ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs

Steve Haroz
Northwestern University
isotype@steveharoz.com

Robert Kosara
Tableau Research
rkosara@tableau.com

Steven L. Franconeri
Northwestern University
franconeri@northwestern.edu

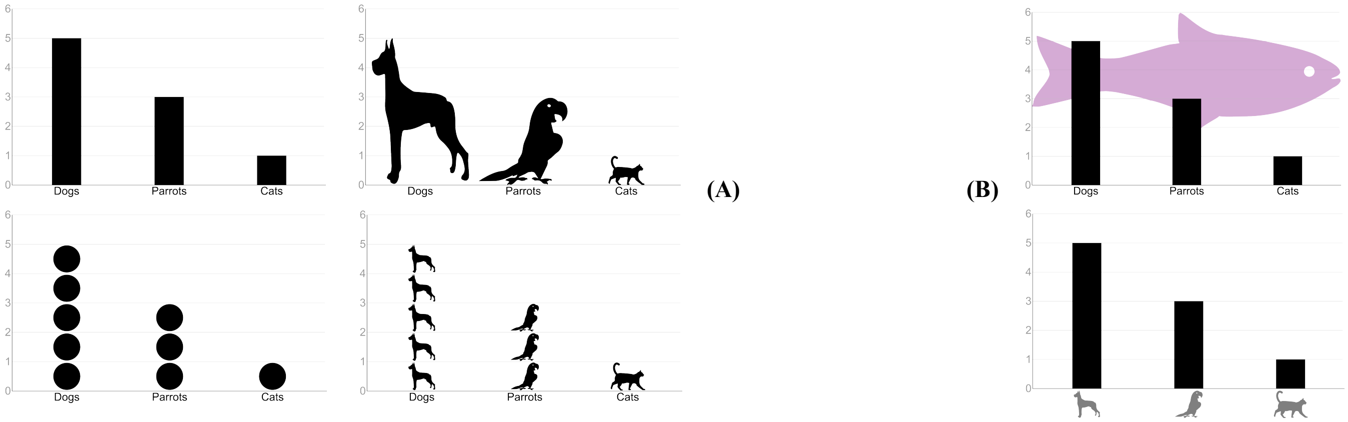

Fig. 1. Pictographic charts have been used for decades. (A) Which chart above most effectively conveys information? Which data is easiest to remember during a demanding task? Which is most engaging? (B) How integrated must a pictograph be to benefit the user? Do purely decorative background images offer the same benefits as simple axis labels? Or must they be used to convey data?

ABSTRACT
Although the infographic and design communities have used simple pictographic representations for decades, it is still unclear whether they can make visualizations more effective. Using simple charts, we tested how pictographic representations impact (1) memory for information just viewed, as well as under the load of additional information, (2) speed of finding information, and (3) engagement and preference in seeking out these visualizations. We find that superfluous images can distract. But we find no user costs – and some intriguing benefits – when pictographs are used to represent the data.

Author Keywords
Visualization; Psychophysics; Working Memory; User Performance; Pictograph; Embellishment; ISOTYPE

INTRODUCTION
The International System Of TYpographic Picture Education (ISOTYPE) uses simple pictographic elements to convey many types of information, including numerical data. Otto and Marie Neurath defined the term in the 1920s [21], though this type of chart was first described by Willard Brinton in 1914 [6]. Together with Gerd Arntz, the Neuraths created many ISOTYPE designs over several decades [1].

The goal was a universally understandable system for communicating quantities of commercial, social, or economic information (e.g., automobile production or number of children born per year). Symbols, each representing a fixed quantity, were stacked to provide an intuitive representation of a total amount (Fig. 2). Gerd Arntz's pictographs – simplified icons with minimal color – are highly recognizable and are still used in signs, traffic icons, and warning labels.

While the design community has largely embraced the simple style of ISOTYPE for pictographic embellishments [7, 17], the visualization and HCI communities tend to regard pictographs as 'chart junk' – a distraction from the data itself [24]. Here we examine how ISOTYPE-style embellishment affects viewer memory, speed, and engagement within simple visualizations.

Recent work suggests that extraneous pictographic information can indeed improve the effectiveness of visualizations. Bateman et al. found that visualizations that integrate data with illustrations yielded better memory of the data compared to minimalistic outlines of bar charts [2]. Borkin et al. found a related result – that people can better recall having seen a visualization that includes pictures [4], though it is not clear if they would better recall the data, per se. When images and clip-art are embedded in the visualization's data representation, Borgo et al. found occasional impact to work-

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702275

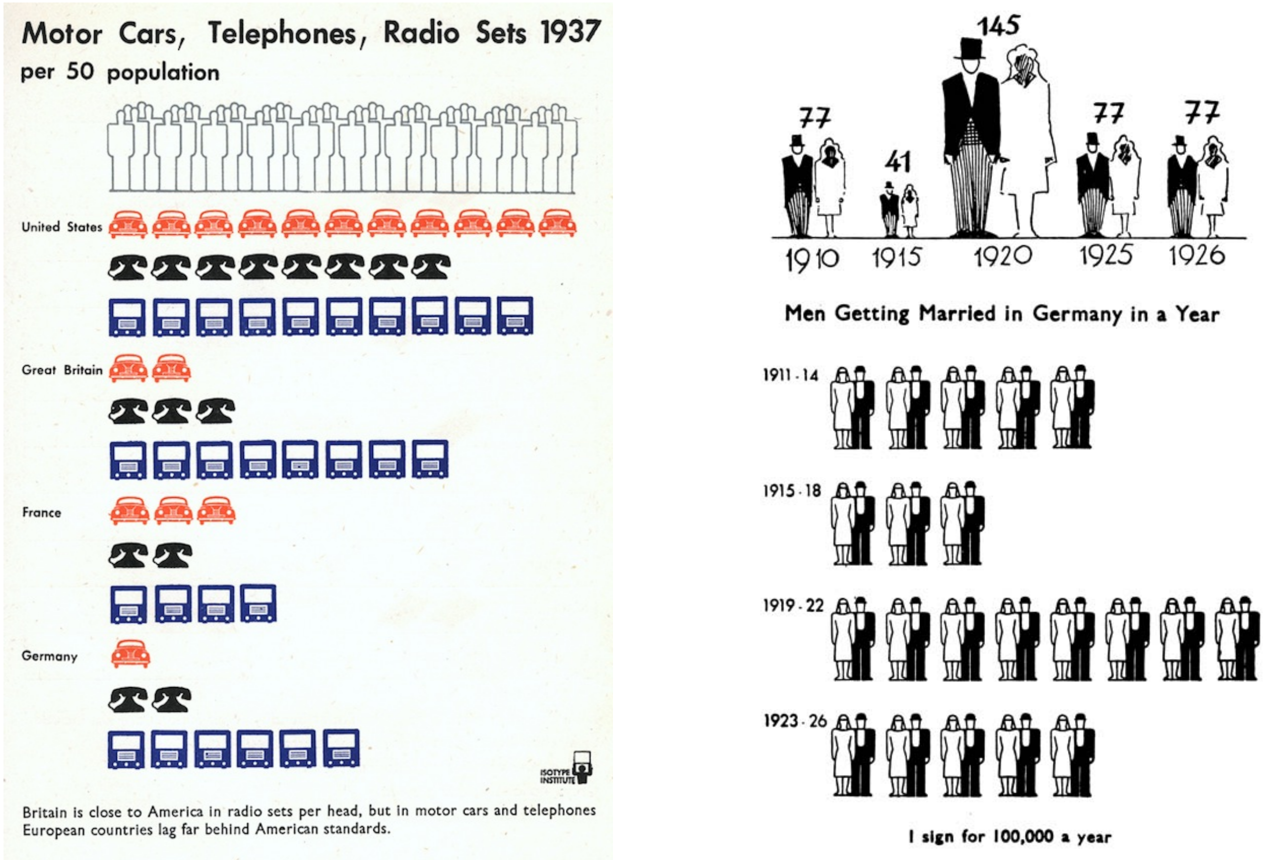

Fig. 2. Both of these example visualizations were made by Otto Neurath, a proponent of using arrays of simple pictographs to present quantitative information. The left image – published in 1937 – uses rows of pictographs to visualize the number of cars, phones, and radios in different countries. The right image – published in 1936 – shows two visualizations of the same data. Neurath insisted that stretching one pictograph (top) was inferior to stacking multiple small pictographs (bottom) [21].

ing and long term memory performance for some visualizations [3]. Other work shows that embellishing data with colors that are semantically consistent (e.g., blue for data about "oceans") can increase the speed of finding information in a visualization [19].

While past empirical research has shown that pictorial information can be beneficial, it is unclear why and when these advantages occur. In particular, we explore whether, and how, the types of pictorial cues designed by Brinton, the Neuraths, and Arntz might create more effective visualizations. Although past work has studied visualization embellishments that contain varied degrees of color [2, 3], we omit color cues and instead focus on pictorial cues carried by an object's shape

We tease apart several properties of ISOTYPE-style pictographs in our experiments in order to identify which aspects lead to costs or benefits for visualizations. For example, pictographs add identifiable symbols, which might increase memorability or engagement. Do these pictures need to be a part of the data, as in the examples in Fig. 2, or do they carry the same advantage when they merely accompany a chart in the background or as a label (Fig. 1B)? If the symbols are a part of the data, does it matter if the data value is conveyed by the number of stacked symbols, as opposed to the stretched height of a symbol, which Neurath insisted was an inferior cue (see Fig. 2, right)? If the design choice to stack or stretch matters, is that effect specific to pictographs, or would the same pattern emerge for simple non-symbolic shapes, such as a stack of circles vs. a single stretched bar (Fig. 1A, left column)?

We explore these questions across a range of sample measures: memory for previously glanced information, memory under the load of seeing intervening visualizations, speed of data extraction, and the level of observer engagement in a visualization.

CHART VARIATIONS

We manipulate several aspects of the charts independently across experiments, using the classic unembellished bar chart (Fig. 1A top-left) as a baseline.

Variation: Pictographs vs. Simple Shapes

ISOTYPE charts typically rely on pictographic representations of the real objects referred to by the data, instead of simple shapes like the rectangles of a bar chart. To test the potential costs or benefits of pictorial information, we use graphs with either ISOTYPE-style pictographs or simple shapes (see the differences across the columns of Fig. 1A).

Pictographic representation may help overcome limits on working memory, a primary bottleneck in human reasoning [9]. Pictures provide multiple cues for encoding and retrieval of memories, providing a richer set of contextual 'hooks' that allow for broader and deeper encoding of data in memory. For example, when a group of people were asked to memorize a list of words under water (while wearing scuba gear), their recall performance for the words was better when they were tested in the water, reactivating a rich set of associations between the memorized words and the sights, sounds, and emotions of their environment – compared to recalling them on land where these associations were absent [13]. Likewise, the imagery of pictographs may provide richer encoding cues, so that recalling the shape of the icon leads to more associations with the data, compared to recalling a text-based label.

We created the pictograph stimuli by making 43 sets of object categories. Each set contained four image-word pairs, chosen to reflect a category such as pets, desserts, clothes, instruments, or vehicles. All pictographs were black and white SVG files that allow for scaling (available in the demo).

Variation: Stretched vs Stacked

ISOTYPE charts typically rely on stacks of individual items to represent values (Fig. 1A bottom-right), instead of stretching the continuous extent of a single item. Because such discretizing of the values may itself impact performance, we independently manipulate this factor across both pictographs (Fig. 1A bottom-right) and simple shapes (Fig. 1A bottom-left). We also include a pictographic representation that is stretched vertically instead of being stacked (Fig. 1 top-right). We proportionally stretch the object along the x-axis to prevent distortion, but x-axis stretching is limited to prevent objects from overlapping horizontally.

Compared to stretching, stacking has the potential advantage of presenting information in a dual format – both as a height and as a number of objects. While the discrimination precision for size is typically better than for number [11], redundant encoding of both dimensions may be beneficial. In addition, for small numbers (1-4), the visual system can employ a faster and more precise mechanism for number discrimination, compared to larger ranges [8, 18]. Reliance on this more efficient system may also free up additional memory capacity for the depicted information [12]. Accuracy should therefore improve for small numbers of pictographs, even if each pictograph represents more than a single value (e.g., in Fig. 2, right, one couple represents 100,000 marriages).

Variation: Axis Labels

The use of pictographs might impact the effectiveness of a visualization because it allows a visual association for the label of the data, or more specifically, because the picture itself is used as the glyph. Would pictographic symbols have the same impact whether they are used as legend markers, allowing a visual association with the data label, or whether they are used as the markers that depict the data (either stacked or stretched)? Experiment 1 includes a condition where X-axis labels are replaced with pictographs (e.g., Fig. 1B bottom). For charts that already incorporate pictographs (stacked or stretched), the X-axis was left blank to prevent subjects from confusing the label for part of the data.

Variation: Superfluous Imagery

Perhaps pictographs impact performance merely because they affect an observer's level of engagement. While the 'axis label' variation partially distances the pictographs from the data, we also test a 'superfluous' variation that completely separates the pictograph from the data. One past study found an improved ability to recall having seen a visualization when it includes any form of picture in the image [4], but recalling the pictorial information is not the same as recalling the data behind it. Experiments 1 and 4 include the condition of a simple bar graph with a background image of the same category as the data but irrelevant to the dataset (Fig. 1B top).

EXPERIMENT 1: WORKING MEMORY CAPACITY LIMITS

Our working memory is severely limited [5, 12] and can form the bottleneck of our reasoning abilities. Visualizations that place high demand on such limited resources can be substantially more challenging to analyze [15]. Here we test whether our pictograph variations impact working memory performance.

Exp 1: Methods

Twenty-two undergraduate students (12 women) participated in this experiment in return for credit in an introductory psychology course.

In each trial, subjects viewed a chart with 3 values (each randomly selected from 1 to 5) for 1.5 seconds. The chart was then replaced with a 'response' screen that asked the subject to recall each of the values in the chart (Fig. 3). In order to test memory for the values for each category outside of the context of the graph itself, the values were queried in a random order that did not necessarily match the left-to-right ordering of the categories in the graph. For example, a chart with A=2, B=5, and C=3 may be followed by a response screen asking for B, C, and A.

The experiment was split into ten blocks – two sets for each of the five chart types specified below. One of each pair of similar blocks was in the first half and second half of the experiment, so we could analyze any fatigue or learning over a 40 minute duration (we found no change over time). Each block began with a brief reminder of the instructions and included an example chart corresponding to the block's chart type. Each block included trials with every combination of three conditions:

- *5 Charts*:
  2 Shape versus Pictograph (Fig. 1A left-right) ×
  2 Stretched versus Stacked (Fig. 1A top-bottom) +
  1 Stretched bar with a background (Fig. 1B top)
- *2 Axis styles*: whether the x-axis labels are pictographs or text (Fig. 1B bottom)
- *2 Response styles*: To avoid a potential confound of the text axis style matching the response text, we also vary

whether the response screen uses pictographs or text to identify the input fields.

The blocks were randomly ordered, as were the trials in each. Each trial was then randomly assigned one of 42 data categories. Once a category was used, it would not be reused for at least the next 10 trials. A 43rd category was reserved to show a sample chart before each block, so subjects would know what to expect.

The experiment included 200 trials (2 block repetitions × 5 chart types × 2 axis styles × 2 response styles × 5 repetitions). The first 5 trials were considered training and thus discounted in the analysis.

Exp 1: Analysis

We compute error level by averaging the absolute error for each value in a chart. So if a chart's values are 2, 3, 4, and a subject inputs 2, 5, 5, the average error for each value is (0+2+1) / 3 = 1. Each trial therefore yields a per-value error. For each result, we computed an average per subject per condition and computed an ANOVA between these subject means followed by a Tukey HSD correction. Error bars in all figures throughout this paper (except for Fig. 4) show the standard error between subject means. All line charts show linear fits with standard error ribbons.

Compensating for Individual Differences

Fig. 4 shows the substantial differences in memory accuracy across subjects for the unembellished simple bar graph. Some had nearly perfect recall, while others were only slightly better than an optimal guess (the mean of the Y-axis range). To factor out these differences for experiments 1-4, we normalized each subject's results by their performance on the simple bar condition. That is, all subject errors are individually scaled such that simple bar charts have an error of 1.0, and other values are costs or benefits relative to that baseline.

Chart type analysis

For the chart type analysis, we omit the superfluous condition (see the Superfluous pictograph section), allowing us to split the four remaining chart types into a 2x2 combination (stretched vs. stacked; shape vs. pictograph). The result is a more diagnostic 4-way repeated-measures 2x2x2x2 ANOVA with chart segmentation type (stretched or stacked), chart depiction type (pictographic or shape), axis style (pictographic or text), and response style (pictographic or text) as factors. As there were no interactions among any of these factors (all F<1), the following section examines the main effect of each factor.

Exp 1: Results

Pictographs and stacking

As Fig. 5 shows, using pictographs instead of simple shapes led to no difference in error rates ($F[1, 21]<1$, $\eta_p^2=0.002$). But using stacked items was associated with reduced error relative to stretched items ($F[1, 21]=10.0$, $p < 0.005$, $\eta_p^2=0.06$).

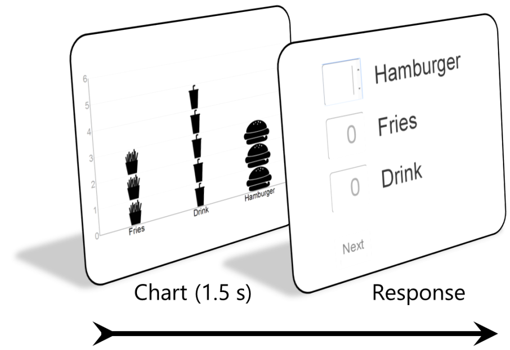

Fig. 3. In experiment 1, a chart is briefly shown then hidden. The response screen then appears where the subject enters the values. Notice that the order of the items is different in the response screen.

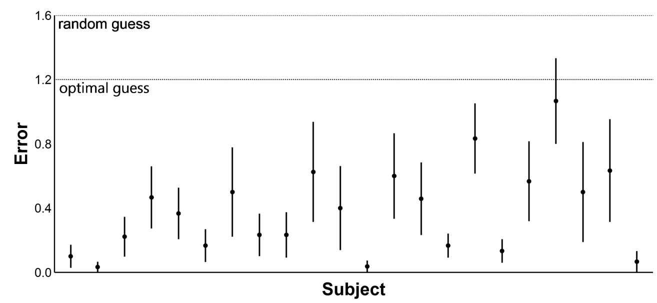

Fig. 4. Each subject's average error (difference from the correct values) for the simple bar graph condition with standard error (SE). Individual differences were large.

We were surprised at the lack of an impact from pictographs, having assumed that the rich perceptual encoding and engaging nature of pictures would be at least partly responsible for any potential advantages. The striking benefit of stacking may be explained by the cognitive advantages for small collections – the visual system has specialized mechanisms for enumerating small discrete quantities (4-5 or fewer objects [8]), and the redundant encoding of length and number may have proven beneficial to memory. Dividing length-defined objects into countable regions may have a benefit, but we predict that this benefit should only arise when stacks are smaller than 4-5 objects. The next experiment will test this prediction by introducing more numerous stacks of objects.

Axis and question pictographs

As Fig. 6 shows, using pictographs as X-axis labels resulted in more error for every type of chart ($F[1, 21]=12$, $p < 0.001$, $\eta_p^2=0.07$). Text labels may be recognized more quickly, as a lifetime of reading renders recognition of single words a surprisingly fast and automatic process. Alternatively, text labels may have led to better memory links between the label and the data value. Intriguingly, this advantage held regardless of whether the question was asked using text or pictograph labels, which showed no difference in performance ($F[1, 21]<1$, $\eta_p^2=0.0003$). There was also no reliable impact on error from the interaction of axis and question pictographs

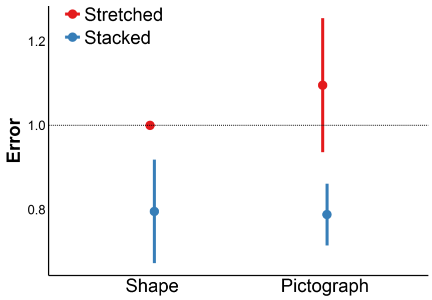

Fig. 5. Error levels are scaled relative to the error level of the simple bar chart, so it has 1.0 error for all subjects (top-left point). Charts with stacked items (blue) produced less error than those with stretched items (red). The use of simple shapes (left) or pictographs (right) had little impact.

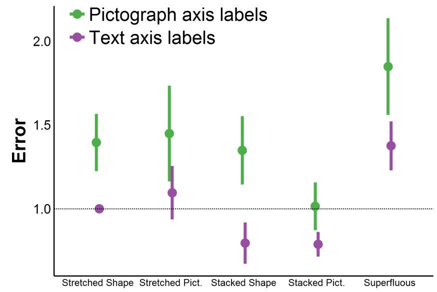

Fig. 6. For every type of chart, axes with text labels (purple) led to less error than those with pictograph labels (green).

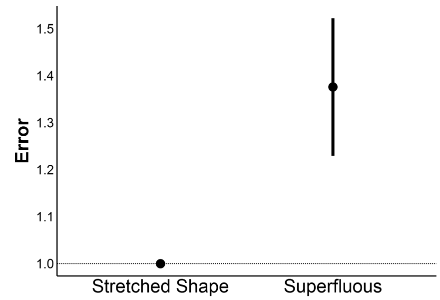

Fig. 7. An otherwise simple bar chart with a superfluous background image yields much higher error levels than an unembellished simple bar chart.

(F[1, 21]<1, η_p^2=0.002). We do note, however, that these labels were succinct and easy to read, which could have played a role in this result.

Superfluous pictograph

We also ran a 3-way repeated-measures 5x2x2 ANOVA with all five chart types, axis style (pictographic or text), and response styles (pictographic or text) as factors. There was a significant main effect of chart type (F[4, 21]=9.0, $p < 0.001$, ηp^2=0.08), driven primarily by the 45% larger error in the 'superfluous' chart type condition, relative to the mean of the other 4 conditions, (t[138]=3.39, $p < 0.001$).

As Fig. 7 shows, the addition of a superfluous pictograph – an embellishment that did not encode any data – dramatically increased recall error relative to a simple unembellished bar chart.

Left to right

Memory for the left bar was substantially more precise than the bar on the right (F[1, 21]=15, $p < 0.001$, η_p^2=0.19). Looking only at the simple bar chart trials, we scaled each subject's performance on the middle and right bar based on their error rate for the leftmost bar. The linear regression in Fig. 8 (left) shows that error level increases from left to right, suggesting that subjects are sequentially inspecting bars from left to right [10, 20] rather than simultaneously perceiving and memorizing all of them [14].

Maximum values

Because the values in the data were randomly selected, not all charts had a maximum value of 5. Some had a lower maximum. Smaller ranges may allow more precise encoding in memory. To determine if subjects were using the entire range of the axis versus the range of the data, we analyzed how the maximum data value correlated with the amount of error.

A linear regression graphed in Fig. 8 (right) shows that as the range of values increases, the amount of error increases (F[1, 21]=41, $p < 0.0001$, η_p^2=0.33). Therefore, recall from working memory becomes less accurate in proportion to the

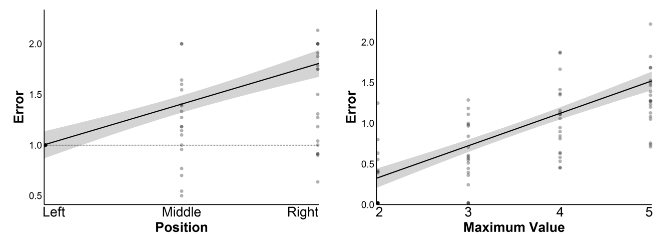

Fig. 8. (Left) Subjects strongly prioritized the leftmost values. The line is a linear fit, and the shaded region is the SE of the fit. (Right) As the maximum value – not the Y-axis range – increased, subjects made increasingly more error.

value range. The next section explores whether this effect of range on error exists at larger scales.

EXPERIMENT 2: LARGER RANGES

The visual system can quickly and precisely encode small quantities of items up to a maximum of about 4-5, an ability known as *subitizing* [8, 15, 18]. As the number of items increases beyond that range, the visual system is forced to shift to either slow counting or noisy estimation. This noisy estimation of large collections has slightly worse precision than length estimation [11].

If the improvements in working memory performance in the stacking conditions of experiment 1 are due to this advantage for number processing in small sets of items, then when larger stacks (more than 5) are displayed, performance in the stacked condition should no longer trump performance in the stretched condition, which relies on length judgment.

Exp 2: Methods

The procedure for this experiment is identical to the first experiment. The primary difference is that while the first experiment only used the range 1-5, this experiment is split into three blocks, which use the ranges 1-5, 2-10, and 3-15. The sizes of the charts on the screen were identical, only the Y-axis scale changed. We did not use a range higher than 15 because fitting more pictographs in a limited space made them difficult to discern.

We excluded the axis and question pictograph conditions, leaving all labels as text. We also excluded the superfluous pictograph condition. The experiment had 144 trials (3 ranges × 2 stacked vs stretched × 2 shape vs pictograph × 12 repetitions) blocked by range and chart type. The first 5 trials for each block were considered training and were discounted.

Due to a limited availability of undergraduate students at the time, this study was run as a web application via Amazon Mechanical Turk [16] with 30 subjects, all from the USA. It took an average of 35 minutes and paid 8 US dollars.

Exp 2: Results

We again found substantial individual differences, so we normalized the error levels by scaling all error values by the amount of error in the simple bar chart condition with a range of 1 to 5. Therefore all results represent an increase or decrease in error relative to that condition. For each result, we computed a mean per subject per condition and performed a Tukey HSD-corrected ANOVA between these subject means. There was no 3-way interaction between the three factors ($F[2, 29]<1$, $\eta_p^2=0.002$), and unless reported below, there were no significant 2-way interactions between factors.

Replicating experiment 1

The leftmost panel in Fig. 9 shows the amount of error for the 1-5 condition. This condition was very similar to experiment 1, and despite not being run in the lab, it produced very similar results. Again, stacking has a reliably large improvement on performance ($F[1, 29]=25$, $p < 0.001$, $\eta_p^2=0.22$). This time we found a small trend for a relative impairment for pictographs ($F[1, 29]=3.2$, $p > 0.05$, $\eta_p^2=0.05$), though it was tiny in comparison to the differences in error associated with different number ranges.

Smaller vs Larger Ranges

The panels in Fig. 9 show error levels for different Y-axis ranges. Although the 1–5 range shows a clear separation between stretched charts and stacked charts, an interaction between stacking and range results in the benefits of stacking disappearing with higher values ($F[2, 29]=4.5$, $p < 0.05$, $\eta_p^2=0.29$).

Maximum value

As with the previous experiment, the maximum value can be lower than the peak of the range. A linear regression of the error as a function of the maximum value (Fig. 10) showed that the maximum was correlated with error rate independent of the Y-axis scale ($F[1, 29]=470$, $p < 0.0001$, $\eta_p^2=0.4$). The lack of an interaction between the maximum and the axis range ($F[1, 29]<1$, $\eta_p^2<0.01$) supports the previous experiment's finding that people adjust their representation of scale based on the actual range of values, instead of the potential range according to the scale of the axis.

EXPERIMENT 3: MEMORY UNDER LOAD

The previous experiments tested memory for briefly glanced information, with an immediate test. Such tests might encourage verbalization strategies, such as repeating the legend names and associated numbers that are less likely to be used in the real world.

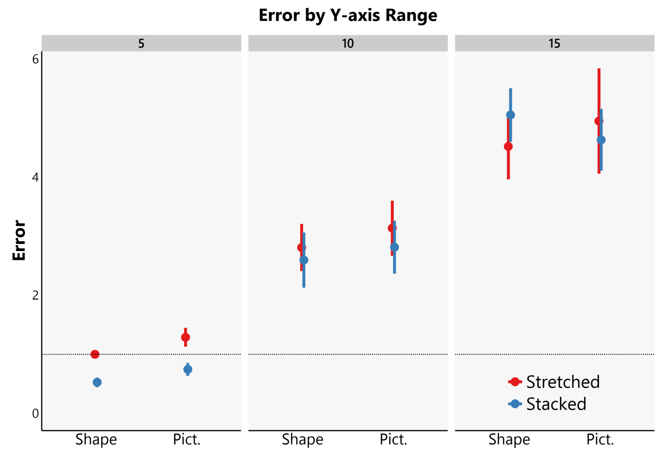

Fig. 9. Each panel shows a different Y-axis range. The amount of error for each subject is scaled by their error in the simple bar chart condition with a range of 1-5. The difference in error between the stretched (red) and stacked (blue) conditions is clear in the smaller range (left) but not for the larger ranges.

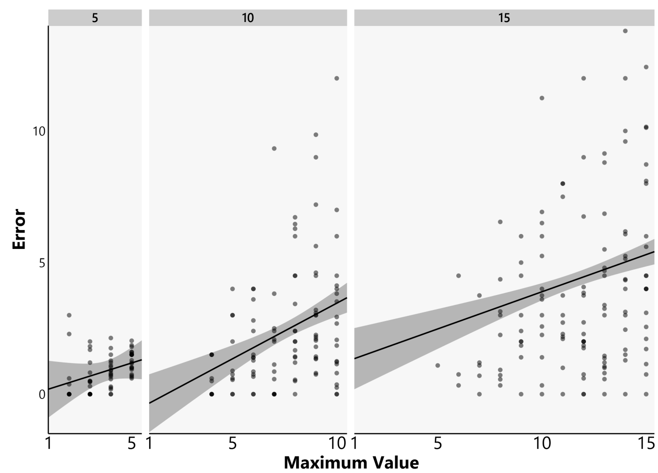

Fig. 10. The similar slopes for all Y-axis ranges show that, like in experiment 1, performance is correlated with the range of values, not just the scale of the Y-axis.

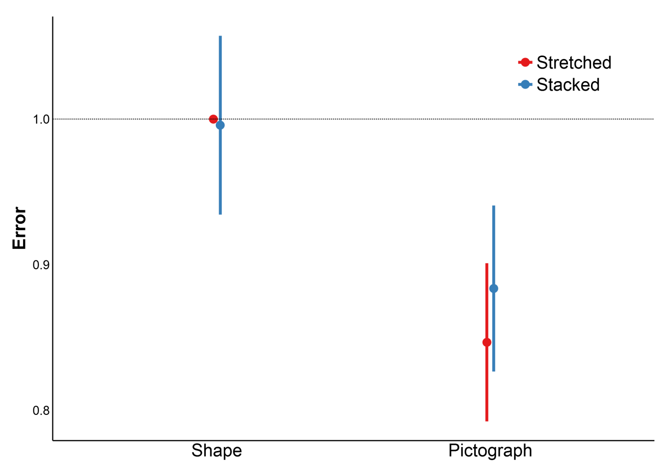

Fig. 11. In exp. 3 (1-back), pictograph charts have less error.

But a single visualization does not typically have exclusive access to working memory – other data and tasks intervene between encoding and recall. An observer might compare two visualizations, and readers of a news article might keep information from the text in mind as they examine a figure. Such situations may be less likely to encourage verbal coding of data, so that viewers rely more strongly on visual encoding of information.

Experiment 3 tests how pictographs affect memory when memory is more crowded. When memory is crowded, similar information, such as a verbal code for one set of numbers and verbal code for a second set of numbers, interferes and becomes noisy or lost [12]. Pictures may help keep information separated by expanding the information encoding space to include associations with identities and shapes that do not mutually interfere (e.g., the stack of dog icons and the stack of parrot icons in Fig 1). Research on picture memory is consistent with their support for rich and robust encoding – photographs can be recognized even after long time periods and hundreds of other viewed photographs [23], and similar recognition advantages occur for visualizations that contain rich pictorial information [4]. Therefore, experiment 3 tests whether pictographs will present an advantage over simple shapes when memory is taxed by the requirement to remember the data from an additional intervening visualization.

Exp 3 Methods

The procedure for this experiment is similar to the previous experiments but uses a *1-back* design. Subjects were asked to remember charts with a range of 1 to 5, but they were always tested on the chart *before* the one that they just saw, introducing the need to store two charts at all times. 20 subjects (12 women) participated in this experiment, which lasted an average of 25 minutes. This experiment was run in the lab, so the experimenter could confirm that subjects understood the relatively complex instructions. All subjects were undergraduates who received credit in an introductory psychology course for participating in the experiment.

For simplicity we excluded the axis and question pictograph conditions, so all labels were text. We also excluded the superfluous pictograph condition.

The experiment included 160 trials (2 stacked vs stretched × 2 shape vs pictograph × 40 repetitions) blocked by condition. The first 5 trials of each block were discounted to build up the load on memory.

Exp 3 Results

When memory is crowded, pictures help (Fig. 11). Using the same scaled measure and analysis as the previous experiments, pictograph charts led to slightly less error than charts with simple shapes ($F[1, 19]=20$, $p < 0.0001$, $\eta_p^2=0.06$).

Interestingly, the advantage found in previous experiments for stacked over stretched representations of number disappeared ($F[1, 19]<1$, $\eta_p^2<0.01$) and no interaction was found ($F[1, 19]=2.2$, $p > 0.1$, $\eta_p^2<0.01$). While this difference merits further study, we suspect that an increase in memory crowding shifted the performance bottleneck, such that the speed and accuracy of extracting values from the graph became less important than the ability of semantic memory to overcome competition from other datasets.

EXPERIMENT 4: SPEED AND PERFORMANCE

Do pictographs impact the speed of information recovery from a chart? Intuition might suggest that they make it easier, but they also might distract – and the results of experiment 1's axis label condition already suggest that text can trump pictographs in the context of a memory task. This experiment simulates the act of rapidly comparing the values of two known variables in a chart.

Exp 4 Methods

Subjects are presented with two targets (Fig. 12), each in the form of a pictograph with the corresponding name under it (e.g., "dog" and "parrot"). In the center is one of two questions: Which has MORE? or Which has FEWER?

A key press starts the trial and the timer. The subject views a chart with three values and presses a key to indicate which of the two pictographs or words best answers the question. The third value in the chart serves to deter people from simply selecting the tallest or shortest bar on the graph.

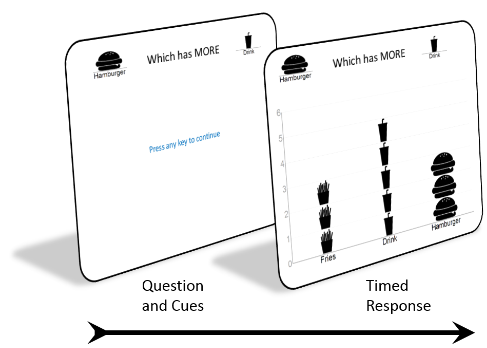

Fig. 12. The procedure for experiment 4. Notice that the order of the targets may be different from their order in the chart.

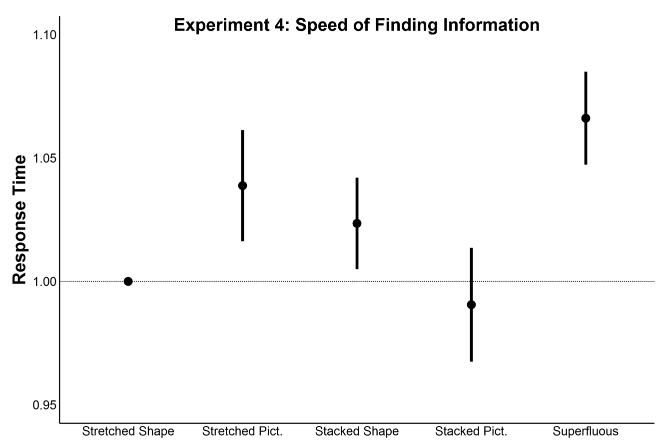

Fig. 13. How quickly subjects answered a question using the different types of charts. There is a performance cost for including a superfluous background image.

We ran 50 subjects on Amazon Mechanical Turk in 200 trials (5 chart types × 2 questions × 20 repetitions) blocked by chart type. Each subject was paid 8 US Dollars for the 30-minute study, and all participants were from the USA.

Exp 4 Results

All subjects showed over 92% accuracy, allowing incorrect responses to be dropped from analysis without substantially affecting statistical power. We also collapsed across the 'More' vs 'Fewer' condition to yield approximately 40 trials per chart type per subject. As with the previous experiments, we analyzed the results within-subject to determine the performance relative to that of the simple bar charts.

We found a main effect of graph type on response time ($F[4, 49]=20$, $p < 0.05$, $\eta_p^2=0.02$). A Tukey HSD-corrected comparison of all the graph types found that only the superfluous condition was significantly different from the standard bar graph ($p < 0.05$) as can be seen in Fig. 13.

This result combined with the results of experiment 1 show that superfluous images hurt both memorability and speed of usability of charts.

EXPERIMENT 5: INITIAL ENGAGEMENT

Although speed can be an important benchmark, the aim of some visualizations is to make people pause and look – as is often the case in news articles. Designers often rely on pictographs because they are thought to draw the attention of a reader. When perusing through a collection of articles, an enticing visualization may increase the likelihood that an article will be inspected more closely. Will an ISOTYPE visualization be better at capturing attention than a simple bar chart? We ran an experiment that simulated how visualizations are commonly encountered in a peripheral glimpse, as thumbnails among a collection of text and other visualizations competing for interest.

Exp 5 Methods

Subjects were presented with a 3x3 grid of items (Fig. 14). Each item included a short title above a small, slightly blurred thumbnail. The thumbnail was either a set of sentences about the topic from Wikipedia or a chart related to the topic. The subjects were given two minutes to look through the thumbnails. They could click whichever item interested them to view the information in full screen without pixilation or blur. Clicking again returned them to the grid, where they could repeat the process. No limit was placed on the number or duration of views for each item. However, after the trial's time had finished, everything was removed from the screen. They were then presented with a button to begin the next trial.

We selected 36 topics from the previous experiments' categories and constructed text, a bar chart, and a stacked pictograph chart for each. Throughout the experiment, each subject encountered each topic exactly once (9 items × 4 trials).

A trial included 3 bar charts, 3 stacked pictograph charts, and 3 pieces of text. We tracked the start time and duration of each view.

10 subjects (4 women) participated in this experiment. Because it was implemented as a Windows desktop application, it was run in the lab. All subjects were undergraduates and were paid 5 US dollars for the 15 minute duration.

Exp 5 Results

We binned the first minute of viewing into one-second intervals and found the portion of subjects viewing each type of item. Fig. 15 shows a linear fit of these results collapsed across trial. For the first few seconds, most are at the selection grid. However, the ISOTYPE visualization takes a quick

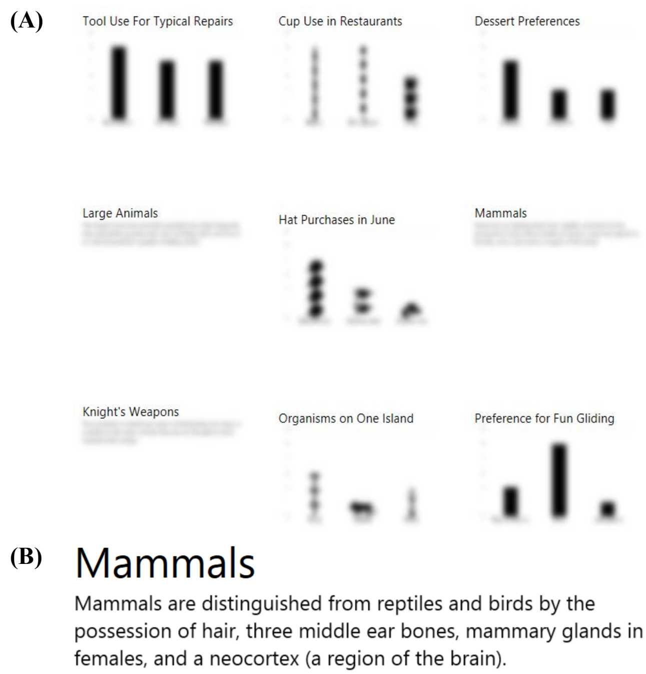

Fig. 14. (A) An example of the selection grid for experiment 5. The title is readable, but the details of the content are unrecognizable beyond the type of information. (B) An example text display that can also been seen in the middle right of the selection grid in (A).

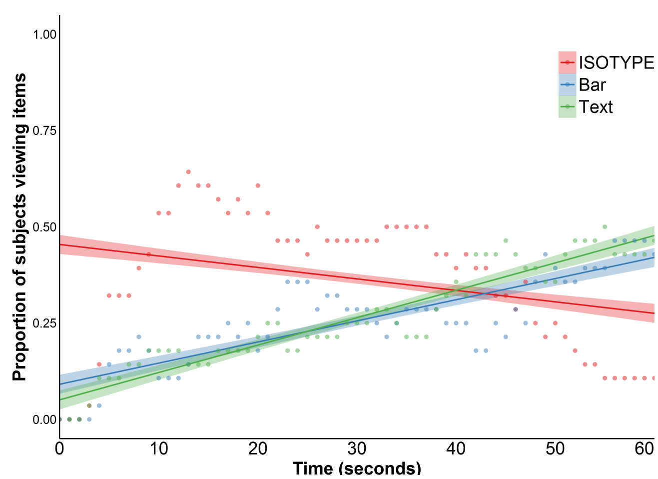

Fig. 15. ISOTYPE charts are best at initially engaging subjects to inspect information more closely.

lead and by 15 seconds has two thirds of the views. After all ISOTYPE visualizations are viewed, subjects proceed to the other items. Because most subject viewed all the items within the first minute, we discounted the second minute which had no interesting result.

Looking at the proportion viewing a given type of item, we found a main effect of item type (F[2, 9]=61, $p < 0.0001$, η_p^2=0.1) and an interaction between item type and time (F[2, 9]=58, $p < 0.0001$, η_p^2=0.08). A Tukey HSD-corrected comparison of item types found that the bar chart was indistinguishable from text (ns) and the ISOTYPE visualization was substantially different from both ($p < 0.0001$).

We also found no significant effect of trial number on ISOTYPE viewing (F[1, 9]<1, η_p^2<0.01), revealing that the initial interest in ISOTYPE-style charts remained consistent through the experiment. Not only are ISOTYPE visualizations highly effective at attracting initial attention, but new ISOTYPE visualizations also continue to engage.

GUIDELINES

Based on our findings in the studies, we suggest the following guidelines for using ISOTPYE displays.

(1) Superfluous pictographs are a distraction

Pictographs do not impair the viewer as long as they are used to represent data. But including an unnecessary background image in a visualization appears to be distracting, and it may divert attention away from the data. Even replacing text labels with pictographs makes encoding less efficient (at least when the text labels are unambiguous).

(2) Redundantly code length and (small) number

Break up large length-defined objects (such as the bars in a bar chart) into a few smaller items. One way of doing so is to use ‘Tufte-style’ gridlines [24], which are white lines superimposed over a (black) bar chart. This approach divides the bars at regular intervals allowing a user to also make a number estimate rather than only a length judgment. For small values below 4-5, number estimation is quick and accurate. However, because this performance diminishes rapidly for larger values, gridlines that break the bars into more than a few sections are unlikely to be beneficial.

(3) Use pictographs for demanding tasks

When working memory is under load, the data in ISOTYPE visualizations is recalled more accurately than with simple bar charts. Presenting successive visualizations of different information (such as visualizing sales of different products or showing food preferences in different regions) may benefit from ISOTYPE. In spite of the additional visual complexity, the information is recalled more accurately.

(4) ISOTYPE engages readers

Visualizations rarely exist in isolation. They are often embedded among additional content such as text and other visualizations that compete for a user, reader, or viewer’s attention. ISOTYPE visualizations offer a way through this assortment to engage with a potential viewer. People are inclined, at least initially, to direct their focus towards a visualization with pictographic data compared with a simple bar chart or text.

CONCLUSIONS AND FUTURE WORK

Our initial exploration used simple visualizations with tightly controlled experiment parameters in hopes that this work can be extended to a more diverse set of contexts. For example, ISOTYPE visualizations have been used with multiple colors, as stacked bar charts, for fractional values, to represent vary large values, and with more than three bars. In addition, the present work uses short and abstracted tasks that allow exploration of a large parameter space. Future work should confirm that ISOTYPE visualizations help viewers better encode the relations and patterns in a dataset, extending beyond the encoding of data values that we tested here. Furthermore, because our tasks were designed to simulate communication of data, and not analytics, it is unclear whether ISOTYPE visualization would be useful for tasks such as search or group segmentation, which are common in exploration. Recent advancements in automatic pictograph selection [22] adds impetus to the need for a better understanding of how pictographs impact chart usability.

We have four main conclusions. (1) Only pictographs embedded as part of data mapping are beneficial (or at least, not harmful). Superfluous pictographs and label images are distracting and confusing. (2) Discretizing a bar into a collection of small items improves encoding and recall but only for small values. (3) Pictographs can help people remember information during demanding tasks. (4) Pictographs entice people to inspect a visualization more closely.

We found no strong evidence that using pictographs to communicate data hurts performance on any of our tasks. Our work adds more evidence to the claim that not all kinds of ‘chart junk’ are necessarily detrimental. The space of questions around the efficacy of embellishments in visualization is still largely unexplored, and it appears to be a fertile ground for further research.

DEMO AND CODE

Chart examples generated in D3, practice versions of the experiments, and all of the pictographs are available at:

`http://steveharoz.com/research/isotype`

ACKNOWLEDGEMENTS

Thanks to Kevin Hartstein for helping run the experiments and Darko Odic for suggesting references. This work was partly funded by NSF grants IIS-1162067 and BCS-1056730.

REFERENCES

1. Annink, E. and Bruinsma, M., eds. *gerd arntz: graphic designer.* 010 publishers, 2013.
2. Bateman, S., Mandryk, R., Gutwin, C., Genest, A., McDine, D., and Brooks, C. Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. *ACM Conference on Human Factors in Computing Systems*, (2010), 2573–2582.

3. Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P.W., Reppa, I., and Floridi, L. An Empirical Study on Using Visual Embellishments in Visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2759–2768.
4. Borkin, M.A., Vo, A.A., Bylinskii, Z., et al. What Makes a Visualization Memorable? *Transactions on Visualization and Computer Graphics*, 6 (2013).
5. Brady, T.F., Konkle, T., and Alvarez, G.A. A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision 11*, 5 (2011).
6. Brinton, W.C. *Graphic Methods for Presenting Facts.* The Engineering Magazine Company, New York, 1914.
7. Chevalier, F., Vuillemot, R., and Gali, G. Using concrete scales: a practical framework for effective visual depiction of complex measures. *IEEE transactions on visualization and computer graphics 19*, 12 (2013), 2426–35.
8. Choo, H. and Franconeri, S.L. Enumeration of small collections violates Weber's law. *Psychonomic bulletin & review 21*, 1 (2014), 93–9.
9. Cowan, N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and brain sciences 24*, (2001), 87–114; discussion 114–185.
10. Dickinson, C.A. and Intraub, H. Spatial Asymmetries in Viewing and Remembering Scenes: Consequences of an Attentional Bias? *Attention, Perception, & Psychophysics 71*, 6 (2009), 1251–1262.
11. Droit-Volet, S., Clément, A., and Fayol, M. Time, number and length: similarities and differences in discrimination in adults and children. *Quarterly journal of experimental psychology 61*, 12 (2008), 1827–46.
12. Franconeri, S.L. The nature and status of visual resources. In *Oxford handbook of cognitive psychology*. 2013.
13. Godden, D.R. and Baddeley, A.D. Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology 66*, (1975).
14. Halberda, J., Sires, S.F., and Feigenson, L. Multiple spatially overlapping sets can be enumerated in parallel. *Psychological science 17*, 7 (2006), 572–6.
15. Haroz, S. and Whitney, D. How Capacity Limits of Attention Influence Information Visualization Effectiveness. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2402–2410.
16. Heer, J. and Bostock, M. Crowdsourcing graphical perception. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, ACM Press (2010), 203–212.
17. Holmes, N. Pictograms: A view from the drawing board or, what I have learned from Otto Neurath and Gerd Arntz (and jazz). *Information design journal 10*, 2 (2001), 133–144.
18. Kaufman, E.L. and Lord, M.W. The discrimination of visual number. *The American journal of psychology 62*, 4 (1949), 498–525.
19. Lin, S., Fortuna, J., Kulkarni, C., Stone, M.C., and Heer, J. Selecting Semantically-Resonant Colors for Data Visualization. *EuroVis*, (2013), 1–10.
20. Michal, A.L. and Franconeri, S.L. The order of attentional shifts determines what visual relations we extract. *Journal of Vision 14*, 10 (2014), 1033–1033.
21. Neurath, O. *From hieroglyphs to Isotype: a visual autobiography.* Hyphen Press, 2010.
22. Setlur, V. and Mackinlay, J.D. Automatic generation of semantic icon encodings for visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2014).
23. Standing, L. Learning 10,000 pictures. *The Quarterly journal of experimental psychology 25*, 2 (1973).
24. Tufte, E.R. *The visual display of quantitative information.* Graphics press, 1983.