



The Suitability of Data Visualization for Domestic and International Serial Publications

Seolyung Choi^{1,2}, Eunryung Hyun^{3*}

¹Education & Training Division, Korea Institute of Design Promotion, Assistant Researcher, Seongnam, South Korea

²Major in Cognitive Information Design, Department of Learning Science, Ph.D. Candidate, Hanyang University, Seoul, South Korea

³Department of Applied Art Education in College of Education & Department of Cognitive Information Design, Department of Learning Science, Professor, Hanyang University, Seoul, South Korea

Abstract

Background With the development of information and communication technology, the scale of information has increased, and we are in a big data environment. Data visualization designers in this era must have the ability to understand the nature of the data, to select the right chart, to communicate the data accurately, and to review and evaluate the data.

Methods To find out how to evaluate data visualization suitability, we established data visualization suitability evaluation criteria through literature research. For each major error category, we organized the details of each error. Based on the derived data visualization suitability evaluation criteria, we evaluated 382 data visualization outputs of six domestic and international serial publications.

Results Data visualization errors consist of misrepresentation, lack of clarity, counterintuitive, and visual clutter. The evaluation results showed that distortion of information and lack of clarity accounted for the highest percentage, followed by visual confusion and counterintuition. The most frequent error items were different for each medium, and it was inferred that the reason for such sporadic errors is the lack of accurate data visualization principles. The completeness of visualization depends on the competence of the designer for each medium.

Conclusions Accurate and correct data visualization goes beyond the visualization skills of the designer and requires data literacy skills to read the data, to select charts that communicate it effectively, and to ensure that the charts are error-free. There is also a need to develop data visualization principles and checklists.

Keywords Data Visualization, Information Graphics, Date Literacy, Design Guide

*Corresponding author: Eunryung Hyun (pariosa@hanyang.ac.kr)

Citation: Choi, S., & Hyun, E. (2024). The Suitability of Data Visualization for Domestic and International Serial Publications. *Archives of Design Research*, 37(2), 415-431.

<http://dx.doi.org/10.15187/adr.2024.05.37.2.415>

Received : Nov. 18. 2023 ; **Reviewed** : Mar. 15. 2024 ; **Accepted** : Mar. 16. 2024

pISSN 1226-8046 **eISSN** 2288-2987

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

1. 서론

1. 1. 연구 배경과 목적

정보통신 기술의 발달로 정보의 규모가 증가하면서 우리는 빅데이터(Big Data) 환경에 놓이게 되었다. 이 시대에 텍스트를 통한 정보의 전달은 그 한계가 분명하게 있으며, 방대한 내용을 효과적으로 전달할 수 있는 데이터 시각화의 중요성이 높아지고 있다. 데이터를 활용한 사회, 경제, 문화 등의 부가가치 창출도 늘어나면서 데이터 시각화는 경영, 언론, 산업계, 과학계 등 다양한 분야에서 주목받고 있으며(Kim, 2018). 전문적인 데이터 시각화 디자이너의 인기도 높아지고 있다(Meeks, 2019). 데이터 시각화 디자이너는 기초 데이터의 특성을 이해한 다음, 여러 데이터 값을 정직하게 전달할 수 있는 레이아웃을 정하고, 데이터를 인코딩하여 최종 결과물을 도출해야 한다(Wanly 외, 2019). 데이터 시각화의 중요성이나 학술 연구가 주목받고 있음에도 불구하고, 실무자가 실제로 전문적인 환경에서 디자인하는 기준이나 원칙에 대해서는 정립된 바가 많지 않다(Parsons, 2022). 데이터를 시각화하는 데 있어서 시각적인 효과를 내는 방법에 대한 논의는 많지만, 어떤 시각적 효과가 왜 유용한지, 그렇지 않은지에 대한 평가 연구는 많지 않다(Claus, 2019). 따라서 본 연구에서는 데이터 시각화의 적합성 여부를 평가할 수 있는 기준을 도출하고, 이 기준에 따라서 국내의 연속간행물인 신문과 잡지의 데이터 시각화 적합성을 판별한다. 이를 기반으로 우리가 실생활에서 접하는 데이터의 시각화 수준을 알아보고, 데이터의 의미를 정확하게 분석하는 데 기초자료로 활용하고자 한다.

1. 2. 연구 방법과 범위

미국 *The Wall Street Journal*의 데이터 시각화 전문가인 Dona Wong이 출판한 ‘정보 그래픽 가이드’에서 소개하는 기본원칙을 중심으로 국내의 참고문헌을 기반으로 데이터 시각화 결과물 평가 기준을 수립하였다. 이를 기초로 국내의 연속간행물인 일간신문과 주간 잡지의 데이터 시각화 적합성을 알아보기 위하여 총 382점의 데이터 시각화 결과물을 수집하고 평가하였다. 해외 자료로는 미국의 일간신문인 *The Washington Post*, *The Boston Globe*, 주간 잡지인 *Bloomberg Businessweek*에서 사례를 수집하였다. 국내 자료로는 일간신문인 《동아일보》, 《한겨레》, 주간 잡지인 《주간경향》을 수집 대상으로 했다. 수집 대상 기간의 간행물에 있는 모든 데이터 시각화 결과물을 평가 대상으로 하였다. 평가 결과를 중심으로 높은 빈도로 발생하는 오류의 사례, 꼭 지켜져야 하는 원칙이 지켜지지 않아 데이터 전달이 왜곡된 사례 등을 정리하였다.

국어나 수학과 같은 과목과 달리 인포그래픽은 학교에서 가르치는 경우가 드물고 그렇다 보니 실무에서 인포그래픽을 제작하는 직원은 정확한 원칙을 이해하고 그리기 어렵다. 이 연구를 읽은 독자는 좋은 차트와 나쁜 차트를 평가할 수 있는 중요한 잣대들을 이해하고 의사소통 수단으로 효과적으로 사용할 수 있게 될 것이다.

2. 이론적 배경

2. 1. 데이터와 정보

데이터는 연구나 조사, 발견, 수집의 결과인 일종의 기초자료로 정보를 만들기 위한 원자재와 같은 것이며, 정보는 데이터에 의미와 가치를 부여해 조직화하고 변화시킨 것으로 형태와 형식을 지닌다(Oh 외, 2023). 정보의 효율적인 전달을 위해 데이터를 조직화하고 의미를 부여하는 것은 감각적인 향상의 차원을 넘어 논리적 차원까지 작용하며 궁극적으로는 정보를 더욱 가치 있게 한다(Richard 외, 2000). 정보는 사용의 주체나 상황에 따라 의미와 가치가 다르므로 정보가 생산되고 사용되는 맥락, 즉 콘텍스트가 중요하다(Oh 외, 2023). 산발적이고 가치를 알 수 없던 정보가 가공되어 필요한 정보의 효율성에 대한 가중치가 생기는 등의 정보화를 개념적으로 정리하면 <Figure 1>과 같다.

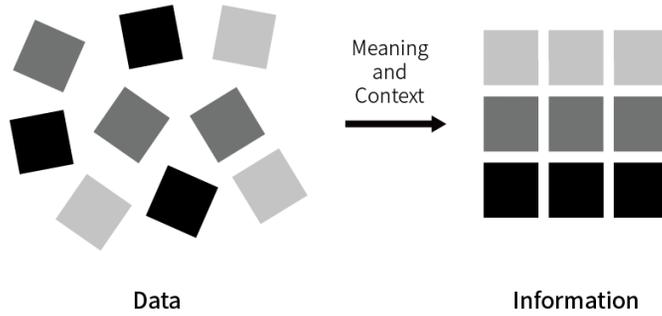


Figure 1 Data and Information

2. 2. 데이터 시각화와 데이터 리터러시

방대한 양의 데이터는 사람의 눈과 두뇌로는 분석, 분류, 판별할 수 없으며, 시각화를 통해 데이터를 읽기 쉽게 만들 수 있고, 시각적 차트를 통해 데이터를 발견할 수 있다(Brath & Jonker, 2015). 여러 정보를 한눈에 알아볼 수 있도록 양적인 데이터를 시각적으로 가공한 것을 데이터 시각화라고 하는데 이를 개념적으로 정리하면 (Figure 2)와 같다. (Figure 2)는 단순하게 나열되어 있던 활자 데이터를 가공하여 한 눈에 보고 이해하기 쉽게 시각적으로 표현한 것이다.

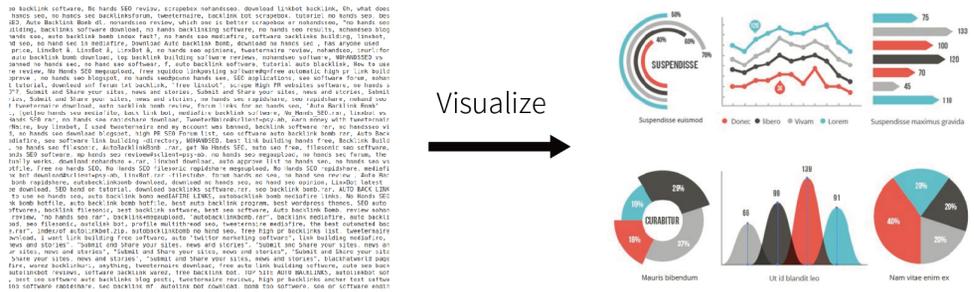


Figure 2 Data Visualization

데이터를 시각화하는 것은 분명 데이터의 의미를 쉽고 직관적으로 전달할 수 있다는 장점을 갖고 있지만 어떻게 시각화하느냐에 따라 정보 전달 효과에 차이가 있을 수 있으며, 특히 잘못된 시각화로 인해 데이터의 의미가 왜곡되어 전달되는 경우 데이터 해석의 오류로 인하여 잘못된 의미 전달이나 의사결정을 내리게 하는 부작용이 발생할 수 있다(Kang 외, 2020).

데이터 리터러시는 데이터를 보고 읽을 줄 알며, 데이터를 활용하는 역량이다. 또한, 시각화를 기반으로 한 데이터 분석을 시각적 분석(Visual Analysis)이라고도 하는데, 시각화가 중요한 역량인 이유는 데이터에 관한 고도의 기술 없이도 쉽게 데이터 분석과 인사이트를 도출할 수 있기 때문이다(Kang 외, 2020). 데이터를 올바르게 판별할 수 있는 기술을 활용하여 데이터와 시각화 자료를 통해 인사이트를 발견하고, 스토리텔링할 수 있다.

2. 3. 데이터 시각화의 유형

데이터 시각화의 유형에는 대표적으로 막대그래프, 선그래프, 파이그래프, 단계 구분도, 지도 플로우, 타임라인, 누적막대그래프, 산점도 등이 있다. 각 유형별 시각화 결과물과 정의는 (Table 1)과 같다.

Table 1 Types of Data Visualization Charts

Index	Chart	Description																				
Bar graph ¹⁾	<p>The Recurring Costs of Modern Life Monthly subscription cost</p> <table border="1"> <thead> <tr> <th>Service</th> <th>Cost (\$)</th> </tr> </thead> <tbody> <tr> <td>Washington Post</td> <td>10</td> </tr> <tr> <td>Spotify</td> <td>11</td> </tr> <tr> <td>Disney+</td> <td>14</td> </tr> <tr> <td>Planet Fitness</td> <td>14</td> </tr> <tr> <td>Netflix</td> <td>15</td> </tr> <tr> <td>Max</td> <td>16</td> </tr> <tr> <td>Mint Mobile</td> <td>17</td> </tr> <tr> <td>Facebook and Instagram</td> <td>20 (proposed)</td> </tr> <tr> <td>New York Times</td> <td>27</td> </tr> </tbody> </table> <p><small>DATA: COMPANY WEBSITES, BLOOMBERG. DISNEY+, NETFLIX AND MAX ARE AD-FREE SUBSCRIPTIONS; ONE-TIME AND ANNUAL FEES ARE AMORTIZED OVER A ONE-YEAR PERIOD. FACEBOOK AND INSTAGRAM COSTS ARE AS PROPOSED TO EUROPEAN REGULATORS, ACCORDING TO THE WALL STREET JOURNAL.</small></p>	Service	Cost (\$)	Washington Post	10	Spotify	11	Disney+	14	Planet Fitness	14	Netflix	15	Max	16	Mint Mobile	17	Facebook and Instagram	20 (proposed)	New York Times	27	<p>A graph plotted with rectangular bars whose height or length is proportional to the value of each quantity.</p>
Service	Cost (\$)																					
Washington Post	10																					
Spotify	11																					
Disney+	14																					
Planet Fitness	14																					
Netflix	15																					
Max	16																					
Mint Mobile	17																					
Facebook and Instagram	20 (proposed)																					
New York Times	27																					
Line graph ²⁾	<p>Disney's quarterly net income</p> <p><small>9/2018 9/2023 ESTIMATE</small></p>	<p>A graph plotted with rectangular bars whose height or length is proportional to the value of each quantity.</p>																				
Pie graph ³⁾	<p>WORKERS BY JOB TYPE, 2021</p> <p>□ White ■ People of color</p> <p>Executives Managers</p> <table border="1"> <thead> <tr> <th>Job Type</th> <th>White (%)</th> <th>People of color (%)</th> </tr> </thead> <tbody> <tr> <td>Executives</td> <td>74%</td> <td>26%</td> </tr> <tr> <td>Managers</td> <td>63%</td> <td>37%</td> </tr> </tbody> </table>	Job Type	White (%)	People of color (%)	Executives	74%	26%	Managers	63%	37%	<p>A graph that plots the entire range of values in a circle, dividing the area of the circle by the proportion of each item.</p>											
Job Type	White (%)	People of color (%)																				
Executives	74%	26%																				
Managers	63%	37%																				
Step diagram ⁴⁾	<p>US Nursing-Home Closures</p> <p>Per 100k residents age 65 and older, 2022-23</p> <p><small>2010 2022</small></p> <p><small>DATA: CENTERS FOR MEDICARE & MEDICAID SERVICES; US CENSUS BUREAU</small></p>	<p>A graph that clusters data by rank, with a step-by-step representation of each rank on the map.</p>																				
Flow map ⁵⁾	<p>Top Ten Corridors on Terrapay India is the biggest destination for international payments</p> <p><small>Source: Terrapay Bloomberg</small></p>	<p>A graph that uses linear symbols to represent movement.</p>																				

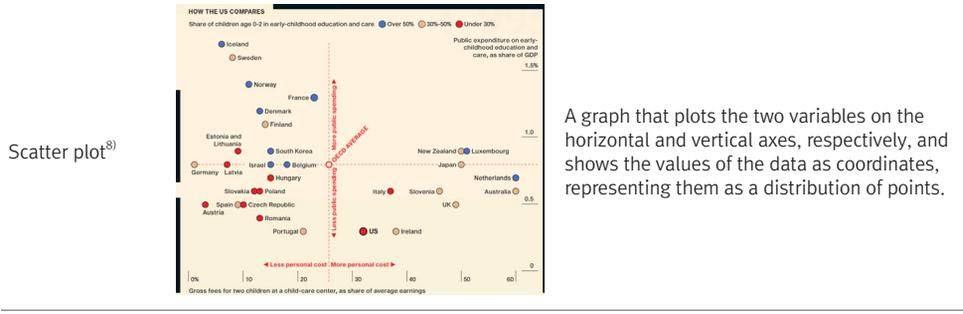
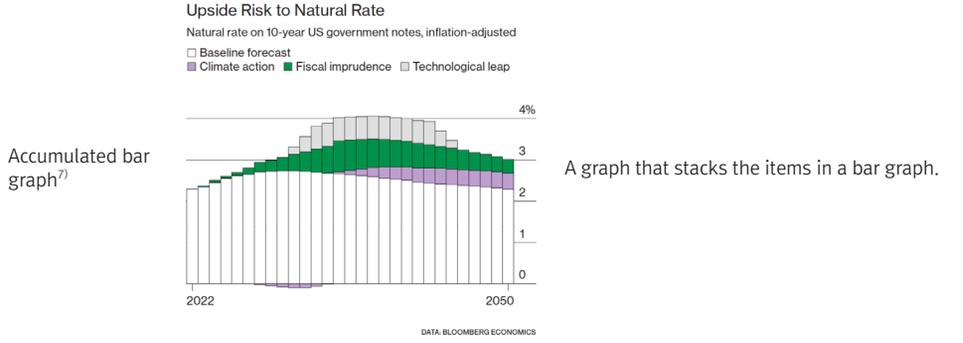
1) Bloomberg Businessweek. (2023.10.16.). The Recurring Costs of Modern Life. <https://www.bloomberg.com/>

2) Bloomberg Businessweek. (2023.10.16.). Disney's quarterly net income. <https://www.bloomberg.com/>

3) Bloomberg Businessweek. (2023.10.9.). Workers by Job Type. <https://www.bloomberg.com/>

4) Bloomberg Businessweek. (2023.10.16.). US Nursing-Home Closures. <https://www.bloomberg.com/>

5) Bloomberg Businessweek. (2023.4.21.). Top Ten Corridors on Terrapay. <https://www.bloomberg.com/>



2. 4. 데이터 시각화 오류 유형

Dana Wong은 *The New York Times*와 *The Wall Street Journal*에서 약 20년 동안 데이터 시각화 디자이너로 일하며 정리한 원칙들을 *The Wall Street Journal Guide to Information Graphics*라는 책으로 출판하였다. 이 책에서는 주요한 3가지의 사항과 더불어 여러 가지 원칙들을 소개한다.

(1) 직관에 위배

첫 번째는 막대그래프의 기준선을 0으로 해야 한다는 원칙이다. 막대그래프는 개별 수량을 나타낼 때 사용하는 그래프로, 하단부를 잘라내면 <Table 2>의 ① 사례와 같이 각 막대의 전체 값이 명료하지 않아 왜곡하게 된다. 데이터 포인트의 값들이 비슷비슷하여 막대 높기로 차이를 보여주기 어려운 경우에는 차잇값이나 증감률로 차트를 구성하는 것이 더 효율적이라고 설명한다. 두 번째로 파이 차트를 이용할 때는 <Table 2>의 ② 사례와 같이 가장 비중이 큰 값을 12시 기준으로 오른쪽에 두고 두 번째 큰 조각을 12시 기준으로 왼쪽에 두는 것이다. 이는 사람들이 직관적으로 12시부터 시계방향으로 내용을 읽기 때문이다.

(2) 명시성 부족

또한, 파이 조각이 5개를 넘으면 비교 대조가 어려워지므로 비중이 작은 조각들은 합쳐 '기타'라고 명명하는 것이 좋다. 세 번째로는 Y축이 있는 차트의 경우 Y축의 척도는 <Table 2>의 ③ 사례와 같이 자연스러운 증가분을 사용해야 한다는 것이다. 즉, 0, 5, 10, 15 등과 같이 독자가 그리드 사잇값을 능숙하게 파악할 수 있는 증가분을 사용함으로써 그리드 선 사이의 데이터 포인트를 쉽게 이해할 수 있도록 배려하는 것이다. 그 외에도 선 차트의 경우 선 아래의 음영처리는 반드시 차트의 기준선이 0인 경우에만 허용해야 와전되지 않는 것, 정수는 오른쪽 정렬로 작성하며, 소수는 소수점을 기준으로 작성할 것, 파이 차트를 그릴 때 반지름을 비율로 사용하면 원의 넓이는 반지름의 제곱에 비례하므로 부정확한 비례가 생길 수 있으니 정확한 비율을

6) The Washington Post. (2023.9.11.). Matches fixed by the Hossam brothers. <https://www.washingtonpost.com/>

7) Bloomberg Businessweek. (2023.10.16.). Upside Risk to Natural Rate. <https://www.bloomberg.com/>

8) Bloomberg Businessweek. (2023.10.9.) How The Us Compares. <https://www.bloomberg.com/>

위하여 면적을 비율로 사용해야 한다는 점, 비례 파이 차트의 경우 도넛형 파이 차트는 파이 안쪽 원이 나머지의 비율을 왜곡하므로 지양해야 한다는 점, 음수를 나타내는 막대는 기준선 중심으로 좌측이나 아래쪽에 위치해야 직관에 혼란을 주지 않는 점, 데이터 포인트가 10개 미만인 경우 특이하게 높은 막대가 있어도 끊지 말고 특잇값은 특잇값처럼 보이도록 해야 한다는 점 등을 설명한다.

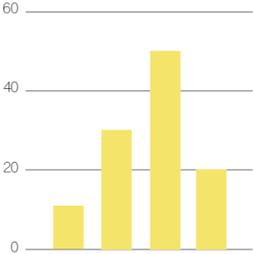
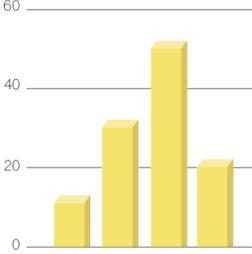
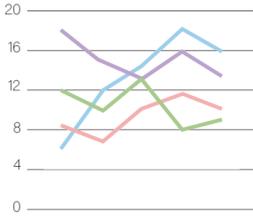
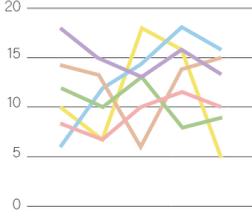
Table 2 The Wall Street Journal Guide to Information Graphics

No.	Right case	Wrong case
①		
②		
③		

(3) 시각적 혼란

이 책에서는 앞서 2.4에서 소개한 원칙 외에도 세부적인 내용을 서술한다. 막대그래프는 3차원 효과나 그림자를 넣을 경우 <Table 3>의 사례 ①과 같이 정확한 데이터 포인트의 값을 인식하는 데 혼란을 주므로 지양해야 한다. 많은 데이터를 비교하는 다중 꺾은선 차트의 경우 <Table 3>의 사례 ②와 같이 꺾은선은 하나의 차트에서 4개 이하로 하는 것이 값을 비교하기에 적당하다. 이 두 가지 내용은 *The Wall Street Journal Guide to Information Graphics*에서도 설명하고 있다.

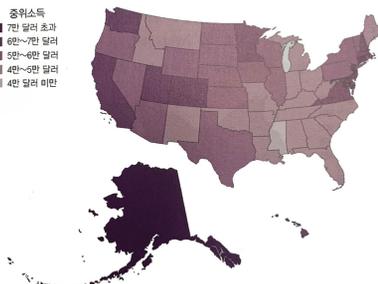
Table 3 Charting & Infographics Data Visualization

No.	Right case	Wrong case
①		
②		

(4) 정보의 왜곡

지도를 기반으로 여러 지역에 걸친 정량적 정보를 표현할 때는 지역별 색을 다르게 칠하는 단계구분도를 사용한다. 이때 유의해야 할 점은 지도상의 면적이 독자에게 직관적으로 인식되기 때문에 정보 전달의 목적이 면적이 아니라 인구 등인 경우는 정보 전달이 왜곡된다. <Table 4>는 미국 모든 주의 중위소득을 나타낸 단계구분도인데 이 지역에서 유독 두드러져 보이는 알래스카는 중위소득이 높지만 실제로 인구밀도는 아주 낮은 지역이다. 이러한 경우는 오른쪽 예시처럼 각 주의 크기를 같은 모듈로 정리하고 실제 인접한 주끼리 위치를 배치한 카토그램 히트맵을 사용하는 것이 보다 정보 전달에 유리하다.

Table 4 Charting & Infographics Data Visualization

Right case ⁹⁾	Wrong case
	

3. 국내외 연속간행물의 데이터 시각화 표현 적합성 평가

3. 1. 평가 대상과 분석 방법

평가 대상으로 국내외의 연속간행물인 일간신문과 주간 잡지에서 총 382점의 데이터 시각화 결과물을 수집하였다. <프레스 리더>라는 국내외의 신문 플랫폼에서 2023년 9월~10월에 발행된 간행물을 전수조사할 수

9) Kang, W. (2020). 데이터가 한눈에 보이는 시각화 [Visualize your data at a glance]. Seoul: Wikibooks.

있는 대상을 위주로 선정하였으며, 시각적 언어로 의사소통을 충분히 하는 매체를 선정하기 위하여 매체별로 2개월의 기간 동안 20건 이상의 데이터 시각화 자료가 있는 매체로 정하였다. *The Washington Post* 66점, *The Boston Globe* 20점, *Bloomberg Businessweek* 72점, 《주간경향》 12점의 데이터 시각화 자료를 수집했다. 《동아일보》 148점과 《한겨레》 64점은 수량이 충분히 확보되어 2023년 9월 발행 간행물만을 대상으로 하였다. 앞서 3장에서 정리한 데이터 시각화 원칙들을 4.2절에서 평가 재구성하여 평가 도구로 활용하였다.

3. 2. 평가 도구 제안

2장에서 정리한 3권의 책의 데이터 시각화 원칙들을 기반으로 평가 도구를 구성하였다. 도구의 항목을 정보의 왜곡, 명시성 부족, 직관에 위배, 시각적 혼란으로 구분하고 각 항목에 해당하는 세부 사항을 정리한 평가 도구를 <Table 5>와 같이 제안한다. 평가도구를 기준으로 각 오류에 대한 조작성 정의를 설정할 수 있다. 정보의 왜곡은 차트를 작성할 때 정보를 있는 그대로 표현하지 않고 강조, 축소, 생략 등 왜곡하는 경우를 의미한다. 명시성 부족은 시각적 표현 과정에서 한눈에 정확하게 인지되도록 하는 인지적 정확성이 부족하여 잘못된 정보가 전달되는 경우를 의미한다. 직관에 위배는 직관적으로 이해하게 되는 과정에서 오류가 생기는 전달을 의미한다. 시각적 혼란은 시각적 전달 과정에서 간결성 등이 부족하여 정보 전달에 혼란을 주는 경우를 의미한다.

Table 5 Data visualization evaluation tool

Index	Details
정보의 왜곡	막대그래프의 기준선은 0으로 작성
	지도 기반 시각화에서 면적이 주요 정보가 아닌 경우 주요 정보를 드러낼 수 있는 다른 형태의 차트로 구성
	선 차트의 경우 선 아래의 음영처리는 반드시 차트의 기준선이 0인 경우에만 허용
	파이 차트를 그릴 때 반지름을 비율로 사용하면 원의 넓이는 반지름의 제곱에 비례하므로 부정확한 비례가 생길 수 있으니 정확한 비율을 위하여 면적을 비율로 사용
	비례 파이 차트의 경우 도넛형 파이 차트는 파이 안쪽 원이 나머지의 비율을 왜곡하므로 지양
명시성 부족	데이터 포인트가 10개 미만인 경우 특이하게 높은 막대가 있어도 끊지 말고 특잇값은 특잇값처럼 보이도록 작성
	막대그래프를 3차원이나 그림자로 하는 경우 데이터 포인트를 정확하게 알 수 없으므로 막대만 표현
직관에 위배	타임라인 등에서 정보를 겹치지 않게 표기함으로써 각 데이터의 값을 명확하게 전달
	음수를 나타내는 막대는 기준선 중심으로 좌측이나 아래쪽에 위치 Y축은 자연스러운 증가분을 사용
시각적 혼란	파이 그래프에서 12시를 중심으로 오른쪽이 제일 큰 값, 왼쪽이 두 번째 값, 나머지는 시계방향으로 배열하고 작은 값은 기타로 작성
	꺾은선 그래프는 하나에 4개 이하로 작성 정수는 오른쪽 정렬로 작성하며, 소수는 소수점을 기준으로 작성

3. 3. 분석 및 평가 결과

(1) 매체별 데이터 시각화 활용 현황

매체별로 데이터 시각화의 차트 종류별 활용 현황을 살펴보면 <Table 6>과 같다.

Table 6 How data visualization is being used

구분		막대	선	파이	단계 구분도	지도 플로우	타임 라인	누적 막대	산점도	기타	합계
동아 일보	건수(건)	58	62	13	1	0	0	3	0	11	148
	비율(%)	39.2	41.9	8.8	0.7	0.0	0.0	2.0	0.0	7.4	100.0
한겨레	건수(건)	24	20	5	1	0	0	10	0	4	64
	비율(%)	37.5	31.3	7.8	1.6	0.0	0.0	15.6	0.0	6.3	100.0
주간 경향	건수(건)	4	3	3	0	0	0	0	0	2	12
	비율(%)	33.3	25.0	25.0	0.0	0.0	0.0	0.0	0.0	16.7	100.0
소계	건수(건)	86	85	21	2	0	0	13	0	17	224
	비율(%)	38.4	37.9	9.4	0.9	0.0	0.0	5.8	0.0	7.6	100.0
The Washington Post	건수(건)	21	16	0	14	4	2	8	1	0	66
	비율(%)	31.8	24.2	0.0	21.2	6.1	3.0	12.1	1.5	0.0	100.0
The Boston Globe	건수(건)	10	5	1	0	1	0	3	0	0	20
	비율(%)	50.0	25.0	5.0	0.0	5.0	0.0	15.0	0.0	0.0	100.0
Bloomberg Businessweek	건수(건)	18	33	3	2	0	1	7	2	6	72
	비율(%)	25.0	45.8	4.2	2.8	0.0	1.4	9.7	2.8	8.3	100.0
소계	건수(건)	49	54	4	16	5	3	18	3	6	158
	비율(%)	31.0	34.2	2.5	10.1	3.2	1.9	11.4	1.9	3.8	100.0
합계	건수(건)	135	139	25	18	5	3	31	3	23	382
	비율(%)	35.3	36.4	6.5	4.7	1.3	0.8	8.1	0.8	6.0	100.0

총 382건 중에서 선그래프가 139건, 막대그래프가 135건, 누적 막대그래프가 31건, 파이 그래프가 25건, 단계구분도가 18건 활용되었다. 이 다섯 가지 종류의 차트가 상위 5가지를 차지하고 있다. 누적 막대그래프를 막대그래프의 한 종류로 포함하는 경우 막대그래프가 가장 많이 활용되는 것으로도 해석할 수 있다.

(2) 데이터 시각화 평가 결과

총 6종의 연속간행물의 데이터 시각화 결과물 382건을 <Table 5>에서 정리한 평가 도구를 활용하여 평가하였다. 평가 결과는 <Table 7>과 같다. 382건 중 13.1%인 50건의 시각화 결과물에 오류가 있었다. 항목별로 살펴보면 정보의 왜곡과 명시성 부족이 각각 동일하게 19건인 38%로 가장 높은 비율의 오류를 차지하였다. 다음으로 시각적 혼란이 8건으로 16%, 직관에 위배가 4건으로 8%를 차지했다. 각 차트의 종류별로 오류 항목을 살펴보면 <Table 8>과 같다.

Table 7 Data visualization evaluation results

구분	정보의 왜곡	명시성 부족	시각적 혼란	오류 건수	데이터 시각화 건수	오류 비율
동아일보	13	11	0	29	148	19.6
한겨레	1	3	0	5	64	7.8
주간경향	0	0	0	1	12	8.3
The Washington Post	4	1	1	6	66	9.1
The Boston Globe	0	2	1	3	20	15.0
Bloomberg Businessweek	1	2	2	6	72	8.3
합계	19	19	4	50	382	13.1
오류 항목 비율	38.0	38.0	8.0	-	-	100.0

Table 8 Errors by chart type

동아일보						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	58	13	4	7	-	2
선	62	8	7	-	-	1
파이	13	4	-	2	-	2
단계구분도	1	1	1(%)	-	-	-
지도플로우	0	0	-	-	-	-
타임라인	0	0	-	-	-	-
누적막대	3	1	-	1	-	-
산점도	0	0	-	-	-	-
기타	11	2	1	1	-	-
합계	148	29	13	11	0	5
오류율/합계		19.6				29
항목별 비율			44.8	37.9	0.0	172
한겨레						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	24	4	1	3	-	-
선	20	0	-	-	-	-
파이	5	1	-	-	-	1
단계구분도	1	0	-	-	-	-
지도플로우	0	0	-	-	-	-
타임라인	0	0	-	-	-	-
누적막대	10	0	-	-	-	-
산점도	0	0	-	-	-	-
기타	4	0	-	-	-	-
합계	64	5	1	3	0	1
오류율/합계		7.8				5
항목별 비율			20.0	60.0	0.0	20.0
주간경향						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	4	0	-	-	-	-
선	3	1	-	-	-	1
파이	3	0	-	-	-	-
단계구분도	0	0	-	-	-	-
지도플로우	0	0	-	-	-	-
타임라인	0	0	-	-	-	-
누적막대	0	0	-	-	-	-
산점도	0	0	-	-	-	-
기타	2	0	-	-	-	-
합계	12	1	0	0	0	1
오류율		8.3				1
항목별 비율			0.0	0.0	0.0	100.0
The Washington Post						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	21	0	-	-	-	-
선	16	2	1	-	1	-
파이	0	0	-	-	-	-
단계구분도	14	3	3	-	-	-
지도플로우	4	0	-	-	-	-

타임라인	2	1	-	1	-	-
누적막대	8	0	-	-	-	-
산점도	1	0	-	-	-	-
기타	0	0	-	-	-	-
합계	66	6	4	1	1	0
오류율		9.1				6
항목별 비율			66.7	16.7	16.7	0.0
The Boston Globe						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	10	3	-	2	1	-
선	5	0	-	-	-	-
파이	1	0	-	-	-	-
단계구분도	0	0	-	-	-	-
지도플로우	1	0	-	-	-	-
타임라인	0	0	-	-	-	-
누적막대	3	0	-	-	-	-
산점도	0	0	-	-	-	-
기타	0	0	-	-	-	-
합계	20	3	0	2	1	0
오류율		15.0				3
항목별 비율			0.0	66.7	33.3	0.0
Bloomberg Businessweek						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	18	0	-	-	-	-
선	33	2	-	1	1	-
파이	3	1	-	-	-	1
단계구분도	2	1	1	-	-	-
지도플로우	0	0	-	-	-	-
타임라인	1	1	-	1	-	-
누적막대	7	0	-	-	-	-
산점도	2	0	-	-	-	-
기타	6	0	-	-	-	-
합계	72	5	1	2	1	1
오류율		6.9				
항목별 비율			20.0	40.0	20.0	20.0

위의 <Table 8>에서 매체별 오류율은 높은 순서대로 《동아일보》 19.6%, *The Boston Globe* 15%, *The Washington Post* 9.1%, 《주간경향》 8.3%, 《한겨레》 7.8%, *Bloomberg Businessweek* 6.9%의 순서로 나타났다. 매체 간에 주요하거나 유사하게 반복되는 오류가 있는 것이 아니라 산발적으로 여러 항목의 오류가 발생하고 있었다. 이는 통일된 데이터 시각화의 가이드와 원칙이 없이 매체별로 시각화 디자이너의 개인 역량에 따라 작성하기 때문으로 유추된다.

Table 9 Suitability of data visualization

총합						
구분	활용 건수	오류 건수	정보의 왜곡	명시성 부족	직관에 위배	시각적 혼란
막대	135	20(14.8%)	5(25.0%)	12(60.0%)	1(5.0%)	2(10.0%)
선	139	13(9.4%)	1(7.7%)	1(7.7%)	2(15.4%)	2(15.4%)
파이	25	6(24.0%)	0(0%)	2(33.3%)	0(0%)	4(66.7%)
단계구분도	18	5(27.8%)	5(100.0%)	0(0%)	0(0%)	0(0%)
지도플로우	5	0(0%)	-	-	-	-
타임라인	3	2(66.7%)	0(0%)	2(100.0%)	0(0%)	0(0%)
누적막대	31	1(3.2%)	0(0%)	1(100.0%)	0(0%)	0(0%)
산점도	3	0(0%)	-	-	-	-
기타	23	2(8.7%)	1(50.0%)	1(50.0%)	0(0%)	0(0%)
합계	382	49	19(38.8%)	19(38.8%)	3(6.1%)	8(16.3%)
오류율/합계		12.8				
항목별 비율			22.4	38.8	6.1	16.3

6개의 연속간행물에 대한 데이터 시각화 사례를 살펴보면 선그래프와 막대그래프의 사용이 많음을 알 수 있다. 또한, 오류 사례를 정리하면 <Table 9>와 같다. 전체 382건의 데이터 시각화 사례 중 12.8%에 해당하는 49건에 오류가 있었다. 오류 중 가장 정보의 왜곡과 명시성 부족이 38.8%의 비율을 차지하고 있었다. 다음은 16.3%로 시각적 혼란, 6.1%로 직관에 위배되는 오류가 있었다. 차트 타입별로 많이 발생하는 오류를 살펴보면 막대그래프의 오류 20건 중 명시성 부족이 12건으로 60%의 오류가 발생하고 있었다. 선그래프에서는 직관에 위배와 시각적 혼란이 각각 2건으로 15.4%를 차지한다. 파이그래프의 오류 6건 중 4건인 66.7%가 시각적 혼란의 오류를 범하고 있었다. 단계구분도는 5건 중 5건 모두에 정보의 왜곡이 있었다. 타임라인에서는 2건 중 2건 모두에 명시성 부족이 있었다. 누적막대의 오류 1건은 명시성 부족이었다. 즉, 전체적으로는 정보의 왜곡과 명시성 부족의 오류가 많은 비중으로 발생하고 있었으며 차트의 종류별로 자주 발생하는 오류의 종류가 달랐다. 막대그래프에서는 정보의 왜곡과 명시성 부족 현상이 주로 발생하고, 선그래프에서는 직관에 위배되거나 시각적 혼란을 주는 오류 사례가 빈번했다. 또한 단계구분도에서는 모든 오류 건에서 정보의 왜곡이 일어나고 있었다.

3. 4. 데이터 시각화 오류 사례

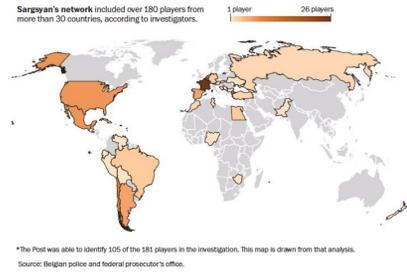
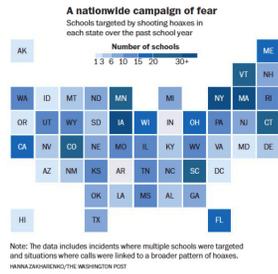
(1) 정보의 왜곡

정보 왜곡의 대표적인 사례는 <Table 10>의 단계구분도이다. 단계에 따라 색상으로 구분하는 경우, 실제로는 낮은 단계를 전달하고자 하는 것임에도 넓은 면적을 칠함으로써 수치가 왜곡되어 전달된다. 단계구분도 사례의 왼쪽 차트는 미국의 각 주를 정사각형으로 표현한 것이다. 각 주의 면적을 동일하게 표현하고, 색상으로 단계를 전달함으로써 정보 왜곡의 오류를 방지하였다.

선그래프에서는 오른쪽 차트와 같이 특잇값을 나타낼 때 줄임표시로 데이터 값을 줄여 표현하는 경우가 잘못된 경우다. 이는 지면에 할애할 공간이 부족하기 때문에 나타나는 현상으로 추정할 수 있는데, 이렇게 특잇값을 줄여서 표현하는 차트의 경우 정확한 수치를 전달하지 못하므로 왜곡된 정보를 전달하게 된다. 줄임표시를 사용한 막대그래프의 경우 대부분 가로로 긴 형식의 레이아웃을 활용하였다. 같은 면적을 할애하더라도 세로로 긴 형식의 레이아웃을 사용하면 해당 오류 없이 차트를 작성할 수 있다.

Table 10 Example of skewed information

Right case	Wrong case
Step diagram ¹⁰⁾¹¹⁾	



Line graph ¹²⁾¹³⁾	
------------------------------	--



(2) 명시성 부족

〈Table 11〉에서 타임라인의 바른 사례와 오류 사례를 확인할 수 있다. 오른쪽 사례는 두 개의 범례가 서로 명확하게 구분되지 않으며, 각 지점별로 일어난 사건에 대하여 말풍선의 위치가 제각각으로 시기별로 일어난 사건이 명확하게 전달되지 않는다. 막대그래프 예시에서 오른쪽 차트는 증가율을 나타내는 반원이 막대그래프를 하단부를 가리고 있어서 막대그래프의 수량을 한눈에 파악하는 데 방해가 될뿐더러 반원의 크기가 모두 같아 증가율에 대한 명시적 전달되지 않는다. 이 경우 별도의 막대그래프를 추가하여 증가율을 비교할 수 있도록 하는 등 적절한 형태의 그래프를 선택하여 표현하는 것이 옳은 작성이다. 〈Table 10〉의 왼쪽 차트는 친환경차 판매 대수를 나타내는 막대그래프 위에 점유율의 추이를 표현하기 위하여 선그래프를 추가하였다. 이처럼 여러 종류의 정보를 전달하고자 하는 경우 적합한 그래프를 선택하여 별도로 작성하는 것이 중요하다.

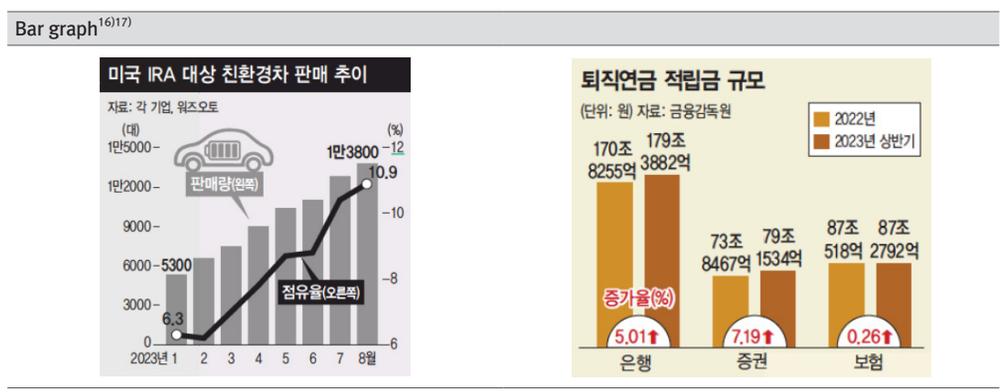
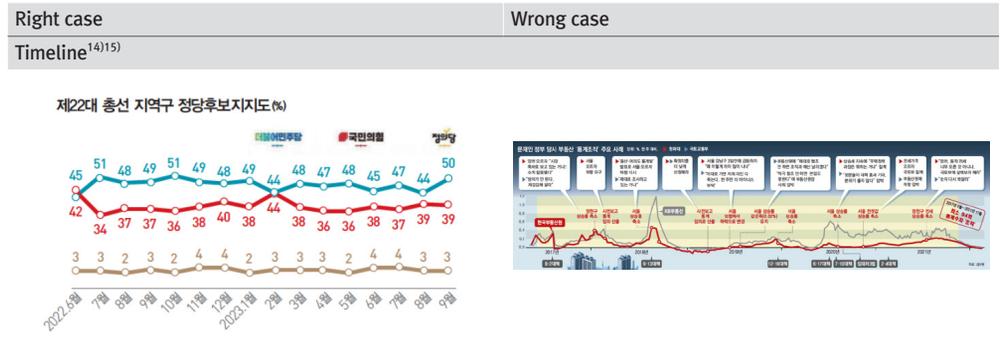
10) The Washington Post. (2023.10.5.). A nationwide campaign of fear. <https://www.washingtonpost.com/>

11) The Washington Post. (2023.9.11.). Sargsyan's network. <https://www.washingtonpost.com/>

12) Bloomberg Businessweek. (2023.10.2.) Smallest available semiconductor, in nanometers. <https://www.bloomberg.com/>

13) Dongailbo. (2023.9.1.). Net buying of Japanese stocks. www.donga.com

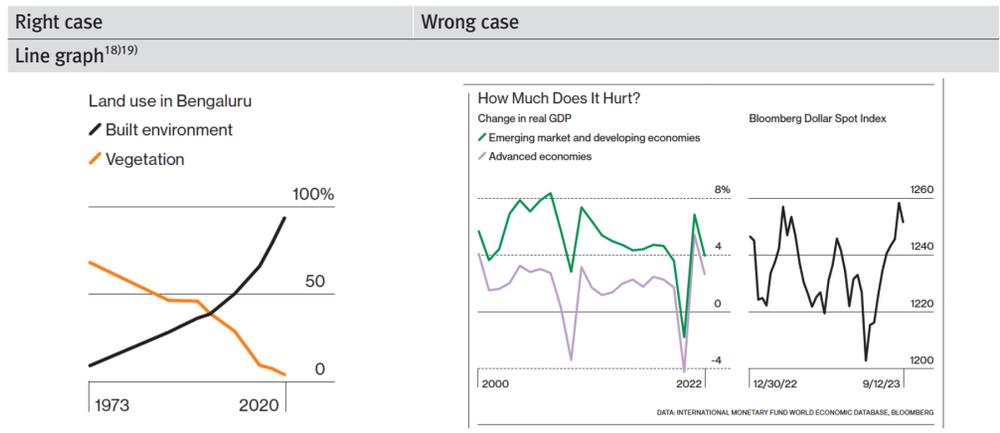
Table 11 Example of lack of explicitity



(3) 직관에 위배

<Table 12>의 오른쪽 차트에서 Y축 증가분은 6의 단위로 직관적으로 양을 추정하기에 어려운 단위를 활용하였다. Y축의 증가분은 0, 2, 4, 6이나 0, 5, 10, 15 등 사잇값을 직관적으로 추정할 수 있는 값을 기준으로 해야 한다.

Table 12 Examples that are counterintuitive



(4) 시각적 혼란

<Table 13>에서 파이그래프는 가운데 공간을 비워 글을 전달하도록 함으로써 면적 전달에 오류를 범하고 시각적 혼란을 준다. 또한, 파이그래프에서 자주 발생하는 오류는 면적이 아니라 지름을 기준으로 표현하는 경우이다. 지름을 기준으로 하는 경우 면적 비율로는 제곱이 되므로 시각적 혼란을 준다. 이런 오류를 범하지 않고 표현한 그래프가 파이그래프 사례의 왼쪽 그래프이다. 선그래프의 오른쪽 차트는 11개 항목의 추이를

14) WeeklyKyunghyang. (2023.10.16.). 22nd General Election Candidate Endorsements. <https://weekly.khan.co.kr/>

15) Dongailbo. (2023.9.16.). Major cases of real estate winterization during the Moon Jae-in administration. www.donga.com

16) Hankyeore. (2023.9.20.). Green car sales to US IRBs. <https://www.hani.co.kr/>

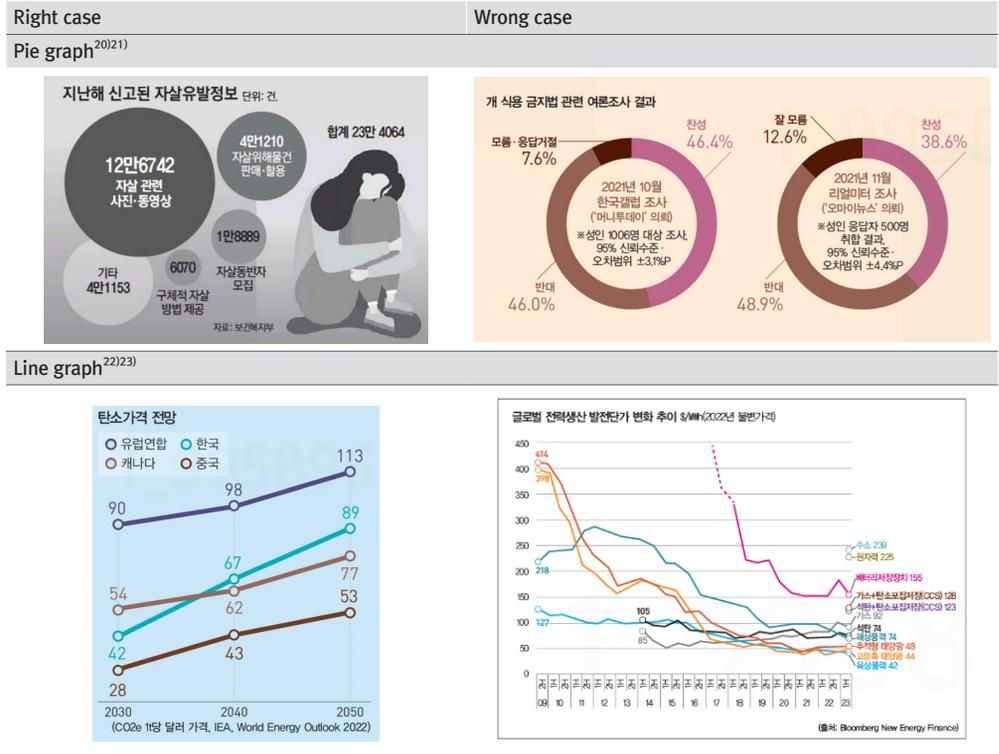
17) Hankyeore. (2023.9.18.). Retirement plan contribution size. <https://www.hani.co.kr/>

18) Bloomberg Businessweek. (2023.10.2.) Land Use in Bengaluru. <https://www.bloomberg.com/>

19) Bloomberg Businessweek. (2023.9.25.) How Much Does It Hurt?. <https://www.bloomberg.com/>

작성하였고, 많은 항목의 정보를 넣다보니 범례 간의 간격이 거의 없어 시각적으로 혼란을 준다. 각 선그래프의 변화가 잘 읽히지 않을뿐더러 범례와 선을 구분하는 것도 쉽지 않다. 선그래프로 추이에 대한 정보를 전달하고자 할 때는 4개 이하의 항목으로 정리하는 것이 좋다. 또한, 색상은 같은 색상에서 톤만 다른 색을 활용하되, 강조할 선에만 다른 색을 사용해주는 것도 효과적인 방법이다.

Table 13 Examples that are counterintuitive



4. 결론 및 제언

본 연구는 데이터를 가공한 정보를 시각적으로 의미미하고 원활하며 정확하게 전달하는 것에 대한 중요성을 인식하며 출발하였다. 문헌조사를 통하여 데이터 시각화의 평가 원칙을 도출하고, 이를 기초로 국내외 연속간행물의 데이터 시각화 현황과 적합성을 평가하였다. 데이터 시각화의 원칙에서는 정보의 왜곡, 명시성 부족, 직관에 위배, 시각적 혼란을 오류 발생의 요소임을 정리하였다. 이를 기초로 데이터 시각화 적합성을 평가한 결과 국내외 6개의 연속간행물에서 여러 가지 유형의 오류가 산발적으로 발견되었으며, 이는 통일된 가이드나 원칙이 없이 데이터 시각화 디자이너 개인의 역량에 따라 작성하기 때문으로 유추하였다.

정확하고 올바른 데이터 시각화를 위해서는 디자이너의 시각화 역량을 넘어 데이터 리터러시 능력이 겸비되어 데이터를 읽고, 이를 효율적으로 전달할 수 있는 차트를 선정하며, 차트 작성에 오류가 없도록 하는 책임이 필요하다.

본 연구를 통하여 매체 간 분석의 결과가 산발적으로 여러 항목에서 오류가 발생하는 것으로 드러나면서 다른 기준을 통한 분석이 보다 유용한 결과를 도출할 수 있다는 것을 알 수 있다. 따라서 후속 연구로 현재 분석한 382개의 결과물에 차트 유형별 분석 결과를 알 수 있는 결과물을 추가하여 데이터의 내용 또는 형식으로 결과물을 분류하고, 유형별로 어떤 오류가 발생했는지 분석함으로써 시각화 가이드라인을 수립하는 것이 필요하다. 이를 통해 연속간행물에서 자주 활용되는 차트의 종류별로 데이터 시각화 원칙을 수립하고 이를 확인할 수 있는 체크리스트도 개발할 수 있을 것이다.

20) Dongailbo. (2023.9.9.) Suicide attempts reported in the past year. www.donga.com

21) WeeklyKyunghyang. (2023.9.25.). Poll results on dog food bans. <https://weekly.khan.co.kr/>

22) WeeklyKyunghyang. (2023.9.18.). Carbon Price Outlook. <https://weekly.khan.co.kr/>

23) WeeklyKyunghyang. (2023.10.30.) Global electricity generation unit cost trends. <https://weekly.khan.co.kr/>

References

1. Clous, W. (2019). *Fundamentals of Data Visualization*. O'Reilly.
2. Dona, W., (2010). *Information Graphics*. W.W. Norton & Company.
3. Kang, W., Lim, J., Choi, H., & Newsjelly. (2020). 데이터가 한눈에 보이는 시각화[Visualize data at a glance]. Paju: Wikibooks.
4. Kim, H. (2018). *고등학교 미술과 정보디자인 교육으로서 데이터 시각화 수업 실행에 관한 질적 연구. [A Qualitative Study on Implementing Data Visualization Class in High School Art and Information Design Education]* (Doctoral dissertation). Seoul National University.
5. Lee, H. (2020). *차트 제작 & 인포그래픽 데이터 시각화 [Charting & Infographics Data Visualization]*. Seoul: Sungandang.
6. Meeks, E. (2019). "Annual Data Visualization Survey Results" in Nightingale. *The Journal of the Data Visualization Society*.
7. Oh, B., & Kang, S. (2023). *정보 디자인 교과서 [Information Graphic Textbook]*. Seoul: Ahngraphics.
8. Parsons, P. (2022). Understanding Data Visualization Design Practice. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 665–675.
9. Richard, W., Loring, L., & David, S. (2000). *Information Anxiety 2*. Portland: c. Book News.
10. Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2019). *Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff*. IEEE Transactions on Visualization and Computer Graphics.

국내의 연속간행물의 데이터 시각화 적합성 평가 연구

최서령^{1,2}, 현은령^{3*}

¹한국디자인진흥원 인재육성실 주임연구원, 성남, 대한민국

²한양대학교 대학원 러닝사이언스학과 인지정보디자인전공 박사과정, 대한민국

³한양대학교 사범대학 응용미술교육과 & 대학원 러닝사이언스학과 인지정보디자인전공 교수, 서울, 대한민국

초록

연구배경 정보통신 기술의 발달로 정보의 규모가 증가하면서 우리는 빅데이터(Big Data) 환경에 놓이게 되었다. 이 시대의 데이터 시각화 디자이너는 데이터의 특성을 이해하고 올바른 차트를 선정하며, 이를 정확하게 전달하고, 검토·평가할 수 있는 능력을 갖추어야 한다.

연구방법 데이터 시각화 적합성 평가 방법을 알아보기 위하여 문헌연구를 통해 데이터 시각화 적합성 평가 기준을 수립하였다. 주요하게 발생하는 오류 항목별로 각 오류가 발생하는 세부 내용을 정리하였다. 도출된 데이터 시각화 적합성 평가 기준에 따라 국내외 연속간행물 총 6건의 데이터 시각화 결과물 382점을 평가하였다.

연구결과 데이터 시각화 오류의 항목은 정보의 왜곡, 명시성 부족, 직관에 위배, 시각적 혼란으로 구성된다. 평가 결과는 정보의 왜곡과 명시성 부족이 가장 높은 비율을 차지했으며, 시각적 혼란과 직관에 위배가 그 뒤를 따랐다. 또한, 매체별로 가장 빈번하게 발생하는 오류 항목은 서로 달랐으며 이렇게 산발적인 오류들이 발생하는 이유로는 정확한 데이터 시각화 원칙의 부재로 인하여 시각화 완성도가 매체별 디자이너의 역량에 좌우된다는 것을 그 원인으로 유추하였다.

결론 정확하고 올바른 데이터 시각화를 위해서는 디자이너의 시각화 역량을 넘어 데이터 리터러시 능력이 겸비되어 데이터를 읽고, 이를 효율적으로 전달할 수 있는 차트를 선정하며, 차트 작성에 오류가 없도록 하는 책임이 필요하다. 또한, 데이터 시각화 원칙과 체크리스트 개발의 필요성이 있다.

주제어 시각화, 정보 디자인, 데이터 리터러시, 디자인 가이드

*교신저자: 현은령 (pariosa@hanyang.ac.kr)