

Data Journalism



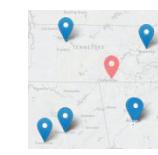
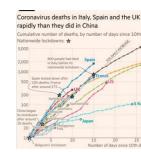
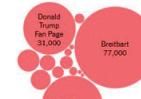
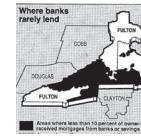
Michael Friendly
Psych 6135

<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



2

What is data journalism?

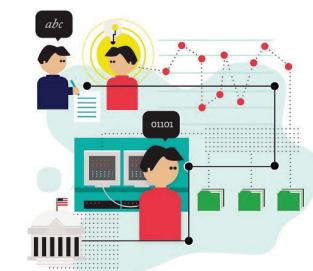
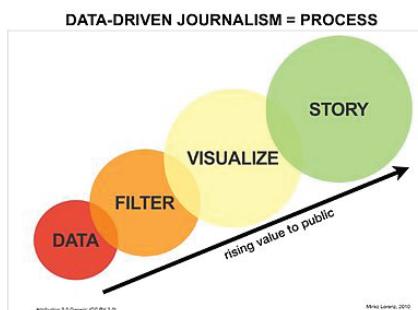
- **Data journalism** reflects the increased role of numerical data for reporting in the digital era. **Data driven** or **DDJ** more useful
 - Reflects increased interaction between journalists and {visualization design, computer science and statistics}.
- Aspects:
 - **Database analysis:** FOI requests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

3

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?

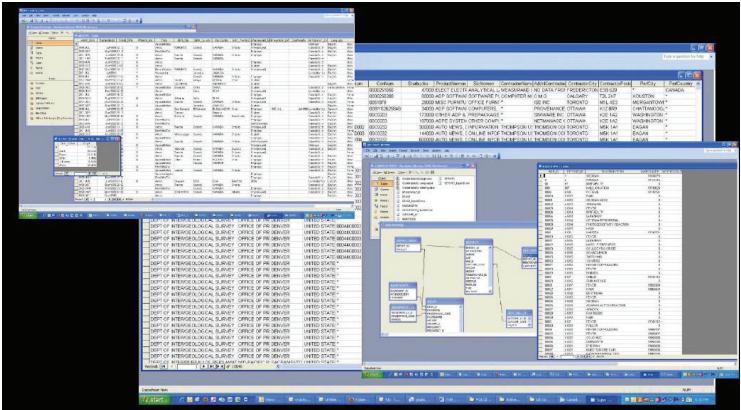


4

“Data” in Data Journalism

Investigative reporting generates messy databases from various sources

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

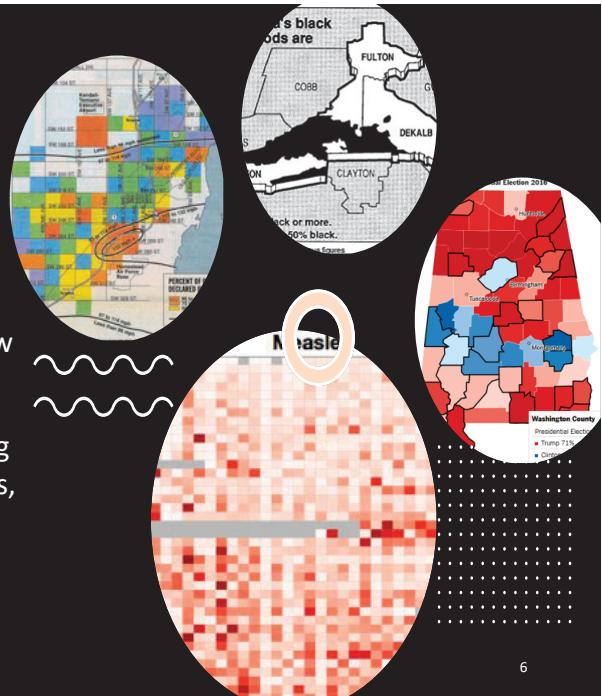


From: Jim Rankin, [Data Successes](#), 6135 Guest lecture, 2020

5

Examples of #DDJ

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



6

Bill Dedman: *The Color of Money*

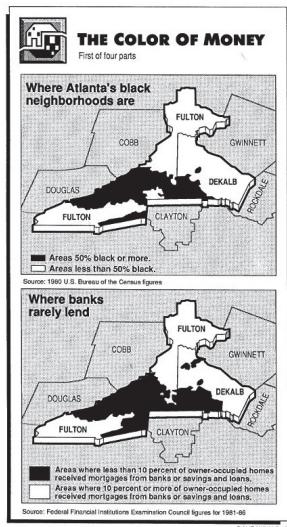
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

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



7

Hurricane Andrew

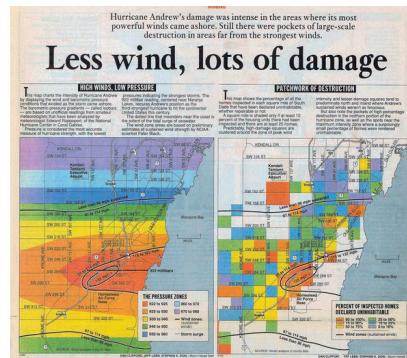
- ❖ In Aug. 1992, a category 5 hurricane hit Miami-Dade county
- ❖ Huge numbers of houses were destroyed, but this varied by neighborhood
- ❖ Investigative reporters at the *Miami Herald* discovered why:
 - Building codes and inspection became lax after 1980.
 - Older homes survived.
- ❖ This report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



Source: <https://www.flickr.com/photos/juggernautco/sets/72157607210036175/>

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



9

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.

10

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>

11

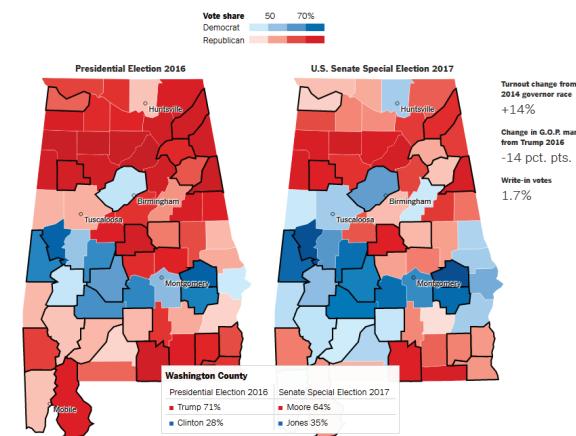
Election reporting: NYT

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?

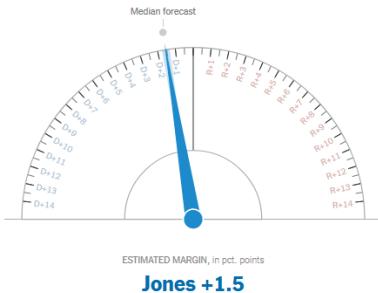


12

Election reporting: Election needle

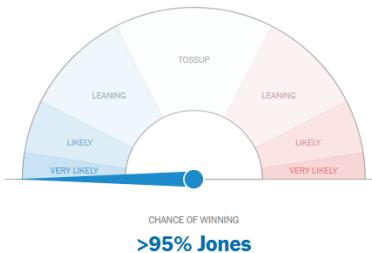
In the 2016 US election, the NYT used **dynamically updated** dashboard displays of the election returns and model-based predictions of the winner of the Alabama Senate race, predicting a win by the Democrat, Doug Jones, over rival Republican, Roy Moore.

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



Two dials:

- Predicted outcome
- Certainty



13

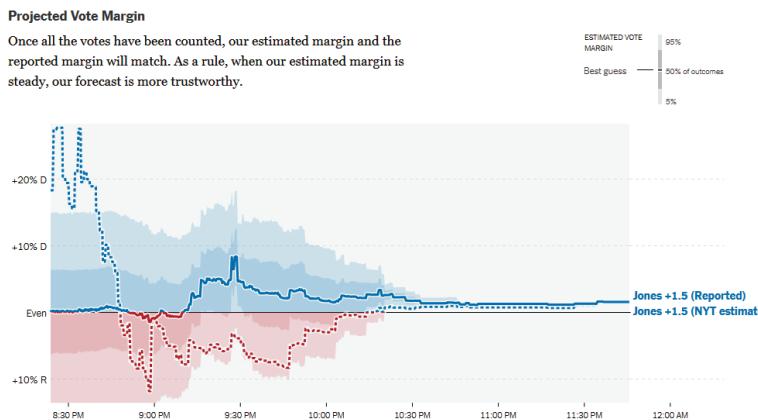
The jittery icon of liberal trauma

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.”



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.



15

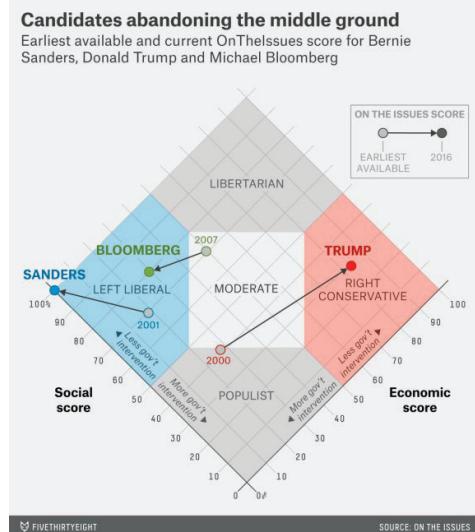
Novel graphic designs

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



16

Measles and vaccines

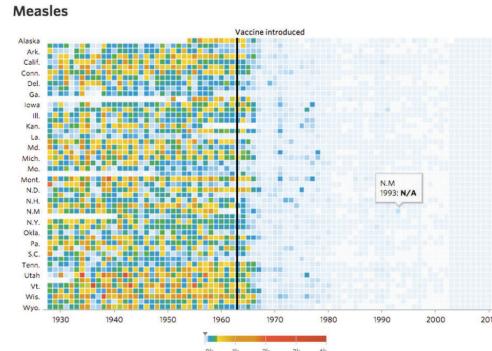
Visualizing the impact of health policy interventions

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.



Source: <http://graphics.wsj.com/infectious-diseases-and-vaccines/>

17

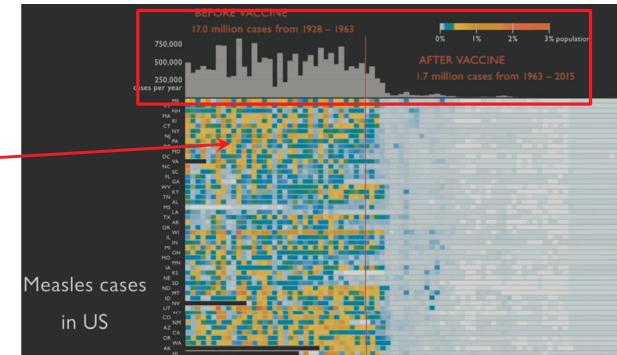
Making the message more explicit

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”



18

Effect ordering & color scales

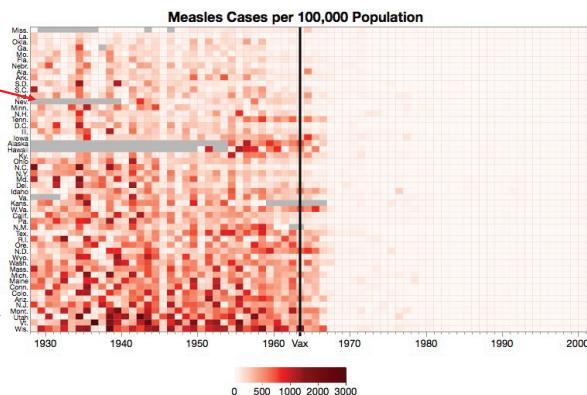
Ordering the states by an overall measure of measles incidence makes it easier to study details.

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?



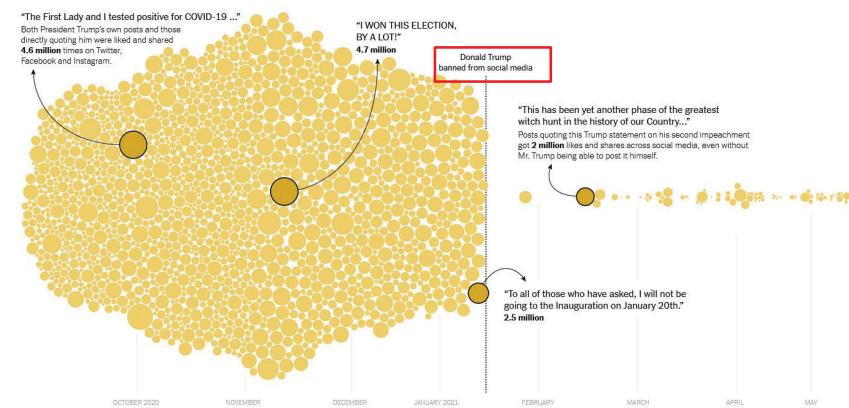
From: <https://communityjmp.com/t5/JMP-Blog/Graph-makeover-Measles-heat-map/ba-p/30550>

19

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

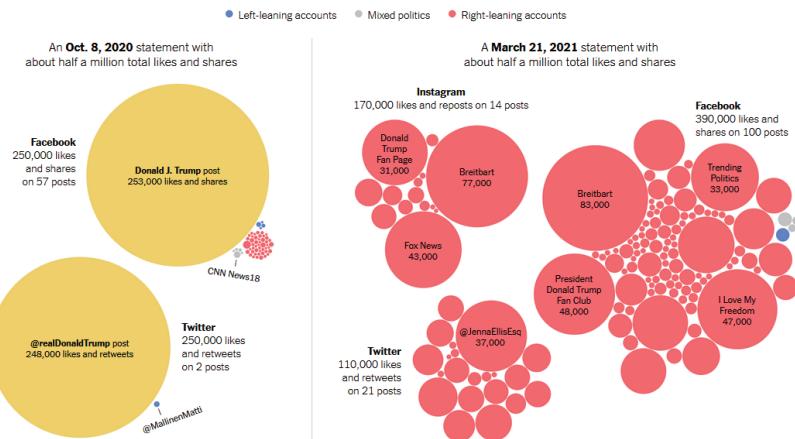


<https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>

20

Before – After: Who pushed the lies?

How/Who Spread Trump statements before and after the ban



21

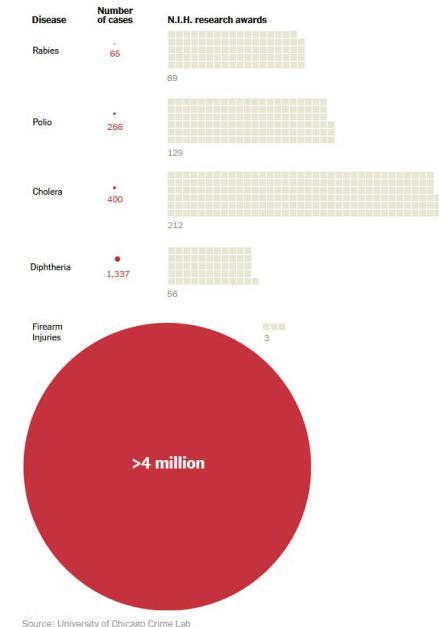
Health & research funding

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?



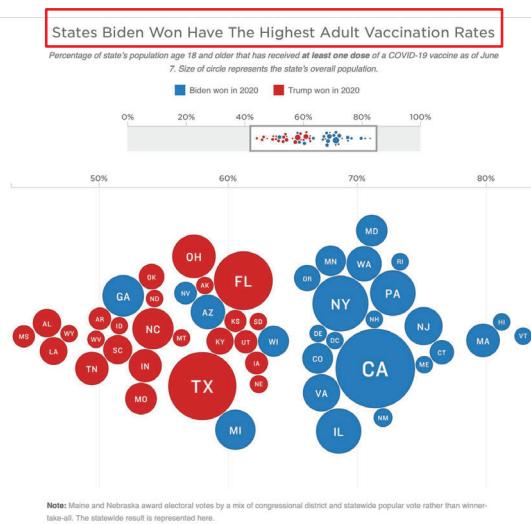
22

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>

23

NYT Infographic: Gun deaths

EACH SYMBOL REPRESENTS 500 GUN DEATHS IN 2016

AN ESTIMATED 22,000 GUN SUICIDES



ABOUT 11,760 HOMICIDES



"a single death is a tragedy, a million deaths are a statistic" – Joseph Stalin

Infographics use simple icons to give emotional impact to raw numbers

But journalists always use the 'single death' in the lede

24

NYT: Gun deaths & policy

Guns per 100 people

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

UNITED STATES 120.5

CANADA	34.7
SWITZERLAND	27.6
SWEDEN	23.1
FRANCE	19.6
GERMANY	19.6
AUSTRALIA	14.5
ITALY	14.4
SPAIN	7.5
ENGLAND, WALES	4.6
JAPAN	0.3

These simple barcharts have a grim conclusion

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

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

25

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.

UNITED STATES 3.4

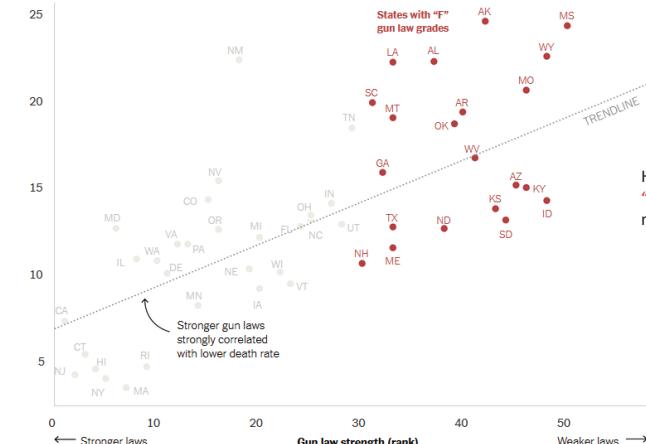
CANADA	0.6
FRANCE	0.4
SWEDEN	0.4
ITALY	0.3
SWITZERLAND	0.2
AUSTRALIA	0.1
GERMANY	0.1
SPAIN	0.1
ENGLAND, WALES	0
JAPAN	0

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 & gun laws

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

Gun deaths per 100,000 people



Source: Giffords Law Center

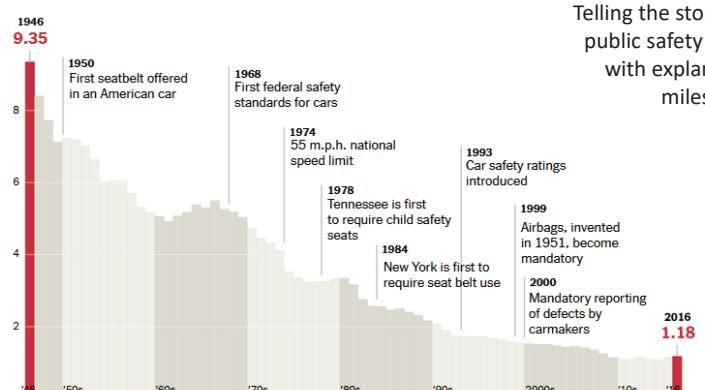
Highlighting the "F" grades to make the point

26

Seatbelts: Story telling

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

Deaths per 100 million vehicle miles traveled



Telling the story of a public safety policy with explanatory milestones

27



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

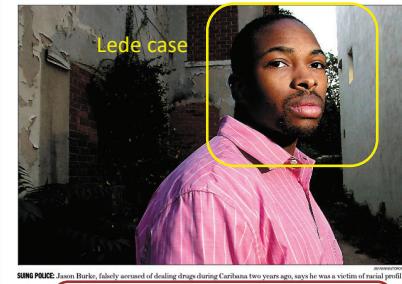
HOW TO WRITE THIS STORY?

THE SATURDAY STAR



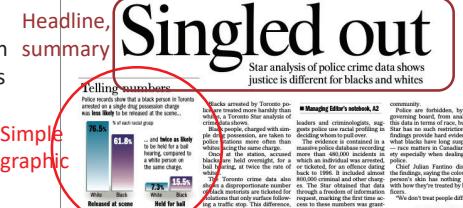
Periods of rain: windy. High: 14 C
October 19, 2002
[thestar.com](#) ONTARIO EDITION

AN INVESTIGATION INTO RACE AND CRIME



Singled out

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



Chrétien expected to keep cabinet minister

Ethics report has 'wiggle room' to save MacAulay

BY TIM HARVEY AND LES WHITTINGTON TORONTO STAR

OTTAWA — Jean Chrétien has told his ethics committee today that it expects to keep his former chief of staff, Lawrence MacAulay, in his cabinet.

Ethics committee chair Howard Sapers said he will make a final decision and deliver it to Prime Minister Paul Martin by Oct. 24, where it was received by Chrétien's office yesterday.

It was thought to be rejected to the ethics committee that MacAulay, a French-speaking nation. It was accepted because the fax arrived as Chrétien was leaving the country.

Source said that right before the fax arrived, Chrétien had a meeting with his general, Lawrence MacAulay, to discuss his appointment to cabinet.

One of the stations, around 600,000 people, is Black. Of the stations, around 400,000 residents are Black. In the same year, Black people were arrested at twice the rate of white people.

Toronto crime data also showed that Black people were stopped, searched and arrested more often than white people.

Police said that only surface-level information was available to the public.

Chief Julian Fantino deplored the statistics, but said they were not surprising, marking the first time such a comparison had been made.

"We don't treat people differently," he said.

IN INSIDE

Barrie 23

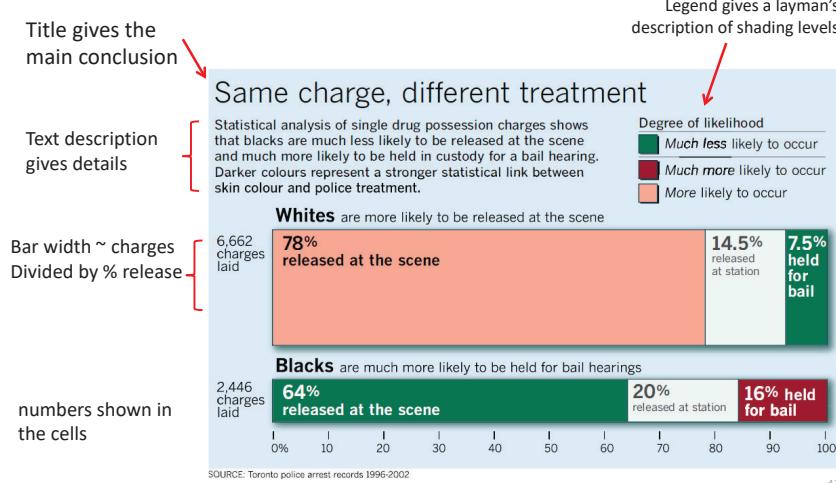
Brantford 47

Brantford 47

Brantford 47

Racial profiling: Presentation graphic

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



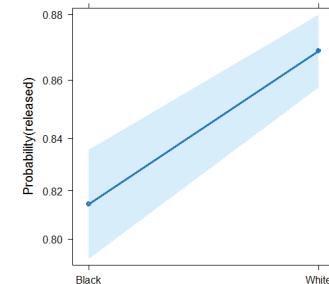
Marijuana data: Analysis

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", ...)
```

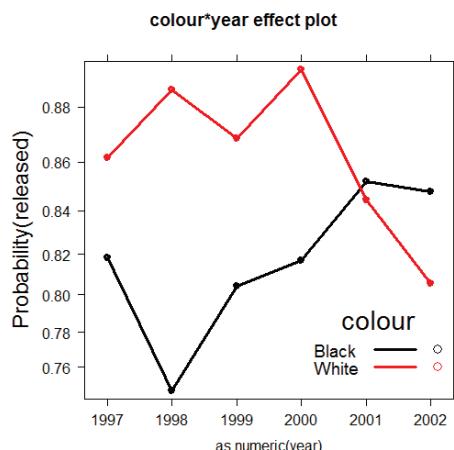


- Effect plot shows average effect **controlling** (adjusting) for **all** other factors simultaneously
- (*Star* analysis controlled for these one at a time.)
→ Strong evidence for different treatment of blacks & whites
- But effect **size** smaller than reported by the *Star*

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

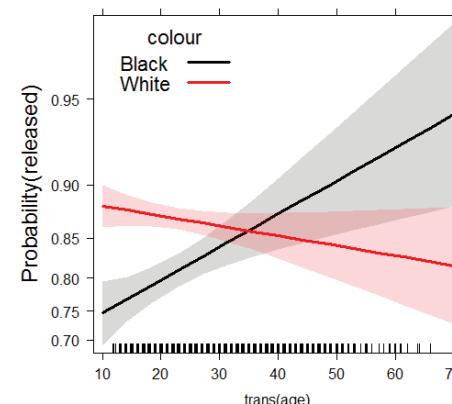
Also: evidence to support Police claim of effect of training to reduce racial effects in treatment

Effect plots: Interactions

A more surprising finding ...

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

Effects of skin colour and age on release



Opposite age effects for blacks & whites:

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

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, ...)

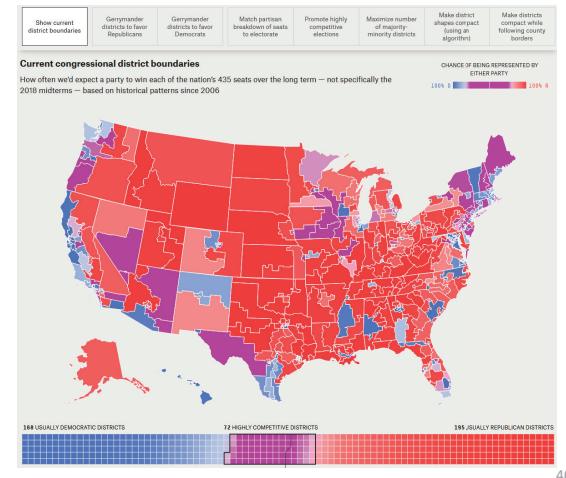
39

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?



40

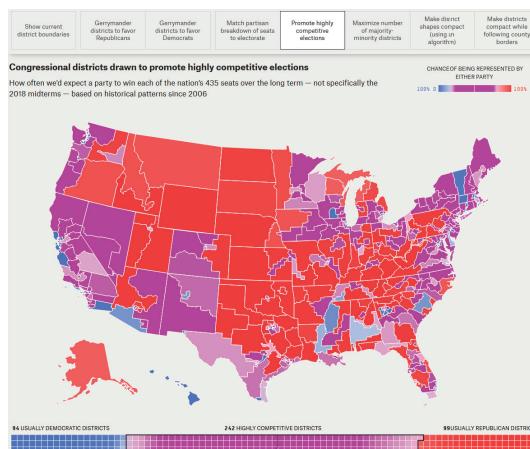
538: Atlas of Redistricting

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/>



41

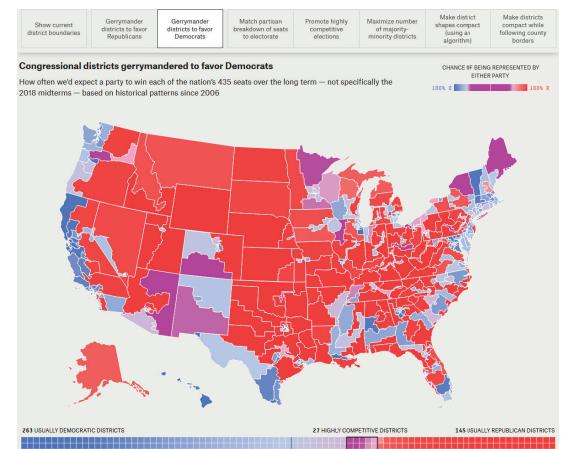
538: Atlas of Redistricting

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/>

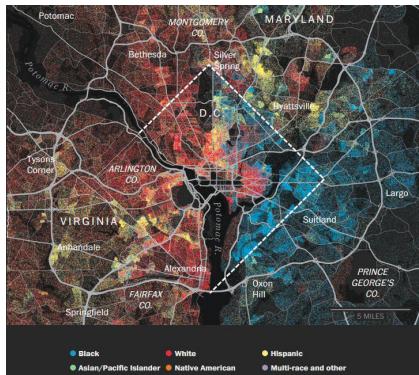


42

WP: Segregation in U.S. Cities

Aaron Williams & Armand Emadjoemeh 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.



From: <https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/>

43

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?

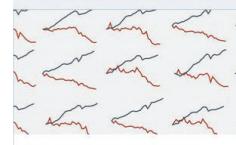
By Hannah Ritchie – June 07, 2021



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.

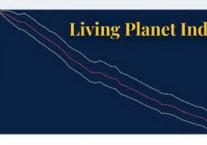
By Hannah Ritchie – June 04, 2021



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.

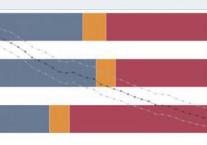
By Max Roser – June 01, 2021



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

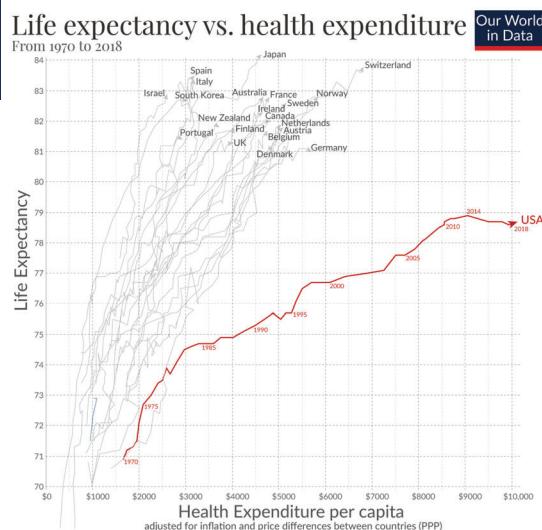
44

Our World in Data

What is the main message of this chart?

Why is it effective?

Would using log expenditure be better or worse?



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.

45

Financial Times COVID tracker

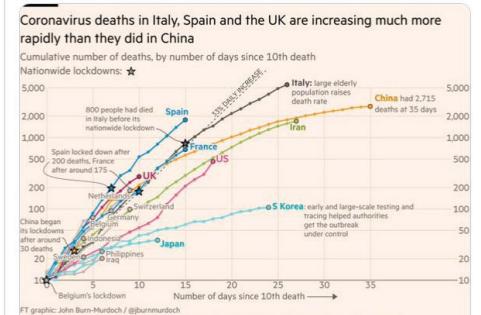
John Burn-Murdoch @jburnmurdoch

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-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

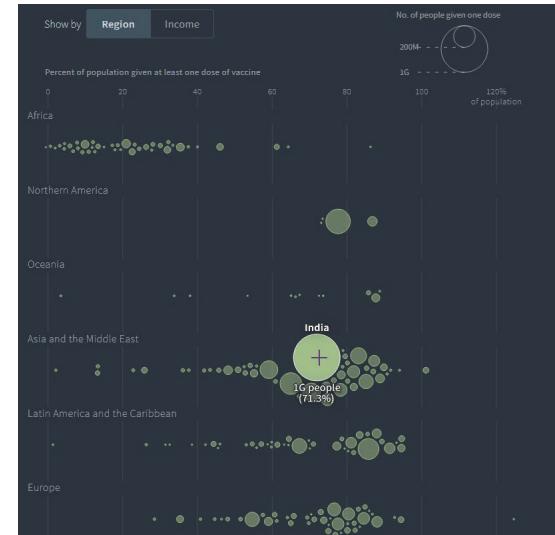
46

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."

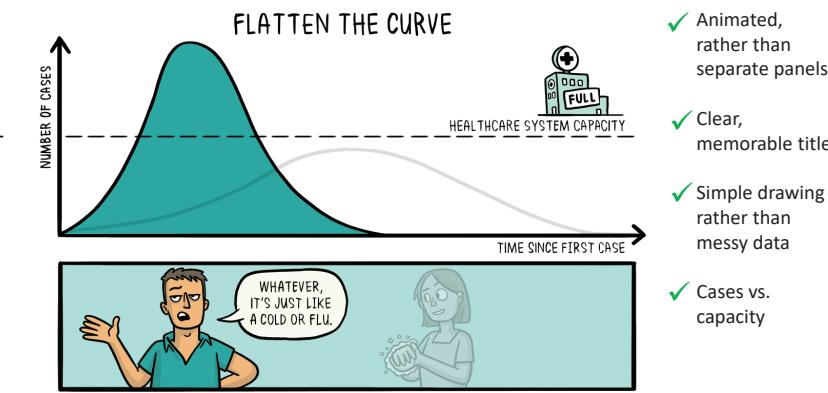
That meant creating a unifying structure to catalogue the different phases of each country's vaccination rollout plan."
-- Prasanta Kumar Dutta

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

48

Flatten the curve

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



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

49

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/>

50

Toronto Star: Trump Lies

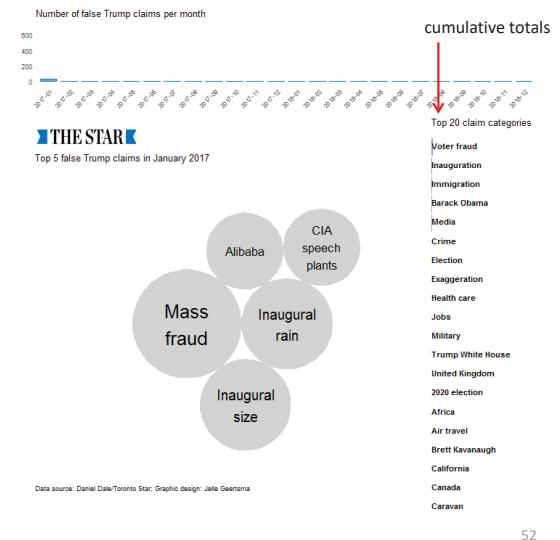
Every false claim is documented: place, claim, truth, "slug", categories

	claim	truth	category
1	1 2017-01-20 Post-inauguration...	"I looked at the rain, which...	Inauguration
2	2 2017-01-20 Post-inauguration...	'Even the media said the...	Inauguration
3	3 2017-01-21 Speech at Centr...	'Honestly, it looked like a...	Exaggerat
4	4 2017-01-21 Speech at Centr...	'It was almost raining, the...	Inauguration
5	5 2017-01-21 Speech at Centr...	Neither of these claims is true. ...	Exaggerat
6	6 2017-01-23 Private meeting ...	Trump's numbers are well off. H...	Media
7	7 2017-01-25 Interview with A...	'So a reporter for TIME ma...	Exaggerat
8	8 2017-01-25 Interview with A...	'It was almost raining, the...	Voter fraud
9	9 2017-01-25 Interview with A...	'They can mean anyone, but n...	Election
10	10 2017-01-25 Interview with A...	'No, no, you have to unde...	Exaggerat
11	11 2017-01-25 Interview with A...	'Trump's victory was not close ...	Election
12	12 2017-01-25 Interview with A...	'Most of the audience was inde...	Exaggerat
13	13 2017-01-25 Interview with A...	'With regard to his speech ...	US intellig
14	14 2017-01-25 Interview with A...	'Healthcare.gov...	Health care
15	15 2017-01-25 Interview with A...	'Regarding Healthcare.gov...	Jobs
16	16 2017-01-25 Interview with A...	'Regarding voting fraud: "...	Military
17	17 2017-01-25 Interview with A...	'There is no evidence of a signif...	Trump White House
18	18 2017-01-25 Interview with A...	'Regarding voting fraud: "...	United Kingdom

51

Jelle Geertsma: Trump Lies

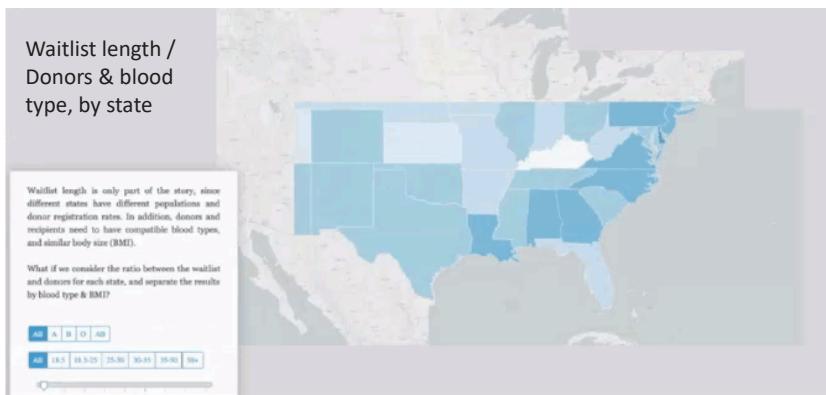
Graphic artist Jelle Geertsma designed this animated graphic showing the top 5 claim "slugs" & the top 20 categories over time, by month.



52

Scrollytelling: The Impatient List

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



From <https://xujunjiejack.github.io/>

53

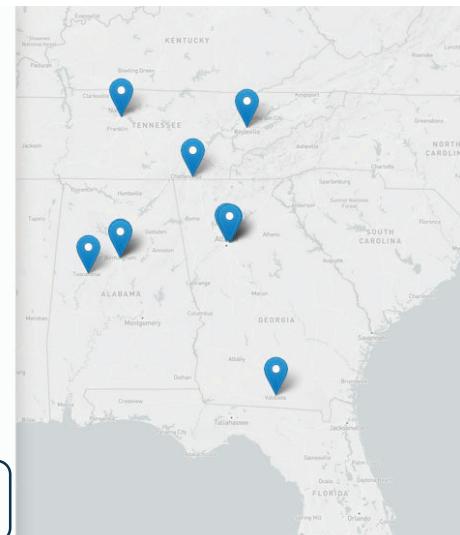
Scrollytelling: The Impatient List

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

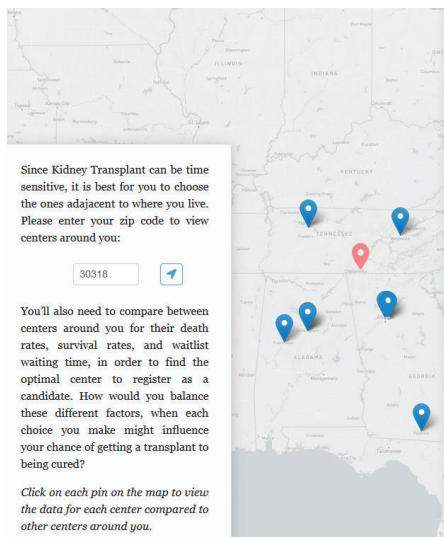
You'll also need to compare between centers around you for their death rates, survival rates, and waitlist waiting time, in order to find the optimal center to register as a candidate. How would you balance these different factors, when each choice you make might influence your chance of getting a transplant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

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



54

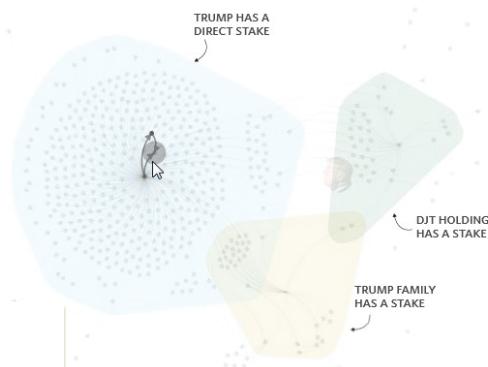
Scrollytelling: The Impatient List



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)	

57

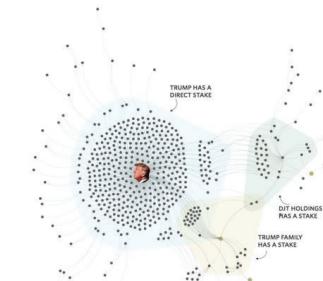
Drill-down to tell a story

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. 18, 2017 at 6:00 a.m. EST



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.comgraphics/donald-trump-potential-conflicts-of-interest/>

56

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

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 charged 'second home' designation four times in four years

The Business Secretary
Mandelson claimed thousands after saying he would quit as MP

58

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.

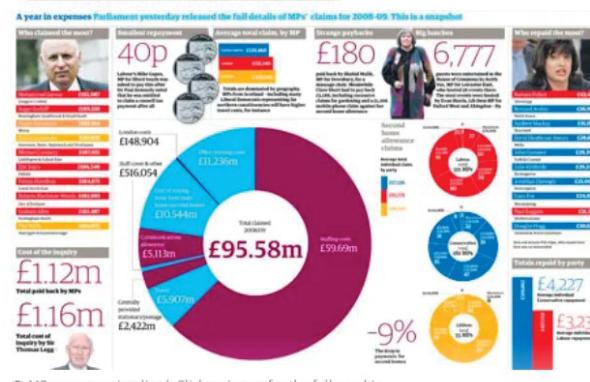
<https://www.theguardian.com/news/datablog/2009/sep/18/mps-expenses-westminster-data-house-of-commons>

59

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

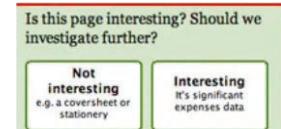
- Who claimed what?
- The Legg report's full list of repayments



61

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.



<https://www.theguardian.com/news/datablog/2009/sep/18/mps-expenses-westminster-data-house-of-commons>

60



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
IJC: 280 top investigative journalists, >100 media partners, >100 countries



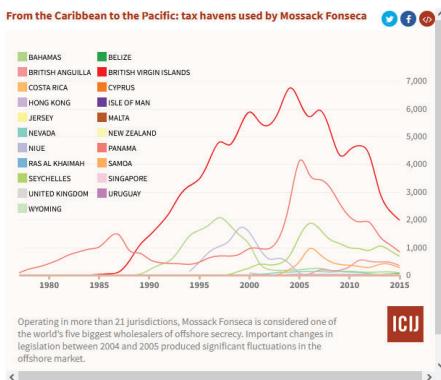
<https://www.icij.org/investigations/panama-papers/>

62

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?

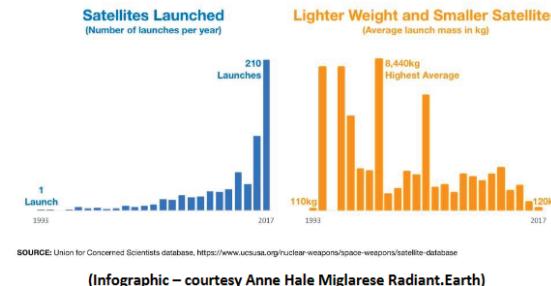
63

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.



(Infographic – courtesy Anne Hale Migliarese Radiant.Earth)

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

64

Satellite Journalism: Smoking Guns

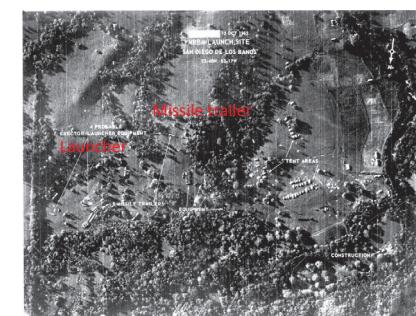
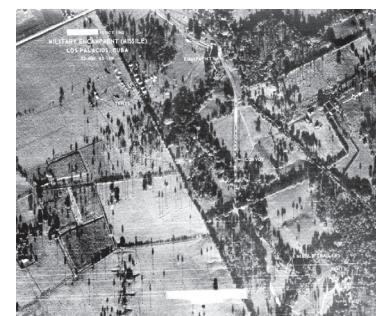
- Satellite imagery has provided 'smoking-gun' evidence for:
 - investigative reporting of human rights abuses (Htusan et al., 2015)
 - exposure of environmental destruction in remote locations (e.g., the Amazon: Weir, 2017)
 - before & after imagery of natural disasters (Murphy, 2018)
 - analyses of 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

65

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, ...

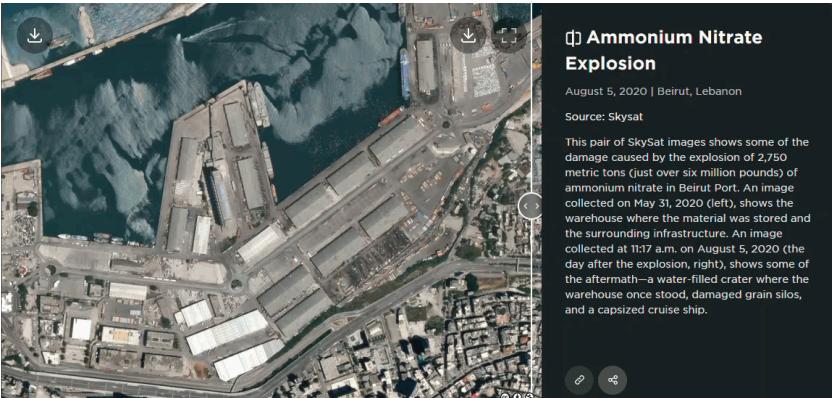
<https://nsarchive.gwu.edu/document/29153-oct-14-first-u-2-photographs-soviet-missiles-cuba-1962>

66

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



<https://www.planet.com/gallery/#!/post/ammonium-nitrate-explosion>

67

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.



January 29, 2016



July 21, 2016

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.



November 4, 2016



January 20, 2017

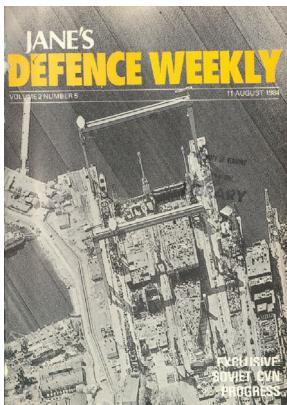
The Peruvian government intervened and actively targeted illegal mining operations inside the reserve.

Images: Planet Labs
<https://www.planet.com/>

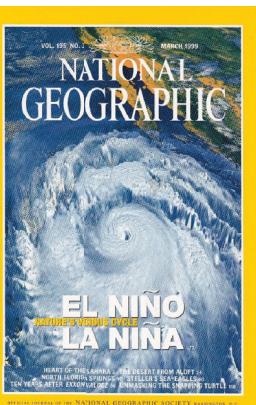
68

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



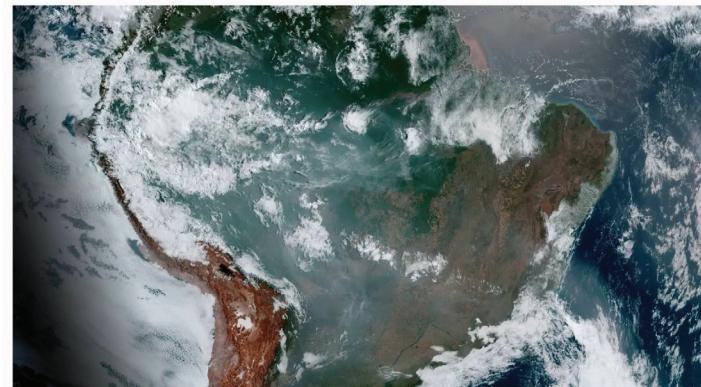
3D rendering of Hurricane Linda (NOAA)

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

69

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.

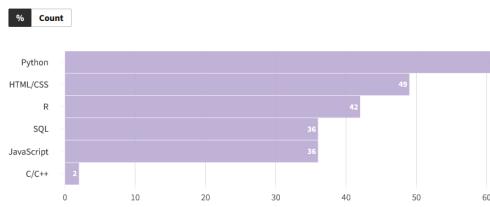
Image: NASA/GOES-16 East

Zoë Schlanger & Daniel Wolfe, [The fires in the Amazon were likely set intentionally](#), Quartz, Aug. 2019

70

Software for DDJ

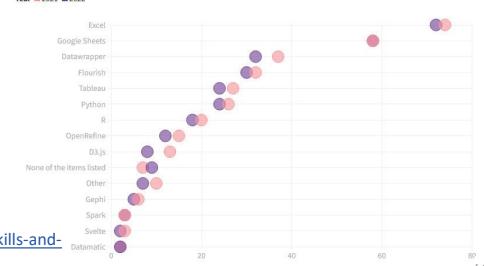
Which programming languages do you use for your work, if any?



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

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

Year 2021 2022



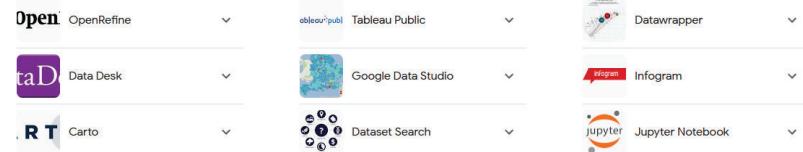
Data analysis & dataviz mostly done with Excel & Google sheets, followed by Datawrapper, Flourish, Tableau

<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":



72

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.

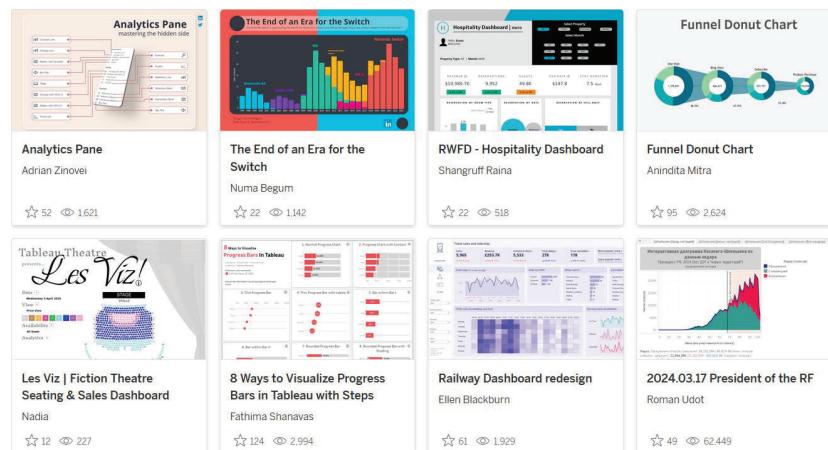
Wikibase

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

73

Tableau Public

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



74

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



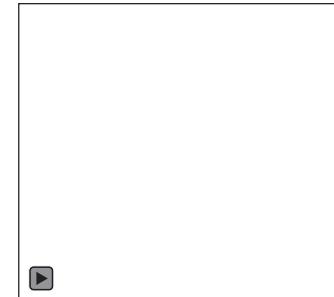
How to choose the right map?

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

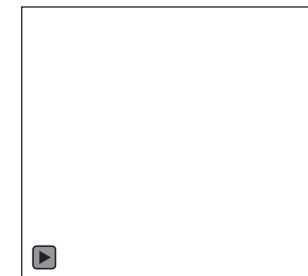
See: <https://flourish.studio/>

75

Drag & drop data import



Animated graphics

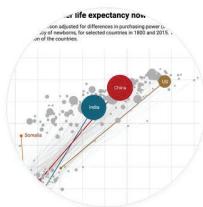


76

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



Charts

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



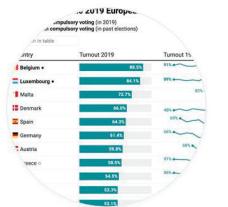
Maps

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

Share via URL fullscreen normal size
<http://www.datawrapper.de/r2FGI/>

Copy Embed code responsive iframe iframe
<iframe title="GDP per capita.com" copy>

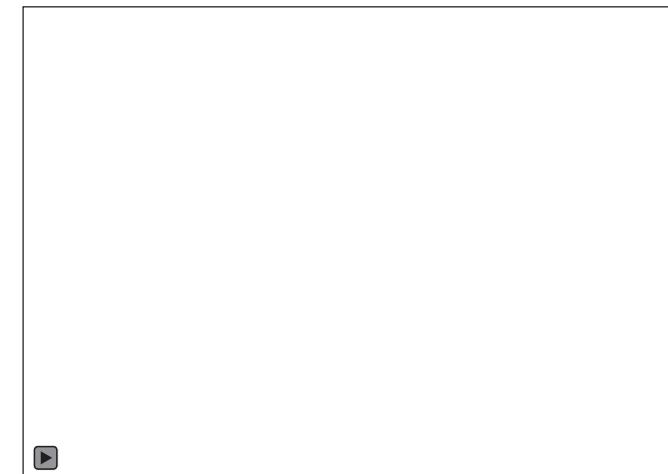
PNG PDF SVG DUPLICATE



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



77

78

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>



79

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

80