Multicollinearity: underlying causes and resulting problem(s)

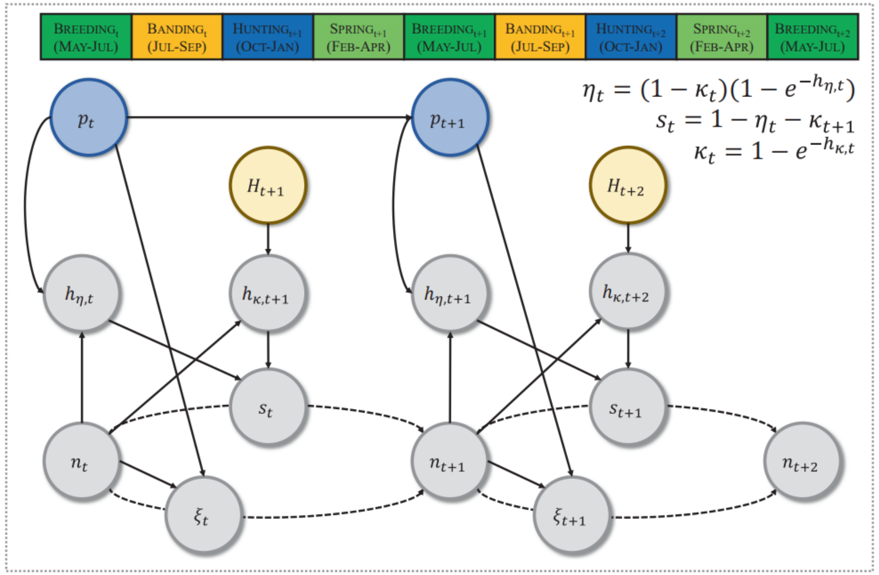
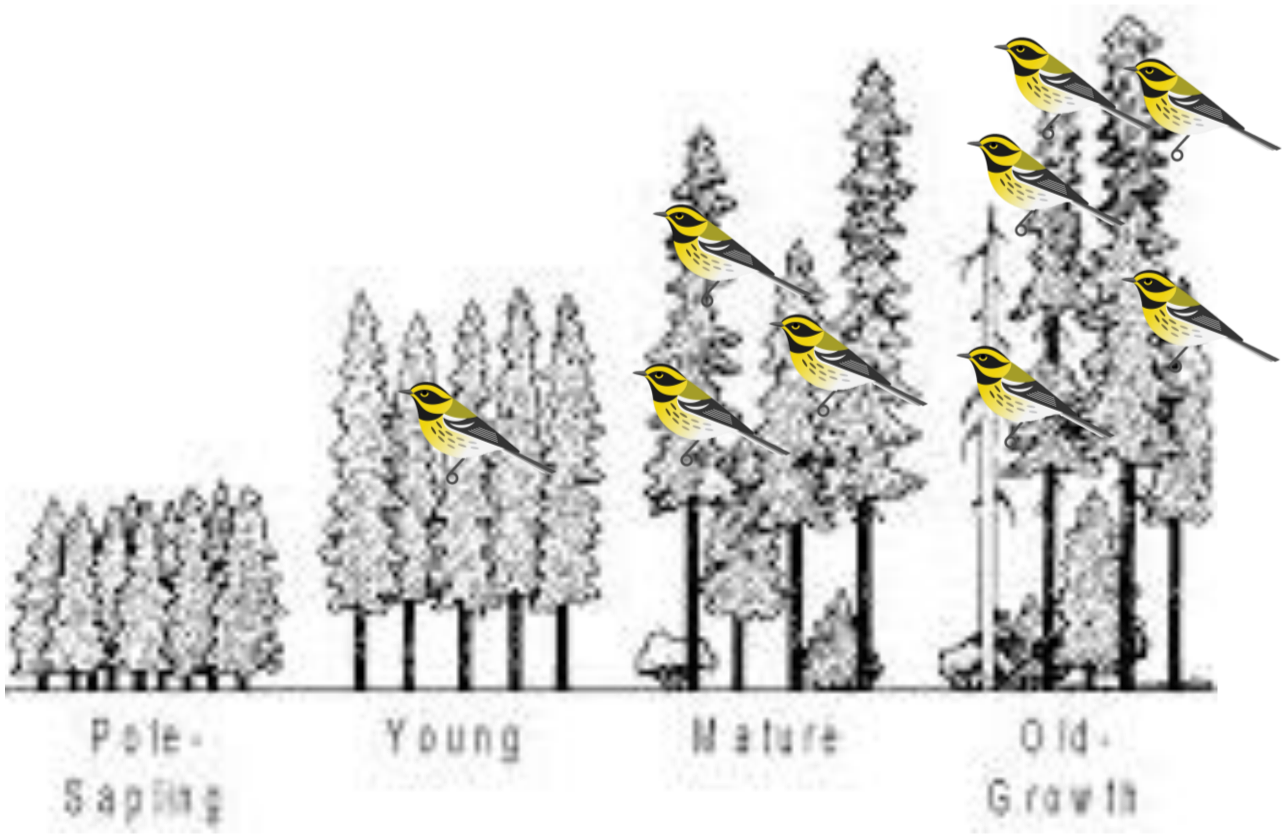


FIGURE 2 A directed acyclic graph demonstrating the relationships among abundance (***n***), ponds (***p***; blue), fecundity (ξ), hunting mortality hazard rate (h_κ), natural mortality hazard rate (h_η), survival (**s**) and the number of duck hunters (***H***; brown) for blue-winged teal breeding in the North American Prairie Pothole Region across the annual cycle (1973–3016). Solid arrows represent estimated directional relationships, and dashed arrows represent processes leading to changes in population abundance.

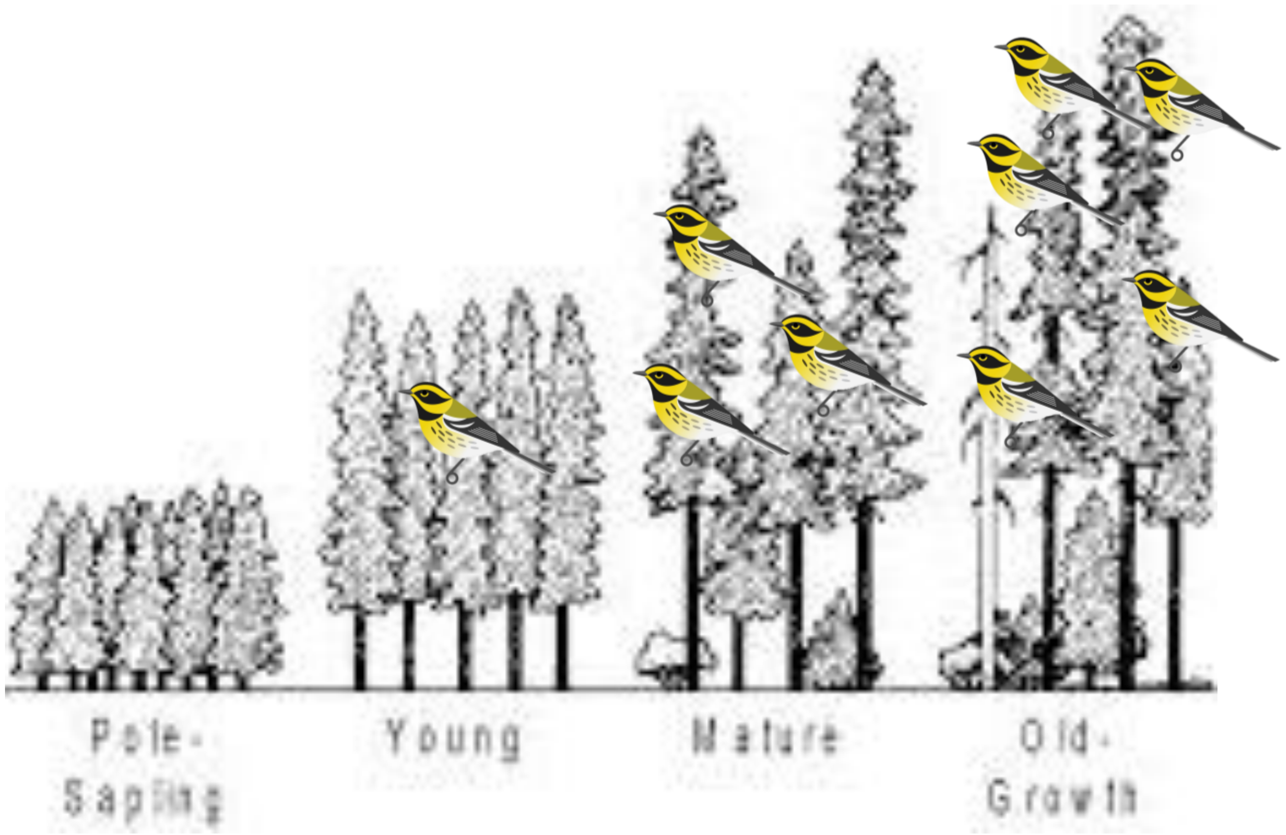
Two simulated datasets

1. Counts of birds in response to forest maturity with two covariates (canopy and sub-canopy height in meters).



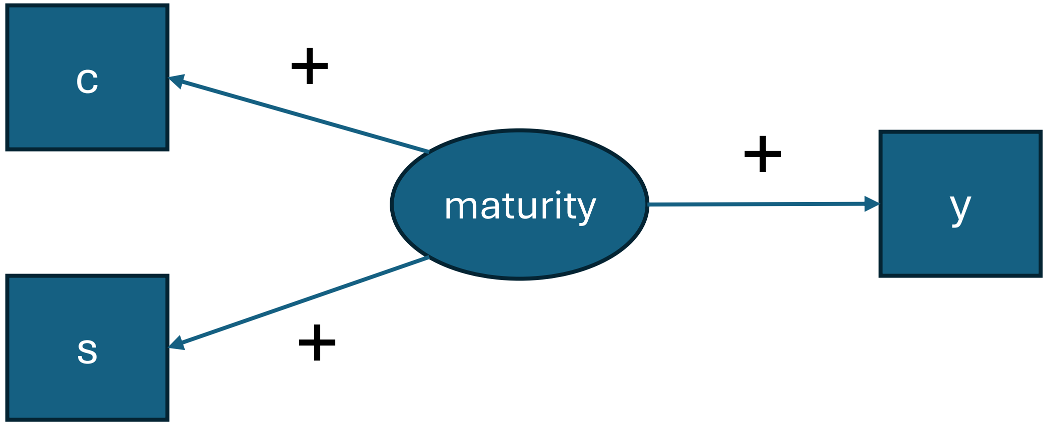
Simulated dataset 1

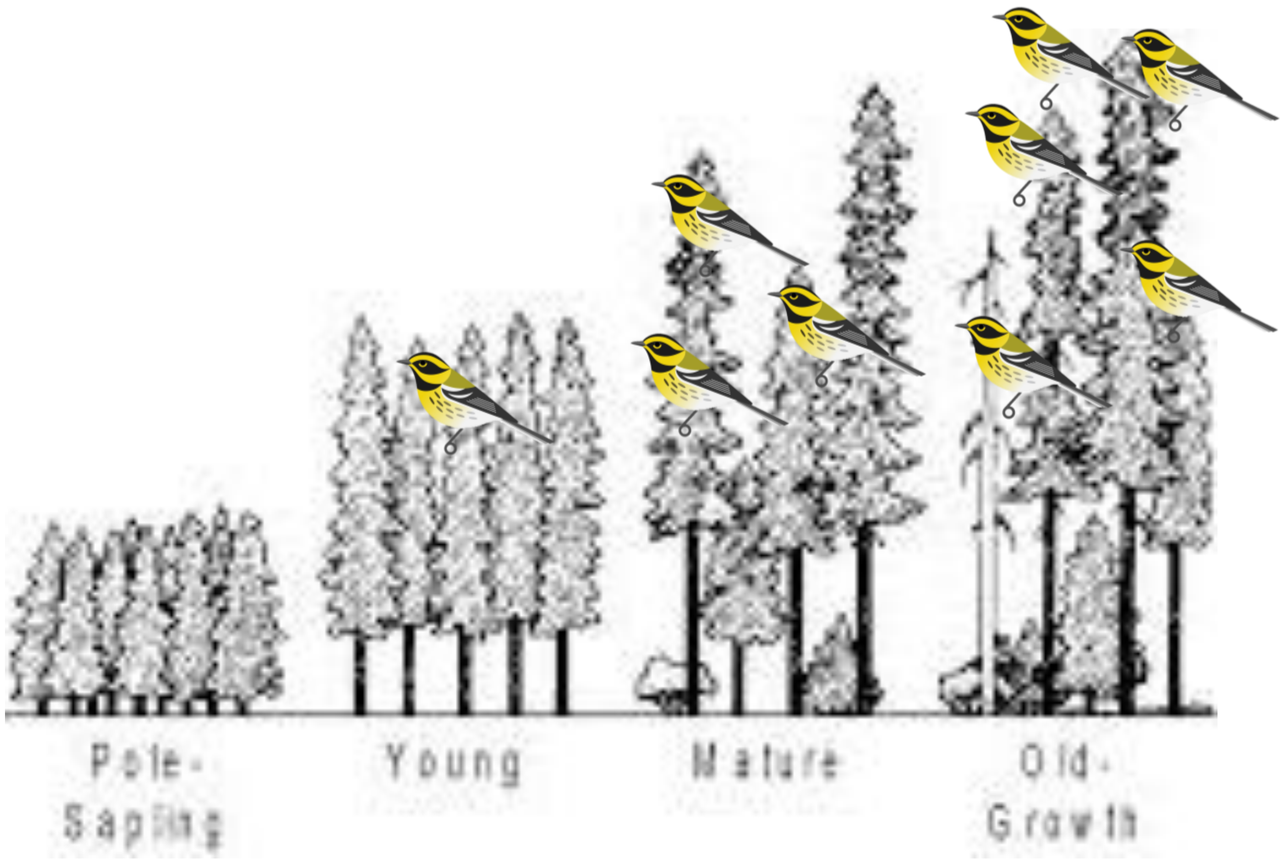
Data: counts (y) of ‘yellow-footed weeble-wobbles’ at sites with different canopy (c) and sub-canopy (s) heights



Simulated dataset 1: both covariates are a function of maturity

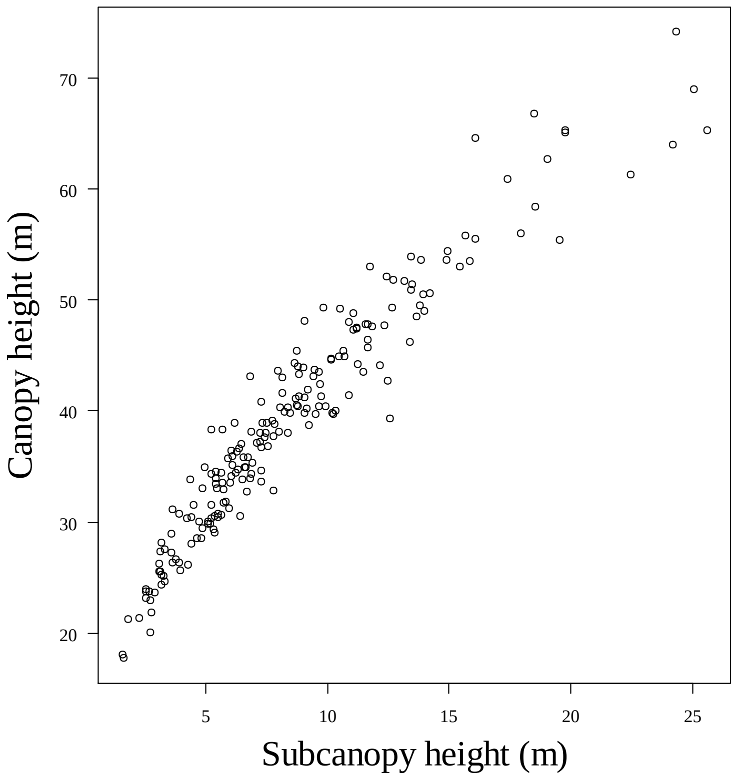
Data: counts (y) of ‘yellow-footed weeble-wobbles’ at sites with different canopy (c) and sub-canopy (s) heights

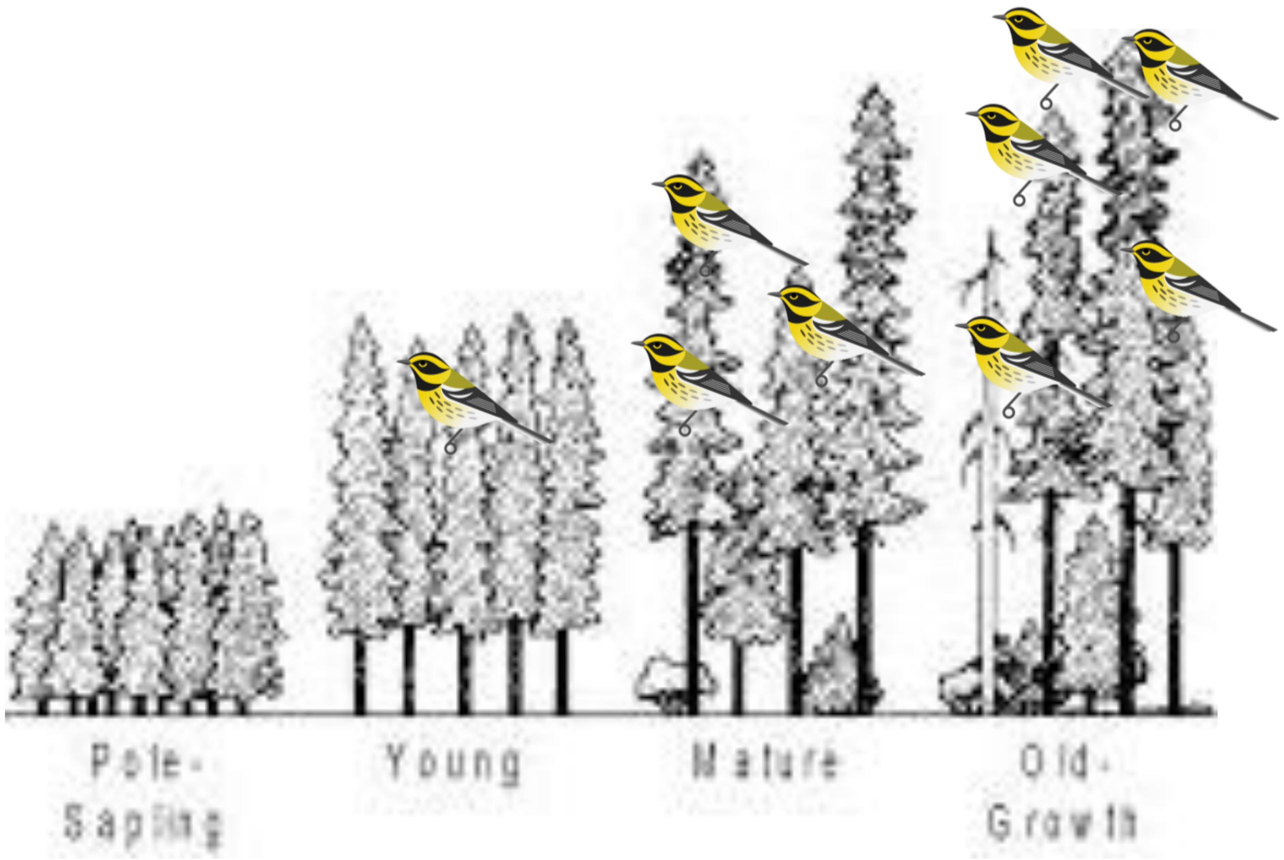




Simulated dataset 1: this induces multicollinearity

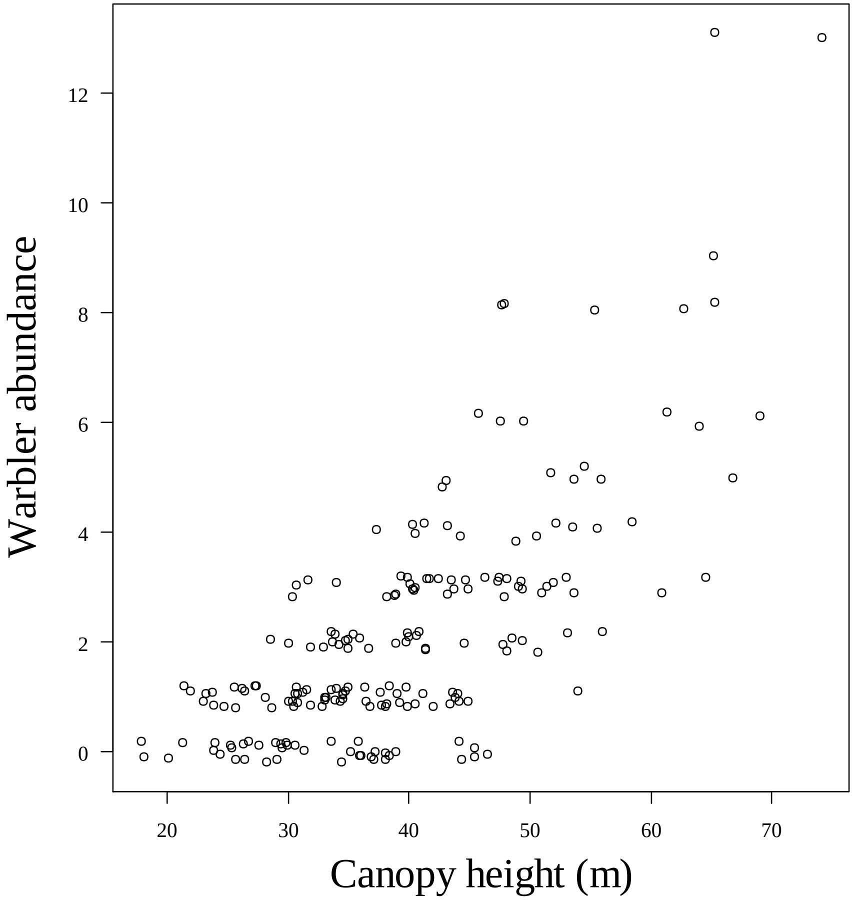
Data: counts (y) of ‘yellow-footed weeble-wobbles’ at sites with different canopy (c) and sub-canopy (s) heights

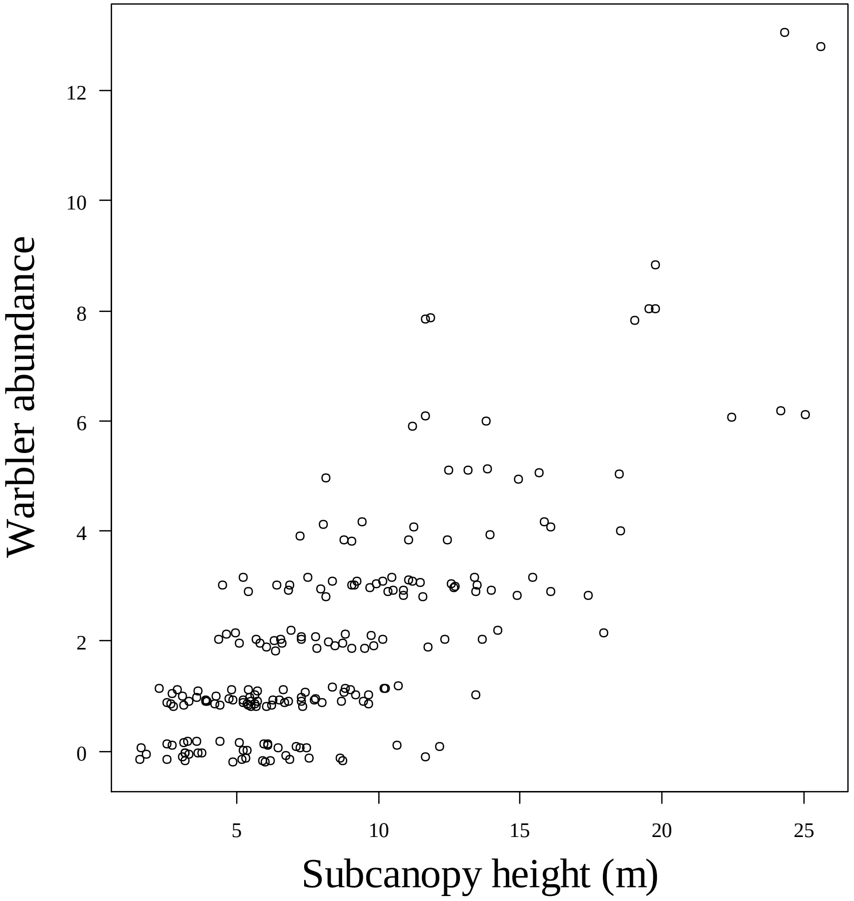




Simulated dataset 1: so, what's the driver?

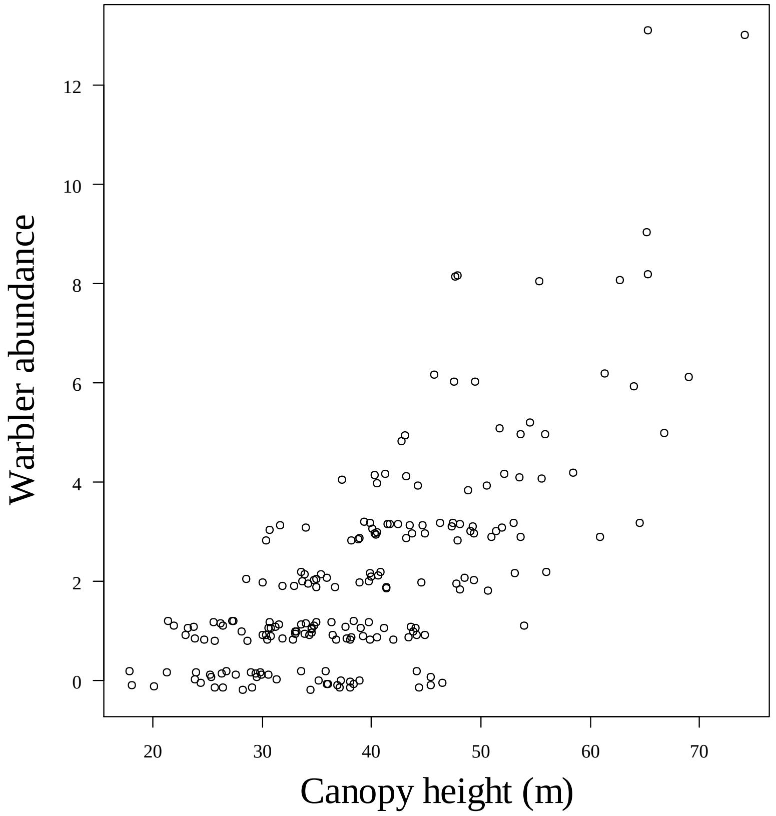
Data: counts (y) of 'yellow-footed weeble-wobbles' at sites with different canopy (c) and sub-canopy (s) heights

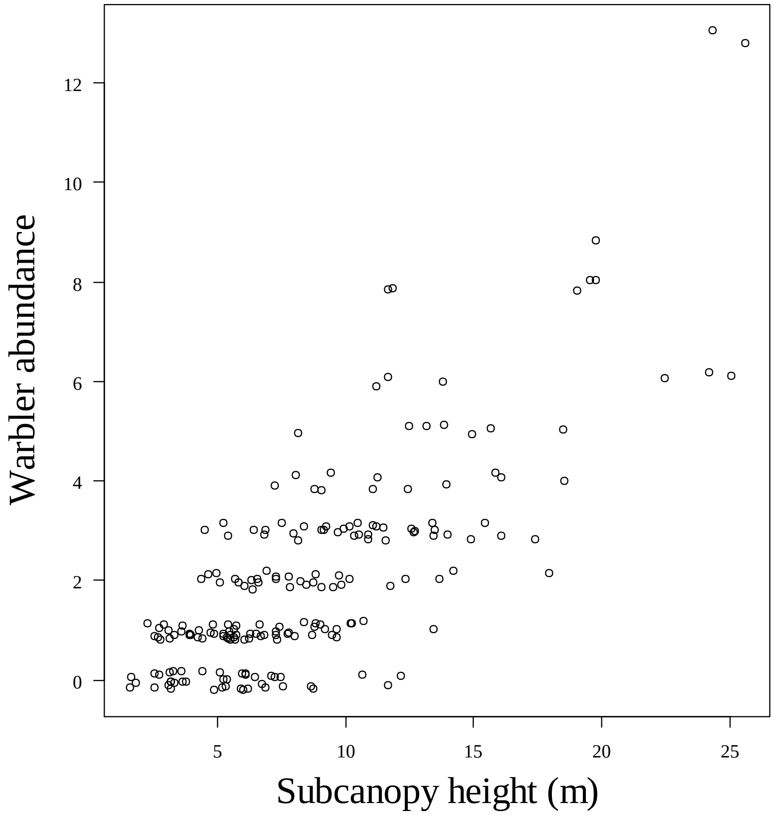




Simulated dataset 1: Our model

$$y_i \sim \mathrm{Poisson}(\alpha + \beta_1 c_i + \beta_2 s_i)$$
$$\alpha \sim normal(1, 0.1)$$
$$\beta_1 \sim normal(0, 0.1)$$
$$\beta_2 \sim normal(0, 0.1)$$

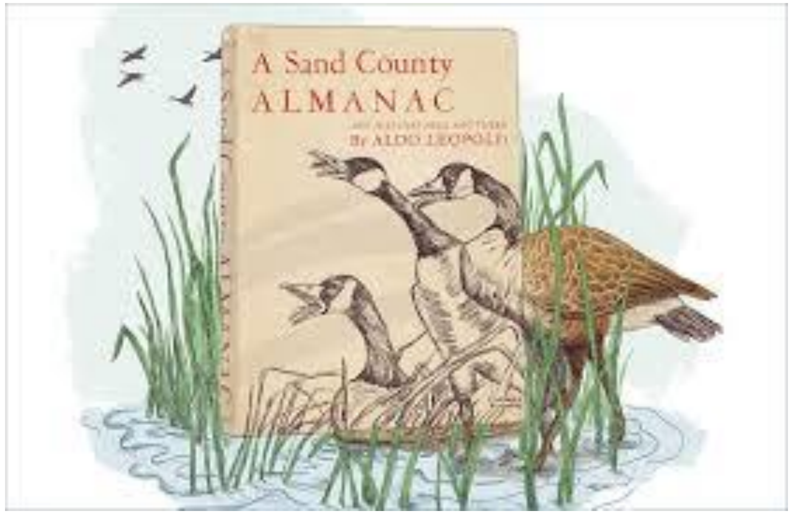


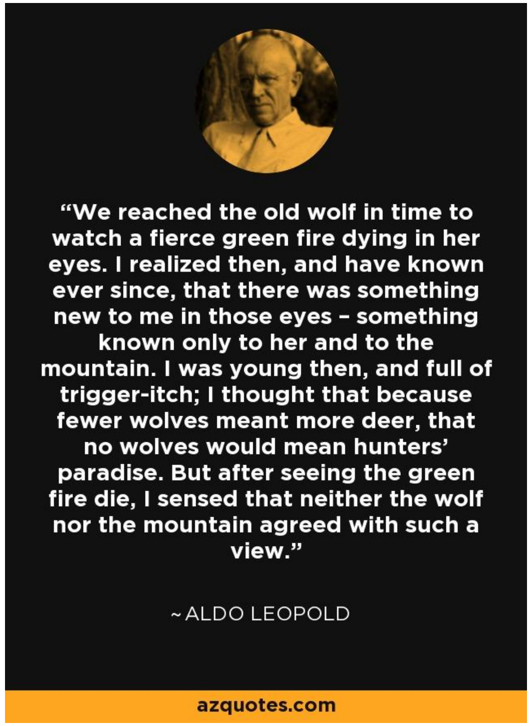


Two simulated datasets

1. Counts of birds in response to forest maturity with two covariates (canopy and sub-canopy height in meters).

2. Available browse (vegetation) as a function of ungulate (deer) and predator (wolf) abundance.

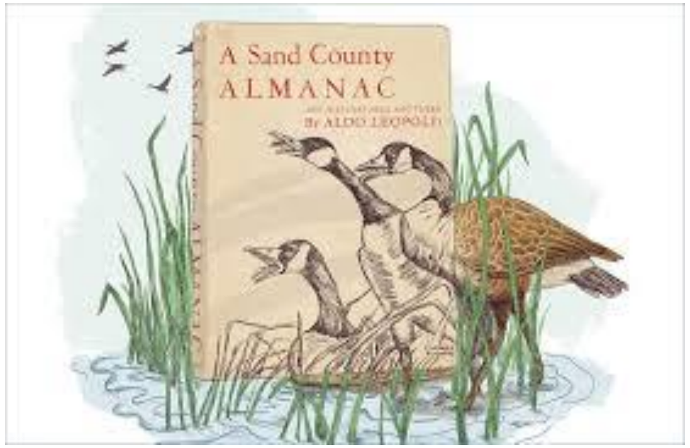




Two simulated datasets

1. Counts of birds in response to forest maturity with two covariates (canopy and sub-canopy height in meters).

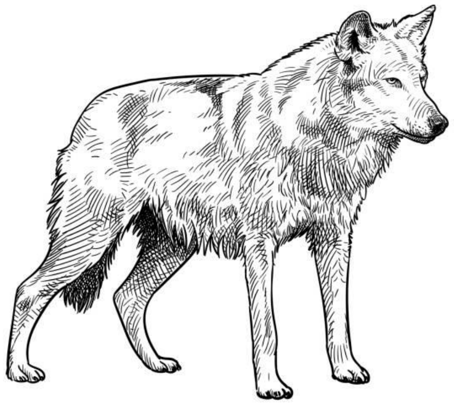
2. Available browse (vegetation) as a function of ungulate (deer) and predator (wolf) abundance.

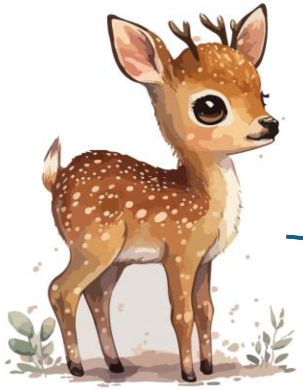


Since then I have lived to see state after state extirpate its wolves. I have watched the face of many a newly wolfless mountain, and seen the south-facing slopes wrinkle with a maze of new deer trails. I have seen every edible bush and seedling browsed, first to anaemic desuetude, and then to death. I have seen every edible tree defoliated to the height of a saddlehorn. Such a mountain looks as if someone had given God a new pruning shears, and forbidden Him all other exercise. In the end the starved bones of the hoped-for deer herd, dead of its own too-much, bleach with the bones of the dead sage, or molder under the high-lined junipers.

Simulated dataset 2

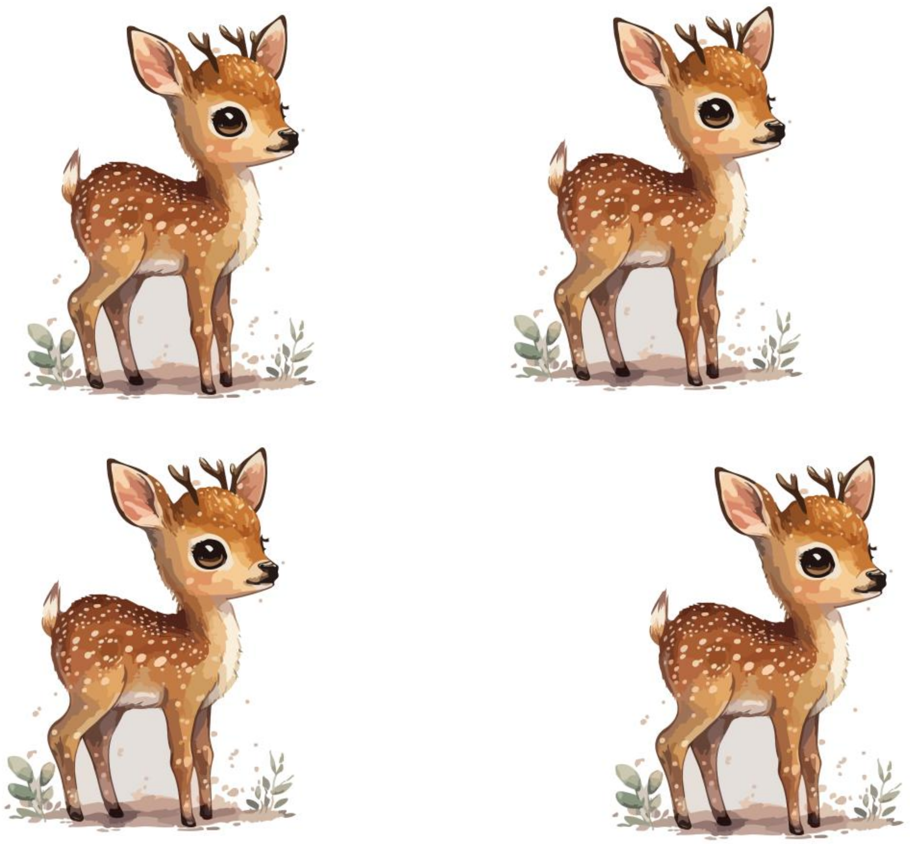
Data: estimates of browse vegetation density (v) as a function of ungulate density (u) and wolf density (w)





Simulated dataset 2

Data: estimates of browse vegetation density (v) as a function of ungulate density (u) and wolf density (w)



Some key take-homes

1. Multicollinearity leads to ‘variance inflation,’ i.e., credible intervals will be extremely broad due to sampling correlation

2. If there are multiple collinear covariates that all affect a response variable, models that only include ONE of those covariates will ‘attribute’ the effect of ALL of those covariates to the single covariate included in the model

3. We can think about modeling this in other ways (and we start next week)!