Sorting Visual Complexity and Intelligibility of Information Visualization Forms

Mingran Li(✉), Wenjie Wu, Yingjie Victor Chen, Yafeng Niu,
and Chengqi Xue

Department of Computer Graphics Technology,
Purdue University, West Lafayette, USA
`li1940@purdue.edu`

Abstract. Faced with the challenge of information explosion, almost everyone have been exposed to some kind of information visualization in varies of forms. Understanding how people read, understand, interpret and distinguish various forms of visualizations helps designers and developers think about how to improve the designs from the perspectives of users. This paper applied a mixed research method of quantitative and qualitative to explore how the designs of visualization forms evolved, and whether those kinds of graphs and charts are easy for users to understand, and how much information the users can get from the visualization. By testing if users are able to easily and accurately reach the information and providing the scales of simple to complex, and easy to hard, we see that basic visualizations like bar, pie, bubble, line, and scatter charts have been distributed in areas which are relatively simple in design and easy to read. Nonetheless, visualizations like the tree, parallel coordinate, sunburst, heat map, box plot and Sankey graphs have been concentrated in the regions of relatively complex in design, and are difficult to understand. In addition, the visualizations, including stacked bar, word cloud, box plot, and theme river that frequently appeared in the middle region of the grid, embodied the transitions of visualization design from simple to complex, and easy to hard.

Keywords: Visual Literacy · Visual complexity · Intelligibility · Readability · Sorting · Visualization distributions

1 Introduction

In the era of information explosion, it is inevitable that everyone will be exposed to some kinds of data visualizations, even if they are non-professionals in the field of information visualization (InfoVis) or fields related to information science. In the research on InfoVis, designers try to present information by utilizing different kinds visual forms of charts, graphs, and diagrams, providing readers quick interpretation, visible outliers, and insightful explorations [1]. Information visualization is both science and art. It has solid scientific foundation especially in human perception and cognition, such as preattentive visual processing, Gestalt

laws of perception, and color perception theories. Preattentive processing refers to an initial organization of the visual field based on cognitive operations believed to be rapid, automatic, and spatially parallel [2,3]. Preattentive processing has been well utilized in visualization designs, to enable intuitive high-speed target detection, boundary identification, and region detection [3]. Gestalt principles were developed to help describe and explain the rules of the organization of relative complex visual fields [4,5]. Color perception is additional main research direction in the visualization field. Color selection in data visualization is not merely an aesthetic choice, it is a crucial tool to convey quantitative information [6]. Information visualization is also a design. The same data can be represented in different forms with different colors. How to properly select the right form, right layout, and right colors to convey the underlying information accurately, and create better understandings for users are always challenging and as ongoing research topics for many researchers.

Visual Literacy (VL) is the ability for a human to interpret and make sense from information presented in visual forms. By applying the concept of Visual Literacy, researchers put forward the visual design and composition principles, including clear indication of the nature of the relationship, accurate representations of the quantities, comprehensible comparisons of quantities, obvious hierarchy of values, etc. [7], to evaluate and explore whether the designs achieve the goals. Those principles are of fundamental importance for the production of effective visual instructional material [8].

By referring the past research, we brought forward a new method in this paper that sorting the information visualization forms by visual complexity and intelligibility. In other words, participants would conduct visualization distributions that help with researchers to study what kinds of graphs are easy for users to comprehend and interpret, how accurately the users are able to obtain the information, and how the designs of visualization charts evolved. It is beneficial for users to realize how the understandings they have for each kind of visualization by applying the standards of simple to complex, and easy to hard. Furthermore, the finding of threshold of chart evolution would help designers and developers think about how to improve the designs from the perspectives of user-centered [9]. Reading information visualization as one of the 21st century important skills, inspires designers to study harder on how to match tools with users of different age and knowledge background, tasks and real problems.

We used both quantitative and qualitative research methods to distinguish the readability and intelligibility of each type of chart. There a large amount of results have been collected from total 20 participants' tasks and interviews. By readability, we mean to test if users can obtain the information that's delivered from the charts accurately and easily, which was the important step for our team to verify whether the readers fully understand the graphs. By providing them with scales of simple to complex, and easy to hard, users were able to independently create a clear distribution based on their judgments of the complexity of the charts, and corresponding simplicity of reading, that is the readability and intelligibility in their minds. Through integrating gathered data, we could

explore the reasons why charts found themselves on either side of the complexity threshold, indicate how the complexity threshold looks like in the evolution of graphs, as well as what the common features the threshold has.

The current result shows that visualizations like bar, pie, bubble, line, and scatter charts have been distributed in areas which are relatively simple in design and easy to read. However, visualizations like the tree, parallel coordinate, sunburst, heat map, box plot and Sankey graphs have been concentrated in the regions of relatively complex in design, and are difficult to understand. In addition, the built distributions from most of the participants showed the charts like stacked bar, word cloud, box plot, and theme river were transitions that made users' readings and understandings changed from easy to difficult, and simple to complex.

In the rest of this paper, we present a brief overview of previous research, summarizing the key reasons why sense making problems of familiar and unfamiliar visualizations had important research significance for non-professional areas and readers, and highlighting the institutions of current literature on visual literacy technologies. We then present our study, the procedures of data collection and analysis, and the primary research results from interviewing and observing participants, concluding with a discussion of the implications of such experiment as a research base for the subsequent study of the sense-making problems of familiar and unfamiliar interactive visualizations.

2   Previous Research

Much of the previous research focused on exploring sense-making problem with data through the process of visualization. Past researchers used a variety of methods to investigate and assess the visualization literacy, including the visualizations that people are familiar or unfamiliar with [10,11]. According to Catherine's studies, the challenges of visualization included: how to match tools with users, tasks and real problems, how to improve user testing, including looking categorically at the same data from different perspectives, answering questions participants didn't know they had, factoring in the chances of discovery and the benefits of awareness, and addressing universal usability [12], which were not only for adults, but also for younger groups, i.e. secondary school students [13]. Therefore, Visual Literacy was defined as a significant ability within the scope of 21st century skills [13–15].

Visual Literacy (VL) was first proposed in 1969 by John Debes, which was mentioned by Avgerinou and Ericson in the article of a review of the concept of Visual Literacy [8,16]. In that article, the authors also defined the concepts of Visual Literacy by referring to other researchers' statements, such as "Visual Literacy is the ability to understand (read) and use (write) images and to think and learn in terms of images, i.e., to think visually" [8,17], "Visual Literacy refers to a group of vision-competencies a human-being can develop by seeing, and at the same time having and integrating other sensory experiences. The development of these competencies is fundamental to normal human learning.

When developed, they enable a visually literate person to discriminate and interpret the visible actions, objects, symbols, natural or man-made, that he encounters in his environment" [8,16], and Sinatra's proposition about Visual Literacy should be considered as a prerequisite indispensable to human thinking [8,18].

To be more specific, an early 2016 article investigated how people make sense of unfamiliar visualizations by applying a grounded model of NOVIS [19]. Sukwon put forward NOVIS model consisting of five cognitive activities, including encountering visualization, constructing a frame, exploring visualization, questioning the frame, and floundering on visualization. It emphasizes how the users complete these five activities based on parallel coordinate, chord diagram, and tree map, and observes how participants express their feelings and opinions about impressions. In 2014 and 2015, Boy and Borner used different research methods to investigate and assess the issue of visualization literacy, respectively. Boy focused on building a set of visualization literacy tests for line graphs, bar charts, and scatter plots [20]. He developed the method based on Item Response Theory (IRT) and conducted six specific tasks to get the scores from participants. The authors obtained the most accurate characteristic values for each item according to finding the best variant of the model. Borner found that people were more familiar with basic charts, maps, and graphs, but very few were familiar with network by conducting experiments with 20 information visualizations and 273 science museum visitors [21].

With reference to these previous studies, our team wants to improve upon existing research by conducting a study which involves a more varied range of visualization charts, and demonstrates whether the information the users get is correct on the basis of understanding.

3   Research Questions

Although numerous studies helped us with analyzing and exploring how people read and understand visualization charts, our team not only wants to improve upon existing research by conducting a study which incorporates a more diverse range of visualizations but also indicate the complexity threshold in the evolution of charts, by which we mean the phase of chart development beyond which each chart category only serves to confuse users. In addition, our study explored reasons why charts found themselves on either side of the complexity threshold. Faced with this problem, we tried to produce a useful and understandable visual evaluation study to explore questions such as

– How much information do people obtain accurately, including topics, values, and relationships, etc.?
– How does the visualization distribute based on readability and complexity in users' minds?
– How the complexity threshold in the evolution of charts is indicated? What regular pattern brings out?

And, to help users fully understand the levels of their cognition of each type of chart through dragging and constructing the charts' distributions in the scale grid. The expectation of our research is to get specific experimental results such as

– Based on each chart, how many people (percentage) can/cannot make sense of the chart?
– If he/she can make sense some charts, he/she will go with the specific tasks that we have designed. But not all the answers are correct.
– If he/she can make sense some charts, he/she will go with the specific tasks that we have designed. And, all the answers are correct.
– On the basis of the judgements of visual complexity and intelligibility, how will the participants create charts' distributions?

4   Mixed Research Method of Quantitative and Qualitative

4.1   Participants

Our team focused on collecting 20 participants to take part in the study, and each participant would be asked to provide diverse demographic background, including age, gender, education, and profession. All of our participants had a basic computer operation capability. The majority of our participants are students who came from Purdue University (West Lafayette, IN), representing 10 different majors, and their educational levels scattered from Freshmen to Graduated. Participants in groups also involved several professors who came from the programs of Computer Graphics Technology, and Art & Design at Purdue University.

Using Autodesk Sketch Pro as a fundamental tool, the participants created the distributions with forms by pulling and dragging each graph. Before that sorting, the participants would be asked to make all the visualizations classified that depended on their readings and understandings of each chart. The participants would work up to 2 hours to complete the experiments.

4.2   Data Collection

4.2.1   Experimental Questions

Each participant was provided 54 static visualizations (forms/charts/graphs) as the experimental elements (Fig. 1), and all the gathered visualizations were with full labels in order to help with readings and understandings. Figure 1 shows partial image resources. Name or title of each image file didn't affect the process of the experiment. Based on reading each visualization, the participants were asked to decide whether they could make sense of it within 3 min, and if they could understand, they would be required to answer several particular questions in order to verify whether they got the information accurately. The bullets and the mind map (Fig. 2) below showed the specific experimental process.

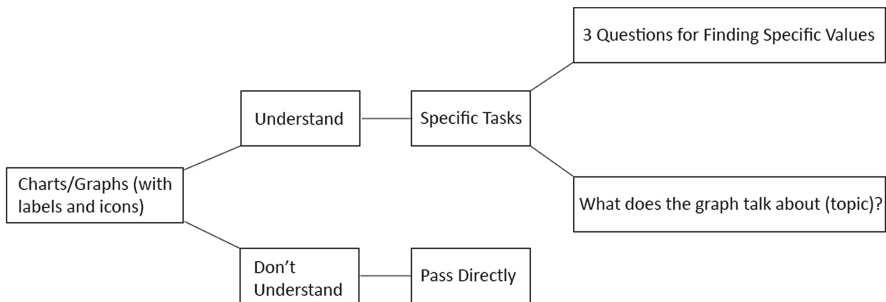| Visualization Name	Visualization	Visualization Name	Visualization
03_Line_Chart_3		23_3D_Pie_Chart	
04_Bar_Chart		24_Donut_Chart	
05_Multi-Set_Bar_Chart_1		25_Multiple_Donut_Chart	
06_Multi-Set_Bar_Chart_2		26_Sunbrust_Diagram	
07_Variant_Bar_Chart		27_Arc_Diagram	
08_Radial_Bar_Chart		28_Scatter_Plot	
09_Radial_Column_Chart		29_Multiple_Scatter_Plot	
10_Spiral_Plot		30_Bubble_Chart	
11_3D_Bar_Chart		31_Bubble_Map	
12_Stacked_to_Grouped_Bar_Chart		32_Heat_Map	
13_Normalized _Stacked_Bar_Chart		33_Choropleth_Map	
14_Marimekko_Chart		34_Area_Chart	
15_Nightingale_Rose_Chart		35_Flow_Chart_1	
16_Span_Chart		36_Flow_Chart_2	
17_Diverging_Stacked_Bar_Chart		37_Flow_Chart_3	
18_Bullet_Graph		38_Parallel_Sets	
19_Waterfall_Chart		39_Sankey_Diagram	
20_Box_Whisker_Plot		40_Bidirectional_Hierarchical_Sankey_ Diagram	

Fig. 1. Visualization forms

Fig. 2. A mind map shows the experimental process

1. Do you think you understand this visualization? If no, skip to next visualization. If yes, continue.
2. Tell us the meaning of this visualization. The participants will verbally explain their interpretations of the visualization. Specifically, what does the graph talk about (topic/Q1)? Under the premise of answering the questions accurately, participants can sum up the theme easily and neutrally.
3. If the visualization encodes some special information, we will ask two or three specific questions related to the visualization, for example:
   – What is the relation between A & B (Q2)?
   – What is the trend of X in recent Y years (Q3)?
   – What is the meaning of the peak value, and why there (Q4)?

We recorded the answers and interactions from each participant, and filled in the following table (Fig. 3). Researchers would help the participants to mark understand or not-understand with Y or N at first, then note the accuracy with correct, incorrect, or partially correct, which was for the Q1 and Q4. In addition, the descriptions of participants about topics were transcribed for researchers to verify if the users fully made sense of the main idea.

Visualization No.	Understand or Not (Y/N)	If Understand									
		Q.1.			Q.2.		Q.3.		Q.4.		
		Correct	Incorrect	Partially Correct	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Partially Correct

Fig. 3. Table for collecting participants' answers

4.2.2   Visualization Sorting

We were asking the participants to rank these visualizations by visual complexity and intelligibility in their minds. By applying Autodesk Sketch Pro, we conducted a blank grid for them to build their own visualization distributions. Those 54 visualizations provided above would be pulled and dragged into the grid area, even if there might be overlap. The grid had been set up by two axises, which specifically presented the levels from simple to complex (the complexity of the charts) of X axis, and the degrees from easy to hard (the ease of reading and understanding) of Y axis.

The initialization of the task was to present the 54 visualizations scattered on the left side of the grid. Then, the participants were asked to complete distributions by how they understand each visualization (Fig. 4). Moreover, we recorded the text descriptions of how the participants described, interpreted and thought

Fig. 4. Participants created visualization distributions by dragging and pulling forms into grid

about each visualization, and the reasons they dragged and took out the visualization to a particular location.

We processed the data analysis for two aspects: one was to verify the correctness of participants' answers, and another one was to transcribe and convert their verbal descriptions that could be used for data analysis. For example, we translated participants' narrations into text files, and presented them with quotes.

"This kind of basic bar chart is easy to understand for me. But if it is added up, the stacked bar chart feels more difficult. One more problem is it's hard for me to actually compare the height of the bar to the corresponding values on the left (Y axis)."

5 Findings

Over 400 hours of data were collected from the 20 users. The interviews with participants, and observations of their actions, revealed some prevailing patterns of visualization distribution and chart evolution threshold. Below we present a discussion of three themes that emerged in our exploration of visualization sorting and distribution studies.

5.1 Answer Accuracy

We used a total of 1080 answer tables in the statistical analysis to record answers of participants, and applied accuracy rate. Figure 5 showed the answers of numerical problems and the interpretations of topics from visualization No. 41. The users who had never been trained to read the relatively complex visualizations of parallel coordinates, could not interpret the meaning of such intensive lines in a short period of time. Moreover participants found the exploration of relationships between each line, and the nodes on each line, difficult to understand.

Visualization No.	Understand or Not (Y/N)	If Understand										
		Q.1.			Q.2.		Q.3.		Q.4.		
41		Correct	Incorrect	Partially Correct	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Partially Correct
	N			✔	✔		✔				✔

Fig. 5. Answer records - visualization No.41

By contrast, the No. 22 visualization obtained a better result in user response. Pie charts are common in daily life; individuals are trained to read, explain, and interpret proportions and distributions of the chart when they were young. Combined with the simple graphic design and years of accumulated knowledge, most of the participants were able to identify the topic issue and verify the correct answers successfully (Fig. 6).

Visualization No.	Understand or Not (Y/N)	If Understand											
		Q.1.			Q.2.		Q.3.		Q.4.			
		Correct	Incorrect	Partially Correct	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Partially Correct	
22	Y	✓			✓		✓		✓			

Fig. 6. Answer records - visualization No.22

5.2   Visualization Distribution

We know from literature that Visual Literacy (VL) as defined by past research is the capability to read, understand, interpret, and make meaning from information presented in the form of images. Our research applied a method by asking users to construct a distribution of visualizations to study and present users' abilities in respect of VL. We have provided users a tool, Autodesk Sketch Pro, to construct their own visualization periodically according to their readability and intelligibility in their minds. After an analysis of completed research we understand how human perception, cognition, and particular mental models work on readings, and understandings, of the visualization. As we can see (Fig. 7) most of bar, pie, bubble, line, and scatter charts have been distributed in the areas that were easy to read because of relatively simple design. However, the majority of the graphs, including the tree, parallel coordinate, sunburst, heat map, box plot and Sankey have concentrated in regions of relative complexity in design and are more difficult to understand.
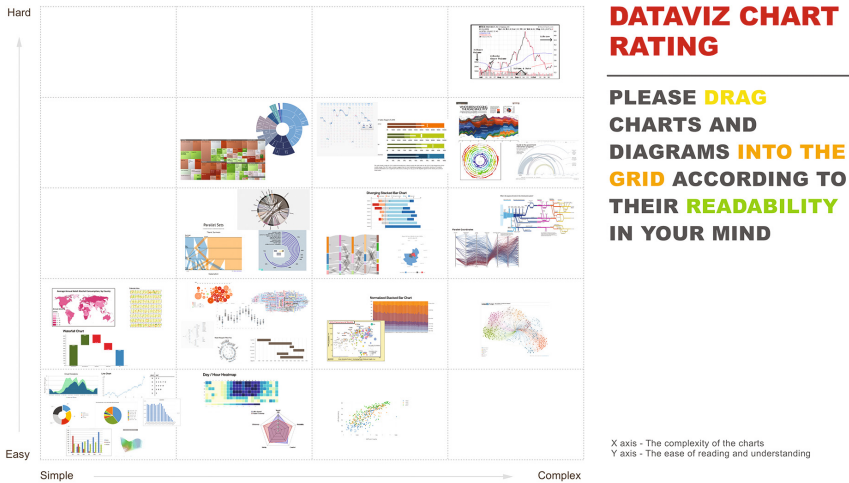

Fig. 7. Visualization distribution

5.3   Chart Evolution

Through integrating gathered data, we could indicate how the threshold looks like in the evolution of graphs, as well as what the common features the threshold has. The builded distributions from most of the participants showed the charts, such as stacked bar, word cloud, box plot, and theme river were regarded as the transitions that made users' readings and understandings changed from easy to difficult, and simple to complex. Figure 8, which congregated the ideas from a majority of participants, gave out an obvious comparison, and a relative expression of evolutionary thresholds.

The reason why most of users put these visualizations in the transitional zone, and treated them as thresholds was those visualizations expressed the kind of the same ideas by more innovative ways. There was a greater difference between these designs and basic knowledge in their brains. Several relevant quotes from the participants:

"Why I put the word cloud in the middle area because this is my first time to see it. I know the visualization wants to express a topic that relates to the words, or texts. But it's a new form so that I can not make an interpretation."

"I have an idea about how to read the line graphs. But this visualization, which is called'theme river' seems be composed with thousands of lines. Then, I don't know how to read that."
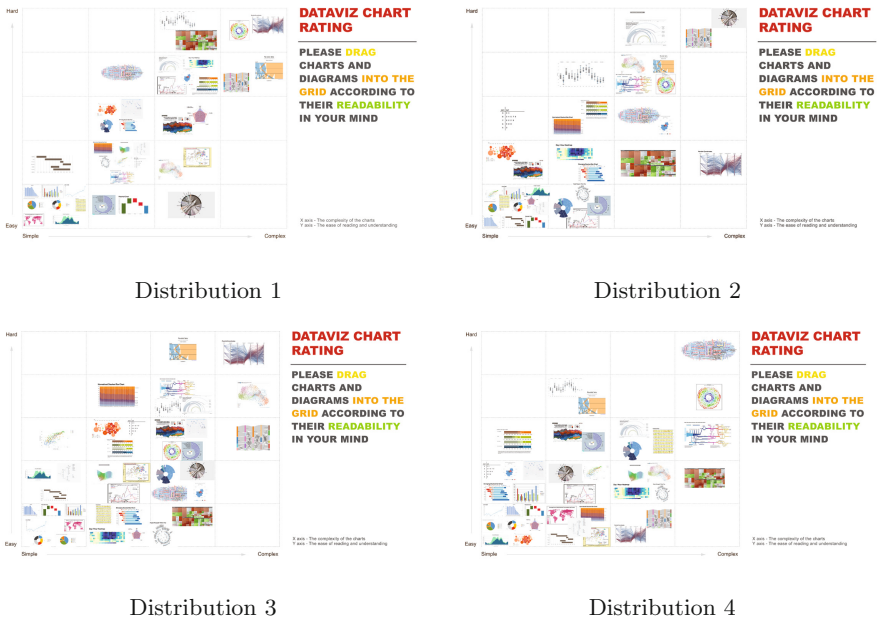

Distribution 1

Distribution 2

Distribution 3

Distribution 4

Fig. 8. Visualization distributions with thresholds

6   Conclusion and Future Work

In this research, we investigated how the visualization design evolved, which specifically measured how the participants sorted the visualizations and built the distributions, and how they thought about the threshold issues based on their readings and understandings. The findings showed that visualizations like bar, pie, bubble, line, and scatter charts have been distributed in areas which are relatively simple in design and easy to read. Contrastively, visualizations like the tree, parallel coordinate, sunburst, heat map, box plot and Sankey graphs have been concentrated in the regions of relatively complex in design, and are difficult to interpret.

There were many past literatures also mentioned interaction plays a very important role in creating a good design of a visualization chart. Designers have always focused on how to better use interactive methods to help users read and understand charts. Based on this, we will consider a more sophisticated way that involves interactive charts in the studies in the subsequent stage. In addition, we will explore in depth the process participants undergo to obtain accurate information, which specifically means how do they judge the topics, values, and relationships through reading and interpreting the visual elements.

References

1. Denovo Group. (n.d.). The Importance of Visualizations. http://www.denovogroup.com/2016/06/02/the-importance-of-visualizations/. Accessed 15 October 2016
2. Healey, C. G.: Perception in visualization (2007). Accessed 10 February 2008
3. Healey, C.G., Booth, K.S., Enns, J.T.: High-speed visual estimation using preattentive processing. ACM Trans. Comput. Hum. Interact. (TOCHI) **3**(2), 107–135 (1996)
4. Nesbitt, K.V., Friedrich, C.: Applying gestalt principles to animated visualizations of network data. In: Proceedings of Sixth International Conference on Information Visualisation, pp. 737–743. IEEE (2002)
5. Todorovic, D.: Gestalt principles. Scholarpedia **3**(12), 5345 (2008)
6. Silva, S., Santos, B.S., Madeira, J.: Using color in visualization: a survey. Comput. Graph. **35**(2), 320–333 (2011)
7. Few, S.: Data visualization for human perception. In: Soegaard, M., Dam, R. (eds.) The Encyclopedia of Human-Computer Interaction. Interaction Design Foundation, Aarhus (2013)
8. Avgerinou, M., Ericson, J.: A review of the concept of visual literacy. Br. J. Educ. Technol. **28**(4), 280–291 (1997)
9. Cognitive testing interview guide. https://www.cdc.gov/nchs/data/washington_group/meeting5/WG5_Appendix4.pdf
10. Abilock, D.: Visual information literacy: reading a documentary photograph. Knowl. Quest **36**(3), 7–14 (2008)
11. Carpenter, P.A., Shah, P.: A model of the perceptual and conceptual processes in graph comprehension. J. Exp. Psychol. Appl. **4**(2), 75 (1998)

12. Plaisant, C.: The challenge of information visualization evaluation. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 109–116. ACM, May 2004

13. Kibar, P.N., Akkoyunlu, B.: Searching for visual literacy: secondary school students are creating infographics. In: Kurbanoğlu, S., Boustany, J., Špiranec, S., Grassian, E., Mizrachi, D., Roy, L. (eds.) ECIL 2015. CCIS, vol. 552, pp. 241–251. Springer, Cham (2015). doi:10.1007/978-3-319-28197-1_25

14. Nuhoğlu Kibar, P., Akkoyunlu, B.: A new approach to equip students with visual literacy skills: use of infographics in education. In: Kurbanoğlu, S., Špiranec, S., Grassian, E., Mizrachi, D., Catts, R. (eds.) ECIL 2014. CCIS, vol. 492, pp. 456–465. Springer, Cham (2014). doi:10.1007/978-3-319-14136-7_48

15. https://github.com/INRIA/Visualization-Literacy-101

16. Fransecky, R.B., Debes, J.L.: Visual Literacy: A Way to Learn–A Way to Teach. Association for Educational Communications and Technology, Washington, D.C. (1972)

17. Hortin, J.A.: Visual Literacy and Visual Thinking (1980)

18. Sinatra, R.: Visual Literacy Connections to Thinking, Reading and Writing. Charles C. Thomas, Springfield (1986). 2600 South First St., PO Box 4709, IL 62708–4709

19. Lee, S., Kim, S.H., Hung, Y.H., Lam, H., Kang, Y.A., Yi, J.S.: How do people make sense of unfamiliar visualizations? A grounded model of novice's information visualization sensemaking. IEEE Trans. Vis. Comput. Graph. **22**(1), 499–508 (2016)

20. Boy, J., Rensink, R.A., Bertini, E., Fekete, J.D.: A principled way of assessing visualization literacy. IEEE Trans. Vis. Comput. Graph. **20**(12), 1963–1972 (2014)

21. Brner, K., Maltese, A., Balliet, R.N., Heimlich, J.: Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. Inf. Vis. **15**, 198–213 (2015)