36. The #ddj Hashtag on Twitter

Eunice Au and Marc Smith

Abstract

How we used the social network analysis and visualization package NodeXL to examine what the global data journalism community tweets about.

Keywords: #ddj, Twitter, social network analysis, data journalism, social media, data analysis

Picking a single term to track the data journalism field is not easy. Data journalists use a myriad of hashtags in connection with their work, such as #datajournalism, #ddj, #dataviz, #infographics, and #data. When the Global Investigative Journalism Network (GIJN)—an international association of investigative journalism organizations that supports the training and sharing of information among investigative and data journalists—first started to report on conversations around data journalism on Twitter six years ago, the most popular hashtag appeared to be #ddj (data-driven journalism).¹

The term “data-driven journalism” itself is controversial as it can be argued that journalism is not driven by data; data merely informs, or is a tool used for journalism. Data consists of structured facts and statistics that require journalists to filter, analyze and discover patterns in order to produce stories. Just as one would not call a profile piece “interview-driven journalism” or an article based on public documents “document-driven journalism,” great data journalism stories use data as only one of their components.

¹ <https://gijn.org/>

the patterns of interconnection that emerge from activities such as replying, @mentioning and retweeting. These patterns highlight key people, groups and topics being discussed.

As an international investigative journalism organization, GIJN is always looking for ways to raise awareness about what is happening in the fields of investigative and data journalism. When GIJN's executive director, David Kaplan, saw Smith's network graph, he proposed to use the map to produce a weekly Top 10 #ddj to showcase popular and interesting examples of data journalism. (He and Smith also tried a weekly round-up of investigative journalism, but no single hashtag came close to doing the job that #ddj does for data journalism.) Although GIJN follows the network graph's suggested findings closely, some human curation is necessary to eliminate duplicates and to highlight the most interesting items.

Since the birth of the series, we have assembled more than 250 snapshots of the data journalism community's discussions featuring the #ddj hashtag over the past six years (GIJN, n.d.). The series now serves as a good quick summary for interested parties who cannot follow every #ddj tweet. Our use of the term "snapshot" is not simply a metaphor. This analysis gives us a picture of the data journalism Twitter community, in the same way that photojournalism depicts real crowds on the front pages of major news outlets.

The Evolution of #ddj Twitter Traffic

To get a sense of how Twitter traffic using #ddj has evolved, we did a very basic and rough analysis of the #ddj data we collected from 2014 to 2019. We selected a small sample of eight weeks in February and March from each of the six years, or 48 weeks. There was a variety of content being shared and engaged with and the most popular items included analysis and think pieces, awards, grants, events, courses, jobs, tools, resources, and investigations. The types of content shared remained consistent over the years.

In 2014, we saw articles that discussed a burgeoning data journalism field. This included pieces arguing that data journalism is needed because it fuels accountability and insights (Howard, 2014), and predicting that analyzing data is the future for journalists (Arthur, 2010). In later years, we observed new topics being discussed, such as artificial intelligence, massive data leaks and collaborative data investigations. There were also in-depth how-to pieces, where data journalists started offering insights into their data journalism processes (Grossenbacher, 2019) and sharing how to best utilize databases (Gallego, 2018), rather than debating whether the media

Figure 36.2. Example of top influencers (from January 1, 2018 to August 13, 2018). Source: NodeXL.

industry should incorporate data journalism into its newsrooms. We also noticed that among the investigations shared there were often analyses of elections, immigration, pollution, climate and football.

GIJN’s weekly #ddj round-up not only highlights the most popular tweets and URLs, but also lists the central participants of the #ddj discussion. Some of the usual suspects at the centre of #ddj discussions include data journalism experts Edward Tufte, Alberto Cairo, Martin Stabe, Nate Silver and Nathan Yau, along with data teams from Europe and North America, including those at *Le Telegramme*, *Tages-Anzeiger*, *Berliner Morgenpost*, *FiveThirtyEight*, the *Financial Times*, and The Upshot from *The New York Times*. Their work can at times be educational and inspiring and trigger further debate. The data journalism community can also take advantage of and network with these influencers.

A number of other hashtags often accompany #ddj, as Connected Action’s mapping reveals, allowing members of the community to seek out similar stories.

By far, the most common hashtags to appear alongside #ddj were #dataviz, #visualization, #datajournalism, #opendata, #data and #infographics. This signals to us that those who are in this field particularly care not just about the availability of public data, but also the way in which data is creatively presented and visualized for readers.

However, the NodeXL #ddj mapping is by no means representative of the entire field as it analyzes only people who tweet. Furthermore, those who

Top Hashtags in Tweet in Entire Graph:

[\[22540\] ddj](#)
[\[6765\] dataviz](#)
[\[1783\] datajournalism](#)
[\[1578\] opendata](#)
[\[1517\] vg](#)
[\[1253\] data](#)
[\[1080\] infographics](#)
[\[589\] opensource](#)
[\[541\] datajournalismawards](#)
[\[534\] journalism](#)

Figure 36.3. Example of top related hashtags (from January 1, 2018 to August 13, 2018). Source: NodeXL.

generally have more followers on Twitter and garner more retweets tend to feature more prominently in our round-up.

We have also noticed that the majority of the top tweets usually come from Europe and the Americas, particularly Germany and the United States, with some smatterings of tweets from Asia and Africa. This could be due to the skew of the user base on Twitter, because other regions have relatively less robust data journalism communities, or because data journalism communities in other regions do not organize through the same Twitter hashtags or do not organize on Twitter at all.

Over the past year, we observed that some work by prominent data journalism organizations that was widely shared on Twitter did not appear in our network graph. This could possibly be due to people not using the hashtag #ddj when tweeting the story, or using other hashtags or none at all. We suspect that Twitter's expansion of the tweet character count from 140 to 280 in November 2017 might also have helped people to choose lengthier hashtags such as #datajournalism.

Fun #ddj Discoveries

While what we find is often powerful journalism and beautiful visualizations, sometimes it is also just plain funny. By way of conclusion, we briefly discuss some of the more entertaining items we have discovered using the #ddj hashtag in the past year.

In an adorable and clever visual essay, Xaquín G. V. (2017) showed what people in different countries tend to search for the most when they want to fix something. In many warmer countries, it is fridges, for North Americans and East Asians it is toilets, while people in northern and eastern Europe

seem to need information on how to fix light bulbs. Next, a chart, found among the Smithsonian's Sally L. Steinberg Collection of Doughnut Ephemera, argues that the size of the doughnut hole has gradually shrunk over the years (Edwards, 2018). In a different piece, graphic designer Nigel Holmes illustrated and explained oddly wonderful competitions around the world, from racing snails to carrying wives, in a book called *Crazy Competitions* (Yau, 2018).

In another piece in our collection, women worldwide already know that the pockets on women's jeans are impractically tiny, and *The Pudding* has provided the unequivocal data and analysis to prove it (Diehm & Thomas, 2018). Finally, is there such a thing as peak baby-making seasons? An analysis by Visme of United Nations' data on live births seems to suggest so. They found a correlation between three different variables: The top birth months, seasons of the year and the latitude of the country (distance from the equator) that may have influence on mating rhythms in different countries (Chibana, n.d.).

Works Cited

- Arthur, C. (2010, November 22). Analysing data is the future for journalists, says Tim Berners-Lee. *The Guardian*. <https://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee>
- Chibana, N. (n.d.). Do humans have mating seasons? This heat map reveals the surprising link between birthdays and seasons. *Visual Learning Center by Visme*. <https://visme.co/blog/most-common-birthday/>
- Diehm, J., & Thomas, A. (2018, August). Pockets. *The Pudding*. <https://pudding.cool/2018/08/pockets/>
- Edwards, P. (2018, June 1). Have donut holes gotten smaller? This compelling vintage chart says yes. *Vox*. <https://www.vox.com/2015/9/20/9353957/donut-hole-size-chart>
- Gallego, C. S. (2018, January 23). How to investigate companies found in the offshore leaks database. ICIJ. <https://www.icij.org/blog/2018/01/investigate-companies-found-offshore-leaks-database/>
- GIJN. (n.d.). Top 10 in data journalism archives. Global Investigative Journalism Network. <https://gijn.org/series/top-10-data-journalism-links/>
- Grossenbacher, T. (2019, March 8). (Big) data journalism with Spark and R. <https://timogrossenbacher.ch/2019/03/big-data-journalism-with-spark-and-r/>
- Howard, A. (2014, March 3). Data-driven journalism fuels accountability and insight in the 21st century. *TechRepublic*. <https://www.techrepublic.com/article/data-driven-journalism-fuels-accountability-and-insight-in-the-21st-century/>

- Smith, M. (2014, January 22). First NodeXL #ddj network graph. Twitter. https://twitter.com/marc_smith/status/425801408873385984
- Xaquín G. V. (2017, September 1). How to fix a toilet and other things we couldn't do without search. <http://how-to-fix-a-toilet.com/>
- Yau, N. (2018, May 21). Nigel Holmes new illustrated book on Crazy Competitions. *FlowingData*. <https://flowingdata.com/2018/05/21/nigel-holmes-new-illustrated-book-on-crazy-competitions/>

About the Authors

Eunice Au is programme coordinator at the Global Investigative Journalism Network, where she produces a weekly Top 10 #ddj round-up of the most popular tweets on data journalism.

Marc Smith is a sociologist building and applying tools to study social media networks.