

A Gentle Introduction to Causal Inference

The schedule for each session...

14:00 -- 14:10 Introduction and Motivations

14:10 -- 14:50 Presentation and Q&A

14:50 -- 15:00 10-minute Break

15:00 -- 15:45 Practical Session

15:45 -- 16:00 Final Wrap-Up





THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



Welcome!



Chris Oldnall



www.cdcs.ed.ac.uk



THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society



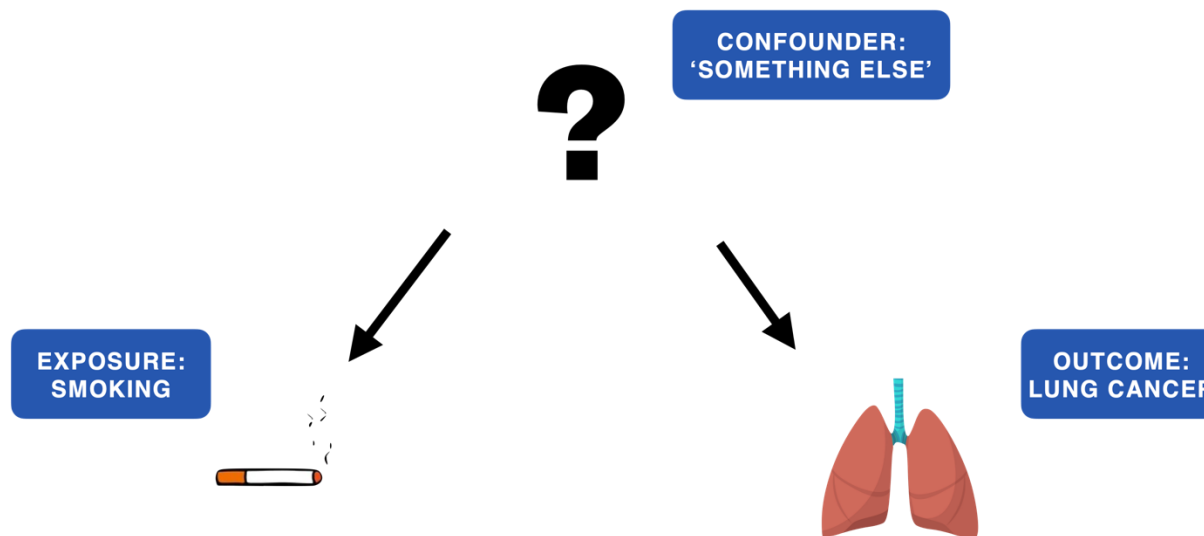
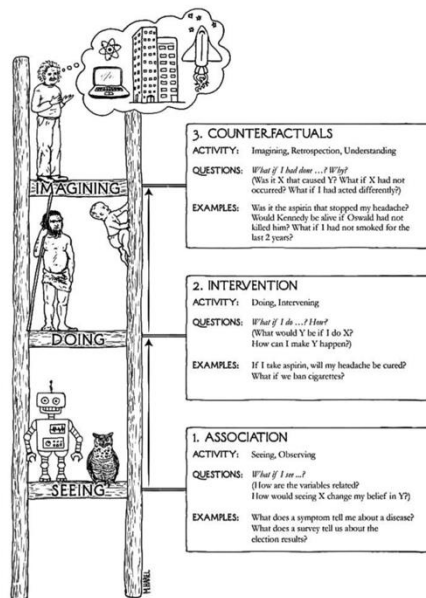
Week 2: Advanced Concepts and Estimation



www.cdcs.ed.ac.uk

What did we look at last time?

Smoking in the multiverse of madness



The problem of confounding



The problem of confounding

If we simply compare communities with and without funding, we may overestimate the effect due to existing differences in community engagement.



Prior
community
engagement

Need for better
methods to **balance**
groups and make fair
comparisons.

Funding for
community
engagement.



Effect of funding on
community engagement.



Engagement
from the
community.

Stable unit treatment value assumption (SUTVA)

- 1. No interference** – One unit's treatment shouldn't affect another's outcome.
- 2. No multiple versions of treatment** – The treatment must be well-defined.

The effect of social media bans on political activism:

If one group loses access to social media but mobilises through other means and influences another group's activism, the assumption is violated.



What is a propensity score?

The probability of receiving an intervention given a set of observed covariates.

$$e(X) = P(T = 1|X)$$

- Creates comparable groups **without needing randomisation**.
- Reduces confounding by balancing covariates.



How do we estimate a propensity score?

Logistic Regression: Predicts treatment assignment based on covariates.

$$\log \left(\frac{e(X)}{1 - e(X)} \right) = \beta_0 + \beta_1 X_1 + \dots$$

Machine Learning Approaches such as Decision Trees, Random Forests, Neural Networks exist and are more flexible but require careful tuning.



Using propensity scores for matching.

Matching Method: Each treated unit is paired with a control unit having a similar propensity score.

Steps:

- Estimate **propensity scores**.
- Match treated and untreated units.
- Assess balance and estimate the causal effect.



Inverse probability of treatment weighting (IPTW)

Instead of discarding data (like matching), IPTW **reweights** observations.

$$w_i = \frac{T_i}{e(X_i)} + \frac{(1 - T_i)}{1 - e(X_i)}$$

Effectively creates a pseudo-population
that mimics randomisation.



Some key considerations on IPTW.

Advantages:

- Uses full dataset.
- More efficient than matching.

Disadvantages:

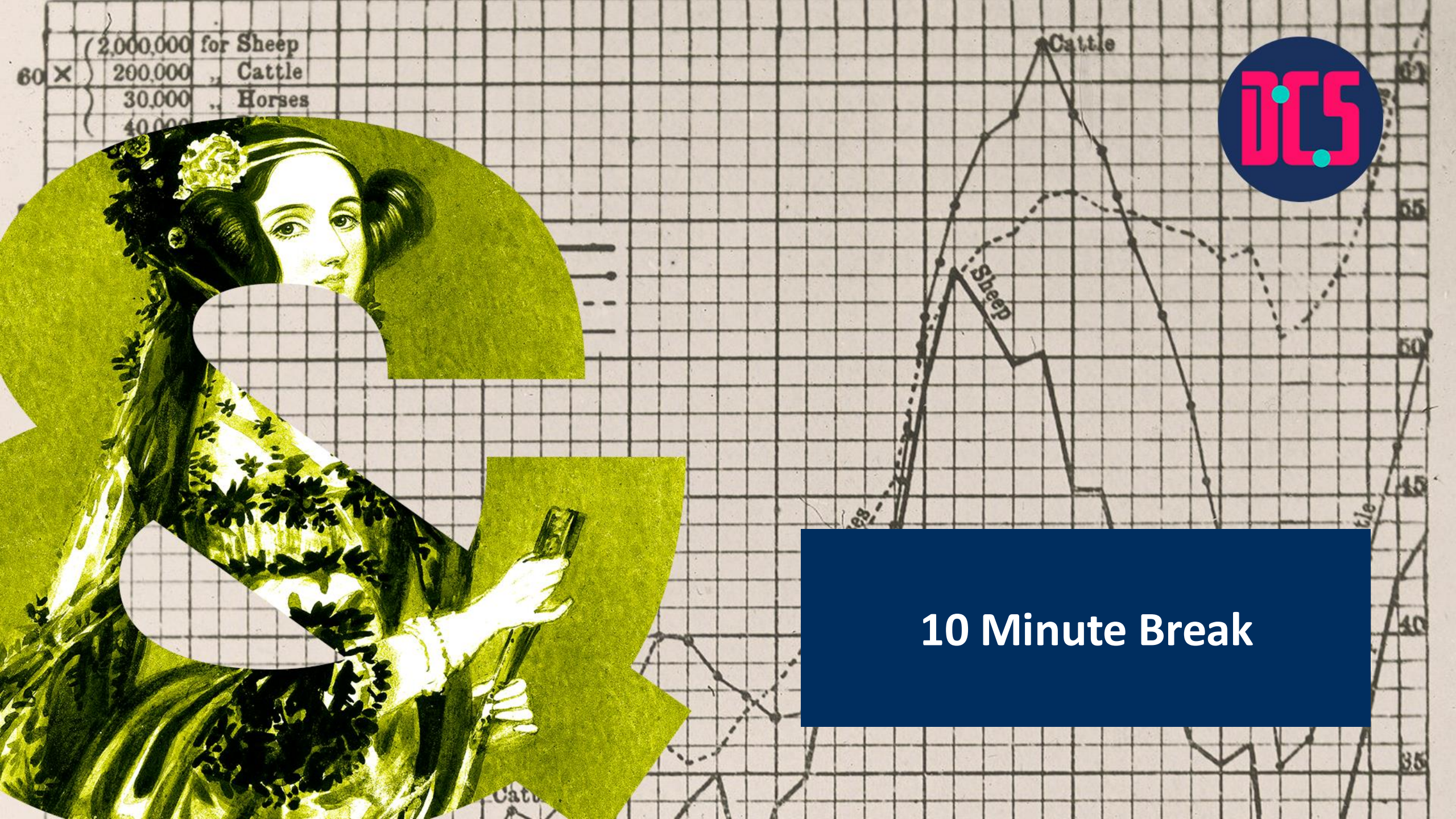
- Sensitive to incorrect model specification.
- Can create extreme weights



When do we use what methods to deal with confounding?

| Method | Strengths | Weaknesses |
|-------------------------------|----------------------------------------------|----------------------------------------------------------------------------|
| Randomisation | Eliminates confounding. | Often infeasible with already collected data / not specifically collected. |
| Matching | Easy to interpret, removes extreme cases. | Data loss. |
| IPTW | Uses full data, handles complex confounding. | Sensitive to model misspecification |
| Instrumental Variables | Can handle unobserved confounding. | Hard to find valid instruments. |







THE UNIVERSITY OF EDINBURGH
Centre for Data, Culture & Society

Practical Session 2



www.cdcs.ed.ac.uk

PAIR PROGRAMMING

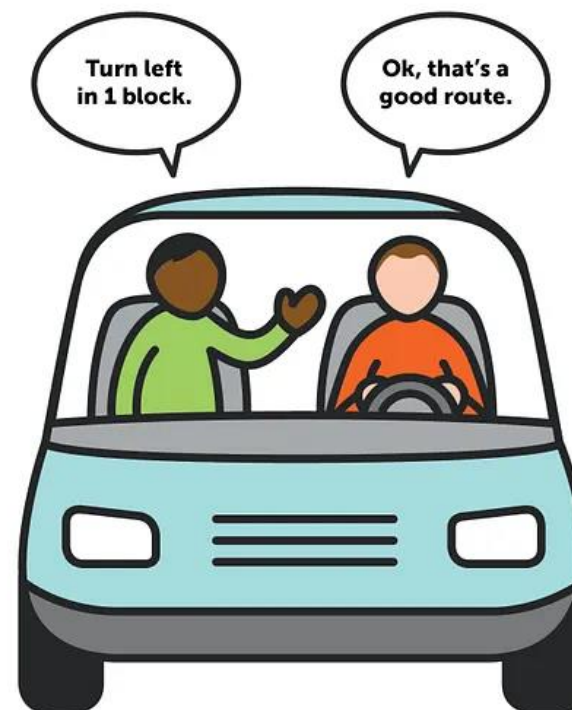
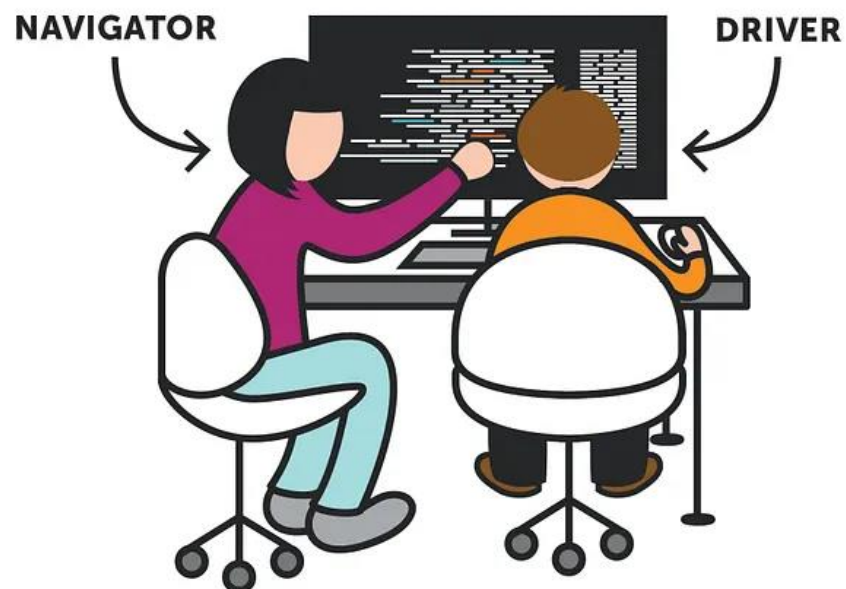


Image from:
https://medium.com/@tomspencer_uk/pair-programming-and-problem-solving-4531ef3bf171

Pair Programming

- find a partner
- switch driver & navigator roles regularly, e.g., after every task in the notebook
- you can work with the same partner throughout the course, or switch between sessions
- ask us for help! (yes, even for small things)



Let's get programming

For Python:

1. Go to <https://noteable.edina.ac.uk/login>
2. Login with your EASE credentials
3. Select 'Standard Notebook (Python3)' as a personal notebook server and press start
4. Click the '+GitRepo'
5. Copy and Paste this repository URL <https://github.com/DCS-training/IntroCausalInference> as the Repository URL - you do not need to add in any other fields.
6. Decide where to locate the folder. By default, it will locate it in your home directory
7. Press 'Clone' Congratulations you have now pulled the content of the repository on your Notable server space.

For R:

1. Go to <https://noteable.edina.ac.uk/login>
2. Login with your EASE credentials
3. Select RStudio as a personal notebook server and press start.
4. Go to File > New Project> Version Control > Git
5. Copy and Paste this repository URL <https://github.com/DCS-training/IntroCausalInference> as the Repository URL (The Project directory name will filled in automatically but you can change it if you want your folder in Notable to have a different name).
6. Decide where to locate the folder. By default, it will locate it in your home directory.
7. Press Create Project Congratulations you have now pulled the content of the repository on your Notable server space.

