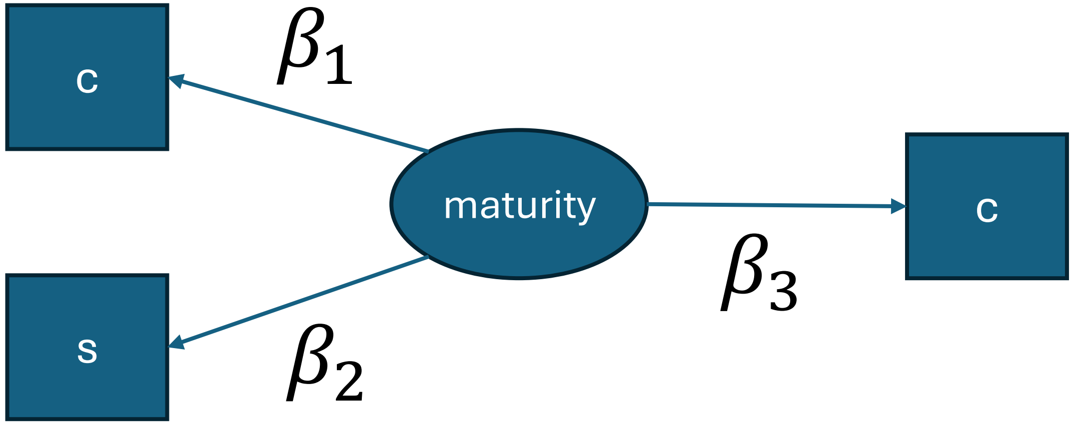
Latent variables: Part I



Quick announcements

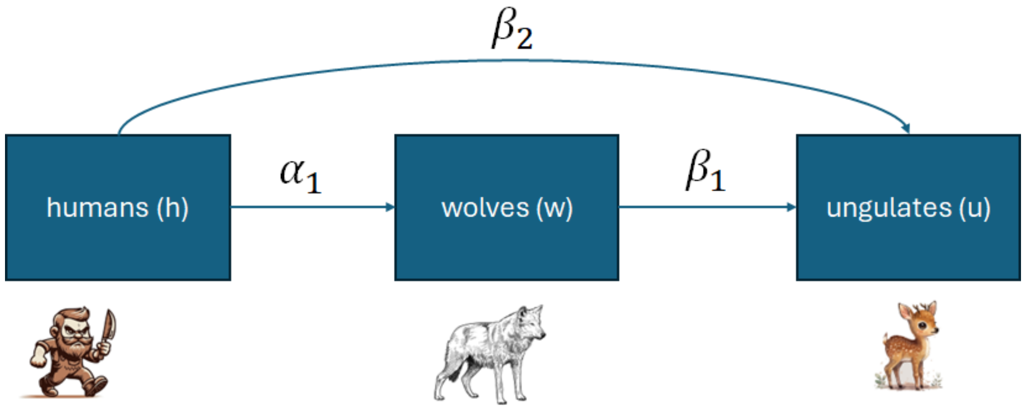
1. Class cancelled (i.e., delayed) on Oct. 10th

2. Final topic due on or around Oct 10th*

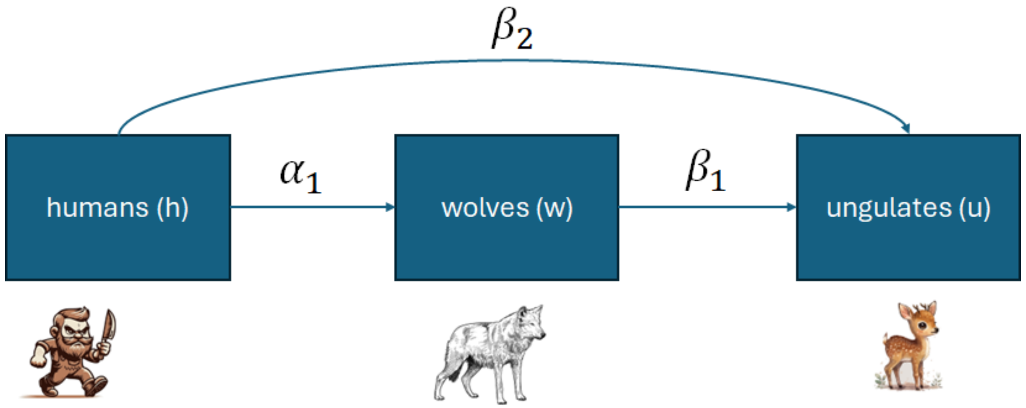
3. Let’s plan to begin 1-on-1s during office hours or by appointment if you’d like to discuss how to set up your data, simulating data, etc.

*you can still change it if there’s an insurmountable problem

Reviewing last week's question: what if one covariate affects another?

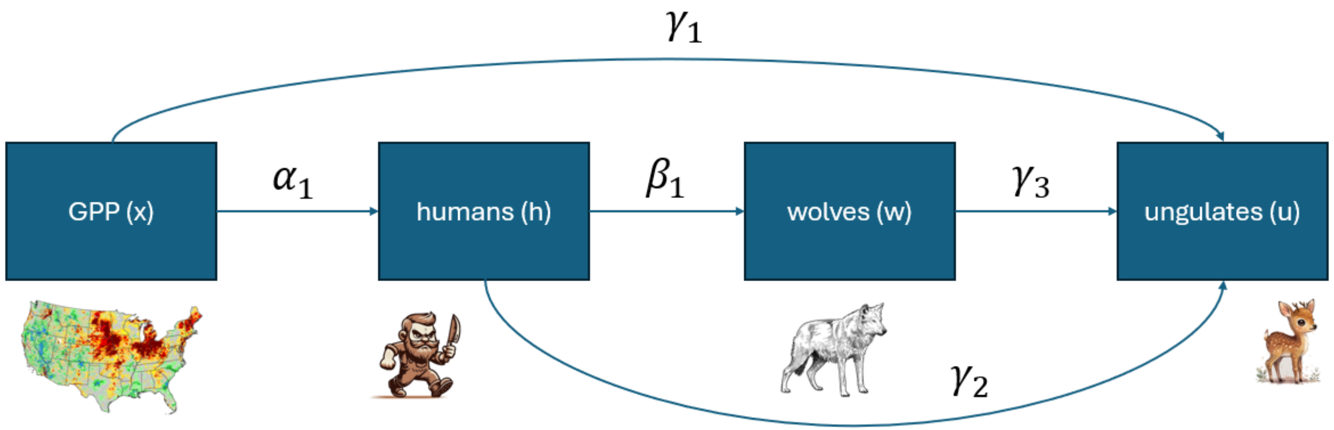


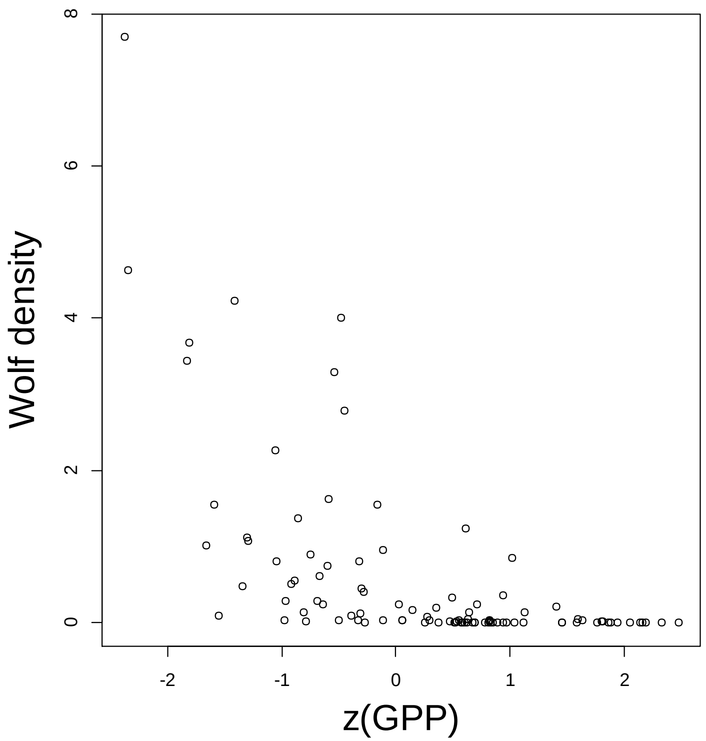
What if one covariate affects another?



Multicollinearity will occur. This can be problematic.

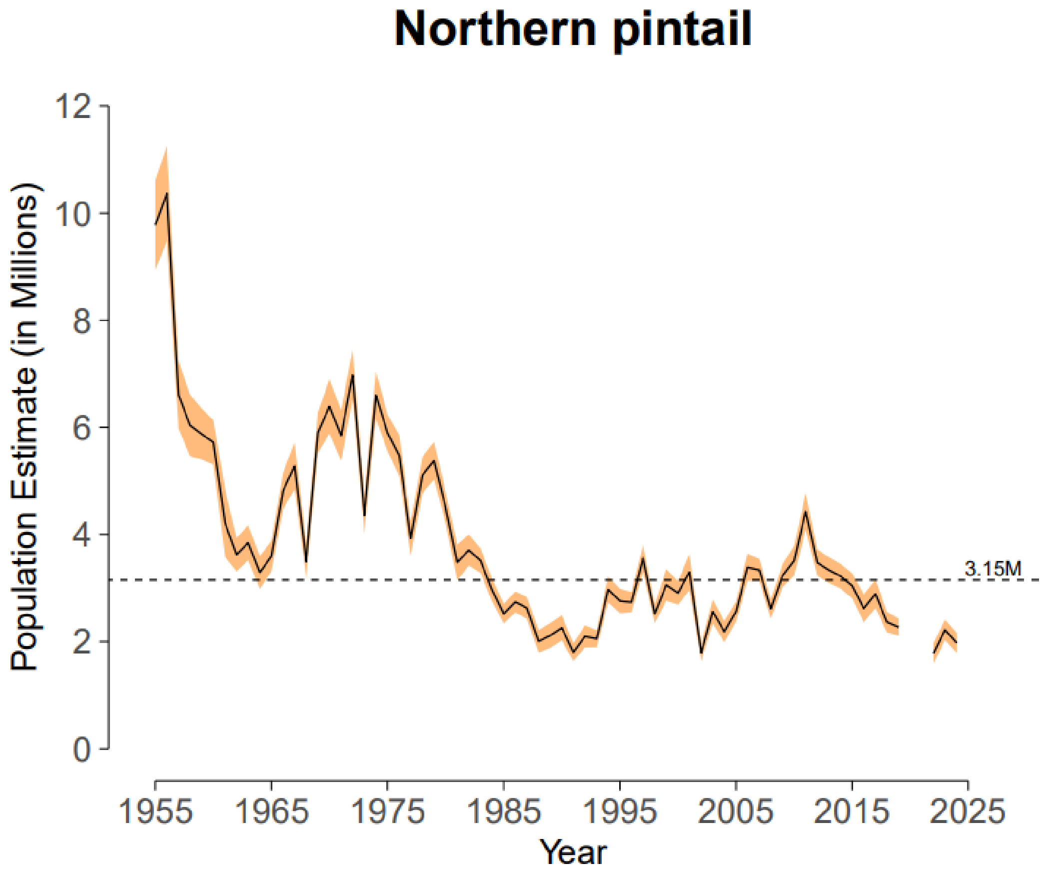
What if one covariate affects another?



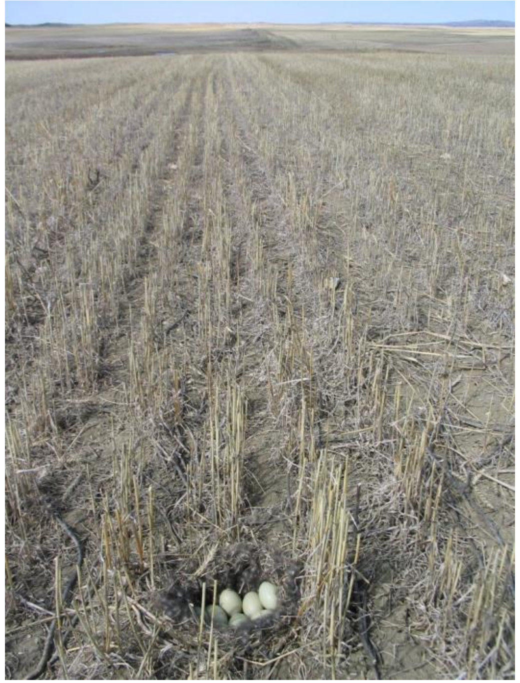


Wild responses can occur!

A quick, real-life example from pintails (not ducks!)



The textbook example of an ecological trap



Devries et al. **(2023)** *Wildlife Monographs*

The textbook example of an ecological trap

Devries et al. **(2023)** *Wildlife Monographs*

95% of nests in spring-seeded crops fail

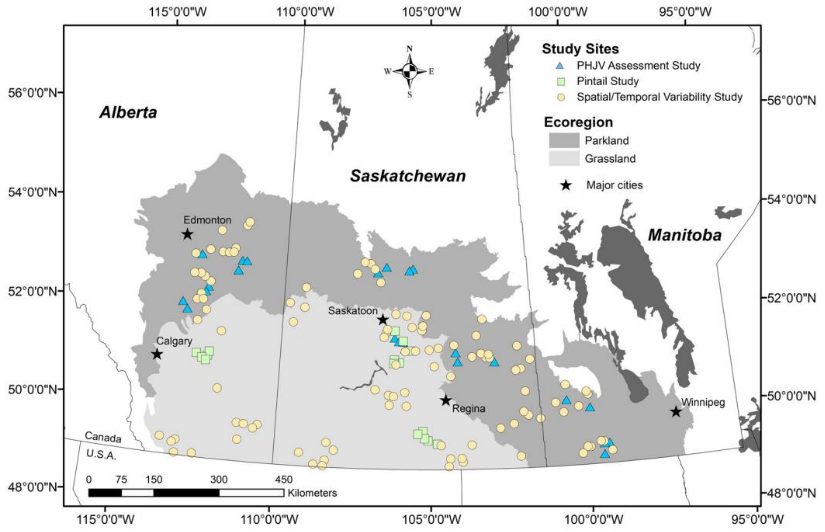


FIGURE 2 Distribution and location of 145 Prairie Habitat Joint Venture (PHJV) Assessment Study, Pintail Study, and Spatial and Temporal Variability Study (SPATS) sites used in the modeling of duck nest survival and nest habitat selection within the Grassland and Parkland ecoregions of the Canadian Prairie Pothole Region (PPR), 1994–2009.

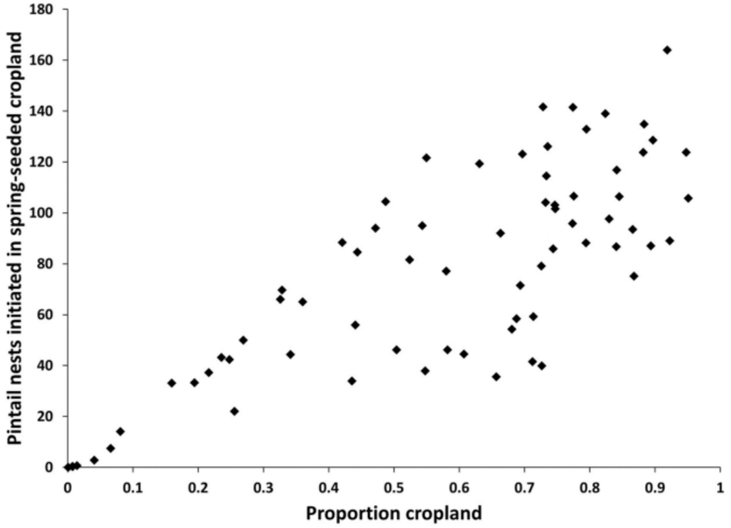
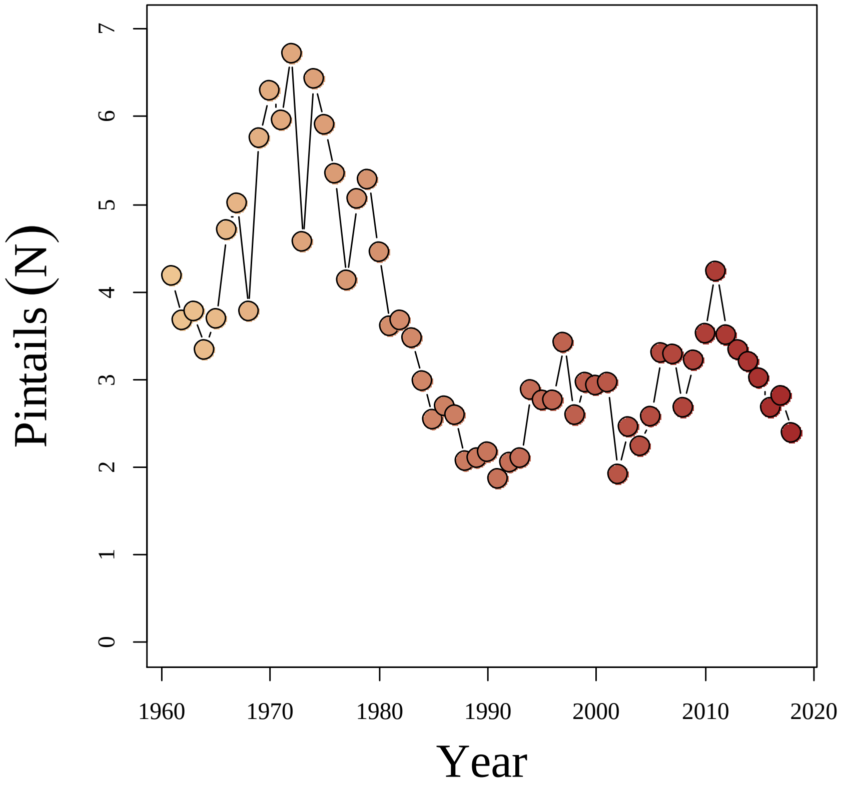


FIGURE 9 Example of the conservation planning model (CPM) estimated number of northern pintail nests initiated in spring-seeded cropland relative to the proportion of cropland present in 71 41-km^2 grids comprising the Coteau North Prairie Habitat Joint Venture (PHJV) Target Landscape in southern Saskatchewan, Canada (circa 2014).

Almost half (~45%) of pintail nests are in spring-seeded crops

Change in agricultural practices led to declines

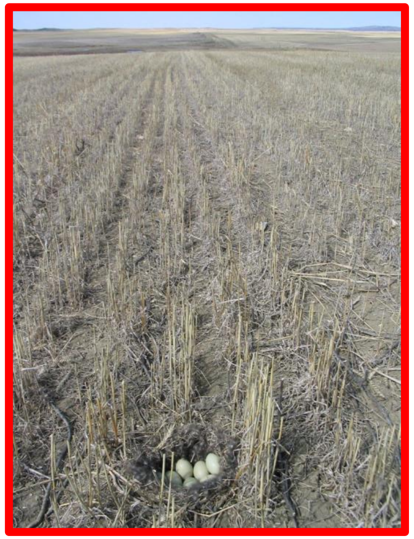
δ_{ag} (x) → Pop. growth (λ)



Changes in population growth rate led to changes in N

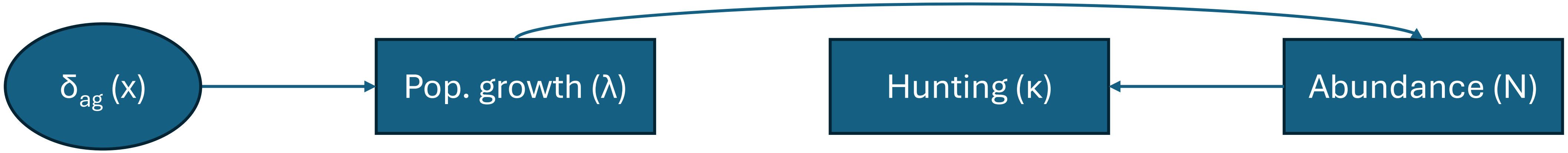
$$N_{t+1} = N_t \times \lambda_t$$

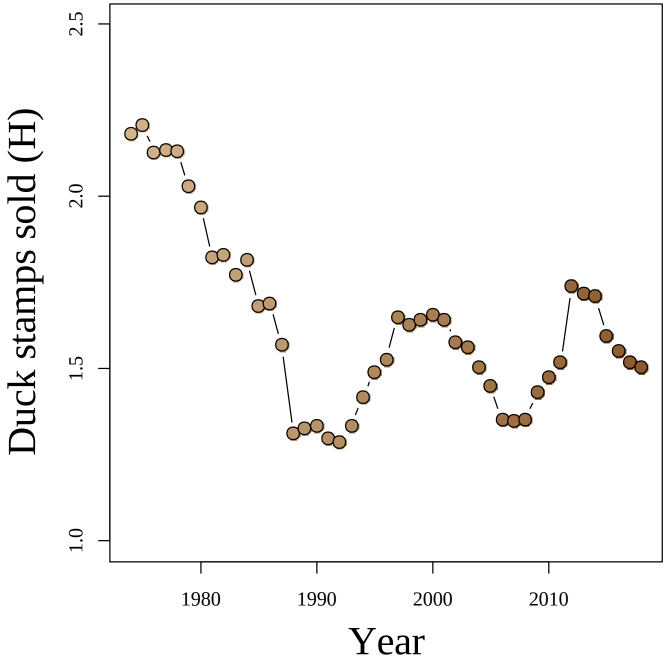
$$\lambda_t = S_t + R_t$$

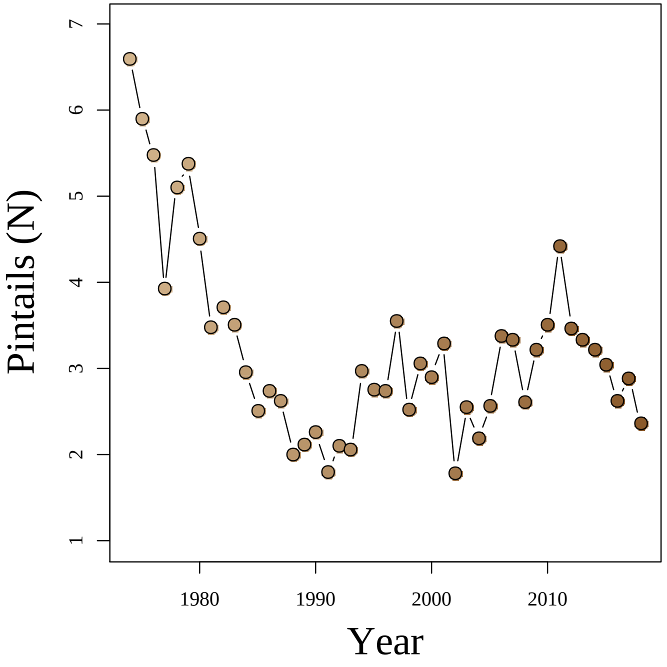


Imagine if 45% of the females of your favorite wildlife species were going to be run over by a tractor every year.

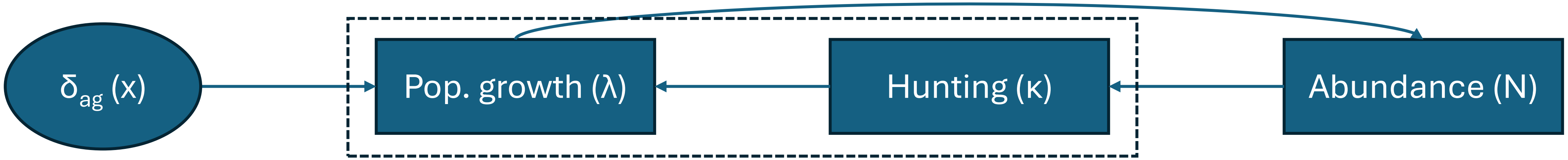
Changes in N lead to changes in hunting regs and effort





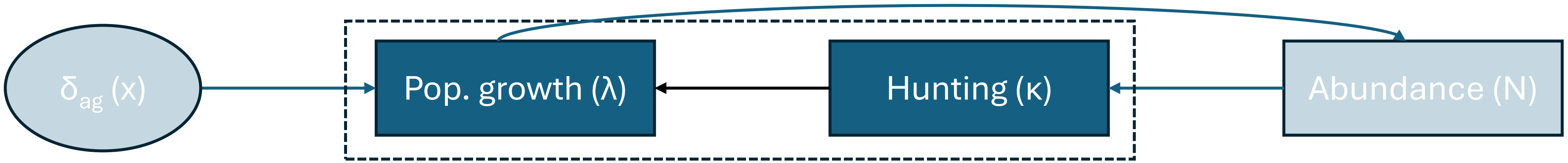


Hunting may also affect population growth rate



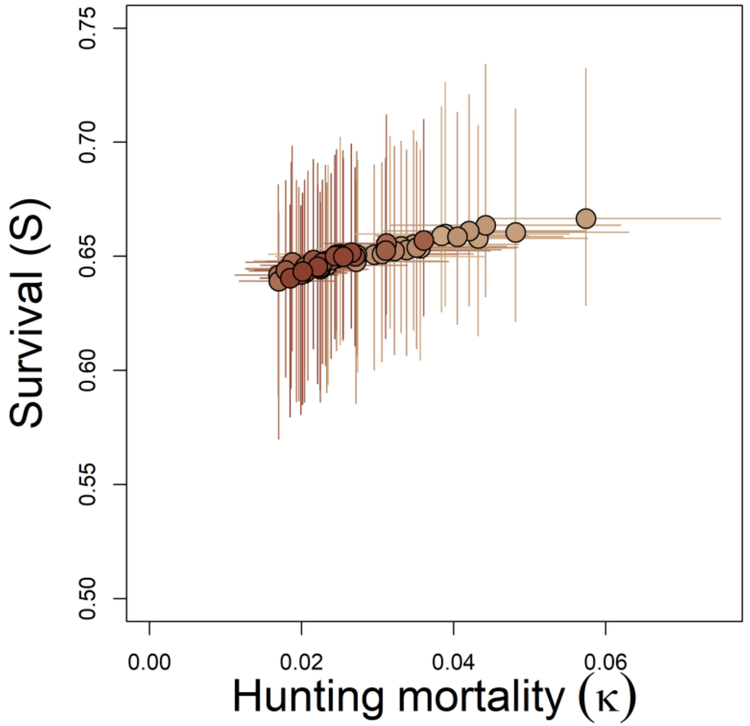
$$\lambda_t = S_t + R_t$$

If we just look at this model…

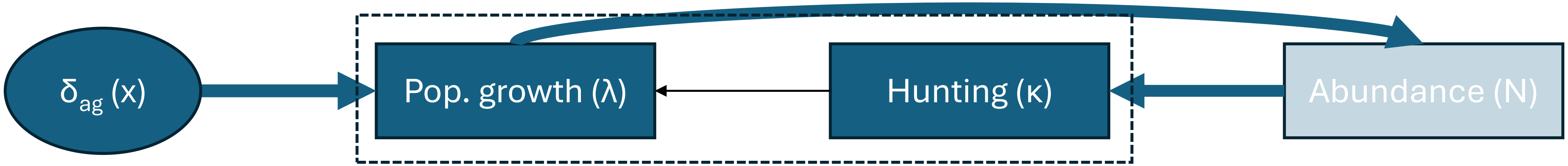


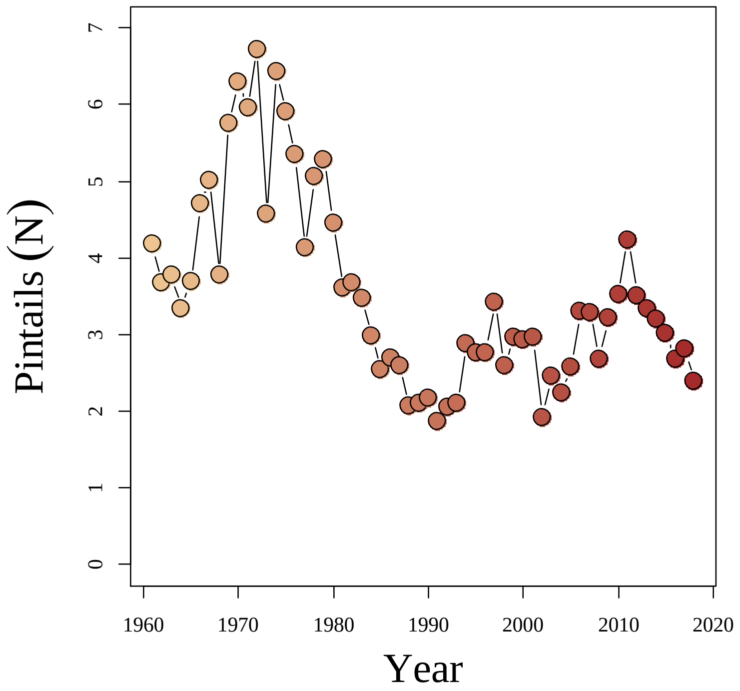
What is the effect of hunting on S?

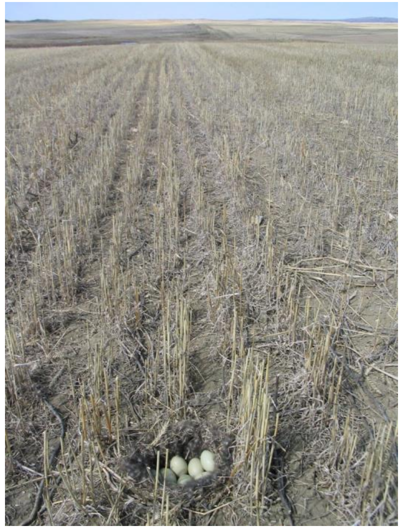
Weakly positive. Freaking great.

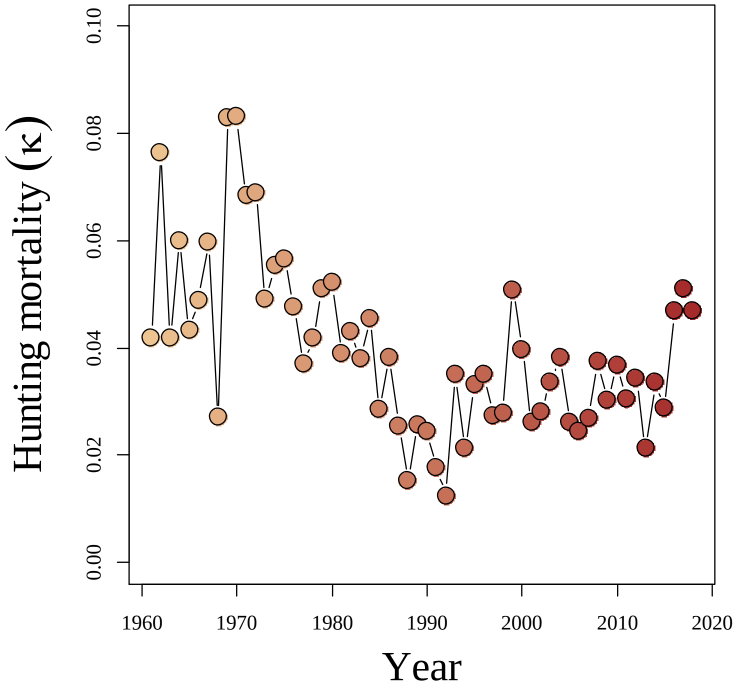


What happened?



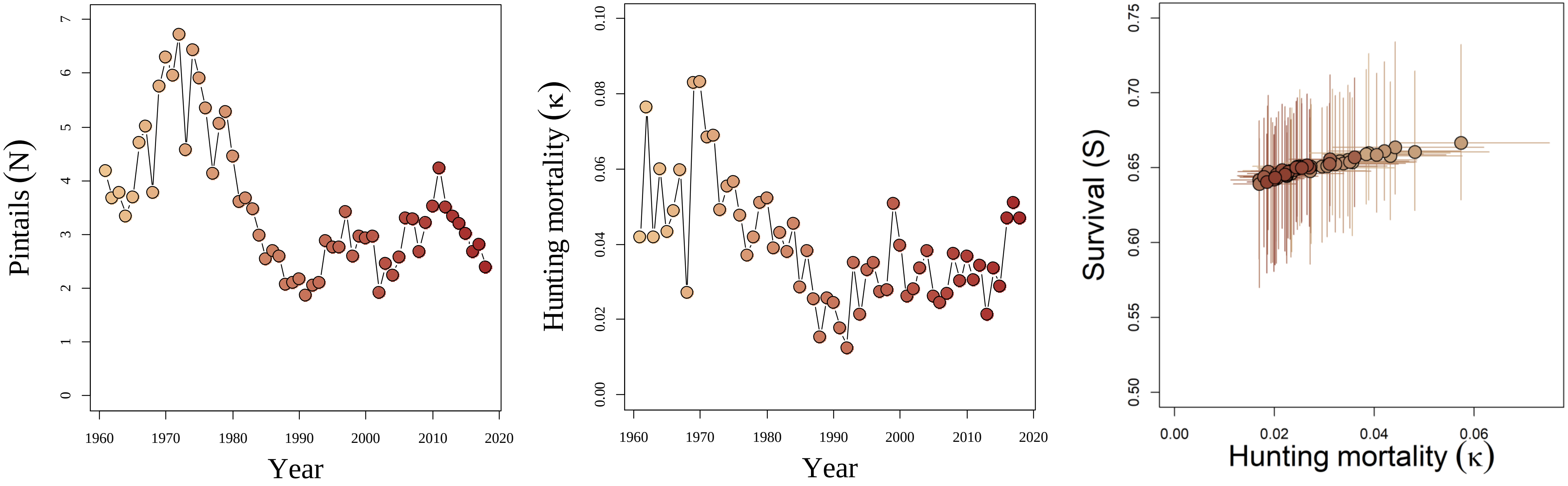


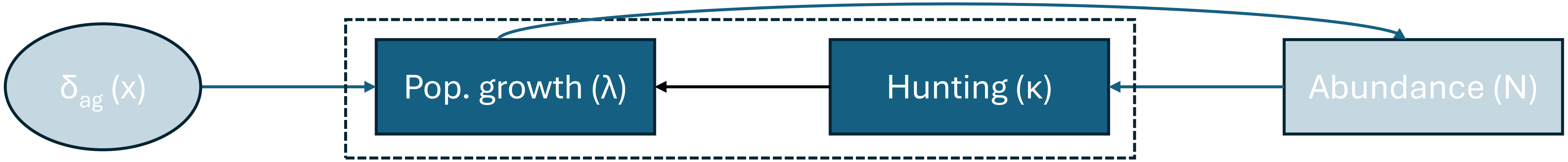




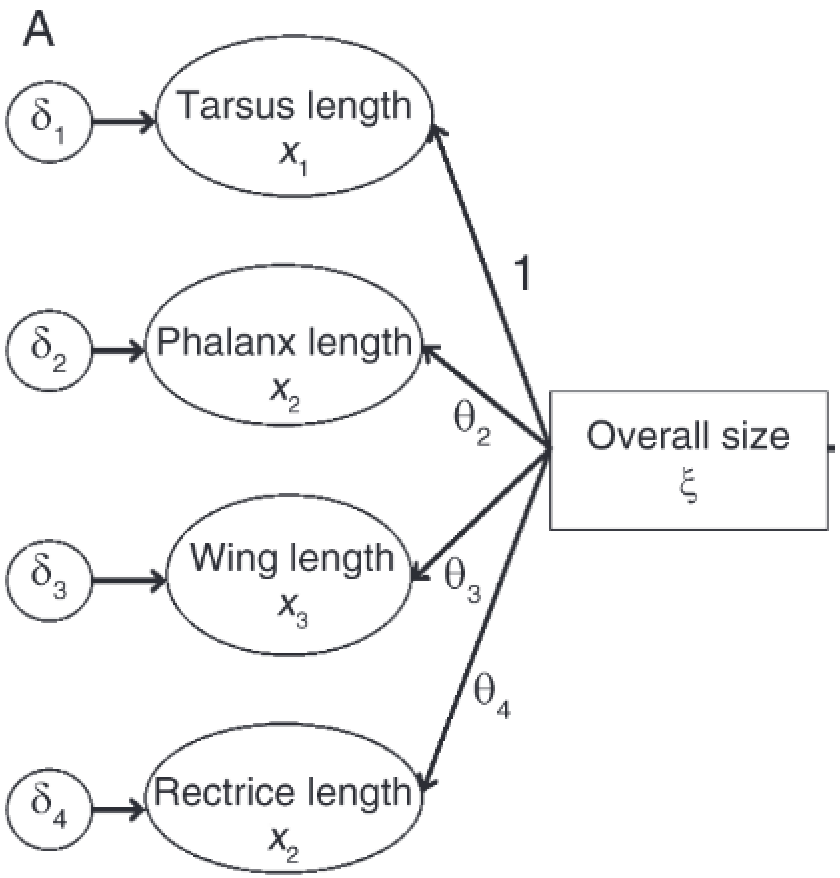
The effect of changes in agriculture masked any effect of hunting

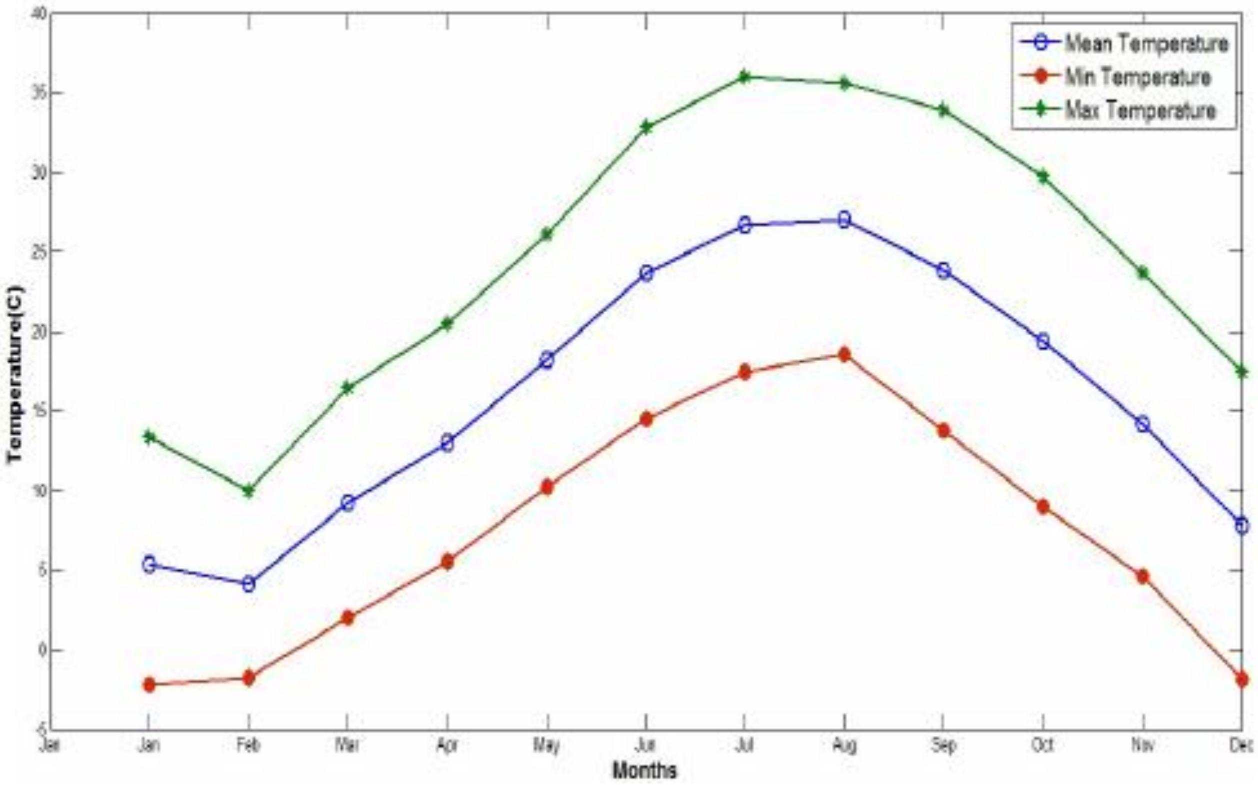
Covariates that affect each other can lead to deeply flawed inference if we don't think carefully



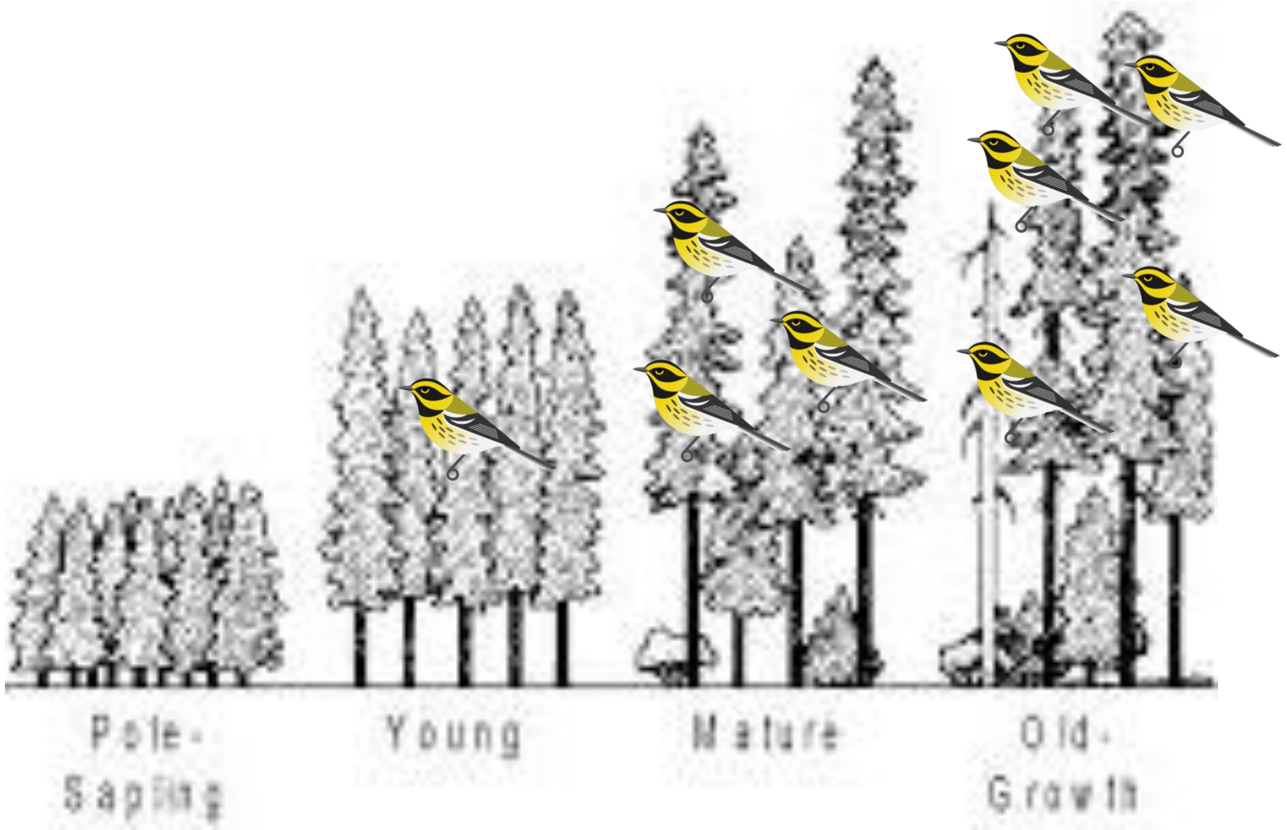


This week's question is similar: What if we're measuring the same thing?





Our first example: forest age as a latent variable

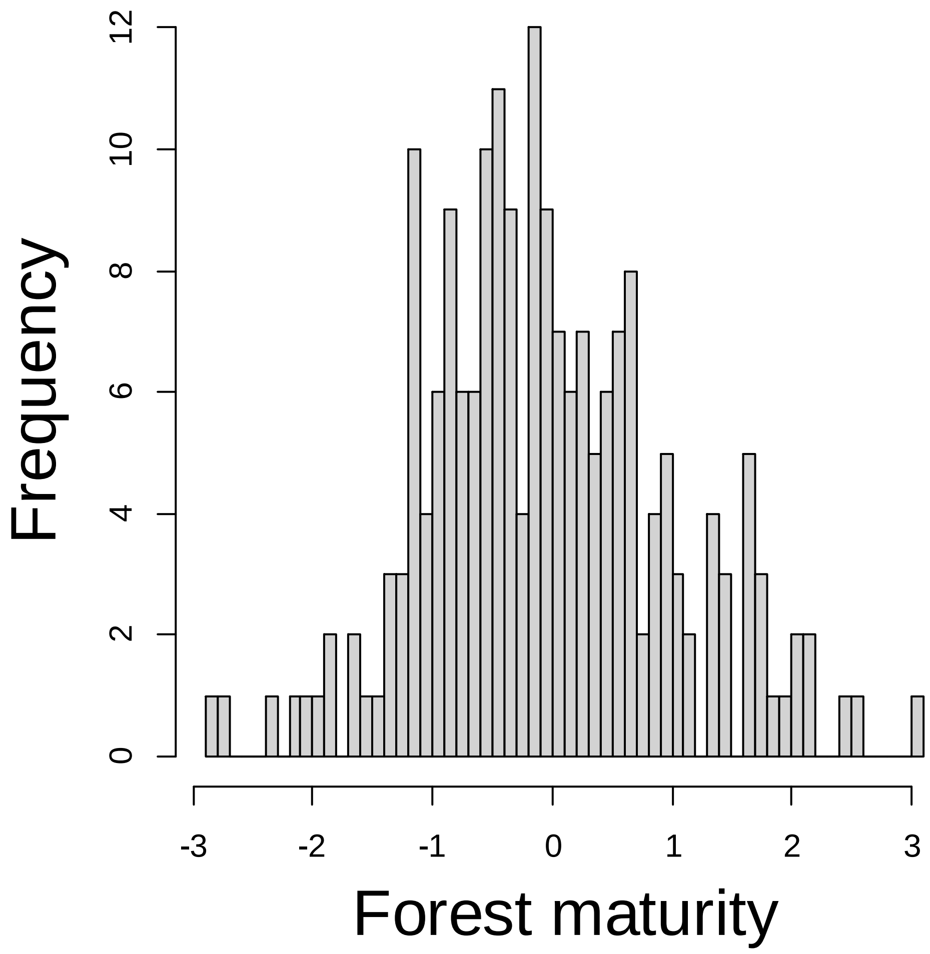


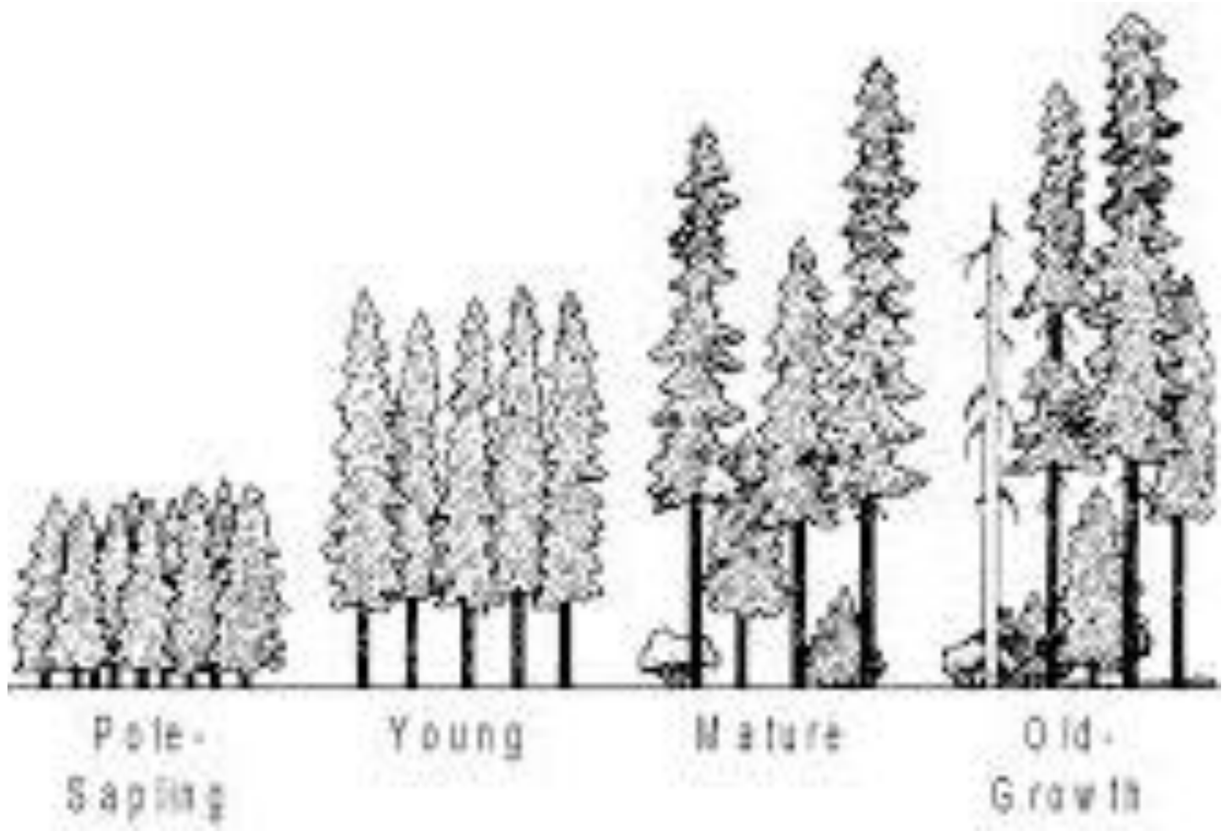
Let’s simulate some data

Data: counts (y) of ‘yellow-footed weeble-wobbles’ at sites with different canopy (c) and sub-canopy (s) heights

Step 1: simulate variation in forest maturity

$$\boldsymbol{m} \sim \text{normal}(0, 1)$$

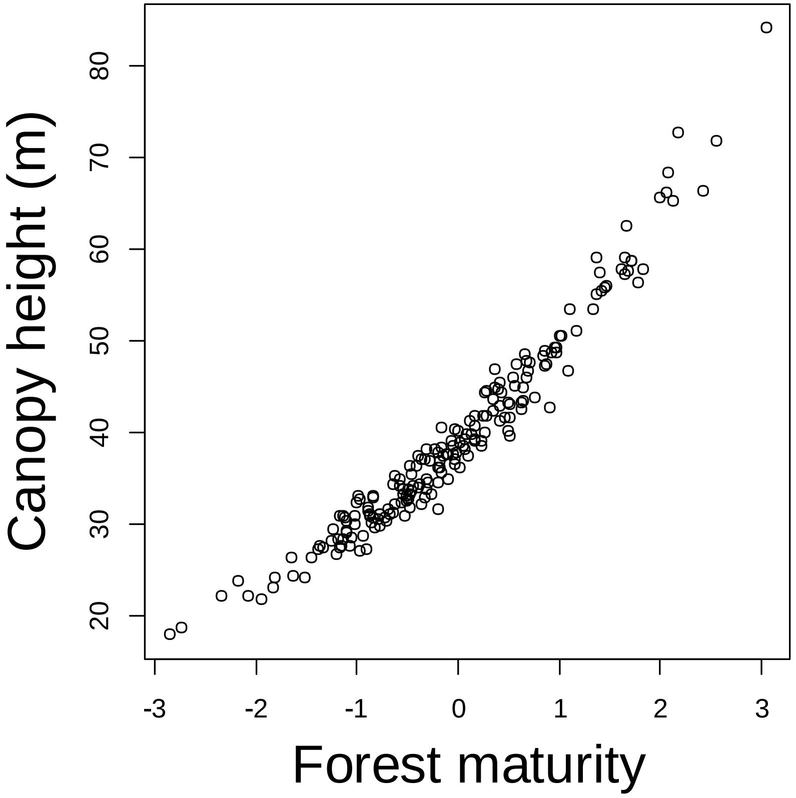


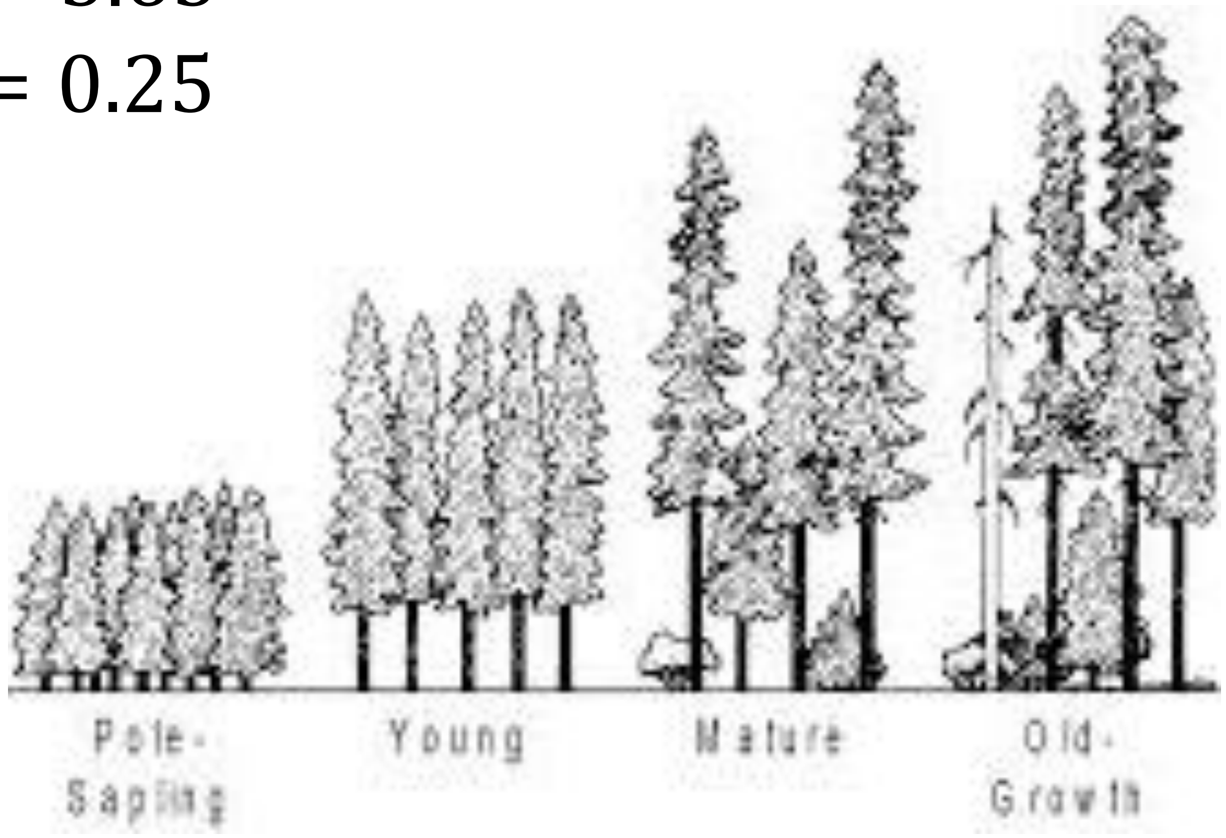


Step 2: simulate variation in canopy height (c)

$$\boldsymbol{c} \sim \mathrm{lognormal}(\alpha_1 + \beta_1 \boldsymbol{m}, \sigma_c = 0.05)$$

$$\alpha_1 = 3.65$$
$$\beta_1 = 0.25$$

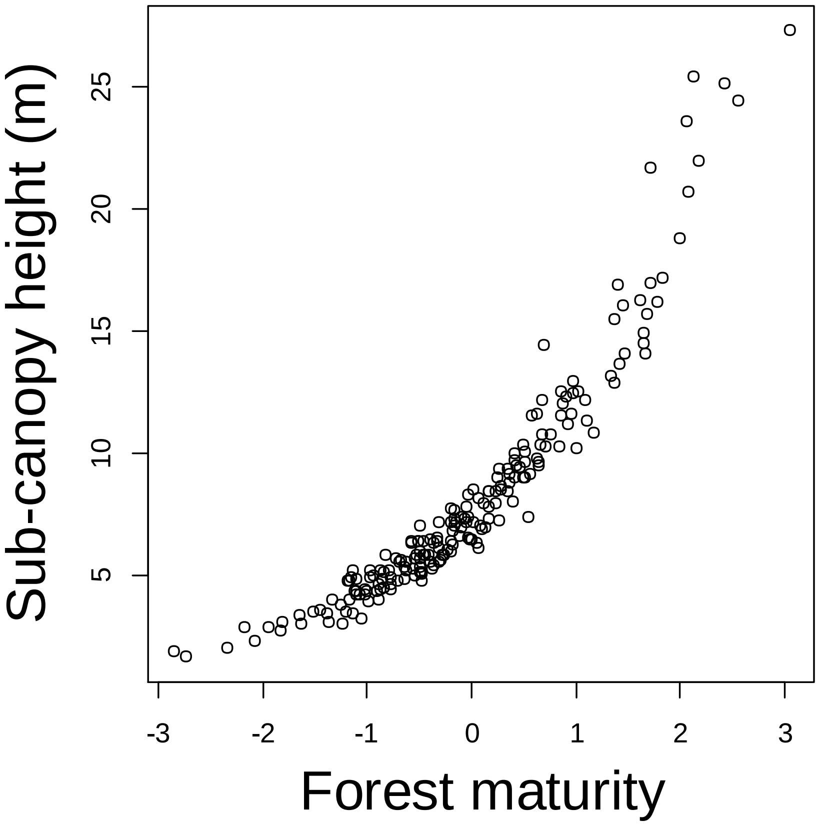


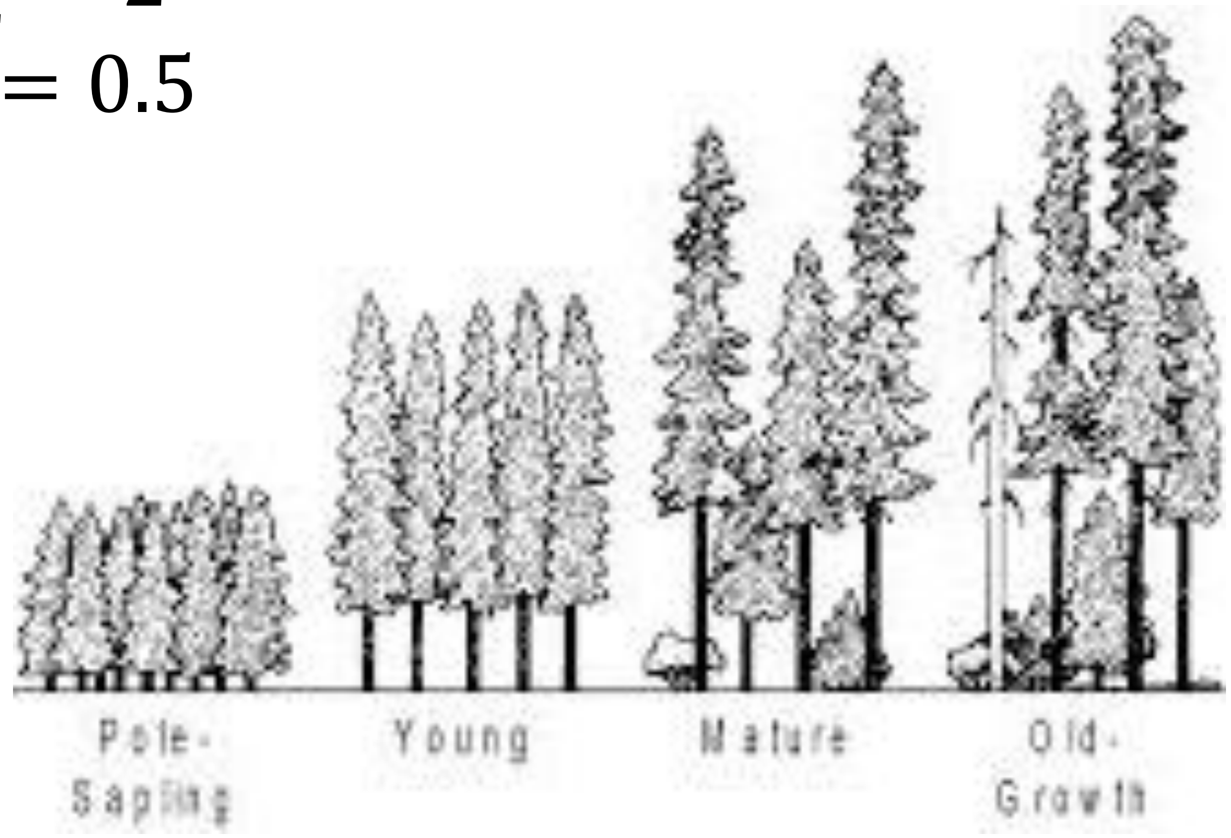


Step 3: simulate variation in sub-canopy height (s)

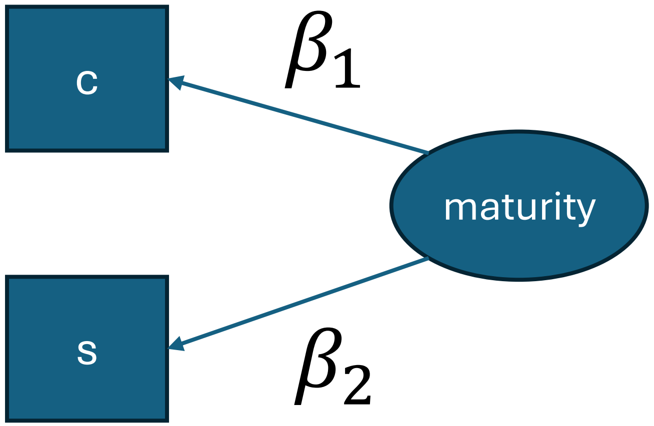
$$\boldsymbol{s} \sim \text{lognormal}(\alpha_2 + \beta_2 \boldsymbol{m}, \sigma_s = 0.05)$$

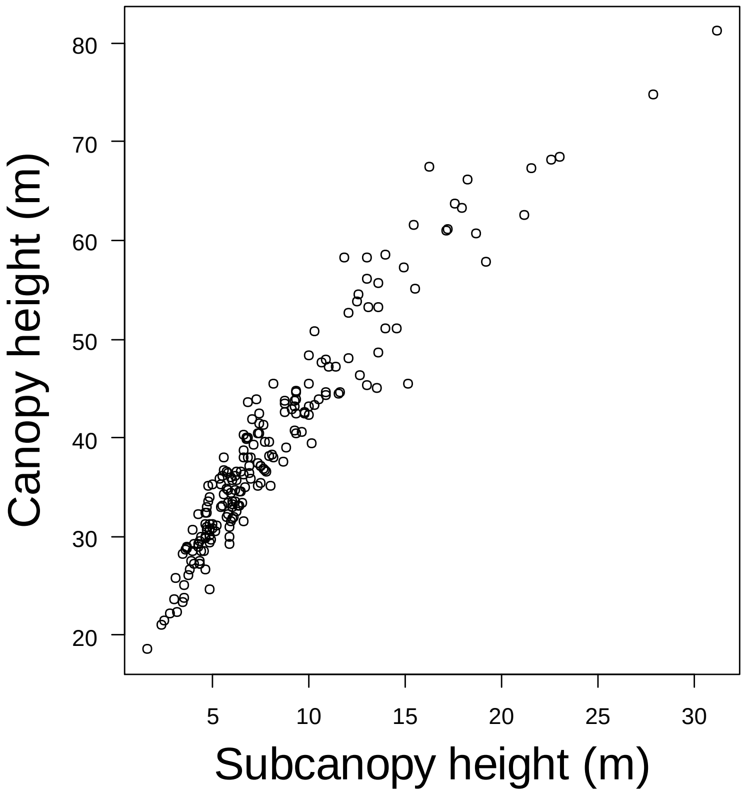
$$\alpha_2 = 2$$
$$\beta_2 = 0.5$$



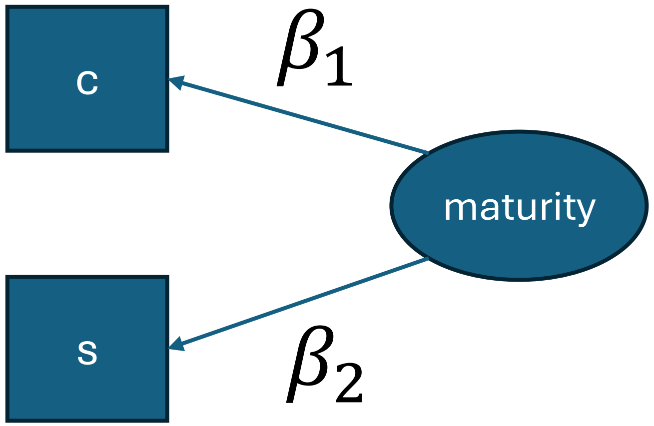


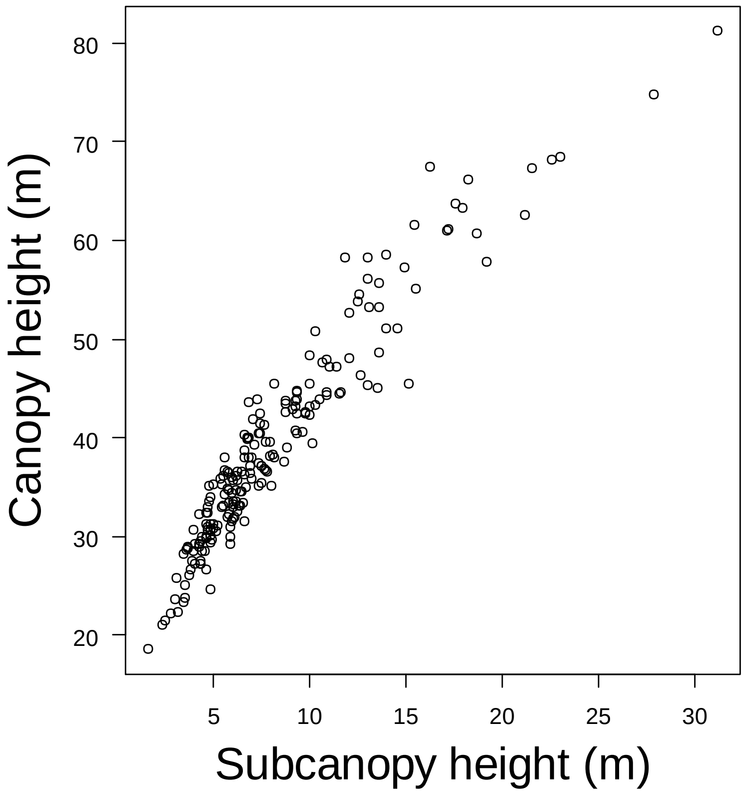
The hypothesis: older forests will have greater canopy heights and greater sub-canopy heights





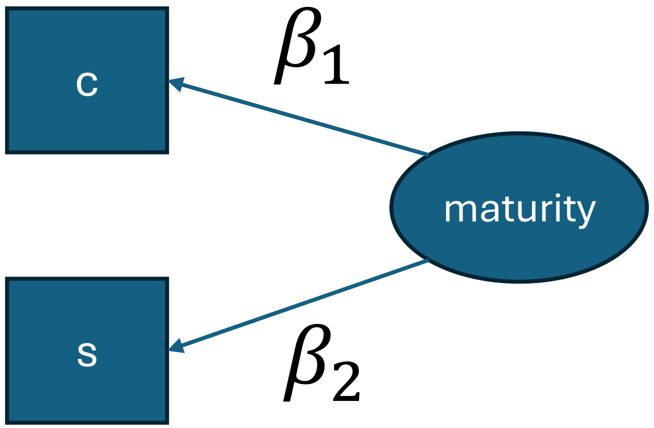
<u>The most important caveat</u>: if things aren't collinear, then you can't assign them to a latent variable





A note on drawing graphs

Squares or rectangles represent measured variables



This oval represents a latent variable

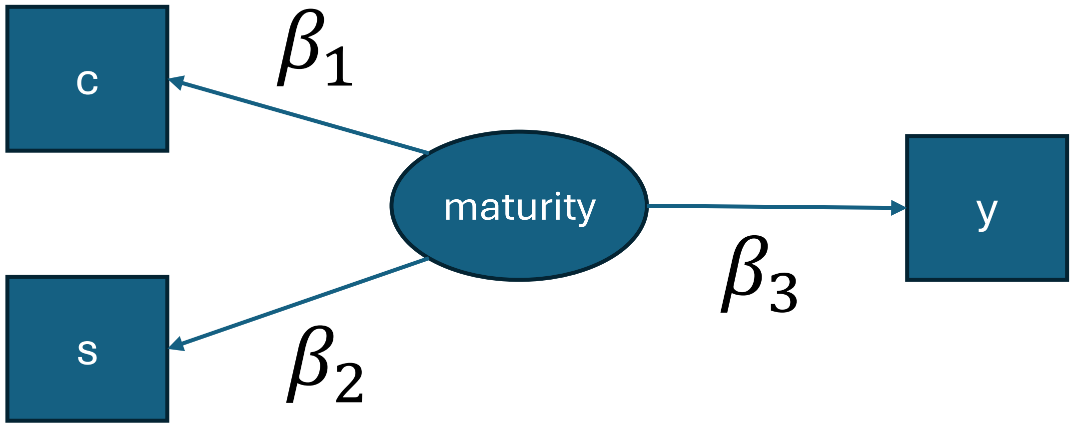
Arrows represent paths (linear models). The direction of the arrow indicates how to parameterize the relationship

Step 4: simulate variation in warbler counts (y)

$$\boldsymbol{y} \sim \text{Poisson}(e^{\alpha_3 + \beta_3 \boldsymbol{m}})$$

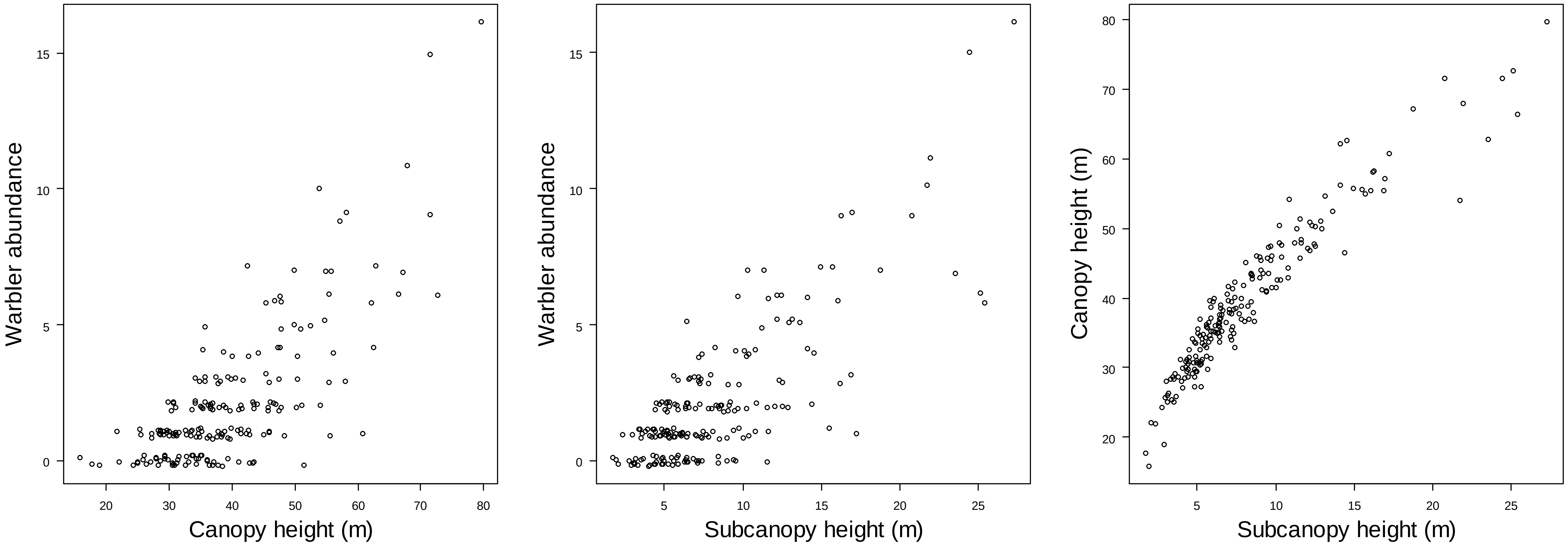
$$\alpha_3 = 0.5$$

$$\beta_3 = 0.75$$



Step 4: simulate variation in warbler counts (y)

$$\boldsymbol{y} \sim \text{Poisson}(e^{\alpha_3 + \beta_3 \boldsymbol{m}})$$



The hypothesis: older forests will have more birds

Our model

$$\boldsymbol{m} \sim \text{normal}(0, \sigma_m^2)$$

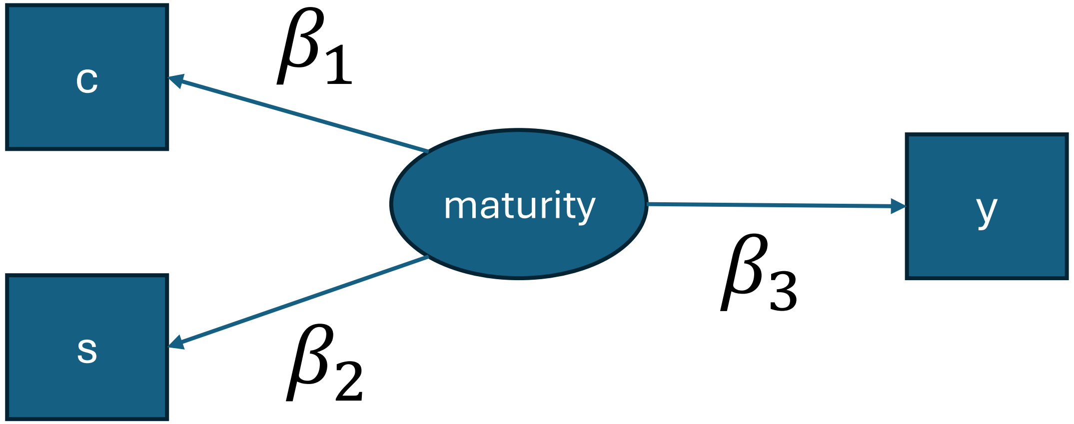
$$\boldsymbol{c} \sim \text{lognormal}(\alpha_1 + \beta_1 \boldsymbol{m}, \sigma_c^2)$$

$$\boldsymbol{s} \sim \text{lognormal}(\alpha_2 + \beta_2 \boldsymbol{m}, \sigma_s^2)$$

$$\boldsymbol{y} \sim \text{Poisson}(e^{\alpha_3 + \beta_3 \boldsymbol{m}})$$

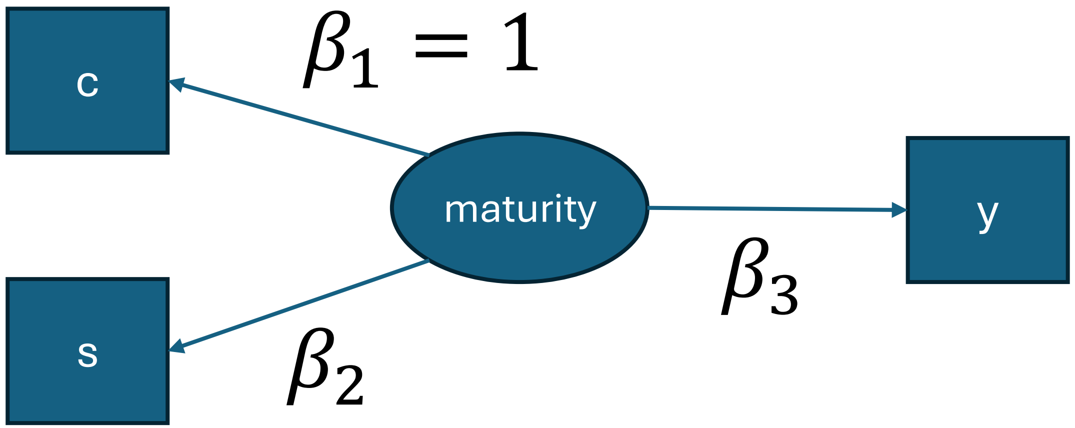
There is one very non-intuitive thing to discuss

We have to fix a ‘loading’ to 1



There is one very non-intuitive thing to discuss

We have to fix a ‘loading’ to 1



Why?!

Well, so the model will be identifiable…

What are the implications of that?

What are the implications of that?

1. The latent variable will be on the same scale as whatever path we fix = 1

What are the implications of that?

1. The latent variable will be on the same scale as whatever path we fix = 1.

2. Our estimates of parameter relationships will be a function of that scale.

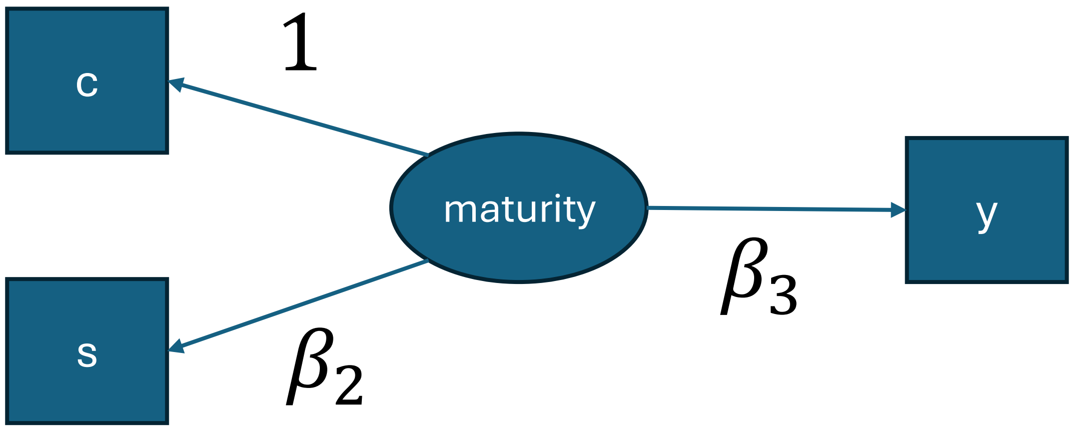
What are the implications of that?

1. The latent variable will be on the same scale as whatever path we fix = 1

2. Our estimates of parameter relationships will be a function of that scale.

3. That’s it. It won’t change our predictions (i.e., warbler counts)

All we’re really assuming when we fix that beta is that there is a positive relationship between our latent variable and the measured variable



Let’s go to the code…