

Glossary

Aerial photography

Photography from airborne platforms.

Boolean logic

A type of data-searching statement that uses the operators "and," "or," and "not." It can be used for selecting certain groups of data, such as all the data that are contained in a certain province, or all the data for women of childbearing age.

Buffer

An area created by specifying a distance from a point, line, or polygon on a map. Can be used to identify geographic features that occur inside or outside a certain distance from another feature.

Buffer tool

Creates buffer polygons to a specified distance around the input features. An optional dissolve can be performed to remove overlapping buffers in a GIS system. (Source: Environmental Systems Research Institute [ESRI])

Catchment area

Catchment area is the area from which a health facility attracts patients. A simple means of estimating a catchment area is to define a radius beyond which individuals are unlikely to access the services offered at that facility.

Choropleth map

A map that uses colors or shading to display attribute data for geographic areas rather than for points. To display values that take into account the size differences of the geographic areas, data should first be normalized (e.g., calculating population density, such as people per square kilometer, rather than using simple population counts). Choropleth maps are most visually informative when they display between two and seven classes of data using colors or shades that gradually darken as values increase.

Cluster

For Demographic and Health Surveys (DHS) data collection, the geographic location is collected based on what is known as the "cluster." DHS clusters are usually census enumeration areas—

sometimes villages in rural areas or city blocks in urban areas—containing the households selected for survey. A single GPS location is recorded at the center of the settlement area of the cluster. Collecting only one point for the cluster greatly reduces the chance of compromising the respondents' confidentiality, but is adequate to allow the integration of multiple data sets for further analysis. (Source: MEASURE DHS)

Confidentiality

The result achieved by protecting data and information that might identify individuals in a way that could cause harm or otherwise violate agreements made with them. For more information, see the MEASURE Evaluation publication, "Overview of Issues Concerning Confidentiality and Spatial Data": <http://www.cpc.unc.edu/measure/resources/publications/wp-08-106>.

Confidentiality agreement

Agreement entered into between a public health organization and an individual regarding the protection and nondisclosure of personally identifiable information.

Coordinate system

A coordinate system is a reference system used to represent the locations of geographic features, imagery, and observations such as GPS locations within a common geographic framework. (Source: ESRI)

CSV (comma separated values)

A type of spreadsheet file extension used commonly by open source software. Data in rows and columns can be exported in this format from Excel and imported into QGIS.

Data demand and use

Demand for and use of data to enhance evidence-based decision making for public health. Activities that foster data demand and use (DDU) involve a systematic approach that applies proven, effective best practices and appropriate tools to help increase demand for health system data and ensure that the information is used in an evidence-based decision making process. (Source: MEASURE Evaluation)

Data dictionary

A data dictionary is a text-based description of tables and fields in a database. It provides a solid foundation for writing data cleaning programs and a common language for facilitating communications between managers and analysts. (Source: U.S. Centers for Disease Control and Prevention [CDC])

Data schema

A data schema is a description of how data in a computer database are organized into tables and fields. It identifies acceptable values for individual fields. A common way to capture a data

schema is in a data dictionary. A proper data schema ensures that data are standardized and complete, and that they can be used to create accurate maps.

Datum

A set of control points, which are points on the surface of the Earth with known locations, and a corresponding mathematical model used to approximate the shape of Earth and to calculate the location of any given point on that shape. The datum used for the Global Positioning System is known as the World Geodetic System 1984 or WGS 84.

Decimal degrees

A numeric format for storing latitude and longitude that makes it easier to import coordinates into a GIS and to make location-based calculations. For example, a comparison of latitude and longitude formats for the location of the Library of Alexandria in Egypt is as follows:

Degrees, minutes, seconds: 31°12'31.93"N, 29°54'33.62"E

Decimal degrees: 31.208870, 29.909339

Dot density map

A map that uses dots to display data on a map. Each dot generally represents a given quantity of a certain occurrence (not necessarily one), such as 10 people per dot.

Environmental Systems Research Institute (ESRI)

The leader in commercial GIS software and services. Based in Redlands, CA (www.esri.com).

Equal intervals

A method of data classification that often works better for data that are continuous—that is, not highly skewed. Each resulting range (interval) of values will be approximately equal, but there may be a very different number of observations per class. Attention will be more focused on outliers. The map resulting from this method of data classification will tend to highlight any data with particularly high or low values, and may show an uneven distribution of colors.

Equator

An imaginary line encircling the Earth, equidistant from the poles, that demarcates zero degrees of latitude. The prime meridian splits the earth into Northern and Southern hemispheres.

Euclidean distance

A distance calculated using a straight line to connect the beginning and ending points.

Excel

A software platform from Microsoft Corporation that allows users to organize data into rows and columns (a spreadsheet) and perform calculations on the data. It comes as part of

Microsoft's famous "Office" software package. Files produced with this software usually have the name extension ".xls" or ".xlsx."

Exploratory spatial data analysis

Exploratory spatial data analysis (ESDA) applies statistical tools to the evaluation of spatial data. A basic technique for ESDA is to link observations in a histogram, box plot, and map to detect spatial patterns, such as outliers (source: Anselin, 2005). A box plot is a graphical summary of the following statistical measures: median, upper and lower quartiles, and minimum and maximum data values (source: NETMBA). A great free tool for ESDA is GeoDa (<http://geodacenter.asu.edu/>).

Extracting

The process of using an algorithm to decompress a "zipped" file, so that it can be used in a program. This must be done before a zipped file can be fully accessed. It is best to extract to a new folder, as more than one file can be zipped together, and upon extraction, they will be put into one manageable location.

File management

The process of storing, naming, and organizing files in a computer system.

Geocodes

Coded variables in a table of geographic data that indicate position, either at a point or within an area. They can be codes indicating latitude/longitude coordinates or administrative areas. This is a coded version of a geographic identifier.

Geodatabase

A newer type of geographic data storage invented by ESRI in the early 2000s. A geodatabase takes on the appearance of a folder filled with many files, when viewed in Windows, and can thus be easily moved or copied. It can comprise various types of geographic data sets that follow specific rules, which can be useful for editing the data and for performing advanced analysis. It is the main type of data currently used in ArcGIS, and its contents can be created and viewed by using an ESRI file management system called ArcCatalog.

Geographic coordinate system (GCS)

A coordinate system based on a three-dimensional, spherical surface. As a result of being defined in relation to the more natural, three-dimensional surface of a globe, a GCS is considered to be "unprojected" rather than "projected."

Geographic data

Information describing the location and attributes of things, including their shapes and representation. (Source: ESRI [ArcGIS documentation])

Geographic identifier

A geographic identifier is any piece of information that indicates the geographic or spatial location of features on the landscape, such as latitude and longitude, street address, or administrative division name (i.e., province, district, county, etc.). Common geographic identifiers play a critical role in joining data from different sources.

Geographic information system(s) (GIS)

A computer-based system used to collect, store, manage, analyze, display, and distribute geographic data (points, lines, and polygons referenced to the surface of the Earth) and their attributes (e.g., unique identifier, name, type, date collected, etc.). (Source: MEASURE Evaluation)

Geographic primitives

A graphic representation of a location: for example, a point to represent the location of a smokestack, or a polygon to represent the location of a toxic plume.

Geography

Geography is the study of patterns on the surface of Earth, and the causes of those patterns. The patterns can be the result of natural forces or human activity. This glossary entry is a synthesis of definitions from several sources, as many definitions of geography emphasize sub-fields of geography, and can be too narrow in scope. (Source: MEASURE Evaluation)

Georeferenced

Assigned to a geographic location.

Global positioning system (GPS)

Satellite-based system originally created by the United States Department of Defense to provide accurate data on position, velocity, and time to both military and civilian users. Coordinates are generally given in digital degrees relative to the equator and prime meridian.

Google Earth

A virtual globe from Google (www.google.com). It has the highest user base and satellite imagery library currently available, and can be downloaded from <http://earth.google.com>.

GPS receiver

GPS stands for “global positioning system” (see separate entry above). A GPS receiver is a hand-held device that can detect signals from the satellites composing the GPS, use them to determine the location of the viewer on the earth’s surface, and store this information for later use.

GPX

A file extension that denotes the GPS Exchange Format. This is a special schema used by many GPS units to store waypoint data. These files can be copied or shared to denote a set of locations on the Earth's surface. They can also be read by QGIS.

Ground truth

Information about a feature on the Earth's surface that is collected in the field. In remote sensing, the process of acquiring ground truth data is referred to as "ground truthing."

Health management information system(s) (HMIS)

A planned system of collecting, processing, storing, disseminating, and using health-related information to carry out functions of management. It consists of people, tools (paper-based and electronic), and procedures to gather, sort, and distribute timely, accurate information to decision makers. (Source: Kotler and Keller, 2006)

Histogram

A histogram is a graphical summary showing the count of data points falling in various ranges. It provides a rough approximation of the frequency distribution of the data. (Source: NETMBA)

IKONOS

The IKONOS Satellite is a high-resolution satellite operated by GeoEye (www.geoeye.com).

Imagery

Pictures or graphical representations. The term is used in remote sensing and GIS to describe digital representations of the surface of Earth. (Source: FWIE)

Isarithmic map

A map that uses contour lines to show change in a continuous variable over the land surface, such as temperature, precipitation, or elevation.

Kernel density estimation (KDE)

A geographic technique that disperses discrete phenomena across continuous space without the constraints of administrative boundaries. It provides a more realistic representation of the spread of people and services across a landscape. (Source: Spencer and Angeles 2007)

KML

KML, which originally stood for Keyhole Markup Language, is an XML-based file format that can incorporate descriptive text, image links, and geographic information associated with points, lines, and polygons. It is an open standard officially named the OpenGIS® KML Encoding Standard (OGC KML). KML files can be read by Google Earth and several mapping software packages. (Source: Open Geospatial)

Land cover

What can be seen remotely, from satellite data or aerial photographs. Current mapping techniques of land cover would not be possible today without milestones such as James Anderson's 1976 publication, *A Land Cover Classification System for Use with Remote Sensor Data*. (Source: CPC and USGS)

Latitude

Angle between a line connecting the center of the Earth to the equator and a line connecting the center of the Earth to a point on the Earth's surface on, north, or south of the equator along a line of longitude. Latitude ranges from 0 degrees at the equator to 90 degrees at the poles. Latitude is positive north of the equator (0 to 90 degrees) and negative below it (0 to -90 degrees). Lines of constant latitude can be visualized as circles drawn around the Earth horizontally in parallel with the equator.

Longitude

Angle between (a) a line connecting the center of the Earth to the equator at a prime meridian, such as the meridian that passes from pole to pole through Greenwich, England (also known as the Prime Meridian or Greenwich Meridian), and (b) a line connecting the center of the Earth to the equator at its intersection with a meridian that passes through the point of interest. Longitude ranges from 0 degrees at the Prime Meridian to 180 degrees along the meridian on the opposite side of the Earth. The 180th meridian roughly parallels the International Date Line, where the date changes as travelers cross going east or west. Lines of constant longitude can be visualized as half circles drawn on the Earth's surface vertically from pole to pole.

Medical geography

Medical geography applies the discipline of geography to the study of patterns of public or human health. (Source: Meade and Emch, 2010)

Metadata

Data about data. Metadata usually contain information about when a data set was collected or created and by whom. Geospatial metadata also contain information about projection and datum. Sometimes metadata are stored as part of the main data file, but they can also be in their own separate file. As a best practice, metadata should be provided with any geographic data set. International standards for geographic metadata are available as ISO 19115.

Multispectral (imagery)

Two or more images taken simultaneously, but each image taken in a different part of the electromagnetic spectrum. The electromagnetic spectrum is the total range of wavelengths or

frequencies of electromagnetic radiation, extending from the longest radio waves to the shortest known cosmic rays. A portion of the electromagnetic spectrum corresponds to light that is visible to the naked eye. (Source: CCRS)

National data infrastructure

The national data infrastructure consists of all the data available to national-level decision makers as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision-making purposes. (Source: MEASURE Evaluation)

National mapping agency (NMA)

National organization responsible for the creation and maintenance of map series and related data for a country. A good source of NMA contacts is the United Nations Second Administrative Level Boundaries (UN SALB) website (www.unsalb.org).

National spatial data infrastructure (NSDI)

The "technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and nonprofit sectors, and the academic community. The goal of this Infrastructure is to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships with [stakeholders] to increase data availability." (Source: FGDC)

Natural breaks

A method of data classification that assigns data to classes such that the variance within classes is minimized while the variance between classes is maximized. The primary advantage of this method is that it takes the natural distribution of the data into account before assigning observations to classes. A disadvantage is that the breaks between the classes could be irregular and therefore not intuitive

Network analysis

A method of geographic analysis that calculates measures along a network of such entities as roads, railroads, or rivers. It can be used to study accessibility of healthcare services.

Pixel

"Picture element" is the ground area corresponding to a single element of a digital image data set. (Source: CCRS)

Plug-in

Any type of added (often third-party) feature in QGIS. These are smaller, separate programs that can be used to enhance the basic features of QGIS. There is a QGIS "Plugin Repository"

where these programs can be uploaded, and where they are vetted by a group that manages the QGIS software itself.

Positional accuracy

The degree to which a user's location can be accurately determined. GPS data results can vary according to the arrangement of satellites overhead.

Prime meridian

An imaginary line passing through Greenwich, England, that demarcates zero degrees of longitude. The prime meridian splits the earth into eastern and western hemispheres.

Projected coordinate system

A coordinate system in which locations on the three-dimensional surface of the Earth are transformed or "projected" onto a flat, two-dimensional surface for display, measurement, or other analysis.

Projection

A process in which locations on the three-dimensional surface of the Earth are transformed onto a flat, two-dimensional surface for display, measurement, or other analysis. Any projection method will in some way ultimately compromise either distance, direction, or shape, but is a necessary step if straight-line distances or 2D (flat) areas are to be calculated during spatial analysis. A common type of projection is UTM (Universal Transverse Mercator), in which distances can be measured in meters.

Proportional symbol map

A map that uses varying sizes of symbols (often circles) to display attribute data for geographic areas or points.

QGIS

A free and open source (FOSS) software program that can read geographical data and display it on a map. Both vector and raster data can be read by QGIS. It is a powerful but user-friendly program that has a growing customer base. It has many features in common with ESRI's ArcGIS program and is also good for spatial data analysis.

Quantiles

A method of data classification that attempts to place an equal number of values into each class. For example, if there were 50 observations, each with a different value so that no duplicates were encountered during data classification, grouping them into five classes (also known as quintiles) would result in 10 observations per class. Quantiles are useful for classifying ordinal (rank-order) data, and for comparing maps with the same number of classes. Outlier values will be less visible and attention will be focused on relative rankings. The map resulting

from this method of data classification will tend to produce an even distribution of map colors. Caution: If the data is highly skewed, the quantiles classification method will still place the data into the number of classes specified. This forcing of unevenly distributed data into classes containing equal numbers of observations could lead to the false impression that the data are normally distributed.

Quartile

The 1st, 2nd, and 3rd quartiles are the 25th, 50th, and 75th percentiles, respectively. (Source: Statistics.com)

Raster data

Spatial data stored in a computer as a series of values in a grid pattern (pixels). This type of data generally requires much more computer storage space than vector data. Larger numbers of pixels over a smaller area provide greater spatial resolution but take up much more memory. This type of data can show continuous change over a surface, such as land cover. Satellites collect data in this format.

Remote sensing

The science, technology, and art of obtaining information about objects or phenomena from a distance (i.e., without being in physical contact with them). (Source: CCRS)

Scale

The Earth is too large to draw on a map without reducing its size. This reduction is expressed as map scale, which is the ratio of the distance on a map to the actual distance on the surface of the Earth. As a result, a small-scale map displays a small amount of detail, but covers a large geographic area. A large-scale map shows a large amount of detail, but for a small area. Scale can be expressed graphically as a scale bar, or in writing using text or numeric forms: (i) Text: 1 inch = 24,000 inches OR 1 inch = 2,000 feet; or (ii) Numeric: 1:24,000.

Shapefile

A spatial data format originally developed by ESRI and in widespread use today. A shapefile is actually a collection of at least three files: (i) Main file (filename.shp), which contains geometric information for the features of interest on a record-by-record basis. (ii) Index file (filename.shx), which identifies the positional offset of each record in the main file from the beginning of the main file. (iii) dBASE file (filename.dbf), which contains a table of attribute data for each geometric feature described in the main file. A shapefile can also have a projection file (filename.prj) to specify the coordinate system and datum. Although a projection file is optional with respect to the shapefile technical specification, it is essential for accurate geographic analysis.

Spatial data

Data that describe the geographic shape and location of entities in relation to the physical space defined by the Earth's surface. In terms of shape, spatial data can take the form of points, lines, or polygons. With respect to location, spatial data are organized and displayed according to coordinate systems and datums.

Unique geographic identifier

A name or code that uniquely identifies a geographic entity. Examples include province or district name for an administrative region; waypoint ID and latitude/longitude combination for a GPS point; P-code (place code) from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). Unique geographic identifiers are essential for distinguishing between individual geographic entities and ensuring the accuracy of attribute data for those entities, as non-unique identifiers can create confusion and produce errors during all phases of data use. (Source: MEASURE Evaluation)

USB (universal serial bus)

The most widely used standard hardware connection for attaching external devices such as hard drives, flash drives, cameras, or phones to a computer.

Vector data

Spatial data stored in a computer as points, lines, and polygons. In the case of a straight line, the coordinates of one point, the distance and direction to a second point, and the coordinates of the second point will all be stored. This is usually the most effective method of spatial data storage.

Virtual globe

A 3D representation of the Earth that provides the ability to zoom in and out through a wide variety of scales and to change viewing angle. Virtual globes often combine satellite imagery collected at varying levels of detail with actual aerial or even street-level photography. They also often allow additional overlays such as points, maps, or images.

Waypoint

A single point location on the earth's surface, stated using an x/y coordinate pair. These points can be stored on a GPS receiver unit.

XML

XML, which stands for eXtensible Markup Language, is a simple, flexible text format that plays an increasingly important role in the exchange of data on the Web. (Source: w3)

Zipping

The process of compressing a file (or group of files) using an algorithm to store data more efficiently. A “zipped” file is generally smaller than an unzipped one, and is a good way to group files in a package, in order to send them to a colleague.