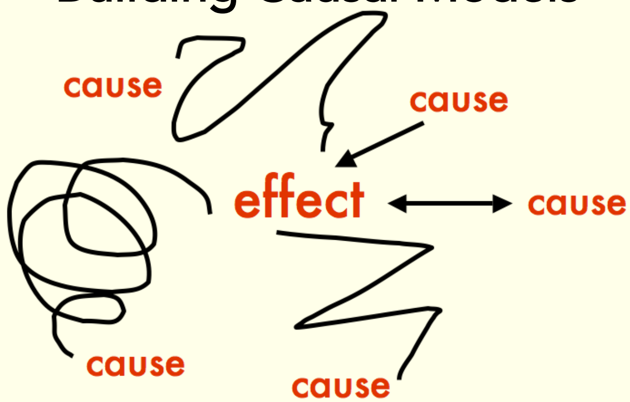
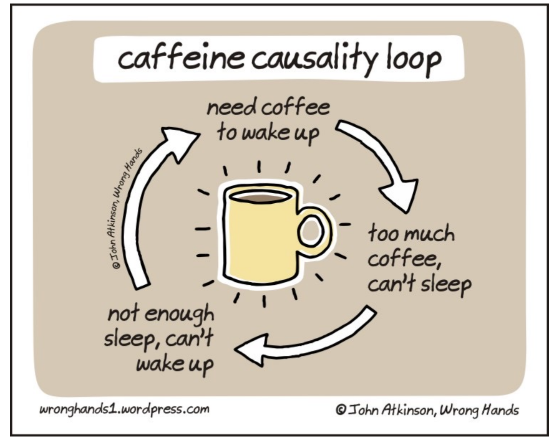
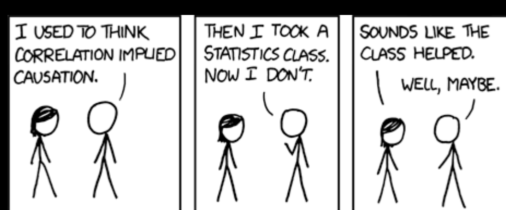
Building Causal Models



1



2



xkcd

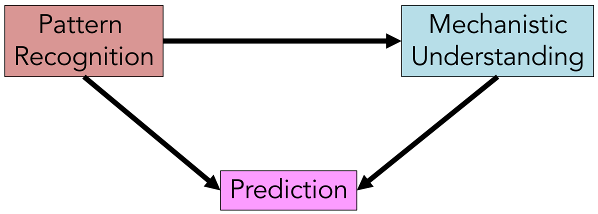
3

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

4

Goals of Science



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

5

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

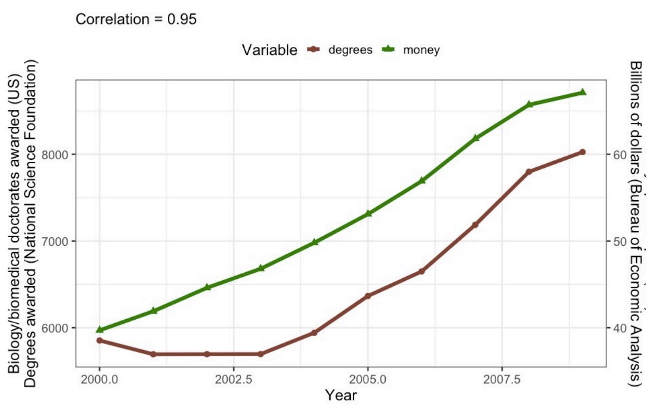
1. Observation – Cause is associated with effect

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

Pearl and Mackenzie 2018

6

Observation: Can We Learn Anything?



http://tylervigen.com/

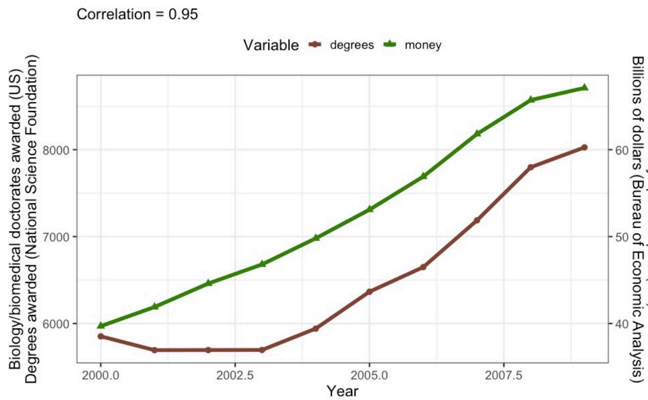
7

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

-Jim Grace

8

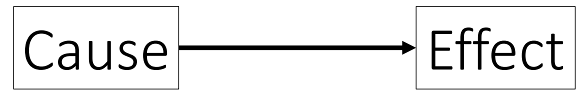
Where's the Fire?



http://tylervigen.com/

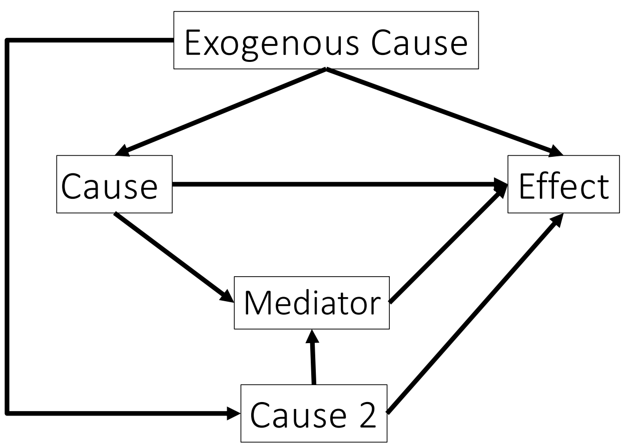
9

What We Want to Evaluate



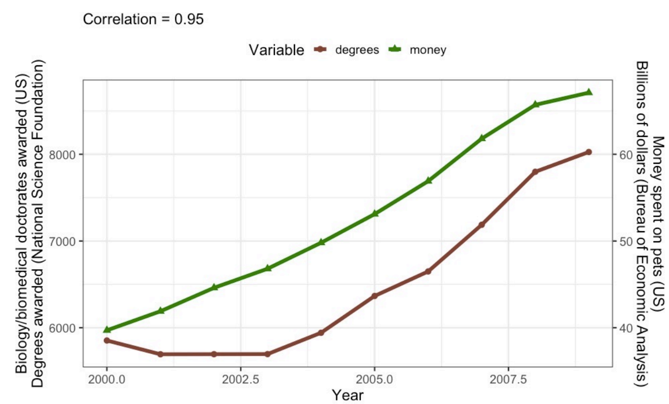
10

But This is the World



11

What Is The World Behind This Association?



http://tylervigen.com/

12

Do You Need to be Doing Causal Inference?

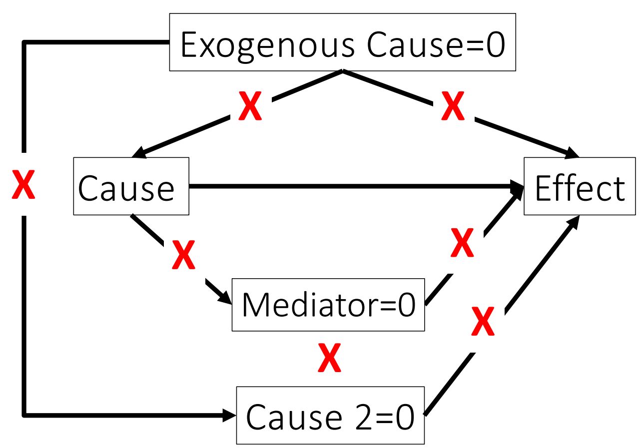
- No!
  - Not all studies will provide causal links between different variables of interest
  - If the study goal is predictive or descriptive rather than causal, this might not be needed
- But…
  - We cannot hope to understand the world without developing an understanding of causal associations
- Indeed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

13

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

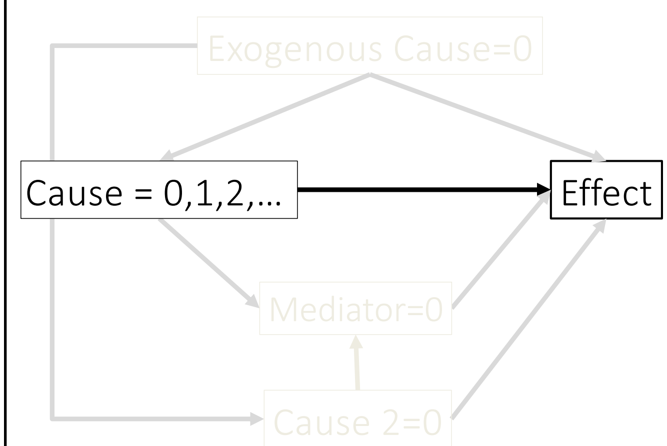
14

Intervention: Experiments! What can we learn?



15

Experiments: Manipulate Cause of Interest

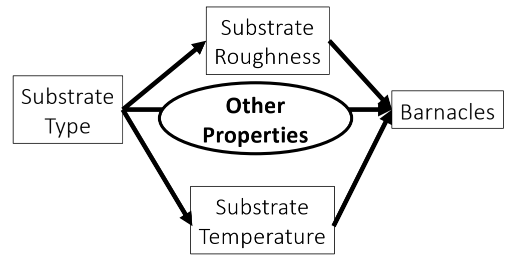


16

Experiments and Causal Diagrams: Substrate and Barnacles

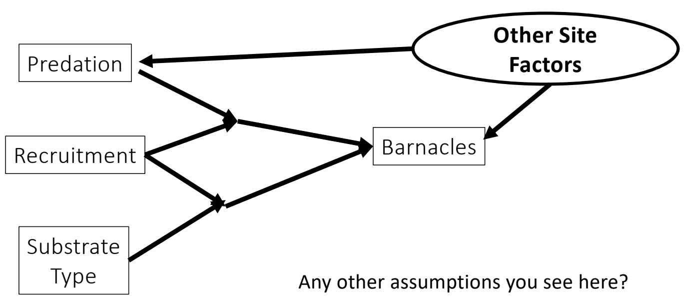
17

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



18

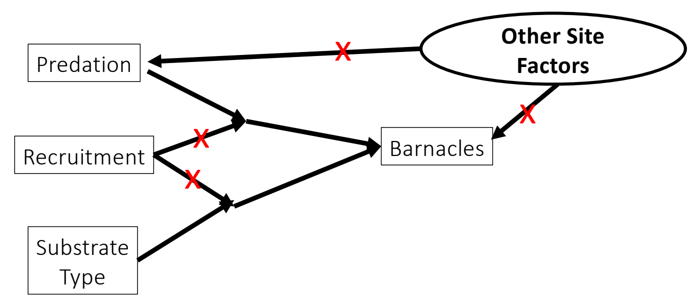
Experiments and Causal Diagrams: Flesh Out the System



Any other assumptions you see here?

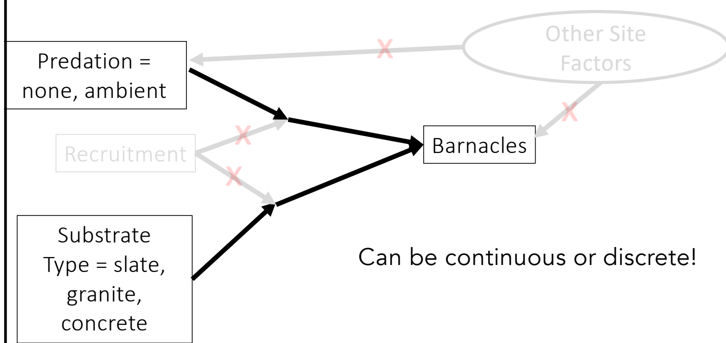
19

Collapsing the Diagram for an Experiment: Use Just One Site



20

Experiments and Causal Diagrams: Setting Treatment Levels?



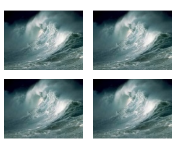
21

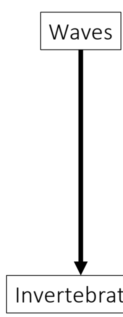
Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

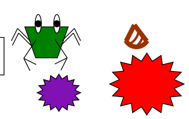
22

Counterfactual Thinking: What would happen if….



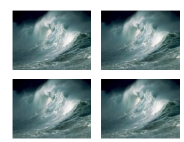


- We want to estimate an **Average CAUSAL Effect** of waves on invertebrates
- We observe Inverts With Waves – Inverts Without Waves
- This is a *POPULATION* phenomenon – the **Average Treatment Effect**

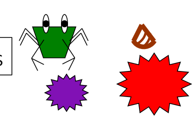


23

Counterfactual Thinking: What would happen if….

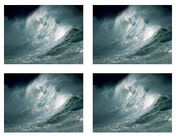


- From our measurements, we only observe what happens with waves or no waves in the sample we have
- What would have happened if those same replicates had opposite the "treatments"? Would our observation hold?

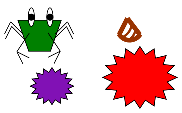


24

The Potential Outcomes Framework for Causal Thinking



Waves → Invertebrates



- ATE = [Observation with treatment –

  What would have happened to treatment plots with no treatment] +

  [What would have happened to treatment plots with no treatment –

  Observation without treatment]
- This latter term is known as ***selection bias***
- This is the **Potential Outcomes** Framework

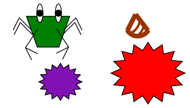
25

Average Causal Effect: Difference Between Observed Reality and Potential Outcome

Observation of inverts at wavy site –
what would have been there with no waves

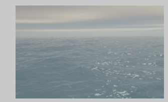
Observed

Sites 1 and 2



–

Imagined



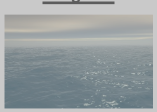
Sites 1 and 2

26

Selection Bias: Difference Between Potential Outcome and Observed Reality

What would have been at wavy invert site if no waves –
observation of inverts at sites without waves

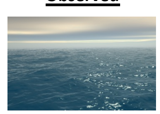
Imagined



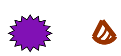
Sites 1 and 2

–

Observed



Sites 3 and 4



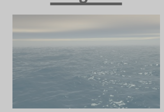
No Selection Bias

27

Selection Bias: Difference Between Potential Outcome and Observed Reality

What would have been at wavy invert site if no waves –
observation of inverts at sites without waves

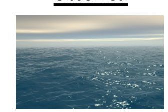
Imagined



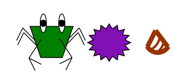
Sites 1 and 2

–

Observed



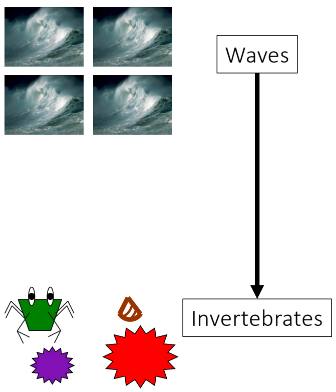
Sites 3 and 4



Some Selection Bias

28

How Do We Overcome Selection Bias?

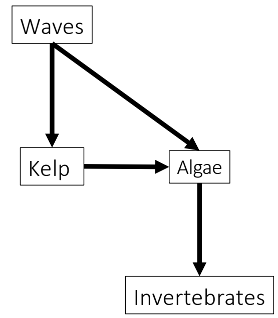


- ATE = Average Causal Effect + Selection Bias
- Our job is to remove selection bias
- **Experiments** let us remove selection bias by removing drivers of bias
- **Observational studies** let us remove selection bias via carefully constructed models based on DAGS

29

Uniting DAGs and Counterfactual Thinking: What would Happen If….for the entire network!

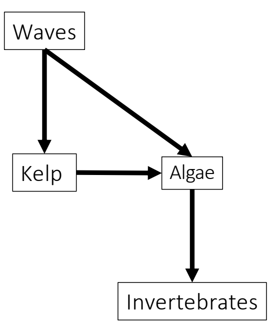
Observation



What We Want to Think About

30

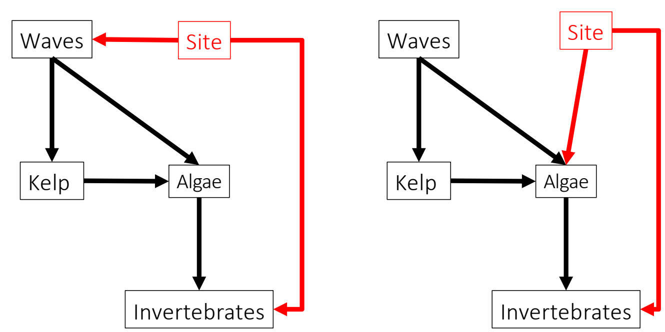
DAGS Show Possible Sources of Sampling Bias



- If we only chose sites with kelp, what would we have missed?
- If we only chose sites with sparse algae, what would we have missed?
- If we have biased sampling, what do we need to bring to our models to make it right and counter attenuation of effects?

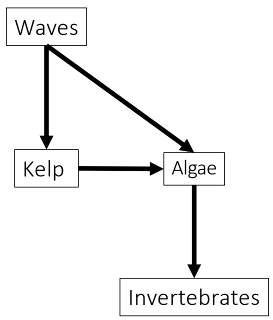
31

DAGS Show Possible Open Back-Doors to Selection Bias



32

DAGS + Counterfactuals = Clear Inference



- With a DAG, we can see that there are no external sources of selection bias
- We can use counterfactual thinking here to understand how changing waves should cascade through the system
- In practice, we can see what variables might obscure our counterfactual inferences

33

What do you need to control for to have valid counterfactual inference?

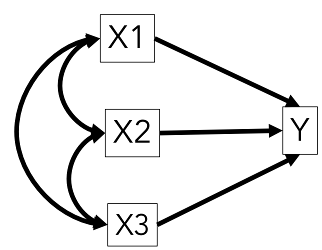
34

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

35

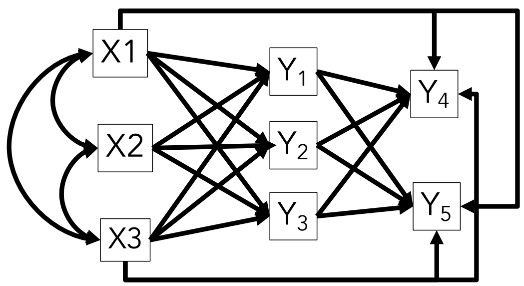
Can We Think of Multiple Regression from a Causal Standpoint?



- 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

36

But….We Want to Avoid This



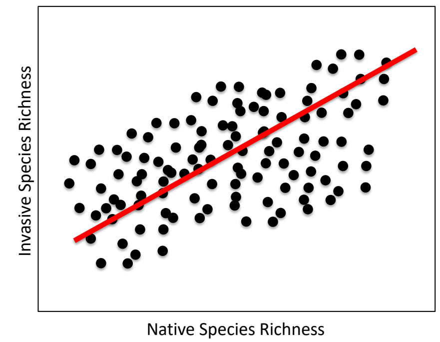
1. What can you actually learn from this?

2. No, everything is not connected to everything

37

Why Use Multiple Predictors: Simpson's Paradox

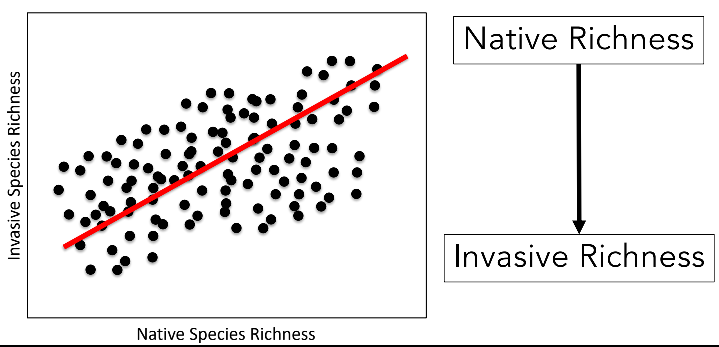
- Classic Problem: Does having more native species hinder invasive species success?



38

Why Use Multiple Predictors: Simpson's Paradox

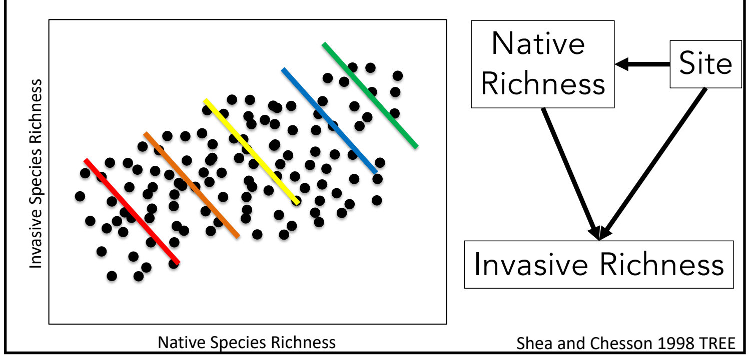
- Classic Problem: Does having more native species hinder invasive species success?



39

Why Use Multiple Predictors: Simpson's Paradox

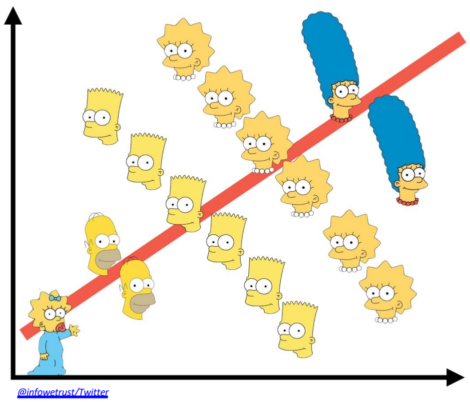
- Simpson's Paradox addresses the influence of **confounders** causing flips in signs of relationships



Shea and Chesson 1998 TREE

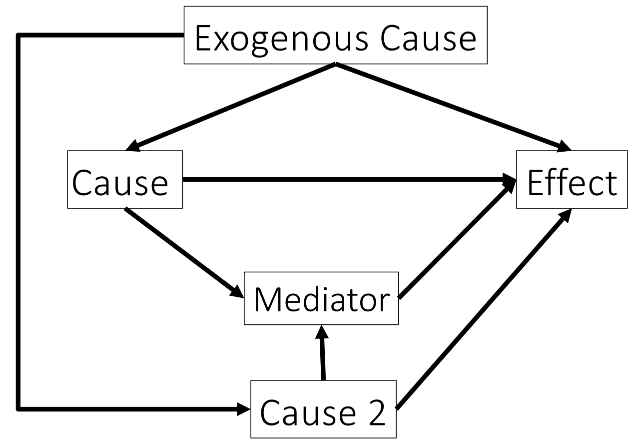
40

Simpson's Paradox is Everywhere



41

DAGs Let us Discover and Disentangle Simpson's Paradox



42

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

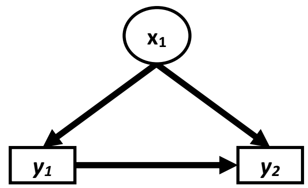
43

What do we mean when we say 'correlation is not causation'?

What is the actual problem?

44

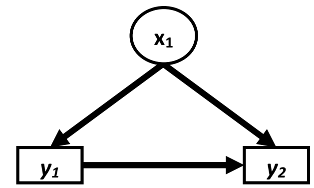
The Back-Door Effect *sensu* Judea Pearl



X1 is a **confounder** - We need to find a way to shut the back door!!!

45

Open Back Doors and Omitted Variable Bias

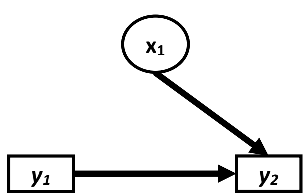


If we omit x1 from a model, our results will be **BIASED**

46

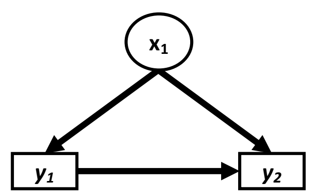
What is Omitted Variable Bias?

Uncorrelated unmeasured driver only causes additional error



- Omitted variables average to 0 with good sampling
- No influence on estimates of effect of y1 on y2
- Downward bias in SE

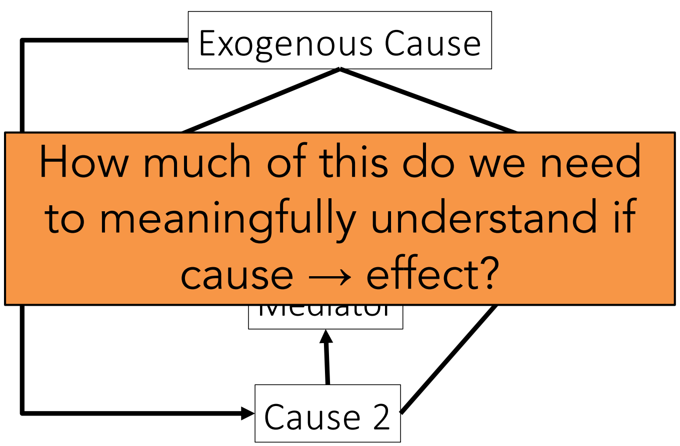
Uncorrelated unmeasured driver causes *bias*



- Omitted variables will correlate with y1
- Will contaminate estimate effect of y1 on y2
- You will not know the direction/magnitude of ***bias***

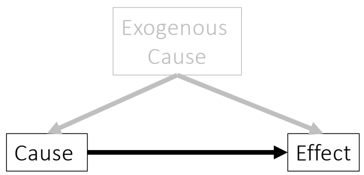
47

How do we grapple with Omitted Variable Bias in this World?



48

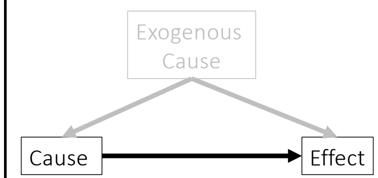
OVB and Causal Identification



This path is not causally identified

49

Causal Identification

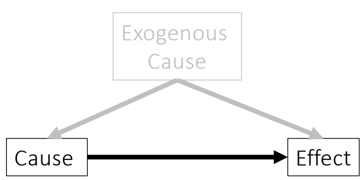


Your model ***need not be causally identified*** – but be specific that you are only talking about associations

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

50

Causal Identification

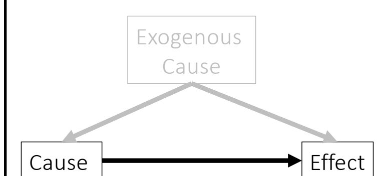


Causal identification does not require knowing ULTIMATE cause

Nor does it require knowing exact mechanisms within a causal pathway

51

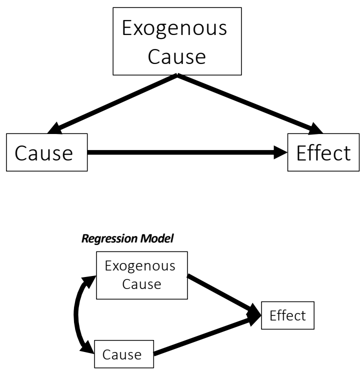
How do we solve this problem?



This relationship is not causally identified

52

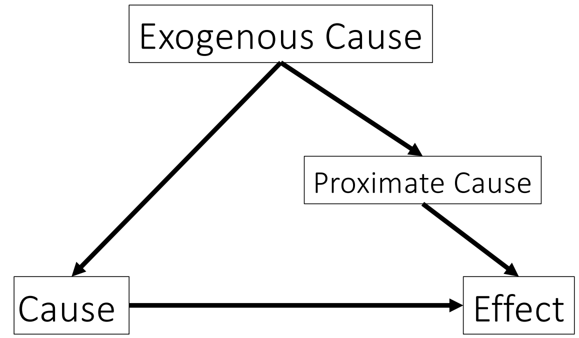
Solution 1: The Backdoor Criteria



- 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!)

53

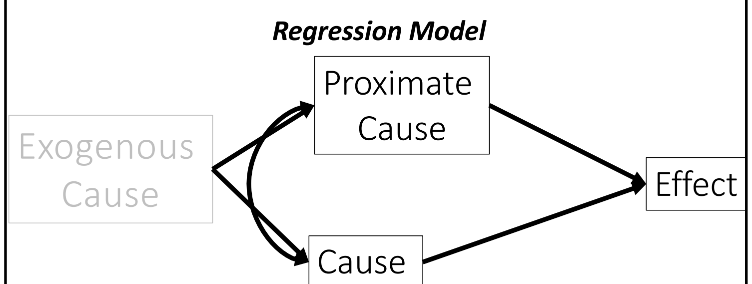
Proximate Backdoors



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

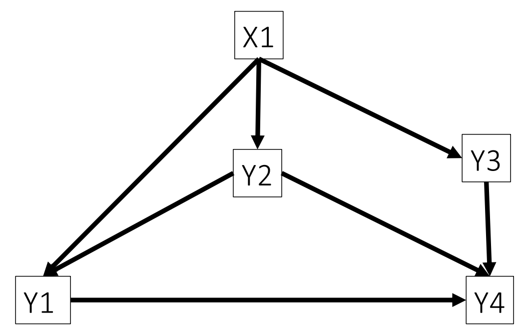
54

Proximate Backdoors and Regression



55

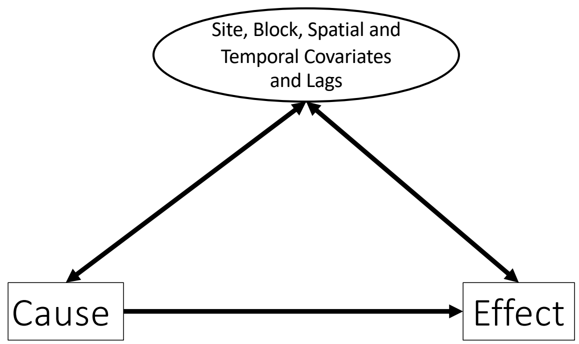
What Variables Block the Back Door?



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

56

Space and Time Live in the Backdoor



57

Finding Backdoors with dagitty

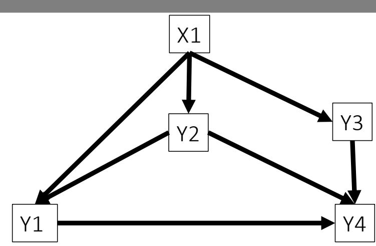
- Great package for graph prototyping
- Many ways to analyze graphs as well!

To build a DAG

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

58

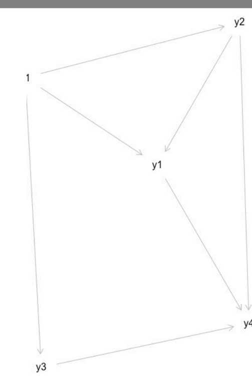
Building a DAG with dagitty




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

59

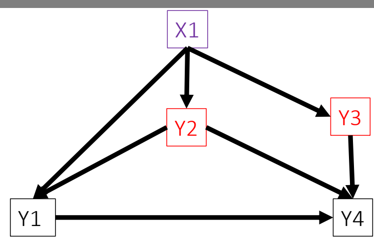
Plot your DAG!




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

60

How to Shut the Back Door




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

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

61

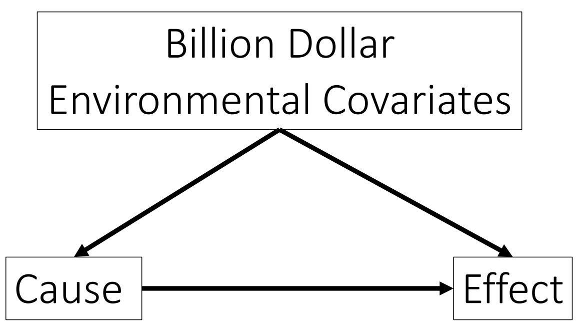
Exercise: daggity

- 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

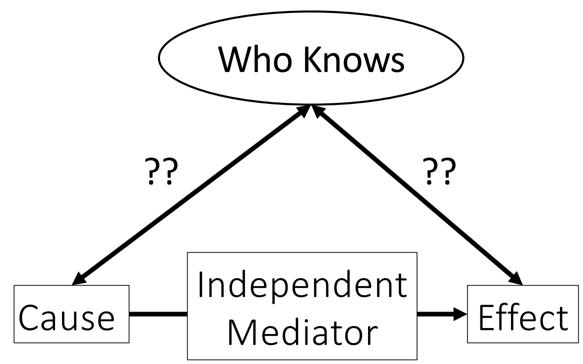
62

Sometimes We Cannot Shut the Backdoor



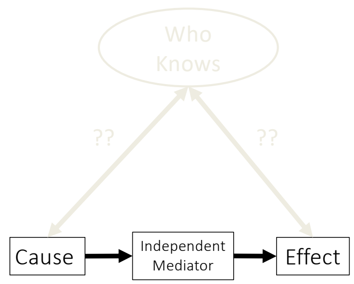
63

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



64

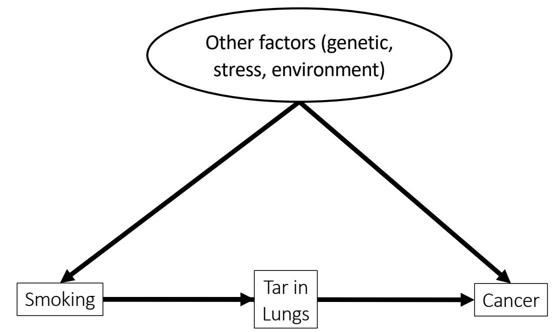
Solution 2: The Front-Door Criterion



- 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.

65

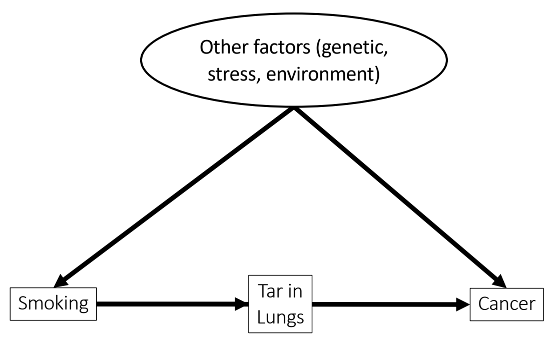
Example: Smoking and Cancer



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

66

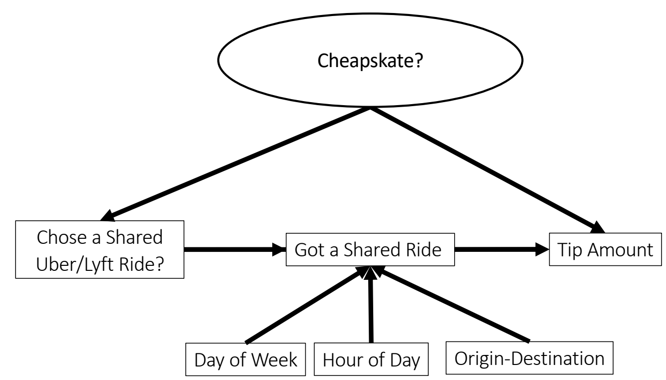
Example: Smoking and Cancer



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

67

Example: Ride Sharing and Tipping

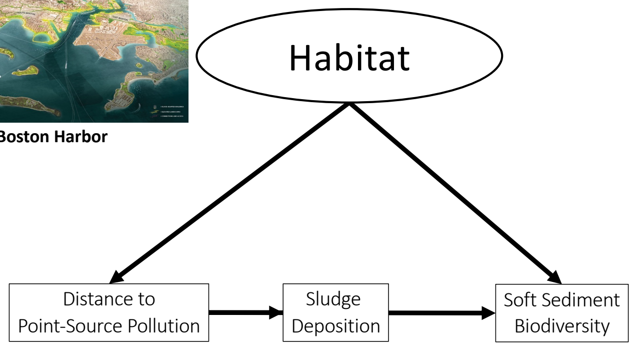


Bellemare et al. 2022 – The Paper of How

68

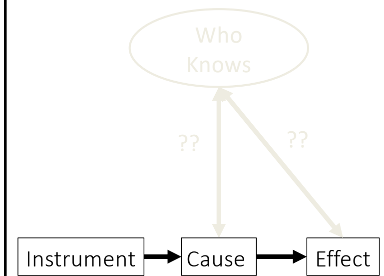
Example? Pollution and Impact

Boston Harbor



69

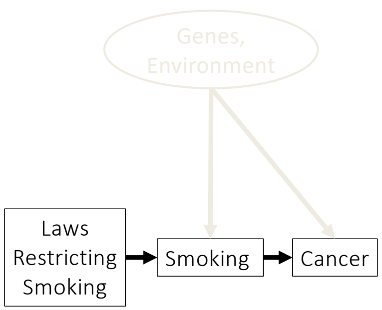
Front Doors and Instruments



- 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

70

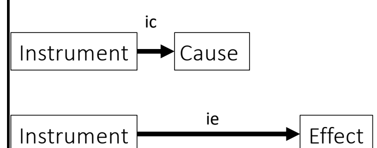
Smoking and Cancer as Classic Example



- 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

71

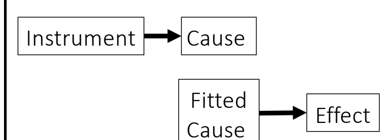
Two Approaches in Instruments



1. Fit two models, leveraging the fact that

IE = IC * CE

so, CE = IE/IC

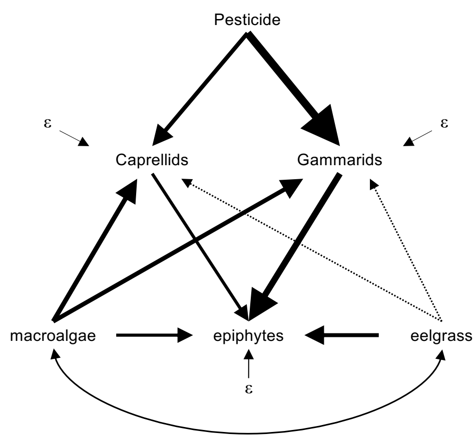


2. Fit two models, but only use fitted values of the cause for the second

See `ivreg`

72

IV In Context with Seagrass



73

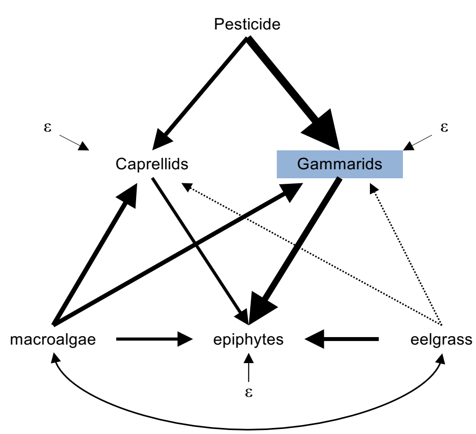
What is a Good Instrument?

- Fully exogenous to system
  - This is VERY HARD to determine
  - Think of an experimental manipulation
- Has a causal effect on cause of interest
- Or – at minimum least no correlation with response
  - Can incorporate other covariates, but, **weak instrument**

See Kendall 2015 A statistical symphony: Instrumental variables reveal causality and control measurement error

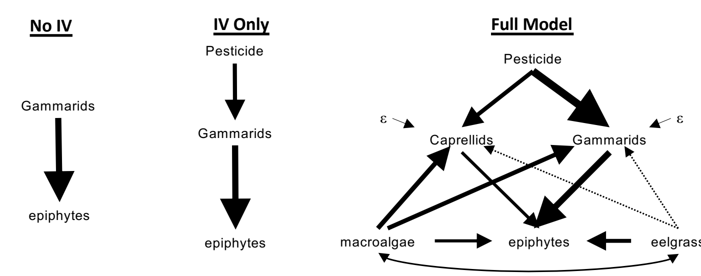
74

Who are the Instruments for Gammarids?



75

The Effects of IV Analysis in SEM



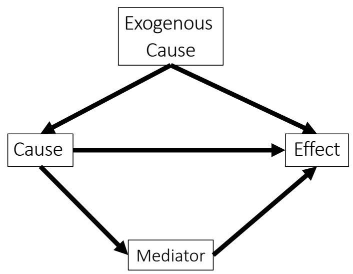
Effect of Gammarids on Epiphytes:

No IV: -0.703 ± 0.115
IV Only: -0.996 ± 0.167
Full Model: -0.886 ± 138

See Grace 2021 MEE

76

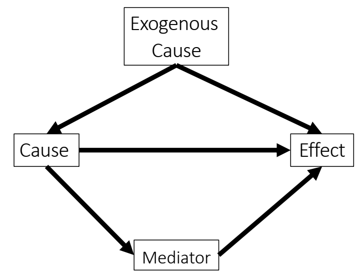
Path Represent Causal Relationships – but how solid is our inference?



- 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

77

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

78

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

79

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

80

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

81

The Continuum of SEM

structural equation modeling

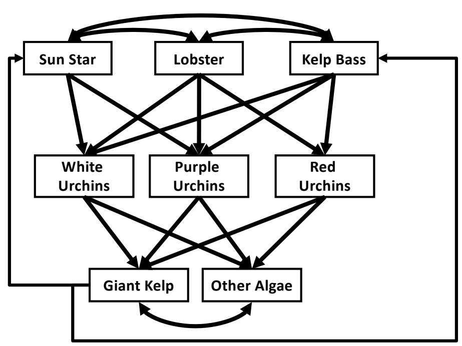
exploratory/ model-building applications → **confirmatory/ hypothesis-testing/ model-comparison applications**

Your goals will inform how you build your model

82

What are the purpose of your modeling?

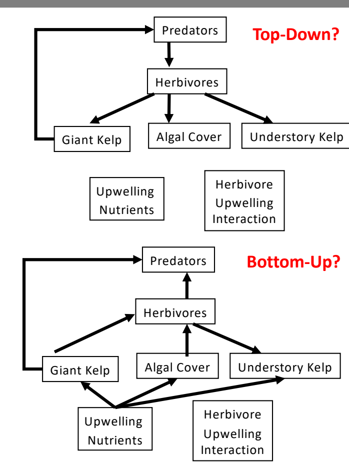
- Discovery?
- Hypothesis testing?
- Making predictions?



83

What are the purpose of your modeling?

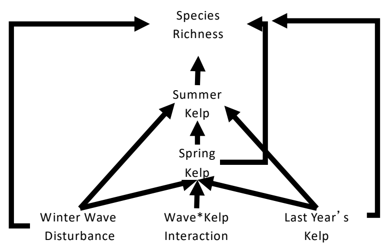
- Discovery?
- **Hypothesis testing?**
- Making predictions?



84

What are the purpose of your modeling?

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

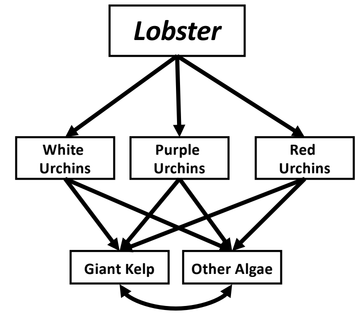


	High Kelp	Low kelp
High waves	?	?
Low Waves	?	?

85

What is the focus of your modeling?

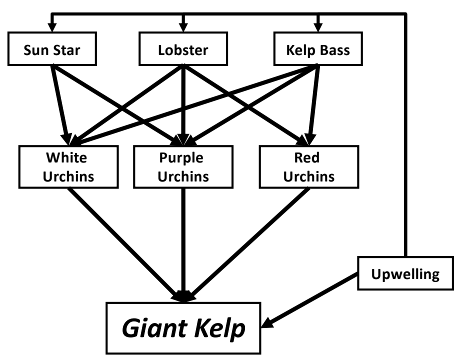
- Driver
- Response
- Mediation
- Theory Testing



86

What is the focus of your modeling?

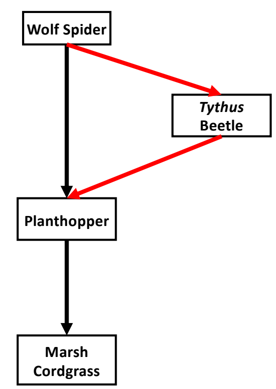
- Driver
- **Response**
- Mediation
- Theory Testing



87

What is the focus of your modeling?

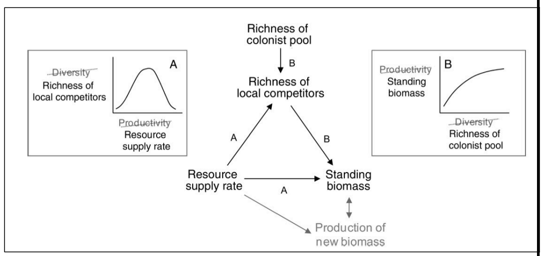
- Driver
- Response
- **Mediation**
- Theory Testing



88

What is the focus of your modeling?

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

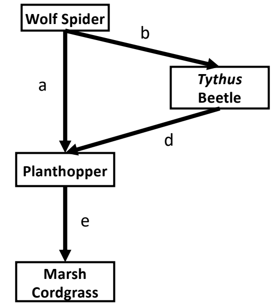


Cardinale et al. 2008

89

What is the span of your inference?

- Local estimation
- Learning about processes

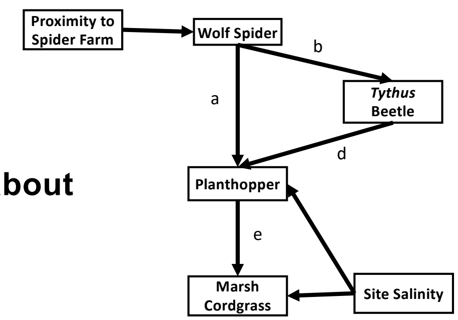


*What are a,b,c,d, and e in *THIS* marsh?*
(e.g., for biocontrol)

90

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?*

91

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?

92

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

93

The continuum of SEM

Structural Equation Modeling

exploratory/
model-building
applications

→

confirmatory/
hypothesis-testing/
model-comparison
applications

It all starts with an underlying model!

94

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

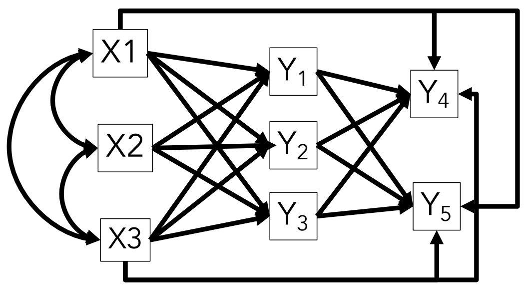
95

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

96

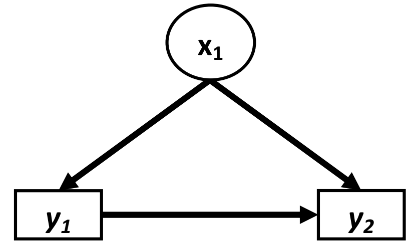
The First Elephant in the Room: The Kitchen Sink Model



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

97

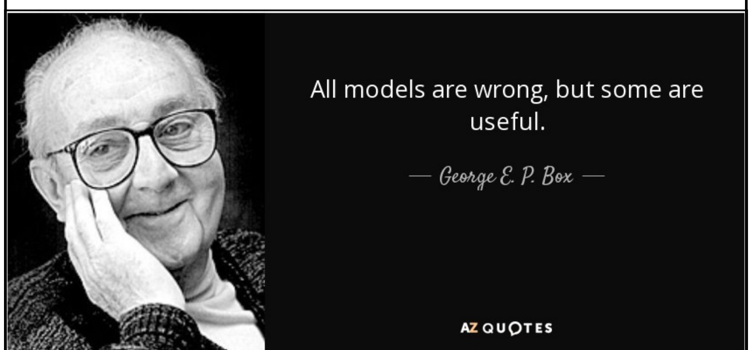
The Second Elephant in the Room: You Will Not Measure Everything



- 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***

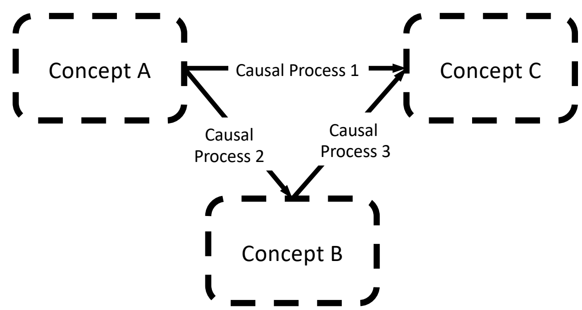
98

The Final Elephant in the Room



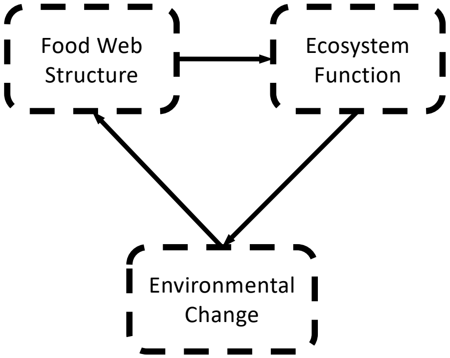
99

The Structural Equation Meta-Model (SEMM)



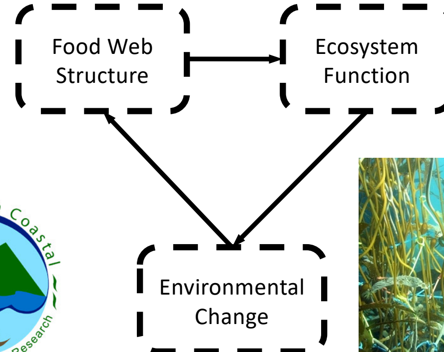
100

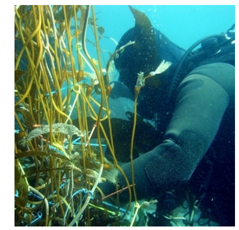
My Research Program as a Meta-Model



101

Targeting Your Question

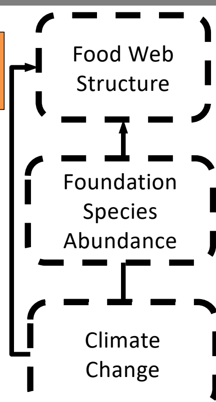




102

Targeting Your Question

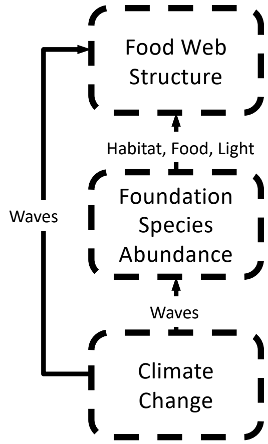
Focused on driver & Mediator



103

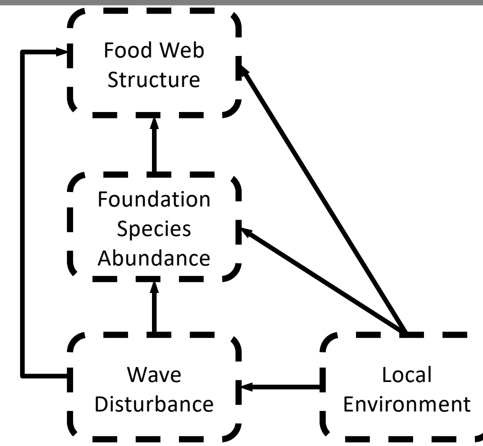
104

Identify Relevant Processes



105

Do you need to shut a conceptual backdoor?



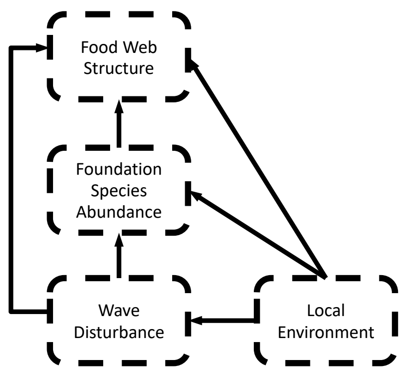
106

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



107

Meta-Model Your Research

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

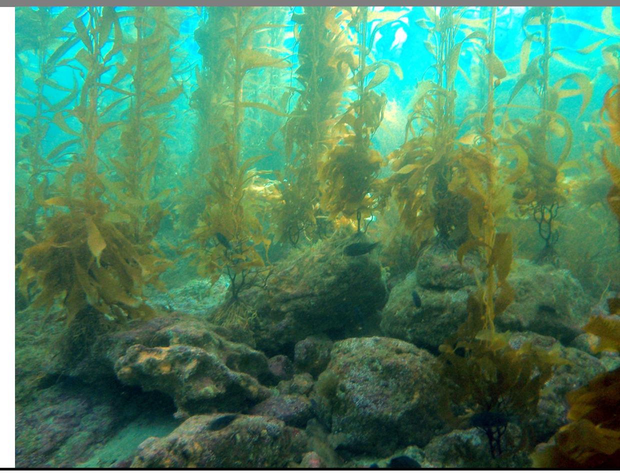
108

Overview

1. What is Causality?
2. Causality and Counterfactual Thinking
3. Simpson's Paradox and Causal Thinking
4. Causal Identification Methods
5. Choosing how to design a model
6. Starting with Meta-Models
7. Realizing Your Model

109

Complex Systems are Complex

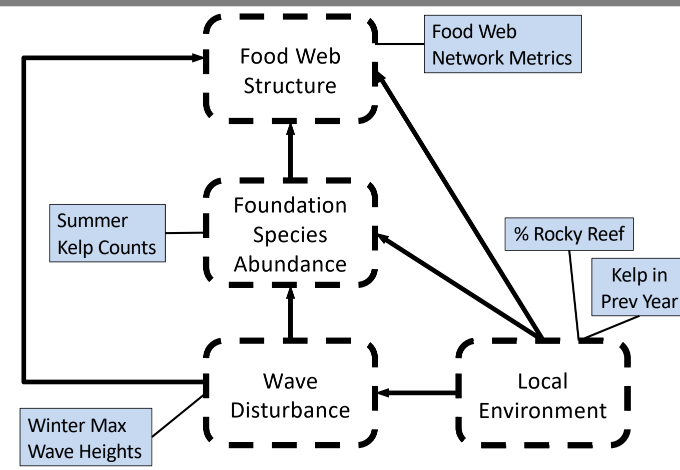


110

Sampling of Rocky Reefs 2000-2009

111

Matching Data to Concepts



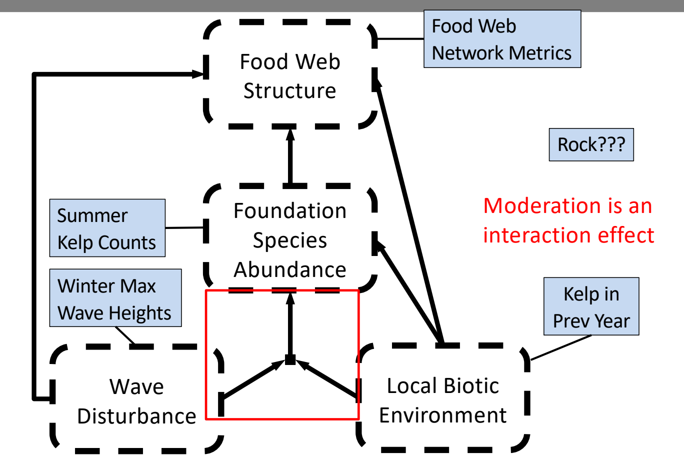
112

Adding Biological Realism

Problem 1: Kelp moderates disturbance
- More Kelp = Smaller Disturbance?
- BUT no effect on kelp that isn't present…

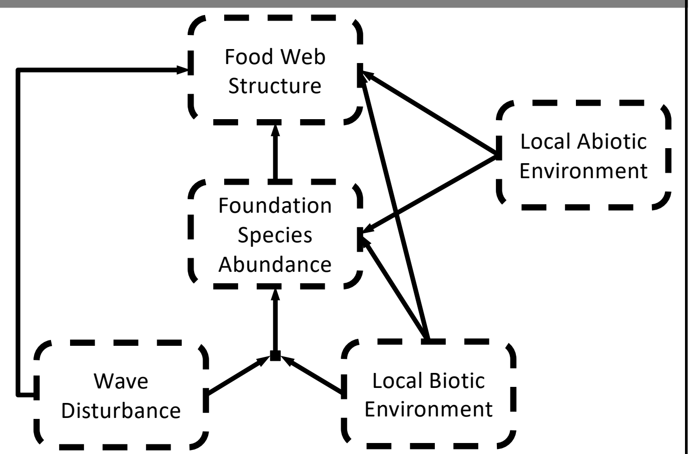
113

Solution 1: Moderation



114

But: Maintain Backdoor Blockage!



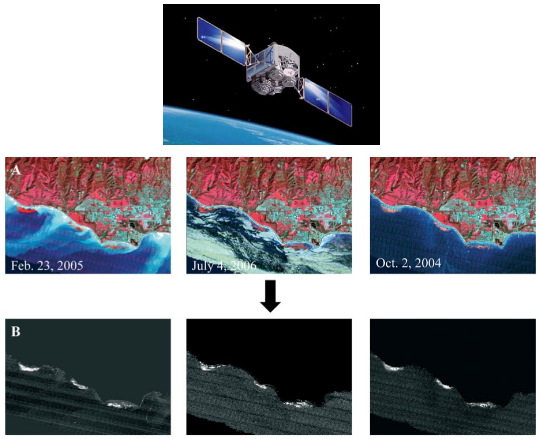
115

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!

116

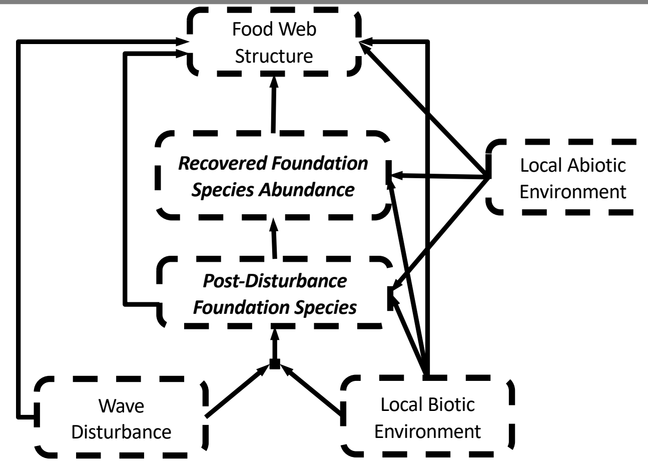
Measuring Realized Disturbance via Satellite Measurements



Cavanaugh et al 2011 MEPS

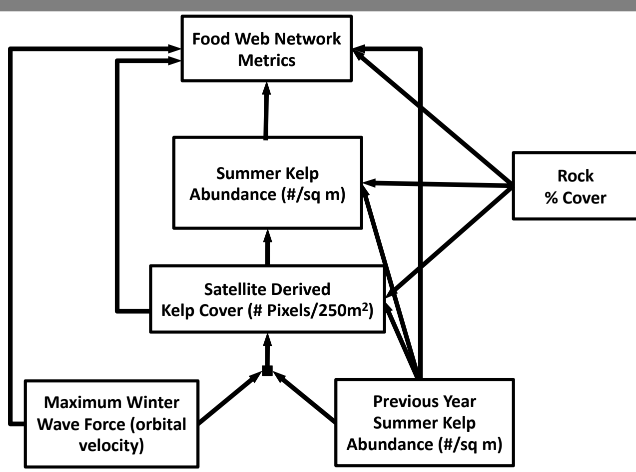
117

Incorporate Natural History of Disturbance



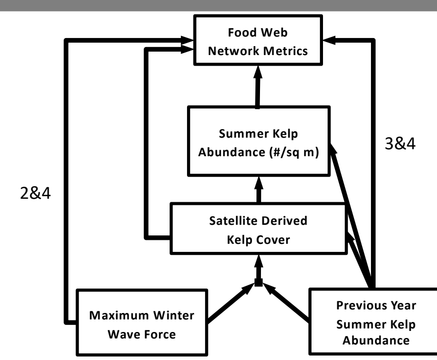
118

Model with Observed Variables



119

Goal: Hypothesis Evaluation



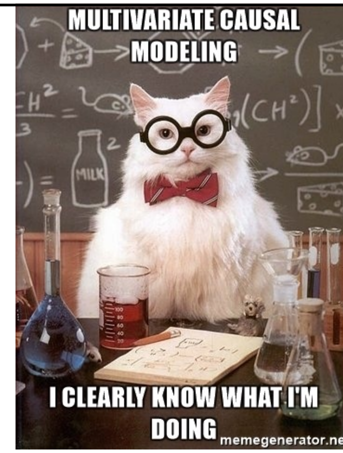
1. Full Model
2. No waves -> Food Web effect
3. No effect of previous year's kelp -> Food Web
4. No effect of either waves **or** previous kelp -> Food Web

120

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

121



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

122