

Data Journalism



Michael Friendly
Psych 6135

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



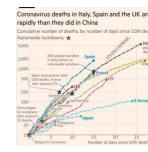
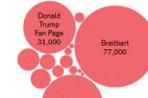
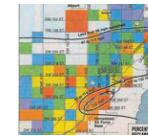
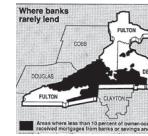
What is data journalism?

- **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

Topics

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

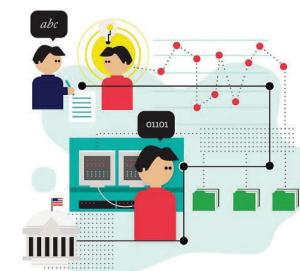
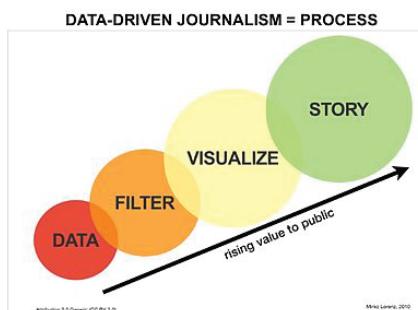


2

What is data journalism?

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?

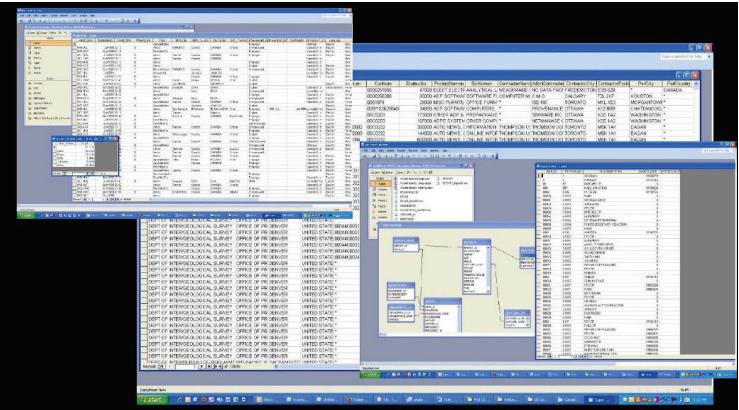


4

“Data” in Data Journalism

Investigative reporting generates messy databases from various sources

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

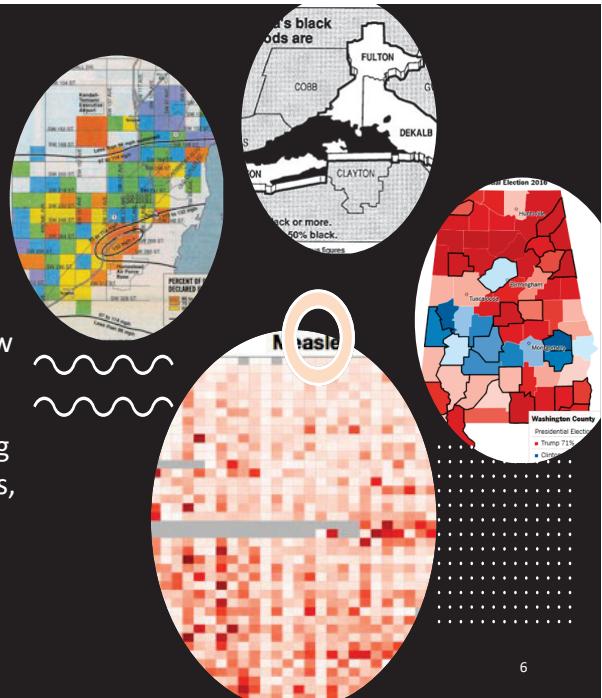


From: Jim Rankin, [Data Successes](#), 6135 Guest lecture, 2020

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*

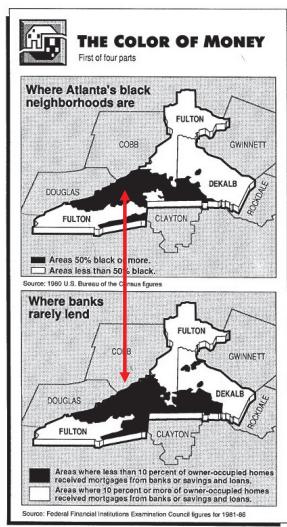
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

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.

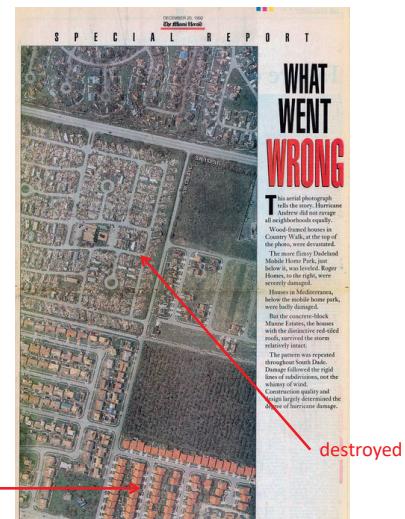
See: <http://powerreporting.com/color/> for all the articles and images



7

Hurricane Andrew

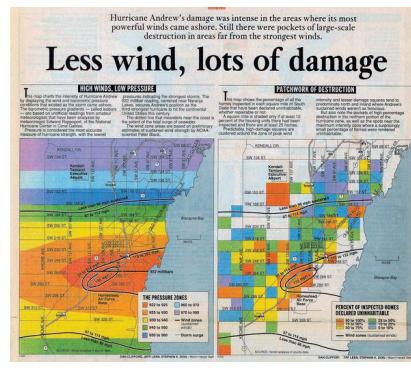
- ❖ In Aug. 1992, a category 5 hurricane hit Miami-Dade county
- ❖ Huge numbers of houses were destroyed, but this varied by neighborhood
- ❖ Investigative reporters at the *Miami Herald* discovered why:
 - Building codes and inspection became lax after 1980.
 - Older homes survived.
- ❖ This report 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
- ❖ Overwhelming evidence for cause of destruction beyond wind
- ❖ 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**"



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

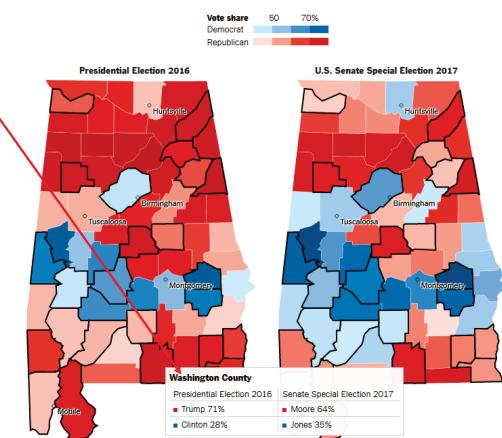
Election reporting: NYT

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?

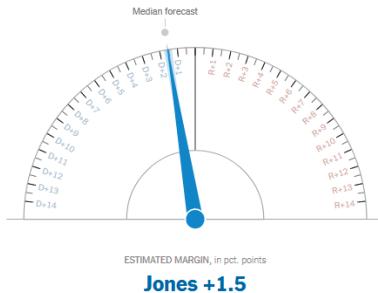


12

Election reporting: Election needle

In the 2017 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**.

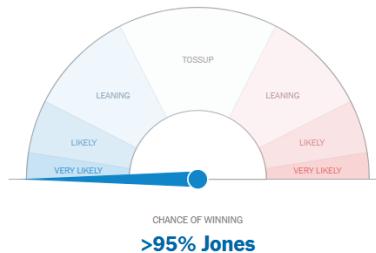
Doug Jones has won the election, according to A.P.



Jones +1.5

Two dials:

- Predicted outcome
- Certainty



>95% Jones

13

The jittery icon of liberal trauma

But in the 2016 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.'**"



Trump +0.0

FORECAST, in pct. points

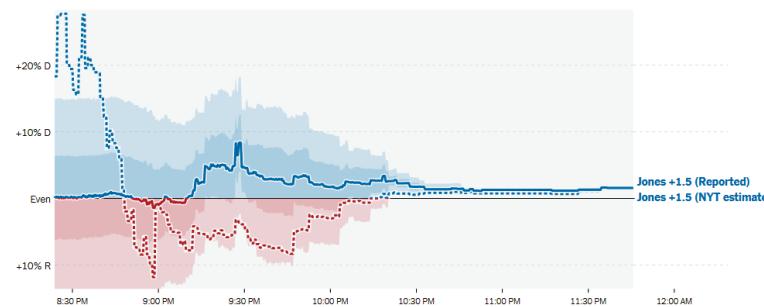
The moment our long national nightmare began

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

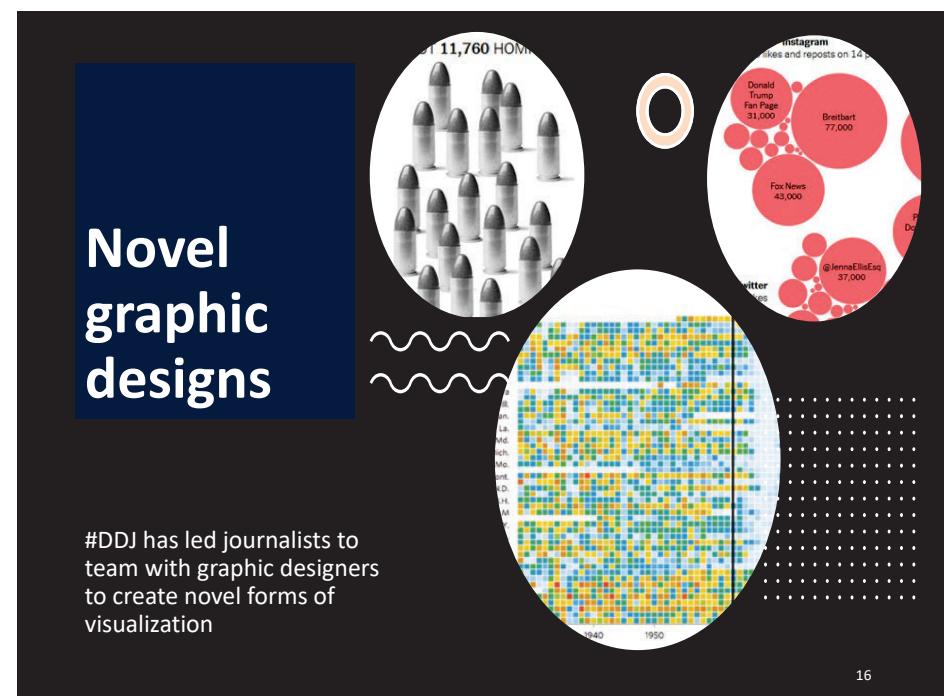
Once all the votes have been counted, our estimated margin and the reported margin will match. As a rule, when our estimated margin is steady, our forecast is more trustworthy.



15

Novel graphic designs

#DDJ has led journalists to team with graphic designers to create novel forms of visualization



16

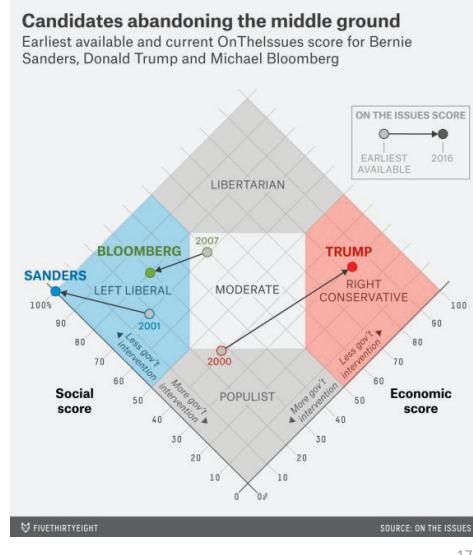
Novel graphic designs: Zone chart

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



17

Measles and vaccines

Visualizing the impact of health policy interventions

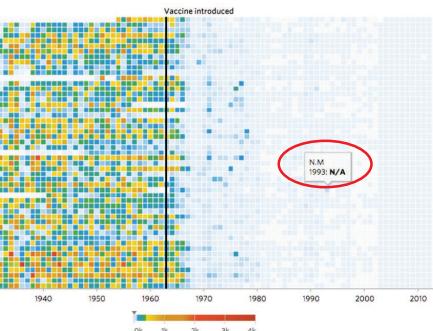
In 2015 Tynan DeBold & Dov Friedman in the *Wall Street Journal* tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases

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.

Measles



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

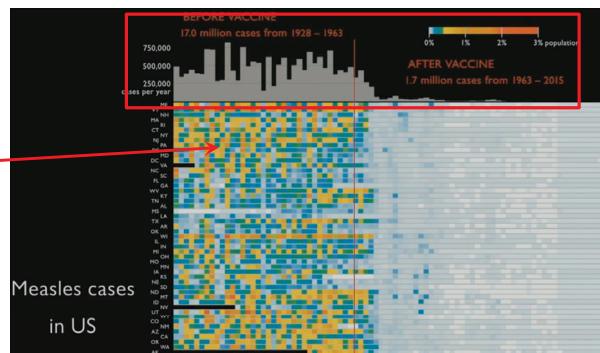
18

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”

19

Effect ordering & color scales

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

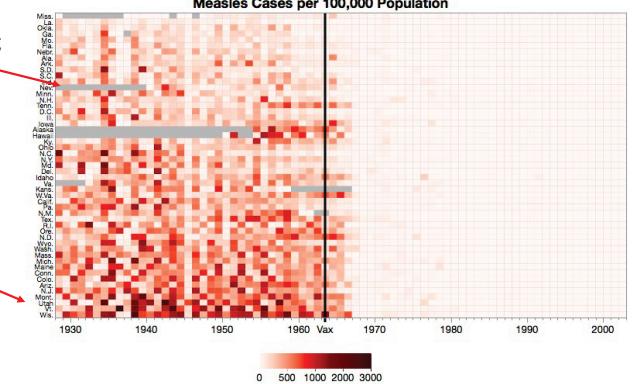
A uni-dimensional scale of **color** is more effective than the ad-hoc scale used in the original

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?

Measles Cases per 100,000 Population



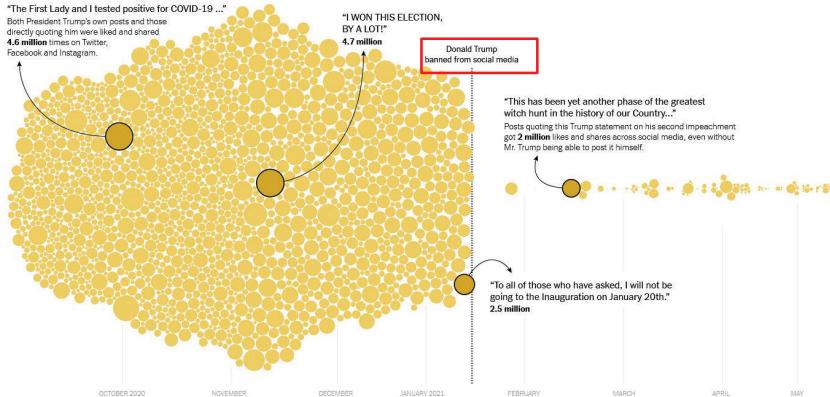
From: <https://communityjmp.com/t5/JMP-Blog/Graph-makeover-Measles-heat-map/ba-p/30550>

20

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



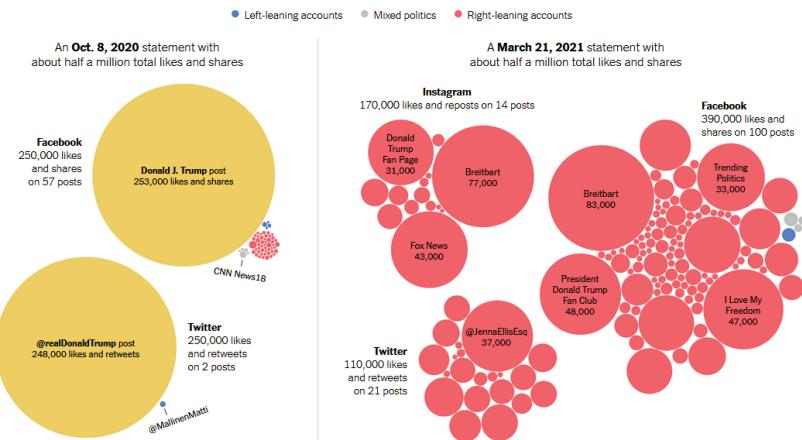
<https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>

21

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



22

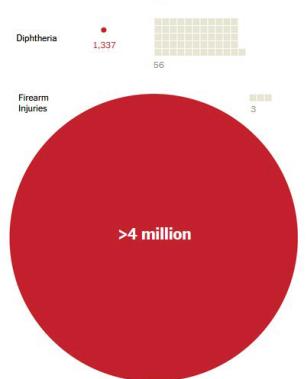
Health & research funding

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

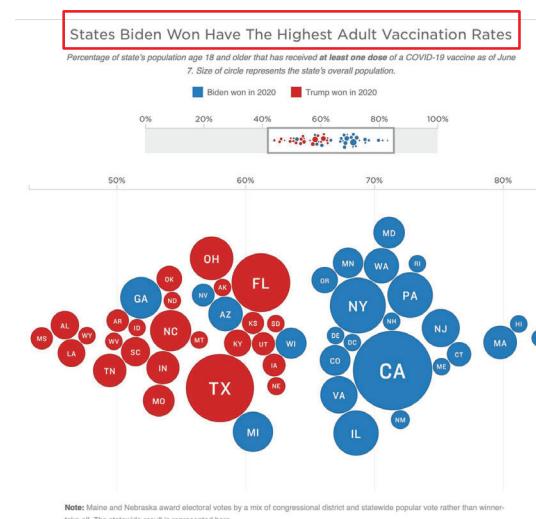
23

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>

24

NYT Infographic: Gun deaths

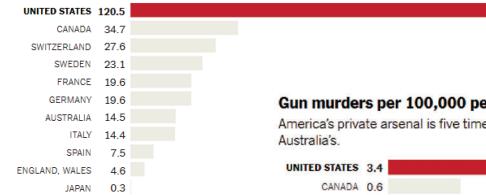


25

NYT: Gun deaths & policy

Guns per 100 people

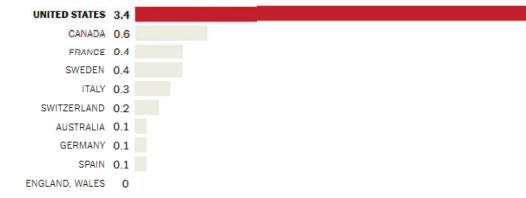
Among developed countries, the United States has by far the highest rate of firearms ownership.



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.



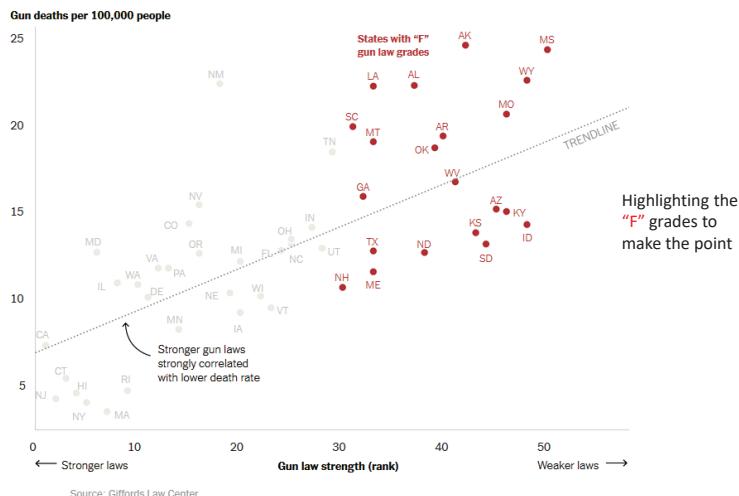
Sources: Small Arms Survey, United Nations Office on Drugs and Crime. • Ownership rates are for 2017. Murder rates for the U.S., Canada, Sweden, Switzerland, Australia and Spain are from 2016; otherwise, the latest available rates are used.

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

26

NYT: Gun deaths & gun laws

This graph directly linked gun deaths to the strength of states' [laws on gun ownership](#)

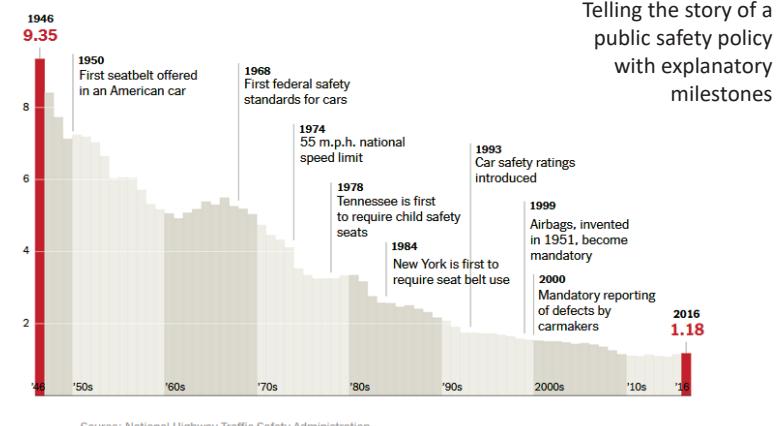


27

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

28

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?

Headline, summary
Simple graphic

THE SATURDAY STAR

The photo that never was
SAINTS WORLDS, C3

Periods of rain; windy. High 14°C

October 19, 2002

thestar.com ONTARIO EDITION

AN INVESTIGATION INTO RACE AND CRIME

Lede case

Chrétien expected to keep cabinet minister

This report has 'wiggle room' to save MacAulay

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

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 is arrested more easily than a white person, and less likely to be released at the scene.

White	Black
70.5%	61.6%
15.6%	13.5%
2.3%	2.0%
held for bail	held for trial

IN SIDE
Editorial C2
Books C2
Entertainment C2
Home Team C2

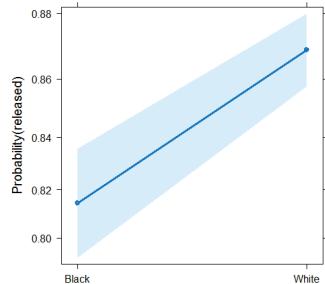
Marijuana data: Analysis

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**

33

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

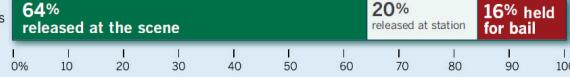
Same charge, different treatment

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.

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

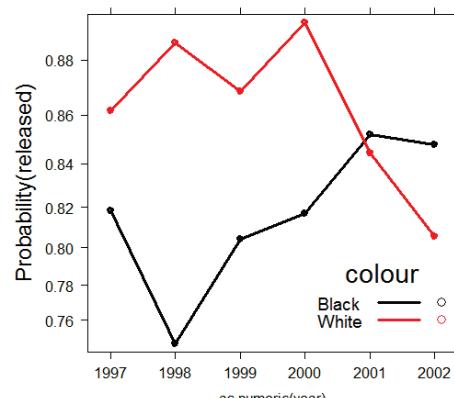
32

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, ...)
```

colour*year effect plot



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

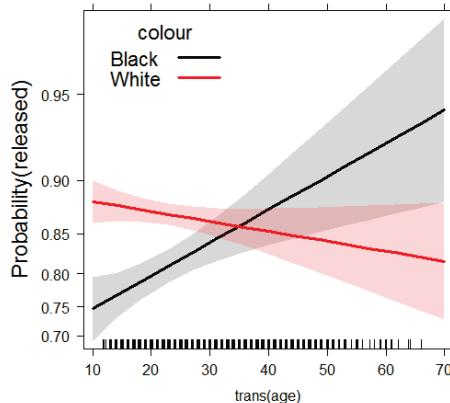
34

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

35

Traffic stops

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

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



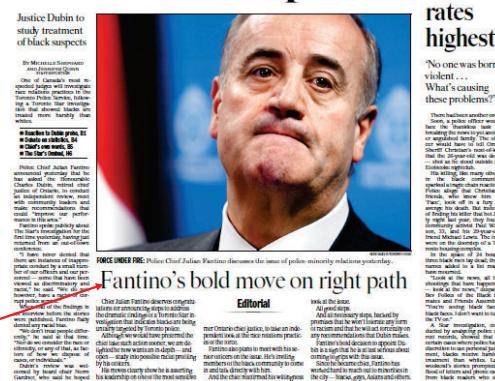
It's assumed random checks would generate a pattern of charges that mimics the racial distribution of drivers in society as a whole. So the rate at which minority drivers are charged is often used, in the U.S., as a bellwether for racial profiling.

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



37

But not without pushback

Goal? "Libel-chill to cause media to think twice..."

Judge: TOSSED!
"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

Police union sues Star over race-crime series



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

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

Danison said regardless of whether the association loses the case against the association, the paper may still be losing and that's not even trying."

When asked whether the lawsuit, regardless of its outcome or not, may encourage other police unions and media to think twice about criticizing institutions whose employees have power and influence, Danison said that concern "sounds speculative."

He added that the Star has gone well beyond fair comment in its series, to a degree that it may well tick to the quintessential core of what it is to be a police officer."

The Star's racial profiling series may align all key officers

consulting service, found the paper's libelous content.

In an in-depth analysis of the data, published in the paper's police section on Dec. 10, Peter, author of *The SAS System for Statistical Graphics and Visualizing Categorical Data*, and Jim Raskin, author of *Statistical and Graphical Techniques for Data Analysis and Visualization*, found that the Star's racial profiling series had a significant effect on the public perception of police stops. Civil libertarians and constitutional lawyers argued that the paper's effect on the public perception of police stops was unconstitutional.

Raskin said he hoped the Star will apologize for the error.

Following appeals and mediation, a copy of the database, excluding personal information, was given to the Star in May, 2002. It records more than 400,000 incidents in which an individual was stopped by police, from about 800,000 criminal and other charges. It served as the basis of the investigation into policing, crime and race.

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

38

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 @leerankin . 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?

39

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, ...)

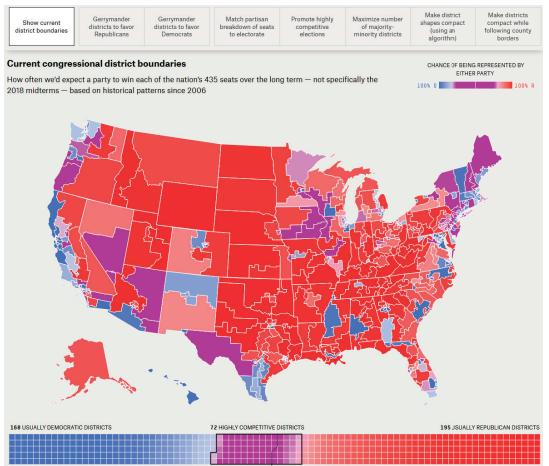
40

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?



41

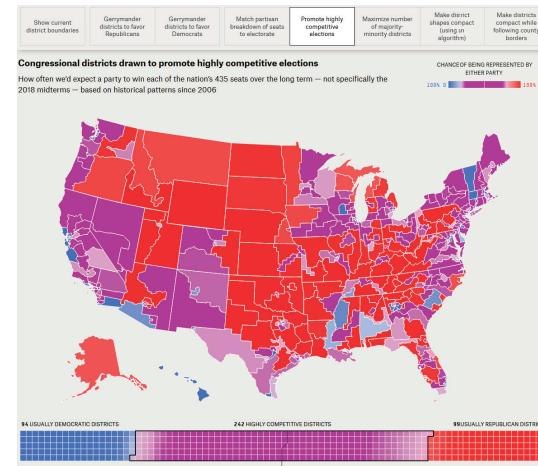
538: Atlas of Redistricting

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/>



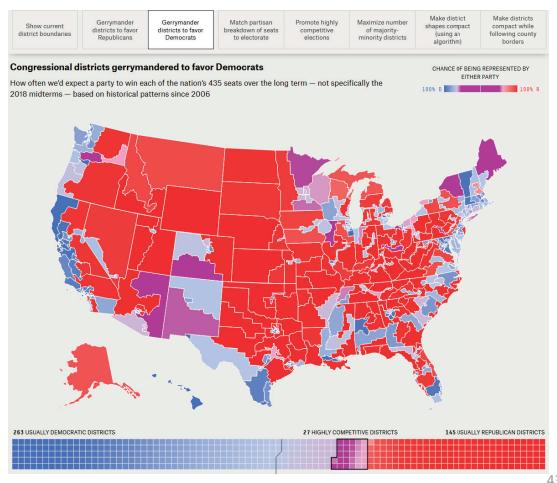
42

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



43

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

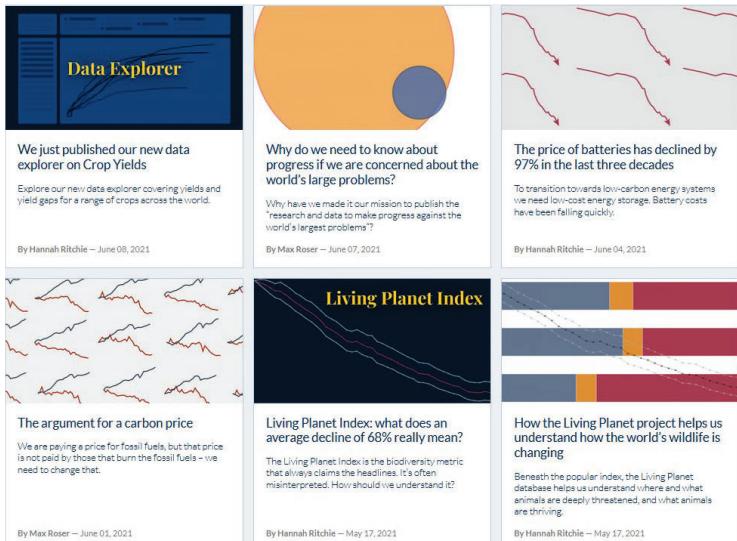
They relate these to demographic, policy, and urban/suburban issues using well-designed thematic maps.



44

Our World in Data

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

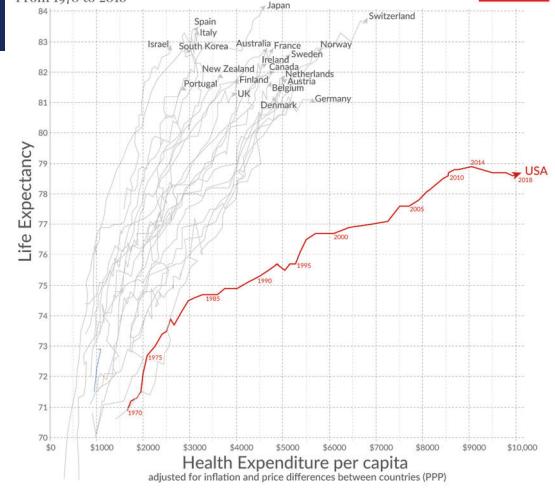


45

Our World in Data

Life expectancy vs. health expenditure

From 1970 to 2018



46

Financial Times COVID tracker

John Burn-Murdoch 
@jburnmurdoch

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

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

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

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



7:20 AM · Mar 23, 2020

47

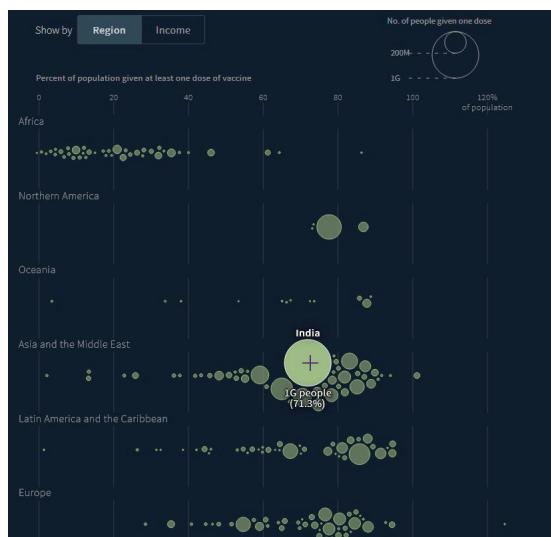
Reuters COVID tracker

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.



48

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.”

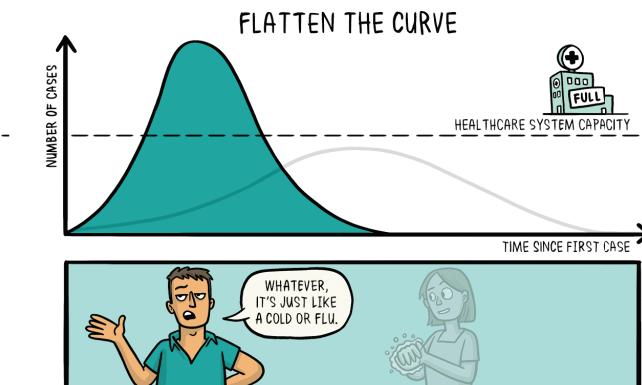
-- Prasanta Kumar Dutta

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

49

Flatten the curve

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



@SIOUXSIEW @TOTL @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>

50

- ✓ Animated, rather than separate panels
- ✓ Clear, memorable title
- ✓ Simple drawing rather than messy data
- ✓ Cases vs. capacity

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/>

51

Toronto Star: Trump Lies

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

claim	truth	inauguration
1 2017-01-20 Post-inauguration ...	The rain began right at the begin...	Inauguration
2 2017-01-20 Post-inauguration ...	Even the media said the c...	Inauguration
3 2017-01-21 Speech at Centr...	The major media reported that ...	Exaggerat...
4 2017-01-21 Speech at Centr...	"Honesty, it looked like a ...	Inauguration
5 2017-01-21 Speech at Centr...	"It was almost raining, the ...	Exaggerat...
6 2017-01-23 Private meeting ...	Neither of these claims is true. ...	Inauguration
7 2017-01-23 Interview with A...	"So a reporter for TIME m...	Media
8 2017-01-23 Interview with A...	Trump's numbers are well off. H...	Exaggerat...
9 2017-01-23 Interview with A...	"In terms of a total audience, ...	Election
10 2017-01-23 Interview with A...	"They" can mean anyone, but t...	Exaggerat...
11 2017-01-25 Interview with A...	"I think you're demeaning ...	Media
12 2017-01-25 Interview with A...	With regard to his speech ...	Exaggerat...
13 2017-01-25 Interview with A...	Most of the audience was inde...	CIA spe...
14 2017-01-25 Interview with A...	"No, no, you have to unde...	Media
15 2017-01-25 Interview with A...	Trump's victory was not close. L...	Trump White House
16 2017-01-25 Interview with A...	Regarding Healthcare.gov ...	Intellig...
17 2017-01-25 Interview with A...	Healthcare.gov did not cost \$5 ...	Exaggerat...
18 2017-01-25 Interview with A...	"Now you're telling me Pe... Trump was trying to use a 2012 ...	Election
	"Pew Fr... These large numbers of illegal v...	Media
	"There is no evidence of a signific... Mass fr...	Chicago
	"Regarding voting fraud: '... There is no evidence of a signific... Mass fr...	Voter fraud
	"Now you're telling me Pe... Obama ... no	NRA
	"Look, Barack Obama -- if ... This is a gross mischaracteriz... Obama ... no	Barack Obama
	"So, look, when President ... There were not only no homicid... Obama ... no	Crime

media voter fraud

52

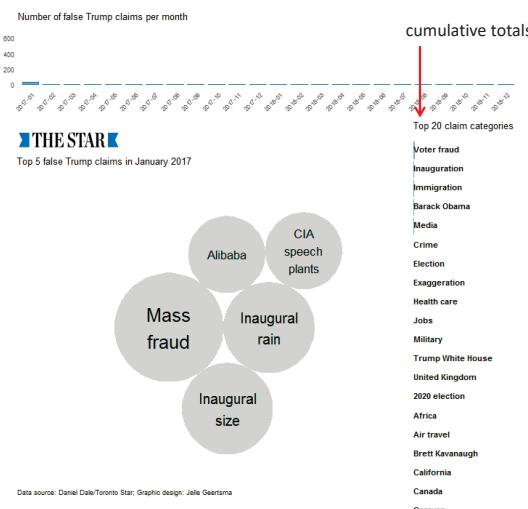
Jelle Geertsma: Trump Lies

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 gganimate package.

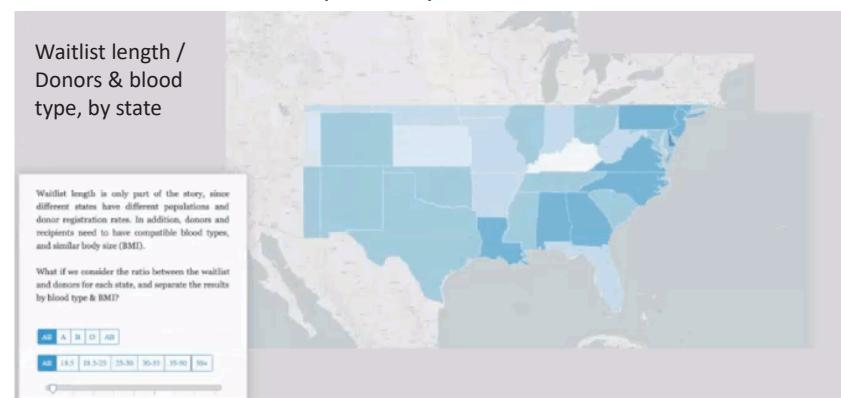
A TrumpLies package is in progress. **Ugh! Not so.**
<https://github.com/friendly/trumpLies>



53

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/>

54

Scrollytelling: The Impatient List

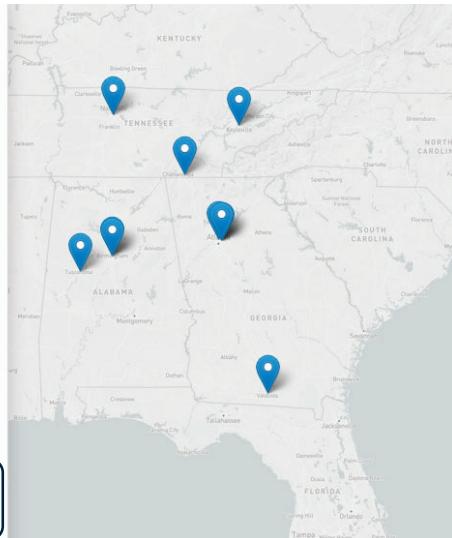
Since Kidney Transplant can be time sensitive, it is best for you to choose the ones adjacent to where you live. Please enter your zip code to view centers around you:

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.



55

Scrollytelling: The Impatient List



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

30318



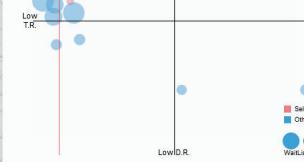
You'll also need to compare between centers around you for their death rates, survival rates, and waitlist waiting time, in order to find the optimal center to register as a candidate. How would you balance these different factors, when each choice you make might influence your chance of getting a transplant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

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

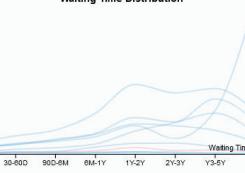
Death Rate (D.R.) Vs Transplant Rate (T.R.)

High D.R.

National Avg. D.R.



Waiting Time Distribution



56

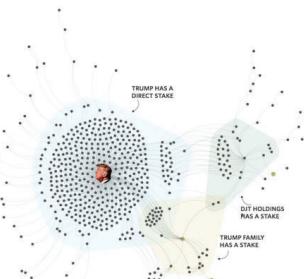
Drill-down to tell a story

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



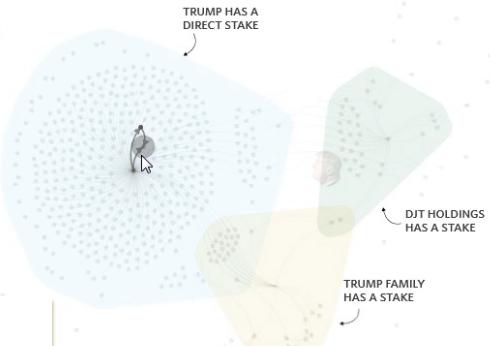
President Trump's network of conflicts of interest. Credit: WSJ.

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

57

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 \(May 2016\)](#)

58

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'

Friday May 8, 2009 BRITAIN'S BEST-SELLING QUALITY DAILY 10p No. 47,377

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

59

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](#)

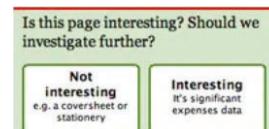
Please read our [privacy policy](#) to find out how we use your data. You must also read our [terms of service](#). By reviewing pages, you are agreeing that you have read the terms of service, and that you agree to them.

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

60

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.

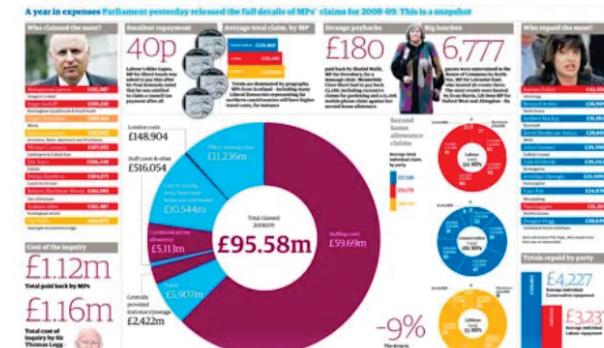


61

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



MP expenses visualised. Click on image for the full graphic

62

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



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

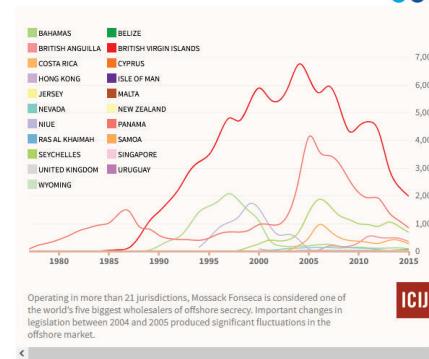
63

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

From the Caribbean to the Pacific: tax havens used by Mossack Fonseca



How could this graph be improved?

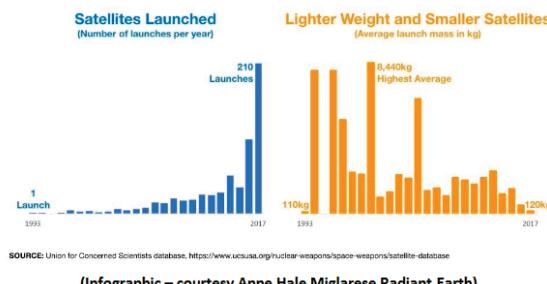
64

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.



65

Satellite Journalism: Smoking Guns

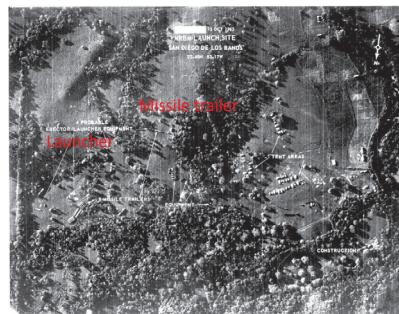
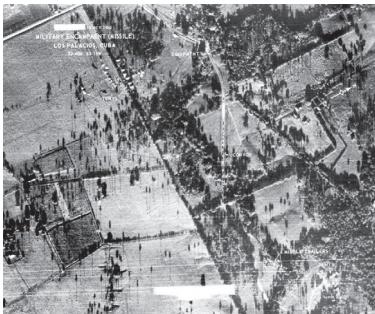
- Satellite imagery has provided ‘smoking-gun’ evidence for:
 - investigative reporting of human rights abuses (Htusan et al., 2015)
 - exposure of environmental destruction in remote locations (e.g., the Amazon: Weir, 2017)
 - before & after imagery of natural disasters (Murphy, 2018)
 - analyses of 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%20Satellite%20Journalism_1.pdf

66

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, ...

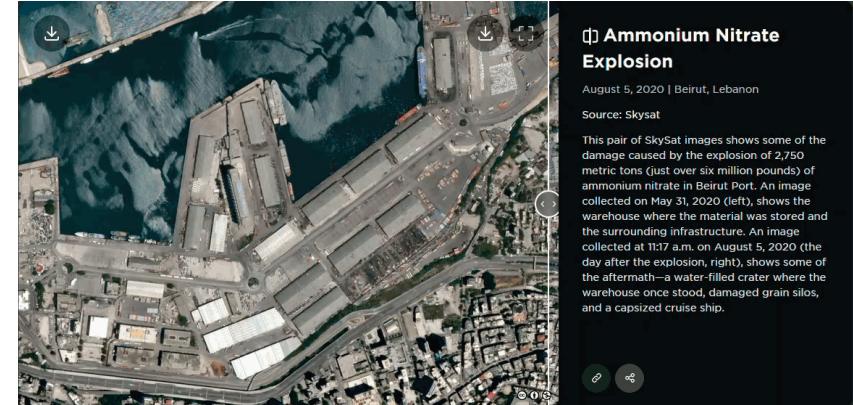
<https://nsarchive.gwu.edu/document/29153-oct-14-first-u-2-photographs-soviet-missiles-cuba-1962>

67

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.

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

68

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/>

69

Remote sensing tools for telling stories

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



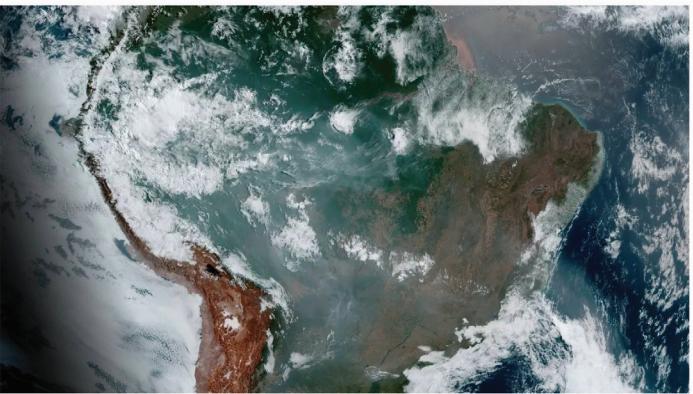
U.S. spy satellite image of a Soviet shipyard 3D rendering of Hurricane Linda (NOAA)

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

70

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

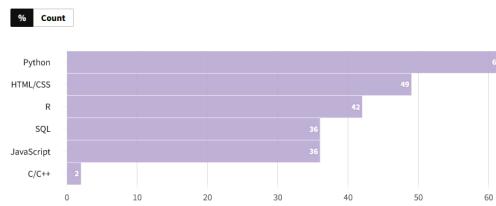


A view of wildfires burning in the Amazon rainforest.

Zoë Schlanger & Daniel Wolfe, [The fires in the Amazon were likely set intentionally](#), Quartz, Aug. 2019 71

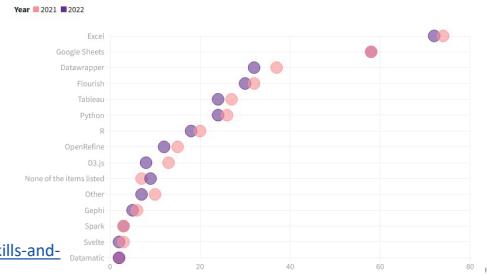
Software for DDJ

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?

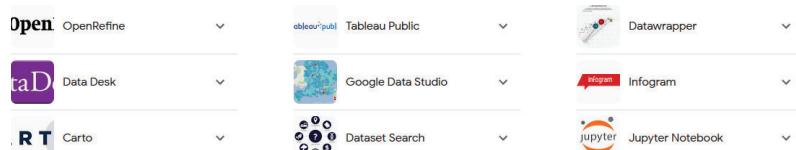


<https://datajournalism.com/survey/2022/skills-and-tools/>

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.

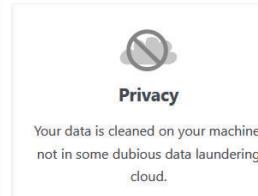
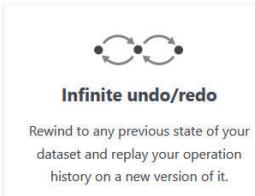
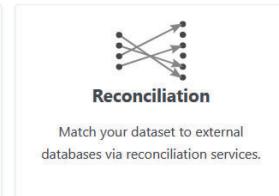
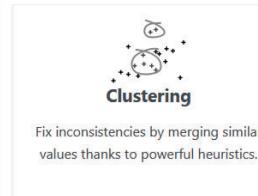
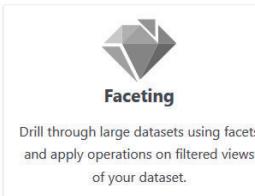
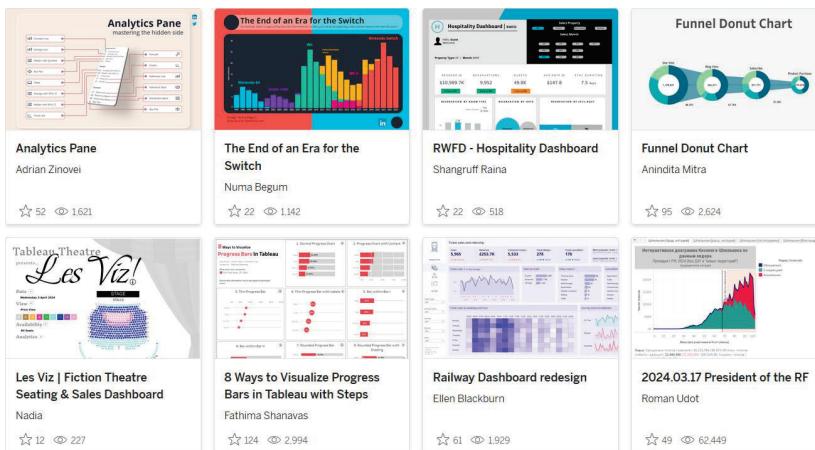


Tableau Public

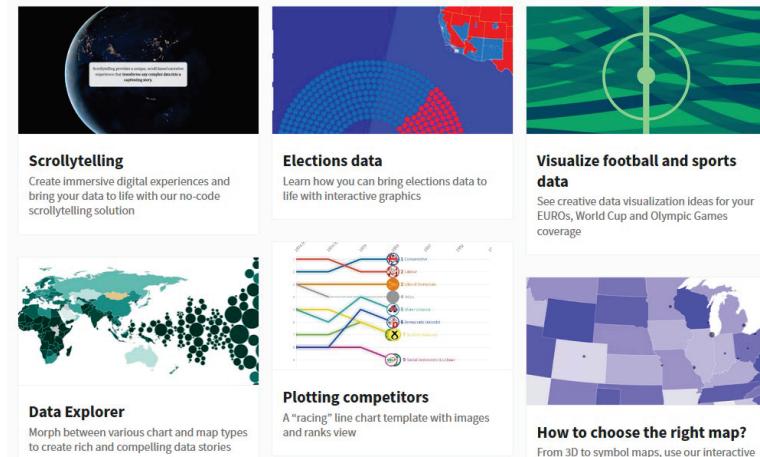
A free platform to explore, create, and publicly share data visualizations online.



75

Flourish

Popular DDJ tool for creating interactive, dynamic stories and datavis



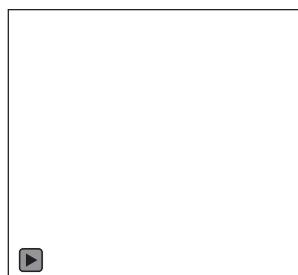
76

Flourish

Drag & drop data import



Animated graphics

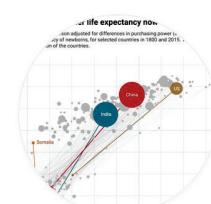


77

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



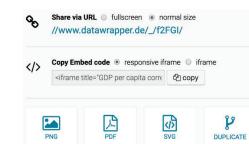
Charts

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.

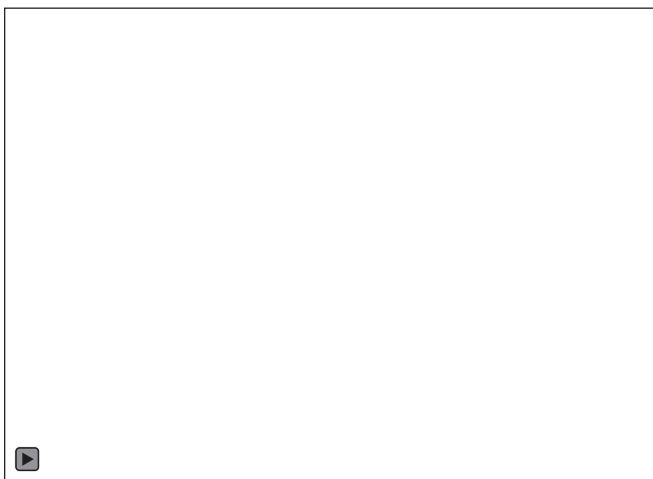


Tables

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.

78

Datawrapper demo



79

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>



80

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

81