Rotterdam School of Management
Erasmus University

RSM Erasmus

# Causal Inference

## Lecture 1
## Introduction to the course

Dr. Xi Chen
6/10/2023

# Welcome!

# What is causal inference?

Inferring the effects of any treatment/policy/intervention/business strategy/marketing campaign etc.

Examples:

...Omitted because there are too many of 'em...

# Why causal inference?

Correlation is not causation.

A thought on causation:

- Causation is a limiting case of correlation
  - A & B must be correlated.
  - A always precedes B.
  - If A does not happen, B will not happen.

How do you make of it?

# Correlation is not causation

This is the idea from Karl Pearson!

"If the unit A be always preceded, accompanied or followed by B, and without A, B does not take place, then we are accustomed to speak of a causal relationship between A and B." (Pearson and Lee, 1897, p. 459)

"It is the conception of correlation between two occurrences embracing all relationship from absolute independence to complete dependence, which is the wider category by which we have to replace the old idea of causation." (Pearson 1910, p. 157)

# Correlation is not causation

## The origin of “correlation is not causation”

## Karl Pearson’ work on correlation

- “Independent contributory causes” of two correlated observables (Pearson 1896, p. 261-263)
- Spurious correlation (Pearson 1897)
- The search for “organic relationships” (Pearson and Elderton 1923)

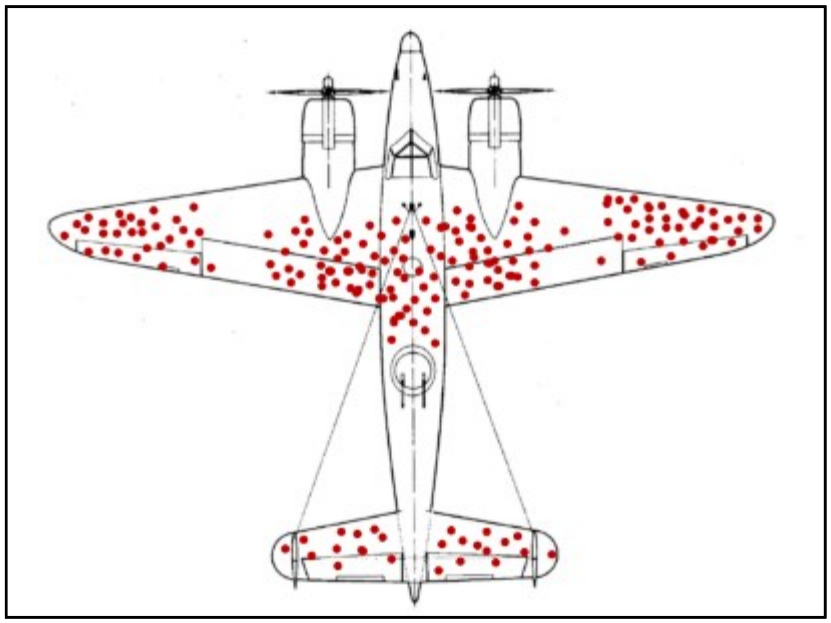
# A quote from Pearson

Beyond such discarded fundamentals as “matter” and “force” lies still another **fetish** amidst the **inscrutable arcana** of modern science, namely, the category of cause and effect.

-- Karl Pearson

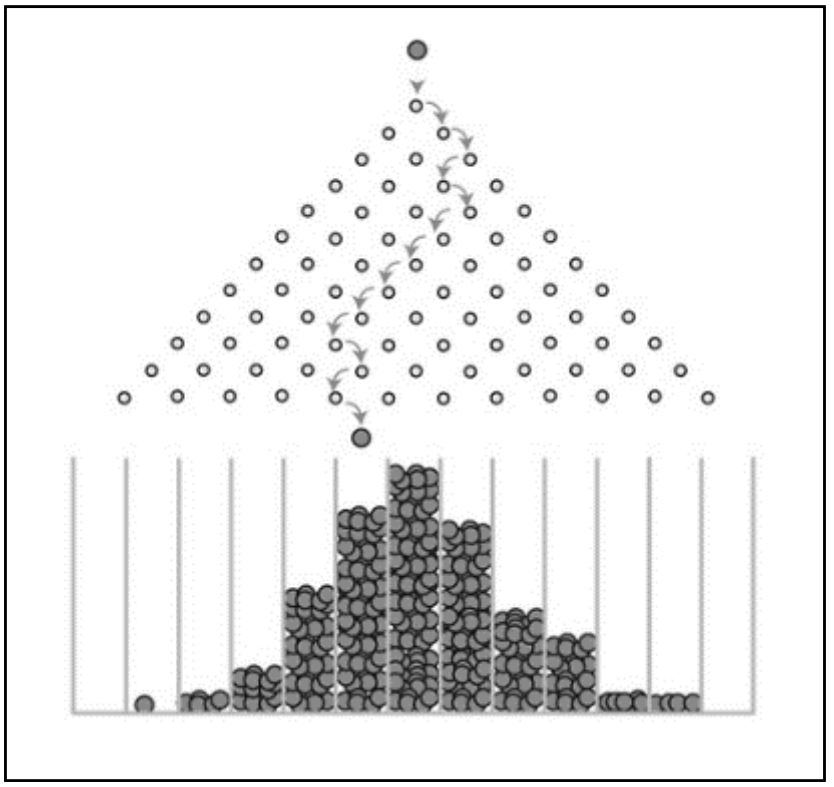
# The importance of knowing causality

## “Cliche” examples



Wald, Abraham. (1943). *A Method of Estimating Plane Vulnerability Based on Damage of Survivors*. Statistical Research Group, Columbia University.

Survival Bias



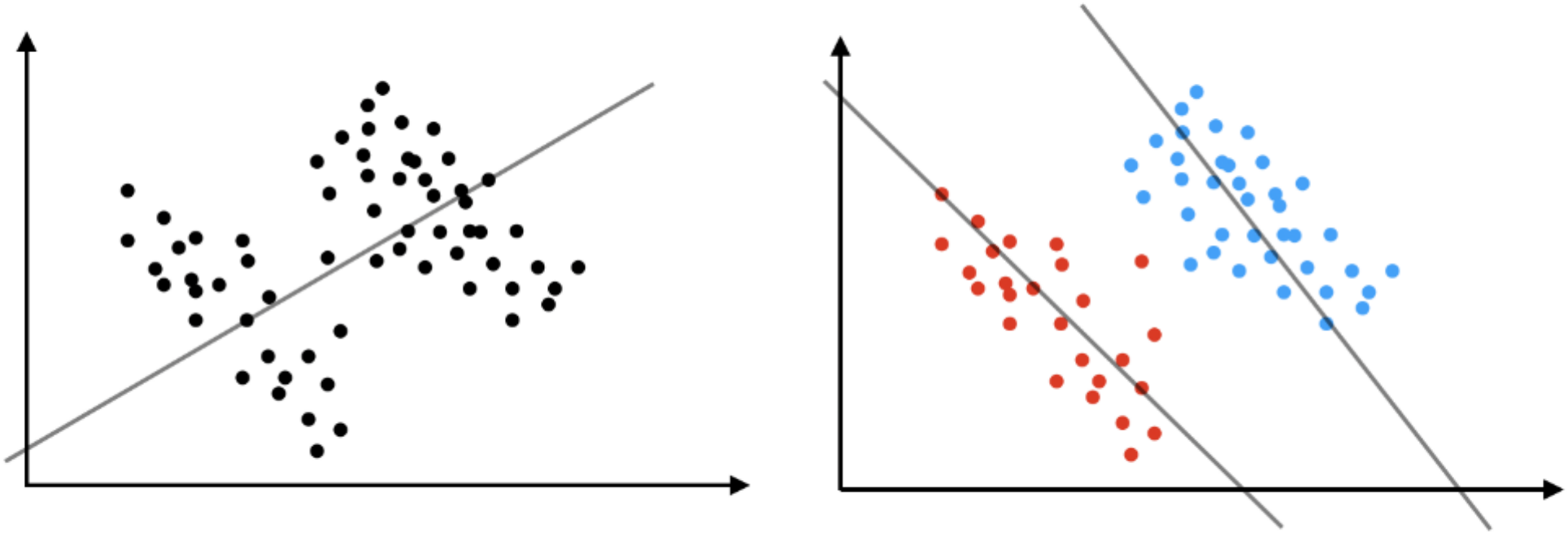
Galton, F. (1894). *Natural inheritance*. Macmillan and Company.

Regression to the Mean

# The importance of knowing causality

## “Cliche” examples

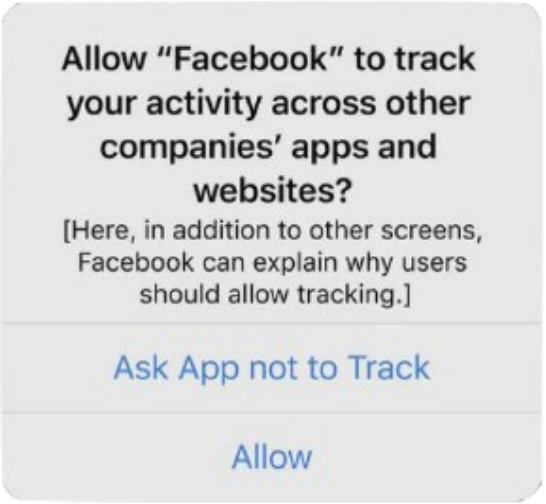
Simpson, Edward H. (1951). "The Interpretation of Interaction in Contingency Tables". Journal of the Royal Statistical Society, Series B. **13**: 238–241.



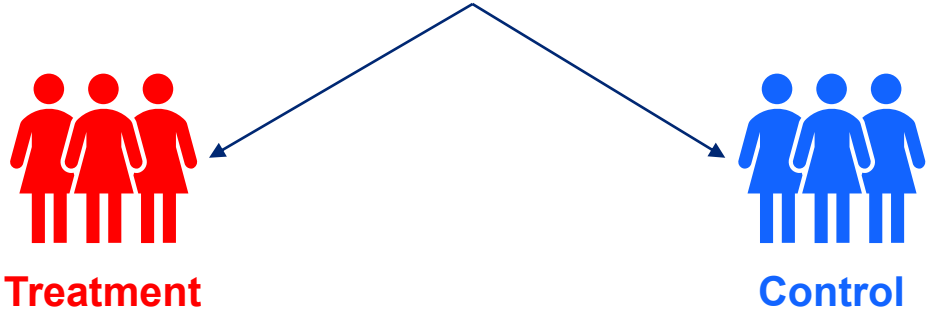
Simpson’s Paradox

# More modern examples…

Facebook vs. Apple



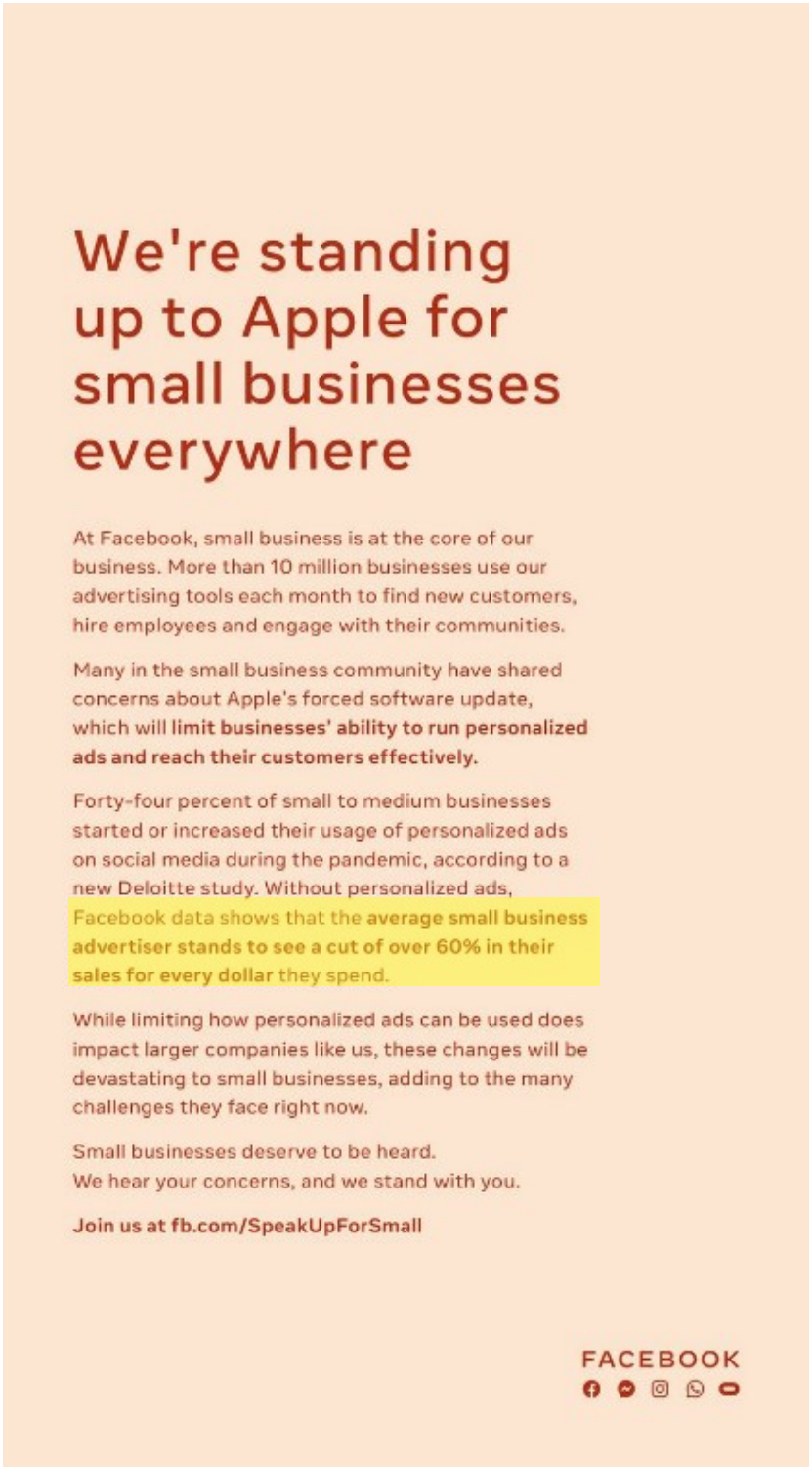

**2,5000 ad campaigns**



**Treatment** **Control**

**For Clicks** **For Purchases**

For more, see here: https://hbr.org/2021/02/facebooks-misleading-campaign-against-apples-privacy-policy



## We're standing up to Apple for small businesses everywhere

At Facebook, small business is at the core of our business. More than 10 million businesses use our advertising tools each month to find new customers, hire employees and engage with their communities.

Many in the small business community have shared concerns about Apple's forced software update, which will **limit businesses' ability to run personalized ads and reach their customers effectively.**

Forty-four percent of small to medium businesses started or increased their usage of personalized ads on social media during the pandemic, according to a new Deloitte study. Without personalized ads, Facebook data shows that the **average small business advertiser stands to see a cut of over 60% in their sales for every dollar** they spend.

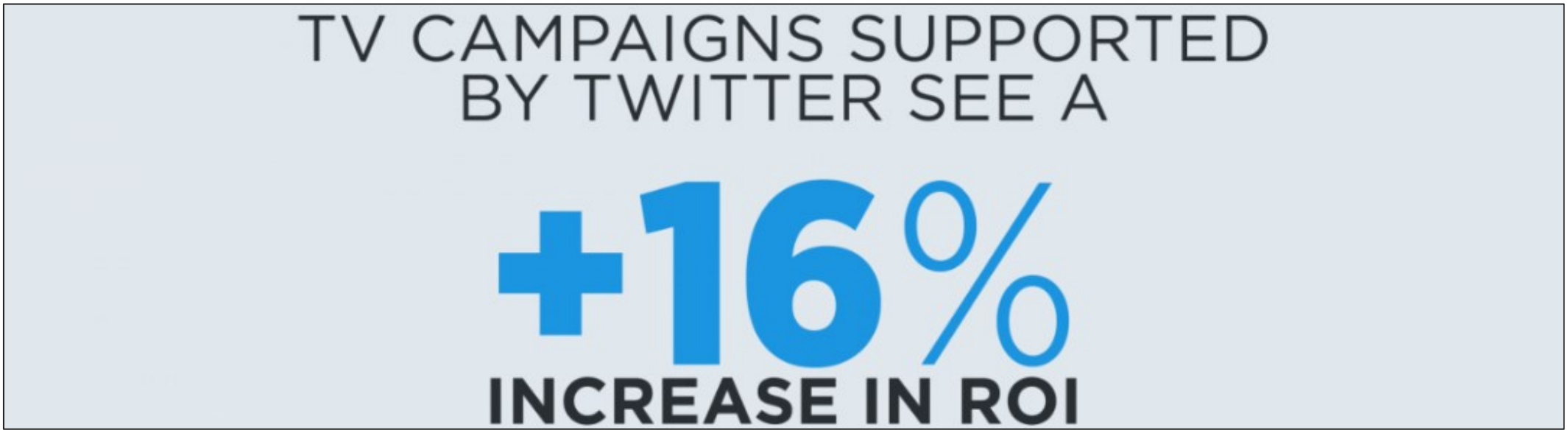
While limiting how personalized ads can be used does impact larger companies like us, these changes will be devastating to small businesses, adding to the many challenges they face right now.

Small businesses deserve to be heard.
We hear your concerns, and we stand with you.

**Join us at fb.com/SpeakUpForSmall**

FACEBOOK

# Business examples




**The study's design**

Studios tend to spend most of their advertising budgets around a film's theatrical release date. To get a comprehensive view of performance, we looked at nearly four years of paid theatrical history data — from December 2011 to June 2015. The research team used multivariate regression analysis (a process that measures and predicts the sales impact of various media channels) to understand the effects of changes in Twitter media for movies. Researchers also looked specifically at comedy and action genres in the US, as these categories were among the biggest box office performers.

Source: https://blog.twitter.com/en_us/a/2016/new-movie-marketing-research-reveals-twitter-ads-deliver-increased-ticket-sales

## Self-introduction

Xi Chen

quantitative marketing

digital marketing

policy evaluation

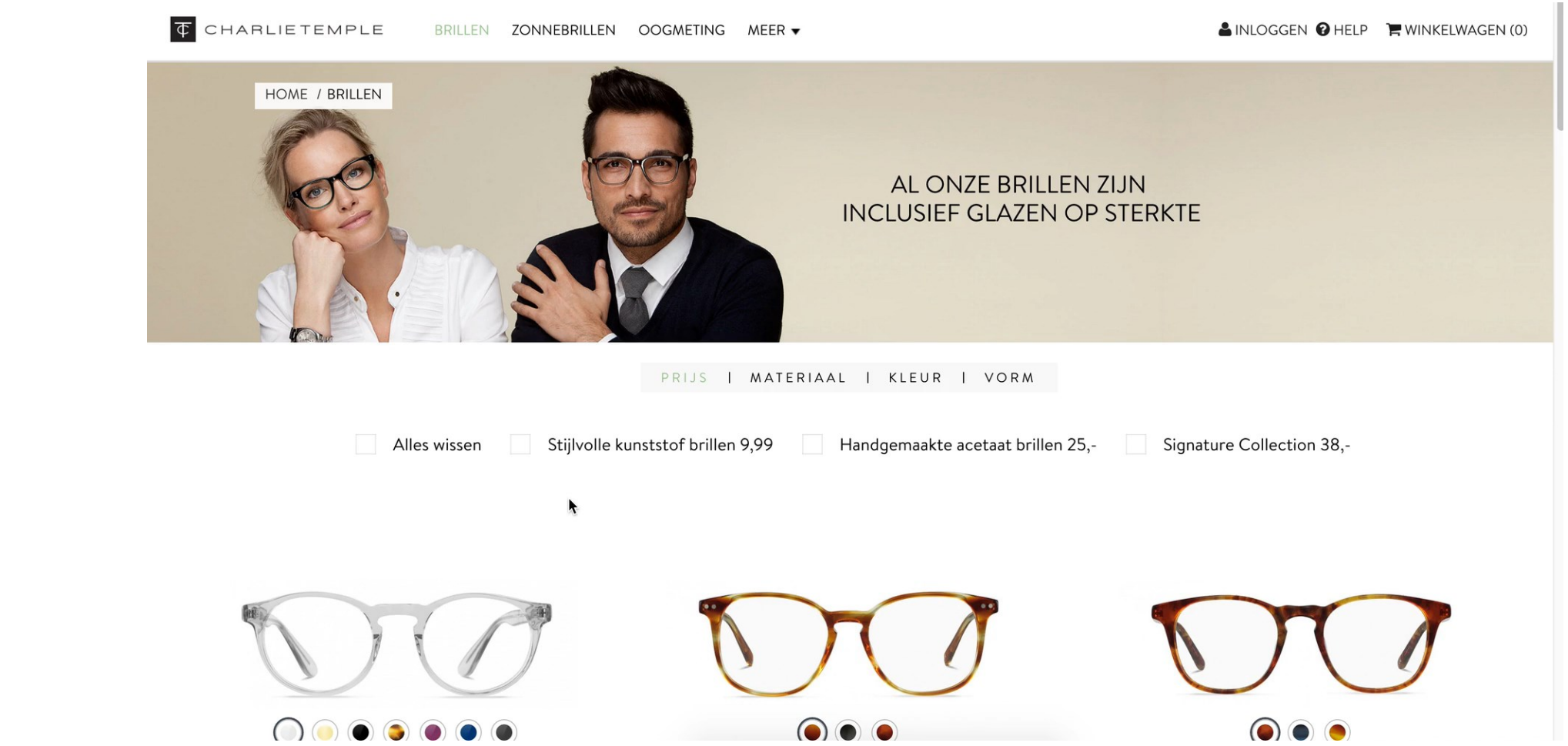
rich experiences with data analysis

working with companies

# AR Fitting on Product Returns

## At Charlie Temple

Video link: https://vimeo.com/175709390



## To get acquainted…

A short self-intro with

- Your name
- Your background

# Ask/share anything about causality or causal inference

# 21 Questions Game: causal inference ver.

No causation without correlation

Running a regression gives you effects

- e.g. Sales = b*Price

Adding more variables always gives you better estimates

- e.g. Sales = b*Price + Seasonality +…

# 21 Questions Game: causal inference ver.

"No causation without manipulation"

OR "Only experiments can establish causality"

- "Causes are only those things that could, in principle, be treatments in experiment" (Holland 1986, p.954)
- "Such questions ['no series of actions can be inferred from the description of the treatment'] have no causal answer within our framework" (Rubin 1978, p.39)

How experiments establish causation?

- What's so unique about experiments?

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, *6*(1), 34-58.

# 21 Questions Game: causal inference ver.

A or B?

- A. Causation can be inferred solely from data.
- B. Causal inference always requires assumptions on top of data.

Any other questions?

# Overview of the course

# The thinking behind the course design

The decision problem:

$$\textbf{max} \quad LearningExperiences$$

$$\textbf{subject to}: \begin{cases} 5\ days\ and\ 25\ hours \\ Compact\ Learning \\ A\ Complex\ Topic \\ Diversity\ of\ Our\ Group \end{cases}$$

# Your expectations from the pre-survey

1. Methods in causal inference
   - RDD and Diff-in-Diff
   - Causal forest
2. Addressing causality issues
3. Theoretical aspects
4. Applications or coding aspect
5. Running field experiments and experiments
6. Conceptualizing causal relationships
7. Understanding assumptions
8. Secondary data

# The thinking behind the course design

To integrate different perspectives for problem-solving

To focus on the in-depth understanding

Only on essential technical details

To communicate “principles” of causal inference

For your critical thinking

To use blended learning methods

# The thinking behind the course design

First-principles based approach

We go back to the source!

1

Start from the basic problems

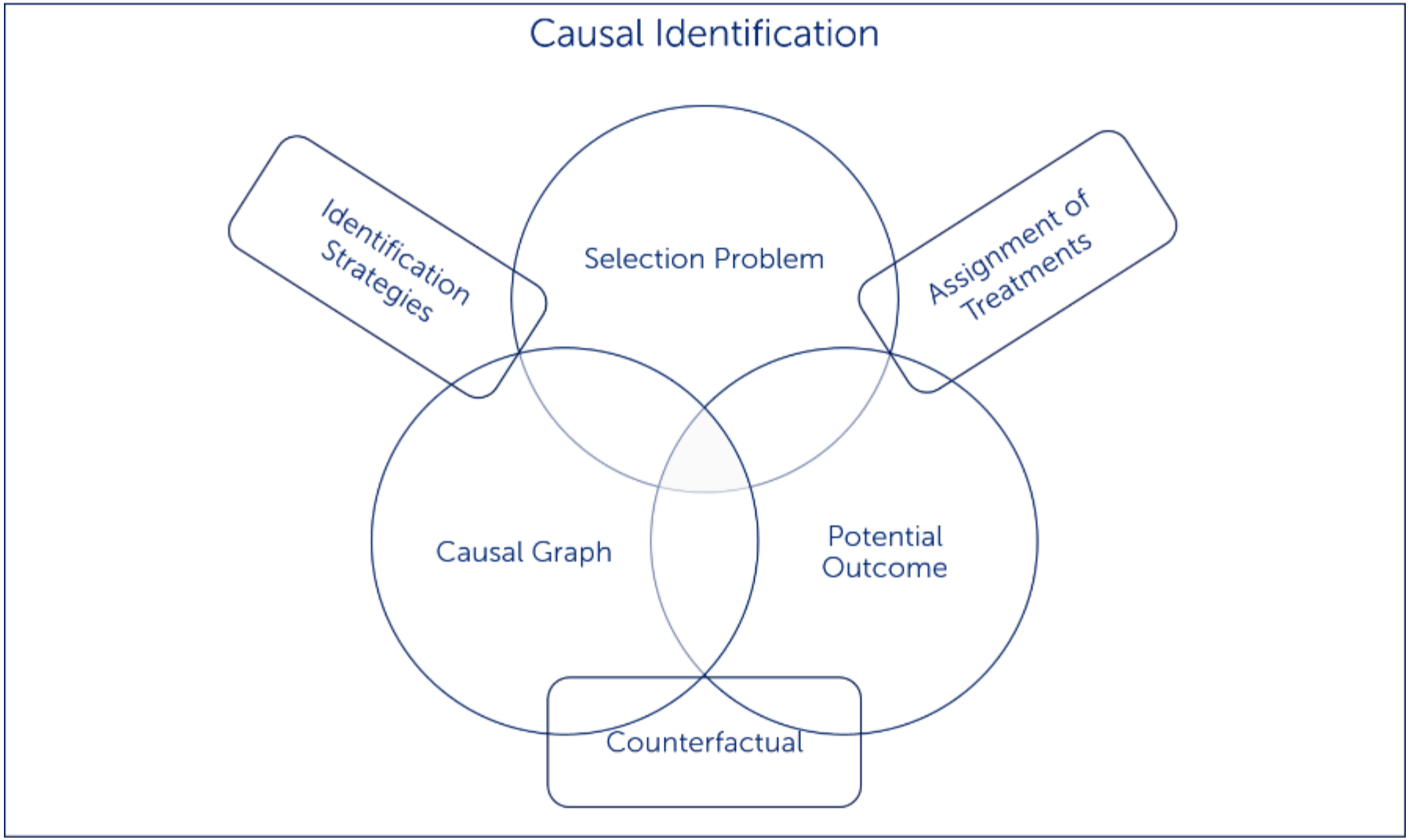
2

Break them down into fundamental blocks

3

Derive the methods / solutions

# Logic of the contents of the course

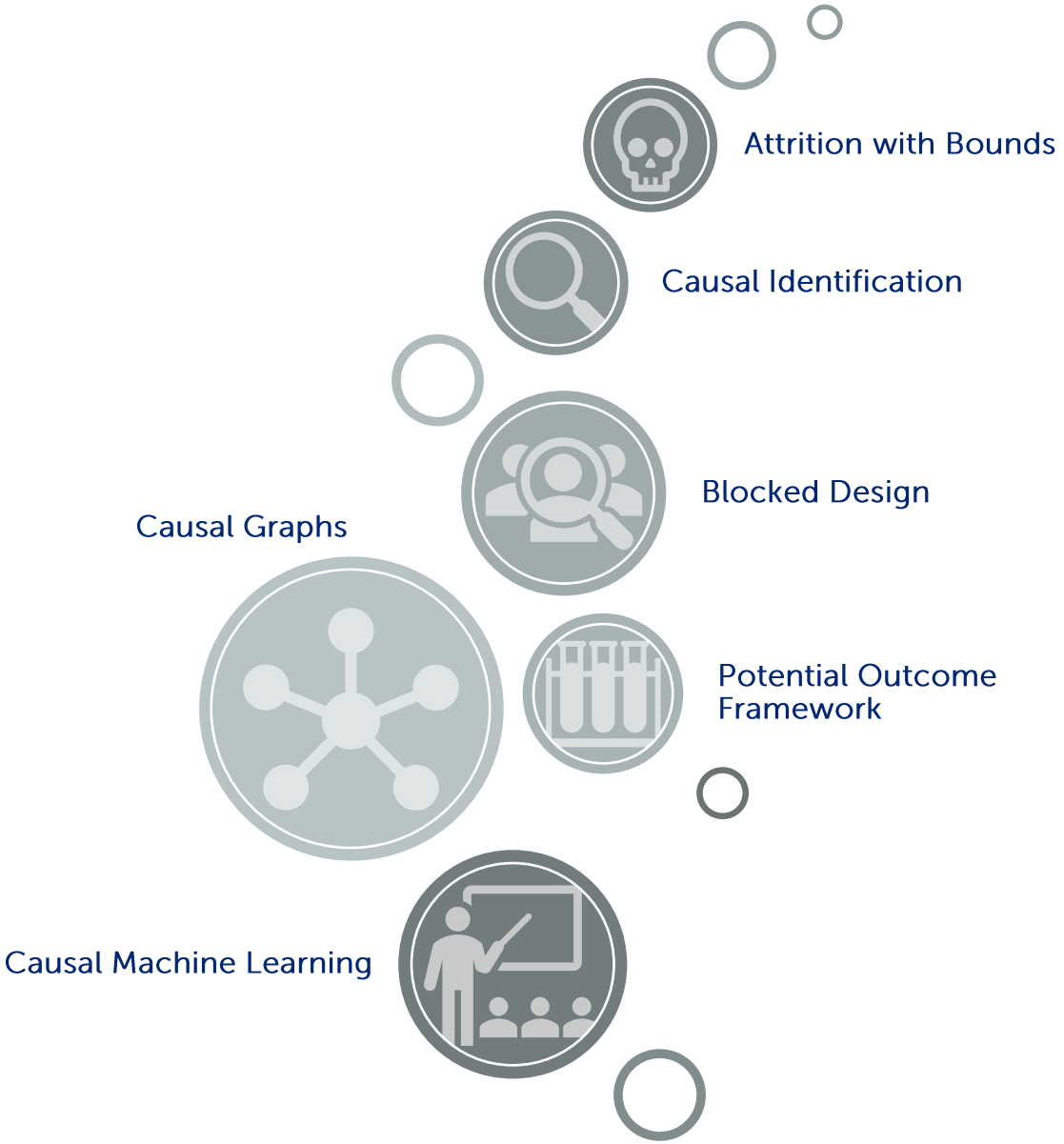



Statistical and Machine Learning Methods

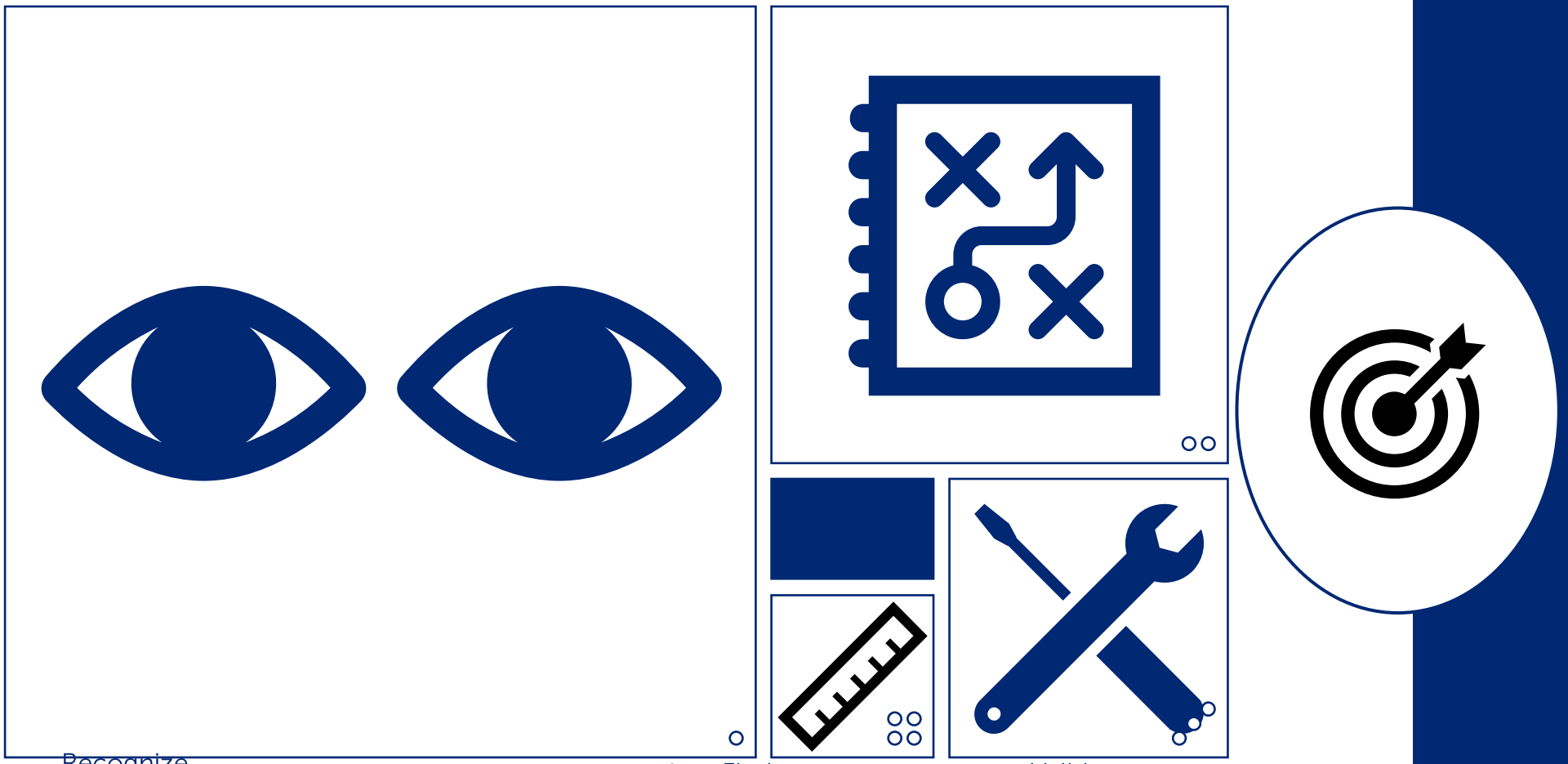
# Design logic of the course

| Topics | Controls |
| --- | --- |
| Completely randomized experiments | Full control<br>Full information |
| Matching and weighting | Partial control<br>Full information |
| Non-compliance in experiments | Partial control<br>Partial information |
| Attrition in experiments | Partial control<br>Partial information |
| Quasi-experiments | No control<br>Partial or no information |

# Some topics


Attrition with Bounds
Causal Identification
Blocked Design
Causal Graphs
Potential Outcome Framework
Causal Machine Learning

# “True” objectives



Recognize causal problems

Propose an identification strategy | Data + Theory

Find or create an appropriate method

Validate your results truthfully

# Blended learning
## Lectures

Lectures

- The basic concepts and understanding
- Systematic way of implementation
- Limitations of tools

Principles

- Principles of causal inference

# Blended learning
## Discussions

Discussion

- Lecturing is “dull”
- Of selected topics
- Based on papers

Active participation is important

- At least skim through the papers

# Blended learning

## Cases

Cases

- To apply the models to solve real problems
- Data- and model-driven
- Step-to-step guide for doing projects

# Course materials

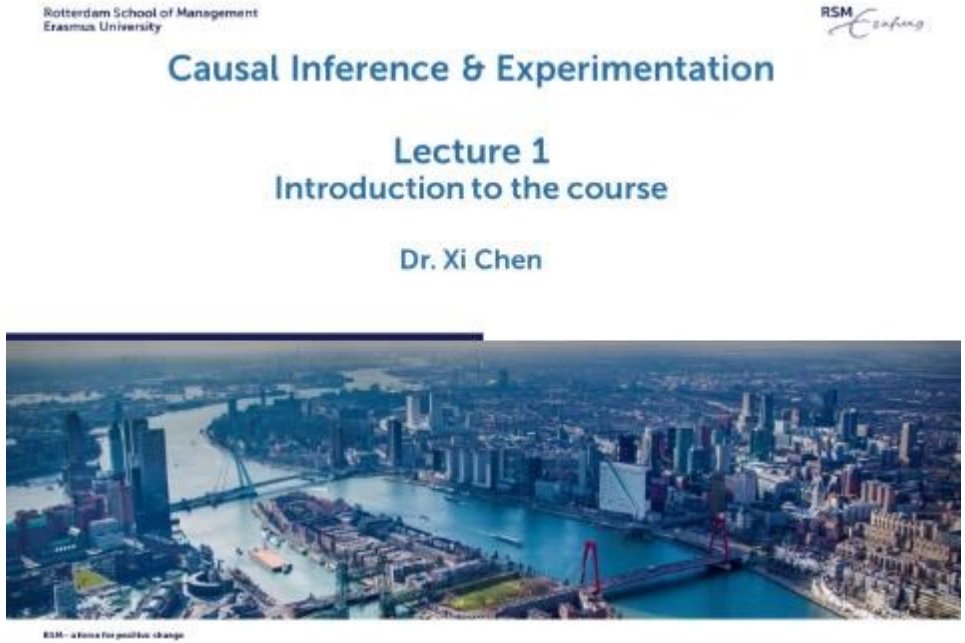
**Textbooks (see course manual)**

***(**optional, reference only)***

You do not have to read them if you do not feel like it.

**Slides**

**(most important)**



# Course materials

## Papers (for discussion, on Canvas)

***(**required)***

<u>You'd better read them or else…</u>


## Extended readings

*(**optional, if you are interested)*

At your leisure time…

# Assessment

## Integrative Assignment (three weeks)

- Online in the last week (data and questions on Canvas)
- Data analytics for causal inference problems
- 2 types of questions
  - Analytical & interpretation
  - Open conceptual or managerial questions
- Focus on problem-solving
  - If needed, pseudo-codes and analytical procedure will be given.

Any other questions?