

Data Journalism

Michael Friendly

Psych 6135

<http://euclid.psych.yorku.ca/www/psy6135>

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

Topics

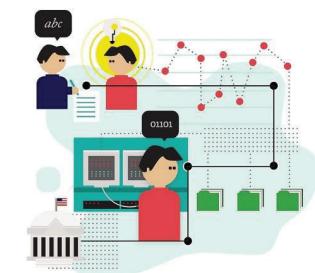
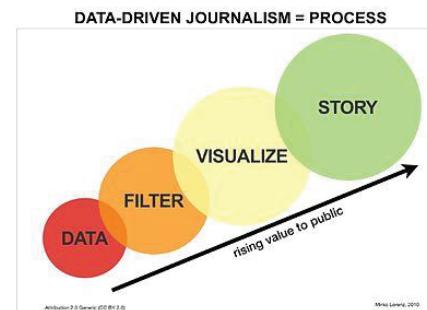
- What is DDJ?
- Some early examples
- Modern examples
 - election reporting
 - health statistics
 - social media
- Best examples??
- Now trending

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

Bill Dedman: *The Color of Money*

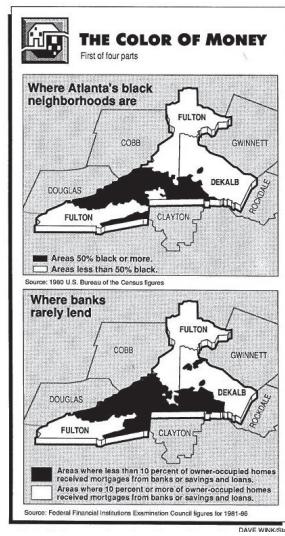
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

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

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

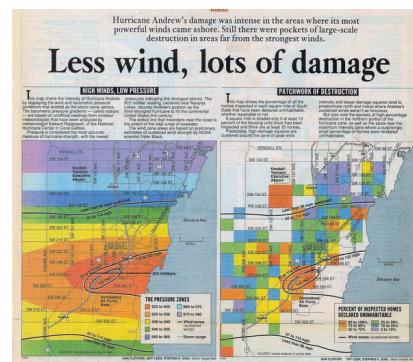
It helped establish the idea of computer-assisted reporting and the importance of compelling visualizations in telling the story.

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



Hurricane Andrew

- The report in the *Miami Herald* contained many detailed analysis maps
- There was compelling evidence for the cause of destruction.
- The combination of investigative reporting, analysis and graphics was seen as powerful



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/>

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.

8

CNN Magic Wall

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

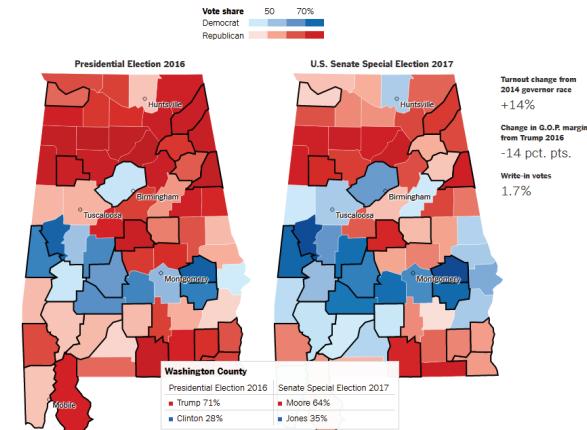


Video: <https://www.youtube.com/watch?v=rpLHnK-MSy4>

9

Election reporting: NYT

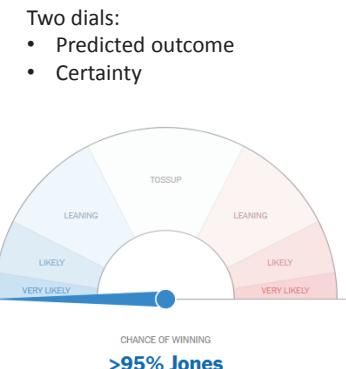
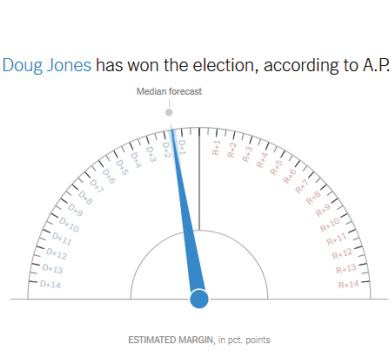
In the special election for the Alabama U.S. Senate, Dec. 12, 2017, *The New York Times* used dynamic, interactive maps to compare the vote to the 2016 Presidential election



10

Election reporting: Dashboards

Throughout the evening of Dec. 12, the NYT dynamically updated dashboard displays of the election returns and model-based predictions of the winner



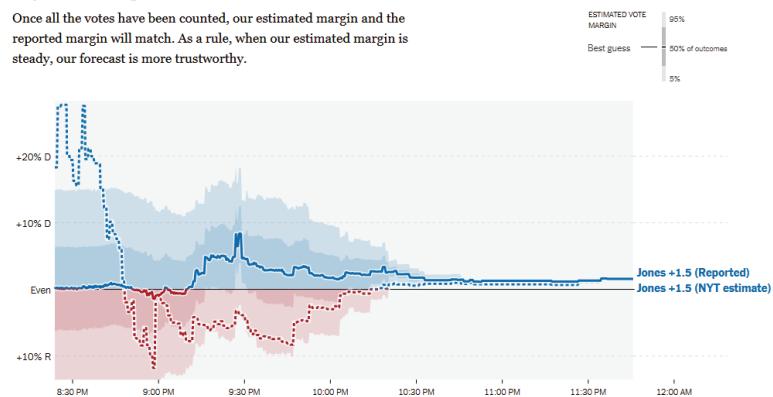
11

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.



12

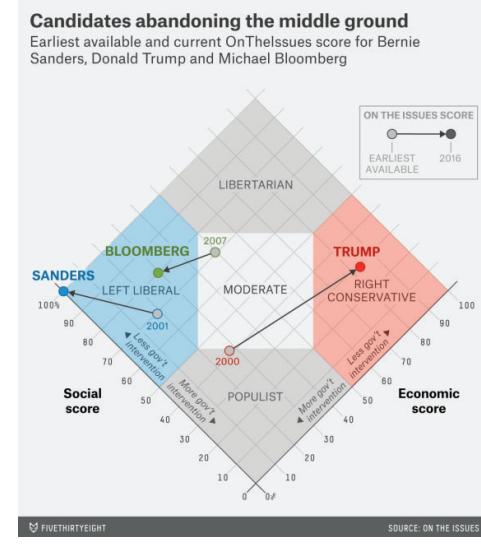
Novel graphic designs

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



13

Measles and vaccines

Visualizing the impact of health policy interventions

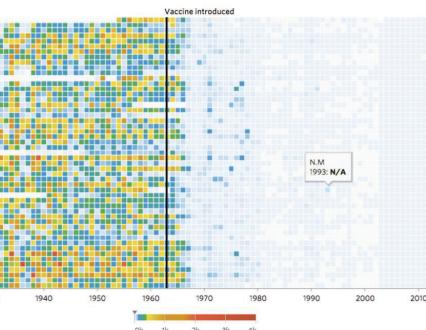
In 2015 Tynan DeBold & Dov Friedman in the *Wall Street Journal* tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases

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/>

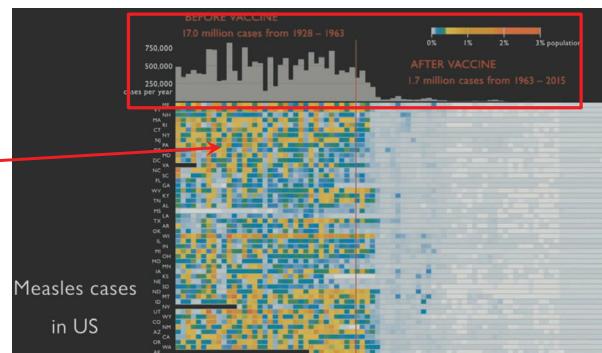
14

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



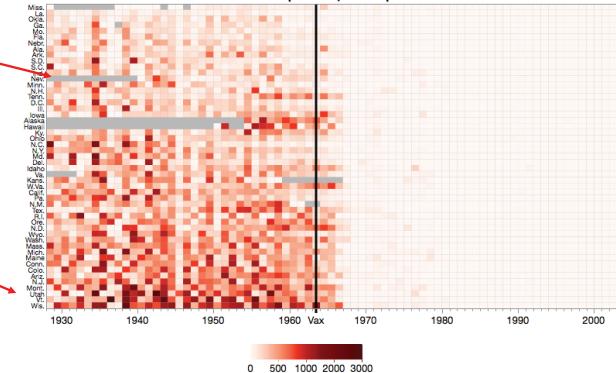
15

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

Measles Cases per 100,000 Population

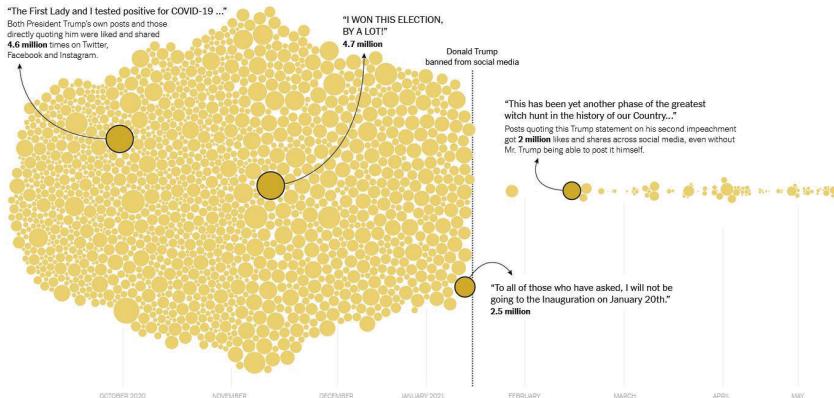


16

Before – After Circle Chart

What Happened When Trump Was Banned on Social Media?

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



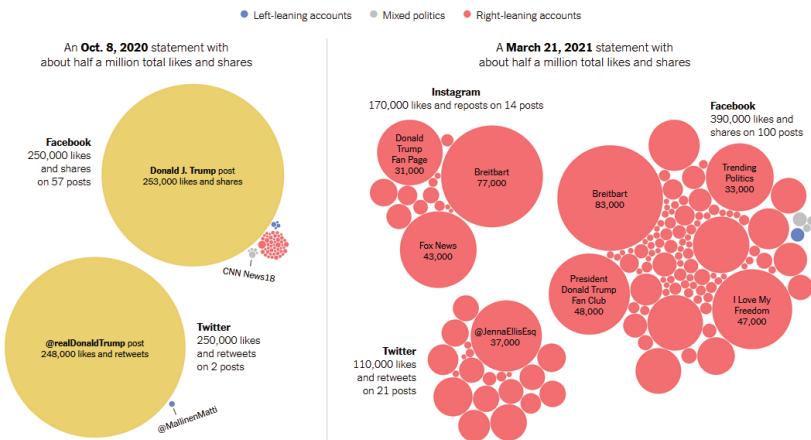
<https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>

17

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



18

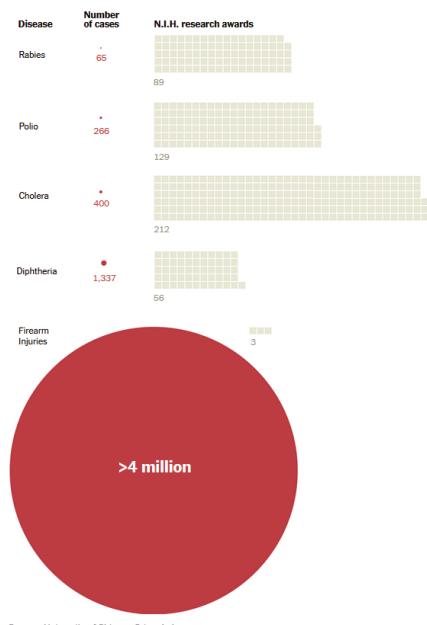
Health & research funding

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

19

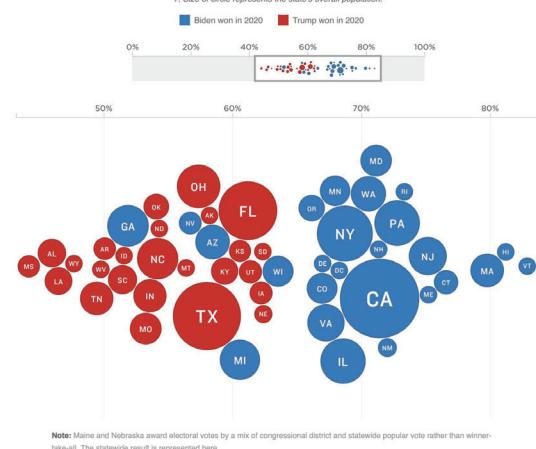
Votes for vs. COVID vaccinations

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

circle size ~ state pop.

States Biden Won Have The Highest Adult Vaccination Rates

Percentage of state's population age 18 and older that has received *at least one dose* of a COVID-19 vaccine as of June 7. Size of circle represents the state's overall population.

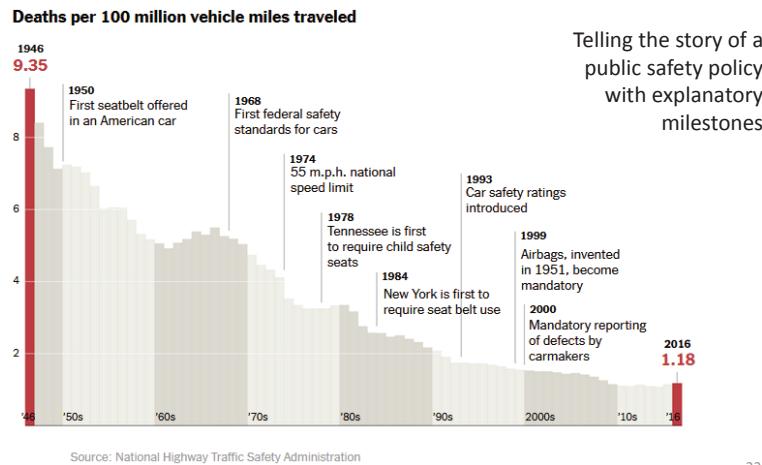


From: <https://twitter.com/kihealy/status/1402691186826039301>

20

Seatbelts

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



22

Best examples of DDJ? Ask Twitter



Michael Friendly @datavisFriendly · 22h

1/ Thinking about an upcoming class in my #psy6135 #datavis course on the topic #datajournalism, with a guest star @jlerankin . 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...

TPTimes's tampabay.com/projects/2015/...

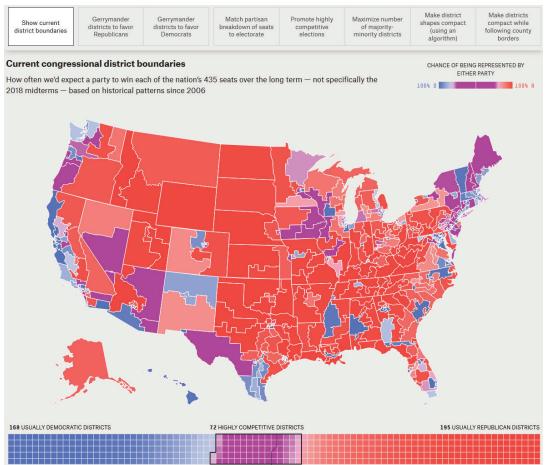
23

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?



24

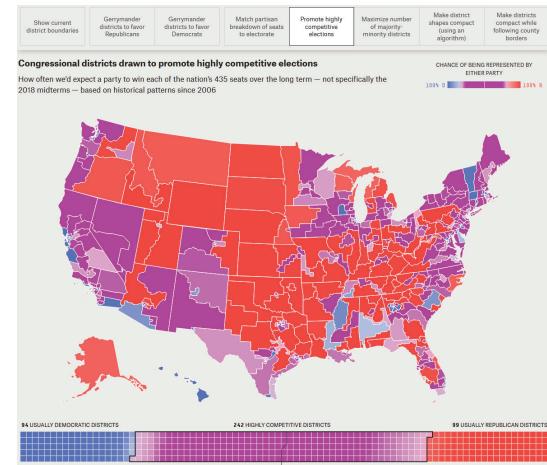
538: Atlas of Redistricting

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/>



25

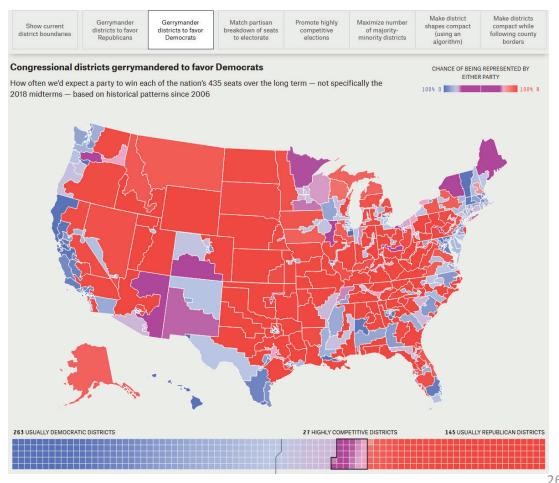
538: Atlas of Redistricting

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

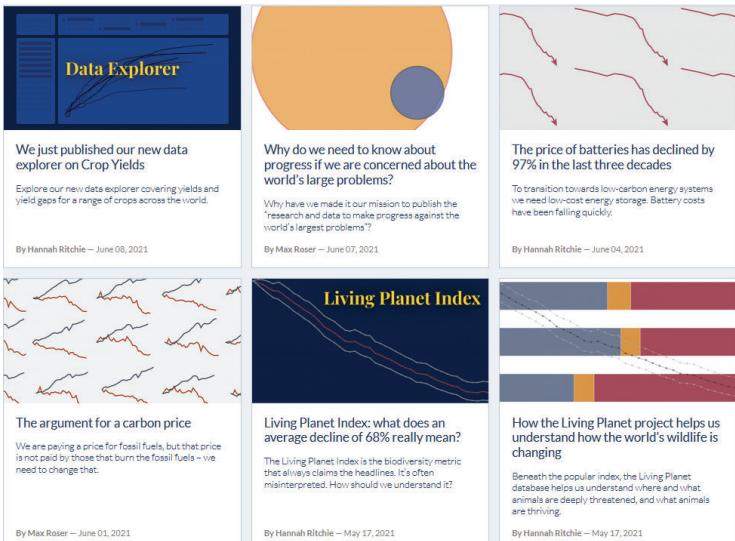
From:
<https://projects.fivethirtyeight.com/redistricting-maps/>



26

Our World in Data

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

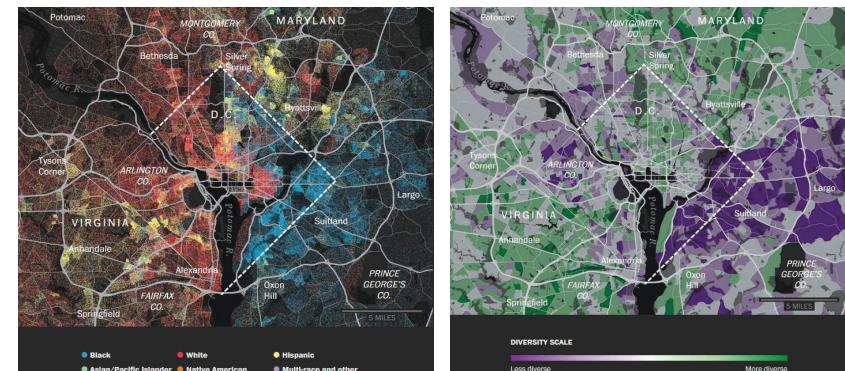


28

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.



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

27

Financial Times COVID tracker

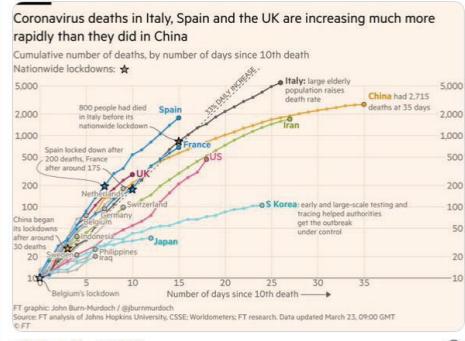
John Burn-Murdoch @jburnmurdoch

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

29

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/>

30

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 Weather
2 2017-01-20 Post-inauguration ...	The major media reported that ...	Inauguration Exaggerat
3 2017-01-21 Post-inauguration ...	The crowd, which may not have ...	Inauguration Exaggerat
4 2017-01-21 Speech at Centr...	Neither of these claims is true ...	Inauguration NA
5 2017-01-21 Speech at Centr...	Trump's numbers are well off. H...	Media Exaggerat
6 2017-01-23 Private meeting ...	This claim, also reported by mu...	Voter fraud Election
7 2017-01-25 Interview with A...	"They" can mean anyone, but it...	Inauguration Exaggerat
8 2017-01-25 Interview with A...	Trump's victory was not close. L...	Election Exaggerat
9 2017-01-25 Interview with A...	I think you're demeaning ...	Media Exaggerat
10 2017-01-25 Interview with A...	With regard to his speech ...	CIA spe... no
11 2017-01-25 Interview with A...	Most of the audience was inde... CIA spe... no	Trump White House US intellig
12 2017-01-25 Interview with A...	Regarding Healthcare.gov ...	Healthc... no
13 2017-01-25 Interview with A...	Mueller called the author ...	Grovell... no
14 2017-01-25 Interview with A...	Groveling means to draw bac...	Election
15 2017-01-25 Interview with A...	Now you're telling me Pe... Trump was trying to use a 2012 ...	Pew fra... no
16 2017-01-25 Interview with A...	Regarding his false claim ...	These large numbers of illegal v...
17 2017-01-25 Interview with A...	No, look, when President ...	Mass fr... yes
18 2017-01-25 Interview with A...	So, look, when President ...	Obama ... no
		Trump/Barack Obama
		Chicago
		NA
		Voter frau
		Crime

media

voter fraud

31

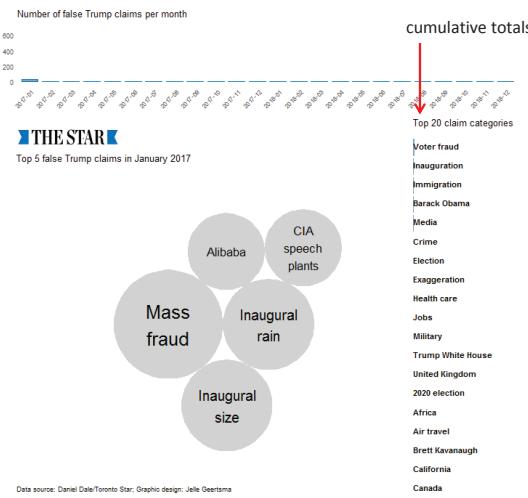
Jelle Geertsma: Trump Lies

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.

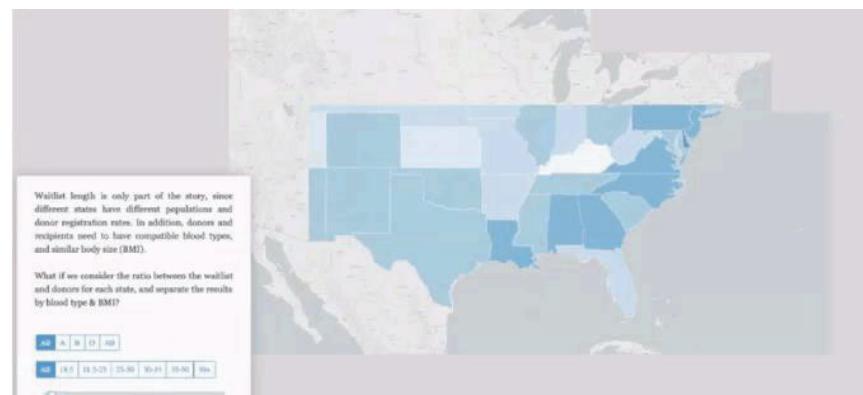
A TrumpLies package is in progress. **Ugh! Not so.**



32

Scrollytelling: The Impatient List

What it's like waiting for a kidney transplant in the U.S.



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

33

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 focus for data visualization
 - New online forms: scrollytelling
- Open data
 - Maps-GIS (Google, Open Street maps)
 - Gov't data: census, health, crime, ...
 - NGO initiatives: Our World in Data
- Software infrastructure
 - Javascript as base for online, interactive methods
 - Libraries, frameworks building on this (D3.js, Processing, ...)