

A Guiding System of Visualization for Quantitative Bigdata Based on User Intention

Jung Yun Byun[†] · Young B. Park^{††}

ABSTRACT

Chart suggestion method provided by various existing data visualization tools makes chart recommendations without considering the user intention. Data visualization is not properly carried out and thus, unclear in some tools because they do not follow the segmented quantitative data classification policy. This paper provides a guideline that clearly classifies the quantitative input data and that effectively suggests charts based on user intention. The guideline is two-fold; the analysis guideline examines the quantitative data and the suggestion guideline recommends charts based on the input data type and the user intention. Following this guideline, we excluded charts in disagreement with the user intention and confirmed that the time user spends in the chart selection process has decreased.

Keywords : Bigdata Visualization, Open Source Visualization Tool, Visualization for Quantitative Data, Visualization Guideline, Chart Recommendation

사용자 의도 기반 정량적 빅데이터 시각화 가이드라인 툴

변정윤[†] · 박용범^{††}

요약

기존의 다양한 데이터 시각화 툴에서 제공하는 차트 추천 방식은 사용자의 의도를 고려하지 않은 상태로 차트를 추천한다. 일부 시각화 툴에서는 세분화된 정량적 데이터 분류 체계를 따르지 않기 때문에 명확한 데이터 시각화가 이루어지지 않고 있다. 본 논문에서는 입력된 정량적 데이터를 정확하게 분류하고, 사용자 의도를 반영하여 효율적으로 차트를 추천하는 가이드라인을 제안한다. 가이드라인은 데이터를 분석하는 분석 가이드라인과, 입력된 데이터 타입과 사용자의 의도를 반영하여 차트를 추천하는 추천 가이드라인으로 구성되어 있다. 이러한 가이드라인을 통해 차트 선택 과정에서 사용자의 의도에 부합하지 않는 차트를 배제하였고, 사용자가 차트를 선택하는데 소요되는 시간이 감소하였음을 확인하였다.

키워드 : 빅데이터 시각화, 오픈 소스 시각화 툴, 정량적 데이터 시각화, 시각화 가이드라인, 차트 추천

1. 서론

시각화는 데이터를 단시간에 이해할 수 있는 가장 효과적인 방법이지만, 대부분의 사람들은 자신의 의도에 맞게 차트를 생성하는 시각화 전문 지식이 부족하다[1].

기존의 다양한 데이터 시각화 툴에서는 차트 추천 방식을 제공하고 있지만 이들은 사용자의 의도를 고려하지 않은 상태로 차트를 추천한다. 비전문가들은 시각화 설계 과정에서 자신의 기반 지식에 따라 차트를 선택하게 된다. 이러한 과

정에서 사용자는 본래 표현하고자 했던 목적과 부합하지 않는 차트를 선택하게 되어 원하는 차트를 생성하기 위해 소모적인 과정을 반복하게 된다. 의도에 맞게 적절한 차트를 선택하는 작업은 사용자의 지식에 의존하므로, 이를 지원할 필요가 있다[2]. 따라서 본 논문에서는 사용자의 의도를 고려하여 차트를 효율적으로 선택할 수 있도록 돕는 차트 추천 가이드라인을 설계하였다.

시각화 분야에서는 정량적 데이터 시각화에 대한 연구가 많이 이루어지고 있지만 일부 시각화 툴에서는 세분화된 정량적 데이터 분석 방법을 정의하지 않아 명확한 데이터 시각화가 이루어지지 않고 있다[2-5]. 본 논문에서는 Stevens가 정의한 데이터 분류 체계를 따라 정량적 데이터를 정확하게 분류하여 더욱 명확한 데이터 시각화를 지원하였다[6].

보다 더 효율적인 차트 추천을 위해 본 논문에서는 다음과 같은 내용을 포함한다. 첫째, 입력된 정량적 데이터를 분석하

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 고용계약형 SW석사 과정 지원사업(H0116-16-1015)의 연구결과로 수행되었음.

† 준회원: 단국대학교 컴퓨터학과 석사과정

†† 종신회원: 단국대학교 전자계산학과 교수

Manuscript Received: May 2, 2016

Accepted: May 30, 2016

* Corresponding Author: Young B. Park(ybpark@dankook.ac.kr)

여 알맞은 타입으로 분류하는 분석 가이드라인을 설계하였다. 둘째, 비전문가들이 쉽게 이해할 수 있는 범위 내의 차트를 선정하여, 사용자 의도에 맞는 차트를 선택할 수 있도록 차트 추천 가이드라인을 설계하였다. 셋째, D3.js로 개발된 오픈소스 시각화 툴 RAW에 개선된 분석 가이드라인과 추천 가이드라인을 추가하여 정량적 시각화를 위한 사용자 의도 기반 시각화 가이드라인 시스템으로 발전시켰다[7, 8]. 마지막으로, 가이드라인을 통해 선택된 차트들을 평가하여 차트 추천 시스템의 유용성을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 시각화 설계 관련 지식과 정량적 데이터에 대해 설명하고, 본 논문에서 사용한 오픈 소스 소프트웨어를 소개한다. 3장에서는 정량적 데이터 시각화를 위해 설계된 분석 가이드라인과 추천 가이드라인을 정의한다. 4장에서는 기존 오픈 소스 시각화 소프트웨어 Raw에 가이드라인을 적용한 시스템을 구현한다. 이에 대한 결과 및 분석을 5장에서 보이고, 6장에서는 결론 및 향후 연구 방향을 제시한다.

2. 연구 배경 및 관련 연구

2.1 빅데이터 시각화

빅데이터 시대에 점점 데이터의 양이 방대해지고 있지만, 가공되지 않은 원천 데이터(raw data)만으로는 의미 있는 정보를 얻을 수 없다. 최근에는 데이터를 재구성하여 의미를 찾아내고자 하는 기술이 주목 받고 있다[9]. 시각화 기술을 활용하면 다양하고 방대한 양의 데이터를 복잡한 가공 작업 없이 더욱 쉽고 빠르게 살펴볼 수 있다[10]. 시각화를 통해 데이터를 효과적으로 이해할 수 있도록 도와주며, 특히 직관적 통계 분석이 어려운 빅데이터 분석에서 데이터간 관계를 직관적으로 제공하여 올바른 의사 결정을 할 수 있도록 지원한다[9].

2.2 데이터 시각화 설계

데이터 시각화는 데이터 속성을 적절한 시각화 요소로 매핑시켜 정보를 전달하는 것이다. 정확한 정보 전달을 위해서는 체계적인 시각화 설계 과정을 따라야 한다[2].

기존 시각화 연구에서 시각화 설계 과정은 다양하게 정의되고 있다[2, 11]. Fig. 1은 Andy Kirk의 시각화 설계 과정을 나타낸 그림으로, 효과적인 정보 전달을 위해 1) 시각화 목표 설정 단계, 2) 시각화 데이터 파악 단계, 3) 시각화 설계 단계, 4) 시각화 구현 및 평가 단계를 제시하였다[2].



Fig. 1. Andy Kirk's Visualization Design Process

2.3 정량적 데이터

1) 정량적 데이터 분류 체계

시각화 분야에서 데이터 분류 체계에 관련된 연구는 오랜

역사를 지니고 있다[2, 3, 4, 5, 12, 13, 15]. 데이터를 적절한 시각화 요소로 매핑시키기 위해서는 먼저 데이터 속성의 기본 개념을 이해하는 것이 중요하다[2].

많은 시각화 연구에서 Stanley Smith Stevens가 정의한 데이터 분류 체계를 따른다[2-5, 12-15]. Stevens는 데이터를 크게 정성적 데이터(Qualitative Data)와 정량적 데이터(Quantitative Data)로 분류하였다[6]. 이러한 데이터 분류 체계는 과학 분야 및 공학 분야에서 역시 널리 적용되고 있다[13]. 시각화 분야에서 정량적 데이터는 다음과 같이 두 분류로 정의된다[2, 13].

1. 정량-등간 데이터(Quantitative-Interval Data)
2. 정량-비율 데이터(Quantitative-Ratio Data)

정량-등간 데이터는 값들의 '차이'에 의미가 있는 데이터로 영점(zero point)이 절대적이지 않기 때문에 비율적으로 의미가 없는 데이터다. 반면, 정량-비율 데이터는 값의 '비율'에 의미가 있는 데이터로 절대적인 영점이 존재한다[10].

2) 정량적 데이터 시각화 요소

시각화 분야에서 정량적 데이터 시각화 설계에 관한 연구가 활발하게 이루어지고 있다[3-5, 12-14]. 데이터 분석을 위한 다양한 시각적 요소가 존재하지만, Stephen Few는 대부분의 정량적 데이터가 오직 네 가지 시각적 요소만으로도 충분히 표현될 수 있다고 주장하였다. Table 1에서는 Stephen Few가 제시한 정량적 데이터를 위한 시각적 요소와 각 시각적 요소의 표현 목적을 나타낸다[3].

Table 1. Visual Object and Purpose of Visual Object for Quantitative Data Visualization by Stephen Few

| 시각적 요소 | 표현 목적 |
|--------|-------|
| Bar | 비교 |
| Box | 분포 |
| Line | 패턴 |
| Point | 관계 |

시각적 요소 Bar를 활용하면 데이터값의 크고 작음을 비교할 수 있다. Box의 경우 Bar와 유사하게 동작하지만, 값들의 분포를 나타내는 데 특히 유용하게 쓰인다. Line의 경우, 데이터 셋의 전반적인 추세나 시간에 따른 변화 패턴을 파악하는 데 도움이 된다. 산점도와 같이 Point 요소를 사용하여 나타난 차트에서는 각 요소들의 위치, 크기에 따라 값들의 관계를 파악할 수 있다[3].

2.4 오픈 소스 시각화 소프트웨어: Raw

Raw는 DenstiyDesign 연구소에서 개발한 웹 기반 오픈 소스 시각화 소프트웨어다. Raw는 D3 자바스크립트 라이브러리를 활용하여 사용자 정의 백터 기반 시각화를 지원한다. Raw의 설계 목적은 스프레드시트(e.g. Microsoft Excel, Apple Numbers, Google Docs, ...)와 백터 그래픽 에디터(e.g. Adobe Illustrator, Ink scape, ...) 사이의 연결을 제공해주는 것이다[8]. 즉, Raw는 입력된 데이터 변수와 시각적 요소를 매핑해주는 데이터 형상화 작업을 지원해주는 툴이다[2].

2.5 오픈 소스 시각화 라이브러리: D3.js

D3는 Data-Driven Documents의 약자로, Mike Bostock에 의해 개발된 오픈 소스 시각화 라이브러리다[7]. D3에서 데이터(data)는 사용자가 입력하는 데이터를 의미하며, 문서(document)는 HTML이나 SVG처럼 웹 브라우저에서 렌더링할 수 있는 웹 기반 문서를 의미한다[17].

3. 사용자 의도 기반 시각화 가이드라인 설계

3.1 차트 범위 한정

본 논문에서 제안하는 시각화 가이드라인 시스템은 D3.js로 개발된 오픈 소스 시각화 툴 RAW에 분석 가이드라인과 추천 가이드라인을 추가하여 구현하였다. D3는 SVG를 통해 시각적 요소를 생성하고 조작하며, SVG로 표현할 수 있는 시각적 요소는 rect, circle, ellipse, line, path, point 등이 있다[17].

정량적 데이터 시각화를 위하여 Stephen Few가 정의한 네 가지 시각적 요소로 차트 범위를 한정시켰다[3]. 네 가지 시각적 요소 Bar, Box, Line, Point는 각각 SVG 요소와 대응된다. Bar와 Box는 rect로, Line은 line으로, Point는 point로 매핑된다.

시각화 전문 지식이 부족한 사용자들에게 익숙한 차트를 제시하여 사용자가 목표한 데이터 시각화 결과를 더욱 효과적으로 이끌어낼 수 있도록 한다. D3 'Basic Chart'에 구현되어 있는 차트 중 rect로 생성할 수 있는 차트는 (1) Bar Chart, (2) Stacked Bar Chart, (3) 100% Stacked Bar Chart, (4) Box Plot가 있다. line을 이용하여 생성할 수 있는 차트는 (1) Line Chart, (2) Area Chart, (3) Stacked Area Chart가 있다. 마지막으로, point로 생성할 수 있는 차트는 (1) Scatter Plot이 있다. 본 논문에서는 Table 2에 정의된 8개의 차트로 범위를 한정하여 정량적 데이터 시각화를 위한 가이드라인을 설계하였다.

Table 2. Types of Chart for Quantitative Data Visualization

| 시각적 요소 | SVG 요소 | 차트 이름 | 표현 목적 |
|--------|--------|------------------------|-------|
| Bar | rect | Bar Chart | 비교 |
| | | Stacked Bar Chart | |
| | | 100% Stacked Bar Chart | |
| Box | | Box Plot | |
| Line | line | Line Chart | 패턴 |
| | | Area Chart | |
| | | Stacked Area Chart | |
| Point | point | Scatter Plot | 관계 |

3.2 사용자 의도 정의

기존의 시각화 설계 과정은 다양하고, 많은 연구들에서 제시된 시각화 방법론 또한 수없이 많이 존재한다[2, 3, 10, 11, 14, 16]. Andy Kirk가 제시한 스토리 발견 방법[2], Stephen Few의 7가지의 시각화 방법론[3], Data Visualization Catalogue에서 정의한 16개의 데이터 표현 기능[18]에는 비교, 관계, 패턴이라는 공통 요소가 발견된다. 본 논문에서는 세 가지 공통 요소를 시각화 가이드라인 시스템의 '주 의도(Main Intention)'로 설정하였다. 본격적인

시각화 설계에 앞서, 사용자는 자신의 의도를 선택함으로써 더욱 명확한 시각적 표현을 할 수 있다[2]. 추가적으로 위에서 제시된 시각적 방법론들을 조합하여 각 '주 의도'를 여러 가지의 '자세한 의도(Detail Intention)'로 상세화하였다. Table 3에서는 주 의도와 주 의도에 해당되는 자세한 의도를 나타낸다.

Table 3. Main Intention and Detail Intention

| 주 의도 (Main Intention) | 자세한 의도 (Detail Intention) |
|--------------------------|------------------------------------|
| 비교 (Comparisons) | 카테고리별 비교(Comparison by Categories) |
| | 시간에 따른 비교(Comparison by Time) |
| | 비율(Proportion) |
| | 범위(Range) |
| 패턴 (Patterns) | 분포(Distribution) |
| | 전반적인 추이(Overall Trend) |
| 관계 (Relationships) | 패턴(Pattern) |
| | 관계(Relationship) |
| | 전체-부분(Part-to-Whole) |

3.3 데이터 분석 가이드라인 생성

본 논문에서는 정량적 데이터를 정량-등간 데이터 타입과 정량-비율 데이터 타입으로 세분화하였다. 입력된 데이터에 따라 매핑되는 시각적 요소는 다르며, 선택될 수 있는 차트 종류 또한 다르게 나타난다.

Fig. 2는 입력 데이터를 정량-등간 데이터 타입과 정량-비율 데이터 타입으로 분류하는 알고리즘을 의사 코드로 설명한 것이다. 입력된 데이터의 영점 비율이 일정 수준(threshold) 이상 되는 데이터 셋의 경우 정량-비율 데이터 타입으로 분류하고, 그렇지 않은 경우 정량-등간 데이터 타입으로 분류한다.

Algorithm 1: Data Analysis

```

Input: DataSet
       입력 데이터 셋
Output: DataType
        정량적 데이터 타입

Algorithm:
p=Proportion_of_Zeropoint(DataSet)
IF (p > θ)
    RETURN DataType(DataSet)=Quantitative-
    Ratio Type
ELSE
    RETURN DataType(DataSet)=Quantitative-
    Interval Type
ENDIF
WHERE θ is a Threshold
    
```

Fig. 2. Algorithm of Data Analysis

하지만 이러한 분류는 사용자가 데이터를 보는 관점에 따라서도 달라질 수 있다. 데이터 분석 시, 값들의 차이와 비율 중 어느 쪽에 더 중점을 두느냐에 따라 데이터 타입이 달라질 수 있다. 이러한 경우 분석 알고리즘에 따라 정량-비율 타입으로 간주된 데이터일지라도, 추후에 사용자가 정량-등간 타입으로 변경할 수 있도록 설계하였다.

3.4 차트 추천 가이드라인 생성

입력 데이터 타입과 데이터 표현 목적에 따라 선택할 수

있는 차트의 종류는 다르다[2, 3, 10, 13]. 본 논문에서는 입력된 데이터의 타입을 분석하고, 사용자의 의도를 입력받아 적절한 차트를 추천하는 시각화 가이드라인 틀을 제안한다.

Fig. 3은 분석된 정량적 데이터 타입과 사용자가 선택한 의도에 따라 차트를 추천하는 알고리즘을 의사 코드로 설명한 것이다. 선택된 주 의도와 자세한 의도, 그리고 입력된 데이터 타입에 따라 보여지는 차트의 종류가 다르게 나타난다.

```

Algorithm 2: Chart Recommendation
Input: Main Intention, Detail Intention
      주 의도와 자세한 의도
Output: Print Types of Chart
        입력 데이터와 선택된 의도에 따른 차트 종류

Algorithm:
GET Main Intention, Detail Intention
SWITCH (DataType (DataSet))
CASE Quantitative-Interval Data Type
IF (Main Intention으로 Comparisons가 선택되고 Detail Intention으로 Comparison by Categories가 선택됨)
RETURN Print Line Chart, Stacked Area Chart
ELSE IF (Main Intention으로 Comparisons이 선택되고 Detail Intention으로 Comparison by Time이 선택됨)
RETURN Print Line Chart, Area Chart, Stacked Area Chart
.
.
END IF
END CASE

CASE Quantitative-Ratio Data Type
IF (Main Intention으로 Comparisons가 선택되고 Detail Intention으로 Comparison by Categories가 선택됨)
RETURN Print Bar Chart, Stacked Bar Chart, 100% Stacked Bar Chart, Box Plot, Line Chart, Stacked Area Chart
ELSE IF (Main Intention으로 Comparisons이 선택되고 Detail Intention으로 Comparison by Time이 선택됨)
RETURN Print Line Chart, Area Chart, Stacked Area Chart
.
.
END IF
END CASE
END SWITCH
    
```

Fig. 3. Algorithm of Chart Recommendation

1) 입력 데이터 타입에 따른 차트 추천 룰

입력 데이터와 시각화 요소 매핑에 대해 다양한 연구들이 존재한다[2, 3, 10, 11]. Table 4에서는 Andy Kirk가 정의한 시각화 설계 기법을 따라 분석된 입력 데이터에 따른 차트 추천 룰을 제시한다.

Table 4. Rules of Chart Recommendation based on Input Data

| 입력 데이터 | 차트 종류 |
|--|--------------|
| 1x범주형 데이터, 1x정량-비율 데이터 (1xCategorical, 1xQuantitative-Ratio) | Bar Chart |
| 1x정량-등간 데이터, 4x정량-비율 데이터 (1xQuantitative-Interval, 4xQuantitative-Ratio) | Box Plot |
| 1x정량-등간 데이터, 1x정량-비율 데이터, 1x범주형 데이터 (1xQuantitative-Interval, 1xQuantitative-Ratio, 1xCategorical) | Line Chart |
| 2x정량 데이터 (2xQuantitative) | Scatter Plot |

2) 사용자 의도 선택에 따른 차트 추천 룰

Table 5에서는 입력된 의도에 따른 차트 추천 룰을 제시한다. 입력된 의도뿐 아니라 세분화된 정량 데이터 타입에 따라 추천될 수 있는 차트의 종류는 다르다. Table 5에 제시된 룰은 Table 2에서 제한한 차트 범위와 Table 4에서 제시된 입력 데이터에 따른 차트 가이드라인 룰을 조합하여 생성하였다.

Table 5. Rules of Chart Recommendation based on User Intention

| 주 의도 | 자세한 의도 | 정량적 데이터 타입 | 차트 종류 | |
|------|-----------|------------------------|------------------------|------------------------|
| 비교 | 카테고리별 비교 | 정량-등간 | Line Chart | |
| | | | Stacked Area Chart | |
| | | 정량-비율 | Bar Chart | |
| | | | Stacked Bar Chart | |
| | | | 100% Stacked Bar Chart | |
| | | | Box Plot | |
| | 시간에 따른 비교 | 정량-등간 | Line Chart | |
| | | | Area Chart | |
| | | 정량-비율 | Stacked Area Chart | |
| | | | Line Chart | |
| | | | Area Chart | |
| | | | Stacked Area Chart | |
| 비율 | 정량-등간 | - | | |
| | | Stacked Bar Chart | | |
| | 정량-비율 | 100% Stacked Bar Chart | | |
| 범위 | 정량-등간 | - | | |
| | | 정량-비율 | Box Plot | |
| | 분포 | 정량-등간 | - | |
| 패턴 | 전반적인 추이 | 정량-등간 | Line Chart | |
| | | | Area Chart | |
| | | 정량-비율 | Stacked Area Chart | |
| | | | Line Chart | |
| | | | Area Chart | |
| | | | Stacked Area Chart | |
| | 패턴 | 정량-등간 | Line Chart | |
| | | | Area Chart | |
| | | | Stacked Area Chart | |
| | | 정량-비율 | Scatter Plot | |
| | | | Bar Chart | |
| | | | Stacked Bar Chart | |
| 관계 | 관계 | 정량-등간 | 100% Stacked Bar Chart | |
| | | | Box Plot | |
| | | 정량-비율 | Line Chart | |
| | | | Stacked Area Chart | |
| | | | Area Chart | |
| | | | Scatter Plot | |
| | 전체-부분 | 정량-등간 | - | |
| | | 정량-비율 | Stacked Bar Chart | |
| | | | | 100% Stacked Bar Chart |

4. 구현

본 논문에서 제안하는 ‘사용자 의도 기반 정량적 빅데이터 시각화 가이드라인 툴’은 Raw에 개선된 분석 가이드라인과 추천 가이드라인을 추가하여 구현되었다.

4.1 입력 데이터 분석 가이드라인 개선

Raw에는 입력 데이터에 대한 분석이 있다. Raw에서 지원하는 데이터 타입은 date, string, number다. 본 논문에서는 정량적 데이터 타입을 세분화하여 사용자가 더욱 명확한 데이터를 표현할 수 있도록, 3.3절에서 생성한 분석 가이드라인을 통해 기존 Raw의 분석 알고리즘을 개선하였다. Table 6은 Raw에서 지원하는 데이터 타입과 본 논문에서 제안하는 데이터 타입을 나타낸다.

Table 6. Data Types

| Raw에서 지원하는 데이터 타입 | 본 논문에서 제안하는 데이터 타입 |
|-------------------|-----------------------|
| date | date |
| string | string |
| number | quantitative-Interval |
| | quantitative-Ratio |

4.2 차트 추천 가이드라인 추가

Raw의 기본 구성은 다음과 같다.

1. 데이터 입력 섹션
2. 차트 리스트 섹션
3. 데이터 변수와 시각적 요소를 매핑하는 섹션
4. 차트 사용자 정의 섹션
5. 다운로드 섹션

본 논문에서는 Raw 기본 구성에 추가적으로 (1) 주 의도 선택 섹션과 (2) 자세한 의도 선택 섹션을 추가하여 총 7개의 섹션으로 시각화 가이드라인 시스템을 구성하였다. Fig. 4는 주 의도를 선택하는 섹션이고, Fig. 5는 자세한 의도를 선택하는 섹션을 나타낸다.

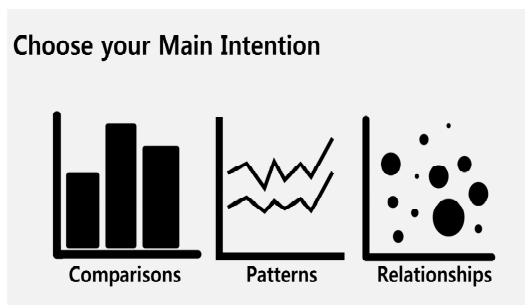


Fig. 4. Main Intention Section (Icon was Designed by Freepik)

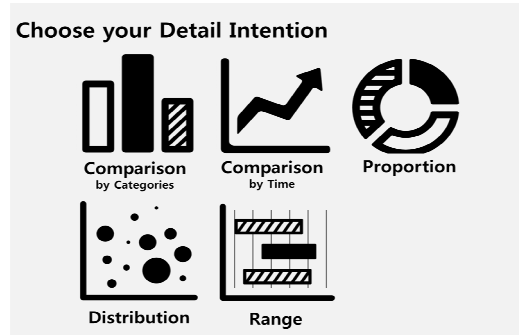


Fig. 5. Detail Intention Section (Icon was Designed by Freepik)

5. 구현 결과 및 분석

Raw에 가이드라인이 추가되어 시스템 구성이 4단계에서 6단계로 늘어났지만, 차트 선택 과정에서 사용자 의도에 부합하지 않는 차트를 배제함으로써 사용자가 차트를 선택하는 시간이 감소하였음을 확인하였다.

가이드라인 없이 차트를 선택하는 과정에서 ‘차트 선택 단계’에 나타나는 차트의 개수는 8개다. 사용자는 시각화 관련 사전 지식을 통해 8개 중 한 개의 차트를 선택해야 한다. 비전문가들은 시각화 설계 과정에서 자신의 기반 지식에 따라 차트를 선택하게 된다. 이러한 과정에서 사용자는 본래 표현하고자 했던 목적과는 부합하지 않는 차트를 선택하게 되고, 원하는 차트를 생성하기 위해 소모적인 과정을 반복하게 된다[2].

반면, 본 논문에서 설계한 차트 추천 가이드라인을 통해 ‘차트 선택 단계’에 나타나는 차트의 개수는 평균 5.3개다. 추가적으로 정량적 데이터를 세분화하여 사용자가 더욱 명확한 데이터를 표현할 수 있도록 분석 가이드라인을 구현하였다. 이러한 분석 가이드라인을 통해 ‘차트 선택 단계’에 나타나는 차트의 개수는 평균 2.6개다. Fig. 6은 분석 가이드라인과 차트 추천 가이드라인에 따라 추천되는 차트의 개수를 비교하여 나타낸다.

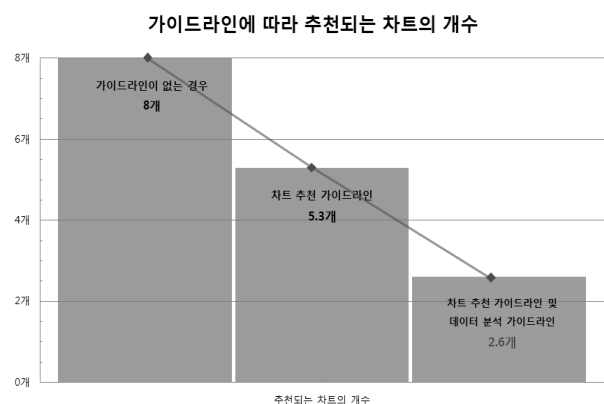


Fig. 6. The Number of Recommended Charts based on Guideline

6. 결론 및 향후 연구 방향

본 논문에서는 정량적 데이터를 세분화하여 사용자가 더욱 명확한 데이터 시각화를 할 수 있도록 분석 가이드라인을 설계하였다. 또한, 입력 데이터와 사용자의 의도를 고려하여 적절한 차트를 추천하는 시각화 가이드라인 툴을 설계하고 구현하였다.

결론적으로, 제안된 추천 시스템을 통해 선택된 차트들이 별도의 가이드라인 없이 선택된 차트들보다 더욱 효과적으로 사용자의 의도에 맞게 데이터를 표현하였음을 확인하였다. 본 시스템의 유용성을 시간적 효율성 측면에서만 확인하였지만, 이에 그치지 않고 추천된 차트에 대한 설문조사, 사용자 인지도 조사까지 확장하여 보다 명확한 유용성 검증이 이루어져야 한다.

본 시스템에서는 정량적 데이터 타입을 이론적 정의로 분류하였다. 하지만 현실 세계의 많은 데이터들을 모두 이론적 정의로만 분류할 수는 없다. 정량적 데이터를 정량-등간 데이터와 정량-비율 데이터로 분류할 수 있는 더욱 효과적인 알고리즘 개발에 대한 연구가 필요하다. 나아가 정형 데이터 뿐 아니라 비정형 데이터까지도 분석이 가능한 시각화 가이드라인을 설계한다면 빅데이터를 더욱 효율적으로 분석할 수 있다.

References

[1] Jock D. Mackinlay, Pat Hanrahan, and Chris Stolte, "Show me: Automatic presentation for visual analysis," *Visualization and Computer Graphics, IEEE Transactions on*, Vol.13, No.6, pp.1137-1144, 2007.

[2] Andy Kirk, "Data Visualization: a successful design process," Packt Publishing Ltd., 2012.

[3] Stephen Few, "Now you see it: simple visualization techniques for quantitative analysis," Analytics Press, 2009.

[4] Stephen Few, "Show me the numbers," Analytics Press, 2004.

[5] Edward R. Tufte and P. R. Graves-Morris, "The visual display of quantitative information," CT: Graphics press, 1983.

[6] Stanley Smith Stevens, "On the theory of scales of measurement," *Science, New Series*, Vol.103, No.2684, pp. 677-680, 1946.

[7] Mike Bostock, D3: Data-Driven Documents [Internet], <https://d3js.org/>.

[8] DensityDesign Lab, RAW [Internet], <http://raw.densitydesign.org/>.

[9] Hyungnyun Kim, "Case Study of Bigdata Visualization -Centre around the Visual Representation Form-," *Journal of Integrated Design Research*, Vol.13, No.4, pp.125-136, 2014.

[10] 모리후지 다이치, 안티베이지안, 엔지니어를 위한 데이터 시각화: D3.js로 배우는 데이터 시각화 이론과 12가지 사례, 김성재 옮김, 한빛 미디어, 2015.

[11] Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment," O'Reilly Media, Inc., 2007.

[12] Munzner, Tamara. "Interactive visualization of large graphs and networks," Ph.D. dissertation, Stanford University, 2000.

[13] Senay, Hikmet and Eve Ignatius, "Rules and principles of scientific data visualization," Institute for Information Science and Technology, Department of Electrical Engineering and Computer Science, School of Engineering and Applied Science, George Washington University, 1990.

[14] Colin Ware, "Information visualization: perception for design," Elsevier, 2012.

[15] Jacques Bertin, "The Semiology of Graphics," University of Wisconsin Press, 1983. (First edition 1967).

[16] Michael Friendly, "A brief history of data visualization," Handbook of data visualization. Springer Berlin Heidelberg, pp.15-56, 2008.

[17] Scott Murray, "Interactive data visualization for the Web," O'Reilly Media, Inc., 2013.

[18] Severino Ribecca, The Data Visualisation Catalogue [Internet], <http://www.datavizcatalogue.com/>.



변 정 운

e-mail : byunit@naver.com
 2015년 단국대학교 컴퓨터학과(학사)
 2015년~현 재 단국대학교 컴퓨터학과
 석사과정
 관심분야 : 소프트웨어 품질, 빅데이터
 시각화



박 용 범

e-mail : ybpark@dankook.ac.kr
 1985년 서강대학교 전자계산학과(학사)
 1987년 폴리테크닉대학교 전자계산학(석사)
 1991년 폴리테크닉대학교 전자계산학(박사)
 1993년~현 재 단국대학교 전자계산학과
 교수
 관심분야 : 지능형 소프트웨어 공학, 소프트웨어 보안, 소프트웨어 품질