Making Policy with Data

An Introductory Course on Policy Evaluation

Final Review

Instructor: Prof Yiqing Xu

June 8

Final Exam

June 13 (Tuesday)

Time: 9:30—11:00 AM

Location: WLH 2111

Key Concepts: First Half of the Course

1. The potential outcome model
2. Causal effects
3. The fundamental problem of causal inference
4. Causal estimands: ATE, ATT, ATC
5. Selection bias
6. Identification problem
7. Estimators, estimates, and statistical inference

Selection bias

- Comparisons between observed outcomes of treated and control units can often be misleading term unlikely to be 0 in most applications

$$
\begin{aligned}
E[Y_i|D = 1] - E[Y_i|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{\text{ATT}} + \underbrace{\{E[Y_{i0}|D_i = 1] - E[Y_{0i}|D_i = 0]\}}_{\text{BIAS}}
\end{aligned}
$$

- Bias term unlikely to be 0 in most applications

- Selection into treatment is often associated with the potential outcomes.

Key Concepts: Randomized Experiments

1. Identification under random assignment
2. Estimation: Difference-in-Means
3. Estimation: Regression
4. Experimental Design (e.g. SUTVA)
5. Inference (standard errors, hypothesis testing)

Identification Strategies (and Estimators)

1. RCT (random assignment)
    - Difference-in-Means
    - Regression
2. Selection on observables
    - Matching
    - Regression
3. Difference-in-Differences
    - Double-Difference
    - Regression

Selection on Observables

Recall that randomized experiments work because:

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i$$

Assumption: Conditional Ignorability

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i \mid X_i = x \quad \text{for any} \quad x \in \mathcal{X}$$

(a.k.a. exogeneity, unconfoundedness, selection on observables, no omitted variables)

Read: Among units with same values of X_i, D_i is "as-if" randomly assigned.

Assumption: Common Support

$$0 < \Pr(D_i = 1 \mid X_i = x) < 1 \quad \text{for any} \quad x \in \mathcal{X}$$

Read: For any value of X_i, unit could have received treatment or control

“As-if” randomization

- **“Find strata of X in which you think an experiment is occurring”**
- Approximate a randomized experiment within subgroups
- Plausibility of SOO: can you argue that variation in treatment status within strata of X is random?
- Placebo/Falsification test to alleviate concerns of omitted variables

Matching and Regression

- **Matching**
  - Nonparametric and transparent
  - Overlap is guaranteed
  - Curse of dimensionality: lose data as we have more and more covariates
- **Regression**
  - Easy to implement; no waste of data
  - Functional form assumptions: constant treatment effect, linearity
  - Extrapolation

Omitted Variable Bias (OVB) Equation

- Wages on schooling (S), controlling for ability (A)

$$Y_i = \alpha + \rho S_i + A_i'\gamma + \epsilon_i$$

- Ability is hard to measure. What if we leave it out?

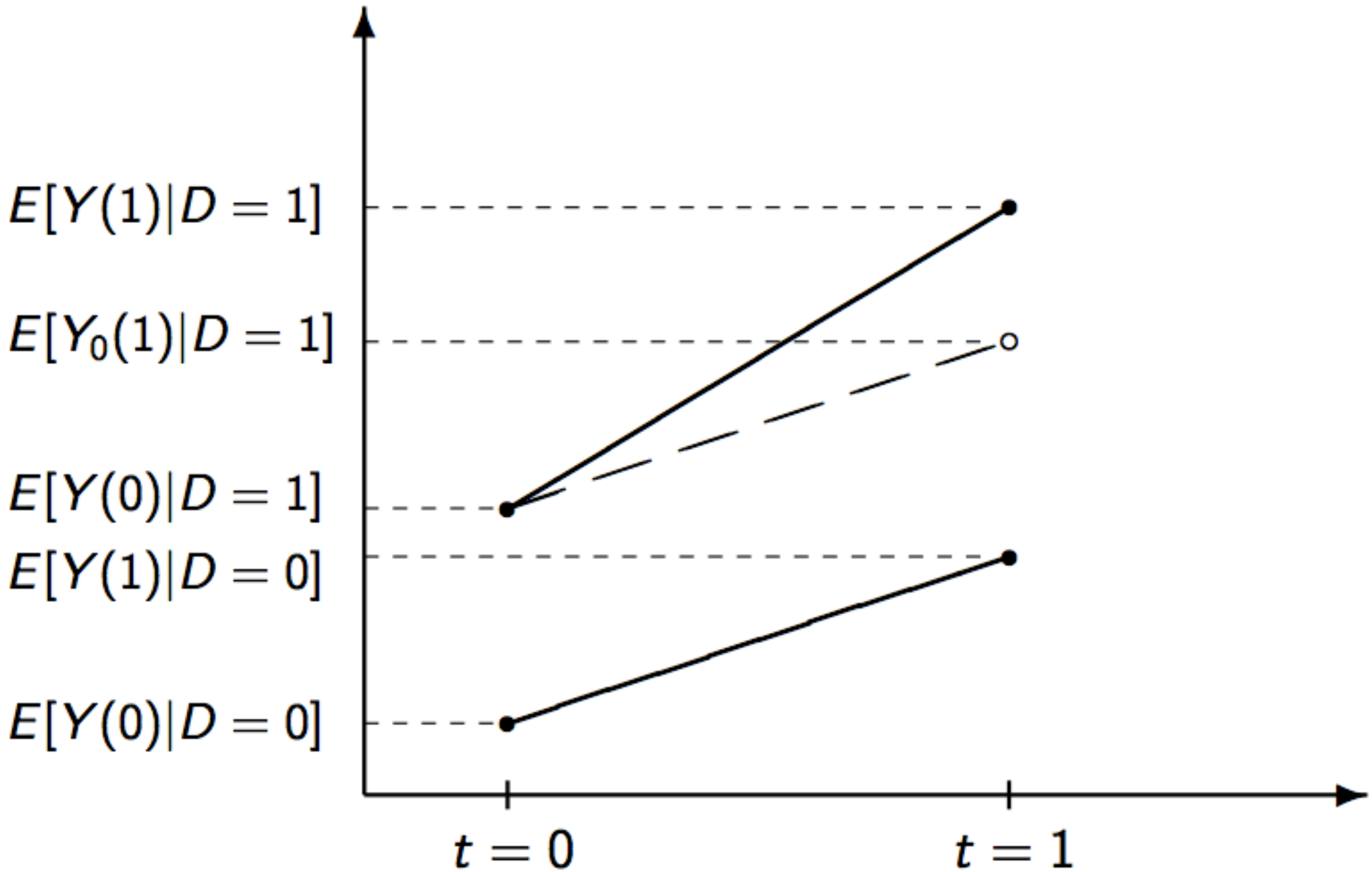
- Omitted variable bias = The effect of the omitted ×
  The correlation between the omitted (A) and the included (S)

$$\frac{Cov(Y_i, S_i)}{V(S_i)} = \rho + \gamma'\delta_{AS}$$

Difference-in-Differences

- Diff-in-Diffs: An extremely popular strategy when there is longitudinal data (panel or repeated cross-sections) and the treatment is one-shot
- Allow selection on time-invariant unobservables and time-varying observables; but not on time-varying unobservables (which lead to unparalleled trends)
- ***Parallel trends:*** unobserved confounders must be additive and time-invariant
- Always be cautious about the assumptions you make. Better to have multiple periods
- Common setup:
  - One difference: before and after
  - Another difference: across units (between treated and controls)

Difference-in-Differences



Identification with Diff-in-Diffs

Identification Assumption (parallel trends)

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

Identification Result

Given parallel trends the ATT is identified as:

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Estimation

Estimand (ATT)

$$
\begin{aligned}
E[Y_1(1) - Y_0(1)|D=1] \;&=\; \Big\{E[Y(1)|D=1] - E[Y(1)|D=0]\Big\} \\
&-\; \Big\{E[Y(0)|D=1] - E[Y(0)|D=0]\Big\}
\end{aligned}
$$

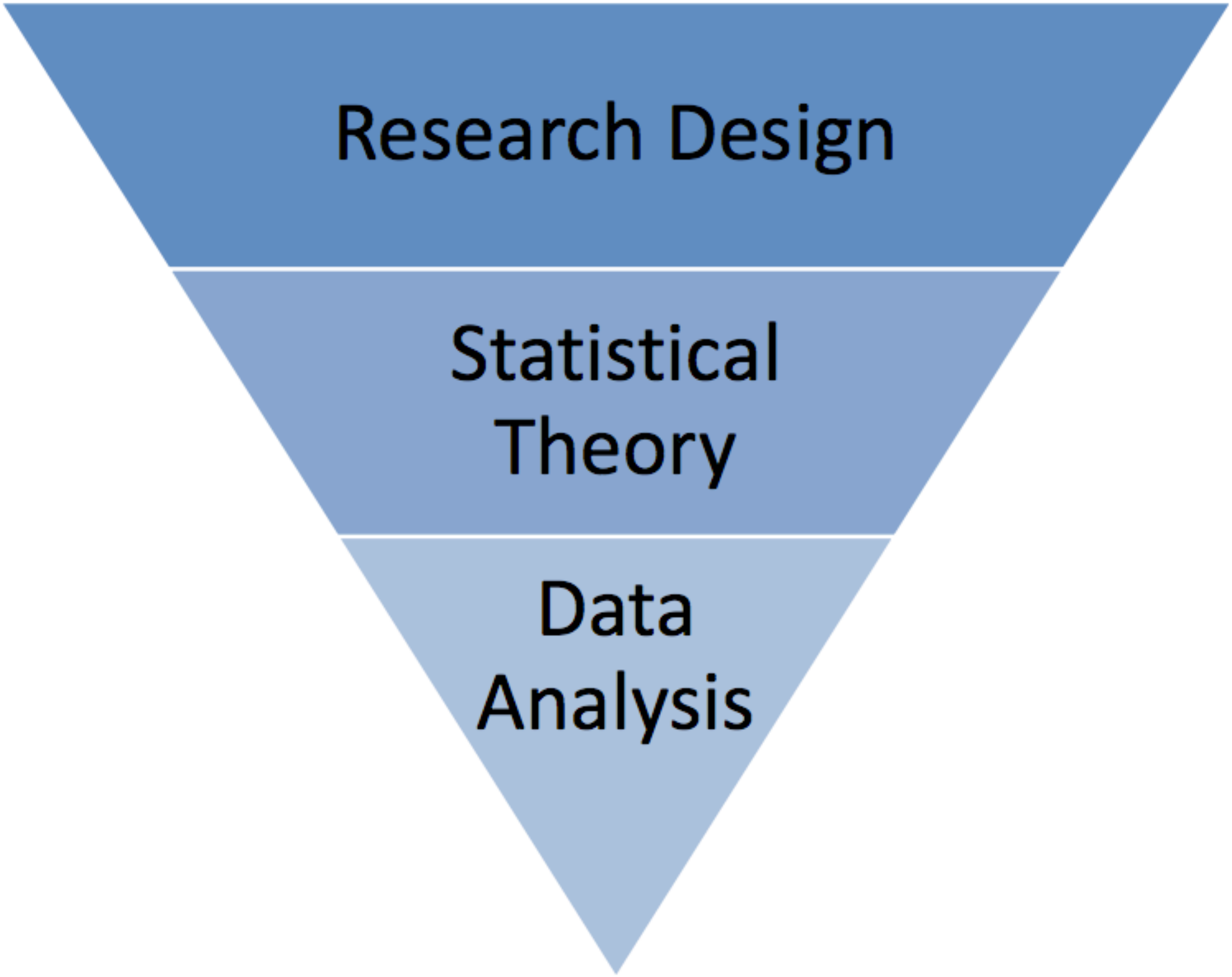
Estimator (Sample Means: Panel)

$$
\left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(1) - \frac{1}{N_0}\sum_{D_i=0} Y_i(1)\right\} - \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(0) - \frac{1}{N_0}\sum_{D_i=0} Y_i(0)\right\}
$$

$$
= \left\{\frac{1}{N_1}\sum_{D_i=1}\{Y_i(1) - Y_i(0)\} - \frac{1}{N_0}\sum_{D_i=0}\{Y_i(1) - Y_i(0)\}\right\},
$$

where N_1 and N_0 are the number of treated and control units respectively.

So What Do We Do?



Where Are We?

- **Causal Inference**
  - You're gonna forget all the Y1, Y0 stuff
  - But you've seen how good researches are done
- **Statistics**
  - You're gonna forget bias correction and clustered SEs
  - But you know good statistical analysis is not scary
- **R**
  - You're gonna forget all the messy options
  - But hopefully you're not afraid at writing code anymore

The Take-aways

- **Correlation is not causation**
    - Mainly because of selection bias
- **Compare like with like**
    - Find methods to eliminate selection bias
- **Think of the counterfactuals**
    - Use statistics to predict counterfactuals

What are Left Out?

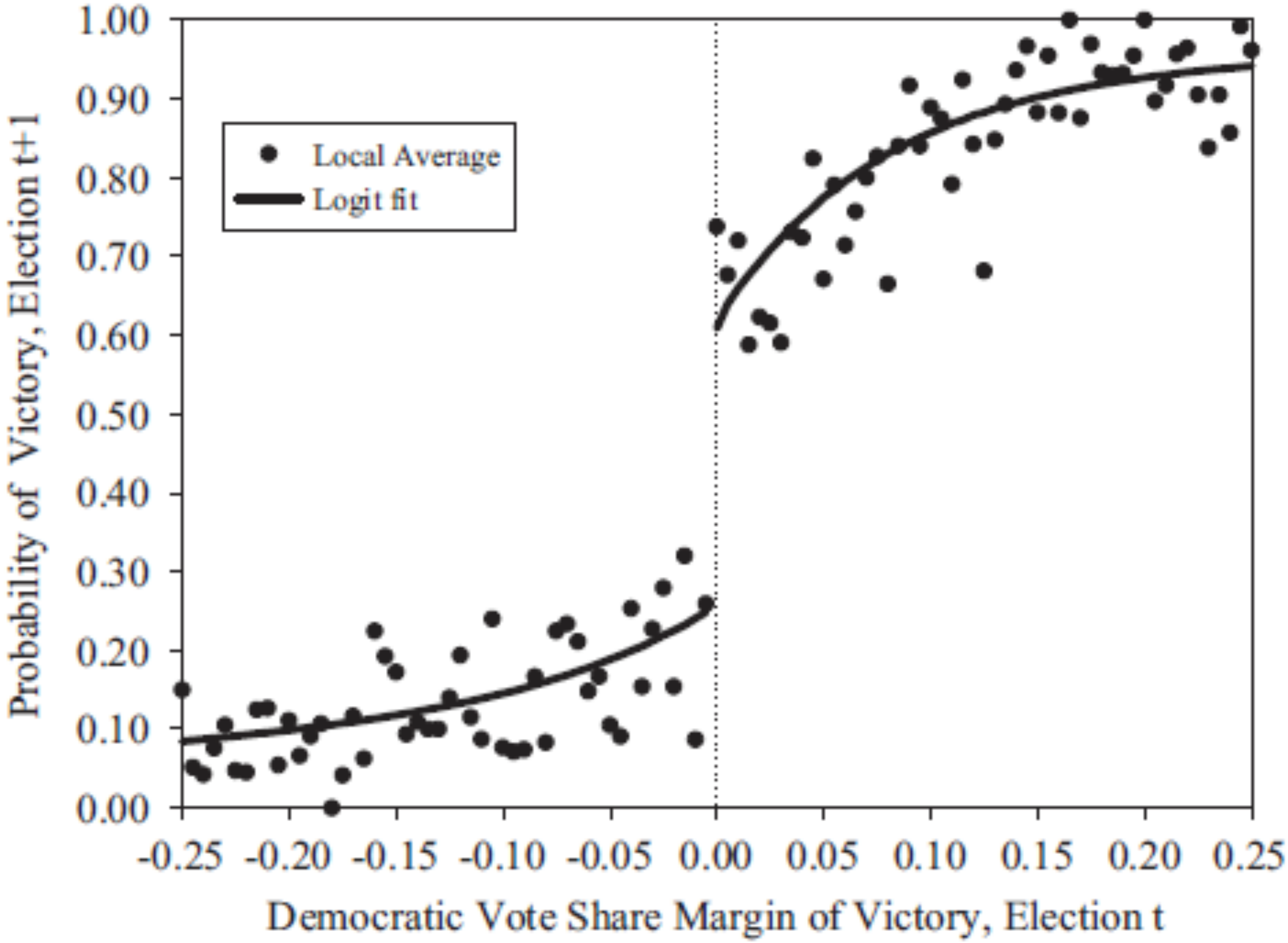
Instrumental Variables (AP Chapter 3)

- "No compliance" issue
  - In experiments, we often cannot force subjects to take specific treatments
  - Units choosing to take the treatment may differ in unobserved characteristics from units that refrain from doing so

Example: Non-compliance in JTPA Experiment

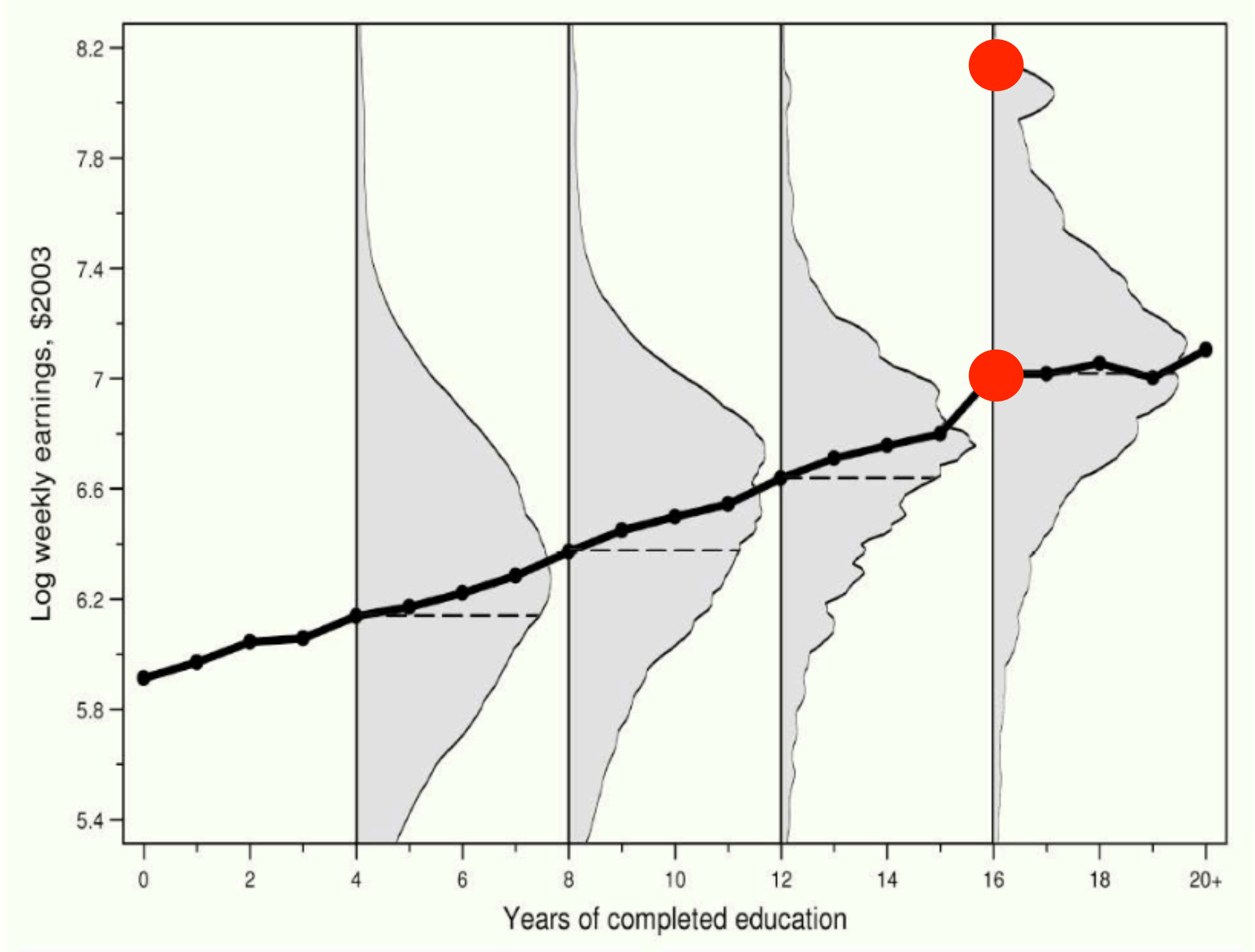
	Not Enrolled in Training	Enrolled in Training	Total
Assigned to Control	3,663	54	3,717
Assigned to Training	2,683	4,804	7,487
Total	6,346	4,858	11,204

Regression Discontinuity Designs (AP Chapter 4)

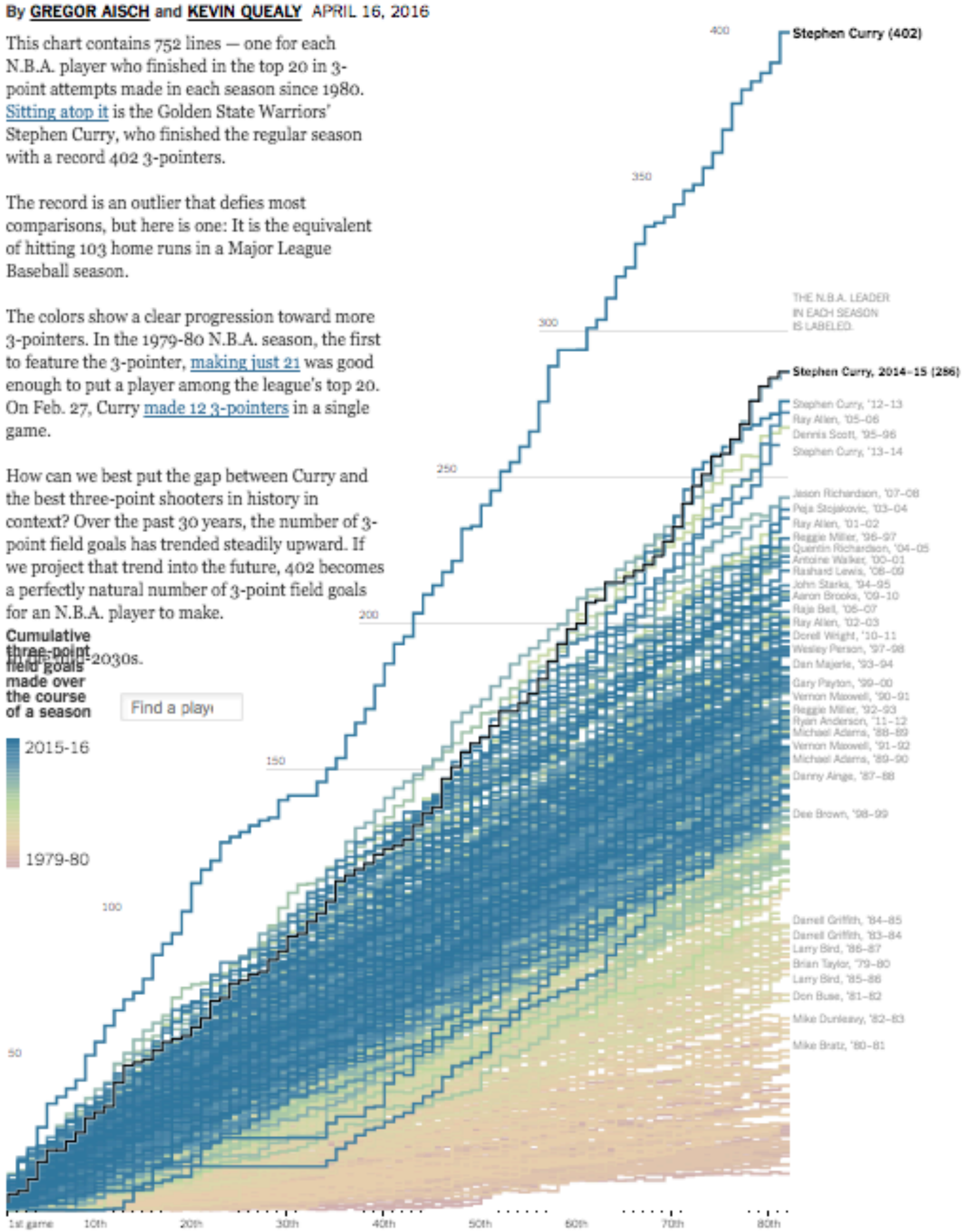


Last Word

Defying the "Expectation"



“Off the Chart”


By GREGOR AISCH and KEVIN QUEALY APRIL 16, 2016

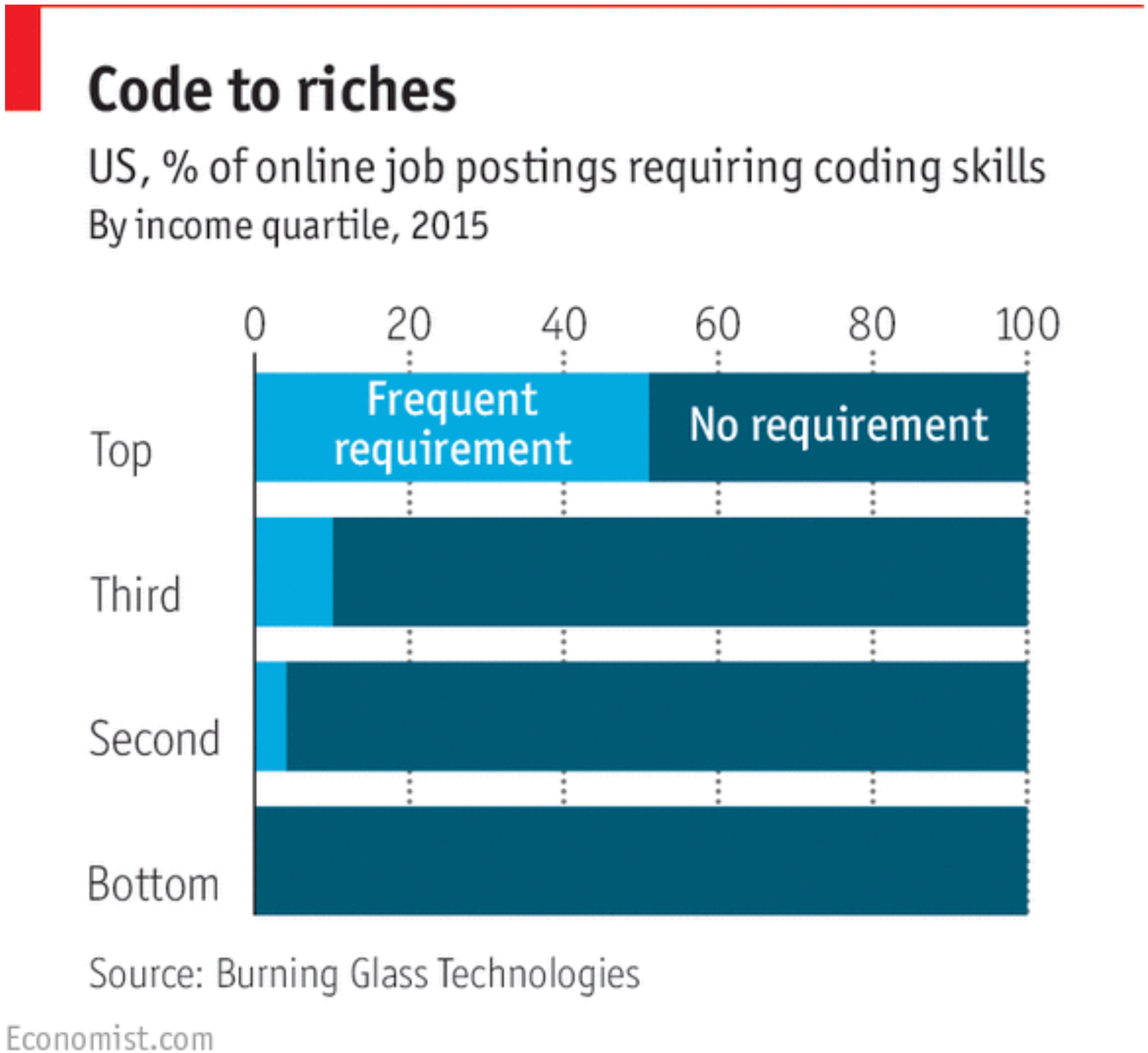
This chart contains 752 lines — one for each N.B.A. player who finished in the top 20 in 3-point attempts made in each season since 1980. Sitting atop it is the Golden State Warriors' Stephen Curry, who finished the regular season with a record 402 3-pointers.

The record is an outlier that defies most comparisons, but here is one: It is the equivalent of hitting 103 home runs in a Major League Baseball season.

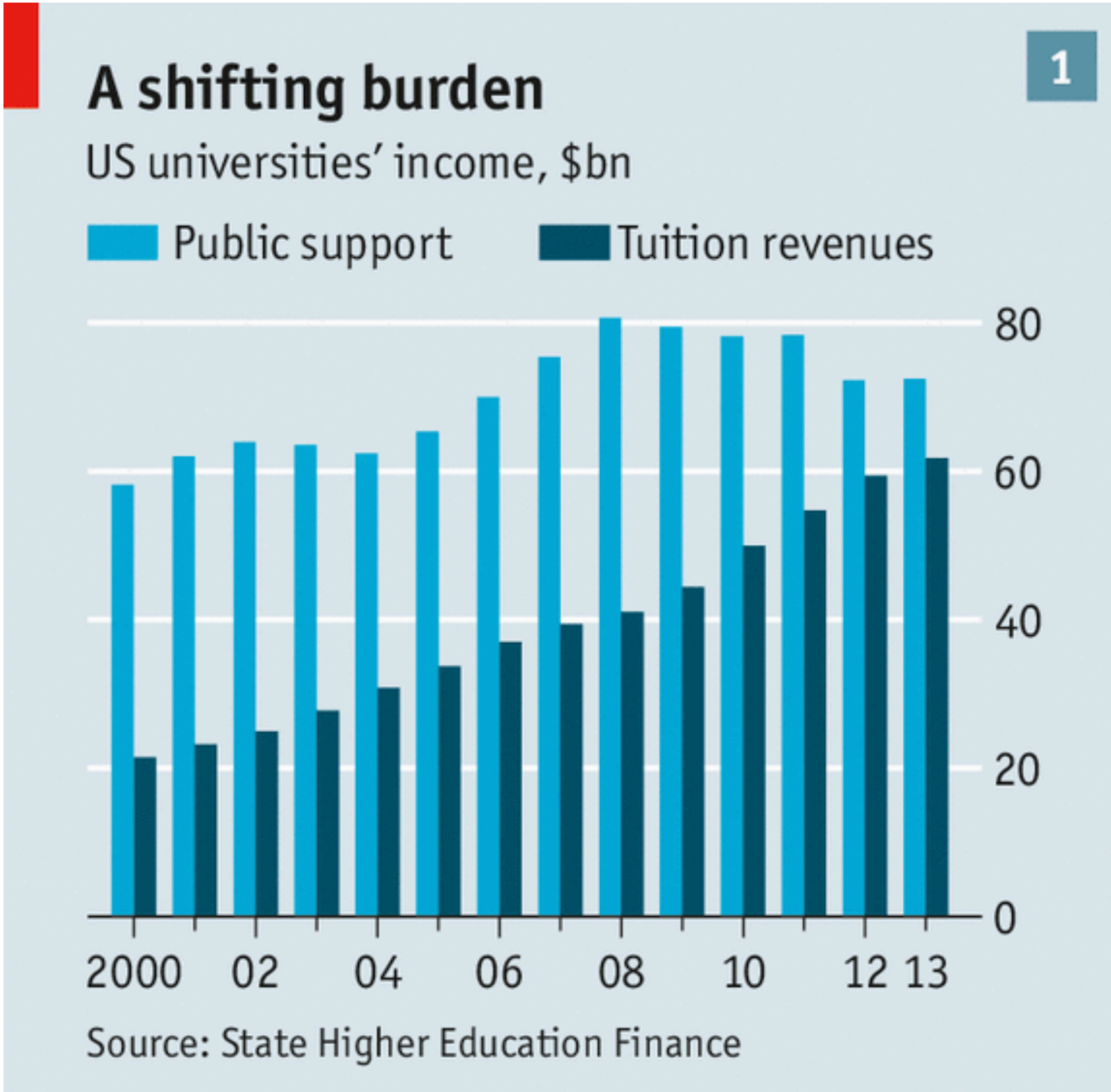
The colors show a clear progression toward more 3-pointers. In the 1979-80 N.B.A. season, the first to feature the 3-pointer, making just 21 was good enough to put a player among the league's top 20. On Feb. 27, Curry made 12 3-pointers in a single game.

How can we best put the gap between Curry and the best three-point shooters in history in context? Over the past 30 years, the number of 3-point field goals has trended steadily upward. If we project that trend into the future, 402 becomes a perfectly natural number of 3-point field goals for an N.B.A. player to make. -2030s.

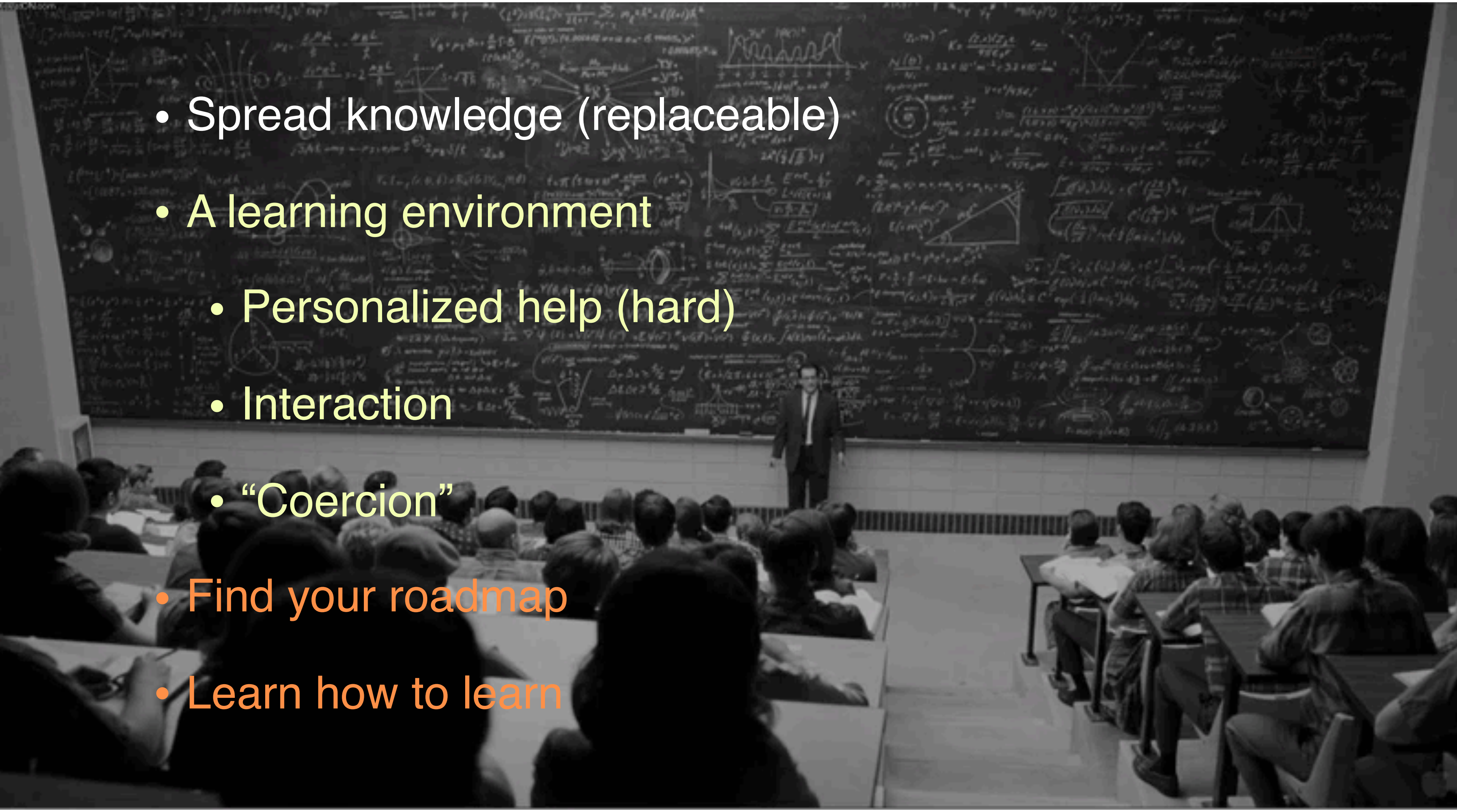
“Lifelong learning is becoming an economic imperative.”
— *Economist*



Colleges becoming more expensive (and offering less)…

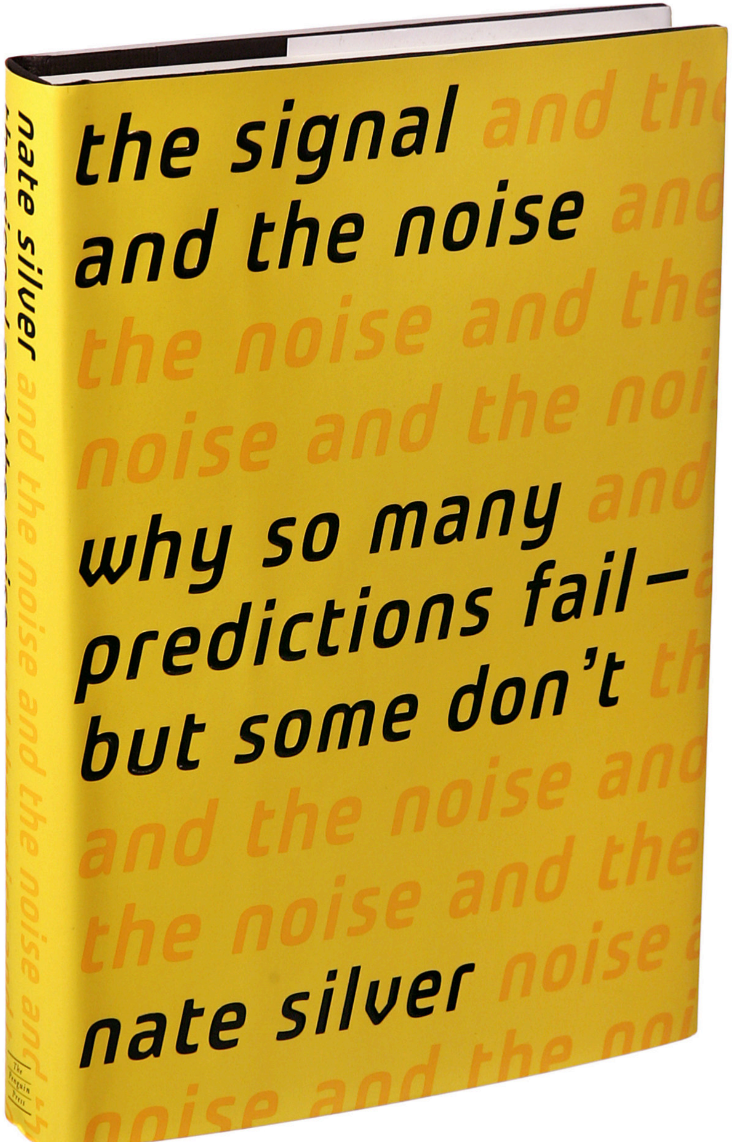


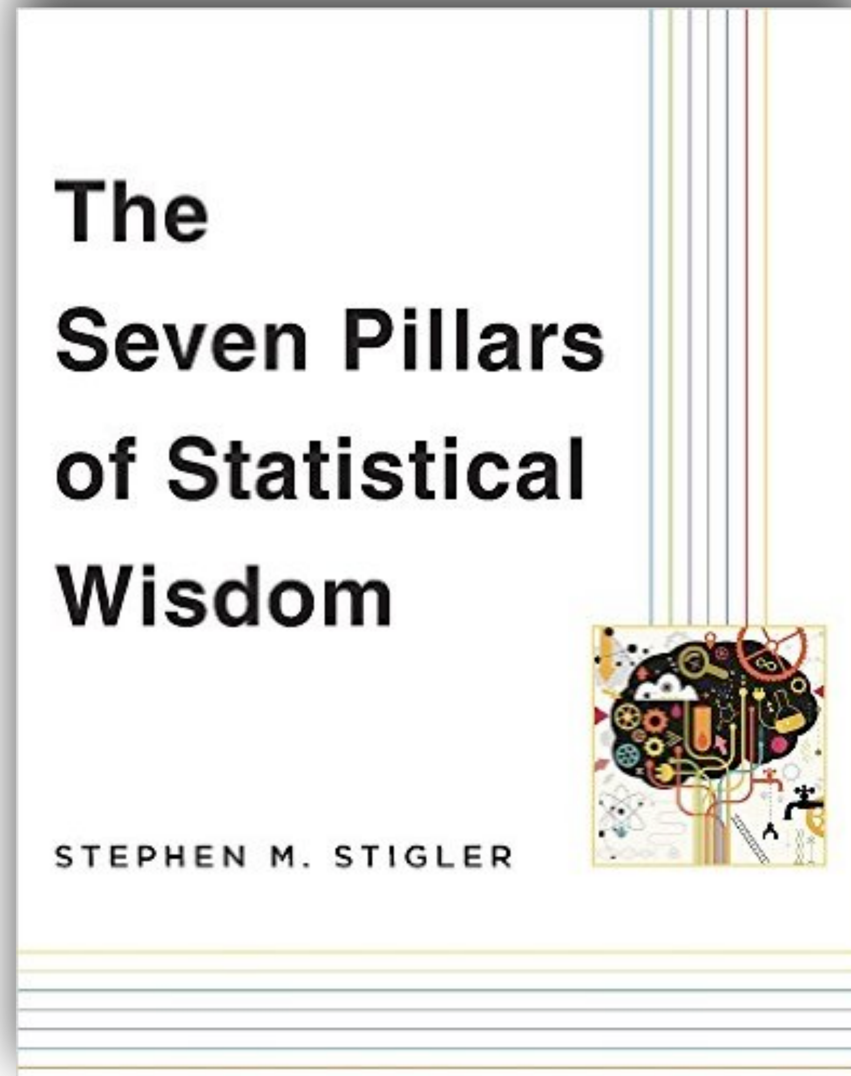
The Nature of Education



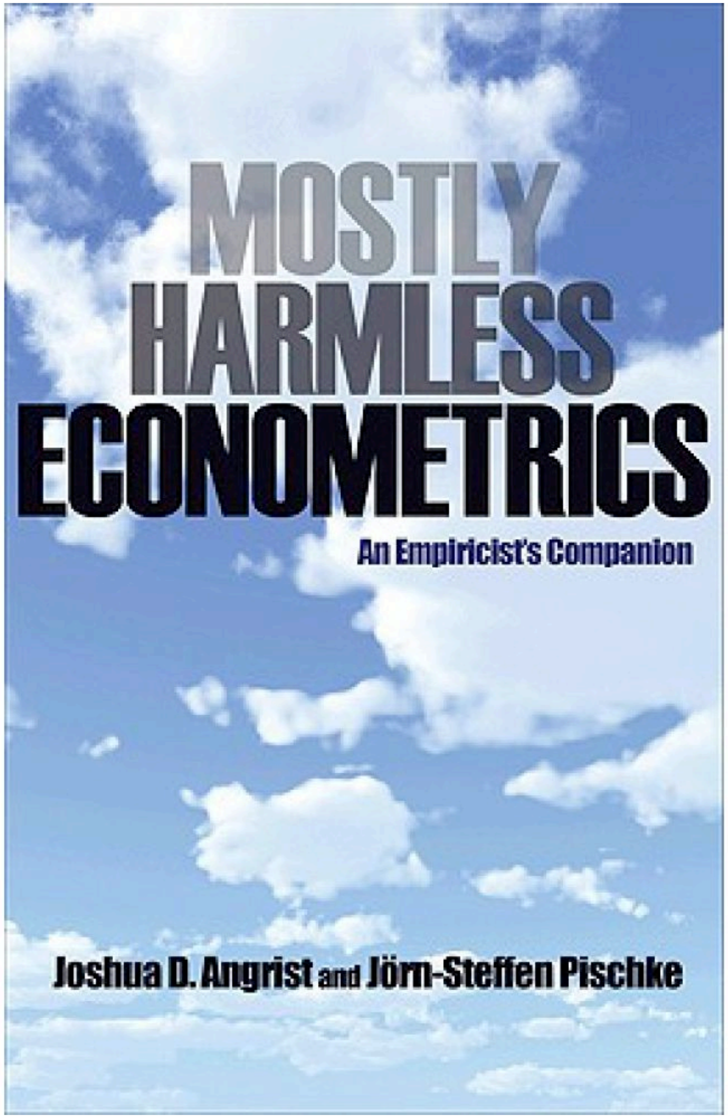
- Spread knowledge (replaceable)
- A learning environment
  - Personalized help (hard)
  - Interaction
  - “Coercion”
- Find your roadmap
- Learn how to learn

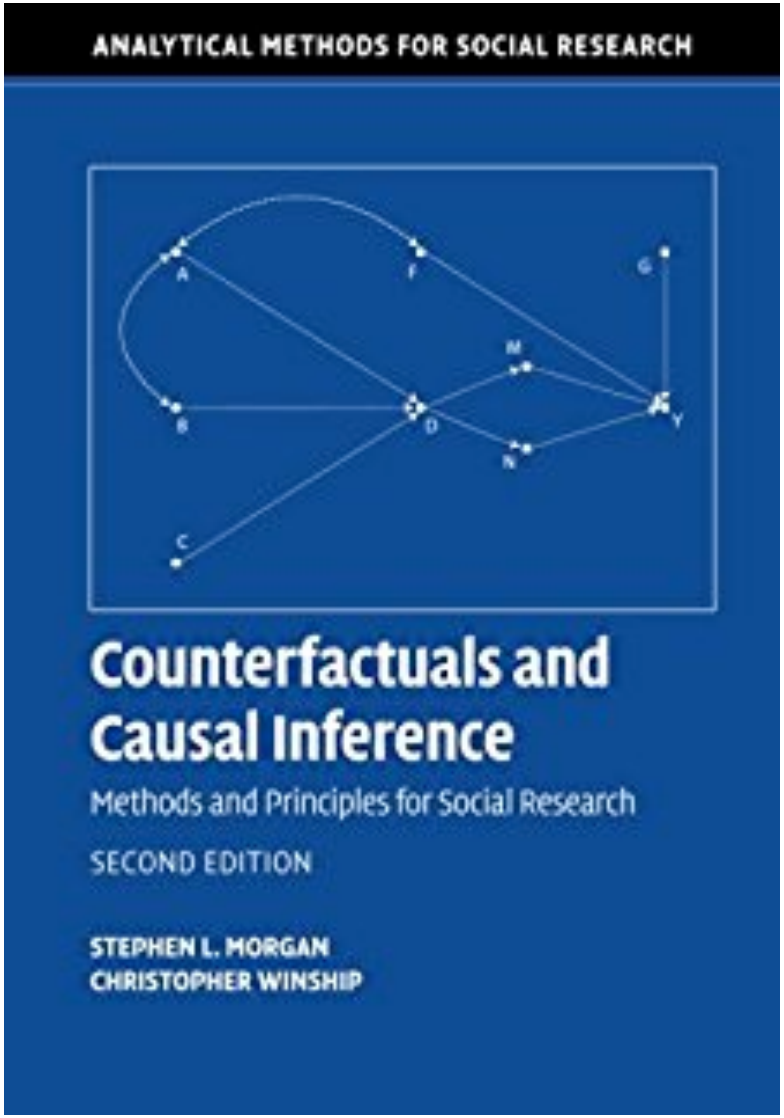
Recommended Books — Casual Reading



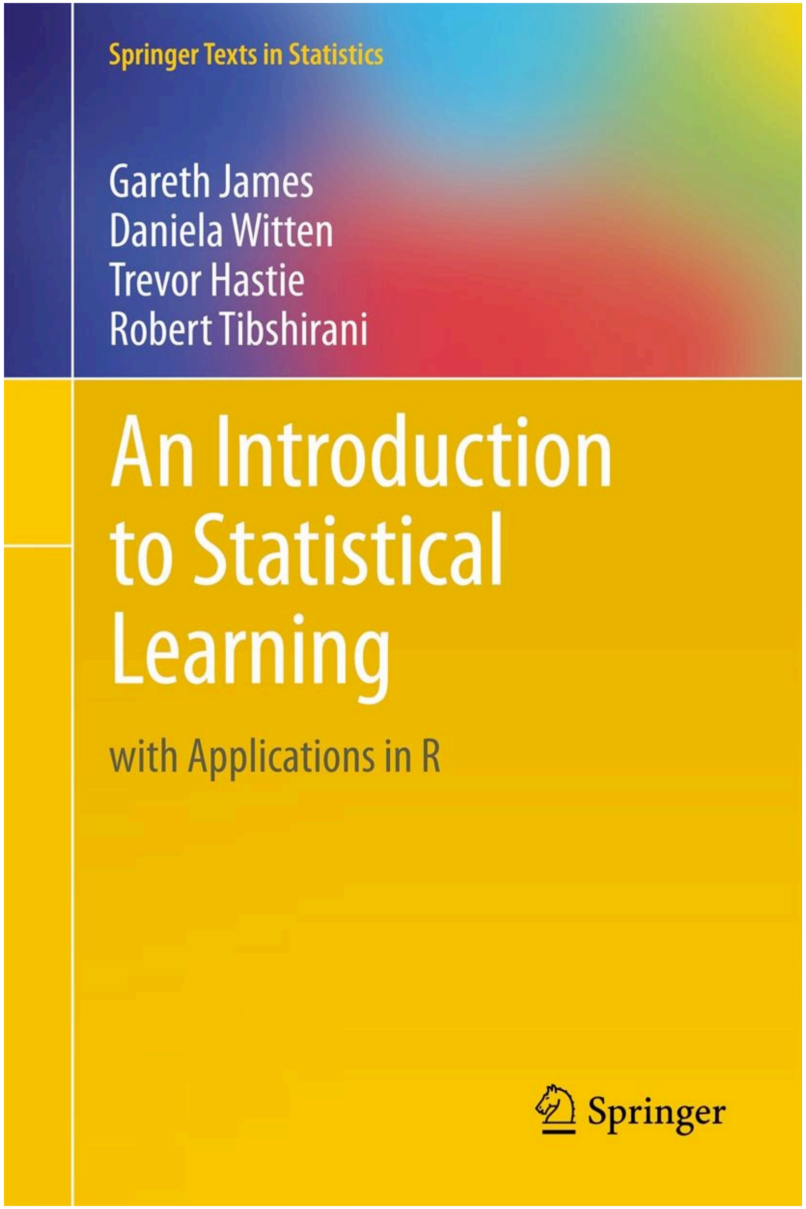


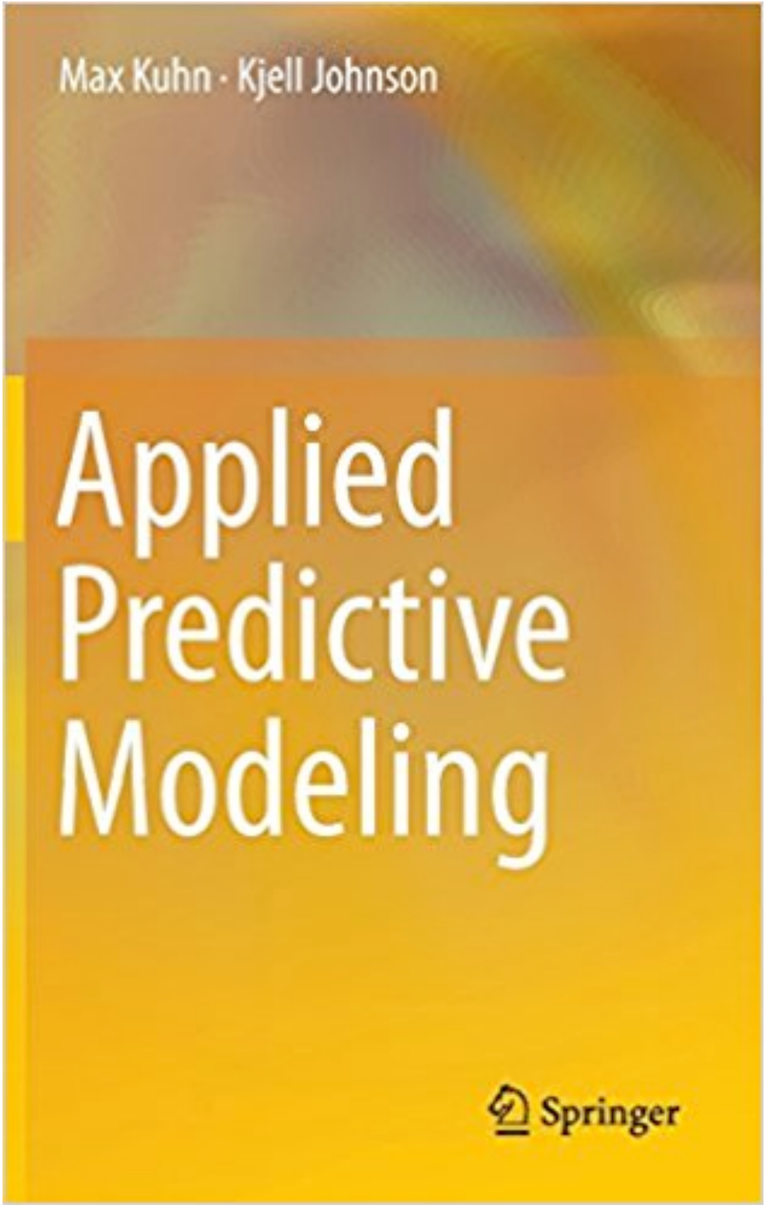
Recommended Books — Causal Inference





Recommended Books — Intro to Machine Learning





A *Mostly Harmless* Data Analyst!

- Keeping an open mind
- Ready to be convinced by data

Thank you

Yiqing Xu (徐轶青)

yiqingxu@ucsd.edu

yiqingxu.org

Hope to see you again!