INVESTIGATING INEQUALITY USING COMPUTER-ASSISTED REPORTING

Andrew Kreighbaum

David Herzog, Professional Project Chair

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

¹ http://www.cjr.org/analysis/counting_how_many_people_are_killed_by_police_requ ires_more_than_arithmetic.php
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 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 that 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 schools, 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.