

Reflexión

Introduction to DATA Journalism

Tom Sanderson
tom@tcij.org

Recibido: 27 de junio de 2018 | **Aceptado:** 20 de julio 2018

DOI: <https://doi.org/10.18272/pd.v2i0.1332>

Referencia de este artículo:

Sanderson, T. (2018). Introduction to Data Journalism. *#PerDebate, volumen 2* (pp. 58-67). Quito: USFQ Press.

Tom Sanderson es editor de información del Centro de Periodismo de Investigación (CIJ). Investigador independiente y activista voluntario con la campaña End Child Detention Now. El CIJ es un grupo de reflexión, una universidad alternativa y un laboratorio experimental creado para capacitar a una nueva generación de profesionales en periodismo investigativo. Sanderson tiene una larga trayectoria en la investigación en red y periodismo de datos.



Abstract

Journalistic research is based on data rather than statements. In the digital age, the challenge of doing data journalism is global visualization. It not only implies showing numbers and statistics, but also customizing them for audiences, to discuss them in their social networks. Journalists must find a story to show that extensive information or big data in a digestible way. Here the keys that Sanderson explained during a workshop in Quito at the International Journalism Congress in Debate 2017.

Keywords

data journalism, data visualization, precision, journalistic research, globalization.

Introducción al Periodismo de Datos

Resumen

La investigación periodística se basa en datos más que en declaraciones. En la era digital, el reto de hacer periodismo de datos es la visualización global. No solo implica mostrar números y estadísticas, sino personalizarlos para las audiencias, que las discutan en sus redes sociales. Los periodistas deben encontrar una historia para mostrar esa amplia información o *big data* de forma digerible. Aquí las claves que Sanderson explicó durante el taller en Quito durante el Congreso Internacional Periodismo en Debate 2017.

Palabras claves

Periodismo de datos, visualización de datos, precisión, investigación periodística, globalización.

What is data journalism?

- Gathering Data, scrapers.
- Understanding Data: 'cleaning', 'analysing' or 'interviewing' to find leads or evidence.
- Presenting Data: visualisation, interactivity and personalisation

Not just numbers and statistics

Data is everywhere: audio, video, connections on social networks, photographs, documents, financial information, likes on Facebook, web search results, numbers on spreadsheets.

Also Metadata

When information was scarce, most of our efforts were devoted to hunting and gathering. Now it is abundant, processing is more important.

We process at two levels: (1) analysis to bring sense and structure out of the never-ending flow of data and (2) presentation to get what's important and relevant into the consumer's head.

Philip Meyer
Professor Emeritus, University of North Carolina at Chapel Hill

Software is your friend

It helps you to: gather data, clean data, analyse the data, find connections, find patterns or changes, find the story, tell the story.

Computer Assisted Reporting (CAR)

- Pioneered in the US because Philip Meyer's coverage of the 1967 Detroit riots.
- Bill Dedman's revelations of racial bias in lending policies of banks in the 80s



The racial segregation was very strong in Detroit to the point that the owners put this sign in 1942. Credit: Wikipedia.

- Steve Doig's analyses of the damage from Hurricane Andrew in the 90s
- Police raided after hours drinking club in predominantly black neighbourhood.
- Deadliest riots in US history 43 dead, 467 injured, 7,200+ arrests 2,000+ buildings destroyed.
- Psychology professor Nathan Caplan and journalist Phillip Meyer used social science techniques to look for root cause.
- Rioters thought to have low economic status or educational levels.
- Also blamed recent immigrants from the South.

Among the Findings

There was no correlation between economic status and participation in the disturbance.

College-educated residents were as likely as high school drop-outs to have taken part.

Recent immigrants from the South had not played a major role; in fact, Northerners were three times as likely to have rioted.

Precision Journalism

“Precision journalism was a way to expand the tool kit of the reporter to make topics that were previously inaccessible, or only crudely accessible, subject to journalistic scrutiny. It was especially useful in giving a hearing to minority and dissident groups that were struggling for representation.”

Philip Meyer

Author of *Precision Journalism:*

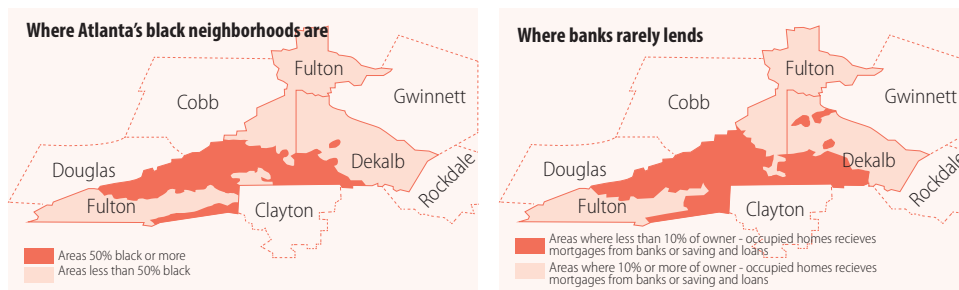
A Reporters Introduction to Social Science Methods (1973)

80s - Racial Bias in Lending Policies

- Through a FoIA request to the Federal Home Loan Bank Board loan Bill Dedman obtained applications
- Analysed 10 million applications for loans in the US
- Derived rejection rates from a national analysis by race, sex and marital status
- Revealed in much of the US high-income blacks were rejected at the same rate as low-income whites
- Blacks were rejected more than twice as often as whites
- Race better predictor of success than sex or marital status

The Color of Money

First of four parts



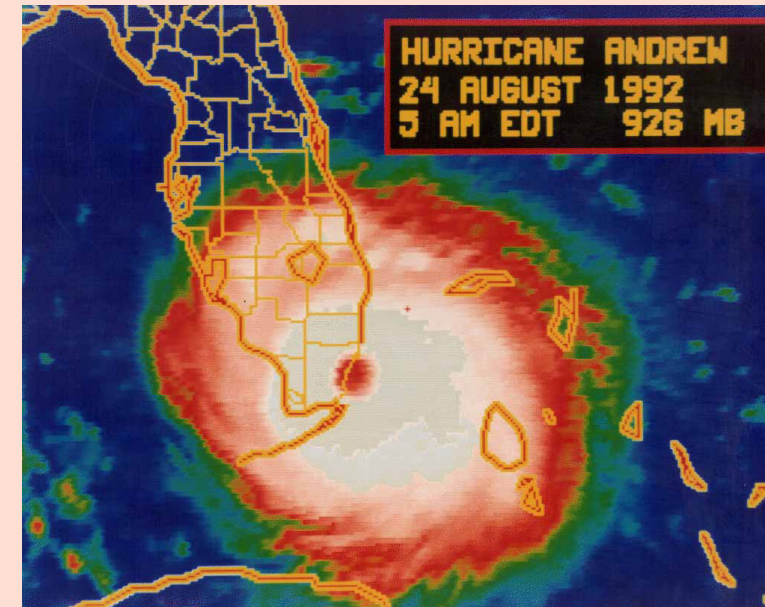
Source: 1980 U.S. Bureau of the Census figures.

Source: Federal Financial Institute Examination Council figures for 1981 - 86

The racial segregation was evidenced in the mortgage loans in the city of Atlanta in the 80s.

What went wrong in 90s with Hurricane Andrew? Stephen Doig, Miami Herald

- Compared dates, damage and paths of other Florida hurricanes
- Merged damage reports with the property-tax roll
- Newer houses more significantly damaged
- Discovered that during Florida's 27-year hurricane-free period, houses were not built to code
- Analysed of campaign contributions showed one out of four campaign dollars from building industry
- “Instead of just supplying anecdotes,” said Doig. “We provided powerful evidence with data analysis”



Considered the most destructive tropical cyclone in the United States during the 20th century, hurricane Andrew affected Florida, the Bahamas and Louisiana. Hundreds of homes destroyed in Florida but others in path untouched.

“CAR is the use of computers and social science methods to acquire and analyze information to tell stories that otherwise would be difficult or impossible”.

Stephen Doig

New Approaches to Journalism

- New technology
- A move towards greater openness
- Application of social science technologies to news gathering and reporting
- Greater computer access and faster processing
- Faster, cheaper, more readily available software
- Larger data sets
- More people accessing news digitally therefore more scope different ways of telling stories

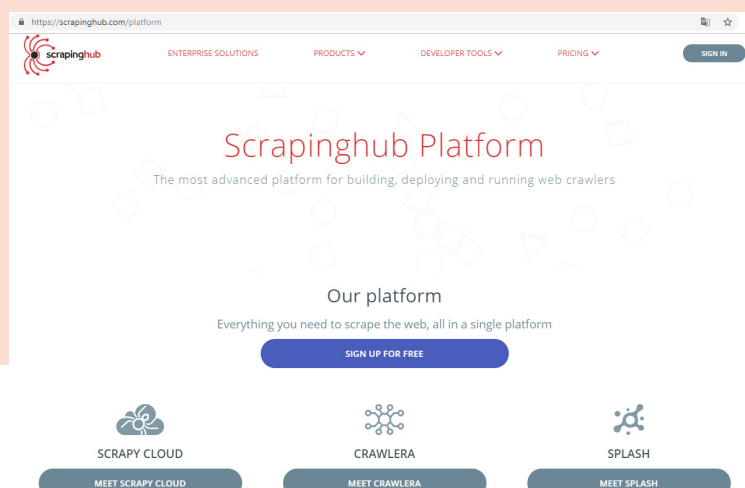
Modern Data Journalism

- Automate data gathering (scraping)
- Find stories in data sets using spreadsheets (Excel) or databases (Access and SQL)
- Data visualisation and infographics (Google Fusion, Tableau, Car-toDB)
- Explain how stories affect an individual

Data Scraping

Automates the process of gathering data: Helium, Outwit, Scraperhub, Quickcode, Python, Ruby, R.

On the web there are scraper platforms like Scraping Hub. Screenshot.



It Starts with a Question

Crime rates, financial information, public services, environment, planning and building regulations, connections, education, health.

Interviewing Data

- A dataset is a source like any other, though with very specific knowledge and an infallible memory for facts
- To interview it, you must learn to ask questions in a language it understands
- It cannot lie to you, but be aware that it may have been lied to by others...

More examples of data journalism

Data of the Week <http://gijn.org/2014/09/29/top-ten-ddj-the-weeks-most-popular-data-journalism-links-29/>

What is Big Data? <http://gijn.org/2014/09/09/what-is-big-data/>

Tableau Gallery <http://www.tableausoftware.com/public/gallery>

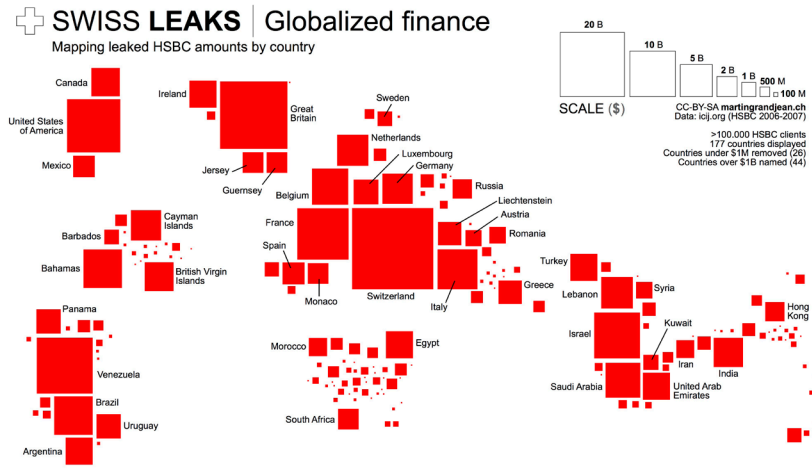
DataBlog The Guardian <http://www.theguardian.com/news/datablog>

Country Statistics <http://www.nationmaster.com/>

Spurious Correlations <http://www.tylervigen.com/>

Where do we find data?

- Published data: open government data, datasets published by NGOs, researchers.
- Requested data: freedom of Information Law, scraping.
- Leaked data: whistle-blowers, inside sources.
- Created data.



The *International Consortium of Investigative Journalists* published this global map of tax evasion based on data visualization in February 2015. Credit: Martin Grandjean

[Journalism is] going to be about poring over data and equipping yourself with the tools to analyse it and picking out what's interesting. And keeping it in perspective, helping people out by really seeing where it all fits together, and what's going on in the country."

Tim Berners-Lee

More webgraphy about data journalism

16 useless infographics. (August 1st, 2013). The Guardian. www.theguardian.com/news/datablog/gallery/2013/aug/01/16-useless-infographics

Gray, J.; Bounegru, L. y Chambers, L. (2012). The Data Journalis. Open Knowledge Foundation. <http://datajournalismhandbook.org/1.0/en/index.html>

Data Journalism Blog. A news site tackling innovative projects made with data, in the newsroom and elsewhere. <http://www.datajournalismblog.com/>

BBC College of Journalism. Journalism blog of BBC Media. <http://www.bbc.co.uk/blogs/collegeofjournalism/entries/89cf3a79-1a82-3b83-8cad-67451db-cee95>

Weaver, D. y McCombs, M. (January 1st, 1980). Journalism and Social Science: A New Relationship? *Public Opinion Quarterly*, Vol. 44, Issue 4, 1 pp. 477–494. <https://doi.org/10.1086/268618>