Santa Clara University

Markkula Center for Applied Ethics

Jennifer LaFleur

Senior Editor for Data Journalism

Center for Investigative Reporting

Go for accurate, not "good enough"

Just like we background check human subjects of stories, we need to background check our data

Integrity checks for every data set

- Read the documentation. Understand the contents of every field.
- Know how many records you should have.
- Check counts and totals against reports.
- Are all possibilities included? All states, all counties, correct ranges?
- Check for missing data, duplicates, internal problems

No data are perfect

(Note the inconsistencies)

DALLAS
DALLAS CITY
DALLAS TX
DALLAS,
DALLLAS

	Practice City	NU PRACTICE	Practice State	Practice Zip	Birth Year	Birthplace
					1880	
T	ELECTRA		TX	76360	1882	
N	HOUSTON		TX	77019	1884	
					1885	MISSOURI
T	DALLAS		TX	75251	1886	TEXAS
N	DALLAS		TX	75201	1886	
M	IRVING		TX	75061	1887	
T	POST		TX	79356	1887	
T	BRIDGEPORT		TX	76026	1887	
T				76039	1887	
N	VAN NUYS		CA	91406	1888	WISCONSIN

NAME
LOW ENERGY SYSTEMS
LOWE S
LOWELL B RICHARDSON
LOWER COLORADO RIVER AUTHORITY
LOWE'S
LOWE'S BUSINESS ACCOUNT
LOWES COMPANIES INC
LOWE'S COMPANIES INC
LOWE'S CREDIT SERVICES
LOWES HOME CENTER
LOWE'S HOME CENTER
LOWES HOME CENTER 1582
LOWE'S HOME CENTERS INC
LOWES HOME CENTERS INC #0103
LOWES HOME IMPROVEMENT WAREHOUS
LOWE'S MARKETPLACE
LOWE'S WFT #075
LOYD ENTERPRISES
LP SPECIALTIES
LRC ELECTRONICS CO
LRC LLC

Offer audiences transparency: Tell them what you know and what you don't

"Simply putting data online is not journalism."

- Vet your data and interpret it.
- Provide a detailed methodology about the data and your process.
- In cases of more complicated analyses, write a white paper about your research.
- Ask experts to review your analysis. Seek experts with various interests and expertise.
- Invite feedback. Include a mechanism to submit changes.

EXAMPLE: Explanation and invitation

AG	AH	AI	AJ	AK	AL
ALT_DI	ALT_OT	GTE_PR	PROVIDE _AP	NUM_AP_COURSES	STUDENT _AP_CHO ICE
0	0	1	1	22233	1
0	0	0	1	1717	0
0	0	0	1	1397	1
0	0	1	1	1212	1
0	0	1	1	1212	1
0	0	0	1	1111	1
0	0	1	1	666	1
0	0	1	1	545	1
0	0	1	1	532	1
0	0	0	1	527	1
0	0	0	1	404	1

The data were reported by schools and districts to the Office of Civil Rights. ProPublica spent several weeks verifying the accuracy of the data. Where we were able, we corrected extreme outliers and contacted hundreds of schools to verify their data. Because of some of the problems we found in the initial data, Assistant Secretary for Civil Rights Russlynn Ali said that the office is revamping its process for gathering and verifying their data. We also vetted our analysis with education research experts.

We may not have accounted for every problem in the data and welcome feedback from schools and districts.

Check accuracy across stories, graphics, maps and apps

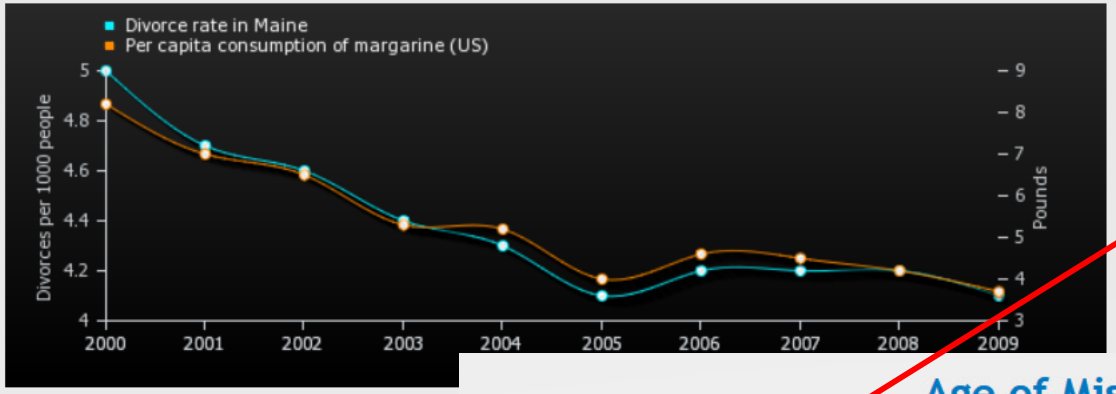
- For example, if you're mapping something and you weren't able to pinpoint the location of some of the data – does that make the map inaccurate?
- Do you have a problem of "tiny Ns"? (Where the population is so small that any change seems big.)

Watch your words

- "Significant", "likely" and "correlate" actually mean something. Use them wisely.

- Beware the spurious correlation
- While your analysis may generate lots of Rs and Ps, descriptives are easier to readers to understand.

Divorce rate in Maine
correlates with
Per capita consumption of margarine (US)

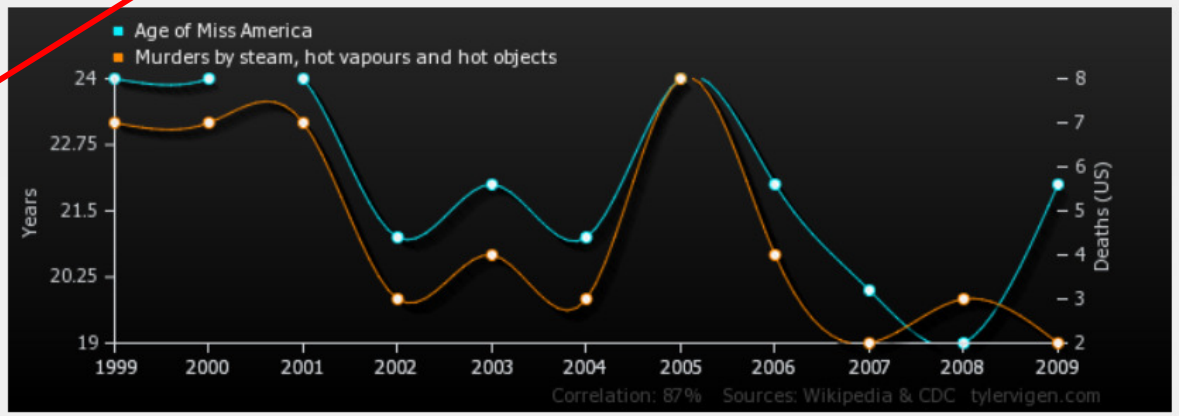


	20(
Divorce rate in Maine *Divorces per 1000 people (US Census)*	5
Per capita consumption of margarine (US) *Pounds (USDA)*	8.

Correlation: 0.992558

Source your information

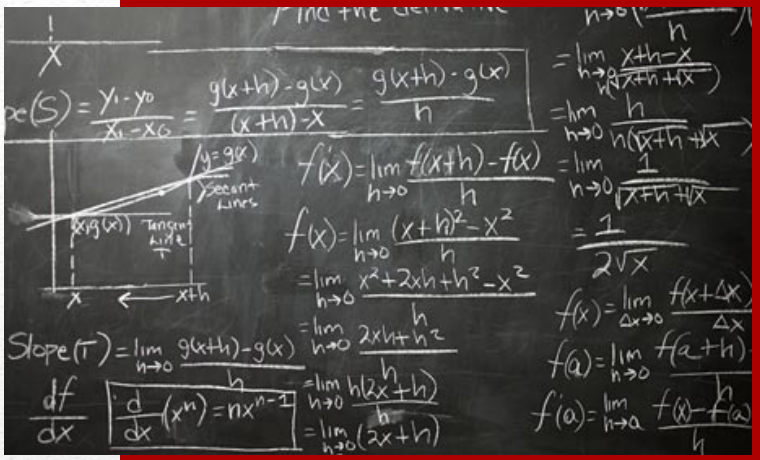
Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Age of Miss America *Years (Wikipedia)*	24	24	24	21	22	21	24	22	20	19	22
Murders by steam, hot vapours and hot objects *Deaths (US) (CDC)*	7	7	7	3	4	3	8	4	2	3	2

Correlation: 0.870127

From http://www.tylervigen.com/



Watch your math

- Adjust money over time (a dollar really doesn't buy what it used to)
- Use rates rather than raw numbers
- Use median when averages might be skewed

PRO PUBLICA | Journalism in the public interest

Presidential Pardons Heavily Favor Whites

(Brendan Smialowski/Getty Images)

by Dafna Linzer and Jennifer LaFleur
ProPublica, Dec. 3, 2011, 11 p.m.

31 Comments | Republish | Email | Print

First of two parts. Part two here. This story was co-published with The Washington Post.

955 Tweet | 68 Share | 23 +1 | 18 Like

White criminals seeking presidential pardons over the past decade have been nearly four times as likely to succeed as minorities, a ProPublica examination has found.

Editor's Note

Blacks have had the poorest chance of receiving the president's ultimate act of

Tell Us Your Story

EXAMPLE:
Database of 500 people who had been granted or denied presidential pardons from list of 2,000.

We found that even after controlling for other factors, whites were more likely to get a pardon.

How ProPublica Analyzed Pardon Data

By Jennifer LaFleur, Director of Computer-assisted Reporting, ProPublica

Collecting the data

ProPublica's project on presidential pardons relied on data about individuals who were denied and granted pardons during the George W. Bush administration. As a matter of practice, his advisers said, President Bush relied almost exclusively on recommendations from the Office of the Pardon Attorney inside the Justic
assess the office's impact on final pa

Through a Freedom of Information A
who were denied pardons during Bu
petitioners who received pardons ca
random sample of 500 names from t
final sample numbered 494. For eacl
pardon office's recommendation to

Regression variables

Nagelkerke pseudo R square: .29
Hosmer and Lemeshow Goodness of fit: p=.42

Variable	B	S.E.	Wald	Sig.	Exp(B)	Reference category
Non-Hispanic White	1.31	0.58	5.18	0.02	3.71	All minorities
Probation only	0.81	0.38	4.51	0.03	2.25	
Military-related crime	1.05	0.77	1.85	0.17	2.84	
Female	0.77	0.52	2.24	0.13	2.17	
No subsequent crimes found	0.49	0.51	0.92	0.34	1.63	
Correspondence written on petitioner's behalf	1.14	0.48	5.74	0.02	3.12	
Married	0.73	0.46	2.51	0.11	2.08	
No bankruptcy found	1.06	0.67	2.52	0.11	2.89	

STRIKING DIFFERENCES

A process of juror elimination

RICHARD MICHAEL PRUITT/Staff Photographer

Prosecutors excluded blacks from juries at more than twice the rate they rejected whites, a study of felony trials showed.

Key Findings

Prosecutors and defense attorneys in Dallas County exclude jurors on the basis of race, despite Supreme Court bans on discrimination in jury selection, a two-year investigation by *The Dallas Morning News* found. Beginning today, *The News* examines the practices of prosecutors, defense attorneys and judges. The key findings:

- Dallas County prosecutors excluded black jurors at more than twice the rate they rejected whites.
- Defense attorneys excluded whites at more than three times the rate they rejected blacks.
- Even when blacks and whites gave similar answers to key questions asked by prosecutors, blacks were excluded at higher rates.
- Blacks ultimately served on juries in numbers that mirror their population primarily because of the dueling prosecution and defense strategies.

Dallas prosecutors say they don't discriminate, but analysis shows they are more likely to reject black jurors

Racial discrimination was once so raw in Dallas County that a black college president who tried to serve on a jury was flung headfirst down the courthouse steps while sheriff's deputies watched.

This past March, nearly 70 years later, a young black man had to show a judge his teeth in order to serve.

The all-white jury — that enduring image of Jim Crow justice — is a fading sight around the Frank Crowley Courts Building. But while times, laws and leaders have changed, race still matters.

Prosecutors excluded eligible blacks from juries at more than twice the rate they rejected eligible whites, *The Dallas Morning News* found. In fact, being black was the most important personal trait affecting which jurors prosecutors rejected, according to the newspaper's statistical analysis. Jurors' attitudes toward criminal justice issues also played an important role, but even when blacks and whites answered key questions the same way, blacks were rejected at higher rates.

District Attorney Bill Hill denied that his prosecutors exclude, or strike, jurors on the basis of race.

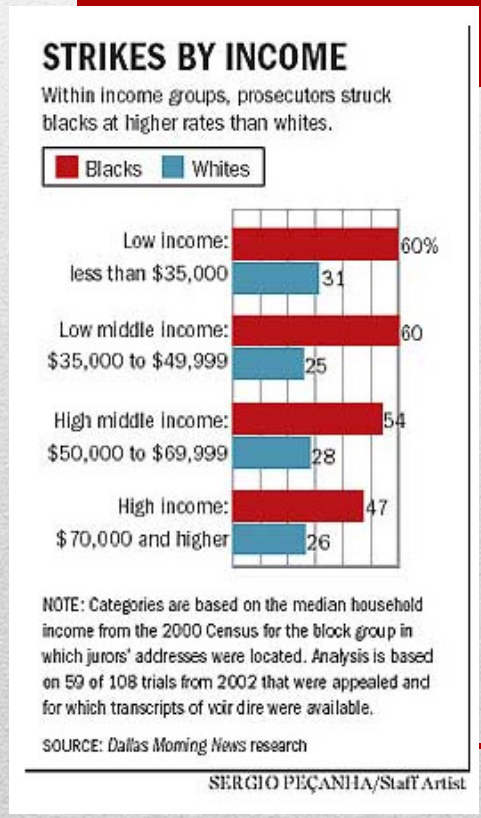
See HILL *Page 16A*

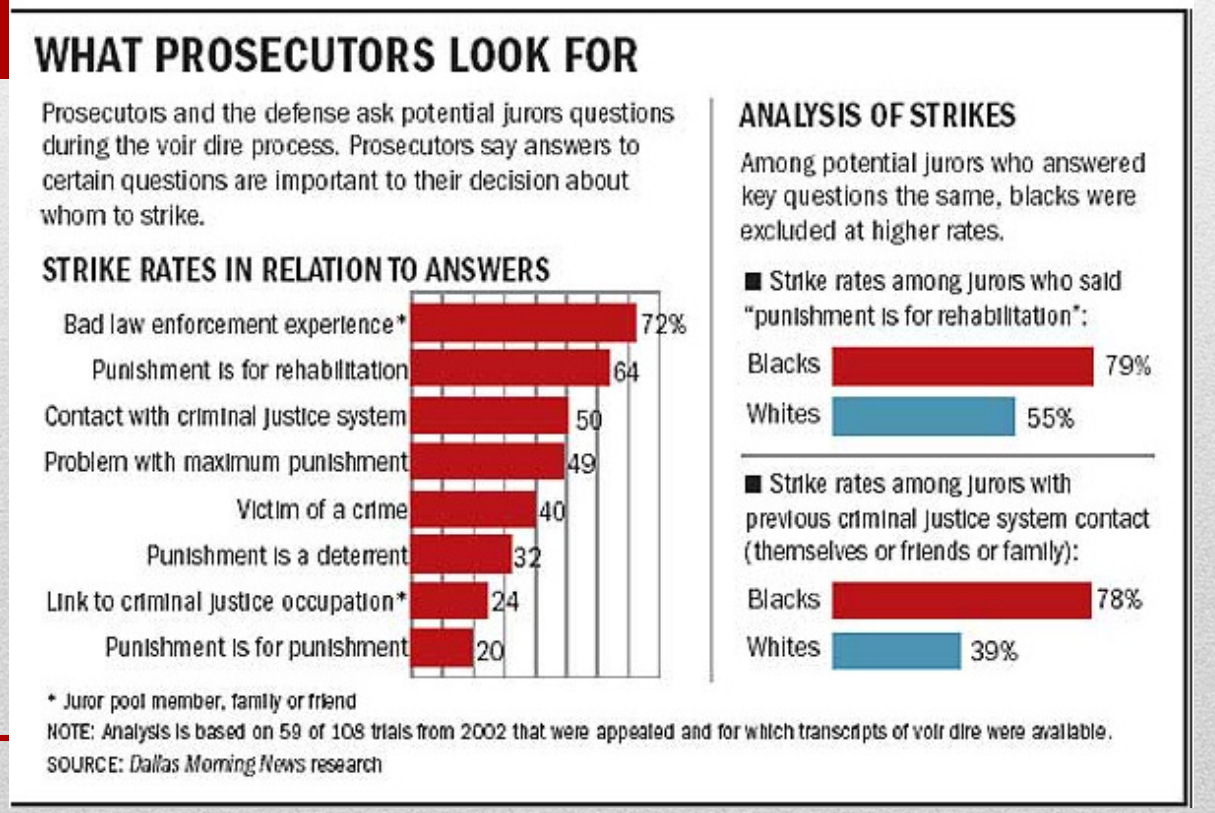
Inside

EXAMPLE: *The Dallas Morning News* used a random sample of non-capital murder cases to build a database and to study jury strikes:

1. Demographics
2. Voir dire
3. Socioeconomics

Juror race	B	S.E.	Wald	df	Sig.	Exp(B)	Reference category
Hispanic	0.44	0.23	3.79	1	0.052	1.55	
Black	1.14	0.17	43.93	1	0.000	3.12	
Other	-0.60	0.46	1.71	1	0.191	0.55	
Race unknown	-0.02	1.51	0.00	1	0.988	0.98	White





MILWAUKEE·WISCONSIN
JOURNAL SENTINEL

Disparities in water usage

- "Water use highest in poor areas of the city"
- Mapping and statistical analysis

Low-income Milwaukee neighborhoods use more water on average

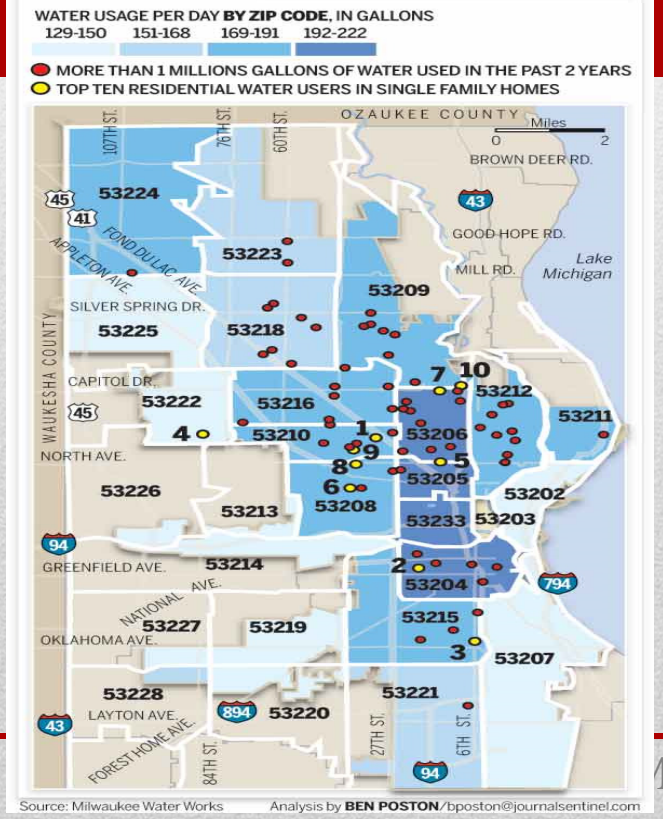
A Journal Sentinel analysis found that single-family homes in low-income neighborhoods tend to use more water on average than wealthier ones. The dots represent the 70 single-family homes in Milwaukee that used more than 1 million gallons of water in the past two years. Of those, nearly two-thirds are located on the near north and near south sides.

The 10 single-family home customers who used the most water from June 2008 to June 2010 all consumed at least 2.3 million gallons during that time. Most of the highest water users had severe leaks or broken pipes, city data shows.

Top 10 Milwaukee residential water users in single family homes

	ADDRESS	GALLONS PER DAY	AMOUNT BILLED
1	2722 N 34TH ST.	6,601	$7,761
2	1426 S 23RD ST.	6,553	$7,085
3	3159 S 9TH ST.	5,817	$6,649
4	2872 N 85TH ST.	4,842	$5,682
5	2139 N 16TH ST.	4,556	$5,583
6	4135 W Vliet ST.	3,837	$4,719
7	3835 N 17TH ST.	3,675	$4,496
8	2135 N 40TH ST.	3,655	$4,368
9	2466 N 41ST ST.	3,195	$3,840
10	3940 N 11TH ST.	3,174	$3,845

Note: Gallons per day and amount billed are based on usage from June 1, 2008, to June 1, 2010. Amount billed is for water usage only.



Source: Milwaukee Water Works Analysis by **BEN POSTON**/bposton@journalsentinel.com

Steps to vet your analysis

- Do a gut check
- Ask: What else could explain my findings?
- Ask: Did I fill in all possible holes?
- Ask: Did I collect all the data I needed to?
- The analysis is just the beginning. Once you start reporting, ask: Is it consistent with my findings?