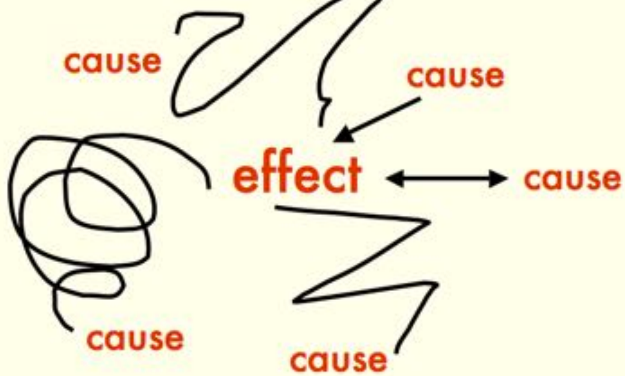
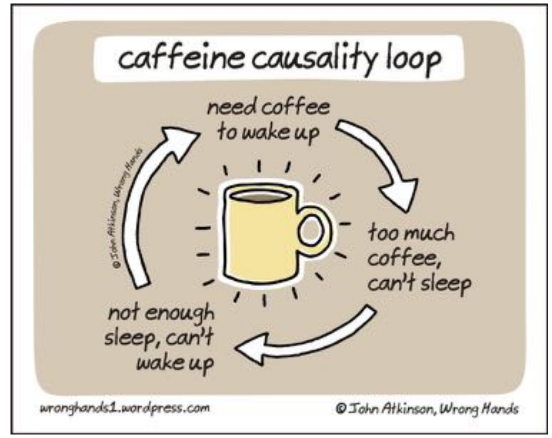
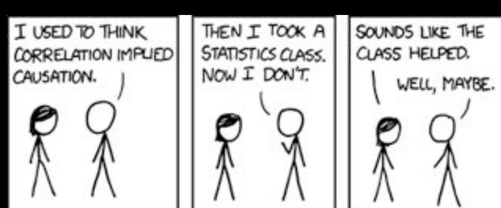
Building Causal Models





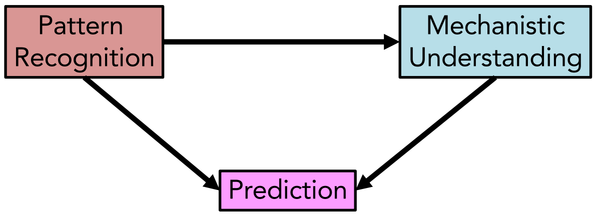


xkcd

Overview

1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

Goals of Science



All are valid and useful in particular contexts – What are ***YOU*** seeking to do?

Pearl's Ladder of Causality

3. Counterfactual – Can imagine what would happen under unobserved conditions

Prediction
- Requires model of a system
- Requires identification of causality

2. Intervention – Understand what happens you do something

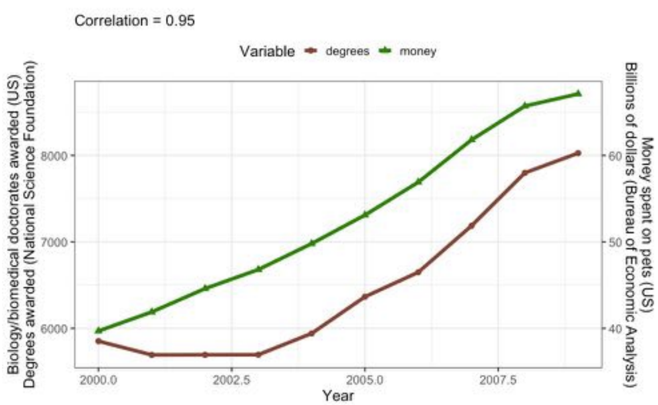
Mechanistic Understanding
Experiments
Provides evidence of causal link

1. Observation – Cause is associated with effect

Pattern Recognition
- Correlation
- Can only predict within the range of data

Pearl and Mackenzie 2018

Observation: Can We Learn Anything?

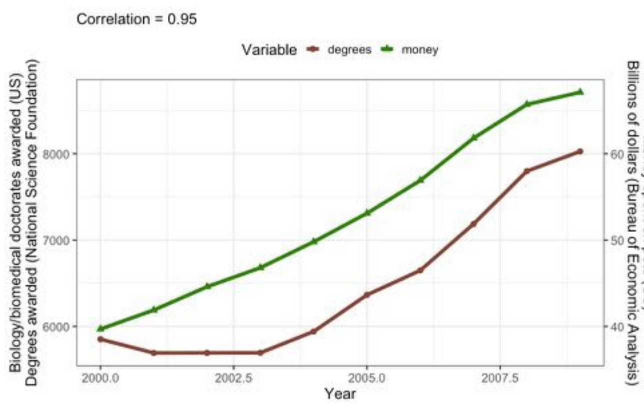


http://tylervigen.com/

Correlation does not equal causation… but where there's smoke, there's fire.

-Jim Grace

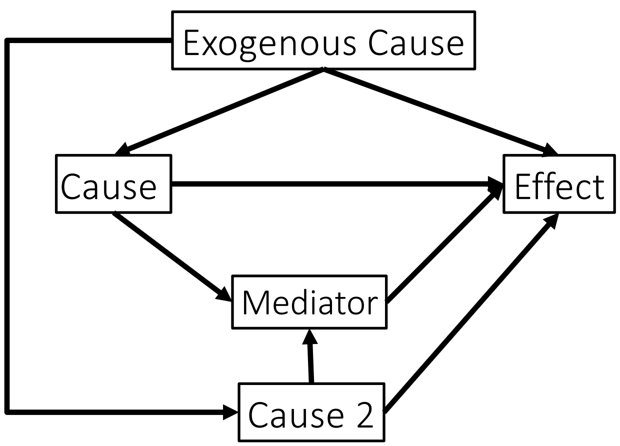
Where's the Fire?



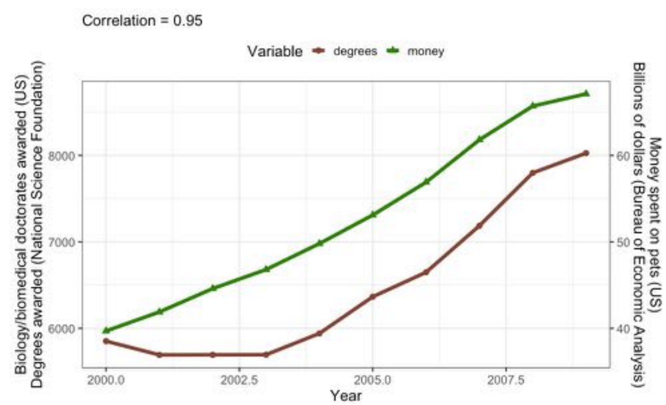
http://tylervigen.com/

What We Want to Evaluate

But This is the World



What Is The World Behind This Association?



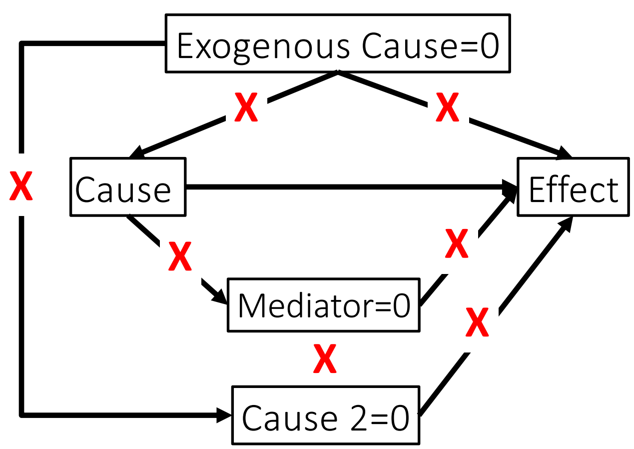
http://tylervigen.com/

Do You Need to be Doing Causal Inference?

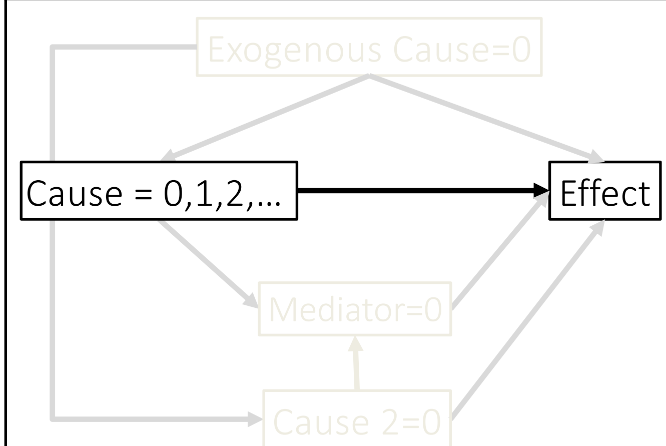
- No!
  - Not all studies will provide causal links between different variables of interest
  - If the study goal is predictive or descriptive rather than causal, this might not be needed
- But…
  - We cannot hope to understand the world without developing an understanding of causal associations
- Indeed
  - Understanding the clockwork machinery of the universe is an end goal of science – one which we can never achieve, but strive for!

WHAT IS YOUR QUESTION? IS IT FUNDAMENTALLY CAUSAL? OR NOT?

Intervention: Experiments! What can we learn?

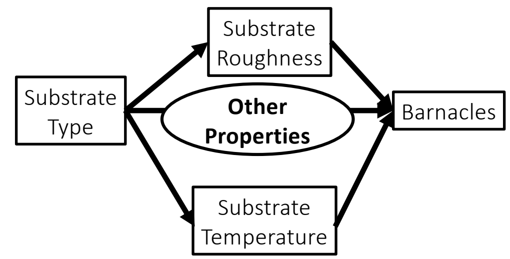


Experiments: Manipulate Cause of Interest

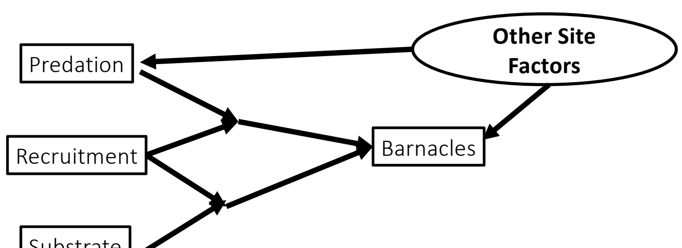


Experiments and Causal Diagrams: Substrate and Barnacles

Experiments and Causal Diagrams: Substrate and Barnacles – Mediators Creep in!

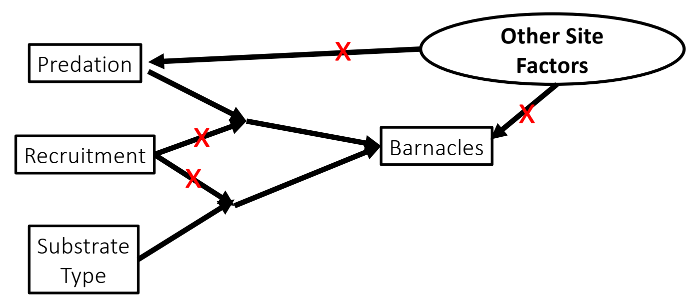


Experiments and Causal Diagrams: Flesh Out the System

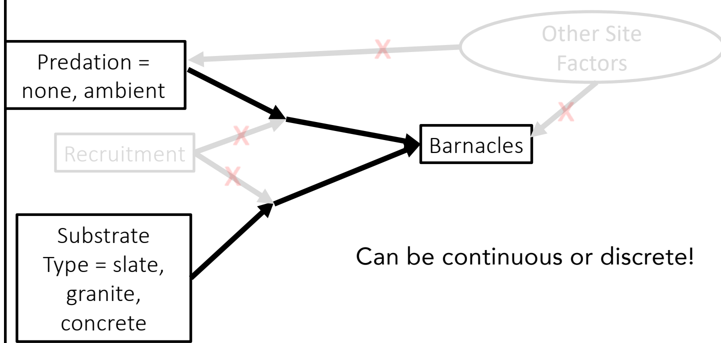


Any other assumptions you see here?

Collapsing the Diagram for an Experiment: Use Just One Site



Experiments and Causal Diagrams: Setting Treatment Levels?

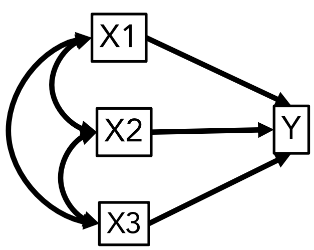


Can be continuous or discrete!

Overview

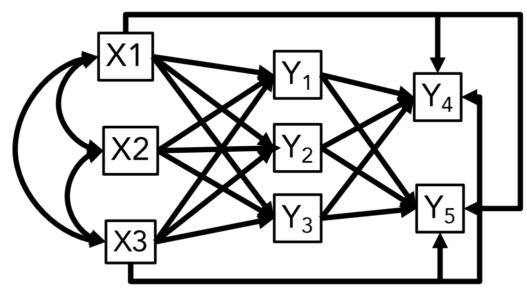
1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

Can We Think of Multiple Regression from a Causal Standpoint?



- We estimate the effect of exogenous variables ***controlling*** for all others
- Covariances implied
- Not controlling for the right variables = bad inference
- Controlling for the wrong variables = bad inference

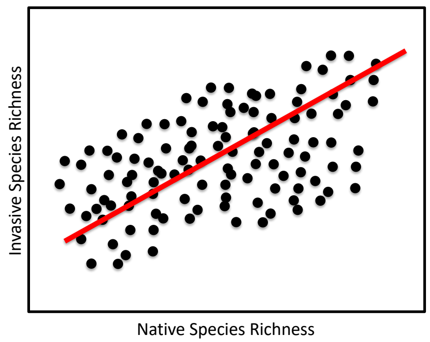
But….We Want to Avoid This



1. What can you actually learn from this?
2. No, everything is not connected to everything

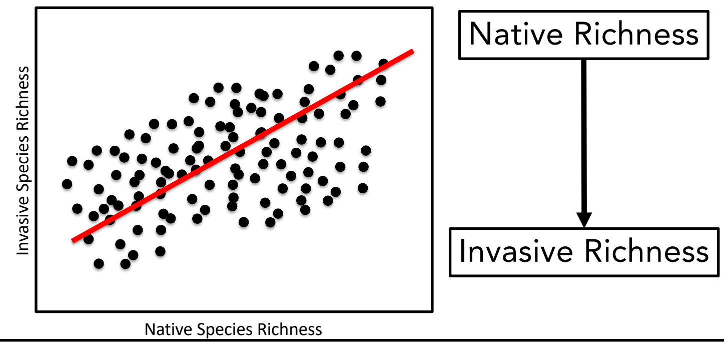
Why Use Multiple Predictors: Simpson's Paradox

- Classic Problem: Does having more native species hinder invasive species success?



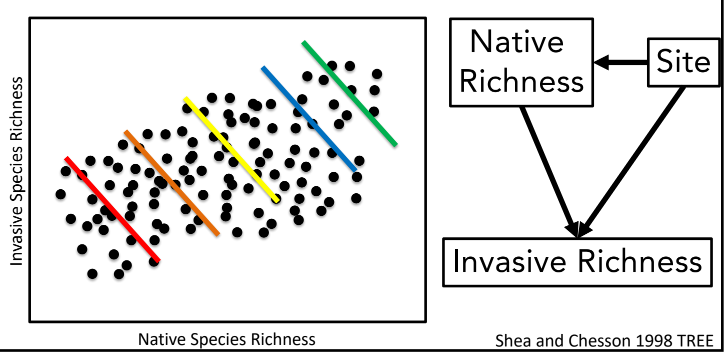
Why Use Multiple Predictors: Simpson's Paradox

- Classic Problem: Does having more native species hinder invasive species success?



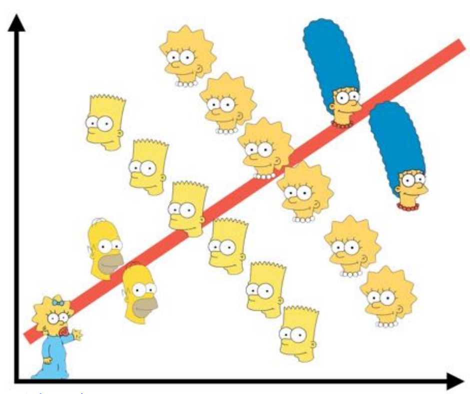
Why Use Multiple Predictors: Simpson's Paradox

- Simpson's Paradox addresses the influence of **confounders** causing flips in signs of relationships

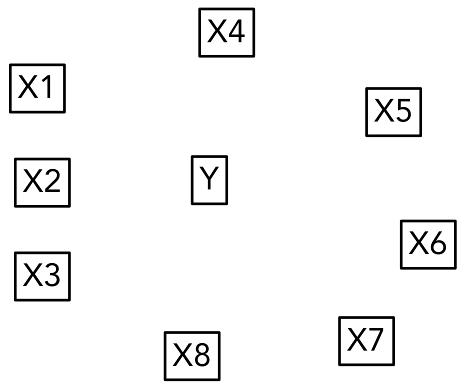


Shea and Chesson 1998 TREE

Simpson's Paradox is Everywhere



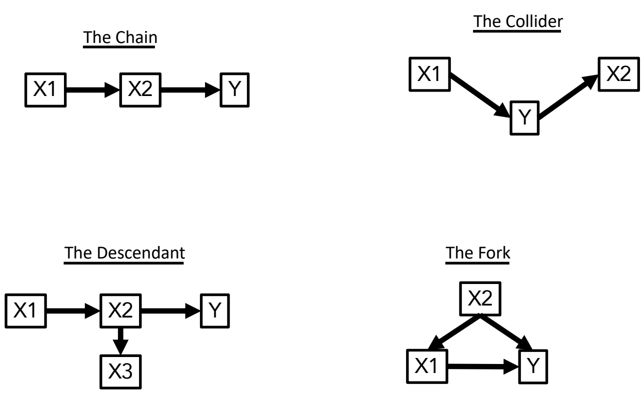
But Which Predictors? All? Some? Which?!?!?!?!



Why Not to Use All Predictors

1. Increased SE on all coefficients
2. Spurious correlations might give false signals
3. Including variables that can obfuscate the signal of causation due to causal graph structure!

Features of Causal Graphs to Watch Out For and Why to use SEM

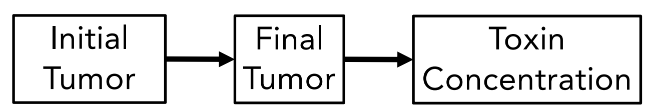


Chain of Foolish Inference

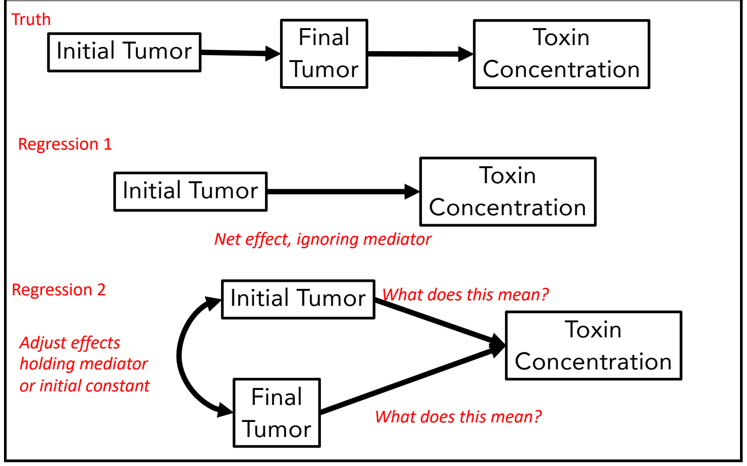
Consider the following:

You are following a cohort of patients with tumors to see if tumor size influences buildup of blood toxins. You measure differences in tumor sizes at t0, and then again at t1, when you take blood samples

Which tumor size(s) do you use as predictors of blood toxin concentrations? Initial, final, or both?



Consider the Consequences

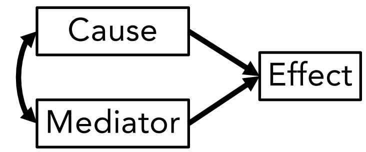


The Chain

Causal Model

In experiments, this is controlling for post-treatment effects

What it can do to Multiple Regression

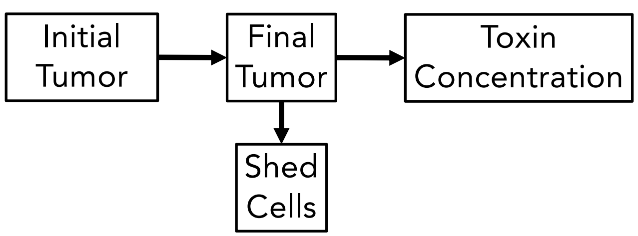


- Mediator blocks cause
- Looks like no or opposite link between cause and effect
- Can be OK if you are also analyzing the cause -> mediator relationship

Will a Child Set us Free?

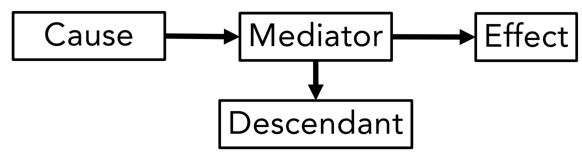
Consider the following:

What if we knew the final tumor was shedding cells, and we had their concentration. Should we control for that?

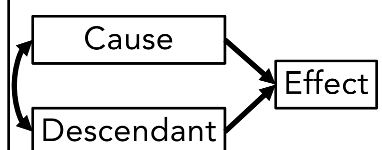


The Descendant Problem

Causal Model



What it would do to Multiple Regression

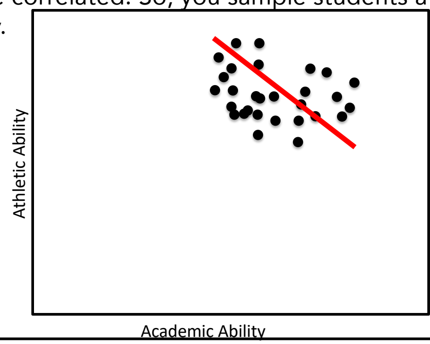


- Descendant still partially blocks cause
- Masks effect of cause – many possibilities

A Collision of Sampling and Regression

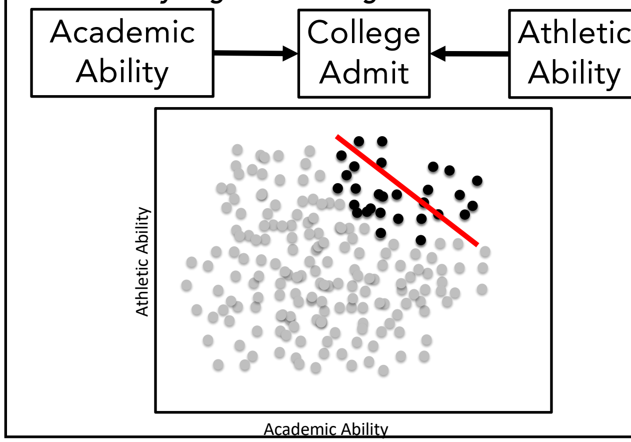
Consider the following:

You are interested to see if academic ability and athletic ability are correlated. So, you sample students at your university.



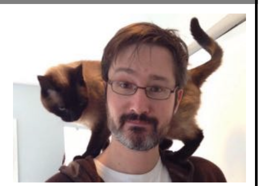
A Collision of Sampling and Regression

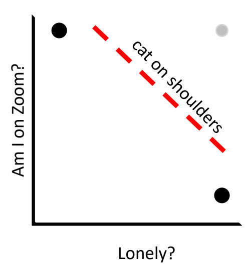
But how do you get into college?



Example of Collider Bias

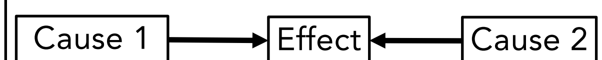
- My cat jumps on my shoulders when she is lonely, I am on Zoom, or both.
- It's more likely she is satisfying one of those conditions at any one time, rather than both
- Thus, if I am on zoom, I know she is likely not lonely
- I'd falsely conclude there's a relationship between my being on zoom and if she is lonely



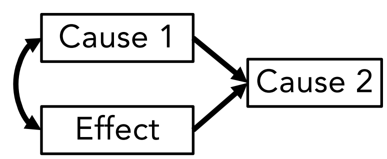


The Collider

Causal Model



What it would do to Multiple Regression

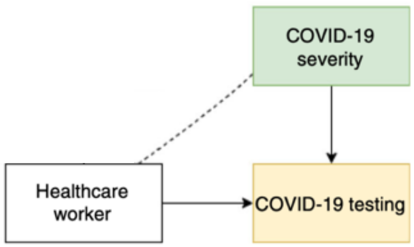


- Conditions on effect, opening path between causes
- Creates spurious correlations

Surely We Wouldn't Fall Prey To Collider Bias…

Preliminary studies suggested that
a) Healthcare workers were likely to get less severe Covid infections
b) Covid infection was lower among smokers

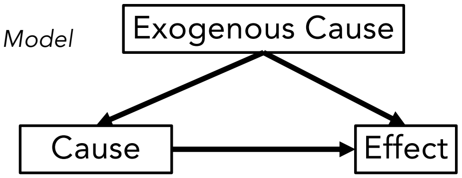
WHY?



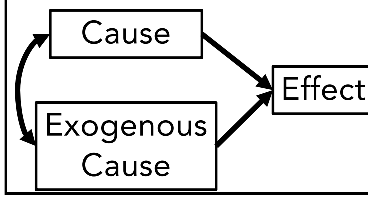
Griffith et al. 2020 Nat. Com.

The Fork

Causal Model



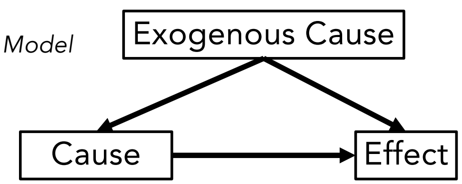
What it would do to Multiple Regression



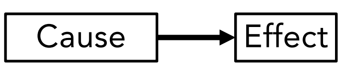
- ***Nothing – this is great!***
- ***The core of piecewiseSEM!***

The Bigger Problem…

Causal Model

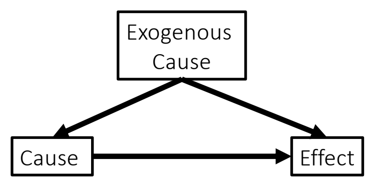


Univariate Regression



- ***The relationship is contaminated – results unreliable!***

The Omitted Variable Bias Problem

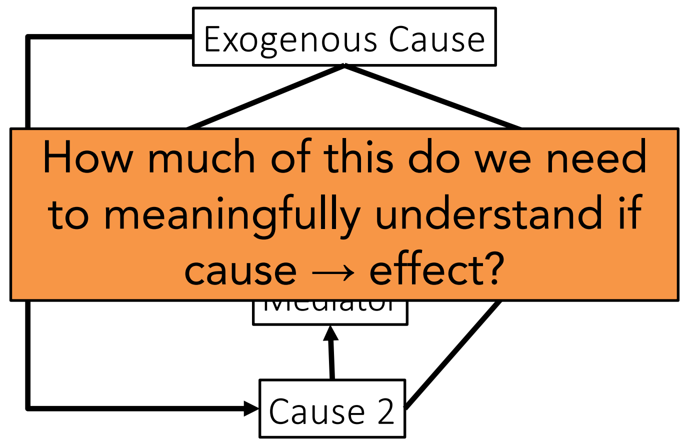


- We assume that sampling means that omitted variables average to 0
  - Omission produces downward bias in SE of coefficients
- But, if omitted variables are correlated causally with a predictor, they likely are not averaged out
- This ***will*** bias your estimates
  - You will not know in what direction

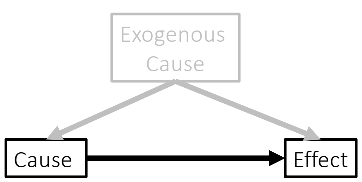
Overview

1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

Build The World to Make Counterfactual Predictions

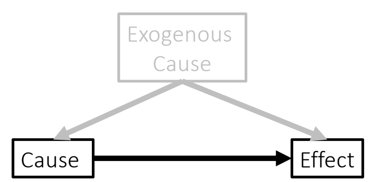


OVB and Causal Identification



This path is not causally identified

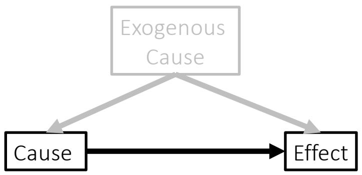
Causal Identification



Your model ***need not be causally identified*** – but be specific that you are only talking about associations

You can only make counterfactual statements if you are confident in causal identification

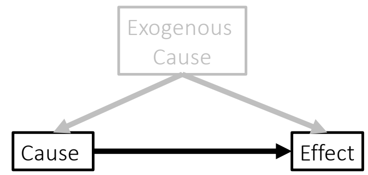
Causal Identification



Causal identification does not require knowing ULTIMATE cause

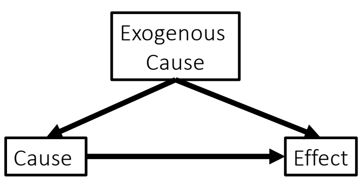
Nor does it require knowing exact mechanisms within a causal pathway

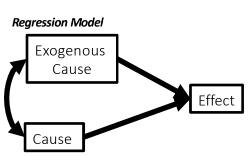
How do we solve this problem?



This relationship is not causally identified

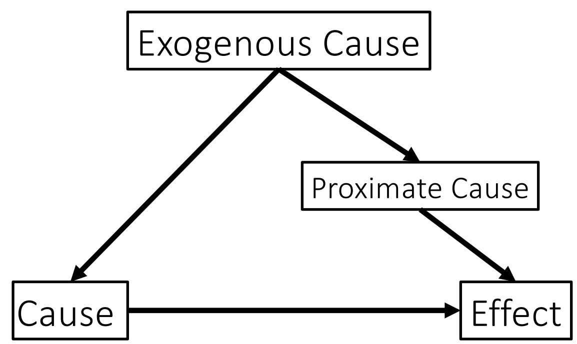
Solution 1: The Backdoor Criteria





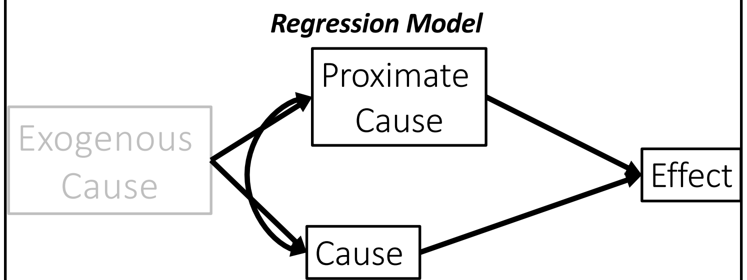
- If we want to know the link between cause and effect given variables that affect both, we must include all variables with a path ***into*** the cause
- So, variables must block all backdoor paths from treatment to outcome
- AND variables must not be descendants of the cause (i.e., no mediators – see the pipe!)

Proximate Backdoors

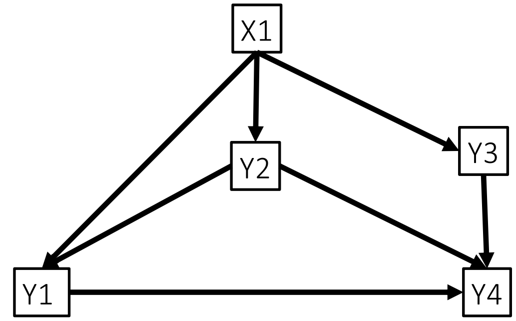


Often we only have proximate variables in a backdoor path. Controlling for just them is sufficient.

Proximate Backdoors and Regression

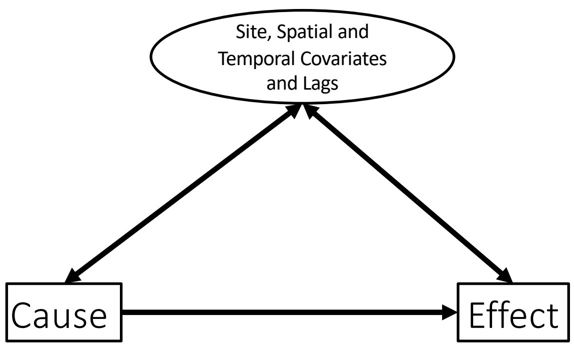


What Variables Block the Back Door?



There are two ways to build a multiple regression with closed back doors to determine if Y1 -> Y4. What are they?

Space and Time Live in the Backdoor



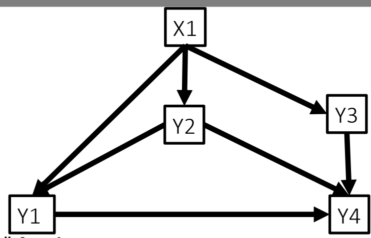
Finding Backdoors with dagitty

- Great package for graph prototyping
- Many ways to analyze graphs as well!

To build a DAG

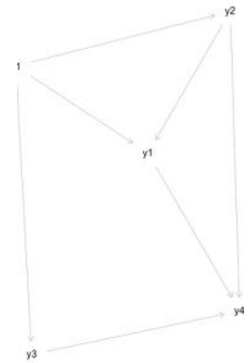
```
g <- dagitty("dag{
     ...
             }")
```

Building a DAG with dagitty



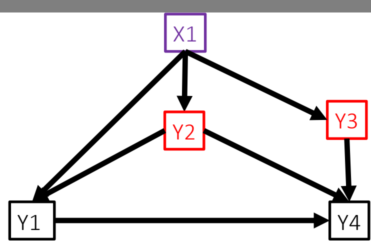

```
g <- dagitty("dag{
  x1 -> y1
  y2 -> y1
  x1 -> y2
  x1 -> y3
  y3 -> y4
  y2 -> y4
  y1 -> y4
           }")
```

Plot your DAG!



```
plot(graphLayout(g))
```

How to Shut the Back Door



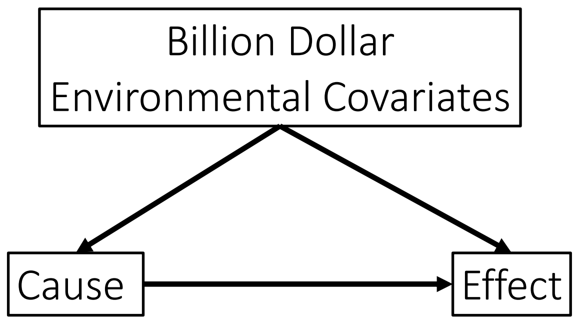

```
> adjustmentSets(g, exposure = "y1",
                 outcome = "y4")

{ y2, y3 }
{ x1, y2 }
```

Exercise: daggity

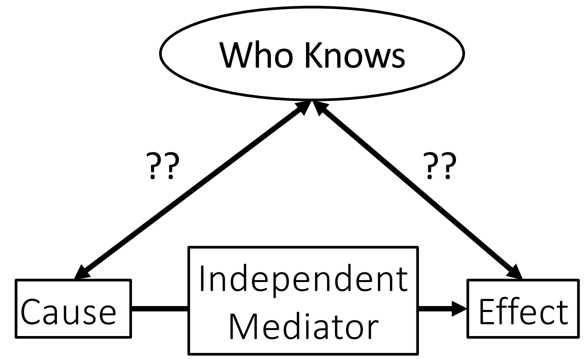
- Sketch a model of 4-5 variables in your system
  - Don't think to hard (that's for later!)
- See if you can figure out how to close any backdoors
- Use `daggity` to find the back doors between a chosen pair

n.b. can represent chains as: a `-> b -> c ->d`
or colliders as: a -> b <- c

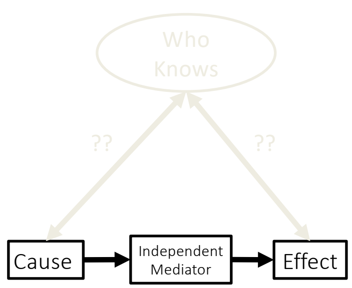
Sometimes We Cannot Shut the Backdoor



Or, we suspect, but don't know, of backdoors

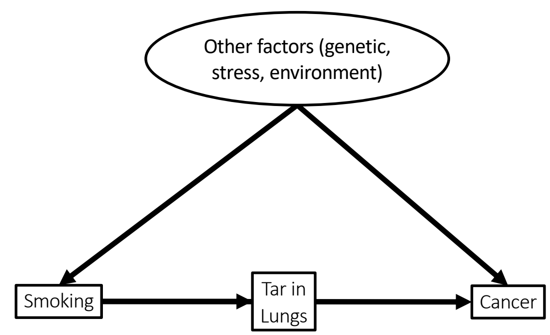


Solution 2: The Front-Door Criterion



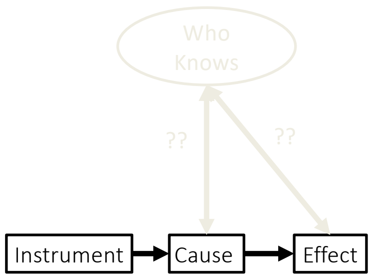
- A variable satisfies the front-door criteria when it blocks all paths from X to Y.
- In practice, you need a causally identified mediating variable unaffected by anything else.
- Thus, the influence of the cause is felt by the effect solely through its mediator.

Example: Smoking and Cancer



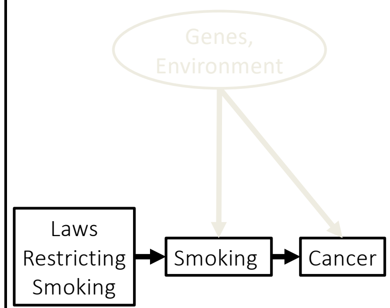
See Pearl's books and papers for the do calculus of this

Front Doors and Instruments



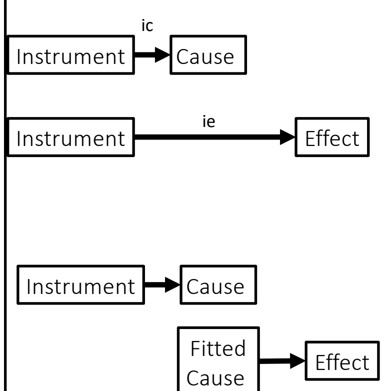
- Instrumental variables are those that *****only***** affect the cause, and have no relation to anything else
- By examining their relationship to both the cause and effect, we can derive an estimate of the causal effect
- Also useful when cause and effect involved in a feedback

Smoking and Cancer as Classic Example



- BUT – are laws a TRUE exogenous variable
- If laws are in response to cancer rates, they might not be
- Unless *a priori* cancer is the same everywhere
- Multiple methods to test validity of instrument

Two Approaches in Instruments



1. Fit two models, leveraging the fact that

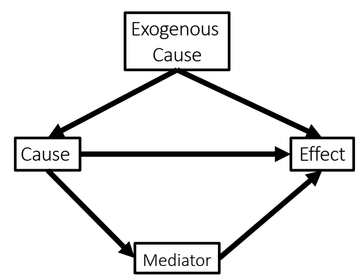
IE = IC * CE

so, CE = IE/IC

2. Fit two models, but only use fitted values of the cause for the second

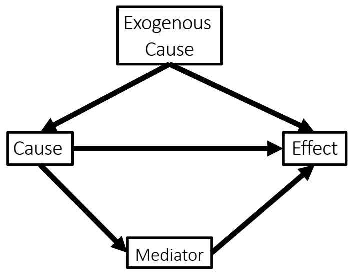
See `ivreg`

Path Represent Causal Relationships – but how solid is our inference?



- We state that a direct link between two variables implies a causal link via a dependence relationship
- We estimate the strength of that relationship
- This is a ***soft*** causal claim

Conditional Independence and Hard Causal Claims



- We assume that two variables not connected are independent, conditioned on their parent influences

Mediator⊥Exogenous Cause | Cause

- This is a HARD causal claim, setting a path to 0
- Testable

Making Sure Pieces of your Model are Causal

- Are there omitted variables?
- If so, are they collinear with included variables?
- Can you shut the back door?
- Can you shut the front door?
- Can I support all causal independence statements?
- Be bold yet honest about causal interpretations!
  - Science advances by others noticing what you left out

Causal Diagrams and Modeling Observational Data

- Yes you can!
- Causal diagrams guide you to the appropriate set of predictors – and fend off testy reviewers
- Sometimes, your model is non-causal, and that's OK!
- If you begin by thinking in terms of a causal system, you will produce more robust meaningful inference

Overview

1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

The Continuum of SEM

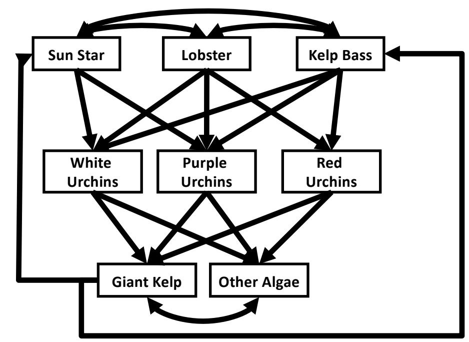
structural equation modeling

exploratory/ model-building applications → confirmatory/ hypothesis-testing/ model-comparison applications

Your goals will inform how you build your model

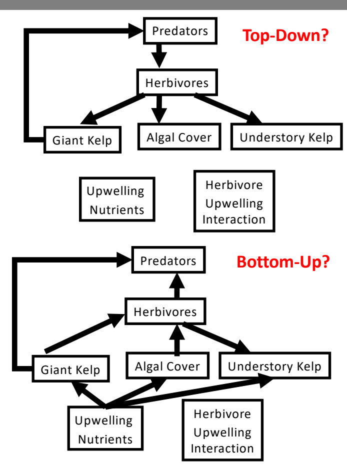
What are the purpose of your modeling?

- Discovery?
- Hypothesis testing?
- Making predictions?



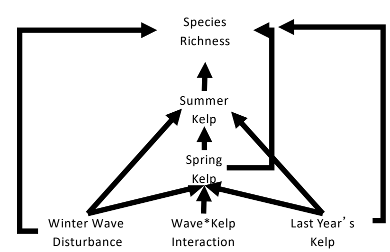
What are the purpose of your modeling?

- Discovery?
- **Hypothesis testing?**
- Making predictions?



What are the purpose of your modeling?

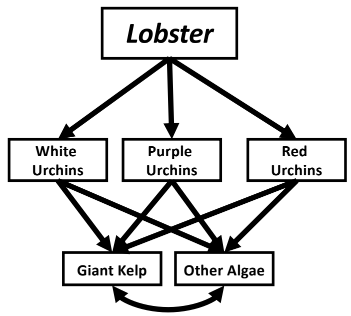
- Discovery?
- Hypothesis testing?
- **Making predictions?**



	High Kelp	Low kelp
High waves	?	?
Low Waves	?	?

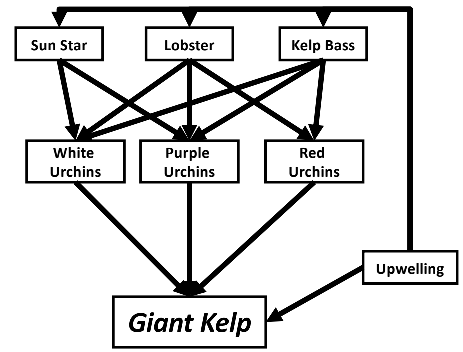
What is the focus of your modeling?

- Driver
- Response
- Mediation
- Theory Testing



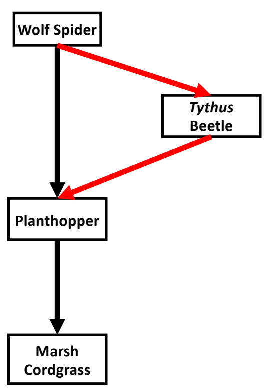
What is the focus of your modeling?

- Driver
- **Response**
- Mediation
- Theory Testing



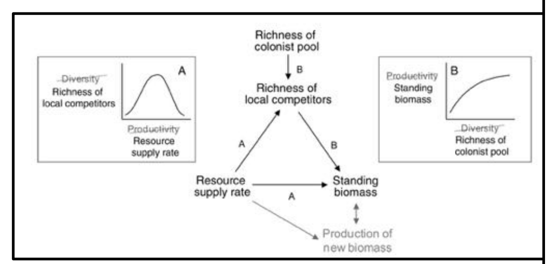
What is the focus of your modeling?

- Driver
- Response
- **Mediation**
- Theory Testing



What is the focus of your modeling?

- Driver
- Response
- Mediation

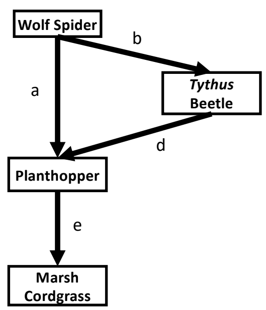


Cardinale et al. 2008

- **Theory Testing**

What is the span of your inference?

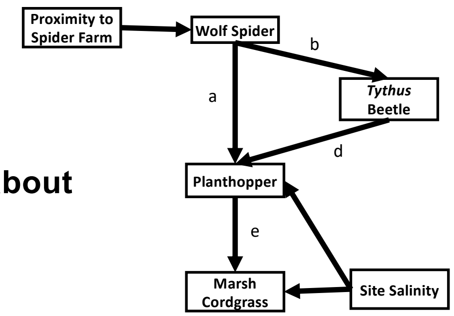
- Local estimation
- Learning about processes



*What are a,b,c,d, and e in *THIS* marsh?*
(e.g., for biocontrol)

What is the span of your inference?

- Local estimation
- **Learning about processes**



Across marshes, what is the relative importance
*of a versus b*d versus site-influences?*

What are you doing this week?

Purpose of modeling effort:
- discovery?
- testing hypotheses?
- making predictions?

Focus of modeling effort:
- driver focused?
- response focused?
- mediation focused?
- theory testing focused?

Span of inference:
- doing inferential estimation?
- learning about processes?

Overview

1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

The continuum of SEM

Structural Equation Modeling

exploratory/
model-building
applications → confirmatory/
hypothesis-testing/
model-comparison
applications

It all starts with an underlying model!

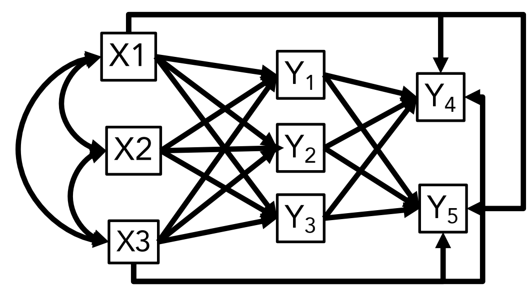
Exploratory SEM

- Evaluate *multiple models*, tweaking along the way
- Suspected causal relationships, testing strength of paths and if they are effectively zero or not
- Results should be proposed as *preliminary* until further confirmatory testing can be conducted

Confirmatory SEM

- Evaluate *a priori* models and attempt to falsify
- Interested in *strength* of relationships
- If models fails, can go to *Exploratory* – or, falsification is good in and of itself
- *Nested comparisons* can test multiple hypotheses about how systems work
- *Model comparison, Cross-Validation, etc.* also possible

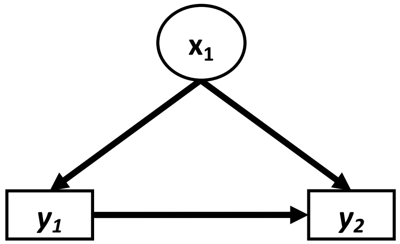
The First Elephant in the Room: The Kitchen Sink Model



1. What can you actually learn from this?

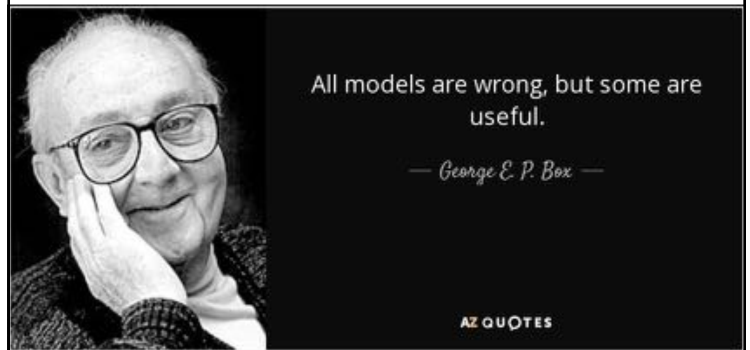
2. No, everything is not connected to everything

The Second Elephant in the Room: You Will Not Measure Everything

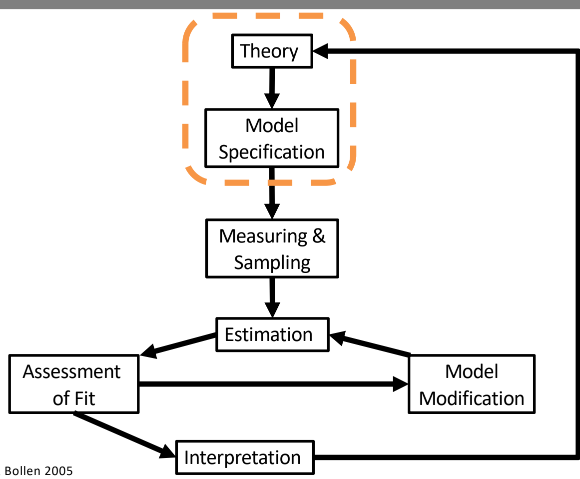


- You will not be able to measure everything
- But, build an initial model that shows you what you HAVE to measure to achieve causal validity
- See also coping with ***Omitted Variable Bias***

The Final Elephant in the Room

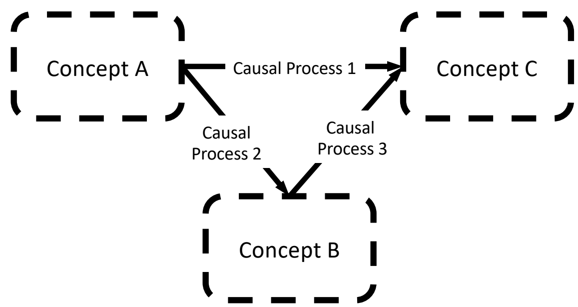


SEM as a Unifying Process



After Grace & Bollen 2005

The Structural Equation Meta-Model (SEMM)



My Research Program as a Meta-Model

Food Web Structure → Ecosystem Function

Environmental Change → Food Web Structure

Environmental Change → Ecosystem Function

Targeting Your Question

Food Web Structure → Ecosystem Function

Environmental Change → Food Web Structure

Environmental Change → Ecosystem Function

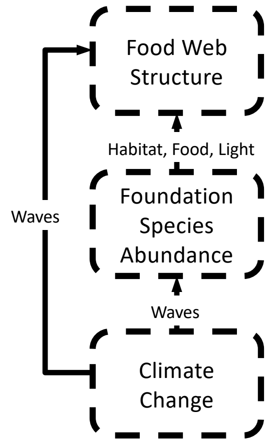
Targeting Your Question

Focused on driver & Mediator

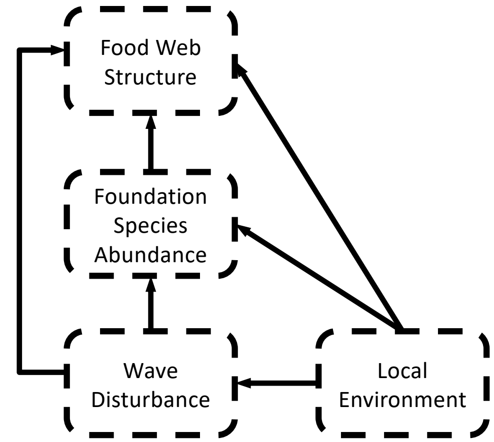
Climate Change → Foundation Species Abundance → Food Web Structure

Climate Change → Food Web Structure

Identify Relevant Processes



Do you need to shut a conceptual backdoor?

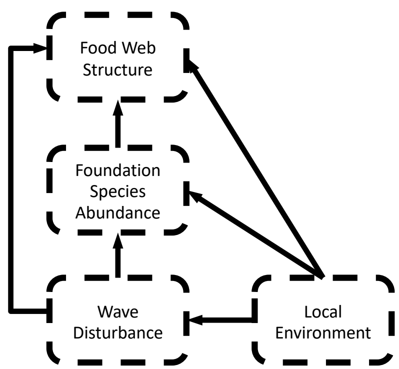


What did I want to do?

Purpose of modeling effort:
- testing hypotheses
- making predictions

Focus of modeling effort:
- driver focused
- mediation focused

Span of inference:
- learning about processes



Meta-Model Your Research

http://bit.ly/sem-eeb-models-2021

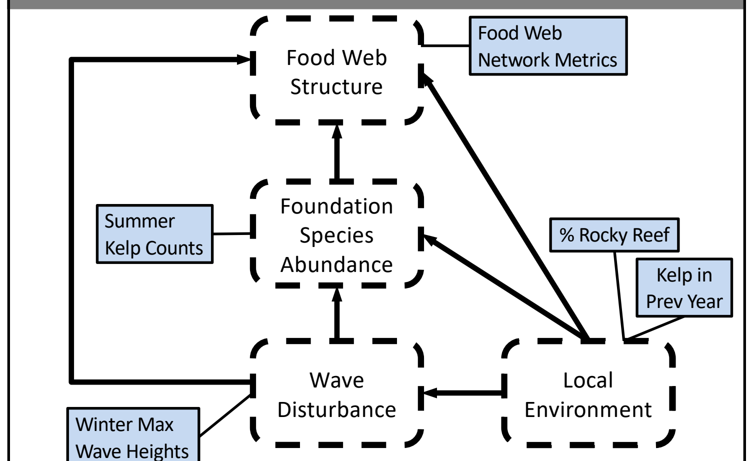
Overview

1. What is Causality?
2. Thinking about causal model structure to build valid inference
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

Complex Systems are Complex

Sampling of Rocky Reefs 2000-2009

Matching Data to Concepts

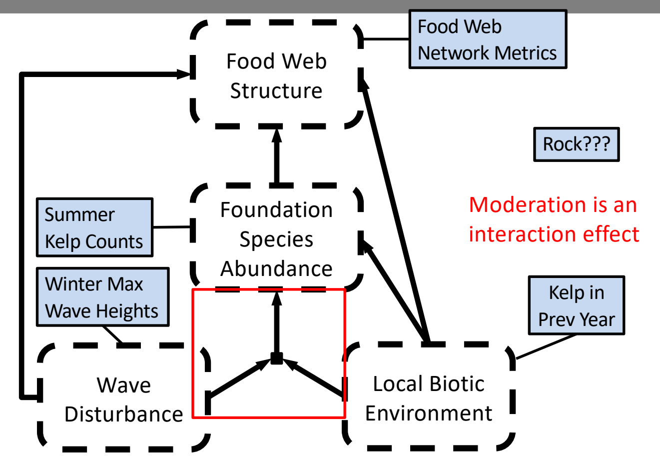


Adding Biological Realism

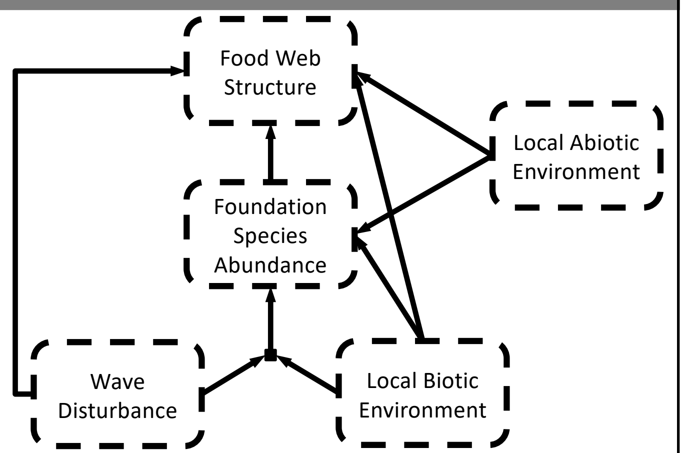
Problem 1: Kelp moderates disturbance

- More Kelp = Smaller Disturbance?
- BUT no effect on kelp that isn't present…

Solution 1: Moderation



But: Maintain Backdoor Blockage!

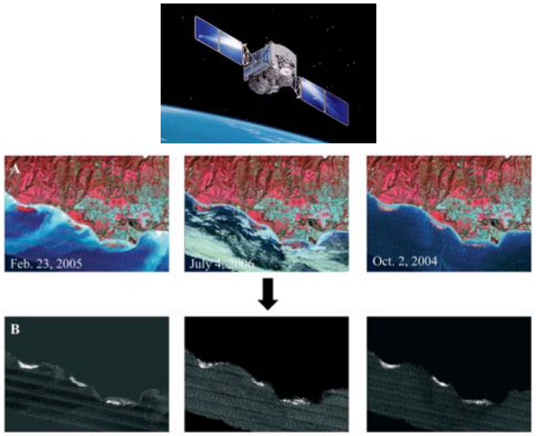


Natural History Creates Problem

Problem 2: Kelp regrows quickly

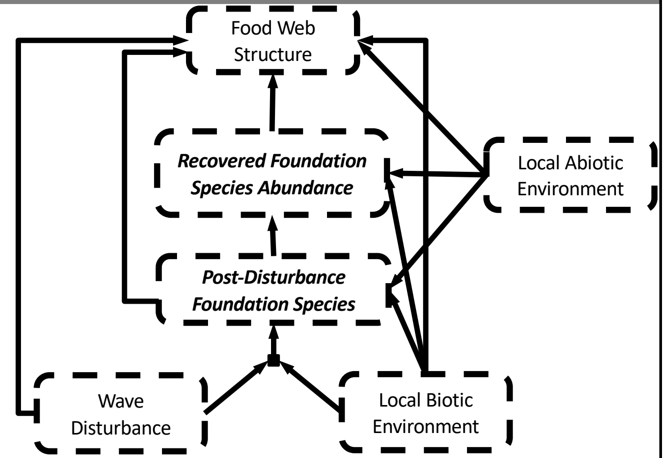
- It's a jungle by summer if nutrients are present
- Need to see if kelp was actually removed in winter!

Measuring Realized Disturbance via Satellite Measurements

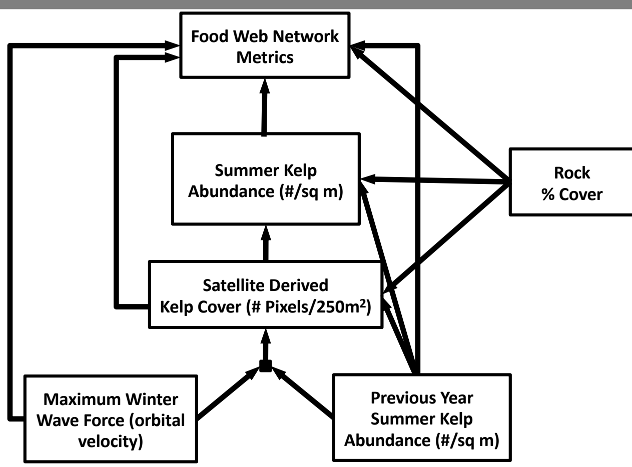


Cavanaugh et al 2011 MEPS

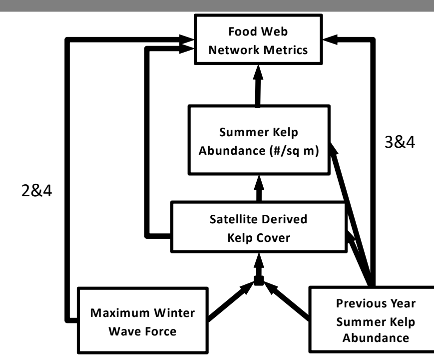
Incorporate Natural History of Disturbance



Model with Observed Variables



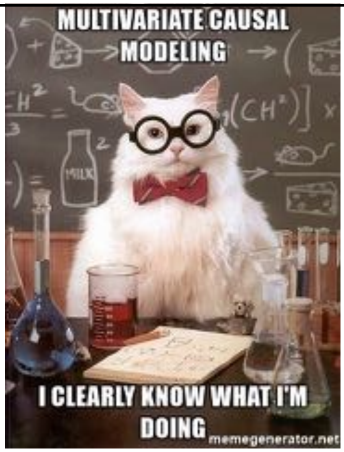
Goal: Hypothesis Evaluation



1. Full Model
2. No waves -> Food Web effect
3. No effect of previous year's kelp -> Food Web
4. No effect of either waves **or** previous kelp -> Food Web

The Process of Model Building

1. Make a conceptual meta-model
2. Ensure meta-model's causal structure meets your research goals
3. Reify your model based on system natural history (a bigger model!) and available data
4. Ensure causal structure is still intact



Make your model based on data!
http://bit.ly/sem-eeb-models-2021