

Chapter 1 – Introducing GIS Analysis

Intro –

We analyze all the time, but just don't know it.

So Mitchell looks at 6 of the most common types of analysis in a more formal way:

- Mapping where things are
- Mapping the most and the least
- Mapping density
- Finding what's inside
- Finding what's nearby
- Mapping change

These types of analysis may require different techniques, diff extensions, diff approaches.

This book will help the user start developing basic analytical skills. Over time, more books on higher levels of analysis, modeling, etc, are becoming available (e.g., Modeling Our World, also by ESRI).

I. What is GIS Analysis?

“chapters in the book follow the process you go through when performing an analysis”

1. Frame the question – analysis often starts with a question:
“where were most burglaries last month?”
“how much forest is in this watershed?”
“what parcels are within 500 feet of this liquor store (WWM just did this for Communipaw Ave Block Assn for Liquor Control Board hearing....)”

usually, the more specific your question, the more well-defined your analysis and the better you can choose your methods of analysis

2. Understand your data – you need to understand what you have, its limitations, and what you might need to add to it (create new data). Creating new data could be just making new fields in the attribute table, or could involve more raw data collection.
3. Choose a method of analysis – “there are almost always 2-3 ways getting the info you need”. One method may be quick, give general results. Another method may require more data, more processing, and give more precise results.
4. Process the data – after you select the method, perform the steps. This book is geared toward helping the user understand what the GIS is doing for the more common types of analysis. Also, book will help you choose parameters.
5. Analyze results – results can be shown in many ways. Could be: a map, values in a table, a chart or graph. Now, you have to decide what info to use, how to group the info, how to present the info so others understand it.

II Understanding Geographic features

A. first need to understand that there are different types of “geographic features”

- discrete
- continuous
- summarized by area

1. Discrete features – individual points or lines or polygons – features have specific locations, can be pinpointed exactly (**fig p.12, left**)

2. Continuous features – (precipitation, temperature, elevation, air and water quality, etc) “blanket the area” – no gaps. NOTE – the actual measurement of these continuous phenomena, however, are usually performed at discrete locations (rain gauges, thermometers, air sensors, water quality Hydrolabs). Then these discrete measurements are extrapolated throughout the “blanket” by creating TINs (or by making raster meshes), and then these surfaces are contoured, like a topographic elevation map. (**fig p.12, right**)

Note that some continuous data, like soil types, can be separated from each other with boundaries (**fig p.12**).

3. features summarized by area

examples:

- no. of businesses in each zip code
- total length of streams in a watershed
- no. of households in each census tract

these features “represent the counts (ie, the actual number of indiv features) or density of features within a polygon”

this can also be the sums of like features from polygons, like no. of households in a census tract can be summed for all census tracts in a Hudson Co. to yield no. of households in Hudson Co.

For features summarized by area, remember that the data value applies to the entire area...it is an average value. **Fig UL p13** shows how indiv business locations (left) can be transformed into a density or counts map based upon the area of each zip code (right).

This leads us to a simple but good exercise...**Exr 1**...what is the housing (household) density of YOUR census tract and YOUR county?) use Arc8 and TIGER.

final product will be an .mxd file (like a .apr), containing a map and an attribute table showing the number of households in each census tract in your county, with your census tract highlighted, and a box showing the avg no. of households/tract in your county)

Mitchell notes that “a lot of data already comes summarized by area, especially demographic data” like the TIGER data you’ll use in Exr. 1.

For example, there are totals (total pop., total h-hold) and percentages (% over 65 yrs old, % Hispanic residents).

What if you have other data that you might wish to summarize by area? Poorly written, but Mitchell implies you work with an external table and then do a “join”.

If the features are already “tagged” (somehow linked) to an area (eg, each business is listed with its zip code, then you can just join this external data table to an attribute table using zipcode as the primary key. **(Fig UR p.13)**

GOOD POINT:

Mitchell notes in **Fig LR p.13** that if you do NOT have the businesses linked to the zip codes already, you can use overlay and spatial join techniques (Ch 5). We did this in NJCU urb design project Sp 2003, where we joined point data from NJDEP Contam Site list with the block-lot tracts from City of Jersey City.

B. Two ways of representing geog features:

1. vector

each feature takes up a row in a table, feature shapes are defined by x&y (or long&lat) coords – pt is x,y; line is between 2 pts; polygon is collection of lines

“when you analyze vector data, much of your analysis involves working with the data in the layer’s data (attribute) table” – **there are multiple attributes in that table**

2. raster

“features are represented as a matrix of cells in continuous space”

“each layer represents **one** attribute” – the raster layer does not HAVE an attribute table per se.....it really only has a “key” or legend that shows what number or color in each cell means. Therefore, most of your analysis involves the user combining layers in order to create new layers that have the combinations of attributes you want.

It is NOT an overstatement to say that vector and raster models, and the way that they handle attributes, are radically different, although it appears that they do very similar things (**figs p15 Left**). As you’ll see time and time again, most analysts use:

vector for discrete features (indiv points, lines, polygons),

raster for continuous features (topography, chemical concentrations, aerial photographs)

as you learned in GIS I, raster cell size is critically important to analysis...too large a cell and too much detail is lost. Too small a cell size and processing takes exponentially longer to do.

(Fig LR p. 14)

final point in this section: sometimes you may want to characterize discrete features with a raster system in order to produce a raster layer – this is because raster is particularly straightforward to use (at least in terms of visualization & theory) in overlay analysis involving multiple layers, a task which is performed frequently in GIS

C. Map projections and coordinate systems

You know from GIS I that all data layers MUST be in the same map projection and in the same coordinate system, otherwise things don’t “line up” properly. You’ve seen this when you try to overlay unprojected lat-long data with projected State Plane or UTM data. **Fig UR p.15** shows this as well.

Mitchell explains these two components:

“map projection translates the locations on the globe to a flat surface”

“a coordinate system specifies the units use to locate ftrs in 2D space and also the origin point of those units”

One potentially big benefit of Arc8 (vs AV3) is that it will re-project, on the fly, data from one system so that it overlays the first, controlling layer in a data frame. We will test this in **an extra credit exercise (set a time/date for GPS collection)**, where we will bring in GPS data from a lat-long system and place it on the Jersey City block-lot shapefile, which is in State Plane, NAD 83 coordinates in both AV3 and AV8, and see what happens.

III. Understanding geographic attributes

Recall from GIS I the “NOIR” measurement system for classifying attributes:

Nominal – naming or categorizing (simple labels, no ordering implied)

Ordinal – ordering (1st, 2nd, 3rd; oldest to youngest; largest to smallest; etc)

Interval – (like the degrees on the Fahrenheit and Celsius scales)

Ratio – mathematical operations can be performed because msmt system has a common zero

A. Mitchell lists several types of attributes values:

- Categories
- Ranks
- Counts
- Amounts
- Ratios

“**Categories** are groups of similar things” – categories are examples of the Nominal system of msrmt.

Freeways, highways, or local roads are all “roads” and can be shown with lines.

Burglaries, thefts, assaults are all crimes and can be shown with points.

Category values can be represented using “numeric codes or text”...this works for either vector or raster.

Note the example in **fig LL p.16**...the “type” category identifies the type of crime with a 3-digit number.

“**Ranks** put features in order, from high to low”.

This follows the Ordinal system of measurement. **Fig UR p.16** shows streams ranked by recreation value. Mitchell points out that this does NOT tell an analyst how much higher or lower one feature is ranked wrt another

Realize that ordering things frequently requires some type of judgment or interpretation that is introducing an “opinion”. Also realize that not all people share the same opinion. (eg, who ranks what college football team #1?) opinions vary WIDELY.

Ranks usually are assigned on the basis of another attribute....example **fig LR p.16** shows how soils are ranked according to how suitable (good) they are for growing a certain crop

“**Counts and amounts** show you total numbers”.

Count is the actual number of features on the map. Amount can be “any measurable quantity associated with a feature, such as number of employees at a business.” These are in the Ratio system of measurement, because the measurable quantities have a common zero.

“Using count or amount shows actual value of certain attributes, as well as magnitude compared to other features.” **See fig UL p.17.**(shows # of employees with progressively larger dots)

“**Ratios** show the relationship between 2 quantities, and are calculated by dividing one quantity by another, for each feature.” Of course, ratios are in the Ratio system of msrmt.

Example: In a given census tract, total no. of people / total no. households = avg no. people per 1 household. See **fig LL p.17**

In AV, this operation to get a ratio is often accomplished in the table legend by using “normalize”.

2 special ratios are mentioned by Mitchell:

- Proportions
- densities

proportions show “what part of a total each value is” .

Example: in a census tract, the number of 18 to 30 year olds / total population = proportion (%) of 18 to 30 yr olds (**fig UR p.17**)

“densities show distribution of values or features per unit area”

Example: **fig LR p.17** shows population density for a census tract, in no. of people/sq mile. This requires you to make decisions about what “pop density” means, and to make sure you have the data you need to calculate it (here, the area in sq. miles for each census tract)

The attribute types listed above can also be split into two major categories:

- Categories & ranks = noncontinuous values
A “set” number of values in the layer – only so many values are present
- Counts, amts, ratios = continuous values
A feature may have a unique value anywhere in a range from highest to lowest....not necessarily limited to a “set” number of values

B. working with data tables

Mitchell notes 3 main oprns that an analyst typically does with GIS –

- Select
- Calculate
- Summarize

Selecting is essentially querying the database, perhaps multiple times or stringing queries together. Mitchell example is finding land use = commercial and parcels > 2 acres.

Calculation is frequently performed upon attributes in order to generate new data (new attributes). This is done with the Calculator in AV3 and AV8. Note that the analyst often has to create a new column first, with the name of the new attribute, before actually performing the calculation. Mitchell example is the **table UL p.19**. and also **table LL**. In the latter, analyst creates a “people per household” column, then performs the calc using two existing columns in attrib table: population / households.

Summarizing

Here you “summarize attribute values to get statistics”. Sometimes you’ll create a new table to house the summary data. In the tables **R p.19**, it appears that Mitchell takes specific landuse data in a floodplain area (**LR**), then adds the data (sums it) in order to create a sum of land value, count of features, mean land value, and max land value (**UR**).