GISCO Desktop Mapping guide

Desktop Mapping guide, version 2

February 1999

Geographical Information System for the Commission
Directorate D - Unit D2 Regional Statistics and Geographical Information

Work on this site started when after a long waiting period ArcView has been selected as the desktop mapping software for the Commission. Eurostat and more particularly GISCO has been active in this domain participating in all phases of the selection. This second version of the Desktop Mapping guide shows our continuous effort to support the users of the GISCO database and ArcView mapping software. GISCO aims to promote the appropriate use of maps for visualizing statistics but also to provide non-experts with some basic guidelines on designing thematic statistical maps and how to avoid the most common errors.

It is obvious that in such a difficult task we cannot get it right from the beginning and it will be only your comments that will allow us to correct our errors and improve the quality of this guide. Our e-mails are at your fingertips, please take a few minutes to send us your comments. This site is under continuous development.

The current version has been realized thanks to the constructive comments of all GISCO team members and especially Mrs Natalie Zimmer who reviewed part of the text and contributed in choosing the right maps to use. We wish to explicitly thank Mr Alistair Calder from the Office for National Statistics of UK and Mrs Ulla-Maarit Saarinen from Statistics Finland who allowed us to copy and use parts of StatMapWeb which is the result of a joint project started in spring 1998 between Statistics Finland and the Office for National Statistics, United Kingdom.

Purpose of this site

This site represents our second effort to introduce the concept of maps in Eurostat’s working environment. Statisticians - and not only - need maps to explore and understand their data and to confirm and refine their hypotheses. Beforehand, there is need to understand the principles and limitations of statistical maps, which are not as straightforward as many software products suggest. Our purpose is to provide potential GISCO database users with a quick reference on cartography, thematic mapping and examples of thematic maps. Moreover, we have concentrated on the use of ArcView as the desktop mapping software to be used by all Commission’s users. All examples have been created using this software.

We aim to present some of the main principles of mapping (cartography and thematic mapping) along with some practical suggestions on what to do and what to avoid. Starting from very general cartography we will then concentrate on thematic mapping. It has often been said that maps are used for decoration in publications! We aim to demonstrate that they offer an interesting alternative in validating, viewing and analysing statistical data. As an extension to the analysis of statistical data, our experience has shown how much further, graphical, exploratory and spatial analysis yields benefits. We hope that users will benefit from reviewing our on-line guide and soon create their own maps for their analysis and/or publications. "Confronted” by the voluminous documentation and bibliography on the subject, we had to be very subtractive in order to present the necessary information for a “quick start”.

Why do we use maps?

Maps are a great way of displaying statistical data.

They can present complex data clearly and compactly.
They can be a great help in spotting patterns within data

They are accessible -

people understand maps (or at least think they do)
people like maps
maps attract attention and brighten up presentation
But maps of statistics do present a number of problems

A map always generalises and simplifies information.

Maps can end up as decoration - unless you are careful sometimes the appearance of the map can become more important than its value and validity for presenting statistics.

Information on a map is always interpreted information. Maps can mislead as well as provide useful information. Bad design can provide completely the wrong impression of the data. There is always the risk of unintentionally lying with maps.

Avoiding these problems and making sure that maps inform the reader, release new information from the data and present the statistics in valid way isn’t difficult. It does however require that you are logical, careful and think hard about what you are doing.

What is GIS?

A Geographical Information System (GIS) is a computer based set of tools (a collection of computer software) for capturing - collecting, editing, storing, integrating, analysing and displaying data in association with their geophysical location

Much of the research in GIS has been concerned with computational geometry, spatial discretisation, spatial analysis models, presentation of georeferenced statistical data etc.

Today we suspect that there are only a few - if any - scientists working with multiple data sets, collected in the natural environment, who do not use GIS. Geological, environmental, medical, demographic, market analysis and political sciences applications, just to mention few, use GIS tools.

Earlier manifestations of statistics for spatial data have appeared in the form of data maps. Spatial models appeared much later.

Further need for political decisions based on regional or in general spatial phenomena manifest the need of including GIS in every aspect of statistical analysis. Eurostat has developed statistical methods where time is the principle dimension. Time series are therefore manipulated in order to assist in the validation of data; the creation of series with estimated data; and the calculation of derived series etc. The usage of GIS tools requires the development of similar statistical models, comprising spatial elements in order to carry out similar tasks. For example, the interpolation of mean rainfall, in order to calculate the expectation of crop production, requires an adequate GI statistical model.
There are two principal GIS data-models in widespread use, which are termed vector and raster. They differ in how they conceptualise, store and represent the spatial locations of objects. (More information can be found in the glossary under the terms topology and raster). In GISCO Reference database all available information is in vector format. Nevertheless, exceptions exist. Landcover coverage is in raster format. ArcView gives the possibility to access coverages in both formats.

And in a few words ..............

Economic, social and natural actions and phenomena all have a spatial component. By coupling statistical information with geographical territories we enhance the effectiveness with which they are presented or analysed.

What is a thematic map?

In the past, map production was rather exclusive, but today everyone with a PC and mapping software can. We then use the map to communicate with other people, and we want them to receive the message the way it was meant to be received.

We can distinguish between different types of maps: Topographic-, technical- and thematic maps. A road map or survey map is a good example of a topographic map. Technical maps are those you receive from the technical division in the commune, they describe the border of your site, and where to find technical equipment on your site.

In this guide we refer only to thematic maps. This is a map where we connect non geographical data sets (ex. economic, social, demographic traffic data) with an indirect geographical reference (ex. region code, commune code, road number) to the map. This could be a starting point for future analyses, where the producer and/or the reader want to increase insight into the data set during a cartographic presentation.

Thematic maps take their bases from existing topographic maps but they are distinguished further by the subject matter which usually is not the physical earth or locations upon it. The subject may be some distillation of physical phenomena, such as average annual temperature or precipitation values. Commonly, though, the subjects mapped are both abstract and non-physical, like crude birth rate per thousand inhabitants.

The concern of thematic mapping is for a sound presentation of the essence of some distribution. We consider a thematic map as the primary component of any spatial analysis, presenting statistical information on "how much" or "how many", but also "where" a phenomena occurs.
A strong sense of “visual logic” is vital, and a knack for choosing the right words to accompany the graphics is equally important. Thematic Maps - Their design and production by David J. Cuff and Mark T. Mattson

Introduction to mapping concepts

In order to create a complete map, several important mapping concepts should be followed, such as:

map features

map characteristics

structure of a thematic map

Beside the choice of the right symbol (point, line, area) describing a specific theme analyzed, the map characteristics (projection, scale,…) are essential in order to form the base elements of a map.

As far as the structure of a map is concerned, certain elements such as title, legend, … are absolutely necessary to create a clear elaborate map.

In the following paragraphs all these ingredients, making a successful map, will be explained more in detail.

Map features

The information conveyed by a map is represented graphically as a set of map components. Location information is usually represented by points, lines and areas.

Point feature

A point feature is represented by a single location. It defines a map object of which the boundary or shape is too small to show as a line or area feature.

Line feature

A line feature is a set of connected, ordered coordinates. It represents the linear shape of a map object that may be too narrow to be displayed as an area, such as a road, or a feature that has no width, such as a contour line.

Area feature

An area feature is a closed figure whose boundary encloses a homogeneous area, such as a state, country, soil type or lake.

Map characteristics
Map projection

Each map projection is the location framework of a thematic map. It is a systematic arrangement of the earth’s meridians and parallels onto a plane surface. We have got different types of projections, but each generates automatically some distortions of the area, distance, shape and direction. There is no transformation process which can completely eliminate simultaneously all these distortions. So the user has got to select the most appropriate projection depending on the map’s message.

Map scale

Map scale is the extent of reduction required to display a portion of the Earth’s surface on a map. It can be expressed as a representative fraction, which is a ratio of the distance on the map page to distance on the ground. Larger scale maps show features in greater detail but represent less area. Smaller-scale maps show larger area but represent less detail. It is important to remember that only maps of the same scale should be used as overlays. Maps in different scales serve different needs. By no means a map of 1:3,000,000 should be used in order to depict the location of wells in a region. However, it is adequate for presenting the ports of France on an A0 format poster.

Map resolution

The resolution of a map is the accuracy with which the location and shape of map features can be depicted for a given map scale. Scale affects resolution. In a larger-scale map, the resolution of features more closely matches real-world features because the extent of reduction from ground-to-map is less. When using a map we should always think about the scale and resolution.

Other map characteristics are the map accuracy and the map extent.

Structure of a thematic map

A complete map contains 5 elements: Title, legend, scale, textual information, and the actual map. A map should be as self-explanatory as possible, so that a reader immediately sees what the map is all about without consulting the legend (e.g. for quantitative data). This is obtainable if we follow the visual rules according to the thematic information and the used visual variables.

Title

The title should identify which theme variables are involved, what the map is all about. Very often we need a long title, and in this case we should use a short main title and a subtitle. The subtitle should contain information about the area the map covers and in the case of statistical information the reference period of the data. An indication of the NUTS level showing the breakdown of the regional data is obligatory.

Legend

The legend should identify each of the theme variables used in the map as well as which visual variable corresponds to which theme variable. In simple words the variable which has been used for mapping should be explicitly stated and not mixed with the "Title" of the map i.e what does the line or point show?; what is the difference between the blue and the red line etc. The unit of measurement of the variable is obligatory.

Scale

The scale is one of the most important elements on a map. So scale selection has got an important consequence for the map’s appearance and its potential as a communication device. On a map, however, use a graphic scale bar rather than a numerical scale (1:50 000), because any reduction of the map will not correspond anymore to the reality.

Textual information
This could be subtitles or footnotes connected to the map as well as a declaration of the statistical and geographical data sources, the date of production of the map itself and the geographic orientation (N). Orientation need not always be shown by an arrow, you can also use the graticule (parallels and meridians) or grid ticks. In case of statistical data from different sources, estimates or with a different reference period should be explicitly declared. Any exceptions of the NUTS nomenclature should be mentioned.

The actual map

This is the thematic map produced from geographical information (ex. NUTS boundaries, commune boundaries etc) and statistical information for the NUTS regions, communes etc.
Visualization of information

Visual variables

The presentation of information is a question of using the shape, the size, the orientation, the value, the texture and the hue - called visual variables - of objects (triangles, circles, rectangles etc.) in order to visually communicate a message.

Bertin recognised two kinds of visual variables: retinal variables and locational variables (namely horizontal and vertical co-ordinate axis of two dimensional data graphs or longitude, latitude in the case of maps).

Cartographic treatments of Bertin’s visual variables commonly ignore the locational variables as having little to do with symbolisation.

Variation of shape

There exists an infinite number of shapes. The variety of shapes is, above all, related to characteristic of a point (center of a region, port, city, airport etc). The changes of line shapes are very few and in a zonal layout, the "point" elements distributed within the zone are those that vary.

The variation of shape neither expresses an ordinal relation nor a quantitative change. The variation of form has an associating quality; it has a very small (restricted) differential character. The rapid selection (choice) of different forms depends on the number of symbols used, their density and their selective capacity. Hopefully, you can distinguish the triangle, the square and the circle on the map of France corresponding to the qualitative characteristics of the regions.

Variation of orientation

A line can vary in its orientation. It can be vertical, horizontal or follow any other axis. The change of orientation is best used in a ponctual implantation. Its undeniable efficacity is nevertheless restricted on a differential level and to the maximum use of four directions.
The variation has a poor (limited) differential possibility in a zonal layout. On the other hand, when combined with variables of quality (value) and texture, the variation of orientation enhances (increases) visual differences (contrasts).

Variation of hue

Hue is that aspect of colour associated with wavelength in the comparatively tiny portion of the electromagnetic spectrum visible to the human eye. So hue is the name we give to the various colours we perceive. The different colours, ranked from blue to green and red, do represent a certain order. The human eye actually perceives the variation of value before the change of colour.

Variation of texture

The size of elements forming (making up) the texture may vary without any change of the white/black relation. It is variation of texture. The perfect variation of texture is achieved by enlarging or reducing by photography the size of a given texture. In this case, the variation of texture is orderly (ordered) and selective.

The considered variation of texture is expressed by a more or less large fineness of the structure elements of the given plane. In this case we speak of fine grain and of thick (coarse) grain. The combination of coarse grain to a variation of shape is particularly selective but not orderly. So there is no change of value. The double quality (propriety) of selectiveness by using coarse grain of different shapes and order by variation of texture inside the same shape, allows to represent opposite phenomena varying on both sides of a common origin.

The variation of texture is best represented in a zonal layout and remains restricted to a limited number of levels (three more or less) in a punctual or linear layout.

Variation of value

The concept of value is expressed by a relation between quantities of black and white on a given surface. The variation of orderly (ordered) value. The continuous progression goes from white to black by passing through an infinity of intermediate greys.
The change of value is monochrome. A variation from pure red to white, or from pure green to white is a variation of value but not of colour.

The variation of value is most effective in zonal layouts. In a punctual or linear layout, it is difficult to distinguish between more than four changes of value without having to give a considerable surface to the point or line.

The variation of value is also differential, but it is not quantitative. It is not possible to assign a quantity, or even a quantitative relation to grades of different values without resorting to a legend. The change of value permits to say "more" but not "how much more".

Variation of size

The variation of size corresponds to a change in length or of surface. In a zonal layout, the surface of the zone does not vary, the change of size relates to the elements, punctual or linear, inscribed (entered) in a zone.

A variation of size is differential and ordered. But, above all, the change of size is quantitative. It is the only variable that expresses quantities properly.

Choosing and using colours

Visual variables

The choice and use of colours is one of the most important elements of map design. Good colours help make a map useful and attractive - the wrong colours can make it unusable and ugly. Although what makes for a good choice of colours is subjective - people see and have different reactions to different colours - there are a number of considerations which can guide our choice.

**Colour association and the meaning of colours**

People see colours differently and have different reactions to colours. However it is important that you think about how the user is going to interpret and react to the colours and also to consider the associations that users will make with certain colours. It is quite difficult to make rules in this area but here are some ideas .....  

In general it is a good idea to use darker more intense values for high values. This means that the dark areas on the map are those where values are high. This normally gives the right overall impression.

Try to pick a colour which will have the right association - here for example live births are normally a good thing so a bright colour has been chosen ....
Mortality rate is always bad (!) so here a duller, darker colour has been used. High mortality areas will have a but dull colour.

If you can, pick a colour with a real association......be aware though that these can vary from country to country but these ones should be safe enough...

However things sometimes get more complicated ..... In this case high values of literacy will appear dark and dull on the map .... this isn't quite right ....

One trick that you sometimes see in this case is the renaming of the subject illiteracy (just take the percentages away from 100!) - now the map will work fine - dark values are for high values and the worse areas of the map will appear dark .....

However this is clearly a difficult area - there is a fine line between helping the reader of the map and making moral judgements.

The shades on the right are a summary of the "good practice" shades. However, the lowest shade should not be completely white since it is reserved exclusively for the non available data (no values, confidential etc). More examples of shades taken from GISCO Maps request guide:

Black and White shades
Colour shades for 4 ranges
Colour shades for 5 ranges
SUMMARY

Think about colour association and use it to your advantage.

Simply stated:

Use bright, nice colours for good things.

Use dark and ugly colours for bad things.

Normally use high values of the dominant colour for the higher values.

Just try and give the right impression when the user first looks at the map.

This is a tricky area - sometimes we are in danger of making moral judgements about what is good and bad - and these may vary from culture to culture - just be aware and take as much care as you can.

**Colours must be clear and different**

The key role of colours on a statistical map are to distinguish the values of different areas. It is critical that the user can read colours clearly from the map and identify matching colours in the legend.

This clearly means that there is going to be a limitation on the number of colours that can be used. With black and white maps it is possible to produce 3 or 4 different distinguishable greys without using patterns. (See the section on B&W maps for more details and suggested ranges). If you try and use any more the distinction between areas will almost certainly disappear.

With full colour printing a total of between 15 and 20 distinguishable colours are possible - though you should avoid doing so if you can!

In either case, however, a good deal of care must be taken to ensure that the colours really will be different when produced on the final output device - carry out tests to make sure colours work together and critically that the differences can really be seen. If you are using an external printer you should be aware that the results you obtain will be different from those on an inkjet or laser printer. It is possible to obtain a much more subtle variation with professional printing but the results will be different - if possible get hold of a colour chart produced on the final printer and pick your colours from that.

If you find that the limitation on the number of colours imposed by the production method is a problem think very hard. If you need more than 15 colours you are almost certainly moving towards a badly designed map. In most cases you do well to aim for 5 or 6 distinguishable colours - if you can't get the message over with this many colours you are probably not trying hard enough.

If you really do need to display more colours than these you will need to resort to splitting the data onto two maps, reclassifying the source data or introducing patterns on the map.

Maps on screen are a particular issue here - if the map is to be viewed on your own screen the only limitation must be whether the colours look different enough to be distinguishable - again perhaps around 15 significantly different colours are possible. If, however, the map is to be viewed elsewhere - say as part of a CD product or on the Internet it is critical that things are kept much more simple. Different setups and pallettes on different machines will mean the way the colours will be displayed are liable to vary. It is a good idea to stick to the 256 colour pallette (or better the 216 colour pallette supported by Internet Explorer and Netscape) and keep colours as simple as possible. Again perhaps 5 or 6 is a good target number.

Right ... now we have identified the limitations - the fact that the production method and the need to maintain the difference between colours limit our choice. The rest of this section is more positive and covers the more interesting area of actually choosing some colours .....
Choice of what colours work well together is subjective - people respond differently to different colours and have different tastes and cultural associations. Some people are good at picking colours, some not so good. If you are good, congratulations, if you are not perhaps some of these hints will be of value ....

Use colours to reflect the data

Colours are one of the key elements in providing an overall impression of your data.

The simplest case is where there is no order within the classes of data - for example if your map shows different types of landuse, vegetation or ethnic origin. In this case it is reasonable to pick a range of different colours (though probably with a similar visual intensity).

If, on the other hand, you are producing a choropleth map and your data ranges from high to low you should aim to reflect this in the data by choosing colours which vary in intensity from dark to light or strong to weak. In most cases it is best to choose a range within a single colour - like from dark green to light green.

It is a good idea to try and make the strength of colour reflect the ranges - If one range is significantly higher than the others make this colour significantly darker - you can be very subtle this way. A range within a single colour is normally a better solution than a range from one colour to another - like from blue to green or red. The reason for this is pretty obvious. It is clear at a first glance that dark green represents a higher value than light green but with green and blue the relationship, and which represents the higher values, is much less obvious.

In addition if you use two colours you are implying that the two ends of the scale are fundamentally different. Of course, there are cases where a range from one colour to another is the right decision. If the values at each end of the scale are fundamentally different - like increase and decline in population - or if you particularly want to highlight two ends of a scale - it is right to use two opposite colours. Normally you will want the colour to change at zero or at the mean value ....

...... notice by the way that red and blue work well together to represent increase and decrease because of our old friend ...... colour association.

One thing that you really should avoid, though lots of GIS packages, notably ArcView, like to offer them, is a range of colours which wanders through a spectrum of colours.

This choice is fine to represent distinct values as on a landuse map (and can be great for remote sensing or visualising spatial models) but is quite unsuitable for representing a range of related values on a choropleth map. Avoid ranges with colours in an apparently random order ...

... this type of range will represent all of the values correctly but can only obscure any patterns in your data.

Systems for defining colours

Most software allows you define colours in variety of ways. These are the most common:

RGB - Red Green Blue - the colours which are displayed on screen
CMYK - Cyan Magenta Yellow Black - the four colour plates which make up conventional 4 colour printing

HSB - Hue Saturation Brightness - change just the hue to produce a range of colours of a similar weight

Most books on cartographic production contain details on how these systems work - have a look if you really want to know more.

With experience these numbers can be of value and you can use them to subtly adjust colours and pick ranges of colours which work well together. However it is essential to use your eyes rather than the numbers which define colours to tell you what colours work well together. If you are not sure ask someone else to have a look. Once you have set of colours you like it is a good idea to write down their values for the next time - or if your software allows save them as a palette.

Sometimes colours are defined in terms of colour systems like the Pantone system which allows precise matching of colour during the printing process. If your maps are going to be printed using such a system talk to your printer of someone who knows something about this for some advice.

Testing out colours

There is no substitute for trying colours out together. If you can, print out a sheet with all of the colours you are considering or get hold of a colour chart from your printer and pick from that. Don't ever trust the colours you see on screen (unless this is how the map will appear!) - print out a proof and check colours that way.

So what colours should you pick?

Bold bright colours are sometimes invaluable on maps. They can be useful for maps on screen and are great for picking out small areas or symbols. However, unless you are confident with what you are doing, you are normally safer with a range of lighter colours. Use bold, strong colours on lines, graphs and symbols and weaker, though different, colours for background areas.

It should be obvious but pick attractive colours - do your best to avoid using 'muddy' dull colours - unless you want to give that impression for these areas. Try an pick colours which work well together - red and yellow, blue and green. Avoid colours which obviously clash - like green & red - unless you want the areas to contrast or stand out.

Even when picking individual colours always think about colour association - avoid using blue for small areas - they will look like lakes. Avoid picking colours which will cause an international incident - think about cultural and religious significance of the colours you pick.

As ever, if you get stuck ask someone else what they think about your colours - they probably won't like them because everyone has different ideas but you might get some good suggestions.

Patterns

In the past patterns were often used on thematic maps - largely because they were easy to produce using conventional techniques. There is much less need for them now but there are still occasions on which they can be invaluable. If the map has to be printed in a very limited number of colours or has to be photocopied patterns can be great for extending the number of categories. They can also be the only way of identifying complex overlapping areas.
The disadvantage of using patterns is that, unless used very carefully they can look really ugly. For this reason it is probably best to avoid the using them unless you need to. ........ but when you do ..... try and avoid using just the angle of lines to distinguish between patterns. These look fine in the key but can be difficult to identify on the map and (worst of all) can end up looking very ugly indeed. This type of symbolisation will certainly not provide the overall impression of the data that we are trying to create with our map ......

If there is a variation in value from high to low in the data being presented it is much better to combine different angles with density to provide a proper impression of the relative values. Using variation in density is generally a good way of distinguishing between patterns. Use denser patterns for higher values.

If there isn't any variation in values, like on a landuse map, it would be better to use different types of patterns in order to avoid giving a false impression of the relative importance of different categories. See the map on the right as an example.
If you can, think about using colours and other types of patterns to help distinguish clearly between areas.

However - it is probably best to avoid very graphic patterns. This type of symbolisation is commonly used for geological or agricultural mapping but unless there is need for it can look very messy on statistical maps.

1. Professional colour printing

The number and range of colours available to you if your map is being professionally printed will be dictated by the number of colours (or number of printing plates) being used for printing. If the map is being printed in black and white (just one plate) you can probably manage 3 or 4 significantly different shades (see the separate section on black and white maps for an explanation of why your range is so limited)

How the number of printing colours limits the range available to you

If you have two colour printing available for your map - this normally means it will be printed in black and another colour (orange in this case) it is possible to produce 4 or 5 colours just by producing shades of the orange colour. (Incidentally the percentages suggested on the left are reasonable choices - there is an explanation of why in the section on B&W maps)

However it is possible to increase the number of colours available by using shades of grey as well as of your second ink colour as here .....Even with only two inks it should be possible to get to around 8 or 9 significantly different colours if you really need them.
With 3 colours (here black, blue and pink) much more sophisticated mixes of colours can be attempted and so the range of colours possible extended. However take care - this sort of thing takes practice. If you do want to mix colours like this is is a very good idea to try and get a test colour chart to pick colours from.

If you have access to full 4 colour printing (normally cyan, magenta, yellow and black - called 4 colour process printing) then you can mix almost any colour you want - the only real limit is the number of colours that can be distinguished.

In some cases maps might be printed in more than 4 colours - with individual 'spot' colour inks for each colour - if you are in this position you need to talk to your printer about what you want to do.

The advantages and disadvantages of professional printing

Professional printing allows output to a very high quality - lines will be sharp and clean and colours are clear and even. However proper printing is obviously more expensive than printing proofs on your own printer - and you have less control.

Because of the different way it works there is always a degree of variation between the results you will get from your own colour printer and the final printed results. Just try and allow for this by making the percentage screens you choose for your colours as different as possible - a variation of 30% in at least one colour is a good idea. And do not expect the results to be the same as on your own printer - this is the one case when it is better to use you head (and experience) as well as your eyes to pick colours.

Do not expect an external printer to be able to print a finer screen than 10% - it might not print at all and you will not be popular. In general it is best to only use steps of 10% when defining colours - this is safest since the film imagesetters upon which the printing separations are created will probably be calibrated for these values (but not for example for 43%).

GISCO has a printer which can be used for calibrating your colours. The principle that this printer works is the same as the colour separation principle that most printer houses work when using Cyan-Magenta-Yellow-Black colour scheme, while producing the films. The quality of the shades is the closest possible to the one a printer house can produce. Speaking from experience it advised that after delivering your material to a printer house, to ask for an example produced by their films of all maps in order to verify that the quality of the colour shades is as expected.

2. Maps produced on laser or inkjet printers

In some respects this is the simplest case - with a modern colour printer it is possible to create thousands of different colours. So there is no real limitation from the production method.

Critically if you have access to the printer it is possible to produce test sheets to see how colours work together. Why not produce a sheet of colours which work well together and use these as standard colour sets when designing new maps. Those test sheets should be tested on GISCO's special printer.
3. Black and white output

Often it is necessary to produce maps in just black and white - either because the map is to be included in a simple publication, or because you can't get to a colour printer. However, this type of map presents a particular set of problems - producing good, attractive maps in black and white is a real art.

The main problem here is that it is very difficult to get the range of 'colours' and symbols you need to show several different things using only black ink.

The key rule here must be to keep things as simple as possible - reduce the number of ranges as far as possible and think about alternative ways of displaying your data - proportional symbols for example will normally work fine in a single colour.

Shades or patterns?

If you do need to produce a choropleth map or one where several different colours are required the key decision is whether to use shades of grey or different patterns to symbolise the values in different areas.

If the map is fairly simple and you can get away with a relatively small number of different shades it is probably best to use shades of grey. This is the more attractive option.

If you need lots of different area symbols or if the map has to be photocopied it might be best (but not advisable) to make use of patterns.

Choosing greys

If you decide that you want to use different shades of grey at first sight there might not seem to be too much of a problem. Most mapping packages will allow you to adjust shades of grey at intervals of at least 10%. So it is possible to show 9 or 10 different colours of grey.

NO IT IS NOT - because it is essential that it is possible to identify the differences in the shades and be able to identify individual values in the key you are much more limited than this.

Unless you are very sure that the differences are clear enough when printed you should not attempt to produce more than 4 colours of grey. It is probably better to use just 3 if you can (particularly if there is a possibility that the map will be copied). In addition the 'colour' of solid black (100%) is not normally available for area fills as it will make it impossible to see the lines that define each area (which are normally in black).

If you want 4 different greys and are in doubt about what percentages to pick try using: 10%, 30%, 50%, 70% Black

This seems a strange choice - a more obvious one would seem to be 20%, 40%, 60%, 80%. However the suggested range takes account of the fact that on most printers 10% black looks significantly different from white and that shades of 80 or 90% can often fill in and print like a solid. (This is true of any colour - so it is probably best to avoid the very highest percentages in any colour) If you want 3 greys why not use 10%, 40%, 70% (or 20%, 50%, 80% depending on the printer).
As with all other colours the suitability of your choice will depend upon how the map looks when it is printed. Try your choice out by printing a proof if you possibly can.

Patterns

Patterns provide much more freedom in terms of the number of different symbols that can be managed in a single colour and have other advantages for mapping. They deserve some discussion in their own right.

Black & white maps - Summary

Black and white maps are difficult but here are the key points:

keep the map as simple as possible - if it gets complicated split it into two maps

think about different ways of symbolising data which might work better in black and white - like with a proportional symbol or graph map

use simple and clearly different shades of grey

or if you must - use patterns (but use them carefully)

4. Maps on screen

Maps on screen are a completely different matter from those that appear in a publication. The different medium (and particularly the resolution) means that it is often best to have several simple maps which you can switch between on screen rather than one complex and illegible one.

The colour palettes available are also completely different - and the final results will be influenced by the different machines on which the maps are eventually displayed. You are best to design with the simplest (256 colours) palette in mind and keep things bold & simple.

If you are designing for yourself - just visualising data on screen - do whatever you want - as long as it works for you and you can see the difference between colours that is fine. If you are designing for an intranet or the Internet you need to be much more careful ........

**Influence of production method on choice of colours**

I know it's boring but ...

Before begining to think about choosing colours we need to be aware of the limitations placed upon us by the production method we are using to print our maps. Think first of all about the final product:

Is it to be viewed on screen or printed out?

Is it going to be printed on a black & white or colour printer?

Is the map going to be included in a publication or produced by a professional printer?

To some extent our answer to these questions dictate the range of colours we can choose from.

Summary on colours

Before you start think about the method of production - this limits the range of colours available to you.

The first role of colour is to identify areas and their values - don't try and squeeze too many colours onto your map and keep them different. Always print them out if you can to check.
Think hard about colour association and the effect colours have on the user - does the map give the correct overall impression of the data?

Try and make sure that colours are properly balanced and work well together. Unless you are confident with what you are doing keep colours light - remember that strong, bright colours will be dominant - use them for symbols or small areas.

Try your best to make colours look attractive together - and if you are not confident get someone else to look at them.

A strong sense of "visual logic" is vital, and a knack for choosing the right words to accompany the graphics is equally important.

David J. Cuff and Mark T. Mattson: Thematic Maps - Their design and production
Choice of cartography - Types of information

The first real decision we have to make in designing a map is choosing what type of map to use - referred to as choice of cartography. This is an important decision and one which is often not given enough attention. For reasons that will become clear later in this section by choosing the wrong type of map it is possible to completely misrepresent your data and most software packages are happy to assist you in making this mistake. Regardless of the type of the information sometimes we might feel more comfortable and experienced to use a type of statistical map which is not suited at all for the data.

The list below is an attempt to identify the main types of data from the statistical as well as their geographical reference, point of view

Always remember: The choice of cartography is dependent upon the type of data

1. Qualitative data with:
   points e.g. position of nuclear power stations
   lines e.g. rivers, roads
   area e.g. communes, counties

2. Quantitative data either absolute values or ratios, with:
   points e.g. waste water treatment plants capacity in communes
   lines e.g. traffic capacity of road network
   area e.g. population density

3. Ordinal (or ranked data) with:
   points e.g. high-, medium-, low capacity of waste water treatment plants
   lines e.g. high-, medium-, low capacity of road network per type of road
   area e.g. high-, medium-, low density regions

Qualitative data with points.

Symbol map.

A symbol map is simply showing qualitative characteristics of geographical positions represented as points. If all the symbols are equal we talk about a localising map, and if the point symbol shows more than one quality in each point it is called a diagrammed map.

The variation in shape and colour are the main visual variables. Point symbols could be divided into three different kinds: geometric, pictures and associative symbols. On this example we have used geometric symbols, where the crosses for example could present railway dense areas, line could be industry sites and points could be harbour sites.

Geometric symbols.
Geometric symbols could be for example circles, triangles and squares. Their meaning may not necessarily be obvious without an explanation. This is found in the legend. Note that it is easier for the eye to separate geometrical symbols from the map, than pictograms shown below.

- •
  - ×

Picture symbols (pictograms).

A pictogram is a simplified drawing of the object you want to present. It may be simplified in such a way that it may be possible to see what it means without explanation. These symbols are used for displaying information on touristic maps and drawings more than information on statistical maps.

Associative symbols

Associative symbols are even more simplified than pictograms. A number of geometrical symbols are combined to produce these symbols. We can associate what the symbol means.

Qualitative data in line format.

The basic map contains "line" elements: coastline, rivers, borders etc., so if qualitative data are to be presented on line format, these lines should be stronger than the basic line elements in the map. For example, in a map we can use solid line for roads and dotted lines for railways or as in the example on the left blue, red and green for existing, under construction and planned roads respectively. The width of the lines in this case represents a quantitative characteristic i.e traffic, or ranking of the roads but not a qualitative one.

Qualitative data in surface (area) format.
A distribution map shows the distribution of different themes in a map. For example it can be used to identify different kind of forest (oak, pine etc.). It is therefore possible to show where wood could be found and which kind of wood is inside this area. One way of doing this is by using one texture for forest, and different colours for different species of tree. If we want to show the distribution of pine wood and the distribution of an animal species, we could use two different textures for the different species. Our example uses in raster format different colours to show the qualitative characteristics of each grid cell.

The example on the left shows different areas of the New York City indicating with brown colour the business centers, with purple colour the commercial centers, with blue other commercial zones and green the parks.

Quantitative data with points

A different kind of map exists for visualisation of quantitative data sets. As for qualitative data there are three kinds of symbols: point, line and surface

Maps with area proportional circles

The circled area is proportional to the size of the data value. Area proportional circles are suited for data sets with wide ranging values. The size of the circles should be calculated so that the area of these circles does not cover more than 5-10 % of the area of the map. A circle which represents all the values in the data set should not cover more than 10 % of the mapped area.

Do not arrange area proportional circles according to classes. This will reduce the information value.

The following example shows the number of inhabitants for the agglomerations of New England. Circles are proportional to the number of inhabitants.
Dot density map

A dot density map shows quantitative data connected to areas according to the density of dots. The dots are of equal size and represent equal values. This example uses the same data set as the map with area proportional circles (just above). The size of the dots should be calculated so that in the most dense area of the map they appear to almost float together. The dots are distributed in the area they represent. If it is known how the data are distributed within each area, the dots are placed accordingly. The number of dots increases where density of the data is greatest. In this way the dots are connected to a localised point. Dot density maps could also be produced as a combination of dots and area proportional circles. In this example, an area proportional circle is used in the most dense areas of the map. Dot density maps may often lead to mis-interpretations. It was relatively rare until technology made it easy to produce and it is now offered by most mapping packages. However this type of map is very dangerous (or as dangerous as maps get) and in fact so open to misinterpretation that is probably best avoided. It doesn't really add anything that you can't get from the other types. SO AVOID unless you know what you are doing.

The following example is a dot density map representing the population distribution. One point corresponds to 500000 inhabitants. The bigger circles show major agglomerations in the map extent of the map.
This map shows data in line format using lines of different width. The width of the line is proportional to the data value of the point on the line. This kind of map is most frequently used to present roads, railways, rivers etc. The map on the right could also show the number of people transported on different road segments in the road network during one day.

The following example shows the traffic flow of main road axes (additional information on airport links exists as well) in France. The width of lines is proportional to the mean annual traffic flow.

Quantitative data on area form

Choropleth maps

The name originates from the Greek noun "choro" for the area, and "plethos" for the number or the amount. A choropleth map presents areas where density is dependent upon the class of the data value for each area. The choropleth map represents quantitative data with use of a ranged variable density, and reduced in this way, the information level of the data set.

The readability of the map depends upon the range between the least dense (lightest) area and the most dense (darkest) area. Choropleth maps may only be used for ratios or proportional values, for example percentage, per km, per inhabitant. This kind of map cannot be used to show absolute values.

This is the most frequently used map, and we have a lot of information on how this kind of map is perceived. The map specifies average values for the area, and does not show the variance within the area. This could give big differences in density from one area to the next, even if the differences are not so big. In a "correct" map, the areas presented on the map should be of equal size. Different sized areas give different impressions of values, and could therefore give the wrong impression of the average value for the total area.

For example, communes of large areas and low population densities cover a bigger part of the map than communes with small areas and high population density. This creates a false impression of the real situation. Refer to the bottom of the page to have a look for another problem as well ...

Dasymetric map. This is a choropleth map where the area units have been manipulated according to knowledge of the geography. "Dasy" is Greek and means dense or density and "metro" means measuring. For example, areas with zero population density (geographical knowledge) are known, these could be deleted from the map. The map will then show areas where people live, and how densely populated these areas are. In this example the dasymetric used reduces the visual dominance of rich communes with low population density.
Square net maps

This is a Choropleth map where all the areas are of equal size and shape. Because all the areas are equal, the map could be used to present absolute values. By using squares it is easy to aggregate into different levels. Note that the squares along the coastline have been cut through by the coastline and thus have less land area than the cells entirely on the land. This will effect the value of the cells in the case of calculations. The squares can be replaced by triangles and hexagons though these do not appear often in the commonly used software.

Ordinal (or ranked data)

Depending on the message of the map, the intention of the author and the type of the data to be mapped; ranked data are presented accordingly. Ordinal data result from quantitative data showing a specific message not present when using quantitative point, line or area cartography. A map presenting qualitative information such as road type - presented as double line, single line, dashes, dots - can be enhanced by presenting information on the capacity. Accordingly high-, medium- and low capacity road segments may be coloured with red, orange and yellow respectively.

The following example shows points ordered according to the population size (small, medium, high) in the agglomerations while the shades in the background show inhabitants per square kilometer. The different size population groups correspond roughly to the importance and functional level of the cities.

The following example represents ordered French domestic air traffic lines. From thin to wide they represent low, medium and high traffic flows.
The following example, though very similar to a choropleth map, shows ordered (5 categories) unemployment ratio in some European countries. In fact, there is no difference between this map and its choropleth version except that the legend comprises different explanations. In this case we do not know "how much" less or more is the unemployment ratio between the different classes.

So how do you choose what type of map to use?

Making the right decision here is a matter of experience and common sense but the decision will be dictated by nature of the data, the audience and the purpose of the map. Sometimes the nature of the data or the purpose of a map will dictate a particular approach.

If you have accurate point data a dot location map will probably be best - perhaps with proportional symbols or graphs for each point. If you have values for lines or flows a map where different colours or thicknesses of lines or arrows is the obvious choice.

Remember: The nature of the data or the requirement of the map will dictate the type of map which is possible.

However, it is not always so easy ....Many statistics presented on maps relate to areas rather than points or lines and here the choice is a little less obvious.

The two main ways of representing data are by Choropleth (coloured areas) or proportional symbol maps and in most cases either method could be used. However choosing the wrong type is a major mistake and yet one of the most common encountered. The next few pages provide an example of the problem here and discuss this issue in more detail ....

**Analysis of the data set - Choice of data ranges**

It is important to be conscious about what you want to show with the map, which questions you want to answer. In this case it is important to know the quality of the visual variables, and to choose the right variable for the message you want the map to give.

Our effort should be towards the enhancement of maps' efficiency and communicative characteristics.
Mapping industrial companies, could be described as qualitative mapping (car-, paper-, furniture industry), or quantitative (industry with small, average and big importance) from a national economical viewpoint. Remember that in the first example you should use visual variables where you separate the different groups of industry (shape or colour). In the second example we are talking about a ranged data set, and to present this kind of data we use visual variables like size, value or texture.

Classes (intervals or ranges).

We can use class division on data sets containing quantitative data values - either in the form of absolute numbers or more often ratios. Quantitative data values can be expressed by using variation in size. Size is a visual variable with limited possibilities if used on an area (choropleth maps versus proportional symbols). That is why value and texture are used as visual variables.

In most Eurostat choropleth maps the number of classes portrayed is four or five and in a few cases the number of classes, is three. This assumption has been found to be appropriate in relation to the (A4) sized maps and the required level of generalization of the represented phenomena. Generally, the number of classes for all choropleth maps should remain between 4 and 5 depending as well to the possibilities of the different colours to use.

Remember: The choice of number of classes is a compromise between a wish of detailed information and the map user's ability to distinguish classes from each other. The human eye is not capable of distinguishing more than 8 different steps when we are talking about a black and white scale. This means that it is not possible to use more than 8 classes. Good advice would be not to use more than 5 classes and as few classes as possible.

Using class division reduces the data values from quantitative data to an arranged level, so that values are grouped into classes. Information becomes more generalised and simplified. The accuracy of the original values is lost, but the overview of the data set is better. This simplified information is easier to recognise for the reader of the map. A good class division will focus on what is the main content of the data set, and in this way minimise the loss of accuracy by generalisation.

A user can easily develop a procedure on his own for the identification of the most appropriate data classification method. The procedure is based on the interpretation of graphs (histogram of data values –HDV and frequency distribution diagram -FDD) and takes into account a number of factors:

Critical values (i.e. increasing – decreasing population)

Class limits with rounded numbers

Series of maps with common classification scheme (i.e.: portrayal of the same phenomenon on maps displaying different country)

A data set can be analyzed through the two graphs (histogram of data values –HDV and frequency distribution diagram -FDD). The use of HDV graph is auxiliary, and its interpretation is utilized for data having irregular frequency distribution.

![Histogram of data values - HDV](image)
The different class division methods are adapted to the different distribution that data sets can have. By using these methods on data sets, the data values will be "spread" to all classes.

None of the class division methods presented here (or even in existence) will satisfy all demands required by the data set. The choice of method should be taken having identified these demands. Of course it is also possible to combine different methods. Further down you will find four different class dividing methods ranging data values.

**Natural Break**

This method sets the breakpoint in the data set to "natural points" in the data set. The strength of this method is that it increases the information content, because it is adapted to the map users understanding of the map. A viewer of the map will recognise all values in one class to be "equal".

The range breaks are determined according to an algorithm such that the difference between the data values and the average of the data values is minimised on a per range basis. This reduces error and enables you to obtain a truer representation of your data.

**Equal Count or Quantile.**

Each range contains approximately the same number of records. With 5 classes, each contains 20% of the total number of the data values.

This method is suited for comparing one data set with data sets from other themes. If the data deviate from a linear distribution, the absolute class width will show large variations.
Equal Ranges

The difference between the top and bottom values in each range is the same. This means that we can use values like 0 -20, 20 - 40 etc. or calculate the width of the data set, and divide by the number of classes wanted. In this case the lowest class will start with the lowest value, the width between the classes will be the same, and the top of the highest class will be the highest value in the data set.

This method is suited for data sets with a smooth linear distribution. If the method is used on data set that are not linear distributed, you will have some classes with many values, and others with few or no values.

Standard Deviation

The class borders are calculated from the mean value and the standard deviation. Standard deviation is an expression for describing dispersion of the data values compared to the mean for the data set. The width of the class is equal to the standard deviation.

This method is suited for normal distributed data sets only. The method will then give equal number of values in each class. It method is not suitable for data sets that have extreme values or uneven distribution.
Finally, the defined class limits are refined through the critical values filter and rounded number filter (nobody remembers decimals, try to round the intervals). In some cases critical values can be easily be located, otherwise collaboration with experts of the domain is required.

### Regional Statistical data and various issues of common sense

We feel that there should be a place in this guide where we talk about regional data. Currently, Eurostat collects spatial statistical data based on NUTS (Nomenclature des unités territoriales statistiques). Accordingly, in the GISCO database all references for administrative boundaries use the same nomenclature as published in March 95. It is obvious that only data with direct reference to NUTS can be presented to a map.

Soon a revised version of the NUTS will become available, NUTS 1999. Main differences with the previous one occur in Eastern Germany (reform of regions in Sachsen), in Sweden (counties merged), in Finland (partitioning of the capital region and transfer of communes between regions) and in United Kingdom. GISCO will prepare the new geographical coverages corresponding to the boundaries of the new version of the nomenclature.

Administrative boundaries: A map showing boundaries (usually administrative) of regions is an essential document for maintaining the definition of those regions. A very accurate definition is also available from the sequence of real-world co-ordinates (latitude-longitude pairs) that define lines on the earth’s surface. However, from a visual point of view, this definition is sterile. Economists, demographers, epidemiologists, ecologists, sociologists, and so on, have need of regional phenomena to pinpoint where, and then to explain why, certain phenomena (unemployment, crime, migration, etc.) are exhibiting spatial dependence. Administrative boundaries are not always ideal for, say, epidemiological study, but for bureaucratic reasons the data are often collected in this way.

The challenge: Having established a regional data set, it is tempting to present the data on the regional map by colour or grey-tone or with overlaid symbols proportional to the size of the variable. Of course if the variable is, for example, the number of unemployed people in the 1994 civilian labour force by NUTS administrative regions level 2 in the EU, then unless it is standardised by the countries’ total civilian labour force, such a regional map does not mean very much. In other words, it is the percentage unemployed that is the variable of interest. The calculation of the ratios is referred to as standardisation.

Think first: This standardisation is essential for comparison from region to region. A large amount of regional data are counts from a base that is itself variable, so that although the standardisation in some sense yields comparability of means, the unequal base from region to region results in unequal variances. Additionally, most rates or proportions calculated for areal units are susceptible to the small numbers problem. This arises from the fact that minor fluctuations in the number or size of events will have a bigger impact on rates and proportions when the denominator is ”small”. One could claim that the very act of shading the standardised variable on a regional map and then looking for geographical clusters is an attempt to assimilate a statistical model. However, further analysis of this specific problem does not fall under the scope of this handbook.
The aim: We consider a thematic map as the primary component of any spatial analysis, presenting statistical information on "how much" or "how many", but also "where" a phenomena occurs. By coupling geography and statistics correctly we make maps talk themselves.

It is important to mention that GISCO maintains information on a number of themes such as rivers and lakes, capital cities and settlements, roads and railways, land use etc. All this information is stored in separate coverages in a consistent manner and updated regularly. The presentation of such information on thematic maps yields obvious benefits for further analyses. Requirements on the contents of GISCO database should be addressed to the GISCO secretariat.

Data from New Cronos and other sources, validation of codes

Introduction

Here we give an example of making thematic maps using statistical data from New Cronos and geographical data from the GISCO database for Arcview. We identify problems that might occur and some hints on how to get around them. Our intention is to create a map of "Population density of the average total population" for 1995 in NUTS level 2. Although some knowledge of ArcView is required our main purpose is to show the sequence of actions to be taken and not to introduce the software. However, an "assistance" interface has been developed in order to facilitate the creation of a choropleth map. This GISCO mapping tool is a customized Arcview Project, to be used with ArcView 3.0a or later versions. It is build in Arcview’s programming language AVENUE and dialogs are made in the Dialog Designer Extension for Arcview. The tool provides a mapping menu for non-experienced GIS or ArcView users that makes it possible to combine statistical NUTS data with geographical NUTS boundary data into a thematic map. All standard Arcview functionality is still available for more experienced users. Normally this extension should already be installed if you have ArcView. If not please contact estat-gisco@cec.eu.int

The steps to be followed in order to produce a map with ArcView are:

- Extract data from New Cronos
- Validate the codes found
- Use GISCO mapping tool
- Improve the layout

Extract data from New Cronos

Data can be extracted from the New Cronos reference environment or a local reference environment provided that in both cases the NUTS geographical codes are used. No matter how simple the extraction and loading of data sounds, there are intermediate steps to be followed in order to have the least possible problems at a later stage.

Data is exported from New Cronos using the standard menu of the "New Cronos Menu" choosing the DIF option for the format of the data to be extracted. Data for "Population density of the average total population" are under:

Theme 1 General Statistics

REGIO Regional statistics

Demography Demographic statistics

Tables at NUTS level 3

Population density of the average total population (file:d3densit.dft)

After raising the data on screen, choose the option Export/Table and then format DIF. Save the extraction to a file with a filename extension DIF (Data Interchange Format). Initially we were exporting the data with format EXCEL. However, when importing them into EXCEL the codes of the new Länder in Germany - DEC01,
DEC02, DEC03, DEC04, DEC05, DEC06 were interpreted as dates due to a bug in Excel. Additionally, it took more time to define the delimiters and the columns format than with DIF.

Data is then imported to Excel as DIF file. In Excel empty lines, description of regions, columns and other information not in use should be deleted. The final data set should be a table of n lines per m columns starting from cell A1 up to Smn. The first line should have the headers of the columns starting with the NUTS codes header.

Example of data file with data for mapping

Worksheet #1 of the user's data file in Excel format

It is advisable that in the same worksheet as the data, no other information should appear. However, mapping preferences, choice of colors, data classes and other useful comments should be noted in another worksheet in the same workbook. This will help documenting your work and tracing back points that need to be changed.

In our case, where we have only one variable to map - population density for 1993 - data should look as on the left (top and bottom of the data set):

Until now the validation of the regional codes and other stylistic changes had to be done manually by the GISCO team using Excel facilities. The increasing demand for GI services obliged us to automate part of the controls to be done. The result is a utility which can be used by you interested in producing a thematic map. The utility is organized in independent modules which are described briefly in the following paragraphs.

Notice: The latest version supports both versions of NUTS (March 1995 and 1999). It is not installed by default in your environment. Bare in mind that it is not a robust software and it is meant to ease a few transformations of the data set to be mapped. Additionally it verifies if the codes used can eventually be mapped.

Change Users’ data

This module makes the following changes to the data set:

- replace lower case letters (normally the NUTS codes and the column headers) with capitals

Attention DEC01 ...DEC06 are still interpreted as dates. Change them back manually to their original NUTS code in capital letters.

- convert all cell values to numbers with two decimal places

- replace all values not converting to numbers to -999 (i.e. for data NOT AVAILABLE)

- fix the column width to maximum for viewing

As the changes are permanent it is advisable to work on a copy of the data.

Check double Codes

This module finds the NUTS codes repeated in the data set. The results are noted to the spreadsheet "Errors Found". Any codes reported should then be manually removed from the data set.
Validation of EC15 codes

Where NUTS version 6 classification is used as a reference for the data set, this module checks if all codes used are valid and can be linked to the geographic codes in the available GISCO geographical coverages. Invalid codes are noted to the spreadsheet "Errors Found". These codes will not be encountered when extracting the data for mapping.

An additional module checks version 7 (1999)

Analyze ISO codes in users' data

This module is merely used for analysis purposes. The table on the right gives an overview of the data to be mapped. We can easily judge if the data is suitable for mapping per level of NUTS and country. Columns 2, 4, 6, 8 refer to the codes existing in the complete classification list for NUTS level 0, 1, 2 and 3 respectively. Columns 3, 5, 7, 9 refer to the codes found in the data set for NUTS level 0, 1, 2, 3 respectively. In order that data is mapped properly per country and NUTS level; the codes found must equal the codes existing for a given NUTS level and country. This table may also be used for documentation purposes.

In order to certify the continuity of the REGIO chronological series, new codes are inserted in the NUTS classification list (i.e. be02_94, it113_94). These codes will have been rejected at a previous stage. Since there is no real reference to any geographical code their values cannot be mapped. During analysis these codes will be ignored. In our example the validation table will be:

As shown in the above table, one code from DE is missing. It concerns DE601 Hamburg which has been inserted in NUTS version 6 classification, level 3, by error. The corresponding NUTS code should be DE6 for all levels. However, it needs to be corrected in our data set so that it corresponds to the correct region code of GISCO geographical coverage. For similar cases of missing data the NUTS codes need to be present with value -999 as this will be interpreted as data not available.

Extract data per country and NUTS-ISO

This module is used to make the final selection of the codes (and data) to be mapped. The previous table provides all the information needed to have the best representation of the data per country and NUTS level. It is our intention in the future to show "data not available" cases explicitly. It is merely a problem of workload. Extracted data will be saved to a separate workbook with two worksheet names; data and analysis. In our example we shall extract all NUTS 2 codes for all EU countries. The file must be saved as an DBF file. Dbase files can be read by ArcView.