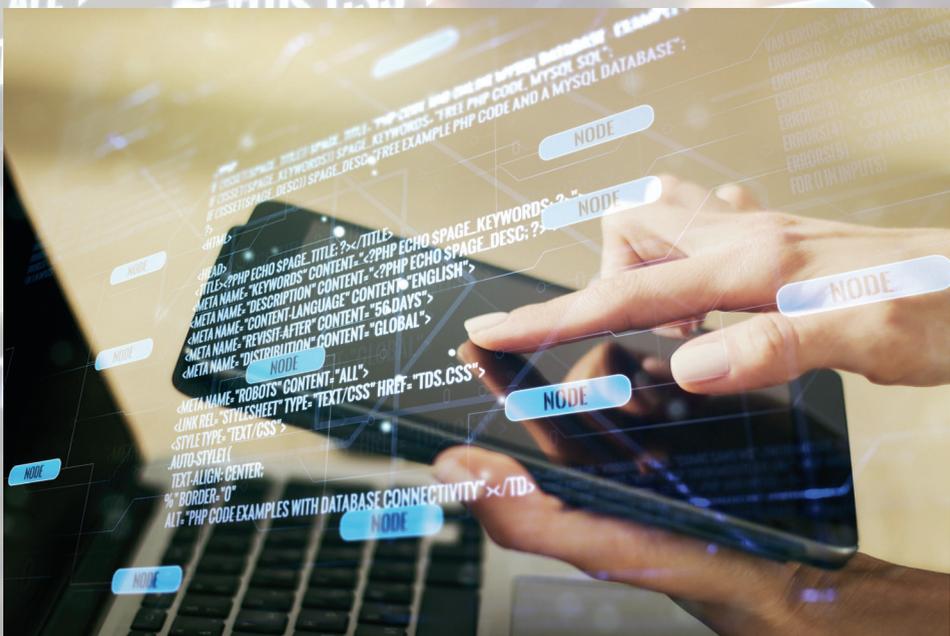


# BIG DATA AND JOURNALISM: HOW AMERICAN JOURNALISM IS ADOPTING THE USE OF BIG DATA

Andrew M. Clark  
Julián Rodríguez





# BIG DATA AND JOURNALISM: HOW AMERICAN JOURNALISM IS ADOPTING THE USE OF BIG DATA

ANDREW M. CLARK

JULIÁN RODRÍGUEZ

UNIVERSITY OF TEXAS AT ARLINGTON

---

## *Abstract*

This research uses in-depth interviews with three data journalists from the Houston Chronicle and the New York Times in the United States to describe the role of data journalists, and to illustrate how and why they use big data in their stories. Data journalists possess a unique set of skills, including the ability to find and gather data and use that data to tell a compelling written story in a visually coherent way. Results from our interviews and research show as newspapers move to a digital format, the role of data journalists is becoming more essential, as are laws like the Freedom of Information Act, which enable journalists to request and use data to inform the public and hold accountable those in power.

**Keywords:** Big data; Data Journalism; Reporting; Newspapers; FOIA, Freedom of Information Act

**Authors:** Dr. Andrew M. Clark is an Associate Professor in the Department of Communication at the University of Texas at Arlington. He is also the university's Quality Enhancement Plan Director and Associate Director for the Center for Research on Teaching and Learning Excellence. His research interests include international broadcasting, propaganda, public opinion, and broadcast history. He has published in journals such as the Journal of Radio and Audio Media, International Communication Gazette, American Journalism, and Digital Journalism. His teaching experience includes courses in broadcasting, communication law, and qualitative research methods.

Julián Rodríguez is a professor in the Department of Communication at the University of Texas at Arlington. Julián teaches broadcast journalism with an emphasis on Spanish-language media, researches U.S. Hispanic media and the adoption of media technologies for the development of systems awareness. Julian is the director of the UTA Hispanic Media Initiative, a program focused on advancing Spanish-language media education, journalism and research. For further information, please visit [www.utahispanicmedia.com](http://www.utahispanicmedia.com).

**Received:** July 24, 2020; **evaluated:** August 26, 2020; **accepted:** October 1, 2020.

---

# BIG DATA Y PERIODISMO: CÓMO EL PERIODISMO ESTADOUNIDENSE ESTÁ ADOPTANDO EL USO DE BIG DATA

ANDREW M. CLARK

JULIÁN RODRÍGUEZ

UNIVERSITY OF TEXAS AT ARLINGTON

---

## Resumen

Esta investigación utiliza entrevistas en profundidad con tres periodistas de datos de los periódicos estadounidenses el *Houston Chronical* y el *New York Times* para describir el papel de los periodistas de datos e ilustrar cómo y por qué usan *Big Data* en sus historias. Los periodistas de datos cuentan con un conjunto de habilidades único, lo que incluye la capacidad de encontrar y recopilar datos, para después escribir una historia impactante de forma visualmente coherente. Los resultados de nuestras entrevistas e investigaciones muestran que a medida que los periódicos pasan a un formato digital, el papel de los periodistas de datos es cada vez más esencial, al igual que leyes como la Ley de Libertad de Información, que permiten a los periodistas solicitar y utilizar datos para informar al público y hacer rendir cuentas a los que están en el poder.

**Palabras claves:** Big data; periodismo de datos, periodismo, periódicos, FOIA, Freedom of Information Act (Ley de Libertad de Información)

**Los autores:** El Dr. Andrew M. Clark es profesor asociado del Departamento de Comunicación de la Universidad de Texas en Arlington. También es el Director del Plan de Mejora de la Calidad de la universidad y Director Asociado del Centro de Investigación sobre la Excelencia en la Enseñanza y el Aprendizaje. Sus intereses de investigación incluyen la radiodifusión internacional, la propaganda, la opinión pública y la historia de la radiodifusión. Ha publicado en revistas como *Journal of Radio and Audio Media*, *International Communication Gazette*, *American Journalism* y *Digital Journalism*. Su experiencia docente incluye cursos de radiodifusión, derecho de la comunicación y métodos de investigación cualitativa.

Julián Rodríguez es profesor del Departamento de Comunicación de la University of Texas en Arlington. Julián enseña periodismo televisivo con énfasis en los medios de comunicación en español, investiga los medios hispanos estadounidenses y la adopción de tecnologías mediáticas para el desarrollo de sistemas de conciencia. Julián es el director del UTA Hispanic Media Initiative, un programa enfocado en el avance de la educación, el periodismo y la investigación de los medios de habla hispana. Más información en [www.utahispanicmedia.com](http://www.utahispanicmedia.com).

**Recibido:** 24 de julio de 2020; **evaluado:** 26 de agosto de 2020; **aceptado:** 1 de octubre de 2020.

## BIG DATA E JORNALISMO: COMO O JORNALISMO ESTADUNIDENSE ESTÁ USANDO BIG DATA

ANDREW M. CLARK

JULIÁN RODRÍGUEZ

UNIVERSITY OF TEXAS AT ARLINGTON

### *Resumo*

Esta pesquisa utiliza entrevistas em profundidade com três jornalistas de dados dos jornais estadunidenses *Houston Chronical* e *New York Times* para descrever o papel dos jornalistas de dados e ilustrar como e por que usam o *big data* em suas histórias. Os jornalistas de dados contam com um conjunto de habilidades único, o que inclui a capacidade de encontrar e reunir dados para depois escrever uma história impactante de forma visualmente coerente. Os resultados das entrevistas e pesquisas mostram que, à medida que os jornais passam a um formato digital, o papel dos jornalistas de dados é cada vez mais essencial, além da implementação de leis, como a Lei da Liberdade de Informação, que permitem aos jornalistas solicitar e utilizar dados para informar o público e fazer com que os que estão no poder prestem contas.

**Palavras-chave:** *big data*; jornalismo de dados, jornalismo, jornais, *Freedom of Information Act* (Lei da Liberdade de Informação)

**Os autores:** Dr. Andrew M. Clark é professor associado no Departamento de Comunicação da Universidade do Texas em Arlington. Ele também é Diretor do Plano de Melhoria da Qualidade da universidade e Diretor Associado do Centro de Pesquisa em Excelência no Ensino e Aprendizagem. Seus interesses de pesquisa incluem transmissão internacional, propaganda, opinião pública e história da transmissão. Ele publicou em revistas como o *Journal of Radio and Audio Media*, *International Communication Gazette*, *American Journalism*, e *Digital Journalism*. Sua experiência de ensino inclui cursos em radiodifusão, lei de comunicação e métodos de pesquisa qualitativa.

Julián Rodríguez é professor no Departamento de Comunicação da Universidade do Texas, em Arlington. Julián ensina jornalismo de radiodifusão com ênfase na mídia em espanhol, pesquisa a mídia hispânica dos EUA e a adoção de tecnologias de mídia para o desenvolvimento de sistemas de consciência. Julián é o diretor da Iniciativa de Mídia Hispânica da UTA, um programa focado no avanço da educação, jornalismo e pesquisa de mídia em língua espanhola. Mais informações em [www.utahispanicmedia.com](http://www.utahispanicmedia.com).

**Recebido:** 24 de julho de 2020; **avaliado:** 26 de agosto de 2020; **aceito:** 1 de outubro de 2020.

## Introduction

The advancement of technology has made it easier for individuals, businesses, and organizations, both public and private, to collect, store, and disseminate data. Gathering and managing data empowers institutions like the Center for Disease Control (CDC) to identify and address public health issues in near real-time. For instance, the end of winter and beginning of spring marks the beginning of influenza (flu) season in the United States. But how many people get the flu? How many strains are there? Who is most at risk? What about on a world-wide scale? An organization like the CDC keeps statistical information that would be useful for people, governments, and healthcare providers. Similarly, with the outbreak of COVID-19 worldwide, big data gives government agencies and independent organizations the ability to compile statistics and track trends around the world.

According to SAS (2020) big data refers to data that is “so large, fast or complex that it’s difficult or impossible to process using traditional methods...the concept of big data gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three V’s: volume, velocity and variety.”<sup>1</sup> Essentially, the three V’s describe how organizations collect large amounts of data from a variety of sources, at incredible speed and in all types of formats, both structured and unstructured. (Grable and Lyons 2018) Grable and Lyons (2018) note the value of data sets comes from “the vast information hidden within the data structures. When analyzed computationally, big data can provide more precise insights into hidden patterns, trends, and associations, especially in the context of human decision making” (17).

Stuijs, Braaksma and Daas (2014) note this information can be used "in public discussion, forms the basis of policy decisions, is required for business use, feeds scientific research, is used in education and so on" (1).<sup>2</sup> They argue there must be "collaboration between National Statistical Institutes, Big Data holders, businesses and universities" (1) to make the best use of big data. Nonetheless it is difficult for the average person to make sense of the information, in addition to knowing where to find it or how to obtain it. Journalists play a pivotal role in not only knowing how to obtain the information, but also making sense of it for their readers, viewers, and listeners, often in an interactive graphic format.

---

<sup>1</sup> (SAS 2020).

<sup>2</sup> (Struijs, Braaksma and Daas 2014).

For journalists, big data can also be used to track the effectiveness of what they write. Ferrer-Conill (2017) states companies that employ journalists can use big data to track metrics, and then reward journalists whose articles meet those metrics with various incentives.<sup>3</sup> Furthermore, in mass media, radio stations use big data to ascertain the station's effectiveness in reaching its audience.<sup>4</sup> It is useful not only for station programming, but also for providing valuable information to advertisers. Thus, media uses big data to develop content that targets audiences locally and globally.<sup>5</sup>

Big data can be collected in many ways, ranging "from satellite and sensory data, to social network and transactional data," among others, according to Stujis et.al. (2014) (2). It could be census data, polling data, geographic data, or social media data, but its volume is so large that powerful computers are needed to compile it.<sup>6</sup> For an individual journalist, or indeed a team of journalists, collecting and making sense of data is a challenging task.<sup>7</sup> Nguyen and Lugo-Ocando (2015) argue journalists have a hard time interpreting and understanding the power of statistics and how these statistics shape everyday life. In other words, it is one thing to identify the source of the data, but it is another thing entirely to understand the data's relevancy and interpret it in a way the public can understand it.<sup>8</sup>

Gaining access to data, particularly if it is held by a government agency or private entity, can prove to be difficult, perhaps impossible, unless there is a legal mechanism to facilitate the request and access to public or public-private data. Governments and private entities executing public contracts are often reluctant to provide data when this information exposes performance issues. Understanding and interpreting data comes through experience and training, but the right to request access to government held data comes from legislation similar to the federal Freedom of Information Act (FOIA).<sup>9</sup>

---

<sup>3</sup> (Ferrer-Conill 2017).

<sup>4</sup> (Ziegler 2016).

<sup>5</sup> (Arsenault 2017).

<sup>6</sup> (Kitchin and McArdle 2016).

<sup>7</sup> (Fairfield and Shtein 2014).

<sup>8</sup> (Nguyen and Lugo-Ocando 2015).

<sup>9</sup> (Baack 2015).

## Freedom of Information Act (FOIA)

In the United States, press freedoms have their root in the First Amendment to the Constitution, which states "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances."<sup>10</sup> Congress, which by virtue of the 14<sup>th</sup> Amendment, refers to all federal and state government institutions, shall not restrict the press. In *Houchins v. KQED*, the Supreme Court established "Neither the First Amendment nor the Fourteenth Amendment mandates a right of access to government information or sources of information within the government's control."<sup>11</sup> (id., at 15). The press may battle secrecy through legislation like the federal Freedom of Information Act (FOIA), but there is no guarantee of success. For the press, and anyone else seeking information through FOIA, the government has a comprehensive website ([www.foia.gov](http://www.foia.gov)), that includes everything from information on the cost of FOIA requests to government departments, to details on how to request information.

Similarly, states also have Freedom of Information rules that govern who can ask for information and how to obtain it.<sup>12</sup> These pieces of legislation are crucial to journalists, because the data entities collect is not always readily available. In some cases where if the data highlights an issue the organization does not want investigated or contains information considered sensitive to national security, the entity may be reluctant to share the data or have the right to deny access to it. At the federal level, FOIA gives journalists the right to request information and the organization must provide it unless it shows the data requested is protected by one of nine statutory exemptions. Most states provide for similar exemptions. Even if the organization holding the requested information claims the data is exempt, for example to protect national security, journalists may file lawsuits and ask the courts to intervene if they believe the organization's exemption claim is unlawful.

---

<sup>10</sup> U.S. Const. amend. I.

<sup>11</sup> *Houchins v. KQED*, 438 U.S. 1 (1978).

<sup>12</sup> For example, the State of Texas has the Texas Public Information Act <https://statutes.capitol.texas.gov/SOTWDocs/GV/htm/GV.552.htm>.

## Research Questions

Considering the challenges journalists face in gathering and using big data described above, two research questions have been developed to guide this study:

RQ1: How do newspaper journalists in the United States use big data?

RQ2: What is the role of data journalists at newspapers in the United States?

## Methodology

To research the role of data journalists, their backgrounds, and the issues they face gathering data, the authors interviewed three data journalists: John Harden, who at the time worked at the Houston Chronicle (he is now at the Washington Post), and Tim Wallace and Derek Watkins who work for the New York Times. They were selected for several reasons. Harden has won several regional and national awards for his innovative work using big data. Wallace and Watkins work for a newspaper that sets the agenda in the United States and around the world. In addition, they have created innovative graphics for many national and international stories requiring the use of big data. For example, they led the way in creating illustrations on the impact of Hurricane Harvey in Houston, Texas. What sets them apart from other journalists is they both have degrees outside of the field of journalism, which demonstrates the multifaceted approach that newspapers are taking to cover stories in an immersive way.

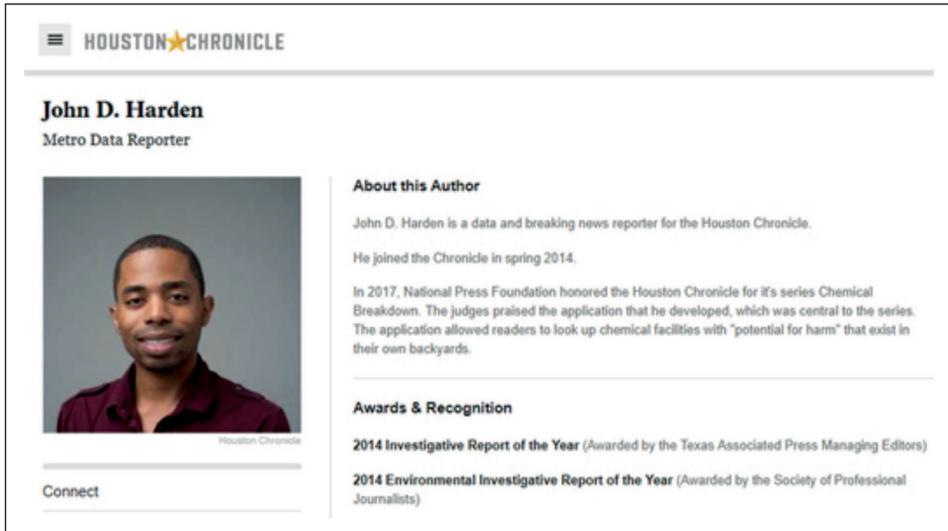
The authors conducted the interviews via Skype, and transcribed and reviewed them for accuracy. Qualitative research seeks to understand the human experience with methods that explore participants' thoughts and experiences that are difficult to measure quantitatively. In-depth interviews were the most appropriate way to gather data, as it was important to hear directly from reporters about their experiences, the process they undergo, their challenges, the use of laws such as FOIA, and how they use big data to illustrate stories.

A survey of a larger sample of data journalists was not feasible because little was known about how data journalists use big data and the challenges they face. However, the data gathered in these interviews will enable further research on a larger sample of data journalists in the United States and around the world. Such a study across various countries would more completely illustrate the

commonalities and challenges journalists face in dealing with different laws and varying access to information.

## Data Journalism

A good example of how journalists are using big data to investigate and develop stories comes from the *Houston Chronicle* in Texas, the fourth largest city in the United States, and its former award-winning Metro Data Reporter, John Harden.<sup>13</sup>



The image shows a screenshot of a web page for John D. Harden, a Metro Data Reporter at the Houston Chronicle. The page features a header with the Houston Chronicle logo, a profile picture of John D. Harden, and sections for 'About this Author' and 'Awards & Recognition'. The 'About this Author' section states that John D. Harden is a data and breaking news reporter for the Houston Chronicle, joined in spring 2014, and mentions a 2017 award from the National Press Foundation for his work on the 'Chemical Breakdown' series. The 'Awards & Recognition' section lists two awards from 2014: 'Investigative Report of the Year' and 'Environmental Investigative Report of the Year'.

Harden regularly uses big data to generate story ideas and create interactive graphics to illustrate stories.<sup>14</sup> In a conversation, Harden stated he files at least one FOIA request each week and often more. For the most part, his requests are to state agencies and fall under the State of Texas FOIA laws, not the federal law.<sup>15</sup> By state law, entities in Texas have 10 days to respond to a FOIA request and to let him know what the charge will be. Harden relies on these requests to give him access to data to help create visualizations for his stories. He uses specialized software to help with the visualizations, but he also uses freely available tools like those provided by Google. Indeed, Google not only provides tools to help interpret and visualize big data, it also collects and disseminates big data through applications like Google Earth ([www.google.com/earth](http://www.google.com/earth)) and Google Public Data ([www.google.com/publicdata](http://www.google.com/publicdata)).

<sup>13</sup> John Harden now works as data reporter for The Washington Post.

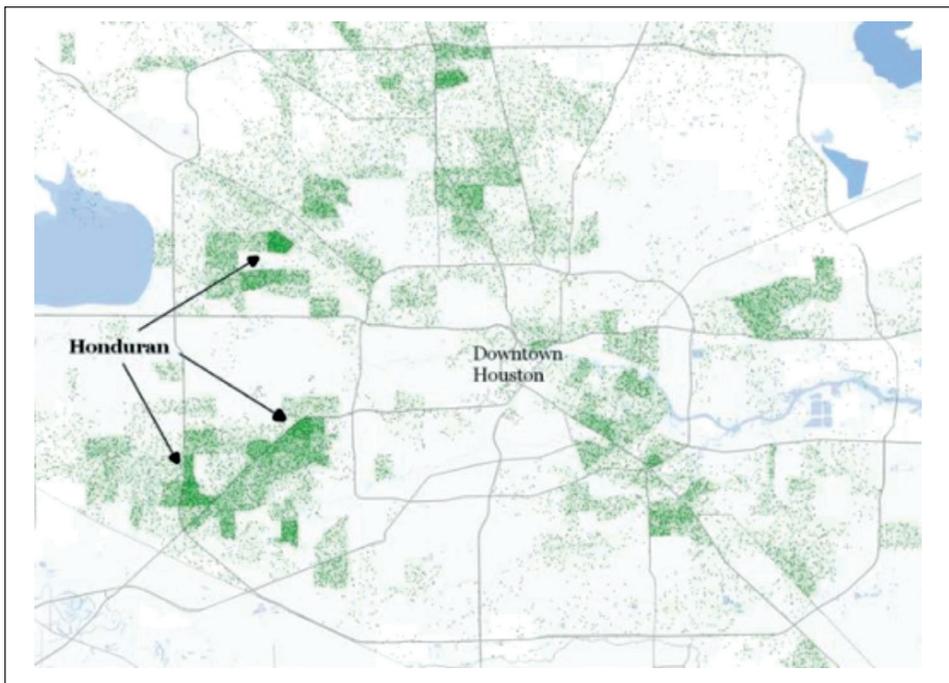
<sup>14</sup> (Houston Chronicle 2018).

<sup>15</sup> (Harden, Data Visualization and Journalism 2018).

Harden says the demand for data journalists is growing among newspapers, magazines, and native online publications. The ever-changing environment of data reporting demands journalists stay up to date. Harden is part of several social media groups of data journalists who share information and communicate about changes in the industry, tools, cases, and collaborations, among other topics. Companies like Google regularly offer continuing education through its Training Program, with classes on using its tools for analyzing and visualizing big data.

One story Harden wrote that illustrates the use of big data to provide visualizations for a story mapped multi-national communities in the Houston metropolitan area. Using big data from sources including the Census Bureau ([www.census.gov](http://www.census.gov)), he was able to illustrate changes in Houston's population in 2017, and where various ethnic communities are located throughout the city.<sup>16</sup> Harden created interactive graphics, such as the one illustrated below, that allow readers of the article to more clearly understand where these communities exist (See Graphic 1).<sup>17</sup>

Graphic 1: Honduran Communities in the Houston Metropolitan Area

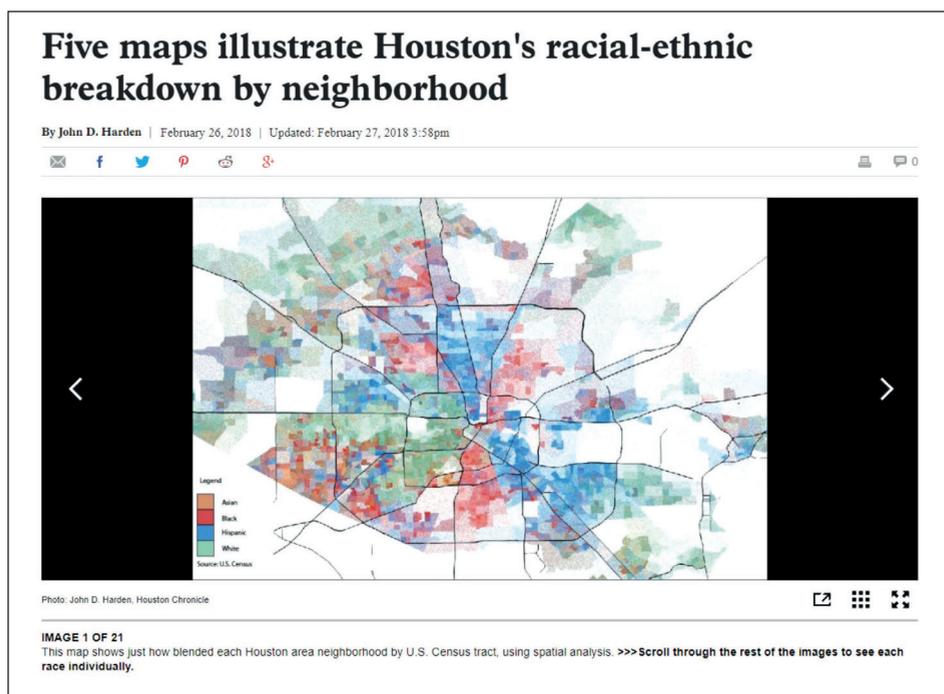


<sup>16</sup> (Harden, Houston region's population growth decelerated in 2017, Census figures show 2018).

<sup>17</sup> (Harden, Five maps illustrate Houston's racial-ethnic breakdown by neighborhood 2018).

These interactive maps also enable Harden to include an interactive data analysis. For example, through his visualizations he is able to break down data into individual identification of communities by ethnic background, or create composite views of multiple ethnic groups in the Houston area (See Graphic 2).<sup>18</sup> The analytical and graphic tools and the availability of big data from the Census Bureau enable him to tell a visually compelling, interactive, and in-depth story that can be explored through digital formats.

Graphic 2: Racial-Ethnic Communities in the Houston Metropolitan Area

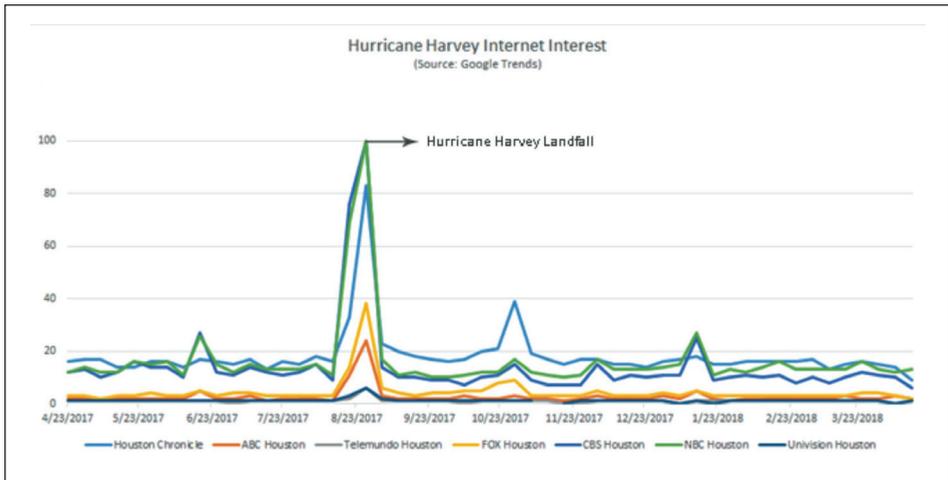


Data visualization allows the reader to engage with the story and create their own analysis based on Harden's work. It is important to note that Harden is reporting on an interesting and important city phenomenon: the growth of ethnic communities. Big data helps him extrapolate what growth may look like over the next years and its potential effects. This would not have been possible even just a few years ago. The accessibility of these tools and the data itself make his stories more valuable to readers.

<sup>18</sup> (Harden, Five maps illustrate Houston's racial-ethnic breakdown by neighborhood 2018).

There are times when an event is of such magnitude it affects the entire country, and thus attracts the attention of national publications, as evidenced in the Graphic 3 below, which shows the spike in public interest across Houston media companies.<sup>19</sup> Hurricane Harvey was such an event that affected Houston and the surrounding areas.

Graphic 3: Hurricane Harvey Internet Interest



According to the United States National Hurricane Center ([www.nhc.noaa.gov](http://www.nhc.noaa.gov)), damage from Hurricane Harvey exceeded USD 125 billion. Its devastation highlighted flaws in Houston community planning and development, which caught the attention of Tim Wallace and Derek Watkins, data journalists for the *New York Times*. Wallace and Watkins' backgrounds are not that of traditional journalists, but their education gave them the tools and the passion necessary for their jobs. Tim Wallace studied archeology and developed an interest in maps. Similarly, Derek Watkins studied geography, with a focus on maps and geographic visualization. Watkins said as he progressed through graduate school, he realized a career in academia was not for him and, with his interest in maps, he was progressing toward visual communication.

<sup>19</sup> (Google 2018).



Tim Wallace - Graphics Editor, New York Times



Derek Watkins - Graphics Editor, New York Times

Although data journalism was not the primary goal for either, it was a natural product of their varied interests, particularly in cartography. They admired the visual work they saw in the *New York Times*, and, in the words of Tim, "I was irritated that I didn't know how to do what they were doing."<sup>20</sup> Apart from their interest in maps and visualization, they both also started at the New York Times as interns.

Both Wallace and Watkins work at the "graphics desk." "Desk" in newsroom lingo refers to a specific department. The graphics desk is very much a part of the newsroom, with around 40 employees, but unlike other "desks", it is "organized around ways of presenting information as opposed to a specific topic or beat," such as Education, Healthcare, or Crime. Journalists on the graphics desk use tools to present stories, including maps, charts, satellite imagery, 3-D models, virtual reality, augmented reality, and drone footage. Wallace and Watkins note the common theme for those working in graphics is visualizing data in journalism. Not every story lends itself to visualization in the sense that some stories are better covered using traditional written word. Wallace and Watkins consider that if it gets too difficult to visualize stories, that is an indication to use a more traditional format to present the information. However, according to Tim Wallace, "at least once a day they check in as a team to see what they can do visually, based on what the top stories for that day are."

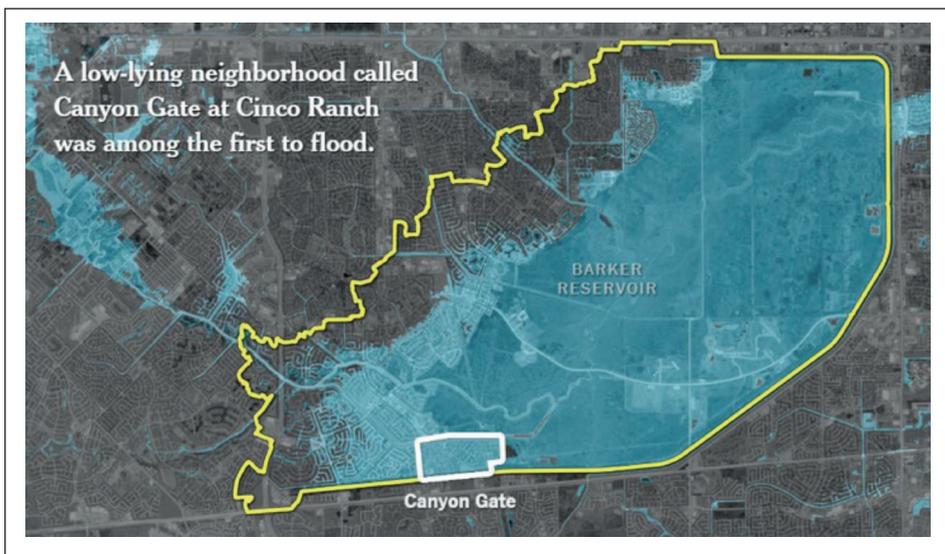
---

<sup>20</sup> (Wallace and Watkins 2018).

Watkins and Wallace see a clear distinction between what they do as data journalists, and what is known as computer assisted reporting. Unlike data journalism, computer assisted reporting is a term that has been used in the *New York Times* newsroom for quite a few years, and many reporters would be considered computer assisted reporters.<sup>21</sup> As Watkins and Wallace point out, there are really "two sides: the way you collect the data, and the way you present the data. You need to choose the techniques that tell the story in the best way." Neither of them is a member of or active in data journalism organizations, but they do attend conferences and keep in touch with other data journalists, in particular through Twitter and other online networks.

Together Watkins and Wallace created stories that visualize the vulnerability of certain communities to a once-in-a-century event like Hurricane Harvey. To create their visualizations, they used big data from a variety of sources. One of the most notable pieces they worked on in the past year was a visualization of the effect of Hurricane Harvey on a specific housing area in Houston built inside a reservoir (See Graphics 4 & 5).<sup>22</sup>

Graphic 4: Canyon Gate Built Inside Barker Reservoir



<sup>21</sup> (Hammond 2015).

<sup>22</sup> (Park, et al. 2018).

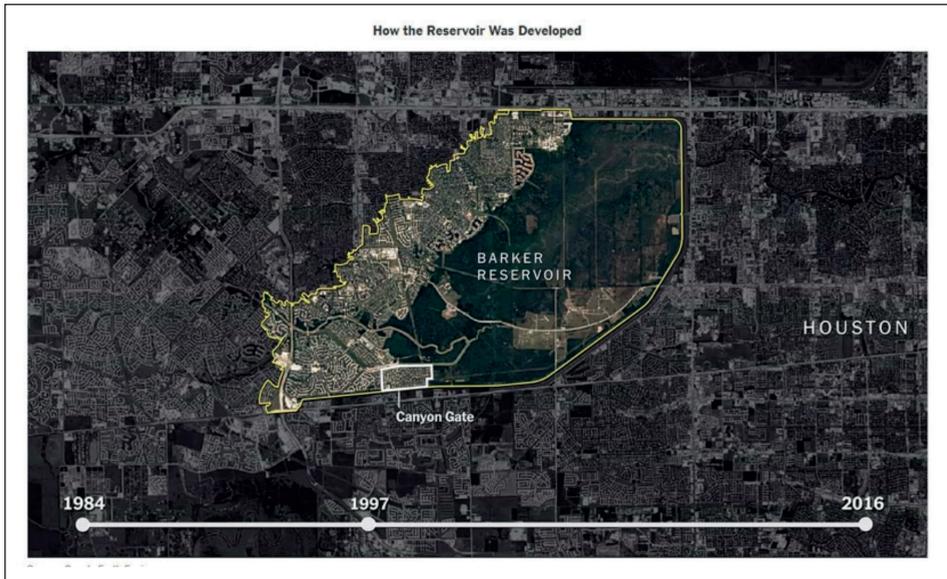
It is a perfect example of the power of visualization and the impact it can add to traditional reporting. Watkins and Wallace estimate their piece *How One Houston Suburb Ended up in a Reservoir* on this reservoir and the issues caused by the hurricane was about a quarter of the whole *New York Times* package (print stories, videos, data pieces etc.). In addition to their visualization, which included time progression with satellite imagery, there was also a video piece and two written pieces. Sometimes the written word can paint a very clear picture of something, such as a natural disaster and residents' reactions to it. However, as in the case of this story, seeing the effects of the storm and changes over time paints a whole new picture.

Graphic 5: Canyon Gate Flooded by Hurricane Harvey



As does Harden, data journalists at the *New York Times* rely on their relationship with Google and the tools Google offers. For example, the tool used to create the time lapse in the Hurricane Harvey story was Google Time Lapse, which uses satellite imagery with 30-meter resolution, with images from all over the world from 1986 forward (See Graphic 6).

Graphic 6: Time Lapse of Barker Reservoir Development (1984-2016)



Such tools and information are publicly available, which cuts down on the need for FOIA requests to obtain data. If a FOIA request is needed, employees at the New York Times have the knowledge and skills to make requests efficiently. Being at the New York Times may have its advantages; although Watkins and Wallace cannot say for certain being employed at the Times means they have greater access to information, they recounted an incident that happened when they was still interning at the paper. They said they had tried for a long time to obtain information using their university email accounts, but never had any success. However, with the approval of their supervisor, Watkins and Wallace asked for that data using their NYT email and received the information the next day.

Public data is something they use every day. Some of the data comes from maps and satellite data that is made publicly available by the U.S. Government and the European Space Agency.<sup>23</sup> Watkins and Wallace note there are many satellites traversing the sky at any given time, generating all types of high volume data, and two of the largest uses of satellite data are in agriculture and meteorology. The key seems to be creativity and the capacity to see and develop ways of using publicly available information to illustrate issues in a way that effectively informs audiences.

<sup>23</sup> (Batty 2013).

When developing stories, senior editors at the graphics desk check their work, and their colleagues also edit their stories, "they keep an eye out for each other." If a story they are working is in collaboration with another desk, "they have the full editing power of the other desks." Watkins and Wallace said opinion data journalism (data op-ed, is relatively new. Those involved in opinion pieces, even in terms of visualization, "don't sit in the newsroom... there is purposeful separation from the news side... they don't work hand-in-hand."

## Conclusion

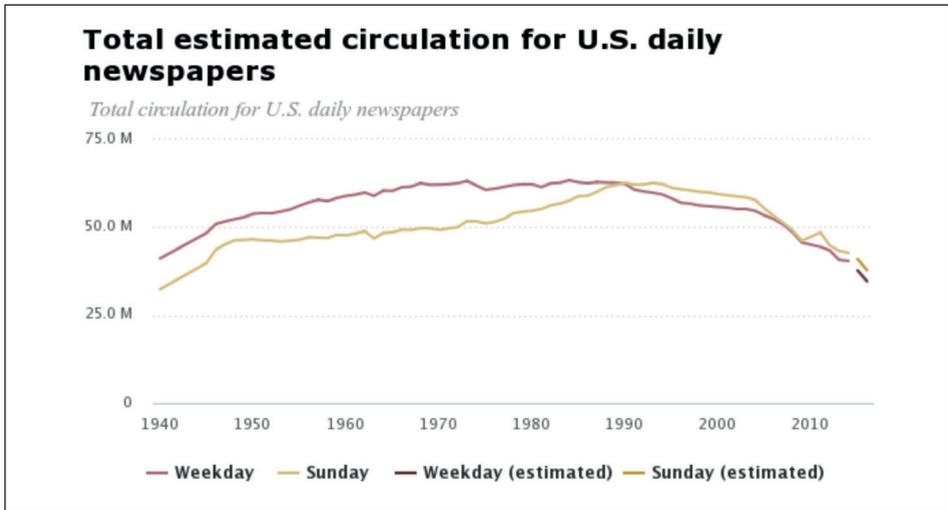
This research was guided by two overarching questions that asked how U.S. newspaper journalists use big data, and what the role of data journalists at newspapers in the United States is. Results from the interviews show how through technological advances, big data is more available and increasingly important to data journalists. Data journalists may follow a traditional journalism path, as is the case with John Harden, who obtained a degree in journalism before working at a series of newspapers, or may follow a more unique path, like Wallace and Watkins, whose background in geography and cartography has provided them unique skills in interpreting and illustrating big data. For national or global events, big data is invaluable as reporters seek to illustrate the enormity of the event, or the impact on specific groups of people. Journalists would have covered such occurrences in the past, but today, big data enables reporters and data journalists to offer completely new perspectives with interactive graphics generated from the data set.

The issue of big data and the ability to create interactive graphics to illustrate a story is in many ways an extension of the way newspapers have always been viewed. Broadcast was the medium of choice for breaking news, and newspapers were the choice for more in-depth analysis of events. Both mediums are changing, but newspapers are moving from print medium to a digital platform. Statistics from a Pew Research Center report demonstrate this change (see Graphic 7), "The estimated total U.S. daily newspaper circulation (print and digital combined) in 2016 was 35 million for weekday and 38 million for Sunday, both of which fell 8% over the previous year. Declines were highest in print circulation: Weekday print circulation decreased 10% and Sunday circulation decreased 9%."<sup>24</sup>

---

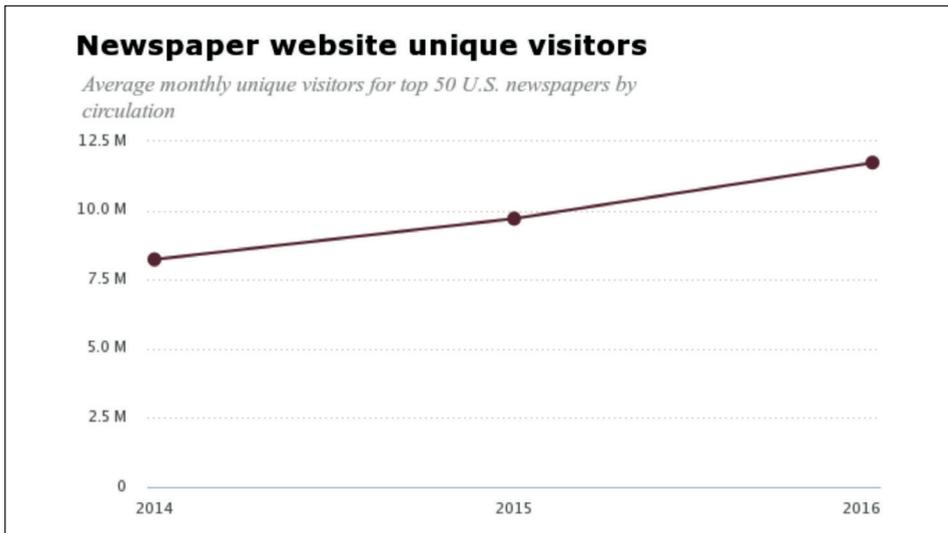
<sup>24</sup> (Pew Research Center 2017).

Graphic 7: U.S. Newspaper Circulation, Pew Research Center



The digital platform has the potential to open many opportunities for newspaper companies, as illustrated by the increase in data journalists. The ability to harness the power of big data and create in-depth analysis and visualization may be a catalyst a struggling industry. People may not buy the print edition as they once did, but they are reading papers online (see Graphic 8) and the visualization and insight data journalists create is part of the reason some papers are surviving.

Graphic 8: Newspaper Website Traffic, Pew Research Center



The role of data journalists is increasingly important in the newsroom. This role requires a particular skillset. Data journalists must be skilled at creating graphics; finding and gathering data, and using it to tell a compelling, visually coherent, and appealing story. As Watkins and Wallace noted, much of what they use is publicly available, but it is important data remain accessible, as are legal means like FOIA journalists may use to request data to inform the public and hold accountable those in power.

## References

- Arsenault, Amelia H. 2017. "The datafication of media: big data and the media industries." *International Journal of Media & Cultural Politics* 7-24.
- Baack, Stefan. 2015. "Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism." *Big Data & Society* 1-11.
- Batty, Michael. 2013. "Big data, smart cities and city planning." *Dialogues in Human Geography* 274-279.
- Fairfield, Joshua, and Hannah Shtein. 2014. "Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism." *Journal of Mass Media Ethics* 38-51.
- Ferrer-Conill, Raul. 2017. "Quantifying journalism? A study on the use of data and gamification to motivate journalists." *Television and New Media* 706-720.
- Google. 2018. *Google Trends*. Accessed April 17, 2018. <https://trends.google.com/trends>.
- Grable, John E., and Angela C Lyons. 2018. "An Introduction to Big Data." *Journal of Financial Service Professionals* 17-20.
- Hammond, Philip. 2015. "From computer-assisted to data-driven: journalism and big data." *Journalism* 408-424.
- Harden, John, interview by Julian Rodriguez. 2018. *Data Visualization and Journalism* (March 23).
- . 2018. *Five maps illustrate Houston's racial-ethnic breakdown by neighborhood*. February 26. Accessed June 26, 2018. <https://www.houstonchronicle.com/houston/article/Five-maps-illustrating-Houston-s-racial-breakdown-12711221.php#photo-15142320>.
- . 2018. *Houston region's population growth decelerated in 2017, Census figures show*. March 22. Accessed June 26, 2018. <https://www.houstonchronicle.com/houston/article/Houston-region-s-population-growth-decelerated-in-12772015.php#photo-15270096>.
- Houston Chronicle. 2018. *John D. Harden*. Accessed June 20, 2018. <https://www.houstonchronicle.com/author/john-d-harden/>.
- Kitchin, Rob, and Gavin McArdle. 2016. "What makes big data, big data? Exploring the ontological characteristics of 26 datasets." *Big Data & Society* 1-10.

- Nguyen, An, and Jairo Lugo-Ocando. 2015. "The state of data and statistics in journalism and journalism education: issues and debates." *Journalism* 3-17.
- Park, Haeyoun, Anjali Singhvi, Tim Wallace, Derek Watkins, and Josh Williams. 2018. *How One Houston Suburb Ended Up in a Reservoir*. March 22. <https://www.nytimes.com/interactive/2018/03/22/us/houston-harvey-flooding-reservoir.html>.
- Pew Research Center. 2017. "Newspapers Fact Sheet." June 1. Accessed June 6, 2018. <http://www.journalism.org/fact-sheet/newspapers/>.
- SAS. 2020. *Big Data, What it is and why it matters*. Accessed 20 02, 2020. [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html).
- Strujis, Peter, Barteld Braaksma, and Piet JH Daas. 2014. "Official statistics and Big Data." *Big Data & Society* 1-6.
- Wallace, Tim, and Derek Watkins, interview by Andrew Clark and Julian Rodriguez. 2018. *Data Visualization and Journalism* (May 1).
- Ziegler, Lady Dhyana. 2016. "Radio as numbers: counting listeners in a big data world." *Journal of Radio & Audio Media* 182-185.