

빅 데이터 가시화 기술을 적용한 공공데이터 콘텐츠 구현 - Map가시화 기법

박선희^{1*} · 김정호^{2**} · 유현배^{3***}

¹아이에이치테크 / (주)유토비즈

²한밭대학교 컴퓨터공학과

³나사렛대학교 교양교육학부

Implementation of public data contents using Big data Visualization technology - Map visualization technique

Seon-Hui Bak^{1*} · Jong Ho Kim^{2**} · Hyun-Bea You^{3***}

¹Ihtech / Utobiz

²Computer Engineering, Hanbat National University, Daejeon 126-17, Korea

³Division of Liberal arts education, Korea Nazarene University, Cheonan 31172, Korea

[요 약]

4차 산업혁명의 가속화로 인하여 우리 주변의 데이터가 급속도로 증가하였다. 이에, 데이터를 수집하는 것 이상으로 데이터 분석을 통해 얻어진 데이터의 성질과 의미를 보다 손쉽게 파악하고 데이터의 가치판단에 유연하게 적용할 수 있어야 한다. 가시화 표현기술은 현재 많은 분야에서 관심을 받고 있다. 가시화는 데이터 분석 결과를 보다 쉽게 이해할 수 있도록 그래프, 차트 등으로 사용자가 보다 쉽게 데이터의 정보를 파악할 수 있어, 즉각적인 판단이 가능하게 하여 빠른 의사결정을 할 수 있도록 한다. 그중에서도 사용자들의 활용 가치가 높은 공공데이터를 활용한 가시화에 대한 관심이 높다. 이에 본 논문에서는 가시화를 표현할 수 있는 다양한 소프트웨어들 중에서 R 라이브러리와 R Studio를 활용하여 전국 자전거 보관소 설치 장소의 공공데이터를 가시화 하는 콘텐츠를 구현하였다.

[Abstract]

Due to the acceleration of the 4th industrialization, the data around us rapidly increased. Therefore, it is necessary to be able to more easily grasp the nature and meaning of data obtained through data analysis than to collect data, and apply it flexibly to the value judgment of data. Visualization technology is now attracting attention in many fields. Visualization allows the user to more easily grasp the information of the data with graphs, charts, etc. so that the data analysis result can be understood more easily, so that the user can make an immediate judgment and make a quick decision. Among them, there is a high degree of interest in visualization using public data, which is highly useful to users. In this paper, we implemented R - library and R Studio to visualize public data at the installation sites of bicycle storage sites among various software that can express visualization.

색인어 : R, 공공 데이터, 빅 데이터, 가시화, ggmap

Key word : R, Public Data, Big Data, Visualization, ggmap

<http://dx.doi.org/10.9728/dcs.2017.18.7.1427>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 30 October 2017; Revised 09 November 2017

Accepted 25 November 2017

*Corresponding Author; Hyun-Bea You

Tel: +82-10-9754-3200

E-mail: hbyoo@kornu.ac.kr

I. 서론

IT 정보화 기술이 발전함에 따라 우리는 주변의 넘쳐나는 정보들로 인해 골머리를 앓고 있다. 정보는 매 시각 급격하게 증가하는데 비해 정보 처리, 분석 기술은 그 속도를 따라가지 못하고 있는 실정이다. 특히 파일, 텍스트, 이미지, 음성 등 비정형 데이터들은 더욱더 처리가 곤란하다. 때문에 분야와 상관없이 조직들은 보유하고 있는 많은 양의 정보 데이터들을 효율적으로 가공하여 정보기술(IT)을 보다 창의적으로 사용하기 위해 빅 데이터(Big data) 개발에 심혈을 기울이고 있다[1].

IT 기술에서 빅 데이터란 기존 데이터베이스 관리능력을 넘어서는 수십 테라바이트의 정형·비정형의 데이터를 분석하는 기술이다[2], [3]. 빅 데이터는 다양한 형태(Variety), 빠른 생성 속도(Velocity), 데이터 양(Volume)의 성향을 지니 3V라고 불리며[4] 여기에 시간이 지나면서 매우 빠르게 변화하고 전파되는 빅 데이터의 특성상, 데이터 전체를 분석하고 패턴을 발견하기 어려워짐에 따라 가치(Value)를 더한 네번째 특징으로 4V라고 불리기도 한다[5], [6].

빅 데이터는 과거에도 매우 유용하게 활용되어 왔는데 대표적인 예로 미국의 제 44대 대통령 버락 오바마는 선거 활동 시 유권자의 기본 인적사항과 과거 투표 여부, 사용하는 핸드폰 기종, 자주 방문하는 장소까지 유권자의 성향 정보를 모아 유권자 데이터베이스를 확보하고 활용하여 경쟁자인 롬니를 제치고 미국의 제 44대 대통령이 될 수 있었다. 이렇게 빅 데이터는 디지털 환경에서 생성되는 거대한 양의 데이터로 다양하고 정확한 분석을 통해 미래를 예측하고 가치를 창출한다.

하지만 많은 양의 빅 데이터는 일일이 검증하기가 불가능하며, 검증되지 않은 저질 정보의 범람으로 잘못된 정보를 가공, 활용할 경우 돌이킬 수 없는 문제가 발생하기도 한다. 때문에 빅 데이터 중 공공 데이터(Public Data)의 가치는 매우 높다. 공공 데이터란 빅 데이터 중 정부나 공공 기관의 업무 과정에서 발생한 데이터로서 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 하는 데이터이다[7]. 검증된 데이터라는 점에서 일반적인 데이터보다 상대적으로 높은 가치를 지니며[8] 누구나 공공데이터를 제공 받고 사용할 수 있다.

이러한 빅 데이터에도 한 가지 문제점이 있다. 빅 데이터는 데이터를 수집하는 것 이상으로 데이터 분석을 통해 얻어진 데이터의 성질과 의미를 보다 손쉽게 파악하고 데이터의 가치 판단에 있어 보다 유연하게 적용할 수 있어야 한다. 그러

나 SNS의 발달과 동시에 파일, 텍스트, 이미지, 음성과 같은 비정형 데이터가 계속해서 증가하고 있는 추세이다. 이러한 비정형 데이터는 사용자가 데이터의 특성 파악과 의미 전달을 하는데 있어 어려움을 겪게 된다.

이 문제는 가시화(Visualization)를 통해 해결할 수 있다. 가시화는 데이터 분석 결과를 보다 쉽게 이해할 수 있도록 그래프, 차트 등 가시화 기법을 활용하여 데이터를 가시화함으로써 사용자가 보다 쉽게 데이터의 정보를 파악할 수 있으며, 즉각적인 상황 판단이 가능하게 하여 의사결정을 빠르게 한다[9], [10].

본 논문에서는 빅 데이터의 시장 동향과 공공데이터의 활용법, 또 이를 가시화하는 기법과 솔루션에 대해 알아보고 마지막으로 직접 공공데이터를 활용한 가시화 콘텐츠를 구현하여 가시화 된 데이터의 효용성을 파악하고자 한다. 데이터는 자전거 보관소 관련 공공데이터를 사용하였으며, 개발 가시화 솔루션은 작업환경이 자유로운 R-Studio IDE와 R라이브러리를 사용하였다. 또한 가시화 기법 중 하나인 Map을 활용한 가시화 기법을 적용하여 자전거 보관소 위치를 보다 직관적으로 파악할 수 있도록 하였다.

II. 본론

2-1 관련 연구

1) 빅 데이터 시장 동향

빅 데이터는 미래 IT산업을 선도할 4차 산업혁명의 핵심 기술이다. 디지털 기기 및 센서 등의 보급화를 통해 보다 많은 양의 데이터를 주고받을 수 있는 환경이 확대됨에 따라 데이터 생산은 매 시각 증가 하고 있다. IDC(International Data Corporation)에서 발표한 '디지털 유니버스 연구 보고서'에 따르면 전 세계 디지털 데이터 양은 2020년에는 약 40제타 바이트가 될 것이라고 밝혔으며, 이는 데이터가 약 2년마다 2 배씩 증가하고 있음을 의미한다. 즉 데이터의 증가는 더욱더 가속화 될 것이며 이에 맞추어 빅 데이터의 중요성은 더욱 커질 것임을 알 수 있다.

IDC의 자료에 의하면 전 세계 빅 데이터 인프라 시장은 2019년 까지 486억 달러 규모(연평균 성장률 23.1%)에 이를 것으로 전망하였으며, 미국의 지디넷은 빅 데이터 시장이 2019년 까지 1,879억 달러 규모로 연평균 50% 성장률을 보일 것으로 전망했다.

국내시장은 한국과학기술정보연구원(KISTI; Korea Institut

e of Science and Technology Information)의 ‘2015년~2020년 국내 빅 데이터 시장 전망’에 따르면, 국내 빅 데이터 시장 규모는 2015년 2억 6300만 달러에서 2020년 8억 9380달러로 증가할 것으로 전망하였다.

이처럼 데이터의 증가에 따라 빅 데이터 시장 또한 함께 성장하고 있다. 이러한 흐름에 발맞추어 본 논문에서는 빅 데이터 중 공공데이터를 활용하여 직접 가시화 콘텐츠를 구현하였다.

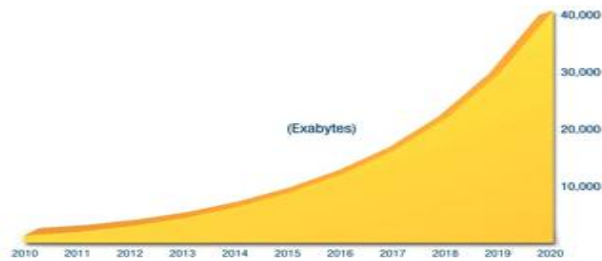


그림 1. 디지털 데이터의 증가 추세, IDC
 Fig. 1. Increasing trend of digital data, IDC

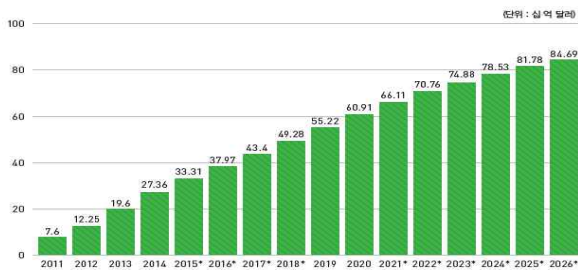


그림 2. 세계 빅 데이터 시장 동향 및 전망, Statista, Wikibon
 Fig. 2. World Big Data Market Trends and Forecasts, Statista, Wikibon

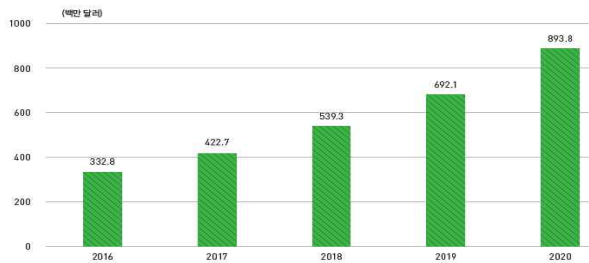


그림 3. 국내 빅데이터 시장 전망, 한국과학기술정보연구원
 Fig. 3. Big Data Market Forecast, Korea Institute of Science and Technology Information

2) 공공데이터의 활용

공공기관의 업무과정에서 얻어지는 공공데이터는 검증된 데이터이다. 검증된 데이터라는 점에서 공공데이터의 가치에 대한 관심이 급증하고 있으며, 특히 공공데이터의 전략적 활용이 국가경쟁력을 향상 시킬 수 있다는 가능성과 기대에 따라 공공 데이터의 활용을 극대화 하고자 하는 노력이 적극적으로 이루어지고 있다.

공공데이터 활용의 좋은 예로 우루과이의 공공의료기관의 ‘ATSERVICIO.UY’가 있다. 우루과이의 공공의료기관은 접근성이 굉장히 낮아 대부분 광고나 마케팅에 의존했는데, 공공기관의 평균 진료 대기시간, 진료비용, 환자 만족도 등의 의료 정보들이 공개하였고 150만 명의 환자들이 자신에게 적합한 공공의료기관을 선택, 공공기관의 의료 정보 다운로드 수가 68배 증가하는 등 국민들의 의료생활의 큰 변화를 이끌어 내었다[11].

또한 미국에서 전기명세서에 있는 숫자는 사람들에게 아무 의미 없는 데이터라고 받아들여지고 있었는데, 미국의 에너지 테크 기업인 Opower는 가정의 에너지 사용량과 함께 다른 가구들의 에너지 사용량을 비교할 플랫폼을 제시하여 에너지 절약을 권고 하였고, 그 결과 Opower는 시간당 3테라 가량의 에너지와 4천 600만 파운드의 이산화탄소를 감소시킬 수 있었다[12].

이처럼 공공데이터는 활용에 따라 사회적, 경제적 가치를 가지며 본 논문에서는 자전거 보관소 관련 공공데이터를 활용하여 사회적, 경제적 가치를 높이고 공공데이터의 사용을 더욱더 활성화 시키고자 한다.

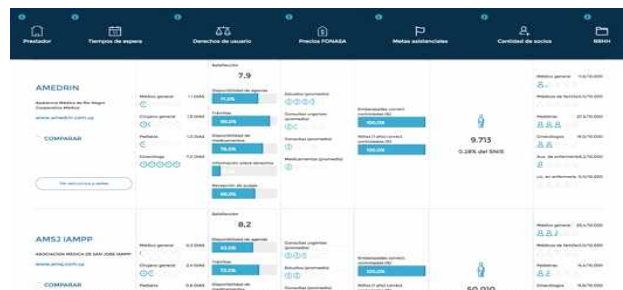


그림 4. 우루과이 보건당국의 의료서비스 비교 플랫폼 , ATSERVICIO.UY
 Fig. 4. Uruguay Health Authority Healthcare Comparison Platform , ATSERVICIO.UY

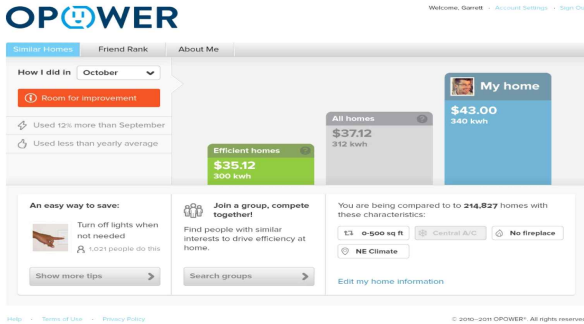


그림 5. 기타 가정용 전기 사용량 표시 및 절전 권장 결과, Opower
 Fig. 5. Other household electricity usage indication and power saving recommended results, Opower

3) 가시화 기법

표 1. 주요 가시화 기법

Table 1. Major visualization techniques

Classification	Major visualization techniques
Time visualization	Bar graph, Stacked bar graph, Point graph, Time-series graph
Distribution visualization	Pie chart, Donut chart, Tree map, Cumulative continuous graph
Relationship visualization	Scatter plot, Bubble chart, Histogram
Comparative visualization	Heat map, Star chart, Parallel coordinate system, Multidimensional scaling
Space visualization	Map

데이터를 가시화 하는데 있어 그 용도에 맞게 활용하기 위해서는 어떠한 가시화 기법을 사용했는지가 중요하다. 가시화 기법은 그 쓰임새에 따라 다양한 차트, 그래프 등이 존재하며 이를 5가지 가시화 기법으로 분류하여 기술한다[13].

시간 가시화는 시계열 데이터(Time series data)를 대상으로 한 가시화 기법이다. 시계열 데이터의 가장 큰 특징은 트렌드(Trend) 경향성이며 이산형, 연속형으로 나뉜다. 이산형의 경우 특정 시점의 데이터 값을 막대그래프, 점그래프, 누적 막대그래프 등으로 표현한다. 연속형의 경우 차량 속도 변화하는 데이터를 시계열 그래프, 계단식 그래프 등으로 표현한다.

분포 가시화는 분포 데이터(proportional data)를 대상으로 한 가시화 기법으로 시계열 데이터와 비슷하지만, 분포 데이터는 전체 분포와 시간에 따른 분포로 구분한다. 전체분포는

최대, 최소, 전체 분포를 나타내는데 각 부분이 차지하는 정보를 도넛 차트, 파이 차트, 누적 막대그래프 등으로 가시화한다. 시간에 따른 분포는 시간의 흐름에 따른 변화를 누적 연속 그래프, 누적 영역 그래프, 선 그래프 등으로 가시화한다.

관계 가시화는 변수들 사이에 존재하는 관계를 찾는 가시화 기법으로 상관관계(Correlation), 분포, 비교로 구분할 수 있다. 상관관계는 하나의 변수에 변화가 일어났을 때, 다른 변수에 어떤 영향을 미치는지를 파악하여 한 변수의 값의 변화를 통해 다른 변수의 변화를 예측할 수 있도록 하며 스캐터 플롯, 버블차트 등으로 가시화한다. 분포는 데이터들이 평균, 중앙값 등 어떠한 값을 기준으로 데이터가 분포되어 있는지 표현하는 것으로 스텝 플롯, 히스토그램, 밀도함수 그래프 등으로 가시화 할 수 있다. 비교는 분포를 여러 개 배치하여 서로 비교할 수 있도록 하는 가시화 기법으로 히스토그램으로 가시화 한다.

비교 가시화는 여러 변수의 데이터 값들을 비교하는 방법으로 히트 맵, 스타 차트, 평행 좌표 그래프 등이 사용된다. 또한 군집분석과 같이 각각의 개체를 측정한 뒤 개체들 사이의 유사성과 비유사성을 측정하여 가시화 하는 다차원 척도법이 사용된다.

공간 가시화는 좌표 값을 가진 데이터를 지도상에서 가시화하는 기법으로 x, y 좌표가 아닌 위도와 경도를 사용하여 좌표를 정의한다. 지도상에서 두 지점 간의 연결 관계나 거리, 더 나아가 이동시간까지 측정이 가능하다. 야후, 마이크로소프트 등 글로벌 인터넷 업체들과 네이버나 카카오 등 국내 업체들에서 지도를 제공하고 있으며, 지도상에서 다양한 정보를 가시화 할 수 있도록 지원한다.

본 논문에서는 다양한 가시화 기법 중 자전거 보관소 관련 공공데이터를 가시화하기에 적절한 공간 가시화 중 지도기반의 Map 가시화 기법을 활용하여 데이터를 가시화를 하였다.

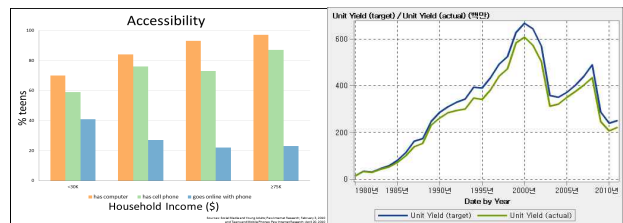


그림 6. 시간 가시화, 막대그래프(좌)-시계열그래프(우)
 Fig. 6. Time visualization, Bar-graph(left)-Time-series graph(right)

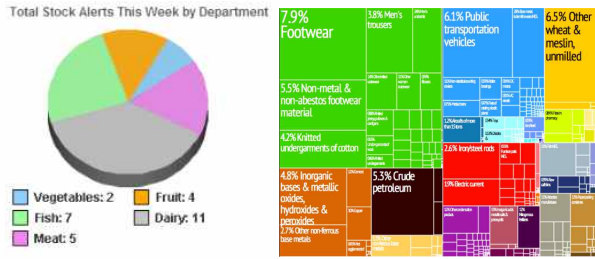


그림 7. 분포 가시화, 파이차트(좌)·트리맵(우)
 Fig. 7. Distribution visualization, Pie chart(left)·Tree map(right)

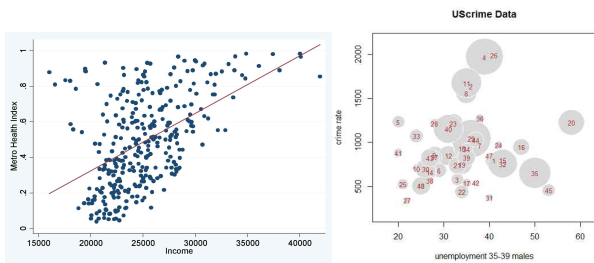


그림 8. 관계 가시화, 스캐터 플롯(좌)·버블차트(우)
 Fig. 8. Relationship visualization, Scatter plot(left)·Bubble chart(right)

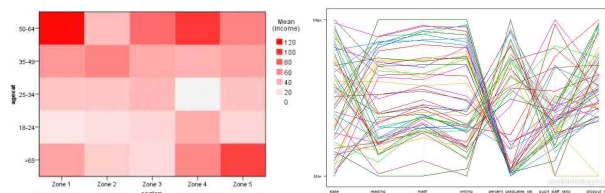


그림 9. 비교 가시화, 히트맵(좌)·평행 좌표계(우)
 Fig. 9. Comparative visualization, Heat map(left)·Parallel coordinate system(right)



그림 10. 공간 가시화, 지도 맵핑
 Fig. 10. Space visualization, map mapping

4) 가시화 솔루션

표 2. 주요 가시화 솔루션

Table 2. Major visualization Solutions

Classification	Contents	Coding
R Studio	R is an integrated development environment (IDE) for programming and has a convenient working environment	O
D3.js	It is a JavaScript visualization library for HTML & SVG that is a good solution for representing data on the web	O
Datawrapper	An online tool for creating interactive charts, allowing users to download original data	X
Tableau	VizQL and data engine enable real-time analysis and results validation	X

데이터 가시화 솔루션은 크게 코딩이 요구되는 솔루션과 코딩이 필요 없는 솔루션으로 나뉜다. 이를 2개씩 나누어 4개의 가시화 솔루션을 기술한다.

(1) 코딩이 요구되는 개발자를 위한 도구

- R Tools Technology, Inc.에서 개발된 R Studio는 리눅스, 맥, 윈도우 등 멀티 플랫폼 지원되는 IDE다. 에디터, 콘솔, 명령어 히스토리, 가시화, 파일탐색, 패키지 관리 등을 한 화면에서 보여준다. 클라이언트-서버 환경으로 설정해두면 서버 리소스를 원격에서 활용할 수 있으며 특히 웹 브라우저 자체적으로 연결이 끊겨도 작업 내용과 환경을 그대로 유지하고 있어 인터넷만 된다면 자신이 작업하던 내용, 환경을 그대로 불러올 수 있다. R Studio에서 사용하는 R은 데이터 분석을 위한 언어로 빅 데이터 분석 기술을 통해 분석된 데이터를 가시화하기 위한 대표적인 프로그래밍 언어다[14].

- D3.js는 'Data Driven Documents'의 줄임말로 데이터 가시화를 떠올리면 가장 먼저 생각 날 정도로 유명한 도구이다. 차트와 다이어그램을 만들기 위해 HTML, CSS, SVG를 사용한다. 패키지로 제공되며 상호 작용이 풍부한 것이 특징이다. 대부분은 무료이자 오픈소스로 제공 된다. D3.js는 미리 제작된 차트를 외부로 출력할 수는 없지만, 자체적인 갤러리 기능을 갖고 있다. 하지만 D3.js는 대부분의 웹 브라우저 중 최신 표준 기술이 업데이트 된 웹 브라우저에서만 구동된다. 때문에 오래된 브라우저에 게시되는 우려가 없을 때 사용하여야 한다.

(2) 코딩이 필요 없는 비 개발자를 위한 도구

- Datawrapper는 인터랙티브 차트를 만들기 위한 온라인 도구이다. CSV파일을 이용하거나, 직접 데이터를 붙여넣기만 하면 Datawrapper가 자동으로 막대, 선 그래프를 비롯한 가시화기법을 사용하여 데이터를 가시화 시켜준다. 대부분 기자, 뉴스관련 조직들이 기사에 차트를 넣기 위해 Datawrapper를 사용하며, 매우 간편하며 누구나 원본 데이터를 다운로드 할 수 있다는 특징이 있다.

- Tableau는 데이터베이스 질의 결과를 가시화 하는 VizQL엔진을 사용한 가시화 솔루션으로 차트, 그래프, 지도 등 데이터 가시화 솔루션 중 가장 유명한 도구이다. 무료로 제공되며 차트와 그래프를 실시간 데이터 분석 및 결과를 확인 할 수 있으며 Drag-and-Drop방식으로 빠른 설계가 가능하다는 특징이 있다. 또한 공유가 쉽고 공유한 대시보드(dashbord)를 통해 사용자들은 가시화 방식으로 데이터 확인이 가능하다[8].

본 논문에서는 편리한 작업환경을 제공 하며 대표적인 가시화 프로그래밍 언어인 R을 사용하는 R Studio를 사용한다.

아하는 라이브러리이다. 필요한 라이브러리 로드가 완료된 후에는 read.csv() 명령어를 통하여 working directory에 있는 파일을 읽어온다. 이 때, 한글보다는 영문으로 된 것이 훨씬 어려가 적게 발생하고 접근하기 편하므로 파일명과 컬럼 값을 영문으로 바꾸어준다.

CSV 데이터가 올바르게 Import 되었는지 확인하는 방법에는 View() 명령어와, str() 명령어가 있다. View() 방식은 데이터를 테이블 형태로 확인이 가능하고, str() 방식은 String 형태로 Console에서 직관적으로 확인이 가능하다[15].

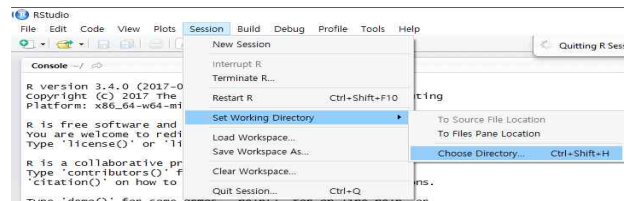


그림 11. 작업 디렉토리 설정
Fig. 11. Set Working Directory

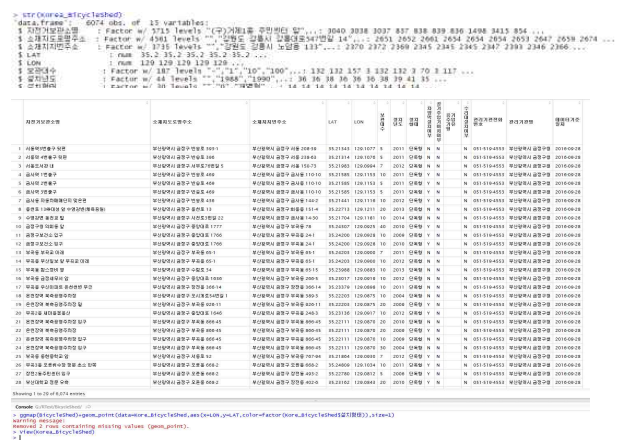


그림 12. str()과 View() 명령어의 차이
Fig. 12. Difference between str() and View() commands

2-2 콘텐츠 구현

빅 데이터 가시화 방법에는 시간, 분포, 관계, 비교, 공간 가시화가 있다. 본 논문에서는 소셜 네트워크가 모바일 중심으로 발전하면서 활발하게 공유되고 있는 공간 가시화 중 지도 기반의 Map 가시화 방법을 적용하여 구현하였으며, 다양한 분야의 빅 데이터 중에서도 앞서 일반적인 데이터 보다 상대적으로 가치가 높은 데이터로 공공데이터를 직접 활용하여 구현하였다.

1) 공공 Open API 다운 및 R Studio 작업환경 구축

먼저 공공데이터 포털(data.go.kr)에 접속하여 전국 자건거 보관소 설치 장소 표준데이터를 다운받는다. 다음 R Studio에 접속하여 [Set Working Directory]-[Choose Directory]를 눌러 작업 디렉토리를 설정하고, 다운받은 표준 데이터를 작업 디렉토리에 저장한다. 이후 install.package() 명령어를 입력하여 ggplot2와 ggmap 라이브러리를 다운로드 한 후, library() 명령어를 통해 다운받은 패키지들을 로드한다. 라이브러리는 데이터 가시화에 가장 중요한 라이브러리이며, ggmap 라이브러리는 맵 데이터를 가시화할 때 필수로 설치해

2) 지도의 위치정보 설정 및 Map 기반 가시화

데이터가 올바르게 들어왔는지 확인 후, 지도의 위치정보를 얻기 위해 google map API에서 제공하는 geocode() 함수를 통해 경도와 위도 값을 설정한다. 이때, 위도와 경도 값은 string으로 들어오기 때문에 as.numeric()함수를 통해 수치형 벡터로 만들어 준다.

다음 get_googlemap()함수를 통해 중심점, Map 형식과, Base Layer의 color, Map size, zoom의 크기를 지정한다. 본 논문에서는 중심을 서울로 지정하였으며, map type을 'roadmap'

으로 지정하였고, Base Layer는 흑백으로 지정하고, Map의 size는 2로 지정하였으며 zoom 값은 7로 지정하였다.

마지막으로 import한 데이터를 기반으로 Map 기반의 가시화를 위하여 ggmap()함수와 geom_point()함수를 사용한다. ggmap() 함수는 가시화하고자 하는 map data로 본 논문에서는 Base Layer를 사용하였다. geom_point() 함수의 파라미터에는 사용하고자 하는 데이터, 위도, 경도, 표시하고자하는 데이터, 점의 크기가 있다[13].

논문에서는 설치 형태에 따른 자전거 보관소 장소에 대한 정보를 점의 형태로 찍기 위해 데이터는 읽어들인 Csv를 지정하였고, Csv에 입력되어있는 위도와 경도 값을 각각 저장하고, Factor 값을 설치형태로 주었으며, 점의 크기는 1로 지정하였다. 결과는 그림 14와 같다.

```

> library(ggplot2)
need help? try the ggplot2 mailing list: http://groups.google.com/group/ggplot2.
> library(ggmap)
Google Maps API Terms of Service: http://developers.google.com/maps/terms.
Please cite ggmap if you use it: see create() or map() for details.
> gc=geocode("seoul,korea","source=google")
Information from url: http://maps.googleapis.com/maps/api/geocode/json?address=seoul&sensor=false
> center=as.numeric(gc)
> #
lon lat
[1] 126.9730 37.56654
> center
[1] 126.9730 37.56654
> map=ggmap(ggmap(center=center,nptype = "roadmap",language = "ko-kr",color="bw",scale=2)
Map from url: http://maps.googleapis.com/maps/api/staticmap?center=37.566535,126.973698&zoom=10&size=640x640&scale=2&nptype=roadmap&language=ko&sensor=false
> ggmap(map,extent = "device")
warning message: panel.margin is deprecated. please use panel.spacing property instead
> korea_bicycleshed.csv<korea_bicycleshed.csv">
> #
str(korea_bicycleshed)
data.frame: 6074 obs. of 15 variables:
 $ 자전거보관소명 : factor w/ 5713 levels "(주)기림들 자전거보관소",...: 3040 3038 3037 437 438 439 836 1498 3415 854 ...
 $ 도로명주소 : factor w/ 4501 levels "", "경기도 고양시 인문로547번길 14",...: 2051 2652 2661 2654 2654 2654 2653 2647 2639 2674 ...
 $ 도로명주소 : factor w/ 3735 levels "", "경기도 고양시 인문로 117",...: 3704 2352 2369 2345 2349 2345 2347 2389 2346 2366 ...
 $ LAT : num 35.2 35.2 35.2 35.2 35.2 ...
 $ LON : num 129.129 129.129 129.129 ...
 $ 보행인 : factor w/ 187 levels "", "1", "10", "100",...: 132 132 137 3 132 132 3 70 3 117 ...
 $ 장애인 : factor w/ 46 levels "", "1", "100",...: 146 146 146 146 146 146 146 146 146 146 ...
 $ 장애인보행 : factor w/ 30 levels "", "0", "100",...: 14 14 14 14 14 14 14 14 14 14 ...
 $ 장애인보행 : factor w/ 6 levels "", "0", "100",...: 3 3 3 3 3 3 ...
 $ 장애인보행 : factor w/ 5 levels "0", "10", "100",...: 3 3 3 3 3 3 ...
 $ 장애인보행 : factor w/ 4 levels "0", "10", "100",...: 3 3 3 3 3 3 ...
 $ 장애인보행 : factor w/ 2 levels "0", "10", "100",...: 1 1 1 1 1 1 ...
 $ 장애인보행 : factor w/ 10 levels "0", "100", "1000",...: 424 424 424 424 424 424 424 424 424 ...
 $ 장애인보행 : factor w/ 710 levels "자전거보관소 명칭",...: 239 239 239 239 239 239 239 239 239 ...
 $ 장애인보행 : factor w/ 136 levels "2016-08-11", "2016-09-02",...: 29 29 29 29 29 29 29 29 29 ...
> #korea_bicycleshed.csv<korea_bicycleshed.csv">
map From url: http://maps.googleapis.com/maps/api/staticmap?center=seoul&zoom=7&size=640x640&scale=2&nptype=roadmap&language=en&sensor=false
Information from url: http://maps.googleapis.com/maps/api/staticmap?center=seoul&zoom=7&size=640x640&scale=2&nptype=roadmap&language=en&sensor=false
> ggmap(@korea_bicycleshed)geom_point(data=korea_bicycleshed,aes(x=LON,y=LAT,color=factor(korea_bicycleshed$자전거보관소명)),size=2)
warning message: removed 2 rows containing missing values (geom_point).
>
  
```

그림 13. 명령어 입력
Fig. 13. Input Command

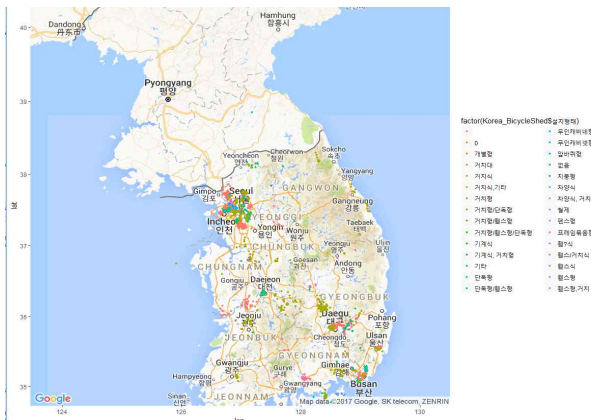


그림 14. Map을 이용한 가시화
Fig. 14. Visualization with Map

III. 결론

본 논문에서는 R 라이브러리와 R Studio를 통해 공공데이터를 가시화하는 방법에 대해 제안하였다. 급변하는 디지털 정보사회에서 처리해야 할 많은 데이터의 유형은 다양해지고 증가하면서 사람들은 유용한 정보를 직관적이면서 쉽고 빠르게 습득하기를 원한다. 그 중에서도 데이터의 활용 가치가 높은 공공데이터에 대한 관심도는 날도 증가할 것이다. 본 논문에서는 실제 전국 자전거 보관 장소에 대한 공공 데이터를 기존의 텍스트 형태로만 단순히 보았던 데이터를 사람들의 눈으로 데이터를 직접 확인함으로써, 해당 데이터가 갖는 가치 판단에 좀 더 유연하게 적용 해볼 수 있었다.

적용 가시화 결과로서 자전거 보관소는 대부분 도심에 많이 몰려있는 것을 확인할 수 있었고, 주로 개별 형이 많은 것을 확인할 수 있었다. 이처럼, 무수히 많은 데이터를 일일이 다 확인 할 필요도 없이 가시화를 통해 한눈에 알아 볼 수 있도록 하여 기존의 틀을 깨고 심층적으로 데이터를 분석 할 수 있다는 점에서 데이터의 가시화가 앞으로도 많이 활성화 될 전망이다. 향후 연구로는 단순히 가시화만을 표현 한 것이 아닌 새로운 의미를 부여할 수 있는 재사용성이 가능한 콘텐츠 구현으로 쓰일 수 있도록 할 예정이다.

참고문헌

- [1] Sang-Chul Kim, Kwang-Ho Kim, "A Study on factors affecting the viewer rating of My Little Television : Focusing on SNS Big Data", The Journal of Digital Contents Society, Vol. 17, No. 1, pp. 1-10, Feb 2016.
- [2] James Manyika, Michael Chui, Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, p. 1, 2011.
- [3] John Gantz, David Reinsel, "Extracting Value from Chaos", 5th annual Digital Universe study from IDC, p. 6, June 2011.
- [4] Laney, Douglas (2001, February). 3D Data Management: Controlling Data Volume, Velocity and Variety[Internet]. Available: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3-D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- [5] Villanova University. What is Big Data[Internet]? Available:

- <https://www.villanovau.com/resources/bi/what-is-big-data/>
- [6] Boris Evelson, Brian Hopkins (2011, September). Big Data Enlarges Digital Horizons[Internet]. Available: <https://www.forrester.com/report/Expand+Your+Digital+Horizon+With+Big+Data/-/E-RES60751?objectid=RES60751>
- [7] Seung-Yeol Bang, Hyo-Dong Ha, Chang-Jae Kim, "A Study on Big Data-based Software Architecture Design for Utilizing Public Open Data", The Korea Information Technology Society, Vol. 13, No. 10, pp. 99-107, Sep 2015.
- [8] Mun-Doo Hwan, Do-Nam Chul, "Big Data", Korean Society of CAD/CAM Engineers, Vol. 18, No. 2, pp. 26-29, August 2012.
- [9] Deok-Young Jeong, Jun-Suk Lee, and Sang-Sung Park, "Technology Trend Analysis Using Word Cloud", The Korea Intelligent Systems Society, Vol. 26, No. 1, pp. 17-18, April 2016.
- [10] Jung-Woo Ryu, Jin-Hee Song, "Visualization for Big Data", The Korea Contents Association, Vol. 12, No. 1, pp. 21-26, March 2014.
- [11] ATUSERVICIO. Uruguay Health Authority Healthcare Management Platform[Internet]. Available: <http://atuservicio.uy/>
- [12] Opower. US Opower's National Home Energy Utilization Comparison Platform[Internet]. Available: <https://social.opower.com/maintenance.html>
- [13] Nathan Yau, Visualize this : the FlowingData guide to design, visualization, and statistics, JohnWiley & SonsInc, pp. 120 ~ 403, 2011.
- [14] Hee-won Jeon, R visualization of data: Visualizing data to convince customers Practical know-how, Hanley Media, pp. 9-15, 2013.
- [15] Seon-Hui Bak, Hyun-Bea You, Jong-Hwan Bae, and Tae-Jun Choi, "Implementation Of Public Data contents Using Big Data Visualization Technology – Focusing on Utilization of Map visualization Technique", The Proceedings of the 2017 KIIT·DCS Summer Conference, Kumho Institute of Technology, pp. 143-144, 2017.



박선희(Seon-Hui Bak)

2007년 : 공주대학교 영상예술대학원 (공학석사)
2016년 : 부산외국어대학교 대학원 (공학박사-ICT창의융합)

2012년~2017년: 아이에이치테크 이사

현재: ㈜유토비즈 이사

※관심분야 : HCI, 빅데이터(Bigdata), ICT, 가상현실, 증강현실, 인터랙티브 등



김정호(Jong Ho Kim)

1980년:경북대학교 전자공학과(공학사)
1983년:경북대학교 전자공학과(공학석사)
1994년:단국대학교컴퓨터공학과 (공학박사)

1983년-1996년:한국전자통신연구원(실장)
1989년 정보처리기술사

1996년~현재: 한밭대학교 컴퓨터공학과 교수

※관심분야 : 데이터통신, 정보보호, 데이터의 시각화 등



유현배(Hyun-Bea You)

2000년 : 일본국립쓰쿠바대학교 대학원 (공학박사-지능기능공학전공)

2002년~현재: 나사렛대학교 교양교육학부 교수

※관심분야 : 신학, 유니버설디자인, 증강현실(AR), 빅 데이터(Big data), 인터랙티브, 영상정보처리