GEOGRAPHY AND STATISTICS - SOME CHALLENGES AND POSSIBILITIES

Mr. Jan Byfuglien, Statistics Norway

Summary

This paper will discuss some issues in the borderline between geography and statistics, mainly based on experiences from Norway, but also with some reflections related to the European situation. First of all some issues related to linking of statistics to regional units at different levels will be addressed. Further some methodological questions using and presenting statistics linked to regional units, for instance in the form of maps, will shortly be discussed, and not least some new possibilities provided by technological development mentioned. Finally, issues related to standardisation and comparability will be highlighted. It is underlined that there strong user requirements for statistics with a detailed geographical breakdown. But it is suggested that there is a need focus on improving the basis for regional statistics as well as on tools and methods for analysis and presentation of regional statistics.

1. USER REQUIREMENTS FOR REGIONAL STATISTICS

It should be fairly obvious that there are strong and diversified user requirements for regional statistics, meaning statistics with a geographical breakdown ranging from major regions of a country to small local areas, or what is sometimes called 'small area statistics'.

One of the challenges in this area is to clarify the diversity of user requirements and be able to meet those requirements when disseminating regional statistics.

The formulation and implementation of regional policies both at the European level and at national level requires statistical information on relatively high regional levels (equivalent to NUTS II or III).

National policies may require much more detailed information. In Norway there is a need for a set of annually updated indicators for 434 municipalities and partly for around 13 000 basic statistical units as a basis for regional monitoring and a redistribution policy. For local planning regular small areas statistics is needed, at least covering basic social and demographic statistics (see for instance (1) for initiatives in UK to meet these requirements).

The development of environmental statistics has lead to an increased focus on geographical information and the needs for a detailed breakdown of statistics in order to be able to analyse the environmental status of different areas properly.

Business companies are (or should be!) major users of regional statistics in order to evaluate their development and marketing strategy.

For news media, students and many citizens' statistics for their region or locality often will appear as relevant and useful.

Thus there is a complex user situation; users want different type of statistics for different types of geographical breakdown. For the National Statistical Institutes this poses considerable challenges in relation to dissemination: how to allow maximum flexibility and at the same time safeguard basic quality requirements and protect confidentiality. As there are potentially large amounts of data, care should also be given on how to provide search and find facilities related to regional statistics.

2. LINKING STATISTICS TO GEOGRAPHY

The way primary statistical sources are linked to geography is fundamental for how the statistical data can be further utilised and presented (see also (2)).

By geographic referencing is meant any method of relating statistics to geographical entities (addresses, street segments, coordinates, administrative units etc.). This is used synonymous with georeferencing.

Several solutions for a more detailed georeferencing have been discussed and tried:

* Use of point referencing of addresses/buildings
* Use of grid squares for referencing
* Use of lines/segments
* Use of small areas/polygons

These different strategies also have advantages/disadvantages:

Point referencing has been the ambition of many attempts to build new geoinformation systems - but has often taken rather long time to develop and mature. In Norway a plan to establish and update a building register with coordinates in the early 1970s is only recently more or less operational. It has also been experienced that although point referencing gives maximum flexibility, it does not solve the problem of building basic analytical units, and it may cause problems related to the treatment of confidential information.

Grid squares may be seen as point referencing on a more aggregate level. It offers less flexibility that points on a detailed level, but it seems attractive as it is supposed to be easier to establish and update, and fits well into raster oriented mapping systems. But one problem is that grid squares have no relation to natural, cultural, administrative objects and boundaries, and it also may be problems connected to comparability and the handling of confidential information.

Lines/segments, which is known from census statistics in some countries, seems to be suitable with a regular street pattern, but must be difficult to implement on a more general basis also in rural areas. In relation to analysis and mapping the use of lines/segment does not solve the problems related to comparability, statistical significance and privacy.

Areas/polygons is the traditional way of referencing in statistics known from e.g. enumeration districts. One advantage is that these units can be defined and observed in reality

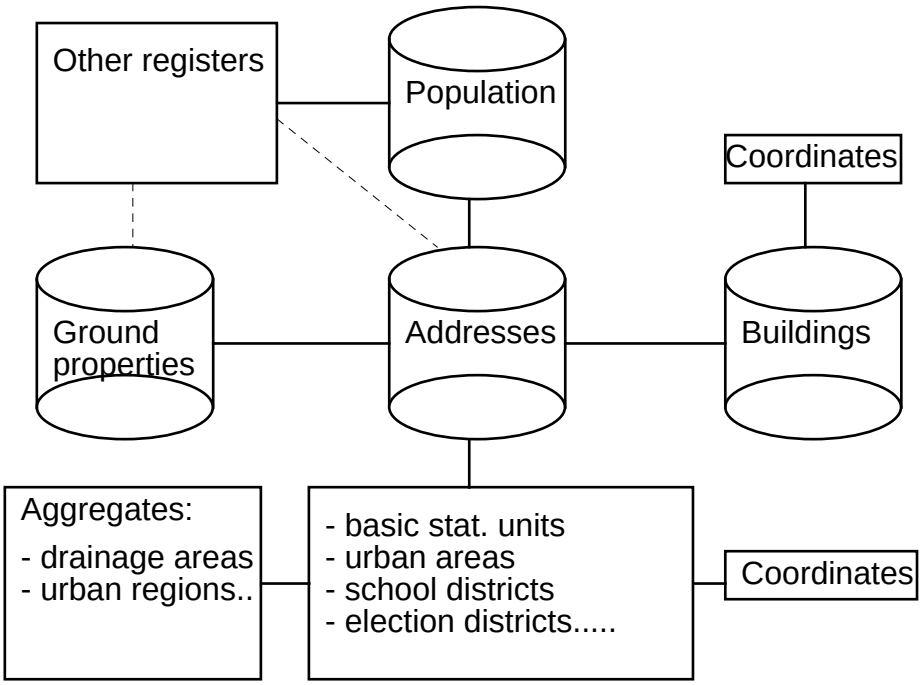
and often having links to administrative files. The possibilities of these units for use in GIS systems and for mapping depend on the detail and the quality of the basic unit division.

During the last 30 years a basic unit system has been established in Norway on the basis of former enumeration districts, but with the aim to be more suitable for statistics and geoinformation on a more regular basis. With linkage to an address and a population register, different statistical information can be given for this level on an annual basis, and also more periodically, like the censuses. These units have an average population of about 300, and are documented on local maps as well as in the form of digital files, which can be used as a basis for thematic maps. A stable system on this level may make it easier to control statistical significance and privacy.

The central part of this solution consists of an integrated register system for Ground properties, Addresses and Buildings (the GAB system). These registers have mainly been developed for administrative purposes, but Statistics Norway have been actively involved to include possibilities for the extraction of statistics, e.g. on building activity or ground property markets.

Seen from a GIS/statistical viewpoint the key register is the **address register** which should contain official positioning addresses (mostly street names/codes + house/entrance numbers), but which is also functioning as a register over geographical areas, like basic statistical units ("enumeration districts"), urban areas, school districts, election districts, parishes and even special local planning units.

Figure 1 Overview of the main elements of the basis for regional statistics in Norway



To conclude: a combination of strategies for georeferencing may be the best, and that the "traditional" polygon-referencing still may have some advantages also for use in GIS and mapping systems.

Detailed geographical breakdown of statistics is normally dependent on full coverage censuses or registers. The development of a system administrative registers with unique identifiers and a legal framework for their use in public statistics is thus the basis for the Norwegian solution described above.

However, there have been several efforts in order to develop and improve estimation methods to provide breakdown of sample data (see for instance the 5th Framework Eurarea project, coordinated by Office of National Statistics).

3. GEOGRAPHICAL UNITS AS BASIS FOR ANALYSIS AND PRESENTATION

It is important to make a distinction between basic geographical referencing units and geographical units for analysis and mapping. If statistical data only are referenced to the primary administrative unit (for instance a municipality) this restricts the potential usage of these data. Any further analysis or mapping will then tend to use these units as a basis for analysis. However, the problem can often be that these fundamental units in reality are not comparable for use in analysis and mapping.

In the Norwegian context it is for instance e.g. hard to tell what is common between municipalities varying in population from 200 to more than 510 000! It is a pure statistical observation that small municipalities will tend to have extreme values on indicators. Combined with very varying area (between 6 to 9.000 sq.km) and differences in population distribution, such a geographical basis causes problems for mapping - and for regional analysis. A thematic map - and a regional analysis - will contain a certain level of "statistical noise"; much of the observed regional variations and patterns may be a function of the structure of the geographical units, while "real" variations are somewhat hidden. But still we too often use these non-comparable units as a basis for ranking based on a selection of indicators!

The heterogeneity of the basic geographical mesh influences the comparability of the regional statistics and analysis in general across Europe. Within EU15 for instance, the average size of local units vary from 15 to 782 sq. km with an average population ranging from 1500 (France) to 117 800 (United Kingdom, districts).

The same problem applies to the NUTS II and III regional divisions, where there is a considerable variation of area and population within and between countries.

When using these geographical entities for defining other spatial zones, such as 'towns' or 'functional urban regions' as in a recently published 'Urban audit' (see: http://europa.eu.int/comm/eurostat/Public/datashop/print-product/EN?catalogue=Eurostat&product=3-25062004-EN-BP-EN&mode=download) the

result can be strongly influenced by the heterogeneity of the primary geographical units (see also (3)).

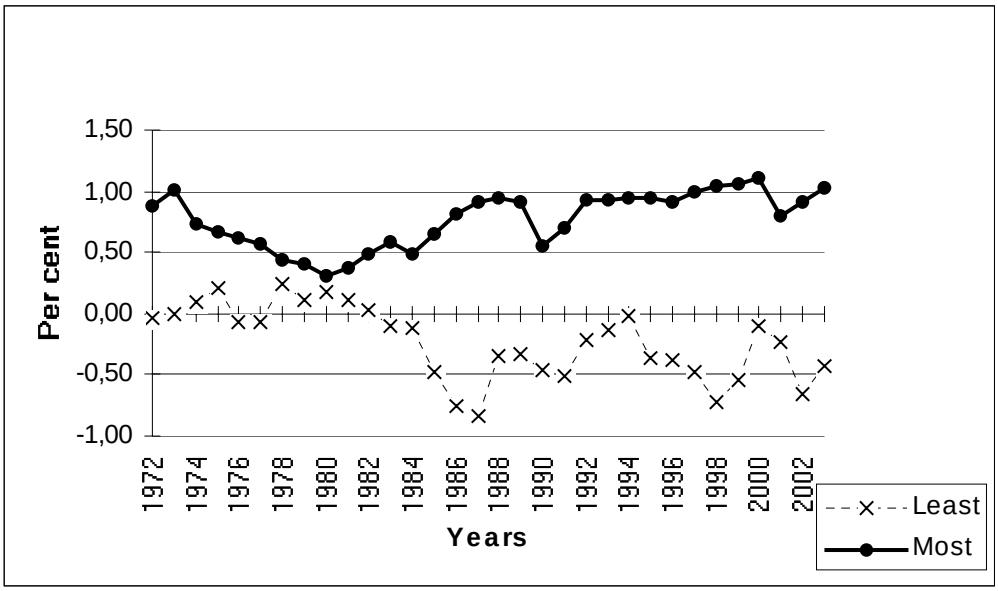
This situation underlines the need to work on international regional standards for statistical usage, not only 'administrative' standards such as NUTS, but also specific targeted standards for urban areas, functional zones etc. One approach is illustrated in (4) analysing some of the problems with heterogeneous regions and the inadequacy of traditional spatial analysis. One approach is to break down the larger municipalities into sub areas - also to be used for public statistics in general, even if one is not able to georeference statistics on a more detailed level (see chapter 1).

Further, methods for aggregation/classification/clustering are important in order to construct comparable units and units relevant for the topic in question. This is especially relevant when performing regional analysis or producing thematic maps. Catalogues for aggregation of the primary regional units should therefore be available as part of a regional database.

In Statistics Norway there are several such catalogues for the grouping of municipalities included in the statistical database, also available on the web. One catalogue is a classification of municipalities by centrality, based on distance from major city centres. The most central group consists of municipalities where the travelling distance of the centre of an urban settlement with at least 50 000 inhabitants (five such centres in Norway). The least central group consists of municipalities outside commuting distance to urban settlements with a population of at least 5000. In addition there are intermediary groups classified according to the size of the commuting centre (5000-10000 and 10000-50000) (see (5) and (6)).

One important issue in Norwegian political discussion is to what degree there is a regional centralisation process. The following graph provides one partial answer to this issue. Apart from one period around 1980 when there was a relatively balanced regional population development; the overall tendency is that the most central municipalities gain and the most peripheral municipalities loose population.

Figure 2. Population growth in municipalities Norway 1972-2003. Most central and least central



4. PRESENTATION OF REGIONAL STATISTICS

One obvious way of presenting regional statistics is by using thematic maps; especially as IT tools now are available making this relatively easy, also via the web. However, the availability of simple graphical tools make it even more easy to 'lie with statistics' also on maps. There are numerous examples on the misuse of mapping tools, for instance by using choroplet maps (which are easy to produce, see figures 5 and 6) also for absolute figures, giving a totally wrong picture in as situation with a wide variation of the size and population density of the geographical units. A basic understanding of the strengths and weaknesses of graphical tools is thus necessary.

The map gives a small or large picture of "reality". One often has to make rather strong generalisations and simplifications, and the question is how this process is done. In the mapping of natural phenomena this is often done in an intuitive way, e.g. by combining small areas. On the other hand there are geometrical methods for the smoothing of lines and the elimination of details - which may not be relevant to retain a maximum of information when working with statistics on society.

The question is also: What level of generalization gives the "right" picture of reality; is it the scale 1: 5000 meters, where details may be depicted. Or what use can be made of maps with at high level of generalization, e.g. scale 1: 3 000 000 meters. Some of the discussions on the value of thematic maps are based on different opinions on this point: Surveyors e.g. prefer maps "where you can know where you are", while geographers also like maps on a high level of generalization, and seen more as a regional model.

As a space oriented graphic medium the map has some strengths, but it also suffers from some disadvantages (see also (2)):

Advantages are amongst others:
- Gives a quick overview: Much information on little space.
- Makes use of primary and intuitive graphical understanding
- Focuses on geographical patterns and relationships
- Facilitates the geographical integration of phenomena

Disadvantages/problems are:
- Dependent on the quality and the documentation of the basic data
- The chosen map units may cause "noise" for interpretation and analysis
- Statically and visually oriented; what about processes and interactions?
- Geometrically oriented; is this relevant for social phenomena with varying "intensity" in space?
- The chosen projection influences the interpretation (especially on the world map level)
- Very dependent on the graphical methods chosen (patterns, colours etc.). Wrong use may give "noise" or lead to misleading interpretation
- The map picture may give imprecise information and it may be difficult to make exact comparisons

Figure 3. Basic statistical units in the centre of the municipality of Kongsvinger

Figure 4. The municipality of Kongsvinger with major statistical subdivision and population density for grid squares 250 * 250 metres

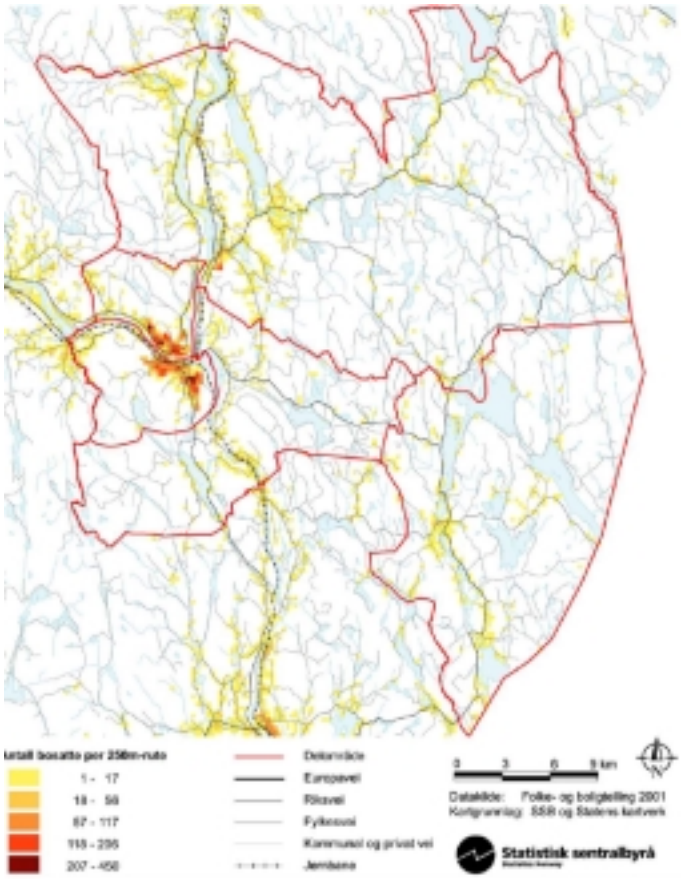


Figure 3 provides an example illustrating the granularity of the basic units in Norway in the city centre of the municipality of Kongsvinger (population around 11 000 in dark grey urban area). In figure 4 the whole municipality is included showing only the major (statistical) subdivisions with an illustration of population distribution based on density in squares 250x250 metres (colour scale not easy to see in a black and white representation). These illustrations are from a booklet for each municipality produced with population census results.

The example in figure 5 is showing the proportion of people 65 years and more in all Norwegian municipalities - not easy to interpret, partly because the more densely populated areas around Oslo have municipalities with small areas but relatively large population. This map also illustrates the problem making good legends; black is very strong while white areas is more or less equal to not existing information.

Figure 5. The percentage of persons 60 years and older. 1 Jan. 2004. Municipalities. Norway

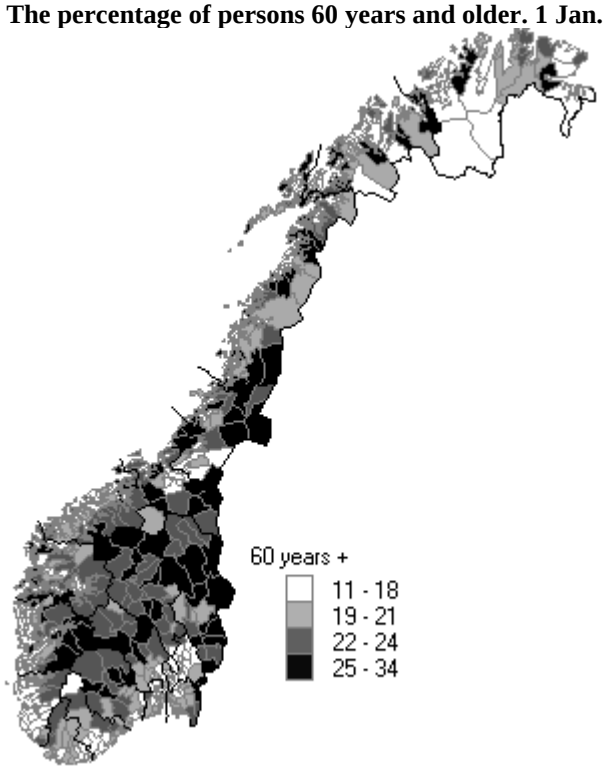
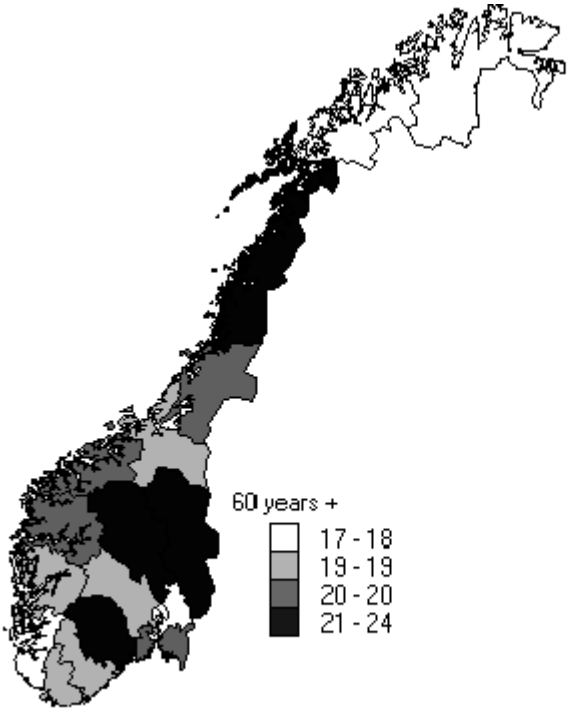


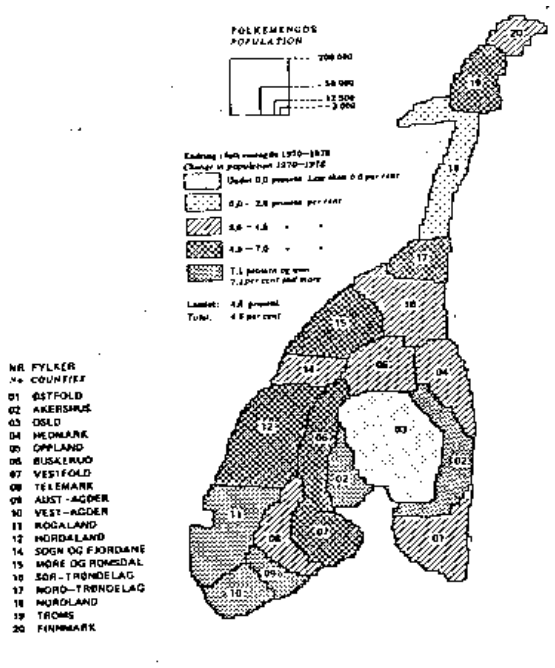
Figure 6. The percentage of persons 60 years and older. 1. Jan. 2004. Counties. Norway



A simplified picture is illustrated in figure 6 with the percentage of persons 60 years and more on county level. This level hides important differences, and a level between the counties and the municipalities, might thus in this case be most relevant.

One way to counteract the deficiencies of the (too?) often used choroplet map is to use symbols proportional to the population, or even to use a 'transformation' where the area of the basic geographical units are proportional to the population, as in figure 7. These types of maps may give insight into regional patterns, even with a distorted geometrical picture. One drawback is that existing IT tools cannot produce such maps.

Figure 7. Population 1970 and population change 1970-1978. Area of counties proportional to population 1970



5. SOME CONCLUDING REMARKS

There are strong user requirements for regional statistics, ranging from statistics for very small areas to larger regions. The availability of data depends to a large degree on full coverage censuses or development of registers. New technology helps in making available regional data bases with flexible search and manipulating facilities. Tools for aggregation and classification of regional data should be developed and made available. There is a strong plea for some basic international standardization related to urban areas and other basic regional units. Methodological issues both related to the non-comparability of basic regional units and to methods and tools for analysis and map presentation should also be addressed.

References

1. Virdee, D. 2001. *Strategies and approaches for small area statistics. UNECE Conference of European Statisticians, Geneva, 11-13 June 2001*
2. Byfuglien, J. 1993. *Thematic maps and the presentation of statistics: some problems and possibilities. UNECE Work session on Geographical Information Systems. Ottawa, Canada, 28-30 September 1993*
3. Byfuglien, J. 1994. *Some issues in relation to the definition of urban units and other units for geographical analysis at the European level. UNECE Work Session on GIS. Vorburg, Netherlands, 27-30 September 1994*
4. Grasland, C. 2000. *Spatial homogeneity and territorial discontinuities. UNCE Work Session on Methodological issues involving the integration of statistics and geography. Neuchâtel, Switzerland, 10-12 April 2000*
5. Byfuglien, J. 1986. *The analysis of the settlement pattern in relation to planning problems. Norsk geogr. tidsskr. Vol 40. 187-194. Oslo. ISSN 0029-1951*
6. Byfuglien, J. 1995. *Urbanisation and centralisation in Norway. Norsk geogr. tidsskr. Vol. 49, 35-44 Oslo ISSN 0029-1951*