CS 171 Process Book

Team: US Healthcare
Members: Daniel Dong, Jake Dorab, Wentao Xu

- Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

Overview

There are many issues with the US healthcare system. Healthcare costs have been rising faster than inflation over the past fifty years, and spending now account for a fifth of US GDP. Despite this, the World Health Organization ranked the US 37th in quality of medical services. To help reign in on wasteful spending and to increase overall transparency, the US government has established an open data initiative to make to make hospital spending and performance easily accessible to the public. As a result, we decided to look into the publicly released data on www.data.gov to explore the relationship between hospital spending and the effectiveness of healthcare. By creating a data visualization tool based on these data, we hope to help anyone looking for cost-effective healthcare to contrast and compare hospitals on the local, state, and national level. We hope that in addition to helping patients make a better choice when selecting hospitals, we want to show trends highlighting the factors associated with the poor performance of some hospitals, and so reveal patterns in the data that could be important for policymakers.

Project Objectives. Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

Hospital spending and healthcare effectiveness vary greatly across different geographic regions. This geographic variation is affected by demographics, socioeconomic status and government policy, to name a few. Through our project, we hope to first shed light

on the differences in the effectiveness and cost of healthcare, which will be immediately helpful in allowing potential consumers compare and “shop” various hospitals in their region. Specifically, we hope to answer the following questions through our visualization tool:

- Which hospitals are high-performing? Which ones are low-performing?
- Which hospital within the selected county is the best hospital to go to?
- How does the selected hospital compare to all the other hospitals in the same county, state, or country?
- How does the spending and effectiveness differ among different types of hospitals?
- What are the similarities between high-performing hospitals? And what are the similarities between low-performing hospitals? What can the low-performing hospitals learn from high-performing ones?

Data Source

Demographic

A. Demographic Data 1
    a. Description: The database contains 2010 county level data on population distribution by age, race, and poverty level.
    b. Source:http://catalog.data.gov/dataset/community-health-status-indicators-chsi-to-combat-obesity-heart-disease-and-cancer
B. Demographic Data 2
    a. Description: 2005 - 2010 Data
    b. Source: http://catalog.data.gov/dataset/american-community-survey
C. Income Distribution
    a. Description: The database contains the distribution of incomes at the county level. It also provides data on unemployment and number of people on welfare.
    b. Source: http://www.irs.gov/uac/SOI-Tax-Stats-County-Data-2012; http://www.irs.gov/uac/SOI-Tax-Stats-County-Data
D. Healthcare Coverage
    a. Description: 2010 Healthcare coverage at county level
    b. http://catalog.data.gov/dataset/2010-small-area-health-insurance-estimates-sahie-data

Finance

E. Hospital Spending Data
    a. Description: The database contains the Medicare Hospital Spending by each care episode.
    b. Source:https://data.medicare.gov/Hospital-Compare/Medicare-Hospital-Spending-by-Claim/nrth-mfg3
F. Hospital Type
    a. Description: The database contains hospital type information (Government, Non-profit, Proprietary) for about 4800 Medicare certified hospitals.
    b. Source: https://data.medicare.gov/data/hospital-compare
G. Provider Utilization
    a. Description: 2012 Charge Data
    b. Source: http://catalog.data.gov/dataset/medicare-provider-charge-data

Big Picture

H. US Healthcare Spending
    a. Description: The database contains the breakdown of US healthcare spending from 1960 to 2013.
    b. Source:http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html'
I. CMS Summary Statistics
    a. Description: For general understanding of how Medicare and Medicaid works
    b. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/Downloads/CMS_Stats_2014_final.pdf

Data Processing

A. Demographic Data
    a. The database contains information about age and race distribution in columns in which each row represents a county. Thus, we would need to iterate through each row and store the distribution data in a dictionary where the key is the county name and value is an array of dictionaries.
B. Income Distribution

    a. This will involve extensive data processing as the income data is separated into different files for every state. Furthermore, within each file, the income data is not structured in a way that makes it easily accessible (no consistent index values, except for a country FIPS code.

C. Hospital Spending
    a. The database contains the total hospital spending for each episode of care. Therefore, we would need to iterate through each row and store the information about the hospital spending in a dictionary where the key is the hospital name and value is the spending amount.

D. Hospital Type
    a. The database has about 4800 rows with each row representing different hospital. Thus, we would need to iterate through each row and store the information about the hospital type in a dictionary where the key is the hospital name and value is the type.

E. US Healthcare Spending
    a. The database contains annual spending for each year from 1960 to 2013 in columns. Thus, we would need to iterate through each column and store the spending information in a dictionary where they key is the year and value is the spending.

-
- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

Implementation

We successfully acquired and cleaned the data to be formatted into JSON format used in the visualization. We also mapped the state and county population onto the US Map. Then, we've also created a basic bar chart to show the demographics information by state level. From here, we intend to layer the initial heat map of the US with the different

statistics on demographics and income levels. Users can navigate between the different statistics through a switch panel. We also plan to supplement the heat map for some of the variables such as income, with a side panel showing the wage inequality at a specific district level (through a simple histogram).

Evaluation

Peer Evaluation by Marcus Schrow, Dennis Cui, Hiro Yamada

- Who is the target audience?
    - If it's to help people find hospitals around them, would it help to have a filter to enter a zip code/address and return all within a certain radius?
        - Counties seems a little arbitrary - what if they live on a edge of a county / state.
    - If it's policy makers, some sort of comparison tool would help
        - Aggregate by county / state / etc.
        - As you explain that use case, probably most important thing to show is average spend
        - Probably show average spend, could have rank-ordered list of hospitals inside that county.
- Scale
    - May get cluttered if it's dense - Will Boston be covered by dots?
    - What if counties or even states have no or little data?
- Visual encoding
    - Can you get at spending of the hospitals on the main map page? I feel that's the most important aspect
        - Size may be a bit messy though…
        - Color already used for type of hospital :(
    - I guess it's okay if separately there's a bar graph or something that will highlight as you hover over a point on the map.
- Interaction seems good and pretty intuitive. May need to spec out how you'll do comparisons a bit more to make sure this doesn't get cluttered.
- Don't really see how animations can help unless you have data over time.