

® GRAPHARTI

GREEN, Hilary
Macquarie University
Australia

In this paper, I introduce a graph which I call Grapharti. Grapharti is designed to display data obtained from surveys, opinion polls or evaluations. The data are retrieved from a database and displayed on a web page. The purpose of this paper is to show that grapharti can enable objective interpretation of the data; however, it also aims to invite subjective interpretation of the data.

INTRODUCTION

“Information Visualization extends traditional scientific visualization of physical phenomena to diverse types of information (e.g. text, ...) from large heterogeneous data sources” (Wise et al., 1995). It is in this context that I introduce *grapharti*. I have designed *grapharti* to visualize data obtained from surveys, opinion polls or evaluations, large heterogeneous data sources. The name *grapharti* has been coined to incorporate three features of the display. The first is that the display is meant to graph the data. The second is that the display is designed to be visually attractive, and lastly, that the display is meant to communicate opinion and be visible to onlookers, in the sense of graffiti.

I suggest some ways in which *grapharti* can be used in the context of statistical education at various levels. I also show that *grapharti* has a use in the wider community.

Since a *grapharti* is a visual summary of data, then it is a type of statistical graph. There may be several variables attached to each item in a survey: the question itself; the date of the question; the number of responses; the number of “yes” responses; the primary topic of the question and perhaps a secondary topic of the question. Using *grapharti*, the user is able to select variables and generate various graphs using html and scripting functionality. In this paper, the principles on which *grapharti* is based are explained and the method is demonstrated using data from the opinion polls carried out on the [ninemsn](#) website. The questions may be grouped by a primary topic and then ordered by number of responses or date. A simple image is allocated to each question. The size of the image is determined by the number of responses and the colour determined by the proportion of “yes” responses. These images are tiled alongside each other and form a tapestry. *Grapharti* results in an interactive and colourful display. By rolling the mouse over the image the user is able to extract extra information as required. “The spatial representation may then be visually browsed and analysed in ways that avoid language processing and thus reduce the analyst’s mental workload” (Wise et al, 1995).

Graffiti, in today’s context, is often a mixture of words and pictures displayed at public places, such as walls, with the intention of communicating strong political and/or social opinions of individuals or groups. Typically, graffiti is easily understood, topical and usually emotive. Like graffiti, the objective of *grapharti* is to communicate opinions in an art form. Both are intended to produce a visual image which captures some essence of the data, as well as evoke subjective responses.

Consequently, *grapharti*, as an art form, is a means of communicating information derived from the data as a display. The viewer is prompted to actively investigate this information while judging the attractiveness of the generated display. The “interweaving of interior mental action and external perception (and manipulation) is no accident. It is the essence of how we achieve expanded intelligence.” (Card et al., 1999). As a statistical graph, *grapharti* is a means of summarizing a large amount of data concisely. Either as a piece of art or a statistical display, *grapharti*, should provoke further investigation of the data and so is a means of communicating with statistics.

THE DATA

Each day thousands of people respond to online surveys. A common goal of online surveys is to maximize the response rate and one way to ensure this is to provide the facility for quick and painless responses. Multiple choice questions appear to be the most popular format for online surveys. Examples are many. Typically, results from individual questions in these surveys are presented in bar charts, as are the daily results on the [ninemsn](http://news.ninemsn.com.au) website. Bar charts are effective in displaying results from one or two variables; however, even clustered or divided bar charts or mosaic charts are limited to displaying only a few variables at a time. The purpose of grapharti is to organize and summarize a large amount of data, comprising of several variables, one of which is text, in a simple display in order to promote and facilitate investigation of the data.

In this paper the methods are applied, with permission, to the data from the ninemsn surveys, where there is a binary response to a different question posed daily. After choosing to vote in a poll, a small pop up window opens with YES and NO buttons. Mouse clicking one of these options triggers the current results being displayed numerically and in horizontal bars. (See Figure 1). It would be a fairly tedious and fruitless task to vote again on the same question on the same computer.



Figure 1. Screen shots of pop-up windows with (left) option to vote and (right) results.

The user may choose to view the results of previous surveys. These are listed, showing the dates, the questions, the numbers and percentages of yes and no responses, ordered by date. It is the information from this list that I use to demonstrate grapharti. I have included two other variables, `topic_1` and `topic_2`, to indicate the subject/s of the questions. Some of these topics include: party politics, sport, refugees, environment, family issues, education, health, media and so on. Questions have been allocated to one or two of these categories.

The [ninemsn](http://news.ninemsn.com.au) website claims to capture the largest online audience in Australia. In 2004, an average of about 42,000 readers responded to each online survey on this website. Figure 4 shows that the number of respondents to the surveys has been increasing since their inception in December, 2000. I assume that the respondents to individual questions are samples which could be considered as representative of some fairly significant, but unknown, subsection of the Australian population. To obtain an overview of this population we could investigate the results of the surveys using tables or some sort of display. Grapharti is an attempt to assemble the large amount of poorly connected data into single displays with the intention of revealing hidden patterns. 'If the visual task is contrast, comparison and choice then the more relevant information within the eyespan the better' (Tufte, 1990).

THE DISPLAY

To organize and summarize the large amount of data in a simple display, I have chosen to represent the results from each individual survey question with a colour-coded bar positioned on a web page: the results from all the selected survey questions form a wall of bricks. Details for a

particular question are found by rolling over the particular brick with a mouse.

The bars are adjoining and wrap left to right. The height of the bars is constant and the width is proportional to the number of respondents. The proportion of *yes* voters is proportional to a number corresponding to a hexadecimal value of a colour. For example, a very high proportion of *yes* voters may realise dark red while a very low proportion of *yes* voters may realise a deep blue. Colour schemes are chosen so that the resultant image could resemble a scene or be otherwise aesthetically satisfying. This image of the ordered colours appears on the index page of the site. An example of an index page appears in Figure 2. The colours should also translate to colours on the greyscale of corresponding brightness, to accommodate those with colour vision impairment. The tiles are an example of small multiples which have the effect of “visually enforcing comparisons of changes, of the differences amongst objects, of the scope of alternatives” (Tufte, 1990).

From the index page, the user has the option of displaying the data in tables or as grapharti. The index page also includes options, such as to display all data, data from polls over a certain time period or data referring to selected topics. The user may also be given the option to order the data by date or number of respondents, rather than by proportion of *yes* voters (default). To motivate the user to investigate the data, he/she should be able to obtain the desired result quickly. It is important that the index page have clear instructions, be uncluttered and make use of features familiar to most users. The index page includes a link to a page explaining the idea of grapharti and how to interpret the display. In a later development the image on the front page may change so that it corresponds to the option selected.

When the grapharti is displayed in full in a new window the user is prompted to roll the mouse over the image. A rollover text appears above bars showing the question, and more information such as the date and the number of respondents. The user can choose whether to display borders (default) or not. The border colour of the cells is thin and grey and assists the viewer in separating the individual records. However, if omitted, the graph may look more like a picture.

Analyse results from ninemsn opinion polls with Grapharti

Make a selection from the list below to enable you to interact with the [ninemsn questions](#) database.

Tables

- List Questions in date order
- List Questions sorted by number of voters
- List Questions sorted by number of "yes" responses
- List Topics

Grapharti

- Grapharti by date
- Grapharti ordered by proportion of "Yes" voters
- Grapharti and politics
- Show another Grapharti table of colours

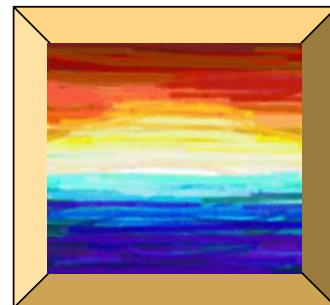


Figure 2. Screen shot of a web page introducing grapharti image and display options.

In developing grapharti I have stored the data in Microsoft Access and I have used html, JavaScript and VBScript to construct the website from which the grapharti is accessible. At this stage of development the page has been tested on Internet Explorer.

COLOUR SCHEME

The choice of colour scheme is most important in making grapharti both attractive and effective in communicating meaning from the data. Tufte (1998) recommends using visual elements that make a clear difference but no more, contrasts that are definite, effective and minimal. In the grapharti of the ninemsn data, colour is used to display the proportion of *yes* votes, by default. The colour scale that has been used in the image in Figure 3 ranges from deep red, used to represent a high proportion of *yes* votes, $P(\text{yes})$, through orange, yellow and white representing a proportion of 50%, followed by pale blues through to deep blues which represent a high proportion of no votes. By using hexadecimal numbers on the blue scale that mirror the hexadecimal numbers on the red scale I have obtained contrasting colours of the same hue. I have obtained 100 hexadecimal numbers and ordered them to correspond to the proportion of *yes* voters rounded to 2 decimal places. It is not really necessary to include a legend on the display, as the proportions are revealed by mousing over bars.

EXAMPLE

It is clear from the grapharti in Figure 3, that a high proportion of respondents believe that the cost of university education has become too expensive. Questions that returned the highest proportion of *yes* responses were: *Should people smugglers be jailed? Should the employment of security personnel be more tightly regulated? Do you believe there is corruption in the horse racing industry? Should the Reserve Bank force banks to cut credit card fees? Should One.Tel's founding directors surrender their bonuses? Do you think Zimbabwe's election was rigged? Should there be regulation to limit bank fee increases?*

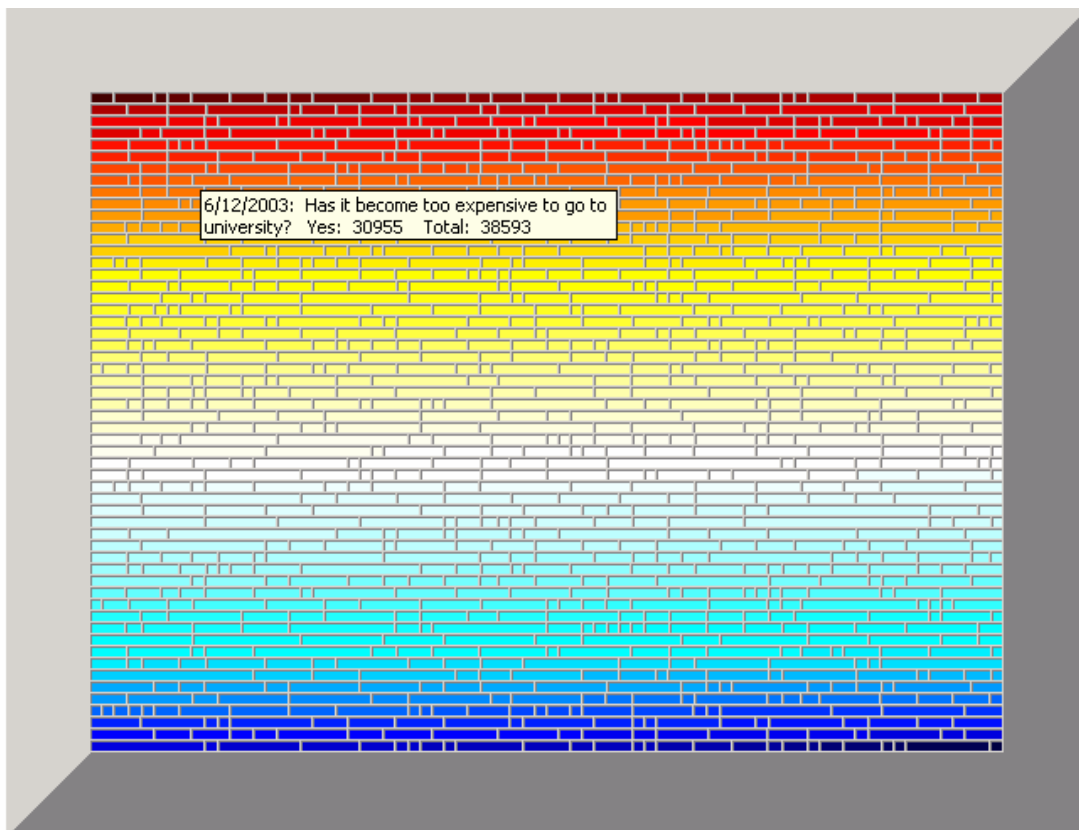


Figure 3. An example of grapharti of the data from the ninemsn website. The default colour scheme and ordering has been used.

Some questions that attracted the highest number of respondents were: *Should the PM resign if he lied about the children overboard affair?* $P(\text{yes}) = 0.49$. *Do you now support a war against Saddam Hussein?* $P(\text{yes}) = 0.45$. *Do you have sympathy for war-on-terror detainee David Hicks?* $P(\text{yes}) = 0.69$. *Do you believe Howard was unaware that boat-children claims were false?* $P(\text{yes}) = 0.45$. *Is Pauline Hanson's three year jail sentence too severe?* $P(\text{yes}) = 0.73$.

Some questions attracted an unusually large number of responses; such as *Should America retaliate with extreme force?* ($P(\text{yes}) = 0.49$) on the day following 9/11/2001. Some questions attracted an unusually small number of responses, such as *Do you think the US is getting too involved in Australian domestic politics?* ($P(\text{yes}) = 0.66$) on the 6th June, 2004. A reason for this question having such a low number of respondents is that it only appeared on the ninemsn website for a few hours.

By selecting a particular topic, such as *refugees*, displaying the grapharti in date order, it may be possible to determine whether the opinion of the target population has remained constant over time. And by displaying the same questions in order of $P(\text{yes})$, one could gain further insight into the target population's feelings on the issue.

MESSAGE

The responses to grapharti are subjective. Each user independently analyses the data and focuses on aspects of particular interest. Nevertheless, independent findings may provoke further analysis or discussion. Because the data have been obtained from opinion polls the validity of the data as a representative sample of the Australian population, is suspect. The disclaimer on the ninemsn website reads "ninemsn excludes any warranty that these poll results were obtained using a reliable statistical sample." The same disclaimer should appear with the grapharti image. The results are thus invalid for scientific research. However, do they tell us something? Up to 130,000 people respond to a ninemsn survey in one day and the number has been increasing since it began. Issues relating to sample size, representative or biased samples are fundamental to sensible interpretation of the data. Grapharti presents a situation where the lay user is also exposed to these issues.

AUDIENCE

A target audience for this grapharti would be participants in the online surveys and visitors to the site. Grapharti presents the data in a simplified format and so is suitable for a lay audience. Users would interact with grapharti enabling them to search for visual hints to help them decipher the multiple images. The non-threatening format should entice users to extract various layers of information from the display.

Grapharti may be of use as a tool for data mining, which involves sorting through data to identify patterns and establish relationships. There are initially no research questions; the user is engaged in extracting meaning from the data, rather than applying statistical methods involving inference or hypothesis testing. Potentially, however, there is a vast number of research questions that could be postulated by the user as he or she reads the grapharti and on reflection.

Grapharti may be of use as a tool for exploratory data analysis (EDA) (Tukey, 1977) where the primary goal is to maximize insight into a data set and into the underlying structure of a data set. "Graphs allow us to explore data to see overall patterns and to see detailed behaviour; no other approach can compete in revealing the structure of data so thoroughly"(Cleveland, 1994, p. 5).

IN THE CLASSROOM

There are many ways that grapharti may be of use as an educational tool for statistics students, and as a pedagogical device. For example, grapharti of the ninemsn data could be used in the statistics classroom to stimulate discussion and research on a wide range of issues and some suggestions follow. Initially, the target population would need clarifying, its composition and whether or not it is constant from day to day or from topic to topic. How typical is the target population of the Australian population in general? There are many problems regarding the samples; for example, can we assume there is no significant difference in the nature of the respondents from day to day? Is it a problem that it is possible, though time consuming, for a person to vote more than once on the same question? The idea of independence could follow from

discussions about samples. The degree to which the user believes in the reliability of the results is fundamental to the interpretation of the grapharti. Do the large sample sizes have any bearing on this reliability? Are the results biased, and if so, is it possible to measure this bias?

The idea of variables can be introduced. What has been measured or counted? Are the data continuous or categorical? What other graphs can be used to display the data from one or two of the variables? What are the features of a good graph? How does grapharti rate as a graphic? How have the data been organised in order to display the grapharti? Should the proportion of *yes* votes be used or would the odds of *yes* votes be more meaningful?

What are the research questions? Should these be postulated before looking at the data? What does data snooping mean? What are the similarities and differences between data snooping, data mining, statistical data analysis and statistical inference?

The deep, fundamental question in statistical analysis is *Compared with what?* (Tufte, 1998). This statement could motivate much discussion with respect to grapharti. For example, can we compare the results with results from similar surveys, election results and popular opinion? How can these comparisons be made? The notion of hypothesis testing could follow. Are the responses consistent over time? Can a change in opinion of the population to various issues be detected over time?

The scatterplot in Figure 4 reveals that since the inception of the ninemsn opinion polls the number of respondents has been increasing. This could trigger discussion relating to linear regression. The notion of outliers could be introduced, and investigation into the questions that provoked the unusual numbers of responses should also prove interesting. Discussion could also arise as to whether the data should be detrended when presented as grapharti so as not to distort its interpretation. A large number of respondents in 2000 would represent a small number of respondents in 2004.

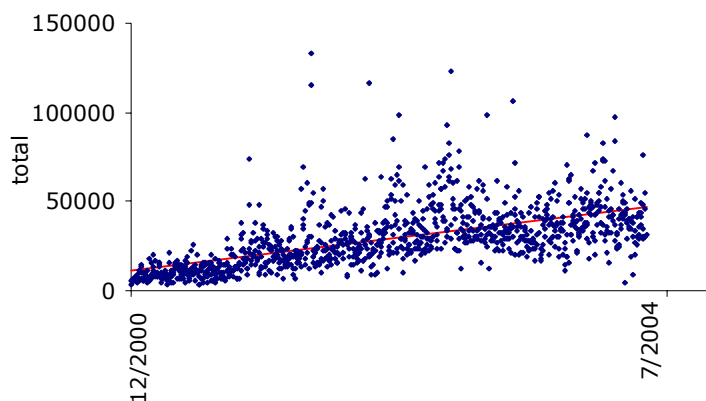


Figure 4. The number of respondents has tended to increase over time. Some questions have attracted an unusually large number of respondents.

The extraction and analysis of the data from a database would be of interest to statistics units dealing with data mining and also graphics. The design of the web page displaying the grapharti should be of interest to students of web database engineering.

Grapharti should also suggest other questions of interest such as; what questions prompted the largest/ smallest number of responses? What topics attract the largest number of responses? On what issues are the respondents equally divided? Has there been a change in attitude (eg towards refugees) over time? What issues are not covered in the surveys? And finally, what are the main characteristics of the target population?

DISCUSSION AND CONCLUSION

The intention of grapharti is to enable a visitor to a website to derive further meaning from results of online surveys. It is not meant as a scientific tool but rather as one that facilitates

exploration of the questions and responses. This is achieved by the design elements which provide the user with choices as to how the data are to be grouped and displayed.

At this early stage of development stage grapharti as a tool for statistical communication has neither been trialled nor evaluated and this is obviously a necessary step in the project. However, I believe that grapharti could be used effectively as a motivational tool not only for students in statistics at various levels, but also for the general population. At the most basic level, grapharti presents the data as a picture, where the attributes of colour and size of each bar add meaning to the question which the mouse rolls over. By grouping the data in various ways the user is likely to acquire an idea or a picture of the voting population. On the other hand, it would be interesting to discover whether students and teachers of statistics regard grapharti as having any merit as a statistical graph. Such a debate would necessarily include many issues that are fundamental in the communication of statistics.

REFERENCES

- Card, S.K., Mackinlay, S.K., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers, San Francisco, Calif.
- Cleveland, W.S. (1994). *The elements of graphing data*. AT&T Bell Laboratories. Murray Hill, N.J.
- Ninemsn. *Welcome to ninemsn.com.au*. <http://portal.ninemsn.com.au/default.aspx> (accessed 21/02/05)
- Tukey, J.W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, Mass.
- Tufte, E.R., (1990). *Envisioning Information*. Graphics Press, Cheshire, Conn.
- Tufte, E.R., (1998) *Visual explanations: images and quantities, evidence and narrative*. Graphics Press, Cheshire, Conn.
- Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995). *Visualizing the Non-Visual: Spatial analysis and interaction with information from text documents*. In S.K. Card, S.K. Mackinlay, and B. Shneiderman (Eds), *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers, San Francisco, Calif.