Teaching Data Journalism

Cheryl Phillips

Abstract
Teaching data journalism begins with teaching critical thinking.

Keywords: critical thinking, data journalism education, programming, collaboration, data practice, researcher–journalist collaborations

At Texas State University, Professor Cindy Royal teaches web development.¹ A few thousand miles east, at the University of Florida, Mindy McAdams, the Knight Chair of Journalism Technologies and the Democratic Process, and Associate Professor Norman Lewis, teach a variety of classes from coding to traditional data journalism and app development. Alberto Cairo, the Knight Chair of Visual Journalism at the School of Communication at the University of Miami, teaches an entire programme focused on data visualization.

Go north and students at Columbia University and CUNY take classes taught by practicing data journalists from NBC and The New York Times, learning the basics of investigative reporting along with data analysis. At the University of Maryland, media law classes regularly go through the process of submitting public records requests for journalism projects. In Nebraska, Matt Waite teaches students to visualize data using Legos. At Stanford University, we teach basic data analysis, coding in Python and R and basic data visualization, more for understanding than presentation.

Data journalism professors—many of whom got their start as practitioners—teach in a variety of ways across the world (and the examples above

¹ Credit for this chapter is due to Charles Berret, co-author of Teaching Data and Computational Journalism, published with support from Columbia University and John S. and James L. Knight Foundation.
are just from programmes in the United States). Which programme is true data journalism? Trick question: All of them are. So how to teach?

The same way we teach any type of journalism class. Any specialization—from sports journalism to business reporting or science reporting—has domain-specific skills and knowledge that must be learned. Yet each rests on the fundamentals of journalism.

In the same way, data journalism education should begin with the fundamentals. By that, I don’t mean learning spreadsheets, although I do think it can be ideal for understanding many basic tenets of data journalism. There’s nothing like understanding the inherent messiness of entered data by having students embark on a class exercise that involves entering information into little boxes on a computer screen. I also don’t mean learning a particular way of coding, from Python to R, although I do think both languages have many benefits. There’s nothing like seeing a student run a line of code and get a result that would take four or more steps in a spreadsheet.

Learning about data journalism begins with understanding how to think critically about information and how it can be collected, normalized and analyzed for journalistic purposes. It begins with figuring out the story, and asking the questions that get you there.

And journalism educators likely already know the form those questions can take:

– Who created the data?
– What is the data supposed to include?
– When was the data last updated?
– Where in the world does the data represent?
– Why do we need this data to tell our story?
– How do we find the answers to the questions we want to ask of this data?

So, build the curricula using spreadsheets, or SQL, or Python, or R. It doesn’t matter. Just as it doesn’t matter that I once knew something called Paradox for DOS. What matters is knowing the steps to take with collecting and analyzing data. Visualization is key both in analyzing and presenting, but if visual analysis for understanding comes first—then presentation follows more easily.

This chapter contains a variety of approaches and starting points regarding how to teach data journalism, based on who you are, what level of programme you have and how you can build collaborative efforts. After introducing the “suitcase” approach to teaching data journalism, it explores one-course models, flipped classroom models, integrated models and experiments in co-teaching across different disciplines.
One Course Is All You Can Do: Packing the Suitcase

When we go car camping, we always make the joke that we pack everything, including the kitchen sink. The trick is knowing what you can pack and what would overload you to the point of unproductiveness. That kitchen sink is actually a small, foldable, cloth-based bowl.

If you are teaching just one class, and you are the solo data journalism educator—don’t try to pack in too much, including data analysis with spreadsheets and Structured Query Language (SQL), data processing using Python, analysis using R and data visualization design using D3, all in one quarter or semester.

Pick the tools that are vital. Consider making the class at least partly project-based. Either way, walk through the steps. Do it again and keep it simple. Keep the focus on the journalism that comes out of using the tools you do select.

In 2014 and 2015, Charles Berret of Columbia University and I conducted a survey and extensive interviews with data journalists and journalism educators. Most of those who teach data journalism reported that beginning with a spreadsheet introduces the concept of structured data to students in a way that is easy to grasp.

Another step is to ramp up the complexity to include other valuable techniques in data journalism: Moving beyond sorts and filters and into “group by” queries, or joining disparate data sets to find patterns otherwise undiscovered.

But that doesn’t mean adding a myriad number of new tools, or even picking the newest tool. You can introduce students to that next level using whatever technology works for you and your institution’s journalism programme. If it’s a university programme where every student has MS Access, then use that, but go behind the point-and-click interface to make sure that students understand the Structured Query Language behind each query. Or use MySQL. Or use Python in a Jupyter Notebook. Or use R and R Studio, which has some great packages for SQL-like queries.

The goal is to teach the students journalism while helping them to understand what needs to happen and that there are many ways of achieving similar operations with data in the service of telling a story.

Again, keep it simple. Don’t make students jump through hoops for tech tools. Use the tools to make journalism more powerful and easier to do. To go back to that car camping analogy, pack just what you need into your class. Don’t bring the chainsaw if all you need is a hatchet, or a pocketknife.

But also, once you have the one class established, think beyond that one-class model. Think about ways to build in data journalism components
throughout the department or school. Find shared motivation with other classes. Can you work with colleagues who are teaching a basic news reporting class to see where they might be interested in having their students learn a bit more about integrating data?

Some journalism professors have experimented with “flipped classroom” models to balance skills instruction, critical thinking and theoretical reflection. Students can take tutorials at their own pace and focus on problem-solving with instructors during class as well as learning other methods for tackling a variety of data journalism challenges. Professor McAdams from the University of Florida follows a flipped classroom model for her designing web apps class, for example.

One benefit for this type of classroom is that it accounts for journalists of many different skill levels. In some instances, a journalism class may draw interest from a student who is adept at computer science, and, at the same time, a student who has never used a spreadsheet.

But teaching data journalism goes beyond flipped classrooms. It means thinking about other ways to teach data journalism concepts. At SRCCON, a regular unconference, Sarah Cohen, the Knight Chair in Data Journalism at Arizona State University, and a Pulitzer Prize-winning journalist most recently at The New York Times, advocated using other analogue activities to engage students. Cohen and Waite, a professor of practice at the University of Nebraska, were introducing the idea of a common curriculum with modules that can be used by educators everywhere. The goal is to create a system where professors don’t have to build everything from scratch. At the conference in summer 2018, they led a group of participants in contributing possible modules for the effort. “We are trying not to have religion on that stuff [tools],” Cohen told the group, instead arguing that the focus should be on the “fundamental values of journalism and the fundamental values of data analysis.”

Now, a GitHub repo is up and going with contributors adding to and tweaking modules for use in data journalism education.² The repo also offers links to other resources in teaching data journalism, including this handbook.

A few possibilities for modules include interpreting polls or studies. Basic numeracy is an important component of journalism courses. Finding data online is another quick hit that can boost any class.

It also doesn’t mean you have to give up all your free time for the cause. Build a module or tutorial once and it can be used over and again by others.

² https://github.com/datajtext/DataJournalismTextbook
Or tap into the many free tutorials already out there. The annual conferences held by Investigative Reporters and Editors (IRE) and the National Institute for Computer-Assisted Reporting (NICAR) yield even more tutorials for their members on everything from pivot tables to scraping and mapping.

I guest-teach once a quarter for a colleague on finding data online. The benefits include creating a pipeline of students interested in exploring data journalism and being part of a collegial atmosphere with fellow faculty.

If possible, consider building modules that those colleagues could adopt. Environmental journalists could do a module on mean temperatures over time using a spreadsheet, for example. Doing so has one other potential benefit: You are showing your colleagues the value of data journalism, which may also help to build the case for a curriculum that systematically integrates these practices and approaches.

More on an Integrated Model, or Teaching Across Borders

A fully integrated model means more than one person is invested in teaching the concepts of data journalism. It also has potential to reach beyond the bounds of a journalism programme. At Stanford, we launched the Stanford Open Policing Project and partnered with Poynter to train journalists in analyzing policing data. Professors in engineering and journalism have worked together to teach classes that cross boundaries and educate journalism students, law students and computer science students. This is important because the best collaborative teams in newsrooms include folks from multiple disciplines. More recently, academic institutions are not only adopting such integrated models, but producing work that reaches into newsrooms and teaching students at the same time.

Just this month, the Scripps Howard Foundation announced it is providing two $3 million grants to Arizona State University and the University of Maryland, which will launch investigative reporting centres. Those centres will train students and produce investigative work, taking on the role of publisher as well as trainer.

Classes that have a mission and that move beyond the classroom are more compelling to students and can provide a more engaging learning experience. One of the most successful classes I have been a part of is the Law, Order &

---

3 To learn more about the grants for launching investigative journalism centres, see Boehm, J. (2018, August 6). Arizona State University, University of Maryland get grants to launch investigative journalism centers, AZCentral. https://amp.azcentral.com/amp/902340002
Algorithms class taught in spring 2018 by myself and Assistant Engineering Professor Sharad Goel. The class title is Goel's, but we added a twist. My watchdog class by the same name met in concert with Goel's class. Between the two classes, we taught computer science and engineering students, law students and journalism students. The student teams produced advanced statistical analysis, white papers and journalism out of their projects. Goel and I each lectured in our own area of expertise. I like to think that I learned something about the law and how algorithms can be used for good and for ill, and that Prof. Goel learned a little something about what it takes to do investigative and data journalism.

As for the students, the project-based nature of the class meant they were learning what they needed to accomplish the goals of their team's project. What we avoided was asking the students to learn so much in the way of tools or techniques that they would only see incremental progress. We tried to pack in just what was necessary for success, kind of like those car camping trips.

About the Author

Cheryl Phillips teaches data journalism at Stanford University and is the founder of Big Local News (biglocalnews.org), a collaborative data-sharing effort and platform for journalists.