FROM JOURNALIST TO CODER:
THE RISE OF JOURNALIST-PROGRAMMERS

A Professional Project
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
YANG SUN
David Herzog, Project Supervisor
DECEMBER 2017
ACKNOWLEDGEMENTS

I would like to gratefully thank my committee of David Herzog, Beverly Horvit and Barbara Cochran, who provided generous guidance and support for my project and internship in Washington D.C.

My sincere gratitude also goes to everyone at the Investigative Reporting Workshop, where I had a wonderful summer and fall. I would not have learned so much from this internship without their help, trust and encouragement.

I also want to thank my awesome parents who support my every choice unconditionally.

At last, I would like to thank the Missouri School of Journalism. It changed my life and it changed how I think of journalism.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..............................................................................................................ii

Chapter

1. INTRODUCTION ................................................................................................................1
2. WEEKLY REPORTS ..............................................................................................................4
3. EVALUATION ....................................................................................................................25
4. PHYSICAL EVIDENCE ......................................................................................................32
5. ANALYSIS .......................................................................................................................57

APPENDIX ...........................................................................................................................69
CHAPTER ONE: INTRODUCTION

In my second year as a journalist in Beijing for Xinhua News Agency, the biggest state-owned news organization in the country, I traveled to a small town in southeast China to report on the local government's crackdown on crime. Like many other reporting trips I had taken, it was arranged by the central government in Beijing.

I talked to a victim in his late 50s who had just lost all his savings to telecommunication fraud. Because it was a virtual crime that was conducted by chains of criminals, it was highly possible that he wouldn't ever determine the final destination of his stolen funds. He felt desperate. I also interviewed a felon who was in his early 20s. He said he joined the fraud network because he needed money to pay his father's medical bills. He ended up in prison and feeling a sense of utter despair.

That trip was truly an eye opener for me. For the first time, I felt that as a journalist it was impossible for me to do real reporting if I continued living my seemingly comfortable life in Beijing. Most importantly, I could not continue to view myself as an independent journalist if my every interview was arranged by a third party.

After much deliberation and discussion with my family and friends, it took me two years to finally decide to leave my job and move to Columbia, Missouri, in the United States with a hope to learn real journalism.

The Missouri School of Journalism definitely fulfilled my hope.

It was during Professor Herzog's Computer-Assisted Reporting class that I learned that I wouldn't put myself in trouble if I requested documents from the
government. I also learned that data analysis is an integral, powerful component of a well-sourced story.

The convergence reporting class offered me a chance to apply all my journalistic skills. I spent many sleepless nights in the Futures Lab. The experience was intense but rewarding. When I was in China, I had dreamt of reporting on the stories of ordinary people who had troubles to deal with every day. This dream finally came true. I profiled a Vietnam War veteran who has been fighting with PTSD for years. I interviewed a woman suffering with a chronic disease without health insurance. I visited a single mother who was struggling to find an affordable childcare service.

In addition to discovering the human-interest element of journalism, I was impressed by the role of technology in the business. Getting into the community of the National Institute for Computer-Assisted Reporting (NICAR) was my best Mizzou experience. My two NICAR conferences awoke my underlying nerdy genes from a deep slumber. I became fascinated by programming because I found that this unconventional skill empowers journalists.

I met journalists at NICAR who spent many hours every day writing both machine-readable and human-readable languages. As a result, I became fascinated by their stories of learning to program as a working journalist; the curiosity their stories spurred within me is what eventually gave rise to my graduate project.

I am glad that nine journalist-programmers took my invitation to be interviewed and were willing to share their stories. Each of them had a unique learning experience, but there is one thing they had in common: they all embody the spirit of continuous
learning. From each of these individuals, I gained what was perhaps my most valuable insight I learned at Mizzou: never stop learning and always be curious.
CHAPTER TWO: ACTIVITY LOG

Week 1.

I started to intern with the Investigative Reporting Workshop at American University this Monday. The newsroom is not a new environment for me because I already had an eight-week internship there this summer. I felt grateful for the summer experience because it helped me to cut to the chase and get into the business directly. However, I did not have a very productive week because I was very jet-lagged. I arrived in Washington D.C from Beijing on Saturday, but I guess my internal clock still ran Beijing time and I had to spend the first two days getting used to the eastern time.

Before I left the Workshop in summer, I was informed by Lynne Perri, the managing editor, that we are expecting a new data editor, Jennifer Lafleur, this fall. She will start next Monday, and we will talk about the Accountability Project, which I have been working on since summer. Lynne also hoped that I would keep a closer eye on post-Charlottesville protests and rallies in Washington D.C. for multimedia stories. There are already four rallies scheduled in September and October in the National Mall, from white-nationalists to anti-Trump groups.

I pitched a story idea on how journalists are covering activities from hate groups. I came up with this pitch because reporting on hate crimes and hate groups has become a new normal in newsrooms. Covering this topic precisely and with balance is challenging for reporters. Lynne really likes this idea and suggested that I start by talking to media ethics professors and Southern Poverty Law Center, a legal organization that has been tracking hate crimes for years.
The Washington program had its first seminar on Friday, which I enjoyed and found challenging. Alex Mallin, a D.C. program alumnus, currently a digital reporter with ABC news covering Trump Administration, revealed a lot of inside stories around the controversies of the President, which I barely knew from media publications. His presentation made me realize that verification and fact-checking are unprecedentedly important at a time when the top leader of the country keeps labeling the media "fake news." His fast-paced narrative from one politician to another and one story to another is hard for me to follow sometimes, reminding me I should be more immersed in the news from Washington.

**Week 2.**

I made some progress on my last week's story pitch by interviewing Rachel Glickhouse, a ProPublica reporter and project manager of Documenting Hate, one of the few media efforts to track hate crimes nationwide. We were on the phone for 20 minutes, and she introduced this collaborative project with many local media outlets and technology companies. The core mission of it is to fill the gap between federal statistics on hate crimes and the reality. Glickhouse stressed the importance of this project, particularly when such incidents spiked after the election and Charlottesville. ProPublica hopes that this hate crimes database can make the life of reporters covering hate incidents in their local communities much easier.

I am also working on scheduling more interviews next week. The communication associate from SPLC already contacted me and will connect me with one of their Hatewatch Project researchers. I also emailed Christopher Benson, a journalism professor at the University of Illinois whose research emphasis is on media and race. I talked to
Lynne that I would try to finish this story and publish it before September 16, when both a pro-Trump and an anti-Trump rallies are going to occur in National Mall. I am going to the Mall to cover that.

As for the Accountability Project that I have been working on since summer, not much happened. Chuck Lewis, the founder of the Workshop, Josh Benson, data reporter and I were supposed to meet with the new data editor Jennifer Lafleur; but since this was her first week, she didn’t have time to go over the project with us. Jennifer didn’t have much time to talk to me either, which made me a little bit sad. I hope there are data-driven projects that I can work with her since this is one of the reasons I chose to intern with IRW.

I already have sent out interview inquiries to four data reporters for my research analysis, but I only had one response. I am frustrated with such an outcome, and I hesitate whether I should send out follow-up emails to them next week. Patrick Galvin, a visual journalist at the Boston Globe, replied to me and we are going to talk after Labor Day. Galvin was a guest lecturer in my infographic class. He talked about how his programming knowledge helped him seize a job opportunity at one of the most reputed newspapers in the country. He didn't go into much details on his career story in that class, but we will talk more about it next week.

The D.C. Program class made a trip to the Newseum which is kind of a routine for Mizzou students in D.C., but this trip is different as its President stepped down a few days ago due to Newseum's financial challenges. I disagree with Politico's Jack Shafer's idea that it is not worthy of Newseum spending $500 million to have a building facing the Capitol Hill and educate the value of the First Amendment. I believe that such lecture to
the public is important in Trump's era. Not to mention that, as Gene Policinski, Newseum’s CFO, said, it is not just about journalism freedom but all kinds of freedom. Then how can you put a price tag on freedom? Good luck to Newseum!

**Week 3.**

I had mixed feelings about my three-day week at the Workshop. I was pleased with the progress I’d made on my pitched story, but found myself frustrated with the lack of progress on the Accountability Project.

I interviewed three sources for the story on guidelines for covering hate. Before Labor Day, I talked to a reporter from ProPublica who provided the perspectives of national media practitioners. I also needed community media practitioners' thoughts. Therefore, I reached out to The Daily Progress, the only daily newspaper in the vicinity of Charlottesville. I talked to a reporter and an assistant editor there who had been following the event since its beginning. Aside from media practitioners, I also interviewed Christopher Benson, a lawyer and a professor of African-American studies and Journalism at the University of Illinois-Urbana Champaign. After my interview with an expert from the Southern Poverty Law Center next Monday, I will be ready to write the story.

I was a bit frustrated as the Accountability Project did not go well this week. I joined in a meeting with Lynne Perri, Chuck Lewis, and Josh Benson. Jennifer LaFleur, the Workshop's new data editor, attended the meeting to offer a fresh perspective. We agreed in the summer that some improvements on the site's search functionality and interface design could better the project overall, yet Jennifer posed a question on the accessibility of the website which made us re-think our vision of this project.
She raised the concern that she did not see originality from it, because the website, as a hub of hundreds of online database's URLs, only returns the users to the original websites for search and analysis. Though it sounded harsh, Jennifer had posited a valid argument. As it currently exists, our project has reached an impasse. At this point, there is relatively little I can do before the editors find a solution.

I had scheduled an interview with Patrick Garvin, a visual journalist for The Boston Globe. However, due to the high volume of news stories which broke this week, he could not allot an appropriate amount of time for an interview. As such, we have postponed the interview until next week. In addition to Patrick Garvin, I will be interviewing Chris Amico, PBS Frontline's interactive editor next week. I also sent out four emails, anticipating that more interviews will be scheduled over the coming weeks. My goal is to conduct six interviews by the end of September.

Our class reconvened on Friday to meet with John LaBombard, communications director, and Sarah Feldman, press secretary of Senate Claire McCaskill. I was impressed by how accessible high-ranking officials in U.S. are to the public. Being a political reporter for four years in Beijing, I had never stepped into any central-government official's office. John gave us some great advice on what to do if we are covering news on Capitol Hill or need comments from politicians. I am appreciative of the dynamic interaction between politicians and media. But for a government and its mouthpiece, the relationship mainly is about ordering and obeying. That just sounds familiar to me.

**Week 4.**

Despite the fact that I changed my plans a little bit due to hurricane Irma, I had a productive week.
I had agreed with a communications associate of Southern Poverty Law Center that I would be connected to an expert for an interview this Monday. However, their Alabama offices were closed this week because of the hurricane. Unfortunately, I wasn't able to conduct the interview which is the most important aspect of my "hate beat" story. Despite that, I transcribed other interviews and outlined the story. I read all of the Hatewatch newsletters and Intelligence Reports of SPLC and familiarized myself with the data they collected on hate crimes and hate groups after last year's election. I tried to write a few paragraphs despite my difficulties, and quoted the reports. This led me to the realization that I still need stronger and more interesting quotes from a human source. Lynne Perri, the managing editor of the Workshop, agreed that my story writing should wait until I talk to someone from SPLC since they have the most expertise in tracking hate.

Another job I did at the Workshop this week was preparing for my next story. Our data editor Jennifer LaFleur mentioned that many media used drones to acquire footage of the aftermath of Harvey and Irma and we could write something about it. I didn't have much time to work on it this week, except for watching some of the footage. My plan next week for this story is to study the FAA's regulation on drones and find a good story angle. I may call Professor Bill Allen who taught the Missouri Drone Journalism Program for his advice.

I also made progress on my research project this week. I interviewed Patrick Garvin, a visual journalist for the Boston Globe, and Chris Amico, an interactive editor for PBS Frontline. Garvin started to learn how to code because he knew this skill could help him secure his dream job. Amico moved to programming because he had to make
his way to survive the economic recession. The two had different reasons for transitions but had the same takeaway: Learning how to learn is essential. Jennifer suggested that I talk to some female journalist-programmers, as at that time I didn't have many females on my contact list and their alternate perspective might lend to the overall process of learning.

The D.C. class visited Politico's office this Friday. Angela Keane, Politico Pro's Deputy Technology Editor, a Mizzou alum, gave some great suggestions on how to pave a career path. A key point that she emphasized is that journalists should grow expertise. Her career benefited from her specialty on transportation industry and policies. I spent the whole day thinking about what my expertise was and how I could find my selling point. Answers to these questions are imperative because the job-hunting season is ahead of me.

**Week 5.**

I had a quite slow-paced week at the Workshop as I spent several days waiting for my last interviewee to call me. The Southern Poverty Law Center was the most valuable content for my hate beat story, but the organization was very protective to its staff. The way their communications team works is that I can only get connected with their employees through a communication associate and wait for an expert to call me. There was not much I can do until I received the phone call. It took me two weeks to finally talk to someone there. Such communication mechanism frustrated me because I can't imagine how desperate I would be if I had a hard deadline for the story.

I interviewed Heidi Beirich, the director of SPLC’s Intelligence Project. It was worth waiting because Beirich is for sure the most prominent figure in term of monitoring hate crimes and white supremacy in the country. The talk with her reminded me that it
was not only media that should hold the account for failing to properly report hate, but the whole country, which is reluctant to admit white supremacy as a societal problem, should be blamed.

Beirich brought a lot of depth to this story but also challenged my ability to clarify such a complex topic. I spent rest of the week drafting the story and had the first draft before the end of the week. I was not satisfied with this version, but Lynne Perri, the managing editor and I will sit together and edit it on Monday.

As for my research project, I interviewed Greg Linch and Ben Welsh this week. Greg made a complete transition since he studied print journalism and worked as a reporter, now as an back-end developer at McClatchy. Ben is a data editor at LA Times, who applies coding to his day-to-day job and teaches many journalists how to do the same. He is a well-known figure in the NICAR community because he explored coding several years before the industry realized that this was a necessary skill for journalists.

The D.C. class did not have a seminar this week.

**Week 6.**

I spent most of my week at the Workshop editing the hate beat story. Lynne Perri, the managing editor, is a highly detail-oriented editor. So the editing took a longer period than I expected. I worked closely with another editor, Margot Susca, in the summer, whose editing style was different from Lynne's. She was a quick thinker and straightforward. She would polish or rewrite a story all by herself and inform me her reasons afterwards. It normally took her a few hours to finish editing a long-form story. Lynne, on the contrary, would go over every quotation and sentence with me to make sure that they make sense, a process that takes a lot longer. Unfortunately, she had a
fully-scheduled week and we weren’t able to finish everything. I am now working on the fourth draft and am glad to see the improvements.

I did not have an interview lined up for my research project. The fact that so few people responded to my interview inquiries frustrated me. Since I didn’t interview, I managed to transcribe the four interviews I had conducted and had some reflections. I had a list of questions for each interview which I could follow. I realized in transcribing them that I stuck so much to the list that I sometimes forgot to ask follow-up questions to get more details. This reminds me of the conversation with Chuck Todd after the D.C. class watched Meet the Press at the NBC studio Sunday morning. In response to the question on the evolvement of his interview style, Chuck said he now focuses more on the conversation flow instead of cornering the interviewee for an answer, and looks forward to what’s being said, but not who said what. His words helped me realize that a great interview should be a natural and engaging conversation, not a list to score point. I will pay closer attention to improving my interview skills, and also next week my biggest priority is to schedule interviews.

The D.C. class had a seminar this Friday with Marina Walker Guevara, deputy director of the International Consortium of Investigative Journalists. Marina and her team coordinated more than 370 reporters in 76 countries to work on the leak of the Panama Papers. It was easier to appreciate the great efforts these journalists made to shake governments and businesses across the world. But I did not know the inside stories of how courageous and committed they were before talking to Marina. I was impressed by the story of an Icelandic reporter, who had to quit his reporting job to investigate secretly. He spent a year staying at home to figure out the connections between Icelandic leaders
and tax havens. In countries with small communities like Iceland, family and friends around him started to be concerned that he was a psychopath. To make matters worse, being unemployed for a year, he couldn't afford Christmas gifts. Eventually, the hardship and sacrifice were paid off by the Prime Minister of Iceland’s resignation.

**Week 7.**

At the beginning of this semester when I pitched the story idea on how journalists deal with the rise of hate incidents and crimes, particularly after the deadly riot in Charlottesville, I didn't expect that it would take me six weeks to finish. It wasn't until the 7th week of D.C. Program that I could confidently say that this story was nearly complete. I worked closely with Lynne Perri, the Workshop's managing editor, for two weeks to polish this story. Looking through six drafts I have worked on, I am glad to see my improvement and progress. This is the first time that I have written a long-form story on complex social issues independently in my second language, and I must say I am happy with how it turned out. Though Lynne hasn't had a fixed run date for this story yet, I will work with Workshop's graphics editor next week to add some visual elements to the story.

I interviewed Serdar Tungoren this week for my research project. As one of the earliest journalists who explored the world of web development, Serdar started his journalism career as a print reporter and now is a technical leader at the Associated Press. When he began to learn to program in 2005, he even enrolled in a programming class at a community college. After reflecting on the last week's interviews, I kept a copy of Serdar's bio with me for the interview instead of a list of pre-determined inquiries, so that I could spontaneously generate questions. The interview flowed nicely as we engaged
each other in a two-way conversation, as opposed to about a Q&A session. I was happy with this as it felt more natural than having approached the situation with ready-made materials. Serdar even emailed me afterward which said he really enjoyed our talk.

The D.C. class visited NPR this Friday, and we had a discussion on Diversity in Journalism with Keith Woods, NPR’s Vice President for Newsroom Training and Diversity. Before this trip, I only related journalism diversity to the racial and gender inclusiveness in the newsroom. Yet Keith pointed out a broadened concept which includes diversities on ideology, socio-economic status, and geology. Though the statistics of several studies demonstrate that journalism is pacing toward a more inclusive workplace, I learned from working on the hate beat story that the journalists tend to pattern minority attackers as radicalized terrorists and fail to treat hate crimes from white perpetrators as institutional ones. This demonstrated that journalists still maintain a narrow and singular ideology toward complex social issues, which is a sign of lack of diversity.

**Week 8.**

I was not as productive as I wanted to be this week since I had the flu. I stayed at home on Monday so that I wouldn't spread the germs in the office. During rest of the week, I worked with Lynne Perri to pitch my hate beat story to Workshop's partner media for co-publishing. We reached out to Columbia Journalism Review and Poynter, and only the latter one expressed the interest in running it. Observing how Lynne negotiated with other outlets was interesting because she showed me the intricacy of finding the fine line between giving out enough information to intrigue their curiosity and not feeding them too much content of the story.
I am wrapping up my current project and in need of more assignments. I talked with Lynne about my plan on writing a blog post for the Workshop about the Kalb Report at the National Press Club. The event next Monday will discuss the new administration's threats to the press and the relationship between Trump and media with the executive editors of The New York Times and The Washington Post. I also managed to have talks with Jennifer LaFleur, the Workshop's new data editor, for possible projects that I can work on. She is still in the process of planning some data projects that the Workshop can partner with the Associated Press and The Washington Post. But she didn't recommend me jump into any of them before they carrying out a solid story idea. Instead, she suggested me think of some quick turnover data-related stories that I can work on my own. I like this idea and will go over my data inventory next week to look for pitches.

I conducted two interviews for my research project, and I liked the contrast between two interviewees' working environments. I talked to Darla Cameron, a front-end developer and graphic designer at The Washington Post, and David Montgomery, a former political reporter at Pioneer Press in Minnesota. Working for one of the most prominent media organizations in the country, Darla said the Post hired many coding talents and regularly held sharing meetings on the latest trends in journalism-programming. Whereas, in David's newsroom, he was a "lonely coder" that he encountered many eyebrow-raising moments when he pitched code-heavy projects. David quit his job at the Pioneer a few weeks ago and is now in the process of being hired. He admitted that his data-and-coding literacy definitely makes him more competitive in the marketplace. His remark resonated my conversation with Jennifer when she said that her years of hiring experience showed that journalists who can treat
data critically and combine that capability with story-telling skill are more valuable and harder to find than news developers. I have scheduled the eighth interview for next week. While I am planning to arrange a few more interviews, I will start to work on the first draft.

This week's seminar was at the Weekly Standard, a conservative magazine sponsored by Rupert Murdoch originally. We spoke with Fred Barnes, one of its founding editors. I am not a big fan of conservative or right-wing publications, but I have to admit that I respect Fred as a veteran journalist. He is honest and straightforward with his political opinions and obviously very knowledgeable about Washington. Though Fred said they put a lot of efforts in digitalizing the magazine and attracting younger readers, after browsing their website and periodicals I felt that they could try harder. One thing I noticed during our short stay in the Standard's office that there was rarely minority among its staff, which is drastically different from NPR's racially diverse newsroom.

**Week 9.**

Overall, I had a productive week. The Workshop and Poynter co-published my [explanatory story](#) on how to cover hate this week. Though both Lynne and I had felt satisfied with the story before we reached out to Poynter, Anne Glover, its senior editor, suggested that I reconstruct the story a little bit to make the story more tailored for her organization’s audience. Prior to the publication on Tuesday, I worked with Anne to have some final touches on the story. This was the official end of my first independent project at the Workshop, which taught me how to report complex social issues, fact check and write better. During the rest of the week, I wrote and published a [blog post](#) on an event I attended at the National Press Club.
In addition, I did some research for a possible project. I talked with John Sullivan, an investigative reporter at The Washington Post and a senior editor at the Workshop. John told me that he has been thinking about a project on investigating the EB-5 visa program for a few weeks. This is still an idea at its novice level, but John recommended me do some research or even FOIA some government documents if I am interested.

I interviewed Sisi Wei, a data reporter/news developer at ProPublica for my research project. Sisi humbly considered herself a "lucky early bird" who ventured into the career trajectory as a journalist-programmer. I also spent a few hours this week to transcribe the interviews.

I have conducted eight rounds of interviews so far, and I am glad that I have a diverse group of interviewees, which includes both male and female, front-end designers and back-end developers, journalism from national and regional news organizations. Some journalists I interviewed also hold part-time teaching positions in journalism institutions, teaching students how to incorporate programming skills into their future career. But I want to have insights from some full-time journalism faculties, and I am still searching for such candidates.

Our D.C. class reconvened Monday evening at the National Press Club, watching a panel discussion featuring the executive editors of The New York Times and The Washington Post on the topic of journalism in Trump era. I think such a topic has been regularly discussed in D.C. since the rise of Trump as a presidential candidate. But this was the first time that journalism community hear comments from the top editors at the country's two leading organizations, which are also considered as President Trump's two most significant media punching bags. I think my biggest takeaway from the event was
Week 10.

I started this week by attending the International Women's Media Foundation’s 2017 Courage in Journalism Award ceremony in the Newseum. I had a blast that evening listening to the courageous and inspiring stories of several female journalists who have covered politics in the capitals of some of the biggest developed countries and brutality of wars in some of the most impoverished countries on the planet. I would like to give special thanks to my advisor Barbara who helped me to nab a seat in the last few hours before the event; thanks to her generosity, I got the chance to interact with some of the greatest women in journalism. I also wrote a [blog post](#) about the IWMF event for the Workshop.

During the rest of the week at the Workshop, I worked on the Accountability Project, which started with a massive Excel spreadsheet listing over 400 websites/database on money in politics. My job in the summer was to enter each database into the project’s data entry interface. The Workshop’s data editor Jennifer LaFleur wanted me to find out whether there were open-source tools on this topic at the state level that we didn’t include in our project.

The reason for the additional scrutiny was that she thought the Accountability Project was too ambitious to include both state and federal level contents. The California Civic Data Coalition that Jennifer was involved has spent several years cleaning the data on politics only in California. She thought that people from other states might have made similar attempts that we could learn from. Hoping to glean guidance from prior attempts,
I reached out to NICAR email list and several data journalism slack groups. I appreciated the assistance from journalists at the NICAR community who gave me insight into what resources were available, such as campaign data scrapers in Virginia, Florida and Georgia. Notably, I am experiencing frustration with our project because I feel that it still lacks an achievable goal.

The seminar this week was the best so far in my opinion, because our guest speaker Donna Leinwand Leger, the managing editor of USA Today, gave us so many great tips on ways to excel as a journalist. Some of the advice was so readily applicable that I could implement it directly to my next story. For example, Donna recommended against long and excessive direct quotes when writing a story. On reviewing some of the articles I had previously written, I found that their flow could be improved by incorporating more concise phrasing and instead of using block quotes. Donna also shared with us her adventures in Iraq and Haiti as a foreign correspondent, which I found to be engaging and thought provoking.

I also spent several hours this week transcribing the interviews for my research component. I reached out to Matt Wait, a professor at the University of Nebraska-Lincoln, but to date he has not replied.

**Week 11.**

I can’t believe how fast time flies that this is almost the end of my last academic semester. We are now in the most colorful and beautiful season in Washington D.C and also nearly the end of our D.C. program.

I spent this week at the Workshop blogging and working on the Accountability Project. Our director Chuck Lewis was invited to a project launch at the Newseum by
Laurent Richard, a French investigative reporter who visited the Workshop last year. But he had a busy schedule and couldn't make it. Therefore, I got the chance to attend the event on behalf of the Workshop and I wrote a blog post about it. The project Forbidden Stories aims to help finish an uncompleted story when a reporter is silenced, and was founded by Reporters without Borders (RSF).

Having gone to a journalism school in a country which is considered an authoritarian one by western standards, I was educated that RSF was a politically biased and anti-China group who puts groundless accusations against human rights and press freedom in China when I was a undergraduate.

Yet I was glad that I managed to put aside my pre-set ideas and participate in the event with an independent mind. I strongly agree with the core concept of this project that when the topics that investigative journalists are dealing with have become more complex, it is more than ever important for journalists across countries to act like a team.

As for the Accountability Project, I am still searching for more open-source tools that can help journalists and scholars understand money in politics better. I think it is a good idea to go over the GitHub accounts of newsrooms from different states and experiment with tools developed by news organizations. My data editor, Jennifer LaFleur, thought this was a good idea and considered it a long-term assignment. I think I will be working on it in the following weeks.

I didn't hear back from Matt Waite, a possible interviewee for my research component, until I resent an email to him. Somehow my first email fell through the cracks and he didn't receive it. We scheduled the interview for next Monday. I spent a couple of hours during the weekdays and most of the weekend analyzing the rest of
interview scripts and planned out the final report. I will be able to finish the analysis and writing next week.

During this week's seminar, we spoke with Kurt Wimmer, Steve Weiswasser and Hannah Lepow, three attorneys who have profound experience in media law. To be honest, I didn't enroll in any classes on communication law at Mizzou. My scarce knowledge on media law in the United States was from my journalism history class in China. So this session was definitely informative to me, especially when the three were able to explain complicated legal concepts and cases in laymen’s terms.

However, the best part of this seminar was our in-depth and candid discussion on sexual misconduct in the workplace given the sexual harassment allegations on Michael Oreskes, NPR's lead of news operation, which came out this week. Inappropriate misconduct in the workplace toward both men and women, particularly women, is not only an institutional problem in the U.S. It is the same problem in China, which was one reason I quit my job in Beijing because I found that my newsroom culture was hostile to young, single female employees.

It is an open secret in my home country that females are treated unfairly and unwelcome sexual advances are rampant in the workplace. Unfortunately, Chinese, as well as people from other Asian countries, are more introverted and reserved, and therefore fewer victims are willing to disclose their encounters. However, I am glad to see that females in this country are more vocal, as can be seen from the case of Harvey Weinstein and the #Metoo campaign on Twitter. As the attorneys said, we can see that the country is undergoing a gender reform considering that more victims feel able to speak out.
Week 12.

This has been a stressful week for me at the Workshop. On Monday morning, I met Emma Schwartz, the PBS Frontline's investigative reporter, at the Frontline's studio, which locates in the same building of the Workshop. The Frontline team had five-year files of campaign donation to Ed Emmett, a Republican politician in Houston, Texas. They needed to find out all the campaign contributors from real estate/ construction business who contributed more than $1000 to him. Then these records should be entered into an Excel spreadsheet for analysis. Lynne sent me to help Emma's team, and I was willing to work with a new group.

The files were scanned and stored in PDF format. I tried multiple tools to convert them into machine-readable text files but failed. This meant that I had to enter everything manually. To make the matter worse, the documents didn't contain information on contributors' job titles and names of their employers. As a result, I had to Google each contributor's occupation and affiliation according to their registered addresses. Emma didn't know how long it would take me to finish the job but wished that it could be done by the end of the week. Due to the unfriendly format and missing information, the workload turned out more substantial than I had anticipated. I spent the early days of the week doing nothing but scrutinizing the files, only finding out that this assignment would take me a whole week.

Everyday from 10 am to 6 pm at the Workshop, I helped the PBS Frontline build the database. In the evening, I rushed home to work on my project until late night. Though being deprived of sleep, I finally managed to finish both tasks.
The PBS assignment truncated the hours that I could have worked on my project. But I still reached out to Matt Waite, the last interviewee for my project. Also, I finished writing the project analysis and a few other chapters. I will keep completing my project and preparing for my defense meeting.

This week in the seminar, we visited the Washington Post's investigative unit and spoke with Jeff Leen, the managing editor, David Fallis, the deputy editor and Steven Rich, the data editor. Steven, for sure, as one of the proudest Washington Program alum, gave the most relevant and applicable advice. He said that the key to being a good journalist is to find a goal and be excel at it. Besides, he suggested not get lost in the data analysis, instead, let it serve the purpose of storytelling. The team we visited on Friday perhaps gathers the most distinguished journalists in the country. I may not be a part of the team eventually. But I would be thrilled if I could be half accomplished as a journalist as anyone from the unite.

**Week 13.**

I can’t believe this is my last weekly report. It was a very quiet week at the Workshop as half of the staff were gone for business trips. I spent Monday wrapping up building the database for PBS Frontline. During the rest of the week, I researched open-source tools for money in politics. After a few rounds of search on Github, I found that most regional newspapers built online scrapers, in order to fetch data on election results and campaign finance. I only found a few interactive databases that focus on state elections. These projects mainly were built and maintained by non-government political organizations or civic hacker groups.
For my research project this week, I edited the first draft of my project and sent the second to my chair.

We went to the Meet the Press Film Festival on Monday as this week's seminar. We watched two short documentaries: Fight for the First and Gavin Grimm VS. The first film addresses the freedom of the press by profiling several students at the Missouri School of Journalism. The documentary was shot in Missouri when a series of campus protests took place in the fall of 2015. At that time, I just started school in Mizzou and had limited knowledge of the First Amendment. As a result, I witnessed the demonstrations as an outsider. Now I am almost finishing my education at Mizzou, and I know better the rights and responsibilities of a journalist. The comparison is tangible evidence on how the Missouri School of Journalism transformed the meaning of journalism to me.
My internship in Washington D.C. lasted longer than the duration that the program required. I spent eight weeks in the summer interning at the Investigative Reporting Workshop, and continued my position at the Workshop after my summer break. Lynne Perri, the Workshop's managing editor and my on-site supervisor, once joked that she already saw me as a member of their full-time staff due to my near-constant presence. I am really grateful for all the knowledge that I gained during this extended internship; I feel that this was an integral part of my growth process and the lessons I learned will continue to serve me going forward.

I was mainly hired to work on the Workshop's Accountability Project, a web-based tool that provides journalists, researchers and the general public with extraordinary access to data, documents and other forms of information about those in power. The concept originated from months-long deliberation in early 2015 between Chuck Lewis and David Donald, the executive editor and the former data editor of the Workshop, respectively. Sadly, David passed away in December 2016 after a year-long battle with cancer. To commemorate his dedication and contribution to journalism, particularly data journalism, the Workshop funded a data journalism fellowship in his name, and I became its first recipient.

I was very nervous about this project at the beginning. To be completely honest, I have very limited knowledge of the complex inner workings of elections in the United States. Although I was previously a political journalist in China, a bipartisan and democratic political system is foreign to me. I was grateful that Chuck and Josh didn't
push me to finish my job right away and let me spend weeks familiarizing myself with the filing regulations, lobbying process, donations and all the nuances of American politics. I also appreciated that they didn't make me feel embarrassed once I asked stupid questions.

I spent most of my summer weeks working on the project's data-entry interface. My responsibility was to research each database listed by the project proposal. I needed to find all the information about a database's content, its landing page URL, publisher, time span and data format, etc. Then I had to enter all the information.

There were 232 databases. Therefore, I repeated this process 232 times. Though to some, this may be perceived as tedious and trivial, that's the real life of a journalist. It is because of the repetitive and unglamorous nature of this job that I learned how to be detail-oriented and meticulous.

Unfortunately, the project didn't proceed significantly in the fall. The main reason is that Jennifer LaFleur, the Workshop's new data editor, provided a fresh insight into the project. She raised the concern over the project's usability, a big flaw that we all neglected before she came. While the Workshop's editors continued to think about whether they should hire new developers to make the project more engaging and meaningful, there was not much I could do.

I spent the rest of the summer working on an explanatory piece about the District government's open data initiative. The idea started with my assignment to update a Workshop project from 2010, which reviewed 54 DC government agencies' FOIA response. But one week later, my editor, Margot Susca, wanted me to shift the focus to DC's open government campaign, which was launched earlier this year.
I was glad that I could apply the skills that I learned from Mizzou to this project. Thanks to the enterprising spirit I gained from the reporting classes, I reached out vigorously to attorneys, civic hackers and open government activists, anyone who could help me to gain more insight about the topic of open data. I used the skills I learned from David's Computer-Assisted Reporting class to analyze DC government's FOIA reports in ten years and five-year budget on technologies.

I also learned the importance of being persistent. For two weeks, I sent a daily email and left two voice messages for the communication director at the Office of the Chief Technology Officer, all in the hopes of pinning him down to verify the District's spending on this initiative. Finally, the director called back and we got answers to our questions and published the story one day before I wrapped up my summer internship.

I came back to IRW after the summer break only finding that there was no project yet for me to work on. I was so eager to get as much experience as I could that I pitched a story to Lynne the next day.

This was just two weeks after the turmoil in Charlottesville, Virginia. The press and the public were still reacting to this tragedy. Therefore, I wanted to write a story on how the media was handling the contradictory aspects of this story. I specifically elected to focus on the irony of this story. On one hand, the rising tide of hatred within certain groups in the country deserved more reporting, so that its dynamic evolution could be more clearly understood; on the other, most newsrooms in the country are struggling with a tight budget and the lack of staff, making this level of investigation a herculean task.
The topics of race and hate are consistently difficult to tackle, even for a veteran journalist, because of its intricacy and complexity. That fact made the experience of writing this story both overwhelming and rewarding.

Lynne definitely gave me guidance every step along the way. She gave me suggestions on where to start with a broad topic such as this. She caught and corrected each detail that could make the story unbalanced and biased. Also, she showed me how to make the story more relevant when it was already one month after the Charlottesville riot. The story took much longer to produce than I had anticipated. During this process, I learned how to research on a complicated topic; how to balance the sources; and how to be meticulous in my writing. I wrote more than five drafts. The word count was trimmed down from 3000 to 1300 when it was eventually published. However, what I learned far surpasses the measurements of drafts and word counts.

Besides the long-form story, I attended several journalism events and wrote blog posts for the Workshop's website. I was honest with Lynne at the very beginning of my internship about my biggest struggle: I had a difficult time writing in a foreign language and I regretted that I dodged the chances I could have practiced before. Being aware of my predicament, Lynne found as many opportunities as possible for me to produce clips and practice my writing in English. I have written four blog posts so far, including one during the summer. Thanks to Lynne, they all felt more like a story assignment and my confidence increased dramatically with each completed piece. Most importantly, I now feel more confident in writing in English. I won't say I am perfectly proficient as a writer in the English language, but I am certainly better than I was previously.
My only regret is that I didn't work on any data-driven reporting aside from entering data. For this I can blame nothing but the timing; at the point that I arrived at the Workshop, there was not a project of this type available. Luckily, I will be able to remain there until the end of this year; this leaves me a few more weeks to work on a few small data projects. I already have a story idea and am completing research on it. I will pitch it to Jennifer once I feel it is ready. I am sure I will keep learning until my last day at the Workshop.

Though I have enjoyed my stay at the Workshop a lot, I need to say goodbye to everyone there by the end of the year. As a student visa holder, I can apply for a year-long Optional Practical Training. Therefore, I plan to find a journalism internship to continue honing my reporting skills while my paperwork for the permanent residency in Canada is in progress. For my long-term goal, I hope to keep practicing journalism in Canada. Although the media industry there is not as dynamic as it is in the U.S., there are many topics, such as government transparency, immigration, environmental protection and inclusiveness of first nations, that are worthy of reporting. I believe that my bilingual advantage and journalism skills that I learned in the U.S. will help advance my journalism career in Canada.

**On-Site Supervisor Evaluation**

Yang Sun, our first David Donald Fellow in data journalism, has been an outstanding Workshop staffer these last six months.

She is smart, persistent, conscientious and diligent in all that she has done on multiple topics and projects. We're looking forward to her return for several weeks after
her graduation from Missouri because we hope to have her start researching a joint-project that we're undertaking with the AP.

Yang is currently researching open-source tools and sites for data in money and politics, which we'll use as part of our long-range accountability studies project.

She also created a database for the FRONTLINE team based at the Workshop for a future program on hurricane relief and recovery programs (making sense out of data related to wealthy Houston residents and their donations to political candidates.

In addition, she has covered several stories and issues for the Workshop that we published as blog posts or stories within the "ilab" section of the site, where we often write about trends in the media industry. These include:

• A memorial to fallen journalists:
  http://investigativereportingworkshop.org/blogs/shop-notes/posts/2017/jun/06/journalists-deaths/

• A new website to honor fallen or endangered journalists' work:
  http://bit.ly/2gZ7UIR

• Journalists recognized for their bravery:
  http://bit.ly/2hD2iEQ

• Top editors saying they're not at war with Trump:
  http://bit.ly/2yWw3ej

And she wrote about covering hate, a very strong story that was co-published on the Poynter.org website:

http://investigativereportingworkshop.org/ilab/story/media-covering-hate/
She's been a wonderful addition to our staff and accomplished even more than we had expected.

Lynne

Lynne Perri
Managing Editor
investigative-reportingworkshop.org
Journalist-in-Residence
School of Communication/American University
CHAPTER FOUR: PHYSICAL EVIDENCE

My assignments at the Investigative Reporting Workshop can be divided into two categories: database management and reporting. The two database projects, the Accountability Project and PBS Frontline investigation, have not been published yet. Hereby, I would like to share with you my published stories in the summer and fall.

Open data benefits many, but cost breakdown unclear

BY CLAIRISSA BAKER, YANG SUN    Thursday, July 20th, 2017

(This article was co-published with Investigative Reporters & Editors)

A new citywide data policy in Washington, D.C., shows there is no simple way for cities to clearly budget open data initiatives.

Meanwhile, as the city works this summer to implement its newly formed data policy and decide what’s releasable, experts say when more data sets are made available online the result will be better access to information, better journalism and more government transparency.

For investigative journalists, the upcoming data release — expected within nine months after the issuance of the order and adding to the 900 sets already available online — means greater access to information on everything from traffic patterns to invaluable health statistics.

“Just knowing the government has that data is a huge step,” said Kate Rabinowitz, founder of the DataLensDC blog that works to help citizens better understand and access the city’s open data.
But Washington officials can’t seem to determine how much the massive data availability will cost taxpayers. The work being done now at the Office of the Chief Technology Officer is in response to Mayor Muriel Bowser’s April 27 executive order, which calls for an inventory and classification of all government data.

Open data matters differently to each citizen, but it can matter a lot.

One example is Google Maps. Many use this tool every day, and it is based on open transit data from city governments, allowing people to determine best routes and commute times.

Rabinowitz also acts as one of the co-captains of Code for DC, a civic hacker community that translates data sets into content usable to average citizens.

One project Code for DC works on is called Housing Insights. The city has a massive amount of data on affordable housing, “but they are all over the place and it’s kind of messy,” Rabinowitz said.

A team of coders and data scientists collected all the relevant data sets and created interactive visualizations, allowing policy makers and the public to understand what affordable housing looks like in the District and what the challenges and opportunities are.

By sharing civic data, people will become more informed of city services, journalists will tell better stories of the city and institutes will advance their research, said Stephen Larrick, the open cities director for the Sunlight Foundation.

Quantifying the benefits of open data, however, can be as hard as measuring the costs.
“How can you put a price on an informed public,” Larrick said. “And how can you put a price on people having the facts that they are relevant to the decisions that are being made?”

It’s not the first time the city has made data public, but this new policy is an important step in making the government more transparent and accountable, experts say.

A decade after the debut of the city’s online data portal, there are more than 900 data sets available online on a range of subjects from information about 311 calls to crime reports to others on health care and government spending.

As city leaders work to implement the policy, there’s no clear cost associated with the rollout.

“It can be a real irony,” said Larrick. “Many open data programs are about being transparent about things like cost,” yet the cost of the program is obscure.

A number of aspects contribute to the inability to quantify costs of open data.

“Open data is a new thing, and very often it is the thing that is a part of someone’s job, but it’s not someone’s full-time job,” Larrick said.

Employees might work on data infrastructure or web services, among other assignments. Some of these tasks fall into the costs of open data, but unless employees track exactly how many hours they spend working on open data, it is not clear how much.

Not knowing how much open data programs costs is often a barrier to implementing policies.

The city is not at fault for lacking a concrete budget, Larrick said, but “the government should do a better job” of examining and listing these costs.
An analysis of the 2017 budget and the 2018 proposed budget for the Office of the Chief Technology Officer shows multiple line items related to open data, including “data transparency and accountability” and “data governance and analytics.” Even in those line items, however, it’s unclear what is related to the mayor’s new initiative.

Each agency is responsible for finding and categorizing its own data, so the costs are spread out and vary widely depending on the city, a Sunlight Foundation survey found.

One of the biggest differences between cities is whether they use a contractor — two include Socrata and Junar — to host the websites that contain this data.

Washington creates everything in-house; the city pays its staff to create and maintain a website to host the city’s data.

In January 2016, Bowser announced the Open Data Initiative and created the Chief Data Officer position.

About seven months later, Barney Krucoff filled the position, leading a team of more than two dozen people who will reach out to each agency and coordinate data collection.

Before Bowser’s 2017 executive order, technical teams were already in place. After, those were rearranged with staff from Business Intelligence, Geographic Information System and the Citywide Data Warehouse.

“D.C. was the leader of open data in general, but we didn’t have a data policy,” Krucoff said.

He said the District posted its data sets in the early 2000s and built a website hosting data sets published by the government in 2007, both among the earliest in the
country. The city even led a hackathon, called Apps for Democracy, in 2008, which Krucoff described as “new and novel of the time.”

US City Open Data Census, research co-conducted by the Sunlight Foundation, Code for America and Open Knowledge International, in 2015 ranked D.C.’s data openness 27th out of 100 cities, as San Francisco, Las Vegas and New York City took the top three positions.

Washington leaders hope to get back to the cutting edge of government transparency with the new data policy. Krucoff said what separates this administrative action from others is that it’s a data policy rather than an open data policy.

The government will not only log and categorize all of the data, but also create a system in which enterprise data is “freely shared among district agencies, with federal and regional governments,” and with the public when the information allows it, according to the data policy.

All city agencies’ data sets will be classified on a scale from zero to four, where level zero data sets have no confidentiality concern and should be completely disclosed to the public.

Significant steps will be taken to ensure the safety of information with privacy and security concerns. The policy includes a host of security protocols for agencies to follow while handling sensitive data sets.

Feedback from the public, such as transparency advocates and civic hackers, helped shape the District’s final version. The drafting team also looked at New York City, San Francisco and the state of Maryland, Krucoff said.
The government will proactively publish a whole class of non-confidential information. This will complement, but not replace, the Freedom of Information Act.

FOIA legally requires government’s reaction on individual requestors and covers items ranging from hard-copy documents to videos.

“I think FOIA will always cover a wider set of material, and open data will cover a more specific set of what we can be proactive,” Krucoff said.

However, there is a gap between the technical language of open data and the accessibility of it by citizens without a computer background.

To bridge the gap, data intermediaries, such as researchers and developers, play an important role. They use the data to make recommendations and tools that the public can understand and use.

Journalists use this resource to find information about their communities. Having information available online makes journalists’ jobs a little easier because the government can place data online that is asked for many times over, instead of responding to requests every day or every month.

“That transparency leads to, I think, a better relationship between government agencies and the public,” said Charles Minshew, data services director at the Investigative Reporters & Editors.

Opening up data is beneficial to both the government and the public, Minshew said, and “it is a true public service.”

Besides the promise of a citywide data inventory, the city also redesigned the online data portal by incorporating more functionality, including visualizations, search functions and interactive tools.
Michael Kalish, Rabinowitz’s counterpart at Code for DC, appreciates the city’s efforts in increasing the website’s usability.

“So I think they’ve very quickly went from something that was not very user friendly to a very approachably useful tool to the community,” Kalish said.

More needs to be done, however, to make data truly open.

Rabinowitz encounters data inconsistencies and missing records while working with city's open data.

In terms of improving the open data quality, Larrick’s primary suggestion for cities is to reach out to communities and listen to their needs.

“It really makes the benefits of open data a lot more tangible,” Larrick said.

Krucoff hopes going forward that the data policy will empower analysts of each agency to explore the value of data and develop a community in which agencies have a common set of tools and data-minded individuals.

“We generally believe that data is sort of an important asset to the city that we’ve never really known,” Krucoff said.

'Their sacrifices will not be forgotten'

BY: Jerrel Floyd and Yang Sun June 6, 2017

The murder of journalists throughout the world has become a technique for control, said National Public Radio’s Michael Oreskes at the annual Newseum Journalists Memorial event.

Forty-eight journalists were killed on the job in 2016, according to the Committee to Protect Journalists. The Newseum event on June 5 marked the addition of 14 of those
names, including NPR photojournalist David Gilkey, and all were recognized for their service.

On June 5, 2016, Gilkey and his translator, Zabihullah Tamanna, died in a Taliban ambush in Afghanistan.

“He [Gilkey] really didn’t photograph war,” said Oreskes, senior vice president. “He photographed the men and women, soldiers and civilians caught in these wars.”

Jeffrey Herbst, the Newseum’s president and chief executive officer, opened the event, which will add Gilkey’s and Tamanna’s names to a wall listing more than 2,000 fallen journalists’ names. The memorial emphasizes the unprecedented danger reporters face as they seek to provide information.

“Our sacrifices are extraordinary and will not be forgotten,” Herbst said.

Twelve of the 14 journalists honored died while reporting on their own countries — including Brazilian journalists João Miranda do Carmo and Sagal Salad Osman, one of few female voices on the radio in Somalia. Others were killed covering wars in Iraq, Libya and Syria.

“Whether it’s drug, gang and corrupt police in Mexico or the self-proclaimed Islamic State in Syria, murdering journalists has become a technique of control,” Oreskes said.

There are powerful organizations and governments killing or threatening journalists to keep them from uncovering the truth, he said. That is why the journalists honored at the event believed that their mission was to serve the world by helping to gather and spread knowledgeable information, he added.
“As we re-dedicate this wall by adding their names, we must also re-dedicate ourselves to the cause they died for. Because it is a cause to write, to gather and to distribute the news,” Oreskes said.

Claudine Kent, Gilkey’s girlfriend, said the event was a great way for the public to celebrate journalists’ work.

“It just honors him and his sacrifice, and brings to the public’s attention his work’s significance, and the importance of keeping his and others’ work in the public eye,” Kent said.

The Newseum Journalists Memorial is a two-story glass panel at the Pennsylvania Avenue museum in Washington. The panel is etched with the names of 2,305 fallen reporters, photographers, broadcasters and news executives. Names date back to 1837, when Elijah Lovejoy was shot protecting one of his publication’s presses from a pro-slavery mob. Dutch photographer Jeroen Oerlemans, who was shot while covering Libya’s war with ISIS, is the most recent addition.

**Media outlets struggle to develop strategies for covering hate**

BY YANG SUN  Oct. 17, 2017

(This story was co-published with poynter.org.)

Most news organizations, if asked, would be hard-pressed to articulate a strategy when it comes to covering hate crimes. But that’s emerging as a rising concern as the number of such crimes rises around the country.

Nearly 900 hate incidents against minorities were documented by the Southern Poverty Law Center during the 10 days after Donald Trump’s election victory.
But even as hate crimes occur with more frequency, newsrooms for the most part tend to treat them as individual problems, not as systemic problems that require better follow-through and focus.

That’s a problem, said Christopher Benson, an associate professor of African-American Studies and Journalism at the University of Illinois Urbana-Champaign.

“Hate crime is treated as a single event by a perpetrator or a group of perpetrators, instead of looking at it in the context of a lot of systemic problems,” Benson said, because many of those in the media have “a belief that the system is fundamentally sound.”

Heidi Beirich, the director of the SPLC’s Intelligence Project, said she doesn’t think this is just about the media. Instead, she said, there’s a “huge reluctance from part of the public to be willing to accept that there could be something born and bred in America ... creating the same kind of violence” as seen in some other countries.

Beirich said many people can understand the Boston Marathon bomber, the shooter at the Orlando nightclub and the gunman who opened fire outside of an Arkansas recruiting station because they see them as people radicalized by terrorists overseas.

“But when it comes to domestic terrorism committed by white people, they're treated as a whole bunch of disparate attacks,” Beirich said. “That’s a huge problem.”

Part of the problem is that reliable data is often hard to come by. Though it’s the primary government effort on tracking hate crimes, the federal database is widely considered inadequate because its data solely depend on voluntary reports from state and local law enforcement.
The 2015 FBI report counted 5,850 hate crime incidents. But surveys from the Bureau of Justice Statistics estimated that the real totality is likely closer to around 208,000 for that year.

**TRYING TO MAKE SENSE OF IT ALL**

“We are trying to investigate all the reasons why this data fall through the gap,” said Rachel Glickhouse, a journalist and the partner manager for ProPublica’s Documenting Hate project, said.

Documenting Hate is a collaboration of more than 100 newsrooms nationwide. The nonprofit news organization launched it in January hoping to “collect better data and try to get a better sense of how many hate crimes and bias incidents really happen in the U.S. each year,” Glickhouse said.

Playing the role of a “tip broker,” ProPublica compiles incident reports from news headlines, social media, civil-rights groups and law enforcement records, as well as hate-crime victims and witnesses. Each incident is treated as a “tip” and entered into a database. ProPublica shares the “tips” with its partner news organizations and relies on them for follow-up reporting.

This coalition has enabled many local media outlets to either report on single incidents that took place in their communities or to add additional sources. Its national partners even found patterns in the data analysis. For example, BuzzFeed published a series of stories on school children “using racist language to bully and harass other students,” Glickhouse said.
The job of documenting hate can be exhausting and frustrating. Glickhouse said that cleaning the database was demanding and her team faces a big challenge to verify as many tips as they can since more are coming in every day.

**TRYING TO TRACK THE HATE**

Beirich and the SPLC have been monitoring domestic extremism activities since the 1980s and contend that Trump’s campaign and presidency emboldened nationalist groups. She said the president, whose controversial remarks on immigration and racist comments on minorities pandered to the ideology of white supremacists, have made the hate groups more politically active than ever.

The SPLC counted 917 hate groups nationwide in 2016. However, that list that has been criticized by some who think descriptions of hate have gone too far, including 47 conservative leaders.

More activities among people with racist ideas have occurred in the past year, including campus hate incidents and marches using or advocating keeping the Confederate flag and monuments in place. Such events culminated in Charlottesville, which Beirich described as “an amazing example of white supremacists and malicious groups all coming together.”

The federal government documented a similar trend.

Months before the August Charlottesville white-supremacist rally, in which one person was killed and 19 injured, the FBI and the Department of Homeland Security warned in a joint intelligence bulletin that the attacks conducted by white supremacists had outnumbered ones by other groups in the country from 2000 to 2016.
“We assess lone actors and small cells within the white supremacist extremist movement likely will continue to pose a threat of lethal violence over the next year,” the bulletin read.

The FBI’s annual report on hate crime statistics also demonstrates a growing frequency of hate crimes within the past six years.

Researchers at the SPLC have to immerse themselves in a world full of page after page of racist language, an experience that Beirich described as “disgusting.”

MORE COVERAGE AT A CRUCIAL TIME

There were not as many organizations making such an effort in previous years. But more media outlets have realized that this problem needs to be addressed after last year’s election.

Journalism organizations have reporters now regularly covering protests and hate crimes. “When I talk to reporters today,” Beirich said, “as opposed to two years ago, they are absolutely more knowledgeable about this than they were.”

Such a shift in journalism is partially because of the fact that the country is undergoing a demographic transition. The Census Bureau said that babies of color outnumbered non-Hispanic white babies in 2015, and estimated that white Americans will lose majority status by 2042. This social reality has caused white supremacists’ anxiety and contributed to the rise in hate groups, Beirich said.

Benson said it is journalist’s responsibility to grow a certain level of cultural awareness in order to help the public reach a deeper understanding of these changes. “Until we are able to appreciate that context ourselves as media professionals,” Benson said, “we are not going to be able to clarify that for the public.”
A diverse society demands higher skills from journalists, making many media outlets shoulder heavier burdens. The Daily Progress, the 20,000-circulation daily newspaper that covers nine counties in Charlottesville, is no exception.

Allison Wrabel, a reporter for the Progress, wished she could have stayed closer to Emancipation Park during the morning of Aug.12 to gain more firsthand knowledge of the “Unite the Right” march and the ensuing chaos. She was one of six reporters from the news organization who covered the rally when usually only one is on duty on Saturdays. And the staff continued to work around the clock for the following two weeks, she said.

Assistant city editor Mark Newton said he was proud of their work ethic and what they accomplished, which included publishing nearly 200 articles relevant to the rally in the two weeks after the event and taking a photo that symbolized the march worldwide. But he said he regretted that no one from the newsroom caught the scene of a white nationalist firing toward a counter-protester during the rally until the American Civil Liberties Union released the footage.

When the rally occurred, the community news staff felt constrained because it had only six reporters due to a recent staff cutback. “I think the regrets were the limitations of the size of our newsroom,” Newton said. “By day two, three, four, the exhaustion had set in.”

Beirich said that it is so important for community media to cover hate crimes, which still remain underreported. “They can document it and draw attention to it,” Beirich said. “Maybe because of their reporting, something can end up in the federal hate crime statistics.”

Editors: 'We are not at war with Trump'
Top editors of two of the country’s leading news organizations said that their reporters were at the forefront of covering the Trump administration, but they were not competing with each other to topple the presidency.

Dean Baquet and Marty Baron, executive editors of The New York Times and The Washington Post, respectively, said that a journalist’s mission of pursuing truth and fairness hasn’t changed in the Trump era.

In an event titled “Guardians of the Fourth Estate” at the National Press Club, the two talked with legendary journalist Marvin Kalb as the part of the 24th season of The Kalb Report, a public television, radio and online series.

Kalb started the discussion with the proposition that democracy and freedom of the press are closely linked. But such a link is in danger now because President Trump constantly challenges the legitimacy of journalism by calling the media “fake news” and “enemies of the American people.”

It’s different but it’s still the same

Journalists are facing a different political climate now, not only because Trump escalated the usual quarrels between the administration and the press but also because he threatened to imprison journalists or revoke broadcasting licenses if their reporting displeased the White House.

Despite a more hostile and threatening environment for journalists, adherence to the core of journalism values remains unchanged, both editors said.

Baron said that when he walks into the newsroom every day, he can see the principles of The Washington Post on the wall. “The very first principle, which has been
around there for more than eight decades, is to tell the truth as nearly as the truth may be ascertained,” he said.

Baquet said the arrival of the digital age has already meant threats to traditional journalism standards in some institutions. When the Trump administration causes a significant shift in the culture of Washington, Baquet said, “That calls for us to stick to our principles even more.”

“I think, over the long haul, if you tell the truth, if you're accurate, if you're aggressive, and if you're fair and you hold on to your principles, I think in the end that's the only way we can cover (him),” Baquet said.

The president’s attacks on the press have become so frequent that news organizations cannot afford the time to respond to each of them. All they can do is to treat them as some degree of background music — an unpleasant one — Baron said.

“On his first day in office, he went to the CIA headquarters and said, ‘I have a war with the press,’” Baron said. “The reality is we don’t have a war with him. We’re not at war. We are at work.”

He said, “If we were to react to this every single day and spend our time making an issue out of it all the time, we wouldn't be able to do our jobs.”

How to rebuild the trust

Kalb asked the editors how to rebuild trust when Trump repeatedly berates the press. A recent Gallup poll showed that the American public’s trust in mass media continues to drop.
Baron said a large segment of the population considered the media a “partisan enterprise,” and added that people's confidence in media reports hit bottom during the Watergate investigation.

“Then it turned out that reporting was validated, and the approval ratings for the press after Nixon’s resignation went sharply up, to the highest point we’ve seen,” Baron said.

Calling himself “maybe naive,” Baquet said as long as news organizations aggressively play the role of government watchdog and hold onto their values, trust will be restored.

Baquet and Baron also agreed that transparency will help re-establish the public’s confidence.

“People are entitled to know who we are. They’re entitled to know how to reach us,” Baron said. “We can show more of the documents that we depend on in doing our reporting. We can disclose more full transcripts or full audio of interviews that we’ve conducted.”

**Mixed feelings for media’s business future**

Baquet and Baron contemplate the current challenges for funding local news organizations.

Asked about the single biggest challenge The New York Times encounters, Baquet cited the financial pressure three years ago, which he said is no longer the case.

In fact, the two organizations have seen a dramatic increase in their readership since the rise of Trump.
Their financial model now relies more on paid subscriptions, a departure from one that depended heavily on advertising. Baquet described such a transition as a “remarkable and terrific development.”

“I think people want to read strongly reported and well-reported journalism,” Baquet said. “They are flocking to us.”

Baron said that Trump’s campaign and presidency shifted how Americans treat the news media.

“People don’t take the press for granted anymore,” he said. “If they don't support quality journalism by paying a very little amount of money, they will not get quality journalism.”

While the big news platforms, such as the Times and the Post, saw a record of revenue growth after last year’s election, the decline of local news organizations seems inevitable.

The financial struggles forced regional media to shrink their newsroom staffs, leaving many public institutions and individuals uncovered, including congressional delegations in Washington, state legislatures, county commissions and school boards.

The two said this lack of coverage is their biggest concern.

Baquet said that in the next four to five years many local newspapers are going to go out of business, and the consequences of this are catastrophic.

“We are in the middle of a crisis that people have not woken up to,” he said.

**Journalists rewarded for their bravery**

BY: YANG SUN   Oct. 27, 2017
Every year, many female journalists are harassed, detained, beaten or even killed just for doing their jobs. Despite many efforts made for improvement, the situation hasn’t changed.

“The attitudes toward women in politics and women in journalism today is more aggressive, is more sexually threatening, more abusive and dismissive than any time that I can recall,” said Andrea Mitchell, chief foreign affairs correspondent at NBC News, in a ceremony honoring courageous female journalists last week.

The Oct. 23 event was the part of the International Women’s Media Foundation’s annual celebration to honor female journalists all over the world who demonstrated courage, commitment and sacrifice in pursuing the truth.

Since its foundation in 1990, the foundation has honored more than 100 women in journalism from 56 countries.

A legend in broadcasting journalism, Mitchell used to be the only female broadcaster during the 1960s and 1970s in Philadelphia interviewing politicians and public figures.

“Now we are seeing the same kind of insults, all against political opponents and journalists, men as well as women, but particularly women,” Mitchell said in her speech to accept the Lifetime Achievement Award, “whether they are just described as wacky or weak — or as Bill O’Reilly called me out for trying to ask Secretary of State a question ‘as an unruly woman.’”

And while journalists in Washington are under attack for doing their jobs by holding public officials accountable, reporters outside of Washington are braving danger every day, Mitchell said.
The death of freelance Swedish journalist Kim Wall, the killing of Daphne Caruana Galizia, who covered Malta's Panama Papers link and the violent attack on Tatyana Felgenhauer, a Kremlin critic radio journalist in Russia, show that female journalists are facing unprecedented challenges today.

Deborah Amos, NPR’s Middle East correspondent, is no stranger to such dangers. During her 40 years of experience reporting from conflict zones, Amos was detained in the Balkans and Iran and kidnapped in Somalia.

Accepting the Courage in Journalism Award, Amos said, “No one gives you a courage test the first time you are given a dangerous assignment because there is no reliable measurement.”

Her job is to ensure that the experience of civilians who endure the brutality of war gets the attention it deserves, she said. “It can get risky at times. But I have support and encouragement, and I always knew that I could walk away if it ever got too tough,” she added.

But women reporting from their home countries cannot walk away from suspicion, censorship and even persecution. This includes Hadeel al-Yamani and Saniya Toiken, two other Courage in Journalism Award recipients.

Hadeel al-Yamani is the first woman to become an Al Jazeera Arabic television correspondent in Yemen, covering the country’s civil war from the front lines. Being a war correspondent in a country in which a large number of people are deprived of basic human rights, al-Yamani said she has to wear body armor with her abaya and hijab during work.
But al-Yamani was denied a visa, which prompted comments from several honorees.

“Hashtag ‘Hadeel’s visa.’ I think we should make more of the noise about that with all the recipients here in Washington, so close to the place where visas are given out,” Amos said on stage.

In her news program on MSNBC the next day, Mitchell asked, “What threat can a 25-year-old woman journalist who is braving the civil war in Yemen possibly present to the United States? If she doesn’t embody U.S. values, I don’t know who does.”

In the acceptance remarks sent to the foundation, al-Yamani wrote, “Journalism, on its basic level, is to report news without biases. However, where we work, it’s become about dealing with the bodies and injuries of children, dodging snipers’ bullets aimed at you or at your colleague or covering the news of your colleague’s abduction.”

Al-Yamani said "my courage is not even comparable to that of a woman who bids farewell to her husband or child as if it’s their last when he goes to defend his land."

Saniya Toiken is always ready to say goodbye to her family and friends in Kazakhstan because she is a favorite target of the dictatorial government and law enforcement.

A reporter for Radio Free Europe/ Radio Liberty, Toiken refuses to turn a blind eye to the injustice and corruption, as most local media outlets do, she said. As an independent journalist in Kazakhstan for two decades, Toiken is accustomed to government’s harassment, threats and imprisonment.
She said many people don’t understand why she constantly puts her life at risk. Even her son called her once when she was detained and interrogated, and asked, “Mom, why are you doing this? Take it down a notch!”

“I am happy when heroes of my stories call me and thank me for covering their stories,” Toiken said. “For me, journalism is my life, and I cannot change it.”

**New site keeps journalists' stories alive**

BY: YANG SUN  Nov. 2, 2017

On March 2, 2017, Cecilio Pineda, a Mexican investigative reporter, posted a video about the close tie between a drug cartel leader and a local politician on his Facebook account. Two hours later, he was murdered.

Pineda’s investigation, along with ones of two other fallen journalists on drug cartels in Mexico, has been translated into nine languages by Forbidden Stories, a newly launched website, so their work can reach as many people as possible.

Laurent Richard, the founder of Forbidden Stories, a renowned TV journalist from France, presented the project at the Newseum on Monday in Washington.

The nonprofit online platform serves as a safe-box for journalists in danger to upload sensitive information via an encrypted communication channel. In addition, staff at Forbidden Stories will complete the work if a reporter is imprisoned or killed.

“If you are a journalist facing threats, it’s very important for you to protect your stories. We provide a way for journalists in danger to make some backups,” Richard said. “In case something happens, we will be able to complete, to follow, to fact-check and to publish the story broadly.”
Digital security is key to the project, which uses open-source technologies, such as Signal, a secure messaging application for the smartphone, and SecureDrop, a whistleblower submission system, Richard said.

Richard developed this idea two years ago when 12 people working in the newsroom of Charlie Hebdo, a French satirical weekly in Paris, were killed by terrorists.

The offices of Richard’s company, Premières Lignes, were next door. He and his colleagues were the first to enter the newsroom after the killing, helping those who were still alive. “That event made me more aware than ever about the fragility of the free press,” Richard said.

His vision to help journalists at risk was not only spurred of the death of journalists at home. Three months after the attack, Richard received a piece of paper from Khadija Ismayilova, an Azerbaijani reporter who was imprisoned in Baku, the country’s capital, because of her investigation into corruption among politicians in Azerbaijan and throughout Europe.

The note reminded Richard of the importance of continuing Ismayilova’s work, and also convinced him of the dire need to create a platform to keep forbidden stories alive.

Independent journalists all over the world are frequently harassed, threatened, jailed or killed for doing their jobs. Nearly 50 media professionals have been killed so far this year, and 300 professional and citizen journalists are currently imprisoned, according to statistics from Reporters without Borders, a partner of Forbidden Stories.

Meanwhile, the crimes and wrongdoings that journalists are revealing don’t stop at national borders. Reporting on the environmental crisis, money laundering and
corruption has become so complicated that such issues are rarely only of local and regional concern.

Globalization of reporting — especially in light of global censorship — is more critical than ever, said Dana Priest, a longtime investigative reporter at The Washington Post who moderated the panel discussion after Richard’s presentation.

Yet cooperation among journalists from different media outlets and countries can be a huge challenge.

“We are not born or taught to collaborate,” said Marina Walker Guevara, deputy director of the International Consortium of Investigative Journalists (ICIJ). "This is not in a reporter’s DNA. We were taught to scoop one another.”

But the journalist’s lone-wolf mentality is no longer applicable when the stories are too complex and dangerous for a single media organization or reporter to tackle, she said.

Her organization, a global journalism network founded two decades ago by Charles Lewis as part of the Center for Public Integrity, proved the power of collaboration by coordinating reporters from 110 media outlets in more than 80 countries to produce the award-winning Panama Papers. ICIJ has now spun off as an independent network of more than 200 organizations in 70 countries.

But when compared to journalism groups from Africa or Latin America — where media outlets are facing stricter censorship and limited resources — news organizations in the United States are generally more self-sufficient and less willing to collaborate, Walker Guevara said.
When ICIJ contacted several American media outlets at the beginning of the Panama Papers project, she said they were met with reluctance.

“They told us, ‘That sounds complicated and time-consuming. I’ve been very successful just working on my own,’” Walker Guevara said.

The nonprofit ProPublica website is now making an effort to work with other partners and also creating its own local chapters; ProPublica Illinois recently launched.

Walker Guevara said that American media now show more interest in working with others.

“I think we are going to see more and more collaboration in the U.S. and global cooperation with the U.S.,” she said.

The concept of continuing a fallen journalist’s story is not new, Walker Guevara said. One of the earliest examples of journalists working together can be traced to 1976, when investigative reporter Don Bolles was murdered for investigating organized crime in Phoenix.

In response to Bolles’ death, 38 journalists from 28 news organizations across the country traveled to Phoenix and produced a 23-part series in 1977 exposing widespread corruption in the state.

Forbidden Stories, equipped with technology on information safety, now takes this model to a larger and more globalized scale, Walker Guevara said.

“When journalists join together and move together,” said Walker Guevara, “if something happens to one of the reporters, the other reporters are going to react as a team.”
CHAPTER FIVE: ANALYSIS

As the newspaper industry started to collapse in 2006 and the whole U.S. economy fell into recession in 2008, Chris Amico placed his bets on computer coding.

The economic downturn made the journalism job market more competitive. As a result, the career path of moving up from a small community newspaper to a major news organization, as Amico said, "seemed really unlikely" to him.

Living in Dalian, a coastal city in northern China, he kept thinking about one question, “Am I going to have a journalism career going forward?” Fortunately, daliandalian.com, a website project he was involved in, strengthened his interest in computer technologies and introduced him to the possibility of combining journalism and programming.

Amico taught himself how to code and gradually shifted his career as a print reporter into a news developer. After returning to the United States, his programming skills helped him land jobs at PBS NewHour and NPR. Now he is an interactive editor for PBS Frontline.

Journalist-programmers, including Amico, are advocates for journalism incorporating computer technologies. As media outlets have grown savvier about the web in recent years, newsrooms have offered more positions that require people fluent both in journalism and web development. Computer technical skills expand journalists’ career possibilities.
However, learning how to code without any technical background can be overwhelming and frustrating, particularly for working journalists already burdened with regular assignments.

That challenge prompted me to speak with media practitioners who have made the shift from being a conventional journalist into a journalist-programmer, to gain perspective on the motivation behind their transition and their career trajectory.

I interviewed eight journalists from national and regional news organizations, and a journalism educator who used to be a journalist-programmer. Their day-to-day jobs include interviewing, writing, crunching numbers, typing code and designing graphics. All of the journalists enrolled in journalism programs without any programming courses or started their careers as traditional reporters. Some journalists dabbled in programming as full-time reporters. Some younger journalists, who had seen the rise of digital production, made the transition while attending college.

There were two female journalist-programmers among my interviewees. They both said that they were not treated differently based on gender in the workplace. The annual American Society of News Editor Census shows that men greatly outnumber women in the newsroom in the U.S.

No matter when and how the transition into a journalist-programmer occurred, the journalists I talked to agreed on three main points. First, the desire to create distinctive journalism pieces was their primary motivation. Second, the most effective way to learn is by working on journalism projects. Third, networks played a significant role in their career paths. The journalists had mentors, friends and co-workers who influenced their learning.
**Getting started**

The journalists who made a transition nearly a decade ago were initially impressed by the potential of computer-assisted reporting (CAR). They picked up how to analyze a dataset in an Excel spreadsheet or database manager software. The next logical step down this road was programming for more advanced CAR.

Ben Welsh, a data editor from the Los Angeles Times, said that he became attracted to the combination of computer technologies and journalism while he was helping two legendary reporters -- Carol Marin and Don Moseley -- on an investigative project for the DePaul Documentary Project in 2003.

During this project, Welsh made his first public records request and did his first data analysis in a Microsoft Excel spreadsheet.

The local government of a Chicago suburb, he recalled, sent him and his reporters the documents in the form of a massive printout. His responsibility was to hammer all the numbers from the printout into a spreadsheet line by line. After hours of tedious, manual data-entry, he added up the final figure and a story on corruption in the local government came out. The moment Marin read the numbers on camera, it clicked to Welsh about his journalism career.

“I learned that technical skills were a way I could contribute,” he said. “Even [the digital tools] just as simple as a spreadsheet kind of opened my eyes.”

Like Welsh, Serdar Tumgoren, a data journalist from The Associated Press, was also fascinated by how journalists could tell a different story by working with spreadsheets.
He considered his journey to programming a traditional "CAR path," a road that was expanded from spreadsheets to database manager software, programming languages and web development. He said that it took him nearly a decade to completely transform from a print reporter to a news developer.

For journalists who made the transition, the realization that programming could make their work stand out sparked their motivation to continue learning.

Matt Waite, now a professor of practice at the University of Nebraska-Lincoln, saw some news publications put searchable databases online in the ‘90s. Then he told himself that such transparency could be the future of investigative journalism. It inspired him to pitch a similar idea to his newsroom, the St. Petersburg Times, but nobody there knew how to publish an interactive database on the Internet. To create such a project, Waite started to teach himself how to build a web application from scratch, bit by bit, line by line, item by item.

In 2007, Waite finished his first web application: He was one of the primary programmers for PolitiFact, an online platform to fact check politicians’ statements. He said that the experience of learning Django, the web framework used to build PolitiFact, was transformative and changed his thoughts on journalism. In 2009, PolitiFact became the first website awarded a Pulitzer Prize.

Adopting programming changes how journalism can be presented to the public. In addition, it alters a journalist’s workflow. Instead of constantly checking a website's update manually, he or she could automate this task by writing lines of code. Rather than depending on a source for a scoop, journalists can use Pandas, a Python library, to
concatenate different datasets and find stories that nobody else can. With the assistance of JavaScript, a journalist can add interactive narratives to storytelling.

Before learning how to program in R, an open-source computer language for statistical analysis and graphics, David Montgomery, a former political reporter from The St. Paul Pioneer Press in Minnesota, had a toolbox which he thought obsolete to him now -- Excel, Adobe Illustrator and Tableau, etc.

His first venture into programming came as he was working on a personal project for which he needed to make a graphic for each state. Generating 50 graphics one by one in Excel, a method he had been familiar with, was possible but tedious. Instead, he found an R script online that could achieve the same goal by automation.

Now Montgomery could create a word cloud based on the governor's speeches, scrape data from websites and analyze spatial data, tasks that he couldn’t do without knowing how to code. “It [ability to code] made my job more efficient and sophisticated,” he said.

For younger journalists, the story is different. They went to college when online journalism was on the rise, learned basic HTML and CSS in class and moved on to acquire more sophisticated programming languages through their internships. This was the path for Sisi Wei, an investigative reporter from ProPublica, and Darla Cameron, a visual journalist from The Washington Post.

**Getting your hands dirty**

For anyone who wants to get their feet wet with coding, he or she has ready access to learning resources. There are many tutorials online, teaching various programming languages to beginners, mostly for free. However, for a working journalist,
the biggest constraint is time. Journalists are already under the stress of needing to meet deadlines constantly. It seems impossible for them to squeeze out a few hours every day to sit in front of a computer and follow the tutorial videos.

The most effective way to learn, nearly every journalist I interviewed said, is a project-oriented approach. They said the learning experience should be modified for a news environment, and a journalist could learn by working on a journalism project.

Welsh saw every opportunity to work on a news project as a chance to stretch his programming skills and pick up something new. He learned how to analyze geospatial data by creating Mapping L.A., an interactive online platform displaying the neighborhoods in Los Angeles. Later, he applied the knowledge to analyze emergency response time in Los Angeles.

He said he didn't have any secret to learning how to program. All he did was to “aggressively experiment with them on really practical projects, not on things that are academic or unrelated to journalism.”

Montgomery also valued every possibility to transform his programming knowledge into a tangible product, rather than attend a lecture full of the nitty gritty of programming. The combination of reporting and programming skills allows him to create distinctive journalism products. The better his story turned out to be, the more time his editor would grant him to experiment, a cycle that Montgomery described as “a snowball effect.”

Acquiring new knowledge while having a full-time job means sacrificing a lot of personal time and having strong determination. Several journalists said that the early days of learning how to code were the most intense period of their lives.
Amico learned web development as being a freelance reporter in California. “I went out to report a story for a newspaper, wrote it and sent it off,” he said. “Then I came home and stayed up late at night, doing Python tutorials or building little web applications.”

Tumgoren also attributed his learning experience to lots of late-night hacking. He recalled his life in those days as “a reporter by day and a coder by night.”

He started tinkering with programming by himself around 2005, but the result didn’t go very well. He enrolled in a Perl class at community college. After the class, he continued to teach himself at night, intensively, for a number of years.

“I was totally obsessed, and I didn't have a girlfriend who was close to me at the time, to be perfectly honest,” Tumgoren said. “I'm not recommending that approach for people today. I think there are much more sane ways to do it.”

Several journalists I interviewed, including Tumgoren, talked about the work-life balance during their transitions. They said that being single and without a family gave them more time to code outside of their daily jobs.

Cameron, who just had a baby girl, said that she experimented with more new digital tools and skills in the early stage of her career.

"I've definitely gone through a period that I worked all the time," she said. "But now I always have to leave at 5 to pick up my child."

The journalists said that one of the worst ways to learn programming is to learn it like you’re studying an academic subject.
Greg Linch, a news developer at the McClatchy Group, still remembered the first time he learned how to code in JavaScript, a programming language commonly used to create interactive web visualizations.

“I was reading a physical book. And I would read it like I was studying history or science,” Linch said. Soon after, he realized that reading books was not working and felt the need to go back to the computer.

The learning experience is a loop of fumbling with typing lines of code, reading others’ documentation, figuring out the reasons behind the code and experimenting with a new project. “Repeat. Repeat. Repeat. It’s a constant process,” he said.

Despite the misconception of learning programming as a subject, another common mistake that novice coders may make is to set a goal of learning as many digital tools as they can - ambitious, but impossible and unnecessary.

The journalists I spoke with suggested that beginners set an achievable goal, make medium- and long-term plans and then find ways to make them happen. “Every time I learn a new language or a new concept, it’s because I need it,” Wei said. Working on projects frequently is the best way to sharpen one’s programming skills, she added.

**Everybody needs a network**

During my conversations with these journalists, I noticed everyone mentioned they had mentors and a strong community that they could turn to for help. How they connect to a community varies according to the size and structure of the newsrooms.

The national news organizations may have a larger group of newsroom programmers than smaller or regional media. Therefore, journalists there have more opportunities to build their skills through in-person interaction.
Linch said his path to becoming a journalist-programmer was different from the ones of many other journalists. The transition took place through an internal experimental fellowship when he worked at The Washington Post.

He had many friends who were journalist-programmers, and he was always fascinated by their jobs.

The idea came out on the day that he was watching a baseball game with Jeremy Bowers, who was a senior developer at the Post.

“He said, ‘Why don’t you pitch an internal fellowship to work on?’” Linch said. “It was experimental and we wanted to try this and see how it worked.”

After the Post’s technology department approved the proposal, Linch was embedded in the newsroom development team, a mixed group of people with technical and journalism backgrounds. During the following six months, Linch was kept busy with building tools for the newsroom and asking questions, an experience he considered “very important” for his transition.

"It was something that nobody had done before, and I don't think anybody has done since," Linch said.

Cameron, who works on the Post’s graphics desk, told me that her team was very supportive of journalists learning to program. She said that the desk holds an hour-long weekly meeting called Skillshare, where every member can present a new project or share fresh knowledge.

The team also has a Slack channel where they can communicate virtually.
Comparatively, journalists at smaller newsrooms have less of a chance to be around someone that they can turn to for correcting a coding error. The solution is to go to the Internet.

The lone coder in his newsroom, Montgomery is active in the Lonely Coders Club, a Slack channel filled with like-minded people. He posts questions there and tries to answer others’.

“I like sitting down to hash out a problem by myself,” Montgomery said. “And I turn to the Internet for help if I get stuck.”

However, a bigger hump for the lone coders to tackle is how to let the rest of the newsroom understands their jobs. For Montgomery, the approach is “very much finished-project-based.” He said that it was especially necessary during the early stage that he tried to establish himself as a journalist-programmer.

“So a lot of time what I do is to work on something in my personal time, and the next day I would show my bosses something quasi-finished,” Montgomery said. “I couldn't describe what I wanted to do if no one understood the method.”

The debate persists

When Adrian Holovaty called for more programmers in the newsroom in 2006, he didn't get much response. A decade later, after the decline of the newspaper industry and the rise of digital media, more journalism professionals have awakened to the advantages of learning web development. However, whether all journalists should learn how to code is an ongoing debate.

An article published by The Atlantic in 2014 said that the knowledge of how to code cannot advance someone who wants to do serious reporting. As the marketplace for
a reporter has become increasingly competitive, journalists should not waste time picking up programming languages because they would never be as good as professional programmers, the author wrote.

The article ignited fierce debate among media professionals, including some who believe that coding skills are essential for anyone who wants to survive in the news industry in the future.

“You will be scooped by a reporter who knows how to program,” Scott Klein, the deputy managing editor from ProPublica wrote in the same year. He was against the idea in The Atlantic article that jobs such as web design and data visualization should be performed by professional programmers.

The journalists I talked to agreed that it is unnecessary, as well as impossible, for all journalists to be expert coders, yet they should at least be literate on data and programming.

The availability of massive amount of data, from governments, organizations and individuals, has made the world that journalists cover more complicated. The lack of data and programming skills leaves many stories out of reach for many journalists.

“If we choose not to have the skills to verify that information, we are completely reliant on what is given to us.” Wei said, “It seems to me the equivalent of saying, ‘Well, just believe everything your sources say.’”

**Conclusion**

The journalists I interviewed, regardless of their different job titles and various digital tools they use, demonstrate a common set of qualities: being forward-thinking and willing to learn.
The journalist-programmers, particularly the pioneering ones who made the transition early, saw the power of computer programming. To stay at the forefront of media innovation, they sought every opportunity to apply their new skills and expand their knowledge.

As Waite said, curiosity should be at the center of a journalist's career, regardless of whether he or she learns programming.

The media industry is rapidly changing and competitive. Perhaps in a few decades, newsrooms will be saturated with journalists who can code. Journalist-programmers perhaps will no longer represent a niche in the job market. Instead of discussing which particular set of journalistic skills distinguishes a journalist from the crowd, it will be more meaningful to emphasize the traditional values: keep learning and always be curious.
APPENDIX

FROM JOURNALIST TO CODER:
THE RISE OF JOURNALIST-PROGRAMMERS

Project Proposal

Introduction

Before I enrolled in the journalism graduate program at the University of Missouri in 2015, I had worked as a political reporter at the Xinhua News Agency, China’s biggest news outlet. I witnessed that the journalism industry in China, facing the trends of digital media, was in dire need of journalists capable of both reporting and analyzing data. However, the Chinese media industry and journalism schools there are still experimenting to train such journalists effectively. So, without hesitation, four years after I started my career as a reporter, I quit my job and moved to the United States to study at the Missouri School of Journalism, one of the top institutions in the world for data journalism education.

Once I enrolled at the journalism school, I saw a different world open up to me. Excited by the various courses, my data skills have grown rapidly. David Herzog’s class on computer-assisted reporting was my very first opportunity to train in the data mindset. I learned how to use open records laws to request data from the government and analyze them with SQL (Structured Query Language). From the advanced data journalism class, I
acquired knowledge on how to use computational techniques to scrape online data and analyze them. During my spare time, I taught myself several programming languages, including Python and JavaScript, to visualize data and develop news applications. Additionally, I studied the latest trends in the industry, including convergence journalism and virtual reality journalism. I applied this knowledge to my day-to-day jobs while I was working at the Columbia Missourian and KBIA, the NPR affiliate radio station here. I produced multiple data-driven stories that could be published across different media platforms, including an in-depth story investigating affordable health care in Missouri and a data-driven story uncovering income discrepancy of Missouri teachers.

Learning these skills and having opportunities to apply them to professional work reinforced my plan for after graduate school --- to be a data journalist, particularly a hybrid data journalist, who can do traditional reporting as well as write code. First of all, developing multiple skills can make me more versatile and employable in a newsroom. I will have a better opportunity to work in a newsroom where my mother tongue is not the working language. As a non-native speaker, writing complicated and long-form stories can be more challenging for me. Therefore, I want to stress my coding and multi-media producing skills. I can maintain my position at the newsroom as a technological savvy person and a visual storyteller, compensating for my weakness as an international employee.

Most importantly, journalism is a constantly evolving field influenced by emerging technologies. Graduating from the Missouri School of Journalism is not the end of my journalism education. Being a hybrid journalist means being adaptive to the rapidly changing industry, and it requires a strong ability for continuous learning. Mastering
coding skills opens the door to me to better tell stories because computer science can expand the journalism narrative into a more interactive and engaging presentation. My reading of scholars in the field shows that many newsrooms nowadays put emphasis on journalists’ technical skills since media is closely tied to technology. Journalists are usually right-brain thinkers as the right side of the brain relies more on intuition and creative thinking. Programming heavily relies on the left brain that is more apt for logic thinking and analyzing. Being both a journalist and a programmer/coder requires the use of both sides of the brain. The transition into a journalist-coder shows a journalist's ability to think both analytically and intuitively to thrive in the new era of journalism. Therefore, it is meaningful and relevant to examine how data journalists who used to work in traditional journalism fields picked up computational skills via daily workflows and transformed into hybrid journalists.

**Professional skills component**

I will be working at the Investigative Reporting Workshop in Washington, D.C. for the professional component of this project. IRW is a nonprofit news organization based at American University that publishes data-centered and investigative stories specializing in accountability of government and public agencies. The newsroom is project-oriented and has several teams working on a couple of long-term projects.

My internship at the IRW consists of two periods. My eight-week summer internship will start on May 23, 2017. My 14-week project time will start on August 21, 2017. My position requires me to work 32 hours per week. I will start to work on my research component while I am interning in the summer. In the fall semester, I will attend
the Washington program's seminar on Friday mornings alongside my internship and research component during the rest of the time.

During my 22 weeks at the IRW, I will acquire more hands-on experience. I believe that an extended period of internship will grant me full immersion into a project. Working with an established data team will give me valuable experience in the field. I plan to follow the whole process from data request, through data analysis, interview, and production. I will spend 32 hours per week working at the workshop. I will spend the remaining hours each week working on the research component of this project. I have not been assigned to any particular position or project yet. I explained my coding skills and multimedia producing experience to the managing editor, Lynne Perri. She promised that there would be a project to which I can apply all my journalistic skills.

Theoretical framework

The research component of this project will explore the transition from journalist to journalist-programmer - specifically a journalist who used to be a traditional journalist interviewing people and writing stories, then acquired programming skills and now have to regularly apply such skills to day-to-day work. Professional publications like Columbia Journalism Review, Digital Journalism, American Journalism Review or the IRE Journal may be interested in publishing this analysis. Three research questions are proposed as follows:

RQ1: What are the motivations for a traditional journalist to transition to a journalist-programmer?

RQ2: What is the path for a journalist to move into being a journalist-programmer?
RQ3: What do the journalist-programmers do to stay on top of their craft?

The integration of computer programming and data analysis in news reporting can be traced back to 1952 when CBS employed a computer program to predict the outcome of 1952 presidential election (Cox, 2000). Over the past decades, data journalism has become more prominent than ever due to the advent of the big data era that commercial and governmental agencies release a massive amount of data on a daily basis. As more computer technologies are adopted to help create complex data-driven journalism projects and produce quality online content, a growing number of journalists have picked up programming skills through their daily workflows, some even have become full-blown programmers in the newsrooms.

This research topic is timely because whether all data journalists should learn to code and how much coding data journalists should know are still ongoing debates in the data journalism field (Du, 2013). This research is relevant and important because nowadays more journalists and news organizations have realized how computational and data techniques can be indispensable for powerful storytelling and compelling digital products. The number of NICAR (National Institute for Computer-Assisted Reporting) conference attendees increased by almost 50% from 1999 to 2017 (Zhang, 2017), showcasing that more journalists have realized how essential data skills are to their career. The computational tools also evolved very fast. Statistics show that in 1999, the NICAR conference had hands-on sessions on six programs, however, in 2017, 22 programs were taught at NICAR. Many journalists, dealing with heavy job loads every day, are struggling to find a balance between maintaining their routine jobs and keeping abreast of new technology. The exploration of the experiences of data journalists who
have managed to overcome the struggle can be rewarding to the journalism community. This topic also fits in with my career plan, as I am intending to be a journalist-programmer. I will encounter the same struggles and hurdles that many journalist-programmers had during their transitions.

*The diffusion of innovations*

The journalists’ transition into journalist-programmers is a process of adopting innovations. The research on such transitions can be governed by the theory of diffusions of innovations. This mass communication theory examines individuals or an organization's motivation in and the process of learning and adopting new technologies.

The earliest studies on this theory can be traced back to research done by European sociologists and anthropologists in the late 19th century (Rodgers, 2003). The theory of innovations diffusion hadn’t fully developed until Rodgers, a rural sociologist, published his seminal work: *Diffusion of Innovations* in 1963. Rodgers (2003) breaks down the diffusion of innovations into four main elements: “…… (1) an innovation (2) that is communicated through certain channels (3) over time (4) among the members of a social system” (P36).

According to Rodgers, the innovations include two components: hardware, which is tangible tools, and software, knowledge on the utilization. As journalists transition into journalist-programmers, the innovations they have to adopt to include computational tools for data analysis, programming languages for scraping online datasets, visualization software for data presentation, as well as being able to look at the world with a coder's point of view.
Since this research is going to explore how journalists adopted computational skills, it is important to understand the concepts of the adopters and the innovation-decision process of the theory. Rodgers (2003) categorizes the innovation adopters into five groups: the innovators, early adopters, early majority, late majority and laggards. He concluded that the reasons determining the adoption rates are different socio-economic status, personality variables, and communication behaviors. Rodgers noted that making the decision on whether to adopt the innovation is not an instantaneous act. Instead, it is a process involving active information-seeking and information-processing behaviors. During the process, an individual is exposed to innovative knowledge at first, then he/she comes up with an attitude toward it that leads to making a decision on adoption or rejection. The adoption finishes with an implementation of the new idea and a confirmation of this decision (Rodgers, 2003).

The innovations diffusion theory has been applied to studies in many fields, such as agriculture, economics, sociology and communication. To inform this study, it is helpful to analyze how journalism scholars studied the innovations diffusion in newsrooms. When the internet and personal computers started to become accessible to the public in the late 20th century, the theory was the most widely used theoretical paradigm to understand how the technologies were spread and adopted in the newsroom, and how the roles and skill sets of journalists were changed. Research by Garrison (2001) explored how computers and the internet were adopted in the newspaper newsrooms in the United States during the 1990s. The study used a series of national mail surveys of newspaper staff to observe the spread of internet technologies from 1994 to 1999 and found that the diffusion model followed the classic S-shaped curve proposed by Rodgers.
(2010). Though the questionnaires were mailed to individuals in the newsroom, "respondents were asked to respond on behalf of the entire newsroom." The newsroom as a whole was the analysis unit, and the perceptions of journalists as individual adopters were not addressed.

As internet technologies developed, various tools were available to journalists. Later studies put more focus on the specific innovations. Swasy (2016) looked at the factors that influenced the spread and adoption of Twitter among four metropolitan newspapers because Twitter is a technological change that can advance how journalists "gather and disseminate information" (p. 645). In this research, Swasy treated each newsroom as a social system within which the adoption of Twitter has been diffused.

By interviewing 50 journalists, Swasy found that the most significant factor was peer influence. Words were spread as journalists saw others using Twitter to find story ideas and sources. The majority of journalists were hesitant to incorporate Twitter into their daily jobs right away because they found it was challenging to navigate "the line between their personal and professional lives" (p. 649) when they started to experiment. However, Swasy found such hesitation and resistance did not last long. People changed their minds quickly because Twitter allowed journalists to see its benefits immediately, such as their stories getting more attention or having more followers. Swasy also ascribed such rapid change to the diffusion of mobile devices, since they made Twitter more compatible with journalists’ daily jobs. She concluded that this "shows another factor in the spread of the innovation: other technologies that will enhance and speed up diffusion". (p. 654)

Applying the diffusions of innovations to data journalism
Compared to conventional journalism, data journalism, specifically, the phenomenon of adopting computational skills, is an innovation in the newsrooms. Scott Maier (2000) analyzed the diffusion of computer-assisted reporting, the practice that data journalism originated from, in the newspaper industry. Similar to other earlier journalism scholars guided by the theory, Maier referred to e-mail communication, internet search, data analysis, and mapping software as the CAR skill set, as well as innovations.

By surveying 28 CAR trainers from 27 news organizations, the overall finding of Maier’s research was that the adoption rates of CAR resembled a normal distribution curve. Maier concluded that the adoption of CAR already passed the innovator stage and was spreading into the early majority stage. In general, the author took a pessimistic stance towards CAR adoption by stating that many reporters still maintained resistance and hesitancy and “the laggards outnumbered the innovators” (p. 102).

Maier found that mapping and statistical software were the least used CAR applications in newsrooms because of the high technical complexity. This finding is consistent with the diffusions theory principle that complexity is a key attribute that negatively affects the adoption rate (Rodgers, 2003). The theory stresses the communication feature of innovation adoption, meaning that the flow of information is more crucial than utilization of new tools. In Maier’s research, the survey result showed that peer communications and in-house training were ranked the top two motivations of CAR diffusion.

At the time when Gynnild (2013) conducted research on the innovation brought by data journalism, the media industry had fully recognized the salience of data journalism. The innovation indicated in Gynnild’s research is a set of more complicated
computational skills to manipulate larger databases, such as using algorithms and programming new online tools. Though not directly framed by the diffusions theory, Gynnild borrowed some key concepts of innovations to analyze the impact of innovators and how data journalism can open the door to a wider range of innovations in the journalism industry.

Similar to Maier, Gynnild emphasized the importance of software on the innovation over the adoption of technological tools and skills. Gynnild claimed that innovations in the data journalism field primarily depend on computational thinking, meaning the capabilities to solve problems, abstract overloaded information, and decompose complexity. By analyzing the production of several successful data journalism projects, Gynnild gave credit to the innovators who are more open-minded and curious towards new ideas, risks, and challenges.

**Literature Review**

The 2016 Oscar-winning movie Spotlight depicted a team of journalists from the *Boston Globe* investigating the cases of child sex abuse by Roman Catholic priests in the Boston area. It showed old-school computer-assisted investigative reporting to a global audience. Over the past decades, this genre of journalism has become more prominent than ever due to a significant amount of databases available to the public and the growing open-source movement. The movement primarily focused on technologies. Its core concept is that information and knowledge should be free. The open-source software applications offer newsroom of different sizes the equal opportunities to create quality content. Free tools online help medium and small newsroom visualize large datasets as efficiently as big news organizations (Washington, 2013).
The media industry has seen a clear division in recent years between traditional computer-assisted reporters and the emerging group of data journalists. Early CAR adopters treat data analysis as an empirical means that can provide more depth and context to storytelling. The data journalists innovated the CAR practice by integrating journalism and programming. They are influenced by the open-source movement and treat journalism as structured and quantified information. The split in NICAR conference attendees between seasoned CAR journalists who usually tend to be older looking at data and the younger group who are more tech-savvy (Fink & Anderson, 2015) reflects such division. A more extreme view proposed by Adrian Holovaty, formerly at The Washington Post, almost subverted the conventional definition of journalism by claiming that facts should be delivered to the public through statistical data instead of narration (Gynnild, 2013).

Despite the ongoing debate, there is no doubt that the development of data journalism has transformed how journalists work. Coddington (2015) points out that the traditions of journalism, building around textual and visual elements, lead U.S. journalists to downplay the importance of numbers, making it notoriously difficult to present numerical data accurately. Tim Berners-Lee, the founding father of the World Wide Web, claims that journalists should “no longer find tips by chatting with people in smoky bars,” instead, they have to “equip with the tools to analyze data” in order to “help people out by really seeing where it all fits together, and what’s going on in the country” (Arthur, 2010).
By analyzing scholars on data journalism, this review will attempt to see how emergent data-driven journalism, particularly the hybrid journalists, have transformed the skill set of journalists and journalism practice.

Clarify the terms

As data journalism has been widely practiced in news organizations, more academic literature has examined the professional practices of data journalists, the newsroom infrastructures that help to connect programmers and journalists, and the development of data journalism (Fink & Anderson, 2015). The majority of data journalism-oriented scholars found it challenging to define data journalism and use a consistent term to refer to this type of journalism.

Early research (Ciotta, 1996; DeFleur and Lucinda, 1993; Garrison, 2001) frequently used “computer-assisted reporting” to refer to journalism that was mostly driven by numerical information. In the 1990s, CAR was a unified concept. As more scholarly work emerged in recent years, Coddington (2015) concluded that a couple of ambiguous related terms were used by the researchers, including computational journalism, open-source journalism, and data journalism.

Closely associated with precision journalism founded by Phillip Meyer, the emergence of CAR was mainly due to the development of social science studies and the introduction of personal computers in the 1980s and 1990s. Both Coddington (2015) and Felle (2016) summarized that CAR mainly served as an investigative tool to aid in long-term, public-affairs journalism projects. CAR is composed of a set of practices: searching the internet for information, sending the e-mail to communicate, requesting data from government agencies and analyzing the data by adopting “social science-based statistical
methods” (Coddington, 2015, p. 334). When more powerful computers and software were applied to reporting and investigating, the practices of CAR also expanded.

Several scholars (Du, 2014; Coddington, 2015; Felle, 2016; Craig, Ketterer & Yousuf, 2017) agreed on the notion that data journalism evolved from CAR because of the development of online journalism. Unlike old-school CAR practices, data journalism offers tools for investigation and story-telling, but as Coddington (2015) stated, it was less closely tied to investigative journalism than was CAR. Comparatively, the most significant feature of data journalism is audience engagement. It uses “interactive maps and graphics and applications on news website that allows for audience choice to tailor stories” (Felle, 2016, p. 91). Du (2014) defined data journalism as a combination of reporting and programming of data, featuring interactivity, statistics, a multi-modular approach, and audience participation.

The development of computational journalism comes along with the prevalence of the open-source movement and programmers, computer scientists and data scientists seeking careers in the media industry. Comparing to CAR and data journalism, it is more computer technology-oriented. Computational journalism centers on applying computational thinking to journalistic process. Wing (2006) defined computational thinking as the human brain’s ability to abstract and decompose data to solve complex tasks, which is also the fundamental concept of computer science. The technological innovations apply to computational journalism include machine learning, artificial intelligence, content analysis, visualization, as well as information science. For example, the *Los Angeles Times*’ 2015 project on LAPD crime classification used machine learning technology to textually analyze more than 400,000 crime records (see Exhibit 1). The
programming scripts (see Exhibit 2) consisted of functions to clean up the crime
descriptions and reclassify records by looping through each word. The data analysis
found that LAPD misclassified an estimated 14,000 serious assaults as minor offenses.

Exhibit 1. The sample data of LAPD crime classification project

<table>
<thead>
<tr>
<th>ID</th>
<th>crime_code</th>
<th>narrative</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>120132594X</td>
<td>DO O1 ACCT W/ HOMICIDE UNDER ARREST FOR BM</td>
<td>Minor</td>
</tr>
<tr>
<td>3</td>
<td>120132595X</td>
<td>DO O1 ACCT W/ HOMICIDE UNDER ARREST FOR BM</td>
<td>Minor</td>
</tr>
<tr>
<td>4</td>
<td>120132596X</td>
<td>DO O1 ACCT W/ HOMICIDE UNDER ARREST FOR BM</td>
<td>Minor</td>
</tr>
<tr>
<td>5</td>
<td>120132597X</td>
<td>DO O1 ACCT W/ HOMICIDE UNDER ARREST FOR BM</td>
<td>Minor</td>
</tr>
<tr>
<td>6</td>
<td>120132598X</td>
<td>DO O1 ACCT W/ HOMICIDE UNDER ARREST FOR BM</td>
<td>Minor</td>
</tr>
</tbody>
</table>

Exhibit 2. The sample code of LAPD crime classification project

```python
def tokenize(description):
    ...# first lower case and strip leading/trailing whitespace
description = description.lower().strip()
    # kill the 'do-'s and any stray punctuation
description = description.replace('do-', '').replace('.', '').replace(',', '').replace('!', '')
    # make a list of words by splitting on whitespace
    words = description.split()
    # Make sure each 'word' is a real string / account for odd whitespace
    words = [i for i in words if i]
    # let's see if anything improves the accuracy
    bigrams = [i for i in words if STEMMED_BIGRAMS.get(i)]
    # nltk trainer expects data in a certain format
    return dict([(i, True) for i in bigrams])
```
Data journalism, as Coddington (2015) claimed, is a more broadly and ambiguously-defined concept, however, it includes more concepts and practices than CAR, and strives for a more journalistic end than computational journalism. Therefore, this project is going to use data journalism to refer to all practices involving journalistically analyzing and presenting quantitative data.

New skill set on demand

Giving a clear-cut definition of data journalists and their organizational roles is not a simple task. Several scholars selected qualitative interviews to research data journalism. Fink and Anderson (2014) used an open approach to identify data journalists: “either self-defining as data journalists or pointing us (the researchers) to those who fit their already preconceived definitions” (p. 469). The journalists they interviewed had job titles including “database editor,” “interactive news editor,” “infographic design editor,” and “computer-assisted reporting specialist.” These studies found that data journalists have widely-varied career paths and they all have very heterogeneous educational backgrounds. Many data journalists started their careers as beat reporters. They gradually picked up data skills as they were working on data-driven stories. The old-school CAR reporters believe that mastering basic data skills, such as searching data online, spreadsheet analyzing public records, and processing geographical information along with solid investigative skills are the key to quality reporting. Data journalism, a more
extended approach than CAR, requires journalists to create a story that can by supported
by numbers instead of sole narration and description (Fink & Anderson, 2014).

The development of data journalism leads to an expansion of journalists’ skillsets.
The most significant trend is the integration of journalism and computer science.
Programming skills (JavaScript, Python, PHP, and Ruby, etc.), which previously served
as means to buttress the operation of the media industry, now are a crucial part of the
journalism craft (Royal, 2010). Fink and Anderson (2014) found a more technological
trend in data journalism by looking at the changes of the NICAR conference, which has
been seeing more attendees each year who have web, information technology, and
graphics skills.

Parasie and Dagiral (2012) studied the growing number of technology-savvy
people hired by news organizations in recent years. These “programmer-journalists” are
teamied with other data journalists to design “news applications” and produce innovative
online content. As a contrast, there are "journalist-programmers," who shifted their
careers from writing stories into writing code in the newsrooms. The early journalist-
programmers included “Adrian Holovaty formerly, at The Washington Post, Aron
Pilhofer, formerly at The New York Times, Matt Waite, formerly at the St-Petersburg
Times and Ben Welsh, at the Los Angeles Times” (p. 860-861). They all had a traditional
journalism background and have moved into programming during their journalistic
careers. This group of journalists advocates that “journalism and programming should be
viewed as inseparable in the newsroom” (Parasie & Dagiral, 2012, p. 862). In a
qualitative study of data journalists at the New York Times, Royal (2010) concluded that
journalists, nowadays, without a decent understanding of modern storytelling and deliver
methods, are at risk of becoming irrelevant in the newsroom (Royal, 2010). A Pew Research study (Project for Excellence in Journalism, 2008) showed that 96 percent of respondents agreed on the necessity for journalists gaining computational skills, indicating the industry's acknowledgment of the power of computational tools to assist storytelling.

The Interactive News Technology department of The New York Times, among one of the earliest dedicated interactive data teams in the country, requires its staff to be proficient with both storytelling and coding. By observing in the newsroom and interviewing employees, Royal (2010) found that most people did not have an educational background in information technology, and the majority had working experience in journalism. Most described that they learned how to program on their own, mainly out of curiosity and interest. These hybrid-skilled data journalists bring a unique perspective to journalism.

There is no right answer for questions as what tools and programming languages are regularly used by data journalists. The reasons are two-fold. First of all, the technologies and tools available are rapidly evolving in the field. It is neither possible nor necessary for a data journalist to know them all. The choices of tools vary by person. What works best for one is not necessarily best for someone else. Secondly, the size of news organizations as well determines data journalists’ toolkits. In the large news outlets that have a dedicated data journalism team, people specialize in various sub-areas of data journalism. Data journalists in small and medium news outlets may often find themselves the lonely “unicorn” in the newsroom (Hermida & Young, 2017), creating data stories from A to Z all by themselves.
The new practices of data journalism

Journalists and programmers have distinctive professional norms and values. As a result, when computational skills are introduced into media practices, the established proposition of journalism will be challenged and transformed. Carrying out a complex data journalism project demands various skills and relies on teamwork of people from different fields.

By studying the operation of Hacks/Hackers networks, an international grassroots organization bringing journalists (“hacks”) and technologists (“hackers”) together, Lewis and Usher (2014) recognized that journalists and programmers shared a similar purpose, which is “to figure out how to use software to tell a story and do it in a more comprehensive way than maybe had been done previously” (p. 389). Difficulties of integration existed. For example, developers have long-term and problem-solving mindsets, whereas, journalists, particularly those working for the local newspaper, are more concerned with quick turnover stories. The study concluded that sufficient mutual understanding is the key to successful collaboration.

The New York Times’ award-winning investigative reporting project on contaminated drinking water is the result of successful hybridity and cross-disciplinary teamwork. The team consisted of investigative reporters, database editors, programmers, and multimedia producers. The project started with an investigation on using the relevant data and public records, but only leading to “a fractured collection of databases” (p. 720). The team used data mining, data scraping, and some other computational techniques to build its own database and published an online tool for the public to search data of every state and every city in the country (Gynnild, 2013). Gynnild concluded that this project
was an interesting case to examine because it shows the new requirements for data journalists: the ability to think like a computer scientist and work as a professional journalist, as well as being able to get along with a team of people with diverse skills.

The Homicide Report project at the *Los Angeles Times* is another manifestation of successful interdisciplinary teamwork. The project originally was carried out by reporters manually publishing posts on each violent death case in L.A. and it later went on hiatus for 14 months. The project relaunched as a data journalist and a programmer-journalist joined the team (Young & Hermida, 2015). This cross-disciplinary project emerged alongside the CAR expertise that is deeply rooted in the L.A. Times and was enhanced by the incorporation of computational technologies. Ken Schwencke, who wrote the algorithms, told the authors in an interview that the algorithms could write the first paragraph of a story on homicide, including basic facts. Journalists can expand the entry as long as the detailed information is available. Besides, the algorithms can detect abnormal levels of crimes and alert the newsroom. The algorithms replaced much of the manual labor that was required for a traditional reporter. As a result, each homicide case occurred in the city that could be covered, and the newspaper’s bias in reporting homicide cases could be corrected. Therefore, the "central tendency" and the outliers can both be included in the project.

As the searchable database becomes a common method, Young and Hermida (2015) pointed out that data journalism uplifted the transparency of journalism as it provides raw data and journalistic research methodology to the audience. Influenced by the open-source movement, the data journalism practitioners call for more engagement from the public by granting them with the access to the raw data. Gynnild (2013) claimed
that *The Guardian* was the early adopter of such practice. The newspaper opened its content API and Data Store to the public and encouraged them to filter and search the raw datasets to meet their personal goals. Holovaty even said that storytelling is no more important because as long as the public have access to databases, they can generate customized news by searching for whatever they need (Gynnild, 2013).

**Conclusion**

As more news organizations recognize the salience of data journalism, the academic field has seen a growing number of scholars on data-driven journalism. As innovation occurs, human practices are always ahead of naming the actions (Gynnild, 2013). Both the industry and academy are still defining the terms. Compared to computer-assisted reporting and computational journalism, data journalism includes more sub-categories of practices and concepts. Alongside the development of data journalism, a group of hybrid journalists appeared in the newsrooms. They are either programmers pursuing their career goals by practicing journalism, or journalists learning programming skills to be better storytellers. Such hybridity challenges the established norms of journalism. More news organizations incline to build a dedicated data team to have people from various fields. Meanwhile, under the influence of open-source movement, data journalism increases audience engagement by proving them with access to the raw data.

**Methodology**

As data journalism develops and the possibility for digital news presentation expands every year, the demand for hybrid data journalists is on the rise. Programmers are participating in the editorial process in the newsroom and journalists are moving into
programming. The journalist-programmers, who have no educational background in computer technologies and used to practice journalism with pens and notepads, have picked up programming skills to interrogate data and reconstruct static text stories into dynamic, interactive online content. Many scholars explored the development of data journalism and the innovations along with it. Only a small number of them tracked the emergence of hybrid journalists, including studies (Royal, 2010; Parasie & Dagiral, 2012) on the programmer-journalists of *Chicago Tribune* and the interactive news department at *The New York Times*. These studies analyzed this phenomenon from an institutional level, and the subjects were mainly programmers having positions in newsrooms. This project aims to take a closer look at how programming techniques as an innovation are adopted by journalists by conducting several semi-structured interviews with data journalists who took a journey from journalists to coders.

Personal accounts are critical to social research because of the power of language to illuminate meaning (Legard, Keegan & Ward, 2003). Qualitative interviews gather data such as personal accounts in a naturalistic and interactive manner. The research topic of this study centers on a particular group of people and their accounts. Qualitative interviews suit such a purpose the best since they provide rich information on the people who are being studied through normal human interaction. It is a kind of “spontaneous exchange” that “offers possibilities of freedom and flexibility for the researchers and narrators alike.” (Perks & Thomson, 1998, p. 197)

Both unstructured and standardized interviews have pros and cons. The unstructured interview has no specifically prepared questions. The flow of information relies on the subject’s responses (Ryan et al., 2009). The open conversation helps to build
better rapport. This allows the researcher the opportunity to probe for information that does not seem relevant before the interview (Corbin & Morse, 2003). However, this type of interview is very time consuming and increases the opportunity for gathering biased data. On the contrary, the standardized interviews aim to ensure that each interview is conducted with the same questions in the same order. This method is commonly employed by quantitative research in which data are easier to code and analyze (Doody & Noonan, 2013).

I choose semi-structured interviews to carry out this project because this format can guarantee the flexibility of the interview and the consistency of data collection. The interview consists of open-ended questions that are directed by a prepared guideline. It assumes that the subjects have different understandings on the research topic because of their individual perspectives and personal experiences (Doody & Noonan, 2013). In comparison to standardized interviews, the open-ended questions can maintain the open minds of the interviewees, allowing new ideas to be brought up during the process. Meanwhile, the interview guide can be handy in keeping the flow of conversation on track. I will be attentive during the interviews to ensure the data-collection process progresses in as organized a way as possible.

Since this project is going to explore how journalists adopted computational skills and made a transition into journalist-programmers, the candidate interviewees should be now working at large to medium news organizations and whose jobs include a significant portion of coding and using advanced data skills. Their job titles may include but not be limited to, data reporter, data editor, data analyst, graphics reporter, graphics editor, digital producer or data specialist. The interviewees should be journalists who started
their careers as reporters or editors working for traditional journalism sectors, or had no educational background on computer technologies.

For interviewing a specific and narrow scope of subjects such as for this type of research, Guest et al. (2006) found that data saturation started to occur within twelve interviews, although the basic elements for the research topic were evident in as early as six. The number of subjects I plan to interview is six to twelve, but the final number will depend on when my data start to show a noticeable amount of overlay and repetition.

The main challenge of this research is two-fold. First, the specificity of this research topic narrows the potential subjects to a niche group of journalists. This may make it difficult to reach enough interviewees. The second challenge is accessibility. This research is heavily dependent on participants’ time availability. Journalists are always under intensive deadlines and juggling multiple tasks. It may be challenging for the researcher to have in-depth and time-consuming interviews with them.

As the country's biggest network for data journalism practitioners, IRE and NICAR is the best starting point to seek out potential interviewees. Constantly, there are reporters posting questions on programming on the NICAR email list and conversations about programming and reporting. By browsing the email list entries and searching for more information about active email-listers, I can spot some journalists whose jobs require them to know how to code and reach out to them.

I will start contacting interviewees this summer when I am interning at the Investigative Reporting Workshop. I hope my position there will help me identify several eligible interviewees. I also hope I can get some referrals while working with many
experienced data journalists at the IRW. Also, I may seek assistance from my committee chair, Professor David Herzog, for recommendations and guidelines.

Each interview has to be carried out via face-to-face conversation or Skype video call and last between 30 to 60 minutes. The face-to-face is a more desirable interview method because an in-person interview is going to capture both verbal and non-verbal ques. The body languages of the interviewees can indicate a level of their enthusiasm for the topic. This interview method is only applicable to interviewees located in Washington, D.C.

To operationalize the interviews, questions will be placed into three specific groups. Group one will be questions on demographic information and their careers before they began coding. Group two will focus on the internal elements of the interviewees, such as motivations. Group three will be questions on external factors shaping interviewees’ transitions, such as encouragement from newsrooms and challenges.

**Interview Guide**

- What did you study when you were in college?
- How many years have you been working in the news industry?
- What was your journalism experience before becoming a journalist-programmer?
- When did you realize that coding skills were getting important to your job?
- When did you transfer your focus on data journalism?
- How long did it take for you to recognize the need to code to actually learning it?
• Which programming languages do you often use?
• How did you balance learning new skills and completing daily work?
• How did you keep your programming skills sharp and up to date?
• What programs or tutorials are most helpful for you to learn coding?
• What opportunities did your news organization provide you to learn more coding?
• How does coding skills elevate your journalism?
• What were the challenges for your transition?
• What were the hurdles holding you back from learning new programming skills?
References


Zhang, C. (2017). CAR through the ages [PowerPoint slides]. Retrieved from https://docs.google.com/presentation/d/1s2p8epTH572K2gXhoI3NFILUh5jZaC4BscmHhCIfE3c/edit#slide=id.p