universität
wien

# BAK3: Introduction to Quantitative Methods

## Week 2: Introduction to Research Design

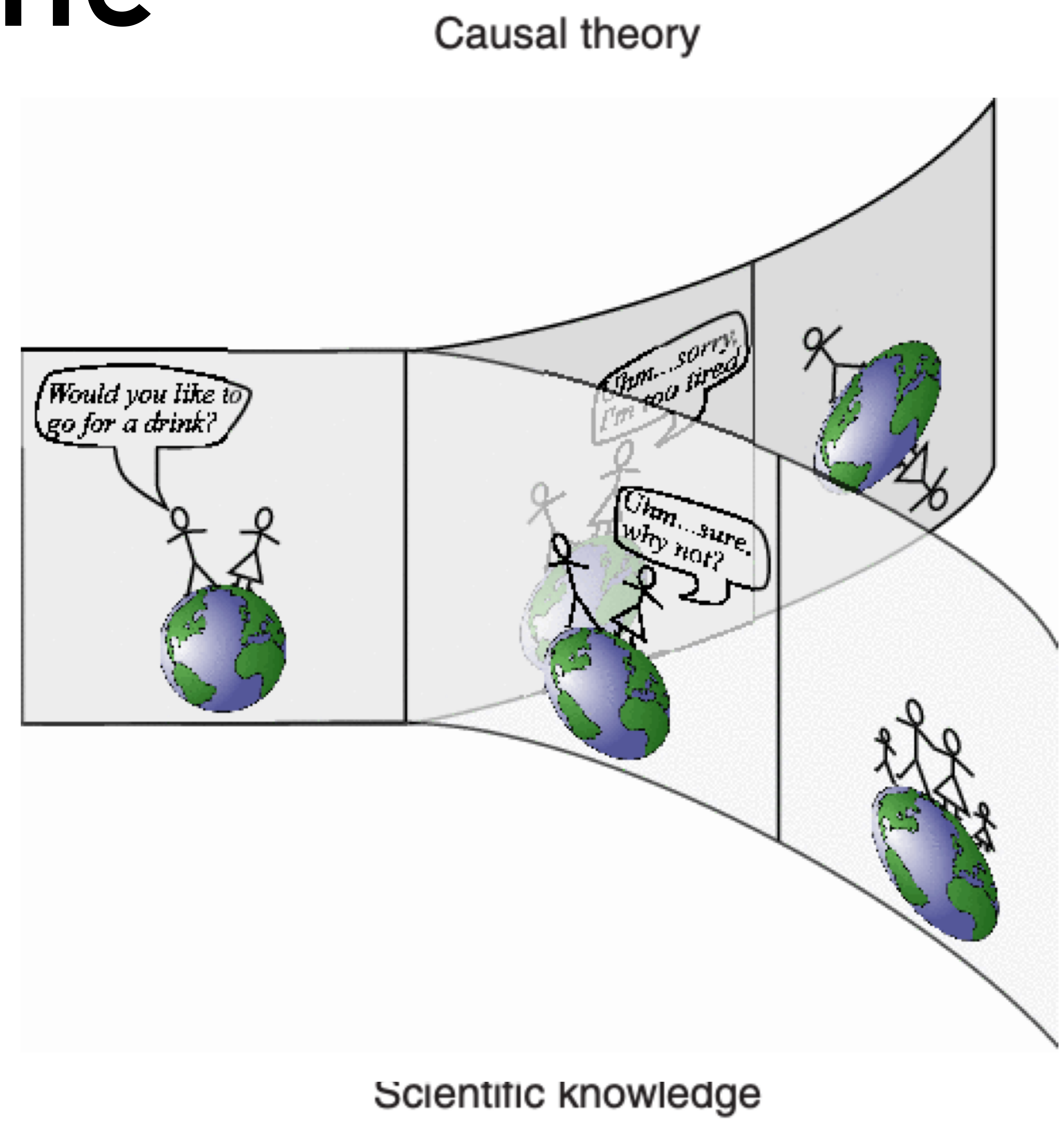
Leonardo Carella

# The Plan for today

- Research Design:
  - Causal theories and hypotheses.
  - Causality in experimental and observational research.
- Coding in R:
  - More on vectors and dataframes (finishing week 1).
  - Importing datasets in R.
  - Describing and creating variables in R.

# Last Time

- **Scientific method**: a way of learning about politics via the test of statements derived from **theory** against empirical evidence.
- **Causality**: “X causes Y if, in absence of X, Y would not have happened”. The problem: we don’t observe the **counterfactual**.



universität wien

# Variables and Observations

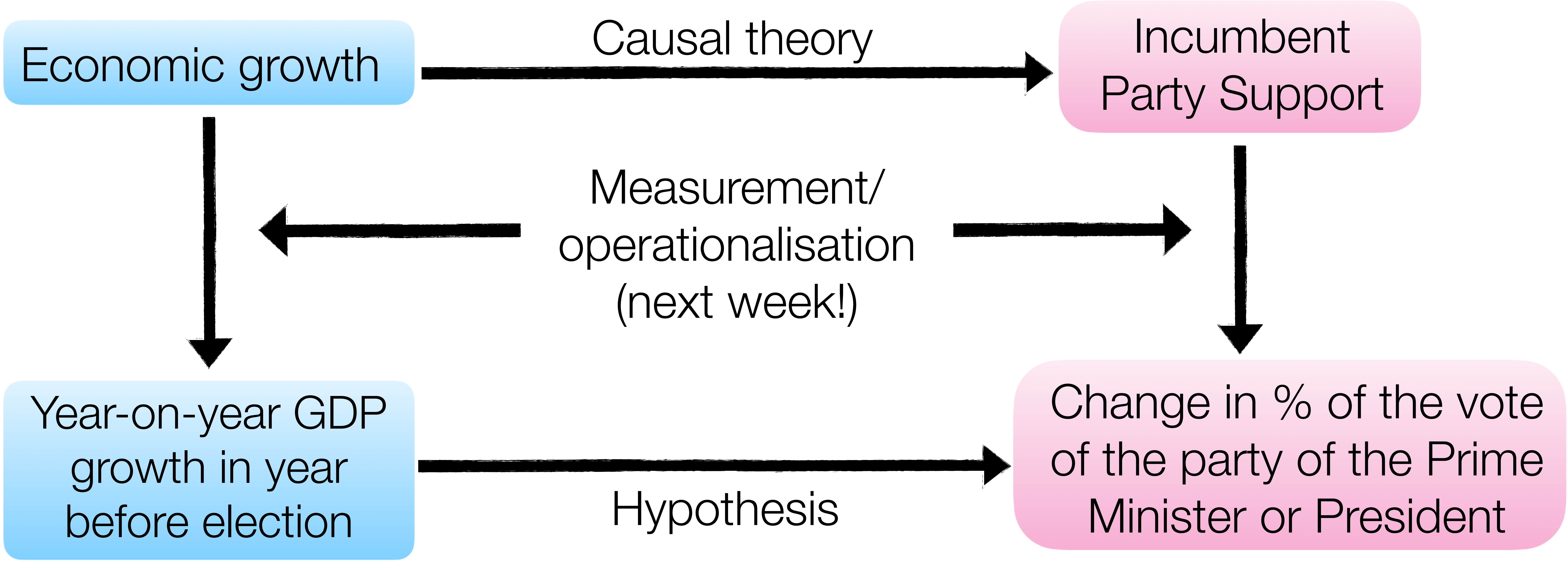
- Political scientists see the world in terms of **variables.**
- Variables have a single **label**, and take multiple **values** for each **observation** (an individual, a country, an election, a town etc.)
  - Label: “Vote Choice”, Values: {SPÖ, FPÖ, ÖVP, ÖVP, Other, SPÖ…}
  - Label: “Unemployment rate”, Values: {4.2%, 6.5%, 5.6%, 7%, 1.8%, 3.5%…}
  - Label: “Presidential System”, Values: {0, 0, 0, 1, 0, 1, 0, 1…}
  - Label: “Support for Gay Marriage”, Values: {Very Favourable, Opposed, Don’t know, Favourable, Very Opposed, Very Favourable, Favourable, Don’t Know}

# Some Causal Theories…

- Economic voting theory (Kramer, 1971)
  - Higher economic growth *causes* higher support for the incumbent.
- Democratic peace theory (Doyle, 1983)
  - Liberal democracy *causes* lower occurrence of interstate war.
- Duverger's Law (Duverger, 1964)
  - Plurality electoral systems *cause* smaller size of the party systems.

- Two types of variable:
- Independent variable: "the cause"
  - a.k.a. "X", "explanatory variable", "predictor","regressor".
- Dependent variable: "the effect".
  - a.k.a. "Y", "outcome variable", "response", "explained variable".

# From Theory to Hypothesis




- Observations: e.g., elections held between 1945 and today in democracies.

# What makes for a good theory?

- Generates directional, probabilistic hypotheses, not deterministic:
  - ***Not great***: “under plurality electoral systems, there can only be two-party systems” [sufficient cause], “two-party systems occur only under plurality systems” [necessary cause]
  - ***Better***: “plurality electoral systems reduce the number of parties” or “plurality electoral systems increase the probability that a country has a two-party system”.

# What makes for a good theory?

- Its variables express “portable” concepts, not proper nouns:
  - ***Not great***: “Trump won the 2024 US Presidential elections because the price of egg was too high”.
  - ***Better***: “higher consumer-goods inflation reduces support for incumbents” or “Experiencing economic hardship increases the probability of voting for a challenger candidate”.

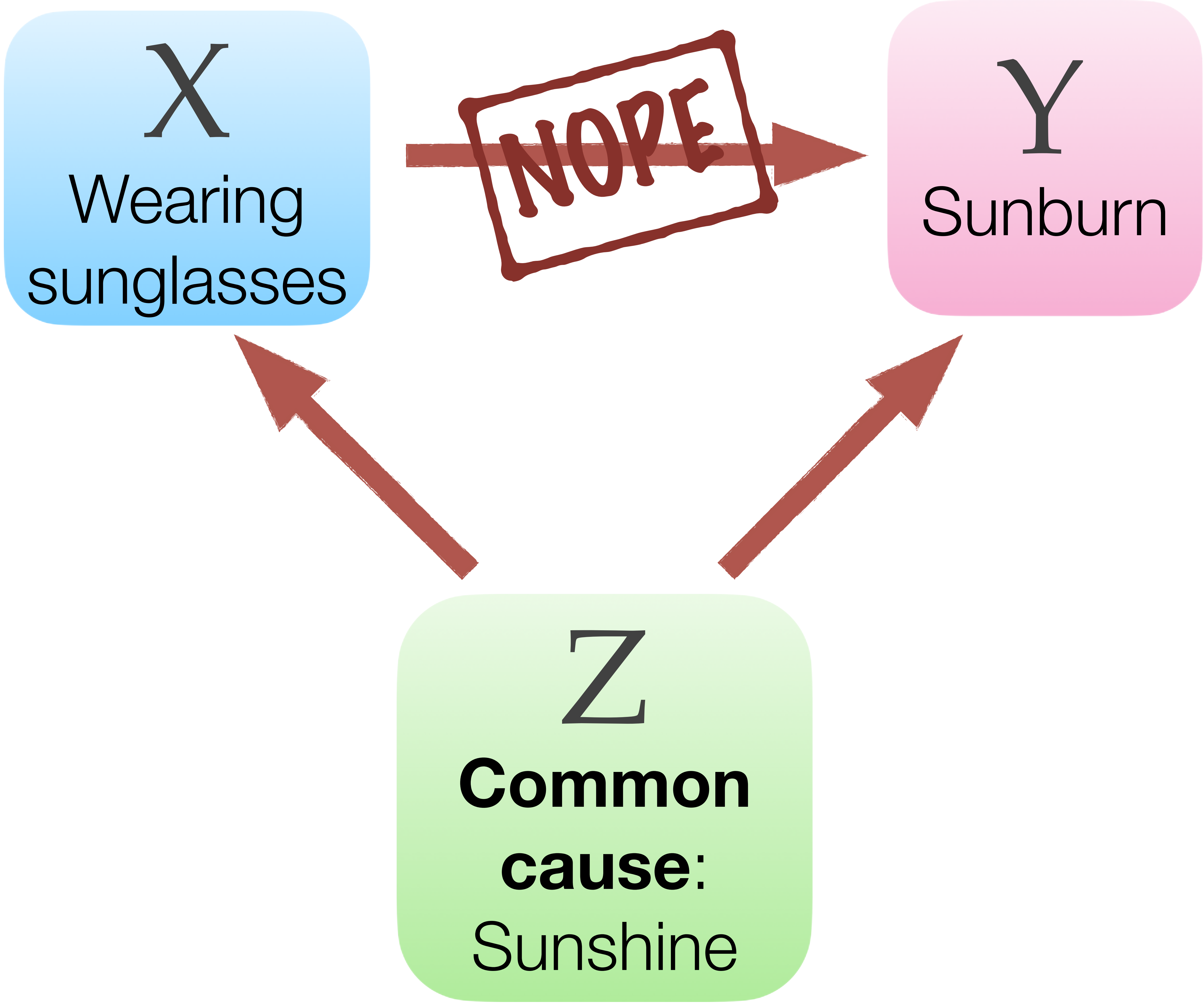
# What makes for a good theory?

- Answers the question “Does X have an effect on Y?”
  - ***Not great***: “why do countries go to war?” Or “how do wars start?”
  - ***Better***: “does having liberal democratic institutions reduce the probability that countries engage in interstate war?”

# Assessing Causality

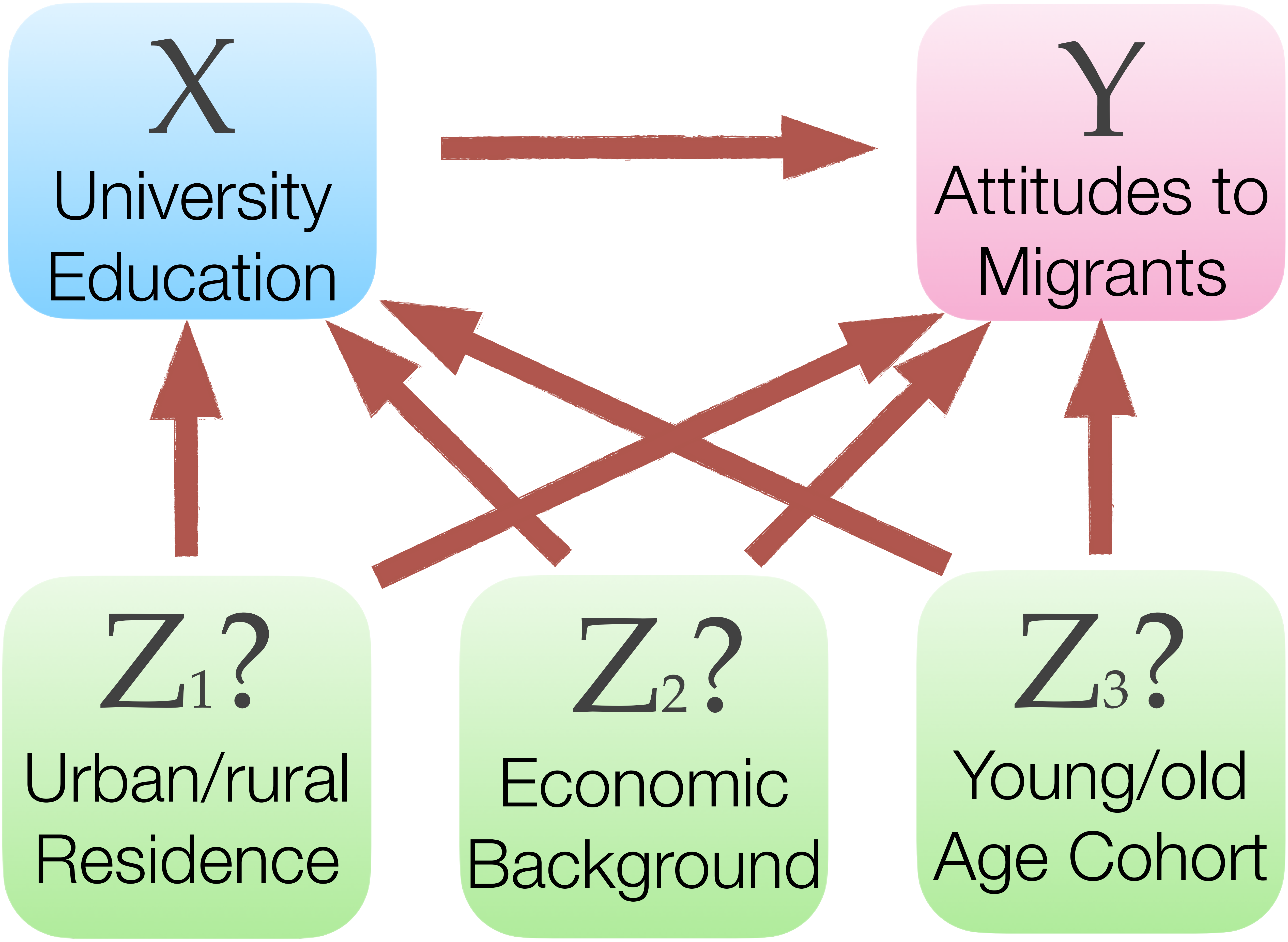
- How can we say that the independent variable X has a causal effect on the dependent variable Y? Three requirements:
  - ***Time order***: X must precede Y.
  - ***Covariation***: changes in X are associated with changes in Y. Much of the statistics part of this course is about how to estimate this.
  - ***Unconfoundedness***: the relationship between X and Y is not driven by other variables that have causal effects on both X and Y. (*This is the big, bad monster*.)
- Note: in your reading, Kellstedt and Whitten (2018) also add a fourth requirement: "credible causal mechanism". Strictly speaking, it's not a requirement, but making claims that intuitively make sense certainly helps!

# Confounders: Intuition



# Confounders are everywhere

- A confounder (or confounding variable) is a variable that affects both the dependent variable (Y) and the independent variable (X).



# Confounders and causation

- Ideally, we would compare the immigration attitudes of the *same* people in a world where they graduated from university, and in a counterfactual world where they did not.
- In reality, often all we can do is comparing *different* people, some who are graduates and some of whom are not.
- But these people are *different* for all sorts of reasons! So the “non-graduate” group is *not* a good comparison for the “graduate” group.
- How can we find a good comparison group?

# Experiments

- Core advantage of experiments: we can **randomise** the value of our independent variable X across groups.
- Because X is **randomly** assigned by the researcher to some observations and not to other, the difference in X across the two groups cannot be "caused" by something else.
- In this sense, experiments are the **gold standard** for evaluating causal theories. (Arguably.)

# Experiments

- For instance, Bertrand and Mullainathan (2004) wanted to find out whether employers discriminate based on applicant race. They sent fictional CVs to 1300 real help-wanted ads in Boston and Chicago newspapers.
- Random assignment of X: some CVs have typically Black names (Lakisha, Jamal), others have typically white names (Emily, Greg, etc.). The CVs are otherwise ***identical*** (i.e. there are no confounders).
- Measurement of the dependent variable Y = call-back rates from the employer.

| | *Call-Back Rate for White Names* | *Call-Back Rate for African American Names* |
|---|---|---|
| Sample: | | |
| All sent resumes | **10.06%**<br>[2445] | **6.70%**<br>[2445] |

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American economic review, 94(4), 991-1013.

# Observational Studies

- Experiments are great, but they not always feasible (for logistic or ethical reasons) or realistic with respect to the real world (how would you randomly assign “democracy” or “economic growth”?).
- Often, we have to work with messy real-world (**observational**) data, ridden with pesky confounding variables.

# Observational Studies (1)

- 'Old school' approach to observational studies: cross-sectional comparisons with **statistical controls** (last part of this course).
- The idea: comparing the value of Y across observations that take different values on the variable X but are similar with respect to the values of the observed (potential) confounders $Z_1, Z_2, Z_3$.
- (One of) the problem(s): **unobserved** confounders.

# Observational Studies (2)

- Modern causal inference:
  - ***Differences-in-differences***: compare observations at different points in time, *before and after* some change in X, with observations at the same points in time, but where X remains constant.
  - ***Near (or quasi-) experiments***: real-world cases where X can be assumed to have been assigned ‘as if random’.
- ***Not*** the focus of this course, but we will come back to these ideas.

# Summing up

- Political scientists think of the world in terms of variables and causal relationships between variables.
- Good causal theories are generally probabilistic and directional, answering questions like “does X have an effect on Y?”
- Causality requires (1) time order, (2) co-variation, and (3) unconfoundedness (i.e. no confounding variables).
- Because they allow researchers to assign X at random, experiment are the gold standard for dealing with causality, but not always feasible.
- Now, let’s open RStudio…!