

## Module 4.2 - Data tools & resources

Welcome back for the second video of module four. Today, we're going to talk about the technical tools you need for data journalism and some resources to help you get started in those.

Now, all of you are bringing different skill levels into this course. Some of you might just be starting out with data journalism and some of you might be more advanced. So, the goal today is to give a broad overview of some of the most commonly used tools out there. So, we're gonna start at the beginning and we're gonna talk about the advanced tools as well. If you're just getting started, don't get overwhelmed by the amount of tools here. You do not have to learn all of these to do data journalism. I highly recommend that you start with spreadsheets and build from there and pick the tools that we discuss that might be useful to the particular project that you were working on.

Also, remember that these slides are available online and they have links to all of these different tools so you can work with them on your own time. And we're also going to give you resources that can help you learn all of these different tools.

The good news in data journalism is that there has been a big trend in the last few years toward more open source and freely available tools. It used to be very expensive to get started and data journalism. You needed a really powerful computer and you needed to pay for all of these tools that can be very expensive. You could get into the thousands of dollars very quickly. But now there are so many free alternatives that you really just can get a cheap laptop and use some of these free tools and you will do just fine.

As I mentioned, spreadsheets are really the gateway to data journalism and they will be really useful to you and all of your projects, even if you're going to use more advanced tools because so much data is available in spreadsheets. Many government agencies use them and so you might get an Excel sheet or be able to pop it into Google Sheets. Spreadsheets can help you sort your data, filter it, aggregate it and group it, add things together, and it can even build charts for you to visually analyze your data. So, spreadsheets can take you really, really far. And excel these days, the latest versions, can handle up to about a million rows of data. So, in many cases, spreadsheets may be all you need to learn.

But if you're getting into larger and larger data sets and data sets that are more complicated, that they have multiple tables, that you need to work with all of them together. Those are called relational databases. And that's when you need to learn SQL or structured query language SQL is a programming language that helps you deal with those large datasets. And SQL is also used by many government agencies. So, you might get data in this format as well.

There's lots of different flavors of SQL. You might hear of MySQL or SQLite or PostgreSQL. But one of the most commonly used types is SQLite and there are a lot of free tools that can help you work with SQLite. I recommend getting started with DB Browser, which is linked here on this slide. It's a free open source tool and it can deal with data sets basically as large as your computer can handle. And it's really easy to setup, much easier than other types of SQL.

Mapping is another good skill to have in data journalism. If you have a data set that has a geographical component, you might want to do spatial analysis on it. This means actually

analyzing your data in a spatial way. An example of this might be that there is a power plant and you might want to analyze how many people live within a certain distance from that power plant. These first few programs listed here. QGIS and ArcGIS online are the two programs that can help you do that kind of spatial analysis. Carto, Mapbox and Datawrapper are some really great online mapping tools that can help you build interactive maps for your projects.

When you're using data in your stories, you also might want to visualize it in some way. Now this could be everything from a simple bar chart that helps people compare numbers to an online map that helps them see spatial patterns or even an entire dashboard that helps them explore a large dataset in multiple different ways. All of these programs listed here on this slide are freely available and require no programming to make some cool custom interactive visualizations.

TableauPublic is a great place to get started. It does have a pretty steep learning curve, but once you learn that, then you can make visualizations pretty quickly and easily and then they are embeddable into your web site. Flourish is a really cool, fairly new tool that can make some really interesting interactives, including these bar chart races or line chart races that build animations into your stories and add a lot of interest. TimelineJS and StoryMapJS are templates that can help you build timelines or maps with different with different elements in them, so videos or audio or photos bringing together many multiple many multimedia into your story.

TimelineJS and StoryMapJS are templates that allow you to build multimedia timelines or story maps that include photos, videos, audio and all kinds of different media in one place. Infogram and Datawrapper also build lots of general-purpose types of charts and graphs that can be embedded into your stories.

The next group of tools, we'll talk about our programming languages. Many journalists today are learning these programming languages to help them with different tasks, such as wrangling very large data sets that you can't use in any other tools. Building some really cool interactive visualizations, scraping data from Web sites, or dealing with very repetitive tasks such as standardizing data over and over.

Popular programming languages really change over time. But the three most popular today are Python, R and JavaScript, and each one of them has their own specialty, although they can do a lot of general-purpose tasks as well. Python is considered the most generally useful and is typically used when scraping web sites or building some automated tasks.

R is really great at statistics and visualization as well. And JavaScript, particularly the D3 library with JavaScript is used for building visualizations and doing a lot of interactivity online.

Now, no matter what tools you're going to be using for analyzing and visualizing your data, you will have to do some steps to gather your data. And it's great if it comes in a spreadsheet or a text file that can be imported into any of these other tools. But in a lot of cases, you might get PDFs where your data is has been converted from a spreadsheet into a PDF and it has been locked in there. The tools on this slide can help you convert that PDF back into data. I highly recommend Tabula, especially because it is built by journalists for journalists and it is not online. It's not in the clouds. So, it's a much more secure option if you have some sensitive data.

But some of these other tools might be useful in different situations as well. Each one of them works a little bit differently, so some of them work on some PDFs and not others for example. I would also like to mention on the slide DocumentCloud, which is a free tool for journalists to upload PDFs and annotate them and then embed them within their story. So, for example, if you're covering a lawsuit, you could upload that PDF, make notes about some of the highlights in it and then put that in a window within your story so that people can see where you're getting your information. Or they can read the entire lawsuit if they are interested in it. It's a really great way to be transparent in our reporting.

Those are some of the large groups of tools and skills that we'll talk about. I have a couple of other tools I want to mention that can be useful for really any project. OpenRefine is one of the best tools out there for helping you clean and reformat your data. We mentioned in the last video that making sure your data is clean is a really crucial step before you analyze it. An OpenRefine can help you make that process easier and quicker and smoother. There are some great videos online on the OpenRefine site that's linked here to help you learn this tool.

Workbench is an exciting new tool that brings together all the pieces of a data project from gathering data to analyzing it to visualizing it even and embedding it into your Web site and it documents your entire process as well so that anyone can come back and check your work. It's a really great tool. There are some great tutorials online for learning this as well.

Collaborate is another exciting new tool that's available from ProPublica. Now, they're very well-known at ProPublica for collaborating with other news organizations. And this is the tool they built to help them do that. It links up to tools that help you crowdsource data or help you import data. And then you can verify data and communicate with the other people who are working with you on this story, even if they're not in your newsroom or even in your same country.

That's an overview of some of the most commonly used data journalism tools today. Now let's talk about some resources that can help you learn those tools. There really is no better way to learn data than to sit in a computer lab and actually do it hands on. So, I highly recommend coming to one of our IRE Boot camps to help you get started with data. We have boot camps on several different tools and we have several of them a year. We have a basic boot camp that starts with Excel and moves into SQL. We have a bootcamp on Python, one in R and one in data visualization using Tableau. We also have many different fellowships that can help you pay for this trip to come to bootcamp. So please check those out on our Web site, IRE.org. And we have two conferences here that also have computer labs and Hands-On data classes.

If you need some online resources, if you aren't able to make it in person and you need some hands-on resources, the Knight Center just had a wonderful course at the end of last year on using free data tools for analysis and visualization. That one's really great. You should check that out.

The Global Investigative Journalism Network has a wonderful list of data journalism resources that was put together by some of the best in the field that can also take you from beginner to advanced. All different levels and lots of different tools and concepts as well available at that Web site.

And then the Google News lab has some great online training modules as well. Google builds many of the free tools that are available to us as journalists and they have some really great training.

And that's the end of our lectures for module four. We'll see you on the discussion boards and we'll see you online.