Mapping Disease: Geographic Data Visualizations Displaying Human Disease

Justine Ledochowski

PSYC6135

Mapping Disease: Geographic Data Visualizations Displaying Human Disease

John Snow's 1854 map of the cholera outbreak in Soho, London is one of the most widely cited maps of human disease. Since Snow's cholera map, the development of visualization methods of human disease has proliferated (Carroll et al., 2014). Data visualizations are used by public health professionals to display and analyze complex factors of human disease in order to use the data for hypothesis testing and decision making (Carroll et al., 2014). The advent of the Internet and improved sharing and communication of national and international health-related data has facilitated the development and usage of novel disease-related data visualization methods. Furthermore, the Internet opened the accessibility of health-related data visualizations to the general public. This paper will begin with a historical review of Snow's map and the methodology he used. The remainder of the paper will focus on contemporary medical maps such as medical atlases, online platforms, and the use of geographic information systems in visualizing health-related data. Examples will be provided to illustrate how visualizations can be used to synthesize information relating to surveillance of human disease.

John Snow's Cholera Map

Perhaps one of the most well-known visualizations of disease is John Snow's cholera map (1854). Snow was first exposed to cholera when he practiced as a medical assistant working in Newcastle coal mines during the first cholera outbreak affecting England in the 1800s (Koch, 2004). During the cholera epidemic beginning in the 1840s Snow was practicing as a physician and anesthesiologist in London (Koch et al., 2004). During this time, the dominant theory was that cholera was spread by *miasma,* or bad air. In contrast, through observation that the symptoms of cholera affected the gastrointestinal system rather than the lungs Snow hypothesized that cholera was caused by something that was ingested rather than inhaled (Snow, 1849 as cited by Koch, 2004). The 1854 cholera outbreak provided Snow with the opportunity to gather data in order to examine his theory.

In one of his studies of the outbreak Snow geographically mapped instances of cholera in the Soho area of London (Figure 1). Snow mapped each death by including a black bar on the map. In cases when there was more than one death per residence, Snow used multiple bars next to each other to indicate this, which allowed for observing clusters in residences. Through geographically mapping cholera deaths Snow was able to show that the deaths were "most numerous next to the pump in Broad Street" leading Snow to propose this water pump had been contaminated and was the source of this outbreak (Snow, 1855 as cited by Koch, 2004).

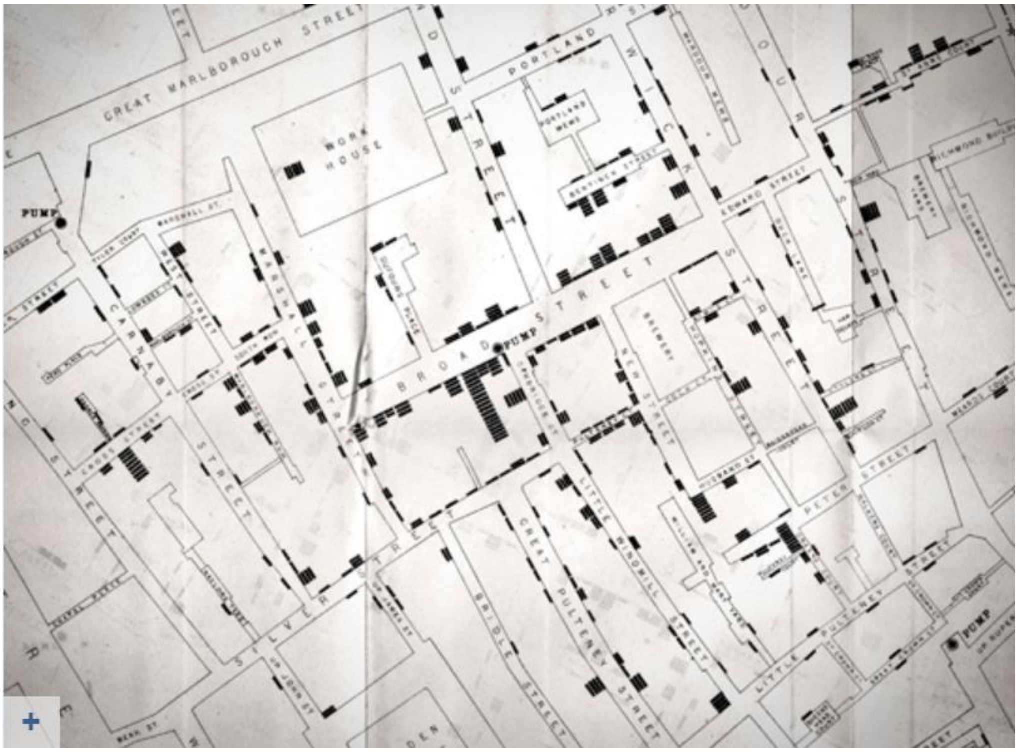

Figure 1. John Snow's Cholera Map showing the spread of cholera around the Broad Street water pump. (*Source*: The Historical Medical Library of The College of Physicians of Philadelphia. Snow J. On the Mode of Communication of Cholera, 2nd ed. London: Churchill, 1855. Accessed via https://www.historyofvaccines.org/content/snows-cholera-map)

It is interesting to consider that Snow was not medical cartographer or any sort of map-maker (Koch et al., 2004). His now famous map was only a portion of his work on cholera, which he used as a "diagram of the topography of the outbreak" alongside a study investigating the correlation between water sources and cholera incidence and further writing including individual case histories (Snow, 1854 as cited by Koch, 2004). For example, Snow also examined anomalies in his map through further investigation. For instance, none of the workers at a brewery located near the Broad Street pump contracted cholera. Through interviewing the manager Snow learned that the brewery had its own water pump that was used by the workers, therefore none of them were exposed to the contaminated pump (Koch, 2004). The original map has now been examined and re-drawn by several researchers including E.W. Gilbert (1958), Andrew D. Cliff and Peter Haggett (1988), Edward Tufte (1983), Monmoneir (1990s), and more recently the US Center for Disease Control (CDC) (Koch, 2004). While Snow was not the first to use maps to study disease, he is considered to be one of the first to use proximity as a measurement when analysing intensity and diffusion of disease in an area of this size. His contributions to the fields of epidemiology and public health cannot be understated (Koch, 2004). Snow's cholera map has been cited as "A classic example of the value of a geographically orientated approach in health studies" (Foody, 2006).

Contemporary Medical Maps

In her review of the use of medical atlases to detect patterns in human disease and healthcare, Gunderson (2000) writes "the value of medical atlases lies in their unique ability to provide tremendous amounts of information about the relation between disease and location in a concise, visual, easy-to-read format, primarily maps". An advantage to the visualization of health data over reporting numerical or verbal statistics is the ability to illustrate an image which otherwise would be multiple pages of text, allowing the viewer to more quickly synthesize the information provided and observe patterns and trends that would likely be undetectable to the human eye in text format (Gunderson, 2000). Medical Atlases produced into the 1990s were painstakingly drawn by hand, and therefore difficult to update (Koch & Denike, 2004). However, as computer methods developed medical maps can now display geographic patterns of disease alongside variables such as age, race, and sex; allowing for the analysis of patterns in health care, disease, and mortality through the use of multiple layers overlaid on a single map (Gunderson, 2000). For instance, lung cancer rates can be overlaid with smoking rates, mortality from lung cancer, and smoking habits per specific demographics such as white women (Gunderson, 2000). Investigation of the different variables can be then used for generating hypotheses or making decisions about health-related services (Gunderson, 2000). Gunderson (2000) provides an example of a geographic pattern of cancer mortality from lung cancer, which can be seen to be clustered around the Atlantic and Gulf coats (Figure 2.) This geographic pattern initiated a research program to uncover the underlying cause of cancer mortality in these regions, which resulted in the discovery of a link between asbestos exposure during World War II from shipyards in those counties. This serves as a poignant example of a health map being used to spur data analytic questions and leading to meaningful results that can be acted upon to better human health (e.g., if this area is known to be a risk then attempts can be made to limit people living nearby).

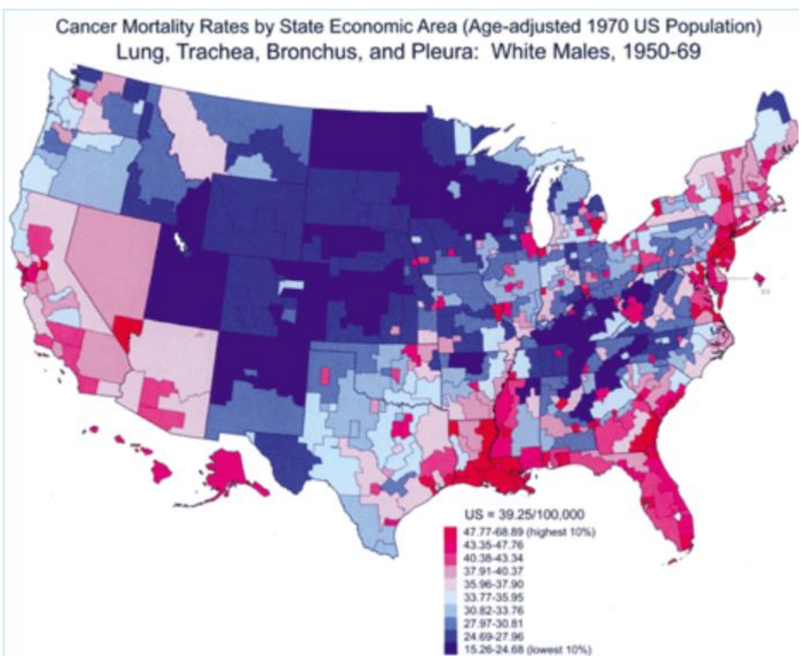

Figure 2. Map displaying rates of male cancer mortality from the Atlas of Cancer Mortality in the United States, 1950-94 as reproduced by Gunderson (2000) accessed via: https://annals.org/aim/article-abstract/713689/mapping-out-using-atlases-detect-patterns-health-care-disease-mortality)

Medical maps are also provided for public access via websites such as World Mapper (www.worldmapper.org). World Mapper uses cartograms, a type of data visualization in which territories on a map are resized in order to demonstrate the variable of interest (World Mapper, 2019).

That is, higher levels of that variable are indicated by an increase of that territory, while lower levels of the variable are demonstrated by a shrinking of the territory on the map. World Mapper plots several types of data, not only human disease, however many cartograms demonstrating world prevalence of disease as well as health related outcomes (e.g., infant deaths, childhood obesity) are available, for example, this visualization of Adults with Diabetes in 2017 (Figure 3). The visualization demonstrates that the absolute highest absolute number of adults with diabetes can be found in China, followed by India, and then the United States of America, which is visible by the swelled appearance of those locations. Additionally, areas with low prevalence such as Greenland are shrunken on this visualization. However, cartograms can be somewhat misleading as focusing on the absolute highest number does not consider the base population (e.g., prevalence rates). This is clarified in the accompanying text, which explains that among the countries with he highest absolute number of adults living with diabetes the countries with the highest prevalence rates in order are Mexico, United States, China, Brazil, and India. This clarification is not noticeable through looking at the cartogram alone. However, the advantage of visualizations provided through World Mapper includes increased public awareness of global health issues. This may stimulate interested viewers to further examine the topic. Each cartogram also includes a short summary and further resources, which is helpful in this endeavor.

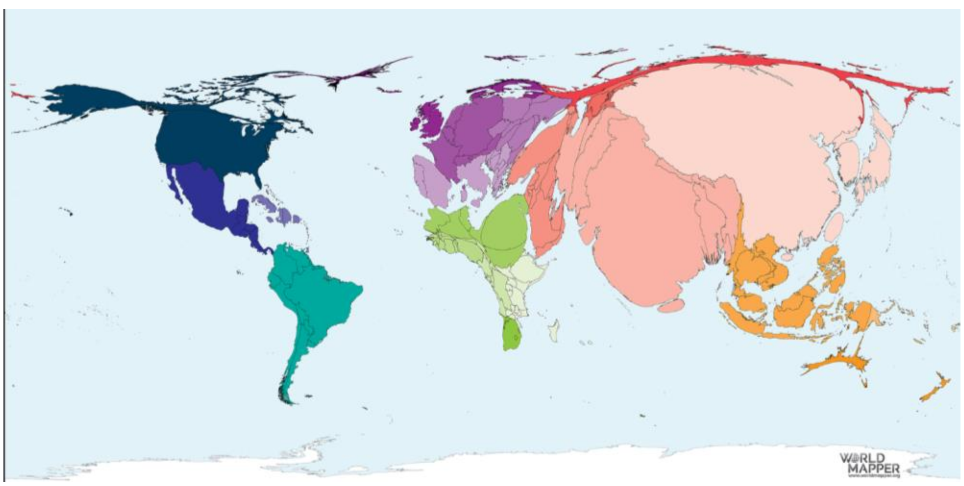

Figure 3. Adults with Diabetes in 2017, cartogram sourced and accessed from: https://worldmapper.org/maps/disease-adults-diabetes 2017/ ?sf_action =get_data&sf_ data=results&_sft_product_cat=disease

Health Map (www.healthmap.org) is another example of publicly accessible health-related data visualization. Health Map is a freely accessible, automized, surveillance system that mines media sources for health-related data in near real-time (Brownstein, Friefeld, Reis, & Mandl, 2008). Capitalizing on the on the vast health-related information available through the Internet to create an interactive map of disease outbreak Health Map collects information from a variety of informal sources such as Google News, expert-curated accounts such as Program for Monitoring Emerging Diseases (ProMED Mail), and validated alerts (e.g., World Health Organization announcements) (Friefeld, Mandl, Reis, & Brownstein, 2008) and uses a Bayesian machine learning algorithm to classify alerts by location and disease, and then overlay them on an interactive geographic map (Friefeld et al., 2008). An evaluation of the system shows that the classifier performed with 84% accuracy with higher accuracy for ProMED Mail than Google News, which the authors explained by referencing the message structure of the two sources

(Friefeld et al., 2008). The overarching goal of Health Map is to report on potential threats to public health and reduce information overload that may result from consulting several competing web sources (Friefeld et al., 2008). Viewing the current homepage of the website for my location (Toronto) I can see nearby outbreaks as dots on the map, as well as text-based alerts (Figure 4).

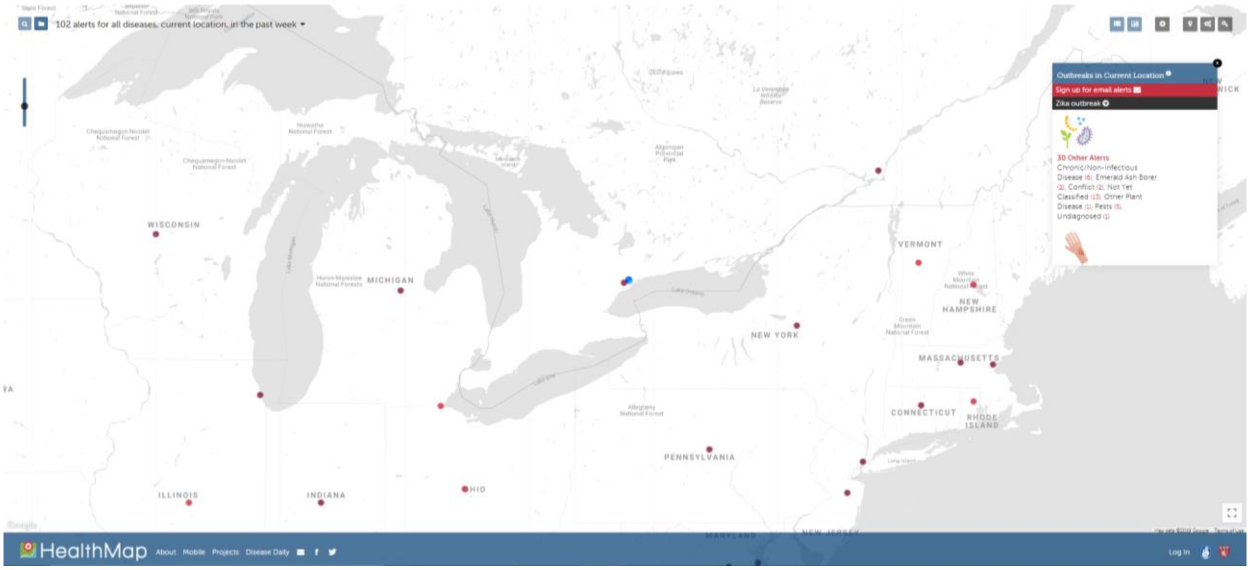

Figure 4. Screenshot taken of the home screen of www.healthmap.org visible to the author. The blue dot is the author's location and the maroon dots represent media coverage of various health-related news. For example, the maroon dot nearby the author's location takes one to an article about measles cases and vaccinations in Toronto.

Interestingly, internet-based health surveillance systems such as Global Public Health Intelligence Network and ProMED alerts have been long been available for public health authorities. However, more recently websites sharing information about infectious disease designed for general public are becoming more popular. These websites are able to increase awareness of public health issues as well as provide critical information for people travelling internationally (Friefled et al., 2008). Additionally, difficulties with current surveillance networks used by public health professionals include gaps in geographic coverage and poor information flow across international borders (Sturtevant, Anema, & Brownstein, 2007). The benefit of informal Internet information is that it can provide current and highly local information about disease outbreaks, even in areas that may not be as easily accessible to global public health efforts (Woodall, 1997). An issue with internet-sourced data includes misinformation or reporting bias (Brownstein et al., 2008). Specifically, sources may be less reliable, lack appropriate training, and report stories without adequate confirmation of presence of the disease (Brownstein et al., 2008). Additionally, political bias such as state censorship may also be a factor. The Health Map team has made attempts to better understand these issues through ongoing analysis and evaluation research (Brownstein et al., 2008). World Map may be particularly helpful for sharing information in under-resourced areas and future developments include additional languages and using low band-width display options such as mobile phone alerts (Brownstein et al., 2008).

Geographic Information Systems

Geographic information system (GIS) is a visualization method that is extends beyond traditional cartography (Muehrcke, 1990). GIS consists of a base map to which data can be added and integrated as needed for the desired visualization via advanced computing methods. GIS is a useful tool for public health professionals to utilize for tracking disease and has allowed for the abilities to detect disease clustering, analyze spread of disease in communities and territories, and predict outbreaks (Carroll et al., 2014). Ultimately, the goal is to utilize GIS to stimulate analyses of human disease that lead to more effective control measures and interventions. As disease phenomena are strongly associated with spatial (e.g., case location, disease diffusion) and temporal factors (e.g., transmission rates, mobility of susceptible populations, temporal cluster shift [Foody, 2006]) GIS a uniquely suited tool for visualizing this type data (Gao, Mioc, Anton, Yi, & Coleman, 2008). GIS systems can help identify cases and exposures to disease, correlate spatial data, and test hypotheses (Carroll et al., 2014). GIS systems provide an effective method to manage, store, analyze, and map disease information as well as other pertinent data such as demographic and environmental variables (Gao et al., 2008). A further advantage of Internet-based GIS is the rapid speed of information sharing, which can allow public health professionals to communication and make decisions about disease outbreaks (Gao et al., 2008).

One example of using GIS to plot disease data is dynamic mapping (Castronovo, Chui, & Naumova, 2009), which allows for the simultaneous visualization of spatial and temporal data. Spatial and temporal variables of particular relevance to mapping human disease are seasonal outbreaks of disease, synchronization between disease and exposure, geographic distribution of disease and exposure, whether the disease travels, and disease clusters that remain in the same location (Castronovo et al., 2009). Castronovo and colleagues (2009) used dynamic mapping to view *Salmonella* hospitalization among the elderly in the United States of America alongside environmental exposure factors including livestock (broiler chickens) and temperature (absolute temperature and deviation from the 30-year norm). Temporal data was aggregated to a monthly level, and spatial data to a county level. Hospitalization was visualized using graduated dot symbols, which were selected for quick visual comparison. Choropleth maps were used to show environmental factors. In order to make the maps dynamic monthly maps of county-level *Salmonella* hospitalizations of the elderly were placed over maps on environmental exposures and made into a movie using Adobe Flash. The timing of the movie was chosen so that viewers could have enough time to identify clusters and commit them to short-term memory (Castronovo et al., 2004). Controls are also included so that viewers may pause, replay, and move through the movie at their own pace. Figure 5 shows an example of a section of the dynamic visualization, specifically screenshots taken that show the rates of *Salmonella* hospitalizations of the elderly over the months of May to August 2002 (dots) overlaid on number of broiler chickens sold (choropleth style map). It can be seen that initially the cases appear more distributed but by August (e.g., as temperature increases) *Salmonella* cases appear to be clustered around regions with higher sales of broiler chickens (more blue dots on the darker portion of the map). Overall, the dynamic maps were able to demonstrate more clusters in the summer months and higher density clusters in the Southern United States. Additionally, a relationship was found between high broiler chicken sales and dense disease clusters in the summer (illustrated via the selected screenshots). The authors concluded that dynamic mapping of disease allows one to observe *processes* (e.g., seasonal effects, rates of livestock sales) rather than static patterns (e.g., geographic location) as possible on traditional maps. Through the use of dynamic mapping the authors were able to demonstrate environmental factors that influenced hospitalization rates of *Salmonella.*

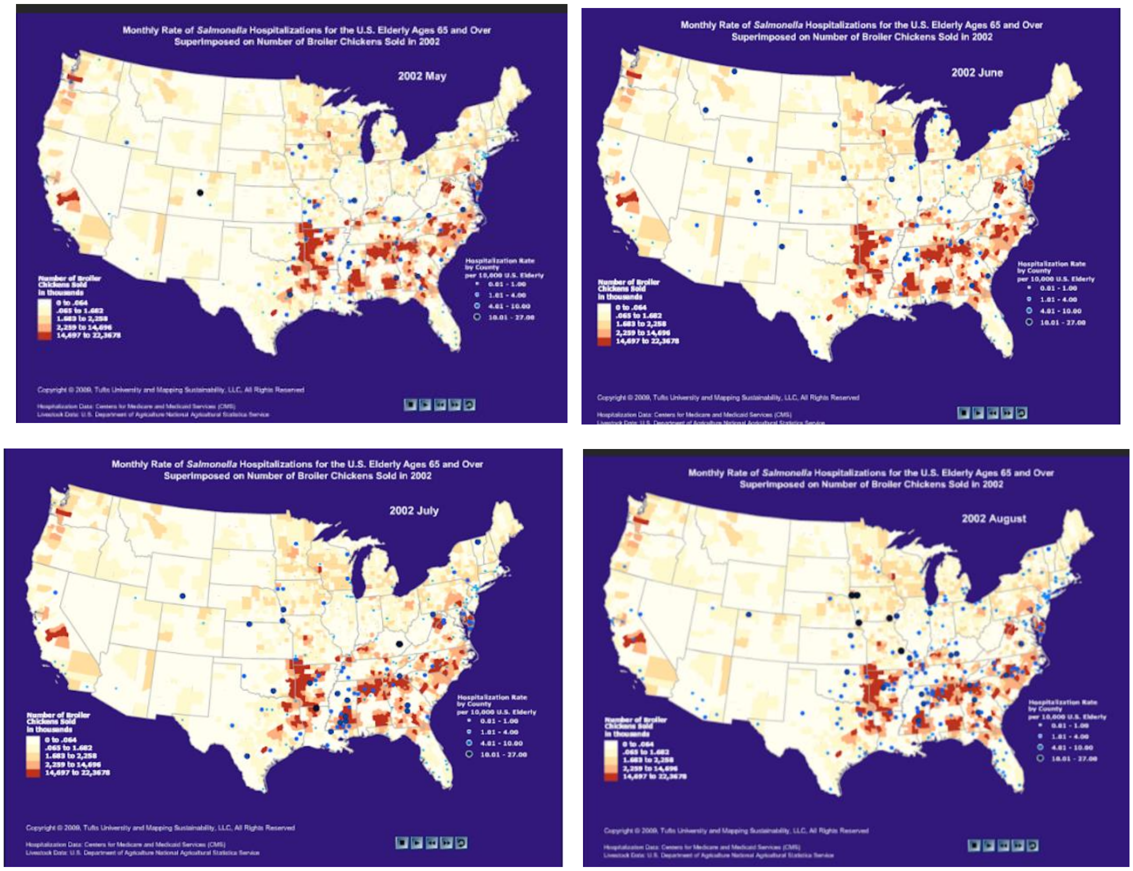

Figure 5. Screenshots from the dynamic map created by Castronovo and colleagues (2009) illustrating *Salmonella* hospitalization rates among the elderly overlaid with broiler chicken sales for the months of May, June, July and August. Accessed via electronic summary material accompanying the article (https://ehjournal.biomedcentral.com/articles/10.1186/1476-069X-8-61)

Conclusion

John Snow likely did not realize the legacy he would be leaving in the domain of disease mapping. His map of the Soho cholera outbreak is now widely recognized as a key example of the utility of geographically mapping human disease. Since the 1854 cholera outbreak in London, methods of mapping human disease have evolved greatly. Mapping disease provides an advantage over raw disease data it allows for a visual method of identifying cause and effect relationships between humans and their environment (Gao et al., 2008). As the Internet developed the mode of sharing information across national and international borders was greatly facilitated allowing for the design of data visualizations spanning countries or the entire world. Public health professionals can more easily communicate about potential outbreaks which facilitates understanding of human disease. Moreover, visualizations are also available to the general public allowing for greater awareness of public health issues. This paper reviewed several maps of human disease and the insight that was gained through geographically visualizing disease.

References

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS medicine*, *5*(7), e151.

Carroll, L. N., Au, A. P., Detwiler, L. T., Fu, T. C., Painter, I. S., & Abernethy, N. F. (2014). Visualization and analytics tools for infectious disease epidemiology: a systematic review. *Journal of biomedical informatics*, *51*, 287-298.

Castronovo, D. A., Chui, K. K., & Naumova, E. N. (2009). Dynamic maps: a visual-analytic methodology for exploring spatio-temporal disease patterns. *Environmental Health*, *8*(1), 61.

Foody, G. M. (2006). GIS: Health applications. *Progress in Physical Geography*, *30*(5), 691-695.

Freifeld, C. C., Mandl, K. D., Reis, B. Y., & Brownstein, J. S. (2008). HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *Journal of the American Medical Informatics Association*, *15*(2), 150-157.

Gao, S., Mioc, D., Anton, F., Yi, X., & Coleman, D. J. (2008). Online GIS services for mapping and sharing disease information. *International Journal of Health Geographics*, *7*(1), 8.

Gundersen, L. (2000). Mapping it out: using atlases to detect patterns in health care, disease, and mortality. *Annals of internal medicine*, *133*(2), 161-162.

Koch, T. (2004). The map as intent: variations on the theme of John Snow. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *39*(4), 1-14.

Koch, T., & Denike, K. (2004). Medical mapping: The revolution in teaching—and using—maps for the analysis of medical issues. Journal of Geography, 103(2), 76-85.

Sturtevant, J. L., Anema, A., & Brownstein, J. S. (2007). The new International Health Regulations: considerations for global public health surveillance. Disaster Medicine and Public Health Preparedness, 1(2), 117-121.

Muehrcke, P. C. (1990). Cartography and geographic information systems. *Cartography and Geographic Information Systems*, *17*(1), 7-15.

Woodall, J. (1997). Stalking the next epidemic: ProMED tracks emerging diseases. *Public Health Reports*, *112*(1), 78.

World Mapper (2019). www.worldmapper.org