Map Construction and Interpretation

M. Monmonier (1996) *How to Lie with Maps* University of Chicago Press, London, 2nd ed

A. M. MacEachren (1995) *How Maps Work:Representation, Visualisation,and Design* Guildford Press, New York

---

- A map can be defined as '*a collection of spatially defined objects*'.

- - A display of the spatial properties of an object set.

This usually implies a two dimensional display of the Cartesian or polar coordinate locations of objects and also their attributes, e.g. a street map displays the locations of streets and houses on these streets, (if the *resolution* of the map is high enough). In addition, the houses may have attributes which relate to the population of each household. Hence a variety of maps could be constructed even from this simple example. We could have a simple street map, a more detailed house map and a map of household attributes at the highest resolution. The display of such varied information in a graphical form has been the concern of *cartography* for a considerable time (see e.g. MacEacheran (1995)). Many of the concerns of those within Statistics about the representation of data in graphical forms have also been explored within Geography for mapped displays. The psychological /visual perceptual implications of chosen mapping methods has been studied extensively (Monmonier (1996),ch. 3-6,) and these issues also apply to the construction of maps of statistical information. Walter (1993) and Pickle (1993-1998) have examined visualisation issues related to medical mapping.

The stages of map construction can each be associated with some form of processing of spatial information and hence can be of concern to anyone wishing to use such methods of presentation.

The main stages are:

- - 1) *choice of scale*
  - 2) *choice of symbolisation or representational processing*
  - 3) *further processing required to construct a suitable map.*

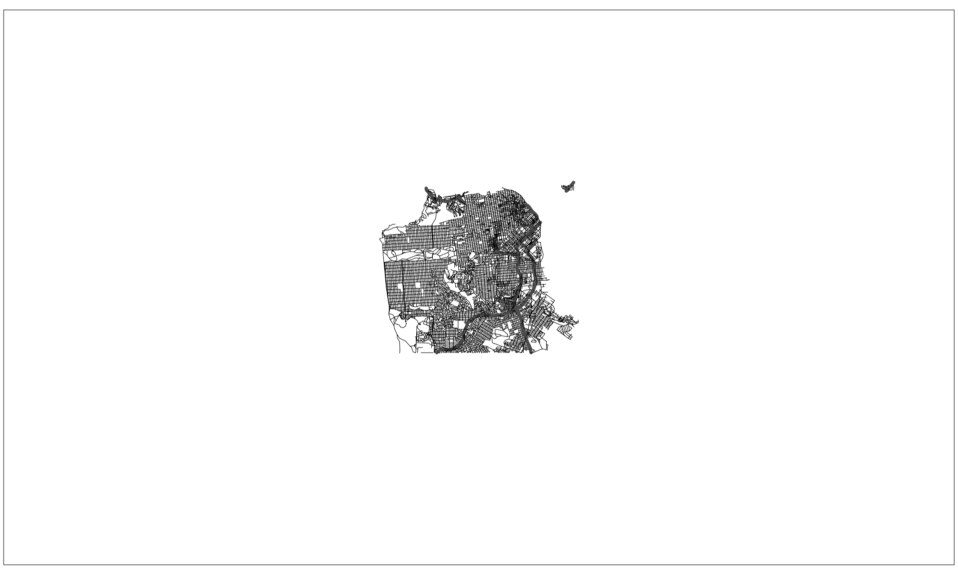
- In stage 1, a suitable scale for the map must be chosen. Any choice of scale, however, inevitably leads to a process of *averaging* of spatial information from higher levels of resolution. For instance, a street map of a city will usually be represented as sets of linear features depicting street locations , but if a larger country scale was to be used, within which the city was but a small part, then the city streets could be represented by a dot(!). Hence in this case, the scale change has resulted in averaging of the spatial information .

- Stage 2, is also represented in the street map example. At the detailed scale, linear features represent the streets while at the country scale, the whole city is represented by a dot. This represents a change in symbolic representation as well as scale change. This both can have a visual perceptual effect for the map user and is an averaging of spatial information.

- Stage 3, that of further processing, can occur when information on the spatial structure of the objects and/or attributes is not available in the form required by the representational system. For example, often we want to or need to compute a map representation from a set of sampling points which are predefined, whereas we need to have measurements at the intersections of a fixed grid which don't correspond to the sampling points. This arises in many statistical mapping problems and leads to the use of *interpolation* or *smoothing* of data. Another example of further processing is the use of *transformations* of the mapped data to represent some feature of the spatial structure. Map projections(Monmonier (1996),ch. 2) are a classic example of transformation. Schulman et al (1988) give an example of using projection and transformation in a medical statistical application.

Hence in two of the three stages of map production, some form of statistical processing of the spatial information usually occurs. This applies in most forms of mapping exercise and hence it can be claimed that map construction is, for a large part, a statistical processing task. The figures (city of San Francisco) display the transition between streets and city level representation. The need for symbol change is also apparent when some scale changes occur

Street level map



City level zoom scale change.

Statistical Maps and Mapping

- Object maps: the study of the statistical properties of locations is characterised by models for the discrete locational information. (Stochastic Geometry). Often these maps of discrete objects are converted to continuous surfaces (contoured) by calculating a measure of the local intensity/density of objects and then mapping this density/intensity
- Geostatistical data: any data which consists of observations on a continuous surface over space can be considered as Geostatitsical data. For example, air pollution measurements or chromium concentration in soils are observed at locations but are thought to be spatially-continuous.

In applications in disease mapping, some of each data type may be encountered:

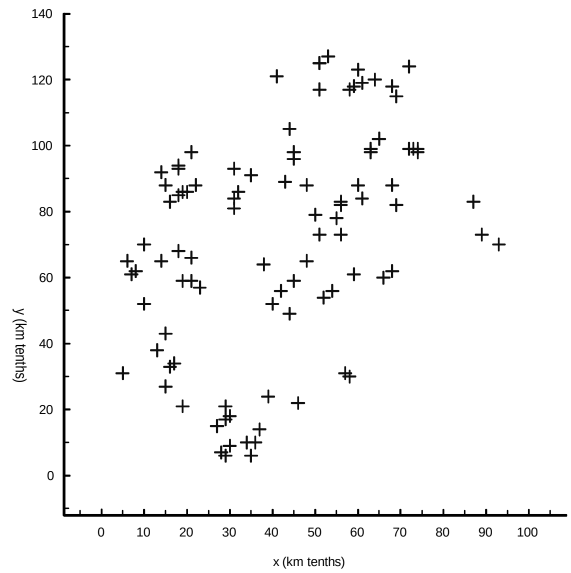
- Maps of case events are object maps

- Covariates which are measured at spatial sampling sites can be regarded as Geostatistical data.

- Count data observed within *fixed* arbitrary administrative regions (such as census tracts) are 'averages' of an object process.

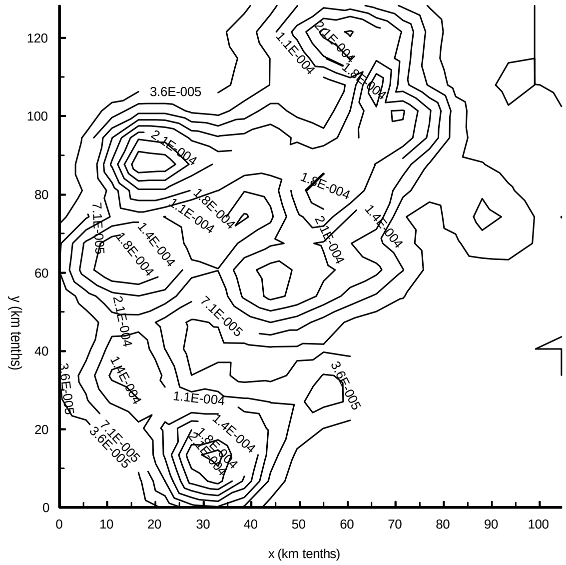
Object Process Mapping

- A simple example has a point location as its observation unit and the realisation of point locations are the objects. For example, the address locations of cases of a disease form a point process and a map of all addresses of disease within the study region would be a mapping of the process. The following figure depicts a case address map for respiratory cancer in a small Scottish town for the period 1966-1976.

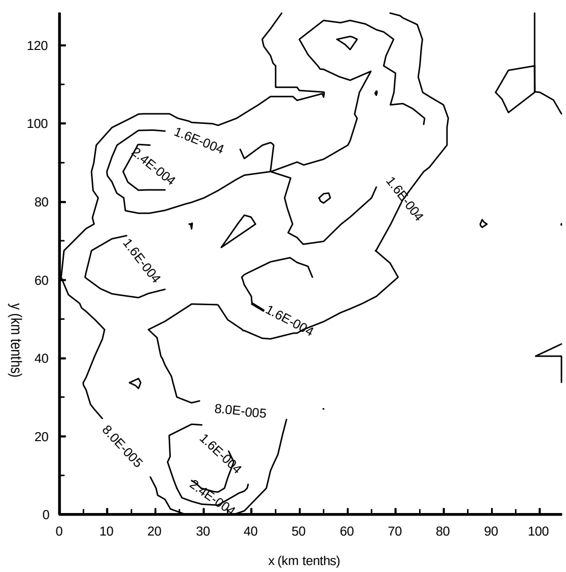


Cases of respiratory cancer

- Density/intensity estimate: Often it is important to transform an object map by converting the object locations into a continuous surface representation of the objects. This kind of transformation can be achieved by computing the *local density* of objects. Density estimation can be used to provide such local densities and the resulting density surface can be mapped over the study window.
- Contouring: Usually such a surface is displayed as a contour plot or, in 3 dimensions, as a surface perspective view. The contour plot is often preferred, as some spatial information is hidden in perspective views.
- Map distortion: To demonstrate how scale and symbolisation affects such mapping, the contour plot of a density estimate of the case event data, has been drawn for two different contour densities (10 and 5 heights). Note that the arbitrary choice of fewer contours effectively produces a smoother surface and can change the perception of the object map.
- Derivation stages: The derivation of these contour maps has proceeded through a number of stages which may affect the final visualisation. First the process of density estimation involves the production of estimates in a grid mesh (interpolation) and the choice of a smoothing constant (bandwidth) which controls the smoothness of the gridded data. Then a graphical package has constructed contours using a further interpolation/smoothing step.



10 contour heights



5 contour heights

Geostatistical Mapping

- Geostatistical data differs from the above in that a network of sites are usually used to sample or measure some spatially distributed variate. For example, the early Geostatistical work related to estimation of Geological structures in mining applications where concentrations of particular minerals were sampled at fixed locations.
- In principle, the basic mapping considerations apply in this case also: for visualisation, the data can be displayed as an object map with each sample site becoming the location of an object representing the measurement at that site. For example, a circle of radius equal to the measurement could depict the distribution.
- Other display forms are available, such as needle plots where vertical lines of length scaled to represent the measurement are drawn at the sites. Often a surface interpolated from the measured data is to be constructed. This surface also requires an interpolation or smoothing step to provide a gridded data set, which can be subsequently contoured. Such interpolation can be achieved by a wide range of smoothing techniques.
- The method of *Kriging* was developed within Geostatistics to provide such processing. Other notable forms of smoothing available for such data are: *nonparametric regression* or *kernel smoothing* and *thin plate splines*.

Map Interpretation

- The visual representation of aggregated count data has been the focus of study for some time. Often the ready availability of aggregated count data for diseases has led to the widespread use of visualisation to depict spatial distributions. Often the purpose of mapping count data is to display the spatial variation of disease so that interpretation of disease variation can be made. Variation of interest to, say, public health workers, could be the identification of ‘clusters’ of high incidence of disease or the isolation/identification of areas of similar incidence. In the first case some public health intervention may result from the identification. In the second case, allocation of public health resources to ‘like’ areas may be the focus

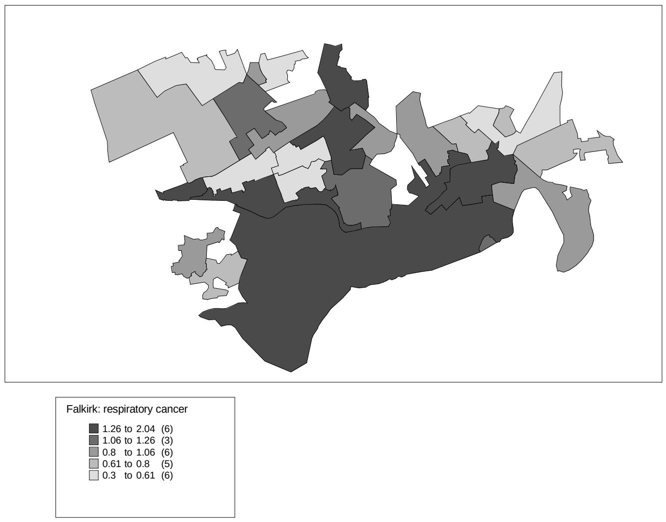
.

- Once statistical processing of the aggregated data has been performed, the resulting map of disease risk (usually relative risk) is often used as the basis of interpretation. Unfortunately, the interpretation of such maps without recourse to additional statistical information relating to estimates and their variability can be extremely difficult. This is akin, in the simpler clinical trials field, to computing the means of a parameter of interest in two dose groups in a trial, and basing judgement of group differences on a visual display of the means. Certainly, this kind of analysis would not pass FDA guidelines!

- The main problem with the use of such maps for these purposes is that the map is a *visualisation* tool, but is being used for an inferential task without recourse to statistical inference procedures. Hence, it is extremely important to present such georeferenced data with all relevant statistical information. At the minimum, any map of relative risk for a disease should be accompanied with information pertaining to estimates of rates within each region as well as estimates of variability within each region. At the other extreme it could be recommended that such maps be only used as a presentational aid, and not as a fundamental decision making tool.

- Some issues relating to disease map interpretation have been studied within cognitive science (see e.g. Pickle and Hermann (1995) *Cognitive Aspects of Statistical Mapping*. NCHS Office of Research and Methodology, CDCP Working paper 18, 321p).

- In these studies, certain aspects of map presentation have been examined in relation to the ability of map observers to detect ‘clusters’ of disease and in their ability to assess given areas of risk. In most of these studies, estimation of observer variability was attempted, but no comparison of observer ability in recovering ‘true’ features were made. These studies have mainly focussed on the construction of thematic maps, i.e. the use of colour schemes or shading to represent the relative risk within regions of the map.


Falkirk: respiratory cancer
1.26 to 2.04 (6)
1.06 to 1.26 (3)
0.8 to 1.06 (6)
0.61 to 0.8 (5)
0.3 to 0.61 (6)

In these studies, it has been established that:

- monochrome thematic maps yield lowest observer variability in detection of areas of risk,
- two colour map schemes have higher variability, but are *preferred* by observer focus groups (of end-users),
- dot density maps tend to emphasise small clusters and yield higher variability in identification of risk areas
- observer focus groups prefer double coloured maps over monochrome maps or contoured risk surfaces,
- these studies also support the use of coloured monochrome maps over grey scale maps.

These results support the use of monochrome colour thematic maps for the presentation of disease incidence. In addition, it has been found that the use of particular colours can reduce interpretational variability. Red monochrome maps appear to be favoured for identification of risk areas. However, it should be noted that although cognitive research has shown that monochrome maps are to be preferred, many observers within focus groups ’prefer’ multiple colour maps. Hence, if atlases of disease risk are produced based on focus group recommendations then these may be sub-optimally designed for interpretational purposes! In addition if an end-user is given a choice of which map to use they are likely to choose a map which is sub-optimal for the purposes of interpretation. This further supports the contention that maps should be used as wallpaper (as observer preference may be for pleasant colour schemes) and not for inferential purposes.

GIS Systems

There are now a large variety of commercially available software packages which provide display and manipulation facilities for georeferenced data. These packages are usually referred to as Geographical Information Systems (GIS).

The fundamental ingredient of these packages is the idea of map layers which contain different information about the mapped area. For example, in one layer might be held the tract boundaries of census small areas, while in another layer some additional information relating to the each tract can be stored and displayed: for example, census small area labels or SMRs or crude counts. Each layer can be manipulated interactively (edited) to provide a composite map. In addition, some packages also provide facilities for selection of sub areas or arbitrary transect displays. The types of display available on the commonest packages is often limited to types of thematic map (choropleth, dot maps etc.), and often contour or interpolation facilities are crude or not available in the basic package. In addition, the ability to handle (point) objects, in a reasonably sophisticated manner, has only recently become available. One *major* drawback of current systems is their lack of spatial statistical tools for analysis of spatial data.

- It is widely regarded that the commonest GIS packages in use currently are MapInfo® and ArcGIS®.

- These packages have been developing over the last 15-20 years and have different market orientations.

- *MapInfo* has as its focus, the manipulation of polygons and their associated data. Hence, small area tract information are well suited to this format, and many business-related applications can be developed with this package. It is also possible to use MapInfo for the analysis of (point) object data via add-on software (e.g. Vertical Mapper®), which can provide interpolated surfaces and compute tessellations. In recent versions this feature has been included.

- *ArcGIS* has focussed on continuous surface modelling and mapping functionality, and therefore finds considerable use in land use assessment and a wide range of environmental applications. It also now has considerable facilities for processing polygons and other forms of spatial data.

- Both packages have links to statistical software packages (e.g. ArcView_S-Plus) and to each other via data transfer facilities, and there are additional facilities which allow user programming of GIS itself.

- However, these packages still await the incorporation of spatial *statistical* tools.

In general both GIS systems have advantages and disadvantages. MapInfo is relatively easy to learn was originally designed for regionalised data.

ArcGIS can handle a variety of data forms. It also has many add-ons which allow more sophisticated analyses (Geostatistical Analyst; Spatial Analyst)

However I have found that MapInfo has a shorter learning curve.

Notes:

- 1) MapInfo and ArcGIS have data exchange facilities
- 2) UK census data (boundary files) can be downloaded from *http://edina.ed.ac.uk* in either MapInfo or ArcGIS formats.
- 3) US census boundary files (Tiger files) can be downloaded from the US census Bureau site: *http://www.census.gov/geo/www/maps/*
- 4) The academic community has favoured ArcGIS
- 5) MapInfo: *http://www.mapinfo.com* ; ArcGIS: *http://www.esri.com*

Quantum GIS

BMTRY 763

Getting MAPS into WinBUGS

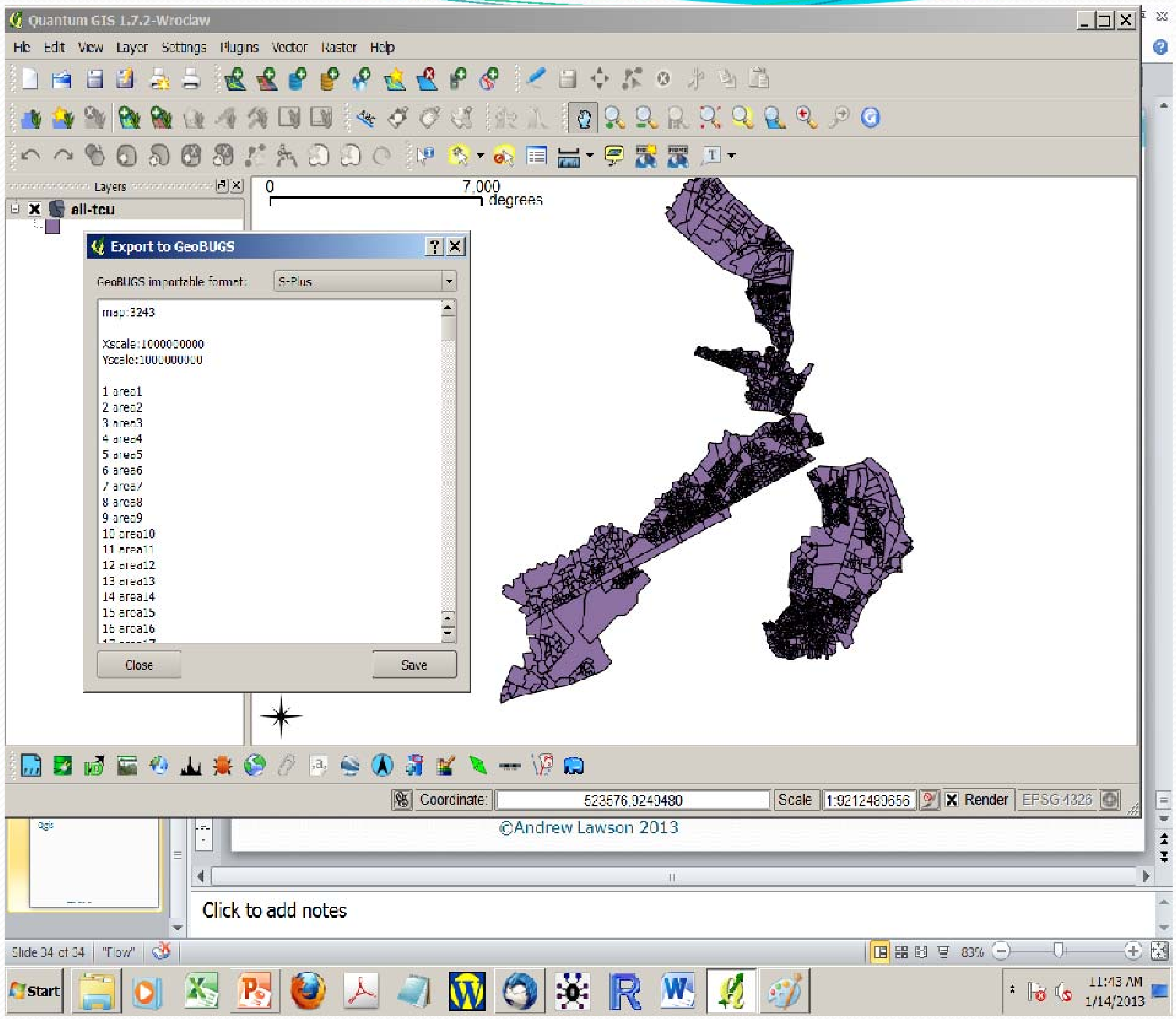
- maps2WinBUGS:
  - http://maps2winbugs.sourceforge.net/
- R functions
- Manual edit
  - ArcINFO format
  - SPlus format
  - Epimap format

QUANTUM GIS

- QGIS now has a plugin for maps2WinBUGS whereby you can create input for .MAP files on GeoBUGS

- http://www.qgis.org/
- http://sourceforge.net/projects/maps2winbugs/

Qgis


Quantum GIS 1.7.2-Wroclaw
File Edit View Layer Settings Plugins Vector Raster Help
Layers
all-tcu
0
7,000
degrees
Export to GeoBUGS
GeoBUGS importable format:
S-Plus
map:3243
Xscale:1000000000
Yscale:1000000000
1 area1
2 area2
3 area3
4 area4
5 area5
6 area6
7 area7
8 area8
9 area9
10 area10
11 area11
12 area12
13 area13
14 area14
15 area15
16 area16
Close
Save
Coordinate:
623676,0240480
Scale
1:9212480656
Render
EPSG:4326
©Andrew Lawson 2013
Click to add notes
Slide 34 of 34
"Flow"
83%
Start
11:43 AM
1/14/2013