Team 007 Final Report

Introduction - Motivation

Quality oral health care is the topmost unmet healthcare need for children in the United States [1] - especially for impoverished children qualifying for public insurance. Publicly insured children are 50% more likely to lack access to healthcare in one or more areas than children covered by private insurance [1]. Currently, less than half of children and adolescents had a dental visit in the past year [6], but the recommended number of visits is two per year. This complete lapse in dental healthcare for children may be due to lack of access to dentists in their area. Almost 25% of US children living in poverty had untreated dental cavities [5] which can lead to problems with eating, speaking, and learning [4]. Increasing access and use of preventive dental care services is a health goal of Healthy People 2030 and the public health community [7]. Closing the gap of unequal access to dental healthcare due to income levels will lead to greater societal health and economic prosperity.

Problem Definition

Finding shortage areas can be difficult due to the large amount of data to parse through to determine how large the supply of dentists in an area is compared to the demand of children who need quality dental care. The goal of this project is to create a web application portal for healthcare professionals and policy makers to identify dental shortage areas for children who are eligible for public insurance (CHIP, Medicaid) with the ability of providing recommended geographical areas to focus policy maker efforts on for maximum impact.

Survey

Current research into dental shortage areas exists [3, 13] but does not address visualization of this issue, and just highlights its importance. Visualization of data related to this pediatric health need is extremely important for identifying at-risk areas that would benefit from more support/funding. The visualization of big health data, especially interactive options, is an ongoing effort for many different areas of healthcare for use by both health professionals, health policy workers, and the general public [2]. Discovering and visualizing areas in the US where there are shortages in dental healthcare could help both inform policy decisions and create a central hub where families can see which dentists in their area accept public health insurance.

From our research, there is no visualization available today that shows the supply and demand of dental care providers accepting public insurance in a specific state's county. Currently, the most popular available tools in dental care availability include https://dentaquest.com/find-a-dentist-gov/ and https://www.zocdoc.com/dentists/medicaid-357d, consumer-focused tools without visualizations that query and display line-by-line provider information with the purpose of allowing users to schedule appointments. The above tools do not help with identifying dental provider shortage areas across geographical ranges. Without any form of visualization designed specifically for public policy purposes, policy makers struggle to gather supporting information that advocates for proposed changes to increase provider availability in shortage areas.

Our approach to visualizing this information is unique as this tool designed specifically for dental care availability doesn't currently exist. While there is an existing resource that has distilled the dental care providers data and summarized important insights by state [13], it does not provide the same level of granularity we are hoping to achieve. The final product will give health policy workers relevant information to identify and approve of different strategies to increase utilization of preventative dental services among children enrolled in public insurance - like evenly distributing providers by incentivizing providers to locate to shortage areas or increasing Medicaid reimbursements.

While making sure to meet all the technical requirements of our project, we will be informing our design on other successful and user-friendly d3 visualizations [8, 15]. This approach is more useful than what already exists in this realm, because it is more granular and the existing resource [3, 13] lacks visualization. Visualization does exist for other healthcare issues, such as the nursing workforce distribution to discover underserved areas, consumer data, visualizing the distribution of individuals with AIDS [9, 11, 12], but the visualization is only on the state's particular county level, not the US as a whole, nor do they tackle multiple states. Most of them are not interactive, as well.

Healthcare systems are becoming more reliant on the intake and storing of sensitive data [10, 14]. However, this mainly applies to patient data instead of policy data, so we do not need to be as concerned about storing sensitive data. The raw data contained some identifying information for individual dentists but that data is scrubbed before entering our application.

Not only is our visualization unique, our analysis is too. There are a few shortage analysis statistics that already exist, such as Health Resource & Services Administration's (HRSA's) threshold [13], but they only use total population to approximate demand for public insurance and do not take into account the number of children that are eligible for public insurance, nor do they account for the level of severity of the shortage. Furthermore, we applied clustering to each state's summary statistics in order to group counties with similar dentist distributions to highlight areas that could benefit from similar policies to address their dental shortage issues. To our knowledge, applying clustering to dentist distributions has not been addressed in literature. Further explanation and discussion of our analysis is below.

Proposed Method(s)

1. Identifying Shortage Areas

We want to measure the dental care availability on a county basis and our approach to shortage area analysis requires a consistent measure of supply and demand.

Determining supply is slightly more complicated than just counting the number of dentist locations in a county—dentists can work in multiple locations across county lines. In order to avoid counting the dentist too many times, we created a model based off dentist hours distributed across all the locations dentists work at. Dentist hours are determined based on the 2010 national average of dentists' work hours: 35.2 hours a week for 42 weeks a year. If a dentist works at four locations across four separate counties, the "supply" of that one dentist would be .25 in each county which equates to .25*35.2*42 = 369.6 hours a year that dentist spends working in each county.

Then we determine demand for dentists in each county by gathering census information about how many children are covered under public insurance and how many children are not covered by any insurance but are eligible for public insurance. Using this total of population children who do not have private dental insurance, we can calculate the approximate demand of dental care by estimating how many dentists are required to serve that population. The suggested number of visits for children is twice a year; assuming each visit is an average of an hour long, each child requires two hours of dental care per year. Therefore, the number of dentists demanded can be determined using the suggested number of visits combined with the aforementioned approach of calculating the supply of dentists. The general equation for the number of dentists demanded is below.

$$\text{Number of Dentists Demanded} = \frac{2 * Number\ of\ Children\ on\ Public\ Insurance + Number\ of\ Uninsured\ Children}{35.2\frac{hours}{week} * 42\frac{weeks}{year}}$$

The dentist supply consists of multiple different dentist types (general, pediatric, and specialist). Using data from the medical expenditure panel survey, an estimation of youth preventative dental care for each dentist type was gathered: general and pediatric dentists spend 22% and 84% of their time on children's preventative dental healthcare, respectively. We then created a Dentist Availability Index based on the ratio of supply of dentists to demand of dentists.

$$Dentist\ Availability\ Index\ (DAI) = \frac{Number\ of\ General\ Dentists * 0.22 + Number\ of\ Pediatric\ Dentists * 0.84}{Number\ of\ Dentists\ Demanded}$$

We then designate a county as a shortage area based on its DAI value. We define a shortage area as an area where demand is twice the supply, thus a shortage area is declared when the county's DAI is less than or equal to 0.5. However, we are aware that there are levels of severity and therefore levels of risk for children. Our levels of shortage are low (DAI between 0.375 and 0.5), medium (DAI between 0.25 and 0.374), and very high (DAI between 0 and 0.24). These levels of severity allow policy makers and healthcare professionals a way to identify the most at-risk areas for children and work on policy solutions to alleviate the problem.

2. Clustering Similarly Distributed Counties

Many counties in a state are likely to have a similar distribution of dentists, particularly dentists that accept Medicaid and/or CHIP, as well as pediatric dentists that mostly see children. Therefore, shortage areas are likely to be easily grouped based on their dentist distributions. It is useful to policy makers to see groups of areas where children are highly likely to lack access to quality dental care because they will be able to target similarly affected areas with policies that are tailored to the group's circumstances while still maintaining some generality. By clustering the counties in a particular state, we can discover similar characteristics between counties that could affect children's access to dentists and thus target clustered areas with similar policies.

The clustering algorithm we implemented was Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which was run on each state's data. DBSCAN detects arbitrarily shaped clusters and did not require any initial cluster centers, and thus was the most suitable for our data. DBSCAN has two main parameters: the size of the epsilon neighborhood, and the minimum number of points required to form a cluster. Experimentation with these parameters is described below in the Experiments/Evaluation section. We implemented the DBSCAN algorithm separately for each state's county summary statistics.

After clustering the counties that had similar dentist distributions, we calculated the average Dental Availability Index (DAI) for each cluster so that we could then designate clusters as shortage areas with the appropriate severity. By clustering the counties in a particular state, we can discover similar characteristics between counties that could affect children's access to dentists, and eventually predict areas that might soon need assistance. Clustering dental data is an innovative way to identify areas that have similar policy needs that has not yet been addressed in literature.

3. Interactive Visualization(s)

For the interactive visualization component of this project, we implemented HTML and d3.js to display our analytical findings and implement interactive data visualization techniques including but not limited to brushing, filtering, drilling, identifying, and scaling. Our visualizations have been completely customized as per the requirements given to us by public policy professionals and currently there is no available tool that addresses such concerns. Our visualizations are not just unique and innovative but are impactful in that they are addressing real needs of policy professionals.

Visualization – Country Level View

The country-level visualization features a map of the United States with two main views. The primary component of the visualization is the US choropleth heatmap. There is additional filtering functionality to filter by county/cluster or insurance type. The *CHIP* and *Medicaid* buttons allow users to filter the data based on how many dentists accept CHIP, Medicaid, or both. There is also a radio button option to switch between the default heatmap view displaying the distribution of total dentists working in each county, or a shortage area view highlighting the various levels of shortage in each county across the country.

The default heatmap corresponds intuitively to the total number of dentists in a given county and state. The darker the color, the greater the number of working dentists in that county. There is also a tooltip which, upon hovering over a county, enables users to discover more information about the identified county displaying summary statistics on that specific county. This type of view should be very familiar to any user that has interacted with a heatmap and should be relatively intuitive.

The country level view can also display the shortage area analysis, if the user selects the Shortage Area option. The color of a particular county represents the severity of the dentist shortage in the area. The color green represents a county that has not been identified as a shortage area and varying levels of red, orange and yellow representing the degree to which there is a shortage in a particular county. Like the standard heatmap view, hovering over a particular county displays summary statistics for that specific county, particularly the DAI value.

When a user is viewing the map at the country level, clicking on a particular county or state will take them to the state-level view enabling users to drill into specific states to explore our analytical findings and statistics in more depth.

Visualization – State Level View

The state level view is the most important part of the visualization as policy decision regarding dental care availability will probably be done at local level. Thus, this view has more functionality and added features compared to the country level in order to enable users to access ample information to support policy decisions.

The default interface of the state level view is similar to the default heatmap of the country level view, as it just depicts the specific state. Offering even more functionality, it currently consists of several interactive visualizations including a choropleth heatmap detailed at a county level, a pie chart (specifically requested by our research advisor - sorry Polo ☹) that indicates the number of dentists by dentist type, and a bar plot that displays the aggregated number of dentists by rurality index. It also allows the user to filter by insurance type and choose a desired scale (log or linear) and allows the user to visualize results by the absolute number of dentists as well as the number of dentists per 100,000.

Not only does a helpful tooltip appear when a user hovers over a particular county, the statistics for that specific county/cluster are displayed alongside the state-level statistics (shown whether the cursor is or is not hovering over a specific county/cluster). This allows the user to compare that particular county's dentist situation to the overall state's; this can be useful to discover characteristics about counties that are struggling or thriving, compared to the state.

Shortage areas can be viewed on both the default county view or in the cluster view. Like the country level view, the color green represents a county that has been identified as not a shortage area and varying levels of red, orange and yellow representing the degree to which there is a shortage in a particular county/cluster. In the default view, the user can select the Shortage Area button and view shortage areas as defined by the county's DAI.

Once the user has selected the cluster view, the choropleth map adjusts to show our clustering results; the color of a particular county corresponds to the cluster it belongs to. Hovering over a particular cluster both pulls up a tooltip with more specific characteristics and adjusts the other appropriate statistics in the other interactive visualization plots (pie, bar, table). Like the other views, the Shortage Area option is available to then discover the shortage areas that are correlated with the cluster view. The shortage area colors are the same as described in the other views, but instead of corresponding to a specific county, the shortage area severity is designated by the average DAI for each county in the specific cluster. This allows the user to view the average shortage severity for similarly distributed counties that have similar dentist distributions.

Experiments and Evaluation

The impact on healthcare policymakers is our ideal measure of success and due to the course timeline limitations and nature of the project, our success will be determined by user feedback given by academic advisors and research professionals in the Georgia Tech Health Analytics department. Visualization requirements implemented above have been carefully chosen and provided by health research professionals. The regular feedback by our research advisor and by public policy professionals has been positive for both the layout and content of our visualizations.

We performed experiments for both the analysis and user experience components of our project. Analysis experiments mainly consisted of parameter tuning for our DBSCAN algorithm. The two main parameters to modify in DBSCAN are the epsilon neighborhood distance, and the minimum number of samples required to designate a group of points as a cluster. We chose the value of two for the minimum number of samples, so that each cluster would have at least two counties. The main testbed of experimentation was modifying the epsilon neighborhood value. We tested values between 0.1 and 1, incremented by 0.1 each iteration. We utilized the average silhouette score as a performance indicator of the particular parameter value. The silhouette score was not the only metric we used; we also looked at the distribution of counties per cluster to make sure that we did not overfit. Silhouette scores continuously increased with each iteration of increasing epsilon, while the cluster distribution became more and more skewed towards one dominating cluster. We determined an epsilon equal to 0.5 to be the optimal value for our purposes.

For a more formal experiment to evaluate the user experience, we released a small user study to the research group we have partnered with, to the members who have not been involved with the creation of the visualization but are still familiar with the subject area. The study ultimately consisted of four users. In an ideal world, we would have many more survey participants, but because our intended audience is policy makers and health research professionals, the pool of users to choose from at our immediate disposal was rather small. Future work includes expanding the survey.

The survey itself was distributed as a Microsoft form, and participants' answers were anonymous. The main features we wanted to assess were:

Is the application easy to learn?

Is the application easy to use?

Is it helpful for discovering shortage areas?

Is information displayed effectively, particularly for each county/state?

Was the cluster view useful?

Participants were asked to rank these attributes of the web portal's user interface, from strongly disagreeing with the attribute to strongly agreeing with the attribute. The feedback from the survey was overwhelmingly positive. Almost of the attributes were ranked with "agree" or "strongly agree", with one "neutral" response. Users also had the opportunity to give free-response feedback to elaborate on what they thought we could do to improve visualization. Comments were mostly positive, with minor suggestions like spacing between text, and some overlapping text within certain features. Obviously, with such a small sample size, results are very likely to be skewed. However, we do believe that the survey results reflect the application with relative accuracy because the comments were very similar.

Conclusions/Discussion

Upon reflection, we are very satisfied with the project outcomes. As stated at the beginning of this report as well as in previous project reports, the goal of this project was to build a web application portal for healthcare professionals and policy makers to identify shortage areas for children who are eligible for public insurance (CHIP and/or Medicaid).

For more than 800 counties spanning 16 states, we cleaned, aggregated, and complied summary statistics whilst also building an interface for policy makers to visualize shortage areas as well as view such statistics. While much of our project went according to expectations, some parts of the process certainly took longer than anticipated and we ran into unforeseeable obstacles.

Data preparation and cleaning stage required many more hours of work than expected given the quality of data we were working with, the availability of the BOD datasets by state and the quality of preparation that preceded our involvement in this project. Although our team would have liked to have compiled statistics for more than 16 states, we are confident that our success with these 16 states will translate into success with more states.

For interactive visualization, we have built a portal that is very much in line what we and our research advisor envisioned at the beginning of the project and that implements lots of interactive visualization techniques. As expected with large projects, we did run into some rather difficult obstacles which required much time and effort to work through. These were most often related to our relative inexperience with d3.js, but both the quality of our work and our familiarity with the library increased exponentially. It was through consistent effort from Week 1 that we were able to build a portal with this level of functionality and are satisfied with the end-result given the allotted time.

We also created an innovative method of quantifying dental care availability shortages using the Dental Availability Index (DAI), as well as implementing clustering in the form of DBSCAN to provide policy makers with groups of counties with similar dentist distributions that would benefit from tailored policy changes. These results should enable further research to look further into both causes of shortage areas and specific policy recommendations.

With the project now complete for our purposes, we will be working with our research advisor on handing this project over to future students who will be focused on compiling statistics for the remaining 34 states. The visualization is ready to be put into production although further changes may be requested in the next few months.

All team members have contributed an equal amount of effort towards this project.

References

[1] Paul W. Newacheck, Dana C. Hughes, Yun-Yi Hung, Sabrina Wong and Jeffrey J. Stoddard

Pediatrics April 2000, 105 (Supplement 3) 989-997;

[2] Ola O, Sedig K. Beyond simple charts: Design of visualizations for big health data. 2016;8(3).

[3] Cao S, Gentili M, Griffin PM, Griffin SO, Serban N. Disparities in Preventive Dental Care Among Children in Georgia. Prev Chronic Dis 2017;14:170176. DOI: http://dx.doi.org/10.5888/pcd14.170176

[4] National Institute of Dental and Craniofacial Research. Oral Health in America: A Report of the Surgeon General. Rockville, MD: US Department of Health and Human Services; 2000. https://www.nidcr.nih.gov/DataStatistics/SurgeonGeneral/Documents/hck1ocv.@www.surgeon.fullrpt.pdf

[5] National Center for Health Statistics. Health, United States, 2015: With Special Feature on Adults Aged 55-64. Hyattsville, MD: National Center for Health Statistics; 2015. https://www.cdc.gov/nchs/data/hus/hus15.pdf.

[6] Griffin, SO, Barker, LK, Wei, L. Use of dental care and effective preventive services in preventing tooth decay among US children and adolescents—Medical Expenditure Panel Survey, United States, 2003-2009 and National Health and Nutrition Examination Survey, United States, 2005-2010. MMWR Morb Mortal Wkly Rep. 2014;63(2):54–60.

[7] Jones, E. (2013). Improving Dental Care Access for Low-Income Populations. Wright State University, Dayton, Ohio.

[8] Gale, Nathan, and William C. Halperin. “A Case for Better Graphics: The Unclassed Choropleth Map.” The American Statistician, vol. 36, no. 4, 1982, pp. 330–336. JSTOR, www.jstor.org/stable/2683080. Accessed 5 Oct. 2020.

[9] Ronald O. Valdiserri and Patrick S. Sullivan (2018). Data Visualization Promotes Sound Public Health Practice: The AIDSvu Example. AIDS Education and Prevention: Vol. 30, No. 1, pp. 26-34 https://doi.org/10.1521/aeap.2018.30.1.26

[10] B. Shneiderman, C. Plaisant and B. W. Hesse, "Improving Healthcare with Interactive Visualization," in Computer, vol. 46, no. 5, pp. 58-66, May 2013, doi: 10.1109/MC.2013.38.

[11] O’Brien, Oliver. “Geovisualisation of Consumer Data.” Consumer Data Research, by James Cheshire et al., UCL Press, London, 2018, pp. 140–151. JSTOR, www.jstor.org/stable/j.ctvqhsn6.13. Accessed 5 Oct. 2020.

[12] Courtney, Karen L. "Visualizing nursing workforce distribution: policy evaluation using geographic information systems." International journal of medical informatics 74.11-12 (2005): 980-988.

[13] Serban, Nicoleta, and Scott L. Tomar. “ADA Health Policy Institutes Methodology Overestimates Spatial Access to Dental Care for Publicly Insured Children.” Journal of Public Health Dentistry, vol. 78, no. 4, 2018, pp. 291–295., doi:10.1111/jphd.12285.

[14] H. Kupwade Patil and R. Seshadri, "Big Data Security and Privacy Issues in Healthcare," 2014 IEEE International Congress on Big Data, Anchorage, AK, 2014, pp. 762-765, doi: 10.1109/BigData.Congress.2014.112.

[15] Sack, Carl & Donohue, Richard & Roth, Robert. (2015). Interactive and Multivariate Choropleth Maps with D3. Cartographic Perspectives. 57-76. 10.14714/CP78.1278.