Lecture 3

Maps and Data Types in GIS

Tomislav Sapic
GIS Technologist
Faculty of Natural Resources Management
Lakehead University
# Data Types

In GIS, data types and data models are forms of attribute representation of observed phenomena.

Data types – representations of values.

<table>
<thead>
<tr>
<th>Data Type Categories:</th>
<th>Numeric</th>
<th>Text (string, character)</th>
<th>Date</th>
<th>Boolean (binary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations:</td>
<td>Integer (whole number): 1, -1, 2, 102, …;</td>
<td>Text (letters like a, b, c, but also characters 1, 2, 3, _, #)</td>
<td>Time and date</td>
<td>True/False (+/-)</td>
</tr>
<tr>
<td></td>
<td>Floating point, double (real number, decimal number): 0.0234, 276.23, …</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Levels of Measurement

- GIS is used to capture any spatially related information, which means almost any type of data.
- Different data types are suitable for different levels of measurement – data type values are assigned to measured variables.

Nominal data: different kinds or categories (e.g., dominant tree species) – could be text or integer values.
Ordinal data: ranking relationship (e.g., stand site class) – usually represented by integer values.
Interval data: have known intervals between data, a zero doesn’t mean zero units (e.g., temperature in °C) – usually decimal values (float or double).
Ratio data: similar to interval data but having an absolute zero (e.g., temperature in Kelvins, tree height, wood volume) – usually decimal values (float or double).
Computer (Digital) Binary System

- The smallest unit of computer memory is byte.
- One byte contains 8 bits.
- Eight bits in the binary system when converted to the decimal system can be a range of numbers from 0 to 255.

\[
\begin{align*}
00000000 & \quad 11111111 & \quad 111111001 \\
2^7 & + 2^6 & + 2^5 & + 2^4 & + 2^3 & + 2^2 & + 2^1 & + 2^0 & \quad 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1 = 255 \\
0 & + 0 & + 0 & + 0 & + 0 & + 0 & + 0 & + 0 & = 0 \\
\end{align*}
\]

- The above approach is used to produce integers, a different approach is applied in storing decimal numbers.

Decimal numbers are saved through binary representations of their components: sign, significand and exponent.

\[(-1)^s \times c \times b^q\]

\[-0.012345 = -1^{-1} \times 12345 \times 10^{-6}\]
Data Types in ArcGIS

- In a GIS software, such as ArcGIS, data types refer to the attributes in the vector data model and cell values in the raster data model.
- In the vector data model, attributes are stored in table fields. A field can accept only one *data type* – e.g., there can’t be a letter saved in an integer field.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Text</strong></td>
<td>Characters such as letters, numerical digits, common punctuation marks. In the computer memory, characters are coded and stored as numbers – e.g. in the <em>ASCII</em> (American Standard Code for Information Exchange) scheme English alphabet characters are given code numbers stored in 1 byte size memory holders. <em>Unicode</em> is, on the other hand, an international encoding standard that codes for characters used in most of world’s writing systems.</td>
</tr>
<tr>
<td><strong>Integer</strong></td>
<td>Whole number – no decimal places. In ArcGIS integers are signed – can be positive and negative, as opposed to unsigned integers. <em>Short signed integers</em> take 2 bytes of space in the memory (storable range -32,768 to 32,768), and <em>long signed integers</em> take 4 bytes of space (storable range -2,147,483,648 to 2,147,483,647).</td>
</tr>
<tr>
<td><strong>Float</strong></td>
<td>Decimal number; single-precision floating point number; 4 bytes; storable range +/- 10^-38 to 10^38.</td>
</tr>
<tr>
<td><strong>Double</strong></td>
<td>Decimal number; double-precision floating point number; 8 bytes; storable range +/- 10^-308 to 10^308.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>Stores dates, times, or dates and times.</td>
</tr>
<tr>
<td><strong>Boolean</strong></td>
<td>Some databases also have an option of specifying the boolean data type, in which case only one binary number is stored (either 1 or 0) and is shown to the user as True or False, respectively.</td>
</tr>
</tbody>
</table>
• *Text* is stored in the computer as a series of characters, each of which is coded in the binary system and occupies 1 byte of memory.

• Character coding specifies the correspondence between the binary pattern and symbols used in the written language.

**ASCII (American Standard Code for Information Interchange) Printable Characters**

<table>
<thead>
<tr>
<th>Binary</th>
<th>Dec</th>
<th>Hex</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0010 0000</td>
<td>32</td>
<td>20</td>
<td>(blank) (□)</td>
</tr>
<tr>
<td>0010 0001</td>
<td>33</td>
<td>21</td>
<td>!</td>
</tr>
<tr>
<td>0010 0010</td>
<td>34</td>
<td>22</td>
<td>&quot;</td>
</tr>
<tr>
<td>0010 0011</td>
<td>35</td>
<td>23</td>
<td>#</td>
</tr>
<tr>
<td>0010 0100</td>
<td>36</td>
<td>24</td>
<td>$</td>
</tr>
<tr>
<td>0010 0101</td>
<td>37</td>
<td>25</td>
<td>%</td>
</tr>
<tr>
<td>0010 0110</td>
<td>38</td>
<td>26</td>
<td>&amp;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bin</th>
<th>Dec</th>
<th>Hex</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0100 0000</td>
<td>64</td>
<td>40</td>
<td>@</td>
</tr>
<tr>
<td>0100 0001</td>
<td>65</td>
<td>41</td>
<td>A</td>
</tr>
<tr>
<td>0100 0010</td>
<td>66</td>
<td>42</td>
<td>B</td>
</tr>
<tr>
<td>0100 0011</td>
<td>67</td>
<td>43</td>
<td>C</td>
</tr>
<tr>
<td>0100 0100</td>
<td>68</td>
<td>44</td>
<td>D</td>
</tr>
<tr>
<td>0100 0101</td>
<td>69</td>
<td>45</td>
<td>E</td>
</tr>
<tr>
<td>0100 0110</td>
<td>70</td>
<td>46</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bin</th>
<th>Dec</th>
<th>Hex</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>0110 0000</td>
<td>96</td>
<td>60</td>
<td>‘</td>
</tr>
<tr>
<td>0110 0001</td>
<td>97</td>
<td>61</td>
<td>a</td>
</tr>
<tr>
<td>0110 0010</td>
<td>98</td>
<td>62</td>
<td>b</td>
</tr>
<tr>
<td>0110 0011</td>
<td>99</td>
<td>63</td>
<td>c</td>
</tr>
<tr>
<td>0110 0100</td>
<td>100</td>
<td>64</td>
<td>d</td>
</tr>
<tr>
<td>0110 0101</td>
<td>101</td>
<td>65</td>
<td>e</td>
</tr>
<tr>
<td>0110 0110</td>
<td>102</td>
<td>66</td>
<td>f</td>
</tr>
</tbody>
</table>
Data Types in Vector Data Model

• In the vector data model, data types refer to designations of the fields in the table – each field is assigned a certain data type and only values of that data type can be stored in that particular field.

• Data types are decided on during the process of the table design and are based on the data that are being stored or anticipated to be stored and the type of analysis one expects to carry out.

• Issues such as expected range of numerical values, length of a text (character) values string are part of the decision.

Ontario_major_cities.shp

• Fields are attributes, meaning, variables, and values for each attribute/variable are stored for individual records (i.e., one or multiple spatial features).

• Fields created of a certain data type occupy the memory space relative to the number of records (rows) and irrespective of the values stored in them.
Data Types in Vector Data Model

- Precision (total number of digits) and Scale (number of decimal places) can be specified for floats and doubles.

- In ArcGIS, text data type can be distinguished from numerical data type by alignment.

Text left-aligned

Numbers right-aligned
Not All Zeros in Shapefile Table Were Created Equal

- Numeric fields (integer, float, double) in shapefile (dbase) tables get populated with zeroes by default, zero then meaning an empty space. However, many data representations use zeroes for specific values (e.g., a 0 site class in FRI represents the best site class), creating a situation where zeros in the same field can have two very different meanings.
- This is solved in geodatabase feature classes with the introduction of the Null attribute, which means no value.

<table>
<thead>
<tr>
<th>DETYPE</th>
<th>OSPCOMP</th>
<th>OYRORG</th>
<th>OLEADSPC</th>
<th>OAGE</th>
<th>OHT</th>
<th>OSI</th>
<th>OSC</th>
<th>OCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt 60Sw 20Bw 10Pj 10</td>
<td>1943</td>
<td>Pt</td>
<td>70</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Sw 40Sb 30Bf 20Pt 10</td>
<td>1983</td>
<td>Sw</td>
<td>30</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sb 80La 10Cw 10</td>
<td>1928</td>
<td>Sb</td>
<td>85</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pt 80Sw 10Bw 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sb 50Pt 40Bf 10</td>
<td>1938</td>
<td>Sb</td>
<td>75</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Bw 30Pt 30Sb 20Bf 20</td>
<td>1943</td>
<td>Bw</td>
<td>70</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

- Not a stand (could be a lake, wetland, road, etc.)
- OSPCOMP – Overstory Species Composition
- OSC – Overstory Site Class
- Productive stands
Data Types in Raster Data Model

- Raster (band) is a variable and cells have numerical values assigned to them.

- Rasters created of a certain data type occupy the memory space relative to the number of cells (columns x rows) and irrespective of the values stored in them.
Maps in GIS

Several definitions from different sources on what a map is:

1. A map is a representation of spatial information.

2. A map is a spatial representation of the environment.

3. A map is a simplified depiction of a space, a navigational aid which highlights relations between objects within that space.
Maps in GIS

Two points to keep in mind regarding maps:

• Maps are a form of communication, which means that what they show should be clear and understandable to a broad range of users. This then invokes symbology, selectiveness and narration.

• Being a simplified spatial representation and (almost exclusively) a scaled-down one, maps have to provide information in a way that will enable the users to understand orientation, spatial characteristics and spatial relations between the represented features (phenomena).
Maps in GIS

• Easiness with which maps can be made with GIS software and abundance of GIS datasets have opened room for an easy creation of maps that often end up being of poor quality in terms of their content and functionality.

• When creating maps, GIS users should at the very least include all map elements, make the content clear and understandable (proper names of legend items, clear labelling, etc.), and use conventional colours for basic features (blue for lakes and streams, black or red for roads, etc.).
Basic Map Elements

- Map body
- Title
- Legend
- Scale
- Projection
- Direction (North Arrow)
- Data source (map author, date of creation, date of data, map projection)
Map Design

• The purpose: to enhance map communication by creating a map that is balanced, coherent, ordered and interesting to look at (Antes et al. 1985).

![Map of Idaho Population 1990-95](image1)

The box around the legend attracts the viewer’s attention to it.

Source: Chang (2008).

![Poorly designed map](image2)
Too many boxes distract the viewer.

The state border in the display on the right is thicker, differentiating it from the county borders.

Source: Chang (2008).
Are all Basic Map Elements Always Necessary?

North Arrow?

Scale?
Map Title

• Generally, the title should indicate the major theme being mapped, the geographic region, and the date of the data.

• It should be placed at the top of the map; if choosing between the left and the right side, preferably it should be on the left side.

Map Legend

• Basic features should be explained if not obvious such as lakes, rivers.

• The legend should have a title if the legend topic is more specific that what the map title suggests.

• Avoid using ‘Legend.’

• Larger values in the legend should be above smaller values because of human perception – up = larger.
FIGURE 2.13  How base information (railroads, in this case) can be combined with a major theme. (Compare with Figure 2.12.)


FIGURE 2.15  An example of poor legend construction and poor positioning of the title, legend, and source. (Compare with Figure 2.12.)

Font

- Use **sans serif**.

- Use only one font typeface for the entire display.

- Legibility is more important than decorative value.

- Use size to distinguish between major elements.

- A combination of uppercase and lowercase lettering is more readable than uppercase alone. Uppercase may be used for short titles.

- Avoid size smaller than 10 points (1 point = 1/72 of an inch (0.35 mm))

- Use bold lettering for important elements only, such as titles.

Many different fonts reduce clarity and cause distraction.

Source: Chang (2008).
Fonts in ArcGIS

• To keep it on the safe side, if not sure how the map file will react to the displaying device, use common fonts, such as Arial or Times New Roman.

• Special characters can be copied over from the operating system character dataset.
Labels

• In GIS, labels usually refer to the feature attributes displayed on their respective features.
• Maps’ primary function is to convey spatial information.
• Labels should be brought down to minimal and tables used instead.
• Sometimes, though, labels are a part of the requirement, e.g. regulations.
General Rules on Labels

- Labels should be easily perceived and unambiguous.
- Avoidance of overlap.
- Conformance to shape and extent of area features.
- Avoidance of excessive clustering of names.
- Conformance to the curvature of the parallels for horizontal placements.
- Conformance to applicable standards and conventions.

Source: Chang (2008).
Printing and Displaying Maps

• Maps are created to be viewed on different media.

• Some examples of the map media are:
  o Paper sheets (e.g., standalone maps, poster and journal figures, etc.)
  o Digital files (e.g., pdf, jpg)
  o Powerpoint presentations

• The spatial size (e.g. map scale) and the amount of the information on a map should reflect the means of viewing by the map viewers (Can one zoom in onto the map, from how close will average viewers be looking at the map (posters behind a presentation booth!), for how long is the map displayed (powerpoint presentation!)?).

• Maps are a message, and the message should be clear and effective.
Printing and Displaying Maps

• The first thing when creating maps is to set the size of the map printout (display).

• Sizes can vastly vary. From a small figure in a report or a journal paper to a several-feet-across standalone map poster.

• The set size will dictate the map scale.

• Letters two inches high are seen at a distance of fourteen feet at approximately the same size as a 10-point type seen at a reading distance of one foot.

• Test seeing the map in a display that is intended for (e.g., projector, black and white printout, etc.)
Simplifying Maps

• Other than controlling the labels and text, maps can be also made more simple by reducing the quantity of features shown or by simplifying (generalizing) the features.

• Care should be taken not to degrade original data.

Reducing the amount of features to only those above a certain size.

Generalizing the features
Colours: Additive and Subtractive Colour Theory

Additive

Subtractive

Paint, print ink, filters

Light, screen displays (computer, TV), projectors.
• Red, green and blue are *primary colours*.

• Yellow, magenta and cyan are also called *complementary colours* – in pairs with their primary colour counterpart they produce white light.

• Black is the absence of all colours.
CMYK Colour Model
(Subtractive)

C – Cyan
M – Magenta
Y – Yellow
K – Key
(Black)
HSV Colour Model

H – Hue (0-360°)
S – Saturation (0 – 100%)
V – Value (0 – 100%)

• Suitable for assigning grayscale colours, setting the saturation of a colour, etc.

Source: Jensen (2007).
• the aim is to create a map that is logical in conveying its message.
Choosing Proper Colour Schemes

Logical versus illogical colour scheme

COLOR PLATE 1.1. The choropleth map: an example of a thematic map. Map A uses illogically ordered hues, while map B uses logically ordered shades of a single hue. Although map A may allow the reader to discriminate easily between individual states, it does not permit the reader to perceive the overall spatial pattern as readily as map B. (Data Source: Famighetti 1993, 583.)
Two Basic Map Types

Topographic (General Reference)  Thematic
Thematic Maps - Choropleth Maps

• Choropleth Maps – created by shading area units according to a descriptive value.

Limitations:

- Lacking possible spatial variation within the units.

- The boundaries of the units might be arbitrary and thus unlikely to reflect major discontinuities.

- Consider data standardization.

Unstandardized versus standardized data representation

**FIGURE 1.5** A comparison of the effect of data standardization. Map A is based on raw-count data (the number of acres of forested land), while map B is based on standardized data (the number of acres of forested land relative to the area of each state). Map A is misleading because states with large areas tend to have more forest. (Data Source: Powell et al. 1992.)
## Levels of Measurement and Symbology

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>Ordinal</th>
<th>Numerical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spacing</td>
<td>P</td>
<td>M(^c)</td>
<td>M(^c)</td>
</tr>
<tr>
<td>Size</td>
<td>P</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Perspective</td>
<td>P</td>
<td>M(^a)</td>
<td>G(^b)</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>G</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Shape</td>
<td>G</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Arrangement</td>
<td>G</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Lightness</td>
<td>P</td>
<td>G(^d)</td>
<td>M</td>
</tr>
<tr>
<td>Hue</td>
<td>G(^d)</td>
<td>G(^d)</td>
<td>M</td>
</tr>
<tr>
<td>Saturation</td>
<td>P</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>

\(^a\) Since height differences are suggestive of numerical differences, use with caution for ordinal data.

\(^b\) Hidden enumeration units and lack of a north orientation are problems.

\(^c\) Not aesthetically pleasing.

\(^d\) The particular hues selected must be carefully ordered, such as yellow, orange, red.

**FIGURE 2.10** Effectiveness of visual variables for each level of measurement for areal phenomena. (After MacEachren 1994a, 33.)

Symbolizing Attribute Data

• All features, i.e., all attributes, in a layer can be displayed with the same symbol, or groups of attributes or selected attributes, can be displayed with different symbols.

• The two main ways in GIS of displaying attributes separately:
  - Categories > Unique Values
  - Quantities > Graduated Colours (Symbols)
Symbolizing Attribute Data

Categories: Unique Values

- Used for discrete data – nominal and ordinal levels of measurement.
Symbolizing Attribute Data Quantities: Graduated Colours (Symbols)

- Data are classed into groups (classes).

- Interval and Ratio levels of measurement.

- Classes can be defined in different ways; ArcMap offers seven classification schemes:
  - Manual
  - Equal Interval
  - Defined Interval
  - Quantile
  - Natural Breaks
  - Geometric Interval
  - Standard Deviation
Classification Schemes - General

Equal Intervals (A)

Quantile (B)

Maximum Breaks (D)

Natural Breaks (E)

Graduated Colours Classification Schemes - ArcMap

Manual classification allows the user to set up her own intervals.

The histogram shows both original data, in gray columns, and created intervals.
Equal Interval breaks the whole range of values into equal intervals (class ranges).
Defined Interval classification breaks the range of values into intervals defined by the user.
Quantile classification breaks the range of values into classes of equal frequencies.
Graduated Colours Classification Schemes - ArcMap

Natural Breaks classification divides the range of values into ‘natural’ classes defined by ArcMap – similar values are grouped together and differences between classes are maximized.
Geometric Interval (Smart Quantile) breaks the range of values into classes of similar frequencies and with fairly consistent changes between intervals.
Standard Deviation classification separates the range of values into classes based on the standard deviations around mean.

Standard deviations and mean shown.
Map Projection

- Maps should be displayed with a proper map projection -- e.g., two popular map projections in Canada are UTM, for the areas south of 84° latitude and with a width around the width, at the latitude in question, of a UTM zone, and Lambert Conformal Conic for areas wider than that.

- Unless there is a very specific and compelling reason, GCS should never be used to display maps.
Forest Management Planning Maps

• Overall directions on maps are provided in the Forest Information Manual (FIM),

• FIM then directs to look for more specifics on map standards and content in other specialized manuals, such as FIM Forest Management Planning Technical Specifications and FIM Base and Values Technical Specifications.
Examples of inventive, interesting and creative maps (referenced mainly through Simon Keustenmacher https://twitter.com/simongerman600: and Paul Banks https://twitter.com/PaulBanks84

https://www.reddit.com/r/MapPorn/comments/70h9n1/i_drew_a_map_of_the_indigenous_people_languages/?st=J7NYQKQB&sh=925ddbb9

https://i.redd.it/6pcqozucc9mz.jpg
Author: Erik Crouch
The 2017 Fall Foliage Map is the ultimate visual planning guide to the annual progressive changing of the leaves. While no tool can be 100% accurate, this tool is meant to help travelers better time their trips to have the best opportunity of catching peak color each year.
94% of China’s population lives on one side, 6% on the other. A visually rich look at the Hu Line by @sixhtone http://interaction.sixhto

Stretching from Heihe on the Russian border to Tengchong on China’s southwestern border with Myanmar, the Hu Line divides China into two parts.

To the east, just over one-third of the nation’s land houses almost 94 percent of the country’s population.

Only 6 percent of citizens — but most of the nation’s ethnic minority groups — share the vast and varied terrain to the west.
I think that Google Earth should add a feature that allows you to exaggerate the reliefs.
Life expectancy gender gap [OC] [1500 x 800]

Female life expectancy is higher than male life expectancy by:
- More than 10 years
- 8 - 10 years
- 6 - 8 years
- 5 - 6 years
- 4 - 5 years
- 3 - 4 years
- 2 - 3 years
- Less than 2 years