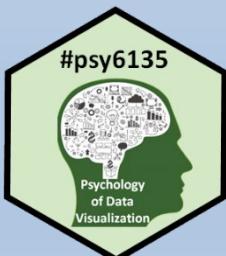
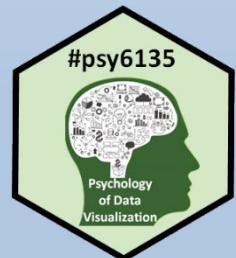


Data Journalism

Michael Friendly Psych 6135

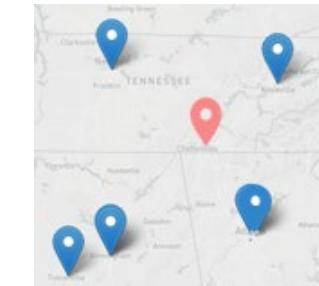
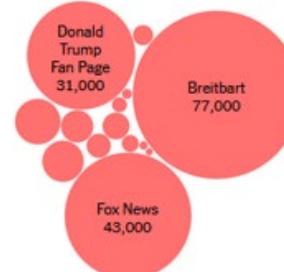
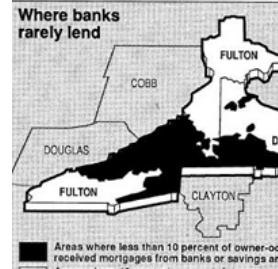


<https://friendly.github.io/6135/>



Topics

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



What is data journalism?

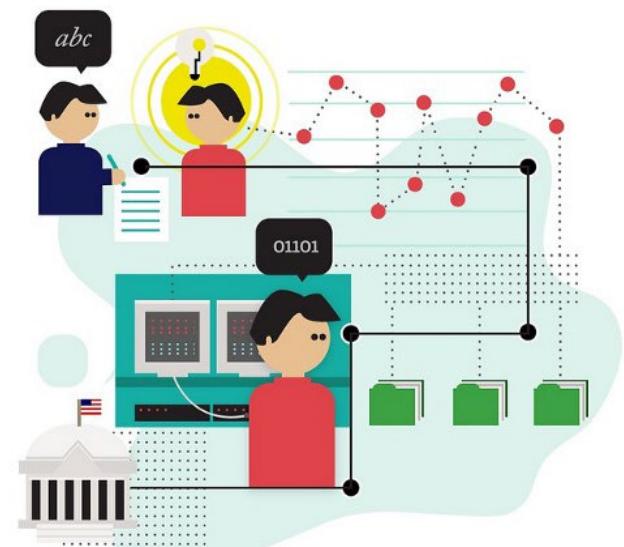
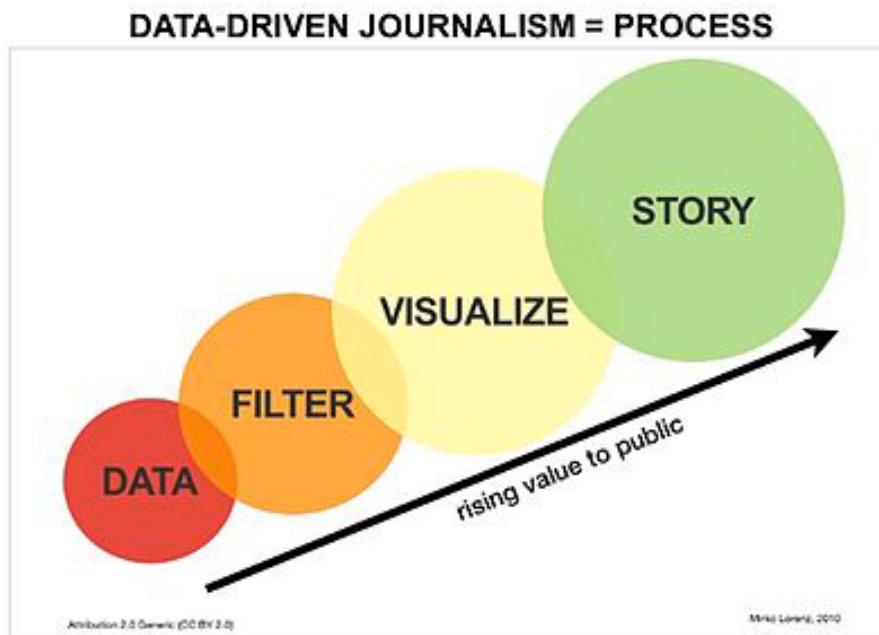
- **Data journalism** reflects the increased role of numerical data for reporting in the digital era. **Data driven** or **DDJ** more useful
 - Reflects increased interaction between journalists and {visualization design, computer science and statistics}.
- Aspects:
 - **Database analysis**: FOI requests 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?



“Data” in Data Journalism

Investigative reporting generates messy databases from various sources

- Challenges: Organize, clean, filter, ...
 - Goals: Find the story; make it credible



From: Jim Rankin, [Data Successes](#), 6135 Guest lecture, 2020

Examples of #DDJ

- The color of money
- Hurricane Andrew
- Predicting elections
- Election reporting
- Measles, vaccines, COVID
- Truth & Lies
- Gun deaths



Bill Dedman: *The Color of Money*

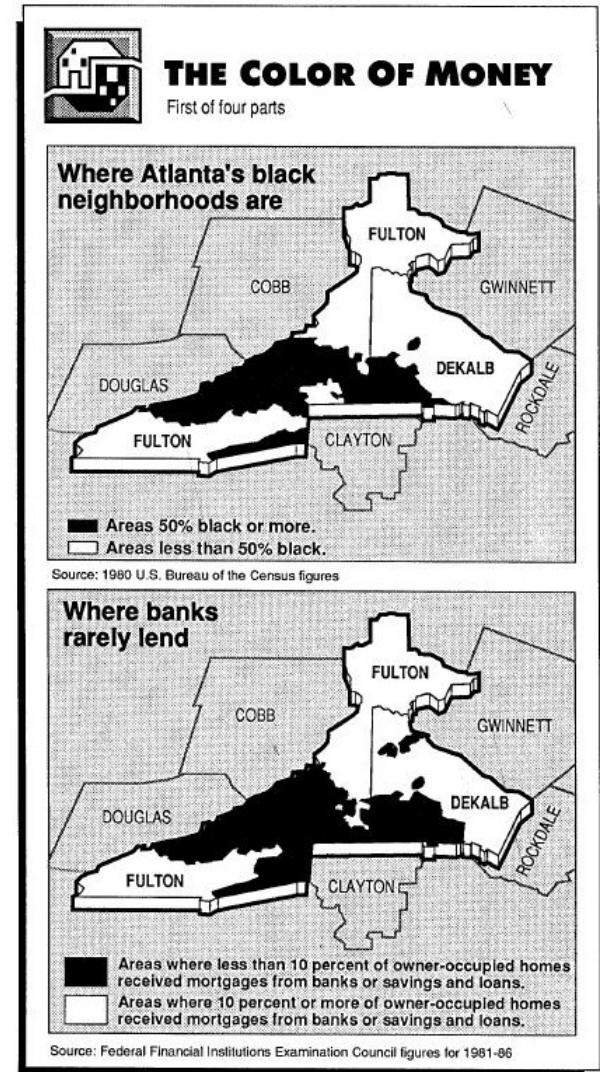
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

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



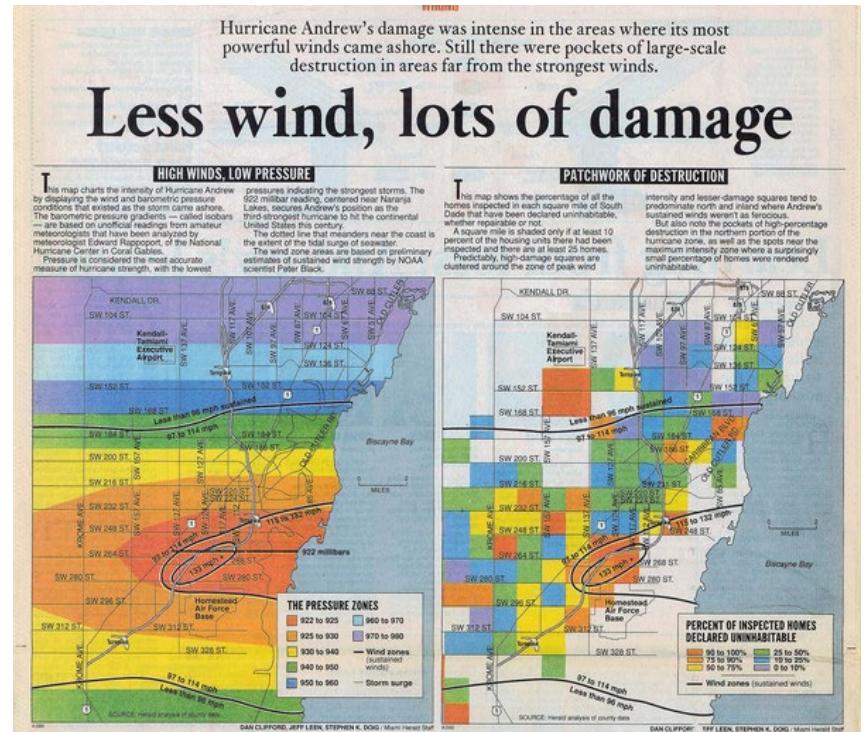
Hurricane Andrew

- ❖ 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>

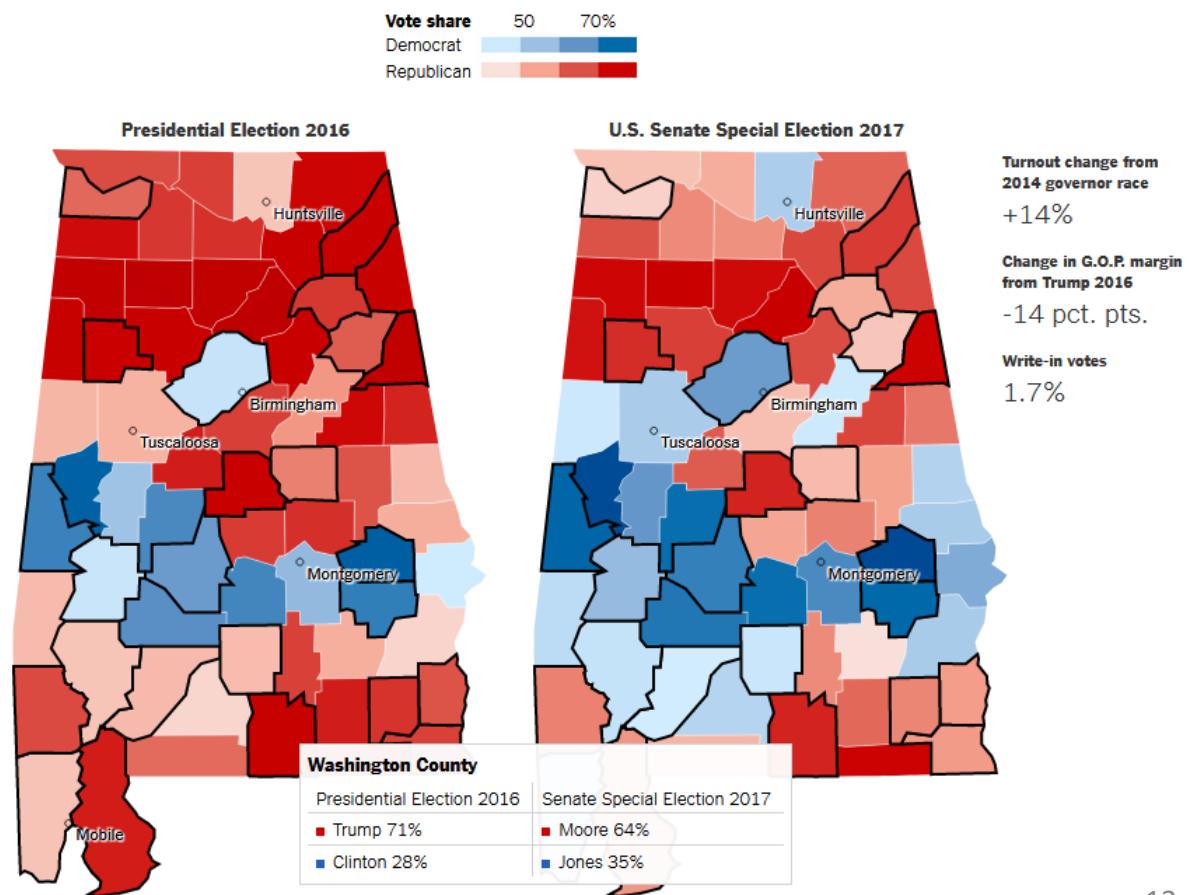
Election reporting: NYT

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

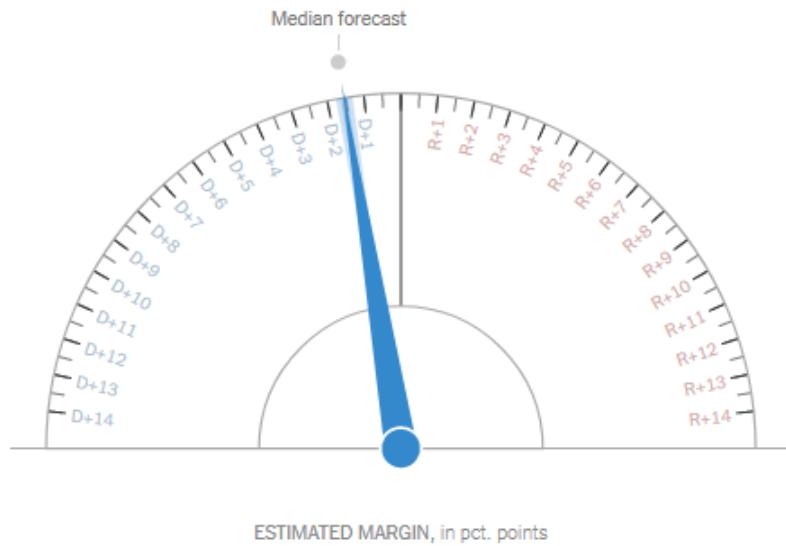
Q: What other graphic might show the comparison better?



Election reporting: Election needle

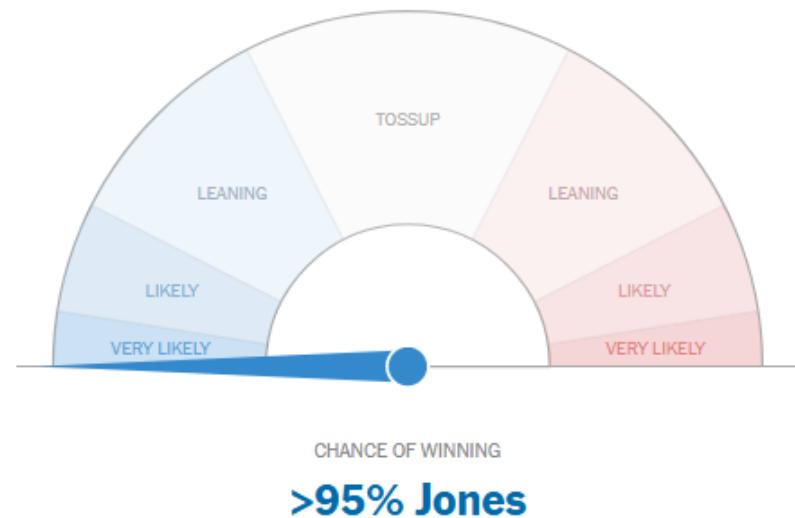
In the 2016 US election, the NYT used **dynamically updated** dashboard displays of the election returns and model-based predictions of the winner of the Alabama Senate race, predicting a win by the Democrat, Doug Jones, over rival Republican, Roy Moore.

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>

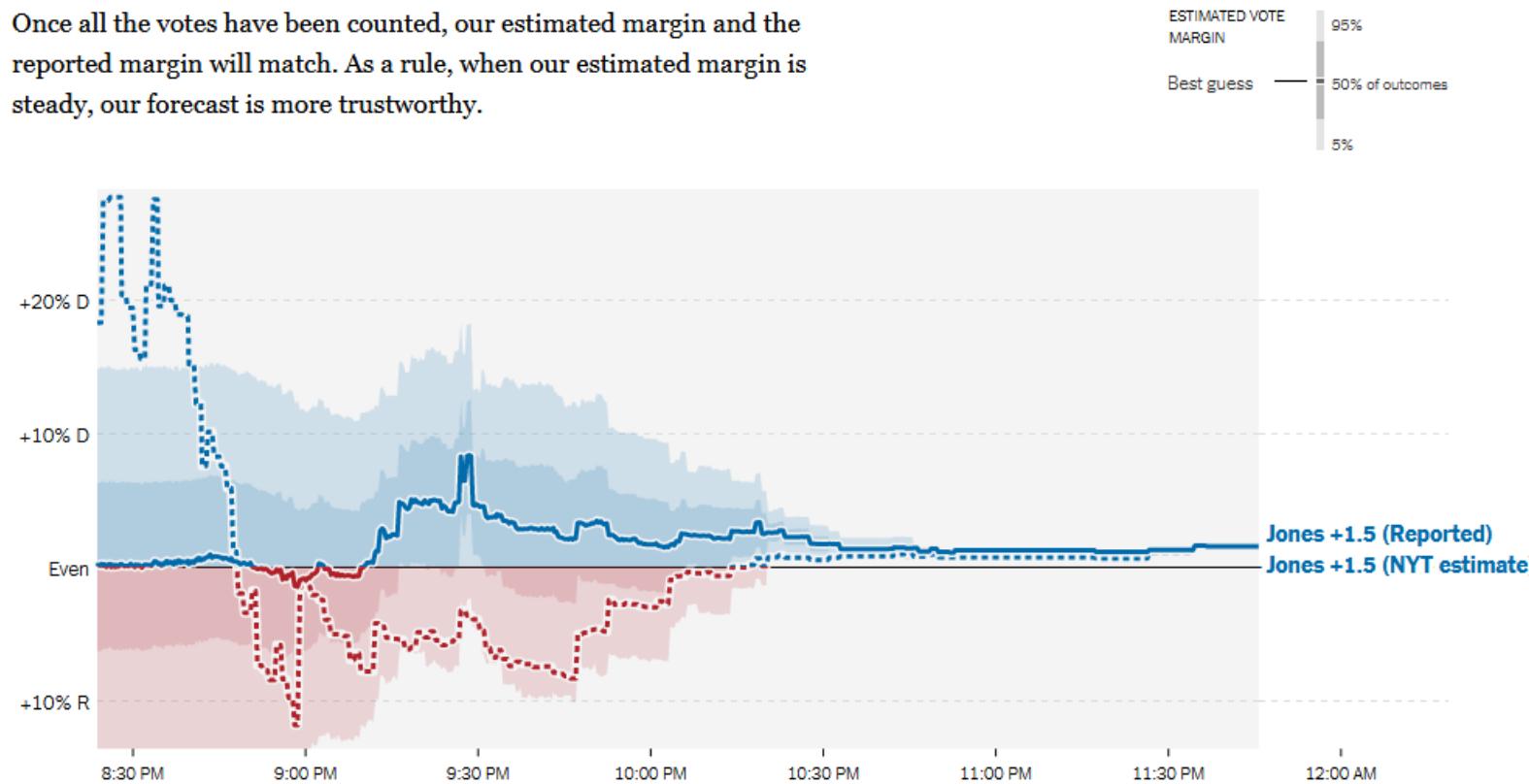
The moment our long national nightmare began

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

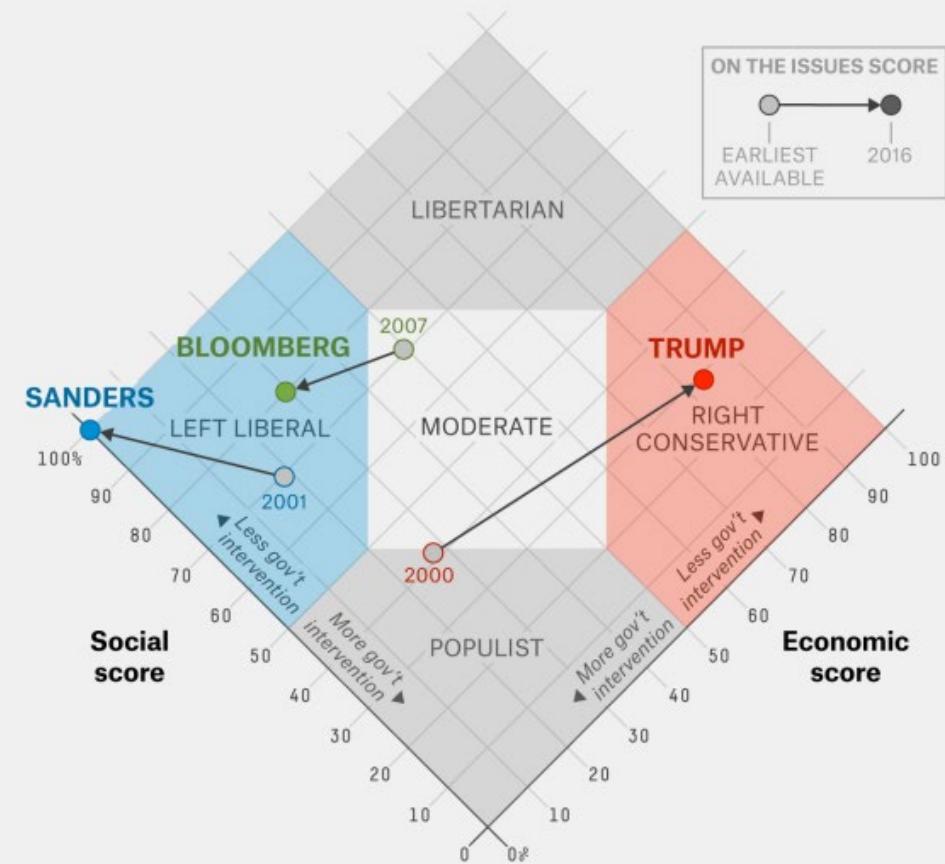
This graph, from fivethirtyeight.com was designed to show how some presidential candidates had shifted positions before the 2016 election.

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

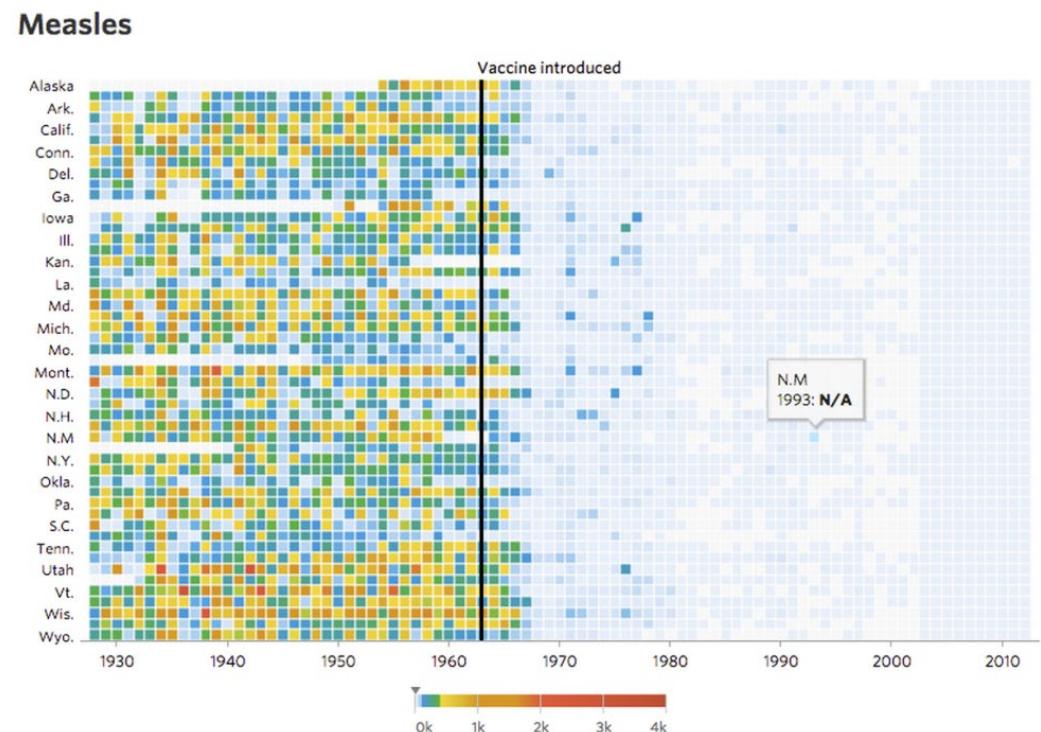
Visualizing the impact of health policy interventions

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.



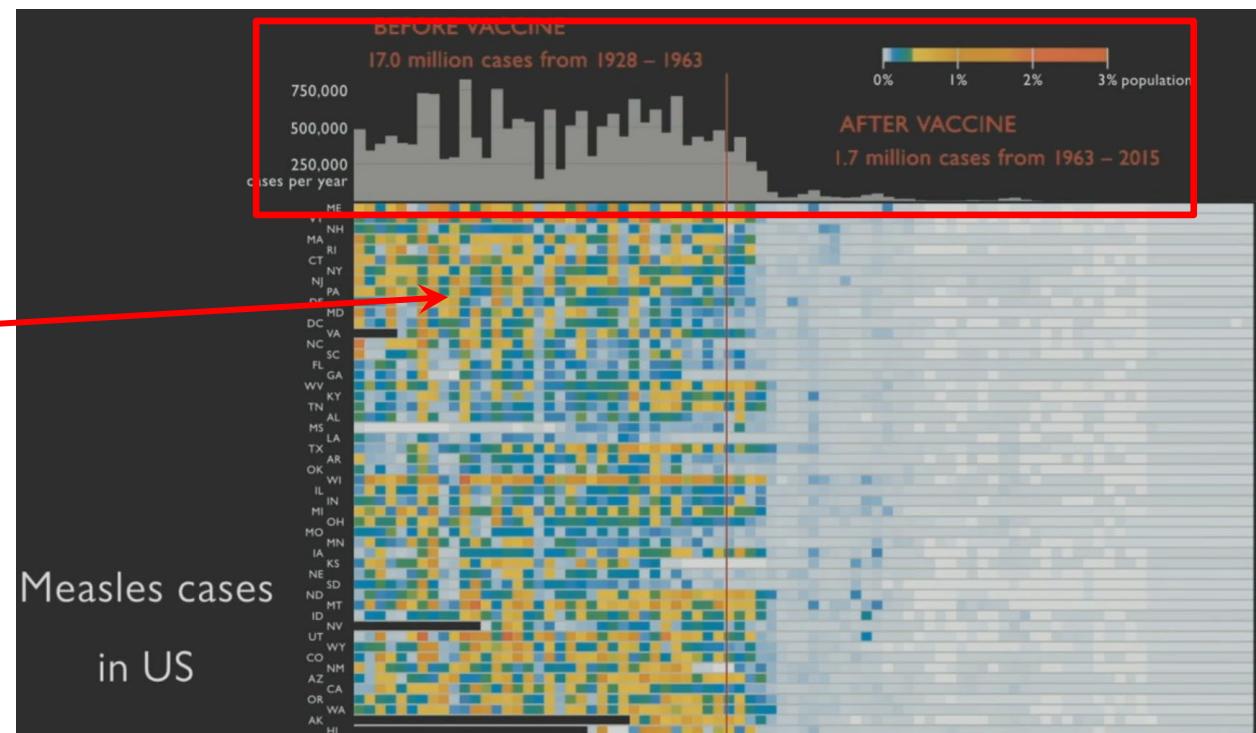
Making the message more explicit

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

Ben Schneiderman:
“Overview first;
details on demand”



Effect ordering & color scales

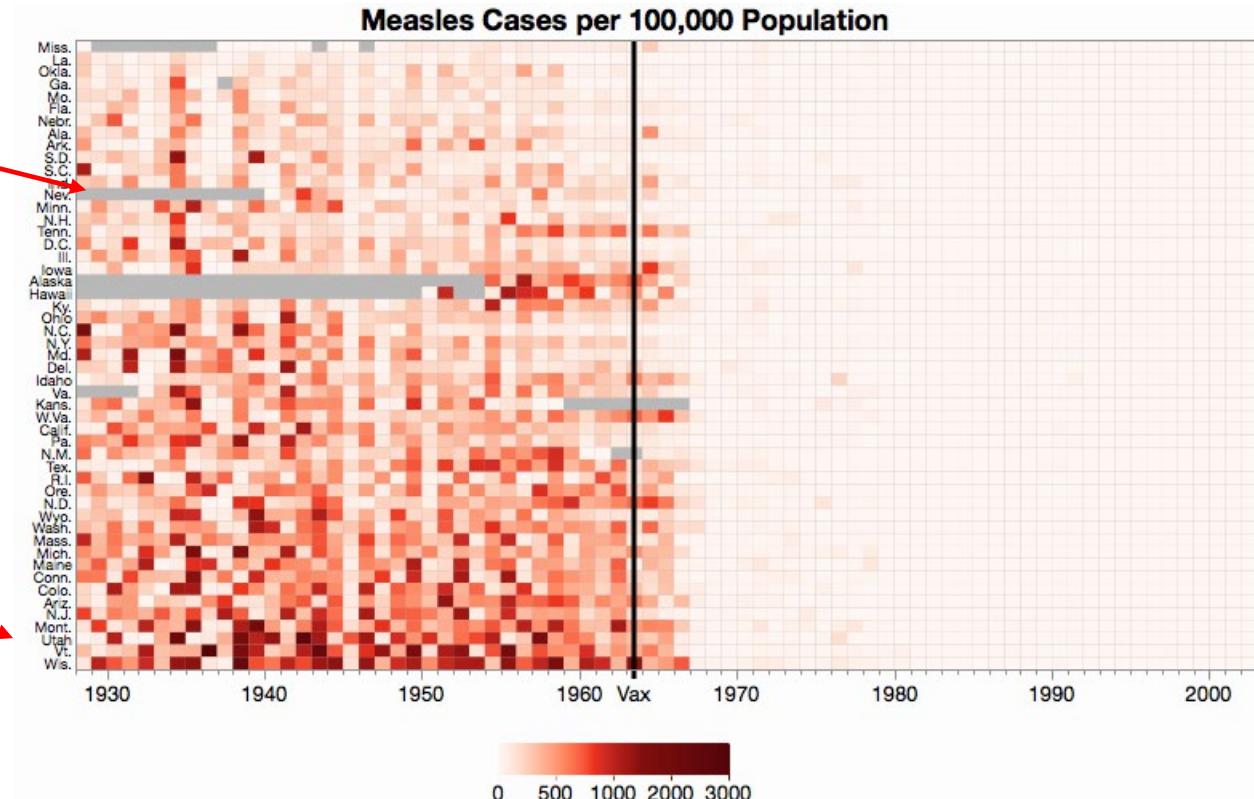
Ordering the states by an overall measure of measles incidence makes it easier to study details.

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.

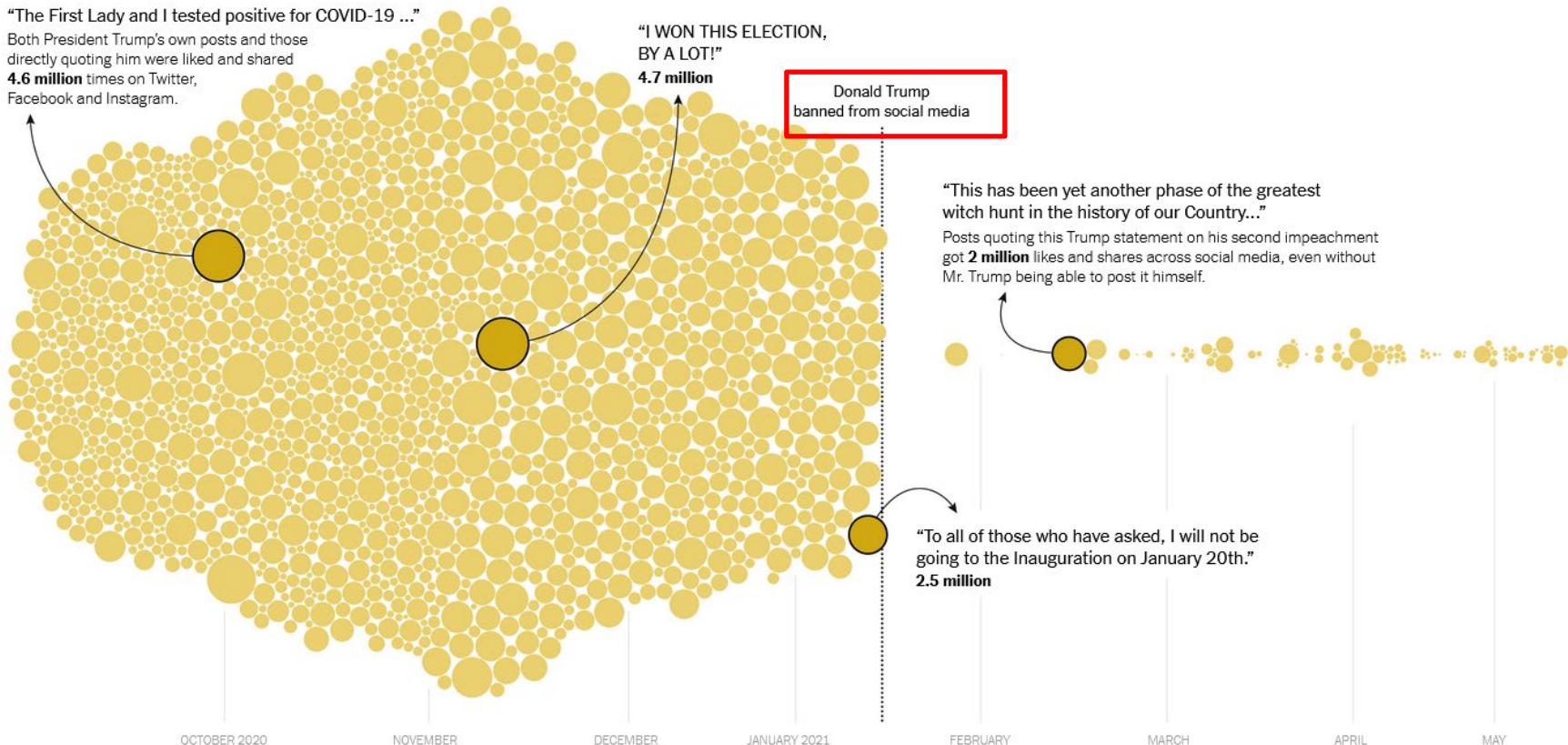
What's wrong in Wis. &
VT?



Before – After Circle Charts

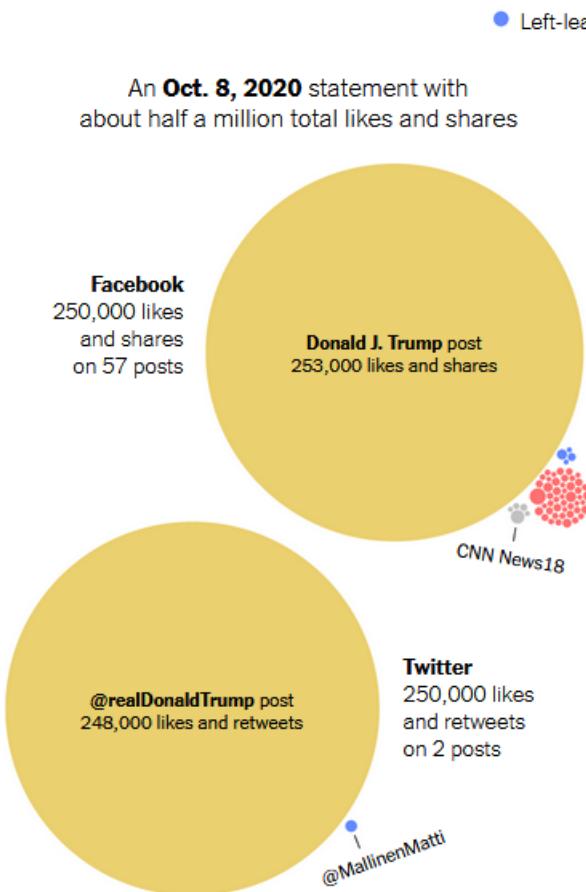
What Happened When Trump Was Banned on Social Media?

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

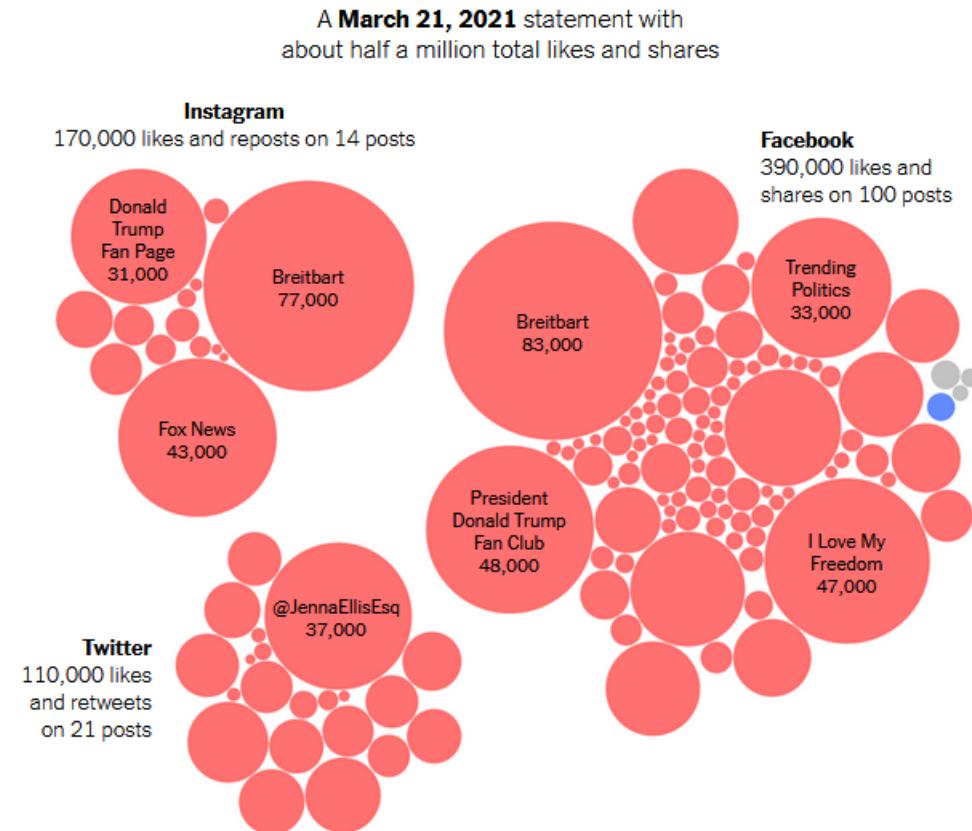


Before – After: Who pushed the lies?

How/Who Spread Trump statements before and after the ban



Posts referencing each Trump statement, sized by total number of likes and shares



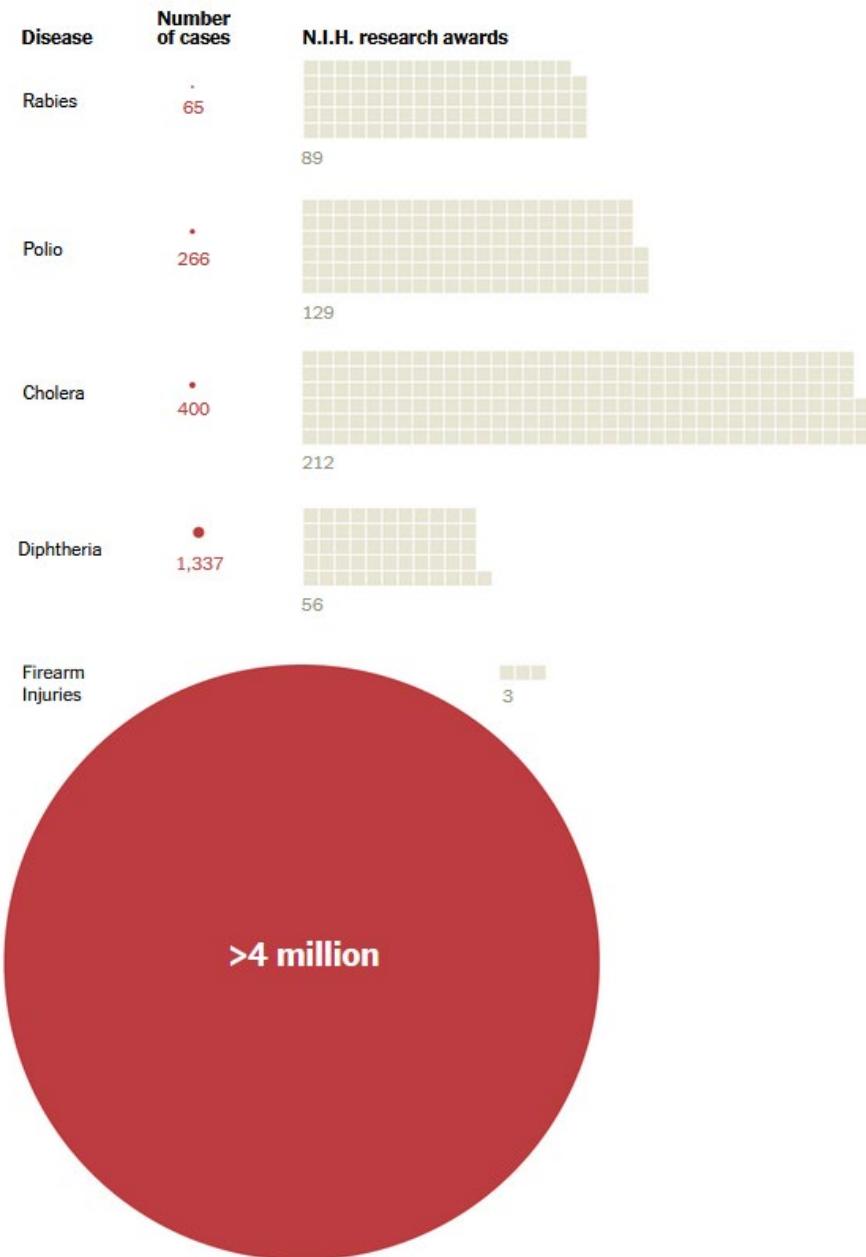
Health & research funding

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?



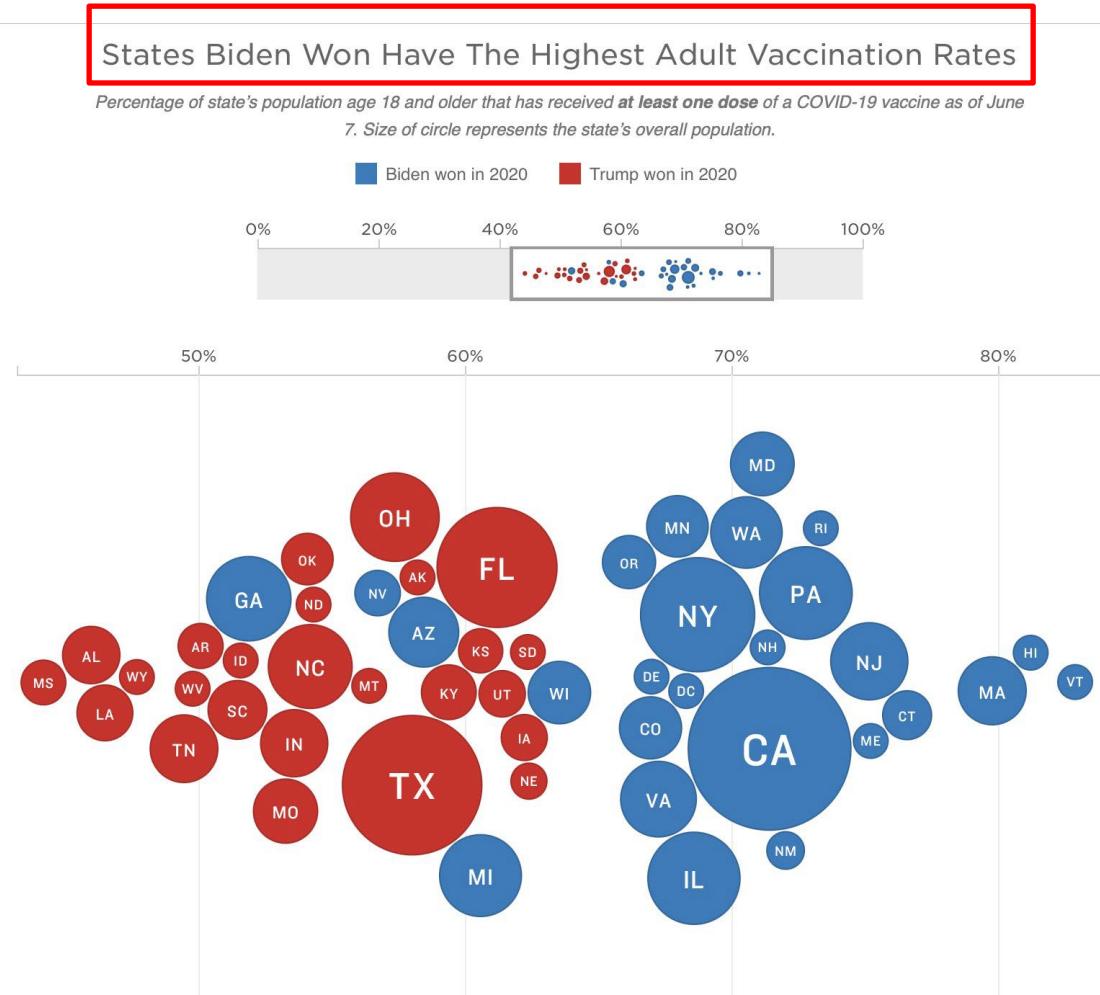
Source: University of Chicago Crime Lab

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>

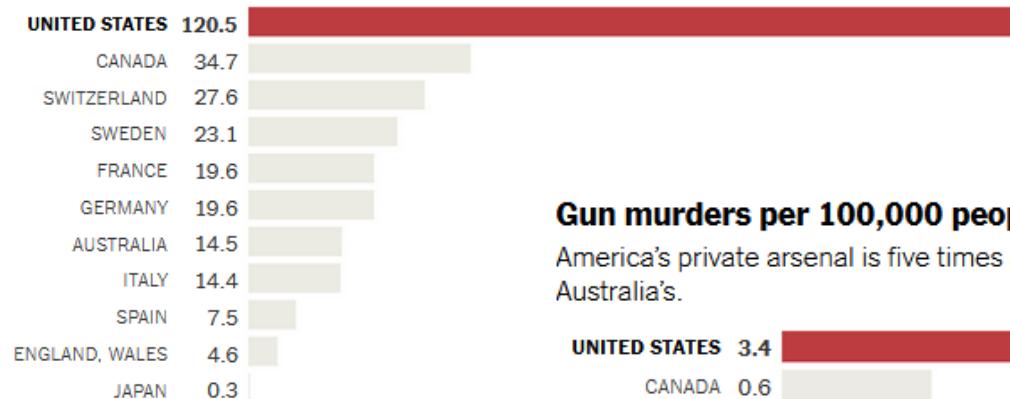
NYT Infographic: Gun deaths



NYT: Gun deaths & policy

Guns per 100 people

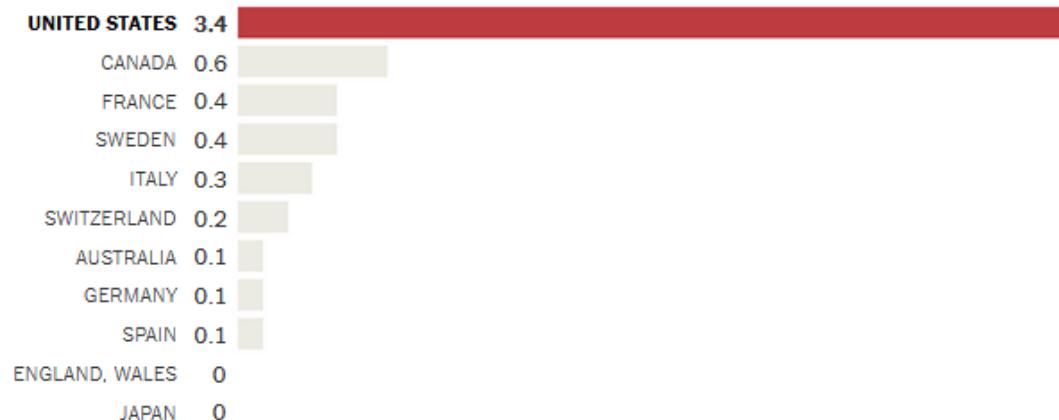
Among developed countries, the United States has by far the highest rate of firearms ownership.



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

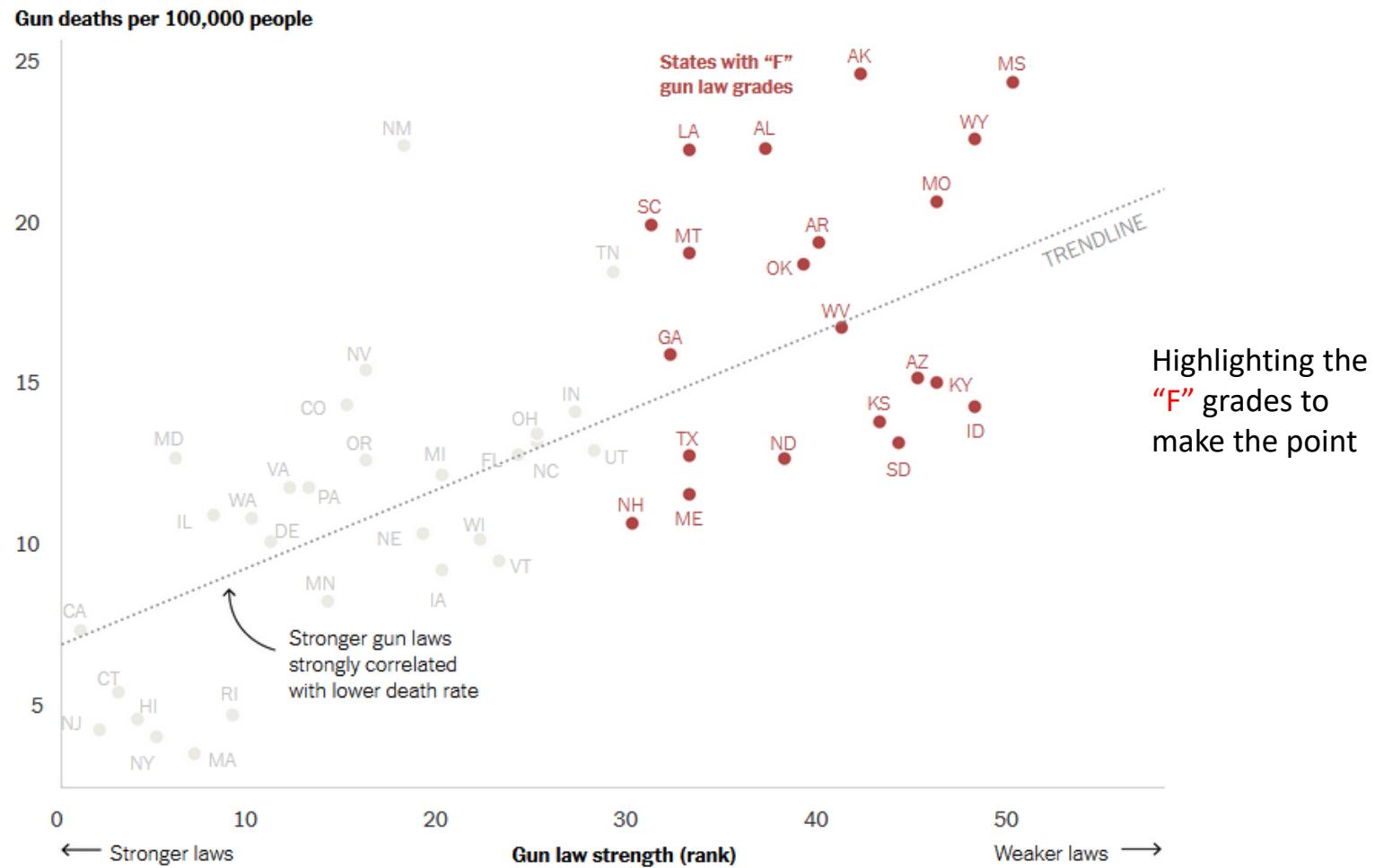
- What makes them effective?
- Could they be done better?

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.

See: NYT, “[Guns in America](#)” for other stories

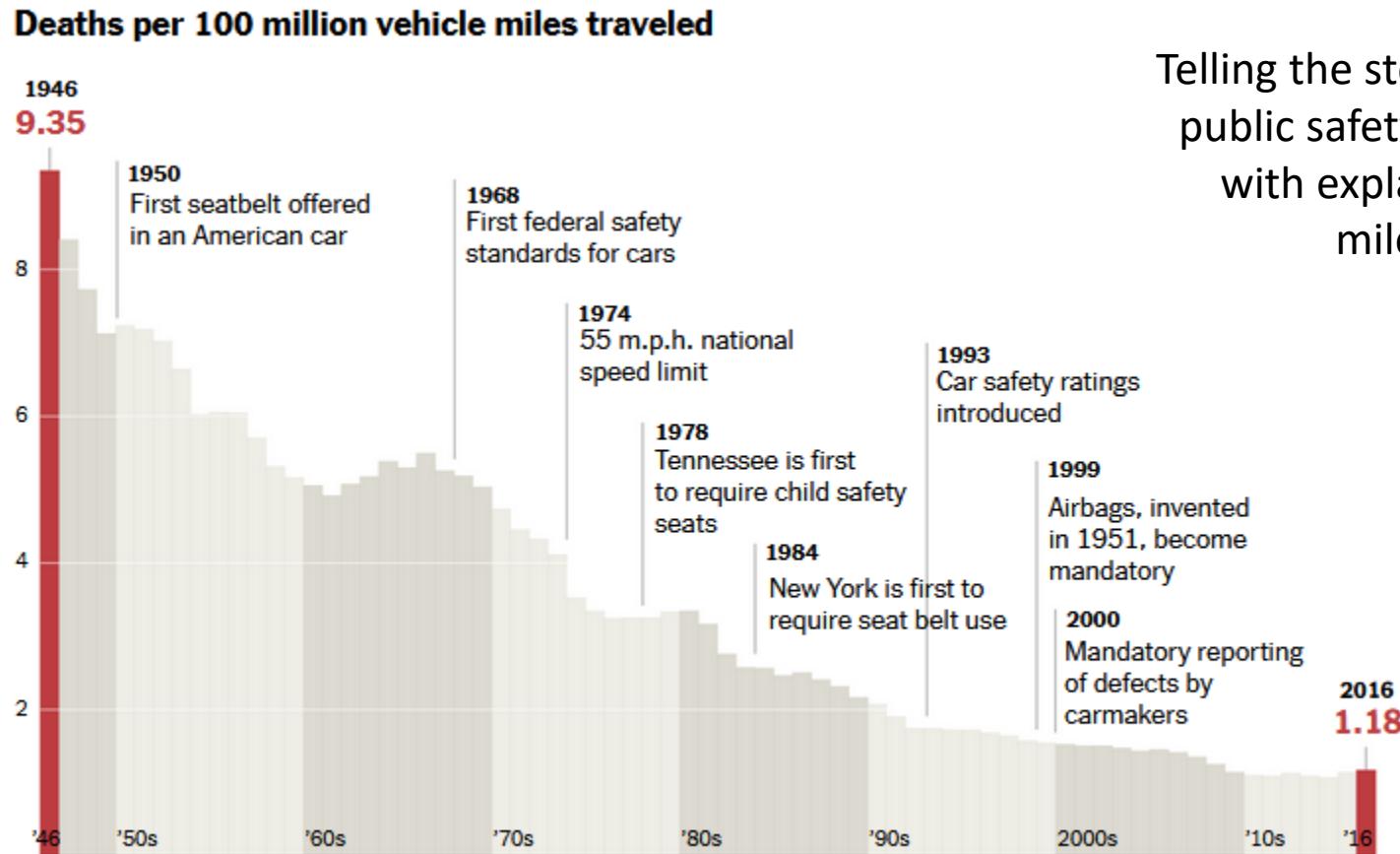
NYT: Gun deaths & gun laws

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



Seatbelts: Story telling

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



Source: National Highway Traffic Safety Administration



Race & Crime

Toronto Star investigation of racial disparities in treatment by Toronto Police Services

FOI request → ½ M arrests, 1997—2002

Evidence for racial profiling?

Only look at discretionary charges:

Simple marijuana possession
Non-moving auto infractions

HOW TO WRITE THIS STORY?

Simple graphic

THE SATURDAY STAR

The photo that never was

GARTH WOOLSEY, C3

Periods of rain; windy. High 14 C

October 19, 2002

Also inside . . .

- Waterfront: Details of what could be, B1, B4-5
- Hydro woes: Insulating against price spikes, E1
- Wheels: The Bug goes roofless, G1
- Paul Martin: The man who would be king, H1
- Carol Shields: Her last novel? Unless . . . J1

thestar.com ONTARIO EDITION

AN INVESTIGATION INTO RACE AND CRIME

Lede case

SUING POLICE: Jason Burke, falsely accused of dealing drugs during Caribana two years ago, says he was a victim of racial profiling.

JIM RANKIN/TORONTO STAR

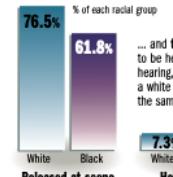
Headline,
summary

Singled out

Star analysis of police crime data shows justice is different for blacks and whites

Telling numbers

Police records show that a black person in Toronto arrested on a single drug possession charge was less likely to be released at the scene...



... and twice as likely to be held for a bail hearing, compared to a white person on the same charge.



Source: Toronto police annual records, 1997-2001, in latest available data.

Blacks arrested by Toronto police are treated more harshly than whites, a Toronto Star analysis of crime data shows.

Black people, charged with simple drug possession, are taken to police stations more often than whites facing the same charge.

Once at the station, accused blacks are held overnight, for a bail hearing, at twice the rate of white people.

Managing Editor's notebook, A2

leaders and criminologists, suggests police use racial profiling in deciding whom to pull over.

The evidence is contained in a massive police database recording more than 400,000 stops, in which an individual was arrested, or ticketed, for an offence dating back to 1996. It included almost 800,000 criminal and other charges.

The Toronto crime data also shows a disproportionate number of black motorists are ticketed for violations that only surface following a traffic stop. This difference, say civil libertarians, community

community. Police are forbidden, by their governing board, from analyzing this data in terms of race, but The Star has no such restriction. The findings provide hard evidence of what blacks have long suspected — race matters in Canadian society, especially when dealing with police.

Chief Julian Fantino disputed the findings, saying the colour of a person's skin has nothing to do with how they're treated by his officers.

"We don't treat people differently."

Please see MacAulay, A8

Chrétien expected to keep cabinet minister

Ethics report has 'wiggle room' to save MacAulay

BY TIM HARPER
AND LES WHITTINGTON
OTTAWA BUREAU

OTTAWA — Jean Chrétien receives a report from his ethics counsellor today that is expected to give him enough "wiggle room" to keep his solicitor-general, Lawrence MacAulay, in the federal cabinet.

Ethics counsellor Howard Wilson completed his report and delivered it to the Prime Minister's Office last night, where it was received by Chrétien's chief of staff, Percy Downe.

It was then to be relayed to Chrétien by secure fax to Beirut, where the Prime Minister is attending a summit of French-speaking nations. It was 1:30 a.m. in Beirut when the fax arrived so Chrétien would likely be reading it this morning.

Senior sources said last night that unless there is a surprise in Wilson's report, the Prince Edward Island minister will remain, Chrétien will return to Ottawa and weather the inevitable storm of opposition and media protest and forge ahead with an ethics package by mid-week.

Wilson has been investigating whether MacAulay broke ethics guidelines for cabinet ministers in the awarding of a contract and extension worth \$100,000 to Everett Roche, a Charlottetown political friend of the solicitor-general's.

Chrétien will not fire MacAulay unless he is given incontrovertible evidence of wrongdoing for two key reasons, sourc-

Please see MacAulay, A8

INSIDE

Barclay L2
Births B7
James Travers H2

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

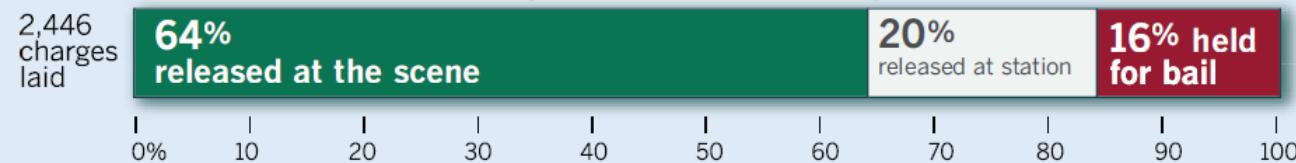
Statistical analysis of single drug possession charges shows that blacks are much less likely to be released at the scene and much more likely to be held in custody for a bail hearing. Darker colours represent a stronger statistical link between skin colour and police treatment.

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

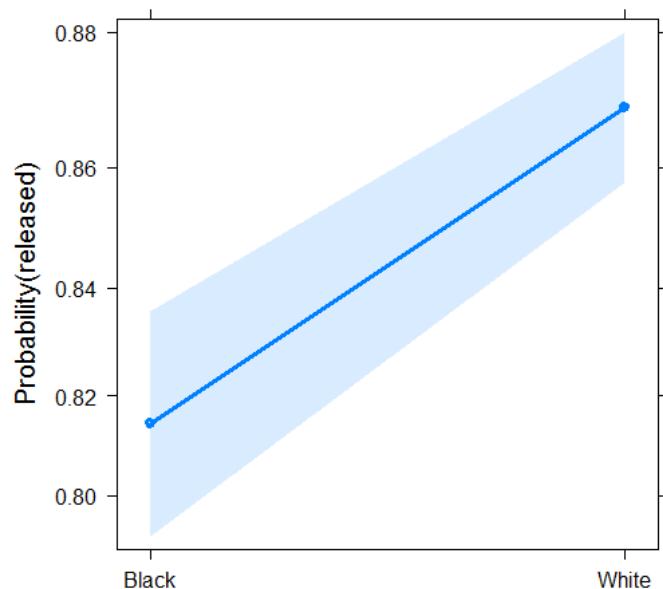
Marijuana data: Analysis

Fit a `glm()` model to account for **all predictors**, plus **interactions** with skin color

```
> data(Arrests, package="carData")
> Arrests$year <- as.factor(Arrests$year)
> arrests.mod <- glm(released ~ employed + citizen + checks +
  colour*year + colour*age,
  family=binomial, data=Arrests)
```

BIG Q: How large is the effect of skin color, controlling for other factors?

```
> plot(Effect("colour", arrests.mod), lwd=3, ci.style="bands", ...)
```

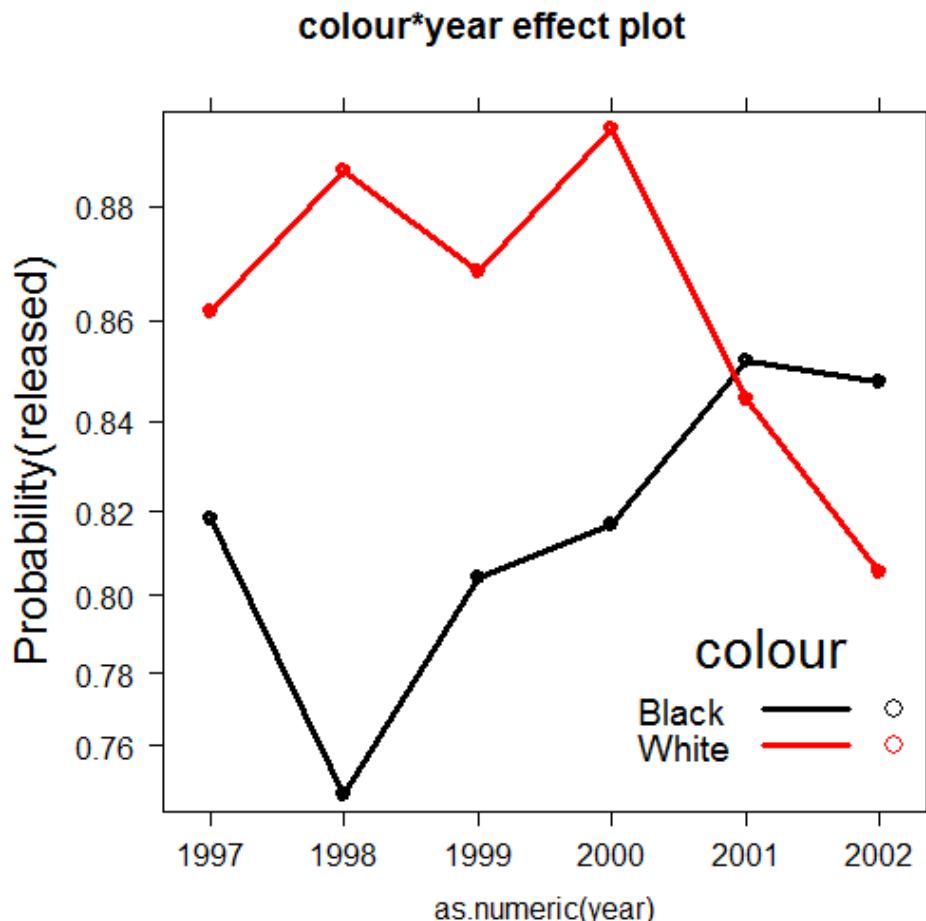


- Effect plot shows average effect **controlling** (adjusting) for **all** other factors simultaneously
- (*Star* analysis controlled for these one at a time.)
→ Strong evidence for different treatment of blacks & whites
- But effect **size** smaller than reported by the *Star*

Effect plots: Interactions

The story turned out to be more nuanced than reported by the *Toronto Star*

```
> plot(Effect(c("colour", "year"), arrests.mod), multiline=TRUE, ...)
```



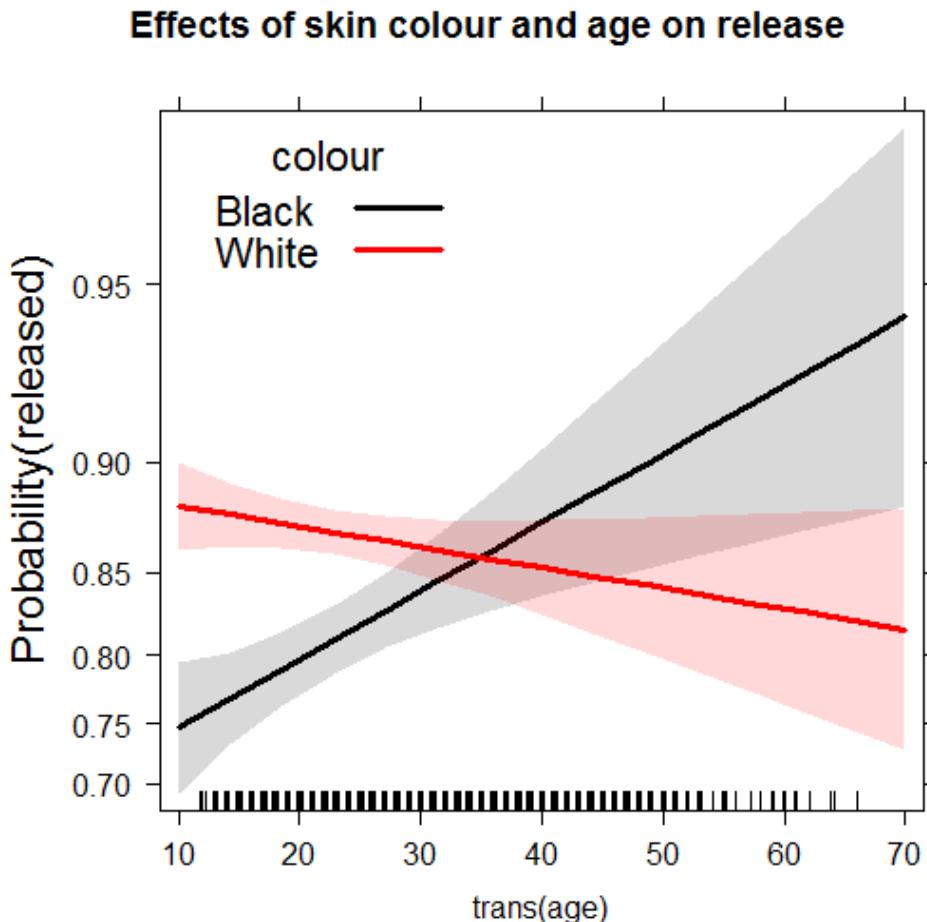
Up to 2000, strong evidence for differential treatment of blacks & whites

Also: evidence to support Police claim of effect of training to reduce racial effects in treatment

Effect plots: Interactions

A more surprising finding ...

```
> plot(Effect(c("colour", "year"), arrests.mod), multiline=TRUE, ...)
```



Opposite age effects for blacks & whites:

- Young blacks treated **more** harshly than young whites
- Older blacks treated **less** harshly than older whites

Traffic stops

Analysis of traffic tickets issued for **discretionary** (non-moving) violations

Argument: Proportion of tickets to Blacks >> proportion in the population



Police target black drivers

Star analysis of traffic data suggests racial profiling

Black drivers confronted by the flashing lights of a police cruiser often worry if they're being pulled over for the colour of their skin.

Officers deny that happens.

■ Reaction, A9 ■ The Star's view, A12

Now, for the first time, empirical evidence suggests police have indeed been targeting black drivers in Toronto.

Police traffic offence data, obtained and analyzed by The Star, shows a disproportionate number of blacks ticketed for viola-

Race and Crime

STAR INVESTIGATIVE TEAM: JIM RANKIN, JENNIFER QUINN, MICHELLE SHEPHERD, JOHN DUNCANSON, SCOTT SIMMIE

tions that routinely surface only after a stop has been made.

These "out-of-sight" traffic offences include failing to update a driver's licence with a change of address, driving without a licence, driving without insurance, or driving while under suspension. Police usually discover such violations only after a motorist has been pulled over. And, in the absence of any other charge, it isn't clear why drivers involved in these offences were stopped in

the first place.

It's assumed random checks would generate a pattern of charges that mimics the racial distribution of drivers in society as a whole. So the rate at which minority drivers are charged is often used, in the U.S., as a bellwether for racial profiling.

raw race-based crime data, arguing racists might use the resulting statistics to stigmatize ethnic communities.

Police follow that rule, and don't record race statistics for the purpose of ethnic analysis. They do, however, list skin colour in most arrest reports when describing a person charged. It's routine in arrests for major

Please see Police, A8

Star gets action

Among the most satisfying outcomes for data journalists is when their stories lead to concrete **attention & action**

Front page editorials are reserved for big events

THE SATURDAY STAR

Morning min. High 18 C

October 26, 2002

thestar.com

ONTARIO EDITION

Race and Crime

UNIDENTIFIED: DR. JAMES DUBIN, MICHELLE SHEPPARD, JOHN GUNNARD, SCOTT RYAN

Black crime rates highest

'No one was born violent ...
What's causing these problems?'

There had been another one. Soon, a police officer would face the thankless task of breaking the news to yet another unarmed family. The officer would have to tell Gmar Sheron Clegg, the most often shot at 25-year-old who was shot as he stood outside an Elstoboker nightclub.

His killing, like many others in the black community, sparked a tragic chain reaction. Paul Watson, one of Clegg's friends, who knew him as 'Pete', took off in a fury to avenge his death. But instead of finding his killer, he found Julian Fantino, 32, and his 23-year-old friend Michael Lewis. The two were on the doorstep of a Toronto housing complex.

In the space of 24 hours, three black men lay dead; their names added to a list many have never heard.

"Look at the news, all the shootings that have happened — look at the news," deplores Dev Polson of the Black Inmates and Friends Assembly. "You're seeing black faces, black men. I don't want to turn the TV on."

A Star investigation, conducted by analyzing police arrest records, showed that in certain cases where police have discretion to use personal judgment, black people receive treatment than whites. Last weekend's stories prompted a flood of letters and phone calls from black readers who say they have been subjected to ri-

Justice Dubin to study treatment of black suspects

By MICHELLE SHEPPARD
AND JENNIFER QUINN

One of Canada's most respected judges will investigate race relations practices in the Toronto Police Service, following a *Toronto Star* investigation that showed blacks are treated more harshly than whites.

- Justice to Dubin probe, \$3
- Details on stats, \$4
- Chief's own words, \$5
- *The Star's* Ombud, \$6

Police Chief Julian Fantino announced yesterday that he has asked the Honourable Charles Dubin, retired chief justice of Ontario, to conduct an independent review along with community leaders and make recommendations that could "improve our performance in this area."

Fantino spoke publicly about the Star's investigation for the first time yesterday, having just returned from an out-of-town conference.

"I have never denied that there are instances of inappropriate conduct by a small number of our officers and personnel, some that have been viewed as discriminatory and racist," he said. "We do not, however, have a racism rampant police service."

Word of the findings in the Star's article before the stories were published, Fantino said, fully denied any racial bias.

"We don't treat people differently," he said at that time. "Nor do we consider the race or ethnicity, or any of that, as a factor in how we dispose of cases or individuals."

Fantino's review was welcomed by board chair Norm Gardner, who said he hoped that there were "no racist

Chief Julian Fantino deserves congratulations for announcing steps to address the dramatic findings of a *Toronto Star* investigation that indicates blacks are being unfairly targeted by Toronto police.

Although we would prefer the chief take such action sooner, we are delighted he now wants an in-depth — and open — study into possible racial profiling by his officers.

His move clearly show he is asserting his leadership on one of the most sensitive issues to affect this city. Its police and its



FORCE UNDER FIRE: Police Chief Julian Fantino discusses the issue of police-minority relations yesterday.

Fantino's bold move on right path

Editorial

look at the issue.
All good steps.

And all necessary steps, backed by promises that he won't tolerate any form of racism and that he will act forcefully on any recommendations that Dubin makes.

Fantino also plans to meet with his senior officers on the issue. He's inviting members of the black community to come in and talk directly with him.

Since he became chief, Fantino has worked hard to reach out to minorities in the city — blacks, gays, Asians and others. But racial bias and prejudice within the

But not without pushback

Goal? “Libel-chill to cause media to think twice...”

Judge: TOSED!
“You can’t libel an entire class”

“many factors other than skin color affect how one is treated, but the effect of skin colour was **persistent** and cannot be dismissed”

News

7200 “members” = \$375K / member
Cops: 5120 + parking enf + cadets + ...

Police union sues Star over race-crime series

7,200-member group seeks \$2.7B damages in class action lawsuit

Star stands by stories as fair, balanced and accurate

PETER SMALL
STAFF REPORTER

The Toronto Police Association has launched a \$2.7 billion class action libel suit against the Star for its series spotlighting the force's treatment of blacks.

“Accusing the members of the Toronto Police Service of racism is a very serious allegation and if such a serious allegation is going to be made, the accusers will be called upon to prove it,” Tim Danson, the association’s lawyer, told a news conference yesterday.

The Toronto Star analysis of police arrest data showed that with charges for simple drug possession, nearly twice as many blacks as whites were taken into custody and held in jail awaiting a bail hearing.

The data also showed a disproportionate number of black motorists were ticketed for offences that routinely would come to light following a traffic stop. Civil libertarians and criminologists say this pattern points to racial profiling, whether conscious or not.

Shortly before Danson spoke at the association’s headquarters, flanked by president Craig Bromell and other union officials, the Star was served with



Police union lawyer Tim Danson says the Star's racial profiling series maligns all city officers.

consulting service, found the paper's findings to be sound.

In an in-depth analysis of the data, presented to the Toronto police services board on Dec. 10, Friendly, author of *The SAS System for Statistical Graphics and Visualizing Categorical Data*, and associate editor of the *Journal of Computational and Graphical Statistics*, said many factors other than skin colour affect how one is treated, but that the effect of skin colour was “persistent, and cannot be dismissed” in the arrest data.

Danson said he hoped the Star will ac-

He said he thought that the case is appropriate for a jury to decide, particularly since it's the community's view that is important. “The police really want to put this in the hands of the public in the form of a jury.”

Danson said regardless of whether the association loses the case against the Star, “there are some things worse than losing and that’s not even trying.”

When asked whether the lawsuit, regardless of its success or failure, may create a libel chill that would cause media to think twice about criticizing institutions whose employees have powerful unions, Danson called that concern “somewhat speculative.”

He added that the Star has gone well beyond fair comment in its series. “This was a direct and frontal attack to the quintessential core of what it is to be a police officer.”

The charges examined by the Star are part of a much-larger Toronto police database that details all arrests made since late 1996. The paper first sought access to the data in a March, 2000, freedom of information request.

Following appeals and mediation, a copy of the database, excluding personal information, was released to the Star in May, 2002. It records more than 480,000 incidents in which an individual was arrested or ticketed, and almost 800,000 criminal and other charges. It served as the basis of the investigation into policing, crime and race.

Using relational database software, Star reporter Jim Rankin spent most of the summer preparing and analyzing the data. The Star study examined

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 @Jleerakin . 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 ...

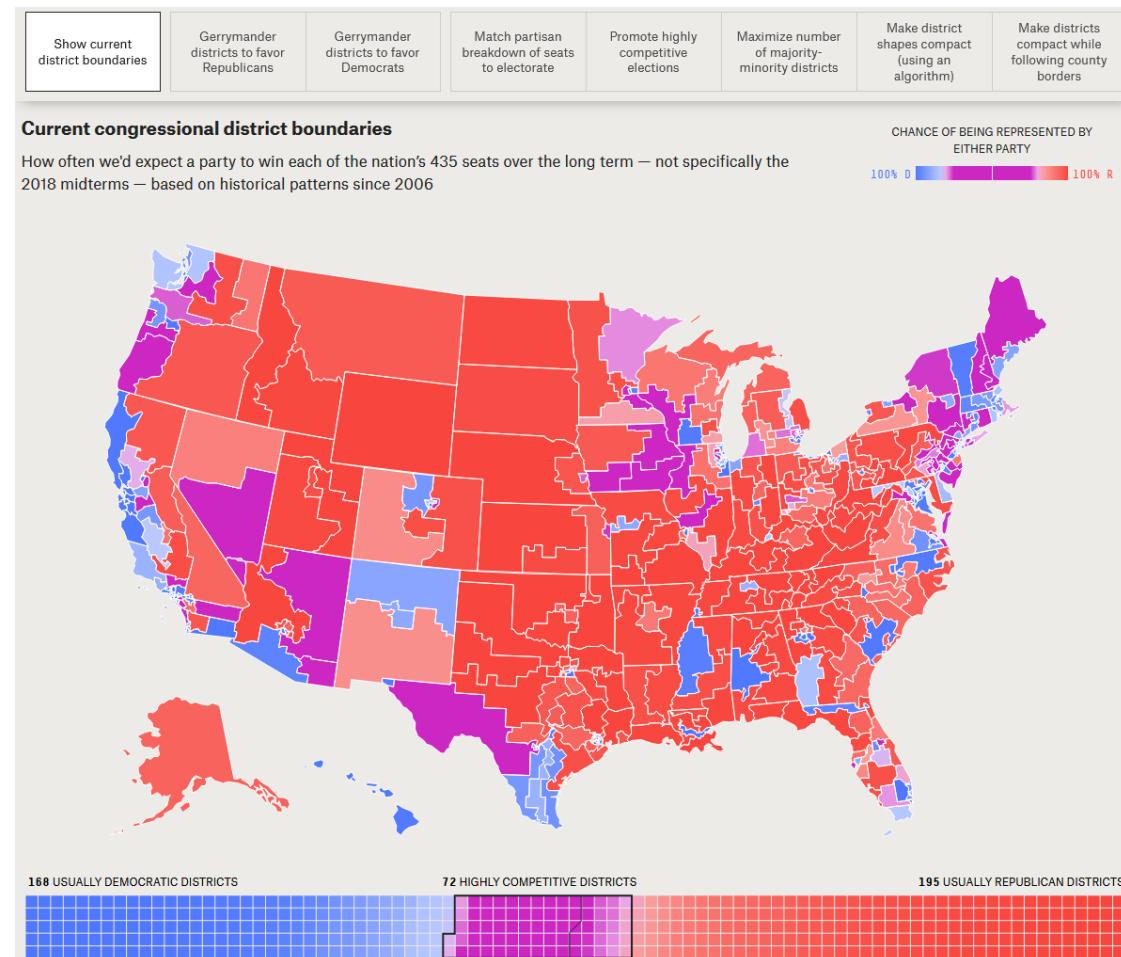
- ❖ Digital publishing and online, interactive media
 - Many news organizations have interactive graphic developers or entire online departments
 - A revenue stream responding to less from print & ads
- ❖ Story telling frames a wider enterprise
 - A new focus and techniques for data visualization
 - 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:

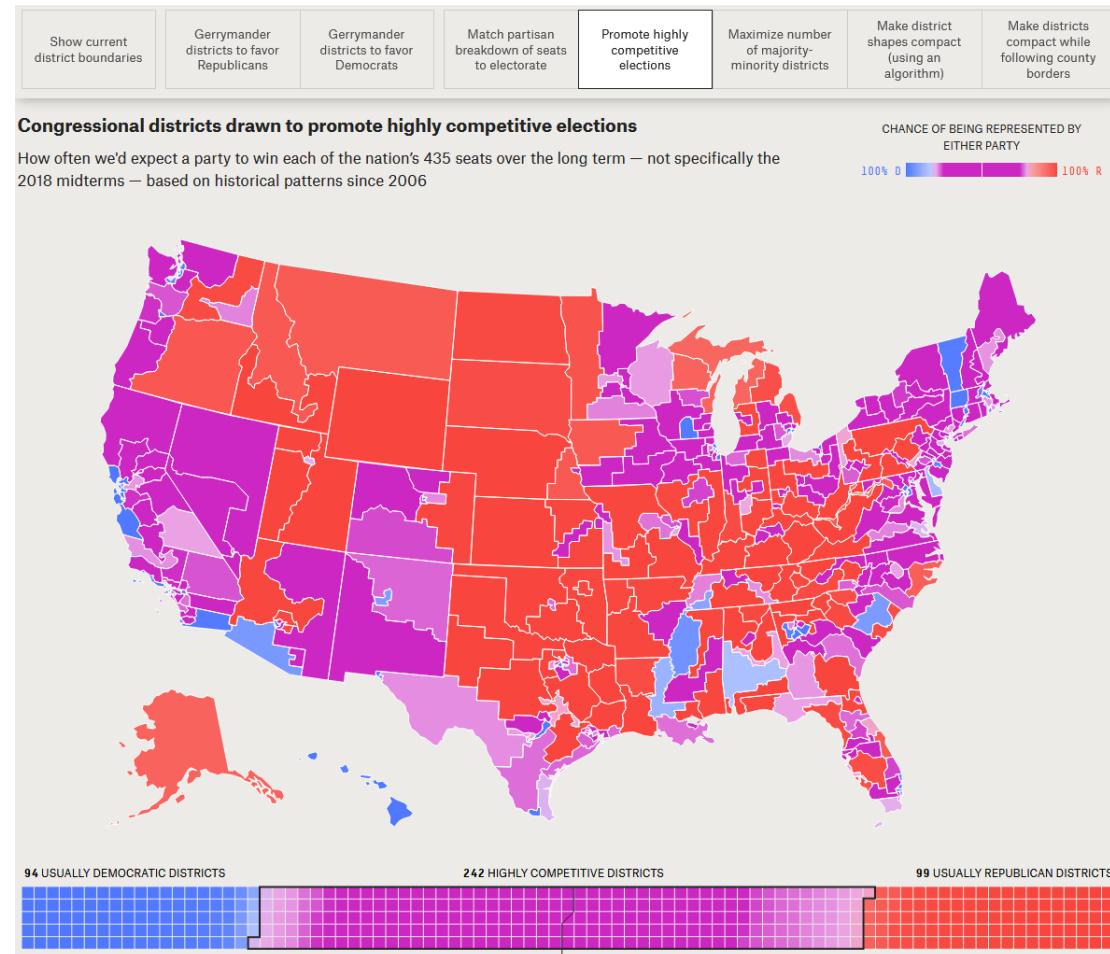
<https://projects.fivethirtyeight.com/redistricting-maps/>

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:

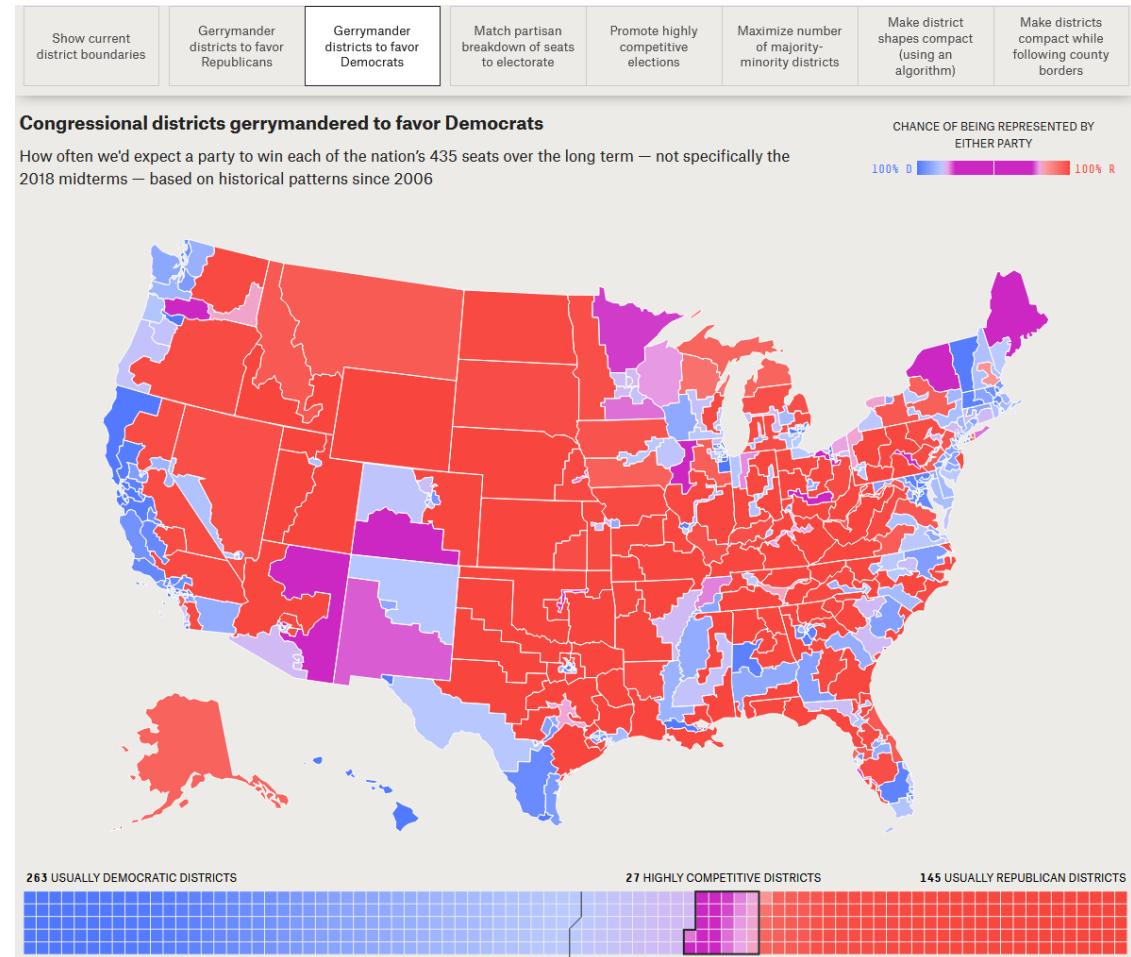
<https://projects.fivethirtyeight.com/redistricting-maps/>

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.

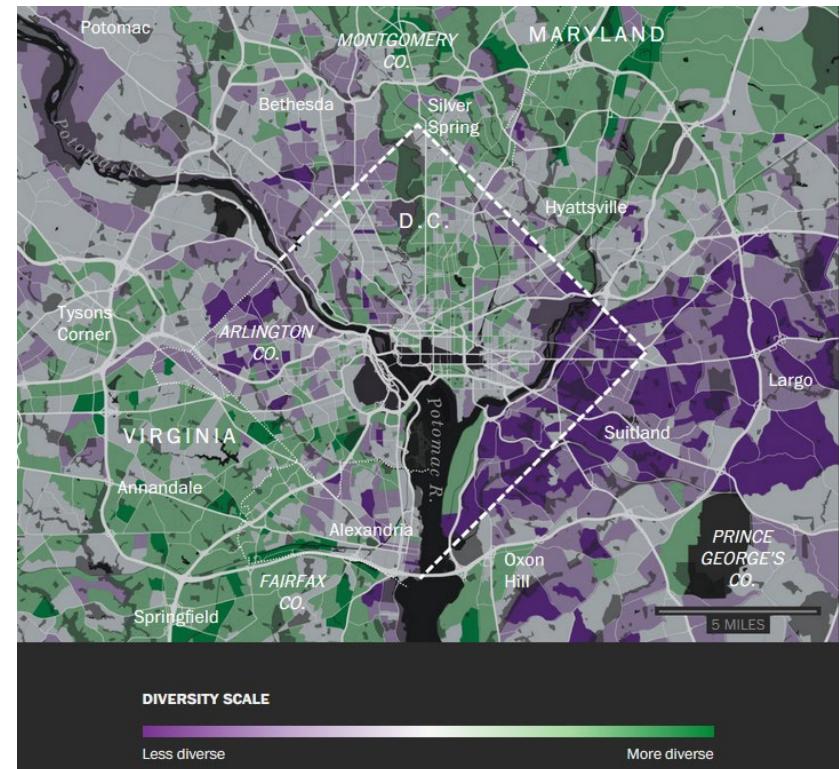
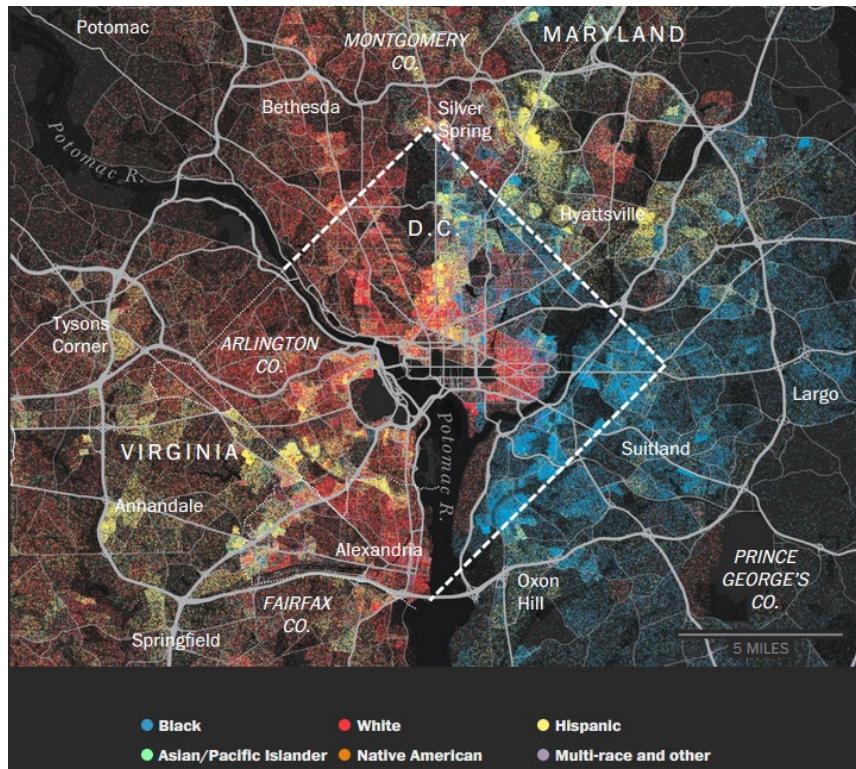
Changes in representation take place largely on the east/west coasts



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

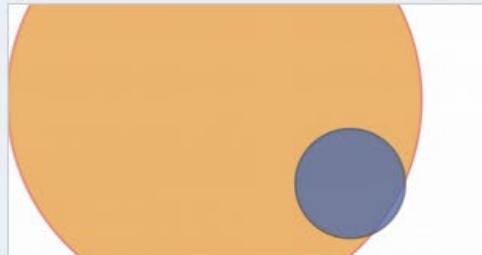
An impressive range of social & environmental issues presented by *Our World in Data*



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

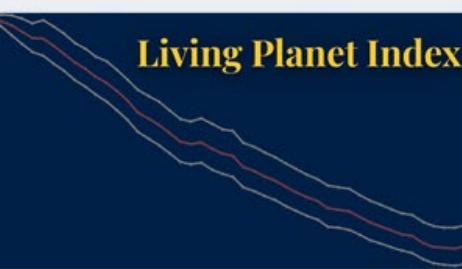
To transition towards low-carbon energy systems we need low-cost energy storage. Battery costs have been falling quickly.

By Hannah Ritchie — June 04, 2021



The argument for a carbon price

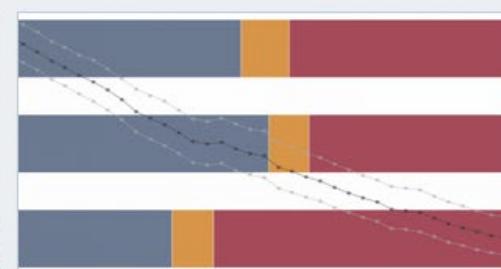
We are paying a price for fossil fuels, but that price is not paid by those that burn the fossil fuels – we need to change that.



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

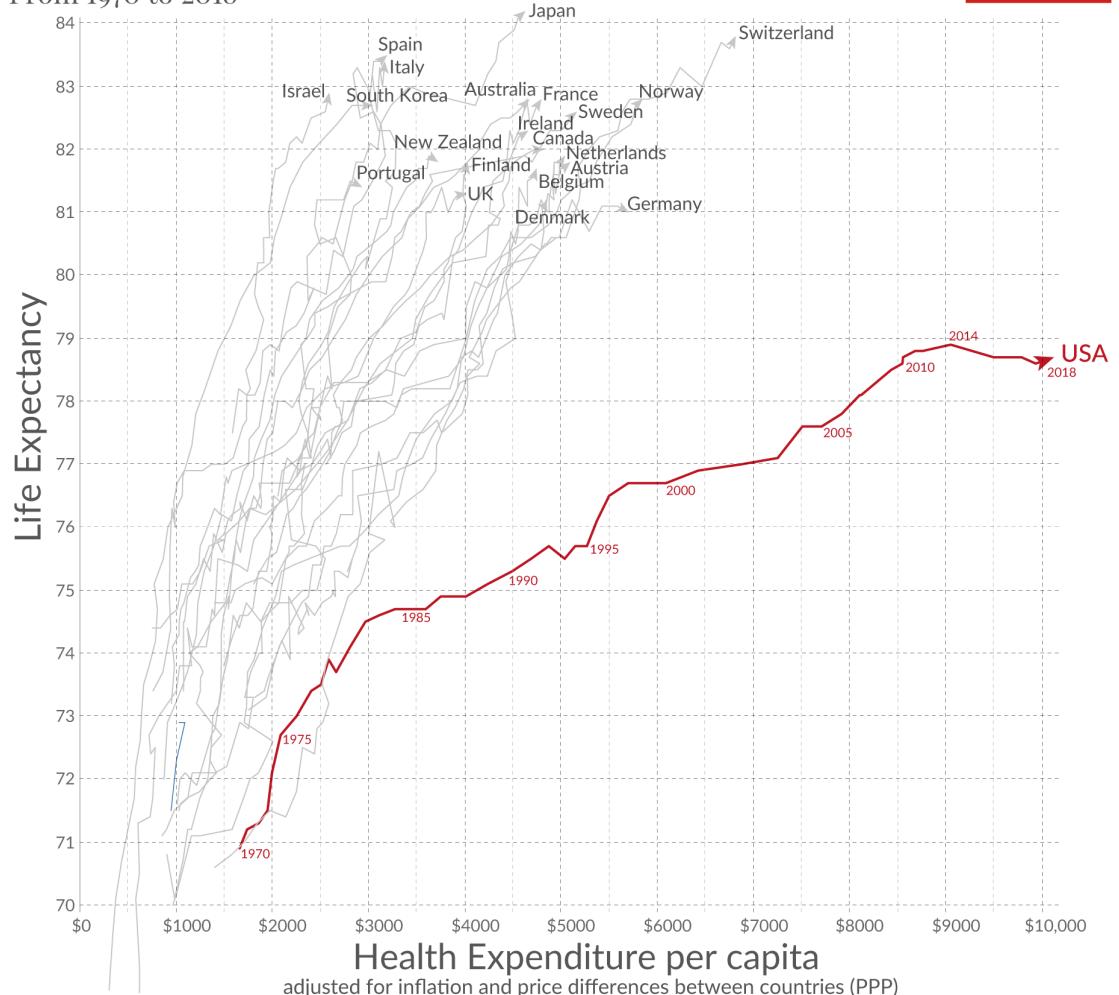
What is the main message of this chart?

Why is it effective?

Would using log expenditure be better or worse?

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



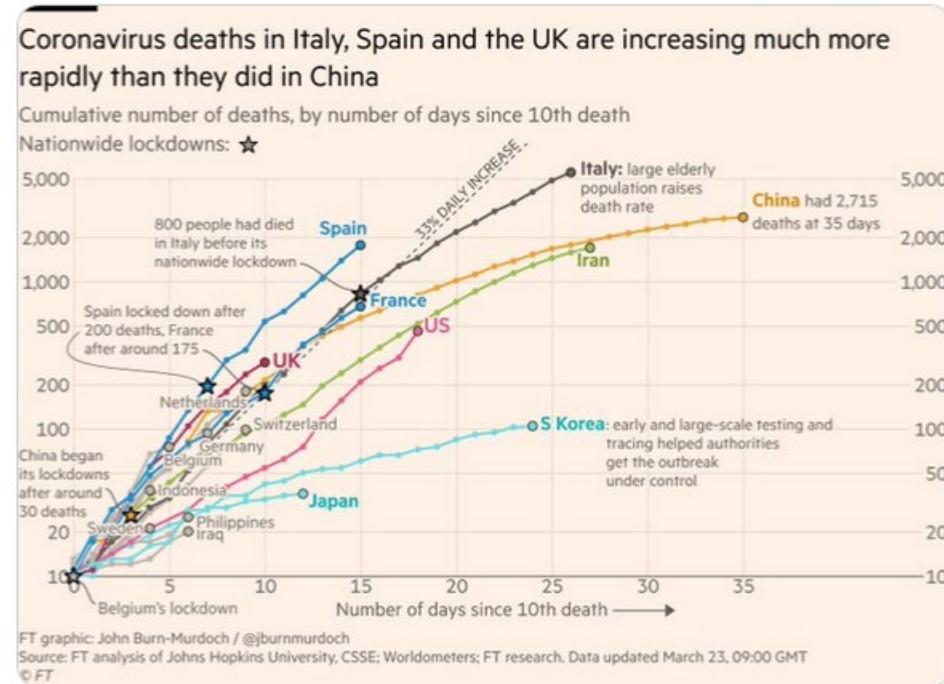
John Burn-Murdoch 
@jburnmurdoch

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 ft.com/coronavirus-la...

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

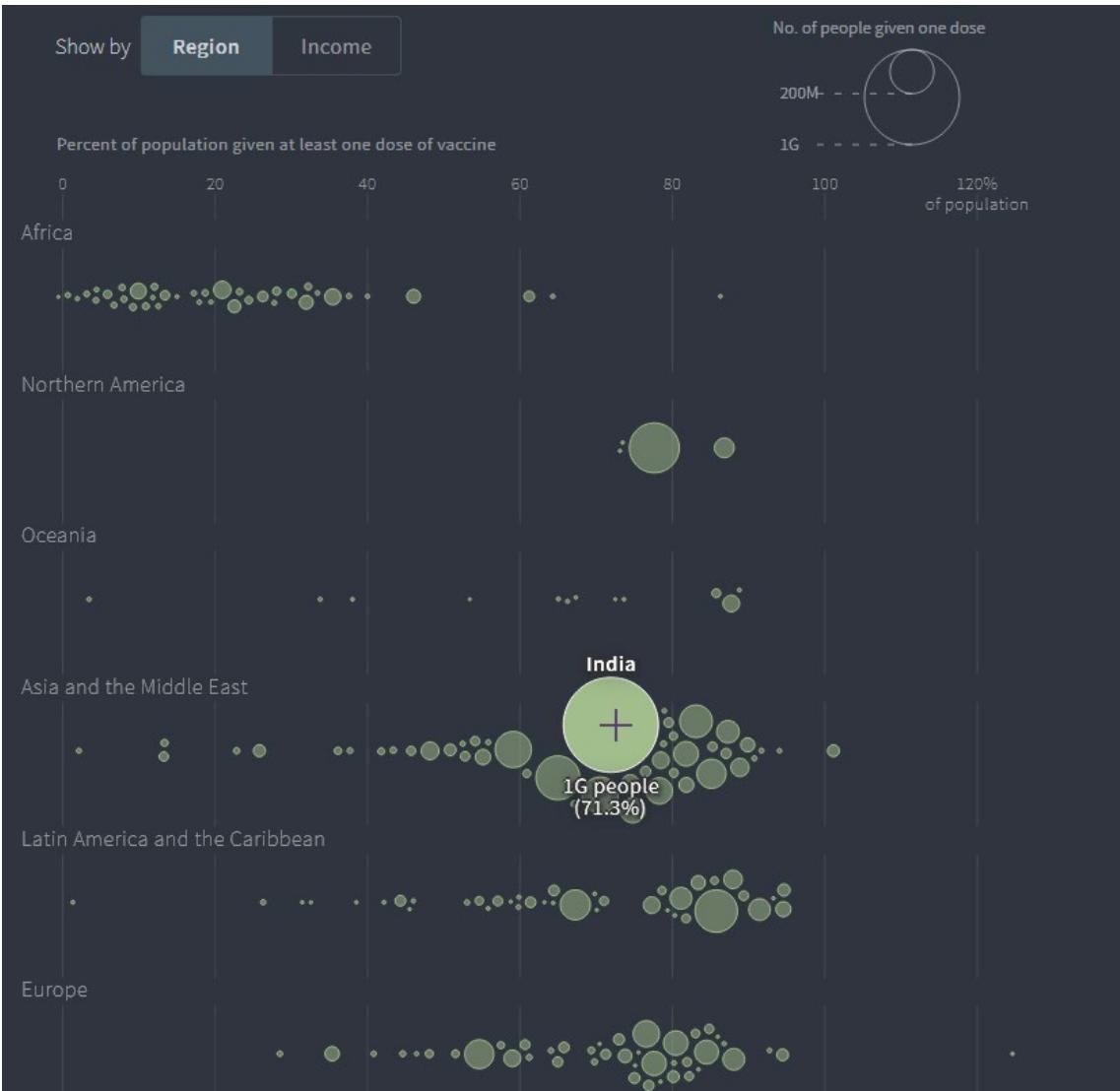


Reuters COVID tracker

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.

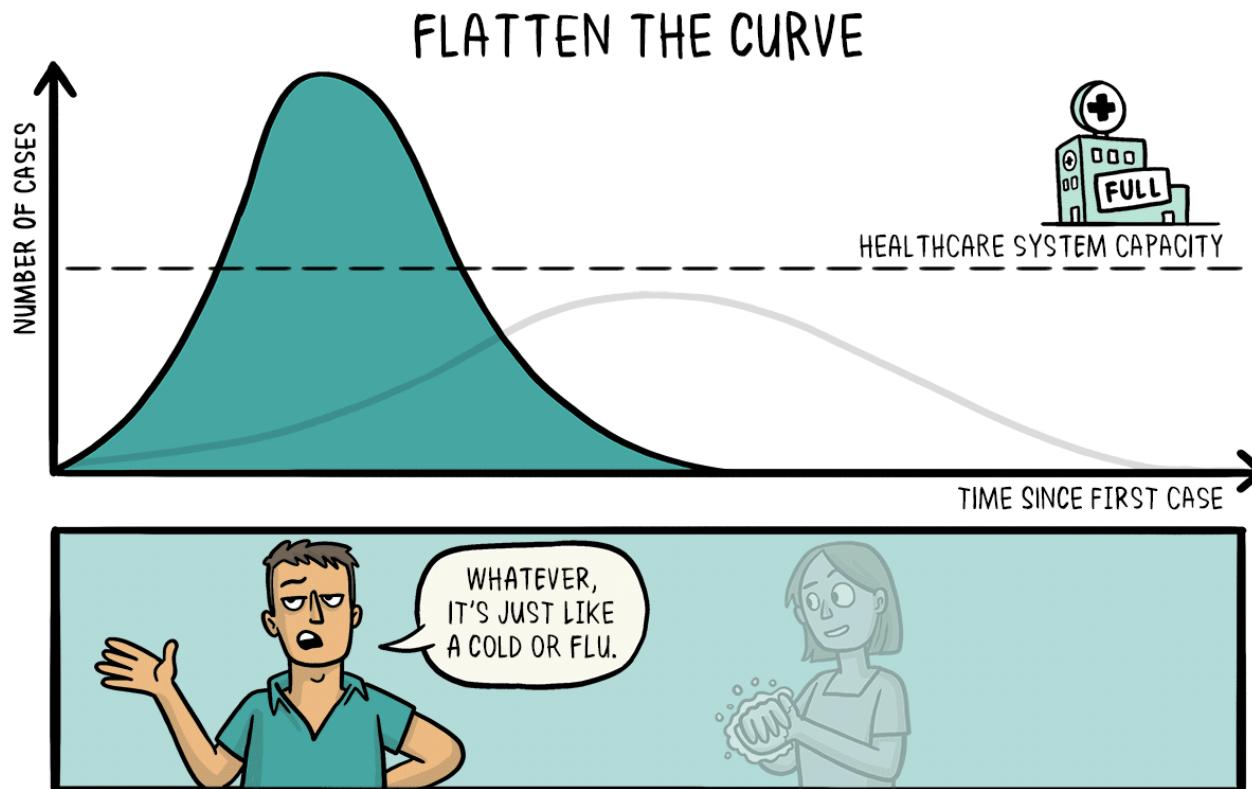
That meant creating a unifying structure to catalogue the different phases of each country's vaccination rollout plan.”

-- 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

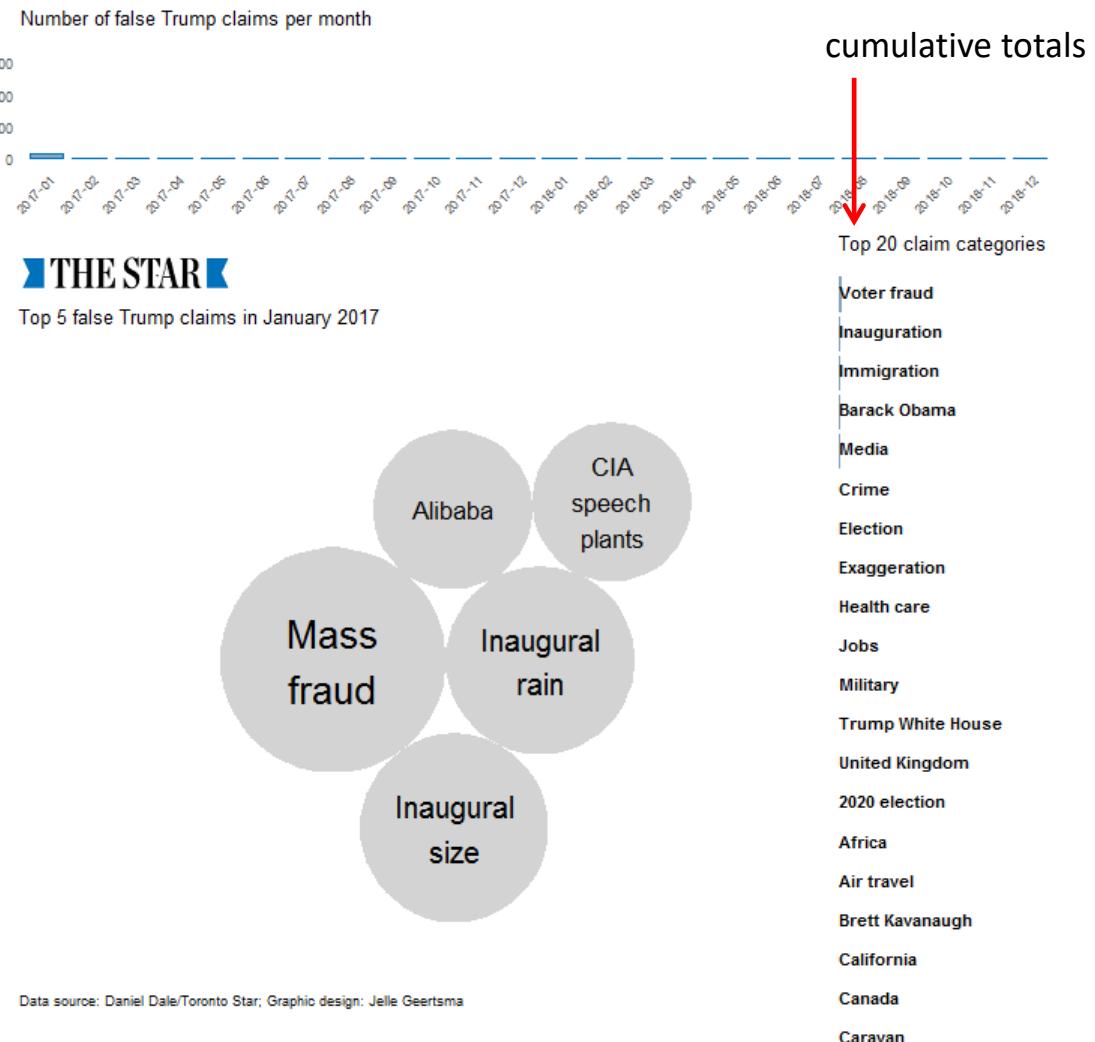
Graphic artist Jelle Geertsma designed this animated graphic showing the top 5 claim “slugs” & the top 20 categories over time, by month.

Bars at right show cumulative totals

This animation uses the ganimate package.

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

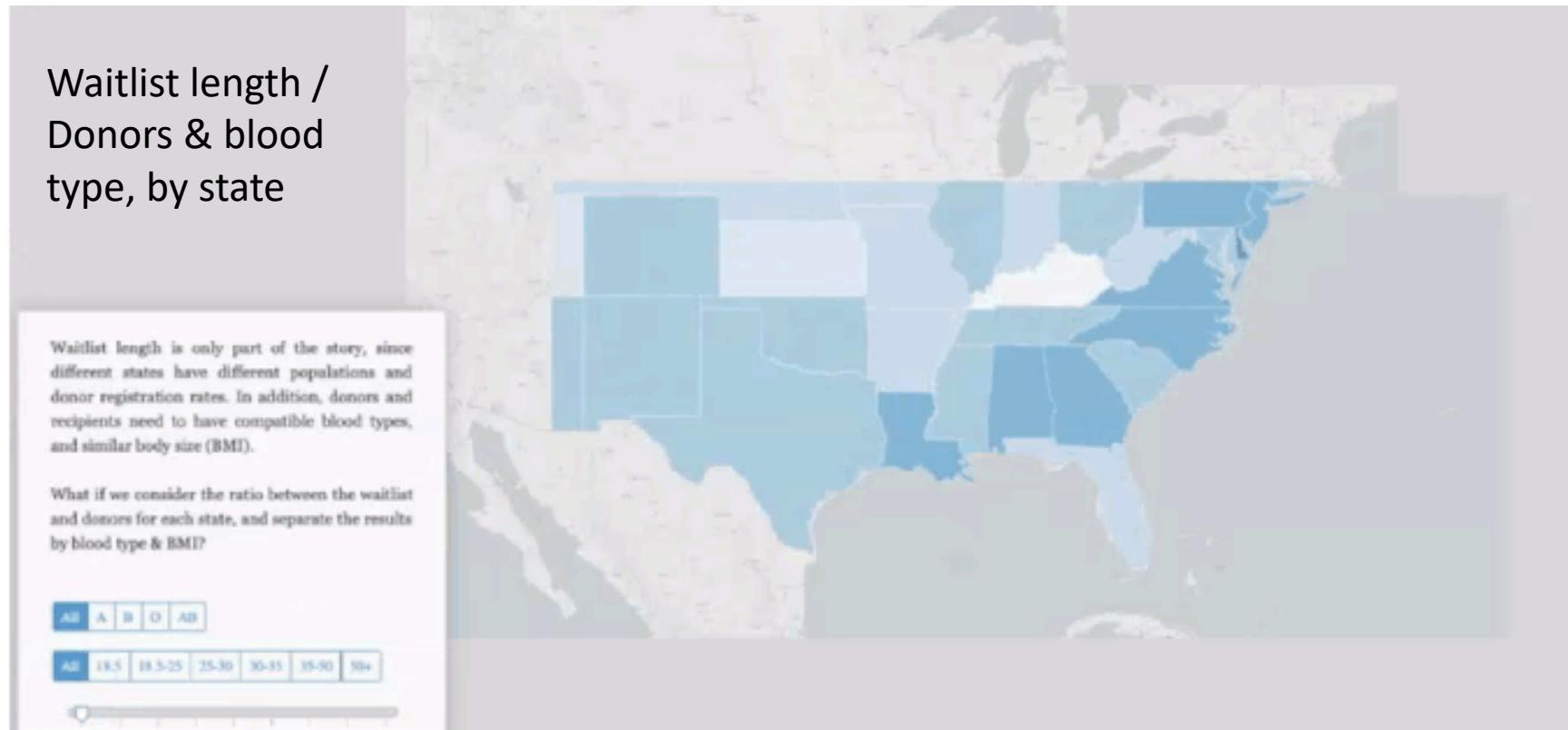
<https://github.com/friendly/TrumpLies>



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

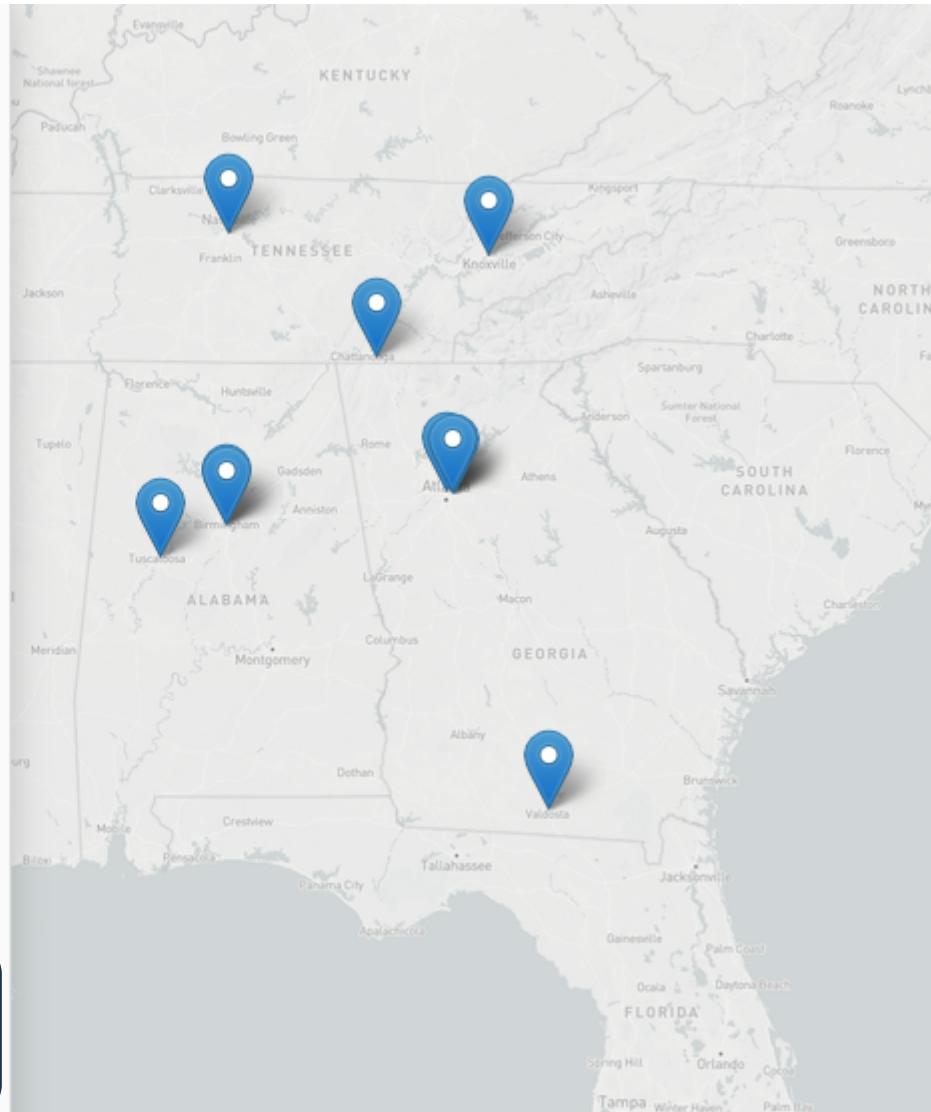
Since Kidney Transplant can be time sensitive, it is best for you to choose the ones adjacent to where you live. Please enter your zip code to view centers around you:

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



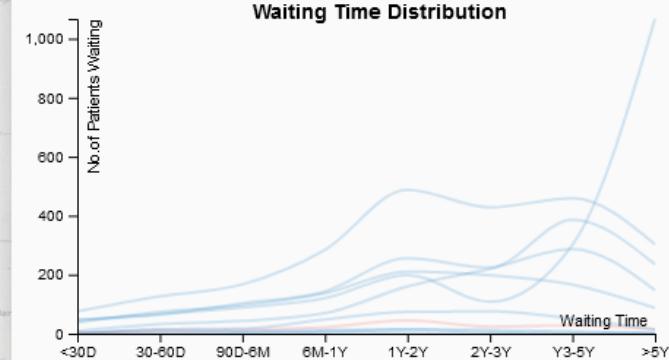
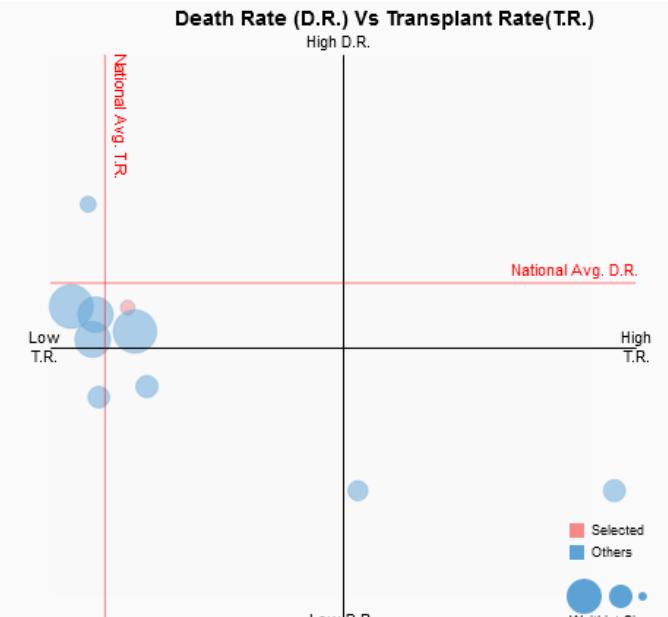
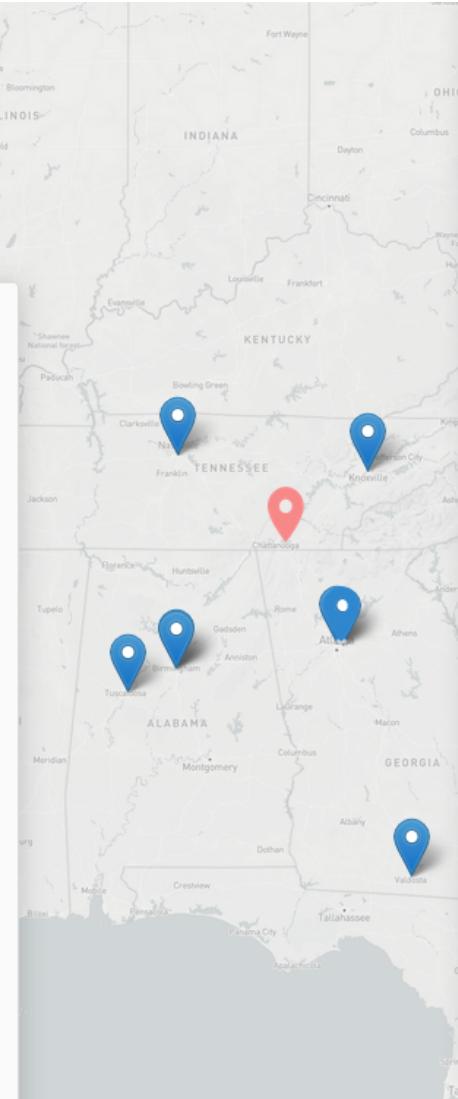
Since Kidney Transplant can be time sensitive, it is best for you to choose the ones adjacent to where you live. Please enter your zip code to view centers around you:

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.



Drill-down to tell a story

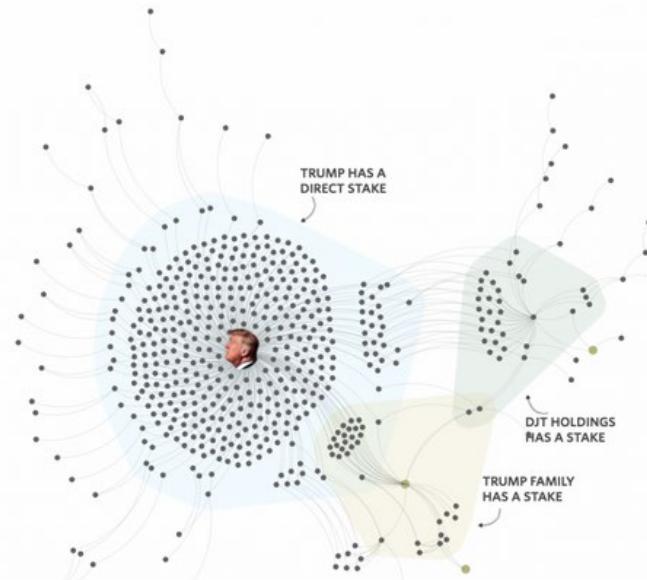
The *Wall Street Journal* published this interactive story of potential conflicts of interest in the business holdings of Donald Trump and his family (2017)

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

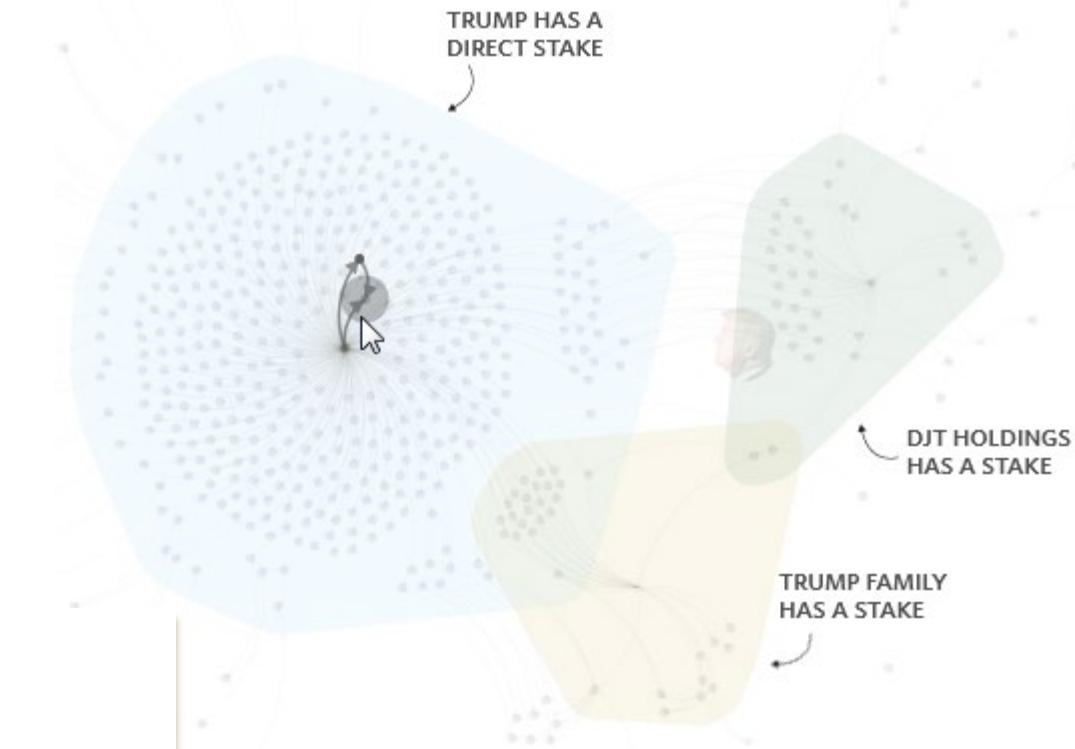


President Trump's network of conflicts of interest. Credit: WSJ.

"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

11th 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](#).

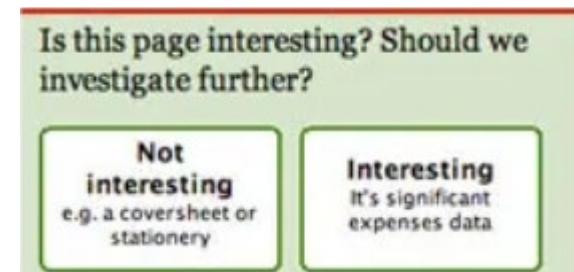
We have **458,832** pages of documents. **32,755** of you have reviewed **225,443** of them. Only **233,389** to go...

[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.

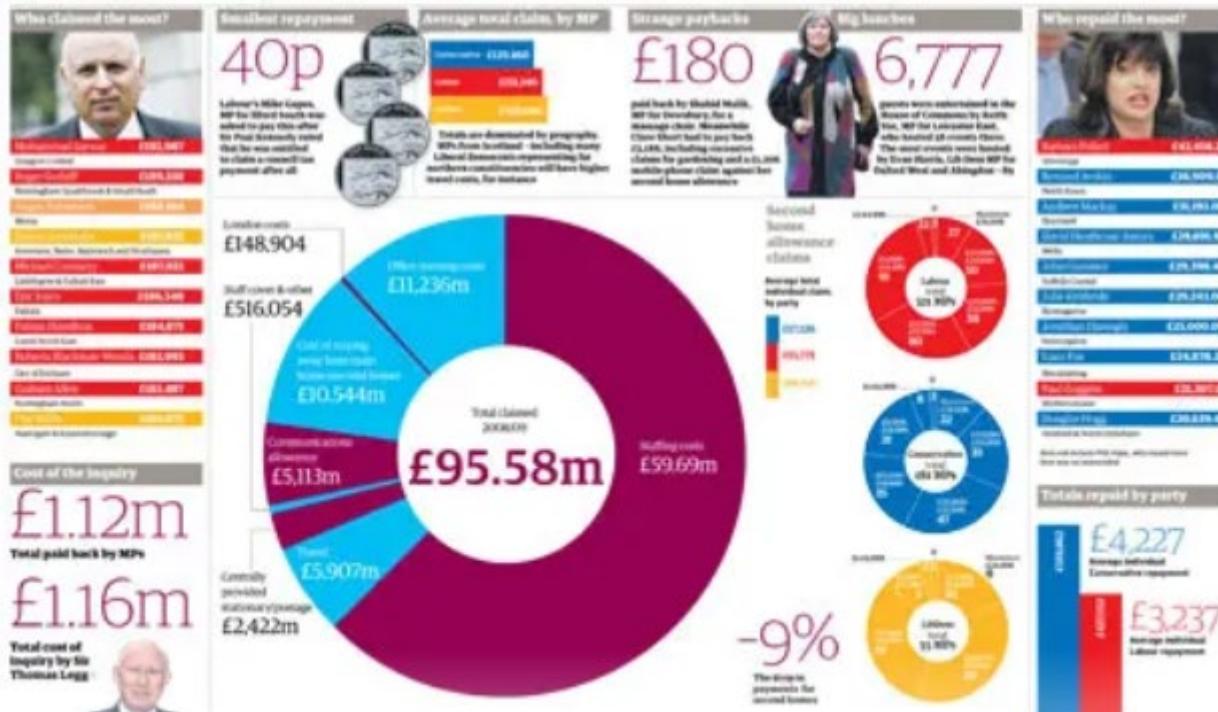


MPs' expenses: how we visualised the data

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.

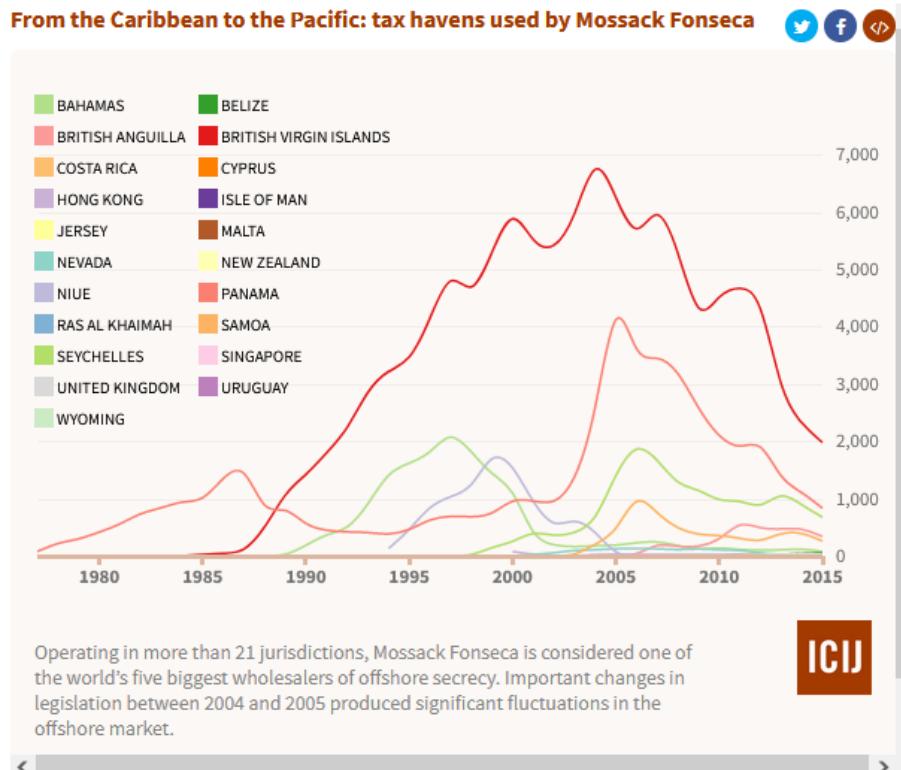
Two terabytes of data, over 40 years, 210,000 companies in 21 offshore jurisdictions
IJIC: 280 top investigative journalists, >100 media partners, >100 countries



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



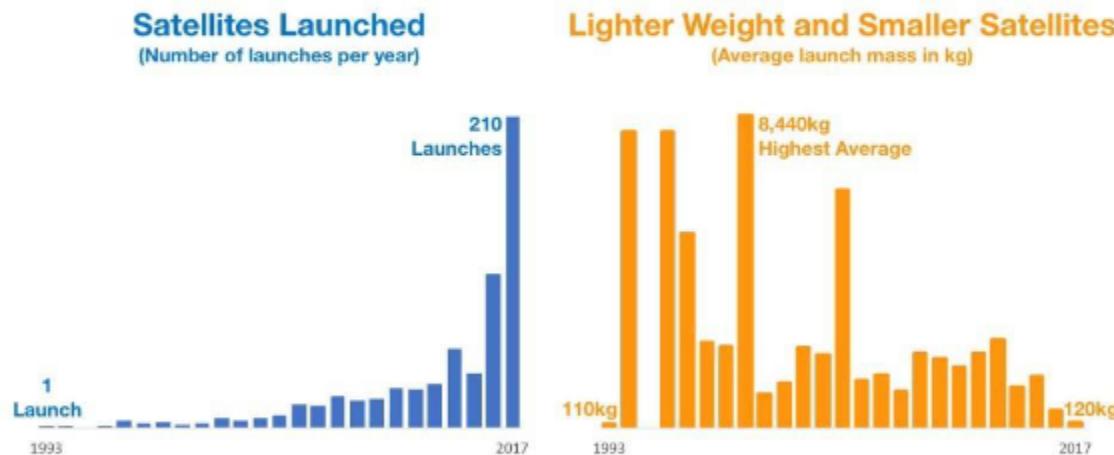
How could this graph be improved?

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 → journalists gaining 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 Miglarese Radiant.Earth)

From: Mark Corcoran (2018),
[Satellite Journalism – The Big Picture](#)

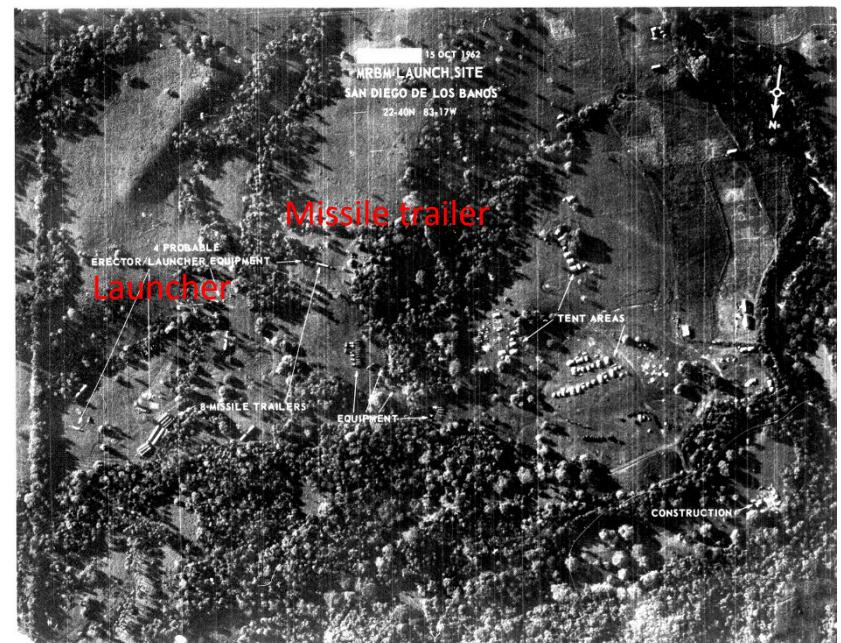
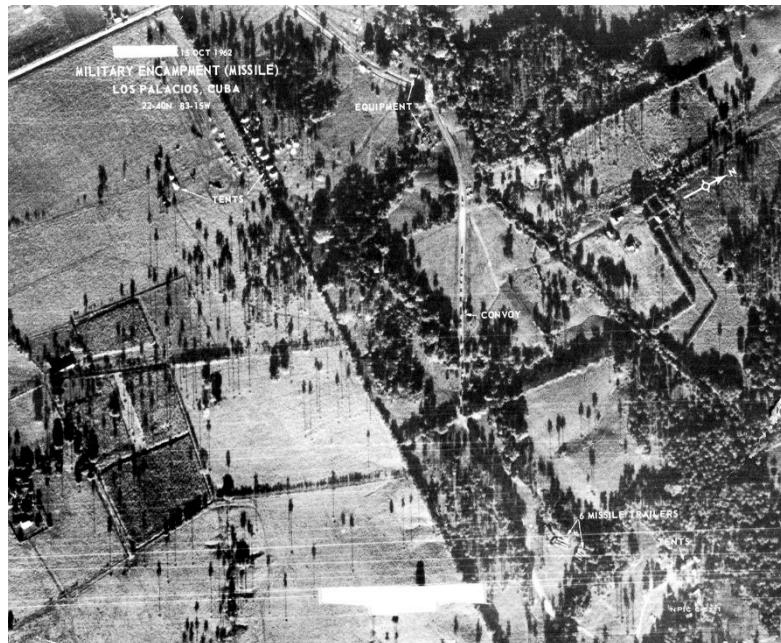
Satellite Journalism: Smoking Guns

- 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

Cuban Missile Crisis: Before vs. After

Not DDJ, but: U-2 photographs on Oct 14-15, 1962 were used to identify the build-up of Soviet medium-range ballistic missiles in sites in Cuba, providing VISUAL PROOF



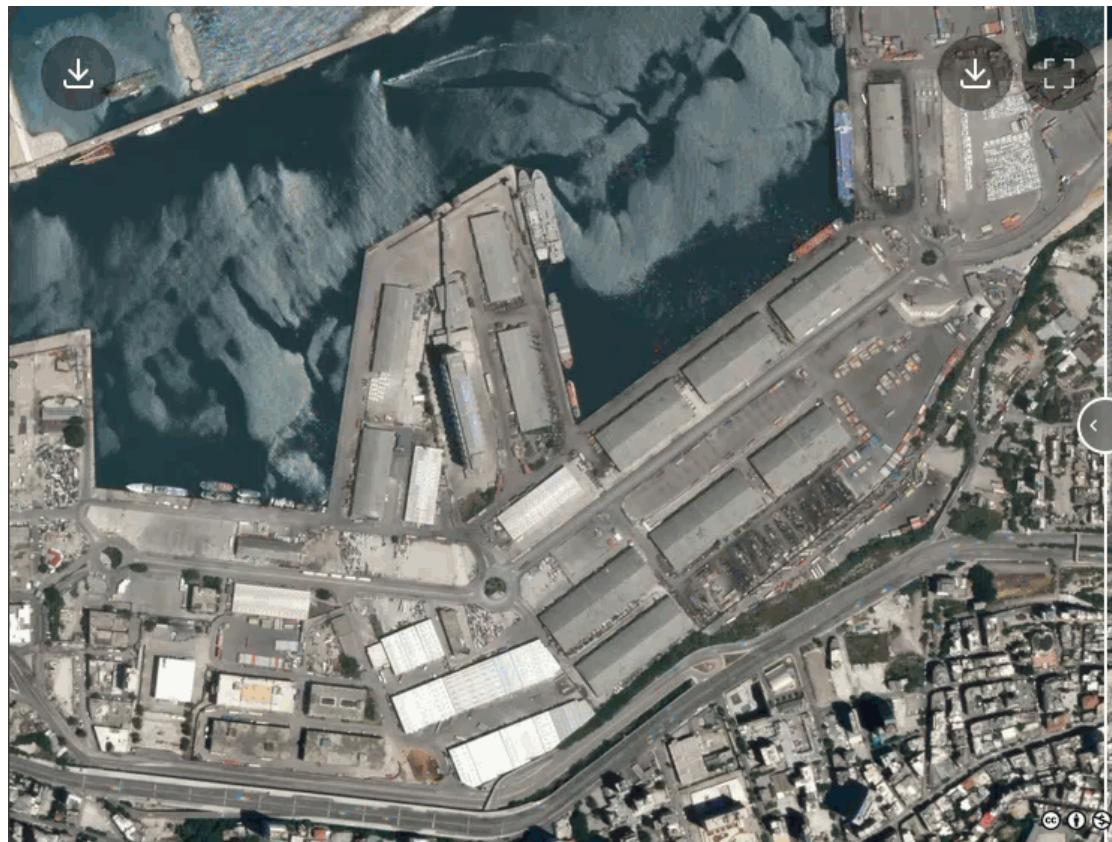
Analysts were able to identify: missile trailers, launching equipment, ...

<https://nsarchive.gwu.edu/document/29153-oct-14-first-u-2-photographs-soviet-missiles-cuba-1962>

Before-After Sweeps

A moving window technique allows one to see before-after comparisons directly

Effect of Ammonium Nitrate explosion at the port of Beirut, Lebanon, Aug. 5, 2020



Ammonium Nitrate Explosion

August 5, 2020 | Beirut, Lebanon

Source: Skysat

This pair of SkySat images shows some of the damage caused by the explosion of 2,750 metric tons (just over six million pounds) of ammonium nitrate in Beirut Port. An image collected on May 31, 2020 (left), shows the warehouse where the material was stored and the surrounding infrastructure. An image collected at 11:17 a.m. on August 5, 2020 (the day after the explosion, right), shows some of the aftermath—a water-filled crater where the warehouse once stood, damaged grain silos, and a capsized cruise ship.

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
<https://www.planet.com/>

Remote sensing tools for telling stories

Satellite data provides [context](#) to the news, [documents](#) events, and a tool for [investigation](#).



U.S. spy satellite image of a Soviet shipyard

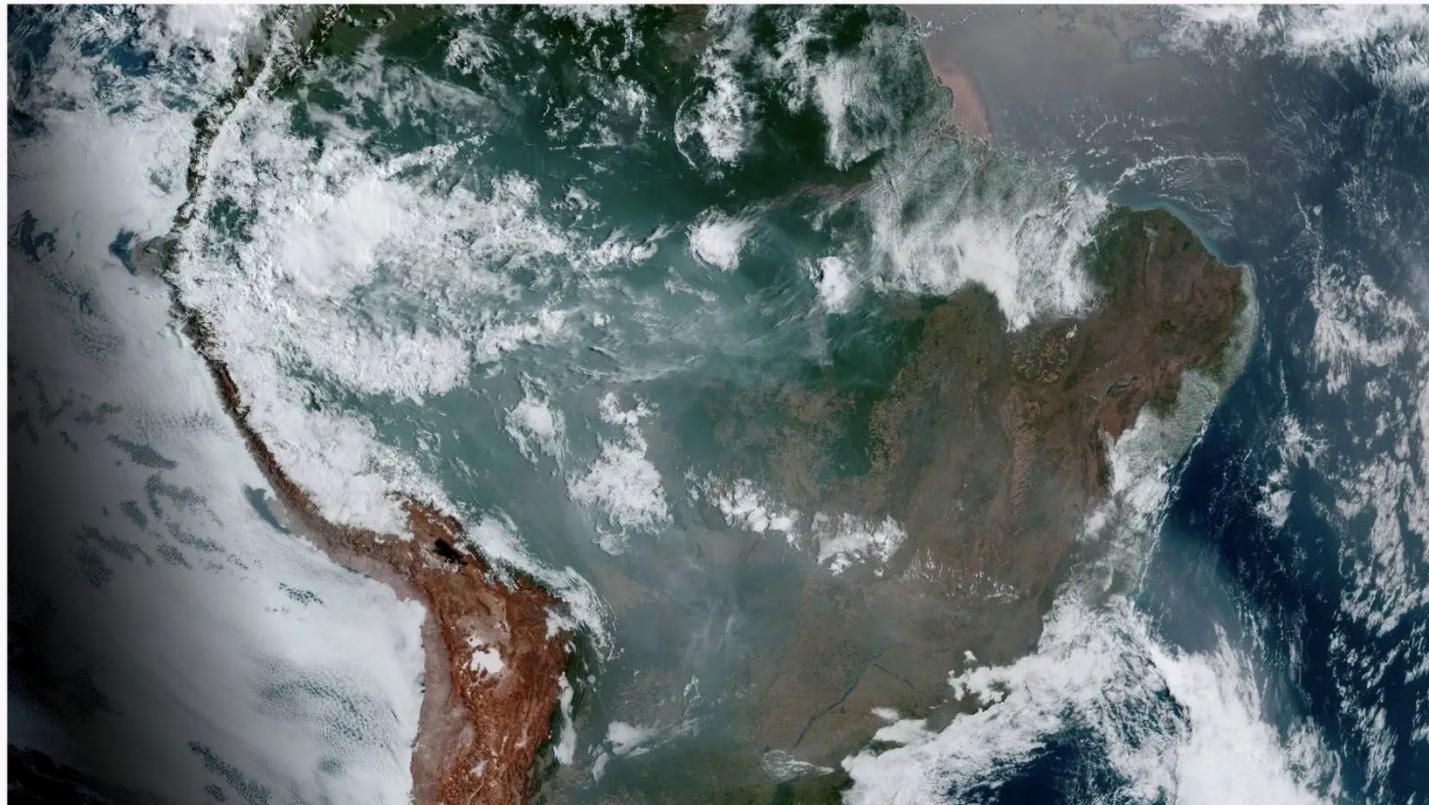
3D rendering of Hurricane Linda (NOAA)



Taken from: Rob Simmon, [From Space to Story in Data Journalism](#)

Satellite image of smoke from wildfires provides the [context & scope](#) for this story, justifying the title

The fires in the Amazon were likely set intentionally

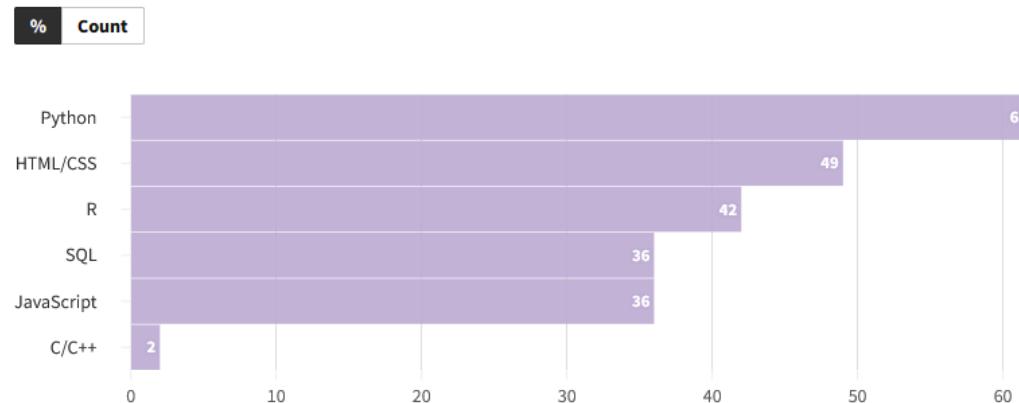


A view of wildfires burning in the Amazon rainforest.

Image: NASA/GOES-16 East

Software for DDJ

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

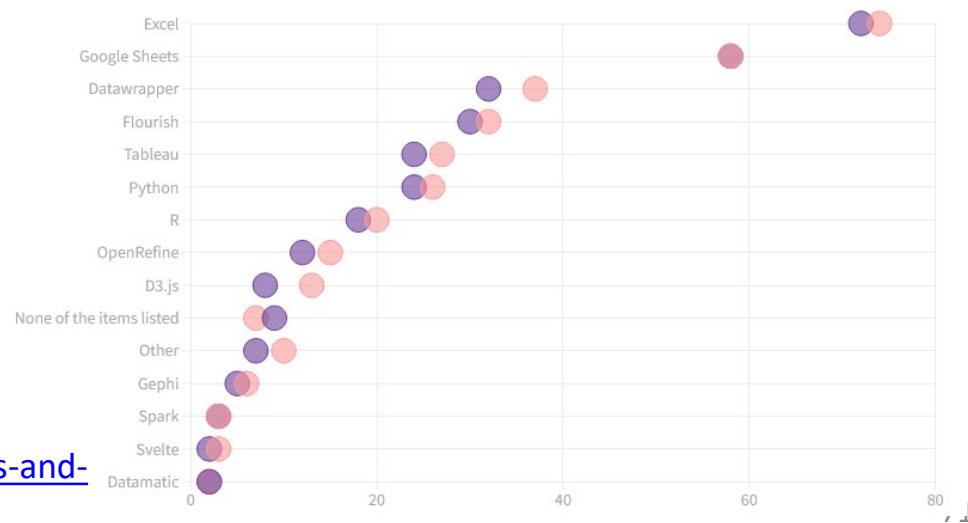


Of those who **program**,
most use Python & HTML,
followed by R

Data analysis & dataviz mostly
done with Excel & Google sheets,
followed by Datawrapper,
Flourish, Tableau

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

Year ■ 2021 ■ 2022

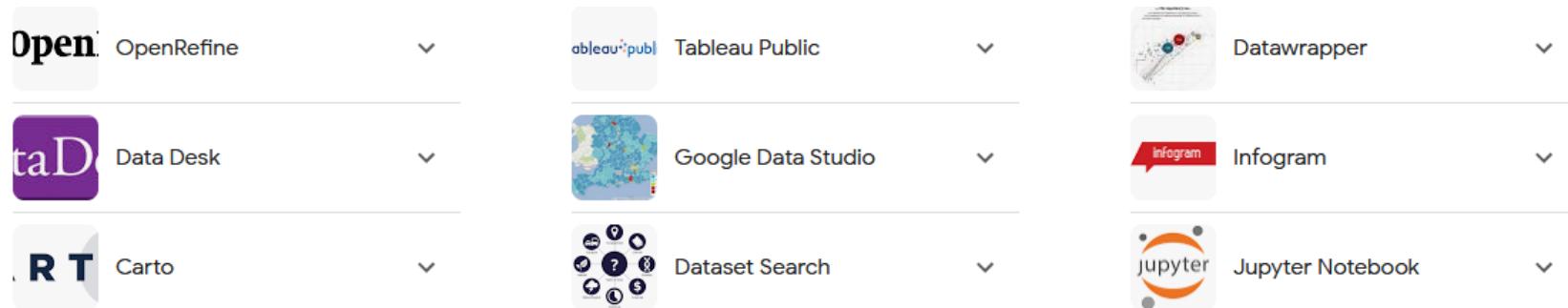


<https://datajournalism.com/survey/2022/skills-and-tools/>

Popular DDJ software

- [OpenRefine](#): Like Excel, but includes data wrangling tools
- [Tableau Public](#): Menu-driven data vis tool
- [Datawrapper](#): Open source tool for easy charts, maps, tables
- [Flourish](#): DDJ tool for creating interactive, dynamic stories
- [Google Data Studio](#): web-based data visualization tool

Ask Google: “software for data journalism”:



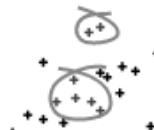
Open Refine

<https://openrefine.org/> free, open source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.



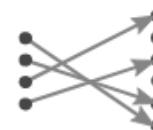
Faceting

Drill through large datasets using facets and apply operations on filtered views of your dataset.



Clustering

Fix inconsistencies by merging similar values thanks to powerful heuristics.



Reconciliation

Match your dataset to external databases via reconciliation services.



Infinite undo/redo

Rewind to any previous state of your dataset and replay your operation history on a new version of it.



Privacy

Your data is cleaned on your machine, not in some dubious data laundering cloud.



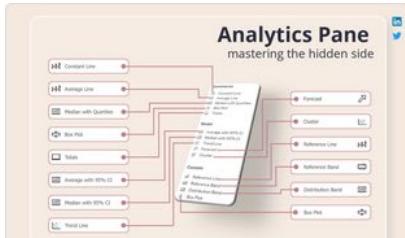
Wikibase

Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

Tableau Public



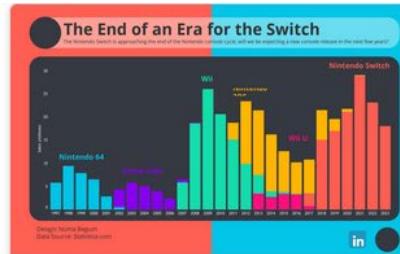
A free platform to explore, create, and publicly share data visualizations online.



Analytics Pane

Adrian Zinovei

☆ 52 ⚡ 1,621



The End of an Era for the Switch

Numa Begum

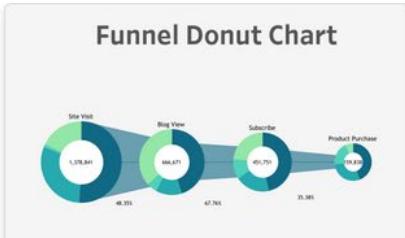
☆ 22 ⚡ 1,142



RWFD - Hospitality Dashboard

Shangruff Raina

☆ 22 ⚡ 518



Funnel Donut Chart

Anindita Mitra

☆ 95 ⚡ 2,624



Les Viz | Fiction Theatre Seating & Sales Dashboard

Nadia

☆ 12 ⚡ 227



8 Ways to Visualize Progress Bars in Tableau with Steps

Fathima Shanavas

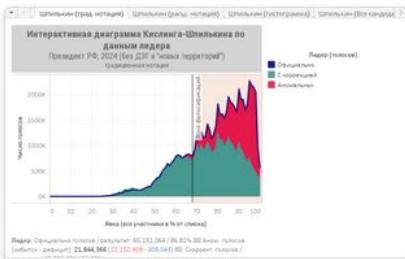
☆ 124 ⚡ 2,994



Railway Dashboard redesign

Ellen Blackburn

☆ 61 ⚡ 1,929



2024.03.17 President of the RF

Roman Udot

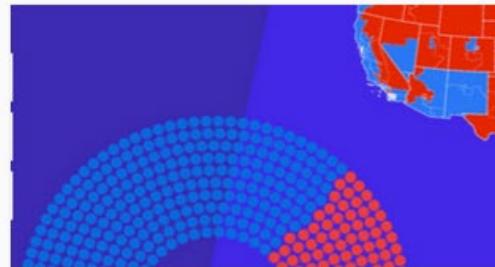
☆ 49 ⚡ 62,449

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



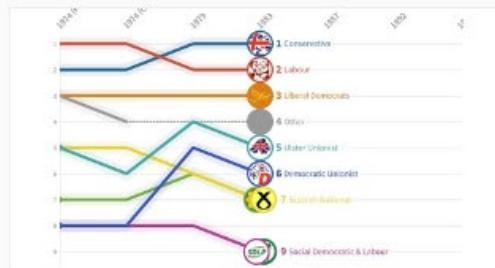
Visualize football and sports data

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



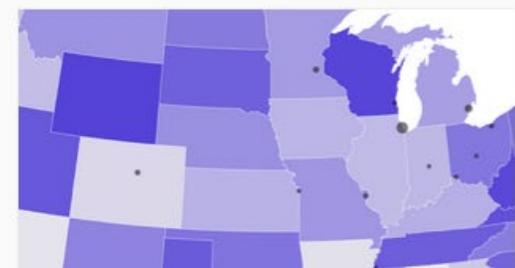
Data Explorer

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

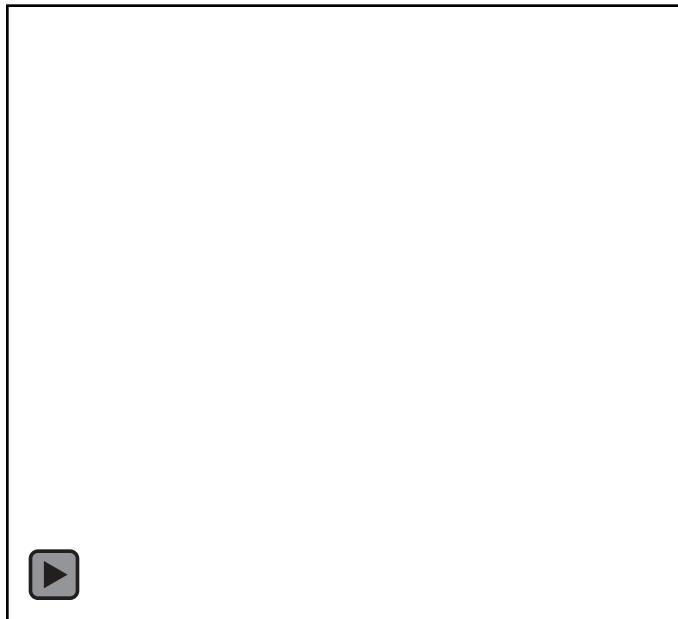


How to choose the right map?

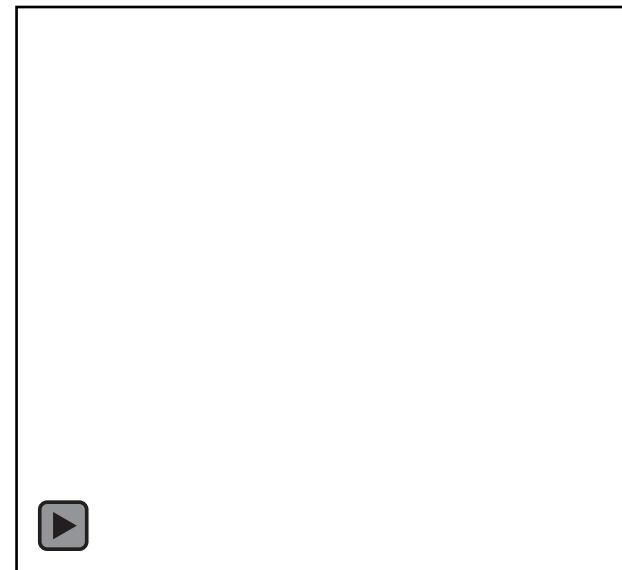
From 3D to symbol maps, use our interactive map templates to visualize geospatial data

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

Drag & drop data import



Animated graphics



Datawrapper

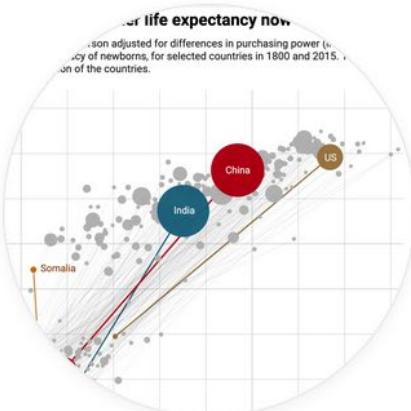
Menu-driven

- Copy/paste, import data
- Choose chart/map/table type, edit (JS under the hood)
- Emphasis on responsive, interactive charts/maps/tables
- Embed responsive code in web or export as static images

Share via URL fullscreen normal size
[//www.datawrapper.de/_/f2FGI/](http://www.datawrapper.de/_/f2FGI/)

Copy Embed code responsive iframe iframe
<iframe title="GDP per capita corr"

PNG PDF SVG DUPLICATE



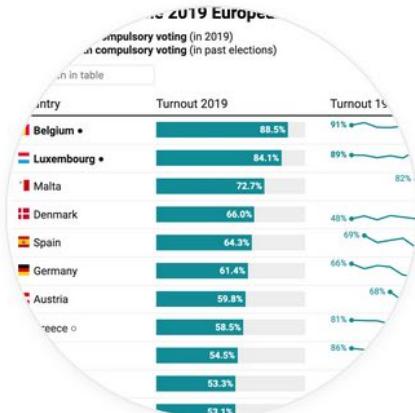
Charts

Use one of **20 interactive and responsive chart types** ranging from simple bars and lines to arrow, range, and scatter plots.



Maps

Three interactive and responsive map types let you create anything from locator maps to thematic choropleth and symbol maps.



Tables

Responsive data tables allow for lots of styling options and let you include bar, column, and line charts, heatmaps, images, a search bar, and pagination.

Datawrapper demo



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**, **data science** and **statistics**} in the digital age.
- New data sources:
 - FOI requests
 - Open Data initiatives
 - Crowd sourcing
 - Satellite journalism
- New graphic forms & reporting
 - Novel graphic displays
 - Interactive story telling
 - Stimulating developments in software
- Powerful impact