

Research design: The Basics

Andrew Roberts

Goals of this class

- Acquaint students with research design and methodology
- Basics and more advanced techniques
- Emphasis on practical issues to help with dissertations

Structure

- 3 meetings including today
 - Others include (i) regression, (ii) experiments, (iii) qualitative methods, (iv) causal inference
- 3 short exercises
- I will also be glad to review methods sections of your dissertations

Tell me about yourself

- Where are you from and what should I visit in your area?
- What are your academic interests/ professional plans?
- Give us a cultural recommendation that you love but we may not be aware of: book, film, TV, album...

About me

- New Jersey, USA. Visit the Grounds for Sculpture in Hamilton.
- Projects on billionaires in politics, how countries view each other, women in politics
- Book: Lee, Pachinko
- TV: Borgen
- Album: Santigold



Topics

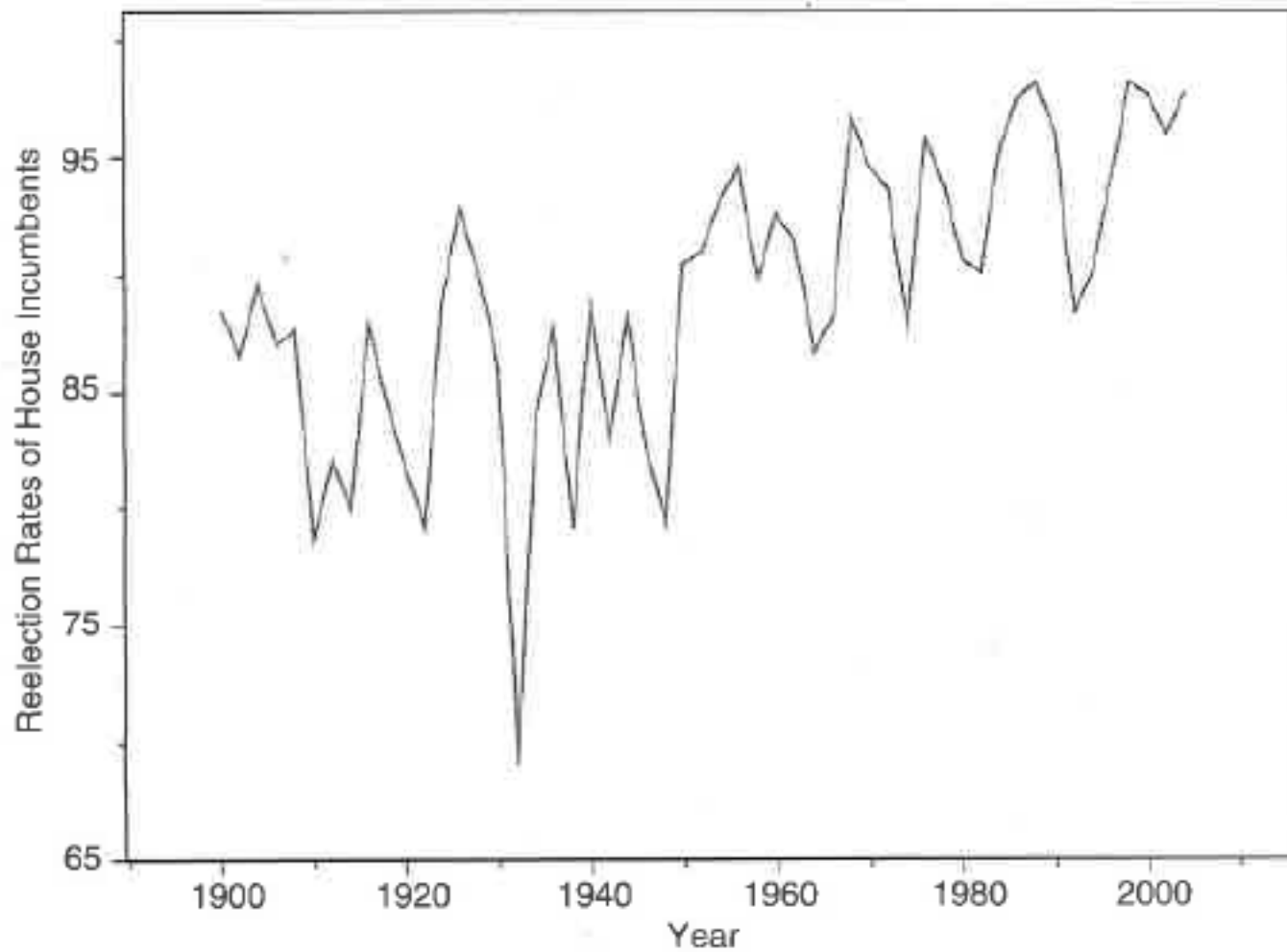
- **What are political scientists trying to do?**
- Fundamental problem of causal inference
- KKV template
- Ways of framing a research question
- Inductive and deductive research
- Quantitative versus qualitative research
- Practical advice

1. What are political scientists trying to do?

Social science versus journalism

- Quick journalism: Factoids
 - 95% of representatives reelected
- Slow journalism: systematic gathering of data
 - Reelection varies but increasing over time
- Social science: causal mechanisms which explain patterns – causes & effects
 - Economy better managed
 - Growth of government: MPs can help constituents

Reelection rates in Congress



Scientific method

- Question or puzzle: Why did that happen?
- Theory or model: outcome as result of some process
- Implications (hypotheses): what should we see if our theory is true
- Test: are implications consistent with evidence; try hard to falsify

The goal of social science

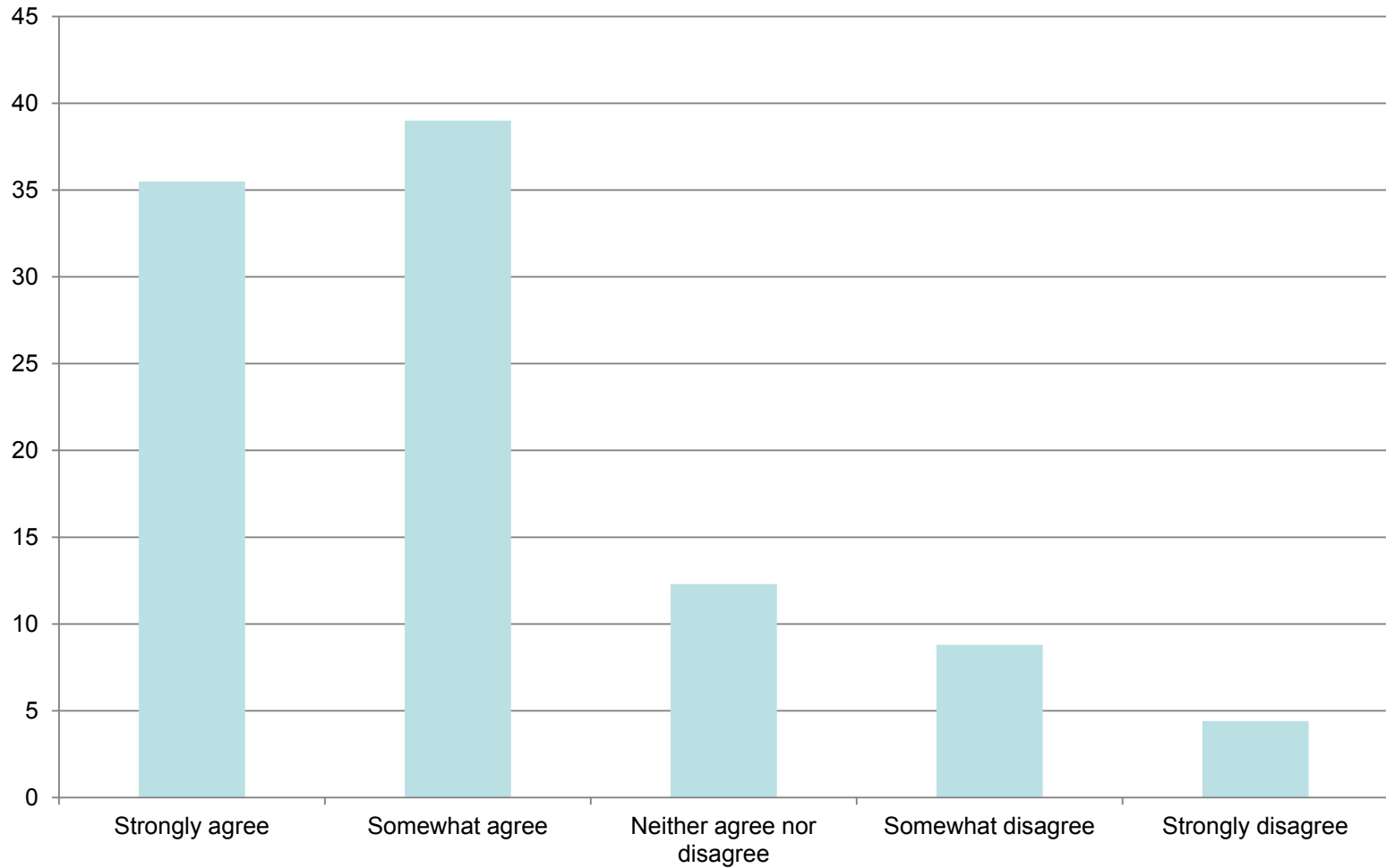
- The ultimate goal is causal inference
 - Cause and effect relationships
- Both qualitative and quantitative researchers agree
 - Main dissent is from interpretivists
 - Also political theorists do something different
- Two elements in most projects
 - Outcome/effect: dependent variable
 - Cause(s): independent variable(s)

This is it

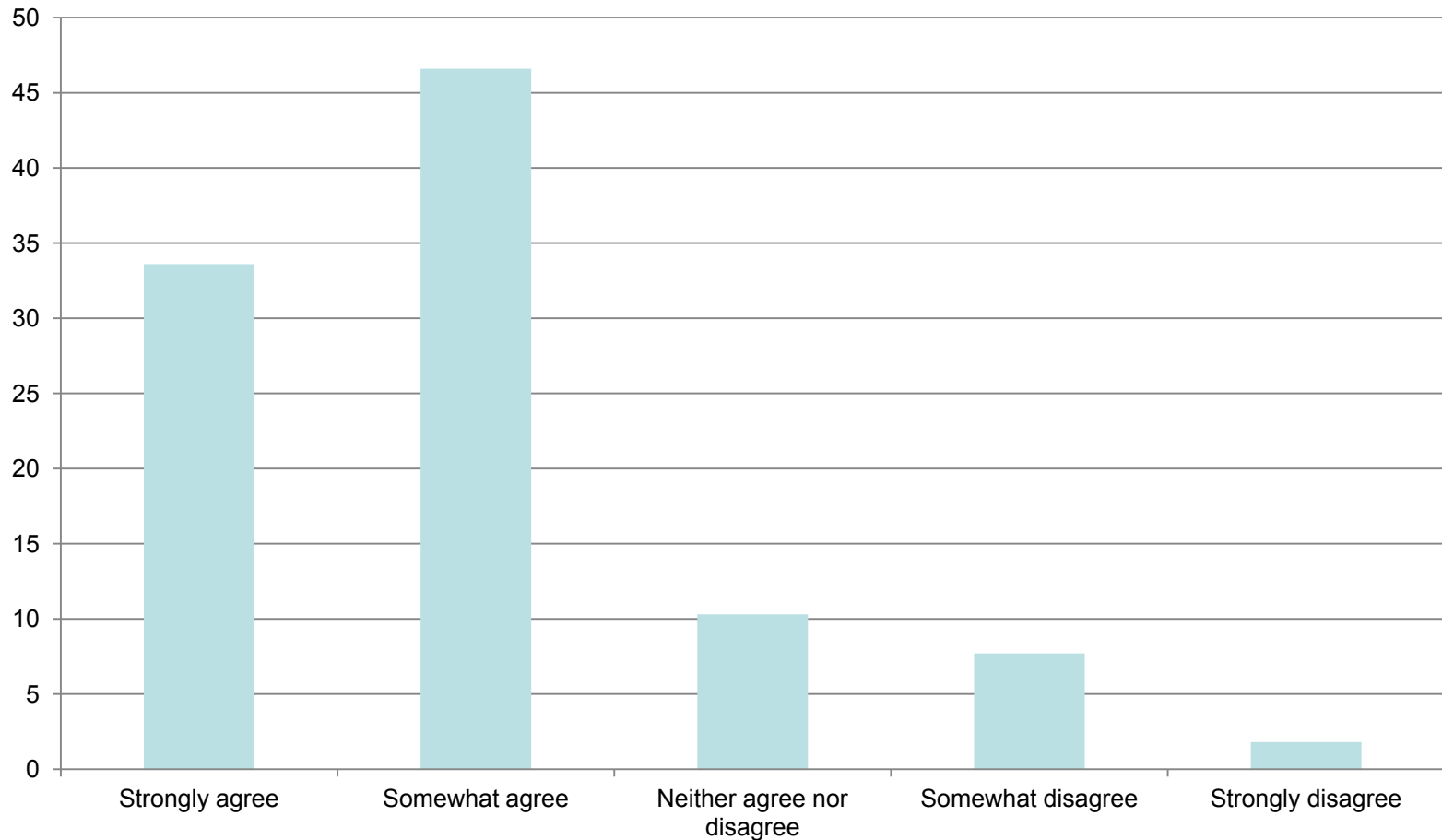
X **→** **Y**

Note: Of course, also much more complicated versions.

Establishing causal relations is the main goal of research in politics



Political scientists should aim to develop broad, general propositions about politics



Why causality?

- Understanding a phenomena means knowing what causes it (or its effects)
- Causality is essential if we want to make interventions to save the world



But: very hard to show causality

- Many studies do something else
 - Description: finding trends & patterns
 - Exploratory study to generate new hypotheses
 - Creation of new concepts or typologies
 - Descriptive inference: generalizing from sample to larger universe (surveys)
 - Explore mechanisms of causal processes
- But keep causality in mind

2. Fundamental problem of causal inference

The philosophy of causality

- Causality related to counterfactual
 - To say that X causes Y means that if we take away X, we don't get Y
 - “Not X” is the counterfactual
 - Causal effect of a drug = health (with drug) – health (without drug)

Fundamental problem

- We can't observe both X and not X at the same time (even in experiment)
- You either get the drug or you don't
- We never really know causal effect

Three solutions

- Experiments
 - Control group is counterfactual
- Comparisons with observational data: countries or regions or groups or individuals
 - Some are treatment, some are controls
- Time series: a case changes over time due to some event
 - Compare before and after an event
 - Event is treatment that might cause change

Experiment as gold standard

- Assign subjects randomly to two groups
 - One group receives treatment, other doesn't
- Any difference in outcomes should be caused by treatment
- Types
 - Lab experiments
 - Survey experiments
 - Field experiments
 - Natural experiments



Observational data

- Gather data from the real world
- Try to determine if outcomes systematically related to key independent variable
 - Eg, do more democratic countries tend to be richer than less democratic ones
 - Eg, does external intervention lead to democracy
- But lots of difficulties in showing causality
 - Main ones: other possible causes & reverse causality

Internal versus external validity

- Internal validity: extent to which we find a true causal relation in our cases
 - Typically high in experiments because of randomization
 - Lower with observational data
- External validity: extent to which our finding generalizes to other situations
 - Worry that low in experiments because situation is artificial and doesn't apply to real world
 - Higher with observational data

3. KKV Template

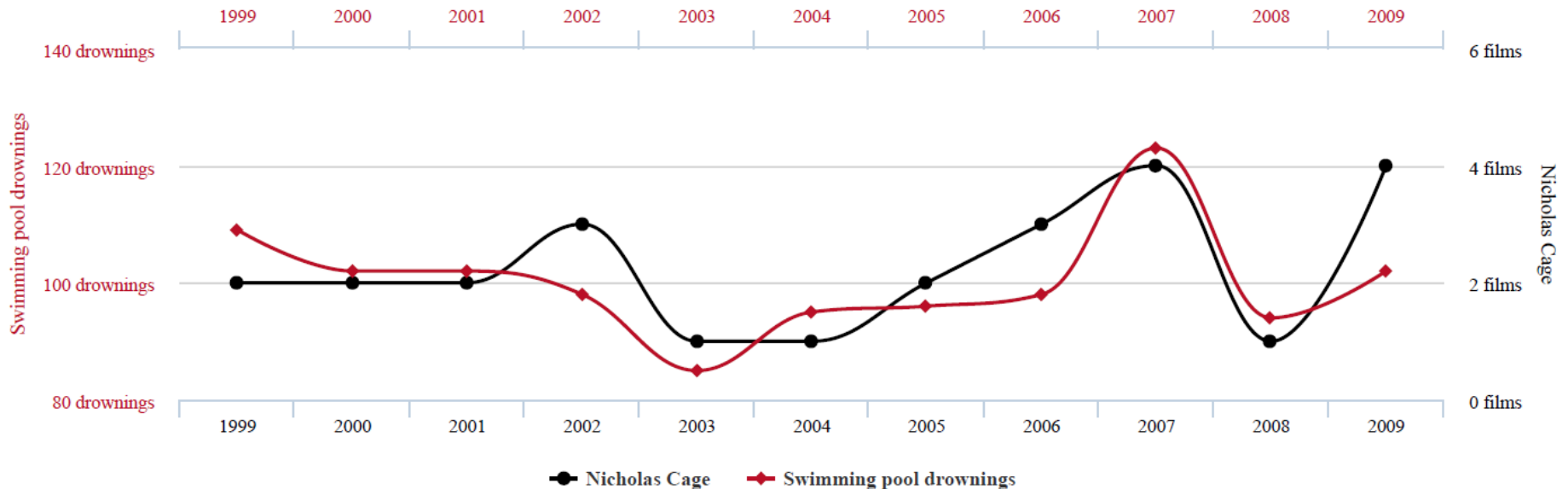
Where to start?

- Let's say we find a relationship or correlation between two variables, X & Y
 - Eg, economic development & democracy
- Does it mean that one causes the other?
- Maybe, but correlation doesn't equal causation
- What can go wrong?

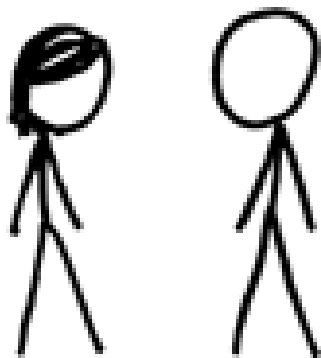
Nic Cage causes drowning?

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in

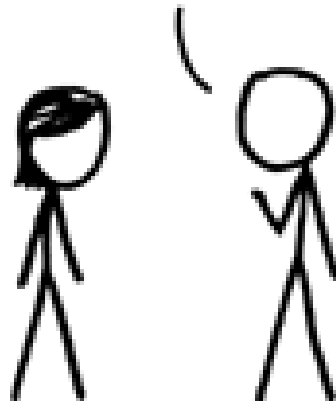
Correlation: 66.6% ($r=0.666004$)



I USED TO THINK
CORRELATION IMPLIED
CAUSATION.

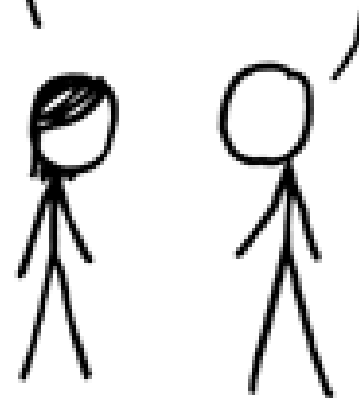


THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



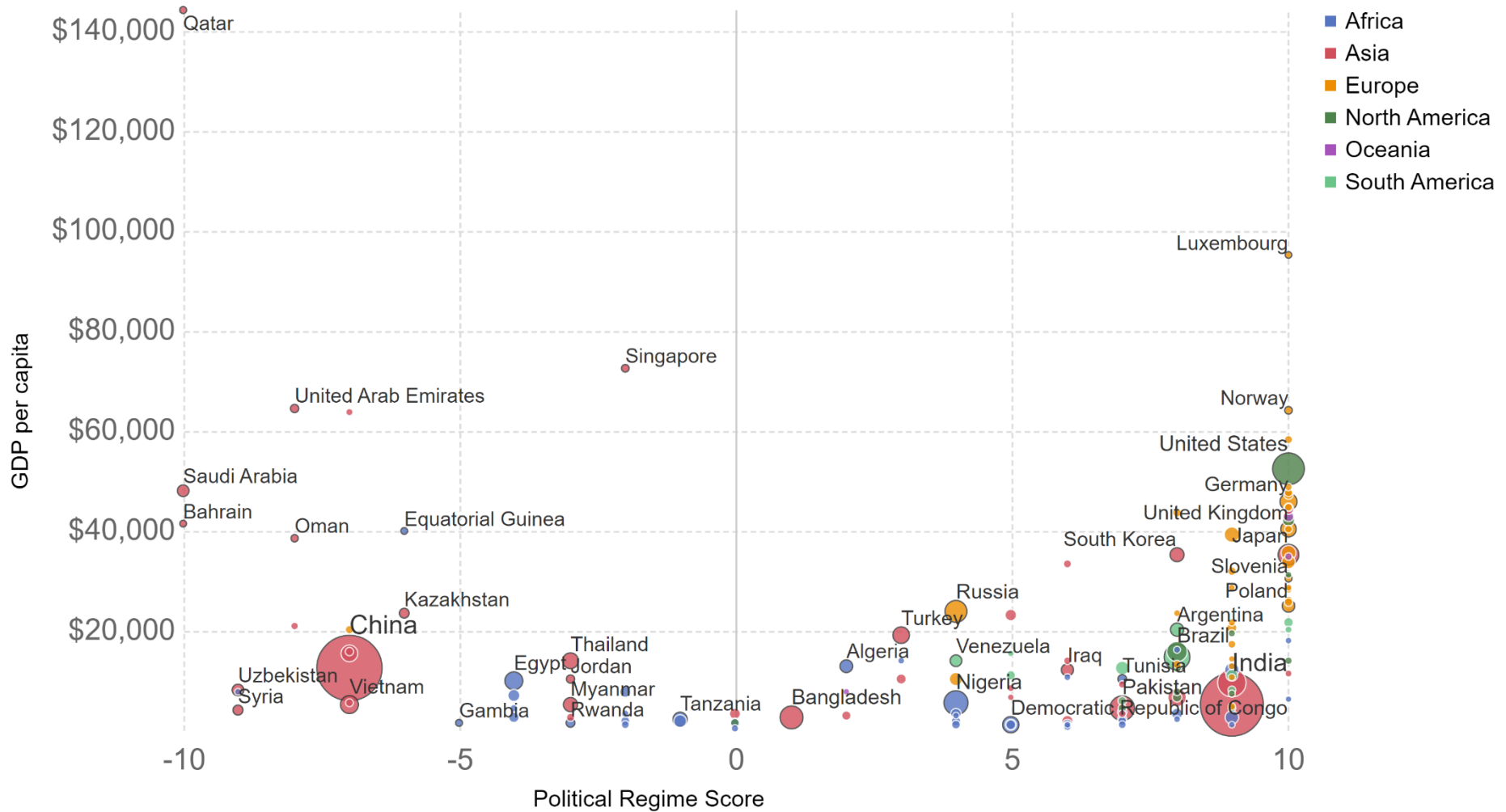
SOUNDS LIKE THE
CLASS HELPED.

WELL, MAYBE.



GDP per capita vs type of political regime, 2014

Political regime are classified on a range from -10 (full autocracy) to +10 (full democracy). GDP per capita is adjusted for price differences between countries to allow comparisons.



Source: GDP per capita, Political Regime (OWID based on Polity IV and Wimmer & Min), Population by country, 1800 to 2100 (Gapminder & UN)

OurWorldInData.org/democracy/ • CC BY

Potential explanations for correlation

All the reasons

From @tstumley



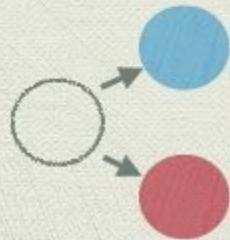
Chance



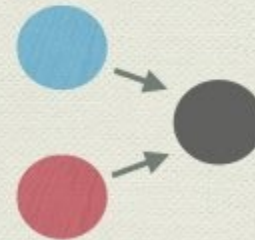
Causation



Reverse causation



Confounding
(including by time)



Selection

Omitted variables

- Is there another variable Z that causes both X & Y ?
 - What factors might cause both democracy & development?
- Solutions
 - Try to control for all of these variables – include them in your statistical or qualitative model

Endogeneity or reverse causality

- Is it possible that Y causes X?
 - Does democracy cause development or does development cause democracy?
- Solutions
 - Timing: X occurs before Y (but maybe previous value of Y causes X)
 - Identify situations where X is exogenous
 - Instrumental variables to create an exogenous X

Selection bias

- Make sure that your cases include variation on the dependent variable
 - Both successes and failures
 - Try not to pick cases according to outcomes
- If only successes, then can't identify causes
 - Can't study causes of democracy by only looking at what democracies have in common
 - You need to find what distinguishes democracies from non-democracies

OVER 15 MILLION SOLD

THE 7 HABITS OF HIGHLY EFFECTIVE PEOPLE

Powerful Lessons
in Personal Change

With a New
Foreword and
Afterword
by the Author

"A wonderful book that could change your life."
—Tom Peters, bestselling author of *In Search of Excellence*

Stephen R. Covey

THE 7 HABITS OF
HIGHLY EFFECTIVE PEOPLE
Powerful Lessons in Personal Change

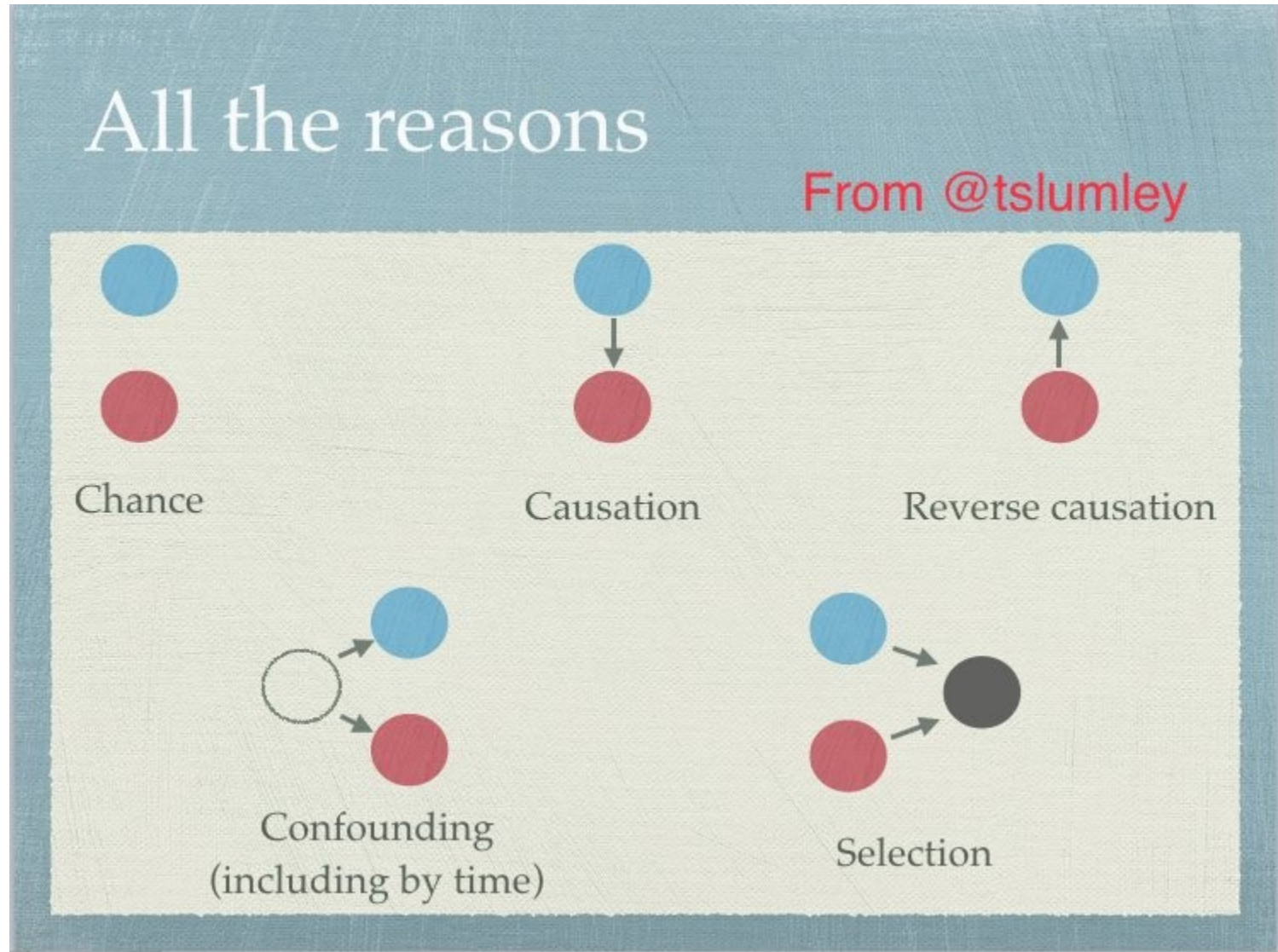
Stephen R.
Covey



Conditioning on a collider

- Cases are selected due to independent variables
 - Eg, there is no correlation between height & success in the NBA
 - NBA teams choose players based on height & other factors like talent
 - If we focus on entire population, we see that height increases basketball success
- Be careful of the selection rule for a sample

Potential explanations for correlation



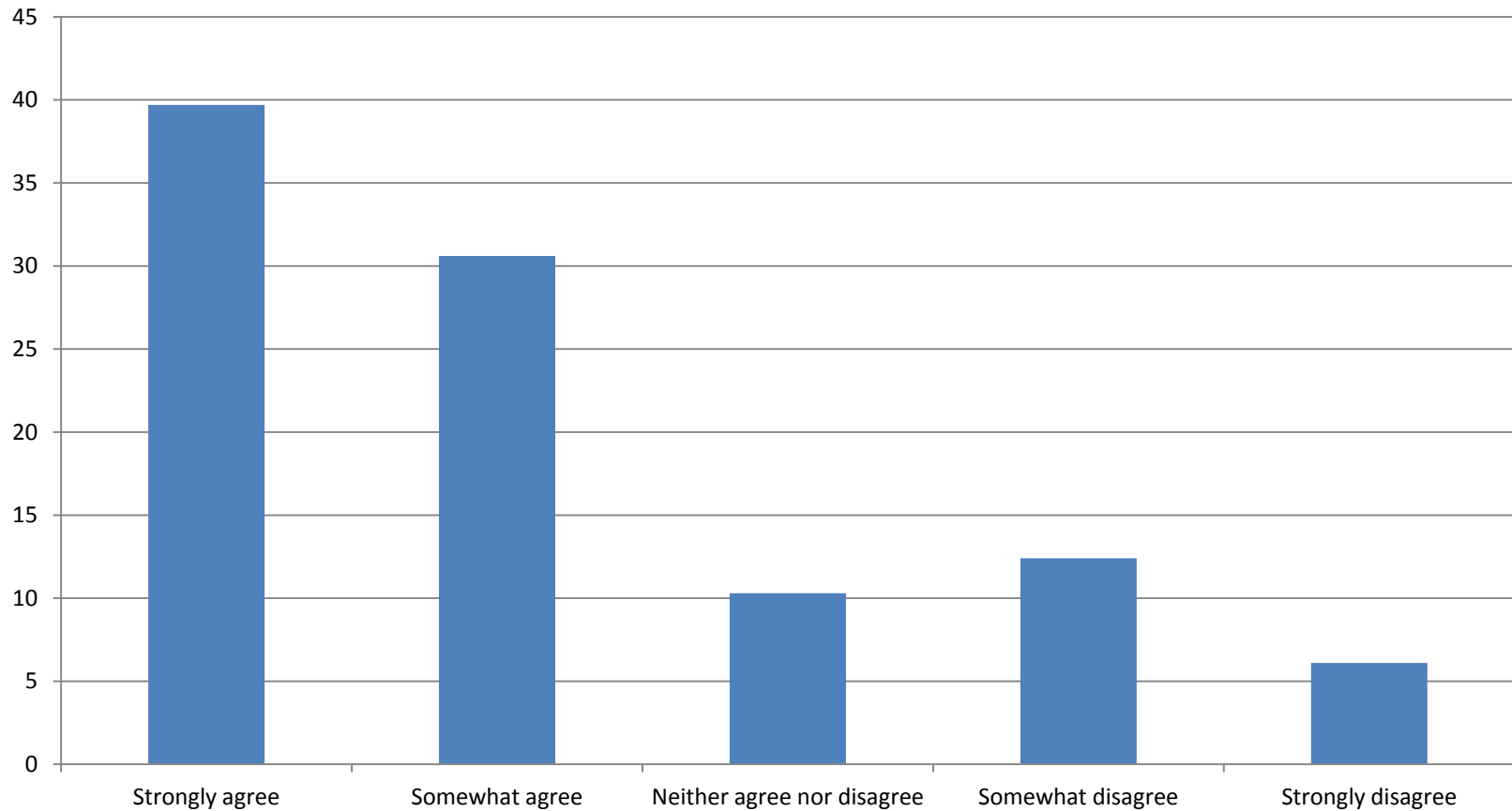
Mechanisms

- Also useful to have evidence of a mechanism in order to demonstrate causality
- How exactly does X cause Y?
 - What are intermediate steps?
 - Do we have evidence for them as well?
- Relationship between economic development & democracy work through education, middle class, etc.

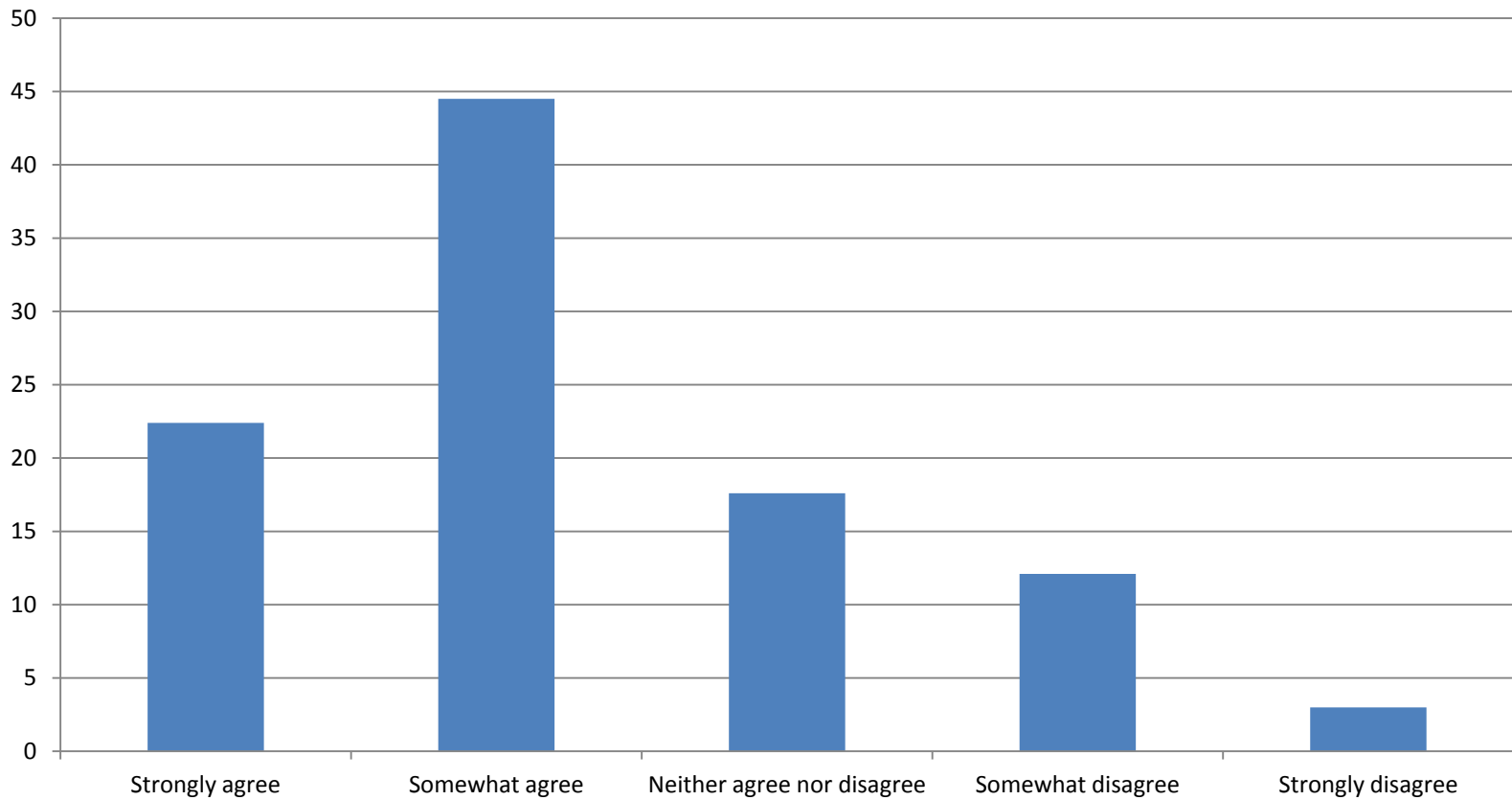
Does development cause democracy?

- Omitted variables: culture, geography, etc.
- Endogeneity: can democracy cause development?
- Selection: are we looking at all countries & all time periods?
- Mechanisms: how exactly does development cause democracy?
- Measurement: how do we measure democracy & development

Empirical research interested in making causal inferences must have variation on their dependent variable



Researchers should usually try to increase their number of observations to achieve better inferences



The big distinction: Experimental versus observational studies

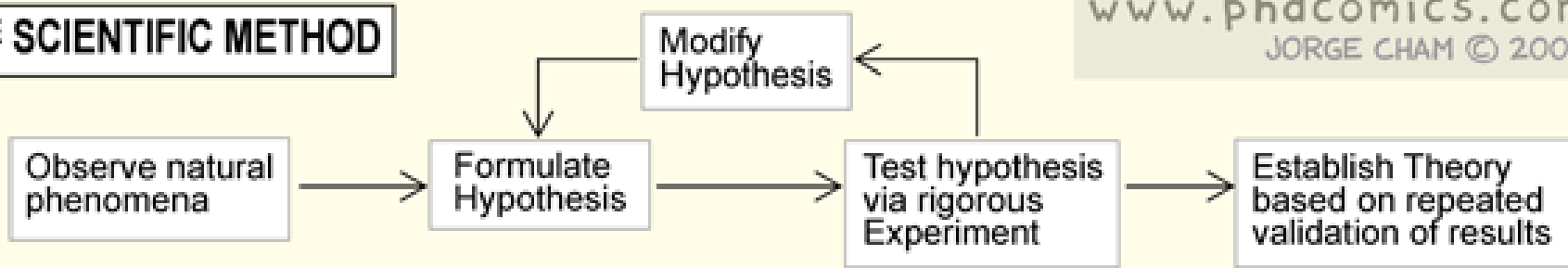
- Experiment usually solves these problems
- But randomization uncommon in real world
 - 1st term of Czech Senate, reserved seats for women in India
- Without randomization hard to show causality
- Two options
 - Search for natural experiments
 - Do the best you can 😊

Even more worries

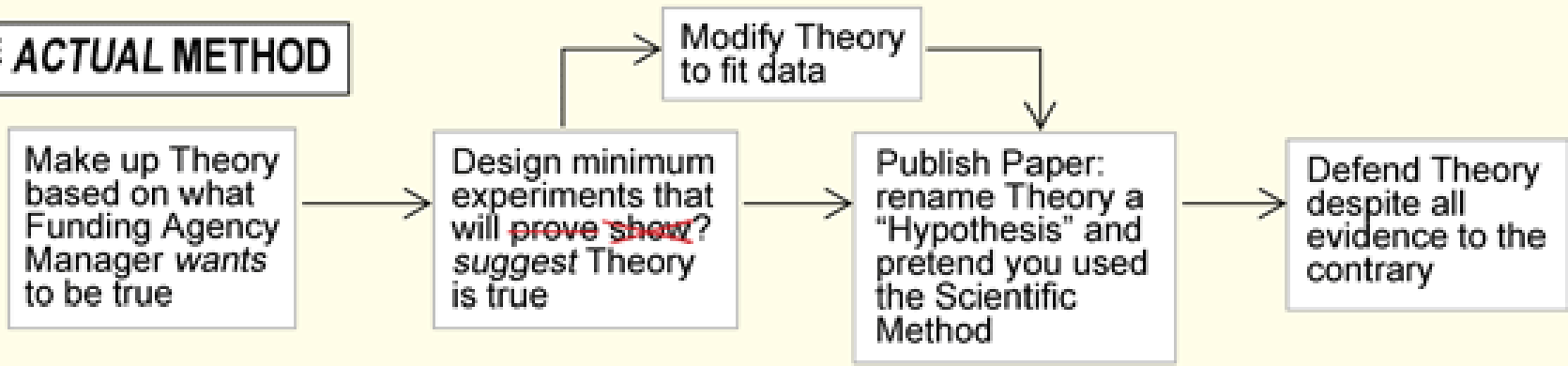
Key assumptions usually aren't met in observational studies

- Independence of observations
 - Is each observation separate? (Eg, time-series)
- Conditional independence of assignment and outcome
 - Is assignment of cases to treatment and control unrelated to other factors that affect outcome?
- Causal homogeneity
 - Does politics work the same in CZ and US?

THE SCIENTIFIC METHOD



THE ACTUAL METHOD



4. Ways of framing a research question

Two sorts of causal questions

- Causes of effects – backwards causality
 - Focus on outcome ($\Rightarrow Y$) and work backwards
- Effects of causes – forwards causality
 - Focus on cause ($X \Rightarrow Y$) and work forwards

Causes of effects

- Examples
 - Why do some people die from Covid?
 - Why are some countries democratic and others not?
- Advantages
 - Most important questions in social science
 - Good for identifying new hypotheses
- Disadvantages
 - Very hard to show causality
 - Hard to eliminate other potential causes
 - Big selection biases & endogeneity

Effects of causes

- Examples
 - Does this vaccine prevent Covid?
 - Do ethnic divisions hinder democracy?
- Advantages
 - Best way to get reliable inferences
 - Approximates experimental design
- Disadvantages
 - Questions are limited



A useful exercise

- What experiment would you conduct to answer your question if you were all-powerful?
 - What is your treatment?
 - How would you assign the treatment?
- This often helps you to figure out how to conduct research in the real world
 - Is there a natural experiment that approximates this hypothetical experiment?

Many conflicts due to different ?s

- A forward causality project: does tutoring improve educational outcomes?
 - Reverse causal objections: you are ignoring inequality, race and gender; does tutoring matter more than these things?
- A backwards causality project: why are some students more successful than others?
 - A forward perspective objection: how do you know that these factors are really causal?

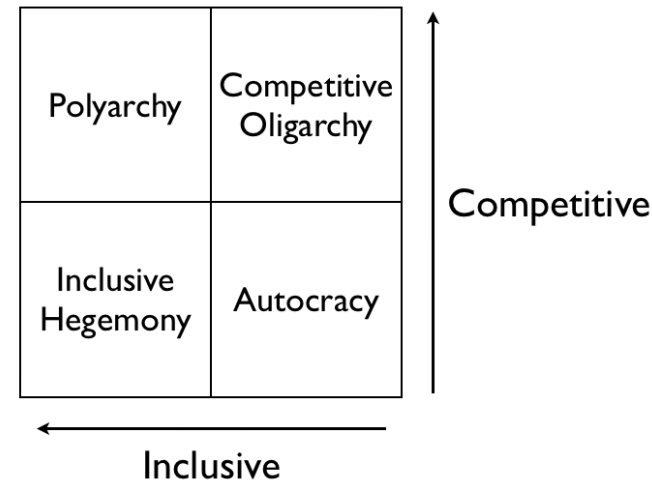
Causality is not all there is

- Systematic description
 - Create new data
 - Uncover new patterns and trends
- Lots of areas where we know very little
- Examples from my own work:
 - What do millionaires think about politics?
 - How much do billionaires participate in politics?
 - Do women have different preferences than men?
 - Are public preferences stable over time?

Also concept formation

- See Goertz, Social Science Concepts & Collier et al., Putting Typologies to Work
 - More on this next time
- Often 2X2 tables but more elaborate ways
 - Over-arching concept
 - Row & column vars

Dahl's Polyarchy



5. Inductive & deductive

Deductive approach

- Create a model of politics
- Often start with assumptions about human nature plus set of constraints
 - Goal-oriented behavior: office, power, ideology
 - Instrumental rationality: actors do the best they can given information
- Then logically reason to outcome

Voter turnout

- Voters weigh costs and benefits when deciding whether to vote
 - Costs of voting = time, energy, money
 - Benefits of voting = benefit if your party wins (B) * probability that your vote matters (p)
- Usually $C > p * B$ (because p very small)
- Implications
 - Most people won't vote
 - Registration rules increase C => less likely to vote
 - Closeness of election increases p => more likely

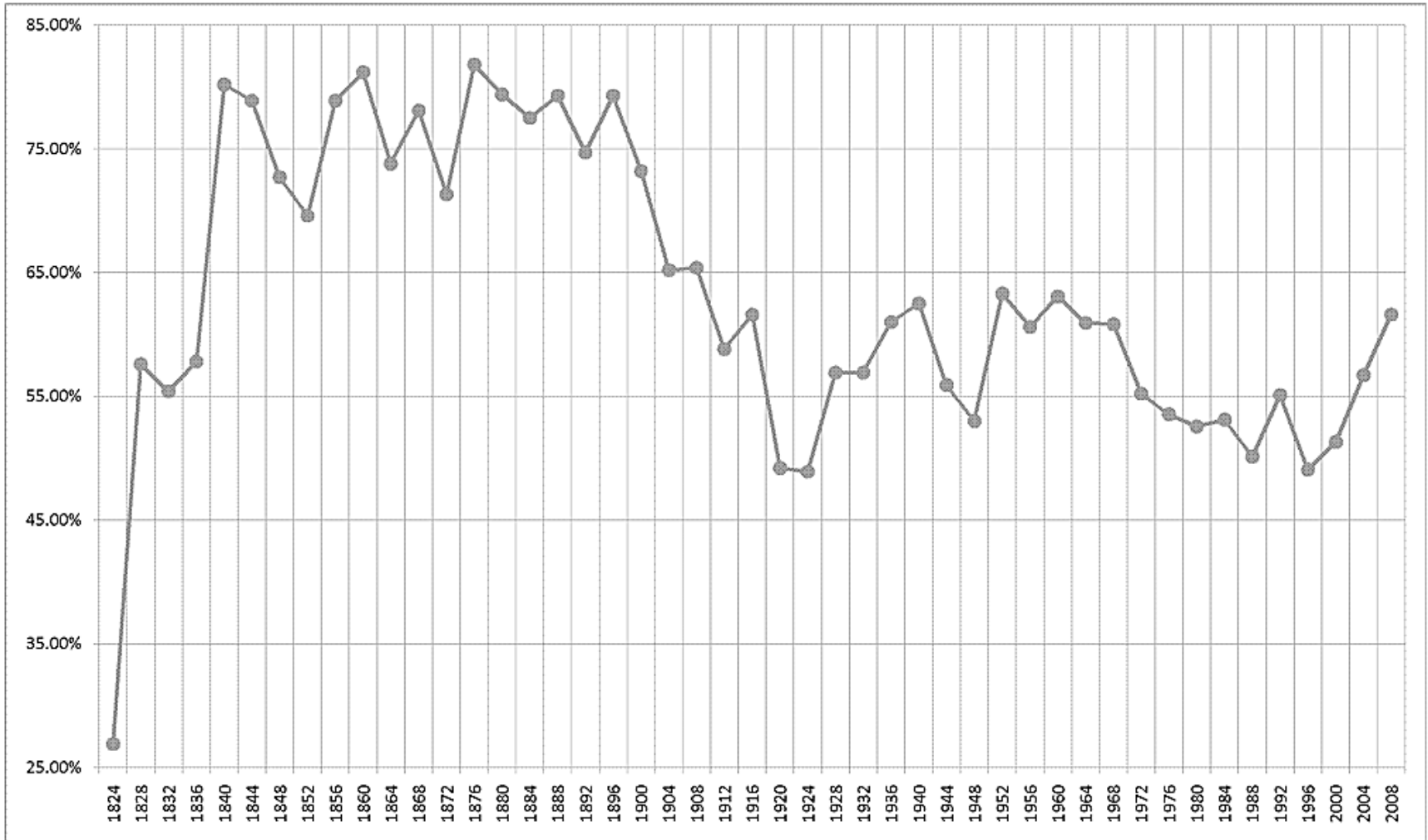
Criticism of deductive models

- Unrealistic, over-simplification of reality
 - Better just to look at facts
 - But what are the facts? Which ones?
 - How do the facts relate to each other?
- But models a useful disciplining force
 - Makes assumptions very clear
 - Shows logic of connections
 - Clear what is missing

Inductive approach

- Search for patterns across independent and dependent variables
 - Different countries, time periods, groups
- Which independent variables related to dependent variable
- But is this relationship enough?

Consider election turnout

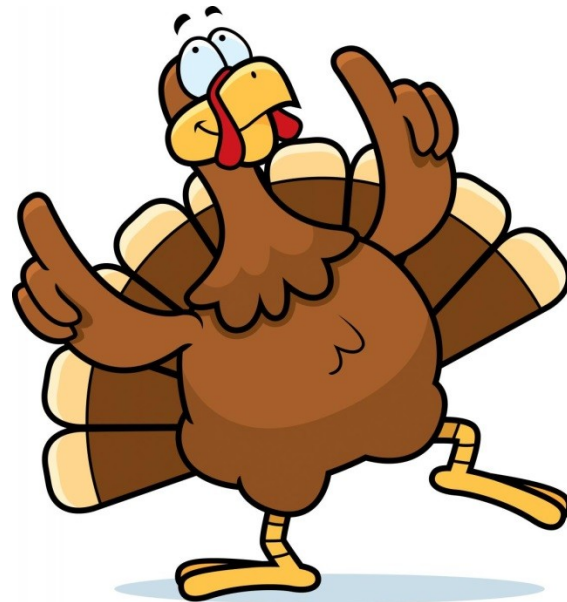


Potential causes

Election	Turnout	Closeness of election	Importance of election	Other things?
US – 2000	50%	Hi	No	
US – 2004	56%	Med	Yes	
US – 2008	57%	Lo	Yes	
US - 2012	55%	Lo	No	

Inductivist turkey

- How to know whether association means causality?
- Which variables to include?
 - Which ones are important?
- Never enough cases
- Lacking a theory



Usually go together

- Need a theory to prevent inductivist turkey
 - Not just machine learning
- Need data to keep theory tied to reality
 - People vote more than theory predicts
 - Add a civic duty term: $p * B - C + D$
- Often go back and forth
- Kant: percepts without concepts are blind, concepts without percepts are empty

6. Quantitative versus Qualitative

Two cultures?

- Do qualitative and quantitative methods have different logics?
 - KKV: single logic for all inference
 - Mahoney and Goertz: two different logics

Approaches and concepts

Quantitative	Qualitative
Large # of cases Usually continuous measures	Small # of cases Often ordinal or cardinal measures
Goal = estimate size of average effects of independent variables on dependent variable, often effects of causes	Goal = fully explain causes of individual cases, typically causes of effects
Probabilistic conception of cause	<ul style="list-style-type: none">• Deterministic conception of cause: necessary and sufficient conditions• Equifinality: multiple discrete causal paths

More qualitative specificities

- Case selection focused on specific, positive cases
 - Negative cases don't tell us as much
 - Some cases more important than others (eg, WWII)
- Each case should be explained correctly
 - Because causation is deterministic
- Few cases => avoid causal heterogeneity
- Sequences and pathways are important
- More attention to concepts, revision of concepts

Data set (DSO) versus causal process (CSO) observations

- DSO = Excel spreadsheet
- CSO = insight or piece of data that provides information about context, process, or mechanism
 - Eg, Nuclear taboo: documentary evidence that leaders thought about taboo
 - Eg, Why do Latin American leaders switch to neoliberalism? Evidence that they spoke with international investors and got scared

Issues with qualitative template

- Is social world deterministic or probabilistic?
 - How many phenomena where cause is necessary or sufficient for outcome?
 - “No bourgeoisie, no democracy”
- Often simply historical or descriptive
- Can it predict?
 - Small scope conditions mean that inferences don't apply outside of particular cases
- Is it useful for making interventions?
 - Can it tell us what we should do?

7. Practical Advice

What makes a good project?

- An interesting empirical puzzle
 - Differences in outcomes: why here, but not there
 - Or why did things change: why then, but not now
- Strong research design
 - Natural experiment
 - Great answer to less important question versus weaker answer to important question
- New data
- Policy & current events relevance

Be clear about your novelty

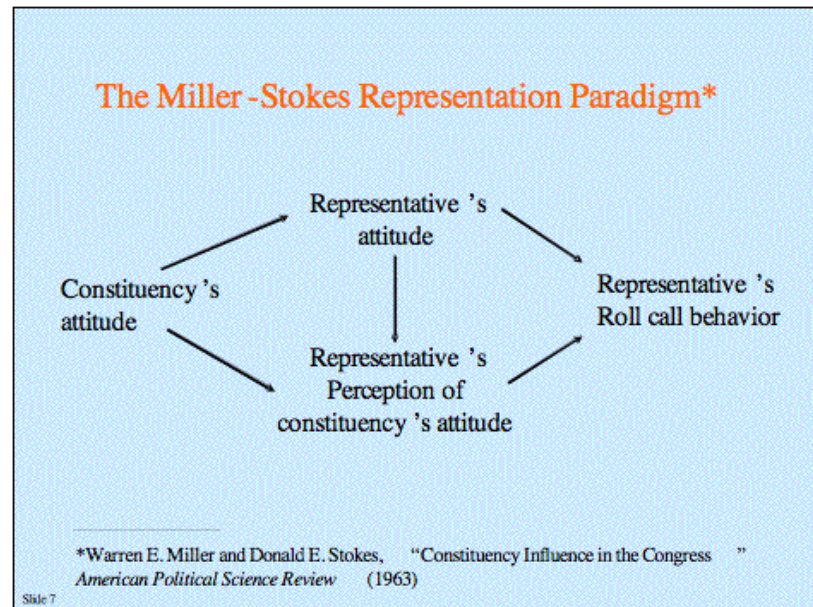
- New question: nobody has investigated this dependent variable before
- New theory: nobody has considered this hypothesis, independent variable before
- New method: this method hasn't been used and is better than existing methods
- New data: I have created or gathered new data (eg, new country, new time period)

Construct falsifiable theories

- Theory should have as many observable implications as possible
- Hypotheses should be as concrete and specific as possible
 - Best to list hypotheses: H1, H2, ...
- Make sure to consider other possible explanations in addition to your preferred one
 - Try to create real contenders rather than “straw men”

Arrow diagrams

- When dealing with multiple variables, important to draw arrow diagrams
- Few readers can follow a causal chain without a graphical picture



Create new data!

- A new data set means an automatic contribution
 - Simple descriptions of new patterns
- Perfect for dissertation: requires lots of time and elbow grease
- Likely to gain citations as others use it
- To successfully reanalyze old data, you need very good skills

The quick and easy way

- Find a really good published article
- Apply its methodology to a new case (eg, CZ) or a new time period
 - Recreate data in new case
 - Use same technique of analysis
- Voila!

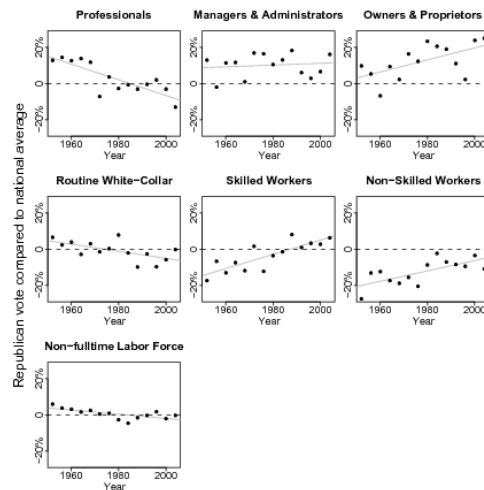


Critique your own research design

- Omitted variables: What other factors might affect X and Y?
- Endogeneity: Could Y cause X?
- Selection bias: Are observations chosen from full range of outcomes? If not, how to justify?
- Mechanisms: What steps connect X to Y?
- Measurement & conceptualization: Are measures & concepts valid and reliable?

Cheap ways to impress

- Lots of pretty graphics, not so many tables
 - Figures in Stata or R
- Use Latex as your word processor
 - Nice typefaces, large margins



Welfare and Paternalism

Stu Jordan
Department of Political Science
University of Rochester

Nolan McCarty
Woodrow Wilson School
Princeton University

March 2, 2010

Abstract

Citizens' desires to assist the poor reflect a mixture of insurance motives, altruism, and paternalism. Consequently, government policies toward the poor have always been a mixture of income supports and regulations of recipient behavior. Unfortunately, there have been very few analyses of the how these distinct motives interact to generate various policies to the poor. In this paper, we develop such a model where a political decisionmaker has policy preferences generated from a mixture of considerations include social insurance, altruism, and paternalism. The decisionmaker's ability to pursue these goals, however, is constrained by the moral hazard and adverse selection inherent in income support policies.

1 Introduction

Debates on welfare have long centered on the effects of its cash transfers and administration on the behavior of the poor. Critics from both the right and left have criticized American welfare policy on these grounds. Conservatives such as Charles Murray (1984) have argued that U.S. welfare programs

Real ways to be smart

- Follow many political scientists on Twitter
- Read outside of political science
 - Psychology, Sociology, Economics, History
 - Again Twitter and blogs
- Take online courses: MOOCs, Coursera, etc
- Learn lots of literatures
 - Annual Review of Political Science/Sociology
 - Journal of Economic Perspectives
 - Oxford Handbooks of...

Next time

- More advanced qualitative techniques
 - Process tracing
 - Structured, focused comparison
 - Case selection techniques
 - Concepts/typologies