

Final Project on Medicare Inpatient Prospective Payment System

OBJECTIVES

- 1) Introduction
- 2) Tidying the data
- 3) Analysing the data
- 4) Conclusion

Introduction

The dataset includes 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. Medicare is for old and disabled people.

About the data set

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 MS-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 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.

TIDYING THE DATA

The dollar symbol has been removed from Average.Covered.Charges, Average.Medicare.Payments and Average.Total.Payments. Since the label of i.DRG.Definition is long, it is hard to display them in plot. Hence the numericals are extracted from the column and then a prefix has been concatenated to it since it is a categorical variable else R Studio considers it as a continuous variable. Then later I have renamed the column names.

```
Inpatient <- read.csv("C:/Users/praneetha/Desktop/Inpatient.csv", sep=";", header=TRUE)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.3.3
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Warning: package 'tibble' was built under R version 3.3.3
## Warning: package 'readr' was built under R version 3.3.3
## Warning: package 'dplyr' was built under R version 3.3.3
```

```

## Conflicts with tidy packages -----
## filter(): dplyr, stats
## lag(): dplyr, stats
library(data.table)

## -----
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
## -----

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
## between, first, last

## The following object is masked from 'package:purrr':
##
## transpose
library(forcats)
## To Remove The Dollar Symbol for Average.Covered.Charges,
##Average.Medicare.Payments and Average.Total.Payments
Inpatient$Average.Covered.Charges <-
  as.numeric(gsub("[\\$,]", "", Inpatient$Average.Covered.Charges))
Inpatient$Average.Medicare.Payments <-
  as.numeric(gsub("[\\$,]", "", Inpatient$Average.Medicare.Payments))
Inpatient$Average.Total.Payments <-
  as.numeric(gsub("[\\$,]", "", Inpatient$Average.Total.Payments))

## To extract the numericals from i..DRG.Definition
Inpatient$i..DRG.Definition<-
  as.numeric(gsub("\\D", "", Inpatient$i..DRG.Definition))

## To add a prefix to i..DRG.Definition
Inpatient$i..DRG.Definition <-
  sub("^", "D", Inpatient$i..DRG.Definition)

## To Rename the Column Names
setnames(Inpatient, old=c("i..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"),
  new=c("DRG_Definition", "Provider_Id", "Provider_Name", "Provider_Street_Address",
  "Provider_City", "Provider_State", "Provider_Zip_Code", "Hospital_Ref_Reg.Desc",
  "Total_Discharges", "Avg_Covered_Charges", "Avg_Total_Payments", "Avg_Medicare_Payments"))

```

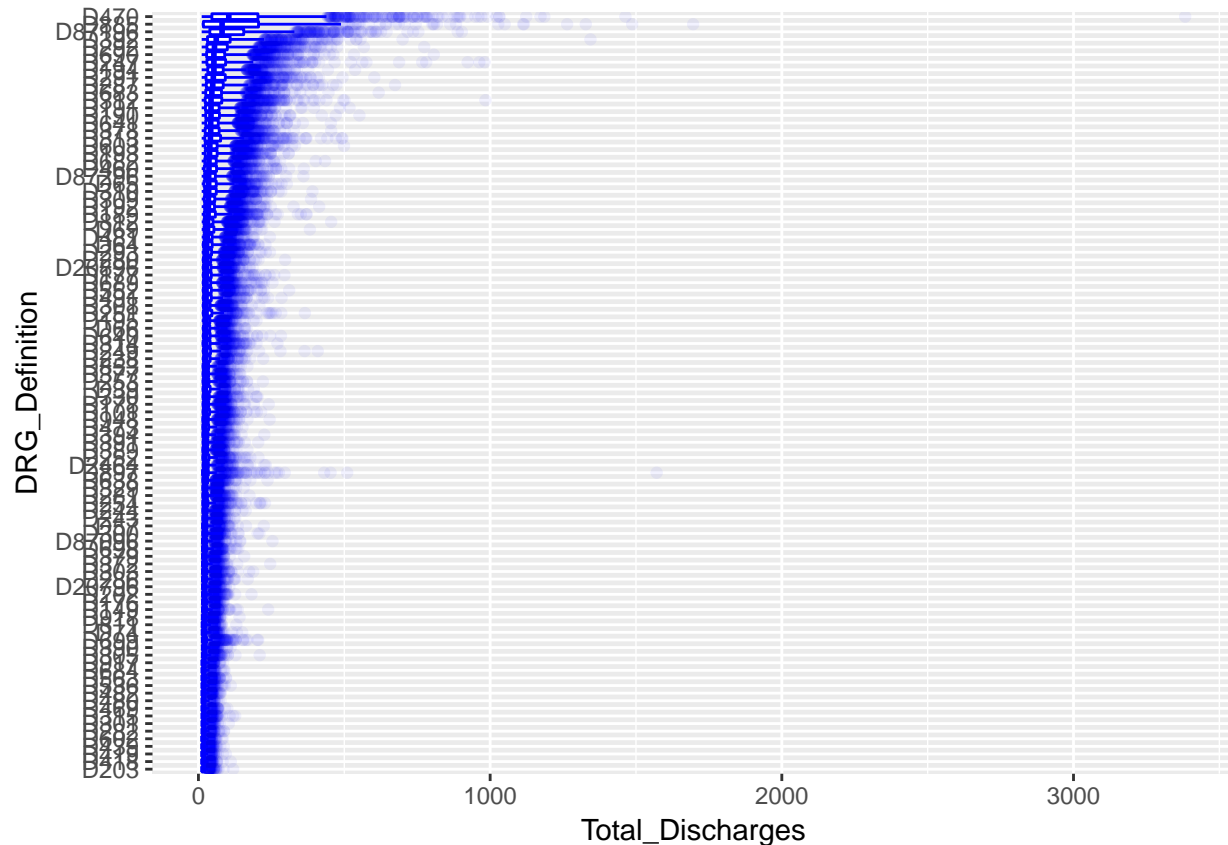
Since the dataset is large a subset of the dataset has been taken, in order to do that I have found the drug that has been used the most based on the Total_Discharges. Total_Discharges gives the maximum number of cases.

To analyse the total discharges for the drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” per state.

Below is a Boxplot of DRG_Definition and Total_Discharges.

```
Inpatient_Drug<-Inpatient%>%filter(DRG_Definition == "D470")%>%
  group_by(Provider_State)

ggplot(data = Inpatient) +
  geom_boxplot(mapping = aes(x = reorder(DRG_Definition, Total_Discharges, FUN = median),
    y = Total_Discharges),color="blue",alpha="0.05") +
  coord_flip()+
  xlab("DRG_Definition") +
  ylab("Total_Discharges")
```



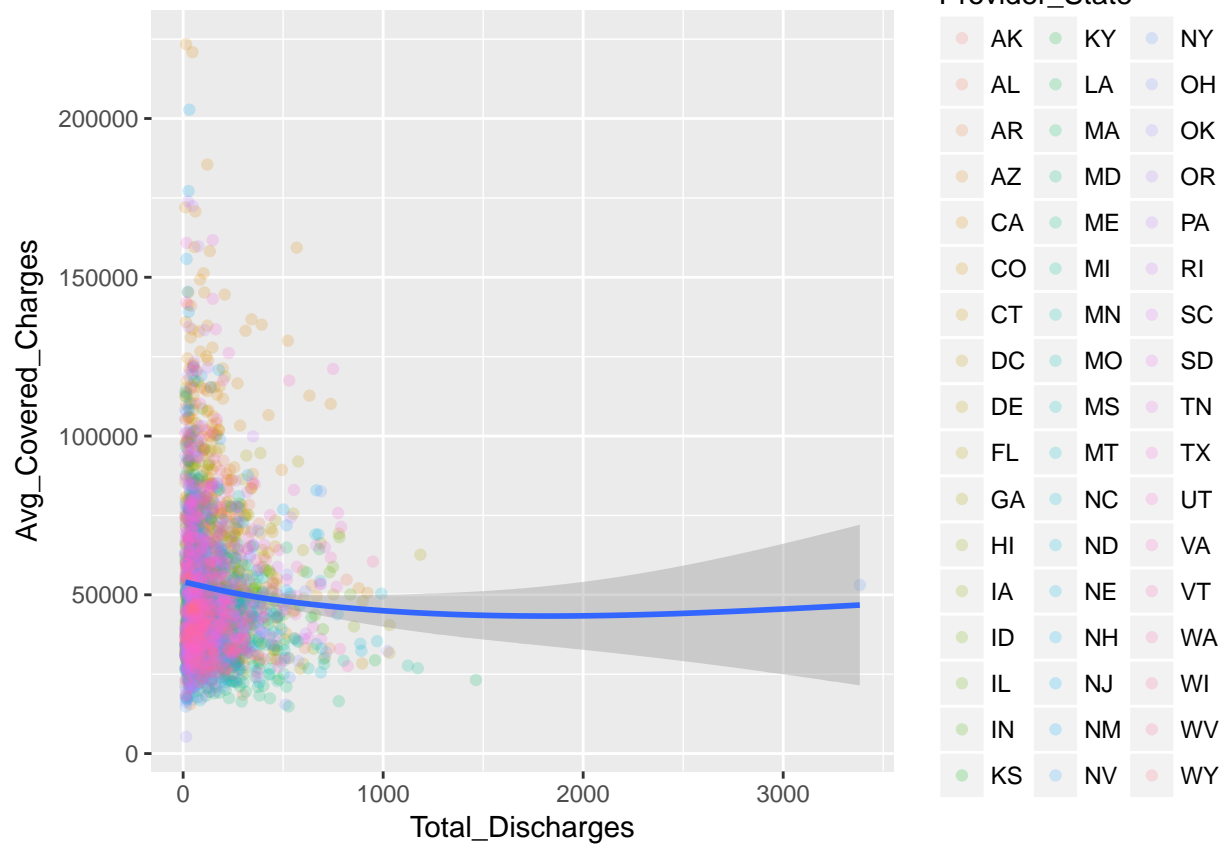
From the graph we see that the drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” has maximum number of discharges and the drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” has the minimum number of discharges. I choose the drug with maximum number of discharges to analyse. I would like to know why this drug has been used the most. As per my understanding since medicare is for old and disabled people there is a possibility that old people tend to have knee problems and so is the case with disabled people.

The next step is to analyse why this drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” has maximum number of discharges. The next step is to create a subset of data based on that drug.

```
ggplot(data = Inpatient_Drug, mapping = aes(x = Total_Discharges, y = Avg_Covered_Charges)) +
  geom_point(mapping = aes(color = Provider_State,se=FALSE),position = "jitter",alpha=0.20) +
  geom_smooth()
```

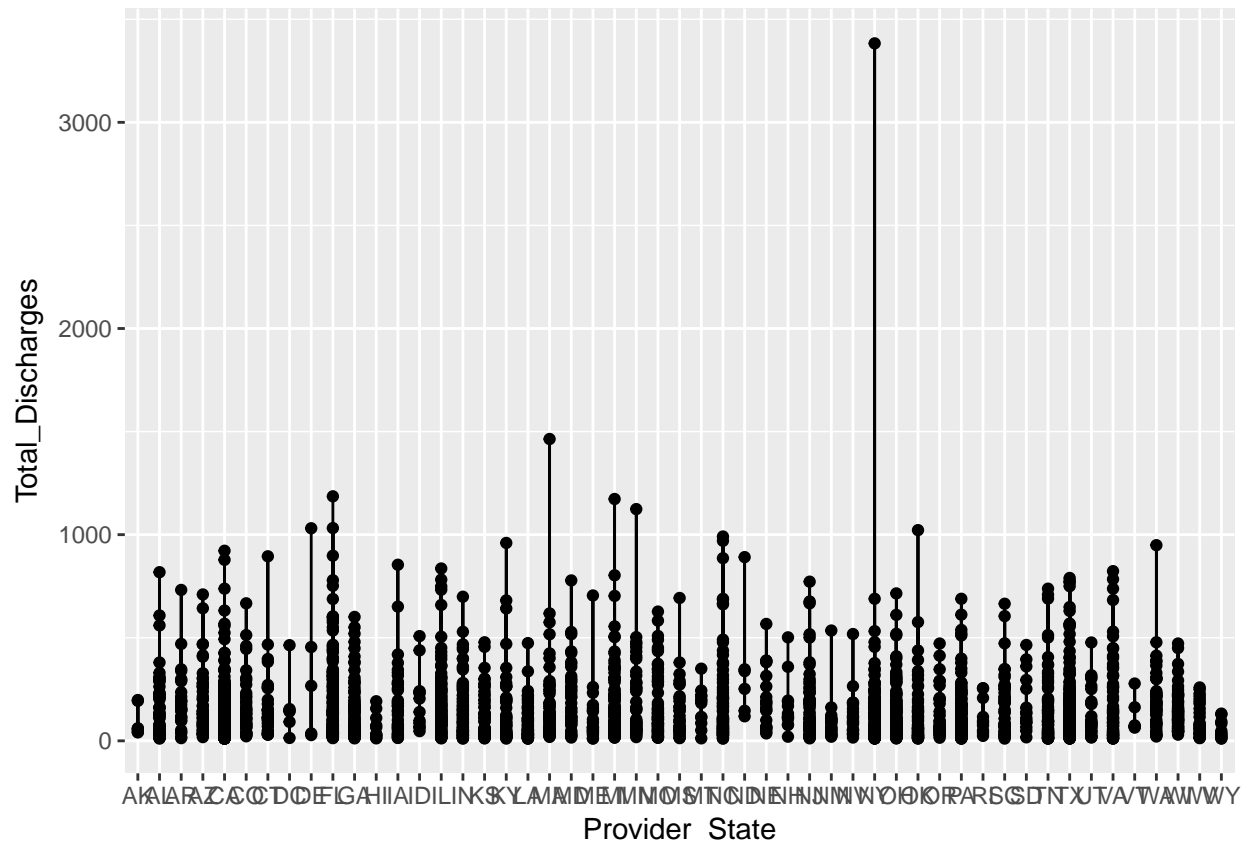
```
## Warning: Ignoring unknown aesthetics: se
```

```
## `geom_smooth()` using method = 'gam'
```



This graph shows that there is an outlier. That point looks like is for newyork but it doesn't provide a clear indication, hence i have another graph as below.

```
ggplot(Inpatient_Drug, aes(Provider_State, Total_Discharges)) +
  geom_point()+
  geom_line()
```



The above graph shows that newyork has most number of cases with an outlier. The next step is to analyse the cause for this. Below is a dataset to analyse it.

```
## To get the data of Newyork city for the drg D470
```

```
Inpatient_Drug_D470_NY<- Inpatient_Drug%>%filter(Provider_State=="NY")%>%
  arrange(Total_Discharges)
```

```
Inpatient_Drug_D470_NY<-select(Inpatient_Drug_D470_NY,
  DRG_Definition,
  Provider_State,
  Provider_Name,
  Total_Discharges,
  Avg_Covered_Charges)
```

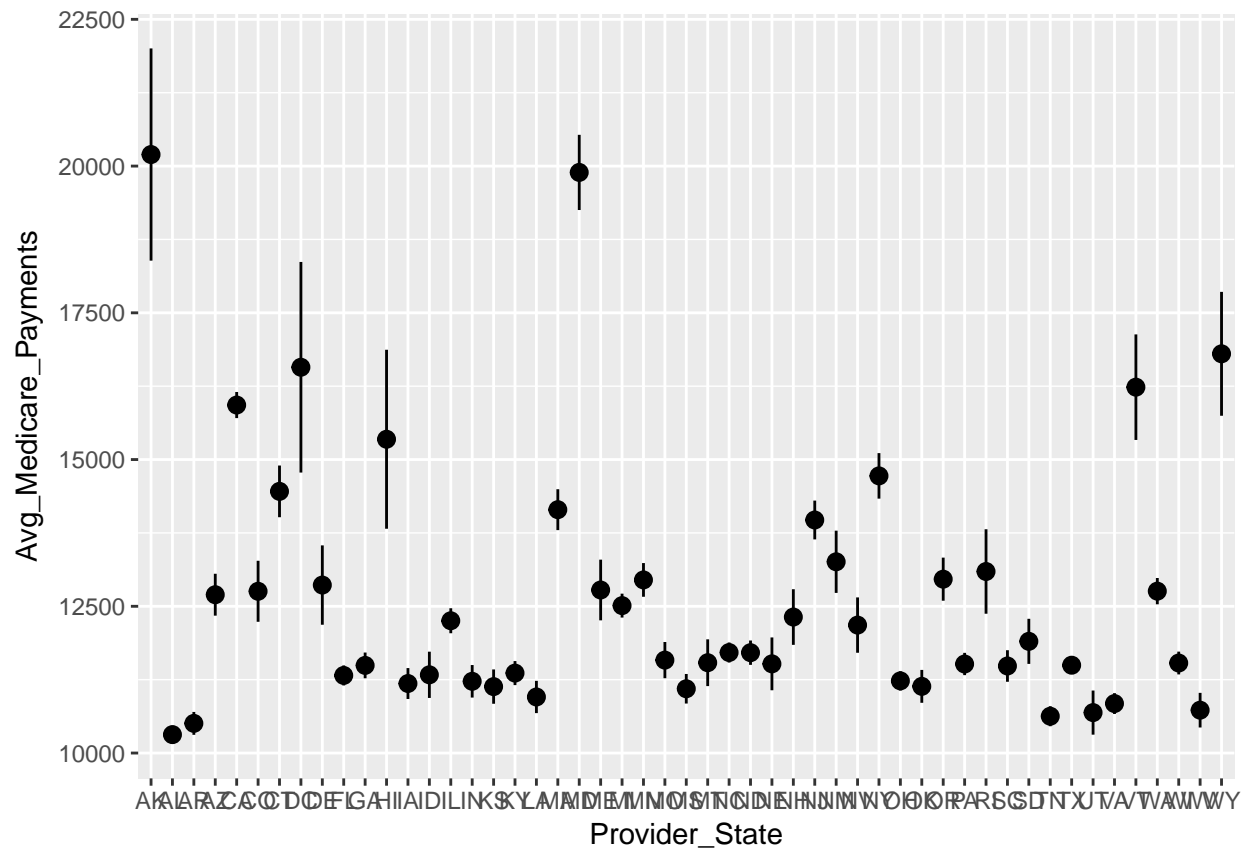
```
View(Inpatient_Drug_D470_NY)
```

From the above results we can say that the total number of dicharges are more at the provider_name=HOSPITAL FOR SPECIAL SURGERY. It could be because there are more old and disabled people in that area and the other possibility could be the population in that area.

To analyse the average medicare payments for the drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” per state.

```
ggplot(data = Inpatient_Drug) +
  stat_summary(
    mapping = aes(x = Provider_State, y = Avg_Medicare_Payments))
```

```
## No summary function supplied, defaulting to `mean_se()
```



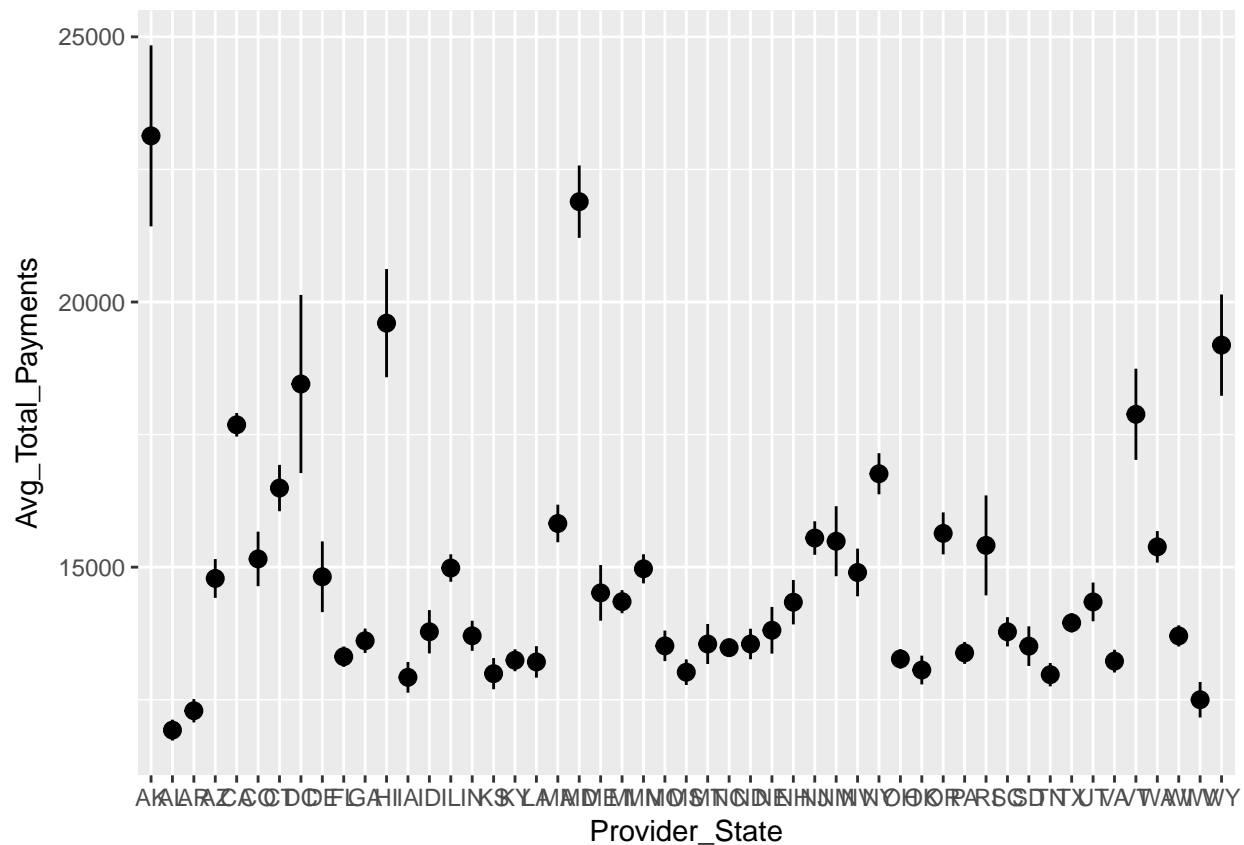
From the above graph we can see that AK(Arkansas) has the maximum average medicare payments. This could be because of the population in those states and the funding is fixed for each state.

```
Inpatient_med_AK<-Inpatient_Drug%>%filter(Provider_State == "AK")%>%
  arrange(Avg_Medicare_Payments)
```

```
View(Inpatient_med_AK)
```

```
## To find the average total payments
ggplot(data = Inpatient_Drug) +
  stat_summary(
    mapping = aes(x = Provider_State, y = Avg_Total_Payments))
```

```
## No summary function supplied, defaulting to `mean_se()
```



The above graph provides information about the amount paid by the individual. From this we can say that the medicare payments are same as the average total payments. I believe that this is because of the population in those states.

Conclusion.

In conclusion the drug “470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC” has maximum number of discharges. Newyork city has the maximum number of discharges. HOSPITAL FOR SPECIAL SURGERY in newyork has maximum number of discharges. Average medicare payments are more in arkansas Average medicare payments and average total payments are same.