

Introduction to R

Capstone Collaborative Exercise



Ben R. Fitzpatrick

PhD Candidate, Statistical Science, Mathematical Sciences School, Queensland University of Technology

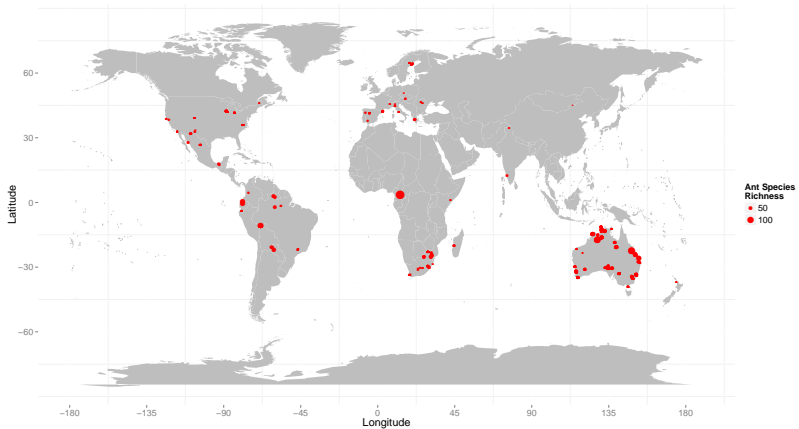
0000-0003-1916-0939

github.com/brfitzpatrick/

@benrfitzpatrick

Modelling Ant Species Richness with Climate & Habitat

Opportunity to practise skills learned in this course



Workflow:

- 1 Summarise & Visualise the Data
- 2 Formulate Analysis Questions
- 3 Choose Model
- 4 Fit Model
- 5 Produce and Interpret Model Diagnostics
- 6 Produce & Interpret Model Summary Statistics
- 7 Predict from Model and Quantify Uncertainty Associated with these Predictions
- 8 Reflect & Revise Analysis Questions
- 9 Potentially, Repeat...

Meet the Data

Ant Species Richness at Sites around the Globe

This exercise uses some publicly available data from the Dryad Repository. Please download a copy of 'Data appendix.xlsx' from:

<http://dx.doi.org/10.5061/dryad.r36n0>

These data were published as part of:

Gibb H, Sanders NJ, Dunn RR, Watson S, Photakis M, Abril S, Andersen AN, Angulo E, Armbrrecht I, Arnan X, Baccaro FB, Bishop TR, Boulay R, Castracani C, Del Toro I, Delsinne T, Diaz M, Donoso DA, Enrquez ML, Fayle TM, Feener DH, Fitzpatrick MC, Gmez C, Grasso DA, Groc S, Heterick B, Hoffmann BD, Lach L, Lattke J, Leponce M, Lessard J, Longino J, Lucky A, Majer J, Menke SB, Mezger D, Mori A, Munyai TC, Paknia O, Pearce-Duvet J, Pfeiffer M, Philpott SM, de Souza JLP, Tista M, Vasconcelos HL, Vonshak M, Parr CL (2015) *Climate mediates the effects of disturbance on ant assemblage structure*. **Proceedings of the Royal Society B** 282(1808).

For today let's focus on Species Richness as a response variable and ignore the hierarchy in the data (as coded by the 'cluster' variable)

If we do this we can consider a preliminary analysis with Multiple Linear Regression or Generalized Linear Models.

This is OK for practise but if we were seriously attempting this analysis for publication etc. we'd need to consider incorporating random effects for cluster or modelling the cluster means...if you feel confident to try one of these options please do...

Lat.	Long.	Mean annual temp.	Total annual precip.	Temp. range	Disturb.	Hemi	Cont.	Pitfall days	Trans. length	Species richness
-22.5	148.3	21.90	615	25.40	Undist.	South	Oceania	450	40	75
-22.5	148.3	21.90	615	25.40	Dist.	South	Oceania	300	40	55
-22.5	148.3	21.90	615	25.40	Dist.	South	Oceania	300	40	50
-22.5	148.3	21.90	615	25.40	Dist.	South	Oceania	300	40	38
-22.5	148.3	21.90	615	25.40	Dist.	South	Oceania	300	40	46
-22.5	148.3	21.90	615	25.40	Undist.	South	Oceania	450	40	87

Some Ideas...

Article title:

Climate mediates the effects of disturbance on ant assemblage structure.

suggests interaction terms may be important here

- as variables are on different scales it will be advisable to recenter and rescale them all to a common mean and magnitude (0, 1) is the traditional choice

Species Richness is not continuous (it's non-negative integers) so:

- a Generalized Linear Model would be a good idea
- however there is quite a range in the response so Multiple Linear Regression will be OK here for our purposes if you'd like to keep it simple

Options for Extension

Model for Probability of Interspecific Encounter (PIE)

PIE is bounded between 0 and 1 (it's a probability!)

- so a logit transformation would be necessary in order to model it with a linear model...

Accounting for Hierarchy in Data

'Because sites were spatially clustered, we used mixed-effects models, with clusters of sites separated by no more than 100 km from each other represented by a single random effect to control for potential autocorrelation between localized sites' - Gib et al.

- Simpler option: calculate cluster means and model those...
- Better option: use linear mixed effects models as Git et al. did

Over to you...

- Please form small groups
- Together I'd like you to collaboratively plan and conduct an exploratory analysis of these data
- Please write neat, well annotated code and share this code with your group members via a GitHub repository
- prior to the end of the course each group will present their methods and code along with their results and interpretation

Some possible Questions

- 1 Is climate a useful predictor of the variation in ant species richness?
- 2 Is habitat disturbance a useful predictor of the variation in ant species richness?
- 3 Does habitat disturbance influence the nature of the correlation between climate and ant species richness?
- 4 Are there non-linear correlations between ant species richness and the continuous covariates?
- 5 Does habitat disturbance influence the nature of the correlation between climate and ant species richness after all other covariate effects have been accounted for?
- 6 Do models that incorporate non-linear effects and interactions out perform model that include only linear effects?
- 7 Can you use a variable selection technique to build a parsimonious model from the full set of linear, non-linear and interaction terms for the covariates included in these data?

A quick note on reading the data into R

The data file available on the Dryad Repository, 'Data appendix.xlsx', has two columns we don't really need.

When using your Spread Sheet program (Excel, Calc, etc.) to convert the .xlsx file to a .csv feel please delete the columns:

``Locality_ID'`

and

``Source'`

for a less text heavy dataframe to read into R.

A code file containing annotated code to complete a multiple linear regression based analysis of the potential relationships between ant species richness, climate and habitat disturbance is available at:

```
https://github.com/brfitzpatrick/Intro\_to\_R/blob/  
master/Collaborative\_Exercise\_Solutions.R
```