Project proposals

Project work

- Performed alone or in pairs

- **Step 1 (Proposal):** Deadline October 6.
  1. Join a Canvas group
  2. Identify a causal inference problem (e.g., in your research)
  3. Define causal inference problem mathematically
  4. Propose a dataset to learn from / experiment to run
  5. Propose identification & estimation strategy

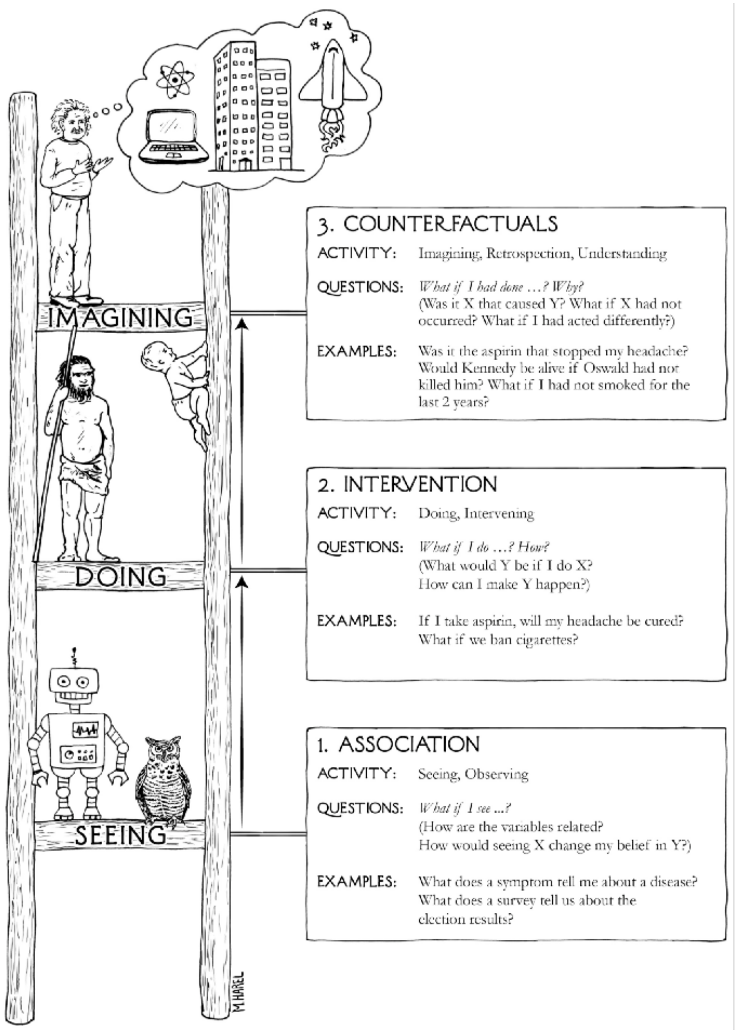
- **Step 2 (Report):** Deadline at the end of the course.
  - Describe the project and its results (more info. on Canvas)

What is a causal inference problem?

Broadly speaking, causal inference problems (in this context) concern rungs 2 and 3 on this ladder

- Interventions
- Counterfactuals

… and the latter is often quite tricky



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

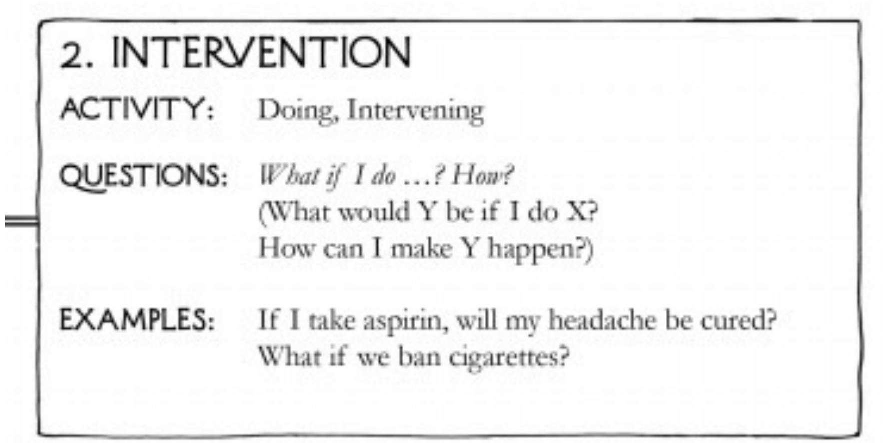
1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see …?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

What is a causal inference problem?

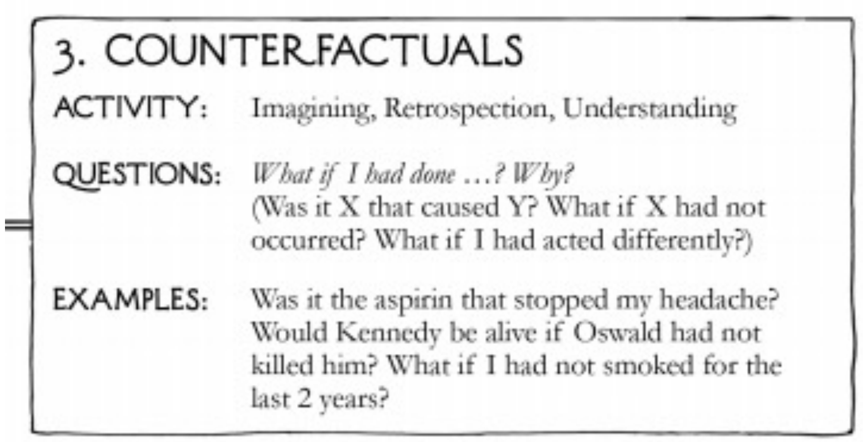


2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

Understanding interventions

Interventions come in many flavors

- Choosing between drug A and drug B for patient X
- Recommending a product on a website
- Controlling a robot in a new environment
- Directing economic policy
- Sentencing criminals

Learning about interventions

Broadly speaking, there are two ways to learn about the outcomes of interventions*

Experiments:
Control the interventions yourself, observe outcomes

Observational studies:
Passively observe interventions controlled by another agent and their outcomes, typically retrospectively

* … that we will study in this class

Toy experiment

Example: You want to find out how satisfied your peers are with their education. You suspect that self-reported answers may be affected by who is asking the question, when it is asked, etc.

Problem: Estimate the effect of time on self-reported responses.

Experiment: Gather 100 students, divide randomly into two groups. Survey one half on Monday, one half on Fridays.

* … that we will study in this class

Experiments

The difficulties of running experiments are both practical…

- Cost, time, ethics,

… and fundamental / philosophical…

- Experiment biases, placebo effects, etc.

* … that we will study in this class

Example of an observational study

Example: You want to evaluate a new policy for recommending products on a website

Problem: Estimate the causal effect of moving from the old policy A to the new policy B on future sales

Data: You have downloaded a dataset of product recommendations and purchase history of users of some company

* … that we will study in this class

Observational studies

The difficulties of performing observational studies include

- Accessing all relevant variables (will be clarified later)
- Creating clear definitions of actions and outcomes (follow-up)

You are welcome to work with synthetic data.

Potential issues in observational studies

What does “treatment” mean?

- When was the treatment prescribed? When was it taken?
- By which policy was it selected?
- Was the same dose given?

What does “outcome” mean?

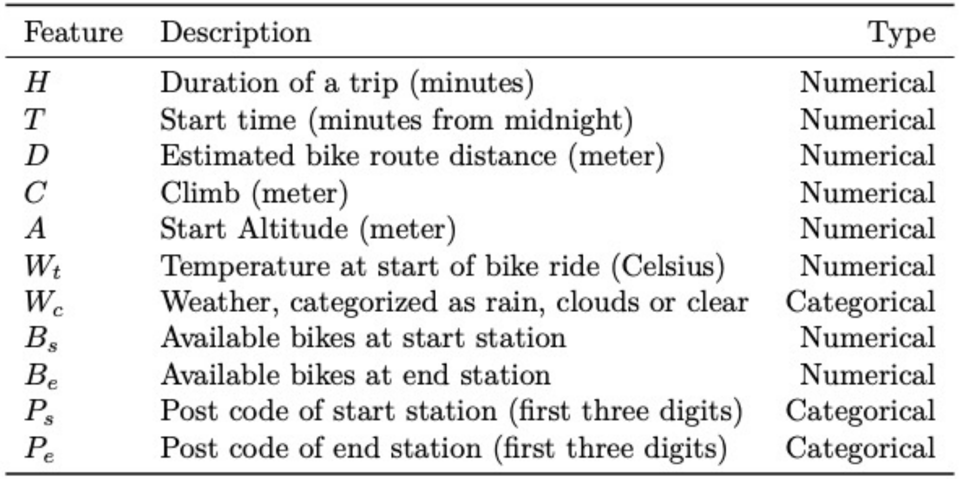
- When was the outcome measured?
- Was it measured using the same equipment?
- What if a patient left the dataset before follow-up?

Example projects from 2020

- Causal discovery in bike sharing service database [Observational study]
- Improving prediction using causal graphs [Experiment / observational study]
- Causes that affect the performance of neural networks on the MNIST dataset [Experiment]
- Causal effects of natural language using Yelp reviews [Observational study]
- Causal effects of multiple concurrent treatments using the MIMIC database [Observational study]
- Counterfactuals in Alzheimer's disease and Lung cancer [Observational study / Simulator]
- Impact of demographic variables on voting [Observational study]

1: Bike sharing service

- Accessed data from the Styr & Ställ API:



Feature	Description	Type
H	Duration of a trip (minutes)	Numerical
T	Start time (minutes from midnight)	Numerical
D	Estimated bike route distance (meter)	Numerical
C	Climb (meter)	Numerical
A	Start Altitude (meter)	Numerical
W_t	Temperature at start of bike ride (Celsius)	Numerical
W_c	Weather, categorized as rain, clouds or clear	Categorical
B_s	Available bikes at start station	Numerical
B_e	Available bikes at end station	Numerical
P_s	Post code of start station (first three digits)	Categorical
P_e	Post code of end station (first three digits)	Categorical

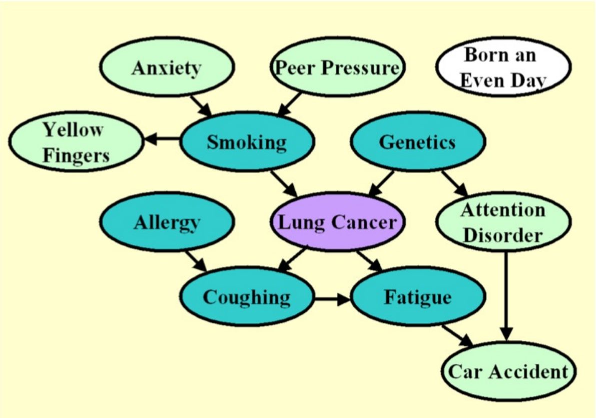
Table 1: *The variables used in our dataset, along with a description and the type of the variable.*

- Tried to learn the causal graph connecting these variables*
- Reported and discussed multiple possible explanations

* Structure learning is not part of this year's course

2: Counterfactuals of lung cancer and smoking

- Studied the LUCAS0 Lung cancer simulator



- Estimated counterfactuals under the monotonicity assumption
- Would someone with lung cancer have been healthy if they didn’t smoke?

Working on the proposal

But Fredrik! How can I write a proposal when I don’t know everything I need to know about causality?

To a large extent, this is about learning the process of working with causal problems, defining them, etc.

I will give you feedback on the proposal and you will iterate.