For one of the chapters of my dissertation, I'm modeling the species distributions of shrub species in the Sierra Nevada, CA. I collected presence-absence data in the field for the past two summers at survey plots. Instead of making species distribution models with just this data, I would like to incorporate the abundant presence-only data (e.g. – herbarium records) that are available as publically downloadable datasets.

Last year a paper came out (Fithian *et al.* 2015) that presented a Poisson process model and an R package for combining presence-absence data with presence-only data. While I could just use their R package, I would much rather understand what they're doing (and why) with it, and write my own code using theirs as a basis for mine. All of their code is up on GitHub. Their main functions are "block.glm.fit", "block.projection", "multispeciesPP", "predict.multispeciesPP" – these can all be found in the "R" folder. They also provide a "man" folder with helpful (but only if you've read their paper) descriptions of what the functions do, their arguments, etc.

(Fithian W, Elith J, Hastie T, Keith DA. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol Evol 6: 424-38.)