INVESTIGATING INEQUALITY USING

COMPUTER-ASSISTED REPORTING

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

ANDREW KREIGHBAUM

David Herzog, Project Supervisor

MAY 2016
ACKNOWLEDGMENTS

I would like to thank the members of my master’s committee for their guidance and patience in the process of completing this professional project: David Herzog, Mark Horvit, Sarah Diem and Barbara Cochran. I first met Herzog and Horvit as an attendee at Investigative Reporters and Editors boot camps. I took away invaluable lessons from their courses at Missouri that inform how I think about approaching both data analysis and investigations. As the rare journalism student in Diem’s politics of education course, I benefitted greatly from opportunities to reflect on and discuss the kinds of policy issues that formed the backdrop to many of the stories I pursued as an education reporter before coming to Columbia. Cochran’s guidance as the director of the Missouri Washington program was crucial to navigating this final semester in the graduate program.

I must also thank Liz Lucas, the director of the NICAR Database Library, whose offer to become a graduate research assistant at the library made my decision to enroll in the master’s program at Missouri a no-brainer. Spending three semesters as a staff member at the database library was as valuable as any course I took on campus for my development as a data journalist.
TABLE OF CONTENTS

ACKNOWLEDGMENTS........................................................................................................ ii

Chapter

1. INTRODUCTION...........................................................................................................1
2. FIELD NOTES............................................................................................................. 3
3. EVALUATION................................................................................................................ 18
4. PHYSICAL EVIDENCE............................................................................................. 20
5. ANALYSIS................................................................................................................ 22

APPENDIX

ORIGINAL PROPOSAL.................................................................................................... 38
Chapter One: Introduction

I entered the journalism graduate program at the University of Missouri in 2014 with two basic motivations — I wanted to become much more proficient practicing data journalism and I wanted to eventually pursue investigative reporting in some form. No journalism program has a stronger reputation than Mizzou for training reporters in either field. But there’s really no substitute for practicing journalism full-time, especially in an organization where every project is a partnership with an outlet itself known for strong watchdog reporting. That’s why it was so exciting to join the Investigative Reporting Workshop as a graduate research assistant in my final semester as a Missouri graduate student.

I was unfortunately unable to spend time working with the Workshop’s data editor, David Donald, this semester because of an illness that kept him away from the office for an extended period. But in my time at the organization, I was able to observe and participate in long-term investigations involving multiple journalists. Combined with the data journalism skills I took from my time as a graduate student and as a graduate assistant at the NICAR Database Library, I’m hopeful that those experiences will make me a more attractive candidate for hiring editors whether they’re looking to add an enterprise or beat reporter.

The projects the Workshop pursued in partnership with local NPR affiliate WAMU and the Washington Post involved very different topics. WAMU approached the Workshop about investigating the political influence of local power companies and public utilities, while the Post looked into the release of officer names after fatal
police shootings. But, like any serious investigative reporting effort, these projects scrutinized the actions of powerful institutions and attempted to understand how they fit within broad trends. I played a different role in each of these projects but they both involved in-depth reporting using documents, data and interviews.

The Post investigation, “Fatal Force,” kicked off the second year of the paper’s ongoing coverage of fatal police shootings. At the heart of much of that coverage has been the issue of inequality — black men have died at disproportionate rates in encounters with police. The professional analysis I pursued in my final semester had much to do with this theme. Protest movements like Black Lives Matter have indisputably shaped the national conversation about use of force by police. And the concerns of those protesters have been reflected in recent investigations by news organizations like the Post and the Guardian.

As a graduate student with a focus on data journalism, I was curious about how journalists are using computer-assisted reporting to cover inequality in a wide range of topics. The analysis project below hasn’t focused solely on reporting about racial discrimination. It’s instead an attempt to account for how newsrooms are using data journalism to expose systemic inequality in an era when the protest movement perhaps best symbolized by Ferguson has put the topic on the national agenda. My research for the analysis and my time at the Workshop have provided me with valuable insights into the investigative process and I look forward to applying those lessons to my work as a full-time journalist after leaving the master’s program.
Chapter Two: Field Notes

Week 1 (January 11-15)

A lot of this week was spent completing new employee tasks, including trips to the Human Resources office or to a separate office for a staff portrait. Finding an office routine has been fairly easy, especially with a recent Mizzou grad, Josh Benson, here. One of the biggest challenges this week has been catching up on the work the Investigative Reporting Workshop has already completed on a few projects I’ve been involved in. The workshop has spent almost two years on a project about the border and the political activities of anti-immigration groups who have helped to shape the national debate on the issue. I spent my first couple days this past week researching the lobbying expenditures over the last decade and a half of two of these groups and completed a memo for Lynne Perri, the workshop’s managing editor.

Previous interns at the workshop have made trips to the National Security Archives to review documents, so I spent the second half of the week reviewing these documents to identify papers that might have renewed relevance. The workshop is apparently discussing some kind of collaboration with a Texas newspaper for this border/immigration project. My hunch is that the Trump presidential campaign as well as a renewed surge of migrant children from Central America will mean there is still plenty of interest in this topic. One challenge I’m finding with this project is that several people have contributed reporting over the last two years, so I’m trying to figure out what original contributions I can make to
the project. I think examining the Congressional lobbying of anti-immigration groups was a good start and we’ll have to see where the project develops from there.

I haven’t had the chance to get involved in any serious data reporting projects yet. But Lynne and I will meet this Thursday with Patrick Madden, a reporter with the local NPR affiliate WAMU. He has a couple projects that he would potentially like to collaborate on with the workshop, including an investigation of political influence in public utilities. From what Lynne and David Donald, the workshop’s data editor, have described so far, it sounds like there could be a heavy data element to that project. My work on the analysis portion of the project so far has consisted of reaching out to possible interview subjects. I’m hoping to start lining up interviews for the end of this month or early February.

*Week 2 (January 18-22)*

I’m writing with an update on my second, abbreviated week at the Investigative Reporting Workshop and my progress on the analysis portion of my master’s project. I spent Tuesday and Wednesday this past week finishing up some research for the border/immigration project the Workshop has been pursuing for the past couple years. Lynne Perri, the Workshop’s managing editor, and I may make a trip next week to the National Security Archives at George Washington University to examine more documents related to anti-immigrant groups like the Federation for American Immigration Reform.

More immediately, I’m working on a project involving money and political influence by political utilities. This is a project we’re collaborating on with Patrick
Madden, a reporter at local NPR affiliate WAMU. Lynne and I met with Patrick on Thursday to go over what he’s hoping to do with this project. I’m starting by gathering campaign finance and lobbying data in the DC/Maryland/Virginia area. It’s been a little bit of a challenge navigating the websites of these state ethics commissions. I’m hopeful that after a few more days to become familiar with those tools, finding data relevant to this project will be a lot easier.

Regarding the analysis project, I completed my first interview this week with Paul Kiel of ProPublica. We spoke for almost 40 minutes, mostly about his work on The Color of Debt (link here), a project that revealed debt collection suits are far more common in black communities than white ones. That pattern has a lot to do with the wealth gap between black and white Americans but those lawsuits are making it impossible for people of color to work themselves out of debt. Paul actually learned a lot of data journalism skills — including SQL code — while working on the project. He said the debt collection data gathered by ProPublica for areas like Chicago; Essex County, New Jersey; and St. Louis County made it pretty obvious there was a racial element to the pattern of the lawsuits. But Kiel and his team at ProPublica assembled a sort of panel of academic experts to consult with on their initial findings to point them towards concepts like the wealth gap to put that data in context.

Students from the Washington program also met this week with Peter Hart, a Democratic pollster. While I like to think I have a pretty good handle on the limitations of polls, it was really useful to talk through some of the pitfalls of reporting specifically on the latest NBC/Wall Street Journal presidential primary
poll. At the very least, our meeting with Hart I think will help me explain some key concepts involving polls if I’m called on to do so in the future.

Week 3 (January 25-29)

I’m writing with an update on yet another shortened work week, thanks to Winter Storm Jonas. The AU campus and our office were closed on Monday and Tuesday thanks to icy road conditions but I got a fair amount accomplished on our joint project with WAMU on the political influence of public utilities. I started by looking into the political activities of power companies in Maryland and DC. They involve both campaign finance and lobbying as well as “revolving door” relationships at the local government level.

Exelon and Pepco, the companies involved in a proposed merger utility merger affecting DC area residents, gave more than $100,000 to state and local candidates’ election committees in Maryland over the last two years. But Exelon by itself spent about six times that lobbying state officials in 2015 -- more than any other firm in the state. Lobbying is where most of the political activities of the power companies have been focused. It’s been frustrating to discover how limited the disclosures of industry lobbyists are at the state level, even though they are required to submit numerous reports to the Maryland Ethics Commission. Most activity reports indicate only that a lobbyist was advocating on an issue such as “Any and all matter relating to electric generation and energy matters.”

But I did manage to dig up records of a lobbyist for Baltimore Gas and Electric, one of Exelon’s subsidiaries, hosting the Senate Finance Committee and House Economic Matters Committee at restaurants like Ruth’s Chris Steakhouse.
Members of those committees were also amongst the biggest recipients of campaign contributions from utility PACs. So, I think we'll be able to show some interesting connections in this project.

Week 4 (February 1-5)

Most of my work on the research project this week consisted of lining up further interviews. I’ve made plans to speak with Michael Berens (formerly of the Seattle Times, now of the Chicago Tribune) about using data to report on the Methadone and Politics of Pain series for the Seattle Times.

The highlight of the past week at work was a meeting with Patrick Madden at the offices of WAMU to discuss the progress we’ve made on the project involving political influence by public utilities in the DC/Maryland/Virginia area. I presented the data I’d gathered on lobbying expenditures, PAC donations and charitable contributions to local nonprofits by public utilities. I was hoping we could figure out one angle or area to focus on for further reporting but the project seemed to get more unwieldy as the meeting went on. Madden wanted to figure out what we could quantify in terms of industry influence in each state. So I used some of my time Friday and Monday making sure we had complete data on PAC contributions to state candidates and committees in Maryland and Virginia, as well as DC.

I also set up calls with Common Cause Maryland on Monday as well as two public policy professors at the University of Maryland to discuss how the lobbying process works and how we should be interpreting the expenditure numbers for Exelon, Baltimore Gas and Electric and other power companies. Our biggest challenge at this point appears to be explaining how the utilities are influencing the
levers of government – other than listing their total expenditures – and explaining why that matters to residents and ratepayers. After speaking with lobbying experts this week, my next set of interviews will likely be with residents of the district who are opposed to the proposed Exelon-PEPCO merger.

As part of the seminar last week, we met with two members of Bloomberg’s White House reporting team. Bloomberg isn’t necessarily a publication I would aim to work for but it was interesting hearing strategies from reporters on finding sources for stories the administration wouldn’t want covered. I’m hoping that insight can come in useful for the project I’m working on now with WAMU.

Week 5 (February 8-12) / Week 6 (February 15-19)

Since I didn’t get around to sending an update last Monday, I’m including notes about what I’ve been up to the last couple weeks. I’ve made some pretty significant progress on the project we’re working on with WAMU as well as completing interviews for the analysis project.

Last week, I interviewed Kendall Taggart of Buzzfeed, Mike Berens of the Chicago Tribune and Michael LaForgia of the Tampa Bay Times about using data journalism to cover issues of inequality. I’ve also got an interview coming up this week with Kimbriell Kelly at the Washington Post. After speaking with her, I’ll have five total interviews (Paul Kiel from ProPublica was the first). I’m planning to get a couple more interviews in first but I’ll also be at the NICAR conference in Denver March 11-13. So that should provide an opportunity to discuss the project with some more folks in the field and possibly arrange additional interviews.
The project on the political influence of public utilities moved from gathering data on campaign finance and lobbying expenditures to conducting interviews with activists and consumer advocate groups like Public Citizen. Reports from Maryland’s State Ethics Commission showed that Exelon – the power generator pursuing a merger with DC’s local public utility – spent more on lobbying in the state last year than any other entity. But Patrick Madden (the WAMU reporter leading this project) and I knew we had to get a better understanding of the connection between those expenditures and the decisions of regulatory bodies like the Maryland Public Service Commission, one of several panels to rule on the merger. We knew that a 47 percent run-up in lobbying expenditures by Exelon and its subsidiaries would jump out at readers but needed to be able to explain the significance.

Tyson Slocum, Public Citizen’s Energy Program director, explained that outreach to elected officials at the state and local level can help sway the process if those officials publicly endorse the agreement. The commission is an independent body but keeps an eye on which way the political winds are blowing, Slocum said.

My interviews with Slocum as well as Anya Schoolman, a grassroots activist who has led local opposition to the Exelon/Pepco merger, and Mike Tidwell, executive director of the Chesapeake Climate Action Network, have really helped to illuminate some of the “big picture” themes of this project. One of the common themes from these interviews was that political money isn’t just a feature of the merger application process, it’s central to the process. They also pointed to an “astroturfing” campaign that’s included public support for the deal from local nonprofits who receive support from Pepco.
Opponents of the merger have also frequently compared the companies’ influence in DC and Maryland to that of Dominion Power in Virginia, where that utility is the biggest campaign donor in state politics. In a meeting at the Workshop last week with Lynne Perri (the workshop’s managing editor), we decided to run a story within the next week about Dominion’s application for new coal ash disposal permits in Virginia that will incorporate some of my reporting on campaign finance in that state. That will set the stage for reporting later on about the influence Exelon has exerted in the Maryland and DC markets. I’m hoping by the time I provide my next update at the end of this week, I’ll have a link to a story to send or at least a run date for the first story on Dominion.

*Week 7 (February 22-26)*

I continued to make progress on the master’s project this week. I also found time to scratch out some more work on the Workshop's collaboration with WAMU while getting roped into an investigative project with the Washington Post.

On Wednesday, I interviewed Kimbriell Kelly from the Washington Post for my analysis project. We talked quite a bit about her contributions to the paper’s *Dashed Dreams* series last year. But some of her best answers actually referenced her experiences covering the foreclosure crisis at the Chicago Reporter before arriving at the Post. She echoed Kendall Taggart from Buzzfeed, who suggested getting in touch with Corey Johnson of the Marshall Project and Topher Sanders of ProPublica. Kimbriell also strongly recommended reaching out to Ron Nixon of the New York Times.
Earlier in the week, myself and fellow Mizzou grad student Karol Ilagan met with John Sullivan of the Washington Post about the paper’s next investigation into police shootings. John splits his time between the Post and the Investigative Reporting Workshop and he’s leading this effort along with another Mizzou alum, Steven Rich. The Post team has built an internal interface for the project that lists each shooting victim by police department. Karol and myself, along with several American University grad students, were assigned multiple states and (where we could) identified officers involved in police shootings using Google searches or LexisNexis. On Monday, we’ll be heading to the Post to meet with John and his team about the project. It sounds like the next stage will involve calling departments where names of officers have not yet been disclosed, followed -- hopefully -- by the chance to contribute data analysis.

While I wasn’t working on the Post database, I continued to conduct interviews for the collaboration with WAMU about the influence of power companies, specifically focusing on the merger of Exelon and DC utility Pepco. I spoke with a Sierra Club attorney challenging the merger in a Maryland appeals court, a lobbyist for a Maryland environmental group, a city council member who was a leader in the opposition to the merger in that state and two commissioners from DC neighborhood advisory councils who have spoken out against the merger in DC. (The merger was rejected a second time Friday by the DC Public Service Commission but isn’t completely dead.) What was notable in these interviews was how much more reserved the Maryland sources were in describing the influence of
the power companies, compared to the sources in DC. That may be indicative of the differences in how Exelon and Pepco advocated for the merger in those locations.

During our seminar meeting Friday, we heard from Mike McCurry, the former Bill Clinton press secretary. I really enjoyed the exchange the class got to have with McCurry and thought he conveyed a real appreciation of the American public’s disillusionment with not just politics but the most prominent institutions in society. I don’t think we agree on the best choice for the next president but I thought he was very thoughtful and really appreciated his perspective.

Week 8 (February 29-March 4)

Karol Ilagan and myself began this week at the Washington Post, where we met with John Sullivan and other members of the paper's investigations team about the police shootings project we will be helping out with for the next several weeks. The paper has set up an online interface for this project that will allow us to enter names of officers for each case as we find them. The time it took to set up the internal interface for this project is one of the reasons we are just now making it to the paper to discuss the project. We spent much of this week reaching out to departments from states assigned to us by Steven Rich, a database editor at the paper and another Mizzou alum.

Week 9 (March 7-11)

After two months in DC, I can report that the first story I've contributed reporting to this semester has seen publication. It’s the first story in what WAMU/the Investigative Reporting Workshop are calling an “occasional series” on the political influence of public utilities in the DC/Maryland/Virginia area.
Additional stories about the long-sought Exelon/Pepco merger will, hopefully, follow soon. I’ve contributed a lot more to the reporting efforts on political influence in DC and Maryland.

Most recently for that side of the story, I began looking into the role of shadow lobbyists — basically advocates on policy outcomes who don’t register as lobbyists — in the efforts to advance the merger application. Before leaving for the NICAR conference in Denver, I interviewed John Wonderlich, Sunlight Foundation interim executive director, about the growth of shadow lobbying at the federal level and whether the phenomenon is filtering down to state and local politics as well. Part of the story on Exelon/Pepco will focus on how much the Chicago-based power generator is spending on the merger process and how murky the reporting is for a good chunk of that money.

Most of the last two weeks have actually been devoted to the Washington Post’s investigation into police shootings. Myself and fellow Workshop staffer Karol Ilagan trekked to the Post’s new offices again at the beginning of last week to review our progress tracking down the names of officers involved in police shootings in California, Michigan, Ohio, Missouri and other states. The story on disclosure practices of police departments is set to run March 20 no matter what, so we’re working hard this week to track down names or get denials from the remaining unresponsive departments in the Post’s database.

On the analysis project front, I’ve got an interview set up this week with Talia Buford of the Center for Public Integrity to talk about her coverage of environmental inequality. Attending NICAR also helped generate some new ideas. The first panel
session I was able to attend was a talk about Investigating Inequality. I was able to touch base with Dallas Morning News reporter Holly Hacker afterward to schedule an interview for the project in the next couple weeks. Her talk focused on a project the Morning News passed on running about how the college system fails to serve poor students adequately. It eventually ran in the Huffington Post in December.

Week 10 (March 14-18)

This past week was probably the last one where I'll be doing any heavy lifting on records requests for the Washington Post before the paper's next story on police shootings is published. "The Unnamed" will look at the various approaches of police departments across the country to naming officers involved in fatal police shootings. On my end, that involved a lot of calls/emails to around 40-50 police departments or sheriff's offices where officer names weren't already publicly available over the last couple weeks. As of last week, I only had a handful of departments remaining where a Freedom of Information request was still pending. Most had released the names and a handful of departments had made denials. It looks like the story, which is the first in a series of planned police reporting projects this year, should run later this week.

I can't say I had more than a small part on this project but it was somewhat gratifying to contribute to greater public transparency on the issue and it was fun being around the Post newsroom for a bit the last few weeks. I'm not sure if there is time left in this semester to become more involved on other investigations at the paper. However, I met with Patrick Madden of WAMU this week along with the Workshop's Lynne Perri. They asked me to start a story outline for our next piece on
the political influence of power companies in the DMV area. I think this next week after we see what we have, we’ll get a better idea of what gaps still remain in the reporting.

On Tuesday, I also interviewed Talia Buford from the Center for Public Integrity about her work on the Environmental Justice, Denied project. Talia really began to learn a lot of her data skills by working on this project. But I thought it was one of the best interviews I’ve had in terms of her responses on how reporters cover inequality and how they can make those stories accessible or reporters.

Week 11 (March 21-25)

I started this week by circling back to the immigration project I spent the first couple weeks of the semester working on. Monday morning, I met Workshop Managing Editor Lynne Perri at the National Security Archives at George Washington University to review additional documents from the archives of the Federation for American Immigration Reform. That afternoon, I went back to the Post for another meeting on the paper’s police shootings investigation for another meeting to check in our progress touching base with various police departments across the country. After initially extending the print deadline, editors have said it will go no later than April 3. So much of this week was spent making any remaining follow-up calls and fact checking the responses we had already recorded.

Week 12 (March 28-April 1)

At the end of this week another project I’ve been involved in was published. The Washington Post investigation into the disclosure of officer names from police
shootings went live on the paper’s website Friday and ran later in the Sunday edition.

It looks like the contributions of fellow Mizzou grad student Karol Ilagan and myself at the Post will be wrapped up after this past week. It’s gratifying to see the story we worked for the past several weeks finally come together. But I think we’re also both ready to focus on the separate projects in progress with the Center for Investigative Reporting and WAMU respectively.

I met again last week with Patrick Madden from WAMU and Lynne Perri, the Workshop’s managing editor, on our story about the political influence of power companies in the DC area. I completed an outline of about 2,500 words of the first draft of our story. The outline sketches out the largest objections to the merger of Exelon Energy Corporation and DC utility Pepco as well as the efforts to advance the deal through campaign finance, lobbying, revolving door relationships and other methods. The outline has helped us identify areas where significant reporting still needs to be completed.

Patrick and I are focused right now on finding voices from the DC community of residents who would be especially hurt by potential rate increases on their utility bill and individuals who supported the merger before it was approved at the end of last month but have changed their minds. I feel like our reporting right now is missing good analysis of what the merger is going to mean for the low-income DC residents, older residents and residents of color. I’m hoping to do a lot more reporting on that angle this week after touching base with some of the most vocal activists on the issue.
Week 13 (April 4-8)

I spoke this week with Dallas Morning News education reporter Holly Hacker, my final interview for the analysis project. It took a few weeks to catch up with her after we spoke at the NICAR conference in Denver. But we were able to talk about her project for the Huffington Post about the low numbers of low-income students served by elite colleges and universities who are increasingly focused on prestige rankings. The most impressive finding in the story Hacker produced with the Hechinger Report’s Jon Marcus was that low-income students are actually subsidizing the college experience of affluent peers thanks to a new emphasis on awarding financial aid based on academic merit and not need. Hacker said she and Marcus used data published online through the National Center for Education Statistics. Their familiarity with this data from covering the education beat helped them make sense of it for this project. Hacker and Marcus weren’t dealing with an entirely new set of data like many of the reporters I interviewed for this project.

I also continued reporting this week for our collaboration with WAMU on the Exelon-Pepco merger. The most notable interview this week was with an expert on utility mergers who has testified on a number of these cases before state regulators as well as the DC Public Service Commission for the Pepco case. He made the argument that the influence and lobbying campaign run by the power companies outside of the PSC’s administrative process was less distressing than the lack of vision displayed by the commissioners overseeing the process.
Chapter Three: Evaluation

During a lunch with the editors, reporters and graduate students who contributed to the Fatal Force project at the Washington Post, the predominant theme of the discussion was the power of working in a team to contribute to a successful project like that investigation, which had run the week before. This lesson seemed to me both useful and limited in application for many newsrooms smaller than the Post.

Most of the professional work I’ve done while in grad school has involved some sort of collaboration — whether I was working with other staff members at Investigative Reporters and Editors on data analysis for a member organization, gathering information on new NCAA cost-of-attendance scholarships while at an internship for USA Today, partnering with WAMU’s Patrick Madden on coverage of power companies or working with the team of journalists on the Fatal Force project. In some respects, getting my master’s has provided me with as much experience working in teams as working with data, my original reason for going back to school. So assessing my work as a graduate researcher at the Investigative Reporting Workshop should account for that fact.

Although I wish that I’d been involved in the Post investigation to an extent that I could have contributed more original reporting, I think I did a fairly effective job of coordinating with other graduate students who were tasked with getting law enforcement agencies to release officer names — or get a denial. This wasn’t an enterprise that involved a lot of handholding from editors or reporters at the paper.
but we managed to get an impressive amount of work done in a matter of weeks. Being able to record a response or denial from those police departments by the end of March meant the paper was able to have the data needed to report one out of five officers involved in police shootings went unnamed last year.

My role with the WAMU partnership at the Workshop involved a different kind of collaborative process. Reporter Patrick Madden came to the Workshop with the original idea for the piece based on his ongoing reporting on the proposed merger of nuclear giant Exelon with Pepco, DC’s local public utility. But I played a bigger part in shaping the direction of the reporting — including identifying potential interview subjects and data sources — and in writing the text of the story itself. Because Patrick often was caught up in daily reporting for the station while this project was underway, I was often left to my own devices on the story for days or weeks at a time. I would have sometimes preferred to be in the same newsroom to do that reporting — instead of a shuttle ride and then another metro stop away — but those circumstances forced me to become a more active communicator about the progress I was making. That’s something I can bring to future collaborations, hopefully in a more systematic and comprehensive way.
Chapter Four: Physical Evidence

Because all of my work this semester was of a collaborative nature and on long-term projects, only two examples of published work are available as of the writing of this report. On those projects, I contributed reporting by pulling data, making calls and filing open records requests. But since my contributions were less substantial than other reporters involved, I received a contributing credit and not a full byline. The first was a story for local public radio station WAMU about the political influence of Dominion Virginia Power by focusing on the power company’s application to state regulators to dump coal ash waste into a tributary of the Potomac River. The second project was a Washington Post investigation of police departments’ transparency on officer-involved shootings that found in one of every five police shooting cases from 2015, the officer’s name goes undisclosed. A team of eight reporters contributed to the project by contacting police departments across the country about nearly 1,000 fatal police shootings, filing open records requests on many of those cases and interviewing law enforcement experts and families of police shooting victims.

The project on which I’ve had the most substantial contributions this spring—another collaboration with WAMU’s Patrick Madden—has yet to run. Because I’ve played a large role in both the reporting and writing of the piece so far, it’s anticipated that I’ll share a byline on the story, which is expected to run online on the websites of both WAMU and the Investigative Reporting Workshop. The
reporting for this story has consisted of interviews and research using documents and data obtained from local ethics commissions.

I also hope to find an outlet or outlets for publication of the research analysis below after it is submitted for fulfillment of the master’s degree, whether that’s on the Investigative Reporting Workshop website, in the IRE Journal or in another industry publication. I think there are valuable insights in the comments from the journalists I’ve interviewed and I hope that my analysis could enlighten other reporters approaching projects with similar concerns.
Chapter Five: Analysis

If newsrooms had not been jolted into a new awareness of inequality by the protests over the death of Trayvon Martin, the issue had become impossible to ignore after events in Ferguson in 2015. Anger over the police shooting of Michael Brown and the handling of the subsequent investigation sparked a protest movement based in the St. Louis suburb. Police violence and officer-involved shootings have received more serious and sustained coverage than ever from national outlets, such as the Washington Post and the Guardian.1 Meanwhile, local media outlets like the St. Louis Post-Dispatch have taken a renewed interest in the discrepancies between the overwhelmingly white local police forces and the largely black communities they patrol.

But some of the best opportunities for reporting went far beyond law enforcement issues. More specifically, they showed the extent to which Michael Brown and black teenagers like him grew up in St. Louis suburbs starkly different from their white counterparts. Or how fines assessed to the area’s poorest residents became a major driver of municipal revenues. Telling the story of inequality isn’t a project that can wait until the next crisis or round of protests. And data journalism is an essential tool for reporters who are engaged in that effort.

Journalists like Phil Meyer have pursued groundbreaking reporting on inequality as far back as 1967. Meyer, then of the Detroit Free Press, used

1 http://www.cjr.org/analysis/counting_how_many_people_are_killed_by_police_requir es_more_than_arithmetic.php
quantitative methods to report on the underlying causes of the Detroit riots and found that residents who attended college were as likely to have taken part in the riots as high school dropouts. In 1989, the Atlanta Journal-Constitution's Bill Dedman used data analysis to reveal that the city's banking and savings and loan industries were denying to middle class and affluent African Americans the kinds of loans extended to even the poorest whites.

Using data to cover inequality isn't a novel pursuit. But recent events in Ferguson and elsewhere have helped to put a new focus on these issues. This analysis will explore how data journalism can elevate the coverage of disparities using several recent examples of reporting projects. It is based on interviews with seven practitioners who used data analysis to provide novel reporting on problems of inequality in their local communities or from a national perspective on their beats.

“You’re seeing a lot more journalists just putting a lot more emphasis on the data and proof behind it,” said Talia Buford, a reporter for the Center for Public Integrity’s environment and labor team. “It's helpful especially with some of these issues that either resonate, or originate, or predominately affect communities of color.”

Having the data to back up reporting or illustrate a problem on a wider scale makes it impossible to dismiss the complaints of people of color as just anecdotes, Buford said. Her reporting with CPI colleague Kristen Lombardi on the Environmental Protection Agency’s Office of Civil Rights showed that the
department had never made a formal finding of a civil-rights violation after reviewing 300 complaints from minority communities over two decades.

Journalists interviewed for this report agreed that to the extent that the problems of minority communities or other marginalized groups are underappreciated or too easily dismissed by the typical reader – or even the typical editor. Having a command of the data gives a reporter more authority to tell that story.

“I think more and more newsrooms have discovered data journalism puts you in control of the story to some degree,” said Chicago Tribune reporter Michael Berens.

His 2011 reporting along with Ken Armstrong for the Seattle Times on the overprescribing of methadone to low-income Medicaid patients showed how they were suffering and even dying because of a decision by doctors and state regulators to steer them toward the drug as a cheaper solution for pain management.

For the purposes of this analysis, I define an investigation of inequality as reporting that examines the causes and consequences of unequal conditions affecting a particular class or group of people. Frequently those stories are concerned with inequality of class, race or gender but often they can involve all of these. This kind of reporting can deal with problems arising from failures of particular institutions or from wider social problems.

Several of the journalists interviewed – Berens, the Tampa Bay Times’ Michael LaForgia and Buzzfeed’s Kendall Taggart – reported stories on specific decisions or missteps by local or state government agencies that led to negative
effects for minority and low-income people. Buford reported on the failures of the EPA’s Office of Civil Rights to serve the interests of minority communities due to mismanagement and turmoil within the organization. Holly Hacker of the Dallas Morning News revealed how low-income students are actually subsidizing more affluent college peers while elite private and public schools across the country are placing a low priority on serving students who won’t contribute to prestige rankings of the institutions.

Kimbriell Kelly of the Washington Post and Paul Kiel of ProPublica reported stories about issues of inequality affecting minority communities on a local level with causes originating from complicated societal and economic factors. Kelly’s reporting on the effects of the foreclosure crisis in Prince George’s County, Maryland, revealed how even affluent African American neighborhoods were entangled in the subprime loan market at disproportionate rates. Kiel’s coverage of debt collection lawsuits showed how those suits were concentrated in African American communities partly as a result of a wealth gap going back generations. Whether these stories involved the failures of a single government agency or the kinds of broad patterns identified by Kelly and Kiel, they used data-driven reporting to reveal unequal conditions and explain their consequences.

Among the key findings of this analysis:

1) These journalists see great data journalism arising from strong beat reporting or familiarity with a subject area

2) The first key to a data-driven investigation is the ability to quantify a problem
3) Data analysis confers additional authority on an investigation, allowing a reporter to “bulletproof” their findings from criticism or false stereotypes.

4) Incorporating characters who are representative of the data is critical to successful narrative storytelling.

5) Finding a “smoking gun” for a problem takes a back seat to identifying and explaining broader trends or problems.

6) While data can’t make the same visual or emotional impact as video, it can help engage news audiences in news on the issues of inequality.

*The data reporting process*

Buford had worked as a beat reporter covering criminal justice at the Providence Journal before starting at the Center for Public Integrity in July 2014. However, she came to the non-profit news organization without any significant experience using computer-assisted reporting methods for newsgathering. The “Environmental Justice, Denied” project she pursued with Lombardi gave her the impetus to attend the NICAR conference in 2015 and enroll in a computer-assisted reporting boot camp. The first steps in the project involved creating a database in Excel using information pulled from a large set of PDF files of complaints obtained through a Freedom of Information Act request. Buford said she began writing SQL code for the first time in subsequent reporting.

The journalists investigating inequality who were interviewed here have markedly different skill levels or expertise in data journalism tools. But a common
theme from the interviews suggested reporters often develop skills or learn these tools because of a specific story or project they want to pursue.

Whether or not reporters had received a tip on a story or simply had a hunch about a potential trend, they started by seeking to quantify an issue. Buzzfeed’s Taggart said she has attempted to find stories by requesting a public database and looking for interesting patterns. But she said she has had more success starting with a story tip and attempting to quantify a problem by looking for what data she could find.

“It’s much easier, I’ve found, basically to hear about a bad anecdote and then say, ‘I wonder how often that happens?’ And then you’ll find a dataset that allows you to quantify that,” she said.

The website’s reporting on indigent Texas teenagers jailed for failing to pay fines for truancy arose out of conversations with juvenile attorneys and other advocates after she read stories about fines levied against low-income residents of Ferguson and adjacent municipalities. Taggart and her colleague, reporter Alex Campbell, requested about 40 jail databases from Texas counties before opting to focus on data gathered from the 10 largest counties. Their analysis showed that more than 1,000 teenagers had been sent to jail in the state in three years over failures to pay fines.

Taggart said it was writing practice queries in SQLite on the El Paso County data for their truancy investigation that led to a striking discovery – driving without a license was the most common charge for individuals sitting in that county’s jail. That led the reporters to request data from the state’s case outcome reporting
system through the Texas Office of Court Administration. When Taggart and Campbell looked at courts with the highest rates of cases resolved with jail credit, El Paso “kind of popped out,” Taggart said. That finding led to their second investigation of inequality in the state’s court system.

The Post’s Kelly begins her story process with an extensive review of what published material or databases are available having to do with a specific story topic.

“I try to find the best data source on whatever it is I’m interested in,” she said. “We have a responsibility to find whatever the best source that is out there.”

Kelly seeks out stories already written on a subject through sources such as the Investigative Reporters and Editors website and LexisNexis. She also looks for data sources by contacting government agencies and sometimes the private vendors. She said working at a paper like the Washington Post has made obtaining data from private sources easier in those cases.

“I go to great lengths to make sure I’m using the best possible data. Most data is not perfect, so you try to get the best of what’s out there,” she said.

Hacker said her reporting on low-income representation at elite higher ed institutions along with the Hechinger Report’s Jon Marcus grew out of a previous collaboration after the U.S. Department of Education began to require colleges and universities to report the full cost of attending their institutions by income bracket.

“It was kind of an offshoot of a story we’d already done,” she said.

Their previous reporting and familiarity with the data available from the National Center for Education Statistics meant they weren't just starting from
scratch on a brand new project but instead providing follow-up coverage to an issue on their beat.

Berens said one of the primary elements of any of his initial investigation is determining what can be quantified on a particular problem. Quantifying the number of deaths resulting from methadone use based on a state coroner’s database was the “secret ingredient” of the Seattle Times project, he said.

At ProPublica, Kiel's investigation of debt collection lawsuits revealed how African American neighborhoods were disproportionately impacted by lawsuits that were perpetuating a cycle of poverty. Seeking to answer some basic questions about debt collection led Kiel to request case level data from 20 courts in 11 different states.

“We suspected that we would find something but we didn’t know what we would find,” he said. “There was an idea that there was some kind of inequality going on but we didn’t know what form it would take. We didn’t have a good idea that assets would be the main driver.”

*Reporting objectives: Forgetting the smoking gun*

In the process of reporting trends of inequality, journalists can sometimes become stymied looking for a way to pin the cause of unequal conditions on the intentions or decision-making of specific actors. In a 2015 article in the IRE journal exploring how journalists can better cover inequality, New York Times Magazine reporter Nikole Hannah-Jones warned against becoming preoccupied with finding a “smoking gun” in their investigations. Establishing intent is less important than showing that policymakers’ decisions have adversely impacted marginalized
communities and that those consequences could have and should have been foreseen, she wrote. Hannah-Jones’ admonishment was made as much to encourage reporters not to be dissuaded from covering topics like inequality as anything else.

Conversations with data journalists indicate a common theme on such objectives — none of the journalists interviewed were more concerned with placing blame on a specific public official or body than they were with constructing an accurate picture of the devastating consequences of the policies that create conditions of inequality in schools, neighborhoods or even health outcomes.

Berens of the Chicago Tribune rejected outright the importance of uncovering a “smoking gun” in the context of a data-driven investigation.

“It’s not critical,” he said. “It’s not necessary. I think the data is your smoking gun. That, to me, is the beauty of database journalism.”

Buford said if she and Lombardi had gone into their investigation of the EPA looking for a smoking gun, they wouldn’t have found one.

“If you’re only looking for the smoking gun, you’re going to miss a lot of the nuanced problems,” she said.

In their investigation, Center for Public Integrity obtained a 2011 Deloitte Consulting report commissioned by EPA showing the agency was aware of the failure to adequately adjudicate complaints made under Title VI of the Civil Rights Act. But the reporters discovered no emails or directives in their investigation that indicated the agency’s leaders had decided not enforce or prioritize the complaints. The story became about what happens to people when an agency that is supposed to
serve minority communities can’t get its act together and enforce a bedrock civil rights law, she said.

In the Tampa Bay Times’ Failure Factories investigation, there was likewise no “smoking gun” in written communications or public statements by the Pinellas school board showing an intent to re-segregate African American students in the district’s worst campuses. But the paper’s reporting demonstrated how decision makers should have anticipated that outcome — and worsening academic performance — when they voted to change school attendance zones. Investigative reporter Michael LaForgia said readers often fall into the trap of believing that large societal problems are inevitable and immutable when that’s never true. The paper’s reporting focused on how the school board twice failed students who otherwise could have succeeded academically: the first time, with a plan to change school zone boundary lines that predictably re-segregated campuses; the second, by failing to take any of the measures pursued by districts elsewhere in the state to remediate the lagging performance of black students.

Revealing new forms of inequality

The killings of Trayvon Martin and Michael Brown were seen by only a handful of witnesses or just the shooter himself. But the new prevalence of smartphones on city streets has led to a wave of recordings of police interactions with black men that led to violent confrontations and often deaths. Hannah-Jones wrote in the IRE Journal that the advent of those videos as an object of mass consumption has created a tipping point in overwhelmingly white communities and
even largely white newsrooms, whose journalists did not previously appreciate the extent of the problem of police violence.

In the same way, data-driven reporting, done properly, can help demonstrate to white audiences how critical issues of inequality and injustice are—even when they themselves are not directly affected.

“I think that’s what doing data stories can illuminate and, yes, (they) can help engage others who may not be part of that struggle of being discriminated against,” Kelly said.

Those stories can create an understanding and interest among readers in a topic they might not have understood before. LaForgia of the Tampa Bay Times said the trick for reporters covering inequality is using the data right.

“It’s hard to look at a cop wrongfully shooting a black guy and come away with anything but (the sense) that this is a terrible injustice. But that’s not true for data,” he said. “Data can mislead, it can be boring, it can be put in the wrong context. It can drown out the message.”

Journalists interviewed for this project described a number of approaches and strategies to ensure that the data in their reporting is working in service of the goal of exposing inequality:

1) Present data in a straightforward, honest and accurate way

2) Bulletproof your analysis so that those being held accountable cannot undermine your findings
3) Use characters in a story who are most representative of the trends in the data and whose struggles could not be pinned on factors other than the problem identified in the story.

4) Tell the stories of people and communities affected in a way that is accessible and relatable to readers.

LaForgia said the most important step a reporter working with data can take is to be honest and direct with readers about what the data is saying and then present it in the most effective way possible. Before the “Failure Factors” series ran in the Tampa Bay Times, his team was most concerned with countering the notion offered by skeptics or public officials that the performance of the majority black Pinellas schools was a reflection of the students and their parents. As LaForgia put it, the attitude was essentially “Oh, black kids have shitty parents.’ So your kids get out of education what you put into it as a parent, right?”

“So we took a data heavy approach,” he said. “We approached this almost exclusively through data on a mission to sort of bulletproof the story from that criticism and I think we did a hell of a job.”

Part of that bulletproofing effort involved the presentation of students and families affected by the re-segregation of Pinellas campuses who were representative of the findings of the investigation and who would speak to the largest possible audience.

“We didn’t want people that even the most affluent reader could come across and then dismiss as someone who’s making their own problems, for example,” LaForgia said.
In producing “Environmental Justice, Denied,” Buford said the CPI team took pains to include stories of community members from states or subdivisions most frequently appearing in their database. They also made sure to select for individual anecdotes within a story only those people whose complaints went unresolved because of the failures of the EPA — not the errors of the complainants themselves.

Buford and Lombardi interviewed members of multiple communities represented in the complaints to the agency before deciding they were not strong enough illustrations of failures in the EPA process.

These standards used by the Center for Public Integrity reporters help make sense of the story the writers are trying to tell using the data. Fundamentally, the reporters are seeking to shape a story using the data, not simply present enough numbers and statistics to bowl over the reader.

“The data isn’t necessarily the story,” said the Post’s Kelly. “The story is what the data means.”

The Post’s analysis of foreclosure rates in Prince George’s County was just the starting point for an investigation into how the mortgage crisis affected African-American families neighborhood by neighborhood. Prince George’s was the wealthiest majority black county in the country. But Fairwood, a booming subdivision in the county before the housing crash, had the second-highest foreclosure rate of any neighborhood in the country with more than 100 loans.

Kelly’s reporting showed than even affluent black neighborhoods were disproportionately affected by the crisis, thanks to a high prevalence of subprime loans. Her deep dive into the story of one family in the Fairwood neighborhood of
Prince George’s revealed how many West African immigrants in particular found themselves caught up in the foreclosure crisis.

For Berens, the collection of individual stories that arose from reporting out the data offered up a chance to educate the audience on the reality of methadone use and push back against perceptions that victims of the drug were just addicts looking to score a high. Many readers associate methadone with heroin users and Berens and Armstrong were well aware of those perceptions. But the reporters found that patients who died from complications related to methadone use were steered toward the drug at a higher rate if they were insured through Medicaid.

In one striking example, a King County 911 dispatcher named Angeline Burrell developed undiagnosed chronic pain after a gall bladder surgery in 2004. As she sought help for the condition, her own doctors began to suspect that she was a prescription addict scamming for drugs. In the two years it took doctors to diagnose surgical-related nerve damage, Burrell lost her job, her house and her private insurance. Among the drugs she was prescribed during that time period was methadone. Despite guidelines warning against prescribing methadone to patients taking anti-anxiety medications, doctors actually upped her dosage of the drug. Eventually, her prescriptions created a “toxic cocktail” that paralyzed her respiratory muscles.

“We tried to show here’s a person that could have been you or me,” he said of Burrell. “Here’s a woman who had a home, children, a common-law husband and was progressing the way every person wants to and through no fault of her own gets switched to methadone.”
Assembling the data allowed the Seattle Times reporters to show that the victims of the state’s methadone policy weren’t just pill seekers or people abusing drugs and alcohol. Berens and Armstrong found case after case where victims of chronic pain suffered negative consequences from being prescribed the drug as a cheap alternative. The concentration of these prescriptions among low-income patients revealed by the data analysis made the story effective.

ProPublica’s Kiel said without helping readers understand the scale of inequality, troublesome trends can be reduced to anecdote or negative stereotypes about the individual people affected. Reporting for ProPublica and This American Life last year, his conversations with St. Louis County residents showed how the sense of shame associated with debt collection lawsuits kept residents affected from talking to neighbors or even family members who had experienced the same circumstance.

“People tend to make character judgments about people who fall into debt. Maybe people are less likely to do that,” he said if reporters can effectively use data journalism to show the truth.

Conclusion

Interviews with the journalists featured in this analysis demonstrate a common set of understandings of the potential of data journalism to power reporting on inequality. Foremost among them is that any investigation of a problem of inequality begins with attempting to quantify that problem. Finding a pattern — showing through an analysis of the data that unequal conditions exist — allows a reporter to describe a situation with authority and get past the constraints
of "he said, she said" stories. The reporter herself can describe unequal patterns in
tschools, the justice system or a housing market and tackle an explanation of the
reasons why.

The interviews conducted for this project showed that journalists must think
as deeply about how they explain the significance of the data as they do about the
analysis itself. A common refrain was that data alone cannot tell a reporter’s story.
Kendall Taggart of Buzzfeed even questioned whether data journalism could ever
have the kind of impact of a video showing an unprovoked killing by police.
Numbers alone can’t tell a story; reporters have to develop strategies to give them
narrative weight for readers. Reporters described how they attempt to account for
those challenges by presenting data clearly but also finding the most compelling and
representative stories of individuals in the communities affected by unequal
systems. The narrative impact of those stories combined with the strength of
comprehensive analysis of wider trends gives these stories extra power to expose
these issues.
Appendix: Original Proposal

Introduction

When I enrolled at the University of Missouri’s journalism graduate program in 2014, I had decided that data journalism was the best way to make inroads in a career as an investigative reporter – at the very least, I thought developing those skills would make me more employable in a newsroom. I previously spent more than three and a half years at small to midsize daily newspapers covering beats like local government and public education. Rather than make another move to a large metro daily, I opted to focus on building data reporting skills at Missouri.

In three semesters at the program, studying data journalism has shaped not just my tool kit as a reporter but also my general thinking about how to approach news coverage. Coursework in programs like Excel, MySQL and Python provided a foundation of skills, while other classes have pushed me to think about how to use these programs to add depth to news coverage. I’ve come to see the context provided by data analysis as essential to understanding what might otherwise be compelling narrative stories. Data reporting can also provide a more nuanced view of how incidents like the killing of Michael Brown in Ferguson, Missouri fit within larger patterns of police violence and black criminalization. Philip Meyer, in many respects the originator of data journalism, used computer-assisted reporting to explore the social and economic backdrop to the Detroit Riots of 1967. From the beginning, the field has been concerned with issues of social equity and inequality. I believe exploring these origins of data journalism practices can help reorient us as a
profession away from a preoccupation with privileged voices and toward a news framework that is concerned with balanced reporting on society, including the voices of the least privileged.

My readings in the emerging scholarship on the field show that more than ever data reporting is taking a variety of forms in professional newsrooms. I remain interested on a personal level in using the data journalism skills I’ve acquired at Missouri for investigative reporting in a narrative format. I’m interested in exploring how those tools are shaping the approach of reporters who continue to tell stories with text-based narratives toward issues of inequality. The growth in accountability reporting from digital news outlets like Buzzfeed, The Intercept and Al Jazeera illustrates the extent to which narrative reporting remains relevant in the new online media landscape. But data reporting abilities have often been essential to the most impactful reporting of news organizations like these (Truancy), as well as many legacy print media outlets (Failure Factories). This project will examine how data journalism is shaping the approach of both current investigative and beat reporters who tell narrative stories. But in acknowledging the origins of the field, it will also explore how data journalism is producing reporting on inequality. It will place these findings in context by providing an overview of the academic literature on the field.

**Professional skills component**

I will pursue the professional component of my project at the Investigative Reporting Workshop in Washington, D.C., where I will intern under David Donald
during the spring 2016 semester. The internship will involve 30 hours per week at the Workshop. I will spend the remaining hours each week on the research component of this project. The Investigative Reporting Workshop is a nonprofit newsroom based at American University specializing in government and corporate accountability. The Workshop has a focus on finding underappreciated angles on topics widely covered. For example, an investigation released this year looked at how infrequently the families of police shooting victims received payment after seeking, and winning, compensation in civil courts. I will contribute literature reviews for investigations, data analysis and occasional reporting. The experience at the workshop will provide me with hands-on experience observing how data analysis and research methods shape long-term news investigations.

**Theoretical Framework**

The research component of this project will explore how data journalism is shaping reporting practices and priorities in U.S. newsrooms – specifically how data reporting is leading journalists to explore the kind of social issues that Meyer pioneered through precision journalism. Publications like *Columbia Journalism Review*, *American Journalism Review* or the IRE Journal may be interested in running this analysis.

Reporters have practiced data journalism for decades but the academic research was fairly limited until recent years, when acknowledging the value of data-driven reporting became mainstream. By 2012, data analysts like Nate Silver were becoming notable even outside the industry (McGregor, 2013). The growth in
scholarship may not reflect that development as much as a shared interest in
computer science thinking by academics and the newest practitioners of data
journalism. (Parasie, 2015, p. 861) Many organizations have at the very least begun
paying lip service to the value of data skills while still struggling to resolve how data
journalism tools should change how reporters approach their jobs at a fundamental
level. Much of the scholarship explores the different strains of quantitative
journalism as a whole and how recent developments reflect the values of either the
“open source” movement of collaborative programming or the social sciences.
(Parasie & Dagiral, 2013, p. 861) Other research, while often drawing on this
scholarship, reflects the challenges in incorporating data journalism practices into
newsrooms that have been traditionally organized around qualitative
newsgathering techniques. (Young & Hermida, 2015, p. 382)

Explorations of these practical challenges illustrate the extent to which
journalists act as gatekeepers of information in their decisions to use data
journalism tools in reporting and presenting stories. Even as data journalism creates
new possibilities for telling stories, journalists also shape those quantitative
methods to meet traditional journalistic requirements.

This research is timely because of the role data has played in the
construction of newer accountability news models as well as the decision by many
traditional newsrooms to re-orient themselves as “digital” organizations. In many
cases this shift in strategy also entails an acknowledgment of how data journalism
tools can make a newspaper a more compelling digital product. This topic will fit in
with my plans for the professional component of the master's project as I am
seeking a position as a data journalist or enterprise reporter at a news organization. In that role, I will encounter many of the same choices working journalists make involving data.

*Gatekeeping theory*

Journalists, even in everyday circumstances, act as gatekeepers when they decide what information the public needs to know and what is superfluous. Those decisions may take place without serious debate in a newsroom or even serious consideration on a conscious level by an individual reporter. Shoemaker writes that gatekeeping is “the process by which the vast array of potential news messages are winnowed, shaped, and prodded into those few that are actually transmitted by the news media.” The theory encompasses more than just the messages selected but also accounts for how they are handled and presented to news consumers. (2001, p. 233)

The first major study of gatekeeping effects was David Manning White's examination of how the decisions made by wire editors at local Midwest newspapers shaped the content of news coverage. (Ernste, 2014, p. 20) Gatekeeping takes various forms in the practical decisions journalists make –from the selection of story topics to the use of specific quotes by officials and members of the public to the filtering of photographic images. Researchers have found that gatekeeping decisions play a role in shaping visual framing of news events as well as the textual framing of events. (Fahmy, 2007, p. 556) Theorists have also observed both individual decisions and routine journalistic forces shaping gatekeeping effects on
the news. (Shoemaker, 2001, p. 235)

Journalists exercise the gatekeeper role in the kind of individual decisions mentioned above but they also act within institutional frameworks that shape those choices. The general mission of a news organization or incentives to reach more readers or land on the front page of a newspaper exercise gatekeeping influence in the same way as a reporter’s individual priorities for news coverage. There has always been a hierarchy of influences on media gatekeepers. As Shoemaker outlined it, this model included the individual level, the media routines level, the organization level, the extramedia level and the ideological level. (Ernste, p. 28)

Ernste builds on Shoemaker's analysis by offering a theory of “networked gatekeeping” that takes account of social networks’ affect on traditional gatekeepers. He rejects the idea that the medium of the Internet, which has allowed any user to consume and share nearly limitless amounts of information, has led to the death of gatekeeping. (p. 31) Rather, the monopoly of news organizations like daily newspapers as gatekeepers has weakened over time because of structural changes in the industry and new challenges from other voices. To say “everyone is a gatekeeper” now is to employ a shallow understanding of the theory that boils it down to any sort of filtering. (p. 33)

Ernste shows that the role of journalist as gatekeeper has undergone continual challenge over the decades, especially in recent years. Journalists once assumed professional authority because of the news monopoly possessed by a handful of organizations. That monopoly was challenged by politically oriented blogs and, later, by social media platforms like Twitter. Breaking news, political
news, and sports and celebrity coverage are available from myriad sources now—including direct communications from candidates, athletes and other public figures.

But the development of specialized data analysis skills in newsrooms suggests, alongside other new digital roles, suggests a persistent position for journalist as gatekeeper. It’s not just a role defined by journalists’ decisions to limit or filter information available to their audiences. Journalists in this gatekeeper role have also expanded the knowledge of their audience by making sense of vast amounts of information. Their skills push open the window of knowledge for their audience. Data journalism has redefined the reporter’s role as a gatekeeper of vital public information. Instead of presenting competing narratives from official sources or checking those claims against documents, journalists can present a comprehensive picture of policy or social issues suing statistical evidence.

The data journalism approach has had other practical effects in news organizations that have made it a practice to publish original documents, data sets and even code available to readers alongside text narratives. In other cases, newsrooms such as the Guardian have sought to actively involve their audience in analyzing documents or records. (Daniel, 2010, p. 4) The growing prominence of data as the raw material of storytelling means more critical reflection on how journalists’ handling of data serves the needs of news consumers.

Although the growth in scholarly output on data journalism reflects the new prominence of the field, the research hasn’t directly examined the practice through a gatekeeping perspective. Much of the scholarship actually examines how the open source technology perspective has shaped quantitative-oriented journalism, few
articles explicitly examine data reporting techniques through the lens of mass media theories like gatekeeping or framing. Even studies with the most practical focus, however, point to the extent to which data journalism tools shape gatekeeping decisions.

*Applying gatekeeping theory to data reporting*

Newsroom leaders have embraced data journalism as a tool for sharpening investigations (Nesmith, 2015). But many of the discussions at the industry level speak of the field in very general terms. The use of “data journalism” as an all-encompassing term fails to take into account for distinct strains of quantitative journalism. The differences in those practices are reflected in the various ways quantitative oriented journalism is used in U.S. newsrooms. They also reflect distinct gatekeeping roles. The scholarship reflects three broad strains of quantitative journalism that Coddington categorizes respectively as computer-assisted reporting, data journalism and computational journalism. To the extent that these strains inform the quantity and kind of information journalists incorporate into their storytelling, they affect the gatekeeper role.

Broadly speaking though, U.S. newsrooms are using data journalism tools to fit traditional journalism requirements for reporting on politics, government and society. This pattern suggests opportunities for expanding data journalism into traditional beat reporting as well as enterprise projects. Sylvain Parasie, in an analysis of a yearlong investigation of earthquake safety standards at California schools by the Center for Investigative Reporting, found a tension between
“hypothesis-driven” and “data-driven” perspectives. The former describes a traditional reporting process where tips from interviews and documents – traditional reporting tools – are followed up on and confirmed through data analysis (2015, p. 366); the latter describes a process where a lead is unnecessary and stories can be generated and explored through analysis of the dataset. The latter model assumes a certain level of completeness in the data. (p. 376)

Parasie’s findings illustrate how journalists continue to play the role of gatekeeper, first analyzing a database from the California Department of Education, and then checking it against separate two separate government databases. Finally, they reported the story using “traditional” methods involving documents and interviews at individual schools. This analysis will attempt to address how data journalism efforts like these are shaping the way newsroom are organized and reporters conceive of their roles. In sharing information with the public, do they see their audience as active participants in consuming the raw materials of reporting, or is the traditional role of media as gatekeeper intact?

**Literature Review**

The scholarship on the field of data journalism acknowledges that there are distinct understandings of how news organizations and reporters obtain and interpret information. Data journalism practitioners from the beginning have challenged established conceptions of the role of journalists. They posited that reporters should acknowledge the typical as much as the unusual. A single “man bites dog” story might be more interesting than a “dog bites man” story but a
hundred of the latter suggest a real public issue. Data journalism assumes a departure from the narrative priorities of most journalism and a focus on finding broad patterns in the aggregate of individual events.

The debate between what could be referred to as “narrative” journalists and data journalists involves both storytelling priorities and the way reporters conceptualize knowledge. Early adopters of computer-assisted reporting saw data as providing a more empirical view of society (Coddington, 2015, p. 333) More recently, a debate has developed within the data journalism field between traditional practitioners of computer-assisted reporting and developer-journalists whose thinking is influenced as much by the open source software movement as much as traditional journalism values and norms. Writ large, these developer-journalists or coder-journalists are less concerned with written narrative than data reporters who come from a traditional journalism background.

These kinds of ideological debates may rarely unfold in a newsroom in any explicit manner. But news organizations seeking to employ data in the service of innovation are confronting practical questions about how to incorporate specialized journalists in their newsroom structure. (Bell, 2012) Among those questions, editors are confronting whether data reporting can only realistically be incorporated into news organizations with significant investigative resources and knowledge or could data help to power daily beat reporting. Should data journalists be housed in projects teams or a specialized data desk? Should they be incorporated into the regular newsroom? These are practical concerns without an apparent theoretical stake but they point to developments in scholars’ understanding of the field. Studies
of contemporary newsrooms show computational reporting methods can expand the reporting power of news organizations. But these studies also show how competing ideological visions of data journalism shape practical newsroom judgments.

For the purposes of this master's project, I am interested in how these practical questions are informed by the theoretical underpinnings of the data journalism field. This literature review will examine the scholarship on computational reporting using two concepts as outlined in the theoretical framework: gatekeeping theory and the evolution of journalistic epistemology as reflected in debates over the role and specific uses of data journalism.

*Defining Data Journalism*

Before exploring how theoretical debates are reflected in newsroom organization, it is useful to discuss what exactly data journalism comprises. The literature uses a (sometimes confusing) variety of terms to refer to the field of data-driven reporting and its various subfields but I find it practical in this paper to refer to the field as a whole as “data journalism.” Early adopters referred to the practice as computer-assisted reporting, a reflection of how novel computer technology was when journalists first seriously attempted to use data for journalism informed by the scientific method. Coddington writes that as recently as the 1990s, CAR was still understood as a “single, unified concept.” But today, the field has split into three areas as reflected in the literature: computer-assisted reporting, data journalism and computational journalism. (2015, p. 332)
Computer-assisted reporting is most closely associated with the precision journalism developed by Meyer and intensive investigative reporting. (p. 334) Coddington describes data journalism as “more loosely coupled with investigative journalism than CAR” and encompassing any activity that uses data toward journalistic ends. Where CAR is almost synonymous with investigations, data journalism has become more integrated into newsroom practices. (p. 335) In one sense, the distinction between data journalism and CAR has as much to do with evolving newsroom practices as any fundamental differences. Computational journalism involves the application of processes like algorithms and automation for journalistic ends. In more general terms, it is influenced by computer science where CAR is derived from a social science tradition.

These classifications seek to make sense of the new developments in data-driven reporting, which in many ways barely resembles the original field. While elements of all three can be found in newsrooms — often simultaneously — these categories are reflective of how both the technologies and conceptions of data journalism have evolved.

These changes are best understood by placing data journalism in the context of the material “objects” that have been central to newsgathering, according to Anderson. (2015, p. 350) He finds that reporters arrived at the conception of “precision journalism” – Philip Meyer’s coinage for social science-driven reporting – after a number of fits and starts. (p. 358) Understandings of the role data and technology play in storytelling remain fluid. Thus, the “center of gravity” in debates over the value and utility of data is likely to continue to shift in the future.
Practical Newsroom Experiences

It’s notable how many case studies of news organizations have found, like Coddington’s, that data journalism emerged in a non-linear fashion in newsrooms. Instead, debates about practical concerns and reporter values shaped how data practices unfolded in these organizations. Journalists like those at the Center for Investigative Reporting have often encountered unexpected challenges in processing that shaped value judgments. In other cases, such as the Los Angeles Times’ Homicide Report blog, journalists with different skillsets and perspectives on the potential uses of data have shaped the same raw material toward remarkably different ends. These and other cases show how journalists’ views on how they attain knowledge about the world have shifted. Yet they also show how journalists continue to use judgment as gatekeepers to shape the information accessible to news consumers.

Parasie examined the role data has played in contemporary newsrooms in two separate studies. At the Center for Investigative Reporting, she examined how databases fit within existing enterprise reporting practices. Her study of the Chicago Tribune, meanwhile, illustrated how programmers with non-industry backgrounds shape journalism practices. Reporters at CIR obtained a database of California public schools’ compliance with seismic safety standards. (2015, p. 370) Efforts to create a more authoritative database by merging this original file with datasets from the California Geological Survey and California Department of Education led to a series of challenges for the reporting team. The messiness of the data – misspelled names,
inaccurate locations, etc. – revealed “a gap between the epistemological framework that the journalists involved had relied on, and the features of the investigation.” (p. 370) In other words, the data didn’t meet the standards of an objective assessment of reality. Reporters modified both their expectations and their journalist approach to account for this reality.

The resulting investigation involved both quantitative analysis of the data and qualitative analysis of the schools represented in the data. Reporters’ attempts to reconcile what was in the database with what the information they were gathering in interviews and documents led them to view the database as a “fragile construction” of real life. (p. 372) This conclusion implicitly rejected the concept of data as encapsulating a “hidden truth” that can’t be captured by synthesizing documents and interviews. (p. 373) Instead, the reporters used the data as a clue to indicate where they should be focusing more shoe-leather reporting. Data wasn’t the reporting method but one of several that CIR combined to produce “On Shaky Ground.” The project found that state regulators hadn’t properly enforced the state’s earthquake safety laws, putting California students at risk. Those reporting efforts were recognized with the top prize that year from Investigative Reporters and Editors. (p. 369) CIR reporters exercised a gatekeeper role in analyzing and fact checking the seismic inspection data. In a certain sense, the availability of those records widened the sphere of knowledge available to CIR’s readers. But the reporters also used judgment calls to decide what information was verifiable and useful.

Parasie argues that there are two paths to for reporters to arrive at justified
beliefs about their own knowledge: one where data helps reporters to confirm a lead and another where data is valuable in and of itself. The reporting process that produced the project reinforced the former view. But the contributions of programmer-journalists added at the Chicago Tribune in the mid to late-2000s serve as an ideological counterpoint to this conservative view of data. Journalism newcomers like programmer Brian Boyer, now with NPR, arrived with certain assumptions about how data-driven journalism can shape democracy. These industry outsiders claimed data would strengthen journalistic objectivity, provide new tools for accountability reporting and increase participation by citizens. (2012, p. 854) As Boyer stated, journalism “is sort of democratically enabling things to get better. It’s about informing people, voting, things like that, and democracy.” (p. 861)

In that respect, computational journalists had much in common with the originators of computer-assisted reporting. The overriding aim of data journalists is to improve democracy by expanding the knowledge available to citizens, so they can make more informed decisions. But computational journalists see the public as active consumers of the news. Parasie finds that that entails a rejection of an idea there is “a story hidden in the data.” (p. 862)

Young and Hermida found both of the perspectives described above at work in the Los Angeles Times’ Homicide Report project. There was no debate between those two camps as the project unfolded. Rather, the power and utility of data was understood differently as the project unfolded thanks to changes in personnel and developments in technology. Crime reporter Jill Leovy wanted to address bias in the paper’s reporting of homicide deaths in Los Angeles County by attempting to report
on each case for a full year. “There’s sort of an upside-down logic of press coverage in homicide where the nature of news is to cover the statistical fringe of homicide,” she told the authors in an interview. “You cover the very unlikely cases that don’t represent what’s happening most days in Los Angeles County.” (Young & Hermida, 2015, p. 386)

Leovy didn’t have the background of a data journalist but she saw the potential for data to correct for these kinds of biases and provide a fuller, more truthful picture of homicides in her city. She was eventually asked to hand over the blog to another reporter and it later went on hiatus for more than a year. When the project was re-launched in 2010 with the help of data journalists Ben Welsh and Ken Schwenke, it was based on an algorithm that replaced much of the labor previously contributed by a traditional reporter. In fact, the algorithm would write the first paragraph of a story about a homicide, including key details about the death. (Young & Hermida, p. 388) But journalists at the newspaper saw the algorithm as a tool to augment, not replace, the contributions of crime reporters. (p. 389)

A theme that’s apparent in several case studies of newsrooms is the extent to which many traditional roles of reporters remain unchanged despite the potential of data journalism tools to alter the way they gather information. The inspection database obtained by the Center for Investigative Reporting offered incredible potential for investigative stories. But members of the CIR team still found themselves using judgment and exercising traditional reportorial skills to determine the accuracy of individual data points and their usefulness to the investigation. At
the Los Angeles Times, data journalists who worked on the Homicide Watch project also wrote algorithms to “alert the newsroom to levels of crime that exceed the norm for an area in another neighborhood crime project, not the Homicide Report.” Reporters could then use qualitative methods like interviews to cover those patterns and why they may be occurring. (Young, p. 389) Data tools can provide “news tips” to reporters to point them in the direction of the story but their traditional roles in many ways remain fundamentally unchanged.

Broussard likewise finds that computational reporting methods can amplify a newsroom’s reporting capacity, especially investigative reporting. One of the biggest problems with newsroom investigations, she writes, is the reactive nature of many projects that results from the knowledge gaps of inexperienced reporters. Multiple theorists have observed the potential of data analysis to find investigative ideas. (2014, p. 2) Broussard describes how those concepts worked in practice by designing an algorithm to determine whether a lack of proper textbooks was hurting Philadelphia public school students’ academic performance. (p. 4) The data analysis conducted with the algorithm would fill her own “knowledge gap” concerning the Philadelphia school system for the Stacked Up project, which sought to address this question.

Most reporters have a narrative view of their field – their job is about the public interest and accountability but, above all, a good story. Broussard has a social science view of reporting, which may not be shaped by her exposure to data journalism but makes her more receptive to its potential uses. “Waiting for inspiration to strike” is not adequate for reporters to fulfill the demands of the job.
But the Stacked Up project demonstrates a method by which reporters can ask questions, seek answers in the data and find a story.

Conclusion

It's apparent from the literature that while data journalism has been practiced in newsrooms in various forms for decades; it's only recently that academics have begun to treat it as a serious object of study, rather than merely a specialized skillset. Scholarship reflects the development of three broad but distinct strains of quantitative journalism – computer-assisted reporting, data journalism and computational journalism – that are informed by social science on one hand and the open source technology movement on the other. Newspapers and other news organizations aren’t typically the forums to directly address these debates but the ideas reflected in these sub-categories of data journalism have shaped practical decisions in reporting and presenting stories. (TV journalism?) Taken together, recent case studies on data journalism in newsrooms suggest that journalists continue to exercise traditional roles as gatekeepers of information. Even when new technologies have replaced some of the labor traditionally provided by reporters, journalists for the most part still shape narratives and play an important role in reporting the meaning of data.

Methodology

The growing amount of qualitative research in recent years has added to our understanding of how data journalism is shaping both the practices and the
philosophies of newsroom. The most useful scholarship has illustrated by both by exploring the decisions of actual journalists using data to advance either investigative or beat reporting. This project will aim to build on the existing qualitative research by conducting several semi-structured interviews with journalists about how data journalism defined broadly has shaped the projects they have pursued and the mission of their newsroom.

Galleta writes that the semi-structured interview allows “for a narrative to unfold, while also including questions informed by theory.” (2012, p. 2) The format also allows interviewers to examine the context of a narrative. And context is essential to understanding how these reporters shape narratives out of data to address a complex topic like inequality. Data doesn’t necessarily come ready made to address the issue. I’m interested in hearing how journalists find such stories in the data or confirm story leads in data they’ve collected and analyzed.

My research will seek to answer the address the question: How do narrative reporters use data journalism to tell stories about social issues like inequality? I will also look to address whether a growing appreciation of data tools has led to a greater emphasis in reporting on systemic issues of inequality. I’m expecting to hear that inequality is not necessarily the highest priority for newsrooms doing data journalism, as opposed to topics like crime or politics. I also expect that journalism nonprofits or organizations devoted to investigative reporting may place a greater emphasis on reporting on inequality than more general interest newspapers or websites.
To get a range of experiences, this project will seek interviews from at least six journalists who have worked at a variety of news organizations. But I will focus these interviews on journalists who use data analysis to report on issues of inequality in a narrative medium. That means I won’t plan to reach out to news application developers on the one hand or narrative journalists who don’t conduct data analysis themselves on the other. I’m tailoring these interviews to journalists – especially investigative journalists – who both tell narrative stories and use data analysis for reporting both because this focus fits my own career path and because I think that focus will better reveal how data journalism can inform the decisions of reporters who cover inequality. To draw a distinction between investigative reporting versus narrative or feature reporting, I would consider the former to include any journalism that delves deeply into a topic of public interest. I would also consider investigative reporting as a form of journalism that uses a variety of tools such as interviews, documents or data, if not all of these. I also would say that investigative reporting usually examines systemic problems with policy, law or politics.

Among the journalists who would fit that description:

- Kendall Taggart, Buzzfeed
- Paul Kiel, ProPublica
- Walker Moskop, St. Louis Post-Dispatch
- Brad Heath, USA Today
- Kimbriell Kelly, Washington Post
- Michael LaForgia, Tampa Bay Times
These interviews will include the following questions about the subjects’ experiences using data to report on inequality:

- How does data analysis allow you to report on inequality in ways you wouldn’t otherwise?
- How do data journalism skills shape the way you conceive of or think about potential projects?
- How do you typically identify a data set for reporting?
- How do you relate findings of data analysis to issues like inequality without politicizing the data?
- What factors are limiting in terms of the types of stories you can pursue with data reporting? In your newsroom, is data seen as a tool to cover issues of inequality or are other topics like crime and elections prioritized?
Bibliography


things-every-journalist-should-know-about-data/.


