SnakeBait

Tad Tatsuya Murao

Steve Ricard

Charlie Ortman

Sasha Bridges

Project Write-Up

Questions

Our big question was “Is Medicare spent differently for each county?” We broke this up into several smaller questions, such as “What are the median/mean incomes of our chosen counties?”, “Is there a difference between the amount of data recorded between counties?”, “Are these counties meeting their estimates?”, “What is the county’s average score for Medicare spending?” and so on.

Process

First, we had to find data files to work with. Tad suggested looking at medical data files. Charlie wanted to look up disparity between Medicare and hospital spending. We looked for data at CMS.gov, Medicare.gov, and the Census website. From there, we found our initial hospital data files. Originally, we intended to only use “Complications and Death – Hospital”, “Readmissions – Hospital”, “Medicare Spending per Patient – Hospital”, and “Readmissions and Death – VA\_07\_10\_2018,” but eventually started looking for more data to compare our counties’ income median and mean. “Readmissions and Death” was eventually dropped for being too narrow and not providing enough data – later we found out that this was because it was *only* data regarding the VA hospitals.

We spent a lot of time trying to figure out what “score” meant for each file. It turns out that the definition was different for each file, and how it was used was *also* different. There was also no documentation about this in the included pdf, meaning we had to ask the person who compiled all this information what the scores meant.

After we finished the initial processes of understanding the basic data, we began the clean up.

Data Clean Up

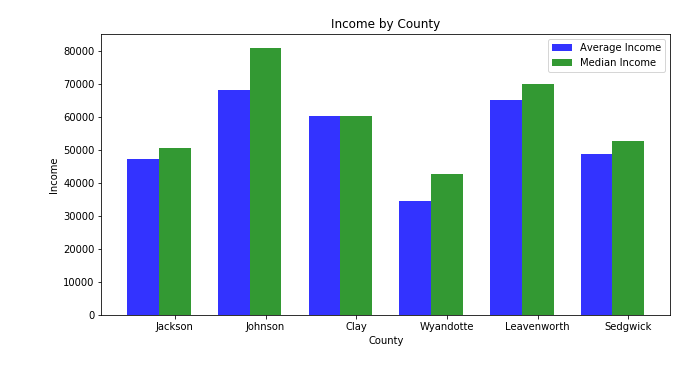
There was a lot of data to sift through. Our four main files shared a lot of common data, such as county names, but also had a lot of useless information for our purposes, such as phone numbers. We cleared out all the data except state, county, score, denominator, and estimates. Then we tried to mean the score – and couldn’t. There were several columns with “Not Available” data written as a string. This wouldn’t parse to an int without the string transforming into a “NaN” type. You can drop the NaN, but it still leaves the column as an object type. Not helpful!

Eventually we went into the csv files and got rid of all the “Not Available” rows. We couldn’t use those rows anyway, since they had no data recorded. But we sure tried.

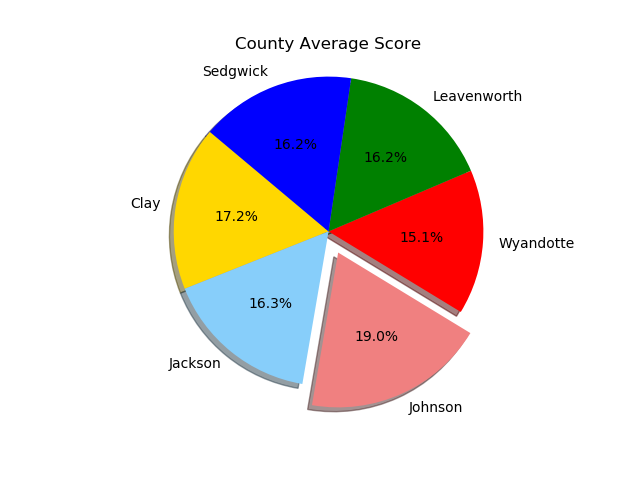
After that, we didn’t have any more issues until we tried to make new graphs. The Y-axis was acting strangely, so we tried to take out DataFrame columns individually and then try to place them into a new DataFrame. We tried to wrangle the data for hours, but alas, it defeated us, and we were forced to edit csv files again.

Data Analysis

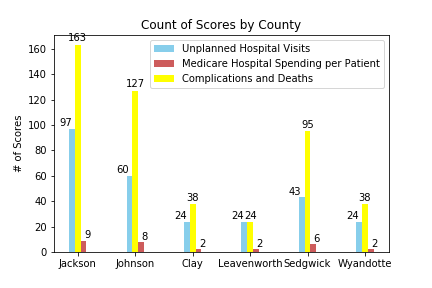
This was the fun part. We didn’t know what we were looking for – just something interesting. And we found out a lot of neat stuff! Johnson has the highest median and mean income. They also have the highest amount of Medicare money spent per patient. Johnson county is the second highest rank in terms of data recording. They also had the highest estimation regarding deaths/complications, and they exceeded that number.



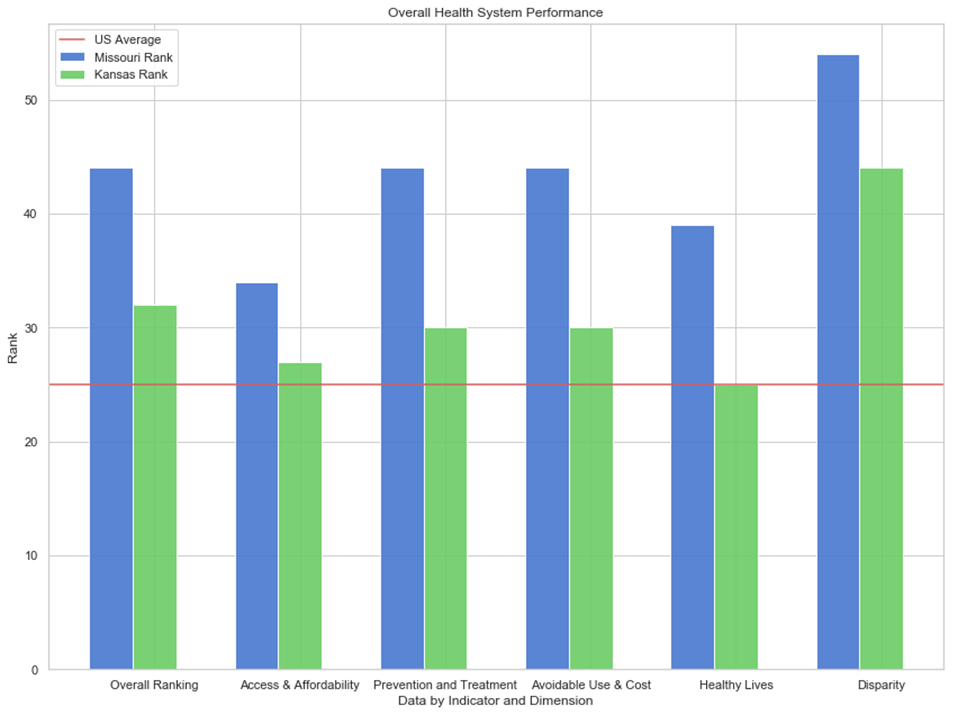
A simple median/mean chart of income by county would give a baseline of how rich each county was.



One of the burning questions that we had was if there was a disparity between how much each county spent per patient through Medicare. It turns out that there is! Wyandotte spends the least amount of money per patient, while Johnson spends the most. Better stay in Johnson county!

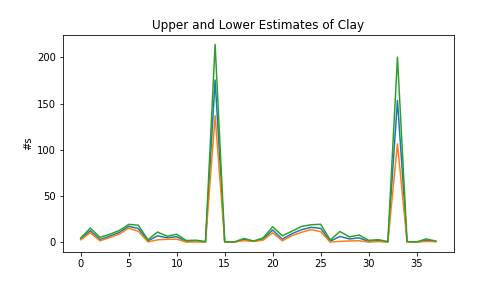


When doing analysis, you also have to consider where your data came from and how much came from a singular location. We considered a simple question: which county recorded the most data? Turns out it was Jackson! They very adamantly recorded what they could. Johnson also recorded a lot of data.

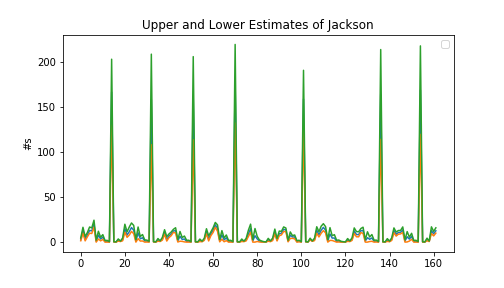


We wanted to compare the states of Missouri and Kansas to the national average. Missouri was above average for everything, while Kansas was above average for everything BUT healthy lives. What a shame, Kansas. What a shame.

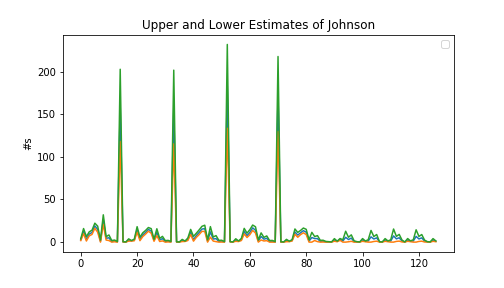
We also wanted to look at the upper estimates of each county and compare how them to how they actually scored, and it was enlightening. The orange lines are the lower estimate, the blue lines are the higher estimates, and the green line is the actual score for each county.



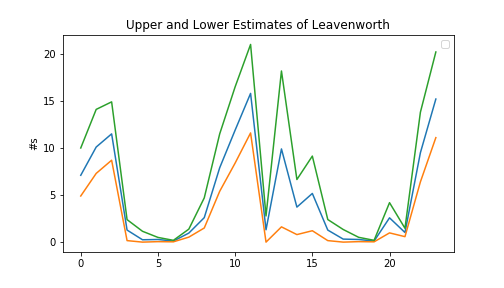
Clay had big, dramatic spikes twice.



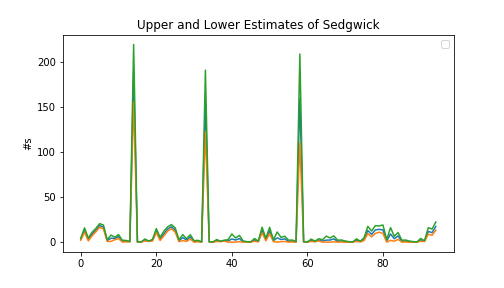
Jackson had a great amount of spikes almost routinely.



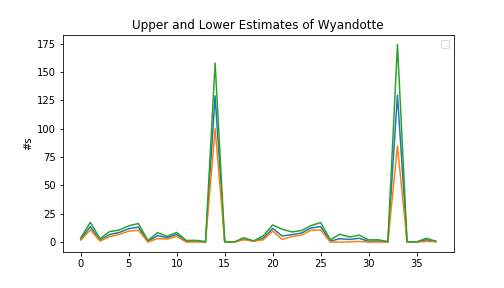
Johnson had a lot of spikes in the beginning, but they tapered off towards the end.



Leavenworth didn’t have high spikes, or much data recorded, but it had the most interesting and dramatic displays.



Sedgwick had three major spikes but was mostly well below 50 during the recorded timeframe.



Wyandotte had interesting data. They were almost entirely above their higher estimate during the data collection time. It spiked *high*, well above their estimate, but was below 25 for everything *but* those two spikes.