

Teaching Data Journalism in a World of Tool and Tech Overload

Rahul Bhargava
School of Journalism
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA
r.bhargava@northeastern.edu

Catherine D'Ignazio
Department of Urban Studies & Planning
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
dignazio@mit.edu

ABSTRACT

Against a backdrop of systemic disruption in the field of journalism writ large, data journalism represents a subarea that is undergoing rapid transformation due to the introduction of new tools and techniques as well as the changes in reporting practices as journalists and newsrooms experiment and innovate. This paper explores the challenges for data journalism educators to teach in such a rapidly shifting landscape. Drawing from our experiences teaching journalism students in higher education, we assert that the goal of data journalism education amidst this complexity is not to teach tech, nor even to teach technical skills, but rather to model for students strategies of dealing with transformation and complexity. These include peer learning, hands-on learning activities, modeling learning and information seeking, and establishing a culture of critique. We introduce a number of activities that put those approaches into practice, drawing on learning literature to support our fellow educators shifting from the “banking model” of education[10] to a learner-centered model[23]. Working with students to co-create knowledge, acting as a “Guide on the Side”[15] can help better prepare students for the constantly evolving ecosystem of technologies and tools that support data journalism.

1. INTRODUCTION

In the introduction of *Journalism & Mass Communication Educator*, Maria B. Marron states, “Look anywhere at media, and there is nothing but disruption.” [19] She is speaking broadly about shifts in business models, user participation and technology. Donica Mensing and David Ryfe similarly make a case for a tri-fold crisis in journalism that consists of “economic disruption, technological disruption and decline in cultural authority” [1, 20]. These authors are preoccupied not just with changes to the profession and practice of journalism but their implications for teaching journalism in the context of higher education against a backdrop of systemic disruption. The field of data journalism is a particularly accelerated subarea in this regard [14]. How does one teach an emerging field whose foundation is undergoing profound change, whose tools and methods shift radically from year to year, and whose mastery involves drawing from quantitative, visual and narrative domains?

In this paper we draw on our experiences teaching non-technical learners about data and computational methods¹.

¹Bhargava teaches data storytelling to undergraduate and

The recent report “Teaching Data & Computational Journalism” [20] by Charles Berret and Cheryl Phillips identified one of the key institutional challenges to educating journalists as the lack of faculty expertise in data journalism. This is logical, given that it was until only recently that journalists were mainly writers working in a specific form of black and white type set on newsprint. Most academic programs still have strong divides between print and broadcast journalism. Only recently have they come to terms with the idea of teaching convergence journalism [16] and restructured curricula to focus on multimedia storytelling. With the mainstreaming of data and computational journalism in the past ten years, leading institutions now recommend adding data scraping and cleaning, data analysis and visualization, coding and other computational methods to the list of skills that journalism programs should be offering [2]. However, research shows that journalism educators, with very good reason, are “stressed out” by these technological changes and the proliferating tools and techniques that accompany them [32]. While there have been numerous studies of teaching visualization and data journalism, few focus on the underlying pedagogical approaches that respond to the overall context of the field [17, 30, 31, 13].

2. TOOL OVERLOAD

Neither journalism professionals nor educators can keep up with the rapid proliferation of tools, techniques, platforms and technology that can be used in support of data-driven storytelling. The co-authors of this paper have, in separate work, logged more than 500 free or “freemium” tools designed for journalists and communicators to collect, analyze, or visualize data. This creates tremendous complexity for journalism educators. Navigating the myriad of tools alone can be a challenge to faculty strapped for time without large professional development funds. Below are some examples of tools used in a data processing pipeline.

graduate media students at Northeastern University and runs the community-centered data consultancy Data Therapy. D'Ignazio previously taught data visualization to undergraduate and graduate journalism students at Emerson College. Both run workshops for professionals in journalism, communications and the non-profit sector.

| Process | Tools |
|---------------------------|---|
| Acquiring | Python scripts, browser extensions |
| Cleaning | OpenRefine, Trifacta Wrangler |
| Exploring and analyzing | Tableau, Excel, R |
| Presenting visualizations | DataWrapper, Observable, Vega-Lite, D3.js |

Should one introduce map-making to students with CARTO, GoogleMaps, ESRI StoryMap, Knight StoryMaps, ZeeMaps, OpenStreetMap, plot.ly, Tableau, D3 or R? The options are intimidatingly large.

While the explosion of tools can be seen as a boon to an emerging field, it also presents challenges to the data journalism educator. The landscape shifts continuously. Businesses that offer free tools that students and professionals come to rely on can easily shift into a paid business model, as DataWrapper did in 2014², or discontinue them altogether. Open source tools maintained by communities of dedicated volunteers can lose steam and have a hard time fixing bugs and creating good user experiences, as has happened with OpenRefine³. Technology that comes out of research or grant-funded projects may thrive for some time but ultimately be left unsupported and unmaintained, as happened with the before-its-time platform ManyEyes⁴. Moreover, questions of data ownership and privacy get complicated when data (which may contain sensitive information about sources) is uploaded to and hosted on proprietary platforms. Too often it is unclear when, for how long, and what purposes uploaded data is stored on cloud services.

3. PEDAGOGICAL APPROACHES

By suggesting classroom strategies for navigating this complexity, we hope to begin to shift the norms of journalism education in a computational and data-driven world. Specifically, we argue that it is not feasible for data journalism educators to master all tools and techniques to transmit them to students, nor should they try. There is a wealth of project- and peer-based learning literature to draw upon that can support educators to make the shift from the “banking model” of education [10] to a learner-centered model[23, 8] in which students co-create knowledge with the professor, who acts as a “Guide on the Side”[15]. Indeed, we assert that data journalism teachers themselves should actively seek to connect with experts outside the classroom, intentionally set up situations in which their students teach them new skills, and model “not knowing but finding out” for their students.

We have implemented a number of methods to tackle the challenges around information overload, technical complexity and tool proliferation in data journalism. This paper reviews four approaches we have taken by providing background pedagogy, describing what we do, and discussing student reaction and impact. Our approaches are deeply rooted in Jean Piaget’s foundational work around the processes of “assimilation” and “accommodation” in learners [23]. His work describes how new information and experiences are ei-

ther assimilated into a learner’s existing theories, or how the new information causes the learner to change her theories to accommodate the new information. This approach helps us value the student as a collaborative learner in educational settings, and respects the experiences and context they bring. Building on this foundation, we draw on a variety of pedagogical theories to inform each of our approaches. These include:

- peer learning
- learning by doing
- modeling learning and information seeking
- establishing a culture of critique

3.1 Peer Learning

Working with data on a budget involves piecing together a patchwork quilt of tools. The sheer number of tools to choose from has made it impossible to introduce even the most important of them all to students. More critically, there isn’t well-established criteria for helping students understand how and when to piece together this quilt. To address this challenge we look to peer learning approaches. We define peer learning as the process of students learning from and with each other, as opposed to from their professor or a third party domain expert. Decades of existing study has held up peer learning as an effective strategy in a variety of educational settings[28]. Existing work also suggests it allows for engagement of a more diverse set of learners[11]. Our courses have integrated peer learning to address tool introductions in two distinct ways.

We have students write reviews of tools for each other as a way to build and distribute expertise across the classroom. In class, students select the top 3 tools they are interested to learn, and then assign each student one tool to test and write a review about. Often these are tools that we, as the professors, have never used. The audience for each review is other learners, including educators and practitioners. We have each student present the tool they reviewed for 5 minutes in class, allowing for broader awareness across the class. After these presentations, students are considered the class “expert” in that tool and over the course of the semester we will often refer other students to that expert when they need to use that tool in their final projects.

In another example, we have students teach tools to each other to establish criteria for assessing when and how to use them in appropriate ways. Bhargava uses this technique to introduce students to useful tools for mapping. He assigns half the class a tutorial that introduced them to CARTO⁵, and assigns the other half a tutorial for mapping in Tableau Public⁶. Each group is assigned the task of analyzing the same dataset geographically using the tool specified. In the next class the groups are paired across tools; each pair comprised of one person that learned CARTO and one that

²<https://www.journalism.co.uk/news/datawrapper-to-launch-paid-for-options/s2/a562961/>

³<http://openrefine.org/>

⁴<http://www.bewitched.com/manyeyes.html>

⁵<http://academy.CARTO.com/courses/beginners-course/>

⁶<http://www.tableau.com/learn/tutorials/on-demand/basic-mapping>

learned Tableau. They are then given 15 minutes each to introduce the other to what they made, how they had done it, and why that tool might be useful. The activity concludes with a short collective discussion comparing the two.

These two examples of peer learning within formal undergraduate classroom settings demonstrate how to start tackling the challenge of introducing the enormous volume of online tools for working with data. We find ourselves constantly referring back to tool reviews as students run into a problem best solved by a tool a different student reviewed. This solidifies the idea that we are all learning this new area together, and reinforces the value of the reviews and learning a student did (even if they don't end up using the tool they reviewed). The paired teaching activity consistently ranks highly in end-of-semester feedback from students, and the post-activity discussion always raises relevant conversations about tool affordances and appropriateness.

However, assessing peer learning can sometimes be a challenge. Fortunately, previous work has laid out motivations and approaches for this[6]. We focus on individual assessment for the reviews, since we have the artifact of the review itself to evaluate. For the paired peer teaching, we use the discussion time afterwards as a group-level assessment for the learning that happened in pairs.

3.2 Learning by Doing

Engaging students in using computer-based tools for story-finding and storytelling requires a rich set of invitations to appeal to different learning styles. Far too many students have described rote introductions to data tools, involving the professor tediously doing worksheet and spreadsheet manipulations of data they did not understand. Those basic activities do little to prepare students for a future of learning new tools to find and tell stories in new ways. The best way to learn how to learn these tools is by trying them out.

Building on Piaget's previously described concepts of "accommodation" and "assimilation", we take inspiration from Seymour Papert's theory of Constructionism [22]. His approach revolves around the idea that optimal learning occurs when people are designing and making things that are meaningful to them or their peers; you learn by doing the task with your peers. This pedagogy values the learner as a rich individual full of experiences, knowledge, and ideas that can be engaged to introduce them to new material. We're inspired by Papert's classic anecdote about wanting to create "soap sculpture math"; a math that allows for the type of engagement children have when carving sculptures out of soap; lovingly attended to over time and celebrated for uniqueness and playfulness [21]. We are looking for a model for what "soap sculpture data analysis" might be. In addition, Papert's intellectual descendants have created a variety of approaches that we can draw from; including design principles for building software for learners [12, 27] and guidelines for creating activities to introduce those tools [26].

We have created a number of hands-on learning activities, and tools to support them, that integrate this playful pedagogy into our data storytelling curriculum.

- Getting started with a dataset is often intimidating.

We introduce students to how to start looking for stories in a spreadsheet by brainstorming questions to ask the data in front of them, and looking for other datasets they might need to answer those questions. Our WTFcsv tool supports this activity by providing a quick overview of a spreadsheet's columns.

- Students often approach text as qualitative data, unaware of the potential of computational text analysis. We introduce quantitative text analysis by having students look for patterns in pop music lyrics, and then sketch what they find as visual stories on large pieces of paper. We created the WordCounter tool to support this activity, by showing the frequency of words and phrases used [3].
- Quantitative text analysis is often more meaningful in comparison, and can involve algorithmic analysis. We introduce the idea of comparing two sets of documents to identify similarities and differences. Students use SameDiff to compare different artists' lyrics and then write an imaginary duet two artists would sing together.

These activities, and others, are collected online as a suite available on our Data Culture Project website⁷. Each offers a fun, hands-on approach to learning concrete skills related to finding and telling stories with data. The tools are available in multiple languages, with culturally appropriate sample data, to invite an even larger audience into the field.

Students consistently respond well to these hands-on activities in our classes. This validates findings from the field of psychology that demonstrate how externalizing learning in drawings leads to more impactful and memorable learning experiences[29]. One commented on appreciating the opportunity to "get our hands dirty in the process"; while another commented that integrating these activities into lectures "is incredibly useful". The artifacts they produce give us a window into the learning that happens, letting us point to sketches, questions, and lyrics to demonstrate their growth in understanding of key concepts. Each activity ends with a sharing time for small groups to describe what they have made, offering opportunities for discussion to open up larger and deeper issues the tiny activities only touch on. We offer activity guides with share-back discussion suggestions and more for other educators who wish to run them in their classrooms. Our approach to developing these tools takes inspiration from a history of involving the target users as co-designers in the software itself, and an awareness that acceptance is driven by features, experiences, awareness, and purpose [18].

3.3 Modeling Learning and Information Seeking

Another challenging aspect of the disruption in journalism is that so much of the conversation in the field is moving at a pace far faster than ever before. Data journalism today is not a static field. It is endlessly dynamic, with innovations emerging from many sectors. Both professors and students need to consider themselves learners in this arena.

⁷<https://databasic.io/en/culture/>

The professor can explicitly position herself as a learner to set an example for students (by frequently saying “I don’t know, but let’s find out together”) and modeling information seeking and technical troubleshooting practices. Introducing students into an evolving landscape like this requires a new set of approaches to how they think of themselves joining the field. We find inspiration here in the emphasis on learning by doing in the “teaching hospital” model of journalism⁸ as well as Vygotsky’s writing about how learners join fields based on apprenticeship[33]. Vygotsky documents how learners’ interactions with more “capable” domain experts can benefit those that are primed to join a field, but need support to continue developing well. This “zone of proximal development” theory offers a methodology that involves respecting both the independent skills of the learner, but also the process of being challenged and supported by more experienced co-learners or facilitators. We attempt to implement this in a variety of ways in our courses.

Our readings are sourced from a variety of online platforms, providing an understanding of where the innovation is happening. More than half of the readings we assign are written in the last 3 years. As Berret and Philips’ report found, there is little agreement among data journalism educators on textbooks[2]. We would offer that that is because the field is moving too quickly and unevenly to capture in a single-authored text. Multiple (often contradictory, interdisciplinary) voices are important. Classic academic papers and books certainly play a role, but are not where one finds the latest responses to the challenges being faced. Videos, podcasts, blog posts, and other multimedia artifacts also set the norm for how new ideas are communicated in the field by professionals. Roughly half of the readings we assign are blog posts or videos. This is built on the rich history of integrating multimedia content into educational settings[5, 24].

Additionally, we require our students to join the conversation online in a variety of ways. At the most basic level, students in both our classes are required to post assignments to a public blog for the class. This builds a practice of writing about ideas still under construction. D’Ignazio’s “Getting in the Flow” assignment requires a number of additional steps:

- Tweeting out ideas and inspirations under the #emersondataviz hashtag on Twitter.
- Joining a number of relevant mailing lists (Data-Driven Journalism, NICAR, Data is Plural, and more).
- Following at least 3 data journalism-related blogs.

The goal of this assignment, which occurs early in the semester, is to put the journalism students in the middle of the evolving conversation among experts. Students see immediately

⁸While the teaching hospital model does take a hands-on approach to learning and uses journalism to begin a conversation with the surrounding community, we think Mensing and Ryfe [20] raise important critiques about its limitations in introducing students to the complexity and uncertainty that currently surround journalistic business models and professional practices.

that professionals don’t know everything. They rely on each other for methodological brainstorming and technical troubleshooting and draw on the knowledge of a global group of peers. This puts students in the right places to understand where the new norms are being created in an emerging field, again echoing Vygotsky’s theory of the social construction of knowledge (as opposed to Piaget’s heavy focus on the individual).

This process of thinking out in the open can be difficult for students to take on. Classrooms are intended to be safe spaces for learners to experiment with ideas, so mainly we focus on introducing students to these digital public spaces as listeners.

3.4 Establishing a Culture of Critique

Just as hearing multiple (interdisciplinary, possibly contradictory) voices in course readings should be seen as a desirable design goal, students should hear multiple authoritative voices in the classroom as well. In art and design schools, the “critique” functions as a reason for learners, professors and invited outside guests to come together around an artifact [7]. We assert that this model has utility for data journalism education in several ways.

First, the artifacts of data journalism are often (though not always) visual and possibly interactive. Rather like a user testing session, the critique offers a way of efficiently gathering multiple people’s perspectives and reflections on the reception and understanding of the story, graphic, visualization or app. A good critique empowers peers to speak equally with professors and invited guests, so this folds back into peer learning and assessment.

Second, because producing data-driven stories is complex - doing it well may involve computational skills, quantitative skills, visual design skills and narrative skills - no single expert will look at the result in the same way. Data scientists will ask questions about the collection and analysis of data and offer suggestions for alternate comparisons, writers will offer better ways to contextualize numbers in the narrative, designers will give advice on ease of use and visual experience, etc. Inviting these experts into the classroom and making the critique a public process empowers students to see their work from these different points of view. For example, a designer attended final project reviews for D’Ignazio’s class in Data Visualization and observed that all projects needed to make color choices more intentionally and to consider the consistency of their color palettes across a story when using multiple graphics. Several teams commented that they had never thought about consistency in color and tuned their final projects accordingly - the designer’s feedback had helped them see their work in a new way.

Finally, inviting expert guests to the class is also an opportunity for the professor to see and learn from these other perspectives as well as to establish relationships with others in the field. Small and continuous opportunities for professional development and connection with the world outside academia should be encouraged and incentivized by the institution⁹, particularly for fields like data journalism that

⁹For example, departments might allot faculty a small guest

are both transdisciplinary and undergoing rapid transformation. This provides an informal check and balance that class approaches and student work are in line with emerging practices and introduce new ways of thinking to the professor. For example, the guest designer's advice about consistency across graphics in a story is something that D'Ignazio had not previously noticed¹⁰. After the guest's visit, she has now incorporated that perspective as basic visual design feedback for student work in data journalism. Bhargava has had similar changes in his approaches to his teaching based on insightful comments from invited guests.

The feedback in student evaluations is definitive on the value of outside voices. One student notes, "I really appreciated how you brought in colleagues to help critique and inform our work." And several students noted that opportunities to learn from people outside the classroom were one of the most rewarding parts of the class; saying "I liked the guest panel" and "the most valuable parts include engagement opportunities with professionals (field trips, visits, MIT presentations)".

4. CONCLUSION

In this paper, we have outlined how tool proliferation, transformation of reporting methods, and new emerging social practices are the new norm in the field of data journalism. We do not expect things to settle down anytime soon. To the contrary, we see journalists taking up emerging computational methods like machine learning [4], reverse-engineering of algorithms [9] and sensor journalism [25] and expect that there will soon be tools that help other journalists and newsrooms take advantage of these methods in a more user-friendly way. The goal of data journalism education in this landscape should not be to teach tools, nor even to teach technical skills, but rather to model for students strategies for dealing with transformation and complexity. These include peer learning, learning by doing, joining the field as a learner and establishing a culture of critique. All of these methods seek to de-center the professor as "all-knowing expert" (still too often the norm in higher education) and move towards a future where professors model practices of information seeking, experimentation and assessment together with students and a community of transdisciplinary experts from outside the institution. This is essential for the creation of a model of journalism education that serves today's students. Ultimately, the value for students is to introduce them to a lifelong set of strategies to deal with a field that is in a permanent state of technological and informational transformation.

5. REFERENCES

- [1] C. W. Anderson. The sociology of the professions and the problem of journalism education. *Radical Teacher*, 99:62–68, 2014.
- [2] C. Berret and C. Phillips. Teaching data and computational journalism, 2016. <https://www.gitbook.com/book/columbiajournalism/teaching-data-computational-journalism/details>.
- [3] R. Bhargava and C. D'Ignazio. Designing tools and activities for data literacy learners. In *Proceedings of the 2015 WebScience conference Data Literacy Workshop*, 2015.
- [4] G. Bhatia. How newsrooms are using machine learning to make journalists' lives easier.
- [5] N. C. Boling and D. H. Robinson. Individual study, interactive multimedia, or cooperative learning: Which activity best supplements lecture-based distance education? *Journal of Educational Psychology*, 91(1):169–174, 1999.
- [6] D. Boud, R. Cohen, and J. Sampson. Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4):413–426, 1999.
- [7] K. Buster and P. Crawford. *The critique handbook: A sourcebook and survival guide*. Prentice Hall, 2007.
- [8] R. M. Cullen, M. Harris, and R. R. Hill. *The learner-centered curriculum: design and implementation*. Jossey-Bass, 2012. OCLC: 753624215.
- [9] N. Diakopoulos. Algorithmic accountability. 3(3):398–415.
- [10] P. Freire. *Pedagogy of the oppressed*. Continuum, 1968. OCLC: 43929806.
- [11] D. Fuchs, L. S. Fuchs, P. G. Mathes, and D. C. Simmons. Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1):174–206, 1997.
- [12] I. R. Harel. *Software design for learning : children's construction of meaning for fractions and LOGO programming*. Thesis, Massachusetts Institute of Technology, 1988.
- [13] J. Hewett. Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, 17(1):119–137. Publisher: SAGE Publications.
- [14] J. Howe, A. Bajak, D. Kraft, and J. Wihbey. Collaborative, open, mobile: A thematic exploration of best practices at the forefront of digital journalism.
- [15] A. King. From sage on the stage to guide on the side. *College teaching*, 41(1):30–35, 1993.
- [16] J. Kolodzy. *Convergence Journalism: Writing and Reporting across the News Media*. Rowman & Littlefield Publishers, 2006.
- [17] L. Y.-H. Lo, Y. Ming, and H. Qu. Learning vis tools: Teaching data visualization tutorials.
- [18] R. R. Mackie and C. Dennis Wylie. Factors influencing acceptance of computer-based innovations. In *Handbook of Human-Computer Interaction*, pages 1081–1106. North-Holland, 1988.
- [19] M. B. Marron. Nothing but disruption. 71(2):131–132.
- [20] D. Mensing and D. Ryfe. Blueprint for change: From the teaching hospital to the entrepreneurial model of journalism education, 2013. http://www.academia.edu/3412374/Blueprint_for_Change_From_the_Teaching_Hospital_to_the_Entrepreneurial_Model_of_Journalism_Education.
- [21] S. Papert. Situating constructionism. In I. Harel and S. Papert, editors, *Constructionism*. Ablex Publishing, 1991.
- [22] S. Papert. *The Children's Machine: Rethinking School In The Age Of The Computer*. Basic Books, revised ed. edition edition, 1994.

honoraria budget each semester in order to invite guest reviewers to critiques.

¹⁰We only reluctantly admit this for the purposes of this paper, and hope that readers will not tell anyone else.

- [23] J. Piaget. *The origins of intelligence in children*, volume 8, No. 5. International Universities Press, 1952.
- [24] R. B. Pipes and J. M. Wilson. A multimedia model for undergraduate education. *Technology in Society*, 18(3):387–401, 1996.
- [25] F. Pitt. Sensors and journalism.
- [26] M. Resnick and E. Rosenbaum. Designing for tinkability. In *Design, make, play: Growing the next generation of STEM innovators*, pages 163–181. Routledge, 2013.
- [27] M. Resnick and B. Silverman. Some reflections on designing construction kits for kids. In *Proceedings of the 2005 conference on Interaction design and children*, pages 117–122. ACM, 2005.
- [28] P. Sanders. Peer tutoring: An effective instructional strategy. 2001. <http://eric.ed.gov/?id=ED457224>.
- [29] S. P. Schmidgall, A. Eitel, and K. Scheiter. Why do learners who draw perform well? investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, 60:138–153.
- [30] S. Splendore, P. Di Salvo, T. Eberwein, H. Groenhardt, M. Kus, and C. Porlezza. Educational strategies in data journalism: A comparative study of six european countries. *Journalism*, 17(1):138–152. Publisher: SAGE Publications.
- [31] G. Treadwell, T. Ross, A. Lee, and J. K. Lowenstein. A numbers game: Two case studies in teaching data journalism. *Journalism & Mass Communication Educator*, 71(3):297–308. Publisher: SAGE Publications Inc.
- [32] P. S. Voakes, R. A. Beam, and C. Ogan. The impact of technological change on journalism education: A survey of faculty and administrators. *Journalism & Mass Communication Educator*, 57(4):318–334, 2003.
- [33] L. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.