

Species-area relationships of Georgia mammals

John M. Drake & Andrew Park

One of the key concepts in biodiversity measurement is the *species area curve*. Particularly, it has been found that species richness almost universally scales with area according to the formula

$$s = ca^z,$$

where s is the number of species, a is the area of interest, and c and z are fit coefficients. Often, the coefficient z is used to explore the biogeographic effects of different environmental variables.

In this exercise, we will explore species area relationships for mammals in Georgia counties. A table, containing presence or absence of mammal species has been provided in the file `ga-mammals.csv`. Some additional covariates (area in km², population size from the 1990 census, and physiographic region with Appalachian Plateau, Valley & Ridge, and Blue Ridge regions combined) are included in the file `covariates.csv`. Each county is also listed with its name and a numerical identifier. The locations of these counties are shown in Figure below.

This exercise will be performed in R. Before you begin, create a directory and an empty markdown document `mammal-species-area.Rmd` to store your work. Place the file `ga-mammals.csv` and `covariates.csv` in your directory, open R, and set this directory as your working directory. To load the data in your workspace, run the following lines.

```
mammals <- read.csv('ga-mammals.csv')
covariates <- read.csv('covariates.csv')
```

Now you are ready to begin the exercise.

1. You can inspect the data by typing `head(mammals)` or `mammals` at the command prompt. To calculate the number of species in each county, sum across columns containing the indicators for presence or absence of species, i.e., `richness <- rowSums(mammals[,3:80])`. To look at the distribution of species richness across counties, plot a histogram (Hint: use the function `hist`).
2. Treat each county as a separate sampling unit. Plot a species area curve. Note that counties in the covariate data file are ordered exactly as in the species composition data file. Next, plot on logarithmic scales. Does the plot look linear? Use the function `lm` to fit a linear regression line to the log-transformed data. What are your estimates of c and z ?
3. There is a lot of variability in this relationship. Possibly this is because data from several physiographical regions have been combined. After all, there is a clear geographical gradient in species richness at this granularity. Explore the effect of these regions by fitting separate models for each. Is the species area relationship more clear at this scale? Are these models better than the model with all counties combined? (Hint: the model returned by `lm` contains a quantity called the *coefficient of determination*. This can be used to quantify the goodness-of-fit of the model.)
4. Another possible explanation for this variability is that most counties in Georgia are quite small. Accordingly, we have looked at a very small range of areas over which it may be difficult to discern a pattern. To explore a larger range of areas, aggregate some counties together in a “nested design” and refit the species area curve. Note that this will require some manipulation of the data. You can’t simply add county species richness together. (Why not?) Now what pattern do you see? Does it make a difference whether the counties you aggregate are adjacent or not? Do this for all counties together and separately for the different regions.

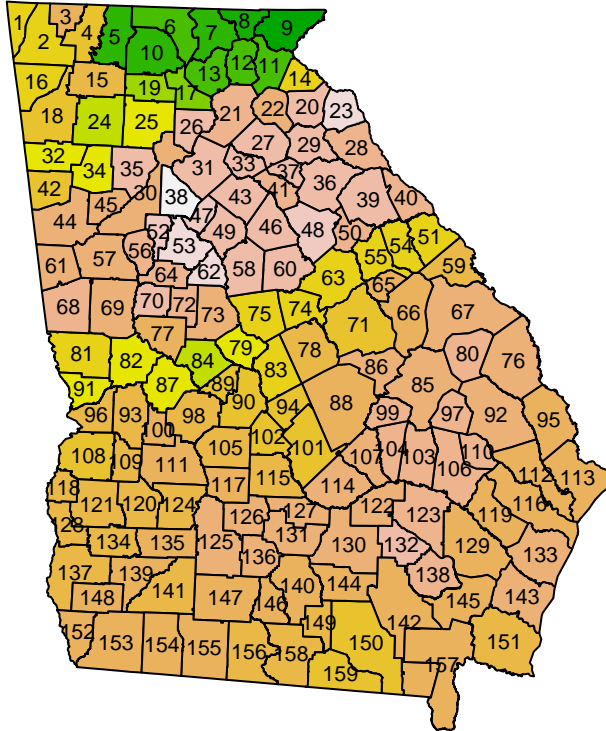


Figure 1: Mammal species richness in Georgia ranges from 41 (white) to 59 (green). Numbers coincide with county designations in the file ga-mammals.csv

5. **Bonus.** Finally, what in addition to area might account for the variation in species richness at the county level? Develop one or more hypotheses. See if you can find some data that may help to test your hypotheses. (The websites <http://georgiainfo.galileo.usg.edu/county.htm> and <http://georgiastats.uga.edu/> may be of use.) Enter these data and combine them with your previous analysis.