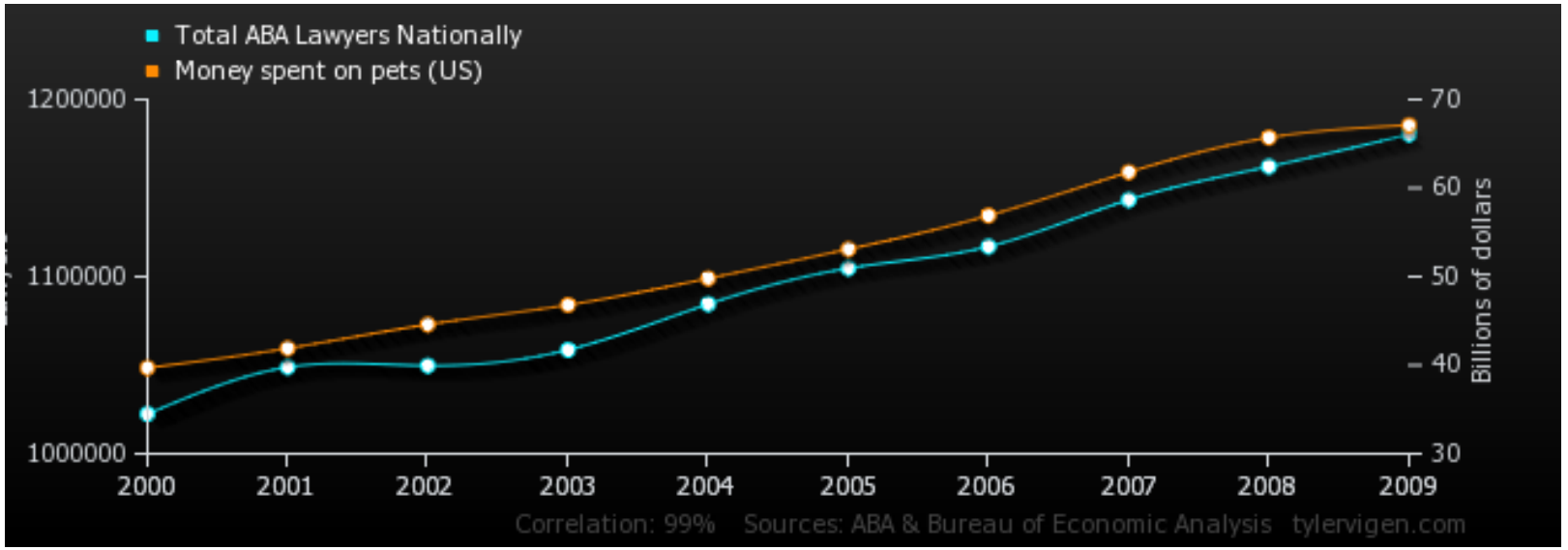
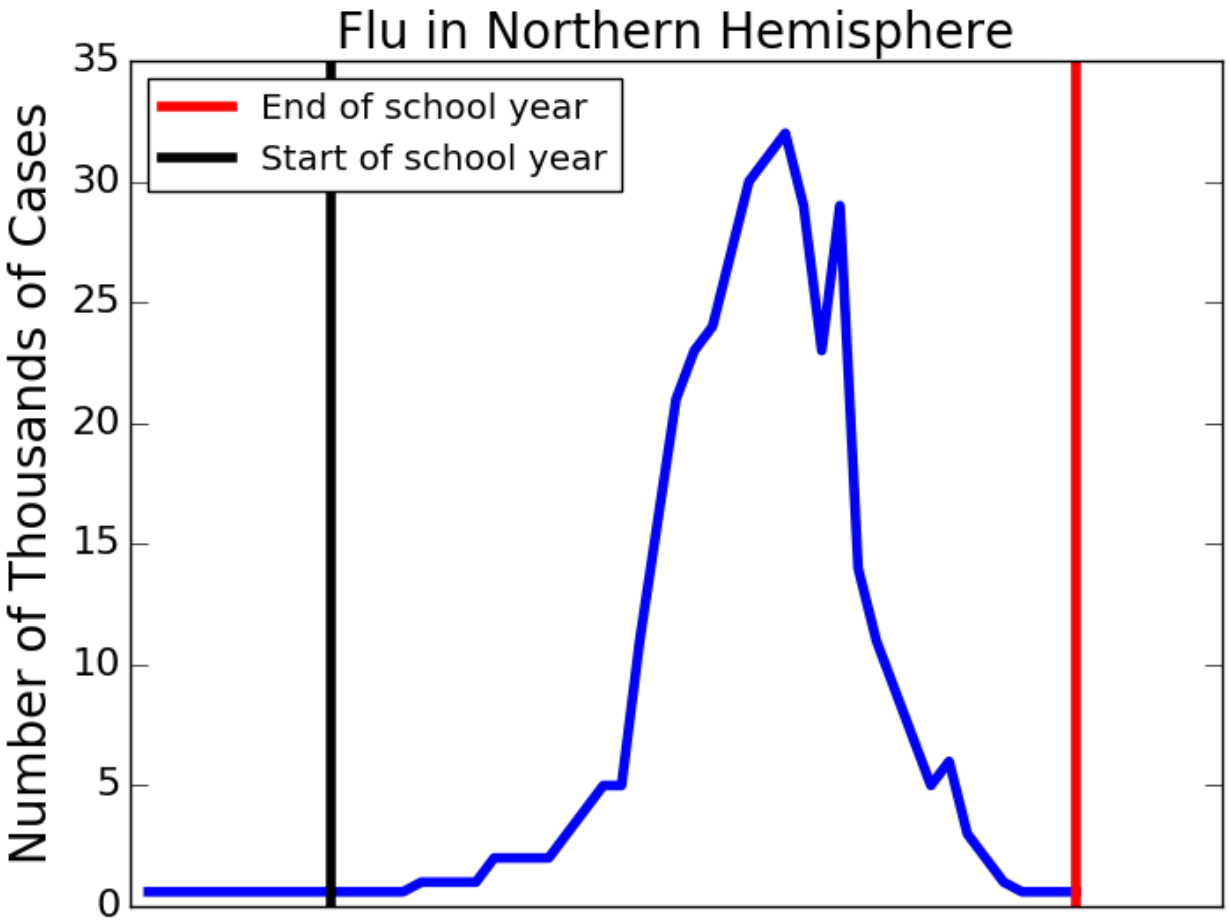
Statistical Abuses and Course Wrap Up

Cum Hoc Ergo Propter Hoc

- With this, therefore because of this
- Humans "wired" to find patterns in information, and like to think causally

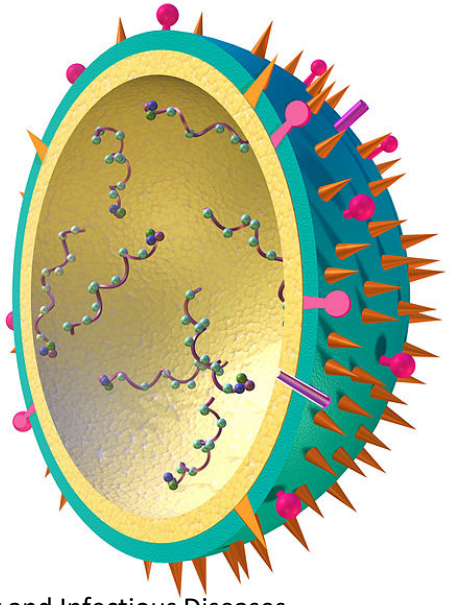


Flu and School Attendance



Flu and School Attendance

- Does going to school contribute to the spread of flu?
- Does the spread of flu contribute to school attendance?
- Or is there some other **lurking variable**?



National Institute of Allergy and Infectious Diseases

Establishing Causation

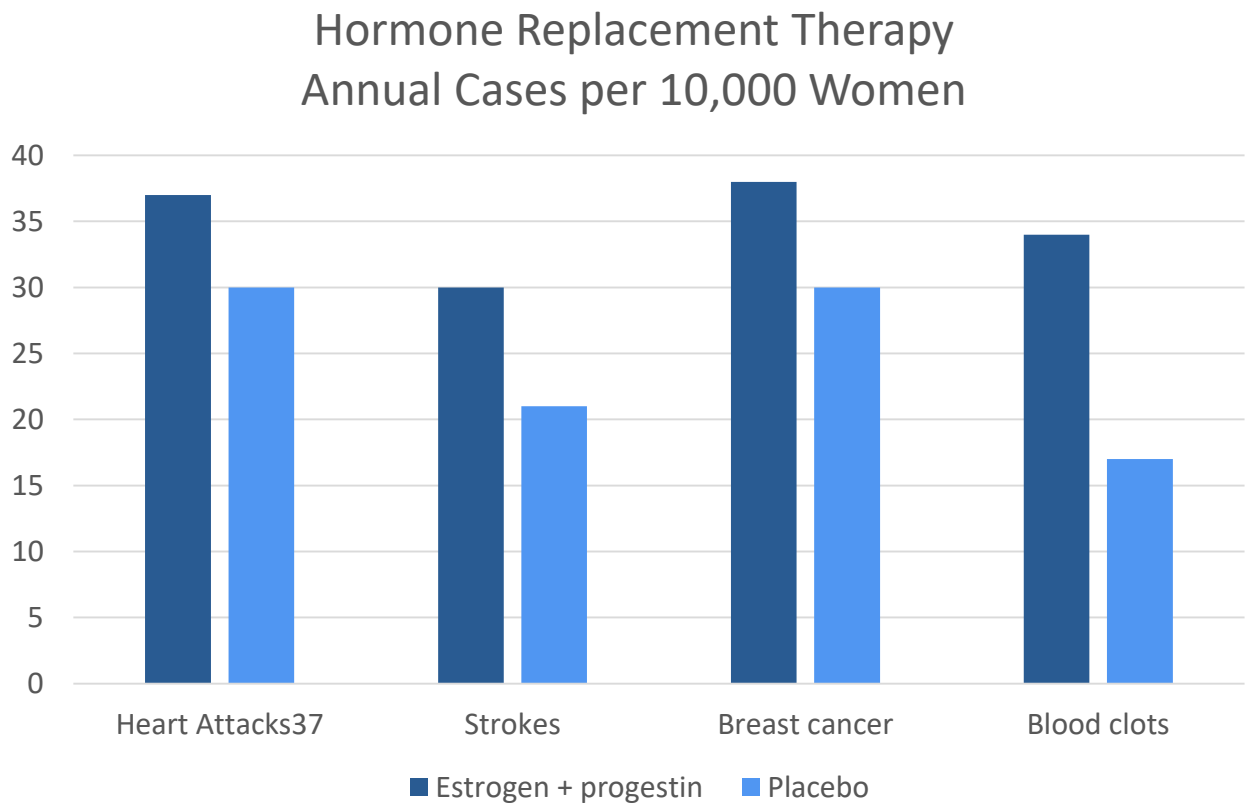
- Attempt to control for all variables other than the variables of interest
  - Rarely possible
- Randomized control studies the gold standard
  - Start with a population
  - **Randomly** assign members to either
    - Control group
    - Treatment group
  - Deal with two groups identically except with respect to the one thing being evaluated
  - Very hard to do

Hormone Replacement Therapy

- "Multiple observational studies suggest a marked reduction in risk of coronary heart disease (CHD) associated with postmenopausal estrogen use."

Annual Review of Public Health, 1998

Hormone Replacement Therapy

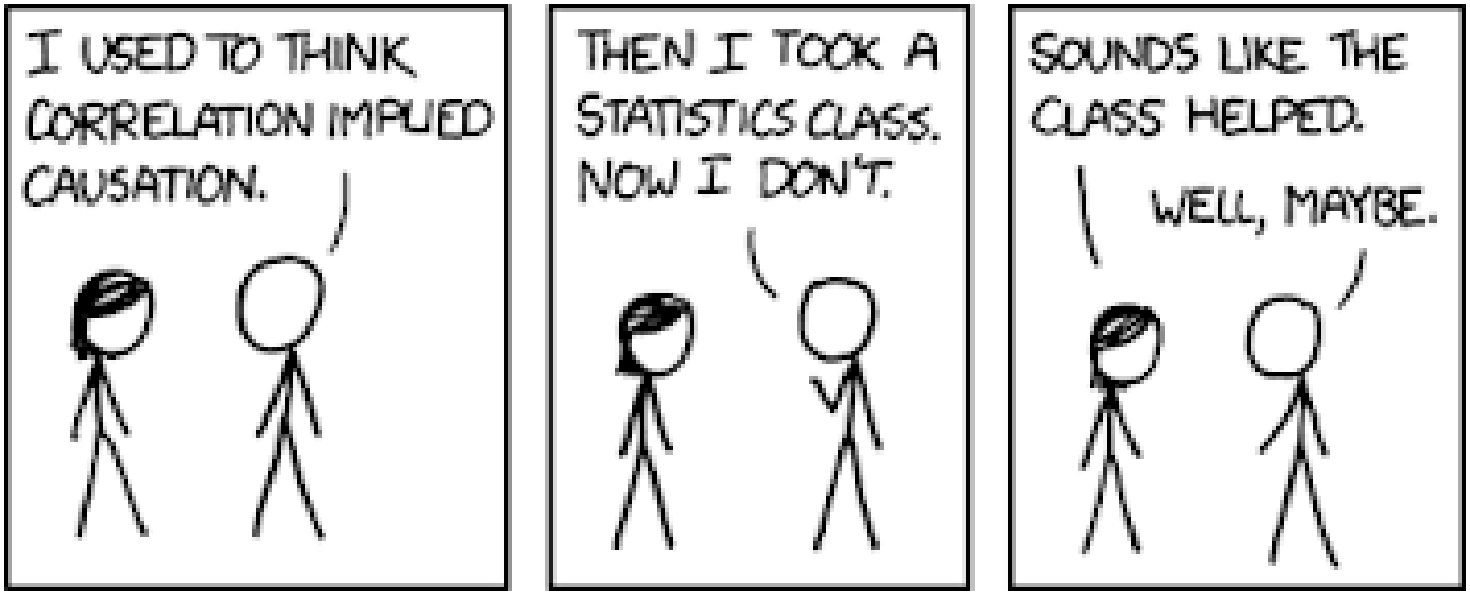


Hormone Replacement Therapy

- "Multiple observational studies suggest a marked reduction in risk of coronary heart disease (CHD) associated with postmenopausal estrogen use."

Annual Review of Public Health, 1998

Hope This Helps



xkcd

6.00.2x Major Topics

- Optimization problems
- Stochastic thinking
- Modeling aspects of the world
  - Random walks
  - Games of chance
  - Making sense of data
- Becoming a better programmer
  - Exposure to a few extra features of Python and some useful libraries
  - Practice, practice, practice

Optimization Problems

- Many problems can be formulated in terms of
  - Objective function
  - Set of constraints
- Greedy algorithms often useful
  - But may not find optimal solution
- Many optimization problems inherently exponential
  - But dynamic programming often works
  - And memoization a generally useful technique
- Examples: knapsack problems, graph problems, curve fitting, clustering

Stochastic Thinking

- The world is (predictably) non-deterministic
- Thinking in terms of probabilities is often useful
- Randomness is a powerful tool for building computations that model the world
- Random computations useful even when for problems that do not involve randomness
  - E.g., integrating curves

Modeling the World

- Models always inaccurate
  - Provide abstractions of reality
- Deterministic models, e.g., graph theoretic
- Statistical models
  - Simulation models: Monte Carlo simulation
  - Curve fitting (linear regression)
  - Machine learning
  - Characterizing accuracy is critical
    - Central limit theorem
    - Empirical rule

What's Next for You?

- Try to use what you have learned
  - Remember that you can write programs to get answers
- Take more courses, read books about CS
- Complete that last problem set, and prepare for the final

Thanks for being one of the survivors