I used to think correlation implied causation.

Then I took a statistics class. Now I don't.

Sounds like the class helped.

Well, maybe.
1. Credible empirical studies are key inputs to making sound policy decisions

Examples:

- What is the impact of secure property rights on natural resource outcomes (i.e., effect of ITQ on fish stocks, harvest decisions, and their economic value)?

- What will be the impact of climate change on human health?

- Are environmental regulations that improve ambient air quality cost-effective?
The field of analysis of empirical policy evaluation methods goes by many labels:

- “Program evaluation”
- “Policy evaluation”
- “Impact evaluation”
- “Causal inference” (of course the inference could be about a question without a direct a policy application)
Motivation (ctd)

2. Designing and implementing a credible empirical study is often difficult ... The context matters a lot

In an purely ecological setting, simple correlation between outcome and “treatment” (predictor of interest) may reveal a causal relationship

- Example: Measuring the effect of temperature on mockingbird song frequency and dB

When humans ‘interfere’ with natural systems, the analysis is typically not as simple

- Example: ITQ regimes are not randomly implemented across fisheries and may be negatively selected on biological outcomes
- Humans exposed to high levels of ambient pollution may be positively selected on SES (e.g., Env Justice literature)
Motivation:

3. Key challenge is that most study designs in social sciences (and natural sciences) are non-experimental:

- “Treatment” status not randomly assigned to subjects
- Subjects often select their treatment status
- Opens up the possibility of biases due confounding, omitted variables, selection, simultaneity causality

⇒ Correlation ≠ causation
- i.e. observing a correlation between two variables does not imply there is a causal relationship between the variables
Note: correlation = 0.96

Source:
http://www.tylervigen.com/
Note: correlation = 0.95

Source:
http://www.tylervigen.com/
What is Causality?

- Causality is complex concept, even if it appears simple and intuitive (ask a philosopher...)

- We will use the counterfactual or potential outcome framework to define causal relationship

- Other frameworks exist, but most policy evaluation techniques have been developed with the potential outcomes framework, so I use it here
Objectives of this class:

1. Provide a better understanding of issues that lead to the correlation ≠ causation problem
   - Of course this also highlights when simple correlations (or naïve regression coefficients) derived from observational data can have a “causal” interpretation

2. Present assumptions and statistical methods that can help identifying causal relationships with non-experimental data
   - Note the underlined above: these methods are never silver bullets that guarantee identification of causal relationships
   - Validity of any particular application is always dependent on context and subject to maintained assumptions
   - How believable are the assumptions?
Outcome: dependent variable of interest in the study
- In the ITQ and fisheries example, the outcomes could be: fish stocks, fish harvests, economic value of harvest
- In the air pollution example, the outcome could be: measures of health, housing values, etc

Treatment: independent variable of interest in the study
- ITQ status = \{0, 1\}
- Ambient air pollution concentrations

Treatment assignment (mechanism): Process by which subject become exposed (or not) to the treatment
- In RCT: random process (i.e. coin flip)
- ITQ: Not sure...
- Air Pollution: Sorting process in which individuals tradeoff housing values and ambient air pollution level
As we will see, the key will be to understand the mechanism by which treatment status was assigned to the subjects we are studying:
- Randomly
- Quasi-randomly / exogenously?
- Selected based on observed characteristics?
- Selected based on unobserved characteristics?

There are assumptions & methods to identify causal relationships for all types of treatment assignment mechanisms:
- Again, the context, and the credibility of the maintained assumptions will distinguish between good and not so good empirical papers.
Class Administrative Details

- Class meets twice a week (Mon/Wed, 12:30-1:45)
- Office hours: Fridays 1:30-3:00, or by appointment (email), North Hall 2050
- Class website: (lecture notes, assignments, data, answer keys, readings, etc)
Class Evaluation

- 1. Class participation: 10%

- 2. Assignments (4) 30% (10% each, will count best 3). Can work individually or in teams of 2

- 3. Final examination (take-home during exam week): 60%
Teaching Assistant

- Dan Ovando
- dovando@bren.ucsb.edu
Textbooks / Readings

- No required textbook

- Recommended #1: “Mostly Harmless Econometrics” by J. Angrist & S. Pischke [A&P]
  - Modern presentation (and not too technical) of most of the material from this class
  - About $25 in paperback

Textbooks / Readings (ctd)

- Recommended #2: Undergraduate textbook “Introduction to Econometrics” by J. Stock & M. Watson [S&W]
  - I will put on copy in reserve at the Library
  - There are 3 editions available, all are fine, and you may find a used version for about $50

- Other readings will be posted on the class website
Outline of class

1. Basics of regression analysis
   - Bivariate regression
   - Multivariate regression
   - Nonlinear specifications, dummy variables, interactions
   - Hypothesis tests

2. Internal / External validity of empirical studies

3. Definition and framework of causal relationships

4. Regression and Matching

5. Propensity Score Methods
Outline of class (ctd)

☐ 6. Instrumental Variables Methods

☐ 7. Regression Discontinuity Methods (if time permits)

☐ 8. Fixed Effects and Difference-in-Difference Methods

☐ 9. Introduction to ‘Big Data’ methods (if time permits)
Disclosure: (1)

- I was training as a labor economist but am now a researchers environmental economics (mostly)

- Most of the “go to” examples and applications I know are from these fields (including the textbooks and readings)
  - This is especially true of the homeworks

- I will try my best to find good applications in other fields of interests for Bren students (ecology, hydrology, etc)

- Feel free to pass along some suggestions to this end!
Disclosure: (2)

- Like most economists, I use the commercial statistical software Stata

- My class demonstrations and homework solutions will use Stata

- It is available in the Bren “GIS” lab and other some labs on campus (including new library)

- Students can use the software of their choice for this class. “R” seems like the best suited one besides Stata
Before next class:

- Review slides from S&W on class website
- Begin reading S&W Chapters 4-7
- Begin reading A&P Chapter 3