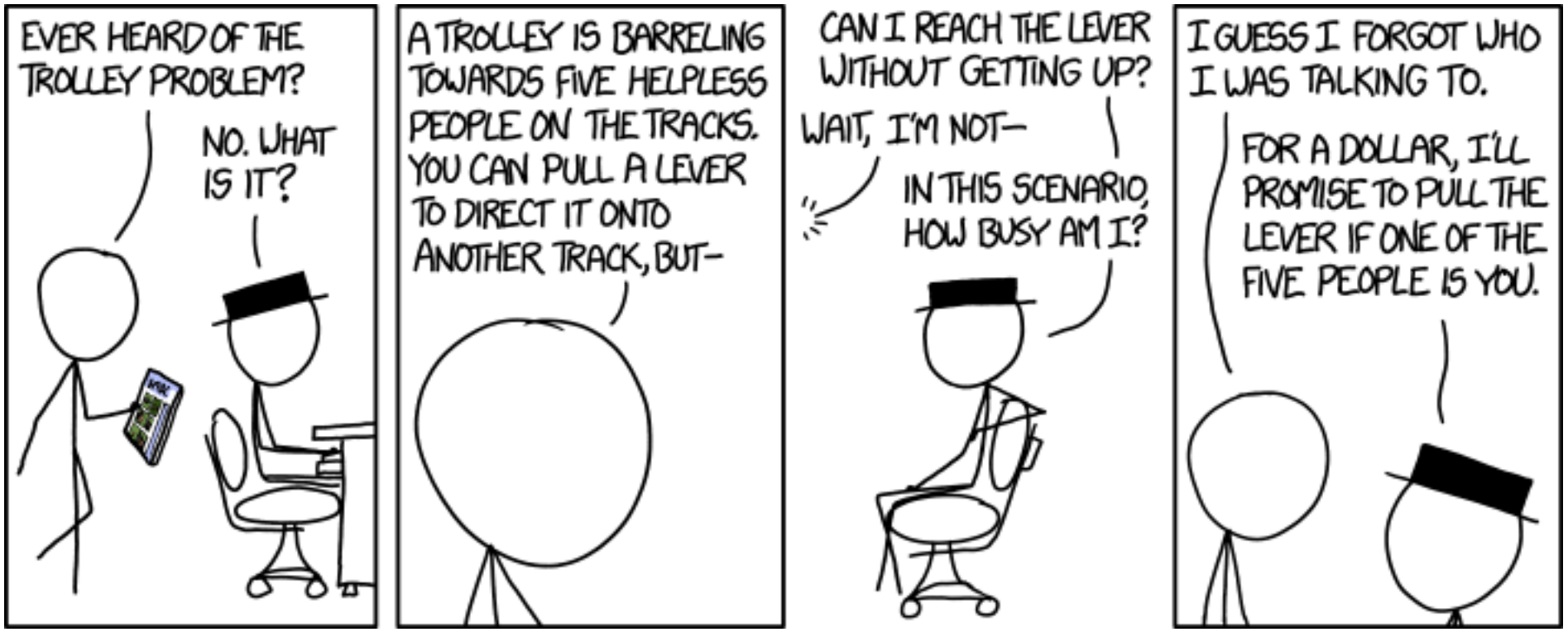
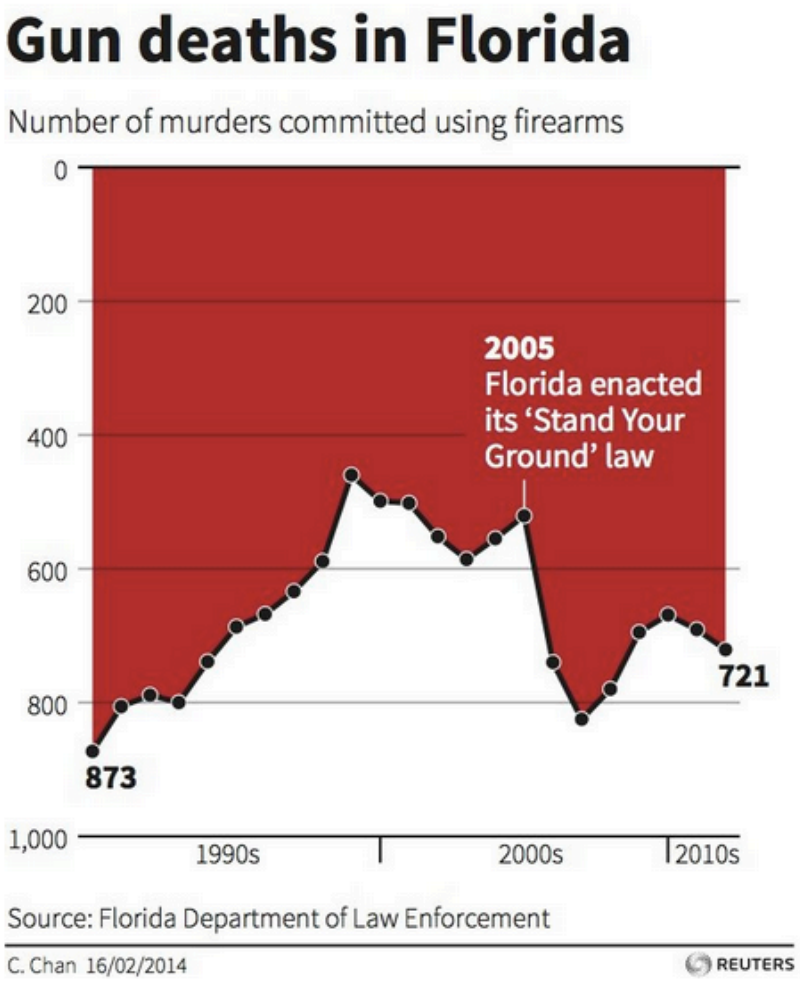
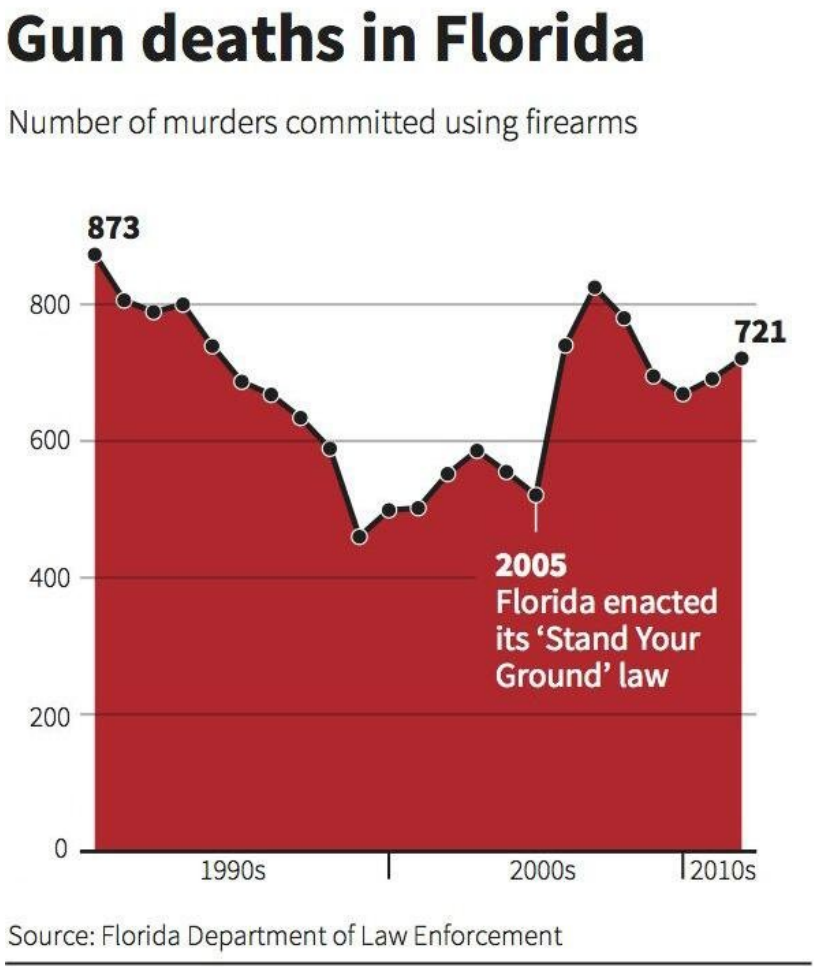
Ethics of Data Visulization



CSE 631 Data Visualization | Steven Bedrick & Jackie Wirz

http://xkcd.com/1455/





http://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2

1) Visualization is the graphical display of information…

2) Graphics act as a cognitive aid

Graphics take the EFFORT out of imagining the information.

Alberto Cairo, Professor of the Professional Practice, University of Miami

The goal of ethical visualizations is to help with cognitive strain, but to stay true to the data and analysis.

Alberto Cairo, Professor of the Professional Practice, University of Miami

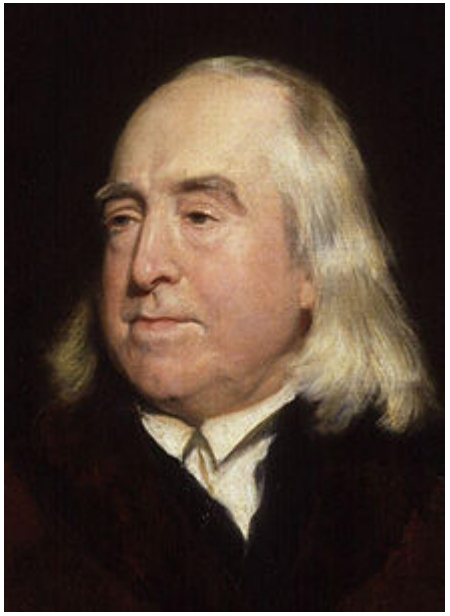
"what I am doing when I create the graphic... Is to take the effort out of your brain and put it on paper"

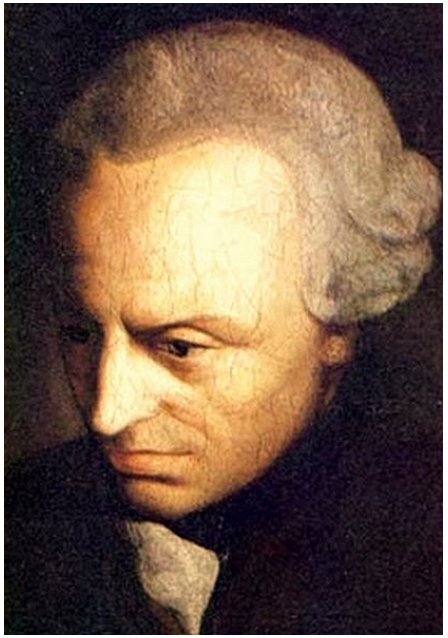
Alberto Cairo, Professor of the Professional Practice, University of Miami

“The goal is not to over simplify the information. We *should* present complexity to readers in an engaging way, people will care.”

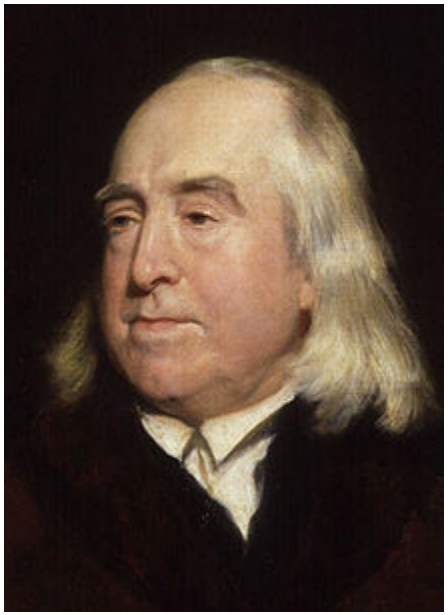
Alberto Cairo, Professor of the Professional Practice, University of Miami

There are two main families of ethical reasoning:





Consequentialism

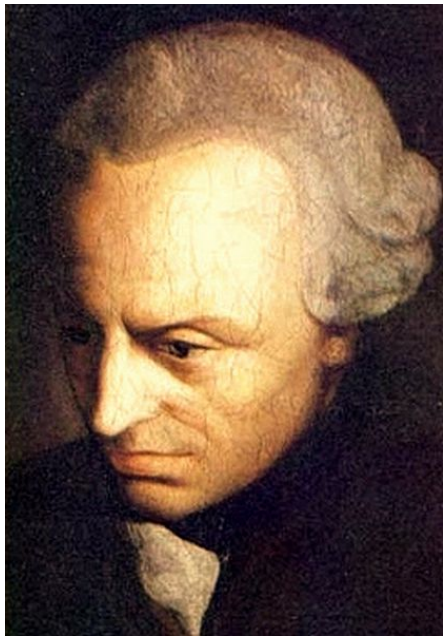


"Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do."

Jeremy Bentham , *Principles of Morals & Legislation*

Image of Jeremy Bentham from wikipedia

Deontology



"Act only according to that maxim whereby you can at the same time will that it should become a universal law without contradiction."

Immanuel Kant, *Grounding for the Metaphysics of Morals*

Image of Immanuel Kant from wikipedia

Modern biomedical ethics recognizes four guiding principles:

Autonomy

Nonmaleficence

Beneficence

Justice

The same principles can be applied to visualization!

What a Good Visualization Should Be

Alberto Cairo, Professor of the Professional Practice, University of Miami

1. Truthful
2. Functional
3. Beautiful
4. Insightful
5. Enlightening

Ethical Correlates

1. Truthful | be honest, avoid lying to ourselves and we avoid lying to others
2. Functional | be clear, don't oversimplify
3. Beautiful | attention to relevant info
4. Insightful | reveal hidden truths
5. Enlightening | attention to relevant info

Visual.ly's corporate code of ethics is instructive:

"As an organization that both practices and recognizes quality data-journalism, Visual.ly subscribes to the code of ethics of the Society of Professional Journalists and agrees to abide by all of its principles.

We also agree to the following principles to support data analysis and visualization:

Data will be accurate and verifiable - Visual.ly will not "lie with statistics."

Proper Sourcing & Attribution - Visual.ly will always give credit where due and will do its own reporting.

Best Practices in Visual Representation - Visual.ly will not exploit idiosyncrasies of the human visual system to exaggerate or misrepresent data.

Most succinctly stated, Visual.ly's policy is one that encompasses accuracy, honesty, and transparency."

Marian Dörk, et al., identify four guiding principles:

Disclosure
Plurality
Contingency
Empowerment

Dörk M, Feng P, Collins C, Carpendale S. Critical InfoVis: exploring the politics of visualization. CHI '13. Paris, France, 2013.

An attempt at distillation:

Our goal in visualizing data is to communicate information to a viewer.

We must strive to do so as accurately and honestly as possible…

… and avoid design decisions that confuse or mislead our viewer.

“A good visualization should tell a story”

“The data should tell the story”

Simplification

Clarification

Simplification

Contextual Granularity: Going Deeper into Data

"The purpose of visualization is insight, not pictures."

Shneiderman B. Research agenda: Visual overviews for exploratory search. Information Seeking Support Systems. 2008;11:4.

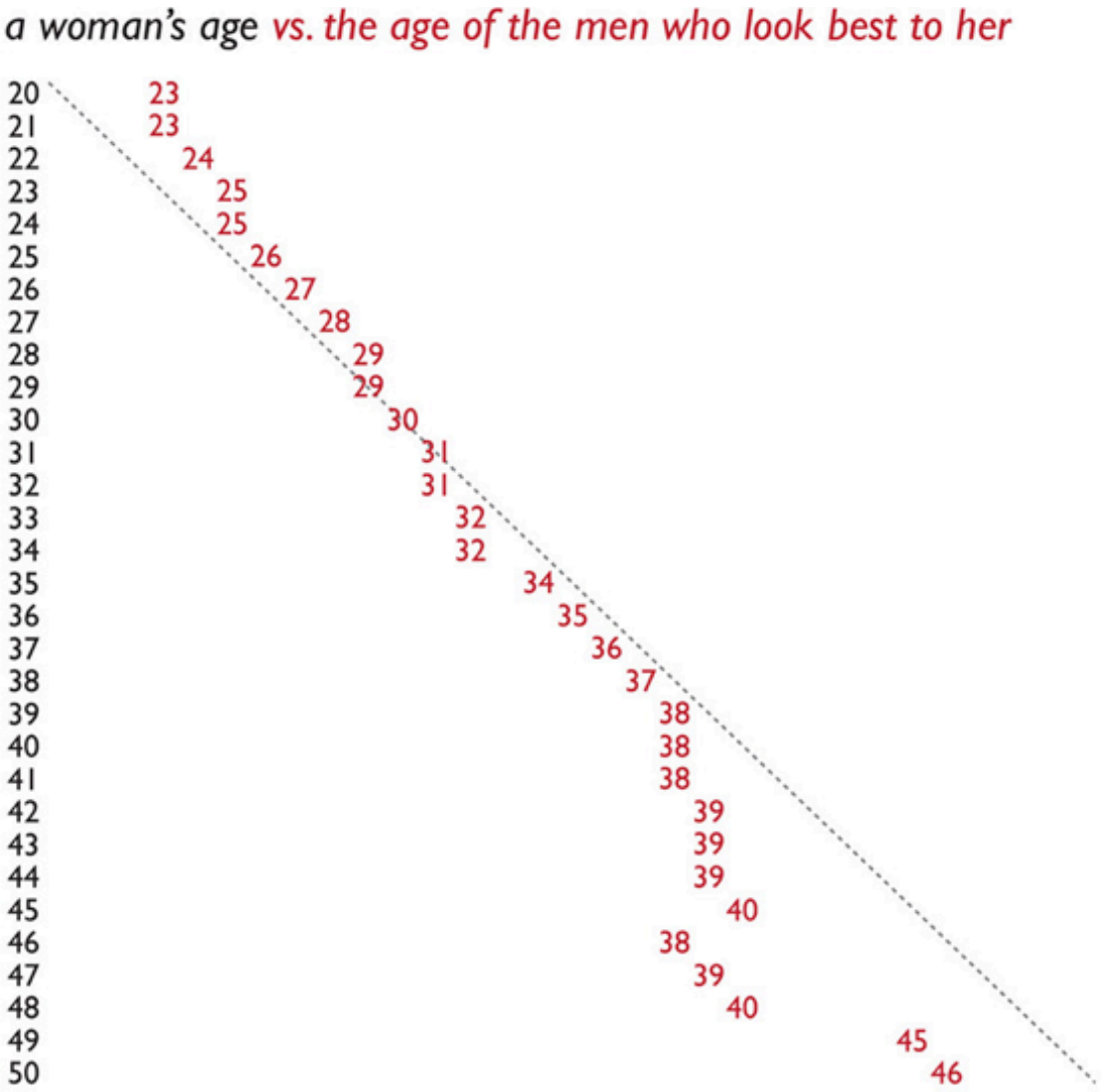
Lie Factor = Size of effect shown in the graphic / Size of effect in data

“Charts, graphs, maps, and diagrams don’t lie. People who design graphics do…

A graphic can *mislead* though.”

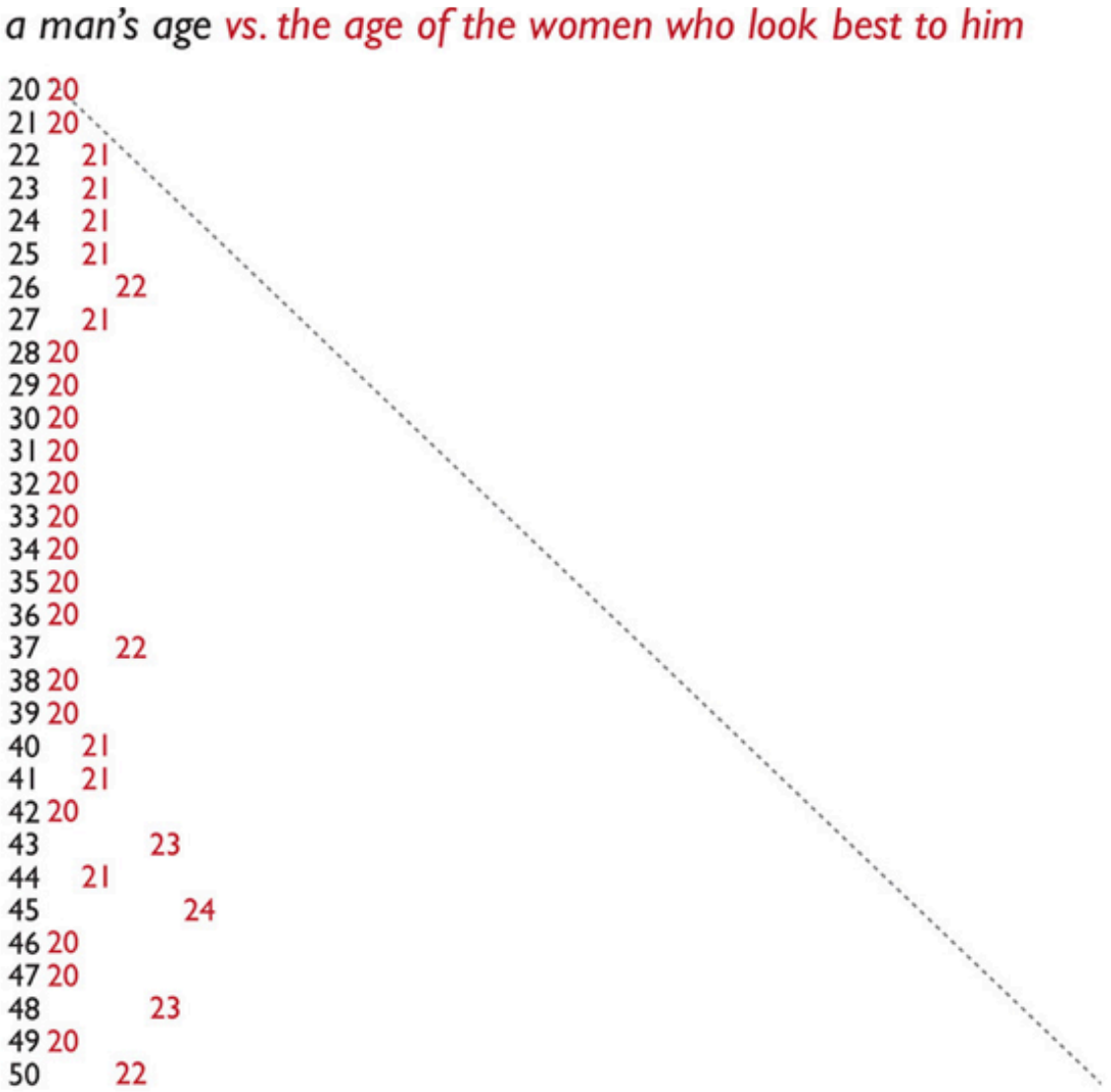
Alberto Cairo: http://www.thefunctionalart.com/search?q=ethics

Case study: OKCupid Ratings



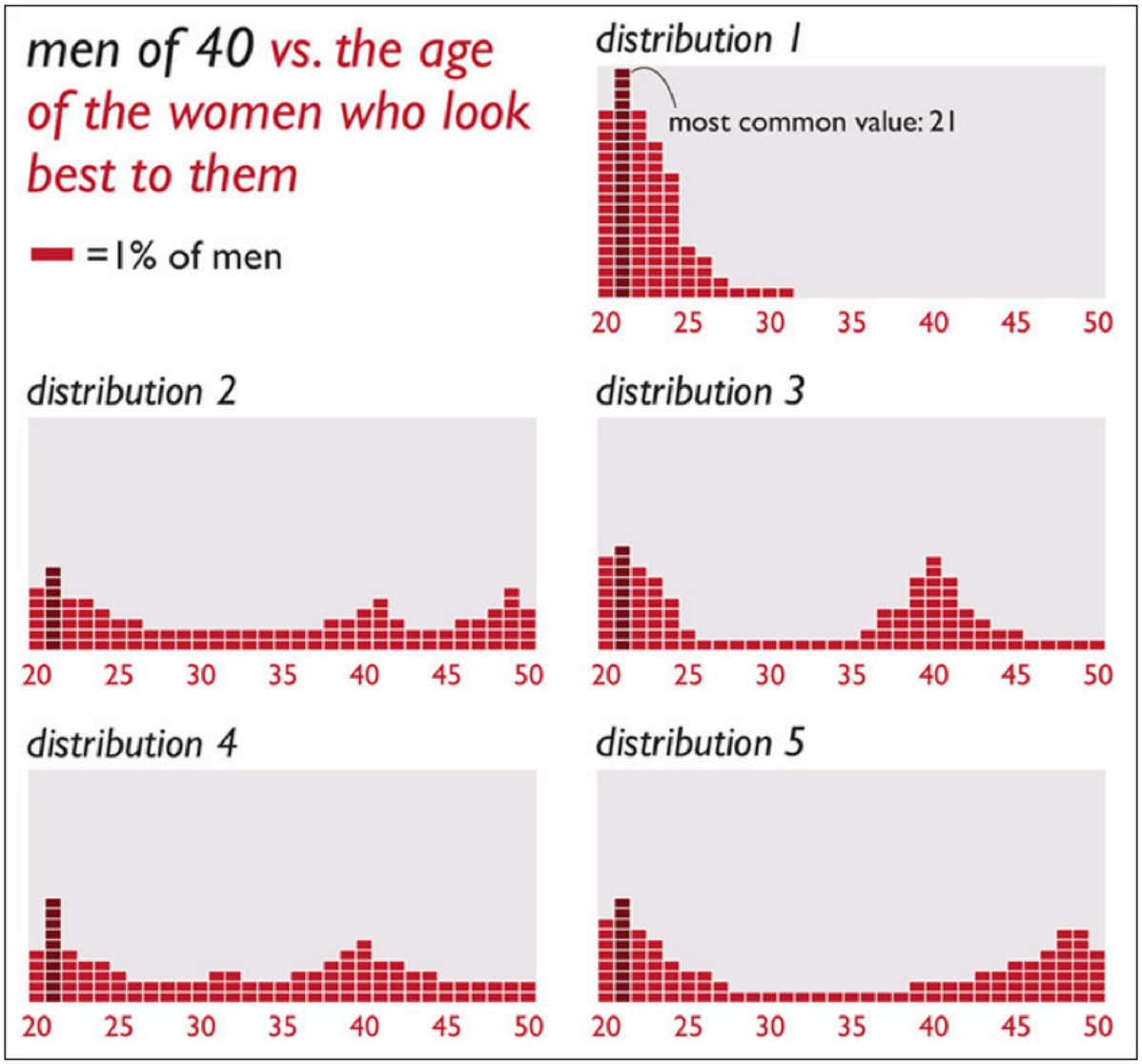
Christian Rudder, *Dataclysm: Who We Are (When We Think No One's Looking)*,
Cited in Alberto Cairo. *The Truthful Art*.

Case study: OKCupid Ratings



Christian Rudder, *Dataclysm: Who We Are (When We Think No One's Looking)*,
Cited in Alberto Cairo. *The Truthful Art*.

Case study: OKCupid Ratings



Christian Rudder, *Dataclysm: Who We Are (When We Think No One's Looking)*,
Cited in Alberto Cairo. *The Truthful Art*.

Information	**Definitely** true **Vitally** important
Uninformation	**Probably** not *untrue* **Probably** not important **Possibly** interesting
Noninformation	**Possibly** true **Possibly** not true **Probably** not important **Possibly** confusing
Misinformation	**Definitely** not true *Important to avoid!*
Disinformation	**Deliberately** not true Used vengefully or tactically to mislead

"Designing Information", by Joel Katz, p.14

“You must not fool yourself, and you are the easiest person to fool”

Richard Feynman

Data Collection

reliable sources|verifiable |cited

Data Analysis

minimize or annotate assumptions

Design

accurate portrayal of data *and* analysis
use hierarchy of visual properties appropriately
share your visualization for critical feedback

Modified from the Visual.ly “Code of Ethics”

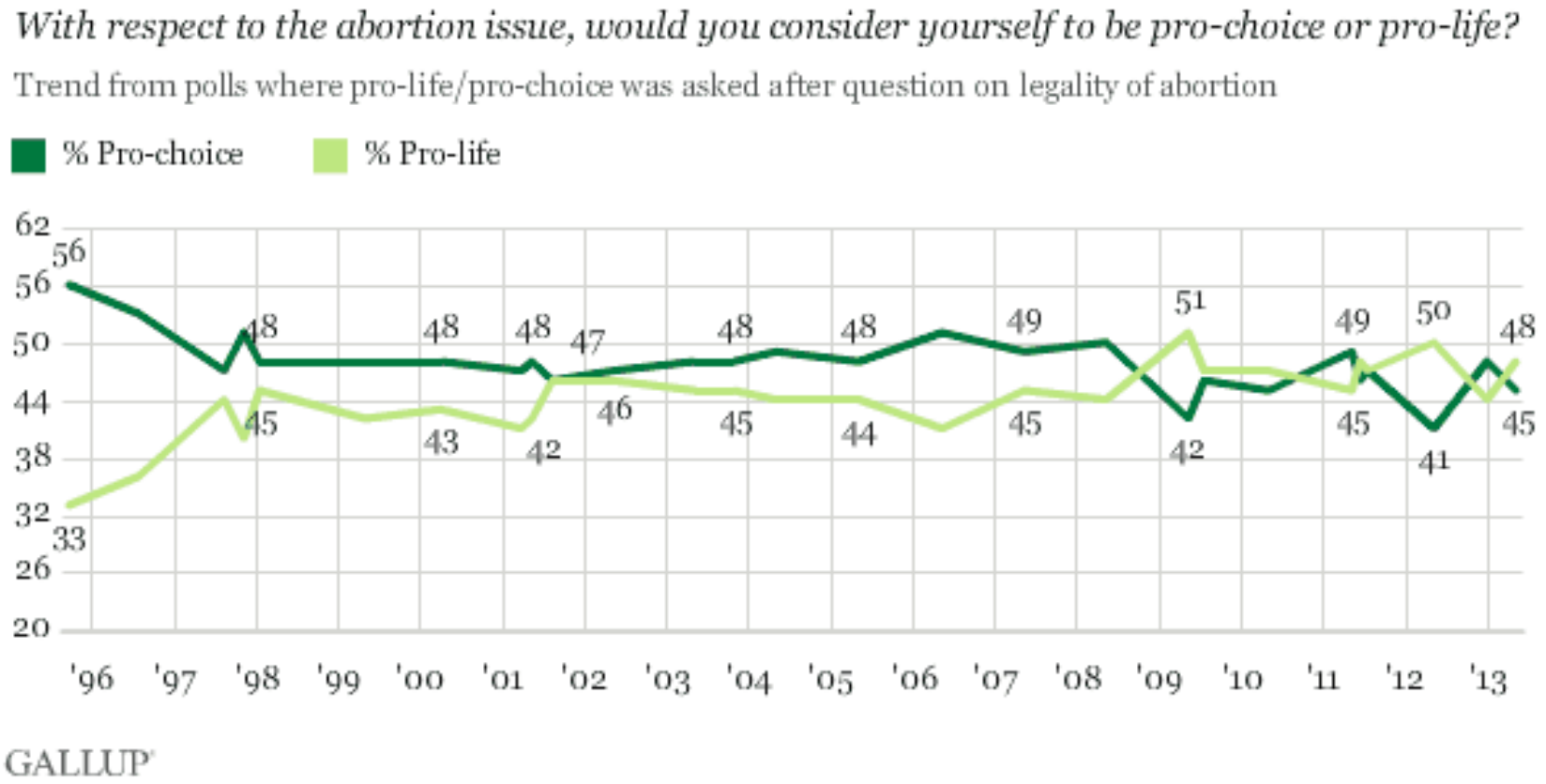
Gallup Abortion Poll: Numbers

With respect to the abortion issue, would you consider yourself to be pro-choice or pro-life?

Trend from polls in which pro-life/pro-choice was asked after question on legality of abortion

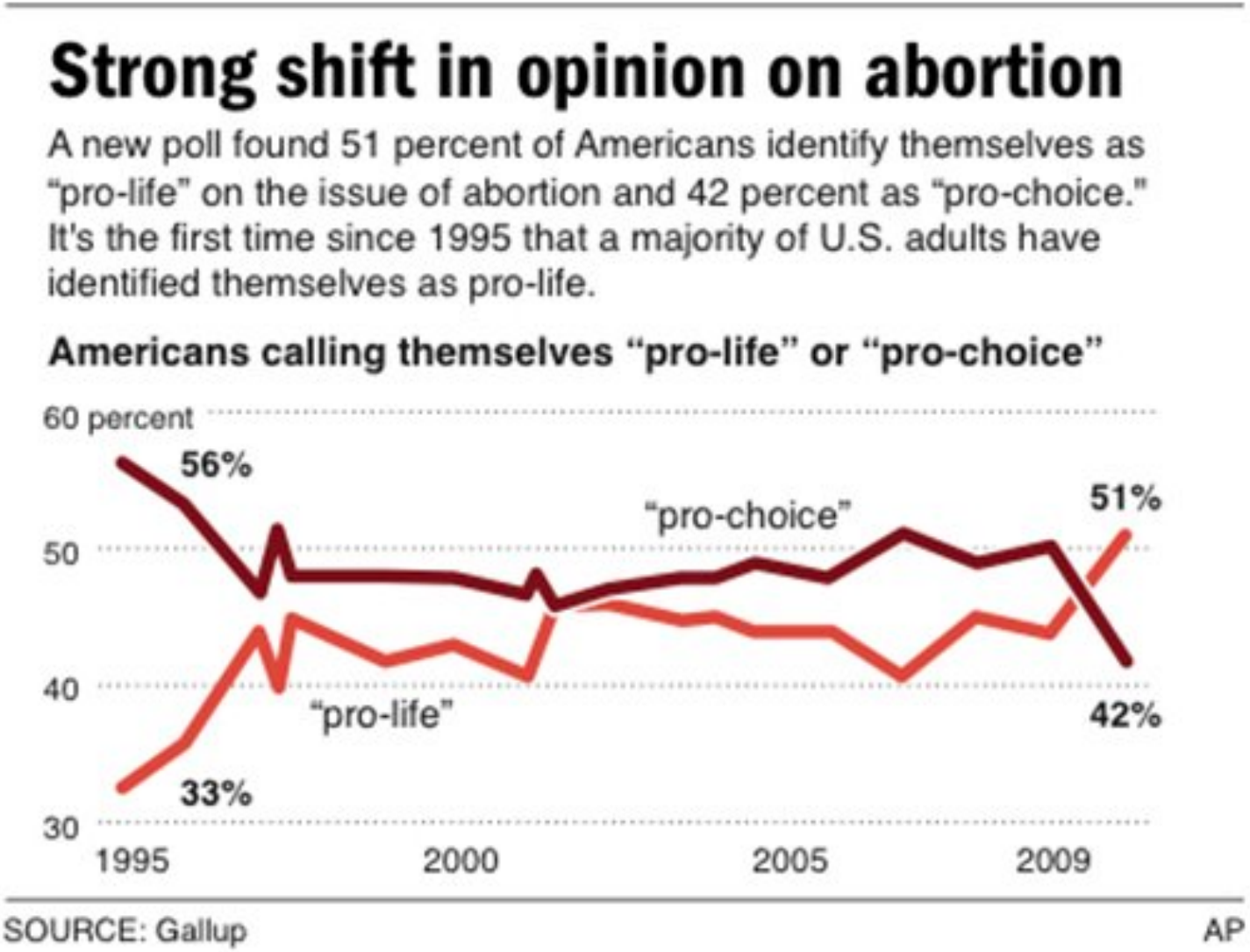
	Pro-choice	Pro-life	Mixed/ Neither (vol.)	Don't know what terms mean (vol.)	No opinion
	%	%	%	%	%
2013 May 2-7	45	48	3	3	2
2012 Dec 27-30	48	44	3	4	1
2012 May 3-6	41	50	4	3	3
2011 Jul 15-17	47	47	2	3	1
2011 Jun 9-12	46	48	2	3	2
2011 May 5-8	49	45	3	2	2
2010 May 3-6	45	47	2	4	3
2009 Jul 17-19	46	47	3	2	2
2009 May 7-10	42	51	2	4	1
2008 May 8-11	50	44	1	2	2
2007 May 10-13	49	45	3	2	2
2006 May 8-11	51	41	4	2	3
2005 May 2-5	48	44	3	3	2
2004 May 2-4	49	44	2	3	2
2003 Oct 24-26	48	45	3	3	1
2003 May 5-7	48	45	2	4	1
2002 May 6-9	47	46	3	3	1
2001 Aug 10-12	46	46	3	3	2
2001 May 10-14	48	42	4	4	2
2001 Mar 26-28	47	41	4	6	2
2000 Mar 30-Apr 2	48	43	2	4	3
1999 Apr 30-May					

Gallup Abortion Poll: Their Graph



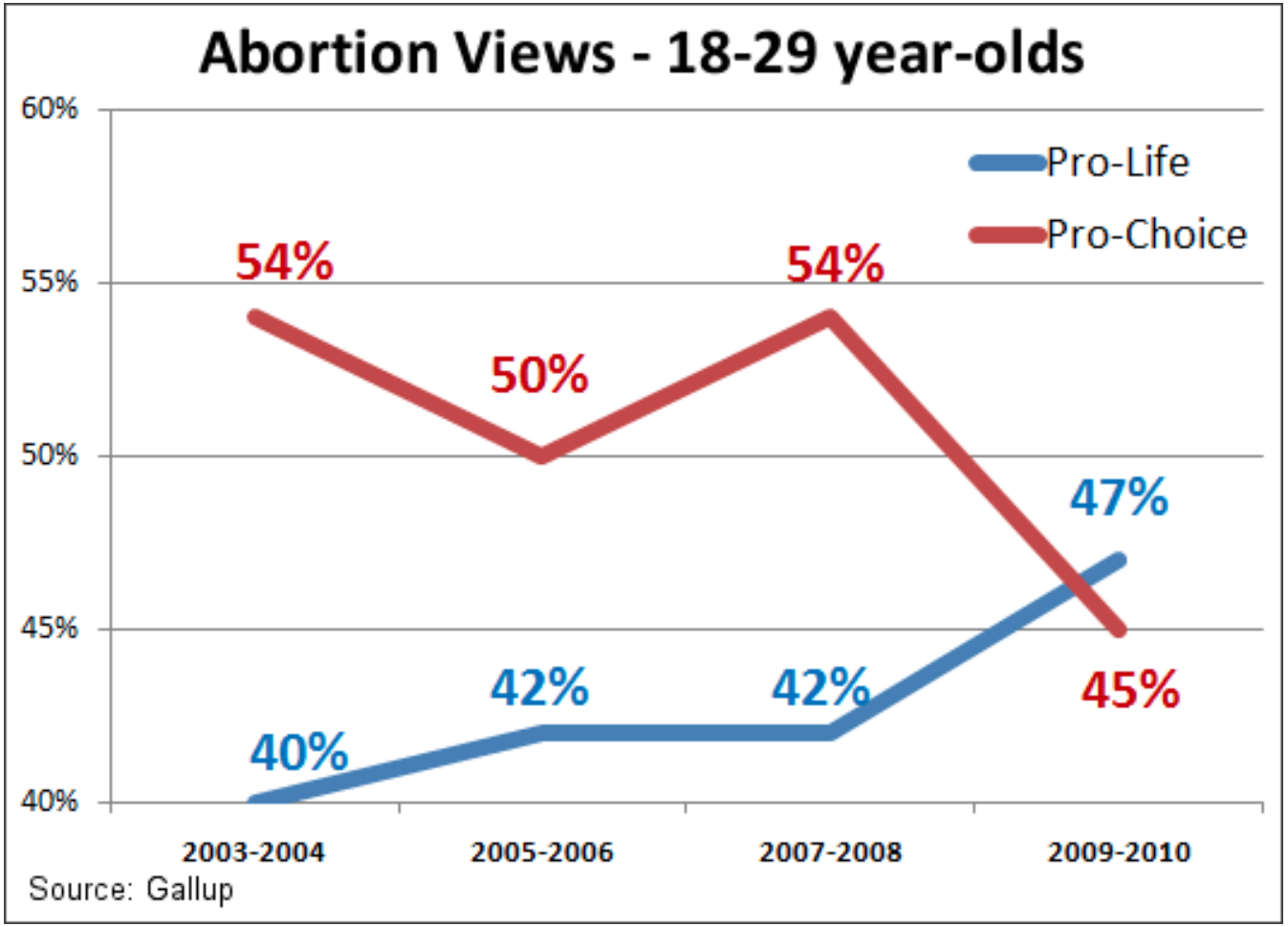
http://www.gallup.com/poll/1576/abortion.aspx

Gallup Abortion Poll: AP



https://visualisingadvocacy.org/blog/disinformation-visualization-how-lie-datavis

Gallup Abortion Poll: LiveAction.org



https://visualisingadvocacy.org/blog/disinformation-visualization-how-lie-datavis

Gallup Abortion Poll: LiveCitizen.com

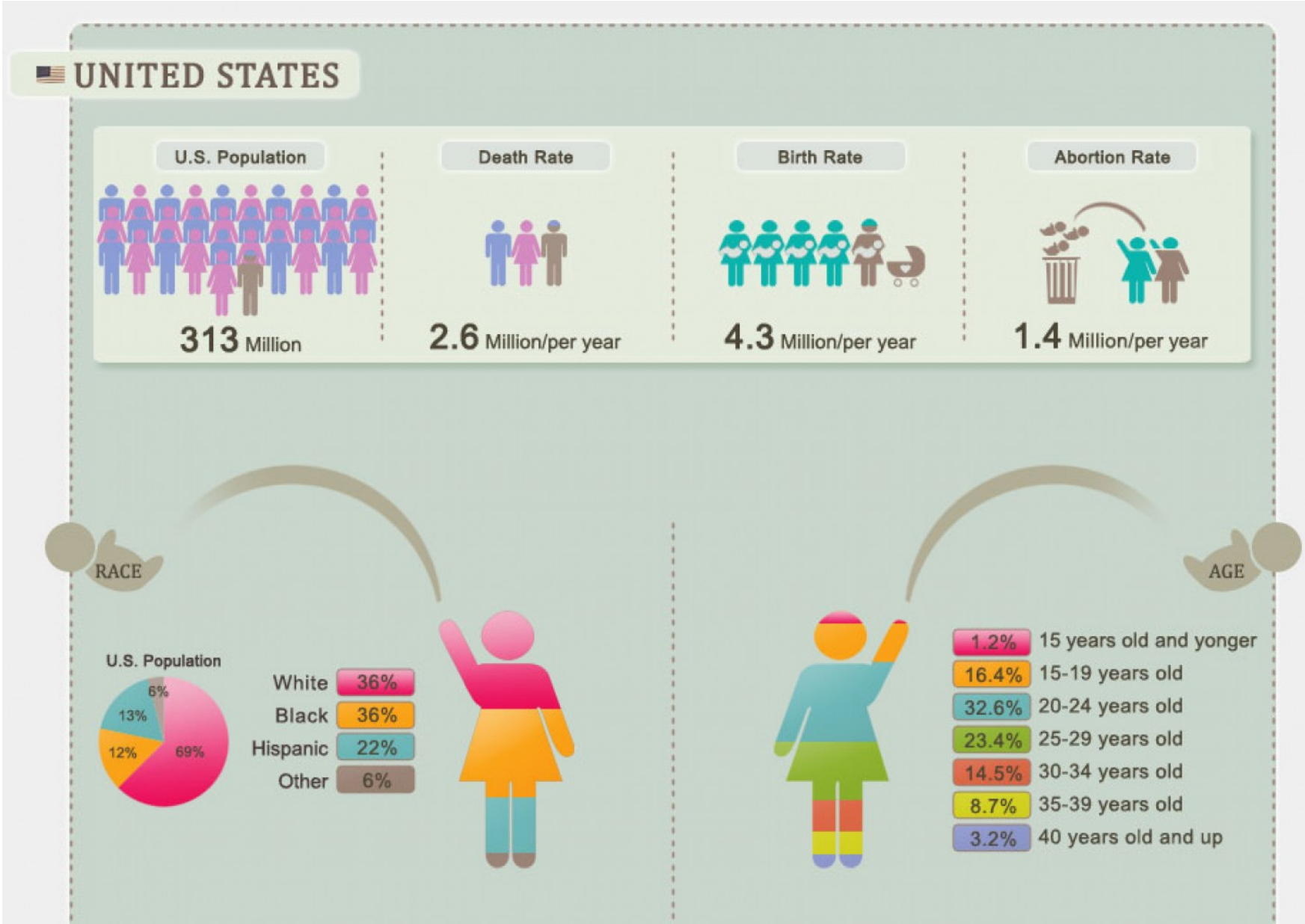
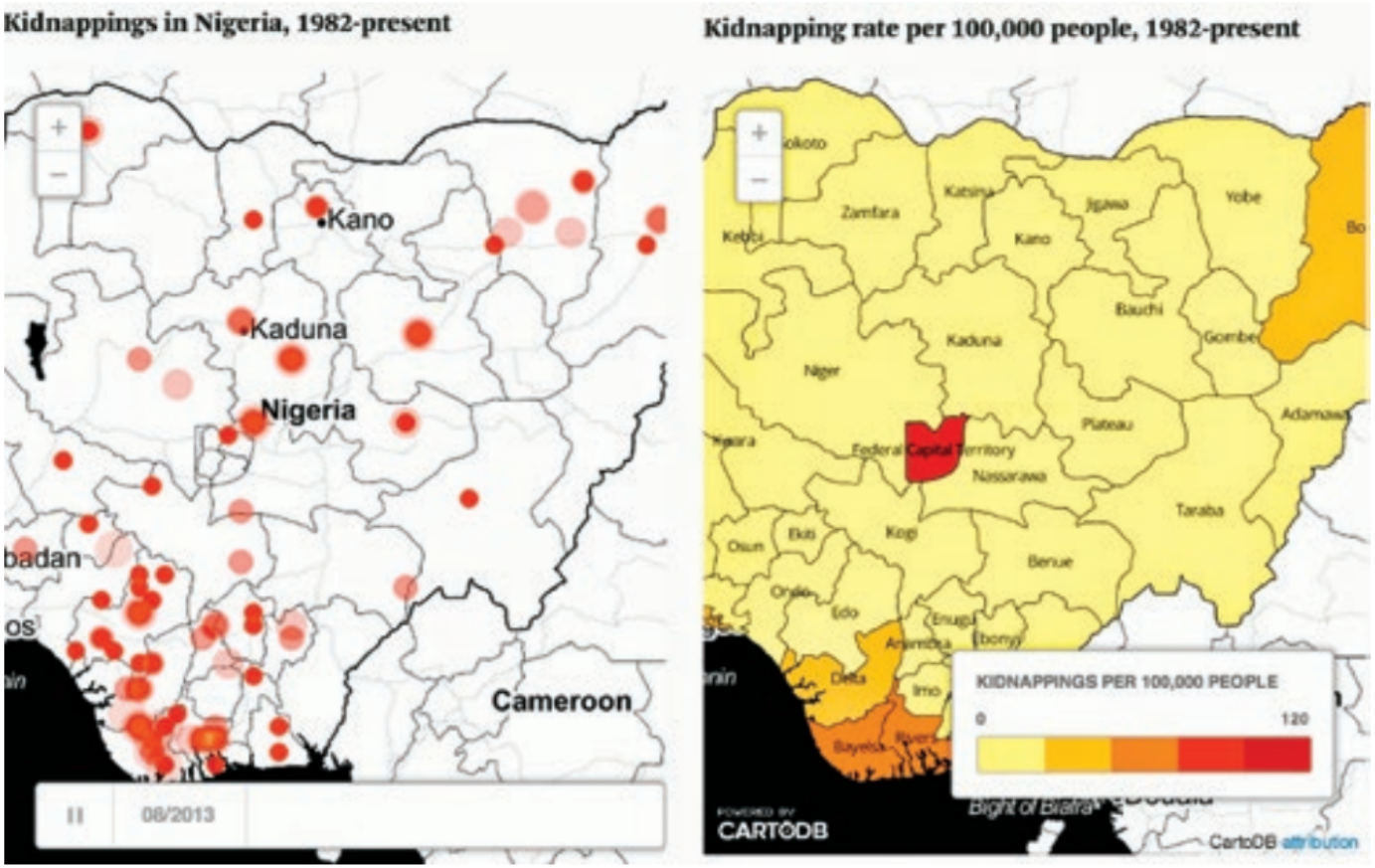


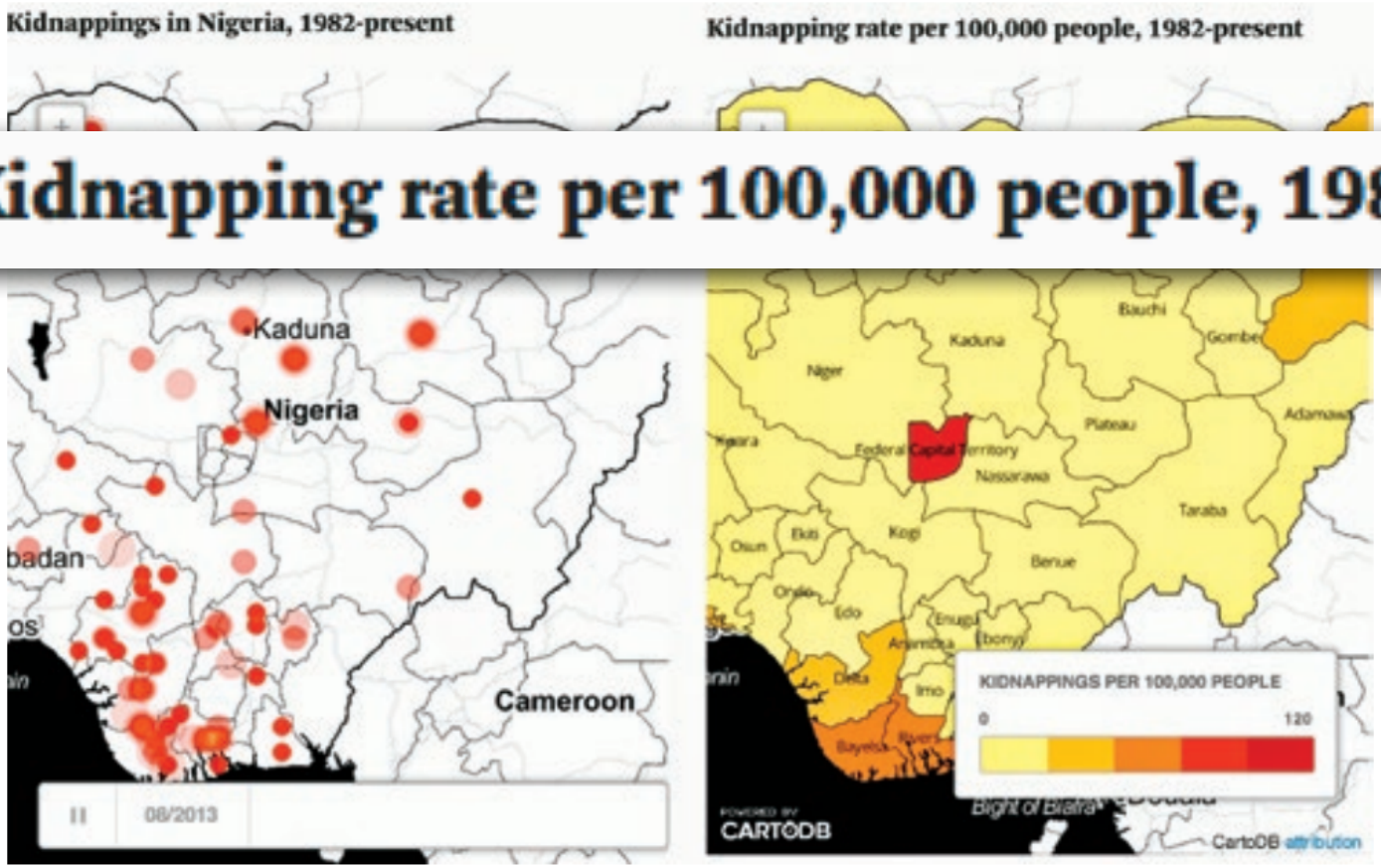
Figure 1



FiveThirtyEight used data from the Global Database of Events, Language and Tone, which tracks news stories about kidnappings, to create both graphics.

FiveThirtyEight

Figure 1

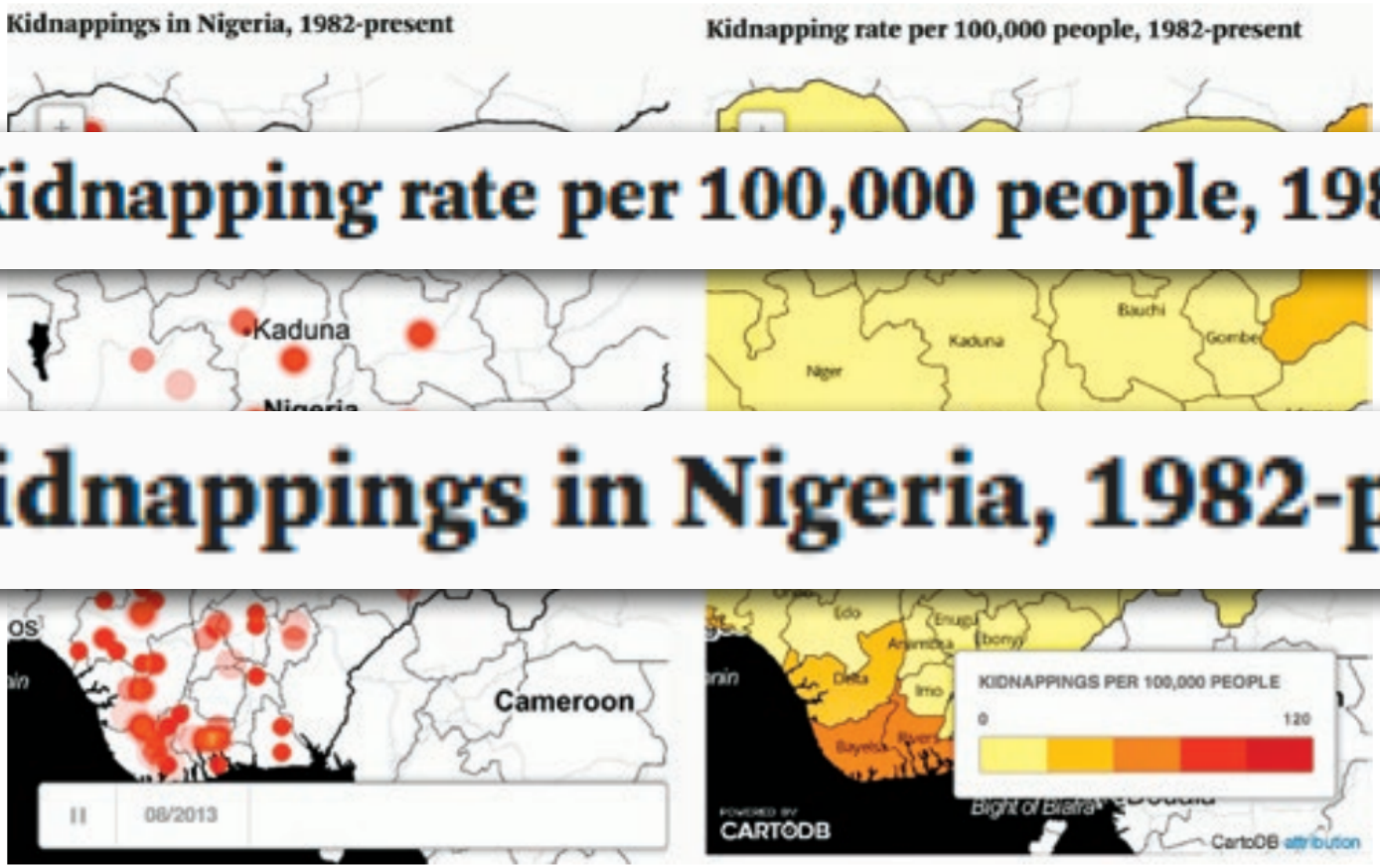


Kidnapping rate per 100,000 people, 1982-present

FiveThirtyEight used data from the Global Database of stories about kidnappings, to create both graphics.

FiveThirtyEight

Figure 1



FiveThirtyEight used data from the Global Database of stories about kidnappings, to create both graphics.

Kidnapping rate per 100,000 people, 1982-present

Kidnappings in Nigeria, 1982-present

The problem: their unit of analysis was “media report in Nigeria about kidnapping”, not “kidnapping”.

Many things could affect how many articles there were about kidnapping in a given time period...

... only one of which is the real item of interest for the map (actual kidnappings)!

“The article repeatedly refers to the number and location of kidnappings. But the Global Database of Events, Language and Tone (GDELT) — the data source for the article — is a repository of media reports, not discrete events. As such, we should only have referred to “media reports of kidnappings,” not kidnappings.”

http://fivethirtyeight.com/datalab/mapping-kidnappings-in-nigeria/

“The animated map also incorrectly locates some reported kidnappings. If the location of a reported kidnapping isn’t in a media report, GDELT defaults the location to the center of Nigeria. So that part of the country is overrepresented in the animated map.“

“This piece did not meet FiveThirtyEight’s standards for publication. We apologize for the mistakes. We will do better. The original article follows below.”

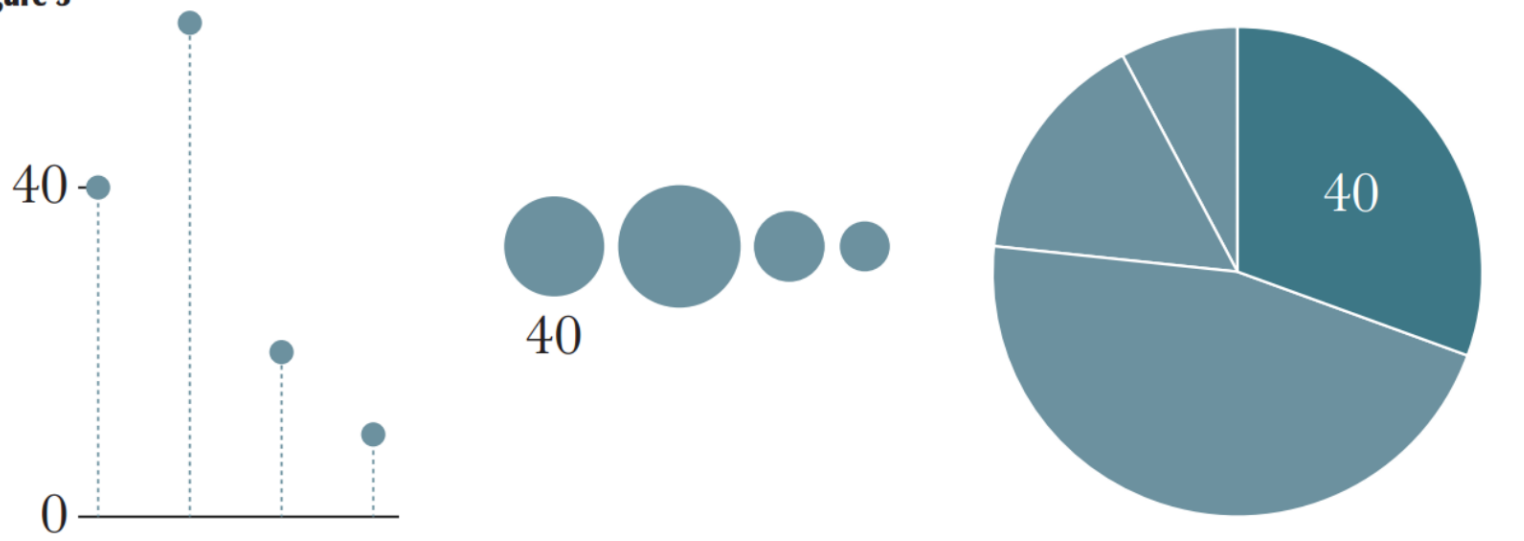
http://fivethirtyeight.com/datalab/mapping-kidnappings-in-nigeria/

Lies in Area

Symbolically Speaking

A major part of the problem…

Figure 3



Try to estimate the values represented on each chart based on the only value shown. You can probably do this quite well on the dot plot, but not on the bubble chart or the pie chart. The values are 40, 60, 20 and 10.

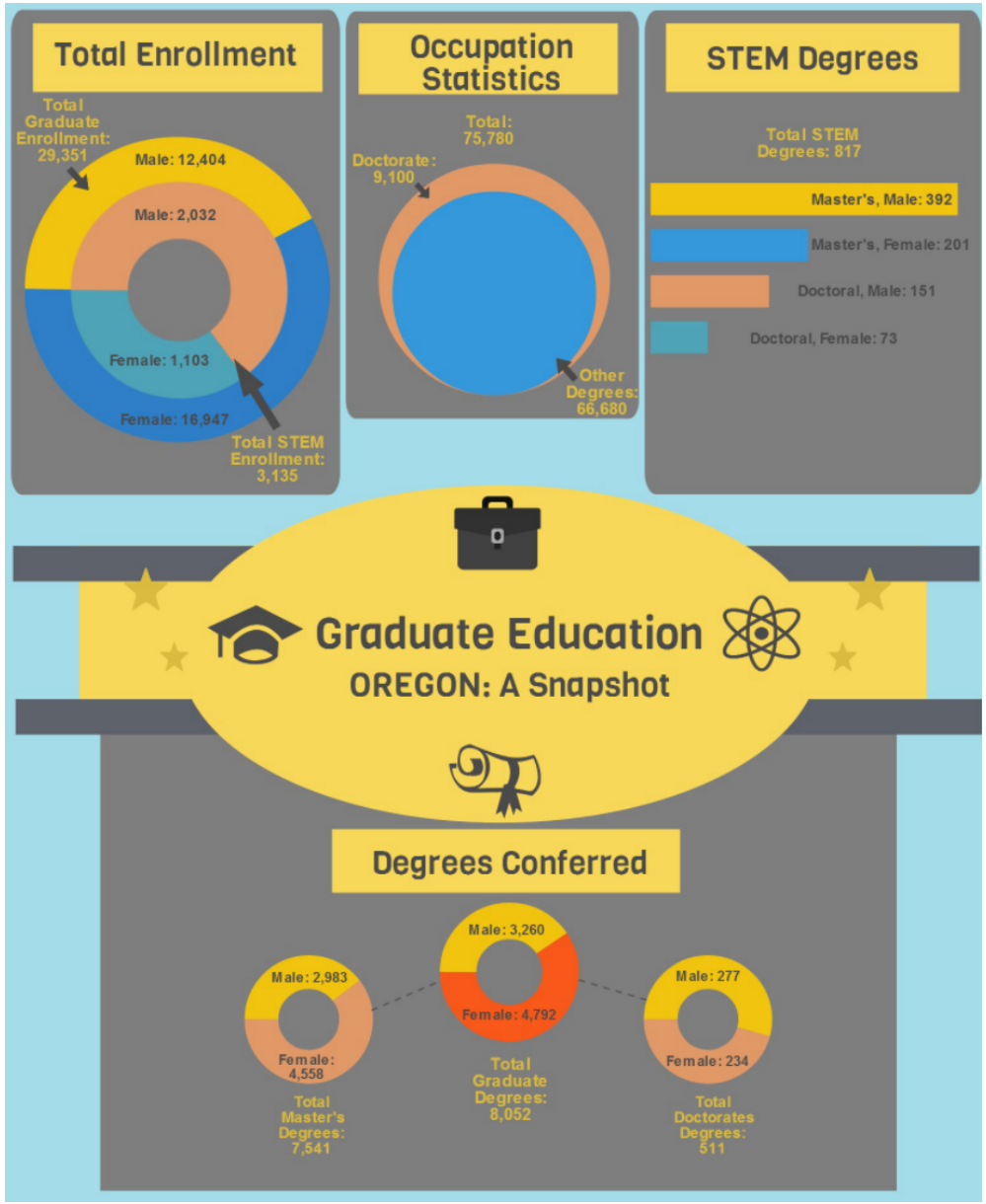
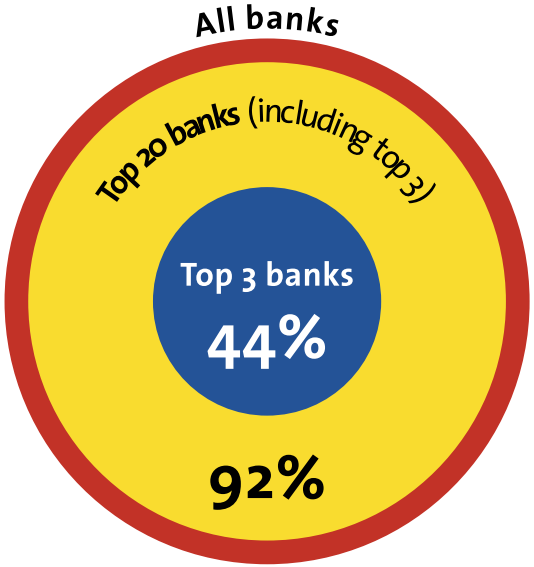
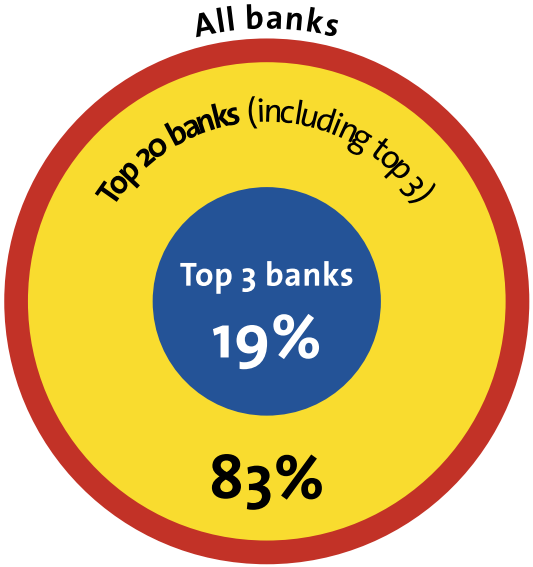


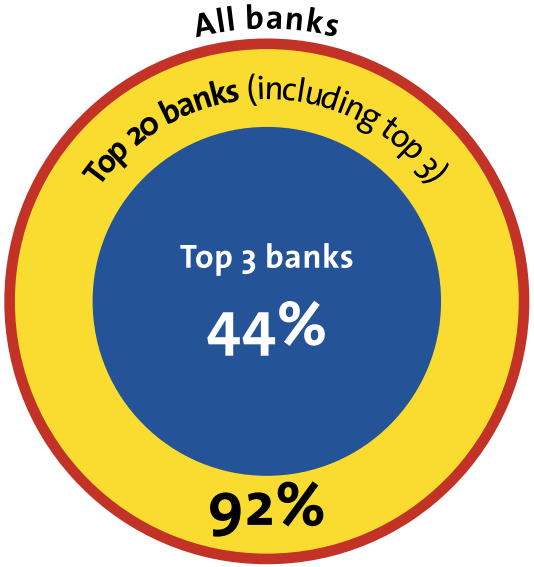
Image from the Conference of Oregon Deans



In *The New York Times*



Actual percentages of *The New York Times* diagrams



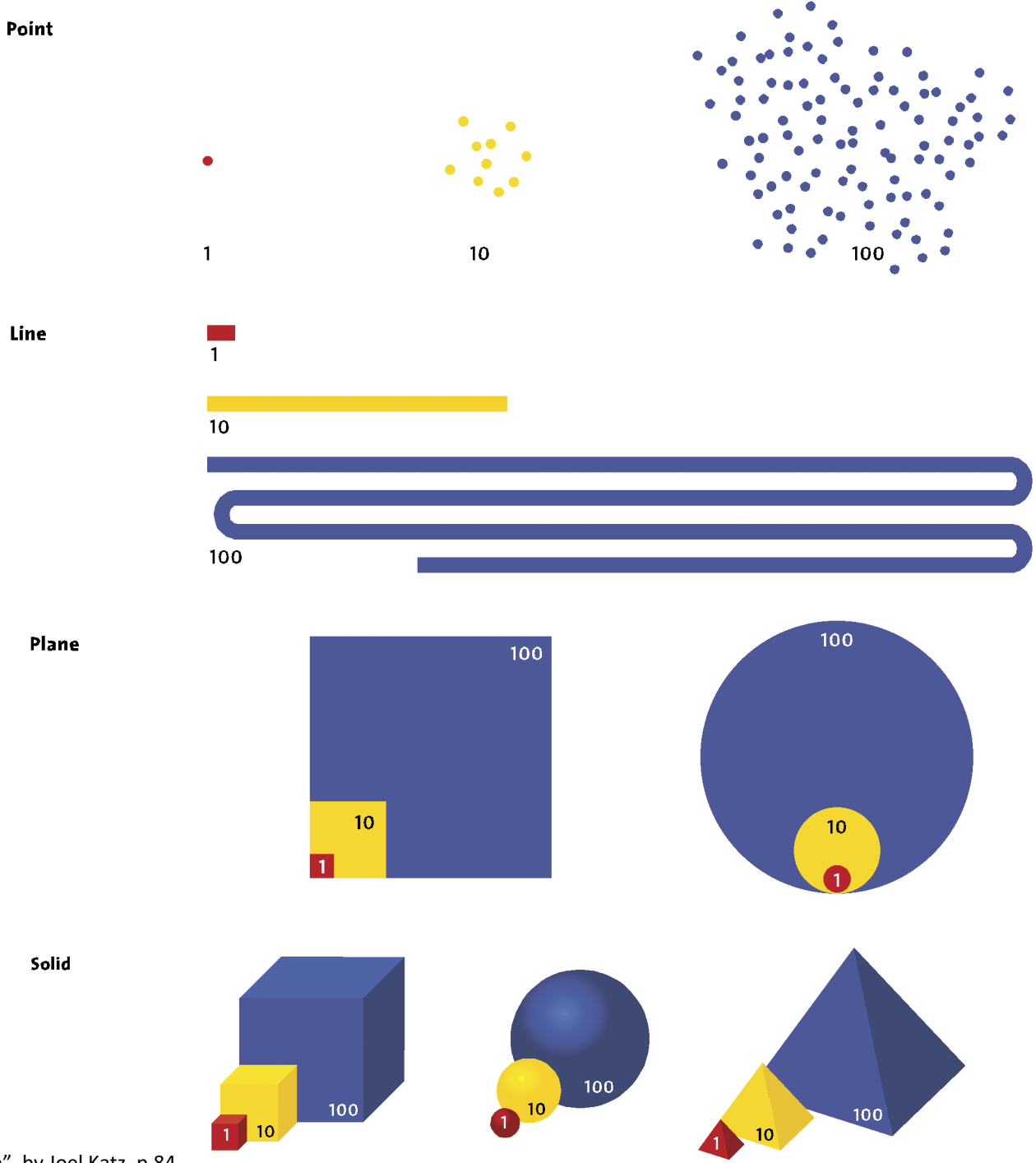
Correct diagram of *The New York Times* data

Sixth-grade math department. *The left circle is a reconstruction of a diagram in* The New York Times, *showing market share among U.S. banks. Unfortunately, the graphic is incorrect, with the percentage relationships determined by diameter rather than by area (πr^2).*

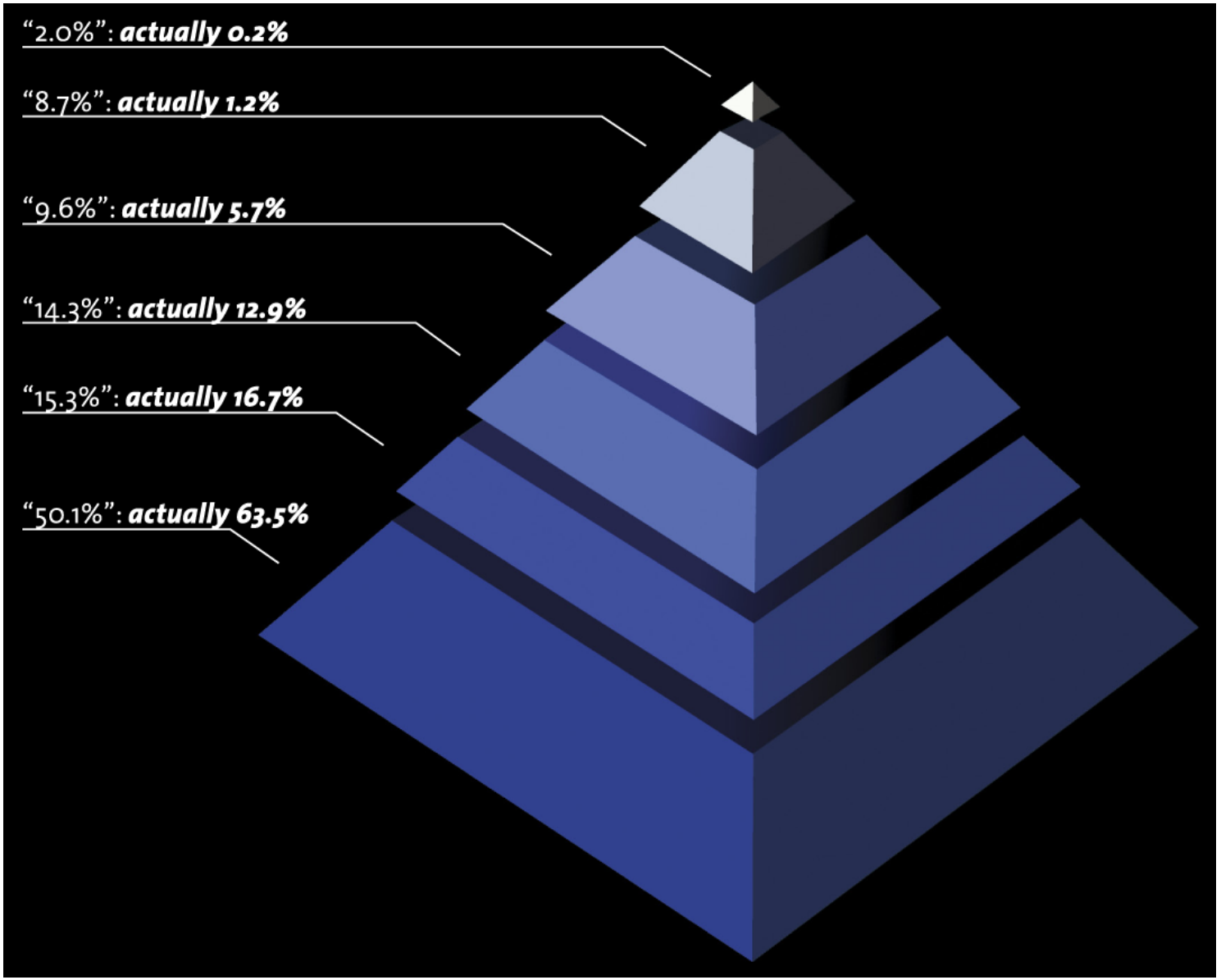
The true percentages represented by The New York Times*'s circles are shown in the center graphic; a graphic accurately representing the percentages is at the right.*

"Designing Information", by Joel Katz, p.85

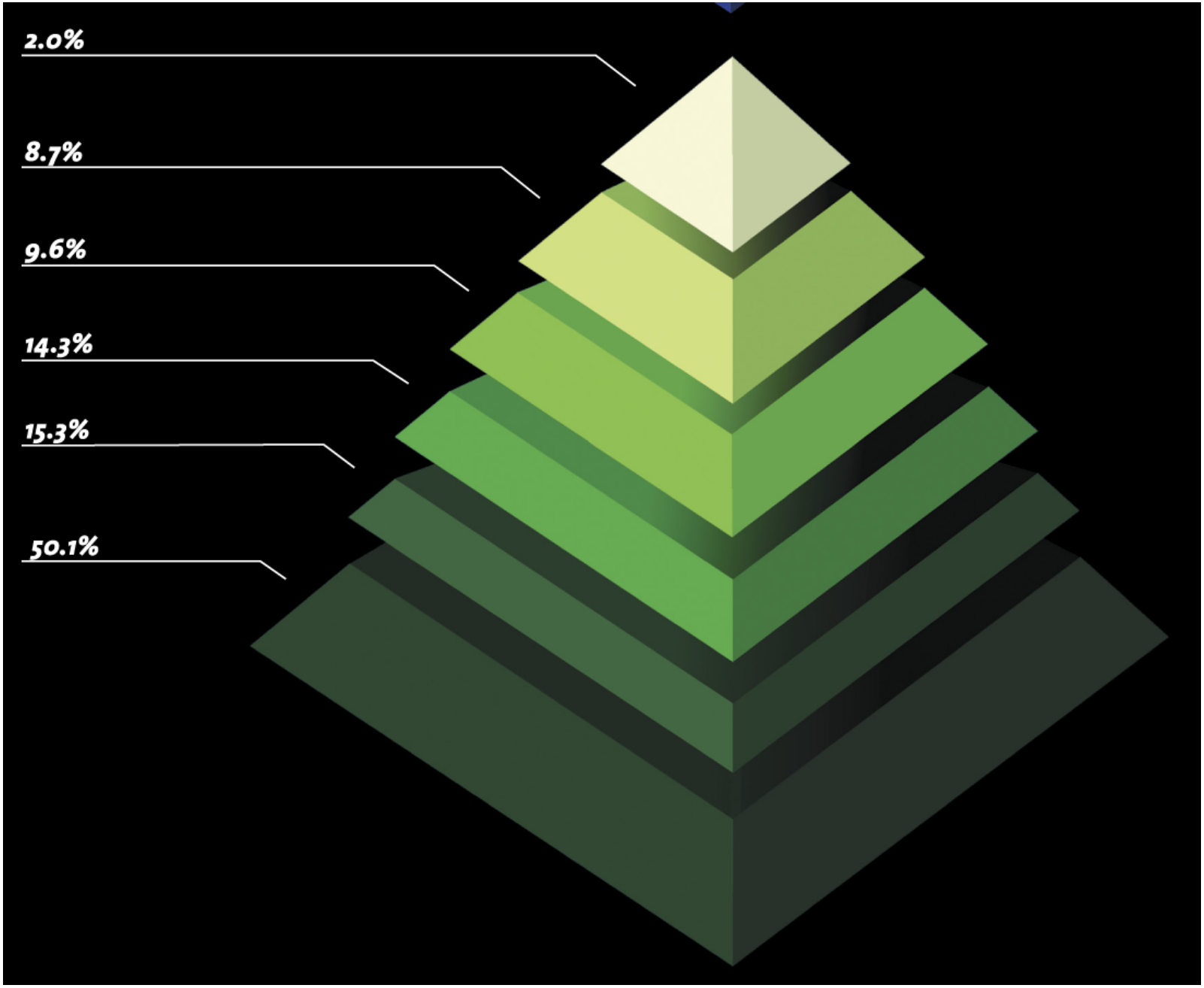
"Designing Information", by Joel Katz, p.85



"Designing Information", by Joel Katz, p.84



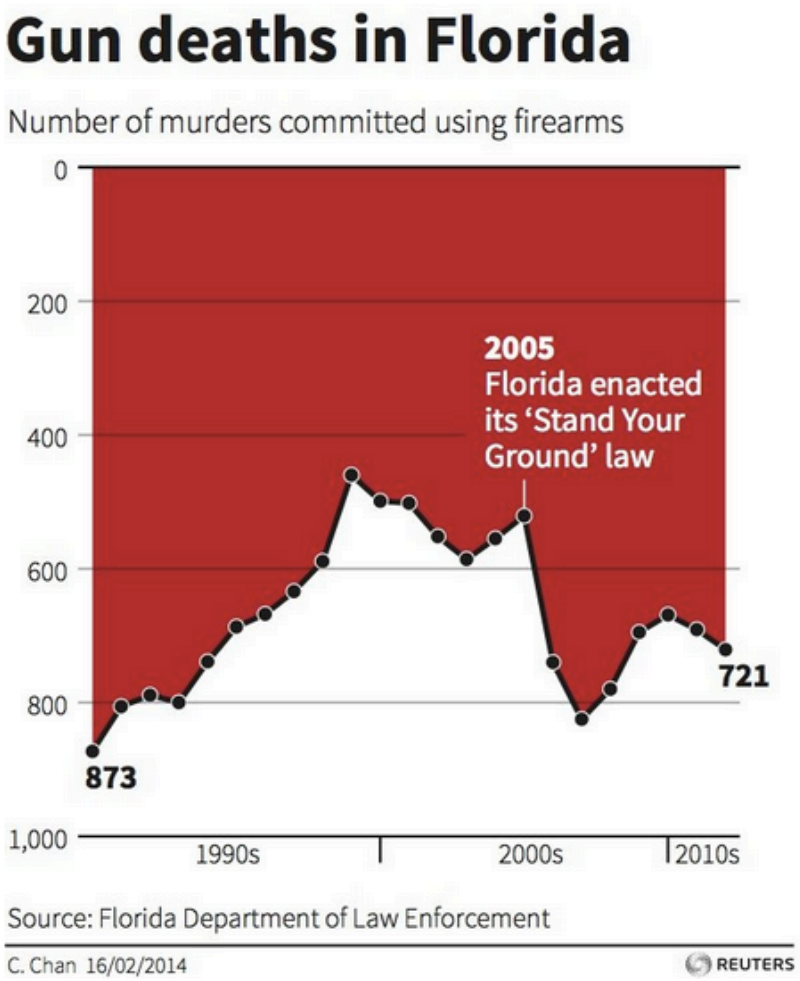
"Designing Information", by Joel Katz, p.86

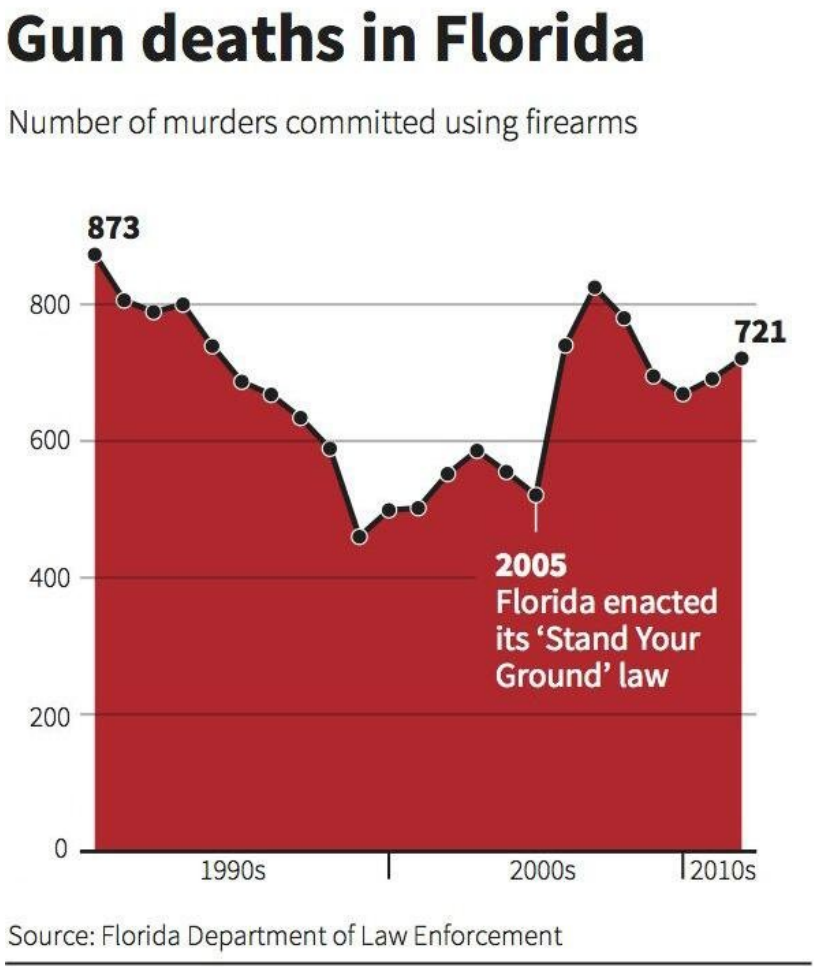


“Designing Information”, by Joel Katz, p.86

Lies in Scale

Don’t be a jerk

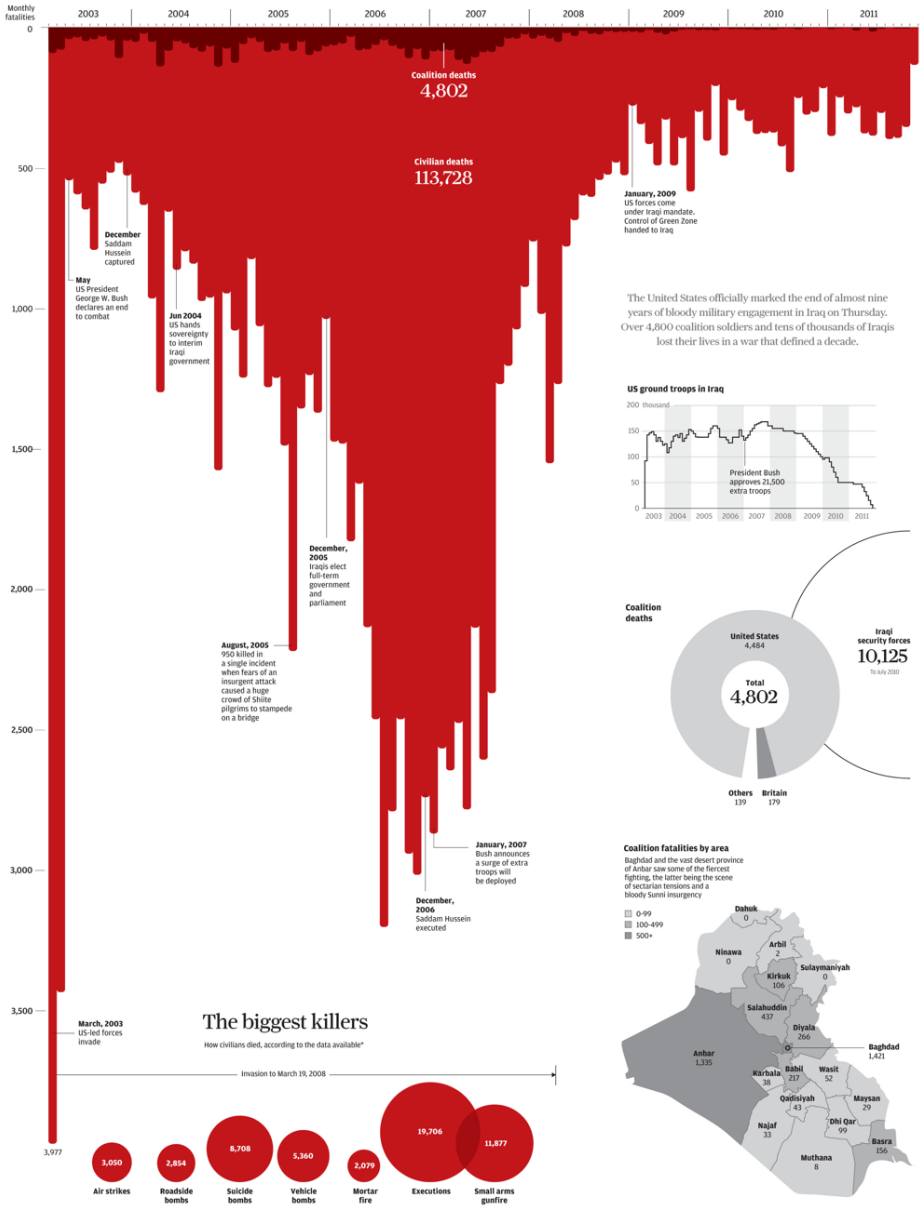




http://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2

Simon Scarr, South China Morning Post, 12/17/2011

Iraq's bloody toll



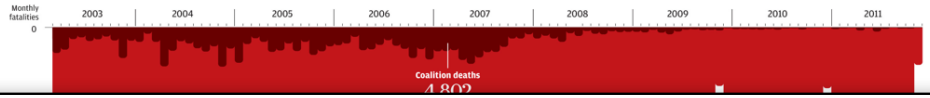
SCMP graphic: Simon Scarr

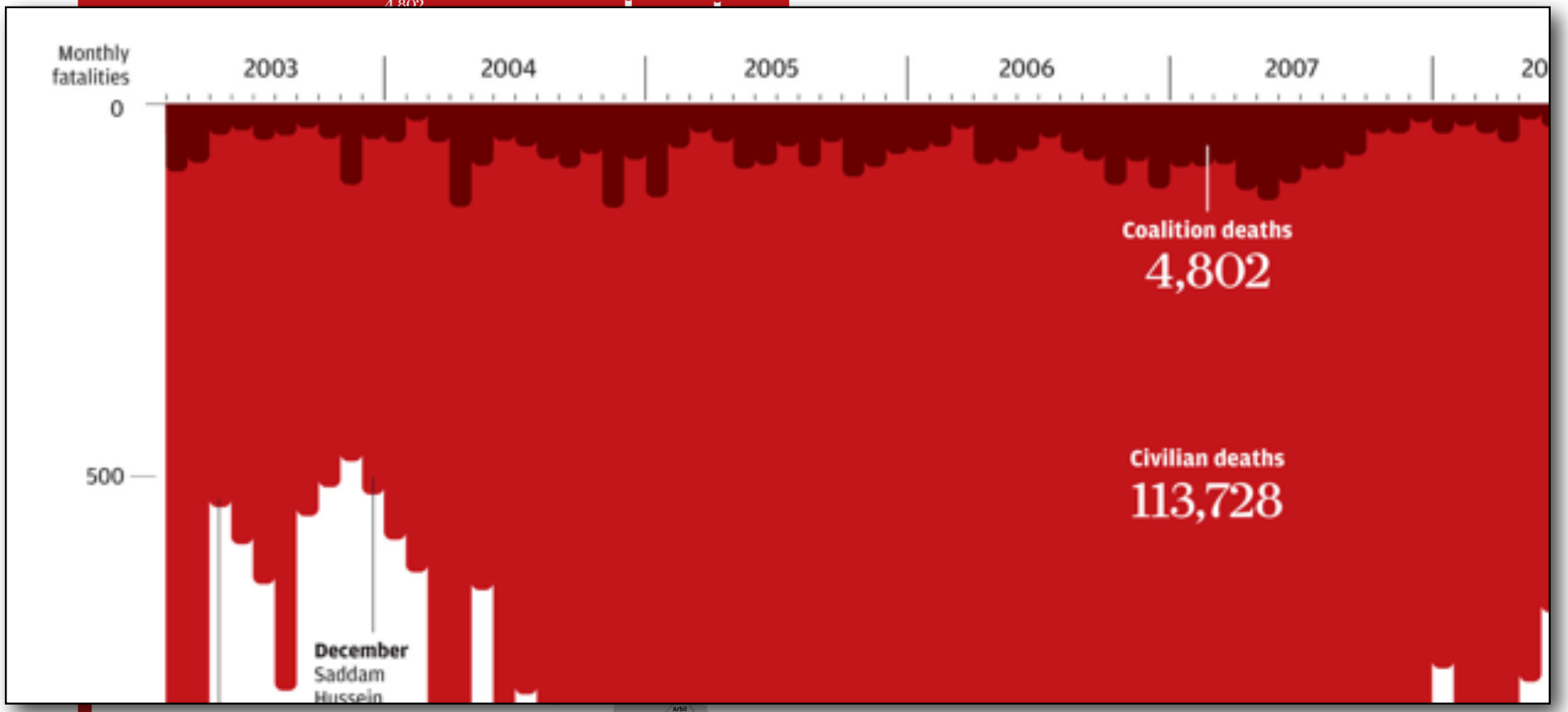
*Deaths from unknown causes are not included. Causes of death accounting for less than 0.5 per cent of killings also not included.

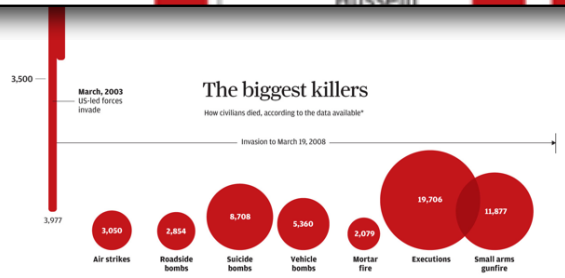
Sources: iCasualties, Iraq Body Count, New England Journal of Medicine, Global Security, Brookings Institution

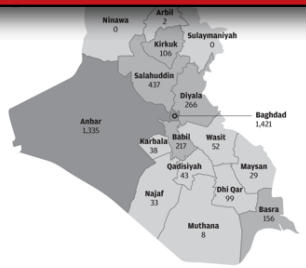
Simon Scarr, South China Morning Post, 12/17/2011

Iraq's bloody toll





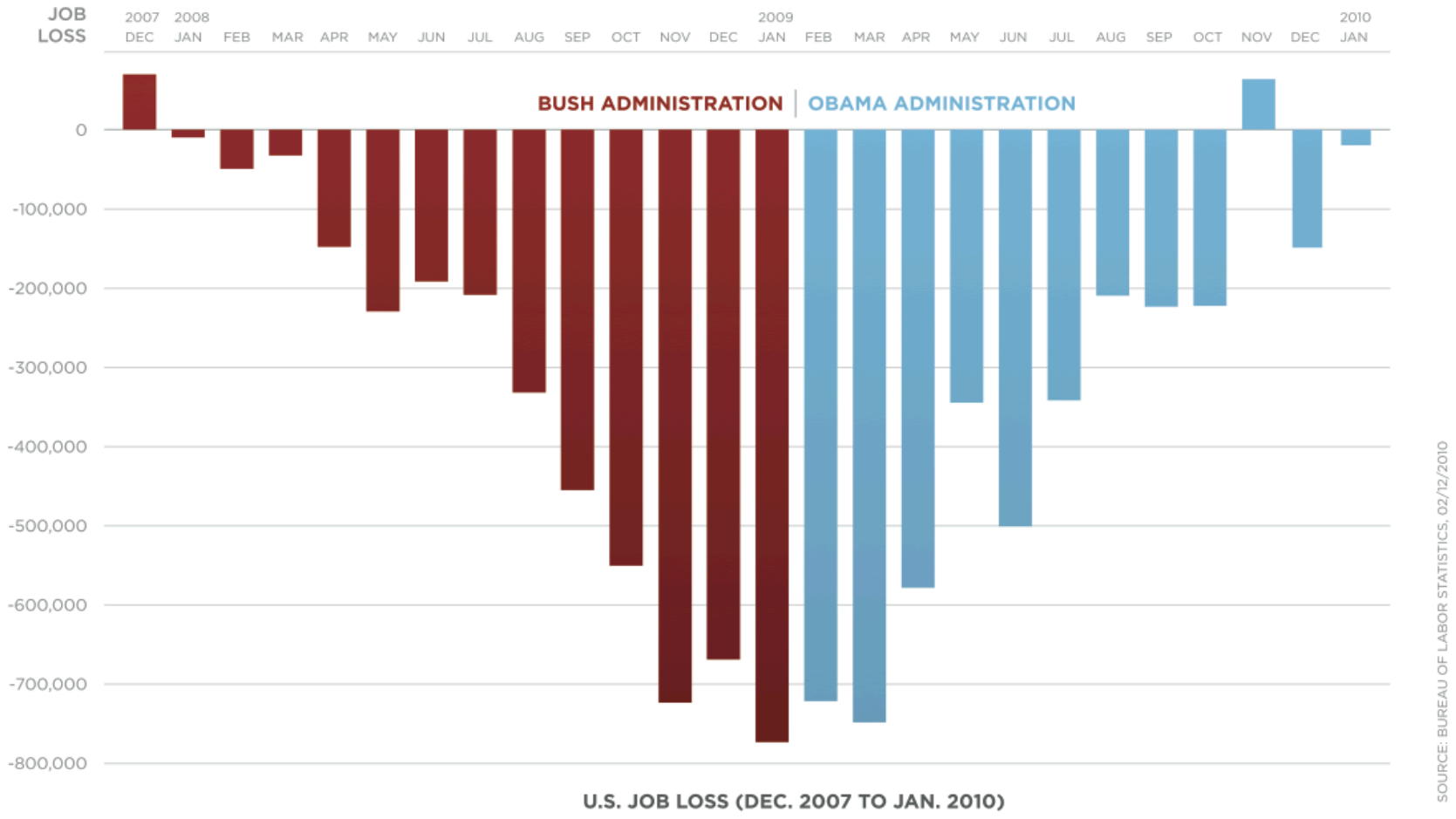




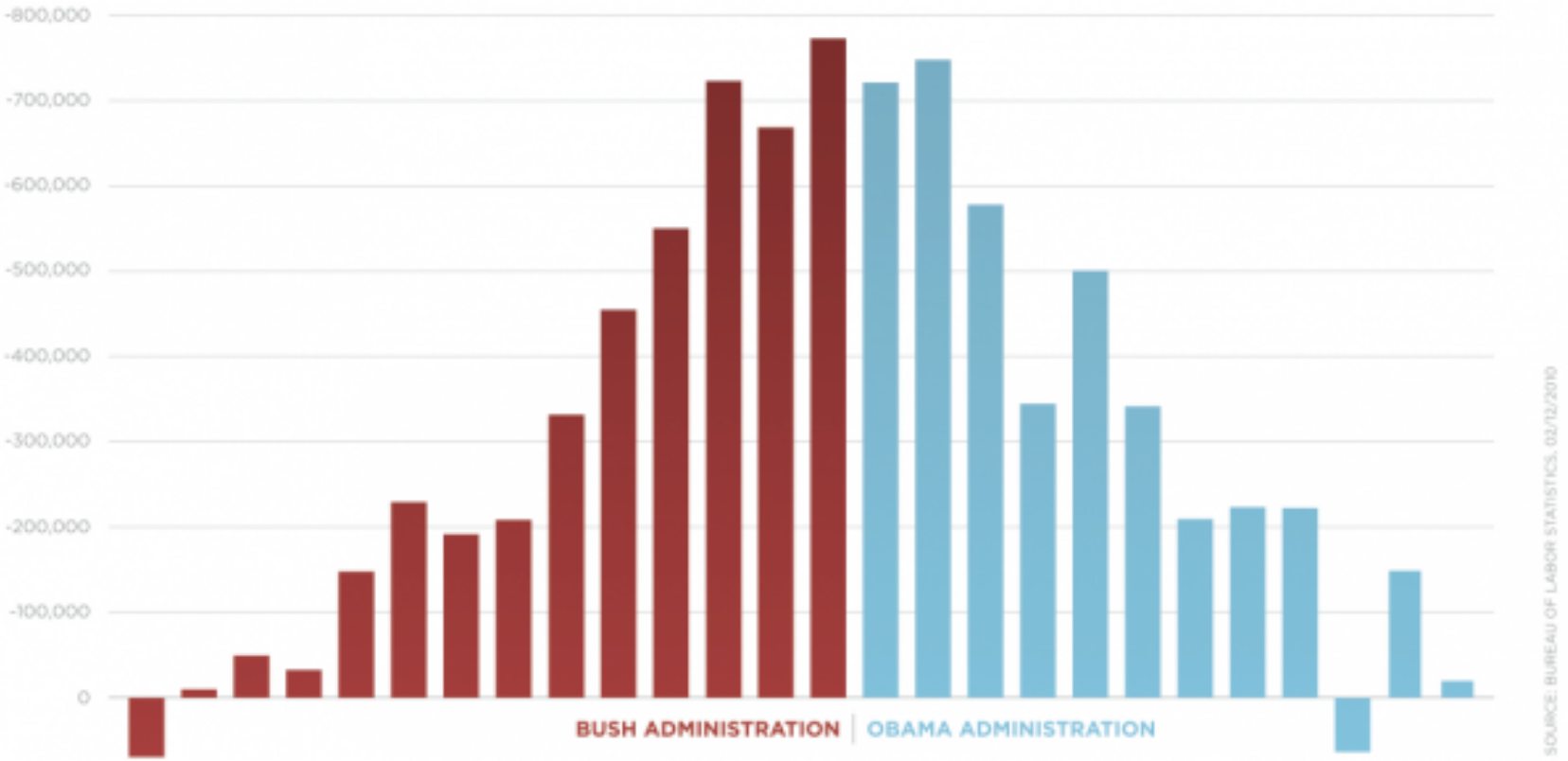
SCMP graphic: Simon Scarr

*Deaths from unknown causes are not included. Causes of death accounting for less than 0.5 per cent of killings also not included.

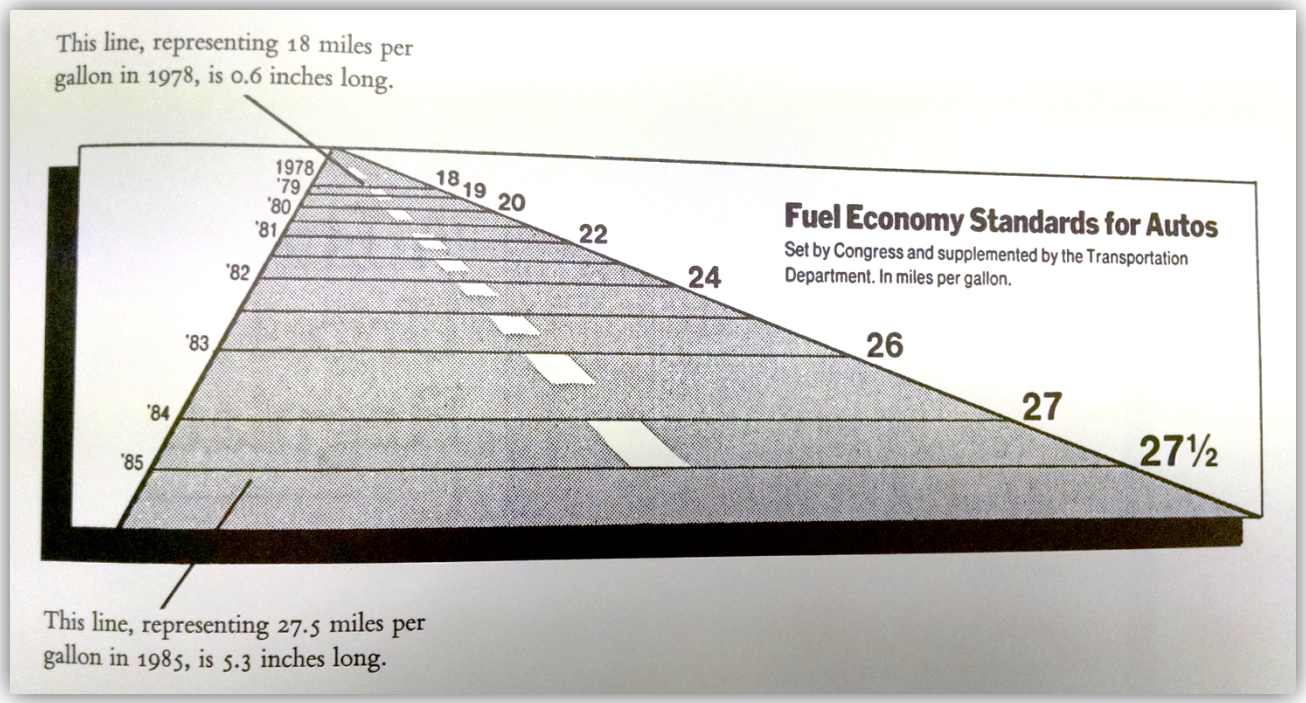
Sources: iCasualties, Iraq Body Count, New England Journal of Medicine, Global Security, Brookings Institution



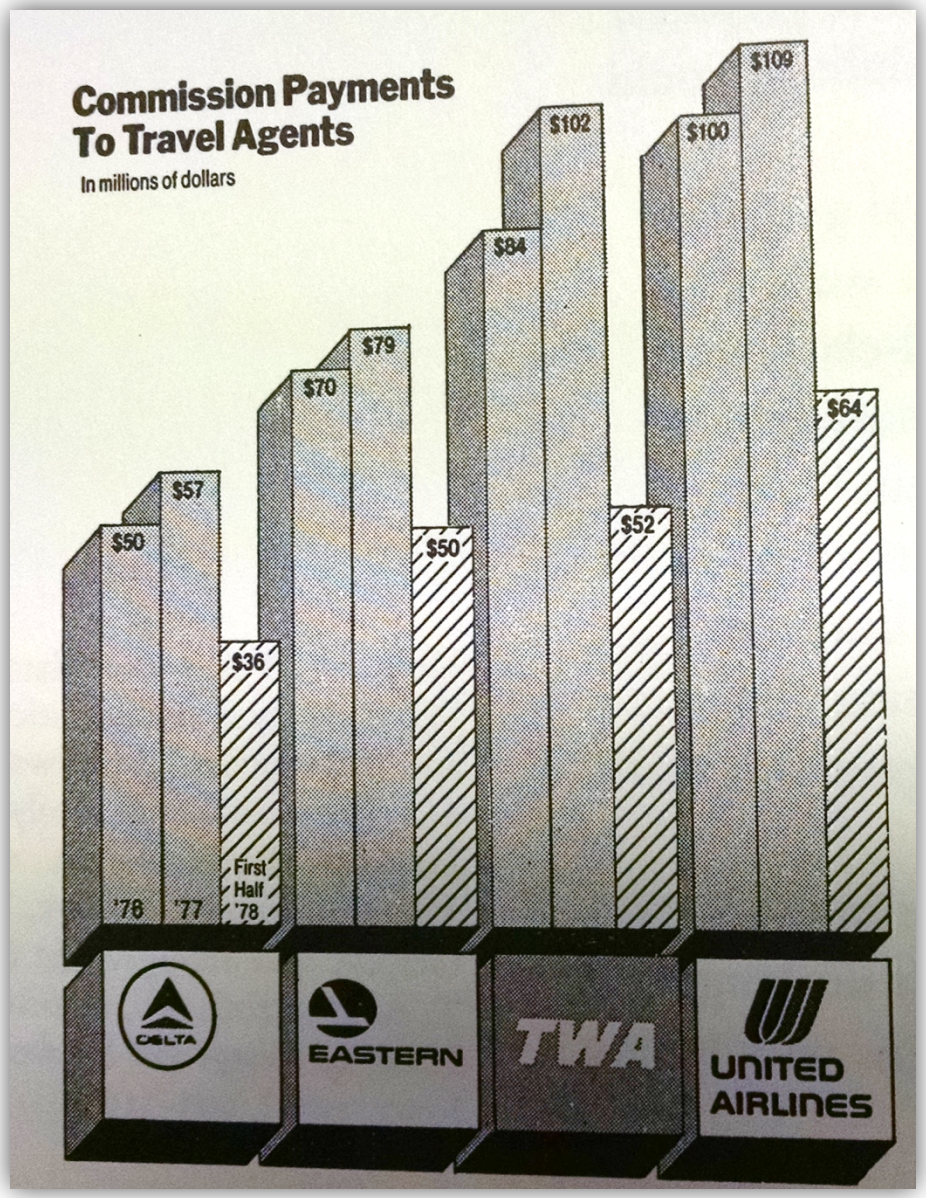
https://eagereyes.org/blog/2012/bikini-chart



https://eagereyes.org/blog/2012/bikini-chart



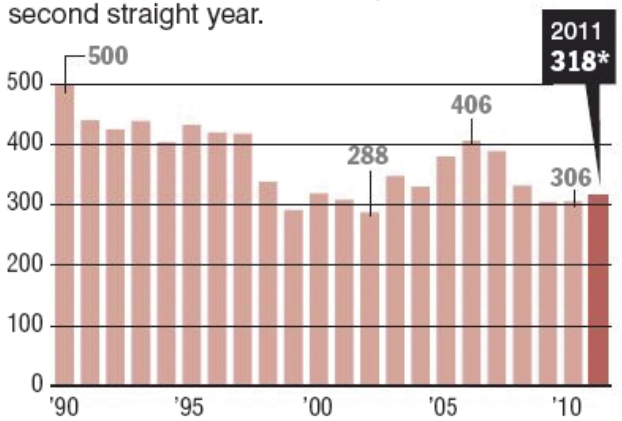
The Visual Display of Quantitative Information 2E, Tufte 2001



The Visual Display of Quantitative Information 2E, Tufte 2001

Phila. homicides by year

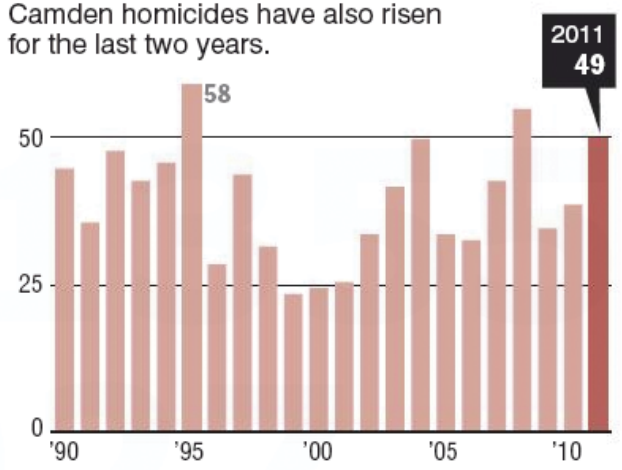
After declining between 2006 and 2009, the number of homicides ticked upward for the second straight year.



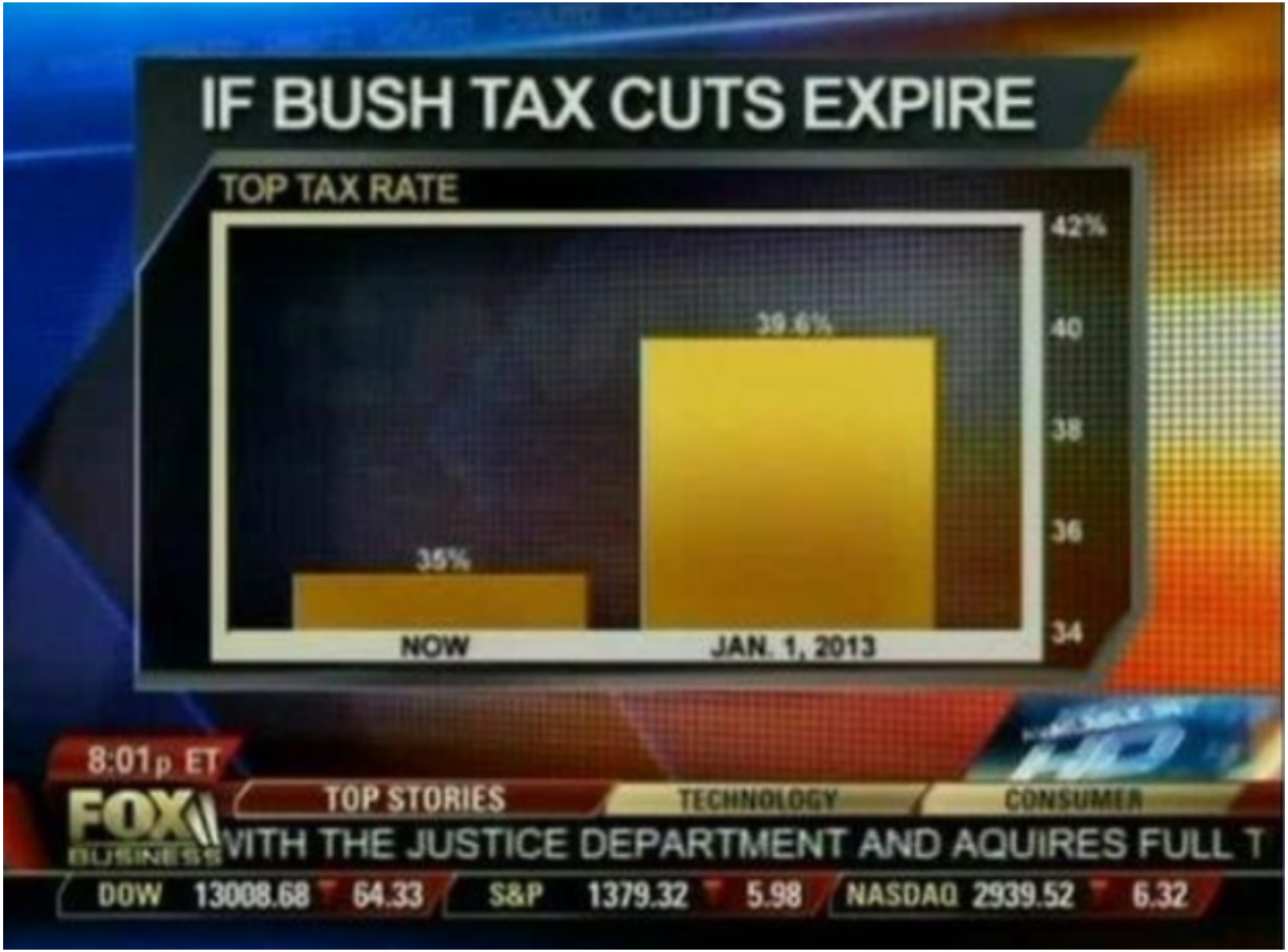
* Figures do not include the eight deaths from the Jan. 19, 2011, grand jury indictment charging abortion doctor Kermit Gosnell with murder. The Police Department adds these deaths to the 2011 total, even though they occurred in previous years, because the FBI requires homicides to be counted in the year in which they become known to the police.

Camden homicides by year

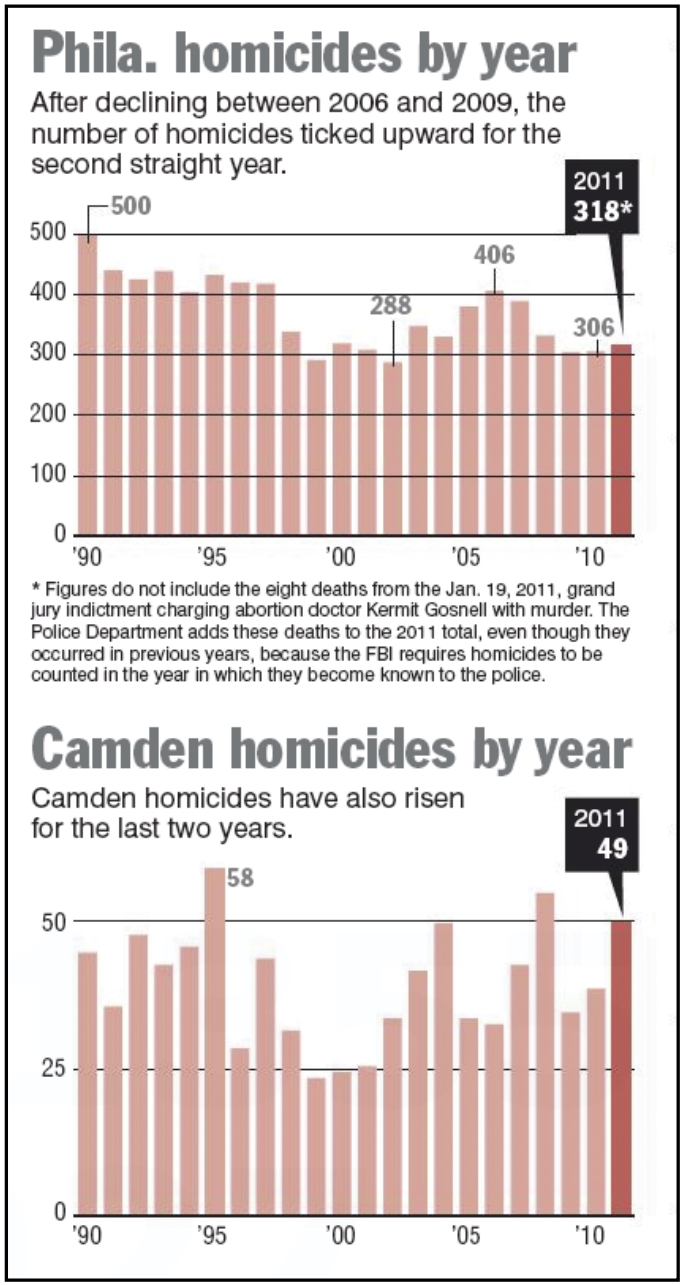
Camden homicides have also risen for the last two years.



http://flowingdata.com/2014/04/04/fox-news-bar-chart-gets-it-wrong/



http://flowingdata.com/2012/08/06/fox-news-continues-charting-excellence/

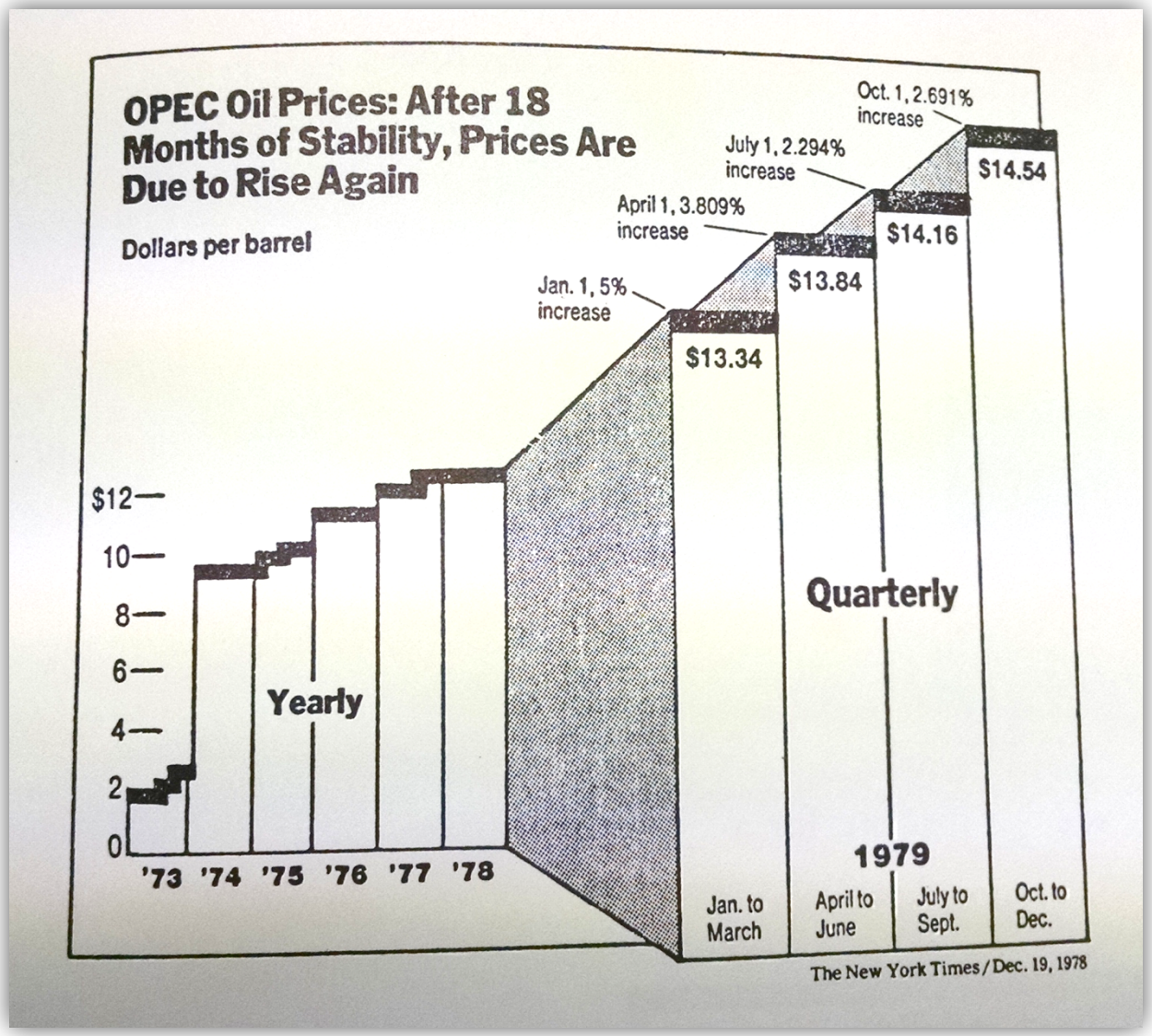


Misinformation…

Disinformation.

Design Variation

Consistency helps

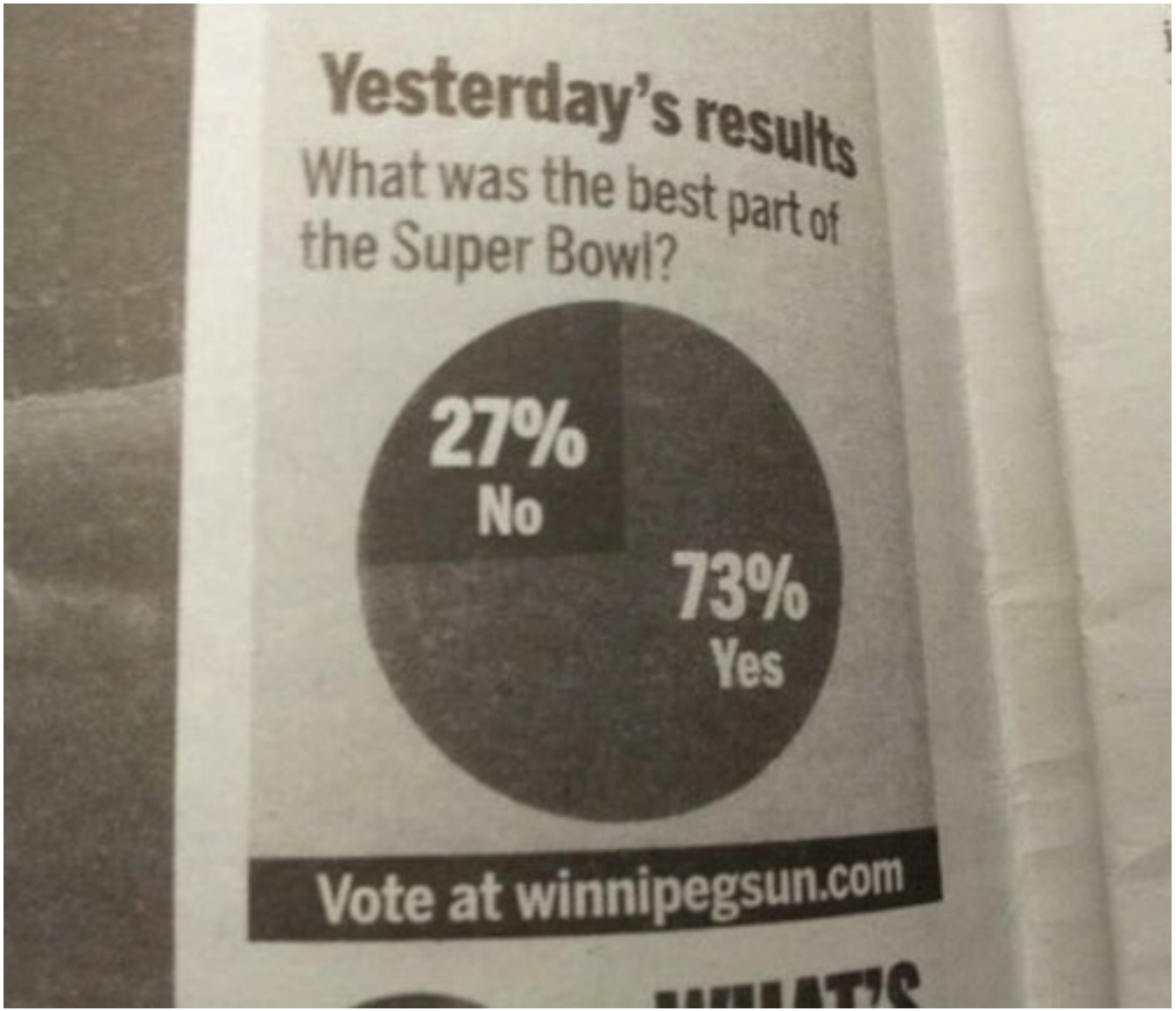


The Visual Display of Quantitative Information 2E, Tufte 2001

"The purpose of visualization is insight, not pictures."

Shneiderman B. Research agenda: Visual overviews for exploratory search. Information Seeking Support Systems. 2008;11:4.

http://flowingdata.com/category/statistics/mistaken-data/



http://flowingdata.com/category/statistics/mistaken-data/

OPINION

Obama's Divided Nation

ama presides over
ica more divided
any time in 50
ry that was riven
g racial lines gath-
2008 to elect its
sident. That presi-
four years dividing
the basis of eco-
The campaign re-
vealed no evi-
dence that Mr.
Obama will close
the chasm he has
created between
his voters and
those he attacked
and vilified.

It may be true
that Mitt Romney
failed to respond

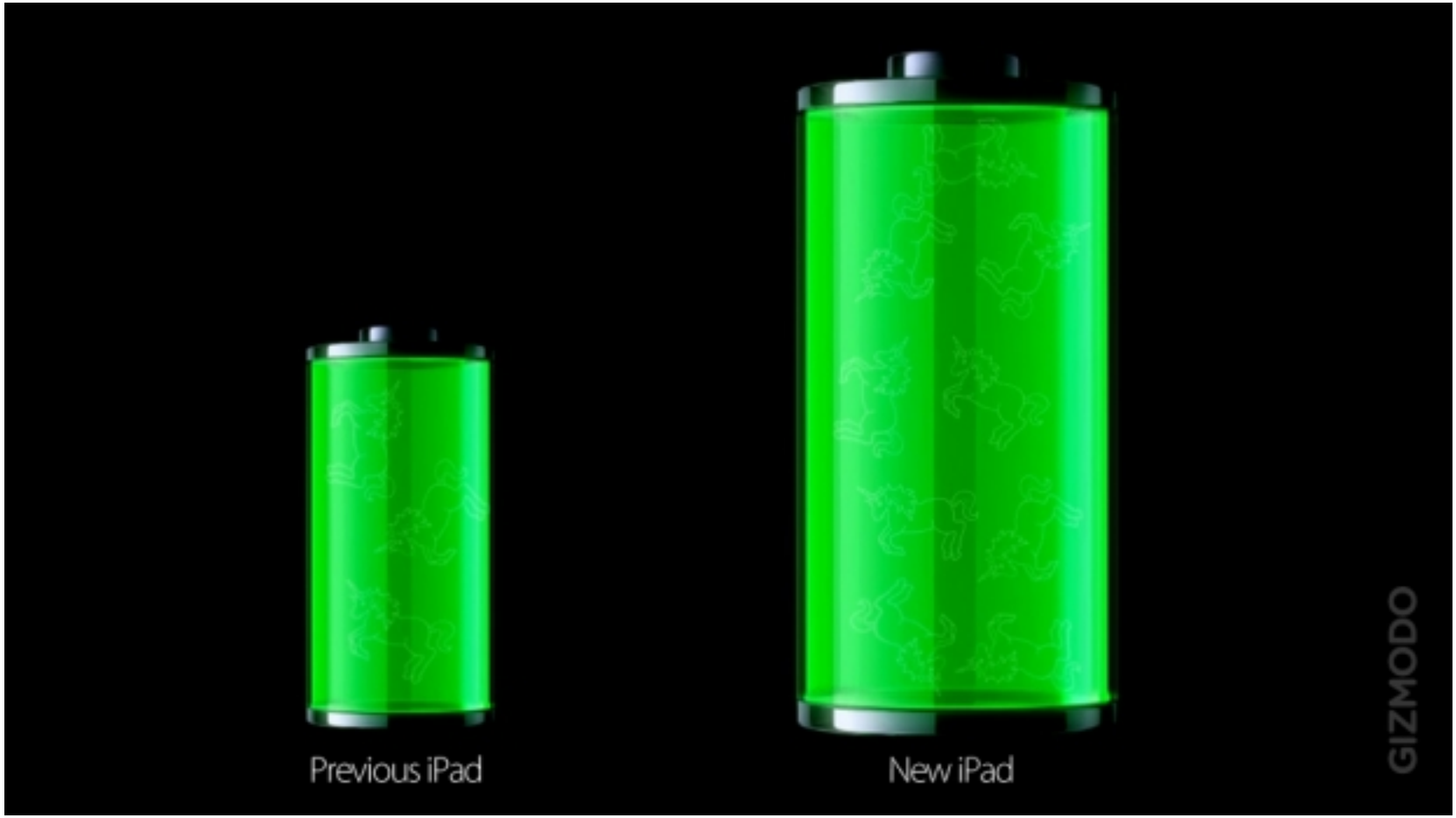
drawn attention to what hap-

Obama spokestotum replied:

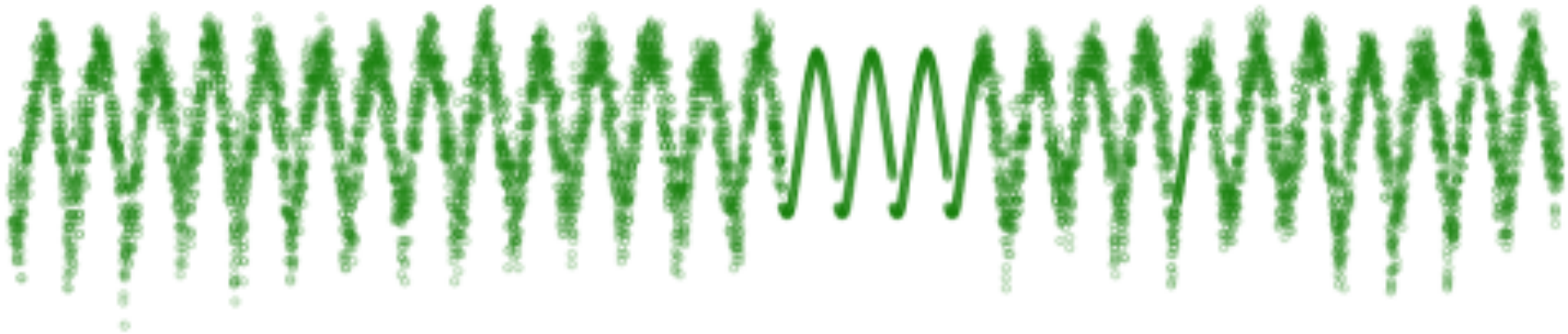
problem with pols, e
verbally facile as Mr.
that in crunch time,
reverts to No. 1. Exi
that 9% of the electo
who to vote for jus
Tuesday; and amo
42% said Mr. Oban
Sandy response—t
tie photo-op—wa
factor. Of those,
voted for Mr. Ob
Mr. Christie is o
politico who is

Yes, Republic
across two pres
that there are
how crudely
issue like ill
Blowing up th
if you thoug
day's results

http://flowingdata.com/category/statistics/mistaken-data/



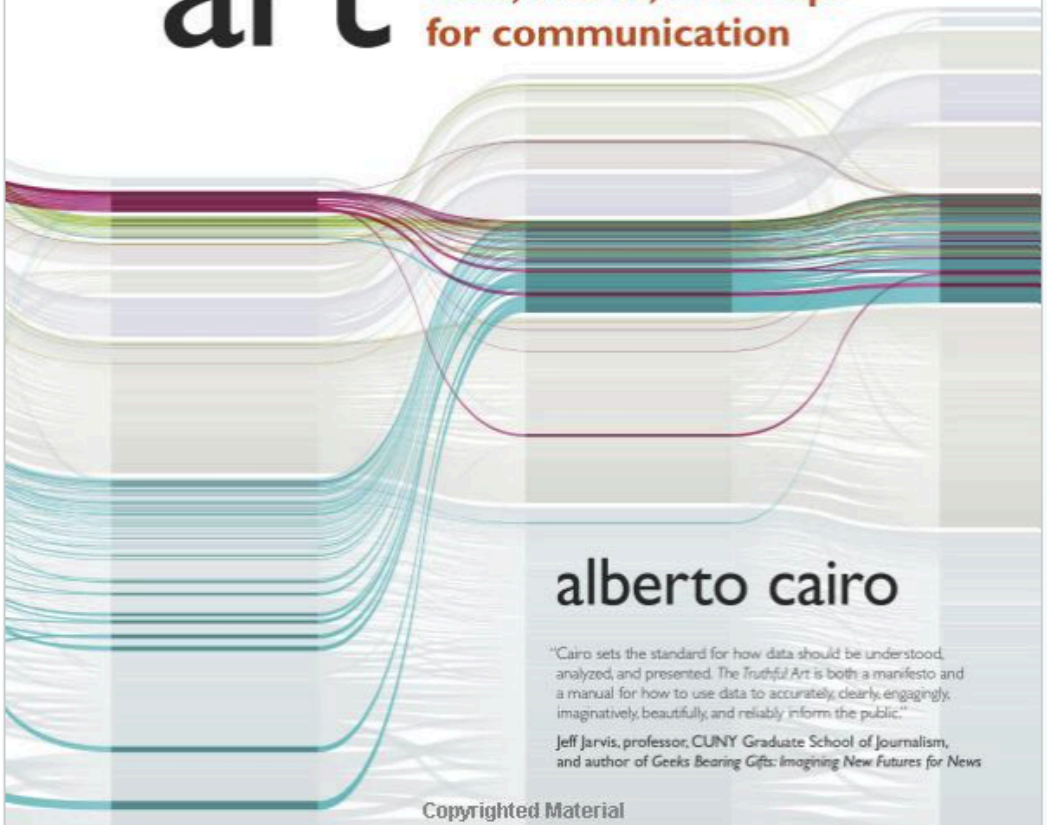
http://flowingdata.com/category/statistics/mistaken-data/



Copyrighted Material

the truthful art

data, charts, and maps for communication



alberto cairo

"Cairo sets the standard for how data should be understood, analyzed, and presented. *The Truthful Art* is both a manifesto and a manual for how to use data to accurately, clearly, engagingly, imaginatively, beautifully, and reliably inform the public."

Jeff Jarvis, professor, CUNY Graduate School of Journalism, and author of *Geeks Bearing Gifts: Imagining New Futures for News*

Copyrighted Material