

빅데이터를 위한 데이터 시각화 방법과 표현 연구 (광주 대중버스노선 이용 실태를 적용한 태블루를 활용한 시각화 표현) (A Study on Visualizing Method and Expression for Big Data)

문희정*

(Hee Jeoung Moon)

요약

하루가 다르게 방대한 데이터가 생겨나고, 데이터를 반영한 정책지원이나 콘텐츠들이 증가하면서 데이터에 대한 중요성이 주목받고 있다. 데이터의 시각화 측면에서 모든 데이터를 살펴보는 것에 제약이 따르기 때문에, 시각화의 기술적인 요소뿐만 아니라 데이터를 요약하고, 이해하기 쉽도록 돕는 시각화의 방법론적 요소가 중요하다고 할 수 있다.

본 논문에서는 현재 많이 사용되는 디자인 중심의 인포그래픽을 개선하고 데이터 중심의 인포그래픽을 제안하기 위해 데이터 시각화의 표현 방법을 정리하고, 시각화의 방법을 Tableau Public을 활용하여 데이터 분석 및 인포그래픽 제작사례를 제시하고자 한다. 인포그래픽 제작을 위해 광주광역시 버스 이용자 데이터가 사용되었고, 분석결과 정류소 사용 총 승객수는 일반 승객수와는 유사한 결과를 보이지만 환승역 이용 승객, 청소년 승차·환승 승객수에 따른 정류소와는 차이를 보였다. 그래프 유형에 따라 시각적인 결과는 다르게 도출되었다. 데이터 중심의 인포그래픽 시각화는 기존의 인포그래픽의 시각적인 역할만 강조한 것에 비해 데이터를 효율적으로 전달하는 것을 물론 과학적 연구를 위한 도구로 활용될 것으로 기대한다.

■ **중심어** : 데이터시각화 ; 빅데이터 ; 인포그래픽 ; 비주얼리제이션 ; 태블루

Abstract

The importance of data is increasing at a high rate as data is massively generated and taken into account in various policy supports and contents. However, because of their speed of growth, it is difficult to find the data that is needed. Both the methodological elements that summarize the data and the technical elements of the visualization that help to see at a glance are important. This paper summarizes data visualization methods to improve the currently used design - oriented infographics and propose data - centric infographics. In addition, we will present examples of data analysis and infographics production using Tableau Public. The Gwangju metropolitan city bus user data was used for infographics production, and the results show that the total number of passengers using the stopping point is similar to that of the general passengers, while it is different from the numbers of transit passengers and teen riding-and-transit passengers. Data-centric infographics visualization, unlike existing infographics that is pronounced only as a visual role, is expected to be used as a tool for scientific research as well as efficiently delivering data

■ **keywords** : Data-Visualization ; Big-data ; Infographics ; Tableau public

I. 서론

최근 빅데이터 기술시장의 성장과 함께 시각화 기술에 관한 관심은 지난 몇 년간 꾸준히 증가해왔다. 일반적으로 시각화라는 것은 정보를 효율적으로 전달하기 위한 표현기법을 의미한다[1][2]. 위키백과에서는 데이터 시각화(data visualization)를 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하

고 전달되는 과정을 말한다. 데이터 시각화의 목적은 도표(graph)라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것이라고 정의하고 있다. 빅데이터의 시각화 측면에서는 모든 데이터를 살펴보는 것에 제약이 따르기 때문에, 시각화의 기술적인 요소와 더불어 데이터를 요약하고, 한눈에 살펴볼 수 있도록 돕는 시각화 방법론적 요소가 중요하다고 할 수 있다. 시각화 관련 연구들에서는 이를 데이터 자체에 대한 시각화 외에 정보 시각화, 지식 시각화 등으로 구분하고 있다[3][4].

* 정회원, 호남대학교 소프트웨어학과

이 논문은 2016년도 호남대학교의 연구지원을 받아 수행된 연구임.

접수일자 : 2018년 12월 10일

수정일자 : 1차 2018년 12월 22일, 2차 2019년 02월 27일

게재확정일 : 2019년 03월 05일

교신저자 : 문희정 e-mail : moon5@honam.ac.kr

데이터 시각화는 철저한 기획을 통해 정보를 표현하는 기법으로 단순히 시각화 도구로 해결될 수 있는 문제가 아니다. 실제로 인포그래픽과 데이터 시각화로 유명한 비주얼 닷 리(visual.ly)는 별도의 도구(Tool)를 사용하지 않고, 자체적으로 기획하고, 디자인도 직접 한다고 한다[5]. 시각화는 색상, 이미지, 숫자, 문자 등 다양한 요소로 표현할 수 있다.

빅데이터 시대에서 기술과 관련하여 수학자, 통계학자들이 주요한 역할을 할 것이고, 데이터 시각화와 관련해서는 스토리텔러, 디자이너들의 역할이 중요하게 될 것이다.

Cairo[6]는 데이터 시각화 복잡성이 독자들의 수준에 따라 달라져야 한다고 했으며, Laine-Hernandez and Särkkä[9]는 Market 데이터 시각화를 두 가지 관점 (과학적인 관점 (scientific orientation), 보도 중점의 관점(journalistic orientation))으로 분류하였다. Cairo는 데이터의 내용에 중점을 두고 있지만, Laine-Hernandez and Särkkä는 데이터의 처리 과정 및 독자에게 초점을 맞추고 있다고 하였다. 우리나라에서 주로 가공되는 인포그래픽은 보도 중점의 관점으로 주로 다루어지지만, 과학적인 관점으로는 부족했다고 볼 수 있다.

본 논문에서는 현재 많이 사용되는 디자인 중심의 인포그래픽을 개선하고 데이터 중심의 인포그래픽을 제안하기 위해 데이터 시각화의 표현 방법을 정리하고, 수많은 데이터 속에서 보이지 않았던 패턴이 발견될 수 있는 시각화의 방법을 Tableau Public을 활용하여 데이터 기반의 인포그래픽 제작사례를 제시하고자 한다.

II. 본 론

1. 데이터와 데이터 시각화

데이터란 Data Object를 모아둔 것을 말한다.

일반적으로 알고 있는 데이터는 주로 표 형태라고 볼 수 있다. 표 형태의 자료들은 처리되지 않은 파일(Raw data)이 아니라 원자료를 특정 조건에 따라 계산하여 정리한 통계 데이터를 보기 쉽게 정리한 것을 말한다.

데이터 시각화는 역사적으로 지리적 시각화(geographic visualization)에서 출발하여 주제도(thematic cartography), 통계 그래픽(statistical graphics)을 거쳐 발전되었다[8]. 데이터 시각화는 정보시각화(information graphics), 인포그래픽(Infographics), 정보 일러스트레이션(information illustration)등으로 사용되고 있으나 명확히 정의된 용어는 없다. 정보 과학, 컴퓨터과학, 인간-컴퓨터 상호작용, 통계, 데이터마이닝, 지도학, 저널리즘 등의 광범위한 학문 분야에서 각기 조금씩 다른 의미로 사용되고 있다[9]. Cairo[6]는 독자의 역할이 수동적인지 또는 능동적인지의 정도에 따라 데이터 시각화를 구분하였는데, 인포그래픽의 경우에는 통계차트, 지도, 다이어그램, 스

토리텔링 등을 통해 정보를 보여주는 방식이지만, 정보 시각화는 사용자가 데이터에 대한 탐색과 분석을 할 수 있도록 시각적 도구를 제공하는 것이라고 정의하였다[8]. 그러나 인포그래픽도 복잡화 정도와 사용자 탐색 및 분석 정도에 따라 정적 인포그래픽, 모션 인포그래픽 그리고 인터랙티브 인포그래픽으로 나눌 수 있으며, 따라서 현재까지는 데이터 시각화, 정보 시각화, 정보 그래픽, 인포그래픽, 정보 일러스트레이션 등을 단일화된 척도로 구분하기는 어렵다[8].

Choi et al.[3]은 데이터 시각화를 정보 시각화, 과학적 시각화, 인포그래픽 등으로 분류했다. Lankow et al.[11]은 시각화의 목적을 매력(appeal), 이해(comprehension), 기억(retention)의 세 가지로 분류하였다. Laine-Hernandez and Särkkä[9]는 세 가지 목적의 우선순위에 따라 과학(scientific), 신문논설(editorial), 마케팅 (marketing), 시장 데이터 (market data)의 4가지 각기 다른 응용 분야를 정의하였다.

Lankow et al.[10]은 인포그래픽 설계를 위한 탐색적(explorative) 관점, 서술적(narrative) 관점을 제시했다. 탐색적 관점은 데이터에 중점을 둔 과학적인 설계방법이며, 서술적 관점은 마케팅 등에 적합한 디자인 중심의 설계방법이다. Segel and Heer[12]는 서술적 시각화(narrative visualization) 설계 방법을 작가 중심(author-driven) 시각화와 독자 중심(reader-driven) 시각화의 양극단 속성으로 분류하고 대부분의 시각화 기법은 그사이에 존재한다고 정의하였다.

Lankow et al.[10]이 분류한 인포그래픽 4분면을 참조하여 본 논문에서는 데이터 중심의 인포그래픽 제작을 위해 데이터 시각화 4분면을 그림1과 같이 제시하였다. 4분면의 X축은 인포그래픽의 콘텐츠가 데이터 지향적인지 또는 디자인 지향적인지로 분류하고, Y축은 인포그래픽의 형식이 수동적인지 또는 능동적인지에 따라 분류하였다.

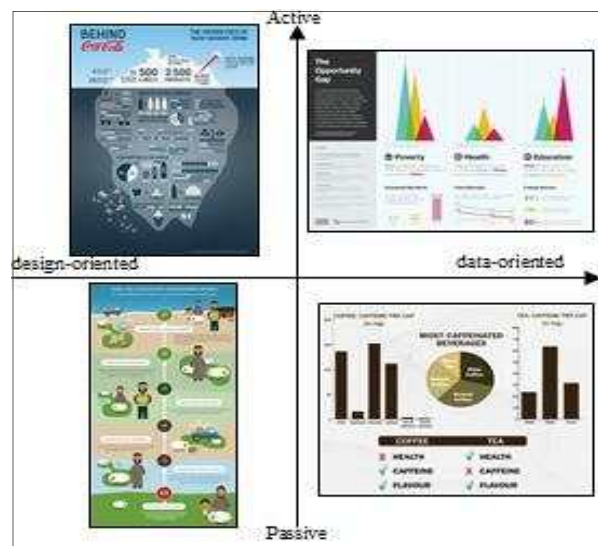


그림 1. Infographics 4분면

1사분면에 제시된 인포그래픽은 코카콜라에 관련된 숨겨진 정보를 물속에 잠긴 빙하로 표현한 인포그래픽으로 아이콘을 중심으로 정보를 전달하고 있다. 2사분면은 통계자료에 색을 효과적으로 적용하여 심미성을 높인 인포그래픽이다. 3사분면은 봉사단체인 Muslim Aid의 생계 관리 프로그램에 관한 인포그래픽으로 이미지 위주로 표현하고 있다. 마지막으로 4사분면은 커피와 차의 효능을 비교한 인포그래픽으로 그래프 위주로 단순하게 표현하고 있다. 인포그래픽 제작을 위해 다양한 방법을 사용할 수 있지만, 빅데이터를 적절히 시각화하는 방법은 바로 2사분면에서 제시되고 있는 데이터 지향적이며 능동적인 형식의 인포그래픽이라 할 수 있다.

우리나라에서 활용되고 있는 대부분의 인포그래픽 제작 설계는 1, 3사분면에 있는 서술식, 보도적 관점의 인포그래픽이다. 그러나 데이터 중심의 인포그래픽은 1, 4사분면에 제시된 과학적인 인포그래픽이며, 이는 다양한 의사결정 시스템에 사용되거나, 과학적인 연구에도 사용될 수 있을 것이다.

2. Tableau Public을 이용한 데이터 시각화

본 논문에서는 현재 대중버스 노선의 이용 실태를 분석하여 정책결정자들을 위한 참고 자료를 제시할 것이며, 이러한 정책 계획과 실행을 결정하기 위한 도구로써 Tableau Public을 이용한 데이터 분석 및 인포그래픽을 제작하여 정책결정자들이 빅데이터에 대한 이해를 쉽게 할 수 있도록 한다.

가. 연구방법 및 데이터

데이터 시각화를 위한 원자료로는 광주광역시 공공빅데이터 오픈소스 - '2015년도 광주광역시 시내버스 정류소별 이용객 수'와 '2016년도 광주광역시 시내버스 정류소별 이용객 수' -를 사용하였다.

이 데이터에서는 정류소명, 정류소번호, 승차 일반, 승차 청소년, 승차 어린이, 환승 일반, 환승 청소년, 환승 어린이의 변수를 보여준다. 이 논문에서는 전체 정류소 이용자 수와 일반 승차, 청소년 승차, 일반 환승, 청소년 환승 승객수의 데이터를 바탕으로 분석하도록 할 것이다. 어린이 데이터는 보호자 즉 일반 데이터와 함께 움직일 가능성을 가지고 있어 배제하도록 하였다. 기본적인 데이터 분석을 위한 시각화는 Tableau Public을 활용하여 데이터를 분석하고 기본적인 데이터 시각화 방법으로 구현하여 데이터 분석결과를 알아보도록 할 것이다. 여기서 나타나는 결과를 활용하여 Tableau Public을 활용한 대시 보드와 데이터 기반의 인포그래픽을 제안하도록 하였다.

나. 연구 결과

(1) 정류소 사용 승객수 비교

변수의 값을 x축과 같은 한 개의 축에 기준으로 하여 표시할

경우 가장 많이 쓰이는 것은 바 차트(bar chart), 라인 차트(line chart), 파이 차트(pie chart)와 같은 기본 차트이다. 가장 단순한 차트 구조인 데이터 값의 크기를 직관적으로 파악할 수 있고, 누구나 쉽게 이해할 수 있다. 따라서 이 논문에서의 데이터 시각화 방법은 막대그래프, 원그래프를 기본적으로 활용하였다.

그림2는 정류소 명과 총계(정류소 사용자 수) 데이터를 근거로 가로막대, 채워진 버블로 표현된 데이터 시각화 예시이다.

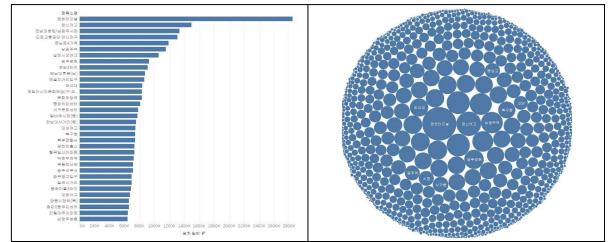


그림 2. Tableau Public을 활용한 시각화 표현1

정류소 사용자 총합계로 본 시각화 결과로 광천터미널이 사용자가 가장 많고, 다음으로 경신여고, 남광주역으로 나타난 것을 알 수 있다. 이렇듯, 변수가 하나일 경우는 비교적 간단한 차트로 구현되며 데이터값 또한 직관적으로 알 수 있다.

그림3은 정류소 명을 기준으로 정류소 총사용자 수와 일반 승차 승객수를 비교하여 나타내도록 데이터 시각화를 표현하였다.

정류소 사용자 총합계로 본 시각화 결과와 비교하여 볼 때 일반 승객 사용자가 가장 많은 곳은 광천터미널이 가장 많고 다음으로 경신여고, 전남대병원으로 대체로 유사한 그래프로 나타났지만 3번째 역을 비롯한 몇몇 역은 총계와 일반 승차 승객수의 차이가 있음을 알 수 있다. 병렬막대그래프, 병렬원그래프, 가로막대그래프로 표현된 데이터 시각화 예시들이다.

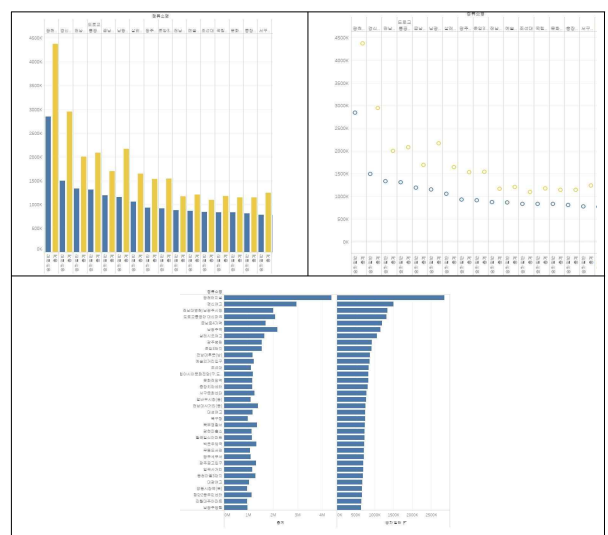


그림 3. Tableau Public을 활용한 시각화 표현2 (정류소 총사용자 수와 일반 승차 승객수)

그림4는 정류소 명을 기준으로 정류소 총사용자 수와 청소년 승차 승객수를 비교하여 나타내도록 데이터 시각화하였다.

정류소 사용자 총합계로 본 시각화 결과와 비교하여 볼 때 청소년 승객 사용자가 가장 많은 곳은 광천터미널, 살레시오여고, 경신여고, 대성 여고 순으로 나타났다. 정류소 승객 총계 그래프와 청소년 승차 승객수의 그래프와는 차이가 있음을 알 수 있다. 병렬막대그래프, 병렬원그래프, 가로 막대그래프로 표현된 데이터 시각화 예시들이다.

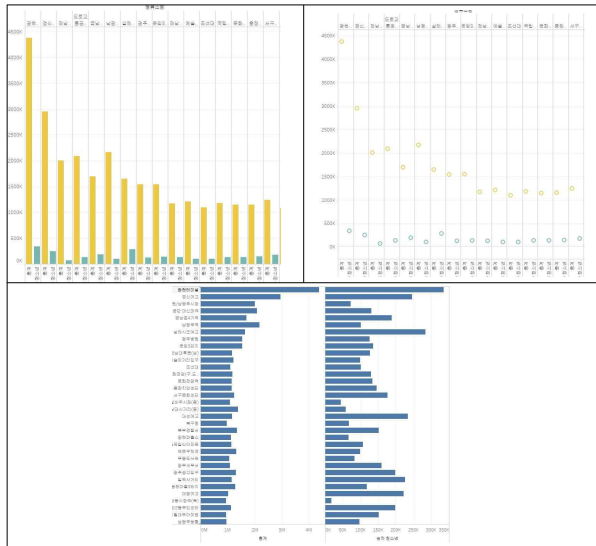


그림 4. Tableau Public을 활용한 시각화 표현3 (정류소 총사용자 수와 청소년 승차 승객수)

정류소 사용자 총 승객수로 본 시각화 결과와 비교하여 볼 때 일반 승차 승객수와는 대체로 유사한 그래프로 나타나지만, 청소년 승차 승객수의 그래프는 차이가 확연하게 나타나는 것을 알 수 있다.

그림5는 정류소 명을 기준(가로축)으로 세로축의 상단은 정류소 총 승객수와 두 번째 블록은 일반 승차 승객수, 세 번째 블록은 청소년 승차 승객수를 표시하여 비교하였다.

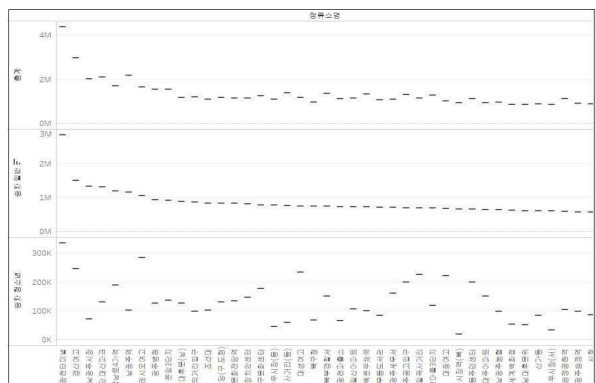


그림 5. 정류소 명을 기준으로 정류소 총 승객수 & 일반 승차 승객수 & 청소년 승차 승객수 분포

그림6은 정류소 사용자 총 승객수 대비 일반 승차 승객수(좌)

와 청소년 승차 승객수(우)의 분산형 차트로 표시하였다. 정류소 사용자 총 승객수 대비 일반 승차 승객수는 비교적 비례하는 그래프로 나타났지만, 정류소 사용자 총 승객수 대비 청소년 승차 승객수는 분산되어 표시되는 것을 알 수 있다. 이 그래프로 일반 승객이 이용하는 정류소는 고르게 분포되어 있지만 청소년 승객이 이용하는 정류소는 특정 지역에 집중되는 있음을 알 수 있다.

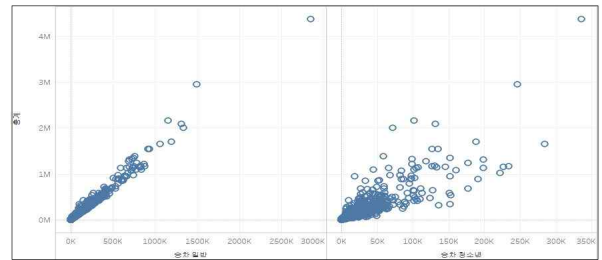


그림 6. 총 승객수 대비 일반승차승객수 & 청소년승차승객수 분산형 차트

(2) 환승 승객수 비교

정류소 총사용자 수와 환승 승객수의 데이터를 비교해 보았다. 환승 승객수가 정류소 총사용자 수와의 연관 관계를 확인해 볼 필요가 있다.

그림 7은 정류소 총사용자 수와 일반 환승 승객 수를 비교하여 데이터 시각화하였다. 정류소 사용자 총합계로 본 시각화 결과로 광천터미널, 경신여고, 남광주역으로 나타난 것에 비해 일반 환승 승객 이용자가 가장 많은 정류소는 경신여고, 광천터미널, 남광주역으로 나타났다. 일반 환승 승객수를 나타낸 그래프를 보면 총 승차 승객수의 그래프와 차이가 있음을 알 수 있다. 환승 승객수를 나타낸 그래프에서는 몇 개소의 정류소가 두드러지게 사용량이 많은 곳이 나타남을 알 수 있다.



그림 7. Tableau Public을 활용한 시각화 표현4 (정류소 총사용자 수와 일반 환승 승객수)

그림 8은 정류소 총사용자 수와 청소년 환승 승객수를 비교하여 데이터 시각화하였다. 청소년 환승 승객 이용자가 가장 많은 정류소는 광천터미널, 경신여고, 남광구역으로 나타났으며, 정류소 총사용자 수의 그래프에 비해 사용자가 큰 정류소와 적은 정류소가 두드러지게 나타남을 알 수 있다. 이는 청소년 승객의 환승은 일정 정류소에서 많이 이루어진다는 것을 알 수 있다.

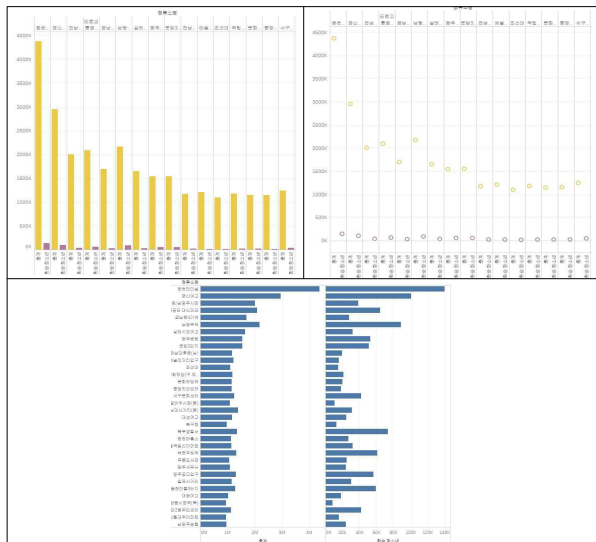


그림 8. Tableau Public을 활용한 시각화 표현5 (정류소 총사용자 수와 청소년 환승 승객수)

그림 9는 정류소 환승 승객수를 일반 환승 승객과 청소년 환승 승객 수를 비교해 보았다. 본 시각화 결과를 그래프의 흐름에서 보면 전체적인 흐름이 비슷하게 나타나는 것을 알 수 있다. 일반 승객과 청소년 승객의 대상에 따른 차이보다는 환승 정류소는 비슷하다는 것을 알 수 있다.



그림 9. Tableau Public을 활용한 시각화 표현6 (일반 환승 승객수와 청소년 환승 승객수)

정류소 사용자 총 승객수로 본 시각화 결과와 비교하여 볼 때 정류소 총사용자 수와 비교했을 때 일반 승차 승객수와 청소년 승객 수는 모두 그래프에서 차이를 보였으며 일반 환승 승객수와 청소년 환승 승객수를 비교했을 때는 대체로 유사한 그래프 유형으로 나타나는 것을 알 수 있었다.

(3) 태블루를 활용한 시각화

Tableau Public을 활용하여 일반적인 접근의 데이터 시각화와 다른 방법의 시각화를 제안해 보도록 하였다.

Tableau Public의 버블채우기 방식을 사용하여 정류소별 승차 승객수의 관계를 살펴보았다. 이때 마크 값의 요소를 색상과 크기로 두고 서로 바꾸었을 때의 결과 값을 비교해 보았다.

버블채우기 결과값의 모양이 비슷한 경우는 두 개의 변수에 따른 결과가 유사한 것이고 서로 다른 모양이 나타나면 결과가 다르다는 것이라고 할 수 있다.

그림 10은 정류소 사용자 총 승객수와 일반 승차 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 총 승객수에 색상을, 일반 승차 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 총 승객수에 크기를, 일반 승차 승객수에 색상을 마크 값으로 준 결과이다. 아주 유사한 결과를 보여주고 있다.

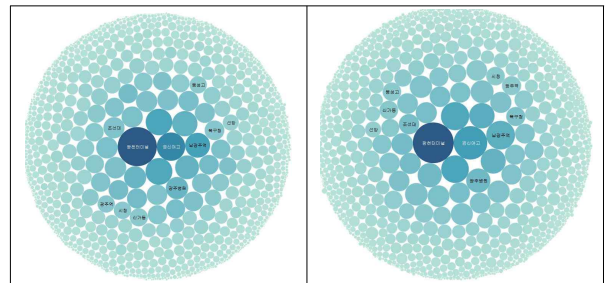


그림 10. 총 승객수(색상) 대비 일반승차승객수(크기) & 총 승객수(크기) 대비 일반승차승객수(색상)

그림 11은 정류소 사용자 총 승객수와 일반 환승 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 총 승객수에 색상을, 일반 환승 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 반대로 총 승객수에 크기를, 일반 환승 승객수에 색상을 마크 값으로 준 결과이다. 시각적으로 차이를 보여주고 있다.

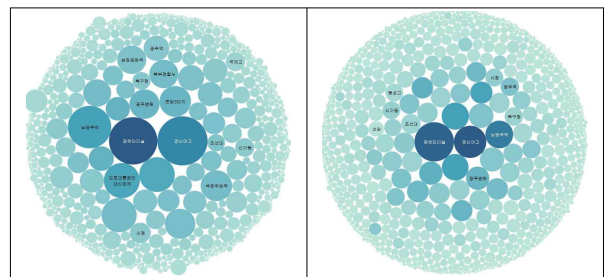


그림 11. 총 승객수(색상) 대비 일반환승승객수(크기) & 총 승객수(크기) 대비 일반환승승객수(색상)

그림 12는 일반 승차 승객수와 일반 환승 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 일반 승차 승객수에 색상을 마크 값으로, 일반 환승 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 반대로 일반 승차 승객수에 크기를, 일반 환승 승객수에 색상을 마크 값으로 준 결과이다. 시각적으로 차이를 보여주고 있다.

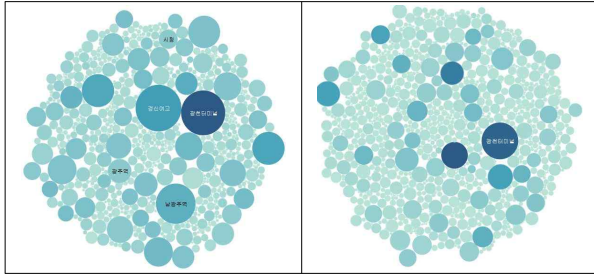


그림 12. 일반승차승객수(색상)대비 일반환승승객수(크기) & 일반승차승객수(크기)대비 일반환승승객수(색상)

그림13은 청소년 승차 승객수와 청소년 환승 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 청소년 승차 승객수에 색상을, 청소년 환승 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 반대로 청소년 승차 승객수에 크기를, 청소년 환승 승객수에 색상을 마크 값으로 준 결과이다. 시각적으로 차이를 보여주고 있다.

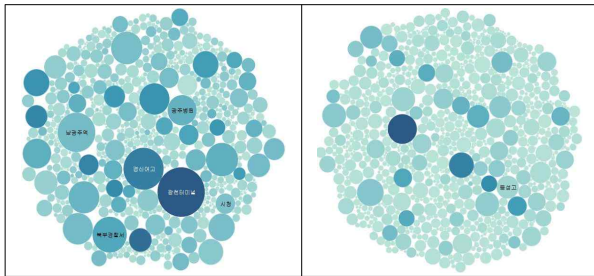


그림 13. 청소년승차승객수(색상)대비 청소년 환승 승객수(크기)& 청소년승차승객수(크기)대비 청소년 환승 승객수(색상)

그림14는 일반 승차 승객수와 청소년 승차 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 일반 승차 승객수에 색상을 마크 값으로, 청소년 승차 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 반대로 일반 승차 승객수에 크기를, 청소년 승차 승객수에 색상을 마크 값으로 준 결과이다. 시각적으로 차이를 보여주고 있다.

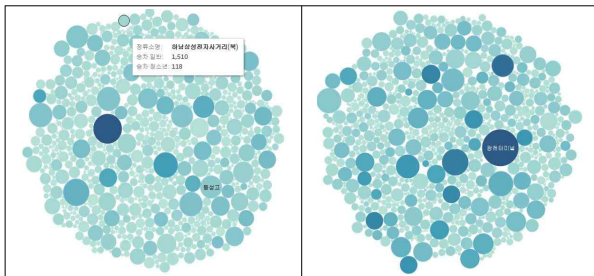


그림 14. 일반승차승객수(색상)대비 청소년승차승객수(크기) & 일반승차승객수(크기)대비 청소년승차승객수(색상)

그림 15는 일반 환승 승객수와 청소년 환승 승객수의 값을 입력했을 때의 결과이다. 왼쪽은 일반 환승 승객수에 색상을 마크 값으로, 청소년 환승 승객수에 크기를 마크 값으로 준 것이고 오른쪽은 반대로 일반 환승 승객수에 크기를, 청소년 환승 승객수에 색상을 마크 값으로 준 결과이다. 시각적인 차이를 보여주고 있다.

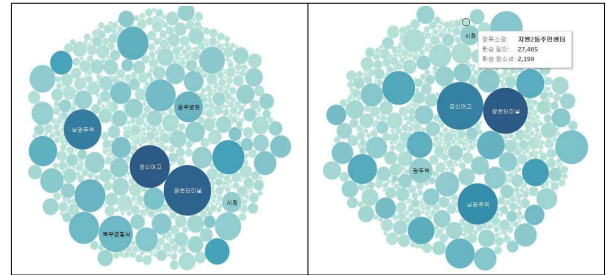


그림 15. 일반환승승객수(색상)대비 청소년 환승 승객수(크기)& 일반환승승객수(크기)대비 청소년 환승 승객수(색상)

그림 16은 태블루에서 대시 보드를 사용하여 시각화한 것이다. 3가지 그래프를 한 번에 보여주면서 복잡한 데이터를 여러 가지 방법으로 동시에 제시함으로써 데이터 이해에 쉬운 방법이라고 할 수 있다.

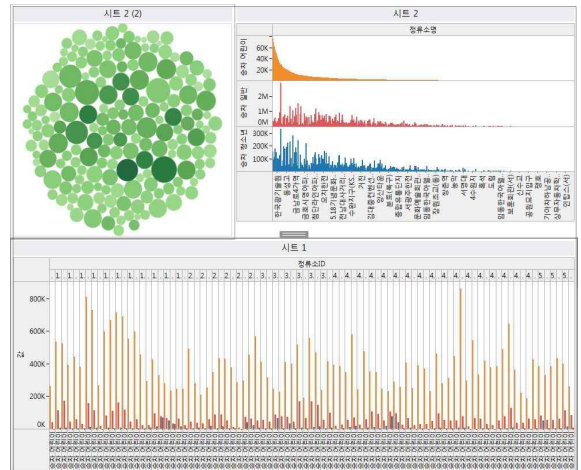


그림 16. Tableau Public을 활용한 시각화 표현_대시 보드

그림 16에서 왼쪽 상단은 채워진 버블그래프, 오른쪽 상단은 병렬막대그래프, 하단에 막대그래프를 배치하였다. 정류소 총 사용자, 일반 승차 승객 수, 청소년 승차 승객 수를 항목별로 고유 색을 지정하여 표시하였다. 이 그래프에서 채워진 버블 그래프는 정류소 총사용자 수는 원의 크기로 확인할 수 있고, 일반 승차 승객 수는 원의 색상(농도)으로 알 수 있으나 청소년 승차 승객 수는 커서를 움직일 때 정류소 명과 나타나는 도구설명에서만 알 수 있는 불편함이 있다.

(4) 다양한 시각화

데이터의 시각화 표현으로 많이 활용되는 것에 인포그래픽이 있다. 기존의 인포그래픽은 데이터를 목적에 맞게 가공하여 시각화에만 중점을 두어 사용하는 경우가 많았다. 그림 17은 이 논문에서 사용된 데이터를 기반으로 하여 다이어그램과 픽토크램 형태의 인포그래픽으로 데이터 시각화하였다.

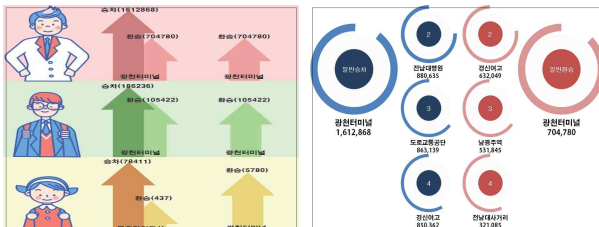


그림 17. 다이어그램과 픽토크램을 활용한 데이터 시각화

그림 18은 Tableau Public의 버블채우기 방식을 응용하여 데이터의 시각화를 제안해 보았다. 변수 값이 한 개였을 경우 버블 채우기 표현의 그래프로 표현된다. 이를 2개의 변수를 동시에 표현하여 비교하기 쉽게 하였다. 각각의 변수에 고유의 색 지정하고 이를 가공하여 시각화하였다. 변수에 따른 대푯값 즉, 일반 승차 및 환승, 청소년 승차 및 환승 승객이 많은 정류소를 전체적으로 한눈에 알 수 있다.

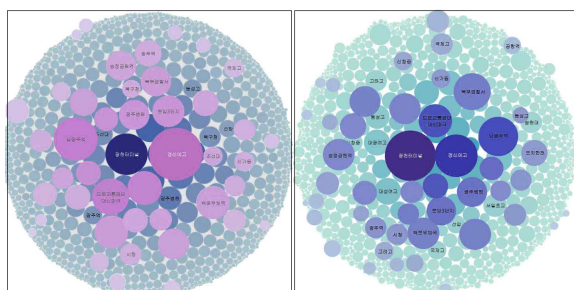


그림 18. 버블데이터채우기를 응용한 데이터 시각화 (일반 승차·환승 & 청소년 승차·환승 승객수)

그림 19는 승차 일반·환승, 청소년 일반·환승 승객수를 다각형 그래프를 응용하여 분류별 색상을 달리하여 표현하였다. 일반, 청소년은 색을 구분하고 해당 변수의 승차, 환승 값은 색상과 투명도를 주어 시각적으로 표현하였다. 분류별 비교가 쉽게 확인됨을 알 수 있다.

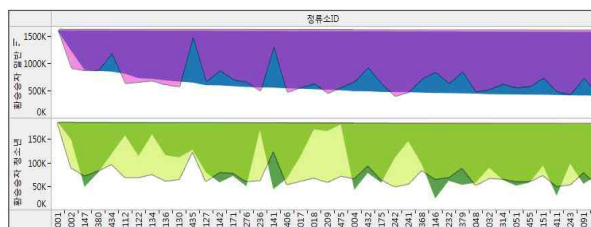


그림 19. 다각형 그래프를 응용한 데이터 시각화 (일반 승차·환승 & 청소년 승차·환승 승객수)

인포그래픽은 디지털 데이터의 규모를 조금이라도 쉽게 가늠할 수 있도록 설명하는 방법이라 할 수 있다.

특정 데이터 혹은 정보의 집합을 시각적인 표나 그림, 그래픽 등으로 전환하여 사용자가 감각적이고 직관적으로 이해할 수 있도록 표현한 것을 의미한다. 인포그래픽은 시각적으로 직관적이며 감성적이어서 광범위한 계층에 효과적으로 정보를 전달할 수 있는 반면에 디자이너의 정보 재가공 과정에서 정보 왜곡의 위험이 있으며, 과도한 시각화는 오히려 정보 독해의 효율을 떨어뜨릴 수 있다. 따라서 인포그래픽을 사용할 때는 시각언어로서의 보편적 표현 양식을 고민해야 할 것이다[9][10].

III. 결 론

앞으로는 지금보다도 더 데이터의 양이 기하급수적으로 늘어날 것이라고 예상된다. 방대한 데이터에서 유익한 정보를 얻기 위해서는 다양한 데이터 분석 기법들이 생겨날 것이다. 너무나도 많은 양의 데이터 때문에 정작 필요로 하는 데이터를 찾는 데 어려움을 겪게 된다. 그리고, 모든 사람이 방대한 데이터를 분석할 수 있지는 않을 것이다. 그렇다면 데이터를 분석하여 사용자(고객)가 이해하기 쉽게 시각화하여 제시하는 것은 매우 중요하다고 볼 수 있다. 또한, 분석된 결과를 한눈에 파악하기 위해서 그에 맞는 적절한 빅데이터 시각화 기법의 연구가 필요하다고 생각된다.

본 논문에서는 데이터 시각화의 방향은 사용자가 원하는 데이터를 사람이 직관적으로 이해하기 쉽고 장기간 기억에 남을 수 있는 이미지로 전환하여 데이터 분석의 결과를 보여주는 것에 목적을 두고 Tableau Public이라는 프로그램을 활용하여 데이터를 왜곡 없이 그래프화 하고 데이터를 단지 그래프화 시키는 것에서 그치지 않고 데이터를 시각화한 본래의 이미지를 훼손시키지 않는 범위에서 그래픽적 요소를 추가하여 일반적인 그래프의 복잡함을 단순화하여 쉽게 이해할 수 있도록 하는 것을 목적으로 하였다. 그래픽 위주의 지금의 인포그래픽에서 쉽게 이해할 수 있다는 장점을 살리고 그래픽 위주의 인포그래픽 일 경우 데이터의 전달이 왜곡되거나 생략되는 단점을 보완할 수 있었다.

데이터를 여러 가지 방향으로 시각화해 보면서 수많은 데이터 속에서 보이지 않았던 패턴이 발견되거나, 데이터에 대한 새로운 질문을 만들어 낼 수 있다[11].는 것에 공감하였다.

기존의 연구들은 데이터 분석방법이나 사람의 시선을 유도하는 등 데이터 시각화 결과물에만 많은 연구를 진행해왔다. 데이터 분석과 시각화의 결과물 자체도 중요하지만, 사용자들의 특성을 고려해서 데이터 시각화를 설계할 필요가 있다. 빅데이터의 필터링과 분석, 이에 따른 사용자 편의를 고려한 시각화까지 연결하는 사고력과 능력을 겸비하여 표현할 수 있어야 할 것이다.

저자 소개

REFERENCES

- [1] 박혜진, “효율적인 인포그래픽 디자인을 위한 정보 시각화 체계와 표현 특성 연구”, *한국브랜드디자인 학연구*, 제15권, 제3호, 185~201쪽, 2017년 9월
- [2] 이지선, “빅데이터를 위한 정보디자인의 시각화 방법 및 표현 연구”, *기초조형학연구*, 제14권, 제3호, 259~269쪽, 2013년 6월
- [3] 이충권, “빅데이터 정보시스템의 구축 및 사례에 관한 연구”, *스마트미디어저널*, 제4권, 제3호, 56~61쪽, 2015년 9월
- [4] 심미희, 류시천, “빅데이터 통계그래픽스의 유형 및 특징-인지적 방해요소를 중심으로-”, *스마트미디어저널*, 제3권, 제3호, 26~35쪽, 2014년 9월
- [5] Visual.ly가 공개한 인포그래픽 자동제작툴 ‘비주얼닷리 크리에이트’ 사용후기(2015년)
<http://viceversablog.co.kr/220416089284>(11, 5, 2018)
- [6] Cairo, A. The Functional Art: An Introduction to Information Graphics and Visualization. Berkeley: *New Riders*, 2012.
- [7] Data Analytics(2015)
[https://blog.naver.com/vipjune85/\(11.5.2018\)](https://blog.naver.com/vipjune85/(11.5.2018))
- [8] Kim, D.H, “A Case Study of Infographics for National Defense - Focusing on the Data journalism of Afghanistan War in Guardian”, *Journal of Korea Spatial Information Society*, vol.22, no.5, pp.43-52, Oct. 2014
- [9] 함유근, 채승병, “빅데이터, 경영을 바꾸다”, *삼성경제연구소*, pp.1-55, 2012년.
- [10] Jason Lankow, Josh Ritchie, Ross Crooks, *Infographics: The Power of Visual Storytelling*, *John Wiley & Sons*, pp.1-263, 2012.
- [11] E Segel, J Heer, “Narrative visualization: Telling stories with data. Visualization and Computer Graphics”, *IEEE Transactions on Visualization and Computer Graphics*, vol.16, no.6, pp.1139-1148, Oct, 2010



문희정(정희원)

1994년 울산대학교 시각디자인학과 학사 졸업.

1998년 일본 국립 큐슈예술공과대학 대학원 정보전달학과 석사 졸업.

2002년 일본 국립 큐슈예술공과대학 대학원 정보전달학과 박사 수료.

<주관심분야 : 인포그래픽, UX/UI,

VR, AR 등>