Methods for Causal Inference
Lecture 1

Ava Khamseh
School of Informatics

2021-2022

Logistics

These lectures are being recorded.

- Lectures: **Mondays** and **Thursdays** at **10:00-11:00am**
- Tutorials: Every other week **Wednesdays 12:00-1:00pm** 40GS_LG.07 Teaching Studio, **first session: 26/01/2022**
- Slides and recordings will appear on Learn
- Office Hours: Wednesdays 15:00-17:00
- Email me any questions, happy to discuss!

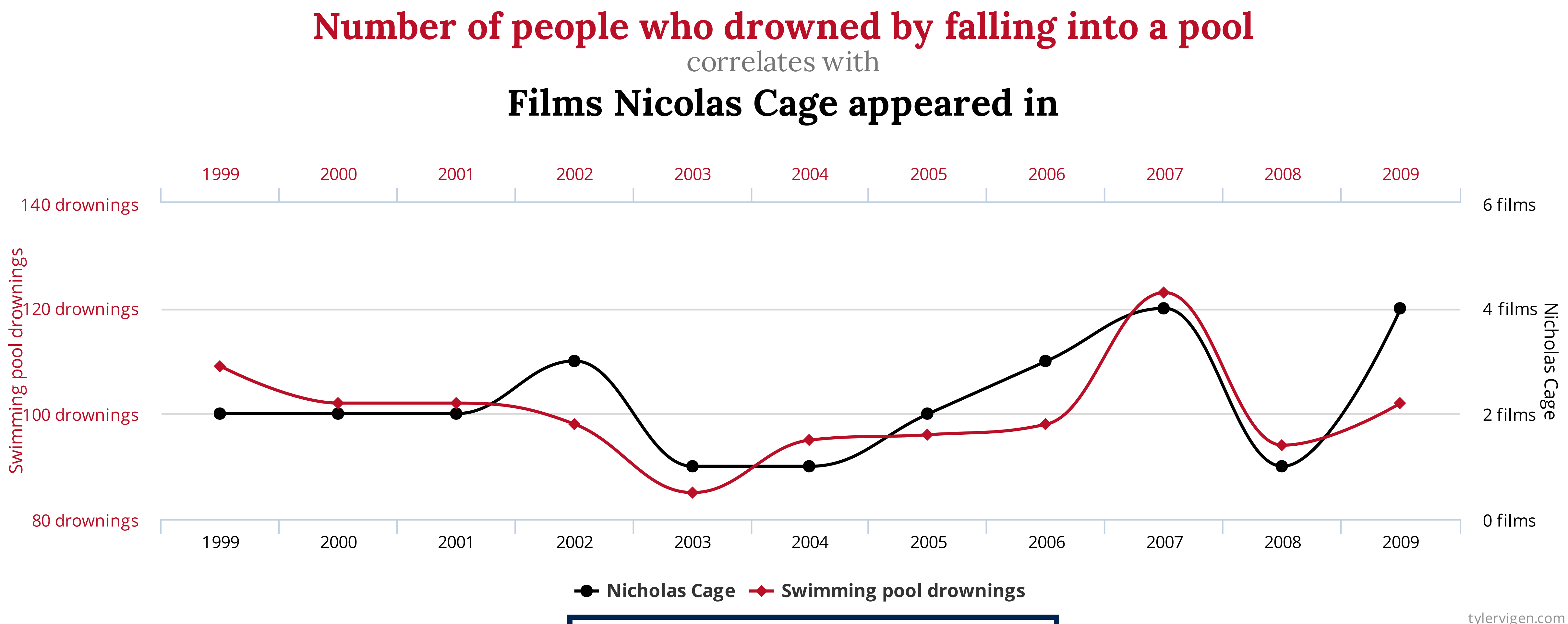
References

- Causal Inference in Statistics: A Primer (Pearl, Glymour, Jewell, 2016)
- More advanced: Causality (Pearl, 2009)
- Elements of Causal Inference: Foundations and Learning Algorithms (Peters, Janzing and Schölkopfk)
- Many other papers from the literature … (will be referenced)

“Correlation does not imply causation”

“Correlation does not imply causation”

Spurious correlation (random coincidence)

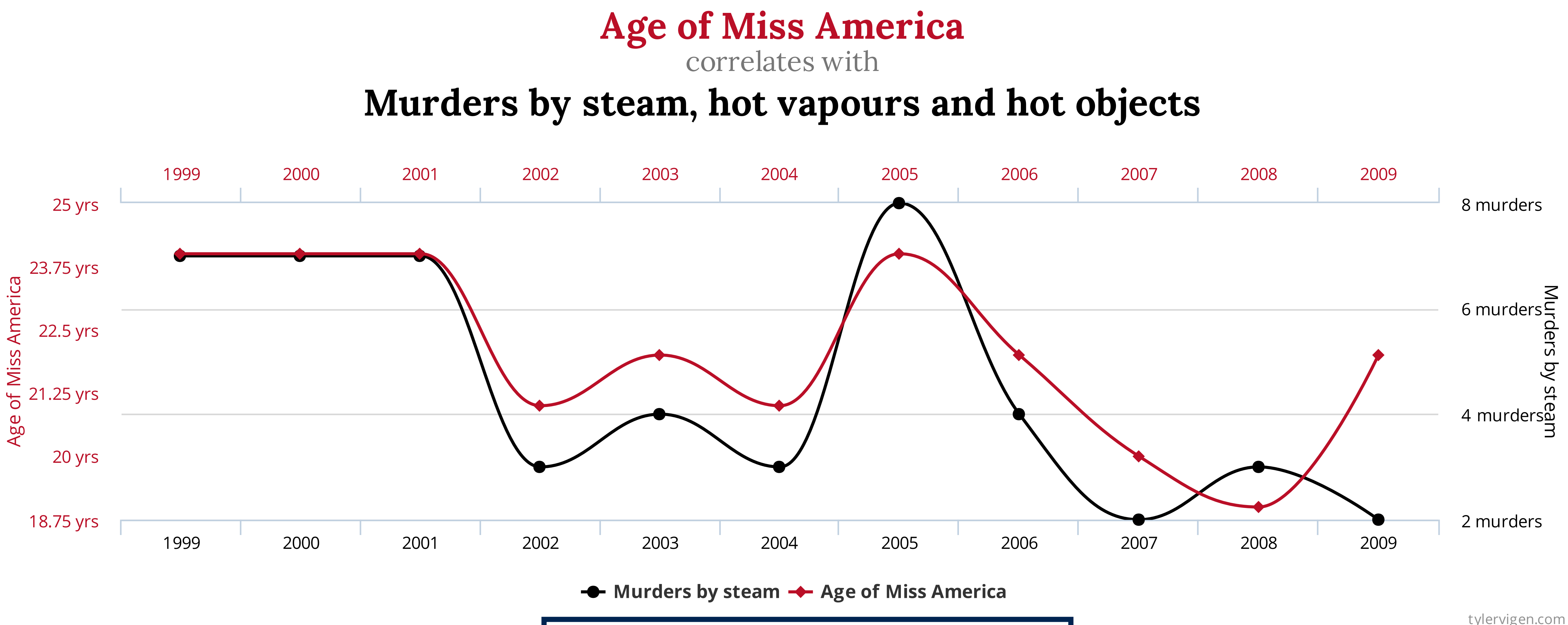


Correlation: 66.6% (r=0.666)

tylervigen.com/spurious-correlations

“Correlation does not imply causation”

Spurious correlation (random coincidence)

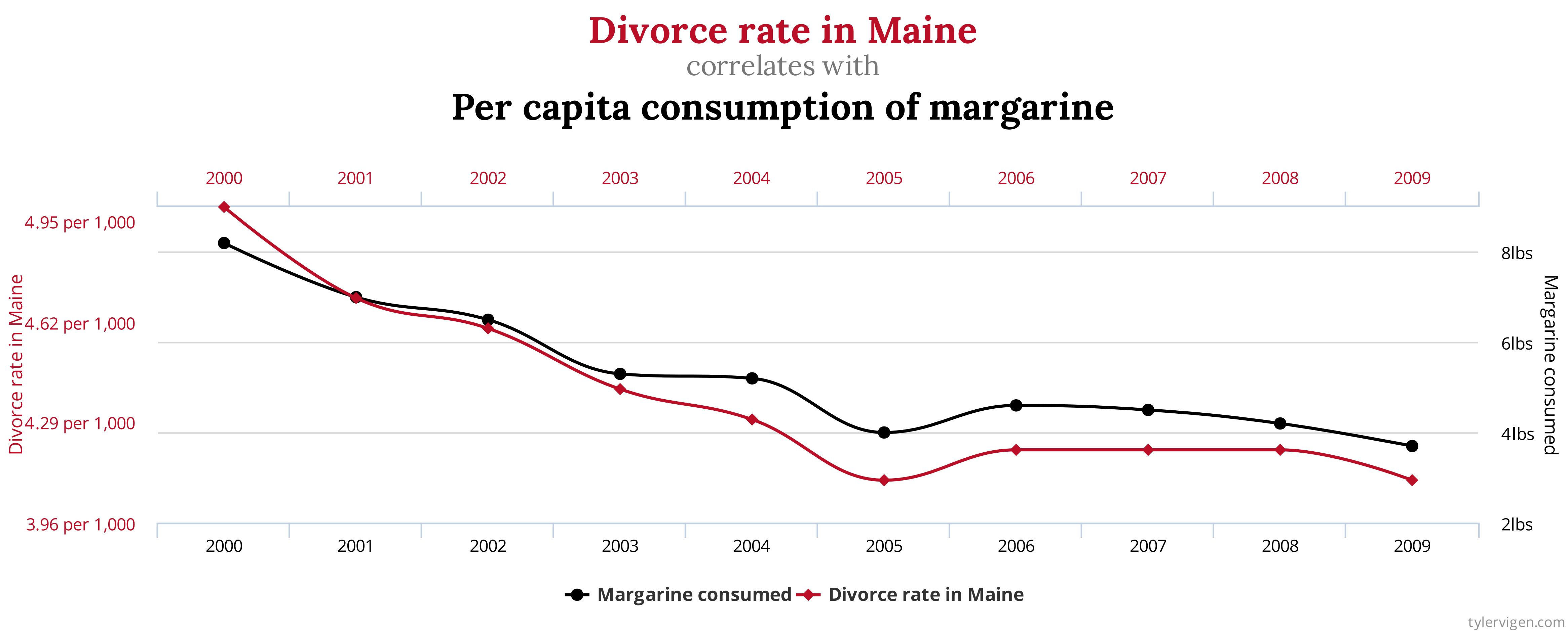


Correlation: 87.01% (r=0.8701)

tylervigen.com/spurious-correlations

“Correlation does not imply causation”

Spurious correlation (random coincidence)



Correlation: 99.26% (r=0.9926)

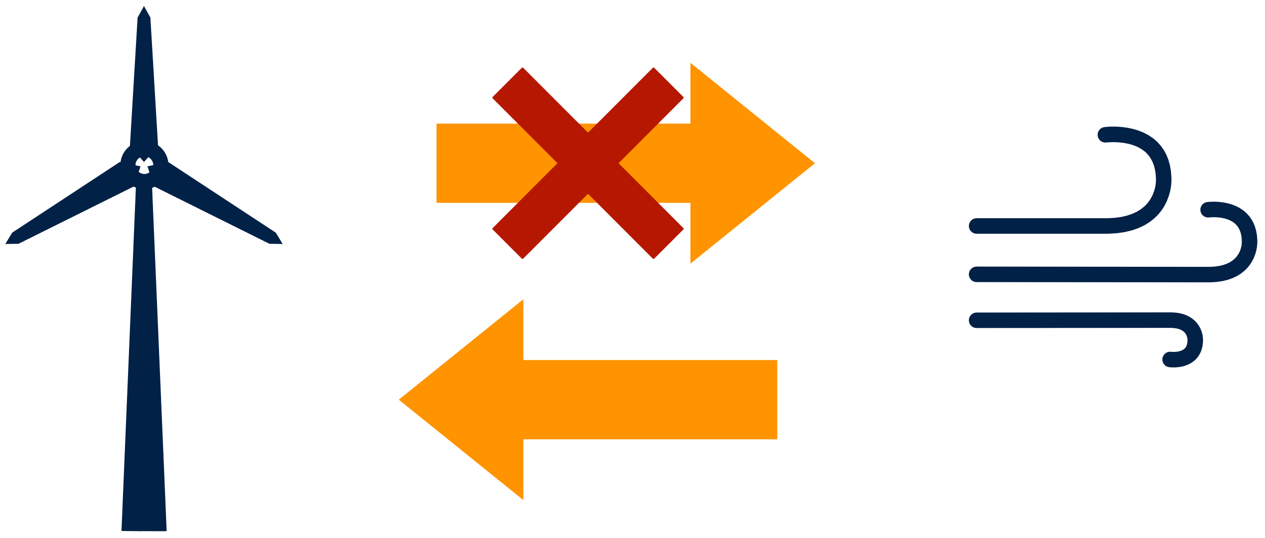
tylervigen.com/spurious-correlations

“Correlation does not imply causation”

Reverse causation:

The faster the wind-turbine rotates, the more wind is observed.

Therefore, rotation of turbines is the cause for winds!



“Correlation does not imply causation”

Circular/bidirectional cause and consequence:

Hours spent on Netflix and weight gain

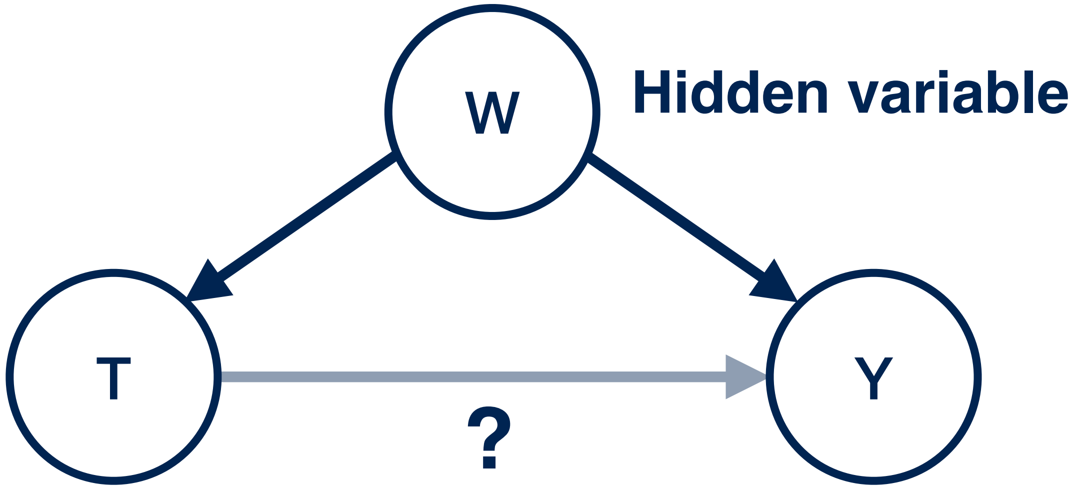
Hours spent on Netflix -> Less activity -> increase in weight

Weight gain -> exercising becomes harder -> more time online as hobby

“Correlation does not imply causation”

Confounding factor:

- Fever is not a cause of sneezing, they are both symptoms of flu (No arrow)
- Treatment outcome relationship confounded by age



Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention:
  “What if I were to act differently?”

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: "What if I were to act differently?"

- Controversial examples:
- **Biomedical**: "Vaccines lead to autism"

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: "What if I were to act differently?"

- Controversial examples:
- **Biomedical**: "Vaccines lead to autism"
- **Political/Economical**: "increases minimum wage, increases unemployment (people become lazy)"

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: “What if I were to act differently?”

- Examples:

- **Biomedical**: What drug, what dose, when, how often, … (see later)

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: "What if I were to act differently?"

- Examples:

- **Biomedical**: What drug, what dose, when, how often, … (see later)
- **Political**: How social media posts from famous individuals (e.g. celebrities, ex-political figures, etc.) influence elections

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: "What if I were to act differently?"

- Examples:

- **Biomedical**: What drug, what dose, when, how often, … (see later)
- **Political**: How social media posts from famous individuals (e.g. celebrities, ex-political figures, etc.) influence elections
- **Environmental**: Is the constant energy consumption in region X due to the regions's energy efficiency standards or due to its mild climate, demographics

Why should we care about **causation**?

- To guide **actions** and **policies**:
- To understand *how* and *why* **interventions** affect outcomes
- **Predict** what would have happened under a different intervention: "What if I were to act differently?"

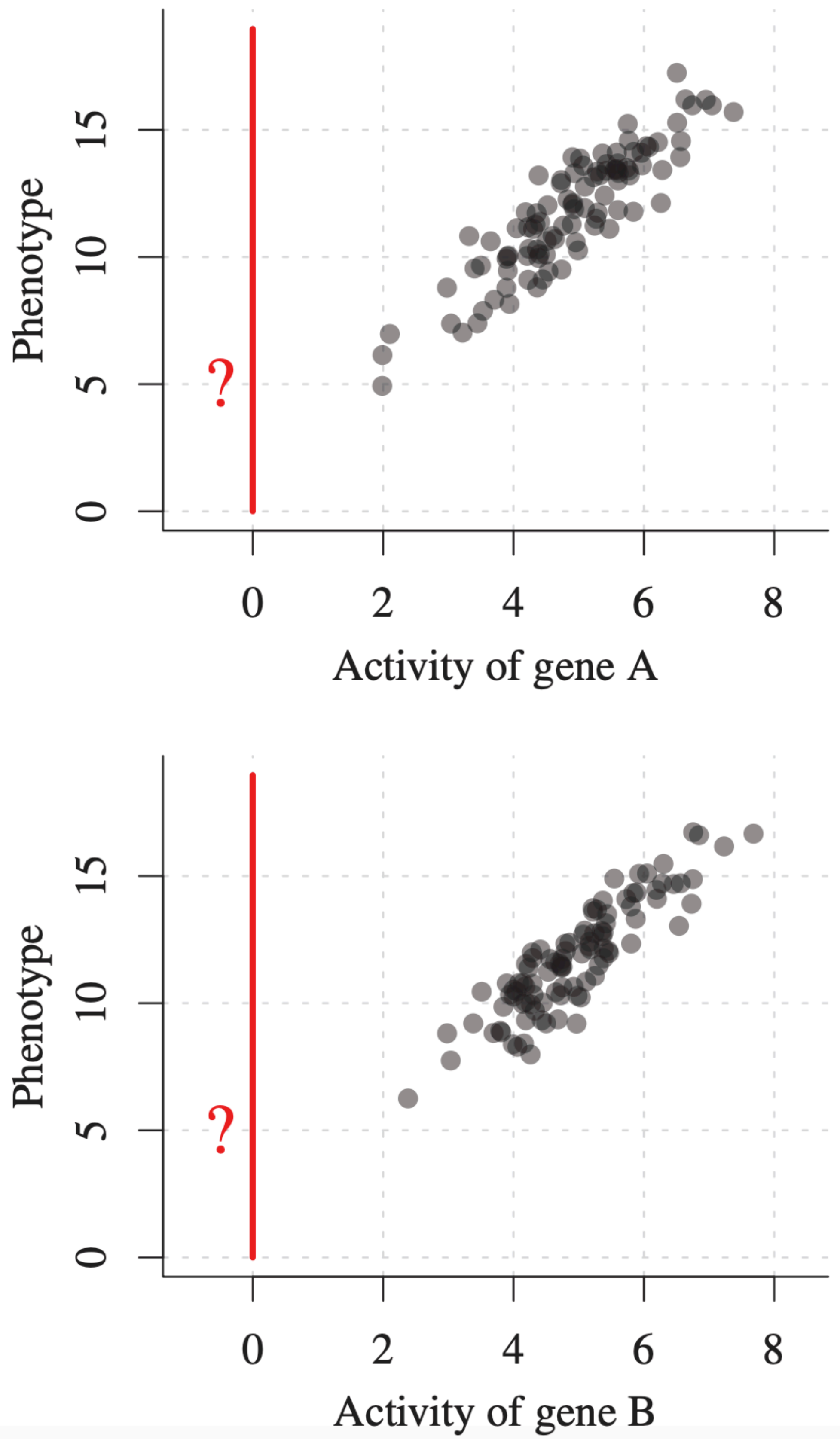
- Examples:

- **Biomedical**: What drug, what dose, when, how often, … (see later)
- **Political**: How social media posts from famous individuals (e.g. celebrities, ex-political figures, etc.) influence elections
- **Environmental**: Is the constant energy consumption in region X due to the regions's energy efficiency standards or due to its mild climate, demographics
- **Education**: People with feature X are more likely to obtain an internship in tech

Personalised Medicine

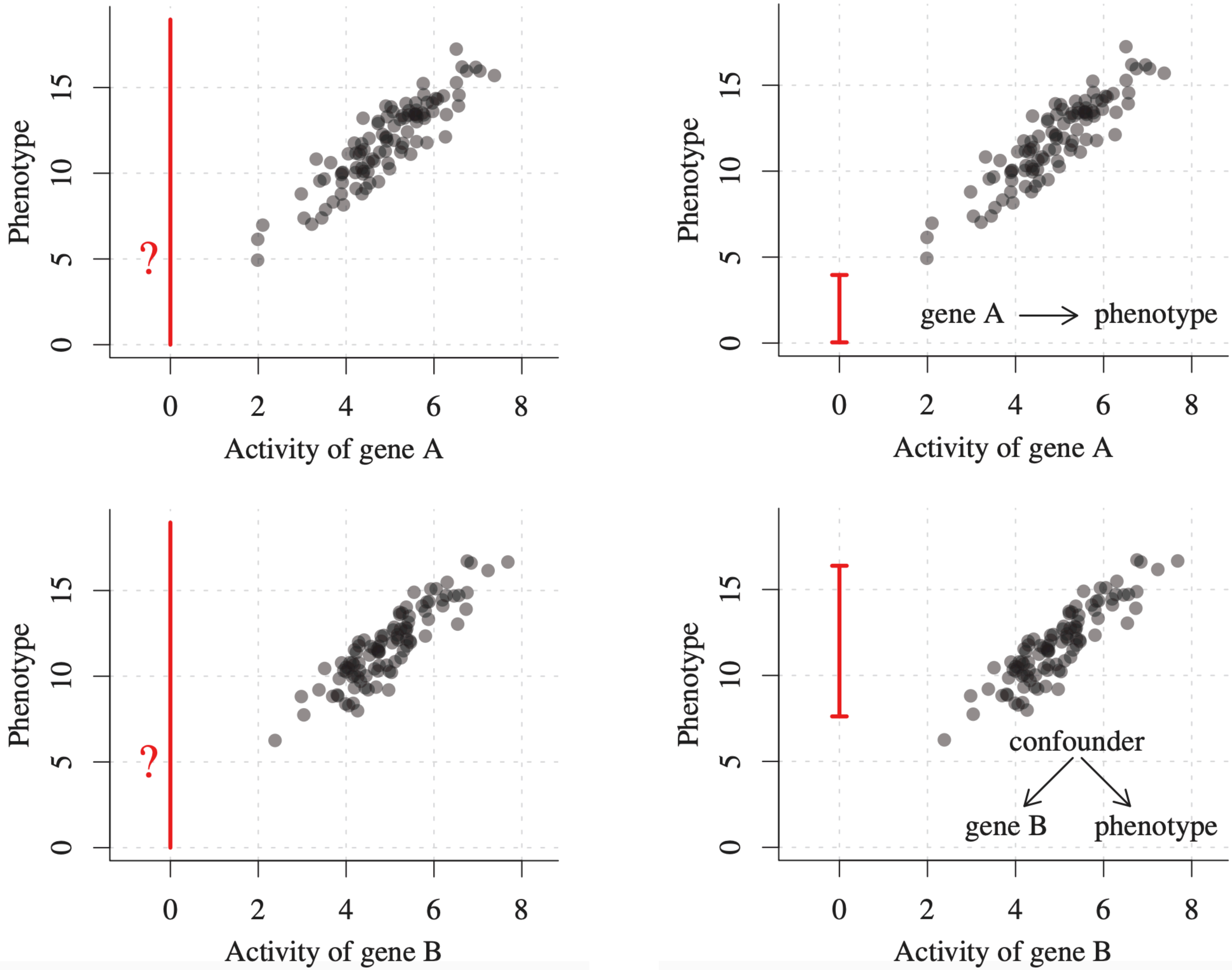
- Patient diagnosed with a particular **disease**
- Certain **baseline covariates** are known, e.g. age, weight, BMI, blood sugar, …
- **Question:** Should **treatment A** or **treatment B** be given
    - What is the causal effect of A vs B
    - Design a **policy**: Features —> {A,B}
    - i.e. best treatment for **a given individual**
- Source: **Electronic Health Records**

Gene Perturbation



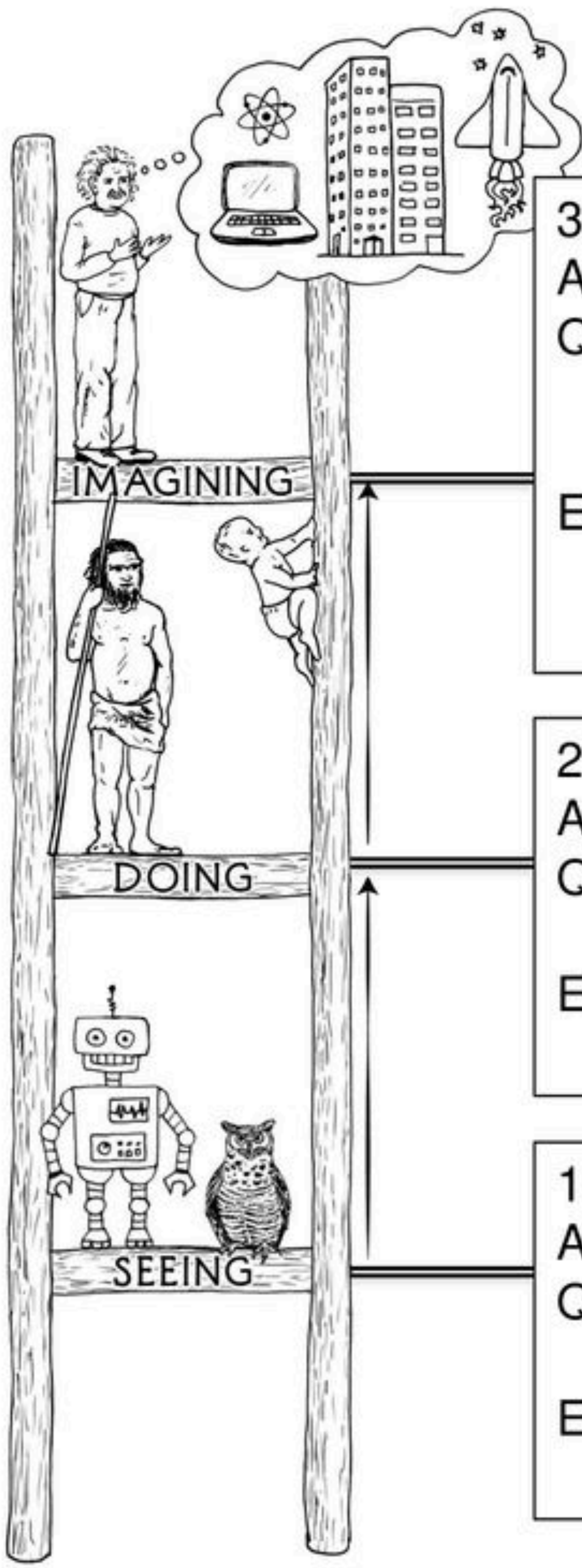
Elements of Causal Inference, Peters et al.

Gene Perturbation



Elements of Causal Inference, Peters et al.

Pearl’s ladder of causation



3-LEVEL HIERARCHY

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 the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do . . . ? How?*
(What would Y be if I do X?)

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

1. ASSOCIATION

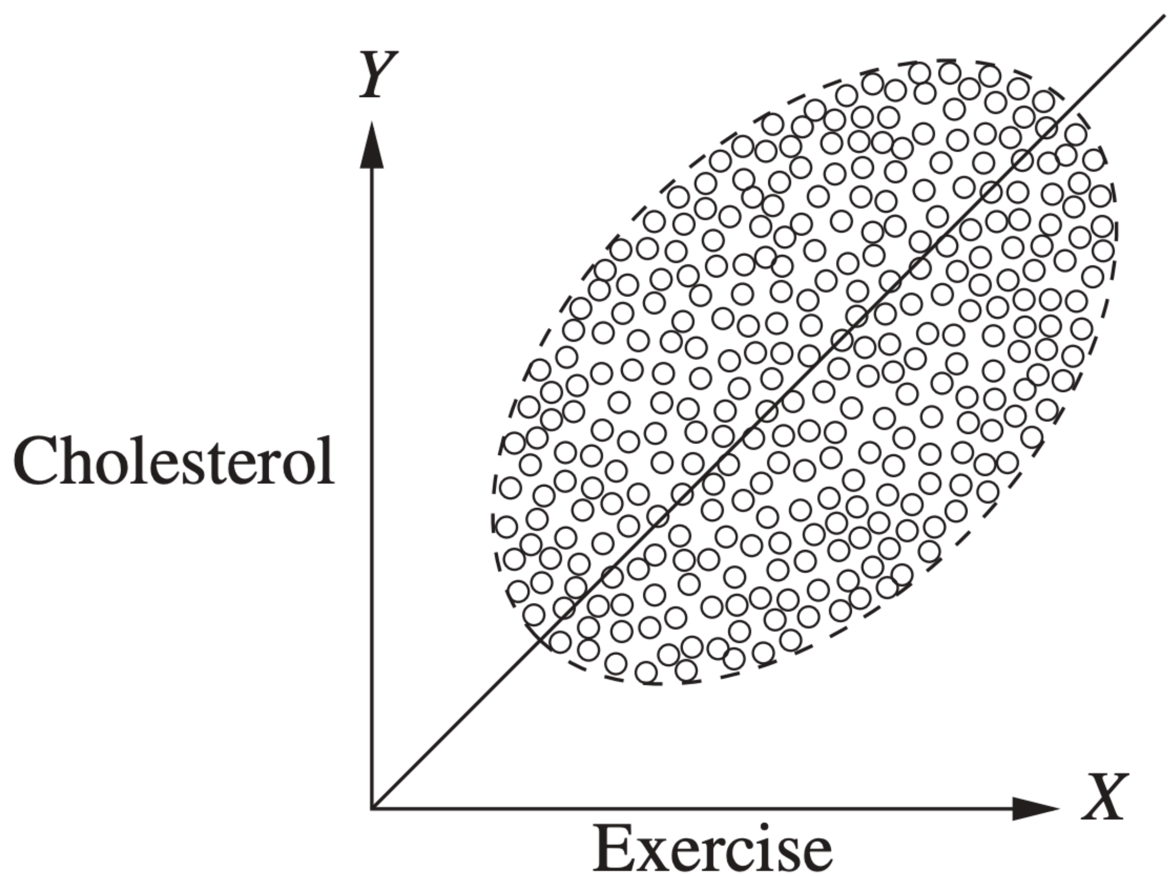
ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see . . . ?*
(How 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?

Simpson's Paradox

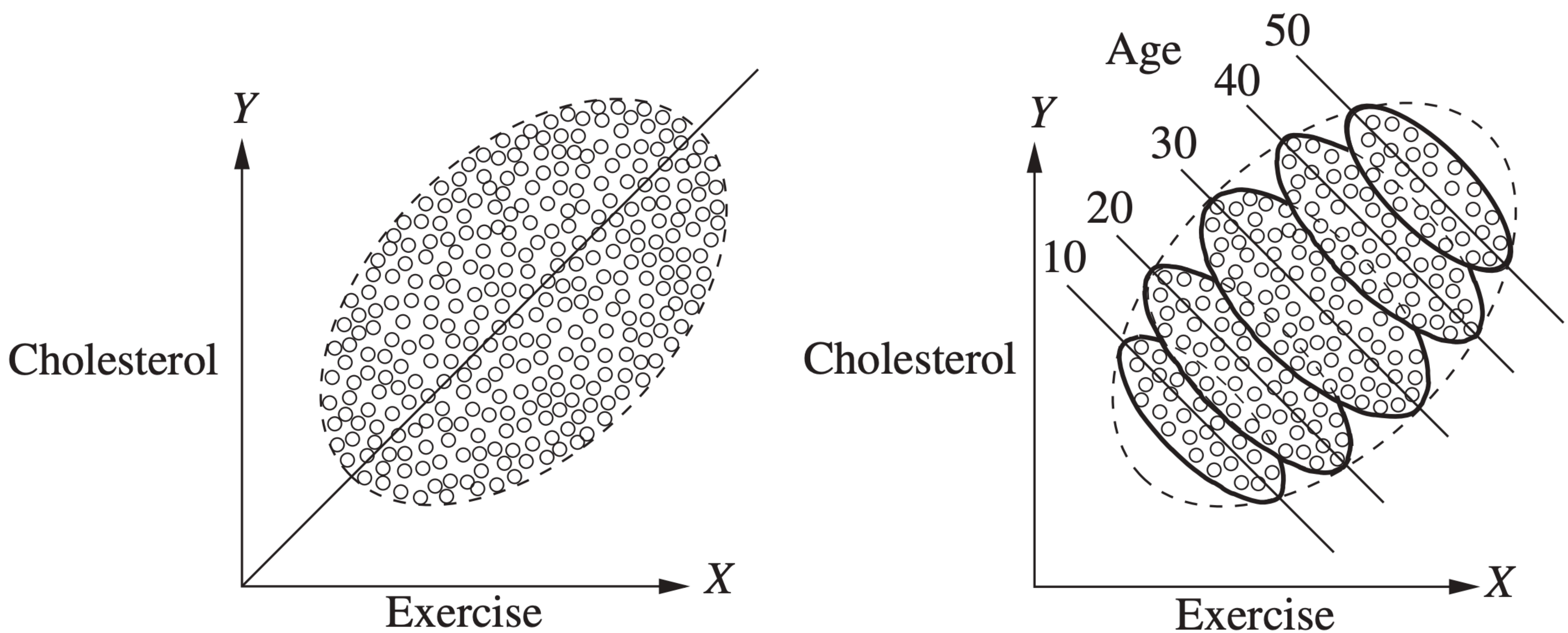
- Why concluding causality from purely associational measures, i.e. correlation, can be **very wrong** (not just neutral): "It would have better not to make any statements!"



Causal Inference in Statistics, Pearl (2016)

Simpson's Paradox

- Why concluding causality from purely associational measures, i.e. correlation, can be **very wrong** (not just neutral): "It would have better not to make any statements!"



Causal Inference in Statistics, Pearl (2016)

Example with numbers

We will come back to this example after having built the causal calculus how the causal effect can be detected accurately (without us suffering …)

Language of causality and the roles of variables

“What intervention”, “how much”, “when”, “how often”, “Control”, “effect of”, “why did”, “what if”, …

Causality language

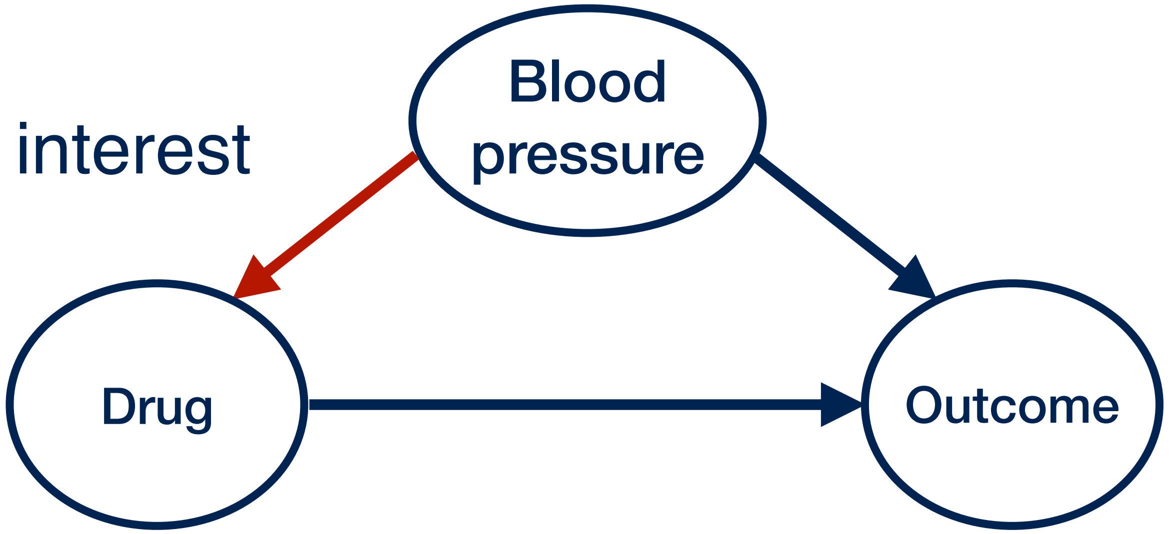
Consider all variables affecting the system of interest and **the role each plays.**

Patient: Info on DNA variants and biomarkers, traits/disease, confounders
Clinician: What drug, what dose, when, how often, …

Pre-treatment

Consider all variables affecting the system of interest and **the role each play** (as far as possible)

Blood pressure is a **confounder** here:



Language of causality and the roles of variables

“What intervention”, “how much”, “when”, “how often”, “Control”, “effect of”, “why did”, “what if”, …

Causality language

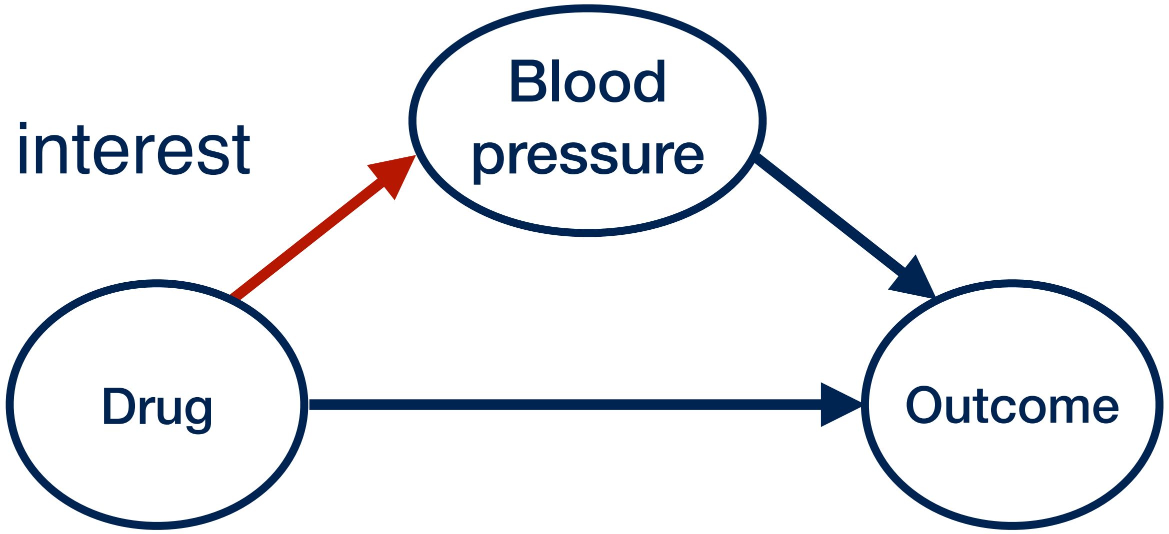
Consider all variables affecting the system of interest and **the role each plays.**

Patient: Info on DNA variants and biomarkers, traits/disease, confounders
Clinician: What drug, what dose, when, how often, …

Post-treatment

Consider all variables affecting the system of interest and **the role each play** (as far as possible)

Blood pressure is a **mediator** here:



Language of causality and the roles of variables

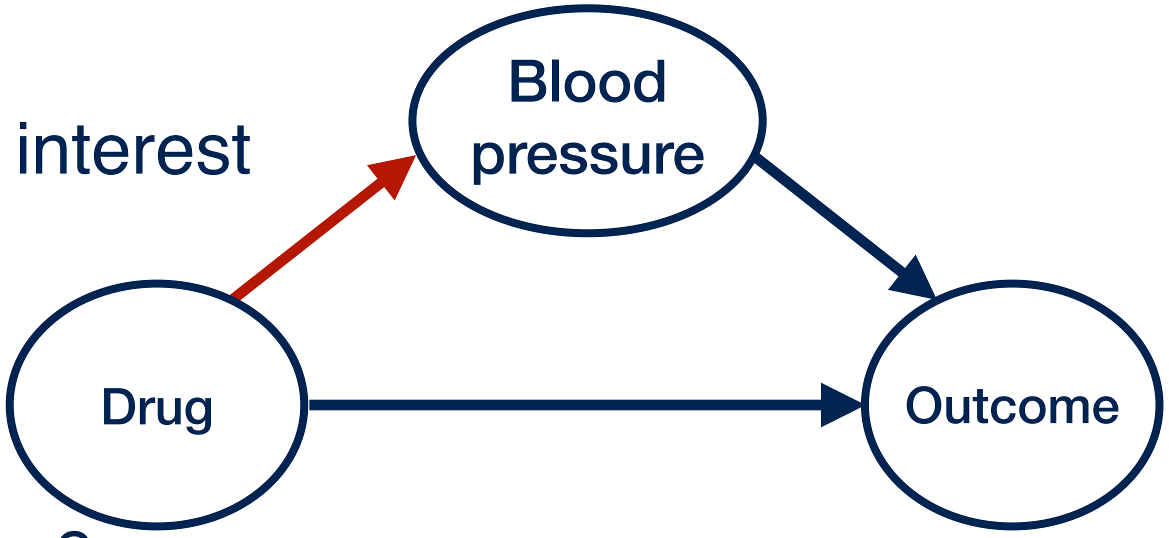
“What intervention”, “how much”, “when”, “how often”, “Control”, “effect of”, “why did”, “what if”, …

Causality language

Consider all variables affecting the system of interest and **the role each plays.**

Patient: Info on DNA variants and biomarkers, traits/disease, confounders
Clinician: What drug, what dose, when, how often, …

Consider all variables affecting the system of interest and **the role each play** (as far as possible)



Blood pressure is a **mediator** here:
What happens when there are lots of variables?

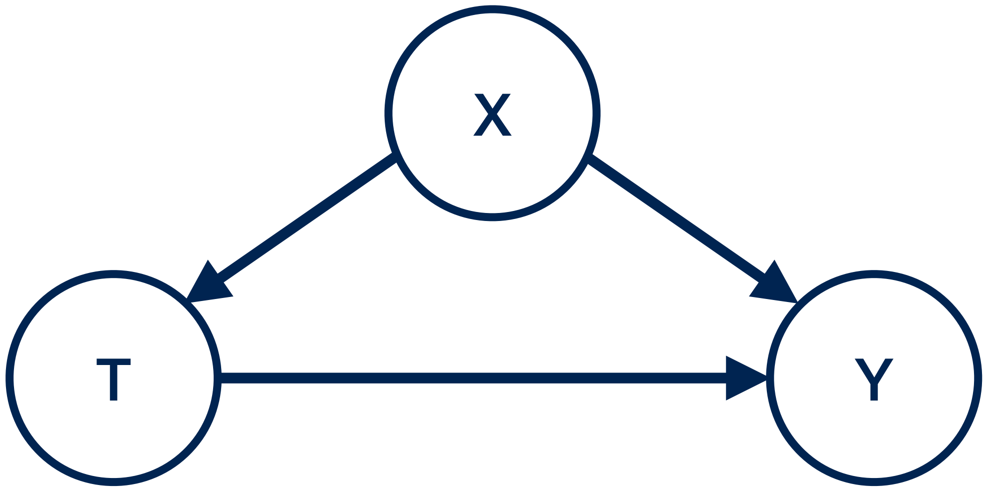
Conventions

- Variable to be manipulated: **treatment (T)**, e.g. drug

- Variable we observe as response: **outcome (Y)**, e.g. success/ failure of drug

- Other observable variables that can affect treatment and outcome causally and we wish to correct for: **confounders (X)**, e.g. age, sex, socio-economic status, …

- Unobservable confounder **(U)**



Causal Estimation of Effects

- Have a prior causal knowledge (may be incomplete) and know the treatment/outcome pair, c.e., weight gain, hours online

- Interested in estimating the **effect size**:

$$\mathbb{E}\big[y_{t=1}(x) - y_{t=0}(x)\big] = \int \big(y_1(x) - y_0(x)\big)p(x)dx$$

Note: The features/confounders x for both treatment and control groups are drawn from the **<u>same</u>** distribution p(x)

- Goal: Find an **unbiased estimator**, e.g. signal/noise ratio

Randomised experiments: Already in causal framework

- In a **randomised experiment**, p(x) is designed to be the same for both treatment groups (t=0 or t=1), typically uniform
- Paired '**clones**' in treatment and outcome groups
- Simply take the difference of the averages:

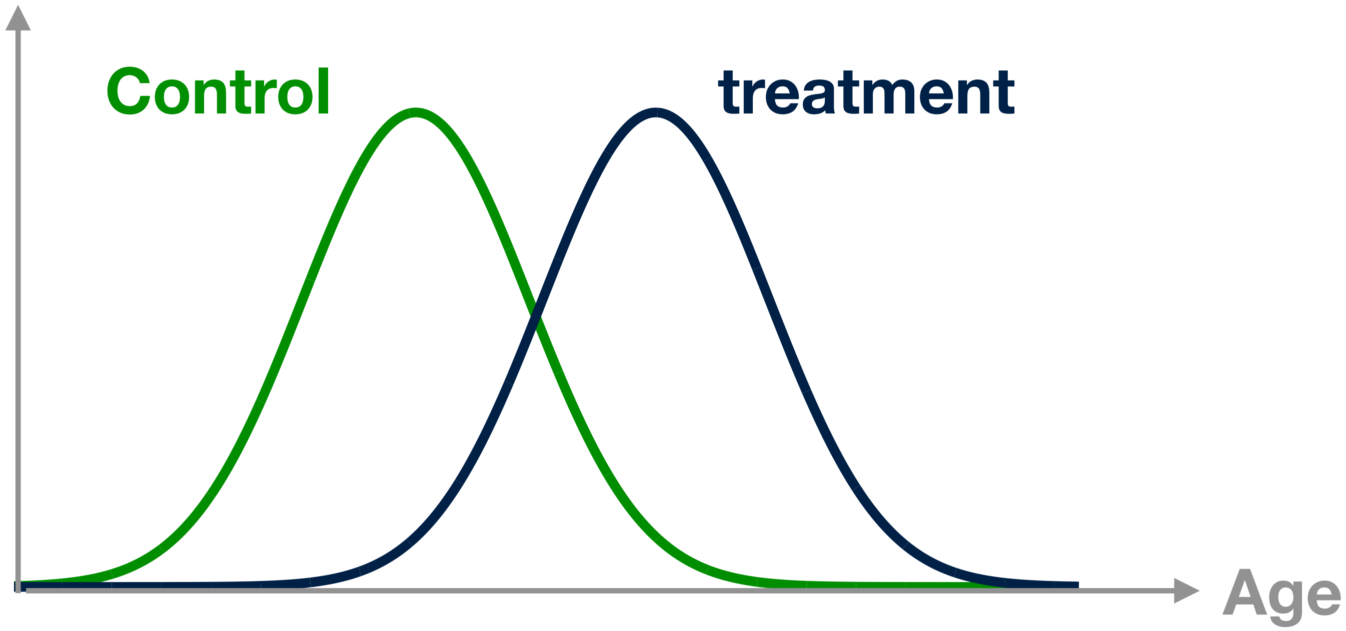
$$\Delta\hat{\mu} = \hat{\mathbb{E}}\left[y_{t=1}(x) - y_{t=0}(x)\right] = \frac{1}{N}\sum_{i=1}^{N}\left(y_1^{(i)}(x) - y_0^{(i)}(x)\right)$$

- Statistical test: e.g. T-test and p-values …

$$\frac{\Delta\hat{\mu}}{\sqrt{\frac{(\hat{\sigma}_{\Delta\mu})^2}{N}}} > t^*$$

Observational data: What goes wrong?

$$p(x|t=1) \neq p(x|t=0)$$



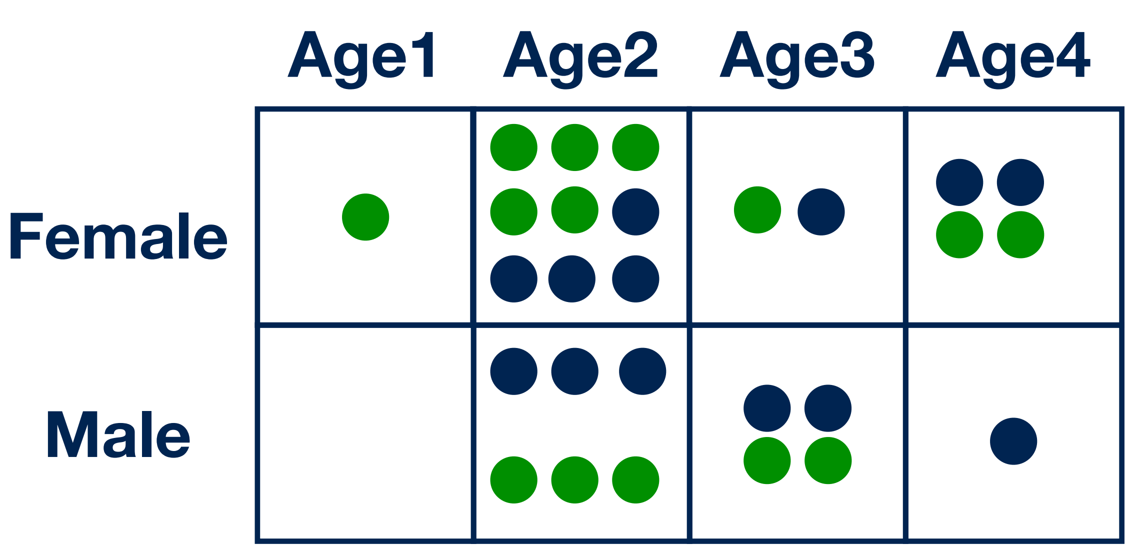
$$\left(\int y_1(x)p(x|t=1)dx \; - \int y_0(x)p(x|t=0)dx\right) \neq \int \big(y_1(x) - y_0(x)\big)p(x)dx$$

Observational data: Stratification

- Measure outcome (success/failure), **within** each of the young/old groups **separately**
- Take weighted average by the probability of being young/old

$$\mathbb{E}(\text{Healed}|t=1) = \mathbb{E}(\text{Healed}|t=1,\text{young})p(\text{young}) + \mathbb{E}(\text{Healed}|t=1,\text{old})p(\text{old})$$

- Disadvantages:
  - All possible confounders need to be observed
  - Assumes overlap between the two distributions (if there is no overlap, sample is not representative, e.g. performing the experiment only for old people)
  - Bad estimates as confounder dimensionality increases

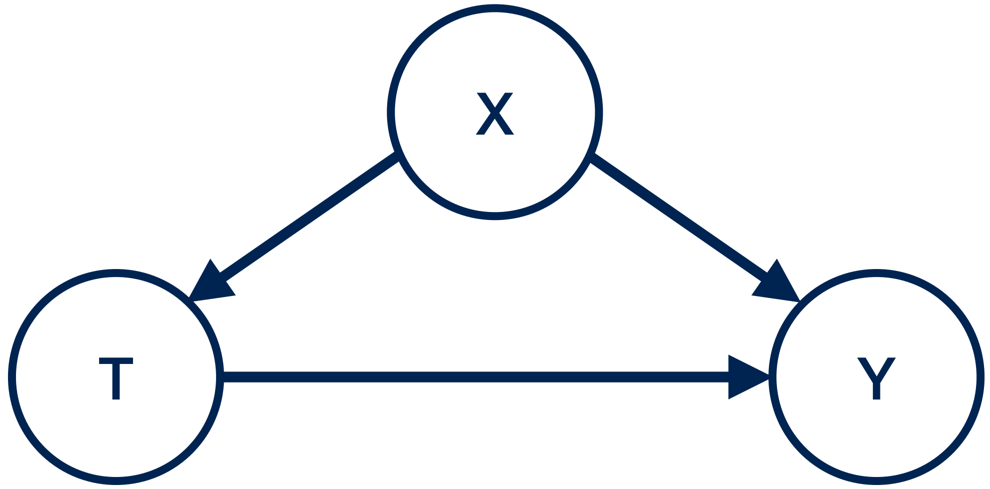


$\Rightarrow$

Need specific causal effect estimation techniques

Two main Frameworks for causal estimation/discovery

- **Potential outcomes (Rubin):**
    - Requires a given treatment-outcome pair (known directionality)
    - Mainly applies to causal estimation (learning effects)
    - More familiar to biologists



- **Structural causal models (Pearl):**
    - Causal graph
    - Structural equations
    - Algorithmic: Causal Discovery

$$x = f_x(\epsilon_x),\ t = f_t(x, \epsilon_t),\ y = f_y(x, t, \epsilon_y)$$

Assumption: Independent noise terms: $\epsilon_x \perp\!\!\!\perp \epsilon_t \perp\!\!\!\perp \epsilon_y$

Extend the language of probability theory:
do-calculus

Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.

Overview of the course

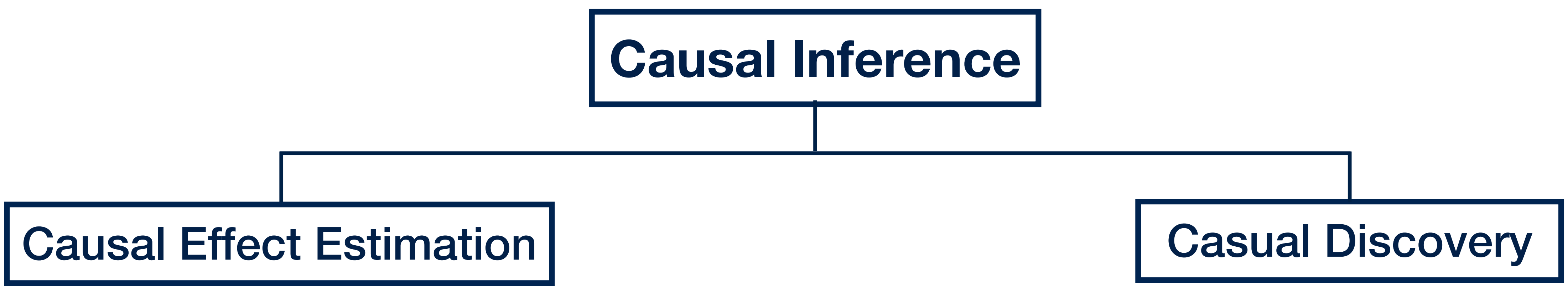
- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule

Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM

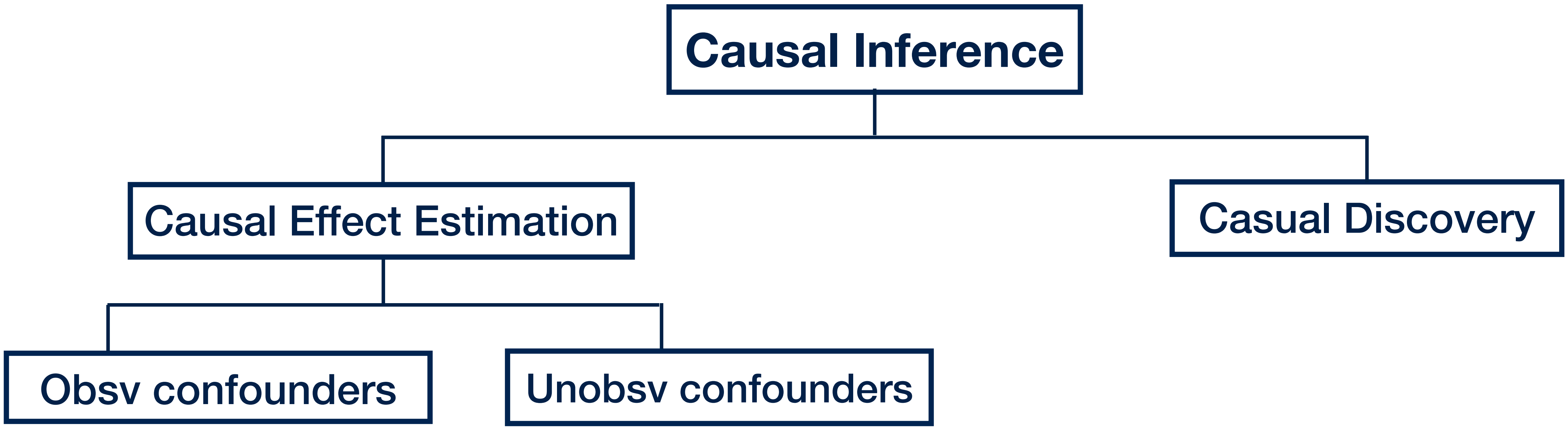
Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes’ rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



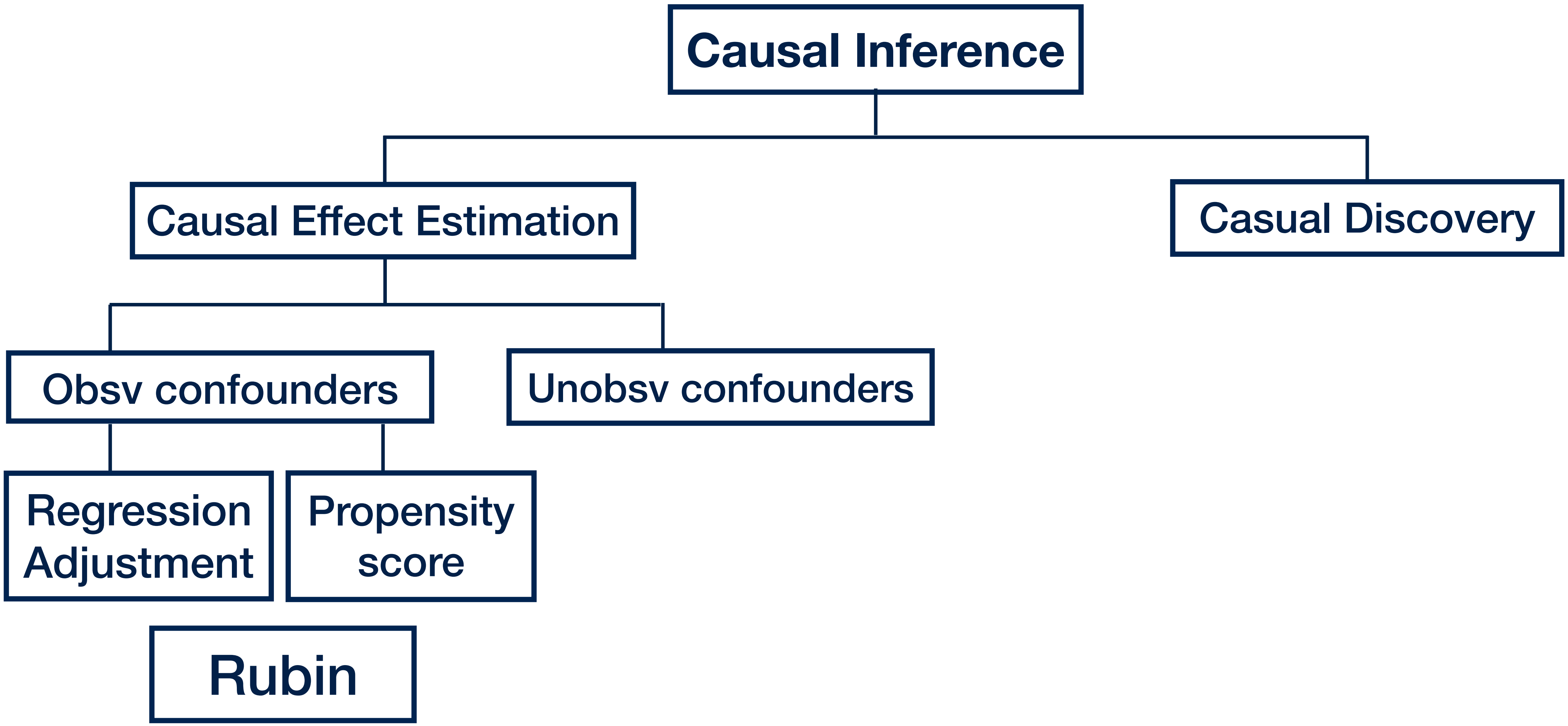
Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



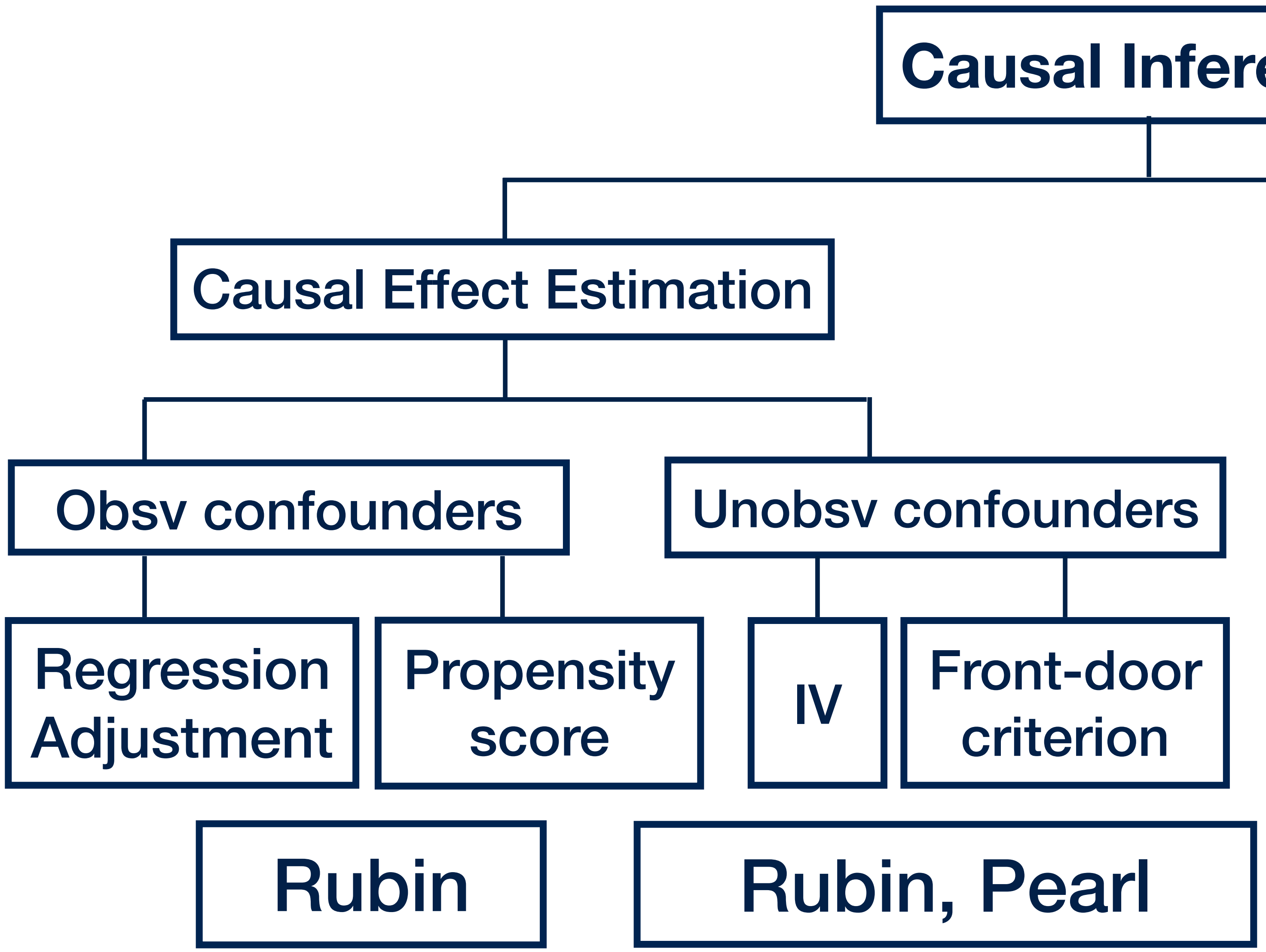
Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



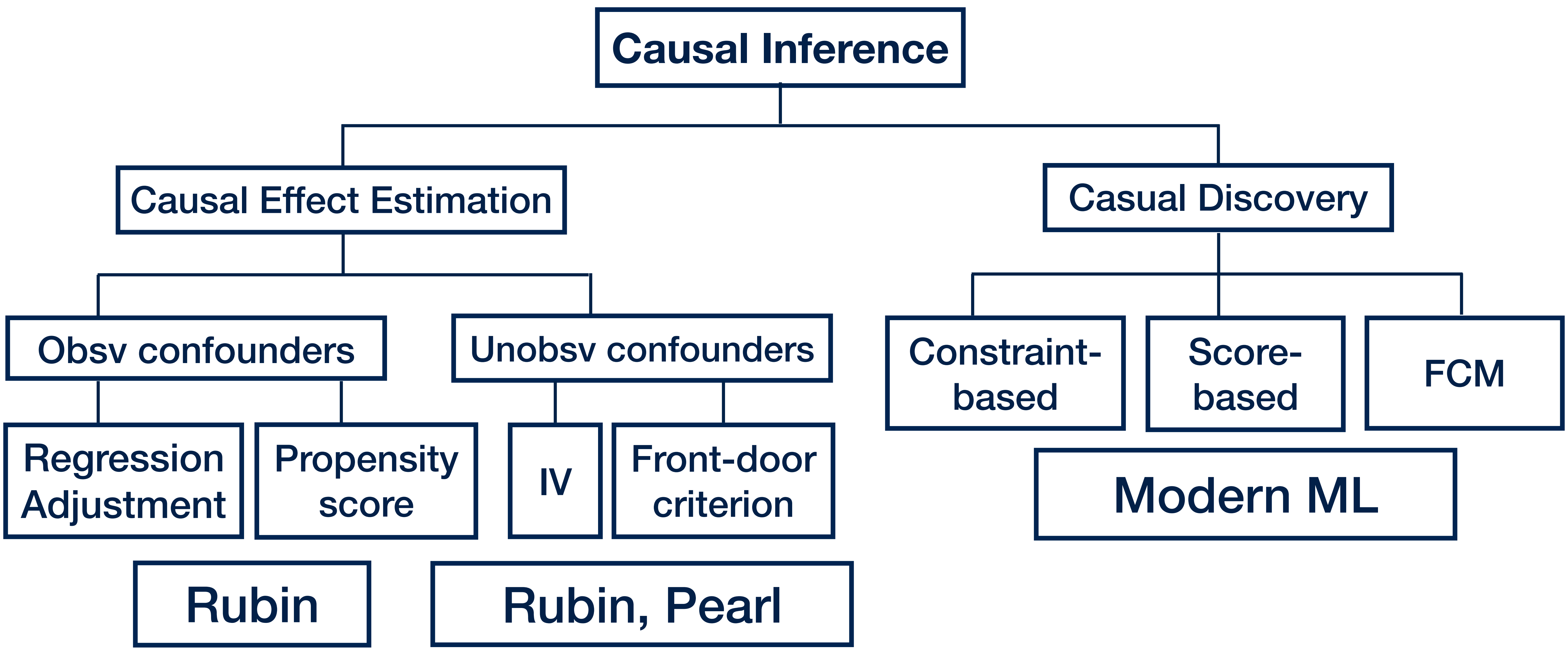
Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



Overview of the course

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



Causal Estimation of Effects vs Causal Discovery

- **How much would some variables (features or labels) change if we manipulate the value of another variable?**
  - Have a prior causal knowledge (may be incomplete)
  - Wish to estimate degrees of causal dependencies

- **By modifying the value of which variables could we change the value of another variable?**
  - Wish to discover the causal graph
  - Apply causal inference