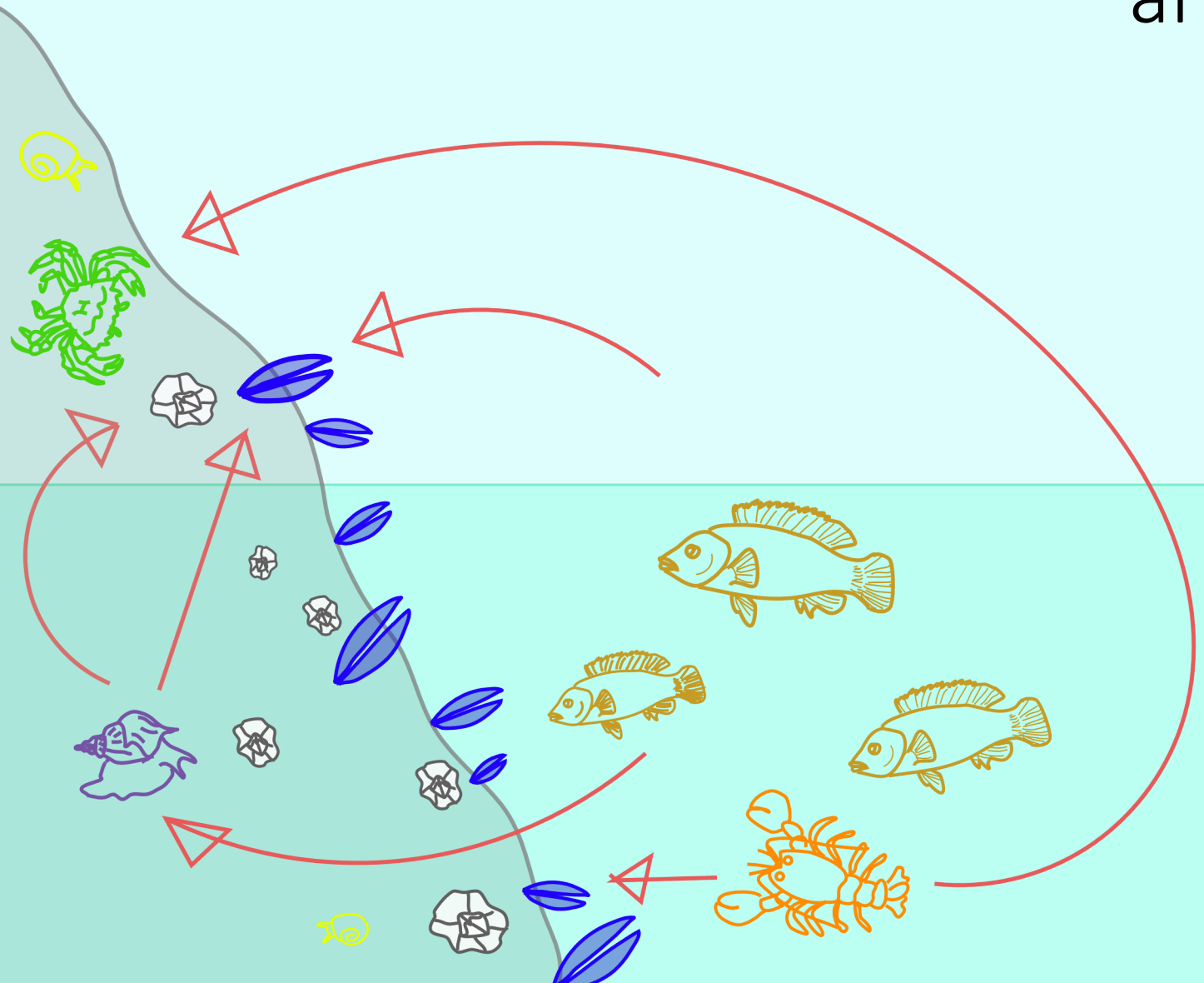
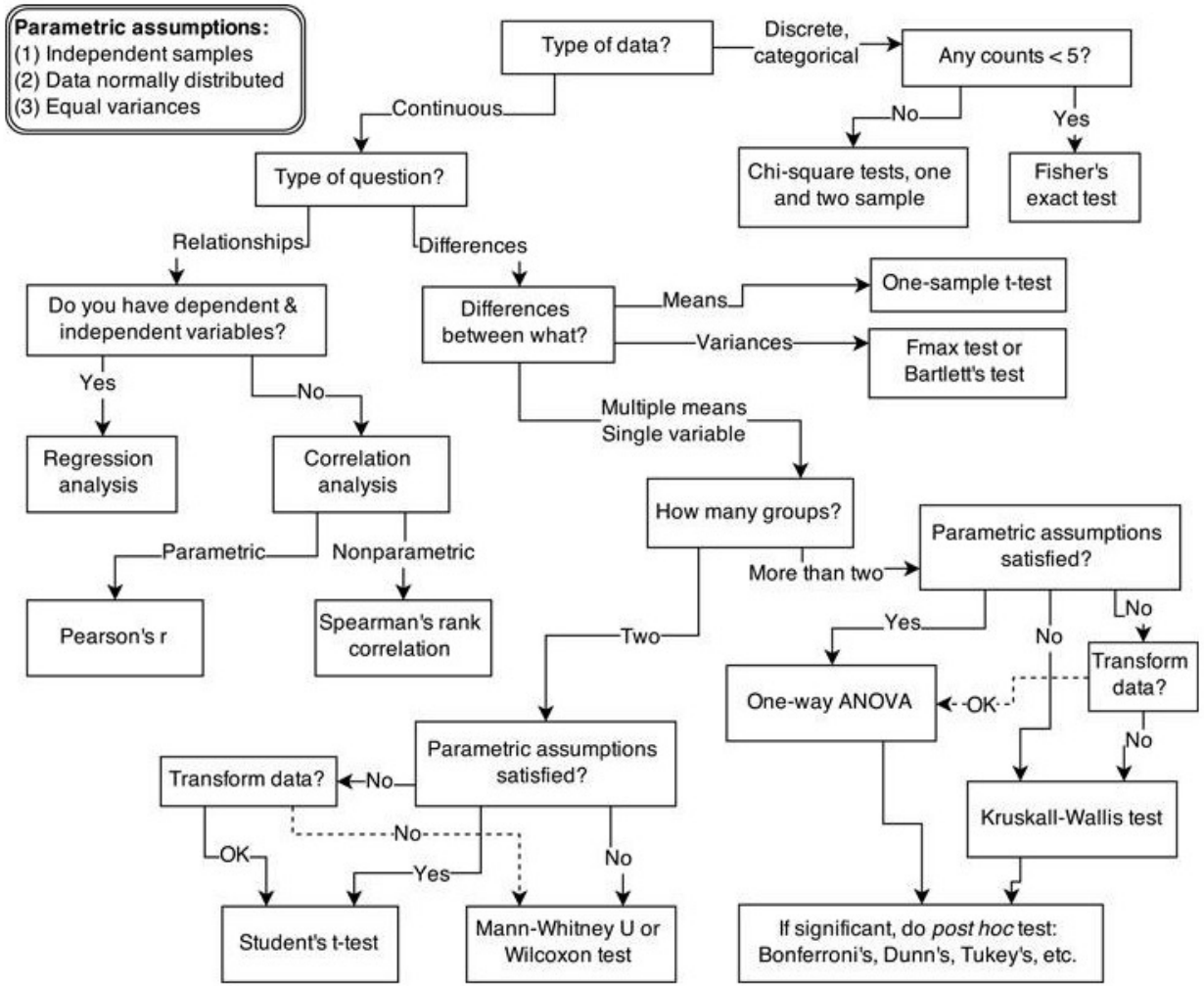
Structural equation modelling and ecological inference

Nicole Knight

QCBS R Symposium, 2022

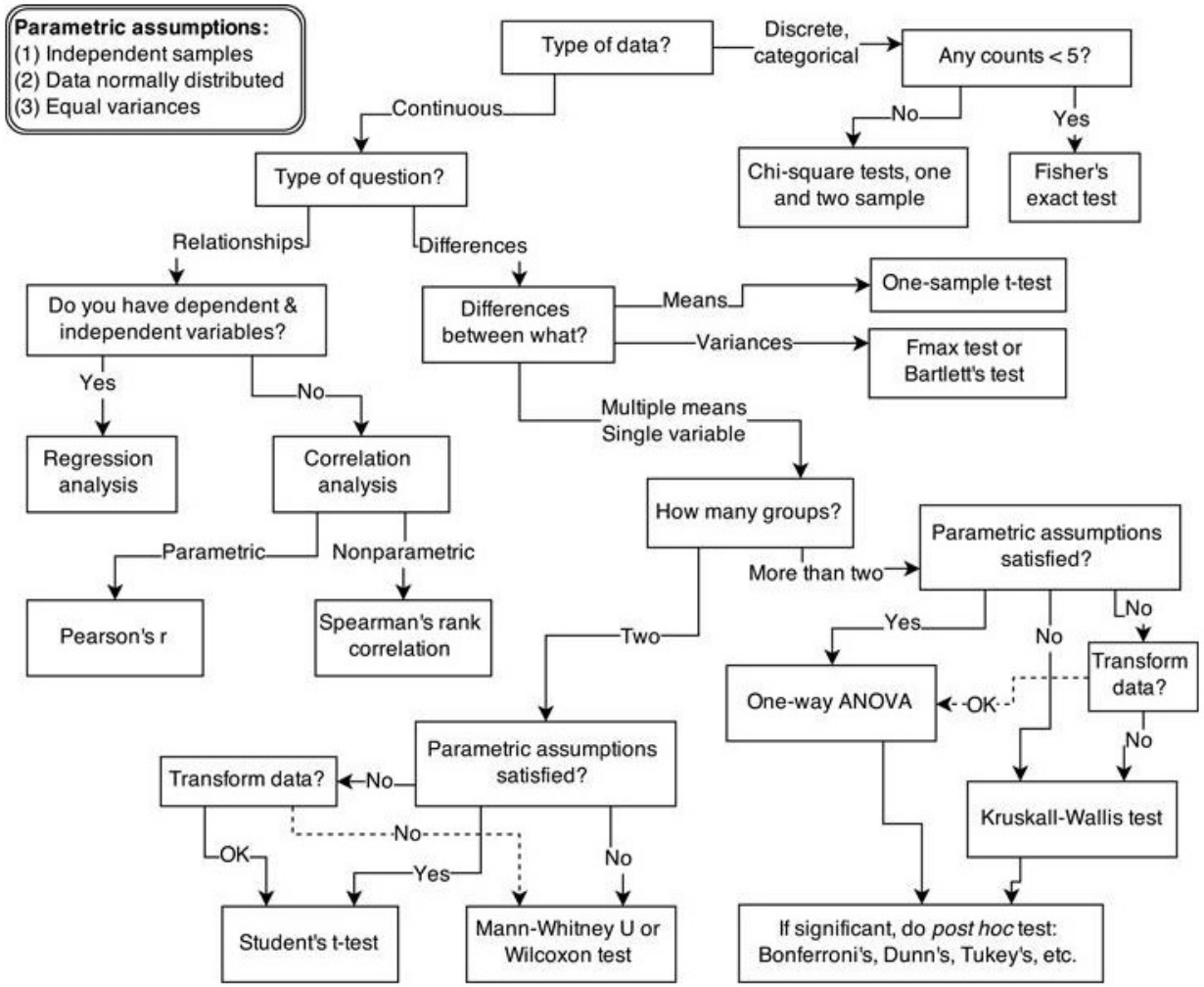


Undergraduate statistics training often looks something like this



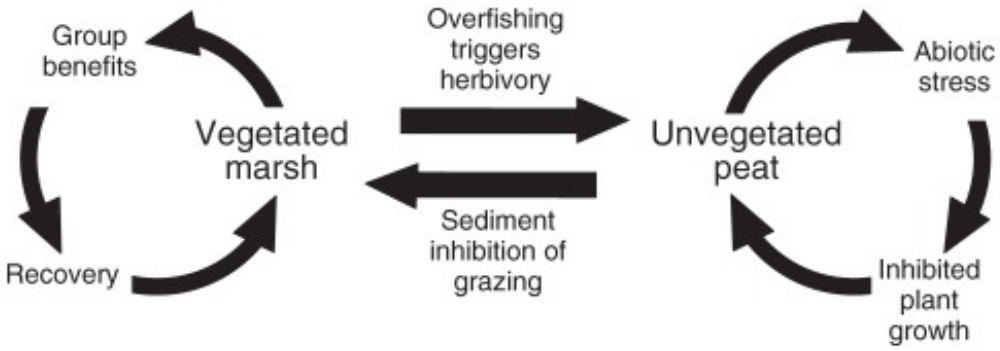
McElreath et al. 2020

Undergraduate statistics training often looks something like this



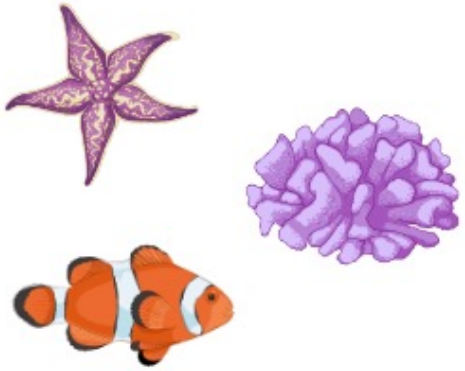
McElreath et al. 2020

But ecology is complicated-



Altieri et al. 2012

Ecosystems are characterized by direct and indirect effects



Not all ecological concepts are directly measurable (e.g., biodiversity, ecosystem function)


Ecologists need a pile of Legos, not a kit

Going from kits to loose Legos is the challenge

Going from kits to loose Legos is the challenge

This is where structural equation models come in

What is structural equation modelling?

Structural equation modelling is a framework for understanding relationships between multiple response variables

Why use structural equation models?

- To model direct *and* indirect effects

Why use structural equation models?

- To model direct *and* indirect effects

- To model theoretical concepts that can’t be measured directly (latent variables)

Why use structural equation models?

- To model direct *and* indirect effects
- To model theoretical concepts that can’t be measured directly (latent variables)
- To do a better job of linking data to conceptual models

How do we build an SEM?

Steps to build a SEM

Step 1: Describe your system with causal diagrams

Step 2: Design your statistical model

Step 3: Code your model

Step 4: Check your model

Step 5: Interpret your model

Step 6: Present your model

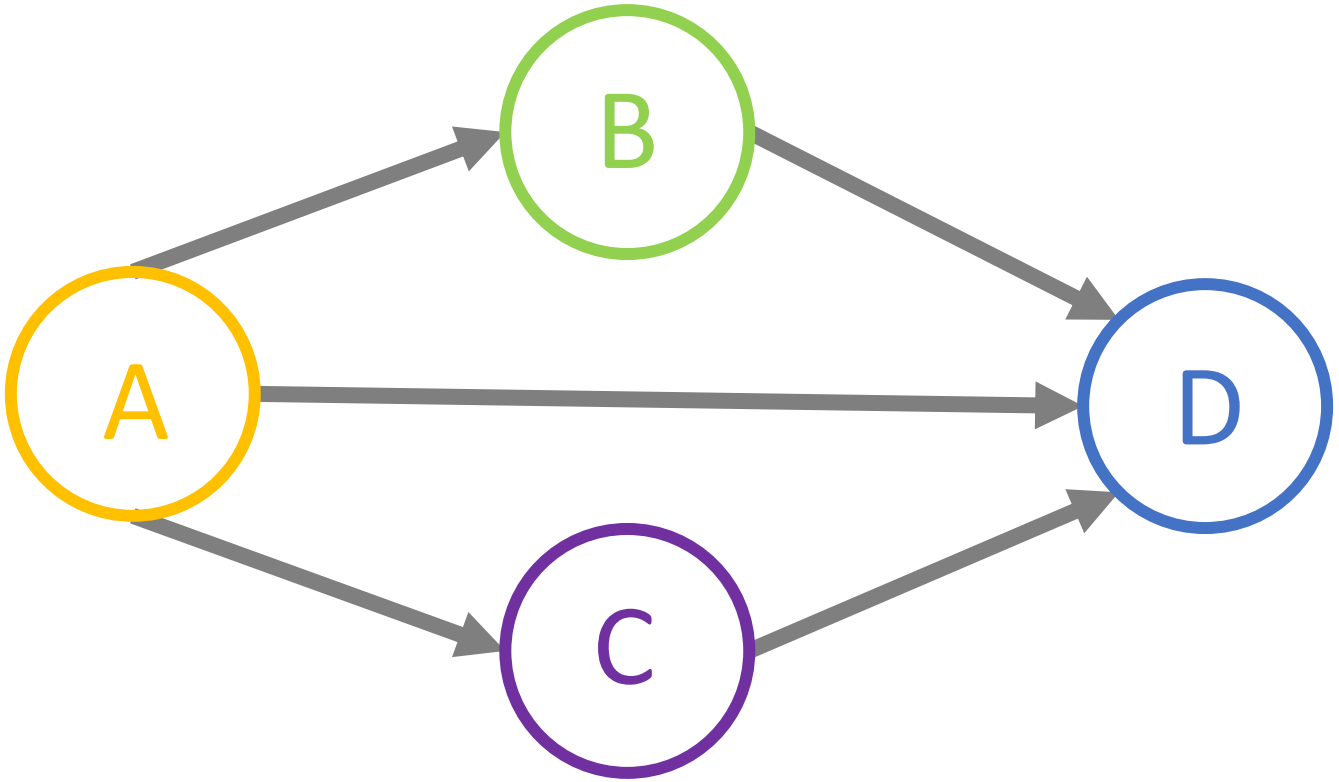
What is a causal diagram?

- A causal diagram is a conceptual model that describes cause-effect relationships in your system

- They can be used for SEMs, designing experiments or observational work, building other types of models, etc.

Step 1: Describe your system

Anatomy of a causal diagram

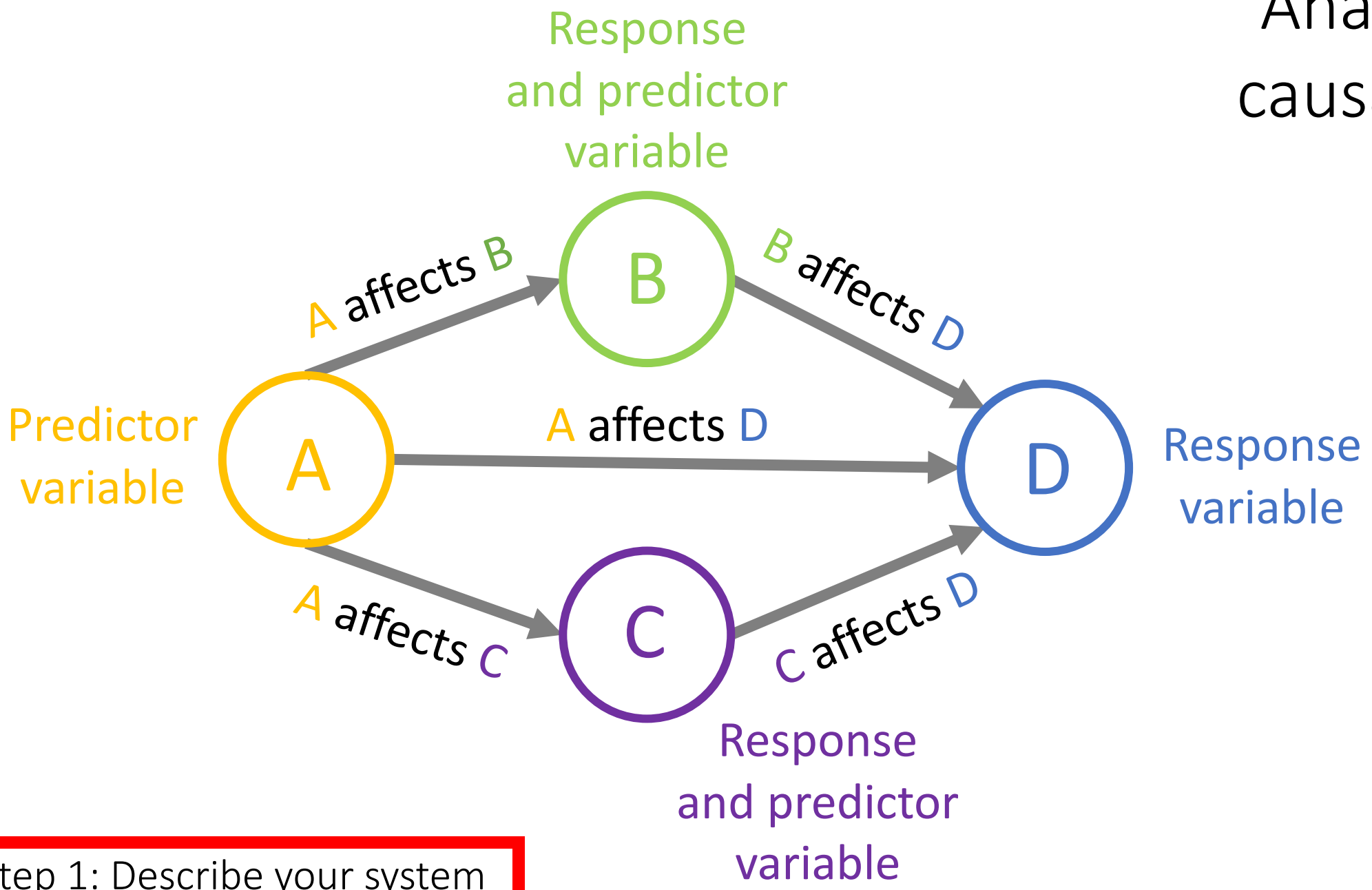


Path: effect or relationship

Variable

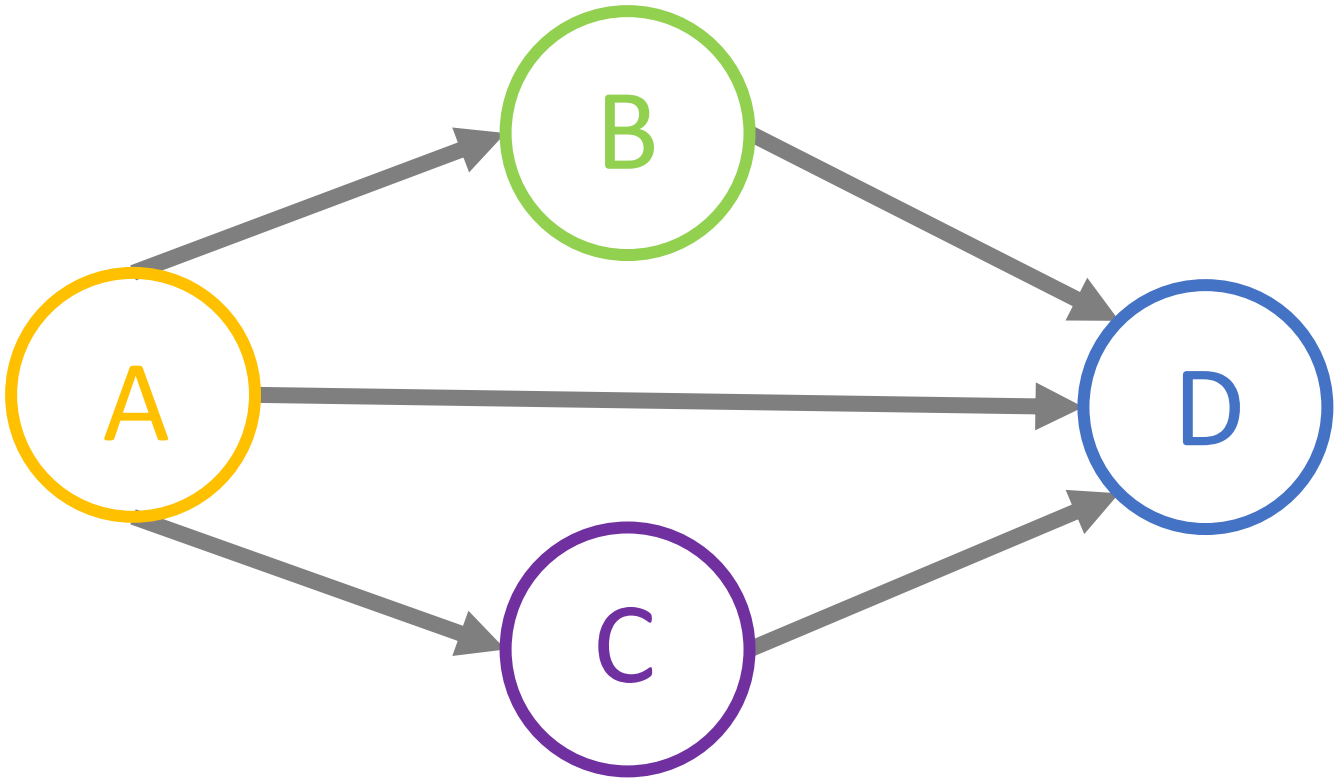
Step 1: Describe your system

Anatomy of a causal diagram



Step 1: Describe your system

Anatomy of a causal diagram



Model equivalent

B ~ A

C ~ A

D ~ B + C + A

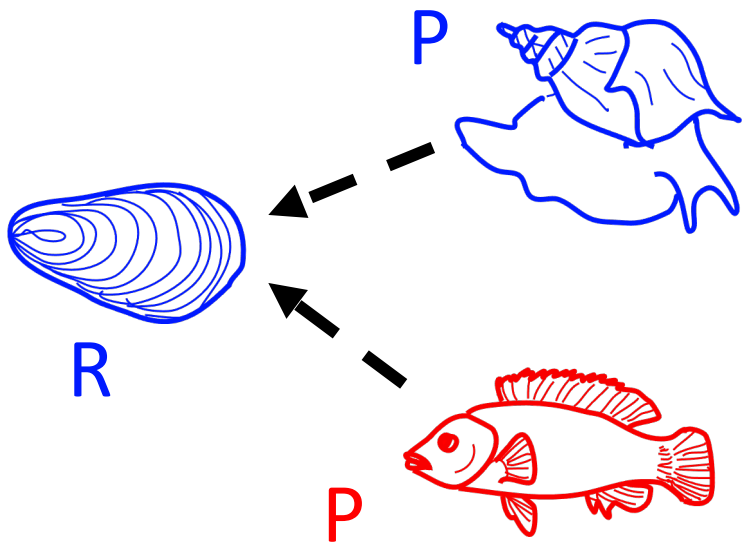
Step 1: Describe your system

What should go into my causal diagram?

- Variables and paths that describe your core research questions/hypotheses
- Potentially confounding variables (Step 4)
- Variables and paths that describe reasonable alternative hypotheses/mechanisms

Example: a simple food web

Traditional model



Estimated effects from this model will NOT be “wrong”, will just give net effect

positive effect

negative effect

R response

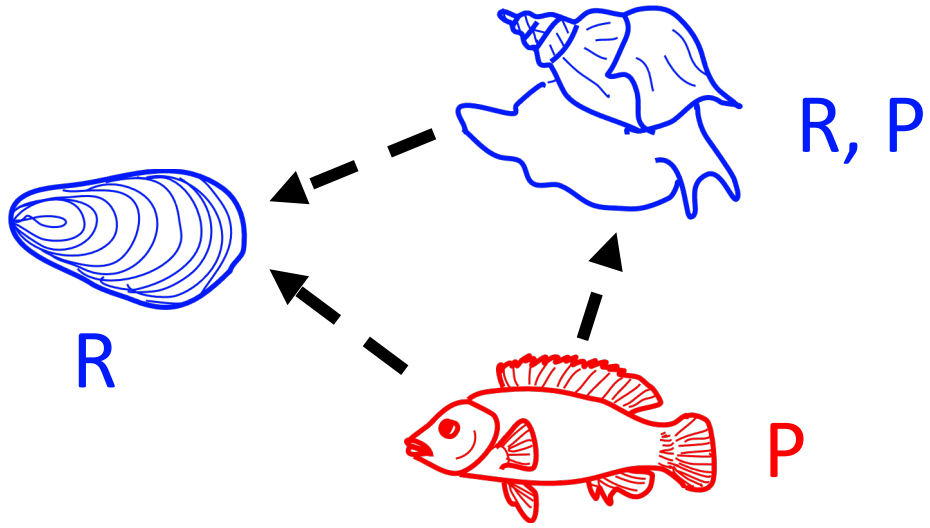
P predictor

Step 1: Describe your system

Example: a simple food web

Building a model like this will allow you to separate direct and indirect effects

Structural equation model



positive effect

negative effect

R response

P predictor

Step 1: Describe your system

Exercise 1:
Describe kelp ecosystems using *dagitty*

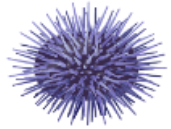
Understory algae

Epiphytes

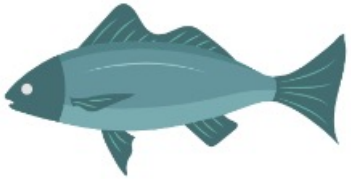
Sea otters

Small predators

Kelp



Urchin



Large predators

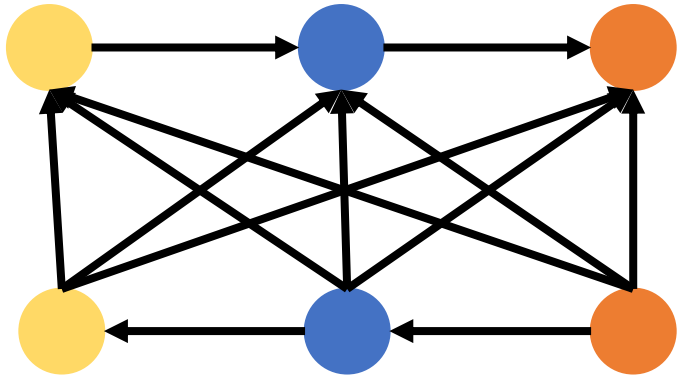
Small grazers

Juvenile fish

Images from BioRender

Common pitfalls to avoid

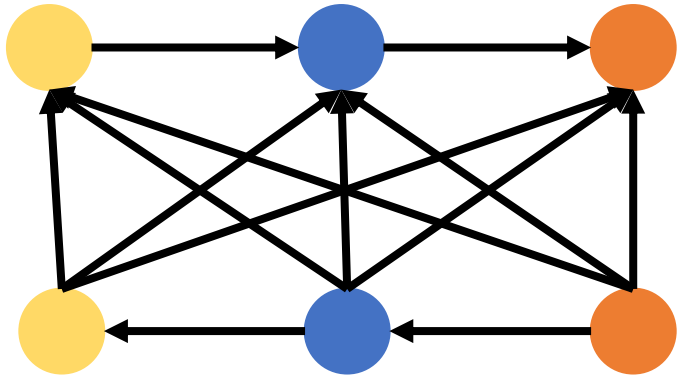
- Including paths between every single variable as a default
  - SEMs aren't meant to be a fishing expedition



Step 1: Describe your system

Common pitfalls to avoid

- Including paths between every single variable as a default*
  - SEMs aren't meant to be a fishing expedition
  - There is a data cost to every path you include (20 data points/path ideal, 10/path good, 5/path absolute minimum)

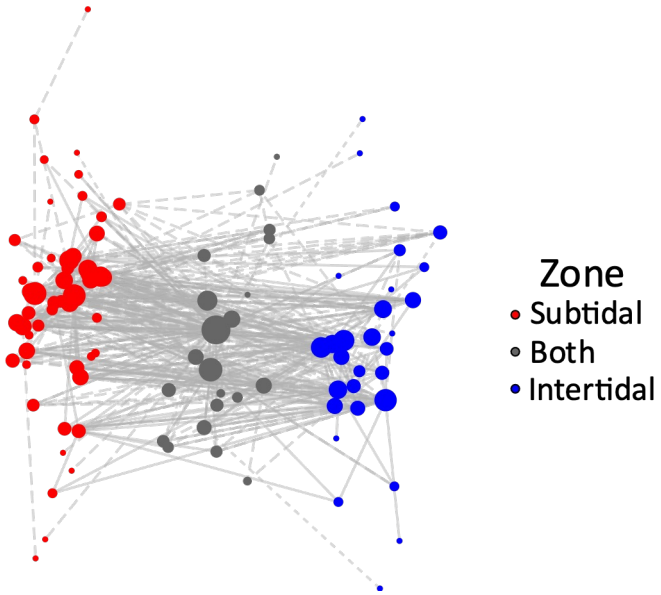


*You will eventually want to check that there aren't any obvious correlations in your data that the model doesn't account for

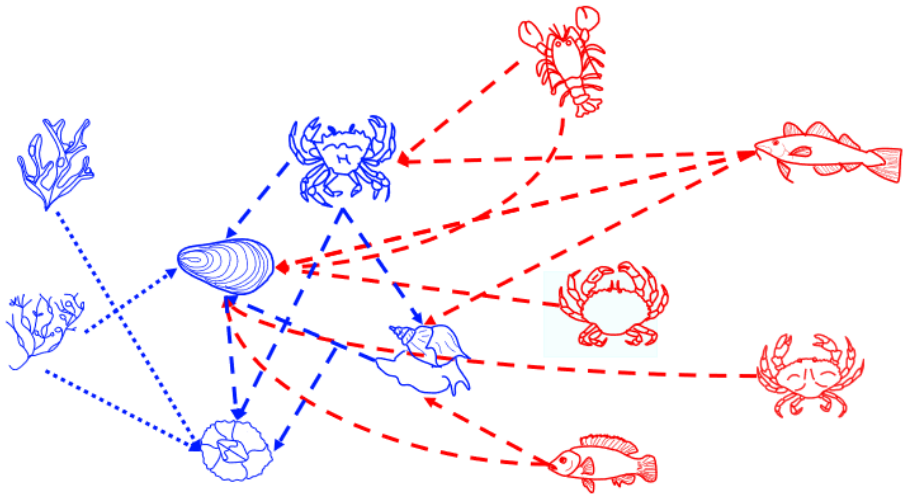
Step 1: Describe your system

Common pitfalls to avoid

- Including paths that are anticipated to be weak/unimportant
  - You can't model everything



Burant, Peller, Beaulieu, Knight et al. *in prep*

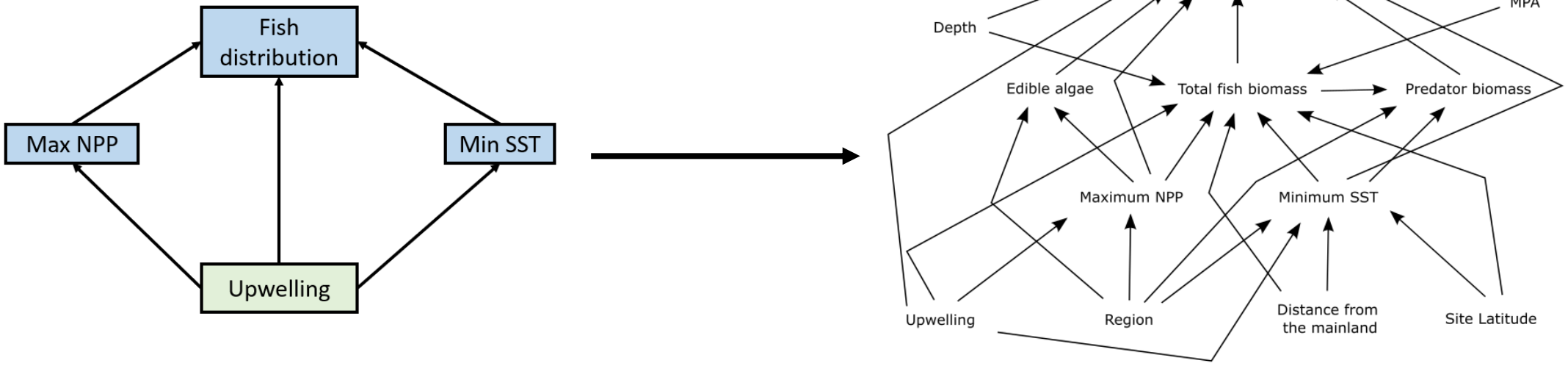


Knight et al. *in prep*

Step 1: Describe your system

Common pitfalls to avoid

- Trying to explain everything in your system
  - Focus on including paths that are relevant to your hypotheses

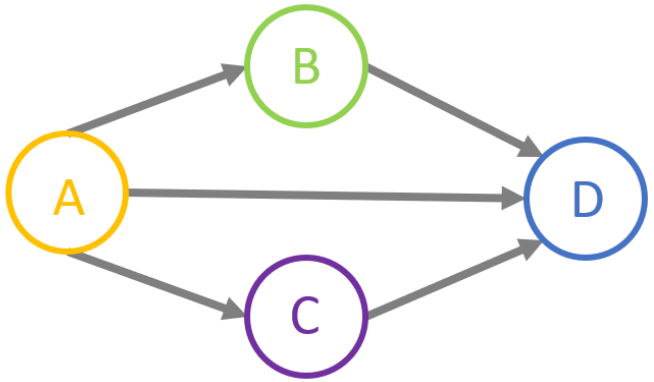


Step 1: Describe your system

A few last tips

- Document how and why you build your causal diagrams (preferably from **Day 1**)
- R Markdown is a great tool for creating this documentation
- Get external feedback on your causal diagram before moving forward

Design your statistical model



B ~ A

C ~ A

D ~ B + C + A

Step 2: Design your model

You know more than you think

The building blocks of SEMs are techniques you may already know

- Linear models
- Generalized linear models
- Generalized additive models
- Multilevel models
- Bayesian models

Some considerations

- Do I have data for all of my variables?

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)
- Are my data spatial or temporal?

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)
- Are my data spatial or temporal?
- Are my data hierarchical?

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)
- Are my data spatial or temporal?
- Are my data hierarchical?
- What probability distributions best describe my response variables?

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)
- Are my data spatial or temporal?
- Are my data hierarchical?
- What probability distributions best describe my response variables?
- Do I have any latent or composite variables in my system?

Some considerations

- Do I have data for all of my variables?
- Do I have enough data? (5/path minimum)
- Are my data spatial or temporal?
- Are my data hierarchical?
- What probability distributions best describe my response variables?
- Do I have any latent or composite variables in my system?

What type of SEM should I build?

All of my response variables are normally distributed and non-hierarchical *because I'm the luckiest person in the world*

Traditional variance-covariance SEM

Step 2: Design your model

What type of SEM should I build?

All of my response variables are normally distributed and non-hierarchical *because I'm the luckiest person in the world*

Traditional variance-covariance SEM

At least one of my response variables is not normally distributed or I have hierarchical data

Piecewise SEM

Step 2: Design your model

What type of SEM should I build?

All of my response variables are normally distributed and non-hierarchical *because I'm the luckiest person in the world*

Traditional variance-covariance SEM

At least one of my response variables is not normally distributed or I have hierarchical data

Piecewise SEM

Step 2: Design your model

What R package should I use?

lavaan

- Frequentist package specific to SEMs
- **Can** accommodate latent, composite variables
- Designed for “traditional” SEMs

What R package should I use?

lavaan

- Frequentist package specific to SEMs
- **Can** accommodate latent, composite variables
- Designed for “traditional” SEMs

piecewiseSEM

- Frequentist package specific to SEMs
- Can’t accommodate latent, composite variables
- Model structure must be acyclic
- Designed for non-normal data, hierarchical data

What R package should I use?

lavaan

- Frequentist package specific to SEMs
- **Can** accommodate latent, composite variables
- Designed for “traditional” SEMs

piecewiseSEM

- Frequentist package specific to SEMs
- Can’t accommodate latent, composite variables
- Model structure must be acyclic
- Designed for non-normal data, hierarchical data

brms

- General Bayesian modelling package
- Cannot easily code latent or composite variables
- Model structure must be acyclic
- Otherwise extremely flexible
- Allows for tools like posterior predictive checks
- Can be slow to run

Step 2: Design your model

Latent variables

Latent variables cannot be observed,
but can be represented by *indicator variables*

Latent variables

Latent variables cannot be observed,
but can be represented by *indicator variables*

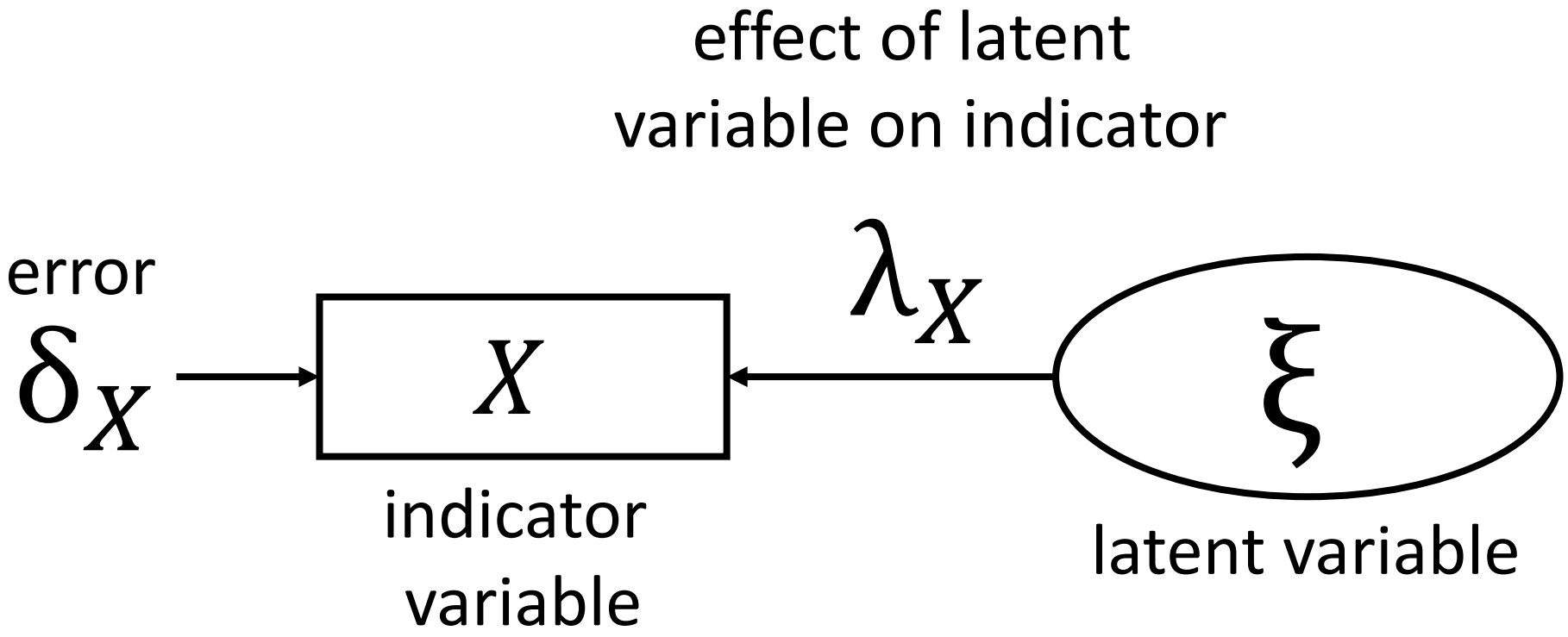
Examples of latent variables:

- body size
- biodiversity
- habitat quality
- ecosystem function
- ecosystem recovery

Latent variable	Indicator variables
body size	body mass, length, height
biodiversity	species richness, species evenness, phylogenetic diversity
habitat quality (seagrass)	shoot density, blade length, patch size

Step 2: Design your model

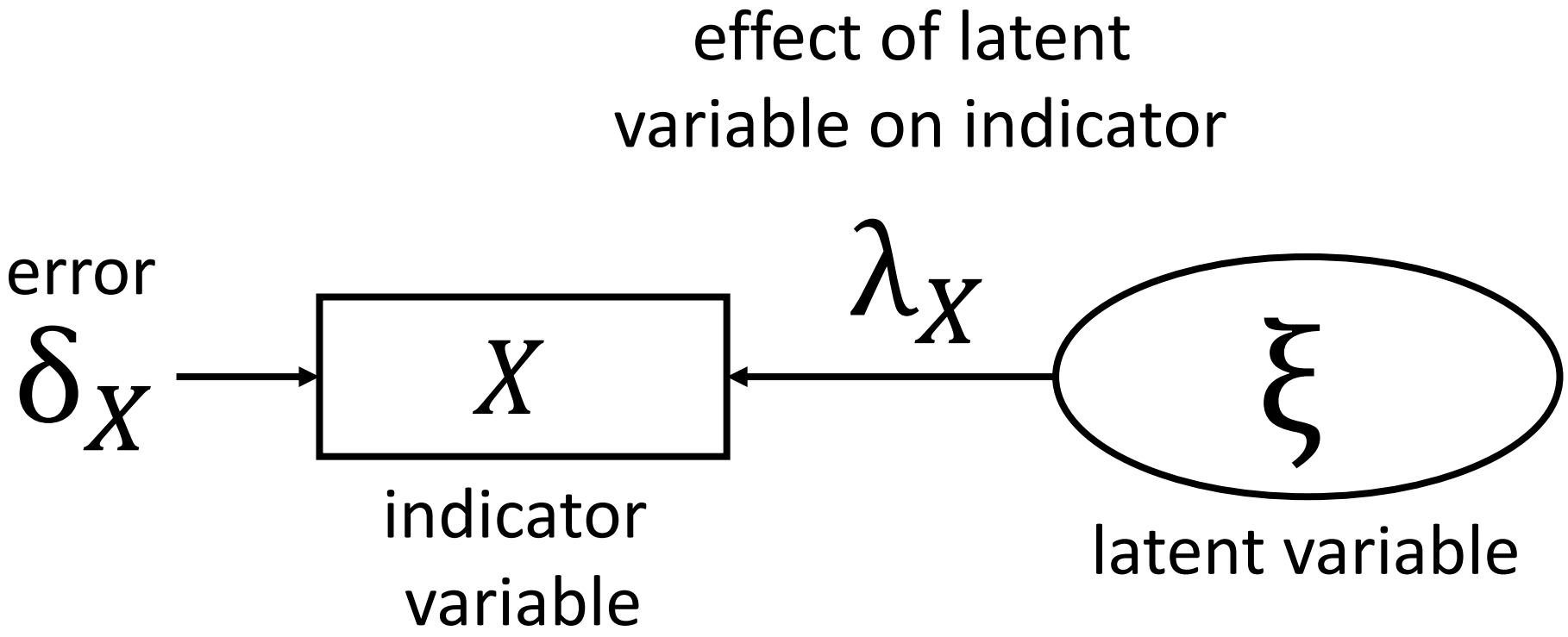
Modelling a latent variable



Step 2: Design your model

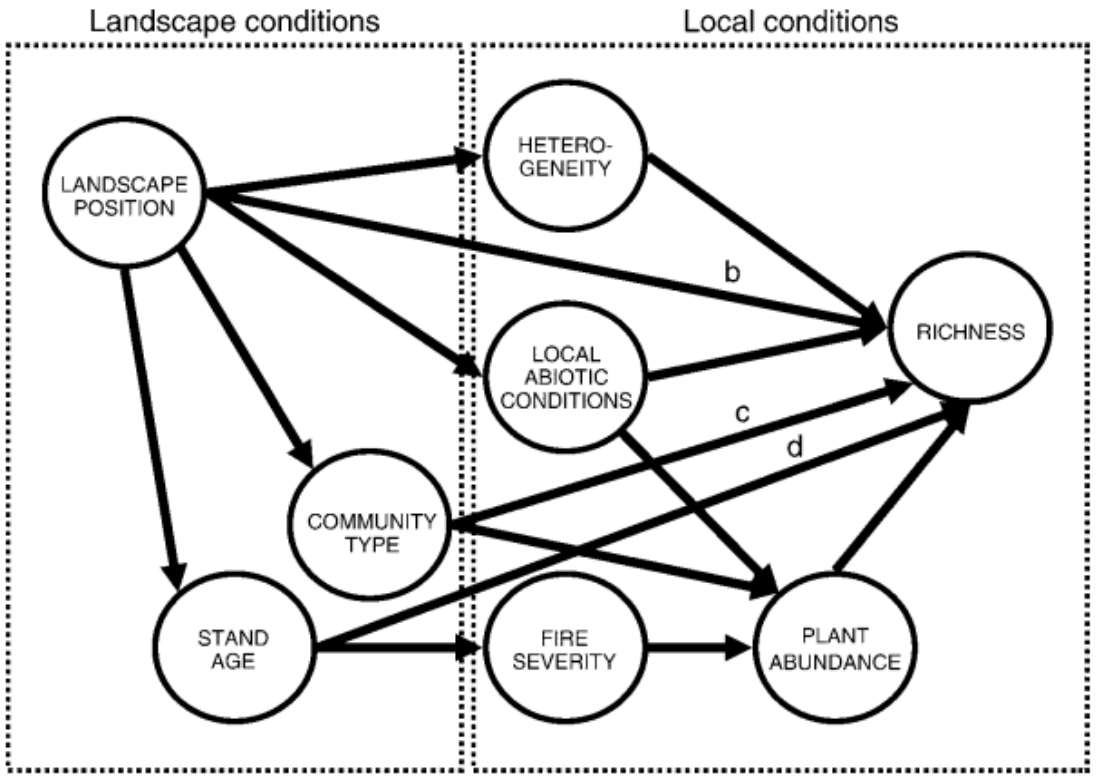
Modelling a latent variable

latent variables can be modelled using the lavaan package



Step 2: Design your model

Exercise 2: Coding a classic SEM using *piecewiseSEM*



How do wildfires affect plant diversity across a landscape?

Grace, J. B., & Keeley, J. E. (2006). A structural equation model analysis of postfire plant diversity in California shrublands. *Ecological Applications: A Publication of the Ecological Society of America*, 16(2), 503–514.

Step 3: Code your model

Checking your model fit

- Check model fit
- Check that your path coefficients make sense!

Check model fit: Tests of directed separation

- Tests whether we are missing any paths
- Fits missing relationships and tests whether path coefficients are significantly different from 0
- If any values $p < 0.05$, need to re-evaluate model

Step 4: Check your model

What if your model fits, but it doesn’t make sense?

Example: intertidal competition

<u>System:</u>

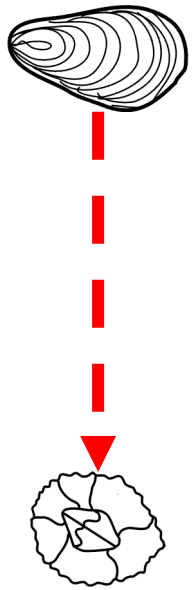
- Barnacles recruit in early spring

Step 4: Check your model

Example: intertidal competition

<u>System:</u>

- Barnacles recruit in early spring
- Mussels recruit after barnacles,
  but are better competitors for space



negative effect

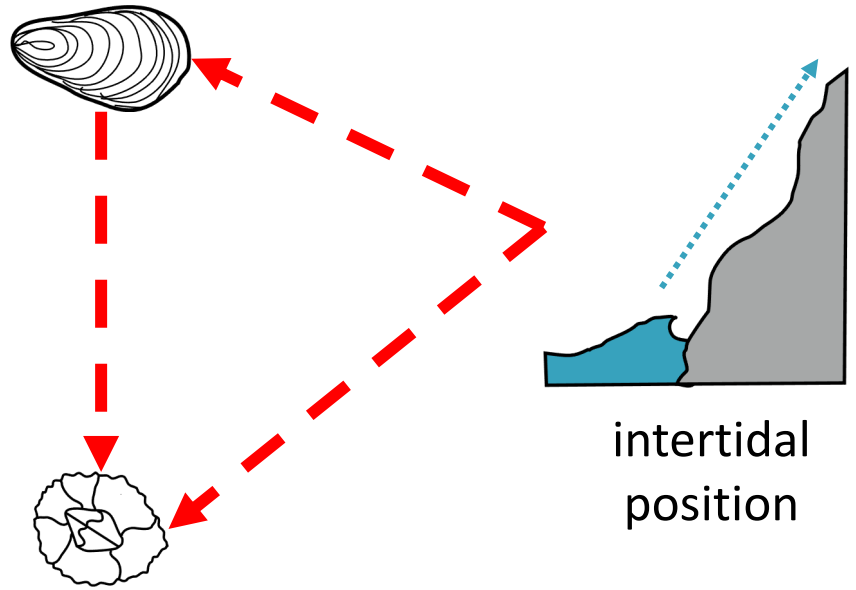
positive effect

Step 4: Check your model

Example: intertidal competition

<u>System:</u>

- Barnacles recruit in early spring
- Mussels recruit after barnacles, but are better competitors for space
- Both species decrease in abundance moving from the low -> high intertidal



negative effect

positive effect

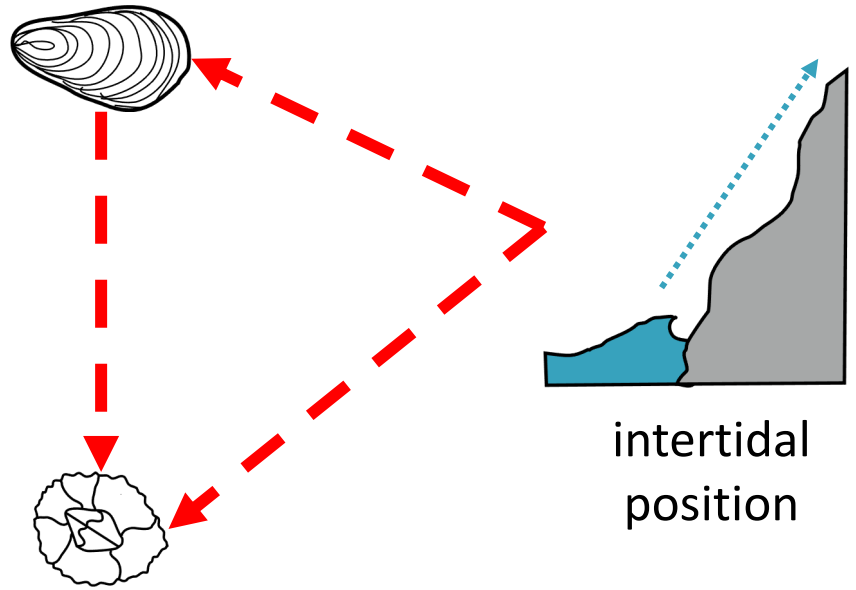
Step 4: Check your model

Example: intertidal competition

<u>System:</u>

- Barnacles recruit in early spring
- Mussels recruit after barnacles,
  but are better competitors for space
- Both species decrease in abundance
  moving from the low -> high intertidal

We collect abundance data over a 10 year period

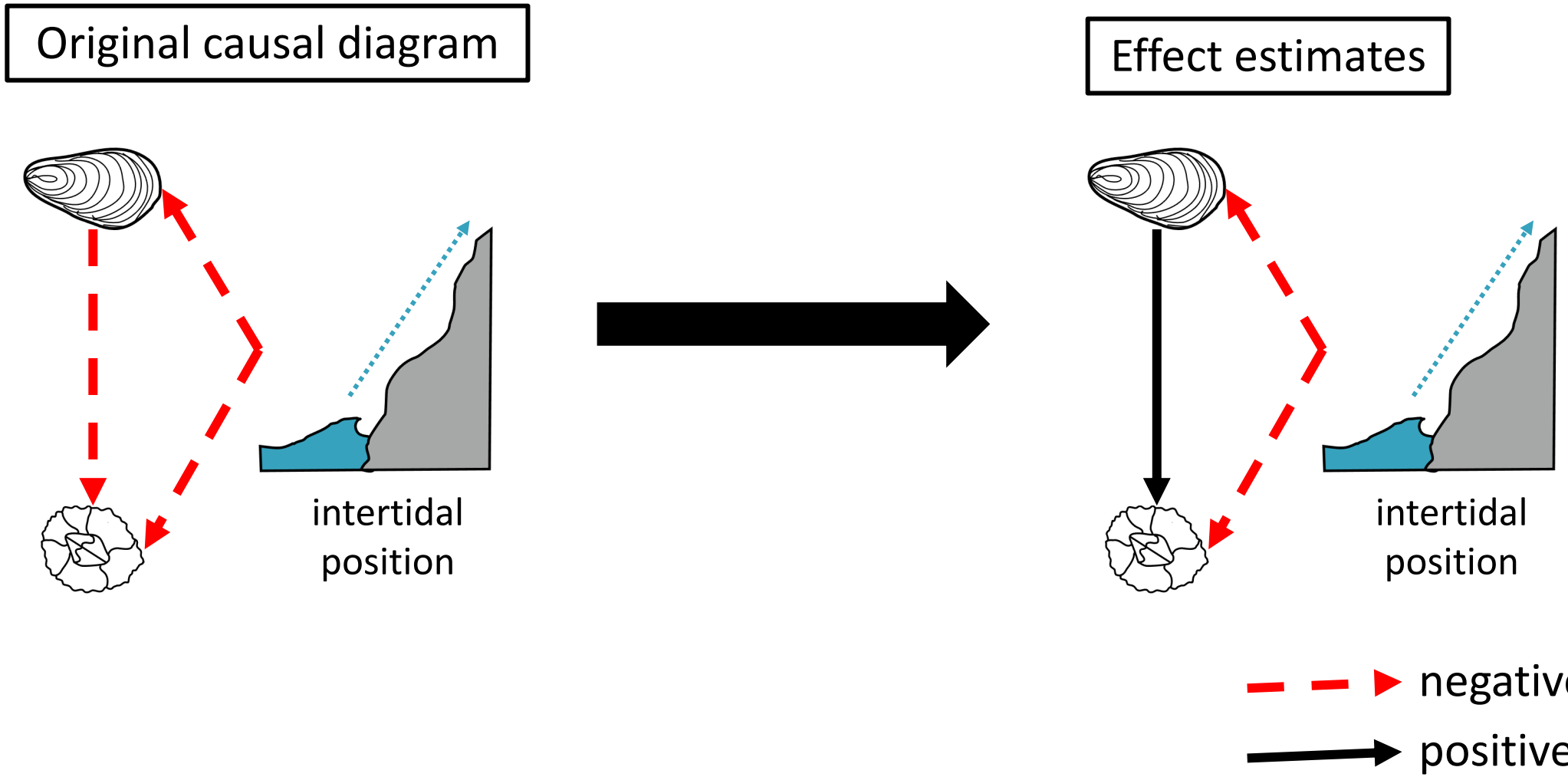


Step 4: Check your model

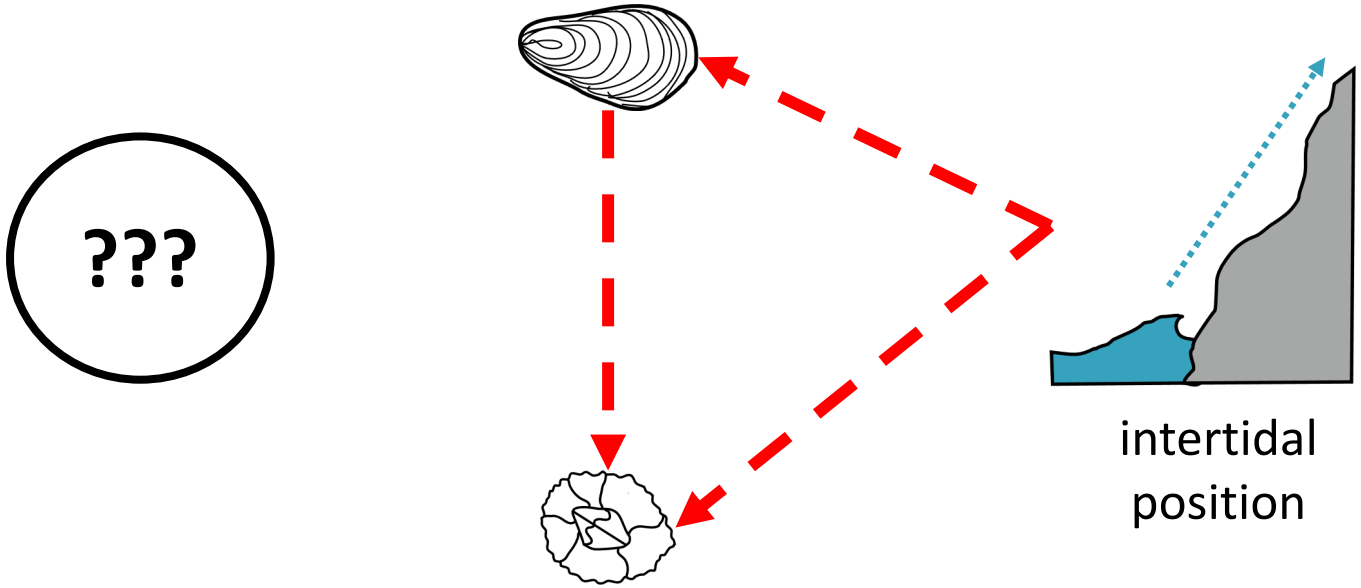
Confounding variables

- An unmeasured variable influencing both a predictor and a response variable

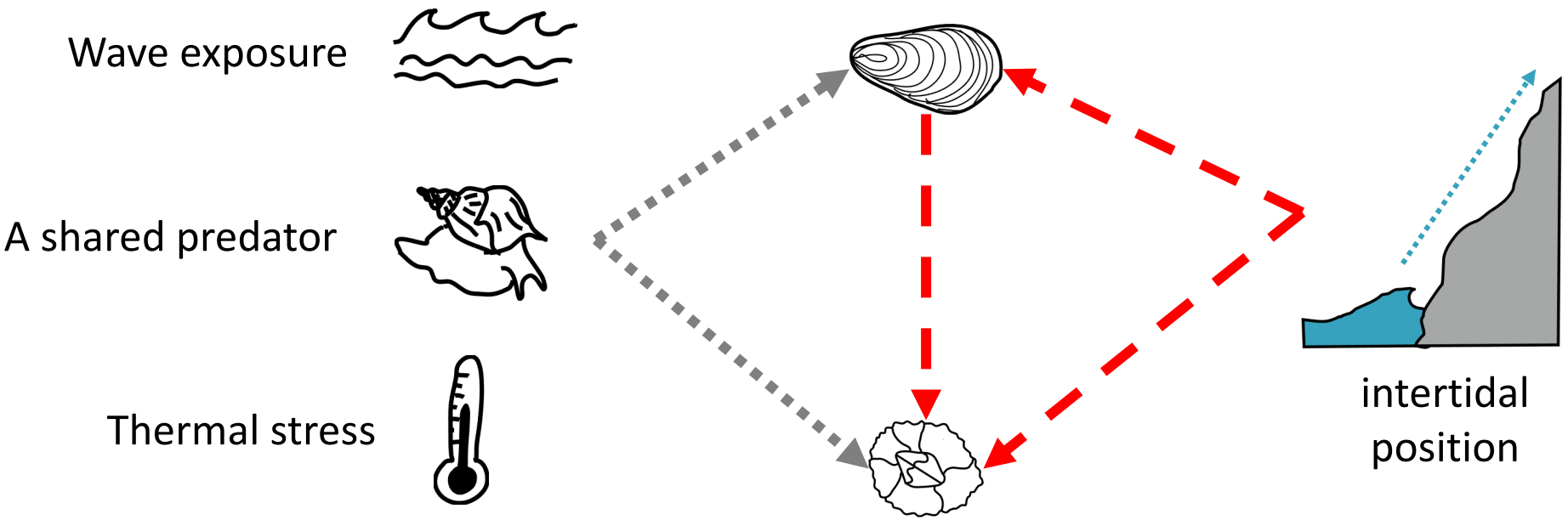
Sometimes you get the opposite effect estimate that you hypothesized



Confounding variables, a phantom menace

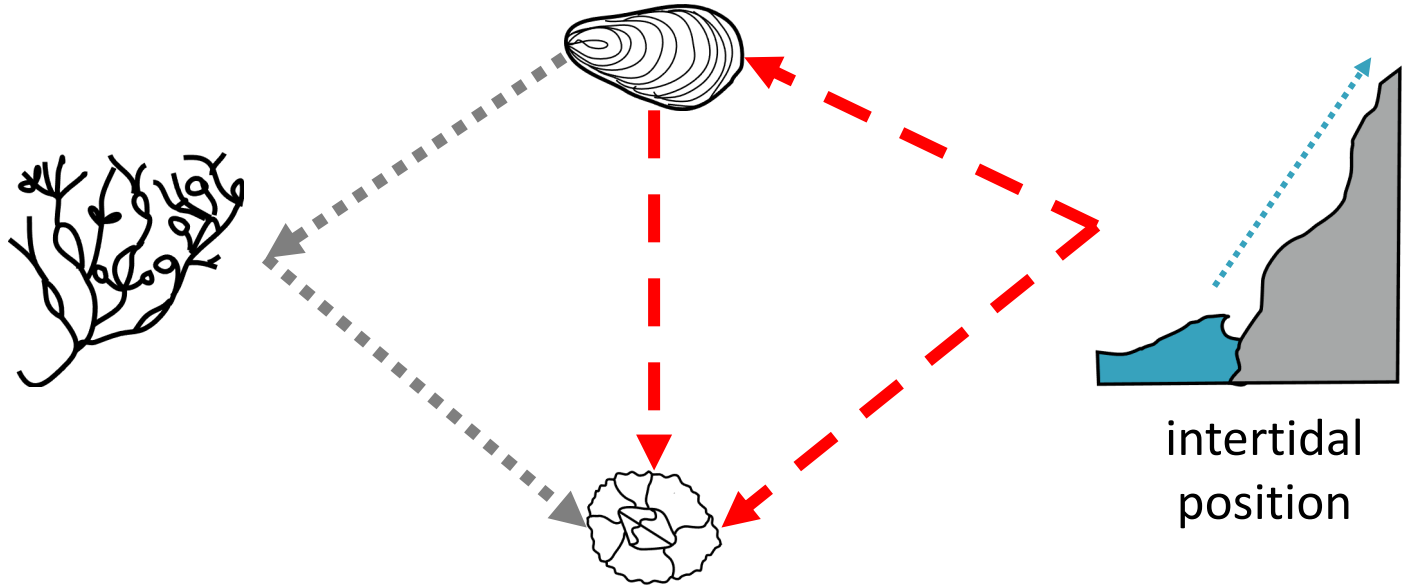


Confounding variables might be a shared predictor



Confounding variables might be a mediator

An algae that facilitates predators



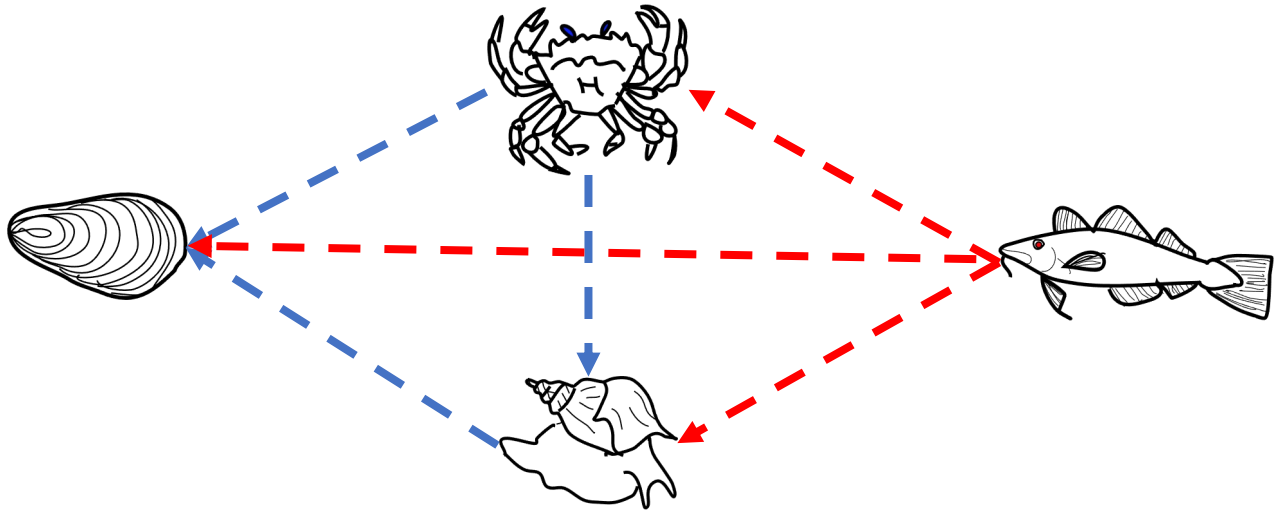
intertidal position

negative effect

positive effect

Even if a confounding variable doesn't change the sign of a path/effect, the strength of the effect might change (and this is harder to detect)

Exercise 5:
A simple intertidal food web in *brms*



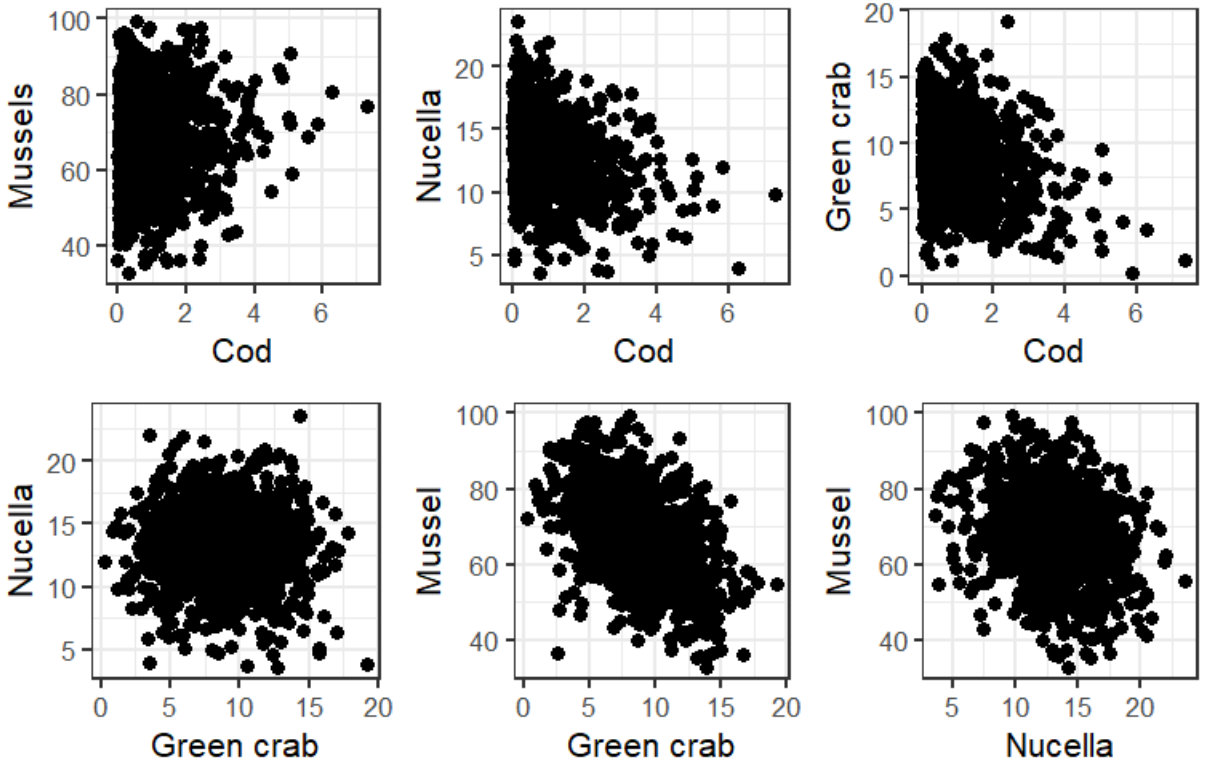
What is the effect
of cod predation
on intertidal communities?

Step 5: Interpret your model

Exercise 5:
A simple intertidal food web in *brms*

Built SEM in brms using fake data

fake_intertidal_data.csv



intertidal_SEM.rds

```
> intertidal_SEM
 Family: MV(gamma, gamma, gamma)
  Links: mu = log; shape = identity
         mu = log; shape = identity
         mu = log; shape = identity
Formula: green_crab ~ cod_CEN
         nucella ~ cod_CEN + green_crab_CEN
         mussel ~ cod_CEN + green_crab_CEN + nucella_CEN
   Data: intertidal_dat (Number of observations: 997)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
greencrab_Intercept      2.19      0.01     2.16     2.21 1.00     8807     2980
nucella_Intercept        2.57      0.01     2.56     2.59 1.00     9765     2798
mussel_Intercept         4.20      0.00     4.19     4.21 1.00     9840     3101
greencrab_cod_CEN       -0.12      0.01    -0.15    -0.10 1.00     8854     3086
nucella_cod_CEN         -0.09      0.01    -0.10    -0.07 1.00     7874     3100
nucella_green_crab_CEN  -0.03      0.01    -0.05    -0.02 1.00     7217     3056
mussel_cod_CEN          -0.03      0.01    -0.04    -0.02 1.00     7705     3346
mussel_green_crab_CEN   -0.10      0.00    -0.11    -0.09 1.00     7886     3300
mussel_nucella_CEN      -0.05      0.01    -0.06    -0.04 1.00     7934     3379

Family Specific Parameters:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
shape_greencrab   7.28      0.32     6.67     7.92 1.00     8268     2927
shape_nucella    17.64      0.79    16.14    19.25 1.00     8204     2898
shape_mussel     44.53      1.99    40.64    48.39 1.00     7907     3047

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Step 5: Interpret your model

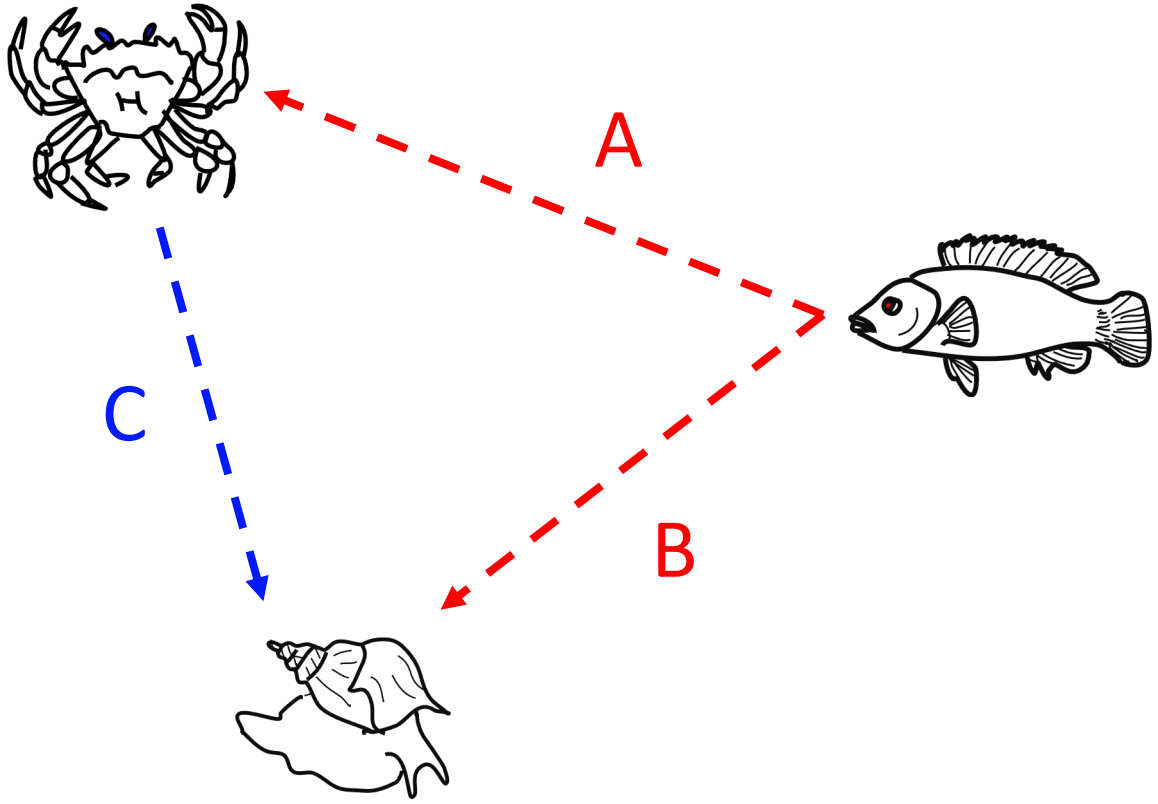
Exercise 5:
A simple intertidal food web in *brms*

Net effect of cunner on mussels

= A + (B * C)

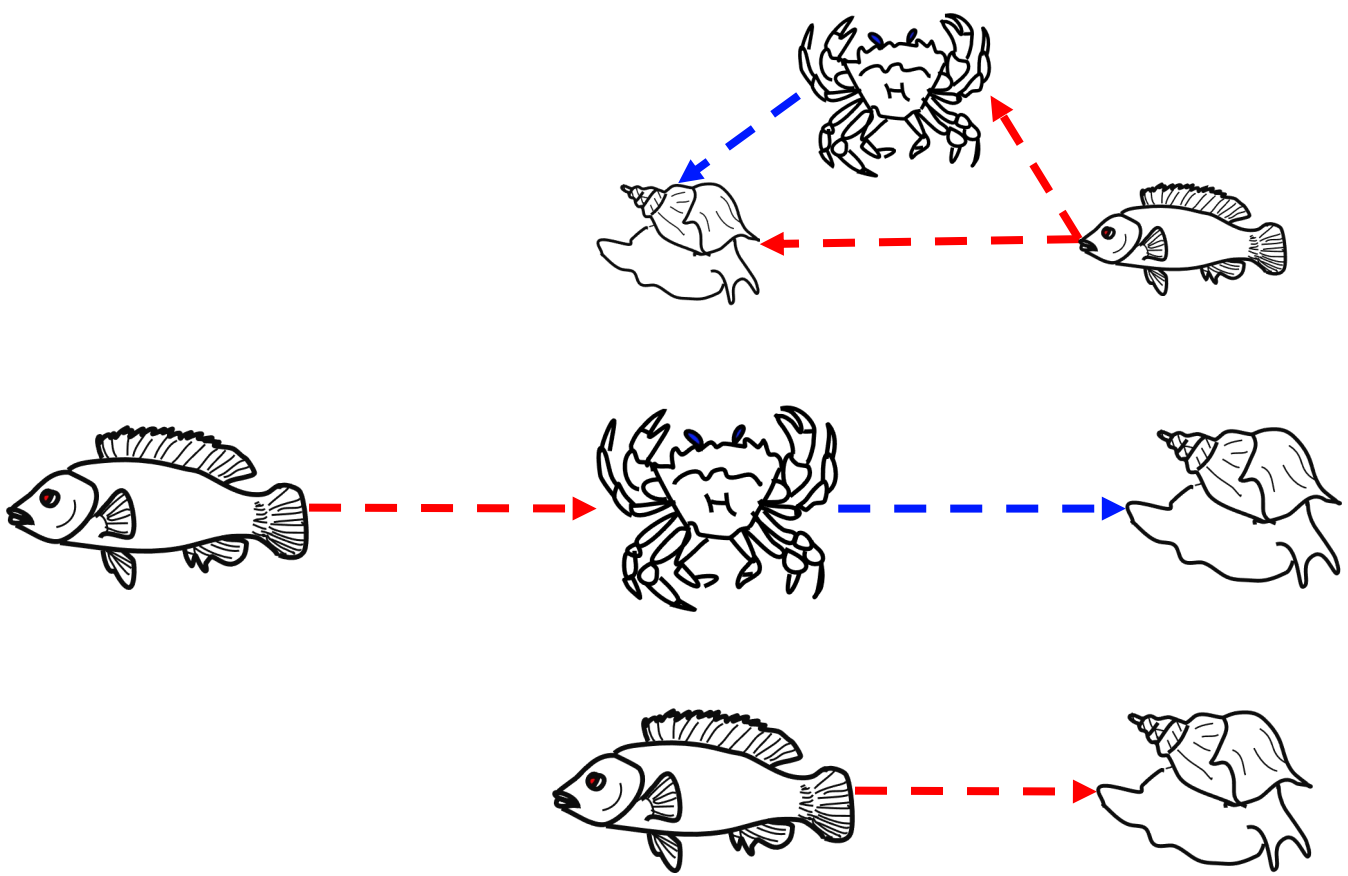
direct effect

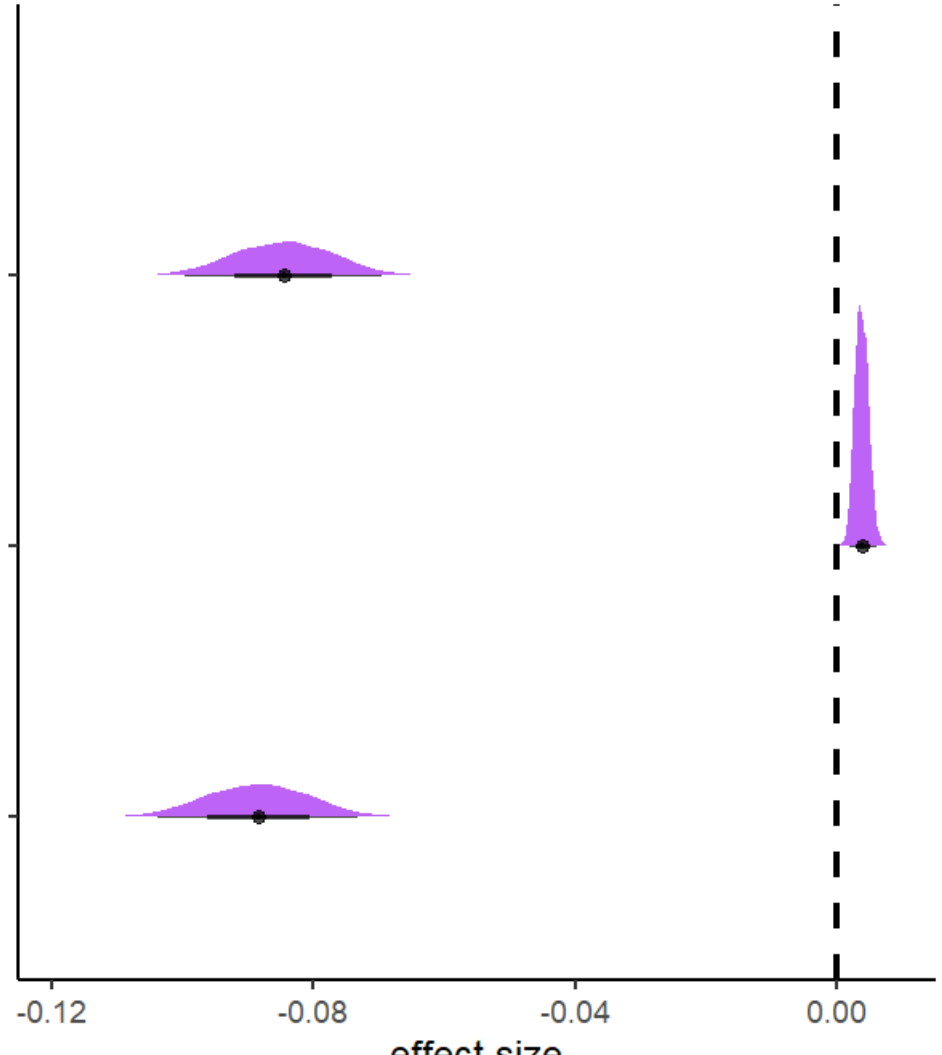
indirect effect



Step 5: Interpret your model

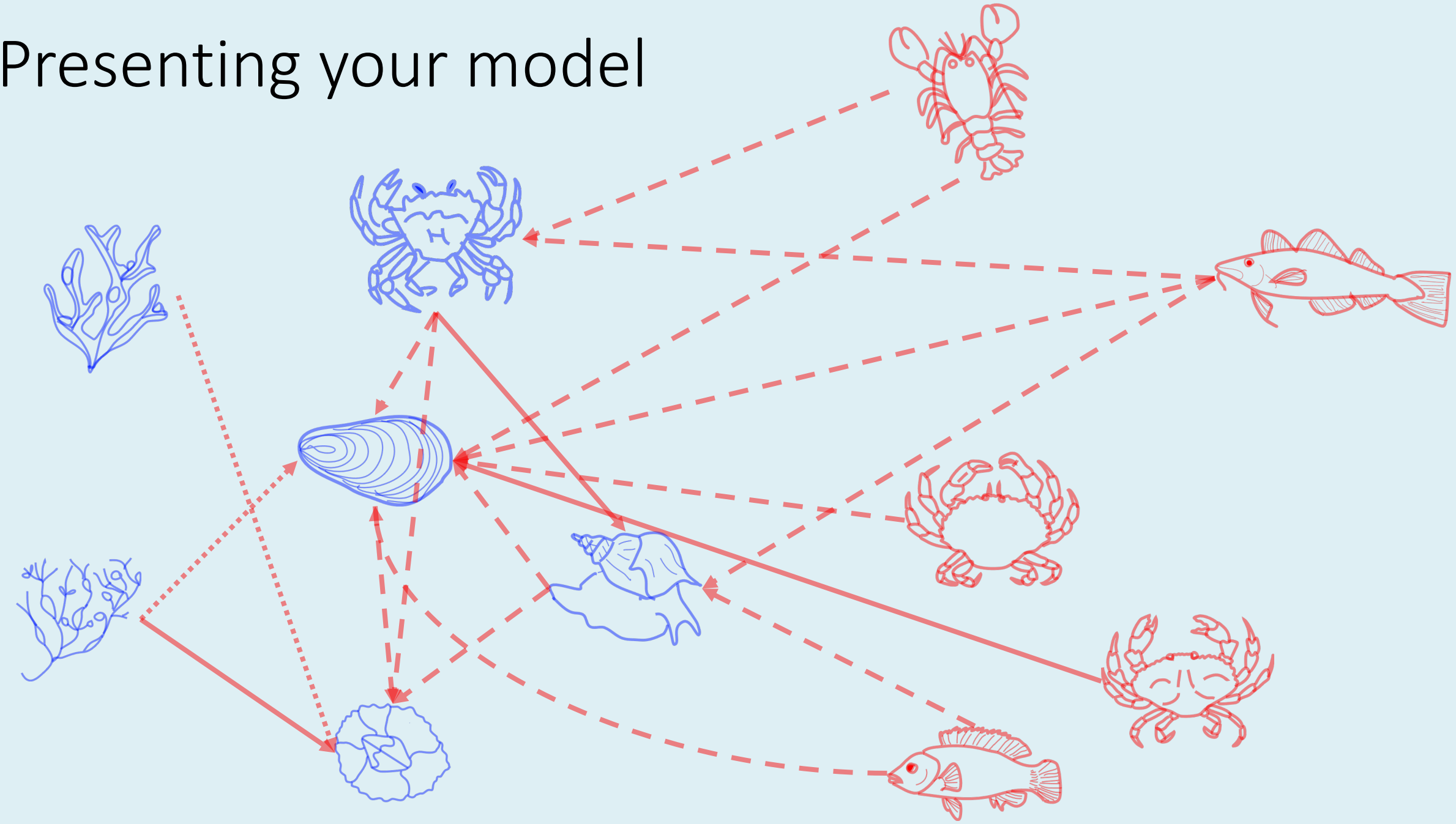
Exercise 5:
A simple intertidal food web in *brms*





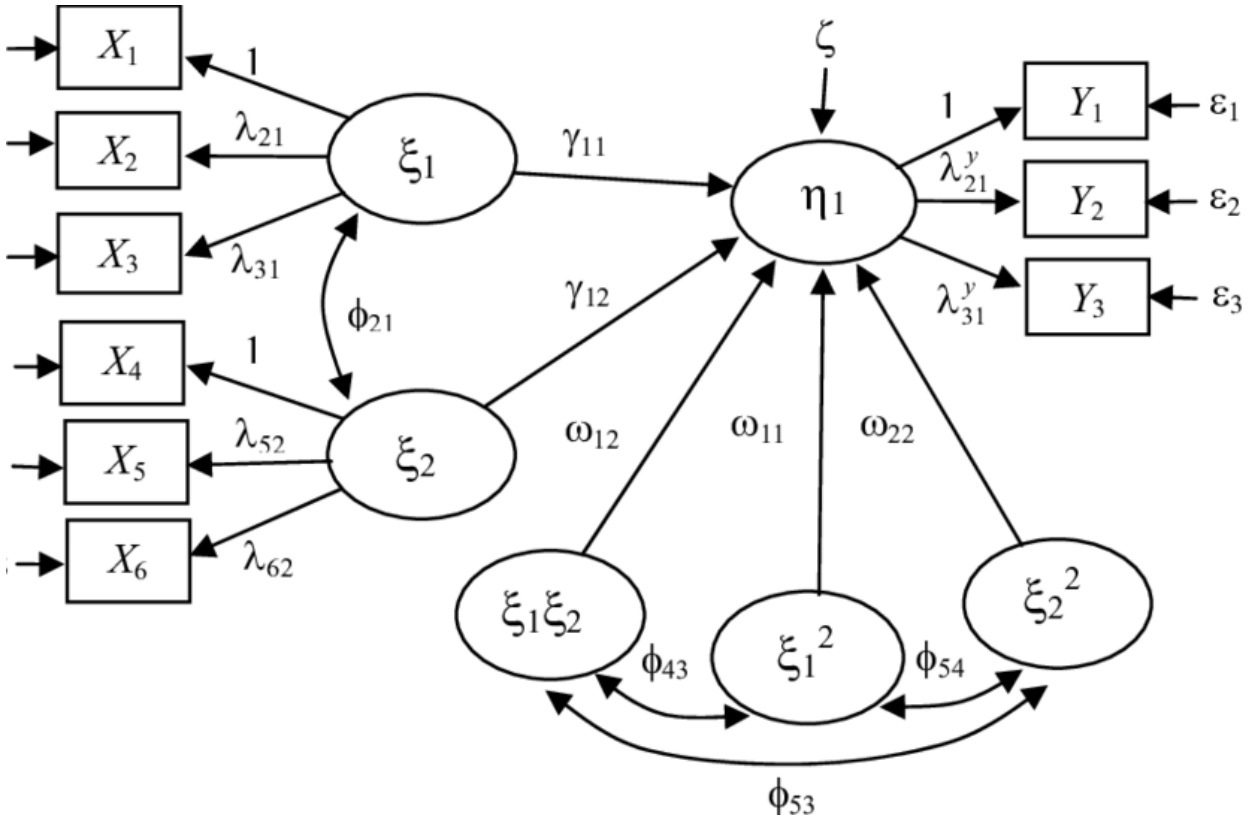
Step 5: Interpret your model

Presenting your model



What not to do

Throw something like this up on the screen and start talking



Step 6: Present your model

What to do

- Explain what a SEM is

Step 6: Present your model

What to do

- Explain what a SEM is
- Explain how your SEM tests your research questions

Step 6: Present your model

What to do

- Explain what a SEM is
- Explain how your SEM tests your research questions
- Build up your SEM slowly and explain the important paths

Step 6: Present your model

What to do

- Explain what a SEM is
- Explain how your SEM tests your research questions
- Build up your SEM slowly and explain the important paths
- Consider not showing control variables/unimportant variables

Step 6: Present your model

What to do

- Explain what a SEM is
- Explain how your SEM tests your research questions
- Build up your SEM slowly and explain the important paths
- Consider not showing control variables/unimportant variables
- If you put up coefficients, explain units

Step 6: Present your model

What to do

- Explain what a SEM is
- Explain how your SEM tests your research questions
- Build up your SEM slowly and explain the important paths
- Consider not showing control variables/unimportant variables
- If you put up coefficients, explain units
- Take your time!

Step 6: Present your model

Thank you for listening!

If you have questions, feel free to contact me

nicole.knight@mail.mcgill.ca