Data Journalism Education in Canada: Challenges and Opportunities

By

Jennifer Leask

Submitted to the Faculty of Extension

University of Alberta

in partial fulfillment of the requirements for the degree of

Master of Arts in Communications and Technology

August 3, 2017

Acknowledgements

Always first in my mind, my family: my husband James, whose support for this endeavor started on a highway between Vancouver and Penticton three years ago and has never wavered. My children, Ellen and Cameron, who make everything more fun, and who see the world through a lens full of love and joy, have been instrumental in making this project successful. I am fortunate to have many strong, smart women in my life. Three who made all the difference are my mother Cathie, who showed me how to never give up; my aunt Laurie, who showed me how rewarding it is to go back to school when your kids are little, and illuminated the challenges and rewards of academic research; and my aunt Gina, who helped me focus this project through our long talks about news media and the state of the world.

My parents cheered me on throughout the MACT program, and Bill, Cathie, and my mother-in law Carol pitched in when I had to be away from home. My extended family of Leasks, McCrearys, Flynns, and Macaulays have helped in countless ways, with long chats, a spare bedroom in Edmonton, and wise advice on everything from software tools to group work dynamics.

Dr. Gordon Gow has been a terrific academic supervisor and I would like to thank him for steering this research project, and for conversations that helped me find the answers within myself. I'd also like to acknowledge the fantastic work of the MACT program staff and instructors who do a fantastic job, in particular: Dr. Tommy Barker for his enthusiastic trial by fire in communications theory; Dr. Stanley Varnhagen for his illuminating evaluation course; Dr. Rob McMahon for being a supportive sounding board and creative instructor; and Dr. Megan Lefebvre for keeping in touch throughout this study and for presenting research methods in a way that inspired me, stuck with me, and was fun.

One of the great strengths of the MACT program is the cohort model. In May 2015, I walked into a room at the University of Alberta filled with strangers from all walks of life and I left with a group of friends, colleagues, and mentors who I know I can reach out to anytime.

There are a few researchers who I would like to thank for giving me insight into the field including Dr. Louise Yarnall and Dr. Sharon Dunwoody, who graciously dug into their archives to inform my research design. Dr. Mary Lynn Young and Dr. Alfred Hermida at the University of British Columbia School of Journalism took the time to chat with me about their research and share their ideas. I would like to thank all the participants who gave their time to this research, and were enthusiastic about it from the start.

I would like to recognize the friends and colleagues in my life who have been instrumental in supporting me through this study, and who listened patiently as I worked through my ideas about grounded theory, Big Data, and the comprehensibility of data journalism for the audience. Finally, my colleagues in journalism both at the Canadian Broadcasting Corporation and elsewhere, with whom I have had the good fortune of working with over the years: you tell tough stories and hold power to account, and I hope in these pages you can find something which will help you keep it up.

Abstract

Data journalism is transforming news work, however there is limited research on how those profound changes are affecting journalism education, particularly in Canada. This exploratory study examines how data journalism is impacting practice, how the shift is translating into journalism schools and what skills are essential for a journalist to tell more quantitatively oriented stories. Interviews with key informants, instructors of data journalism at Canadian post-secondary institutions, were analyzed using a qualitative iterative analysis approach. The findings include the shift in practice in which data journalism increases a sense of control for the reporters both in their ability to choose the stories they tell and by getting them closer to the source of information. Key informants expressed data journalism sets reporters apart from their colleagues, but also encourages collaboration with other journalists and outside experts, leading to a more active voice and less reliance on attribution in storytelling. The lack of a concrete approach in journalism schools to incorporating data literacy, numeracy and statistical and computer language learning was noted as a concern. A typology of beginner data skills for all journalists, intermediate skills for those interested in data journalism and advanced skills for specialization on offer outside of a traditional journalism school setting was produced as a road map to inform educators and newsroom leaders of a path to cultivating journalists who are better able to leverage data storytelling tools for the public good.

Acknowledgements **0**

Abstract **3**

Figures and Tables **6**

Introduction **7**

Literature Review **12**
- **Literature Review Methodology** **13**
- **Discussion** **17**
  - Origins of data journalism 17
  - Theoretical perspectives and research methods 18
- **The Changing Role of the Journalist** **20**
  - Transparency 20
  - Epistemological changes 21
  - Big Data and its ethical implications 22
- **The Gaps Between Emerging Journalism Practice and Journalism Learning** **23**
  - Investigating and finding data 24
  - Statistics and numeracy 25
  - Data visualization and presentation 26
  - Computer science knowledge 26
- **Data Journalism Education Strategies** **28**
  - Hands-on learning 28
  - Collective, collaborative learning 29
  - Technical obstacles 30
  - A Canadian perspective 30
- **Summary** **30**

Methodology **31**
- **Introduction** **31**
- **Design** **31**
- **Participants** **33**
- **Setting** **36**
- **Instrument** **37**
- **Questions informed by literature** **38**
  - Research influences 39
  - Theoretical influences 40
- **Procedures** **41**
  - Notes 42
  - Transcription 42
  - Anticipating roadblocks 42
- **Data Analysis** **43**
  - Coding process 44
  - Validity and reliability 45
- **Summary** **47**

Analysis and Discussion **47**
- **Findings** **47**
- **RQ1: How is the development of data journalism changing journalism practice?** **49**

A shift from attribution to evidence 49
Anecdotes and trends 51
Giving up control to the audience 52
Data journalism: set apart from the pack 52
Collaboration and competition 53
RQ2: How is the development of data journalism impacting journalism education? 54
Can everyone do data journalism? 54
Self-guided learning 56
The classroom experience 57
Motivation 58
RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned? 58
Basic data skills for all journalists 59
Intermediate skills 61
Advanced skills 62
Data Analysis 63
Procedure 63
Validity and reliability 66
Discussion 66
RQ1: How is the development of data journalism changing journalism practice? 68
RQ2: How is the development of data journalism impacting journalism education? 70
RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned? 73
Limitations of the Study 77
Summary 78
Conclusion 79
Summary of Findings 80
Findings in Context 82
Future Direction of Research and Limitations of Study 83
Conclusion 85
References 86
Appendix A: Search Terms Used 94
Appendix B: Personalized Recruitment Email 95
Appendix C: Journalism Schools in Canada Offering Data Journalism Courses 96
Appendix D: Confirmation of Course Offering 98
Appendix E: Information and Consent Letter 99
Appendix F: Interview Script Illustrating RQ Relevance and Influences 102
Appendix G: Transcription Key 107
Appendix H: Codebook 108
Appendix I: Evolution of RQs 118

Figures and Tables

LIST OF FIGURES

FIGURE PAGE

1. Infographic timeline ……....11
2. Timeline illustrating the date of publication of literature review studies ……...15
3. Google map of journalism schools offering data journalism ……..35
4. Coding frequency by number of times used……63
5. Coding frequency by saturation of sources……64
6. Observations generated by RQ1...……...69
7. Observations generated by RQ2...……...71

LIST OF TABLES

TABLE PAGE

1. Range of participant experience……..48
2. Basic skills for all journalists……...73
3. Intermediate data journalism skills……..74
4. Advanced skills learned outside journalism schools.....75

Introduction

Data journalism is something you have probably already seen: maybe it is an infographic which combines pictures and structured numbers, or an investigative project driven by data collected by government. Data journalism could be presented as an interactive map illustrating locations relevant to a feature story, or a heat map showing the voting results in the last election. As the amount of digital data being collected, shared, and stored is exploding exponentially, journalists are using computational tools which have lately become less expensive to newsrooms and more accessible to use. The practice of journalists using these tools to make sense of the world for an audience has developed quickly, unevenly, and seemingly without any observable structure. The genre of data journalism, which includes using structured data to tell stories through visualizations, interactive sites, and other computational methods, has been expanding in recent years. Against this backdrop, this study will explore what data journalism means for news outlets and the post-secondary institutions who train reporters.

Telling journalistic stories has shifted over the years from print, to radio, to television and online, but the core purpose of journalism has remained unchanged: relaying what is new and why it matters to the audience. Data journalism sits at the intersection of communications and technology: news consumption habits are evolving as audiences read more news online, glean meaning from visuals and skim stories for significance. How journalists adapt to these new audience behaviours has implications for how people understand the news. Further, the success of this adaptation may affect the value the public attaches to news in a world where journalistic authority is increasingly under threat by stories of fake news, and changes caused by the dissemination of news through social media platforms. At a moment in history when newsrooms are facing increasing financial pressure to produce more with less, it is important data journalism be practiced in a way that creates accurate news projects that increase transparency and bolster

credibility. However, the underlying skills of accessing, cleaning, and analyzing data in order to come to a journalistic conclusion are not always straightforward to learn. Without training in statistical concepts, numeracy, and data literacy, or without an understanding of how to interpret numbers as visuals, mistakes can be easy to make.

Taking all of this into consideration, my capstone project was designed to answer three questions for the future of journalism:

RQ1: How is the development of data journalism changing journalism practice?

RQ2: How is the development of data journalism impacting journalism education?

RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned?

As will be outlined in the literature review, there has been some academic research into these questions, but there is a gap in literature which looks at how data journalism skills are being introduced in Canadian journalism schools. This study sets out both to add to the literature and to provide a practical framework of use to journalism educators as they plan for the introduction of data journalism concepts in their programs. There are limitations in this study: the literature review has a selective scope and is not intended to be comprehensive; interviews were limited to a relatively small but purposefully selected sample of key informants in the area of data journalism education; and the study does not investigate the positioning of data journalism in post-secondary institutions by including administrators and students as research subjects. However, all of these limitations were addressed in each stage of the study, as will be detailed specifically in each chapter.

The study design was informed by a structured literature review on data journalism adoption including case studies of how it is being adopted in the workplace and introduced in the

classroom. Literature that defines data journalism practice, explores shifting newsroom norms, and calls for improved approaches in the areas of journalistic transparency, algorithmic accountability and collaboration were instrumental in my research design. Of particular note, Coddington's (2015) typology, work by Meredith Broussard (2015, 2015a, 2016) on the practical challenges facing the field of data journalism, Borges- Rey's (2016) and Lewis and Usher's (2014) work on collaboration, and Hermida &Young's (2016) and Tabary, Provost, & Trottier's (2015) research into the field in Canada. Two theories were influential in the research design: Rogers' *Diffusion of Innovations* (2003) and *Knowledge Gap Theory* (Tichenor, Donohue, & Olien, 1970). These theories will be explored further in the methodology and analysis chapters. The work of Berret & Phillips (2016) into the state of data journalism education in the United States provided a framework for my study and is essential reading for any educator interested in introducing data journalism to the classroom.

After carefully considering the literature and how this study would contribute to the field of data journalism education in Canada, an exploratory design was selected as the most suitable approach to answer the research questions and to lay the groundwork for future research. After surveying the data journalism training on offer at post-secondary institutions in Canada, key informants who have worked in the field and teach stand-alone courses were selected and interviewed. While I considered including both a survey for the key informants and an analysis of the course outlines, the interview approach was deemed much more likely to produce rich information and to generate insights that I had not necessarily considered, both of which are important in an exploratory study.

The study will begin with a comprehensive literature review, exploring the state of data journalism, its influence on the workplace and the classroom, as well as the implications for

journalism in the future. Next, the research design and methodology chapter will lay out how the study was conducted, including the selection criteria for key informants and the development of the semi-structured interview guide. The findings and discussion chapter provides insight from the interview analysis and proposes a typology for data journalism skill acquisition, while offering practical information for journalism educators. Next, an overview on the limitations of the study will be presented. Finally, the concluding chapter will summarize what was learned from the research.

This study will reveal the ways in which the adoption of data journalism is shifting the role of the journalist, how that is having an influence on post-secondary journalism programs, and will offer a roadmap of essential skills. This research will contribute to the field by enabling journalists, educators, and students to have a better understanding of how data journalism is changing their field, and how to adapt to the shift. In order to begin that research, the literature that has already been produced by scholars in the field was identified, analyzed, and incorporated into my research. Those in charge of training and hiring data journalists are currently debating if it is better to teach a journalist data skills or to teach someone with a range of data skills journalism. This study does not endeavor to answer that question definitively, but it will inform the opinions of those who are asking the question by laying out a roadmap on how journalism is changing and what skills are needed to meet those challenges.

A note on illustrations used in this report: data journalism tools like Excel, Tableau, Google Fusion Tables and others used for data visualizations are now easier to use and cheaper to access than in the past. Throughout this study, I have used some of these tools to present my research as examples to demonstrate this point, as in Figure 1 below.

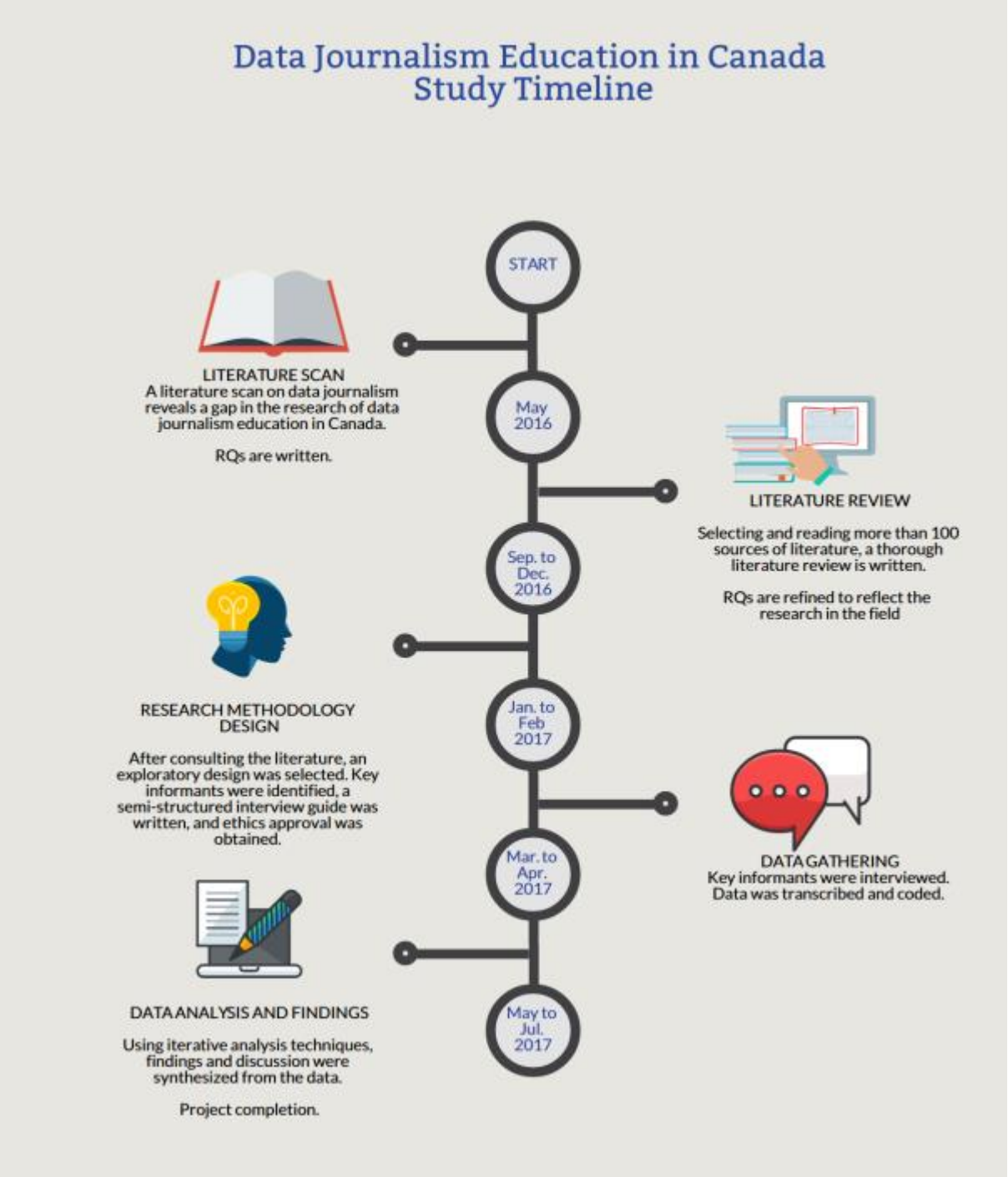

Figure 1. Using easyl.ly an infographic timeline was created to illustrate the process and timeline followed in this study.

Literature Review

Unlike other more established fields of research, the literature surrounding data journalism has only recently begun to include discourse on typology, theoretical perspectives and epistemologies. The terms "computational exploration in journalism" (Gynnild, 2013), "quantitatively oriented journalism" (Coddington, 2015), "data-driven journalism" (Parasie & Dagiral, 2013), and "data journalism" (Bradshaw, 2014; Knight, 2015; Splendore, Salvo, Eberwein, Kus, & Porlezza, 2015; Tabary, et al., 2015; Berret & Phillips, 2016; Treadwell, Ross, Lee, & Lowenstein, 2016) are used in the literature. For the purpose of this review and my capstone project, I will use "data journalism" which is the preferred term in industry (Gynnild, 2013), and is used in the majority of the academic writing surveyed. In their report for the Columbia School of Journalism on the state of data journalism education, Berret & Phillips (2016) define data journalism as:

> using data for the journalistic purpose of finding and telling stories in the public interest. This may take many forms: to analyze data and convey that analysis in written form, to verify data found in reports, to visualize data, or to build news apps that help readers to explore data themselves. This field also encompasses the use of computation—algorithms, machine learning, and emerging technologies—to more effectively mine both structured and unstructured information to find and tell stories (p. 9).

This chapter will begin with an explanation of the literature review search methodology, including the rationale for how research questions were developed for the literature review, my sampling strategy, data classification process, and the reasoning behind the inclusion and exclusion of literature. Next, a robust discussion section will provide a broad understanding of

the literature surrounding the history and development of this specific type of journalistic practice, some of the current thinking around theoretical perspectives, as well as an examination of the fundamental struggles between schools of thought on how Big Data is changing the role of the journalist in society. This section will also include a review of the literature on applied journalism practice, which explores algorithmic and automatic journalism, and the development of new ethical frameworks. After laying the groundwork for an understanding of the wider context of data journalism, the review examines the research into the breadth of perspectives on what are deemed important skills for journalists working with data, as well as the pressures faced in introducing those skills in a journalism classroom. Many of the skills acquired by data journalists are self-taught, or learned in collaboration with computer scientists, so an assessment of the literature exploring current resources available to journalists throughout their careers was performed. In concluding, the gaps in the literature are highlighted as well as how these gaps are explored.

Literature Review Methodology

Setting out to review the current state of the literature, three RQs for the literature review, separate from the study RQs, were developed to improve understanding of how the role of journalists is shifting through the adoption of data journalism skills. The three literature review RQs were:

LRQ1: How is more access to open data, Big Data and the tools to visualize information changing journalism practice?

LRQ2: How is data changing the skills a journalist needs?

LRQ3: How should those skills be learned?

A systematic library search protocol was developed to answer these questions. This literature review intends to give the reader a snapshot into current thinking and research, so the search timeline was limited to articles from 2005 onwards. The literature states the practice of data journalism began to increase in the late 2000s (Coddington, 2015), with data journalism projects perceived as having a greater impact later in that decade. For example, often cited ground-breaking projects include *The Guardian's* MP database project in 2009 (Flew, Spurgeon, Daniel & Swift, 2012; Aitamurto, Sirkkunen, & Lehtonen, 2011; Felle, 2016) and *The New York Times* War Logs project in 2010 (Knight, 2015; Hewett, 2016; Splendore et al., 2016). While some literature older than ten years was included, the vast majority of the literature available and selected is from 2010 onwards, reflecting both the recent increase in academic interest in the area and the coinciding expansion of journalistic practise. In order to illustrate the recent increase in research in the area, an interactive timeline was built using Time Line Curator, a visualization tool created to scrape sites and extract dates for timelines from text. In the visualization the publication of studies in my literature review can be seen by date in Figure 2.

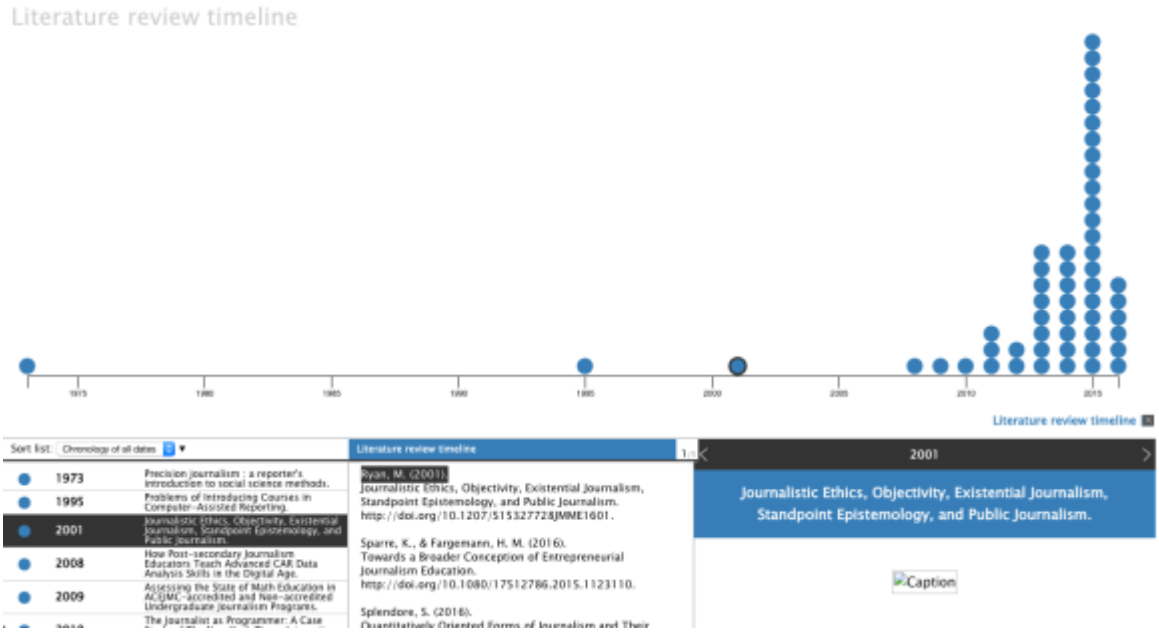

Figure 2. An interactive timeline of the literature can be found at [*http://www.cs.ubc.ca/group/infovis/software/TimeLineCurator/tlcExport/?tl=Literaturereviewtimeline*](http://www.cs.ubc.ca/group/infovis/software/TimeLineCurator/tlcExport/?tl=Literaturereviewtimeline)

Beginning with the search for terms like "data journalism," "data journalism and education," and "computer assisted reporting," and focusing the search in the *Communications & Mass Media Complete* database, several articles and authors were identified as being highly relevant to this study. The special editions of *Digital Journalism* (2015, Vol.3, Iss. 3) and *Journalism Practice* (2016, Vol. 10, Iss. 7), which specifically focused on the changes in journalism created by computational trends, were used to further hone the search terms further. The keywords were then organized into a spreadsheet where the 40 most relevant and frequent terms were identified, and from this list, 18 keyword searches using Boolean logic were identified as the most likely to generate literature relevant to the RQs (see Appendix A).

The search was conducted both through the University of Alberta's Library portal and Google Scholar. Approximately 100 peer-reviewed articles, books, and non-peer reviewed sources were identified in my search. Approximately 75 of these sources were put into a matrix where they were classified according to, when applicable, their academic rigor, theoretical perspectives, research methodology, and ability to answer the RQs. The literature considered most relevant is included in the discussion section below.

It is acknowledged there are limitations to this method because much of the discourse on learning skills exists in online-only formats, and is not exclusively in peer-reviewed journals. The most striking example of this is that this search protocol did not reveal the report by Berret and Phillips (2016), which I discovered much later in the literature review process. Other resources missed by my methodology include the National Institute for Computer Assisted Reporting (NICAR) listserv[1] for members of Investigative Reporters and Editors which has many discussions which are useful to practitioners. Many online-only teaching resources and courses could have been surveyed for further information on data journalism pedagogy, for example the massive open online course on data journalism offered by the University of Hong Kong. Due to the fluid and inconsistent nature of these sources which may change each year, and are available only to those who sign up, they were deemed outside the scope of this review. There are very few of what would be considered traditional data journalism textbooks, however the groundbreaking text by Philip Meyer (2002) on social science methods in news, as well texts on statistics and numeracy in news (Cohn, Cohn Runkle, & Cope, 2011) and data literacy (Herzog, 2016) were noted. There are a few data journalism texts which lay out a more formal approach to skill acquisition *Computer-Assisted Reporting: A Comprehensive Primer (*Vallance-

[1] https://www.ire.org/resource-center/listservs/subscribe-nicar-l/

Jones, Pilhofer, Dowdell, & McKie, 2009), *The Data Journalism Handbook* (Gray, Bounegru & Chambers, 2012), and *The Data Journalist* (Vallance-Jones & McKie, 2017).

It should be noted that my background as a journalist has shaped my RQs: for example, articles which examined this phenomenon through Actor Network Theory were given less weight than articles which explored the practical side of data journalism because they were deemed to be less likely to offer insight into current journalism practices.

Discussion

Origins of data journalism

In order to answer **LRQ1**, it is important to examine the origins and continuing development of data journalism. Journalists have always used data in reporting, but there have been noteworthy historical shifts. For example, the relative values of a "record versus a report" (Anderson, 2015, p. 361) whereby journalism moved away from the use of government documents as primary sources to a reliance on interviews in the mid-nineteenth century (Schudson, 1995, in Anderson, 2015) as journalism changed from a written to an oral form. Some literature states the development of data journalism as a distinct genre emerged in 1952 when CBS used Computer Assisted Reporting (CAR) to predict the outcome of the US Election (Bounegru, 2012, in Felle, 2016). The earliest journalism pioneer is widely considered to be Phillip Meyer. He was an innovator at using sampling and correlation methods as a reporter at Knight Newspapers (Cohen, Hamilton, & Turner, 2011). In 1967, he worked with two scholars from the University of Michigan to analyze survey data on the social profile of participants in the Detroit riots and found that contrary to the assumption that "low education was widely believed to be behind riots" (Nguyen & Lugo-Ocando, 2015, p. 10.), the numbers showed "college educated people were just as likely as dropouts to take part" (Nguyen & Lugo-Ocando, 2015, p. 8). He won a Pulitzer Prize for his work and later started writing about how to apply social

science methods to journalistic inquiry. Meyer's book *Precision Journalism: A Reporter's Introduction to Social Science Methods* (2002) was first published in 1973 and is often cited. In the fourth and newest edition, Meyer (2002) explains why a new set of social science skills is required for today's journalist:

> The world has become so complicated, the growth of available information so explosive, that the journalist needs to be a filter, as well as a transmitter; an organizer and interpreter, as well as one who gathers and delivers facts…a journalist needs to be a database manager, a data processor, a data analyst (Meyer, 2002, p. 1).

Coddington's typology of "quantitatively oriented journalism" (2015) is useful to give an overall sense of the areas of journalism work included in data journalism. He divides the field into three areas: computer-assisted reporting (CAR), data journalism, and computational journalism. Computer-assisted reporting (CAR) includes " data gathering and statistical analysis descended from Meyer's precision journalism" (Coddington, 2015, p. 334). Unlike the catch-all term used in industry, data journalism is defined by Coddington as journalism which can remove the need for data gathering through methods like interviews and instead tells stories only through "examination of the data" (Coddington, 2015, p. 339). Finally, computational journalism is defined by Coddington (2015) as a collective application of computational thinking and networked activities which is more concerned with producing a platform, rather than a narrative, per se. The skills needed to perform computational journalism are more likely to include programming skills and the work can and does exist in other professions outside of traditional journalism roles.

Theoretical perspectives and research methods

A general lack of literature which applies theoretical perspectives to data journalism was noted, as most of the literature is concerned with the practical implications of the adoption of

data journalism to newsrooms, and no studies with theoretical applications to data journalism education were identified.

The most dominant paradigm in the literature surrounding data journalism is Actor Network Theory (ANT) (Splendore, 2016). Some argue ANT is less of a theoretical lens and more of a "particular orientation toward news production that allows us to ask interesting questions in new ways" (Anderson, 2015, p.352). Others believe ANT can interpret the changing dynamics of journalism and its relationship to technology (Primo & Zago, 2015). Seeing a gap in the application of ANT to the study of institutional news production, Lewis and Westlund (2015a) advocate for a move towards a "sociotechnical emphasis" (p. 20) to examine not only the role between human actors and technology, but also to take into consideration the relationships between business, editorial and technological actors in news work. For example, a journalist may see a news audience as a passive receiver, while a businessperson may see it as a commodity. This framework departs from ANT as it does not give technological and human actors equal weight, but rather views technology tools as "inscribed and instructed by humans, socially constructed to suit journalistic, commercial, and technological purposes within news organizations" (p. 24).

A research gap highlighted in the literature is what is described by Broussard as the lack of a "systems paper" (Broussard, 2016, p. 267) for computational journalism scholars. She suggests communications scholars adopt the format from computer science to "facilitate methodological conversations around code" (p. 269) which should, in turn, allow for more accountability and transparency in algorithmic journalism, as called for by Diakopoulos (2015).

The Changing Role of the Journalist

There are two parts to the literature on the craft: the journalist's evolving attitude towards data, and the changing technological tools which journalists use to acquire, interpret, and illustrate data.

Transparency

In the literature, several researchers take the position that the adoption of data journalism practice adds transparency to the journalistic process. When journalists open up the news gathering process, combining the "front-stage" (gathering and processing) and "back-stage" (distribution and presentation) (Karlsson, 2011, p.282), journalists may lose control of the narrative (Karlsson, 2011) while attempting to increase transparency. This friction between opening up the journalistic process in the service of transparency and reinforcing the role of the professional journalist can be seen in the example of *The Guardian's* MP expense reporting whereby it released a data set of expenses on its website and asked readers to analyze the documents and suggest story angles (Rogers, 2003 in Hammond, 2015). While including crowdsourcing gives more muscle to the reporting and is meant to increase transparency, these kinds of efforts may also diminish claims of journalistic authority and the sense of need for professional journalists (Hammond, 2015). Without a journalist to interpret what is important about the data, or to give it context, "the journalist is effectively offloading the responsibility of understanding the data's significance onto the public" (Lesage & Hackett, 2013, p.44), which may lead to a further erosion of trust. The idea that journalists are not objective, but numbers are (Lesage & Hackett, 2013; Splendore, 2016), may also lend a false sense of certitude to data journalism stories. This may prove problematic as much of the information used in these kinds of projects comes from institutional sources which may not be objective in the collection of the data (Parasie & Dagiral, 2013; Tabary et al., 2015). Further, the weakness of statistical reasoning and

numeracy education in most journalism schools is well documented (Bradshaw, 2014; Cusatis & Martin-Kratzer, 2009; Dunwoody & Griffin, 2013; McConway, 2015; Nguyen & Lugo-Ocando, 2015) and without the ability to verify the numbers provided by elites, journalism risks losing its mission as a verification discipline in the long term (Nguyen & Lugo-Ocando, 2015).

Epistemological changes

Journalism, as a field of professional practice, has a distinct epistemology or way of thinking about knowledge, which adheres to certain ideals like objectivity and double-sourcing (Lewis & Westlund, 2015b). The link between these ideals and social science methodologies was expressed by Ryan (2001) in that "objective journalism shares the core values of the scientific method. The overarching value for the objective journalist (or scientist) is the collection and dissemination of information that describes reality as accurately as possible" (p.3). Knowing this, it has been advanced in the literature that one of the biggest challenges in the development of data journalism is shifting the views of the people working in the field; journalists have found it difficult to move away from established news values (Lewis & Westlund, 2015b). At the *Chicago Tribune*, a division between journalists and so-called "programmer-journalists" (Parasie & Dagiral, 2013) was observed. Programmer-journalists come from more of an open government/open software mindset and believe "news in itself should be viewed as computer-processable data, and not only as a story hidden in the data," (Parasie & Dagiral, 2013, p. 862) and that interpretation is unnecessary as long as the data is "granular and complete" (Parasie & Dagiral, 2013, p. 863). By contrast, more traditional journalists felt it was their job to interpret data for an audience by applying social science methods.

These competing epistemological lenses among people working in journalism can also be seen in projects like *Everyblock.org* in which developer and journalist Adrian Holovaty proposed the audience should be given data to "create their own news" through tools which

generate data sets based on a person's zip code (Gynnild, 2013, p. 721), or what Splendore (2016) views as a shift from the idea of journalism as a public service to journalism as an individual service.

It can be further noted that up until now, journalists have been more concerned with what is unique in the world, rather than what is common (Anderson, 2015). If journalists adopt more of a statistician's view, one which is more interested in the averages than the extremes (McConway, 2015), the question arises of how news narratives will reflect the change. This may be difficult to gauge as it has been determined that the adoption of data journalism in newsrooms remains a specialist set of skills, (Hermida & Young, 2016; Royal, 2010; Tabary et al., 2015; Young & Hermida, 2015;) and the wholesale adoption of these attitudes is slow to permeate.

Big Data and its ethical implications

Before beginning a data journalism project, the literature generally calls for a basic understanding of where the data comes from, and advocates for critical thinking about the methods and ethics of the data collection. Tabary, et al. (2015) found limitations in the objectivity of public data available to journalists in Quebec because it is mostly aggregates provided by institutional sources. Parasie & Dagiral (2013) found:

> reporters believe that since data are produced by officials, their categorization might be manipulated. This is why they have long been reluctant to use data from city officials – especially the Chicago Police department, which they have generally suspected of manipulating data (p. 860).

However, Young & Hermida (2015) found the *Los Angeles Times* directly used coroner's office data for their algorithmic *Homicide Report* project. The literature is lacking in the area in terms of how ethics in this area are taught at journalism schools, so it is difficult to conclude how these attitudes are shifting, if at all.

The Gaps Between Emerging Journalism Practice and Journalism Learning

In order to answer **LRQ2,** an exploration of the literature around journalism education that has given priority to traditional skills was undertaken. In their Columbia Journalism School report on the state of data journalism education in the US, Berret & Phillips (2016) write the adoption of computers in journalism education was split into data and digital skills: "Early calls for journalism schools to adapt to changing technological conditions were answered mainly with the addition of digital classes—learning how to build a web page, create multimedia, and curate content" (p. 24).

However, for those schools that do push forward with a data journalism agenda, a fairly standard core skill set including "critical thinking, primary research and storytelling" (Toughill, 2012, p.37) has been taught for decades, along with traditional professional news values including "thoroughness, accuracy, curiosity, independence, and transparency" (Gillmor, 2016), interviewing (Graham, 2015), collaboration (Hewett, 2015), and analytical thinking (Donsbach, 2014). Overall, Berret and Phillips advocate for an education which "reflects a mission to find and tell stories in the public interest—as well as develop partnerships and collaborations with other disciplines" (p. 28). Some also call for an adoption of skills which, though less traditional in scope, would give an edge to a journalist working in the modern media environment. These include entrepreneurship (Gillmor, 2016; Pavlik, 2013; Sparre & Færgemann, 2016), a better comprehension of statistics (Cohn, Cohn Runkle, & Cope, 2011; Nguyen & Lugo-Ocando, 2015), numeracy (Dunwoody & Griffin, 2013; Herzog, 2015), and coding literacy (Broussard, 2015a; Treadwell et al., 2016). Most of the literature centers on traditional journalism school learning, but for many journalists, learning takes place throughout their careers, and there is only peripheral study of this in the literature.

Many journalism schools are trying to prepare students for a career in mainstream media, and while the demand for students with computational journalism skill sets has been established (Hermida & Young, 2016), these same students may not be exposed to work that uses these skills in the context of internships (Hewett, 2015) and thus may not be as interested in them. Cook (2015) established in her research that students who have an aptitude for computer science work would generally be attracted to a computer science program. By contrast, Hewett (2015) found journalism students may not have a good sense of why computer science is beneficial to them and would be less likely to push for or pursue this kind of learning.

Investigating and finding data

Most of the literature skims both the traditional and computer-oriented tools which can lead a journalist to rich source material and is more concerned with analyzing and presenting data. While there is some discussion as to the objectivity (Lesage & Hackett, 2013) and over-reliance of institutional data (Tabary et al., 2015), very few sources in the literature discuss research or methods by which computer programming tools can be used to find data outside of open data repositories with a few exceptions. Three texts, however, stand out in this regard: *The Data Journalist: Getting the Story* (Vallance-Jones & McKie, 2017), *The Data Journalism Handbook* (Gray, Bounegru & Chambers, 2012), and *Computer Assisted Reporting: a comprehensive primer* (Vallance-Jones, et al., 2009), include sections on Freedom of Information requests and advanced internet search methods. Broussard's (2015) paper outlining the prototype of her artificial intelligence software, the "Story Discovery Engine" (Broussard, 2015, p. 815), aims to help investigative journalists sort through data in order to find storytelling opportunities. Hamilton and Turner (2009, in Broussard, 2015) write "the future of watchdog journalism may be found in using algorithms (precisely defined problem-solving procedures) for

accountability" (p. 815), and Broussard's paper explores this notion as well as other challenges and tools for investigative journalists working with data projects.

Statistics and numeracy

The literature identified above describes the shifting epistemologies, key debates and emerging questions of what has become a data-rich environment for journalists. This review now turns to the implications for journalism education. An examination of the challenges holding this education back from formal journalism classrooms and in-career professional development is warranted. The literature establishes that the incorporation of math, numeracy and statistical education in journalism schools is mostly perceived negatively by both instructors and students (Dunwoody & Griffin, 2013; Griffin & Dunwoody, 2015; Hewett, 2015; Nguyen & Lugo-Ocando, 2015; Sparre & Færgemann, 2016). As far back in the literature as 1995, attitudes have stayed relatively static with feelings of apathy and intimidation being problems listed by faculty when introducing data journalism skills in journalism schools (Lee & Flemming, 1995). This "math-phobic" (Dunwoody & Griffin, 2013, p. 535) attitude in journalism students is reinforced in the literature by the example of a undergraduate-level course in New Zealand: when journalism students enrolled in a data journalism course found out they would have to take a math examination, five of the 20 enrolled dropped out (Treadwell et al., 2016). Griffin & Dunwoody (2015) surveyed journalism department heads twice, in 1997 and 2008, to assess shifts in attitudes towards the inclusion of statistical reasoning in the curriculum. Of note, SAT scores from journalism students showed higher actual math aptitude than was appreciated by faculty. The authors determined that over the decade, very little had changed in the reluctance to teach statistical reasoning, and that this stemmed from the belief among educators that students were neither capable nor interested in learning the concepts. The authors did find an increase in the perception of the value of such an education by faculty over time and after their quantitative

analysis, Dunwoody & Griffin (2013) conclude there exists a "self-fulfilling prophesy" (p. 535) of math-phobic journalism programs.

Nguyen and Lugo-Ocando (2015) however, encourage the attitude that statistical analyses are about the application of valid reasoning, and not merely calculation, and as such these skills are a perfect fit for a journalism education:

> If one can add, subtract, divide and multiply, he/she can learn to handle statistics for the news, as long as he or she is willing to apply to data the same probing and enquiring mind that is essential for any other newswork (p. 5).

There is also a call for critical thinking that will allow for journalism students to "learn to treat data in an ethical way, so that rather than bending the data to represent a particular view, the goal is toward truth and accuracy" (Berret & Phillips, 2016, p.41).

Data visualization and presentation

Much has been written about data visualization and so-called infographics (Cairo, 2014; Dick, 2013; Stray, Ma, & Chun, 2016; Tufte, 2006; Vallance-Jones et al., 2009). As well, a set of best-practice principles is beginning to emerge in journalism, consistent with other disciplines with more established histories of visualizing information (Bradshaw, 2014). Adopting these rules from fields like computer science and psychology is seen to be important because "the purpose of journalism is to increase knowledge among the public while minimizing the side effects that making that knowledge available might have" (Cairo, 2014, p.25). Cairo (2014) states not adopting these practices, or being ignorant of them, can itself be unethical as it will lead to confusion instead of clarity.

Computer science knowledge

It is widely agreed knowing how to use a spreadsheet is an essential skill (Graham, 2015; Herzog, 2015; McGregor, 2014; Plaue and Cook, 2015; Stray, et al., 2016.; Vallance-Jones, et

al., 2009, Vallence-Jones & McKie, 2017), however the literature varies in approaches when it comes to the demand for computer coding literacy, and even what "coding" (Broussard, 2015a) means in a journalism school setting. Gillmor (2016) does not advocate that journalists become programmers, but does recommend journalists acquire an understanding of how code works to a level that they be able to communicate with programmers. Berret and Phillips (2016) call for, at a minimum, understanding how coding works, whereas Broussard (2015a) supports the adoption of "CS0"-level course content modeled after "precursor courses often focusing on computer fluency or computational thinking" (ACM, 2013). There are some specific coding languages deemed as valuable in the literature including Java Script (Gillmor, 2016; Stray et al., 2016) Python (Cook, 2015), HTML (Plaue & Cook, 2015) and Cascading Style Sheets (Plaue & Cook, 2015; Stray, et al., 2016), however there is little consensus as to the level of fluency in the languages needed.

The literature also acknowledges (Young & Hermida, 2015) and promotes (Diakopoulos, 2015; Flew et al., 2012) the application of computational thinking in journalism. This is perhaps best described by Gynnild (2013) as thinking which "integrates logical, algorithmic, scientific, and innovative dimensions of human cognition. It includes abstracting and decomposing data when approaching complex tasks, in addition to building algorithms for pattern recognition" (p. 723). However, the literature also establishes the resistance to adopting this kind of learning because journalism students may be afraid of computer science courses (Cook, 2015) or unaware of its value to journalistic pursuits (Hewett, 2015), and educators may be unwilling to make room in a curriculum by pushing out more traditional topics (Yarnall, et al., 2008).

Data Journalism Education Strategies

Having established some of the gaps in both the literature surrounding data journalism education and its practice, we turn now to answering **LRQ3**: the literature focused on creating learning opportunities while overcoming resistance to a new journalism pedagogy.

Hands-on learning

Post-secondary journalism programs almost always adopt a hands-on learning approach to teaching news gathering skills (Gillmor, 2016; Hewett, 2015; Yarnall et al., 2008). Therefore, it is perhaps not surprising that the literature dedicated to revealing the design of courses meant to support data journalism reveals a hands-on approach to teaching the skills while integrating them into more traditional news projects (Graham, 2015; Plaue & Cook, 2015; Treadwell et al., 2016). Exploring political reporting (Graham, 2015) and medical reporting (Plaue & Cook, 2015) courses, the literature reveals challenges to this kind of course delivery in a traditional post-secondary setting, including faculty being not up to date with what is happening in industry, long lead times for curriculum changes, a lack of reference material, and the difficulty in acquiring data sets in the public interest. To overcome these obstacles, the literature suggests creating a curriculum collaboratively and iteratively (Treadwell et al., 2016); customizing textbooks and using online resources (Plaue & Cook, 2015); using existing open-source data for course projects (as opposed to integrating new data acquisition through Freedom of Information) (Graham, 2015); and setting students up as "networked journalists" by pairing them with journalists working in industry (Graham, 2015, p. 247).

The literature advocates for all journalists, whether preparing directly for work in the data journalism genre or not, to endeavor for a kind of mastery in order to understand and report on how data affects the world in which we live. Diakopoulos (2015) calls for a better understanding of how algorithms work: "algorithmic accountability reporting thus seeks to articulate the power

structures, biases, and influences that computational artifacts play in society" (p. 399). Bradshaw (2015) stresses the need for an understanding of how to apply traditional journalism measures of "the public interest" (p. 210) and "minimizing harm" (p. 206) to a data environment.

Collective, collaborative learning

It is well established in the literature that many practitioners of data journalism are self-taught (Lewis & Usher, 2014; Royal, 2010; Young & Hermida, 2015). The literature explores several ways to self-teach computer science skills (Broussard, 2015a) like coding and data visualization through sites like Code Academy (Plaue & Cook, 2015). Considering the collaborative nature of this work, it makes sense that there are many online courses and resources available to journalists at any point in their careers. A complete online course designed by Stray, et al. (2016) including a syllabus, has been posted for use by all educators, students and professionals. Groups like Hacks/Hackers have sprung up around the world, and as the literature points out members are both computer hackers and mainstream journalists and are frequent collaborators (Borges-Rey, 2016). Their meetings sometimes serve as places where journalists and students (Hewett, 2015) can collaborate to learn computer science skills and attitudes (Lewis & Usher, 2014) and pursue the ethical values espoused by hackers: "sharing, openness, decentralization, and improving the world" (Bradshaw, 2015, p. 207).

Finally, there are many organizations that teach data journalism skills outside of an educational institution, including the Canadian Association of Journalists (McGregor, 2014) and NICAR, which has been active for decades (Lee & Fleming, 1995). Both organizations continue to be important players in educating journalists in data-driven storytelling skills (Anderson, 2015; Berret & Phillips, 2016). Through conferences, online newsgroups, and publications, these groups are responsible for educating new journalists and experts wanting a new skill set. There are also numerous massive online open courses (MOOCs) delivered through sites such as

Coursera and schools like the Journalism and Media Studies Centre at the University of Hong Kong, which were not examined in my survey of the academic literature surrounding data journalism education.

Technical obstacles

Broussard (2015a) found that not all online tools could work in a computer lab because of security protocols, and that completing assignments on a cloud server could violate student privacy regulations. Her solution was to work collaboratively with the school's IT department to develop a secure server, and to use tools like WordPress for student projects.

A Canadian perspective

There are many journalism programs at college, university and graduate levels in Canada, practitioners of data journalism working in newsrooms around the country, and Canadian researchers writing on the topic, yet the literature is sparse in terms of Canadian examples of research into data journalism work or education. A study of newsrooms in Quebec (Tabary et al., 2015), and an exploration as to how highly valued data journalism skills are by news leaders (Hermida & Young, 2016) were the only works found which specifically explore the Canadian landscape.

Summary

Through this literature review, I have examined the epistemological shifts taking place in journalism necessitated by the proliferation of data and the increasing availability and application of new computational journalism tools. The literature discusses both the steps taken in an ideal data project and the skills needed to work as a journalist in the era of Big Data. I have also identified several areas where research is lacking, including an analysis of the objectivity of data, transparency, and data acquisition techniques in the context of journalism training programs.

This literature review informs the approach to my capstone project which seeks to fill some of these gaps by examining how the instructors of data journalism learned their skills, how they are teaching these skills in a traditional journalism school setting, and how post-secondary journalism programs are responding to these trends by designing program and course content. Combined with data gathering from key informants, and with a Canadian focus, my capstone will reveal how journalists are leveraging computational tools to increase accountability, inform storytelling, and engage audiences.

Methodology

Introduction

As was explored in the literature review, the introduction and development of data journalism has an effect on journalism practice, yet little has been researched on this effect on Canadian journalism education. It was my intention to explore the phenomenon of data journalism education in order to fill the gaps in the literature and to give some insight to practitioners and instructors in the field. Using an exploratory research design, I interviewed key informants to determine how data journalism is impacting both the perception of journalism, the skills which are currently taught, and their usefulness to journalists pursuing data-oriented stories. This chapter explores both the strengths of exploratory research design, and will reveal how the weaknesses were addressed. Next, the process by which the participants were identified as key informants and recruited will be explained, along with an overview of how I handled data collection. Finally, my approach to data analysis using iterative analysis (Tracy, 2013) will be described in depth.

Design

My research project has been developed to answer the following RQs:

RQ1: How is the development of data journalism changing journalism practice?

RQ2: How is the development of quantitatively oriented journalism impacting journalism education?

RQ3: What are the core skills needed to be a successful practitioner of quantitatively oriented journalism and where are they learned?

Knowing there is much revealed in the literature about the changing roles and responsibilities of journalists, particular attention in developing my methodology was given to **RQ2** and **RQ3** which focuses more specifically on the development and delivery of data journalism education. In order to answer these RQs, an exploratory design was chosen as the most suitable: there are few studies exploring data journalism education in Canada and, as such, academic investigation is in its early stages. An exploratory design aims to understand context and identify key issues (Bamberger, Rugh & Mabry, 2006), is flexible, allows for the development of theories and hypotheses, and informs future research (University of Southern California, n.d.). Exploratory design was also chosen as the most likely to provide "rich description" and because it is "primarily inductive" (Cuthill, 2002, p. 80).

The intention of this study is to provide insight into what leading practitioners feel are essential data journalism skills, and to highlight best practices to deliver that knowledge to journalism students. These findings will help to establish more specific lines of inquiry for future studies and inform discussions about journalism curriculum development.

While it is acknowledged there are limitations to this type of design, primarily that using a small sample of key informants is not reflective of a general population (University of Southern California, n.d.), the field of data journalism in Canada is small, so the commonality in the data gathered will reflect generally and with a fairly high degree of reliability which and how

data journalism skills are being taught. Further, it is not the intention of exploratory research, when the phenomenon is less well known as in the case of data journalism education in Canada, to make absolute conclusions, but rather to explore and describe (Stebbins, 2001). Any "early weaknesses in sampling, validity, and generalizability tend to get corrected over the course of several exploratory studies" (Stebbins, 2001, p. 5).

Participants

In order to best explore the state of data journalism education, a purposeful sample of key informants was identified which includes data journalism instructors who also have experience working as journalists. Purposive sampling aims to group the participants according to criteria which is most relevant to the research question (Mack, et al., 2005), and helps to strengthen the validity of a qualitative design (Bamberger, et al., 2006). While I considered snowball sampling (Edwards & Holland, 2013), purposive sampling was deemed the best approach as it was straightforward to find key informants and is more transparent, which addresses issues of validity in an exploratory study.

I identified my selection criteria as those who are currently teaching a stand-alone data journalism course at a post-secondary institution in Canada, including two-year college diploma, bachelors, and master's level programs, and who work or have worked in the field as journalism practitioners. As my theoretical influences include Rogers' *Diffusion of Innovation*, the criteria were selected partly to include those who fit into the "early adopters" or "early majority" groups (Rogers, 2003, p. 283). This bias in this sample towards current educators who are also practitioners of data journalism should be seen as a strength (Mayan, 2009): as stated above, academic research on how data journalism is taught in Canada is sparse, and as such, choosing this sample offers "the most and best information" (Mayan, 2009, p. 62) about the topic.

The identification of participants began as I looked at my personal professional network, and I was in touch with these key informants via a personalized email (see Appendix B). In an effort to overcome the limitations of the convenience sample, (namely that I have more knowledge of people in my professional network through both the geography of where I live and the media organizations where I worked), a rigorous process was undertaken to ensure key informants were not missed. This process was developed partly in an effort to broaden the demographics of the key informants from predominantly men, which also persists in the general population of data journalists (Berret & Phillips, 2016).

By using the educational institutions list on the Canadian Journalism Project website (J-Source.ca) a total of 20 colleges and 18 universities were identified as offering journalism programs in Canada. Each program's website was checked for a stand-alone course title which included words like data journalism, data visualization, or computer assisted reporting. While it is acknowledged that some data journalism content may be offered in other journalism classes, such as multi-media or investigative journalism, these were excluded; the analysis of courses was not intended to yield an overview of content, but rather to identify key informants. Through this process, five universities and six colleges offering data journalism courses were identified, (see Appendix C) and one additional key informant was found for a total of 12. Next, the institutions that did not explicitly offer a data journalism course were sent a form email (see Appendix D) to confirm that data journalism was not being taught as a specific, stand-alone course with a different name for example one listed as "special topics." Despite a poor response rate, through this effort, one more key informant, who teaches both at a college and a university, was identified for a total of six key informants. It should be noted several programs acknowledged they had courses in development for the future or were interested in the topic.

A Google Fusion Table map showing the distribution of these schools can be seen in Figure 3 below.

Figure 3. A Google Fusion Table illustrating the distribution of journalism schools in Canada; those with stand-alone data journalism courses are shown in yellow. An interactive version is at: https://www.google.com/fusiontables/DataSource?docid=14lqGlBvS2FAwOIjVzVR5V05jO6ERFZQ_CQutitFz

Some exclusions of note: first, one instructor who may have been selected through this process is also a member of my MACT cohort and therefore was not interviewed as a key informant; second, there was a potential for a conflict of interest, whereby I was employed on a freelance basis by a potential key informant, which excluded another post-secondary program fitting the criteria. In one case an institution had more than one instructor teaching data journalism, and the instructor selected from the two was chosen on the basis of teaching the more advanced class (based on Coddington's 2015 typology) as well as being mentioned as a subject matter expert by several other key informants in my professional network.

Following the guidelines of the University of Alberta Research and Ethics Office, each of the identified key informants was sent an information letter and consent form (see Appendix E),

which outlines my research and gives enough information so each participant "can decide in a conscious, deliberate way whether they want to participate" (Mack, et al., 2011).

A note on anonymity: the data journalism field in Canada, as found in Hermida and Young's (2016) study of the professional development of data journalism in Canada, has a small number of practitioners and most of them are known to each other, and to other journalists through reputation. As outlined above, there are only a dozen institutions offering stand-alone classes in data journalism at this time, and some of the instructors teach at more than one of them. Despite the small number of people in this field, I offered anonymity in order to encourage participants to be open during their interviews (Hermida & Young, 2016), when discussing their workplaces and institutions. However, I have made clear to them that I could not guarantee that no one would be able to figure out who they are based on some of their responses.

Setting

In an effort to do research which gives a pan-Canadian picture of data journalism education, key informants were selected from programs in Ontario, Nova Scotia, and British Columbia. It would not be financially or practically feasible to conduct each interview in person, therefore, all of the interviews took place as an audio only interview over Skype or by phone, depending on the preference of the participant. All interviews were recorded on a digital audio recorder for transcription and analysis, and observations from the interview were noted in a notebook during and immediately after the interview.

While ideally a face-to-face interview would be conducted with each key informant, as the production of knowledge in this context is through conversation (Brinkman& Kvale, 2015), and conversations held face-to-face may have a better flow, it is almost certain all of the participants have conducted or been the subject of interviews or conversations through this

technology by virtue of their work. For these reasons, and because of the research of Sturges and Hanrahan (2004 in Edwards & Holland, 2013) which found the depth, quality, and amount of data between telephone and face-to-face interviews did not differ substantially, I do not expect there to be a difference in the quality of the study data generated as a result of this choice.

Each participant was interviewed in a similar setting: namely a relatively quiet and private space of their choice, be it their office, home, or another such place. My intention in offering them the choice was for them to select where they are most comfortable and most likely be able to speak freely, and to have quiet in order to preserve the quality of the recording. Though it can sometimes be desirable to set the interview in a more neutral setting to encourage a participant to speak in more general terms rather than in an organizational context (Edwards & Holland, 2013) the questions in the interview script took this into account and touched on both questions of what the institution is doing generally, as well as what data journalism education should be aspiring to do.

Instrument

It is my intention to gather data through conversation, rather than the "modernist approach to interviewing" (Mayan, 2009, p.70), whereby the interviewer asks questions just to retrieve information from the participant. A semi-structured interview was selected because it leaves more room for exploration as the questions can be varied in order to "obtain the maximum information from each interviewee" (Bamberger et al., 2006, p. 307). This offers more precision than an unstructured interview, and more flexibility than a structured interview. It is also the most suitable when a researcher has "a fair enough idea of what is going on in or with the phenomenon to develop questions about the topic but not enough to predict the answers," (Mayan, 2009, p. 71). I see myself not as an objective observer in this interview, but rather as

"implicated in the process at play" (Edwards & Holland, 2013, p. 2) in a variety of ways: first, I have a thorough understanding of the literature on the topic; second, I share some professional knowledge with the participants as we have all worked as journalists; and third, we all have some experience teaching in the classroom, though admittedly mine is limited to one class on news reporting fundamentals at the college level. In addition to the interviews, each key informant was asked to provide a course outline for the data journalism class(es) they teach before the interview in order to inform my approach to questions.

My semi-structured interview guide (see Appendix F) is more precise than one based only around themes (Edwards & Holland, 2013), but each question has been organized around thematic ideas. It was based on a structure offered by Mayan (2009) which includes four phases: introducing the topic, easy questions which show empathy, tough questions and toning down the interview on a positive note. This is a similar structure that is used in feature length journalism interviews, as is my professional experience, where there is an ice-breaker, context, tough questions, and forward looking/closing questions. The participants responded well to the structure, and the conversations built rapport and generated rich data. In all of the interviews but one, the flow was natural and questions followed the order of the guide. In the remaining case, the questions were in a slightly different order so as to follow the flow of conversation.

Questions informed by literature

The design of the interview guide was informed by the literature review and two theoretical perspectives: Rogers' *Diffusion of Innovations* (2003) and *Knowledge Gap Theory* (Tichenor, et al. 1970). Coddington's typology (2015) was used to place both the participants' programs and their experience on a spectrum in order to analyze the correlation between professional skills and to give an indication of the baseline of programs in Canada. Broussard

(2015a), was used to inform questions about technological barriers in educational settings. The work of Treadwell, et al. (2016), Graham, (2015), Hewett (2015), and Lee and Fleming (1995) was used to explore the practicalities of teaching data journalism in a journalism classroom. Literature by Nguyen and Lugo-Ocando (2015), Dunwoody and Griffin, (2013), and Yarnall, et al. (2008) was used to inform the structure of sections on statistics and numeracy. The research of Plaue and Cook, 2015 was used to inform the section on student demographics and gender bias. Literature exploring the role of the journalist in interpreting data (Lesage & Hackett, 2013), and the objectivity of data (Lesage & Hackett, 2013; Splendore, 2016), were also used to inform questions which attempt to gain insight into the motivation of the instructors and their perceptions of how this kind of news work impacts audiences.

Research influences

During the literature review, two studies stood out as having particular relevance to my work: Dunwoody and Griffin's work *Statistical reasoning in journalism education* (2013) and Yarnall et al.'s *How post-secondary journalism educators teach advanced CAR data analysis skills in the digital age* (2008). Both offered frameworks on how journalism departments perceive statistics and numeracy and how new storytelling technologies and techniques are introduced to the journalism classroom. I was in touch with both Dr. Sharon Dunwoody and Dr. Louise Yarnall via email, and they were both kind enough to provide me with the information on the questionnaires they used to conduct their respective surveys. I have used these instruments to inform some of the questions in my semi-structured interview guide. Their instruments were valuable to me in that they asked questions exploring the challenges, either technological, student-based, or from within a faculty, of delivering data journalism classes. It was not my intention to replicate their studies, but rather to incorporate some of their queries into my research.

Theoretical influences

Rogers' *Diffusion of Innovations* (2003) theory was applied in crafting the interview guide in order to get a sense of how data journalism is diffusing into Canadian journalism programs. Rogers' theory outlines four main elements of diffusion as "(1) an *innovation*, (2) is *communicated* through certain *channels* (3) over *time* (4) among members of a *social system*" (p. 11). I have selected key informants who are considered, for the most part, to be either in the "early adopters" or "early majority" group of *Diffusion of Innovation* (Rogers, 2003, p. 283). Early adopters are defined by Rogers as those who are "not too far ahead of the average individual in innovativeness," yet "help trigger the critical mass when they adopt an innovation" (Rogers, 2003, p. 283).

Early majority members are defined as those who "adopt new ideas just before the average member of a system" and who rarely hold opinion leadership positions in that system (Rogers, 2003, p. 283), and are considered an important link in the diffusion process. Questions in the interview guide probe where the participants will fall on this spectrum, including questions on their histories with data journalism and if they influence others in adopting it.

Knowledge Gap theory (Tichenor et al., 1970) was also a theoretical influence on my research design and can be summarized as:

> The infusion of mass media information into a social system increases, segments of the population with higher socioeconomic status tend to acquire this information at a faster rate than lower status segments, so that the gap in knowledge between these segments tends to increase rather than decrease (p. 159).

While this theory is predominantly concerned with how science and public affairs information is diffused through mass media to different socioeconomic groups, the ideas can be

applied to the diffusion of data journalism knowledge and skills in journalism schools. For example, one tenet is that as people learn more, they are more compelled to learn. One section of the interview guide explores whether some exposure to data journalism techniques may set journalism students up on a course to be not just journalists but also innovators in data journalism. It has been established in the literature that practitioners of data journalism are usually also members of groups of like-minded journalists like NICAR and the Canadian Association of Journalists (CAJ), and the "relevant social contact" (Tichenor, et al., 1970, p. 162) provided by these groups may serve to augment journalism students' knowledge, thus impacting their career opportunities. Last, as the theory states education is "assumed to be a valid indicator of socioeconomic status" (Tichenor et al., 1970, p. 160), and that a growth in knowledge is greater among higher status segments, an exploration of the perceived value of data journalism at the college, bachelor, and master's level will be undertaken.

Procedures

As described in the instrument section above, a semi-structured interview script was designed to gather the data most likely to answer my research questions. In the data gathering phase, I arranged to conduct the interview in advance and recorded each one on an MP3 recorder. Because qualitative data analysis "tends to be inductive" (Schutt, 2012, p. 322), I approached the interviews with no pre-defined hypotheses. My interview script was informed by knowledge I gleaned from theory and the literature review, through casual conversations with practitioners, and through the knowledge I have acquired through my study of data journalism techniques. This approach is more flexible, enabling me to sharpen my design as data is collected and interpreted (Bamberger et al., 2006). The interview guide was tested on the aforementioned

MACT colleague and several adjustments in the number, order and wording of questions were made through several versions.

Notes

Both during the interview and immediately afterwards, I took notes in a field journal. During the interview, I noted phrases and key words which would "trigger" (Mack, et al., 2011, p. 44) my memory when expanding my notes later as I felt taking detailed notes during the interview would be distracting (Edwards & Holland, 2013). Immediately after the interview, I wrote down my thoughts and observations which supported my data analysis process, writing this report, and improved the quality of the interviews which followed. I also jotted down thoughts and comments during both the transcription and analysis phases in order to keep track of ideas and questions for follow up, and to develop ideas around codes and themes.

Transcription

I decided to transcribe the interviews myself because as a television news producer, I have spent countless hours transcribing video and audio and so the thought of accurate transcripts did not daunt me. I transcribed the interviews as soon as practical in NVivo. I used a transcription key based on the key in Mayan (2009) (see Appendix G). This key offers a consistency for each interview which will be important in analysis and provides additional information about the audio that may not be apparent from the transcript. I have also edited the transcripts for clarity (Roulston, 2014) removing words like "um", "uh" and "you know", "kind of" and "like" when they serve only as filler words and do not clarify meaning.

Anticipating roadblocks

There are several issues I was concerned about affecting my data gathering procedures: the first was that although I had spent quite a bit of time reading data journalism literature, had taught one journalism class at the college level, and have worked in investigative journalism, I

did not have direct experience or expertise in teaching data journalism and had limited experience and expertise in publishing data journalism projects. In an attempt to be certain I would have an understanding of many of the key terms which could come up in the interviews, in the lead up to this study I took some introductory classes in the areas of data journalism I did not already have a familiarity with, or spent time refreshing skills I already had. This included an online data journalism course offered by the University of Hong Kong's Journalism and Media Studies Centre, introductory classes in coding languages like Ruby, HTML, and CSS through Ladies Learning Code, and a course in data visualization techniques and data management skills.

Another area of apprehension was that there are few data journalism instructors in Canada, and if I could not get enough people from my key informant sample to agree to participate, I would not have enough participants to ensure a rigorous study. However, the enthusiasm shown by the participants I approached eliminated this problem.

Data Analysis

My research used an interview guide modeled to answer the three broad **RQs**, and through an iterative analysis approach (Tracy, 2013) alternated "between emic, or emergent, readings of the data and an etic use of existing models, explanation and theories" (Tracy, 2013, p. 184). In early planning of this research project, I had intended to use a grounded theory analytic process to "turn raw data into something that promotes understanding and increases professional knowledge," (Corbin & Strauss, 2008, p. 3). However, grounded theory approaches have many variations, and seemed more rigid and in need of longer timelines than this project would allow. I felt the iterative approach as outlined by Tracy, which aims for a "problem-based approach of qualitative data analysis" (Tracy, 2013, p. 184) allowed for more flexibility, was

more reflexive in nature, and allowed for the progressive refinement of emerging insights (Srivastava & Hopwood, 2009, in Tracy, 2013).

Coding process

My intention in my data analysis was to move from speculations to possible explanations (Mayan, 2009) and I accordingly developed a hybrid process of data analysis using paper transcripts, NVivo, and an Excel spreadsheet.

I decided to use NVivo software as part of my data analysis for several reasons: first, the exploratory nature and iterative analysis approach had enough fluidity that using coding software would not be too restrictive. Second, this approach allowed for coding with multiple codes per section; this was a benefit as I have three different, but overlapping, research questions. As the topic of my study is data journalism, I felt NVivo would enable me to visualize qualitative data in a way that would help me stay in the quantitative mindset of a data journalist

The first round of open coding was done by hand on paper copies of the transcripts, which allowed me to slow my reading down for better comprehension (Wolf & Barzillai, 2009). Those codes were put into an Excel spreadsheet where each was defined and put into alphabetical order for easier reference. The next round of open coding was done in NVivo, a recursive process (Bazeley, 2013) where I referenced codes from the first round, and worked through my sources to review the earlier codes (Bazeley, 2013) into a more precise list. As the codes were further refined to reflect the data, they were entered into a spreadsheet which recorded their meaning and provided an audit trail on their use. I also kept memos in NVivo outlining questions and thoughts about the data. The process of analyzing the data helped me to focus on what the participants meant by their words and by doing several rounds, I was able to see patterns.

After two rounds of open coding, my second stage coding began by re-reading the transcript passages for each code to find themes and concepts (Bazeley, 2013) and selecting the themes which best answered each RQ. This codebook can be found in Appendix H.

Validity and reliability

There has been some debate, explored in depth by Mayan, (2009) as to how to define concepts like validity and reliability in qualitative research, and how to ensure these studies are as rigorous as those with quantitative designs. Validity, or trustworthiness, can be described as "getting to the truth of the matter" (Grbich, 2013, p. 5) and reliability, or dependability, can be "viewed as sound research design" (Grbich, 2013, p. 5). Throughout the design, data collection, and analysis of this study, checks were put in place to ensure validity and reliability. Unlike quantitative studies, research efforts using qualitative inquiry do not translate data into numbers nor do they separate themselves from the instrument (Tracy, 2013). Qualitative inquiry recognizes the researcher's mind as a fundamental part of the research (Tracy, 2013) and works "inductively from individual cases" (Mayan, 2009, p.11). However, both qualitative and quantitative research require "careful planning and precise means for collecting evidence" (Merrigan, Huston & Johnston, 2012).

As noted by Arksey and Knight (1999), validity raises "the question of whether you are actually investigating what you claim to be investigating" (p. 51). To that point, each of the questions in the interview guide was created to answer one or more of the RQs, and have been noted as such. Applying knowledge from the literature review and theory while writing the interview guide, and using this information in the data analysis process, serves to triangulate the findings. The evolution of the RQs has also been made transparent and can be seen in Appendix I.

It is acknowledged there are threats to validity with my research design. For example, the key informant exclusions noted above, and the possibility a leader in data journalism education was missed, perhaps because they were on sabbatical or the school website was not up to date and did not show data journalism courses on offer. However, several steps were taken throughout the research design process to enhance validity. These steps include an interview guide structured to build rapport with participants (Arksey & Knight, 1999), consistency in data collection such as asking questions in the same sequence each time (Quinn Patton, 2008), and giving participants a chance to review the transcripts of our interview as a member check.

There exists discussion around what constitutes reliability in qualitative studies. For example, Quinn Patton (2008) writes "A measure is reliable to the extent that essentially the same results can be produced repeatedly, as long as the situation does not change" (p. 402), while Miles, Huberman and Saldana (2014) point to issues of quality and integrity. These questions on reliability have been addressed in this capstone, however, by the study design: the research questions are clear; data collection is appropriate for the scope of the study; my role in the world of data journalism and education has been explained to the reader; and the findings in the data have meaningful commonalities across participants.

As data journalism, and in particular data journalism education in Canada, is a constantly evolving field, it may not be possible to replicate these results exactly five or ten years from now. However, I believe that my study is sound and a similar study taken in five or ten years using this design would be replicable in that it would give a similar type of moment-in-time insight into the state of data journalism education in Canada according to the key informants.

Summary

Through this methodology chapter, I have demonstrated the validity of choosing an exploratory research design for the exploration of data journalism education in Canadian journalism schools, as well as presented a reliable design to gather and analyze data. Using interviews from key informants, I sought answers to the RQs examining how data journalism is changing journalism work and education, and what the most important data skills are for journalists today. Relying on the literature review and previous research, while acknowledging the applicable theoretical frameworks, the information I have collected will provide important insights in the discussions and analysis section.

Analysis and Discussion

This chapter will reveal what was discovered during the data gathering process and how the interviews with key informants and data analysis will answer the three RQs which explore data journalism in journalism practice, its influence in the classroom, and the essential data skill set and where it can be learned. First, I will present the findings from my data gathering, addressing the research questions and highlighting the commonalities and differences among the views of participants. Following that, I will outline the data analysis process, including illustrations of how codes and themes were developed and organized. Next, the discussion section of this chapter lays out how my findings contribute to the field generally, and more specifically offers journalism educators a road map to designing courses that will serve their students. Finally, the limitations of this study and how they were addressed will be presented.

Findings

Throughout the interviewing process, one thing stood out: the participants are all storytellers. When asked a question about data journalism work, teaching, or their personal

motivations to learn data skills, they almost always used a journalistic story they had worked on, or a data story they thought was notable, to explain their answers. Six key informants, five male and one female, were interviewed over a three-week period in April and May, 2017. All of them work as instructors at either the college, bachelor, or master's level at post-secondary institutions in British Columbia, Ontario, or Nova Scotia. Participants B and D teach specific skills in a workshop setting, while Participants A and C both lead intensive bootcamps and workshops, in addition to teaching in a post-secondary setting. All but one (Participant F) has worked as a data journalist in the field. Participant F was included as a key informant because they have experience as an online journalist, as well as significant interest and experience in teaching data journalism skills. Two of them (Participants A and C) have more than ten years' experience teaching, three (Participants B, E and F) have between five and ten, and one (Participant D) had less than five years. The table below illustrates this range of experience:

Participant alias	Data journalism work experience	Program level teaching experience	Years of experience teaching data journalism skills
Participant A	Yes	Bachelor, Masters, Bootcamp	More than 10
Participant B	Yes	Bachelor, Workshops	5-10
Participant C	Yes	Bachelor, Masters, Bootcamp	More than 10
Participant D	Yes	Bachelor, Workshops, Fellowship	Less than 5
Participant E	Yes	College/Bachelor	5-10
Participant F	No	College/Bachelor, Graduate Certificate	5-10

Table 1: Illustrates the range of experience of participants.

Each interview took roughly an hour, and in that time rich and informative discussions were had on who controls the narrative of news, and what kind of skills are needed to tell stories

effectively with data. In all of the interviews but one, the flow was natural and questions followed the order of the guide. In the remaining case, I followed the participant on a tangent where we discussed the challenges of getting Access to Information[2] data from government sources, but I do not feel this had an effect on the completion of all the questions or the resulting data.

In order to understand how the views of participants overlapped and where they differed, as well as to get a sense of what drives their approach to teaching data journalism skills, it was important to ask first how they felt data journalism was changing the way reporters told stories. This is where the interview conversations began, and it offered a great deal of insight.

RQ1: How is the development of data journalism changing journalism practice?

A shift from attribution to evidence

One of the points most emphasized by participants was their feeling that has been a marked shift from journalism stories being primarily an exercise in collecting quotes from experts, to relaying the evidence through data. This change was described by Participant B as "[data journalism] does get you closer to this ideal of being able to actually find out what the truth is." This change was described by Participant B as moving journalism from a model of attribution to a model of evidence, shifting away from relying on the opinions of others towards a more analytical role for the reporter:

> To actually get at data so you can say with some confidence crime is going up, crime is going down, this party has raised the most money from this area, this party has done the

[2] "Access to Information" and "Freedom of Information" are used interchangeably when referring to government information and/or documents acquired under the Canadian Government's Access to Information Act, or similar laws and regulations at the provincial or municipal level.

> most from this. Whereas I feel like a lot of journalism as it is traditionally practiced, I think it is out of necessity it is a he said, she said model.

Participant A cited an example of a data project using inspection reports when explaining that data allowed for a better sense of the details of the story rather than relying on interviewing experts:

> it's a more perfect artifact than someone's memory a lot of the time…I saw the ability of [data] to take the journalist from asking people what's going on to being able to kind of literally sit on the shoulder of every rooming house inspector and see what they'd found.

Participant C agreed this role includes the ability to write in a more declarative way, to discover new insights through data analysis and to have more confidence in the conclusions put forward: "It's not he said, she said, it is not something someone else is saying.

Of particular note in the discussions was the *Globe and Mail's* recent investigative project *Unfounded*, in which reporter Robyn Doolittle gathered police reports on sexual assaults. This was mentioned by Participant A, Participant C, and Participant E as an example of exemplary data journalism work. As described by Participant C, when there is an inkling that something unjust is happening, the proof may be found in the data. This approach was described by Participant C as "the subjectivity of purpose but the objectivity of method":

> Robyn Doolittle knew there was a problem, but her method was very scientific. They filed hundreds of Access to Information, Freedom of Information requests and they crunched numbers. They explained how they did it. Your purpose is very subjective. You feel there is a problem, and there has to be a solution and people need to be accountable. That's a very subjective exercise. How you get to it is very scientific.

All participants stated at some point they did not believe that data could be the only source, but instead saw it as one part of a process which still involves interviews and outside expertise. The participants agreed confirmation of the analysis should be sought, and believed there was still a role for expert opinion. All participants spoke of the need for critical thinking around the origins of the data: "The numbers do not speak for themselves, you have to find out who captured those numbers and why they did that" (Participant F). Participant D noted reporters and students need to "understand that any data release has a political decision behind it, and to ask questions about what that decision is."

Anecdotes and trends

The work of Kevin McConway's *Statistics and the media: a statistician's view (2015)* was used to inform the conversation exploring a shift from using the more noteworthy "extreme values" (p. 50) in a data set to shape a journalistic narrative to one more akin to a statistician's approach which prioritizes averages and distributions. Most participants did not feel that inclusion of statistical analysis was causing a shift away from using anecdotal examples, as is done in traditional journalism stories. Instead, the majority expressed that data stories combine an approach where they present the outliers, while also telling the story of what is happening to the majority. Participant C said "I think that really good data journalism is sort of about the outlier: it puts the outliers in a more understandable, more meaningful context, but it's also about the stuff that's about the other two thirds of the iceberg." An interactive data journalism story that illustrated vaccination rates at schools across a health district was used as an example by Participant B to illustrate this point:

> People were engaged to find out 'what's the vaccination rate at my kid's school?' but then while you are looking at the map for your kid's school, you look at the school beside it

> where vaccination rates are really low. So it makes you aware of the bigger picture at the same time as you are looking at your specific piece of it.

Giving up control to the audience

While the participants discussed how data gave them more control over the narrative, they also talked about how they give up some control to the audience through interactivity in their stories:

> So now you can do a story for Bob, a story for Jane and a story for Carol… And that's huge, it's huge because the previous paradigm, if you can call it that, was that I as a journalist decided what was important (Participant A).

One participant said data journalism allows for an interactivity that other formats of telling stories do not: "Data journalism, to me, is the closest thing that gets to that [point] where you almost are able to create something that answers a very specific question that your audience has, which traditional journalism was never able to do" (Participant B).

Data journalism: set apart from the pack

Throughout the interviews, the participants talked about the ways in which practicing data journalism differentiates them from their peers, either by winning journalism awards or being able to choose their own assignments by generating exclusive story ideas. These original story ideas can help to set participants apart from their co-workers in the newsroom, allowing journalists to work on assignments they felt were more desirable or on stories of their own choosing, rather than being assigned to a routine story by an editor:

> At the end of the day, [data journalists] want to be masters of their own destiny, they don't want to be assigned all the time, they want to be doing their own stories, they want to be doing original stories (Participant C).

Some participants felt there was value in using data journalism skills for daily news as a tool in the same way that photography is considered to be one tool in a journalist's tool box, yet most of the work examples given were of in-depth projects that took more time than a daily news story would. Project examples included interactive maps showing vaccination rates, flood maps using government sensor data, an interactive adverse drug reaction database, information about crime rates, and Taser use by the RCMP. All of these stories highlighted the use of data skills in elevating a story in the news cycle in a way that was exclusive to their media outlet: "A lot in journalism, we are all writing about the same thing, the same event or the same press conference. With data journalism you can do something completely new" (Participant D); "we were able to find and tell stories nobody else had" (Participant A).

Collaboration and competition

Despite the common assertion that practicing data journalism was going to set them apart, there was also much discussion about how data journalism is fostering greater collaboration within the field. Some stated this is because a data journalist is likely to be the only one with that position in the newsroom, so they would look for and find support elsewhere, mostly through online communities like the listserv of the National Institute for Computer Assisted Reporting (NICAR) based in the United States:

> That's the thing that is hard to quantify that data journalism has going for it is everyone kind of at one point was a beginner, so it's a pretty helpful, generous community to newcomers…you almost never get anyone at NICAR saying, 'that's a stupid question.' People are just willing to help out (Participant B).

Participants pointed to a rich field of mentors outside of their organizations, who enable them to push their craft forward through generosity in ways they have not experienced in other media lines. As explained by Participant D: "Nobody holds anything back, there's no feeling of,

it's my trade secret, I'm not sharing it. It's more like yeah, here take it. I busted my ass for months to figure this out, but here you go."

After establishing how participants felt about the changes brought by data journalism practice, the next step was to explore the influence of this in the classroom settings where they teach.

RQ2: How is the development of data journalism impacting journalism education?

The range of teaching experiences of the participants was varied: some have taught for many years at the college and university level, as well as at bootcamps or conferences. Others had more limited teaching experience but are themselves very highly regarded data journalists in the field. Only two of them, Participants A and F, are full time instructors, while the rest primarily work as journalists or freelancers and work part time as instructors.

Can everyone do data journalism?

I began this research project with an inkling that data journalists, and data journalism instructors in particular, would be enthusiastic about incorporating the full range of data skill instruction for all journalism students. However, I did not find this, but instead discovered instructors who are passionate about data journalism, but do not see all of the skills as a good fit for all students, or in all cases. While the participants advocated for basic data journalism skills which would serve all journalists, I was surprised when participants discussed what they felt was a certain aptitude necessary to be a successful data journalist. Participant A said:

> I think that there are some data journalism skills that I think are now skills everyone should have, although there is an argument to be made that if your mind isn't wired that way, maybe you shouldn't be dabbling in things you don't understand.

There was agreement that while some students may have the aptitude, there was limited value in teaching intermediate or advanced skills to others because they are not going to pursue data

journalism work once they are finished school. As Participant E stated "Coming out of the class, we have about 50/50 people who say, you know what? I found it interesting, but this is not for me."

The aptitude question was especially pronounced, according to Participant A, in the difficulty in teaching programming and coding to journalists:

> Some people, no matter how hard you try, they never get those things. I'm not sure it's impossible to teach them, but I think a lot of it has to do with redefining who we are teaching journalism. And we have to move towards bringing more and more people into these programs who have these interests, who might previously have gone into maybe a more technical field, who don't see journalism with a computational element as a valid career choice.

Many of the participants felt that without regular practice of data journalism on the job, the skills would atrophy: "The difficulty comes when once they graduate and they move on, if they don't get a job where they use these skills all the time, then they forget" (Participant C). Participant B suggested:

> It's not the kind of thing where people can do a little of it. In order to be good at it and to avoid major errors, I think you need to be doing it, not necessarily full time but as a significant part of your day to day job.

After reading the research of Plaue and Cook (2015) and Tichenor, et al. (1970), I made an effort in the interview to probe the key informants if they felt there was a demographic with a stronger propensity to show an interest in data journalism or whether they had observed a knowledge gap based on socioeconomic factors. None of the participants expressed they had noticed this in their classes in their teaching experience. All participants were aware of the well-

researched gender bias towards men in computer science fields, however, as was noted by Participant E:

> I think there is also a really strong gendered lens to look through. Usually it's the men in my class who are over confident, who think this is a guy's field. Usually the people who express fear or hesitation or say phrases like 'computers hate me' tend to be the women in the class. One of my jobs is to overcome those preoccupations, and make them confident that they are actually good at this and they can do it.

Participants universally agreed there is a higher number of women than men generally in journalism programs, and that anecdotally they did not feel there was a problem with encouraging women to take up data journalism per se. Participant B observed "our student cohort is more female, but even then I would say the ones who are good at the data visualization class are probably pretty similar to the breakdown of the students as a whole."

Self-guided learning

The instructors I spoke to acknowledged that they try to teach a course which will get students interested in pursuing data journalism skills on their own, as explained by Participant B:

> There are so many resources out there to learn specific things, and I am never going to teach them everything and they can find out some stuff on their own, but it's getting over the hump of not being intimidated by it and this is something I want to continue to learn about.

Perhaps this attitude originated in their experience: all of the participants spoke about how they had learned most of their data journalism skills in a self-directed way: through bootcamps, sessions at conferences, online learning through sites like Coursera and Code Academy, or most commonly, by acquiring the skills they needed to fit a particular story they wanted to tell: "most of it was learning little bits and pieces as I needed them" (Participant B) or "the hard skills you

can learn as you go depending on the types of stories that draw you in" (Participant E). An example of this would be learning how to build and use a database to suit a large data set, or to build a web scraper if the information is not available another way. This path to learning also helps data journalists specialize, and keeps them motivated according to Participant B:

> I think the very best way to learn something is when you have a very specific reason to learn it and you are very motivated to learn something. And anytime I have learned anything, that's when I have learned it, because I need to, and learning that specific thing is going to let me learn it better.

The classroom experience

I spoke to participants about some of the challenges of teaching data journalism in the classroom, and they all expressed similar misgivings about a program's ability to impart to students a wide range of data journalism skills in a single course. Participants said that because of the nature of data journalism instruction, smaller classes are important for success, as was noted by Participant F: "with data journalism there is a lot of hands on labs and exercises, whether it's teaching them the coding or the scraping. Having more than 20 students can be a bit challenging as an instructor." Most suggested a basic course at the beginning of the journalism program with more advanced electives at the end, and including an element of pointing students towards outside resources if they want to go further. As Participant E suggested:

> ideally it would be multi-level, where you could be introduced to the concepts in say first year, and just go over a basic spreadsheet story and then let those concepts settle. Apply them to a fluid environment to a story along the way in your second year, and then come back to advanced data skills in third year. To really let it settle in and realize that you have these skills and they are not scary…. To have it in one semester in the middle of

> your degree makes it really easy to think ‘oh hey that's cool’ and then forget it very easily.

Motivation

Participants felt strongly that the best possibility for employment in the changing journalism landscape comes from acquiring a strong set of data journalism skills and that their applicability on the job should be considered when designing course offerings. As was noted by Participant C “Any journalism program has to have elements of [data journalism]. Absolutely. And for them not to is sheer negligence. That's where the jobs are.”

After having a good sense of the participants’ experience in the classroom and how they approach data journalism skills training, we turned our attention to **RQ3**: distilling what the essential skills are for journalists, and which skills should be taken up in cases where a reporter wants to specialize.

RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned?

There was general agreement among participants on what should be offered to all journalism students, but less so when it came to more specialized data skills. There was agreement that data journalists are unlikely to be masters in all of these areas at once, as observed by Participant A: “Everyone has to have a certain amount of knowledge, but not everyone needs to be a specialist… We can't be everything.” Participant D elaborated on this point:

> I don't think you can be an expert at all of those things, I think you can be pretty good at a few of them, or an absolute master of one or two. But being good at coding, at analysis, at visualization, web presentation, user experience, all of that, that's impossible. That's what you call a unicorn.

Basic data skills for all journalists

Interviewing, writing, researching and critical thinking were cited by all participants as vital skills necessary for all journalists, no matter the media line or genre. Participants viewed leaning how to tell a journalistic story as both the most challenging, and the most important skill for any journalist. Participant C expressed that this was of the utmost importance:

> The thing I really stress, is that it's about storytelling. People don't care how the sausage was made, people don't care how many access to information requests you made. People don't care how long it took you to get the data or how difficult it was to crunch the numbers, what they care about is the story. They care about the accountability…We always have to keep that front and centre because the temptation is that they fall in love with the technology. And fall in love with the process.

All of the participants agreed the stereotype of journalists having a math phobia was well founded, yet many of them felt it should not deter students from exploring data journalism because the pursuit did not require advanced math skills. According to Participant B:

> It's surprising how little math is generally required of data journalism because for the most part data journalism is pretty simple. You are basically adding stuff up, or you are ranking it from biggest to smallest, or you are calculating a percentage. It very rarely gets more advanced than that. I think for good reason: because the more advanced forms of math are harder to explain to readers.

The need for numeracy and data literacy was stated by participants as a high priority, saying the lack of basic numeracy is "problematic because a lot of the institutions that we are reporting on, businesses, government, they use numbers all the time" (Participant B), and that without the ability to understand these numbers, the ability of journalists to hold people to

account would be diminished. Participant C was passionate about this point several times in our interview:

> Rather than be the skeptical tough minded journalist that you think you are because you are asking all these tough questions, you in essence are a stenographer and a patsy and will never hold anyone accountable. So I think we have to become very comfortable with numbers and we should always be asking people how they got their numbers, challenging numbers. That is just basic numeracy.

Without question all participants felt that knowing how to use a spreadsheet program like Excel in a journalistic way is a crucial skill, calling it "the gateway drug of data journalism" (Participant A). However, even if a journalist does not work on data stories, the knowledge of how to manipulate and understand spreadsheets was valued by all participants. Similarly, most participants agreed a certain amount of data literacy is indispensable, pointing to critical thinking about data collection, insisting on raw data rather than relying on aggregates, and knowing how to put data into a spreadsheet and manipulate it properly, as representative of this.

There was less agreement on how much coding or programing should be considered foundational, but participants generally felt students should have some programming awareness. As Participant F put it:

> I tell my students you don't have to be a programmer. Having an understanding of what's possible out there so you can work with a programmer or a graphic designer, is essential. You need to be able to know what's possible so you can create your stories.

In terms of at what time in the program these basic skills should be offered, those expressing an opinion agreed basic skills should be offered "right at the beginning" (Participant C) and that they are "entry level" (Participant A), "basic concepts" (Participant E).

Intermediate skills

Participants talked a lot about encouraging students to uncover data in new places through Freedom of Information requests and open data portals like Statistics Canada, and to have a critical approach to the data they find. Participant D said "That's something really important, to not just accept whatever pre-digested stats are given to you. Ask for the source and do your own calculations." Participant E expanded on this point:

> We talk [in class] about the holes in the data, we talk about method, collection methodology, who collected this data, what was the format it was collected in, was it handwritten and so subject to transcription, is it collected by people who are tapping on iPads in a grocery store, or is it collected by professional statisticians?

Teaching students how to incorporate methodology and demonstrate transparency with their data stories was also important to participants and was seen as a way to get students to show their work and increase the credibility of their stories: "What you do is you put your methodology out there, you explain exactly how you got the story, you allow people to download the data, you explain how you crunch the numbers, you allow them to do it themselves" (Participant C).

While simple-to-use visualization tools like Tableau, Google Fusion Tables, and Data Wrapper were mentioned, most participants agreed that they were not always necessary to tell a data driven story, and those basic skills needed to be well established before they could be translated into effective visualizations. Participant E explained some of the design challenges faced by data journalism students:

> You first have to learn about effective information design, know your bar charts from your line charts, your pie charts from your scatterplots, you have to know which ones tell

different stories. That takes time, you have to know about colour theory, perception theory, you have a long way to go.

In terms of timing this skill set, participants did not express a strong preference as to whether it should be a core or elective course, or when it should be introduced except to say that it should be offered after basic skills. As Participant F put it, "we should have an introduction to data journalism and then something a little bit more [sophisticated]."

Advanced skills

Not all participants felt that an ability to build or use a database was an essential skill; some saw it as quite an advanced pursuit: "I think every journalist can benefit from some data literacy, but not all journalists need to code or be a hardcore data person, like [using] Access databases, or know how to do stats and calculations" (Participant D). Others said it was a necessary tool for those wanting to take on important data projects: "People who want to do serious work should learn how to use a database. My SQL, SQL Lite, Access, should be able to do more sophisticated analysis with large data sets" (Participant A).

There was also little agreement on how much programming should be taught as part of a journalism program, with some considering programming literacy as a basic skill, and others who felt that coding knowledge was important to offer. Participant D advocated for a limited amount of programming knowledge across the board: "I think a bit of data literacy is good for every journalist. I don't think every journalist should know R or Python, because I don't think every journalist needs to," whereas Participant F wanted options offered to journalism students: "One of the things I would like to see is an introduction to programming for the journalism students as a general elective, or statistics programming."

Data Analysis

Procedures

As outlined in the methodology chapter, the transcripts from my semi-structured interviews were analyzed using an iterative analysis approach (Tracy, 2013). The codes were developed from notes taken after conducting the interviews, while transcribing the interviews, and through several rounds of coding the transcripts. Development of the codes was tracked in an Excel spreadsheet, and the final version contains the 46 codes which emerged during the process, as well as 12 themes (see Appendix B), as shown in Figures 4 and 5 below.

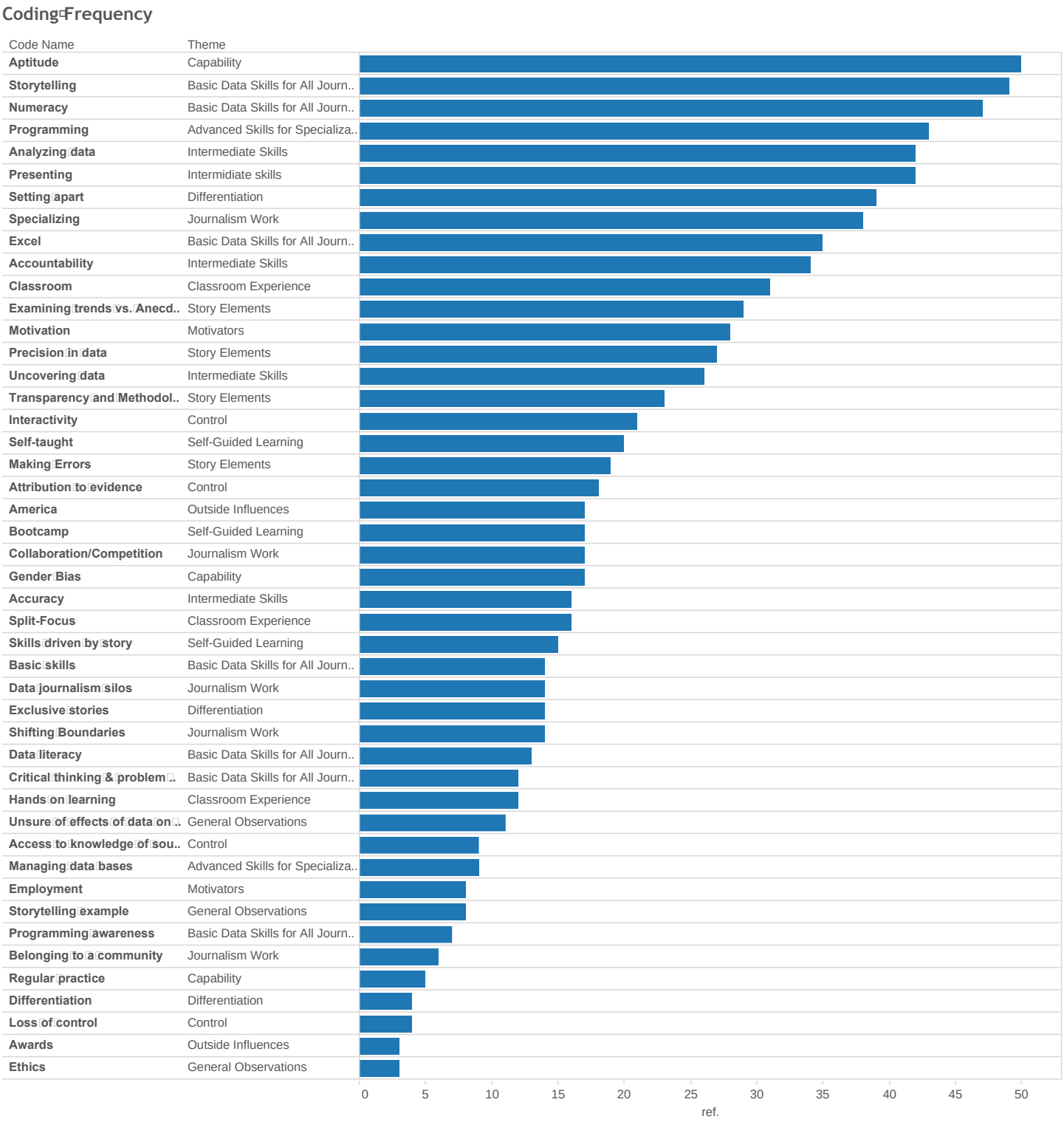

Figure 4. Coding frequency by code and number of instances found in transcripts.

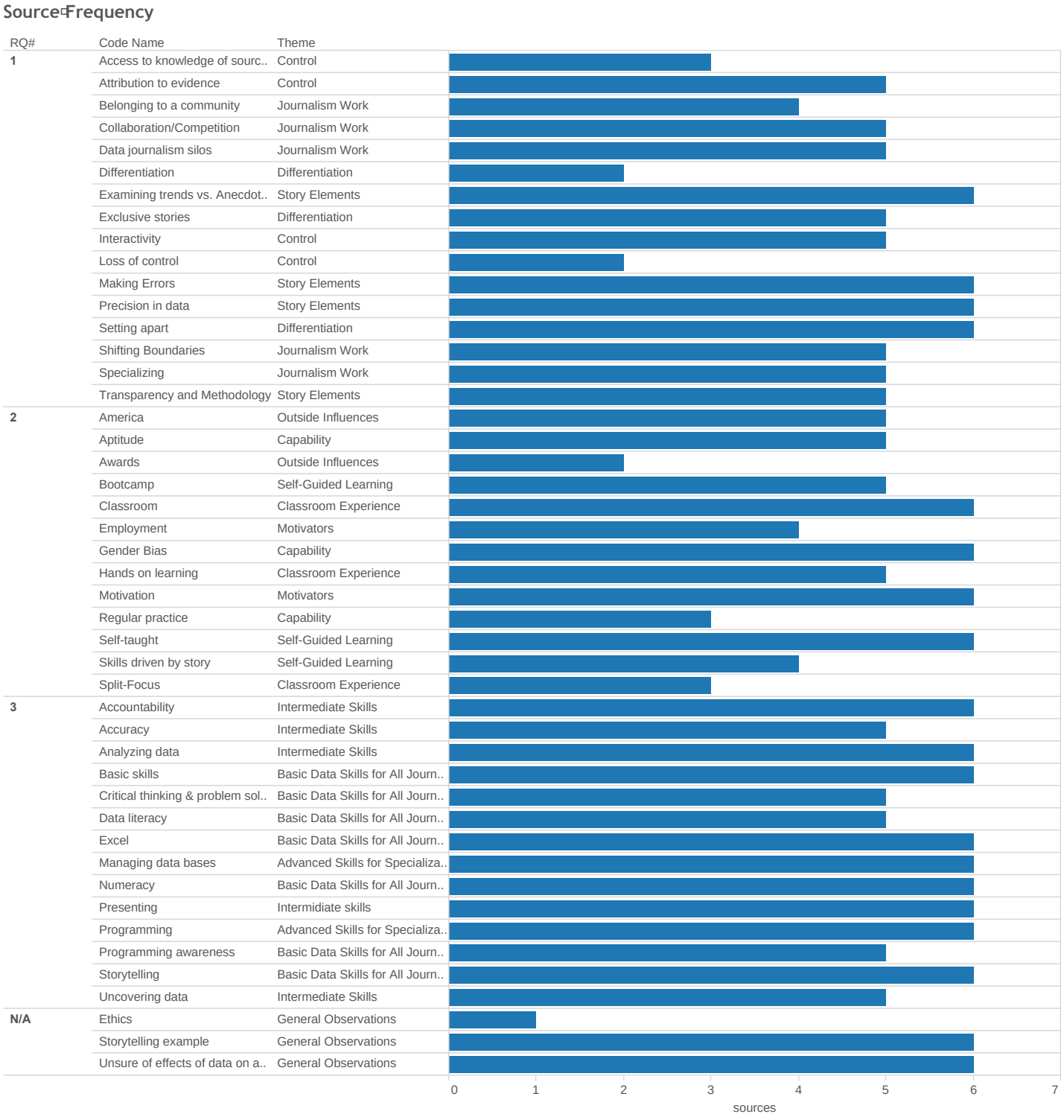

Figure 5. Coding frequency by saturation of sources.

Validity and reliability

As was described in the methodology chapter, validity and reliability have been addressed in each stage of this study. Mayan (2009) writes that validity refers to having "confidence that our conclusions, however unfinalized or problematized they are, come from the independent variable of the data" (p. 106). This confidence in my conclusions was built through my coding process, and as the figures above show the codes and themes which were generated were common to most or all participants and directly related to the RQs. My instrument, analysis and discussion have been triangulated using a thorough literature review. Rigor in qualitative inquiry is also substantiated by its iterative nature: as the data were collected, I reflected on each transcript and began initial coding as the other interviews were being conducted, thus analyzing the results concurrently. This leads to a "mutual interaction between what is being learned and what one needs to learn" (Morse, in Mayan, 2009, p. 110), allowing for adjustments to be made along the way, rather than an approach where all the data is gathered then analyzed as a next step which can compromise rigor in the data analysis phase (Mayan, 2009).

Reliability in qualitative inquiry, as described by Mayan (2009), is found in repetition: "It comes from hearing/seeing, over and over again, similar common experiences among participants, even if there are both common and contradictory experiences in the same data set" (p. 106). The tables above illustrate the saturation of experiences in what was shared by participants; this has been made visible in the codebook (Appendix A). The discussion section which follows will explore how these commonalities in data add to the discourse on the place of data journalism in Canadian journalism education.

Discussion

The scope of this research project was to look at the Canadian landscape of journalism instruction— how the development of data journalism is changing journalism practice and how

these changes are affecting journalism education. My goal was also to develop a typology of data journalism skills and offer insight into how best to acquire them. It is my intention that this project will add to the field of data journalism research by giving a snapshot-in-time overview of the state of data journalism education in Canada, an area of research which is lacking in the literature, as well as to provide guidance for journalism schools wishing to expand or refine their programs in this area. For RQ1 and RQ2, I have used an inductive process to build theory from my observations, modeled after De Vaus (2001). For **RQ3**, I have developed a typology of data journalism skills to guide journalism educators.

My professional background positions me as an "insider" (Mayan, 2009, p.79), which gives me the benefit of my professional connections, the ability to speak the same language as the participants, and common experiences, which enhance my ability to build rapport. While my position as an insider may have influenced both my analysis and conclusions, I am confident that my reflexivity throughout the process has allowed me to be rigorous. If my study was to be replicated by an outsider, it is possible that they would have different results because they may have probed in different areas than I did, and would not have taken the same things for granted (Mayan, 2009).

As was outlined above, Rogers' theory outlines four main elements of diffusion as "(1) an *innovation*, (2) is *communicated* through certain *channels* (3) over *time* (4) among members of a *social system*" (p. 11). I did not find that participants were diffusing their knowledge throughout the post-secondary institutions to other instructors (the post-secondary instructor social system), though they do communicate the skills to their students (the journalism student social system). There seems to be little transfer of the knowledge to other non-data journalists in their faculties for a variety of reasons including that the participants, and many other instructors,

are mostly part-time. I also expected to find more evidence of key informants' opinions that data journalism storytelling itself was an innovation, instead the interviews found key informants felt the core concepts of storytelling remain the same. Further, according to Rogers' theory, the tools by which data journalists tell these stories do not meet all the criteria to be defined as an innovation. Data journalism knowledge does give a *relative advantage* (in this case social prestige and employability) to those who adopt the skill set, and the key informants interviewed see it as having *compatibility* with their existing journalism values in pursuing accountability and good storytelling. The work created using data journalism tools do have a high level of *observability,* which may lead to journalists in the same newsroom (or faculty) to request more information about their use. However, as stated by participants, data journalism tools can have too much *complexity* for every journalist to incorporate into their work, and as has been illustrated it is difficult to dabble in these tools, thus not meeting the criteria of *trialability.*[3]

There was less discussion than expected about the shifting ethics of data journalism practice, however there is a rich exploration of this topic in the literature. I expected the data journalism instructors I interviewed to have a clearer sense of the impact of data journalism on the audience's perception of these stories as being more comprehensible or credible, and yet none of them cited any non-anecdotal evidence of this being so.

RQ1: How is the development of data journalism changing journalism practice?

This question has been addressed at length in the literature review and it was my intention through the semi-structured interviews with participants to gain a sense of how key

[3] Trialability refers to "the degree to which an innovation may be experimented with on a limited basis" (Rogers, 2003, p. 16). In this context, it refers to the ability for journalists to use data skills on a limited basis in some of their work, some of the time. While data journalism can be learned by doing by a journalist who is considering adopting this kind of approach, not all journalism stories are suited to data approaches and so trialability can be limited.

informants perceive these shifts, and how they position them in their teaching. Throughout these discussions, participants talked about how they were able to bypass institutional gatekeepers to produce journalism which can "actually find out what the truth is" (Participant B). This idea of presenting a more truthful narrative is especially important because of the public discourse on fake news: if done properly, with fair analysis and transparent methodology, data journalism projects may be able to give the audience more confidence in the truthfulness of the story. In the interviews, key informants did discuss the necessity in methodological transparency, as was also found by Aitamurto, et al. (2011), to increase the credibility and impact for the audience. Collaboration with others, be it other data journalists or outside experts, is important to check the acuracy of data journalism work.

Though the audience had some control over the narrative through interactivity, key informants advocated against the kind of "data dump" (Participant A) described by Hammond (2015) and Lesage and Hackett (2013). The participants interviewed for this project felt that data journalism skills give them more control over the stories they tell and an ability to be "declarative" (Participant B) and to find story angles which are different from the angles promoted by the institutional sources of the data.

As illustrated by Figure 6 these observations from the data were used to answer **RQ1.** Findings from this study suggest that data journalism work gives reporters more control over the narrative by getting them closer to unfiltered information: they have more confidence in their accuracy through collaboration with other data journalists and use a more declarative voice.

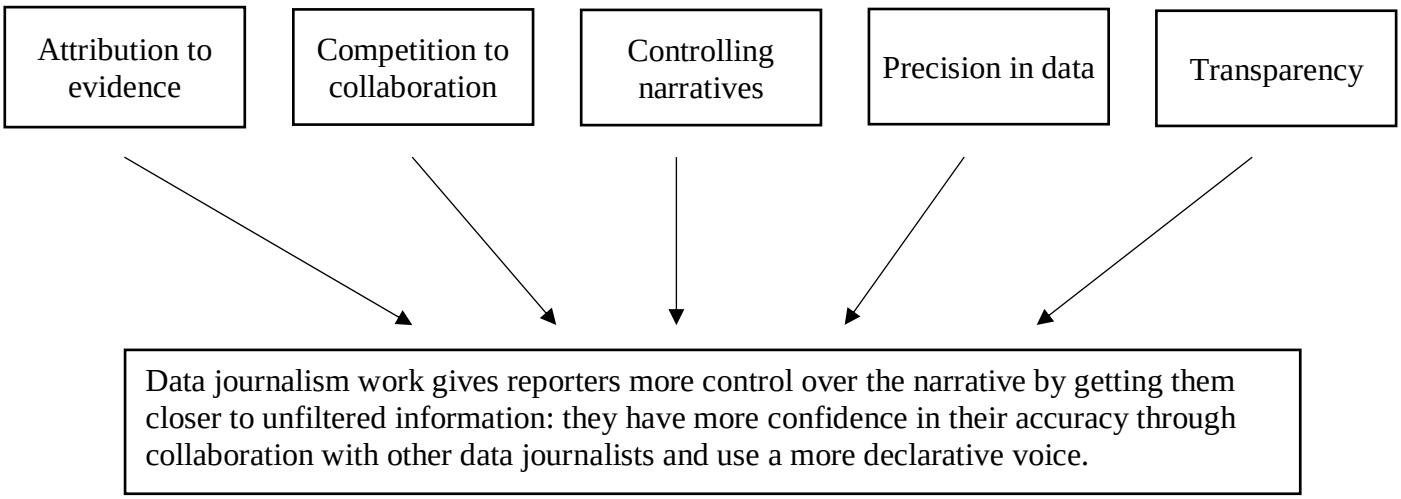

Figure 6. Observations from data on RQ1 and the resulting theory.

Further research in this area should include comparing stories told with aggregated data sources to those where the reporter performed the analysis to validate the belief among study informants that those stories are more precise; a study of the ways in which Canadian journalists are collaborating reporters who would be traditionally considered their competition, as well as with experts outside of their newsrooms similar to the research of Lewis and Usher (2014); and a study of methodological transparency in the sources of information in data-driven stories compared to more traditional news gathering.

RQ2: How is the development of data journalism impacting journalism education?

None of the participants learned their data journalism skills in a traditional journalism school, so perhaps it is not surprising they discussed several ways in which the skills could be learned outside of that setting. For those who want to pursue data journalism, organizations such as NICAR and CAJ offer skills training at their conferences, and institutions like the University of King's College and Carleton University host summer bootcamps. This model, according to

key informants, works to whet the appetite of journalists enough to motivate them to learn more on their own. Observations from the data answering **RQ2** are illustrated by Figure 7 below.

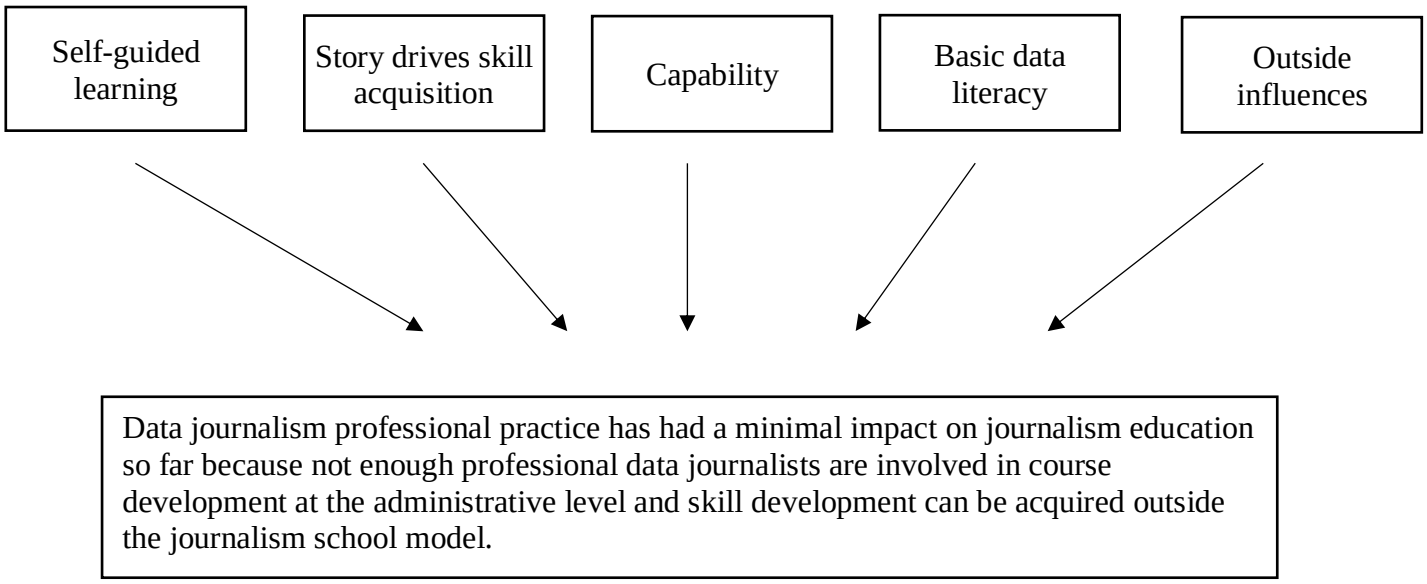

*Figure 7. Observations for **RQ2**.*

The model of learning skills through telling a story maybe the most effective for data journalism instruction: it has a history of use for teaching other journalistic skill sets, is used by the key informants in professional and classroom settings, and was found to be successful in studies by Graham (2015) and Treadwell, et al. (2016).

The question of which students are best positioned to excel in data journalism, and the factors that contribute to (or inhibit) this success, is one that requires further study for the Canadian journalism landscape. Key informants felt there were some people with more aptitude and a better chance of success, but that there was not a clear way to measure or address this programmatically. I did not find evidence of a "knowledge gap" (Tichenor et al., 1970) at work, as none of the participants felt there was a specific group that was more likely to gain a data

journalism proficiency. In the same vein, there was no evidence that students in a four-year university program versus a two-year college program had different levels of exposure to these concepts or were more apt to pursue data journalism successfully.

As suggested by Fink and Anderson (2015), all the participants interviewed felt there should be some basic level of statistics, data literacy, and numeracy education for all journalists. As was predicted in the research of Treadwell et al. (2016), participants expressed that if students knew there was an element of math or statistics in a data journalism course, and especially if this course was an elective, they would be hard presssed to get students to sign up. Instead, it would be more useful to have a core course which keeps "all the numbers in one class" (Participant F) early in the program. This would allow students to apply those numeracy and data literacy skills even if they decide not to specialize in data journalism, and may encourage them to pursue other data journalism skills once they have a chance to overcome their aversion to numbers.

The majority of key informants are not involved in developing the journalism programs as a whole and as such there seems to be limited input from them in terms of program planning. This likely contributes to the low incidence of teaching numeracy to journalism students, as does the likelihood that many of the journalism faculty members were working in the field before the growth of data journalism practice. In order to better understand the role "math phobia" (Participant B) plays in the selection of course content at the department level, a study incorporating elements of research by Nguyen and Lugo-Ocando (2015), and Griffin and Dunwoody (2015), would shed light on how entrentched these ideas are at Canadian journalism schools and may provide suggestions as to how they can be overcome.

RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned?

In order to answer **RQ3**, questions 4 and 5 of the interview guide (see Appendix F) asked participants to articulate what they saw as essential data journalism skills. Using the analysis of the transcript sections about specific skills and learning, I developed a typology which divides the skills into basic, intermediate, and advanced data journalism. My typology is of the skills as they can be taught, rather than how they are applied in the workplace. As was discussed above, participants did not express strong opinions on the timing of these courses; for a more precise prescription of how to design a journalism program with data elements, depending on the type of program and length, Berret and Phillips (2016) report *Teaching Data and Computational Journalism*, is insightful and very thorough. Their report offers model curricula and course outlines to incorporate data journalism skills from introductory to graduate levels, and recommends a foundational course in data journalism which teaches "basic principles of data analysis for the purpose of finding stories while cultivating a sense of the general techniques and possibilities of data-driven reporting" (p. 49). The scope of my research is more narrow in that I am looking specifically at the state of data journalism education in Canada, which includes fewer schools, but there are several similarities in what Berret and Phillips recommend and what participants expressed in our interviews. For example, participants I interviewed felt a similar foundational level of learning was needed across all journalism programs.

The tables below (Tables 2, 3, and 4) illustrate the typology developed from my findings: basic skills which all journalism students should learn so they can fulfill their accountability roles in a world that is becoming increasingly data rich, intermediate skills for those who may want to pursue some aspects of data journalism in the workplace, and advanced skills which may be taught in a higher level university program or outside of the traditional journalism school

model. Storytelling was excluded as a data journalism skill from the tables because it was felt by all participants to be the most important skill for engaging an audience no matter what the genre and so is not considered a separate skill as it is the core of what journalists do.

It is possible that these instructors fill a niche in the data journalism genre, and it remains more relevant for the journalism programs to train people in traditional media lines in which they have established internship programs. The skill sets described as basic amounts to knowledge deemed by the participants to be important for any journalist, and also positions students for further data journalism learning. Many participants talked about introducing basic skills early on in a program. There was less consensus around the appropriateness of intermediate skills for core or elective courses, and no clear agreement on how best (or whether) to integrate teaching of advanced skills within a journalism curriculum.

Skill Name	Specific to Data Journalism?	Significance
Critical thinking about data	No	"You need to learn how to tell good information from bad information. What's credible, what's not" (Participant A).
Data literacy	No	"Getting you to think in a data driven way will benefit you in whatever you do. It should be more analytical, maybe more skeptical" (Participant D).
Excel/ spreadsheets	No	"You want to be able to sort a spreadsheet you want to be able to filter a spreadsheet, you want to be able to do a basic calculation… And I think for most journalists, that's probably sufficient, if they are presented with data they will be able to answer some basic questions about it. If someone is making claims about data then they know enough to be able to say give me the raw data to make sure the conclusions that you are drawing are correct" (Participant B)
Numeracy	No	"If you don't know how to calculate a percentage change and you don't know the difference between a percentage point and a percentage, then you are going to be lost in that and it's going to be very easy for politicians and business leaders to snow you because you are not going to have the skills necessary to call them on it" (Participant C).
Programming awareness	No	"You need to be able to know what's possible so you can create your stories" (Participant F).

Table 2. Basic skills for all journalism students.

Skill Name	Specific to Data Journalism?	Significance
Data analysis	No	There is still a huge roll for journalists to play in interpreting and understanding and selecting information so as to guide the audience in the story, that hasn't changed (Participant A)
Uncovering data	No, it can also be used for investigative projects, navigating open data portals, and institutionally-generated data like statistics.	"I think that one of the concerns I have is that we get into a false sense of security by being comforted by the fact that there's all this open data floating around and all we have to do is sit back and dig into it, when I think we have to be more vigilant than ever. I am not quite sure anything has really changed dramatically, I mean, maybe the door has been pushed open slightly more but I still think we have to barge through it" (Participant C).
Transparency/methodology	Yes	"We use scientific methods to tell really important stories, or to inform the conversations about stories that are already out there" (Participant C).
Collaboration	No	"We are going to have to become more and more familiar with and more and more comfortable with, as a lot of our reporting falls into niche fields, especially as we start talking about internet security, hacking, encryption technology. These are fields that journalists are not by and large well equipped to understand without any external training" (Participant E).

Table 3. Intermediate skills to be offered after basic skills.

Skill Name	Data Journalism Specific?	Significance
Managing databases	Yes	Creating a data base from uncovered data allows for new insights "that allows you to say, not only when you are sighting the source, but this is the first ever analysis of this kind of this data set" (Participant C).
Advanced visualizations	Yes	"A good chart can tell a powerful story, not just throwing all the data at the reader and say, 'here you figure it out.' You are actually making the numbers sing somehow with math or scrollable narratives that are very visual" (Participant E)
Programming/web scraping	Yes	"The whole data journalism community has become more and more made up of developers, people who are coders by training" (Participant B).

Table 4. Advanced skills which can be learned at a higher level or outside of journalism school.

As advanced skills could be learned on the job or as a part of a self-guided process, it may be a useful pursuit for journalism schools to leverage the training capacity of data journalists to offer, or co-create, intensive summer bootcamps for their interested students, which would also serve mid-career journalists in their communities.

Limitations of the Study

It is acknowledged that there are limitations to this approach, namely that an exploratory study is broad in scope and does not attempt to make absolute conclusions, but rather explores and describes (Stebbins, 2001). Another limitation of my research is restricting my participants to data journalism instructors, the majority of whom work part-time in post-secondary education and so may not be as motivated, or able, to contribute to design and development of journalism programs at the department level. Future studies could widen the field to include surveys of other journalism school faculty and administration (including a comparative analysis of institutions

outside Canada), as was done by Dunwoody and Griffin, (2013), and Yarnall, et al. (2008), and/or a more thorough document analysis of course outlines and offerings. As participants were exclusively instructors of stand-alone courses, programs which may have elements of data journalism scattered over several courses were not investigated, which precludes insight into that way of delivering data journalism learning. Future research could also delve into a survey of the experience of journalism students in the classroom, and perhaps even what they are looking for in terms of their journalism education. This study focuses mainly on data journalism skills in a formal post-secondary setting, and as such an exploration of ad hoc ways to learn these skills outside of journalism schools (i.e., by mid-career journalists) in a Canadian context would fortify the literature. A longitudinal study of five or ten years duration following data journalism students into their careers would also benefit the field by following the skills they learn and the success they have in their careers. Finally, a limitation with my approach is that the field of data journalism is rapidly changing, while the process of program reviews at journalism schools happens at a much slower pace, suggesting that programs themselves may represent lagging indicators of student and employer needs.

Summary

In summary, the results of this study suggest that in answer to **RQ1**, participants primarily view data journalism as a toolkit to tell exclusive stories in new ways, and to hold institutions to account by performing original analysis. However, because so few data journalists have leadership roles in journalism faculties, as well as the slow pace of program development and the fast pace of how the field of data journalism is changing, **RQ2** found the adoption of basic data journalism skills has not yet permeated traditional journalism curricula at a level commensurate with the rapidly-evolving state of journalism practice. In answer to **RQ3,** the data

from key informants revealed what they would like to see in terms of data skills for all journalism students, as well as which skills can advance data journalism education. A typology of skills based on these findings suggests journalism educators can scaffold their programs in a way that will set journalism students up to work effectively with data. Following this findings and discussion chapter, my conclusions to the study into data journalism education in Canada will be presented.

Conclusion

The question of how the proliferation of data and its impact in journalism is one that has profound implications on both the way journalism is practiced as well as how data journalism is taught in Canada. The development of data journalism has been unfolding over many years: from the early days of Meyer's *Precision Journalism,* to Computer Assisted Reporting and now the contemporary implementation of new tools for a more quantitative storytelling lens. However, there has been a modest amount of research into the field of data journalism education, and very little academic exploration of the topic in Canada. In order to fill this gap in the literature, three research questions were developed:

RQ1: How is the development of data journalism changing journalism practice?

RQ2: How is the development of data journalism impacting journalism education?

RQ3: What are the core skills needed to be a successful practitioner of data journalism and where are they learned?

The concluding chapter of my study is organized as follows: a summary of the key findings from my research will be presented, as well as a section which places those finding in context for the field; a section which explores the limitations of the study and identifies opportunities for future research will follow the findings, which will include specific recommendations. In concluding

this section, a summary will be presented along with how this study contributes to the professional practice of data journalism instruction.

Summary of Findings

As there were three RQs in this study, the findings for each one of them will be addressed separately. **RQ1** set out to understand how key informants in the field of data journalism instruction felt about how the adoption of data journalism storytelling was impacting journalism work. The key finding from the data was that adopting a data journalism approach gives reporters a feeling of more control: it sets the reporter apart from the pack by giving them more leeway in pitching stories, and elevates their news outlets by having an angle that no one else does. Key informants held that this gives journalists more control as they are closer to the evidence by reducing institutional filters and have the ability to do more analysis and rely less on attribution. By understanding these attitudes on structured data's place in journalism, it informs how the skills are selected and taught in the classroom.

RQ2 asked how the attitudes toward data journalism practice are impacting journalism education and found the majority of advanced data journalism skills are being learned outside of a traditional post-secondary classroom. Participants learned most of their skills in a self-guided way, and the majority of them do not hold positions in their faculties which influence the overall program course composition. There was a consensus in the need for better data literacy and numeracy education in journalism schools. A feeling was expressed that small classes, with many opportunities for hands-on learning, are required in order to facilitate learning complex data journalism skills. The participants advocated for a basic level of math, statistics, data literacy and data journalism skills to be adopted for all journalism students as an enhancement in their ability to hold those in power to account. Participants did feel that reserving more advanced

skills training for other settings such as bootcamps, conferences and specialized workshops would be beneficial as not all journalism students have the required aptitude or interest, nor do all journalism jobs need advanced data skills.

The understanding of what the participants viewed as being an achievable level of learning in the post-secondary setting led to the creation of a typology which was created to answer **RQ3,** which asks which skills are needed and where they are learned. The basic level of skills, includes numeracy, critical thinking about the origins of data, competency with spreadsheets, and an understanding of programing languages and how they work. These basic skills should be offered to all journalism students no matter what their storytelling medium. The intermediate set of skills includes the ability to find data through Access to Information requests and open data portals, and to analyze the information in a scientific way and incorporate the methodology of the project in the reporting. This intermediate set of skills also includes the ability to collaborate and check analysis with others: colleagues in other, sometimes competing, newsrooms, outside experts, and in some cases the institutions who hold the data. These skills can be learned on the job in a story-specific approach, or as an advanced elective in the post-secondary setting. The more advanced skills, which include specializations like advanced data visualizations, programming languages, large-scale web scraping, and managing purpose-built databases, are most likely to be learned as needed to tell data stories or at a more immersive setting like a bootcamp. Participants did not expect data journalists to have all of these skills at an advanced level, but instead advocated for them to specialize in one or two areas.

The findings for all three **RQs** are significant to two principal groups—journalism employers and educators—who can apply them in practical context. Findings give to news leaders in Canada a sense to news leaders in Canada of how data skills impact the field, as well

as the current state of data journalism instruction in post-secondary institutions. This can guide newsroom managers in how best to train their employees if they want to strengthen data journalism competencies within their organizations. My research also offers a road-map of sorts to those who teach and administer journalism programs of how to adjust the skills training they offer in order to leverage data storytelling tools and respond to the challenges a data-rich world presents to journalists.

Findings in Context

The literature review confirmed there exist studies on how data journalism is changing journalism work, as well as some research on data journalism education. However, there were no studies that looked at data journalism education in Canada; this study addresses that gap in the literature. My research also serves to highlight further gaps in knowledge that can be addressed by future research, including whether there is a different effect on the audience for data journalism projects in terms of perceived credibility, and a longitudinal study exploring whether data journalism skills truly lead to better employment opportunities and outcomes. While selecting the key informants, I was in touch with many post-secondary journalism schools, and several of them indicated they were looking for guidance in the area of data journalism education. This study can assist educators and administrators by enhancing their understanding of the shifts underway, as well as a more practical approach to program planning by relaying the knowledge of the key informants in the field.

While the amount of literature on the Canadian landscape is lacking, other work has been done in the area of my research which may provide alternatives to my findings. In particular, the work of Berret and Phillips, whose paper *Teaching data and computational journalism* (2016), studied the landscape of data journalism in the United States and concludes by laying out several

suggestions on how to design data journalism education programs. Donsbach (2014), who has worked predominantly in Europe, advocates for journalists to have deep knowledge and "subject competence" (p.668) which is achieved by double majors at the bachelor level, or graduate study of a specific field before covering that area as a journalist. This was not something that was suggested by key informants in my research, suggesting there may be differences in the Canadian approach to data journalism education in Europe.

Future Direction of Research and Limitations of Study

The following section will identify the limitations of my study and will make suggestions for future research to address them. The first limitation is that exploratory design "is often regarded as little more than a preliminary step toward specific and focused causal research to generate required hypotheses" (Streb, 2012, p. 373) and the flexibility in this design methodology can be seen as a weakness because of a lack of "specific, theory-based prior assumptions" (Streb, 2012, p. 373). Exploratory design is applied as a first step, and is useful to explore an area which does not have much preliminary research (Streb, 2012) as in the case of data journalism education in Canada. The exploratory design offers informed observations which allow for planning future research, suggestions for which I have outlined below.

First, the selection of participants was restricted to journalism instructors who teach a stand-alone course and did not consider the input of administrators and instructors at other post-secondary institutions. A cross-sectional design study could reveal the differences in the outcomes for journalism students at institutions which offer stand-alone classes and those that do not. This would serve to illustrate if it is beneficial to adopt this kind of learning in all journalism schools in Canada.

Second, while my research provides a typology of skills, it does not provide findings on how best to introduce this into a journalism program in terms of practicalities: what other content can be dropped to make room, should students be streamed (and if so how), and what are the implications on internships are a few of the questions that need exploration. A study using action research design could be undertaken to introduce data journalism learning in a post-secondary setting (perhaps at one of the aforementioned institutions which is exploring incorporating this into their programs) in order to have a sense of the challenges introducing this content could present, and to suggest some ways to mitigate these difficulties. This could also be done using a case study design as was done in Treadwell, et al. (2016).

Thirdly, my research is an exploratory design which yields only a snapshot in time. The remedy of this limitation would be a longitudinal study which would explore the changes in attitudes towards data journalism education over time, as was done in Dunwoody and Griffin (2013). Another longtudinal study which would serve to illuminate this topic would be to follow data journalism students into their careers over five or ten years to get a sense of the real world outcomes of this kind of learning.

Finally, the key informants spoke about a lack of numeracy and statistical and data literacy in journalism schools, and often cited the generalization that journalists have a problem with math. These assertions are anecdotal, and a study using experimental design, whereby the numeracy and statistical literacy levels of students in journalism schools could be systematically tested by comparing those programs with more numerical core class content to those with less on offer. This kind of study could reveal if the lack of numeracy in journalism students is because of an aversion in students or, if as was found by Dunwoody and Griffin (2013), there is a lack of will to introduce these subjects on the part of administrators.

Conclusion

The purpose of this study is to explore how journalism is changing with the adoption of data journalism skills, how that change is manifesting in journalism schools, and which skills are most necessary to be a successful data journalist and where they are learned. The key findings of this research have value to journalists who want to learn data skills, the employers who hire them, and the instructors and administrators who plan the programs that train them. My research has found participants are motivated to learn, use, and teach data journalism skills by their increased ability to control the narrative of their stories, their ability to shape their journalistic assignments and to enhance their roles as analyzers of facts instead of attributors. I found the understanding that this skill set is essential has not yet permeated journalism faculties across the country, nor is it certain that all data journalism skills are necessary for all journalism students or reporters in the field. Hands-on learning, small class sizes, and an opportunity to further learning in a self-directed way are all influencing the uptake of data journalism skills, as much of this is also learned on the job for specific stories. A typology of skills was developed in my research in order to lay out a roadmap for educators interested in exploring how to incorporate this learning into their programs. It is my hope that this study will allow journalism educators to have a broad-based understanding of the field of data journalism, and how to teach it in a way that best serves their students and the industry for which they are preparing them.

References

Aitamurto, T., Sirkkunen, E., & Lehtonen, P. (2011). Trends in data journalism. In *Next Media* (0–27). Retrieved from http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In Data_Journalism.pdf

Anderson, C. W. (2015). Between the unique and the pattern. *Digital Journalism, 3*(3), 349–363.

Arksey, H. & Knight, P. T. (1999). Meanings and data analysis. In *Interviewing for social scientists* (149-173). SAGE Publications Ltd.

ACM Computing Curricula Task Force, Ed., (2013*). Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. ACM, Inc. Retrieved from: https://www.acm.org/education/CS2013-final-report.pdf

Bamberger, M., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: working under budget, time, data, and political constraints.* Thousand Oaks, CA.: SAGE.

Bazeley, P. (2013). *Qualitative data analysis: practical strategies*. London: SAGE.

Berret, C., & Phillips, C. (2016). Teaching data and computational journalism. Retrieved from: https://journalism.columbia.edu/system/files/content/teaching_data_and_computational_journalism.pdf.

Borges-Rey, E. (2016) Unravelling data journalism. *Journalism Practice, 10*(7), 833-843.

Bradshaw, P. (2014). What is data journalism? *Ethics for Digital Journalists: Emerging Best Practices,* 202-219. Retrieved from: https://books.google.ca/books?hl=en&lr=&id=QJxeBAAAQBAJ&oi=fnd&pg=PA202&dq=Bradshaw,+P.+(2014).+Data+journalism&ots=Ig3vJ-8WSe&sig=7uH2_xsQBlReV0w64s2b-wU7kSY#v=onepage&q&f=false

Brinkman, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing.* Thousand Oaks, CA.: SAGE.

Broussard, M. (2015). Artificial intelligence for investigative reporting. *Digital Journalism*, *3*(6), 814–831.

Broussard, M. (2015a). Teaching coding in journalism schools: Considerations for a secure technological infrastructure. Retrieved from: http://cj2015.brown.columbia.edu/papers/teaching-coding.pdf

Broussard, M. (2016). Big Data in practice. *Digital Journalism, (4)*2, 266-279.

Cairo, B. Y. A. (2014). Ethical infographics. *IRE Journal*, 37, 25–27.

Coddington, M. (2015). Clarifying journalism's quantitative turn. *Digital Journalism, (3)*3, 331-348.

Cohn, V., Cohn Runkle, D., & Cope, L. (2011). *News & numbers: a writer's guide to statistics.* 3rd ed. / Malden: Wiley-Blackwell.

Cohen, S., Hamilton, J. T., & Turner, F. (2011). Computational journalism. *Communications of the ACM*, *54*(10), 66.

Cook, L. (2015) *Why journalism students don't learn CS: Reporting back from a year of on-campus research*. Retrieved from: https://source.opennews.org/en-US/learning/journalism-students-and-cs/

Corbin, J. & Strauss, A. (2008). *Basics of qualitative research 3rd ed.: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: SAGE.

Cusatis, C. & Martin-Kratzer, R. (2009). Assessing the state of math education in ACEJMC-accredited and non-accredited undergraduate journalism programs. *Journalism & Mass Communication Educator*, *64*(4), 356–377.

Cuthill, M. (2002). Exploratory research: Citizen participation, local government and sustainable development in Australia. *Sustainable Development*, *10*(2), 79-88.

De Vaus, D. A. (2001). *Research design in social research.* London: SAGE Publications Ltd.

Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, (*8)11*, 1–18.

Dick, M. (2013). Interactive infographics and news values. *Digital Journalism*, (*8)11*, 1–17.

Donsbach, W. (2014). Journalism as the new knowledge profession and consequences for journalism education. *Journalism*, *15*(6), 661–677.

Dunwoody, S., & Griffin, R. J. (2013). Statistical reasoning in journalism education. *Science Communication*, *35*(4), 528–538.

Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* London: Bloomsbury.

Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional "fourth estate" function. Journalism, *17*(1), 85-96.

Fink, K., & Anderson, C.W., (2015) Data journalism in the United States, *Journalism Studies, (16)*4, 467-481.

Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, *6*(2), 157–171.

Flick, U. (2014). *The SAGE handbook of qualitative data analysis*. London: SAGE.

Gillmor, D. (2016). Towards a new model for journalism education. *Journalism Practice, (10)*7, 1–5.

Graham, C. (2015). By the numbers: Data journalism projects as a means of teaching political investigative reporting. *Asia Pacific Media Educator*, *25*(2), 247–261.

Gray, J., Bounegru, L., & Chambers, L. (2012). *The data journalism handbook: [how journalists can use data to improve the news]*. Sebastopol, CA: O'Reilly Media.

Grbich, C. (2013). *Qualitative data analysis: an introduction.* 2nd ed. London: SAGE.

Griffin, R. J., & Dunwoody, S. (2015). Chair support, faculty entrepreneurship, and the teaching of statistical reasoning to journalism undergraduates in the United States. *Journalism (17)*1, 97-118.

Gynnild, A. (2013). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism, (15)*6, 713 - 730.

Hammond, P. (2015). From computer-assisted to data-driven: Journalism and Big Data. *Journalism (18)*4, 408-424.

Hermida, A., & Young, M. L. (2016). Finding the data unicorn: A hierarchy of hybridity in data and computational journalism. *Digital Journalism*, *5*(2), 159-176. http://doi.org/10.1080/21670811.2016.1162663

Hewett, J. (2015). Learning to teach data journalism: Innovation, influence and constraints. *Journalism, 17*(1), 119-137.

Herzog, D. (2016). *Data literacy: a user's guide*. Thousand Oaks, CA.: SAGE.

Karlsson, M. (2011). The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority. *Journalism*, *12*(3), 279–295.

Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice, 16*(1), 55-72.

Lee, K. C., & Fleming, C. A. (1995). Problems of introducing courses in computer-assisted reporting. *Journalism & Mass Communication Educator*, *50*(3), 23.

Lesage, F., & Hackett, R. A. (2013). Between objectivity and openness - The mediality of data for Journalism. *Media and Communication, 1(*1), 39–50.

Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism. *Digital Journalism, (2)*3, 383-393.

Lewis, S. C., & Westlund, O. (2015a). Actors, actants, audiences, and activities in cross-media news work. *Digital Journalism*, *3*(1), 19–37.

Lewis, S. C., & Westlund, O. (2015b). Big Data and journalism. *Digital Journalism*, *3*(3), 447–466.

Mack, N., Woodsong, C., McQueen, K. M., Guest, G., & Namey, E. (2011). *Qualitative Research Methods: A data collector's field guide* (pp. 1-120). Family Health International. Retrieved from https://www.fhi360.org/sites/default/files/media/documents/Qualitative Research Methods - A Data Collector's Field Guide.pdf

Mayan, M. J. (2009). *Essentials of qualitative inquiry.* Walnut Creek, CA: Left Coast Press.

McConway, K. (2015). Statistics and the media: A statistician's view. *Journalism, (17)*1, 49-65.

McGregor, G. (2014). Finding the story lost in data. *Media*, *16*(1), 13–14. Retrieved from *caj.ca/images/downloads/Media/media_winter_2014_final.pdf*

Merrigan, G., Johnston, R. T, & Huston, C. Logan. (2012). *Communication research methods.* Canadian ed. Don Mills, ON: Oxford University Press.

Meyer, P. (2002). *Precision journalism: a reporter's introduction to social science methods.* 4th ed. Lanham, MD.: Rowman & Littlefield Publishers.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook.* 3rd ed. Thousand Oaks, CA: SAGE.

Nguyen, A., & Lugo-Ocando, J. (2015). Introduction: The state of statistics in journalism and journalism education - issues and debates. *Journalism, (17)*1, 3-17.

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.

Pavlik, J. V. (2013). Innovation and the future of journalism. *Digital Journalism*, *1*(2), 181–193.

Plaue, C., & Cook, L. R. (2015). Data journalism : Lessons learned while designing an interdisciplinary service course. *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)*, 126–131.

Primo, A., & Zago, G. (2015). Who and what do journalism? An actor-network perspective. *Digital Journalism*, *3*(1), 38-52.

Patton, M. Quinn. (2008). *Utilization-focused evaluation.* 4th ed. Thousand Oaks, CA.: SAGE.

Rogers, E. M. (2003). *Diffusion of Innovations.* 5th ed., New York: Free Press.

Roulston, K. (2014). Analysing interviews. In Flick, U. *The SAGE handbook of qualitative data analysis* (pp. 297-312). London: SAGE Publications Ltd.

Royal, Cindy. 2010. "The journalist as programmer: A case study of The New York Times interactive news technology department." Presented at the International Symposium for Online Journalism, Austin, TX, April 23. Retrieved from: https://online.journalism.utexas.edu/2010/papers/Royal10.pdf.

Ryan, M. (2001). Journalistic ethics, objectivity, existential journalism, standpoint epistemology, and public journalism. *Journal of Mass Media Ethics, (16),1,* 3–22.

Schutt, R. K. (2012). *Investigating the social world: the process and practice of research.* 7th ed. Thousand Oaks, CA.: SAGE.

Sparre, K., & Færgemann, H. M. (2016). Towards a broader conception of entrepreneurial journalism education. *Journalism Practice*, *10*(2), 266–285.

Splendore, S. (2016). Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass*, *10*(5), 343–352.

Splendore, S., Salvo, P. Di, Eberwein, T., Kus, M., & Porlezza, C. (2015). Educational strategies

in data journalism: A comparative study of six European countries. *Journalism, (17)1*,138-152.

Stebbins, R. A. (2001). *Qualitative research methods: Exploratory research in the social sciences.* Thousand Oaks, CA.: SAGE.

Stray, J., Ma, Y. & Chun, R. (2016) Interactive data journalism: A one-semester syllabus. Retrieved from: https://source.opennews.org/articles/interactive-data-journalism-one-semester/

Streb, C. Exploratory Case Study. In Mills, A. J., Durepos, G. & Wiebe, E. (2010). *Encyclopedia of case study research.* Thousand Oaks, CA: SAGE.

Tabary, C., Provost, A.-M., & Trottier, A. (2015). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism 17*(1), 66-84.

Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media flow and differential growth in knowledge. *Public Opinion Quarterly, 34*(2), 159-170.

Toughill, K. (2012). Selling out or saving journalism? *Media, 16*(1), 37-38.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact.* Chichester, UK: Wiley-Blackwell.

Treadwell, G., Ross, T., Lee, A., & Lowenstein, J. K. (2016). A numbers game: two case studies in teaching data journalism. *Journalism & Mass Communication Educator, 71*(3), 297-308.

Tufte, E. R. (2006). *Beautiful evidence.* Cheshire, Conn: Graphics Press.

University of Southern California, (n.d.). *Organizing Your Social Sciences Research Paper: Types of Research Designs.* Retrieved from http://libguides.usc.edu/writingguide/researchdesigns

Vallance-Jones, F., Pilhofer, A., Dowdell, J., & McKie, D. (2009). *Computer-assisted reporting: a comprehensive primer.* Don Mills, ON.: Oxford University Press.

Vallance-Jones, F., & McKie, D. (2017). *The data journalist: getting the story.* Don Mills, ON.: Oxford University Press.

Wolf, M., & Barzillai, M. (2009). The importance of deep reading. *Educational Leadership*, *66*(6), 32-37.

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.

Young, M. L., & Hermida, A. (2015). From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism, 3*(3), 381–397.

Appendix A: Search Terms Used

1. Algorithm or algorithms and journalism
2. "Big data" or data and journalism
3. "Computer assisted reporting"
4. Computational and journalism
5. "Data Analysis" and journalism
6. "Data-based reporting"
7. "Digital journalism" and data
8. Innovation and journalism
9. Inforgraphics and journalism
10. Journalism and education or students
11. Journalism and numeracy
12. Journalism and statistics or "statistical reasoning"
13. Programmer-journalist
14. Journalism and visualization
15. Journalism and science or computer science
16. Networked journalism
17. Journalism and open source
18. Journalism and pedagogy and data

Appendix B: Personalized Recruitment Email

To:
From: Jennifer Leask (jleask@ualberta.ca)
Subject: Data Journalism
(A custom introduction depending if I know the participant personally or am introducing myself to them.)

I've taken a leave to do my MA in Communications and Technology at the University of Alberta and I am hoping you would agree to participate in my capstone project, which focuses on data journalism skills, how they are being acquired by journalists, as well as what skills journalists need to thrive in a world full of data. As one of the leaders in this field in Canada, I am hoping you would agree to participate in my data gathering for my project, which will include doing a semi-structured interview with me on the topic in the coming weeks.

I'm going to interview people who work as and teach data journalism in Canada for my project and I wanted to see if you would be willing to participate. I'm recruiting data journalism instructors, who I feel are key informants in the field, to participate in my research, so if you have any suggestions of other people who should be included I would be very happy to take suggestions.

Because of the way the University research/ethics process is set up, I won't have the go-ahead for data collection for a few weeks yet, at which time I will be sending you a more detailed consent letter which will outline what kind of information I will collect, how I will store it and anonymize it in my final report. Your commitment will be to do an interview by Skype or phone, which I will record and transcribe. You will have the option to check what I will include from your interview before I put it into the report and you can opt out at anytime.

Please let me know if you would be interested in being a part of this when you've had some time to think it over. I can be reached through this email or at 604.312.8526.

Thank you in advance for your consideration.

Warm regards,

Jennifer

Appendix C: Journalism Schools in Canada Offering Data Journalism Courses

School	Location	Course
Algonquin College	Ottawa	JOU 1526 Computer Assisted Reporting
Humber College	Toronto	JRNL 222- Computer Assisted Reporting, JRNL 223- Data Journalism
Lethbridge College	Lethbridge	DCM-2250- News Reporting and Data Journalism I DCM-2254- News Reporting and Data Journalism II
Loyalist College	Belleville	JOPB2008 Data Journalism
Sherridan College	Oakville	JOUR 52500 Data Journalism
St. Clair College	Windsor	JRN 207- Math for the Media JRN 309 Digital Design for Journalists
UNIVERSITIES		
Carleton	Ottawa	(BJ) JOUR4208E Data Journalism and Storytelling
Carleton	Ottawa	(MJ) MJ JOUR5508E Data Journalism and Storytelling
Kwantlen Polytechnic	Surrey	JRNL 4165 Data Journalism

		JRNL 4260 Computer Programming for Journalists
Ryerson	Toronto	JRN 204 Data Journalism
University of Kings College	Halifax	Master of Journalism (Data Specialization)
University of British Columbia	Vancouver	Data visualization
University of Toronto Global Fellowship	Toronto	Data Visualization, Data Journalism

Appendix D: Confirmation of Course Offering

Hello:

I am working on my MA in Communications and Technology at the University of Alberta, and the focus of my capstone project is an exploration of how the development of Quantitatively Oriented Journalism (data journalism) is impacting journalism education.

I have looked through your journalism program website and I do not see a specific course with a title like computer assisted reporting, data journalism or data visualization. I know that sometimes the course titles do not always reflect what is currently being taught, so if you do offer instruction in this area, be it for an elective or as a core course, can you please let me know which course it is and put me in touch with the instructor? I'm hoping to interview several journalism instructors for my data collection phase.

I can be reached through this email or at 604.XXX.XXXX.

Thank you for your time.

Regards,

Jennifer Leask

Appendix E: Information and Consent Letter

Study Title: Data Journalism in Canadian Journalism Schools: Learning Challenges and Opportunities.

Research Investigator:
Jennifer Leask, University of Alberta

jleask@ualberta.ca

Supervisor:
Dr. Gordon Gow, University of Alberta
MACT Program, 10230 Jasper Avenue
Edmonton, AB, T5J 4P6
ggow@ualberta.ca
780-492-6111

Background

You are being asked to participate in this research project because you have been identified as a key informant in the area of data journalism education in Canada. This study is part of the requirement for me to complete my Master of Arts in Communications and Technology at the University of Alberta.

Purpose

My research is to determine what data journalism (or quantitatively oriented journalism, QOJ) instructors believe are essential skills for journalists to work in a world of big data in order to fulfill their role as watchdogs who serve the public. I will also investigate how these skills are learned and what motivates data journalists to specialize, as well as how data journalism is changing the responsibilities and perceptions of journalism work.

Study Procedures

As a participant in this study, I will ask you to do a semi-structured interview with me sometime in March or April, 2017 at your convenience. The interview will be done in person, or via Skype/FaceTime depending on your proximity and preference. It will take between 45 minutes and one hour. I will record the audio of the interview for accuracy, and will transcribe it after our discussion for my data analysis. The interview recording as well as the transcripts will be kept on my encrypted and password protected personal computer, as well both the University of Alberta cloud backup and a hard drive, which is also encrypted and password protected. After the interview has been transcribed, I will email you a copy of the transcription so you can check it for accuracy, and to give you an opportunity to add something you feel was missed (either through a follow up email or interview). I will also ask if you want a copy of the final report to read once the study is complete.

I also ask that before the interview, you share with me a course outline for the data journalism course(s) you teach in order to prepare for our conversation.

Benefits

There are no direct benefits to participants, however you may be prompted to examine what you think is important to teach or learn (as a skill set), which may motivate you to change behaviors in a positive way.

The benefit of this research is that it will inform journalists, as well as practitioners and instructors of data journalism, which may result in more instructors or journalism faculties adopting aspects of the kinds of instruction in the report.

It will build on previous research to give a Canadian focus to the topic as well will be of interest to journalism instructors across the country. There is a significant gap in the literature, especially with regards to Canadian research, which will be partly addressed by my work.

Risks

There may be risks to being in this study that are not known. A risk to reputation could result from a participant being critical of the structure of his/her newsroom or academic institution, and their supervisors being displeased with this if the supervisor reads the report and can identify them.

Confidentiality & Anonymity

It is my intention to use this research for my capstone project at the University of Alberta, and I would like to publish the findings in an academic journal, or use them for further academic research, if that is appropriate.

Your raw data will remain confidential, but may be reviewed by my supervisor, Dr. Gordon Gow, if he feels it is necessary. As the field of data journalism instruction is small in Canada, anonymity cannot be guaranteed, as with so few institutions offering stand-alone classes, it is likely that anyone who wanted to figure out who participated in this study could do so with a little work. I cannot assure anonymity, but I can assure you there will be confidentiality in my process: both in my transcripts and the final report, I will use a pseudonym for your name and will name the institution "Institution A, B, C, etc…" I will also remove any identifying information, for example the city where you live from the final report.

I will use your comments to explore what is happening in data journalism and to come to some general conclusions as to how it is being taught in Canada and why. Some of your comments may appear as direct quotes, after checking with you for accuracy, in my final report. More general comments will be attributed with terms like "an instructor in a graduate program" or "an instructor at the college level," but again, an interested reader could likely deduce who the comments come from.

Voluntary Participation

I'm very thankful for your agreement to participate, but please remember you are under no obligation to do so. Your participation is completely voluntary, and you are not obliged to answer any specific questions even if participating in the study.

Freedom to Withdraw

If you would like to opt out of the study, you can do that at any time without any kind of repercussions. You can also ask that your data be withdrawn from the study and destroyed. I will be submitting the data to my supervisor on April 30, 2017 so if you do not want your data to be included, I ask that you let me know via email by clearly stating you want to withdraw within two weeks of our interview or before April 15, 2017 and I will destroy it and not include it. This includes any supplementary material (course outlines) you have given me.

Future Research

We may use this data in future research, but if we do we will obtain ethics approval at that time.

Further Information

If you have any further questions about this study or your role in it, please get in touch with me and we can discuss them. I can be reached through jleask@ualberta.ca or 604.XXX.XXXX. If you would like to speak to my capstone supervisor, he can be reached through the contact information above.

The plan for this study has been reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Research Ethics Office at (780) 492-2615.

Consent Statement

I have read this form and the research study has been explained to me. I have been given the opportunity to ask questions and my questions have been answered. If I have additional questions, I have been told whom to contact. I agree to participate in the research study described above and will receive a copy of this consent form. I will receive a copy of this consent form after I sign it.

_____ _____

Participant's Name (printed) and Signature Date

_____ _____

Name (printed) and Signature of Person Obtaining Consent Date

Appendix F: Interview Script Illustrating RQ Relevance and Influences

Note: For each interviewee, this guide will be modified by deleting questions that are not relevant to that particular person.

RQ1: How is the development of quantitatively oriented journalism changing the perception of journalist's role?

RQ2: How is the development of quantitatively oriented journalism impacting journalism education?

RQ3: Where are journalists learning the core skills needed to be a successful data journalist?

Start with the definitions:

There are all kinds of names for this kind of journalism, in professional circles it's called data journalism, QOJ, QEJ, and others in the literature. What is your preference?

Questions

Phase 1: Introducing the Topic
(RQ 1, RQ2, RQ3)

1/ Tell me how you got into data journalism

-Can you tell me when you learned your data journalism skills? For example, what were you doing (ie: teaching, working, in school, etc…)

- Have you had formal training in data skills?
- Have you had formal training in journalism?

- What do you like about it?
- What are the challenges?

2/ How do you believe journalism is changing because of the development of these kinds of skills in journalism practice?

- Do you think it is the same core storytelling process with new tools or is it a new innovation? *
- Is it data journalism compatible with more traditional approaches to journalism?*
- Some statisticians say that statistics is about the middle (averages) and journalism is about the outliers. What are your thoughts on that?
- Is it the increasing use of data making journalism more or less humanistic in its approach to storytelling?

3/ Do you think the audience perception of journalism is different when news is presented in this way?

- Do you think the audience understands visualizations in a different way that a straight news story?
- How is the believability or reliability for the audience changed by stories supported by statistics or "data"?
- How do you teach objectivity differently around data?

Phase 2: Easy Questions which show empathy
(RQ3)

4/ Can you describe what you feel are the essential skills for any journalist working today?

- CAR skills

- Coding?
- Excel?
- Stats + Numeracy?
- Computational thinking?
- Visualization?

(RQ2)

5/ How are essential skills in journalism practise reflected in what your program is teaching?

- Is the skill set for journalists working with mostly quantitative data different?
- Have you seen a change in student expectations to learning data/data journalism skills when they start your program?
- What is the advantage to a journalism student or journalist to learning this skill set?*
- Is there an advantage to the audience to have more data journalism skills in the news media?

Phase 3: Tough questions which address more sensitive topics
(RQ1, RQ2)

Much of the literature states that to do a good job with data journalism, a basic understanding of statistical concepts is important for journalists. For example, The US Accrediting Council on Education in Journalism and Mass Communication states in its standards that graduates be able to "apply basic numerical and statistical concepts."

(RQ2)

6/ In your opinion, does your program adequately prepare students with the statistics and numeracy skills they need as journalists?

- The literature shows that there may me a gap between what journalism students are capable of when they get to J-school in regards to statistical literacy and numeracy and the perception of what they are taught- what are your thoughts on this?

- Are these skills taught as part of data journalism teaching? Is it in other courses?

(RQ2)

7/ What are some of the challenges that are presented in teaching this skill?

- Do you have technical challenges in terms of the equipment you have available to you?[#]
- Do you have buy-in challenges from the students?
- If yes, which parts of data journalism (for example mapping, spreadsheets, statistics, coding) are most difficult for the students to learn?[#&]

(RQ2)

8/ Is there a certain demographic of student that is more interested in this kind of knowledge?

- Do you see a difference in socioeconomic status and the students most interested in this?[+]
- Cook's research at the University of Georgia found that female students in particular are less likely to be interested in data journalism because if they had an aptitude for computer science, they were likely to be recruited into the computer science faculty directly. Do you find a difference between the enthusiasm from male students vs female students in the pursuit of data journalism?

Phase 4: Toning down the interview, ending on a positive note
(RQ2, RQ3)

9/ If you teach or have taught at more than one program, what are some of the differences you see in the kind of learning undertaken by students at the college, bachelor or graduate level?[+]

- What do you feel is the ideal setting for data journalsim learning? For example, should it be part of a journalism school's core curriculum or an elective? A graduate program? An online course?#

(RQ1, RQ2, RQ3)

10/ In your opinion, would you like to see more students exposed to this kind of learning?

- What do you think would attract them?
- What keeps them away?
- Do you feel students who are exposed to some data journalism learning are more or less likely to keep pursuing these skills once they've left journalism school?

Phase 5: Leaving the door open

11/ Is there anything else you would like to add at this time?

- I will send you my transcripts so that you can give me feedback that it is accurate and also to give you a chance to let me know if there's anything else you want to add to the knowledge.

- Would you like me to send you my study when it is complete?

*Questions are informed by the Diffusion of Innovations theory (Rogers, 2003).

+ Questions are informed by Knowledge Gap theory (Tichenor, Donohue, & Olien, 1970).

Questions are informed by Yarnall et al. (2009).

& Questions are informed by Dunwoody & Griffin (2013).

Appendix G: Transcription Key

This key it includes codes for participant and interviewer, (P, JL) as well as shorthand for things like pauses (short .. long…) pseudonyms and changes to ensure confidentiality (~Maria. ~ college where I work) inaudible words (inaudible?) (Mayan, 2009, p. 140).

Appendix H: Codebook

*This table was created using Microsoft Excel.

Code Name	Sources	Ref.	Previously used	Description	Theme	RQ#
Access to knowledge of sources	**3**	**9**		describes how data journalists are on more even footing with the institutions and stories they are investigating.	Control	1
Attribution to evidence	**5**	**18**	Selecting and Interpreting, journalist guiding	describing a shift between a journalist's simple attribution of the words of sources, to being able to say what the evidence is after analyzing the data. The journalist selects and interprets the data, rather than the source.	Control	1
Interactivity	**5**	**21**	Loss of control/ journalist guiding	describes when a journalistic story is built to include interactivity from the audience, guiding the audience and losing some control over the narrative	Control	1

Loss of control	**2**	**4**		when the discussing the role of the audience and the loss of control over the narrative.	Control	1
Differentiation	**2**	**4**	New approaches to journalism	when participants talk about data journalism as being different from traditional "pen and ink" journalism	Differentiation	1
Exclusive stories	**5**	**14**		describing the ability to get exclusive stories through data	Differentiation	1
Setting apart	**6**	**39**	Innovating	describing how data journalists are set apart from their peers through their actions	Differentiation	1
Belonging to a community	**4**	**6**		signifies when people talk about being part of a community, a supportive community or a differentiation of groups. It is more precise than the data journalism silos because it speaks to the feeling of being part of a different group, as opposed to a differentiation in the work.	Journalism Work	1

Collaboration/Competition	**5**	**17**	Connects with other DJs	used when discussing the collaboration with other data journalists and outside experts to develop a story or skill set. Losing competition	Journalism Work	1
Data journalism silos	**5**	**14**	Big Tent	used when a talking about all the different kinds of data journalism there are and how they are divided.	Journalism Work	1
Shifting boundaries	**5**	**14**	Fuzzy Boundaries, change in journalism work, workplace conflict	describing how the role of the journalist is changing from bloggers/ traditional journalists/ etc. and how workplace dynamics change	Journalism Work	1
Specializing	**5**	**38**		This is used when a participant talks about how there are some people who are elevated in their data journalism by specializing in one kind. It is a sub group of data journalism silos because it talks about what someone does, as	Journalism Work	1

				opposed to what work they produce.		
Examining trends vs. Anecdotal data	**6**	**29**	Wider view allowed by data, 30,000 foot view, near to far, far to near, context in data, cherry picking data, anecdotal data.	looking at how the middle/average /mean can actually illuminate the trends and the outliers. used as an example of storytelling that does not show "the whole picture" but rather relies on anecdotes to tell the story.	Story Elements	1
Making errors	**6**	**19**		participants discuss the pitfalls of data as being more prone to error than other forms of reporting	Story Elements	1
Precision in data	**6**	**27**	data more precise	participants talk about the more precise picture journalists can paint through complete (accurate, clean) data sets	Story Elements	1
Transparency and methodology	**5**	**23**		discussing how data stories should have transparency of data for the audience, and	Story Elements	1

				to the subjects of the story, and includes the discussion of methodology		
Aptitude	**5**	**50**	Capable, not for everyone	the idea that not everyone has the ability to do data work, that it takes a certain kind of person to learn the skill set	Capability	2
Gender bias	**6**	**17**		describing the gender split in the data journalism field	Capability	2
Regular practice	**3**	**5**		participants talk about the skills need to be used frequently in order to be of use	Capability	2
Classroom	**6**	**31**		code describes classroom or post-secondary learning environments	Classroom Experience	2
Hands on learning	**5**	**12**		classroom experience of hands on learning, and how that approach is different	Classroom Experience	2
Split-focus	**3**	**16**		instructors talk about how there isn't enough time to teach all the data literacy and skills needed because of content that	Classroom Experience	2

				serves other media lines		
Employment	**4**	**8**		when a participant discusses how data journalism education will give a student an edge when it comes to getting a job or employment after school.	Motivators	2
Motivation	**6**	**28**	Goals	participants discuss having goals or a specific motivation changes the way journalists learn data skills	Motivators	2
America	**5**	**17**		tracking the influence of US organizations like NICAR, IRE, etc.… on data journalism education and practice	Outside Influences	2
Awards	**2**	**3**		when a participant talks about how an award or nomination is a mark of quality journalism, in reference to data journalism.	Outside Influences	2
Bootcamp	**5**	**17**		describing a course that is offered in an intense, outside of post-	Self-Guided Learning	2

				secondary setting		
Self-taught	**6**	**20**		used when describing how data journalists get their skills, MOOCs, online, through a story, etc..	Self-Guided Learning	2
Skills driven by story	**4**	**15**		the acquisition of data journalism skills to tell a specific story	Self-Guided Learning	2
Managing data bases	**6**	**9**		skills like Access, SQL, etc., and managing data in it	Advanced Skills for Specialization	3
Programming	**6**	**43**	Developing, manipulating data	this describes journalists who have become software developers in order to tell stories	Advanced Skills for Specialization	3
Basic skills	**6**	**14**		listing basic skills needed by all journalists	Basic Data Skills for All Journalists	3
Critical thinking & problem solving	**5**	**12**		not taking data at face value and knowing when to push for answers the need for problem solving skills in manipulating data and data tools	Basic Data Skills for All Journalists	3
Data literacy	**5**	**13**	basic data skills	participants talk about minimal levels of data literacy	Basic Data Skills for All Journalists	3

				for any journalist		
Excel	**6**	**35**		describes using a spreadsheet like excel to organize or keep track of data	Basic Data Skills for All Journalists	3
Numeracy	**6**	**47**	applied math skills, math phobia	the effect of journalists lacking math skills or having math phobia.	Basic Data Skills for All Journalists	3
Programming awareness	**5**	**7**		describes when journalists don't have programming skills, but do know enough about programming to collaborate on projects.	Basic Data Skills for All Journalists	3
Storytelling	**6**	**49**		the need to be able to tell a journalistic story through data	Basic Data Skills for All Journalists	3
Accountability	**6**	**34**	Questioning	the ability to question the reliability of the data, as well as to question human sources in order to get accountability	Intermediate Skills	3
Accuracy	**5**	**16**	Accountability (split)	the accuracy and questioning of the data sets, using data to teach critical thinking	Intermediate Skills	3

Analyzing data	**6**	**42**	journalist guiding, selecting and interpreting	used when describing the role of performing data analysis, when discussing the role of the journalist to make sense of the data that is later presented to the audience.	Intermediate Skills	3
Uncovering data	**5**	**26**		discussing how data is found for stories, through FOI or open data portals	Intermediate Skills	3
Presenting	**6**	**42**	Visualizing	used when discussing the role of data visualizations in journalistic storytelling	Intermidiate skills	3
Ethics	**1**	**3**		discussing the changing ethical landscape of working with data	General Observations	N/A
Storytelling example	**6**	**8**		when a participant uses an example of a story they have done to indicate the skill they learned, the question they were considering, or in response to the question at hand.	General Observations	N/A

Unsure of effects of data on audience	**6**	**11**	Unsure of effects of data on audience	participants describe what they think the effect of data on the audience is.	General Observations	N/A

Appendix I: Evolution of RQs

RQs when the project began, September 2016:

RQ1: How is the development of quantitatively oriented journalism changing how the role of the journalist is perceived?

RQ2: How is the development of quantitatively oriented journalism impacting journalism education?

RQ3: What are the core skills needed to be a successful practitioner of quantitatively oriented journalism?

RQ4: How and where are journalists learning the core skills needed to be a successful data journalist?

RQs for the Literature Review:

RQ1: How is more access to open data, Big Data and the tools to visualize this information changing journalism practice?

RQ2: How is data changing the skills a journalist needs?

RQ3: How should those skills be learned?

RQs during Methodology Design, March 2017:

RQ1: How is the development of quantitatively oriented journalism changing the expectations of the roles and responsibilities of journalists?

RQ2: How is the development of quantitatively oriented journalism impacting journalism education?

RQ3: Where are journalists learning the core skills needed to be a successful QOJ journalist?

Final RQs for Analysis:

RQ1: How is the development of quantitatively oriented journalism changing journalism practice?

RQ2: How is the development of quantitatively oriented journalism impacting journalism education?

RQ3: What are the core skills needed to be a successful practitioner of quantitatively oriented journalism and where are they learned?