

Enhancing use of geographic statistics through visualisation – the Danish experience (Améliorer l'utilisation des statistiques géographiques grâce à la visualisation – l'expérience danoise)

Thygesen, Lars
Statistics Denmark
Sejroegade 11
2100 Copenhagen O, Denmark
E-mail: lth@dst.dk

Quitau, Jarl
Statistics Denmark
Sejroegade 11
2100 Copenhagen O, Denmark
E-mail: jaq@dst.dk

0. Overview

Official statistics have great potential to generate knowledge and serve as a basis for decisions taken by many actors in society; however, the potential is under-utilised because many people do not know about statistics, do not care about them and do not understand them.

This paper highlights Statistics Denmark's recent work on enhancing the use of geographic statistics through visualisation on the web. The paper demonstrates how a visualisation tool has been implemented as a powerful enabler to different categories of users of statistics, who wish to understand the geographic dimension of societal phenomena and find the facts hidden in the data. The scope of the tool spans from leading edge regional researchers, through local and regional planners and news media, to the general public, e.g. citizens who are interested in what goes on in their own community. Consequently, the tool allows for many different ways of presenting itself and the data, and it is applied to many different levels of geographical detail.

1. The objectives

Official statistics are much less recognised and used than they ought to be, considering the valuable information they offer. Many public policy makers and business decision makers don't know about the official statistics and turn to other sources or decide without sufficient evidence base, and among the informed public they are even less used or known. Thus the potential of statistics is vastly under-utilised.

Statistics have an unfortunate image of being boring – even though some of us know that they are in fact fascinating and exciting. But it is a historical fact that the representations of statistics in different media

do not necessarily lend themselves easily to being absorbed by non-experts. This, in turn, makes people less prone to go and look for statistics.

The primary objective of the introduction of geographic visualization is to avoid boring statistics. We want statistics to be exciting so that they invade people's minds and generate knowledge, thus allowing users to apply new insights as a basis for decisions. In order to do this the visualization should highlight structures and patterns in the data and allow the user to play around and test her own hypotheses and ideas.

It is well known that maps speak to the minds of many people. In addition to maps showing the data, we want to highlight correlations between several variables across the geographic areas, as well as dynamics over time.

2. Characteristics of the tool

The tool chosen by Statistics Denmark is Statistics eXplorer, developed at Linköping University by the NCVA research team led by Professor Mikael Jern. It allows authors to present their geographic data and facilitates sharing of insights through stories linked to selected views on the data. Some of the key features that greatly help turning statistics into knowledge, seen from a user perspective, are described in the following.

2.1 Maps and graphs

With the Statistics eXplorer tool we can easily present statistics by municipalities, counties, parishes, countries or any other geographical division into interactive maps and other graphics, which help users to see and understand the structures and relationships in the statistics. Statistics Denmark uses Statistics eXplorer for two audiences: to present statistics on its own web site, and to assist paying clients who need such a tool to show their statistics in a way suitable to generate knowledge.

Statistics eXplorer is designed with different options for viewing. The full user interface (or standard view) is ideal for expert users who want to analyze and go in depth with data; another view aims for an ordinary citizen needing an easy overview of the geography of a statistic, e.g. ageing of the population.

The standard view on Statistics eXplorer is shown in Fig. 3 below. It is centred around dynamic maps that interact with other dynamic graphs included in the tool. In addition to the dynamic map, with mouse-over showing data, you can choose from the following views: Scatter-plot, Table Lens, Data Grid (i.e. a table), Parallel Axes Chart and Time Graph. All views are fully synchronised and can be shown with time animation.

A user accessing the tool on Statistics Denmark's web site can easily add his own data and analyse them locally but they will not appear on our web site.

2.2 Story-telling

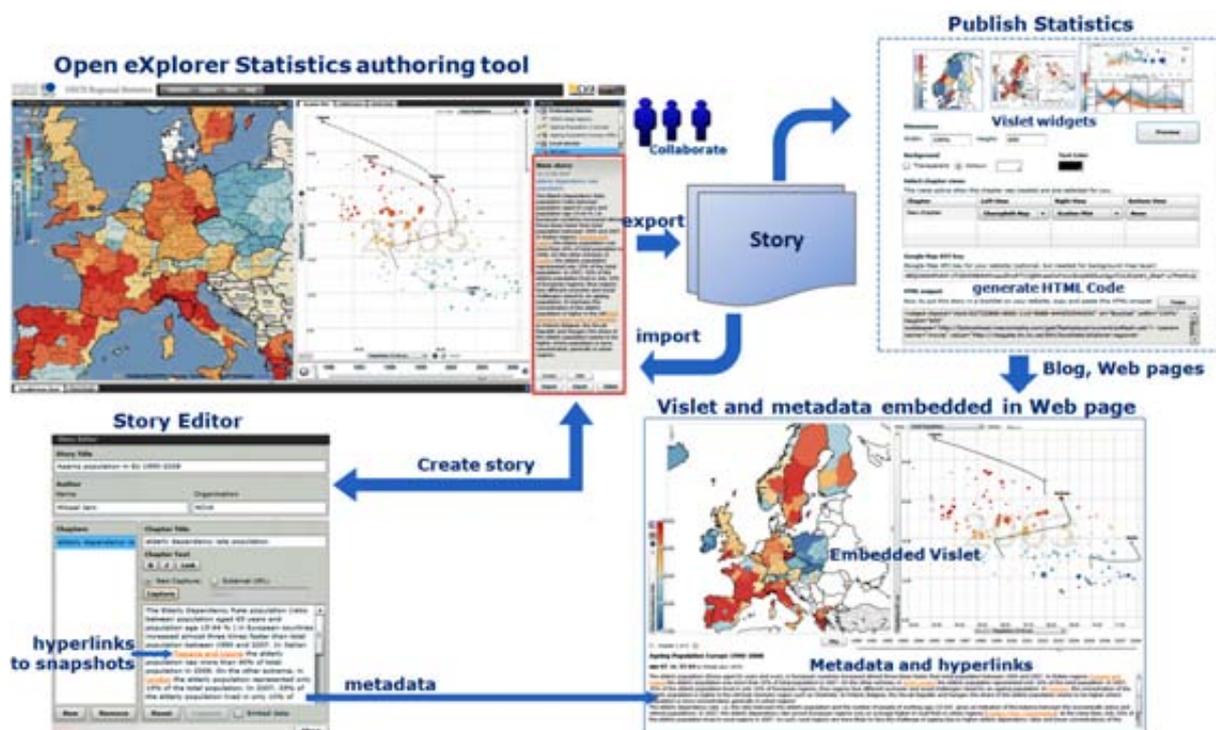
Statistics eXplorer allows the NSI or any other user to analyse the data, find the interesting stories they contain and select the views on data – maps, correlation diagrams, etc. – highlighting such a story; the author can then write the story to come along with the data and present it all together, including animation and links.

If users are to devote interest to statistics, gain knowledge, and use the information for decisions, the statistics must clearly convey a message. The story-telling module therefore is essential. The idea is illustrated in Fig. 2.

2.3 Vislets

A subsystem of Statistics eXplorer (called Vislet) allows you to choose the components you want (e.g. only a map) and "embed them" into a web site. Here it is also possible to have links in the story, which then activate the animation of maps and graphs. The principle can be seen in Fig. 1 below. The Vislet can be a very simple and easy-to-understand site that lends itself very well to users without any special experience in understanding complicated statistical analysis.

Figure 1. Creating stories and vislets



3. Introducing Statistics eXplorer at Statistics Denmark

Until recently, Statistics Denmark's web site contained only few maps or pictures; content was almost exclusively data and text.

Starting from 2009, Statistics Denmark experimented with the tool to demonstrate how the stories hidden in the data could be unveiled to the public. The first presentations on our web site came out in 2010.

To ensure that users would come forward and contribute to a good implementation of visualization of the transfer and sale, Statistics Denmark established a small group consisting of users with different profiles: regional researchers, municipal planners, school teachers and pupils. The group met in early 2010 and was been invited to test and evaluate solutions based on Statistics eXplorer. The feedback was very positive, as most user categories found that services along these lines would greatly help them understand and use the geographic information. They also gave useful input regarding aspects of the presentation, some asking for additional features, others for simplifications. This confirms the expectation that we must have different levels of service for different segments of users – which is possible to manage with Statistics eXplorer. The producer of Statistics eXplorer, NComVA, has shown great interest in these messages and has proved willingness to act on them.

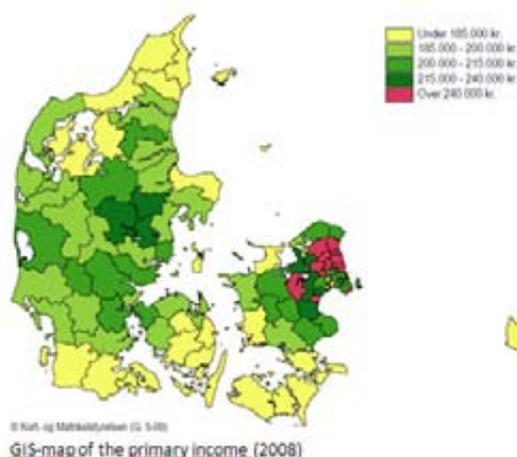
4. Launching visual income statistics

For more than a century Statistics Denmark has published its annual income statistics. This work would consist of a short introduction of the most important concepts regarding income statistics followed by a horde of tables. By 2003 more than 150 pages consisting of hardly anything but numbers showing various types of income tabulated across age, gender, family-types, socio-economical groups, isco-88 codes, municipalities etc. As such the user group primarily consisted of local government officials, social scientists and fellow statisticians.

Around this time many of the same tables were made available in Statistics Denmark's Statbank on the web (www.statbank.dk) and a decision was made to widen the target audience for this publication to include anyone who might have an interest in income statistics. Explanatory text paragraphs were inserted to make the statistics understandable for lay-men. Some tables were converted into graphs and figures. Other tables were aggregated and presented in a more digestible format.

There has always been a large interest for regional income statistics, i.e. various income types tabulated on municipalities. In the new publication format, a few very sought after variables such as the disposable income were shown on static GIS-maps. The remaining types of income tabulated on municipalities were placed in the Annex or omitted.

Figure 2. GIS-map of primary income (2008)



4.1 A student's perspective

Let's look at the income statistics from the viewpoint of a high school student from Lolland, one of the poorer municipalities, having to write a paper on the local living conditions. Looking through the publication he finds a map on the primary income. He locates Lolland in the bracket of municipalities with an average primary income under 185k DKr pr. capita above the age of 15. Just like many of the neighbouring municipalities in southern Denmark. In itself that's not much information. *What's the exact primary income in Lolland?* With some luck he'll see a reference and find the correct table in the annex. If he manages to seek out the correct column and find Lolland between the 98 other municipalities, he'll find

that the average income is actually only 148k DKr. Not only is this below the country average at 201k DKr, it's also significantly lower than the neighbouring municipalities.

Why is the primary income of Lolland so low? Talking with his father he may learn that it has all been downhill since the local shipyard closed. Talking to Co-students, teachers and friends about the matter he may get a range of other explanations, e.g. the financial crisis, high unemployment rates and a relatively high elder / working-age ratio. Some may in addition claim that the situation is not as bad, when you consider other factors such as lower real-estate prices in the municipality, the effect of the progressive taxes which helps equalise the post-tax income and a high level of public transfers and pension payments etc.

Now in true mythbuster style our student has to sort through all these claims and see, which of them hold up to the test, which are plausible and which ones he can bust. He would quickly realise that using a purely literary approach, would lead to a lot of page turning, spanning several publications.

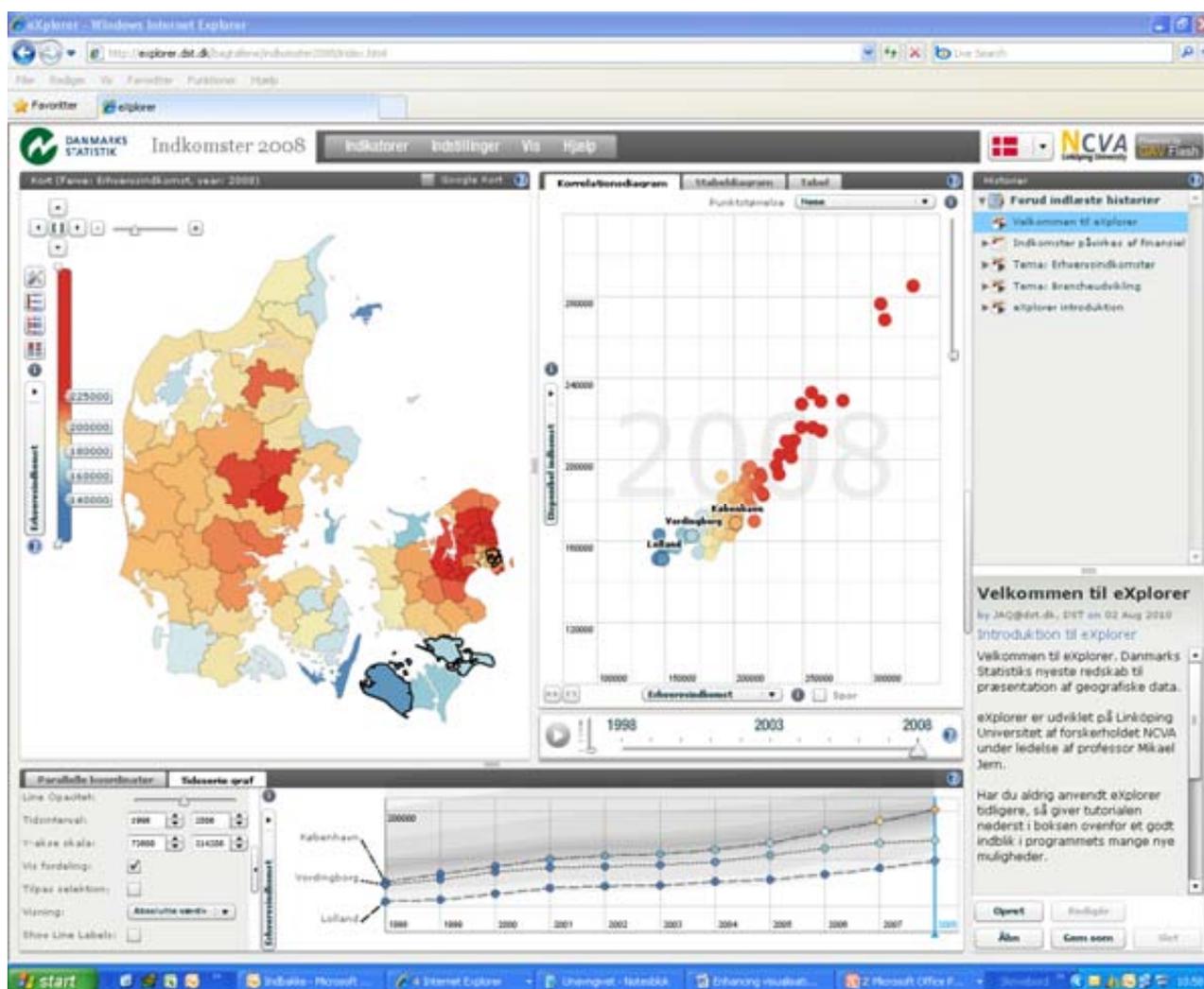
A more surmountable approach would be to visit Statistics Denmark's Statbank on the web and start locating the relevant tables. For an experienced statistician it wouldn't take long to do the necessary investigation into these claims and categorise them as plausible or not. But for lay-men this can be a daunting

task and could quite easily turn into a case of information overload. Luckily for our student a third option is now available on Statistics Denmark's webpage, Statistics eXplorer (www.dst.dk/indkomster2008).

eXplorer will by default present our student with an interactive map of the Danish municipalities. Upon choosing the primary income, he can by holding his mouse pointer over his municipality find the exact primary income of Lolland, and by hitting the play button he can see how the primary income have evolved since 1998.

Next to the map the selected municipalities are highlighted in a range of figures including a scatter plot, histograms, time series graphs and basic tables. Furthermore Statistics Denmark has written a set of stories on how incomes have evolved in the past decade.

Figure 3. eXplorer, primary income (2008)



It quickly becomes evident that Lolland indeed has one of the lowest primary income pr. capita in Denmark. Hardly a day has gone by in the news without a mention of the financial crisis; this may be a good place to start. Looking over the story section, he may find a story focused on the consequences of the financial crisis, showing how income changed in the period 2006-2008 spanning the beginning of the financial crisis.

What he will find is that Lolland actually wasn't affected much in the first period of the financial crisis. The municipalities that took the biggest blow were instead the richest municipalities north of

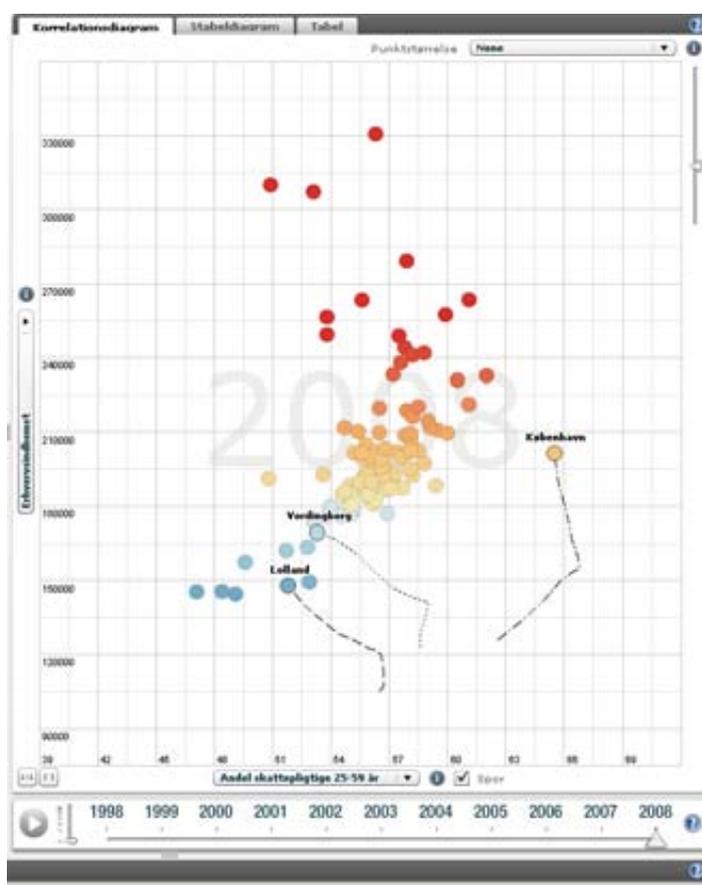
Copenhagen which suffered great financial losses from declining real-estate prices and stock exchange prices. This does not exclude the possibility that primary income in the following period took a blow from the financial crisis and unemployment rose. But it can be concluded that the financial crisis can't explain that Lolland fared poorly on the primary income in comparison with the rest of Denmark in 2008.

Lolland is however characterised by high unemployment numbers, and this is certainly a contributing factor. Taking a look at the line of businesses in which the employed work, it's also apparent that the primary and public sectors are overrepresented on Lolland, sectors which today pay lower average wages and give less profits for the self-employed, than for instance the service sector.

When looking at the percentage of primary income which stems from the industrial sector, our student would find that this is actually relatively constant around 16 percent. This makes his father's idea about the closed shipyard as the cause of low income by 2008 less plausible. Even though his father always talk about it, as though it happened last week, a quick search on the web would quickly reveal that the shipyard in Nakskov closed in 1987, several years before our student was born.

Even if the situation arguably would be better with the shipyard still in function, it can hardly be characterised as one of the sole contributing factors to the low primary income in Lolland.

Figure 4. eXplorer, working-age ratio and primary income 1998-2008



Another important factor however could be the proportion of elderly people compared to the number of people in the working-ages. The working-age-ratio is shown in the scatter plot below along with the primary income. In 1998 56.1 pct. of the population above 15 years old, were in the working age. By 2008 this ratio had plummeted to 51.8 pct. Apart from 3 of the richest municipalities north of Copenhagen, this scatter plot seems to support the idea of a positive correlation between primary income per capita and the percentage of people in the working age. Considering that Lolland has one of the lowest working-age ratios in the country, this also seems to be a plausible contributing factor as to why primary income is so low in this municipality.

4.2 Implementation of eXplorer

Statistics eXplorer was introduced as part of the annual theme in the income publication, about the regional distribution of primary incomes and its sources.

Income statistics as well as many other statistics in Denmark are register based.

The population of the Income Statistics Register includes anyone who has submitted a tax return to the Tax Administration, and as such the register covers anyone who is economically active, with or without residence in Denmark. Merging these data with the addresses from the Central Person Register allows us to make very detailed regional statistics on income. This however raises a range of issues with regard to discretion and protection of personal data.

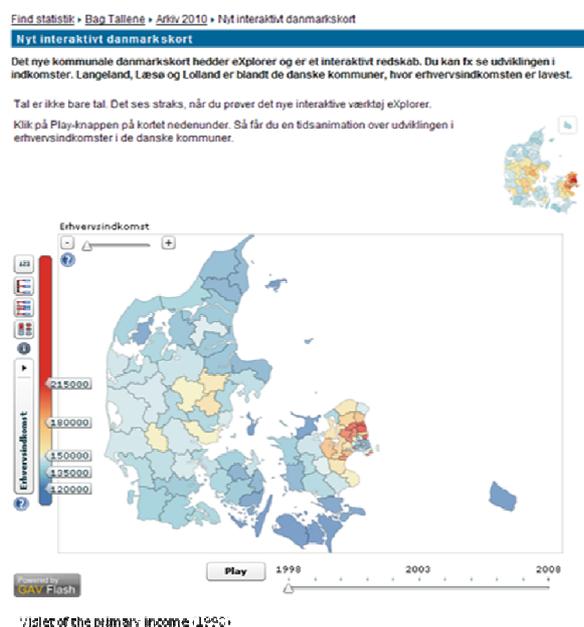
The requirement for discretion did cause a few concerns when it came to the use of eXplorer. In Denmark we have a few very small municipalities. Thus the average income can easily drop or rise by

several percentage points if someone quite rich moves between municipalities or if a local self-employed suddenly has seen his business decline. Even though we from a technical standpoint don't provide any more information than you do by publishing a set of tables tabulated across municipalities, a tool like eXplorer will quickly reveal such sudden changes to the public.

To mitigate this issue in the 2009 version we are currently considering publishing the median and various other percentile limits for the most commonly used income variables instead of using averages. Another often used solution to this issue has been to merge some of the smallest island municipalities with their larger neighbouring municipalities which became the solution in 2008 version.

Judging by the number of journalists and students calling on a regular basis with doubts about which type of income to use for a particular paper, understanding income statistics is complicated enough in it self for lay-men. If you to this add the large number of configurations made available with the eXplorer, chances are that our users would find eXplorer very hard to digest.

Figure 5. Vislet, primary income, 1998



In order to address this we took 3 basic steps. Firstly in the full version of the eXplorer, the default story the user sees, is not about income statistics, but is in stead an introduction to the many options available in eXplorer. Furthermore an introduction to eXplorer was also included in the annual publication *Indkomster 2008*. Finally an introductory article on eXplorer was posted on our website with a few *vislet* editions of eXplorer, where most of the options were removed, for example only presenting the user with the interactive map and the 6 most commonly used income variables.

To make people aware of our new interactive maps we took the following steps. In the annual publication the URL's and hyperlinks underneath each GIS-map would encourage readers to try eXplorer on the web. Furthermore the eXplorer maps were advertised for about a month on Statistics Denmark's front page, containing an article with the *vislet*-edition and a link to the full edition.

Looking ahead we aim to make eXplorer a central part of the income section on Statistics Denmark's new homepage, which we'll launch in the fall of 2011. Furthermore we hope that we one day can make eXplorer or similar maps an integral part of our Statbank. This should enable our users to directly export all available regional data from our Statbank directly onto custom maps.

5. Serving specialised users

In addition to publishing geographic statistics on the web, Statistics Denmark is planning to use the tool also for specialised users, typically customers who are willing to pay for the creation of special datasets, e.g. small area statistics. The idea is that Statistics Denmark will host the data as well as the Statistics eXplorer tool for the customer, who can then analyse, make stories and present results to their internal or external users. This service is in the process of being prepared, and the first customers have shown great interest and will probably have started when this paper is presented in August 2011.

6. Discussion

In this section we present three of the important questions and doubts we have to consider in relation to our work with visualisation of geographic statistics.

6.1 Usability

Can our users digest the rather sophisticated presentations offered by a tool like Statistics eXplorer? As explained above, Statistics eXplorer encompasses many presentation formats working together. While this allows the professional user to see and understand a lot of interconnections between the data, it can be overwhelming to a lay user who just wants to get the essence and doesn't read statistics in bed at night.

We have experienced that the full tool with all its many facilities is clearly attractive and useful to expert users. The story-telling facility makes it possible to communicate the insights found through the data also to more common users. The Vislet feature, which allows to simplify the message and present it on a web site, goes one step further towards making the messages understandable, and yet interesting, to almost any user group.

6.2 Technical platform

Is a tool based on Adobe Flash sufficiently accessible across today's many technical platforms? While most pc users (at least among the users of Statistics Denmark's products) actually use Flash on their computers, or can easily install it if they haven't already done so, it could be argued that smartphones are increasingly becoming a platform that people use for web access, also for serious business like statistics; and Flash is not . However, on the smartphone market there are at present multiple platforms without any one being dominant, so it would be an overwhelming task for an NSI to keep abreast with nice apps for all these platforms. Flash is developing into the smartphone market, it is already usable on some of the important platforms and we are confident that Adobe will work hard on overcoming these obstacles. To our mind, Flash remains a viable de facto standard on which we can fruitfully base our dissemination for the immediate and medium term. What happens after that, no one can see today.

6.3 Who should serve users

Why should an NSI engage in making its output of geographic statistics edible to users? Shouldn't an NSI restrict itself to making the data available on the web in an easily downloadable format (but not necessarily inviting to end users) and rely on private redistributors, or entrepreneurs to present it in nice ways?

We believe that we should apply a two-tier approach: Allow redistributors to develop nice tools while at the same time making the statistics freely available in appetizing forms, thus competing with redistributors for the attention of users. If users prefer to go to commercial entrepreneurs there is no harm done, except that it is an important source of knowledge and innovation for an NSI to be in close contact with real users.

If redistributors misinterpret or misrepresent our statistics, it is our duty to stop it and to publicly denounce it, according to the Law ... if we become aware of it, and that may be a very serious problem!

Redistributors don't have the same responsibility as NSIs regarding credibility. If they make mistakes, they may just disappear from the market (but may re-emerge with a new name). NSIs cannot afford to do that, and NSIs have a responsibility to provide a quality assured, and, we believe, not too boring presentation of statistics to all kinds of users.

References

Jern, M., L. Thygesen, M. Brezzi: A web-enabled Geovisual Analytics tool applied to OECD Regional Data, Reviewed Proceedings in Eurographics 2009, Munich, March 2009

Statistics Denmark (2010): Indkomster (Income Statistics). Copenhagen, Denmark, September 2010

Thygesen, L., B. Sundgren: Innovative approaches to turning statistics into knowledge. Statistical Journal of the IAOS. Amsterdam 2009

Thygesen, L.: Finding the Essence in Geographic Statistics – Promoting Informed Decision Making. International Journal of Spatial Data Infrastructures Research, 2010, Vol.5, 347-364

Résumé (Abstract)

Les statistiques officielles sont une grande source de connaissance et servent de base pour des décisions prises par de nombreux acteurs de la société mais leur potentiel reste cependant sous-exploité car la plupart des individus ne connaissent pas les statistiques, ne s'y intéressent pas et ne les comprennent pas.

Ce document met en évidence des travaux récents de Statistics Denmark sur l'utilisation croissante des statistiques géographiques à travers la visualisation sur le web. Il montre comment un outil de visualisation a été mis en place afin de donner la possibilité à différentes catégories d'utilisateurs de statistiques de comprendre la dimension géographique des phénomènes de société et d'identifier les faits cachés derrière les données. L'outil est destiné aux chercheurs régionaux, aux planificateurs locaux et régionaux ainsi qu'aux médias et au grand public, comme par exemple les citoyens s'intéressant à ce qui se passe dans leur propre communauté. Cet outil offre donc de multiples possibilités de présentation des données et s'applique à différents niveaux de détail géographique.