Eurographics/ IEEE-VGTC Symposium on Visualization 2010
G. Melançon, T. Munzner, and D. Weiskopf
(Guest Editors)

Volume 29 (2010), Number 3

Understanding Interactive Legends: a Comparative Evaluation with Standard Widgets

Nathalie Henry Riche[1], Bongshin Lee[1] and Catherine Plaisant[2]

[1]Microsoft Research, Redmond, WA, USA
[2]University of Maryland, College Park, MA, USA

Abstract

Interactive information visualization systems rely on widgets to allow users to interact with the data and modify the representation. We define interactive legends as a class of controls combining the visual representation of static legends and interaction mechanisms of widgets. As interactive legends start to appear in popular websites, we categorize their designs for common data types and evaluate their effectiveness compare to standard widgets. Results suggest that 1) interactive legends can lead to faster perception of the mapping between data values and visual encodings and 2) interaction time is affected differently depending on the data type. Additionally, our study indicates superiority both in terms of perception and interaction of ordinal controls over numerical ones. Numerical techniques are mostly used in today's systems. By providing solutions to allowing users to modify ranges interactively, we believe that interactive legends make it possible to increase the use of ordinal techniques for visual exploration.

Categories and Subject Descriptors (according to ACM CCS): H.5 [Information Interfaces and Presentation]: Miscellaneous—

1. Introduction

While the main purpose of information visualization is data exploration to discover patterns and exceptions, presenting and communicating results is also important. Interactive visualizations help users gather insights from the data by allowing them to explore it visually from different perspectives. Interaction plays a key role in this exploration by allowing users to perform dynamic queries (i.e. filtering data) or modifying visual encodings. This interaction is usually supported by a set of controls which we will call standard widgets (Figure 1 left). For example, check boxes and lists are used to filter categorical data or sliders to filter numerical data. A few more complex widgets have been designed, such as range sliders (double-sided sliders) to specify ranges, or color wheels to select a specific hue.

When the time comes to communicate their insights and prepare a report, users can use export features or screen capture utilities to illustrate their findings. The static visualizations must stand alone and explain their content and visual encoding. To that end, tools often generate legends

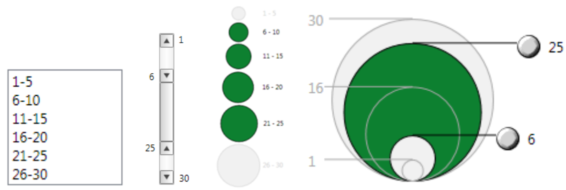

Figure 1: *Standard widgets (left), interactive legends (right)*

and captions, most likely because legends are common on printed materials and seem easier to interpret. Information visualization has now started to reach a wider audience with popular websites such as the Baby Name Voyager [Wat05], Many Eyes [VWvH*07], or maps of election results [PEG]. Some of these websites replace the standard widgets with a new form of controls which we call interactive legends (Figure 1 right). These controls augment legends with interaction mechanisms allowing users to perform queries or to modify the representation. More professional systems such as

Tableau [MHS07] are also integrating widgets that are more similar to what users could find in legends. This new practice raises several research questions:

1. Are interactive legends more effective than widgets at conveying the visual encoding and the current state of interactive visualizations?
2. Are interactive legends more difficult to interact with than standard widgets?
3. Should the use of interactive legend be generalized (i.e., replacing standard widgets)?

In the remaining sections, we first review the related work, define interactive legends and describe our designs. We report the results of two controlled experiments comparing interactive legends and standard widgets for different visual encodings in terms of perception and interaction. We conclude with design implications for information visualization.

2. Related work

Two seminal books present fundamental concepts in perception and design of static visualizations (e.g. prints): Bertin's Semiology of Graphics [Ber83] and Tufte's Visual Display of Quantitative Information [Tuf01]. Both define and present key concepts and illustrate them with a large number of visualizations. They also provide a rich but finite set of commonly used legends in printed communications.

A number of researchers focused on graphical perception and evaluated users' ability to interpret visual encodings such as shape, colors, and position to represent different attributes of the data. Cleveland and McGill [CM85] provide seminal work in this area and rank different visual encodings according to their performances. Ware in his book on perception for design regroups a large amount of research in the area [War04]. Treisman [Tre85] and later Healey [Hea96] led research on preattentive perception. Other researchers worked on evaluating graphical perception for particular types of visualizations such as Spence and Lewandowsky [SL91] for bar charts, pie charts and tables or Purchase [Pur97] for node-link diagrams.

The first 15 years of research in information visualization mostly lead to interactive systems using standard widgets (e.g., check boxes, sliders, lists) to specify zoom and filter operations or to manipulate the visual encodings. Following principles of direct manipulation [Shn87], dynamic queries [AWS92] have become common in application such as Spotfire [Ahl96], Improvise [Wea06], or the infovis toolkit [Fek04]. Beaudouin-Lafon in his article on instrumental interaction [BL00] lists a number of drawbacks of standard WIMP (Windows Icon Mouse Pointer) interfaces. Two of them transfer directly to the use of widgets in information visualization systems:

1. The large amount of screen-estate used by objects that are used to control the visualization but are not, in fact, the main objects of interest.
2. Not so direct manipulation: the sequence of actions breaks the current attention on the visualization. Dynamic queries provide a direct feedback on the visualization but users still use a trial and error process and split their attention between the visualization and the controls as controls provide limited feedback.

While a number of research projects has gone into the design of zooming techniques for large visual spaces [BH94], surprisingly little research attempted to evaluate the performance of standard widgets and improve their design.

Ahlberg and Shneiderman compared Alpha slider designs to select items in large lists [AS94]. Chintalapany *et al.* [Chi04] designed a color binning legend to control the color encoding of a treemap visualization. More recently, Willett *et al.* proposed scented widgets [WHA07], a generalization of augmented scrollbar [HHWM92], and histogram sliders [DHMR99].

The design and study of interactive legends has also been active in the field of cartography. Peterson [Pet99] also proposed changing the color encoding of maps when mousing over the legends. Similar interactive legends can be seen in Many Eyes [VWvH*07]. More complex legends have also been created to interact with spatio-temporal environmental data [REP97, HM99].

Several brushing and linking techniques have been proposed [ABS91, Pet06] to link visual components together (sometimes including legends) and Tudoreanu and Kraemer [TK01] have proposed to use legend as interaction device. However, in most of today's systems traditional widgets remain the standard tools to interact with visualizations. In this paper, we aim at comparing the performance of these standard widgets to interactive legends. Our goal is to provide empirical evidence to guide visualization researchers in the design of effective interaction tools.

3. Interactive legends

Legends constitute a *visual dictionary* of all the elements of a visual representation. We define interactive legends as a class of controls that combine the visual representation found in legends (i.e. integrate visual objects of the representation) and interaction mechanisms provided by widgets (Figure 1) to select or filter out data in visualizations.

We believe that interactive legends have advantages that make them highly desirable for visual exploration.

1. Legends allow users to directly visualize the mapping between the visual encoding and the data value. For example, in Figure 1, one can see the size associated with 30, whereas the widget merely tells you 30 is the largest value. The direct mapping of interactive legends may reduce the amount of attention shifting between controls and visualization, as well as reduce the trial and error process needed to understand that mapping.

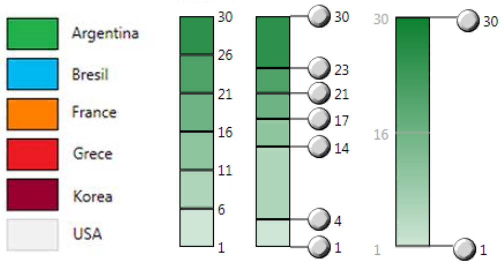

Figure 2: *Interactive legends for color and opacity. Categorical color (left), ordinal opacity without and with handles to modify ranges (middle), numerical opacity (right).*

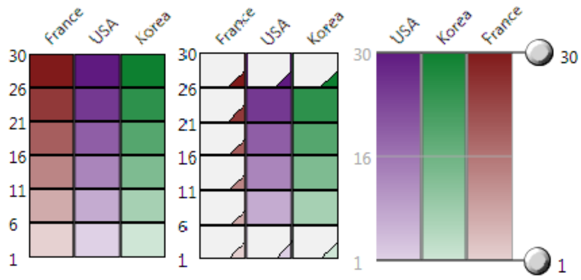

Figure 3: *Combined interactive legends for color+opacity. Categorical color and ordinal opacity (left), feedback on filtered elements (middle), categorical color and numerical opacity (right).*

2. Legends summarize the encodings used in the visualization. Most viewers are familiar with them since reading legends for maps and charts is taught early and regularly in elementary schools. During interactive exploration, filters can be applied multiple times and the visualization state may change across datasets or representations. We believe legends will help users rapidly perceive this state.

However, interactive legends may also have drawbacks.

1. Widgets are familiar to users as most user interfaces include them (maybe with the exception of range sliders). Interactive legends may lack affordances and their use may degrade the interaction performance since users are not as familiar with them.
2. Legends use more screen real estate since they are composed of multiple visual objects.

To better understand if and when to use interactive legends and quantitatively assess the benefit of their use, we compared them to their widget equivalent.

3.1. Our design

In this section, we present the characteristic of interactive legends with respect to the four data types described by Bertin [Ber83]: nominal, categorical, ordinal, and numerical data. Most systems using interactive legends do not make this distinction, which may lead to confusion. For example, Many Eyes uses both ordinal and numerical legends but both are mapped to numerical data.

One important difference between interactive legends and static legends is that static legends in printed materials do not show filtered-out values. When the goal of the presentation is to convey a message in the data, filtered out data is considered noise and thus removed. However, in visual exploration, it is crucial to represent the filtered data to keep users aware of the whole dataset and allow them to remove or update filters. In our design, we selected a light gray color to indicate filtering both in the visualization and the interactive legend.

3.1.1. Nominal and categorical data

Nominal data, such as the label of each item, is generally represented as textual information in the caption of a visual representation. Categories, since they are less numerous, are generally represented by a list. While this list can be ordered alphabetically, it does not represent a unique and universal order conveyed by the data. To express this absence of orders, we separate items in the list with white space. Each category is coupled with the visual encoding representing it (Figure 2 left). This legend is the most commonly used. The same design applies to other visual variables suitable to represent categorical data, such as shapes. The boundary between interactive legend and widget is really thin here since the list controls may integrate visual icons. We argue that when a control contains a visual object used in the representation, it should be categorized as an interactive legend.

The strength of standard widgets is that they convey how to interact with them (i.e., they have good affordances). In the case of categorical data, the user interface would use either check boxes (filtering by unselecting boxes), radio buttons (a single category showed at a time, clicking on a second one filters out the others) or list boxes (act as a radio button but provide shift and ctrl for multi-selection). Each interaction has advantages and drawbacks depending on the number of categories to display and the possible constraint of the system (for example forbidding multiple filtering). In fact, an optimal use of these three types of interaction is to change widgets according to the type of data visualized and the filtering constraints the system has. In practice, system designers take a decision and provide a single widget. As list boxes offer the largest flexibility, it is a common choice. Thus, it is the interaction we decided to offer to interact with the legend. Clicking on an items filters the others out. Shift-click is used to select a range of items. Ctrl-click is used to toggle items to the selection.

3.1.2. Ordinal data

Similarly to the legend representing categorical data, ordinal data is represented by a list. In ordinal data, each category represents a range of values that are all encoded by the same visual encoding (e.g., same size). We considered ranges of numerical values (or binned numerical values) as ordinal data. To indicate the presence of an order, in the case of ranges encoded by opacity (Figure 2), we removed the space between elements and indicated a single value to mark each range boundary. This makes the legend more compact and also makes it possible to display the legend horizontally. In the case of ranges encoded by size (Figure 1), only the increase of numerical ranges and the increase in the size of the visual items mark the ordinal nature of the data. We provide the same interaction as for categorical data.

3.1.3. Numerical data

Numerical data raises the challenge of representing continuous values and visual encodings. Figure 1 and Figure 2 show examples of the visual encoding of numerical data using size and opacity. As we were browsing a large number of visual printed communication, we noticed that legends for numerical data seem far less common than these for categorical and ordinal data. For example, we did not find any example of static legends for numerical data encoded by orientation, angle or curve closure. The most common widgets to filter numerical data in information visualization are the slider and range slider. To enable the same type of interaction, we used lines and handles that users can grab and move to adjust the filter. Figure 5 shows a filtered legend.

3.2. Combining Interactive Legends

An interesting property of legends is their ability to be combined and to provide a more extensive visual dictionary. Furthermore, as they are visual objects, it seems less awkward to drag and drop them on top of each other to combine them than doing so with widgets. In particular, combining legends representing the same attribute offers dual visual coding, while combining two different attributes expands the visual dictionary provided to the viewer.

Dual coding. To reinforce the interpretation between visual encoding and data attribute, it is common to use dual-coding in a visual legend. In most current systems, there is no difference between two visual variables encoding the same attribute or two different attributes, in the sense that there are two separate widgets in both cases. By combining legends representing the same attribute, it is possible to refine the visual dictionary presented to the user. We believe that in addition to gaining some space, combined legends also better convey the current encoding of the visualization and suppress the potential confusion when filtering one of the attributes. Filtering the size widget should also update the filter on the opacity widget.

Combining two legends. To further ease the perception of the visual encodings, we can extend the visual dictionary provided to the user. Figure 3 shows an example in which color and opacity encode two different attributes and are combined in a single legend. The visual dictionary is extended as the legend provides all combinations of color and opacity. To combine categorical and ordinal legends for example, the legend becomes a matrix representation, in which each cell provides the encoding for the corresponding combination of the two attributes values (Figure 3 left).

Combining two different attributes changes the filtering interaction as users may filter one or the other attribute or both. For selecting a given attribute, we attempted to provide a similar interaction as for individual legends. Single click, shift and ctrl to toggle selections are provided for categorical and ordinal data; handles for numerical data. Compared to standard widgets, combined interactive legends reduce the number of interactions for selecting a given combination of color and opacity as it only requires a single click. It also allows more complex filtering. For example, providing the possibility to select different opacity ranges for each color typically requires providing a widget by color while interactive legends achieve this with one coherent control.

4. User Study

To compare users' performance between Interactive Legends (IL) with Standard Widgets (SW), we conducted two controlled experiments using a social network with 30 nodes and 50 links. We designed our experiments to answer the following questions:

- Do interactive legends improve the perception of the mapping between data attribute and size/opacity?
- Are interactive legends harder to use (i.e., slower) than widgets to interact with when filtering the data?

4.1. Tasks

We compared IL with SW with two types of tasks: perception and interaction. Each type included two different tasks based on the direction of the mapping participants had to accomplish: 1) from a data value to a visual encoding and 2) from a visual encoding to a data value.

In the **perception** tasks, we measured how quickly the participant could perform the mapping between data values and their visual encodings.

- Task 1 required participants to assess the value designated by a visual encoding. For instance, we highlighted a node in the visualization and asked participants to estimate its value and either select among multiple choices of ranges (ordinal data), or enter a best estimate within 15% error margin (numerical data).
- Task 2 required participants to assess the size/opacity encoding of a given value. For instance, "Click on a node with a movie count between 5 and 10."

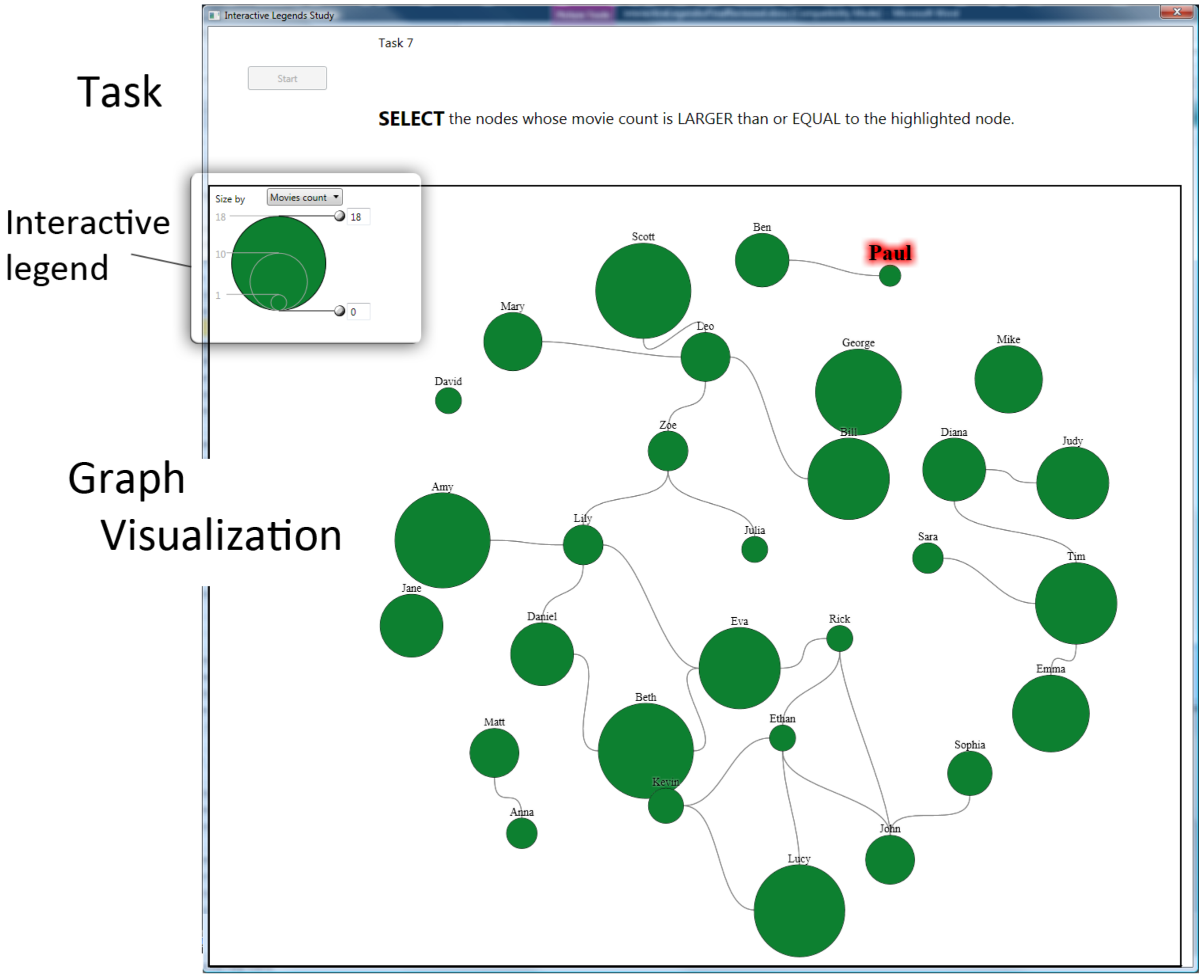

Figure 4: *Experimental software. The top of the screen contains the task and start button. The center of the screen contains the visualization including an interactive legend or a standard widget depending on the experimental condition.*

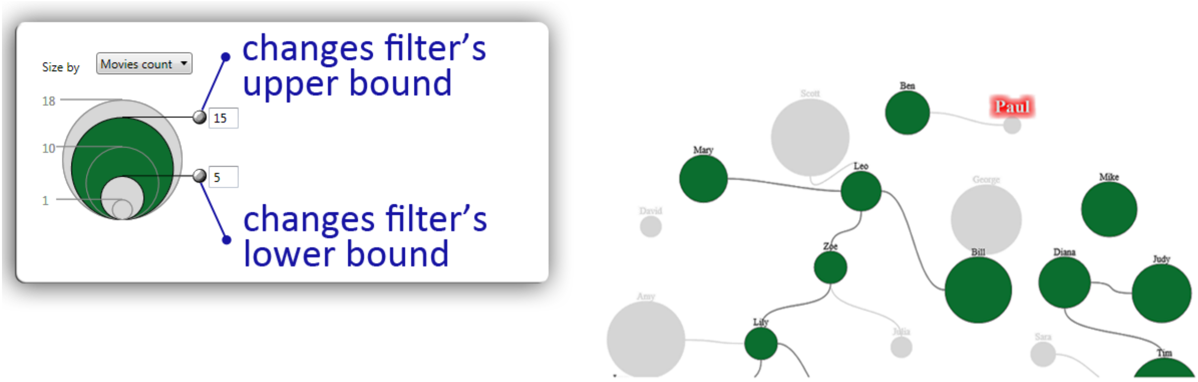

Figure 5: *Interactive legend for controling the size. Handles are provided to filter interactively the visualization.*

Note that during normal use of a visualization users may click on the node or use a popup window to get the exact value instead of estimating its value, but this is not possible when users try to estimate the value of a large number of items at once. Our tasks try to reproduce this situation.

In the **interaction** tasks, we measured how fast participants could interact with the widget to retrieve a particular value or set of values by filtering out data.

- Task 1 required participants to filter the values encoded by a designated size/opacity. For instance, "Select nodes that have a movie count in the same range than the highlighted node (ordinal data)" or "a smaller or equal movie count than the highlighted node" (numerical data).
- Task 2 required participants to filter the size/opacity given one or a range of value(s). For instance, "Filter out nodes that have a movie count within between 5 and 10."

4.2. Participants and Procedure

We recruited 8 participants for each experiment, screened to include people with general computer experience, balanced for age and gender, and not color-blind. For Experiment 1, we had 4 males and 4 females, with an average age of 29.6, ranging from 22 to 33 years of age. For Experiment 2, we had 5 males and 3 females, and the average age was 33.8, ranging from 19 to 48 years of age. Each participant used a 3.00 GHz dual-core PC with 4 GB of RAM, running Windows Vista, and using 21" Samsung monitors at a resolution of 1200x1600; the size of the network was 1000x1000. Figure 3.2 shows a screenshot of the experimental software.

Before each Technique, participants received instruction and practiced the tasks in order to familiarize themselves with both task and interface. Results were recorded by the application used for the experiment. For each task, participants clicked on a button to indicate that they had finished reading the description of the task and were ready to begin, and pressed the space bar when they were done. When they had to enter a value, we stopped the time when they started typing (to exclude this typing time from the results). To keep the study a reasonable length, we limited each task to a maximum of 30 sec. The study lasted approximately 60 min including training. After the experiment, participants filled out a preference questionnaire.

Independent variables in both experiments included: Technique (Interactive Legends vs. Standard Widgets), DataType (Ordinal vs. Numerical), TaskType (Perception vs. Interaction), Task (1 vs. 2.), Range (Small vs. Large). Additionally, Experiment 1 included Size (Small vs. Large) and Experiment 2 included VisualEncoding (Opacity vs. Color+Opacity).

Dependent variables in both experiments were accuracy (from 0 to 1), task completion time, preferences (rank from 1(best) to 4 (worse)) and confidence (Likert from 1 to 7).

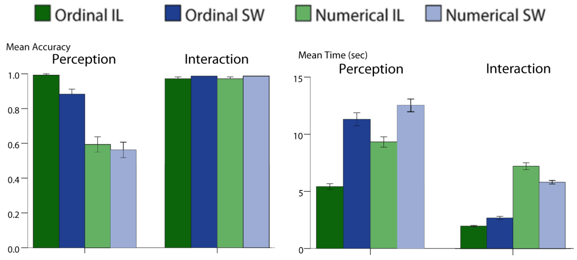

Figure 6: *Bar charts presenting the mean answer (left) and mean time including errors (right) for the four conditions.*

	Accuracy Percep.	Inter.	Time (sec) Percep.	Inter.
Ord IL	0.99 (0.1)	0.98 (0.1)	5.4 (2.8)	2.0 (0.8)
Ord SW	0.88 (0.3)	0.98 (0.1)	11.3 (6.4)	2.7 (1.7)
Num IL	0.59 (0.5)	1.00 (0.1)	9.3 (5.0)	5.8 (1.8)
Num SW	0.56 (0.5)	1.00 (0.1)	12.5 (6.2)	7.2 (3.3)

Table 1: *Mean and standard deviation for accuracy and time for each of the four techniques for experiment 1.*

5. Experiment 1: size

For both experiments, we report results with time including errors. We used non-parametric tests for the accuracy and preference. We analysed performance using a Repeated Measure ANalysis Of VAriance (RM-ANOVA).

5.1. Method

We used a within-subject design: 2 TaskType x 2 Technique x 2 DataType x 2 Task x 2 Range x 2 Size with 2 repetitions. We collected a total of 1024 trials.

We fully counterbalanced the order of the Technique and TaskType and fixed the order for DataType. As the Ordinal case required less training, participants performed Ordinal before Numerical. Participants also always performed Task 1 before Task 2. To reduce memorization effect between trials, we selected two orders (A, B) for the combination Range x Size and alternated them for each Task. Half of the participants performed A then B. Half performed B then A. To reduce the memorization effect, we used the same graph structure but randomly rotated the graph and used different sets of attribute values for each task.

To simplify the description of the results, we define the combination of Technique x DataType as Condition. The four Conditions we use are Ordinal Interactive Legends (Ordinal IL), Numerical Interactive Legends (Numerical IL), Ordinal Standard Widgets (Ordinal SW), and Numerical Standard Widgets (Numerical SW).

5.2. Perception results

H: Legends will be more accurate and faster than widgets in both Ordinal and Numerical cases.

Accuracy: Friedman's test shows a significant difference between the 4 Conditions (p<0.001) (Figure 6). Ordinal IL is significantly more accurate than Ordinal SW . Both Numerical IL and Numerical SW are far less accurate and there is no significant difference between them.

Time: RM-RM-ANOVA shows a significant difference between Conditions ($F_{3,21} = 44.26, p < 0.001$) (Figure 6). Post-hoc pair wise comparison shows that legends (both Ordinal and Numerical) are significantly faster than widgets. Ordinal IL is faster than Numerical IL. There is no significant difference between Ordinal SW and Numerical SW. We found no significant effect of Task, Range or Size (and their interactions with Conditions) on time.

5.3. Interaction results

H: We will not find any significant difference between legends and widgets.

Accuracy: Friedman's test does not show any significant difference between the conditions (Figure 6). Both techniques are very accurate.

Time: RM-ANOVA shows a significant difference by Task ($F_{1,7} = 24.16, p < 0.01$). Task 1 is significantly slower than Task 2. RM-ANOVA also shows a significant difference between Conditions ($F_{3,21} = 153, p < 0.001$) (Figure 6) and two significant interactions: Condition x Task ($F_{3,21} = 77, p < 0.001$) and Condition x Range ($F_{3,21} = 5.55, p < 0.01$). Overall, Ordinal IL is faster than Ordinal SW, which is faster than Numerical SW, which is faster than Numerical IL. The interaction Condition x Task affects the Ordinal and Numerical techniques differently: Ordinal techniques are faster for task 2 whereas the Numerical techniques are slower for this task. Post-hoc pair wise comparison shows that for task 1, there is a significant difference between the fours conditions but for task 2, there is no significant difference between the Ordinal legends and widgets. Numerical SW is still faster than Numerical IL. The interaction Condition x Range also affects the Ordinal and Numerical techniques differently: Ordinal techniques are not affected. Both Numerical techniques are similarly affected and significantly slower (mean raises of 0.5 sec for Numerical IL, mean raises of 0.4 sec for Numerical SW).

5.4. Preference and Confidence

Friedman's test shows a significant difference between users' preference. Wilcoxon's test shows that Ordinal IL is the preferred technique, there is no significant difference between the other ones. Participants reported 70% confidence with Ordinal IL, 67% with Numerical IL, 57% with Ordinal SW and 50% confidence with Numerical SW (Table 2).

6. Experiment 2: opacity

6.1. Method

We used a within-subject design: 2 TaskType x 2 Technique x 2 VisualEncoding x 2 DataType x 2 Task x 2 Range with 2 repetitions. We collected a total of 1024 trials.

We fully counterbalanced the order of Technique and TaskType, and fixed the order of VisualEncoding and DataType. Since the Opacity case required less training than Color+Opacity, participants performed it first. For each VisualEncoding, since the Ordinal case required less training than Numerical, it was performed first. We used Range to reduce memorization effect and alternate Small and Large cases for each Task. To simplify the description of the results, we define Condition as Technique x DataType x VisualEncoding.

6.2. Perception results

H: Legends will be more accurate and faster than widgets for Opacity and Color+Opacity in both Ordinal and Numerical cases. Numerical techniques will be less accurate than Ordinal ones.

Accuracy: Friedman's test shows a significant difference between the 8 conditions (p<0.001) (Figure 7). Wilcoxon's test on pairs shows a significant difference between Ordinal Opacity IL and all other conditions as well as Ordinal Opacity SW and all the other conditions. Ordinal Opacity IL is the most accurate (M=0.81 (0.4)), followed by Ordinal Opacity SW (M=0.7 (0.5)). The other conditions are not significantly different and have a low accuracy (mean varying from 0.41 to 0.52).

Time: RM-ANOVA shows a significant difference between the 8 conditions ($F_{7,42} = 43.85, p < 0.001$) (Figure 7) as well as two significant interactions Condition x Task ($F_{7,42} = 26.5, p < 0.001$) and Condition x Range ($F_{7,42} = 3.7, p < 0.01$). The interaction Condition x Task shows a significant difference between Conditions for Task 1 in the Ordinal case only. Opacity SW (M =5.34sec, SD=2.3) is faster than Color+Opacity SW (6.6 sec (2.4)), whereas there is no significant difference between Opacity IL (4.4 sec (1.8)) and Color+Opacity IL (6 sec (3.5)). Both legends (Opacity and Color+Opacity) are significantly faster than widgets. The interaction Condition x Range reveals a significant difference for the Small Range. Ordinal Color+Opacity SW is slower than all Numerical techniques.

	Confidence (Size)	Confidence (Opacity)
Ord IL	4.88 (1.2)	5.38 (0.8)
Ord SW	4.75 (1.2)	4.75 (1.2)
Num IL	4.00 (1.5)	4.75 (1.7)
Num SW	3.50 (1.5)	4.00 (1.8)

Table 2: *Confidence for both experiments*

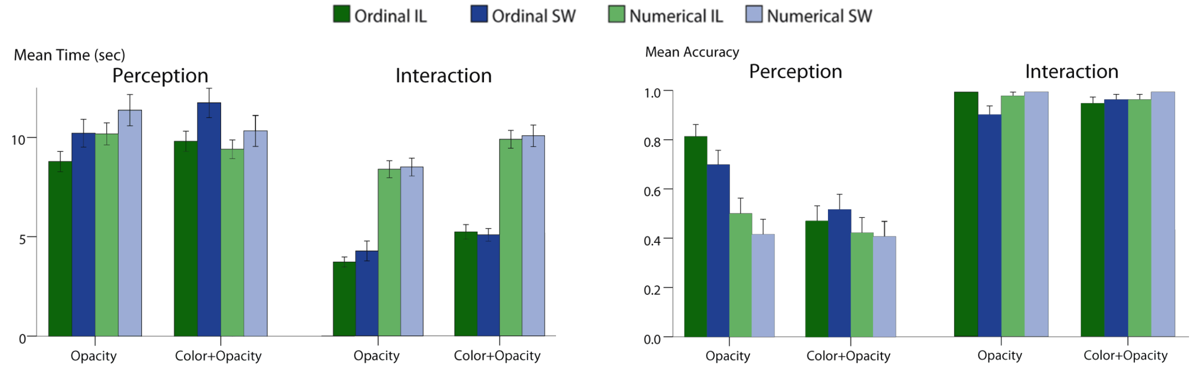

Figure 7: *Bar charts presenting the mean answer (left) and mean time including errors (right) for the eight conditions.*

6.3. Interaction results

H: We will not find any significant difference between legends and widgets for Opacity. Legends will perform faster for Color+Opacity.

Accuracy: Friedman's test shows a significant difference between Conditions (p<0.05) (Figure 7). Wilcoxon's test on pairs shows only a significant difference between Ordinal Opacity IL with 100% accuracy and Ordinal Opacity SW (M=0.9 (0.3)); legends being more accurate than widgets.

Time: RM-ANOVA shows a significant difference between the the Visual Encoding ($F_{1,6} = 68.5, p < 0.001$), the Data Type ($F_{1,6} = 110, p < 0.001$) but not between the Technique (Figure 7). Ordinal techniques are faster than Numerical techniques. Opacity techniques are faster than Color+Opacity techniques. The interaction Condition x Task ($F_{7,42} = 6.3, p < 0.001$) reveals that the Ordinal techniques are faster for Task 2. The interaction Condition x Range ($F_{7,42} = 3.1, p < 0.001$) reveals that the Ordinal techniques are significantly faster for the Large Range.

6.4. Preference

Friedman's test reveals no significant difference between techniques for users' preference. Participants reported that they had 77% confidence with Ordinal IL, 68% with Numerical IL, 68% with Ordinal SW and 57% confidence with Numerical SW (Table 2).

7. Discussion

7.1. Experiment on size

Our results indicate that interactive legends improve the perception of the mapping data value - visual encoding by 10% and they perform twice faster. As participants are able to assess the correct range to filter using legends (instead of clicking successively on widgets), they also perform faster for the interaction tasks.

However, surprisingly, legends do not provide a better perception for the Numerical techniques and they perform significantly slower than widgets for the interaction task. The accuracy drops from around 30% for both widgets and legends in the Numerical case. Confidence results indicate that participants had more confidence in the legends, indicating that they might be misleading or require more usage to be able to estimate their error rate. Legends also suffer from their design for the interaction task. Indeed, the distance between the filtering handles in the legend corresponds to the number of pixels of the larger size. The slider does not suffer from such a limitation since it has an independent fixed width.

7.2. Experiment on color and opacity

Overall, as confirmed by previous perception study [CM85], opacity is difficult to perceive, especially when numerical or combined with colors (both only reach 40% accuracy). Unsurprisingly, we found that Ordinal opacity techniques performed better than the Numerical ones. Legends also improve the perception by 10% in this case and are slightly faster compared to widgets.

Surprisingly, we did not find any difference between the combined legends and the use of two widgets for the interaction tasks. This may be explained by the lack of familiarity with the legend and its interaction; but also by the simplicity of the filter to set. We believe that the combined legend will perform better than widgets in more complex cases (when requiring different opacity filters per color for example).

7.3. Design implications

When information visualization systems integrate interactive legends, they often use them to control numerical data. However, our results show that the benefit of using legends for representing numerical data is low. More generally, our results confirm previous studies [CM85] and clearly show that mapping between data values and visual encodings is very difficult to perceive when using numerical data. Even with size encoding (which is good for perception), participants could not estimate the value correctly in 40% of the cases (within a 15% error margin). Participants were more confident with the numerical legends than the widgets despite similar error rates. While it may be due to the lack of familiarity, this could also suggest that range sliders may be preferable when exploring numerical data, because numerical legends may lead to unwarranted trust in poor estimates. Indeed, participants were well aware of this high error rate when using the widgets as they rated their confidence on 50% in average.

However, our results show that there is a real benefit in using legends for perceiving and interacting with ordinal data. Legends improve the perception accuracy by 10% in both experiments and decreases the perception time by 50% when using size. Participants can better assess the range to filter and improve the interaction with the control too. Surprisingly, ordinal data controls are rarely used in today's interactive visualization systems. This is likely to be caused by the loss of precision when using different ranges (e.g., becomes impossible to distinguish two different values in the same range). In today's systems, modifying ranges on the fly is rarely supported and requires either modifying the data or the program. However, when using interactive legends, it is simple to offer the modification of the ranges. Figure 2 middle shows how we can provide handles between visual objects to control the ranges upper and lower boundaries.

Those results reinforce the recommendation to use ordinal legends in information visualization systems even if using these types of controls comes with a price: scalability. Indeed, ordinal interactive legends require more screen real estate than widgets but we believe that helping users acquire a batter mapping between data value and visual encoding will reduce their cognitive load and improve their visual exploration. Furthermore, there are opportunities to reduce the amount of space used by interactive legends. For example all the values of filtered data can be grouped in a single category when users finish interacting with the legend. Combined legends may also offer a good compromise to reduce the screen real-estate since they do not degrade performances and provide the ability to compose richer filters (e.g., filter out different opacity ranges by color).

8. Conclusion and future work

We explored the design space of interactive legends as a novel class of controls that combine the visual representation found in static legends and interaction mechanisms found in the widgets. Results from our user studies suggest that interactive legends are valuable alternatives to the standard widgets. Because interactive legends provide the possibility to modify parameters of the legend on the fly, one of their largest benefit might be that designers will choose to use ordinal controls instead of the deceptively difficult to use numerical ones. However, we need to point out the limitations of our studies and advocate that further studies should be run to explore the whole potential of interactive legends. Indeed, there are a number of design variations, each with strength and drawbacks. In this paper, we evaluated a small subset of them in very controlled conditions.

Future work might include developing design guidelines for improving the visual affordances of interactive legends. While users are familiar with the interaction mechanisms of the standard widgets the discoverability of interactive legends might suffer from their resemblance to static legends. Three dimensional elements and subtle animations might be helpful. There are also novel legends to design. For example what would be the appropriate interactive legend to represent and control motion or vibration? Finally, hybrid designs might be investigated to address the screen space issues. Using zooming, an interactive legend might become enlarged to reveal more detail and facilitate interaction when the cursor hovers over the controls.

References

[ABS91] A. Buja J. A. McDonald J. M., Stuetzle W.: Interactive data visualization using focusing and linking. In *Vis'91: Proceedings of the 2nd conference on Visualization* (1991), IEEE Computer Society Press, pp. 156–163. 2

[Ahl96] Ahlberg C.: Spotfire: an information exploration environment. *SIGMOD Rec. 25*, 4 (1996), 25–29. 2

[AS94] Ahlberg C., Shneiderman B.: The alphaslider: A compact and rapid selector. ACM Press, pp. 365–371. 2

[AWS92] Ahlberg C., Williamson C., Shneiderman B.: Dynamic queries for information exploration: An implementation and evaluation. ACM Press, pp. 619–626. 2

[Ber83] Bertin J.: *Semiology of Graphics*. University of Wisconsin Press, Madison, WI, 1983. (trans. W. Berg). 2, 3

[BH94] Bederson B. B., Hollan J. D.: Pad++: a zooming graphical interface for exploring alternate interface physics. In *UIST '94: Proceedings of the 7th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1994), ACM, pp. 17–26. 2

[BL00] Beaudouin-Lafon M.: Instrumental interaction: an interaction model for designing post-wimp user interfaces. In *CHI '00: Proc. of SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2000), ACM, pp. 446–453. 2

[Chi04] Chintalapani G. P. C. S. B.: Extending the utility of treemaps with flexible hierarchy. *Proc. International Conference on Information Visualization* (2004), 335–344. 2

[CM85] Cleveland W. S., McGill R.: Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science 229* (1985), 828–833. 2, 8

[DHMR99] Derthick M., Harrison J., Moore A., Roth S. F.: Efficient multi-object dynamic query histograms. In *Proc. of Information Visualization* (1999), IEEE Press, pp. 58–64. 2

[Fek04] Fekete J.-D.: The infovis toolkit. In *INFOVIS '04: Proc. of the IEEE Symposium on Information Visualization* (Washington, DC, USA, 2004), pp. 167–174. 2

[Hea96] Healey C. G.: Choosing effective colours for data visualization. In *In Proceedings Visualization ï£¡96* (1996), pp. 263–270. 2

[HHWM92] HILL W. C., HOLLAN J. D., WROBLEWSKI D., MCCANDLESS T.: Edit wear and read wear. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1992), ACM, pp. 3–9. 2

[HM99] HARROWER M., MACEACHREN A.: Exploratory data analysis and map animation: Using temporal brushing and focusing to facilitate learning about global weather. In *International Cartographic Association Commission on Visualization Workshop* (1999). 2

[MHS07] MACKINLAY J., HANRAHAN P., STOLTE C.: Show me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1137–1144. 2

[PEG] PRESIDENTIAL ELECTION GUIDE 2008 W.: http://elections.nytimes.com. 1

[Pet99] PETERSON M. P.: Active legends for interactive cartographic animation. In *International Journal of Geographical Information Science* (1999), vol. 13, pp. 375–383. 2

[Pet06] PETERSON M. P.: Towards ubiquitous brushing for information visualization. In *International Conference on Information Visualisation (IV'06)* (2006), pp. 151–156. 2

[Pur97] PURCHASE H. C.: Which aesthetic has the greatest effect on human understanding? In *GD '97: Proceedings of the 5th International Symposium on Graph Drawing* (London, UK, 1997), Springer-Verlag, pp. 248–261. 2

[REP97] R.M. EDSALL M.-J. KRAAK A. M., PEUQUET D. J.: Assessung the effectiveness of temporal legends in environmental visualization. In *GIS/LIS'97, Cincinnati* (1997), pp. 677–685. 2

[Shn87] SHNEIDERMAN B.: Direct manipulation: A step beyond programming languages. 461–467. 2

[SL91] SPENCE I., LEWANDOWSKY S.: Displaying proportions and percentages. 61–77. 2

[TK01] TUDOREANU M. E., KRAEMER E.: Legends as a device for interacting with visualizations. Technical Report WUCS-01-44. 2

[Tre85] TREISMAN A.: Preattentive processing in vision. *Comput. Vision Graph. Image Process. 31*, 2 (1985), 156–177. 2

[Tuf01] TUFTE E. R.: *The Visual Display of Quantitative Information, 2nd edition*, 2 ed. Graphics Press, May 2001. 2

[VWvH*07] VIEGAS F. B., WATTENBERG M., VAN HAM F., KRISS J., MCKEON M.: Manyeyes: a site for visualization at internet scale. *Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1121–1128. 1, 2

[War04] WARE C.: *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004. 2

[Wat05] WATTENBERG M.: Baby names, visualization, and social data analysis. In *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization* (Washington, DC, USA, 2005), IEEE Computer Society, p. 1. 1

[Wea06] WEAVER C. E.: *Improvise: a user interface for interactive construction of highly-coordinated visualizations*. PhD thesis, Madison, WI, USA, 2006. Adviser-Livny, Miron. 2

[WHA07] WILLETT W., HEER J., AGRAWALA M.: Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1129–1136. 2