Lecture 1 - Overview and fundamental hurdles

Topics in Econometrics

Vincent Bagilet

2025-09-11

Introduction and Steps of Econometric Analyses

Objective of the class

- Discuss **practical issues** and challenges that one may face when doing applied economics:

  - Help you be aware of some of them

  - Provide you with some tools to be able to spot others by yourself

- These hurdles may arise at any step of the research process

- *Topics* class: will not cover everything but instead pick points within topics

Steps of Applied Economics Analyses

- Define question/topic

- Find, get, wrangle and clean data

- Summary statistics

- Define an identification strategy

- Build a regression model

- Estimate your model

- Specification checks

- Additional inference

- Robustness checks

- Communicate

A More Structured Version

- **Design**: decisions of data collection and measurement

  - *eg*, decisions related to sample size and ensuring exogeneity of the treatment

- **Analysis**: estimation and questions of statistical inference

  - *eg* standard errors, hypothesis tests, and estimator properties

- **Modeling**: define statistical models

  - In between design and analysis

Outline of the class

1. Overview and fundamental hurdles

2. Simulations

3. Design: beyond identification

4. Design: identification (Fixed Effects and related)

5. Data visualization

6. Design: identification (IV and RDD)

7. Modelling

8. Analysis

Research questions

What is a good research question?

- It **can be answered**

  - There is some sort of objective answer

- It should **improve our understanding of the world**

  - Should **inform theory** in some way

  - Takes us from theory to an hypothesis (statement about what we will observe in the world)

- A solid econometric analysis only matters to the extent that you have a good research question (but the opposite might be true as well)

Example

- Impact of the size of motors of boats in Norway and cod catch under a catch cap

- Not that interesting in itself, is it? Would be more interesting if, for instance:

  - Look at this from a game theory and forced technological adoption perspective

  - Find a way to use this case to say something *new* or *different* on management of renewable natural resources

- Can produce radically different papers on the same topic and setting

- **Use theory to put light on your specific case *and* your specific case to inform theory**

Identifying a research question

- Can start with a research question/hypothesis or from theory

- **Or** can find a natural experiment and come up with a question

- Know your literature to identify **gaps**

- We are interested in **why** and not **what**

- **Avoid data mining**: it can help but to identify questions to test on **other** data sets

Is your research question good?

- **Relevance**: is it interesting, important or policy relevant ?

- **Potential results**: what would any result tell you about your theory?

- **Feasibility**: is the right data available?

- **Scale**: how much resources would you need?

- **Research design**: is there a good one that would allow you to answer your question?

- **Keep it simple**: avoid building several questions into one
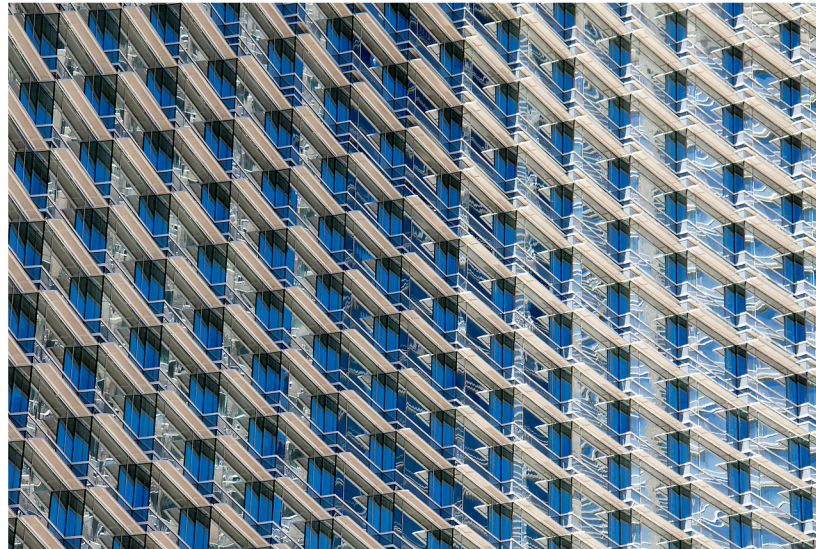
Logistics

Website

https://vincentbagilet.github.io/metrics_m2/

A typical lecture

1. I introduce concepts and intuition

2. We discuss a paper together (when reading assigned)

3. Some R coding together *and* on your own

Grading and assignments

Assignment	Percentage of final grade	Due date
Final report	30 %	November 7, 8pm
Final presentation	20 %	November 4, 8:30am
Participation	10 %	-
Replication	20 %	October 13, 8pm
Homework	20 %	Throughout

Final project

- In pairs

- Build a simulation to replicate an analysis you may do in your master thesis

- Generate realistic fake-data, run your analysis and discuss your results

- Project proposal, short report, presentation

Structure of the final project

- Structure it as short research paper (pitch):

    - Quick motivation and context

    - Research question

    - Data section: describe how you generate your data. Start very simple, complexify later

    - Modeling and analysis: describe your model, your choices and the outputs of your regressions

    - Discussion: what did you learn with this exercise
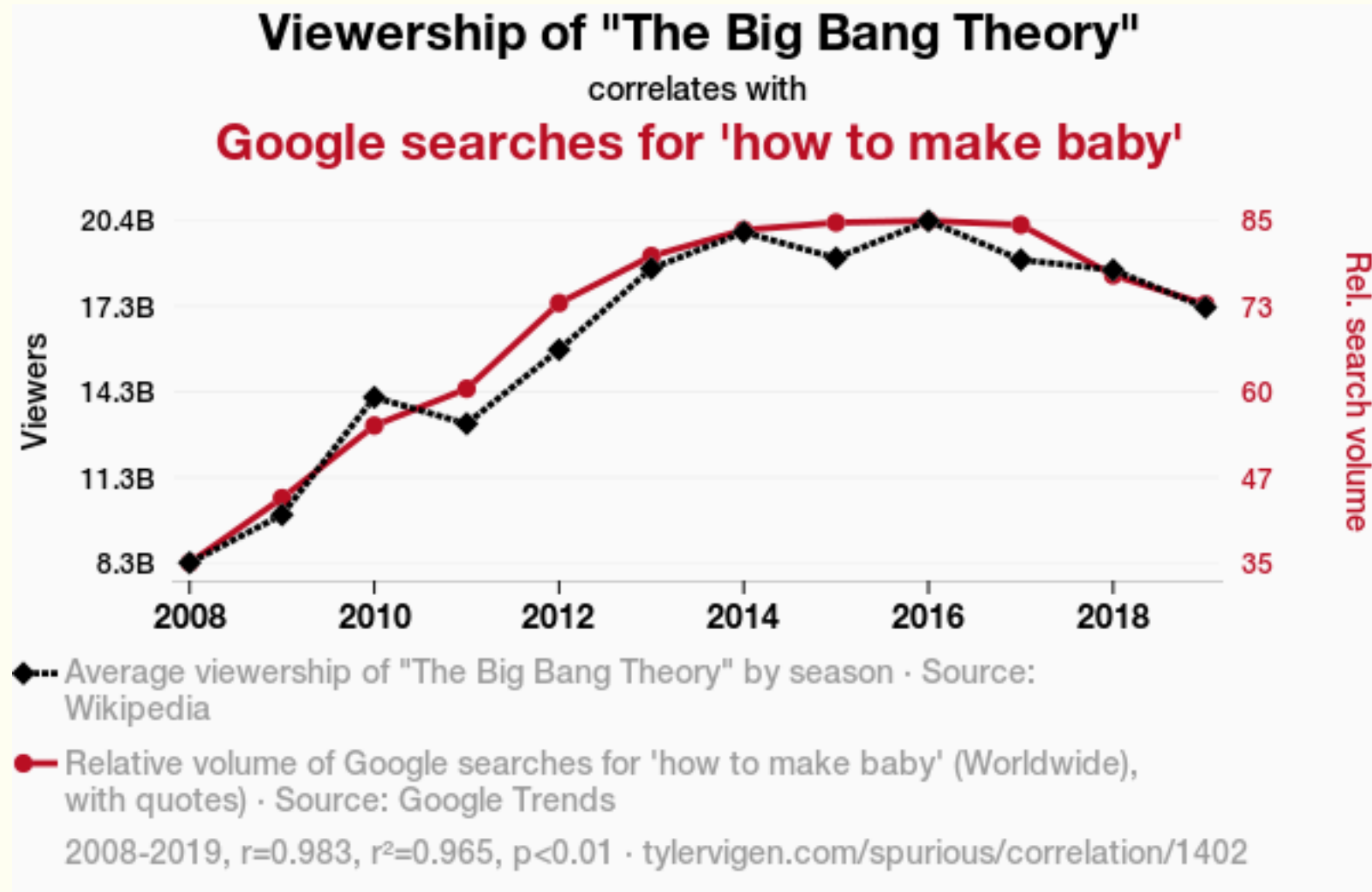
Homeworks (due before lecture)

1. -

2. Non-graded assignment

3. -

4. Graded assignment + reading

5. Graded assignment

6. Graded project proposal + reading

7. Replication + reading

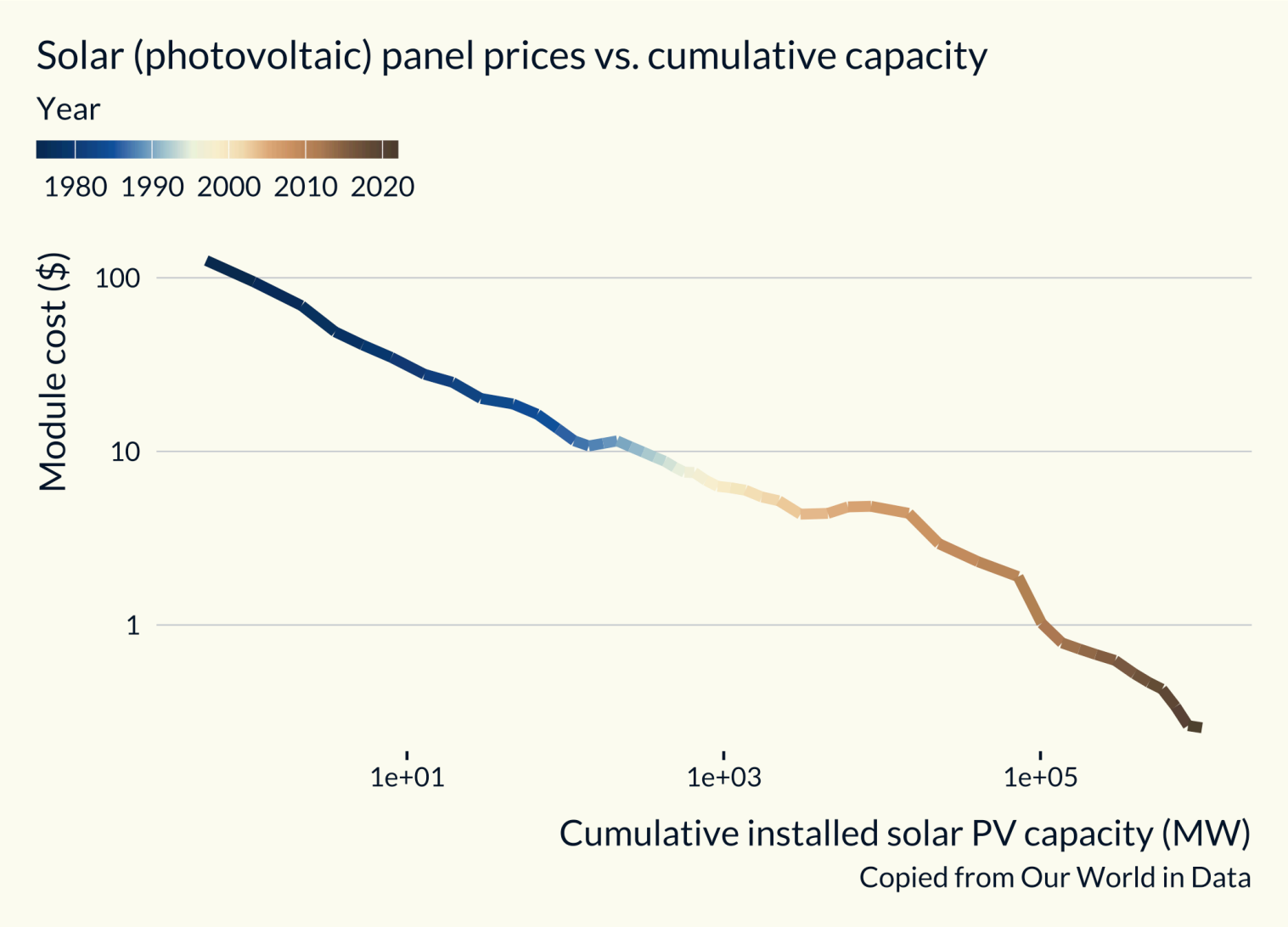8. Graded assignment + reading

Fundamental hurdles

Typical pitfalls in economics research

- Spurious correlation

- Reverse causality

- Confounders

- Model miss-specification

- External validity
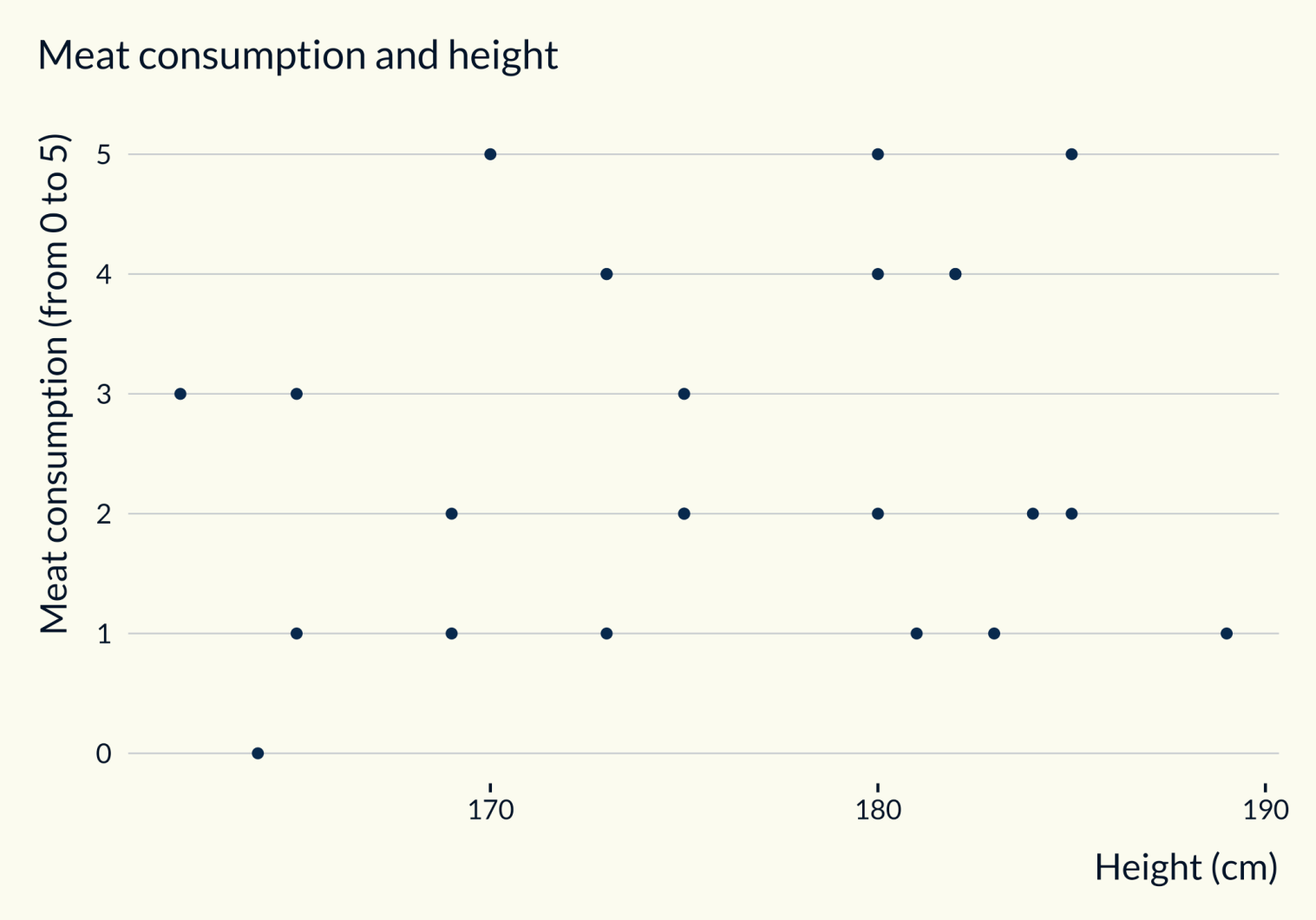
- Insufficient power

Spurious correlation

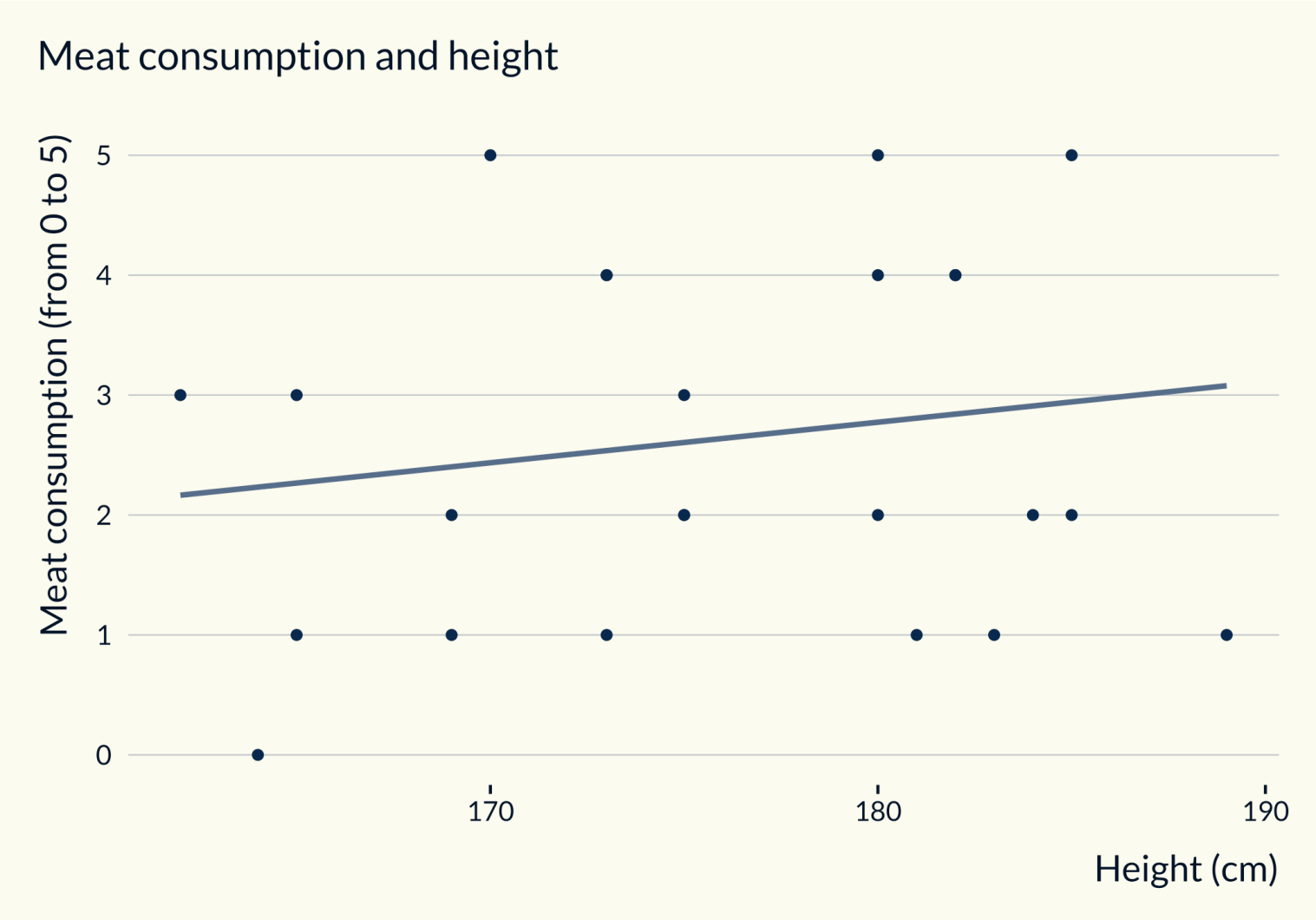Viewership of "The Big Bang Theory" correlates with Google searches for 'how to make baby'

◆ Average viewership of "The Big Bang Theory" by season · Source: Wikipedia

● Relative volume of Google searches for 'how to make baby' (Worldwide, with quotes) · Source: Google Trends

2008-2019, r=0.983, r²=0.965, p<0.01 · tylervigen.com/spurious/correlation/1402

Reverse causality

Solar (photovoltaic) panel prices vs. cumulative capacity

Copied from Our World in Data

Confounders

https://forms.gle/tHcTYqaKPeDTAUn56

Confounders

Meat consumption and height

Confounders

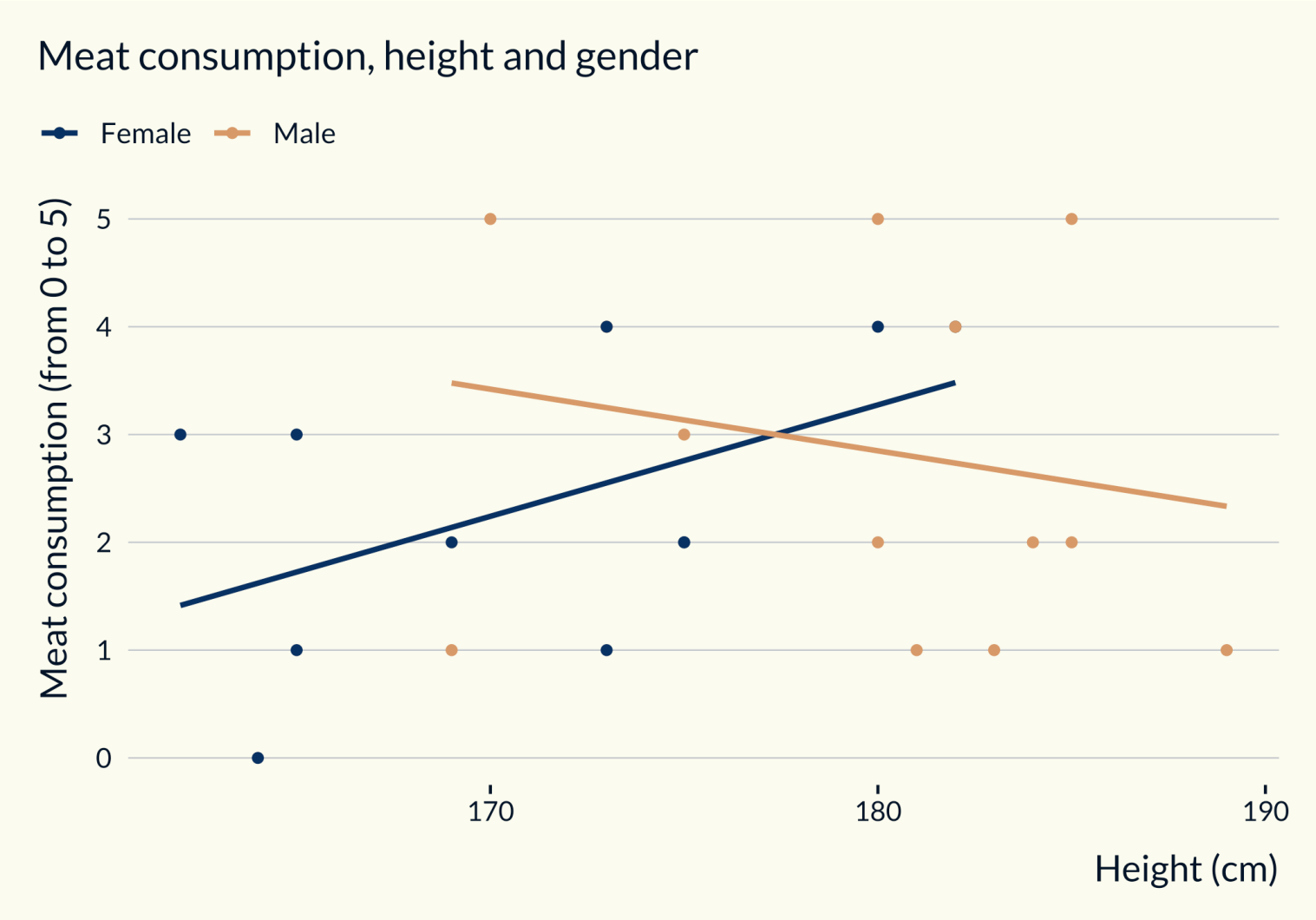Meat consumption and height

Confounders

Meat consumption, height and gender
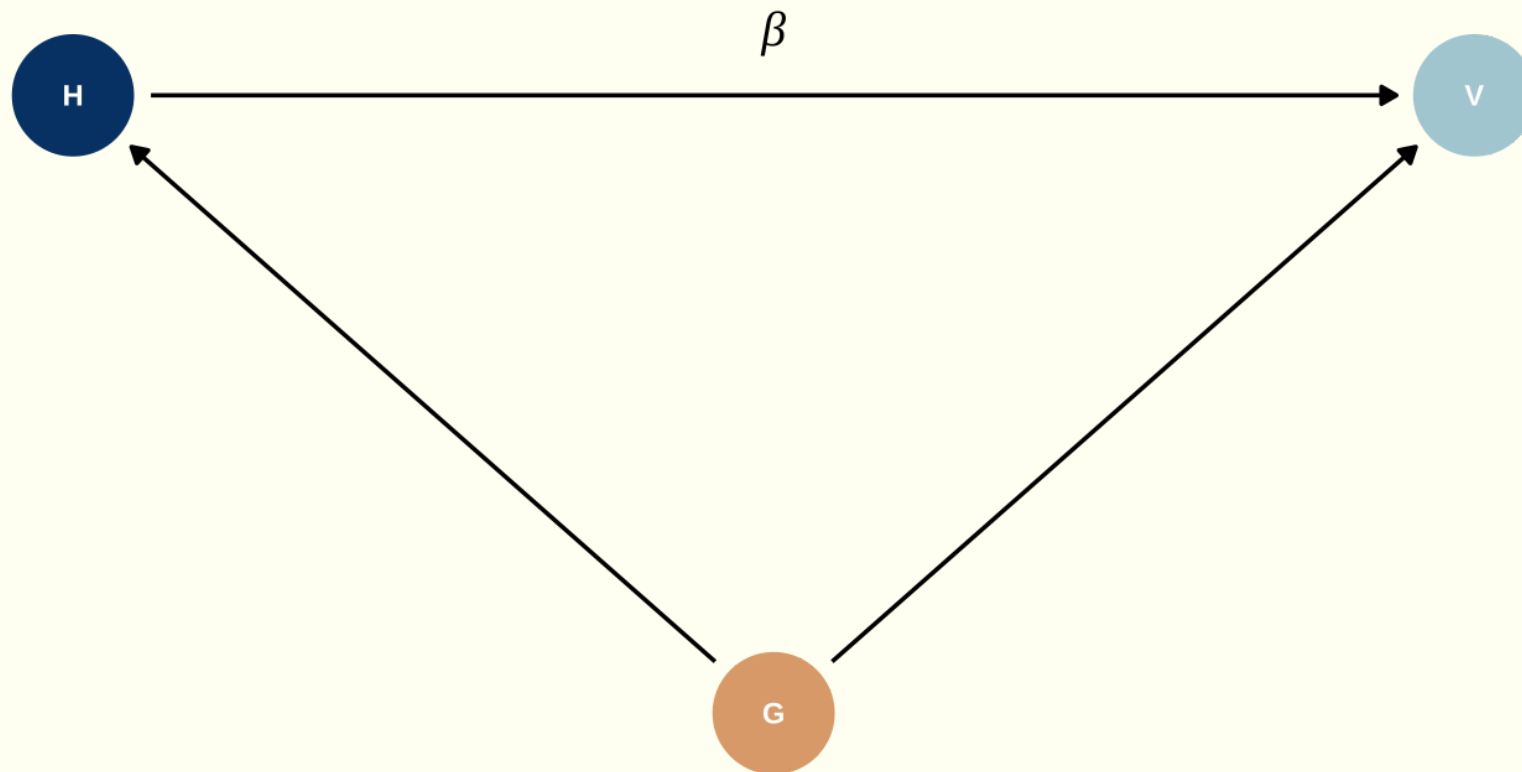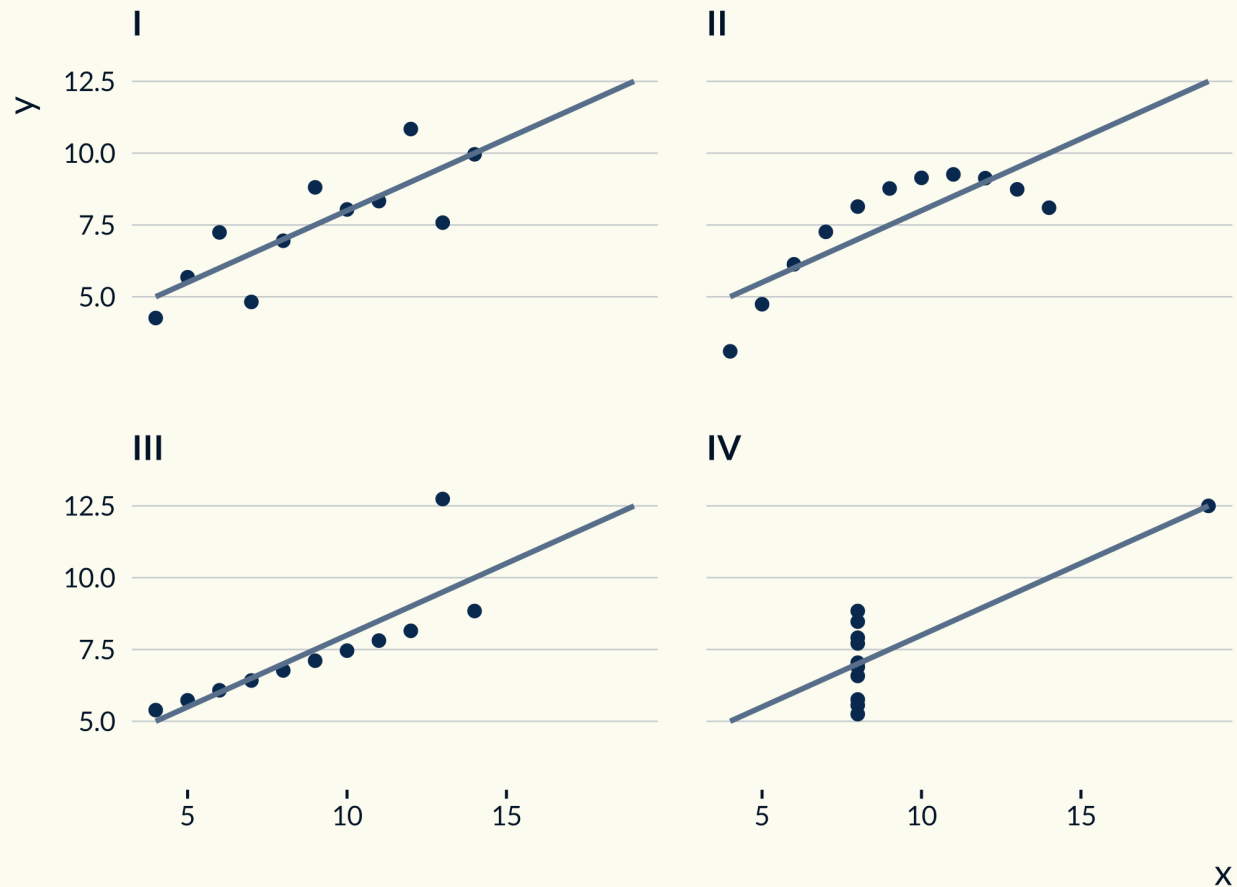
Confounders

Model miss-specification

Anscombe's quartet

Examples of more subtle hurdles

- Bad controls

- Leverage and outliers

- Measurement error

- Clustering level

- How do identification strategies actually work (*eg* FE and TWFE)

- More complex identification strategies (*eg*, shift-share)

Avoiding hurdles

How to avoid hurdles?

Learn, understand metrics and applied research

The Barque of Dante by Manet, after a painting by Delacroix

Replication, a helpful learning tool

How to avoid hurdles?

Derive the maths

- Sometimes relatively straightforward and very illuminating

- *eg* drivers of the variance of your estimator: $\mathbb{V}_{\hat{\beta}} = \dfrac{\sigma_u^2}{n\sigma_x^2}$

- Deriving the maths can be more complex and time consuming

How to avoid hurdles?

Simulations can help

- Super easy to implement simple simulations

- Can be informative of **what does not work**

- Can help you identify **where the issue comes from**

- We will discuss that with an example in a second

How to avoid hurdles?

Implement checks

- Check if the model seems to represent the DGP

- Check if our identification hypotheses seem to hold

- Check if the hypotheses for estimation seem to hold

Look at the consequences if this does not hold

- Robustness checks

- Evaluate the design retrospectively

Objectives for this class

- Build a mindful mindset

- Help you be aware of some of them

- Provide you with some tools to be able to spot others by yourself

- Learn how to implement simulations

Simulations

Usefulness through an example

A simple example: OVB

- How does an **omitted variable** affect our point estimate of interest? Why?

- Under which condition is an omitted variable an issue

- How does it affect the point estimate? The s.e.?

- How does that vary with various parameters? *eg* correlation between variable (sign and magnitude)

- Start very simple and complexify the process

- Let's move to R

Lecture summary

What did we do today?

- Discussed the structure of applied econometric research and where we may encounter hurdles

- Discussed logistics

- Reviewed some of these common hurdles encountered in applied research

- How to implement a simple simulation to understand the impact of an omitted variable

What did you learned, liked, disliked?

Thank you!