Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce 'chart junk' – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors

INTRODUCTION

Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk*, decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates plain and simple charts that maximize the proportion of *data-ink* – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

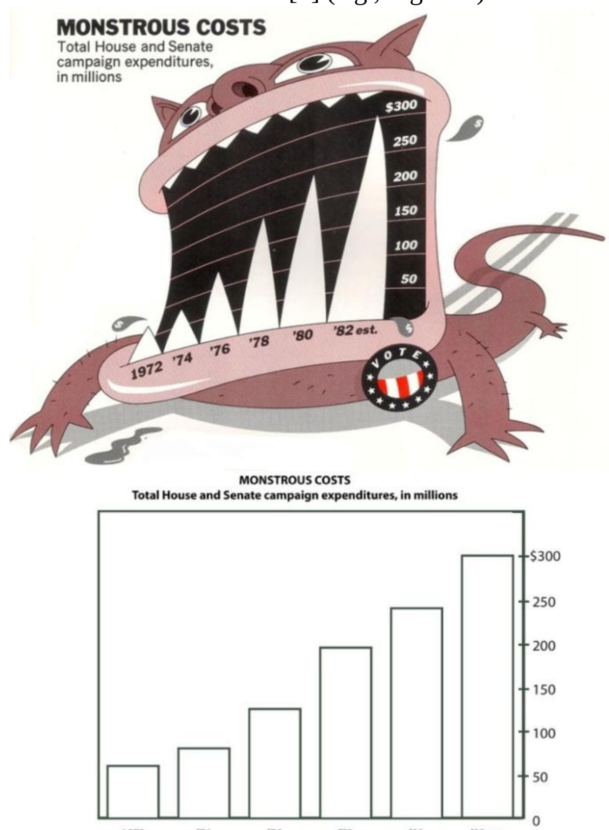

Figure 1. A chart by Holmes [7] (above), and a 'plain' version.

These kinds of charts appear regularly in many mass-media publications, and the widespread use of embellished designs raises questions about whether the minimalist position on chart design is really the better approach. Two issues in particular are raised: first, whether visual embellishments do in fact cause comprehension problems; and second, whether the embellishments may provide additional

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

information that is valuable for the reader. For example, the added visual imagery in a Holmes-style chart could draw the reader's eye, could help to convey a specific message, or could make the chart more memorable.

There is very little evidence from prior research to answer these questions. Previous studies have suggested that minor decoration in charts may not hamper interpretation [1,11,20], and work in psychology has shown that the use of imagery can affect memorability [4,6,14,17], but there is very little work that looks at how chart imagery can affect the way people view information charts.

To answer these questions, we carried out an investigation that tested the effects of visual embellishments on two aspects of interaction with charts: interpretation accuracy, and memorability. We conducted a study in which participants looked at several different minimalist and embellished charts (taken directly from Holmes's work) and answered questions about each chart's topic and details. Participants then answered the same questions again, after a gap of either a few minutes or of around 2 weeks.

Our study showed two main results. First, we did not find any difference in people's interpretation accuracy between the embellished and minimalist charts, despite the fact that the embellished charts contained a much larger amount of non-data-ink. Second, participants were able to remember significantly more about the embellished charts after a gap longer than 12 days, suggesting that the addition of strong visual images does help people to remember both the topic and details of the chart.

We followed up these performance results with an analysis of eye tracking data, which showed that people did spend time looking at the non-data ink in the embellished charts. These results support the idea that people do notice and process the visual imagery, which may be helping to encode the chart in their memory. Further, in their subjective ratings of which type of charts they preferred for reading and remembering participants overwhelmingly chose Holmes-style charts over plain charts.

Our study showed that if memorability is important, visual imagery can help to fix a chart in a viewer's memory. However, we do not advocate this strategy as a general principle, because the use of strong images in charts is contentious – for example, images convey messages that may intentionally or unintentionally bias the viewer's interpretation of the data. More importantly, then, our work shows that there *can* be strong effects from the inclusion of visual embellishments – something that has not been reported before – and that this phenomenon should be better understood by researchers and chart designers.

PREVIOUS WORK

The Debate over Visual Embellishment

In the field of information design there has been a long-standing debate concerning the use of imagery and 'unnecessary' visual embellishments. While some of these arguments are aesthetic in nature, considerable focus is also placed on what presentation style best conveys information to a viewer in graphic form. We divide this work into two sides: those that discourage the use of visual embellishments, and those that encourage their use.

The most well-known proponent of eliminating visual embellishments is Edward Tufte. Tufte takes the stance that "…it's wrong to distort the data measures—the ink locating values of numbers—in order to make an editorial comment or fit a decorative scheme" (p.59, [22]). Tufte proposes the *data-ink ratio*, where he argues that all ink that is not used to present data should be removed [22]. The data-ink ratio can be calculated by dividing the ink used for displaying data (*data-ink)* by the total ink used in the graphic, with the goal of having the ratio as close to 1 as possible. Cleveland [3] similarly proposed a series of recommendations for graph design, based on his research into graphical perception, and emphasized a minimalist approach to presenting data graphically. These minimalist views aim to reduce interpretation effort and increase interpretation accuracy when a user examines an information chart.

Although Tufte's and Cleveland's rules of thumb are well known, designers and publications continue to produce charts with a wide variety of chart junk, decorations, low data-ink ratios, and 3D representations. In a survey of journal publications, magazines, and newspapers from 1985 to 1994, Zacks *et al.* [24] found that there had been little change in the prevalence of graph features that are identified by Tufte, Cleveland, and others as undesirable. This suggests that there may be other reasons for including non-data-ink in graphs.

The other side of the debate, where embellishment of data graphics is encouraged, is exemplified by the work of Nigel Holmes. Holmes states that a data graphic "must engage the reader's interest" ([7], p. 60), and he demonstrates how this can be achieved using graphic imagery. He also states that "the purpose for making a chart is to clarify or make visible the facts that otherwise would lie buried in a mass of written materials" ([7], p. 21).

Interpretation Time and Errors

The minimalist perspective argues that the inclusion of visual embellishments and correspondingly lower data-ink ratios in graphs decreases the interpretability of data. There is some support for this assertion when interpretability is defined as accuracy and speed in answering questions based on graphs. Schonlau and Peters [18] did find a small but significant decrease in accuracy when a third dimension was added to pie charts. Siegrist [19] found that 3D bar and pie charts usually take more time to evaluate, although the accuracy of those evaluations was on par with 2D versions. Gillan and Richman [5] showed that high data-ink ratios were correlated positively with faster response times and greater accuracy, although further investigation demonstrated that varying the particular location and function of the additional ink changed results. For example, background images made interpretation harder, but lines for X and Y axes improved response time. Their findings support Tufte's assertion that data-ink should be maximized, but only "within reason."

However, other research differs, finding no significant differences in interpretability between varying ratios of data to ink. Kulla-Mader [11] found no differences in interpretation errors between three ratios of data-ink; however, ratios varied only in the presence and colour of gridlines, and the existence of chart borders. Blasio and Bisantz [1] found no significant difference in interpretation time between low data-ink ratios and medium ratios in real-time monitoring tasks. They did find a significant difference between their high data-ink condition and the other conditions, but the high data-ink data was presented as a table, while other ratio conditions were charts. In addition, visual psychophysics research into graphing elements has found that graph elements with irrelevant extra dimensions are processed just as quickly, and in some cases more quickly, than simpler elements [20].

Aesthetics and Preferences

Tufte has suggested that even well-known chart designs, such as the rectangular bars of bar charts, contain redundant data. One of his redesigns suggest removing the left vertical line of the bar, leaving only an L-shaped indicator of a data point; he asserts that the left vertical bar is redundant ink. However, some researchers have suggested that moderate increases in visual embellishment such as fully-drawn graph elements (e.g. rectangular bars for values rather than L-shaped representations) or small levels of chart junk provide a valuable aesthetic improvement to the graphic. Inbar et al. [9] found that users preferred non-minimalist graphs; although, they suggested that the participant's unfamiliarity with the minimalist design may have been a confounding factor in the study. They conclude that the problem is complicated, a conclusion supported by research into 3D graphs: Levy et al. [12] found that users have a preference for using 3D graphs for showing information to others and enhancing memorability, but 2D graphs for immediate personal use.

These findings make sense within the context of Tractinsky and Meyer [21], who found that graphs were more likely to be designed with chart junk when designers had persuasion or impressiveness as an objective, particularly when the data in the graph reflected poorly on the graph designer.

Chart Memorability

Ware identifies 'memory extension' as one of the ways in which visualizations extend human cognition [23]. Charts, especially when they are used within the context of persuasion or presentation, are designed to aid in the memorability of the presented data [10], and existing research has supported this fact.

In both recognition and free-recall situations, research suggests that pictures lead to better performance than words alone for experimental learning tasks [6,17]. The impact of pictorial stimuli on written retention has been shown with young children and college students (e.g., [4,14]). A recent study examined the effects of pictorial illustrations on younger and older adults' recall of the content of short sentences, and found that when pictorial stimuli is combined with verbal stimuli, content recall was improved for both groups [2].

The emotional tone of pictures also appears to affect memory. Mather and Nesmith [13] found that participants were more likely to remember the location of positive and negative pictures than non-arousing pictures during an incidental-coding task. Since enhancement of memory occurred for both positive and negative pictures, it can be noted that arousal (rather than valence) is the critical factor.

Blasio and Bisantz [1] argue that in monitoring tasks, such as situational awareness of a control board, interpretability is paramount. Reductions in interpretation in these settings can significantly increase costs for corporations or make typical tasks much more mentally demanding. The majority of charts are generated for static media, but in these situations memorability of a graph may be as, or more, important than the interpretability; as having the static chart media on-hand when it is needed may not be possible. There have been very few studies of memorability in graphs. Kelly [10] found no differences in memorability between high and low data-ink conditions. Kelly's findings may have been confounded by the questions asked in the study, which included estimation-based components that could have biased the results.

COMPARISON OF PLAIN AND EMBELLISHED CHARTS

We designed an experiment to investigate differences in comprehension and recall between charts with two dramatically different levels of visual imagery. We maximized the differences in visual embellishment between two sets of charts while remaining as realistic as possible. For our embellished charts, we selected fourteen charts from Nigel Holmes' book *Designer's Guide to Creating Charts and Diagrams* [7] (see Figure 2). All of the charts incorporated some degree of chart junk, had relatively low data-ink ratios, and often contained images and cartoons that were woven into the presentation of the data. These charts were designed to attract the eye, engage the reader, and sometimes provide a particular value message over and above the presentation of the data itself.

For each of the Holmes charts we created a plain version of the chart (Figure 2), taking care to include exactly the same data in the chart (including type of chart, titles, axis labels, and axis values). In creating the plain charts we did not focus on specifically maximizing the data-ink ratio, but developed charts that reflected Inbar's conjecture [9] about the 'sweet spot' of chart design where high data-ink ratios and familiarity intersect to create comfortably readable graphs. As a result, the plain charts preserved axis ticks, the overall chart style (e.g. bar, line, or pie), and the positioning of axis values and labels as they existed in the original

Holmes version. However, axis lines (both for X and Y) were removed, and the axis titles and chart title were repositioned if necessary to conform to more traditional title positions. Bar and line charts were framed by a black box, and all used grayscale.

Participants and Apparatus

Twenty participants (9 male, 11 female), aged between 18 and 40 were recruited from a local university (7 graduate, 13 undergraduate). Of the 20 participants, 17 said they at least occasionally created charts and 19 said they at least occasionally read or interpreted charts.

The charts were digitized and formatted to be displayable on a computer monitor as full-screen slides, yielding 14 charts, each with a Holmes version and a plain version. A standard Windows PC was used to present the charts on a 24-inch widescreen monitor; we used a Tobii eye tracker to capture eye-gaze data.

Each slideshow presented the fourteen charts, alternating between Holmes and plain versions. Each participant saw only one version of each chart, either Holmes or plain. Presentation order for the versions was counterbalanced, but chart order was not. We treated the first two slides as training. A fixation-cross slide was displayed between each chart to provide the participants with a visual break.

Procedure

The experiment had two parts: a chart reading and description phase, and a recall phase. To prevent intentional learning, participants were not told about the recall phase – they were told that there were two parts to the study and that the first part was unrelated to the second part. Participants were assigned to one of two order conditions, with ten in each condition.

Part 1: Reading and Description Task

Participants were seated approximately 24 inches from the computer monitor, and the eye-tracker was calibrated for each participant. The experimenter instructed the participants that they would be shown a series of charts that they could examine for as long as they needed while answering four questions related to the chart. Participants were informed that they could answer the questions without prompting or ask for assistance in the form of a reminder. The answers were stated verbally to the experimenter. If at any time, the participant strayed from either the prescribed question ordering or subject matter, the experimenter would redirect the conversation to focus on the question.

The experimenter asked the four questions verbally; the list of questions is shown below, followed by extra instructions (in italics) provided if the participant did not understand. Questions represent components of high-level chart comprehension.

Q1–Subject: ‘What is the chart is about?’ *‘Tell me about the basic subject of the chart.’*

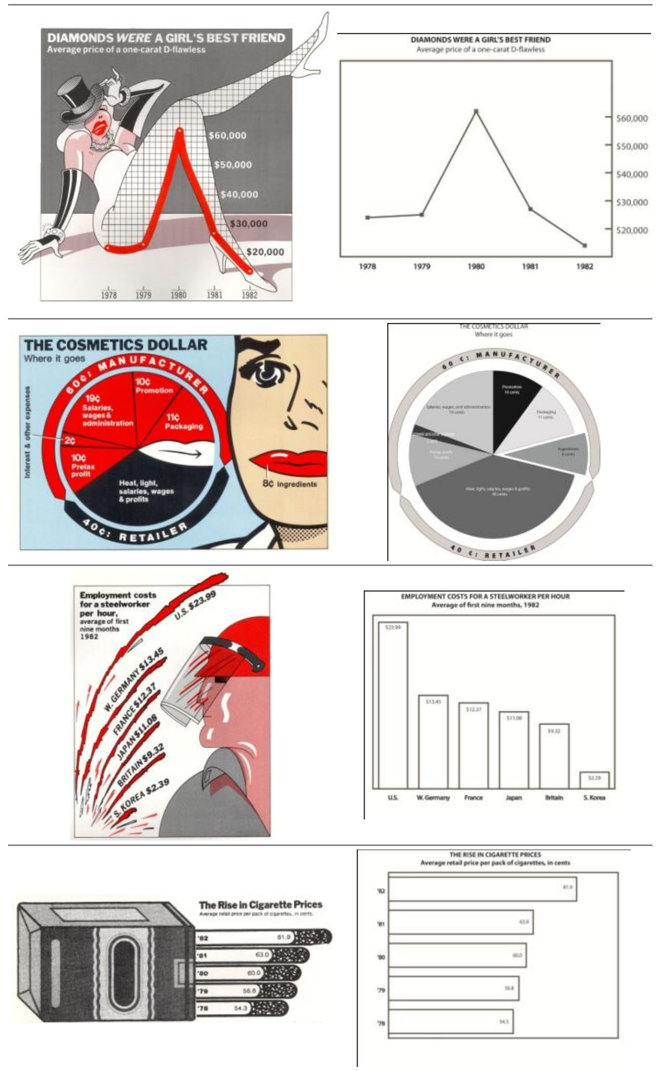

Figure 2. Example charts used in the study: Holmes versions (left) and equivalent plain versions (right).

Q2–Values: ‘What are the displayed categories and values?’ *‘Tell me how the chart is organized and any relevant values.’*

Q3–Trend: ‘What is the basic trend of the graph?’ ‘*Tell me whether the chart shows any changes and describe these changes.’* (Note that this question was not relevant for pie charts.)

Q4–Value Message: ‘Is the author trying to communicate some message through the chart?’ *‘Is the author trying to get across a specific point or is he or she merely presenting objective information?’*

If participants failed to provide a complete answer to any question, the experimenter went through a series of increasingly specific prompts until either the participant sufficiently answered the question or the list of prompts was exhausted. The prompts were as follows:

1. ‘Can you be more specific?’ This prompt was provided as generic reminder to elaborate.

2. ‘Can you be more specific about *X*?’, where *X* was a question-specific prompt. e.g. ‘the basic subject’; ‘how they’ve categorized the data’; ‘any changes the chart shows’; or ‘whether there is a message’.
3. A specific prompt directly related to the component of the participant’s answer currently missing. (E.g. ‘What were the values on the x-axis of the graph?’)

Part 2: Recall Task

After the reading and description task, participants were placed in one of two recall conditions (immediate recall or long-term recall), with ten participants in each. The long-term participants scheduled their second part of the study for two to three weeks later. Participants in both conditions had previously been informed that the second portion of the experiment was unrelated to the first and therefore remained unaware of the recall task.

Participants in the immediate recall condition played a five-minute game after describing the charts to clear their visual and linguistic memory before starting the recall task.

At the beginning of the recall task, whether in the immediate or long-term recall condition, participants were reminded of the four component questions they had been asked for each chart during the description task (subject, values, trend, and message) then instructed to recall as many of the charts as possible. The experimenter recorded any correctly recalled charts (e.g. “I remember one about the price of diamonds”). After the participant had recalled as many charts as possible, the experimenter asked the participants to describe the charts as completely as they could, based on the four component questions. The descriptive recall was performed in the order of chart recall, not experimental presentation. If at any point the participant recalled a new chart, it was added to the end of the recall list, but the participant was redirected to the current chart’s descriptive recall stage. Finally, the experimenter prompted the participant for each un-recalled chart (e.g. “Do you remember the chart called ‘Monstrous Costs’?”) and continued with a descriptive recall stage for each in turn.

Measures

We collected three types of data from the experiment: experimenter-coded response scores, eye-gaze data, and a preference questionnaire.

Response Scores

A single experimenter coded the participant responses. The same scale was used in both description and recall parts of the study. The experimenter updated a checklist as the participant described each graph. For each question in both the description and long-term recall the experimenter scored the participant’s answer in two parts: first, whether a prompt was required (scored as either “Yes” or “No”, regardless of what level of prompting was required; second, the quality of their response (independent of prompting) on the following four point scale.

3 points (all correct): The participant provided all of the relevant information in a given category.

2 points (mostly correct): The participant provided most of the relevant information in a given category, but omitted one or two important details.

1 point (mostly incorrect): The participant provided some correct information in a given category, but omitted most relevant details.

0 points (all incorrect): The participant provided no correct or relevant information in a given category.

0 points (I don’t know): The participant stated that they did not know any of the relevant information in a given category. This was scored as completely incorrect.

An example of scoring: to a participant looking at the Holmes ‘Monstrous Costs’ chart, we would ask question Q3: ‘What is the basic trend of the chart?’ If the participant responded, ‘I don’t understand,’ we would elaborate: ‘Tell me whether the chart shows any changes and describe these changes.’ The participant might answer ‘The teeth get bigger every year.’ This answer would score 1 point, as it is not a complete answer (with incorrect information about the period of the data reported) but provides at least some information that the bars increase. The experimenter would then provide additional prompts starting with ‘Can you be more specific?’ A complete answer scoring four points might be ‘The chart shows that campaign expenditures by the house increased by about 50 million dollars every two years, starting in 1972 and ending in 1982.’

Early in the data collection stage, a second researcher reviewed the primary coder’s results by independently coding a single participant through a review of the interview video. The primary coder also performed an additional coding of the same interview, based only on the video capture. All three versions, the original coding, the second coding by the primary coder, and the alternate coding by the second researcher differed only a few times and then never more than by one point.

Preference Questionnaire

After the recall portion of the experiment, participants were asked to choose their preferred chart format for a number of questions related to both the description of the charts and the recall of the charts. For participants in the immediate recall group, this questionnaire was administered on the same day as the description of the charts; for participants in the long-term recall group, this questionnaire was administered 12-22 days after the initial chart descriptions.

Gaze Data

We used the eye tracker throughout the description task to get an idea at what people looked at on a chart. For each chart, we used the Tobii ‘area of interest’ analysis software to designate each type of content that was in a chart. Chart content was coded by a single experimenter as one of the following (see Figure 3):

- Data: gazes in the region would be for primarily data
- Embellishment: gazes in the region would be for primarily non-data imagery

- Dual coded (Data and Embellishment): it could not reasonably be assumed if gazes in the region would be for data or imagery
- Other: any gaze not falling in a specified region

We removed gaze data for charts where gazes did not register correctly, which would have resulted from the participant shifting or moving out of range of the eyetracker during the study.

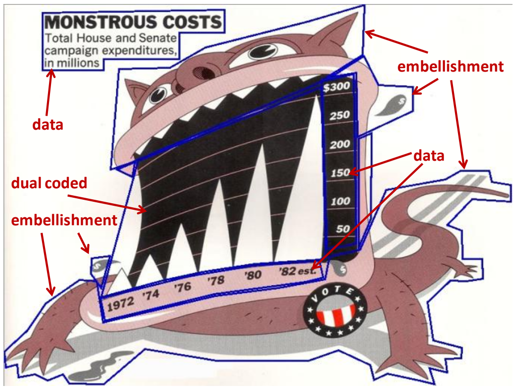

Figure 3. Area of interest analysis: gaze falling in any of the areas defined by the blue borders were labeled as data, embellishment, dual encoded, or other.

RESULTS

We first present the results related to describing the charts, followed by the recall results and the user preferences, and then analyze the participants' gaze data. In all cases we used a t-test for analysis; all of our hypotheses and questions related to a comparison of two groups (e.g. long-term vs. immediate recall). We applied a 1-tailed test where our stated hypothesis was that one group was significantly greater or less for a particular measure.

Description

Description scores were calculated from the coded responses. After removing data for the two training charts, and summing description scores over the remaining charts for each participant, we conducted pairwise t-tests to determine whether the presence of visual embellishments altered the quality of participant verbal chart descriptions. There were no differences in the scores of chart descriptions for the subject of the chart (t_{19}=0.84, p=.412), the categories in the charts (t_{19}=1.38, p=.185), or the trend of the charts (t_{19}=0.23, p=.818), depending on whether or not visual embellishments were present (see Figure 4). There was a difference in the description of a value message of the charts (t_{19}=3.37, p=.003), with better descriptions of a value message for the Holmes charts than the plain charts (see Figure 6).

We also looked at the total completion time for users in completing the description task – calculated from the total onscreen time spent with each chart. A pairwise t-test was used to determine if the presence of embellishments resulted in users spending more time to complete the description tasks. There was no difference in completion time for describing the charts (t_{19}=1.834, p=.082), depending on the presence of visual embellishments (Mean Holmes=2.60min, SD Holmes=0.67; Mean Plain=2.43min, SD Plain=0.68).

Recall

Recall scores were calculated from the coded responses. After removing data for the two training charts, and summing recall scores over the remaining charts for each participant, we conducted pairwise t-tests to determine if users were able to better recall charts that contained visual embellishments. Because our hypothesis was that visual embellishments would aid recall, we conducted 1-tailed t-tests with α=0.05.

For participants in the immediate recall group, there were no differences in recall scores for the subject of the chart (t_9=1.24, p=.124), the categories in the chart (t_9=1.21, p=.129), or the trend of the chart (t_9=0.35, p=.369) between charts that contained visual embellishments and charts that did not (see Figure 5). There was a difference in the recall score for the value message (t_9=2.24, p=.026), see Figure 6.

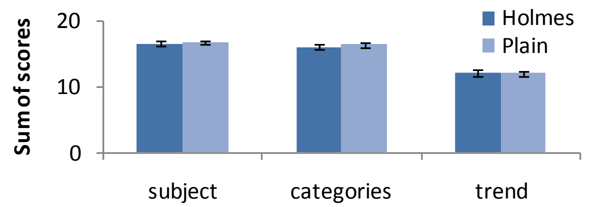

Figure 4. Means ± SE for description scores.

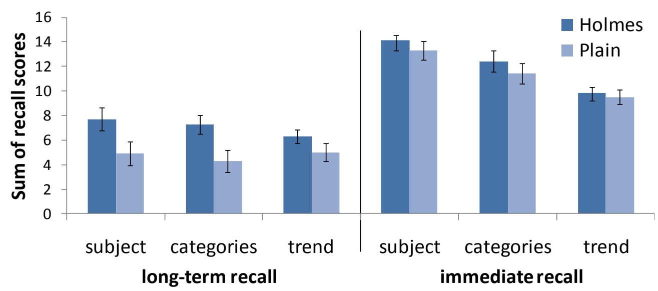

Figure 5. Means ±S E for recall scores for long-term and immediate recall.

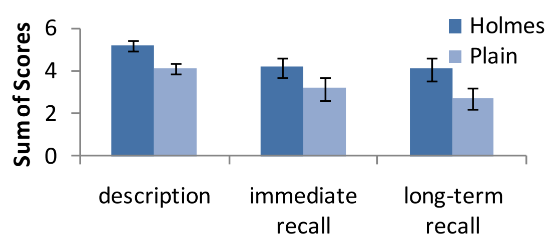

Figure 6. Means ± SE for sum of value message scores.

For participants in the long-term recall group, there were significant differences in recall scores for the subject of the chart (t_9=2.56, p=.015), the categories in the chart (t_9=5.03, p≈.000), the trend of the chart (t_9=1.95, p=.042), and the value message (t_9=2.41, p=.020) between charts with visual embellishments and ones without (see Figure 5 & Figure 6).

In addition, participants in the long-term recall group required more prompting in order to recall subjects of the charts ($t_9=2.67$, $p=.013$), categories in the charts ($t_9=2.81$, $p=.011$), and trends of the charts ($t_9=2.45$, $p=.018$) for charts that did not contain visual embellishments. However, for participants in the immediate recall group, there were no differences in the required amount of prompting to recall subjects of the charts ($t_9=1.41$, $p=.097$), categories in the charts ($t_9=0.32$, $p=.379$), or trends of the charts ($t_9=-0.32$, $p=.379$) depending on whether or not charts contained visual embellishments (see Figure 7).

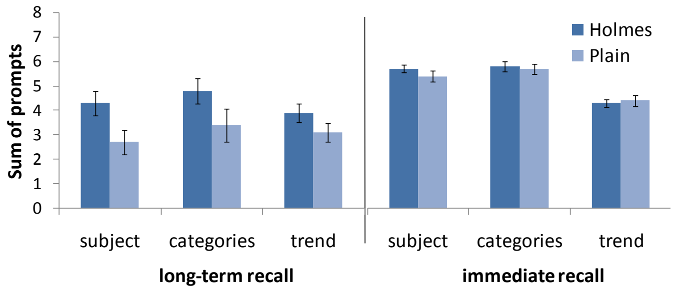

Figure 7. Means ±SE for the number of recall prompts needed for long-term and immediate recall.

User Preferences

After the recall portion of the experiment, participants were asked to choose their preferred chart format for a number of questions. Results are shown in Figure 8. Chi-squared tests showed that participants most enjoyed charts with visual embellishments ($\chi^2=8.9$, $p=.003$) and found them to be the most attractive ($\chi^2=11.8$, $p=.001$). Participants also found it easiest to remember the charts with visual embellishments ($\chi^2=15.2$, $p\approx.000$) and to remember the details of these charts ($\chi^2=8.9$, $p=.003$). Finally, participants felt that it was both fastest to describe ($\chi^2=4.3$, $p=.039$) and fastest to remember ($\chi^2=11.8$, $p=.001$) charts containing visual embellishments. The remaining questions did not produce significant differences.

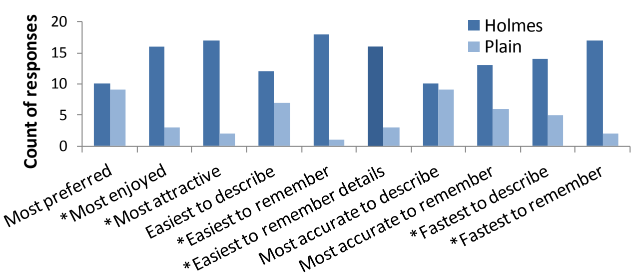

Figure 8. Count of user responses: *indicates significant difference between chart types from chi-squared test at α=0.05

Gaze Detection

During the recall portion of the experiment, we recorded users' gaze data to determine what proportion of time was spent looking at different aspects of the charts, including visual embellishments.

Holmes charts contained data, embellishments, and areas that were both data and embellishments (dual coded). In addition, areas of the chart that were not of interest (e.g., blank space) were categorized as 'other'. Plain charts only contained the data and other categories. Based on how long participants looked at the screen (sometimes they would look off screen or at the experimenter during the study), we calculated the proportion of time spent looking at the different chart elements. As Figure 9 shows, participants spent 67% of their on-screen time looking at data or data and embellishments for Holmes charts as compared to 78% for Plain charts. For Holmes charts, users spent 27% of their on-screen time looking at areas that contained both data and embellishments, and 13% of their time looking at embellishments that did not communicate any data. Although 40% of users' time was spent looking at embellishments for Holmes charts, this did not result in a longer time to describe the charts.

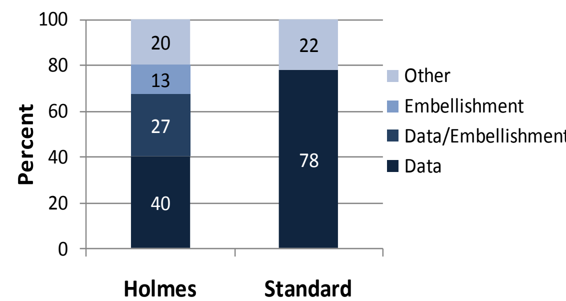

Figure 9. Percentage of on-screen time spent looking at different chart elements for Holmes and Plain charts.

DISCUSSION

The study showed five main findings:

- There was no significant difference between plain and image charts for interactive interpretation accuracy (i.e., when the charts were visible).
- There was also no significant difference in recall accuracy after a five-minute gap.
- After a long-term gap (2-3 weeks), recall of both the chart topic and the details (categories and trend) was significantly better for Holmes charts.
- Participants saw value messages in the Holmes charts significantly more often than in the plain charts.
- Participants found the Holmes charts more attractive, most enjoyed them, and found that they were easiest and fastest to remember.

In the next sections we propose explanations for these results, and then discuss several issues that these results raise for the practice of information presentation.

Why did the extra visual information in the image charts not lead to reduced interpretation and recall accuracy?

Some experts in chart design (and some previous research) suggest that the additional information in the Holmes-style image charts should lead to greater difficulty in correctly interpreting the data in the chart. This did not happen in our study; we propose some possible reasons for our results.

First, the eye-tracker data suggests that people spent similar amounts of time looking at the data portions of the charts, regardless of their style, although with the Holmes charts the data was sometimes presented through imagery (dual coded). If participants' investigations of the data elements in both conditions were equal, it is reasonable that their interpretation of that data would be equally accurate.

It is important to note that we did not constrain the amount of time participants had to examine the charts. If viewing time was limited, it is possible that people could spend less time on the data elements (assuming that they look at both data and image in an interleaved fashion).

Second, the time that people spent looking at non-data elements in the image charts may not have hindered their interpretation as much as might be supposed. This is because the images in the Holmes charts are actually tightly connected to the subject and even the details of the chart. That is, both the overall message (which includes the topic) and some of the structural elements of the chart (such as the trend) are encoded in the image. For example, the image of the monster in Figure 1 evokes "Monstrous" in the chart's title, and the shape of the mouth and teeth show the increasing trend of the bars. the time spent examining the image itself may have helped people remember elements of the chart, possibly overcoming any detrimental effects caused by the increase in non-data information.

Nigel Holmes has in fact stated that one of the reasons for images in his charts is to allow faster pickup of messages:

> I think [Tufte] missed the point of much that I was trying to do: TIME magazine charts were aimed at lay readers, not unintelligent ones, but busy ones. I knew they'd get the point quicker if they were somehow attracted to the graphic. ([8], p. 76).

Whether this comes at the cost of interpretation richness will be discussed below – in terms of topics, categories, values, and trends, our study shows that the basic elements of the chart are correctly interpreted.

Why was long-term recall better for the image charts?

There are several possible reasons why participants were able to recall the Holmes charts better than the equivalent plain charts. First, as shown in previous research (e.g., [4,6,13,14]) the presence of an image along with other content provides an additional encoding in memory which can improve recall. The images in Holmes's charts are also closely associated with the subject of the chart, reinforcing people's memory.

Second, the Holmes images were all fairly different from one another, which may have helped people remember specific charts. The plain bar charts, in contrast, all had a similar visual appearance (as do all bar charts).

Third, the user's emotional response to the imagery in embellished charts may be a hidden factor. Whether from the subjective preference for the charts or the emotions evoked and associated with the chart imagery, it could be that the emotions, in combination with the visual imagery, help to anchor chart details in a viewer's memory.

Finally, it is intriguing that participants also remembered the details of the Holmes charts better than the plain charts. This may have occurred because of how closely the image is integrated into the chart data – for example, the monster's oddly shaped mouth and teeth show the rising line of the data. Remembering the picture, therefore, may also provide access to the details of the chart (e.g., the overall trend).

Why did participants see a value message in the images?

We believe that people saw value messages in the Holmes charts more often because the images do often convey values, and this was likely the intention of the designer. The image of a monster (and the word 'monstrous' in the title) conveys clear negative connotations that are part of the story. It is difficult to imagine using monster imagery with other more positive stories, such as rising vaccination rates or rising life expectancy ("monstrous life expectancy"?).

It is possible to include visual imagery in charts that reinforces the subject of the chart, but does not provide a value judgment. Although this was not true of the images used in our study, we present example image charts with no value message in Figure 10.

Design Questions and Implications

Should imagery be used in charts to improve recall?

The success of visual imagery in our study raises the question of whether designers should be encouraged to use images in their charts. This is a potentially contentious issue, and there is no way to make a clear recommendation either in favour or against visual imagery. In addition, further study is needed to consider other aspects of this practice before drawing strong conclusions; such as different tasks. For example, it could be that a task that required detailed analysis of charts is hampered more by embellishments, rather than the high-level description task that we asked of our participants. One could imagine that in some safety-critical systems, such as those used by flight control systems, limiting the presentation to the salient information would likely be the preferred course.

Perhaps the most important lesson for designers is that there may be more to the usability and utility of charts than is currently captured by minimalist design approaches.

One practical consideration arguing against the use of strong imagery in charts is the creativity, effort, and artistic ability required to do so. Finding an accompanying graphic that fits the story (and does not bring in any unwanted connotations), and finding a way to integrate the image into the representation of data, are likely to be difficult tasks that cannot be done well without a skilled designer. In addition, some charts and some topics may not be amenable to Holmes-style visual presentations.

In the end, the purposes and traditions of the venue are likely to dictate the style of charts. As stated by Rock, publications can be oriented towards reading or viewing:

> Some things are designed for reading: scholarly journals, literary reviews, financial pages, and their

ilk are fairly impenetrable to the casual page flipper. Other objects like USA Today, annual reports, fashion magazines, and so on are for looking. … Then there are the gray areas. These include newsmagazines and textbooks, which imply reading but are increasingly about looking. ([16] p. 84).

It is interesting to consider where the CHI proceedings fits into this continuum of reading and looking - we note that we shied away from presenting our experimental data in Holmes style (although partly from lack of artistic ability).

In our study we intentionally chose the most extreme type of visual embellishment that we could – namely, the full cartoon imagery used by Holmes. However, as stated earlier, there are several other types of visual embellishment that are less extreme, and it is possible that these other types will also help people to remember charts, without requiring such extensive imagery.

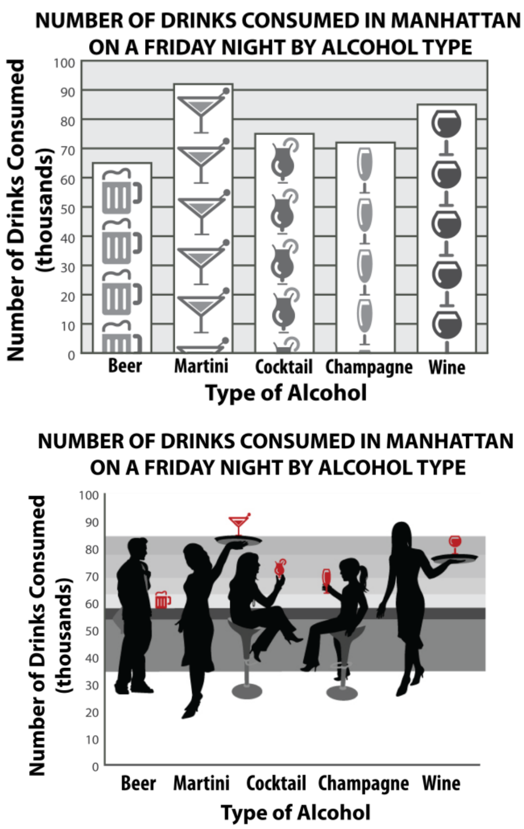

Figure 10. Less extreme visual imagery in charts.

Based on our results, we believe that there will be a difference between embellishments that provide images and those that do not. Psychology research suggests that pictures are useful in terms of memory and recall [4,6,14,17], and it is likely that a larger effect would be seen with charts that use pictures rather than non-image embellishments such as textures, colours, or abstract shapes. Figure 10 shows examples of possible designs that use imagery in less extreme ways.

The wider problem of bias in charts

Charts that include strong imagery have been criticized for introducing bias into the presentation, but it is worth noting that this is only a more explicit attempt to persuade than some of the other rhetorical devices that are used in written and visual communication. There is a tacit assumption in the minimalist approach that by removing embellishment and non-data ink, the presentation is made more objective, and the data 'is allowed to speak for itself.' However, despite their different visual style, the two charts in Figure 1 both tell the same story – that campaign expenditures are rising at a higher-than-linear rate – and both characterize this trend as 'monstrous.' There is therefore no guarantee that minimalist charts are free from bias.

Communication is, in most cases, oriented towards making an argument, and few if any pieces of writing are truly objective. There are many ways that even minimalist charts can be biased or used in support of a particular story – the type of chart, the selection of data, the ordering of bars, the scale of axes, the choice of titles and labels – all can be used to highlight or hide different aspects of the data.

Despite these potential adjustments behind the scenes, minimal charts still maintain the appearance of objectivity – but claiming objectivity is also a common rhetorical technique. As Rock [16] states,

> …images and charts seem to not imply an inherent point-of-view. They radiate a kind of false objectivity because the concept of the image-as-opinion is difficult for most people to grasp. (p. 86)

(Note that here 'image' implies the visual of even a minimalist chart, rather than the image of a Holmes chart).

The illusion of objectivity (as used in minimalist charts) and the use of evocative imagery (as used in Holmes charts) are perhaps just different approaches that work at different ends of the rhetorical spectrum. Designers and readers should remember that a Holmes chart is not necessarily more biased than its plain counterpart – but it may be more effective at conveying the value message that is part of the overall argument.

Stepping back another way, it can be noted that the use of charts altogether has been seen as a practice that takes away from the more expansive, judicious, and nuanced arguments that can be made through writing:

> Charts and diagrams are certainly useful for offering general, relational explications of an issue but they necessarily shave away the ambiguous, nuanced, or obscure aspects of any idea. The information has been preprocessed, prechewed; it can only lead to one conclusion. ([16], p. 85)

Although few practitioners would argue for dispensing with all charts, it is important for designers and readers to understand the wide variety of ways in which charts are used in support of stories and arguments, and to see chart imagery as just one point in this space.

CONCLUSION

Many experts suggest that visual embellishments and 'chart junk' cause interpretation problems and should be removed from information charts, but many examples exist of embellished charts that seem to be successful. To investigate some of the specific costs and benefits of these two approaches to chart design (minimalist and embellished), we compared interpretation accuracy and long-term recall for plain and 'Holmes-style' charts. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three week gap was significantly better. In addition, participants preferred the embellished charts. Although we are cautious about proposing specific design recommendations, it seems clear that there is more to be learned about the effects of different types of visual embellishment in charts. Our results question some of the premises of the minimalist approach to chart design, and raise issues for designers about how charts are designed and used in different publications and different contexts.

ACKNOWLEDGEMENTS

We thank Nigel Holmes for granting us permission to use his images, and the CHI reviewers for their comments. This research was supported by NSERC.

REFERENCES

1. Blasio, A.J. and Bisantz, A.M. A comparison of the effects of data-ink ratio on performance with dynamic displays in a monitoring task. *J. Industrial Ergonomics 30* (2002), 89-101.
2. Cherry, K.E., Dokey, D.K., Reese, C.M., and Brigman, S. Pictorial illustrations enhance memory for sentences in younger and older adults. *Experimental Aging Research 29* (2003), 353-370.
3. Cleveland, W.S. *The Elements of Graphing Data*. Wadsworth, Monterey, CA, USA, 1985.
4. Gambrell, L.B. and Jawitz, P.B. Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly* 28 (1993), 265-276.
5. Gillan, D.J. and Richman, E.H. Minimalism and the Syntax of Graphs. *Human Factors 36*, 4 (1994), 619-644.
6. Hockley, W.E. The picture superiority effect in associative recognition. *Memory and Cognition 36* (2009), 1351-1359.
7. Holmes, N. *Designer's Guide to Creating Charts and Diagrams*, Watson-Guptill Publications, 1984.
8. Holmes, N., *Nigel Holmes on Information Design*, Jorge Pinto Books, New York, 2006.
9. Inbar, O., Tractinsky, N., and Meyer J. Minimalism in Information Visualization - Attitudes towards Maximizing the Data-Ink Ratio. *Proc. ECCE*, ACM Press (2007), 185-188.
10. Kelly, J.D. The Data-Ink Ratio and Accuracy of Newspaper Graphs. *Journalism Quarterly 66*, 3 (1989), 632-639.
11. Kulla-Mader, J. *Graphs via Ink: Understanding How the Amount of Non-data-ink in a Graph Affects Perception and Learning*. Master's Thesis, Dept. of Information and Library Science, UNC, 2007.
12. Levy, E., Zacks, J., Tversky, B., Schiano, D. Gratuitous Graphics? Putting Preferences in Perspective. Proc. *CHI 1996*, ACM Press (1996), 42-49.
13. Mather, M. and Nesmith, K. Arousal-enhanced location memory for pictures. *J. Memory and Language 58*, (2008), 449-464.
14. McDaniel, M. A., and Waddill, P. J. The mnemonic benefit of pictures in text: Selective enrichment for differentially skilled readers. *Advances in Psychology 108*, North-Holland/Elsevier (1994), 165-181.
15. Pressley, M. and Miller, G.E. Effects of illustrations on children's listening comprehension and oral prose memory. *J. Research in Reading 8*, 1 (1985), 45-56.
16. Rock, M., Since When Did 'USA Today' Become the National Design Ideal? *I.D. Magazine*, March/April 1992, 84-86.
17. Sampson, J.R. Free recall of verbal and non-verbal stimuli. *Memory and Cognition* 22, (1970), 215-221.
18. Schonlau, M. and Peters, E. Graph Comprehension: An Experiment in Displaying Data as Bar Charts, Pie Charts and Tables with and without the Gratuitous 3rd Dimension, RAND Working Paper WR-618, (2008).
19. Siegrist, M. The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour and Information Technology 15*, 2 (1996), 96-100.
20. Spence, I. Visual Psychophysics of Simple Graphical Elements. *JEP: Human Perception and Performance 16*, 4 (1990), 683-692.
21. Tractinsky, N. and Meyer, J. Chartjunk or goldgraph? Effects of persenataion objectives and content desirability on information presentation. *MIS Q. 23*, 3 (1999), 397-420.
22. Tufte, E. R. *The Visual Display of Quantitative Information*. Graphics Press, 1983.
23. Ware, C. *Information visualization: perception for design*. Morgan Kaufmann, San Fransisco, 2000.
24. Zacks, J., Levy, E., Tversky, B., and Schiano, D. Graphs in Print. In M.Anderson, B. Meyer, and P. Olivier (Eds.), *Diagrammatic Representation and Reasoning*, London: Springer-Verlag, (2002).