

Student project: overview

Over the course of the quarter, you will learn much of the material by working through examples in the textbook. In addition, you will build towards a complete Bayesian analysis of a dataset, which will serve to synthesize all of this new information.

Your final homework submission will be a pdf of presentation slides with the following components:

- Directed acyclic graph (DAG) that conveys the scientific reasoning for the analysis
- The mathematical expressions for the model (deterministic and stochastic parts)
- Prior predictive simulations
- Traceplots and MCMC checks
- Prior and posterior distributions for one parameter
- Caterpillar plot of the most important parameters
- Plot of posterior predictions
- Your scientific conclusions

You will also submit:

- Data and metadata in .csv format
- Code in .R, .Rmd, .Qmd format

The slides, code, data will be shared with your classmates to facilitate their learning. With your permission, I may also use your example in a future class. For these reasons, you may not want to use your own data for these assignments.

As an alternative, I recommend you find a publication that:

- used a general or generalized linear model with 3 or more predictors
- did provide the data to recreate the analysis
- did not use a Bayesian analysis and/or did not use a DAG to formulate their question (you will do this!)

For example, it would be interesting to apply our Bayesian tools to the kelp dataset [in this paper](#). I'm a little biased, but the dataset in this paper about [coral reefs](#) has potential, and the analysis in this old paper on [temperate reefs](#) could be interrogated from a Bayesian perspective.

You can also use a publicly available dataset; this will work best if you have the appropriate domain knowledge. Here are some ecological examples:

Long-term research on kelp forests:

<https://piscoweb.org/find-pisco-dataone>

Long-term research on coral reefs:

<https://mcr.lternet.edu/>

And the broader LTER network:

<https://lternet.edu/>

The benefit of using a publication is that the context is already provided and the dataset is (close to) ready for analysis; you just need to modify their original analysis (likely fit using maximum likelihood

and / or without explicitly formulating a DAG) into one that reflects our learning. It is important to choose a paper / dataset wisely and I am here to help.