Overview: Visualizing Healthcare Data

Yang (Julia) Zhu

Abstract— Data visualization is receiving increasing attention from many industries, including healthcare. In the healthcare system, visualization plays a critical role in efficiently displaying large quantities of data to make trend or pattern visualization much easier. This paper offers an overview of select visualization idioms used to visualize univariate, bivariate and multivariate health data, and each idiom's respective strengths and limitations.

Index Terms—Information visualization, health care data, visualization idioms

INTRODUCTION

Traditionally, healthcare systems used paper records to document patient's health records, clinical information, or research findings. Left in static form, physicians and researchers spent vast amounts of time paging through paper records to mine for relevant insight. In today's digital age, healthcare systems have been creating vast quantities of healthcare data in digital format, spanning the range of patient information to clinical trials. With the sheer quantity of data generated, the complexity in processing all the information increases. This information overload may cause users to overlook, ignore, or misinterpret crucial information and hence negatively impact interpretation accuracy [1]. As such, there is a critical need to use effective ways of analysing large amounts of medical data [1].

The nature of the massive data movement has influenced the healthcare industry to realize the value in using technology to visualize all the data they have produced. Data-driven health systems are promising applications that can help to develop more personalized treatment plans, to make more precise diagnosis, and to better understand medical outcomes within diverse patient populations [2].

Given the complexity and heterogeneity of health data generated, data visualization tools have the ability to play critical roles [2]. In particular, they use visual representations to explore, make sense of, and communicate quantitative data to patients, physicians and other healthcare workers [3]. They are capable of transforming raw data into relatively straightforward and actionable insights to make understanding data and spotting trends much easier. As such, some researchers have been using visualizations to help them answer medical research questions, while others conduct design studies testing the effectiveness of using visualization technology on health data. This paper aims to discuss select visualizations that have been used to visualize univariate, bivariate and multivariate health data – with a focus on the later – and their respective strengths and limitations.

1 DEFINITIONS

Univariate data is data consisting of only one variable [4]. Univariate graphs provide information about the variable's distribution without including excessive details contained in individual data points [4]. In this way, univariate graphs can provide summative overviews of general trends of structure within the dataset [4]. Bivariate data is data involving two different variables that have a correlative relationship with each other [4]. Multivariate data is data involving three or more variables [4].

2 VISUALIZATION IDIOMS

Data visualization presents raw data in graphical visualization formats that allow users to answer questions and discover insights. As such, there are many different ways to visually present the data. An idiom is the specific way to visually create and manipulate data [5]. Common idioms include bar graphs, pie charts, scatter plots, bubble charts, and heat maps. Each idiom has its own strengths and weaknesses; deciding which idiom to use to represent a dataset depends on the research question and the type of data present.

Common univariate visualization idioms include the bar chart and the pie chart. To visualize bivariate data, graphs are commonly used to provide information on the relationship between the two variables. Scatter plots are one of the most popular options for graphically representing bivariate data. There are many idiom options for visualizing multivariate data, including radar charts and tree maps. Additionally, some multivariate data may also be tightly coupled with geospatial regions. Common idioms used to visualize these kinds of data include spiral theme plots, cartograms, heat maps and ring maps.

3 IDIOMS FOR VISUALIZING UNIVARIATE HEALTH DATA

3.1 Bar Charts

Bar charts are commonly used for comparison tasks because they take advantage of human's inherent abilities to make faster conclusions based on side-by-side length comparisons [6]. They typically display a quantitative data measure at one axis, and a categorical data measure at the other axis [6]. The height of the bars correlates to the magnitude of data [6]. For both medical personnel and patients, bar charts are relatively easy to understand and draw conclusions from compared to other idioms [6]. For these reasons, bar charts are one of the most commonly used idioms in health data visualizations [6].

In Loth *et al.*'s 2016 study, they wanted to know how effective bar charts were at presenting data about the quality of life amongst cancer patients [7]. They interviewed patients with a series of questions on their treatment durations, the magnitudes of tumour pain, and how their pain had changed over the course of one year [7]. After obtaining the data, the researchers visualized them with bar graphs, and asked the interviewed patients to judge how accurate the bar charts were in displaying their feedback [7]. Values below a predefined threshold for clinical importance were displayed as green bars, and those above a predefined threshold were displayed as red bars [7].

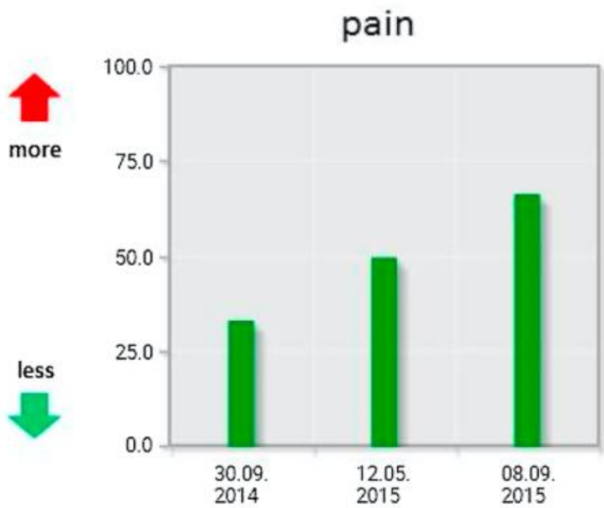


Figure 1: A bar graph showing how pain magnitudes have changed amongst patients from 2014 to 2015.

Almost 97% of patients reported the bar graphs were easy to understand, and 85% of patients believed the visuals were accurate representations of data [7]. The study's results show that bar charts indeed offer simple and accurate ways to visualize and compare counted values; however, they also have many disadvantages as well. For example, they are more appropriate for smaller datasets, as large and complex datasets would require more bars which may clutter the visual display. This problem also alludes to another drawback of the bar chart: scalability. In order for bar charts to be easily understood, there must be enough space between the bars so that they can be distinguishable [8].

Another disadvantage is that their designs can influence accurate interpretation of statistical health data. As bars often start from zero, this encourages users to compare bar sizes relative to zero [8]. This can be problematic when comparing multiple experimental samples to each other because they are usually generated from large and potentially irregular populations with inherent variations [8]. Therefore, rather than representing the distribution of the data used to calculate the means, visualizing their means with bar graphs can misleadingly assign importance to each bar's distance of the mean from zero [8].

Strengths	Limitations
Easy to see data trends	Scalability
Relatively easy and straightforward to understand	Inappropriate for interpreting statistical data
Displays relative numbers and proportions of categories	Limited to relatively small datasets

Table 1. Strengths and limitations of bar charts.

3.2 Pie Charts

Another method for visualizing univariate data is using pie charts. Pie charts are circular graphs divided into individual slices to show the relative contribution that different categories of data contribute to the overall total. Pie charts are popular visualization methods for displaying quantitative data [9]. In fact, compared to other visualization idioms, several psychophysical studies have found pie charts to be most efficient at presenting proportions: because each slice is represented as a fractional part of a whole, it is intuitively easy to understand the visual [9].

In a healthcare setting, pie charts are also frequently used to visualize data. In Ahern *et al.*'s study, they wanted to gain insight into heart disease mortality rates amongst females in both developing and developed countries through the use of a pie chart, and to provide policymakers with data helpful in prioritizing health interventions [10]. For their study, they obtained heart disease mortality data from the World Health Organization. The dataset spanned form 1994 to 2007, representing 115 countries [10]. After performing a series of data analysis, they created two pie charts based on their results: one for heart disease mortality rates for females aged 15 to 49 in developing countries, and another for those in developed countries [10].

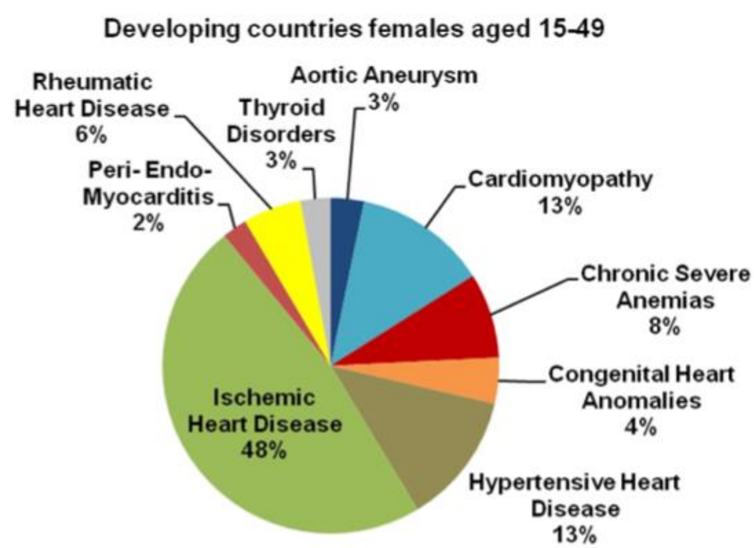


Figure 2. A pie chart displaying heart failure-attributed deaths for females aged 15 to 49 in developing countries.

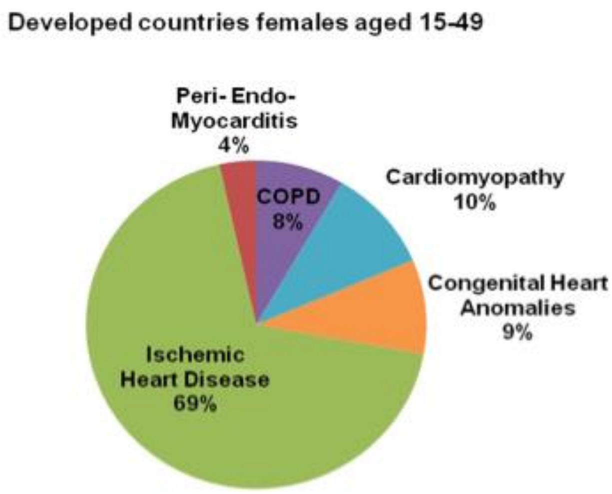


Figure 3. A pie chart displaying heart failure-attributed deaths for females aged 15 to 59 in developed countries.

Each pie chart's slices were color-coded to help ease categorization and visualization. To further help the user understand the chart, researchers provided additional information by labelling each slice with respective names and percentages. This way, each pie chart clearly breaks down the prevalence of heart diseases which led to patient mortality.

The pie charts show that ischemic heart disease was the leading cause of death, followed by hypertensive heart disease. The researchers suggested that because ischemic heart is preventable, prevention efforts, early detection programs and treatment of underlying conditions that lead to ischemic heart disease should be a top priority [10]. They believe this finding should enable policymakers to make more informed decisions in the future regarding priority in health interventions and resource allocations [10].

Even though the pie chart provides a straightforward and clear breakdown of precise causes of death and their rates in each category, this idiom also has many limitations. One of the disadvantages is that had the researchers not included the percentages of each heart failure-attributed deaths, it would have been difficult to discern the differences in some categories. For example, in figure 3 the percentage of deaths due to cardiomyopathy is at 10% and that of congenital heart anomalies is at 9%. Without the percentages being given, it would have been very difficult to see a difference in the sizes of those two slices.

Another limitation of pie charts is that a slice located at a less recognizable position may be more difficult to judge [11]. In a pie chart, humans are better at judging positions at the extreme top, bottom, left and right of a circle. When a slice is located at a 90 degree interval from other slices, it forms a right angle that is naturally easier to judge the size of [11]. Therefore, if a slice was located at a position other than the extreme top, bottom, left or right, its size becomes more difficult to judge [11].

Strengths	Limitations
Straightforward representation of data as fraction of whole	Slices of similar sizes may be difficult to tell apart from
Can accommodate additional information to be added to slices for clearer communication	Data interpretation may depend on slice location in the whole

Table 2. Strengths and limitations of pie charts.

4 IDIOM FOR VISUALIZING BIVARIATE HEALTH DATA

4.1 Scatter Plots

Scatter plots are one of the most frequently used idioms in visualizing bivariate health data. Scatter plots are square or rectangular plots made by plotting measured data results against a parameter in a Cartesian co-ordinate system [13]. They allow users to simultaneously identify a possible relationship between two variables and outliers in the data [12].

In Worku and Woldesenbet's study, they used a scatter plot to visualize a hypothesized link between life expectancy and income inequality in 52 African countries [14]. The researchers obtained their study data from UNICEF and analysed the association between the two variables using statistical methods [14]. They plotted the results on a scatter plot to graphically show the variables' possible correlation [14].

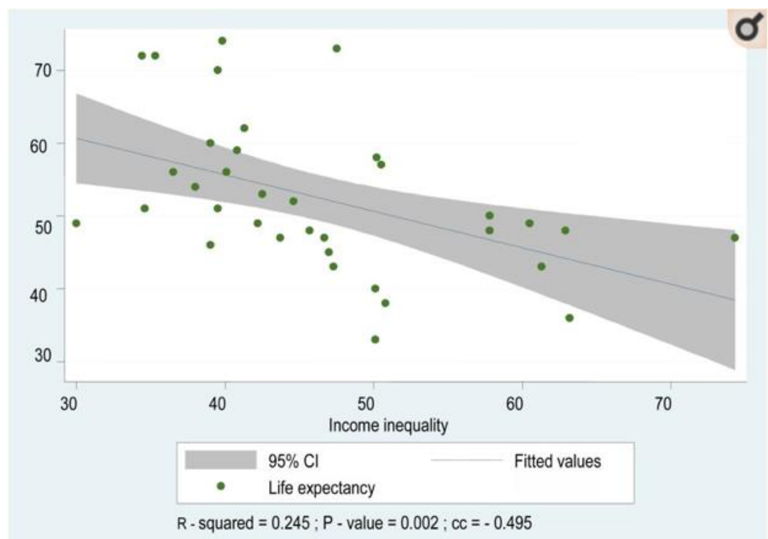


Figure 4. A scatter plot showing the connection between life expectancy at birth and income inequality.

From the scatter plot, the researchers were able to conclude that there is indeed an inverse relationship between life expectancy at birth and income inequality [14]. This finding led them to believe that powerful political and social power structures in African countries shape the distribution of life expectancy and mortality among populations [14]. Based on the results shown in the scatter plot, they believed tackling poverty should be a priority in order to provide better health for those in Africa [14].

Their study shows the advantages of scatter plots as idioms for visualizing bivariate data: they are straightforward to analyse and their scales can be managed such that scatter plots can show a wide range of data – from their minimum to maximum.

However, like other idioms, scatter plots have several limitations. One disadvantage is that even though they show a qualitative correlation between two variables, they do not show a quantitative measure of the relationship between the variables. Another limitation is that scatter plots are limited to only data consisting of two variables.

Strengths	**Limitations**
Can show a range of data (from maximum to minimum)	Shows only qualitative relationship
Observation is straightforward	Limited to bivariate data

Table 3. Strengths and limitations of scatter plots.

5 IDIOMS FOR VISUALIZING MULTIVARIATE HEALTH DATA

5.1 Radar Plots

Also known as radar charts, radar plots offer a circular graphing technique that has a series of rays projecting from a central point, where each spoke represents a different variable label [15]. The lengths of the rays represent the values of the variable [15].

Radar plots are commonly used in clinical studies to show changes over time in multiple variables in a single individual or in different groups, to show multiple treatment-group differences on multiple-outcome measures, or to show differences between disease conditions on multiple variables [15]. They have also been observed to graph basic science data such as molecular or genetic similarity between groups, comparative compositions of pharmaceuticals and biologic marker variability between various tissues or tumours [15].

Radar plots are not yet widely used in the healthcare system; however, there are a few examples of their uses present. One of the most prominent studies is Saary's 2007 study which examined a fictitious scenario showing how radar plots could be used to visualize patient satisfaction with their interaction with the healthcare system [15]. In the study, Saary had patients from three towns (Town A, Town B, Town C) rate their satisfaction levels with the healthcare system on a scale from 5 (most satisfied) to 1 (dissatisfied) on seven variables: physician knowledge (MD knowledge), physician attitude (MD attitude), simplicity of the system (Simplicity), access to emergency care (Access), wait time for tests (Wait Time), cost incurred (Cost) and availability of specialists (Availability) [15]. Total satisfaction was represented by the perimeter (outmost ring) [15].

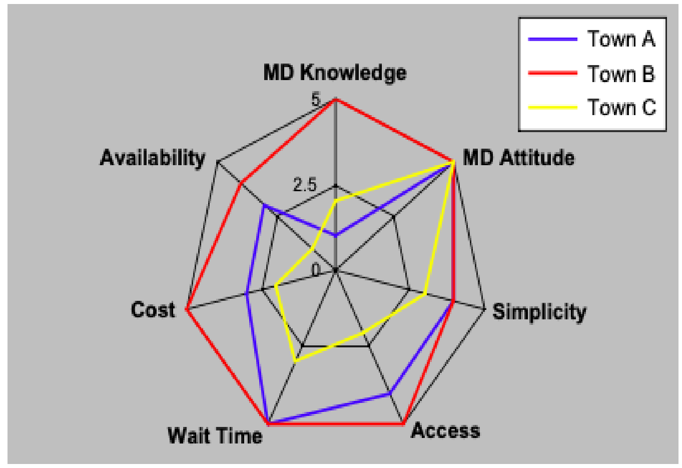
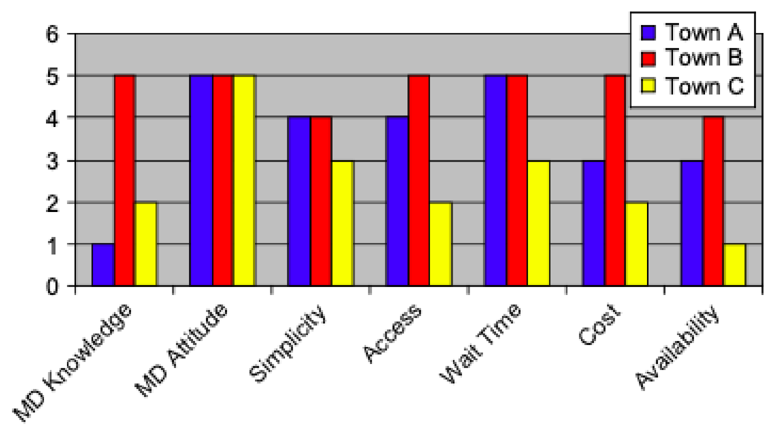


Figure 5. A radar chart comparing three groups on seven variables.

The radar plot clearly illustrates that patients from Town B were more satisfied with their healthcare system compared to patients in Town A and Town C, as Town B reached the perimeter five times out of seven. To emphasize that radar plots are more suitable for visualizing multivariate data, Saary plotted the same data (three groups on seven variables) on a bar graph for comparison.

Figure 6. Bar chart comparing three groups on seven variables.



Comparing the bar chart to the radar chart, the radar chart is noticeably less cluttered, highlighting that a radar chart can more efficiently present the bigger picture when it comes to displaying multivariate data.

Radar charts also have a number of limitations. For instance, they are simpler when associated with statistical analyses, which can be complex because of the multidimensionality [15]. It is believed that this is the main reason why they are not widely used to visualize health data [15]. Another disadvantage is that the order in which variables are arranged can impact data perception: data perception may change when they are arranged in a continuum from descriptive to analytical compared to when they are arranged from analytical to descriptive [15].

Strengths	**Limitations**

Efficient way of showing a wide variety of data in a single picture	Often associated with statistical analysis, making it complex due to multidimensionality
Can accommodate additional information by adding spokes without being overwhelming	Can be very complicated as a mathematical technique to be practical
Effectively presents the bigger picture	Variable arrangements can impact data perception

Table 3. Strengths and limitations of radar charts.

5.2 Tree Maps

Another way of visualizing large amounts of data on a single screen is using tree maps. Tree maps are space-constrained visualizations typically used to visualize genome data or health data statistics [16]. They are characterized by a large window subdivided into multiple parts, where each rectangle is proportional to a specified dimension of the data [16].

Tree maps are becoming increasingly popular in visualizing healthcare data. For example, the Gene Ontology Consortium uses tree maps to visualize genome data, and the Department of Radiology at Northwester Medical School uses them to track the real-time operational statistics of the radiology department [16].

One of the most prominent research studies using tree maps is in Hugine *et* al's 2014 study. In their research study, they examined the effectiveness of using tree map visualizations to help surgeons compare their surgery performances with their counterparts.

Within each surgery department, surgeons can perform various types of surgeries on different patients. To visualize surgical data, the researchers designed two tree maps that looked at the types of surgeries performed, the surgeon ID associated with each surgery, and the outcomes for each surgeon's patient: whether the patient survived the surgery and their length of stay (in terms of days in the hospital) [16]. Then they asked 120 undergraduate and graduate students a series of surgery judgement questions based on the tree maps to analyse their effectiveness and ease of use [16].

Figure 7. Tree map of surgeries performed by surgeons, and the surgery outcomes for each patient.

The traditional tree map in figure 7 shows the surgeon ID of each surgeon displayed on the upper left corner of each block [16]. The size of each rectangle or square represents the length of stay – in days – for each patient [16]. The varied rectangular sizes is also called varied node size [16]. Each rectangle's colour represents the outcome for each patient: red for deceased and green for those who survived the surgery [16].

The researchers then designed a second tree map which visualized the same data used to create the tree map seen in figure 7. Technical aspects of the second tree map are similar to the traditional tree map in figure 7; however, all the rectangles in the second tree map had the same size (equal node size) such that length of stay was not taken into account [16].

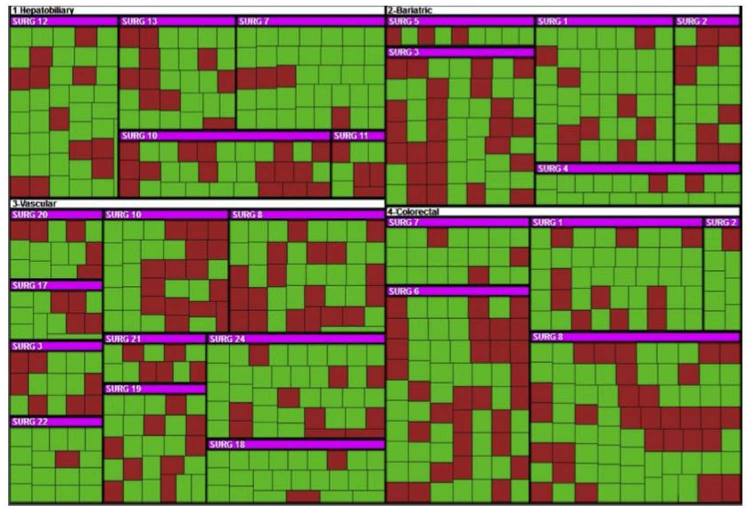


Figure 8. Tree map of surgeries performed by surgeons, and the surgery outcomes for each patient with equal node size.

The researchers asked study participants a series of questions based on the data displayed in each tree map. The results showed that more participants made accurate judgements when the tree map displayed equal node size layout [16]. Participants also recorded faster completion times for the equal node size layout [16]. This illustrates a major downfall of tree map visualizations: their layout design can affect judgement performance.

Other limitations of tree maps include size distortion: the more pixel used to show data hierarchy, the more the size distorts [17]. Eventually, it would only be possible to make size comparisons at one level of the hierarchy at a time [17].

Strengths	**Limitations**
Efficiently displays large datasets on a single screen	Size is not the most accurate channel (size distortion)
Gives clear overview of the entire hierarchy	Interpretation can depend on tree map layout (traditional vs grouped)
Can encode colour and size	Layout design can affect decision-making

Table 4. Strengths and limitations of tree maps.

5.3 Bubble Charts

A bubble chart is a variation of a scatter plot where the data points are replaced with bubbles. Bubble charts compare the relationships between data objects in three numeric-data dimensions: the x-axis data, the y-axis data, and data represented by bubble size [18]. Bubble charts sometimes also visualize a fourth variable through using different bubble colours [18]. Bubble charts allow the comparison of data in terms of their relative positions with respect to their sizes and axis [18].

In Al-Hajj *et al.*'s 2013 study on visual analytics for public health, they used a bubble chart to visualize data on hospitalization rates from injury and to spot hospitalization trends. For their chart, in addition to the categories of data represented on each axis, bubble size and bubble colour also provided additional data dimensions: the bubble's colour represented an injury cause (red for fall-related injury and blue for transport-related injury), and the bubble's size represented the injury cost [18].

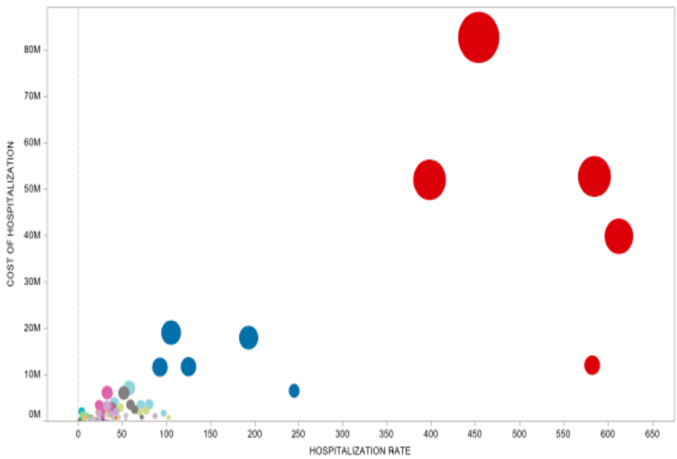


Figure 9. Bubble chart visualization of hospitalization data showing hospitalization rates and hospitalization costs per injury cause.

From the visualization, the researchers were able to conclude that fall-related injury was the most frequent cause for hospitalization, followed by transport-related injury. The red bubbles also revealed that fall-related injuries were the most costly in terms of hospitalization.

Even though bubble charts can conveniently display four-dimensional data, they also have some drawbacks as well. One of the most obvious limitations is bubble overlap. As seen in the lower left corner in figure 9, there are many similar-sized bubbles that have overlapped, as there are also a few bigger bubbles covering smaller ones. Although careful colour choice (large differences in hue) could potentially alleviate this problem, it nonetheless makes precise data visualization more challenging with large datasets.

On a similar note, bubble charts are not ideal for displaying extremely small outliers or negative values because the bubble may be so small that it might be almost impossible to see them amongst larger bubbles.

Another limitation is that determining the value of data associated with bubble size can be difficult. Since the variable's value is depicted by the bubble's area, it requires subjective judgement and comparison with other bubbles to determine those that are relatively larger and smaller.

Strengths	**Limitations**
Displays three to four variables without using 3D graphs	Ability to ascertain actual values is dependent on circle size
Visual size makes it easy to make relative comparisons	Difficult to display large quantities of data
Can be easily made with Excel	Difficult to see small bubbles if overlapped by larger ones

Table 5. Strengths and limitations of bubble charts.

6 IDIOMS FOR VISUALIZING MULTIVARIATE HEALTH DATA ASSOCIATED WITH GEOSPATIAL REGIONS

6.1 Spiral Theme Plots

Population healthcare data are usually tightly paired with geospatial regions. In order to visualize this type of data, the visualization requires integrating geo-visualization, multi-dimensional, and time-variant information visualization [19]. In order to accomplish this, Bloomquist *et al.* developed a visualization tool called the spiral them plot.

The spiral theme plot is a visualization technique that combines other information visualization methods including the theme plot, scatter plot and the spiral plot [19]. Spiral theme plots allow many years of patient data to be visualized such that seasonal or abnormal patterns for seasonal diseases could be spotted [19]. Spiral theme plots have a spiral pattern composed of stacked categories, or themes, along a spiral curve which represents the time axis [19]. Each data point is plotted within the regions of the themes [19].

To test the effectiveness of this idiom, Bloomquist and his team visualized four years of seasonal flu pattern for patients using a spiral theme plot. The flu data was associated with geographic regions in conjunction with the details of the individual and his medical histories over time [19]. The researchers plotted patients' flu data as points in stacked spiral rings, with each point possessing proper visual attributes [19]. For instance, the attribute "age" was represented as radius, while other attributes such as race and gender were represented as color and shape of the dots [19]. Time was represented as a spiral base curve and diseases were represented as stacked themes along the curve [19]. The theme's width at a particular angle was determined by the total occurrence of flu at that particular time [19]. This way, the spiral theme plot allowed for years of patient data to be plotted and visualized for seasonal patterns or diseases [19].

Figure 10. Spiral them plot displaying seasonal pattern of flu over four years.

The researchers tested the idiom's effectiveness by conducting unstructured qualitative interviews with participants. A majority of users were pleased with the visualization of long time-series geospatial health data [19]. However, some users also felt the visualization was very complex and had relatively high information density, which could obfuscate vital information [19]. Some users also preferred to have pop-up descriptions to help interpret the plot [19].

Strengths	Limitations
Can accommodate longer time axis	Can exhibit high information density
Can visualize time-variant geospatial data	May be difficult to understand without in-line guidance or pop-up descriptions
Suitable for visualizing large health datasets	Details in lower prevalence regions may be overlooked

Table 6. Strengths and limitations of spiral theme plots.

6.2 Ring Maps

In addition to using spiral theme plots to visualize geospatial health data, ring maps can also be used to plot health information in conjunction with graphical information systems (GIS).

Ring maps are recent cartographic innovations that can visualize multivariate spatial health data through illustrating individual datasets as separate rings of information [20]. The concentric, segmented rings can be circular or elliptical [20]. The rings surround a base map of a certain geographic region of interest [20]. In this way, the set up allows for the visualization of spatially referenced graphic of multiple layers of data presented in an array of regional attributes [20].

A recent example of using ring maps in visualizing health data is Lopez-De Fede *et al.*'s 2011 study. In their study, the researchers created a ring map to spatially visualize county-level HIV/AIDS, sylphilis, chlamydia and gonorrhoea diagnosis rate data for South Carolina. They constructed a ring map with two principal parts: a central base map and a ring display [20].

The central base map showed South Carolina's 46 counties [20]. Radiating from the base map were 46 spokes, each of equal width, representing each of the state's counties [20]. The ring display consisted of four concentric rings, each representing a separate layer of data [20]. For example, the three outer rings represented chlamydia, gonorrhoea and syphilis diagnosis rates; the inner ring represented HIV/AIDS diagnosis rates [20]. The rings were shaded to depict the prevalence of each illness. [20]

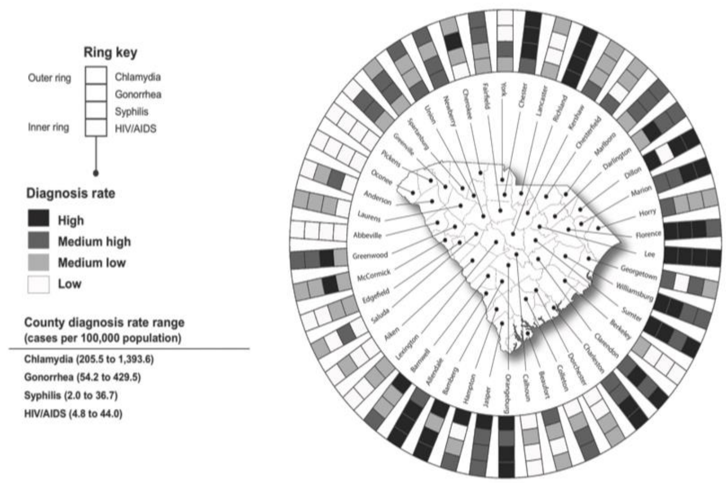


Figure 11. Ring map showing diagnosis rates of HIV/AIDS, syphilis, gonorrhoea and chlamydia in South Carolina counties.

Their ring map revealed large gender and racial disparities in STD and HIV/AIDS diagnosis across the state. This finding was useful in targeting at-risk populations in terms of resource allocation and prevention programs [20].

Ring map visualizations also have a number of strengths and limitations. One of their advantages is that they can be adapted to spatially visualize a wide variety of datasets at multiple geographic scales. In this study the researchers used them to visualize STD and HIV/AIDS diagnosis rates in association with socioeconomic standards. But ring maps could also be used to visualize health data in association with other properties such as linguistic isolation or family fragmentation [20].

Another strength of ring maps is that spatial data represented in rings can be summarized at various geographic levels, including postal code areas or public health service areas [20]. Moreover, ring maps can convey a variety of time-series data for a single health condition [20]. For example, a map with 10 rings could show HIV rates over the course of 10 years, while a map with 12 rings could illustrate the weekly incidence of syphilis outbreak over 12 weeks [20].

A major limitation with ring maps is the loss of data regarding spatial continuity of geographic units in the rings [20]. For instance, in the center of South Carolina lies Richland County. This county is surrounded by six adjacent counties; however, in the rings it was portrayed as having only two adjacent neighbours [20]. In this manner spatial topology was not accurately represented [20].

Another disadvantage is that the number of rings and spokes that can be added are limited due to graphic space and legibility constraints [20]. Finally, the visualization does not convey statistical information about the correlation between geographic area and diagnostic rates in illness.

Strengths	Limitations
Can be used to visualize a variety of datasets at multiple geographic scales	Potential inaccurate representation of spatial topology
Spatial data can be summarized at various geographic levels	Number of rings and spokes can be limited due to space limitations
Can convey different time-series data	Does not convey statistical data on correlation between geography and illness diagnostic rates

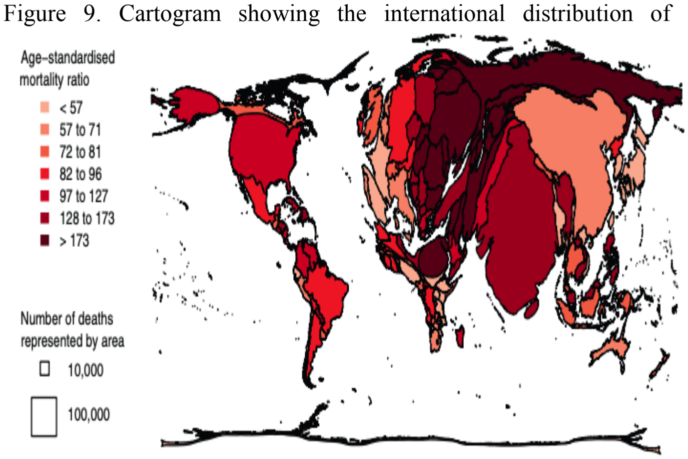
Table 7. Strengths and limitations of ring maps.

6.3 Cartograms

Besides from using spiral theme plots and ring maps to visualize population geospatial health data, cartograms are another option. Cartograms are maps that distort the shape of geographic regions so that the area encodes a data variable. This feature makes them valuable for their abilities to offer spatial representations of a variable of interest while also retaining the semblance of a world map [21].

Nowbar *et al.* used a cartogram to visualize the worldwide distribution of ischaemic heart disease mortality rates in 2010. To do so, they algorithmically transformed the area of each country so that it was proportional to the number of deaths due to ischaemic heart disease in each nation [21]. The researchers then used colour gradients to represent standardized mortality rates in each country, with darker shades of red representing countries with higher standardized mortality ratios [21]. This allowed users to assess both the relative ischaemic heart disease risks of different countries as well as the magnitude of such risks.

Figure 9. Cartogram showing the international distribution of



ischaemic heart disease.

Using the cartogram, researchers were able to deduce that India and China contributed the greatest proportions of ischaemic heart disease deaths in comparison to their true land areas [21]. At the same time, India had the highest standardized mortality rate – as can be seen by its darker shade of red in comparison to China or Western European countries [21]. Based on this finding, the researchers believed that cartographic displays of ischaemic heart disease can assist the recognition of age-standardized mortality rates in individual countries [21].

Indeed, this study has shown that cartograms are convenient ways of showing international geographical distributions of data of interest. The visualization's colour shade feature can be used to encode data quantity and prevalence. Together, they allow users to assess data both qualitatively and quantitatively [21].

However, they also possess many limitations. For example, the scale may be variable such that cartograms are not true maps and hence, the geographic regions they represent may not be truly indicative of a country's area or shape. Additionally, shapes and sizes of nations my be distorted which could make it potentially difficult to visualize countries.

Strengths	Limitations
Convenient way to show global geographical distribution of data	Scale can be variable
Colour can be used to encode quantity	Distorts sizes of countries

Table 8. Strengths and limitations of cartograms.

6.4 Heat Maps

Heat maps are popular visualization techniques to visualize health data. They are often used to show gene expression or geospatial population health data [22]. Structurally, heat maps often have a matrix of cells displayed on a single screen, and use a gradient of coloured cells to visualize the intensity of individual data points to communicate relationships between the dataset [22]. The combination of cell location and colour gradients allow users to generate conclusions about the dataset.

A study using heat maps to visualize geographic variations in health risk factors is seen in Loop *et al.*'s 2017 study. For their research, Loop and his colleagues created several heat maps to visualize the prevalence of hypertension, self-reported smoking and diabetes mellitus among white and black people across American counties [23].

In their heat maps, they used a warm-to-cool colour spectrum to show prevalence (in percentages) of hypertension, diabetes mellitus, and smoking prevalence. Red was representative of high prevalence, and blue was indicative of low prevalence [23].

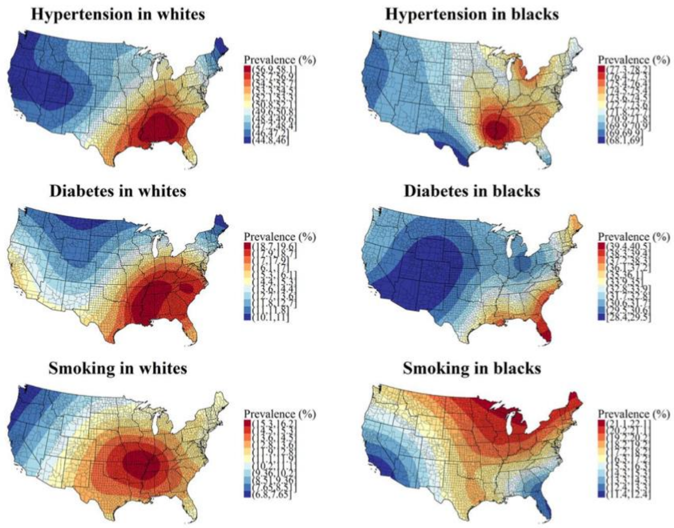


Figure 10. Heat maps of hypertension, diabetes mellitus, and current smoking prevalence among whites and blacks.

Results from the heat maps show geographic heterogeneity of each risk factor, the prevalence of hypertension, diabetes mellitus and self-reported smoking varied across the country, and that the prevalence of the risk factors were not uniformly high across Southeastern United States as the researchers had previously hypothesized [23].

The spatial representation and colour gradients in heat maps makes them suitable for visualizing trends in multivariate geospatial data. On the other hand, if they have an excessive use of colours, it may have a negative impact on the legibility of the graphic. Colour shading also can have a strong influence over the effectiveness of heat maps. When humans perceive shading, they can easily notice sharp contrasts between adjacent cells; it becomes increasingly difficult to compare shading in non-adjacent regions [24]. Known as the Checker Shadow illusion, this phenomenon could potentially mislead users into thinking that some non-adjacent cells are different colours when they are actually the same hue [24].

Strengths	Limitations
Objective spatial representation	Legibility dependent on colours
Colour gradients can ease data interpretation	Checker Shadow illusion may influence data interpretation

Table 9. Strengths and limitations of heat maps.

7 Related works

Currently there are no previous survey papers on visualizing general health data. Instead, there are systematic reviews on visualization and analytic tools for infectious disease epidemiology, electronic health record data, and health informatics. These studies focused on select popular visualization idioms known to have been used in specific medical fields. Each study covered two to three idioms.

8 Conclusion

Data visualization is a powerful tool for realizing trends and outliers in data. As it is gaining popularity in various industries, the health care sector is also beginning to utilize visualization to improve healthcare and policies. Although health data visualization is still a relatively new field, the present work has identified previous studies that utilized a variety of visualization idioms to graphically present univariate, bivariate and multivariate health data. Some of these studies used idioms to visualize health data and subsequently answer research questions, while others were design studies that tested the effectiveness of the chosen idioms. Moreover, this paper also discusses the strengths and limitations of each presented idiom.

Acknowledgments

I would like to thank Professor Tamara Munzner for her guidance and support with this project.

References

[1] Caban, Jesus, and David Gotz. “Visual Analytics in Healthcare – Opportunities and Research Challenges.” *Journal of the American Medical Informatics Association*, vol. 22, no. 2, Mar. 2015, pp. 260–262., doi:10.1093/jamia/ocv006.

[2] Gotz, David, and David Borland. “Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization.” *IEEE Computer Graphics and Applications*, June 2016, pp. 90–96., doi:10.1109/MCG.2016.59.

[3] Hesse, Bradford, et al. “Improving Healthcare with Interactive Visualization.” *Computer*, vol. 46, no. 5, May 2013, pp. 58–66., doi:10.1109/MC.2013.38.

[4] Jacoby, William. *Statistical Graphics for Univariate and Bivariate Data*. Thousand Oaks, CA: SAGE Publications Inc., 1997. *SAGE Research Methods*. Web. Quantitative Applications in the Social Sciences., doi: 10.4135/9781412985963.

[5] Munzner, Tamara. *Visualization Analysis and Design*. CRC Press, 2014.

[6] “Worth a Thousand Words: How to Display Health Data.” *California HealthCare Foundation*, Feb. 2014, https://www.chcf.org/wp-content/uploads/2017/12/PDF-WorthThousandWordsDataViz.pdf.

[7] Loth, F, et al. “Cancer Patients’ Understanding of Longitudinal EORTC QLQ-C30 Scores Presented as Bar Charts.” *Patient Education and Counselling*, vol. 99, no. 12, Dec. 2016, pp. 2012–2017.

[8] “Kick the Bar Chart Habit.” *Nature Methods*, Feb. 2014, p. 113., doi:10.1038/nmeth.2837.

[9] Linden, Sander, et al. “How to Communicate the Scientific Consensus on Climate Change: Plain Facts, Pie Charts or Metaphors?” *Climatic Change*, vol. 126, 2014, pp. 255–262.

[10] Ahern, Ryan, et al. “Improving the Public Health Utility of Global Cardiovascular Mortality Data: the Rise of Ischemic Heart Disease.” *Population Health Metrics*, vol. 9, no. 8, 2011.

[11] Few, Stephen. “Save the Pies for Dessert.” *Perceptual Edge Visual Business Intelligence Newsletter*, 2007.

[12] Keim, Daniel, et al. “Generalized Scatter Plots.” *Information Visualization*, 2010, pp. 301–311., doi:10.1057/ivs.2009.34.

[13] Geladi, Paul, et al. Scatter plotting in multivariate data analysis. *Journal of Chemometrics*, 2003, pp. 503–511.

[14] Worku, Eshetu, and Selamawit Woldesenbet. “Poverty and Inequality - but of What - as Social Determinants of Health in Africa?” *African Health Science*, vol. 15, no. 4, 2015, pp. 1330–1338., doi:10.4314/ahs.v15i4.36.

[15] Saary, Joan. “Radar Plots: a Useful Way for Presenting Multivariate Health Care Data.” *Journal of Clinical Epidemiology*, vol. 60, Apr. 2007, pp. 311–317.

[16] Hugine, Akilah, et al. “Visualizing Surgical Quality Data with Treemaps.” *Journal of Surgical Research*, Mar. 2014, pp. 74–83.

[17] Lü, Hao, and James Fogarty. "Cascaded Treemaps: Examining the Visibility and Stability of Structure in Treemaps." *Proceedings of Graphics Interface*, 2008.

[18] Al-Hajj, Samar, et al. "Visual Analytics for Public Health: Supporting Knowledge Construction and Decision-Making." Jan. 2013, doi:10.1109/HICSS.2013.599.

[19] Bloomquist, Sam, et al. "Healthcare Data Visualization: Geospatial and Temporal Integration." *Imaging and Computer Graphics Theory and Applications*, vol. 2, Jan. 2016, pp. 214–221., doi:10.5220/0005714002120219.

[20] Lòpez-De Fede, Ana, et al. "Spatial Visualization of Multivariate Datasets: An Analysis of STD and HIV/ AIDS Diagnosis Rates and Socioeconomic Context Using Ring Maps." *Public Health Rep.*, Sept. 2011, pp. 115–126., doi:10.1177/00333549111260s316.

[21] Nowbar , Alexandra, et al. "2014 Global Geographic Analysis of Mortality from Ischaemic Heart Disease by Country, Age and Income: Statistics from World Health Organisation and United Nations." *International Journal of Cardiology*, 2014.

[22] Gehlenborg, Nil, and Bang Wang. "Heat Maps." *Nature Methods*, 2012.

[23] Loop, Matthew, et al. "Heat Maps of Hypertension, Diabetes Mellitus, and Smoking in the Continental United States." *Circ Cardiovasc Qual Outcomes*, 2017.

[24] Mohamad, Zaid, et al. "Eye Illusion Enhancement Using Interactive Differential Evolution." *2011 IEEE Symposium on Differential Evolution (SDE)*, 2011, doi:10.1109/SDE.2011.5952064.