Visualising Project Management Dimensions

Luis Copertari
Autonomous University of Zacatecas, México

Abstract

It is easier to decide upon which projects to include in a portfolio if a comprehensive human–computer interface is provided. Two alternative displays were designed according to best practice from the literature and tested according to a simple mathematical optimisation model and the average error from testing subjects and the "optimal" solution to decide which one performs better.

Keywords: Visualisation, human–computer interface, project portfolio

Introduction

Visualisation includes the study of both image synthesis and image understanding, spanning many academic disciplines, scientific fields, and multiple domains of inquiry. Lohse et al. (1994) argue that the need for classification schemes is based on the fact that classification is at the heart of every scientific field. Classifications "structure domains of systematic inquiry and provide concepts for developing theories to identify anomalies and to predict future research needs". Graphs and images can be used to visualise classifications and different dimensions. Graphs and images can be characterised as either functional (focus on the intended use and purpose of the graphic material) or structural (focus on the form of the image rather than its content). Graphs encode quantitative information using position and magnitude of geometric objects. Numerical data in one, two, or three dimensions are plotted on a Cartesian coordinate or polar coordinate system. Common graph types include scatterplot, categorical, line, stacked bar, bar, pie, box, fan, response surface, histogram, star, polar coordinate, and Chernoff face graphs. Preece et al (1996) and Dix (2003) identify seven techniques to represent numeric data: scatterplots, line graphs or curves, area, band, strata or surface charts, bar graphs, column charts or histograms, pie charts, simulated meters, and star, circular or pattern charts. In this work, a modified version of scatterplots in a Cartesian coordinate system is used, with position, length and area judgements for information coding.

Projects are typically specified using three dimensions: time, cost, and performance. Another dimension is added, risk, to consider the probabilistic nature of the portfolio selection process. In the following sections, the reasoning process followed and the interface features of two alternative displays are discussed based on the literature reviewed (theory). This leads to the conceptualisation of two interfaces to choose from (design) is the result. A simple experiment is devised (testing) to determine which interface performs better in each circumstance (discussion and conclusion).

Theory

Bertin (1983) defines understanding as "simplifying, reducing a vast amount of «data» to the small number of categories of «information» that we are capable of taking into account in dealing with a given problem". Preece et al. (1996) and Dix (2003) discuss what is known in the Human–Computer Interaction (HCI) literature as the 7±2 magic number, related to short term memory, which shows that humans are able to recall between 5 and 9 numbers or figures at the same time. That is one of the reasons why a good HCI display is critical, since it allows users to consider several numbers all at once if the display presents information in a meaningful way. Although this concept regarding understanding seems to be most accurate,

the human brain is actually much more capable than Bertin seems to imply. The mind is able to make abstractions, synthesise various elements from reality, and put them together using not only short-term, but also long-term memory. A suitable design is a polysemic graphic system, in which the meaning of the individual signs follows and is deduced from consideration of the collection of signs. For the purposes of this paper, perception deals with the ability of any given individual (or group of) expert(s) to find relationships between the images and the real world, in an attempt to reach the best project portfolio. The information displayed on the screen or printed on a sheet of paper is the result of summarizing in a plot, based on mathematical models, the combination of all the available data from historical records (retrieved from a database) and the input obtained from experts during each session.

For any given project, there are four dimensions: time, cost, performance, and risk. Let t_k, c_k, p_k, and r_k be the time, cost, performance and risk of project k for all k=1…s, where s is the total number of projects (portfolio size). Also, let the zero-one decision variable x_k indicate whether or not project k is selected; if $x_k=0$ project k is not selected, if $x_k=1$ project k is selected. In vector notation, $\mathbf{T} = \{t_1, t_2, …, t_s\}$, $\mathbf{C} = \{c_1, c_2, …, c_s\}$, $\mathbf{P} = \{p_1, p_2, …, p_s\}$, $\mathbf{R} = \{r_1, r_2, …, r_s\}$, and $\mathbf{X} = \{x_1, x_2, …, x_s\}$. Vectors $\mathbf{T}$, $\mathbf{C}$, $\mathbf{P}$, and $\mathbf{R}$ are row vectors, whereas vector $\mathbf{X}$ is a column vector. These estimates are the result of applying mathematical models to the raw (input) data obtained from historical records and/or decision-makers' expertise. For the purposes of this empirical study, these figures are given. The objective is to maximize performance subject to cost and risk constraints. The cost constraint is not to exceed the budget (B); the risk constraint is not to exceed, on average, investor's risk preferences (K). Equations (1) to (3) portray the simplified zero-one integer programming model used in this study to find the optimal solution (see Ghasemzadeh, et. al (1999) for a detailed discussion). Notice that equation (3a) is a modified version of equation (3), which is more helpful because it says that the combined risk of the selected projects should be less or equal than a given threshold K.

Maximize $\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $\mathbf{PX}$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ (1)
Subject to:
$$\mathbf{CX} \leq B \quad\quad\quad\quad (2)$$
$$(\mathbf{R} - K)\mathbf{X} \leq 0 \quad\quad\quad\quad (3)$$
$$\mathbf{RX} \leq K \quad\quad\quad\quad (3a)$$

The challenge is to meaningfully portray 4 dimensions of data. In table format there are s rows and four columns; each row represents a project and each column a project dimension. The problem is what Bertin (1981) calls the impassable barrier: with up to three columns, a data table can be constructed directly as a single image, producing a scatter plot or correlation diagram, in which the objects are in the third (vertical dimension) typically denoted as the z axis. But there are four (not three) dimensions to picture. Is there anything that can be done to avoid sacrificing the overall relationships of the entire set?

Considering what can be represented in a flat sheet of paper, a graphic system can include eight variables besides the two to three axes of the plane or space: a) size, b) value, c) texture, d) colour, e) orientation, and f) shape. Cleveland & McGill (1984) ordered from most to least accurate the 10 elementary perceptual tasks:
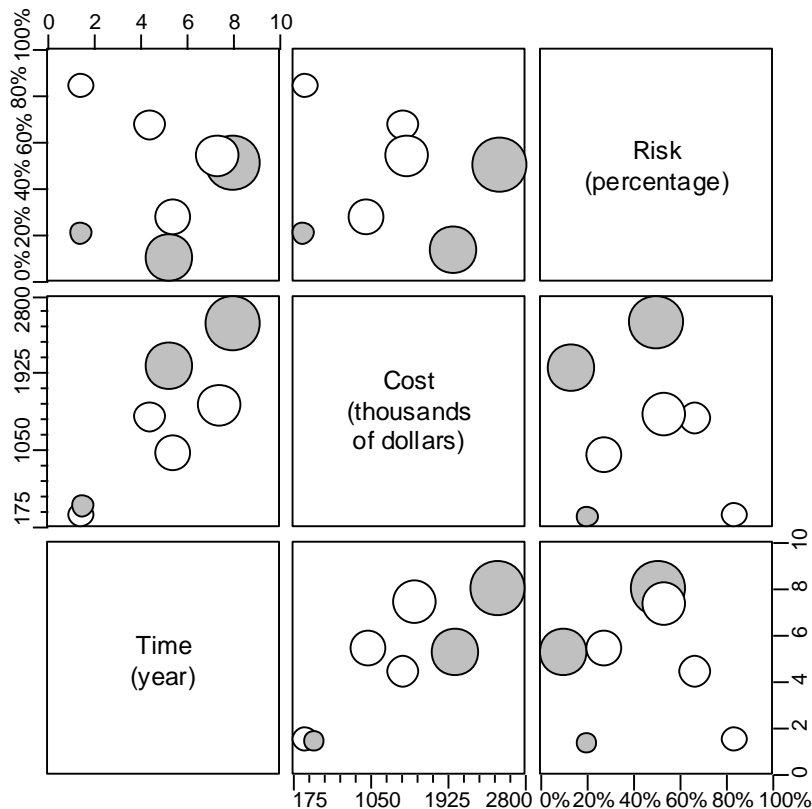1)   Position along a common scale.
2)   Positions along nonaligned scales.
3)   Length, direction, angle.
4)   Area.

5) Volume, curvature.
6) Shading or color saturation.
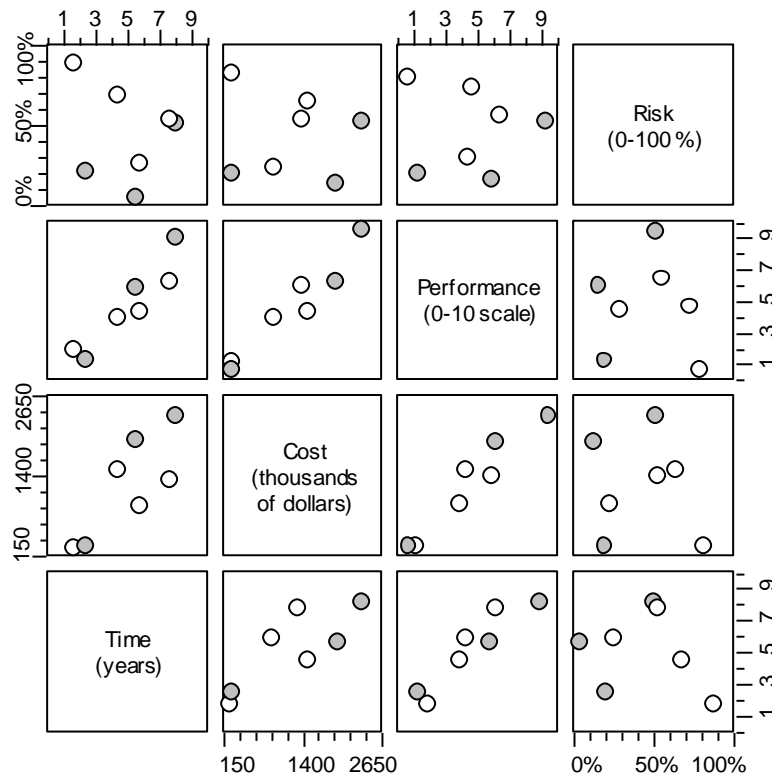
Figure 1. Multipanel Scatterplots.

The main problem is how to portray four dimensions. Cleveland (1993) explores the use of multiple scatter plots in a multi-panel display of four rows and four columns for hyper-variant data. Each pair of variables is graphed on a scatter plot within each panel; the left column is column one and the bottom row is row one. The graphs are arranged in a shared-scales matrix: along each row or column, one variable is plotted against all others, which allows to visually linking features of one scatter plot with another. The data is pictured by relying on position along a common scale: small circles correlate each pair of variables (see Figure 1).

Figure 2. Multipanel Bubble Scatterplots.

The display in Figure 1 shows too much information at once. Bertin (1983) illustrates the use of bubbles to picture position along the row/column common scale in multiple scatter plots. The diameter (as opposed to the area) of the bubbles pictures another dimension because length is more accurate than area. The scatter plot shown in Figure 1 can be modified by portraying the first two dimensions (time and cost) as well as risk along the common scales, depicting performance as the diameter of each bubble. Intuitively, a good portfolio minimises time and cost while maximising performance for a given risk profile. Risk should not be portrayed as area because the risk profile is usually expressed as a risk range (or maximum total risk tolerance, K), and position along a common scale allows a more accurate perception than length. Since time and cost share the same objective (minimisation), it seems reasonable to use the same elementary perceptual task (position along a common scale) in both cases. The remaining variable is performance, which is portrayed using the diameter of the bubble as shown in Figure 2.

Design

Measuring interface efficiency is the objective (Lazar, et al. 2010 and Jacko 2012), where efficiency is defined as the percentage of correct portfolio selections from the user using a given display such as the one from Figure 2 when compared to the optimal solution according to the model from equations (1) to (3a). Although Figure 2 meaningfully displays all four dimensions at once, it would be difficult to make sense of the information for larger portfolios, because the bubbles would interfere with each other or if the display were scaled down, the resulting bubbles would be too small to be clearly visible. Even if each display is clear, excessive information may clutter the computer display and confuse the user. The solution is to choose which dimension to portray. But which of the four dimensions should be chosen for each axis? Looking back at the decision support model from equations (1) to (3a), it is clear that only cost and risk are constraints. Thus, it makes sense to assign risk and cost

to the horizontal and vertical axes, respectively, leaving performance (the objective function) to be depicted by the diameter of the bubbles, ignoring time as being irrelevant in this particular case (time is solved using PERT/CPM). Users will be advised to choose projects as close to the horizontal axis (i.e. as low in cost) and as large in diameter as possible while at the same time keeping the average risk below the maximum allowed.

Another way to improve the display is by substituting framed rectangles for bubbles. According to Cleveland & McGill (1984), position along non-aligned scales is more accurate than length. The height of a rectangle can be used to portray performance instead of the diameter of circles. To portray information more accurately the rectangle is framed within another rectangle in such way that a sense of percentage can be obtained by comparing the relative distances between the inner and the outer rectangles. Let project j be the project of the maximum performance; then the performance of project i is given as a percentage of that as shown in Figure 3.

Project i

Project j
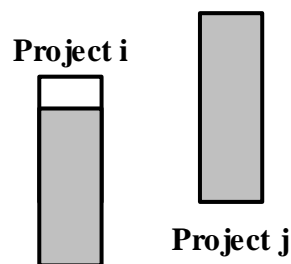
Figure 3. Framed Rectangles.

Consequently, there are two alternative displays. The first display (see Figure 4) is the bubbles chart display and it portrays performance of each portfolio as the diameter of each bubble. The second display (see Figure 5) is the frames chart, portraying project performance as the height of the rectangle, which provides an idea of percentage because the inner rectangle can always be compared to the outer rectangle.
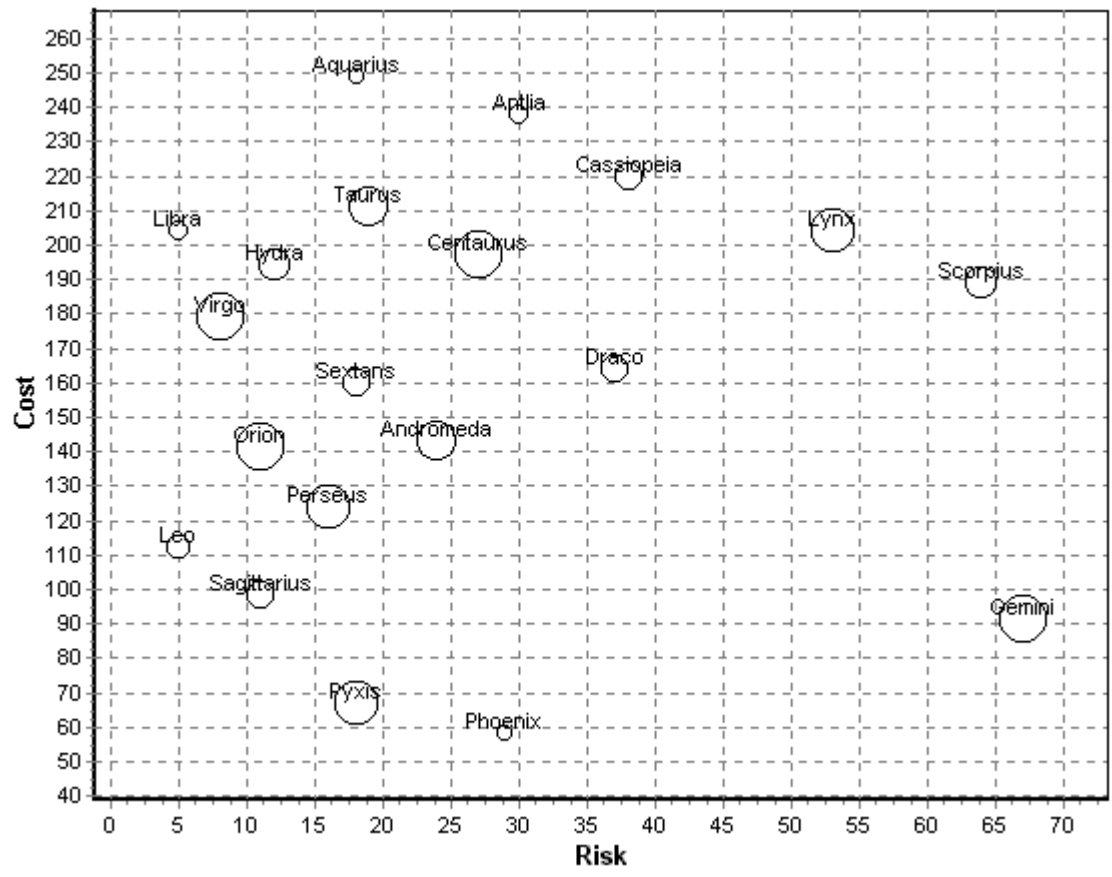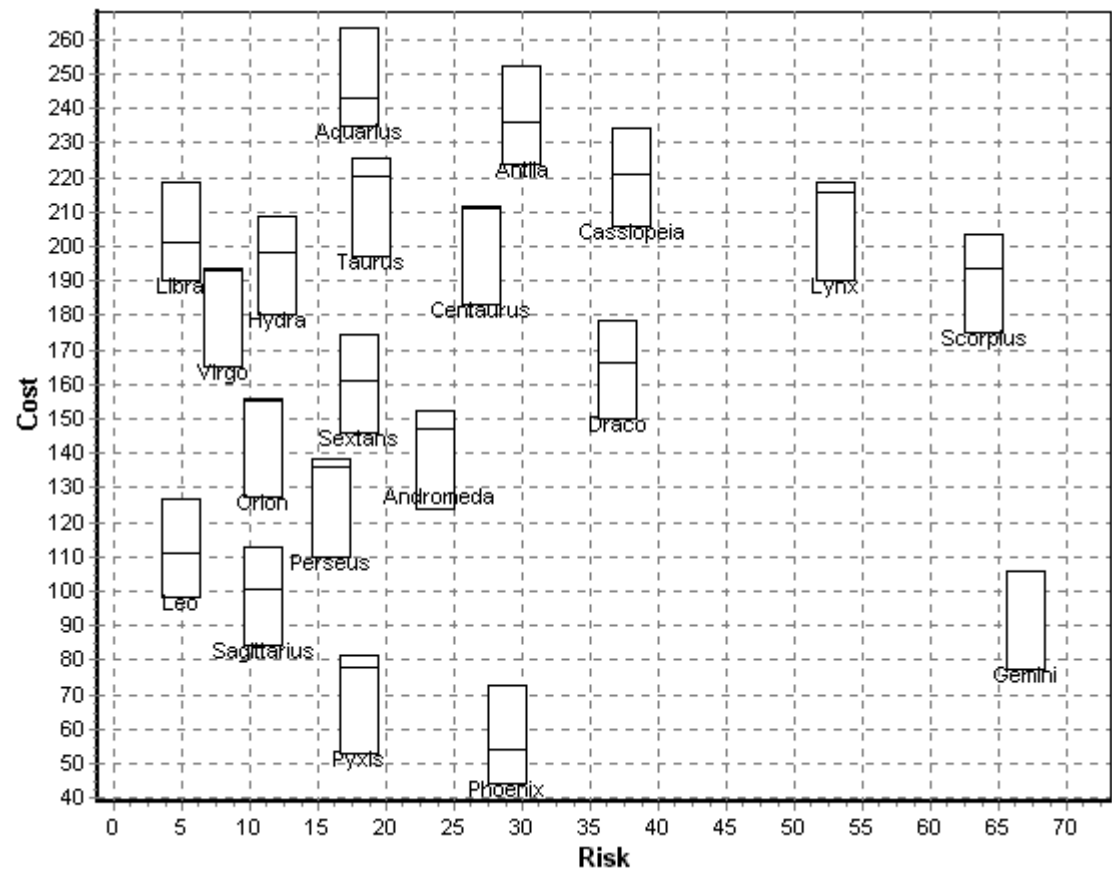
Figure 4. Bubble Chart.

Figure 5. Frames chart.

The hypothesis being tested is that the second display (frames chart) should portray data better for representing performance using positions along nonaligned scales (rectangle-within-rectangle idea) as opposed to the use of length of the first display (bubble chart). Further, such interface improvement should translate into a better ability of the user to interact with the display and to select portfolios closer to the optimal solution.

Testing

A test to decide which interface display (bubbles chart in Figure 4 versus frames chart in Figure 5) performs better was devised. Performing better in the context of this discussion means that, on average, the display leads users to solutions closer to the optimal. Thus, to decide which display is better, the error for a given portfolio and the average error for a given display are needed.

The error (E) for a given portfolio is the sum of the squared differences between the solution vector and the data vector. Solution ($S = \{S_k$ for all k=1…s}) and data ($D = \{D_k$ for all k=1…s}) vectors are binary vectors of size s, where s indicates portfolio size. S_k=1 in the solution vector means that project k is part of the optimal solution; conversely, S_k=0 indicates that project k should be kept out of the solution vector. D_k=1 means that the user selected project k as part of the portfolio; D_k=0 indicates that project k was not included in the portfolio. The error (E_x) as shown in equation (4), indicates the total number of misallocations for experiment x, where x = 1…m, and m is the total number of experiments: either the user selected a project that should have not been selected, or the user forgot to select a project that had to be selected. It is also possible that there was no error for experiment x and project k, that is, either the user selected the project, which was to be selected, or the user did not select the project, which was not to be selected.

$$E_x = \sum_{k=1}^{n} \left(S_k - D_k \right)^2 , \ x = 1…m \tag{4}$$

Values for E_x occur for the data of each and every subject in the experiment. Remember that m is the total number of subjects and E_x denote the error on experiment x. Then, the mean (average) error can be calculated as the average of errors as shown in equation (5).

$$\overline{E} = \frac{\sum_{x=1}^{m} E_x}{m} \tag{5}$$

The average error can be used to compare the two displays because the best display is expected to have a lower error when compared to the alternative display.

Discussion and conclusions

A total of 50 subjects participated in the pilot study (m = 50). The pilot study involved interacting with both displays (one at a time) and deciding, based on the visual information displayed, which projects to select. All subjects were advised of the best strategy: go for the projects with the highest performance first. As it turns out, it is more difficult to identify which projects have the largest performance using the bubbles chart. In fact, 35 out of 50 users said they preferred the frames chart compared to the bubbles chart. The mean error for the bubbles chart was 2.1 whereas the mean error for the frames chart was 1.8. The latter

seems to indicate an advantage of the frames chart compared to the bubbles chart. The number of users who found the optimal solution is larger in the case of the frames chart (20 users found the optimal solution using the frames chart, whereas only 15 users found the optimal solution using the bubbles chart).

Only 10 users complained they did not have enough time to complete the experiment. On a scale from 1 to 7 (where 1 is strongly disagrees and 7 is strongly agrees), the group rated the bubbles chart as 5 and the frames chart as 5.44 in usefulness. In summary, it seems the frames chart is the favourite. The small advantage in the mean error of the frames chart seems sufficiently large to accept the hypothesis, considering the evidence in favour of the frames chart when it comes to its usefulness.

References
Bertin, J. 1981. *Graphics and Graphic Information Processing*, Walter de Gruyter.
Bertin, J. 1983. *Semiology of Graphics: Diagrams, Networks, Maps,* The University of Wisconsin Press.
Cleveland, W.S. 1993. *Visualizing Data*. Murray Hill, AT&T Bell Laboratories.
Cleveland, W.S. & McGill, R.. 1984. Graphical perception: theory, experimentation, and application to the development of graphical methods, *Journal of the American Statistical Association*, Vol. 79, No. 387, pp. 531-554.
Dix, A. 2003. *Human-Computer Interaction*, Prentice Hall.
Ghasemzadeh, F, et. al. 1999. A zero-one model for project portfolio selection and scheduling, *Journal of the Operational Research Society*, Vol. 50, pp. 745-755.
Jacko, J.A. 2012. *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, CRC Press.
Lazar, J., et al. 2010. *Research Methods in Human-Computer Interaction*, John Wiley and Sons.
Lohse, G.L., et al. 1994. A classification of visual representations, *Communications of the ACM*, Vol. 37, No. 12, pp. 36-49.
Preece, J., et al. 1996. *Human-Computer Interaction*, Addison-Wesley.

About the author:

Dr Luis Copertari is an active professor and researcher of the Computer Engineering Department at the Autonomous University of Zacatecas. He has participated as a congress speaker at the Portland International Conference on Management of Engineering and Technology 2001. Since then, he has conducted research in Project Management, Human-Computer Interface, Operations Research and Artificial Intelligence, among others.
Email: copertari@yahoo.com

Appendix. Gathering data from testing subjects

The data for the experiment was gathered between 1999 and 2001 at the Michael G. DeGroote School of Business (McMaster University). A total of fifty subjects from the Master in Administration Program were asked to participate and paid each ten dollars (Canadian, at the time the currency exchange was approximately CAN $1.2 per USD $1), so that they were motivated to conduct the study.

The population comprised male and female MBA students selected randomly from the class. The format of the review was simple. Individuals were asked to participate voluntarily, although the CAN $10 dollars incentive was important so they did not consider the experiment a waste of their time.

Summary of the findings

Total number of experimental subjects:	50
User interface preference:	35 out of 50 users said they preferred the frames chart.
Mean error ($\square$)	2.1 for the bubbles chat and 1.8 for the frames chart.
Optimal solution found:	15 subjects with the bubbles chart and 20 subjects with the frames chart.
Complaints:	10 subjects said they did not have enough time to complete the experiment.
Time for the experiment:	Ten minutes.
Usefulness rating (scale from 1 to 7, were 1 is strongly disagrees and 7 is strongly agrees):	5 for the bubbles chart and 5.44 for the frames chart.