

PSC 400

SYRACUSE UNIVERSITY

DATA ANALYTICS

FOR POLITICAL

SCIENCE

**ESTIMATING CAUSAL EFFECTS WITH
RANDOMIZED EXPERIMENTS**

CAUSAL EFFECT

- **Goal: Estimate causal effect of X on Y**
 - **Y: outcome variable, dependent variable**
 - **X: treatment variable, independent variable**

CAUSAL EFFECT

- **Goal: Estimate causal effect of college attendance on future earnings**

CAUSAL EFFECT

- **Goal: Estimate causal effect of college attendance on future earnings**
 - **outcome variable: earnings at age 30**
 - **treatment variable: attended college or not**

INDIVIDUAL CAUSAL EFFECT

- **Earnings of Mary if attended college - Earnings of Mary if did not attend college**

AVERAGE CAUSAL EFFECT

- **Take average of:**
 - **Earnings of Mary if attended college - Earnings of Mary if did not attend college**
 - **Earnings of Joe if attended college - Earnings of Joe if did not attend college**
 - **Earnings of Lisa if attended college - Earnings of Lisa if did not attend college**
 - **Earnings of Bob if attended college - Earnings of Bob if did not attend college**
 - **...**

FUNDAMENTAL PROBLEM

- **Fundamental problem of causal inference: We can only observe outcome in factual scenario, but never observe outcome in counterfactual scenario**
 - **Observe either income if someone attended college**
 - **Or observe income if they did not attend college**
 - **But never both**

ALTERNATIVE?

- **Mary and Joe attended college, Lisa and Bob did not**
- **Average earnings of Mary and Joe - average earnings of Lisa and Bob**
 - **Does this capture casual effect of attending college on earnings?**

PROBLEM

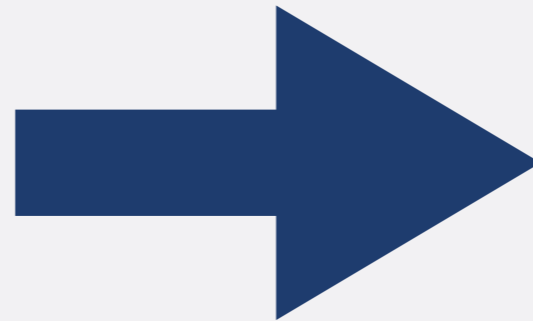
- **Mary and Joe *chose* to attended college, Lisa and Bob *chose* to not attend**
- **People who choose to attend college are different from people who choose not to attend college**
- **These differences interfere with our ability to compute the causal effect of attending college**

PROBLEM

**Academic
aptitude**



**Attending college
or not**



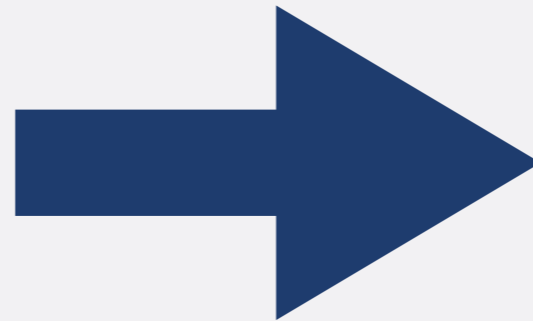
Earnings

PROBLEM

**Academic
aptitude**



**Attending college
or not**



Earnings

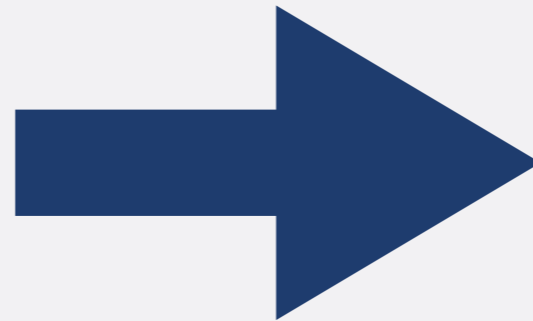
- **People with higher academic aptitude are more likely to attend college**

PROBLEM

**Academic
aptitude**



**Attending college
or not**



Earnings

- **People with higher academic aptitude are more likely to have higher earnings**

PROBLEM

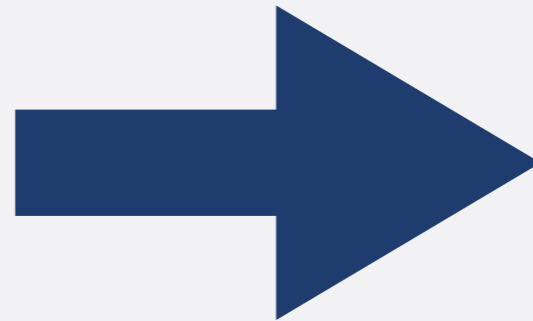
- **So if people who attend college have higher earnings, this could be due to:**
 - **Attending college**
 - **Having higher academic aptitude**
 - **Some mix of the two**

PROBLEM

Parent wealth



**Attending college
or not**



Earnings

- **Many other potential confounders**

EXPERIMENT

- **Randomly assign treatment**
- **Randomly assign people to either attend college or not**

EXPERIMENT

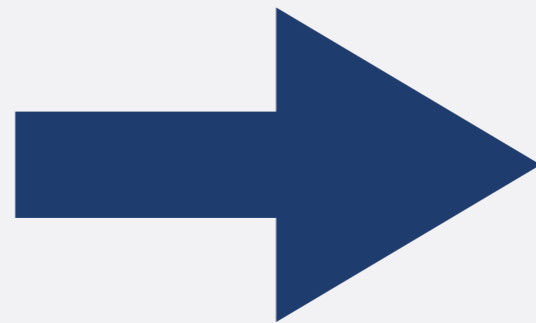
- **People who are randomly assigned to attend college on average will be the same as people randomly assigned not to attend college on everything (except attending college)**
 - **Similar academic aptitude**
 - **Similar parental wealth**
 - **etc.**

AVERAGE CAUSAL EFFECT

- **Average earnings of people randomly assigned to attend college - Average earnings of people randomly assigned to not attend college**
 - **Average causal effect**
 - **Also known as average treatment effect (ATE)**

WHAT WE WANT TO KNOW

**Smaller vs.
larger class size**



**Academic
Outcomes**

PROBLEM

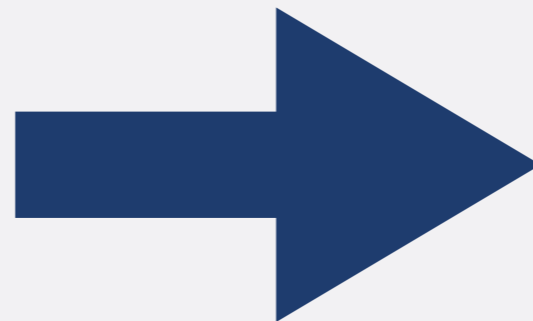
Parents socio-economic background



Smaller vs.
larger class size



Academic
Outcomes



EXPERIMENT

- **Students who are randomly assigned to be in smaller classes will be the same as students randomly assigned to larger classes on everything (except class size)**
 - e.g. similar parental wealth
 - Similar gender composition
 - etc.

AVERAGE CAUSAL EFFECT

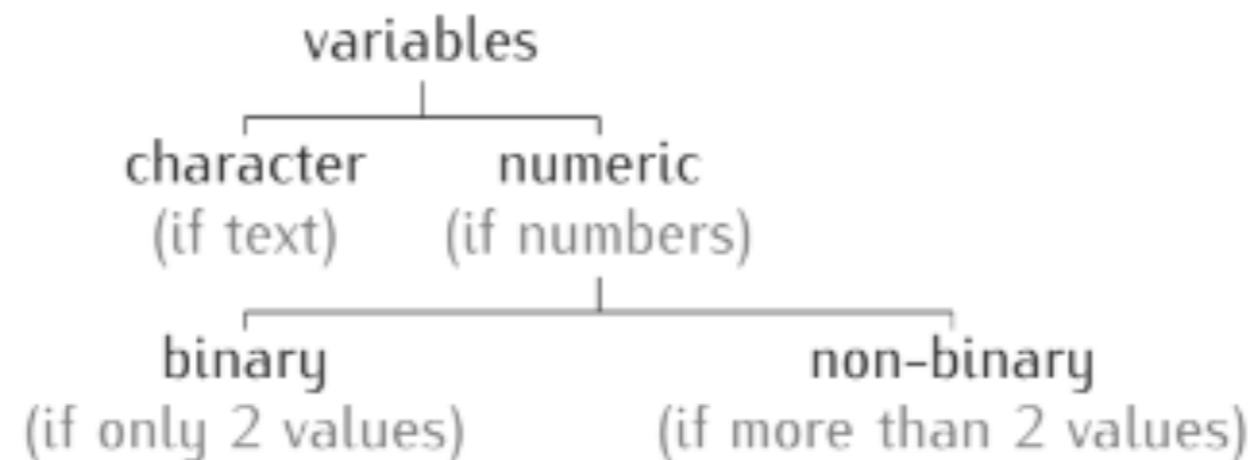
- **Average score of students randomly assigned to small classes - Average score of students randomly assigned to larger classes**
 - **Average causal effect**
 - **Also known as average treatment effect (ATE)**

DATASET: STAR.CSV

variable	description
<i>classtype</i>	class size the student attended: "small" or "regular"
<i>reading</i>	student's 3rd-grade reading test scores (in points)
<i>math</i>	student's 3rd-grade math test scores (in points)
<i>graduated</i>	identifies whether the student graduated from high school: 1=graduated or 0=did not graduate

DATASET: STAR.CSV

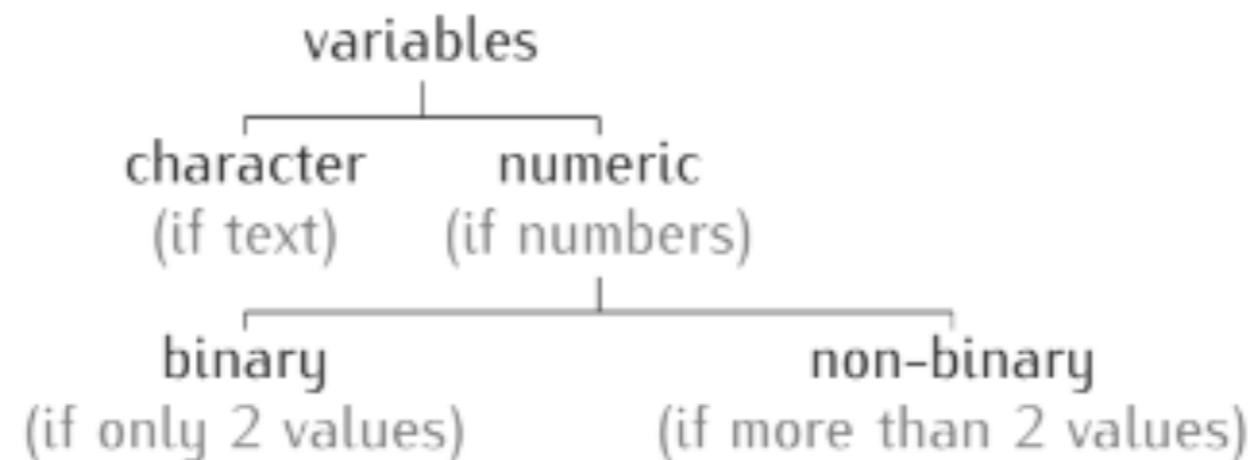
variable	description
<i>classtype</i>	class size the student attended: "small" or "regular"
<i>reading</i>	student's 3rd-grade reading test scores (in points)
<i>math</i>	student's 3rd-grade math test scores (in points)
<i>graduated</i>	identifies whether the student graduated from high school: 1=graduated or 0=did not graduate



- What kind of variable is classtype?

DATASET: STAR.CSV

variable	description
<i>classtype</i>	class size the student attended: "small" or "regular"
<i>reading</i>	student's 3rd-grade reading test scores (in points)
<i>math</i>	student's 3rd-grade math test scores (in points)
<i>graduated</i>	identifies whether the student graduated from high school: 1=graduated or 0=did not graduate



- What kind of variable is graduated?