# Exploring Medicare Hospital Spending and Quality in the Upper Midwest

An Interactive Visualization Dashboard

CIS 671 Final Project

Yu-Ting Liu

## Introduction and User Goals:

Health care payers and analysts are increasingly interested in understanding not only how much is spent on inpatient care, but also whether that spending is associated with high-quality outcomes. In this project, I designed and implemented an interactive data visualization dashboard to explore Medicare inpatient hospital utilization and spending in a five-state region of the upper Midwest: Michigan, Ohio, Indiana, Illinois, and Wisconsin.

The primary user persona for this dashboard is a health-care analyst working for a regional health insurer or consulting firm. This analyst needs first to identify which counties represent the largest inpatient markets and whether they are relatively high-spend or low-spend after standardization. Second, compare hospital quality by CMS star ratings within those markets. Third, understand the clinical drivers of volume at a specific hospital. Fourth, benchmark states against each other on standardized spending. This dashboard supports these tasks through four coordinated visualizations that share common data sources and interactions. The goal is to help the analyst reason about value where high spending does or does not coincide with strong measured quality.

## Data Sources and Preprocessing

All data come from public Centers for Medicare & Medicaid Services (CMS) datasets for the United States:

- Medicare Geographic Variation – County level
  (standardized Medicare spending per beneficiary by county and year)
- Hospital General Information
  (hospital identifiers, location, ownership, certified beds, and overall star rating)
- Medicare Inpatient Hospitals by Provider and Service
  (MS-DRG–level discharges and payments by hospital)
- U.S. Census County shapefile
  (geographic boundaries for counties in all 50 states)

All preprocessing was implemented in Python using pandas and geopandas. Some keys highlighted here:

1. County–hospital crosswalk

   County names in the hospital file were cleaned, such as removing terms like “County” and “Parish,” and then joined to the Census county shapefile using state abbreviations and normalized county names, so that each hospital could be assigned a five-digit FIPS county code.

2. County-level quality and spending measures

   Hospital star ratings were combined into a bed-weighted county quality score. Geographic variation data were filtered to a single year, and I selected standardized per-beneficiary Medicare spending as the main cost measure. For each county, I then computed z scores, including z_spend, which represents standardized spending per beneficiary.

3. Hospital capacity and utilization

   From the MS DRG file, I aggregated total discharges per hospital, computed each DRG's discharge share, and calculated a hospital-level Herfindahl–Hirschman Index for DRG concentration as well as a weighted average Medicare payment per discharge. For each hospital, I also derived the percentage of medical versus surgical discharges based on the DRG text.

4. Regional subset

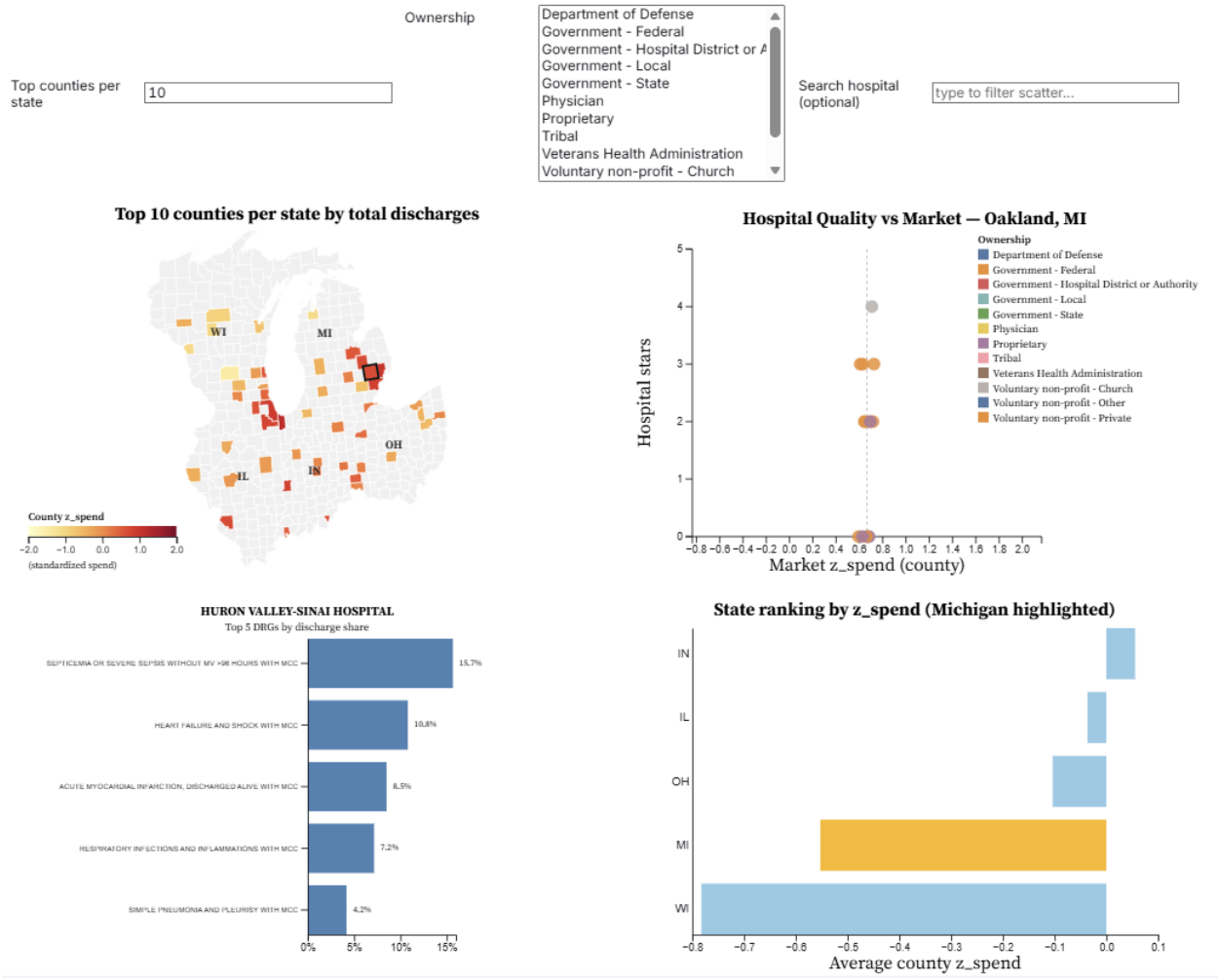
   Counties from the five target states were selected, and for each county, I aggregated beds_sum (total certified beds) and discharges_sum (the sum of hospital total discharges), which are used to rank the "top" counties in the dashboard.

The preprocessing script outputs four files used in Observable:

- counties_2023.csv  that shows county-level spend, quality, and capacity
- hospitals_2025.csv that presents hospital-level attributes and county FIPS
- hospital_drg_top5.csv that demonstrates the top five DRGs by discharge share for each hospital
- counties_geo_region.json that is GeoJSON of county boundaries for the five-state region

## Visualization Design and Implementation

All visualizations were implemented in **Observable** using **D3.js** and Observable's Inputs components. The final dashboard consists of four linked views arranged in a 2×2 grid.




At the top of the dashboard, the user can adjust three main controls: the number of top counties per state, which is used to filter both the map and the scatterplot, a hospital ownership filter that allows selecting multiple ownership types, and a hospital search box that provides an optional text filter on hospital name. Selections in one view, like clicking a county or a hospital bubble, update the other views via a shared selection object.

*Map: Top Counties by Discharges and Standardized Spend*

The top-left panel is a choropleth map of the five-state region. Counties are drawn from counties_geo_region.json and projected with a geoAlbersUsa projection. For each state, I use the Top 10 counties by total discharges (discharges_sum), and dim all other counties in light grey to emphasize the largest markets.

Color encodes z_spend using a sequential YlOrRd color scale, with a fixed domain of –2.0 to +2.0 standard deviations to make colors comparable across views. Darker red counties represent higher standardized Medicare spending per beneficiary; pale yellow counties are lower-spend markets. State labels are drawn at the centroid of each state's counties.

Clicking a county updates the selection and drives the hospital scatterplot and the hospital DRG chart.

*Scatterplot: Hospital Quality vs. Market Spending*

The top-right panel shows hospital quality within the selected market. In this scatterplot, the x-axis represents the county's z_spend value, and all hospital points are jittered slightly around this value, so they still align vertically with the market's cost level. The y-axis shows each hospital's overall CMS star rating from 0 to 5. Circle size scales with the number of certified beds to emphasize larger facilities, and color encodes ownership type, with a legend on the right to help interpret the categories. Filters at the top allow the user to restrict the plot to particular ownership types, including nonprofit and proprietary, or search for one hospital by name. Clicking a hospital bubble sets the selected provider and updates the DRG bar chart.

*Bar Chart: Hospital Top 5 DRGs by Discharge Share*

The bottom-left panel drills down into the clinical case mix of the selected hospital. Using pre-computed aggregations from hospital_drg_top5.csv, the chart displays the top five MS-DRGs ranked by the proportion of total discharges. In this chart, the x-axis shows the percentage of discharges, the y-axis lists DRG descriptions such as "SEVERE SEPSIS WITHOUT MV >96 HOURS WITH MCC" or "HEART FAILURE AND SHOCK WITH MCC," and the percentage values are printed at the right end of each bar. This view helps the analyst understand which conditions drive most admissions and most inpatient spending at that hospital.

*State Ranking: Average County z_spend*

The bottom-right panel is a horizontal bar chart that ranks the five states by average county z_spend. Each bar shows the mean z_spend across counties in that state within the five-state region. The bar for Michigan is highlighted in gold, while the other states are light blue. The x-axis is labeled "Average county z_spend", allowing direct interpretation of whether each state is above or below the national standardized spending level.

## Tasks and Personas

The dashboard is designed around tasks for a regional health-insurance or consulting analyst:

1. "Where are the largest Medicare inpatient markets in our five-state region, and which of them are relatively high-spend after standardization?"
    - Visualization: Top-N county map with z_spend choropleth.
2. "In this county, how do hospitals compare on quality, size, and ownership given the market's spending level?"
    - Visualization: Hospital Quality vs. Market scatterplot.

3. “For an individual hospital of interest, which clinical conditions (DRGs) account for most Medicare discharges?”
    - Visualization: Top-5 DRG bar chart.
4. “Where does Michigan sit relative to its neighbors on standardized Medicare spending?”
    - Visualization: State ranking bar chart with Michigan highlighted.

Thus, these tasks reflect realistic questions an insurer or consultant would ask when thinking about network design, value-based contracts, and regional cost-control strategies.

## Insights from the Visualizations

Using the dashboard, several patterns emerge in the five-state Medicare inpatient data. On the map of top counties, darker red clusters around large metros such as Detroit in Michigan and Chicago in Illinois indicate high spend markets, while many top discharge counties in Wisconsin and Indiana are much lighter, which suggests they handle substantial volume with lower standardized spending per beneficiary. In the hospital Quality vs Market scatterplot, zooming into Oakland County, Michigan, shows that most hospitals are rated only two to three stars, even though the county’s z_spend is above zero. This points to a high spend but modest quality market that could be a priority for targeted quality programs or value-based contracts. A few hospitals appear with zero stars, which in practice usually reflects missing or insufficient data rather than truly zero quality, and reminds us to interpret the rating scale cautiously. At the hospital level, the Top 5 DRG bar chart shows that a small set of conditions, such as sepsis, heart failure, myocardial infarction, infections, and pneumonia, accounts for a large share of discharges. For consultants working with hospitals, these become priority service lines for pathway redesign and readmission reduction, while for insurers or consultants advising patients or employers, they help identify facilities with deep experience in particular clinical areas. Finally, the state ranking chart reveals that most states in the region have negative average z_spend, which means they are below the national standard on average, with Wisconsin lowest, Illinois and Ohio closer to zero, Michigan in the middle, and Indiana slightly above zero. Taken together, these views help payers and consultants see where spending is concentrated, where value may be misaligned, and which markets and hospitals are promising targets for further analysis or value-based care initiatives.

## Limitations and Future Work

Several limitations should be noted in this analysis. First, z_spend is measured at the county level rather than at the hospital level, so a high spend county does not automatically mean that any individual hospital is inefficient, and more detailed hospital-level cost analysis would be needed. Second, CMS star ratings are used as a proxy for quality, but they do not capture every dimension of care, such as access to specialty services, patient mix, or social risk factors. Third, the results focus only on traditional Medicare beneficiaries and may not generalize to commercial insurance or Medicaid populations. Finally, the dashboard uses a single year of data,

so it cannot show whether spending and quality are improving or worsening over time. Future extensions could include incorporating readmission rates, mortality indicators, and post-acute care utilization, and allowing users to switch between multiple years in order to explore trends rather than a single snapshot.

## Conclusion

This project demonstrates how public CMS datasets can be combined into an interactive visualization dashboard that supports realistic decision-making tasks for health-care insurers and consultants. By linking a regional county map, a hospital-level scatterplot, a DRG mix bar chart, and a state ranking chart, the dashboard allows users to move seamlessly from regional patterns to individual hospitals and clinical service lines.

The final visualization helps answer key questions about where Medicare spending is concentrated, how that spending relates to measured quality, and which hospitals and conditions might be priorities for improvement or partnership. For my own learning, the project also reinforced the importance of careful data preprocessing, thoughtful encoding choices, and designing interactions around specific personas and tasks rather than simply plotting every variable.

## Reference


- Centers for Medicare & Medicaid Services Data. (n.d.). Data.cms.gov. https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-geographic-comparisons/medicare-geographic-variation-by-national-state-county
- PQDC. (2025). Cms.gov. https://data.cms.gov/provider-data/dataset/xubh-q36u#overview
- Centers for Medicare & Medicaid Services Data. (n.d.). Data.cms.gov. https://data.cms.gov/provider-summary-by-type-of-service/medicare-inpatient-hospitals/medicare-inpatient-hospitals-by-provider-and-service
- Data.gov. (2023). Data.gov. https://catalog.data.gov/dataset/2023-cartographic-boundary-file-shp-county-and-equivalent-for-united-states-1-5000000
- PQDC. (n.d.). Data.cms.gov. https://data.cms.gov/provider-data/topics/hospitals/overall-hospital-quality-star-rating
- Overall star rating for hospitals. (n.d.). Www.medicare.gov. https://www.medicare.gov/care-compare/resources/hospital/overall-star-rating


## Observable dashboard link:

https://observablehq.com/d/8428ef9bfbf826d9