



Contents lists available at SciVerse ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Community Health Map: A geospatial and multivariate data visualization tool for public health datasets

Awalin Sopan^{a,*}, Angela Song-Ie Noh^b, Sohit Karol^c, Paul Rosenfeld^d, Ginnah Lee^d, Ben Shneiderman^a

^a Human–Computer Interaction Lab, Department of Computer Science, University of Maryland, USA

^b Department of Computer Science, University of Maryland, USA

^c Department of Kinesiology, University of Maryland, USA

^d Department of Electrical and Computer Engineering, University of Maryland, USA

ARTICLE INFO

Available online xxxx

Keywords:

Healthcare data
Geospatial multivariate visualization
Graphical user interface
Information visualization

ABSTRACT

Trillions of dollars are spent each year on health care. The U.S. Department of Health and Human Services keeps track of a variety of health care indicators across the country, resulting in a large geospatially multivariate data set. Current visualization tools for such data sets make it difficult to make multivariate comparisons and show the geographic distribution of the selected variables at the same time. Community Health Map is a web application that enables users to visualize health care data in multivariate space as well as geospatially. It is designed to aid exploration of this huge data repository and deliver deep insights for policy makers, journalists, consumer groups, and academic researchers. Users can visualize the geospatial distribution of a given variable on an interactive map, and compare two or more variables using charts and tables. By employing dynamic query filters, visualizations can be narrowed down to specific ranges and regions. Our presentation to policy makers and pilot usability evaluation suggest that the Community Health Map provides a comprehensible and powerful interface for policy makers to visualize health care quality, public health outcomes, and access to care in an effort to help them to make informed decisions about improving health care.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

As the health care debate raged on in the Senate, there were numerous opinions about what is wrong with health care in the U.S. While everyone agrees that the cost of health care is exploding for both individuals and businesses, the reason why costs rise but quality does not necessarily follow suit appears to be a much more open question. This is further complicated by the fact that costs of health care are rising more quickly in some parts of the country than in others. Portions of the health care system are plagued by waste, inefficiency, and in some cases, fraud (DeMott, 1997; Himmelstein, Wright, & Woolhandler, 2010; Meier & Homann, 2009). With such a large amount of money at stake, commentators on all sides want to ensure that funds are used efficiently and effectively. Medicare is a social insurance program administered by the Health and Human Services (HHS) department of the United States government, and forms a major part of the fiscal expenditure every year. Although detailed metrics are available on how the Medicare money is being spent across various states, Hospital Referral Regions (HRRs) and

counties, only a limited number of tools are available for exploration and visualization of this huge repository of data (Fisher, Bynum, & Skinner, 2009) to aid the policy makers. Community Health Map is a web-based tool designed to visualize health care data that can be of county level or HRR level granularity. During development of this tool, HHS staff provided continuing feedback, along with the raw data used to create the visualizations. The Community Health Map is designed for users with experience in the health care industry and it does not require programming experience. Users should be able to achieve these goals:

1. Visualize health performance, access, and quality indicators for their county or HRR (Hospital Referral Region).
2. Filter the regions by demographic characteristics (such as age, income, poverty, or education). The filters also allow users to compare their zone of interest to the national average.
3. Produce visual results for presentations that would spur policy makers to take action to improve the health and quality of health care in their regions.

The contribution of the Community Health Map is that it enables a wider range of non-technical users, such as mayors, local health officials, journalists, and consumer groups, to dynamically explore the geospatial distribution of health data. The map uses easy to understand color coding and provides detailed information in a multivariate table

* Corresponding author.

E-mail addresses: awalin@cs.umd.edu (A. Sopan), angela@cs.umd.edu (A.S.-I. Noh), skarol@umd.edu (S. Karol), prosenf1@umd.edu (P. Rosenfeld), ginnah@umd.edu (G. Lee), ben@cs.umd.edu (B. Shneiderman).

0740-624X/\$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.giq.2011.10.002

Please cite this article as: Sopan, A., et al., Community Health Map: A geospatial and multivariate data visualization tool for public health datasets, *Government Information Quarterly* (2012), doi:10.1016/j.giq.2011.10.002

and chart simultaneously. Users can compare map regions for an overview and then go to the table and chart for more detail.

The Open Government Directive has produced many initiatives to provide public access to many government datasets, which has triggered a dramatically increased interest in visualization tools to explore these datasets. A prime requirement has been that these visualization tools be powerful enough to enable significant discoveries, while easy enough to use by non-programmers with only moderate technical skills, such as policy makers, journalists, consumer groups, and diverse academic researchers.

This paper reviews existing tools to visualize health care data, along with their capabilities and limitations. Next, it briefly describes the structure of raw data and design inspiration. This is followed by the interface description, and by sample insights obtained with the Community Health Map. Finally, it details the feedback obtained from a usability evaluation conducted with the domain experts, both from the Department of HHS as well as three professionals in the health care industry. The paper concludes with the future directions for this tool.

2. Related work

Data visualization has been used as an important tool to gain insights into health care data sets, which are typically multivariate, discrete, and at different granularity levels. Tominski, Schulze-Wollgast, and Schumann (2008) draw attention to the importance of multivariate data representation with respect to time and space. Showing geographically distributed health related data in a choropleth map using sequential color scheme has been proven very useful for data driven knowledge discovery (Koua & Kraak, 2004; Maceachren, Brewer, & Pickle, 1998). While designing visual analytics tools for healthcare data, it is crucial to maintain simplicity as the targeted users need not have technical expertise. Gil-Garcia and Pardo (2005) identified usability and ease of use as two notable technical challenges for successful e-government. Usability has become a stronger design consideration in government website design producing critical discussions (Fedorowicz & Dias, 2010) and extensive guidelines (Donker-Kuijjer, de Jong, & Lentz, 2010; National Cancer Institute, 2010). Previous works have employed different desktop applications designed for

multivariate analysis to study health care datasets (Keahey, 1998; Madigan & Curet, 2006). However, most of desktop tools for multivariate data visualization typically require users to spend a significant amount of time learning the tool. To have better impact of the available public data, access, usage, and impact of the service are important factors for the end-users (Verdegem & Verleye, 2009) and for these purposes a web-based tool is a better fit. Unfortunately, there are few web applications for health care data visualizations. Among these, most of such tools using geospatial visualizations only have the capability to compare a couple of variables at a given time (Fisher et al., 2009).

One of the most popular tools in this category is the Dartmouth Atlas (Cooper, 1996; The Dartmouth Institute, 2010), which is shown in Fig. 2. The atlas features an interactive map of the United States and divides up the regions based on the states and the Hospital Referral Regions (HRRs). It is designed to visualize only two variables, which can be selected by a drop down menu—either the total reimbursements for the year 2007, or the annual growth rate for the selected region from 2001 to 2007. The map is then colored in five different gradients, depicting the geographical variation for the selected variables. Selecting a particular state or HRR using the cursor also shows the average for the selected region and the corresponding national average in a new layer.

A modified and slightly improved version of the Dartmouth Atlas is available on the *New York Times* website (Regional differences, 2007). Although the interactive map available on this tool only features the HRRs and not the states, variables from three broad categories—reimbursement, enrollees, and surgery rates are visualized. Like the original Dartmouth Atlas, only one variable can be visualized at a given time, and rolling the mouse over a certain region compares the selected variable value of that region to the state average and national average using bar charts. The Health Statistics Map (Plaisant, 1993) uses dynamic query to visualize epidemiological data provided by the National Center for Health Statistics (NCHS) on the U.S. map. When the Interactive Cancer Atlas (InCA) (Interactive Cancer Atlas, 2010) shows only one variable in the table for all the states, the Health Profiles Interactive (Association of Public Health Observatories, 2010) shows multiple variables in the table but only for a selected region along with the map view (Fig. 1).

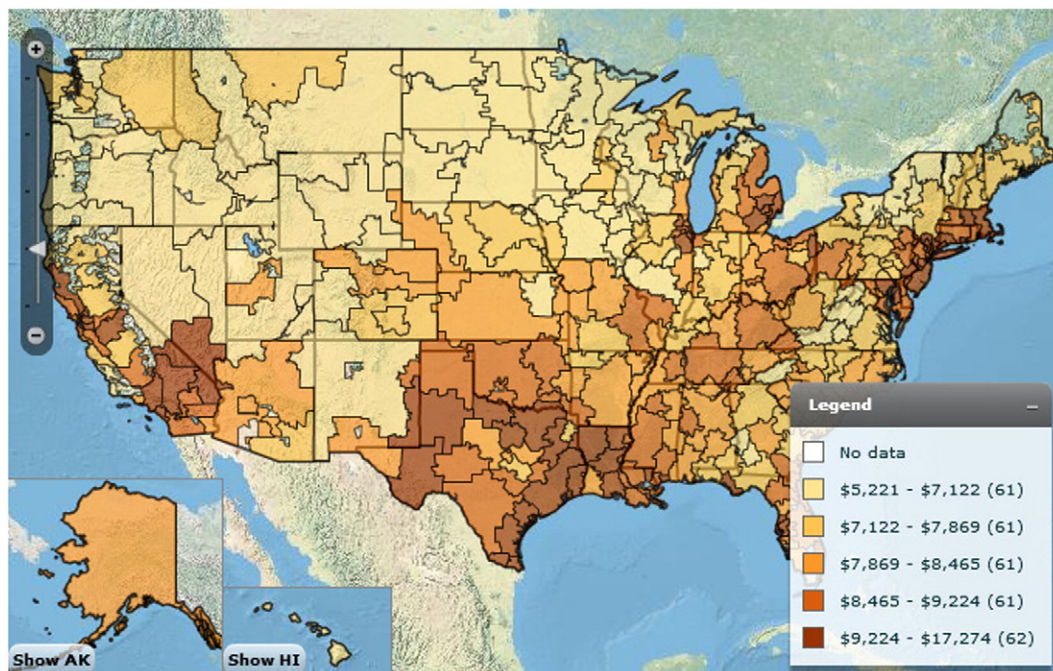


Fig. 1. Dartmouth Atlas showing medicare reimbursement per enrollee distributed over the Hospital Referral Regions.

ToolName	Type	Features	TargetAudience
Dartmouth Atlas	Web Application	<ol style="list-style-type: none"> 1. Map showing HRRs and States. 2. Allows one variable to be visualized at a time. 3. Data from only two variables are available. 	Health care specialists and researchers.
Health Statistics Map	Desktop application	<ol style="list-style-type: none"> 1. Map showing the states. 2. Sliders control the map animation. 3. Temporal changes in a geographical region can be visualized. 4. Uses dynamic query. 	Epidemiologists doing research on cancer.
Hospital Compare	Web Based tool	<ol style="list-style-type: none"> 1. Allows comparisons of up to three hospitals in a given county and three medical conditions, using a drop down menu. 2. Comparisons are not allowed for different geographical regions. 3. No geographical map. 	Patients and general internet audience.
HealthVis PCP	Desktop Application	<ol style="list-style-type: none"> 1. Map as well as scatter plots for visualizing multivariate data. 2. No dynamic query filters. 	Health care researchers.
Pennsylvania Cancer Atlas	Web Application	<ol style="list-style-type: none"> 1. Map and scatter plots. 2. Only one variable can be viewed at a time. 	Cancer researchers.
Community Health Map	Web Application	<ol style="list-style-type: none"> 1. Map, Table, Chart view. 2. Multiple variables in Table and Chart view. 3. Dynamic query filtering. 	Policy makers as well as general audience.

Fig. 2. Comparison among Community Health Map and existing tools for visualizing health care variables.

One of the web-based tools that allows comparisons of different variables and has been used in several studies is the Hospital Compare (Department of Health and Human Services, 2009; Werner & Bradlow, 2006). The tool features an interactive wizard which gives users multiple options to select a geographical region of interest and a single medical condition and returns the most appropriate hospitals based on the input criteria. The tool also allows side by side comparison of up to three hospitals on three selected variables. However, Hospital Compare lacks the ability to compare different geographical regions across the state or the country for different variables, and is restricted only to the comparison of three hospitals in the chosen location at the county level.

The ability to visualize multiple variables on the map and compare them using a table and charts at the same time can provide valuable insights which might not be possible to obtain from current tools. Although a conceptual frame-work and the potential capabilities of such a tool have been discussed in the past, there are no similar applications existing to the best of our knowledge (Lu, 2005). While Geographic Information Systems (GIS) have been found to be a useful method to visualize the health care data (Castronovo, 2008; Villalon, 1999), there also have been concerns that using interactive maps without any other form of data visualization might be subject to misinterpretation (Goodman & Wennberg, 1999). The HealthVisPCP (Edsall, MacEachren, & Pickle, 2001) uses a parallel coordinate plot and scatterplot linked to a GIS. Pennsylvania Cancer Atlas (Bhowmick, Robinson, Gruver, MacEachren, & Lengerich, 2008) deals with the cancer data of Pennsylvania and includes coordinated table, scatter plot, and map views; however, the table can show data of only one type of cancer at a time. As different geographical regions can have different demographical distributions, only the choropleth map may not suffice to provide a deeper understanding for hypothesis generation. It is also very important to show how the regions with similar demographics are doing in different areas of healthcare. Maciejewski, Drake, Rudolph, Malik and Ebert, 2010 show a technique showing similar demographic region with the same color in a choropleth map, though the distribution of health data in similar regions cannot be shown in

the same map. With our dynamic query filtering, we provide a way for the users to select similar regions according to their demographics and visualize the healthcare data for those regions.

Our Community Health Map gains its inspiration both from the Dartmouth Atlas and Hospital Compare. Community Health Map aims to provide the combined features of a simple GIS application having the capabilities of comparing multiple variables at any given time on a map and showing the insights using conventional graphs and tables as well as enabling exploration of data such as zoom, pan, filtering, and access to raw data. Fig. 2 presents a tabular comparison of the existing tools for visualizing health care variables with a Community Health Map.

3. Data and design

Community Health Map manipulates two types of datasets: health-care data and demographic data. These data sets are combined with geospatial data sets which define the boundaries of regions such as counties in order to generate the map. The initial data set came from several sources and in different formats. The geospatial data was distributed as shape files; county level shape files came from the U.S. Census TIGER/Line website.¹ The demographic data used for the filter variables came from U.S. Census datasets distributed in Excel format compiled at the county level. The data filtering mechanism allows users to select counties based on specific demographic variables. The health care data are also broken down to county or HRR level and they were provided by Patrick Conway (Chief Medical Officer at the Department of Health and Human Services) in Excel format. For the final prototype we have used eleven health care variables which are categorized into three functional areas:

- Quality
- Accessibility
- Public Health

¹ <http://www.census.gov/geo/www/tiger/>.

Data for all these variables is derived from the Center for Medicare and Medicaid (CMS) and the Community Health Status Indicator (CHSI) databases. This dataset is a modified and sanitized version of the data provided by the Community Health Status Indicators project.² Before any of this data could be used, all of it was loaded into the PostGIS relational database and the health data and demographic data was cross referenced to the geographic boundaries. The tool can be extended to accommodate any number of health care variables along with demographic variables for filtering purpose. While the data from CHSI were broken down to county level, the CMS data were of HRR level granularity.

One of the most challenging aspects of this project is that the provided data is in two different geographic resolutions. The data filtering mechanism should allow the user to select counties or HRRs based on specific demographic variables. The filter variables were obtained from census data sets which are compiled at the county level. Many of the interesting health data sets were at the HRR level, which is much bigger than a county and does not respect state, county, or ZIP code boundaries. The problem is further compounded by the fact that selection of a region on the map is performed by searching for a ZIP code, which is smaller than a county and an HRR. The question quickly arose: if the dynamic filtering variables are at the county level and the variables to be drawn in the map are at the HRR level, how does one filter HRRs based on county-level variables? For example if we want to visualize on our map the variable *Hearth Attack Mortality rate* which is an HRR level variable, and we also want to filter out the areas where percentage of population over 65 (which is a county level variable) is higher than national average, we will face a problem of filtering appropriate areas. In order to solve this problem, an algorithm was used to compute a population-weighted average of a demographic variable based on the intersections of counties and HRRs. The algorithm is as follows:

For each HRR, run a geospatial query to determine all counties that intersect this HRR

For each intersecting county

 Compute the area of intersection between the county and the HRR

 Compute the total area of the county

 Given the population of county P_{county} , $P_c = (\text{Area}_{\text{intersection}} / \text{Area}_{\text{county}}) \times P_{\text{county}}$

Sum all of these population portions to get the total HRR population, $P_{\text{tot}} = P_{P_c}$.

Finally compute the ratio P_c/P_{tot} and use this as the scaling factor for each filter-variable, then sum over all the counties that intersect this HRR

HRR value for the filter – variable = $P_{(P_c/P_{\text{tot}})} \times \text{CountyLevelValue}$.

Therefore we determine all counties that intersect an HRR, calculate the population of the HRR by summing up the population of all the areas that fall inside that HRR, and compute the value of the filter-variable for the HRRs. This way we can determine the percentage of people over 65 for each HRR though this filtering variable is originally recorded at county level granularity.

Another challenging aspect of this project is to combine health care data with demographic data. The data filtering mechanism should allow the user to select counties based on specific demographic variables. The filter variables were obtained from census data sets which are compiled at the county level. The problem is further compounded by the fact that selection of a region on the map is performed by searching for a ZIP code, which is smaller than a county. Instead of percentages, several

variables' value (for example, Uninsured people) were given in absolute number which is infeasible to compare and detect correlation; therefore, we have derived the percentage by dividing the value with the population count of that county.

Before starting the final implementation we have analyzed the data to see possible correlation among the variables using HCE (Seo & Shneiderman, 2006) to remove variables that may not be interesting in the domain of health care. For example, we have not used the data for homicides in this system. This way the users will not be overwhelmed by lots of variables; rather, they can pay attention to the important ones. Again there were missing values in the provided data; we have used a particular hue (gray) in the map to indicate the unavailability of value of the selected variable for those counties.

One of the design goals of Community Health Map was to simplify usability. Since the tool broadly targets the people in charge as well as people concerned about the health care system of their communities, we do not make any assumptions about technical prowess of these individuals. For this reason we expect that a web application would be the best solution since our target users are not computer experts. Any user with a modern web browser can view the data with different representations such as a heat-map, tables, and charts, which are generated without any additional plugin, binaries to download, or datasets to import. Another important design goal of Community Health Map is to make it an app in the true sense of the word. Community Health Map does far more than just simply push a few tables and map the relevant data for users to look at. Community Health Map can be qualified more as a rich internet application in that all interactions happen within a page itself. Users are free to zoom and pan around the map, sort table columns, or pick a new variable with charts to explore all without ever leaving the page.

4. Community Health Map user interface

Community Health Map embodies three coordinated visualizations: a map display, data table, and chart view. This allows users to see the geospatial as well as multivariate features of the data. Users inputs are provided to the system in three ways: the ZIP code entry box, selection panel, and dynamic query filter panel. The visualization dynamically changes according to these inputs. Fig. 3 shows the Graphical User Interface (GUI) of Community Health Map. Essentially, the interface can broadly be divided in four distinct parts:

- The top left panel consists of a pannable and zoomable interactive map of the United States.
- The middle left panel contains hinted double-sided filters that allow the users to filter the map and data tables simultaneously.
- The bottom left panel consists of data tables. The raw data is displayed in a sortable table with cells colored based on the quintiles of the data.
- The right side vertical selection panel contains a list of variables, classified into three categories, which can be selected using check boxes. Doing a mouse-over on any variable name shows its description, granularity level of the variable, and source of its data in a small bubble. The ZIP code entry box lets users provide a ZIP code that will be selected on the map.

4.1. Map view

When users start the Community Health Map, the ZIP code entry box will be shown. Users will be asked to put an area zip code which they are interested in (see Fig. 4). Then, the arrow shielded black box will come up indicating the selected area in the map. To visualize the data on the map, users begin by clicking the map-icon beside a variable to draw the heat map overlay. Once a variable is selected, the label of the variable becomes bold and the geographical distribution is shown by heat maps varying in 5 different shades.

² <http://www.communityhealth.hhs.gov/HomePage.aspx>.

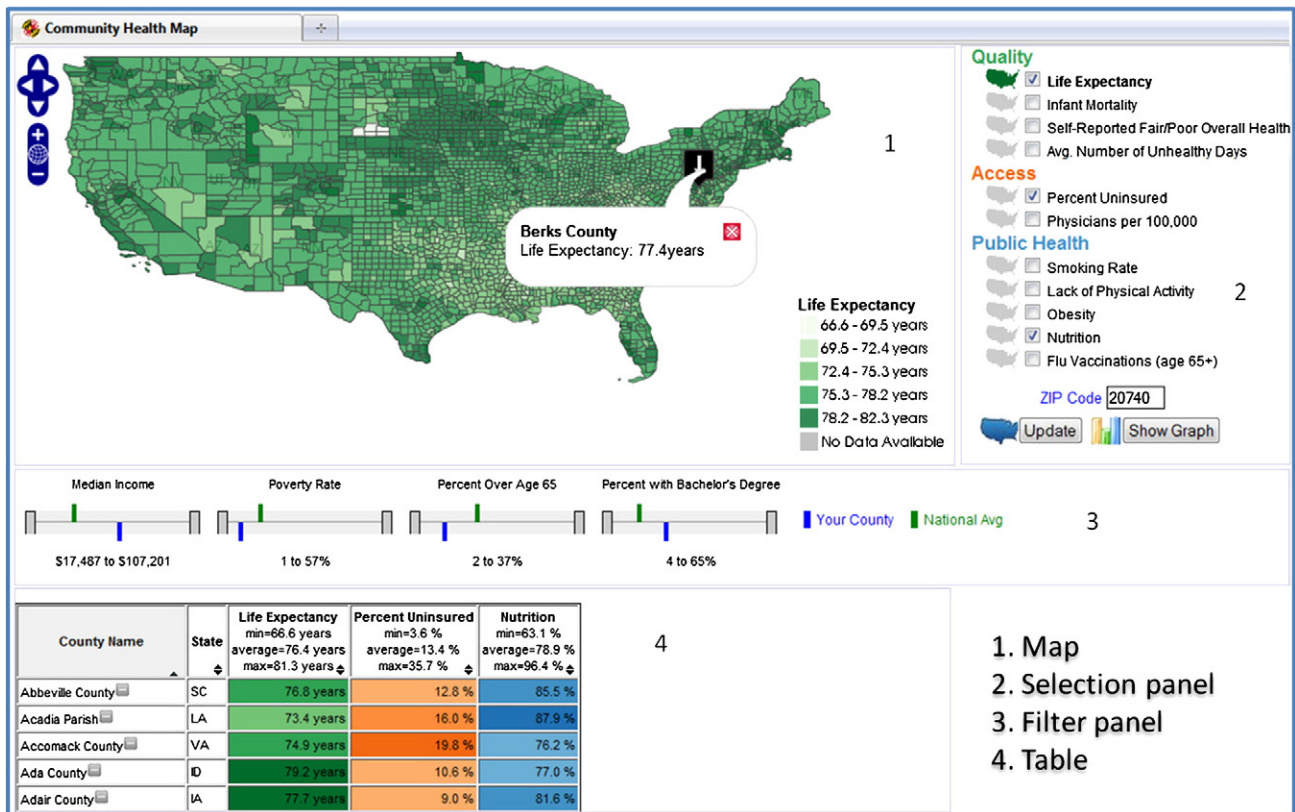


Fig. 3. The Community Health Map interface consists of a map (top), hinted double-sided sliders for map and table filtering (middle), a color coded table (bottom), and the selection panel (right). The selection panel allows users to click the link to render a layer on the map, or check boxes to display a table with the selected variables.

Users can only select one variable at a time to be shown on the map. Users can zoom (using the map controls or mouse wheel), pan, and obtain the variable value of a region by clicking on the map. Clicking on any region on the map shows the name of the selected county, along with its value for the chosen variable in a small bubble.

4.2. Table and chart view

To view multiple variables, the Community Health Map provides two additional views apart from the map: a chart view and a table view. Users can use the checkboxes in the selection panel to check the variables of interest and click the Update button (see Fig. 3). The table shows the values of selected variables within the filter range so that the map and table always stay consistent. Each row of the table represents data from a county or an HRR. If users select both county and HRR level variables then two tables will be generated in two different tabs, one with only county level variables and another one with HRR level variables. By clicking the tabs users are able to switch between the tables and access the desired one. Cells are also color coded for easy comparisons with other variables and

geographical regions. More intense colors in a cell indicate higher value for a variable. The table cells always mimic the map colors for consistency. The table columns are sortable to provide easy comparison of values. The county to which the area with the given ZIP code belongs is highlighted in the table with a yellow background (see Fig. 10). When the selected county is filtered out by dynamic query filter, however, the highlighted color gets to become gray from yellow. Users can also create a bar chart for the geographic area of interest. The show graph button triggers a modal dialog containing a chart to hover above the page while dimming the interface underneath (see Fig. 5). Each chart consists of a blue bar showing the value for the selected variable on that county and a green bar showing its corresponding national average as a benchmark. The charts are generated dynamically, so users can create charts showing various variables for the selected region and their respective national average. The charts are particularly convenient for saving a snapshot of an insight or just having one more avenue to perform comparisons.

4.3. Dynamic query filters

One of the key features of Community Health Map is its ability to employ dynamic query filters. Four filtering options represented by double-sided sliders are present right below the map. The filters represent the Median Income, Poverty Rate, Percent of People Over Age 65, and Percent of People with Bachelors Degree. When the users change the range of the filter variables, both the map and data table update simultaneously by removing entries that are out of the selected range. The double-sided sliders are hinted with benchmark values such as the national average with green and the currently selected region's value with blue (see Fig. 6). By selecting ranges that surround their position on the filter, the users can get a sense of which regions are similar to their own and immediately see how their colors compare to the colors that are remaining on the map and in the table.

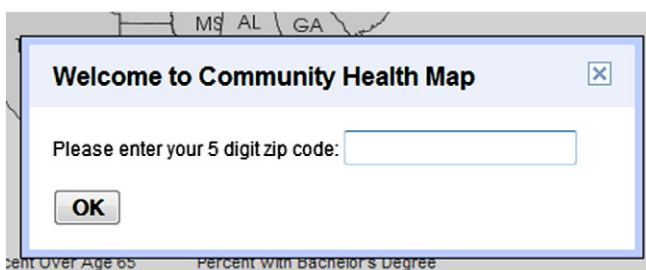


Fig. 4. The welcome screen asks users to put a ZIP code representing their region of interest. Then, the tool will find a matched county for that ZIP code.

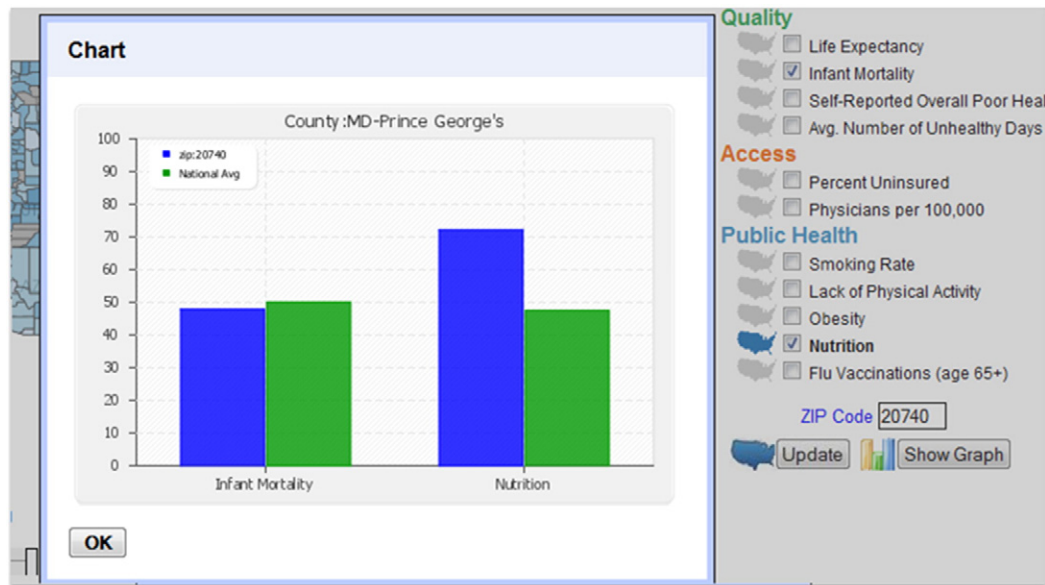


Fig. 5. Another method for comparing multiple variables with their national average is through the graphs. The modal dialog floats above the page so that users do not lose the context of what they were doing previously. Users can save the graphs to record their insights and send to others. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This is an important feature since users might try to use socioeconomic factors as excuses to rationalize poor performance or high cost. The ability to filter out data based on such demographic variables can be helpful in removing the credibility of such excuses. Initially the filter ranges are adjusted to national minimum to national maximum values for each demographic variable. When users move the sliders, the filters show the values for the current selected range below the filters. Each time users pick a new ZIP code, the your region marker shifts to the corresponding position on all filter sliders.

4.4. Color coding

Each category of variables is assigned a particular hue: green for Quality, orange for Access and blue for Public Health. When users are looking at the table they can easily understand which category a column belongs to by the color. Each variable value range is divided into five equal length portions and given a shade of the assigned hue so that difference in intensity corresponds to difference in the values. The sequential color schemes are derived from ColorBrewer Brewer (2004).

5. Application example

This section presents insights that we were able to achieve using the Community Health Map. Having a geographical heat map is effective for showing variables that appear to be clustered around a particular region.

Fig. 7 shows an overview of the entire U.S. showing the Average Number of Unhealthy Days in the Last Month. There is a very clear dark spot that develops around eastern Kentucky into West Virginia

(circled in red). The counties in this area all fall into the 4th and 5th quintiles and no other place on the map has such a dark area. A professor at University of Maryland pointed out that Kentucky and West Virginia are mining centers in the U.S. West Virginia and Kentucky have several underground coal mines, which is a plausible explanation of poor health. Furthermore, after zooming into this area and switching the map to life expectancy (Fig. 7), it becomes very clear that the area of poor health corresponds to low life expectancy. Since the map re-renders completely in place, users can keep their eye on the area of interest to be able to identify correlations with ease. Users can then choose any other variable to draw on the map and look for other correlations. While it is somewhat intuitive that coal miners might have poorer health and lower life expectancy, we are encouraged that the tool allows users to gain these types of insights in a straight forward way.

The map filtering mechanism is also extremely effective in comparing and contrasting different segments of the population, such as wealthier vs. poorer areas. Wealthier areas were those with mean median income above national average and poverty rate below national average, while poorer areas were below the median. These groups were selected by adjusting the double-sided filters for Median Income and National Average. Fig. 8 shows how wealthier and poorer counties report their overall health status respectively. These two maps show the same region of the United States, but we can see that the colors shown on the wealthier map are much lighter than the colors shown on the poorer map. This implies that wealthier counties tend to have better health than poorer ones. Again, while this insight may seem to be obvious, it demonstrates the ability of users to easily compare the quality of health care across socioeconomic classes, which is more of a challenge with other tools. The

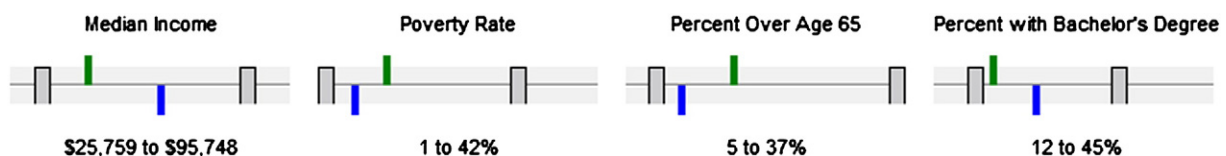


Fig. 6. Double-sided slider for dynamic query filtering, annotated with the national average (green marker) of a demographic variable and the average for the county (blue marker) currently selected. The current range of the selection is shown below each filter; for example, in this figure, Poverty Rate is selected from 1 to 42% so counties with a poverty rate of more than 42% will not be shown in the map or the table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

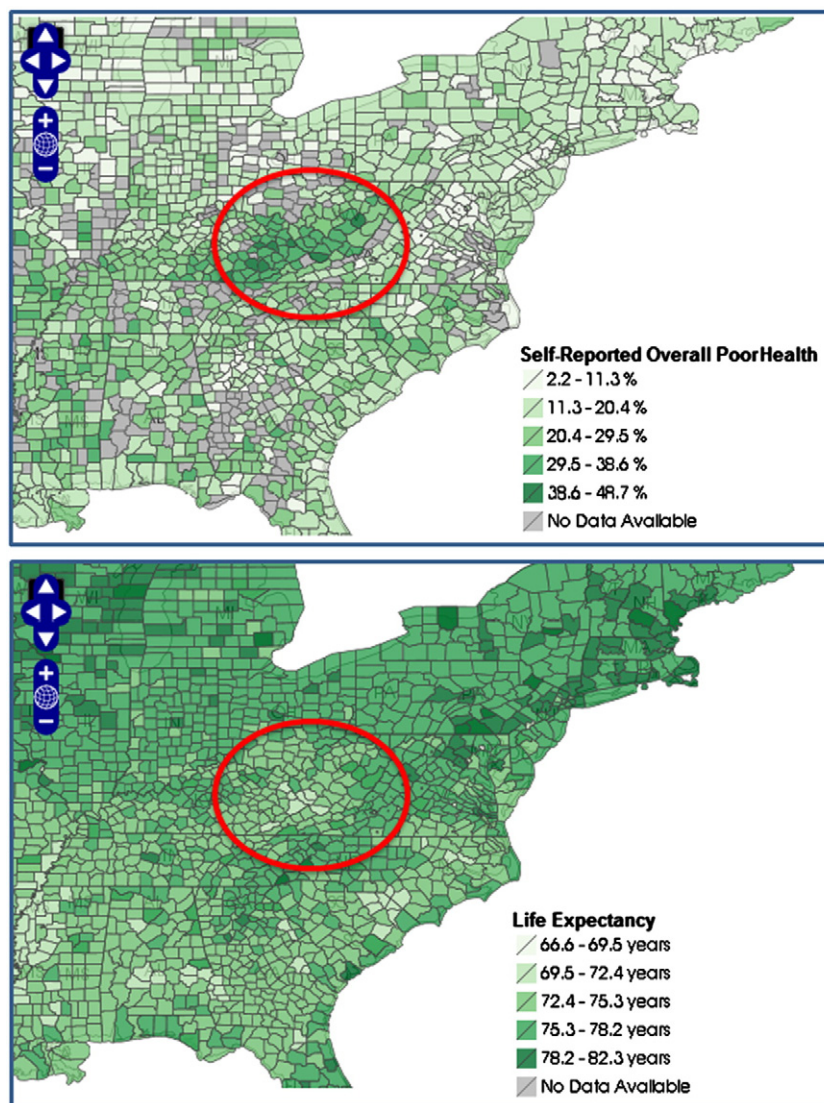


Fig. 7. Top: the circled region having poor health around eastern Kentucky and West Virginia. The dark region circled in red shows the high number of Self-reported Overall Poor Health. Bottom: same area of the map showing Life Expectancy. Lighter color indicates lower value of life expectancy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

filtering mechanism is also useful for trying to pinpoint whether specific demographic variables determine poor quality or high costs. These four variables should allow users to dissect the United States in interesting ways to look for interesting or unexpected patterns.

Fig. 9 shows the distribution of life expectancy over south central states. Life expectancy is relatively very high in Nebraska, Minnesota, Iowa, and Colorado. However, four counties in the middle point of the map show distinct white spots, which represent very low life expectancy. In the detail view, we zoomed in and clicked to see the county names in South Dakota: Shannon, Bennett, Jackson, and Todd.

Then to analyze the correlation with other health indicator variables, we selected infant mortality, percent uninsured, and smoking rate variables. To compare Life Expectancy, we sorted the table by this variable. From Fig. 10, Shannon County in South Dakota, shows the standard we used for comparison. We can easily notice that top four counties with low expectancy are all from South Dakota and those are the counties with white spot from the map view. Except Bennett county, the three other counties also show a high rate of infant mortality. We assume some poor environmental health conditions could influence this outcome. We find out from the table view, that those four counties have a relatively high percentage of uninsured population and a high smoking rate compared to other counties.

We assume some health care access factors could influence these results. Therefore, for a more detailed comparison we selected two more variables, Percent uninsured and Physicians per 100,000, and generated the chart view. From Fig. 11 we find out that Shannon County has more uninsured people and fewer physicians per 100,000 people. This comparison chart supported our hypothesis. Shannon County has twice as high a percentage of uninsured population as the national average. Also it has a significantly lower rate of physicians per 100,000 compared to the national average. These results show that there should be more support on health care access to these counties to overcome poor environmental health conditions.

6. Prototype application architecture

Several open source projects such as Geoserver and OpenLayers form the basis of the overall architecture of Community Health Map and make the system easy to deploy. A block diagram showing how all the pieces fit together is shown in Fig. 12. Geoserver is a powerful map-rendering engine that is committed to strictly following standard sets by the Open Geospatial Consortium.³ Respecting standards allows

³ <http://geoserver.org/display/GEOS/What+is+Geoserver>.

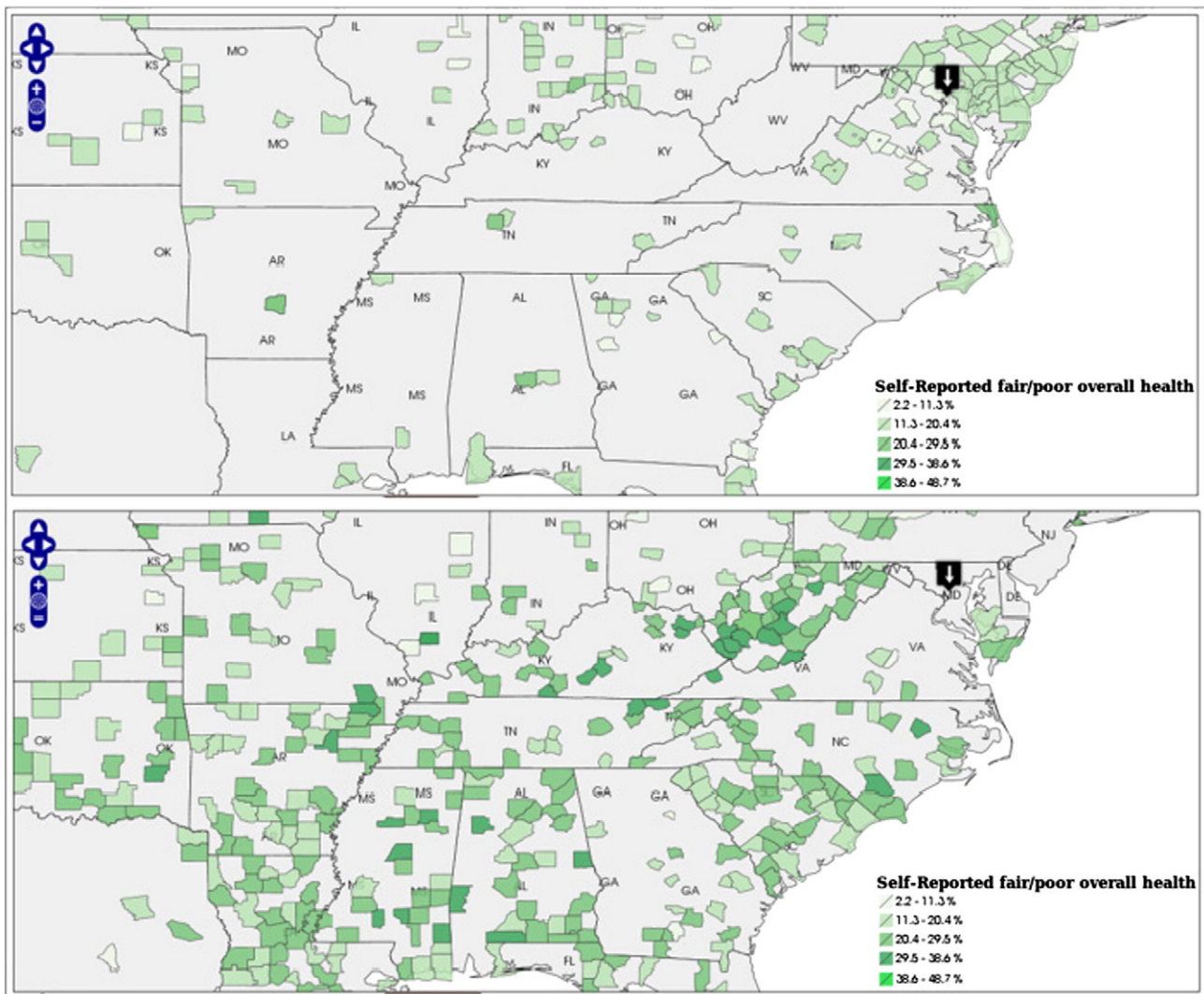


Fig. 8. Top: self-reported Overall Poor Health for wealthy counties of the U.S. (low poverty rate, high median income). Bottom: same variable for the same region, but this time only considering poorer counties. The darker gradient indicates greater percentage of self reported poor overall health.

Geoserver to easily interoperate with any other application compliant with the standard, such as Google Earth, Microsoft Virtual Earth, OpenLayers, etc. To render the maps generated by Geoserver, we have chosen an open source javascript mapping client called OpenLayers. OpenLayers⁴ implements many of the same features as Google Maps, and has a very robust set of features. To generate charts and graphs, we have used an open source PHP library called pChart⁵, which can dynamically draw a plethora of various charts and graphs. All of these tools run on the Tomcat servlet engine (for Geoserver) and the Apache HTTP server. Both of these servers are used extensively all over the internet. PostGIS⁶ is a set of geospatial extensions to the open source relational database PostgreSQL. PostGIS combined with Geoserver form the backbone of all map generation. Web browsers are notorious for being heterogeneous when it comes to their implementations of javascript engines, CSS renderers, and DOM functionality. In order to ameliorate this situation, we leveraged a recent open source javascript library from Google called Closure.⁷ This library has an impressive array of UI elements and other advanced features that are used on a daily basis by Google, therefore it is rigorously tested both for

correctness and cross-browser compatibility. Finally, to complete the picture, all of this software can run on the GNU/Linux operating system, arguably the most widely used open source project.

As this is a prototype we did not address the issues of scalability and security at this stage which could be crucial for a real system that can be accessed by people across the country simultaneously. Again, as the data used to demonstrate our system are only one year, but both health care and demographic data are changing year by year, so a full functional system should incorporate the capability of easy update of data to show the most recent situation, and even better to have the option to select a time frame to visualize health care data during that period.

7. Usability evaluation

During the three months of development of Community Health Map we worked closely with Todd Park (the Chief Technical Officer of HHS), Karen Milgate (Director of Centers for Medicare and Medicaid Services), and Patrick Conway (Chief Medical Officer of HHS) and implemented their suggestions. Their assessment of primary users and use cases guided our development. After completion of each milestone we sought their feedback and made changes accordingly so that the Community Health Map would be useful as well as simple to use. In our final demo to Bill Corr (Deputy Secretary of HHS), Todd Park, and his team,

⁴ <http://openlayers.org/>.

⁵ <http://pchart.sourceforge.net/>.

⁶ <http://postgis.refractory.net/>.

⁷ <http://code.google.com/closure/library/docs/gettingstarted.html>.

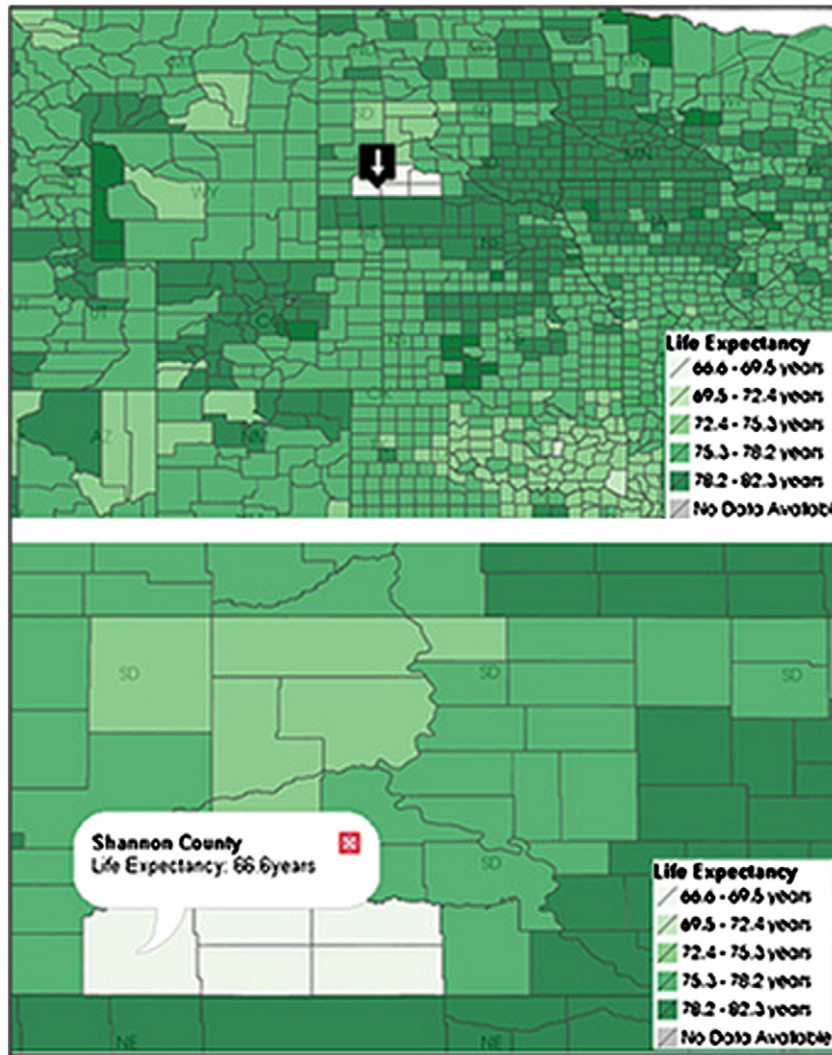


Fig. 9. Top: overview of Life Expectancy map for north central states. Bottom: zoomed version of the map across South Dakota.

County Name	State	Life Expectancy min=66.6 years average=76.4 years max=81.3 years	Infant Mortality min=0.0 deaths average=7.3 deaths max=23.7 deaths	Percent Uninsured min=3.6 % average=13.4 % max=35.7 %	Smoking Rate min=3.6 % average=23.4 % max=46.2 %
Bennett County	SD	66.6 years	3.3 deaths	24.1 %	28.8 %
Jackson County	SD	66.6 years	18.4 deaths	27.2 %	20.2 %
Shannon County	SD	66.6 years	15.7 deaths	29.3 %	42.4 %
Todd County	SD	66.6 years	14.9 deaths	30.9 %	39.9 %
Baltimore city	MD	69.6 years	11.1 deaths	17.1 %	27.8 %
Marlboro County	SC	69.6 years	10.6 deaths	19.2 %	35.7 %
Petersburg city	VA	69.6 years	12.6 deaths	18.1 %	26.0 %
Phillips County	AR	69.8 years	10.4 deaths	26.0 %	20.9 %

Fig. 10. Color coded table shows the correlation among four variables from quality (colored green), access (colored orange) and public health indicator (colored blue). The table is sorted by Life Expectancy. Also, the selected county, Shannon County is highlighted in yellow, which makes users find out easily where their region of interest lies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

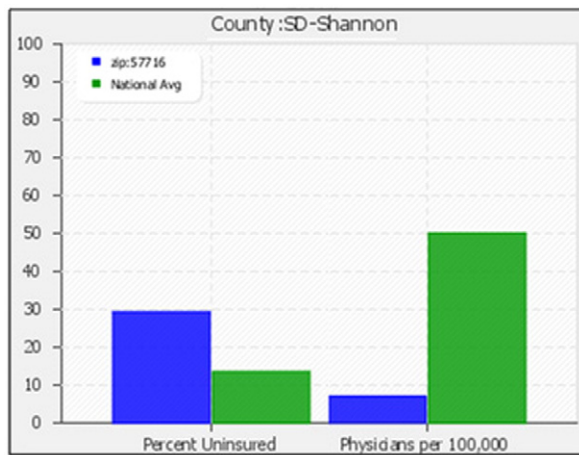


Fig. 11. Several health indicator values of Shannon County are compared to the national average. Blue shows the values of Shannon county and green shows the national average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we received positive feedback on the usability and flexibility of the tool. In spite of their brief use of Community Health Map, they found it to be superior to the many interactive maps they had studied. During our final demo Todd Park remarked:

This is incredibly cool, I think it is spare, elegant, and intuitive. The White House and HHS see this as a powerful seed for changing health care in the United States.

An initial usability evaluation with a user interface expert and some of graduate students specializing in Health Informatics at the University of Maryland, College Park also revealed interesting results. Subjects were given a four minute video demonstration of the Community

Health Map and allowed to use it for 20 min. They were asked to perform five tasks incorporating the use of filter, interactive map, table, and chart:

- Find the areas with similar income to their own ZIP Code (use the dynamic filters to filter out regions that are too high or low income compared to own ZIP code area).
- Identify which one of these areas is richest.
- Find which health care variable in their selected ZIP code area needs the most improvement as it is far below the national average (use chart to draw several variables along with the national average value in bar chart).
- Find which county is performing best for some health care indicator variable and then to find the top 10 other counties that are next to that county for that variable (sort the table according to that variable, select top 10 rows).
- Find which county has most '% of Uninsured People' among all counties that have a poverty rate above national average (coordinated use of dynamic filter to draw the map and the table).

We designed the tasks especially to evaluate whether the users are comfortable enough using different aspects of the interface like clicking on the map to see the value of selected variable, using double ended slider for filtering, sorting tables, and using the chart for comparison. We asked them to think aloud so that we could keep record of their reactions. Finally they were asked to explore the tool on their own and work with variables of their choice, based on their expertise and research interests. They also provided their general feedback at the end of the evaluation.

While one of the subjects had a few minor suggestions about the usability of the interface, they agreed that a small demonstration before the sessions was necessary for them to use the tool effectively. None of the subjects reported any major usability challenges. They all were able to perform all the tasks easily. Especially, they all liked that this tool is a web-based program which means no additional effort for using it such as a difficult installation process. One of the

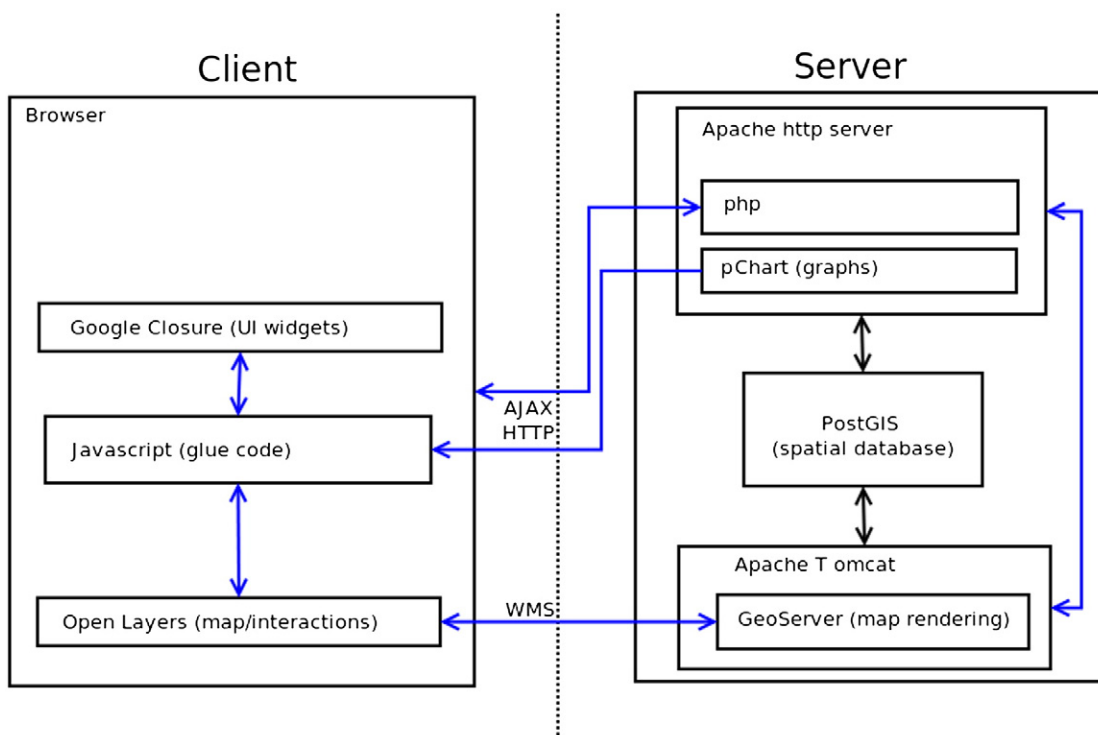


Fig. 12. Block diagram of the application architecture.

subjects, a Health Informatics consultant in the Washington D.C. area, said:

This tool is simple, innovative and powerful. I haven't come across anything like this before and hope professionals in my field will have access to it sometime soon.

Another subject, who is also a PhD candidate in Public Health wanted to verify the hypothesis about the correlation between the lack of physical activity and obesity. The subject was easily able to do his research using the tables which show both of variables at the same time for not only all areas but also only the areas which he was more interested in by filtering, and comparing those numbers with the national average. Further, this subject also appreciated the simplicity of the interface, along with the ability to filter the results using the dynamic query filters. For long tables (where each row of the table represents a county) scrolling up and down is necessary in the current implementation, and one of the subjects mentioned this as a problem.

After getting suggestions from our user interface expert, we modified the interface to make it more intuitive to use. We made the color of the selected county's bar in the chart the same as the color of hinting bar for the selected county's demographic indicator in the query filter sliders (it was different before). We also added the feature that mousing over a variable name in selection view will show a tool tip of the variable's description and data source.

8. Conclusion and future directions

The Community Health Map was designed as a simple web application to enable a broad range of users to explore health care data for the geographical regions of their interest. Some of the key features of Community Health Map are its ability to:

- analyze data at the level of counties
- visualize the data simultaneously on a geographical map, charts, and tables
- use dynamic query filters to selectively visualize healthcare data of regions with similar demographic features and identify who is doing better in this aspect (a region can improve its health care condition by following the strategy of other regions with similar demographic patterns but better health care conditions)
- dynamically change the visualization of the data by selecting a specific range of income, poverty levels, age, and education of the chosen geographical regions
- compare multiple geographical regions as well as variables side by side and
- incorporate data of both county level and HRR level.

The features, along with the simplicity, make Community Health Map a powerful tool to analyze the health care data in United States and assist in discovering important trends and valuable insights. Our clients, the Department of Health and Human Services, envision this as a tool that can spur action by decision makers and streamline health care spending.

This prototype should act as an inspiration to more complete implementations what would include more demographic, outcome, and cost variables, as well as richer geographic region selection features. An expanded implementation would support linking to other datasets, collaboration among users, and history-keeping to enable backtracking during exploration. Moreover, the challenge of managing and coordinating two levels of granularity (HRR and county) demands more attention from the people in charge to organize and manage health care data.

Further testing would be needed to understand user problems and user requests for additional features. While the current design

emphasizes visualization which presents some difficulties for users with visual disabilities, an improved version could support spatial discovery with sonification and tabular exploration with text-to-speech tools. Our user evaluation was designed to see if the users can perform the designated tasks properly and easily; more study can be performed to observe which types of visualizations and interactions are actually important to policy makers so that those particular features could be improved further.

Although Community Health Map was designed for the health care data, it can easily be used to analyze any multivariate data set with a geographic spread, and could be of particular interest for users in the education or finance domains.

Acknowledgments

We would like to thank Patrick Conway and Karen Milgate for taking the time to supply us with clean datasets that were easy to use and Todd Park who provided the initial inspiration for this project and his continued enthusiasm over the tool made it that much more satisfying to develop. Finally we thank Catherine Plaisant and other members of the Human-Computer Interaction Lab of University of Maryland for their valuable suggestions on the design of our user interface and usability test.

References

- Association of Public Health Observatories. (2010). Health profiles interactive. Retrieved from http://www.apho.org.uk/default.aspx?QN=HP_INTERACTIVE
- Bhowmick, T., Robinson, A., Gruver, A., MacEachren, A., & Lengerich, E. (2008). Distributed usability evaluation of the Pennsylvania cancer atlas. *International Journal of Health Geographics*, 7(1), 36.
- Brewer, C. A. (2004). Color research applications in mapping and visualization. *Color imaging conference* (pp. 1–3). IS&T—The Society for Imaging Science and Technology.
- Castronovo, D. (2008). A conceptual framework for visualizing seasonality in spatio-temporal health data using dynamic maps. *American Public Health Association 136th annual meeting*.
- Cooper, M. M. (1996). The Dartmouth atlas of health care: What is it telling us? *Health Systems Review*, 29(3), 44–45 (47).
- DeMott, K. (1997). Healthcare practices vary widely from town to town: Regional Dartmouth atlas. *Health System Leader*, 4(1), 2–3.
- Department of Health and Human Services. (2009). Hospital Compare. Retrieved from <http://www.hospitalcompare.hhs.gov>
- Donker-Kuijjer, M. W., de Jong, M., & Lentz, L. (2010). Usable guidelines for usable websites? An analysis of five e-government heuristics. *Government Information Quarterly*, 27(3), 254–263.
- Edsall, R. M., MacEachren, A. M., & Pickle, L. (2001). Case study: Design and assessment of an enhanced geographic information system for exploration of multivariate health statistics. *InfoVis*, 159–162.
- Fedorowicz, J., & Dias, M. A. (2010). A decade of design in digital government research. *Government Information Quarterly*, 27(1), 1–8.
- Fisher, E. S., Bynum, J. P., & Skinner, J. S. (2009). Slowing the growth of health care costs—Lessons from regional variation. *The New England Journal of Medicine*, 360(9), 849–852.
- Gil-Garcia, J. R., & Pardo, T. A. (2005). e-Government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, 22(2), 187–216.
- Goodman, D. C., & Wennberg, J. E. (1999). Maps and health: The challenges of interpretation. *Journal of Public Health Management and Practice*, 5(4), xiii–xvii.
- Himmelstein, D. U., Wright, A., & Woolhandler, S. (2010). Hospital computing and the costs and quality of care: A national study. *The American Journal of Medicine*, 123(1), 40–46.
- Interactive Cancer Atlas (2010). Interactive cancer atlas. Retrieved from http://apps.nccdc.cdc.gov/DCPC_INCA accessed 11-Aug-2010.
- Keahey, T. A. (1998). Visualization of high-dimensional clusters using nonlinear magnification. *SPIE The International Society for Optical Engineering*, 3643, 228–235.
- Koua, E. L., & Kraak, M. J. (2004). Geovisualization to support the exploration of large health and demographic survey data. *International Journal of Health Geography*, 3(1), 12.
- Lu, X. (2005). A framework of web gis based unified public health information visualization platform. *Computational science and its applications ICCSA 2005* (pp. 265–268). Berlin: Springer.
- Maceachren, A. M., Brewer, C. A., & Pickle, L. W. (1998). Visualizing georeferenced data: Representing reliability of health statistics. *Environment and Planning A*, 30(9), 1547–1561.
- Maciejewski, R., Drake, T., Rudolph, S., Malik, A., & Ebert, D. S. (2010). Data aggregation and analysis for cancer statistics—A visual analytics approach. *System sciences (HICSS)*, 2010 43rd Hawaii international conference on (pp. 1–5).

- Madigan, E., & Curet, O. (2006). A data mining approach in home healthcare: Outcomes and service use. *BMC Health Services Research*, 6(1), 18.
- Meier, B. D., & Homann, D. (2009). Fraud in the health-care system from the perspective of the public health insurance companies. *Gesundheitswesen*, 72(7), 404–411.
- National Cancer Institute. (2010). Nci research-based web design and usability guidelines. Retrieved from <http://www.usability.gov/guidelines/index.html> accessed 04-Dec-2009
- Plaisant, C. (1993). Facilitating data exploration: Dynamic queries on a health statistics map. *Proceedings of the Government Statistics Section, annual meeting of the American Statistical Association* (pp. 18–23).
- Regional differences in costs and care. (2007). New York Times. Retrieved from. http://www.nytimes.com/ref/business/20070611_GAP_GRAPHIC.htm accessed 4-Dec-2009
- Seo, J., & Shneiderman, B. (2006). Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(3), 311–322.
- The Dartmouth Institute. (2010). Dartmouth atlas of health care. Retrieved from. <http://www.dartmouthatlas.org/> accessed 11-Aug-2010
- Tominski, C., Schulze-Wollgast, P., & Schumann, H. (2008). Visual methods for analyzing human health data. *Encyclopedia of Healthcare Information Systems*, 1357–1364.
- Verdegem, P., & Verleye, G. (2009). User-centered e-government in practice: A comprehensive model for measuring user satisfaction. *Government Information Quarterly*, 26(3), 487–497.
- Villalon, M. (1999). Gis and the internet: Tools that add value to your health plan. *Health Management Technology*, 20(9), 16–18.
- Werner, R. M., & Bradlow, E. T. (2006). Relationship between Medicare's Hospital Compare performance measures and mortality rates. *Journal of the American Medical Association*, 296(22), 2694–2702.

Awalin Sopan is a graduate student at the Department of Computer Science at University of Maryland and a research assistant at the Human-Computer Interaction Lab there. Her primary research interests are information visualization and network analysis. She obtained her bachelor's in Computer Science from Bangladesh University of Engineering and Technology, also worked as software engineer in the field of content management system for online publication. Website: <http://www.cs.umd.edu/~awalin/>.

Angela Song-le Noh is a graduate student at the Department of Computer Science at University of Maryland. She is on the way to pursue her PhD. Her research interest is in Human Computer Interaction.

Sohit Karol is a doctoral candidate in the Department of Kinesiology, School of Public Health at University of Maryland, College Park. His research areas include motor behavior, biomechanics, human factors and neuropsychology while using the human hand as a model for investigation.

Paul Rosenfeld is a Graduate student at the department of Electrical and Computer Engineering at the University of Maryland, College Park.

Ginnah Lee is currently doing her PhD in the Electrical and Computer Engineering Department at University of Maryland, College Park, MD. And, she has been working as a graduate research assistant since Jan 2008. Ginnah got her B.S. in Computer Science and Engineering from Seoul National University, Seoul, Korea in Feb 2006. Ginnah's research interests are mainly in network field including network modeling, protocol design and evaluation, distributed systems and Delay Tolerant Network. She is also interested in Human Computer Interaction and parallel computing.

Ben Shneiderman is a Professor in the Department of Computer Science, Founding Director (1983–2000) of the Human-Computer Interaction Laboratory, and Member of the Institute for Advanced Computer Studies at the University of Maryland at College Park (full resume). He has taught previously at the State University of New York and at Indiana University. He was made a Fellow of the ACM in 1997, elected a Fellow of the American Association for the Advancement of Science in 2001, and received the ACM CHI (Computer Human Interaction) Lifetime Achievement Award in 2001. He was elected to the National Academy of Engineering in 2010: "For research, software development, and scholarly texts concerning human-computer interaction and information visualization." He was the Co-Chair of the ACM Policy 98 Conference, May 1998 and is the Founding Chair of the ACM Conference on Universal Usability, November 16–17, 2000. Ben Shneiderman's interest in creativity support tools led to organizing the June 2005 NSF workshop and to chairing the June 2007 Conference on Creativity & Cognition.