Lecture 1. Introduction

Nic Rivers

Advanced quantitative research methods, API6319
Fall 2019

Professor

Dr. Nic Rivers
e-mail: nrivers@uottawa.ca
office: FSS6023
office hours: Thursday 9:00 - 11:00 (or by appointment)

Required materials - textbook

- We will draw from the following text, which is available from the University bookstore:
  - Kellstedt, Paul M. and Whitten, Guy D., 2013, The Fundamentals of Political Science Research, Second Edition. Cambridge University Press.
  - The text is also available as an *online resource* at the library. You can (freely) download all chapters as pdf files. Locate the book by searching by author or this ISBN: 9781139104258. (Note: only one student at a time can access the resource.)
- We will also use the following resource for learning R, which is freely available on-line:
  - Grolemund, Garrett and Wickham, Hadley, R for Data Science (R4DS). `https://r4ds.had.co.nz/`.

Required materials - software

- To learn how to conduct analysis, it is important to get your hands dirty and learn by doing
- We will use R to conduct empirical analysis for the class assignments
- Support for learning R:
  - I will provide a brief introduction to using R for various tasks (each week).
  - There are lots of on-line resources (search: "R merge data" or "R import data", for example).
  - The free book R for data science (R4DS) is an excellent introduction.

Evaluation

Assignments (30%) There will be 10 weekly assignments designed to allow you to practice what we talk about in class. Each assignment is worth 3%. Assignments can be submitted by e-mail and are due before the start of the following class.

Proposal (15%) A two-page proposal describing the research project you will pursue in this class. Details to follow. Due October 8.

Presentation (15%) The last class will be devoted to student presentations of research conducted during the class. Presentations will be 10-15 minutes per student. Details to follow. Due November 26.

Report (40%) A 10-15 page report describing research conducted for this class. Details to follow. Due December 4.

Classes

- Most of the time, classes will be divided into two parts:
  1. A lecture section, in which we will discuss some theory, or a paper, or a new empirical approach,
  2. A lab or workshop section in which we will apply the theory, or have a discussion of research topics, etc. In many cases, it will be useful to bring your laptop to class, so that you can follow along with in-class exercises using R.

Big questions in public and international affairs

- Does development aid help to pull poor countries out of poverty?

  Jeffrey Sachs, End of Poverty: If the wealthy countries of the world were to increase their combined foreign aid budgets to between $135 billion and $195 billion for the next decade, and properly allocate that money, extreme global poverty — defined by the World Bank as an income of less than a dollar a day — could be eliminated by 2025.

  Dambisa Moyo, Dead Aid: In the past 50 years, more than $1 trillion in development-related aid has been transferred from rich countries to Africa. Has this assistance improved the lives of Africans? No. In fact, across the continent, the recipients of this aid are not better off as a result of it, but worse — much worse.

How can we answer this question?

Does development aid help to pull poor countries out of poverty?

- There are two facets to answering this question, which we will cover in class:
  - Come up with a **research design** that can answer the question. The research design will propose a *comparison* that can shed light on the research question. It will consider whether and how it is possible to make a cause-and-effect conclusion. (For example, one possible comparison would be: compare countries that have received foreign aid with countries that have not received foreign aid.)
  - Implement **research methods** to formally implement the comparison, and determine the answer to the research question. The research methods typically involve comparing two or more groups. (For example, researchers might compare the average rate of economic growth in countries that receive aid and countries that don't receive aid.)
- We will briefly discuss **research design** (what comparison to make?); most of the course focuses on **research methods** (how to implement the comparison?).
- We also introduce **data science**: obtaining, organizing, and manipulating data as a precursor to implementing formal research methods.

Big questions in public and international affairs

- Does foreign aid improve a country's economic growth?
- Do ethnic grievances lead to civil wars?
- Does all-day kindergarten increase female employment?
- How big is the gender wage gap in Canada?
- Do conservation programs promote increases in forest cover?
- Do micro-credit programs empower female entrepreneurs?
- Will a changing climate increase the incidence of conflict?
- Do parental-leave policies affect the birth rate?
- Does a carbon tax reduce gasoline consumption?
- Do large class sizes hurt students?
- Are home ownership rates for Millenials lower than prior generations? (what about full-time employment rates?)

Types of questions we answer

- These questions are *positive* as opposed to *normative*

  positive: Positive questions are questions about what *is*.

  normative: Normative questions are questions about what *should be*.

  (You can't answer a normative question with data.)
- These questions focus on description and/or cause and effect
- We can try to apply a scientific approach to answering these questions
- Knowing the answers to these questions is important for formulation of effective strategies and public policies.

What constitutes scientific research in the social sciences?

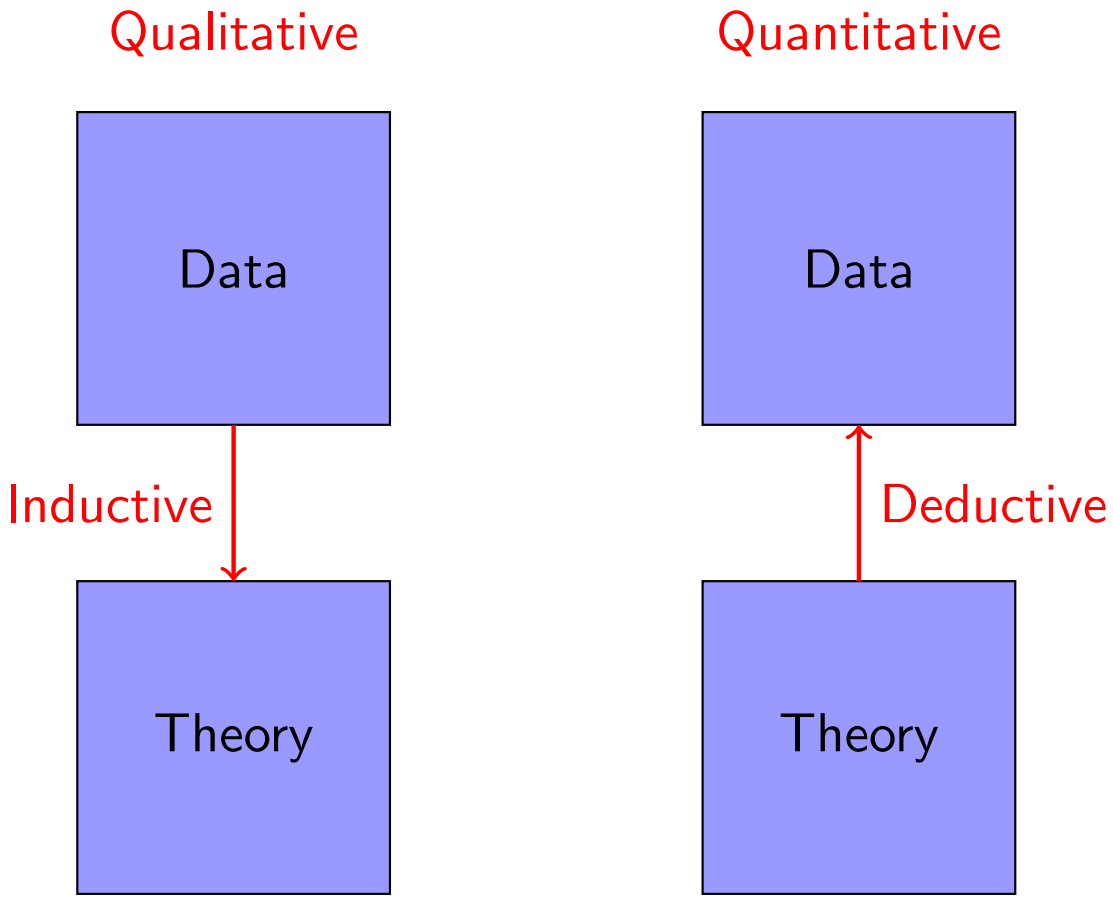
- What do scientists do? Formulate theories about the world and test those theories against data.
- Scientific research in the social sciences involves application of the *scientific method* to questions of importance to social scientists.
- The scientific method involves the use of data to formally test *hypotheses* about the way the world works.

Scientists vs. lawyers

- Scientists come up with theories about the way the world works. They should be detached from their theories, and aim to test them as rigorously as possible - always looking for evidence that they might be wrong.
- Scientists only begin to accept a new theory after hypotheses based on that theory have been subject to a number of tough tests.
- Scientific tests are biased against accepting new theories.
- Contrast to a lawyer: advocates for a particular position; emphasizes research that supports that position.
- When a lawyer faces new evidence that supports her position, she emphazises it. When she faces new evidence that counters her position; she points out weaknesses.

Quantitative (scientific) research

Quantitative research is a form of *deductive reasoning*



Inductive reasoning by Sherlock Holmes:

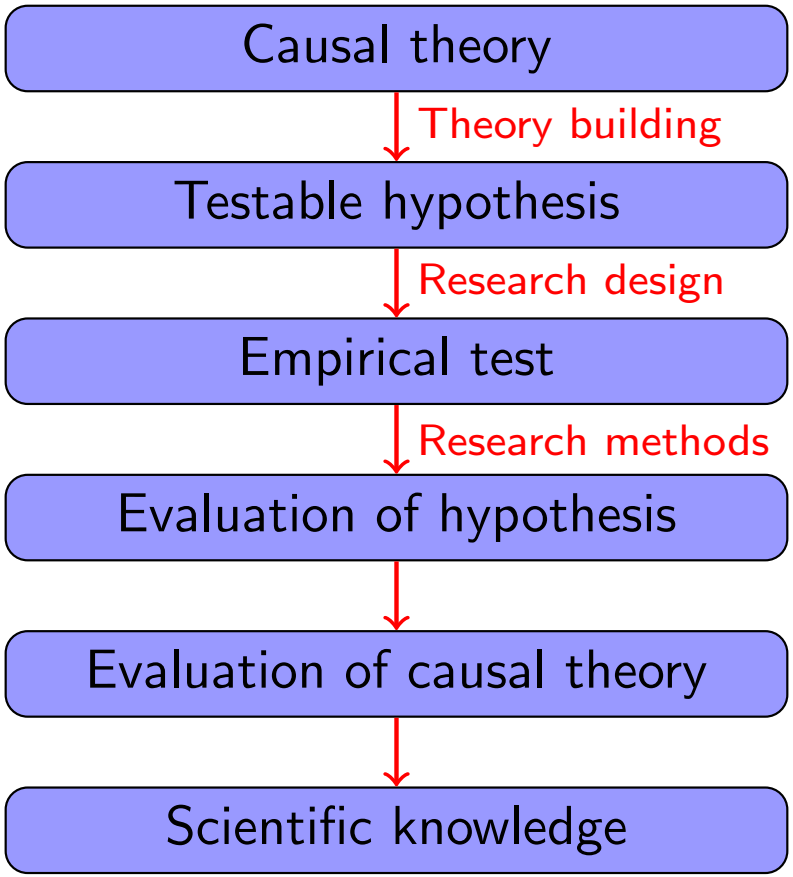
Data (ring tan line) → Theory (was engaged before)

(Bad) deductive reasoning by Monty Python:

Theory (she's a witch) → Data (does she weigh the same as a duck?)

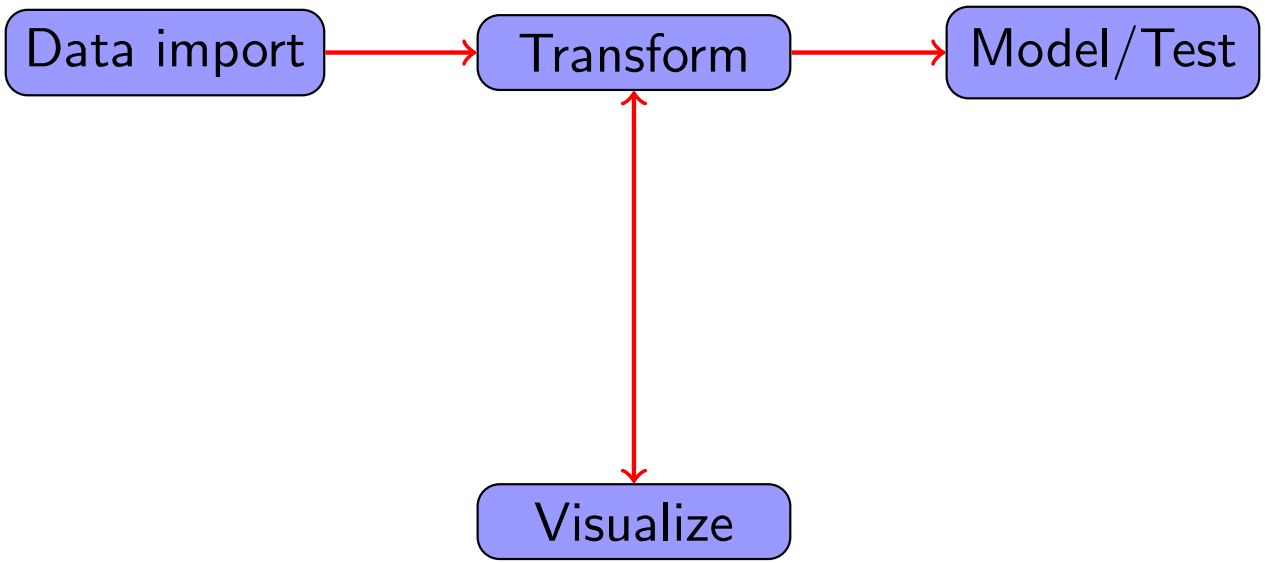
The scientific method

Moving from a general causal theory to scientific knowledge by evaluating the theory against data



Data science

Conducting an *empirical test*



Theory building

- A *theory* is a tentative conjecture that tries to explain some phenomenon.
- The process of theory-building often begins with observation of the real world or is motivated by wanting to know more about the real world.
  - Lots of intra-state conflict seems to occur in countries with rugged terrain. Rugged terrain favours insurgency. Does the presence of rugged terrain increase the likelihood of onset of civil war?
  - Does making schooling compulsory for longer lead to better life outcomes for children (e.g., health, income...)

Operationalizing a theory: Hypotheses

- The aim of the scientific method is to test theory against data.
- In order to test the theory against data, we reformulate the theory into a *testable hypothesis*.
- A hypothesis is a statement about what we would expect to see in the data if our theory were true.
- We can also formulate a *null hypothesis*. A null hypothesis describes what we would expect to see in the data if the theory were not true. Since the hypothesis typically involves a cause and effect relationship between two concepts, the null hypothesis is that there is no relationship between the two concepts.

  Research question Do countries with more rugged terrain experience more civil wars?

  Hypothesis Countries with more rugged terrain are more likely to experience civil wars

  Null hypothesis There is no relationship between terrain and likelihood of civil wars

Operationalizing a theory: variables

- A theory is about relates some phenomenon to some cause.
- A variable is an operationalization of a concept in our theory (e.g., age, conflict, trust, height, corruption). It puts it in the form of something that can be measured in the real world.
- So a theory typically is framed in a way that one variable causes another variable.
- Theories typically specify a causal relationship between an *independent variable* and a *dependent variable*. As shorthand, we sometimes use Y to refer to the dependent variable and X to refer to the independent variable. Our theory is therefore about:

$$X \rightarrow Y$$

(i.e., the effect of X on Y).

Developing a theory

- It is important to recognize that part of the scientific method is testing theories against data.
- Theories that can't be tested against data are not scientific theories.
- It is good practice to think about how a theory (either your own, or in a paper you're reading) would look—in the real world—if it were true.
- Likewise, it is good practice to think about how you might find data to 'disprove' the theory you're considering.

Measurement of variables

- In some cases, how to measure a variable is clear. For example, if our theory is that political party preferences vary by age, then it is clear how we measure age.
- In some cases, however, there is not a perfect link between a concept in our theory and a real-world variable. For example, Sociologist Robert Putnam has advanced a theory that says that more *diversity* in communities causes lower *trust*. How to measure these concepts?

Turning a concept into a variable: Example - trust

In the city or area where you live, imagine that you lost your wallet ... or something holding your identification or address ... and it was found by someone else. Do you think your wallet (or your valuables) would be returned to you if it were found by a neighbor?

(% respondents saying "yes")

	Highest Trustworthiness		Lowest Trustworthiness
Finland	94%	Cambodia	24%
New Zealand	94	Laos	24
Norway	94	Sierra Leone	26
Switzerland	92	Bolivia	31
Ireland	91	Chad	36
Austria	88	Peru	38
Rwanda	88	South Africa	38
Slovenia	88	Benin	39
Mauritania	83	Tanzania	39
Niger	80	Ecuador	40

GALLUP POLL

Turning a concept into a variable: Example - trust

LITERAL QUESTION

Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people?

Values	Categories	N	NW	
1	Most people can be trusted	807	820.7	54.7%
2	You cannot be too careful in dealing with people	672	656.4	43.8%
99	DK/NA	21	22.9	1.5%

Research design

- After we have arrived at a hypothesis that we would like to test against data, we need to consider exactly how we will subject the hypothesis to an empirical test.
- This is referred to as a *research design*: how will we design our research to provide an effective test of our hypothesis? In particular, we would like a research process has a good chance of:
  1. rejecting our null hypothesis if it is actually false, and
  2. not rejecting our null hypothesis if it is actually true.
- Research design consists of choosing and gathering data, thinking carefully about cause and effect claims, and thinking about what statistical tests will be used to test the hypothesis.

Research methods (statistical methods)

- Once we have selected a research design and collected data, we will use the data to subject our research hypothesis to a formal test.
- Formal statistical tests basically ask:

  If the null hypothesis was actually true, how likely is it that data in the real world would be like the data we have in hand?

- If the likelihood is very small, then we judge that the null hypothesis is likely false.
- A significant portion of this course will focus on design and application of these formal statistical tests.

Example: the scientific method

- There is lots of interest in Canada right now about climate change policies. One concern is about whether people will respond to higher gasoline prices by buying more fuel efficient cars or not. This is something we can examine using the scientific method.

  Causal theory : higher gasoline prices lead people to buy more fuel efficient cars.

  Testable hypothesis : when gasoline prices are higher, people will purchase more fuel efficient vehicles.

  Null hypothesis there is no relationship between gasoline prices and the fuel efficiency of vehicles people buy.

  Operationalize : Measure real gasoline prices in each Canadian city over a number of years; compile the average fuel economy of vehicles purchased in each city of those years.

  Empirical test : Test for a statistical relationship between gasoline prices and fuel efficiency (figure next slide) (we will discuss tests later in course)

  Evaluation : Data appears to contradict null hypothesis

New vehicle fuel economy and gasoline price

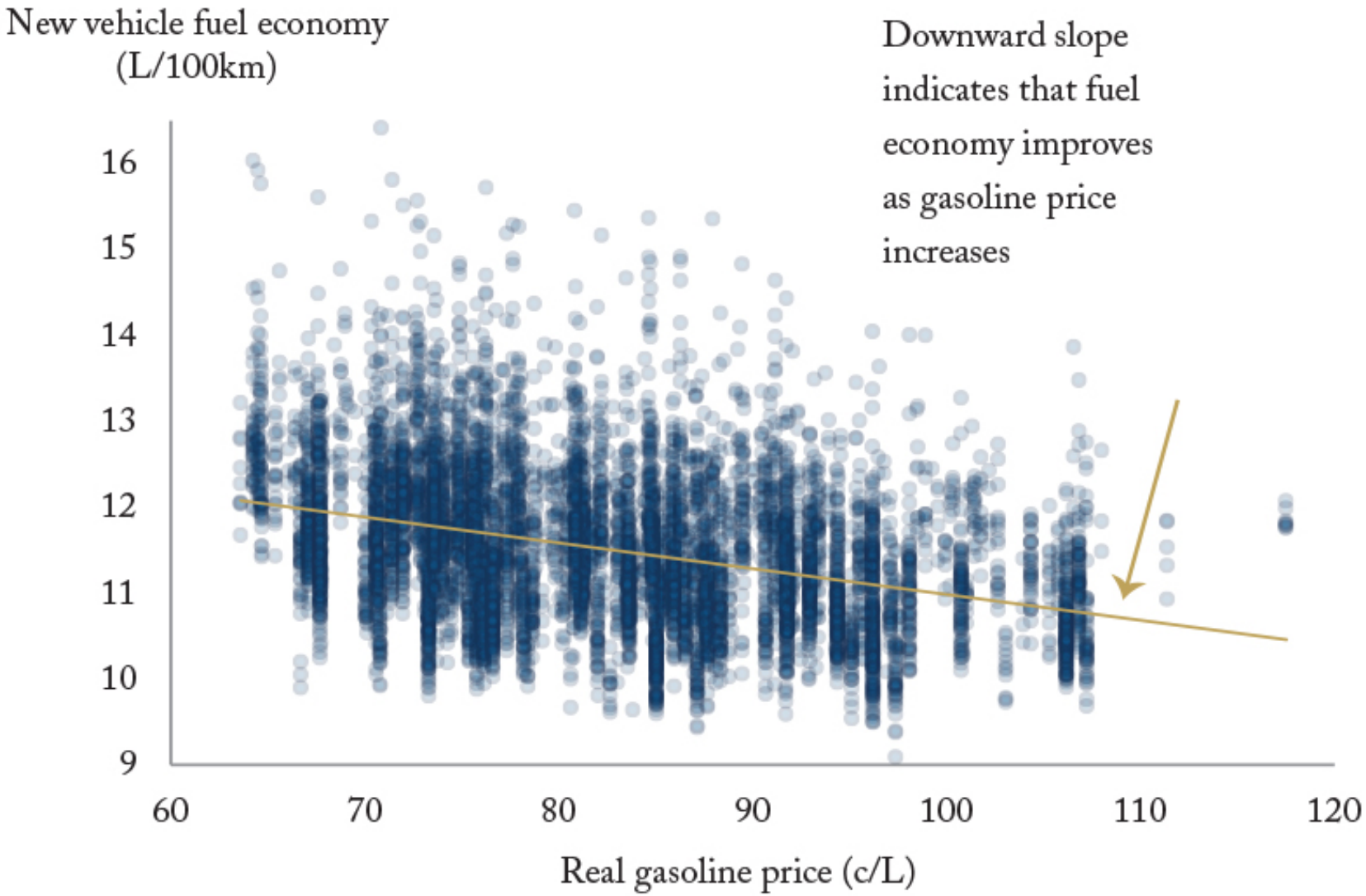


Figure: See: https://www.cdhowe.org/intelligence-memos/nic-rivers-will-consumers-respond-carbon-tax

Course objectives

The aim of this course is to:

1. provide you with the background required so that you are able to understand research that attempts to empirically answer these (and other) critical questions, and,
2. introduce you to the research designs and empirical methods required for conducting empirical research on your own,
3. introduce you to a commonly-used software package for data analysis (R), and,
4. give you practice in conducting your own quantitative research project, from start to finish.