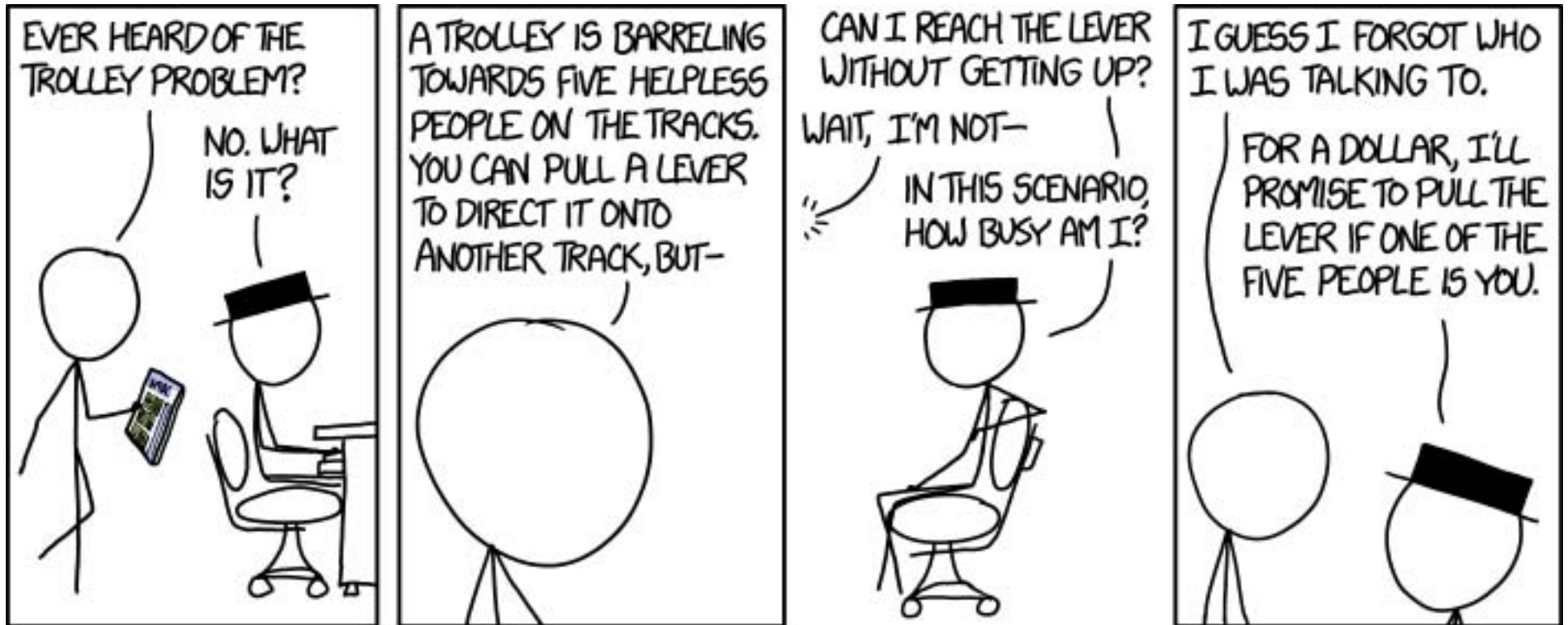


Ethics of Data Visualization

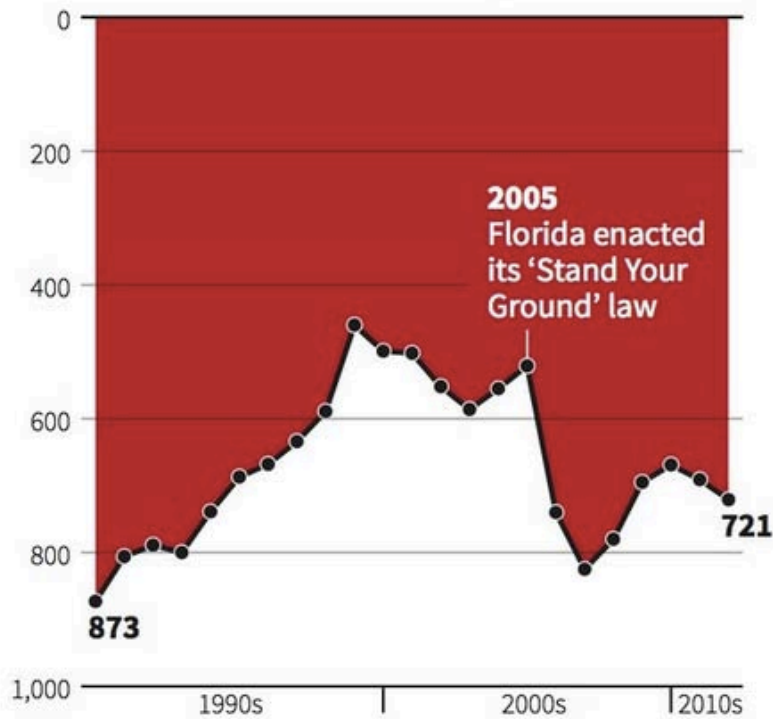


Data Visualization | Steven Bedrick & Jackie Wirz

<http://xkcd.com/1455/>

Gun deaths in Florida

Number of murders committed using firearms



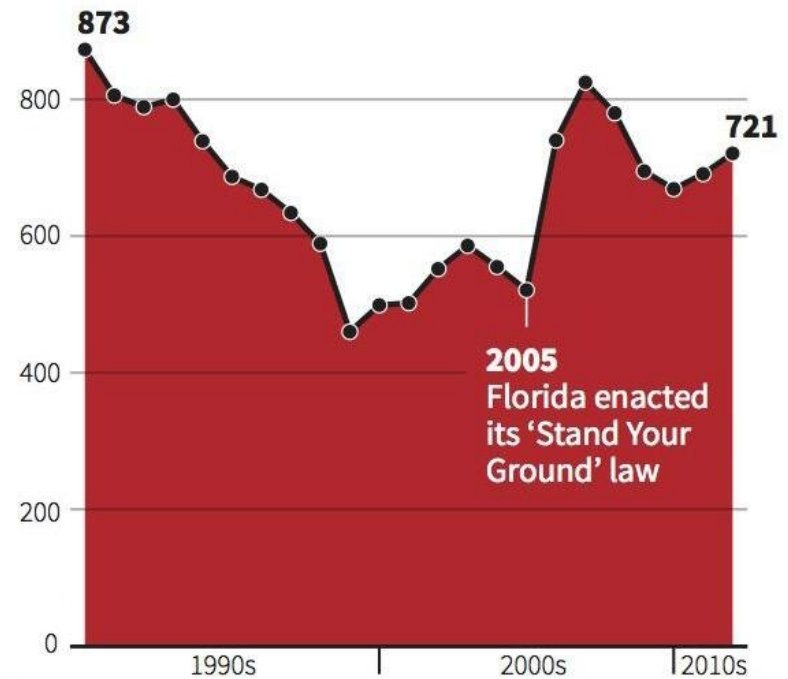
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014



Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

- 1) Visualization is the graphical display of information...
- 2) Graphics act as a cognitive aid

Graphics take the EFFORT out of
imagining the information.

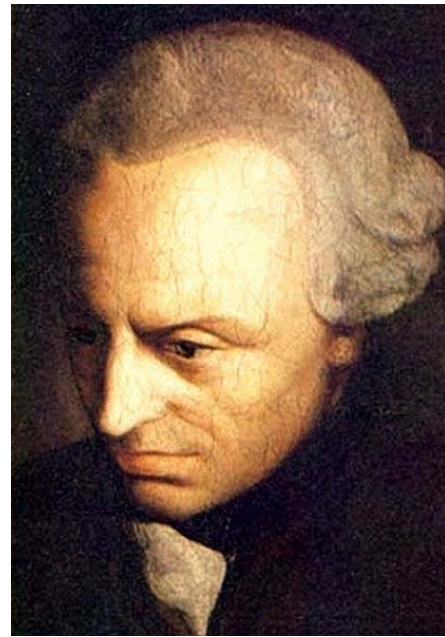
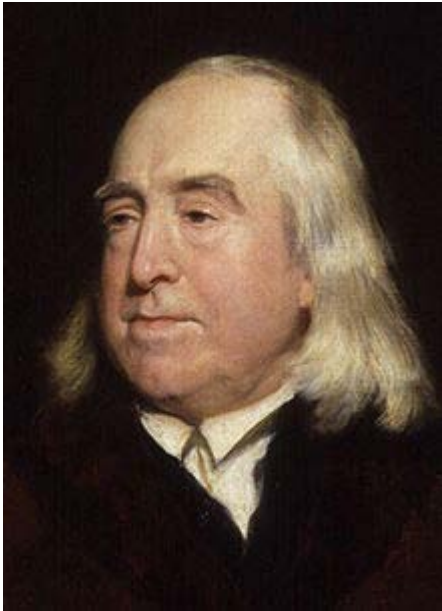
The goal of ethical visualizations is to help with cognitive strain, but to stay true to the data and analysis.

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

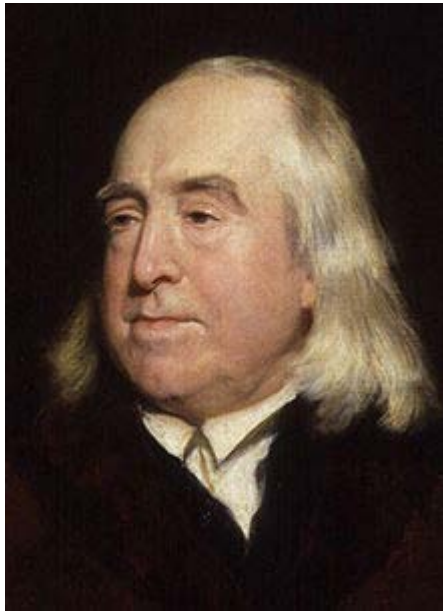
“The goal is not to over simplify the information. *We should* present complexity to readers in an engaging way, people will care.”

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

There are two main families of ethical reasoning:



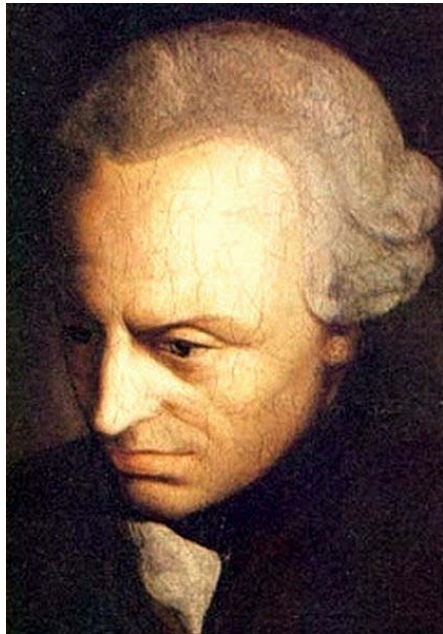
Consequentialism



“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*

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*

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

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

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

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."

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.”

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

*“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

Gallup Abortion Poll: Numbers

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

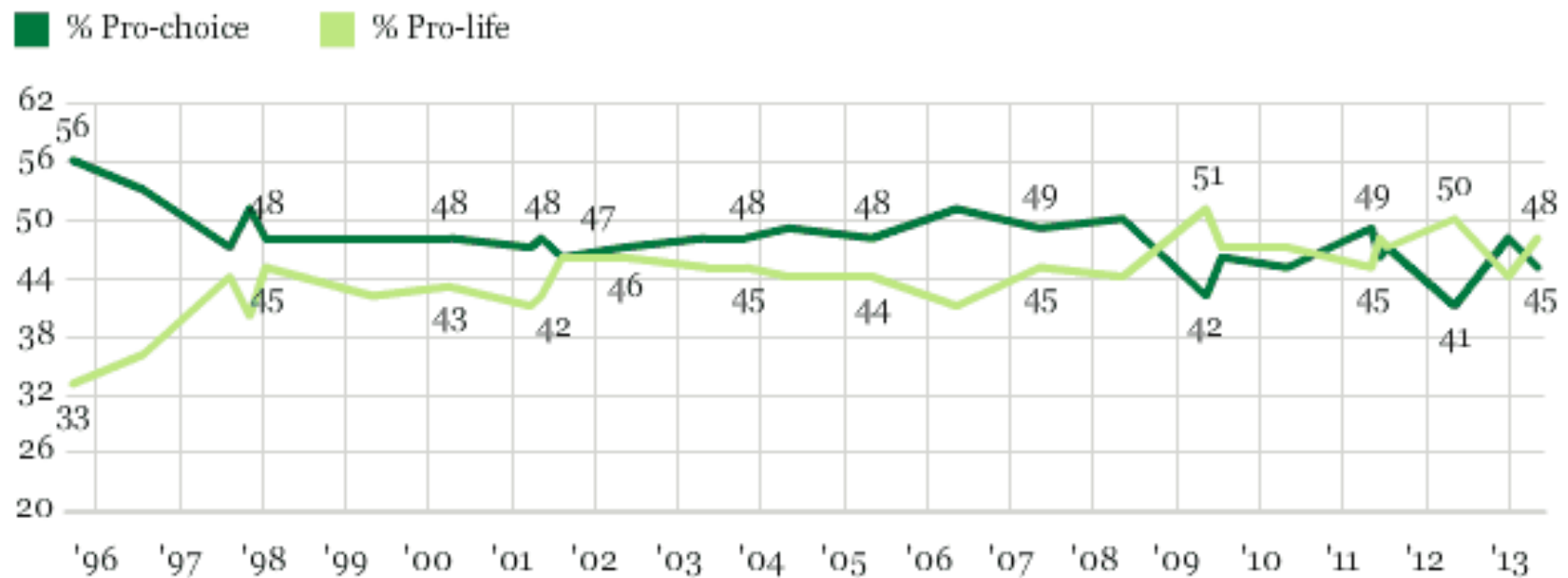
Trend from polls in which pro-life/pro-choice was asked after question on legality of abortion

	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 20-May					

Gallup Abortion Poll: Their Graph

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

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



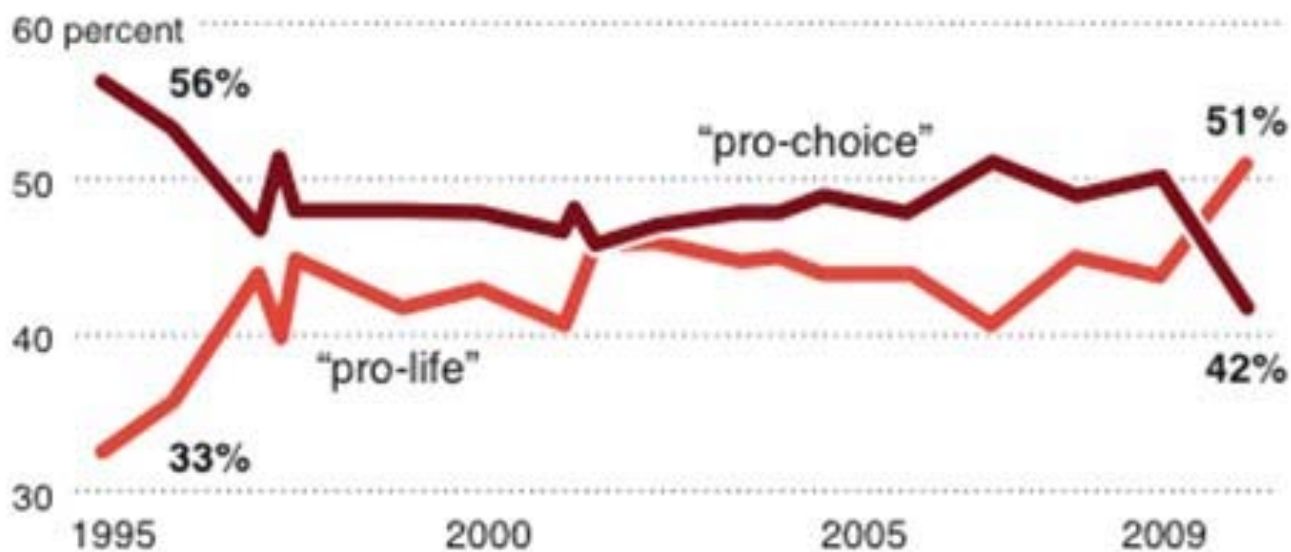
GALLUP

Gallup Abortion Poll: AP

Strong shift in opinion on abortion

A new poll found 51 percent of Americans identify themselves as "pro-life" on the issue of abortion and 42 percent as "pro-choice." It's the first time since 1995 that a majority of U.S. adults have identified themselves as pro-life.

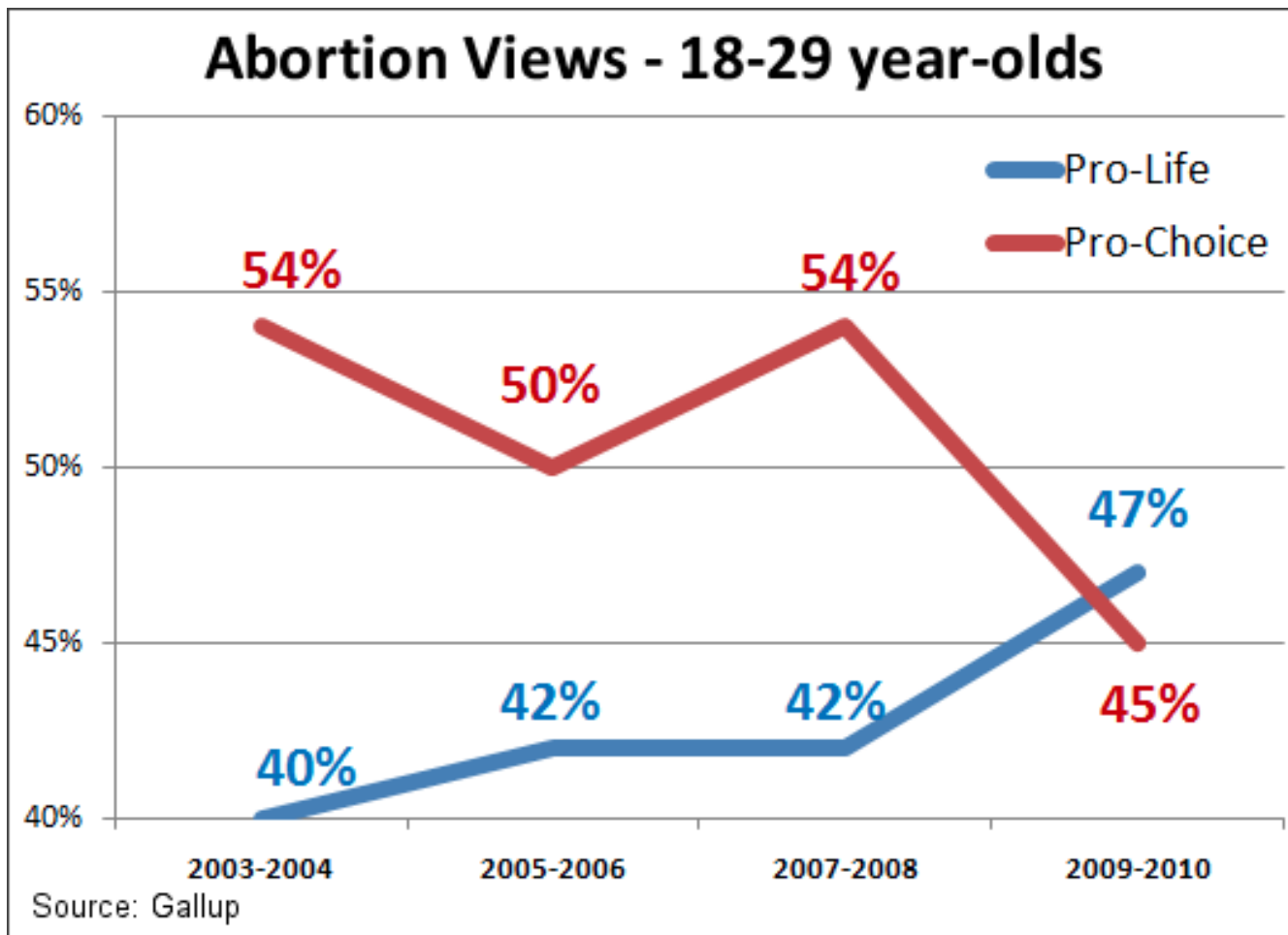
Americans calling themselves "pro-life" or "pro-choice"



SOURCE: Gallup

AP

Gallup Abortion Poll: LiveAction.org



Gallup Abortion Poll: LiveCitizen.com

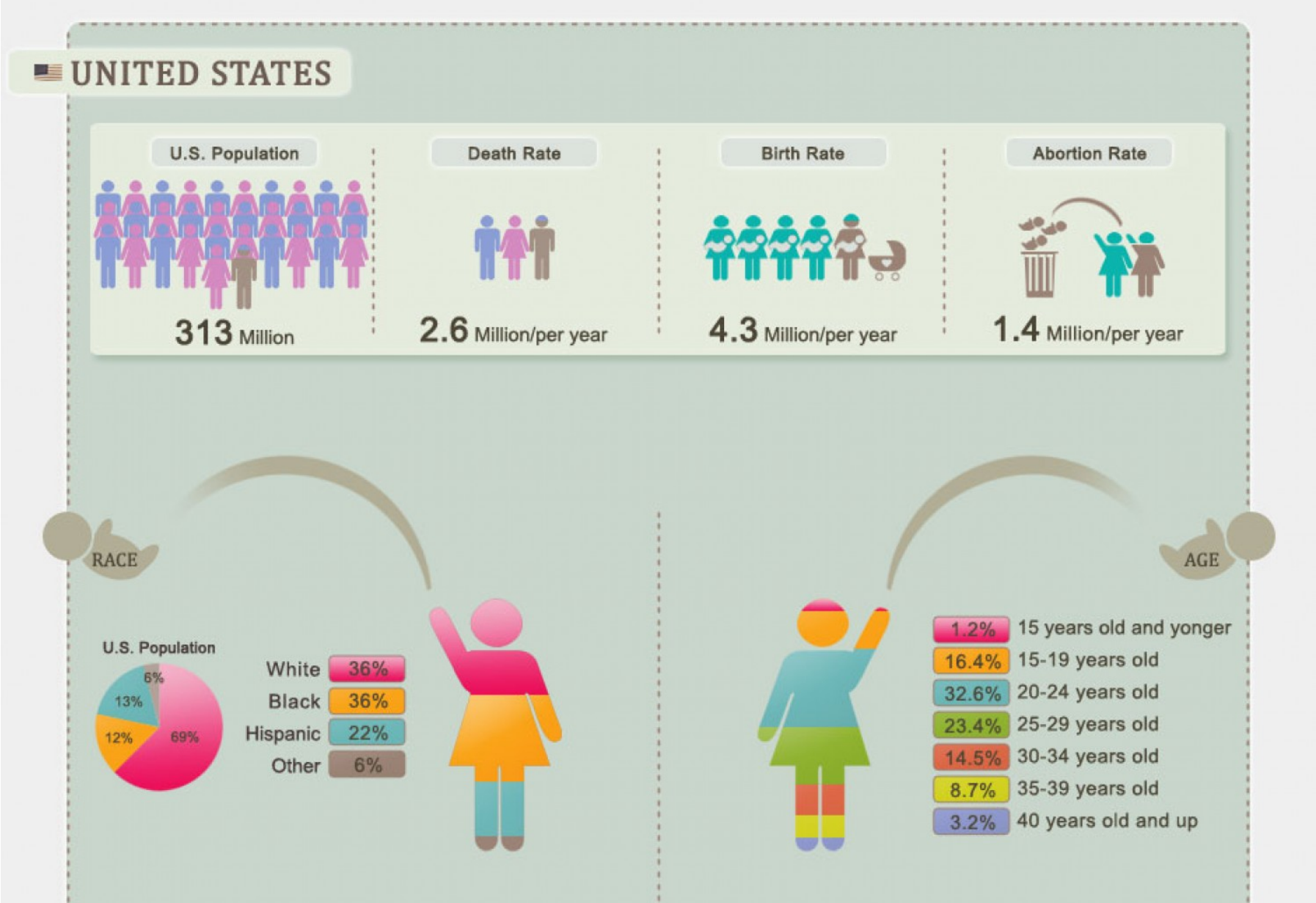
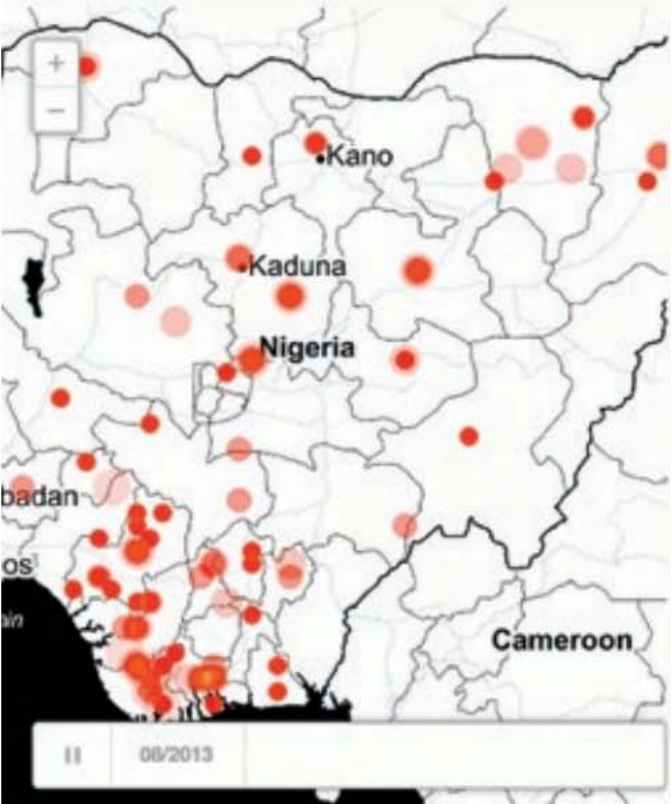
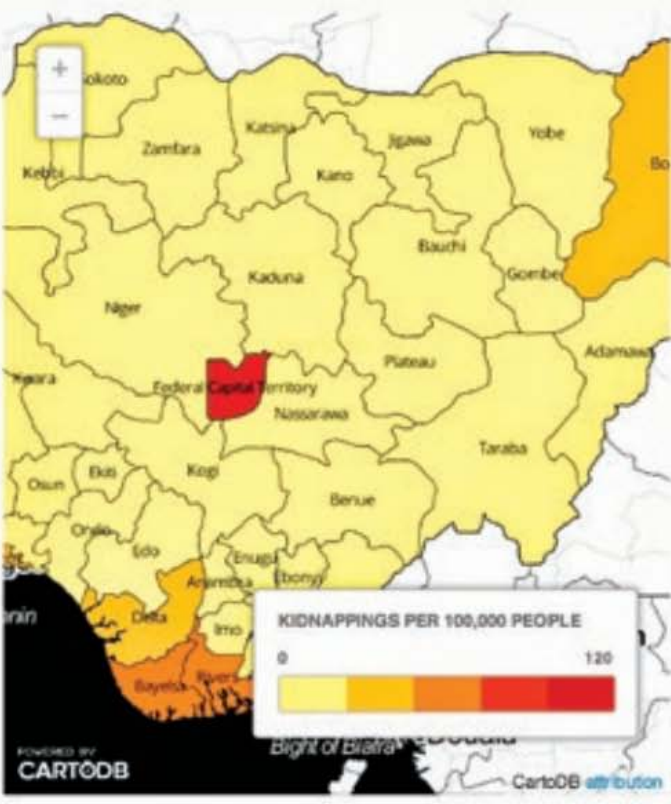


Figure 1

Kidnappings in Nigeria, 1982-present



Kidnapping rate per 100,000 people, 1982-present



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

FiveThirtyEight

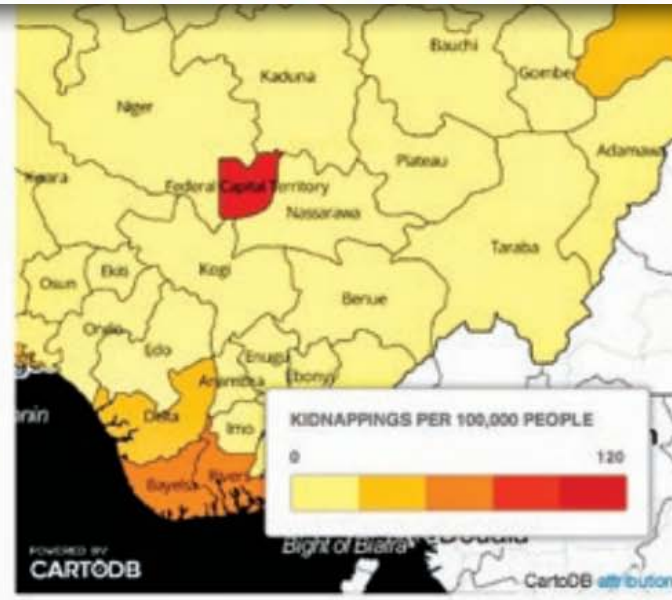
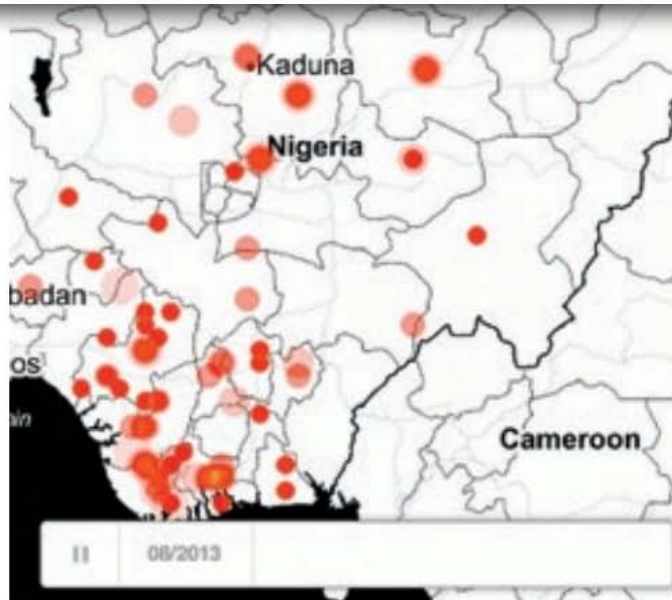
Figure 1

Kidnappings in Nigeria, 1982-present

Kidnapping rate per 100,000 people, 1982-present

FiveThirtyEight used data from the Global Database of

Kidnapping rate per 100,000 people, 1982-present



stories about kidnappings, to create both graphics.

FiveThirtyEight

Figure 1

Kidnappings in Nigeria, 1982-present

Kidnapping rate per 100,000 people, 1982-present

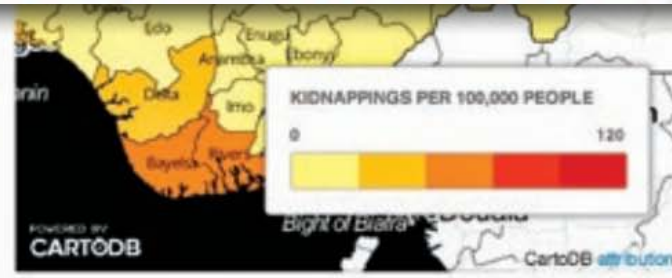
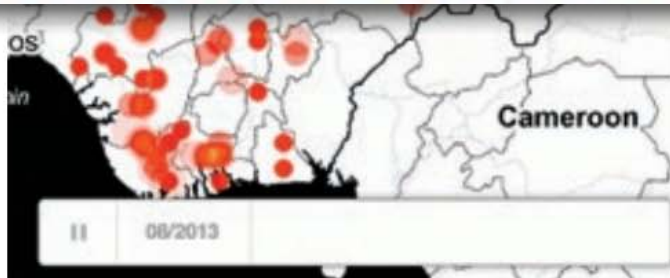
FiveThirtyEight used data from the Global Database of

Kidnapping rate per 100,000 people, 1982-present



stories about kidnappings, to create both graphics.

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.”

“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.”

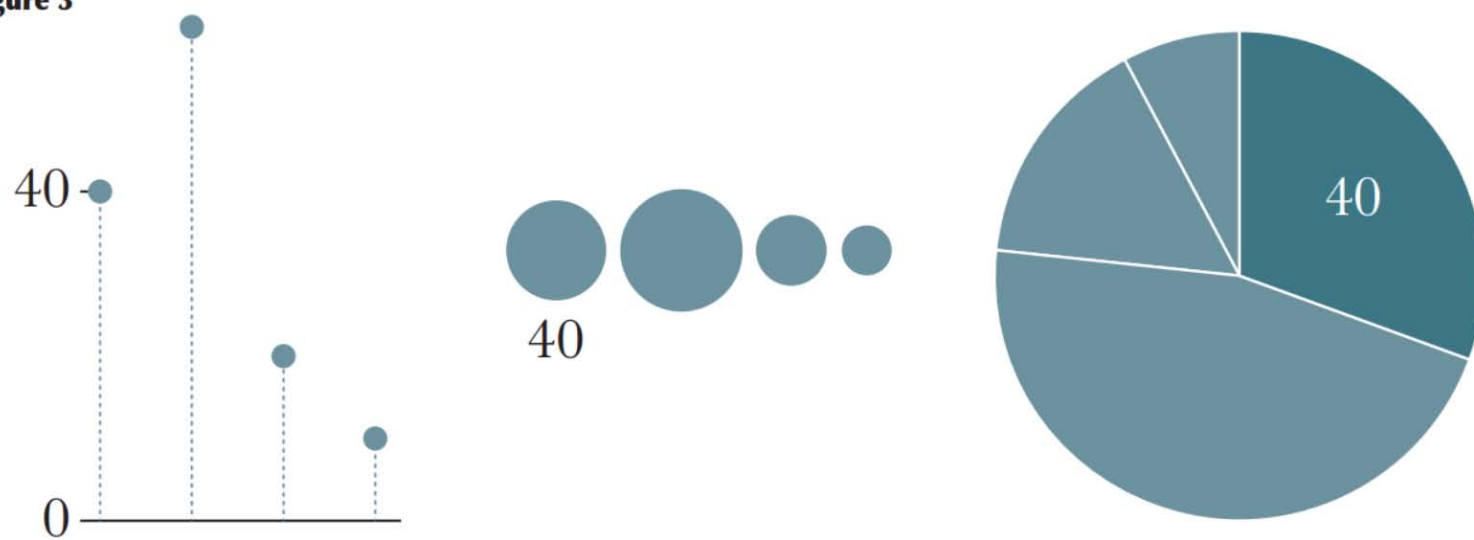
“This piece did not meet FiveThirtyEight’s standards for publication. We apologize for the mistakes. We will do better. The original article follows below.”

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.

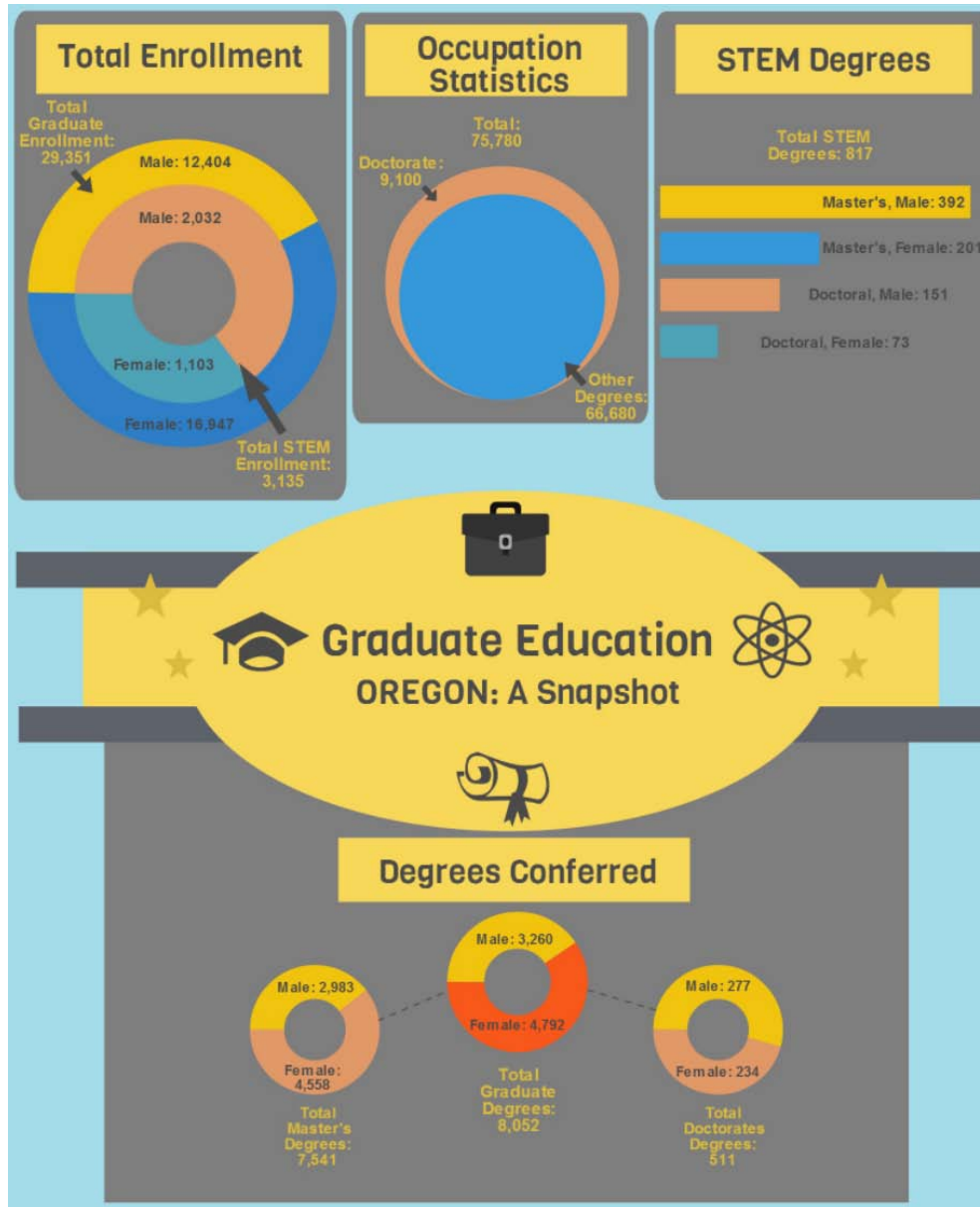
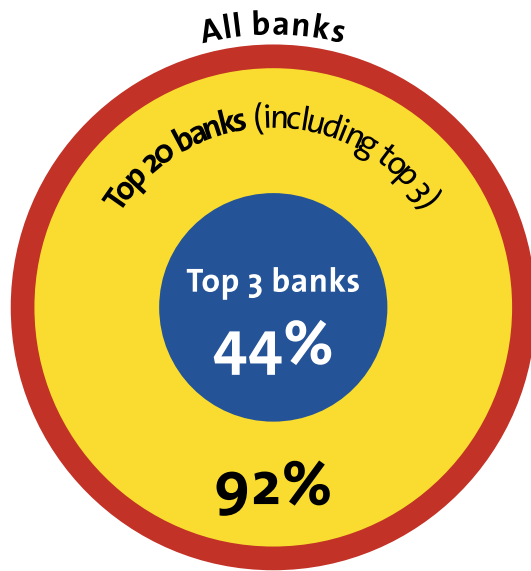
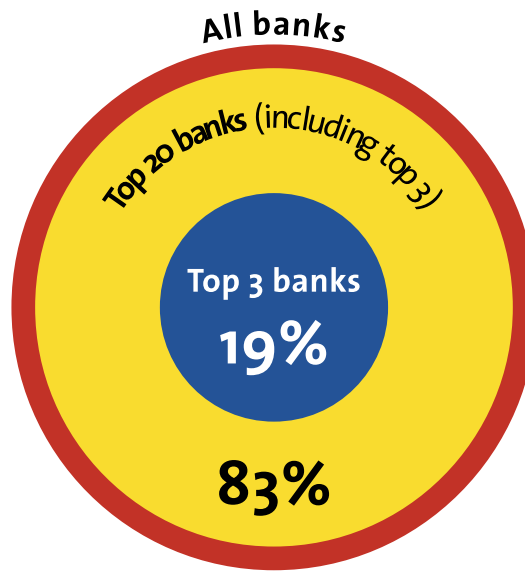


Image from the Conference of Oregon Deans



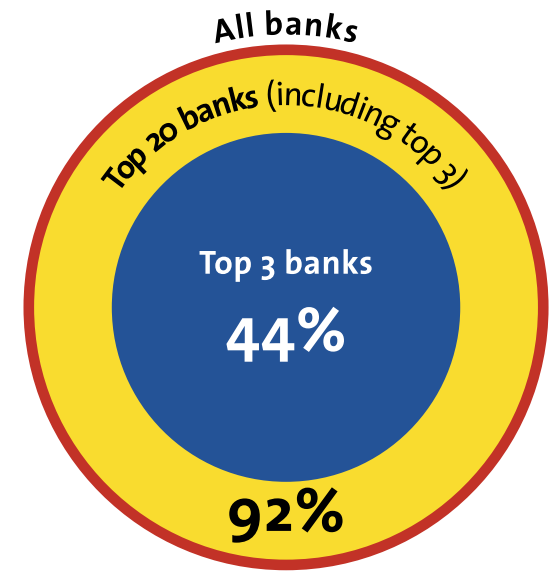
In *The New York Times*

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).*



Actual percentages of *The New York Times* diagrams

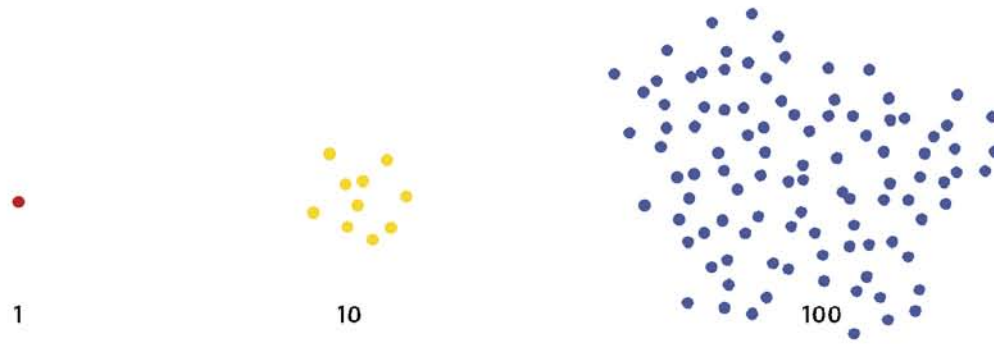
*The true percentages represented by *The New York Times*'s circles are shown in the center graphic; a graphic accurately representing the percentages is at the right.*



Correct diagram of *The New York Times* data

“Designing Information”, by Joel Katz, p.85

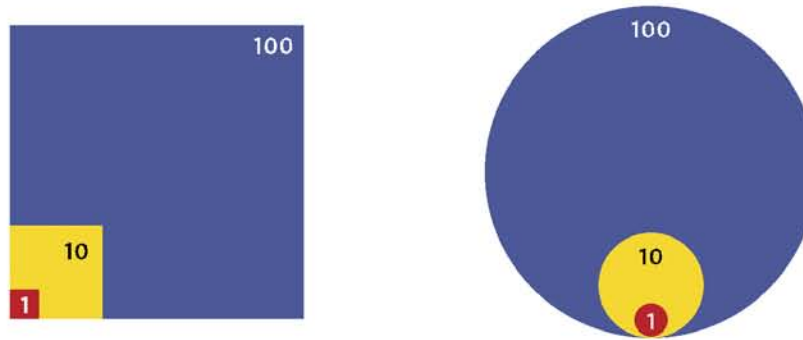
Point



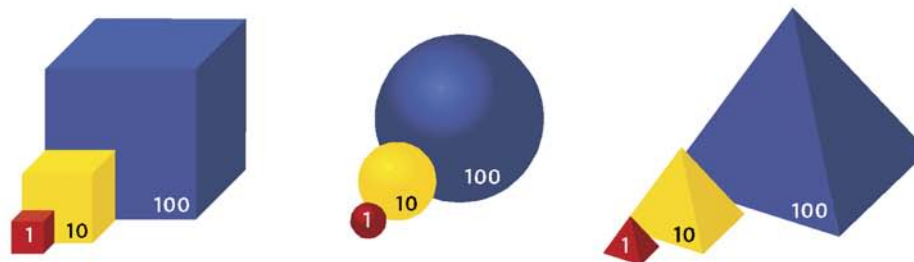
Line

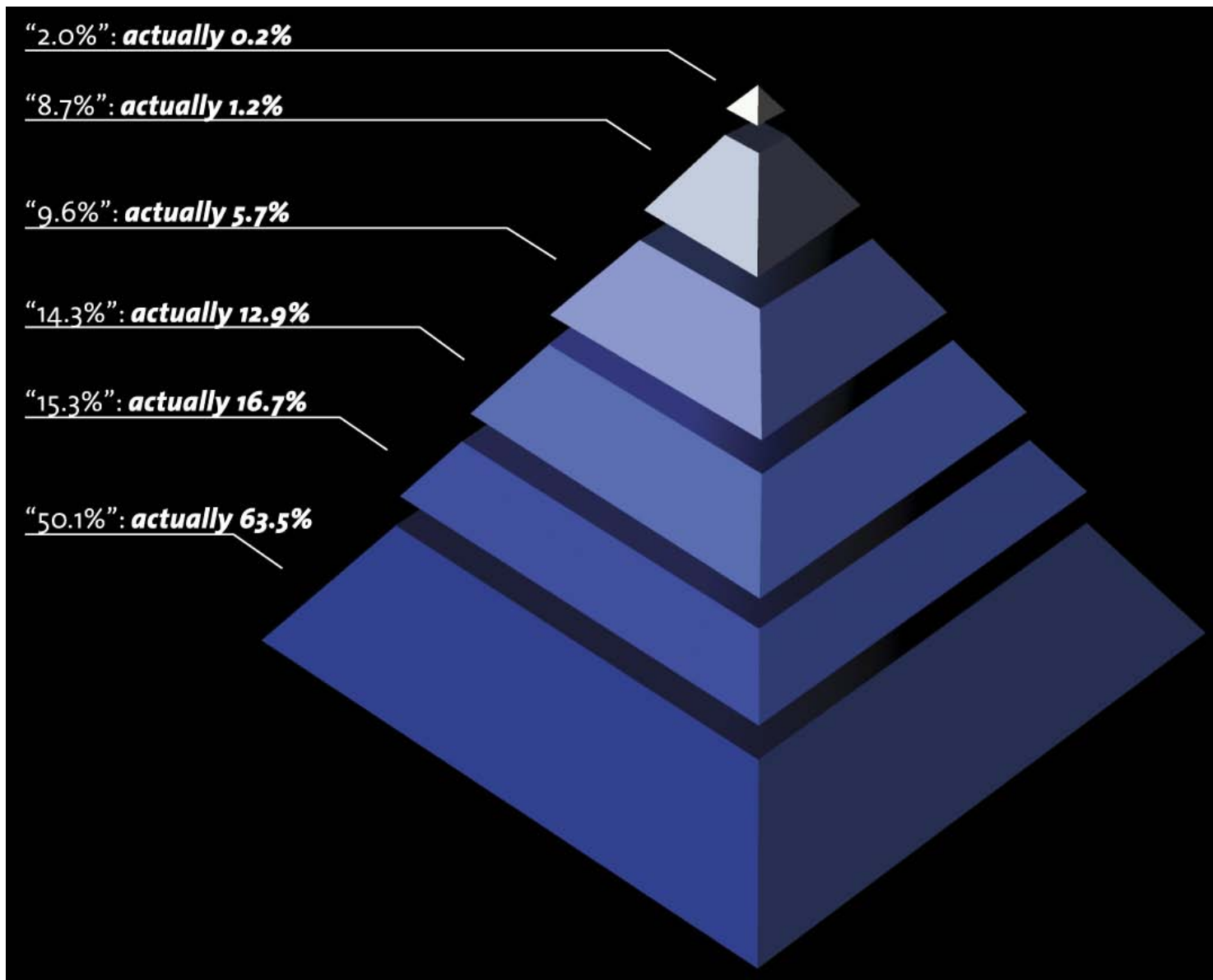


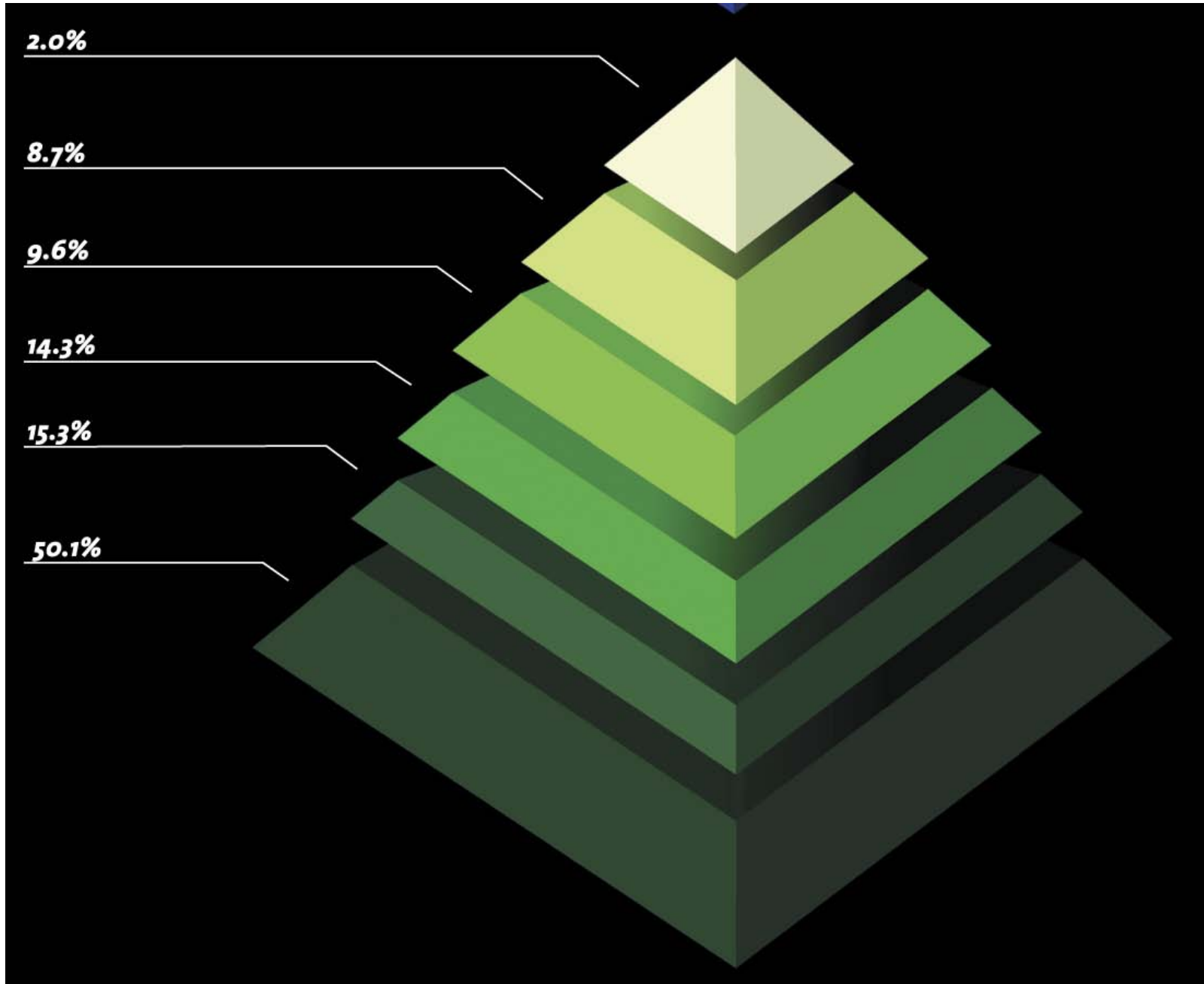
Plane



Solid







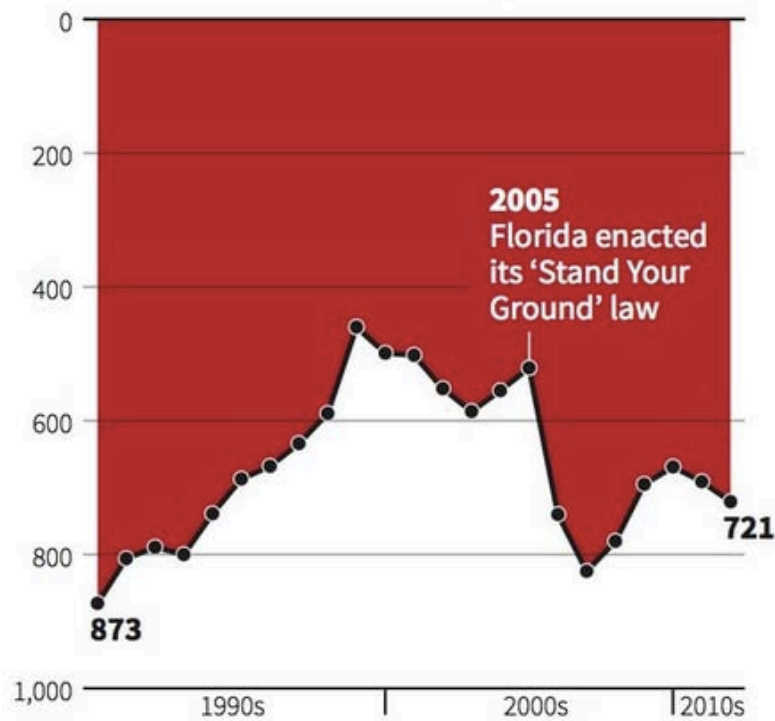
“Designing Information”, by Joel Katz, p.86

Lies in Scale

Don't be a jerk

Gun deaths in Florida

Number of murders committed using firearms



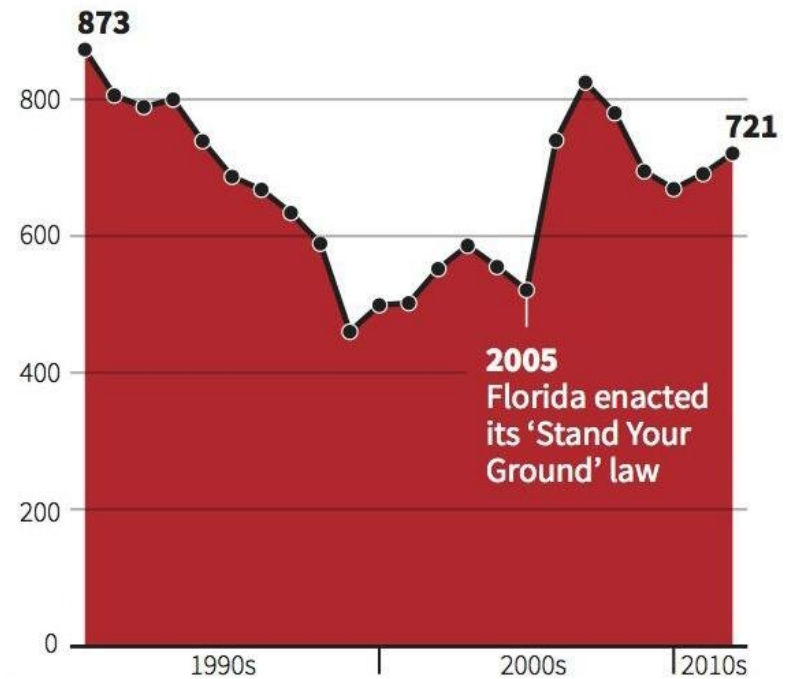
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014



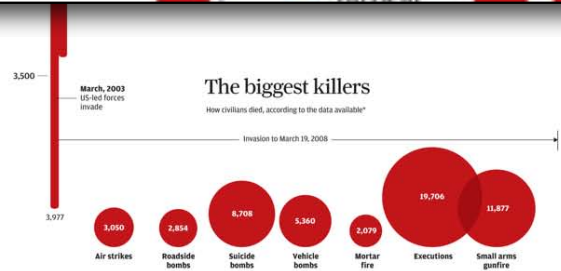
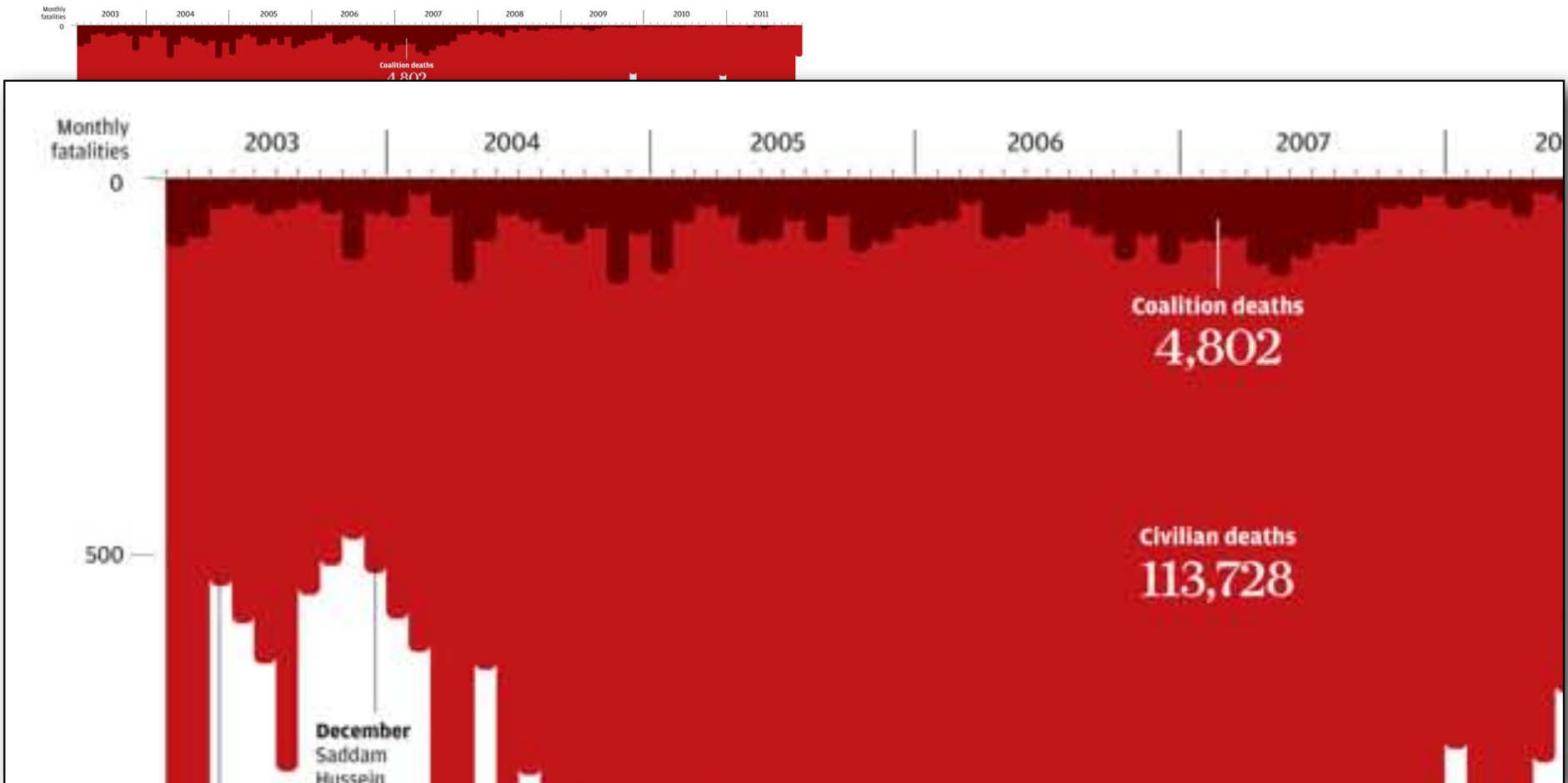
Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

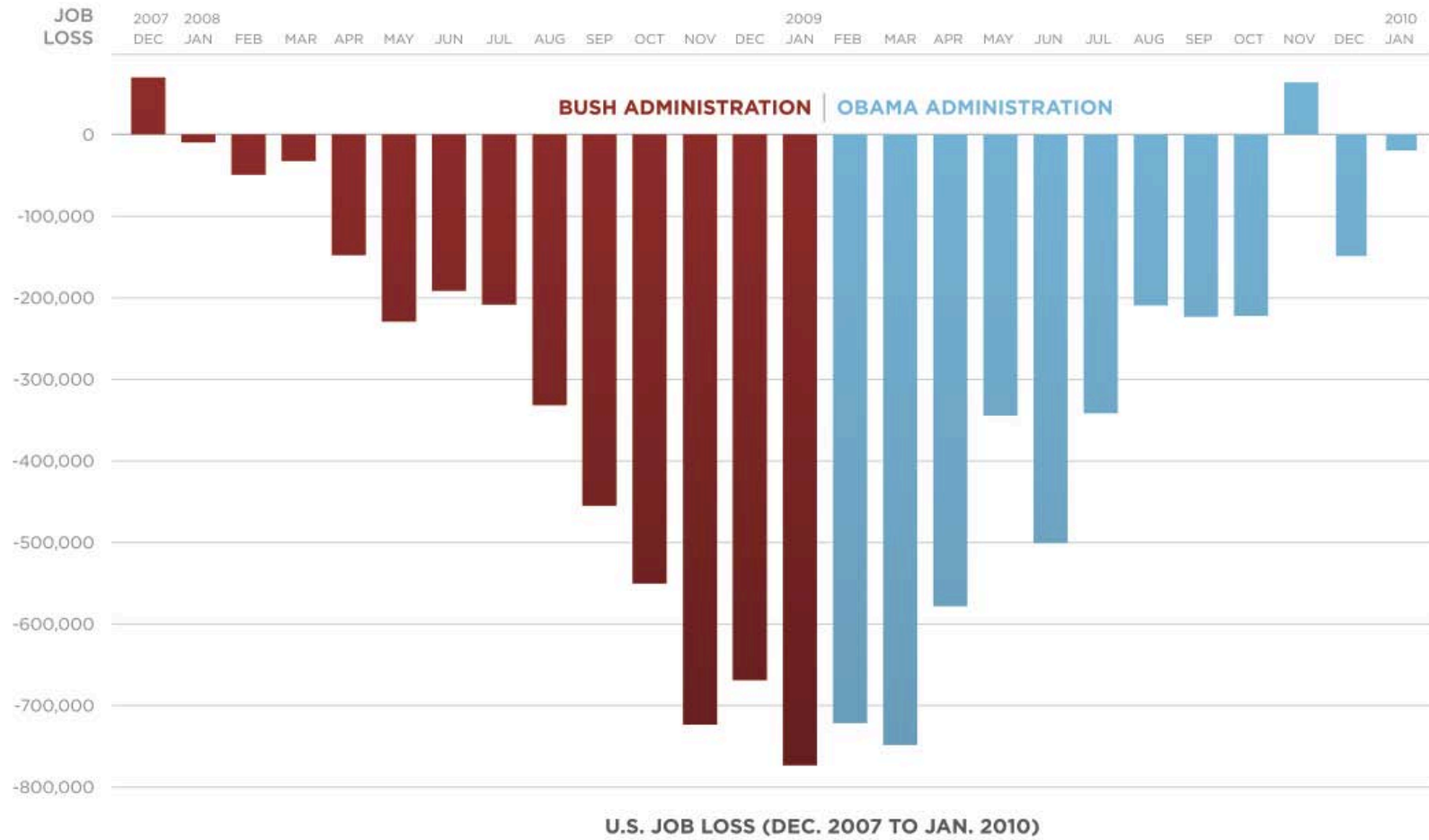
Iraq's bloody toll



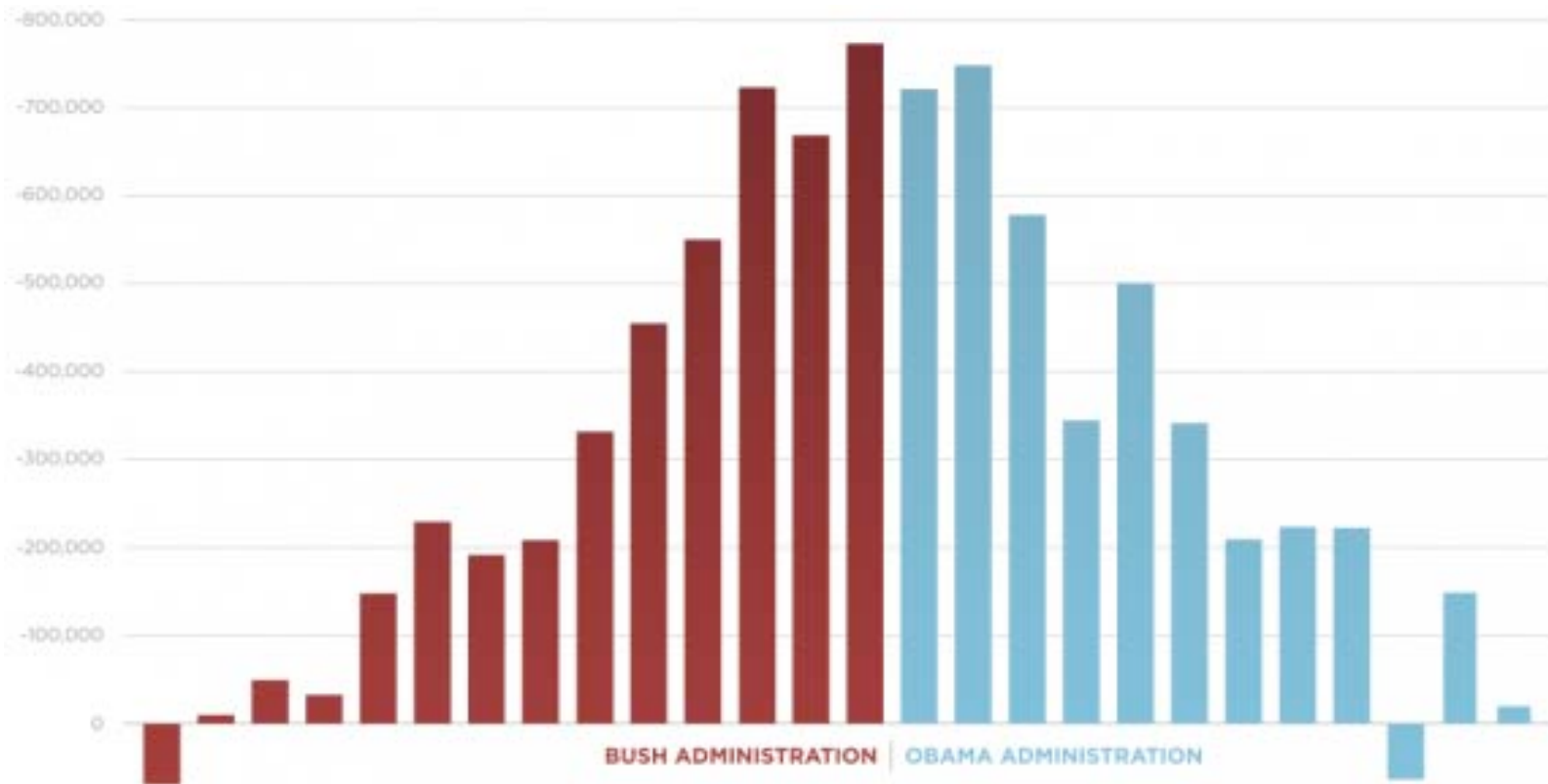
© 2011 Simon Scarr

*Deaths from unknown causes are not included. Causes of death according to the March 03 per cent of killings also not included.

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

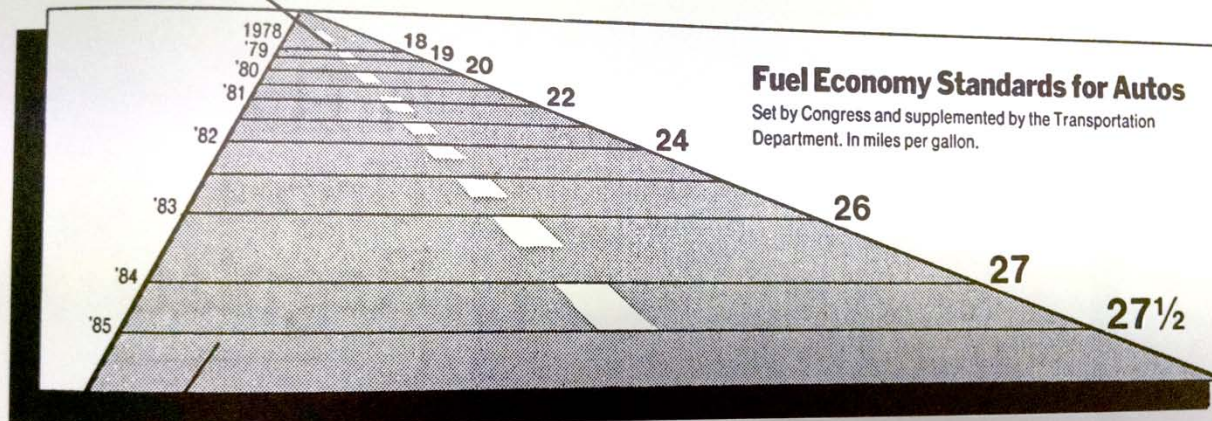


SOURCE: BUREAU OF LABOR STATISTICS, 02/12/2010

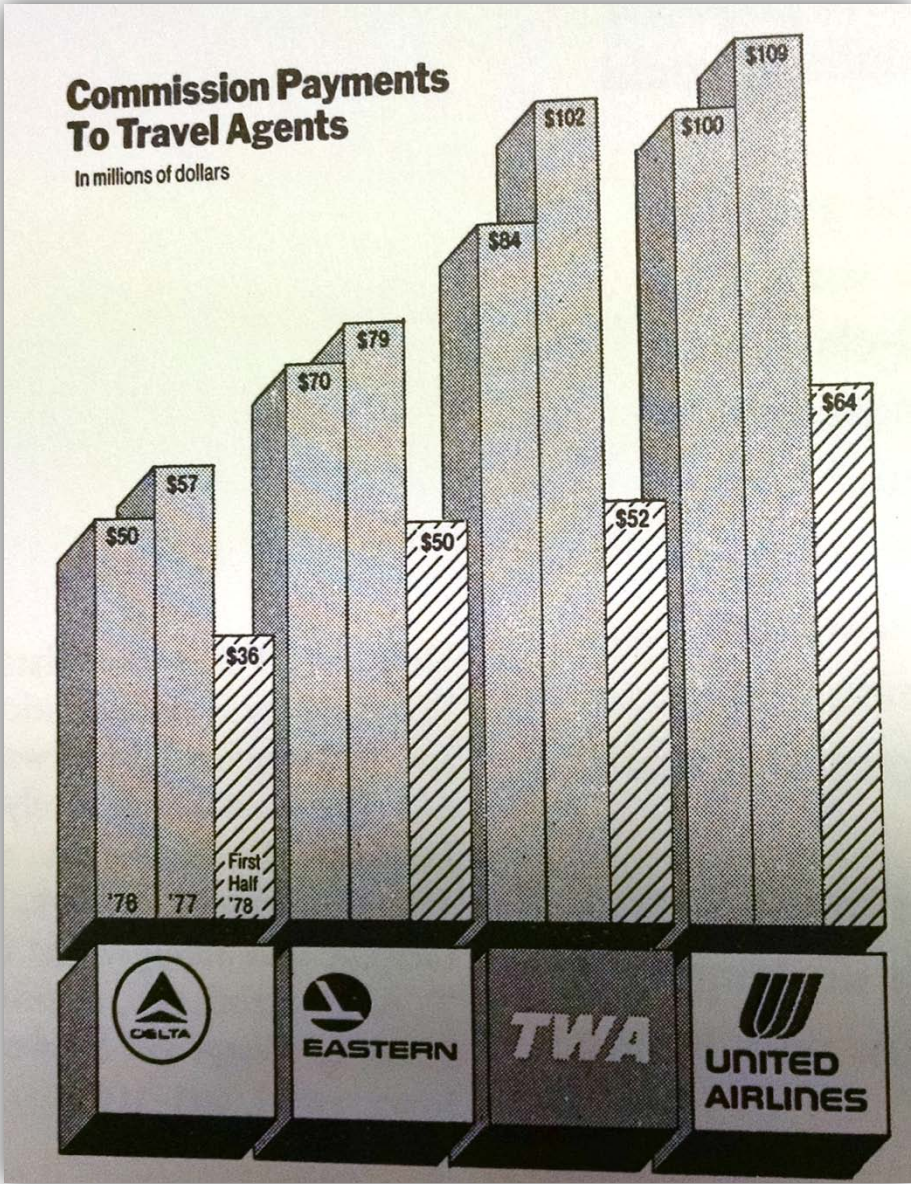


SOURCE: BUREAU OF LABOR STATISTICS, 03/12/2010

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



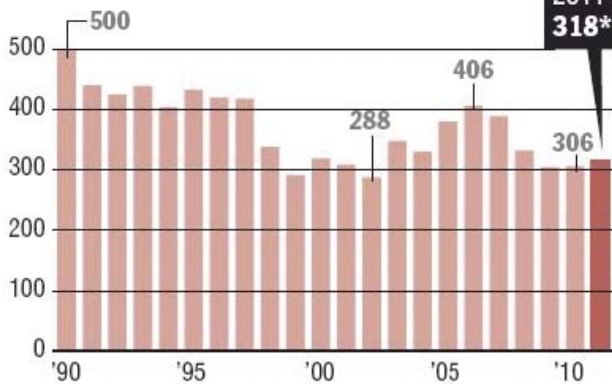
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.



The Visual Display of Quantitative Information 2E, Tufte 2001

Phila. homicides by year

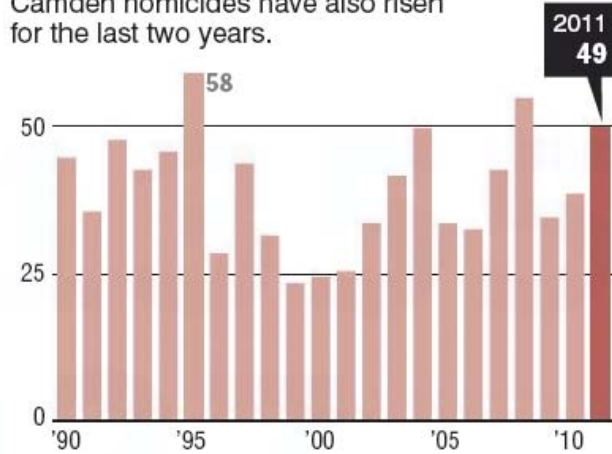
After declining between 2006 and 2009, the number of homicides ticked upward for the second straight year.



* 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.



OBAMACARE ENROLLMENT



AS OF
MARCH 27

MARCH 31
GOAL



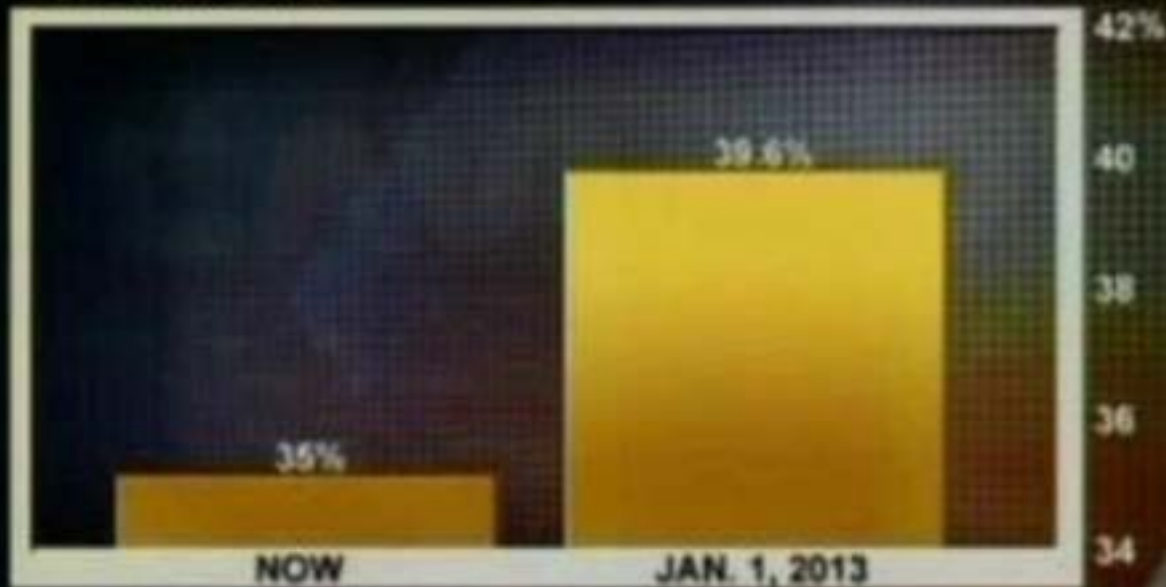
SOURCE: HHS

mediamatters.org

IE IN SECOND PLACE WITH \$26.5 MIL, WHILE "MUPP DOW FUT 16,325.00

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01 p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

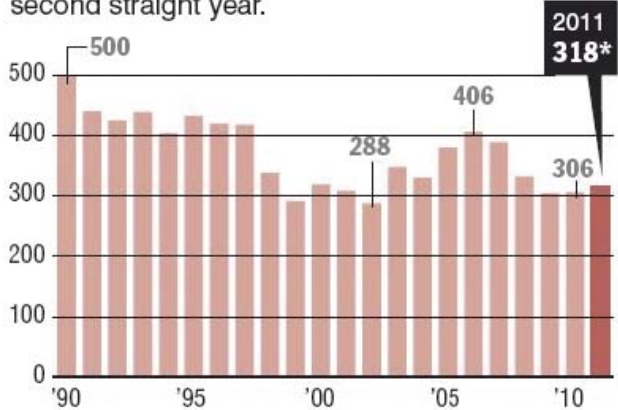
DOW 13008.68 ▼ 64.33

S&P 1379.32 ▼ 5.98

NASDAQ 2939.52 ▼ 6.32

Phila. homicides by year

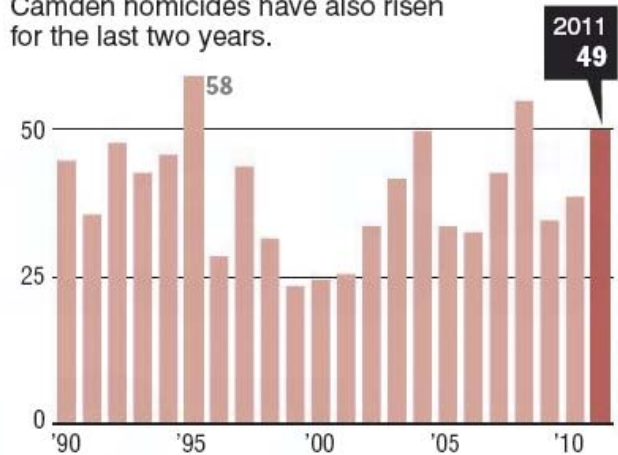
After declining between 2006 and 2009, the number of homicides ticked upward for the second straight year.



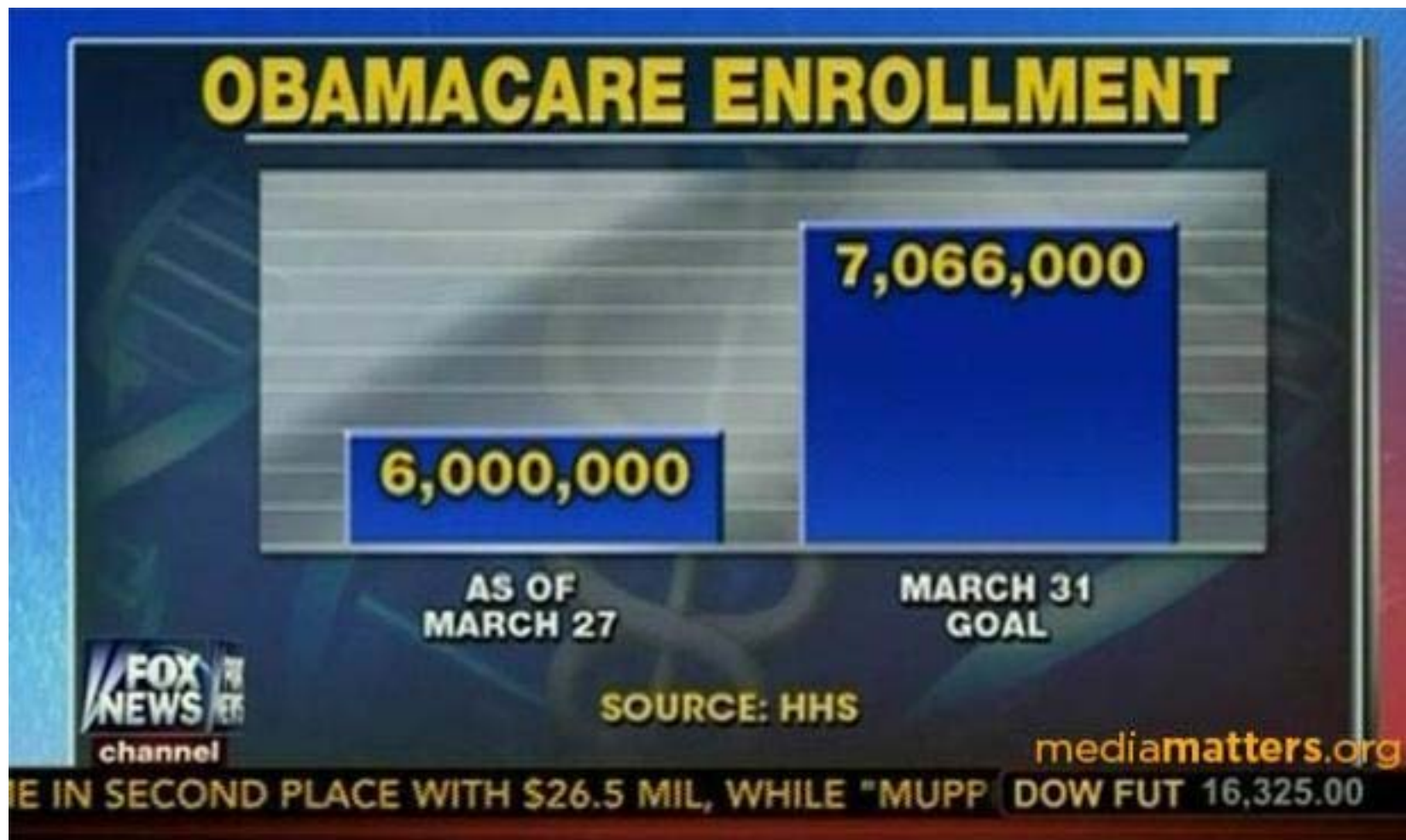
* Figures do not include the eight deaths from the Jan. 19, 2011, grand jury indictment charging abortion doctor Kermit Gosnell with murder. The Police Department adds these deaths to the 2011 total, even though they occurred in previous years, because the FBI requires homicides to be counted in the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

Camden homicides by year

Camden homicides have also risen for the last two years.



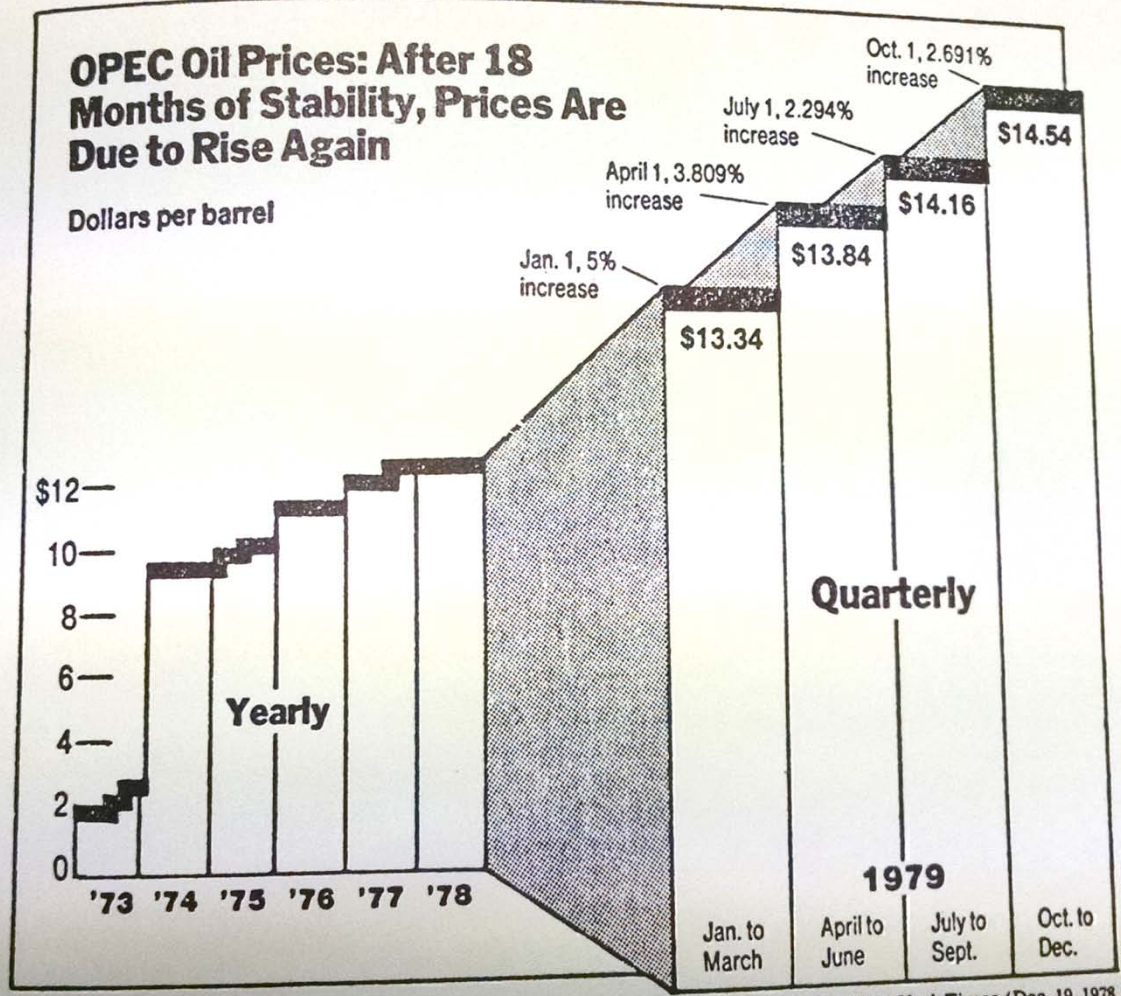
Misinformation...



Disinformation.

Design Variation

Consistency helps



The Visual Display of Quantitative Information 2E, Tufte 2001

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



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

Yesterday's results

What was the best part of the Super Bowl?



Vote at winnipegsun.com

WHAT'S

OPINION

EDITOR

Obama's Divided Nation

Obama presides over a more divided America than any time in 50 years that was riven by racial lines gathering in 2008 to elect its president. That president four years dividing the basis of economic. The campaign revealed no evidence 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-

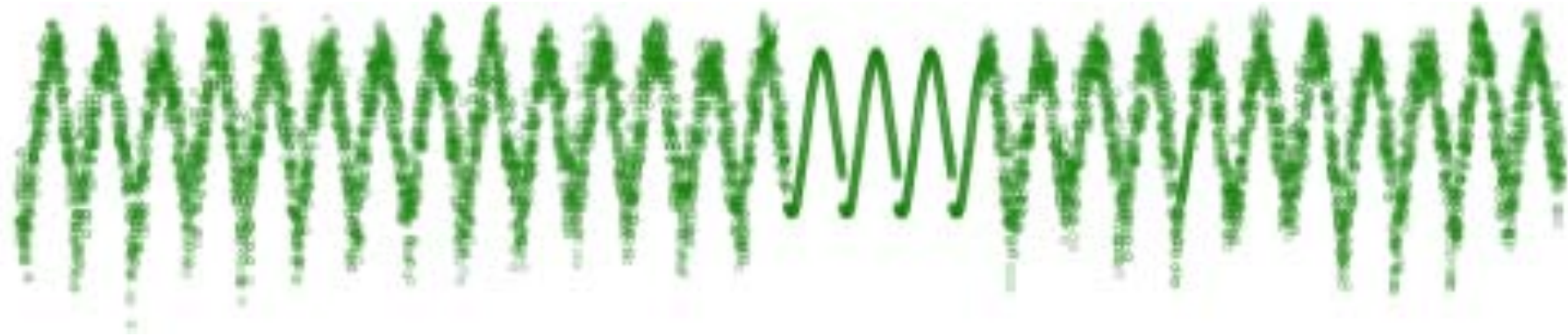
Obama spoketotum replied: if you thought you're wrong. It

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

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



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



Copyright & Trademarks

“It’s fair use, man!”

Copyright

original works of authorship

Writing

Music

Drawings

Art

Movies

Dance

*what can't be
protected?*

Fundamentally Free

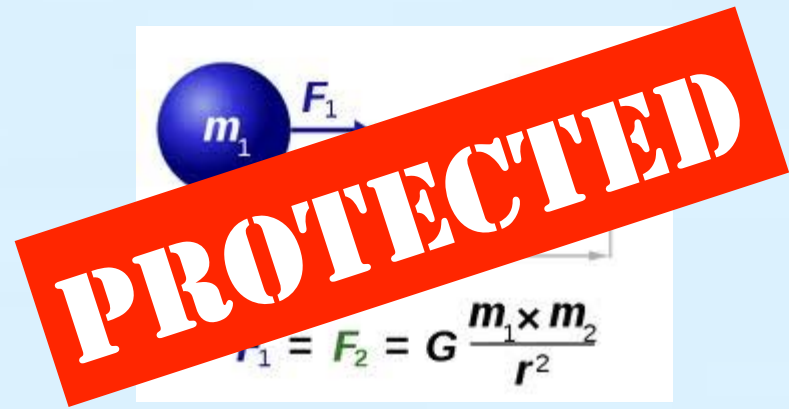
Ideas

Concepts

Facts

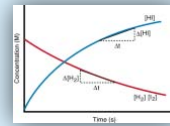
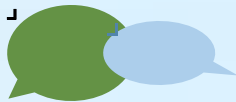
Principles





*when is it
protected?*

*when is it
protected?*



*who is
protected?*

Search: PubMed

[Limits](#) [Advanced search](#) [Help](#)

Search Clear

[Display Settings:](#) Abstract

Adicciones. 2009;21(4):327-45.

'Lager, lager shouting': the role of music and DJs in nightclub disorder control.

[Article in English, Spanish]

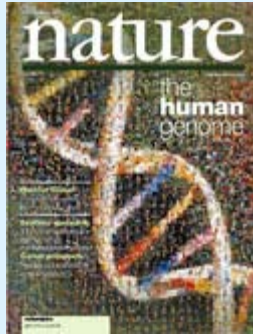
Forsyth AJ.

[Display Settings:](#) Abstract

Exp Clin Psychopharmacol. 2010 Aug;18(4):322-8.

Development of a simulated drinking game procedure to study risky alcohol use.

Correia CJ, Cameron JM.



Human Genome Initial sequencing and analysis of the human genome

Nature (2001) 409, 860-921

Nusbaum¹, Michael C. Zody¹, Jennifer Baldwin¹, Keri Devon¹, Ken Dewar¹, Michael Doyle¹, William FitzHugh¹, R. Funke¹, Diane Gage¹, Katrina Harris¹, Andrew Heaford¹, John Howland¹, Lisa Kann¹, Jessica Lehoczyk¹, Rosie LeVine¹, Paul McEwan¹, Kevin McKernan¹, James Meldrim¹, Jill P. Mesirov¹, Cher Miranda¹, William Morris¹, Jerome Naylor¹, Christina Raymond¹, Mark Rosetti¹, Ralph Santos¹, Andrew Sheridan¹, Carrie Sougnez¹, Nicole Stange-Thomann¹, Nikola Stojanovic¹, Aravind Subramanian¹ & Dudley Wyman¹ for Whitehead Institute for Biomedical Research, Center for Genome Research.; Jane Rogers², John Sulston², Rachael Ainscough², Stepha Beck², David Bentley², John Burton², Christopher Clee², Nigel Carter², Alan Coulson², Rebecca Deadman², Pano Deloukas², Andrew Dunham², Ian Dunham², Richard Durbin², Lisa French², Darren Grafham², Simon Gregory², T Hubbard², Sean Humphray², Adrienne Hunt², Matthew Jones², Christine Lloyd², Amanda McMurray², Lucy Matthews², Simon Mercer², Sarah Milne², James C. Mullikin², Andrew Mungall², Robert Plumb², Mark Ross², Ratu Shownkeen² & Sarah Sims² for The Sanger Centre.; Robert H. Waterston³, Richard K. Wilson³, LaDeana W. Hillier³, John D. McPherson³, Marco A. Marra³, Elaine R. Mardis³, Lucinda A. Fulton³, Asif T. Chinwalla³, Kymberli H. Pepin³, Warren R. Gish³, Stephanie L. Chissole³, Michael C. Wendt³, Kim D. Delehaunty³, Tracie L. Miner³, Andrew Delehaunty³, Jason B. Kramer³, Lisa L. Cook³, Robert S. Fulton³, Douglas L. Johnson³, Patrick J. Minx³ & Sandra W. Clifton³ for Washington University Genome Sequencing Center, Trevor Hawkins⁴, Elbert Branscomb⁴, Paul Predki⁴, Paul Richardson⁴, Sarah Wenning⁴, Tom Slezak⁴, Norman Doggett⁴, Jan-Fang Cheng⁴, Anne Olsen Susan Lucas⁴, Christopher Elkin⁴, Edward Uberbacher⁴ & Marvin Frazier⁴ for US DOE Joint Genome Institute.; Richard A. Gibbs⁵, Donna M. Muzny⁵, Steven E. Scherer⁵, John B. Bouck⁵, Erica J. Sodergren⁵, Kim C. Worley⁵, Catherine M. Rives⁵, James H. Gorrell⁵, Michael L. Metzker⁵, Susan L. Naylor⁶, Raju S. Kucherlapati⁷, David L. Nelson & George M. Weinstock⁸ for Baylor College of Medicine Human Genome Sequencing Center.; Yoshiyuki Sakaki⁹, Asao Fujiyama⁹, Masahira Hattori⁹, Tetsushi Yada⁹, Atsushi Toyoda⁹, Takehiko Itoh⁹, Chiharu Kawagoe⁹, Hidemi Watanabe⁹, Yasushi Totoki⁹ & Todd Taylor⁹ for RIKEN Genomic Sciences Center.; Jean Weissenbach¹⁰, Roland Heilig¹⁰, William Saurin¹⁰, Francois Artiguenave¹⁰, Philippe Brottier¹⁰, Thomas Bruls¹⁰, Eric Pelletier¹⁰, Catherine Robert¹⁰ & Patrick Wincker¹⁰ for Genoscope and CNRS UMR-8030.; André Rosenthal¹², Matthias Platzer¹², Gerald Nyakatura¹², Stefan Taudien¹² & Andreas Rump¹² for Department of Genome Analysis, Institute Molecular Biotechnology.; Douglas R. Smith¹¹, Lynn Doucette-Stamm¹¹, Marc Rubenfield¹¹, Keith Weinstock¹¹, Hong Mei Lee¹¹ & JoAnn Dubois¹¹ for GTC Sequencing Center.; Huanming Yang¹³, Jun Yu¹³, Jian Wang¹³, Guyar Huang¹⁴ & Jun Gu¹⁵ for Beijing Genomics Institute/Human Genome Center.; Leroy Hood¹⁶, Lee Rowen¹⁶, Anup Madan¹⁶ & Shizen Qin¹⁶ for Multimegabase Sequencing Center, The Institute for Systems Biology.; Ronald W. Davis¹⁷, Nancy A. Federspiel¹⁷, A. Pia Abola¹⁷ & Michael J. Proctor¹⁷ for Stanford Genome Technology Center.; Bruce A. Roe²², Feng Chen²² & Huaqin Pan²² for University of Oklahoma's Advanced Center for Genome Technology.; Juliane Ramser²³, Hans Lehrach²³ & Richard Reinhardt²³ for Max Planck Institute for Molecular Genetics.; W. Richard McCombie²⁴, Melissa de la Bastide²⁴ & Neilay Dedhia²⁴ for Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center.; Helmut Blöcker²⁵, Klaus Hornischer²⁵ & Gabriele Nordsiek²⁵ for GBF—German Research Centre for Biotechnology.; Richa Agarwala²⁶, L. Aravind²⁶, Jeffrey A. Bailey²⁷, Alex Bateman², Serafim Batzoglou¹, Ewan Birney²⁸, Peer Bork^{29,30}, Daniel G. Brown¹, Christopher B. Burge³¹, Lorenzo Cerutti²⁸, Hsiu-Chuan Chen²⁶, Deanna Church²⁶, Michele Clamp², Richard R. Copley³⁰, Tobias Doerks^{29,30}, Sean R. Eddy³², Evan E. Eichler²⁷, Terrence S. Furey³³, James Galagan¹, James G. R. Gilbert², Cyrus Harmon³⁴, Yoshihide Hayashizaki³⁵, David Haussler³⁶, Henning Hermjakob²⁸, Karsten Hokamp³⁷, Wonhee Jang²⁶, L. Steven Johnson³¹, Thomas A. Jones³², Simon Kasif³⁸, Arek Kasprzyk²⁸, Scot Kennedy³⁹, W. James Kent⁴⁰, Paul Kitts²⁶, Eugene V. Koonin²⁶, Ian Korf³, David Kulp³⁴, Doron Lancet⁴¹, Todd M. Lowe⁴², Aoife McLysaght³⁷, Tarjei Mikkelsen³⁸, John V. Moran⁴³, Nicola Mulder²⁸, Victor J. Pollara¹, Chris P. Ponting⁴⁴, Greg Schuler²⁶, Jörg Schultz³⁰, Guy Slater²⁸, Ariar F. A. Smit⁴⁵, Elia Stupka²⁸, Joseph Szustakowski³⁸, Danielle Thierry-Mieg²⁶, Jean Thierry-Mieg²⁶, Lukas Wagner²⁶, John Wallis³, Raymond Wheeler³⁴, Alan Williams³⁴, Yuri I. Wolf²⁶, Kenneth H. Wolfe³⁷, Shiao-Pyng Yang³ & Ru-Fang Yeh³¹ for *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings); Francis Collins⁴⁶, Mark S. Guyer⁴⁶, Jane Peterson⁴⁶, Adam Felsenfeld⁴⁶ & Kris A. Wetterstrand⁴⁶ for Scientific management: National Human Genome Research Institute, US National Institutes of Health.; Richard M. Myers¹⁸, Jeremy Schmutz¹⁸, Mark Dickson¹⁸, Jane Grimwood¹⁸ & David R. Cox¹⁸ for Stanford Human Genome Center.; Maynard V. Olson¹⁹, Rajinder Kaul¹⁹ & Christopher Raymond¹⁹ for University of Washington Genome Center.; Nobuyoshi Shimizu²⁰, Kazuhiko Kawasaki²⁰ & Shinsei Minoshima²⁰ for Department of Molecular Biology,