

How to Lie, Cheat, Manipulate, and Mislead Using CTE Statistics and Graphs: A Primer

By

Krishnan Sudharsan

Data, Accountability, and Technical Assessments
Office of Career and Technical Education

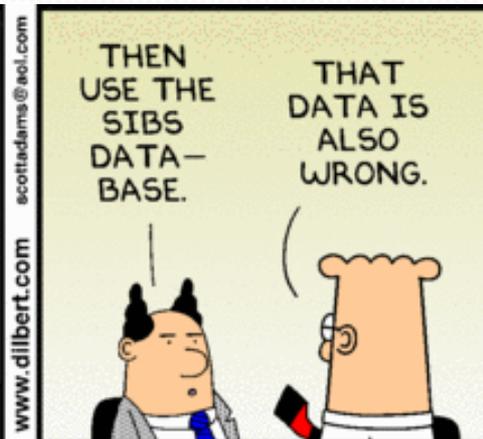
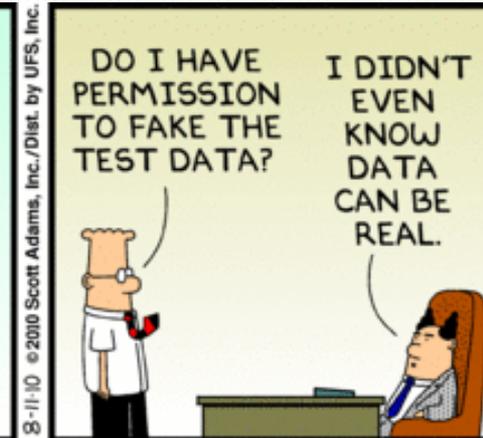
March 12, 2014



Credits

- Huff, Darrell. (1954). How to Lie with Statistics. New York: W.W. Norton & Company Inc.
- <http://cseweb.ucsd.edu/~ricko/CSE3/>
- <http://www.thepsychfiles.com/>

CTE and Data Misuse: Oh!, NEVER

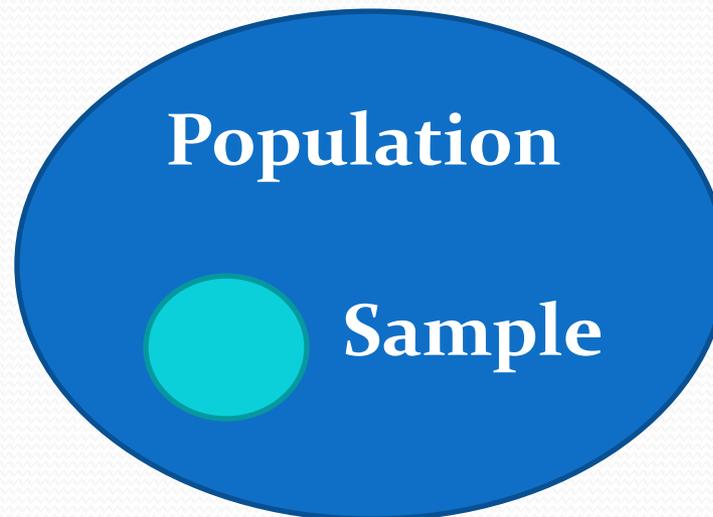


What is Statistics??

- It is the science of the
 - Collection
 - Organization
 - Analysis
 - Interpretation and
 - Presentation of data.

Collection of Data

- For analysis and interpretation of data to occur, the first step is collection.
- Data typically collected is a **sample** and is collected from a **population**



Sampling Bias

- Many different types of sampling bias. Examples include:
 - Self-Selection Bias
 - Area Bias
 - Leading Question Bias
 - Social Desirability Bias
 - Non-response Bias

Self-Selection Bias in CTE

- In Self-Selection Bias a participant's decision to participate may be correlated with traits that affect the study, making the participants a non-representative sample
- Example: What if OCTE allowed students to take an assessment in their area based on their own desire with no mandate? What would happen to 2S1?

Self-Selection Bias in CTE

- CTE Showcase at the Capitol (April 23rd)
 - Why might it be appropriate here?
 - Relates back to purpose/goals

Area Bias in CTE

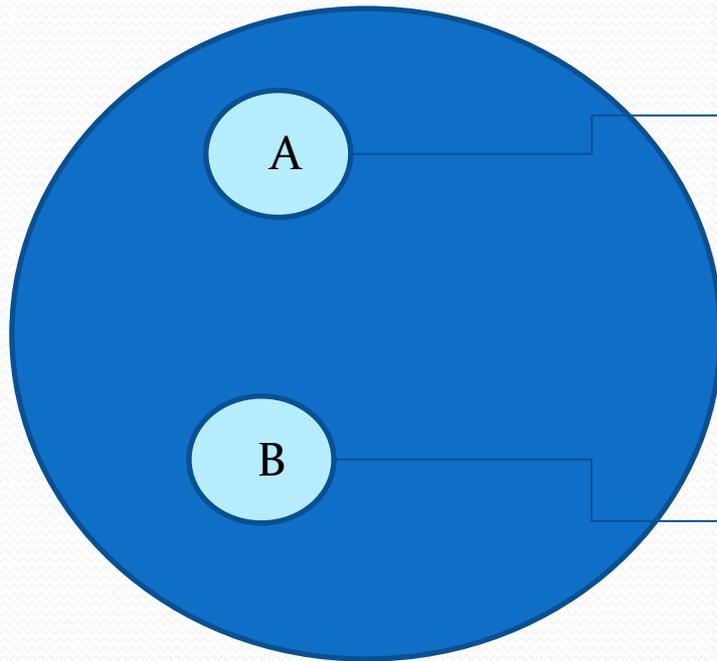


- Area Bias is introduced when the area of the sample is not representative of the study population

Area Bias in CTE

- Let's say you are use KeyTrain software as a formative assessment tool to measure reading/mathematics proficiency
- How can you introduce Area Bias to show improvement in proficiency scores?

Area Bias in CTE



Scores of Students in the
Transportation, Distribution
& Logistics Career Cluster

Scores of Students in the
Information Technology
Career Cluster

Leading Question Bias

- The tone of a question suggests an expected answer thereby biasing results
- Two slides ago, my question was “Why might self-selection bias be appropriate in the CTE showcase?”
- Versus asking “Does self-selection bias occur in the CTE showcase? If so, why?”

Social Desirability Bias

- Data can be biased when asking people questions regarding items considered socially undesirable
- e.g.
 - Do you brush your teeth twice each day?
 - How often do you recycle?

Non-Response Bias

- Non-response bias occurs in statistical surveys when the answers of respondents differs from the potential answers of those who did not respond
- e.g. Follow Up non-respondent survey
 - Conducted to see if answers differ between respondents vs. non-respondents

Non-Response Bias

- On TRAC visits OCTE checks
 - Response Rate (>80%) on the Follow Up survey
 - Proxy Rate (<30%) on the Follow Up survey
- Why??
 - Generally speaking, the lower the response rate, greater the likelihood of a non-response bias in play

What is Data Analysis??

- Data analysis is a process of gathering, modeling, and transforming data to meaningful information with the goal of highlighting this information, suggesting conclusions, and supporting decision making



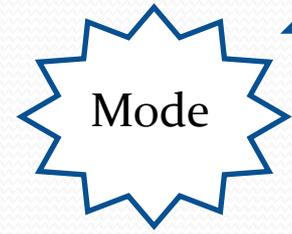
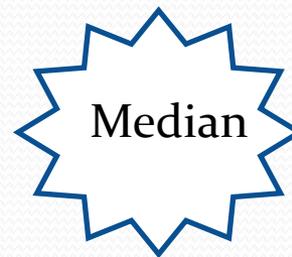
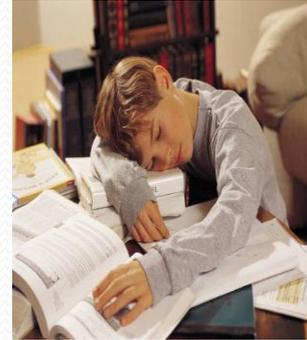
Misleading with Averages

- Lets say you are a class teacher and you are trying to convince students to attend/not attend your class because of mathematics proficiency
- You could, with perfect honesty and “truthfulness”, tell students that the average score in your class is:
 - 50.5%
 - 30%
 - 10%

How? Is This Even Possible?

- The 50.5% figure is the arithmetic mean of all student scores in mathematics
- The 30% figure is the median
- The 10% figure is the mode
- All three: mean, median, and mode are measures of central tendency

How is this Possible?



One Math whiz who scores 100%

Five amazing Math students score 95%

14 math kids score 90%

One sleeper who scores 30%

20 math kids score 10%

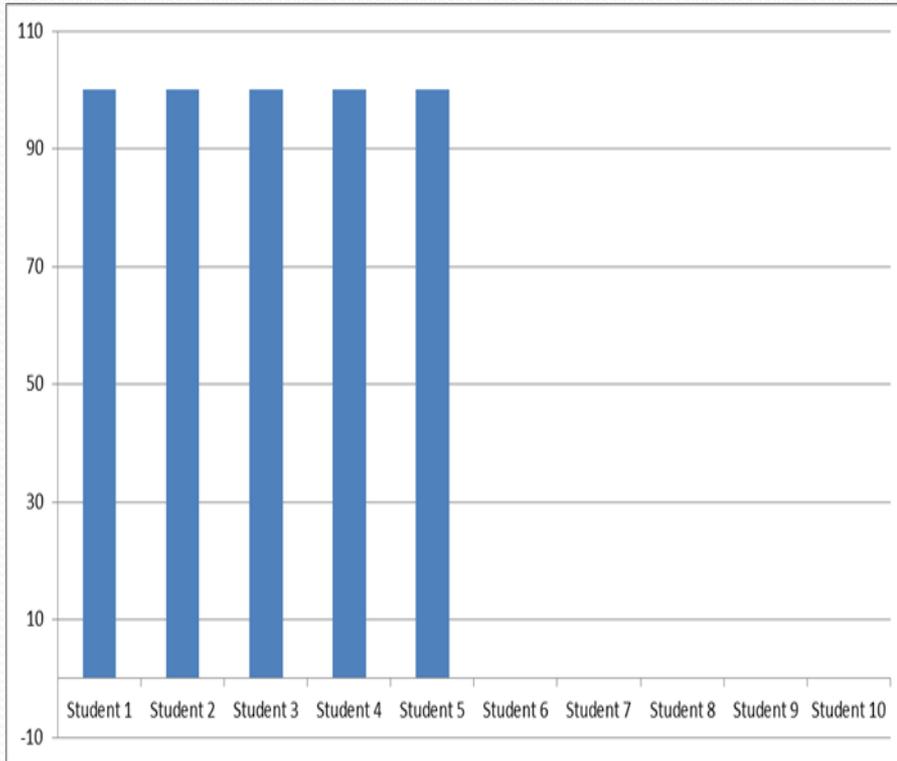


$$= \{(1 \cdot 100) + (5 \cdot 95) + (14 \cdot 90) + (1 \cdot 30) + (20 \cdot 10)\} / 41 = 50.5\%$$

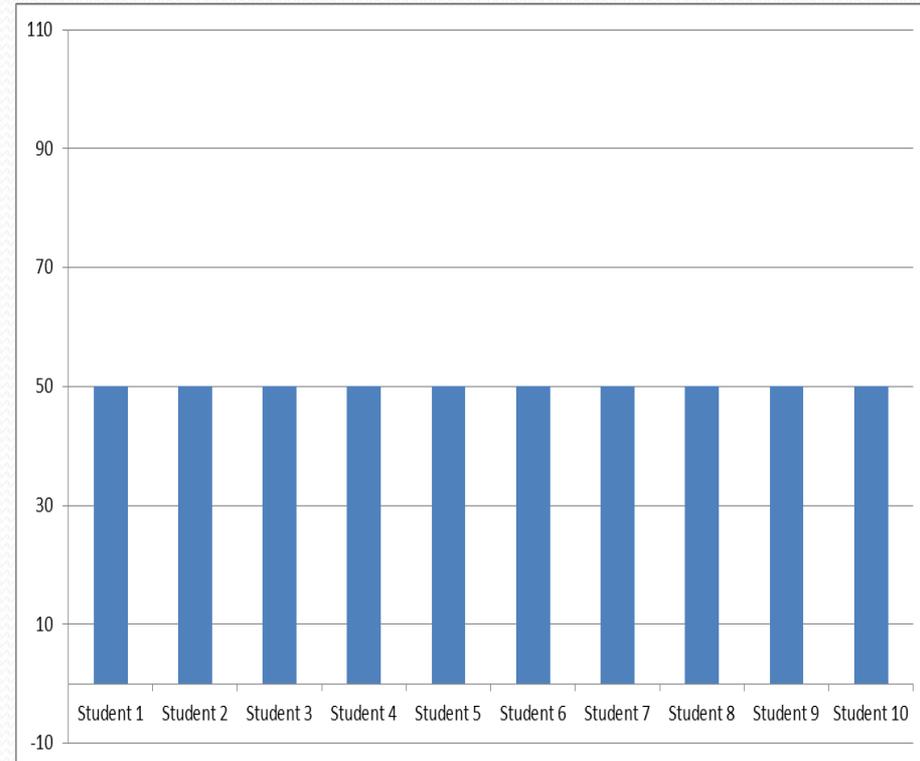
Means Without a Range or Standard Deviation

- Technical Skills Assessment Score = 50%. Is this 50% always created equal?
- A range, standard deviation, variance etc. are measures of dispersion

Means Without a Range or a Standard Deviation



Range = 0 to 100, Stdev = 50



Range = 0, Stdev = 0

Percentages Without Sample Size

- Do you believe the statement that a class is extremely diverse if I said:
 - “It’s composition is 25% African-American students, 25% Asian students, 25% Hispanic students, and 25% White students.”

Significance Testing

- e.g. We found that a relationship exists between X and Y at the $P=0.05$ level.
- Significant does not mean important
- Insignificant does not mean unimportant
 - The Cult of Statistical Significance
 - Wall Street Journal article

Graphical Misplays of Information

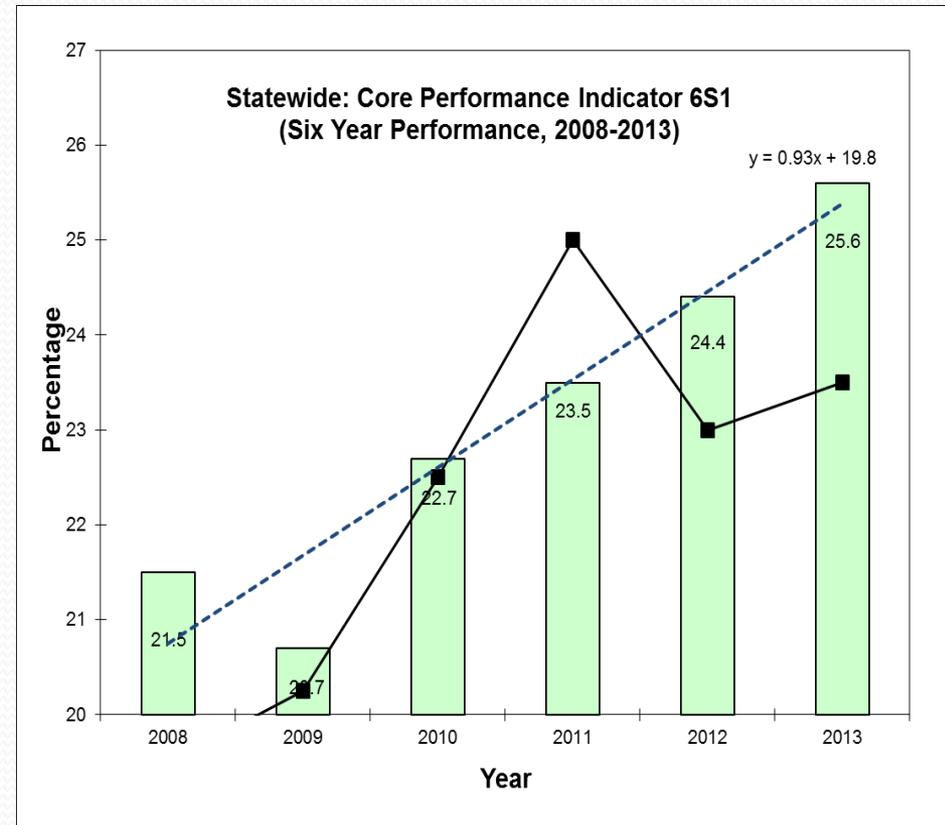
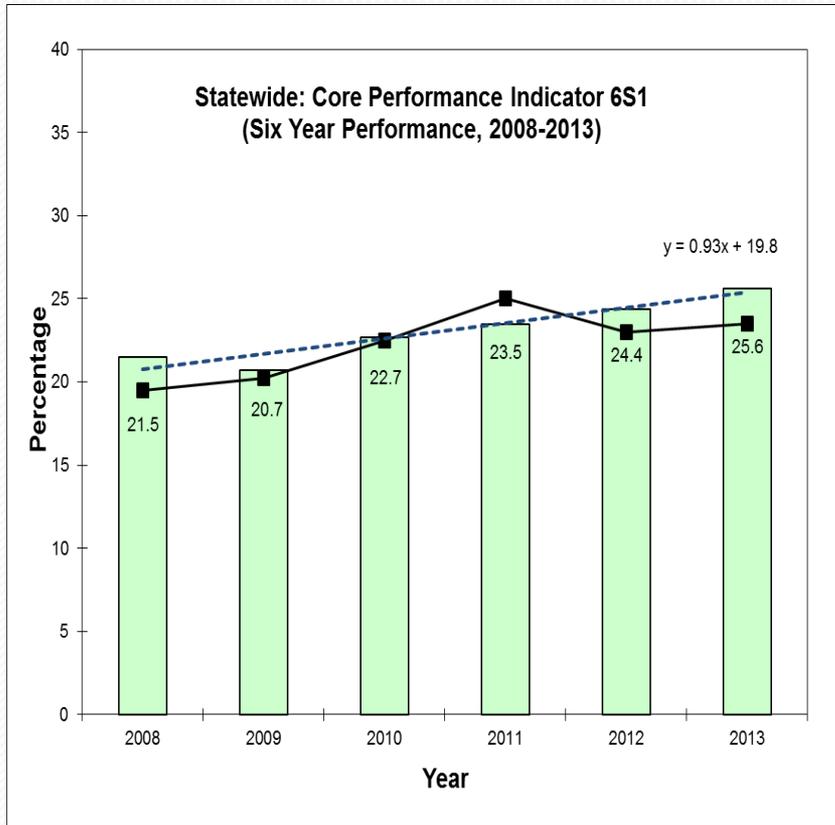
- Graphical misplays, a good friend of



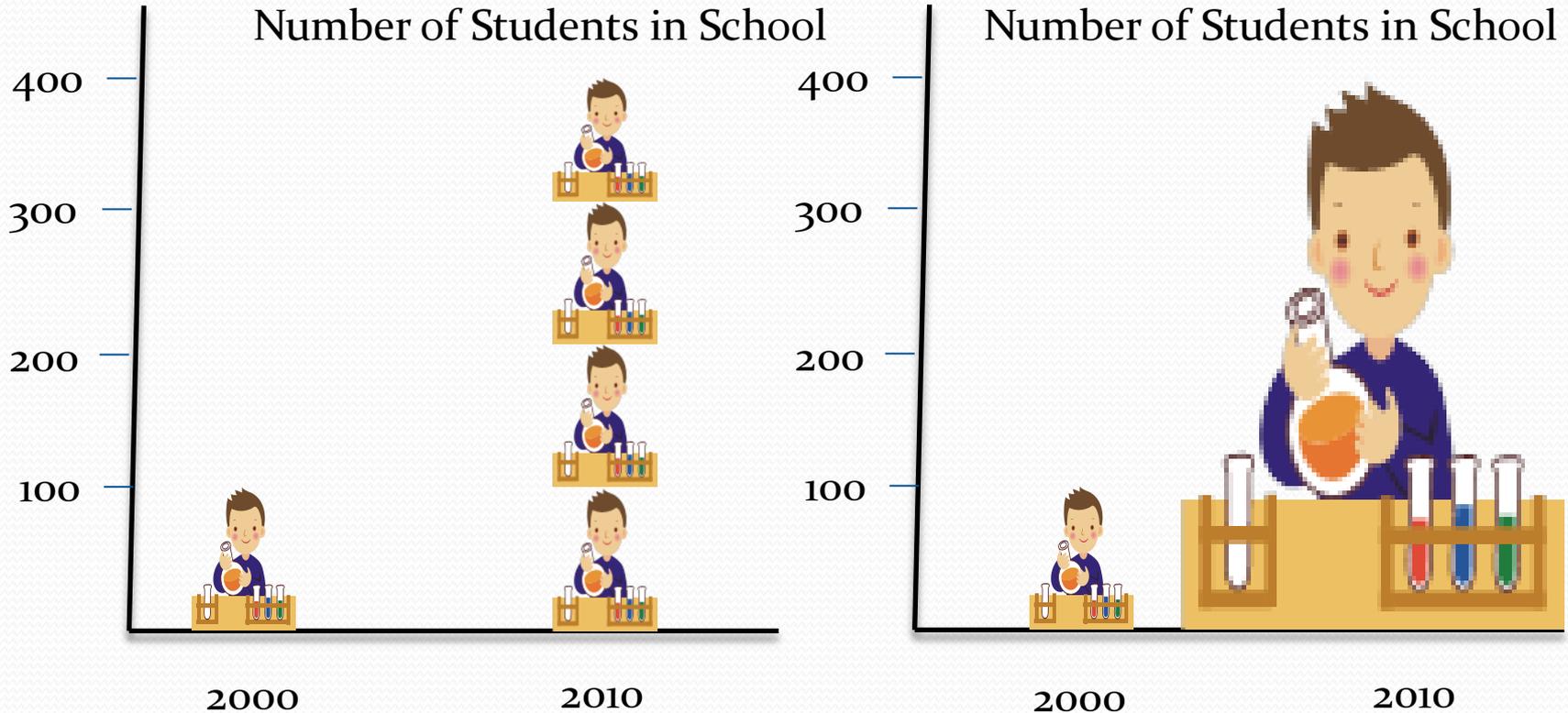
Graphical Excellence

- The principals of Graphical Excellence (GE) are:
 - GE is the well-designed presentation of interesting data – a matter of **substance**, of **statistics**, and of **design**
 - GE consists of complex ideas communicated with **clarity**, **precision**, and **efficiency**
 - GE is that which gives to the viewer the greatest **number of ideas** in the **shortest time** with the **least ink** in the **smallest space**.
- To lie, cheat, manipulate, or mislead, do NOT follow this advice!!!

6S1 Rate for Michigan: Trends



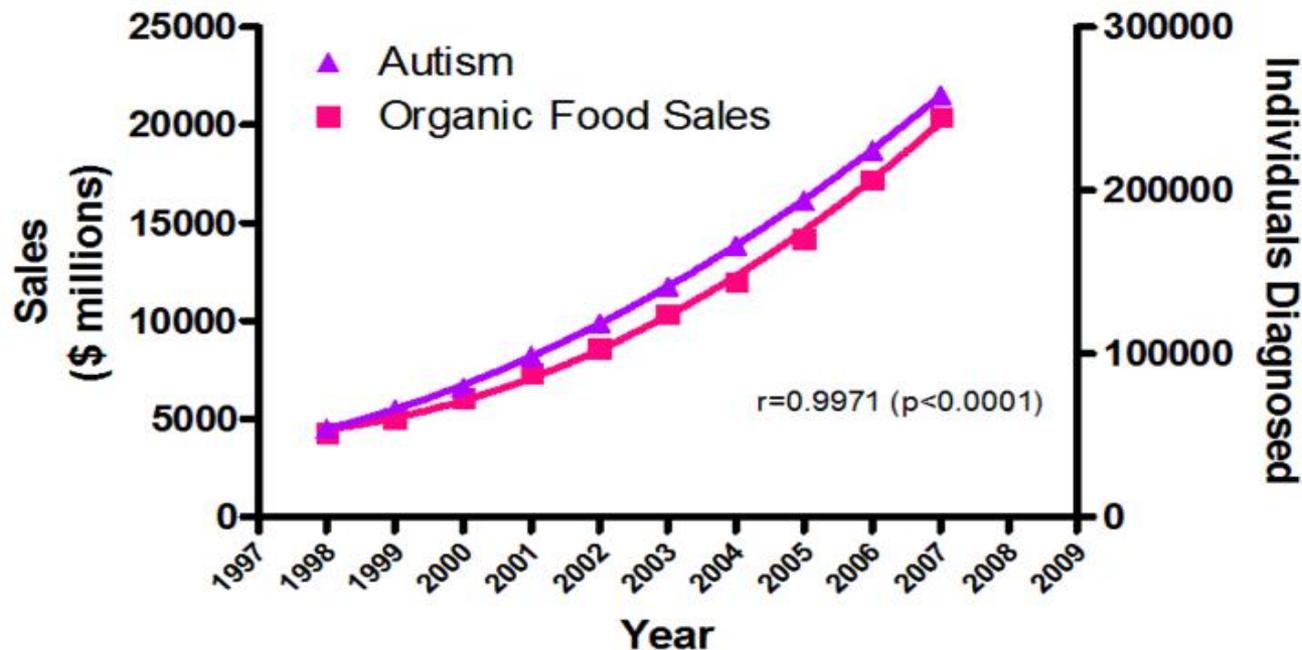
Using Volume Instead of Bars



If we look at the chart heights they show the same information. Volume is one way to mislead

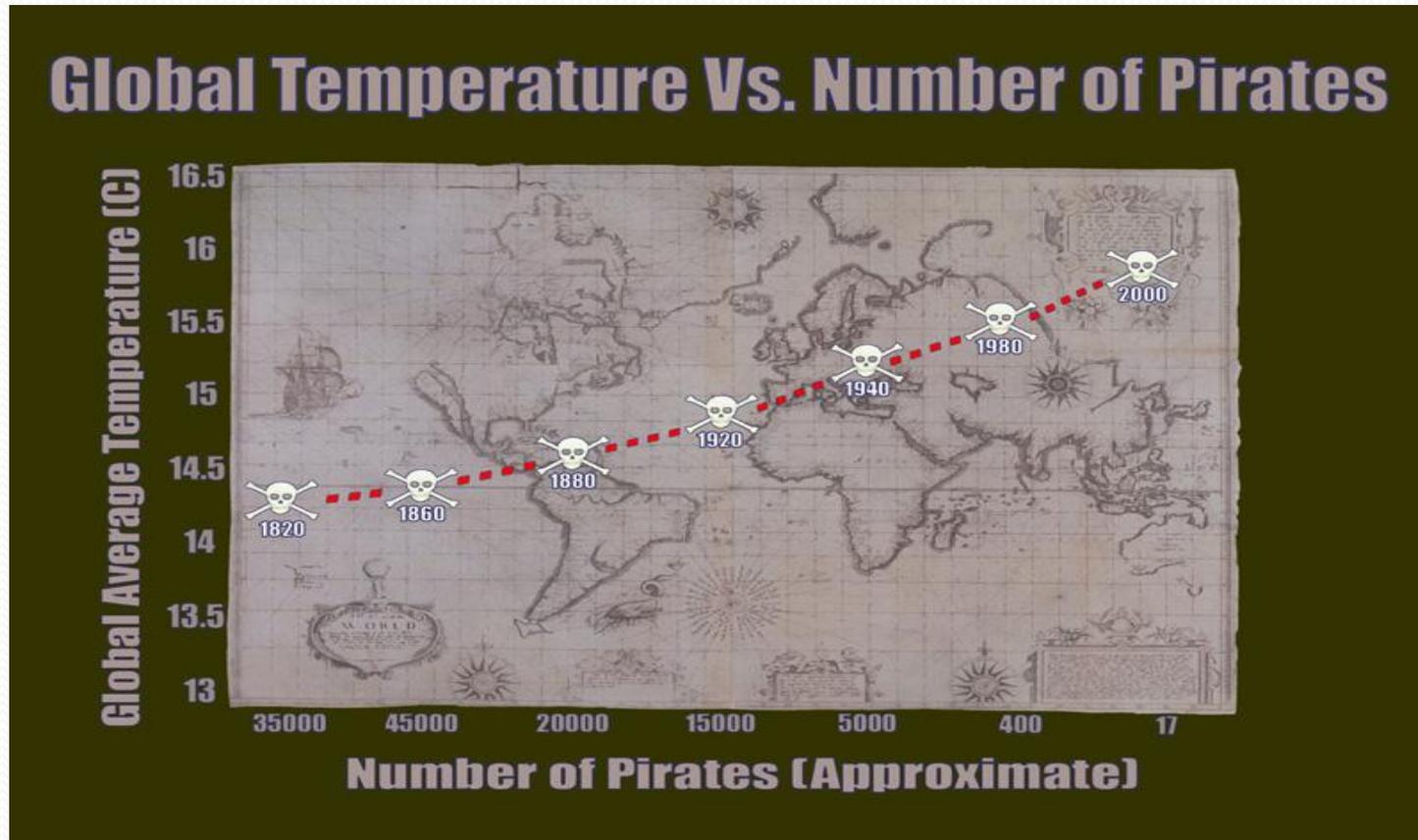
Correlation Does Not Mean Causation

The real cause of increasing autism prevalence?



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act"

Correlation Does Not Mean Causation



Correlation Does Not Mean Causation



TPF_109_Correlations_111009.m4v

- Luckiest people 'born in summer'
(<http://news.bbc.co.uk/2/hi/health/3622817.stm>)
- Eating pizza 'cuts cancer risk'
(<http://news.bbc.co.uk/2/hi/health/3086013.stm>)

Take Away From Today: How to Tell Between Good and Bad Information?

- Look at the data sources. If none are given, do NOT trust the information
- Check to see if there is bias in the data. How and where was it collected?
- Examine closely at the data axes on graphs. Is it following principles of [Graphical Excellence](#)?
- Lastly, do NOT believe everything you are shown just because it is “Science” and “Data”. Try to figure out if the source has some ulterior motive to manipulate your opinion

THANK YOU

- Questions??
- Krishnan Sudharsan
 - sudharsank@michigan.gov
 - 517-241-7652