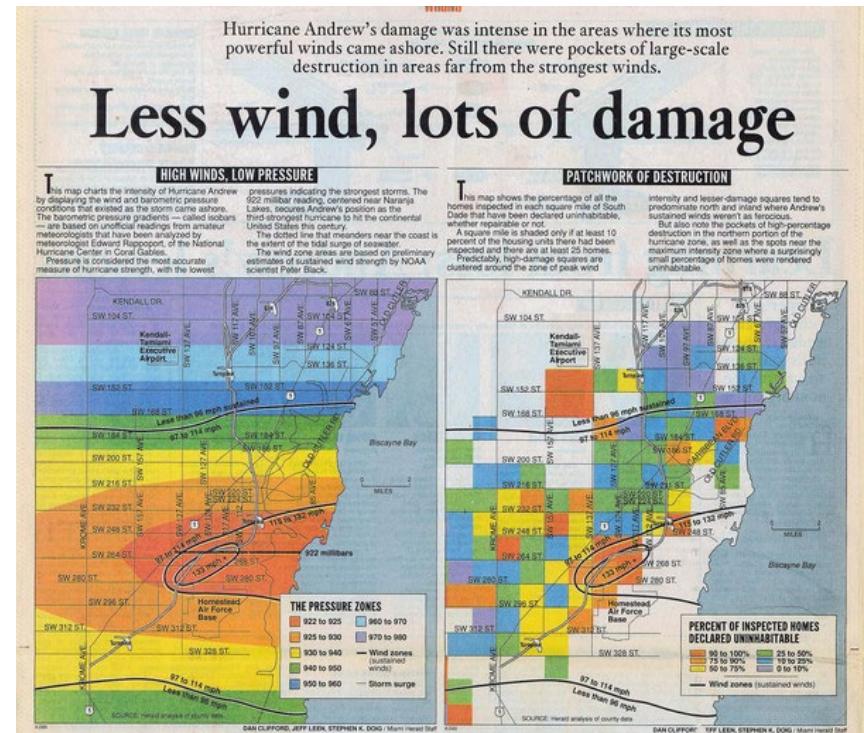
Hurricane Andrew

- ❖ In Aug. 1992, a category 5 hurricane hit Miami-Dade county
- ❖ Huge numbers of houses were destroyed, but this varied by neighborhood
- ❖ Investigative reporters at the *Miami Herald* discovered why:
 - Building codes and inspection became lax after 1980.
 - Older homes survived.
- ❖ This report by Liz Balmaseda et al. won the Pulitzer Prize in 1993.



Hurricane Andrew

- ❖ The report in the *Miami Herald* contained many detailed analysis maps
- ❖ There was overwhelming evidence for the cause of destruction.
- ❖ The combination of investigative reporting, analysis and graphics was seen as powerful & compelling
- ❖ Data-driven journalism had arrived



Predicting elections

- ❖ In 1962, John Tukey and a statistical team joined with NBC to attempt to predict the congressional election
- ❖ Over time, they developed new methods to combine **past results** with **exit polls** and **current returns** to give up-to-date estimates and measures of uncertainty as the numbers came in.
- ❖ Key idea: “borrowing strength”
- ❖ Now known as “hierarchical Bayes”



John Tukey (front, left) and Richard Link (front, right) as part of the NBC Election Day team in Rockefeller Plaza Studios with RCA/NBC staff in the background, 1962. Photo courtesy of the American Philosophical Society, Tukey Archives

NBC execs were freaked out when the Tukey team began to call elections quite early with 95%, then 97.5% certainty. They waited until it reached 99%, and missed the scoop.

CNN Magic Wall

Direct, interactive visualization of election news on TV took a big jump in 2008 with the development of the **Magic Wall**, a large touch-screen display, tied to databases of election results & with on-screen controls to zoom in/out, compare, highlight, etc.

John King
with the
Magic Wall



Video: <https://www.youtube.com/watch?v=rpLHnK-MSy4>

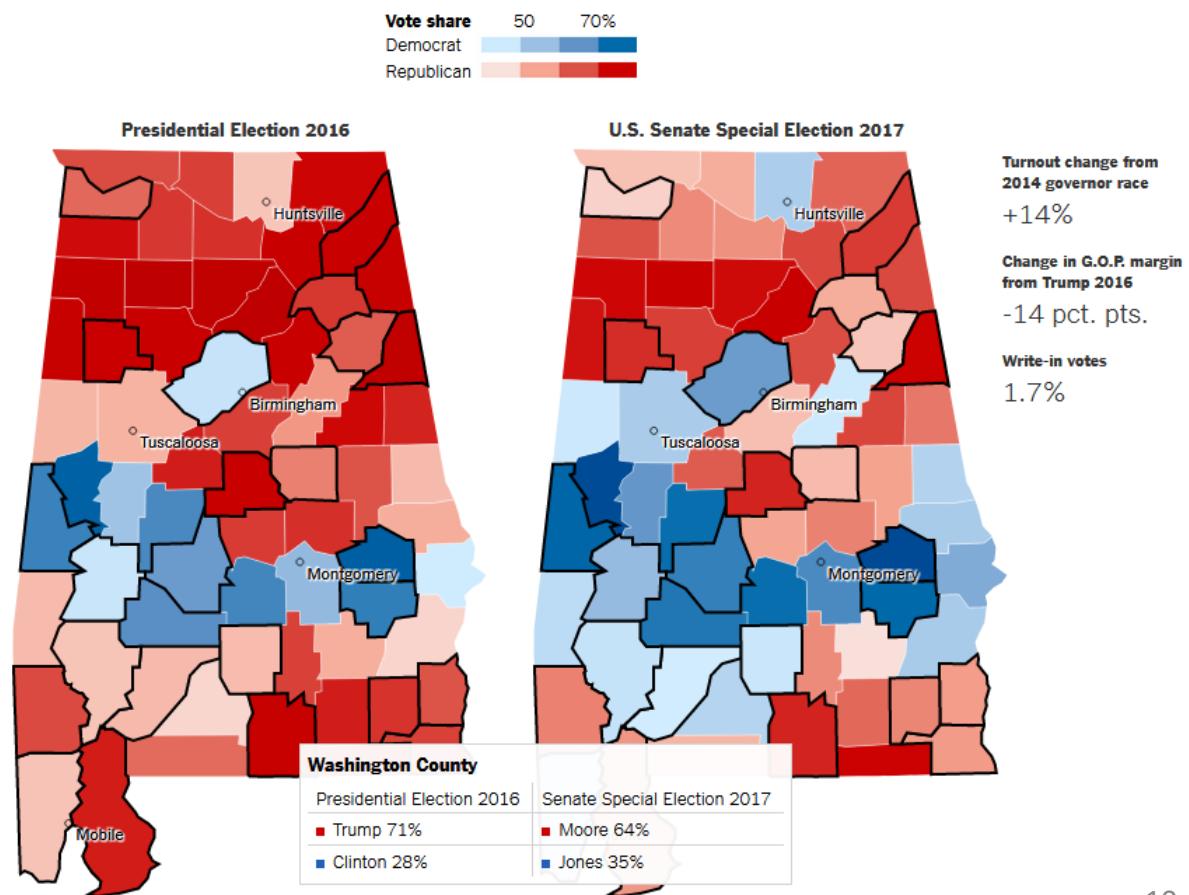
Election reporting: NYT

In the special election for the Alabama U.S. Senate, Dec. 12, 2017, *The New York Times* used dynamic, interactive maps to compare the vote to the 2016 Presidential election

Hovering the mouse over a county showed the details

A map-based display is natural & easy

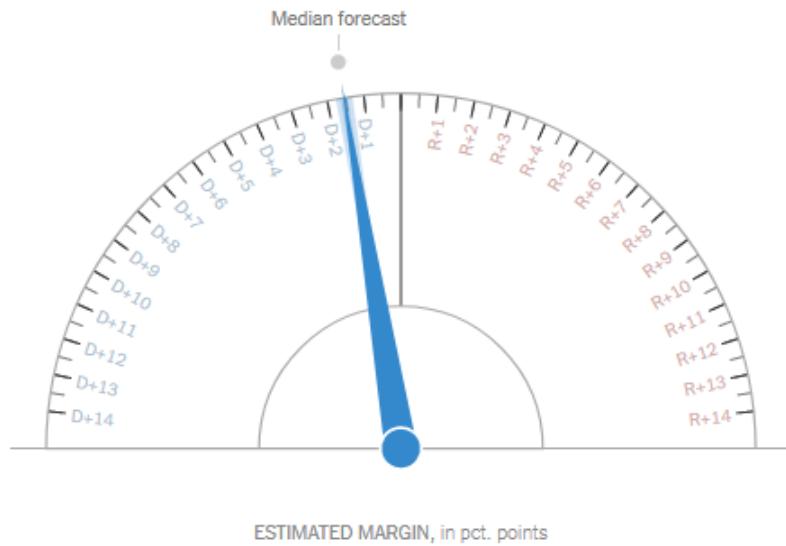
Q: What other graphic might show the comparison better?



Election reporting: Election needle

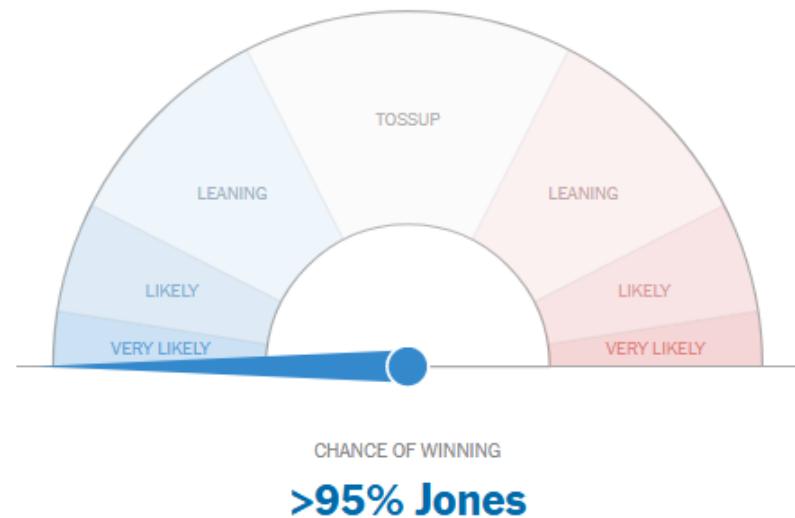
In the 2016 US election, the NYT used **dynamically updated** dashboard displays of the election returns and model-based predictions of the winner of the Alabama Senate race, predicting a win by the Democrat, Doug Jones, over rival Republican, Roy Moore.

Doug Jones has won the election, according to A.P.



Two dials:

- Predicted outcome
- Certainty



The jittery icon of liberal trauma

But in the Hilary Clinton – Donald Trump contest the needle earned a place in media infamy. “For much of the night, ‘the needle’ — a quivering indicator of which candidate is more likely to win— showed Hillary Clinton as the favorite. And then, suddenly, it didn’t.”

“As the needle swung in Trump’s direction around 9 p.m., it became much more than a data visualization. For Clinton voters, it became an encapsulation of everything that was going terribly, unfathomably wrong: The fate of the country and democracy distilled to a lone meter that was suddenly pointing to ‘red alert.’”



Will Oremus, <https://gen.medium.com/how-i-learned-to-stop-worrying-and-love-the-needle-9bfed8f41023>

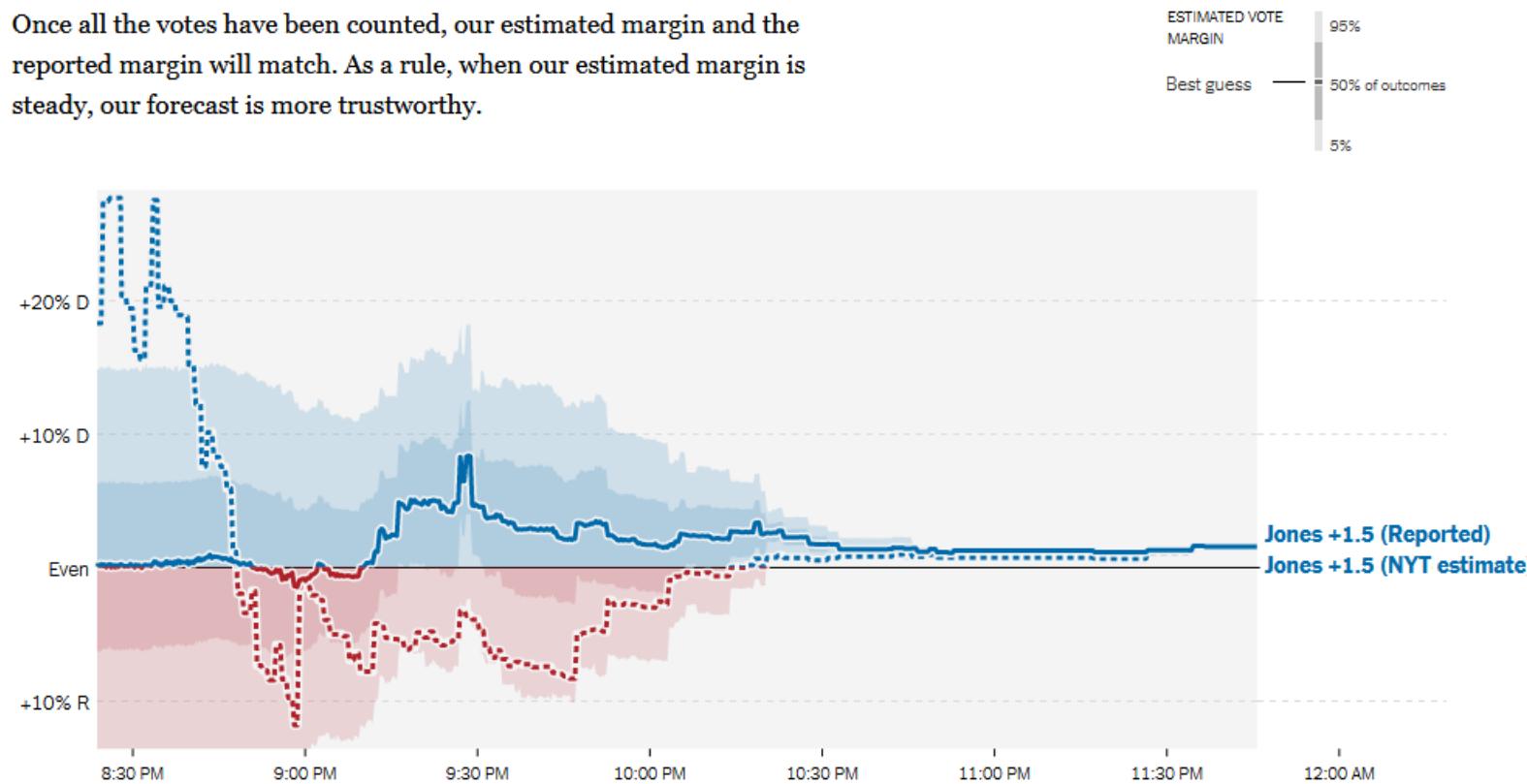
The moment our long national nightmare began

Election reporting: Dashboards

Another dynamically updated chart showed the projected vote margin, together with two levels (50%, 95%) of **uncertainty intervals** for the current estimate.

Projected Vote Margin

Once all the votes have been counted, our estimated margin and the reported margin will match. As a rule, when our estimated margin is steady, our forecast is more trustworthy.



Novel graphic designs

This graph, from fivethirtyeight.com was designed to show how some presidential candidates had shifted positions before the 2016 election.

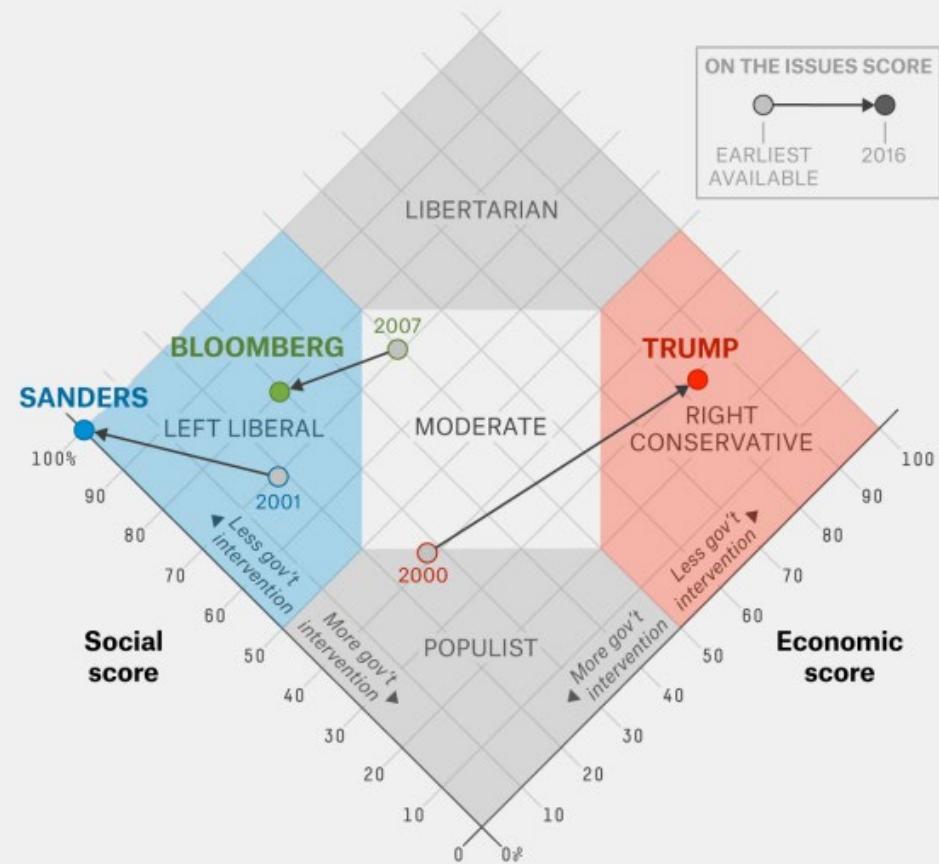
The axes are a score on **social** and **economic** policy, but they rotated the axes by 45° to create zones related to political thought.

This info graphic is eye-catching and self-explanatory:

- colored/labeled zones
- interpretive labels on axes
- arrows showing movement to extremes

Candidates abandoning the middle ground

Earliest available and current OnTheIssues score for Bernie Sanders, Donald Trump and Michael Bloomberg



Measles and vaccines

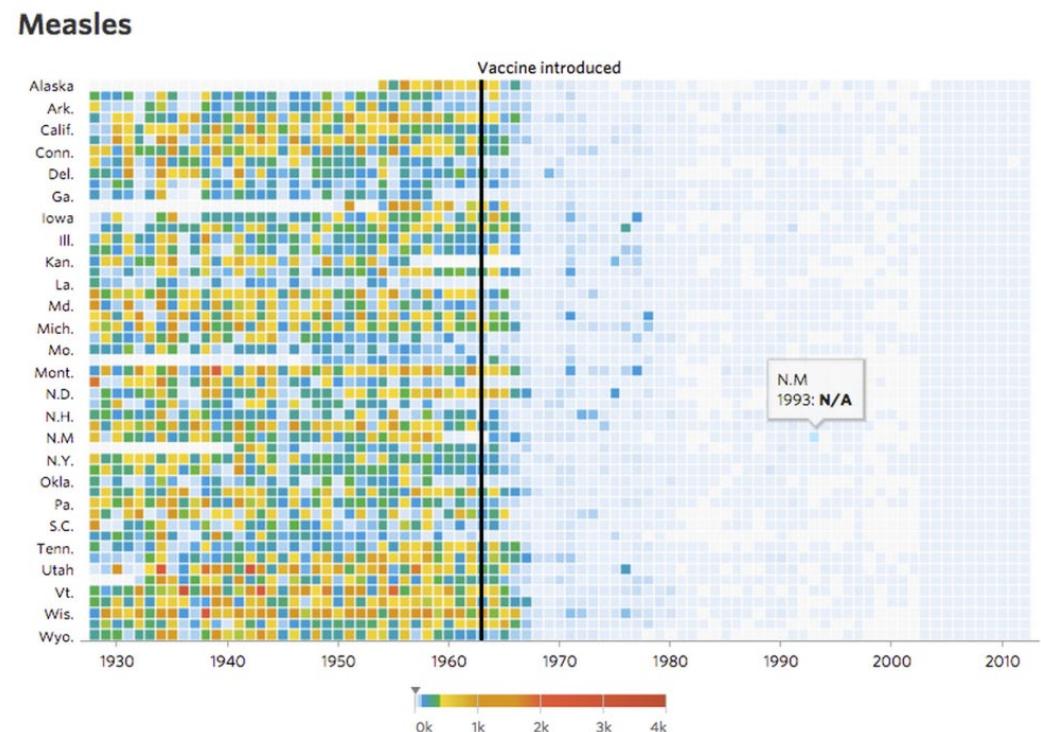
Visualizing the impact of health policy interventions

In 2015 Tynan DeBold & Dov Friedman in the *Wall Street Journal* tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases

The long time series ~70 years made this work.

The heat map color scale is not exemplary, but the message is still clear: disease incidence declined after vaccines were introduced.

The images are “interactive,” in the weak sense that tool-tips are shown on mouse movement.



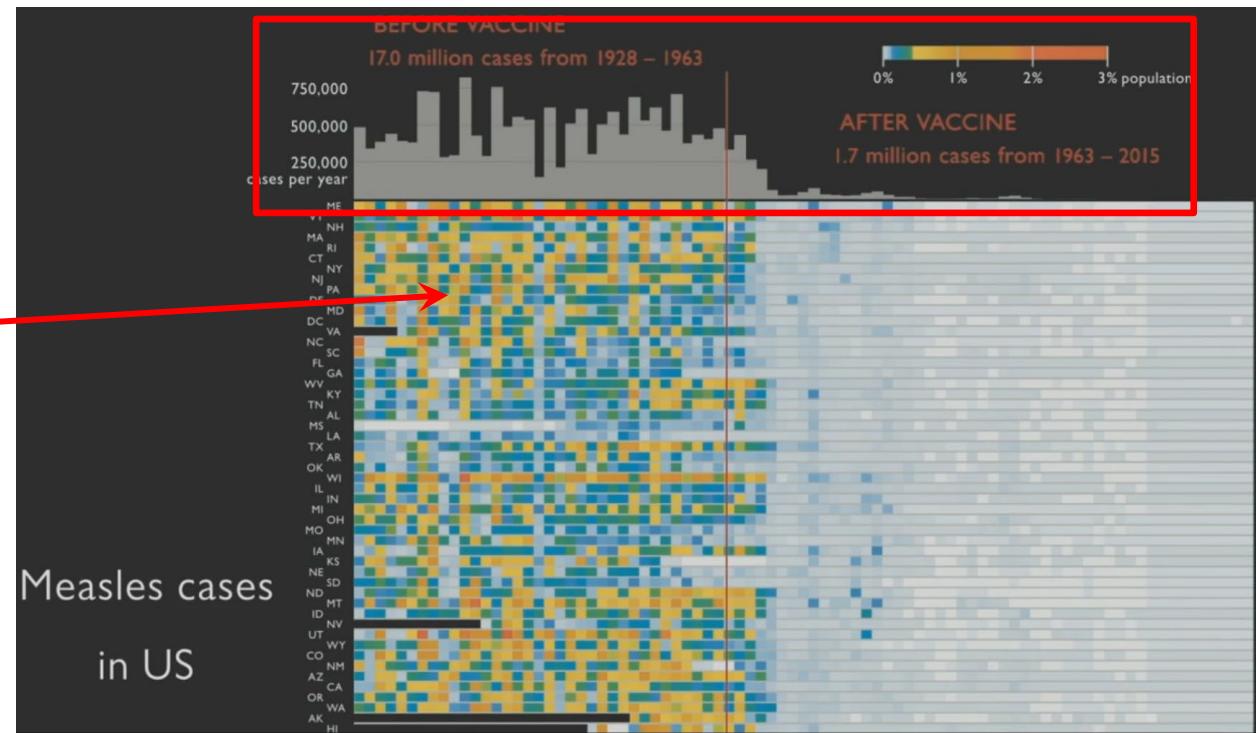
Making the message more explicit

What you should try to show here in an Info graphic is the **overall impact** of vaccination on measles

Ed Tufte did this by adding a histogram at the top showing **total # of cases** by year

With the overall impact made clear, details about individual states can now be explored

“Overview first; details on demand”



Effect ordering & color scales

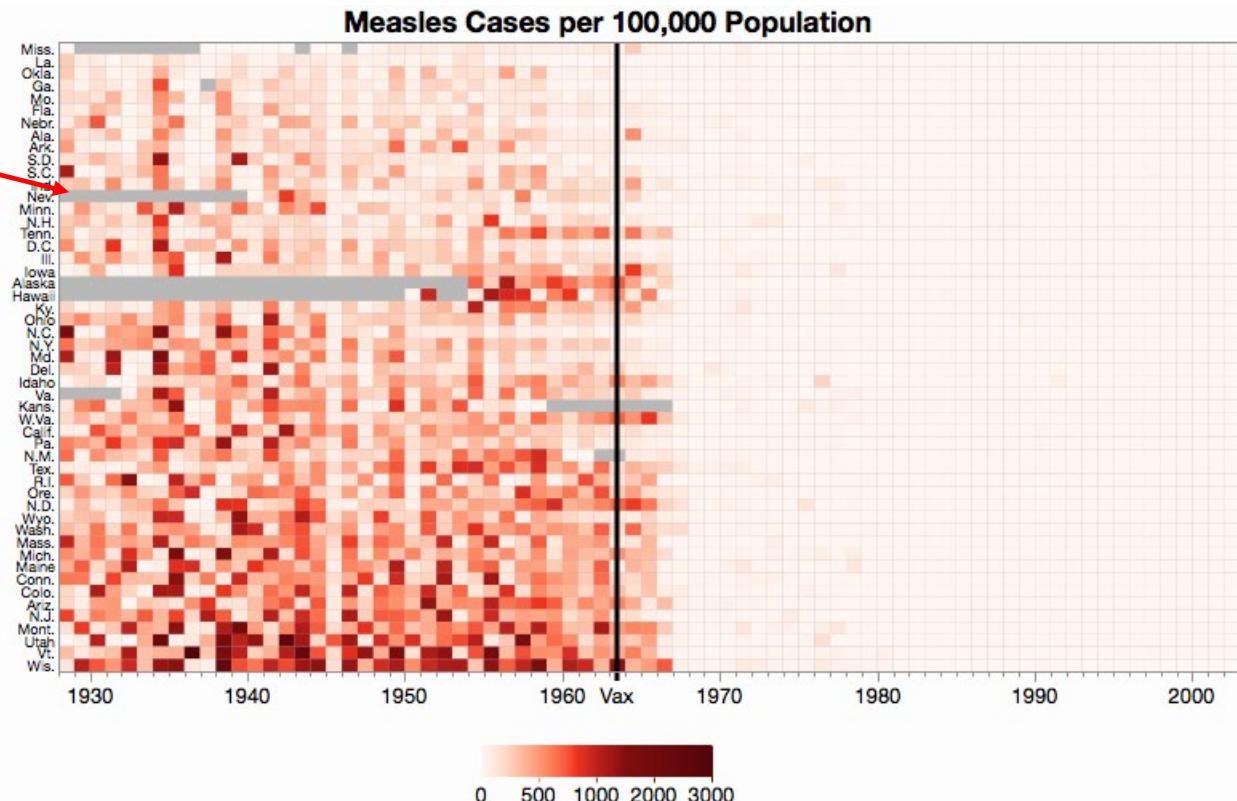
Ordering the states by an overall measure of measles incidence makes it easier to study details.

A uni-dimensional scale of color is more effective than the ad-hoc scale used in the original

This chart, by Xan Gregg
also shows missing data
explicitly.

Color scales for data
graphs is a growing
topic.

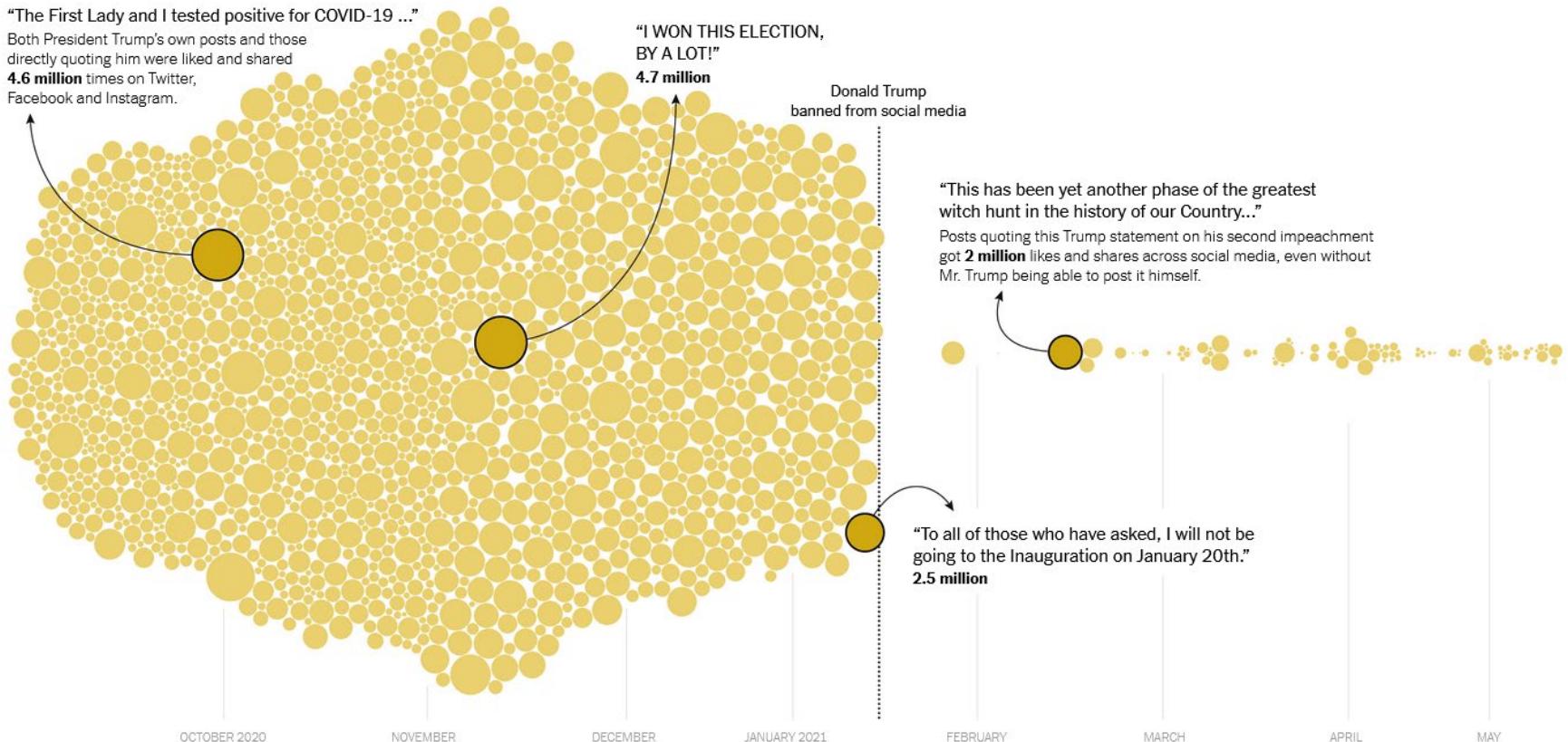
What's wrong in Wis. &
VT?



Before – After Circle Charts

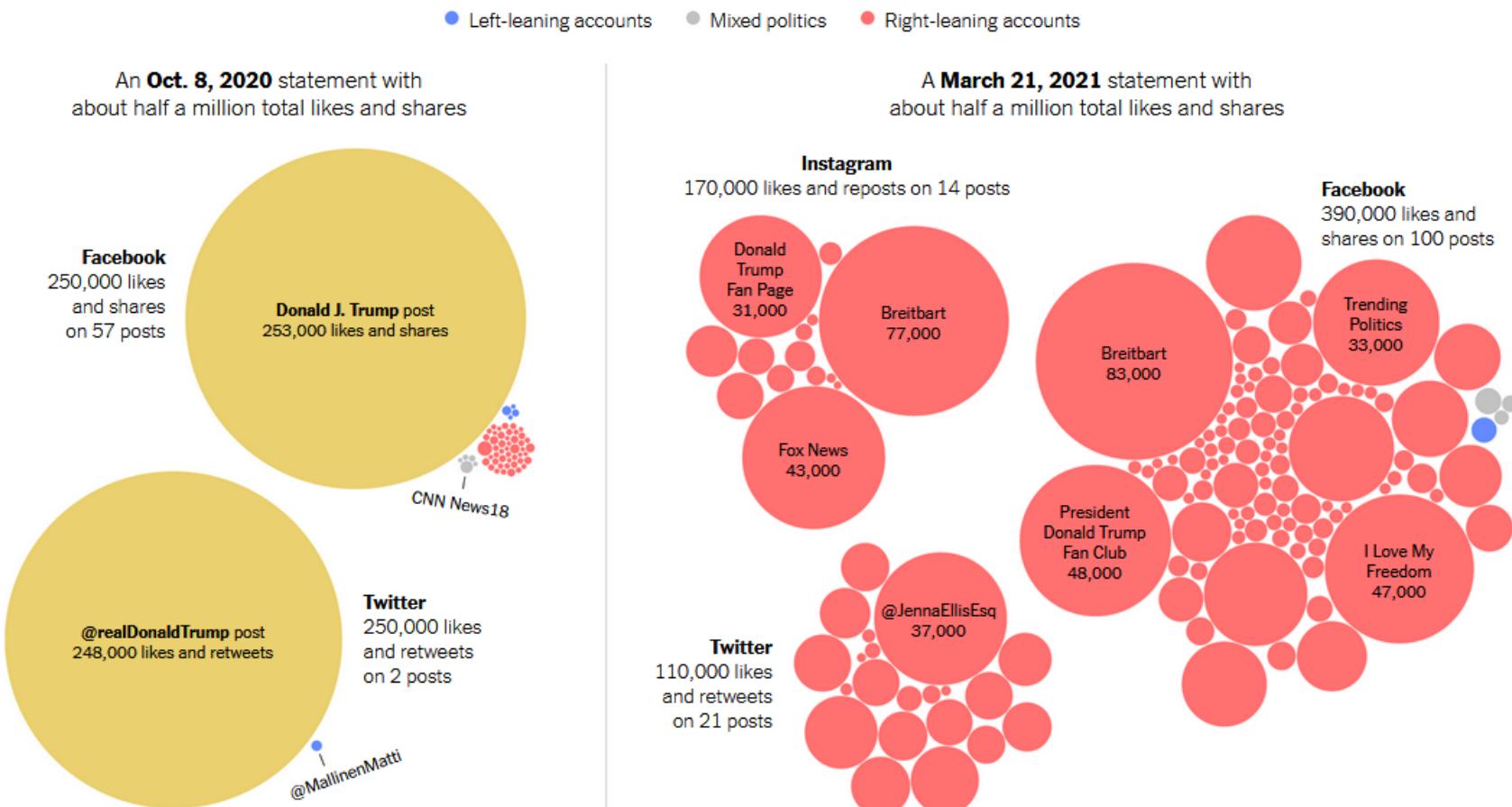
What Happened When Trump Was Banned on Social Media?

- ❖ Circles are statements by Trump.
- ❖ Circle size ~ likes & shares on social media



Before – After: Who pushed the lies?

How/Who Spread Trump statements before and after the ban



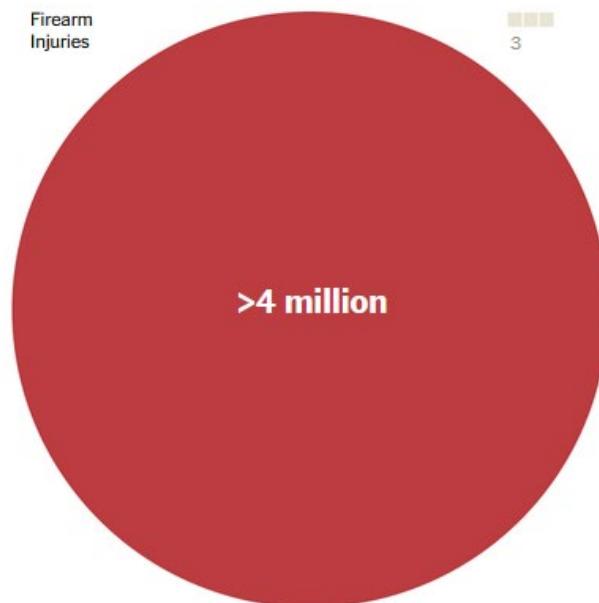
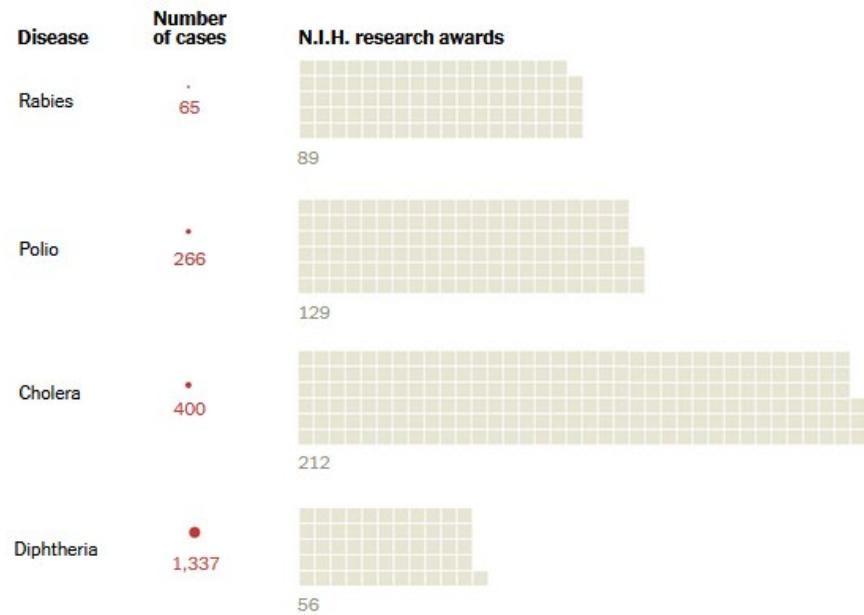
Health & research funding

This powerful NYT chart compares cases of diseases and firearm injuries to the number of N.I.H. research awards.

How well do you like this design?

Is it a fair depiction of the data?

Can you suggest anything different?



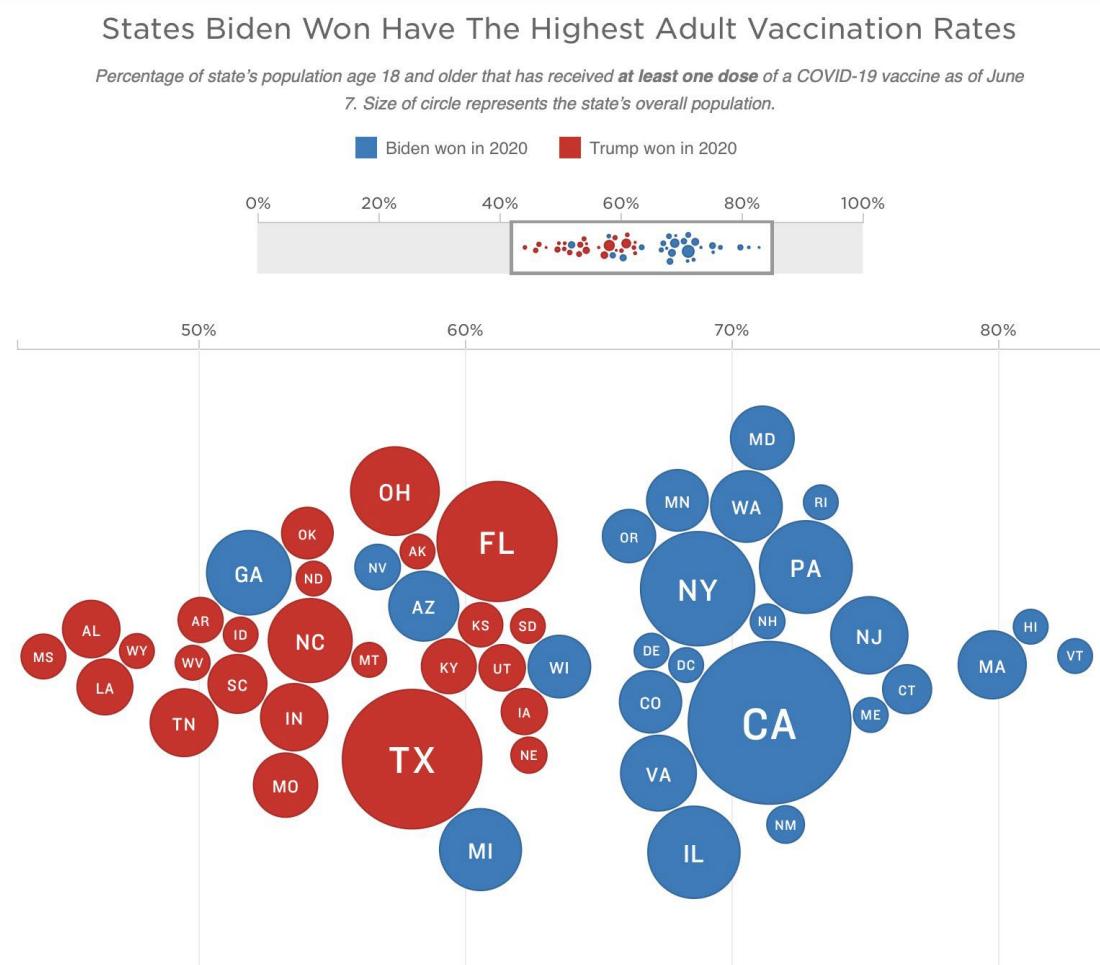
Source: University of Chicago Crime Lab

Votes for vs. COVID vaccinations

% of state pop. with at least one dose, vs. who won (Biden, Trump) in 2020

circle size ~ state pop.

What makes this chart effective?



From: <https://twitter.com/kjhealy/status/1402691186826039301>

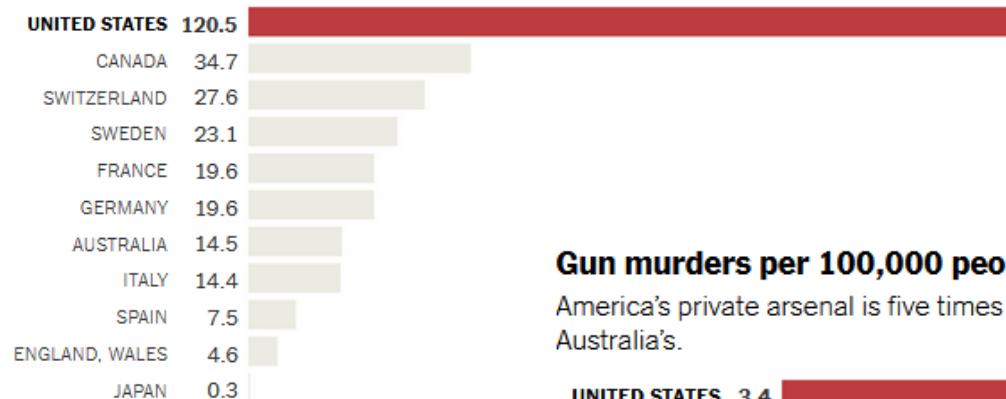
NYT Infographic: Gun deaths



NYT: Gun deaths & policy

Guns per 100 people

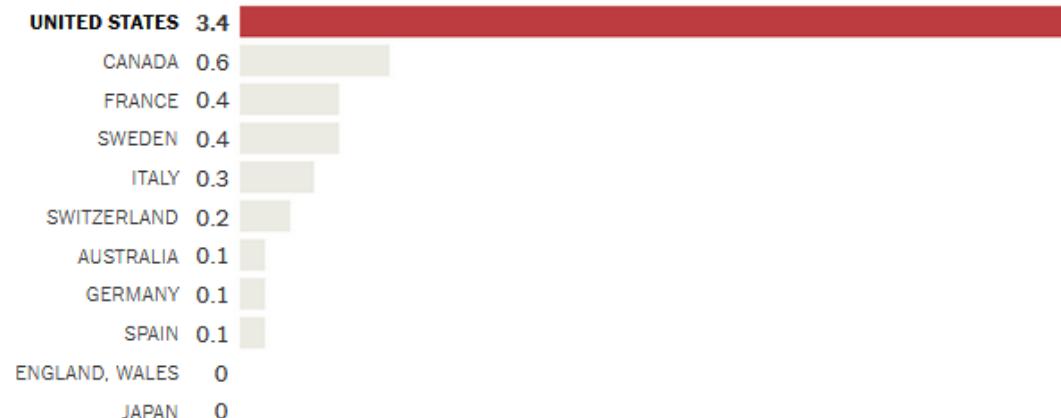
Among developed countries, the United States has by far the highest rate of firearms ownership.



How does the US compare to other Western countries in gun ownership & murders?

Gun murders per 100,000 people

America's private arsenal is five times as lethal as Canada's, and 30 times worse than Australia's.



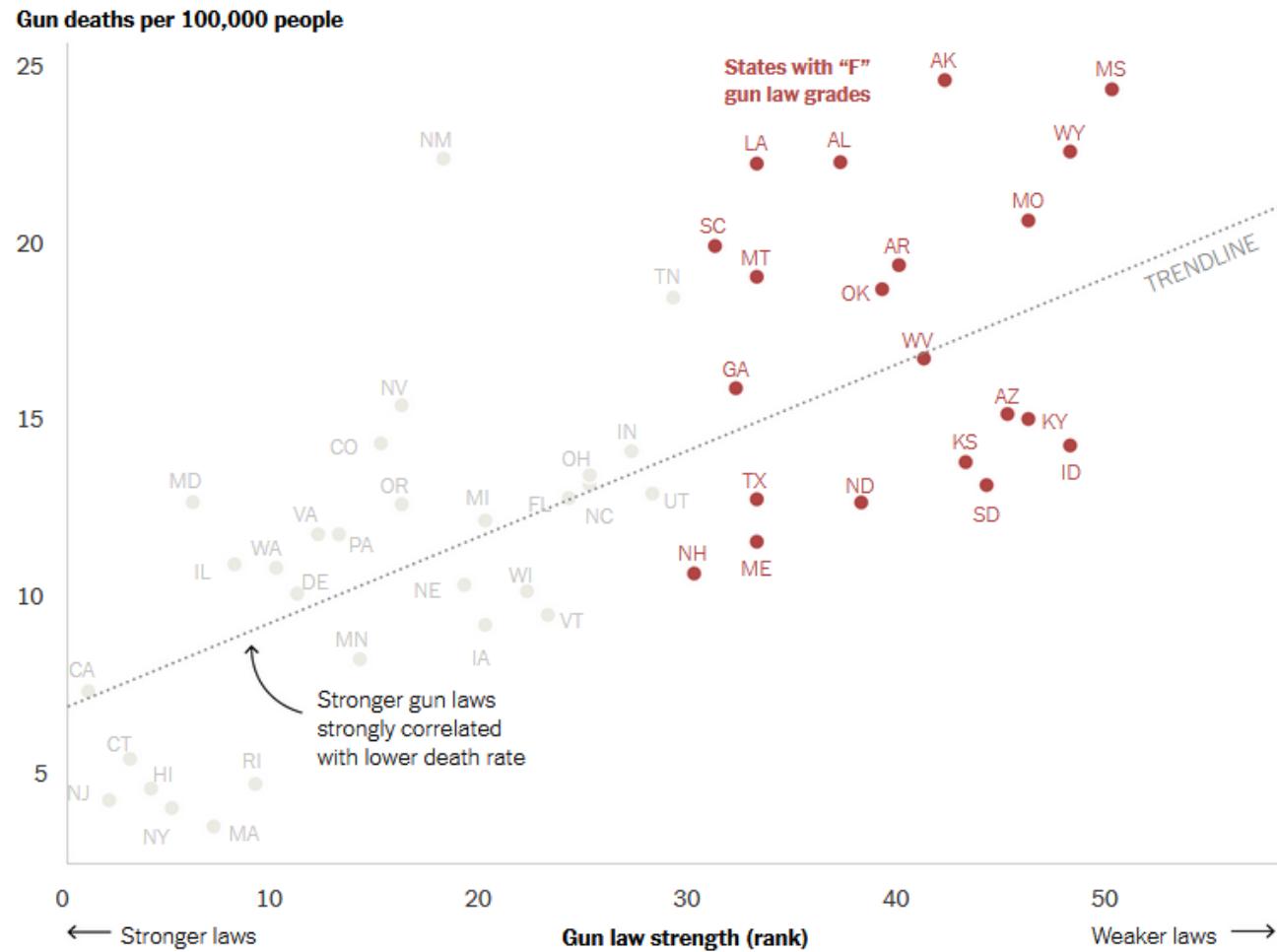
These simple barcharts have a grim conclusion

What makes them effective?

Sources: Small Arms Survey, United Nations Office on Drugs and Crime. * Ownership rates are for 2017. Murder rates for the U.S., Canada, Sweden, Switzerland, Australia and Spain are from 2016; otherwise, the latest available rates are used.

NYT: Gun deaths & policy

This graph directly linked gun deaths to the strength of states' laws on gun ownership

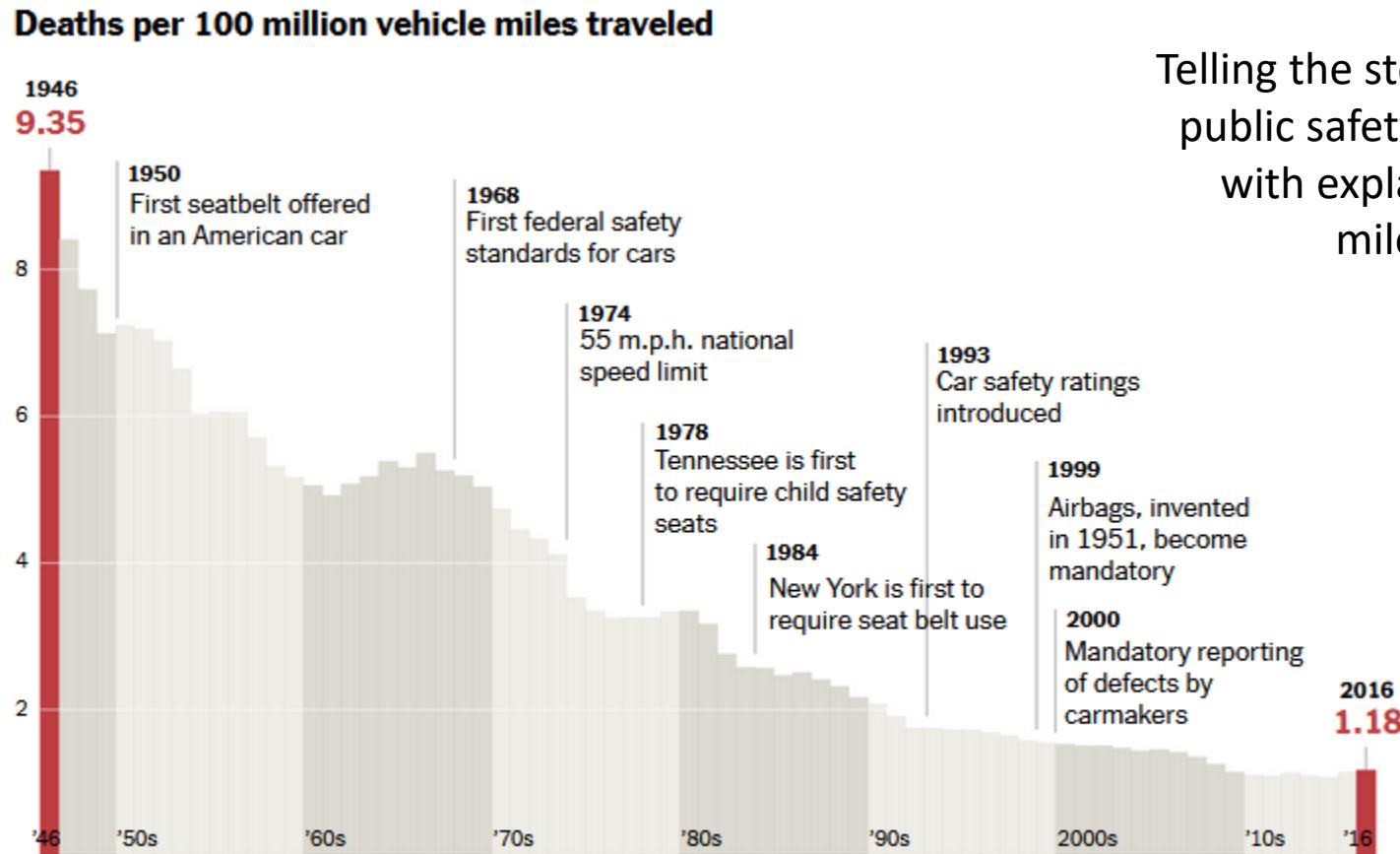


Source: Giffords Law Center

24

Seatbelts: Story telling

New York Times: An annotated chart of vehicle deaths in relation to seatbelts



Telling the story of a public safety policy with explanatory milestones

Source: National Highway Traffic Safety Administration

Racial profiling: The process

How to communicate these results most effectively?

- What is the message? What features are directly comprehensible to the audience?



Man behind the numbers

Racial profiling: Presentation graphic

Together, we created this (nearly) **self-explaining** infographic

Title gives the main conclusion

Text description gives details

Bar width ~ charges
Divided by % release

numbers shown in the cells

Legend gives a layman's description of shading levels

Same charge, different treatment

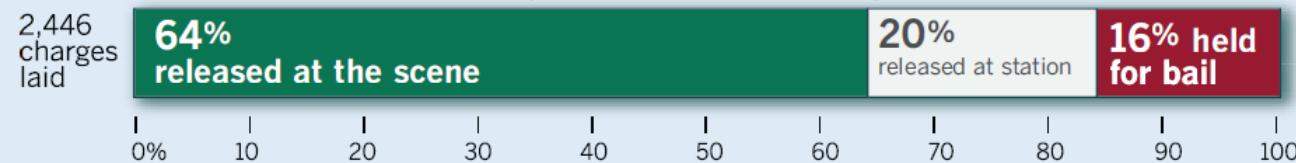
Statistical analysis of single drug possession charges shows that blacks are much less likely to be released at the scene and much more likely to be held in custody for a bail hearing. Darker colours represent a stronger statistical link between skin colour and police treatment.

Degree of likelihood
Much less likely to occur
Much more likely to occur
More likely to occur

Whites are more likely to be released at the scene



Blacks are much more likely to be held for bail hearings



SOURCE: Toronto police arrest records 1996-2002

Best examples of DDJ? Ask Twitter



Michael Friendly @datavisFriendly · 22h

1/ Thinking about an upcoming class in my #psyc6135 #datavis course on the topic #datajournalism, with a guest star @Jleerankin . It got me thinking of the idea of "Great moments in Data Journalism": investigative reporting, incisive analyses, compelling graphic displays.



Michael Friendly @datavisFriendly · 22h

2/ I would love ❤️ for nominations on this topic. @albertocairo and others have done some of this, but perhaps this is a topic on which others could contribute their nominations for "Great moments..."



Alberto Cairo ✅ @albertocairo · 22h

Twitter #ddj and news #dataviz, contribute with nominations.

I'd do @FiveThirtyEight's: projects.fivethirtyeight.com/redistricting-...

Anything by @ProPublica, like @lenagroeger's appropriately creepy: projects.propublica.org/graphics/worke...



Michael Friendly @datavisFriendly · 20h

...

3/

☒ Reciprocity: How has #DDJ contributed to or advanced #datavis itself?
And, vice-versa?

💡 Team work: How do teams of journalists, data people, graphic designers and web developers work together on projects? Anything special/different from other team projects?

Now trending ...

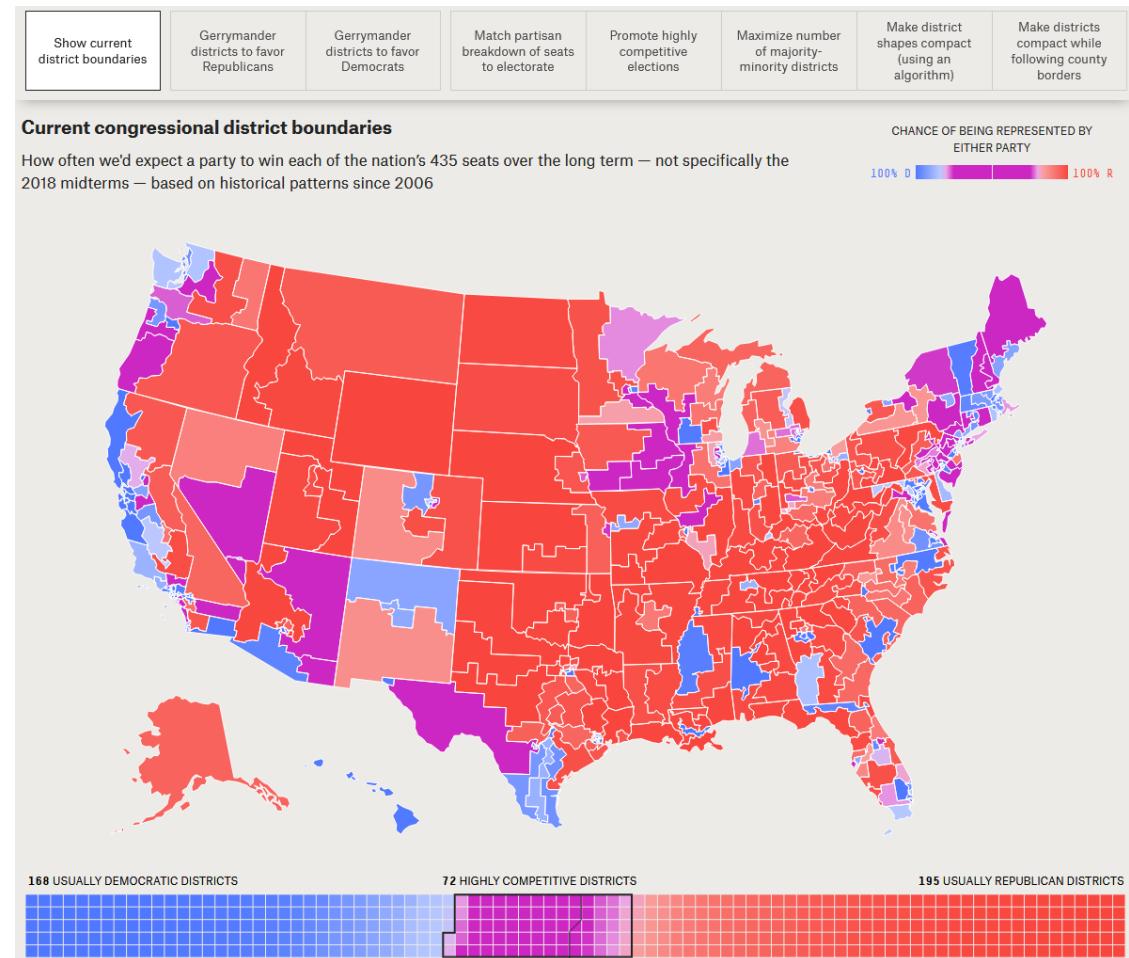
- ❖ Digital publishing and online, interactive media
 - Many news organizations have interactive graphic developers or entire online departments
 - A revenue stream responding to less from print & ads
- ❖ Story telling frames a wider enterprise
 - A new focus and techniques for data visualization
 - New online forms: scrollytelling
- ❖ Opening data
 - Maps-GIS (Google, Open Street maps)
 - Gov't data: census, health, crime, ...
 - NGO initiatives: Our World in Data
 - Unsealed: WikiLeaks, Panama Papers
- ❖ Software infrastructure
 - Javascript as base for online, interactive methods
 - Libraries, frameworks building on this (D3.js, Processing, ...)

538: Atlas of Redistricting

There's a lot of complaining about gerrymandering, but what should US Congressional districts look like? What would they look like to serve various goals?

This map shows the current boundaries, with expected wins by Rep and Dem.

Why do I find this app appealing?



From:

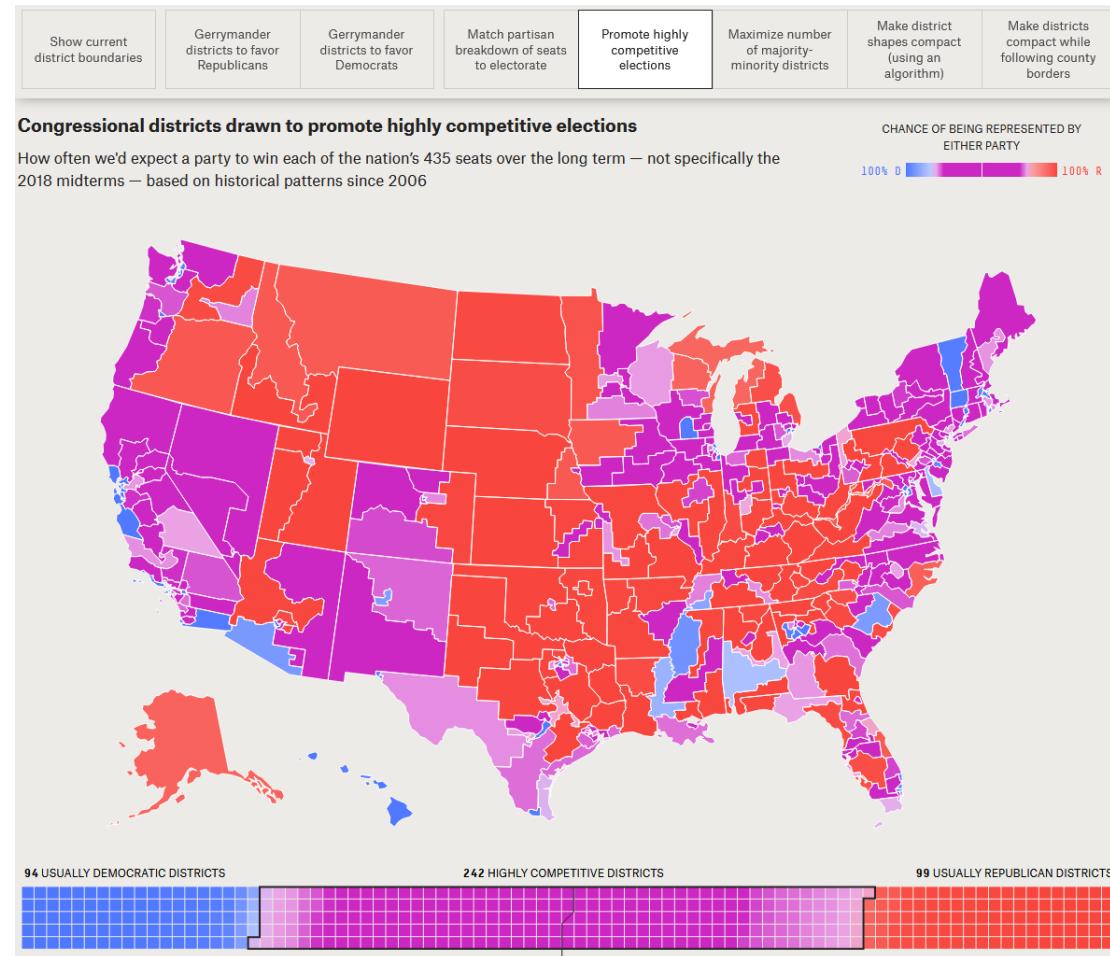
<https://projects.fivethirtyeight.com/redistricting-maps/>

538: Atlas of Redistricting

How could district boundaries be re-drawn to promote more **competitive** elections?

Election district boundaries are consequential & highly political. They can be changed by state legislatures.

Different scenarios show the changes in representation, w/o a single voter switching parties!



From:

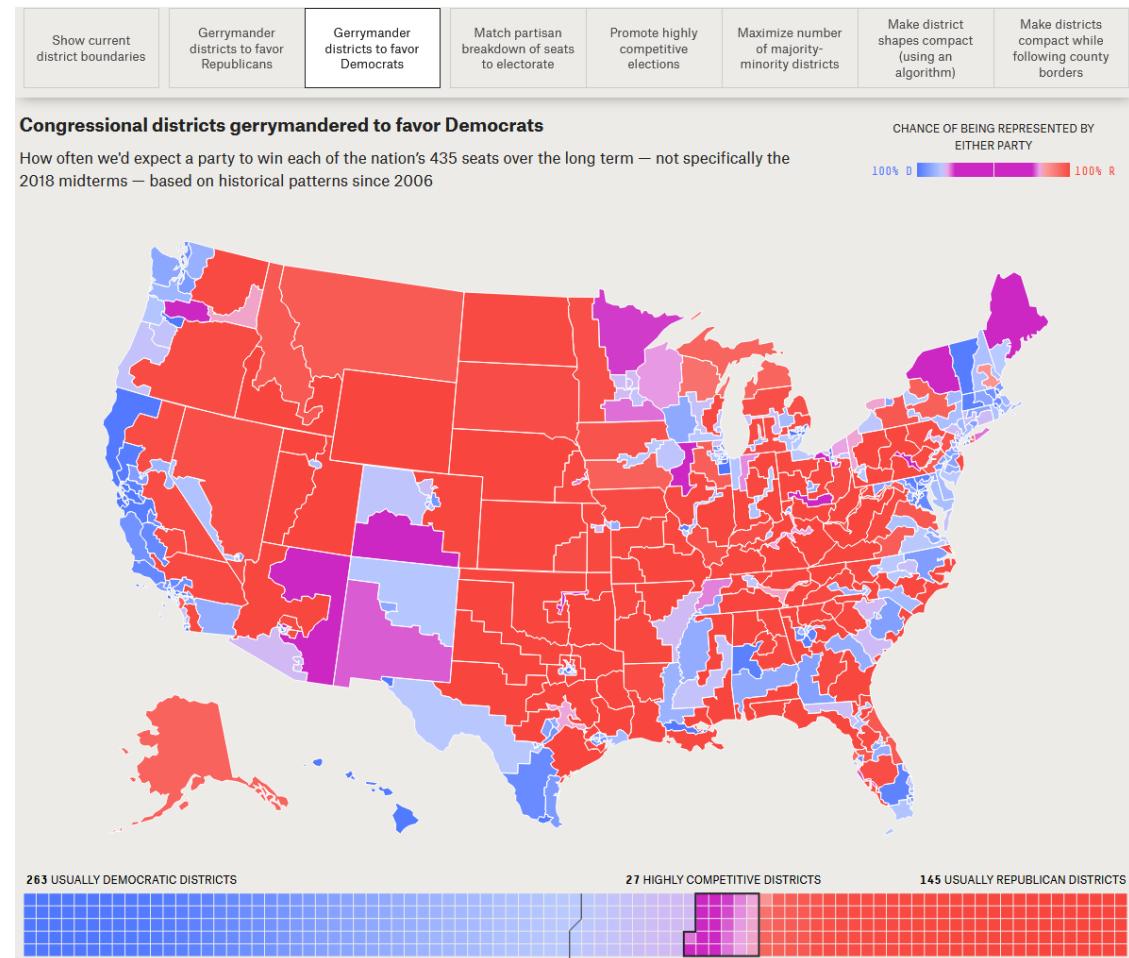
<https://projects.fivethirtyeight.com/redistricting-maps/>

538: Atlas of Redistricting

What if the **Democrats** could control election districts to favor themselves?

Note that the overall impression of the map is still **Rep.** overall.

Changes in representation take place largely on the east/west coasts



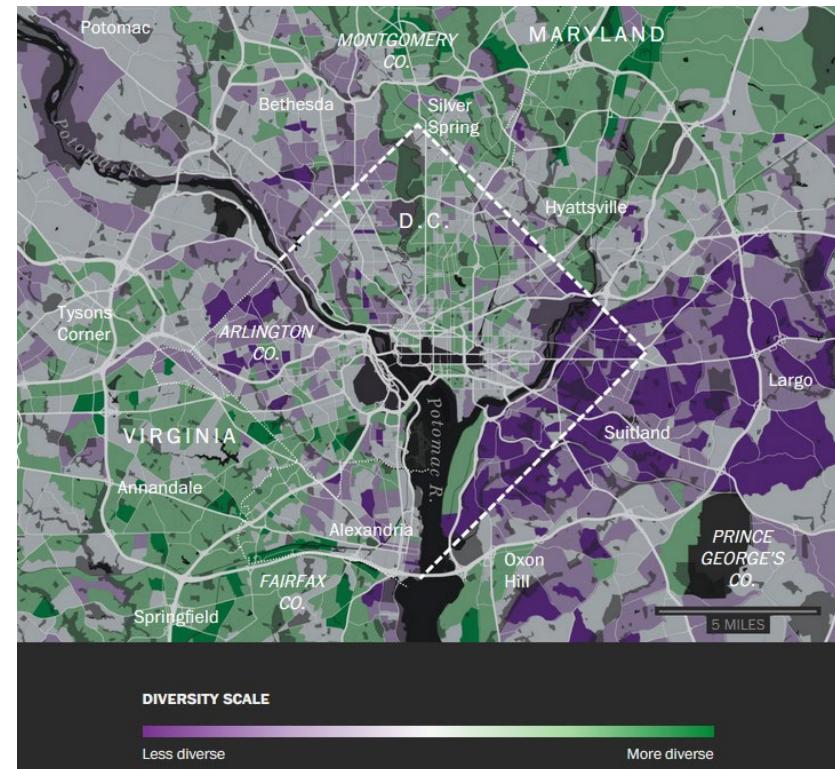
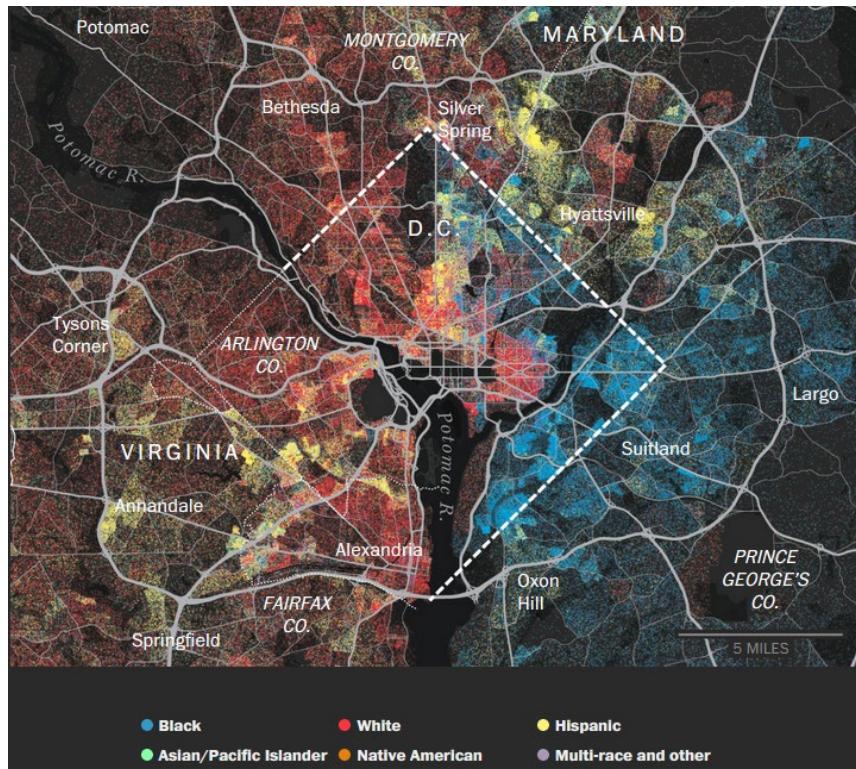
From:

<https://projects.fivethirtyeight.com/redistricting-maps/>

WP: Segregation in U.S. Cities

Aaron Williams & Armand Emamdjomeh at the *Washington Post* studied changes in racial composition and diversity in U.S. cities over 30+ years

They relate these to demographic, policy, and urban/suburban issues using well-designed thematic maps.



Our World in Data

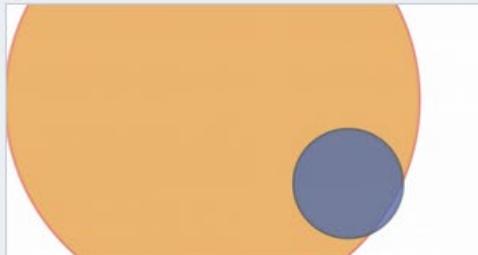
An impressive range of social & environmental issues presented by *Our World in Data*



We just published our new data explorer on Crop Yields

Explore our new data explorer covering yields and yield gaps for a range of crops across the world.

By Hannah Ritchie — June 08, 2021



Why do we need to know about progress if we are concerned about the world's large problems?

Why have we made it our mission to publish the "research and data to make progress against the world's largest problems"?

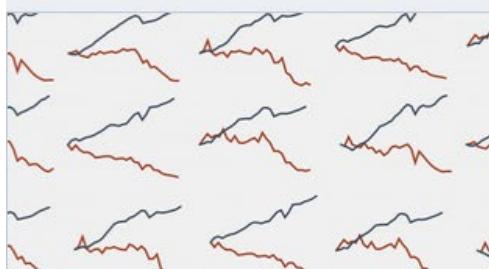
By Max Roser — June 07, 2021



The price of batteries has declined by 97% in the last three decades

To transition towards low-carbon energy systems we need low-cost energy storage. Battery costs have been falling quickly.

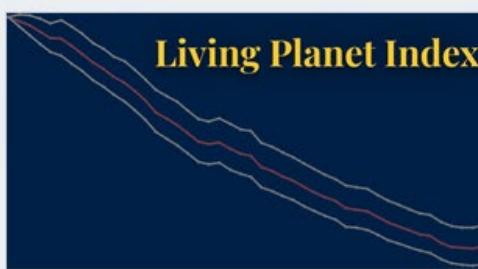
By Hannah Ritchie — June 04, 2021



The argument for a carbon price

We are paying a price for fossil fuels, but that price is not paid by those that burn the fossil fuels – we need to change that.

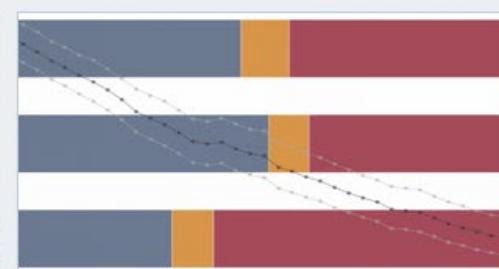
By Max Roser — June 01, 2021



Living Planet Index: what does an average decline of 68% really mean?

The Living Planet Index is the biodiversity metric that always claims the headlines. It's often misinterpreted. How should we understand it?

By Hannah Ritchie — May 17, 2021



How the Living Planet project helps us understand how the world's wildlife is changing

Beneath the popular index, the Living Planet database helps us understand where and what animals are deeply threatened, and what animals are thriving.

By Hannah Ritchie — May 17, 2021

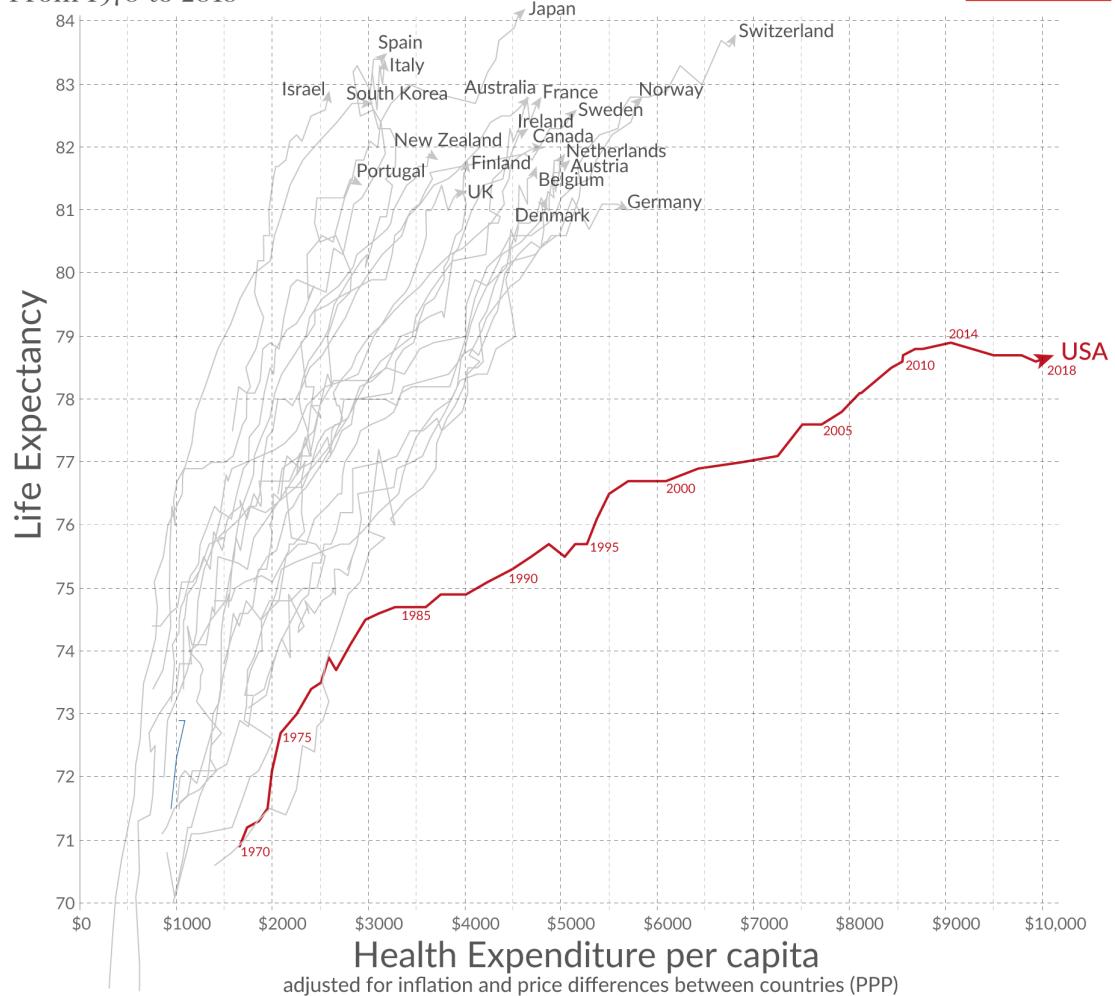
Our World in Data

What is the main message of this chart?

Is it effective?

Life expectancy vs. health expenditure From 1970 to 2018

Our World in Data



Data source: OECD — Note: Health spending measures the consumption of health care goods and services, including personal health care (curative care, rehabilitative care, long-term care, ancillary services, and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments.

Shown is total health expenditure (financed by public and private sources).

Licensed under CC-BY by the author Max Roser.

OurWorldInData.org – Research and data to make progress against the world's largest problems.

Financial Times COVID tracker



John Burn-Murdoch ✓

@jburnmurdoch

This series of charts from the Financial Times was hailed for best-practices in the early stages of the pandemic.

- ❖ Log scale make it easy to compare rates of increase
 - ❖ Visual meaning for “leveling the curve”
 - ❖ Labels for interventions (lockdown)
 - ❖ Sensible (relative) origin – days since 10th death

The [@FinancialTimes](#) coronavirus death & case trajectory trackers are now 🔥 FREE TO READ 🔥 outside the paywall: ft.com/coronavirus-latest

In this morning's update, the US has gone above 470 deaths, bringing it just behind where Iran was at the same stage.



7:20 AM · Mar 23, 2020

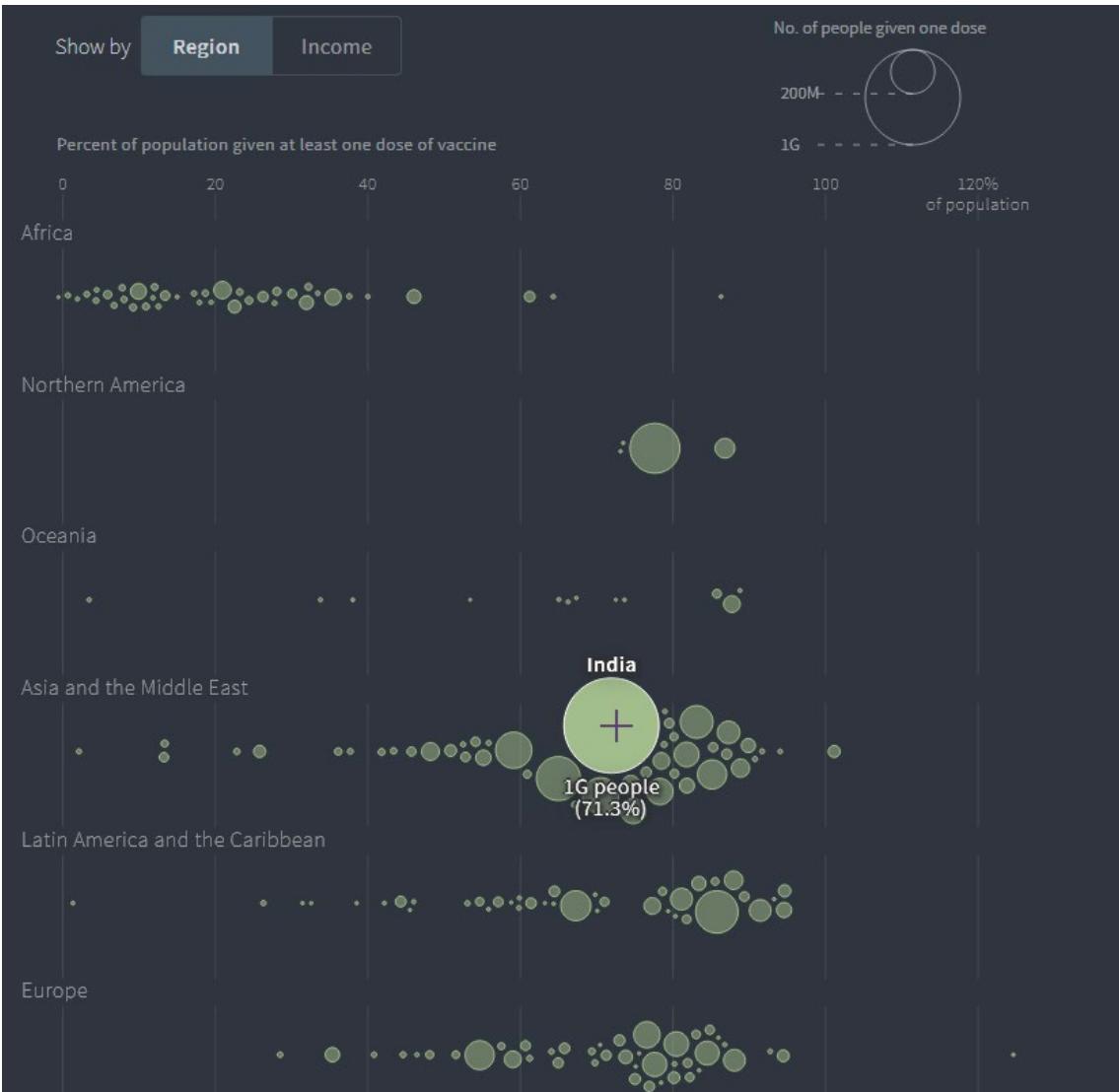


Reuters COVID tracker

Data journalists at Reuters created live trackers, pulling data from different sources to generate real-time graphics with text, telling the story of the pandemic. An interactive globe serves as a novel way to explore vaccination rates.



Other interactive displays provide the possibility to compare countries by geographic region and income level



“Reuters also took on the hard task of tracking the vaccination policies that determined who had access to COVID-19 vaccines in countries around the world.

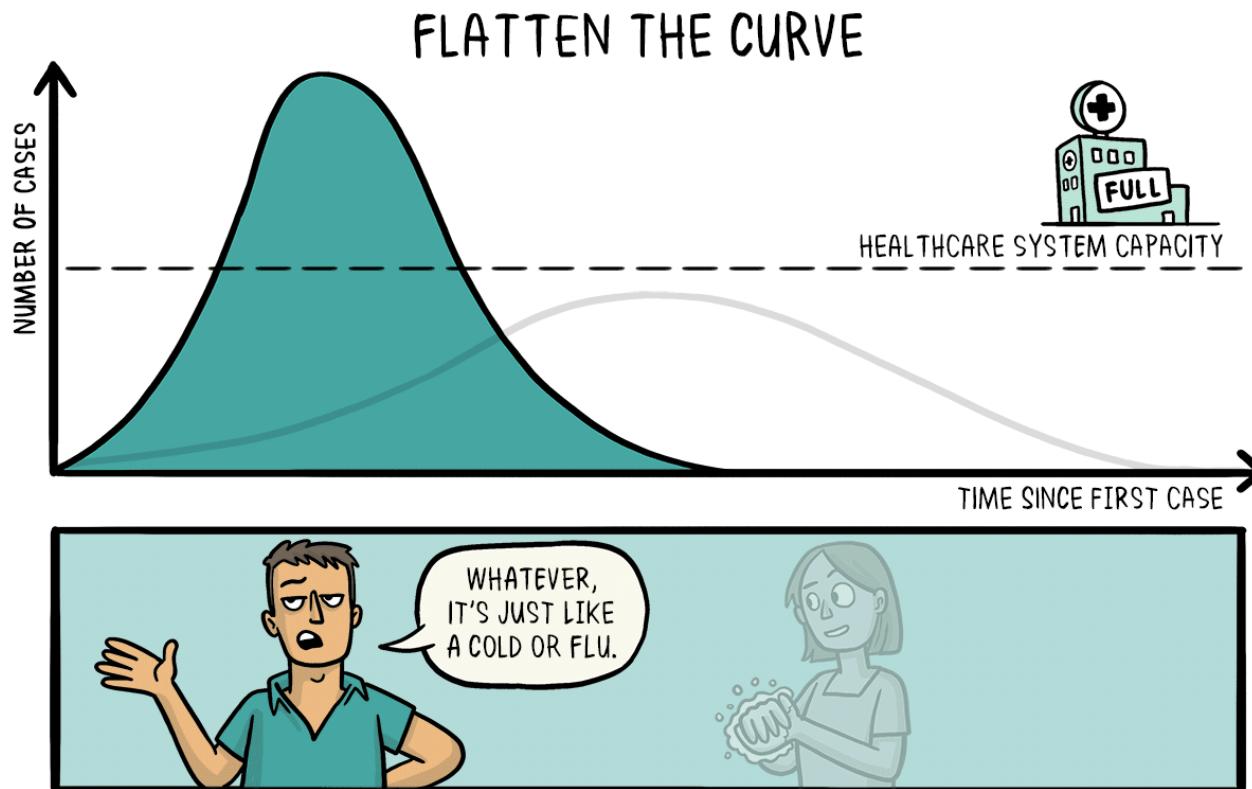
That meant creating a unifying structure to catalogue the different phases of each country's vaccination rollout plan.”

-- Prasanta Kumar Dutta

<https://graphics.reuters.com/world-coronavirus-tracker-and-maps/vaccination-rollout-and-access/>

Flatten the curve

Infographic created by epidemiologist Dr. Siouxsie Wiles & illustrator Toby Morris for the New Zealand publication [The Spinoff](#).



Animated,
rather than
separate panels

Clear,
memorable title

Simple drawing
rather than
messy data

Cases vs.
capacity

@SIOUXSIEW @XTOTL @THESPINOFFTV

'ADAPTED FROM @DREWAHARRIS, THOMAS SPLETTSTÖBER (@SPLLETTE) AND THE CDC'
CC-BY-SA

See: Robert Kosara, <https://eagereyes.org/blog/2020/the-visual-evolution-of-the-flattening-the-curve-information-graphic>

Toronto Star: Trump Lies

Daniel Dale & Ed Tubb at the *Toronto Star* maintained a database of every false claim made by Donald Trump since the Inauguration on Jan. 20, 2017



**Donald Trump has
said 4557 false
things as U.S.
president**

(as of 3-13-2019)

Interactive filter at: <http://projects.thestar.com/donald-trump-fact-check/>

Toronto Star: Trump Lies

Every false claim is documented: place, claim, truth, “slug”, categories

			claim	truth									
number	date	place	claim	infact	slug	repeat	source	category1	category2				
1	2017-01-20	Post-inauguratio...	"I looked at the rain, whic...	The rain began right at the beg...	Inaugu...	Yes	https://www.thestar.com/ne...	Inauguration	Weather				
2	2017-01-20	Post-inauguratio...	"Even the media said the c...	The major media reported that ...	Inaugu...	Yes	https://www.thestar.com/ne...	Inauguration	Exaggerat...				
3	2017-01-21	Speech at Centr...	"Honestly, it looked like a ...	The crowd, which may not have...	Inaugu...	Yes	https://www.youtube.com/w...	Inauguration	Exaggerat...				
4	2017-01-21	Speech at Centr...	"It was almost raining, the ...	Neither of these claims is true. ...	Inaugu...	Yes	https://www.youtube.com/w...	Inauguration	NA				
5	2017-01-21	Speech at Centr...	"So a reporter for TIME ma...	Trump's numbers are well off. H...	TIME co...	Yes	https://www.youtube.com/w...	Media	Exaggerat...				
6	2017-01-23	Private meeting ...	Trump told Congressional ...	This claim, also reported by nu...	Mass fr...	Yes	http://www.foxnews.com/pol...	Voter fraud	Election				
7	2017-01-25	Interview with A...	"In terms of a total audienc...	"They" can mean anyone, but n...	Inaugu...	no	http://abcnews.go.com/Politi...	Inauguration	Exaggerat...				
8	2017-01-25	Interview with A...	"No, no, you have to unde...	Trump's victory was not close t...	Electio...	Yes	http://abcnews.go.com/Politi...	Election	Exaggerat...				
9	2017-01-25	Interview with A...	"I think you're demeaning ...	Saying "you" here, Trump wron...	Media ...	Yes	http://abcnews.go.com/Politi...	Media	Exaggerat...				
10	2017-01-25	Interview with A...	With regard to his speech ...	Most of the audience was inde...	CIA spe...	no	http://abcnews.go.com/Politi...	Trump White House	US intellig...				
11	2017-01-25	Interview with A...	Regarding Healthcare.gov...	Healthcare.gov did not cost \$5 ...	Healthc...	no	http://abcnews.go.com/Politi...	Healthcare	Exaggerat...				
12	2017-01-25	Interview with A...	Muir: "I called the author ...	Grovelling means "to draw bac...	Grovelli...	no	http://abcnews.go.com/Politi...	Politics/transcript-abc-news-anchor-david-muir-aud	NA				
13	2017-01-25	Interview with A...	"Now you're telling me Pe...	Trump was trying to use a 2012 ...	Pew fra...	no	http://abcnews.go.com/Politi...	Interviews/president/story?id=45047602	Aud	Election			
14	2017-01-25	Interview with A...	Regarding his false claim ...	These large numbers of illegal v...	Mass fr...	Yes	http://abcnews.go.com/Politi...	Voter fraud	Chicago				
15	2017-01-25	Interview with A...	Regarding voting fraud: "...	There is no evidence of a signifi...	Mass fr...	Yes	http://abcnews.go.com/Politi...	Voter fraud	NA				
16	2017-01-25	Interview with A...	Regarding voting fraud: "Y...	There is no evidence of a signifi...	Mass fr...	Yes	http://abcnews.go.com/Politi...	Voter fraud	Voter frau				
17	2017-01-25	Interview with A...	"Look, Barack Obama -- if ...	This is a gross mischaracterizati...	Obama ...	no	http://abcnews.go.com/Politi...	Barack Obama	Barack Obama				
18	2017-01-25	Interview with A...	"So, look, when President ...	There were not only no homicid...	Obama ...	no	http://abcnews.go.com/Politi...	Barack Obama	Crime				

media

voter fraud

Jelle Geertsma: Trump Lies

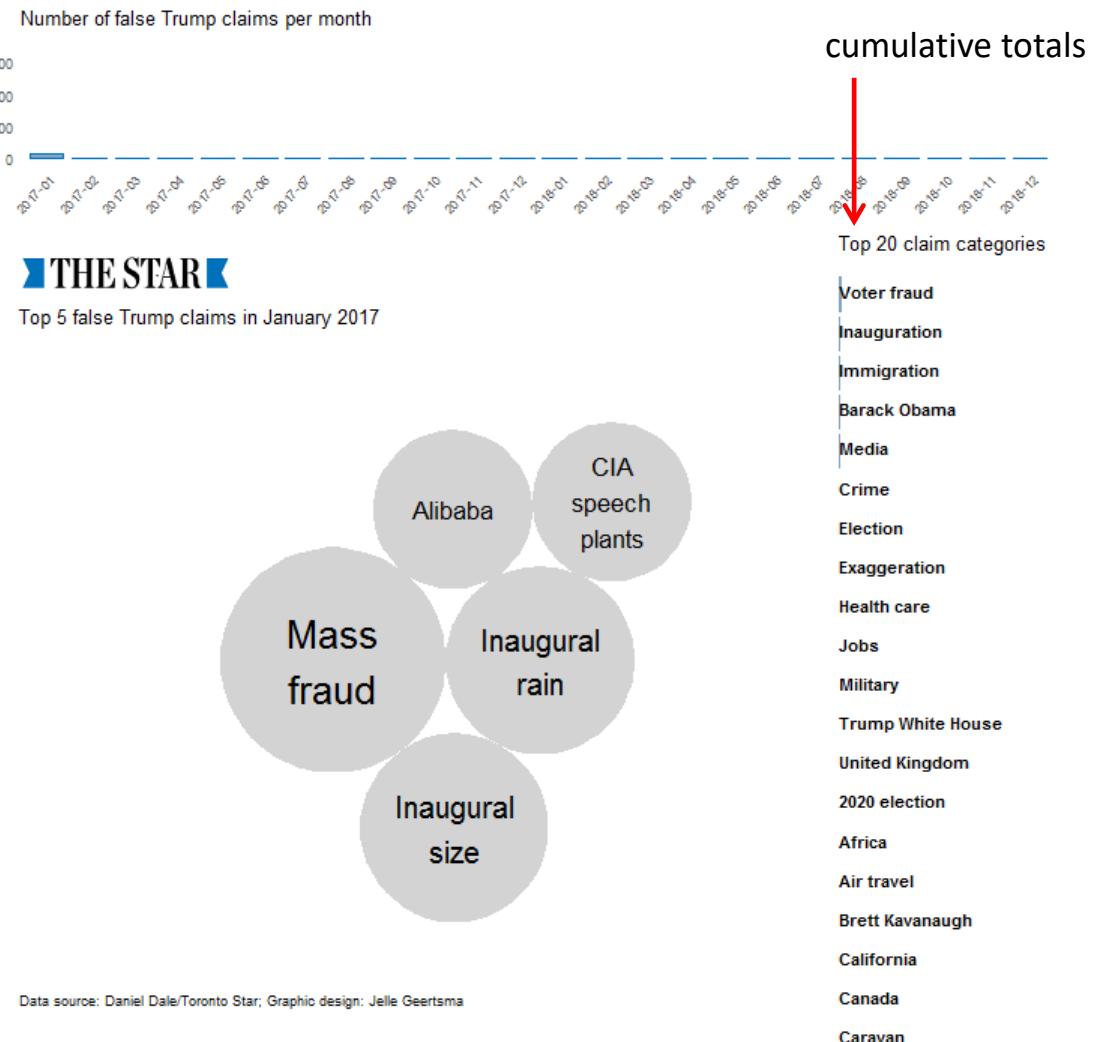
Graphic artist Jelle Geertsma designed this animated graphic showing the top 5 claim “slugs” & the top 20 categories over time, by month.

Bars at right show cumulative totals

This animation uses the ganimate package.

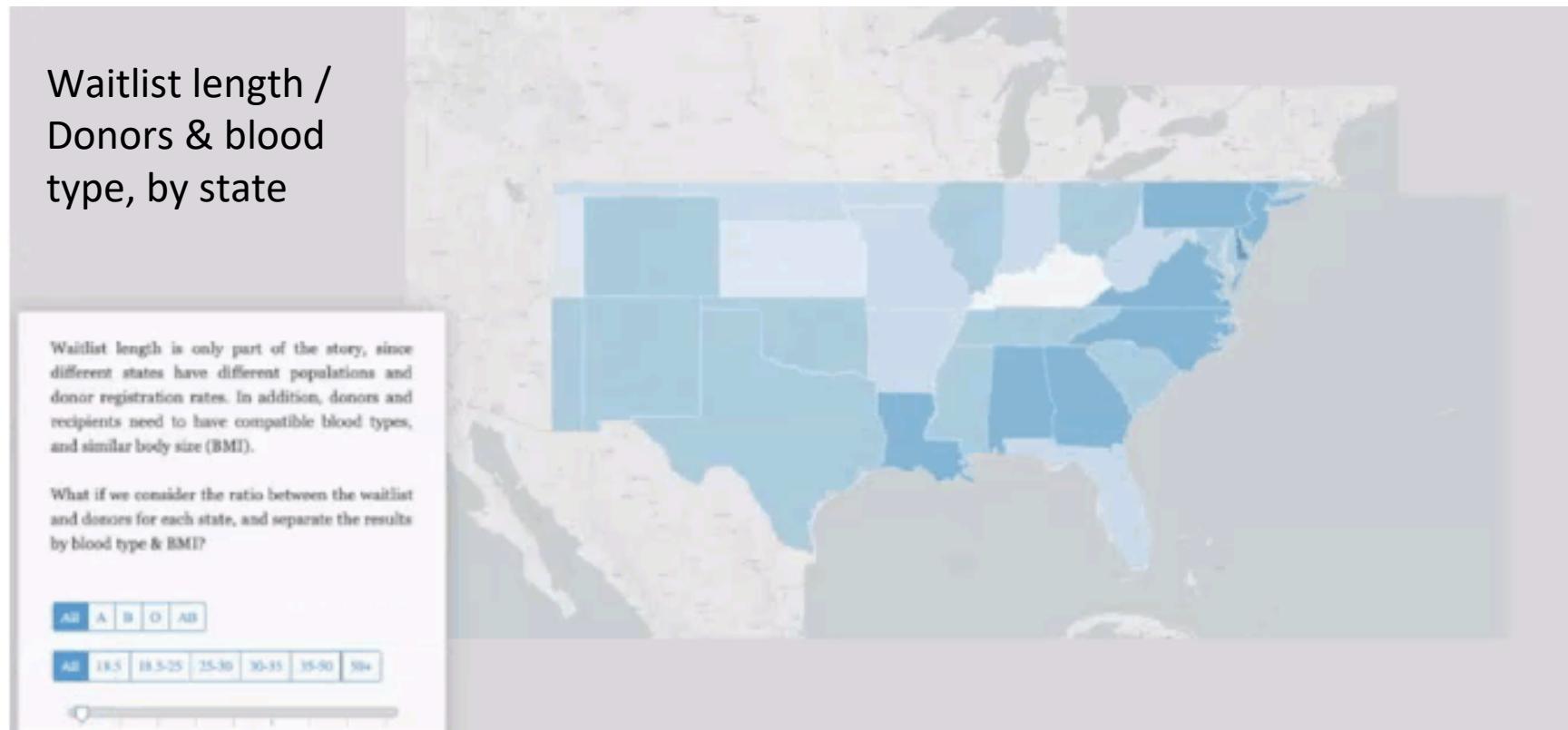
A TrumpLies package is in progress. **Ugh! Not so.**

<https://github.com/friendly/TrumpLies>



Scrollytelling: The Impatient List

What it's like waiting for a kidney transplant in the U.S.?
More importantly: What can I do?



["The Impatient List"](#), an example of Scrollytelling by Gupta et al.

Scrollytelling: The Impatient List

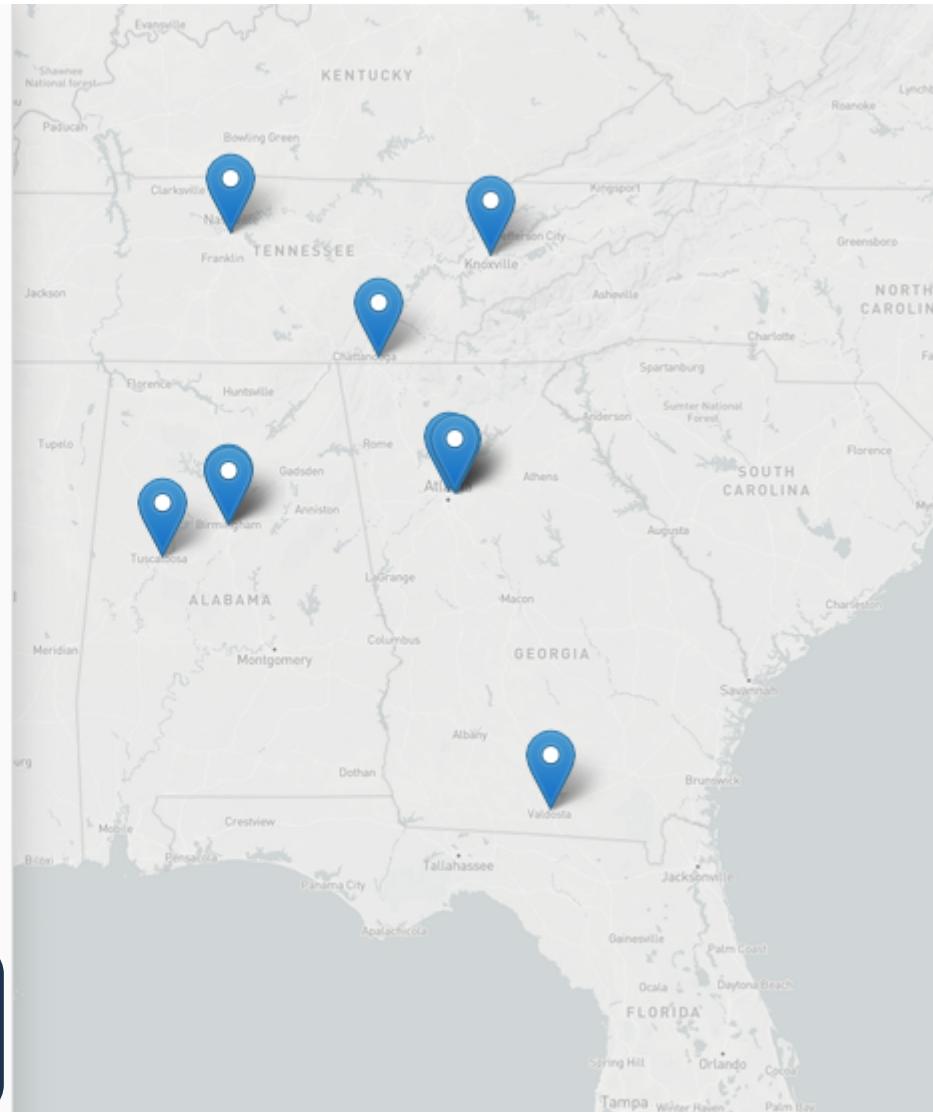
Since Kidney Transplant can be time sensitive, it is best for you to choose the ones adjacent to where you live. Please enter your zip code to view centers around you:

30318



You'll also need to compare between centers around you for their death rates, survival rates, and waitlist waiting time, in order to find the optimal center to register as a candidate. How would you balance these different factors, when each choice you make might influence your chance of getting a transplant to being cured?

Click on each pin on the map to view the data for each center compared to other centers around you.



Scrollytelling: The Impatient List



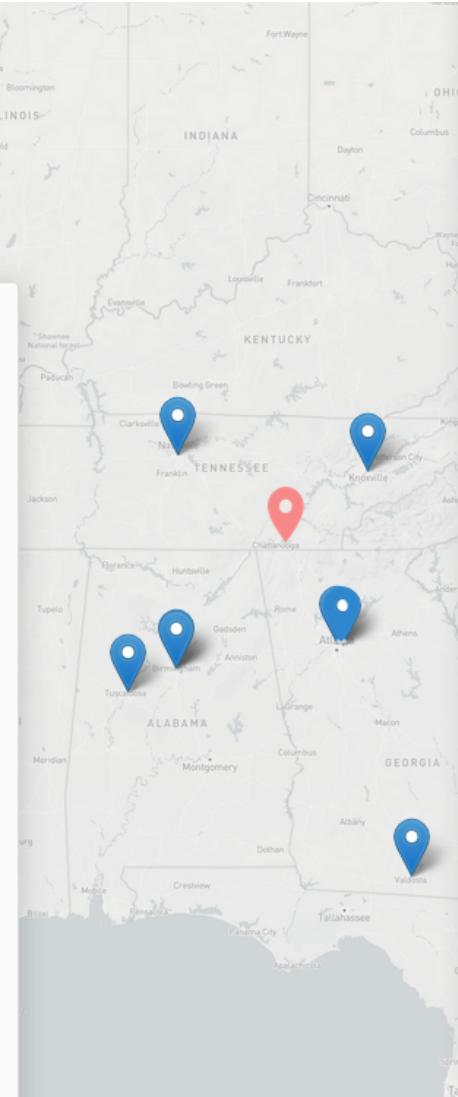
Since Kidney Transplant can be time sensitive, it is best for you to choose the ones adjacent to where you live. Please enter your zip code to view centers around you:

30318

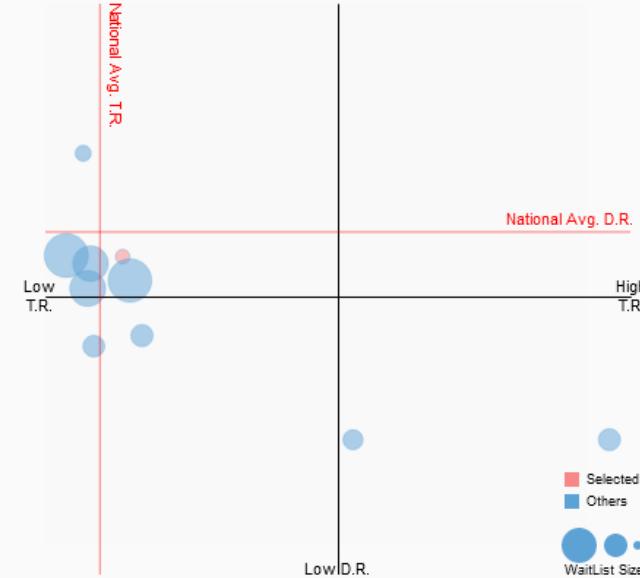


You'll also need to compare between centers around you for their death rates, survival rates, and waitlist waiting time, in order to find the optimal center to register as a candidate. How would you balance these different factors, when each choice you make might influence your chance of getting a transplant to being cured?

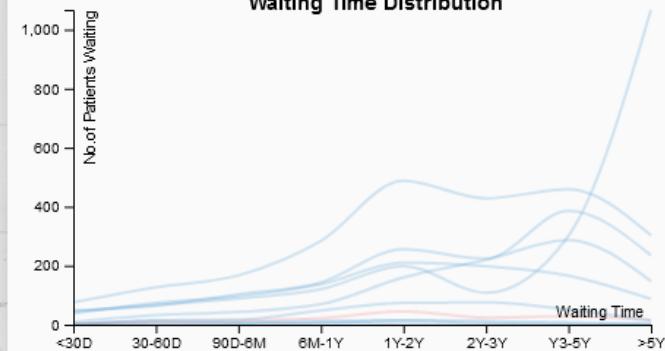
Click on each pin on the map to view the data for each center compared to other centers around you.



Death Rate (D.R.) Vs Transplant Rate(T.R.)
High D.R.



Waiting Time Distribution



Drill-down to tell a story

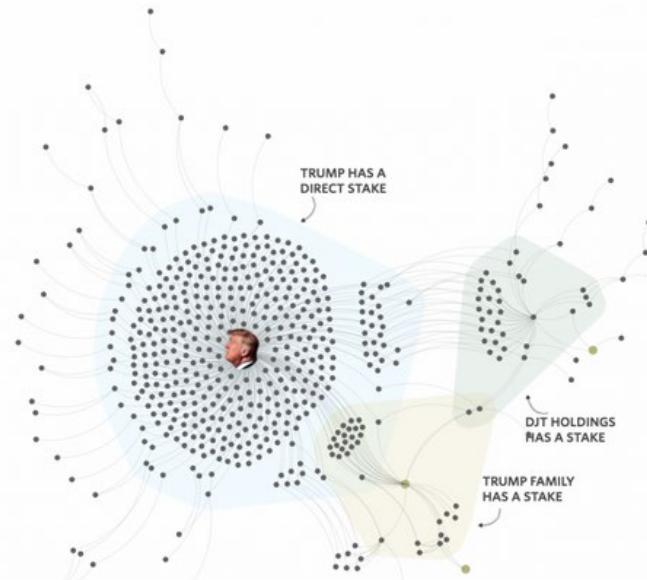
The *Wall Street Journal* published this interactive story of potential conflicts of interest in the business holdings of Donald Trump and his family (2017)

Trump, His Children, and 500+ Potential Conflicts of Interest

President-elect Donald Trump's complex holdings place him in an unprecedented position

By **Joel Eastwood, Coulter Jones and Julia Wolfe**

Published Jan. 19, 2017 at 6:00 a.m. ET

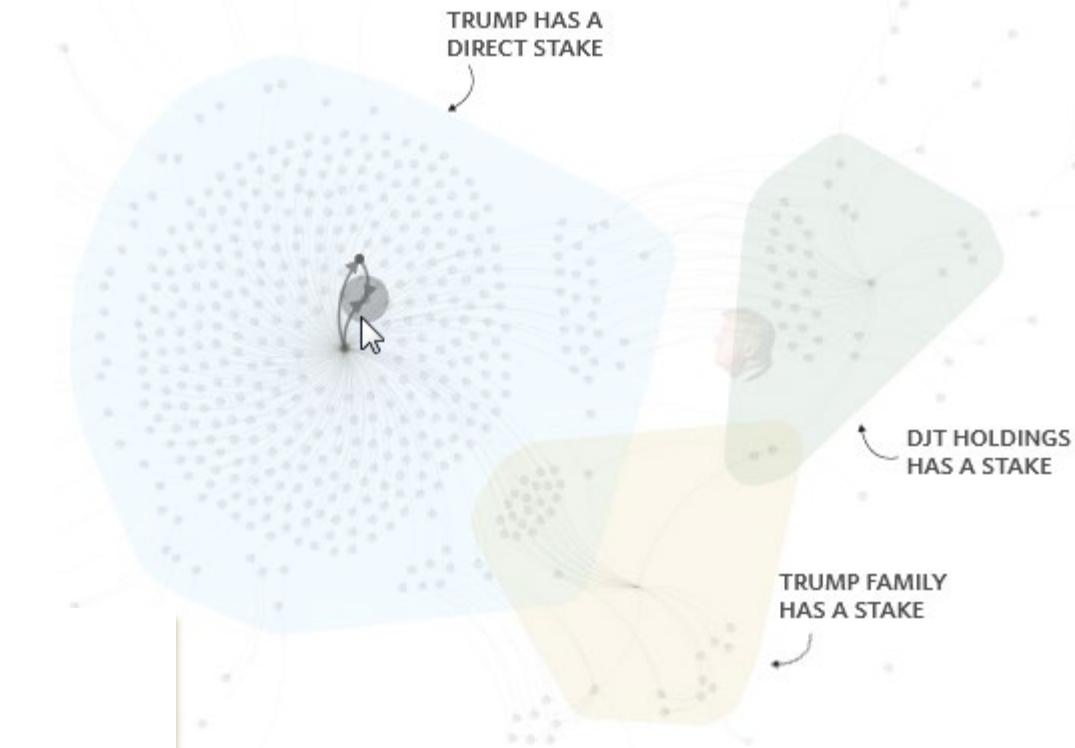


President Trump's network of conflicts of interest. Credit: WSJ.

"The most difficult part of many visualization projects, and certainly of this one, was distilling a complex, messy dataset into a clean visualization with a clear takeaway," says Eastwood.

<https://www.wsj.com/graphics/donald-trump-potential-conflicts-of-interest/>

In each part of the story, hovering with the mouse reveals the details of a particular investment and links to others



EVERY HOLDING IS A POTENTIAL CONFLICT OF INTEREST

Mr. Trump is at the center of a complex network of business holdings that, according to the May 2016 financial disclosures he filed with the federal government, was made up of more than 500 businesses including hotels, condo towers, golf courses and foreign and domestic licensing deals. His personal connections to such a vast empire eclipse the business interests of **any other prominent politician** in modern times.

TRUMP MARKS PANAMA LLC

New York, N.Y. USA

License deal with K GROUP DEVELOPERS INC.

ASSET

Type: license deal | Value: \$15,001 to \$50,000

INCOME

Type: royalties | Amount: \$1,000,001 to \$5,000,000

Trump Financial Disclosure Report (May 2016)

British MP Expense Scandal

In May 2009, an unknown government employee copied a terabyte of documents on MP expenses and released these to *The Daily Telegraph*

MPs' expenses: how scoop came to light – and why journalists fear a 'knock on the door'

The Daily Telegraph's exclusive was offered to at least three other papers by shadowy figure described as a 'rank amateur'

The Daily Telegraph

Friday, May 8, 2009 BRITAIN'S BEST-SELLING QUALITY DAILY 1st Republic £1.10 No. 47,077 90p

The truth about the Cabinet's expenses

The Prime Minister

Brown paid his brother more than £6,000 for 'cleaning services'

The Justice Secretary

Straw had to refund money he claimed in error for council tax

The Chancellor

Darling changed 'second home' designation four times in four years

The Business Secretary

Mandelson claimed thousands after saying he would quit as MP

The Guardian: Crowd sourcing

In June 2009, The British House of Commons released 700,000 documents detailing every expense claim over for 4 years by all 644 MPs.

The Guardian launched a crowd sourced experiment inviting readers to investigate these claims

Investigate your MP's expenses

Join us in digging through the documents of MPs' expenses to identify individual claims, or documents that you think merit further investigation. You can work through your own MP's expenses, or just hit the button below to start reviewing. (Update, Fri pm: we now have a virtually complete set of expenses documents so you should be able to find your MP's) Already created an account? [Log in here](#).

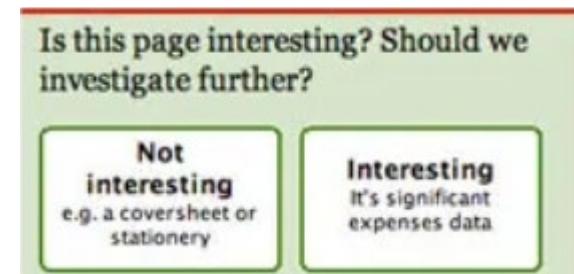
We have **458,832** pages of documents. **32,755** of you have reviewed **225,443** of them. Only **233,389** to go...

[Start reviewing](#)

Please read our [privacy policy](#) to find out how we use your data. You must also read our [terms of service](#). By reviewing pages, you are agreeing that you have read the terms of service, and that you agree to them.

Crowd sourcing: The App

- “Using our system, you will be able to find your MP - or any member - and look at their records directly. For every page for every MP, you will be able to:
 - comment on individual expenses
 - highlight ones of interest
 - tell us how interesting that receipt is
 - Help by entering the numbers on the page
- Using that information, we will begin to be able to piece together a unique picture of how MPs claims have changed over time.

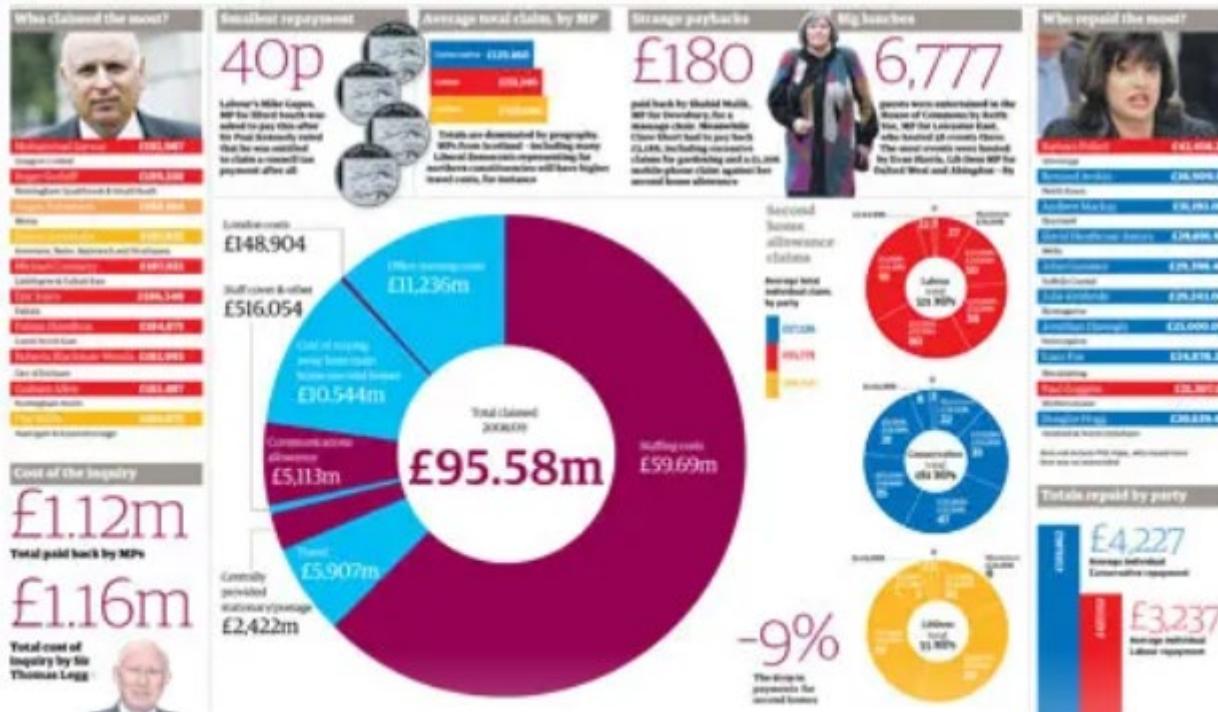


MPs' expenses: how we visualised the data

How do you display a complicated set of numbers on the story of the week? See how we did it

- Who claimed what?
 - The Legg report's full list of repayments

A year in expenses Parliament yesterday released the full details of MPs' claims for 2008-09. This is a snapshot.



MP expenses visualised. Click on image for the full graphic

Panama Papers

In January 2017, The *International Consortium of Investigative Journalists* released the “Panama Papers”, exposing the tax avoidance schemes of the Panama-based firm Mossack Fonseca.

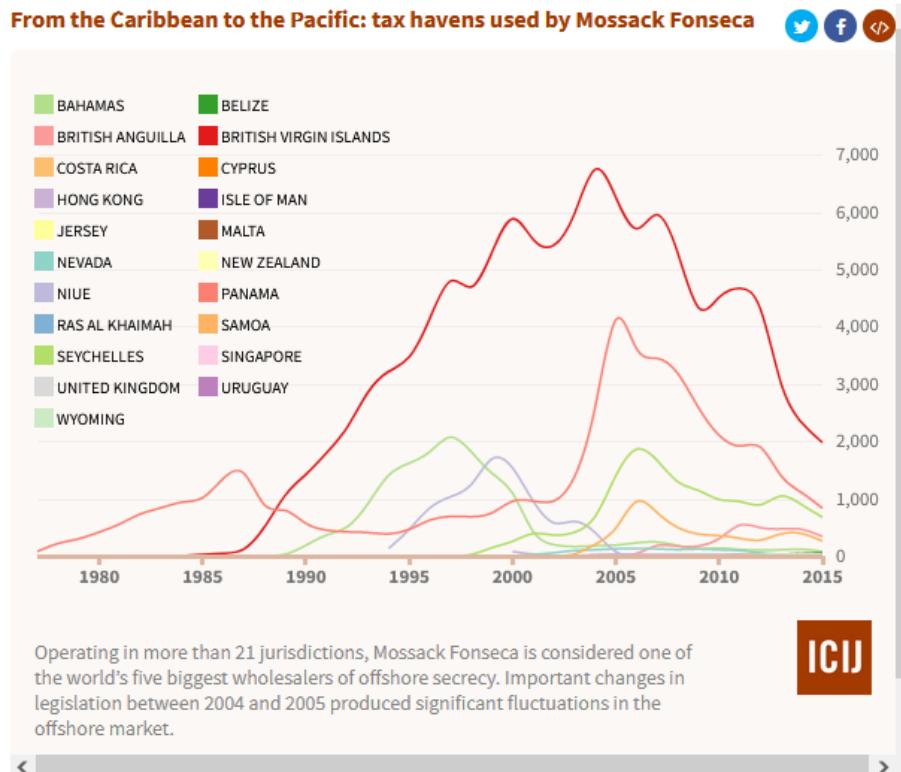
Two terabytes of data, over 40 years, 210,000 companies in 21 offshore jurisdictions
IJIC: 280 top investigative journalists, >100 media partners, >100 countries



Panama Papers

Key findings:

- ❖ Files reveal offshore holdings of 140 politicians & public officials around the world
- ❖ Current & former world leaders include: PM of Iceland, the king of Saudi Arabia, ...
- ❖ >214,000 offshore entities, connected to people in more than 200 countries
- ❖ Major banks have driven the creation of hard-to-trace companies in offshore havens



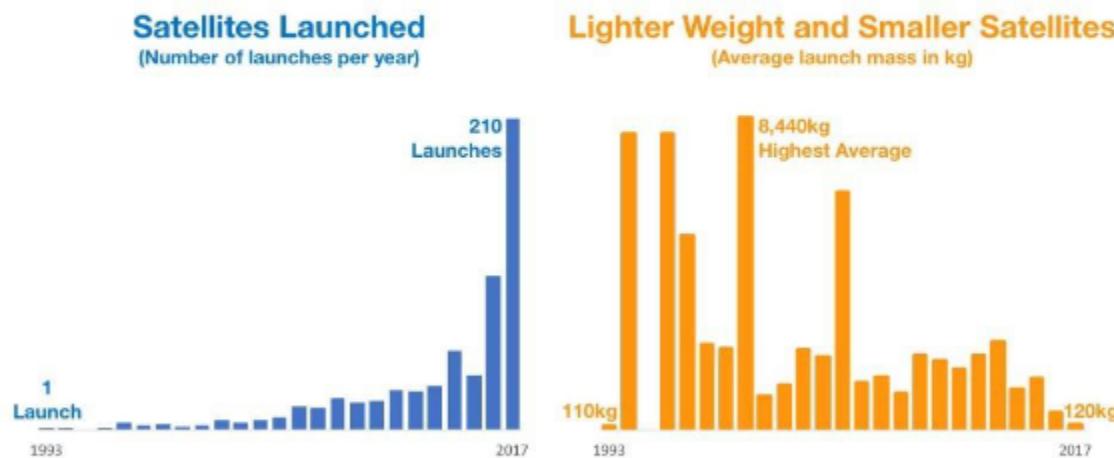
How could this graph be improved?

Satellite Journalism

- An exponential growth in the number of small, privately owned satellites has opened wide new vistas for investigative reporting.
- Media access to satellite company images data bases has resulted in journalists gaining a near real-time reporting source for breaking stories.

Smaller Satellites are a Growing Trend

Between 2003-2017, 594 Earth observation satellites were launched, compared to 26 in the previous decade. While the number of satellites launched has increased significantly, the average size (launch mass) has decreased dramatically.



SOURCE: Union for Concerned Scientists database, <https://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database>

(Infographic – courtesy Anne Hale Migliarese Radiant.Earth)

Satellite Journalism

- Satellite imagery has provided ‘smoking-gun’ evidence for:
 - investigative reporting of human rights abuses (Htusan et al., 2015)
 - exposure of environmental destruction in remote locations (e.g., the Amazon: Weir, 2017)
 - before & after imagery of natural disasters (Murphy, 2018)
 - analyses of military build-ups (e.g.: China’s construction of artificial islands in the South China Sea Wu et al., 2018)
 - missile launches and nuclear testing in North Korea (Fifield and Taylor, 2018)

Source: Mark Corcoran (2018), *Satellite Journalism – The Big Picture*,
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-10/Mark%20Corcoran%2C%20Satellite%20Journalism_1.pdf

The La Pampa gold mine in Peru

In 2016, mining operations were extended S. of the Malinowski River, illegally entering the Tambopata National Reserve—a protected forest.

The Amazon Conservation Association used Planet Labs data to track 100s of hectares of illegal expansion and mapped alterations to the course of the Malinowski River.



January 29, 2016



July 21, 2016



November 4, 2016



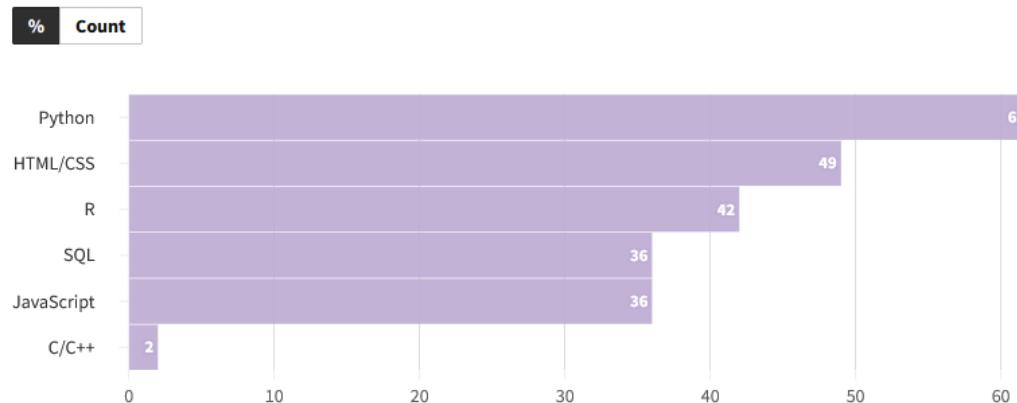
January 20, 2017

The Peruvian government intervened and actively targeted illegal mining operations inside the reserve.

Images: Planet Labs

Software for DDJ

Which programming languages do you use for your work, if any?

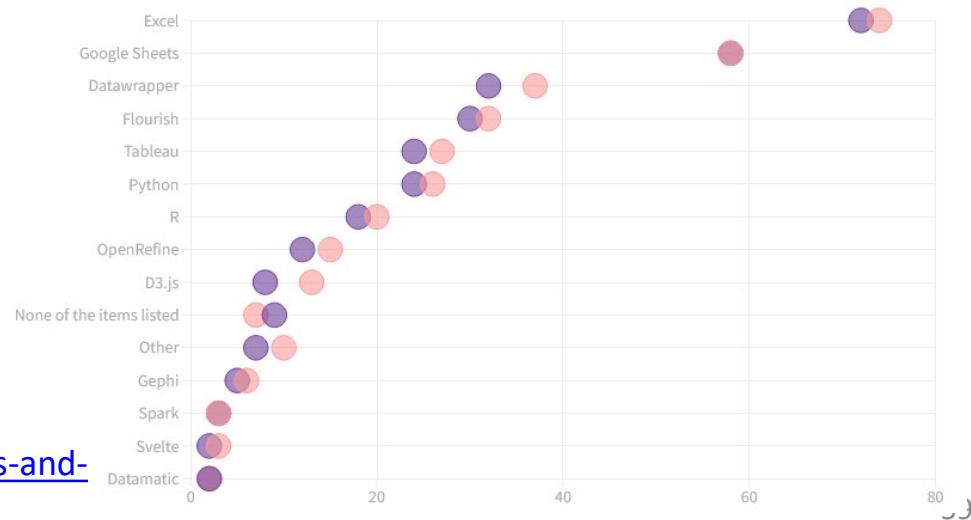


Of those who program,
most use Python & HTML,
followed by R

Data analysis & dataviz mostly
done with Excel & Google sheets

Which of the following data analysis or visualisation tools do you use for your work, if any?

Year ■ 2021 ■ 2022



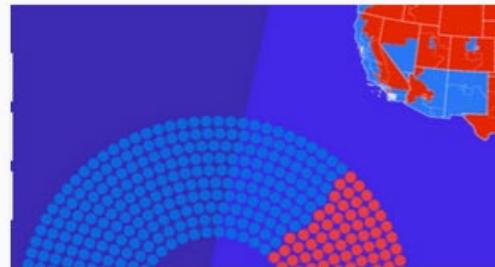
<https://datajournalism.com/survey/2022/skills-and-tools/>

Popular DDJ tool for creating interactive, dynamic stories and datavis



Scrollytelling

Create immersive digital experiences and bring your data to life with our no-code scrollytelling solution



Elections data

Learn how you can bring elections data to life with interactive graphics



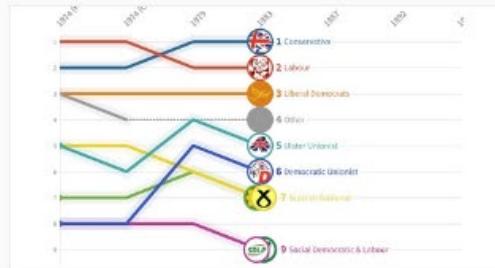
Visualize football and sports data

See creative data visualization ideas for your EUROS, World Cup and Olympic Games coverage



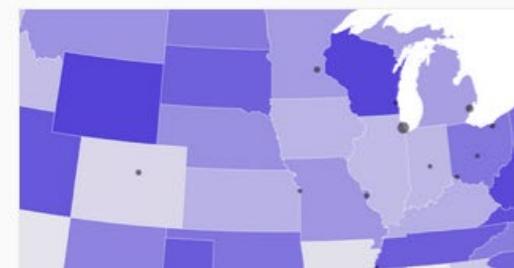
Data Explorer

Morph between various chart and map types to create rich and compelling data stories



Plotting competitors

A “racing” line chart template with images and ranks view



How to choose the right map?

From 3D to symbol maps, use our interactive map templates to visualize geospatial data

See: <https://flourish.studio/>

Learning more

The DataJournalism.com blog presents a roundup of the best data journalism projects in 2021, <https://datajournalism.com/read/blog/best-data-journalism-projects-2021>



Summary

- DDJ reflects increased interaction between journalists and {visualization design, computer science and statistics} in the digital age.
- New data sources:
 - FOI requests
 - Open Data initiatives
 - Satellite journalism
 - Crowd sourcing
- New graphic forms & reporting
 - Novel graphic displays
 - Interactive story telling
- Powerful impact