A Retrospective Study of U.S. Healthcare Cost

Khalil Khouy, Ashrita Challa, Sravani Konjeti, Manoj Barman

Abstract—We keep hearing different highlights and propagandas with mixed sentiments on the news and other social media outlets about issues and deficiencies in the U.S. Healthcare system with no data to back them up. The most major claim is that the U.S. healthcare system is one of the most expensive ones in the world with the least effective care; therefore, we decided to see for ourselves by fetching and visualizing data relevant to that claim and pursuing key drivers to the rising of U.S. national healthcare cost, if there are any.

Index Terms—Healthcare, study, propagandas, effective, least, expensive, deficiencies, visualizing, highlights

1 Introduction

According to a recent commonwealth fund study, the U.S. spends around $2.7 trillion each year on Care, of which $800 billion is wasteful, redundant, or inefficient due to many key drivers. Furthermore, out of 11 developed countries' health care systems, the U.S. is ranked 5th in quality and last in preventing deaths from treatable conditions (e.g., strokes, diabetes, certain treatable cancers).

Our data include hospital-specific charges for more than 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments for the top 100 most frequently billed discharges, paid under Medicare based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for fiscal year 2011. These DRGs represent more than 7 million discharges or 60 percent of total Medicare IPPS discharges. Our data comprises of 163k records and 12 features.

We removed DRG definition, provider ID, provider name, provider street address, provider city, provider zip code, provider HRR variables. We included provider state, total discharges, average covered charges, average total payments and average Medicare payments variables.

It might be surprising of removing address and zip code columns. But we decided on this as we are showing our visualization based on state only. User has the interface to select the state from the drop-down and based on that state, our visualizations are changed according to the data related to that state.

We also have a twitter dataset with 102 tweets extracted and filtered using the twitter API. We had ID, date, user, tweet and location variables in the dataset. We removed ID, date and user variables as we didn't use them for our visualization. We used the Tweet and location variables for our visualization.

We have a third dataset on total national health expenditures ranging from years 1970 to 2016. We have the Present value of these expenditures as per 2016. These values are in US Billion dollars. We tried to plot a time series analysis graph by using these values.
Below are the snapshots of our datasets as samples.

1.1 Extract of Inpatient Prospective Payment System Dataset

State	StateName	Lat	Long	SumofTotalCost	CountofTotalDischarges	SumofAverageCoveredCharges	Variance
AK	Alaska	61.37072	-86.7911	$90,149,869.65	231	$9,320,559.71	10
AL	Alabama	32.80667	-152.404	$1,098,626,721.86	3635	$113,835,339.64	10
AR	Arkansas	34.9697	-111.431	$682,473,694.66	2067	$54,102,745.75	8
AZ	Arizona	33.72976	-92.3731	$1,023,640,012.85	2851	$117,461,379.67	11
CA	California	36.1162	-119.682	$5,862,212,436.91	13064	$881,932,566.42	15
CO	Colorado	39.05981	-105.311	$590,703,840.27	1890	$77,669,807.25	13
CT	Connecticut	41.59778	-72.7554	$1,061,988,486.42	2011	$62,981,322.74	6
DC	District of C	38.89744	-77.0268	$250,592,560.70	462	$18,533,898.61	7
DE	Delaware	39.31852	-75.5071	$276,029,544.25	394	$10,666,249.66	4
FL	Florida	27.76628	-81.6868	$4,499,881,666.21	11155	$513,311,085.66	11
GA	Georgia	33.04062	-83.6431	$1,709,438,163.82	4968	$154,489,562.36	9
HI	Hawaii	21.09432	-157.498	$149,025,177.76	442	$14,221,238.65	10
IA	Iowa	42.01154	-93.2105	$589,341,075.51	1734	$41,908,598.70	7
ID	Idaho	44.24046	-114.479	$184,899,873.61	551	$14,086,616.42	8
IL	Illinois	40.34946	-88.9861	$3,327,673,379.09	7909	$285,213,170.69	9
IN	Indiana	39.84943	-86.2583	$1,573,886,307.71	4260	$119,896,475.44	8
KS	Kansas	38.5266	-96.7265	$540,726,358.09	1638	$51,728,455.50	10
KY	Kentucky	37.66814	-84.6701	$1,254,930,783.81	3229	$79,187,373.35	6
LA	Louisiana	31.16955	-91.8678	$917,263,816.49	3027	$100,149,423.44	11
MA	Massachuse	42.23017	-71.5301	$1,920,569,076.50	3842	$78,891,653.79	4
MD	Maryland	39.06395	-76.8021	$2,043,982,385.22	3330	$44,548,086.62	2
ME	Maine	44.69395	-69.3819	$303,458,283.98	888	$18,110,722.32	6

1.2 Extract of Tweet dataset by State

State	StateName	Lat	Long	TrumpcareTweet1	NontrumpcareTweet2
AK	Alaska	61.37072	-86.7911	The GOP and the D	Soon will come the "de
AL	Alabama	32.80667	-152.404	The Democrats are	Dems R Not in power o
AR	Arkansas	34.9697	-111.431	Just look at what th	Your economy is growi
AZ	Arizona	33.72976	-92.3731	your healthcare co	Have you done ANYTHI
CA	California	36.1162	-119.682	Illegal alliens cost t	The cost of healthcare
CO	Colorado	39.05981	-105.311	People like you wo	Trump sucks. His tax bil
CT	Connecticut	41.59778	-72.7554	Think some areas y	Things that cost more:
DC	District of C	38.89744	-77.0268	Education & health	Corporations recieved
DE	Delaware	39.31852	-75.5071	. . .@SenateMajLd	Hi , Can you explain wh
FL	Florida	27.76628	-81.6868	did you know that	This is more proof that
GA	Georgia	33.04062	-83.6431	So how about more	I had this thought but d
HI	Hawaii	21.09432	-157.498	Have Democrats ev	Thanks for doing absolu
IA	Iowa	42.01154	-93.2105	Job creation that y	Because wealthy corpo
ID	Idaho	44.24046	-114.479	The point is, Obam	He promised repeal an
IL	Illinois	40.34946	-88.9861	Kind of like when C	Only 7% of all healthca
IN	Indiana	39.84943	-86.2583	I would like to exp	Just like he was going t
KS	Kansas	38.5266	-96.7265	If we where to tak	Healthcare cost skyrock
KY	Kentucky	37.66814	-84.6701	Why would u assur	The @GOP @GOPChair
LA	Louisiana	31.16955	-91.8678	The ObamaCare bil	Great healthcare for ev
MA	Massachuse	42.23017	-71.5301	I had a good health	Trump accomplishment
MD	Maryland	39.06395	-76.8021	I would hardly call	I'm not gonna move tar
ME	Maine	44.69395	-69.3819	Unlike 44, isn't tak	How Can healthcare be

1.3 Extract of Time Series Dataset of total national health expenditures vs 2016 Present Value in US $ Billions (1970-2016)

Year	TNHE	PV2016
1970	74.56	370.02
1971	82.73	393.8
1972	92.66	426.48
1973	102.81	449
1974	116.55	460.99
1975	133.28	486.57
1976	152.74	528.6
1977	173.85	564.98
1978	195.33	593.24
1979	221.53	617.99
1980	255.33	643.23
1981	296.16	684.87
1982	334.04	732.07
1983	367.81	772.8
1984	405	819.91
1985	442.9	866
1986	474.69	908.49
1987	516.52	959.34
1988	579.28	1035.56
1989	644.77	1105.08
1990	721.39	1185.09
1991	788.06	1253.5

2 Related Work

2.1 Related Search Paper 1

The below article shows national spending trends from 1960 to 2016 for health care by the payer. It shows the shift over the years how we pay for hospital care, physician services, long-term care, prescription drugs, and other health care services and products in the US. It is relevant to know how much we pay and for what services in order to understand the US healthcare costs dynamics. https://www.chcf.org/publication/us-health-care-spending-who-pays/

2.2 Related Search Paper 2

The following reference is a collection of charts and metrics used to look at how quality of the U.S. Healthcare system compares to other countries, as well as exploring strengths and weaknesses across different dimensions. https://www.healthsystemtracker.org/chart-collection/quality-u-s-healthcare-system-compare-countries/#item-use-emergency-department-place-regular-doctor-visits-common-u-s-comparable-countries

What inspired us to reference the above metrics is that benchmarking U.S. quality measures against those of similarly large and wealthy countries can create a true perspective of what the U.S. healthcare system should be focusing on as areas of improvement.

2.3 Related Search Paper 3

The below paper references how the hospitals are charging outrageous costs to the patients and possible measures to overcome those costs. As our visualization project is looking into the healthcare of USA, this article gives us a good understanding about real costs of the treatments and how the prices are being made-up.

https://www.consumerreports.org/cro/magazine/2014/11/it-is-time-to-get-mad-about-the-outrageous-cost-of-health-care/index.htm

2.4 Related Search Paper 4

The study provides details of the Healthcare analytics spending by End User which can be Payers, Providers and ACOs/HIEs as well as by region. The study performs a retrospective analysis of the healthcare spending and factors impacting these costs. This study provides a detailed understanding of the healthcare system to be able to interpret the data better and enables the story-telling process for our project by stitching together the facts derived from the dataset. https://kxdocuments.accenture.com/contribution/e44865cf-23a0-4361-8c19-76065419d984?referrer=https://search.accenture.com

3 System Design

We used R programming and R-Shiny to develop our system. Our system contains a map and a word cloud. We show the healthcare costs based on the state. With the help of this we can find out which states have low healthcare costs. People can decide to which state they can move to minimize their healthcare costs in the long-run.

We see that the healthcare costs are more in the state of California followed by New Jersey and Rhode Island.

Plus, people in the state talk about the healthcare costs, the facilities that are being provided, the quality of healthcare they are getting out of spending huge amounts on the healthcare. We got this info from the twitter. We collected the tweets based on location and stored in our dataset. When we select the state, we get a word cloud based on the tweets posted from that state.

The words are represented using a color scale ranging from green to red. Most frequent words are represented bigger and in red color. Least frequent words are represented smaller and in green color. Mid-level frequent words are represented in blue color and in mid-size.

This gives a quick view of people's view on the healthcare in their state in a single snapshot. If we see more green words we can say that they are talking less about that topic and if we see more red words we can say that they are talking more about that topic.

We also can change the frequency of the words and the number of words in the word cloud by using slide bars. The words frequency says that we can consider a word to be on word cloud if the particular word matches the word frequency that was chosen from the slider.

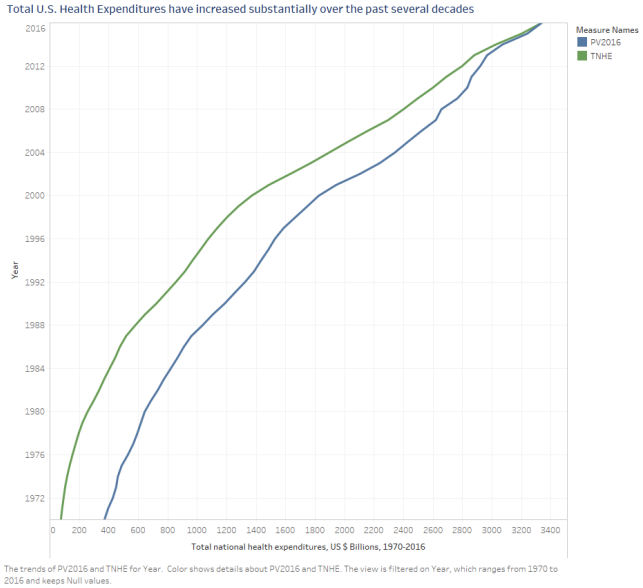
Sometimes user wants to see more frequent words and sometimes they prefer to see less frequent words. This preference is made based on the scenario and the use-case. In such cases this slider is useful.

Also, sometimes user wants to consider more words and sometimes wants to consider less words to be on the word cloud. If the user wants to find out top 20 words or so, then it can be done from the slide bar. Sometimes they prefer to take top 100 words etc. So, to achieve this flexibility we included slide bars with which we can change the number of words to be considered in the word cloud.

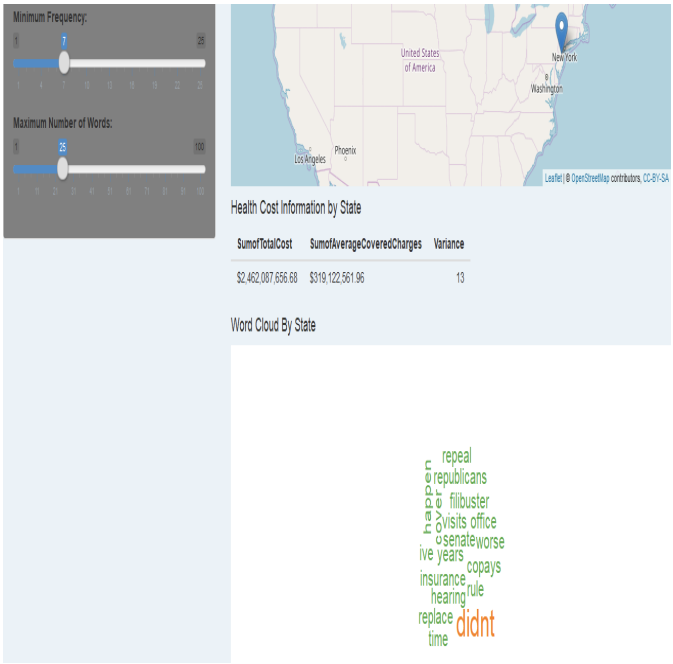
Our time series has green and blue lines. Green line shows the graph of expenditures with their original values ranging from years 1970 to 2016, whereas, blue line shows the Present Values of those expenditures as per 2016. These two lines meet at the year 2016, as obvious.

Below are the snapshots of our final visualizations:

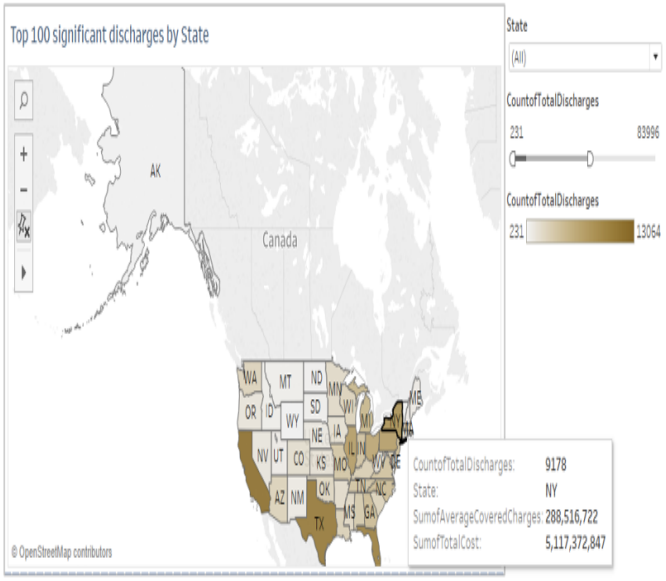
3.1 Interactive Dashboard of U.S. National Health Care Cost Time Series

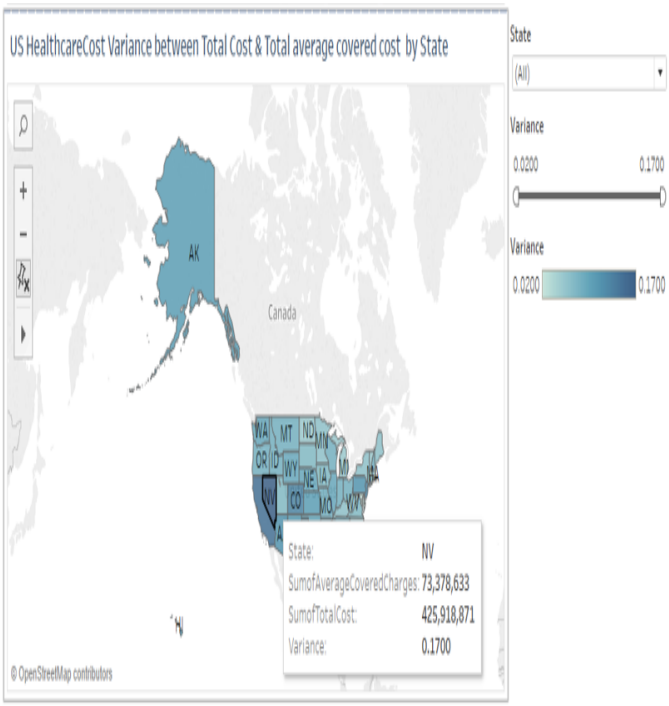


3.2 R Shiny Interactive App Word Cloud of Tweets by State

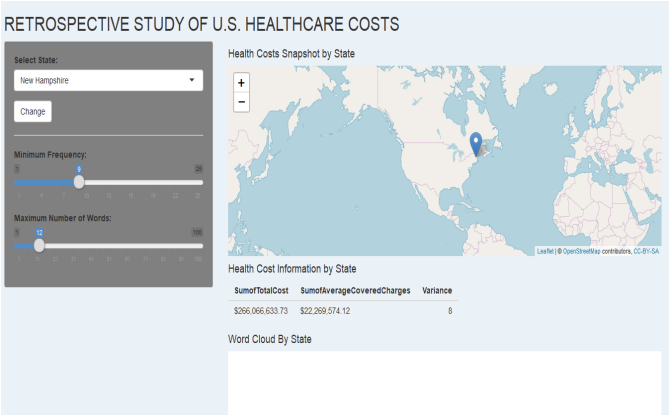


3.3 Interactive Dashboard healthcare cost by state and by top discharges

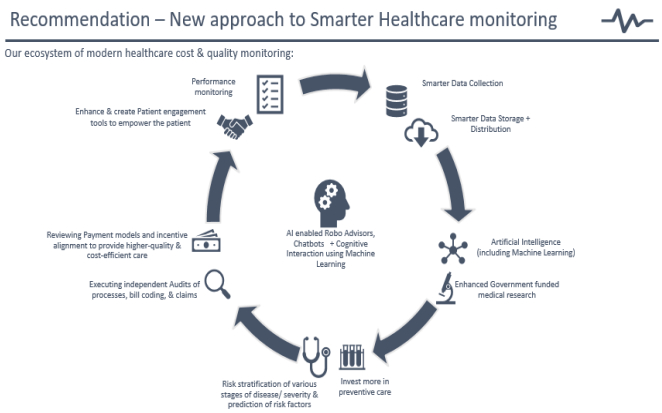




3.4 Interactive R Shiny App of Leaflet Map of U.S. healthcare cost by States



3.5 Recommendations



4 Team Member Contributions

Data set Collection- Khalil Khouy, Ashrita Challa
Visualizations- Ashrita Challa, Sravani Konjeti, Khalil Khouy, Manoj Barman
Project Report- Sravani Konjeti, Khalil Khouy
Presentation- Manoj Barman, Khalil Khouy

5 Conclusion

We would like to conclude our premise, based on our visual scenarios created and analysis, the fact that our new recommended approach to smarter healthcare monitoring should be the ongoing ecosystem of modern healthcare cost and quality monitoring that will help us channel through and focus on the key drivers causing the rising of healthcare costs.

Defensive medicine occurs when doctors order tests, procedures or visits, or avoid high-risk patients or procedures, primarily to reduce their exposure to malpractice liability.

Rises in medical malpractice premiums results in fewer doctors in the workforce, and greater frequency of defensive medicine practices.

We have to concentrate on factors such as, costs of new technologies and medical innovations, drug overprescribing, state-mandated benefits and aging population.

All in all, we have accomplished the tasks that were mentioned in our project proposal. We gave an intro of the current US healthcare economics. We found the insights of underlying factors that influence hospital pricing as mentioned above. For example, the defensive medicine, drug overprescribing etc.

We also did the text analytics on the twitter data on healthcare based on state and displayed in the form of a word cloud. Finally, we came up with some recommendations as discussed above, to minimize profit based unfair practices.

6 References

[1] Josh Cothran, US healthcare spending: who pays, April 2018.
[2] Bradley Sawyer, Selena Gonzales, Kaiser family foundation, how does the quality of US healthcare system compare to other countries, May 2017.
[3] why is healthcare so expensive? why it is so high, how it affects your wallet- and yes what you can do about it, Sept 2014.
[4] Rabah Kamal, Cynthia Cox, how has U.S. spending on healthcare changed over time, Dec 2017.