Visualising official statistics

Statistics New Zealand Working Paper No 11–02

Sharleen Forbes,[1] Martin Ralphs,[2]

Rosemary Goodyear,[1] Nathaniel Pihama[1]

[1] Statistics New Zealand, Wellington, New Zealand

[2] Office for National Statistics, United Kingdom

newzealand.govt.nz

Crown copyright ©
This work is licensed under the Creative Commons Attribution 3.0 New Zealand licence. You are free to copy, distribute, and adapt the work for non-commercial purposes, as long as you attribute the work to Statistics NZ and abide by the other licence terms. Please note you may not use any departmental or governmental emblem, logo, or coat of arms in any way that infringes any provision of the Flags, Emblems, and Names Protection Act 1981. Use the wording 'Statistics New Zealand' in your attribution, not the Statistics NZ logo.

Liability
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation
Forbes, S, Ralphs, M, Goodyear, R, & Pihama, N (2011). *Visualising official statistics* (Statistics New Zealand Working Paper No 11–02). Wellington: Statistics New Zealand

ISSN 1179-934X
ISBN 978-0-478-37719-4 (online)

Published in August 2011 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

info@stats.govt.nz
sharleen.forbes@stats.govt.nz

Phone toll-free 0508 525 525
Phone international +64 4 931 4610

www.stats.govt.nz

Contents

List of figures

Figures by chapter

Abstract

In general, national statistics offices produce a range of key national economic indicators and population estimates. They may also produce other social, cultural, and environmental statistics, but not always. This paper defines the statistics generated by government agencies, whether from administrative or survey data, as official statistics. One way to increase the usefulness of these statistics is to present them in forms that are easy for the user to interpret. Increasing computing power and the Internet have massively increased the ability to store, access, and analyse very large amounts of data. The Internet has itself become an analysis tool. This, together with the availability of many new open-source data visualisation methods and spatial tools, has enabled new ways of accessing and interpreting data. These new tools can open up the wealth of information held in official statistics to less statistically literate users, whether government advisors or members of the public. They are not simply presentation tools; they also enable new levels of data exploration and can be useful for the formulation of hypotheses. This paper demonstrates some recent data visualisation applications that have been used with a variety of official statistics, gives examples of analytical or policy uses, and makes some suggestions for further enhancements. The visualisations include static graphs; dynamic graphs (that show changes over time); interactive graphs (that allow the user to explore and interrogate the data); and mapping tools and software that integrate standard statistical analysis, graphics, and maps.

Keywords: data visualisation, official statistics

1 Introduction

Advances in information technology (IT) and in particular, the Internet, have massively increased the ability to store, visualise, and analyse very large amounts of information or data. This has caused an explosion in the availability of data visualisation and statistical analysis tools that provide new methods for improving access to, and interpretation of data. As the New Zealand Government Statistician, Geoff Bascand, stated "… the burst of creative advancement in presentation and dissemination is a truly very modern phenomenon," (Bascand, 2009). Many of these tools are open-source software able to be downloaded from the Internet at no cost to the user, such as the R Project (www.r-project.org/). Additionally, the web itself has become an analysis tool – the user often only needs an Internet browser to be able to explore data, without the need for expensive statistical software. Tools like Google Maps (http://maps.google.com/) and Google Earth (www.google.com/earth/index.html) have brought spatial data visualisation (geo-visualisation) technologies into the public domain. These tools are not simply for presentation, but also enable a transition from the traditional static graphs to dynamic and interactive graphics that can be used in the formulation of hypotheses. Open-source statistical and visualisation software have revolutionised the ability to analyse data – they have placed sophisticated analytical capability in the hands of the general public without the need for expensive licenses. The caveat is that users also need to know how to use these tools correctly!

A number of national statistics offices (NSOs) make data available in new visual formats. A recent special issue of the *Statistical Journal of the International Association for Official Statistics, 25*(3–4) (2008) was devoted to 'Web 2.0 and Official Statistics'. The papers presented included a discussion on the impact of the Internet on access to and the use of official statistics, the critical steps needed to turn official statistics into knowledge, and the analysis of data visualisation trends.

This paper focuses specifically on new data visualisation tools and their uses. Many of these tools are interactive (users have some element of control) or dynamic (show how data has changed over time). By their nature, these graphics are best demonstrated by playing with them directly through the appropriate website. Therefore, although screen shots of the visualisations discussed here are supplied, the reader is encouraged to explore the visualisations further using the website addresses given. Some of the costs and benefits of data visualisation are also discussed. Comments on different types of visualisations, and possible improvements, are also given together with a number of examples demonstrating how these visualisations could, or have been, used in decision making.

2 Official statistics: The backbone of government policy decision making

Official statistics are, typically, statistics that are collected on behalf of, originating from, and being of key importance to a national government regardless of whether they are generated from survey or administrative data.[1] In some respects, official statistics provide the backbone of social, environmental, and economic policy development. As shown in figure 1, the Census of Population and Dwellings (census) data can be considered the spine of social data as it is the base from which many other social statistics (such as fertility and mortality rates) are derived. The census not only provides a common denominator for comparisons across different areas of social policy (such as education, health, or employment) or across different sections of society (such as ethnic or age groups), but also provides detailed information on the structure and changes in the population.

1 There is no internationally accepted definition for official statistics that explicitly accommodates the considerable variety between countries in the amount and type of statistics collected and disseminated. Even the United Nations Statistical Commission, the first international body established for official statistics in 1947 (United Nations Statistical Commission, 2010) does not provide an explicit definition of this concept. Rather their role is to be the highest decision making body for international statistics setting standards for official statistics, developing principles, concepts and methods and assisting with the implementation of these at national and international levels.

A Dictionary of Sociology (Marshall, 1998) defines official statistics as "Statistical information produced, collated, and disseminated by national governments, their agencies, and the international bodies which link them. These data are almost invariably nationally representative, because they are obtained from complete censuses or very large-scale national sample surveys, and they usually seek to present definitive information conforming to international definitions and classifications or other well-established conventions".

In New Zealand, the 1975 Statistics Act states that "Official statistics shall be collected to provide information required by the Executive Government of New Zealand, Government Departments, local authorities, and businesses for the purpose of making policy decisions, and to facilitate the appreciation of economic, social, demographic, and other matters of interest to the said Government, Government Departments, local authorities, businesses, and to the general public." (New Zealand Government, 1975). Compared with Marshall’s definition, official statistics, as defined in the 1975 Statistics Act also include statistics generated from administrative data sets (such as tax, education, health and social services registers) as well as those from survey data.

Figure 1

Official statistics as the backbone of social policy advice

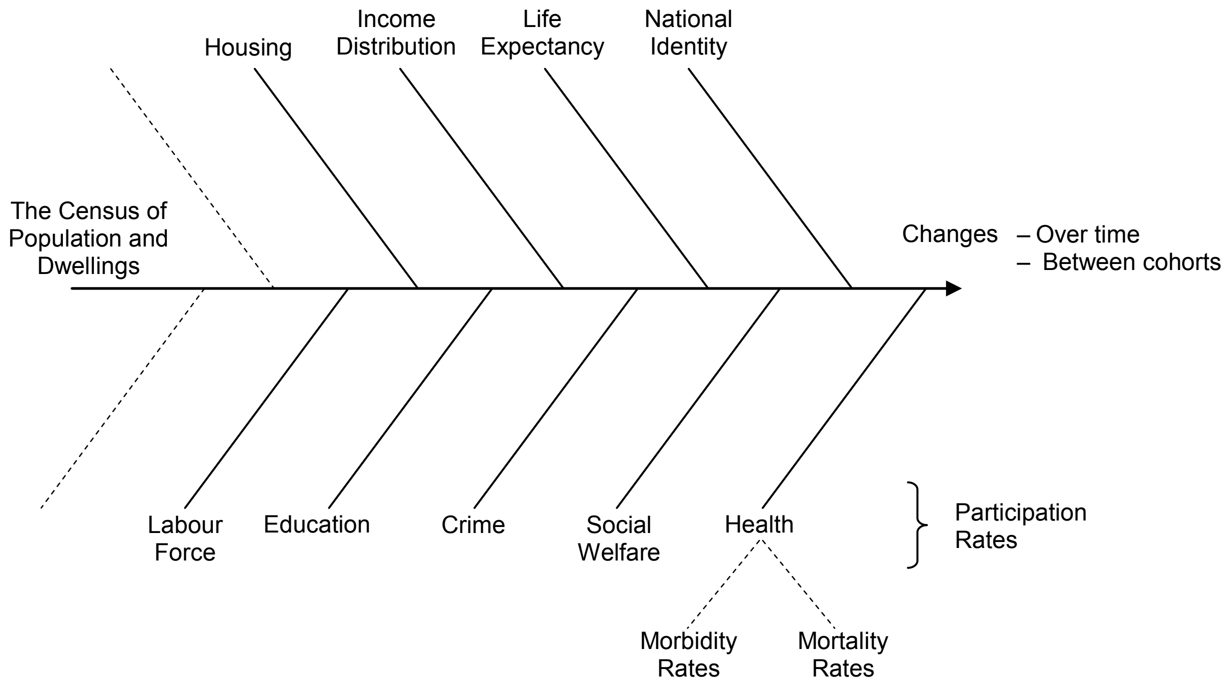
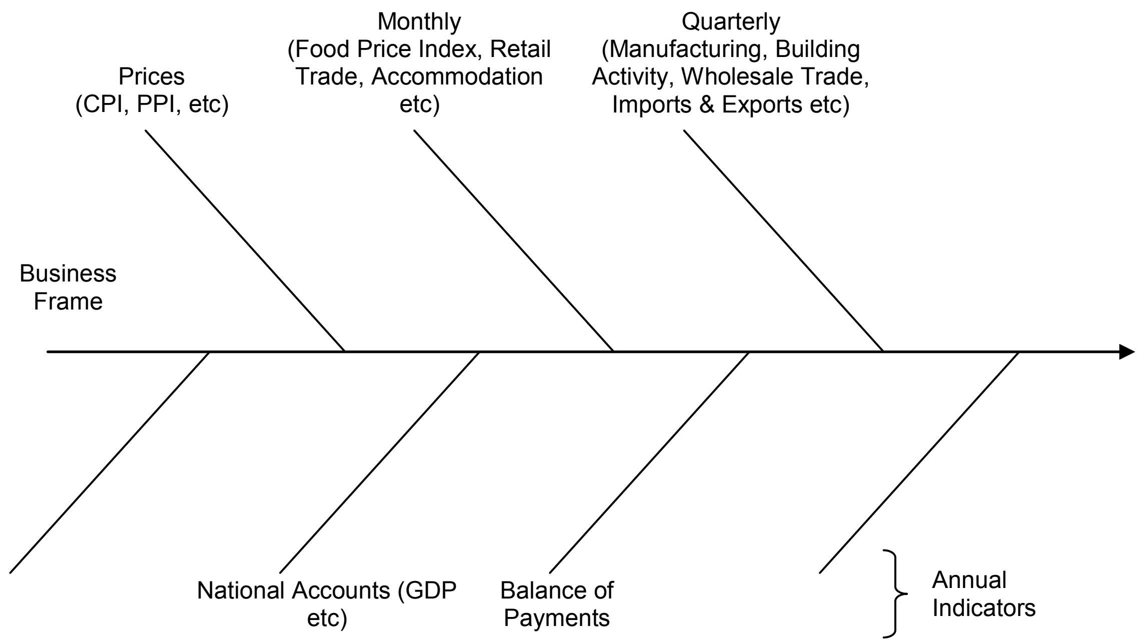


Figure 2

Official statistics as the backbone of economic policy advice



Like the census, business registers comprise a census of the enterprises (eg businesses, schools, hospitals) above a specified trading value that contribute to a country's economy. Business registers or frames can be considered the spine from which the collections (surveys, censuses, or administrative data) needed to collect economic data are designed (figure 2). The information from these surveys is the basis of macroeconomic indicators, such as the gross domestic product (GDP) and the consumers price index (CPI), used, for example, to determine monetary policy.

Government advisors now deal with a wide and increasing range of statistics. In some countries, this includes new statistics created from longitudinal surveys and integrated administrative datasets that are difficult to analyse or interpret. These in particular, may need to be presented in ways that help government advisors, community groups, businesses, and members of the public to understand them and shape their perception of society.

Dennis Trewin (2010), former Australian government statistician and president of the International Statistics Institute, divided official statistics into three levels:

1. indicators – statistics from data collections, with an increasing emphasis on time series
2. intermediate statistics – more detailed aggregate data that investigates relationships between variables, or relationships between data from multiple data collections
3. microdata – 'raw' collection data, often used for complex and/or specific analysis.

Each level of official statistics serves a different role in policy decision making. Trewin (2010) suggested that policy influencers are the major users of macro-data (indicators), policy advisors of intermediate (aggregate) data, and policy researchers of microdata.

3 Obligations on official statisticians

Official statistics are vital to good national and international decision making and there are obligations on official statisticians to ensure that the statistics they produce are of high quality, easily accessible, and interpretable by their users. This international recognition was in part generated by the economic changes of the 1980s with "the need for complete transformation of the national statistical systems. Part of this transformation process was about redefining the role of official statistics, as well as making it clear to governments and other users of statistics that a good system of official statistics must meet certain general criteria" (United Nations Statistical Commission, 2010).

The need for general criteria became the motivation for the Fundamental Principles of Official Statistics, which was adopted by the United Nations Statistical Commission in 1994 (United Nations, 2003). In summary, these principles cover:

- the provision of equal access to official statistics
- the use of professional and scientific processes
- robust analyses and provision of metadata
- official advice on the misinterpretation or misuse of data
- the use of appropriate sources of data to maintain quality but minimise respondent burden
- the confidentiality of individuals' information
- transparency of policies and procedures
- national coordination of official statistics
- international comparability of official statistics
- international cooperation.

While these principles do not explicitly discuss the way statistical information is presented, the first principle is about providing equal access for all to a comprehensive range of social and economic measures. New visualisations of currently available data increase the accessibility of these official statistics. Similarly, Thygesen and Sundgren (2008) state that the Organisation for Economic Co-operation and Development (OECD) "views quality in terms of seven dimensions: relevance; accuracy; credibility; timeliness; accessibility; interpretability and coherence," and noted that "it is not enough to have good data, the official statistics must also perform well on credibility, accessibility and interpretability."

Other types of statistics, such as sport, business, and recreation statistics, are increasingly being presented visually on the Internet. There is a growing expectation for official statistics to be presented in these forms as well. As ten Bosch and de Jonge (2008) state, official statistics 'have a much longer history of dissemination compared with the existence of the internet,' with official statistics being published in many countries for well over a century. While NSOs have been publishing official statistics on the Internet since the 1990s, most official statistics are presented on the web in a form similar to paper publications. There is, however, a growing trend to make use of modern visualisation tools to increase the interpretability of official statistics.

The major investment for the NSO is in staff time – to use these new tools to create appropriate visualisations and animations, or to reformat datasets. The latter is particularly true of visualisations involving maps (geo-visualisations) when geographic locations need to be linked to statistical data. The demand for low-level geographic information in official datasets has been driven by the widespread use of geographic information systems (GIS), as the user interfaces of Google Earth and Google Maps (www.google.com) have become simpler.

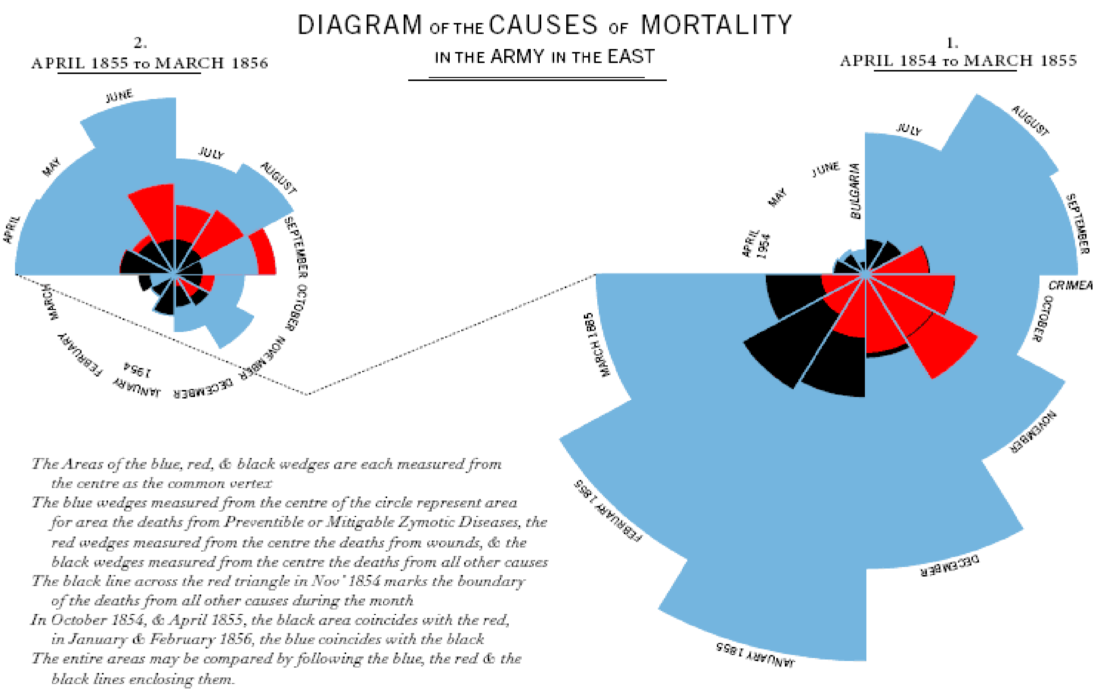
It is acknowledged that there are currently some gaps in official statistics, particularly with global issues. These gaps must be, and are being addressed by international bodies such as the United Nations Statistical Commission and the OECD. But in the long term, ensuring that individuals can access and use currently available official statistics may be of equal importance.

4 Data visualisation, a brief history

Friedman (2008) stated that the "main goal of data visualisation is to communicate information clearly and effectively through graphical means". This is not a new idea. Hidalgo (2010) suggests that one of the first modern users of graphical displays (charts) in England was Joseph Priestley in 1769, followed in 1801 by William Playfair who was credited with the creation of the bar chart and pie chart. Two graphics that were the basis of policy change at this time were John Snow's 1854 'Broad Street Map' which was used to locate the origin of an outbreak of cholera and the 1858 'Rose' of Florence Nightingale. 'Rose' is a diagram of the causes of death in the Crimean war, which showed that most soldiers died from preventable diseases rather than war injuries (figure 3). Small (1998) contends that while Florence Nightingale was not the first person to use graphics to present statistics, "she may have been the first to use them for persuading people of the need for change".

Figure 3

Florence Nightingale's 'Rose' (reproduced from Small, 1988)



Although data can be visualised in a multiple of ways, **good** examples of data visualisation share common elements. A good description of graphical excellence, which can also be generalised to data visualisation, is offered by Tufte (2001, p13). He states:

"Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

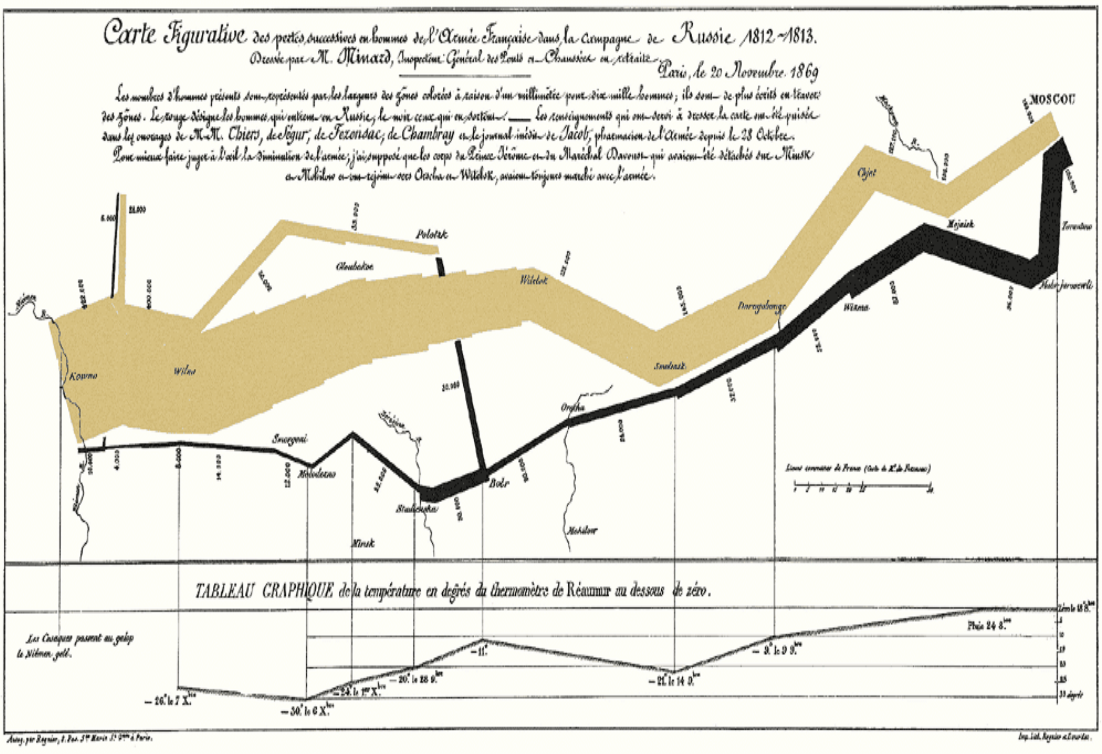
- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

Graphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations."

As an example of graphical excellence, Tufte cites an 1869 graphic by Charles Minard showing Napoléon's march to Moscow (figure 4) as one of the most famous and enduring data visualisations. In summary, good graphs should meet four criteria: simple, easy to interpret, memorable, and tell a story.

Figure 4

Minard's diagram of Napoleon's march to Moscow (reproduced from Tufte, 2001)



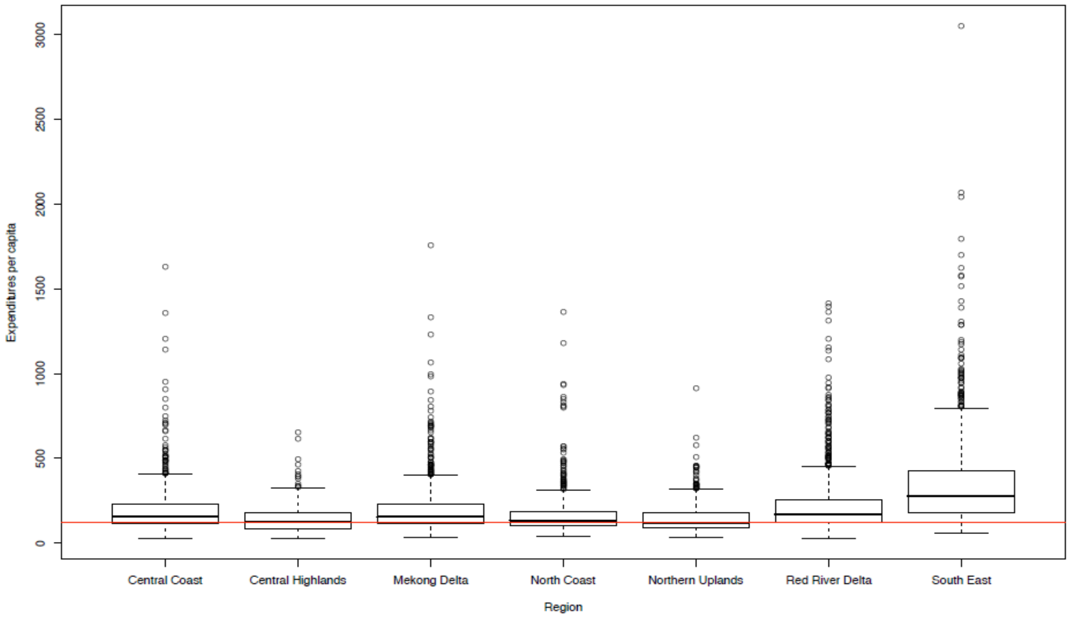
In addition to being a means for communication, data visualisation has also been used for analysis (ten Bosch and de Jonge, 2008). Most visualisations serve both purposes and the focus is largely determined by the viewpoint of the creator. For example, Tukey's (1977) box-plots were designed as an exploratory data analysis tool but are now widely used to communicate statistical results. Tukey himself presented a 'policy question' (eg differences in annual household per capita expenditures in the seven regions of Vietnam, see figure 5) as a box-plot example in his book, suggesting that he saw them as serving the dual purposes of analysis and communication.

While Playfair's pie and bar charts are still commonly used for displaying data (as shown in figure 6) there have been modern graphical innovations. These innovations have enabled visualisations to be used more and more as an analytical tool because patterns, trends, and relationships between variables and distributional properties can be more apparent than in published tables. Among the many possible analytical uses are:

- data modelling
- data interrogation (eg allows the user to explore the data, build their own tables)
- data reduction (reduces complex datasets so that its characteristics can be seen)
- time series exploration
- integration of graphs, tables, statistical analysis, and maps.

Figure 5

Box-plots of per capita expenditure by region in Vietnam (reproduced from Tukey, 1977)



5 New visualisations and their uses

The widespread adoption of computer technology has greatly expanded the methods and tools available for visualising data. The examples above are **static visualisations**. They provide a snapshot of the data and are commonly presented in printed form. Computer tools (often using an Internet connection) allow for more sophisticated, interactive visualisations. These **dynamic visualisations** allow the user to explore data in – potentially – more meaningful ways: customise what they see; look deeper into specific areas of the data; or use motion to track patterns over time and space.

The current trend is away from static to dynamic data visualisations. These new data visualisations can be interactive (the user can change the graphic in some way on demand) or dynamic (animated to demonstrate, for example, changes over time). While there are many papers on data visualisation (eg Cleveland, 1993; Post, Nielson, and Bonneau, 2002; Friedman, 2008) this paper focuses on applications that have been developed for official statistics. Different visualisations are examined for their use with official statistics and their role in policy decision making.

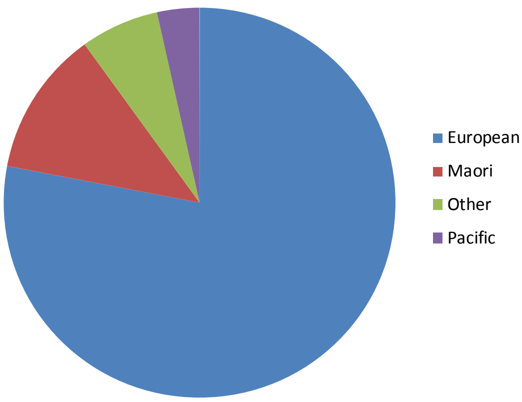
Some, such as the German Price Kaleidoscope, work well at the indicator level. Others, like the Slovenian Interactive Atlas and the UK and New Zealand Commuter View tools, work well at the aggregate and fine aggregate level. The Personal Inflation Calculator is at the microdata level, which allows users to input their own data.

Visualisation tools that reduce complexity or aid exploration of intricate datasets are powerful enablers for users. However, the fundamental principle of confidentiality in official statistics (United Nations, 2003) limits the level of direct user access to microdata, so that most visualisations can be seen only at the fine aggregate level.

The following examples give an overview and indicate the variety of new visualisations currently in use. Together with screen shots of each visualisation, the uniform resource locator (URL) or web address is given so readers can explore the visualisations as they are intended to be used. As with static graphs, good interactive and dynamic graphs should also meet the four criteria of being simple, easy to interpret (and use), memorable, and telling a story. It could be argued that these visualisations add the greatest value to official statistics when they are used as inputs into decision making. The visualisation is just a means of bringing to life the story that the statistics tell. For this reason, some of the practical uses that have, or could have, been made of the visualisations below are given.

Figure 6

Typical static data visualisations: bar chart (top left), scatterplot (top right), pie chart (bottom left) and table (bottom right).



Ethnicity	Average Weekly Income	Frequency
European	$ 585.95	156
Maori	$ 524.79	24
Other	$ 733.46	13
Pacific	$ 219.14	7
Overall	**$ 575.36**	**200**

Source: Synthetic Unit Record File (SURF) for Schools: Income survey, Statistics New Zealand.

6 Static data visualisations

Static data visualisations are still the most common method for presenting data. Software used to manipulate and analyse data typically produces tables and have a range of static data visualisations available, including: scatterplots, bar charts, and pie charts (as shown in figure 6). Most static graphs are easily created by users using readily accessible tools and these will remain in use because of their simplicity and clarity. The recent data visualisation innovations are based on the growing variety of information available, a greater understanding of how we process information, and the availability of new technologies. Data visualisation has begun to grow as a discipline.

An example of recent static graph innovation, created by Hidalgo (2010), displays the human development index (HDI) for a country as either a diamond or a tubular 'development tree'. This has explicitly been designed to communicate the results of a complex 'nested indicator'. The formula used to calculate the HDI has three equally weighted components: an education indicator composed of two elements, adult literacy (with a weighting of two-thirds) and enrolment in primary education (weighting of one-third); a life indicator (proportional to life expectancy); and an income indicator (proportional to GDP), each normalised so that it has a value between zero and one. Hidalgo represents this as a 'tree' with:

1. a trunk given by the value of the HDI
2. three branches, for each of the three components
3. two secondary branches on the education component representing adult literacy and primary school enrolment.

For the 'tubular' version of this tree (see figure 7): the height of the trunk is linearly proportional to the HDI value; the order in which the branches (boxes) come out from the tree indicates the relative contribution of each component (the bottom being the smallest contributor and the top the greatest contributor); the point at which the branch begins is proportional to its contribution to the HDI; the length of the branch is proportional to the actual value of the component; and the colour of the trunk is a weighted average of the colour of the components.

Figure 7

Hidalgo's tubular tree representation of the human development index

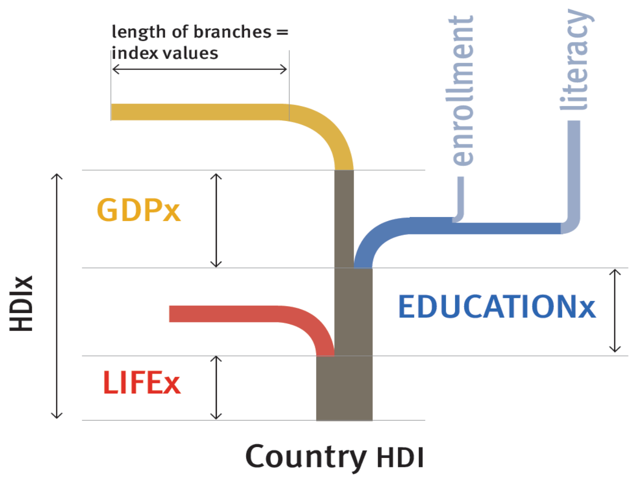
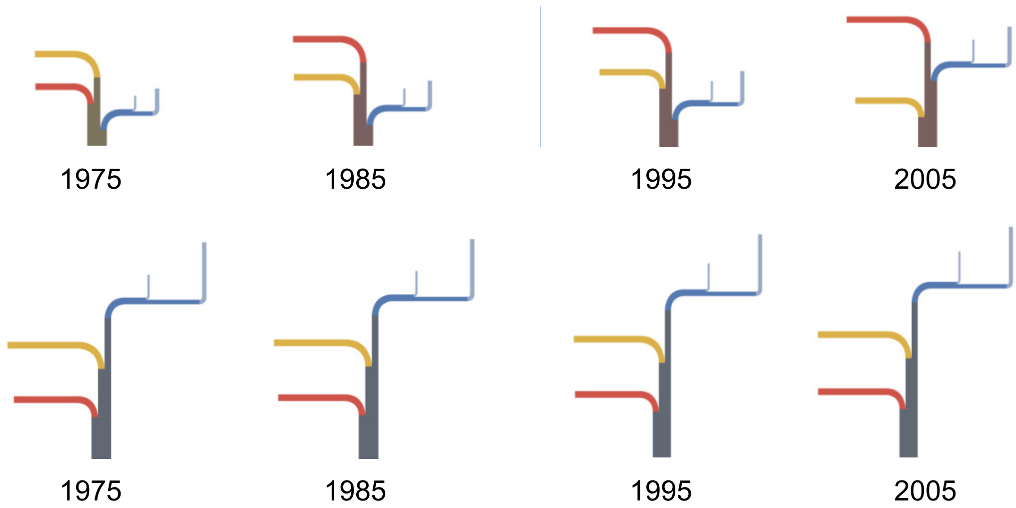


Figure 8

Human development index for Egypt (top) and New Zealand (bottom), 1975–2005



These trees can be used to examine the relative importance of the components over time within a country as in figure 8. This clearly shows Egypt's growth between 1975 and 2005, particularly in life expectancy and education and the change from income to life expectancy as the major contributor to the growth in the value of the HDI. The representation can also be used to compare across countries. Figure 8 compares Egypt's HDI with that of New Zealand, a country with free and compulsory education for children until their late teens, and a relatively high GDP. The HDI representation has another advantage for the user, allowing graphs to be individually created using an Internet link (http://chidalgo.com/Gallery/HDR2010/DevelopmentTreeApplications.zip).

7 Data modelling

The use of graphical techniques to explore the adequacy of statistical models is well known. The production of confidentialised microdata by NSOs for academic or public use, achieves some balance between maintaining confidentiality and providing access to data. An example of how graphs can be used with these types of datasets to develop appropriate models is given using data from a small synthetic unit record file (SURF) created from Statistics New Zealand's 2004 Income Survey, an annual supplement to the Household Labour Force Survey. This SURF retains many of the characteristics of the original survey dataset.

Figure 9a

Weekly income by highest educational qualification and hours worked

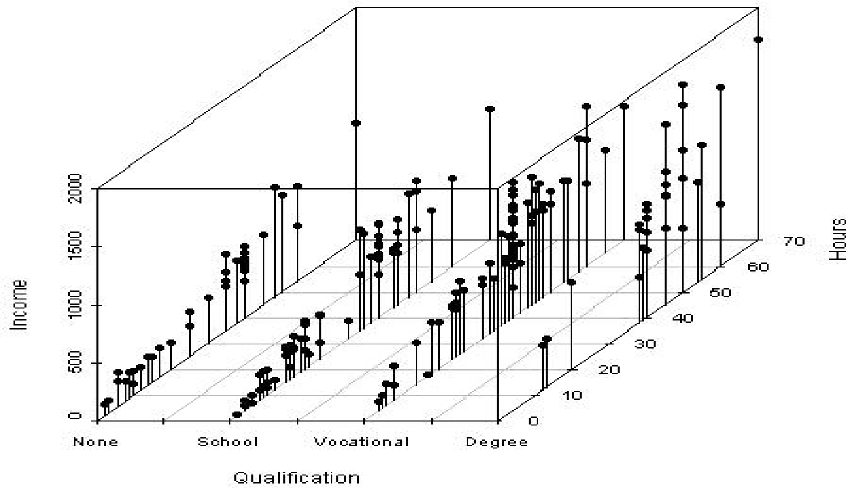
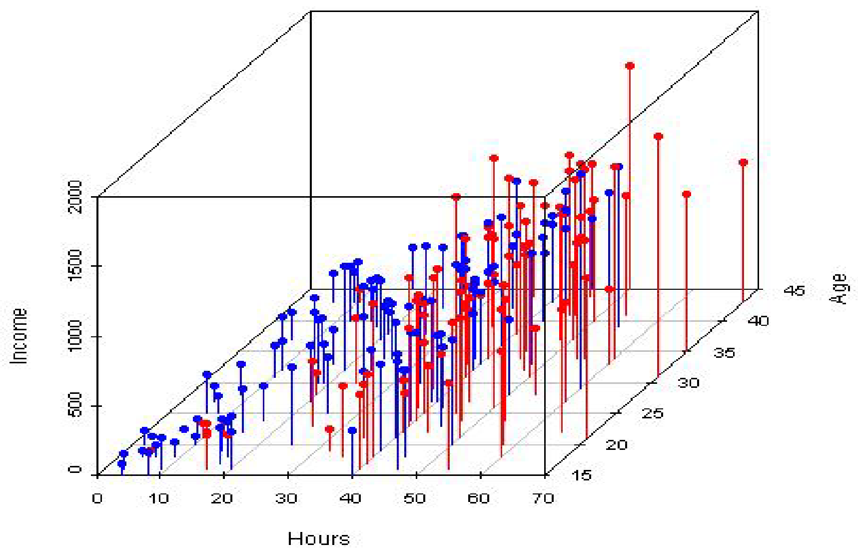


Figure 9b

Weekly income by hours worked and age



Both conceptually and educationally, there is a large step from simple regression using two variables to multiple regression models that may or may not incorporate interactive terms. A visual aid to determine the type of model to use is a three-dimensional (3-D) display (in this case a pin-graph where the heads of the pins form a 3-D scatterplot). Figures 9a and 9b were created using the open-source statistical software package 'R' (www.r-project.org/). These scatterplots allow analysts to visualise the degree of interaction (and decide whether interaction terms should be included in regression models) by looking at the consistency of pattern across highest education classes in figure 9a; and then decide whether separate models should be fitted to subsets of the data as in figure 9b, where part-time work is dominated by females (blue) and full-time work by males (red). Extension from three to four dimensions has been achieved by the use of colour. Other ways of increasing dimensionality are by changing the size, or intensity of colour, of individual data points, or by adding a dynamic feature to move between overlaid static graphs. Both of these methods were employed by Hans Rosling (2007) in the creation of his Gapminder graphs (see figure 14).

Improvements in computer technology have already enabled many national statistics offices to allow users to create their own tables or time series of official statistics (eg Table Builder and Infoshare on the Statistics NZ website, www.stats.govt.nz). The next step was to allow users to interrogate and interact with published data. This is a feature of almost all the visualisations that follow.

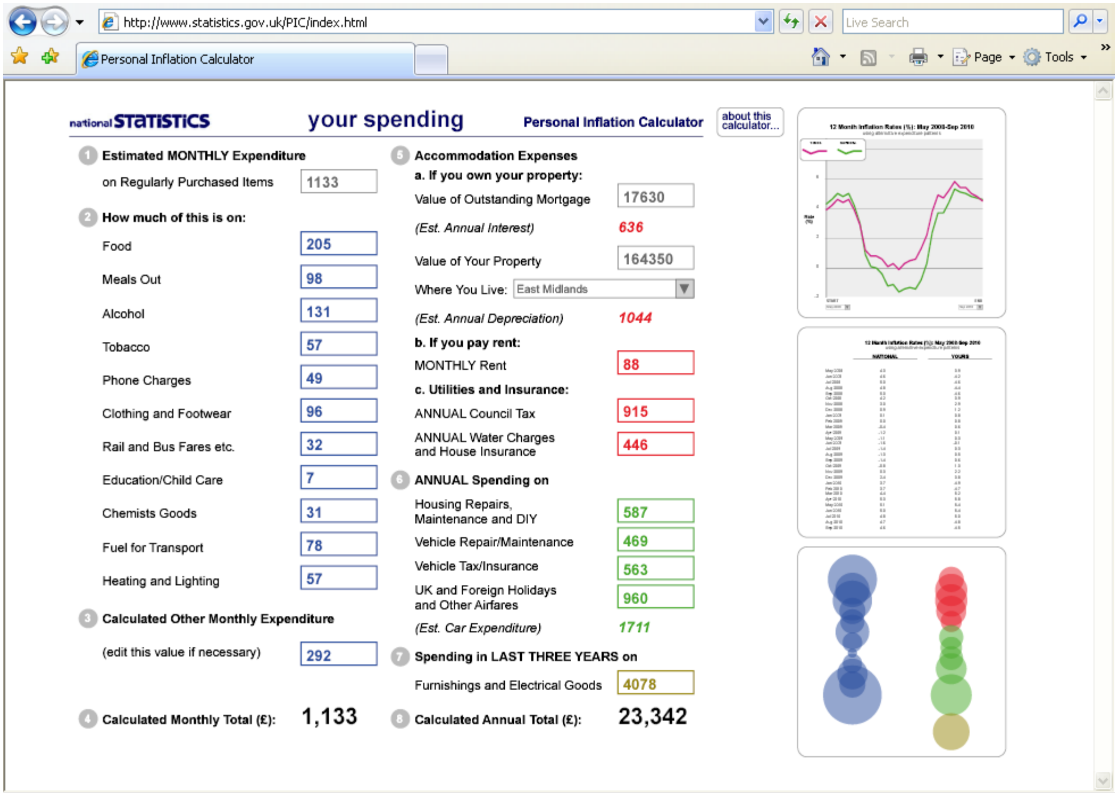
8 Dynamic data visualisations

Data interrogation with interactive visualisations

The number of tools that allow users to interact with, or interrogate, official statistics is growing rapidly, although many are in an early stage of development. Many of the following visualisations were written using scalable vector graphics, an open-source standard for web graphics introduced by W3C in 1999 (World Wide Web Consortium, 2001) and supported natively by almost all modern web browsers, or in the proprietary Adobe Flash format, which has seen widespread support.

Figure 10

United Kingdom Office for National Statistics' Personal Inflation Calculator



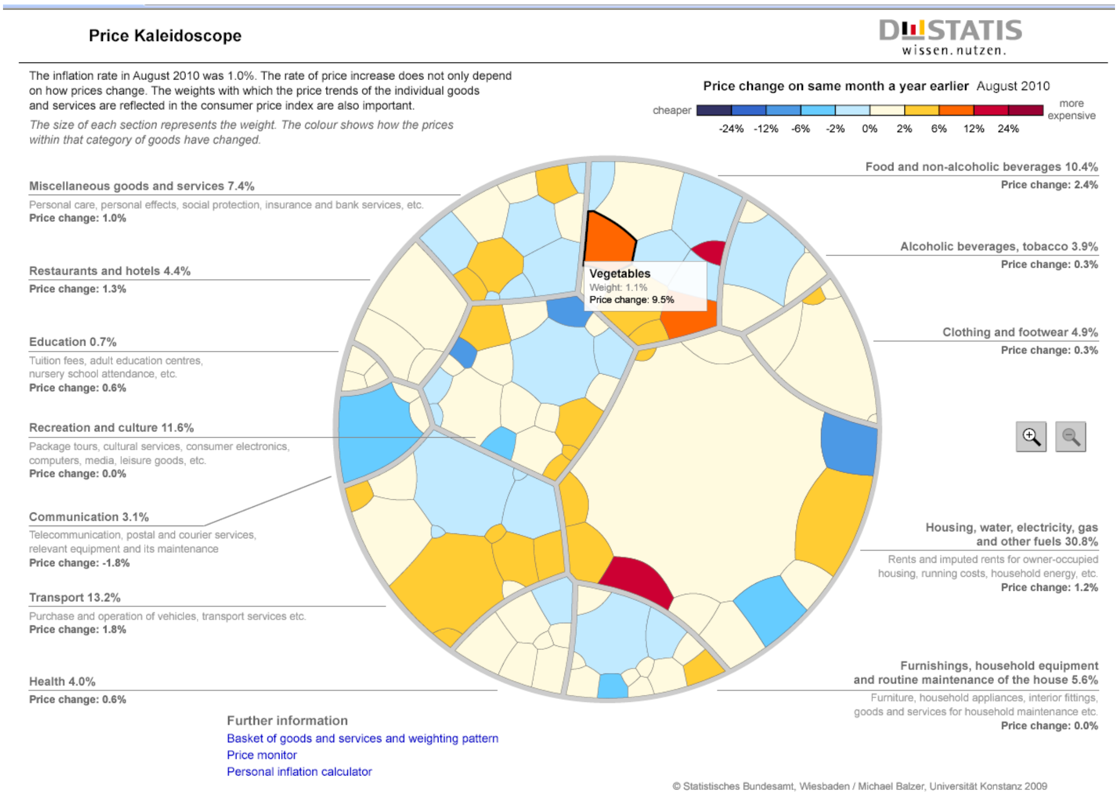
Source: www.statistics.gov.uk/PIC/

The United Kingdom Office for National Statistics (ONS), the Federal Statistical Office of Germany, and Statistics Netherlands have all been active players, developing visualisations that have then been adapted by other national statistics offices. Each of these three agencies offers users a personal inflation calculator similar to the ONS example given (figure 10).

Users can put their personal spending patterns into this tool and get an estimate of how their own inflation differs from the retail price index. It is reported that usage of the tool increased markedly in the UK when it was syndicated to the BBC, which released it together "with a map on which users could 'pin' their personal inflation rates and add comment," (Smith and Rogers, 2010). This type of feature is called a mash-up and although it successfully increased usage, one could debate that it is just another mechanism for attracting users to the website. However, familiarity with one visual tool may increase usage of other tools, and this is just one of a number of tools offered on each of these websites. For example, Statistics Netherlands also allows users to browse and create maps from their statistics dataset StatLine, which provides a business cycle tracer and interactive visualisations of national trade (imports and exports), income distributions, and other national statistics (ten Bosch and de Jonge 2008).

Figure 11

The German Price Kaleidoscope



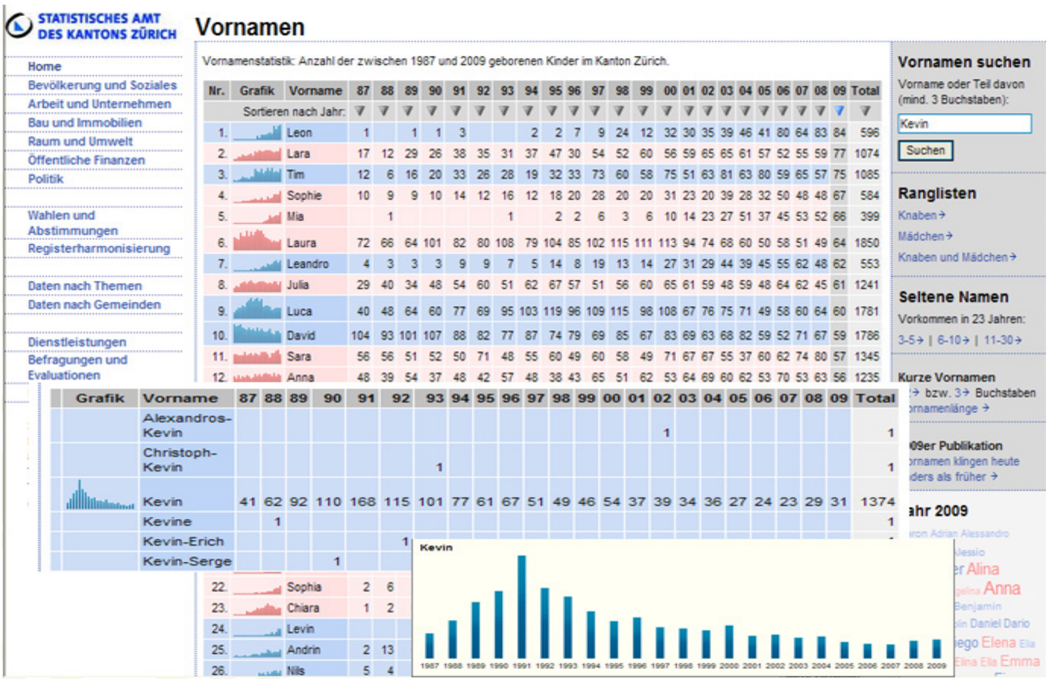
Source: www.destatis.de/e

The Federal Statistical Office of Germany offers dynamic population pyramids (discussed below, the Price Kaleidoscope, Regional Atlas, Business Cycle Monitor, Atlas of Foreign Trade Data, and the Air Transport Atlas). The Price Kaleidoscope (figure 11) is an interactive visualisation of the components of the consumer price index (CPI). The size of each section of the circle represents the contribution of that class of commodities (its weight) to the CPI; and the colour and intensity of colour of each section represents the percentage change since the last quarter. Users can interrogate the kaleidoscope to extract the exact figures for any commodity group (eg the weight and price change for vegetables, as shown in figure 11). The traditional way to release the CPI is as a single figure economic indicator of price change (in this case 1.0 percent) but all the data contained in the kaleidoscope is made available at the same time. This more detailed information is often eagerly sought by the media. A tool such as this kaleidoscope enables them to quickly access the precise data they need. As with all interactive tools, the kaleidoscope is instantly updated on the web as soon as the data is released. The data changes but the tool doesn't. This shows users the living nature of official statistics, constantly being renewed and updated.

While official statistics visualisations are developed for serious purposes, they can also provide an element of fun for the user. For example, a new interactive way of visualising the popularity of baby's names was developed by Zurich's Statistics Office in Switzerland. The tool has miniature bar graphs (called 'spark lines' after Tufte (2006)) inside the league table, which allows people to explore and to make conjectures on the possible reasons behind peaks in the popularity of different names. Alan Smith of ONS's Data Visualisation Centre conjectured somewhat flippantly that the peak in popularity of the name 'Kevin' in the early 1990s in Zurich could be due to the internationally popular Kevin Costner films *Dances With Wolves* and *The Bodyguard*. Of course, this may be completely spurious and would need further investigation, but the ability to 'explore' or 'play with' data is an important feature of both data visualisation and the development of statistical literacy skills (Wild and Pfannkuch, 1999).

Figure 12

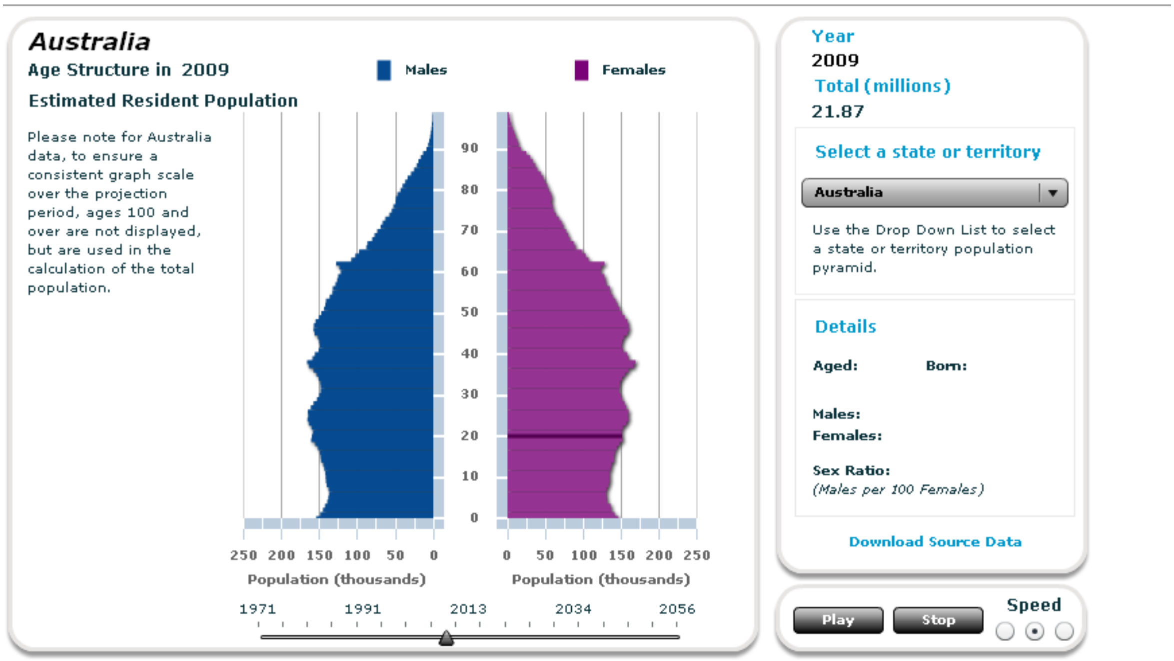
League table of most popular boy baby's names in the Swiss city of Zurich

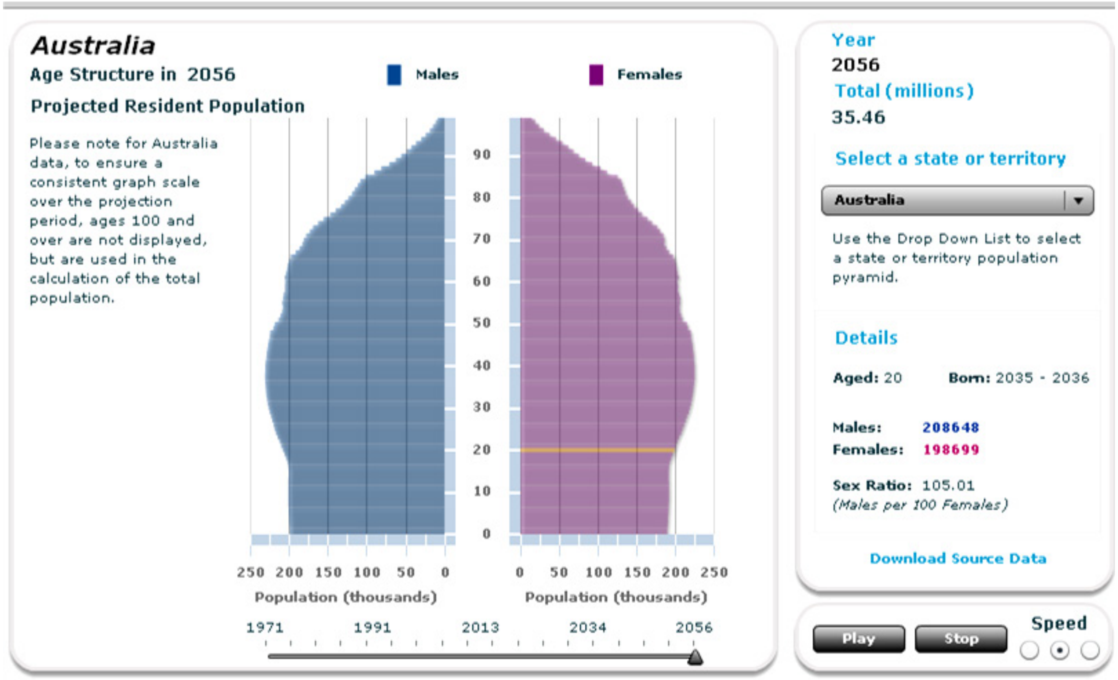


Source: http://www.statistik.zh.ch/vornamen/vorname.php?nam=&M=&F=&year=2006

Figure 13

Snapshot of the Australian Bureau of Statistics' population pyramid at year 2009 (top) and 2056 (below)





Source: www.abs.gov.au/websitedbs/d3310114.nsf/home/Population%20Pyramid%20-%20Australia

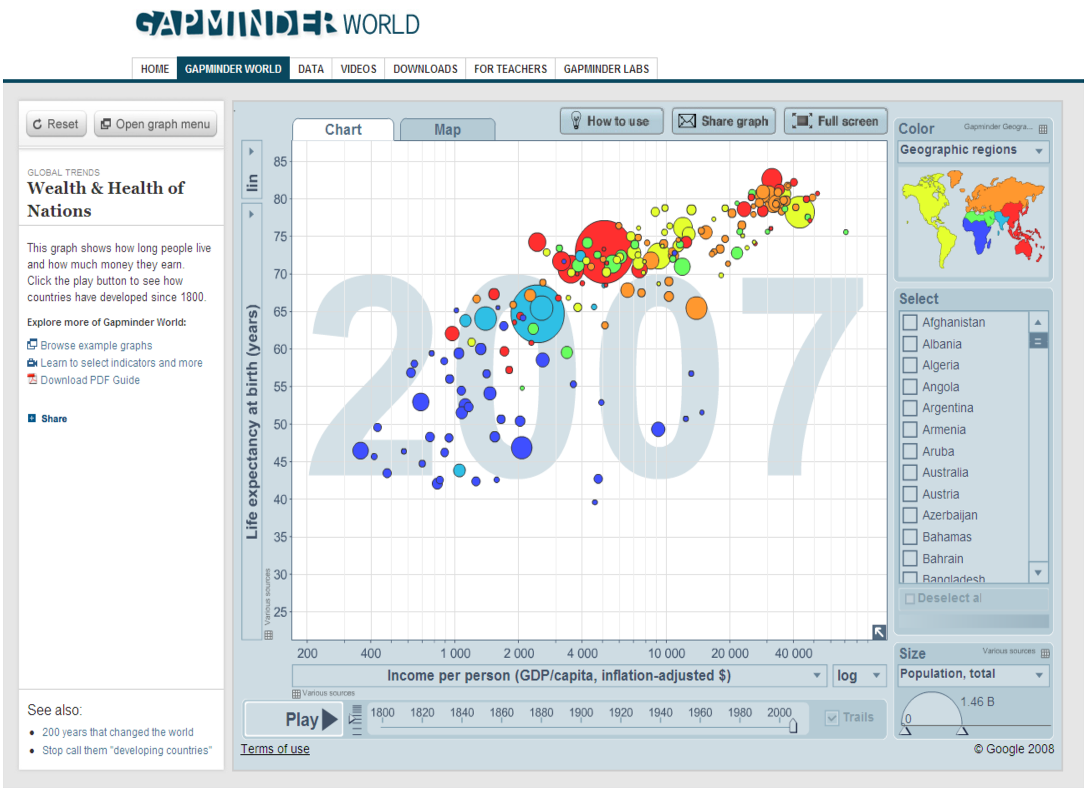
Data reduction and time series exploration

One way to view changes in large datasets over time is to overlap graphs (or maps) taken at different time points and then use an animation tool to 'play' these over time. An example is the Australian Bureau of Statistics' dynamic population pyramid (see figure 13) that clearly shows the changes in population structure as it 'ages'. The lighter shades in the second figure indicate that this is projected data, while the darker shades in the top figure indicates actual data. Dynamic population pyramids that contain overlaid graphs of different census and population projection data are now in relatively common usage by national statistics agencies. By playing these graphs, demographic effects such as momentum (population growth resulting from a youthful age structure, or population decline resulting from an older age structure) can be viewed. As Jackson (2001) says, these "two trends are often on a seemingly unavoidable collision course". Dynamic population pyramids provide a graphic of this process in action, leading up to and beyond the point (different for each country) where natural increase (growth) shifts to become natural decrease (decline). The ability to see differences between age cohorts as they progress over time in these graphs should also provide an incentive for more formal cohort analysis. As with many of these tools, dynamic population pyramids can be used to enhance the understanding and teaching of statistical concepts. This is an area for further research.

The Federal Statistical Office of Germany website contains possibly the most flexible and detailed version of dynamic population pyramids. Users can choose one of four projection models based on different underlying assumptions, select individual years for viewing, and track the progress of birth cohorts by clicking into the pyramid. The population can also be subdivided into three age classes.

Figure 14

Hans Rosling's Gapminder World



Source: www.gapminder.org

Probably the best known dynamic graph is Hans Rosling's (2007) bubble graph (figure 14). This graph makes clever use of colour and size to display five variables simultaneously: region (colour), relative population size of a country (size of bubble), two other variables chosen from a set of OECD indicators (on the vertical and horizontal axes), and time (as the dynamic variable). In a recent presentation to the International Conference on Teaching Statistics, ICOTS8, Rosling (2010) suggested that, not only does the exploration of data using these graphs raise new policy questions, but it can also dispel or challenge old 'myths'. For example, Gapminder World can be used to explore the relationship between life expectancy and gross domestic product (GDP) and demonstrate that, for individual countries, if there is any linear association between these it is only in small localised time periods. Gapminder World can also be used to show the relationship between decreasing child mortality and women's fertility rates, querying whether government efforts to control population growth should be aimed at improving child health and education.

Google (Wikipedia, 2010) acquired the 'Trendalyzer' animation software (www.gapminder.org/downloads/), which is used to create the Gapminder graphs. This software is now available for the public to download for free. Once downloaded and installed, graphs can be created without Internet access. Trendalyzer is used in Google Public Data Explorer (www.google.com/publicdata/home), which contains a range of data from organisations such as Eurostat, the World Bank, and the United States Bureau of Labor Statistics. The public data explorer website states that "As the charts and maps animate over time, the changes in the world become easier to understand. You don't have to be a data expert to navigate between different views, make your own comparisons, and share your findings. …Students, journalists, policy makers and everyone else can play with the tool to create visualisations of public data, link to them, or embed them in their own webpages. Embedded charts and links can update automatically so you're always sharing the latest available data."

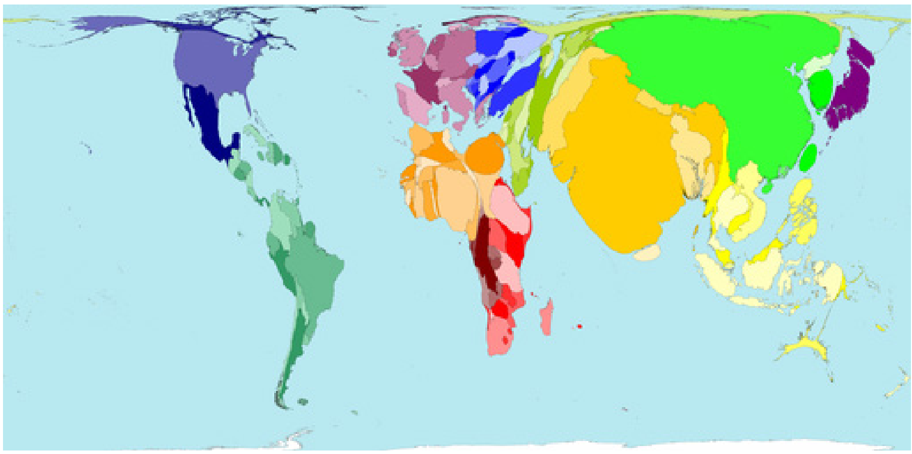
Maps and geo-visualisation

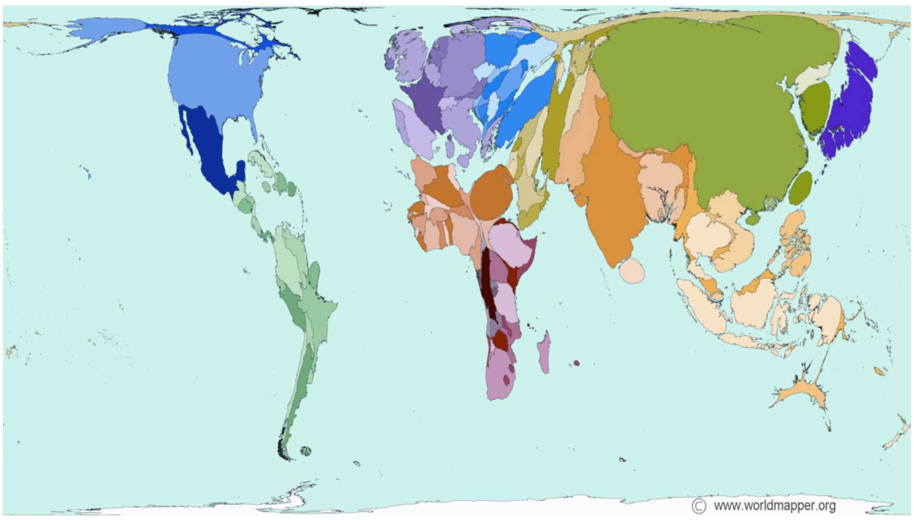
Many official statistics have an associated geography – new ways of visualising these are being explored at the global, national, and city level. Geo-visualisation is "conceptualised as a means of visually representing spatial data to better explore and understand patterns and relationships in the underlying information" (de Róiste et al, 2009). Public domain geo-visualisation tools (such as Google Maps and Google Earth) are only the beginning. The cartogram is a map where another variable is substituted for land area. Each area is resized according to the variable being used. Cartograms can be used to see how closely the variable is related to land area, or to compare areas. The cartograms in figure 15 are reproduced from the Worldmapper website and show: a country's relative contribution to the world's population, access to affordable drugs, and species extinction. Worldmapper also provides some animated graphs, for example, a graph of the world by daily income, age of death, and Internet users between 2000 and 2007. The data used comes from a variety of sources, such as the United Nations, the World Bank, and the World Health Organisation, and may be of variable quality.

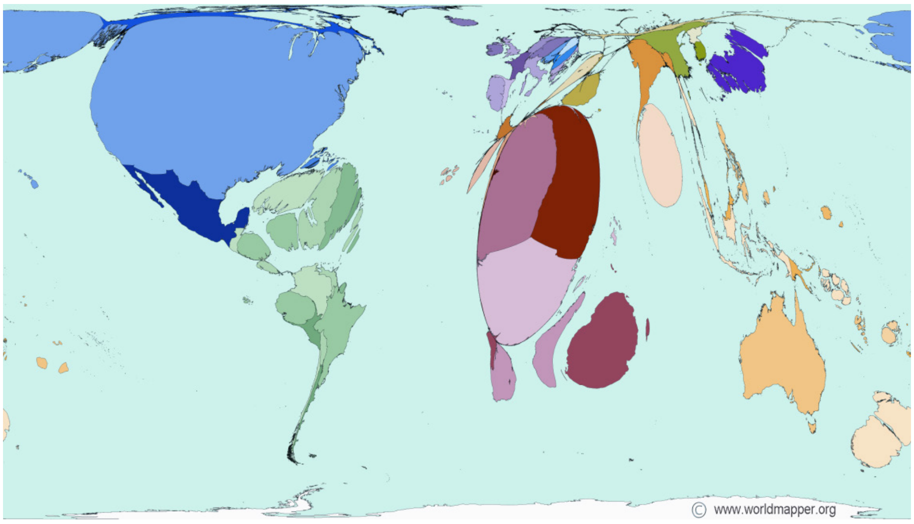
It should be remembered that the quality of the data underlying the visual tool is of critical importance. There have always been examples of misleading graphs, but new dynamic tools provide greater potential than ever to distract and bamboozle users with expensive-looking visualisations that serve to mislead rather than inform. It should be remembered that, if a visualisation looks slick and expensive, many users will assume that both the visualisation and the underlying data are correct (www.worldmapper.org).

Figure 15

The world by population size (top), Population with access to affordable drugs (middle), and Species extinctions occurring 1500–2004 (bottom).





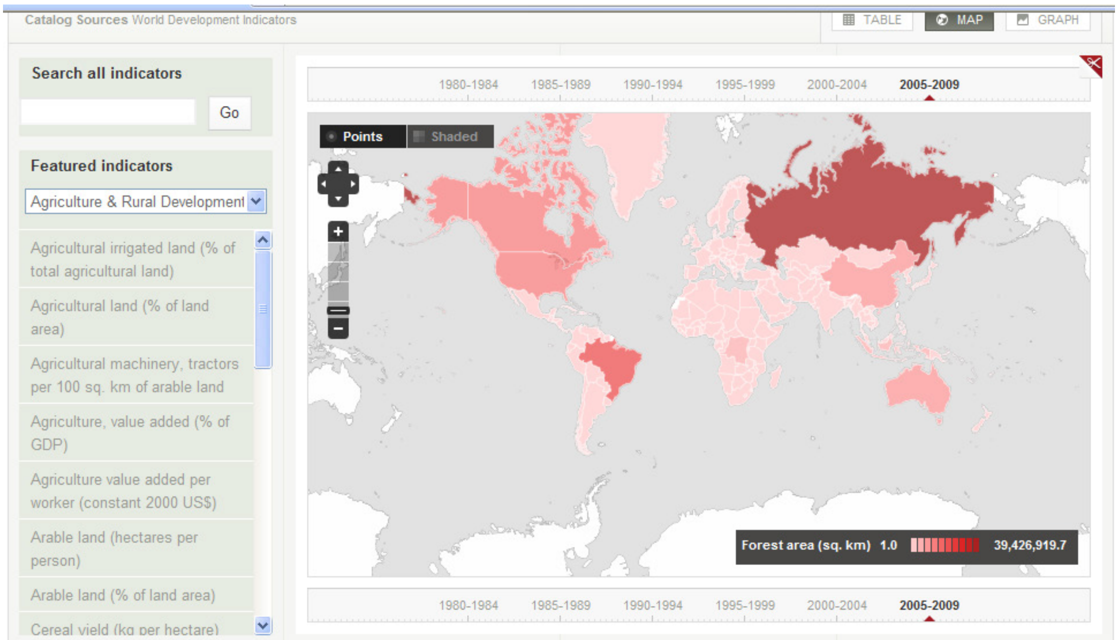


Source: www.worldmapper.org

An example of an interactive map is the World Bank's interactive atlas, which contains a range of demographic, social, economic, and environmental variables. The atlas also allows users to choose between maps, tables, and graphs of the data (figure 16). It has a 'data catalog' that includes a description of the available datasets, tales, and reports, and provides a direct link to the data. The provision of metadata (data about the dataset) is important so that users can understand where the data was produced, the characteristics of the data, and changes to variable definitions that may have been made over time. This information is necessary so that the quality of the dataset can be evaluated.

Figure 16

World Bank atlas



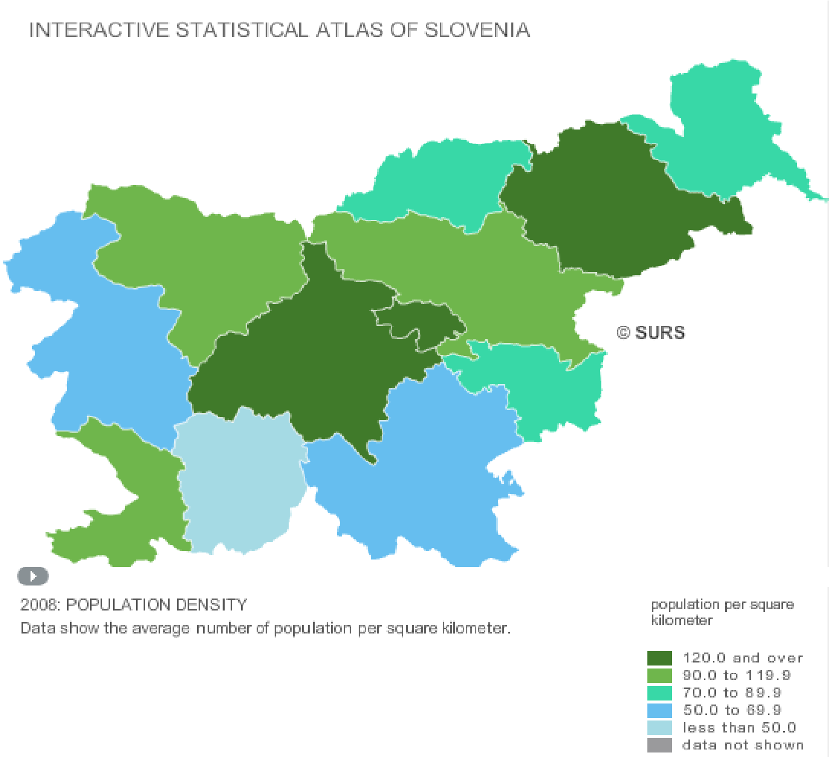
Source: www.worldmapper.org

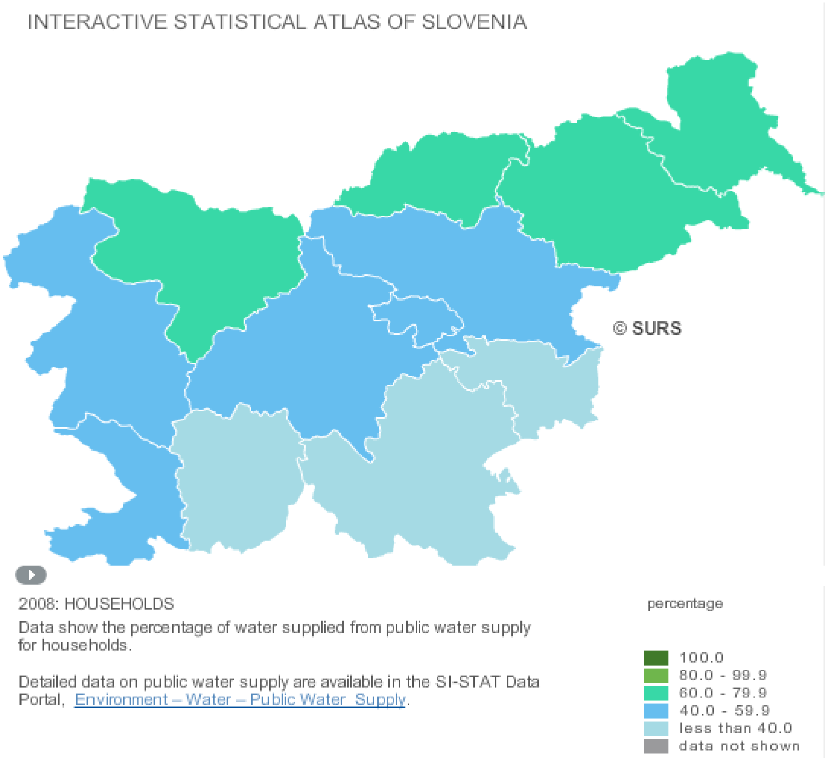
A number of national statistics offices provide a range of indicators in interactive map format (eg Statistics NZ's Interactive Boundary Maps www.stats.govt.nz). These indicators can then be used as input to a range of current policy issues. For example, in some countries, there may be a mismatch between population and water supply. Figure 17 uses the Slovenian Interactive Atlas from the Statistical Office of the Republic of

Slovenia to demonstrate how maps can be used to explore this mismatch and many other issues.

Figure 17

Population density by region in Slovenia (top), Percentage of household water provided from public water supplies (bottom).





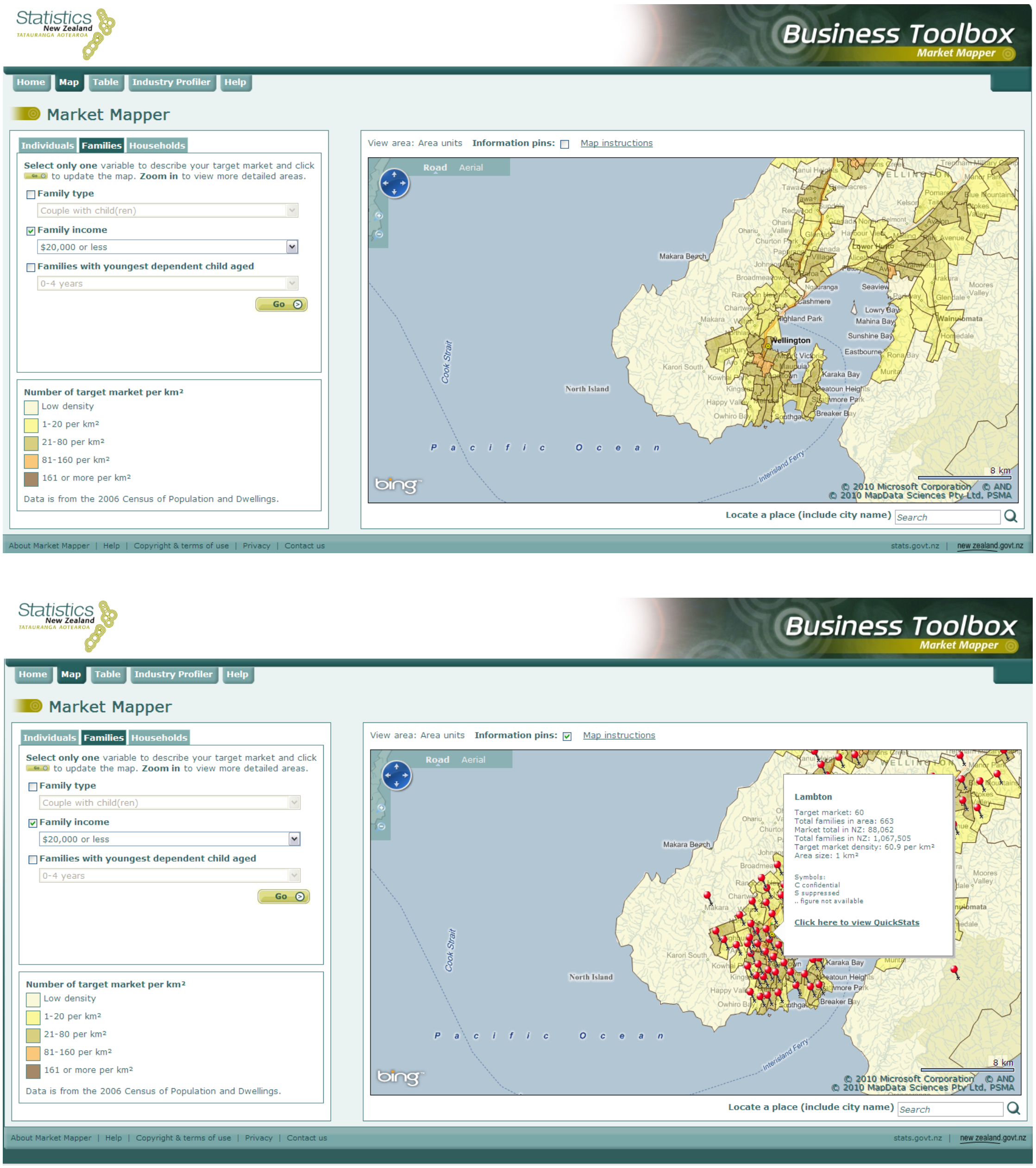
Source: www.stat.si/eng

The Market Mapper facility of Statistics NZ's Business Tool Box gives users the ability to explore the geographic distribution of target market groups (see figure 18 for an example on geographic distribution of couples with children, with a family income of less than

NZ$20,000, and the youngest dependent child aged five). The toolbox also allows users to get basic statistical profiles by industry type. Currently, users have limited choices – they will soon be demanding more refinements, such as individual years of age, ethnicity, and finer breakdowns of income or industry. To reiterate, the constraint on levels of access by users will ultimately be the confidentiality requirements of national statistics agencies.

Figure 18

Statistics NZ's Business Toolbox



Source: www.stats.govt.nz

Data reduction using maps

Maps can be used to reduce, manage, and present very large and complex datasets. An example of this type of data reduction is the interactive mapping tool 'Commuterview', which was released on DVD and distributed by Statistics NZ. This tool, based on a product from the UK Office for National Statistics (Office for National Statistics, 2008), uses scalable vector graphics (SVG). It displays linked information from two 2006 New Zealand Census of Population and Dwellings questions ‘Where do you live?’ and ‘Where do you work?’. Even at a high level of geography, such as territorial authority area of which there are 74 in New Zealand, a table of this data would be very large (over 5,000 cells). A small section of such a table is given in figure 19.

Figure 19

Commuting patterns in New Zealand

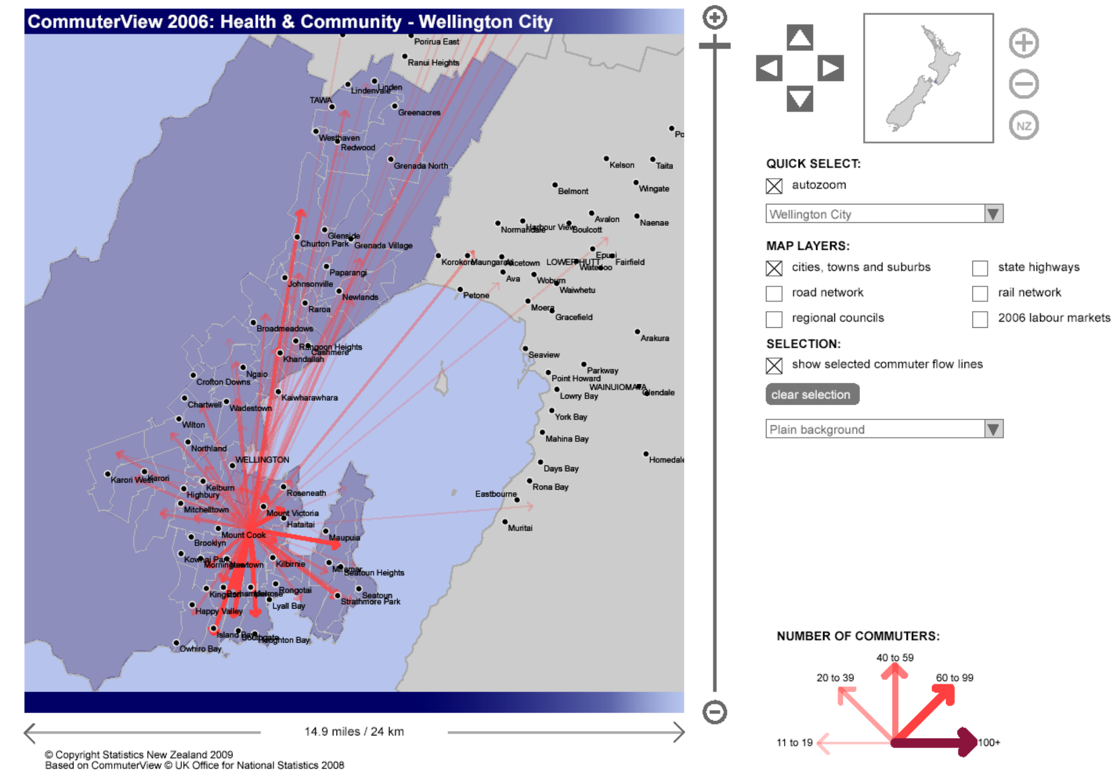
Residential territorial authority area by work territorial authority area

Residential territorial authority area	Work territorial authority area									
	Far North district	Whangarei district	Kaipara district	Rodney district	North Shore district	Waitakere city	Auckland city	Manukau city	Total Auckland	Papakura district
Far North district	16,860	396	36	30	36	12	108	42	201	9
Whangarei district	285	26,379	276	75	57	36	171	54	321	12
Kaipara district	42	327	5,931	315	33	12	69	27	138	-
Rodney district	30	48	126	21,183	6,822	1,701	5,706	627	14,856	54
North Shore district	48	63	24	1,755	58,383	1,905	28,188	2,604	91,077	180
Waitakere city	48	48	12	1,155	4,332	31,794	30,957	3,288	70,371	258
Auckland city	201	141	39	738	7,257	6,183	140,517	16,023	169,983	942
Manukau city	75	69	18	282	1,824	1,050	40,881	66,210	109,962	3,384
Total Auckland	372	321	93	3,930	71,793	40,932	240,543	88,122	441,396	4,764
Papakura district	9	15	-	33	177	84	3,894	5,079	9,231	6,567
Franklin district	12	15	3	48	171	99	3,117	3,720	7,110	1,869

The next level of geography consists of 1,927 area units resulting in a table of more than 3 million cells. The lowest level of geography has 41,392 meshblocks resulting in a table of almost 2 billion cells – too large for the human mind to assimilate. But data at these smaller levels of geography is most useful to city planners so they can provide adequate services (eg public transport). By providing this data on a map, which the user can zoom to an appropriate level, this massive table can be reduced, viewed, and used for planning purposes. The resulting ‘spider’ graphs show commuter flows for selected ethnic groups by modes of transport, industry, and occupation. For example, figure 20 shows the home locations of staff employed by the central public hospital in Wellington city. The map shows that many of the hospital staff live over the hills or at some distance from the hospital, which could be a problem in disaster mitigation planning as the city lies along a major fault line. This type of information is critical for risk management. These visualisations are very useful for any type of flow data, including internal migration data and the geography of the labour market (Ralphs & Goodyear, 2008).

Figure 20

New Zealand commuter flows: Work to home, Wellington city



Source: http://www.stats.govt.nz

Integration of graphs, maps, tables, and analysis

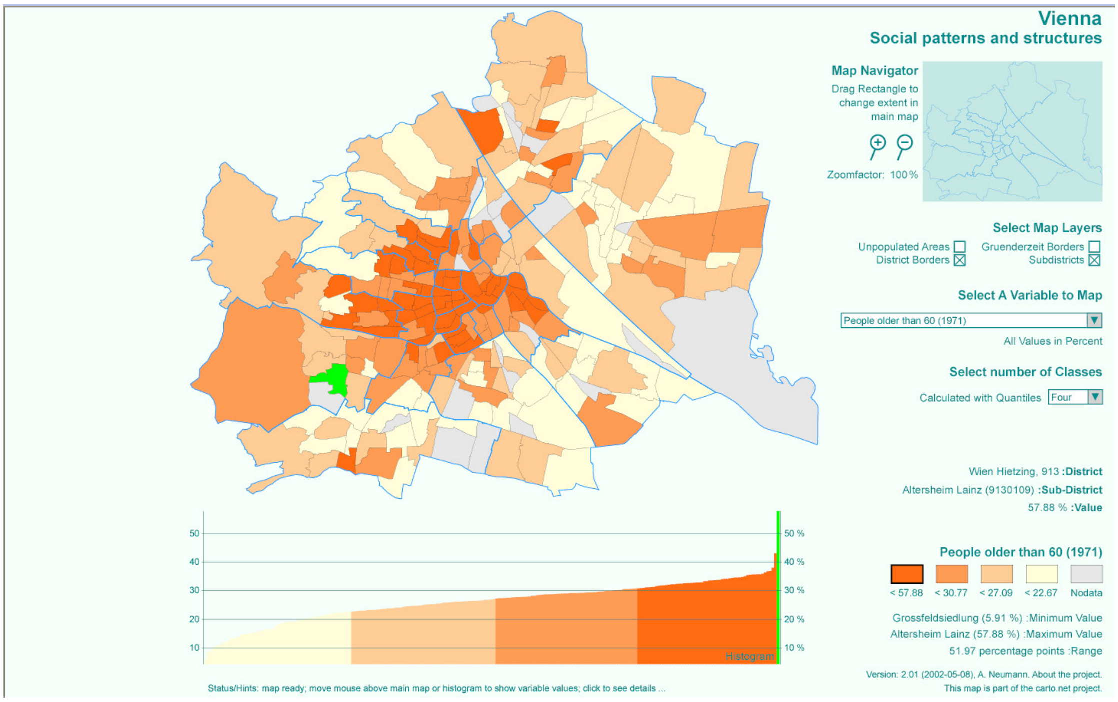
The latest visualisation tools combine graphs, maps, and data analysis. However, official statistics datasets must be geo-coded (the linking of geographic codes with other variable) before these tools can be used. Many national statistics offices store digital boundary datasets separately from their statistical outputs. To input the statistical data into a geographic information system (GIS) form requires some knowledge of, and access to, a GIS software system. Such software can be expensive and could be daunting for untrained staff. However, it is becoming more user-friendly. Apart from Google Maps, there are now several freely available open-source GIS systems available that make GIS technology free at the point of access: Quantum GIS (www.qgis.org), SAGA GIS (www.saga-gis.org), and GRASS GIS (http://grass.osgeo.org). These are exciting trends as they remove barriers to access to sophisticated technology. The 'R' statistical package mentioned earlier also has extensive support for mapping and geographical analysis. Although some agencies already publish digital boundary sets on their websites, there will be an increasing demand for them to provide integrated geospatial and statistical datasets.

Figure 21 shows a visualisation of the proportion of the population aged over 60 in Vienna City (Neumann, 2005). It links the region with its value on an 'ordered' bar graph. Many visualisations challenge our current protocols for graphs and this bar graph is an example. Gone is the idea of grouping into a range of categories; this has been replaced by the creation of a bar for each unique value with a replicate shown on the map. This particular tool contains only a small set of selected demographic and housing variables, some of which are available for three different time periods (1971, 1981, and 1991). It does not include an animation feature. However, the potential use of this type of visualisation for town planning is immense. For example, figure 20 shows that over 30 percent of the residents of many of the inner city areas are older than 60, but if services were only targeted to these areas then the area with the highest proportion (almost 60 percent) of older people (highlighted in figure 21) might miss out.

Interactive mapping tools can be used to show the geographic clustering of groups of particular policy interest (such as low socio-economic groups). According to de Róiste et al (2009) "geovisualisation is already established, mature and has the required theory to guide the production of effective visual displays, leading to insight into pattern and process over a geographic region". They state that "developments on two fronts are leading the expansion of geovisualisation. First, flexible tools have been created to allow users to interact with spatial data in a more fluid environment and second, increase in use of personal GPS (Global Positioning Systems) and web mapping has moved geovisualisation towards the mainstream and affected demand for the visualisation of spatial data". They argue that data exploration is possibly the most powerful use of these visualisations that build on the theoretical foundations of statistical graphics, cartography, and cognitive science.

Figure 21

Vienna City map showing the distribution of the population aged over 60

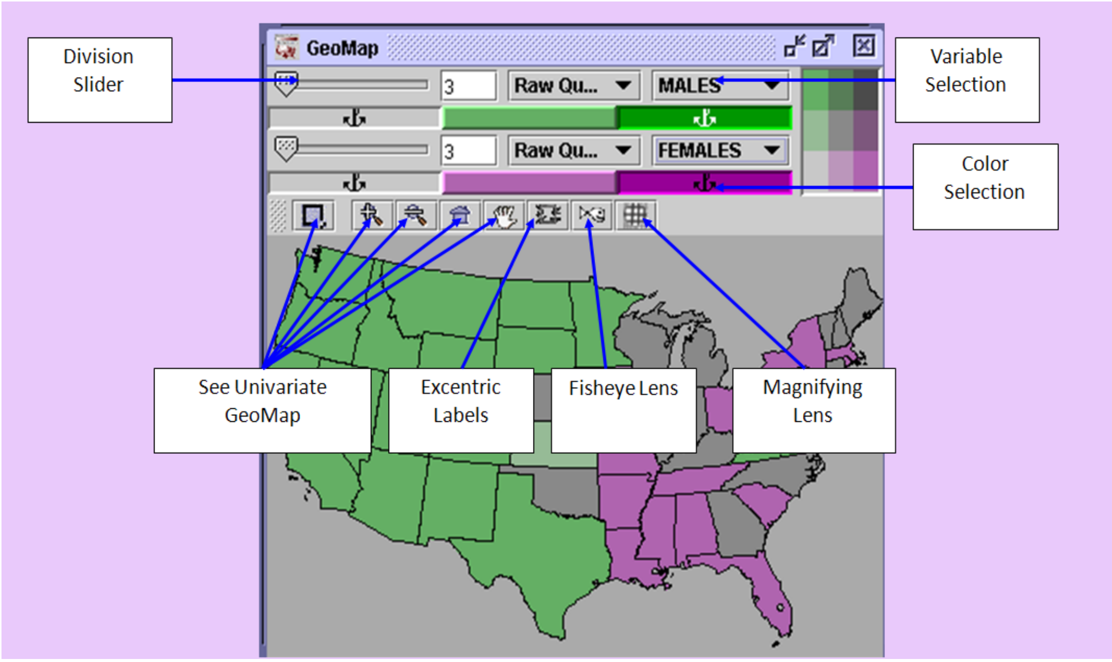


Source: www.carto.net/neumann/cartography/vienna/index_old.html

One of the innovative products for integrating maps and graphs is GeoVista (figure 22), a Java-based product created at the GeoVISTA Research Center at Penn State University (Gahegan et al, 2002). It is available as an open-source software at the Download Geoviz Toolkit webpage, www.geovista.psu.edu/grants/cdcesda/software/.

Figure 22

GeoVista bivariate map tool

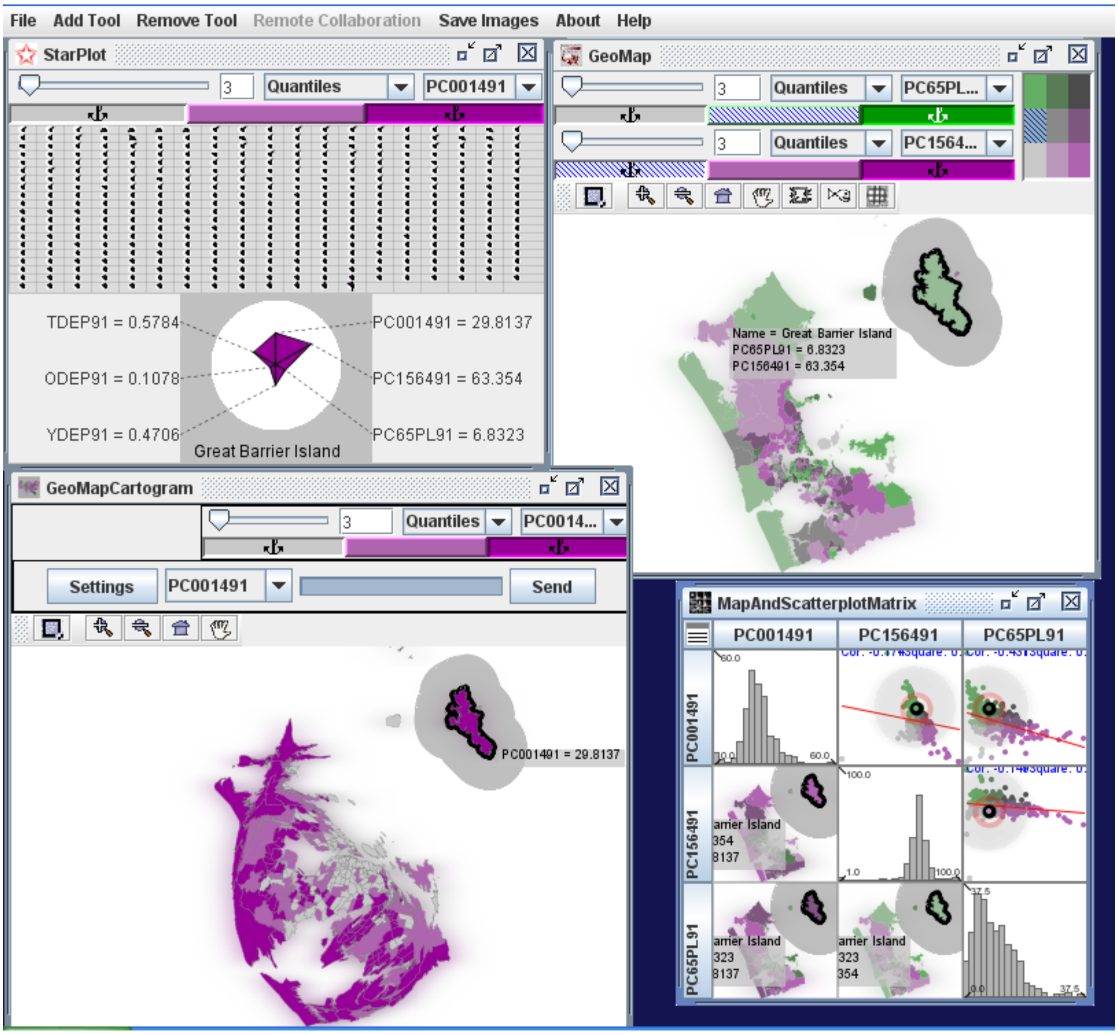


New Zealand census data for a range of derived variables (such as proportions in various age or ethnic groups, with certain household, family, or employment characteristics, median income, and average population growth) for Auckland city was geo-coded at area unit level by Statistics NZ in a trial of the GeoVista software. Figure 23 displays in star plots the proportions of the population by area in Auckland city: children (aged 0–14), working age (15–64), and older people (65 and over). It also shows the ratios of: youth dependency (total children divided by total working age population), old-age dependency (total elderly divided by total working-age population), and total dependency (sum of children and older people divided by total working-age population). In the scatterplot matrix for each of the two-way graphs, each point is linked to its geographic area on a corresponding map. The univariate cartogram indicates that it is the outer areas of the city (the suburbs) that have the highest proportion of children. An even larger number of graphs and maps than the four in figure 23 can be displayed simultaneously but, at some point, there will be a loss of simplicity and interpretability.

Visualisations can become too ‘busy’ to be effective communication tools. GeoVista could be most useful for specialist researchers or academics. However, they can also help policy analysts with some tertiary statistics training or a conceptual understanding of statistics, explore the multivariate characteristics of complex official statistics datasets (such as the Census) and visually link these characteristics to their underlying geographic patterns.

Figure 23

GeoVista display of population age groups in Auckland city



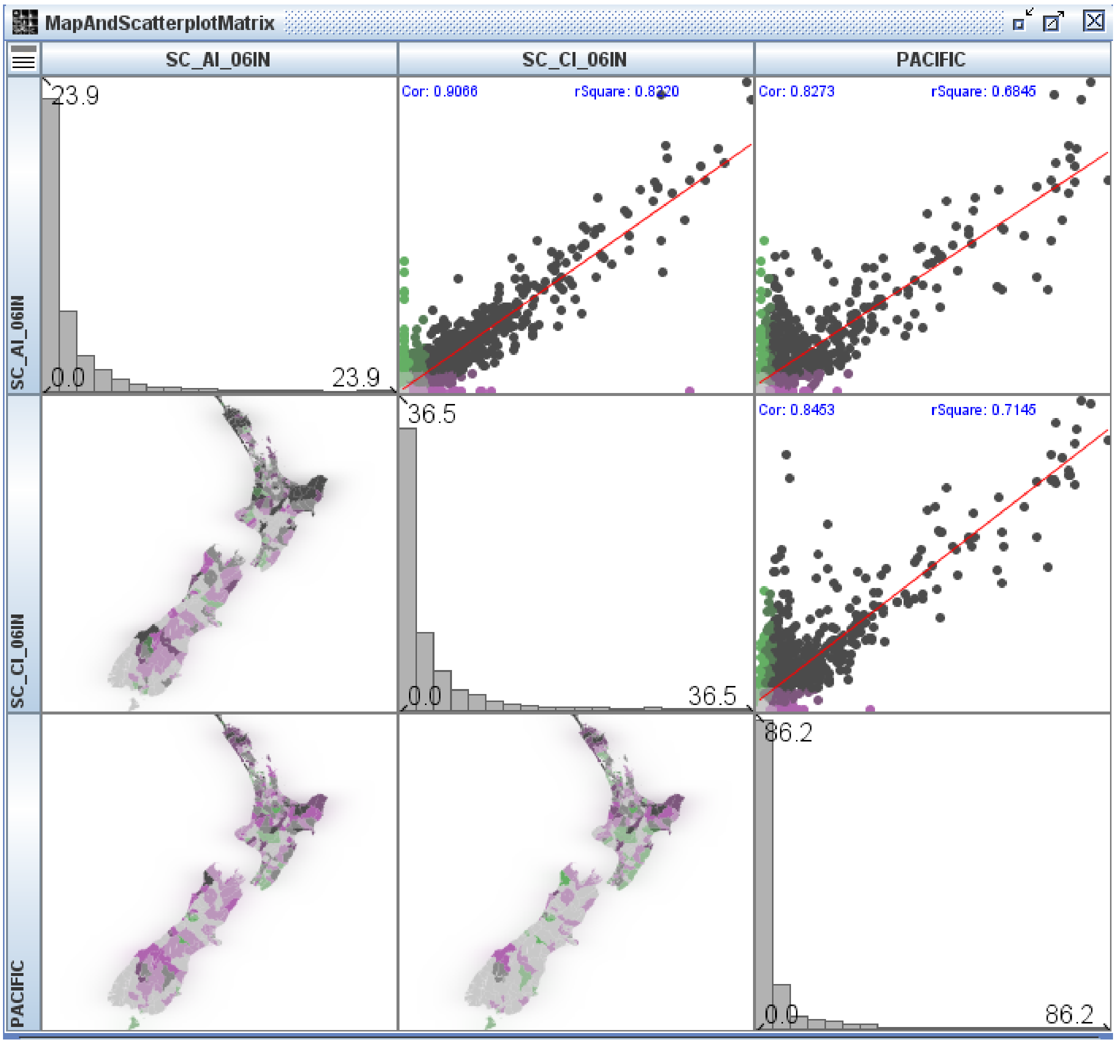
Many issues can be explored using this geo-visualisation tool. For example, de Róiste, et al (2009) used GeoVista with this dataset to investigate four case studies in health, crime, progress towards urban sustainability (looking at the covariation of dwelling occupancy rate and dwelling density), and labour market clearing (the degree to which people can get the employment they want and employers get the labour they need at a local level). The first two used administrative data (cancer rates and reported crime statistics) with statistics from Statistics NZ. In one study, the statistics software Stata and the GeoViz Toolkit were used. The authors stated that "the two packages are complementary… when used together however, it is possible to reap the benefits of both". They also noted that time dimension can be an important element in research.

Figure 24

Map and scatter plot matrix showing the relationship between area units

Proportion of severe crowding and proportion Pacific peoples ethnic group

2006 New Zealand Census



The exploration of crowding in New Zealand given in Goodyear et al (in draft) and Goodyear (2010) is another policy application. Crowding is used at a small area level to identify areas of high housing need as a variable within deprivation indexes. It is also used to explore the link between crowding and infectious disease. This study compared two indices, the American crowding index that counts people per room[2] and the Canadian national occupancy standard crowding, which is a more complex measure[3].

---

[2] This index measures the number of people living in a dwelling divided by the number of rooms. According to this index, dwellings with more than 1 person per room are classified as crowded, while those with more than 1.5 people per room are severely crowded.

[3] Under the CNOS measure, a household is said to be crowded if the dwelling requires extra bedrooms in order to meet the following criteria:

- There should be no more than two people per bedroom; parents or couples share a bedroom
- Children aged less than five years, either of same or opposite sex, may reasonably share a bedroom.
- Children aged less than 18 years of the same sex may reasonably share a bedroom
- A child aged five to 17 years should not share a bedroom with one aged under five of the opposite sex
- Single adults aged 18 years and over and any unpaired children require a separate bedroom.
- A one bedroom deficit is considered crowded and a two or more bedroom deficit is considered severely crowded.

GeoVista was used to examine the geographical distribution of crowding and the relationship between variables such as ethnicity and crowding at area unit level. The histograms in figure 24 show that crowding within New Zealand is very skewed, with most area units experiencing little or no crowding, and a very few area units experiencing high levels of crowding, regardless of which index is used. The distribution of severe crowding at an individual level was very similar whichever measure was used and was highly correlated with areas that have high proportions of Pacific peoples. Identifying geographic clusters is important when policy interventions are being designed.

New policy questions may arise as a result of explorations using tools such as GeoVista. It is likely that as users become more familiar with this type of software, there will be a demand for a greater range of analytical tools to be available on, or linked to it. The benefit of having a dynamic facility added so that there is automated viewing of changes over time is already apparent. Both longitudinal and integrated datasets (containing data often obtained from a number of administrative data sources) are becoming more common in official statistics. The potential use of new visualisations with these types of datasets has yet to be realised.

9 Costs, benefits, and other issues

There are costs associated with the development of interactive, dynamic, or integrated visualisation tools, but none with the software needed. Most visualisation tools are easy to access and are free to download from the Internet. For a national statistics office, it will cost to develop staff capability and capacity and, in the case of geo-visualisation, construct the underlying geo-coded datasets. These are generally one-off costs that need to be weighed against: the benefits of improved evidence-based policy advice to government; and against the cost of inaccurate policy advice caused by an inability to "see the whole picture" (Ugo Panizza, 2010, at the Inaugural IDSC Working Paper conference in Cairo, March 2010). One of the benefits of presenting complex datasets visually is that it increases not just the range of users of that data, but also of uses made of the data.

An area that has not yet been adequately researched is the level and type of skill that users need to be able to interpret these graphics correctly. There is a long history of evidence from statisticians that pictorial representations are better retained than numerical summaries (eg Tufte, 2001) but there is little international research evaluating which groups of users interpret or use these new representations correctly. De Róiste, et al (2009) reviewed literature within management science, decision science, and systems examining the benefits of graphs over tables and vice versa and found the results to be conflicting. Further research is needed to determine who understands what type of visualisation and how these visualisations are used in practice.

A number of authors have called for a cross-disciplinary approach to the creation of good visualisations. For example, Smith and Rogers (2008) note that there is a need for more research into how users perceive, interpret, and learn from these tools "some general aspects of the relationship between perception, particularly visual perception, and quantitative information have long been included in reference books on effective data presentation". Although many of the graphics presented here have been designed specifically for use with official statistics, it should be noted that they have much wider applicability. As Hal Varian (McKinsey & Company, 2009), chief economist for Google, stated, "The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that’s going to be a hugely important skill in the next decades ... I think statisticians are part of it, but it’s just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively".

Bill McLennan (2005), former Australian and British Government Statistician, said "to be useful official statistics need to be used". These new visualisations do enable us to access, see, interpret, and use official statistics in new ways, hence, increasing the value of those statistics. Although almost all national statistics offices now provide access to their data outputs on the Internet, they are at very different stages of development in terms of data visualisation. Already, there is sharing of visualisation expertise between national agencies. This sharing is helped to some extent by the fact that most agencies provide both national language and English versions of their websites. This collaboration should be encouraged further.

10 Conclusions

Official statistics clearly need to be used to be useful. Their potential is currently under-utilised. New interactive and dynamic visualisations provide a variety of ways for users to access official statistics already produced by national agencies. They increase the range of topics that can be addressed, thereby increasing their value. They may also increase users' understanding of the inter-relationships inherent in the statistics and enable them to make better decisions. New visual tools enable official statistics to come to life, acting as either a telescope or a microscope on our data. The user can either see the bigger picture or drill down to the finer details. However, there is still work to be done, in terms of realising their potential with longitudinal and integrated datasets or within statistics education.

There are some issues associated with the provision of these tools that need consideration. The criteria for graphical excellence need to be remembered and it is important to provide adequate metadata. User access to data through interrogative tools needs to be balanced with the confidentiality constraints that national statistics agencies work under. Statisticians need to be vigilant to detect and comment on misleading visualisations but the demand for these tools will only increase as they become more and more familiar to people in other aspects of their lives. Their use may lead to new insights, discovery of new trends in data or new exploratory and analytical methods. In addition, good visualisations may lead to greater usage of good methods of analysis.

As this paper shows, there are already a variety of visualisations that have been developed for and tested with multivariate official statistics. We are already seeing the globalization of official statistics in other areas, such as website release. Progress in data visualisation could be achieved through inter-agency collaboration, secondment of individuals with experience and expertise or by use of the expertise existing outside national statistics offices. However it is achieved the provision of access to official statistics through new visual tools will need to be explored by national statistics offices so that they stay in touch with the world of their users and do not become irrelevant.

References

Adobe Systems Incorporated (2010a). Adobe SVG Viewer 3.03 (www.adobe.com/svg/viewer/install/)

Adobe Systems Incorporated (2010b). Adobe Flash Player (http://get.adobe.com/flashplayer/)

Bascand, G (2009). Paper presented to '*The Role of Official Statisticians Roundtable'* at the Third OECD World Forum on '*Statistics, Knowledge and Policy'* October 2009. Busan, Korea.

Cleveland, WS (1993). *Visualizing Data*. Hobart Press.

de Róiste, M, Gahagen, M, Morrison, P, Ralphs, M, Bucknall, P (2009). *Geovisualisation and policy: exploring the links*. Official Statistics Research Project Report. Statistics New Zealand.

Forbes, S (2010). *Getting better value from official statistics by increasing and improving their use*. Working Paper No. 3 The Egyptian Cabinet – IDSC. Cairo. June 2010.

Friedman V (2008). Data Visualization and Infographics. *Graphics*, January 2008.

Gahegan, M, Takatsuka, M, Wheeler, M and Hardisty, F (2002). GeoVISTA Studio: A geocomputational workbench. *Computers, Environment and Urban Systems* 26: 267–92.

Goodyear, R (2010). *Using Geovis to determine which crowding index works best in New Zealand*. Internal report for Statistics New Zealand, Wellington, New Zealand.

Goodyear, R, Fabian, A, Hay, J (in draft). *“How many is a crowd?” Which crowding index works best for New Zealand?* Working paper for Statistics New Zealand, Wellington, New Zealand.

Hidalgo, CA (2010). *'Graphical Statistical Methods for the Representation of the Human Development Index and its Components'* United Nations Development Programme, Human Development Reports, Research Paper 2010/39. United Nations, New York.

Jackson, NO (2001). *The policy-maker’s guide to population aging: key concepts and issues*, Policy Research Paper No 13. Department of Family and Community Affairs. Commonwealth of Australia.

Marshall, G (1988). *A Dictionary of Sociology 1998.* Oxford University Press 1998.

McKinsey & Company (2009). Hal Varian on how the Web challenges managers. *McKinsey Quarterly*. Jan. 2009. Retrieved 25 October 2010. www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286

McLennan, W (2005). Verbal presentation to Statistics New Zealand staff (as past Australian Government Statistician).

Microsoft Corporation (2010). *Internet Explorer 9 Beta Guide for Developers*, Microsoft Developer Network. http://msdn.microsoft.com

Neumann, A (2005). *Social patterns of Vienna - an SVG example*, available at www.carto.net/papers/svg/samples/wien.shtml

Office for National Statistics (2008). *CommuterView: Understanding commuting patterns from the 2001 Census, User Guide.* Office for National Statistics, United Kingdom.

Panizza, U (2010). *The financial crisis and the role of finance in economic development: Lessons for developing and emerging market countries*. Keynote address. The Inauguration Conference Celebrating the launching of IDSC-WPS. Cairo. 2010.

Penn State University (2001). GeoVista Studio. (www.geovistastudio.psu.edu/) Copyright: Geography department, Penn State University 2001.

Post, FH, Nielson, GM, & Bonneau, G (2002). *Data Visualization: The State of the Art.* Proceedings of the 4th Dagstuhl Seminar on Scientific Visualization. Kluwer Academic Publishers, Boston/Dordrecht/London.

Ralphs, M, & Goodyear, R (2008). *The daily commute: An analysis of the geography of the labour market using 2006 Census data*, Labour Employment and Work Conference Proceedings, Victoria University of Wellington. New Zealand.

Rosling, H (2007). Gapminder. GapMinder Foundation. www.gapminder.org

Rosling, H (2010). *What showbiz has to do with it*. Plenary paper presented to the Eighth International Conference on Teaching Statistics (ICOTS8). Lbujljana. Slovenia. July 2010.

Small, H (1998). *Florence Nightingale's statistical diagrams*. Presentation to Research Conference organised by the Florence Nightingale Museum, St. Thomas's Hospital, 18 March 1998.

Smith, A (2008). Personal communication, Office for National Statistics, United Kingdom.

Smith, A & Rogers, S (2008). Web 2.0 and official statistics. The case for a multi-disciplinary approach. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics, 25*(3–4). IOS press.

ten Bosch, O, & de Jonge, E (2008). Visualising official statistics. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics, 25*,(3–4). IOS press.

Thygesen, L & Sundgren, B (2008). Innovative approaches to turning statistics into knowledge. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics, 25*(3–4). IOS press.

Trewin, D (2010). *Official Statistics in changing times*. Paper presented to Official Statistics Forum 2010: Realising the value of official statistics. Wellington New Zealand March 2010.

Tufte, E (2001). *The visual display of quantitative information*. Graphics Press, UK

Tufte, E (2006). *Beautiful evidence*. Graphics Press, UK

Tukey, JW (1977). *Exploratory Data Analysis*, Addison-Wesley.

New Zealand Government (1975). *Statistics Act 1975*, No 1, Public Act. Wellington.

United Nations (2003). *Handbook of Statistical Organisation*, Third Edition. The operation and organization of a statistical agency, Statistics Division, Studies in Methods, Series F, No 88, United Nations.

United Nations Statistical Commission (2010). http://unstats.un.org/unsd/statcom/commission.htm. Retrieved 25 October 2010.
Wild, C. and Pfannkuch, M. 1999. Statistical Thinking in Empirical Enquiry (with discussion). *International Statistical Review*, *67*(3): 223–265.

Wikipedia. The Free Encyclopedia (2010). Hans Rosling. http://en.wikipedia.org/wiki/Hans_Rosling Retrieved 25 October 2010.

World Wide Web Consortium. 2010. *Scalable Vector Graphics (SVG) 1.1 (Second Edition) W3C Working Draft* 22 June 2010. Available from www.w3.org/TR/SVG/. Retrieved 25 October 2010.