

DDJ: Interviewing data

Data as a source is like any other source

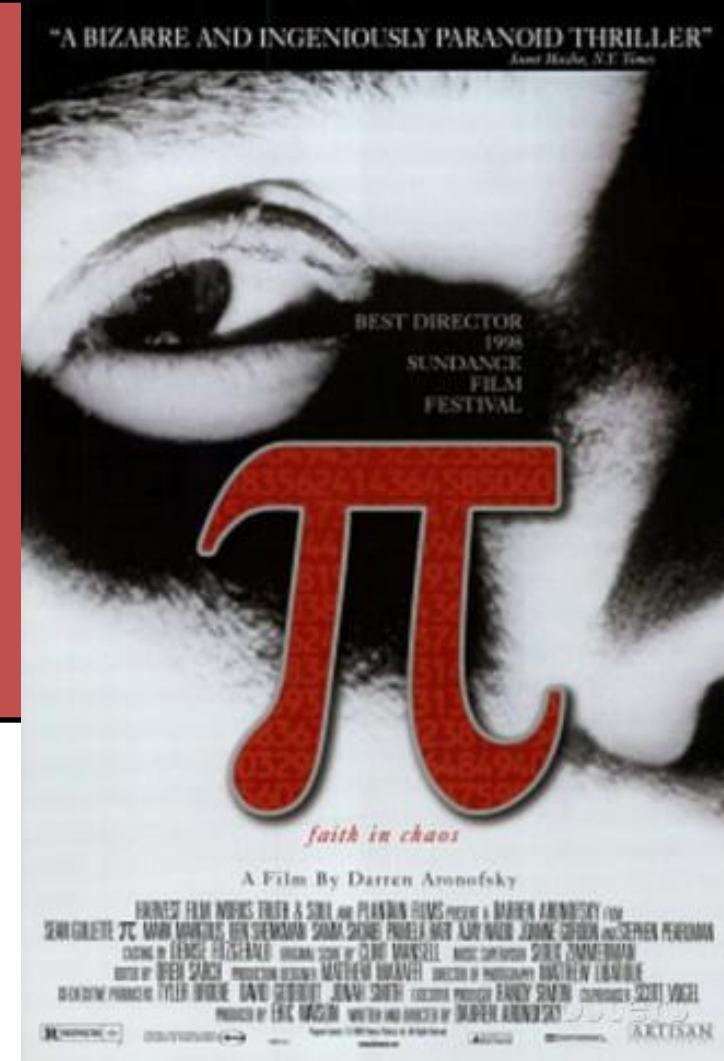
Hassel Fallas – Data Journalist @Nacion.com – Costa Rica – Twitter: @hasselfallas

Introductions



"11:15, restate my assumptions:

1. Mathematics is the language of nature.
2. Everything around us can be represented and understood through numbers.
3. If you graph these numbers, patterns emerge. Therefore: There are patterns everywhere in nature".



Maximillian Cohen in Pi, film by Darren Aronofsky, 1998.



74 million



4,8 million

22

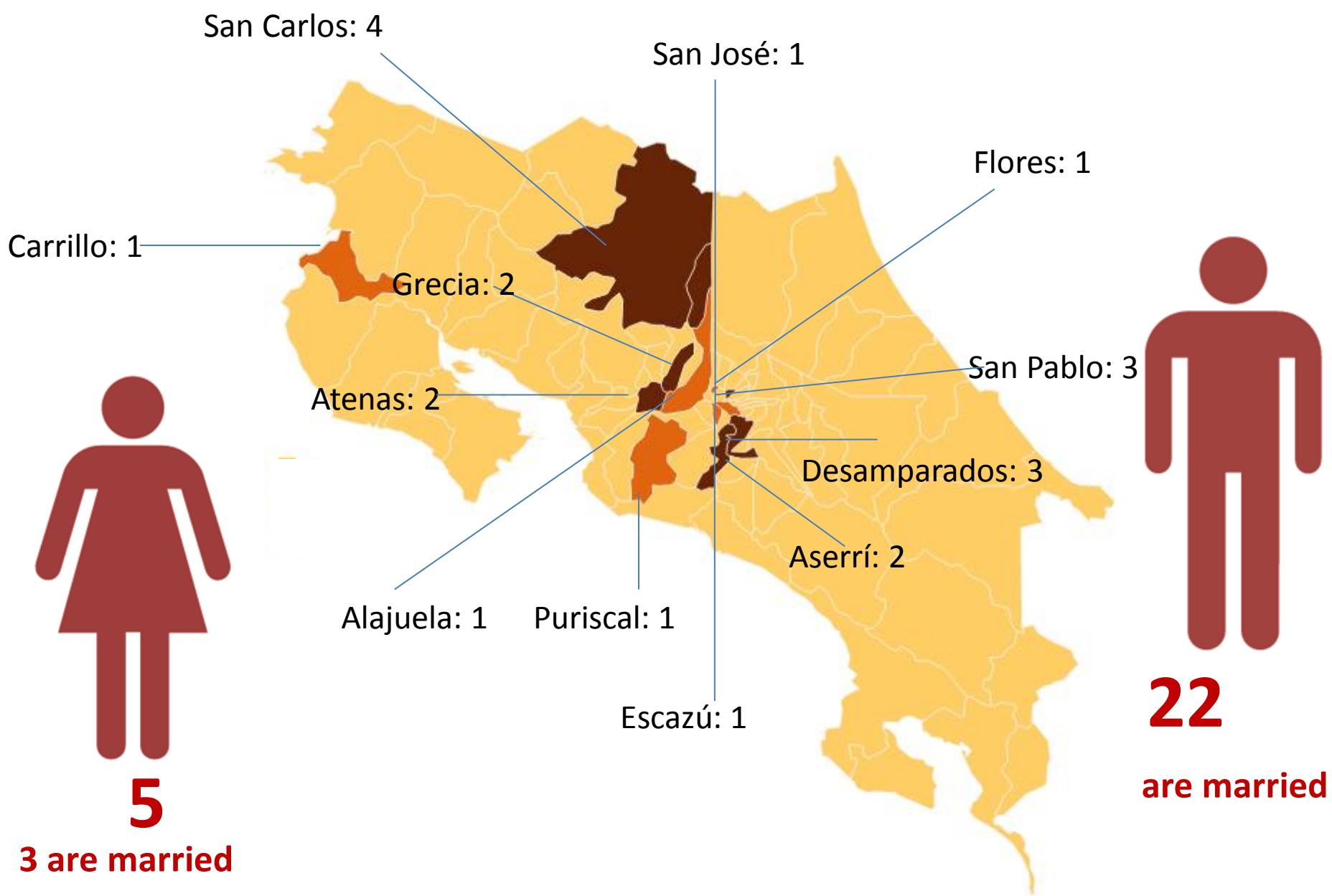
People

3

between 1990-1999

19

between 2000-2001



Who they are? Where do they live?

**Data Driven Journalism is so complicated
that only geeks and unicorns can do it!**

Really?

What is data?

“Database journalism allows reporters to investigate in literally a whole new universe of unfiltered data. If we know how to ask the **right questions**, this data will expose unimaginable stories, never revealed before even by our most effective source”

Giannina Segnini, data driven journalist

What is data?

“Data is a source like any other source. It's **fallible**, it's **incomplete**. Just like a human being, it's sometimes hard to know where the incompleteness and the lies are”

Aron Pilhofer, associate managing editor of digital strategy at The New York Times

What is data?

“Numbers are like **fire**. They can be used for good or for ill. When **misused**, they can create illusions of certitude and importance that render us irrational”.

Meyer, Philip Precision Journalism: A Reporter's Introduction to Social Science Methods

What is data?

“You should pay attention to your **gut feelings** and **critically question** the data. The big risk is that, when you go back to the government with something they don’t know, they have to believe you”

*Sarah Cohen, computer assisted reporting editor for
The New York Times*

Where to start?

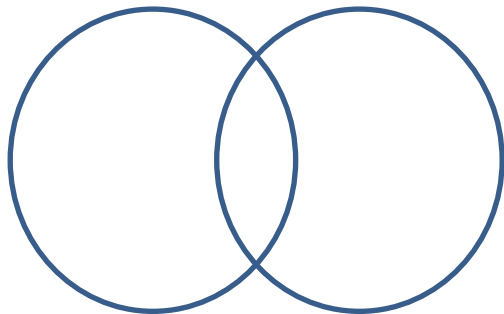
**Good interview =
Good interviewer**



1. Skepticism.
2. Openness.
3. An instinct for operationalization.
4. A sense of the tentativeness of truth.
5. Parsimony.

*Meyer, Philip Precision
Journalism*

Journalists and scientists have more in
common than you know



Why this story matters?



PEOPLE

“To engage people with things that matter, we need to explain why they should care - and often that’s about engaging the human side of things”

Paul Bradshaw, Data Journalism Heist

First questions about data

- Where this data come from?
- How reliable is the information?
- Is the data set complete, with all the variables you asked or were contained in the original?
- If there are missing parts, why?
- How this data was collected? What was the [methodology](#)?
- Does it provide me all the data I need or do I have to look for other data sets or build my own?

Where to start?

“Knowing what to do with data is the essence of the new **precision journalism**. The problem may be thought of as having two phases: the input phase, where data are **collected and analyzed**, and the output phase, where the data are prepared for entry into the reader’s mind”

Meyer, Philip Precision Journalism: A Reporter's Introduction to Social Science Methods

Where to start?

“Once we’re in the data, we need to know where we’re going. We don’t want to spend a minute longer than we have to. **Every dataset will have stories it can tell**”

Paul Bradshaw, Data Journalism Heist

Making questions to data

- Who is top or bottom?
- What is data saying about the country, region, industry, etc?
- Is there any unusual change in the sequence of numbers?
- How much the numbers have change between one year and another or even in a period?
- What is going up and down?
- Who are the winners and losers?
- Can I corralete this data with any other one?

Practical example

Half of the Municipalities throw away Home Recycling Efforts

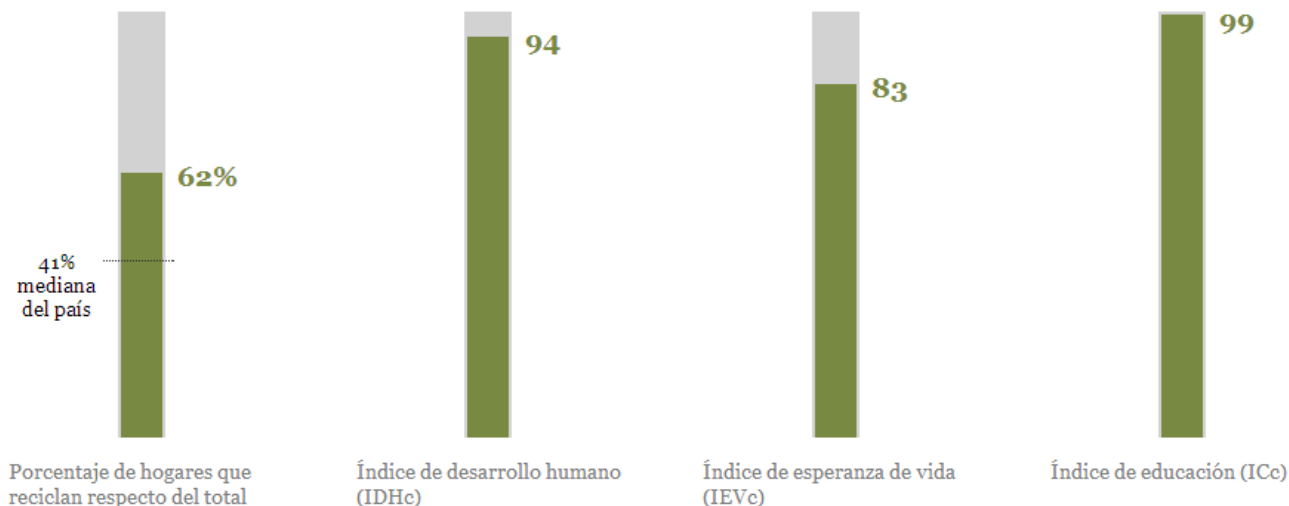
VISUALIZACIÓN

MAPA GRÁFICO PROBLEMA DOTA

RANKING GENERAL

INDICADORES POR CANTÓN

Belén



**MUNICIPIOS RECICLAN
SOLO 2% DEL TOTAL DE**

Consulte el porcentaje de hogares que reciclan y otras variables de desarrollo por cantón, según el Programa de Naciones Unidas para el Desarrollo.

The process

F2		fx =E2/D2				
	A	B	C	D	E	F
1	State	Urban or rural area	Town	Households	Households that recycle	% of households recycling
2	San José	2	Dota	1952	1489	76%
3	Alajuela	2	Zarcelero	3333	2253	68%
4	Cartago	2	Jiménez	4113	2748	67%
5	San José	1	Escazú	16565	10639	64%
6	Heredia	1	Belén	6011	3722	62%
7	Guanacast	2	Abangares	5311	1434	27%
8	Guanacast	2	Bagaces	5615	1461	26%
9	Alajuela	2	Orotina	6024	1498	25%
10	Guanacast	2	La Cruz	4732	1121	24%
11	Guanacast	2	Cañas	7172	1673	23%
12						

Data extracted from the 2011 CR Census.

The process

	A	B	C	D	E	F	G	H	I
1	State	Urban or rural area	Town	Tons of any residuals	Tons of ordinary trash per month	Tons of recyclables saved per month	% of recyclables from total trash	Are recyclables being pick them up in separated way	Do you have recycling programs
45	Puntarenas	2	Puntarenas	2213	2210	3	0%	NO	YES
46	San José	1	Aserri	843	840	3	0%	NO	YES
47	Alajuela	1	Palmares	703	700	3	0%	NO	YES
48	Guanacaste	2	Bagaces	166,8	166	0,8	0%	NO	YES
49	Puntarenas	2	Parrita	265,6	265	0,6	0%	no	YES
51	Alajuela	2	Los Chiles	150,5	150	0,5	0%	NO	YES
52	Guanacaste	2	Nandayure	83,3	83	0,3	0%	NO	YES
56	Puntarenas	2	Buenos Aires	408	408	0	0%	NO	NO
57	Cartago	1	Cartago	642	642	0	0%	NO	YES
59	Heredia	1	Flores	525	525	0	0%	NO	YES
60	San José	1	Goicoechea	3200	3200	0	0%	NO	YES
63	Alajuela	2	Guatuso	132	132	0	0%	NO	YES

And the human side of this story?

The human stories

Café impulsa reciclaje masivo en hogares de Dota



Sonia Madrigal trabaja en el centro de acopio de Coopedota. De lunes a viernes abre bolsas y separa el plástico, cartón y aluminio que los miércoles les envían desde sus casas los pobladores del cantón de Dota. DIANA MENDEZ.

Coffee promotes Mass Recycling in Dota Homes

Last year, the 800 coffee producers of Coopedota received a half million dollar prize for producing a bean in harmony with the environment. The international markets paid them \$10 more per each one of the 50,000 quintals produced in 2012.

In Dota, coffee constitutes a family business, from which most of the 1,900 homes of the county live, directly or indirectly. That is one of the reasons that explains why, in 8 out of 10 of those houses, paper, plastic, glass, and aluminum are separated from the common garbage.

The human stories

Recyclers ask for Fair Prices and to be Appreciated

Maria Gomez arrives at the CoopeDota recycling center at six in the morning. Her job is the same until noon, from Monday to Friday: to open the bags with recyclable materials sent by the Dota homes. Sometimes, Gomez opens a bag and gets badly surprised. “I am good at finding toilet paper and other rubbish. It pisses me off when they send such things on the recycling day,” she claims.



“Data-driven journalists need to spend some time on the street seeing how the data works in the three-dimensional world. The same happens with reporters that work on street. They need to spend some time with the data to see how it is represented. Any record is actually part of something that is happening. And without any of those perspectives, I think you can lose a lot”

[Sarah Cohen mentioned in Data Driven Journalism Secrets, Hassel Fallas .](#)