We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

GIS and Big Data Visualization

Junghoon Ki

Abstract

Geographic information system (GIS) has expanded its area of applications and services into various fields, from geo-positioning service to three dimensional demonstration and virtual reality. Big data analysis and its visualization tools boosters the capacity of GIS, especially in graphics and visual demonstration. In this chapter, I describe major traits of big data and its spatial analysis with visualization. And then I will find a linkage between big data and GIS. There are several GIS-based software and geo-web that deal with big data or similar scaled databases, such as ArcGIS, Google Earth, Google Map, Tableau, and InstantAtlas. For these software and websites are developed based on geography or location, they still have some limits in visualizing big data or persuading people with maps or graphics. I will search a way out of this limitation of GIS-based tools and show an alternative way to visualize big data and demonstrate thematic maps. This chapter will be a useful guide to lead GIS people into a new horizon of big data visualization.

Keywords: GIS, big data, visualization, infographic, Tableau, Chernoff face

1. Introduction

For decades, geographic information system (GIS) has expanded its area of applications and services into various fields, from geo-positioning service to three dimensional demonstration and virtual reality. It is a tremendous progress of GIS since its burgeoning as a combination of map and database. Today, everyone in the world is living, working, and resting under the umbrella of GIS applications and services in the form of navigation system, the Google Earth, GPS, and even Pokémon GO.

Stronger and more fundamental changes are asked in GIS development when big data emerged in the early 2010 [1, 2] (see **Figure 1**). Characterized with a large volume, a vast variety, and a fast velocity, big data has been releasing the explosive datasets in social media and other complex platforms. Is big data just good news to the GIS community? According to Sanderson [3], there are still some hurdles that preventing GIS and big data from joining together. They are related with big data's unstructured data structure, real time data production, accuracy, and scale. Not only these obvious limitations of big data, it also neglects locations of datasets frequently. Big data deals mainly with information, not necessarily geography.

What is a pivotal role of big data in GIS development? It is necessary to take a look at a process of big data production to find out a connection between two fields. The big data production process consists of data collection, storage, computing & batching, analysis, and visualization & demonstration. Among the process, visualization and demonstration could provide an effective and efficient way with GIS people in terms of new interpretation and creative advertisement.
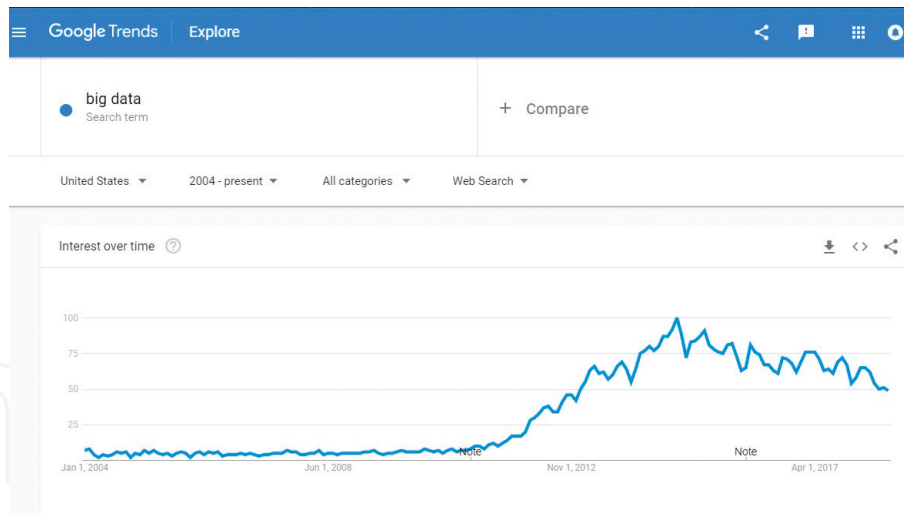
Figure 1.
Google Trends Search for big data since 2004 to present.

Big data's several or more visualization tools with their software are creating a lot of wonderful GIS masterpieces recently. Thereafter, I examine those tools and find some implications from them. Can big data visualization overcomes the limitations of GIS and opens a new horizon? This chapter would provide answers to this question.
The followings constitute sub-sections of the chapter.

- What is big data

- Big data and geographic information system (GIS)

- Big data as an alternative visualization tool for GIS

- Can big data visualization overcome GIS limitations?

- Conclusion

2. What is big data?

2.1 Big data's characteristics and components

Big data can be defined as datasets which have various data styles, fast processing speed, and are hard to be managed and analyzed with existing data systems. These characteristics of big data are summarized with '3V', which denotes volume, variety, and velocity [4].
First, big data deals with large volume datasets, usually more than terabyte size that usually comes from Global Positioning System (GPS), social media, and other sensors. A terabyte is a unit of information equal to one million * million (10^{12}) bytes, or 1024 gigabyte. The brand 'big data' itself implies a size of datasets is very huge compared to past datasets.
Second, big data deals with a variety of datasets such as sound, picture, video stream, map and even social media text message. Big data targets not only structured datasets but also unstructured ones that were usually out of interest to data workers. Its range is beyond our imagination and different kinds of datasets are integrated to

generate new types of database. Big data systems use a computer clouding and other platform such as Hadoop for data combination and integration (see **Figure 2**).

Big data's third characteristic is velocity because it's very fast in generating, spreading, and applying in the real world. Big data's speed in generation, spread, and application can be accelerated with social media or social network services such as Facebook or Twitter [6]. When people post photos in Facebook, those are recorded as datasets, which offer the useful real-time evidence of locations, preference, and other personal information (see **Figure 3**). This information will be used for marketing and sales by private business or policy measures by government sector.

Although a narrow definition of big data emphasizes data source, collection, storage and other technical issues, its wider definition embraces analysis and demonstration aspects. In summary, big data is defined as very large-sized, various-formatted datasets and analytic methods based on engineering technology and social network services, including statistical fusion and new visualization.

Major components of big data are resource, technology, and human capital [4]. Resource here indicates data acquisition and quality management. Big data technology denotes its platform that refers to data storage, management, processing, analysis, and visualization. Human capital in big data is called data scientists who have an ability of mathematics, engineering, economics, statistics, and psychology. They are also asked to have a capacity of communicating with other people, making a creative storytelling, and visualizing their big data contents effectively.

2.2 Big data's data process and analysis techniques

In **Figure 4**, I briefly show big data process with its elements in which the process has data source, collection, storage, processing, with analysis and visualization. Each step of process has a considerably different elements from the past database systems that generally dealt with structured datasets.

First, big data's data sources come from institutions' or organizations' internal database, or external database such as Twitter or Facebook, or pictures and video streams. Generally, urban and geographic researches and projects use a large scale spatial database [7], which can be called big data.

Second, in the collection process, big data utilizes a crawling method with search engine to get Internet data. It also uses Internet of Things (IoT) based sensors to collect data. This step makes a huge difference to big data from the past data collection traditions.
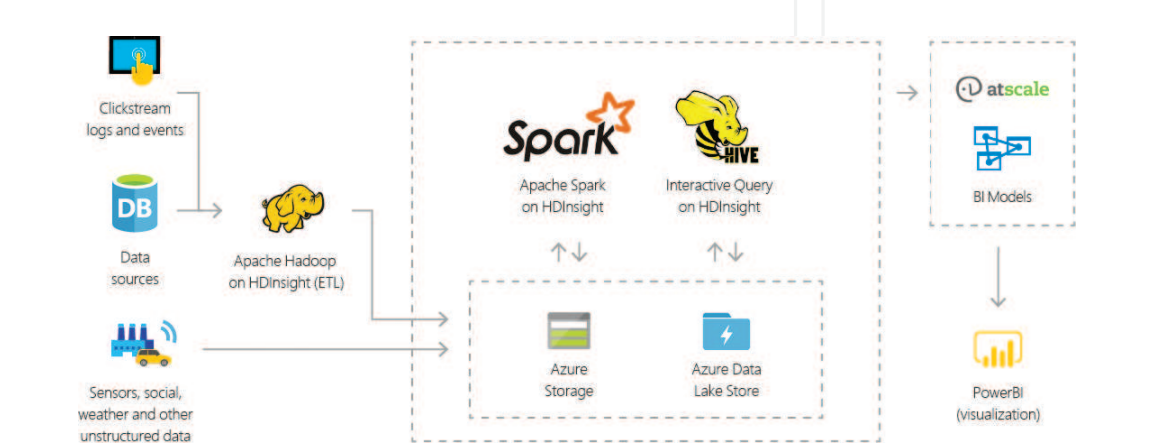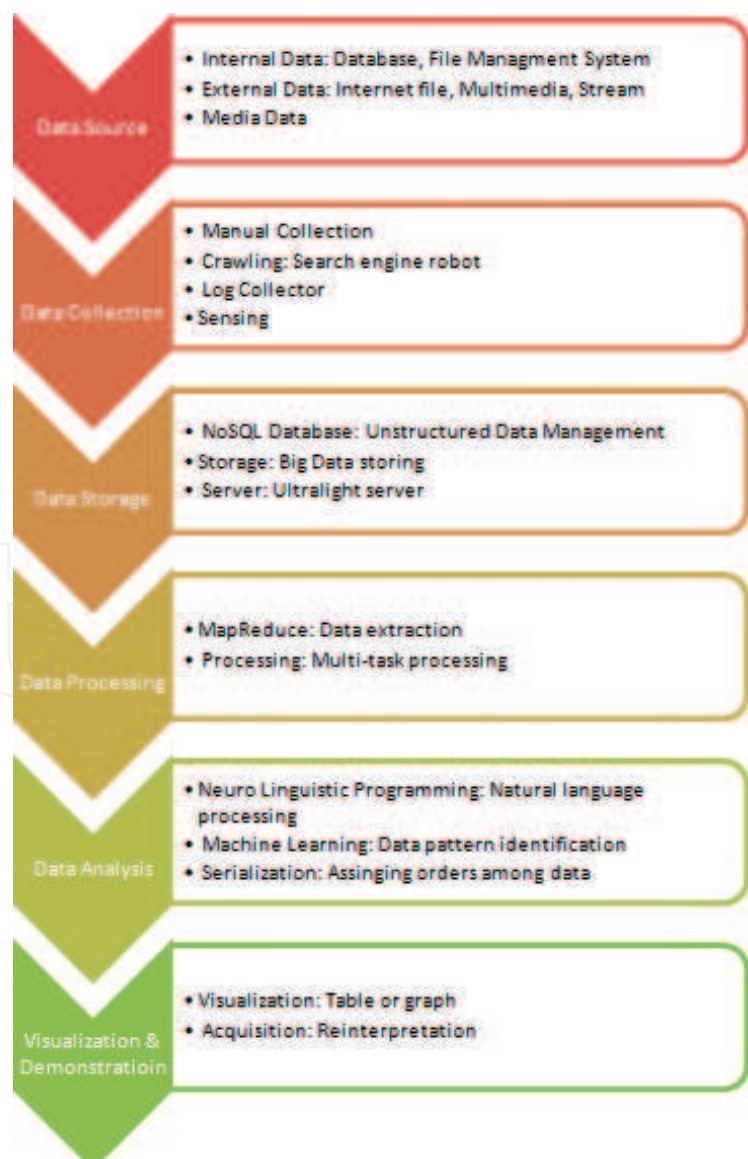

Figure 2.
Data warehousing with Apache Hadoop [5].

Figure 3.
Twitter image that contains some personal information.

Figure 4.
Big data processing and elements [4, 9].

Third, data storage is a step that engineering technologies are concentrated. Big data managers have to control unstructured data with Not Only SQL (NoSQL), extract data with MapReduce, and execute a distributed parallel processing with Hadoop.

In big data analysis, researchers use neurolinguistic programming for natural language processing, machine learning for data pattern identification, and serialization for assigning orders among data. Researchers pay attention to R programming to conduct big data analysis because it is an efficient statistical tool compared to other packages. Many statistical packages begin to equip themselves with big data analysis modules recently.

Big data visualization and demonstration is a process that analyzed datasets are expressed with graph or table format. Merits of big data visualization in comparison with traditional data visualization is that the former uses word/text/tag clouds, network diagrams, parallel coordinates, tree mapping, cone trees, and semantic networks [8] more often than the latter because its data source format and their needs. R, Tableau, Python language are getting a new attention as effective visualization tool for big data demonstration.

In the next section, I find out a relationship between big data and GIS in terms of these six steps of big data processing.

3. Big data and geographic information system

Big data and GIS are able to share several aspects together because they are similar in elements of data processing. In **Figure 5**, I show GIS data processing with its elements. There are popular open source or commercialized software and web-based online GIS systems, which play an important role in processing and analyzing GIS data.

First, GIS uses data that contains a location or space, therefore it is displayed in a map or picture form. Recently, aerial or satellite data becomes more and more important as new technologies are introduced. As a location based data, GIS data is usually large-sized as is big data.

Second, GIS collects field data such as street information, Closed Circuit TV (CCTV), or other location-based datasets. If the datasets do not provide location information, GIS technicians should perform a geo-coding process to convert into GIS datasets. People's participation is also an important way to get GIS data; so the participatory GIS system becomes a significant field of GIS. Crawling with search engine robot is also useful tool for obtaining data in GIS.

Third, GIS has web server, geospatial data server, or cloud server for its data storage. These servers can be overlapped one another sometimes, but they have their own territories that cannot be shared. In **Figure 6**, I introduce a basic principle of geo-database for single-user and multi-users with the ESRI's official website information. Geo-database system is crucial to manage complicated structured GIS datasets and their attributes.

Fourth, GIS desktop and online software plays a pivotal role in the rest of process including data processing (building), analysis, and visualization. In the GIS data processing (building), efficient systems included are ArcGIS Online, Google Maps JavaScript API, Here Maps JavaScript API, Microsoft Bing Geocode Dataflow API, and US Census Geocoder. They are helpful for building up geo-coding and mapping coordinates in the database.

Fifth, GIS data analysis contains several functions as **Table 1** briefly shows with ArcGIS analysis toolbox summary. Similar analyses are conducted with other software such as ArcGIS, QGIS, GRASS GIS, GeoDa, CartoDB, Mapbox, and the other desktop or online GIS systems.
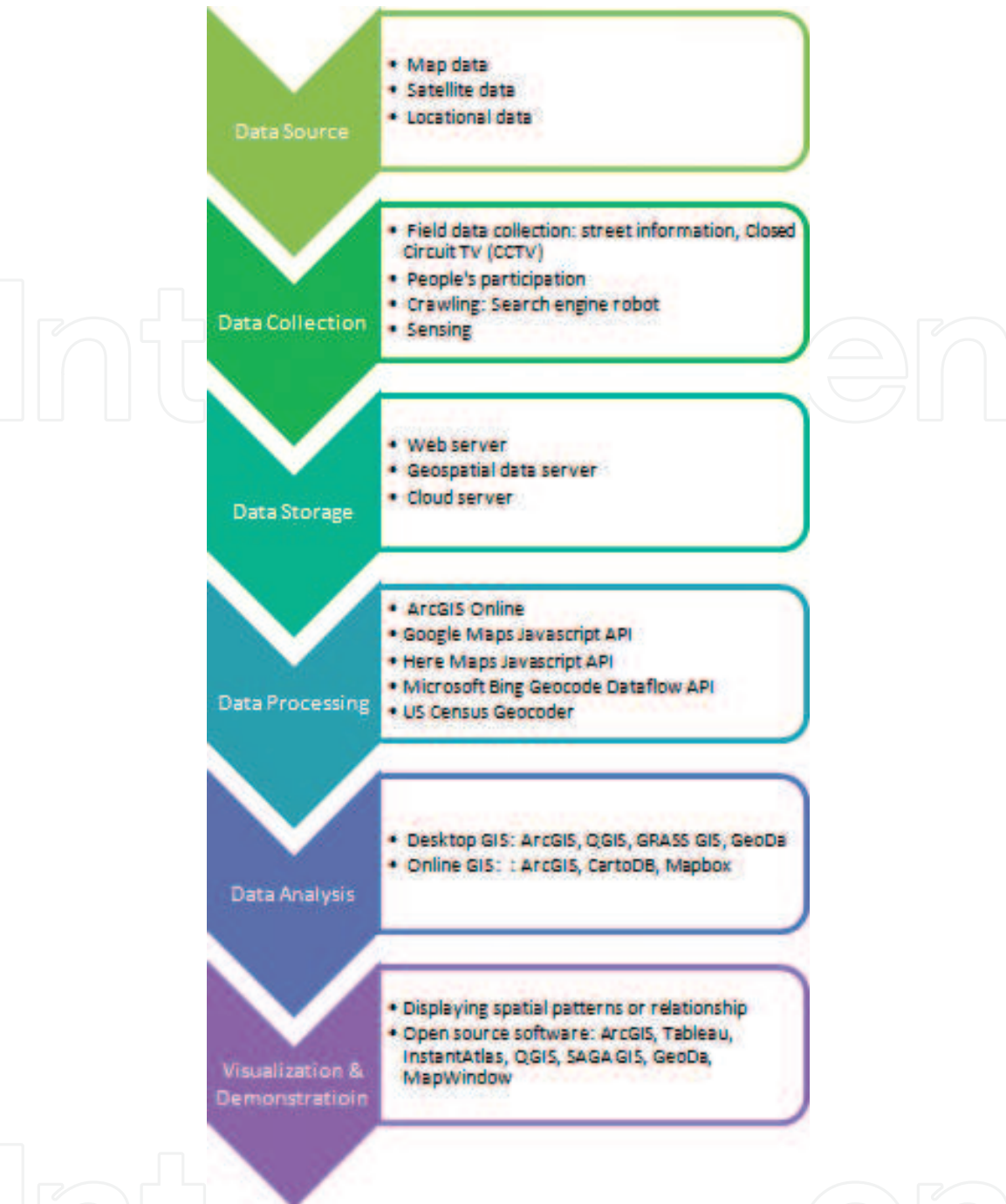
Figure 5.
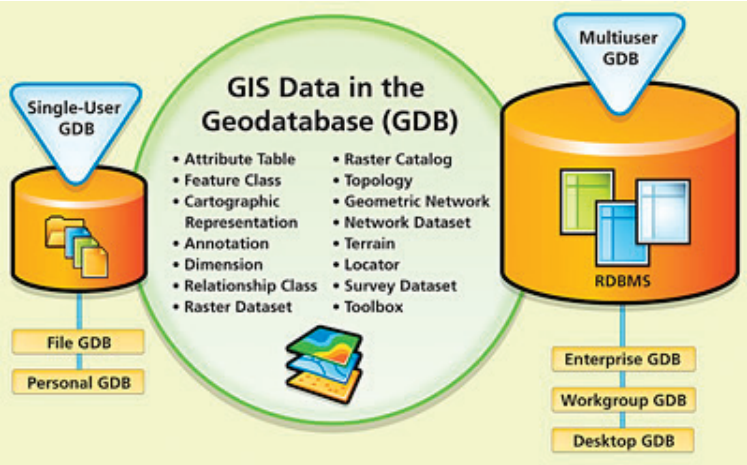*GIS data processing and its elements.*

Figure 6.
GIS data in the geo-database (GDB) [10].

Toolsets	Description
Extract	GIS datasets often contain more data than you need. The Extract tools let you select features and attributes in a feature class or table based on a query (SQL expression) or spatial and attribute extraction. The output features and attributes are stored in a feature class or table.
Overlay	The Overlay toolset contains tools to overlay multiple feature classes to combine, erase, modify, or update spatial features, resulting in a new feature class. New information is created when overlaying one set of features with another. There are six types of overlay operations; all involve joining two existing sets of features into a single set of features to identify spatial relationships between the input features.
Pairwise Overlay	The Pairwise Overlay toolset provides an alternative to some of the tools in the Overlay toolset.
Proximity	The Proximity toolset contains tools that are used to determine the proximity of features within one or more feature classes or between two feature classes. These tools can identify features that are closest to one another or calculate the distances between or around them.
Statistics	The Statistics toolset contains tools that perform standard statistical analysis (such as mean, minimum, maximum, and standard deviation) on attribute data as well as tools that calculate area, length, and count statistics for overlapping and neighboring features. The toolset also includes the Enrich tool that adds demographic facts like population or landscape facts like percent forested to your data.

Table 1.
GIS data analysis toolbox example with ArcGIS [11].

Sixth, GIS data visualization intends to display spatial patterns or relationship between or among locations. Popular open source software included here are ArcGIS, Tableau, InstantAtlas, QGIS, SAGA GIS, GeoDa, and MapWindow. These tools are
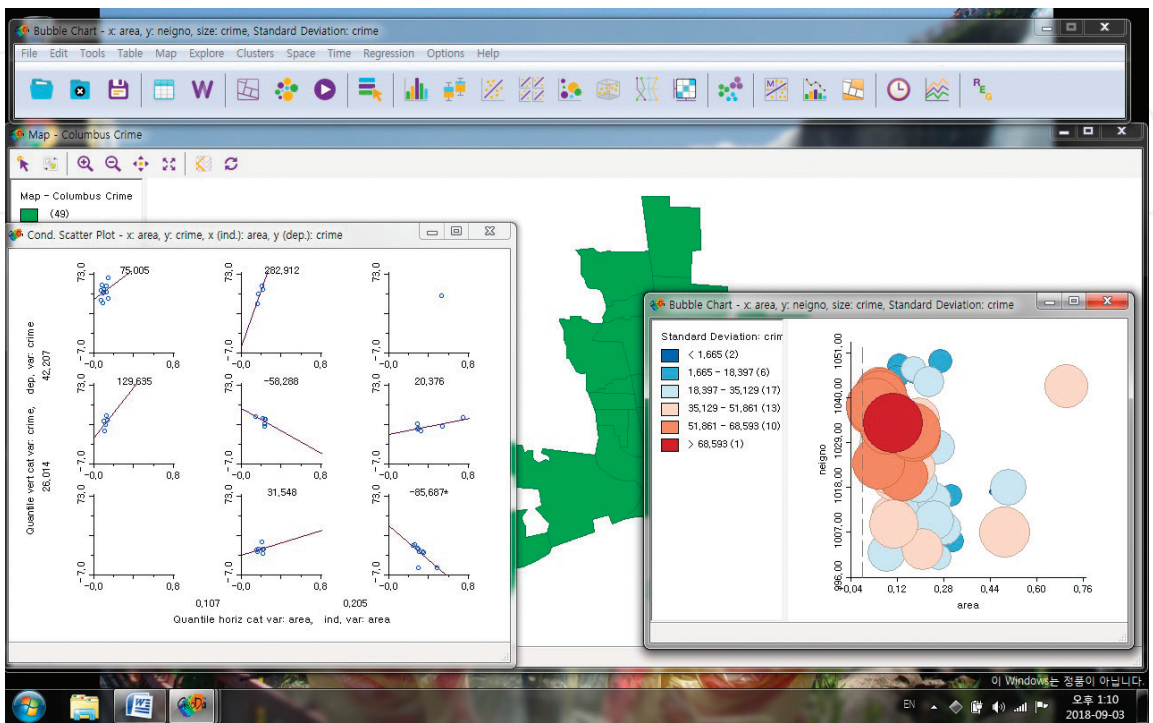
Figure 7.
GeoDa desktop software's visualization windows with crime data.

actively adapted to big data based software or systems to build up location oriented systems as well as more persuasive graphic works. **Figure 7** shows visualization windows with GeoDa desktop software.

## 4. Big data as an alternative visualization tool for GIS

Can big data be an alternative tool for visualizing GIS and mapping works? Does big data plus location data equal to GIS data? Does big data visualization have any hidden card that surpasses GIS visualization and mapping? I will find answers to these questions in writing this section. Big data's potential for an alternative visualization tool for GIS is to be drawn from several examples in big data technology.

In the visualization and demonstration technology, big data and GIS share together in some aspects. However, there is each field's original aspect that cannot be shared or come together (see **Figure 8**). In **Figure 8**, there are three areas defined: (A) as a GIS visualization's exclusive area, (C) as a big data visualization's exclusive area, and (B) as an overlapping area between two technologies.

GIS visualization's exclusive area (A) indicates that visualization takes places based on location or map with geographic coordinates. Meanwhile, big data visualization's exclusive area (C) means a visualization demonstration without a location or map, which denotes no spatial context are provided. Many big data visualization outcomes do not have any geographic traits or variables and belong to this exclusive area.

Figure 9 is an exemplary map of the area (A), while **Figure 10** is an instance of the area (C). **Figure 9** shows US cities by their elevation in which larger bubble implies higher the city location. I can create this figure using US city and state shape (.shp) files with ArcView GIS software. **Figure 10** shows a gender and ethnicity in tech companies with online Tableau public. In this visualization, there is no evidence of location or mapping technology. This is a pure big data visualization area that is not related with a spatial context or geographic coordinates.

What is the overlapping area (B) that both GIS and big data work together or cooperate? In the (B) area, locations or geographic coordinates are important factor, and big data visualization technologies are also playing a crucial role in demonstration. In **Figure 11**, I provide an example of area (B) with the Chernoff face and US map, in which the Chernoff face denotes multivariate big data visualization using human face-like variables with SAS or R programming. There are many other visualization examples available if any big data expressions are embedded in maps or spatial context. **Figure 11** is also a good example of area (B) because it is clearly telling the location although it does not use a map. **Figure 12** shows how much population is moving from a continent to another with big data visualization technology of Tableau software.
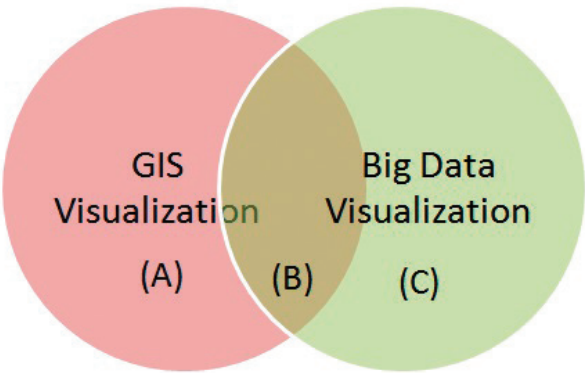
Figure 8.
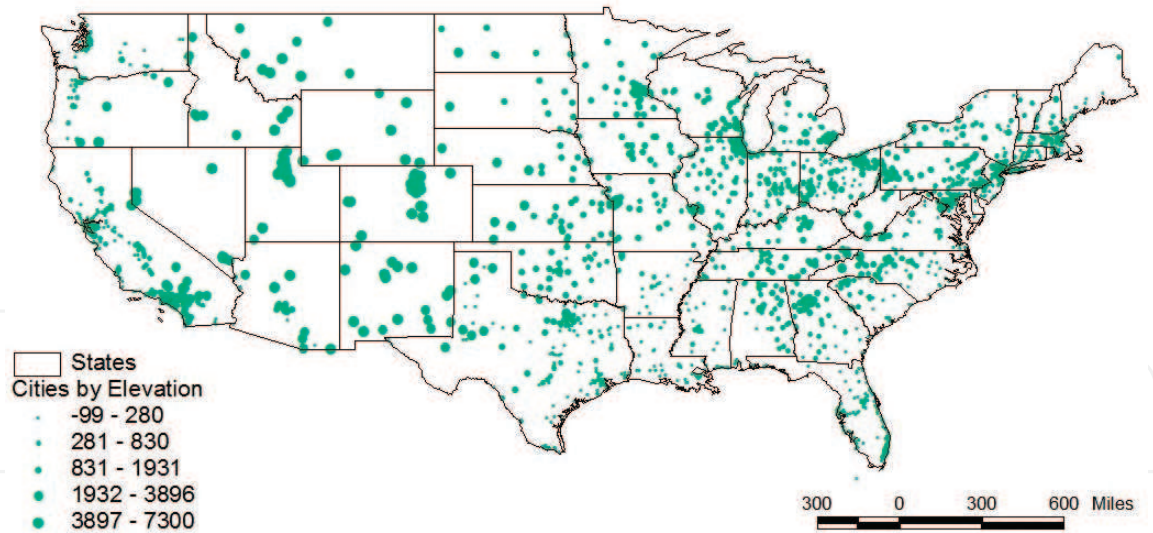*GIS visualization and big data visualization Venn diagram.*

Figure 9.
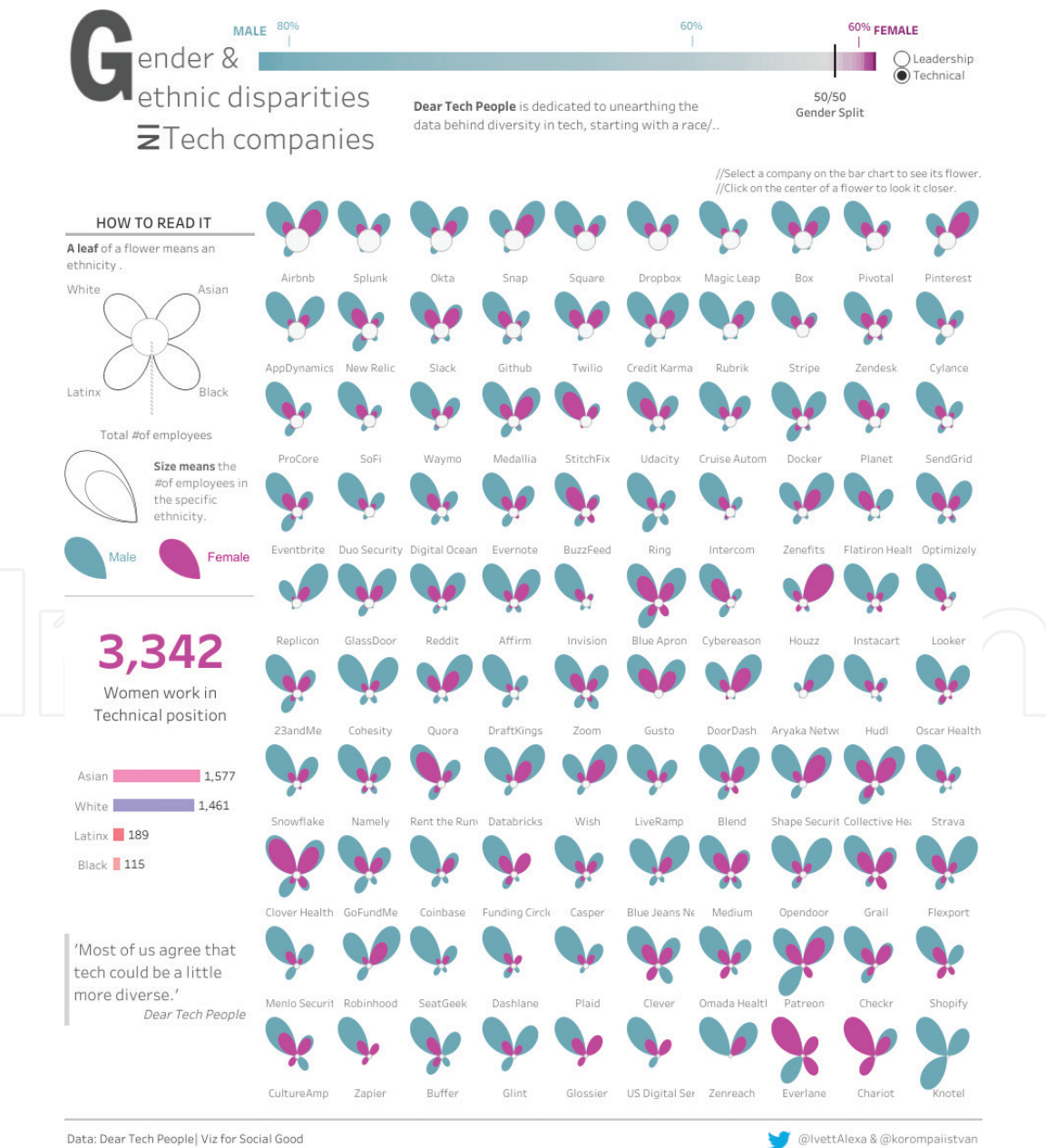*GIS visualization example: US cities by their elevation with ArcView GIS.*

Figure 10.
Big data visualization example: gender and ethnicity in tech companies with online Tableau public [12].

9

Figure 11.
US states' death penalty executions since 1976 [13].
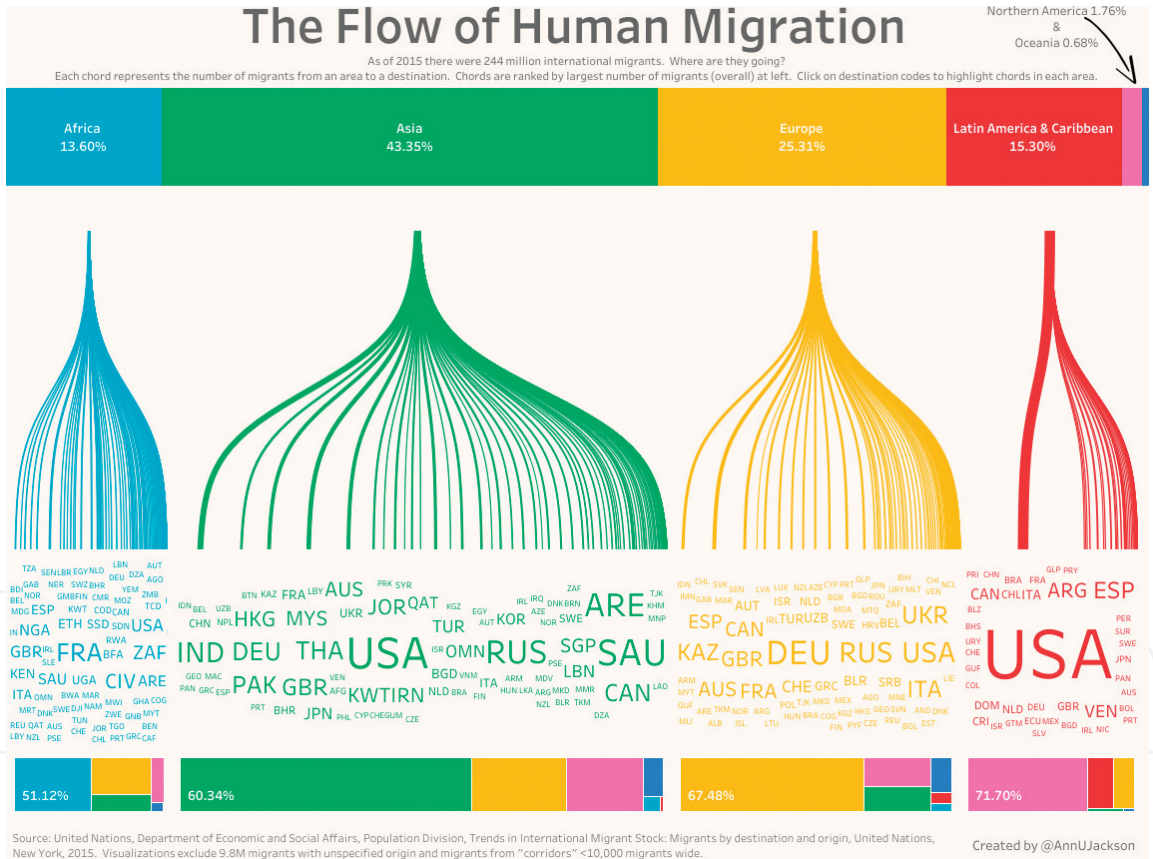

Figure 12.
The flow of human migration with online Tableau public [12].

Does big data visualization overcome GIS and its limitation? About this issue, I describe some insights in the following section.

5. Can big data visualization overcome GIS limitations?

GIS visualization has a limitation since it is basically rooted at the spatial context and geographic maps. GIS visualization's first priority tends more to be geographic

than to be informational or graphical. Location matters at GIS visualization as it did at mapping and geography.

Big data visualization opens a new horizon in GIS visualization because it does not just strengthen the spatial context, but also it gives new meanings and insights to GIS maps and demonstration. As is compared in **Figures 8** and **10**, dots in GIS visualization turn into human faces in big data visualization. **Figure 11** implies that locations can be read without a map. More big data visualization skills and their outcomes will be brought out with more abundant insights and implications to GIS visualization.

However, there are some risks of big data visualization in applying to GIS visualization because their fundamental approaches are different in some ways.

First, big data's engineering technologies tend to be ignorant to geographic perspectives. Big data engineers and visual technicians are not necessarily geographers, spatial experts, or even urban planners. Big data visualization workers if loaded with GIS related jobs should be aware of basic spatial principles and mapping process.

Second, GIS experts who is creating big data related visualization should be ready to adapt themselves to engineering guidelines that ask them set their spatial norms aside to set up new GIS-based big data visualization works. When GIS professionals get a step back, they will experience a power of big data visualization technology.

Third, GIS and big data visualization works should be multidisciplinary projects or research, in which all possible fields of study are involved in the final production. Social scientists, data engineers, medical & health experts, graphic designers, and other research fields' professionals can join to generate meaningful GIS visualization performance [14].

Big data visualization can be a good measure if people involved are deliberately designed, called, instructed, and allocated.

6. Conclusion

Big data is defined as very large-sized, various-formatted datasets and analytic methods based on engineering technology and social network services, including statistical fusion and new visualization. A narrow definition of big data emphasizes data source, collection, storage and other technical issues, but its wider definition embraces analysis and demonstration aspects.

Among big data' data processing, visualization is a process that analyzed datasets are expressed with graph or table format. Big data's advantage in visualization in comparison with traditional data visualization is that the former uses word/text/tag clouds, network diagrams, parallel coordinates, tree mapping, cone trees, and semantic networks [Miller] more often than the latter because its data source format and their needs. R programming, Tableau software, and Python language are getting a new attention as effective visualization tool for big data demonstration.

GIS data visualization displays the spatial patterns or relationship between or among locations. Popular open source software included here are ArcGIS, Tableau, InstantAtlas, QGIS, SAGA GIS, GeoDa, and MapWindow. These tools are actively adapted to big data based software or systems to build up location oriented systems as well as more persuasive graphic works.

Big data visualization opens a new horizon in GIS visualization because it does not just strengthen the spatial context, but also it gives new meanings and insights to GIS maps and demonstration. More big data visualization skills and their outcomes will be brought out with more abundant insights and implications to GIS

visualization. Especially, big data visualization can be a good measure if people involved are deliberately designed, called, instructed, and allocated.

Acknowledgements

Conflict of interest

No potential conflict of interest was reported by the author.

Author details

Junghoon Ki
Myongji University, Seoul, South Korea

*Address all correspondence to: johnki70@gmail.com

IntechOpen

References

[1] Ki J. A big data analysis of urban statistics expression—Chernoff face-based expression of local community health index in Korea. Space and Environment. 2016;**26**(1):336-358

[2] Sui D. The 'G' in GIS: Big data in a small and divided world: Implications for GIS and geography. GEO world. Business Insights: Global. 2012;**25**(3):12-13

[3] Sanderson M. Big data: GIS at the crossroads. GEO: Connexion. Business Insights: Essentials. 2018;**12**(1):6

[4] Jeong J. Three Major Factors for a Successful Big Data Application. Seoul: National Information Society Agency; 2012

[5] Data Warehousing with Apache Hadoop [Internet]. 2018. Available from: http://www.azure.microsoft.com [Accessed: Sep 1, 2018]

[6] Song K. Understanding Society through SOCIALmetrics. Seoul: Daum Soft; 2011

[7] Gao X, Cai J. Optimization analysis of urban function regional planning based on big data and GIS technology. Technical Bulletin. 2017;**55**(11):344-351

[8] Miller JD. Big Data Visualization. Birmingham: Packt Publishing; 2017

[9] Warden P. Big Data Glossary. Sebastopol: O'Reilly Media; 2011

[10] GIS Data in the Geo-database [Internet]. 2018. Available from: http://www.esri.com [Accessed: Sep 1, 2018]

[11] GIS Data Analysis Toolbox Example [Internet]. 2018. Available from: http://www.pro.arcgis. com [Accessed: Sep 1, 2018]

[12] The Flow of Human Migration [Internet]. 2018. Available from: http://www.public.tableau.com [Accessed: Sep 1, 2018]

[13] Huffman, D. On the abuse of Chernoff faces. Cartastrophe [Internet]. 2010. Available from http://cartastrophe.wordpress.com/2010/06/16/on-the-abuse-of-chernoff-faces [Accessed: Sep 2, 2018]

[14] Hesse BW, Moser RP, Riley WT. From big data to knowledge in the social sciences. The Annals of the American Academy. 2015;**659**:16-32