

Centro Universitário de Brasília - CEUB

R introduction for Data Science

Raul Lima Vieira

Database Description and Characterization Medicare
Inpatient Hospitals - by Provider and Service



Brasília

2023

1. Background

The Centers for Medicare & Medicaid Services (CMS) has prepared a public dataset, the Medicare Inpatient Hospitals, with information on services and procedures provided to Medicare beneficiaries by hospital facilities. The Inpatient dataset contains hospital-specific charges for the more than 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG).

2. Data Contents

The original database, free for download on the internet, has a larger number of variables. For this analysis, only a few variables of interest were determined among them (Providers code, FIPS and ZIP codes, street addresses and RUCA codes were removed). The official data dictionary is also available from the source. For the variables used in this analysis, a brief description follows:

Medicare Inpatient Hospitals - by Provider and Service		
Variable Name	Term Name	Definition
Provider_Name	Provider Name	The name of the provider
Provider_City	Provider City	The city in which the provider is physically located
Provider_State	Provider State	The state abbreviation for the state in which the provider is physically located.
Diagnosis_Definition	Diagnosis Related Group (DRG) Definition	Classification system that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay.
Diagnosis_Description	Diagnosis Related Group (DRG) Description	Description of the classification system (DRG) that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay.
Total_Discharges	Total Discharges	The number of discharges billed by all providers for inpatient hospital services
Avg_Covered_Charges	Average Covered Charges	The average charge of all providers' services covered by Medicare for discharges in the DRG. These will vary from hospital to hospital because of differences in hospital charge structures
Avg_Total_Payments	Average Total Payments	The average total payments to all providers for the DRG including the MS-DRG 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.
Avg_Medicare_Payments	Average Medicare Payments	The average amount that Medicare pays to the provider for Medicare's share of the MS-DRG. 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.

3. Characterization of the Database

The database, already filtered as explained above, consists of a .csv file, containing 151,989 lines of records in 9 columns of variables. 4 of these variables are of type 'numeric' (Discharges, Covered Charges, Total Payments, Medicare Payments), the rest are of type 'character'.

There are 51 different states (all from the USA) registered, as well as 1853 cities, and 538 different types of diagnosis/services. Here is a summary of all variables from this dataframe:

```
Provider_Name      Provider_City      Provider_State      Diagnosis_Definition
Length:151989      Length:151989      Length:151989      Length:15199
Class :character    Class :character    Class :character    Class :character
Mode :character     Mode :character     Mode :character     Mode :character

Diagnosis_Description  Total_Discharges  Avg_Covered_Charges  Avg_Total_Payments
Length:151989          Min. : 11.00        Min. : 3341          Min. : 2384
```

```

Class :character      1st Qu.: 14.00    1st Qu.: 32085    1st Qu.: 8218
Mode  :character      Median : 20.00    Median : 53338    Median : 12327
                        Mean  : 34.56    Mean  : 83116    Mean  : 18183
                        3rd Qu.: 35.00    3rd Qu.: 94661    3rd Qu.: 20002
                        Max.   :2650.00    Max.   :3697459    Max.   :632473

Avg_Medicare_Payments
Min.   : 132.9
1st Qu.: 6500.6
Median : 10196.0
Mean   : 15282.8
3rd Qu.: 16732.2
Max.   :610619.9

```

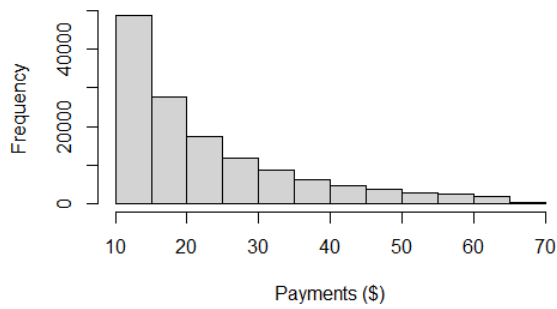
Some interesting insights about this data:

- LIVER TRANSPLANT WITH MCC OR INTESTINAL TRANSPLANT is the most expensive service registered, charged in \$ 3697459,00;
- ALCOHOL, DRUG ABUSE OR DEPENDENCE, LEFT AMA is the cheapest service registered, charged in \$ 3340,84;
- Grouping the data by Region, South has more number of discharges than the others, and West has less;
- South region also has more providers, Northeast has less;
- For higher costs of providers services, less discharges are registered. Cheaper treatments are more common.
- There is a high positive correlation (+0.81) between charges and medicare payments, it means that for higher charges, higher are medicare payments;
- There is a very high correlation between total and medicare payments (+0.98), as expected;
- In northeast, New York is the city with more providers, 1383 registered;
- In south, Houston is the city with more providers, 1039 registered;
- In midwest, Chicago is the city with more providers, 1213 registered;
- In west, Los Angeles is the city with more providers, 902 registered;
- RESPIRATORY INFECTIONS AND INFLAMMATIONS WITH MCC is the most provided service, by 2830 hospitals;
- Total number of Hospitals registered in this base: 2979;
- There was a 92.29 billion dollars total receipt from all hospitals;
- South has the major receipt (over 30 billion dollars), while west has the lowest (over 17 billion dollars)

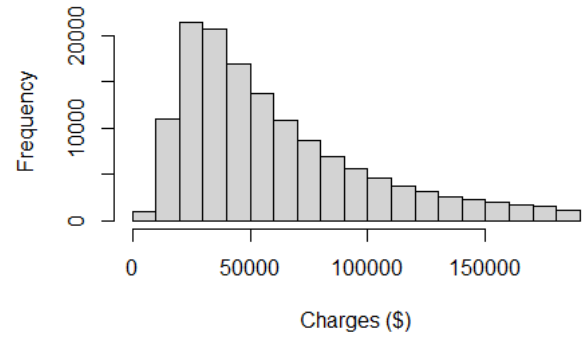
For the next topic, in order to have a better data visualization, axes were adjusted to certain values, excluding outliers from the view. The main purpose is to understand data in its general and most frequent values.

4. Data Visualization

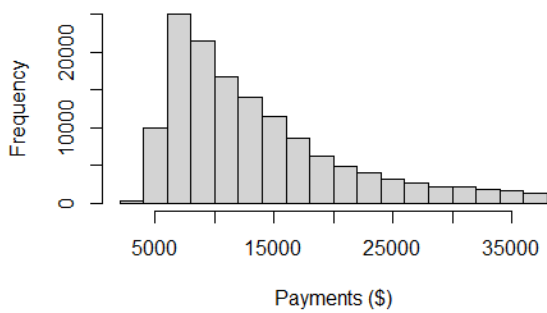
Total Discharges w/o Outliers



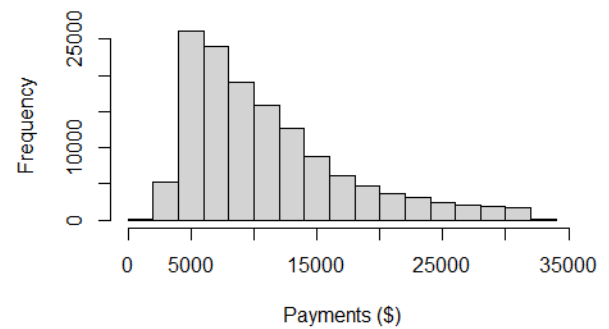
Average Covered Charges w/o Outliers



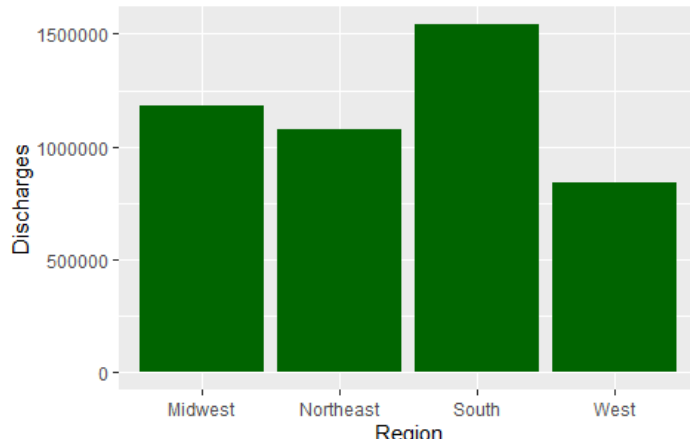
Average Total Payments w/o Outliers



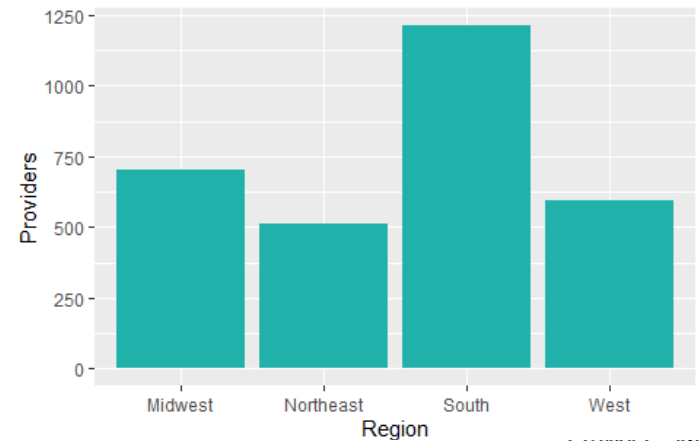
Average Medicare Payments w/o Outliers



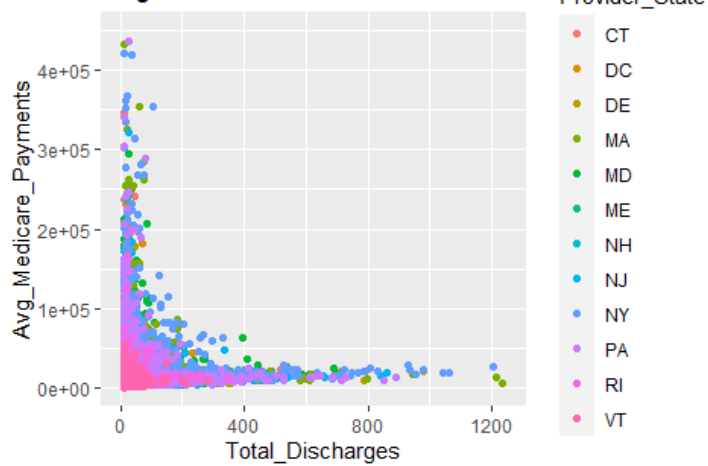
Discharges per Region



Providers per Region



Region: Northeast



Region: South

