Just Plain Data Analysis:
Common Statistical Fallacies in Analyses of Social Indicator Data

Gary Klass
Department of Politics and Government
Illinois State University
gmklass@ilstu.edu

Abstract
This paper presents a short summary of the most common statistical fallacies found in public debates employing social indicator data as the evidentiary premises of arguments about politics and public affairs. The purpose is to offer students a convenient framework for evaluating, and developing their own, arguments relying on social indicator data.

1. Statistical Fallacies in Social Indicator Arguments

Most data-based public affairs writing involves informal logic and argumentation. Since Aristotle, the generally accepted method of evaluating informal arguments has been to analyze them for common forms of fallacious reasoning. "Statistical" fallacies occur when an argument's conclusions are not supported by the numerical evidence provided as premises.

Several of the statistical fallacies commonly found in analyses of social indicator data analysis are forms of the "threats to internal and external validity" identified in the work of Donald T. Campbell and Julian Stanley (1966) on quasi-experimental design. Several of Campbell's threats, such as sample mortality, testing, and Hawthorne effects, are not common in social indicator arguments (in part because such data do not involve repeated measurements of the same sample) and other statistical fallacies are peculiar to the interpretation of social indicator measurements.

Statistical fallacies fall roughly into two categories: those related primarily to the measurement and the collection and construction of the social indicator being used and those related to causation. One statistical fallacy, perhaps the most common in public debate using social policy indicators, can involve data misinterpretation related to both measurement and causation: Cherry Picking.

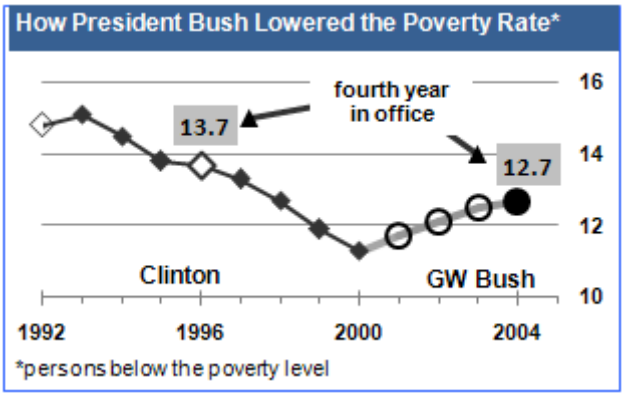

Figure 1: Cherry Picking
http://lilt.ilstu.edu/jpda/asa/1.xlsx

Cherry Picking. Biased selection of social indicator data to support preconceived ideas is probably the most common statistical fallacy in public debate, a phenomenon fostered by the increasing availability of alternative social indicator measurements and annual time series data that permits the researcher to choose beginning and ending points for comparison.

One brazen example of cherry picking was talk show host Bill O'Reilly's attempt to argue that the Bush administration deserved credit for lowering the poverty rate. "The only fair comparison," of poverty rates, O'Reilly insisted, "is halfway through Clinton's term, halfway through Bush's term"(Media Matters, 2005), O'Reilly was correct, sort of, but failed to see that a measure of the change in the poverty rate over the first four years of the presidents' terms would tell an entirely different story.

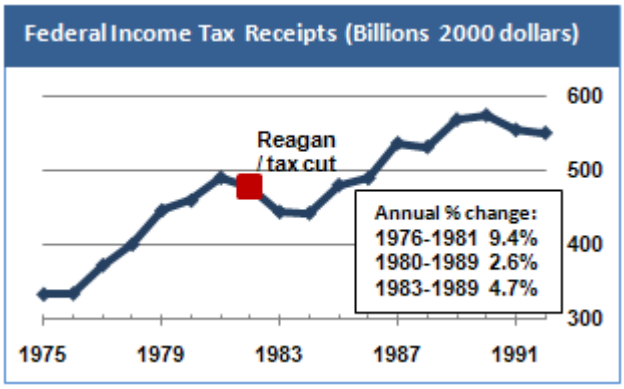

Figure 2: Cherry Picking
http://lilt.ilstu.edu/jpda/asa/2.xlsx

Advocates of the "Laffer Curve" –the idea that cutting taxes will increase government revenue –often cite the beneficial impact of the Reagan (and also Kennedy) administration tax cuts that were partially implemented in the 1982 fiscal year and fully implemented in 1983. Heritage Foundation economist Daniel Mitchell (2003) argues the point: "Once the economy received an unambiguous tax cut in January 1983, income tax revenues climbed dramatically, increasing by more than …28 percent after adjusting for inflation."

Mitchell's 28 percent is the change from 1983 to 1989: Instead of a before-and-after measurement, he has done an "after-and- long-after measurement" – ignoring the two years of income tax revenue reductions that took place while the tax cuts were at least partially in effect – and the higher rates of revenue growth that took place in the Carter years.

1.1 Measurement Fallacies.

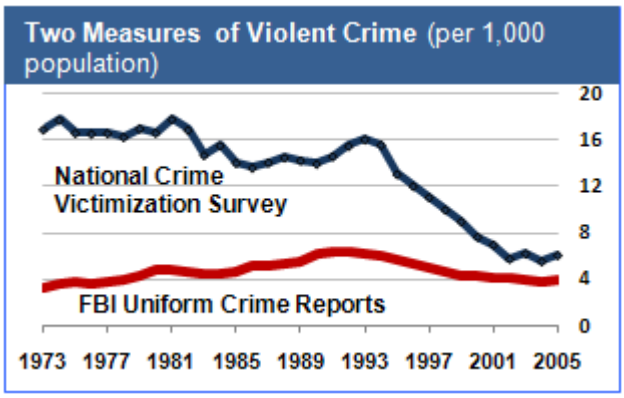

Figure 3: Instrumentation
http://lilt.ilstu.edu/jpda/asa/3.xlsx

Instrumentation. Instrumentation error suggests that the observed numerical comparison may be due to measurement unreliability.

Throughout the 1970s and 80s, the FBI measure of violent crime, based on police reports of violent crime, increased, while the National Crime Victimization Survey, based on an annual survey of personal crime victimization, indicated that the rate was falling. The FBI measure is generally regarded as less reliable, however, as reporting of crimes to police has increased and police departments have improved their record keeping. Oddly, it is because the FBI data have become more reliable over time that time series analysis of the data has become suspect.

Instrumentation can also affect conclusions drawn from cross sectional numerical comparisons. The high U.S. infant mortality rates, often cited as a product of the lack of universal health insurance, may also be at least partly due to the way the U.S. counts live births.

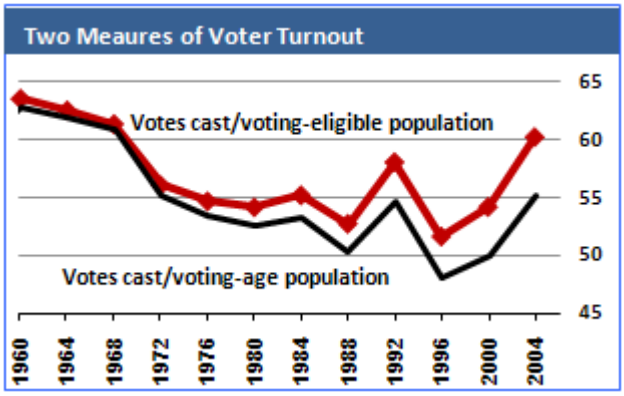

Figure 4: Measurement Validity
http://lilt.ilstu.edu/jpda/asa/4.xlsx

Measurement Validity –how well a measurement measures the concept of interest –falls under Campbell's threats to external validity and is a form of the logical fallacy of hasty generalization.

When voter turnout, traditionally measured as the percentage of the voting age population that votes, fell below 50 percent in the 1996 election, political commentators blamed turned-off voters, partisan politics (the only kind of politics), negative campaigning and the rise of conservative talk radio.

In 2001, political scientist Michael McDonald compiled new data suggesting that the talk about the vanishing American voter was "a myth" (2001, 963). McDonald's analysis called attention to the denominator in the voting turnout statistic: voting age population, and argued that we should instead use the voting-eligible population.

Over recent elections, an increasing percentage of the American voting age population has not been eligible to vote. Mostly this is because of increasing immigration. In addition, in all but two states, prisoners are not allowed to vote and in 12 states even ex-felons are disenfranchised. Because the percentage of the American population that either is incarcerated or has ex-felon status has gone up dramatically since the 1980s, an increasing percentage of the voting age population cannot vote. Taking the votes cast as a percentage of the voting eligible population (the "VC/VEP" trend in figure 4.6) as our measure of turnout, we see no general decline in voter turnout since 1972, when 18 years olds were given the franchise.

So which is the better measure of voter turnout? If you look at the voter turnout as a measure of how democratic a society is, the traditional voting-age numbers have greater validity. Although voting-eligible turnout is increasing and at a long time high, this is true because so many young black males (unlikely voters to begin with) have been put in jail and, in many states, denied the right to vote for the rest of their lives and because so many of our nation's poor are not citizens. If all young voters (the age group least likely to vote) were incarcerated and all the poor (the economic group least likely to vote) were declared non-citizens, the American voter turnout rate would be among the highest in the world, but the United States would not be a more democratic society.

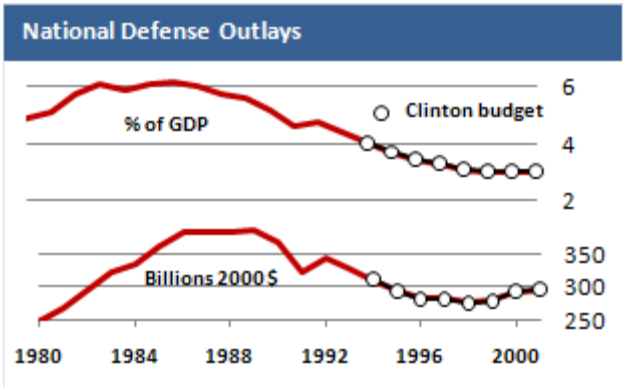

Figure 5: Dominant Denominator
http://lilt.ilstu.edu/jpda/asa/5.xlsx

The Dominant Denominator. Most social indicators are ratio measures – Infant deaths as a percent of live births or traffic fatalities per million miles travelled – and occasionally analysts can be deceived by a change in an indicator that is more a function of a change in the denominator than in the numerator. This can be a particular problem in the case of divisors, such as GDP, that are more volatile than the indicator's numerator.

In 2004, a Wall Street Journal editorial criticized the Clinton administration for cuts in the defense budget: "Bill Clinton and a GOP Congress balanced the budget by withdrawing a 'peace dividend' at a time when al Qaeda was declaring war" (2004). Their evidence was a chart showing the declines in defense spending as a percent of GDP. From fiscal year 1993, the budget year before Clinton took office, to 2001 (the year before the 9/11 increases) defense spending fell from 4.4% of GDP to 3 %, a dramatic and steady decline. Much of the decline, however, was due less to cuts in military spending and more to the dramatic economic growth and the increase in GDP in the Clinton years. In real dollars (adjust for the GDP price deflator), military spending actually increased in Clinton's second term, a dramatic turnaround from the post Cold War decline that began in the first Bush administration.

A general inattention to denominators, other than the basic per capita measures, is the cause for much statistical misinformation. The most commonly used adjustment for inflation, the consumer price index, overestimates inflation by an estimated one percent per year (Boskin, 1996). As a result many monetary measures –including income, spending, and prices –are actually growing faster than the inflation adjusted measure would indicate. Because the CPI results in an artificially higher poverty threshold, the overestimates has the opposite effect on poverty, making it appear that poverty is higher over time.

1.2 Causal Fallacies.

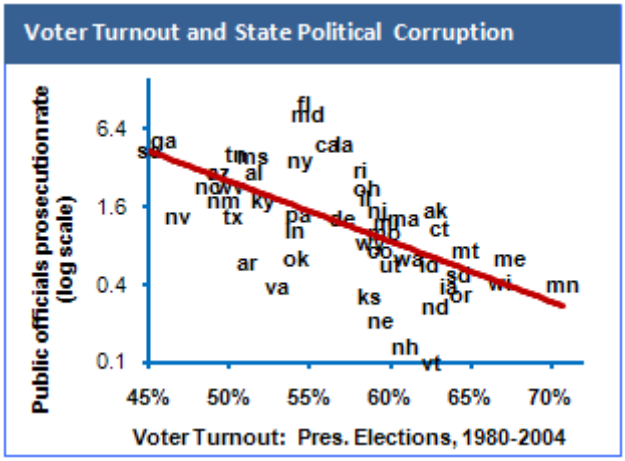

Figure 6: Reverse Causation
http://lilt.ilstu.edu/jpda/asa/6.xlsx

Reverse Causation (*Post hoc ergo propter hoc*)**.** A relationship between two variables in and of itself does not reveal which might be the cause of the other, even if the variables are measured at different points in time. Roosters do not cause the sun to rise.

In the U.S., states with the highest rates of political corruption tend have low rates of voter turnout (note: check the validity of the measure of political corruption). Does the low voter turnout foster greater corruption or vice versa?

Similarly, countries with corrupt regimes tend to have higher poverty rates, but there is considerable debate over what causes what.

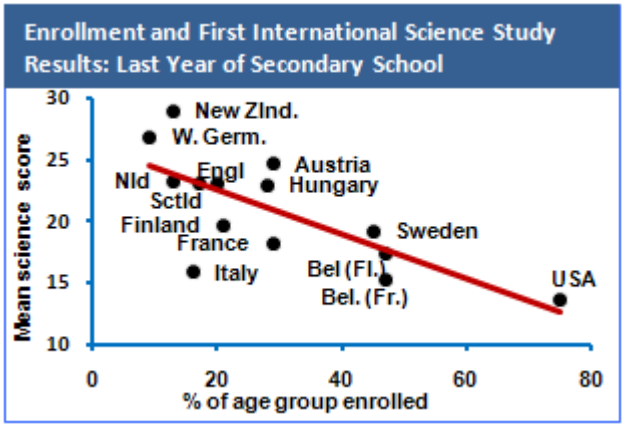

Figure 7: Population Mortality
http://lilt.ilstu.edu/jpda/asa/7%20.xlsx

Population mortality. Although sample mortality is usually not a concern with social indicator measurements, because they do not involve repeated measurements of the same sample, changes and differences in the inclusiveness of the populations surveyed can have significant effects on data comparisons. This is especially true in the case of measures of educational achievement where, sometimes deliberately, the students who are least likely to do well on the tests are often excluded from the testing.

In 1983, the "A Nation at Risk" report began with disturbing evidence of the weak performance of American students on international academic achievement tests (National Commission). Among the evidence cited, American high school seniors recorded the lowest grades on the First International Science Study, but at least part of the reason for the low U.S. scores had to do with the relatively high rates of U.S. students completing high school (Medrich and Griffith,1992).

In some situations there may also be the reverse population mortality effect. Although black 4th grade and 8th grade reading scores have improved in recent years, 12th grade scores for black students have not. Part of the reason the 12th grade scores have not gone up, is due to the decline in the black high school dropout rate (Klass, 2008, 108).

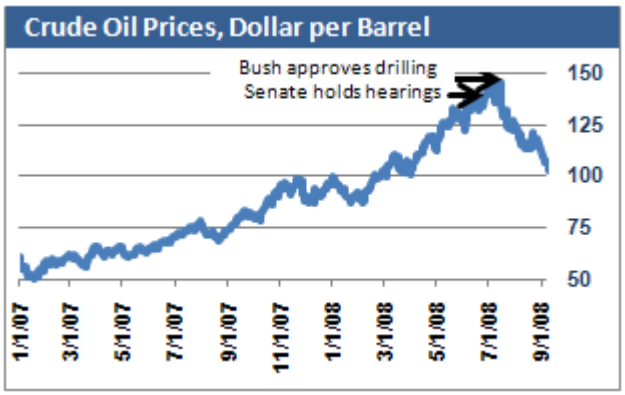

Figure 8: Regression Artifact
http://lilt.ilstu.edu/jpda/asa/8.xlsx

Regression Artifact. Measuring before-and-after change from a base year with unusually low or high values risks a regression artifacts fallacy as the indicator "naturally" regresses to the mean. When oil prices topped four dollars a gallon in the Summer of 2008, Congress was quick to identify the culprit: oil price speculators who had bid up the price of oil to nearly $140 a barrel (the price does not even include the cost of the barrel). The Bush administration insisted that restrictions on off-shore oil drilling were to blame. Congress held hearings. President Bush rescinded the executive order banning off-shore drilling (although the law that also banned the drilling remained in effect).

When the bubble broke, both sides took credit for the lower prices. Liberal bloggers claimed that the hearings

scared away the speculators (Johnston, 8/1/8); John McCain argued that Bush's executive order was responsible (Associated Press, 7/23/08).

Regression artifacts are a common explanation for the apparent effectiveness of the most dubious medical treatments: people feel better after treatment because the underlying condition is one that typically gets better by itself. In a sense regression artifacts such as these are a form of reverse causation: it is the sharp spike in the indicator that causes the policy action.

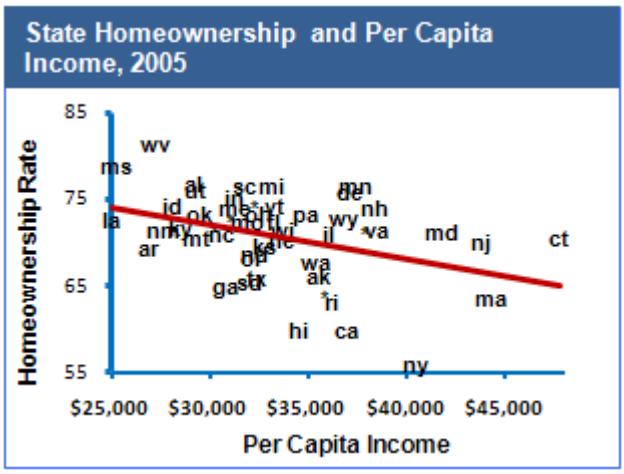

Figure 9: Ecological Fallacy
http://lilt.ilstu.edu/jpda/asa/9.xlsx

Ecological Fallacy. The ecological fallacy occurs when drawing a conclusion about individuals from aggregate data. It is related to the logical fallacy of division. States with higher per capita incomes generally have lower rates of homeownership. The fallacy here would be to conclude that wealthier individuals are less likely to own homes.

A recent study (Galbraith and Travis Hale, 2006) finds that the Democratic Party receives a higher share of the presidential vote in wealthier states and states with greater income inequality. It would be an ecological fallacy to conclude from the first finding that wealthier voters vote Democratic or from the second finding that the rich and poor voters voter Democratic.

Sociologist William S. Robinson coined the term in a 1950 article in which he observed that states with the highest rates of foreign born population also had the highest literacy rates, even though the foreign born had lower literacy rates than the native born population.

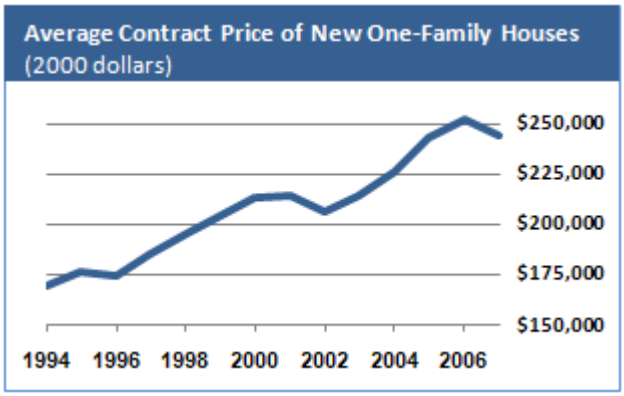

Figure 10: Trending Upward
http://lilt.ilstu.edu/jpda/asa/10.xlsx

The Trend is not Your Friend. Wise investors know this but when things are looking up it is easy to forget.

It seemed to be a good time to buy a home in 2006 (and in 2000). Adjustable no-interest, no down payment, loans were cheap and with rising home prices the prospects of getting another loan before the rate-adjustment kicked in were good. Some, but not enough, of those who profited off the expectation that the trend would continue upward are now on their way to jail.

Unfortunately, discerning which trends are merely speculative bubbles soon to be corrected by market forces, and which represent inexorable forces is no easy task. The upward trend in university tuition in our next example is of the inexorable variety.

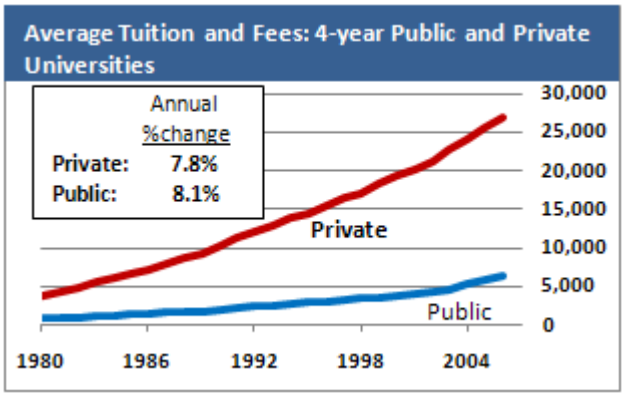

Figure 11: Graphical Distortion
http://lilt.ilstu.edu/jpda/asa/11.xlsx

Graphical Distortion. Poor graphical construction can distort statistical relationships.

A quick glance at the chart on the left would lead one to conclude that private university tuition and fees, which rose from a little over $3,800 in 1980 to almost $27,000 in 2006, is increasing at a much faster rate than public university charges (rising from $840 to $6,400). The chart illustrates a common graphical scaling distortion, similar in some respects to the rate of change fallacy. In fact, public universities are increasing their tuition at a slightly faster rate and both at rates are going up as faster than health care costs or gasoline prices (Gasoline hit $1 a gallon in 1980; had it increased at the same rate as tuition it would now be over $6 a gallon).

Although much of the writing on graphic design focuses on distorted presentations of data, more often, poor graphical display and poor tabular construction work to hide critical relationships.

2005-6 Chicago and Illinois ISAT Results: Grades 3-8

	2005	2006	Net Change	% Change
Passing rate				
Chicago Public Schools	48%	62%	+14	+29%
All Illinois Public Schools	69	77	+8	+12
Failure rate				
Chicago Public Schools	62	48	-14	-27
All Illinois Public Schools	31	23	-8	-26

Figure 12: Rate of Change Fallacy
http://lilt.ilstu.edu/jpda/asa/12.xlsx

Rate of Change Fallacy. The rate of change fallacy occurs when comparing rates of change (usually rates of changes in rates) in two numbers that start out at different levels. The misinterpretation is so common in the interpretation of educational statistics [often involving conclusions that disadvantaged kids are improving at a fast rate] that one educational researcher, Stephen Gorard (1999), has given it a name: the "politician's error".

In 2006, Chicago public school officials trumpeted the apparent gain in student scores on the annual ISAT tests mandated by the No Child Left Behind law. The state pass rate on the exams had increased 8 points –most probably because of revisions to the test and changes passing score for the 8th grade math tests –while the Chicago pass rate increased 14 percentage points.

But is an increase from a 48 to a 62% pass rate a bigger change than an increase from 69 to 77%? Consider a more extreme example: would a student who increased his test score from 40 to 64 be improving at a faster rate than one who increased his score from 91 to 99?

One way of testing for the rate of change fallacy is to take the inverse of the data: using the failure rate rather than the pass rate. In the case of the Illinois data, both state and the city have seen similar percentage declines in their students' rate of failure.

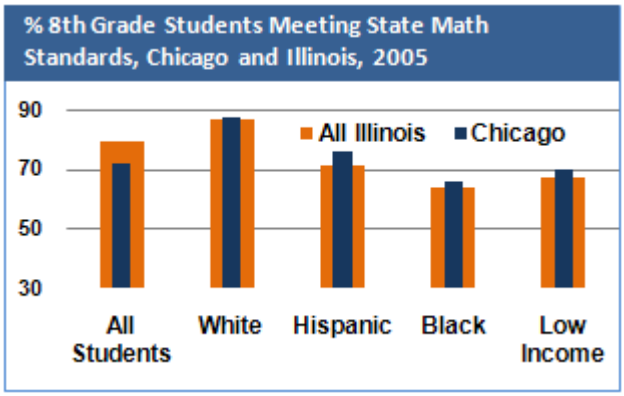

Figure 13: Simpson's Paradox
http://lilt.ilstu.edu/jpda/asa/13.xlsx

Other confounding effects That some third variable might account for an observed relationship is the reason we have social scientists. Many of the statistical fallacies described above (instrumentation, population mortality, regression artifact, and measurement validity) are actually forms of a general spurious effect. What Donald Campbell calls the "history" threat to internal validity and the phenomenon of a "Simpson's Paradox" are two general categories of other confounding effects

In the case of the relationship at left, the city of Chicago has a lower pass rate on the 8th grade math test than the state as a whole (note: Chicago is included in the state totals), but for every demographic group of students, Chicago outperforms the rest of the state.

2. Constructing Good Arguments:

In the case of all argument fallacies, it is generally the reasoning, not the premises, that is at fault. Bill O'Reilly's data were, as he often says, "a fact". Wealthier states do in fact have higher homeownership rates.

Identifying and avoiding statistical fallacies requires subject matter expertise more than trainable statistical skills. It requires a good understanding of probable alternative explanations for numerical findings and a good understanding of the social indicators employed. Familiarity with the indicators requires an understanding of how the data have been collected, how the indicator is constructed, and how readily available alternative measures might produce different results.

In most instances, the key to indentifying statistical fallacies is more data. An extended time series will often expose cherry picking and regression artifacts. Alternative measures can expose instrumentation and measurement reliability problems. In most cases, calculating an inverse of the base statistic will expose the rate of change fallacy.

The best analyses of social indicator data rely on a variety of different measurements and controls for spurious relationships both to avoid the risk that any single piece of numerical evidence might prove fallacious, but also to illustrate the arguments in more detail.

In *Bowling Alone* (2000), Robert Putnam employs a wide range of measures of civic participation, social and family life and political attitudes –including the declining participation in league bowling –to make his case that American civic engagement is in decline. Putnam describes his strategy as attempting to "triangulate among as many independent sources of information as possible" based on the "core principle" that "no single source of data is flawless, but the more numerous and diverse the sources, the less likely that they could all be influenced by the same flaw" (415).

Kevin Phillips, *The Politics of Rich and Poor* (1990) and Paul Krugman's *Conscience of a Liberal* (2008) make essentially the argument: that, since the Reagan Administration the United States has entered a new "Gilded Age," the rich are getting richer and the poor and middle class haven't been doing so well. Both cite an extensive array similar evidence to support their conclusion, such as changes in shares of income earned by the top and bottom income brackets, the ratio of corporate executive to worker earnings

Merely using multiple –even wildly different –measures, however, does not preclude flaws that might affect them all. On a whole host of economic and social indicators the Clinton administration has outperformed the G.W. Bush administration: poverty, stock prices, gas prices, budget deficits, health care spending, and crime. To conclude that Clinton and Bush administration policies were responsible for the measures risks several fallacies. Clinton came into office at the tail end of a recession and some indicator growth may have been a regression artifact, and much of the economic growth may have been due to the dot.com stock market bubble. Bush came into office at the very beginning of an economic downturn and the economy suffered greatly from the 9/11 attacks.

References

Associated Press 2008. "McCain Credits Bush for Oil Price Drop," http://elections.foxnews.com/2008/07/23/mccain-credits-bush-for-oil-price-drop/

Boskin, Michael, et al., 1996. ''Toward a More Accurate Measure of the Cost of Living," report of the Advisory Commission to Study the Consumer Price Index to Senate Finance Committee (December 4).

Campbell, Donald. T. and Julian C. Stanley. C. 1966. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally, 1966.

Duffett, Ann, Steve Farkas, Tom Loveless. 2008. High-Achieving Students in the Era of No Child Left Behind Thomas B. Fordham Institute (June 18). http://www.edexcellence.net/detail/news.cfm?news_id=732&id=92

Bureau of Labor Standards. 2007. "Chapter 17, The Consumer Price Index," *BLS Handbook of Methods*, June 26.

Galbraith, James K. and Travis Hale. 2006, "State Income Inequality and Presidential Election Turnout and Outcomes," The University of Texas Inequality Project Working Paper 33 (March 3) http://utip.gov.utexas.edu/papers/utip_33.pdf

Gorard, Stephen 1999. "Keeping a Sense of Proportion: The 'Politician's Error' In Analysing School Outcomes," *British Journal of Educational Studies*, 47, no. 3, (September): 235–46.

Johnston, Donnie 2008. "Our greed is fuelling gas price increases," The Free Lance – Star 8/1/09. http://fredericksburg.com/News/FLS/2008/082008/08012008/399406/index_html

Klass, Gary M. 2008. Just Plain Data Analysis (Rowman & Littlefield).

McDonald, Michael P. and Samuel Popkin. 2001. "The Myth of the Vanishing Voter." *American Political Science Review* 95(4): 963-974, LINK.

Media Matters for America, 2005. "O'Reilly Defended False Clinton-Bush Poverty Comparison as 'the Only Accurate Measuring Stick'" (September 16) http://mediamatters.org/items/200509160002

Medrich, Elliott A. and Jeanne Griffith, 1992. International Mathematics and Science Assessment: What Have We Learned? Report No. CNES 92-011 (Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement) http://nces.ed.gov/pubs92/92011.pdf

Mitchell, Daniel J. 2003. "The Historical Lessons of Lower Tax Rates," The Heritage Foundation <http://www.heritage.org/Research/Taxes/wm327.cfm> (August 13)

National Commission on Excellence in Education. 1983. *A Nation at Risk:The Imperative for Educational Reform* (April) http://www.ed.gov/pubs/NatAtRisk/index.html

Putnam, Robert D. 2000. *Bowling Alone*. New York: Simon and Schuster.

Robinson, William S. 1950. "Ecological Correlations and the Behavior of Individuals". *American Sociological Review 15:* (June): 351–357.

Tufte, Edward. 1983. *The Visual Display of Quantitative Information*. Cheshire: Connecticut: Graphics Press.

_____, 2006. *Beautiful Evidence* Cheshire: Connecticut: Graphics Press.

Von Kanel, Joe 2008. "McCain: Give Bush credit for oil price drop" CNN Politics.com Political Ticker (July 23), http://politicalticker.blogs.cnn.com/2008/07/23/mccain-give-bush-credit-for-oil-price-drop/

Wainer, Howard, 2008. *Picturing the Uncertain World: How to Understand, Communicate, and Control Uncertainty through Graphical Display.* Princeton, N.J.: Princeton University Press. (forthcoming).

Wall Street Journal. 2004. "Fiscal Follies: Clinton balanced the budget by cutting the military. That's not an option now" editorial (August 9) http://www.opinionjournal.com/editorial/feature.html?id=110005456