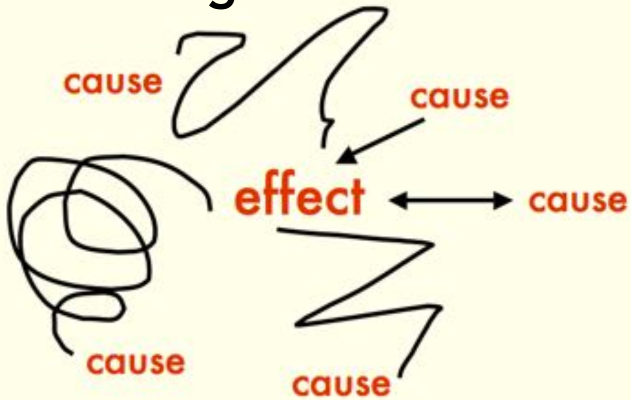
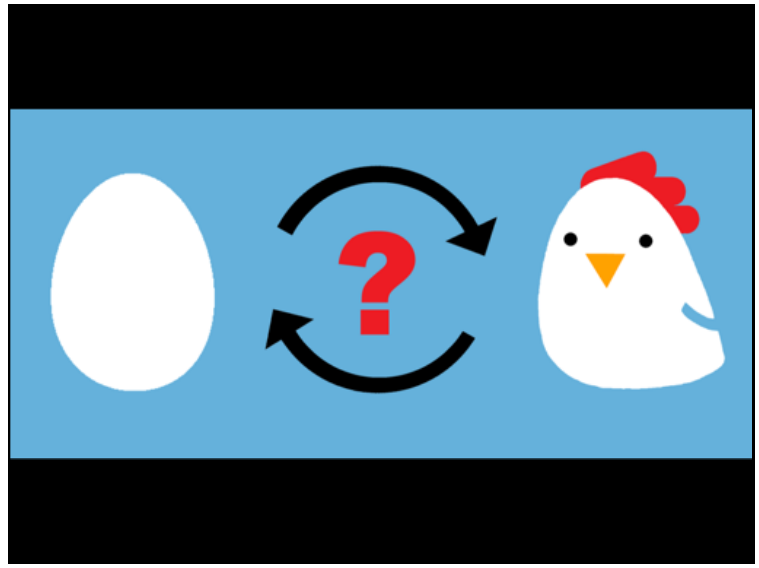
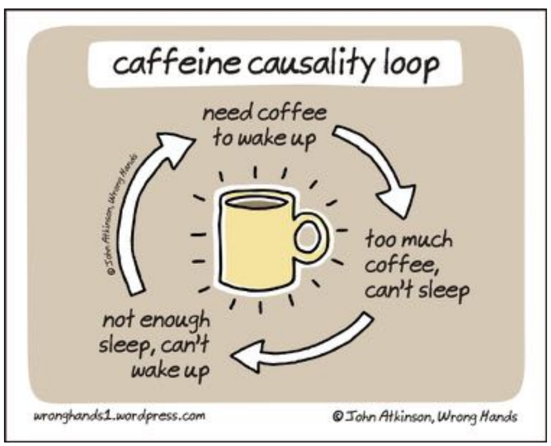
# Building Causal Models



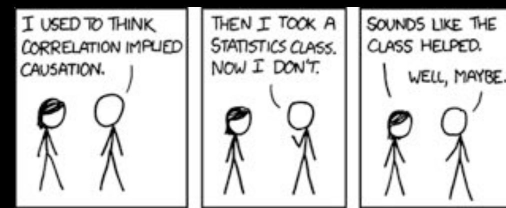




## Overview

1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

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

-Jim Grace




xkcd

## Pearl's Ladder of Causality

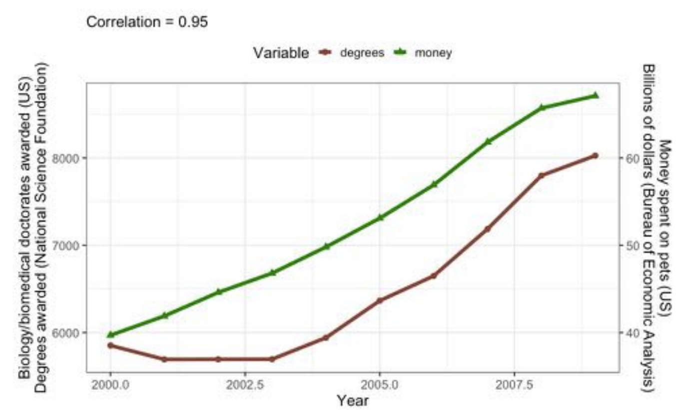

**3. Counterfactual – Can imagine what would happen under unobserved conditions**
- Requires model of a system
- Requires identification of causality

**2. Intervention – Understand what happens you do something**
- Experiments
- Provides evidence of causal link

**1. Observation – Cause is associated with effect**
- Correlation
- Can only predict within the range of data

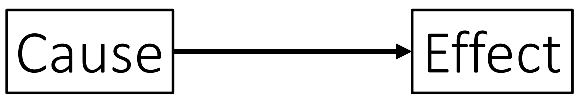
Pearl and Mackenzie 2018

## 1. Observation: What can we learn?

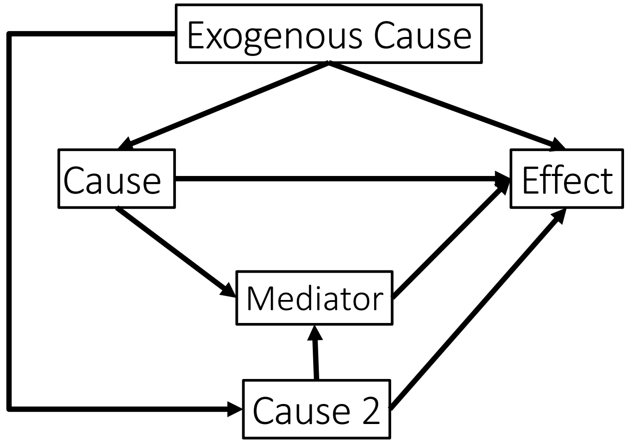



http://tylervigen.com/

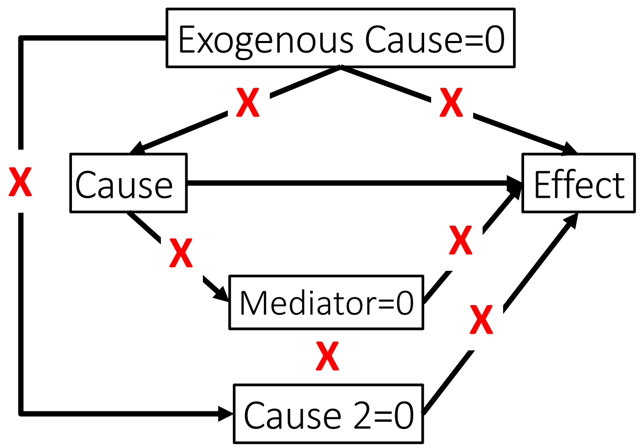
## What We Want to Evaluate



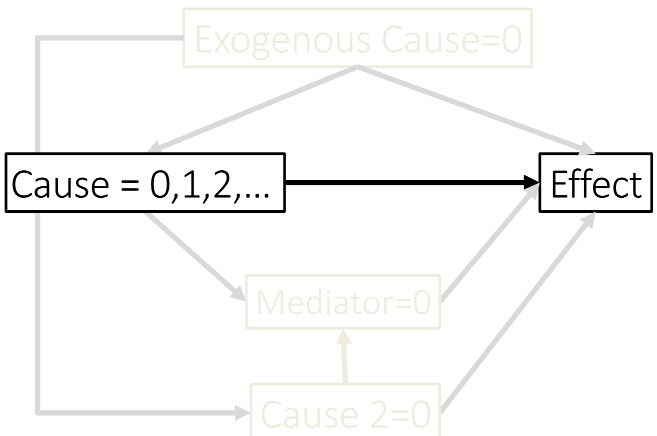

## The World




## Intervention: Experiments! What can we learn?

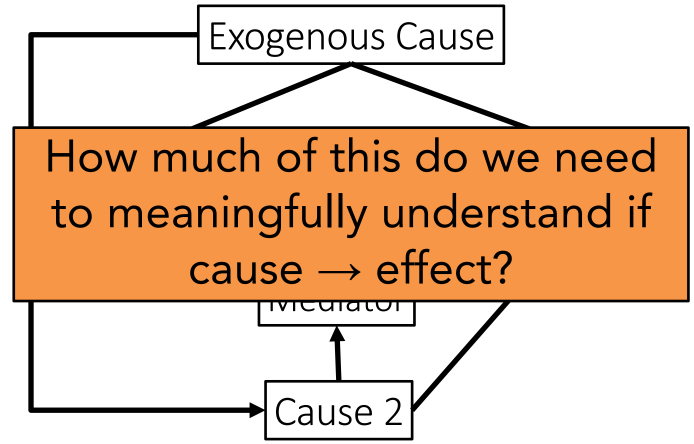



## Experiments: Manipulate Cause of Interest



## Reality Check: Lots of Things Happen in an Experiment

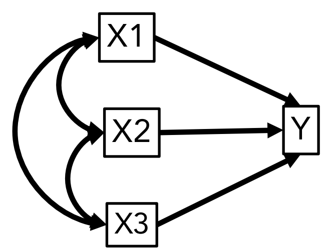
## Build The World to Make Counterfactual Predictions




## Overview

1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

## Why move beyond multiple regression to causal models?



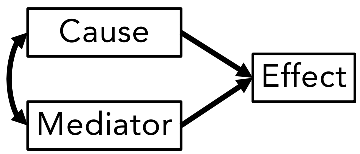

- We estimate the effect of exogenous variables ***controlling*** for all others
- Covariances implied
- Not controlling for the right variables = bad inference
- Controlling for the wrong variables = bad inference

## The Chain

*Causal Model*

***Be wary of this with post-treatment effects!***


*What it can do to Multiple Regression*

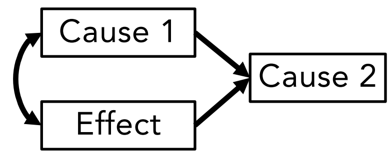



- Mediator blocks cause
- Looks like no or opposite link between cause and effect

## The Collider

*Causal Model*

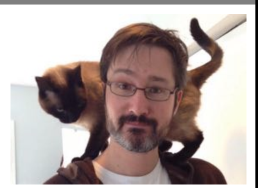

*What it would do to Multiple Regression*

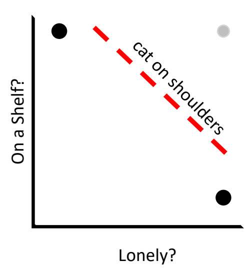



- Conditions on effect, opening path between causes
- Creates spurious correlations

## Example of Collider Bias

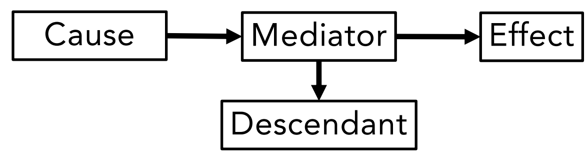
- My cat jumps on my shoulders when she is lonely, on a shelf next to me, or both.
- It's more likely she is satisfying one of those conditions at any one time, rather than both
- Thus, if she jumped from a shelf to my shoulders, I know she is likely not lonely
- I'd falsely conclude there's a relationship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



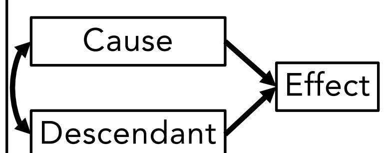



## The Descendant

*Causal Model*



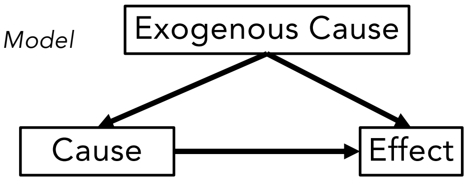

*What it would do to Multiple Regression*



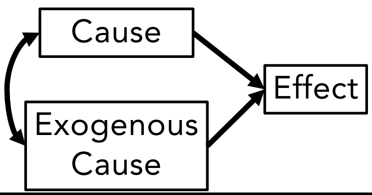

- Descendant still partially blocks cause
- Masks effect of cause – many possibilities
- Same with descendent of collider

# The Fork

*Causal Model*



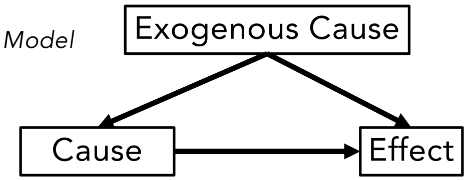

*What it would do to Multiple Regression*



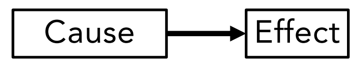

- ***Nothing – this is great!***

# The Bigger Problem…

*Causal Model*

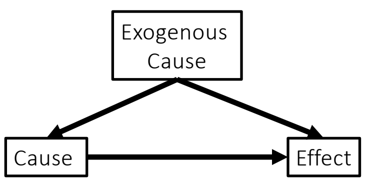



*Univariate Regression*




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

# Omitted Variable Bias

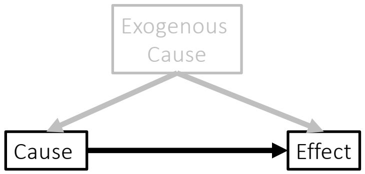



- We assume that sampling means that omitted variables average to 0
  - Omission produces downward bias in SE of coefficients
- But, if omitted variables are correlated causally with a predictor, they likely are not averaged out
- This ***will*** bias your estimates
  - You will not know in what direction

## Overview

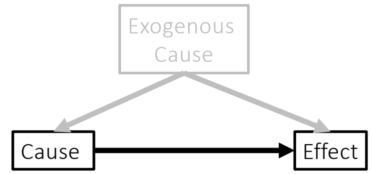
1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

## Bigger Problem




*This path is not causally identified*

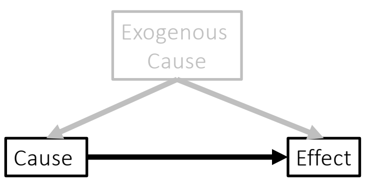
## Causal Identification




All of your model need not be causally identified - some may describe associations

But, you can only make counterfactual statements about parts that are

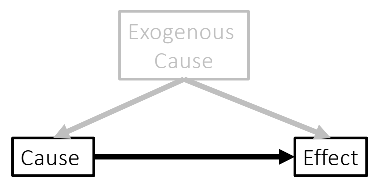
## Causal Identification




Causal identification does not require knowing ULTIMATE cause

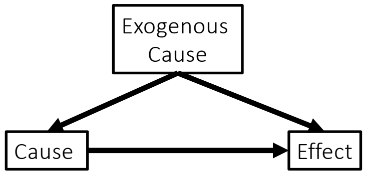
Nor does it require knowing exact mechanisms within a causal pathway

## How do we solve this problem?



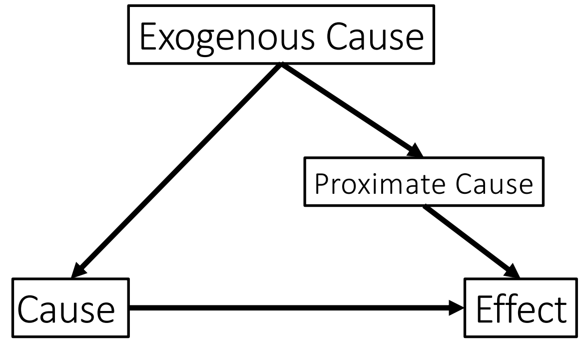

*This path is not causally identified*

## Solution 1: The Backdoor Criteria



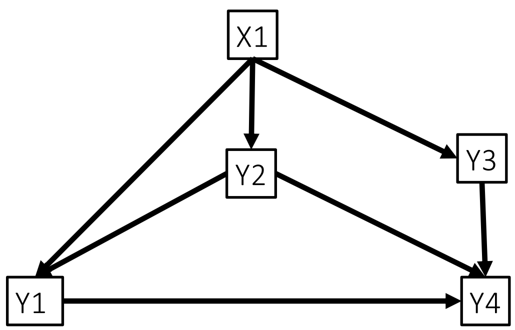

- If we want to know the link between cause and effect given variables that affect both, we must include all variables with a path ***into*** the cause
- So, variables must block all backdoor paths from treatment to outcome
- AND variables must not be descendants of the cause (i.e., no mediators – see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

## Proximate Backdoors



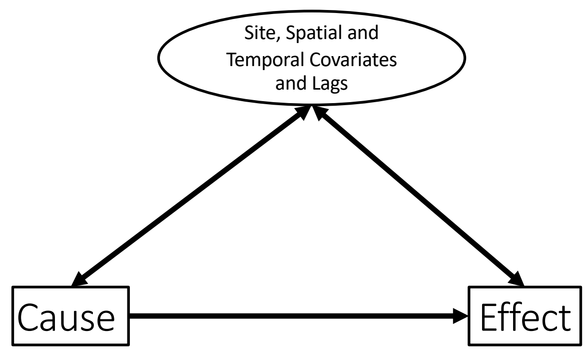

Often we only have proximate variables in a backdoor path. Controlling for just them is sufficient.

## 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




**We will talk more of this with respect to random effects, and autocorrelation on day 4!**

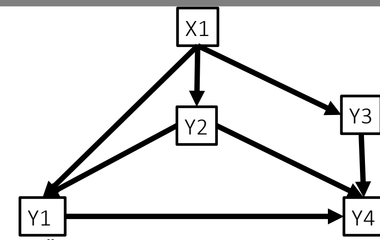
## Finding Backdoors with dagitty

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

*To build a DAG*

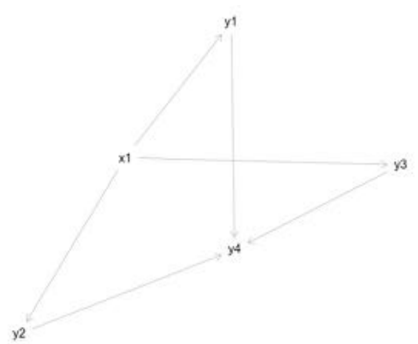
```
g <- dagitty("dag{
     ...
            }")
```

## Building a DAG with dagitty



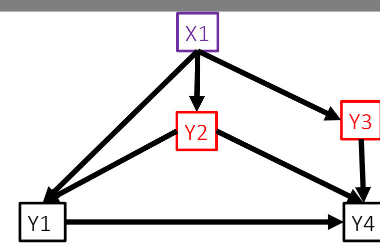

```
g <- dagitty("dag{
  x1 -> y1
  x1 -> y2
  x1 -> y3
  y3 -> y4
  y2 -> y4
  y1 -> y4
            }")
```

## Plot your DAG!




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

## How to Shut the Back Door



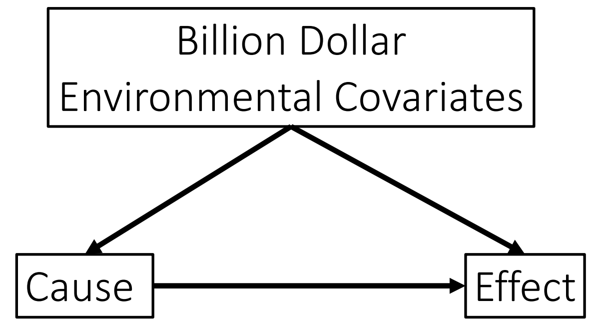

```
> adjustmentSets(g, exposure = "y1",
                 outcome = "y4")

{ y2, y3 }
{ x1 }
```

## Exercise: daggity

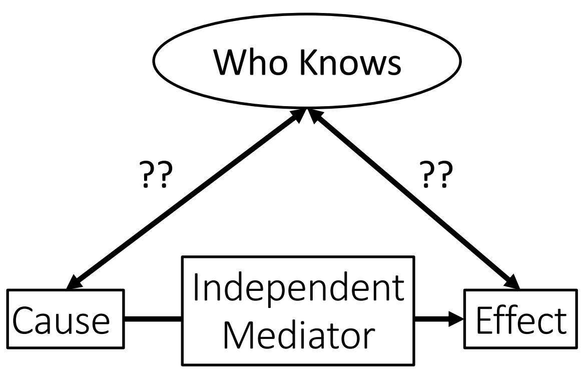
- Sketch a model of 4-5 variables in your system
  - Don’t think to hard (that’s for later!)
- See if you can figure out how to close any backdoors
- Use `daggity` to find the back doors between a chosen pair

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

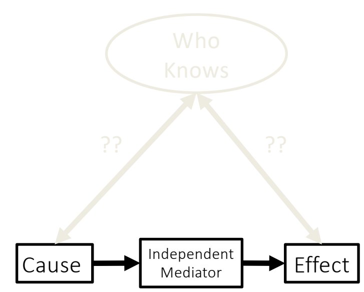
## Sometimes We Cannot Shut the Backdoor




## Or, we suspect, but don’t know, of backdoors

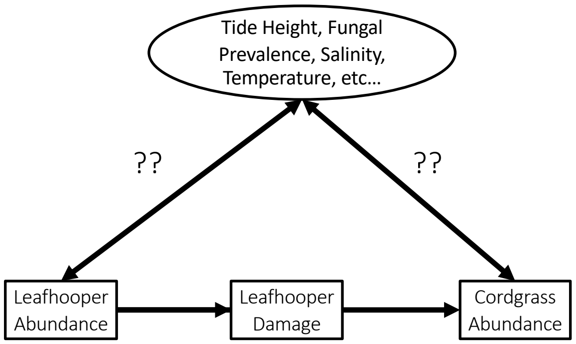



## Solution 2: The Front-Door Criterion

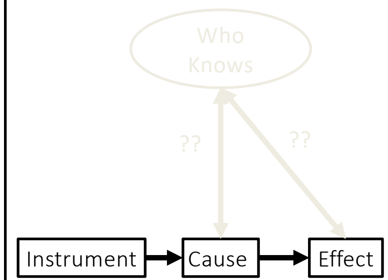



- A variable satisfies the front-door criteria when it blocks all paths from X to Y.
- In practice, you need a causally identified mediating variable unaffected by anything else.
- Thus, the influence of the cause is felt by the effect solely through its mediator.

## Example: Herbivory Through the Front Door

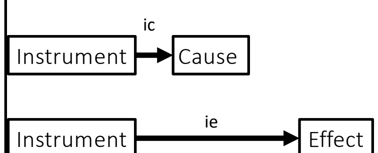



## Front Doors and Instruments




- Instrumental variables are those that <u>*****only*****</u> 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

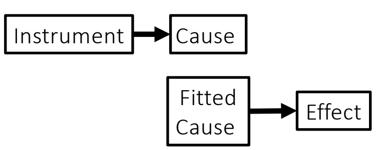
## Two Approaches in Instruments




1. Fit two models, leveraging the fact that

IE = IC * CE

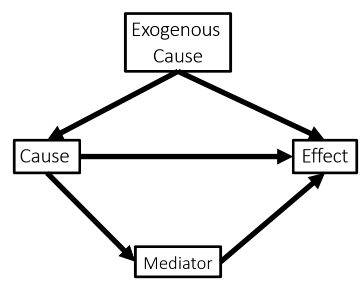
so, CE = IE/IC




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

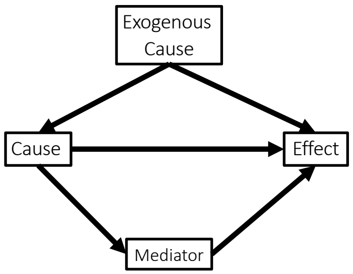
See `ivreg`

## 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

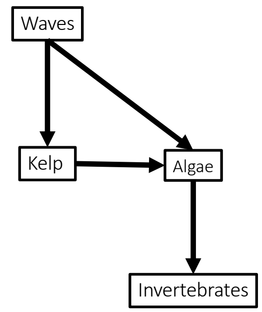



- We assume that two variables not connected are independent, conditioned on their parent influences

Mediator⊥Exogenous Cause | Cause

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

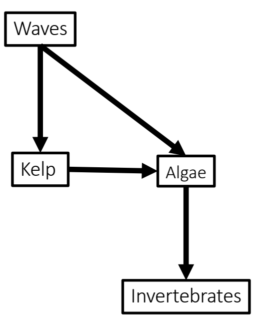
## Check yourself with `daggity`




```
forest_mod <-  dagitty("dag{
  waves -> kelp -> algae
  algae -> inverts
  waves -> algae
}")
```

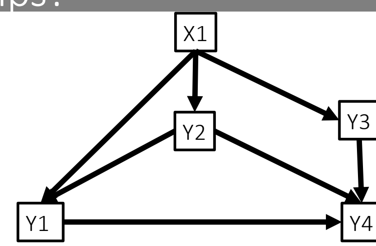
## Check yourself with `daggity`

```
> impliedConditionalIndependencies(forest_mod)
```




```
inverts _||_ kelp | algae
inverts _||_ waves | algae
```

## Exercise: What are your independence relationships?




```
> impliedConditionalIndependencies(g)
x1 _||_ y4 | y1, y2, y3
y1 _||_ y2 | x1
y1 _||_ y3 | x1
y2 _||_ y3 | x1
```

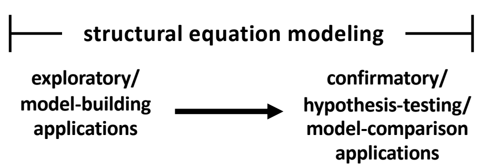
## Making Sure Pieces of your Model are Causal

- 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

## Overview

1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

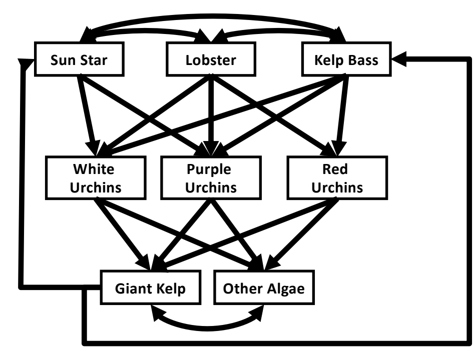
## The Continuum of SEM




***Your goals will inform how you build your model***

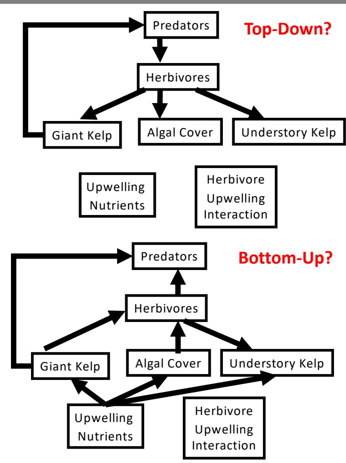
## What are the purpose of your modeling?

- Discovery?
- Hypothesis testing?
- Making predictions?



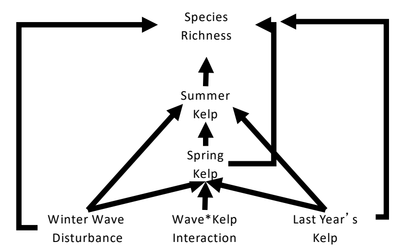
## What are the purpose of your modeling?

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




## What are the purpose of your modeling?

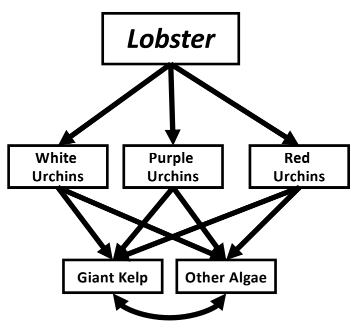
- Discovery?
- Hypothesis testing?
- **Making predictions?**




| | High Kelp | Low kelp |
|---|---|---|
| **High waves** | ? | ? |
| **Low Waves** | ? | ? |

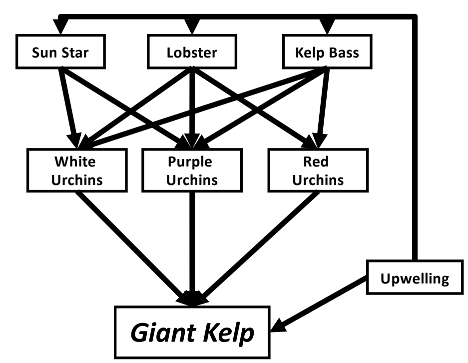
## What is the focus of your modeling?

- Driver
- Response
- Mediation
- Theory Testing



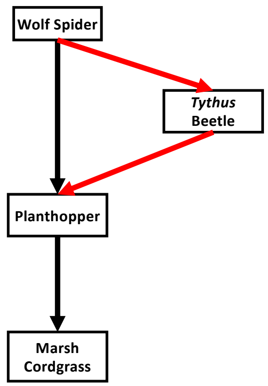

## What is the focus of your modeling?

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



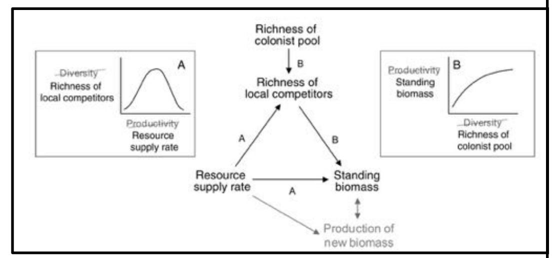
## What is the focus of your modeling?

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




## What is the focus of your modeling?

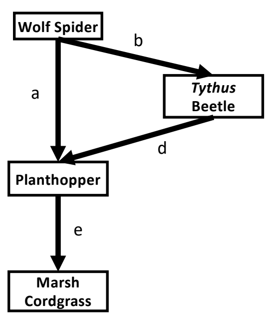
- Driver
- Response
- Mediation
- **Theory Testing**




Cardinale et al. 2008

## What is the span of your inference?

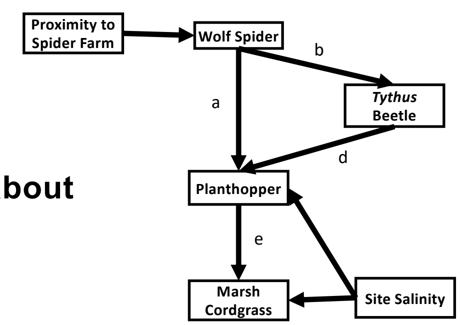
- Local estimation
- Learning about processes




*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?

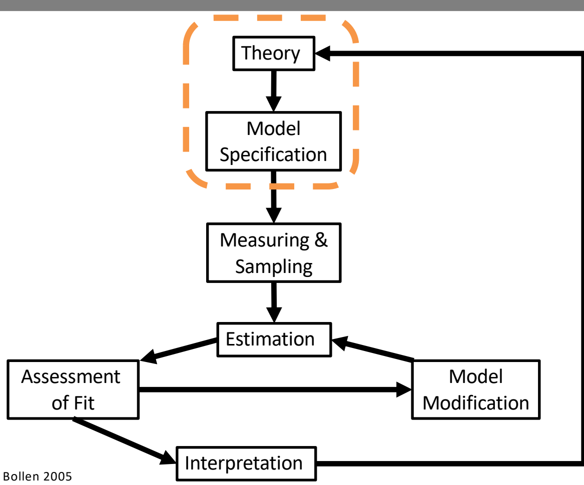
Focus of modeling effort:
- driver focused?
- response focused?
- mediation focused?
- theory testing focused?

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

## Overview

1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
3. Causal Identification
4. Choosing how to design a model
5. Starting with Meta-Models
6. Realizing Your Model

## SEM as a Unifying Process

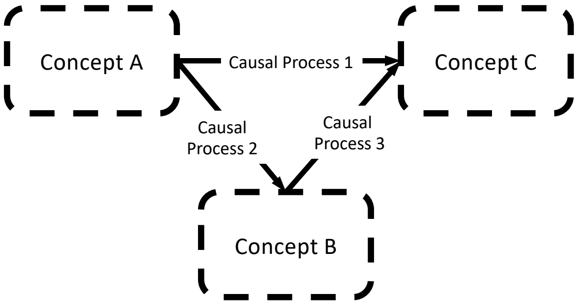



After Grace & Bollen 2005

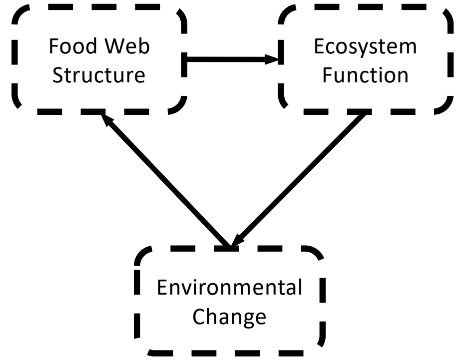
## Model Building

1. What is Causality?
2. Causal Building Blocks
3. Research Goals and Model Structure
4. Starting with Meta-Models
5. Realizing Your Model

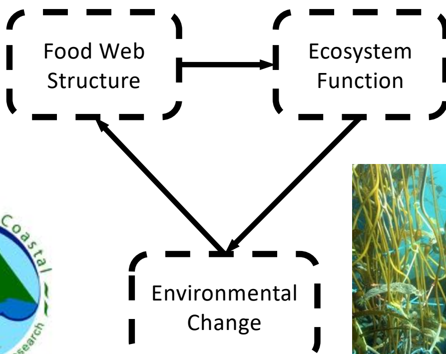
## The Structural Equation Meta-Model (SEMM)




## My Research Program as a Meta-Model

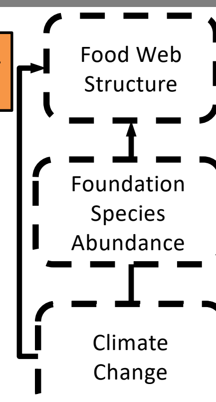



## Targeting Your Question

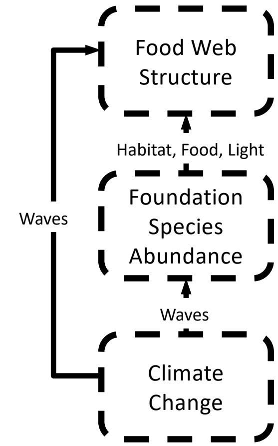



## Targeting Your Question

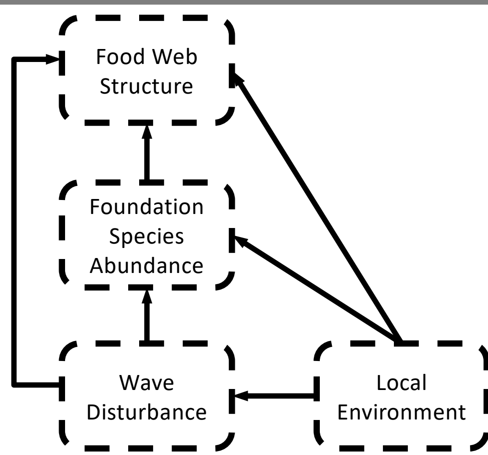
Focused on driver & Mediator




## Identify Relevant Processes




## Do you need to shut a conceptual backdoor?

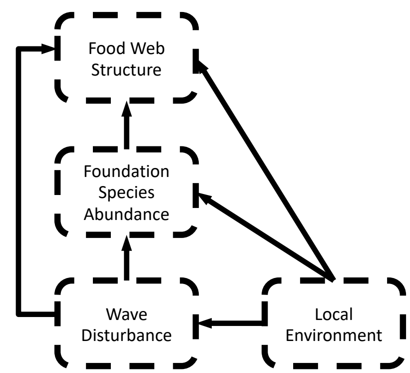



## What did I want to do?

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

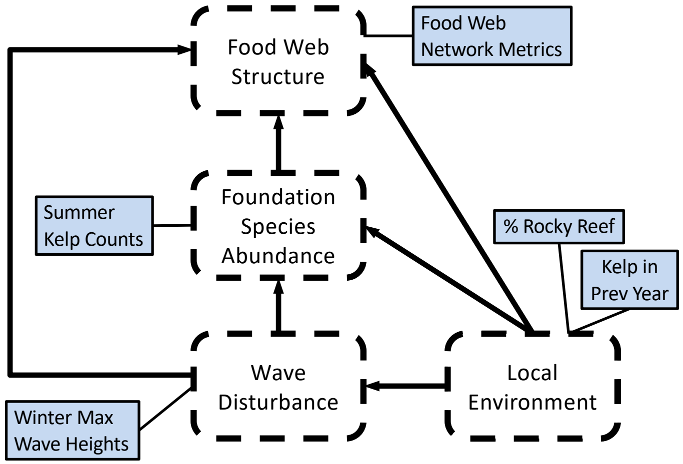
## Overview

1. What is Causality?
2. Benefits of thinking in causal models over multiple regression
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

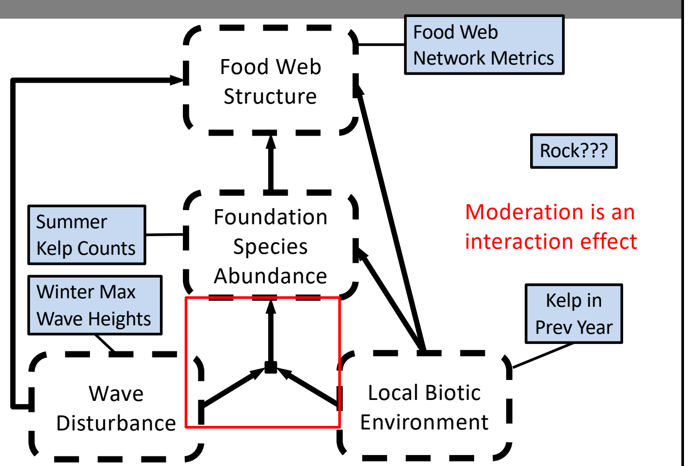
## Matching Data to Concepts



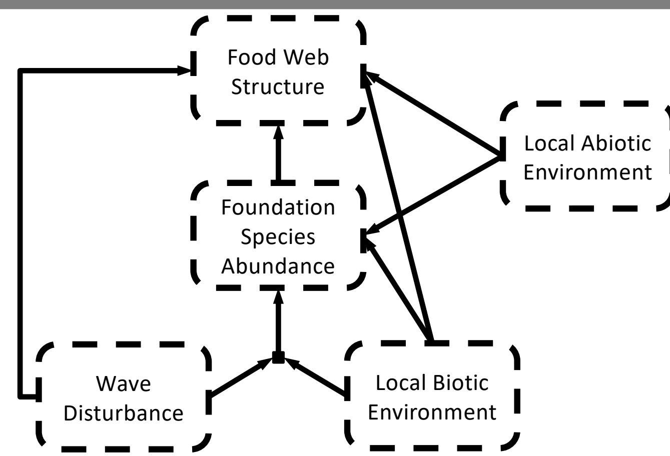

## Adding Biological Realism

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!

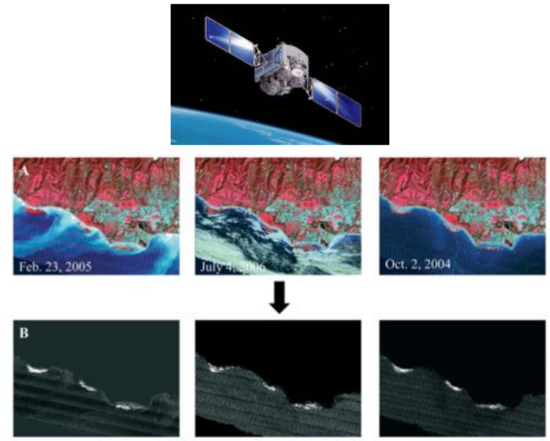


## Natural History Creates Problem

Problem 2: Kelp regrows quickly

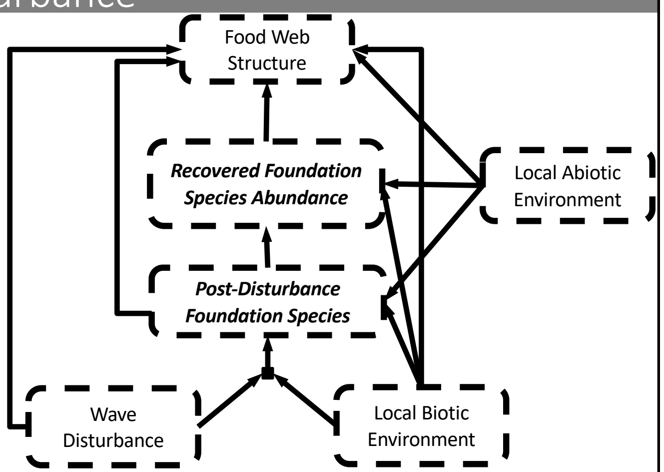
- It's a jungle by summer if nutrients are present
- Need to see if kelp was actually removed in winter!

## Measuring Realized Disturbance via Satellite Measurements

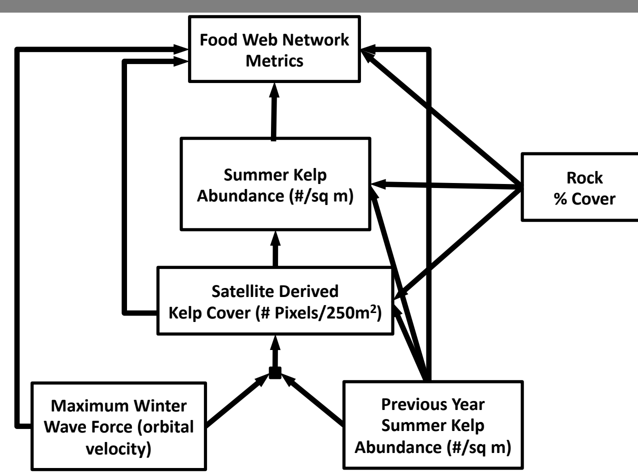



Cavanaugh et al 2011 MEPS

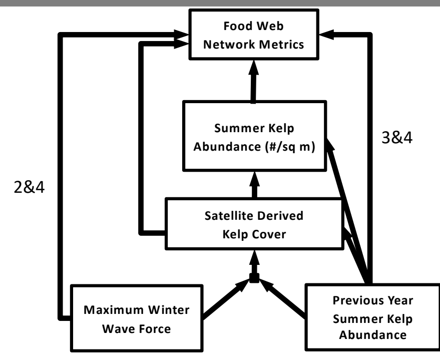
## Incorporate Natural History of Disturbance




## Model with Observed Variables



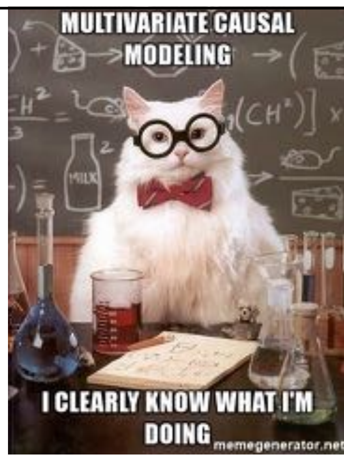
## Goal: Hypothesis Evaluation




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!