

Anyone can be a data journalist

Hello, my name is Simon Rogers, and I'm going to talk today about how anybody can be a data journalist. It shouldn't be something that's sort of a scary and inaccessible and difficult because really now we're at a stage where it's very easy for anybody to begin and anybody to use tools to tell stories in new and interesting ways.

My background is very much a newspaper background. I was at The Guardian. I was, in fact, the launch editor of The Guardian News Unlimited. I also set up the data blog at The Guardian, and I have written a couple of books, including something called Facts are Sacred, which you can still get now and really kind of tells the story of data journalism and certainly from its early days of its most recent incarnation. I worked at Twitter where I was the first ever data editor at the company, and now I work at the Google News Lab where I'm a data editor.

The Google News lab is really a kind of a global team that works within Google to help the news industry. And we're like a bridge between the news industry and on the company. And my area is really around data journalism and newsroom innovation. So I work a lot with visualizations, data, storytelling and newsroom innovations around things like machine learning and VR and so on and so on.

I'm gonna talk a little bit now about my background, how I got into data journalism, because certainly when I was young, data was the last thing that I ever thought of working in. I wasn't very good at math at school. I was kind of put off by the idea of working in numbers. I wanted to work with words. But really, when I look back at the stuff that I kind of got into at the time, I can see that there was definitely something there about telling stories in new ways that was exciting to me. I very much grew up in a newspaper household. We had The Guardian six days a week and The Observer on Sundays and very much we work have always consuming news. But part of that for me was wanting to understand why things are happening. I was very much into Richard Scarry when I was young, and Richard Scarry is a children's illustrator who was very good at explaining how the world worked. And that's what I wanted to do, I wanted to explain how the world worked. Not just because it looked cool being a reporter in Richard Scarry, but you got to run around with a giant carrot pen. But mainly because of the way he would explain stories like this.

So this is a picture of a ship. It's very cartoonish. Lots of funny things going on. But if you look at this diagram, you would understand how that ship worked, how the internal combustion engine worked on that ship. So even those drawn in a very kind of cartoonish and not, you know, in a 100 percent accurate way. It was a way of explaining the story in a way that was accessible and understandable to me as a child and to anybody else that might want to understand it.

So this kind of led into me, certainly writing and working with designers on children's books. And when I first did these books, I thought, well, this is kind of interesting. But now I realised actually for a lot of children, really understanding the world visually is a way of understanding it better because the world is gray and difficult to comprehend. And actually using numbers can help make that more accessible to help make that easier to comprehend. Certainly when you're when you're young, because we are engaged by pictures long before we learn to read. We understand visuals way better than we

understand words and way better than we understand raw numbers. And that can really help you to kind of comprehend how data can be used to tell stories in better ways.

So part of that is really how we're at a stage now where anybody can do data journalism, anybody can get into it and able you can begin. Doesn't necessarily mean there's going to be great. It doesn't matter. It's the fact that anybody can get into this field now. This is a fanzine from 1976 called Dire Sideburns was a punk fanzine. And this is the sideburns theory of punk, which here's a chord. Here's another. This is the third. Now former band. And there is believe it or not a data journalism version of that, which is this the date set. Here's another date set. Here are some free tools. Now become a data journalist.

We're really at a stage where that is possible and there are lots of reasons for that happening. And we've seen this around the world now. Data journalism has truly become mainstream. We did a survey of newsrooms and we found over half of all newsrooms have a dedicated data journalist, particularly in digital newsrooms where it's 60 percent. It's a global phenomenon. This from the data journalism awards. You can really see how we had hundreds of entries. They were from round the world, half of the where from small newsrooms. There were 57 countries. And we've seen entries. And winners. It's like The Wall Street Journal from. We have had winners from Serbia, from Brazil, from Colombia, from Germany, from Russia and Ukraine. It's really a global phenomenon now. And part of that is because of the accessibility of the tools needed for data journalism. Those tools are getting better and better.

We've come a long way since things like this. Where essentially you could only make these kind of charts to showcase a story. This is something that was made using a tool called Carto. It shows sun rising around the world whilst people are tweeting about it. Very, very simple, but unimaginable a few years ago that anybody could just make that without being a hard core developer.

This map shows the Earth's wind patterns. Really worth checking out. But the person who did this taught themselves how to code, taught themselves how to build that design. And that's an incredibly powerful strength that we have. It's like a superpower. You have the superpower now to create data journalism and we can see people doing that. John Bird-Murdoch is visual journalist at the Financial Times, who taught himself how to design sophisticated and complicated visuals. From a journalism background that would never have been possible a few years ago.

But life as a data journalist can be isolating. Often people are working on their own in a sole practitioner places. There's no career path often for the data journalist. When the data journalist leaves, no two really knows how to replace them. So not feeling isolated is really important. And I think that has led to a feeling among data journalists of the importance and power of working together. And there are certainly things to help you do that. There's the Data Journalism Handbook. Just been published in a brand new edition, which really brings together the world's best practice and the best ways to produce data journalism today. And the way we work now, so we don't have to feel on your own.

And really, it's not about being a coder anymore. It can help. But it's not just about doing that. We have tools like Flourish, which you can learn about on this course elsewhere and Flourish is an amazing tool for creating visuals without being a developer or a designer or a coder and allows you to select different templates and really tell a story in interesting and exciting visual ways.

Things like tile grams, which is just a Github tool which allows anybody to make these kind of maps where you're really showing our geography in different ways. Sort of thing that would have been very tricky before now is freely and openly available to interactive for yourself. Now I work a lot with new data sets. You know, a few years ago, we didn't have access to, say, Twitter data or Google data, which is what I work with. So I really can talk a little bit now about that data and how it can tell stories in interesting ways and read doing that as a way to kind of illustrate the accessibility of data now.

First thing to say about Google data is there is a lot of it. It's very, very big. There are billions of searches every single day. And I made this map actually using Carto. Where I geographically posted every single city around the world where you can see Google data, where you can actually learn about Google data. It doesn't go below city level, but it goes from national, global, regional down to down city level. And I removed the map in the background. So really it's just a set of dots which are kind of like the world's brain that tells us where we are. And our natural kind of light guide to the world. Certainly is beyond the kind of echo chamber of social media. I would say and really gives you an idea, very kind granular level of what people care about, what people are interested about. It's incredibly honest.

This is a data set that I pulled last year showing on average where people are searching for gun shot versus gun control, people for buying guns vs. controlling them. And you can see how some states are are colored up in different ways there. Because you're never as honest as you are with your search engine. But after one particular shootings in Philadelphia, the map changed. It shifted straight away. And that honesty is really interesting to see.

We did this project with The Washington Post, where we looked at the top searches that were put in by people looking for same sex marriage in 2004 and 2015. And you can see how people's attitudes towards same sex marriage changed over time as the country kind of chilled out about it and became less of a huge deal. And right now, in 2019, it's hard to imagine some of the things that people were searching in 2004, because you know our attitudes change over time.

You can find a lot of this stuff yourself trends.google.com. It's really a site where it's very, very easy to go and search for yourself and really see how people are searching for particular issues and topics over time. They don't have to all be heavy and difficult subjects. This is a map showing searches for donuts versus searches for *cuissant* around the world. Let's see if you can guess which color is which over time.

And those kind of issues are very, very easy to do with Google Trends. And theres an immediacy to it, which is really interesting. This was a visualization we put together after the Paris attacks in 2015. You could see how people were starting to search for this story around the globe within minutes of it happening.

We also try and surface other data sets within Google. Some of them are more available than others. This is a map that shows visits to different kinds of restaurants over time and across the country. See where people care about pizza or coffee or Mexican food or Italian food or sandwiches or seafood.

And this is really about us trying to make that data more accessible and easier to get at. And there is really a thing with data journalism, which is very important was the vital necessity of being open. It's very easy to just produce something and do something clever and produce a visualization and then not make it necessarily open for other people to use. I've always wanted to produce data in a way that's reproducible by others that they could take that data and tell a story perhaps in a better way. And that is very, very important then to be open and make that data as accessible as possible.

And you can see that now around the world where people are taking public data sets and then making them accessible and making them open, like IndiaSpend, which takes public data within India and then produces really interesting websites, allowing people to explore it for themselves because often public data was produced originally in ways that are difficult to use.

And you can see now that every mainstream data journalism publication will have an open data site where you can go and get that data for yourself and then go and play with it for yourself. And you might be able to tell our story in a more interesting way than the upshot or having the posters done, because you'll come to that data with your own set of preconceptions, but also your own set of interests about what might tell a story in an interesting way.

I'm going to talk a little bit now about process, because for me, this is how I work. Certainly when I was at The Guardian, but it's also how we work now. And working with data can be tricky. So this is really how it works for me. Often we begin with the data itself. Might that data may have come from Google Trends or it may be you may be in send that data, there may be a recurring event. Something as interesting. It might be a theory that you want to explore. And often data journalism really is not necessarily a very sophisticated statistical operation. Often what you're doing is telling stories. You may be saying this thing is big on that thing or this thing has changed over time or this is how one place compares to another. So that work doesn't necessarily have to be super complex at this point. But knowing roughly what you want to do with the data can be really important. Do you want to compare or show change? What does the data mean? What other data sets can you mash up and bring together with it? And often two data sets are almost always better than one. Where you have to bring things together. There can always be a better way to tell story.

And often then you have this whole process where you really have to kind of get rid of the messiness of the data and process it and work with and often a lot of the work that data journalists spend a lot of time doing is that backhand mashing data sets up getting data sets to work together, removing the merge cells or the unnecessary columns of data that people put into often make spreadsheets look prettier.

But at the end of that, you can perform those calculations and that really leads to output. And output is a weird term. It sounds a bit like a kind of engineering, programming term. In a way that's what you're doing because there is no set way to tell a story with data journalism. You might produce a video, you might produce an animation, you might produce just a story of just publishing on social media. The fact that you discover that this thing is bigger than that thing might be enough. I might be all you need that to do that day. Or it might be a complex interactive. It's that variety of storytelling methods that makes data journalism so compelling and so interesting and so powerful right now.

I'm going to show you this link, too, because this is often how you start. You need to begin with data, right? So we actually publish all the data that we work with on a Github page and you can see it there. And that's really where you could begin. You could go there, find some data that we've already worked with. It's already being kind of processed and begin there. Begin with that storytelling process.

I always am with James Cameron, who is British journalist in the 50s and 60s, partly because he's an amazing journalist. I'm a really big fan of his work. Both sides kind of think this. This quote from him is very apposite, that "The new world be a place of answers and no questions, because the only questions left will be asked by computers, because only computers will know what to ask." And I take that to mean that we're at a stage now where we can ask questions of data and do stuff with data we never could before because of the access to technology that we have now. That's an incredibly powerful place to be. But what that means is that you can kind of get carried away and produce things that are over compacts or just because they're pretty and they don't really do anything. So there's a kind of simple thing that's really important to remember that what we're doing is all about telling stories in the best way possible. But it's about that raw journalistic exercise of telling stories and making that data as accessible as possible. Thank you. Now onto the course.