

데이터 시각화 유형 및 툴의 기능 분석

서미란*, 김희진*, 최은영*, 최유주***, 서정근***°,
 *서울미디어대학원대학교 뉴미디어학부 미디어공학전공
 **서울미디어대학원대학교 실감미디어 연구소, °교신저자
 e-mail : sm11013@naver.com, ggaegury@naver.com, jennychoi1707@naver.com,
 yjchol@smit.ac.kr, jksuh@smit.ac.kr

Analysis of Data visualization types and tools

Miran Seo*, Hee-Jin Kim*, Eunyoung Choi*, Yoo-Joo Choi ***, Jung-Keun Suh ***.°
 *Dept of New Media, Seoul Media Institute of Technology
 **Dept of Media Lab, Seoul Media Institute of Technology, °Corresponding Author

요 약

본 논문에서는 데이터 시각화 목적에 따른 시각화 표현 방식과 6개의 데이터 시각화 툴들의 비교 분석을 통하여 목적에 맞는 차트 유형의 선택의 중요성과 각 툴들이 제공하는 기능과 차이점을 분석해 보았다. 부분적이지만 선정된 각 툴들의 비교를 통해 데스크톱 기반인 환경에서는 비교적 많은 시각화 유형(차트)의 제공이 가능하지만 모바일 환경에서의 제약 사항들을 알 수 있었다. 또한 시각화 툴들이 점차 웹 기반 서비스로 진화하고 있으며 클라우드를 제공하여 결과물들을 쉽게 저장 공유하고 팀플로우 기능이 점점 더 요구될 것으로 보인다. 이에 향후 필요로 하는 다양한 초점의 데이터 시각화 연구를 고찰해보았다.

1. 서론

요즘은 인터넷, SNS 및 스마트 기술들이 보편화되면서 무수한 데이터들이 생성되어 축적 되고 있다. 데이터는 정보를 담고 있고, 다양한 정보들의 집합이 지식이 되며 인간의 사고와 해석, 평가 등이 더해질 경우 지혜가 되는 기반을 제공하고 있다[1]. 그런 무수한 데이터들로부터 유용한 정보를 표현하고 추출하는 것이 무엇보다 중요해졌으며, 데이터 정보를 표현하는 가장 대표적인 방법이 데이터 시각화이다. 데이터 시각화란 장시간 누적된 혹은 수집된 많은 양의 데이터를 기초 자료로 하여 쉽게 이해할 수 있도록 시각적으로 표현하고 전달하는 것을 말한다[2]. 데이터의 시각화는 많은 이점들을 가지고 있으며, 표1은 조사의 응답자 비율에 따라 데이터 시각화의 이점을 보여준다[3].

<표1> 데이터 시각화의 이점

Benefits	Percentages (%)
Improved decision-making	77
Better ad-hoc data analysis	43
Improved collaboration/information sharing	41
Provide self-service capabilities to end users	36
Increased return on investment (ROI)	34
Time savings	20
Reduced burden on IT	15

데이터 시각화의 사용 목적은 최적의 데이터분석을 수행할 수 있도록 지원하며 그래프나 이미지 형태를

이용하여 정보를 전달하는데 효율적으로 전달하는데 있다. 그 시도들은 MS 스프레드 시트나 여러 형태로 널리 보급되고 있으며 많은 데이터 시각화 툴들이 나오고 있다. 이에 데이터 시각화 유형과 툴들을 비교분석하고 방향성을 제시하고자 한다.

데이터 시각화 유형 및 툴들의 기능분석을 위한 연구 범위와 방법은 다음과 같다. 첫째, 시각화 목적에 따른 데이터 시각화 그래프 유형을 분류하고자 한다. 둘째, 현재 많이 쓰이는 데이터시각화 툴 6개(Tableau, Datarapper, Infogram, RAW, ChartBlcok, EmberChart)를 선정하여 기능 분석하고자 한다. 셋째, 현재 데이터 시각화 툴들로부터 나온 분석 결과를 토대로 데이터 시각화를 위해 추가적으로 필요로 하는 기능 및 방향성을 고찰하고자 한다.

2. 데이터 시각화 유형

데이터 시각화는 원 데이터를 알맞게 시각화하여 정보를 전달하는 것에 큰 목적을 가지고 있기 때문에 시각화의 유형의 선택이 중요하다. 같은 원 데이터라고 해도 어떤 시각화 유형을 선택하느냐에 따라 다른 시각화 결과를 얻을 수 있으며 결과의 해석이 다를 수도 있기 때문이다. 데이터를 시각화하는 유형은 다양한 표현 방법이 존재하지만, 가장 일반적인 방법은 ‘차트(chart)’로 표현하는 방법이다. 이에 본 연구에서는 차트를 [4]에서 제시한 5개의 분류 방법에 의해 아래와 같이 분류하였다.

- 분류에 대한 비교: 하나 또는 여러 개의 값을 특정 임계치나 다른 항목들과의 비교를 통해 사용자에게 통찰을 얻을 수 있도록 하는 차트군으로 그림 1은 히스토그램의 사례를 보여 주고 있다..

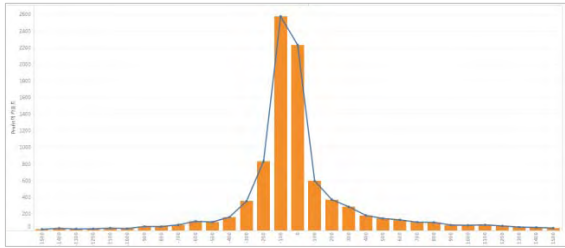


그림 1. Profit 히스토그램
출처: Tableau public

- 구성에 대한 비교: 각 항목들 간에 값의 표현이 서로 영향을 받도록 구성하여 서로의 상대적인 값을 표현하도록 하는 차트군으로서 그림 2는 스쿼어차트의 사례를 보여주고 있다.

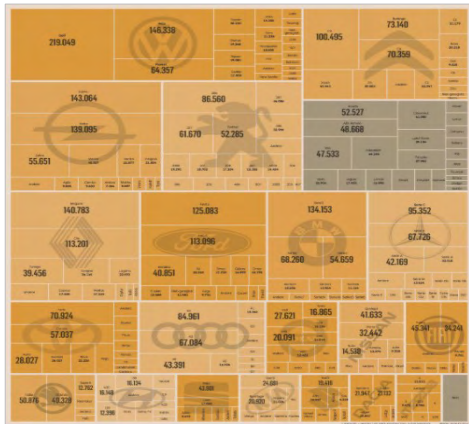


그림 2. 벨기에 자동차 파크하우스 브랜드 및 모델별 현황. 출처: Raw Galley

- 시계열 데이터의 평가: 시간이 경과함에 따라 일정한 간격을 기준으로 측정되는 데이터를 평가할 수 있는 차트군으로서 그림 3은 꺾은선 차트의 사례이다.

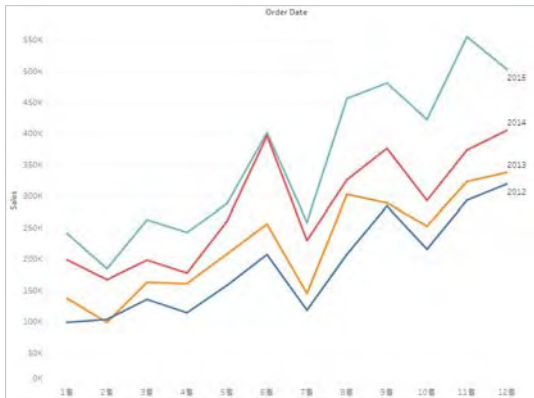


그림 3. 연도별/ 월별 주문량 꺾은선 차트
출처: Tableau public

- 산포도의 관계와 연결: 차트에 산포 되어 있는 데이터를 표현하고 서로의 상관관계, 군집을 찾아내어 통찰하기 위한 차트군으로 그림 4는 분산형차트 예를 보여주고 있다.

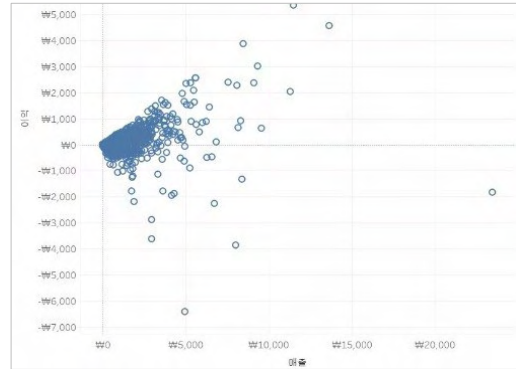


그림 4. 제품, 고개 별 매출과 이익 분산 차트
출처: Tableau public

- 지리적 표현: 지도 그래픽요소와 데이터를 연결하여 위치 정보에 따른 데이터를 통찰하기 위한 차트군으로 그림 5는 지도맵차트의 예이다.

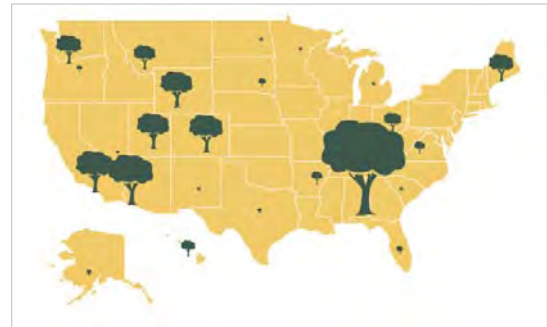


그림 5. 미국 국립 공원지도. 아이콘 크기는 2016년 레크리에이션 방문자 수에 비해
출처: Infogram 블로그

위의 5가지 분류 방법에 의해 가장 보편적으로 사용되고 있는 그래프 유형을 분류하였으며 표 2는 각 군에 포함되는 차트종류를 보여주고 있다.

표 2. 시각화 목적에 따른 그래프 유형

시각화 목적	차트 유형
분류에 대한 비교	막대그래프, 테이블차트, 히스토그램
구성에 대한 비교	파이&도넛차트, 누적막대그래프, 히트맵, 블릿차트, 트리맵, 하이라이트 테이블
시계열 데이터의 평가	라인차트, Gantt뷰, 영역차트
산포도의 관계와 연결	분산형차트, 박스플롯, 히스토그램, Dot plot
지리적 표현	히트맵, 지도맵

3. 데이터 시각화 툴 분석

현재 많이 쓰이는 데이터시각화 툴 6개를 선정하여 구동 하드웨어 플랫폼, 기본적으로 제공하는 데이터 시각화 유형(차트), 사용자 관점의 편의성(클라우드, 팀플로우) 부분을 비교 분석하였다. 분석 대상 6개 툴은 아래와 같다.

- Tableau: 차트, 그래프, 지도를 포함한 다양한 그래픽 기능을 제공하는 비즈니스 전용 시각화 툴이다.[5]
- Datawrapper: 데이터를 업로드하여 차트나 맵으로 레이아웃을 제공하며 주로 출판사와 저널리스트들이 많이 사용한다.[5]
- Infogram: 실시간으로 데이터와 인포그래픽 비주얼 라이제이션을 연동해주는 것이 가장 특징적인 툴이다.[5]
- Raw: 사용자가 단순히 데이터를 붙여 넣을 수 있는 웹 기반 도구로 D3.js 라이브러리를 기반으로 하는 툴이다. [6]
- ChartBlock: 코딩이 필요 없는 스프레드 시트, 데이터베이스 및 라이브 피드에서 시각화를 작성하는 사용하기 쉬운 온라인 도구입니다.[5]
- EmberChart: Ember.js 프레임 워크와 D3.js 를 기반으로 하여 시계열, 막대, 파이 및 분산 형 차트를 제공하는 툴이다.[5]

(1) 구동 하드웨어 플랫폼

구동하드웨어 플랫폼은 크게 데스크톱과 모바일로 나누었으며, 데스크톱은 설치형과 웹(온라인)기반, 모바일은 전용 어플리케이션과 반응형 모바일로 나누어 분류했으며 그 결과는 표 3 과 같다.

표 3. 시각화 툴 별 구동 하드웨어 플랫폼

시각화 툴	데스크톱	모바일
Tableau	데스크톱 설치형 데스크톱 웹(온라인)	전용 어플리케이션
Datawrapper	데스크톱 웹	반응형 모바일 (기능제한)
Infogram	데스크톱 웹	반응형 모바일 (기능제한)
RAW	데스크톱 웹	반응형 모바일 (PC동일작동)
ChartBlock	데스크톱 웹	웹기반 (결과 값만-반응형)
EmberChart	데스크톱 웹	반응형 모바일

(2) 데이터 시각화 표현 유형(차트)

6 개의 툴에서 기본적으로 제공하는 데이터 시각화 표현 유형(차트)를 분류 하였으며 그 결과는 표 4 와 같다.

표 4. 시각화 툴 별 데이터 시각화 차트 유형.
A:Tableau, B:Datawrapper, C:Infogram, D:Raw,
E:ChartBlock, F:EmberChart

차트 유형		A	B	C	D	E	F
테이블	텍스트테이블	Y	Y	N	Y	Y	
	하이라이트테이블	Y	N	N	N	N	
히트맵		Y	N	Y	N	N	N
지도맵	기호맵	Y	Y	N	N	N	
	채워진맵	Y	Y	Y	N	N	
파이차트		Y	Y	Y	N	Y	Y
막대차트	가로막대	Y	Y	N	Y	Y	
	누적막대	Y	Y	N	Y	Y	
	Bullet 막대	Y	Y	Y	Y	Y	
	병렬막대	Y	N	Y	N	N	
트리맵		Y	N	Y	Y	N	N
원뷰	원뷰	Y	N	N	Y	Y	
	병렬원	Y	Y	N	N	N	
라인형 차트	라인(연속형)	Y	N	N	N	N	
	라인(불연속형)	Y	N	N	N	N	
	이중라인	Y	Y	N	Y	Y	
	영역차트(연속형)	Y	N	Y	N	N	
	영역차트(불연속형)	Y	N	Y	N	Y	
이중조합		Y	Y	Y	N	Y	Y
분산형차트		Y	N	Y	Y	N	N
히스토그램뷰		Y	N	Y	N	N	N
박스플롯		Y	N	Y	Y	N	N
Gantt 뷰		Y	N	Y	Y	N	N
불릿 그래프		Y	N	Y	N	N	N
채워진 버블		Y	Y	Y	Y	N	N

(3) 사용자 관점의 편의성

사용자 관점의 편의성은 크게 사용자에게 클라우드를 제공하여 시각화 결과물을 저장 보관하고 자유롭게 수정할 수 있는 지 여부와 시각화한 결과가 다른 사람들에게 공유되어 팀워크를 위한 팀플로우가 가능한지의 기준으로 분류 하였다

표 5. 시각화 툴 별 클라우드제공

시각화 툴	클라우드 제공
Tableau	Y
Datawrapper	Y
Infogram	Y
RAW	N
ChartBlock	Y
EmberChart	N

표 5에서 알 수 있듯이 Tableau, Datawrapper, Infogram, ChartBlock의 4개 툴에서는 사용자 계정을 통한 데스크톱 웹기반에서 데이터 시각화 결과를 저장, 수정할 수 있게 사용자 클라우드를 제공하고 있으며 Raw, EmberCharts는 라이브러리 기반의 툴로 데이터 시각화 결과를 별도로 내보내기, 저장해야 하며 개별 사용자 클라우드를 제공하지 않는다.

표 6. 시각화 툴 별 팀플로우


시각화 툴	팀플로우
Tableau	Y
Datawrapper	N
Infogram	Y
RAW	N
ChartBlock	N
EmberChart	N

표 6과 같이 Tableau, Infogram 툴만이 데이터 시각화 결과를 다른 사람들에게 공유하여 수정, 보완, 피드백을 받을 수 있는 팀플로우가 가능하다

4. 실험

본 연구에서는 6 개 툴의 분석을 토대로 가장 기능이 많은 Tableau 를 이용해 공공데이터 중 지하철의 혼잡도를 가지고 데이터의 시각화를 진행하였다. 시간대별 승하차 인원수 비교는 적은 수의 다른 항목간 비교에 유리한 히스토그램을 사용하였고, 지하철 출퇴근-시간대별 승하차 인원수는 인원수 합계에 유리한 막대 그래프를 사용하였으며, 지하철 역별 승하차 인원수는 다수그룹의 상대비교에 유리한 버블형 그래프를 사용하였다.

표 7. 지하철의 혼잡도 데이터 분석 방법
서울 열린 데이터광장 오픈데이터이용

Row Data	서울시 지하철 호선별 역별시간대별 승하차인원정보에서 사용월,호선명, 지하철역, 시간대별 승차 인원수, 시간대별 하차 인원수
데이터 필터링	2016.03-2017.03 / 1-9 호선 (01 시-04 시 제외)
데이터 선택	Raw 데이터에서 행열 좌우변환: 열(시간대별 승하차 인원수)을 행으로 변환 시간대별 승차인원수+시간대별 하차인원수 = 시간대별 승하차 합계 인원수 호선(2,5,7,9)별이미지 X, Y 축값입력
데이터 시각화 결과	 <p>1)시간대별 승하차 인원수비교 2)지하철 출퇴근-시간대별 승하차 인원수 합계 3)지하철역별 승하차 인원수 합계:버블형</p>

5. 결론

본 논문에서는 데이터 시각화 목적에 따른 시각화 표현 방식과 6개의 데이터 시각화 툴들의 비교 분석을 통하여 목적에 맞는 차트 유형의 선택의 중요성과 각 툴들이 제공하는 기능과 차이점을 분석하였다. 부분적이지만 선정된 각 툴들의 비교를 통해 데스크톱 기반인 환경에서는 비교적 많은 시각화 유형(차트)의 제공이 가능하지만 모바일 환경에서의 제약 사항들을 발견할 수 있었다. 또한 시각화 툴들이 점차 웹기반 서비스로 진화하고 있으며 클라우드를 제공하여 결과물들을 쉽게 저장 공유하고 팀플로우 부분에 점점 더 많은 사용자들의 요구가 있을 것으로 보인다.

이에 향후 연구로서 다음과 같은 다양한 초점에서 데이터 시각의 연구를 수행하여 분야별 특성화된 데이터 시각화 도구의 설계를 위한 기반 연구를 진행하고자 한다.

- 빅데이터에 알맞은 데이터 시각화 표현방법
- 데이터 표현 방식의 다양화(3차원, 다차원 표현 등)
- 색상에 따른 시각화 기대 효과
- 더욱 다양한 분야(바이오, 헬스케어,VR/AR)에서의 시각화 방법 고찰
- 팀플로우의 실시간 SNS 또는 팀워크의 활용에 최적화 할 수 있는 플랫폼 개선

참고문헌

[1] 안춘모, 한국전자통신연구원, 데이터 시각화의 현황과 전망에 관한 연구, 2016년 한국통신학회 하계 종합학술발표회

[2] 두경일, 세명대학교 디지털콘텐츠창작학과 빅데이터의 효과적 시각화를 위한 인포그래픽 연구

[3] V. Sucharitha, S.R. Subash and P. Prakash , Visualization of Big Data: Its Tools and Challenges, International Journal of Applied Engineering Research, 9(18), 2014, pp. 5277-5290.

[4] 다양한 데이터 시각화의 표현방법, <http://www.data.r.co.kr/?mid=textyle&vid=osung&page=3>

[5] Big Data Visualization: Review of the 20 Best Tools By Edoardo L'Astorina in Graphics

[6] R.S. Raghav, Sujatha Pothula, T.Vengattaraman, , Dhavachelvan Ponnurangam : A Survey Of Data Visualization Tools For Analyzing Large Volume Of Data In Big Data Platform, 2016