Data Journalism

Michael Friendly
Psych 6135

http://euclid.psych.yorku.ca/www/psy6135
Topics

• What is DDJ?
• Some early examples
• Modern examples
  ▪ election reporting
  ▪ health statistics
  ▪ social media
• Best examples??
• Now trending
What is data journalism?

• **Data journalism** reflects the increased role of numerical data for reporting in the digital era. Data driven or DDJ more useful
  ▪ Reflects increased interaction between journalists and {visualization design, computer science and statistics}.

• Aspects:
  ▪ **Database analysis**: FOI requests on topics like crime, racial discrimination, judicial sentencing, toxic pollution, ...
  ▪ **Visualization**: infographic→data vis., emphasis on the **story**
  ▪ Increasing use & development of data vis. **software tools**
    • analysis: R, python, ...
    • visualization libraries: R, D3, javascript, ...
  ▪ Increasing use of **dynamic updating** of data sources
    • election polls / prediction
    • disease tracking
What is data journalism?

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

Two motivational infographics of #DDJ as a process
❖ How does it actually work?
❖ What makes it successful?
Bill Dedman 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/](http://powerreporting.com/color/) for all the articles and images
Hurricane Andrew

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

Source: https://www.flickr.com/photos/juggernautco/sets/72157607210036175/
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.
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.
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
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?
Throughout the evening of Dec. 12, the NYT dynamically updated dashboard displays of the election returns and model-based predictions of the winner.

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

Two dials:
- Predicted outcome
- Certainty

**Jones +1.5**

>95% Jones
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.
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
Visualizing the impact of health policy interventions

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

The long time series ~70 years made this work.

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

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

Making the message more explicit

What you should try to show here in an Info graphic is the **overall impact** of vaccination on measles
Ed Tufte did this by adding a histogram at the top showing total # of cases by year

With the overall impact made clear, details about individual states can now be explored
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://community.jmp.com/t5/JMP-Blog/Graph-makeover-Measles-heat-map/ba-p/30550
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

“The First Lady and I tested positive for COVID-19...”
Both President Trump’s own posts and those
directly quoting him were liked and shared
4.6 million times on Twitter,
Facebook and Instagram.

“I WON THIS ELECTION,
BY A LOT!”
4.7 million

Donald Trump
banned from social media

“This has been yet another phase of the greatest
witch hunt in the history of our Country...”
Posts quoting this Trump statement on his second impeachment
got 2 million likes and shares across social media, even without
Mr. Trump being able to post it himself.

“To all of those who have asked, I will not be
going to the Inauguration on January 20th.”
2.5 million

Before – After: Who pushed the lies?

How/Who Spread Trump statements before and after the ban

An **Oct. 8, 2020** statement with about half a million total likes and shares

- **Facebook**: 250,000 likes and shares on 57 posts
- **Twitter**: @realDonaldTrump post 248,000 likes and retweets on 2 posts

A **March 21, 2021** statement with about half a million total likes and shares

**Instagram**: 170,000 likes and reposts on 14 posts
- Donald Trump Fan Page 31,000
- Breitbart 83,000
- Breitbart 77,000
- Fox News 43,000
- Trending Politics 33,000
- President Donald Trump Fan Club 43,000
- I Love My Freedom 47,000

**Facebook**: 390,000 likes and shares on 100 posts

**Twitter**: 110,000 likes and retweets on 21 posts
- @JennaEllisEq 37,000

Posts referencing each Trump statement, sized by total number of likes and shares
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?
Votes for vs. COVID vaccinations

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

circle size ~ state pop.

From: https://twitter.com/kjhealy/status/1402691186826039301
New York Times: An annotated chart of vehicle deaths in relation to seatbelts

Telling the story of a public safety policy with explanatory milestones
Best examples of DDJ? Ask Twitter

**Michael Friendly** @datavisFriendly · 22h
1/ Thinking about an upcoming class in my #psyc6135 #datavis course on the topic #datajournalism, with a guest star @JLeeRankin. It got me thinking of the idea of "Great moments in Data Journalism": investigative reporting, incisive analyses, compelling graphic displays.

**Michael Friendly** @datavisFriendly · 22h
2/ I would love ❤️ for nominations on this topic. @albertocairo and others have done some of this, but perhaps this is a topic on which others could contribute their nominations for "Great moments..."

**Alberto Cairo** @albertocairo · 22h
Twitter #ddj and news #dataviz, contribute with nominations.

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

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

TPTimes's tampabay.com/projects/2015/...
There’s a lot of complaining about gerrymandering, but what should US Congressional districts look like? What would they look like to serve various goals?

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

Why do I find this app appealing?

From: https://projects.fivethirtyeight.com/redistricting-maps/
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/
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/
Aaron Williams & Armand Emamdjomeh at the Washington Post studied changes in racial composition and diversity in U.S. cities over 30+ years. They relate these to demographic, policy, and urban/suburban issues using well-designed thematic maps.

Our World in Data

An impressive range of social & environmental issues presented by Our World in Data
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
Donald Trump has said 4557 false things as U.S. president (as of 3-13-2019)
Every false claim is documented: place, claim, truth, “slug”, categories

<table>
<thead>
<tr>
<th>number</th>
<th>date</th>
<th>place</th>
<th>claim</th>
<th>intact</th>
<th>slug</th>
<th>repeat</th>
<th>source</th>
<th>category1</th>
<th>category2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2017-01-20</td>
<td>Post-inauguration</td>
<td>&quot;I looked at the rain which...&quot;</td>
<td>The rain began right at the beginning</td>
<td>Inauguration</td>
<td>Yes</td>
<td><a href="https://www.thestar.com/news-politics/">https://www.thestar.com/news-politics/</a></td>
<td>Inauguration</td>
<td>Weather</td>
</tr>
<tr>
<td>2</td>
<td>2017-01-20</td>
<td>Post-inauguration</td>
<td>&quot;Even the media said the c...&quot;</td>
<td>The major media reported that...</td>
<td>Inauguration</td>
<td>Yes</td>
<td><a href="https://www.thestar.com/news-politics/">https://www.thestar.com/news-politics/</a></td>
<td>Inauguration</td>
<td>Exaggerated</td>
</tr>
<tr>
<td>3</td>
<td>2017-01-21</td>
<td>Speech at Center</td>
<td>&quot;Honestly, it looked like a...&quot;</td>
<td>The crowd, which may not have...</td>
<td>Inauguration</td>
<td>Yes</td>
<td><a href="https://www.youtube.com/watch?v=">https://www.youtube.com/watch?v=</a></td>
<td>Inauguration</td>
<td>Exaggerated</td>
</tr>
<tr>
<td>4</td>
<td>2017-01-21</td>
<td>Speech at Center</td>
<td>&quot;It was almost raining, the...&quot;</td>
<td>Neither of these claims is true...</td>
<td>Inauguration</td>
<td>Yes</td>
<td><a href="https://www.youtube.com/watch?v=">https://www.youtube.com/watch?v=</a></td>
<td>Inauguration</td>
<td>Exaggerated</td>
</tr>
<tr>
<td>5</td>
<td>2017-01-21</td>
<td>Speech at Center</td>
<td>&quot;So a reporter for TIME ma...&quot;</td>
<td>Trump's numbers are well off.</td>
<td>TIME.co...</td>
<td>Yes</td>
<td><a href="https://www.youtube.com/watch?v=">https://www.youtube.com/watch?v=</a></td>
<td>Media</td>
<td>Exaggerated</td>
</tr>
<tr>
<td>6</td>
<td>2017-01-23</td>
<td>Private meeting</td>
<td>Trump told Congressional...</td>
<td>This claim, also reported by nu...</td>
<td>Mass fr...</td>
<td>Yes</td>
<td><a href="http://www.foxnews.com/politics/voter-fraud">http://www.foxnews.com/politics/voter-fraud</a></td>
<td>Election</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;They can mean anyone, but...&quot;</td>
<td>Inaugur...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/inauguration">http://abcnews.go.com/politics/inauguration</a></td>
<td>Exaggerated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;No, no, you have to unde...&quot;</td>
<td>Trump's victory was not close...</td>
<td>Electio...</td>
<td>Yes</td>
<td><a href="http://abcnews.go.com/politics/election">http://abcnews.go.com/politics/election</a></td>
<td>Exaggerated</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;I think you're demeaning...&quot;</td>
<td>Saying &quot;you here, Trump won...&quot;</td>
<td>Media...</td>
<td>Yes</td>
<td><a href="http://abcnews.go.com/politics/media">http://abcnews.go.com/politics/media</a></td>
<td>Exaggerated</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>With regard to his speech...</td>
<td>Most of the audience was ind...</td>
<td>CIA spe...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/trump-white-house">http://abcnews.go.com/politics/trump-white-house</a></td>
<td>US intelligence</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>Regarding Healthcare.gov...</td>
<td>Healthcare.gov did not cost SS...</td>
<td>Health...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/">http://abcnews.go.com/politics/</a></td>
<td>Exaggerated</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;I called the author...&quot;</td>
<td>Grovelling means &quot;to draw bac...</td>
<td>Grovelli...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/">http://abcnews.go.com/politics/</a></td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;Now you're telling me Fe...&quot;</td>
<td>Trump was trying to use a 2012...</td>
<td>Pew fr...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/">http://abcnews.go.com/politics/</a></td>
<td>Election</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>Regarding his false claim...</td>
<td>These large numbers of illegal...</td>
<td>Mass fr...</td>
<td>Yes</td>
<td><a href="http://abcnews.go.com/politics/">http://abcnews.go.com/politics/</a></td>
<td>Exaggerated</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>Regarding voting fraud; &quot;...&quot;</td>
<td>There is no evidence of a signifi...</td>
<td>Mass fr...</td>
<td>Yes</td>
<td><a href="http://abcnews.go.com/politics/voter-fraud">http://abcnews.go.com/politics/voter-fraud</a></td>
<td>Chicago</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>Regarding voting fraud; &quot;...&quot;</td>
<td>There is no evidence of a signifi...</td>
<td>Mass fr...</td>
<td>Yes</td>
<td><a href="http://abcnews.go.com/politics/voter-fraud">http://abcnews.go.com/politics/voter-fraud</a></td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;Look, Barack Obama -- if...&quot;</td>
<td>This is a gross mischaracteriz...</td>
<td>Obama...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/barack-obama">http://abcnews.go.com/politics/barack-obama</a></td>
<td>Crime</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2017-01-25</td>
<td>Interview with ...</td>
<td>&quot;So, look, when President...&quot;</td>
<td>There were not only no homicid...</td>
<td>Obama...</td>
<td>No</td>
<td><a href="http://abcnews.go.com/politics/barack-obama">http://abcnews.go.com/politics/barack-obama</a></td>
<td>Crime</td>
<td></td>
</tr>
</tbody>
</table>
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.
Scrollytelling: The Impatient List

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

From https://xujunjiejack.github.io/
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, ...)