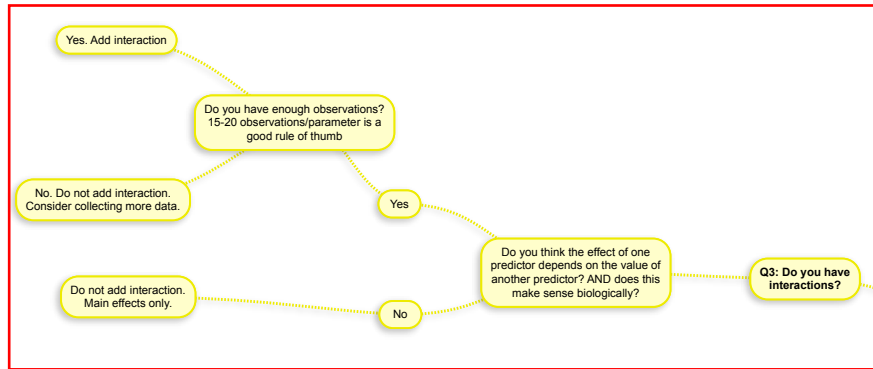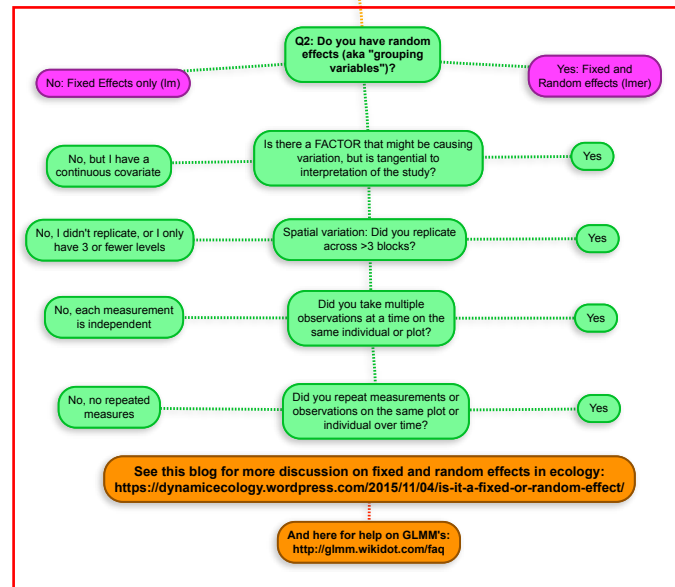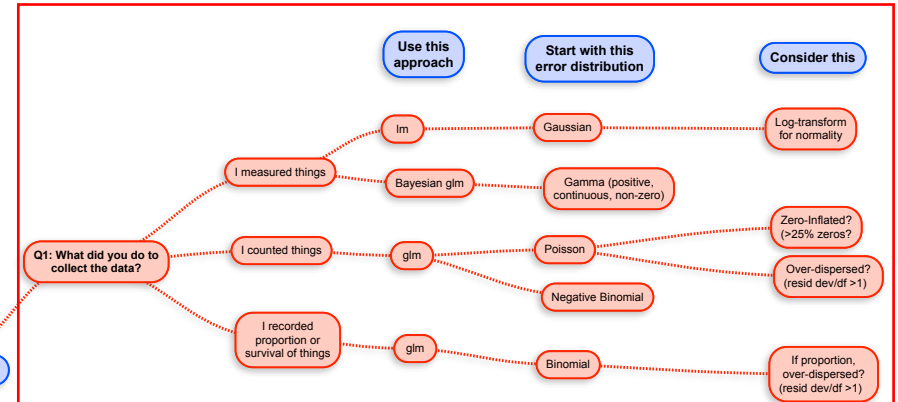# Linear models: selecting the appropriate structure

Notes: Assuming you have a univariate response, a linear relationship, and have explored data for normality, outliers, collinearity, start here to determine your model structure.

Fixed Effects

- Yes. Add interaction
- Do you have enough observations? 15-20 observations/parameter is a good rule of thumb
- No. Do not add interaction. Consider collecting more data.
- Do not add interaction. Main effects only.
- Yes
- No
- Do you think the effect of one predictor depends on the value of another predictor? AND does this make sense biologically?
- **Q3: Do you have interactions?**

Error Distribution

Use this approach	Start with this error distribution	Consider this
lm	Gaussian	Log-transform for normality
Bayesian glm	Gamma (positive, continuous, non-zero)	
glm	Poisson	Zero-Inflated? (>25% zeros?
	Negative Binomial	Over-dispersed? (resid dev/df >1)
glm	Binomial	If proportion, over-dispersed? (resid dev/df >1)

- **Q1: What did you do to collect the data?**
- I measured things
- I counted things
- I recorded proportion or survival of things

What's my model?

Random Effects

- **Q2: Do you have random effects (aka "grouping variables")?**
- No: Fixed Effects only (lm)
- Yes: Fixed and Random effects (lmer)
- No, but I have a continuous covariate
- Is there a FACTOR that might be causing variation, but is tangential to interpretation of the study? — Yes
- No, I didn't replicate, or I only have 3 or fewer levels
- Spatial variation: Did you replicate across >3 blocks? — Yes
- No, each measurement is independent
- Did you take multiple observations at a time on the same individual or plot? — Yes
- No, no repeated measures
- Did you repeat measurements or observations on the same plot or individual over time? — Yes
- See this blog for more discussion on fixed and random effects in ecology: https://dynamicecology.wordpress.com/2015/11/04/is-it-a-fixed-or-random-effect/
- And here for help on GLMM's: http://glmm.wikidot.com/faq