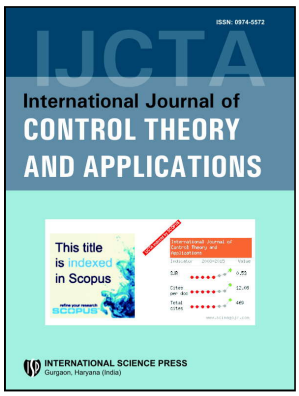

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 18 • 2017

Exploratory Spatial Data Analysis of Public Health Facilities of Chhattisgarh Using Geo-Visualization

Kuhu Gupta[1], Rekh Ram Janghel[2], Ranjeet Kumar[3], Akash Soni[4] and Amir Aman[5]

[1,2,4,5] Department of Information Technology, Emails: kuhu.gupta08@gmail.com, rrjanghel.it@nitrr.ac.in, asoni7667@gmail.com,amanamir78@gmail.com
[3] National Institute of Technology, Raipur, Chhattisgarh, India, Department of Electronics & Communication Engineering, PDPM Indian Institute of Technology, Design and Manufacturing, Jabalpur, Madhya Pradesh, India, Email: ranjeet281@gmail.com

Abstract: Each year major part of a country's wealth is spent on public health. The government organizations keep daily records of large volumes of public health-related data resulting in voluminous multivariate dataset. These datasets are spatially varying which poses a challenge for proper analysis. These datasets cannot be ignored since these are potentially useful for decision makers and public health policy makers to understand the past occurrences, monitor the present, and predict the future of the health scenario in a country. Currently available visualization tools for such datasets fail to give clear insights on multivariate comparisons and spatial distribution of public health resources at the same time. The work propounds an ingenious use of geo-visualization for management of Chhattisgarh public health-care. The organizational aspects of public health resources in Chhattisgarh state are studied with an intention to recognize the districts that stands aberrant regarding availability and accessibility of public health centers for the population. The outcomes aids in decision making for distribution of available resources by health-care officials. Additionally, the cardinal methodological contribution of the work is the introduction of advanced visualization techniques that could be applied to a multitude of fields.

Keywords: geo-visualization, public health, spatial patterns, web-based, visualization, patterns

1. INTRODUCTION

Exploratory Data Analysis (EDA) is a collection of descriptive and graphical statistical tools for detecting interesting patterns, investigating deviant and unexpected attributes, differentiating outliers from salient features and formulate hypothesis [1]. Exploratory data analysis employs relatively simple transformation techniques on different types of data, such as, temporal, spatial and others. In the present days, a plethora of government institutions, research labs, private companies and individuals generate and collect an increasing amount of geospatial data [2]. Traditional exploratory data analysis techniques fails to unveil the trends and untap the complexities of spatial information, therefore, the notion of exploratory spatial data analysis (ESDA) was

introduced. Exploratory spatial data analysis (ESDA) is a subset of exploratory data analysis (EDA) to detect the spatial features of dataset in which every variable value is related to a locational datum [3]. This datum corresponds to the area, place or point to which the attribute refers while other attribute depicts social, demographic or economic aspects. These attributes show an inherent spatial autocorrelation. However, the ESDA techniques focuses at observing spatial patterns in data, clusters or hot spots, assessing spatial models and propose any form of spatial heterogeneity which are based on the geography of the data [4]. In present days, a plethora of scientific and governmental institutions, research labs and private companies generate and collect increasing amount of geospatial data. The traditional methods of ESDA fail to convey understanding for such voluminous amount of data therefore novel ESDA methods were introduced. These methods uses maps as an interactive interface along with an amalgam of human perception and dynamically linked statistical graphs for allowing the user to manipulate various views of the data [5]. Concisely, this is termed as 'Geo-visualization'. The exploration of geospatial data, presentation of knowledge through findings and unearth knowledge through interaction by well versed scientists an students some of the benefits that has made geo-visualization widely known. On creating visual representations, certain goals of geo-visualization must be addressed. The goals are to unveil the unknown attribute and create new knowledge, interaction with visualization interface and the accessibility to users [6]. Keeping these goals in mind, the proposed work aims at combining the disciplines of geo-visualization, public health and informatics to facilitate better decision making for health policy makers.

Decision-making is an important aspect across all health system for policy development and implementation, governance and regulation, health research and health education and training. For healthy decision-making, sound and reliable information is needed which is provided through health information systems [7]. The main task of Health Information System is to collect data from several health sectors, to analyze it and assure the comprehensive quality and significance for easy transformation of data into information for better decision-making processes [8].

Though, in India health is a state subject, the constitution vests the responsibility of health on both, Central as well as State Governments. The Centre is primarily responsible for provisions while the state is responsible for public health, sanitation and medical care, hospitals and dispensaries. Realizing the importance of sound health information system as a source of concrete, consistent and current information, it acts as a yardstick for health service providers at all stages for eliminating the gaps in the system by initiating appropriate actions based on evidence. This idea was conceptualized in the National Rural Health Mission (NRHM), which was

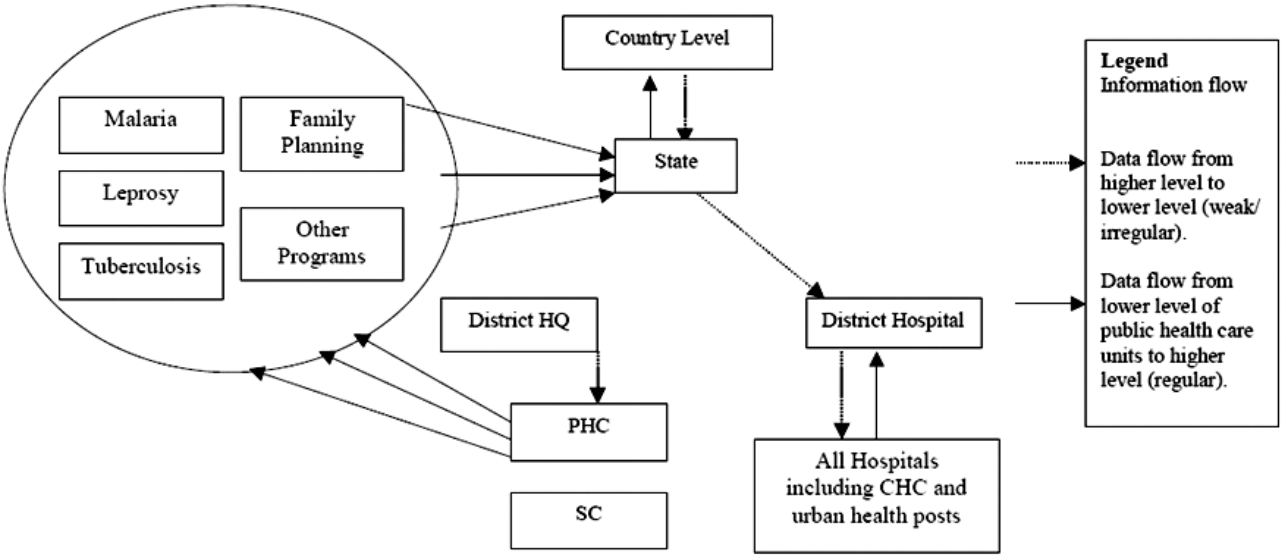

Figure 1: Structure of Information Flow in Public HealthCare System inIndia.

launched by the Government of India in April 2005. NRHM has full authority to achieve impressive advancements in the health systems and health conditions of the people. In addition to several procedures incorporated under NRHM, the basic approach adopted is "Strengthening capacities for data collection, assessment and review for evidence based planning, monitoring and supervision" [8]. In order to give a practical structure to this theoretical approach, the Ministry of Health and Family Welfare, GOI, introduced a Health Management Information System (HMIS) portal for all public health issues. In India, the infrastructure of health facilities is divided into three subdivisions. These are—

1.1. Primary Health Care

Primary health care is concerned with the care of the mother and child. In India, primary health care is provided through a network of Sub Centers (SC) and Primary Health Centers (PHC) in rural areas. These centers act as first level of interaction between individuals and families and the health system. The administration structure at the primary level has 6 Sub centers supervised by a single primary health center. In PHC, Medical Officer and other paramedical staff serves a population of 30000 in the plains and 20000 in hilly, tribal and backward areas while the Sub Center is staffed by Auxiliary Nurse Midwife and Multipurpose Health worker which serves a population of 5000 in plains and 3000 in hilly and tribal area.

1.2. Secondary Health Care

When patients from primary health care need advanced treatment and care they are referred to specialists in better hospitals for treatment. In India, Sub District Hospitals (SDH) and Community Health Centers (CHC) at the block level shapes the secondary health care. Secondary Health Care serves at the second tier of health system.

1.3. Tertiary Health Care

When patients need specialized consultative care, they are referred to District Hospitals by primary and secondary medical care. Specialized Intensive Care Units, cutting edge diagnostic support services and trained medical personnel are some of the fundamental features of tertiary health care. In India, medical colleges and advanced medical research institutes provide tertiary health care services.

Due to recent boost in the use of data in the international public health community, major investments are done in data collection for public health programs for ease of understanding and better visualization. However,

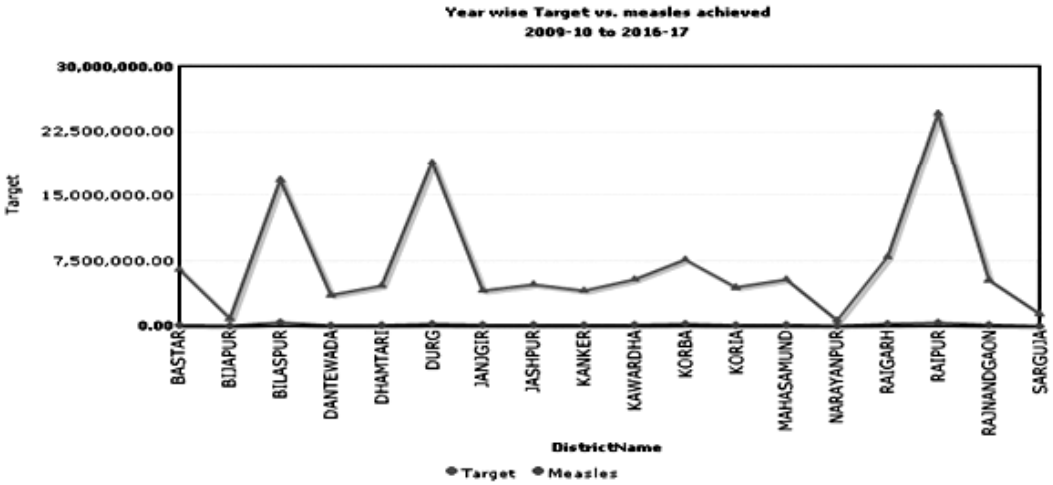

Figure 2: Graphical Representation of Public Health Report on Measles. (www.cghealth.nic.in)

there is concern that data in its raw form and information lacks value unless they are used for making informed decisions. Data use, therefore, is critical to improve the effectiveness and sustainability of the health system. In Chhattisgarh, there are certain challenges such as, poor data quality and low application of accessible data in health sector, data unavailability for 18 districts and unavailability of shape files. At the level of Indian organization, less concentration is given to outcomes and the decisions of resource allotment are based on regular methods that supply little support for evidence-based decision making while at the individual level, dearth of skills to examine and use data are some of the other obstructions. The present HMIS in the state uses traditional line graphs to communicate key findings which fails to provide better understanding in less time as shown in figure 2. Therefore, the current study aims to explore Chhattisgarh's public health facilities data, substantiating public health issues and also convey discoveries and key messages by interactive maps to the target audience, including health authorities, policy makers in health, and the general public to facilitate proportionate distribution of resources. Basically, aim is to create a better visualization and easy to understand graphical representation of current health facilities

2. MATERIALS AND TECHNIQUES

2.1. Site Description

In India in the year 2000, Chhattisgarh was carved out of former Madhya Pradesh as 26th State of Union on the demand of tribal people. The state of Chhattisgarh has an area of 1,35,191 sq. km. , a population of 25.54 million and a population density of 154 per sq. km [9]. It is one of the few landlocked states of the country with 27 districts, 146 blocks, and 20,308 villages. Remote, tribal and extremist affected areas are few classes under which the districts are classified [10]. The state is currently witnessing challenges in ameliorating the health and nutritional status. The surging load of long term non-communicable diseases and ascending elderly population is leading to an epidemiological and demographic transition respectively. The intriguing health indicators, overburdened health systems and ongoing problems of maternal and child mortality and HIV/AIDS pandemic require greater attention to improve public health scenario in Chhattisgarh than it currently is [11].

2.2. Geo-Visualization

Geographic visualization has been in existence for the past years and creates a strong knowledge base of cartographic knowledge on which the computer based visualization flourishes. In accordance with the 2001 research agenda of the International Cartographic Association (ICA) Commission on Visualization and Virtual Environments, the formal definition of geo-visualization is "Geovisualization integrates approaches from visualization in scientific computing (ViSC), cartography, image analysis, information visualization, exploratory data analysis (EDA), and geographic information systems (GISystems) to provide theory, methods and tools for visual exploration, analysis, synthesis, and presentation of geospatial data." [12]. The concept of visualization of spatial phenomena provides a novel technique to convey the results of data collection, data analysis and data simulation [13]. It gives a completely different outlook in the field of research as well as in presenting results and predictions to laymen. In the proposed work, geo-visualization is used for examining epidemiological data and unveiling patterns by human cognition that may be cloaked in textual and tabular form view [14].

2.3. Technologies Used

In the field of Indian public health research, Geographic Information Systems (GIS) are found to be potential communication tools. However, this technology being complex in nature requires expertise in skills to operate and understand the working. Meanwhile, the people has grown interest in geospatial information. Therefore, the need to access geospatial knowledge by a global audience with precise GIS knowledge and less cost was fulfilled by web-based GIS using Scalar Vector Graphics (SVG) technology. Scalability, code readability, compressibility,

open standard and inclusion of animation are some of the benefits of SVG which are tapped in this work by giving interactive maps to larger audience through Internet [15]. The technologies used in the proposed are—

2.3.1. Tableau

It produces interactive data visualization products focused on business intelligence [16]. On inquiring the relational databases, cloud databases, and spreadsheets, the software createscharts and graphs in accordance to the nature of the data. Tableau even facilitates customized mapping by plotting latitude and longitude coordinates. But the software has been denounced for being US-centric.

2.3.2. D3.js

D3.js stands for Data-Driven Documents which is a JavaScript library for transforming documents based on data. The library gives great control over visual outcomes by generating transitional, dynamic and powerful data visualizations in web browsers [17]. To showcase the data as interactive, data is bound to DOM elements and wide use of HTML5, SVG and CSS standards is adopted.

2.3.3. ESRI Shape File

The ESRI shapefile stores geospatial vector data, along with nontopological geometry and details of variable for spatial properties in a data set. The shapefile format give base to points, lines, and areas by representing as closed loops and double digitized polygons. This format is open and popular for data transfer since it can be manipulated by computer needs according to requirement of the project [18].

3. MATERIALS AND TECHNIQUES

In the proposed study, an algorithm is followed in order to create visualization in order to identify interesting and unusual trends in data, to detect spatial outliers and for formulating hypothesis from the data. The steps are:-

3.1. Step 1: Choosing the Public Health Issue

The area of public health has several issues which need to be addressed as soon as possible; therefore a careful study of all prevailing problems needs to be conducted. Since, the focus is public health of Chhattisgarh and the first stage in any public health issue are the health centers availability, therefore, this was primarily chosen for exploration. Once the research problem is chosen, it is necessary to communicate with people who have expertise in the domain of the problem. For the proposed work, the authors approached the DataMeet community to understand the concepts of Open Data, certain visualization companies for the choice of visualization tools and consultants who work on health policy making to identify issues of utmost importance. After careful study and consideration of several resources, the chosen issue is to perform spatial data analysis of the distribution of hospitals and health centers in districts of Chhattisgarh.

3.2. Step 2: Identification of Data Sources

A reliable dataset is the primary need of an insightful visualization. Several reports are available on the web specifying the indicators and summarized tables and textual data regarding several aspects of public health. But these resources do not seem to be reliable. In India, health care is a vital sector that needs to be addressed cautiously. Keeping this in context, Government of India introduced a flagship program known as National Rural Health Mission (NRHM) in April 2005. Under this program, Health Management Information System (HMIS) was set up to enable the flow of information from the District level to the state HQ and the Centre. The online portal consolidates several mechanisms that facilitate monitoring and evaluation systems through performance statistics, surveys, community monitoring and quality assurance. Thus, the portal provides a database

of spreadsheets with information about the number of hospital and health centers at each district in each state. The detailed dataset for the study is taken from the Health Management Information System (HMIS) of National Rural Health Mission (NRHM).

3.3. Step 3: Selection, Application and Analysis of Analytic Method

The spreadsheets available on the online portal needs data formatting to be done for better analysis and visualization according to the chosen public health issue. The data in its raw form may not be able to yield good insights, therefore application of certain statistical methods could simplify the complexities of data. To facilitate better understanding and comparison, ratio of the number of hospitals to population is taken into consideration and the other measure as ratio of the number of hospitals to the area is taken. After the application of analytic methods, it was observed that ratio of the number of hospitals by area denotes a completely different scenario. This variance was due to large forest cover of 44.8% in Chhattisgarh. Therefore, ratio of number of hospitals by population is taken as the measure for comparison for all districts of the state. This measure denotes that how many hospitals serve one person of that district.

3.4. Step 4: Exploratory Data Analysis Using Tableau

Exploratory data analysis (EDA) is considered to be a vital step which is preceded by recognizing the features and acquiring data and it should be done before any modeling [19]. It is crucial for anyone to comprehend the nature not on the basis of assumptions but evidences. The goal of EDA is to find hints about the disposition of the data, its quality to devise hypothesis of the examination done to give better decisions by the application of brief statistics and visualizations. In EDA, certain assumptions are made on the basis of exploratory visualizations and then build some models. The results of the model are further visualized and are tuned according to it. This establishes the notion of EDA being iterative in nature [20]. In this work, Tableau Software is used for exploratory data analysis.

3.5. Step 5: Dynamic and Interactive Visualization -Presentation of Designs and Findings

Once the basic essence of data is identified by exploratory data analysis, dynamic and interactive approaches are employed for data analysis. The utilization of innate capabilities of human brain to observe patterns and trends and thereby gain better insights than it is possible with static and rigid displays is motivation for adopting such approach [21]. It enables the user to discover patterns by allowing transformation of data points in the form of immediate change in the form of selection, deletion and rotation [22]. In the proposed work, the final visualization design is interactive in nature and is built in D3.js which is a JavaScript library used for making interactive data visualization. The final design aims to remove the loopholes found in exploratory data analysis by employing better techniques. The purpose of the final design is to provide better user experience, clear insights of the data and easily formulate conclusions.

3.6. Step 6: Usability Testing and Deployment

Usability of a product is defined as "if a product responds in a way the user expects it to behave without any obstacles and hesitation". The process of usability testing engages people as testing participants who represents target audience to assess the degree to which the usability goals are achieved [23]. In the field of data visualization, at the center of any successful visualization is the ease of learning, ease of use, user satisfaction and minimal user frustration [24]. Therefore, the usability professionals focus on researching the best usability testing methodologies and practices to provide end users easily understandable visualization design regardless of their technical expertise. In order to achieve usability goals, the visualization went through several changes based on the user studies done on professors, government officials and health policy makers [25]. After usability testing, the final visualization design is deployed at public health authorities, government offices as

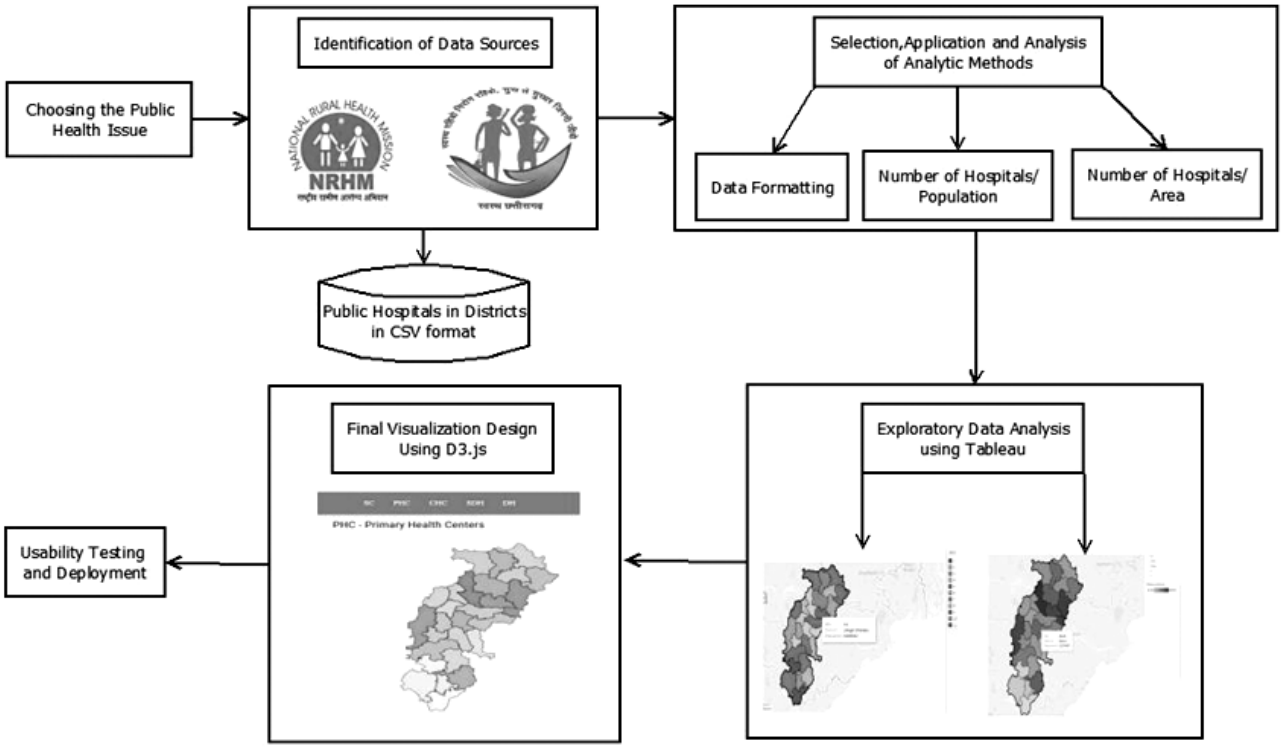

Figure 3: Proposed Methodology block diagram

well as to the health policy makers for understanding whether the visualization cater to their needs and visibility in their decisions.

4. RESULTS

The application of the above algorithm has offered many novel prospects for geographical visualization. The amalgamation of geo-visualization and public health facilities data opens up new avenues for better understanding, interpretation and assessment of textual data managed by government organizations. The use of geo-visualization can represent a powerful technique that can be of strategic importance to public health research. The creation of spatial visual representations or we could say interactive maps tries to achieve certain goals. The goals are to unveil the unknowns, forming new knowledge, the interaction with visualization interface and the visualization users. The proposed work aims to achieve these goals by designing an interactive map interface to access and explore geospatial data. The public health facilities data is conjoined with geospatial dataset to delineate the boundaries of regions for generating the map. Since, design is an iterative process the visualization map of public health facilities require lot of iterations. The three iterations are—

4.1. First Visualization Design Iteration-

The first map is made in Tableau. In this map, each number was assigned a different color which was envisaged on the geographic map as shown in Figure 5. It has a legend at the side of the map to read the visualization and on hovering the mouse on the districts tooltips pop up showing the Number of Hospitals, Population and Name of the District. But there were several issues in the design, such as, along with the map of Chhattisgarh there were maps of Jharkhand, West Bengal and Orissa (other Indian States) visible in the visualization which may deviate the attention of the user. Secondly, color hue is a visual encoding and it is preferred that not more than 3 hues of color are used. In this map, there are 11 colors which is difficult to remember the mapping of color and

number of health center, therefore, every time we need to look for the legend. Thirdly visualization takes a lot of time in loading because in Tableau all the data is stored on the server. The visualization does not give a room to compare among other types of health centers.

4.2. Second Visualization Design Iteration

The second map made in Tableau solves the major issue of the first one by reducing eleven number of color hues to one color. The map uses different saturations of a single color, such type of map is known as choropleth map as shown in Figure 6. A choropleth map is a thematic map in which shading, patterning, coloring and placing of symbols within areas is done in relation to the measurements of the statistical variable. The choropleth map gives an elementary way to visualize the variation of values over a geographical area which depicts the trends across displayed location. It also has a legend at the side of the map to read the visualization and on hovering the mouse on the districts tooltips pop up showing the Number of Hospitals, Population and Name of the District for better exploration. Still there are certain issues with this design, such as, the map of other states along with

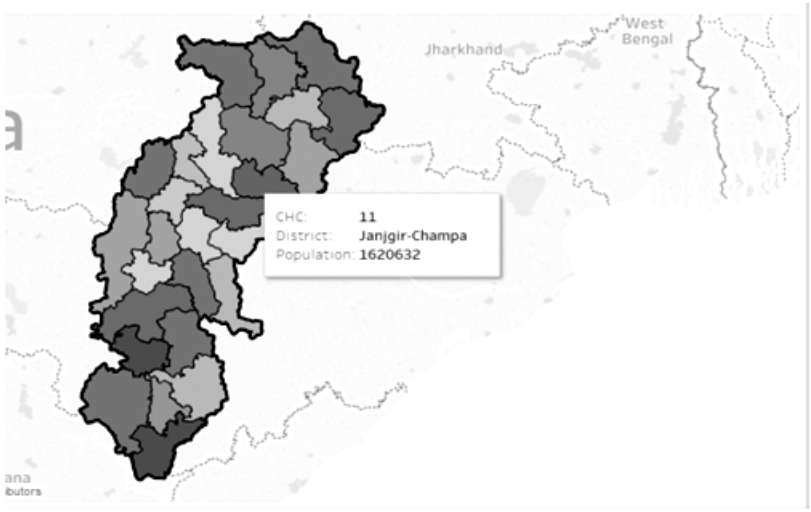

Figure 4: First Visualization Design Iteration-Assigning different colors to different number of hospitals

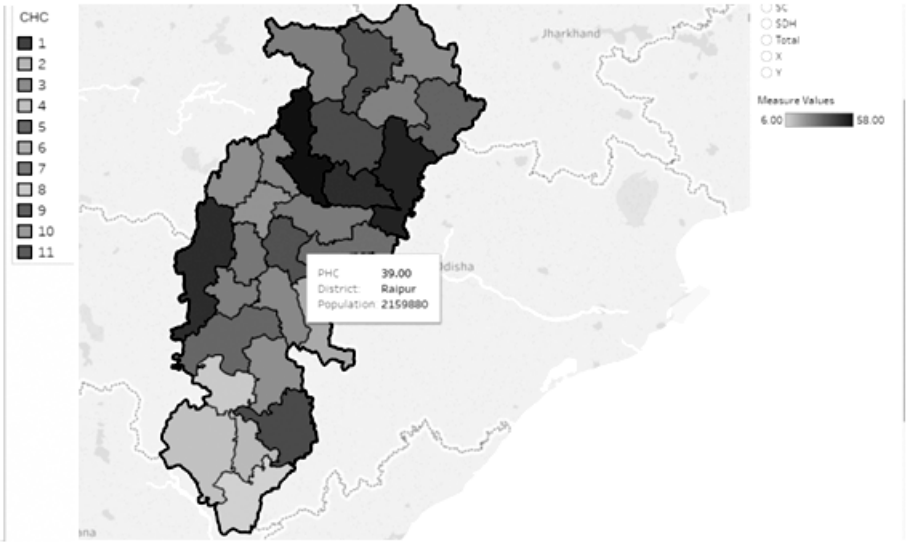

Figure 5: Second Design Iteration-Different Saturation levels of single color is use

Chhattisgarh and the use of Tableau. The major issue is that it fails to achieve the major goal of visualization that is comparison, the user fails to compare among different health centers on a single map.

4.3. Third Visualization Design Iteration

The third iteration of geo-visualization is web based and coded in D3.js. As seen in Figure 7, all the issues of previous versions of visualization design are solved. A menu is provided to select the type of health center with variations on a single map. Since it is web based, it takes less time to load and D3.js handles the data and give insightful visualizations. The map is built using the shape file of Chhattisgarh which is in JSON format so that there is no other states visible except the state of Chhattisgarh. A choropleth map facilitates easy understanding of the map by visual perception. It also has a legend at the side of the map to read the visualization and for better exploration tooltips pop up showing the Number of Hospitals, Population and Name of the District on mouse hover.

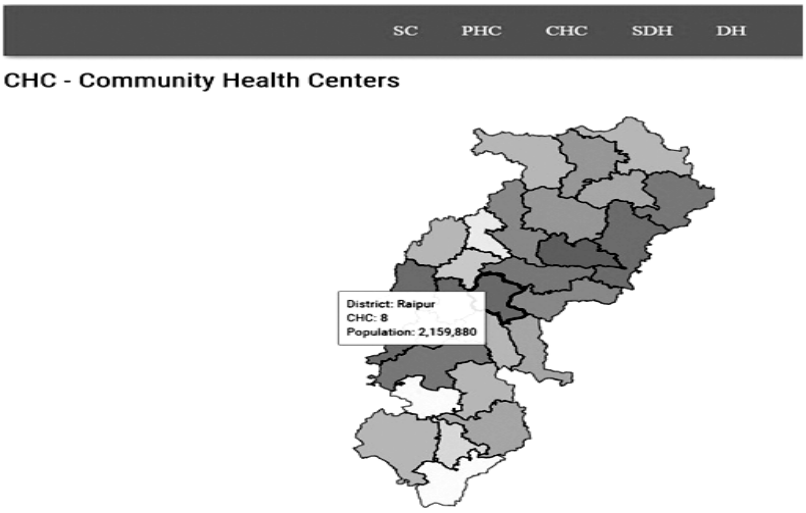

Figure 6: Third Visualization Design of Community Health Centers of Chhattisgarh

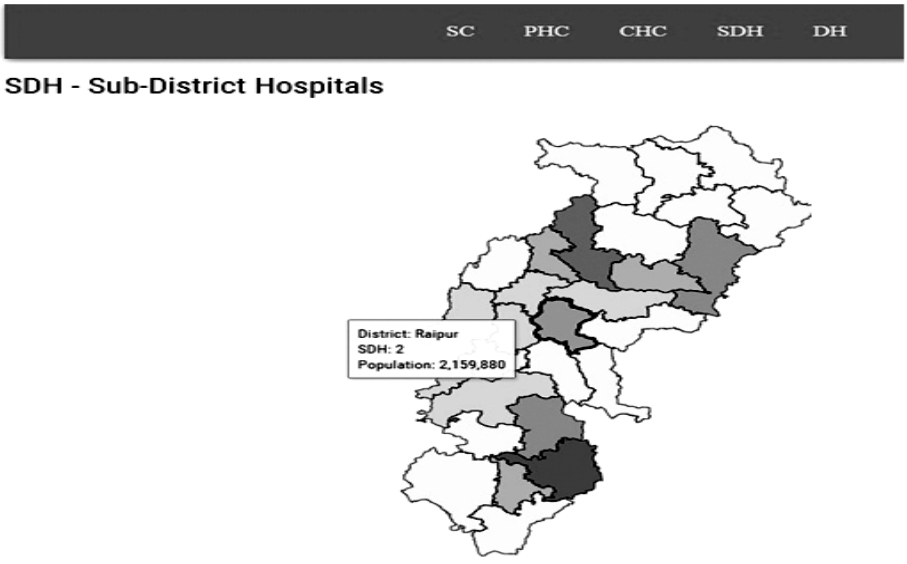

Figure 7: Third Visualization Design of Sub District Hospitals of Chhattisgarh

5. CONCLUSION

In this way, the visualization of public health facilities data aims to solve the problems of health policy makers, public health strategy makers, government agencies and public health authorities to make better decisions for evidence based planning and equitable distribution of the scarce resources in India. The visualization map is designed as a web page to provide a wider audience with a yardstick to investigate public health issues in geographical regions of their interest. Some of the significant characteristics of the work are to aid in examining the data at the granular level of administration, i.e., districts. The visualization map is not only manipulated dynamically by selection but also envisages data simultaneously on maps, charts and tables simplifying the comparison of multiple geographical regions in accordance with the variables. The tool supports in unveiling significant patterns and valuable insights. The web based geospatial system has not only addressed the rise of Internet but also supported the transition from technology driven visualization to more human centered methods. This model is an incentive to build complex implementations that would require more demographic outcomes and variables as well as rich geographical region selection features. The proposed methodology can bring valuable answers to different research fields such as epidemiology, criminology, economy, archaeology, wildlife biology etc. The steps introduced can act as yardstick in making visualization that find applications not only in the area of public health but a multitude of fields.

REFERENCES

[1] Anselin, Luc. "Interactive techniques and exploratory spatial data analysis." Geographical Information Systems: Principles, Techniques, Management and Applications, eds., P. Longley, M. Goodchild, D. Maguire, and D. Rhind. Cambridge: Geoinformation Int (1999).

[2] Jebb, Andrew T., Scott Parrigon, and Sang Woo. "Exploratory data analysis as a foundation of inductive research." Human Resource Management Review (2016).

[3] Sim, Sunhui, and Andy Miller. "Exploratory Spatial Data Analysis of the Distribution of Multiple Crimes: A Case Study of Three Coastal Cities." Papers in Applied Geography 2.1 (2016): 79-84.

[4] Haining, Robert, Stephen Wise, and Jingsheng Ma. "Exploratory spatial data analysis." Journal of the Royal Statistical Society: Series D (The Statistician) 47.3 (1998): 457-469.

[5] Kwan, Mei-Po, and Jiyeong Lee. "Geovisualization of human activity patterns using 3D GIS: a time-geographic approach." Spatially integrated social science 27 (2004).

[6] Kraak, Menno-Jan. "The space-time cube revisited from a geovisualization perspective." Proc. 21st International Cartographic Conference. 2003.

[7] Hafner, Tamara, and Jeremy Shiffman. "The emergence of global attention to health systems strengthening." Health policy and planning 28.1 (2013): 41-50.

[8] World Health Organization. 2008 Report on the Global AIDS Epidemic. World Health Organization, 2009.

[9] Census of India, 2011, Registrar General of India, Governmentof India, 2011.

[10] Available from: http://www.cg.gov.in/Archive. [Last accessed on 2016 Dec 10].

[11] Galhotra, Abhiruchi, et al. "Mapping the health indicators of Chhattisgarh: A public health perspective." Population 65.62.2 (2014): 82-7.

[12] Nöllenburg, Martin. "Geographic visualization." Human-centered visualization environments. Springer Berlin Heidelberg, 2007. 257-294.

[13] Yasobant, Sandul, et al. "Geovisualization: a newer GIS technology for implementation research in health." Journal of Geographic Information System 7.1 (2015): 20.

[14] Forsythe, K. Wayne, et al. "Using geovisualization to assess lead sediment contamination in Lake St. Clair." The Canadian Geographer/Le Géographe canadien (2016).

[15] Kamadjeu, Raoul, and Herman Tolentino. "Web-based public health geographic information systems for resources-constrained environment using scalable vector graphics technology: a proof of concept applied to the expanded program on immunization data." International Journal of Health Geographics 5.1 (2006): 1.

[16] Chabot, Christian, Chris Stolte, and Pat Hanrahan. "Tableau Software." Tableau Software (2003).

[17] Bao, Fan, and Jia Chen. "Visual framework for big data in d3. js." Electronics, Computer and Applications, 2014 IEEE Workshop on. IEEE, 2014.

[18] Anselin, Luc, Ibnu Syabri, and Youngihn Kho. "GeoDa: an introduction to spatial data analysis." Geographical analysis 38.1 (2006): 5-22.

[19] Alspaugh, Sara, et al. "Building blocks for exploratory data analysis tools." Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics. ACM, 2013.

[20] Bailey, Trevor C., and Anthony C. Gatrell. Interactive spatial data analysis. Vol. 413. Essex: Longman Scientific & Technical, 1995.

[21] Swayne, Deborah F., Dianne Cook, and Andreas Buja. "XGobi: Interactive dynamic data visualization in the X Window System." Journal of Computational and Graphical Statistics 7.1 (1998): 113-130.

[22] Sopan, Awalin, et al. "Community Health Map: A geospatial and multivariate data visualization tool for public health datasets." Government Information Quarterly 29.2 (2012): 223-234.

[23] Battleson, Brenda, Austin Booth, and Jane Weintrop. "Usability testing of an academic library web site: a case study." The Journal of Academic Librarianship 27.3 (2001): 188-198.

[24] Hertzum, Morten. "Usability Testing: Too Early? Too Much Talking? Too Many Problems?." Journal of Usability Studies 11.3 (2016): 83-88.

[25] Lamarr, Michael, et al. "How Usability Testing Resulted in Improvements to Ground Collision Software for General Aviation: Improved Ground Collision Avoidance System (IGCAS)." (2016).