Teaching Data Journalism

Assessing the current state of data journalism education, Bahareh Heravi proposes a way forward for developing data skills in journalism programmes and beyond.

Introduction

Data journalism is a relatively new term, yet there are multiple definitions at play. Before we delve into a discussion of data journalism and pedagogy in this chapter, I’ll specify my usage: I define data journalism as finding – in data – stories that are of interest to the public and presenting them in the most appropriate manner for public use and reuse.

Similar to any other journalistic work, data journalism puts the tenets of journalism first: it is about the investigation, the story, and communicating that story to the public. In data journalism, data is the source, and computational methods and applications are the tools to aid journalists in their work.

The rapid emergence of data journalism in newsrooms, and the challenges this poses, calls for a review of the educational offerings for journalists. The limitless amount of data generated every day, the vast number of available data sources, and the abundance of data and computational tools and applications available, demand new skills and training. To facilitate the demand for ‘upskilling’ in the data realm, we must pay attention to how journalists are trained. For the purposes of this paper, I will focus on data journalism in a higher education/university context.

Teaching data journalism

Teaching data journalism is often conceived of as teaching data to journalists and rarely the other way around, i.e. teaching journalism to data specialists. I will follow this distinction, and for the purpose of this article, I divide the teaching of data journalism in
two main categories: teaching data to already-journalists and teaching data to to-be journalists.

Teaching data to already-journalists is often at the postgraduate level through professional certification programmes or individual courses where existing journalists, or journalism graduates, can enhance their data skills. Examples are the Lede programme at Columbia’s Graduate School of Journalism, or the postgraduate professional certificate programme in data journalism at University College Dublin. The latter specifically requires that the participants have a degree in journalism or have worked as professional journalists.

Teaching data to to-be journalists is different in that it aims to incorporate data into journalistic pedagogy from the beginning. In this case, a complete programme is designed around teaching both journalism and data at the same time. These are often undergraduate programmes or Master’s programmes where both journalists and non-journalists are accepted. Examples are Cardiff University’s Computational and Data Journalism MSc and Columbia University’s dual degree (undergraduate) programme in Journalism and Computer Science. Both programmes teach journalistic as well as data skills as part of their curricula.

We return to these two categories later in this chapter. In the following section, I briefly consider what data journalism programmes are available and what courses are being taught around the globe in this area.

**What is out there?**

Berret and Phillips (2016) studied 113 American Journalism programmes, accredited by the Accrediting Council on Education in Journalism and Mass Communications, to find computer-assisted reporting (CAR) and data journalism programmes. They report that nearly half these programmes offer no course on data journalism related subjects. Out of the 59 programmes which do offer data journalism classes, 27 offer just one course, usually foundational, and 14 offer two courses. Only 18 of the 59 programmes are reported to offer three or more courses in this subject. They further report that the courses offered are ‘largely introductory and the need is still largely for the basics, such as knowing how to use a spreadsheet, understand descriptive statistics, negotiate for data, clean a messy data set, and then interview it to find a story’ (ibid: 9).

In a separate ongoing study, I compiled a dataset of data journalism courses and programmes globally. Setting aside the details of this study’s methodology or research
results, it became clear that the US has the largest offerings in data journalism-related courses and programmes while in Europe only a scattered number of such courses and programmes exist. Table 1 provides an overview of the number of data journalism-related subjects globally.

Table 1: Number of data journalism-related courses and programmes in various countries

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>Canada</th>
<th>China</th>
<th>Ireland</th>
<th>Netherlands</th>
<th>UK</th>
<th>Australia</th>
<th>Italy</th>
<th>Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>153</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sudan</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

The data suggests that, out of all European countries listed in Table 1, only the UK (three universities), Ireland (one university) and Spain (one university) offer data journalism within a self-contained programme, with more than one course dedicated to data journalism or offerings of postgraduate programmes in data journalism. The rest of the countries in Europe only provide one or two courses in this area.

When it comes to instructors of these data journalism courses and programmes, in line with the study underlining Table 1 and of Splendore et al. (2016), Berret and Phillips (2016) report that many journalism programmes do not have a faculty member skilled in data journalism. The faculty breakdown differs by region – the dataset behind Table 1 suggests that while most American programmes are taught by professionals, in Europe many university programmes in data journalism are taught by academics, with professional journalists providing occasional training.

Data journalism skills

A systematic approach to planning a data journalism programme requires learning about the skills that prospective students already present and the skills they would find most useful to learn. To understand these aspects, I used data collected as part of the Global Data Journalism Survey 2017 (Lorenz 2016, Heravi 2017). As part of this survey, we asked participating journalists about their educational background, existing skills and skills they expressed interest in acquiring for their future work. This study (3 December 2016 to 10 May 2017) was open to anyone who worked as a journalist or a data journalist in the previous year. Some 206 journalists from 43 countries participated in the survey with 181 respondents filling it out to completion (Heravi 2017). In the remainder of this
section, I present the partial results concerning data journalism education, from the initial survey results previously reported in Heravi (2017).

While the survey was open to all journalists, 86 per cent of participants considered themselves to be a data journalist. The survey results, we believe, are potentially biased towards journalists who have some interest in data journalism as they were more likely to participate in the survey to begin with. Despite the high percentage of participants reporting themselves as data journalists, in terms of data journalism proficiency, only 18 per cent rated themselves as experts in data journalism, while 44 per cent of respondents identified as having a better than average knowledge in data journalism and 26 per cent identified as having average knowledge in the field. Some 13 per cent identified as novice or below average level of expertise in the field. Only half of our participants (50 per cent) had formal training in data journalism.

In terms of education level, 97 per cent had a university degree, with 40 per cent graduate (bachelor) level, 54 per cent postgraduate and 3 per cent with a doctorate or above degree.

Studying the academic background of participants, 62 per cent were formally educated in Journalism at the university level. While Journalism was by far the most prevalent higher education degree, it was followed by Politics (15 per cent), Computer/Information/Data Science/Engineering (12 per cent) and Communication and Language/Literature, each with 10.5 per cent. Some 26 per cent of respondents said they were educated in ‘other’ degrees.

In terms of formal training in knowledge areas used in data journalism, most of our participants demonstrated a high degree of formal training and proficiency in journalism, with smaller and varying degrees of formal training in the more data oriented and technical aspects such as data analysis, statistics, coding, data science, machine learning and data visualisation (Heravi 2017). This breakdown of formal training in various related fields between our participants is depicted in Figure 1.
What data skills are journalists interested to learn?

A remarkably high portion of participants in the survey (98 per cent), expressed interest in acquiring further skills to practise data journalism, with 81 per cent indicating they were very interested. Some 42 per cent said they were interested in formal higher education degrees in this area. However, if the training offered were to be shorter-term or more flexible, a striking 74 per cent of participating journalists expressed interest in formal training in higher education, e.g. a postgraduate certificate or higher education diplomas.

In terms of specific data skills journalists are interested to acquire, data analysis presented itself as the top skill, with 64 per cent expressing they were interested in learning about it. This was followed by learning ‘how to programme/code’ at 63 per cent and ‘visualising data’ at 51 per cent. These top three data skills were followed by another three skills: ‘how to clean data’, ‘how to develop data-driven applications’ and to learn ‘how to check if data is reliable’, with more than 48 per cent of journalists expressing interest in each.

In summary, these results show that most participating journalists had formal higher education training in Journalism and related areas but these same journalists lacked training in data skills. Shorter, targeted higher education programmes would be the most attractive offering for increasing their skills in data, and many expressed a desire to make such improvements. The most important topics to be taught, according to the survey, are data analytics skills, followed by coding skills. Indeed, these figures signify how important training, and particularly data journalism training, is when it comes to formal higher education training.

What to teach?

As mentioned earlier, data journalism is about journalism first, and then it is about data. Hence, the first skills that any student must have are journalistic and investigative skills. After these, the most important topics to cover would be familiarity with data and data sources, an understanding of the lifecycle of data journalism projects, skills for data wrangling and, most importantly, data analysis skills, including sufficient knowledge of statistics. More advanced data visualisation, programming and other advanced topics may follow.

To discuss this in more detail I will return to our two groups: the ‘already journalist’, and the ‘to-be journalist’.
The ‘already-journalists’

‘Already-journalists’ are individuals who have trained in journalism and possibly have worked as professional journalists. They are familiar with the news production cycle and the aspects involved in ‘doing’ journalism: finding story leads, securing sources, verifying information, storytelling, writing and so on. Individuals in this group often require training in the data aspect and not much training in journalism. They may have further specific requirements in terms of the structure of the courses, as well as class timetables, in order to fit with their professional lives.

To this end, the first course any journalist interested in data journalism should take is an introductory course to data journalism, to understand the various aspects of working with data. Such a course could start with the possibilities inherent in working with data, including critical questions around the nature of ‘empirical’ work. The course could begin by examining exemplary data stories to unpack their components and determine what makes them stand out as compelling and convincing stories. Basic technical skills also need to be surveyed: learning where to find data, the basics of collecting and cleaning data, entry level statistics and how to use tools such as Excel or ‘off-the-shelf’ data visualisation tools (e.g. Datawrapper, Infogram or Tableau). By the end of this course, the participants should be able to create data stories of their own, even if these stories are basic in their statistical analysis or contain rudimentary visualisations. An example of this type of work is evident from my Journalism MA students at the National University of Ireland in Galway (Newslab.ie 2017).

Learning the basics of the data journalism cycle enables students to understand the components first, as opposed to pedagogical approaches that throw students into advanced or more technical subjects from the beginning, such as coding, without a clear sense of where these skills can take them. It is important for journalists to know how a complete data-driven story is created, even if they are not visualisation experts, statisticians or computer programmers. By actively engaging in the lifecycle of a data-driven story, they will learn better how to collaborate with programmers or graphic designers. This will also help the journalists to understand the value of each of these steps in a data-centric investigative project, even if in a professional capacity – back in the newsroom – they may not perform some of these steps in person.

Following this foundational introduction, students could proceed to studying data analytics and statistics in more depth (and these skills reflect the desired education listed by journalists participating in the Global Data Journalism Survey). A particularly effective approach to familiarising journalists with data work would be a practice-driven
training where they can put the data skills gained immediately into action, incorporating data into other investigations they may be working on professionally, or could use in their professional context.

In summary, I propose the following order for training ‘already-journalists’ in data-driven investigations:

1. Introduction to data journalism: including learning to find data, clean data, merge data, basic Excel and data analytics and basic data visualisation using tools such as Datawrapper, CARTO, Highcharts and Tableau.

2. Data analytics, including statistics. Using R as a statistical tool is my recommendation as the students not only learn to run statistics but also they learn to code using R programming language. Two in one.

3. Hands-on practical production of data-driven journalism work, such as a data journalism studio.

4 (or 5). Data visualisation and information design.

5 (or 4). Programming in a language such as Python.

After these five courses, students can continue in any area towards advanced level training.

Thus far, I have focused on ‘upskilling’ journalists in the emerging areas of data journalism. However, there are many individuals entering data journalism who do not hold a formal background in journalism. The training for these could follow a more varied path, incorporating data skills alongside training in journalism.

The ‘to-be journalist’

In contrast to the ‘already-journalist’ discussed above, these are individuals who are not trained, or do not work, as professional journalists. Regardless of where they are in their careers, they are only starting their journalism career and, in many cases, are just starting their undergraduate education. I split this group to three sub-groups, and briefly propose an approach to training for each below:

A. Starting a BA in Journalism

These individuals, in most cases, are coming straight from high school / secondary school, eager to learn everything about journalism. They are best positioned to learn data
skills alongside their training in journalism and this can be accomplished by incorporating data courses into degree requirements or electives. Suggestions for data-oriented courses include: Introduction to data journalism, Advanced data journalism, Data-driven investigative journalism, Introduction to programming, Advanced programming (or a specialisation, such as database programming, interactive/web-based programming), data visualisation, Information design, and Statistics.

B. Holds an undergraduate degree outside Journalism, but in a related area

This group is often educated in degrees that provide complementary skills to that of a journalist, such as English, Politics, and Communication Studies. Individuals in this group may be trained in studying socio-economic phenomena, or trained in non-journalistic genres of writing. The suggested training for this group would incorporate that of Group A above, alongside ‘traditional’ journalism skills (writing for the news, investigation, sourcing and verification, etc.).

C. Educated in Computer Science, Information Science or Data Science

These individuals come from a technical background: they know about methods in some, but generally not all, of the following areas, depending on their degree: data analysis, quantitative and qualitative methods, and programming. This group often lacks training in writing. What they need, then, is a Journalism degree or programme and an introduction to the tenets of data journalism so they can put their technical skills to work in the specific domain of journalism. It’s also important that statistics training is incorporated when an individual’s technical training has not included this.

Conclusion

The time to include data journalism in all journalism programmes has arrived. While many journalism programmes still do not cover data skills, many have introductory offerings on the topic and a few have more advanced offerings. As a general observation, journalists and journalism graduates lack sufficient data skills. However, they express interest and the need to further their skills in these areas.

This gap calls for a reform in journalism programmes across the world. Data journalism is an interdisciplinary field and, while I expect that we will see an increase of data offerings in journalism programmes in the upcoming years, we require a broadened approach to data journalism training. This approach must be mature enough to facilitate entrants from the various disciplines that converge to create data journalism professionals.
References


Note on the contributor

**Bahareh Heravi** is an Assistant Professor in Data Journalism at the School of Information and Communication Studies at University College Dublin (UCD). Prior to joining UCD, she was the head of Insight News Lab at the Insight Centre for Data Analytics at the National University of Ireland, Galway. She is the former Lead Data Scientist at The Irish Times, where she spearheaded the Irish Times Data. Bahareh is the Director of the Postgraduate Certificate programme in Data Journalism at UCD and has lectured on Data Journalism in several other Irish universities and institutions, including the National University of Ireland Galway, Dublin City University and Irish Times Training. She is the co-chair of the European Data and Computational Journalism Conference and the founder of Hacks/Hackers Dublin. Twitter: @Bahareh360.