

By Bahareh R. Heravi

The Version of Record of this manuscript
has been published and is available in

Journalism Practice

Date of publication: 18 April 2018

DOI: 10.1080/17512786.2018.1463167

3Ws OF DATA JOURNALISM EDUCATION

What, Where and Who?

Bahareh R. Heravi

School of Information and Communication Studies

University College Dublin

Bahareh.Heravi@ucd.ie

ORCID: <https://orcid.org/0000-0002-0587-1046>

This paper explores data journalism education, with a particular focus on formal training in the higher education sector globally. The study draws on data from: (1) The 2017 Global Data Journalism Survey, to study the state of data journalism education and the requirements in terms of training and (2) A dataset of 219 unique modules or programmes on data journalism or related fields that were curated and examined in order to understand the nature of data journalism education in universities across the world. The results show that while journalists interested in data are highly educated in journalism or closely related fields, they do not have a strong level of education in the more technical areas of data journalism, such as data analysis, coding and data visualisation. The study further reveals that a high proportion of data journalism courses are concentrated in the US, with a growing number of courses developing across the world, and particularly in Europe. Despite this, education in the field does not have a strong academic underpinning, and while many courses are emerging in this area, there are not enough academically trained instructors to lead and/or teach such interdisciplinary programmes in the higher education sector.

KEYWORDS data journalism, data-driven journalism, data journalism education, computational journalism, journalism education, journalism tools

Introduction

Journalism produces a day-to-day account of the history of societies and nations. Mass communication is a primary channel to inform the public, as well as forming public opinion for the better or for worse. In addition to writing history in a decentralised manner, journalists are the watchdogs of society and are there to hold

This document is a post-print and there might be differences between this version and the published Version of Record:
<https://doi.org/10.1080/17512786.2018.1463167>
Journalism Practice, 2018

To cite this article:

Bahareh R. Heravi (2018): 3WS of Data Journalism Education, *Journalism Practice*, DOI: 10.1080/17512786.2018.1463167

those in power to account. This means that what journalists produce has a direct impact on our lives, on our societies, and our decisions. Hence it is of utmost importance that their content is accurate and reliable, and that they depict an honest reflection of the truth, which can be consumed and understood by the public.

Data journalism is an emerging area of practice and study, which draws on knowledge from several disciplines, including journalism, information science, social sciences, data and computer sciences, data analytics, information design, and storytelling. Having been built on the foundations of precision journalism, data journalism promotes a fact-based and scientific approach to journalism. This approach calls for journalism to be treated and practised in the same way as scientific investigations, inviting scientific methods, scientific objectivity, scientific transparency, scientific reproducibility -- in general the ideals of the scientific method -- to the process of journalism and mass communication (Meyer 1973). These scientific methods include quantitative and qualitative data analysis for investigation, producing journalistic content and communicating these content to the public. This new focus on data calls for a radical reconsideration of journalism programmes across the world, to include training on such methods that equip journalists with skills required for finding the facts in today's data intensive economy, to understand them, scrutinise them, and communicate them to the public in the most appropriate and understandable manner.

The term 'data journalism' is a relatively new term, yet there are multiple definitions at play (briefly discussed in the Related Work section). Before delving into a discussion of data journalism education, I will specify my usage: I define data journalism as finding stories in data -- stories that are of interest to the public -- and presenting these stories in the most appropriate manner for public use and reuse (SiliconRepublic 2016; Heravi 2017). Similar to any other journalistic work, data journalism puts the tenets of journalism first; it is about the investigation, the story, and communication of 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 (SiliconRepublic 2016; Heravi 2017).

Data journalism has its roots in computer-assisted reporting (CAR). CAR became popular in the United States in the 1960s, and primarily referred to the use of computers -- and specifically databases -- in journalistic work. In the 1970s, a new term, "Precision Journalism", was coined by Philip Meyer, which encouraged the use of statistics and social science methods in journalism (Meyer 1973). Meyer's specific view on precision journalism was that journalists would be wrong less often if they used a scientific approach to analysing data, employing social science research methods (Meyer 2002).

"Computational Journalism" is another field of study closely related to data journalism. It is described as "the application of computing and computational thinking to the activities of journalism including information gathering, organisation

and sense making, communication and presentation, and dissemination and public response to news information” (Gynnild 2013). Computational journalism combines algorithms, data, and knowledge from social sciences to enable journalists to explore increasingly large amount of structured and unstructured information as they search for stories (Flew et al. 2012).

Data journalism, as we know it today, has been growing for the past 10 years. We are at a stage where many media organisations have data journalists in their newsrooms and organisations and/or are increasingly interested in hiring journalists with data skills. Similarly, academic and educational institutions are making concerted efforts to include data journalism in their programmes, which has lead to a surge in the number of data journalism courses and programmes in the past three to five years. The new and interdisciplinary nature of data journalism requires educators with a diverse set of skills, alongside a deep understanding of both academic and practical aspects of disciplines involved.

Despite the rapid growth of data journalism practice in newsrooms and the inclusion of related subjects in universities, there is limited research in this area. More research is required into (a) newsroom best practices, and the skills required for successful data journalism practice, and (b) the current state of data journalism education globally.

This paper is an attempt to address these gaps by identifying the skillset that data journalists (who participated in Global Data Journalism survey) currently have, and the skills the industry considers important to acquire. This paper further studies the state of data journalism higher education globally by curating and analysing a list of 219 unique data journalism modules¹ and programmes offered across the globe.

The research questions addressed in this paper are as follows:

- *RQ1: What* skills are already amongst the top skills possessed by many data journalists; **what** data-related skills are considered most important for practicing data journalism to acquire, and **which** data skills are currently taught in journalism schools?
- *RQ2: Where* is data journalism being taught across the world?
- *RQ3: Who* is teaching data journalism?

Related Work

Data journalism is defined as a “field [that] encompasses a suite of practices for collecting, analyzing, visualizing, and publishing data for journalistic purposes” (Berret & Phillips 2016, 15). Another definition by Howard (2014) puts journalism next to data science and defines data journalism as the “application of data science to journalism, where data science is defined as the study of the extraction of knowledge

¹ Please note the terms ‘course’ and ‘module’ are used interchangeably throughout this paper.

from data” (Howard 2014, 4). Howard considers data journalism as a process that encompasses “gathering, cleaning, organizing, analysing, visualizing, and publishing data to support the creation of acts of journalism” (ibid). As identified earlier, this paper defines data journalism as “finding stories in data – stories that are of interest to the public – and presenting these stories in the most appropriate manner for public use and reuse”.

Ausserhofer et al.’s (2017) analysis of data journalism literature suggests an increase of research publications on data journalism and related fields since 2010. They note that although CAR has been practised since the 1960s, the scientific investigation of it has also started only recently.

A search on Google Scholar for the term “data journalism” returned 3,150 documents in November 2017. Only 168 documents out of these 3,150 were produced before 2010. Similar Google Scholar searches on “data journalism” returned 3,120 documents in August 2017 and 2,910 documents earlier in April 2017. This shows the fast growing interest in the field: nearly 250 documents generated in 6 months between April and November 2017, containing the word “data journalism” compared with 168 produced before 2010 altogether. Similarly a similar search on the Scopus bibliographic database in November 2017 returned only 103 documents as those containing “data journalism” in their titles, abstracts or keywords. Of these 103 documents, 84 were produced since 2015 (24 documents in 2015, 35 documents in 2016 and remaining 25 documents in 2017 by November). There were no such academic publications prior to 2012.

Since 2010, there has been a growing body of studies on the practice of data journalism in various countries or in specific geographical locations. Examples are studies on these practices in Sweden (Appelgren & Nygren 2014), Norway (Karlsen & Stavelin 2014), Belgium (De Maeyer et al. 2015), Canada (Young et al. 2017), Russia (Radchenko & Sakoyan 2014), the United Kingdom (Dick 2013; Knight 2015; Hannaford 2015; Borges-Rey 2016), the United States (Fink & Anderson 2014; Parasie 2015; Parasie & Dagiral 2013), and Germany (Weinacht & Spiller 2014; Stalph 2017).

Additionally there have been a smaller number of studies, examining the specific characteristics of good data journalism. These include studying nominees or winners of various awards and competitions such as the Data Journalism Awards, e.g. (Loosen et al. 2017; Ojo & Heravi 2017; Young et al. 2017), or studies of specific news organisations data journalism and data visualisation practices, e.g. (Stalph 2017).

Data Journalism Education

Incorporating data journalism in education and training curricula is a rather new development, and only in the past five years have we witnessed a growing number of data journalism modules and programmes in university curricula, the

establishment of Massive Open Online Courses (Howard 2013), the publication of books and, in very few cases, textbooks on data journalism, e.g. (Miller 2013; Gray et al. 2012; Vallance-Jones & David 2017; Mair et al. 2013; Felle et al. 2015; Mair et al. 2017).

There are a limited number of publications studying educational needs and the state of data journalism education.

In a 2008 study on CAR, Yarnall et al. (2008) surveyed 232 journalists from 33 countries - from programmes accredited by the Accrediting Council on Education in Journalism and Mass Communications (ACEJMC) for the US programme and a convenient sampling of non-US journalism programmes -- to study their education in relation to CAR and data analysis skills. Their study revealed that in 2008 journalism programmes in the United States offered fewer opportunities for data analytics related education in their departmental offerings than their non-US counterparts. Despite this, and while there were fewer offering on such courses covering such skills in the United States, Yarnall et al. (2008) report that a higher percentage of United States institutions offer data analytics competencies to their journalism students, in comparison to their non-US counterparts. While this may sound contradictory, it suggests that a higher proportion of US Journalism schools sent their students to obtain such skills through elective courses outside of journalism schools. In their study they further revealed that a higher proportion of non-US universities had written admission guidelines for students in relation to data skills in 2008, as well as having formal proficiency tests and exams in data related proficiencies in order to graduate (Yarnall et al. 2008). This study may suggest that while a higher proportion of US students exited university with training in data related skills in 2008, a higher proportion of non-US school had data related skills formally instituted in their journalism schools' curriculum.

A survey by the European Journalism Centre in 2011 (Lorenz 2011) revealed that while journalists in Europe are eager to enhance their data skills, the lack of available training was a key barrier. The study expressed an urgent need for further education in this domain, and argued that a future journalist (in 2011) should be a researcher, programmer, and designer: three university degrees in one. Significant training is required to meet the demand for these skills.

There has been a significant shift in adoption of data journalism in newsrooms across the world since 2011, and consequently many journalism and communication schools have started to embed new modules and courses in this area in their academic curriculum. Now in 2018, data journalism education is still in its infancy, but we can observe rapid growth.

In 2016 Berret and Phillips (Berret & Phillips 2016) conducted a comprehensive Data Journalism education study, looking at 113 American Journalism programmes accredited by the Accrediting Council on Education in Journalism and Mass Communications (ACEJMC) to find Computer Assisted Reporting (CAR) and

data journalism programmes. They report that nearly half of these programmes offer no course on data journalism related subjects. Out of the 59 programmes that 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 (Berret & Phillips 2016). 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” (Berret & Phillips 2016, 9).

Splendore et al. (2016) studied educational strategies in relation to data journalism education in six European countries. While their study was on a limited number of universities (and in some cases I was not able to locate the courses listed), they report that data journalism education appears to be a very young discipline that frequently neglects some fundamental journalistic topics. They further expressed that most data journalism curricula can often be divided in curricula that focus on “(1) practical versus theoretical contents and (2) statistical and mathematical skills versus journalistic skills” (Splendore et al. 2016, 145). With regard to the practical versus theoretical perspective, they observed differences between university/higher education courses and modules and training programmes offered by other institutions. To this end university-based courses in their sample proved to have a stronger focus on a more holistic approach to data journalism, while training programmes offered by others and in particular by media companies have a stronger focus much more on specific practical skills such as data gathering, elaboration, analysis, and presentation (Splendore et al. 2016).

In terms of educators of data journalism courses, Splendore et al. (2016) found that in their sample most educators did not have a media-related background (10 out of 16 interviewed did not have such background). These educators started their studies and career in other areas such as “political science, pedagogy, and international law or even agriculture or microbiology” and became acquainted with data journalism in their higher education studies, often by researching the topic (Splendore et al. 2016, 146).

Reflecting on his experience in establishing a data journalism module in the United Kingdom, and in the context of learning to teach data journalism, Hewett (2016) argues that a smooth integration of data journalism with existing modules proves to be challenging. Observing the limited set of publications on data journalism education, he calls for further investigation in the area, and specifically for an international comparative study of different programmes that focus on or include data journalism. Similarly, reporting on two case studies that employ data journalism exercises in two postgraduate and undergraduate programmes in New Zealand, Treadwell et al. (2016) express that a set of issues, including a lack of scholarship in

data journalism education, are faced when incorporating data journalism in traditional journalism programmes.

Nguyen & Lugo-Ocando (2016) put statistics at the heart of journalism, and discuss the lack of statistical literacy amongst journalists. Highlighting that it is rather uncommon to find statistics courses in university journalism programmes, they express that “Journalists’ traditional ‘number phobia’ is not because of statistics per se, but because their nature is either vastly misunderstood or too narrowly understood” (Nguyen & Lugo-Ocando 2016, 4). This calls for demystification of statistics and numbers in journalism education.

Discussing potential academic routes a student/learner may take towards data journalism education and practice, Heravi (2017) provides a set of sample curricula for data journalism education in universities, taking into account the potential diverse background and trajectory of students. Acknowledging that data journalism education is often referred to teaching ‘data’ to ‘journalists’, this study broadens the horizon by providing examples of data journalists who came to the field from disciplines other than journalism, communications or closely related fields, such as computer science, information science and social sciences. In this study, the belief is that to train future data journalists, teaching journalism to individuals with data, technical or social sciences backgrounds could be as effective as teaching data to journalists (Heravi 2017).

The rapid emergence of data journalism in newsrooms, and the challenges this poses, calls for a review of the educational offerings for journalists. To this end this paper studies the data skills associated with and needed for performing data journalism projects in today’s competitive and fast changing newsrooms, and the available offerings globally. In order to gain a better understanding of the higher education landscape, this paper investigates data journalism courses and programmes offered in 219 unique data journalism modules and programmes around the world, and examines the educational and professional characteristics of instructors of these courses.

Method

This study draws on data from two sets of research: (1) Results from the Global Data Journalism survey, and (2) a comprehensive study of 219 unique data journalism modules and programmes around the world.

(1) The Global Data Journalism Survey

The 2017 Global Data Journalism survey is an attempt to learn about the state and practices related to data journalism globally. The survey was designed in collaboration with Mirko Lorenz, Founder of Datawrapper and Innovation Manager at

Deutsche Welle Innovation, following a set of interviews with industry experts. The survey consisted of 48 questions in 7 sections.

Following ethical approval from University College Dublin, the survey was launched on the 3rd December 2016 and remained open until the 10th May 2017. It was open to all data journalists and journalists globally, but limited to those who identify as having worked as a journalist or a data journalist in the past year.

Two hundred and six (206) participants from 43 countries participated in the survey, with 181 respondents filling it out to completion. For the purpose of analysing the results, only responses completed to the end were considered and the rest were discarded.

(2) A study of Data Journalism modules and programmes worldwide

To understand the state of data journalism education, I compiled a comprehensive global dataset of data journalism courses/modules and programmes, or closely related courses. This dataset was compiled using data collected from a variety of existing data sources, as well as a comprehensive additional search. In the following I explain this process:

As a starting point I used Dan Nguyen's crowd-sourced curated list of data journalism syllabi (Nguyen 2017). At the time of writing this paper Nguyen's dataset had 199 courses listed. Nguyen's dataset, however, lists one row for every instance of a module per year. For example if module A was taught in university B from 2013 to 2017, five rows for that specific module would be listed in his dataset. Even though this gives great detail on the instructor and syllabus for each year, for consistency I needed one row for each module or programme in a single university. To transform Nguyen's dataset to my desired format, I carefully studied all instances of every course in his dataset, chose the latest instance of a single course to remain in the dataset and deleted all other instances of that courses or programme for various years for one university/institution. In order to not to lose this information entirely, I initially added two columns to Nguyen's dataset at this stage: 'first offering' and 'latest offering'. Using these two columns I could track the history of a module/programme, where there was more than one instance for a course in Nguyen's dataset. Following this method, the number of unique courses in his dataset was reduced to 144.

As an attempt towards a more complete dataset, I added a number of other new variables (columns) to record additional information. These include the 'country' where the programme is being offered, 'level' (undergraduate vs. postgraduate), 'profession' of the instructors defined in two broad categories ('academics' or 'professionals'), and also the 'highest education level' of the module instructors/leaders. This new information was collected and added for all 144 courses borrowed from Nguyen's dataset.

While Nguyen's dataset covers a number of universities and institutions outside of the United States, his list is mainly focused on North American programmes/courses. To complete this dataset further, I added a number of data journalism programmes that I was familiar with in Ireland, the United Kingdom and Europe. Following this I added an additional set of programmes and courses from academic courses listed in Splendore et al. (2016). Splendore et al. (2016) listed 17 academic programmes in their paper. Out of these 17 I included only 10 in my dataset, and discarded another 7. The reason for this was that (a) I either was not able to find sufficient information on the programme/course through information provided in the paper, or through an internet search, or (b) I decided that the course listed was not directly relevant to data journalism, as per definition adopted in this paper, even if they had some elements of data.

The next step was to add relevant listings from the "CAR and Data Syllabi" list, maintained by the Investigative Journalism Education Consortium (IJEI n.d.). I studied all courses listed, identified those not already in my dataset, checked whether or not the listing was up-to-date, added them to the dataset and extracted the above additional information from the webpage of each listing and added to the dataset.

As an additional attempt to enhance the dataset I used an older, but still relevant list of 24 courses, curated by Daniele Palumbo in 2013 (Palumbo 2013). The last step in data collection was to conduct an extensive search on data journalism and closely related courses on Google, add the newly found courses and programmes to the list, and complete the information accordingly.

A final check was conducted on the dataset to ensure that no repeating modules or programmes had been added to the initial list of 144. The final dataset was composed of 219 unique entries, including data journalism programmes and data journalism modules, or closely related modules.

To enhance the dataset with information on the content of the courses in the list, I performed a content analysis on the names and content of the syllabi for all 219 entries in the dataset, and added a new column called "theme" to record this data in a structured manner. Given that the names of the modules were diverse, and to be able to make a useful analysis of the types of skillsets these modules were covering, I defined 17 themes and coded the entries accordingly. These themes were derived from a content analysis of titles and syllabus of the modules. These 17 themes are: CAR, coding, computational journalism, data analysis, data journalism, data science, data visualisation, digital journalism, digital media, investigative journalism, journalism, online journalism, precision journalism, research, web programming, other and unknown.

To identify the tags and code the appropriate courses accordingly, the title of the course was the first deciding factor. If the title of a course/programme was a direct variation of one of the 17 defined tags, that tag would be selected for that course. For example if the title of a course was "data journalism" or "data driven journalism" the

tag “data journalism” would be selected for that course without studying the syllabus. If the title of the course did not present a direct similarity to any of the 17 tags, the syllabus and/or structured of that course/programme was studied in detail to determine the appropriate tag.

In certain cases, a course could be tagged as two or more categories. An example is when a module focused on both data journalism and data visualisation. In such cases a hierarchy of importance was derived, and the module was tagged as the category which was considered an appropriate overarching category in this case. For example if a module focused on data journalism with specific attention to data visualisation, the module was tagged as ‘data journalism’. However, if the main focus of a module was on data visualisation, in the greater context of data journalism (with little on data analysis and other aspects of data journalism), it would be tagged as ‘data visualisation’.

The final dataset is composed of the following fields/variables: ‘id’, ‘title’, ‘theme’, ‘organisation’, ‘school/sub-org’, course listing/code, type (full programme or module), level (UG/PG), credit, start year, latest offering, homepage, instructor 1, instructor 1 highest education level, instructor 2, instructor 2 highest education level.

While a great deal of effort and attention was paid toward compiling a comprehensive dataset, by no means do I claim that I have created a complete list of data journalism courses and programmes across the world, nor do I believe such a complete list is easily achieved. In particular, despite best efforts, finding courses, or finding sufficient information on programmes in non-English speaking countries was difficult and not straightforward.

Results

A systematic approach to studying data journalism education would require learning about the skills that prospective students present, and the skills they would find most beneficial to learn. To study the data journalism skills possessed, needed and taught, i.e. RQ1, I drew on data from both the Global Data Journalism survey, as well as the study of data journalism courses globally. To this end I split this question to three sub-sections: (a) data related skills already possessed by journalists interested in data journalism, (b) data journalism skills most interesting/useful to acquire, and (c) data journalism skills taught as part of data journalism courses across the world.

(a) Data Journalism skills

As part of the survey participating journalists were asked about their educational background, existing skills, and skills they believe to be important to acquire for their future work.

While the survey was open to all journalists, 86% of participants identified as data journalist. However, in terms of data journalism proficiency, only 18% rated themselves as experts in data journalism. Another 44% identified as having a 'better than average' knowledge in data journalism and 26% identified as having average knowledge in the field. Nearly 13% of participants identified as novice or below average level of expertise in the field.

In terms of formal training, half of the participants (50%) responded that they had formal training in data journalism. With regard to a wider understanding of formal training in knowledge areas used in data journalism practices, most participants demonstrated a high degree of formal training in journalism, while they presented a lower 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. Figure 1 depicts the breakdown of formal training in various related fields between participants.

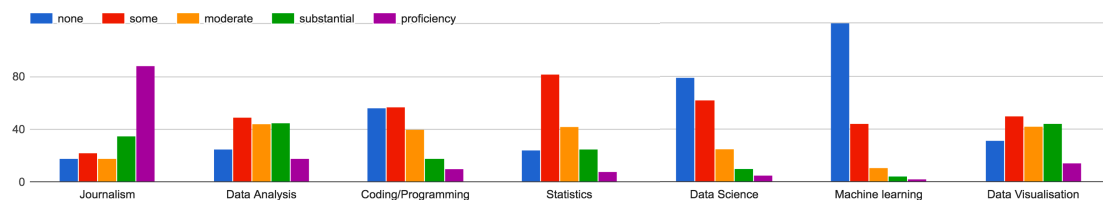


Figure 1. Level of formal training in related knowledge fields, 2017 Global Data Journalism Survey, $N=181$

In terms of education level, 97% of respondents had a university degree, with a breakdown of 40% university graduate (bachelor) level, 54% postgraduate level and 3% with a doctorate or above degree. This shows that the data journalism community is a highly educated community composed of 96% university graduates, 50% of whom have a postgraduate university degree. Looking into the degrees obtained by these participants, 62% were formally educated in Journalism at the university level. While Journalism was by far the most prevalent higher education degree amongst survey participants, this was followed by a combination of other degrees: Politics (15%), Computer/Information/Data Science/Engineering (12%), Communication and Language/Literature each 10.5%. Twenty-six per cent (26%) listed a combination of other degrees.

This shows that while most participating journalists had formal higher education training in Journalism, Communication, Politics and related degrees such as Literature, only 12% had higher education training that deals directly with data and technical topics. It further denotes that formal training between the participants seems to have been mainly obtained through higher education and university degrees, and highlights the importance of including data related courses and modules in relevant higher education Journalism and Communication programmes.

(b) Data Journalism Skills to acquire

With respect to the knowledge gap identified in the Global Data Journalism survey, perceived education requirements in this sector were examined through the data collected from the survey. A remarkably large majority of participants in the survey (98%), expressed that they were interested in acquiring further skills to practice data journalism, with 81% being **very** interested. While nearly all participating journalists were interested in acquiring further skills, merely 42% expressed that they are interested in more formal higher education degrees in this area. However, if the training offered is shorter-term or more flexible, a striking 74% of participating journalists express interest in formal training in higher education, e.g. a postgraduate certificate or higher education diplomas.

In terms of specific data skills journalists expressed interest in acquiring, data analysis presented itself as the top skill, with 64% of individuals expressing their interest in learning it. This was closely followed by learning “how to programme/code” at 63% and visualising data at 51%. 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 over 48% of journalists expressing interest in each. Figure 2 presents results from the Global Data Journalism survey on expressed interest in acquiring skills in various knowledge areas.

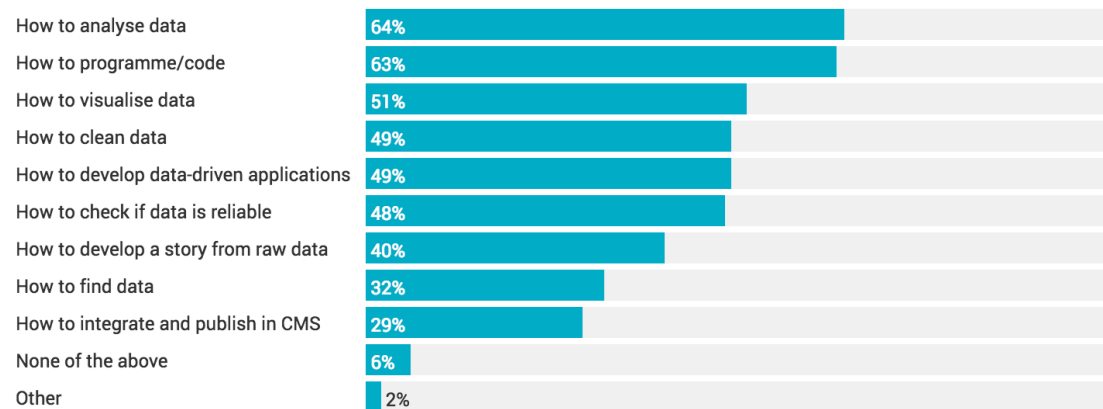


Figure 2. Interest in acquiring skills listed skills (%), 2017 Global Data Journalism Survey, N=181

(c) Data Journalism Skills taught journalism programmes

Moving from the results of the survey to data journalism courses taught around the world, to understand the types of data related courses and programmes taught across the world, I categorised the 219 modules to 17 themes. As a reminder, these 17 themes are: CAR, coding, computational journalism, data analysis, data journalism, data science, data visualisation, digital humanities, digital Journalism, digital media, Investigative journalism, Journalism, online journalism, precision journalism, Research and Web programming, other and unknown.

The results of the 219 module analysis show that most of the courses listed are, not surprisingly, focused on ‘data journalism’ as an overarching topic. These courses have a primary focus on the complete workflow of data journalism, from finding, collecting and cleaning data, to analysis, visualisation and communication. This category forms the over half (55%) of data-related courses taught in journalism programmes across the world. These courses, while often on the less advanced end of the spectrum, prepare their students to complete a data journalism project and produce data journalism output individually or in small groups, even if they are not using advanced tools and methods.

This category of courses may also be prevalent because they provide a foundation for more advanced courses in some universities. In many cases, the universities who have more advanced courses, such as coding, computational journalism or data analysis and data science related courses, list a foundational data journalism course as a prerequisite of these more advanced courses.

Computer Assisted Reporting (CAR) is the second most prevalent data-related category of topics taught in journalism programmes, with 10% of courses being tagged as ‘CAR’. The syllabi of the courses in this category are often not too far away from those tagged as ‘data journalism’, but CAR courses have a stronger focus on databases than on data analysis and visualisation. These courses often include lectures on databases, as well as SQL and other methods and tools for creating and interrogating databases. While SQL and similar database query languages are sometimes present in ‘data journalism’ modules, they often are not the main focus in that category. CAR courses tend to have less focus on data and statistical analysis than their ‘data journalism’ tagged counterparts, but that is not always the defining factor.

Except for one case (which was in Thailand, and is no longer an active course), the courses tagged as CAR are all taught in North America, with 18 being in the United States, and 2 in Canada. Those in the United States are mostly courses that came to existence before the more recent popularity of data journalism as a whole. In some cases the content of the modules under the CAR tag had changed since the time they were initiated, and are closer to the content of the ‘data journalism’ category at the present time. However, if the content was closer to ‘data journalism’ but the name was a variation of CAR, I kept the tag as CAR.

Overall, ‘data journalism’ and ‘CAR’ categories are both foundational and in many cases similar topics, and while I have put them in two separate categories, we can conclude that an overarching 65% of the 219 courses form a variation of foundational data journalism and CAR, despite the content and historical differences.

Data journalism and CAR are followed by courses which have a stronger focus on ‘data visualisation’ (7%), and then courses focused on ‘coding’ (5%) and ‘computational journalism’ (4%) respectively. The computational journalism tag is used for courses that pay strong attention to the process of data journalism, while the

main focus is on computational methods, as opposed to data analysis. This is distinguished from those tagged as ‘coding’, which are courses focused solely on teaching coding and programming, such as Python, and less so on the process of data or computational journalism (Figure 3).

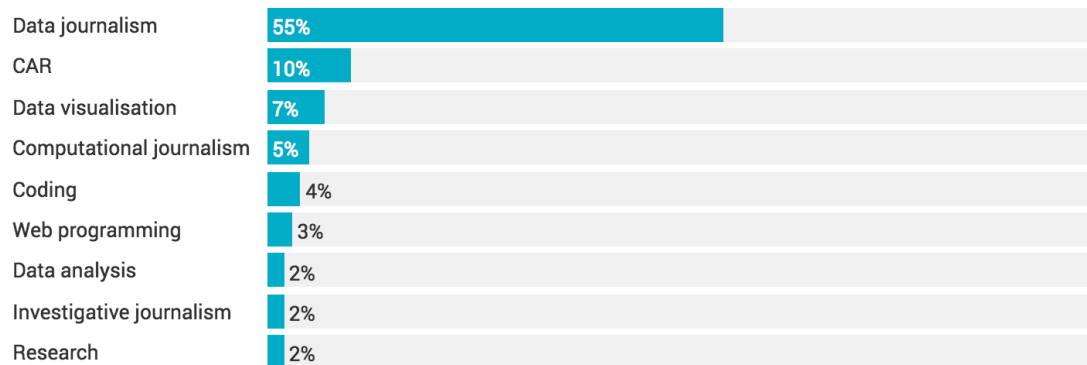


Figure 3. Themes with more than 5 courses in the dataset. Tags with less than 5 courses include precision journalism and journalism, digital journalism, digital media, data science, digital humanities, online journalism, unknown. Six courses were tagged as ‘other’. $N=219$

Courses focused primarily and only on data analysis, as opposed to the basic data analysis covered in foundational data journalism courses, form only 2% of the courses in the dataset. Putting this next to the results gathered from the Global Data Journalism survey (Figure 2), it is immediately apparent that while most journalists are interested in learning about data analysis, apart from the basic data analytics covered in foundational ‘data journalism’ courses, very few courses on more advanced data analytics are offered in journalism and related programmes. Perhaps this is the reason why most journalists feel they need to learn more about data analysis, and a reason behind journalists’ ‘number phobia’ highlighted by Nguyen & Lugo-Ocando (2016). This results, put next to former calls on the importance of statistics and data analysis in journalism, e.g. (Nguyen & Lugo-Ocando 2016), calls for an immediate attention in covering more data analytics and statistics courses in journalism and data journalism higher education programmes.

Apart from the apparent gap in courses covering ‘data analysis’, the rest of the categories and areas that interest journalists are not too far away from programmes offered on the topics across the globe.

In summary, these results of the Global Data Journalism survey with respect to training and education shows that most participating journalists had formal higher education training in Journalism and related areas, but these same journalists lacked training in data skills. Shorter and/or targeted higher education programmes appear to 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 results, are data analytics skills, followed by coding skills.

The content analysis of the topics and syllabi of data-related courses taught in journalism and related programmes indicates that courses covering a complete data

journalism workflow are the most popular courses taught around the world, and are considered foundational for any data-related specialisation in journalism and communication studies. The biggest gap and problem comes into play when it comes to courses covering more advanced data analysis: While learning ‘how to analyse data’ is listed the most pressing skills to acquire, data journalism related courses around the world overall fail to provide such courses, above the very basis statistical analysis, to their students. The second most popular skills to learn according to participating journalists was ‘coding and programming’. These topics are covered by courses tagged as ‘coding’ and in most cases courses coded as ‘computational journalism’. In comparison to the alarming gap between journalists’ appetite in learning about data analysis and offered courses on the topic, a considerably higher number of journalism programmes appear to cover topics on coding and programming, than they do on data analysis.

Where is Data Journalism being taught?

To examine RQ2, and to gain an understanding of the geographical distribution of countries and institutions involved in teaching data journalism across the world, I returned to my curated dataset of 219 data journalism modules and programmes. Immediately it is clear that the United States has the largest offerings in data journalism related modules and programmes, while in Europe only a scattered number of such courses and programmes exist: 148 of the courses are in the United States, 8 are in Canada, which, excluding online courses, leaves only 63 courses and programmes outside of North America offering data journalism related topics altogether. These new figures are plausibly in line with a study from a decade ago, which looked at CAR in Journalism schools (Yarnall et al.'s study of CAR in Journalism schools (2008)). New figures show that while the United States still has a higher (and in fact in 2018 a considerably higher), coverage of data-related skills in their journalism programmes in comparison to non-US schools, unlike in 2008, such data related skills are now an integral part of the structure of many journalism programmes in the United States, and to a smaller extent across the world in 2018.

Outside of North America, the United Kingdom, the Netherlands, Ireland and Australia are the countries with the highest number of data journalism related modules and programmes. The number of courses per country, where two or more modules/programmes are available in a country, is presented in Table 1.

Amongst these 219 modules and programmes listed in the dataset, only 25 are programmes fully and specifically on data journalism. In other words only 25 universities around the world provide degrees or programmes dedicated to data journalism. Despite this, many universities consider data journalism an important topic, and there are 153 instances of stand-alone modules on data journalism in varying non-data journalism focused programmes. The rest are online, vocational or undefined.

Table 1. Number of data journalism related courses and programmes in each country, where more than two modules or programmes are present in a country, excluding the online courses.

USA	UK	Canada	Netherlands	Ireland	Australia	Italy
146	12	8	7	6	4	4
Switzerland	China	Germany	Hong Kong	Spain	Greece	
4	3	2	2	2	2	

Note: The remainder of the countries are Chile, France, South Africa, Sudan, Sweden, Thailand and Turkey, Romania, Poland and Russia, each with one module in this dataset. Online courses are not included in this table (hence the minor difference between number of modules taught in text and in the Table). $N=219$

Overall 24 countries are present in this dataset, out of which only the United States, United Kingdom, Ireland, Germany, Canada, Spain and Hong Kong present a strong focus on data journalism as a programme of its own, with more than one module dedicated to data journalism, or having postgraduate programmes in data journalism. Out of all European countries listed in Table 1, only the United Kingdom (three universities), Ireland (one university) and Spain (one university) present a strong focus on data journalism as a self-contained programme. The rest of the countries in Europe only provide one or two courses in this area.

In terms of level of education -- undergraduate level module/programme or postgraduate level -- 48% of these modules or programmes are postgraduate courses, 35% are undergraduate courses, and the rest are undefined, used in both undergraduate and postgraduate, vocational or online courses.

Who is teaching Data Journalism?

When it comes to instructors of the above listed data journalism courses and programmes, Splendore et al. (2016) and Berret and Phillips (2016) report that many journalism programmes do not have a faculty member skilled in data journalism. To investigate this further (RQ3), I examined data on the education background of instructors of modules and programmes in the curated list of 219 modules and programmes.

Overall, where information on module instructors was obtainable from module information/department pages (149 cases out of 219), 40% of instructors present to have PhD/Doctorate level education, while 34% have Masters level, and 26% have Bachelor's level education.

Looking more closely into each country, less than a quarter (23%) of the United States' data journalism instructors have PhD level education, 23% have Masters level and another 23% BA level education. In 30% of the cases instructors were not specified. On the other end of the spectrum, Australia's higher education data journalism courses (of which there are 4), appear to be taught by PhD level instructors in 100% of the cases. Following Australia, Ireland's data journalism courses are taught by PhD level instructor(s) in 83% of the cases, with the remaining 17% by Bachelor level (professionals). In the Netherlands 43% of such courses are

taught by PhD level instructors and the rest are unspecified. In the United Kingdom 33% of the courses have PhD level instructors, 17% Masters and another 17% Bachelor level; the rest are unspecified. Canada presents the lowest degree of PhD level instructors (14%), while Masters level instructors teach 57% of courses, and 29% of courses have unspecified instructors.

It is important to note here that these figures do not specify the breakdown of education level of instructors. Instead they present what percentage of modules taught by certain education level instructors. Given the small number of cases in each country, in some cases a whole education level is referring to one single instructor, or only a couple of them, specifically in countries where data journalism education is run by a limited number of educators. For example, in Canada, out of eight data journalism courses listed, only five had the name of instructors specified, out of which three are taught by David McKie. This means David McKie's education level dictates the bulk of the Master's level educated instructors in Canada. Ireland presents a similar pattern.

While many universities, particularly in Europe, require PhD level educators for teaching university courses, and in particular postgraduate university courses, the data shows that 60% of data journalism courses are taught by instructors with Masters or Bachelor level education. This is partially due to the fact that many schools across the world, and particularly in the United States, have professionals as their instructors when it comes to data journalism education. Specifically speaking, just over half (55%) of these courses are taught by academics, 36% are taught by professionals, and the rest by individuals who are a combination of academic and professional (e.g. ex-professionals who are now academics, or individuals who work both as academics -- often in part time or adjunct positions -- and as professionals. Another contributing factor may be that due to the recent arrival of data journalism as an academic discipline, there are not yet many academics in this area with PhD level education.

Discussion and Conclusion

Data journalism is an emerging area of practice and study, which has evolved tremendously in the past ten years. Data journalism practices are rapidly becoming an integral part of many newsrooms in the last few years, and data journalism education has been increasingly finding a place in the formal structure of Journalism programmes across the globe. Despite this growth, there is little known about the knowledge sets, skills, and available trainings globally, and the ways in which journalism schools are adding data journalism in their existing curricula, or the addition of new data journalism programmes in universities. To further this knowledge, this paper conducted a study into 219 modules and programmes, which are specifically on data journalism, or considered to be close in terms of content and direction. These data were combined with results from the 2017 Global Data

Journalism survey, as part of which educational background and data journalism related skills amongst participating journalists, as well as data related skills they considered important to acquire for their future work, were studied. Two hundred and sixty participants from 43 countries participated in this study.

The results show that the data journalism community is a highly educated community, while it has its roots mostly in Journalism and Communication degrees, and less so in data/information and computer related disciplines. Most participating journalists had formal higher education training in Journalism and related areas, but these same journalists lacked training in data skills. While technical, data analytics and statistical skills do not appear to be the strength of participating journalists when compared to their journalism background, it appears that most journalists express interest in learning more advanced data skills. Shorter, targeted higher education programmes appears to 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.

Data Journalism and CAR courses are the most frequently taught data related course in journalism school, forming 65% of all courses listed in the dataset. Such courses have their focus on a complete workflow of data journalism, from finding, collecting and cleaning data, to analysis, visualisation and communication, without diving deep into more complex topics or tools. They are sometimes also a prerequisite for more advanced courses. Courses focusing on data visualisation, coding and computational journalism are the more advanced courses in the offering in the courses in the dataset. This highlights an alarming gap that while many journalists expressed lack of skills in data analysis skills, and expressed such skills to be the most important/interesting skills to learn, data journalism related courses around the world overall fail to provide such courses at a level above very basic data analysis. This finding calls for immediate attention in offering more advanced data and statistical analytics courses in journalism and data journalism programmes around the world.

The results show that while we have witnessed the emergence of full programmes focused on data journalism and related topics, only 25 offerings out of these 219 offerings were complete data journalism programmes, and the rest were modules/courses on related topics. This means that, according to this dataset, only 25 universities around the world offer complete programmes on data journalism. A smaller part of these 25 also are smaller, non-full degree, postgraduate certification programmes, and the rest are mostly Masters programme. No undergraduate programme on data journalism exists to-date, while many schools are increasingly integrating data journalism modules as part of their undergraduate education on journalism related topics.

In terms of the geographical location of data journalism related courses in the offering, North America, and specifically the United States has by far the largest

number of courses in the offering. Outside of North America, The United Kingdom, the Netherlands, Ireland and Australia are the countries with the highest number of data-journalism related modules and programmes.

In terms of the instructors of these programme, and despite these programmes being university programmes, only the teachers of 58 of these modules and programmes had PhD-level education. A mere 49 has Master's/postgraduate degree and 38 were taught with individuals with Bachelor's/undergraduate level education. This shows that while many academic/university courses are emerging in this area, not many instructors have the normal academic credential in this area, which, considering the age and (lack of) maturity of the field as an academic discipline, is not surprising. As a result, also as a common norm in journalism discipline, many of these courses are taught by professionals, or by academics with Master's level education, which is unlike many other academic disciplines such as social sciences or data and computer science. This calls for more academics to be trained and educated in data journalism and related field, to help advance the field of data journalism as an academic discipline of its own right, and expand on academic research in the area to inform the training and practices in this rapidly emerging area.

This study reiterates that the time to include data journalism in journalism programmes has arrived. We are heading to an era where every journalist would be expected to have a minimum level of data literacy, and to be able to find, write and communicate stories, rooted in reliable facts and figures.

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 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

- Appelgren, E. & Nygren, G., 2014. "Data Journalism in Sweden." *Digital Journalism* 2(3): 394–405. doi:10.1080/21670811.2014.884344.
- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiassek, S., Goldgruber, E. 2017. "The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive." *Journalism*. Advance online publication. doi:10.1177/1464884917700667.
- Berret, C. & Phillips, C. 2016. *Teaching Data and Computational Journalism*,

- Available at: <https://www.gitbook.com/book/columbiajournalism/teaching-data-computational-journalism/details>.
- Borges-Rey, E., 2016. "Unravelling Data Journalism." *Journalism Practice* 10(7): 833–843. doi:10.1080/17512786.2016.1159921.
- De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., Le Cam, F. 2015. "Waiting for Data Journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium." *Digital Journalism* 3(3): 432–446. doi:10.1080/21670811.2014.976415.
- Dick, M., 2013. "Interactive infographics and news values." *Digital Journalism* (January 2015): 1–17. doi:10.1080/21670811.2013.841368.
- Felle, T., Mair, J. & Radcliffe, D. eds. 2015. *Data Journalism: Inside the global future*. Suffolk, UK: Abramis Academic Publishing.
- Fink, K. & Anderson, C.W. 2014. "Data Journalism in the United States." *Journalism Studies* 16(4): 467–481. doi:10.1080/1461670X.2014.939852.
- Flew, T., Spurgeon, C.L., Daniel, A. (2012) "The promise of computational journalism". *Journalism Practice* 6(2): 157–171. doi:10.1080/17512786.2011.616655.
- Gray, J., Bounegru, L. & Chambers, L. eds. 2012. *The Data Journalism Handbook*, Sebastopol, CA, United States: O'Reilly.
- Gynnild, A. 2013. "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets." *Journalism* 15(6): 713–730. doi: 10.1177/1464884913486393.
- Hannaford, L. 2015. "Computational Journalism in the UK newsroom: hybrids or specialists?" *Journalism Education* 4(1): 6–21.
- Heravi, B., 2017. *Teaching Data Journalism*. In J. Mair et al., eds. *Data Journalism: Past, Present and Future*. Suffolk, UK: Abramis Academic Publishing.
- Hewett, J. 2016. "Learning to teach data journalism: Innovation, influence and constraints." *Journalism* 17(1):119–137. doi:10.1177/1464884915612681.
- Howard, A. 2013. Data journalism MOOCs offer new options for distributed learning. In: Blog, Tow Center for Digital Journalism. *Tow Centre for Digital Journalism*. <https://towcenter.org/data-journalism-moocs-offer-new-options-for-distributed-learning/>.
- Howard, A., 2014. *The art and science of data-driven journalism*, New York. <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>.
- IJEC, CAR and Data Syllabi. *Investigative Journalism Education Consortium*. <http://ijec.org/car-and-data-syllabi/>.
- Karlsen, J. & Stavelin, E. 2014. "Computational Journalism in Norwegian Newsrooms." *Journalism Practice*, 8(1): 34–48. doi:10.1080/17512786.2013.813190.
- Knight, M. 2015. Data journalism in the UK: a preliminary analysis of form and

- content. *Journal of Media Practice*, 16(1): 55–72. doi:10.1080/14682753.2015.1015801.
- Loosen, W., Reimer, J., De Silva-Schmidt, F. 2017. "Data-driven reporting : An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013 – 2016." *Journalism*. Advance online publication. doi:10.1177/1464884917735691
- Lorenz, M. 2011. Training data driven journalism: Mind the gaps. *DataDrivenJournalism.net*.
http://datadrivenjournalism.net/news_and_analysis/training_data_driven_journalism_mind_the_gaps.
- Mair, J. et al. eds. 2013. *Data Journalism: Mapping the future*, UK: Abramis Academic Publishing.
- Mair, J. et al. eds. 2017. *Data Journalism: Past, Present and Future*, UK: Abramis Academic Publishing.
- Meyer, P., 1973. *Precision Journalism: A Reporter's Introduction to Social Science Methods*, Bloomington and London: Indiana University Press.
- Meyer, P. 2002. *Precision Journalism: A Reporter's Introduction to Social Science Methods* 4th ed., Oxford, UK: Rowman & Littlefield Publishers.
- Miller, C., 2013. *Getting Started with Data Journalism Writing data stories in any size newsroom*, Leanpub. <http://leanpub.com/datajournalism>.
- Nguyen, A. & Lugo-Ocando, J. 2016. "The state of data and statistics in journalism and journalism education : Issues and debates." *Journalism*, 17(1): 3–17. doi:10.1177/1464884915593234.
- Nguyen, D., 2017. Computer-Assisted Reporting and Data Journalism Syllabuses. <https://github.com/dannguyen/journalism-syllabi>.
- Ojo, A. & Heravi, B. 2017. "Patterns in Award Winning Data Storytelling". *Digital Journalism*. Advance online publication. doi:10.1080/21670811.2017.1403291.
- Palumbo, D. 2013. Data journalism courses, a map in progress. <https://fusiontables.google.com/DataSource?docid=1EtzkyR8KOsgaHW3CkzAcRoZHIYSuuFCtLFi7xex0#rows:id=1>.
- Parasie, S. 2015. "Data-Driven Revelation?" *Digital Journalism*, 3(3): 364–380. doi:10.1080/21670811.2014.976408.
- Parasie, S. & Dagiral, E. 2013. "Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago." *New Media & Society*, 15(6): 853–871. doi:10.1177/1461444812463345.
- Radchenko, I. & Sakoyan, A. 2014. The View on Open Data and Data Journalism: Cases, Educational Resources and Current Trends BT - Analysis of Images, Social Networks and Texts. In D. I. Ignatov et al., eds.: 47–54. *International Conference on Analysis of Images, Social Networks and Texts_x000D_*. Cham: Springer International Publishing.
- SiliconRepublic. 2016. "How Is Data Journalism Changing the Newsroom?" Q&A

- with Bahareh Heravi. Burke, E. ed. Silicon Republic. <https://www.siliconrepublic.com/innovation/data-journalism-media-newsroom>.
- Splendore, S., Di Salvo, P., Eberwein, T., Groenhardt, H., Kus, M., Porlezza, C. 2016. "Educational strategies in data journalism: A comparative study of six European countries." *Journalism*. 17(1): 138–152. doi:10.1177/1464884915612683.
- Stalphy, F. 2017. "Classifying Data Journalism: A content analysis of daily data-driven stories." *Journalism Practice*. Advance online publication. doi:10.1080/17512786.2017.1386583.
- Treadwell, G., Ross, T., Lee, A., Lowenstein, J. K. 2016. "A Numbers Game : Two Case Studies in Teaching Data Journalism." *Journalism & Mass Communication Educator* 7(3): 297–308. doi:10.1177/1077695816665215.
- Vallance-Jones, F. & David, M. 2017. *The Data Journalist: Getting the Story*. Ontario, Canada: Oxford University Press.
- Weinacht, S. & Spiller, R. 2014. "Datenjournalismus in Deutschland." *Publizistik* 59(4): 411–433. doi:10.1007/s11616-014-0213-5.
- Yarnall, L., Johnson, J. T., Rinne, L., Ranney, M. A. 2008. "How Post-secondary Journalism Educators Teach Advanced CAR Data Analysis Skills in the Digital Age." *Journalism & Mass Communication Educator* 63(2): 146–164. doi:10.1177/107769580806300204.
- Young, M. L., Hermida, A., Fulda, J. 2017. "What Makes for Great Data Journalism? A content analysis of data journalism awards finalists 2012 – 2015". *Journalism Practice* 12(1): 115–135. doi:10.1080/17512786.2016.1270171.