Regional Comparison of Facility Charge Amounts for Inpatient Acute Myocardial Infarction Care

An analysis performed by:
Andi McCollam

March 22, 2016

INTRODUCTION

One feature of the United States health care system is the unpredictability of costs associated with health care delivery. Prices for the same services can vary wildly depending on factors such as the setting where care is delivered or who is ultimately paying for the care. These costs are especially difficult to predict during emergency situations.

Inpatient hospital charges are often presented as Diagnosis Related Groupings (DRGs). Medicare uses DRGs to determine how much a hospital should be reimbursed for services relating to a particular diagnosis under its Inpatient Prospective Payment System (IPPS). While Medicare only pays hospitals this IPPS amount for services, hospitals initially present Medicare with bills for their full chargemaster amounts. A chargemaster amount is like the “sticker price” on a car – it’s the hospital’s asking price, but not necessarily an amount they expect to receive in payment. However, the chargemaster amount still has an impact on the price of healthcare services, especially for uninsured people who are often forced to pay the full amount.

A common emergency which can result in an inpatient hospital say is an acute myocardial infarction (AMI). The most serious AMIs are those with multiple comorbid conditions (DRG 280). The objective of this analysis is to determine if there are statistically significant differences in chargemaster prices for DRG 280 between different geographic regions.

METHODS

To discover any statistically significant regional differences in chargemaster prices for DRG 280, CMS IPPS data was reviewed. Specifically, the data set "Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011" was downloaded from Data.CMS.gov for analysis (*Table 1*). This data lists charge information for every inpatient facility nationwide that delivered care for the 100 most common DRGs billed to Medicare in fiscal year 2011. It includes geographic information for each facility, such as state and ZIP code. It also lists the number of times each DRG was billed by each facility in fiscal year 2011, the average covered charges (chargemaster rate) for these DRGs, and average Medicare payments.

After some initial data manipulation was performed[1], a high level analysis was performed on data at the state level. A new table was created with the mean covered charge amounts for the facilities in each state. As individual claim data was not available, these state mean covered charge amounts are the average of every facility's average charge amounts[2]. The national mean (40,443.13) and standard deviation (16,353.51) were also calculated[3].

After performing this initial review at the state level, the original IPPS data for DRG 280 was assigned into four geographic regions based on provider state: Northeast, North Central, South, and West. This was accomplished by using R's built-in US State Facts and Figures data. Washington, DC needed to be separately assigned to the South region as it was not included in the state data[4].

A series of statistical tests of equivalence were then performed to compare the mean average charge counts for inpatient care associated with an AMI with multiple comorbid conditions.

RESULTS AND DISCUSSION

To understand the shape of the state level data, a bar chart was created, graphing all 50 states plus Washington, DC with their mean covered charge amounts (*Figure 1*)[5]. Prices for this same DRG ranged from a low of $17,692.27 in Maryland to a high of $89,859.87 in New Jersey, resulting in a range of over 5 times the minimum value.

To better understand the individual state data regionally, a heat map was created of the lower continuous 48 states (*Figure 2*). This was accomplished using the ggplot2 and maps libraries in R. Additional data manipulation was needed in order to use the mapping functions. For example, the geom_map function could not use two-letter state abbreviations and needed the full state names to be provided in all lower-case letters. This modification was made using the US State Facts and Figures data sets available in R by default.[6] At a glance, the west coast appears to have a cluster of states with high charge amounts for DRG 280.

To compare the equivalence of mean charge amount across different regions, an ANOVA test would typically be performed. However, the ANOVA test has three assumptions which must be met:

- Observations must be independent
- The dependent variable must be normally distributed for each group
- The dependent variable must have equal variance across groups

The observations are based on a collection of individual medical claims, and are most likely independent. To visually determine if the data appeared to meet the assumptions of normality and equal variance, , a series of QQ plots were generated for each category of data[7]. None of the regions appeared to meet these last two assumptions, so formal tests of normality and equal variance were performed.

To test for the assumption of normality, a Shapiro-Wilk normality test was performed for each region[8]. All four regions rejected this test's null hypothesis at a high level of significance that they were normally distributed (Northeast p<2.2e-16, North Central p=4.258e-14, South p<2.2e-16, West p=1.999e-11).

To test for the assumption of equal variance, a Bartlett test of homogeneity of variances was performed. This test has the null hypothesis that all four regions have equal variance. The null hypothesis was rejected at the p<2.2e-16 level with 3 degrees of freedom[9].

Since the assumptions of normality and equal variance were not met, an ANOVA is no an appropriate test of equality. The nonparametric Kruskal-Wallis test is more appropriate in this case.[10] The Kruskal-Wallis null hypothesis that mean charge amounts are equal across regions was rejected at a high level of significance (p<2.2e-16, 3 df).

Since it has been determined that inpatient charge amounts for AMI with multiple comorbid conditions are not equal at a high level of statistical significance, a boxplot was generated to visually compare the data from different regions (*Figure 3*)[11]. The West region appears to have a distribution different from the other three regions. To formally determine which regions had means different from each other region, a series of Wilcoxon Rank Sum Tests were performed[12]. The Wilcoxon null hypothesis that the means between regions are equal was rejected at the $p<2.2e-16$ level comparing the West region to all three other regions. The null hypothesis was rejected that the South and North Central had the same prices at the $p=4.734e-08$ level. The null hypothesis that the Northeast and North Central had equal means was rejected at the 95% confidence level ($p=0.03255$). However, the null hypothesis of equal means was not rejected for the Northeast and South regions ($p=0.1168$).

CONCLUSIONS

After analyzing the facility pricing data, it was determined that there is a statistically significant difference in chargemaster price for AMI with multiple comorbid condition inpatient care across different regions of the United States, with the exception of the Northeast and South regions. However, the Northeast and South data indicate that there is still extensive price variation between different states and even different facilities in these two regions. For example, New Jersey in the Northeast region had the highest average charge amounts of any state studied, while Maryland, in the South region, had the lowest average charge amounts.

This extensive unpredictability of charge amount has far-reaching consequences for the American Healthcare system and the overall system costs borne by everybody. Variance on similar services is indicative of waste and opportunity for improvement.

Table 1

Columns	***Description***
DRG Definition	Description of Presenting Diagnosis
Provider ID	Provider ID number
Provider Name	Facility Name
Provider Street	Facility Address
Provider City	Facility City
Provider State	Facility State
Provider ZIP	Facility ZIP Code
Hospital Referral Region Description	Greater Metropolitan Region
Total Discharges	Number of Discharges for DRG
Average Covered Charges	Average Chargemaster Prices
Average Total Payments	Average Payments for Services (all sources)
Average Medicare Payments	Average Medicare Payments for Services

Figure 1

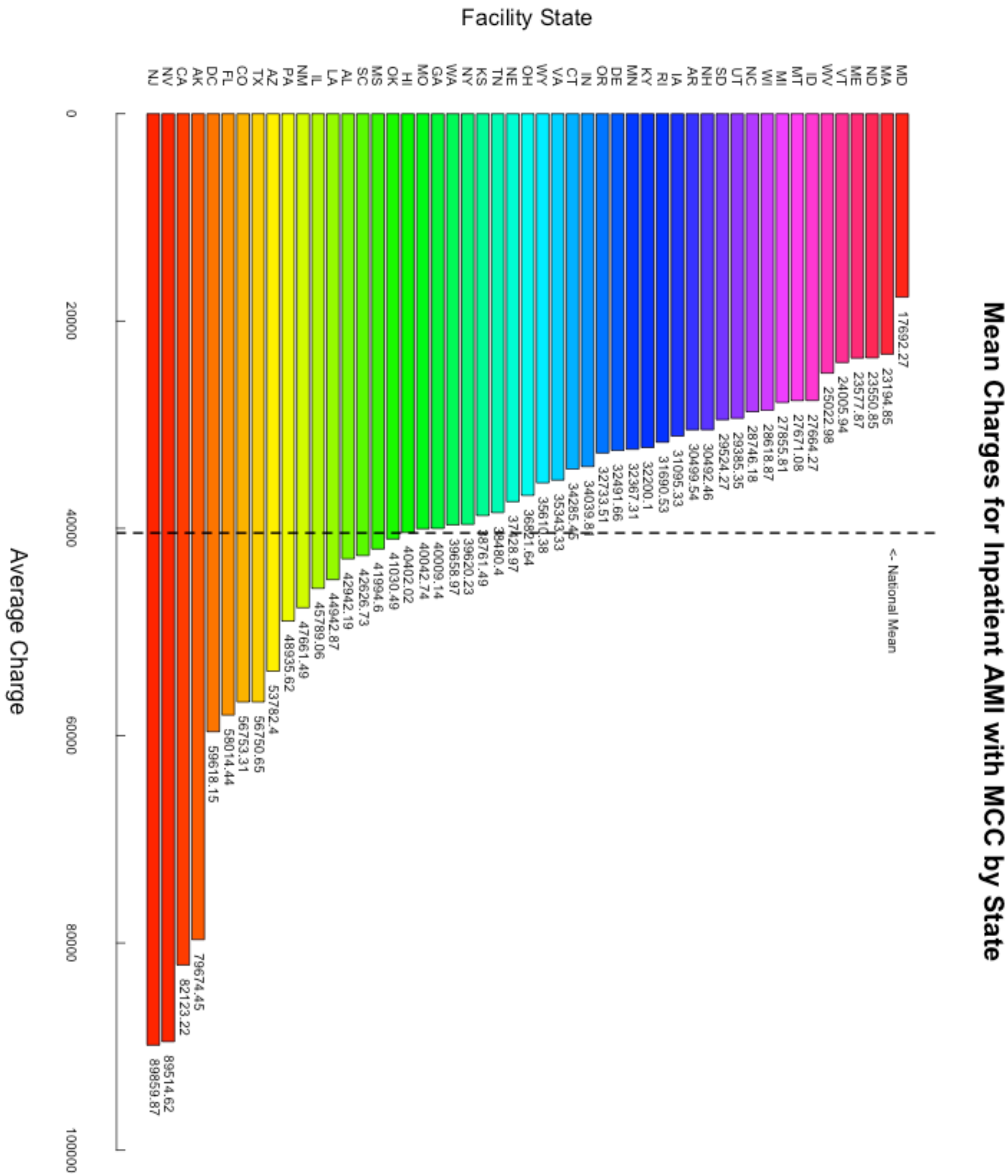

Mean Charges for Inpatient AMI with MCC by State
Facility State
Average Charge
<- National Mean
0
20000
40000
60000
80000
100000

Figure 2

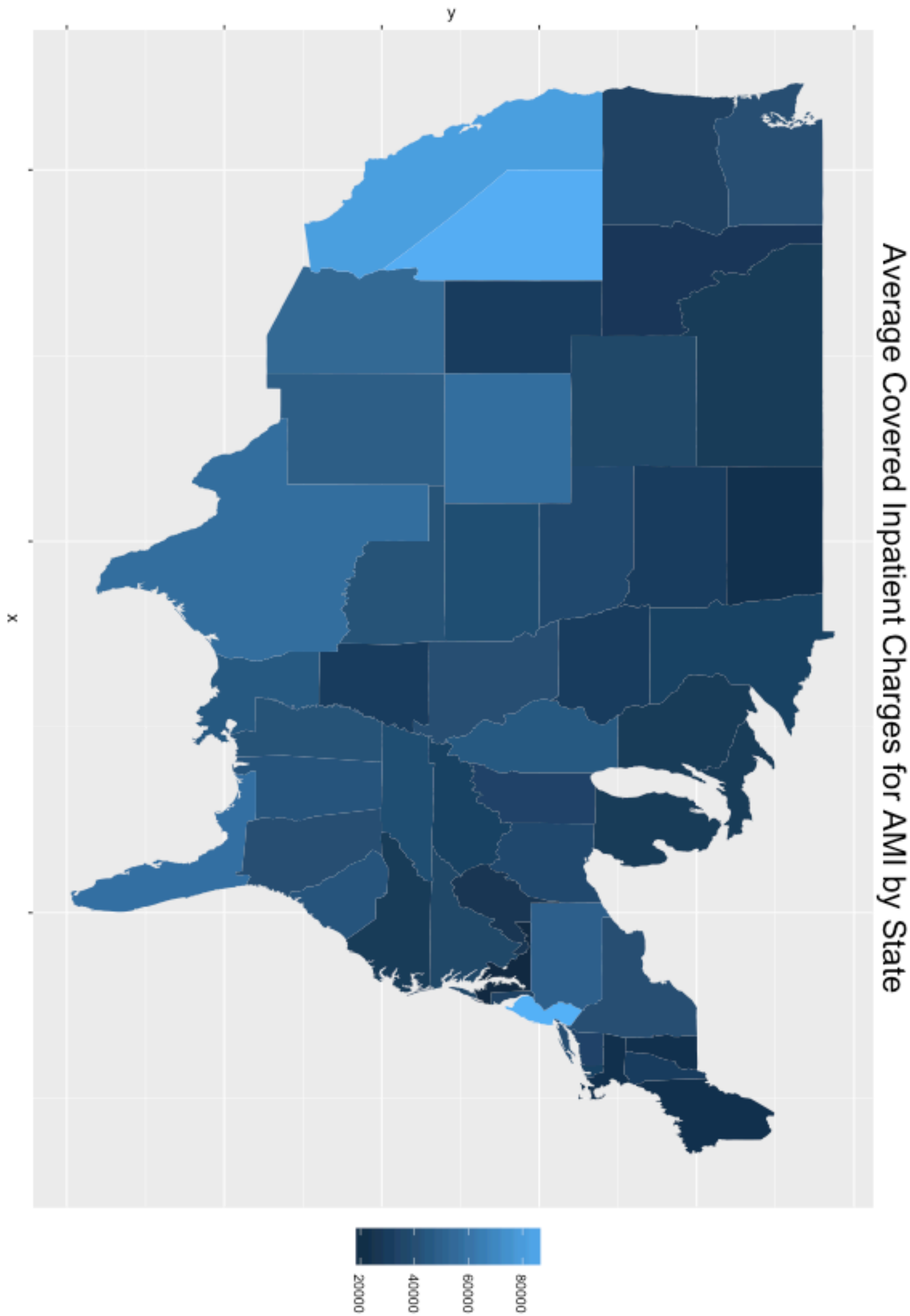
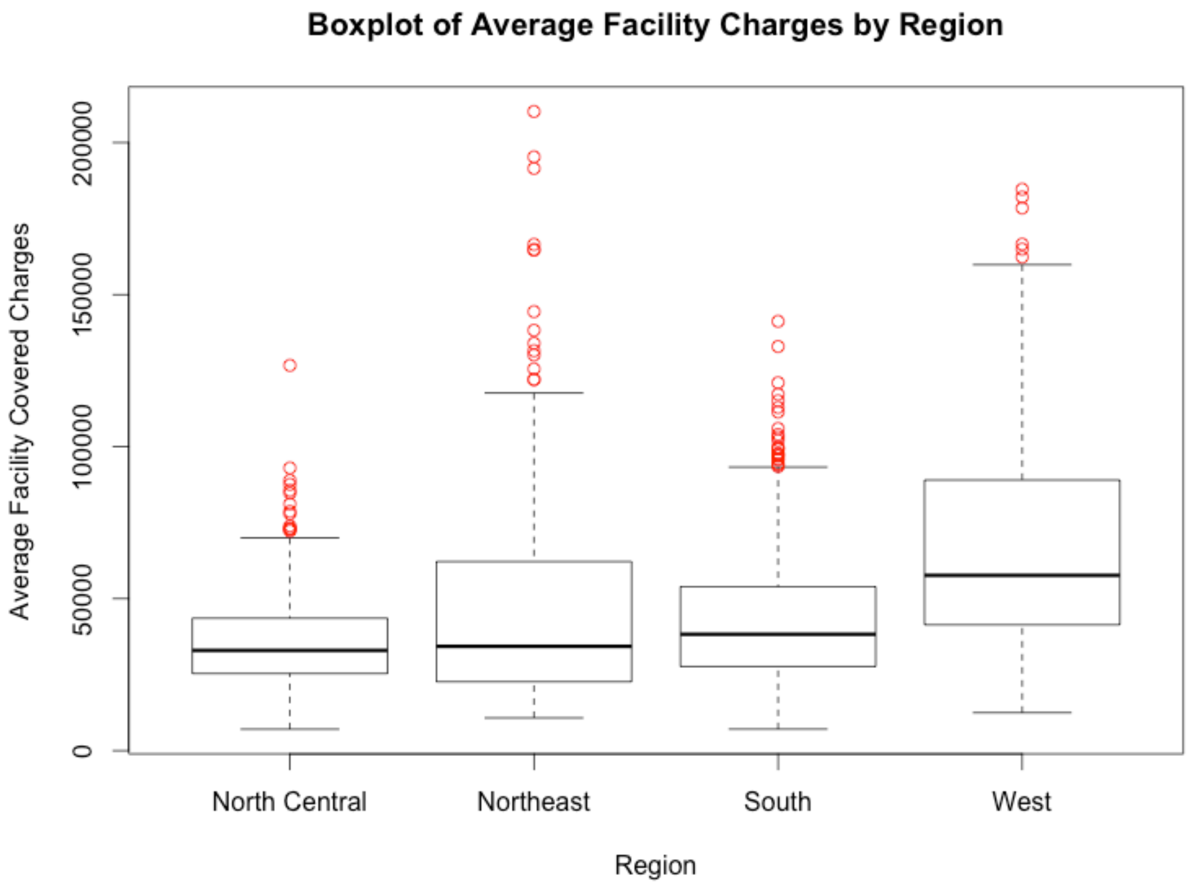

Average Covered Inpatient Charges for AMI by State
y
x
20000
40000
60000
80000

Figure 3


Boxplot of Average Facility Charges by Region
Average Facility Covered Charges
0
50000
100000
150000
200000
North Central
Northeast
South
West
Region

R CODE

[1]#The first problem I encountered was that Average.Covered.Charges, my dependent variable, were structured as Factor and had leading $ signs, so I couldn't do math on it.

```
>IPPS$CovChg<-(substring(Average.Covered.Charges,2))
>IPPS$CovChg<-as.numeric(IPPS$CovChg)
> str(CovChg)
num [1:4553] 40237 49636 47469 44164 17021 ...
```

```
#I created a subset of the data using only the most complex AMI cases, DRG 280.
> IPPS_MCC<-subset(IPPS,DRG.Definition=="280 - ACUTE MYOCARDIAL INFARCTION, DISCHARGED ALIVE W MCC")
```

[2] #Created a table of 51 observations with the mean of the Average.Covered.Charges for each observed facility (individual claims information was not available).

```
> MCC_barchart<-tapply(IPPS_MCC$CovChg,IPPS_MCC$Provider.State,mean)
```

[3]

```
> mean(CovChg_means_MCC)
[1] 40443.13
> sd(CovChg_means_MCC)
[1] 16353.51
```

[4]

```
#using state.name function to assign region
IPPS_MCC$Prov.Region <-
state.region[match(IPPS_MCC$Provider.State,state.abb)]

#Need to manually assign DC to "South"
#Setting region info to Vector so region names preserved for next steps
IPPS_MCC$Prov.Region<-as.vector(IPPS_MCC$Prov.Region)

#Creating a column of region names
DCMatrix<-ifelse(IPPS_MCC$State.Vector=="DC","South",IPPS_MCC$Prov.Region)

#Replacing IPPS_MCC$Prov.Region with DCMatrix
IPPS_MCC$Prov.Region<-DCMatrix
```

[5]

```
#Barchart code
> MCC_barchart<-tapply(IPPS_MCC$CovChg,IPPS_MCC$Provider.State,mean)

#Sorting table to have data show trend
> MCC_barchart<-sort(MCC_barchart, decreasing=TRUE)
> barplot(CovChg_means_MCC,horiz=TRUE,cex.names=.4,las=2,tck=1,main="Mean Charges for Inpatient AMI with MCC by State",ylab="Facility State",xlab="Average Charge")

#This sets the location of the value labels for all states.
> Plotbars<-
barplot(MCC_barchart,horiz=TRUE,cex.names=.6,cex.axis=.5,las=1,tck=1,xlim=c(0,100000),col=rainbow(51),main="Mean Charges for Inpatient AMI with MCC by State",ylab="Facility State",xlab="Average Charge")
```

```
#This is the main plot. Chose rainbow color gradient with 51 levels.
>
barplot(MCC_barchart,horiz=TRUE,cex.names=.6,cex.axis=.6,las=1,tck=0.01,xlim=
c(0,100000),col=rainbow(51),main="Mean Charges for Inpatient AMI with MCC by
State",ylab="Facility State",xlab="Average Charge")

#Places a mean value line
> abline(v=40443.13,par(lwd=2,lty=2))
> text(47000,60,"<- National Mean",cex=.6)

#got rid of scientific notation on x-axis
> options(scipen=5)

#displays values for each row
> text(y=Plotbars,x=MCC_barchart,label=round(MCC_barchart,digits=2), pos=4,
cex=.6)


[6]#Map Code - map of the United States describing average price for each state
using ggplot2 library
#Load ggplot2
> library(ggplot2)

# To get state geography information (i.e. latitude and longitude) for ggplot
to pull from, need to install and call maps package
> install.packages("maps")
> library(maps)

# ggplot is going to be looking for the appropriate level of mapping
information in a variable called map.  Calling this at the state level.
> map <- map_data(“state”)

#To get this to work, I need to transform CovChg_means_MCC
> CovChg_means_MCC<-as.data.frame(CovChg_means_MCC)

#CovChg_means_MCC data frame has row names as two-letter state abbreviations
and the associated means as values.
#Can’t reference row names in plot, so creating column with state
abbreviations as values.
> CovChg_means_MCC$state<-row.names(CovChg_means_MCC)
> CovChg_means_MCC$charges<-CovChg_means_MCC$V1

# Geom_map function can’t use state abbreviations, and requires that all
states be referenced as lower case full state names.  R has a built-in
state.name crosswalk for this.
#note - DC was not considered a state
> CovChg_means_MCC$statename <-
tolower(state.name[match(CovChg_means_MCC$state,state.abb)])

#ggplot code to generate map.
>
ggplot(CovChg_means_MCC,aes(fill=CovChg_means_MCC$charges))+geom_map(aes(map_
id=CovChg_means_MCC$statename),map=map)+expand_limits(x=map$long,
y=map$lat)+ggtitle("Average Covered Inpatient Charges for AMI by
State")+theme(plot.title=element_text(size=20),axis.text.y=element_blank(),ax
is.text.x=element_blank(),legend.title=element_blank());
```

```
[7]#Ran QQ plots with normal reference line for all regions
attach(West_MMC)
qqnorm(CovChg, main = "QQ-plot for Average Charges")
qqline(CovChg)

[8] #Doing formal Shapiro tests
> shapiro.test(Northeast_MMC$CovChg)

Shapiro-Wilk normality test

data:  Northeast_MMC$CovChg
W = 0.83319, p-value < 2.2e-16

> shapiro.test(Northcentral_MMC$CovChg)

Shapiro-Wilk normality test

data:  Northcentral_MMC$CovChg
W = 0.92673, p-value = 4.258e-14

> shapiro.test(South_MMC$CovChg)

Shapiro-Wilk normality test

data:  South_MMC$CovChg
W = 0.92484, p-value < 2.2e-16

> shapiro.test(West_MMC$CovChg)

Shapiro-Wilk normality test

data:  West_MMC$CovChg
W = 0.9223, p-value = 1.999e-11

# Reject H0 for all regions that they are are normally distributed.



[9] #Checking for equal variance using Bartlett test
bartlett.test(IPPS_MCC$CovChg~IPPS_MCC$Prov.Region)

Bartlett test of homogeneity of variances

data:  IPPS_MCC$CovChg by IPPS_MCC$Prov.Region
Bartlett's K-squared = 327.29, df = 3, p-value < 2.2e-16

#Reject H0 that the variability in charge amount is equal for all region
categories.


[10]#In this case, we should really use a nonparametric alternative to ANOVA.
#Running a Kruskal-Wallis Test

#Test doesn't like the format of Prov.Region data, so adding a new
#column with Prov.Region as a factor:
IPPS_MCC$Prov.Region.factor<-as.factor(Prov.Region)
```

```
kruskal.test(IPPS_MCC$CovChg~IPPS_MCC$Prov.Region.factor)

Kruskal-Wallis rank sum test

data:  CovChg and Prov.Region.factor
Kruskal-Wallis chi-squared = 206.62, df = 3, p-value <
  2.2e-16

#Reject H0 that prices are equal across regions.

[11]> boxplot(IPPS_MCC$CovChg~IPPS_MCC$Prov.Region,outcol="Red",main="Boxplot of
Average Facility Charges by Region",ylab="Average Facility Covered Charges",
xlab="Region")


[12]wilcox.test(Northeast_MMC$CovChg,Northcentral_MMC$CovChg)
#p-value = 0.03255
wilcox.test(Northeast_MMC$CovChg,South_MMC$CovChg)
#p-value = 0.1168
wilcox.test(Northeast_MMC$CovChg,West_MMC$CovChg)
#p-value < 2.2e-16
wilcox.test(Northcentral_MMC$CovChg,South_MMC$CovChg)
#p-value = 4.734e-08
wilcox.test(Northcentral_MMC$CovChg,West_MMC$CovChg)
#p-value < 2.2e-16
wilcox.test(South_MMC$CovChg,West_MMC$CovChg)
#p-value < 2.2e-16
```