Geographic Information Systems 2018/19
Week 3, Topic 3

Mark Foley
School of Computing
DIT, Kevin Street

mark.foley@dit.ie

In this session...

Attribute data

Symbology

Classification

Attribute data (DBM revisited)

Spatial and non-spatial attributes of a *Feature* can be split (Shapefile approach) or stored together in a single row (PostGIS approach)

Spatial data stored as an abstract data type (of type *Geometry*). The specific geometry type stored is as per the SFS model referred to in Topic 1.

Spatial data is serialized into a binary format (PostGIS)

Shapefile stores spatial attributes in proprietary format in **.shp** files and non-spatial attributes in **.dbf** files (dBase IV format)

Either way spatial and non-spatial data are related and can be thought of as elements of a single ***feature***.

By using PostgreSQL/PostGIS (or any other spatially-enabled database, we get all of the advantages of the relational database approach to managing data in the general sense.

Attribute data

Topological data and raster data can be stored in database tables but, for now, we consider simple features only.

PostGIS and Sptialite behave in a similar fashion but Spatialite is a single-file, single-user database and so can dispense with any functionality relating to concurrency, locking and access control. This gives Spatialite files a very small size. Spatialite is to SQLite as PostGIS is to PostgreSQL and follows a similar design pattern.

Attribute data

Attributes give context to spatial entities

- How does a linestring know that it's a road? The fact is encoded in its attributes.

Attributes as measures and classifiers

- Types
  - **Nominal** values – any class or category – usually arbitrarily assigned, difference is important
  - **Ordinal** values – any rank – not an absolute measure of anything but the *order* is important
  - **Interval** values – numeric measure where the difference or order is important but the zero value is arbitrary – example is temperature in Celsius: 20^0 is hotter than 10^0 but not twice as hot because 0^0 has been arbitrarily assigned
  - **Ratio** values – numeric measure where ratio and difference make sense – Zero is not arbitrary but indicates absence – 10 miles from a point *is* twice as far from the point as 5 miles
  - **Cyclic** values – where we go back to zero after a certain value – degrees of the compass or time (1 o'clock follows 12 o'clock) – these occur frequently in GIS
- Attribute type has a bearing on how data is handled and displayed in GIS

Classification schemes

Equal Interval: arranges a set of attribute values into groups that contain an equal range of values. This can help show different groups when they are close in size. However, this doesn't often occur in geographic phenomena. Take the range of your data (maximum - minimum) and divide by your chosen number of categories.

Quantile: divides the attribute values equally into a predefined number of classes. The attribute values are added up, then divided into the predetermined number of classes. In order to do this, you take the number of total observations and divide that by the number of classes resulting in the number of observations in each class. One of the advantages to using this method is that the classes are easy to compute and each class is equally represented on the map. Ordinal data can be easily classified using this method since the class assignment of quantiles is based on ranked data

Jenks Natural Breaks: The Jenks Natural Breaks Classification,(or Optimization) system is a data classification method designed to optimize the arrangement of a set of values into "natural" classes. This is done by seeking to minimize the average deviation from the class mean while maximizing the deviation from the means of the other groups. The method reduces the variance within classes and maximizes the variance between classes.

Geometric Interval: This classification method is used for visualizing continuous data that is not distributed normally. This method was designed to work on data that contains excessive duplicate values, e.g., 35% of the features have the same value.

Standard Deviation: The Standard Deviation Classification method finds the mean value of the observations then places class breaks above and below the mean at intervals of either .25, .5, or 1 standard deviation until all the data values are contained within the classes. This classification method shows how much the feature's attribute value varies from the mean. Using a diverging color scheme to illustrate these values is useful to emphasize which observations are above the mean and which observations are below the mean.

Symbology

The symbology of a layer is its visual appearance on the map. The basic strength of GIS over other ways of representing data with spatial aspects is that with GIS, you have a dynamic visual representation of the data you’re working with.

Therefore, the visual appearance of the map (which depends on the symbology of the individual layers) is very important. The end user of the maps you produce will need to be able to easily see what the map represents. Equally as important, you need to be able to explore the data as you’re working with it, and good symbology helps a lot.

Values in your attributes can dynamically determine the displayed symbology. For example, a mpa that shades areas by some attribute of eacfh area is known as a choropleth map.

Choropleth map example

Comparison of choropleth class definition schemes: natural breaks, quantile, equal interval, standard deviation. The data are Mobile Homes Density for North Central USA, 2004.

Choices in choropleth mapping:

Shading patterns and colour

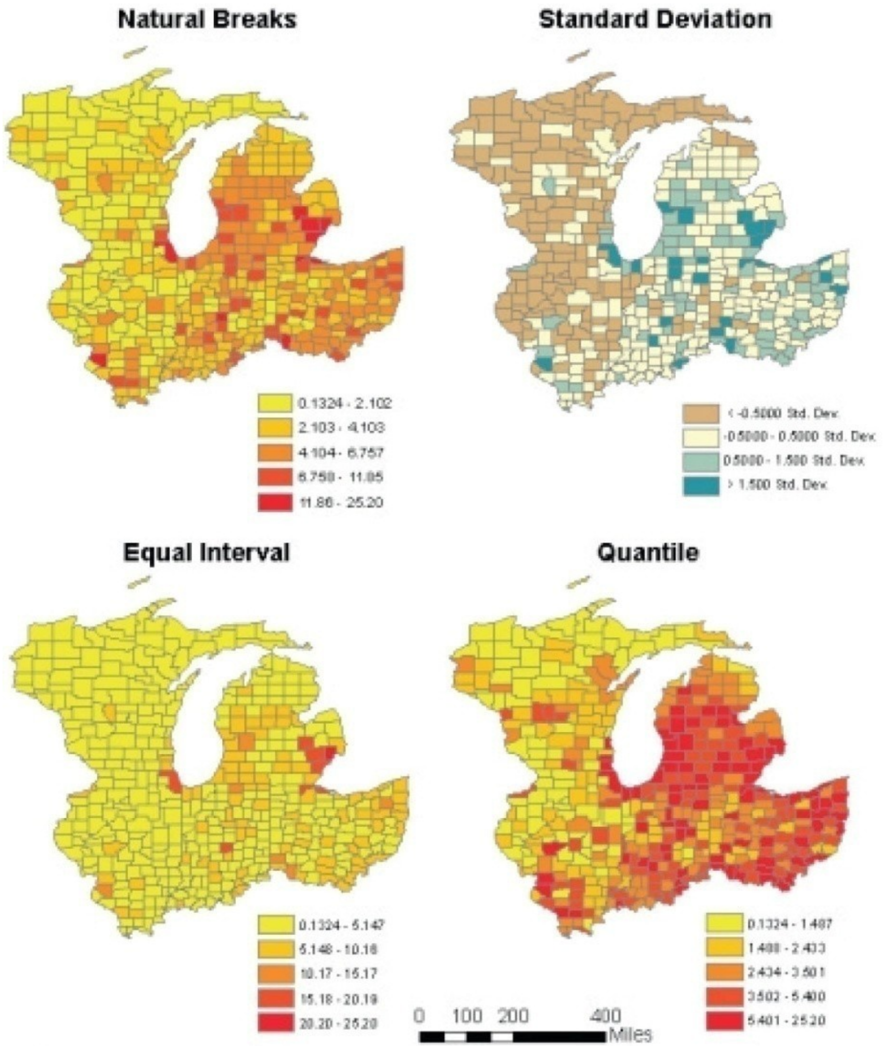
- Denser/darker = more of

Relationships between attribute values

- Nominal -> contrasting colours/patterns
- Ordinal -> ramped colour

Classification system

- Don't overdo number of classes
- Choose class intervals carefully



Sybmbology

Note how easy it is to “lie” with maps

We’ll return to this theme when we consider cartography and map design

Coming next...

Raster Data

Idea of continuous data

Raster types

Cell values

Resolution