GIS Basics, Experimental Design and Getting your data into shape

A. Michelle Lawing
Ecosystem Science and Management
Texas A&M University
College Station, TX 77843
alawing@tamu.edu

michellelawing.info/Reshape.R
Schedule for today

• My presentation
  – The basics of a GIS
  – Experimental Design
  – The reshape package

• GIS in R and Contingency Analysis
• GIS in R and Contingency Analysis Demonstration

• BREAK
• GIS in R and Contingency Analysis Tutorial
GIS Basics
What is GIS?

• Geographic Information System

• Geographic Information Science

• Geographic Information Services

• Basically, it is a collection of data and tools used to visualize and analyze data in a spatially explicit manner
What is a GIS?

Database
The components of GIS

- Spatial data
- Software
- Hardware
- User (you)
Creating a GIS

- Makes you think about spatial relationships
- Helps you to formulate hypotheses to test
- Helps identify potential spatial issues (and biases)
Spatial autocorrelation

- Similar things are near, different things are far
- Violates assumptions of independence
- Methods to deal with this (Moran’s I, Mantel test, covariate)
Spatial autocorrelation

- Moran’s I, parametric test, correlation weighted by inverse distances, effect size
- Mantel test, semi-parametric test, correlation between two distance matrices, effect size
Spatial autocorrelation

- include a covariate in your model (temperature)
- draw a variogram (demonstrate the degree and range of sa in your data)
variogram
variogram

\[ \gamma(h) \]

sill (C)

distance (h)

range of influence (a)

spherical model

exponential model
adjust for or predict with spatial autocorrelation

• adjust
  – factor out the variation
  – use residuals

• predict (interpolate)
  – neighborhood
  – bilinear
  – krige (uses semivariance)
GIS calculations

• measure things that would otherwise be difficult and time consuming

• How much of your study area consists of a specific habitat type?

• How much is under 10m water depth?

• How much has a slope greater than 5 degrees?
Other calculations...

- size of the home range of an individual
- total area occupied by a specific species
- how long are your survey tracks
- how much survey effort was put in to different parts of your study area
make new variables

• slope and aspect of terrain from elevation info

• slope and aspect of a seabed from water depth info

• distance between things (distance to nearest water source)

• community- (or assemblage-) level metrics

• instantaneous velocity of a variable (change over XX)
linking data together

• statistical analysis

• single table (attributes table)

• each row is a sample
viewing data on maps

- traits
- distances
- effects
- variance
- multiple variables (size and color)
experimental design

• deciding where to collect data

• coverage (breadth)

• variation (depth)

• random sampling points (unbiased sampling)
experimental design

• must be at least two treatments or treatment and control

• observational studies, nature does the assigning of treatments

• research has no control over which subject gets which treatment (random effects)
replication

- assign each treatment to multiple, independent experimental units

- larger sample size leads to higher probability of getting the correct answer

- larger sample size means more information, more information means better estimates
 replication
pseudoreplication

• Visscher et al. (1996) compared the effects of two methods of removing a barbed honeybee stinger from a victim
• pinching or scraping with credit card
• n = 40 stings, half removed with each method
• data were the size of the welt after 10 minutes
• pinching led to smaller welt, but not significantly smaller
• all 40 measurements came from TWO volunteers
balance

• all treatments have the same sample size

• smallest sampling error occurs on balanced designs (mathematical artifact)

• ANOVA is robust to departures from equal variance IF balanced

• balance, though is not as important as replication
blocking

- grouping experimental units
- within each unit, treatments are randomly assigned
- difference between treatments are evaluated within blocks, difference between blocks is discarded
blocking: paired vs. unpaired

• consider a two-treatment experiment to estimate the effect of clear cutting on salamander density
blocking: paired vs. unpaired

• paired design, adjacent plots are not independent
• account for non independence with block
• more powerful
randomized complete block (rcb)

- paired is a special case of rcb
- any number of treatments repeated in each block
factor

- a factor is a single treatment variable whose effect is of interest to the researcher

- a factorial design, every possible combination of factors in investigated
observational studies

• combine as much good experimental design as possible

• can not reproduce random assignment
  – use matching (such as paired design)
  – control for known confounding variables (analysis of covariance)
planning your sample size
reshape
load library

> library(reshape)
look at data

```r
> head(iris)
   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1         5.1         3.5          1.4       0.2   setosa
2         4.9         3.0          1.4       0.2   setosa
3         4.7         3.2          1.3       0.2   setosa
4         4.6         3.1          1.5       0.2   setosa
5         5.0         3.6          1.4       0.2   setosa
6         5.4         3.9          1.7       0.4   setosa
```
transpose

• built-in functionality
• t()

> t(iris)[,1:5]

| Sepal.Length | "5.1" | "4.9" | "4.7" | "4.6" | "5.0" |
| Sepal.Width   | "3.5" | "3.0" | "3.2" | "3.1" | "3.6" |
| Petal.Length  | "1.4" | "1.4" | "1.3" | "1.5" | "1.4" |
| Petal.Width   | "0.2" | "0.2" | "0.2" | "0.2" | "0.2" |
| Species       | "setosa" | "setosa" | "setosa" | "setosa" | "setosa" |
```r
> head(iris_m <- melt(iris, id = 5))

<table>
<thead>
<tr>
<th>Species</th>
<th>variable</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>5.1</td>
</tr>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>4.9</td>
</tr>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>4.7</td>
</tr>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>4.6</td>
</tr>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>5.0</td>
</tr>
<tr>
<td>setosa</td>
<td>Sepal.Length</td>
<td>5.4</td>
</tr>
</tbody>
</table>
```
cast

# use cast to look for sample sizes (balance)

> cast(iris_m, Species ~ variable, length)

<table>
<thead>
<tr>
<th>Species</th>
<th>Sepal.Length</th>
<th>Sepal.Width</th>
<th>Petal.Length</th>
<th>Petal.Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>setosa</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>versicolor</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>virginica</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
cast

# use cast to look for ranges

```r
> cast(iris_m, Species ~ variable, range)

       Species Sepal.Length_X1 Sepal.Length_X2 Sepal.Width_X1 Sepal.Width_X2
1     setosa         4.3           5.8          2.3            4.4
2 versicolor         4.9           7.0          2.0            3.4
3   virginica         4.9           7.9          2.2            3.8

       Petal.Length_X1 Petal.Length_X2 Petal.Width_X1 Petal.Width_X2
1            1.0         1.9           0.1            0.6
2            3.0         5.1           1.0            1.8
3            4.5         6.9           1.4            2.5
```
cast

# use cast to look for total range

```r
> cast(iris_m, Species ~ ., range)
     Species  X1  X2
1     setosa 0.1 5.8
2 versicolor 1.0 7.0
3  virginica 1.4 7.9
```
# use cast to compute margins

```r
> cast(iris_m, Species ~ variable, mean, margins=T)

<table>
<thead>
<tr>
<th>Species</th>
<th>Sepal.Length</th>
<th>Sepal.Width</th>
<th>Petal.Length</th>
<th>Petal.Width</th>
<th>(all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>setosa</td>
<td>5.006000</td>
<td>3.428000</td>
<td>1.462</td>
<td>0.246000</td>
<td>2.5355</td>
</tr>
<tr>
<td>versicolor</td>
<td>5.936000</td>
<td>2.770000</td>
<td>4.260</td>
<td>1.326000</td>
<td>3.5730</td>
</tr>
<tr>
<td>virginica</td>
<td>6.588000</td>
<td>2.974000</td>
<td>5.552</td>
<td>2.026000</td>
<td>4.2850</td>
</tr>
<tr>
<td>(all)</td>
<td>5.843333</td>
<td>3.057333</td>
<td>3.758</td>
<td>1.199333</td>
<td>3.4645</td>
</tr>
</tbody>
</table>
```
reshape

- two simple functions
- allows you to reshape your data as you need