

Tableau Public을 이용한 인포그래픽 제작 사례연구

A Case Study of Producing Infographics Using Tableau Public

김동환*

Dong Hwan Kim

요약 최근 디지털 데이터가 폭발적으로 증가함에 따라 빅데이터, 데이터시각화, 정보시각화, 인포그래픽 등에 각종 언론매체 및 기관에서 관심을 기울이고 있다. 국내에서는 조선닷컴, 한겨레 신문 온라인판 등에서 데이터 시각화 및 인포그래픽을 선도적으로 다루고 있으며, 국외에서는 영국의 가디언, 월스트리트저널, 뉴욕타임즈 등에서 데이터 시각화 분야에서 두각을 나타내고 있다. 우리나라의 경우 데이터시각화 분야는 아직까지 디자인적인 측면에 중점을 두고 있다. 그러나 최근 주목받고 있는 데이터시각화 프로그램 중 Tableau Public 프로그램을 이용하면 좀 더 효율적인 인포그래픽을 제작할 수 있다. 본 연구에서는 Tableau Public을 통해 데이터 분석 및 인포그래픽을 제작하였다. 인포그래픽 제작을 위해 세계은행 데이터가 사용되었으며, 전세계자동차보유율을 분석하여 두 가지 결과가 도출되었다. 첫째, 고소득국가 그룹에서는 1인당 국민총소득이 증가할수록 기울기는 감소추세를 보였고, 중위소득국가에서는 증가추세를 보였다. 둘째, 기준기간에 비해 금융위기 기간에는 1인당 국민총소득 증가율에 따른 전년대비 자동차보유율 증가율이 1.7배 높게 나타났다. 본 논문에서는 Tableau Public을 이용한 데이터 시각화 사례연구를 통해, 기존 디자인 중심의 인포그래픽 제작환경에서, 향후 데이터 중심의 인포그래픽 제작 설계를 활성화시키고자 하였다. 그리고 더 나아가 데이터 중심의 인포그래픽은 과학적 연구, 그리고 정책결정을 위한 도구로써 널리 이용되어야 할 것이다.

키워드 : 인포그래픽, 빅데이터, 데이터 시각화, 타블로 퍼블릭, 정책결정

Abstract Recently, according to the increasingly populated data, many media and organizations focus on big data, data visualization, information visualization and infographics. Domestically, Chosun.com and Hankyoreh online have improved on the data visualization field and internationally, the Guardian, Wall Street Journal, and New York Times are the leading companies on that area. Until now, many people have recognized infographics as a design-oriented product in Korea. However, one of significant data visualization programs, Tableau Public, can visualize data more efficiently. In this paper, Data Visualization Methods Quadrant for Policy Making is defined, and data analysis and producing infographics are executed. As used data, World Bank open source was adopted and using the number of passenger cars per 1,000 people, two analysis results are extracted. First, in high income group, the more GNI per capita, the lesser Slope is represented and in mid income group, the more GNI per capita positively affects to Slope. Second, in the global finance crisis, the car ownership rate was about 1.7 times than the usual state in the global economy. Through the case study, this paper suggests that the direction of producing infographics should be changed from design-oriented to data-oriented. Moreover, the data-oriented infographics should be propagated as means of scientific research and policy making.

Keywords : Infographics, Big Data, Data Visualization, Tableau Public, Policy Making

1. 서론

최근 빅데이터, 데이터 시각화, 정보 시각화, 또는 인포그래픽이라는 단어가 급부상하고 있다. 이러한 트렌트의 중심에는 데이터가 있다. Evans[4]는 브리태니커 백과사전과 위키피디아의 비즈니스모델을 예로 들며, 인터넷의 등장과 지식의 빠른 보급을 통하여 브리태니커 백과사전은 몰락의 길을 걷게 되었고, 위키

피디아는 브리태니커 백과사전을 대체하기에 이르렀다고 설명했다. 또한 2000년대 초반에는 아직까지도 아날로그 데이터(종이, 필름, 오디오 테이프)가 대중적이었으나 2007년 이후 세계의 데이터양이 폭발적으로 증가했으며, 디지털데이터(CD, DVD, 서버, PC storage)의 절반이상이 IP주소를 보유하여 다른 데이터와 서로 정보 전달을 할 수 있게 해 준다고 했다. 또한 인터넷을 통한 데이터의 창작 및 분배 등 거래비

* Donghwan Kim, Assistant Professor, Dept. of Civil Engineering and Environmental Sciences, Korea Military Academy.
donghwan.kma@gmail.com (Corresponding Author)

용의 하락으로 많은 사람들이 정보과정 - 요구(requirements), 수집(collection), 분석(analysis), 생산(production), 전파(distribution) - 에 능동적으로 참여할 수 있게 되었다.

IDC(International Data Corp)가 제공한 자료에 따르면 2010년에 생산된 정보만 하더라도 1,200 엑사바이트(exabytes, 10^{18} bytes)에 이른다[2]. 폭발적으로 증가한 데이터를 유의미한 형태로 전파하기 위해서는 반드시 유통 가능한 방식으로 데이터를 생산해야 한다. 데이터와 데이터를 시각화하는 기술은 더 이상 분리된 과정이 아니며, 연장선상에 있다. 그리고 데이터 시각화 기술에 대한 수요도 꾸준한 증가추세에 있다. Flowingdata.com에 따르면, 정보를 시각화하는 도구로 어떠한 것을 사용하고 있는지에 대한 조사에서 Excel, R, Tableau 등이 각각 31%, 21%, 6%로 1, 2, 3위를 차지했다[13]. 또한, Market 데이터를 시각화하는 도구로써 D3, Tableau 등이 소개된 바 있다[9]. 그 중 Tableau는 여타의 다른 소프트웨어 프로그램과는 달리 스크립팅이 필요하지 않고 단순한 drag-and-drop 기능을 이용하여 데이터를 시각화할 수 있는 장점이 있다. 최근 인터넷만 한겨레 신문에서는 ‘자료로 본 한국, 한국인’이라는 블로그를 통해 데이터 시각화를 구현하고 있는데 2014년부터는 Tableau를 통해 우리나라의 정치, 경제, 생활 등의 주제에 대한 인터랙티브 인포그래픽을 제공하고 있다[6].

Cairo[2]는 데이터 시각화 복잡성이 독자들의 수준에 따라 달라져야 한다고 했으며, Laine-Hernandez and Särkkälä[9]는 Market 데이터를 두 가지 관점 - 과학적인 관점(scientific orientation), 보도중점의 관점(journalistic orientation) - 으로 분류하였다. 전자는 데이터의 내용에 중점을 두고 있는 반면, 후자는 데이터의 처리과정 및 독자에게 초점을 맞추고 있다. 최근 다루어지고 있는 데이터 시각화 및 인포그래픽은 위 두 가지 관점 중 보도중점의 관점으로 주로 다루어지고 있는 반면, 과학적인 관점에서의 접근은 부족했다.

본 논문에서는 현재 우리나라에서 널리 사용되고 있는 디자인 중심의 인포그래픽 제작환경을 개선하고, 데이터 중심의 인포그래픽 제작여건을 조성하기 위하여 Tableau Public을 이용한 인포그래픽 제작사례를 제시한다. 데이터 중심의 인포그래픽 제작 활성화는 공공기관의 효율적인 정책결정을 지원하고, 다양한 분야의 과학적인 연구에도 활용될 수 있을 것이다.

2. 빅데이터 시각화와 정보과정

2.1 빅데이터 시각화

데이터 시각화(data visualization)는 역사적으로 지리적 시각화(geographic visualization)에서 출발하여 주제도(thematic cartography), 통계 그래픽(statistical graphics)을 거쳐 발전되었다[7]. 데이터 시각화는 정보 시각화(information visualization), 정보 그래픽(information graphics), 인포그래픽(Infographics), 정보 일러스트레이션(information illustration) 등으로 사용되고 있으나 현재까지 명확히 정의된 용어는 없으며, 정보과학, 컴퓨터과학, 인간-컴퓨터 상호작용, 통계, 데이터 마이닝, 지도학, 저널리즘 등의 광범위한 학문분야에서 각기 조금씩 다른 의미로 사용되고 있다[9,11]. Cairo[2]는 독자의 역할이 수동적인지 또는 능동적인지의 정도에 따라서 데이터 시각화를 구분했는데, 예를 들면, 인포그래픽의 경우에는 통계차트, 지도, 다이어그램, 스토리텔링 등을 통해 정보를 보여주는 방식인 반면, 정보 시각화는 수용자가 데이터에 대한 탐색과 분석을 할 수 있도록 시각적 도구를 제공하는 것이라고 정의했다[9]. 그러나 인포그래픽도 복잡화 정도와 사용자 탐색 및 분석정도에 따라 정적 인포그래픽, 모션 인포그래픽 그리고 인터랙티브 인포그래픽으로 나눌 수 있으며, 따라서 현재까지는 데이터 시각화, 정보 시각화, 정보 그래픽, 인포그래픽, 정보 일러스트레이션 등을 단일화 된 척도로 구분하기는 어렵다[7]. Figure 1은 Lankow et al.[10]이 분류한 인포그래픽 형식이다. 인포그래픽 4분면에 따르면 인포그래픽은 정

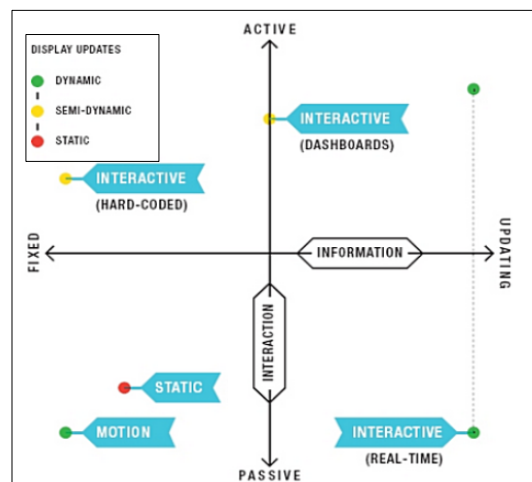


Figure 1. Infographics Formats Quadrant[10]

적, 모션, 그리고 인터랙티브 인포그래픽으로 나누어 질 수 있으며, 특히 인터랙티브 인포그래픽의 경우 업데이트 정도에 따라서 세 가지 형태 - hard-coded, dashboards, real-time - 로 분류된다.

Table 1. Basic Provisions of Communication Methods[10]

Methods	Effects
appeal	· engage a voluntary audience
comprehension	· effectively provide knowledge that enables a clear understanding of the information
retention	· impart memorable knowledge

Table 2. The Order of the Objectives of Visualizations in Different Application Areas[9]

App.	1 st priority	2 nd priority	3 rd priority
scientific	comprehension	retention	appeal
editorial	appeal	comprehension	retention
marketing	appeal	retention	comprehension
market data	comprehension	appeal	retention

Table 3. Approaches to Infographic Design[10]

	Explorative	Narrative
Characteristics	minimalist	illustrative
	only includes elements that represent data	design-focused
	seeks to communicate information	seeks to appeal to viewer with engaging visuals
	in the most clear, concise manner	informs and entertains
Applications	academic research	publications
	science	blogs
	business intelligence	content marketing
	data analysis	sales and marketing materials

Table 4. Properties of Two Stories[11]

Author-driven	Reader-driven
Linear ordering of scenes	No prescribed ordering
Heavy messaging	No messaging
No interactivity	Free interactivity

Choi et al.[3]은 데이터 시각화를 정보 시각화, 과학적 시각화, 인포그래픽 등으로 분류했다. Lankow et al.[10]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 응용분야를 정의하였다.

Lankow et al.[10]은 인포그래픽 설계를 위한 두 가지 관점 - 탐색적(explorative), 서술적(narrative) - 을 제시했다. 탐색적 관점은 데이터에 중점을 둔 과학적인 설계방법이며, 서술적 관점은 마케팅 등에 적합한 디자인 중심의 설계방법이다. Segel and Heer[11]는 서술적 시각화(narrative visualization) 설계방법을 작가중심(author-driven) 시각화와 독자중심(reader-driven) 시각화의 양극단의 속성으로 분류하고 대부분의 시각화 기법은 그 가운데의 스펙트럼 사이에 존재한다고 정의하였다.

본 논문에서는 데이터 중심의 인포그래픽 제작을 위한 데이터시각화 4분면을 Figure 2와 같이 제시하였다. 4분면의 X축은 인포그래픽의 콘텐츠가 데이터 지향적인지 또는 디자인 지향적인지를 나타내며, Y축은 인포그래픽의 형식이 수동적인지 또는 능동적인지를 나타낸다. Figure 2에 제시되어 있는 4가지의 인포그래픽은 Kim[7]의 논문에서 살펴볼 수 있다. 3사분면에 제시되어 있는 인포그래픽은 2014년 대한민국 정부부처 업무보고시 미래창조과학부의 업무보고 인포그

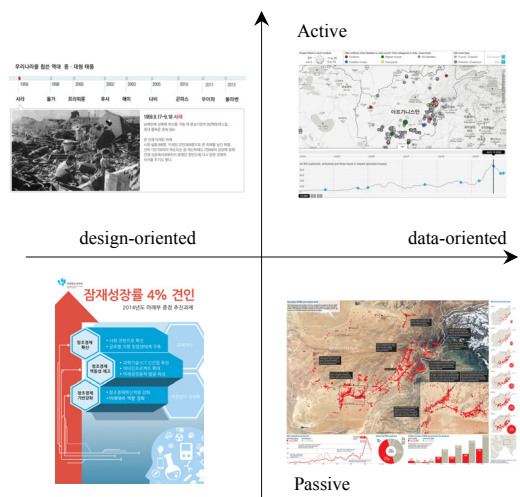


Figure 2. Data Visualization Methods Quadrant for Producing Infographics

래픽이며, 이것은 디자인 지향적이며 정적인 형태의 인포그래픽이라고 할 수 있다. 2사분면은 조선닷컴의 ‘우리나라를 휩쓴 역대 중대형 태풍’이라는 제목의 인포그래픽으로 디자인 지향적이나 일부 인터랙티브 인포그래픽 기능을 제공한다. 4사분면은 영국의 가디언지에서 제작한 아프가니스탄전쟁 관련 급조폭발물 폭발 사건 인포그래픽으로 정적이나 그 콘텐츠는 많은 데이터를 포함하고 있다. 마지막으로 1사분면은 역시 영국 가디언지에서 제작한 급조폭발물 폭발일지를 인포그래픽으로 나타낸 것인데, 모션 인포그래픽 및 인터랙티브 인포그래픽 기능을 제공하면서 또한 콘텐츠적인 측면에서 많은 데이터를 포함하고 있다. 인포그래픽 제작을 위해 다양한 방법을 사용할 수 있지만, 빅데이터를 적절히 시각화할 수 있는 방법은 바로 1사분면에서 제시되고 있는 데이터 지향적이며 능동적인 형식의 인포그래픽이다.

우리나라에서 활용되고 있는 대부분의 인포그래픽 제작 설계는 2, 3사분면에 위치하고 있는 서술식, 보도적 관점의 인포그래픽이다. 그러나 데이터 중심의 인포그래픽은 1, 4사분면에 제시된 인포그래픽과 같은 탐색적이고, 과학적인 인포그래픽이며, 이는 다양한 의사결정 시스템에 사용되거나, 과학적인 연구에도 사용될 수 있을 것이다.

2.2 정보과정

미 야전교범 ‘정보’에 따르면 ‘정보란 외국, 적대세력 또는 잠재적 적대세력 및 조직, 작전지역 및 잠재적 작전지역과 관련된 가용한 첩보의 수집, 처리, 통합, 평가, 분석 그리고 해석 등으로 얻어진 생산물’이라고 정의하고 있으며, 정보과정은 ‘요구(requirements), 수집(collection), 분석(analysis), 생산(production), 그리고 전파(distribution)의 5단계로 이루어진다’라고 정의하고 있다[8]. 정보과정의 가장 첫 단계는 바로 요구 단계이다. 정책결정자의 정보 필요 우선순위에 따라 정보수집이 이루어지게 된다. 정보수집은 그 출처에 따라 크게 5가지 - 영상정보, 신호정보, 계측 및 기호정보, 인간정보, 공개출처정보 - 로 이루어진다. 정책결정자에게 필요한 정보가 되기 위해서는 이러한 경로로 수집된 첩보들이 처리되고 가공되어 생산, 전파되어야 할 것이다.

2.3 Tableau Public

Tableau는 미국 시애틀에 본사를 둔 소프트웨어 업체로, 미국 스탠포드 대학교 교수와 연구진이 데이터

를 일반인에게 좀 더 쉽게 이해할 수 있도록 하기 위해서 VizQL (visual query language)을 개발한 것이 시초가 되었다. Tableau는 다른 데이터 시각화 프로그램과 같이 프로그래밍이 필요 없이 간단한 drag-and-drop으로 데이터를 손쉽게 시각화할 수 있는 특징을 가지고 있다. Tableau는 5가지 제품을 제공하는데, 그것은 Tableau Desktop, Tableau Server, Tableau Online, Tableau Reader 그리고 Tableau Public이다. 그 중 Tableau Public은 데이터를 시각화하여 배포할 수 있는 무료 프로그램이다. 국내에서는 Baik and Cho[1]가 ‘인천/나리타 공항의 여객기 출 도착데이터를 이용한 항공노선 분석 연구’에서 Tableau Public을 이용했다. 항공네트워크 분석용 데이터 마트 설계를 위해 원시 데이터를 수집하고, 이를 요구사항에 맞게 처리하여 시각화 분석 및 해석을 실시하였다. 이를 바탕으로 항공관계자들의 피드백을 반영하여 데이터 마트를 개발하였다. 이러한 개발과정에서 중요한 역할을 한 소프트웨어는 바로 Tableau Public이다. 최근 모든 영역에서, 다루는 데이터의 크기가 방대해지고, 복잡한 테이블 형식의 많은 통계치를 직관적으로 분석하고 해석하기 위해서는, 그에 적합한 시각화 도구가 필수적이다. 위 논문에서는 한국과 일본의 항공노선 분석에서 운항편수, 운항비중, 항공거리, 지연시간 등 다양한 속성정보들을 Tableau Public을 통해 지리정보화 시켰다. 또한 Tableau Public은 다양한 운항정보를 다차원 분석환경을 통해 비교 분석하는데 사용되었다. 그러나 우리나라에서는 아직까지 Tableau Public에 대한 활용이 학계 전반에 걸쳐 활발하지 않은 반면, 해외의 경우, 경영, 마케팅, 데이터시각화 등 다양한 분야에 Tableau Public이 활용되고 있다. 본 논문은 간단한 데이터 분석과 데이터시각화를 통해 Tableau Public이 공공정책결정과 과학적 연구의 도구로 활용될 수 있는 기초적인 정보를 제공할 것이다.

3. Tableau Public을 이용한 연구사례

3.1 개요

최근 도시 이동수단(urban mobility)에 대한 관심이 매우 높아졌다. 유럽의 경우, 인구의 60% 이상이 도시 지역에 거주한다. 한편, 초대형 도시에는 환경오염과 교통체증 큰 문제가 되고 있다. 몇몇 도시에서는 사람들이 삶의 질을 향상시키기 위해서 지하철 등 공공교통수단을 확장하고 있다. 예를 들면, 핀란드의 수도인 헬싱키에서는 승용차로부터 유발되는 많은 문제들

을 해결하기 위해서 2025년까지 공공 교통시스템을 향상시키기로 했다[5]. 이러한 사업의 이상적인 목적은 누구도 차를 가지지 않고 불편함이 없이 도시에서 교통수단을 해결할 수 있는 시스템을 만드는 것이다. 그러나 세계은행에 따르면, 2011년 기준 핀란드의 자가용 보유율은 인구 1,000명당 약 475명으로 매우 높은 수준이다.

최근 한 연구 결과에 의하면, 인구통계학적 변화가 자동차 보유율에 영향을 미치는 한 요소라고 밝히고 있다. 미국을 제외한 몇몇 선진국에서는 인구고령화가 자동차를 이용한 여행에는 부정적인 영향을 끼치는 반면, 자동차 보유율에는 긍정적인 영향을 끼친다고 했다[12]. 그러나 몇몇 선진국, 노르웨이, 영국, 미국의 경우 젊은 연령층에서는 자동차 보유율이 감소하고 있으며, 중년층에서는 독일과 미국에서만 자동차 보유율의 감소 추세를 보이고 있다[12]. 위 연구결과에서 보이는 공통적인 현상은 국가의 경제적인 수준과 자동차 보유 및 사용경향의 상관관계가 점점 약화되는 추세에 있다는 것이다.

많은 나라들은 환경 친화적인 정책 시행을 계획하고 실행하려고 노력하고 있으며, 그 중 가장 대표적인 것이 교통시스템이다. 본 논문에서는 경제수준별 자동차 보유율의 변화추세 및 경제상황에 따른 자동차 보유율의 차이를 연구하여 정책결정자들을 위한 참고 자료를 제시할 것이며, 이러한 정책계획과 실행을 결정하기 위한 도구로서 Tableau Public을 이용한 데이터 분석 및 인포그래픽을 제작하여 정책결정자들이 빅데이터에 대한 이해를 쉽게 할 수 있도록 도울 것이다.

3.2 연구방법 및 데이터

자동차 보유율은 인구 1,000명당 보유한 승용차의 대수를 말한다. 자동차 보유율은 세계은행(World Bank)이 인터넷에 공개 데이터로 제공한 것을 사용했다. 그 자료에는 200여개 국가의 2000년부터 2011년까지의 자동차 보유율을 포함하고 있다. 그 중 최빈국을 제외한 후, 12개년도의 자료 중 10개년도 이상의 데이터를 포함하고 있는 79개의 국가만을 대상 데이터에 포함시켰다.

첫째, 79개 국가의 12개년 자동차 보유율을 선형 회귀분석 방법을 이용하여 추세선의 기울기를 구하였다. 한국의 경우, 기울기는 9.5876이 구해졌으며, 이것은 매년 인구 1,000명당 자동차 보유대수가 약 9.5876대가 증가했다는 것을 의미한다. Figure 3는 Tableau Public을 이용하여 2000년부터 2011년까지 자동차 보

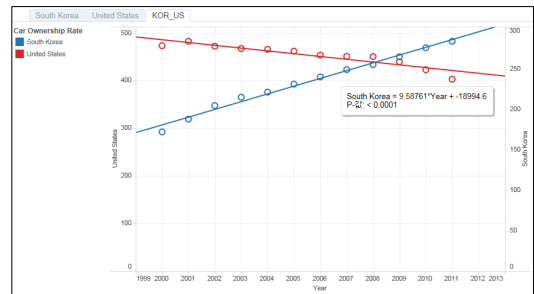


Figure 3. Car Ownership Rate (South Korea and the U.S.) Produced by Tableau Public

유율 및 선형 회귀분석 결과 추세선을 나타낸 것으로 각각의 탭은 한국, 미국, 그리고 한국 및 미국의 자동차 보유율을 나타낸다.

둘째, 79개국의 기율기를 1인당 국민총소득(GNI per capita)을 기준으로 4개 그룹으로 나누고, 그룹별로 분석하였다. 4개 그룹은 최빈국을 제외하고 OECD국가, OECD가 아닌 고소득 국가, 중상위소득국가, 중하위소득국가로 나누었다. 이러한 국가별 소득분류자료는 세계은행 데이터에서 제공하고 있으며, 일반적으로 중위소득국가의 1인당 국민총소득은 약 1,000~5,000 달러이다.

셋째, 경제상황과 자동차보유율과의 관계를 연구하기 위하여, 특정기간 자동차보유율의 변화율을 분석했다. 이를 위해 두 기간을 선정했는데 기준기간은 2000-2001년이고, 다른 기간은 전 세계적으로 금융위기가 있었던 2007-2008년이다. 두 기간의 자동차 보유율 누락데이터로 인해 79개국 중 65개국만 표본국가로 선정하였다.

3.3 연구결과

본 절에서는 Tableau Public을 이용하여 인포그래픽을 제작하고, 기율기 분석, 특정기간 자동차보유율의 변화율을 분석한 후, 이러한 인포그래픽들이 데이터 시각화 4분면에서 어디에 위치하는지 확인하고, 그 유용성을 연구할 것이다.

3.3.1 Tableau Public을 이용한 기율기 분석

총 79개 국가의 경제규모별 자동차보유율을 분석한 결과는 Figure 4와 같다. X축은 2005년 기준 1인당 국민총소득이며, Y축은 2000년부터 2011년까지 12개년의 자동차보유율을 선형 회귀분석하여 구한 기율기이다. 전체적으로 1인당 국민총소득과 기율기는 음의 관계를 나타낸다. 한편 79개의 국가를 고소득국가와

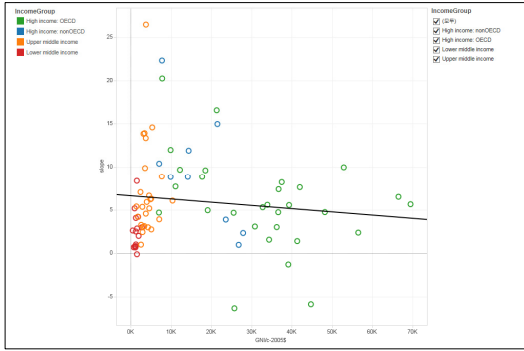


Figure 4. Slope to GNI per capita: 79 Countries

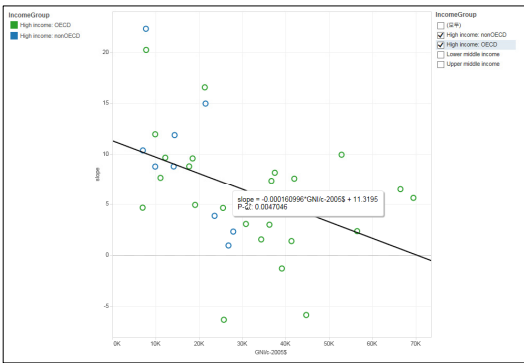


Figure 5. Slope to GNI per capita: High Income Countries

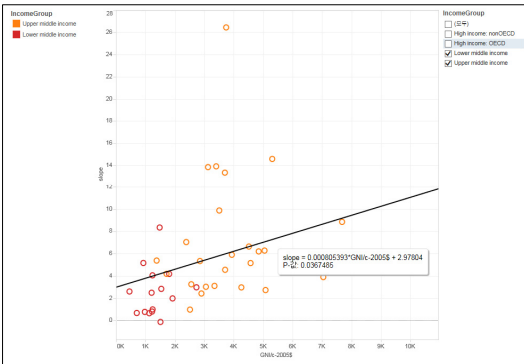


Figure 6. Slope to GNI per capita: Middle Income Countries

Table 5. Statistics from Figure 5 and 6

Group	X-coefficient	P-value
High Income Countries	-0.0001	0.0047
Middle Income Countries	0.0008	0.0367

중위소득국가로 구분하여 나타내면, 각각 Figure 5와 6로 나타낼 수 있으며, 각각의 기울기는 서로 다른 부호를 나타낸다. 즉 고소득국가에서는 소득이 높을수록 기울기의 크기가 작아지지만, 중위소득국가에서는 소득이 높을수록 기울기의 크기도 커진다. 이것은 고소득 국가에서는 일반적인 예측과는 반대로 소득이 높은 국가일수록 자동차보유율의 증가율이 작아짐을 의미한다. Table 5는 Figure 5와 6의 X축 계수 및 P값을 나타낸 것이다. P값이 0.05 이하이므로 통계적으로 유의하다고 할 수 있다.

3.3.2 Tableau Public을 이용한 특정기간 자동차 보유율 변화율 분석

두 기간 동안 총 65개 국가가 데이터를 이용하여, 1인당 국민총소득과 자동차보유율의 변화율, 두 변수의 상관관계를 분석하였다. Figure 7에서 X축은 전년 대비 2001년의 1인당 국민총소득 증가율이며, Y축은 2000년에서 2001년 사이의 자동차 보유율의 변화율을 나타낸다. 즉 Figure 7은 각국 기준기간(2000-2001)의 1인당 국민총소득 증가율과 자동차보유율의 변화율과의 상관관계를 나타내며 Figure 8은 금융위기시

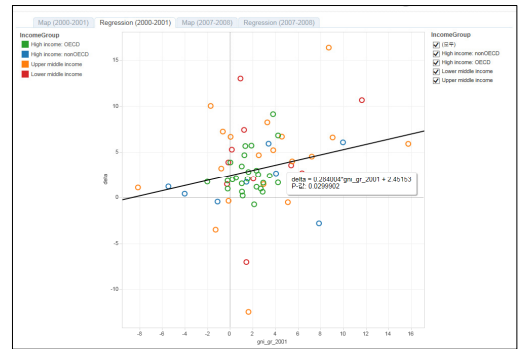


Figure 7. Changes in Car Ownership Rate: 2000-2001

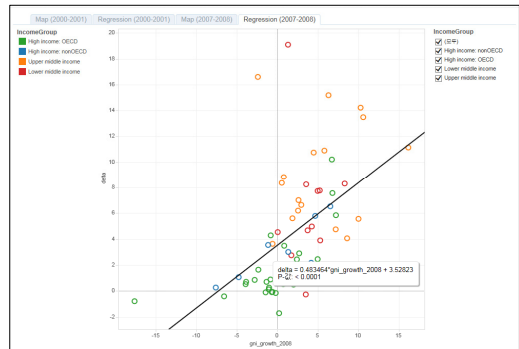


Figure 8. Changes in Car Ownership Rate: 2007-2008

(2007-2008) 1인당 국민총소득 증가율과 자동차보유율의 변화율과의 상관관계를 나타낸다. Figure 7과 8에서 X축 계수는 각각 0.2840과 0.4834이다. 2008년 이 금융위기 시기임을 고려할 때, 2007-2008년에 자동차보유율의 변화율이 기준기간보다 약 1.7배 크다는 것을 알 수 있다. 이것은 세계 각국의 자동차 보유율 또는 자동차 구매율이 금융위기에 영향을 크게 받는다고 예측 할 수 있다.

3.3.3 Tableau Public을 이용한 데이터 시각화

3장에서는 Tableau Public을 이용한 연구사례를 들었다. Tableau Public은 Figure 3~8에서 보는 바와 같이 scatter plot 및 선형회귀분석 등 간단한 통계분석 기능뿐만 아니라 데이터가 포함하고 있는 지리정보를 자동적으로 지도에 표시하고, 인터랙티브 인포그래픽을 생성해준다. Table 7는 세계은행 데이터로부터 편집한 데이터 구조로써 A열의 Country Name과 D열의

Table 6. Statistics from Figure 7 and 8

Period	X-coefficient	P-value
2000-2001	0.2840	0.0299
2007-2008	0.4834	P-value < 0.0001

Table 7. Contents of EXCEL Data

Column	Contents
A	Country Name
B	GNI per capita
C	Slope
D	Region
E	Income Group

Region은 각각 위치정보를 포함하고 있다. Figure 4와 Figure 9은 같은 EXCEL 데이터에서 생성된 인포그래픽이다. Figure 4의 경우에도 Income Group을 필터링해서 그룹별 추세선 분석을 통해 결과값을 얻을 수 있고, Figure 9과 같은 경우에도 총 7개의 지역을 필터링해서 데이터를 손쉽게 알아볼 수 있다. Figure 9을 확인해보면, 기울기 값이 음수인 국가(적색계열: 미국, 영국, 뉴질랜드)를 시각적으로 매우 쉽게 발견할 수 있다.

Tableau Public을 이용한 데이터 시각화는 몇 가지 장점을 가진다. 첫째, 기존 상용프로그램에 비해 인터랙티브 인포그래픽을 제작할 수 있다. EXCEL과 같은 프로그램은 데이터를 고정된 그래픽으로 생산하기에 유용했지만 Tableau Public의 경우에는 데이터를 그룹화하고 사용자가 원하는대로 간단한 클릭으로 그룹별 결과값을 제공한다. 또한 제작된 인포그래픽은 웹으로 게시하거나, 이메일로 공유할 수 있는 등 자유로이 배포할 수 있다.

둘째, 기존 디자인 중심의 인포그래픽 제작에서 데이터 중심의 인포그래픽 제작이 가능하다. 즉 Tableau Public은 과학적 연구방법 등에 활용이 가능하다. 우리나라에 대중적으로 소개되어 있는 인포그래픽은 디자인 중심의 인포그래픽이 대부분이며, 데이터 중심의 인포그래픽 제작을 하더라도 보도 중심의 인포그래픽이 주를 이루었다. 그러나 Tableau Public을 이용하면 방대한 데이터를 손쉽게 시각화할 수 있으며, 특히 지리적 공간과 관련된 데이터일 경우 OpenStreetMap을 통해 인포그래픽 제작이 가능하여, 정책결정자들에게 쉽게 정보를 제공할 수 있다. Tableau Public에서는 “country”, “state”와 같은 지리공간과 관련된 필드 값을 자동적으로 인식하여 지도로 나타내준다.

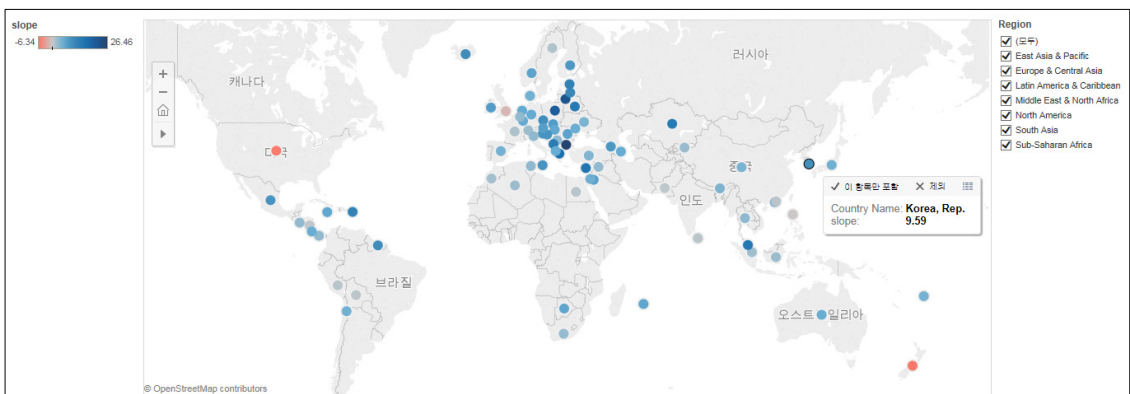


Figure 9. Slope to GNI per capita: Geographical Visualization of 79 Countries

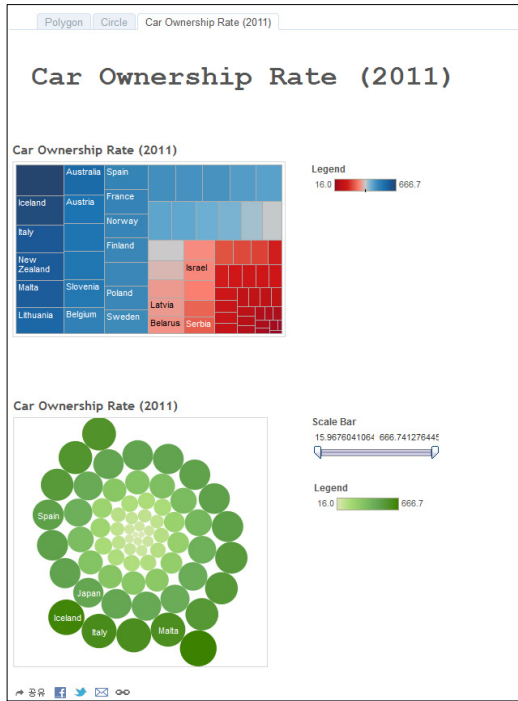


Figure 10. Car Ownership Rate (2011)

셋째, 프로그래밍이 필요없는 무료 프로그램이다. 대표적인 통계 및 시각화 프로그램인 R 등은 프로그래밍이 필요한데 비해, Tableau Public은 drag-and-drop 만으로 데이터를 사용자가 원하는 대로 시각화 할 수 있다.

넷째, 다양한 종류의 인포그래픽 제작이 가능하다. Figure 10은 2011년의 전 세계 자동차 보유율을 인포그래픽으로 제작한 것이다. 각각 Polygon과 Bubble 형태로 제작하였는데, 면적에 따라 각국의 자동차 보유율을 쉽게 이해할 수 있으며, 특히 아래 Bubble 형태의 경우, 오른쪽에 scale bar를 넣어 사용자가 자동차 보유율의 크기에 따라 필터링하여 나타낼 수 있는 인터랙티브 인포그래픽으로 제작하였다. 또한 Figure 10은 dashboard 형태로 제작하여 웹 게시 후 자유롭게 공유가 가능하다.

3장에서 살펴본 인포그래픽을 데이터 시각화 4분면 중 어디에 위치하는지 살펴보면 주로 X축의 위 방향 또는 Y축의 오른쪽 방향, 즉 active 및 data-oriented 성향을 띤다.

4. 결 론

최근 디지털 데이터가 폭발적으로 증가함에 따라

빅데이터, 데이터 시각화, 정보 시각화, 인포그래픽이라는 단어가 급부상하였다. 이에 따라 각종 언론 매체 및 정부기관 등에서 데이터 시각화 및 인포그래픽 사용빈도가 증가하고 있다. 국내에서는 조선닷컴, 한겨레 신문 온라인판 등에서 다수의 인포그래픽을 선보이고 있으며, 특히 한겨레 신문은 2014년부터 Tableau 프로그램을 이용하여 데이터 시각화를 보여주고 있다. 외국에서는 영국의 가디언, 월스트리트 저널, 뉴욕타임즈 등이 인포그래픽을 선도적으로 생산하고 있다. 그러나 이러한 인포그래픽은 주로 보도적인 관점에서 생산된 인포그래픽이며, 수많은 데이터를 시각적으로 재생산해놓은 인터랙티브 인포그래픽이라고 할 수 있다.

본 논문에서는 인포그래픽에 대한 논문들에 대한 선행연구를 바탕으로, 정책결정을 위한 데이터 시각화 및 인포그래픽 설계를 위해 데이터 시각화 4분면을 Figure 2와 같이 제시하였다. 데이터 시각화 4분면의 한 축은 콘텐츠의 지향방향이며, 다른 한 축은 데이터 시각화 포맷 형태이다. 콘텐츠 지향방향은 design-oriented와 data-oriented를 양 극단으로 하며, 데이터 시각화 포맷 형태는 passive와 active를 양 극단으로 하여, 모든 인포그래픽은 그 사이에 위치할 수 있다.

특히 3장에서는 Tableau Public을 이용한 연구를 실시하였다. 세계은행에서 제공하는 오픈데이터를 바탕으로 전세계 자동차보유율을 조사하여 자동차보유율 증가추세를 각 국가별 소득수준, 1인당 국민 총소득 증가율, 특정 기간별 그리고 지역별로 분석하였다. 분석결과 12년 동안 각 국가의 자동차보유율 기울기는 고소득국가 그룹에서는 1인당 국민총소득이 증가할수록 감소추세를 보였으나, 중위소득국가에서는 증가추세를 보였다. 또한 특정기간 중, 1인당 국민총소득 증가율에 따른 전년대비 자동차보유율 증가율은 기준기간(2000-2001) 대비 금융위기기간(2007-2008)에 1.7 배 높게 나타났다. 이러한 데이터 분석은 각국 정책결정자들이 경제·환경·교통 정책 계획을 수립하고 시행하는 데 유용하게 이용될 수 있다.

최근 우리나라에서 인포그래픽이 널리 이용되고 있지만, 데이터 시각화 4분면에서 분석해보면, 주로 디자인 중심의 인포그래픽이 주류를 이루었다. 그러나 데이터가 범람하는 시대에 많은 데이터를 처리하고 정책결정자가 그것을 이해하기 위해서는 직관적이고 또한 분석적인 인포그래픽 사용이 필요하다. 이를 위해서는 미적으로 훌륭한 인포그래픽보다는 데이터 중심의 인포그래픽이 필요하며, 본 논문에서는 Tableau Public을 이용한 데이터 시각화에 대한 기초적인 사례

연구를 제시하였다. 향후 좀 더 복잡하고 방대한 양의 데이터를 기초로, Tableau Public을 사용한 데이터 시각화 연구가 필요하다. 뿐만 아니라 이러한 인포그래픽이 정책결정자의 분석과 이해에 미치는 영향에 대한 연구도 발전되어야 한다.

인포그래픽 사용에는 보도, 홍보 뿐만 아니라, 과학적 연구, 그리고 정책결정을 위한 도구로써 인포그래픽이 널리 이용되어야 할 것이다.

References

- [1] Baik, E. Y; Cho, J. H. 2013, Analysis of Airline Network Using Incheon and Narita Passenger Flight Origin-Destination Data, *Journal of Information Technology Applications and Management*, 20(1):87-106.
- [2] Cairo, A. 2012, *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley: New Riders.
- [3] Choi, K. S; Ham, K. Y; Kim, S. H. 2013, Big Data Visualization, *Korea Society of Computer Information*, 21(1):33-43.
- [4] Evans, P. 2013, How Data will Transform Business, TED.com, Accessed December 15. Available: http://www.ted.com/talks/philip_evans_how_data_will_transform_business.
- [5] Greenfield, A. 2014, Helsinki's Ambitious Plan to Make Car Ownership Pointless in 10 Years, *The Guardian*, Accessed December 15. Available: <http://www.theguardian.com/cities/2014/jul/10/helsinki-shared-public-transport-plan-car-ownership-pointless>.
- [6] Hani.co.kr. 2014, Data Visualization, Hankyoreh [online], Accessed December 31. Available: <http://www.hani.co.kr/arti/ISSUE/127/>.
- [7] Kim, D. H. 2014, A Case Study of Infographics for National Defense – Focusing on the Data-journalism of Afghanistan War in Guardian, *Journal of Korea Spatial Information Society*, 22(5):43-52.
- [8] Kim, D. H. 2014, A Study on the Defense Geospatial Intelligence Governance – Focusing on the Intelligence Community and LandWarNet, *Journal of Korea Spatial Information Society*, 22(1):19-26.
- [9] Laine-Hernandez, M; Särkkä, N, 2013, Market Data Visualization–Concepts, Techniques and Tools. Accessed December 15. Available: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2013/D1.2.3.3_eReading_%20Market%20Data%20Visualization-Concepts,%20Techniques%20and%20Tools.pdf.
- [10] Lankow, J; Ritchie, J; Crooks, R. 2012, *Infographics: The Power of Visual Storytelling*. New Jersey: John Wiley & Sons.
- [11] Segel, E; Heer, J. 2010, Narrative visualization: Telling stories with data. *Visualization and Computer Graphics*, IEEE Transactions on, 16(6), 1139-1148.
- [12] Van Dender, K; Clever, M. 2013, Recent Trends in Car Usage in Advanced Economies – Slower Growth Ahead?: Summary and Conclusions, *International Transport Forum Discussion Papers*, No. 2013/09, Paris: OECD Publishing.
- [13] Yau, N. 2010, What do you use to analyze and/or visualize data?, flowingdata.com, Accessed December 15, Available: <http://flowingdata.com/2010/09/28/poll-results-what-do-you-use-to-analyze-and-or-visualize-data/>.

Received : 2015.01.06

Revised : 2015.03.05

Accepted : 2015.04.16