# The Effect of Multiple Visual Variables on Size Perception in Geographic Information Visualization

Yun Lin [iD], Chengqi Xue[✉], Yanfei Zhu, and Mu Tong

School of Mechanical Engineering, Southeast University, Nanjing 211189, China
`ipd_xcq@seu.edu.cn`

Abstract. Flow maps are a common form of geographic information visualization to show the movement of information from one location to another. Visual variables are essential elements of information representation. Previous studies have evaluated the perceived effectiveness of individual visual variables in a flow. However, the composition of flows usually includes multiple visual variables, and how to evaluate the usability of multiple visual variables has not been well documented. In this study, we investigated the influence of four visual variables (Thickness, Length, Brightness, and Hue) in flows on size perception. The results showed that length and brightness had a significant influence on the accuracy of size perception, while hue had no significance. Furthermore, multiple visual variables might involve a higher visual complexity and have a significant effect on response efficiency. In addition, we found that different references significantly affect the results of comparison judgments. The basic study could help improve the usability of flow maps in geographic information visualization.

Keywords: Flow maps · Multiple visual variables · Geographic information visualization

1 Introduction

Flow Maps are a common form of geographic information visualization. Flow maps geographically show the movement of information or objects from one location to another and their amount, see Fig. 1. Typically, flow maps are used to display migration data of vehicles, trade, and animals, etc. Qualitative and quantitative data information is conveyed with line symbols of different thicknesses. In practical use, complex geographic information is usually presented to the user in visualization using a combination of visual variables. For example, using shapes (straight lines and curves) to avoid overlap, color (hue) to distinguish between different types. However, these visual variables do not represent amount information, and it may interfere with the user's perceived performance. What we are interested in is exploring the influence of multiple visual variables on size perception in flow maps.

Shape (straight lines and curves), size (length and thickness), and color (brightness and hue) are three critical visual variables for encoding flow information [1, 2]. According

to the previous research [3] on the perceptual discriminability of visual variables, we know that size has the highest level of the number of perceptible steps. In flow maps, size is also the most commonly used variable to encode magnitude information. Besides, to visualize numbers across large magnitudes, visualization designers often redundantly represent information with size and color. In this study, we conducted two user studies to evaluate the effect of four common visual variables and reference in flow maps on size perception. The results of this study will provide concrete guidelines for visualization designers in the design of flow maps.
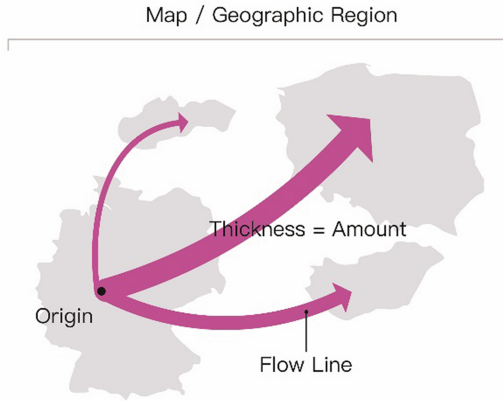
Fig. 1. The flow map

2   Background

2.1   Visual Variables in Flow Maps

The visual variable system was first established by Bertin [4]. He defined seven basic visual variables: position, size, shape, color brightness, color hue, orientation, and grain. On this basis, scholars have expanded the system, such as color saturation, arrangement [5], fuzziness, resolution, and transparency [6, 7]. Recent researches [8] even have gone beyond the traditional static 2D display, such as motion, depth, and occlusion.

There are five types of flow maps, which are distributive, network, radial, continuous, and telecommunications flow maps [9]. Size, color, and shape are three fundamental visual variables in flow maps. Dong et al. [10] evaluated the usability of flow maps through comparisons between (a) straight lines and curves and (b) line thickness and color gradients. Holten et al. [11] made an evaluation for the effectiveness of six directed-edge representations in path-finding tasks.

Though a lot of efforts to evaluate visual variables comprehensively, little attention has been paid to the interaction of multiple visual variables. Bertin's [12] notion of "disassociativity" and Garner's [13] concept of "dimensional integrality" pointed out that different visual variables are either "associative" or "disassociative." Bertin believed

that brightness and size are disassociative: because these variables affect the visibility of symbols, it would be tough to ignore their variations [12]. As we know, a flow consists of many visual variables, but not all of them are used for numerical representation. We are interested in whether the effect of multiple visual variables on magnitude perception is promotion or interference. For example, the length of flows varies depending on origins and destinations in practical use. When users compare the long and thick flow with the short and thin flow, will they make greater estimation error because of stronger visual contrast? Based on previous studies, we investigated the influence of four visual variables (Thickness, Length, Brightness, and Hue) in flows and their combinations on estimation. The study of multiple visual variables includes the following two aspects, redundant variables, and interference variables. Redundant variables are all encoded to represent values, while interference variables are independent of numerical representation.

2.2   Graphical Perception of Size

Graphical perception of size studied the psychophysical relation between perceived and physical magnitudes. Spence [14] experimentally explored the apparent and effective dimensionality of representations of objects. One of the most relevant studies on size perception is the visual cue. The visual cue is an essential factor in part-to-whole comparison [15]. The visual cue may be reference objects or invisible perceptual anchors. For reference objects in visualization perception, Steven's law [16] indicated that when an object is seen in the context of other larger objects, it appears larger itself. In contrast, it seems smaller. Jordan and Schiano [17] found that changing spatial separation between lines produced assimilation or contrast effect. Simkin and Hastie [18] ran studies that appear to confirm that people use perceptual anchors as part of the estimation process. Spence [19] held that the pie chart has four natural anchors at 0%, 25%, 50%, 75%, and 100%, while the stacked bar chart has two anchors at 0%, 50%, and 100%, respectively. Stephen [15] explored the impacts of visual anchors on estimation in part-to-whole comparisons employing Amazon's Mechanical Turk service.

In our investigation, we found that users would use different visual cues unintentionally to make an estimation. In the practical use of flow maps, users may estimate the value of the target flow by referring to the values of adjacent flows, the maximum or minimum values in the global, rather than comparing with the same reference every time. Inconsistent references might be one of the main factors leading to instability in user performance. We suspected that different reference objects might affect the user's estimated performance.

On the other hand, color mapping is a very important visualization technique. The color size effect [20] indicated that the color appearance is affected by the physical size of the color. A common explanation for this effect is that the cones and rods are not evenly distributed throughout the human retina, resulting in a difference in color vision between the fovea and the peripheral retina [20]. Changes in color size appearance are mainly determined by two factors, hue, and brightness. K Xiao [21] revealed the relationship of the changes of color appearance between different sizes through user experiments and found that with the increase of stimulus size, the color appears lighter and brighter. Tedford et al. [22] found that warm colors such as red, orange and yellow appear larger than cool colors like green and blue. Warm and bright colors make objects appear larger

and have a sense of expansion, while cold and dark have a sense of contraction and make objects appear smaller [23]. Visualization designers usually use the color size effect to obtain a visual balance, but it also may bring some potential perception problems. In this study, we studied the influence of length, brightness, and color on size perception based on previous studies.

3   Study 1: Effect of Multiple Visual Variables on Size Perception Processing

Based on the previous research, experiment 1 studied the characteristics of the individual size perception on flow encoded with multiple visual variables. The experiment adopted a within-subjects design. The independent variable was multivariate encoding type, including interference-free, interference, and redundant variable encoding. See Fig. 2 for more details about the levels of each factor.
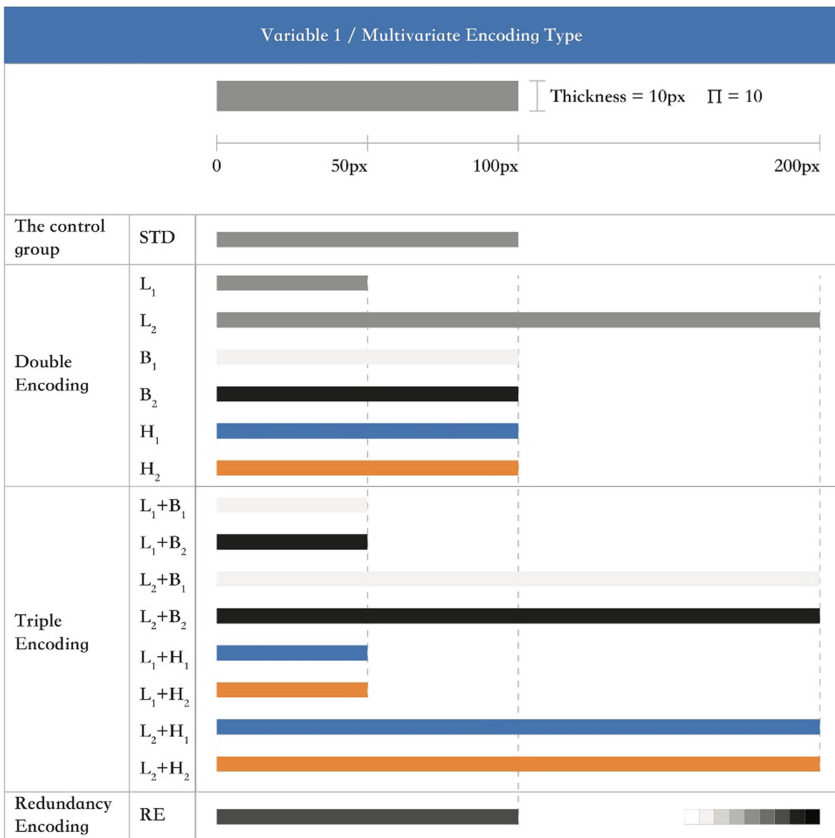

Fig. 2. Independent variables in experiment 1.

3.1   Participates

A total of 19 male and 29 female graduate students aged 23–27 (M = 23.3, SD = 2.6) were enrolled. The participants experienced one or more forms of information visualization on an occasional basis from sources such as the Internet, books, news-papers, and academic articles, etc. All participants had a normal or corrected vision, no color blindness or color weakness. Participants who completed the entire experiment would receive a reward of 30 RMB.

3.2   Experimental Materials

To avoid the influence of irrelevant factors in practical flow maps, such as place names, point size, and so on, we dealt with the experimental materials. In this basic research, we simplified the flow into a straight line with a length of 200 pixels as the standard stimulus. Thickness (2–20 px) of the standard stimulus represented corresponding values from one to ten.

For stimulus encodings with interference variables, there were two types, that is, Double encoding and Triple encoding. Double and Triple meant the number of visual variables that encoded the flow. For example, double encoding referred to the flow encoded with two visual variables (thickness and length), in which thickness represented magnitude information, and the length was an interference variable. The interference variables included length, brightness, and hue. There were two levels for each variable. For length, the short level was half the length of a standard stimulus, and the long level was twice the length of a standard stimulus. For brightness, the light and dark levels were two gradients higher and lower than the brightness of the standard stimulus, respectively. The brightness gradient picked from COLOR BREWER [24]. For hue, we selected only two representative colors, and the cold (#3b7494) and warm (#fd7f0a) levels were taken from the study of Tedford et al. [22].

For redundancy encoding, both thickness and brightness encode quantitative information. The ten brightness gradient was defined according to the HSL color space, and mapped values from one to ten.

3.3   Procedure

We performed a laboratory control web experiment consisting of a simple human-computer interaction process to collect behavioral data from participants. Web experiment has been widely used to evaluate the effectiveness of the interface in recent years. The experiment was carried out in the HCI lab of Southeast University under normal lighting condition (about 500 lx). The stimuli were generated by a computer running the Mac OS operating system with a 2.6 GHz Intel Core i7 processor. The monitor used was a 27-in. IPS monitor with a 4K resolution (Dell U2718Q). The viewing distance used was 50 cm.

The task for the participants was to estimate the value of the stimulus encoding ac-cording to the reference object and submit the answer. The experimental interface screenshot is shown in Fig. 3. Before the formal experiment, participants need to complete two practice trials to familiarize themselves with the functions and interactions

of the test interface. In the middle of the experiment, participants were allowed to rest for two minutes to stay relaxed. To avoid the familiarity effect, the order of the trials appeared random.

Participants needed to complete a total of $16 \times 4 = 64$ trails. The whole experiment took about 15 min. The following data were collected: the physical size for each trail (Π), the perceived size submitted by the participants (P), and the response time (T). Finally, we evaluated the effect of different combinations of visual variables on the usability of the flow maps based on the results.
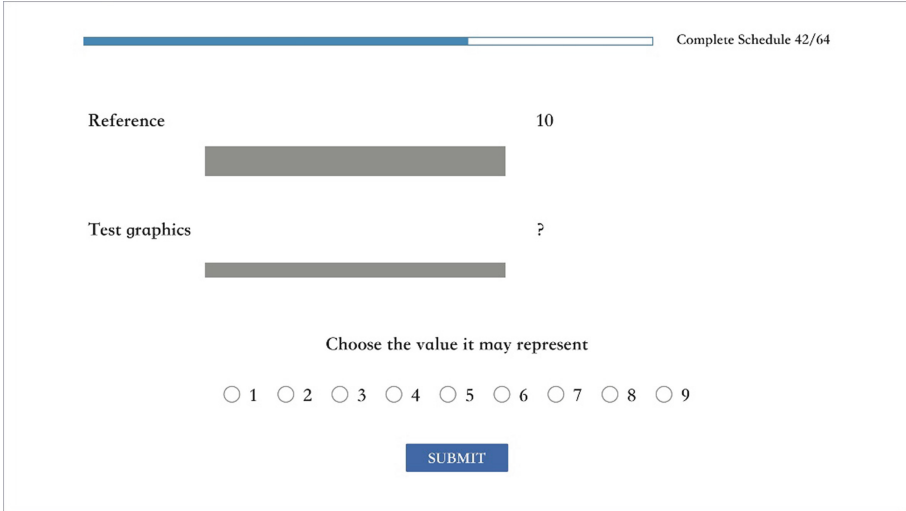

Fig. 3. Screenshot of the experimental interface.

4 Results

In this section, we described an overview of our analysis result. A total of 3072 groups of data (64 trials × 48 participants) were collected from the experiment. We measured both accuracy and response time for each trial. Accuracy percentage was measured by subtracting the percentage of response error from 100, where the response error is:

$$Response\ Error = \left| \frac{P - \Pi}{\Pi} \right| \times 100\%$$

4.1 Task Performance: Results Overview

To detect the effects of multiple visual variables, we turned to an analysis of variance (ANOVA). Before testing, we checked whether the data collected meets the assumptions of an appropriate statistical test. Shapiro-Wilk test showed that the residuals were close

to the normal distribution ($P > 0.05$), and the Levene test showed data of this study had equal variance ($P = 0.061$). In this section, we describe the results of statistical tests by independent variables and their interactions.

ANOVA detected significant main effects for both accuracy ($F_{(15,3008)} = 16.980$, $p < .001$) and response time ($F_{(15,3008)} = 14.234$, $p < .001$), and we followed up with Bonferroni-corrected post-hoc comparisons; see Fig. 4.
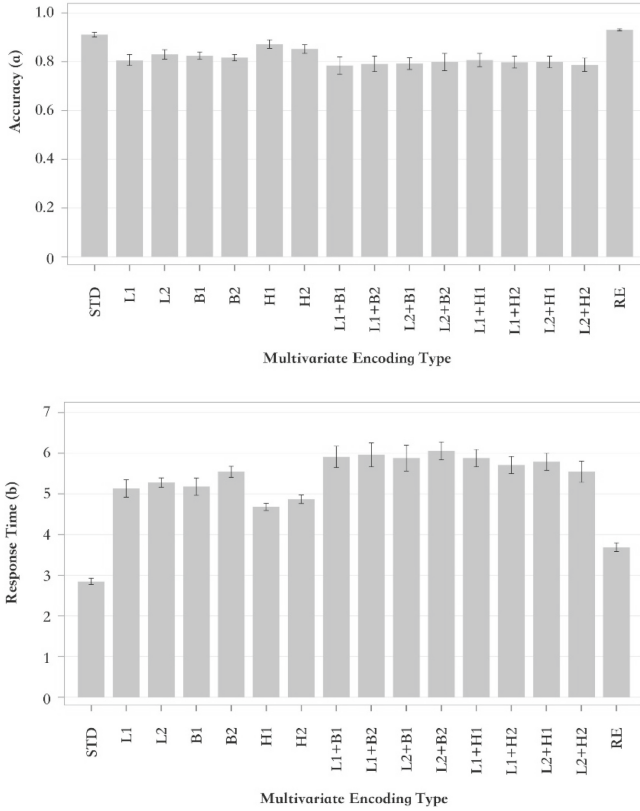
Figure 4 shows that the accuracy of STD ($M = 91.1\%$) and RE ($M = 93.1\%$) was the highest (all $p < 0.05$). The pairwise comparison between STD and RE showed no significant difference ($p = 0.347$). Compared with interference-free encoding (STD), the addition of the interference variable, except for hue, made a worse magnitude perception performance. Pairwise comparisons between encodings with interfering variables mostly did not detect significant differences. Except for hue, pairwise comparison identified no significant difference between triple encodings and double encodings; however, we noticed that the number of errors distributed in triple encodings was the most, and triple encodings had a more significant deviation.

For response time, the pairwise comparisons showed that encoding with interference-free variables (STD) had the shortest response time, and the addition of the interference variable made a longer response time. Redundancy affected the response time. Compared with STD, the response time of RE was significantly longer, but the amount exceeded was minimal. Triple encodings had a significantly longer response time than double encodings (all $p < .05$). See Fig. 4 for more details about the response time ranking of different encodings.

4.2 Estimation Bias Analysis

We ran chi-square tests to investigate whether users have a tendency of overestimation or underestimation to different multivariate encoding (See Table 1). For double encoding, the results indicated that the size perception of stimuli encoded by interference variables (length and brightness) were significantly biased. At the same time, the hue had no significant effect on the estimated bias. For triple encoding, the significances of the estimation bias were detected for all stimulus encodings except L1+B1 and L2+B2. This might be interpreted that there was a conflict when the length and brightness variables encoded stimuli simultaneously. For the variable combination of length and hue (L+H), the estimated biases are consistent with that of length in double encoding.

In general, encodings with significantly overestimated biases are L2, B1, L2+B1, L2+H1, and L2+H2; encodings with significantly underestimated biases are L1, B2, L1+B2, L1+H1, and L1+H2. Standard stimulus and redundant encoding had the smallest estimation bias and the best accuracy. Although these results showed that multiple visual variables affect biases for over- or underestimation of size perception, further researches will be needed to analyze the causes of these biases.

342   Y. Lin et al.

Fig. 4. Accuracy and response time of different multivariate encodings along with statistical results. Mean accuracy is shown in (a), and mean response time is shown in (b).

342   Y. Lin et al.

(a)

Test of Within-Subjects Effects
Multivariate Encoding Type($F_{(15,3008)}$ = 16.980, p < .001)

Pairwise Comparisons (ranked from most accurate to least)
p values are corrected using Bonferroni correction.

Encoding Type			
RE	v s .	L1, L2, B1, B2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H1, L1+H2, L2+H1, L2+H2	(p < .05)
ST D	v s .	L1, B1, B2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H2, L2+H1, L2+H2,	(p < .05)
H1	v s .	B1, B2, L1+B1, L1+B2, L2+B1, L1+H1	(p < .05)
H2	v s .	B2	(p < .05)

(b)

Test of Within-Subjects Effects
Multivariate Encoding Type($F_{(15,3008)}$ = 14.234, p < .001)

Pairwise Comparisons (ranked from most fastest to least)
p values are corrected using Bonferroni correction.

Encoding Type			
ST D	v s .	L1, L2, B1, B2, H1, H2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H1, L1+H2, L2+H1, L2+H2, RE	(p < .05)
RE	v s .	L1, L2, B1, B2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H1, L1+H2, L2+H1, L2+H2	(p < .05)
H1	v s .	B1, B2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H1, L1+H2, L2+H1, L2+H2	(p < .05)
H2	v s .	B2, L1+B1, L1+B2, L2+B1, L2+B2, L1+H1, L2+H2	(p < .05)
B1	v s .	L2+B2	(p < .05)

Fig. 4. Accuracy and response time of different multivariate encodings along with statistical results. Mean accuracy is shown in (a), and mean response time is shown in (b).

Table 1. The percentage of over- or underestimation when estimating different encodings. Chi-square tests compared the frequencies of overestimation and underestimation to detect estimation bias for different encodings. Significant differences are indicated by asterisks ($\star$).

Encoding	Over	Under	Chi-squareed test
STD	8.9%	7.8%	$\chi 2 = 0.107, p = 0.74$
L1$\star$	11.5%	26.6%	$\chi 2 = 8.448, p < 0.005\star$
L2$\star$	27.6%	12.5%	$\chi 2 = 7.887, p = 0.005\star$
B1$\star$	25.5%	13.5%	$\chi 2 = 5.110, p < 0.05\star$
B2$\star$	14.6%	26.0%	$\chi 2 = 4.442, p < 0.05\star$
H1	12.5%	8.3%	$\chi 2 = 1.326, p = 0.25$
H2	14.6%	7.8%	$\chi 2 = 3.221, p = 0.073$
L1+B1	15.1%	26.0%	$\chi 2 = 3.979, p = 0.05$
L1+B2$\star$	11.5%	27.1%	$\chi 2 = 8.892, p < 0.005\star$
L2+B1$\star$	25.5%	11.0%	$\chi 2 = 8.302, p < 0.005\star$
L2+B2	24.0%	14.1%	$\chi 2 = 3.601, p = 0.06$
L1+H1$\star$	13.0%	24.5%	$\chi 2 = 4.923, p < 0.05\star$
L1+H2$\star$	11.5%	22.4%	$\chi 2 = 4.172, p < 0.05\star$
L2+H1$\star$	23.4%	12.0%	$\chi 2 = 5.294, p < 0.05\star$
L2+H2$\star$	23.4%	11.5%	$\chi 2 = 5.900, p < 0.05\star$
RE	5.7%	6.8%	$\chi 2 = 0.148, p = 0.70$

5   Discussion

Experiment 1 compared the user performance of flows encoded by multiple visual variables (length, brightness, and hue). First of all, in terms of accuracy, accompanied by the addition of three visual variables, the results showed a decrease in effectiveness. When visual variables were added in the form of a combination, the accuracy decreased even lower. However, the accuracy of redundant encoding seemed improved. This result may partly be explained that redundant encoding could be used to improve discriminability between stimulus and reference while multiple visual variables increase graphic complexity [3].

After that, the results showed growth in response time when visual variables were encoded in the form of either interference or redundancy. Among them, triple encoding had the lowest efficiency, followed by double and redundant encoding. This supported Tufte's hypothesis to maximize the "data-ink ratio [25]." Tufte considered that Non-Data-Ink is to be deleted everywhere where possible to avoid drawing the attention of viewers of the data presentation to irrelevant elements. An explanation for this might be the information processing theory. Symbol excess and symbol redundancy increase graphic complexity [3]. Flows encoded by more visual variables take up more cognitive

resources (attention and understanding) of participants, further affect the efficiency of their response.

Last but not least, the results of estimation bias analysis showed that except hue, both length and brightness affected the estimation bias for size perception to varying degrees. The effect of length can be explained as its influence on the apparent thickness of flow. The increase in length changes the aspect ratio of flow lines, resulting in a thinner apparent thickness, which further affects the size perception of users. It should be noted that the present study was designed to determine the effect of multiple visual variables on size perception. Thus this study is a qualitative study of flow graph, not a quantitative study.

The results showed that the bright colors make the appearance of objects larger, while the dark colors are the opposite. This result supports evidence from previous observations [23].

The results showed not significant for the effect of hue on the estimation bias, which differs from the findings presented by Tedford et al. [22]. It seems possible that this result is due to our experimental materials are lines, not areas, and the stimulation intensity to participants is not as sufficient as previous studies.

Besides, the results showed that length and brightness have a compound effect on size perception. When they are coded together in different combinations, they can amplify or offset the estimation bias to varying degrees, as different visual variables have different capacities [4].

6   Study 2: Effect of Reference Objects on the Size Perception Processing

To investigate whether reference objects influence on size perception processing, we conducted a user study on the basis of experiment 1. The two factors of experiment 2 were reference objects and physical size. The former had two levels, and the latter had ten size levels, as shown in Fig. 5.

6.1   Participates

A total of 15 male and 25 female graduate students aged between 23 and 27 years (M = 24.1, SD = 2.4) who used computer almost every day were recruited. All participants had normal or corrected vision without color blindness or color weakness.

6.2   Experimental Materials

The experimental material was the same as experiment 1. Participants estimated the size of the standard stimulus based on different levels of reference objects. The physical size of reference objects in experiment 2 has two levels; the physical size of the large one was 10, and the small one was 2. The physical size of the stimulus in experiment 2 has ten levels, covering from 1 to 10.
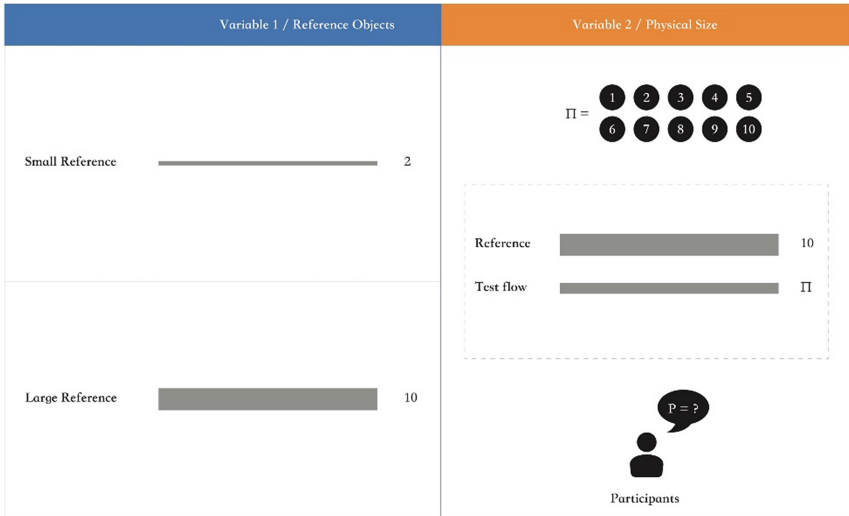
Fig. 5. Independent variables in experiment 2.

6.3 Procedure

The experimental procedure and the collected data were the same as experiment 1. Participants needed to complete a total of $2 \times 10 = 20$ trails.

7 Results

A total of 800 groups of data (20 trials $\times$ 40 participants) were collected from experiment 2.

7.1 Task Performance: Results Overview

We conducted a two-way ANOVA to investigate the effect of reference objects and physical size on size perception. The results showed as follow:

Reference Objects. There were significant differences between the two groups for both accuracy ($F_{(1,780)} = 10.095, p < .01$) and response time ($F_{(1,780)} = 20.677, p < .001$). The large reference group reported significantly more accuracy and shorter response time than the small reference group.

Physical Size. The tests detected a significant effect of physical size on response time ($F_{(9,780)} = 1.973, p = 0.044$), but no significant effect on accuracy ($F_{(9,780)} = 1.217, p = 0.281$).

Reference Objects $\times$ Physical Size. There was a significant interaction between reference objects and physical size orientation for both accuracy ($F_{(9,780)} = 9.318, p < .001$) and response time ($F_{(9,780)} = 8.516, p < .001$). The rankings of accuracy and response

time obtained by pairwise comparison are shown in Fig. 4. In terms of accuracy, the results indicated that the accuracy of the small reference group was significantly higher than that of the large group when physical size was 1 and 2. When the physical size was 7, 8, 9, and 10, the results were the opposite. Although not all pairwise comparisons between the large reference group were significant, it seemed the accuracy of the small reference group was poor when the physical size increased. At the same time, the large reference group was the opposite. Even so, the large reference group was more usable, since its overall accuracy was significantly better than the small reference group.

On the other hand, concerning response time, we found that the small reference group had significantly longer response time when physical size was 5, 7, 9, and 10. And the large reference group had a significantly longer response time when physical size was 1. In the small reference group, participants took longer to answer when the physical size was 5, 6, 9, and 10. The small reference group had a longer response time when the physical size was larger, while the large reference group seemed to have little difference in each physical size. See Fig. 6 for more details.

7.2   Estimation Bias Analysis

We conducted chi-square tests to check estimation bias with different reference objects (see Table 2). The results show that for the small reference object, participants significantly underestimated the size of the stimulus, starting from 5. For the large reference object, estimation bias only happened at 6, and the other comparisons were not significant.

### 7.3   Psychophysical Relations Between Perceived Size (P) and Physical Size (Π) for Flow Lines

To explore the Psychophysical relations between perceived size (P) and physical size (Π) for flow lines, we performed a linear regression analysis on user performance data. The linear relationship between P and Π in the small reference group can be expressed as:

$$P = K_S \times \Pi + b$$

The Small Reference Group. By least-square fitting of perceived size, the slope $K_S = 0.739$, intercept $b = 0.668$, and the fitting error $= 0.840$.

The Large Reference Group. By least-square fitting of perceived size, the slope $K_S = 0.989$, intercept $b = -0.072$, and the fitting error $= 0.948$.

As shown in Fig. 7, the perceived size of the small reference group was lower than the physical size, which may be evidence to support the results of previous studies by Steven et al. [16].
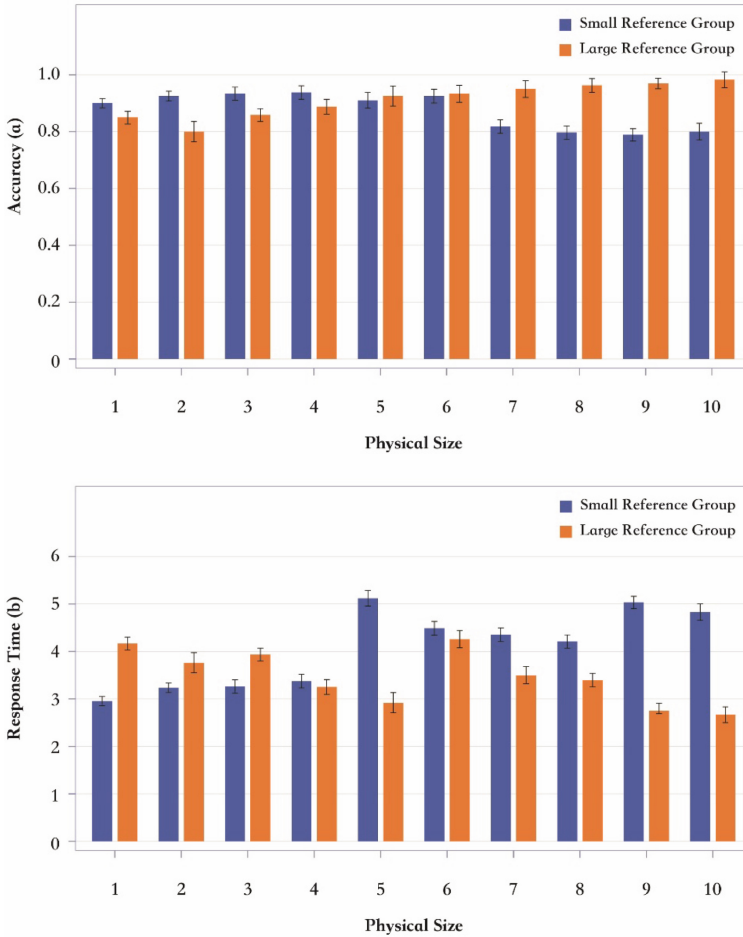
Pairwise Comparisons (ranked from most accurate to least)
p **values are corrected using Bonferroni correction.**

Physical Size ($F_{(1,780)} = 10.095$, p < .01)			
1,2	Small	> Large	(p < .05)
7,8, 9,10	Large	> Small	(p < .05)

Reference Objects ($F_{(9,780)} = 8.516$, p < .001)			
Small	2	8,9,10	(p < .05)
	3	8,9,10	(p < .05)
	4	8,9,10	(p < .05)
	6	8,9	(p < .05)
Large	10	1,2,3	(p < .05)
	5,6,7,8,9	1,2	(p < .05)

Pairwise Comparisons (ranked from slowest to fastest)
p **values are corrected using Bonferroni correction.**

Physical Size ($F_{(1,780)} = 20.677$, p < .001)			
1	Large	Small	(p < .05)
5, 7, 9, 10	Small	Large	(p < .05)

Reference Objects ($F_{(9,780)} = 1.973$, p = 0.044)			
Small	10	> 1,2,3,4	(p < .05)
	9	> 1,2,3,4	(p < .05)
	5	> 1,2,3,4	(p < .05)
	6	> 1	(p < .05)
Large	1	> 10	(p < .05)
	6	> 9,10	(p < .05)

Fig. 6. Accuracy and response time of different reference objects groups along with statistical results. Mean accuracy is shown in (a), and mean response time is shown in (b).

Table 2. The percentage of over- or underestimation when referring to different references. Significant differences detected by the Chi-square tests are indicated by asterisks ($\star$).

Physical size	Reference objects	Over	Under	Chi-squareed test
1	Small	65.0%	5.0%	$\chi2 = 0.049, p = .826$
	Large	30.0%	47.5%	$\chi2 = 0.899, p = .343$
2	Small	62.5%	10.0%	$\chi2 = 0.091, p = .763$
	Large	27.5%	42.5%	$\chi2 = 0.762, p = .383$
3	Small	67.5%	7.5%	$\chi2 = 0.041, p = .839$
	Large	25.0%	45.0%	$\chi2 = 1.363, p = .234$
4	Small	60.0%	5.0%	$\chi2 = 0.153, p = .695$
	Large	22.5%	30.0%	$\chi2 = 0.282, p = .596$
5	Small	57.5%	10.0%	$\chi2 = 4.333, p < .05\star$
	Large	15.0%	42.5%	$\chi2 = 3.445, p = .063$
6	Small	47.5%	15.0%	$\chi2 = 8.680, p < .01\star$
	Large	45.0%	12.5%	$\chi2 = 8.873, p < .05\star$
7	Small	25.0%	20.0%	$\chi2 = 12.692, p < .001\star$
	Large	27.5%	12.5%	$\chi2 = 1.626, p = .202$
8	Small	20.0%	22.5%	$\chi2 = 12.433, p < .001\star$
	Large	22.5%	25.0%	$\chi2 = 0.036, p = .850$
9	Small	10.0%	12.5%	$\chi2 = 9.716, p < .01\star$
	Large	20.0%	22.5%	$\chi2 = 0.041, p = .839$
10	Small	20.0%	17.5%	$\chi2 = 13.823, p < .001\star$
	Large	12.5%	15.0%	$\chi2 = 0.071, p = .789$
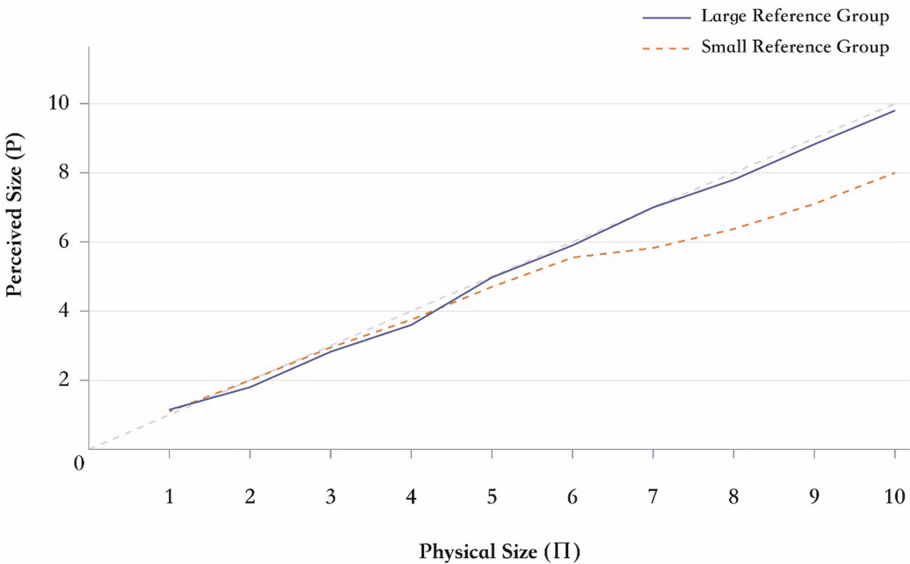
Fig. 7. Psychophysical relations between perceived size (P) and physical size (Π) for flow lines.

8   Discussion

Experiment 2 compared user performance in size perception when referring to different reference objects. Initially, the results showed that the performance in the large reference group was better in terms of accuracy and response time. A possible explanation for this might be that humans are better at estimating the proportion within 100%, because of the assistance of natural anchors [19], such as 25%, 50%, and 75%. However, when participants perform a proportion estimation in which the stimulus-to-reference ratio exceeds 100%, there might be two sources of error. On the one hand, if participants make a whole-to-part comparison, they need to make multiple estimations, and the error superposition may magnify the error. On the other hand, if participants make a part-to-whole comparison, more complex conversions and higher accuracy requirements seem to be factors that cause more significant errors. Both situations may cause errors. However, with a small sample size, caution must be applied, as the findings might not be the evidence that large references perform best. There is a possibility that a reference performs best when the ratio of it to the maximum value falls somewhere in the middle, such as 50%. We will continue to study the effect of different proportions of reference and maximum value on size perception in our future work.

Besides, we found that the small reference group had significant underestimation bias in size perception compared with the large reference group. This finding is consistent with that of Steven [16], who indicated that when an object is seen in the context of other smaller objects, it appears smaller itself. This also accords with the earlier observations of Jordan and Schiano [17], which showed that increasing spatial separation between lines produced the opposite effect.

9   Conclusion

In this study, two experiments were conducted to investigate the effect of multiple visual variables on size perception in flow maps, and the following conclusions were drawn:

1. Different lengths of flow involve the judgment of apparent thickness and then affect size perception. However, what aspect ratio has the most significant effect on size perception requires further quantitative research.
2. As a result, it was detected that brightness had an influence on the estimation bias, while hue had no effect.
3. Redundant encoding of flows improves the accuracy to some extent without losing much efficiency at the same time.
4. A comparison judgment with varied references in flow maps may cause higher errors. However, further research is needed to find out the best proportion of reference for comparison judgment.

Some suggestions for the geographic information visualization designer:

When the value represented by the flow across varying magnitudes, redundant coding can be selected for its better discriminability. Since different references will cause greater errors, designers should guide the user's estimation behavior and recommend a uniform

reference. Although the structure of the flow graph determines that some visual variables, such as length and lightness, cannot be omitted, their impact should be considered when designing visualizations of crucial information.

References

1. Bertin, J.: Graphics and Graphic Information Processing. Walter de Gruyter, Berlin (1981)
2. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nat. Rev. Neuros **5**, 495–501 (2004)
3. Moody, D.: The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Trans. Softw. Eng. **35**(6), 756–779 (2009)
4. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. Univ. Wisconsin Press, Madison (1983)
5. Morrison, J.L.: A theoretical framework for cartographic generalization with the emphasis on the process of symbolization. Int. Yearb. Cartogr. **14**(1974), 115–127 (1974)
6. MacEachren, A.M.: Visualizing uncertain information. Cartograph. Perspect. **13**(13), 10–19 (1992)
7. MacEachren, A.M.: How Maps Work: Representation, Visualization, and Design. Guilford Press, New York (2004)
8. Carpendale, M., Sheelagh T.: Considering visual variables as a basis for information visualisation (2003)
9. Jenny, B., et al.: Design principles for origin-destination flow maps. Cartogr. Geogr. Inf. Sci. **45**(1), 62–75 (2018)
10. Dong, W., et al.: Using eye tracking to evaluate the usability of flow maps. ISPRS Int. J. Geo-Inf. **7**(7), 281 (2018)
11. Holten, D., et al.: An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In: 2011 IEEE Pacific Visualization Symposium (2011)
12. Reimer, A.: Squaring the circle? Bivariate colour maps and Jacques Bertins concept of disassociation. In: Proceedings of the International Cartographic Conference 2011, pp. 3–8 (2011)
13. Garner, W.R.: The Processing of Information and Structure. Psychology Press, New York (2014)
14. Spence, I.: The apparent and effective dimensionality of representations of objects. Hum. Factors **46**(4), 738–747 (2004)
15. Redmond, S.: Visual cues in estimation of part-to-whole comparisons. In: 2019 IEEE Visualization Conference (VIS). IEEE (2019)
16. Stevens, S.S.: On the psychophysical law. Psychol. Rev. **64**(3), 153 (1957)
17. Jordan, K., Schiano, D.J.: Serial processing and the parallel-lines illusion: length contrast through relative spatial separation of contours. Percept. Psychophys. **40**(6), 384–390 (1986). https://doi.org/10.3758/BF03208197
18. Simkin, D., Hastie, R.: An information-processing analysis of graph perception. J. Am. Stat. Assoc. **82**, 454–465 (1987)
19. Spense, I.: No humble pie: the origins and usage of a statistical chart. J. Educ. Behav. Stat. **30**, 353–368 (2005)

20. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd edn. Wiley, New York (2000)
21. Xiao, K., et al.: Investigation of colour size effect for colour appearance assessment. Color Res. Appl. **36**(3), 201–209 (2011)
22. Tedford, W.H., Berquist, S.L., Flynn, W.E.: The size-color illusion. J. Gen. Psychol. **97**, 145–150 (1977)
23. Marks, L.E.: The Unity of the Senses: Interrelations Among the Modalities. Academic Press, Cambridge (2014)
24. COLOR BREWER Homepage. http://colorbrewer2.org. Accessed 21 Nov 2019
25. Tufte, E.R.: The Visual Display of Quantitative Information Graphics Press. Connecticut, Cheshire (1983)