Group Members
Juliana Paun
Sahiba Dogra
Samantha Tuite
Prithvi Koka
Scott Chase
Caleb Macias

Proposal

Our selected topic was American healthcare and the categories include rates of mortality and discharge as well as columns for gender and cost of operation. In addition to the data having limited null values, we all have interest in the healthcare system in America. It is hard to be an American and not face issues with said system and the charges that come with it, even with insurance. A source wrote “If surgery is involved, hospital costs soar through the roof. Some of the most common surgeries have price tags that top $100,000”(Milliken). Even if this issue has not affected you personally, it most definitely has affected someone you love. People should not have to go into debt simply to save their own lives. With our information being federally sourced, this allows us to specifically look at American data from a credible source. Our objective is to take a closer look at the specific procedures that impact American healthcare day to day and understand whether or not the costs are relevant or necessary. Our goal is to analyze the mortality rates of certain procedures and diseases, in conjunction with billing, to understand the discrepancies in the risks and expense of those procedures and diseases.

Data Sources:

The Healthcare Cost and Utilization Project (HCUP) is a nationwide effort to aggregate and store healthcare data relating to hospital stays, surgeries, procedures, diagnoses, and emergency care. The HCUP consists of several databases sponsored by the Agency for Healthcare Research and Quality (AHRQ). The databases are populated by information derived from administrative and anonymous patient data, and were accessible through Wharton Research Data Services. We decided to use two datasets from the HCUP database. The first is the HCUP Diagnoses dataset and the second is the HCUP Procedures dataset.

The HCUP Diagnoses dataset provides information on seventeen categories of diagnoses from infectious and parasitic diseases to injury and poisoning. For each category, there are subcategories of specific diagnoses. For each diagnosis, there are statistics such as the number of discharges, percent of discharges, length of stay, percent died, percent male, mean charges, and mean costs.

The HCUP Procedures dataset provides information on sixteen categories of operations from operations of the nervous system to operations on the integumentary system. For each category, there are subcategories of specific procedures. For each procedure, the same statistics as defined in the HCUP Diagnoses dataset are available.

Projected Timeline:

Already Done	Chose our dataset and described the question that we want to explore
Week 1	● Cleaning and joining the tables ● Brainstorming possible visualizations ● Rank the impact of visualizations ● Caption them
Week 2	● In depth analysis on the visualizations that we ended up choosing ● Creating more visualizations based on the previous ● Peer review and revise ● Start outlining our report and presentation
Week 3	● Begin work on our presentation and paper ● Section off parts of the presentation and paper to people and have peer review on them so we're coherent
Week 4	● Continue work on the paper and presentation and finish both ● Peer review within our own team
Week 5	● Peer review with another team in the class and the professor. ● Make finishing touches ● Presentation

Proposed evaluation metrics:

Our visualizations will be evaluated by ease of gleaning the main idea from each visualization and the insight that the visualization offers into understanding our key question of cost in healthcare diagnoses and procedures. We will check and ensure that our visualizations present the right amount of information without including too much or too little information. Further, the standard deviation for some of the metrics can be assessed including mean cost per procedure and mean length of stay. The standard deviation can help assess the significance of the differences observed in the data.

By using evaluation metrics such as accuracy (standard deviation), clarity (ease of understanding visualization), and readability (correct use of color and visual appeal), we will ensure that our visualizations are accessible and appropriate to answer our desired research questions.

Anticipated challenges:

We anticipate challenges in joining the data and cleaning the data. The data is in two separate sheets that may be difficult to join given that both tables are different sizes. The analysis/clean up of both sheets will likely need to happen separately. However, the data is connected and by understanding them both separately first, we may be able to come up with new insights when integrating them together.

The data also contains nulls in several columns. We will likely need to exclude columns and rows and determine the best way to do so.

Summary

Certain procedures and diagnoses can have higher costs associated. Even with insurance, many people struggle to stay on top of healthcare expenses. By better understanding what procedures and diagnoses are more likely to incur cost, we can better allocate funds towards lowering the costs of those conditions, as well as strengthening insurance to effectively aid people with these conditions. There's also procedures and diagnoses that males or females are more likely to receive. Through our analysis, we can pinpoint whether there's a trend between the sex of the majority of the patients with a procedure/diagnosis, and whether that translates to an increased cost. Identifying these problems helps to make healthcare more equitable, ensuring that one sex isn't incurring higher healthcare costs.

References

Data Sets:

"Wharton WRDS ." Wharton Research Data Services, Wharton University of Pennsylvania , 6 Apr. 2018, https://wrds-www.wharton.upenn.edu/pages/get-data/public-data/healthcare/hcup-procedures/.

"Wharton WRDS." Wharton Research Data Services, Wharton University of Pennsylvania , 6 Apr. 2018, https://wrds-www.wharton.upenn.edu/pages/get-data/public-data/healthcare/hcup-diagnoses/.

Other Articles:

Milliken, Maureen. "Hospital and Surgery Costs – Paying for Medical Treatment." Debt.Org, 30 Nov. 2023, https://www.debt.org/medical/hospital-surgery-costs/#:~:text=The%20average%20hospital%20stay%20is,limited%20budgets%20or%20no%20insurance.

"Healthcare Cost and Utilization Project (HCUP) | Agency for Healthcare Research and Quality." Home | Agency for Healthcare Research and Quality, HCUP, 2022, https://www.ahrq.gov/data/hcup/index.html.