Welcome to Unifying Data Science!

Nick Eubank

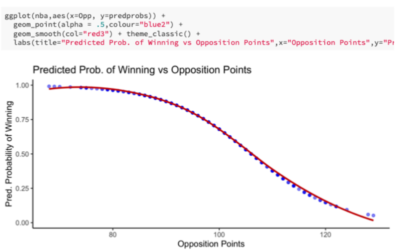
Three Goals of the Course

By the end of this course, you will:

1. Be able to critically evaluate causal claims, and develop research designs to answer causal questions.
   Causal Inference
2. Understanding how different approaches to data science relate to one another, and know when to employ different toolsets.
   The "Unifying" in Unifying Data Science
3. Execute a data science project from conception to delivery
   Complete with step-by-step models

Part One: Causal Inference

Modeling and Representation of Data



You learned a *lot*:

- Model selection
- Interpreting BIC, AIC, AUC, R-squared
- Residual Plots

$\Rightarrow$ Develop model to faithfully represent patterns in the data

Causal Inference

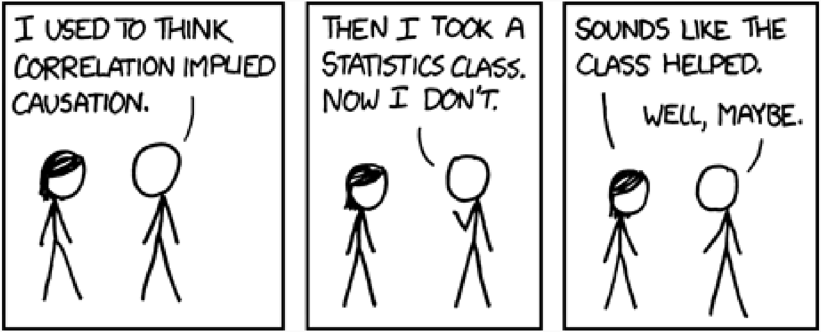
We're focused on what comes next.
Assume our model faithfully represents the data.
$\Rightarrow$ Given those models, what can we conclude **about the world**?

Suppose we find a correlation between car advertising and consumer spending across neighborhoods in North Carolina.

- Does that imply that more advertising would increase spending further?
  In other words, based on this model, do we think advertising is *causing* more consumer spending?

Causal Inference

Correlation does not imply causation


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Causal Inference

Does advertising cause increased consumer spending?

- “Well, correlation does not imply causation, so I can’t say.”
- “Well, correlation does not imply causation, *but* yea probably.”

Causal Inference

Correlation does not *necessary* imply causation,

- but when certain assumptions are met, correlation *does* imply causation.

By learning the *assumptions* that are required for a correlation to be a good estimate of a causal effect, you can:

- Evaluate whether those assumptions are likely to be met,
- Come up with different research designs whose assumptions *would* be met.

Causal Inference

Modeling

Developing Model to Faithfully Represent Data

$\Downarrow$

Inference

Interpreting Model Parameters

Causal Inference

While we will use a "statistical framework" (Potential Outcomes Framework) to help us be rigorous in our thinking...

There are **NO** statistical tests that will tell you if your model is estimating a true causal effect.

- *Fundamental Problem of Causal Inference*

Causal inference is unavoidably about:

- Critical thinking
- Case knowledge

Causal Inference

After introducting Potential Outcomes, we'll explore a range of causal research techniques:

- Experiments (Randomized-Control Trials, or RCTs)
- Linear Regressions as Causal Tools
- Matching
- Differences in Differences
- Natural Experiments

$\Rightarrow$ Each of these designs will provide causal estimates if certain assumptions are met,

- But it will always be is up to you, the researcher, to evaluate whether those assumptions are reasonable!

Causal Inference

By the end of this course, you will:

- Understand why causal inference is hard,
- Be able to critically evaluate causal evidence collected by others,
- Articulate causal questions,
- And develop research designs to answer those questions.