AIT580 Data Analysis Project

Final Project Report

on

Inpatient Prospective Payment System (IPPS)

Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011

Abstract

The key reason for choosing this dataset for my project is to bring this huge dataset into an understandable format by performing visualization and proving the correlations between multiple attributes by using the statistical models. This dataset has a whopping amount of data that has 163,065 records which can be used in many ways to show different statistics and multiple varieties of visualizations. This project aims to show how the attributes like total discharges from a hospital, total payments made to a health care provider, total medicare payments, and total covered charges by getting it all together into few slides providing a clear picture by making use of the visualization techniques.

Introduction

This data will show the hospital-specific charges for more than 3000 hospitals in the U.S that receive Medicare Inpatient Prospective System (IPPS) Payments charged by the hospitals on patients for the common services used for the fiscal year 2011. We have the providers i.e. hospitals revealing the data of charges for various categories depending on the diagnosis-related groups.

(Inpatient Charge Data FY 2011, n.d.)

For this huge amount of data, based on the DRG (Diagnosis-related groups), we can find out what is the most occurred case i.e. with which issue is most of the cases coming to the health care providers. This can be filtered with the states, regions, and ZIP codes. Apart from this, the data gathered also reveals how much payment has been done by the patients.

This dataset has 12 columns. The column names, their data types and Description are discussed below

1. **DRG Definition**

   Data type: The data type of the column name DRG definition is **Nominal** because each procedure, in this case, is considered as a category and few visualizations are also done by applying the group by function on the DRG definition column.

Description: Code and description identifying the DRG. DRGs are a classification system that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay.

2. **Provider ID**

   Data type: The data type for the column with Provider ID is **Nominal** because these values represent labels that distinguish one from another.

   Description: Provider Identifier billing for inpatient hospital services.

3. **Provider Name**

   Data type: The data type of the column name Provider name is **Nominal** because each health care provider has a unique name and it can't act as a rank or number.

   Description: It is the name of the provider.

4. **Provider Street Address**

   Data type: The data type of the Provider Street Address is **Nominal** because it is unique and is a label for the health care providers.

   Description: It is the street address in which the provider is physically located.

5. **Provider City**

   Data type: The data type of the Provider City is **Nominal** as it is same as the provider street address. The city acts as a categorical data where there can be more than one health care provider in the same city.

   Description: City in which the provider is physically located.

6. **Provider State**

   Data type: The data type of the Provider State is **Nominal** as it is a part of address and even this is a category.

   Description: State in which the provider is physically located.

7. **Provider Zip Code**

   Data type: The data type of the Provider Zip Code is **Nominal** as they are unique, and no city or area can have a same zip code

   Description: Zip code in which the provider is physically located.

8. **Hospital Referral Region Description**

   Data type: The data type of Hospital Referral Region Description is **Nominal** as it is just a name of the region to where the providers are mapped to.

   Description: Hospital Referral Region in which the provider is physically located.

9. **Total Discharges**

   Data type: The data type of the column total discharges is **Ratio** as it is a numeric field where the count of patients will be recorded and can be used for statistics.

   Description: The number of discharges billed by the provider for inpatient hospital services.

10. **Average Covered Charges**

    Data type: The data type of the column Average covered charges is **Ratio** as it is a number which makes sense by having a meaning attached to it.

    Description: The provider's average charge for services covered by Medicare for all discharges in the DRG. These will vary from hospital to hospital because of differences in hospital charge structures.

11. **Average Total Payments**

    Data type: The data type of the column Average Total Payments is **Ratio** as it is a number and both the differences and ratio of two values are meaningful.

    Description: The average of Medicare payments made to the provider for the DRG including the DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Also included are co-payment and deductible amounts that the patient is responsible for

12. **Average Medicare Payments**

    Data type: The data type of the column Average Medicare Payments is also **Ratio**.

(Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011, 2014)

Research questions

1. Which procedure in the Diagnosis related groups is most common?
2. What interpretation can be drawn from the regression models/ diagnostic plots?

Review of Literature

Medicare payment for acute care hospital inpatient stays depends on set rates under Medicare Part A. The framework for payments, known as the Inpatient Prospective Payment System (IPPS), sorts cases into diagnoses-related groups (DRGs) that are then weighted dependent on resources used to treat Medicare beneficiaries in those groups. The Centers for Medicare and Medicaid Services (CMS) refreshes the IPPS guidelines yearly, with remark periods open preceding implementation of the final regulations. Since a high level of surgical care is taken in the inpatient hospital environment, the American College of Surgeons (ACS) has a solid interest in CMS' IPPS and the health care providers' quality improvement efforts addressed in the IPPS rules.

(Inpatient Prospective Payment System Rule, n.d.)

Hospitals decide what they can charge for patients' products and services and those rates are the sum of hospital bills for an item or service. The net amount of the invoice includes the MS-DRG sum, the total amount of the bill per diem, the recipient's primary payer claim amount, the recipient's Part A coinsurance amount, the recipient deductible amount, the recipient blood deductible amount and the DRG outlier amount.

Average costs, overall average expenses, and average Medicare charges at the individual hospital level are evaluated for these DRGs. Users would be able to make comparisons for facilities that could be given in conjunction with a specific hospital stay with the price paid by individual hospitals on local markets and nationally.

(Inpatient Charge Data FY 2011, n.d.)

Methods

The data that has been gathered is already so clean that it didn’t have to go through any data preparation methods where the information will be tested for fraudulency, validated by checking if it is complete and suits the research criteria and no cleaning methodologies were required to clean the data. The data has been still checked for any missing values throughout all the records thoroughly in R language by using any(is.na()) function with which a “false” has been returned which implies that there is no cell that is empty with.

The process of coding is done in R, Python, and SQL which covered the statistical and basic data visualization part and few visualizations have been done on Tableau. All these visualizations created by this data clearly shows which diagnosis-related groups stand in which position, the count of discharges from each healthcare provider, and also the average covered charges, total payments, and the average Medicare payments.

R language

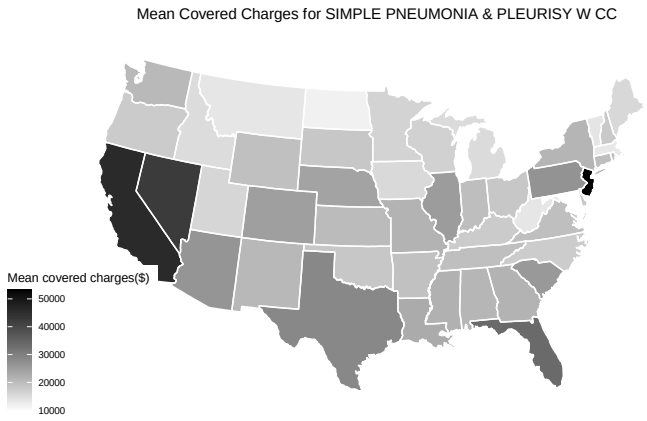
Imported all essential libraries for this project such as tidyverse, lubridate, maps, ggplot2, purr, mapproj, dplyr, stringr, ggthemes, viridisLite, viridis, hrbrthemes, corrplot. The data has been read into R studio as "listings" using the read_csv function and was checked for missing values using any(is.na()) function and no missing values were found. All the columns were left untouched except the column with name "Hospital Referral Region Description" to which a line of code was run to remove the first 5 characters from all the cells in that particular column which would alter the values such as "AL - Dothan" to "Dothan" where "AL" is the state code. This step has been done in order to reduce the complexity as there is already a column for the state with all the state codes.

The column names were changed from DRG Definition, Provider Id, Provider Name, Provider Street Address, Provider City, Provider State, Provider Zip Code, Hospital Referral Region Description, Total Discharges, Average Covered Charges, Average Total Payments, Average Medicare Payments to drg_def, prov_id, prov_name, prov_address, prov_city, prov_state, prov_zip, referral_reg, total_discharges, mean_covered_charges, mean_total_payments, mean_medicare_payments respectively for the coding convenience purposes.

The preloaded data state.abb and state.name in R which has all the state codes and state names have been appended with the state code DC and District of Columbia respectively as these are not included in that preloaded data and a new column with name "Region" where the values were filled by mapping the "prov_state" values with state codes such as AL and TX to the values with state.name and the column values of "Region" will be Alabama and Texas respectively.

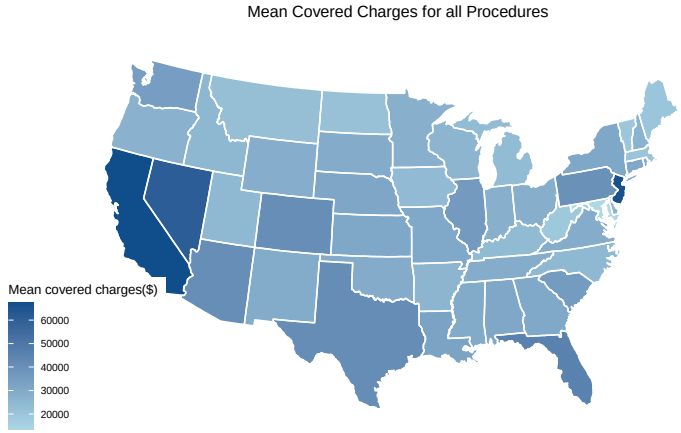
To display the topmost 10 common DRG procedures, the dataset was firstly grouped by the “drg_def” column and a total for each Diagnosis-related group was generated which gathered the data from all the providers for a particular procedure by using the summarize function and the result has been arranged in decreasing order and then the head() function was used to generate the first 10 procedures. It has been found out that the procedure “193 - SIMPLE PNEUMONIA & PLEURISY W MCC” is the first most common procedure that was provided by most of the health care providers.

A choropleth map has been used to represent the information of the mean covered charges of all the hospitals in every state just for the first most common procedure “193 - SIMPLE PNEUMONIA & PLEURISY W MCC”. The output is generated on the United States map with color encoding where the dark gray is the highest mean covered charge over 50000 and lowest being 10000. By seeing the map, it can be concluded that the state of New Jersey stand first where the maximum of mean covered charges have been gathered over $50000 in the entire country where California stands next to and then comes Nevada. A function with name plot_state has been created which will hold the basic data of all the parameters and other functions required to design a map such as colors, longitude, latitude and ggplot, geom_polygon, geom_path, scale_fill_continuous and many others respectively which will act as a package/template that can be readily used to plot the data on the map along with the desired attribute. The function “plot_state” is used to make the below chart.

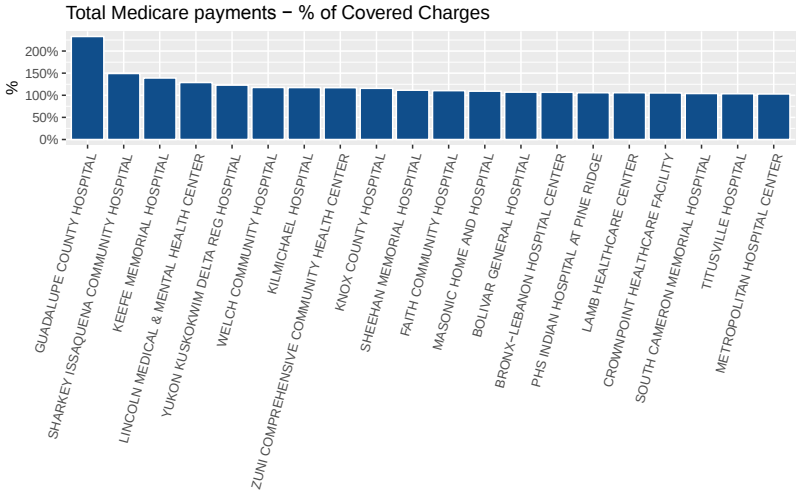


Another Choropleth map has been visualized which represents the mean covered charges by all the hospitals/ providers for all the diagnosis-related groups/ procedures concerning the

states. For this map, unlike the previous ones, no new function was created as it was not that complicated. So, the function group_by is used to group all the records by the newly added region column that has all the state names and is then summarized with the mean of the "mean covered charges" column and the visualization functions such as ggplot, geom_polygon, geom_path and scale_fill_continuous, coord_map and theme_map are used to generate the below map. The interpretation from this visualization would be that the state New Jersey stands first where the maximum of mean covered charges have been gathered over $65000 in the entire country where California stands next to and then comes Nevada.

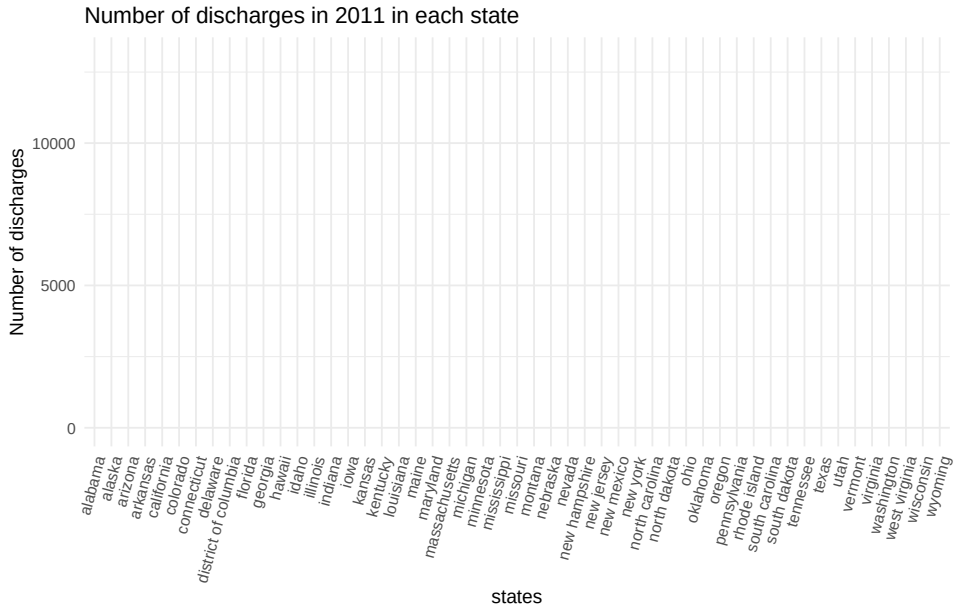


Another visualization in the form of a barplot has been attempted where the information regarding the hospitals/ health care providers which have the highest ratio of medicare payments to the total covered charges. There is math involved in this where a formula is used to calculate how much percentage of the covered charges is covered by the medicare payments.



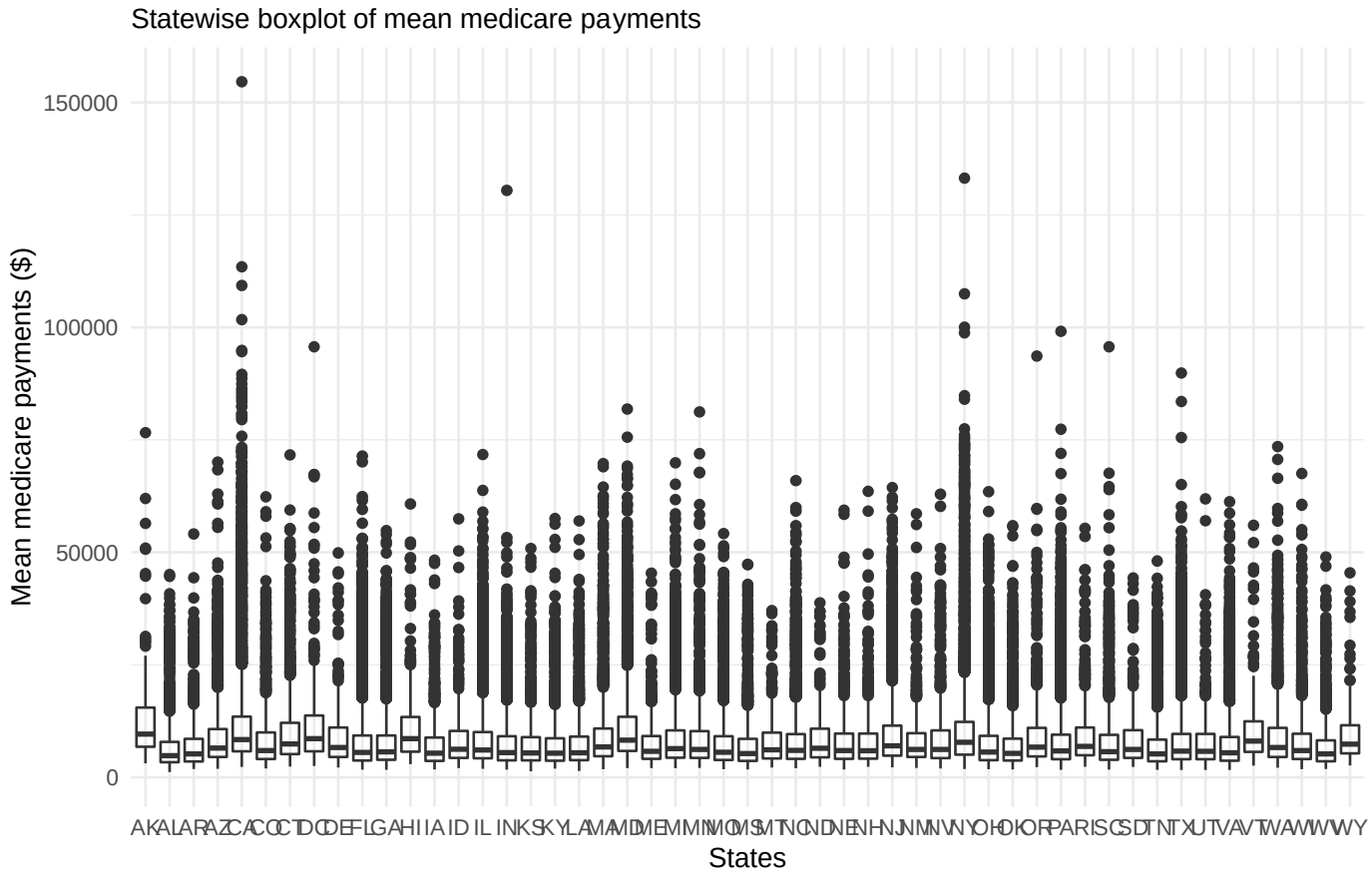
Another bar plot that represents the total number of discharges in the fiscal year of 2011 from each state is visualized in which the group_by function was applied on the region column and the total discharges have been summarized according to the states and it is concerted to the data frame format so to make it easy to display as a table. A sample of random 10 states and their discharge count has been displayed by using the head() function.

The functions ggplot, geom_bar are used to plot this data and the result is as follows where on the X-axis, the state names are placed and on the Y-axis, the discharge information has been placed. It can be concluded that the California state has the highest number of discharges followed by Texas and then Florida.



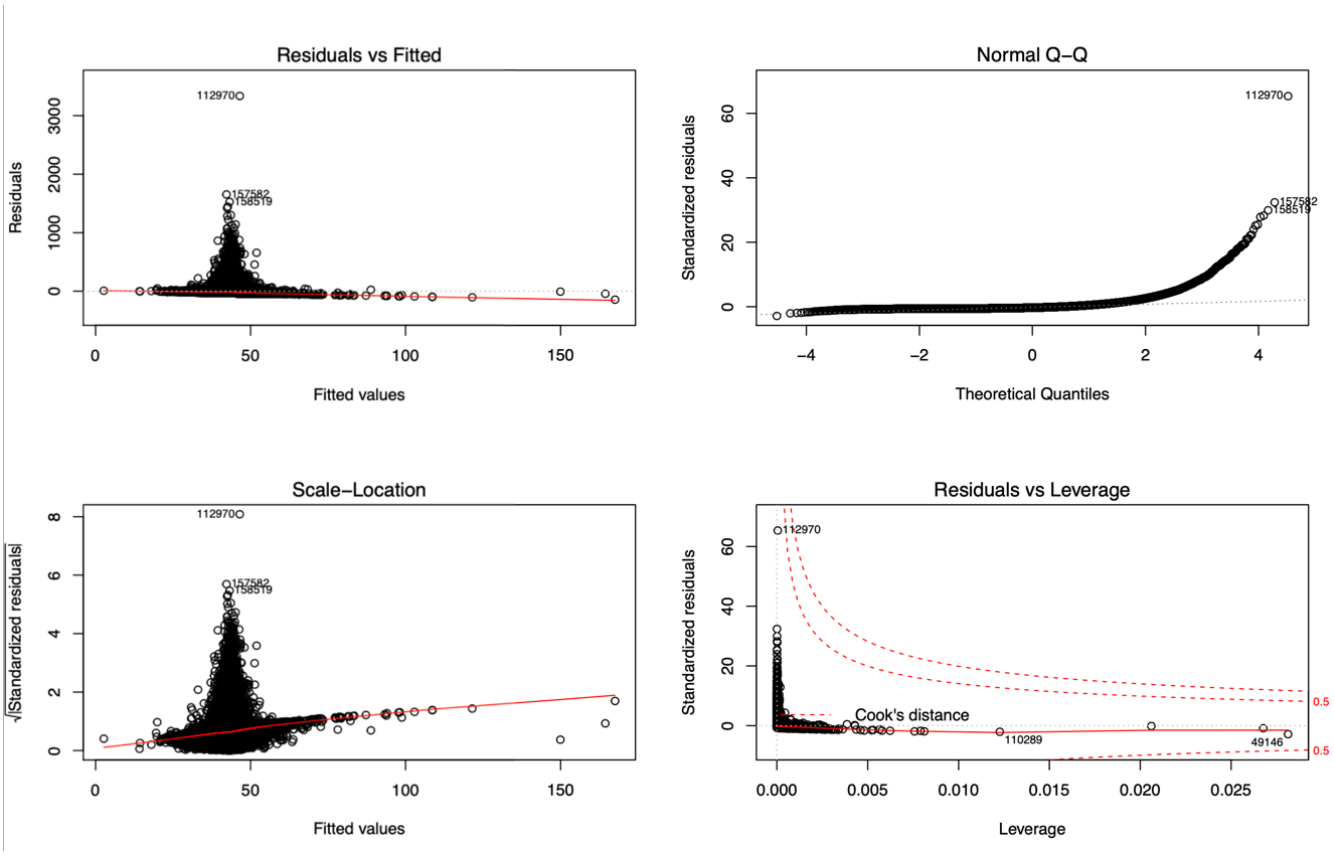
A box plot has been generated between the provider states and mean Medicare payments. The mean, median, and quartile values are affected by the high-end values which are supposed to be considered as outliers, but we cannot eliminate them as the other result interpretations of the visualizations will be affected. If we closely observe the box plot below, the California state has an outlier value of mean medicare payments of above 150000 and most of the rest values are very low and that is how the outliers are impacting the mean, median, and quartile values. The same is the case with other states.The box plot has been plotted by using the

functions ggplot, geom_boxplot, scale_fill_viridis, theme_minimal, theme functions, and ggtitle function is used to labels the names on the graphs.



A linear model has been created with a formula that includes only ratio data types and the formula framed will be used to plot the diagnostic plots and summary. These diagnostic plots include Residuals vs fitted, normal Q-Q plot, Scale-Location plot, and the Residuals vs leverage plot. From these plots, we can conclude that the model that has been created is working well. There are no outliers that are detected. The code for assigning the model to a variable and plotting the diagnostic plots is as follows:

```
mod=lm(listings$total_discharges~listings$mean_covered_charges+listings$mean_total_payments+listings$mean_medicare_payments)
par(mfrow=c(2,2))
plot(mod)
```



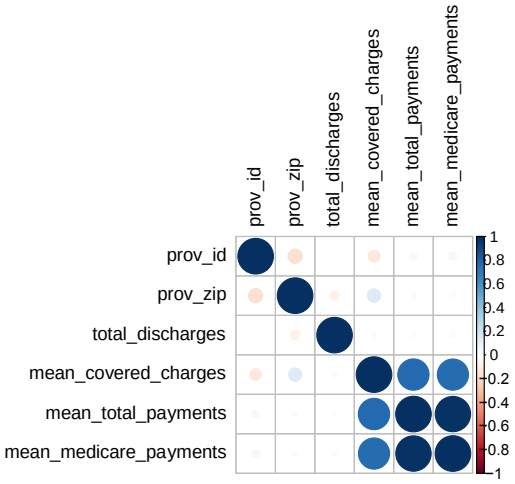
summary(mod)

```
Call:
lm(formula = listings$total_discharges ~ listings$mean_covered_charges +
    listings$mean_total_payments + listings$mean_medicare_payments)

Residuals:
   Min     1Q Median     3Q    Max
-144.5  -25.8  -15.6    6.3 3336.5

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    4.268e+01  2.192e-01 194.719  < 2e-16 ***
listings$mean_covered_charges -2.759e-05  5.699e-06  -4.842 1.29e-06 ***
listings$mean_total_payments   1.672e-03  1.146e-04  14.587  < 2e-16 ***
listings$mean_medicare_payments -1.781e-03  1.190e-04 -14.971  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 51.06 on 163061 degrees of freedom
Multiple R-squared:  0.001803,  Adjusted R-squared:  0.001785
F-statistic: 98.19 on 3 and 163061 DF,  p-value: < 2.2e-16
```

Correlation matrix and plot

The correlation matrix has been generated for all the attributes that have the number datatype and the output is displayed below. Ignoring the prov_id and prov_zip attributes, it has been noticed that the mean medicare payments and mean total payments are highly correlated. The same has been graphically represented below the correlation matrix.

```
> cor(cordata)
                          prov_id    prov_zip total_discharges mean_covered_charges mean_total_payments mean_medicare_payments
prov_id                1.000000000 -0.16023696      0.000455668          -0.11248951         -0.04005931            -0.04484641
prov_zip              -0.160236964  1.00000000     -0.071803522           0.14065846          0.02749599             0.02226216
total_discharges       0.000455668 -0.07180352      1.000000000          -0.02076584         -0.01602010            -0.02129904
mean_covered_charges  -0.112489510  0.14065846     -0.020765842           1.00000000          0.77411184             0.76892679
mean_total_payments   -0.040059305  0.02749599     -0.016020105           0.77411184          1.00000000             0.98936198
mean_medicare_payments -0.044846414 0.02226216     -0.021299043           0.76892679          0.98936198             1.00000000
```



Hypothesis test

mean(listings$total_discharges)

```
[1] 42.7763
```

sd(listings$total_discharges)

```
[1] 51.10404
```

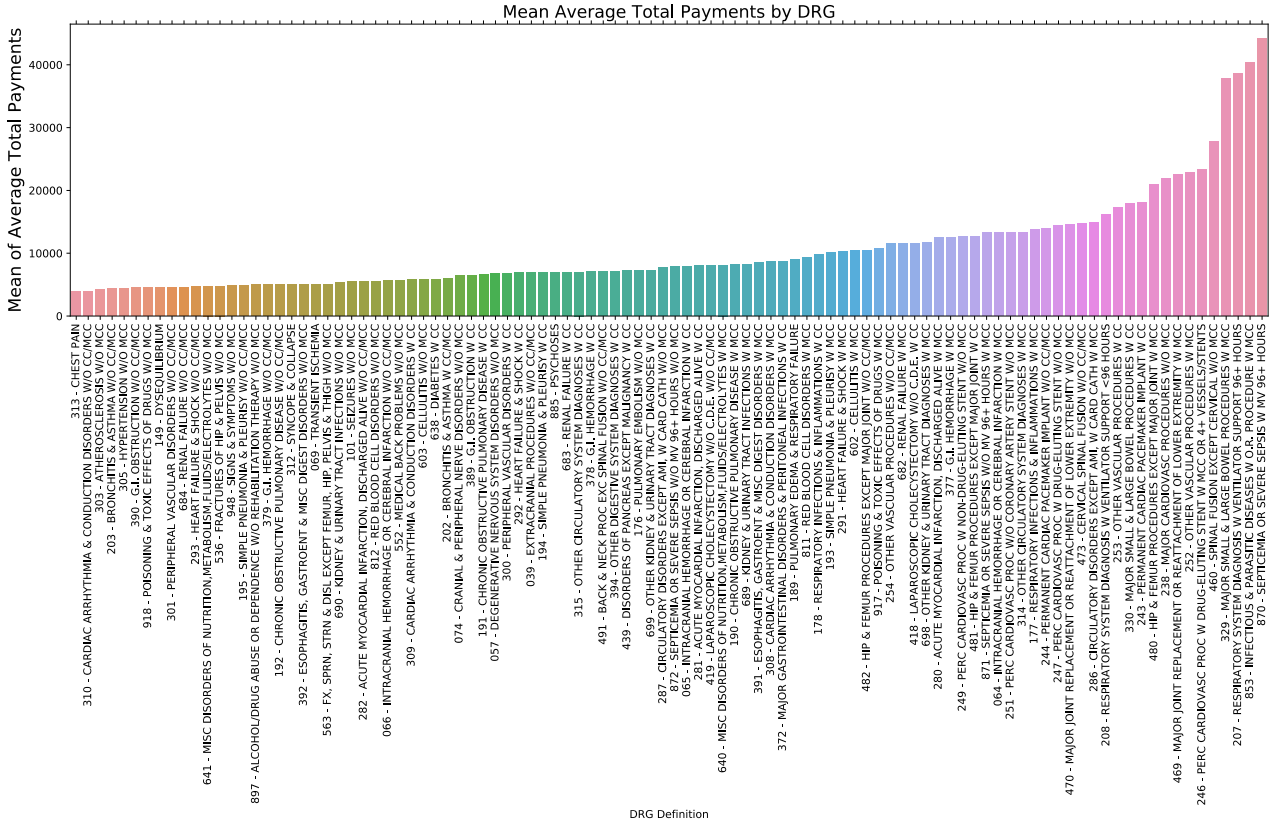
t.test(listings$total_discharges,mu = 0.5)

```
        One Sample t-test

data:  listings$total_discharges
t = 334.06, df = 163064, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0.5
95 percent confidence interval:
 42.52826 43.02435
sample estimates:
mean of x
  42.7763
```

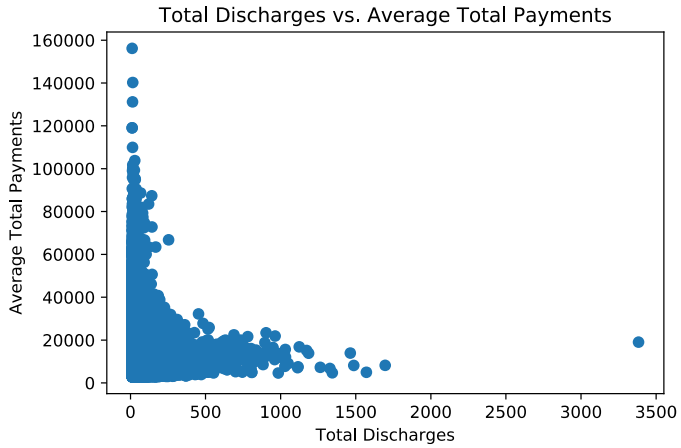
Python

The packages which are needed such as pandas, seaborn,matplot.pyplot and warnings have been imported and then the .csv file was read into spyder .

Firstly, the data loaded was tuned to visualize the table format of the DRG procedures, and the average total payments made to each procedure. Then a function with name plt_setup that has parameters related to labels for the bar plot is setup. The previously displayed data that is in the table format is now displayed in the form of a bar plot where on the X-axis, all the DRG procedures and on Y-axis the average total payments made are placed. The interpretation that can be drawn from this plot could be that the procedure with the name “853 - INFECTIOUS & PARASITIC DISEASES W O.R. PROCEDURE W MCC” has received the highest amount as the total payments.

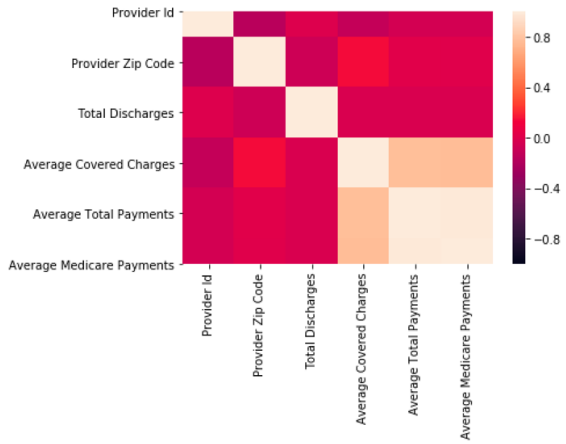


A simple scatter plot has been plotted between the columns Total discharges and Average Total Payments by using the plt.scatter() function and the scatter plot is as follows.



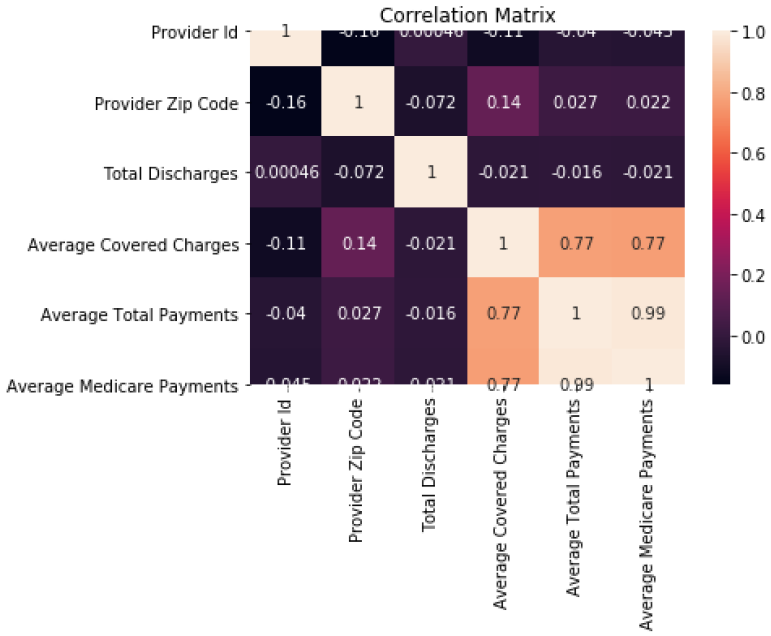
A correlation matrix has been generated using the corr() function which showed the correlation between all the numeric fields and it has been identified that the columns Average total payments and average Medicare payments have the highest correlation with a value of 0.989362. The next aim was to create a correlation plot that visualizes the same data. This correlation function has been assigned to a variable and a heat map is plotted using the heatmap function and the output is as follows

```
In [11]: df.corr()
Out[11]:
                           Provider Id  ...  Average Medicare Payments
Provider Id                   1.000000  ...                  -0.044846
Provider Zip Code            -0.160237  ...                   0.022262
Total Discharges              0.000456  ...                  -0.021299
Average Covered Charges      -0.112490  ...                   0.768927
Average Total Payments       -0.040059  ...                   0.989362
Average Medicare Payments    -0.044846  ...                   1.000000

[6 rows x 6 columns]
```



Displaying a correlation plot along with the correlation value

The below plot will visualize the correlation matrix in the form of a correlation plot which will also display the correlation values.



SQL

A table has been created with name “inpatient” in SQL developer. The table structure is as follows.

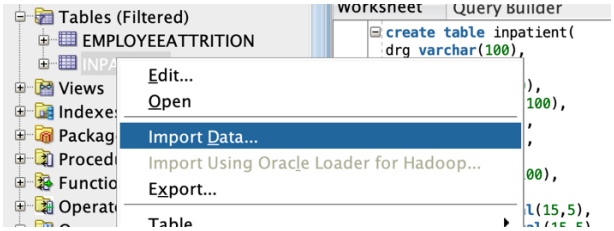
Welcome Page | chai | *Sadasivuni_Final project SQL.sql* | INPATIENT

Columns | Data | Model | Constraints | Grants | Statistics | Triggers | Flashback | Dependencies | Details | Partitio

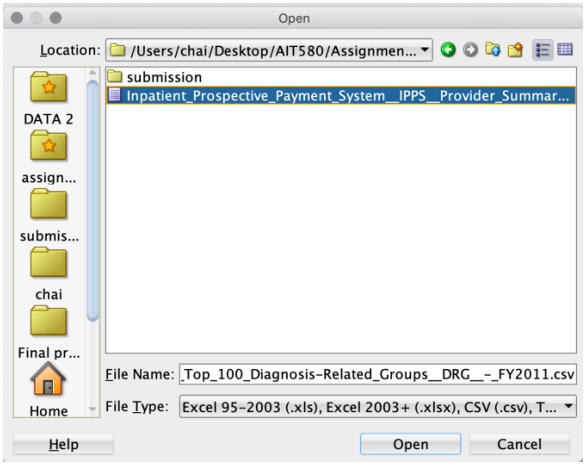
Actions...

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	DRG	VARCHAR2(100 BYTE)	Yes	(null)	1	(null)
2	PID	NUMBER(38,0)	Yes	(null)	2	(null)
3	PNAME	VARCHAR2(100 BYTE)	Yes	(null)	3	(null)
4	PADDRESS	VARCHAR2(100 BYTE)	Yes	(null)	4	(null)
5	PCITY	VARCHAR2(50 BYTE)	Yes	(null)	5	(null)
6	PSTATE	VARCHAR2(5 BYTE)	Yes	(null)	6	(null)
7	PZIP	NUMBER(38,0)	Yes	(null)	7	(null)
8	PREGION	VARCHAR2(100 BYTE)	Yes	(null)	8	(null)
9	DISCHARGES	NUMBER(38,0)	Yes	(null)	9	(null)
10	AVGCHARGES	NUMBER(15,5)	Yes	(null)	10	(null)
11	AVGPAYMENTS	NUMBER(15,5)	Yes	(null)	11	(null)
12	AVGMEDICAREPAYMENTS	NUMBER(15,5)	Yes	(null)	12	(null)

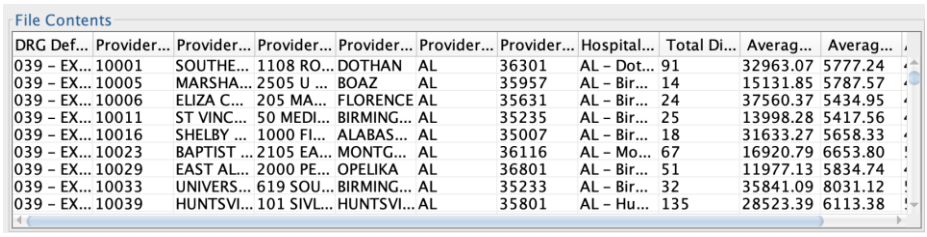
Then the 163065 records of data are imported into the table created.



Selection of .csv file

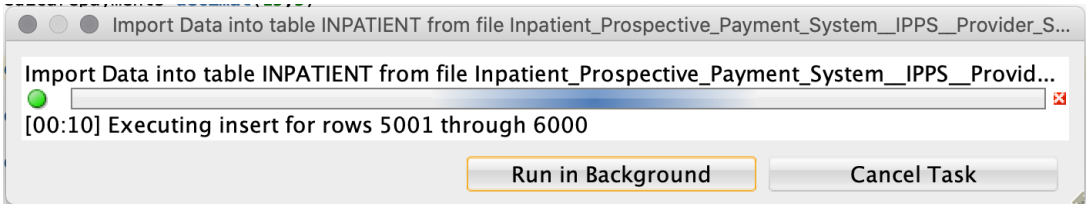


The file contents that will be imported

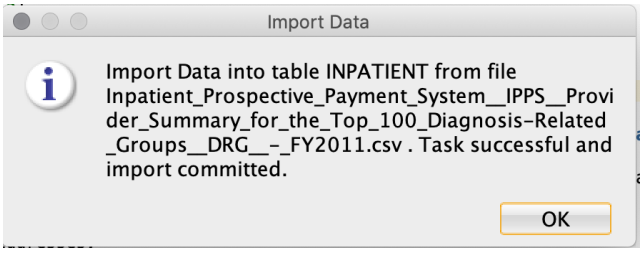


DRG Def...	Provider...	Provider...	Provider...	Provider...	Provider...	Provider...	Hospital...	Total Di...	Averag...	Averag...
039 - EX...	10001	SOUTHE...	1108 RO...	DOTHAN	AL	36301	AL - Dot...	91	32963.07	5777.24
039 - EX...	10005	MARSHA...	2505 U ...	BOAZ	AL	35957	AL - Bir...	14	15131.85	5787.57
039 - EX...	10006	ELIZA C...	205 MA...	FLORENCE	AL	35631	AL - Bir...	24	37560.37	5434.95
039 - EX...	10011	ST VINC...	50 MEDI...	BIRMING...	AL	35235	AL - Bir...	25	13998.28	5417.56
039 - EX...	10016	SHELBY ...	1000 FI...	ALABAS...	AL	35007	AL - Bir...	18	31633.27	5658.33
039 - EX...	10023	BAPTIST ...	2105 EA...	MONTG...	AL	36116	AL - Mo...	67	16920.79	6653.80
039 - EX...	10029	EAST AL...	2000 PE...	OPELIKA	AL	36801	AL - Bir...	51	11977.13	5834.74
039 - EX...	10033	UNIVERS...	619 SOU...	BIRMING...	AL	35233	AL - Bir...	32	35841.09	8031.12
039 - EX...	10039	HUNTSVI...	101 SIVL...	HUNTSVI...	AL	35801	AL - Hu...	135	28523.39	6113.38

Data import in progress

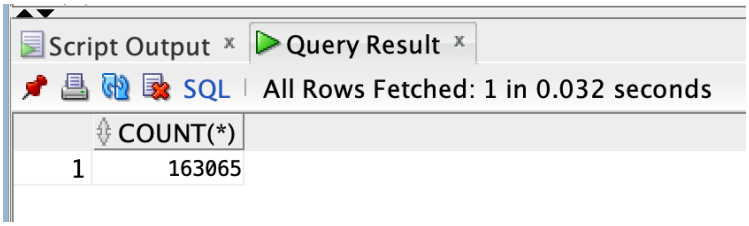


Data import into table successful



The below command counts the number of records that are present in the dataset

select count(*) from inpatient;



The below command gives the sum of all the discharges in each state with the state code for all the Diagnosis-related groups together and orders in from lower sum to higher.

select sum(discharges),pstate from inpatient group by pstate order by sum(discharges) asc;

	SUM(DISCHARGES)	PSTATE
1	6142	AK
2	6535	WY
3	10071	VT
4	11712	HI
5	15705	MT
6	16425	ND
7	18295	ID

The below command groups all the Diagnosis-related groups and takes the average of all the average total payments for each Diagnosis-related group from all the providers and rounds in off and displayed them in ascending order.

select drg,round(avg(avgpayments)) from inpatient group by drg order by round(avg(avgpayments)) asc;

	DRG	ROUND(AVG(AVGPAYMENTS))
1	313 - CHEST PAIN	3912
2	310 - CARDIAC ARRHYTHMIA & CONDUCTION DISORDERS W/O CC/MCC	3967
3	303 - ATHEROSCLEROSIS W/O MCC	4156
4	203 - BRONCHITIS & ASTHMA W/O CC/MCC	4390
5	305 - HYPERTENSION W/O MCC	4403
6	390 - G.I. OBSTRUCTION W/O CC/MCC	4480
7	918 - POISONING & TOXIC EFFECTS OF DRUGS W/O MCC	4485
8	149 - DYSEQUILIBRIUM	4589
9	301 - PERIPHERAL VASCULAR DISORDERS W/O CC/MCC	4600
10	684 - RENAL FAILURE W/O CC/MCC	4618

Using the below command, a view can be generated with the name “phoenix_addresses” which displays the Diagnosis-related groups with the provider name, address, zip code, and region of the records which are having Arizona as the state and Phoenix as the city.

create view phoenix_addresses as select drg,pname,paddress,pzip,pregion from inpatient where pstate='AZ' AND pcity='PHOENIX';

The below command will display all the records from the view that has been created above.

select * from phoenix_addresses;

	DRG	PNAME	PADDRESS	PZIP	PREGION
1	039 – EXTRACRANIAL PROCEDURES W/O CC/MCC	BANNER GOOD SAMARITAN MEDICAL CENTER	1111 EAST MCDOWELL ROAD	85006	AZ – Phoenix
2	039 – EXTRACRANIAL PROCEDURES W/O CC/MCC	ST JOSEPH'S HOSPITAL AND MEDICAL CENTER	350 WEST THOMAS ROAD	85013	AZ – Phoenix
3	039 – EXTRACRANIAL PROCEDURES W/O CC/MCC	PHOENIX BAPTIST HOSPITAL	2000 WEST BETHANY HOME ROAD	85015	AZ – Phoenix
4	039 – EXTRACRANIAL PROCEDURES W/O CC/MCC	JOHN C LINCOLN DEER VALLEY HOSPITAL	19829 NORTH 27TH AVENUE	85027	AZ – Phoenix
5	039 – EXTRACRANIAL PROCEDURES W/O CC/MCC	MAYO CLINIC HOSPITAL	5777 EAST MAYO BOULEVARD	85054	AZ – Phoenix
6	057 – DEGENERATIVE NERVOUS SYSTEM DISORDERS W/O MCC	BANNER GOOD SAMARITAN MEDICAL CENTER	1111 EAST MCDOWELL ROAD	85006	AZ – Phoenix
7	057 – DEGENERATIVE NERVOUS SYSTEM DISORDERS W/O MCC	JOHN C LINCOLN NORTH MOUNTAIN HOSPITAL	250 EAST DUNLAP AVENUE	85020	AZ – Phoenix

The below command will generate the maximum count of discharges taken place in any hospital in all the states.

select max(discharges) from inpatient;

	MAX(DISCHARGES)
1	3383

The below command will display the names of all the diagnosis-related groups from the dataset where the number of discharges is greater than 1000.

select drg from inpatient where discharges>1000 group by drg;

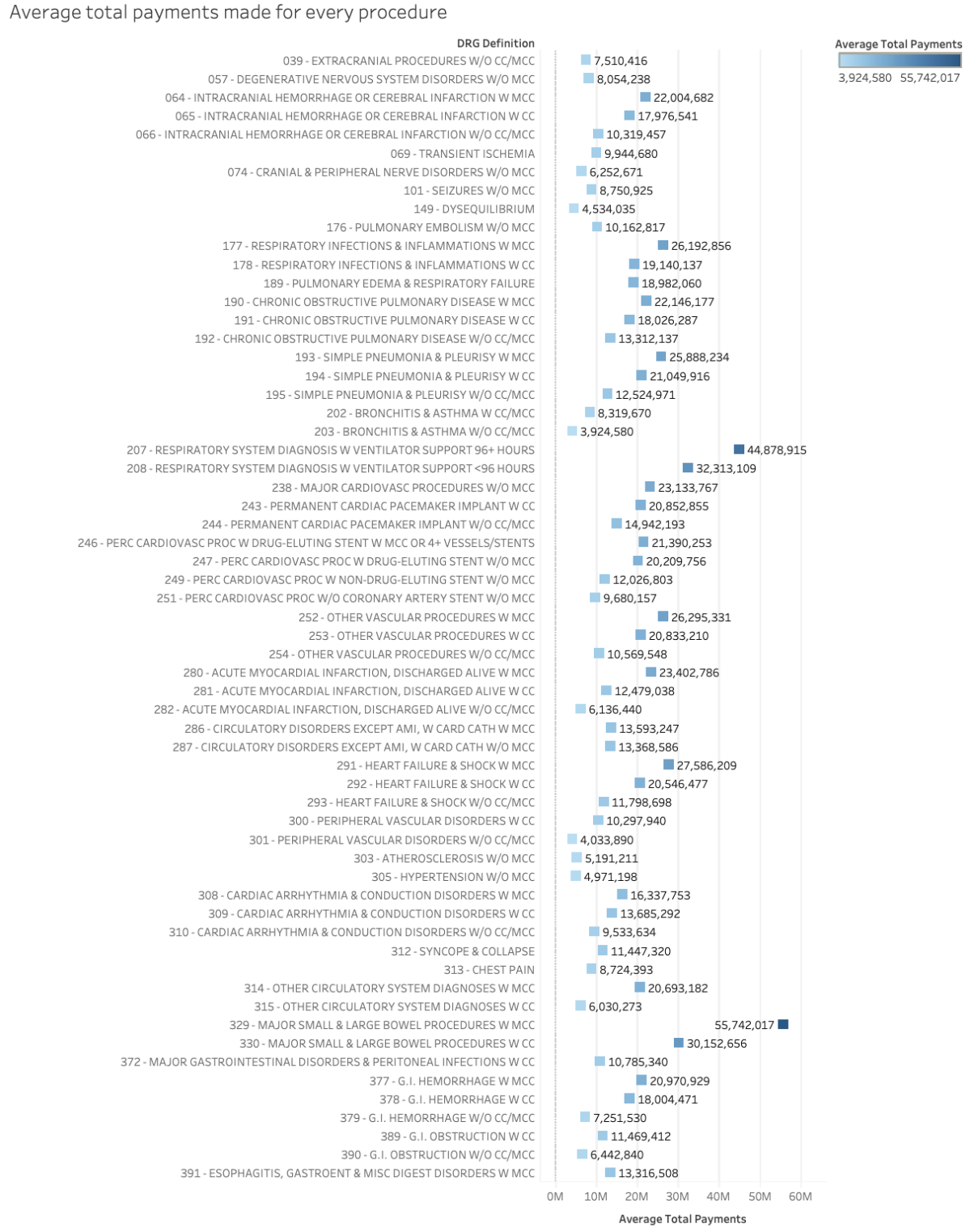
	DRG
1	897 – ALCOHOL/DRUG ABUSE OR DEPENDENCE W/O REHABILITATION THERAPY W/O MCC
2	885 – PSYCHOSES
3	392 – ESOPHAGITIS, GASTROENT & MISC DIGEST DISORDERS W/O MCC
4	470 – MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC

Tableau

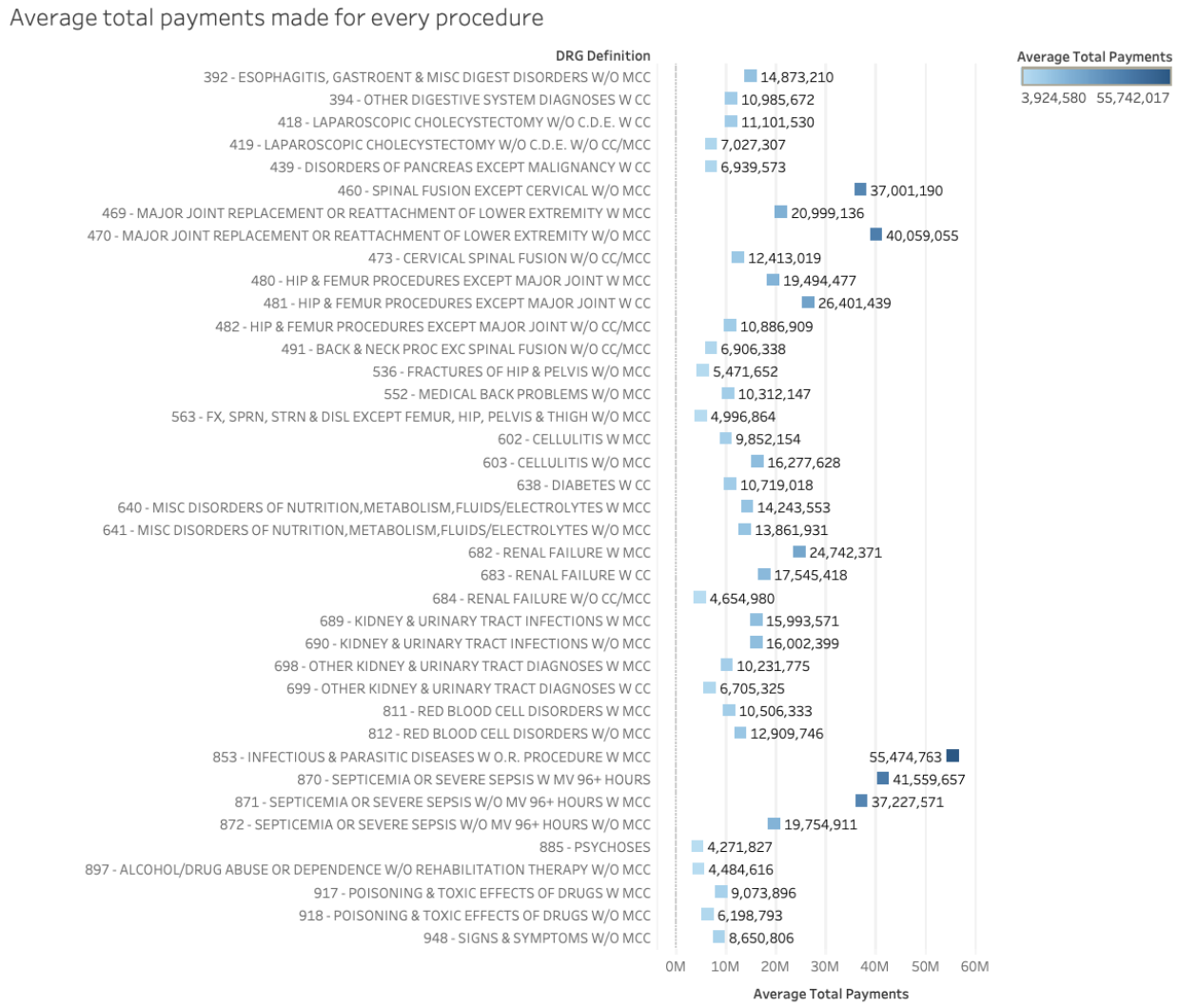
Tableau desktop has been used to create few basic visualizations. Tableau has a mapping functionality and is able to plot latitude and longitude coordinates and connect to spatial files to display custom geography. There are three output sheets generated from tableau visualization tool which are representing Average total payments made for every procedure, Total discharges from each state, Average Medicare Payments, and Total discharges state-wise.

(Tableau Software, n.d.)

Average total payments made for every procedure



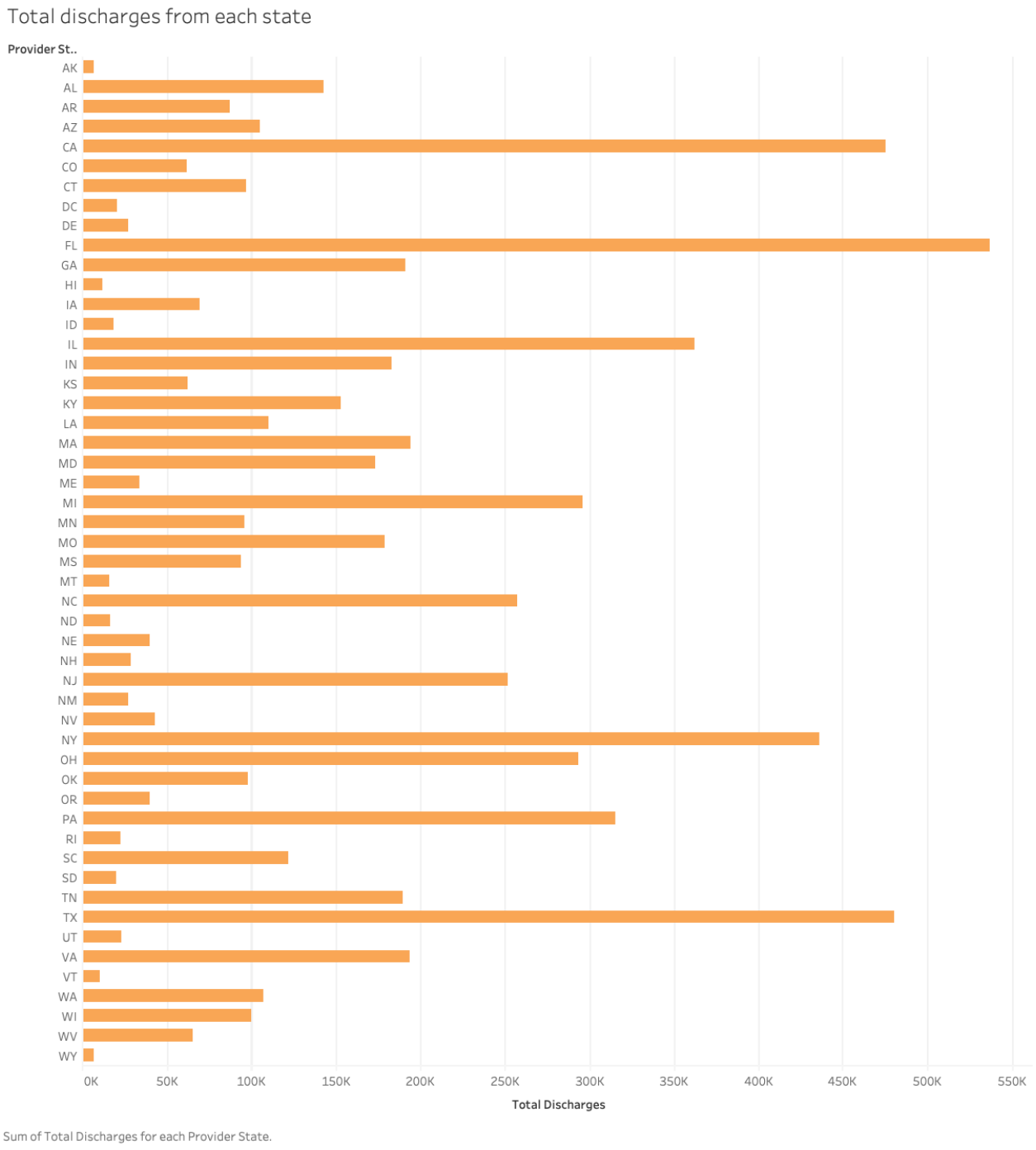
Sum of Average Total Payments for each DRG Definition. Color shows sum of Average Total Payments. The marks are labeled by sum of Average Total Payments.



In this visualization, the Average total payments made to each of the Diagnosis-related groups and the color-coding given to each value can be read and understood from the legend on the top right to know at what level that particular DRG procedure stands.

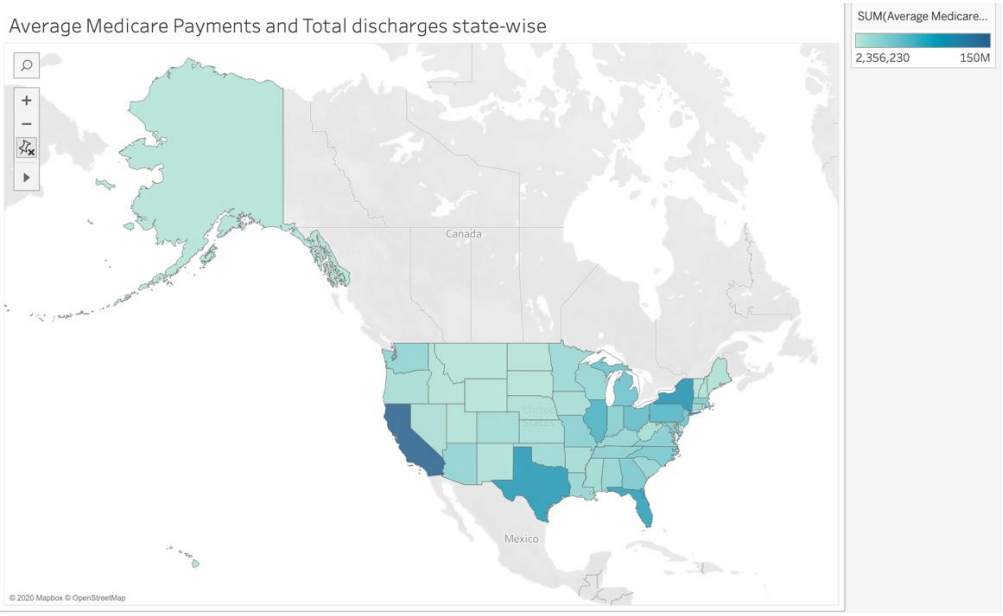
It has been visualized by placing the DRG Definition column on the rows field and sum of average total payments on the columns field. From this visual, it can be interpreted that the highest payments are made for the “329- major small and large bowel procedures W MCC” with $55,742,017.

Total discharges from each state

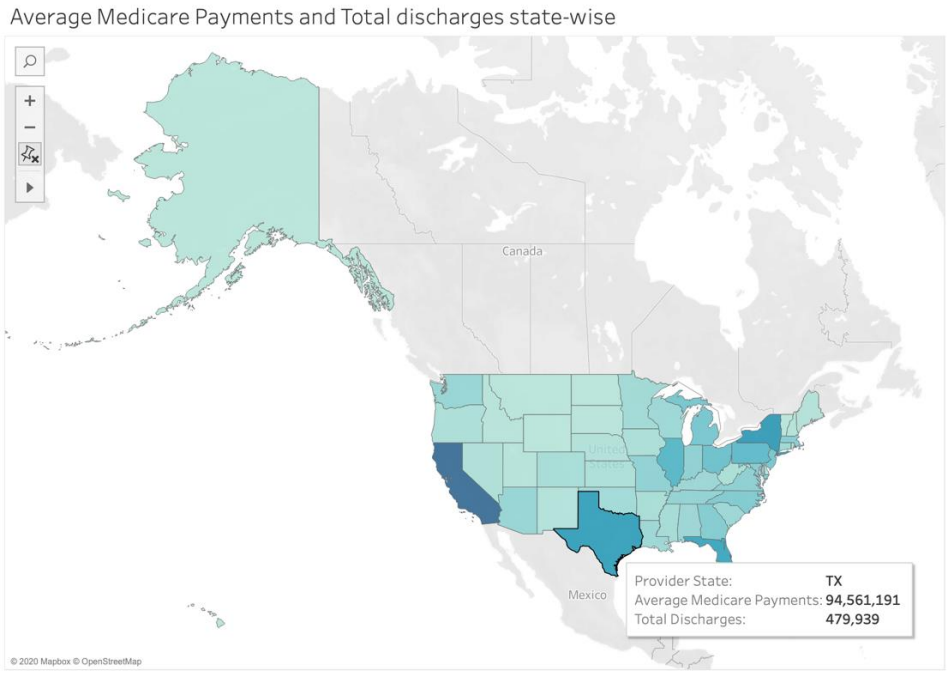


In this visualization, we can see that all the state codes have been displayed on x axis and total discharges on the y axis which are visualizing the total discharges for all procedures from each state. It can be clearly interpreted that maximum number of discharges are from Florida state which have crossed 500,000 and almost was nearing 550,000.

Average Medicare Payments and Total discharges state-wise



This is an interactive map which is representing the sum of average Medicare payments made to the provider by each hospital in every state. The color density attached to the states indicates the maximum to a minimum sum of the amount paid to the providers. It can be matched and checked with the legend on the top right. This map has turned out to be interactive whereby hovering over the state, the state code, total discharges, and the sum of average medicare payments of that particular state will be displayed.



Results and Conclusion

From the analysis we performed, first, we got to find out that at which level a procedure from Diagnosis-related groups stands, what is the average of the covered charges by each provider state-wise, percentage of medicare payments covered, an average of the covered charges by each provider state-wise on the most performed procedure. The visualization of the complicated data has become very simple and the results can be easily interpreted from all kinds of plots (choropleth maps, bar plots, boxplot, diagnostic plots, correlation plots, scatter plots). This analysis that has been performed answers all the research questions.

References

Inpatient Charge Data FY 2011. (n.d.). Retrieved from cms.gov: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient2011

Inpatient Charge Data FY 2011. (n.d.). Retrieved from cms.gov: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient2011

Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011. (2014, June 02). Retrieved from data.cms.gov: https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3

Inpatient Prospective Payment System Rule. (n.d.). Retrieved from facs.org | American college of surgeons: https://www.facs.org/advocacy/regulatory/medicare-payment/ipps

Tableau Software. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Tableau_Software#Software_products