Journal
Technology Innovations in Statistics Education, 6(1)

Author
Forbes, Sharleen Denise

Publication Date
2012

DOI
10.5070/T561012851

1. THE PROBLEM AND ITS IMPORTANCE

In New Zealand, as in many other countries, the government's policy advisers are a powerful influence in the national decision making process. While some of these advisers are highly skilled at using quantitative information, in recent years there has been growing concern at the variable, or lack of, statistical capability in many state sector agencies. Evidence for this lack comes from a pilot study (Macky & Saffron, 2004), national statistics office consultation with statisticians and policy managers in 12 state sector agencies in 2008, subsequent endorsement by state sector chief executives and consultation undertaken by Victoria University's School of Government (Forbes, 2011a). Both the School of Government and the national statistics office, Statistics New Zealand, have responded to these needs by providing a number of training opportunities for government employees (Forbes, 2011b). Most of the learners enrolling in these courses work in agencies where they have their own computer, ready access to the internet and routinely use Google as their first means of accessing research information. Many government advisors have post-graduate degrees in Public Policy or Public Administration but they also have highly variable mathematics backgrounds. Some of the students in the first of the two courses discussed in detail in this paper, the Masters in Public Policy, had little exposure to statistics in their formal education. This variable mathematics background is one of a number of challenges for teachers of statistics. Another is how to motivate these adult learners and demonstrate the importance of statistics to them. Motivation is a key factor in the learning of statistics. As long ago as 1801, Playfair stated '*For no study is less alluring or more dry and tedious than statistics, unless the mind and imagination are set to work, or that the person studying is particularly interested in the subject; which last can seldom be the case with young men in any rank of life'*. (p16). This issue appears not to have abated over time with Peebles stating in 2011 that, *'...statistics is considered by many students to the most challenging, anxiety producing and therefore least popular of their subjects.'* (p1).

The New Zealand government recruits statisticians, many as new graduates, who have had little exposure to the large national data sets (official statistics) routinely used by government. As Tucker (2010) stated, '*government agencies are interested in statisticians who know how to approach real-world problems analytically'* (p459). How can we train government statisticians to be statistically literate in the sense defined by Gal (2002)? That is, as having '*the ability to interpret, critically evaluate, and communicate about statistical information and messages'*. One solution is to provide students with specific training in the final (honours) year of their degree. With this in mind, an honours course in official statistics was trialled with newly available technology – an access grid that provided high-quality audio and real-time video, allowing groups at multiple sites to interact simultaneously, sharing both data and information, and that facilitated simultaneous teaching across New Zealand universities in 2011. The course was a joint winner of the Best Cooperative Project in Statistical Literacy award from the International Statistical Literacy Project in 2011 and is the second course discussed in the paper.

Data visualisation tools have three major advantages: they are free, easy to access (available either on-line or as open-source software) and present complex information in an easy to interpret visual format. It is important for both those creating and using official statistics that they can make sense of new visualizations, typically of very large data sets, of data. As Peebles (2011) states, much research has been done over the last thirty years looking at the perceptual and cognitive processes involved in interpreting graphs, and also at the representational properties of different forms of graphs. He concludes that people are better at comparing and evaluating quantities using bar graphs, identifying trends using line graphs and interpreting interaction better if it is represented by bars. It is also well known that some pairs of colours are more distinguishable than others (e.g. Cui, 2007). This paper discusses the use of a number of recent data visualisations, with particular application to the large data sets that arise in official statistics, in statistics teaching with students in 2011 in both a Masters of Public Policy course run by Victoria University of Wellington and the above new applied statistics honours paper in official statistics. The first criteria for selection of a visualisation tool for classroom use was that it could be applied to official statistics and the second that it visually demonstrated a particular statistical concept. For example, a dynamic (moving) tool was selected to show momentum, a tool that showed relative size as area was chosen to investigate weights and a mapping tool to display geographic data. The visualisations were used to help develop students' conceptual understanding of statistics while minimising the mathematical content being taught. Three of the visualisations were also presented in a basic descriptive statistics course in 2012. In all the courses, statistics were put into the real world of the student by using classroom examples of official statistics that are routinely reported in the media.

2. BACKGROUND AND PREVIOUS WORK

There is already a large and still growing body of research around the use of real data in statistics education (from Dunkels, 1994 through to recent papers at the 2011 International Association of Statistics Education conference such as Payne, 2011; Harraway, 2011; Crawford & Marriott, 2011 and Easterling, 2011). There is a similar body of literature on the development and use of visual tools for communicating and analysing statistics from Tufte (1983) onwards and more recent work on the availability of data visualization software that provide new methods for access to and interpretation of official statistics (SJIAOS, Vol.25, No.3-4., 2008; Forbes et al, 2011). Data visualisation technology is changing so rapidly that it is likely that only professional producers and users of statistics remain completely up-to-date with new tools for displaying data. New static and interactive graphics are now being routinely used by agencies such as the United Nations (UN) and the Organisation for European Co-operation and Development (OECD) as shown in Figures 6 and 7 below. Many new dynamic and interactive visualisations are widely available on the internet (e.g. Hans Rosling's (2007) Gapminder 'bubble' graphs that use OECD data www.gapminder.org) and geo-visualisations such as the real time display of the recent Christchurch earthquakes created by Paul Nicholls of Canterbury University (http://www.christchurchquakemap.co.nz/) are in common use by the media.

As the availability and spread of data visualisation increases so does the likelihood of it being part of the real world of the student. It may, however, be less likely to also be part of the real world of their teachers. Chick & Pierce (2010) noted that teachers do not always see the teaching opportunities offered by real-world data and those presented by new forms of data visualisation may also not always be obvious, but, as Wild et al (2011) stated '*For statistics education, technology is the ultimate game changer. Its biggest pedagogical implications come from the fact that it allows us to conceptualize, in ways that were previously unavailable, potentially providing access to concepts at much earlier stages of development'*. (p248). They suggested that technology could make the landscape of statistics education unrecognizable. Simulation is one of the tools technology is making available in the classroom, data visualisation is another.

Simulation packages (such as Ridgeway et al, 2007 and Stirling, 2011) already exist to assist with understanding variation and estimation. This paper does not discuss the educational uses of these types of visualisation but instead focuses on the use of some of the recent visualizations (usually of aggregates) of very large data sets that can be used to explore the patterns and relationships inherent in the data. Much of the discussion in the literature of these new tools has been on their use to summarise, display and disseminate data. But these visualisations also provide useful real world examples for teaching, giving *'a motivating context for learning and can demonstrate the utility of statistical concepts'* (Chick & Pierce, 2010). They provide students with '*practical, investigational, real, purposeful, appropriate, cross-curricula and enjoyable'* (Davis, cited in Joliffe, 1994, p171) data handing opportunities. Geo-visualisations in particular, help demonstrate the inter-disciplinary nature of official statistics and were used to both capture the interest of students and show the links between statistics and other disciplines and between statistics and the real world. The use of data visualisation as a teaching tool has the potential to change both how and what is taught in statistics. However, it is acknowledged that the teacher plays an important role to bring meaning to these pictorial representations (as Wild et al, 2009 state) and one aim of this paper is help teachers see some of the learning opportunities that these tools offer.

3. COURSE DESCRIPTIONS

The major difference between students in the Masters of Public Policy course and those in the applied statistics honours paper in official statistics is in their previous level of mathematics and statistics. Almost all of the students in the Master of Public Policy course are either government policy analysts or students intending to be policy analysts. In order to provide evidence-based advice to government they need to have basic competency in statistics generally and, in particular, be able to understand and interpret the official data of government (Tucker, 2010) but both the students' age and their mathematics background vary (from secondary school level to early undergraduate mathematics or statistics). The emphasis in the Masters course is on understanding and interpreting statistical concepts rather than the mathematics underlying them. As Libman (2010) suggested '*innovative and effective instructional strategies will continue to be a challenging endeavour, especially for "occasional users" taking statistics'*. The students in the honours course in official statistics were all majors in either statistics or applied statistics (doing honours, post-graduate diploma or Masters degrees). In both courses in 2011 I used data visualisations for three purposes:

- to motivate students in their learning of statistics by providing them with real data to play with (in the sense described by Wild & Pfannkuch, 1999),
- as classroom examples that students were also likely to encounter in their everyday working life and
- as a way of demonstrating the mathematics in action of statistics concepts.

Official statistics provide an important information source for government and were the theme of the honours paper, so this paper focused on those data visualisation tools in common use with official data sets. The interactive and dynamic features of recent official data visualisations were used in the classroom to explore patterns and relationships in the multivariate data sets and to demonstrate statistical concepts. Some visualisations not only present and allow easy manipulation of massive official statistics data sets but also enable new types of information to be obtained.

Although the visualisation tools are easy for students to use, in the Masters course, time constraints mean that they were only used by the teacher to demonstrate possible data manipulations and analysis. However, in this course, examples of some of the policy uses these tools have already been applied to in New Zealand were also given. In the honours paper in official statistics, data visualisations were used in many of the modules (see Appendix) including the overview of official statistics, demography, geographic information systems and price indices. In 2011, one of the assignments (that counted towards the final mark) contained a question that required students to download both a geo-visualisation tool and a given data set and use them to explore geographic distributions of variables and relationships between variables in the data set.

In the short course on basic descriptive statistics in 2012 the three-dimensional scatterplot (Figure 3c) was presented as a graphical extension of two-dimensional scatterplots. Dynamic population pyramids and the Price Kaleidoscope were used to reinforce their related concepts (population momentum and weighting) as described below for the other two courses.

Official statistics data sets are by nature large, sometimes even when presented in aggregate form as in the the Commuterview visualisation provided by Statistics New Zealand (Figure 1) that displays linked information from two 2006 New Zealand Census of Population and Dwellings questions: ‘Where do you live?’ and ‘Where do you work?’. At the level of geography displayed in Figure 1 an equivalent national table would contain more than 3 million cells so this information can only realistically be displayed and explored graphically. At the finest level of geography an equivalent table would contain approximately 2 billion cells. This visualization is an example of an interactive map allowing the viewer to highlight or ‘zoom in’ on part of the graphic to obtain greater detail. Although students could use this visualisation themselves to formulate and investigate their own policy questions it was mainly introduced so students could see the potential of data visualisation as an analytical tool for policy development. The Master of Public Policy students in particular saw immediate applicability of the tool for urban transport design, disaster risk management, etc. For example, the map shown in Figure 1 is of the commuting patterns of health workers in the central business district (CBD) of New Zealand’s capital city, Wellington. The CBD is surrounded by hills and was used to discuss the city’s vulnerability to earthquakes. The linking of statistics with their underlying geography is not new but the wide availability of tools like Google Maps (http://maps.google.com/) and Google Earth (http://www.google.com/earth/index.html) have brought spatial data visualization (geo-visualisation) technologies into the public domain. One could argue that the basic use of geo-visualisation tools should, therefore, be taught in any course that aims to give students the fundamental statistical skills that they will need for their future careers and life decisions.

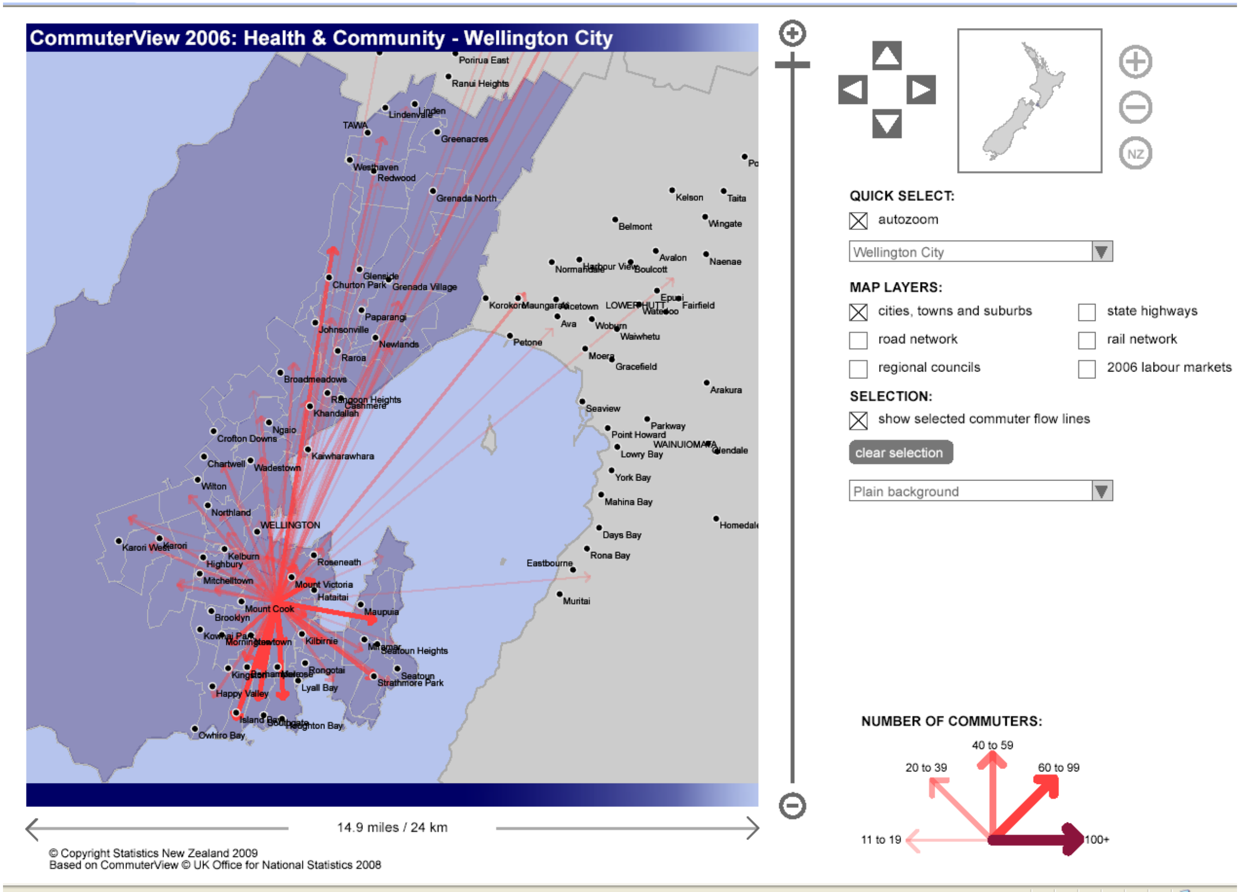

Figure 1: New Zealand Commuter Flows: Work to Home, Wellington City

4. USE OF GEO-VISUALISATION IN THE CLASSROOM

The Commuterview tool above only displays geographic count data so students were limited to description and comparison of data flows, but more sophisticated tools such as geo-visualisation software that integrates maps, graphs and analysis also exist and can be used to introduce simple statistical analyses (such as correlation and regression) together with geographic analysis.

For example, GeoVista (open-source software, GeoVISTA Center, (2012),) was used in both courses to explore the geographic distribution of Census variables. Figure 2 displays the proportions of the population in Auckland city that are, according to the 2006 New Zealand Census:

- children aged 0-14 years (PC001491),
- working age (15-64 years, PC156491)
- elderly (65 years and over, PC65PL91) together with

- the youth dependency (total children divided by total working age population = YDEP91),
- old age dependency (total elderly divided by total working age population = ODEP91) and
- total dependency (sum of children and elderly divided by total working age population = TDEP91) ratios,

.

The software provided an opportunity for the teacher to demonstrate and discuss the use of simple graphics in common usage (such as the histograms and scatterplots shown in the Map and Scatterplot matrix in Figure 2) and also to introduce new graphics such as the cartogram (lower left hand corner of Figure 2) that shows the higher density of children in the outlying suburbs of Auckland. In each case, the data points on the graph are linked to their geographic location as is demonstrated in the linked star plot and map at the top of Figure 2. This linking allowed students to easily identify and explore the nature of outliers, such as the island in the Hauraki Gulf that is highlighted in the top right of Figure 2.

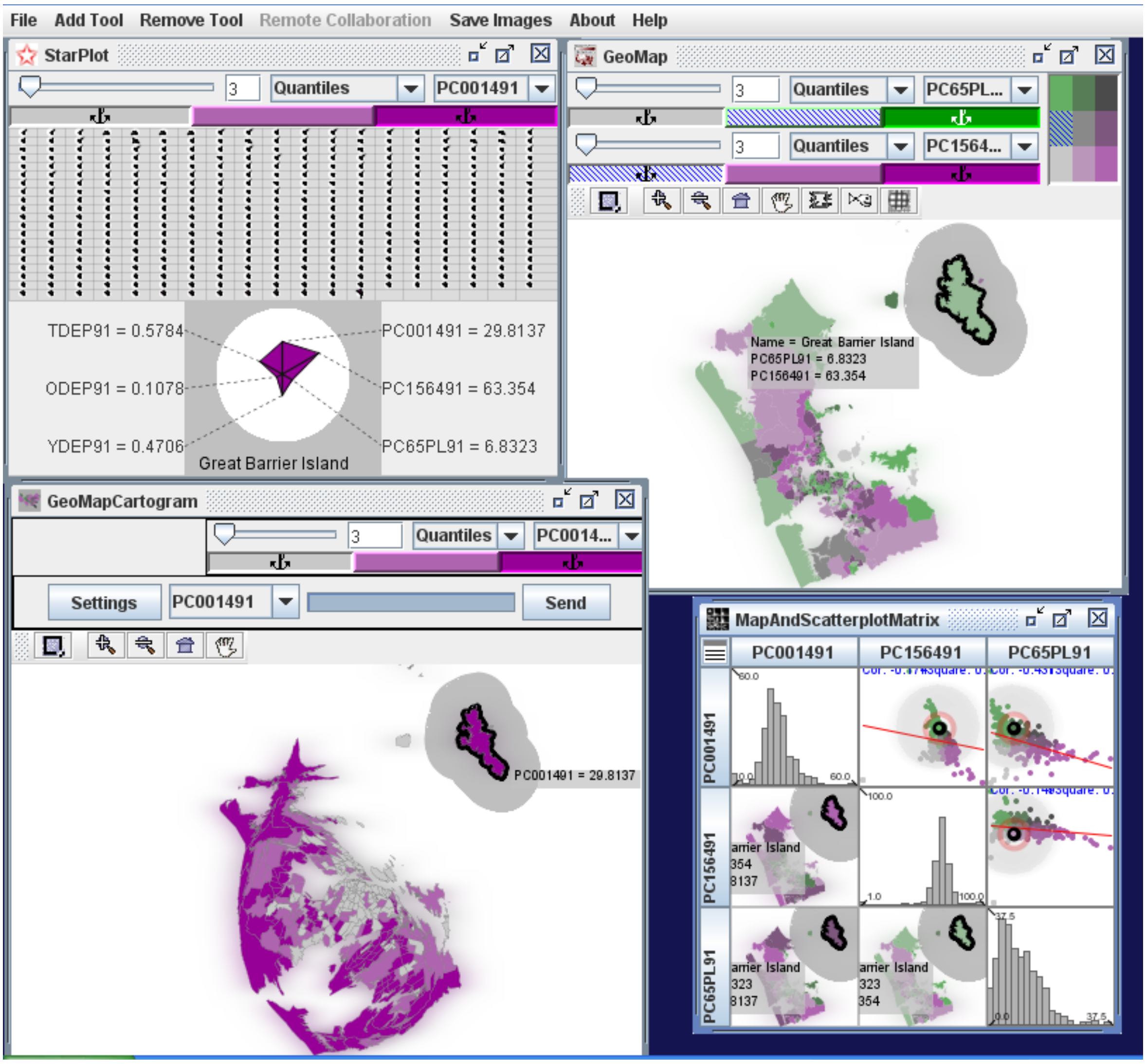

Figure 2: GeoVista display of population age groups in Auckland city (1991 Census data)

The GeoVista software package provided students with a range of simple tools that could be used to '*look for patterns, raise questions, build models, strategies and approaches*' (Libman, 2010, p3). In the Masters course, some of the policy applications that the tool has already been applied to were also presented, including four case studies on health (alternative presentations of cancer data), crime (geographical correlation between the fear and the occurrence of crime), the relationship between dwelling density and occupancy rates, and an analysis of the labour market (de Roiste et al, 2009) together with another paper on the distribution of household crowding (Goodyear, 2010). These gave rise to student-lead discussion on other potential policy questions that could be investigated using this type of software.

In the honours course in 2011, students each downloaded the free open-source GeoVista software together with a data set containing confidentialised Auckland city data derived from the 2006 New Zealand Census of population and dwellings and used this to answer their assignment question. All the 29 students participating in the honours course completed the data visualisation assignment and gained over half marks, and several attained the highest possible score of 10 (median = 7, mean = 7.2, standard deviation = 1.5). Only five of the 29 students (17%) used fewer features of the tool than they had been shown in class, 12 (41.5%) used exactly what they'd been shown and 12 (41.5%) undertook further exploration of the tool using other features that had not been demonstrated. Students were asked to describe the geographic distributions and relationship between at least two of the Census variables. Of the 25 students who attempted to describe the univariate geographic distributions, seventeen (roughly two thirds) adequately described the geographic trends with three of these also discussing geographic clusters. Another four described geographic clusters only. Four (16%) of the students appeared confused about what a geographic distribution was (although this had been discussed in class). These students focussed on single points of interest or a seemingly random selection of areas rather than examining north-south or east-west trends, inner city clustering, etc. Only the very best (five – the top 20%) of the students described both the geographic distribution together with our usual understanding of statistical distribution using histograms to describe central tendencies and shapes of distributions. In the bivariate case however, 19 (65.5%) students discussed either the joint geographic trend or clustering and 12 (41.5%) combined this with calculation of the standard Pearson's correlation coefficient.

One interesting point that arose was that several students were confused between variables and categories of variables, and attempted to analyse the correlation between two (dependent) categories of a single variable, such as the proportion of children and the proportion of working age population (different proportions of the same total). Perhaps this distinction is something that we should give greater emphasis to in our teaching.

5. USE OF DATA VISUALISATIONS TO TEACH STATISTICAL CONCEPTS

As well as presenting data well in their advice to Government, policy users also need to interpret research information together with nationally important social and economic measures. This requires a conceptual understanding of 'traditional' statistical concepts and methods such as confidence intervals, *p*-values and regression together with an understanding of concepts specific to official statistics such as those underlying population dynamics (demography) and the macro-economy. As described earlier, simulation packages are often used to reinforce 'traditional' statistics concepts and data visualisation tools can be used in a similar way in the classroom to develop understanding of official statistics concepts as highlighted by the following examples.

5.1 The Interaction Concept in Regression

There is a large step from simple bivariate regression to multiple regression models that may or may not incorporate interactive terms (where one variable acts as a modifier on another). Three-dimensional scatterplots were used to assist student to understand this concept as shown in Figures 3a, 3b and 3c. All these graphs were created using a small synthetic unit record file created from the 2004 New Zealand Income Survey (to maintain confidentiality) but could similarly have been created using the original data set. Figure 3a is the ordinary scatterplot and simple regression line created in EXCEL. Figures 3b and 3c are pin-graphs where the heads of the pins form a 3-D scatterplot created using the software package 'R' (R core team, 2012). Students were asked to discuss the degree of interaction present by looking at the consistency of pattern (for example, across highest education classes in Figure 3b). Extension from three- to four-dimensions was achieved by the use of colour as shown in Figure 3c. This graph was used to discuss whether separate models should be fitted when the pattern is different for different groups (as in this case, with females dominating part-time work and males full-time work). Further increases in dimensionality could be achieved by changing the size, shape or colour intensity of data points or by adding a dynamic feature (such as time) across overlaid static graphs as used by Hans Rosling (2007) in his Gapminder 'bubble' graphs.

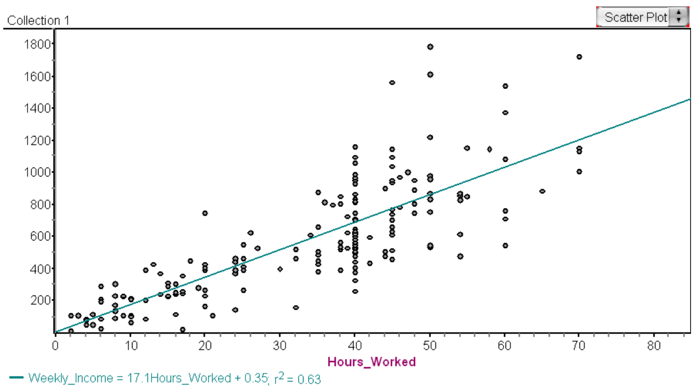

Figure 3a: Simple regression: Weekly income by Hours Worked

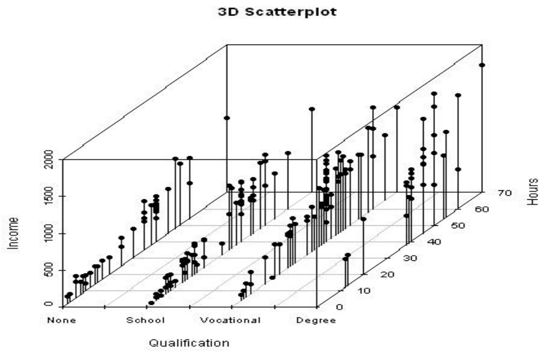

Figure 3b: Weekly Income by Highest Educational Qualification and Hours Worked

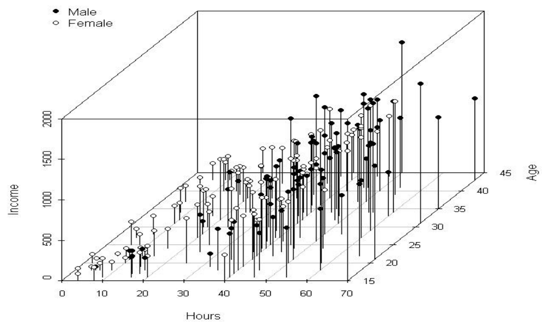

Figure 3c: Weekly Income by Sex, Age and Hours Worked

5.2. The Momentum Concept in Demography

Dynamic population pyramids (Figure 4) can be used by both teachers and students to explore the *momentum* concept in demography (population growth resulting from a youthful age structure or population decline resulting from an older age structure). Students traditionally derive changes in population age structures by following a cohort through time adjusting for known or assumed fertility, mortality and migration rates using a spreadsheet or the like. This was done as part of the demography assignment for students in the honours paper in official statistics. However, it is my experience that students often get focussed on the calculation task in this exercise, losing sight of the *momentum* inherent in the data.

Dynamic population pyramids created from cross-sectional static data (using Census data and official population projections), overlaid and played over time using new visualisation technology were used to demonstrate this *momentum* concept in action, allowing learners to see population growth resulting from high birth rates, leading up to and beyond the point (different for each country) where natural increase (growth) shifts to become natural decrease (decline) as the aging population dies off.

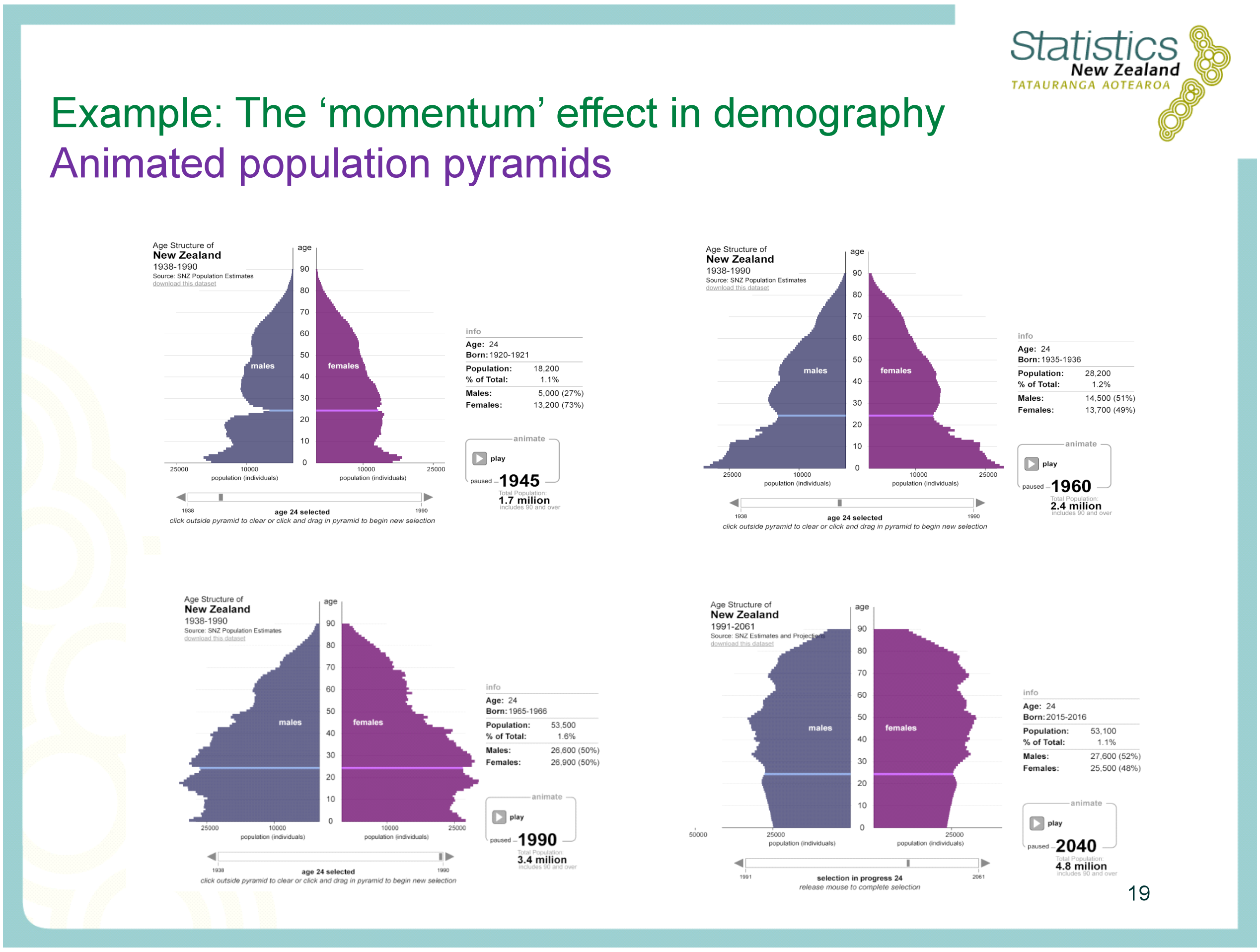

Figure 4: Dynamic population pyramids showing changes in the New Zealand age structure over time.

Students in all the courses were very engaged in this exercise, asking for various ages to be tracked over time (as in the highlighted 24 year olds in Figure 4 above) and commenting on the size of the post-war "baby-boomers" and subsequent bulges. This visualisation seemed to help the Masters students in particular to understand, for example, that almost all the people that will give birth in the next twenty years are already born, and that we can use this information to project future population size and structure.

5.3 The concept of weighting in price and composite indices

When teaching index numbers, weighting seems to be a difficult concept for students to grasp. This is true for even a simple index, such as the Laspeyres formula for the Consumers Price Index. The Laspeyres formula calculates the index for period *t* on base period *o* by

$$Index = \frac{\sum P_t Q_o}{\sum P_o Q_o} \times 1000$$

where Q is the quantity (weight), P the price of the item and the summation is over the groups (and subgroups) of commodities in the Consumer Price Index.(Statistics New Zealand, 1999).

5.3.1 Price indices

The Consumer's Price Index (CPI) measures the price change of representative goods and services (the basket of goods) purchased by households and, in New Zealand, is used to set monetary policy, to adjust benefit rates and to negotiate wages. It is frequently reported by the media. The term is therefore familiar to students but often they have no real understanding of how it is derived. The interactive Price Kaleidoscope shown in Figure 5 produced by the Federal Statistical Office of Germany (http://www.destatis.de/Voronoi/PriceKaleidoscope.svg) is a voronoi treemap. According to Balzer et al (2005), these are heirachy-based visualizations that use iterative relaxation of voronoi tessellations to form arbitrary polygons that subdivide an area without producing any holes or overlaps. Voronoi treemaps were developed from rectangular-based treemaps to overcome the recognized perceptual flaws arising from the creation of numbers of long, thin rectangles that among other things make it hard to tell if neighbouring objects are close in the hierarchy of groups and subgroups.

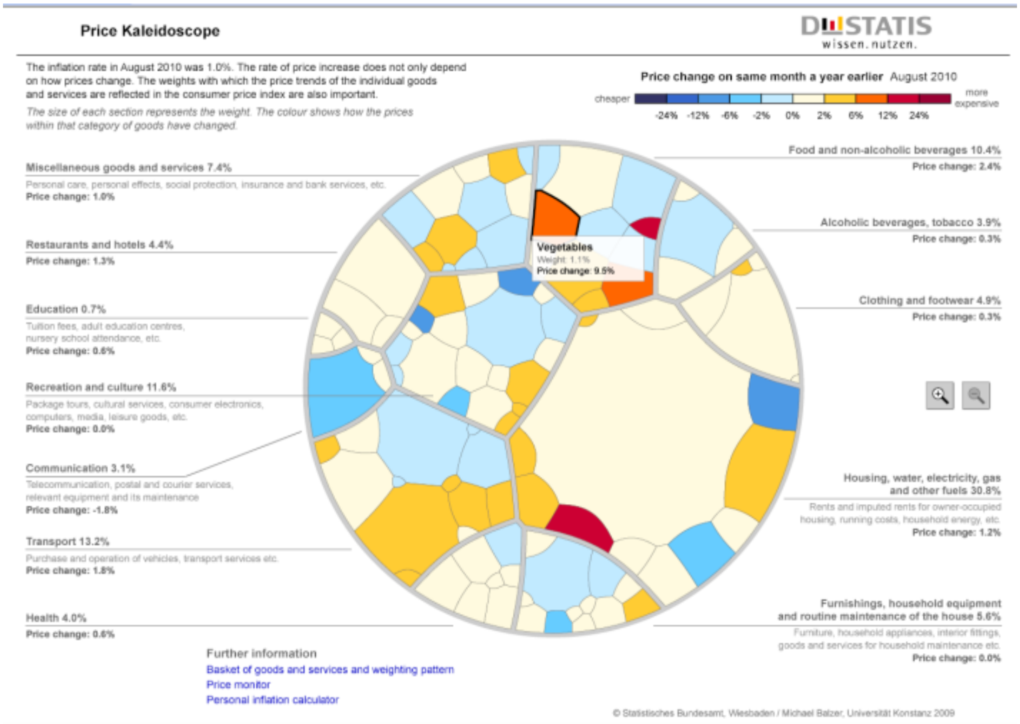

Figure 5: Federal Statistics Office of Germany Price Kaleidoscope

The Price Kaleidoscope demonstrates how the (expenditure) weight assigned to each group or subgroup each good or service impacts on the overall value of the index. In the visualisation, the CPI is represented by a circle and within this each group (and subgroup) has an area proportional to its weight. Clicking on an area in the circle displays its weight and quarterly change in prices, allowing students to explore these. By moving back-and-forth between the region (group) and its weight as represented by its area, and the overall CPI students develop an understanding of the impact of the weight, observing that while some subgroups have large positive or negative price changes the effect of each price change on the total CPI is determined by its associated weight. For example, in Figure 5, Vegetables had a price change of +9.5% but with a weight of only 1.1% made only a 1.0% impact on the overall CPI .

The Price Kaleidoscope helped students to interpret and understand the formula given above and its use led naturally to a student initiated discussion of how the weights are derived (from the Household Expenditure Survey), providing the teacher with the opportunity to introduce and discuss potential sources of error in the measurement of both the weights and prices. One benefit of using this tool to explore the *weighting* concept was that students quickly became actively engaged with the tool and enjoyed using it.

In the honours paper, students were asked to collect their own expenditure data over a two month period, then were given an assignment in the first half of which they had to derive their own weights, compare these to the official CPI weights and calculate the Laspeyres index using their weights. In the second half of the question they were asked to discuss the uses of the CPI in New Zealand. Again all the students attempted the question and all correctly calculated their personal weights. Twenty-eight of the 29 gave sensible reasons why their weights differed from the official weights and 23 used their weights to correctly calculate a Laspeyres index. The results out of a possible score of 20 marks (for both sections) were median = 16, mean = 15.3 and standard deviation = 2.03. While there is no directly comparable evidence with traditional teaching of the Laspeyres index, from the teacher's perspective this was a positive result.

5.3.2. Composite indices

Once the concept of *weights* was established using the Consumers Price Index, other forms of indices that also contained weights were introduced, such as the composite indices used by international agencies such as the UN and OCED, together with new visualisations of these indices. Examples were given that the public policy students in particular were likely to come across. These included Hidaglo's (2010) tree representation of the UN Human Development Index (HDI) that is used to examine the relative importance of the components over time within a country and also to compare across countries (Figure 6). The HDI tool has an internet link that could be used by students to create individual graphs (http://chidalgo.com/Gallery/HDR2010/DevelopmentTreeApplications.zip).

A similar, but interactive representation, the Better Life Index, launched by the OECD in 2011 (OECD, 2011) and shown in Figure 7 was also given. With the composite indices the class discussion centred not only on the construction of the index but also on the clarity of the its visual representation with students having a clear preference for the simplicity of the flower representation below.

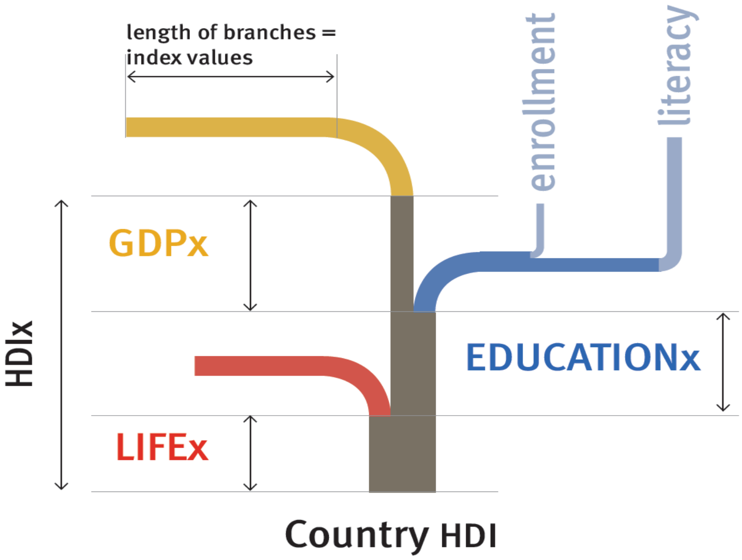

Figure 6: Hidaglo's Tubular tree representation of the HDI

Figure 7. 'Flower' representation of the OCED Better Life Index

6. STUDENT FEEDBACK

Feedback was formally requested from the students in the Honours paper using a standard university questionnaire with some additional opened ended questions. Only 10 of the 29 students completed the questionnaire. Responses to the question '*list two or three specific things about this course that most stimulated/helped you to learn*' mainly concerned aspects of the course such as the format of the assignments and the on-line availability of the (video-recorded lectures). There were five topic related responses of which three were '*data visualisation', 'real-life examples'* and '*being able to see the application of official statistics'.* There were seven separate topics mentioned in response to the question '*Which topics, if any, would you like to have spent more time on in class?*'. Although data visualisation was not listed as a topic in the course outline, one response was '*data* visualisation' and another '*the GIS topic*' (which focused on geo-visualisation). Formal feedback was not obtained from the Master of Public Policy students but a number commented verbally how much they had enjoyed the session and one stated in an email that the "*visual tools helped to further my understanding of different methods for presenting complex data….I have found in my work that how the information is presented to decision makers is sometimes as important as how the data was collected or analysed."* An indirect form of feedback from the honours class was that over 40% of these students did more exploration of the geo-visualisation tool (discussed in section 4 above) than was required for their assessment and roughly two-thirds could describe geographic trends. For most of them the concept of a geographic distribution associated with statistical data was new but their results suggest that the tool allowed them to quickly develop this concept.

After each visualisation was demonstrated in the basic statistics course the 17 students were asked verbally whether they liked the visualisation and whether they found it helpful to understand the concept being discussed. There was unanimous agreement to both questions for both the dynamic population pyramids and the price kaleidoscope, but only two of the seventeen agreed (again to both questions) for the three-dimensional scatterplot. This is perhaps not surprising as this graph was introduced as a extension of two-dimensional scatterplots and not to reinforce the interaction concept as in the other two courses. Interestingly, the entire class also responded that they enjoyed and could understand the multi-dimensional plots created by Hans Rosling in his Gapminder software (Rosling, 2007). It does appear that our students have clear preferences for some types of visualisations over others.

In summary, feedback from these courses suggested that the use of data visualisation in the classroom helped make the students' statistics learning easier and more enjoyable (that is, met Davis' *practical, investigational, real, purposeful, appropriate, cross-curricula and enjoyable* criteria).There has, however, not yet been formal empirical research into whether or not the use of data visualisation tools has a greater impact on students learning and retention of statistical concepts than more traditional tools This is an area for further research.

7. IMPLICATIONS FOR CURRICULAR DEVELOPMENT, TEACHING OR LEARNING

Although the visualisations above were useful teaching tools for demonstrating the patterns and relationships between variables, there is an important caveat when using them either for presentation or for teaching. That is, that they can mask both variation and data quality. If, as is the case with many of these visualisations (including the popular Hans Rosling Gapminder graphic (Rosling, 2007)), the visualisation is of data aggregates (e.g. counts, proportions or point estimates such as sample means) then the underlying variablity in the data is not apparent to the user and needs careful explanation to the statistics learner. This may not be a problem for Census and administrative data sets but is more of a concern with visualisations of survey data. A second issue is that if the visualisation looks slick and professional then the data contained within it may be simply accepted at face value. It needs to be reinforced with both teachers and learners that data visualisations are merely tools. As with any other graphic or statistic, its value depends on the quality of the data within it.

Increased internet usage is likely to result in these types of data visualisation becoming more and more commonplace in everyday life. The burden will be on teachers of statistics to ensure that students have the basic skills and conceptual understanding to interpret them and use them in their decision making.

8. SUMMARY

The data visualisations used in the courses discussed:

1. Provided student motivation through the use of:
   a. real world examples and classroom data
   b. simple tools to explore patterns and relationships in large data sets
   which, together, enabled students to ask and explore new types of questions.
2. Allowed discussion of new, as well as familiar, graphical representations of data.
3. Reinforced official statistics concepts such as:
   a. *interaction* between variables in regression
   b. population *momentum* in demography
   c. *weighting* in price and composite indices.

In addition, they were hands-on and fun tools for students to use. This student-accessibility is one reason, but not the major reason, to make more use of data visualisation in our classrooms. These visualisations also provide teachers with tools that show, rather than simply describe in mathematics, statistics concepts to learners. They help teachers make abstract ideas concrete. Another major advantage of the tools is that they demonstrate real-life applications of statistics.

In the case of official statistics, the underlying data sets may be huge but recent visualisations are simple enough for students to use and allow them to explore data in new ways. In particular, while the growth of geo-visualisation in popular media helps reinforce the multivariate and interdisciplinary nature of official statistics, it also means that we can no longer ignore the inherent dimension of geography in either our teaching or analysis of statistics. New forms of data visualisation are already impacting how we teach statistics but it is likely that they will also alter what we teach in the future. There are many issues in having data associated with place and distance that make learning more complicated but providing students with experience of time and space linked data through visualisation tools should make these aspects easier to understand. The potential educational use of data visualization within statistics has yet to be realized, as does its importance, but it can no longer be ignored.

Note: YouTube videos of how to use the Dynamic Population Pyramids, the Price Kaleidoscope, Commuterview and GeoVista are available on the Statistics New Zealand website, http://www.stats.govt.nz/surveys_and_methods/methods/research-papers/working-papers/visualising-official-statistics.aspx. However, the videos do not demonstrate the classroom use of these visualisations. Separate urls for each of the above data visualisation are given below:

- The Price Kaleidoscope http://www.youtube.com/watch?v=s4y8rE3Bfts
- Population pyramids http://www.youtube.com/watch?v=PHwrH4C6JME&feature=plcp.
- Commuterview http://youtu.be/KxHJB0Y31zk
- GeoVista http://youtu.be/MoO23H5rz-I

REFERENCES

Crawford, E. and Marriott J. (2011). Teaching Statistics through problem solving: using real time data retrieval. In P. Bidgood (ED.), *Proceedings of the 2011 IASE Satellite Conference: Statistics Education and Outreach.* (In press).

Chick, H. and Pierce, R. (2010). Helping teachers to make effective use of real-world examples in statistics. In Reading, C. (Ed). *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8).* International Statistics Institute, Vorrburg, The Netherlands.

Cui, W. (2007). A survey on graph visualization, Phd qualifying exam (pqe) report, Computer Science Department, Hong Kong University of Science and Technology.

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.

Dunkels, A. (1994). Interweaving numbers, shapes, statistics and the real world in primary school and primary teacher education. In Robitaile, D.F., Wheeler, D.H. and Keirnan, C. *Selected lectures from the Seventh International Congress on Mathematics Education.*Les Presses de l'Universite Laval, Quebec.

Easterling, R. (2011). Passion- Driven Statistics. In P. Bidgood (ED.), *Proceedings of the 2011 IASE Satellite Conference: Statistics Education and Outreach.* (In press).

Forbes, S. (2011a). *Official Statistics – training and education needs* Report to the School of Government Trust, Victoria University of Wellington.

Forbes, S. (2011b). Collaboration and co-operation: the key to reaching out. In P. Bidgood (ED.), *Proceedings of the 2011 IASE Satellite Conference: Statistics Education and Outreach.* (In press).

Forbes, S., Ralphs, M., Goodyear, R., Pihama,N. (2011). *New ways of visualising official statistics.* Working paper series. No 20. Information and Decision Support Centre. Cairo, Egypt.

Gal I. (2002). "Adult's statistical literacy. Meanings, components, responsibilities." *InternationalStatistical Review*, 70(1), 1-25.

GeoVISTA Center, (2012). Department of Geography, The Pennsylvania State University, University Park, PA, http://www.geovista.psu.edu/grants/cdcesda

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

Harraway, J. (2011). Use of case studies and new software to motivate statistics teaching and learning at school and undergraduate level. In P. Bidgood (ED.), *Proceedings of the 2011 IASE Satellite Conference: Statistics Education and Outreach.* (In press).

Hidalgo, C. A. (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.

Jolliffe, F. (1994). Report on the session, *Probability and Statistics for the Future Citizen*. In C. Gaulin, B.R. Hodgson, D. H. Wheeler, and J. Egsgard. (Eds.), *Proceedings of the Seventh International Congress on Mathematical Education*. Les Presses de l'Université Laval, Québec, 1994, p. 495.

Libman, Z. (2010). Integrating real-life data analysis in teaching descriptive statistics: a constructivist approach. *Journal of Statistics Education,* Vol 18, No 1. www.amstat.org/publications/jse/v18n1/libman.pdf

Mackey, S. & Saffron, K. (2004). *A qualitative study of training needs of statistical analysts in the public sector.* Victoria University, Wellington, New Zealand.
OCED (2011), *Better Life Initiative: Measuring Well-being and Progress* http://www.oecd.org/document//

Payne, B. N. (2011). Changing Challenge and Classroom Olympics: Competitive and cooperative hands on data collection activities. In P. Bidgood (ED.), *Proceedings of the 2011 IASE Satellite Conference: Statistics Education and Outreach.* (In press).

Peebles. D. (2011). The effect of graphical format and instruction on the interpretation of three-variable bar and line graphs, University of Huddersfield, UK.

Playfair, W. (1801),*The Statistical Breviary,*. Accessed 28/08/12 via Google Books, http://books.google.co.nz/books/download/The_statistical_breviary_shewing_the_res.pdf

R Core Team (2012). *R: A Language and Environment for Statistical Computing*},R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org.

Ridgeway, J., Nicholson, J. and McCuske, S. (2007). Child's Play – Reasoning with Multidimensional Data. Presentation to the OECD/ISTAT Seminar on "Dynamic Graphics for Presenting Statistical Indicators" Rome**.**

Rosling, H. 2007. Gapminder. GapMinder Foundation www.gapminder.org
Statistics New Zealand (1999). *A Layperson's guide. CPI. All about the Consumers Price Index.* Statistics New Zealand. Wellington. New Zealand. www.stats.govt.nz

Statistical Journal of the IAOS. (2008). *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics.*Vol.25, No.3-4. IOS press.

Stirling, D. (2010). Mastery tests to cope with mixed backgrounds in an introductory statistics course. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Tucker, C. (2010). Recruitment, Training and Retention of Statisticians in the U.S. Federal Statistical Agencies. *Journal of Official Statistics*, **26** (3), 455–464.

Tufte, E. (1983). *The visual display of quantitative data.* Graphics Press, UK

Wild, C. and Pfannkuch, M. 1999. Statistical Thinking in Empirical Enquiry (with discussion). *International Statistical Review*, 67(3): 223-265.

.
Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2009). *Precursor statistical inferences*. Unpublished paper.

Wild, C. J., Pfannkuch, M., Regan, M., Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* Vol. 174, 2. pp 247–295. Royal Statistical Society.

Appendix:

Honours paper in Official Statistics – Course Outline

Teaching institution and department.	**Topic.**
Statistics New Zealand	Overview, underlying principles, key case studies (observational, not experimental) advantages and limitations
Victoria University of Wellington Statistics dept.	Administrative and survey data – sample/population, measurement and framework concepts. Case studies for variables, collection
	ASSIGNMENT – 20%
Waikato University Population Studies dept.	Demography – fertility, mortality, migration – age structure analysis
	Demography –population projections, policy implications, life tables, cohorts
	ASSIGNMENT – 20%
Auckland University Statistics & Social Science depts.	Health statistics – age standardization, morbidity statistics, registers, data sets, data access, relative risks, odds ratios, other risks, confounding
	Other social statistics– registers, data sets and data access
	ASSIGNMENT – 20%
Canterbury University Geography dept.	Data visualisation and GIS
Victoria University of Wellington Statistics dept.	Survey design and analysis (cross-sectional, longitudinal, rotating panel), data cleaning, editing/imputation, post stratification, survey weights
Auckland University Statistics dept.	Data matching
	ASSIGNMENT – 20%
Statistics New Zealand	Economic statistics, time series, seasonal adjustment, indices - CPI, PPI National accounts GDP
	ASSIGNMENT – 20%