

Abstract:

Presentation of Danish method to take care of confidentiality providing statistical data on grids

Erik Sommer, Statistics Denmark
Chief Adviser,
Virhe. Kirjanmerkkiä ei ole määritetty., eso@dst.dk

**Nordic Forum for Geostatistics 2007:
Helsinki, Finland 19-21 September 2007.**

The focus in this presentation is on our handling of data within Social Statistics - persons, families and households.

1. Recommended discretionary requirements – Social Statistics

The work on Information Bearing Cells is subjected to Statistics Denmark's discretionary requirements. There must be a sufficient number of households in the combined cells until statistical data can be released. For example, key figures (average figures) are released when there is at least 20 households in an area or in a cell.

Number of households (clusters)	Data from Statistics Denmark
1-19 households	No data
20-49 households	Key Figures
50-99 households	Few intervals
100-149 households	More intervals
150+ households	Statistical datasystem

Figure 1: Discretionary requirements – Statistics Denmark

Statistics Denmark releases a dataset with the number of persons and households on 100 x 100m cells, which form the basis for the work performed by the data user in combining the cells.

Grid ID	Municipality	Households	Population
61901_7126	190	1	6
61902_7126	190	3	8
61903_7126	190	12	26
61904_7126	190	3	7
61905_7126	190	3	8
61909_7126	190	5	10
61910_7126	190	3	6

Figure 2: Building block of households and persons (population).

2. Clustering

If there is not a sufficient number of households in the cell and a process of smoothing or aggregated values for cells are not applied, the cells must be combined into clusters. Overview of distribution of households in Denmark 100x100 meter Grid 1. januar 2007 is listed here:

Level	Householdsintervals	No. Cells	% Cells	No. Households	% Households
1	1-19	396035	94,70%	1452101	57,35%
2	20-49	15193	3,63%	456376	18,02%
3	50-99	5052	1,21%	343045	13,55%
4	100-149	1221	0,29%	146801	5,80%
5	150-399	673	0,16%	128461	5,07%
6	400+	8	0,00%	4001	0,16%
0	Not placed	0	0,00%	1210	0,05%
Total	Denmark 1.1.2007	418182	100,00%	2531995	100,00%

Figure 3: Distribution of households – 100x100 meter 1. January 2007, Denmark.

There are differences in how each individual cluster is generated, partly depending on the hypothesis put forward for the task, the underlying method of registration and partly for reasons of the discretionary requirements. For example, if an analysis of crimes is to be conducted, we want our clusters to consist of neighbouring cells (hypothesis: that crimes are related to specific areas). In the case of a Direct Marketing (DM) campaign, we will make attempts to ensure that our cells are significant with respect to a number of background variables, and this implies that they need not be neighbours.

Cluster presentation of households income for part of Copenhagen, Denmark. The darker the colour the higher is the household income level. The first figure is a presentation on a zipcode level (postal district) and the second figure is based on the use of grid data.

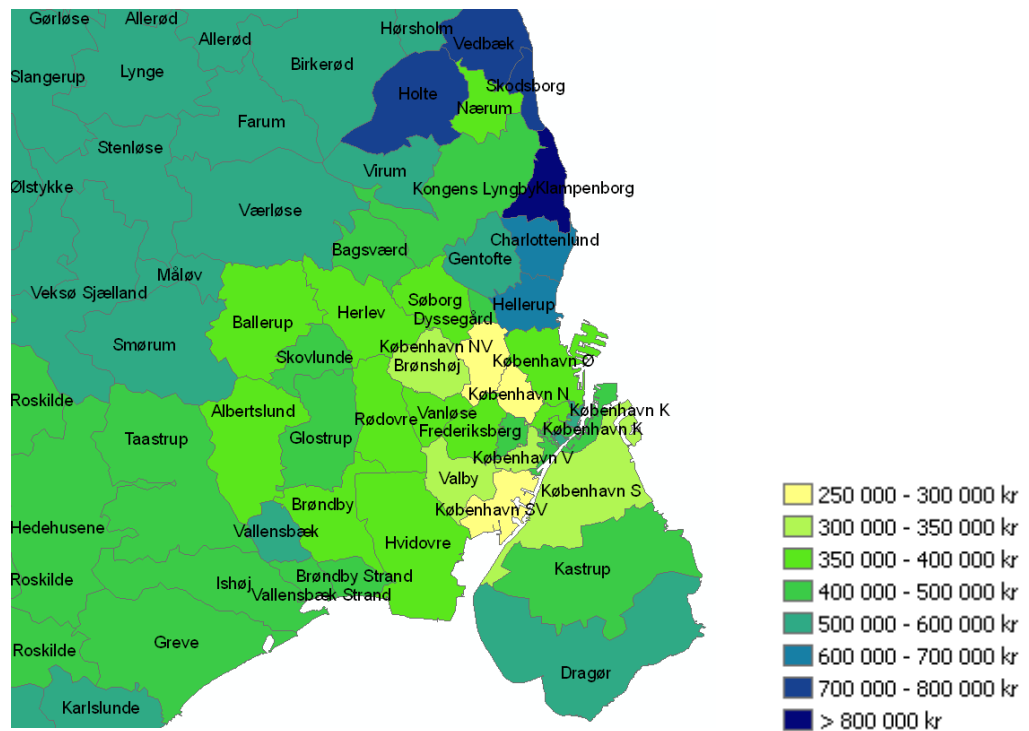


Figure 4. Gross Income Copenhagen Area on postal districts. Average household income taxyear 2005

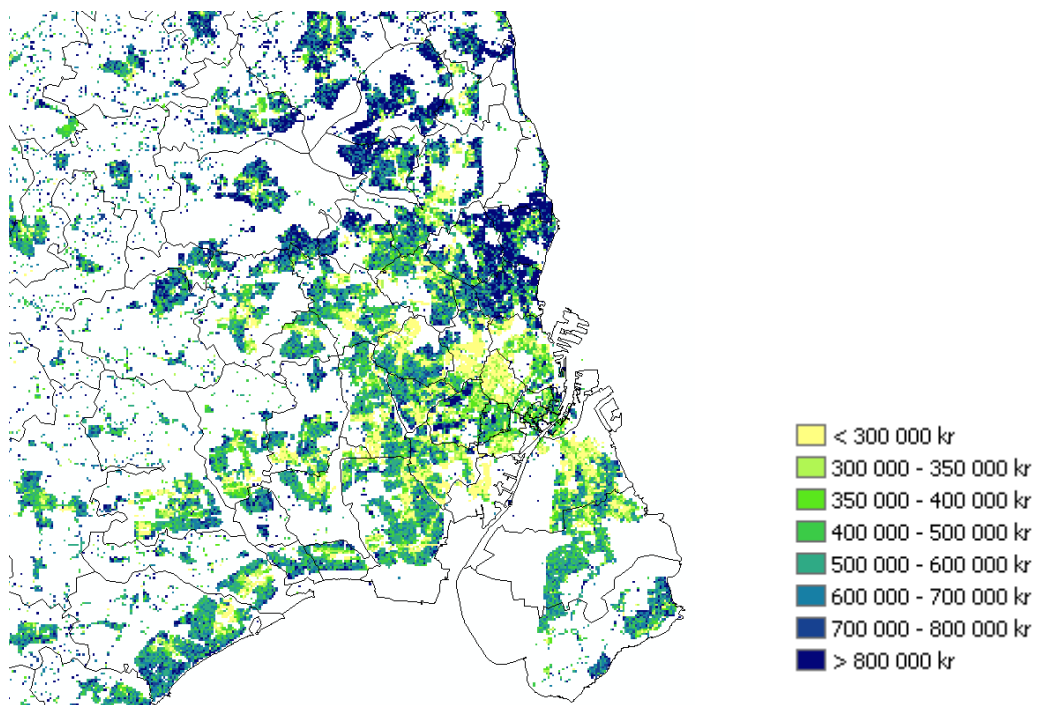


Figure 5: Gross Income Gridclusters with a minimum of 20 households. Average household income taxyear 2005

3 Future Work:

A number of things have to be considered if we want to further develop the use of grid data especially if we want to increase the integration between existing grids for regions across national boundaries.

- Integrating statistics in regional boarder areas
- Integrating business statistics
- Common Disclosure policy?
- Common Pricing policy?
- Common Dissemination of data?
- How do we cooperate with the National Mapping Agencies
- How do we cooperate with private partners (value added data providers)?