

Need for Interactive Data Visualization in Public Health Practice: Examples from India

Abstract

The world is full of data which is increasing by leaps and bounds. In health care, big data is becoming common with increased electronic health data accumulation and/or accessibility to public data previously held under lock and key. At the same time, health data visualization applications have become popular over recent years. Against this background, a review was done to summarize the application of data visualization in public health & the challenges faced. Peer-reviewed original research articles and review articles searched in Google Scholar and Pubmed databases that were indexed in the last ten years period, using the keywords “Big data” or “data visualization” or “Interactive visualization techniques.” Other related information in books, blogs, and published documents were searched in Google search engine using the same keywords. Contents from the downloaded documents were presented and discussed under three headings viz. (a) the visualizations that are still current and how they have evolved further, (b) tools or methods that can be used by end-users to make their own modifications, (c) the platforms to disseminate them. Usage of different plots in public health is explained with suitable examples using the data from public health datasets. From the discussion it can be understood that when big data is visualized well, it can identify implementation gaps and disparities and accelerate implementation strategies to reach the population groups in most need for interventions. As health administrator may come from diverse specialties, robust training and career development for big data in public health is the need of the hour.

Keywords: *Big data, data display, health information systems, public health informatics, visualization*

Introduction

The world is full of data which is increasing by leaps and bounds. From climate change to space science to economics, big data has become important not only to various groups of professionals but also to individuals and communities. Big data is useful in understanding the situation, identification of underlying problems, planning to overcome those problems, and evaluation of implemented plans in each field. At the individual level, big data is creating awareness about different topics and giving readymade solutions to the activities in their daily life. The health sector too has seen phenomenal growth in data gathered from millions of people, each uniquely identifiable or anonymized and pooled together. When properly organized and analyzed data uncovers unseen problems, relationships, trends, and causes that lead to solutions. It empowers professionals such as researchers,

administrators, policymakers, public health agencies, and clinicians to make better decisions.

It is difficult to define “Big Data”. It cannot be defined in terms of gigabytes or terabytes as data stores are constantly growing and so is computing power and what seems a lot of data now may appear perfectly normal in a year or two.^[1] Big data have been characterized in terms of volume, velocity, and variety. These three characteristics cause the many challenges that organizations face in handling data. The most important among them are veracity, variability, and value. Veracity refers to the data quality i.e., dealing with the data growth and validating the data and data value i.e., generating insights in a timely manner. Variability is linked to integration of disparate data sources. Added to these are the resistance of stakeholders to change and the expertise required to handle data.^[1,2]

Summarising it and making meaningful interpretation is another challenge with data,

How to cite this article: Narayan KA, Nayak MS. Need for interactive data visualization in public health practice: Examples from India. *Int J Prev Med* 2021;12:16.

**K. A. Narayan,
M. Siva Durga
Prasad Nayak¹**

*Department of Community
Medicine, Mahatma Gandhi
Medical College and Research
Institute, Pondicherry,
Department of Siddhartha
Medical College, Vijayawada,
Andhra Pradesh, India*

Address for correspondence:

*Dr. M. Siva Durga Prasad
Nayak,
Flat no 503, Satya towers,
10th line, Bharathi Nagar,
Vijayawada, Krishna District,
Andhra Pradesh 520007, India.
E-mail: sivadurgaprasadnayak@
gmail.com*

Access this article online

Website:
www.ijpvmjournal.net/www.ijpvm.ir

DOI:
10.4103/ijpvm.IJPVM_171_20

Quick Response Code:



This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

especially with big data. Therefore, a seventh V of big data, i.e., Visualization could be added to the above. Quantifiable imagery and numbers have served, side by side, in basic graphic forms like the table, coordinate system, and map, and in derivative forms like the line graph, histogram, and scatterplot.^[3] However, statisticians predominantly use them to validate the data rather than summarise and convey information. Scientific journals reiterate the importance of quantitative graphics whether it was Sir Edmund Halley's graphical analysis of Barometric pressure as a function of altitude published in 1686 or the latest advertisements for computer graphic technology.^[3] Academic statisticians view graphical methods as a diversion from "serious analytic techniques". Outside of statistics, though, infographics and data visualization are more important. An idea of the size of the data, relations between variables and the complexity of data and functional relationships can be obtained by graphics.^[4]

Information visualization uses data graphics more generally as a means of communication. In the field of public health, one of the most powerful visualizations has been Dr. John Snow's map of the Cholera Outbreak in London of 1854.^[5] But his work as proof of cholera being a waterborne disease did not satisfy most of his contemporaries.^[6] Probably less known visualizations are H.W. Acland's chart containing multiple histograms, also released in 1854, of cholera and diarrhoea cases in relation to weather conditions such as high and low temperatures, rain, atmospheric pressure, and wind^[7,8] and Florence Nightingale's famous polar bar diagram visualization of 1858, of deaths in the Crimean War which brought about the modern army hospital system.^[4,9]

In healthcare, big data is becoming common with increased electronic health data accumulation or accessibility to public data, previously held under lock and key.^[10,11] With an increase in computing power and availability of data visualization software to end-users, health data visualization applications have become popular over recent years; the ultimate example being of Dr. Hans Rosling who brought "data to life" with the Gapminder software which helps people search and explore global public data.^[12] Against this background, the goal of this paper is to summarize the developments in data visualization, application for public health and the challenges faced.

Methods

The current study is a systemic review article discussing different data visualization techniques that can be used in the public health sector. Peer-reviewed original research article and review articles were searched using google scholar and PubMed databases using the keywords "Big data" or "data visualization" or "Interactive visualization techniques". Articles indexed in the last ten years, i.e., from 2010 to 2020 were included in the study. Other related information in books, blogs, and other related published

documents were searched in google search engine using the same keywords. Collected articles and other information were reviewed to discuss the development of quantitative graphics key concepts in the modern-day visualization. Some examples of visualization using the public health data sets of India are mentioned in the article. Examples are drawn from the analysis of public health data sets by the authors.

This work contributes to various modern and interactive data visualization techniques for public health practice. In this paper, the development of quantitative graphics and their usage in modern days is presented under three headings viz., A) the visualizations that are still current and how they have evolved further. B) Tools or methods that can be used by end-users to make their own modifications and C) the platforms to disseminate them.

A) The visualizations that are still current and how they have evolved further: Visualization techniques or quantitative graphics developed from time to time, based on the need and innovative efforts of different scientists since the 16th century. Time series charts, Pie charts, and bar diagrams continue to be the most popular data visualizations since the olden days.^[13] The bar chart has seen several practical variations in the course of time such as percentage bar, three-dimensional bar, horizontal bar, side by sidebar, negative-positive bar, etc.^[14]

During the process of development of quantitative graphics, five key concepts evolved in modern-day visualization.^[15] The first is that it is computer-based, the second is using visual representations and the third representing underlying data, which is mainly an abstraction rather than a true representation such as the image of a skull in X-ray. The challenge, therefore, is to choose the appropriate visualization. The fourth concept is that data could be made interactive. Especially with large data the user could interact with the data to control or modify the visualization so that they can change not only what is visualized but also how it is done. The last concept is that the goal is to amplify cognition. This means the end-user can solve problems better by observing and understanding invisible patterns of data when it is tabular form. For this, we can use "cognitive artefacts" – tools such as pivot tables, filters, and slicers, through which the end-user can better understand the visualization and make their own interpretations.^[16] The following paragraphs describe cognitive artifacts and some other interactive visualization techniques with examples.

Data for these examples have been drawn from public domain. Data of dengue fever in Kerala state was collected from the website of the Directorate of public health of Kerala state.^[17] Data on maternal services in Andhra Pradesh was collected from the National Rural Health Mission Health Management Information

System (NRHM HMIS) reports.^[18] All the data files and source codes used to generate different graphs are uploaded in the github.com^[19]

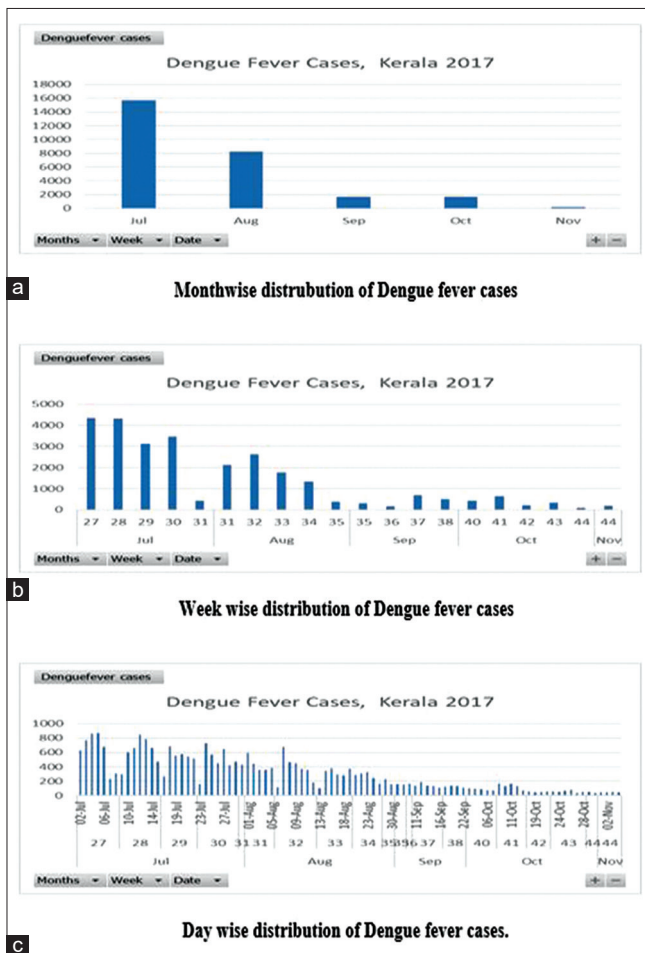
i) The "Drill Down" or "Scroll and Zoom" plots: Computer programs such as Microsoft Excel have a feature of "pivot table", with an option to "Drill Down/Drill" up feature, that summarises the extensive data from a database

Bar diagrams were plotted using the data downloaded from the website of the directorate of public health services, Kerala, with Microsoft Excel software.^[17,19] Distribution of dengue cases by month, week, and day in Kerala were depicted as Figure 1a-c. Panel a and b clearly show an epidemic pattern of dengue. However, panel c shows a peculiar pattern of occurrence not in consonance with epidemiological trends. On closer examination, the troughs correspond with Sundays or closed holidays. This could reflect either a lack of treatment facilities during weekends and/or a weakness in the reporting system. Such visualizations would help administrators to make decisions. Scroll and Zoom plots can also be used for this purpose. Handles in

the chart enable the end-user to scroll the graph in a given time series and to zoom and see the in-depth changes in a shorter time period

ii) The "slicer" function: The "slicer" functions are visual filters. Most spreadsheets like Microsoft Excel, Calc in Open office and Google Spreadsheets provide the "Slicer" function. The biggest advantage is that Slicers can be linked to multiple tables or graphs. However, Slicers are useful when the number of options are limited. As an example line diagrams were created and depicted as Figure 2a and b using the same dengue fever incidence data of Kerala state.^[17,19] Figure 2 shows the incidence of dengue in Kerala for July to November. Three "Slicer" filters have been added namely district, month, and week

iii) Doughnut chart: The advantage of the doughnut chart over the pie diagram is that it can depict more than one series for comparison in one diagram instead of multiple pies. Figure 3 shows a variation of the doughnut diagram. It was created using the data downloaded from NRHM HMIS annual reports using Microsoft Excel software.^[18,19] The diagram presented as a dashboard using slicers shows the percentage coverage of several maternal health indicators over 5 years and by district. The diagram has been made dynamic so that the end-user can select an



Figures 1: (a-c) Distribution of dengue cases by month, week and day in Kerala

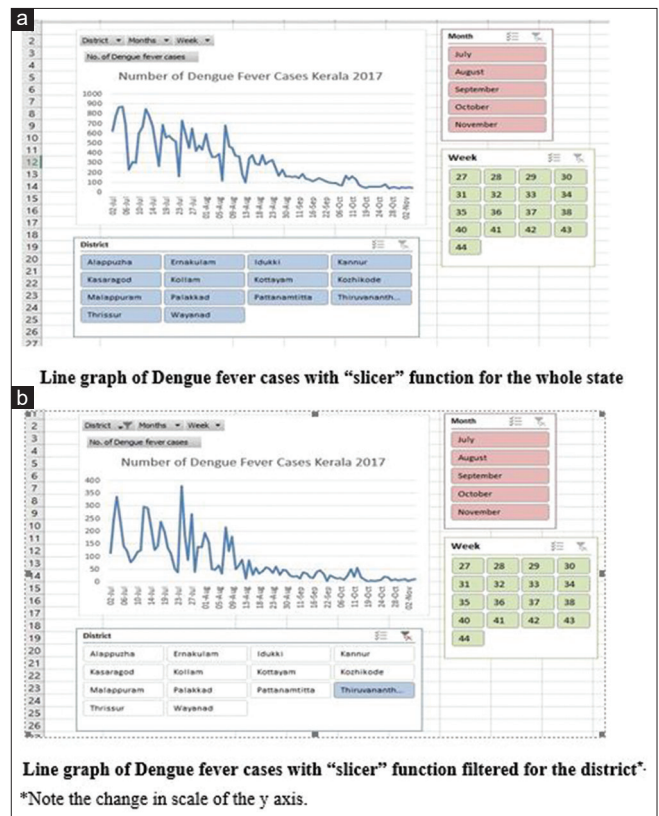


Figure 2: Line graph of Dengue fever cases with "slicer" function. (a) for the whole state (b) filtered for the district*
*Note the change in scale of the y axis.

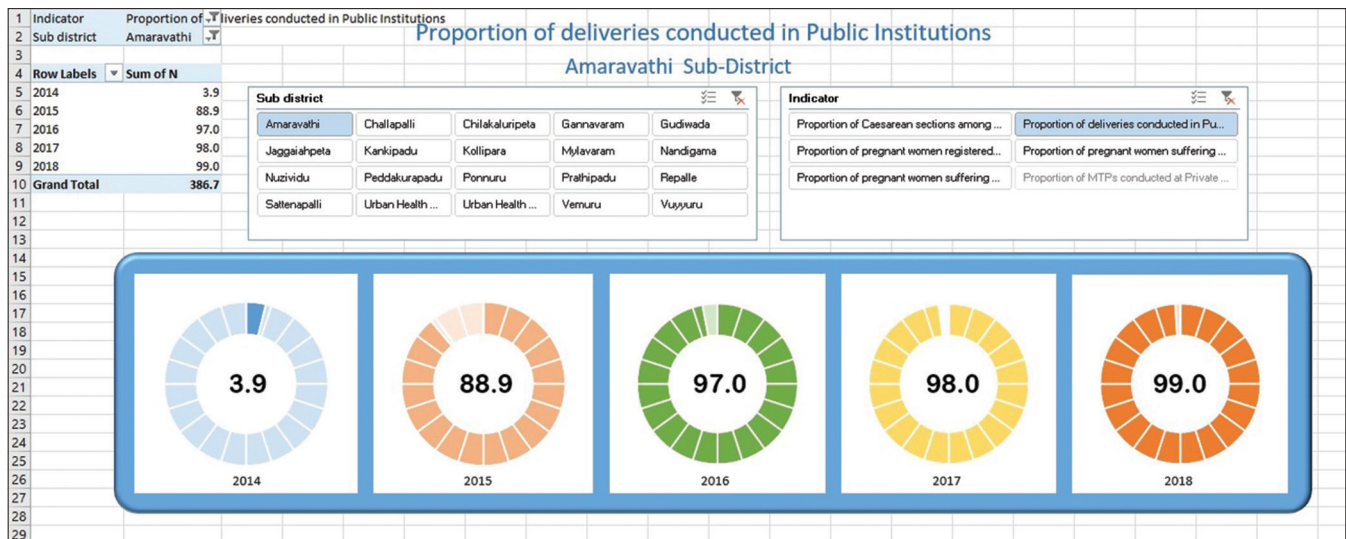


Figure 3: Doughnut chart of proportion of Maternal Care Indices by year - Andhra Pradesh

indicator for a given sub-district. The main difference between basic data visualization and dashboards is its dynamic nature. Data visualizations will be just generated from data sets whereas dashboards are regularly updated according to dataset modifications^[20]

- iv) **Bubble Chart:** A variation of the scatter plot is the bubble chart, which replaces the data point by bubbles. A third variable determines the size of the bubble. Figure 4 depicts the bubble chart of incidence of Dengue by altitude and population density in Kerala state, which is created by triangulating the downloaded data^[17,19]
- v) **Motion Chart or Animated Bubble Chart:** Popularised by Hans Rosling, motion charts are essentially animated bubble charts with changes over time being depicted by the movement of the bubbles in a two-dimensional space and by size or colour. The motion chart is now available as Google Visualization
- vi) **Matrix Plot:** Matrix plots are used to assess the relationships between several pairs of variables at once. Figure 5 depicts the matrix plot showing the crude birth rate in different subdistricts of Amaravathi capital region of Andhra Pradesh state from the year 2014 to 2018, which was created using NRHM HMIS annual reports.^[18,19] Cells filled with red colour had values less than -2SD or more than 2SD, cells filled with yellow colour had values in between -1SD to -2SD and 1SD to 2SD. Cells filled with green colour had values -1SD to 1SD
- vii) **Choropleth Map:** Choropleth Maps display divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. Figure 6 depicts the spatial representation of the proportion of diabetes patients among all outpatients in the Amaravathi capital region of Andhra Pradesh during the period April 2018 to March 2019, which was created using QGIS software^[18,19]

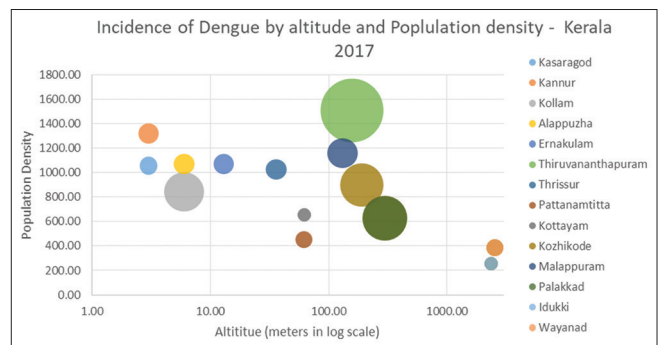


Figure 4: Incidence of Dengue by altitude and population density in Kerala state

- viii) **Spiral Plot:** A common question in time series data is whether the data is periodic and, if yes, what the period is, which can be answered by a spiral plot. A spiral plot of daily incidence of dengue fever was created using R software and depicted in Figure 7.^[17,19] It shows an example of daily incidence of dengue fever has an interesting periodic pattern: in 28 days, there are four periods, which means that there is a weekly pattern: dengue fever incidence is zero in many Sundays, indicating poor surveillance system on Sundays
- ix) **Predictive interactive graph:** Figure 8a and b depict the observed and forecasted number of dengue fever cases with confidence limits from January 2007 to May 2019. Forecasted values were calculated using SPSS trial version 21 using the data collected from the website of the directorate of public health services, Kerala.^[17,19] The first graph used the actual number of dengue fever cases up to December 2017 and the second graph used the data up to December 2018. These graphs can be plotted automatically in the dashboards and made interactive and are useful to program officers for better planning. In graph a, The prediction for subsequent

| Urban Health Equity Assessment and Response Tool | | | | | | | | | | | | | | | | | | | | |
|--|------------------|----------|-----------|---------|---------|--------|----------|-----------|----------|----------|---------|---------|---------|-----------|---------|----------|----------|----------|--------|---------|
| Select indicator from drop | Crude Birth rate | | | | | | | | | | | | | | | | | | | |
| Sum of N | Sub districts | | | | | | | | | | | | | | | | | | | |
| Year | Amaravathi | Challapa | Chilakala | Gannava | Gudivad | Jaggai | Kankipac | Kollipara | Mylavara | Nandigar | Nuzvidu | Peddaku | Ponnuru | Prathipac | Repalle | Sattenap | Urban He | Urban He | Vemuru | Voyyuru |
| 2014 | 24.44 | 2.52 | 26.06 | 6.63 | 8.51 | 4.14 | 2.23 | 17.05 | 2.31 | 4.72 | 16.94 | 22.08 | 18.95 | 15.12 | 16.37 | 15.97 | 1.71 | 2.49 | 15.77 | 5.41 |
| 2015 | 1.02 | 1.13 | 3.43 | 0.87 | 4.05 | 1.77 | 25.64 | 5.75 | 2.18 | 1.73 | 8.23 | 1.15 | 0.63 | 29.44 | 5.16 | 8.25 | 0.05 | 0.36 | 1.30 | 1.97 |
| 2016 | 0.85 | 0.71 | 1.93 | 0.97 | 4.75 | 2.14 | 28.81 | 7.35 | 2.15 | 3.95 | 8.53 | 0.70 | 0.50 | 25.99 | 4.66 | 10.51 | 0.74 | 0.33 | 1.08 | 1.58 |
| 2017 | 1.04 | 0.67 | 2.52 | 1.27 | 3.82 | 1.15 | 28.41 | 7.92 | 1.62 | 2.86 | 8.82 | 0.91 | 0.86 | 23.05 | 5.75 | 7.11 | 2.08 | 1.71 | 1.36 | 1.77 |
| 2018 | 0.83 | 0.67 | 1.93 | 1.22 | 3.34 | 0.96 | 29.97 | 9.11 | 1.23 | 2.64 | 8.00 | 0.67 | 1.25 | 28.93 | 5.03 | 7.63 | 1.23 | 1.24 | 1.10 | 1.15 |

Figure 5: Matrix plot showing the Crude birth rate in different sub districts of Amaravathi capital region of Andhra Pradesh

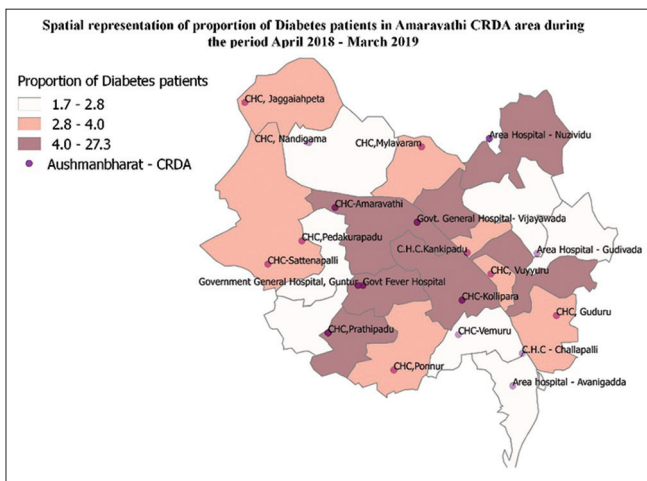


Figure 6: Spatial representation of Diabetes among all OP patients in Amaravathi capital region of Andhra Pradesh during the period April 2018-March 2019



Figure 7: Spiral plot showing the daily incidence of dengue fever in Kerala state from 1st May 2017 to 31st October 2017 in 28 days cycle

years is high. On the addition of actual data for 2018, the prediction is lower, helping administrators in programme evaluation

B) The tools needed to develop Visualization: For most non- technical analysts, the ubiquitous Microsoft

Excel provides great visualizations tools as described above. Other Office Suites such as Libre Office, Open Office also offer the same tools. Statistical software such as SPSS, JASP, and ‘R’ are extensively used for statistical computing. These software offer interactive visualizations systems. If the data is distributed in space and time such as disease surveillance data in public health, these methods would require significant processing time that in turn depends upon the hardware and software used. Academicians and public health analysts are at a disadvantage because they may not have the financial resources for expensive business intelligence “point and click” tools, nor the expertise to do the required coding.^[21,22] Microsoft Power BI and Tableau are the examples for these type of tools. However open-source software, such as Lucky Orange, Pentaho, Elastic search, Anaconda, do help in organizing and analyzing data. Hadoop is an open-source data processing platform for big data analytics in healthcare. Python, known as the “Swiss Army Knife” of programming languages, is powerful scientific computing and visualization framework. Among many things, it supports machine learning, web development, web scraping, desktop applications, etc.^[22] The R package has fantastic visualization and professional document generation capabilities. The Shiny package for R, allows R programmers to interactively show the output for R programs to Web browsers. By using widgets such as sliders and radio buttons it updates output in response to changing inputs by the user.^[22]

C) Platforms to disseminate information interactively: For effective and prompt use of data visualization it is also important that the information generated is disseminated to the stakeholders. In the field of public health, monitoring the movement of outbreaks is of vital importance. But many stakeholders may be involved and each may be managing their own data at various stages of an outbreak. Sharing and collaborating on one platform with various partners is a challenge as, due to limited resources, developing and managing a central platform may not be possible.^[23]

Many cloud-based platforms are now available which permit the generation and hosting of graphics to improve scientific reproducibility by bundling scientific figures with their underlying data, version history, and discussion

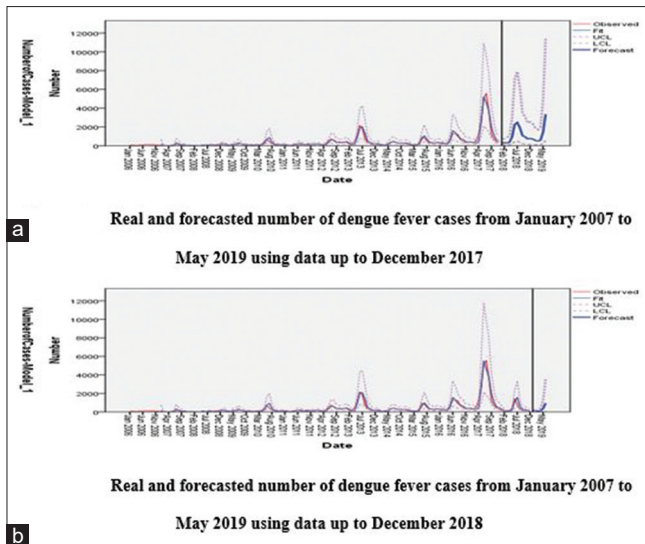


Figure 8: Real and forecasted number of dengue fever cases from January 2007 to May 2019

among collaborators. Many such as ParaView, Gephi, and Weave are available as Open source or free. Some versions of business software such as Tableau Public are available free of cost albeit, with limitations.^[24]

Plotly integrates a wide variety of existing tools and workflows, and is language flexible, allowing coding analysts to use the most popular data science languages and libraries such as R, Python to create, share and edit charts. For example, an interactive chart created in R can be downloaded and edited in Python using Plotly. The government of India has set up the Open Government Data (OGD) Platform India for visualizing data.^[11] The portal, though intended to be used by the Government of India's Ministries/Departments and their organizations, to publish datasets, documents, services, tools, and applications collected by them for public use has, a visualization engine that registered users can use to visualize their own data.^[25]

Discussion

Each visualization technique has its own advantage. Pivot table and Chart with The Drill Down will make the end-user can visualize the data based on the filters. The “drill down” and the “slicer” puts control into the hands of the end-user permitting sub analysis and interpretations. The biggest advantage is that Slicers can be linked to multiple tables or graphs. However, Slicers are useful when the number of options are limited. They have less advanced conditional options than filters. Doughnut chart is a variation on a Pie chart having an extra edge to depict more than one series for comparison in one diagram instead of multiple pies. These Doughnut charts can also be integrated with pivot table options to make it more dynamic. Bubble Charts can be used to depict the variations in health-related states in relation to their determinants. It is easy to understand and many online public health dashboards using it to display

the burden of diseases. Motion chart will add animation to it to understand the changes over a time. Scroll and Zoom charts are also useful to understand the changes in the incidence of disease in the given time.

Colours in the matrix plot will enable the end-user to understand the variation of morbidity and mortality rates or outliers in the given data. It will give a quick understanding of poor performing and good performing geographical units. Choropleth maps are another type of matrix plot in which geographical units are filled with colours based on the given data. It will enable end-users to understand the relation with geographical location and their health outcomes. Choropleth maps can be converted into heat maps by using interpolation analysis techniques in geographical information system tools to identify the actual hot spots. Voronoi polygons can be overlaid on the choropleth maps or heat maps to make assumptions about the possible risk factors that are causing health inequalities in the given area. Spiral Plot can be used to show changes in the health outcomes in a prolonged period in a short place. Specific temporal patterns in health data can be understood with the help of spiral plots. Simple line diagrams can also be converted into interactive predictive graphs, with help of forecasting analysis techniques such as Autoregressive integrated moving averages techniques, to make assumptions about the future burden of the disease in the same geographical area.

Although different software and platforms came into existence for creating these types of visualizations of data and for disseminating information interactively, Information visualization or InfoVis is yet to have clearly defined boundaries. It is not about pretty pictures or statistical graphics. One perfect visualization configuration may not answer a question. The process requires trying out different visualization techniques, settings, filters, and using interaction to probe the data.^[26] Visualization and visual analytics use images, but the images are only one part of the visualization.^[4] Andrew Gelman and Antony Unwin re-plotted Nightingale's data. Though less visually striking, it showed the data and their patterns much more clearly. The trends and departures from trend were clearer when plotted directly as time series. However, Florence Nightingales' graph did the job of attracting attention to the problem, highlighting the need to balance the aesthetics and the underlying data in visualizations.^[4]

Hesse *et al.* envision three domains for health-related data 1) Personal Health Information: data collected by individuals on their own health practices with body monitors e.g., continuous blood glucose monitoring 2) Clinical Health Information with Electronic Health Record (EHR) systems 3) Public Health Information.^[27] Visualization of these data is becoming extremely useful for the first two domains. For example, the FreeStyle Libre Glucose monitoring systems visualization is changing the

way diabetics can be diagnosed and treated.^[28] Interactive visualizations help clinicians and patient safety managers query EHR databases to understand the patterns of use and increase patient safety. For clinical trials dependant on large warehouses of patient data, visualization can reveal data quality problems, support powerful temporal queries or even identify cohorts of patients matching selection criteria.^[26] An important subset of Clinical Health Information is that of Genomic Data.

In the field of public health, data visualization has always been important during outbreaks for interpreting data in order to direct response actions. Several variables may need to be considered. Analysts need to interpret variables and provide decision-makers with contextual, actionable data so that they can coordinate an effective response. Effectively using the advanced feature of cross-filtering, wherein a filter can work across several graphs or tables simultaneously, visualizations have been effectively used in the fight against Ebola.^[21,23]

National health systems are moving away from a focus on disease-specific health responses to the comprehensive strengthening of health systems, Global and national health agendas as a way to improve health outcomes. Strengthening of health systems has become a top priority for countries. The World Health Organization has identified six attributes for a health system of which quality and timely data from health information systems (HIS) are the foundation of the overall system and inform decision making in each of the other five building blocks in the health system.^[29]

In India the Ministry of Health and Family Welfare along with the state governments have established a Health Management Information.^[30] Several analytical reports have been generated for all states and for the country as a whole. Composite indices, scorecards and dashboards have been created for the use of stakeholders. However, these are static visualizations. The focus is on maternal and child health. As a policy, the Government of India is committed to developing information databases on a wide variety of areas that researchers can share. This includes ensuring that all unit data of major publicly funded surveys related to health are available in the public domain in a research-friendly format.^[31] In this scenario, all health professionals must be conversant with quality data and visualization techniques to enable them to make informed decisions.

Health data and information lack value unless they are used to inform decisions. Two surveys have been conducted in India, one by the World Bank (2009) and the other by Monitoring and Evaluation to Assess and Use Results (MEASURE) framework of the United States Agency for International Development (USAID) 2010. The first survey found that data use was below the halfway mark. The second survey saw better utilization of data.

Yet there were several basic issues – a work culture less focused on results, result allocation based on normative practice rather than evidence-based decision making. At the individual level, impediments included lack of skills to analyze and use data, absence of incentives for data use and lack of recognition for performance at the workplace.

The Way forward

Several visualization techniques are now available which can give composites of different visualizations through the use of cross filters, clicking on an element on any panel alters all the visualizations in the dashboard. Social media are opening remarkable possibilities for health and healthcare researchers. Using Network Analysis and Graph Theory for what we do online can be analyzed. Data from Twitter is especially interesting for public health analysts since it is publicly available to enable studies of how medical trends spread. Using clustering active communities of discussions for polio and measles have been identified.^[26,32]

There are several databases in the public domain which offer huge volumes of data. The census, Ministry of Health and Family Welfare. Integrated Disease Surveillance Programme, National Vector Borne Disease Control Programme and the Open Government Data Programme offer such data in India. However, exploring these datasets and correlating between them to derive insights remains a challenge. But, data quality is also dependant on the local level. Therefore, there is a need for the development of analytical and interpretative skills at the lower level of the hierarchy.

The McKinsey report on data analytics identifies five elements for successful data and analytics transformation. (1) Using Cases or Sources of Value 2) an Internal and External Data ecosystem. (3) Modelling Insights by way of Data Modelling Black Box and Heuristic Insights by Smart Boxes (4) Work Flow Integration by Process Redesign and Tech Enablement and (5) Adoption by Capability Building and Change Management.^[33]

Conclusion

In the age of big data, public health impact and implementation needs can be assessed in different ways using different interactive visualization techniques. When visualized well, big data can identify implementation gaps and disparities and accelerate implementation strategies to reach population groups in most need for interventions. For precision public health to succeed, advances in predictive analytics, and practical tools for data integration and visualization are needed.^[34] As Health Administrators may come from diverse specialties and most, including those from public health, will not be well versed in big data science, robust training and career development for big data in public health is the need of the hour.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Received: 06 Apr 20 **Accepted:** 11 Sep 20

Published: 24 Feb 21

References

1. Harvey C. Top Big Data Challenges. Available from: <https://www.datamation.com/big-data/big-data-challenges.html>. [Last accessed on 2019 Oct 06].
2. Ristevski B, Chen M. Big data analytics in medicine and healthcare. *J Integr Bioinform* 2018;15:20170030. doi: 10.1515/jib-2017-0030.
3. Beniger JR, Robyn DL. Quantitative graphics in statistics: A brief history. *Am Stat* 1978;32:1-11.
4. Heiberger R, Symanzik J. A Word from our 2011 Section Chairs. *Statistical Computing Statistical Graphics* 2011;22, 1-22. Available from: https://pdfs.semanticscholar.org/34d5/f2c24233ec5d1d6963d2dda147820825bcd6.pdf?_ga=2.242215042.1284189770.1573808515-1789433939.1555155768.
5. Adhikari A, John D. John Snow and the Broad Street Pump. Available from: <https://www.inferentialthinking.com/chapters/02/1/observation-and-visualization-john-snow-and-the-broad-street-pump.html>. [Last accessed on 2019 Oct 19].
6. Koch T, Denike K. Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. *Soc Sci Med* 2009;69:1246-51.
7. Mason B. The Topography of Disease. *Scientific American Blog Network*. Available from: <https://blogs.scientificamerican.com/observations/the-topography-of-disease/>. [Last accessed on 2019 Oct 19].
8. Cairo A. Heroes of Visualization: John Snow, H.W. Acland, and the Mythmaking Problem | | Peachpit. Available from: <http://www.peachpit.com/articles/article.aspx?p=2048358>. [Last accessed on 2019 Oct 19].
9. Lienhard JH. No. 1712: Nightingale's Graph. *Engines of our Ingenuity*. Available from: <https://www.uh.edu/engines/epi1712.htm>. [Last accessed on 2019 Oct 20].
10. Bärtschi M. Health Data Visualization A review. DIVA Group University of Fribourg 2011;4.
11. Cmsadmin. About Open Government Data (OGD) Platform India. data.gov.in. Available from: 2013. <https://data.gov.in/about-us> [Last accessed on 2019 Oct 20].
12. History. About Gapminder. Available from: <https://www.gapminder.org/about-gapminder/history/>. [Last accessed on 2019 Oct 21].
13. Taylor CK. Pie Charts, Histograms, and Other Graphs Used in Statistics. ThoughtCo. Available from: <https://www.thoughtco.com/frequently-used-statistics-graphs-4158380>. [Last accessed on 2019 Oct 23].
14. Straker D. The Quality Toolbook: Practical variations on the Bar Chart. Available from: http://syque.com/http://www.syque.com/quality_tools/toolbook/Bar/vary.htm. [Last accessed on 2019 Oct 23].
15. Bertni E, Felix C. Key Concepts and Definitions-Introduction to Information Visualization. Coursera. Available from: <https://www.coursera.org/lecture/information-visualization-fundamentals/key-concepts-and-definitions-kGpnn>. [Last accessed on 2019 Oct 23].
16. Cognitive Artifacts. Available from: <https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/cognitive-artifacts>. [Last accessed on 2019 Dec 31].
17. Public Health. Available from: <http://dhs.kerala.gov.in/index.php/publichealth.html>. [Last accessed on 2019 May 05].
18. Standard Reports. Available from: https://nrhm-mis.nic.in/hmisreports/frmstandard_reports.aspx. [Last accessed on 2020 Jan 04].
19. Sivanayak/publichealth: Contains public health data sets and codes used to create different types of analysis. Available from: <https://github.com/Sivanayak/publichealth>. [Last accessed on 2020 Jan 04].
20. Smith A. Why Data Visualization and Dashboards Matter. Available from: https://blog.dataiku.com/2016/03/10/dashboards_and_data_viz. [Last accessed on 2020 Jan 01].
21. Interactive Data Visualization for Public Health Applications: An Evaluation of Plotly's Capabilities. B.Next. Available from: <https://www.bnext.org/article/interactive-data-visualization-for-public-health-applications-an-evaluation-of-plotlys-capabilities/>. [Last accessed on 2019 Sep 29].
22. Healthcare M. How Data Visualization is Helping in Healthcare Decision Making? MarksMan Healthcare Communications. 2017. Available from: <https://marksmanhealthcare.com/data-visualization-helping-healthcare-decision-making/>. [Last accessed on 2019 Oct 10].
23. Fast, Larissa, Waugaman A. Fighting Ebola With Information Learning from the Use of Data, Information, and Digital Technologies in the West Africa Ebola Outbreak Response. 2016.
24. Harris D. A Guide to Free and Open Source Data Visualization Tools. Software Advice. 2019. Available from: <https://www.softwareadvice.com/resources/free-open-source-data-visualization-tools/>. [Last accessed on 2019 Oct 26].
25. Government of India, A digital India Initiative. Visualization Engine v3.0. Visualization Engine v3.0, An initiative of Data.gov.in, Open Government Data (OGD) platform of India. Available from: <https://visualize.data.gov.in/>. [Last accessed on 2019 Oct 27].
26. Ben S, Catherine P, Bradford W. H. Improving Healthcare with Interactive Visualization-IEEE Journals & Magazine. *IEEE Xplore Digital Library* 2013;46:58-66.
27. Shneiderman2013Improving.pdf. Available from: <https://www.cs.umd.edu/~ben/papers/Shneiderman2013Improving.pdf>. [Last accessed on 2019 Oct 19].
28. Fokkert MJ, van Dijk PR, Edens MA, Abbes S, de Jong D, Slingerland RJ, *et al*. Performance of the FreeStyle Libre Flash glucose monitoring system in patients with type 1 and 2 diabetes mellitus. *BMJ Open Diab Res Care* 2017;5:e000320.
29. Nutley T, Reynolds HW. Improving the use of health data for health system strengthening. *Glob Health Action* 2013;6:20001. doi: 10.3402/gha.v6i0.20001.
30. NRHM-HMIS. NRHM-HMIS. Available from: <https://nrhm-mis.nic.in/SitePages/Home.aspx>. [Last accessed on 2019 Nov 11].
31. National Health Policy 2017. Ministry of Health and Family welfare, Government of India. Available from: <https://mohfw.gov.in/sites/default/files/9147562941489753121.pdf>.
32. Tang L, Bie B, Zhi D. Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease. *Am J Infect Control* 2018;46:1375-80.
33. Henke N, Bughin J, Chui M. The age of analytics: Competing in a data-driven world. McKinsey Global Institute 2016. Available from: <https://www.mckinsey.com/~/media/McKinsey/>

Business%20Functions/McKinsey%20Analytics/Our%20Insights/
The%20age%20of%20analytics%20Competing%20in%20a%20
data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.
ashx. [Last accessed on 2019 Oct 18].

34. Khoury MJ, Engelgau M, Mensah GA, Chambers DA. Can Big
Data Science Deliver Precision Public Health? | | Blogs | CDC.
Available from: [https://blogs.cdc.gov/genomics/2019/07/23/
can-big-data-science/](https://blogs.cdc.gov/genomics/2019/07/23/can-big-data-science/). [Last accessed on 2019 Nov 11].