데이터의 시각화의 유형과 표현 방법

이영주*
*청운대학교 멀티미디어학과
e-mail:yjlee@chungwoon.ac.kr

Types of Visualization of Data and Methods of Expression

Young-Ju Lee*
*Dept. of Multimedia, Chungwoon University

요 약
본 논문에서는 작은 디바이스가 널리 사용되고 있는 요즘 다량의 데이터를 쉽게 구현하고 사용자가 직관적을 이해할 수 있도록 데이터에 맞는 시각화 적용 방법에 대해 알아보았다. 데이터의 시각화 우형은 도표형, 타임라인형, 스토리텔링형, 비교분석형 그리고 지도형으로 구분할 수 있으며 데이터가 가진 정보가 수치적 의미를 가지는 경우는 도표형을 사용하고 개념적 의미를 가지는 경우는 타임라인형과 스토리텔링형을 사용할 필요가 있음을 알 수 있었다. 또 데이터 시각화의 단순화와 직관적 이해를 돕기 위해서는 적은 수의 컬러를 적용하고 굵고 큰 폰트로 데이터의 핵심 키워드나 수치를 강조함으로써 사용자와의 커뮤니케이션을 돕는 다는 것을 알 수 있었다. 레이블링과 데이터 수치의 간격은 데이터의 오류를 줄이는데 도움을 주며 데이터의 종류에 따라 시각화 유형을 선택하고 적용할 필요가 있음을 알 수 있었다.

1. 서론

스마트폰은 데스크톱과 달리 디바이스의 크기가 작아 사용자의 사용성을 높이기 위해 많은 노력을 기울여야 한다. 특히 빅 데이터를 분석하고 응용하는 사례가 늘어나면서 사용자에게 보다 쉽게 데이터를 이해할 수 있도록 하는 노력들은 인포그래픽을 발전시켜왔다. 이제는 우리가 흔히 사용하던 막대그래프나 숫자로 된 엄청난 양의 데이터를 스마트폰에서 보는 일은 줄어들었다. 인간의 시각은 다른 감각보다 더 많은 정보를 습득한다. 따라서 인포그래픽을 통해 데이터를 시각화함으로써 사용자는 작은 디바이스에서도 쉽게 데이터를 읽게 하고자하는 노력을 기울이고 있지만 어떤 데이터를 어떻게 시각화 하는가 하는 방법론적인 부분에 대해서는 그 논의가 적은 편이다.

데이터의 시각화는 바른 수행이 전제되었을 때 효과적이다. 사용자는 표시되는 정보와 표시되는 방법 또는 내용에 대해 시각화를 해석할 수 있어야 한다. 또, 데이터는 왜곡되거나 변질되어서는 안 되며 오류가 발생해서도 안 된다. 특히 데이터의 시각화 과정에서 오해가 발생하거나 데이터 값의 오염이 발생하는 것도 위험하다. 따라서 본 연구에서는 다양한 데이터의 시각화 유형에 대해 알아보고 각 데이터에 맞는 시각화를 적용하는 방법에 대해 알아보고자 한다.

2. 데이터 시각화의 정의

데이터의 시각화는 사용자가 쉽게 이해하고 해석할 수 있는 방법으로 시각전 변화를 통해 커뮤니케이션하는 것을 말한다. 위키피디아에서는 그래픽에 포함된 점, 선, 막대와 같은 시각적 개체를 인코딩하여 데이터 또는 정보를 전달하는데 사용되는 기술을 데이터 시각화로 정의하고 있으며 테크노피디아는 그래픽차트, 그림 및 막대에 데이터 또는 정보를 표시하는 프로세스로 정의하고 있다. 데이터를 시각화 하는 것은 데이터 내에서 스토리를 전달하는 방법을 통해 사용자와 이해 관계자 또는 의사결정 지원을 위해 좀 더 쉬운 방법으로 직관적으로 알 수 있는 스토리텔링을 통해 커뮤니케이션 하는 방법이 필요하다고 한다[1,2].

올바른 데이터의 시각화를 위해서는 다양한 유형의 데이터를 각각의 특징에 맞도록 인포 그래픽이나 차트 또는 다이어그램으로 전환할 필요가 있다. 그를 위해서는 기본적으로 데이터가 가진 정보에 대해 분석해야 한다. 데이터의 잘못된 시각화 선택은 데이터의 오류가 발생되어 잘못된 정보를 제공할 수 있기 때문에 컨텍스트에 맞는 내러티브를 강화하여 데이터에 의미를 부여할 필요가 있다[3]. 이를 위해서는 데이터의 정보가 가지는 종류에 따라 다양한 형태의 시각화 유형 중 맞는 형태를 선택하여야 하며 데이터의 오류를 제거하고 사용

자가 직관적으로 이해할 수 있는 시각화 방법을 고려할 필요가 있다.

3. 데이터 시각화의 유형

복잡한 데이터를 사용자가 직관적으로 이해할 수 있도록 하는 데이터 시각화의 방법 중 하나로 인포그래픽은 자주 사용되는 방법 중 하나이다. 많은 데이터 정보를 압축하고 직관적으로 정리하여 시각적 트리거를 제시함으써 사용자가 한눈에 정보 습득이 가능하도록 도와준다. 인포그래픽은 주로 도표형, 타임라인형, 스토리텔링형, 비교 분석형 그리고 지도형으로 구분할 수 있다. 올바른 데이터의 시각화를 표현하기 위해서는 우선 정보가 가진 데이터가 개념적인지 데이터 기반인지를 구분할 필요가 있다. 정보가 개념적이라면 타임라인형이나 스토리텔링형으로 구현하고 데이터 기반이라면 도표형 또는 비교 분석형의 유형을 선택하는 것이 좋다. 또 데이터 기반인 경우에는 몇 개의 변수를 표시하는지, 변수의 항목의 개수와 그룹간의 값의 표시 등에 따라 신중하게 선택하여야 한다. 각각의 유형은 다음과 같은 특징을 가진다.

3.1 도표형

도표형은 모든 종류의 수치 데이터 표현이 가능하며 막대형, 라인형, 산점도, 스파크라인, 파이가 대표적이다. 막대형은 빠른 정보를 스캔하기 쉬운 경우 사용하며 동일한 카테고리에서 둘 이상의 값을 비교하거나 전체의 일부를 비교할 때 사용한다. 다만 그룹이 너무 많지 않아야 하며 관계를 쉽게 식별할 수 있도록 일관된 색상과 레이블의 사용이 필요하다. 라인차트는 데이터의 추세나 패턴 및 변동을 비교하기 위해 사용한다. 비교하는 각 카테고리에 다른 색상을 사용하는 것과 함께 실선을 사용하여 꺽은 선형 차트를 명확하고 간결하게 유지해야 하며 4개 이상의 카테고리를 비교하는 것은 어렵다. 산점도는 흩뿌려진 형태로 보여지기 때문에 서로 다른 데이터 요소가 많고 데이터 세트에서 유사점을 강조하는 경우 적합한 시각화 방법이다. 어떤 특이 치를 찾거나 데이터의 분포를 이해하는데 유용하다. 또 두 변수 간의 관계를 표시하거나 정보를 빠르게 스캔할하기 위해서는 적합하지 않으며 추세선을 사용하는 경우 혼동을 피하기 위해 1개 또는 2개의 추세선 이상은 사용하지 않는 것이 바람직하다. 모바일에서 가장 많이 볼 수 있는 형태인 스파크라인은 크기가 작기 때문에 추세를 표시하는데 가장 적합한 시각화 방법이다. 스파크라인은 특정 기간 동안 추적 된 상태값이 있는 메트릭스와 페어링이 가능하며 여러 시리즈를 플로팅하거나 정확한 데이터 포인트가 필요한 경우의 사용은 바람직하지 않다. 파이는 전체적인 부분의 관계가 매우 분명한 상태에서 상대 값의 비교나 전체의 일부를 비교하기 위해 사용한다. 파이는 조각의 합이 100%가 되어야 하며 숫자값과 백분율을 추가하여 사용하는 것이 좋다. 비교 범주는 최대 5개 이내의 경우에 사용한다.

3.2 타임라인형

타임라인형은 스파크라인과 같이 특정 기간 또는 시간의 변화 흐름을 가지며 주제에 대해 히스토리나 전개 양상을 시간의 흐름에 따라 보여주게 된다. 한 가지 키워드에 대해 저니맵과 같은 형태로 시간이나 이동에 대한 동선의 설계를 우선시 할 필요가 있다. 이때 전체 이미지 속에서 사용자의 시선의 흐름을 고려해 경로를 설계해 주어야 한다.

3.3 스토리텔링형

눈에 띄는 도표나 수치가 없는 스토리텔링형은 사건이나 주제에 대해 스토리를 구성하여 인포그래픽 형태로 제공해 준다. 키워드와 관련된 줄거리에 핵심 요소에 따른 사건을 이야기를 들려주듯이 전개하며 서사 구조를 가지면 사용자가 오래 기억할 수 있다는 장점이 있다.

3.4 비교 분석형

비교 분석형은 일반적인 데이터 수치 값보다는 대척점에 있는 키워드를 비교하는 방식으로 구성된다. 페이스북과 트위터 또는 코카콜라와 펩시콜라처럼 대척점에 있는 브랜드나 기업을 비교하기에 적합하다. 주로 제품이나 개념을 비교하는데 사용하게 되며 특정 카테고리의 특성을 한눈에 비교 가능하다.

3.5 지도형

지도는 위치를 기반으로 국가나 지역별 통계 자료에 대한 정보를 나타내는데 용이하다. 주로 선거철에 득표율을 내거나 키워드의 선호도나 분포도 등을 나타낼 때 선택 가능한 유형이다.

4. 데이터 시각화의 표현

데이터의 시각화는 모바일에서 대시보드형태로 사용자의 참여를 통해 효과적인 커뮤니케이션을 유도한다. 따라서 정보의 종류에 따라 시각화 유형이 결정되었다면 유형에 맞는 시각적 표현 방법이 적용될 필요가 있다.

우선 정보의 시각화를 위한 컨텍스트를 고려해야 한다. 데이터의 시각화는 내러티브를 강화하고 제시하는 데이터에 의미를 부여하기 위함이기 때문에 명확한 맥락이 추가되어야 사용자가 쉽게 이해할 수 있다. 컨텍스트를 부여하기 위해서는 레이블이나 백분율의 강조만으로도 충분한 핵심 메시지의 전달이 가능하다. 포인트를 강조하시 위해서는 컬러의 사용을 단순화 하여 정보를 분류하고 굵은 폰트로 데이터의 포인트를 구분하거나 강조할 수 있다. 비교가 필요한 데이터는 단일

색상으로 구분할 필요가 있지만 연속 데이터의 경우는 그레이디언트로 구현하는 것도 가능하다.

사용자의 혼란을 야기 시키지 않기 위해서는 단순화를 추구하는 것이 좋지만 필요에 따라 3D 차트를 사용하는 경우에는 값을 추가 하지 않는 방법으로 사용자의 혼란을 제거해 주어야 한다. 너무 많은 컬러의 사용이나 장식용 글꼴 및 아이콘은 사용자의 주의를 산만하게 하기 때문에 기타 요소로 사용은 간단하고 읽기 쉽게 만드는 데이터의 시각화 목적을 저해하는 요소가 된다.

마지막으로 데이터의 왜곡을 방지하는 것은 너무나도 자명하게 중요한 일이다. 데이터의 시각화의 설계에 몰두하여 데이터를 왜곡하거나 잘못된 제시를 하지 않아야 하며 데이터의 종류에 따른 시각화 유형을 고려하여 정확한 데이터를 표시해야 한다. 규칙적인 레이블의 추가와 수직 축의 값을 0에서 시작하도록 하는 등의 노력은 데이터의 오류를 줄여주는 좋은 방법이 될 수 있다.

5. 결론

데이터의 시각화는 작은 화면에서 사용자가 효율적이고 직관적으로 이해하기 위해 대시보드의 형태로 제공되어져 왔다. 대표적인 데이터 시각화의 유형은 도표형, 타임라인형, 스토리텔링형, 비교분석형 그리고 지도형으로 구분할 수 있으며 도표형은 정보가 가진 데이터를 수치화 할 수 있을 때 주로 사용되며 정보가 개념적인 경우에는 타임라인형과 스토리텔링형의 유형을 주로 사용한다.

데이터의 시각화는 레이블링, 데이터의 수치 간격, 컬러, 그리고 폰트를 통해 구분하거나 강조할 수 있으며 적은 컬러 수의 사용과 크고 굵은 폰트를 통해 데이터의 시각화를 단순화하고 사용자와의 직관적 이해를 도울수 있다.

본 연구는 데이터의 시각화에 있어각 유형별 분석을 통해 더 면밀한 방법론적 적용 사례 연구가 없음을 한계로 가지지만 올바른 유형을 선택하고 구현하기 위한 기초 연구로서의 가치가 있다.

참고문헌

[1] https://en.wikipedia.org/wiki/Data_visualization
[2] https://www.techopedia.com/definition/30180/data-visualization
[3] 기무라 히로유키, “인포그래픽스”, 미술문화, pp. 23-32, 9월, 2014년.