

Bad Statistics

Recognizing Junk

Anders O.F. Hendrickson

Concordia College
Moorhead, MN

February 10, 2010

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts

Outline

- 1 Dirty Data
 - Study (What they asked)
 - Sample Bias (Whom they asked)
 - Sample Size (How many they asked)
- 2 Useless Numbers
 - Raw Numbers vs. Proportions
 - Pointless Numbers
 - Lack of context
- 3 Visualization Gaffes
 - Axes of Evil / Original Sin
 - Area is Quadratic
 - Chart Junk
- 4 Correlation & Causation
- 5 Defense Against the Dark Arts

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

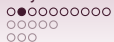
- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

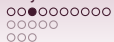
5 Defense Against the Dark Arts



Study Questions

Principle

What questions they asked influence what answers they got.



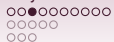
Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- Shocking!
- But what is “bullying”?



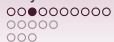
Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- Shocking!
- But what is “bullying”?



Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- Shocking!
- But what is “bullying”?

Example: Bullying

A Federal definition

Bullying can take three forms:

- *physical (hitting, kicking, spitting, pushing, taking personal belongings);*
- *verbal (taunting, malicious teasing, name calling, making threats);*
- *and psychological (spreading rumors, manipulating social relationships, or engaging in social exclusion, or intimidation).*

U.S. Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.

- Think back to *your* junior high days.
- Doesn't 30% seem awfully low?



Example: Bullying

A Federal definition

Bullying can take three forms:

- *physical (hitting, kicking, spitting, pushing, taking personal belongings);*
- *verbal (taunting, malicious teasing, name calling, making threats);*
- *and psychological (spreading rumors, manipulating social relationships, or engaging in social exclusion, or intimidation).*

U.S. Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.

- Think back to *your* junior high days.
- Doesn't 30% seem awfully low?

Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- So the definition of bullying is awfully broad.
- Moreover, this 30% figure includes both bullies *and* victims. Victims alone are 17%.
- Moreover, the study counted anyone who encountered “moderate” bullying—once a month or so. Only 8.4% were victims of weekly bullying.
- The study’s authors chose the figures that would produce a more sensational number.



Example: Bullying

Bullying Behaviors Among U.S. Youth

*Nearly 30% of U.S. children in sixth through 10th grades **have been bullied, or bully other students themselves.***

Journal of the American Medical Association, April 25, 2001

- So the definition of bullying is awfully broad.
- Moreover, this 30% figure includes both bullies *and* victims. Victims alone are 17%.
- Moreover, the study counted anyone who encountered “moderate” bullying—once a month or so. Only 8.4% were victims of weekly bullying.
- The study’s authors chose the figures that would produce a more sensational number.



Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- So the definition of bullying is awfully broad.
- Moreover, this 30% figure includes both bullies *and* victims. Victims alone are 17%.
- Moreover, the study counted anyone who encountered “moderate” bullying—once a month or so. Only 8.4% were victims of weekly bullying.
- The study’s authors chose the figures that would produce a more sensational number.



Example: Bullying

Bullying Behaviors Among U.S. Youth

Nearly 30% of U.S. children in sixth through 10th grades have been bullied, or bully other students themselves.

Journal of the American Medical Association, April 25, 2001

- So the definition of bullying is awfully broad.
- Moreover, this 30% figure includes both bullies *and* victims. Victims alone are 17%.
- Moreover, the study counted anyone who encountered “moderate” bullying—once a month or so. Only 8.4% were victims of weekly bullying.
- The study’s authors chose the figures that would produce a more sensational number.



Example: Stimulus jobs saved

Analysis finds stimulus confusion

The administration reported Friday that stimulus recipients reported having created or saved 640,329 jobs this year, a figure it said buttressed its contention that the \$787 billion package has had a significant economic impact.

USA Today, 11/18/2009

- The federal government sent Bob Bray \$26,174 in economic stimulus aid to fix a fence and replace the roofs on public apartments in Blooming Grove, Texas, a town of fewer than 900 people outside Dallas. He hired five roofers and an inspector to do the job. But the number of jobs he reported to the government looked very different—450 jobs. ... He said that he told the government that he had created six jobs but that a federal official told him that wasn't right. So he reported the number of hours the roofers worked instead.

Example: Stimulus jobs saved

Analysis finds stimulus confusion

The administration reported Friday that stimulus recipients reported having created or saved 640,329 jobs this year, a figure it said buttressed its contention that the \$787 billion package has had a significant economic impact.

USA Today, 11/18/2009

- The federal government sent Bob Bray \$26,174 in economic stimulus aid to fix a fence and replace the roofs on public apartments in Blooming Grove, Texas, a town of fewer than 900 people outside Dallas. He hired five roofers and an inspector to do the job. But the number of jobs he reported to the government looked very different—450 jobs. ... He said that he told the government that he had created six jobs but that a federal official told him that wasn't right. So he reported the number of hours the roofers worked instead.



Example: Stimulus jobs saved

Analysis finds stimulus confusion

The administration reported Friday that stimulus recipients reported having created or saved 640,329 jobs this year, a figure it said buttressed its contention that the \$787 billion package has had a significant economic impact.

USA Today, 11/18/2009

- The police department in Plymouth, Conn., claimed in its report that a \$15,355 grant used to buy new computers had created or saved 108 jobs. The department had 22 law enforcement officers last year, according to the FBI. Mayor Vincent Festa said that the town . . . had no plans to lay off any of its police officers, even without the stimulus. He said he could not explain the report, and the town's police chief did not return telephone calls Monday.



Example: Stimulus jobs saved

Analysis finds stimulus confusion

The administration reported Friday that stimulus recipients reported having created or saved 640,329 jobs this year, a figure it said buttressed its contention that the \$787 billion package has had a significant economic impact.

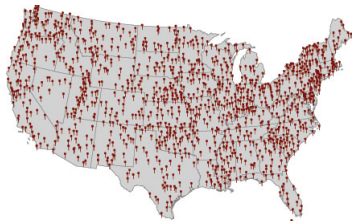
USA Today, 11/18/2009

- The Southwest Georgia Community Action Council, which employs about 500 people in its Head Start preschool program, reported creating or saving 935 jobs with about \$1.3 million in funding. Beverly Wise, the group's fiscal officer, said she multiplied the 508 employees by 1.84, the percentage pay raise they received, and came up with 935 jobs saved. "I thought it was high," Wise said of the number she reported, adding that the process was confusing.

Example: Climate Data

One of the chief sources for statistics about climate in the United States is a network of 1,218 sensors across the country run by the U.S. Historical Climatology Network (USHCN).

<http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html>



Each station is a thermometer that records the daily maximum and minimum temperatures.

Example: Climate Data



USHCN station in Morris, MN

- Note how the Morris station is in an open field.
- It should get an accurate reading.

Example: Climate Data



USHCN station in Detroit Lakes, MN

- Notice the two air conditioning units, one 6 ft. from station and one 8 ft away.
- Readings show a 4°C jump about 1999, when the A/C units were placed in this location

Example: Climate Data



Marysville, CA

- Asphalt, concrete, and more air conditioners

Example: Climate Data



Hopkinsville, KY

- Asphalt, brick, and a Weber grill

Example: Climate Data



Redding, CA

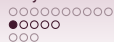
- A light bulb inside the thermometer shelter

Example: Climate Data



Tahoe City, CA

- A burn barrel?!?



Outline

1

Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2

Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3

Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4

Correlation & Causation

5

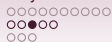
Defense Against the Dark Arts



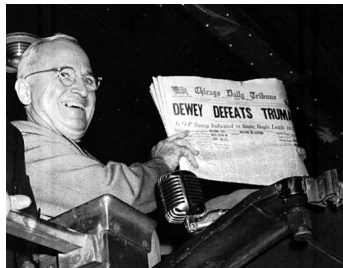
Sample Bias

Principle

Whom they asked influences what answers they got.



Example: Dewey Beats Truman



- Telephone polls in 1948 predicted Dewey would win.
- Dewey's support was urban, and Truman's support was rural.
- Thus the sample included more Dewey supporters than the population at large.



Example: Online Polls

CONSUMERS ARE MORE CONCERNED OVER PAYING DOWN DEBT THAN HOLIDAY SPENDING

Silver Spring, MD – 12/1/2009 – An online poll conducted in November on the National Foundation for Credit Counseling's (NFCC) Web site, www.DebtAdvice.org, revealed that consumers are much more concerned about paying down their existing debt load than they are with spending on holiday purchases.

- The sample: people who visited www.debtadvice.org
- The population: “consumers”—i.e. all consumers
- The sample is probably biased!



Example: Online Polls

CONSUMERS ARE MORE CONCERNED OVER PAYING DOWN DEBT THAN HOLIDAY SPENDING

Silver Spring, MD – 12/1/2009 – An online poll conducted in November on the National Foundation for Credit Counseling's (NFCC) Web site, www.DebtAdvice.org, revealed that consumers are much more concerned about paying down their existing debt load than they are with spending on holiday purchases.

- The sample: people who visited www.debtadvice.org
- The population: “consumers”—i.e. all consumers
- The sample is probably biased!



Example: Online Polls

CONSUMERS ARE MORE CONCERNED OVER PAYING DOWN DEBT THAN HOLIDAY SPENDING

Silver Spring, MD – 12/1/2009 – An online poll conducted in November on the National Foundation for Credit Counseling's (NFCC) Web site, www.DebtAdvice.org, revealed that **consumers** are much more concerned about paying down their existing debt load than they are with spending on holiday purchases.

- The sample: people who visited www.debtadvice.org
- The population: "**consumers**"—i.e. all consumers
- The sample is probably biased!



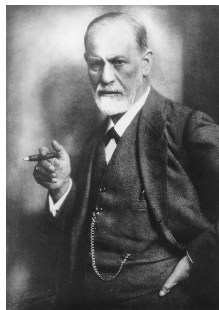
Example: Online Polls

CONSUMERS ARE MORE CONCERNED OVER PAYING DOWN DEBT THAN HOLIDAY SPENDING

Silver Spring, MD – 12/1/2009 – An online poll conducted in November on the National Foundation for Credit Counseling's (NFCC) Web site, www.DebtAdvice.org, revealed that consumers are much more concerned about paying down their existing debt load than they are with spending on holiday purchases.

- The sample: people who visited www.debtadvice.org
- The population: “consumers”—i.e. all consumers
- The sample is probably biased!

Example: Freud



- Sigmund Freud made sweeping claims about human psychology: id, ego, and superego, love and death wishes, Oedipus and Electra complexes, etc.
- These were based on lengthy “analysis” of 145 patients
- But these patients were all upper class and Viennese!

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts



Sample Size

Principle

How many people they asked affects the reliability of their conclusions.

If you want to draw conclusions about a whole population by studying only a sample, that sample needs to be large.



Example: Toothpaste Ad

Dr. Cornish's Tooth Powder ad

Our product has shown considerable success in correction of
... cavities.



Example: Toothpaste Ad

Dr. Cornish's Tooth Powder ad

Our product has shown considerable success in correction of . . . cavities.

- This claim was based on just one experiment, and it involved only six patients.

Outline

1

Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2

Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3

Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4

Correlation & Causation

5

Defense Against the Dark Arts

Reporting Useless Numbers

Bad statistics also arise when the reporter, politician, or analyst chooses the *wrong statistics* to report, or carefully picks the one statistic that shows the desired result.

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts

Example: Stimulus bonuses



[Home](#)
[Business](#)
[Business & Markets](#)

Insurance giant AIG to pay \$165 million in bonuses

By Associated Press

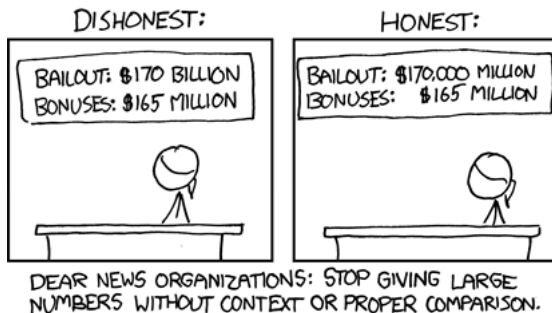
Sunday, March 15, 2009 - Added 330d 12h ago

[E-mail](#)
[Print](#)
[\(5\) Comments](#)
[Text size](#)
[Share](#)
[Buzz up!](#)

WASHINGTON — American International Group is giving its executives tens of millions of dollars in new bonuses even though it received a taxpayer bailout of more than \$170 billion dollars.

AIG is paying out the executive bonuses to meet a Sunday deadline, but the troubled insurance giant has agreed to administration requests to restrain future payments.

Example: Stimulus bonuses



- So that's less than 0.1% of the total.
- It's still a bad thing, but not as bad as it sounded.

Example: Raw dollars and inflation

UM-Twin Cities Tuition

raw dollars

1960–61	\$213
2009–10	\$10,320
% increase	4,745%

Source: <http://www.irr.umn.edu/tuition/TuitionUMNTC.pdf>

oooooooooooo
 oooooo
 ooo

ooo●o
 oo
 ooo

ooooo
 oooo
 ooooooooo

Example: Raw dollars and inflation

UM-Twin Cities Tuition

	raw dollars	2009 dollars
1960–61	\$213	\$1,544
2009–10	\$10,320	\$10,320
% increase	4,745%	568%

Source: <http://www.irr.umn.edu/tuition/TuitionUMNTC.pdf>



Example: Raw dollars and inflation

UM-Twin Cities Tuition

	raw dollars	2009 dollars
1960–61	\$213	\$1,544
2009–10	\$10,320	\$10,320
% increase	4,745%	568%

Fortunately, most charts do account for inflation—
but always be sure to check!

Source: <http://www.irr.umn.edu/tuition/TuitionUMNTC.pdf>



Example: US budget freeze

The New York Times

Politics

WORLD
U.S.
N.Y. / REGION
BUSINESS
TECHNOLOGY
SCIENCE
HEALTH
SPORTS
OPINION

Search Politics
Go

The Caucus »

The 44th President
 The 111th Congress

Obama Budget Freezes Much Domestic Spending

By JACKIE CALMES and ROBERT PEAR
Published: January 30, 2010

WASHINGTON — [President Obama](#) will send a \$3.8 trillion budget to Congress on Monday for the coming fiscal year that would increase financing for education and for civilian research programs by more than 6 percent and provide \$25 billion for cash-starved states, even as he seeks to freeze much domestic spending for the rest of his term.

[Enlarge This Image](#)

Chip Somodevilla/Getty Images
Douglas W. Elmendorf, head of the Congressional Budget Office, at a Senate hearing on Thursday.

The budget for the 2011 fiscal year, which begins in October, will identify the winners and losers behind Mr. Obama's proposal for a three-year freeze of a portion of the budget. Many programs at the [National Institutes of Health](#), the [National Science Foundation](#) and the Energy Department are in line for increases, along with the [Census Bureau](#).

☒ SIGN IN TO RECOMMEND
☐ TWITTER
☐ SIGN IN TO E-MAIL
☐ PRINT
☐ REPRINTS
☐ SHARE

CRAZY HEART
NOW PLAYING
3 ACADEMY AWARD NOMINATIONS

Outline

1

Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2

Useless Numbers

- Raw Numbers vs. Proportions
- **Pointless Numbers**
- Lack of context

3

Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4

Correlation & Causation

5

Defense Against the Dark Arts



Example: Lake Pollution

Acid Rain Report

More than 1,000 lakes 10 acres or larger damaged [by acid rain] in the USA and thousands more in Canada.

USA Today, March 1990

- So what?
- How many lakes of that size *are* there in the USA?
- (Minnesota alone has 11,842 lakes over 10 acres.)
- Isn't something wrong with a statistic that treats Silver Lake of Hawley, MN (114 acres) the same as Lake Superior (20,000,000 acres)?

Example: Lake Pollution

Acid Rain Report

More than 1,000 lakes 10 acres or larger damaged [by acid rain] in the USA and thousands more in Canada.

USA Today, March 1990

- So what?
- How many lakes of that size *are* there in the USA?
- (Minnesota alone has 11,842 lakes over 10 acres.)
- Isn't something wrong with a statistic that treats Silver Lake of Hawley, MN (114 acres) the same as Lake Superior (20,000,000 acres)?

Example: Lake Pollution

Acid Rain Report

More than 1,000 lakes 10 acres or larger damaged [by acid rain] in the USA and thousands more in Canada.

USA Today, March 1990

- So what?
- How many lakes of that size *are* there in the USA?
- (Minnesota alone has 11,842 lakes over 10 acres.)
- Isn't something wrong with a statistic that treats Silver Lake of Hawley, MN (114 acres) the same as Lake Superior (20,000,000 acres)?

Example: Lake Pollution

Acid Rain Report

More than 1,000 lakes 10 acres or larger damaged [by acid rain] in the USA and thousands more in Canada.

USA Today, March 1990

- So what?
- How many lakes of that size *are* there in the USA?
- (Minnesota alone has 11,842 lakes over 10 acres.)
- Isn't something wrong with a statistic that treats Silver Lake of Hawley, MN (114 acres) the same as Lake Superior (20,000,000 acres)?

Example: Lake Pollution

Acid Rain Report

More than 1,000 lakes 10 acres or larger damaged [by acid rain] in the USA and thousands more in Canada.

USA Today, March 1990

- So what?
- How many lakes of that size *are* there in the USA?
- (Minnesota alone has 11,842 lakes over 10 acres.)
- Isn't something wrong with a statistic that treats Silver Lake of Hawley, MN (114 acres) the same as Lake Superior (20,000,000 acres)?



Example: Whisky Ads

An ad for Cutty Sark in the 1980s

In blind taste tests, a surprising *half* of non-Cutty Sark drinkers preferred the taste of Cutty Sark.



Example: Whisky Ads

An ad for Cutty Sark in the 1980s

In blind taste tests, a surprising *half* of non-Cutty Sark drinkers preferred the taste of Cutty Sark.

- In other words, it was just random.



Outline

1

Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2

Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3

Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4

Correlation & Causation

5

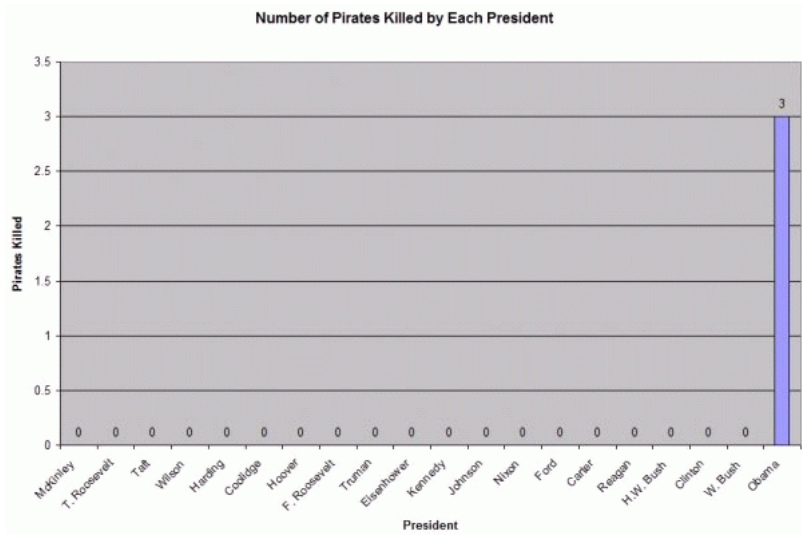
Defense Against the Dark Arts

oooooooooooo
oooooo
ooo

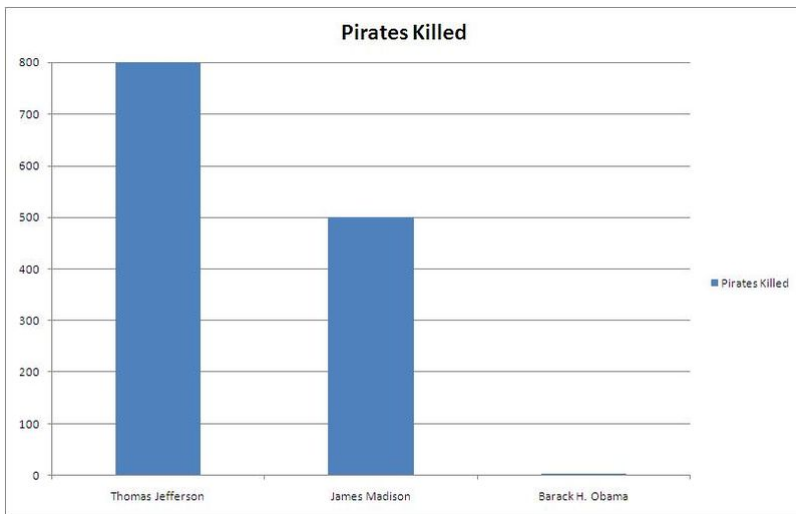
ooooo
ooo
o●o

ooooo
oooo
ooooooooo

Example: Pirates and Presidents



Example: Pirates and Presidents



Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts

Visualization Gaffes

- These days, statistics are often presented to us visually in charts and graphs.
- Computer software has made it lightning-fast to create them.
- Sadly, often very little thought goes into designing them.

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts

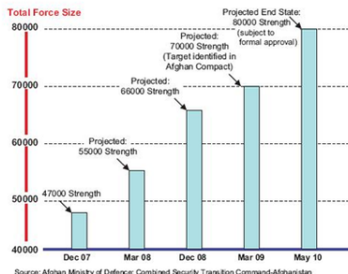
oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

o●ooo
 oooo
 ooooooooo

Example: Afghan Army

FIGURE 1: PROJECTED AFGHAN NATIONAL ARMY GROWTH



Source: Afghan Ministry of Defence; Combined Security Transition Command-Afghanistan

- Looks like dramatic growth, doesn't it?
- But look at the vertical axis! It starts at 40,000, not zero.
- So the growth is really only 70%—not 471%.
- Also look at the horizontal axis—the times aren't evenly spaced.

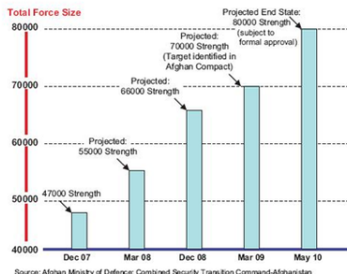
oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

o●ooo
 oooo
 ooooooooo

Example: Afghan Army

FIGURE 1: PROJECTED AFGHAN NATIONAL ARMY GROWTH

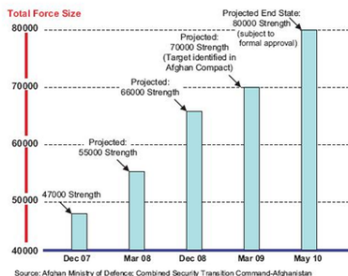


- Looks like dramatic growth, doesn't it?
- But look at the vertical axis! It starts at 40,000, not zero.
- So the growth is really only 70%—not 471%.
- Also look at the horizontal axis—the times aren't evenly spaced.

○○○○○○○○○○
○○○○○
○○○○○○○○
○○○
○○○●○○○
○○○
○○○○○○○

Example: Afghan Army

FIGURE 1: PROJECTED AFGHAN NATIONAL ARMY GROWTH

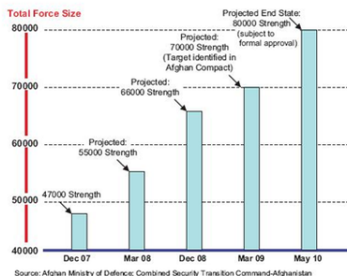


- Looks like dramatic growth, doesn't it?
- But look at the vertical axis! It starts at 40,000, not zero.
- So the growth is really only 70%—not 471%.
- Also look at the horizontal axis—the times aren't evenly spaced.

○○○○○○○○○○
○○○○○
○○○○○○○○
○○○
○○○●○○○
○○○
○○○○○○○

Example: Afghan Army

FIGURE 1: PROJECTED AFGHAN NATIONAL ARMY GROWTH



- Looks like dramatic growth, doesn't it?
- But look at the vertical axis! It starts at 40,000, not zero.
- So the growth is really only 70%—not 471%.
- Also look at the horizontal axis—the times aren't evenly spaced.

oooooooooooo
ooooo
oooo
ooo

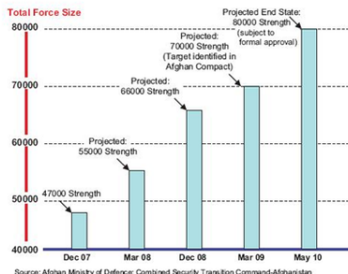
ooooo
ooo
ooo
ooo

●●○○○
○○○
oooooooo

Example: Afghan Army

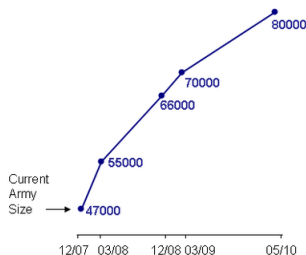
BAD

FIGURE 1: PROJECTED AFGHAN NATIONAL ARMY GROWTH



BETTER

Projected Afghan National Army Growth



- Looks like dramatic growth, doesn't it?
- But look at the vertical axis! It starts at 40,000, not zero.
- So the growth is really only 70%—not 471%.
- Also look at the horizontal axis—the times aren't evenly spaced.

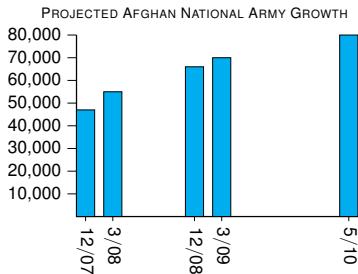
oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

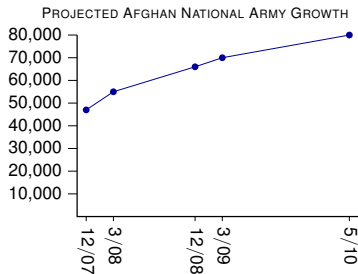
oo●oo
 oooo
 ooooooooo

Example: Afghan Army

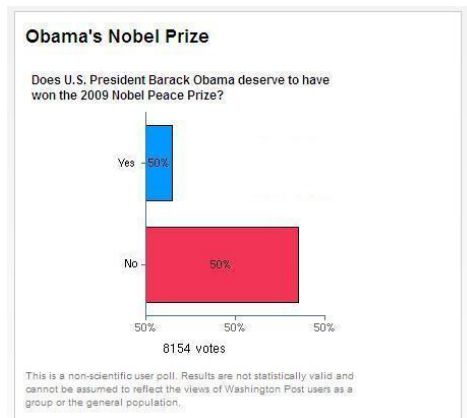
BEST



or



Example: Nobel Prize



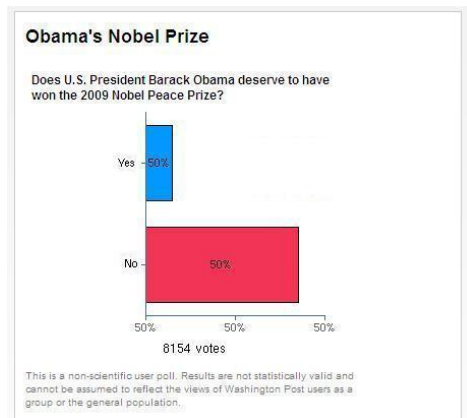
- The percentages are actually equal! Look at the axis.
- This automatically generated chart is very bad.
- The *Washington Post* eventually fixed it.

oooooooooooo
oooooo
ooo

ooooo
ooo
ooo

ooo●o
ooo
oooooooo

Example: Nobel Prize



- The percentages are actually equal! Look at the axis.
- This automatically generated chart is very bad.
- The *Washington Post* eventually fixed it.

oooooooooooo
 oooooo
 ooo

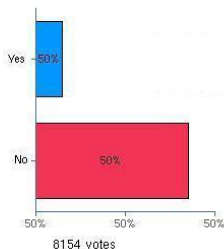
ooooo
 ooo
 ooo

ooo●o
 oooo
 ooooooooo

Example: Nobel Prize

Obama's Nobel Prize

Does U.S. President Barack Obama deserve to have won the 2009 Nobel Peace Prize?



This is a non-scientific user poll. Results are not statistically valid and cannot be assumed to reflect the views of Washington Post users as a group or the general population.

- The percentages are actually equal! Look at the axis.
- This automatically generated chart is very bad.
- The *Washington Post* eventually fixed it.

oooooooooooo
 oooooo
 ooo

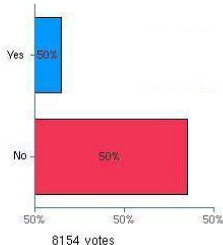
ooooo
 ooo
 ooo

ooo●o
 oooo
 oooooooo

Example: Nobel Prize

Obama's Nobel Prize

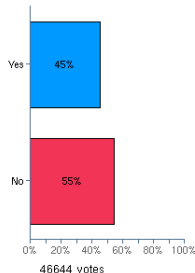
Does U.S. President Barack Obama deserve to have won the 2009 Nobel Peace Prize?



This is a non-scientific user poll. Results are not statistically valid and cannot be assumed to reflect the views of Washington Post users as a group or the general population.

Obama's Nobel Prize

Does U.S. President Barack Obama deserve to have won the 2009 Nobel Peace Prize?



This is a non-scientific user poll. Results are not statistically valid and cannot be assumed to reflect the views of Washington Post users as a group or the general population.

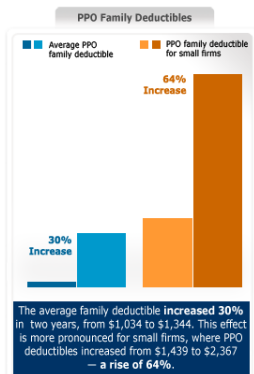
- The percentages are actually equal! Look at the axis.
- This automatically generated chart is very bad.
- The *Washington Post* eventually fixed it.

oooooooooooo
oooooo
ooo

ooooo
ooo
ooo

oooo●
oooo
oooooooo

Example: Health Care Costs



Source: www.healthreform.gov

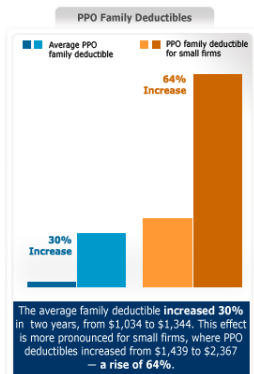
- The words say 30%, but the picture says 912%.
- The words say 64%, but the picture says 211%.

oooooooooooo
oooooo
ooo

ooooo
ooo
ooo

oooo●
oooo
oooooooo

Example: Health Care Costs



Source: www.healthreform.gov

- The words say 30%, but the picture says 912%.
- The words say 64%, but the picture says 211%.

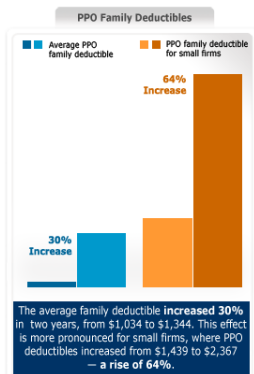
oooooooooooo
ooooo
ooo

ooooo
ooo
ooo

oooo●
ooo
oooooooo

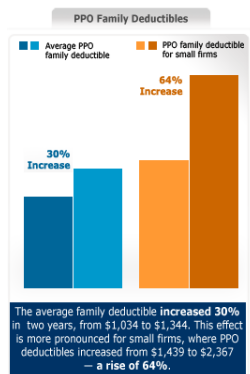
Example: Health Care Costs

WRONG



Source: www.healthreform.gov

FIXED



Source: www.politicalmathblog.com

- The words say 30%, but the picture says 912%.
- The words say 64%, but the picture says 211%.

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

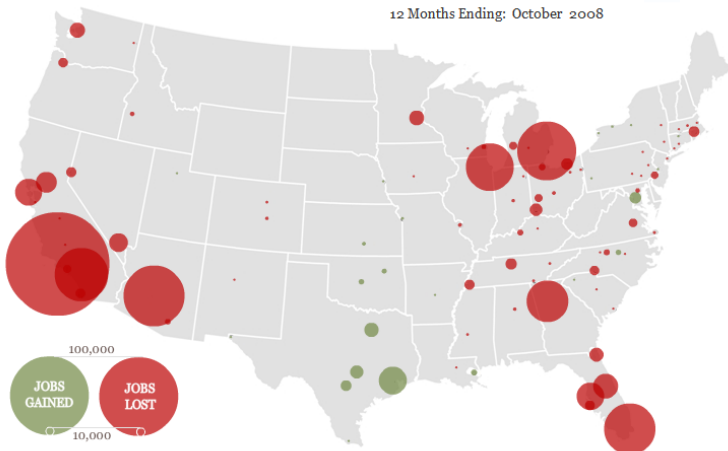
5 Defense Against the Dark Arts

oooooooooooo
oooooo
oooooooo
ooo
oooooooo
o●ooo
oooooooooooo

Example: Job Loss October 2008



Net Job Gains / Losses by Metropolitan Statistical Area,
12 Months Ending: October 2008



Note: Top 100 Metropolitan Statistical Areas in the contiguous United States based on number of jobs as of December 2008

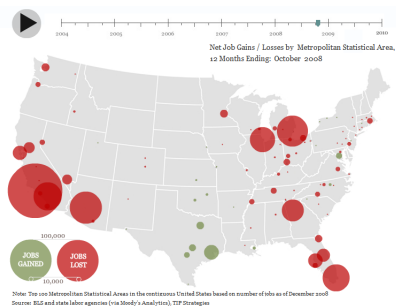
Source: BLS and state labor agencies (via Moody's Analytics). TIP Strategies

oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

ooooo
 ooo
 oooooo

Example: Job Loss October 2008



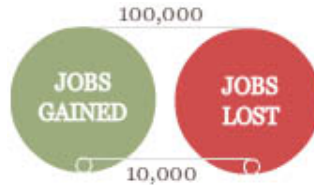
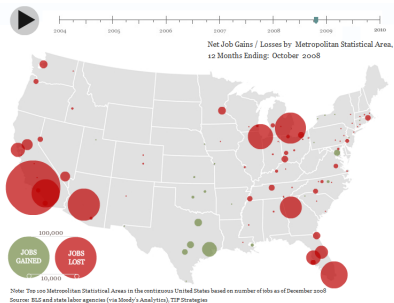
Source: <http://tipstrategies.com/archive/geography-of-jobs>

oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

ooooo
 ooo
 ooo
 oooooooo

Example: Job Loss October 2008



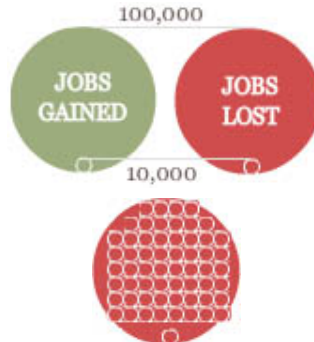
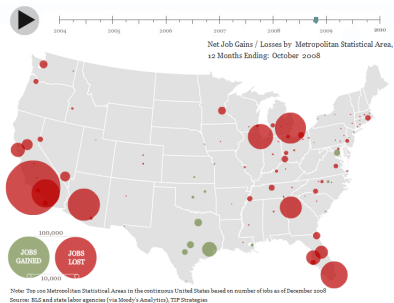
Source: <http://tipstrategies.com/archive/geography-of-jobs>

oooooooooooo
 oooooo
 ooo

ooooo
 ooo
 ooo

ooooo
 ooo
 ooo
 oooooooo

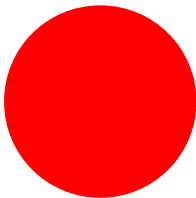
Example: Job Loss October 2008



Source: <http://tipstrategies.com/archive/geography-of-jobs>

oooooooooooo
ooooo
oooooooo
ooo
oooooooo
ooo●
ooooooooo

Example: Job Loss October 2008



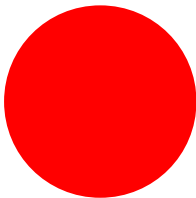
Los Angeles 127,200

oooooooo
ooooo
oooooooo
ooo
oooooooo
ooo●
oooooooo

Example: Job Loss October 2008



Minneapolis 18,400



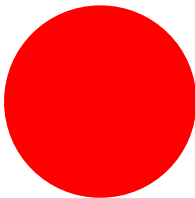
Los Angeles 127,200

oooooooo
ooooo
oooooooo
ooo
oooooooo
ooo●
oooooooo

Example: Job Loss October 2008

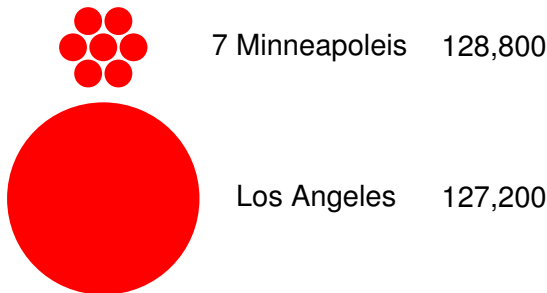


7 Minneapolis 128,800



Los Angeles 127,200

Example: Job Loss October 2008



- If L.A.'s job loss is only seven times Minneapolis's, shouldn't L.A.'s circle look only seven times as big?

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

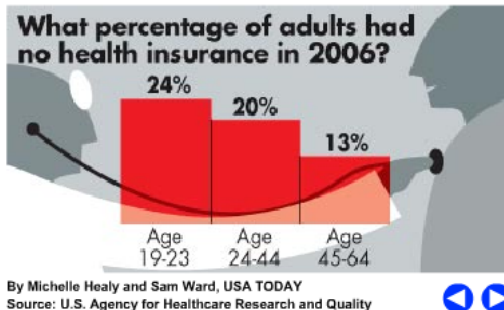
4 Correlation & Causation

5 Defense Against the Dark Arts

Chart Junk

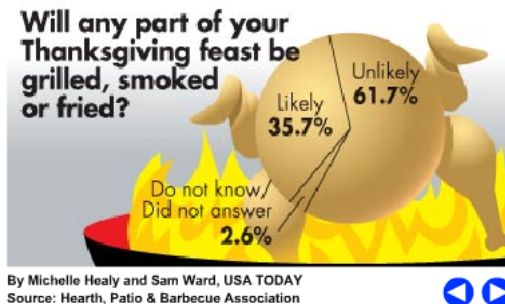
- “Chart Junk” is a name for the useless clutter newspapers use to dress up their charts to look more eye-catching.
- The end result is that they obscure the information.
- (The chief offender is *USA Today*.)

Example: Health Insurance



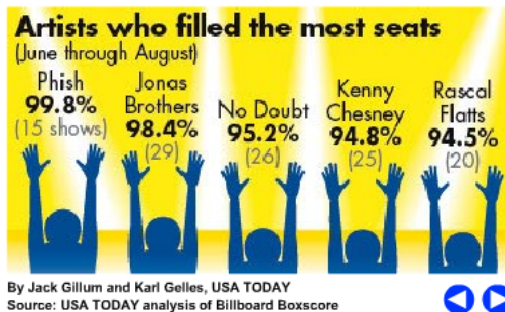
- The curved lines influence the eye.
- We should be comparing the heights of the bars, but instead our eyes compare the dark red regions.

Example: Turkey



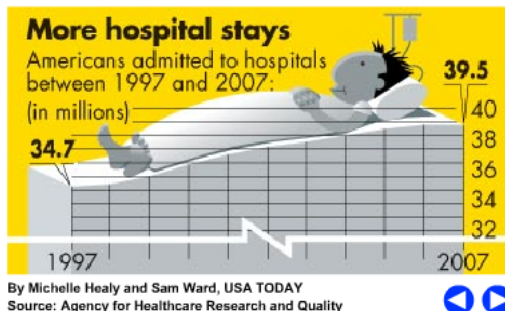
- This is an attempt at a pie chart.
- Note how the turkey's limbs throw off the proportions.
- So "Unlikely" seems even larger than it is.
- To which category does the lower left leg belong?

Example: Concert Ticket Sales



- Er... is this a bar chart of some sort?
- If so, where is the origin? Where are the scales?
- Should my eye be comparing the dark blue areas, or the heights of the fingertips, or what?

Example: Hospital Stays



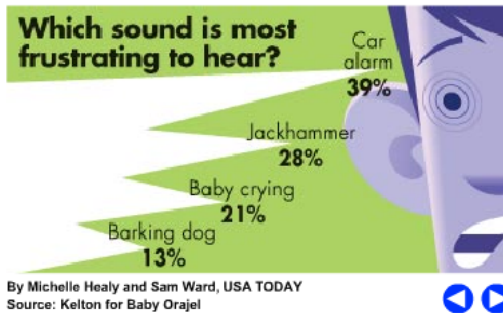
- All that junk just gets in the way.
- Note too an axis of evil:
it looks like the number of stays has doubled!

oooooooooooo
oooooo
ooo

ooooo
ooo
ooo

ooooo
oooo
oooooooo●o

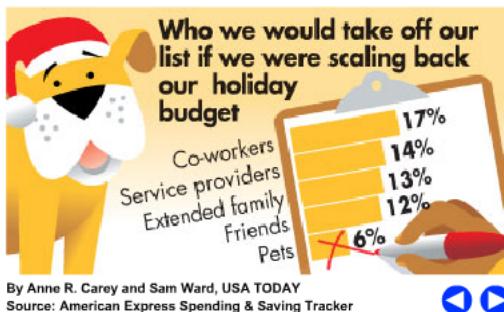
Example: Frustrating Sounds



- Where is the origin? What am I comparing?
- There are no clear areas to compare.
- As your eye travels left, the categories blend into one another.

Example: Pet Gifts

Among holiday shoppers who are scrimping on holiday gifts, few plan to spend less on their pets.



- The numbers and words say that very few people will stop giving gifts to pets.
- The red X suggests that most people are crossing pets off their Christmas lists.

Outline

1 Dirty Data

- Study (What they asked)
- Sample Bias (Whom they asked)
- Sample Size (How many they asked)

2 Useless Numbers

- Raw Numbers vs. Proportions
- Pointless Numbers
- Lack of context

3 Visualization Gaffes

- Axes of Evil / Original Sin
- Area is Quadratic
- Chart Junk

4 Correlation & Causation

5 Defense Against the Dark Arts

oooooooo
oooooo
oooooooo
ooo
oooooooo
oooo
oooooooo

Example: Candy and Violence

Sugar Rush to Prison?

Study Says Lots of Candy Could Lead to Violence

Children who eat too much candy may be more likely to be arrested for violent behavior as adults, new research suggests.

British experts studied more than 17,000 children born in 1970 for about four decades. Of the children who ate candies or chocolates daily at age 10, 69 percent were later arrested for a violent offense by the age of 34. Of those who didn't have any violent clashes, 42 percent ate sweets daily.

CBSnews.com, 10/1/2009

Example: Candy and Violence

Sugar Rush to Prison?

Study Says Lots of Candy Could Lead to Violence

Children who eat too much candy may be more likely to be arrested for violent behavior as adults, new research suggests.

(from further in the article)

The researchers said the results were interesting, but that more studies were needed to confirm the link. "It's not that the sweets themselves are bad, it's more about interpreting how kids make decisions," said Simon Moore of the University of Cardiff, one of the paper's authors.

Moore said parents who consistently bribe their children into good behavior with candies and chocolates could be doing harm. That might prevent kids from learning how to defer gratification, leading to impulsive behavior and violence.

Example: Candy and Violence

Sugar Rush to Prison?

Study Says Lots of Candy Could Lead to Violence

Children who eat too much candy may be more likely to be arrested for violent behavior as adults, new research suggests.

(from further in the article)

“It’s not that the sweets themselves are bad, it’s more about interpreting how kids make decisions.”

- How strong is this correlation? (What’s the r -value?)
- The headline asserts something the actual researchers don’t say.

Example: Candy and Violence

Sugar Rush to Prison?

Study Says Lots of Candy Could Lead to Violence

Children who eat too much candy may be more likely to be arrested for violent behavior as adults, new research suggests.

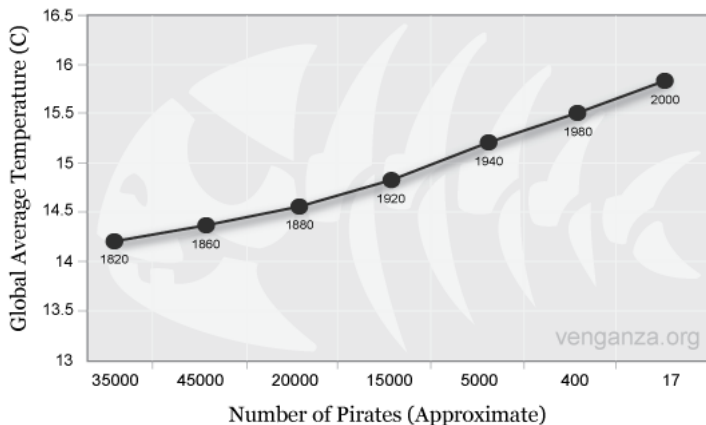
(from further in the article)

“It’s not that the sweets themselves are bad, it’s more about interpreting how kids make decisions.”

- How strong is this correlation? (What’s the r -value?)
- The headline asserts something the actual researchers don’t say.

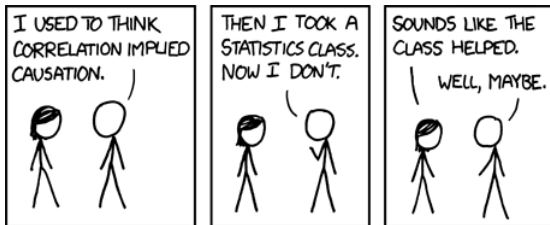
Example: Pirates & Global Warming Climate Change

Global Average Temperature Vs. Number of Pirates



oooooooooooo
oooooo
oooooooo
ooo
oooooooo
oooo
ooooooo

Correlation and Causation



oooooooooooo
ooooo
ooo

ooooo
ooo
ooo

ooooo
oooo
ooooooooo

Key Principle

Causation is very, very hard to prove.

Outline

- 1 Dirty Data
 - Study (What they asked)
 - Sample Bias (Whom they asked)
 - Sample Size (How many they asked)
- 2 Useless Numbers
 - Raw Numbers vs. Proportions
 - Pointless Numbers
 - Lack of context
- 3 Visualization Gaffes
 - Axes of Evil / Original Sin
 - Area is Quadratic
 - Chart Junk
- 4 Correlation & Causation
- 5 Defense Against the Dark Arts

Caution With All Statistics

- Examine the source of the study. Is it biased?
- Examine the wording of survey questions.
- Read the fine print.
- Examine the sample: how many people were studied, and how were they chosen?
- What do these numbers look like in proportion?
- If a poll calls itself “unscientific,” it’s not worth an instant of your time.

Caution With Graphs

- Look carefully at the origin and axes.
- Compare any numbers given to you on the graph.
- Spend as much time “reading” the graph as you would a lengthy description in words.

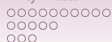
oooooooooooo
ooooo
ooo

ooooo
ooo
ooo

ooooo
oooo
ooooooooo

Caution With News Media

- Disbelieve the headline
- Quotes from the researchers are more likely to be accurate; when the journalist is summarizing, beware!



Dear News Media,

When reporting poll results, please keep in mind the following suggestions:

1. If two poll numbers differ by less than the margin of error, it's not a news story.
2. Scientific facts are not determined by public opinion polls.
3. A poll taken of your viewers/internet users is not a scientific poll.
4. What if all polls included the option "Don't care"?



Signed,

-Someone who took a
basic statistics course.

JORGE CHAM © 2010