

데이터 시각화의 사용자 선호도 분석: 분포 시각화 방식을 중심으로

김 하 늘¹ · 김 성 희^{2*}¹동의대학교 IT융합학과 석사과정^{2*}동의대학교 산업ICT기술공학과 조교수

Analysis of User Preference of Data Visualization: Focusing on Distribution Visualization Methods

Ha-Neul Kim¹ · Sung-Hee Kim^{2*}¹Graduate Student, Department of IT Convergence, Dong-eui University, Busan, 47340, Korea^{2*}Assistant Professor, Department of Industrial ICT Technology, Dong-eui University, Busan, 47340, Korea

[요 약]

데이터 시각화는 사람들에게 시각적 요소를 통해 데이터 인사이트를 빠르게 인지하고 기억할 수 있도록 도와주어 데이터 시각화의 활용 범위가 다양해지고 있다. 데이터 시각화의 관심이 높아지면서 뉴스와 기사뿐만 아니라, 의료 분야에서도 데이터 시각화를 활용하여 정보를 제공하고 있다. 올바른 데이터 인사이트를 제공하기 위해서 사용자를 위한 적합한 시각화가 필요하다. 본 논문에서는, 데이터 시각화에 대한 사용자의 인식을 조사하기 위해 사용자의 데이터 시각화 선호도에 대한 실험을 진행하였다. 본 실험은 분포 시각화 방식의 시각화를 한정하여 5개의 시각화와 4개의 Task를 이용한 데이터 시각화의 상대적, 객관적인 선호도를 측정하여 진행하였다. 실험을 통해 얻어진 사용자의 시각화 선호도 결과를 각 시각화별, Task별로 나눠서 분석하였다. 분석한 결과를 고찰하고 이에 따른 데이터 시각화의 적합한 사용 방향을 제시하고 논의하였다.

[Abstract]

Data visualization helps people quickly recognize and remember data insights through visual elements, therefore, the scope of data visualization is diversifying. As interest in data visualization increases, it is exposed to the general public through not only news articles, but also the medical field as personal data are represented through visualizations. To provide the proper data insights, we need to provide effective visualizations for your users. In this research, an experiment was conducted on the user's data visualization preference to investigate the user's perception of data visualization. This experiment was conducted by measuring the relative and objective preference of data visualization using 5 visualizations and 4 tasks by limiting the visualization of the distribution visualization method. The results of user's visualization preference obtained through the experiment were analyzed separately for each visualization and each task. The results of the analysis were reviewed, and accordingly, the appropriate direction of data visualization was suggested and discussed.

색인어 : 데이터 시각화, 데이터 시각화 선호도, 사용성 평가, 분포 데이터, 시각화 태스크

Keyword : Data Visualization, Data Visualization Preference, Usability Experiment, Distribution Data, Visualization Task

<http://dx.doi.org/10.9728/dcs.2023.24.1.129>



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 non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 28 October 2022; **Revised** 15 November 2022

Accepted 12 December 2022

***Corresponding Author; Sung-Hee Kim**

Tel: +82-51-890-2366

E-mail: sh.kim@deu.ac.kr

1. 서론

빅데이터를 통해 트렌드를 파악하고 인사이트(Insight)를 발견하면서 새로운 부가가치를 창출하는 시대로 접어들에 따라, 데이터의 활용도를 높일 수 있는 데이터 시각화의 역할이 중요해졌다. 데이터 시각화는 데이터의 패턴, 추세 및 상관관계를 쉽게 이해할 수 있도록 도와준다. 정량적인 데이터를 시각적 요소로 표현하기 때문에, 인간의 두뇌에서 데이터의 특징을 쉽게 파악할 수 있는 것이다[1]. Aaker은 데이터 시각화를 통해 사람들이 인사이트를 빠르게 인지하고 기억할 뿐만 아니라, 메시지 전달에도 효과적으로 사용될 수 있다고 설명하고 있다[2].

데이터 시각화는 수학적, 통계적 지식 없이도 시각적 패턴을 통해 누구나 쉽게 이해할 수 있어, 데이터 시각화의 활용 범위가 다양하다. 예를 들어, 사람들에게 정보를 제공하는 매체인 뉴스와 기사 등에서 효과적으로 정보를 제공하기 위해 데이터 시각화를 활용하는 것을 쉽게 볼 수 있다. 또한, 의료 분야에서도 의사와 환자 사이의 의사소통 수단으로 데이터 시각화를 활용하면서, 원하는 조건에 맞춰 의료 정보 결과를 제공하고 있다[3]. 데이터 시각화의 활용 범위가 다양해지고 사용 대상이 많아지면서, 사용자의 시각을 파악하여 적합한 시각화를 제공하는 것은 매우 중요하다. 데이터 시각화는 다양한 방식을 통해 데이터를 표현할 수 있어, 사용자들에게 제공하고자 하는 의도와 다르게 전달되어 정확한 데이터를 제공하지 못하게 되는 경우도 발생할 수 있다[4]. 사용자의 시각에 맞춰 데이터 시각화를 제공하기 위해서, 사용자가 데이터 시각화에 대해 느끼는 인식을 파악한다면, 보다 효과적으로 데이터 인사이트를 전달할 수 있을 것이다.

따라서, 본 연구에서는 동일한 데이터에 대하여 다양한 방식의 시각화로 표현하여, 사용자의 데이터 시각화에 대한 선호도의 인식을 조사하는 실험을 진행하였다. 분포 시각화 방식을 한정하여 진행하였으며 시각화 종류, Task에 따라 어떤 시각화가 효율적으로 표현되었다고 느끼는지 알아보았다. 본 실험에 사용되는 시각화 요소는 데이터 시각화 활용 능력을 측정하기 위해 개발된 VLAT(Visualization Literacy Assessment Test)를 참고하여 구성하였다. 본 실험은 2가지의 Survey로 나눠서 진행되었다. 먼저, 첫 번째 Survey 1은 시각화 형태가 다른 2개의 데이터 시각화를 비교하여 상대적인 선호도를 측정하는 실험으로, 선호 시각화를 선택하고 선호하는 이유를 정성적 방법을 이용하여 데이터를 수집하였다. Survey 2는 각 데이터 시각화에 대한 객관적인 선호도를 측정하기 위해, 7점 리커트 척도(Likert scale)를 이용하여 수치화된 선호도를 측정하였다. 이에 수집된 데이터를 분석하여 분석 결과에 대해 고찰하고, 데이터 시각화의 적합한 사용 방향에 대해 논의하였다.

II. 관련 연구

2-1 데이터 시각화 가이드라인 현황

데이터 시각화는 데이터를 기반으로 객관적으로 표현하는 정보형 메시지를 전달하거나, 데이터를 기초로 해석된 의미를 포함하는 설득형 메시지를 전달하기 위해 사용된다. 데이터 시각화는 디자이너의 수작업이 첨가되는 인포그래픽스와 프로그래밍에 의해 자동으로 생성되는 데이터 시각화로 나눌 수 있다. 데이터의 양이 방대해지고 빠른 주기로 변화함에 따라, 디자이너의 중심으로 만들어지는 인포그래픽보다는 데이터 시각화의 활용도가 높아지고 있다[5]. 하지만, 데이터 특성을 반영하여 적합한 시각화를 개발하기 위한 가이드라인에 대한 연구는 많이 이루어지고 있지 않고 있다. Bahador Saket(2019)은 Task, 즉 시각화를 통해 수행되는 작업에 따라 시각화 효율성이 크게 달라진다고 주장하였다[6]. 또한, Aditeya Pandey(2020)는 시각화에서 Task가 중요한 역할임에도 불구하고, Task 기준, Task에 따른 시각화 모범 사례에 대한 기존 논문들의 가이드라인이 부족하다고 주장하였다[7]. 효과적인 시각화를 제공하기 위해 Task의 기준이 필요한 상황이다.

2-2 데이터 시각화 사용자에 대한 연구

데이터 시각화는 사용자와 상호작용을 통해 데이터의 인사이트가 전달되면서, 데이터 시각화를 활용하기 위해서는 사용자의 경험을 고려하여 접근해야 한다. 사용자 경험(UX)을 중심으로 데이터 시각화 유형에 따라 빅데이터 시각화를 연구하기 위해, 추진기(2019)는 시각화 유형을 시간, 분포, 관계, 비교, 공간 시각화로 구분하여 진행하였다[8]. 사용자 경험의 측정 요소는 피터모빌의 사용자 경험 측정의 육각 구조(User Experience Honeycomb)를 통해 유용성, 사용성, 호감성, 인지성, 접근성, 신뢰성+가치성의 여섯 가지 차원을 사용하여 설문지를 구성하였다. 해당 연구는 5점 리커트 척도에 맞춰 데이터 시각화 유형을 설문지를 통해 평가하는 방법으로 진행되었다. 해당 연구 결과는 시각화 유형에 따라 설문 결과를 분석하여 제공하고 있다. 여기서, 데이터 시각화의 디자인적 요소 이외에 나아가 시각화 기능에 초점을 맞춰서 실험을 진행해야 하는 필요성이 있다.

데이터 시각화의 유용성에 대한 비교 실험 평가는 어렵다고 알려져 있다[9]. 이는 시각화가 사용자의 전문 지식, 데이터의 속성, 사용되는 환경 등 다양한 요소에 영향을 받기 때문이다. 실제 데이터 시각화가 사용되는 환경을 재현하기 위해 외적 타당도를 중시하면 정량적인 실험 결과를 도출하기 힘들며, 내적 타당도를 위해서 맥락 및 시각화에 대해 제한적인 실험을 해야 한다. 따라서, 많은 실험들은 Bar Chart(바 차트)와 Line Chart(라인 차트) 비교와 같이 잘 알려져 있는 제한된 시각화에 대해서 이루어지고 있으며, 정확도나 걸린 시간과 같은 지표를 측정하고 있다.

3-2 실험 설계

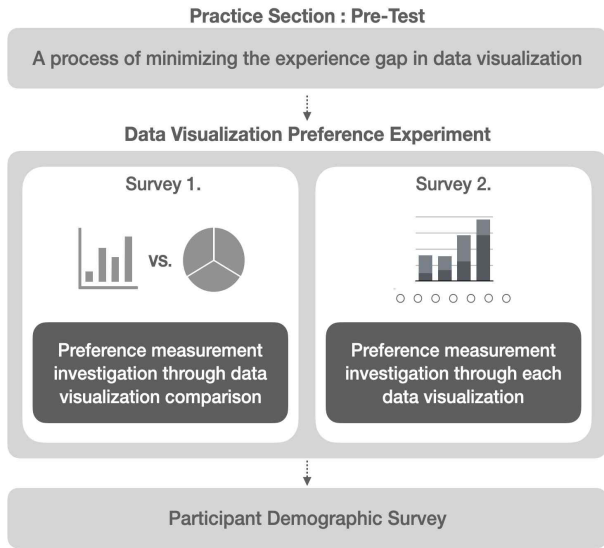


그림 2. 실험 구성을 위한 프레임워크
 Fig. 2. Framework for constructing Experiment

본 실험은 분포 시각화 방식의 데이터 시각화에 대해 사용자의 선호도를 파악하기 위해 그림 2와 같이 설계하였다. 선호도 실험에 앞서, 먼저 해당 데이터 시각화에 대한 경험의 차이를 줄이고자 시각화의 문해력을 요구하는 연습 문제를 통해 데이터 시각화를 충분히 접할 수 있도록 설계하였다. 연습 문제는 각 시각화에 대해 4개의 Task를 이용한 질문으로 이루어져 있으며, 총 20문제로 구성되었다. 데이터 시각화 선호도를 측정하는 실험은 상대적 선호도, 객관적 선호도를 수집하기 위해 2개의 Survey로 나눠 설계하였다. Survey 1은 서로 다른 2개의 시각화의 비교를 통한 상대적인 선호도를 측정하고, Survey 2는 각 시각화의 객관적인 선호도의 수치를 측정하기 위한 실험이다.

Survey 1은 각 세트에서 2개의 데이터 시각화를 비교하여 상대적인 선호도를 측정하고, 3개의 세트로 구성하였다. 세트 1은 수량을 나타내는 데이터 시각화를 비교하는 세트로, 스택형 차트와 클러스터형 차트로 구성하였다[13]. 세트 1은 같은 데이터에서 막대 그래프를 스택으로 쌓아서 표현하거나 동일한 기준점에서 펼쳐서 보여주는 표현 방법을 비교하기 위해 선정하였다. 세트 2는 비율형 데이터를 이용하여 실제 수량을 비율로 계산하여 나타내는 데이터 시각화를 비교하는 세트로, 스택형 차트와 클러스터형 차트로 구성하였다[13]. 세트 2에서는 같은 데이터에 대해 100%를 기준으로 면적을 나타내거나 개별 그래프로 분할하여 면적을 나타내는 표현 방법을 비교하기 위해 선정하였다. 세트 3은 수치형 데이터를 이용하여 수량을 나타내는 데이터 시각화를 비교하는 세트로, 스택형 차트와 계층적 구조의 차트를 구성하였다. 세트 3은 같은 데이터에서 수량을 스택형으로 쌓아서 나타내거나 계층

적 구조를 가지고 나타내는 표현 방법을 비교하기 위해 선정하였다. 각 세트에서는 4개의 Task를 이용한 질문으로, 총 12번의 선호하는 시각화를 선택하고 선호하는 이유를 작성하도록 설계하였다.

Survey 2는 각 5개 데이터 시각화의 선호도를 7점 리커트 척도를 이용해 객관적인 선호도의 수치를 측정하기 위해 설계하였다. Survey 2는 각 데이터 시각화에 대해 4개의 Task를 이용한 질문을 통해 총 20개의 질문으로 구성되어 설계되었다.

표 1. 분포 시각화 방식의 데이터 시각화 비교 세트
 Table 1. Data Visualization Comparison Set of Distribution Visualization Methods

Set	Type	Visualization	Expression Method
1	Quantity	Stacked Bar Chart	Express a bar graph by stacking them
		Multi-set Bar Chart	Express a bar graph spread out at the same reference point
2	Ratio	100% Stacked Bar Chart	Express the area based on 100%
		Multi-set Pie Chart	Express of dividing an area into individual graphs
3	Quantity	Stacked Bar Chart	Express a quantity by stacking
		Treemap	Express quantity with hierarchical structure

3-3 실험 방법

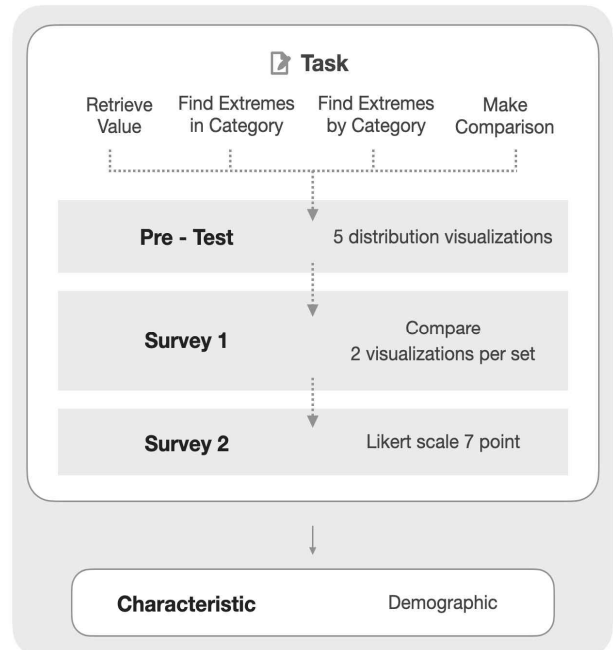


그림 3. 실험 흐름도
 Fig. 3. Experiment Flow Chart

본 실험은 웹을 통해 그림 3과 같이 이루어져, 약 35-40분 정도 소요되었다. 실험 웹사이트는 Django, D3.v4.js를 이용해 개발되었고, 실험의 목적과 안내 사항을 충분히 숙지하도록 공지하였다. 먼저, Pre-Test는 총 20개의 질문으로, 그림 4와 같이 100% Stacked Bar Chart의 시각화와 ‘프랑스의 1인당 연간 곡물류 식품 소비율을 고르시오’와 같은 질문을 통해 진행되었다.

Survey 1은 표 1과 같이 각 세트에서 4개의 Task 질문을 통해 총 12개의 질문으로 시각화를 비교하여 응답하였다. 그림 5와 같이 ‘미국에서 사망원인이 암질환의 사망자 수 찾기’와 같은 Task 질문에 대해 2개의 시각화 중 선호하는 시각화를 선택하고 그 이유를 작성하도록 하였다.

Survey 2는 그림 6과 같은 화면을 통해, Survey 1과 같은 Task의 질문으로 시각화 선호도를 응답할 수 있도록 7점 리커트 척도를 선택하도록 하였다. 마지막으로, 인구 통계학적 특성의 설문 항목을 통해 성별, 연령대, 전공, 최종학력을 입력하도록 진행하였다.

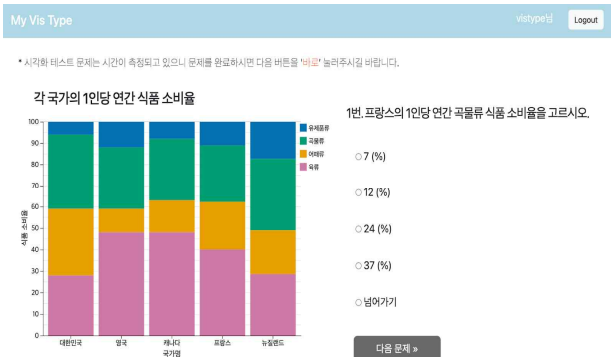


그림 4. 테스트 페이지 화면의 예시
Fig. 4. Example of a Test page screen



그림 5. 선호도 페이지1 화면의 예시
Fig. 5. Example of the preference page 1 screen



그림 6. 선호도 페이지2 화면의 예시
Fig. 6. Example of the preference page 2 screen

IV. 실험 결과

4-1 실험 대상

본 실험은 30명의 참가자를 대상으로 그림 3과 같은 과정을 통해 오프라인으로 같은 종류의 각 PC를 사용하여 실험을 진행하였다. 본 실험에 앞서, 참여자에게 연구의 목적에 대해 충분히 설명하고 테스트를 진행하였다. 실험 참가자 인구 통계학적 특성으로는, 성별은 남자 21명(70%), 여자 9명(30%), 평균 연령은 24.9세(min : 21, max : 30), 전공 분야는 모두 공과대학으로, 최종학력은 대학교 재학 18명(60%), 대학교 졸업 2명(6.7%), 대학원 재학 9명(30%), 대학원 졸업 1명(3.3%)으로 나타났다.

4-2 실험 결과

1) Survey 1의 실험 결과

본 논문에서 분포 시각화의 데이터 시각화 선호도 실험에 앞서 연습 문제를 통한 사용자의 문해력 결과는 대체로 높은 정당률을 보였고, 평균 93.75%(min : 80, max : 100)로 나타났다. 다음은, 그림 4과 같이 2가지의 데이터 시각화를 비교하여 선호도를 측정하는 실험 결과로 3가지 요인(데이터 시각화, Task, 선호하는 이유)으로 나뉘서 분석하였다.

각 세트의 데이터 시각화별 선호 빈도수를 분석한 결과는 그림 7에서 확인할 수 있다. 데이터 시각화 세트 1은 Multi-set Bar Chart의 선택 빈도가 86, Stacked Bar Chart의 선택 빈도가 34로 Multi-set Bar Chart의 선호도가 상대적으로 더 높게 나타났으며, 비교 세트 2에서는 100% Stacked Bar Chart의 선택 빈도수 59, Multi-set Pie Chart의 선택 빈도수 61로 다른 세트들에 비해 선호도 차이가 상대적으로 작다. 비교 세트 3은 Stacked Bar Chart의 선택 빈도수 47, Treemap의 선택 빈도수 73으로 Treemap의 선호도가 상대적으로 더 높았다.

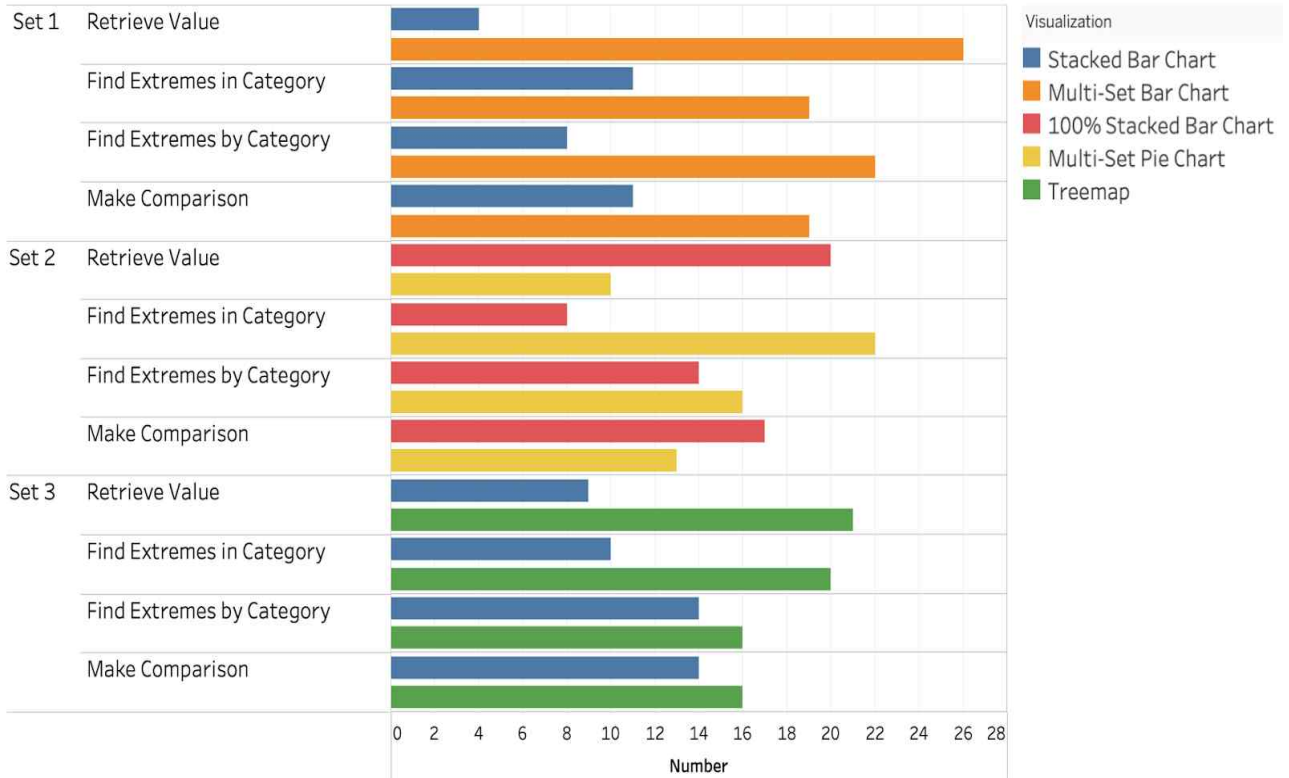


그림 8. 각 세트에서 Task에 따라 데이터 시각화 선호도의 결과
 Fig. 8. Results of data visualization preference by Task in each set

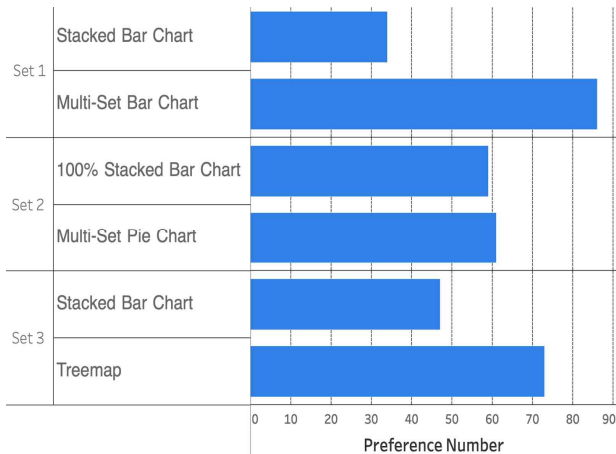


그림 7. 각 세트에서 사용자의 데이터 시각화 선호도 결과
 Fig. 7. Results of user's data visualization preference in each set

각 세트 내에서, Task별 선호 분석 결과는 선호 빈도 수를 분석하여 그림 8과 같은 결과가 도출되었다. 세트 1은 모든 Task의 선호도에서 Multi-set Bar Chart가 더 높은 것으로 보인다. 반면, 비교 세트 2에서는 Retrieve Value와 Make Comparison에서 100% Stacked Bar Chart의 선호도가 높게 나타나고, Find Extremes in Category와 Find Extremes by Category에서 Multi-set Pie Chart의 선호

도가 더 높게 나타난다. 비교 세트 3은 모든 Task에서 Stacked Bar Chart보다 Treemap에서 선호도가 더 높게 나타났다, 예를 들어 Retrieve Value와 Find Extremes in Category는 선호도 빈도 수가 10 이상의 차이를 보였다.

마지막으로 선호하는 이유로 수집된 주관식 답변을 통해 사용자의 인식을 이해하는 측면으로 사용될 수 있다. 분석 결과는 2명의 저자가 합의하여 군집화 과정을 거쳐 표 2와 같이 나타났다. 데이터 시각화 비교 세트 1에서 Multi-set Bar Chart를 선호하는 이유로는, High Visibility(29명), Easy to Compare(22명), X-Axis of starting point 0(16명)등의 순위로 의견을 제시하였다. Stacked Bar Chart의 경우에는, High Visibility(15명)으로 선호하였다. 예를 들어, 선호되는 이유 중 'Multi-set Bar Chart가 그래프의 차이를 한눈에 볼 수 있어 더 보기가 편하다'라는 의견을 통해 High Visibility의 응답 결과로 분류하여 군집화하였다. 비교 세트 2에서는 Multi-set Pie Chart의 Easy to understand Ratio(23명), 100% Stacked Bar Chart의 Convenient to understand Numbers(22명)으로 선호하는 이유로 언급되었다. 비교 세트 3에서는 압도적으로 Treemap의 Convenient to understand Numbers(49명)의 군집화로 분류된 답변이 가장 많은 비율로 차지하고 있다. 대체로 'Treemap은 사망 수의 정확한 수치가 텍스트로 나타나서 더 효과적이다'라는 의견을 통해 Convenient to understand Numbers의 군집화에 해당하는 응답 결과가 도출되었다.

표 2. 데이터 시각화 선호도 비교 세트의 주관식 응답 결과
Table 2. Response results for data visualization preference comparison set

Set	Visualization	Type	Count
1	Stacked Bar Chart	High Visibility	15
		Easy to Compare	6
		Similar Positions on the Y-Axis	4
		Easy to understand Ratio	3
		High Visibility into Proportions	3
		Convenient to understand Numbers	2
		Intuitive Comparison Convenience	1
	Multi-set Bar Chart	High Visibility	29
		Easy to Compare	22
		X-Axis of starting point 0	16
		Convenient to understand Numbers	12
		Intuitive Comparison Convenience	6
		Similar Positions on the Y-Axis	1
2	100% Stacked Bar Chart	Convenient to understand Numbers	22
		High Visibility	10
		Easy Location Comparison	7
		Easy to Compare	5
		Easy to understand Ratio	5
		Similar Positions on the Y-Axis	5
		Color Palette Exists	2
		X-Axis of starting point 0	2
	Multi-set Pie Chart	Easy out-of-group Comparison	1
		Easy to understand Ratio	23
		High Visibility	10
		Easy to Compare	9
		Intuitive Comparison Convenience	8

		Easy to Compare within Groups	5
		Easy Location Comparison	4
		Convenient to understand Numbers	1
		Easy out-of-group Comparison	1
Set	Visualization	Type	Count
3	Stacked Bar Chart	Easy to understand Ratio	7
		Easy color Comparison	7
		Easy to Compare	6
		Similar Positions on the Y-Axis	4
		Easy Location Comparison	4
		Easy out-of-group Comparison	3
		High Visibility	3
	Treemap	Intuitive Comparison Convenience	3
		Convenient to understand Numbers	1
		Convenient to understand Numbers	49
		Easy to understand Ratio	11
		Easy to Compare	4
		High Visibility	4
		Intuitive Comparison Convenience	3

2) Survey 2의 실험 결과

그림 6과 같이 각 데이터 시각화에 대해 7점 리커트 척도 (Likert scale)를 이용하여 선호도를 측정하는 실험으로 IBM SPSS Statistics을 사용하여 분석하였다. 먼저, 7점 리커트 척도(Likert scale)를 이용해 데이터 시각화별 선호도 평균 차이가 있는지 분석하였다. 데이터는 정규성 검정을 충족하지 않아, Kruskal-Wallis 검증을 적용하였다. 시각화별 유의한 차이가 존재했으며(p<0.000), Bonferroni의 보정을 통해 사 후 비교를 하였다. 그림 9와 같이 box plot으로 결과를 확인할 수 있으며, Multi-set Bar Chart의 선호도 수치가 평균 6.0(min : 3, max : 7)로 가장 높게 나타나고, Stacked Bar Chart가 비교적 낮은 선호도의 수치의 평균 4.3(min : 1, max : 7)로 나타났다.

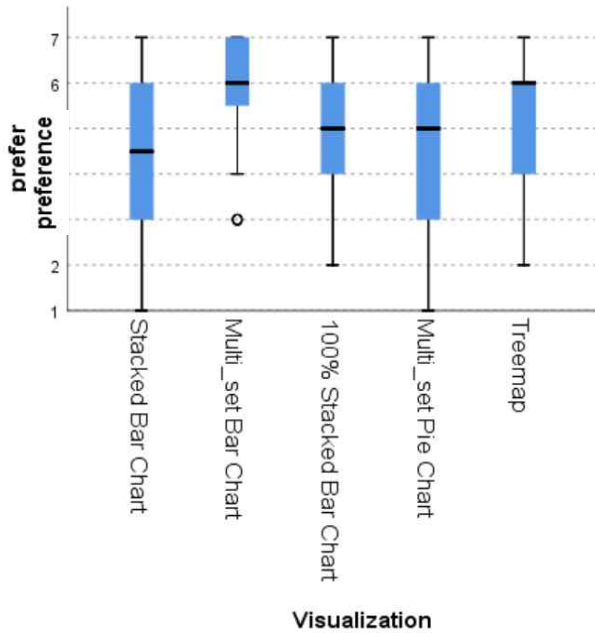


그림 9. 데이터 시각화별 선호도
 Fig. 9. Preference by data visualization

또한, 각 데이터 시각화에서 Task별 선호도 평균 차이가 있는지 분석한 결과, Stacked Bar Chart의 Task별 선호도 평균 차이의 결과는 시각화별 유의한 차이가 존재하지 않았으며 그림 10과 같이 나타났다.

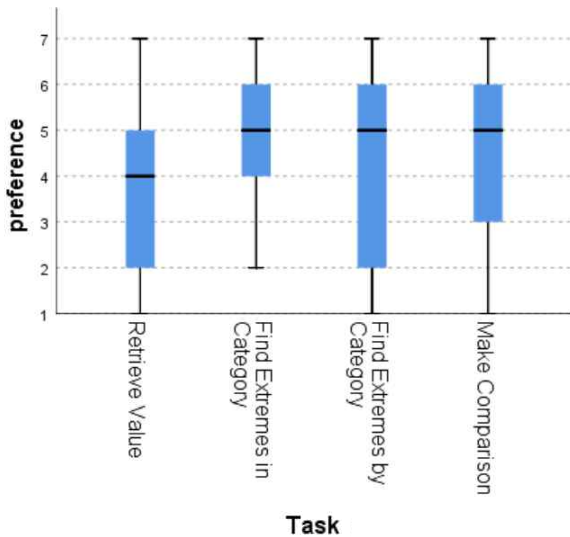


그림 10. Stacked Bar Chart의 Task별 선호도
 Fig. 10. Preference by Task in Stacked Bar Chart

Multi-Set Bar Chart의 Task별 선호도 평균 차이의 결과는 그림 11과 같이 나타났다. 시각화별 유의한 차이가 존재하고($p < 0.050$), 대체로 선호도가 높게 나타났으며, 특히 Find Extremes in Category에서 가장 높았다.

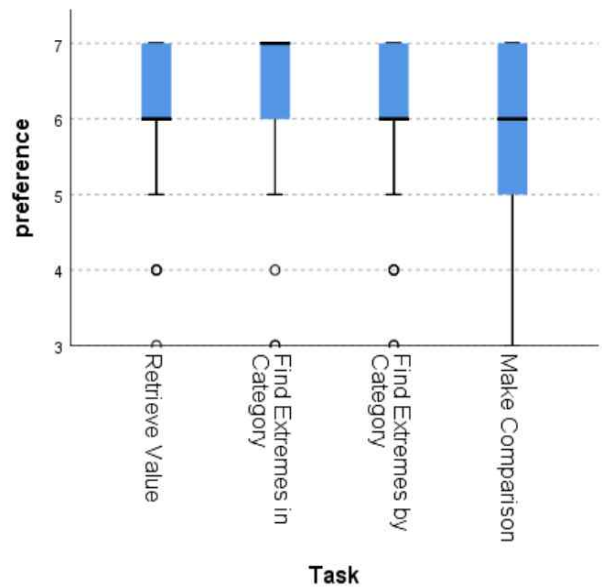


그림 11. Multi-Set Bar Chart의 Task별 선호도
 Fig. 11. Preference by Task in Multi-Set Bar Chart

100% Stacked Bar Chart의 Task별 선호도 평균 차이의 결과는 그림 12와 같이 나타났다. 시각화별 유의한 차이가 존재하고($p < 0.000$), Find Extremes by Category와 Make Comparison 높은 선호도를 보였고, Find Extremes in Category에서는 대체로 낮은 선호도가 나타났다.

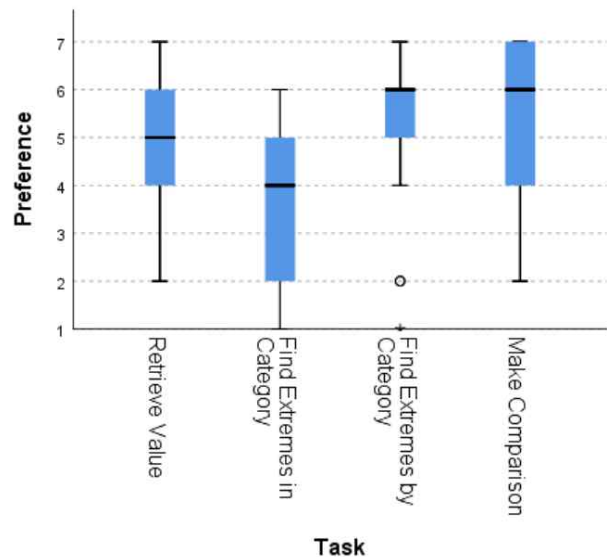


그림 12. 100% Stacked Bar Chart의 Task별 선호도
 Fig. 12. Preference by Task in 100% Stacked Bar Chart

Multi-Set Pie Chart의 Task별 선호도 평균 차이의 결과는 그림 13과 같이 나타났다. 시각화별 유의한 차이가 존재하고($p < 0.000$), Find Extremes in Category에서 가장 높은 선호도가 나타났고, Retrieve Value는 대체로 높은 선호도를 보였지만 선호도 점수가 고르게 분포되어 있다.

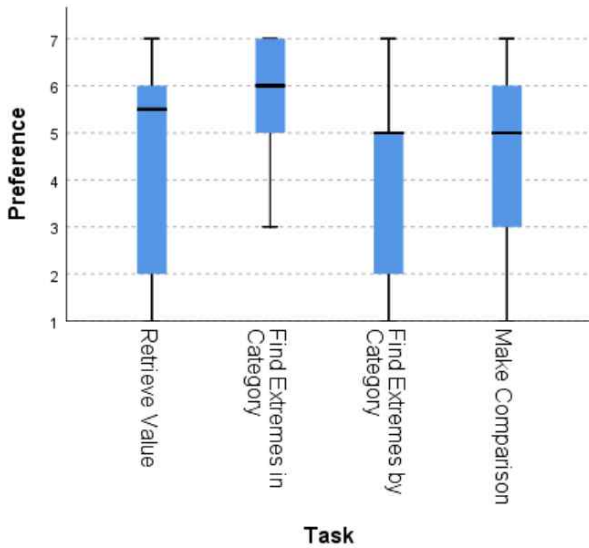


그림 13. Multi-Set Pie Chart의 Task별 선호도
 Fig. 13. Preference by Task in Multi-Set Pie Chart

Treemap의 Task별 선호도 평균 차이의 결과는 그림 14와 같이 나타났다. 시각화별 유의한 차이가 존재하고 ($p < 0.050$), Retrieve Value와 Find Extremes in Category에서 높은 선호도가 나타났다.

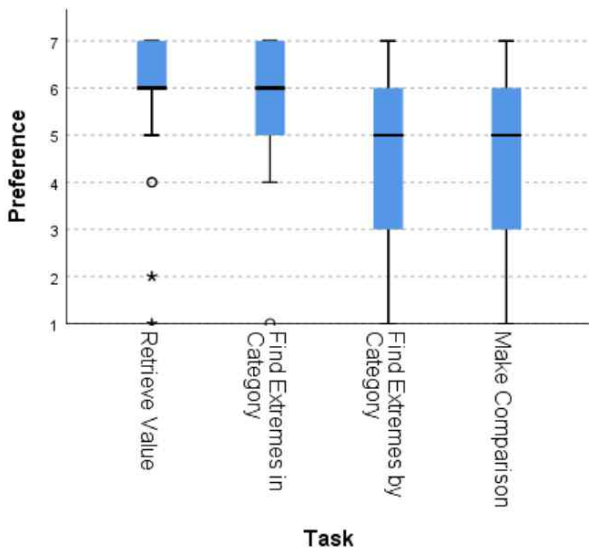


그림 14. Treemap의 Task별 선호도
 Fig. 14. Preference by Task in Treemap

V. 결론 및 고찰

데이터 시각화의 활용 범위가 다양해지고 사용 대상이 많아지면서 사용자의 시각을 파악하기 위해, 본 연구에서는 분포 시각화 방식의 데이터 시각화에 대한 사용자 선호도의 실

험을 진행하고 각 시각화별, Task별로 분석을 진행하였다.

Survey를 종합한 결과로, 분포 시각화 방식의 데이터 시각화 중 Multi-Set Bar Chart가 가장 높은 선호도를 얻었다. 사용자들이 다른 시각화들보다 Multi-Set Bar Chart가 가시성이 높고, 값을 비교하기에 편리하다고 생각하기 때문에 높은 선호도를 얻을 수 있었다. 또한, Set 1의 결과로 Multi-Set Bar Chart가 Stacked Bar Chart보다도 주어진 4개의 모든 Task를 인식하기에 더 쉽다고 느끼고, 특히 Find Extremes in Category에서 값을 인식하기에 유리하다는 결과가 도출되었다. 이러한 결과는 Multi-Set Bar Chart가 절대적, 상대적 선호도가 높다는 것을 보여주고 있다. Set 1의 결과를 토대로 Set 2의 결과에서 Multi-Set Pie Chart가 상대적으로 선호도가 높을 것으로 예상하였다. 하지만, Set 2의 결과는 100% Stacked Bar Chart와 Multi-Set Pie Chart의 각 시각화 선호도 차이가 거의 없었지만, 그림 8과 같이 Task별로 분석하였을 때는 선호도의 차이가 존재한다는 것을 볼 수 있다. 100% Stacked Bar Chart는 Retrieve Value의 Task를 수행하기에 편리하여 상대적으로 높게 선호하였고, 이러한 결과는 객관적으로 선호도를 측정했을 경우에도 비슷한 경향을 보여주고 있다. 또한, Multi-Set Pie Chart에서는 Find Extremes in Category의 Task를 수행하는 것에서 상대적으로 높은 선호도의 경향을 보였고, 그림 13과 같이 제일 높은 선호도의 Task로 Find Extremes in Category를 볼 수 있다. 따라서 Task에 따라 사용자의 데이터 시각화 선호도가 달라지는 것을 확인할 수 있다. 마지막으로, 세트 3의 Treemap은 사용자들이 수치를 파악하기 편리하기 때문에 Stacked Bar Chart보다 더 명확히 선호된다는 것을 표 2를 통해 확인할 수 있다. 이러한 결과는 그림 14에서도 수치 파악과 가장 관련이 높은 Retrieve Value의 Task에서 가장 높은 선호도를 보여주고 있다.

이와 같이 여러 단계의 Task와 3개 이상의 시각화 비교에 대한 연구는 많이 이루어지지 않았다. 빅데이터를 분석하는 경우, 탐색적 데이터 분석이 많이 이루어지고 있으며, 이는 데이터에 대한 분포 확인, 데이터간 비교, 특정 값 추출 등 다양한 Task를 수행하게 된다. 각 단계에 맞춰 적합한 시각화를 보여주는 것이 중요하며, 이에 대한 학문적 연구가 더 필요할 것으로 보인다. 또한, 실제 대시보드를 개발하는 경우에는 다양한 시각화를 제시하는 것이 중요하며, 각 시각화의 목적에 대해서 정의를 내리고 분석 프로세스에 맞춰 제공하는 것이 중요할 것으로 보인다.

데이터 시각화는 현재 사회적으로 주목받고 있는 분야로, 사용자의 접근성이 높아지고 있어 사용자 시각에 맞춰 적합한 시각화를 제공하는 것이 중요하다. 따라서, 본 연구는 사용자의 시각에서 데이터 시각화에 대한 사용 방향에 대해 분석하였고, 이는 데이터 시각화의 적합한 사용 방향을 설정하기 위한 시작점이라 할 수 있다. 차후 범용적인 기준으로 데이터 시각화의 사용 방향을 설정하기 위해서는 사용자의 대상 범위를 넓혀서 실험을 진행하고, 데이터 시각화의 선호도 분석

에 인구 분포학적 요소를 추가하는 연구가 필요하다. 또한, 분포 시각화 방식뿐만 아니라 다양한 방식의 데이터 시각화에 대해 자세한 연구가 필요하다.

감사의 글

본 연구는 “본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성 (GrandICT연구센터)사업의 (IITP-2022-2020-0-01791) 연구 결과로 수행되었으며 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 3단계 산학협력 선도대학 육성사업(LINC 3.0)의 연구결과입니다.

참고문헌

- [1] M. Islam and S. Jin, "An Overview of Data Visualization," *International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1-7, 2019. <https://doi.org/10.1109/ICISCT47635.2019.9012031>
- [2] D. Aaker and J. L. Aaker, "What are Your Signature Stories?," *California Management Review*, Vol. 58, No. 3, pp. 49-65, 2016. <https://doi.org/10.1525/emr.2016.58.3.49>
- [3] R. R. Austin, M. A. Mathiason, R. A. Lindquist, S. K. McMahon, D. S. Pieczkiewicz and K. A. Monsen, "Understanding Women's Cardiovascular Health Using MyStrengths+MyHealth: A Patient-Generated Data Visualization Study of Strengths, Challenges, and Needs Differences," *Journal of Nursing Scholarship*, Vol. 53, No. 5, pp. 634-642, 2021. <https://doi.org/10.1111/jnu.12674>
- [4] J. Y. Ji, H. M. Kang and Y. S. Hong, "Bias in Data Visualization: within-the-bar bias in Bar Graphs," in *Proceeding of HCI KOREA*, South Korea, pp. 743-747, 2020.
- [5] J. Lee, "A Study on Visualizing Method and Expression of Information Design for Big Data," *Journal of Basic Design*, Vol. 14, No. 3, pp. 259-269, 2013.
- [6] B. Saket, A. Endert and Ç. Demiralp, "Task-Based Effectiveness of Basic Visualizations," in *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, No. 7, pp. 2505-2512, July 2019. <https://doi.org/10.1109/TVCG.2018.2829750>
- [7] A. Pandey, U. H. Syeda and M. A. Borkin, "Towards Identification and Mitigation of Task-Based Challenges in Comparative Visualization Studies," *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 55-64, 2020. <https://doi.org/10.1109/BELIV51497.2020.00014>
- [8] J. K. Choo, "A Study on Big Data Visualization Strategy Based on Social Communication :Focusing on User Experience (UX) based on Big Data Visualization Types," *The Journal of the Korea Contents Association*, Vol. 20, No. 1, pp. 142-151, 2020. <https://doi.org/10.5392/JKCA.2020.20.01.142>
- [9] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE transactions on visualization and computer graphics*, Vol. 18, No. 9, pp. 1520-1536, 2011. <https://doi.org/10.1109/TVCG.2011.279>
- [10] B. Saket, A. Endert, and J. Stasko, "Beyond usability and performance: A review of user experience-focused evaluations in visualization," in *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 133-142, October 2016. <https://doi.org/10.1145/2993901.2993903>
- [11] S. K. Lee, S. H. Kim, and B. C. Kwon, "VLAT: Development of a Visualization Literacy Assessment Test," *IEEE Transactions Visualization and Computer Graphics*, Vol. 23, pp. 551-560, Jan 2017. <https://doi.org/10.1109/TVCG.2016.2598920>
- [12] M. Okabe and K. Ito, "Color Universal Design (CUD) How to make figures and presentations that are friendly to Colorblind people," 2002. Available: <https://jfly.uni-koeln.de/color/>
- [13] Stacked Chart or Clustered? Which One is the Best?, January 2017. Available: <https://radacad.com/stacked-chart-or-clustered-which-one-is-the-best>



김하늘 (Ha-Neul Kim)

2021년 : 동의대학교 (공학학사-산업
ICT기술)

2006년~현 재: 동의대학교 IT융합학과 석사과정
※ 관심분야 : 인공지능(AI), 데이터시각화(Data Visualization),
HCI



김성희 (Sung-Hee Kim)

2008년 : 이화여자대학교 컴퓨터공학과
대학원 (공학석사)
2014년 : 퍼듀대학교 산업공학과 대학원
(공학박사)

2015년~2017년: 삼성전자
2017년~현 재: 동의대학교 산업ICT기술공학 조교수
※ 관심분야 : 데이터 시각화, 인간-컴퓨터 상호작용, 인공지능