Analysis of Medicare Provider Data

University of California, Berkeley
MIDS, Fall 2017
W200 Python, Section 1, Project 2
Team: Cameron, Ehsan, Josiah, Sharad

Executive Summary

This project explores variations in hospital provider charges and procedures covered for Medicare patients across the country using descriptive statistics. It focuses on an FY 2011 dataset that contains hospital-charges for approximately 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments.

The table below summarizes the key questions and answers from this data set:

Question	Finding
What medical conditions do patients visit hospitals for the most / least?	Circulatory Conditions - 1,876,016 discharges Ear, Nose, Mouth and Throat - 23,427 discharges
What medical conditions are covered by hospitals most and least frequently?	Respiratory Conditions - 93.5% of providers Mental Diseases and Disorders - 18.4% of providers
What's the difference between what hospitals charge, what they receive, and what Medicare pays?	For Medicare-covered treatments, hospitals receive only a fraction of what they charge (23% on average). Total payment received by providers is the Medicare amount plus the out-of-pocket expenses ($1,284 on average).
Which providers charge the highest amount of covered charges?	Providers associated with university research hospitals and providers with a brand recognition tend to charge more for medical care.
How do hospital payments vary by state?	Average payments range widely by state, from under $8,000 to over $14,000. High cost areas include Alaska, Hawaii, the Northeast, and the West Coast. States in the Deep South are among the least costly.

To what extent do various population demographics affect hospital payments?	Demographics studied include population, urban vs. rural, age, income, education, poverty, unemployment, and health insurance coverage, all of which had weak to no relationship with payments. The team hypothesizes that differences from patient to patient are a much greater driver of payments than the average demographics of hospitals.
Which state has the highest density of providers?	Oklahoma - 1.38 per 10,000 Medicare beneficiaries
Which state's providers cover the most medical conditions on average?	Maryland - 87% of medical conditions covered per provider on average
Are there economies of scale in Healthcare	Even after controlling for average income within a zip code, there is still a positive correlation between Average Total Payments or Amount Billed and Total Discharges

Medicare Provider Search Tool

In addition to answering the research questions, the team built a tool to search for Medicare providers that treat a specific medical condition within a given radius of a user's zip code, returning results from least-to-most costly.

Audience

This report focuses on technical data exploration, and thus it assumes the reader is analytically advanced and familiar with more sophisticated data analysis and plotting techniques, along with basic statistical analysis tools such as box plots and correlation coefficients.

Introduction and Data Overview

Primary Data Overview

The data is organized by a combination of hospital providers and classifications of covered hospital procedures, also known as Medicare Severity Diagnosis Related Groupings (MS-DRGs). Attributes pertaining to 3,000 hospital providers and 100 DRGs are captured in this dataset containing over 163,000 records.

Fields in this dataset include:

- DRG Definition (medical condition)
- Provider Id
- Provider Name
- Provider Street Address
- Provider City
- Provider State
- Provider Zip Code
- Provider HRR (hospital referral region)
- Total Discharges (number of patients treated)
- Average Covered Charges (what the hospital billed)
- Average Total Payments (what the hospital received)
- Average Medicare Payments (what Medicare covered; the patient pays the rest)

Full definitions for each data field are in the appendix.

Addition data sets the team merged for the analysis are outlined in the next section.

Data Exploration Techniques and Challenges

Weighting

Because each row of data represents an average charge / payment for a single provider-medical condition combination, along with the number of total discharges (number of patients treated), calculations throughout the report use averages weighted by number of total discharges, unless otherwise noted.

DRG to MDC Grouping

The main, Inpatient Charges data classifies medical treatments into Diagnosis-Related Groups (DRGs), with the data set containing 100 different DRGs. The research team noticed many DRGs were similar, and searched to find a mapping table (from the Centers for Medicare & Medicaid Services) that grouped DRGs into Major Diagnostic Categories (MDCs). The team joined the data (a left join from the inpatient charges to the mapping table on the unique DRG code integer) to reduce medical conditions into a much more manageable 16 MDC categories. The bulk of this report focuses on MCDs.

Zip Code Data

In addition to our primary Inpatient Charges data set, the team also gathered demographic data from zip codes for analysis (chosen, as opposed to metropolitan statistical areas, for instance, because zip codes were the available field in the Inpatient Charges data). This demographic data, sourced from the American FactFinder resource within the US Census, proved more difficult than expected to find and retrieve. Once found, we loaded data from 7 separate csv tables, each representing a different demographic category (e.g., age, income, education) with dozens to hundreds of columns and over 33,000 rows (one row per zip code). We then combined all 7 tables into a single zip code demographic table using an outer join. Next, we joined this single zip code table into our main table of inpatient charges and costs, using a left join to preserve the inpatient data and ignore zip codes without medical providers.

Medicare Beneficiary Data

The team gathered 2011 Medicare beneficiary population sizes for each state and joined that with our primary dataset. This data was gathered from the Henry J. Kaiser Family Foundation, but the original source was Centers for Medicare & Medicaid Services (CMS) Program Statistics. Including this information allowed us to calculate the density of Medicare providers within each state in our primary dataset.

Internal Data Inconsistencies

The research team addressed several data inconsistencies, the most notable of which are listed in the following table:

Observation	Solution / Ramifications
In 583 (0.36%) of the 163,065 records, Avg. Total Payments exceeded Avg. Amount Billed, and in 356 (0.22%) records, Avg. Medicare Payments exceeded Avg. Amount Billed. All 356 cases were a subset of the 583 records.	We considered this to be negligible. A possible reason is if corrections / refunds / discounts occurred across years (e.g., a 2010 overpayment resulting in a 2011 correction).
172 provider zip codes (0.52%) were not located in the 33,120 zip codes in the US Census dataset.	We considered this to be negligible.
Demographics are for the zip code of the provider location, and not necessarily exactly representative of patient demographics.	Absent having patient data (which would grossly violate HIPPA laws), the research team felt provider location demographics would be reasonably representative, as zip codes in close proximity generally have similar demographics.
US Census data for zip code demographics comes from the Zip Code Tabulation Area (ZCTA), as opposed to the US Postal Service zip code. In nearly all cases, these are identical or nearly the same.	Given the large similarity between the two zip code definitions, the fact that this report uses broad, average demographic statistics, and the similarly in demographics for nearby zip codes, the research team considered the difference to be negligible. More information can be found at this website: https://www.census.gov/geo/reference/zctas.html

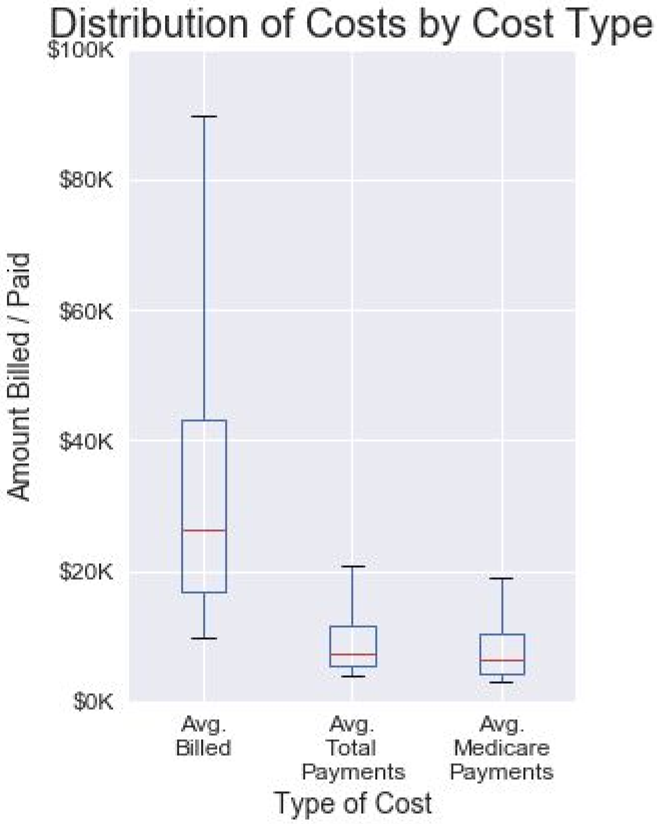
The Medicare inpatient data is from 2011, but not all demographic data was available in this same year. Specifically, percent of urban population came from the decennial 2010 census, and education, poverty, income, employment, and health insurance came from 2015. Age data was available in 2011.	Given the relatively slow changes of a population's demographic, a 1 or 4 year difference between inpatient data and demographic data was considered negligible. Supplemental data on Medicare Beneficiary population sizes were gathered from 2011 as well, to prevent any inconsistencies in analysis of density of providers.

Payment Types

At a high level, there are three types of costs in the data:

1. **Average Amount Billed**: What the provider (hospital) billed for the service.
2. **Average Total Payments**: The total payments the provider received for the service.
3. **Average Medicare Payments**: The payments Medicare paid to the provider. The patient pays the difference between Average Total Paid and Average Medicare Paid.

The graph below shows the range of these costs:



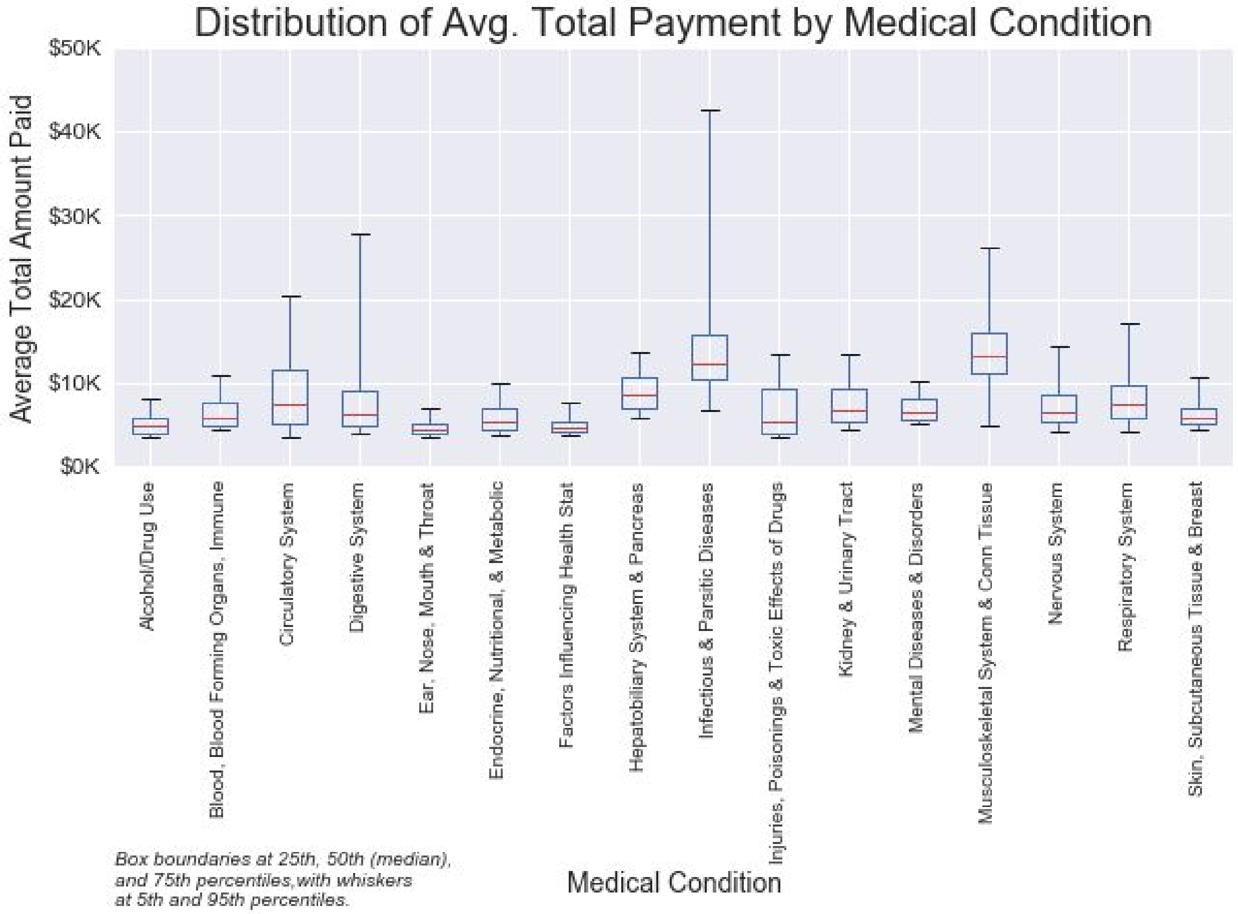
Box boundaries at 25th, 50th (median), and 75th percentiles, with whiskers at 5th and 95th percentiles.

As the graph shows, Average Amount Billed is significantly higher than Total or Medicare Payments. From our research and experience, Average Amount Billed is analogous to a "rack rate" of a hotel room or airplane ticket, and thus not very useful for further analysis, as it is not at all reflective of either cost of providing service or payment received. These rates exist to help capture higher rates from commercial (i.e., non-Medicare) patients, and to maximize revenue by realizing the full amount charged for the rare patient who lacks insurance coverage and can afford these high rates.

Average Medicare Payments are very similar to Average Total Payments, with Medicare paying nearly all of the cost of service, less the average out-of-pocket expenses of $1,284. That said, it's difficult to draw specific conclusions about the proportion of the Average Total Payments Medicare pays because of numerous factors that vary by patient, including patients with differing levels of Medicare coverage (especially Medicare Parts C and D: Medicare Advantage and Medicare prescription drug coverage), varying percentages of deductibles met, and patient premium payments that are not captured in this data set.

Payments by Medical Condition

The graph below shows how costs vary by medical condition:



Key observations include:

- Most conditions have a small cost distribution for the majority of patients
- Most conditions also have a “long tail” of high-cost outliers. This presumably represents complex cases and/or comorbidity factors (patients with multiple conditions).
- Medical conditions with the highest probability of costs exceeding $10,000 are infectious and parasitic, and musculoskeletal, circulatory, and digestive diseases and disorders.

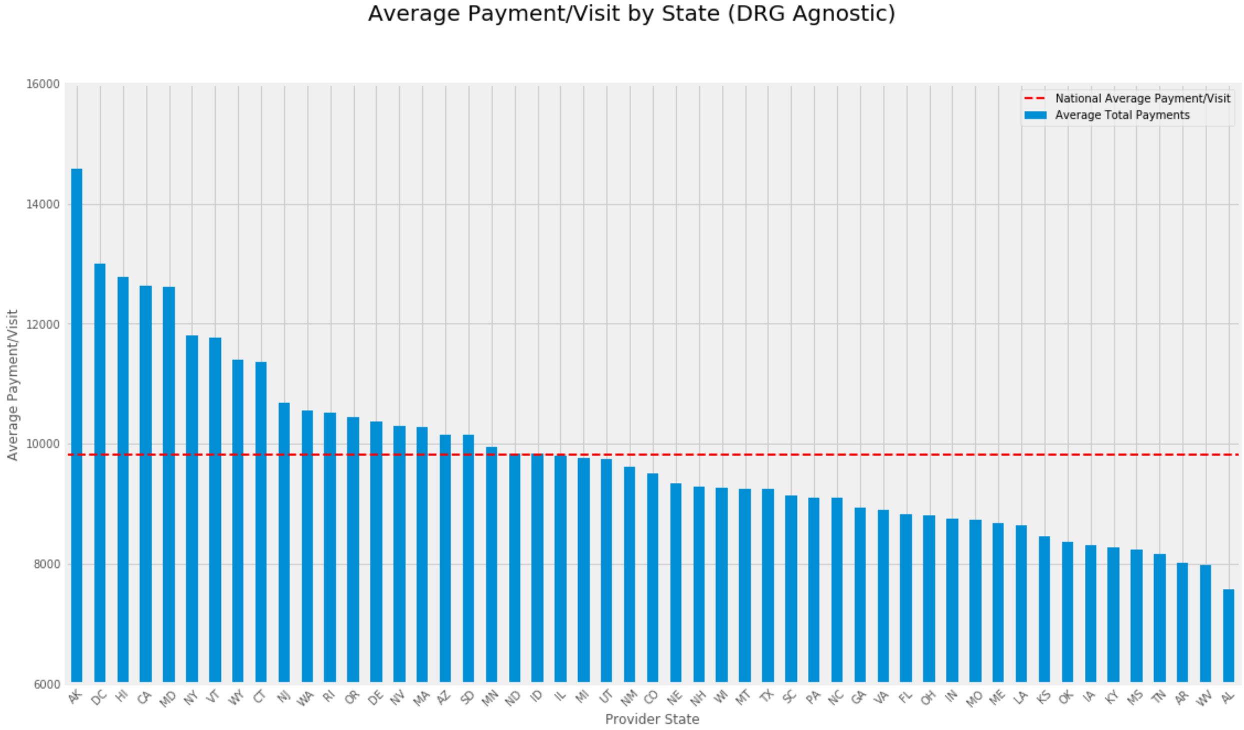
- There is a strong correlation between length of hospital stay and cost. Additionally, MDCs with high costs also tend to have longer hospital stays. For instance, infectious and parasitic diseases often require an extended hospital stay. These extended hospital stays tend to rack up the most in hospital charges as compared to services that are performed and then the patient is quickly discharged. Extended hospital stays can be a lucrative business for hospitals.

Costs by Medical Condition and Geography

This section explores the Average Total Payments providers receive for each treatment, and how they vary by both the medical condition being treated and the geography where they're treated. Unless otherwise stated, costs in this section do not include Average Amount Billed or Average Medicare Payments.

Costs by Geography

This section explores how costs vary by geography. The graph below shows Average Total Payments by state:



Key Observations:

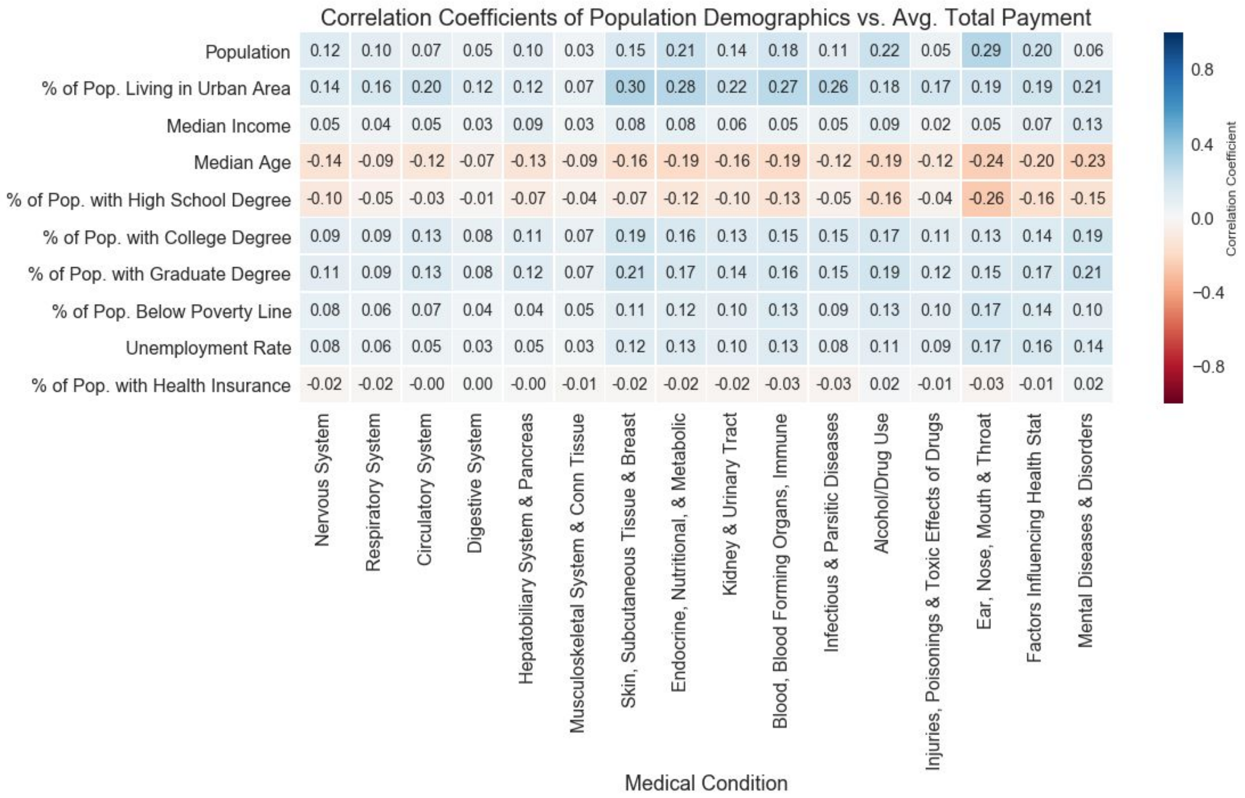
- Average payments range widely by state, from under $8,000 to over $14,000.
- High cost areas include Alaska, Hawaii, the Northeast, and the West Coast
- States in the Deep South are among the least costly

Analysis by Demographics of Provider Zip Code

In this section, we explore the effects of various demographics on Average Total Payments. Note, these demographics are for the zip code of the provider location, and not those of the patients. Demographics included in this study are:

- Population
- Percent of population living in urban areas
- Median income
- Median age
- Percents of population who completed high school, college, and graduate school
- Poverty rate
- Unemployment rate
- Percent of population with health insurance coverage

The heatmap below provides a summary of the relationships, using correlation coefficients between Average Total Payments and selected population demographics, broken down by medical conditions. Though busy, it provides a quick synopsis of the demographic analysis.

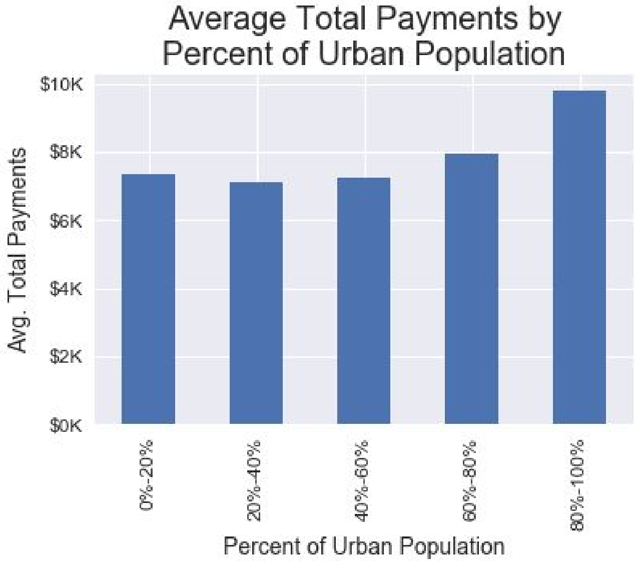


Key Observations:

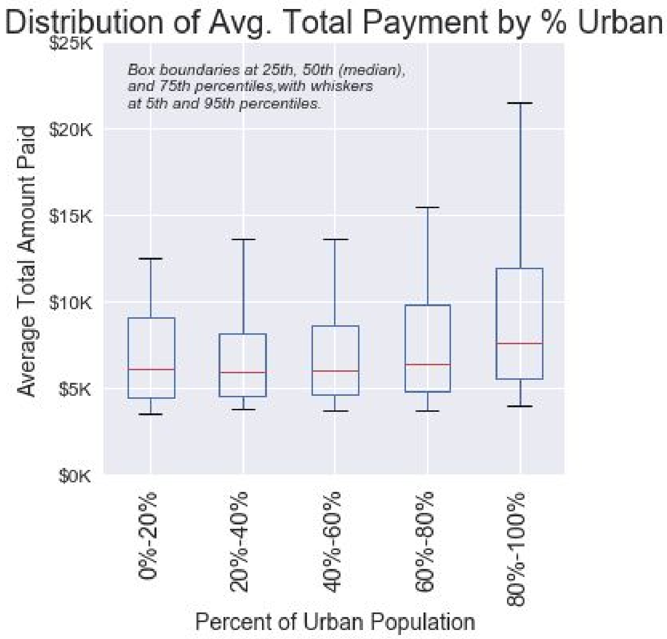
- None of these demographics have a high correlation with Average Total Payments (|0.30| is the largest for any medical condition). This finding suggests hospitals do not receive more revenue based on location, nor does Medicare discriminate based on location. In other words, costs do not fluctuate based on a patient's ability to pay.

- These findings also generally hold true across all medical conditions.
- The low correlation coefficients also suggest that other factors contribute more to the variance in costs. The research team hypothesizes variations in disease severity and complexity are more dominant factors in determining cost than disease categories or demographics, though the data set studied is not sufficient to prove or disprove this hypothesis.
- Since payments are relatively similar across geographies, a hospital seeking to maximize profits from Medicare patients could be best served by locating in a lower-cost area, though other factors such as commercial patient profits and patient volume would also be obvious considerations.
- Though correlation coefficients were generally low across categories, locations with more urban populations have somewhat higher costs. This phenomenon could be due to providers in highly urban areas experiencing higher operating costs (e.g., labor costs, taxes, rent), with Medicare providing higher reimbursements in these areas.

We further explore the effect of percentage of population living in urban areas in the following graph:



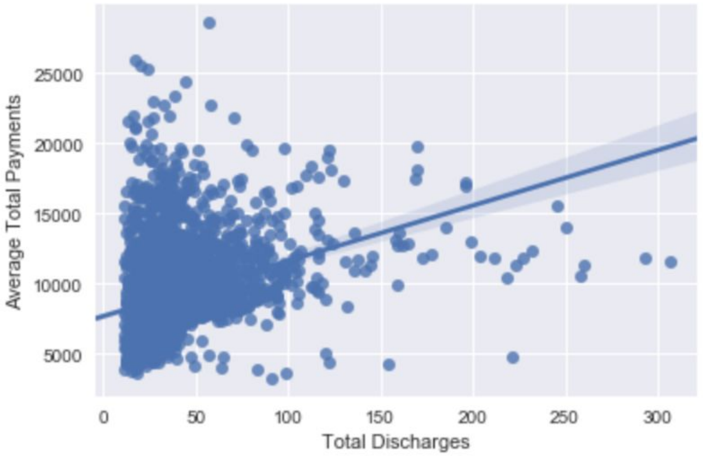
The figure above shows a clear difference between providers in the more rural areas (just over $7,000 on average), as compared to highly urban areas (almost $10,000). This finding then begs the question, if there's such a notable difference in payments between urban and rural populations, why is the correlation coefficient so low? The answer lies in the distribution of these costs, as shown in the figure below:



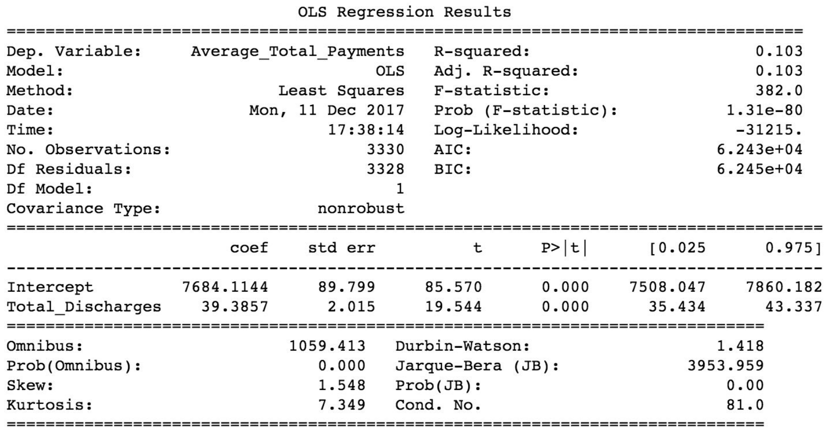
This figure clearly demonstrates the wide distribution of costs within each category of urban population percentage. The red, median lines have a similar pattern to the averages in the Average Total Payments by Median Income graph above, but the variation within each group is much larger than the difference in the average (or median) values among the groups. In other words, although average costs vary a little based on percentage of urban population, they vary much more on an overall basis, which supports our earlier hypothesis that variations in disease severity and complexity are more dominant factors in determining cost than disease categories or demographics.

Other demographics showed similar patterns, but in the interest of brevity, additional findings by zip code demographics are contained in the Appendix.

Economies of Scale in Healthcare

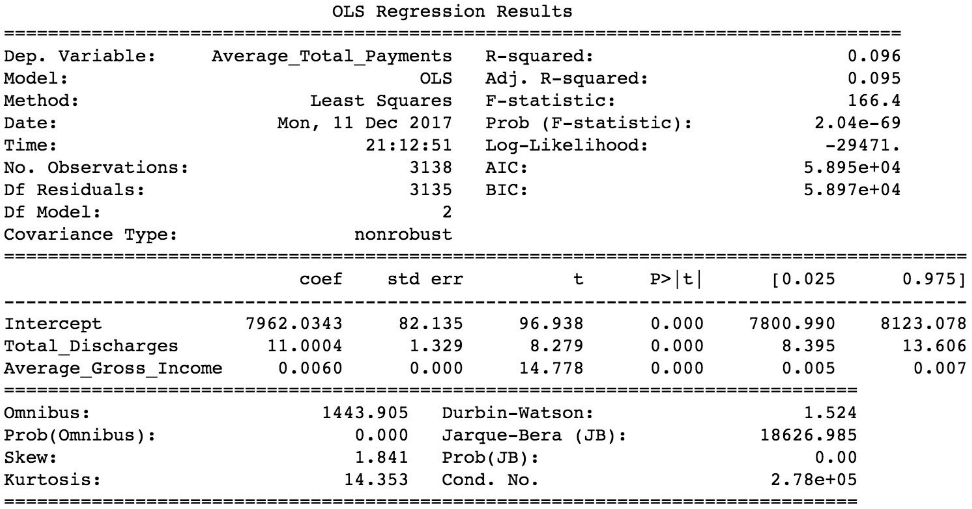


Another question we examined was around the cost efficiency of treatment. Given the importance of economies of scale for efficiency in most sectors, we wanted to see if number of discharges is correlated with either what the hospital billed, or what was paid. In our first regression we simply regressed Average Total Payments on Total Discharges at the provider level. Results are displayed in the figure below, but in essence we found that total discharges is correlated with average total payments. We also ran the same regression with Total Billed as the dependent variable but the results were similar with a slightly lower R Squared.



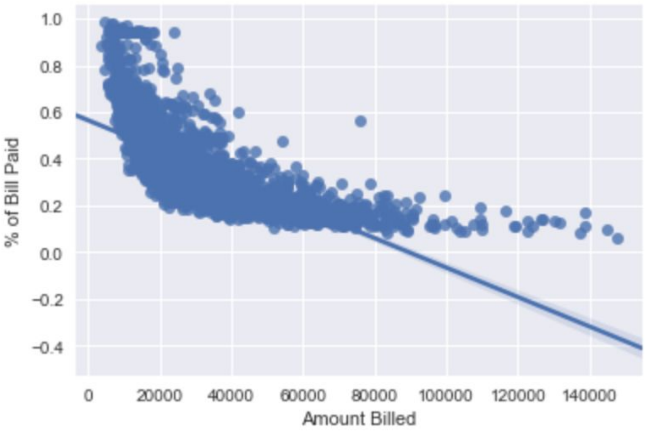
```
                              OLS Regression Results
==============================================================================
Dep. Variable:     Average_Total_Payments   R-squared:                       0.103
Model:                                OLS   Adj. R-squared:                  0.103
Method:                     Least Squares   F-statistic:                     382.0
Date:                    Mon, 11 Dec 2017   Prob (F-statistic):           1.31e-80
Time:                            17:38:14   Log-Likelihood:                -31215.
No. Observations:                    3330   AIC:                         6.243e+04
Df Residuals:                        3328   BIC:                         6.245e+04
Df Model:                               1
Covariance Type:                nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept        7684.1144     89.799     85.570      0.000    7508.047    7860.182
Total_Discharges   39.3857      2.015     19.544      0.000      35.434      43.337
==============================================================================
Omnibus:                     1059.413   Durbin-Watson:                   1.418
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3953.959
Skew:                           1.548   Prob(JB):                         0.00
Kurtosis:                       7.349   Cond. No.                         81.0
==============================================================================
```

These results aren't surprising, given that urban areas have more discharges and higher cost of living. To control for these higher costs, we joined our healthcare data to data provided by the IRS, which details adjusted gross income by zip code. With this information we were able to control for the income level within a zip code by including it as a regressor. In our first iteration we didn't control for MDC and that led to statistically significant results of a positive relationship between average total payments and number of discharges (printout shown below). Once we accounted for MDC in our group by statement, the R squared feel slightly but the correlation remained positive. These results are surprising in that, even after controlling for income level, average total payments is still positively correlated with total discharges, meaning that there doesn't seem to be any economies of scale at play. This could be one of the reasons for the extremely high cost for routine operations in the United States compared to other countries -- as our hospitals get larger, they don't become more efficient.



```
                              OLS Regression Results
==================================================================================
Dep. Variable:     Average_Total_Payments   R-squared:                       0.096
Model:                                OLS   Adj. R-squared:                  0.095
Method:                     Least Squares   F-statistic:                     166.4
Date:                    Mon, 11 Dec 2017   Prob (F-statistic):           2.04e-69
Time:                            21:12:51   Log-Likelihood:                -29471.
No. Observations:                    3138   AIC:                         5.895e+04
Df Residuals:                        3135   BIC:                         5.897e+04
Df Model:                               2
Covariance Type:                nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             7962.0343     82.135     96.938      0.000    7800.990    8123.078
Total_Discharges        11.0004      1.329      8.279      0.000       8.395      13.606
Average_Gross_Income     0.0060      0.000     14.778      0.000       0.005       0.007
==============================================================================
Omnibus:                     1443.905   Durbin-Watson:                   1.524
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18626.985
Skew:                           1.841   Prob(JB):                         0.00
Kurtosis:                      14.353   Cond. No.                     2.78e+05
==============================================================================
```

Something important to remember in the economies of scale conversation is outcomes or quality of care. Our data contain no information about of the operation was successful or if the patient was happy with the care they received. Since this is the case, it's possible that large hospitals have a more specialized doctors which cost more, but have better outcomes. Since we don't have any data on this subject, it is beyond the scope of this analysis, but it is an interesting question for future researchers.



After looking at economies of scale, we also wanted to look at how living in a high income zip code impacts both % of total payments that were out of pocket, and % of the bill that was paid. Before discussing the results, it's important to note that you must be over 65 or have a disability to be eligible for medicare. An older individual might be living in a high income area, but not be as financially well off as other people in their zip code. Our analysis found that there is a slightly negative correlation between % of bill paid and average income in a zip code. These results held when we included MDC, but like similar regressions, the inclusion of MDC decreased the R Squared. These results also made more sense when we looked at Average Income in a zipcode vs Billed Charges. Higher income areas, on average, have higher Bills Charged, furthermore, another regression showed that a smaller % of the bill is paid, as the bill amount increases.

Medical Condition (MDC) & Provider Cost Analysis

This section covers the cost analysis of both MDC procedures and providers. During the proposal process we were primarily concerned with the analysis of the cost of MDC at the provider levels. Specifically, we were interested in analyzing the cost of Average Medicare Payments, Average Total Payments, and Average Covered Charges across both the MDC and providers.

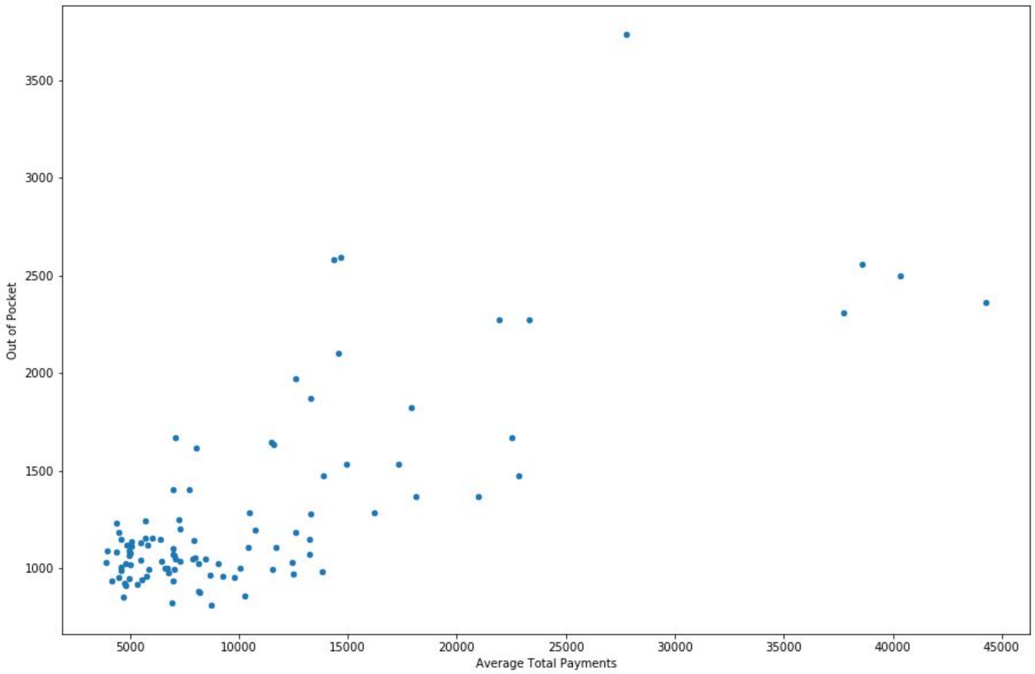
During the initial analysis, I quickly noticed a relationship between two variables: Average Medicare Payments and Average Total Payments. In the official description of the variables one of the major differences between Average Medicare Payments and Average Total Payments is the inclusion of the copayments and deductible amounts for the patient, which I would like to call them 'out-of-pocket' expenses. In figure 3.1, you can see the summary of the average difference between the 'Average Medicare Payment' and 'Average Total Payment".

Figure 3.1 Average difference between 'Average Medicare Payment' and 'Average Total Payment' for each MDC

count	100
mean	1,284.25
std	495.53
min	810.6
25%	1,002.04
50%	1,098.12
75%	1,368.31
max	3,735.07

Given that the only difference between the two variables is the copay and deductible, you can deduce that the average deductible for each Medicare payment is $1,284.25. This number is the total out-of-pocket payments for patients. Looking more closely at this, in figure 3.2, you can see a loose positive correlation between the out-of-pocket charges and the 'Average Total Payments'. As the 'Average Total Payments' increases, so does the out-of-pocket charges. This can be due to the fact that certain hospital services charge a deductible equal to a percentage of the covered charges.

Figure 3.2 Scatter plot of the patient out-of-pocket expenses and the "Average Total Payments"

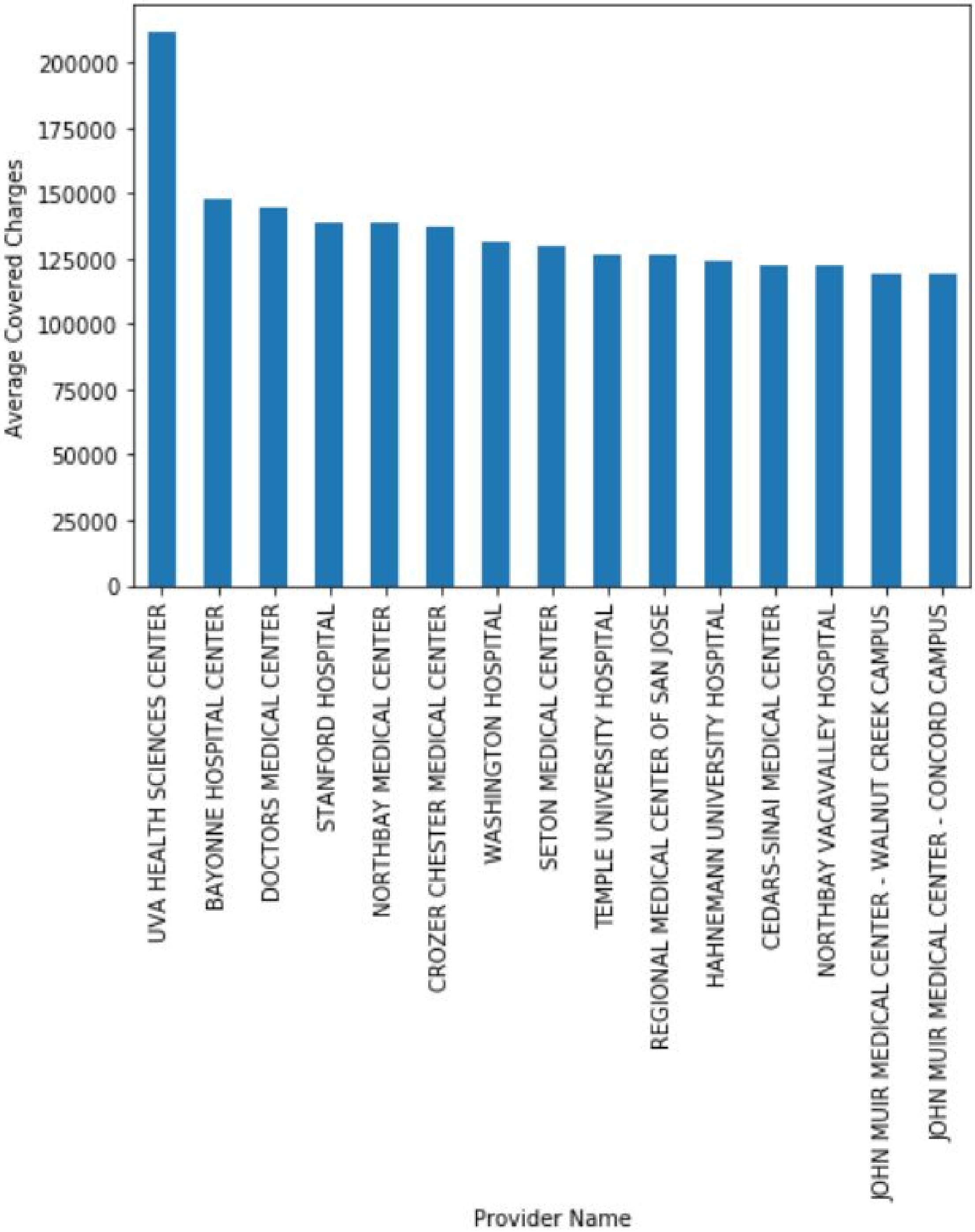


The next interesting finding in the relationship between the 3 cost variables is the relationship between the ‘Average Medicare Payment’ and the ‘Average Covered Charges’. The Average Covered Charges is the total amount that the provider charges for the services. This number is up to the cost structure of each individual hospital. After comparing the two data sets for each MDC, I noticed that there was a direct correlation between the two amounts. In figure 3.3 you can see a tightly correlated relationship between the two. After running an analysis, I discovered that the average percent of Average Medicare Payment to Average Covered Charges comes out to 23%. This number reflects the Medicare payments reimburse a very low amount of total expenses for the provider. It’s difficult to conclude whether Medicare is an unattractive reimbursement option for the provider or that the covered charges are artificially inflated by the provider since provider have arbitrary ways of coming up with their cost instructors. Regardless, I would have think that private insurance plans have a higher percentage of paying for total costs than Medicare does.

Figure 3.3 Average Medicare Payment vs Average Covered Charges for each MDC

Now let’s dive a bit deeper into the analysis of MDC costs themselves. I wanted to take a deeper look at the range of MDC costs for each of these providers. In Figure 3.4, you can see the category Average Total Payment for each of the MDCs. One thing that stood out immediately is how the costliest group, Infectious and Parasitic Diseases, stood out far more than the rest of the group.

Figure 3.5 Average Covered Charges of Provider across all MDCs

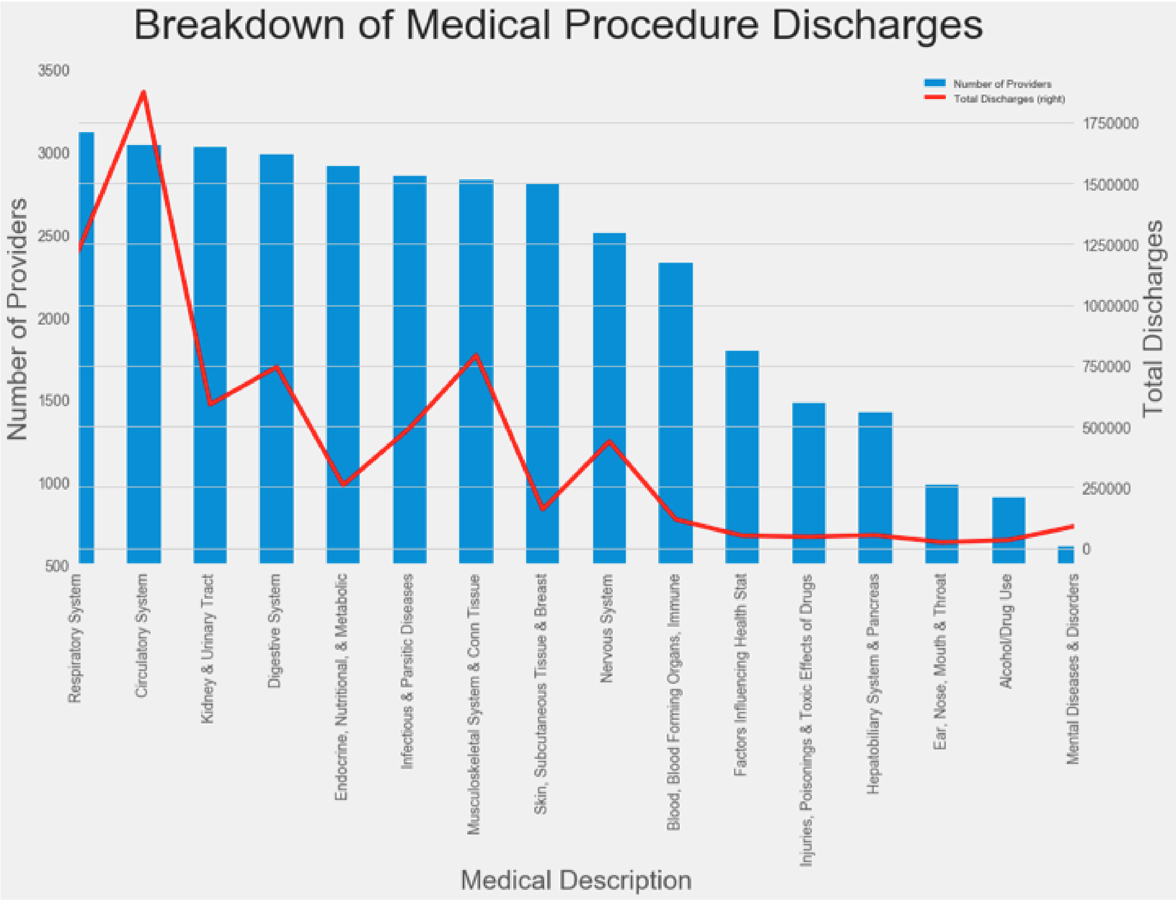


The last part of the MDC & Provider analysis corresponds to the Providers themselves. I was keen on finding out which hospitals were charging the most in Average Covered Charges, because as you recall, this variable is subjective to the providers own cost structure. After averaging the MDC across each provider, I plotted the results of the data in figure 3.5. These results show the top 15 providers by Average Covered Charges. My initial assessment of these results is that the hospitals in the list are predominantly what I would say are more well-known hospitals and/or large university hospitals. I think it's safe to assume that a hospitals brand and reputation trickles down to a higher average covered charge.

Medical Condition (MDC) Coverage by Hospital and Geography

This section explores relationships and trends outside of the currency fields in the dataset. We seek to uncover relationships between Medical Conditions, Discharge Totals, and Medicare Beneficiary Populations.

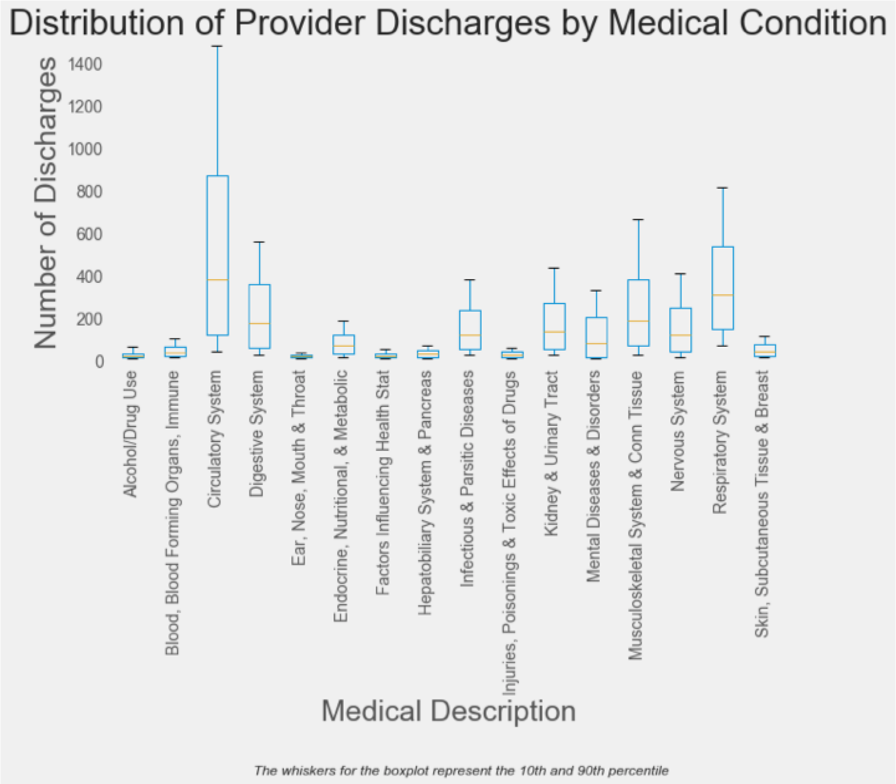
Breakdown of Medical Procedure Discharges



The graph above highlights the total number of providers that cover each of the major medical conditions in our dataset. By adding a secondary y-axis showing Total Discharges, we uncovered the following findings:

- Medicare patients requiring respiratory and circulatory system procedures had far more discharges in 2011 when compared to other medical conditions. Given that people over the age of 80 made up 25% of the Medicare population in 2011, it makes sense to see a large number of discharges pertaining to these conditions, as opposed to conditions on the other end of the spectrum, such as alcohol/drug use.
- On the contrary, it was surprising to see a relatively low number of Nervous System discharges, given an estimated 5.3 million Americans over the age of 65 have Alzheimer's. Perhaps not all of those citizens are Medicare beneficiaries.
- While there is indeed some positive correlation, at the national level, between Total Discharges for a Medical Condition and the Number of Providers that offer coverage for that condition, the spikes in the graph indicate that it is not that strong. Our Pandas pairwise correlation between the two variables corroborated this assumption (corr = 0.66).

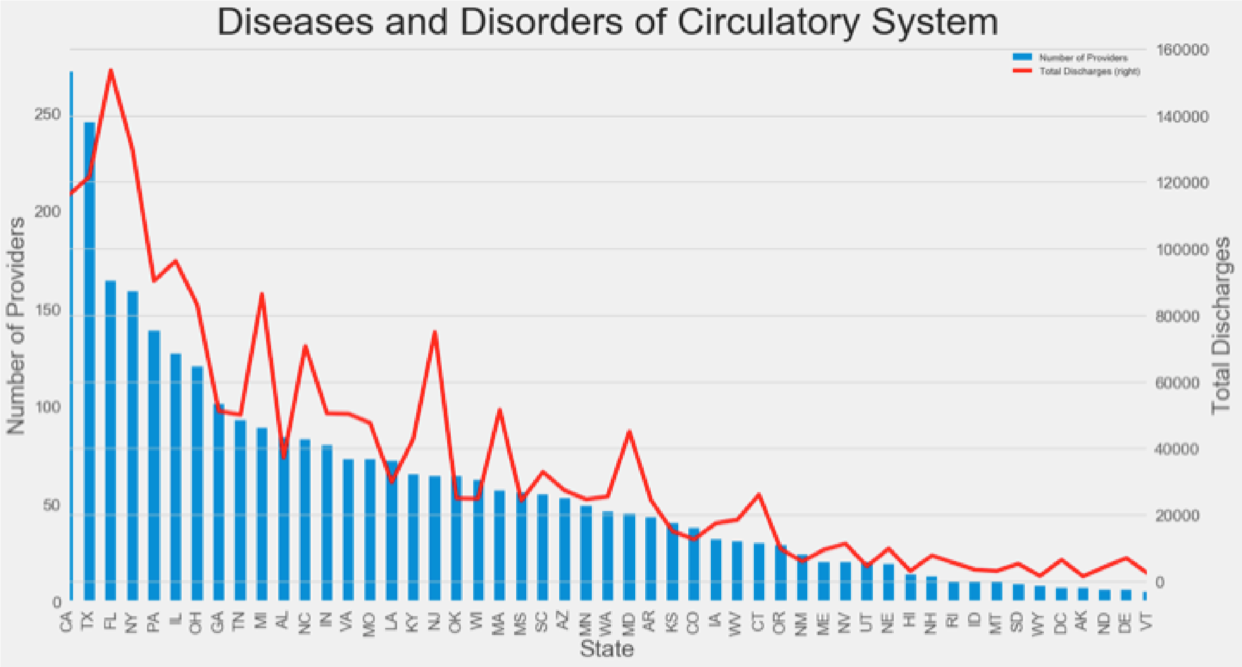
While the graph above is certainly informative, it is depicting aggregate totals and does not give much insight into the distribution of discharges per condition. It could very much be the case that a majority of discharges are occurring in just a handful of hospitals. Due to this speculation, we created the figure below:



The distribution among Circulatory System discharges immediately stands out:

- There is a wide discharge distribution for Circulatory System conditions, with a maximum number of provider discharges exceeding 1600 for this condition, and a minimum number of provider discharges for this condition totaling 11
- The highest number of provider discharges for nearly all other medical conditions does not exceed the third quartile of Circulatory System discharges
- There were many outliers for each condition, presumably representing specialty clinics.

From the box plot above, our team concluded that conditions relating to the Circulatory System were more prevalent across the country for Medicare beneficiaries than any other medical condition in our dataset. To get a state-by-state breakdown on discharges for this condition, we created the plot below:

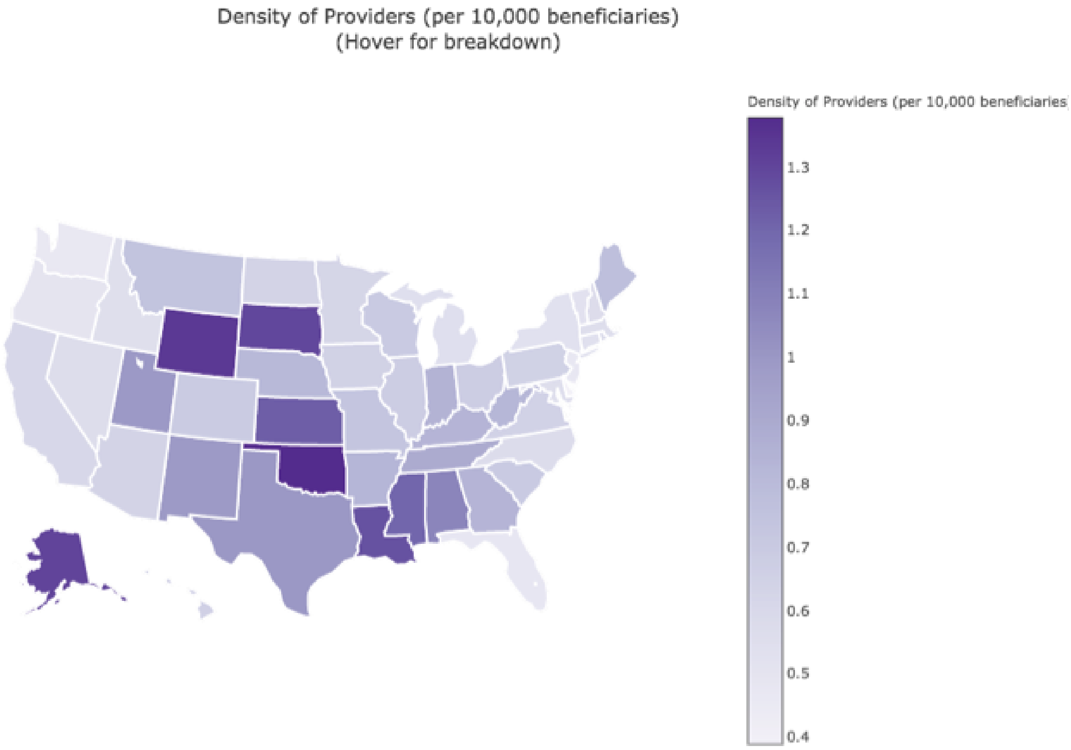


The biggest takeaway from this plot is that Florida owned the highest number of discharges for Medicare patients with Circulatory System conditions. At first glance, we expected this, as Florida is an extremely popular landing spot for post-retirement; people over the age of 60 make up 23% of Florida's population. However, after further research, we discovered that California outnumbered Florida in Medicare beneficiaries by over 1.6 million people. To motivate future exploration, listed below are our speculations as to why Florida outnumbers California in discharges for this condition:

- Perhaps California's smaller population density points to less accessibility?
- Maybe geographic factors play a role in prevalence of circulatory conditions?
- Maybe more outpatient discharges for this condition in California, as opposed to inpatient discharges (which is captured in this dataset)

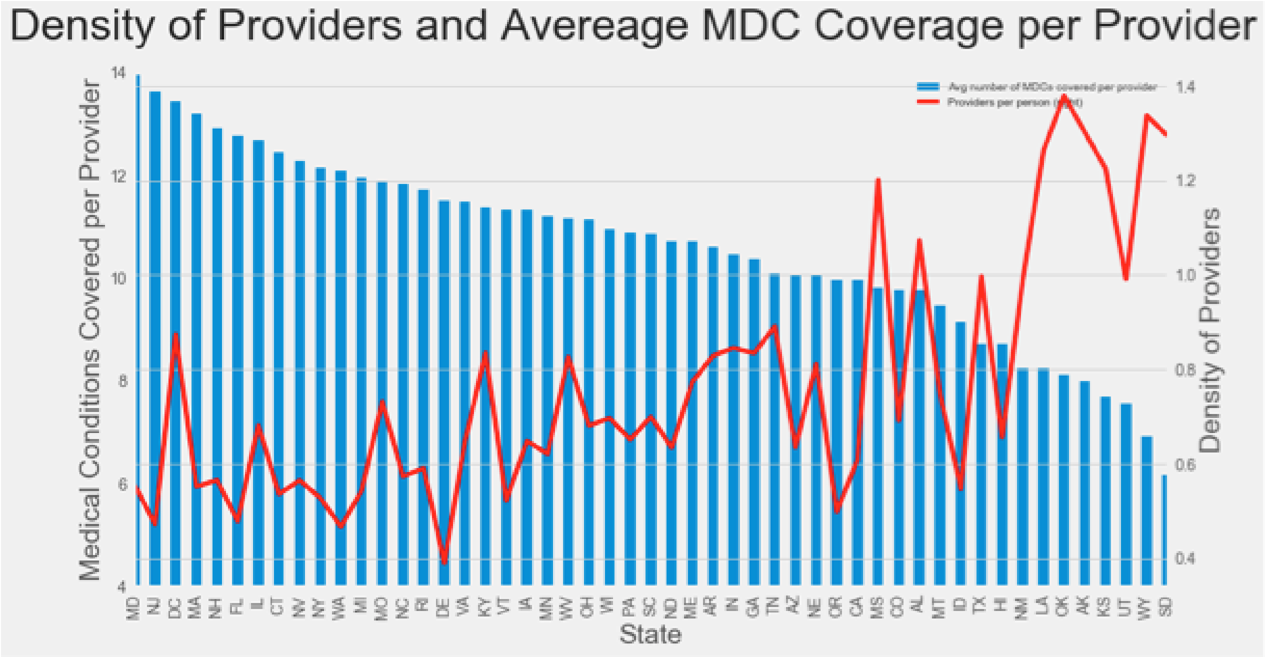
Beneficiary Data

For this section, we extended our dataset by joining Medicare beneficiary population sizes by state in the year 2011. To get a better understanding of provider accessibility, we first calculated the density of providers (per 10,000 beneficiaries) per state that offer coverage for Medicare. Using Plotly, an interactive graphing library for Python, we were able to build a US heat map below to display the different provider density values per state.


Density of Providers (per 10,000 beneficiaries)
(Hover for breakdown)
Density of Providers (per 10,000 beneficiaries
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4

To view this map in interactive mode and be able to hover over states for breakdowns of density values, please go to **https://plot.ly/~sharadv/4/density-of-providers-per-10000-beneficiaries-hover-for-breakdown/**

While this map is useful with regards to density, it does not shed light on the number of conditions that each of these providers cover on average. Therefore, we created the bar chart below:



Surprisingly, we observe a relatively negative correlation between the density of providers per state and medical conditions covered per provider per state. This was confirmed when calculating the pairwise correlation between the two variables (corr = -0.76). Looking at both ends of the spectrum:

- In Maryland, each provider covers on average 14 conditions for Medicare beneficiaries, or 87.5% of the conditions represented in our dataset. However, there is also an underwhelming 0.55 providers per 10,000 beneficiaries, one of the worst ratios in the country based on our findings
- In South Dakota, providers cover on average 6 conditions for Medicare beneficiaries, or 40% of the conditions represented in our dataset. However, there are 1.3 providers per 10,000 beneficiaries, one of the best ratios in the country based on our findings.

Our findings here imply that the care that a Medicare beneficiary requires may not be easily accessible for two reasons:

- There are many Medicare-covered providers located in your state, but they may not cover the specific condition you suffer from
- There may be many conditions covered by providers in your state, but not as many provider options to service the Medicare population.

Medicare Provider Search Tool

The team built a tool to search for Medicare providers that treat a specific medical condition within a given radius of a user's zip code, returning results from least-to-most costly.

Tool Inputs:

- User zip code
- Medical condition
- Search radius

Tool Outputs:

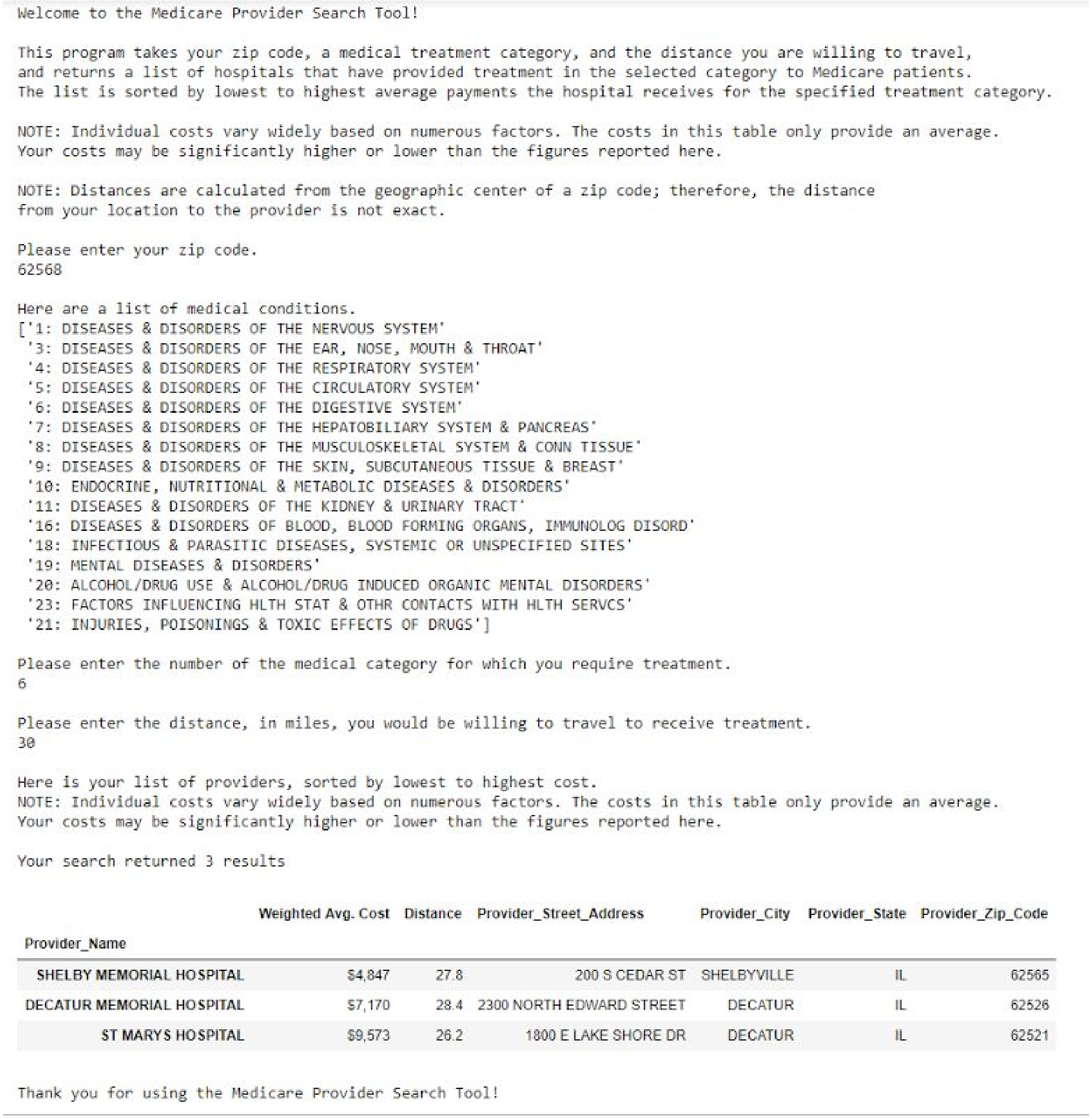
- List of providers within the radius that treat the medical condition selected.

The screenshot below illustrates the tool's functionality.

At this point, the tool has minimal error checking, and is designed only to run in Jupyter Notebook, but it would be easy to port to a command line version, or presumably port to a web-page interface.

Tool speed could also be improved considerably, by any of the following techniques:

- Using multiple processors
- Fetching zip code distances immediately after a user enters their zip code, while the user is entering a medical condition and radius (again using parallel processing)
- Generating a master lookup table of distances among all zip codes



```
Welcome to the Medicare Provider Search Tool!

This program takes your zip code, a medical treatment category, and the distance you are willing to travel,
and returns a list of hospitals that have provided treatment in the selected category to Medicare patients.
The list is sorted by lowest to highest average payments the hospital receives for the specified treatment category.

NOTE: Individual costs vary widely based on numerous factors. The costs in this table only provide an average.
Your costs may be significantly higher or lower than the figures reported here.

NOTE: Distances are calculated from the geographic center of a zip code; therefore, the distance
from your location to the provider is not exact.

Please enter your zip code.
62568

Here are a list of medical conditions.
['1: DISEASES & DISORDERS OF THE NERVOUS SYSTEM'
 '3: DISEASES & DISORDERS OF THE EAR, NOSE, MOUTH & THROAT'
 '4: DISEASES & DISORDERS OF THE RESPIRATORY SYSTEM'
 '5: DISEASES & DISORDERS OF THE CIRCULATORY SYSTEM'
 '6: DISEASES & DISORDERS OF THE DIGESTIVE SYSTEM'
 '7: DISEASES & DISORDERS OF THE HEPATOBILIARY SYSTEM & PANCREAS'
 '8: DISEASES & DISORDERS OF THE MUSCULOSKELETAL SYSTEM & CONN TISSUE'
 '9: DISEASES & DISORDERS OF THE SKIN, SUBCUTANEOUS TISSUE & BREAST'
 '10: ENDOCRINE, NUTRITIONAL & METABOLIC DISEASES & DISORDERS'
 '11: DISEASES & DISORDERS OF THE KIDNEY & URINARY TRACT'
 '16: DISEASES & DISORDERS OF BLOOD, BLOOD FORMING ORGANS, IMMUNOLOG DISORD'
 '18: INFECTIOUS & PARASITIC DISEASES, SYSTEMIC OR UNSPECIFIED SITES'
 '19: MENTAL DISEASES & DISORDERS'
 '20: ALCOHOL/DRUG USE & ALCOHOL/DRUG INDUCED ORGANIC MENTAL DISORDERS'
 '23: FACTORS INFLUENCING HLTH STAT & OTHR CONTACTS WITH HLTH SERVCS'
 '21: INJURIES, POISONINGS & TOXIC EFFECTS OF DRUGS']

Please enter the number of the medical category for which you require treatment.
6

Please enter the distance, in miles, you would be willing to travel to receive treatment.
30

Here is your list of providers, sorted by lowest to highest cost.
NOTE: Individual costs vary widely based on numerous factors. The costs in this table only provide an average.
Your costs may be significantly higher or lower than the figures reported here.

Your search returned 3 results
```

	Weighted Avg. Cost	Distance	Provider_Street_Address	Provider_City	Provider_State	Provider_Zip_Code
Provider_Name						
SHELBY MEMORIAL HOSPITAL	$4,847	27.8	200 S CEDAR ST	SHELBYVILLE	IL	62565
DECATUR MEMORIAL HOSPITAL	$7,170	28.4	2300 NORTH EDWARD STREET	DECATUR	IL	62526
ST MARYS HOSPITAL	$9,573	26.2	1800 E LAKE SHORE DR	DECATUR	IL	62521

```
Thank you for using the Medicare Provider Search Tool!
```

Appendix

Bibliography

- The primary data is located here: https://www.kaggle.com/speedoheck/inpatient-hospital-charges/data
- The data was originally collected by the Centers for Medicare & Medicaid Services (CMS) and posted on data.gov at

https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3

- A “Methodological Overview” for this data is located here: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Inpatient_Methodology.pdf
- Zip code demographic data was sourced from the US Census’ American FactFinder tool at https://factfinder.census.gov/.
- Medicare Beneficiary Population Size data was sourced from here: https://www.kff.org/medicare/state-indicator/total-medicare-beneficiaries/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Data Fields

Definitions from Methodological Overview in the Bibliography section.

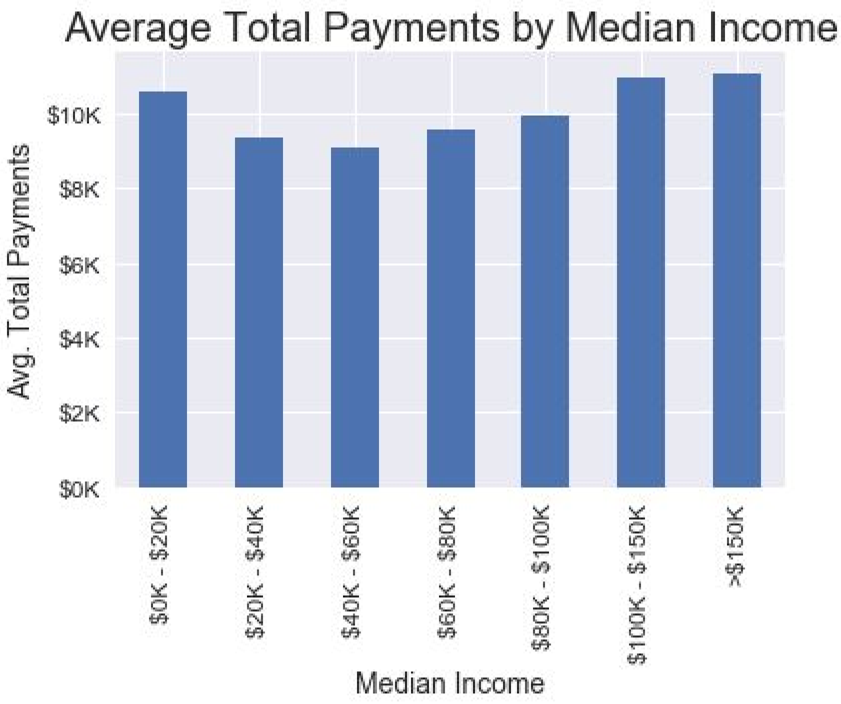
- **DRG Definition**: The code and description identifying the MS-DRG. MS-DRGs are a classification system that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay.
- **Provider Id**: The CMS Certification Number (CCN) assigned to the Medicare certified hospital facility.
- **Provider Name**: The name of the provider.
- **Provider Street Address**: The provider’s street address.
- **Provider City**: The city where the provider is located.
- **Provider State**: The state where the provider is located.
- **Provider Zip Code**: The provider’s zip code.
- **Provider HRR**: The Hospital Referral Region (HRR) where the provider is located.
- **Total Discharges**: The number of discharges billed by the provider for inpatient hospital services.
- **Average Covered Charges**: The provider's average charge for services covered by Medicare for all discharges in the MS-DRG. These will vary from hospital to hospital because of differences in hospital charge structures.
- **Average Total Payments**: The average total payments to all providers for the MS-DRG including the MSDRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Also included in average total payments are co-payment and deductible amounts that the patient is responsible for and any additional payments by third parties for coordination of benefits.
- **Average Medicare Payments**: The average amount that Medicare pays to the provider for Medicare's share of the MS-DRG. Average Medicare payment amounts include the MS-DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Medicare payments DO NOT include beneficiary co-payments and deductible amounts nor any additional payments from third parties for coordination of benefits. Note: In general, Medicare FFS claims with dates-of-service or dates-of-discharge on or after April 1, 2013, incurred a 2 percent reduction in Medicare payment. This is in

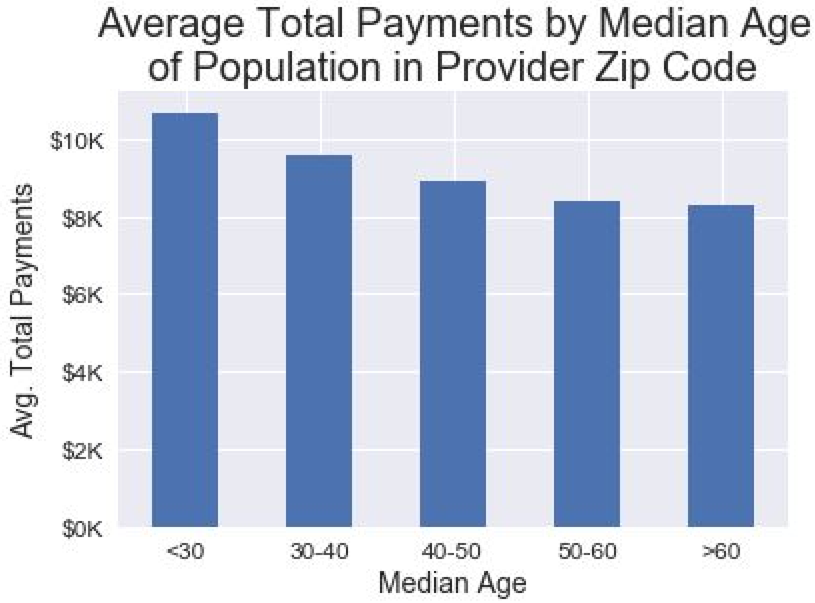
response to mandatory across-the-board reductions in Federal spending, also known as sequestration. For additional information, visit http://www.cms.gov/Outreach-andEducation/Outreach/FFSProvPartProg/Downloads/2013-03-08-standalone.pdf

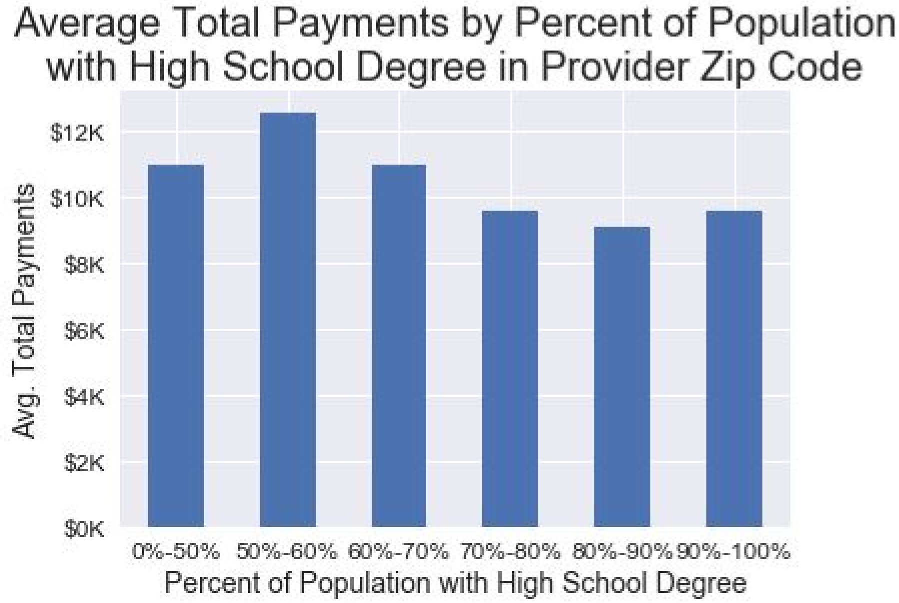
Additional Findings by Zip Code Demographics

This section contains additional charts depicting the relationship to Average Total Payments and various demographics at the zip code level.

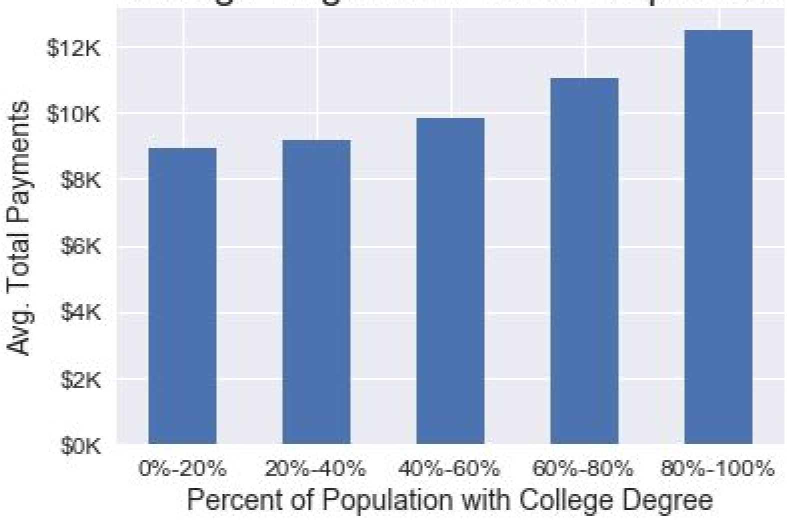
These charts are not supported with individual comments, but the general finding remains the same as in the body of this document, which is that while Average Total Payments vary across some demographics (though not others), all have more variation within their categories than across their categories.



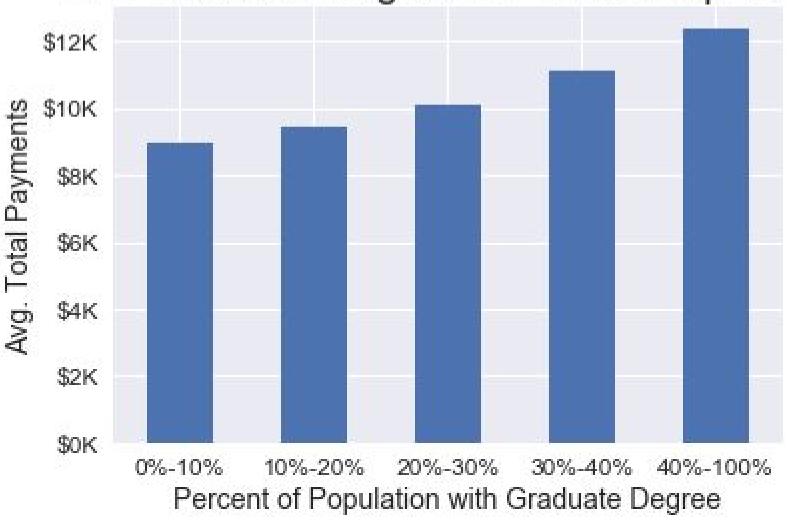

Average Total Payments by Median Age
of Population in Provider Zip Code
Avg. Total Payments
$10K
$8K
$6K
$4K
$2K
$0K
<30
30-40
40-50
50-60
>60
Median Age

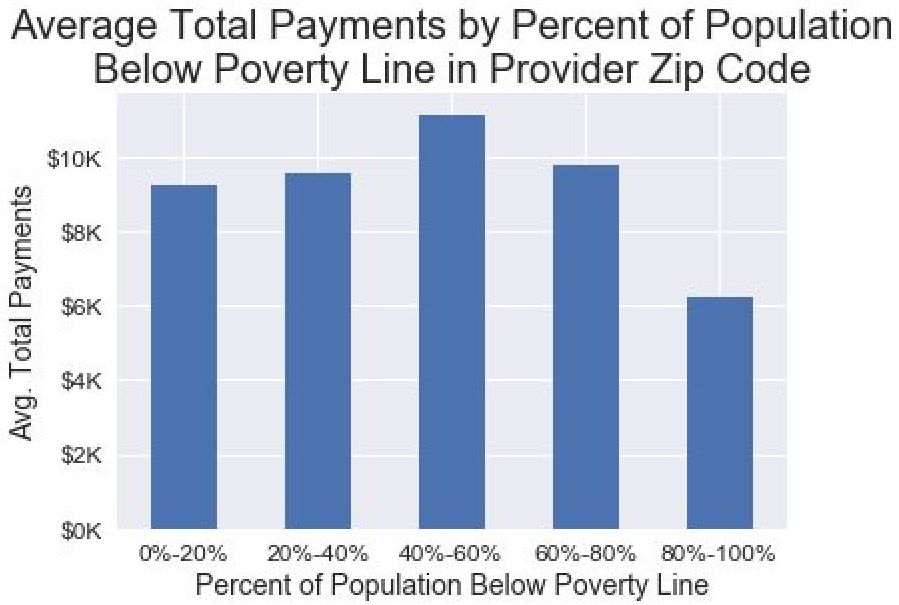

Average Total Payments by Percent of Population
with High School Degree in Provider Zip Code
Avg. Total Payments
$12K
$10K
$8K
$6K
$4K
$2K
$0K
0%-50%
50%-60%
60%-70%
70%-80%
80%-90%
90%-100%
Percent of Population with High School Degree

Average Total Payments by Percent of Population with College Degree in Provider Zip Code

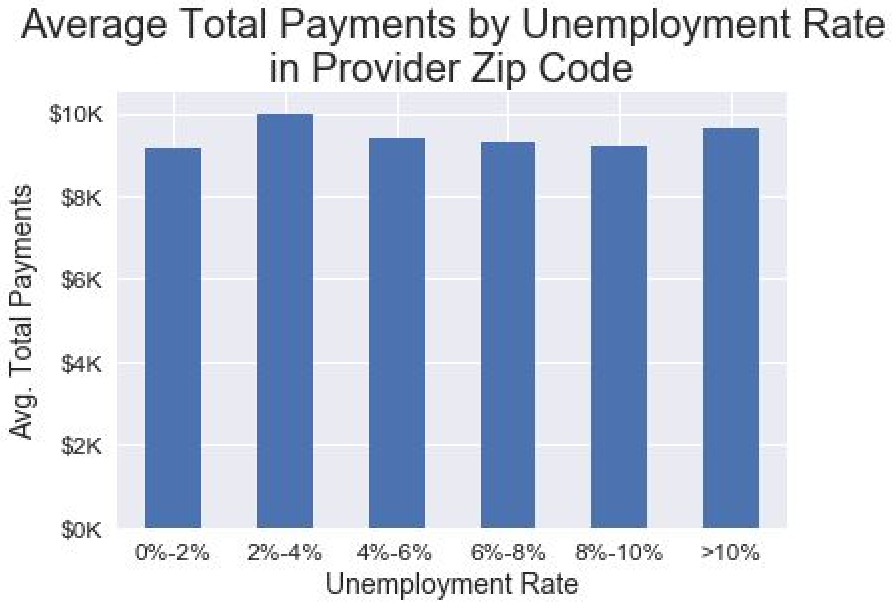


Average Total Payments by Percent of Population with Graduate Degree in Provider Zip Code

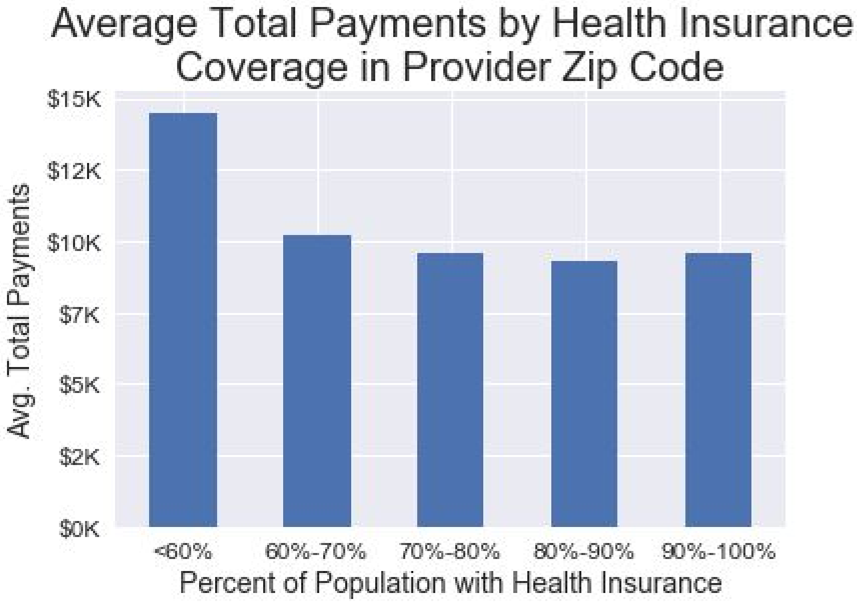




Average Total Payments by Percent of Population Below Poverty Line in Provider Zip Code

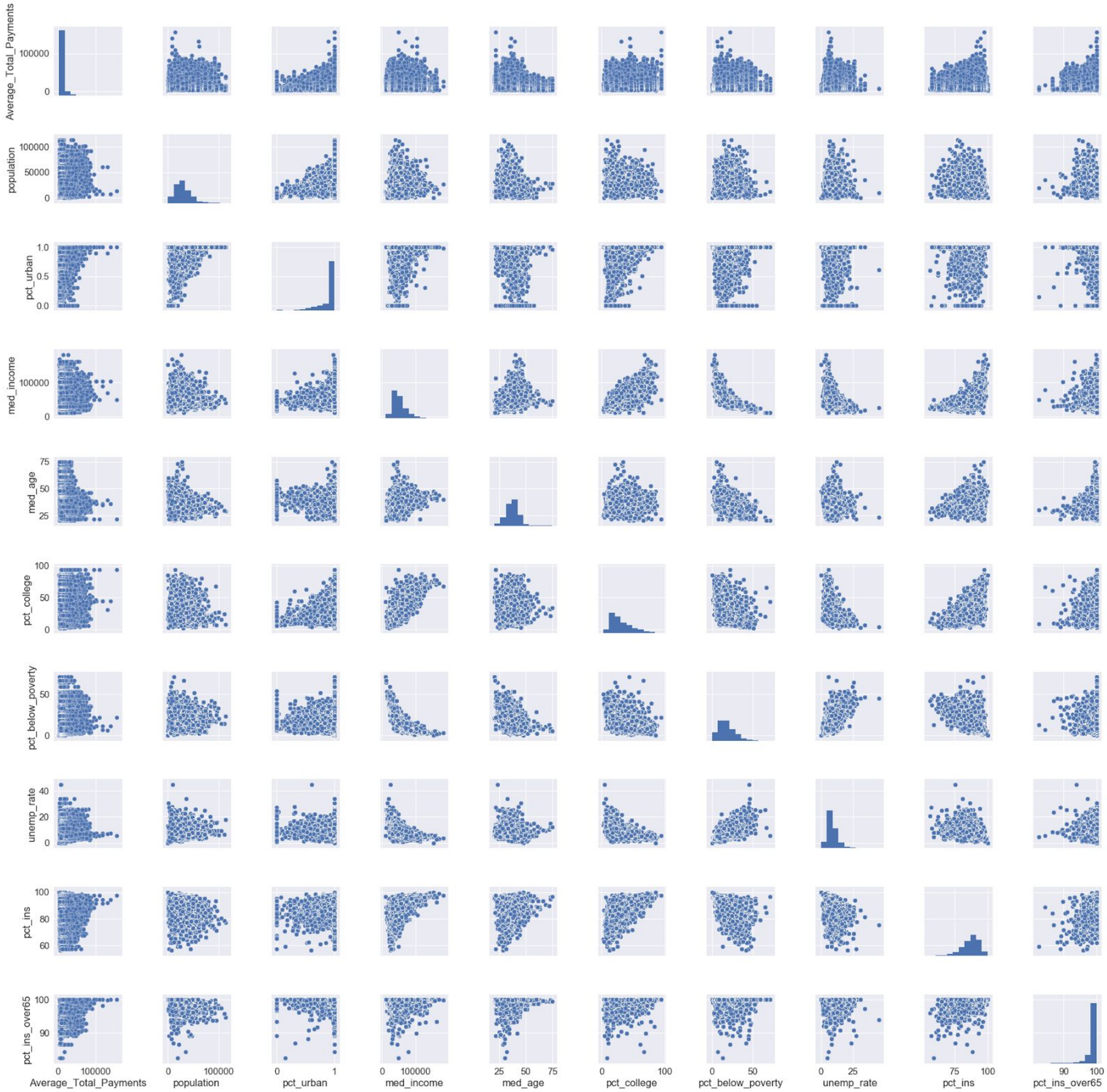


Average Total Payments by Unemployment Rate in Provider Zip Code



Finally, the pairplot below proved useful in data exploration, showing its grid of scatter plots across several continuous variables. Individual variables are not meant to be fully readable in this report, but at a high level, the first row / top column represent Average Total Payments, while the remaining rows represent the numerous zip code demographics. The main finding here is that nearly every scatter plot in the Average Total Payments row / column shows an indistinct “blob,” further supporting the lack of relationship between Average Total Payments and zip code demographics.

This was also a useful tool to “sanity check” the data and proved useful in uncovering errors.


Average_Total_Payments
100000
0
population
100000
50000
0
pct_urban
1.0
0.5
0.0
med_income
100000
0
med_age
75
50
25
pct_college
100
50
0
pct_below_poverty
50
0
unemp_rate
40
20
0
pct_ins
100
80
60
pct_ins_over65
100
90
0
100000
Average_Total_Payments
0
100000
population
0
1
pct_urban
0
100000
med_income
25
50
75
med_age
0
100
pct_college
0
50
pct_below_poverty
0
25
unemp_rate
75
100
pct_ins
90
100
pct_ins_over65