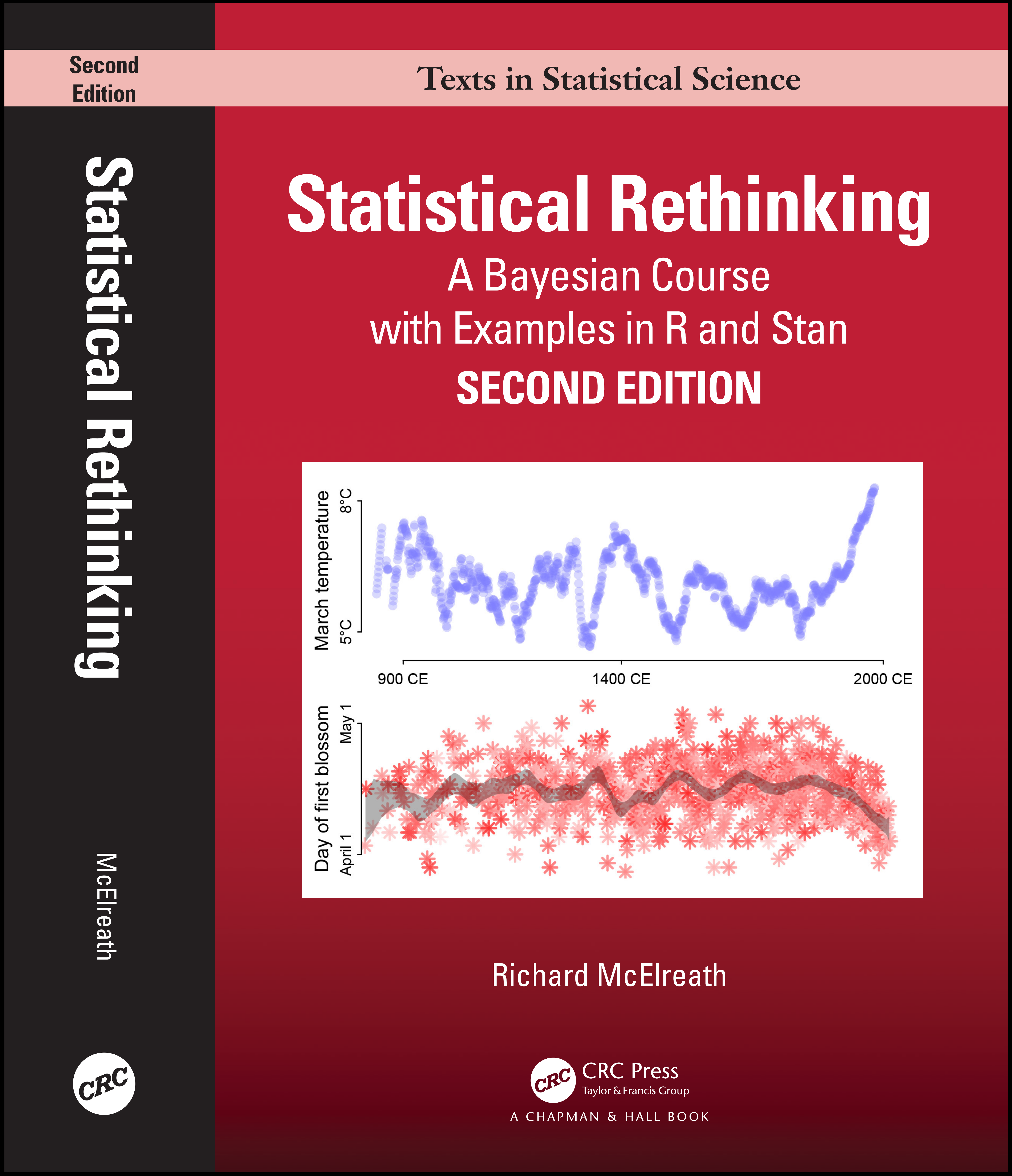
Statistical Rethinking

2022

(1) The Golem of Prague


Parametric assumptions:
(1) Independent samples
(2) Data normally distributed
(3) Equal variances
Type of data?
Discrete, categorical
Any counts < 5?
No
Yes
Chi-square tests, one and two sample
Fisher's exact test
Continuous
Type of question?
Relationships
Differences
Do you have dependent & independent variables?
Yes
No
Regression analysis
Correlation analysis
Parametric
Nonparametric
Pearson's r
Spearman's rank correlation
Differences between what?
Means
One-sample t-test
Variances
Fmax test or Bartlett's test
Multiple means Single variable
How many groups?
More than two
Parametric assumptions satisfied?
Two
Yes
No
No
Transform



Rethinking the role of statistical analysis in research

20 lectures

GOLEMS

OWLS

DAGS

Prague 16th century

Art from: “Breath of Bones: A Tale of the Golem” (2014)

Golems

Clay robots

Powerful

No wisdom or foresight

Dangerous

“Breath of Bones: A Tale of the Golem” (2014)

Statistical Models

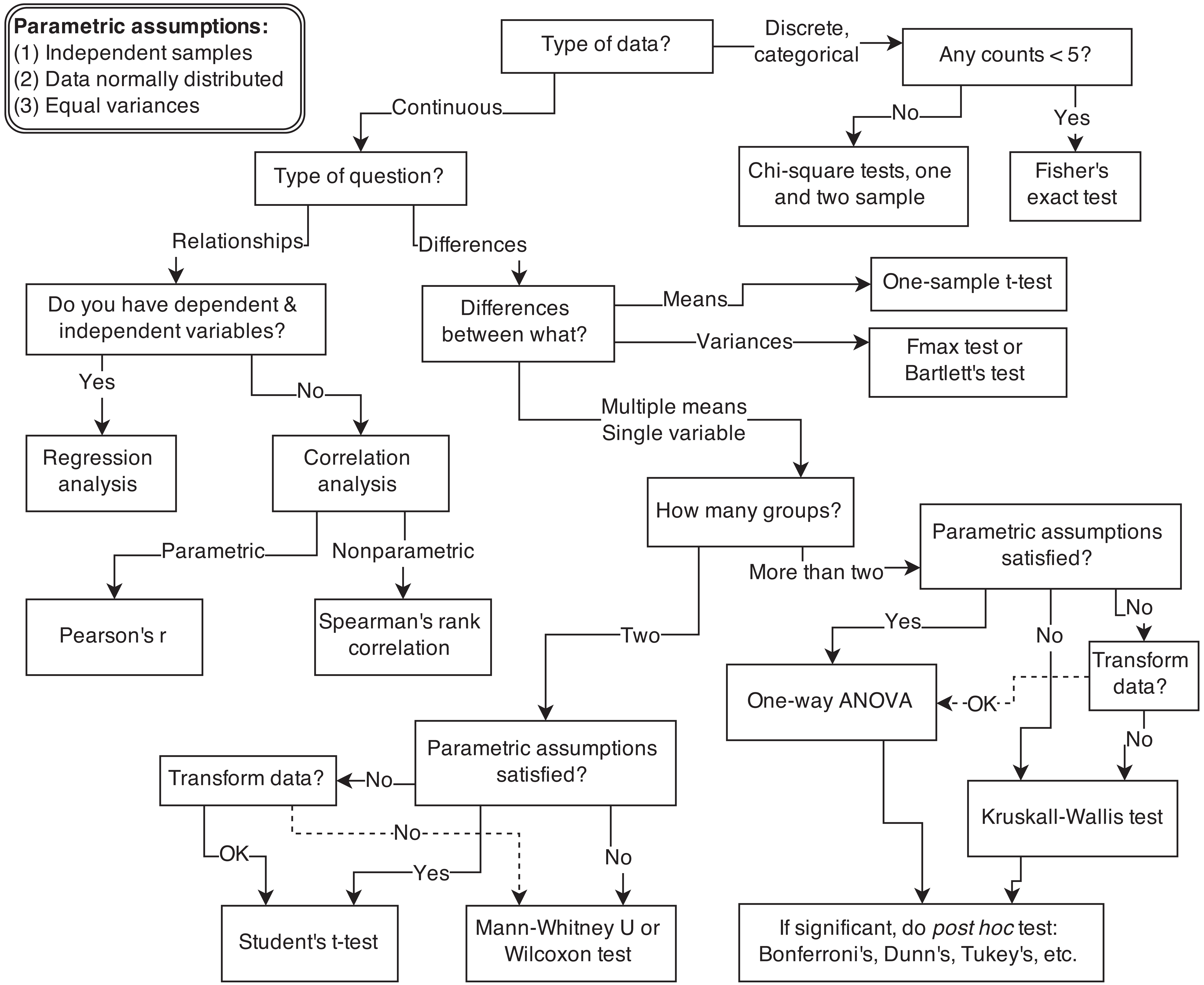
Like Golems

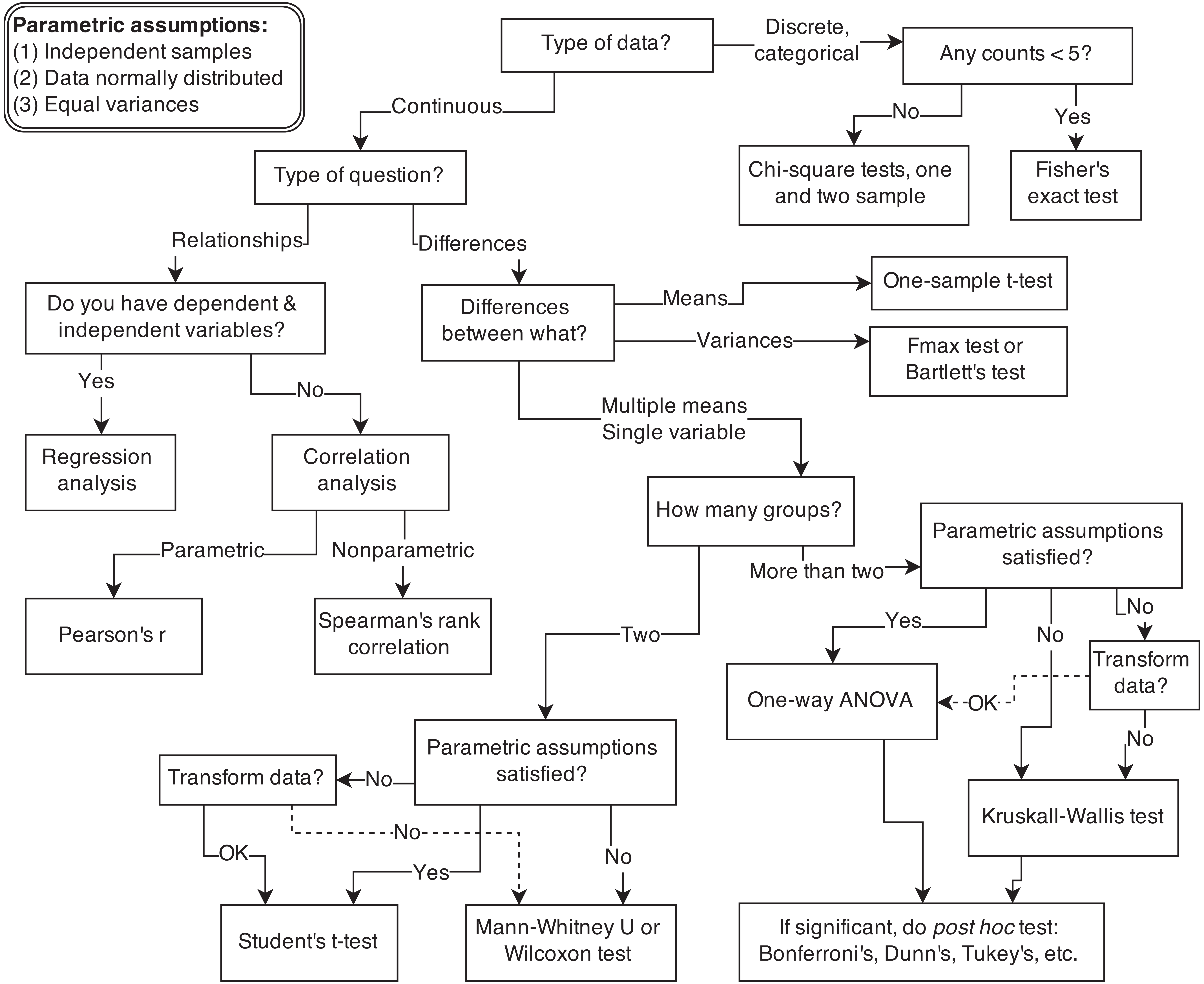
Clay robots

Powerful

No wisdom or foresight

Dangerous





Parametric assumptions:
(1) Independent samples
(2) Data normally distributed
(3) Equal variances
Type of data?
Discrete, categorical
Any counts < 5?
No
Yes
Chi-square tests, one and two sample
Fisher's exact test
Continuous
Type of question?
Relationships
Differences
Do you have dependent & independent variables?
Yes
No
Regression analysis
Correlation analysis
Parametric
Nonparametric
Pearson's r
Spearman's rank correlation
Differences between what?
Means
One-sample t-test
Variances
Fmax test or Bartlett's test
Multiple means Single variable
How many groups?
More than two
Parametric assumptions satisfied?
Yes
No
No
Transform data?
OK
One-way ANOVA
No
Kruskall-Wallis test
Two
Parametric assumptions satisfied?
No
Transform data?
OK
No
Yes
No
Student's t-test
Mann-Whitney U or Wilcoxon test
If significant, do post hoc test: Bonferroni's, Dunn's, Tukey's, etc.

Figure 1.1

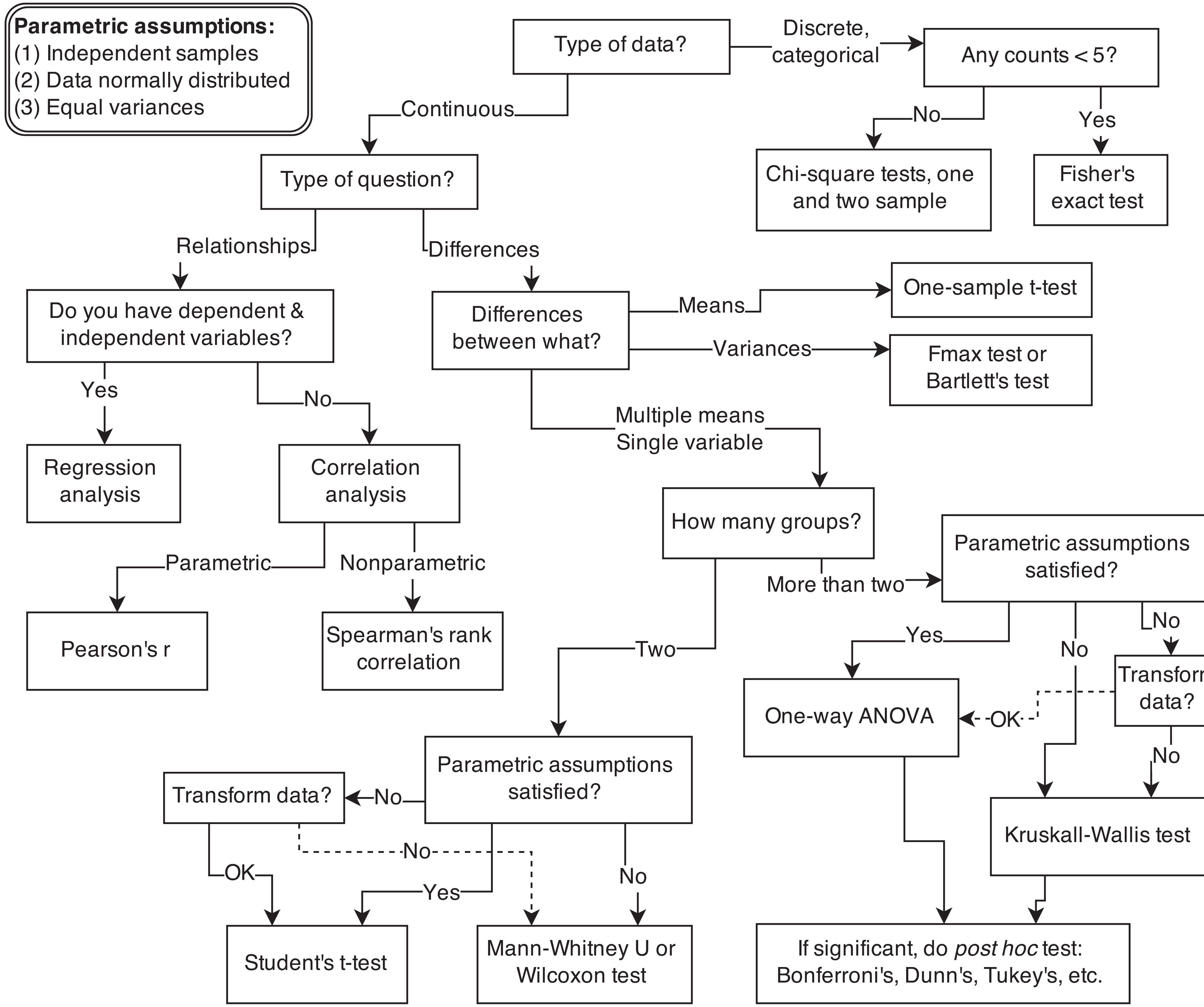
Statistical Models

Incredibly limiting

Focus on rejecting null hypotheses instead of research hypotheses

Relationship between hypothesis and test not clear

Industrial framework



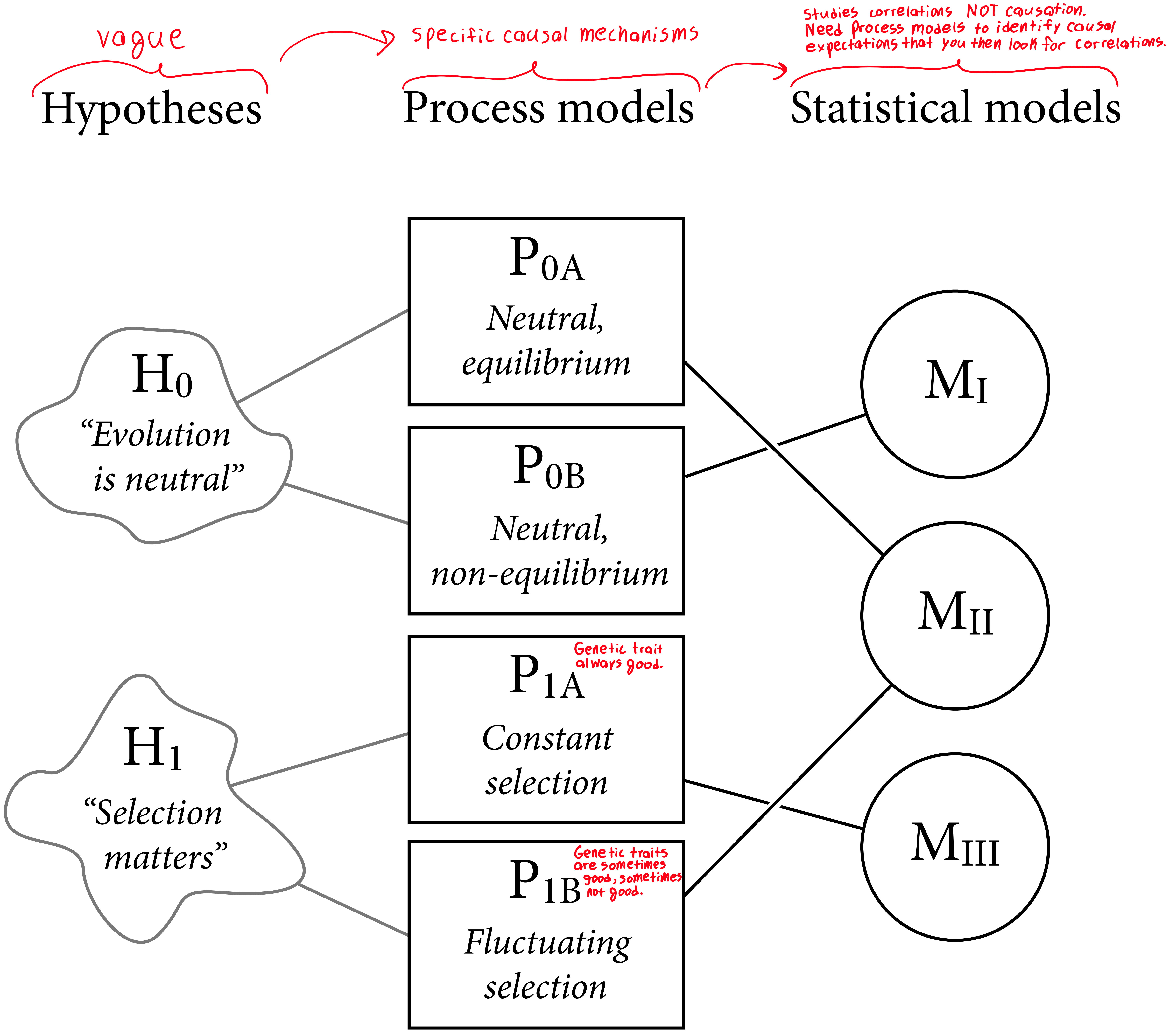

vague
specific causal mechanisms
Studies correlations NOT causation.
Need process models to identify causal
expectations that you then look for correlations.
Hypotheses
Process models
Statistical models
H_0
"Evolution is neutral"
H_1
"Selection matters"
P_{0A}
Neutral, equilibrium
P_{0B}
Neutral, non-equilibrium
P_{1A}
Genetic trait always good.
Constant selection
P_{1B}
Genetic traits are sometimes good, sometimes not good.
Fluctuating selection
M_I
M_{II}
M_{III}

Figure 1.2

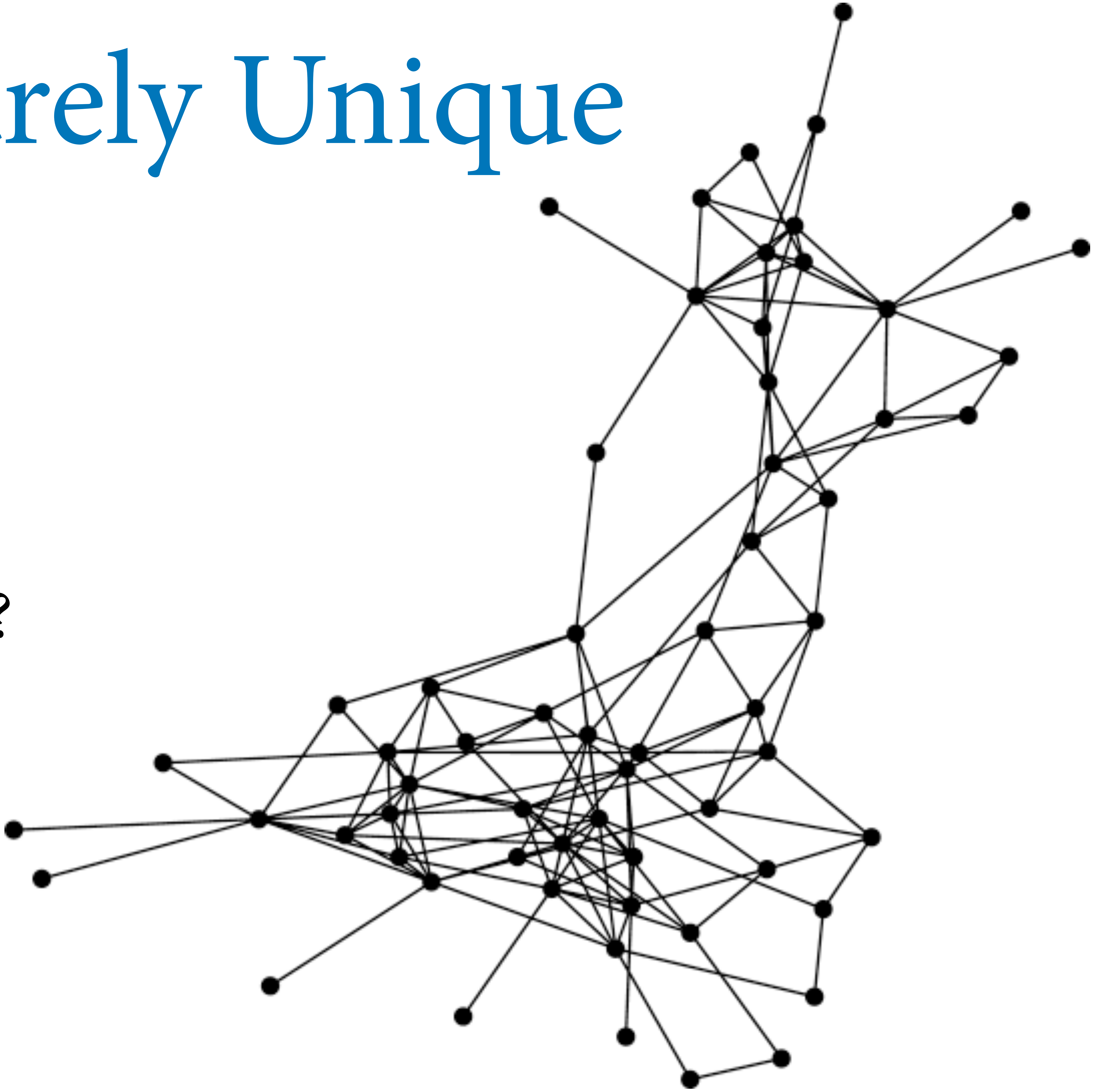
Null Models Rarely Unique

No such "null" hypothesis to randomize to

Null phylogeny?

Null ecological community?

Null social network?



Hypotheses and Models

Research requires more than than tiny null robots

Also requires:

Precise process model(s)

Statistical model (procedure, golem) justified by implications of process model(s) and question (estimand)

OWLS

HOW TO DRAW AN OWL (Joke)

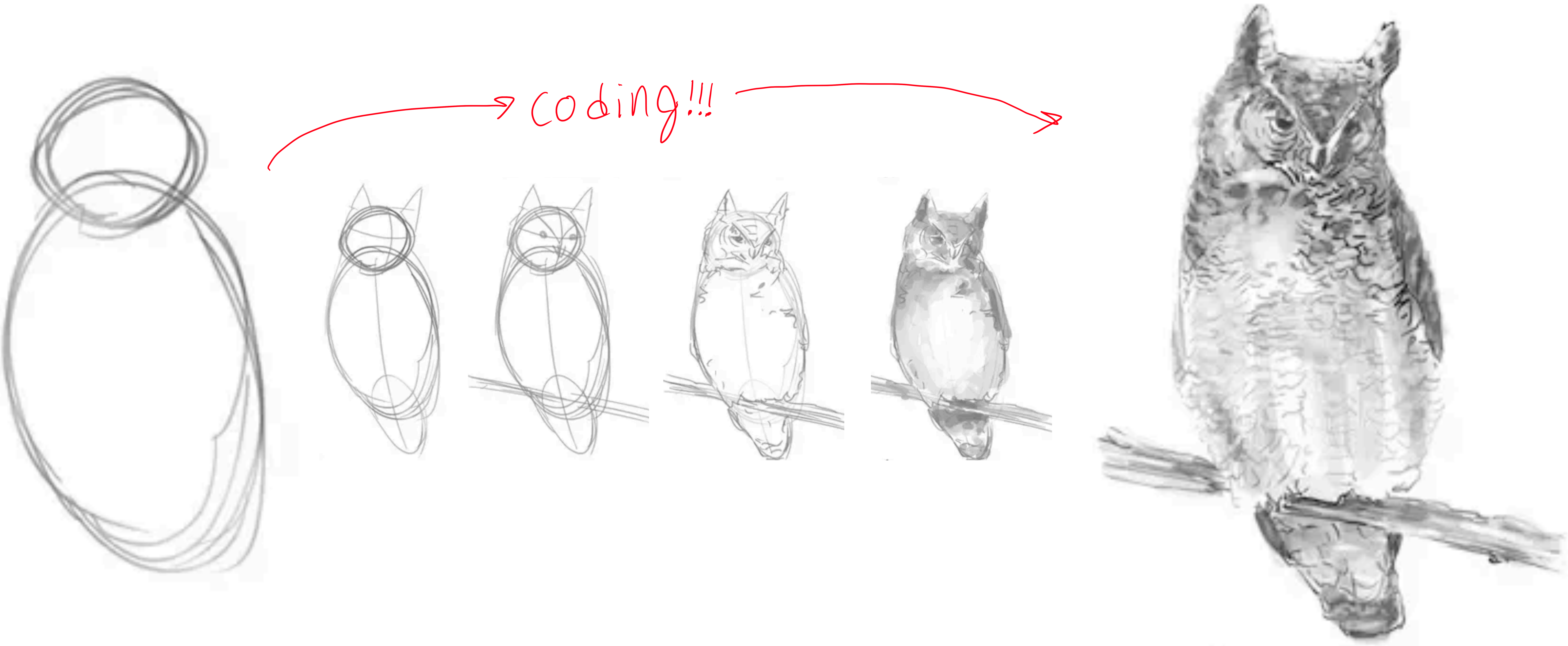
1. Draw some circles

HOW TO DRAW AN OWL (Joke)

1. Draw some circles

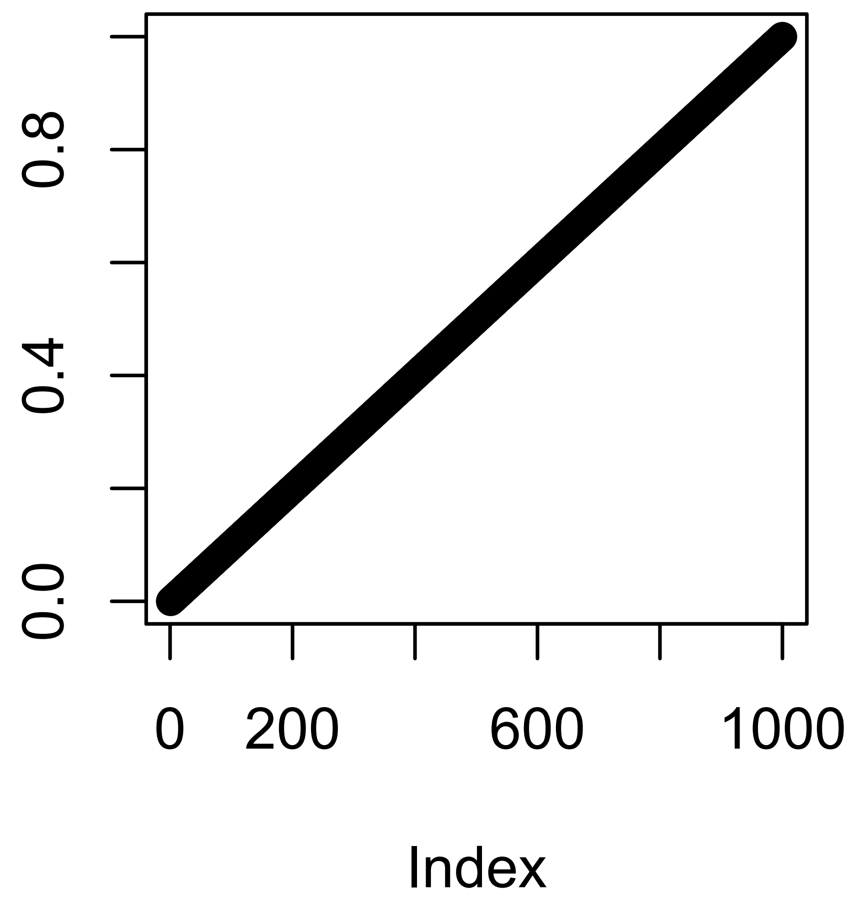
2. Draw the rest of the owl

HOW TO DRAW AN OWL

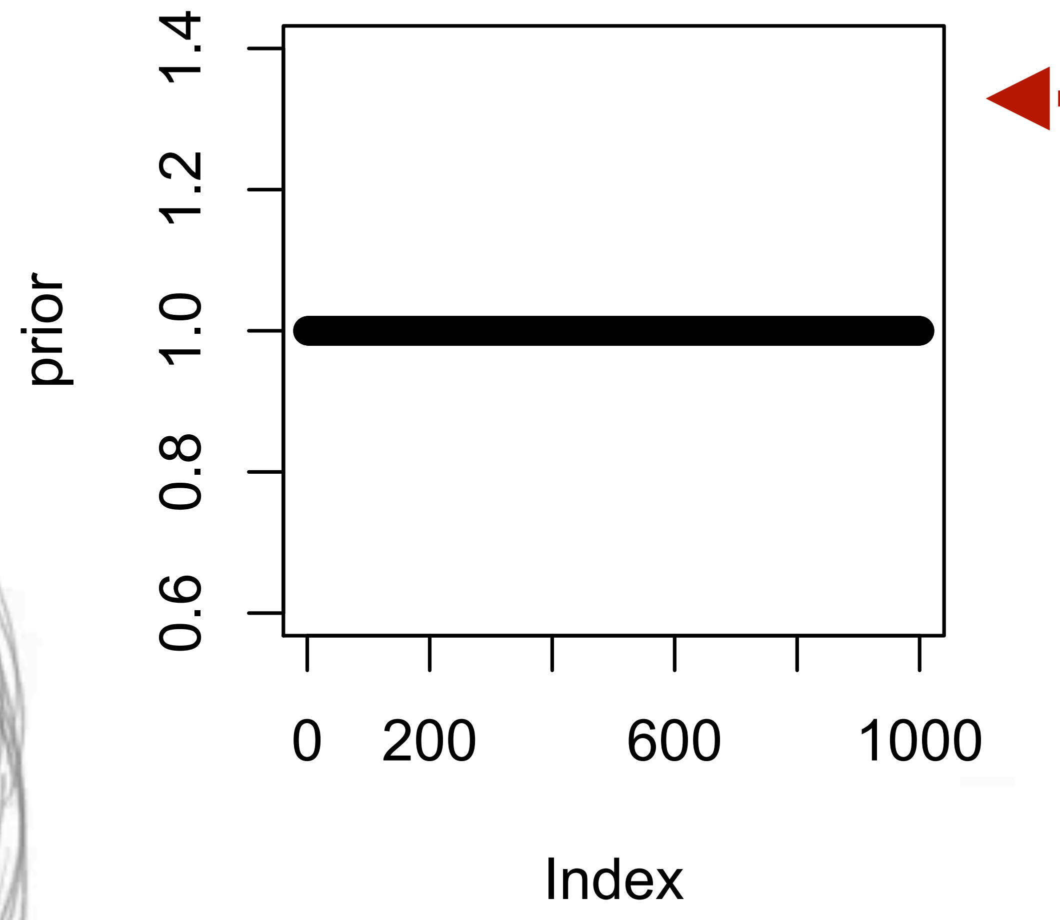


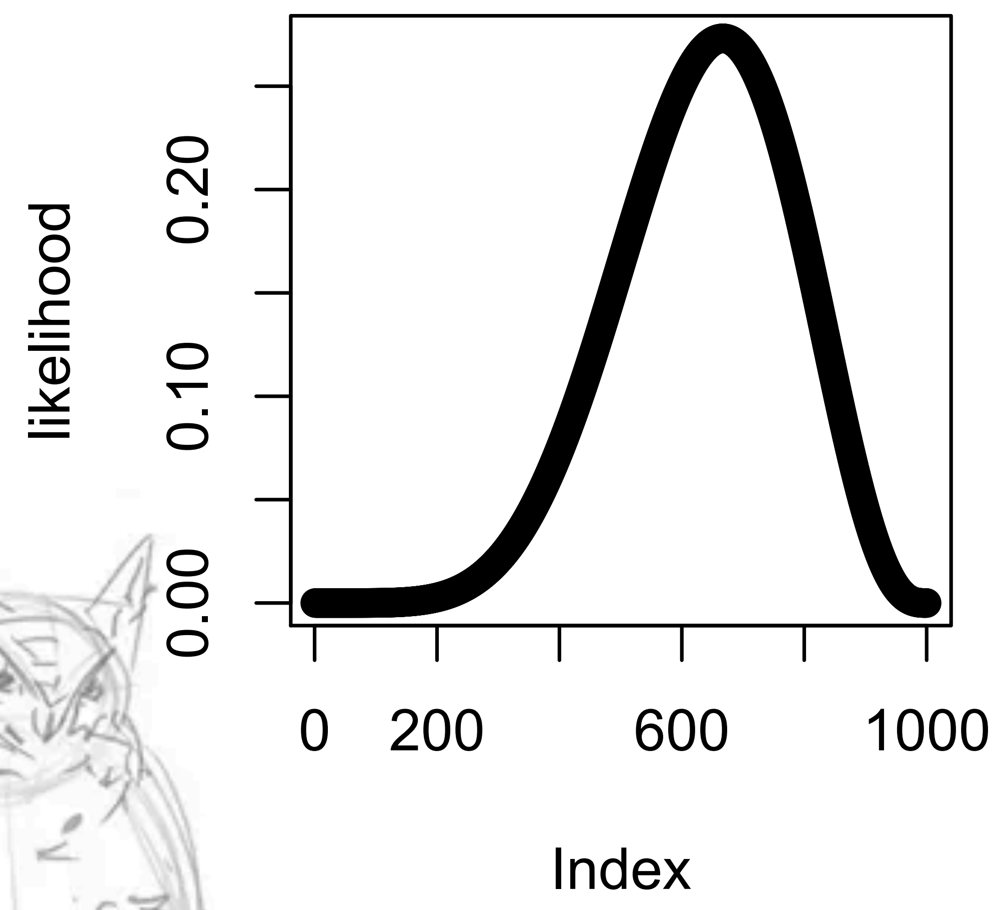
1. Draw some circles

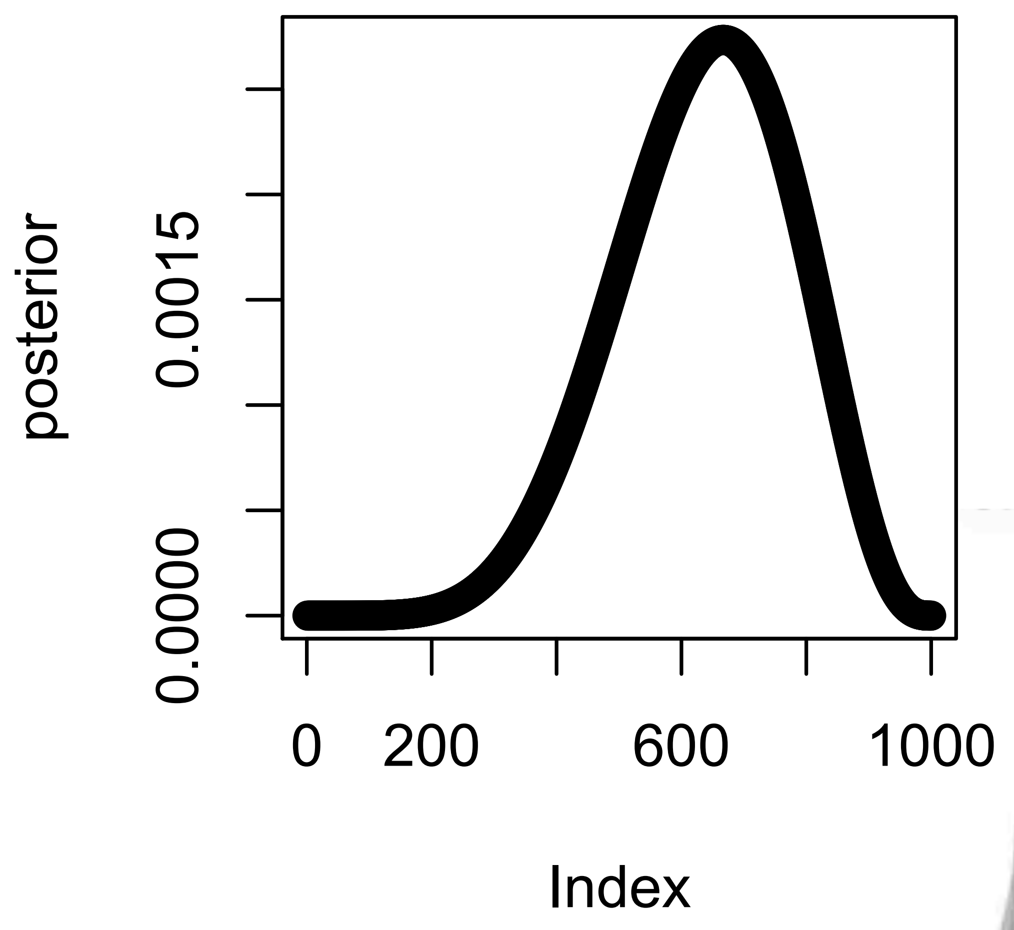
2. Draw the rest of the owl




```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```







Drawing the Bayesian Owl

Three modes:

Understand what you are doing

Document your work, reduce error

Respectable scientific workflow

Drawing the Bayesian Owl

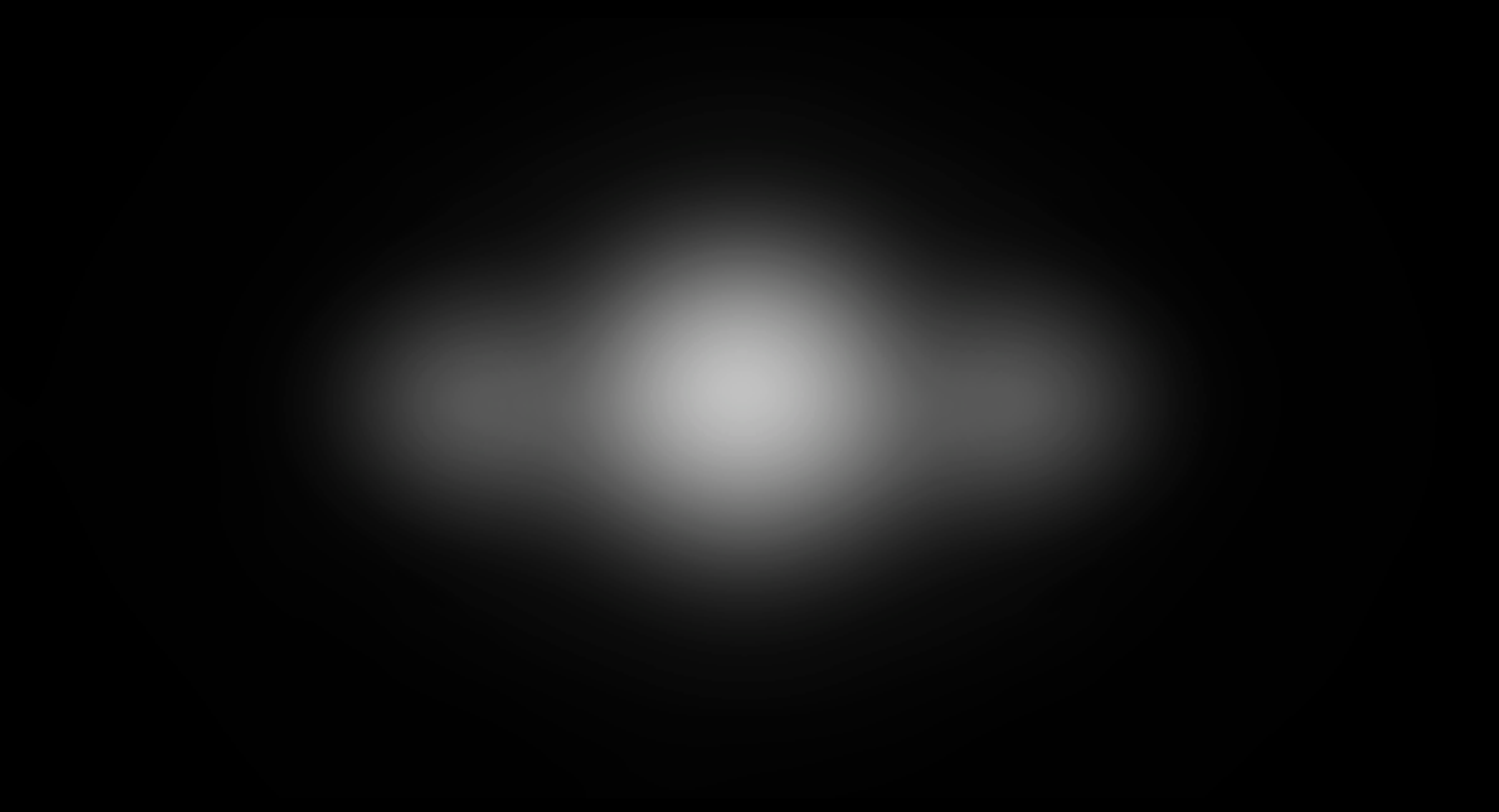
1. Theoretical estimand

2. Scientific (causal) model(s)

3. Use 1 & 2 to build statistical model(s)

4. Simulate from 2 to validate 3 yields 1

5. Analyze real data



Saturn, Galileo 1610

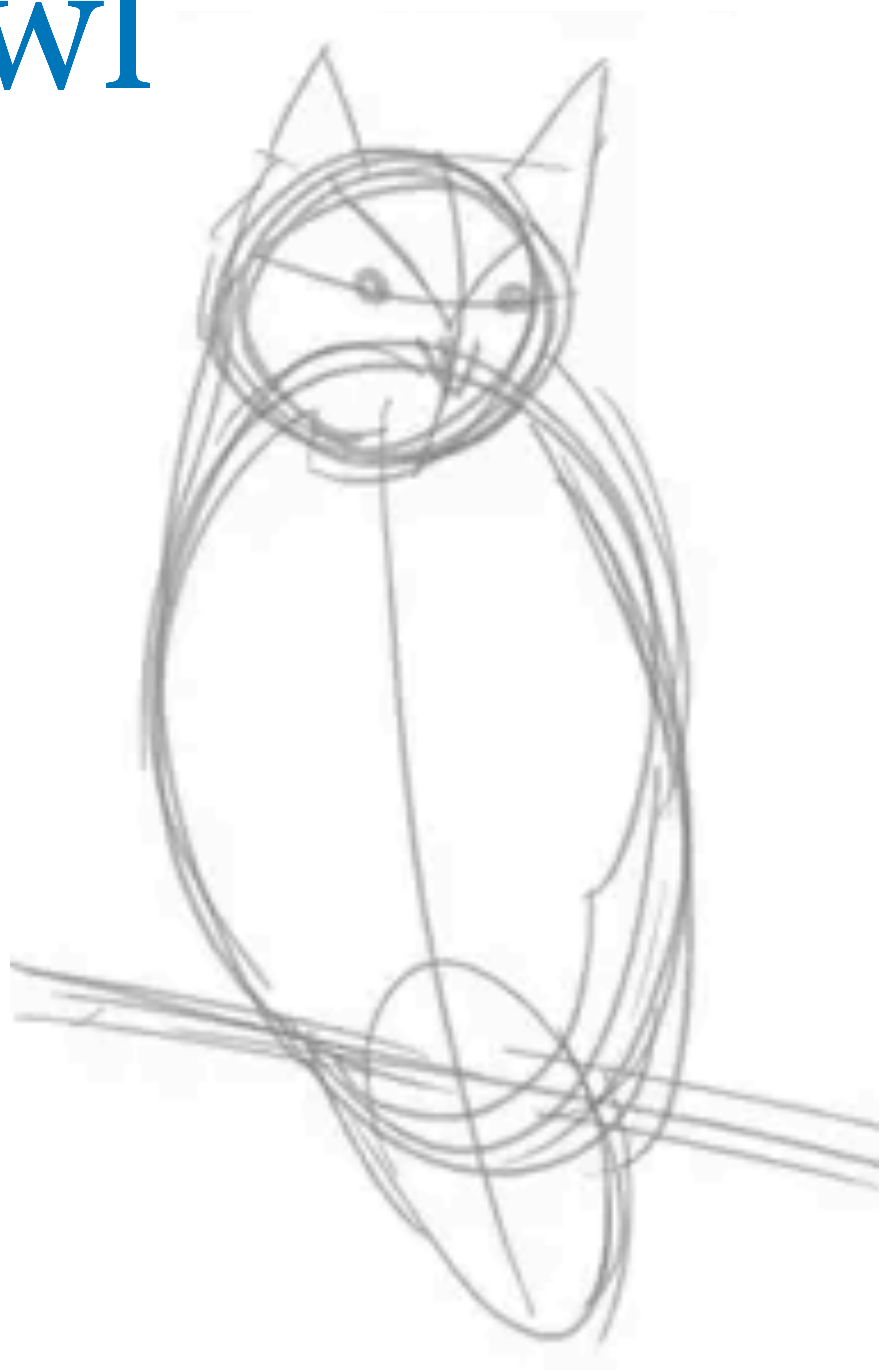
Drawing the Bayesian Owl

Bayesian approach is permissive, flexible

Express uncertainty at all levels

Direct solutions for measurement error, missing data

Focus on scientific modeling



DAGS


BAYES
FREQUENTISM

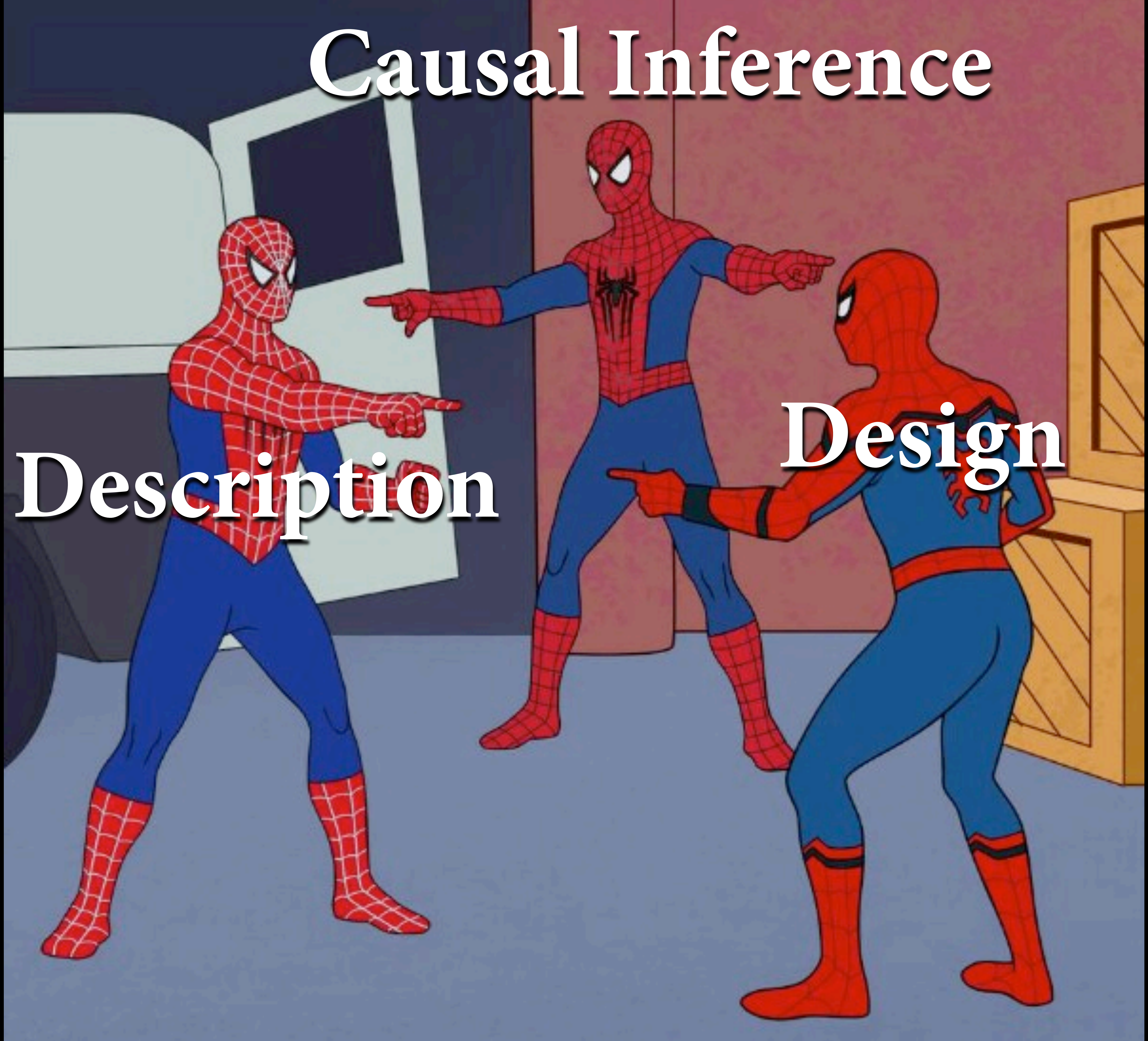

CAUSAL
INFERENCE
imgflip.com

Science Before Statistics

For **statistical models** to produce scientific insight, they require additional **scientific (causal) models**

The **reasons** for a statistical analysis are not found in the data themselves, but rather in the **causes** of the data

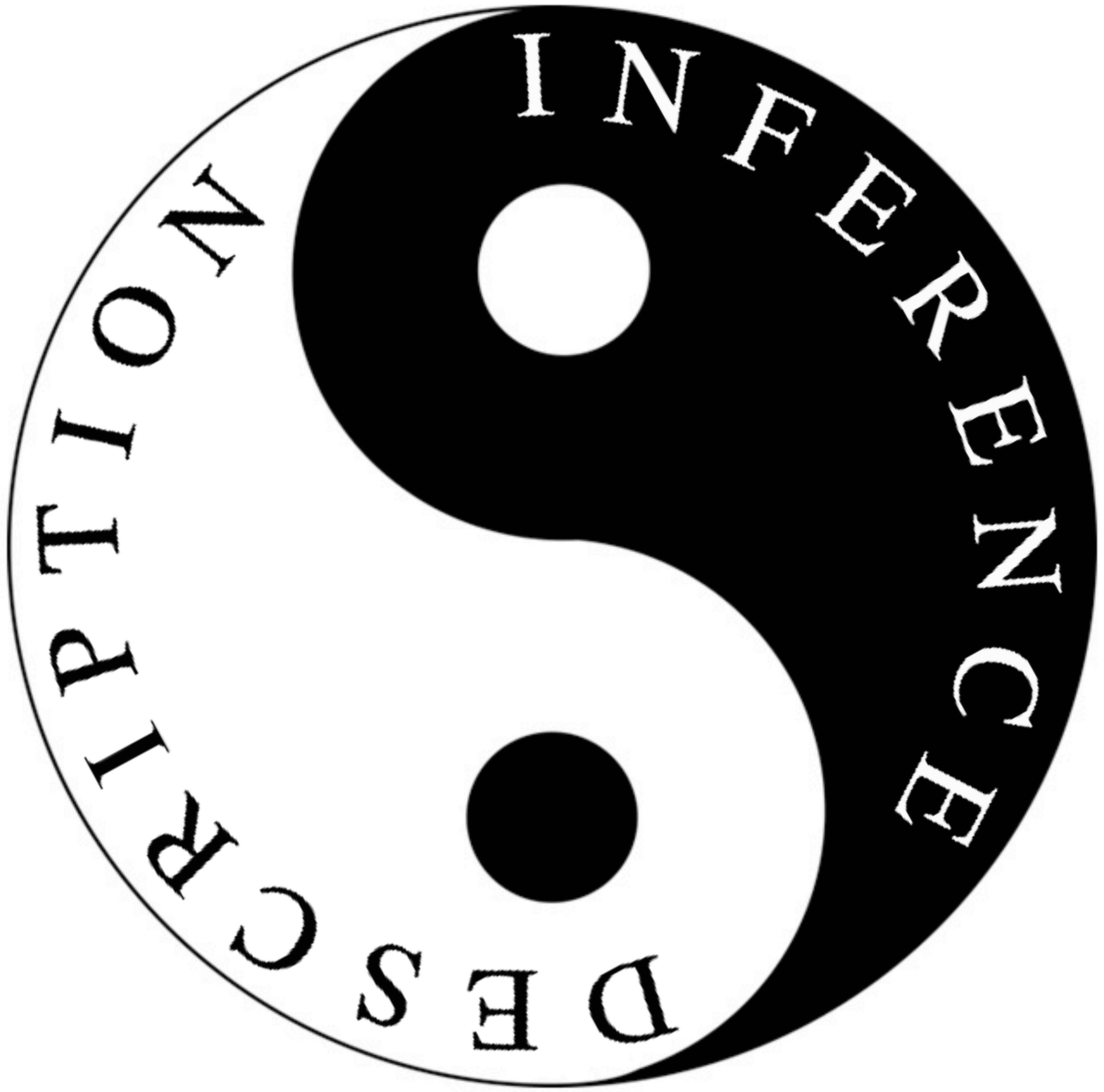
The **causes** of the data cannot be extracted from the data alone. No causes in; no causes out.


Causal Inference
Description
Design

Causes Are Not Optional

Even when goal is **descriptive**, need causal model

The **sample** differs from the **population**; describing the population requires causal thinking



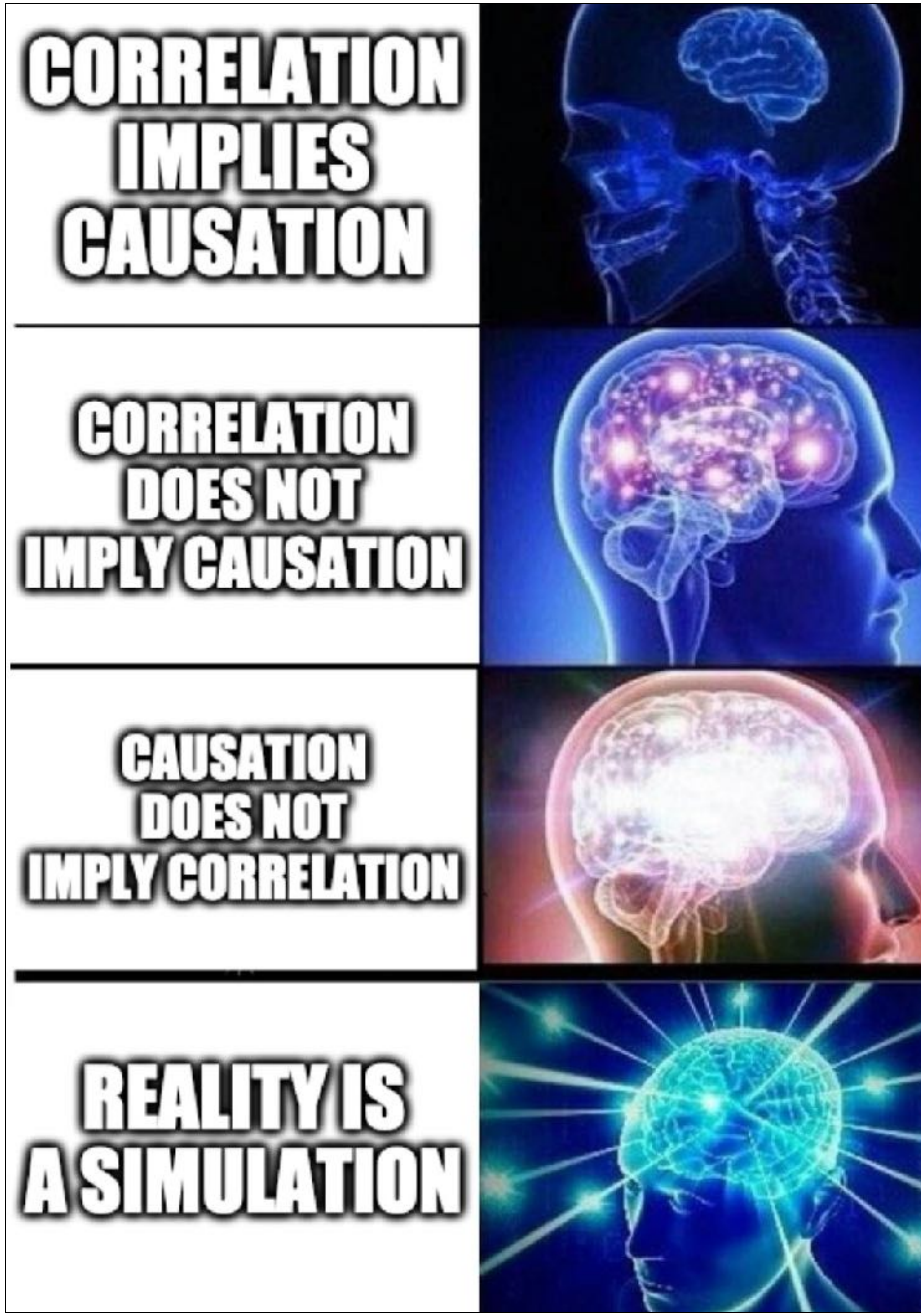
What is Causal Inference?

More than **association** between variables

Causal inference is special kind of **prediction** of intervention

counterfactual imagining of something that might have happened.

Causal inference is **imputation** of missing observations



Causal Prediction

Knowing a **cause** means being able to predict the **consequences** of an **intervention**.
What if I do this?

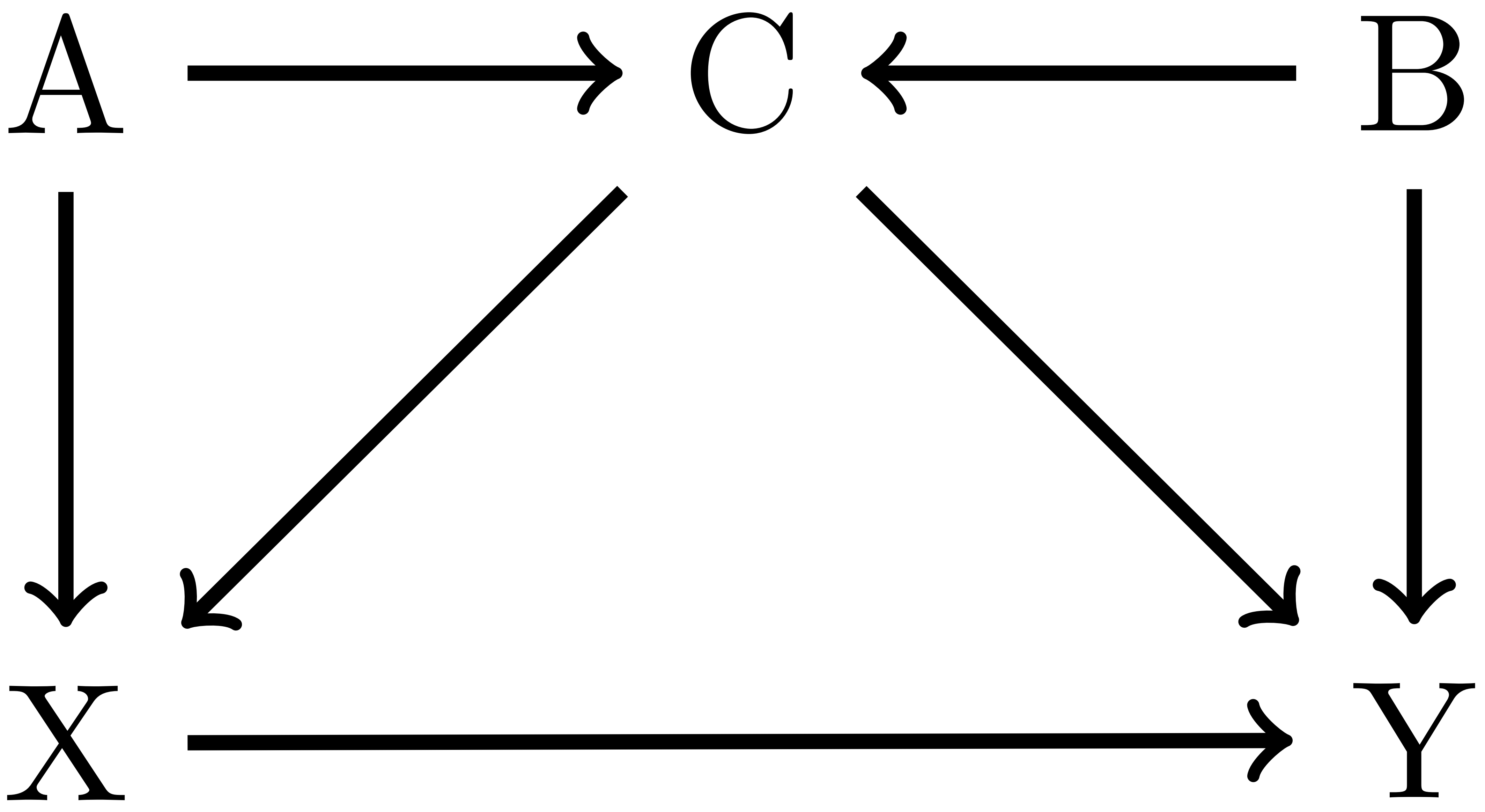
Causal Imputation

Knowing a **cause** means being able to construct unobserved **counterfactual outcomes.**

What if I had done something else?

↳ Then causal inference means that you know what would have happened

DAGs



Directed Acyclic Graphs (DAG)

Heuristic causal models

Clarify scientific thinking

Analyze to deduce appropriate statistical models

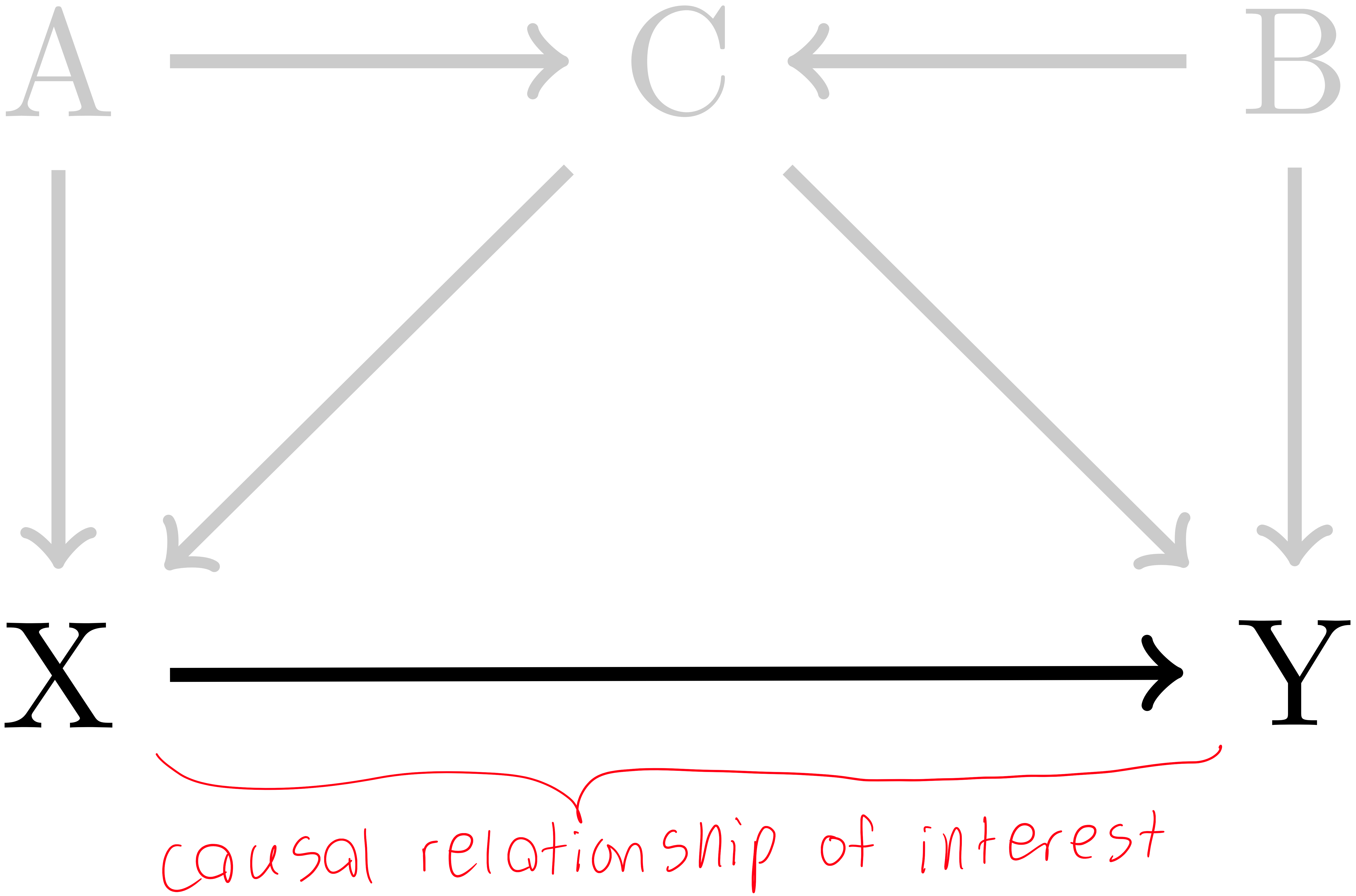
Much more as the course develops

DAGs

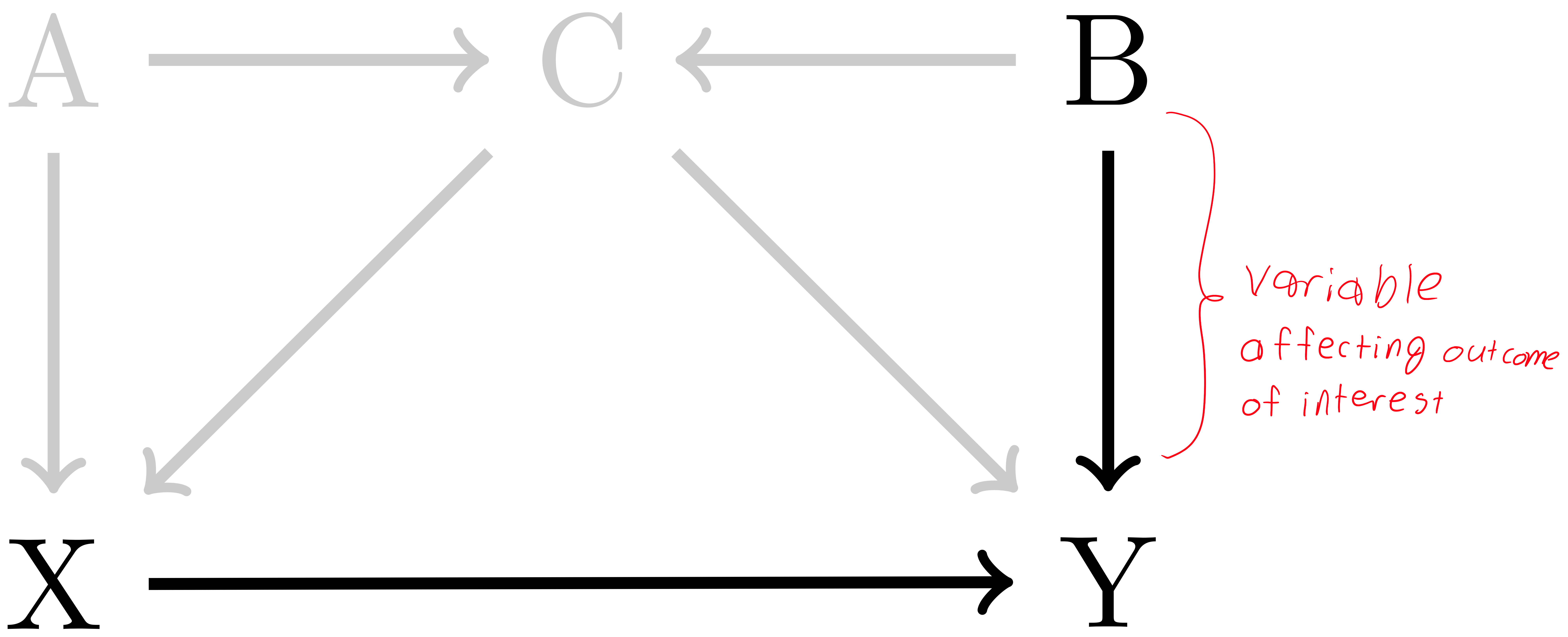
Each letter = variable
Each arrow = causal relationships

$$\begin{tikzcd}
A \arrow[r] \arrow[d] & C \arrow[dl] \arrow[dr] & B \arrow[l] \arrow[d] \\
X \arrow[rr] & & Y
\end{tikzcd}$$

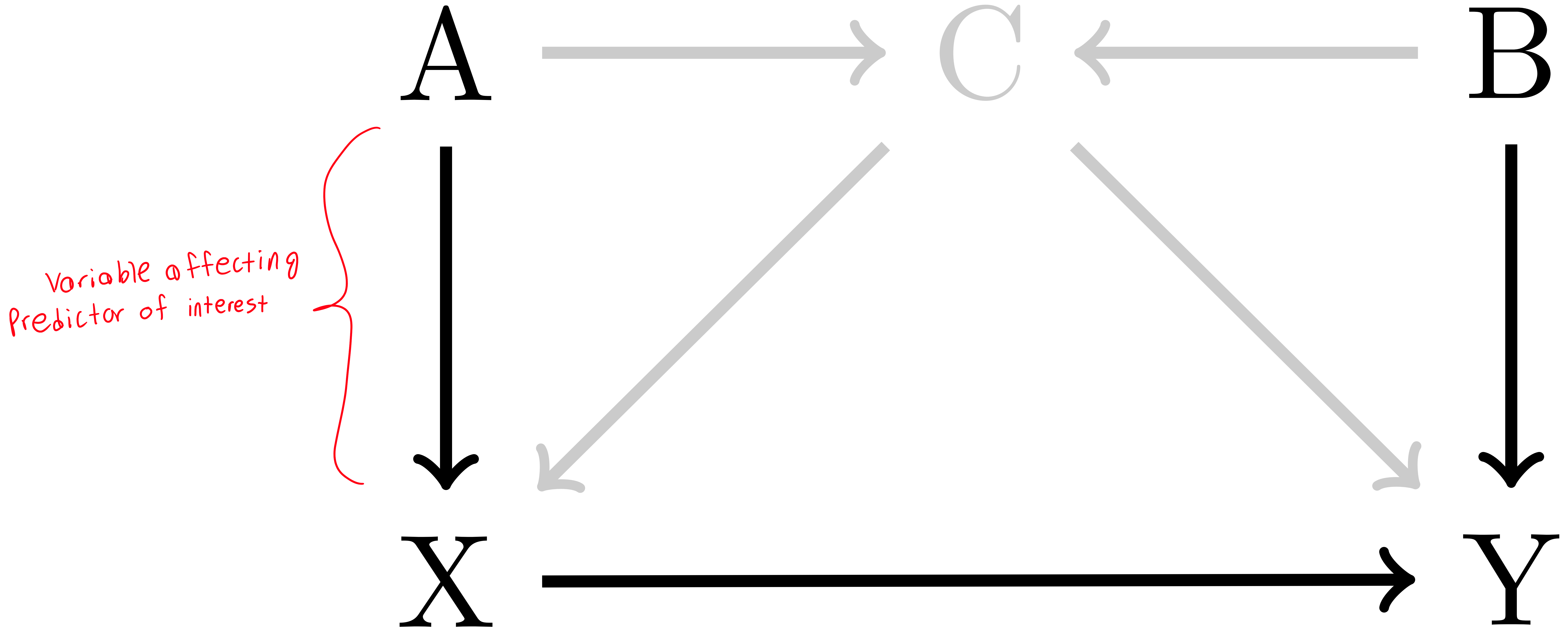
DAGs



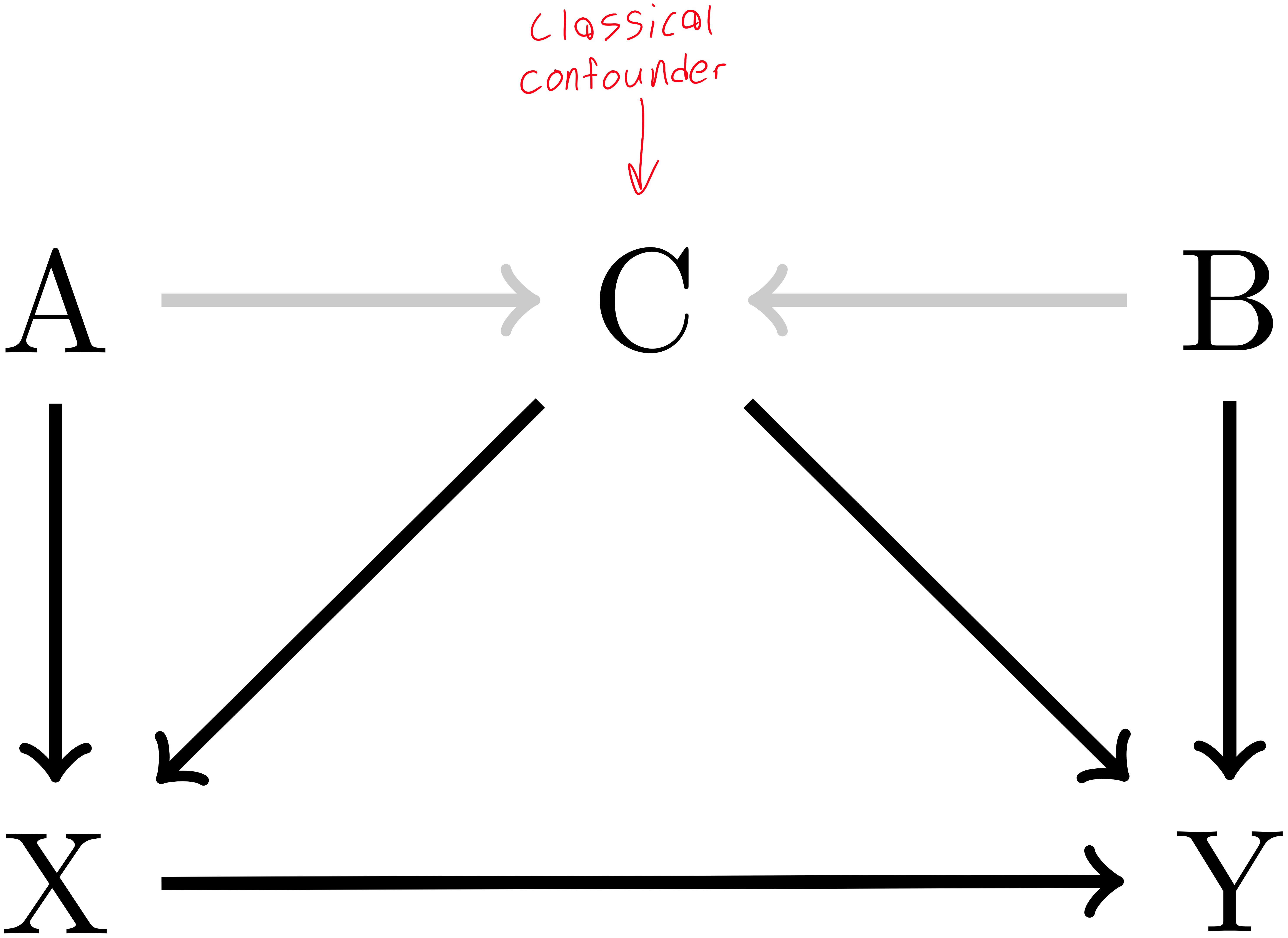
DAGs



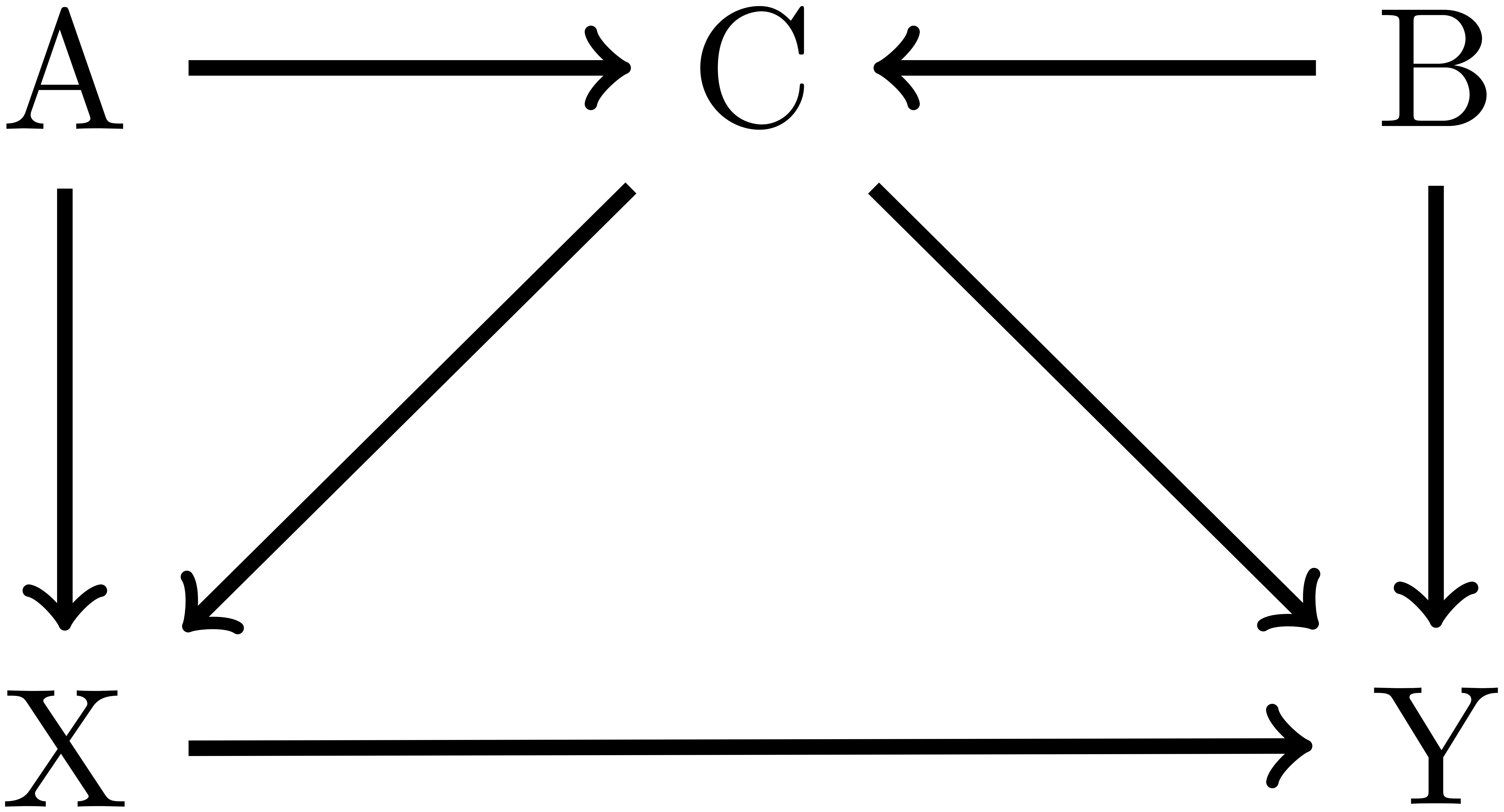
DAGs



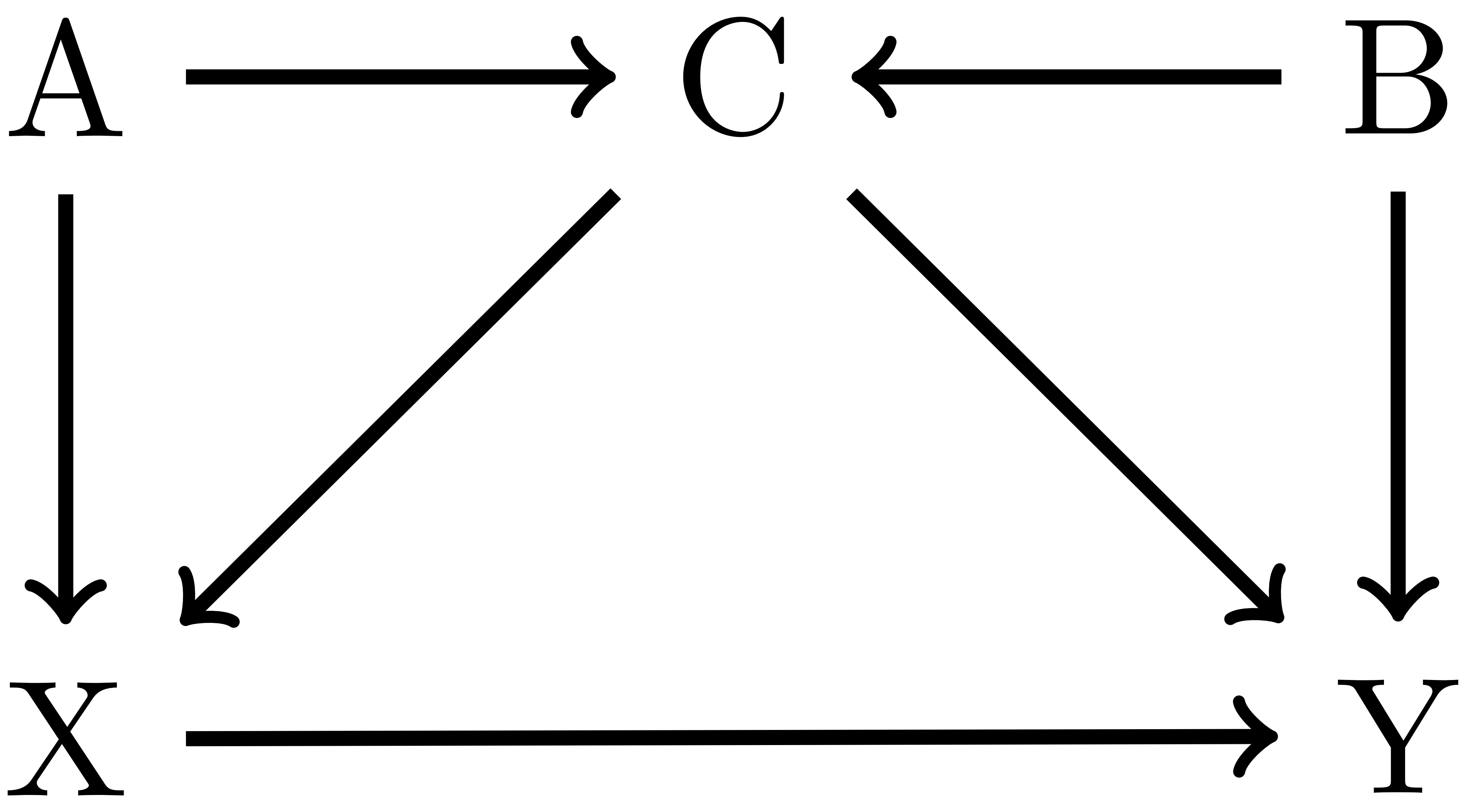
DAGs



DAGs



DAGs



Different queries, different models

Which control variables?

Absolute not safe to add everything — bad controls

How to test the causal model?

With more scientific knowledge, can do more

Golems, Owls, DAGs

Golems: Brainless, powerful statistical models

Owls: Documented, objective procedures

DAGs: Transparent scientific assumptions to
justify scientific effort
expose it to useful critique
connect theories to golems

Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / Interactions	Chapters 7 & 8
Week 5	MCMC & Generalized Linear Models	Chapters 9, 10, 11
Week 6	Integers & Other Monsters	Chapters 11 & 12
Week 7	Multilevel models I	Chapter 13
Week 8	Multilevel models II	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

`https://github.com/rmcelreath/statrethinking_2022`