Project 1

Owen McGrattan

3/7/2017

Question #1

```
#Read in the data and subset by the specific diagnoses
library(readr)
data<-read_csv("~/stat28/projects/data/combinedData.csv")
```

```
## Parsed with column specification:
## cols(
##   Provider.State = col_character(),
##   DRG.Definition = col_character(),
##   Provider.Id = col_integer(),
##   Provider.Name = col_character(),
##   Provider.City = col_character(),
##   Total.Discharges = col_integer(),
##   Average.Covered.Charges = col_double(),
##   Average.Total.Payments = col_double(),
##   Average.Medicare.Payments = col_double(),
##   Provider.Zip.Code = col_integer(),
##   regions = col_character(),
##   Urban = col_integer()
## )
```

```
diagnosed<-data[data$DRG.Definition==c("192 - CHRONIC OBSTRUCTIVE PULMONARY DISEASE W/O CC/MCC","293 - I
```

```
## Warning in data$DRG.Definition == c("192 - CHRONIC OBSTRUCTIVE PULMONARY
## DISEASE W/O CC/MCC", : longer object length is not a multiple of shorter
## object length
```

```
#Create new variables
diagnosed$PatientPays<-diagnosed$Average.Total.Payments-diagnosed$Average.Medicare.Payments
diagnosed$PctPatientPays<-diagnosed$PatientPays/diagnosed$Average.Total.Payments

#Omit na values before creating factors
diagnosed<-na.omit(diagnosed)
#Assign character variables for urban
diagnosed$urbanchar<-ifelse(diagnosed$Urban==0,"mix",ifelse(diagnosed$Urban==1,"rural",ifelse(diagnosed$
urbanFactor<-factor(diagnosed$urbanchar)
urbanRegion<-factor(diagnosed$regions)

#Cross Factors and drop levels from dataset
diagnosed$urbanByRegions<-urbanFactor:urbanRegion
diagnosed<-droplevels.data.frame(diagnosed)

summary(diagnosed)
```

```
##  Provider.State     DRG.Definition      Provider.Id     Provider.Name
##  Length:1820        Length:1820        Min.   : 10001   Length:1820
##  Class :character   Class :character   1st Qu.:120022   Class :character
##  Mode  :character   Mode  :character   Median :250069   Mode  :character
```

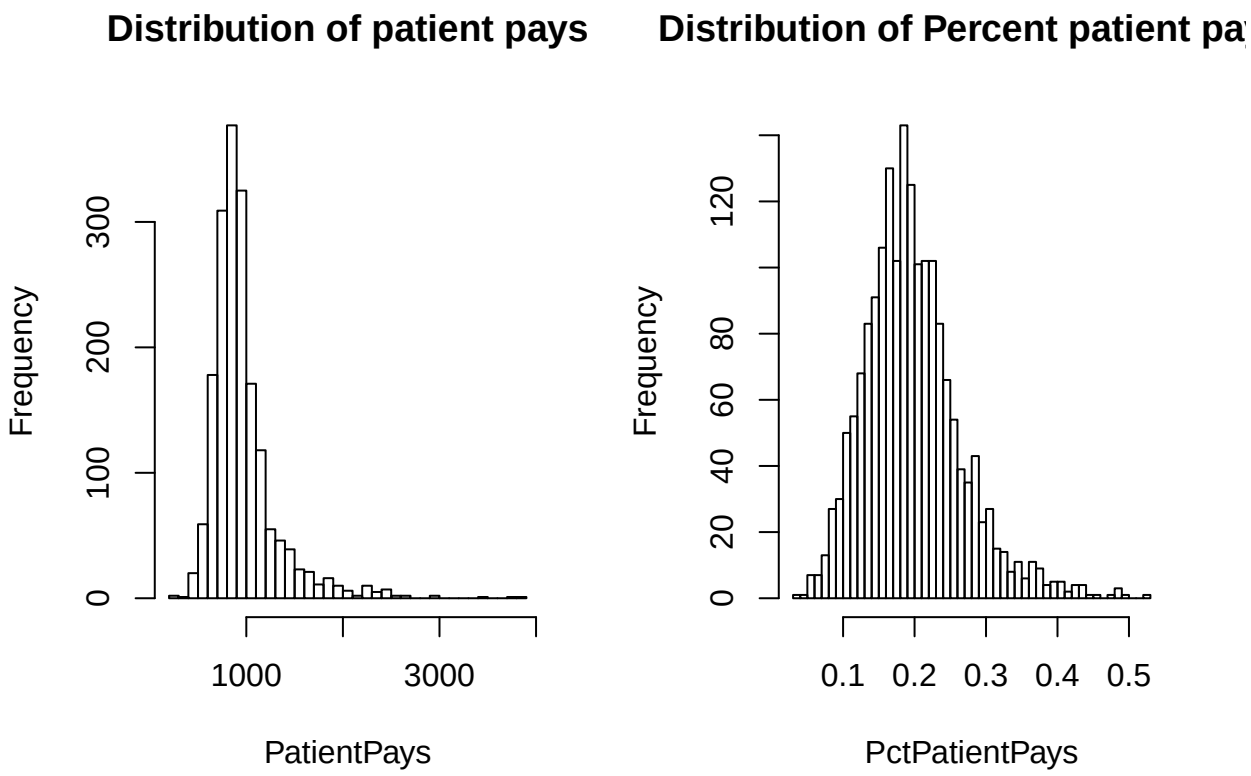
```
##                                      Mean   :258159
##                                      3rd Qu.:390045
##                                      Max.   :670071
##
##  Provider.City      Total.Discharges Average.Covered.Charges
##  Length:1820        Min.   : 11.00   Min.   :  3134
##  Class :character   1st Qu.: 17.00   1st Qu.: 11304
##  Mode  :character   Median : 25.00   Median : 15465
##                     Mean   : 33.41   Mean   : 18475
##                     3rd Qu.: 41.00   3rd Qu.: 22130
##                     Max.   :248.00   Max.   :130690
##
##  Average.Total.Payments Average.Medicare.Payments Provider.Zip.Code
##  Min.   : 3144          Min.   : 2233             Min.   : 1082
##  1st Qu.: 4207          1st Qu.: 3257             1st Qu.:27607
##  Median : 4723          Median : 3759             Median :44634
##  Mean   : 5061          Mean   : 4096             Mean   :48551
##  3rd Qu.: 5544          3rd Qu.: 4537             3rd Qu.:73702
##  Max.   :11989          Max.   :11288             Max.   :99645
##
##    regions              Urban        PatientPays     PctPatientPays
##  Length:1820        Min.   :0.000   Min.   : 277.7   Min.   :0.03898
##  Class :character   1st Qu.:0.000   1st Qu.: 770.2   1st Qu.:0.15223
##  Mode  :character   Median :2.000   Median : 890.3   Median :0.18978
##                     Mean   :2.237   Mean   : 964.9   Mean   :0.19750
##                     3rd Qu.:5.000   3rd Qu.:1045.9   3rd Qu.:0.23182
##                     Max.   :5.000   Max.   :3869.8   Max.   :0.52611
##
##   urbanchar                urbanByRegions
##  Length:1820        mix:south      :285
##  Class :character   rural:south    :272
##  Mode  :character   urban:south    :205
##                     mix:midwest    :172
##                     urban:northeast:161
##                     urban:midwest  :142
##                     (Other)        :583
```

Question 2

Patient pays: Patient pays is the dollar amount that each patient pays out of the total expenses. The mean payment amount is $960 while the median is $890. The payments in the group range from $277 to $3869

PctPatientPays: The percent of the total bill that the patient themselves pay. The mean is 19.75% while the median is 18.97% with the maximum and minimum ranging from 52% to 3%.

```
par(mfrow=c(1,2))
hist(diagnosed$PatientPays,xlab="PatientPays",main="Distribution of patient pays",breaks=50)
hist(diagnosed$PctPatientPays,xlab="PctPatientPays",main = "Distribution of Percent patient pays",breaks
```

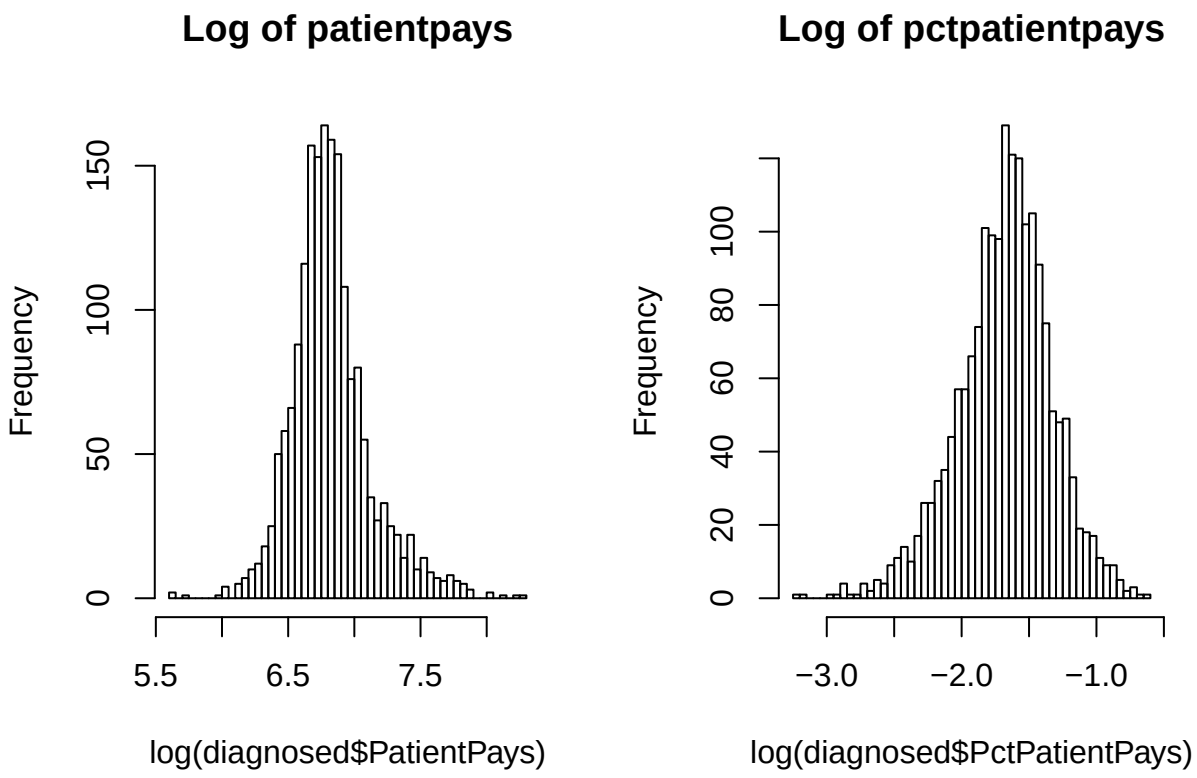


Patient pays has a heavy right hand skewness with much of the data being centered around the $900 area but with outliers that pulls the mean to be greater than the median.

PctPatientPays has a more uniform distribution with a small bit of right hand skewness. There are fewer outliers that drag out and significantly alter the shape of the curve compared to Patient pays.

For PatientPays a logarithmic transformation would be useful given to the high number of values that are around or less than 1000 that aren't even on a visible part of the scale.

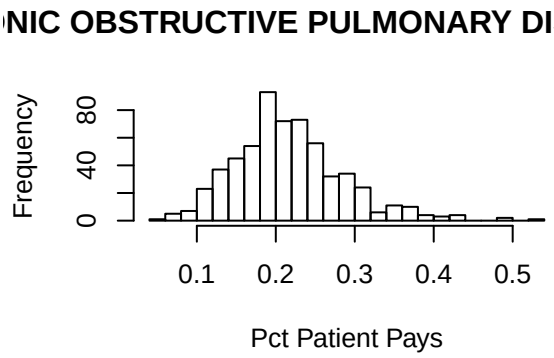
```
par(mfrow=c(1,2))
hist(log(diagnosed$PatientPays),breaks=50,main="Log of patientpays")
hist(log(diagnosed$PctPatientPays),breaks=50,main="Log of pctpatientpays")
```

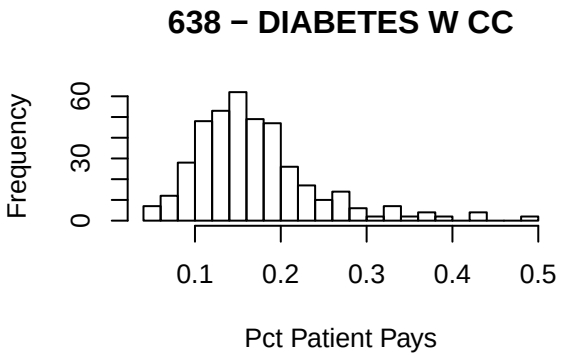


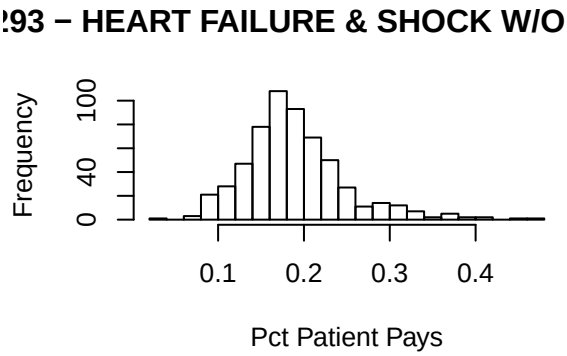
When running this kind of transformation on PatientPays we see a far more normal distribution that isn't being influenced so much by the outliers. We can understand that the highly centralized area from the original distribuiton does make up the bulk of the data and that the data should be seen as something more of a normal distribution around $850 or $950.

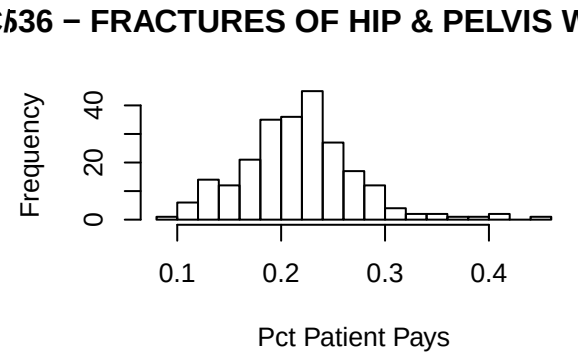
For pctPatientPays we do not see much change since the original distribution was close to normal, only now there is a small skewness to the left.

```
#Create histograms for the four different treatments by patient pays
par(mfrow=c(2,2))
for (i in unique(diagnosed$DRG.Definition)){
  variable<-diagnosed[diagnosed$DRG.Definition==i,]
  hist(variable$PctPatientPays,main=i,xlab="Pct Patient Pays",breaks=25)
}
```









Given the distributions of the individual treatments there is no reason to run any particular transformations for the individual data sets. Although their prices may differ, the distribution of the percent patients pay is similar for each of the treatments and there is no heavy skewness.

```
table(diagnosed$Urban,diagnosed$regions)
```

```
##
##     midwest northeast south west
##   0     172       128   285  132
##   1       2         3    21    0
##   2     109        44   251   41
##   3       0         0     2    0
##   5     142       161   203  124
```

When looking at the contingency table between urban and region we see that the urban values 1,3, and 4 are small and almost nonexistent. This is largely because urban areas will have more people and will require more hospitals and other hospitals will be built on the outskirts of urban areas to serve rural areas. Rural areas only (1) will not have hospitals since there are low numbers of people and there will not be much need for a high number of hospitals. Urban clusters only (3) are most likely near rural areas as well so there will not be many hospitals in those urban clusters only.

Question 3

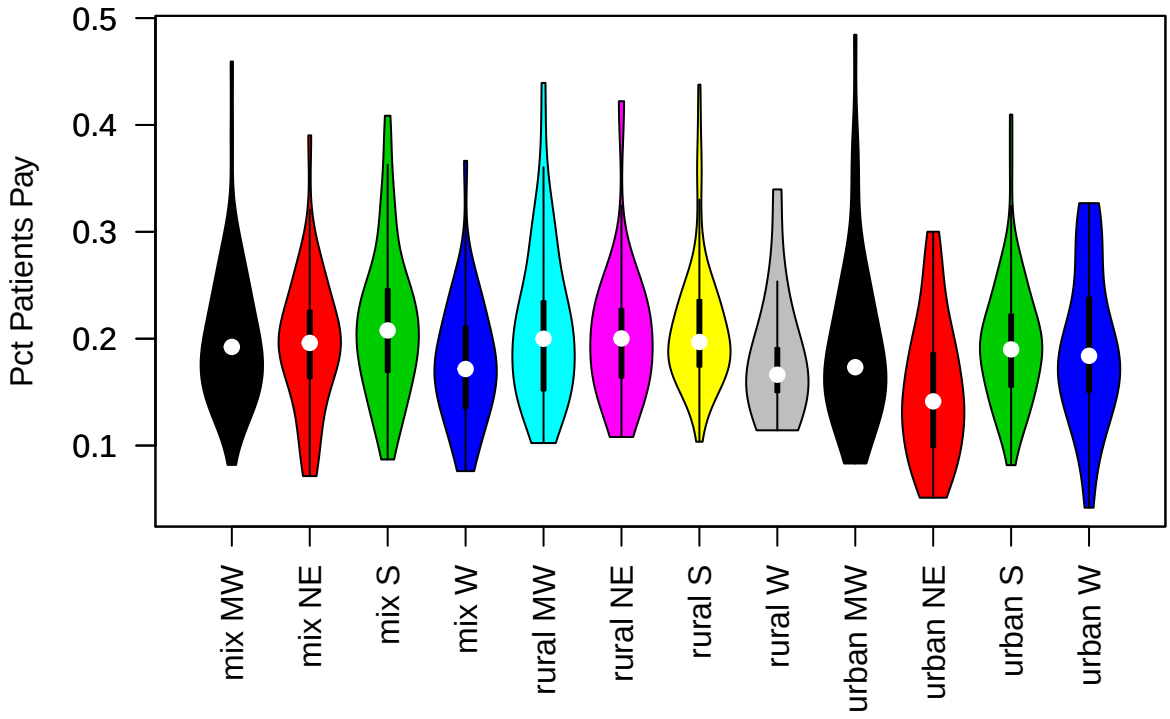
```
#Exclude urban values of 1,3,4
excluded<-diagnosed[diagnosed$Urban==c(0,2,5),]
```

```
## Warning in diagnosed$Urban == c(0, 2, 5): longer object length is not a
## multiple of shorter object length
```

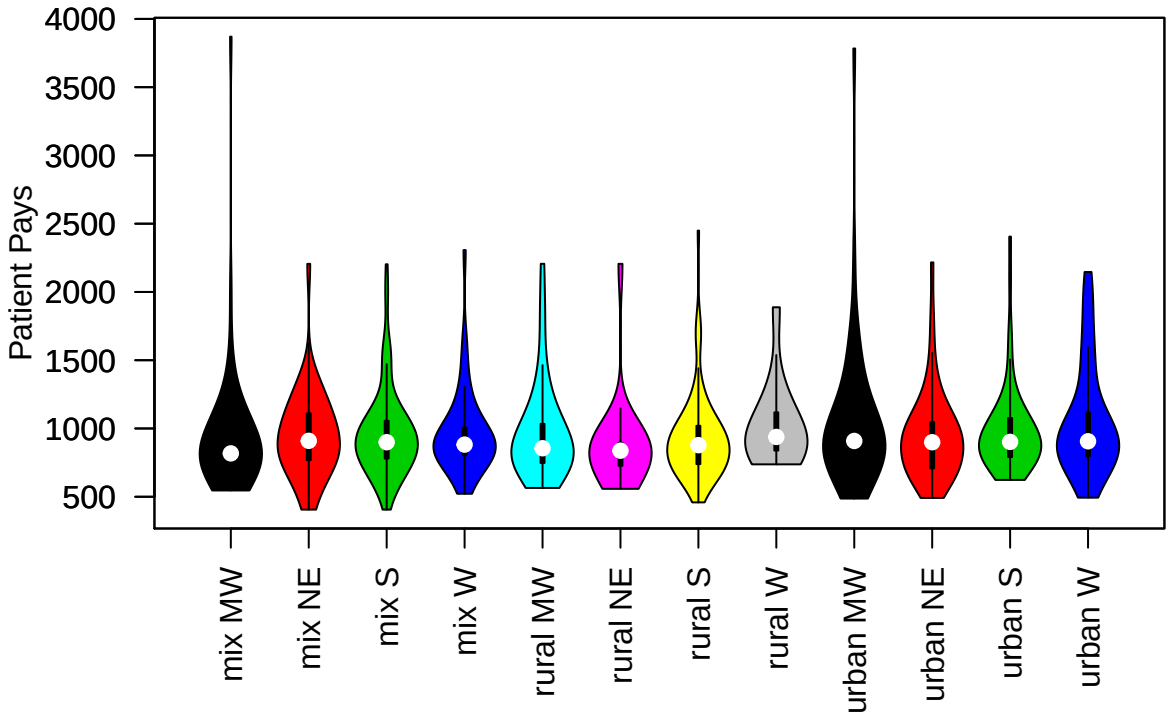
```
#Create shorter factor names for presentation of graphs.
excluded$smfacvar<-factor(excluded$urbanByRegions,labels=c("mix MW","mix NE","mix S","mix W","rural MW"
```

```
#Create violin plots on pctpatientpays and patientpays for each of the factors names in urbanByRegions
source("~/stat28/projects/myvioplot.R")
vioplot2(excluded$PctPatientPays,excluded$smfacvar,las=2,ylab="Pct Patients Pay",col=palette())
```

```
## Loading required package: vioplot
## Loading required package: sm
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
```




```
vioplot2(excluded$PatientPays,excluded$smfacvar,las=2,ylab="Patient Pays",col=palette())
```



The violin plots of pctPatientsPay and PatientsPay show the distributions of pctPatientsPay and PatientsPay for the different regions and area factors.

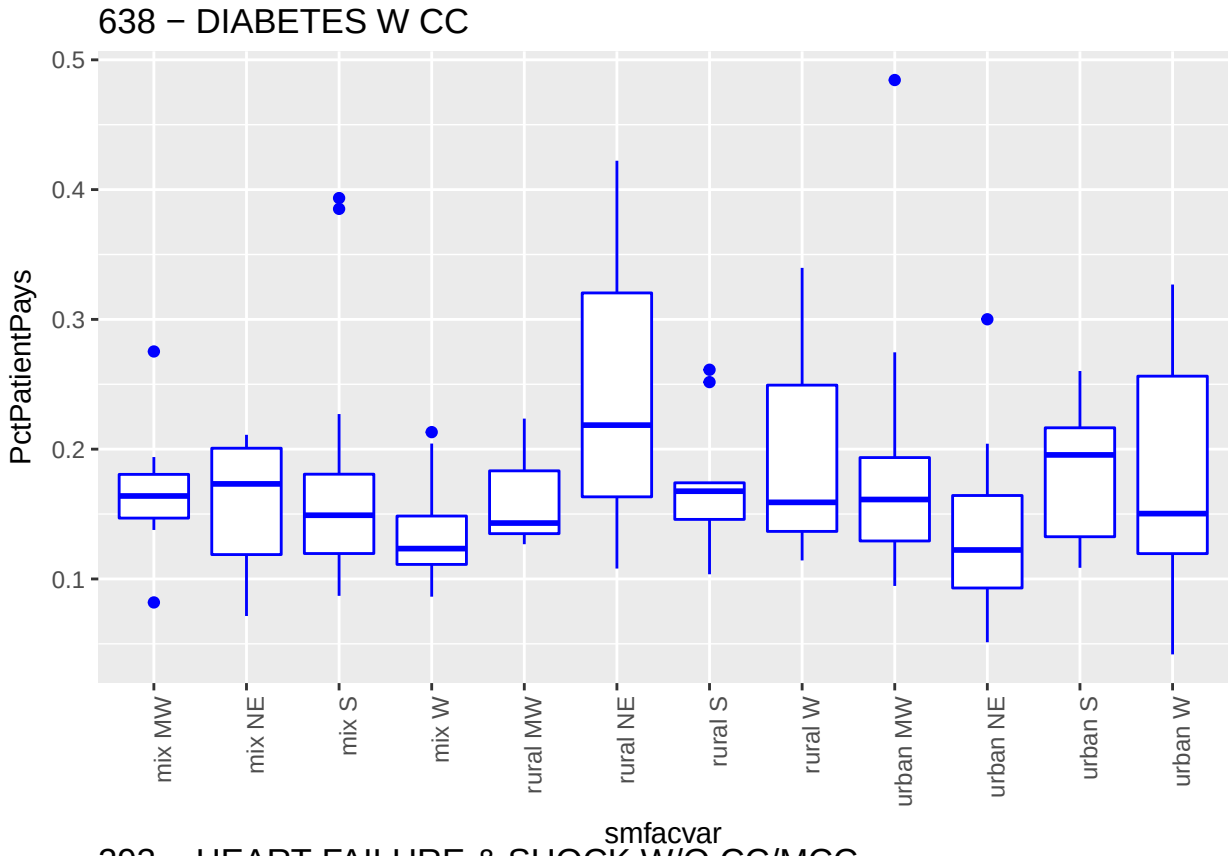
There is not much we can see by looking at the violin plot of Patient pays. The distributions are largely the same with similar spreads as well with the exception of the two outliers in the mix NW and urban NW regions.

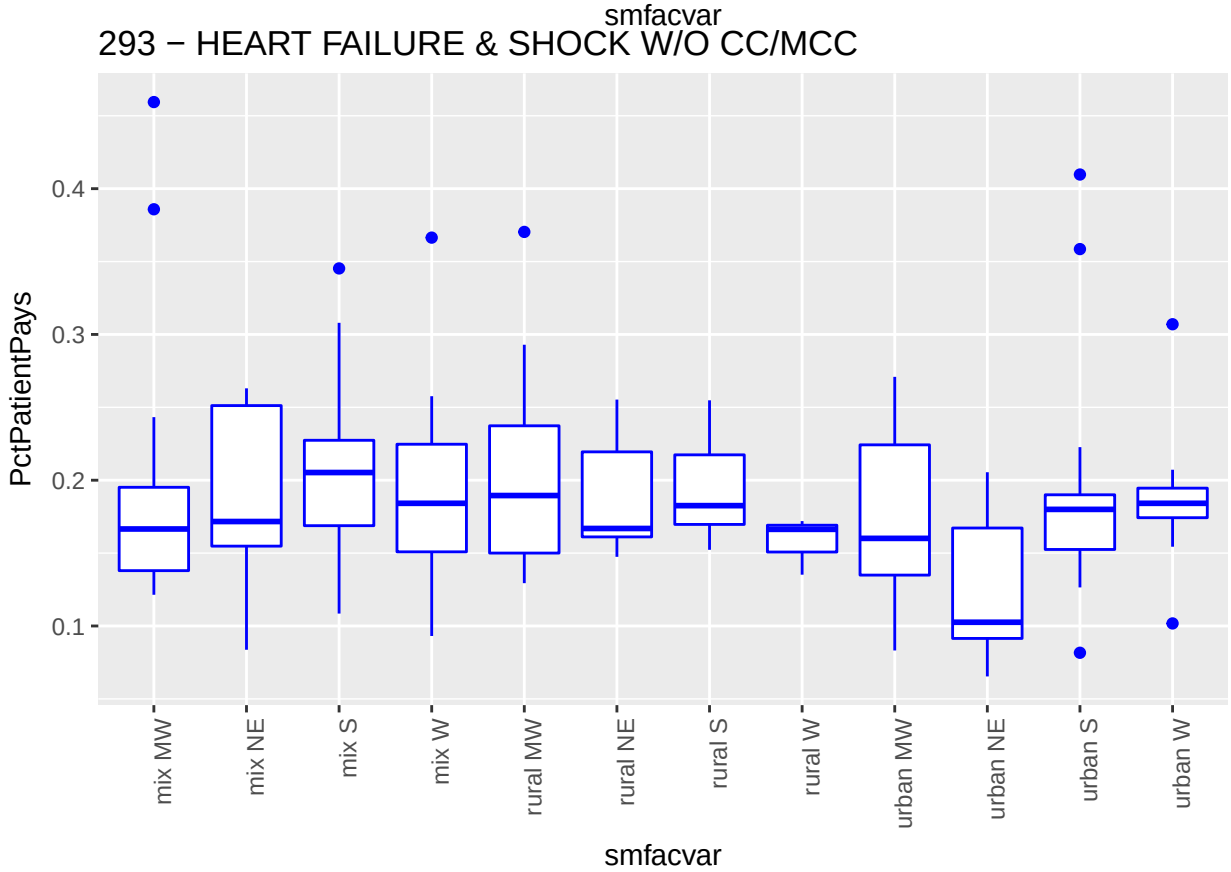
For the pctPatientsPay distribution, the different southern groups all had higher centers and distributions than the other regions. The distribution, center, and spreads for the west groups are all more favorable (lower) compared to the other groups. Even for the urban:west values despite having a higher spread, they have a uniform distribution. The urban:northeast group comes in considerably lower than the other groups with the lowest center and a slightly skewed distribution.

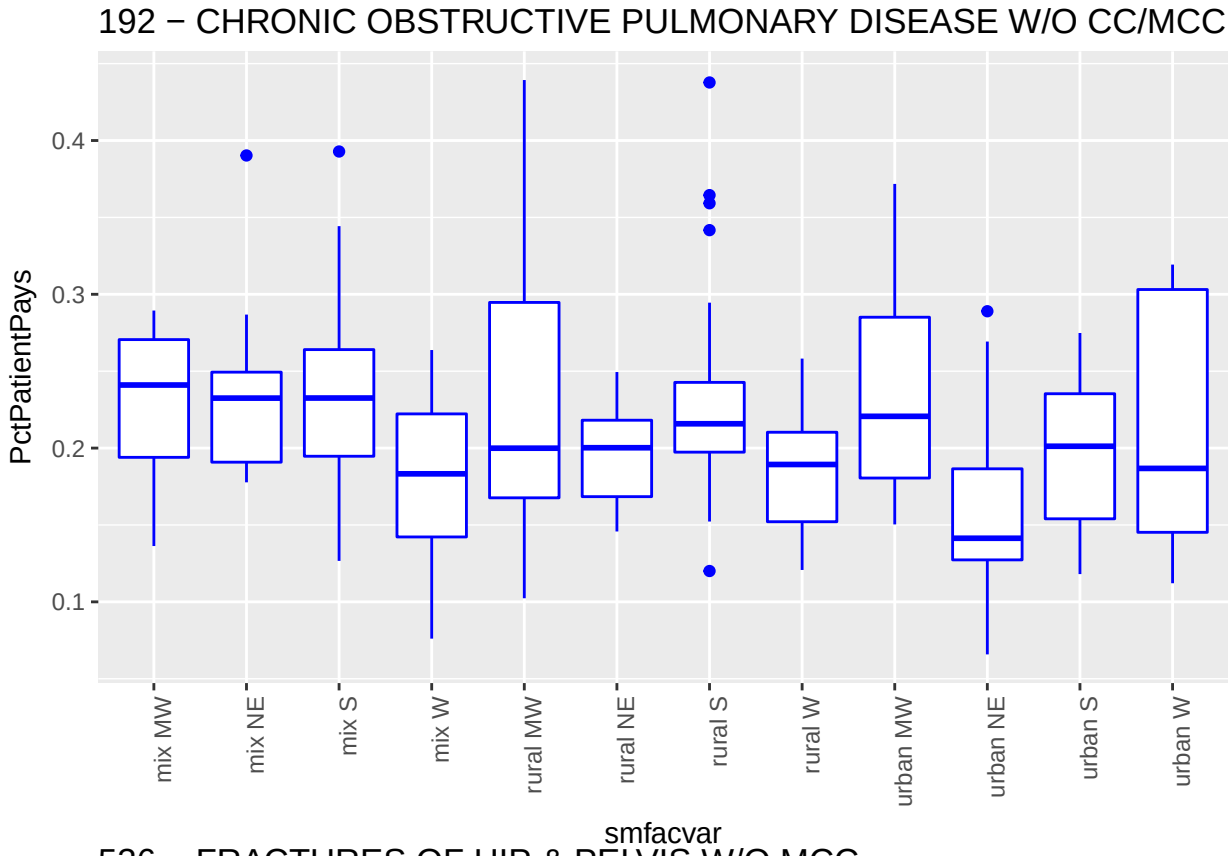
```
#Plot patient pays for each of the diagnoses by urbanByRegions
library(ggplot2)
```

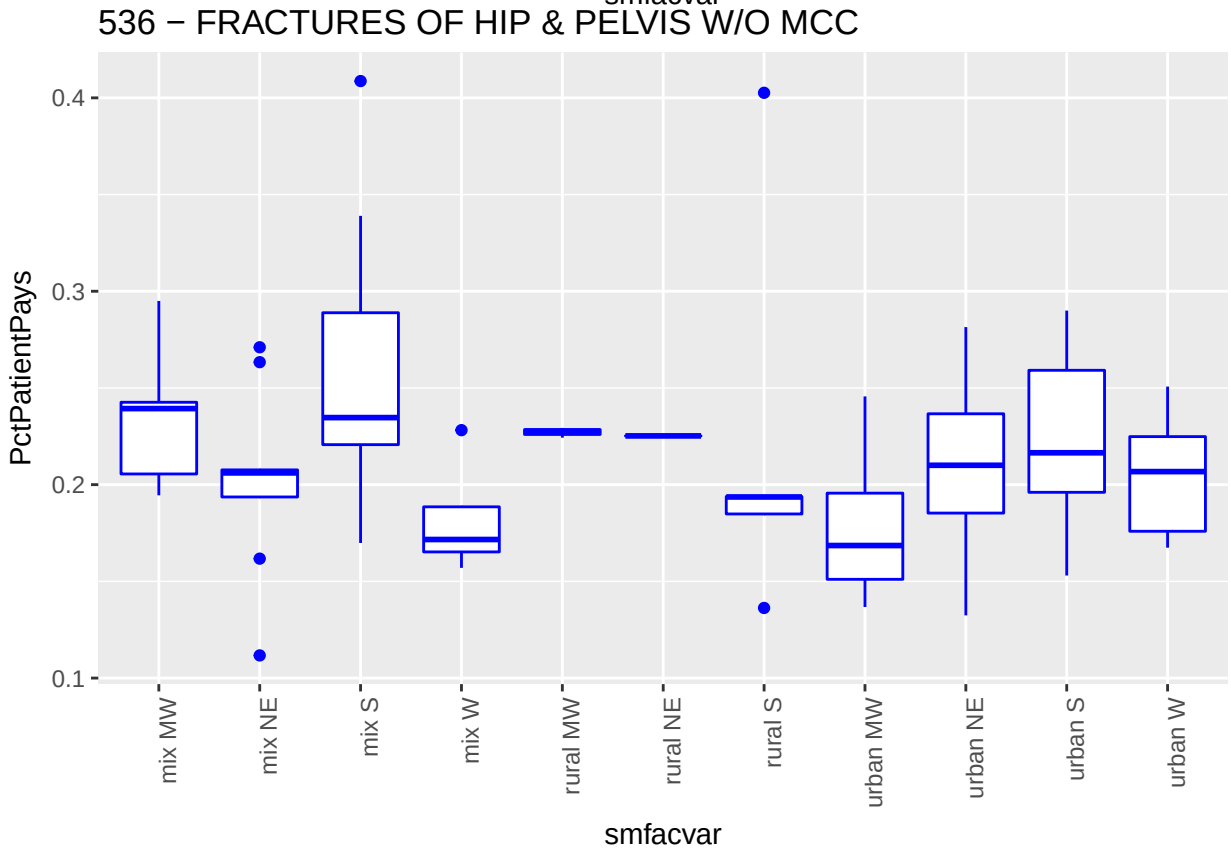
```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
for (i in unique(excluded$DRG.Definition)){
  variable<-excluded[excluded$DRG.Definition==i,]
  print(ggplot(variable,aes(x=smfacvar,y=PctPatientPays))+geom_boxplot(color="blue")+ggtitle(i)+theme(a
}
```


638 − DIABETES W CC
PctPatientPays
0.5
0.4
0.3
0.2
0.1
mix MW
mix NE
mix S
mix W
rural MW
rural NE
rural S
rural W
urban MW
urban NE
urban S
urban W
smfacvar


293 − HEART FAILURE & SHOCK W/O CC/MCC
PctPatientPays
0.4
0.3
0.2
0.1
mix MW
mix NE
mix S
mix W
rural MW
rural NE
rural S
rural W
urban MW
urban NE
urban S
urban W
smfacvar





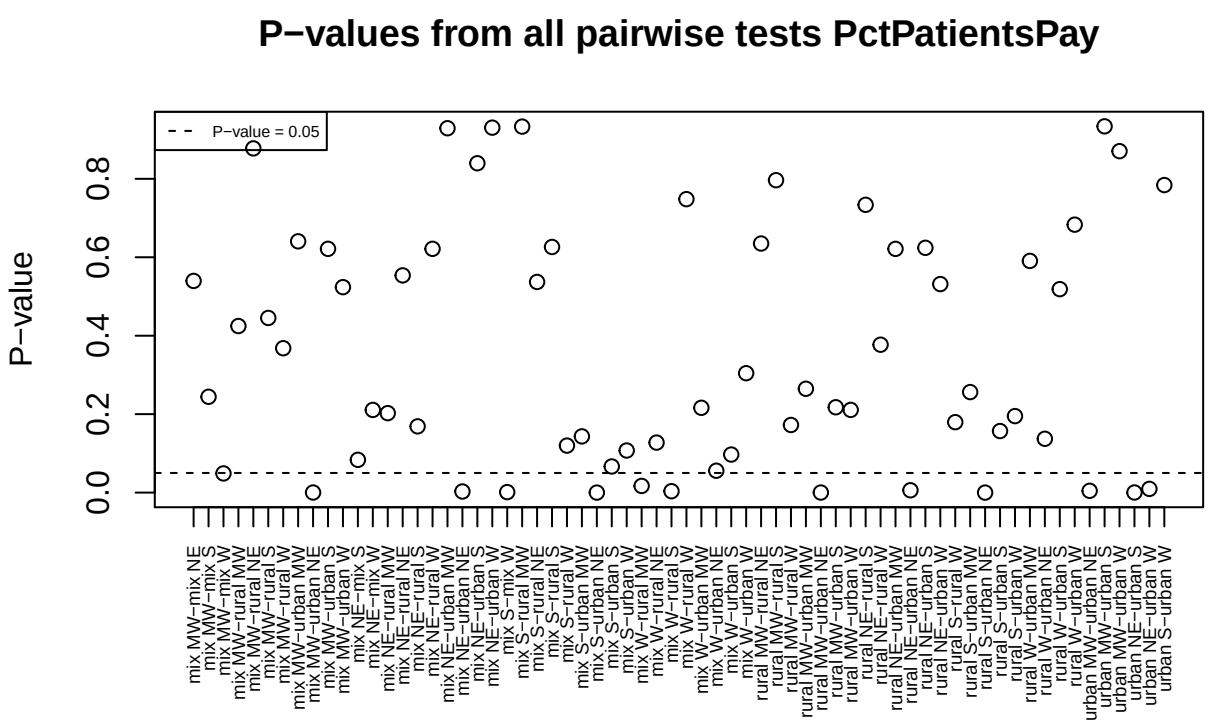
As far as whether or not there should be transformations done on the individual groups, it wouldn't be a

good idea to perform transformations because distributions are not heavily skewed.

Question 4

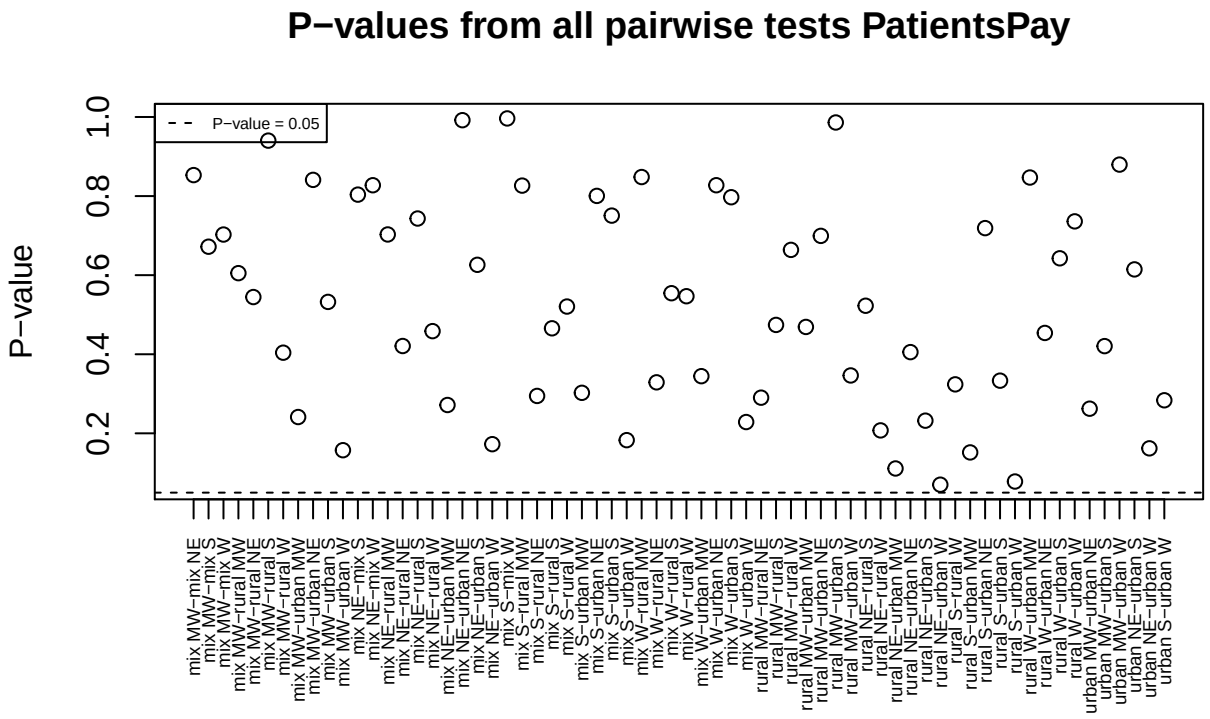
We will proceed to do parametric t-tests because we are working with data that have fairly normal distributions and a data set that is large enough to work with for a parametric t.test.

```
#each pair, do t-test for PctPatientPays
ttestFunPctPP<-function(x,variableName){
    tout<-t.test(excluded$PctPatientPays[excluded[,variableName] == x[1]],excluded$PctPatientPays[exclu
    unlist(tout[c("statistic","p.value")])
    #unlist makes it a vector rather than list
}
#Create pairs and run ttestFun on each pair
factors<-levels(as.factor(excluded$smfacvar))
pairsoffactors<-combn(x=factors,m=2)
t.testPairsPctPP<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestFunPctPP,variableName="smfacvar")

par(mar=c( 8.1,4.1,4.1,1.1))
plot(t.testPairsPctPP["p.value",],main="P-values from all pairwise tests PctPatientsPay",xaxt="n",xlab=
abline(h=0.05,lty=2)
legend("topleft",legend="P-value = 0.05",lty=2,cex=0.5)
colnames(t.testPairsPctPP)<-paste(pairsoffactors[1,],pairsoffactors[2,],sep="-")
axis(1,at=1:ncol(t.testPairsPctPP),labels=colnames(t.testPairsPctPP),las=2,cex.axis=0.6)
```




```
#Repeat the same process for PatientPays, change variable in ttestFun
ttestFunPP<-function(x,variableName){
    tout<-t.test(excluded$PatientPays[excluded[,variableName] == x[1]],excluded$PatientPays[excluded[,va
    unlist(tout[c("statistic","p.value")])

}

factors<-levels(as.factor(excluded$smfacvar))
```

```
pairsoffactors<-combn(x=factors,m=2)
t.testPairsPP<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestFunPP,variableName="smfacvar")

par(mar=c( 8.1,4.1,4.1,1.1))
plot(t.testPairsPP["p.value",],main="P-values from all pairwise tests PatientsPay",xaxt="n",xlab="",ylab
abline(h=0.05,lty=2)
legend("topleft",legend="P-value = 0.05",lty=2,cex=0.5)
colnames(t.testPairsPP)<-paste(pairsoffactors[1,],pairsoffactors[2,],sep="-")
axis(1,at=1:ncol(t.testPairsPP),labels=colnames(t.testPairsPP),las=2,cex.axis=0.6)
```



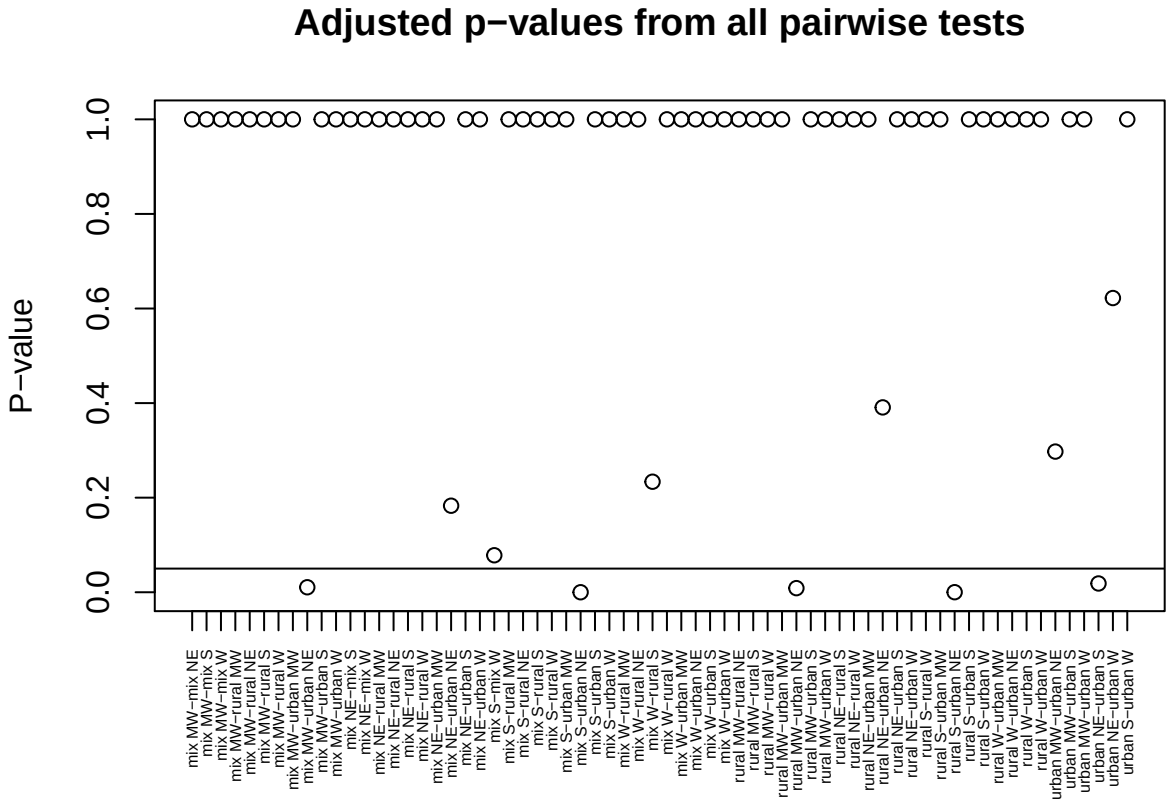
For pctPatientsPay we see that we have a number of significant values with a number of p-values<0.05. However with PatientsPay there are no significant values. Before I move on with analyzing the pctPatientsPay group I will run multiple testing correction to try and correct for any testing error.

```
nfactors<-length(factors)
npairs<-choose(nfactors,2)
cat("Number found significant after Bonferonni (Pct Patient pays): ",sum(t.testPairsPctPP["p.value",]<0
```

```
## Number found significant after Bonferonni (Pct Patient pays):  5
```

Now to plot the adjusted p-values for pctPatientPays by doing bonferonni correction.

```
#Bind the adjusted p.values to testPairsPctPP
t.testPairsPctPP<-rbind(t.testPairsPctPP[1:2,],"p.value.adj"=pmin(t.testPairsPctPP["p.value",]*npairs,1)
plot(t.testPairsPctPP["p.value.adj",],main="Adjusted p-values from all pairwise tests",xaxt="n",xlab=""
abline(h=0.05,las=2)
axis(1,at=1:ncol(t.testPairsPctPP),labels=colnames(t.testPairsPctPP),las=2,cex.axis=0.5)
```



After running through bonferonni correction on the pctPatientPays group we have 5 significant differences left: mix MW-urban NE, mix S-urban NE, rural MW-urban NE, rural S-urban NE, urban NE-urban S. There are very few significant differences for so many pairs, but it is interesting that all 5 are tied to the urban NE. When taking a quick look at the group, it is quickly apparent that the urban NE pctPatientPays average is much lower than the other groups as seen in the violin plot above.

```
#Check for number of significant values after Bonferonni correciton for patient pays

nfactors<-length(factors)
npairs<-choose(nfactors,2)

cat("Number found significant after Bonferonni (Patient pays): ",sum(t.testPairsPP["p.value",]<0.05/npa
```

```
## Number found significant after Bonferonni (Patient pays):  0
```

There were no significant values in the original patient pays group and there is nothing new after the bonferonni correction.

```
#Use similar function as before but extract the estimate and conf.int from the t.test
ttestCI<-function(x,variableName){
    tout<-t.test(excluded$PctPatientPays[excluded[,variableName] == x[1]],excluded$PctPatientPays[exclu
    unlist(tout[c("estimate","conf.int")])
    #unlist makes it a vector rather than list
}
#Create pairs and run ttestCI on each pair
factors<-levels(as.factor(excluded$smfacvar))
pairsoffactors<-combn(x=factors,m=2)
t.CIPairs<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestCI,variableName="smfacvar")
colnames(t.CIPairs)<-paste(pairsoffactors[2,],pairsoffactors[1,],sep="-")
t.CIPairs<-rbind(t.CIPairs,diff=t.CIPairs["estimate.mean of x",]-t.CIPairs["estimate.mean of y",])
```

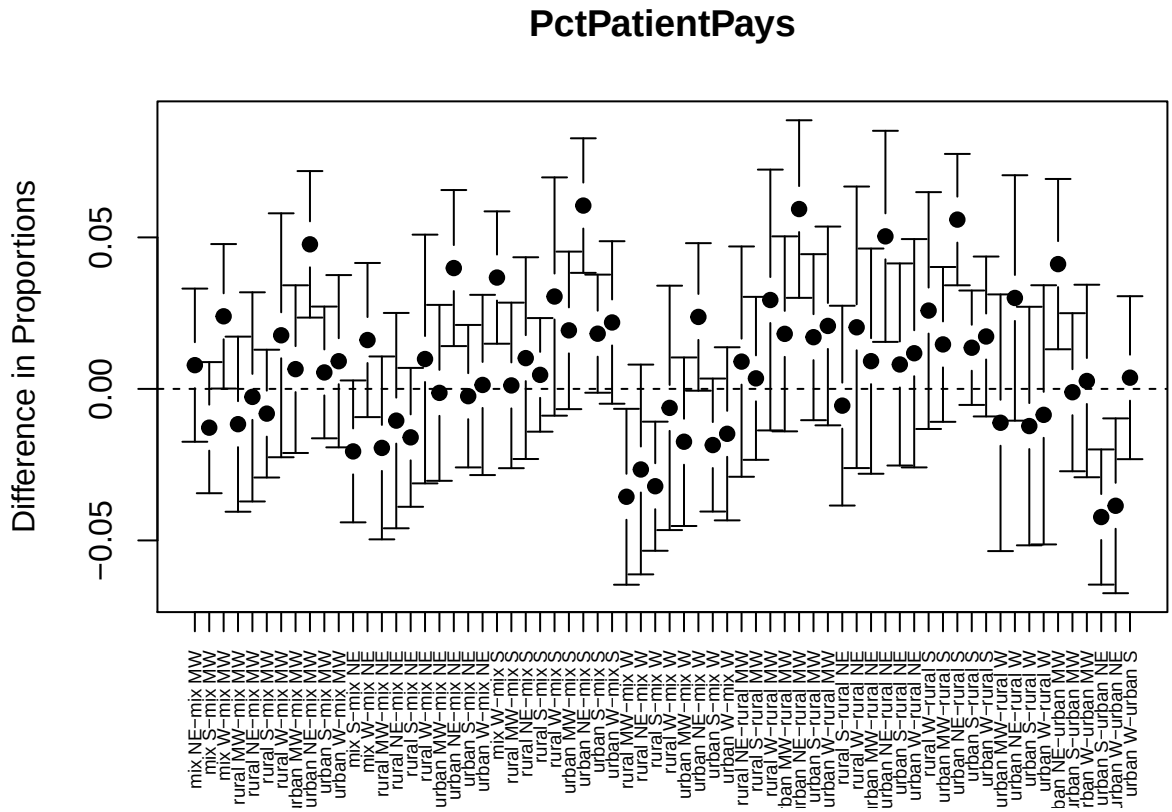
```
#Plot the differences in percentages
require(gplots)
```

```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'
```

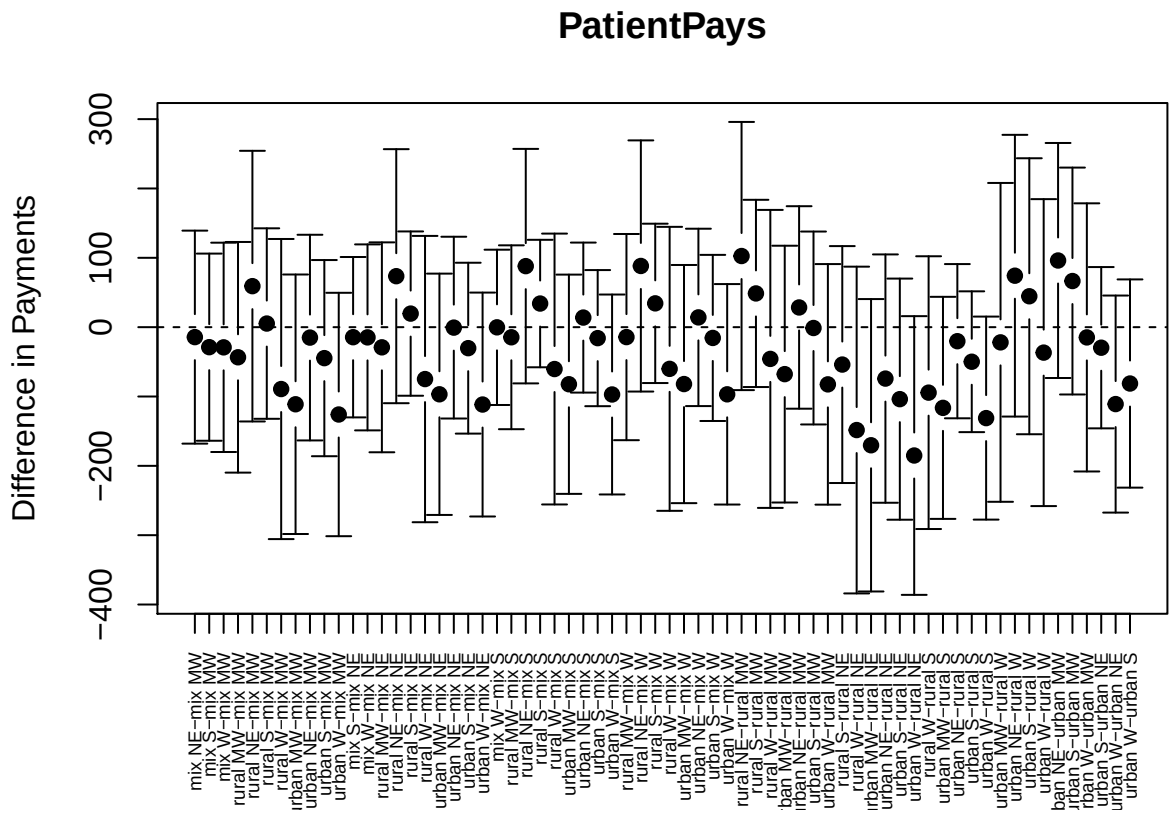
```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
plotCI(x=t.CIPairs["diff",],li=t.CIPairs["conf.int1",],ui=t.CIPairs["conf.int2",],ylab="Difference in P
axis(1,at=1:ncol(t.CIPairs),labels=colnames(t.CIPairs),las=2,cex.axis=0.6)
abline(h=0,lty=2)
```



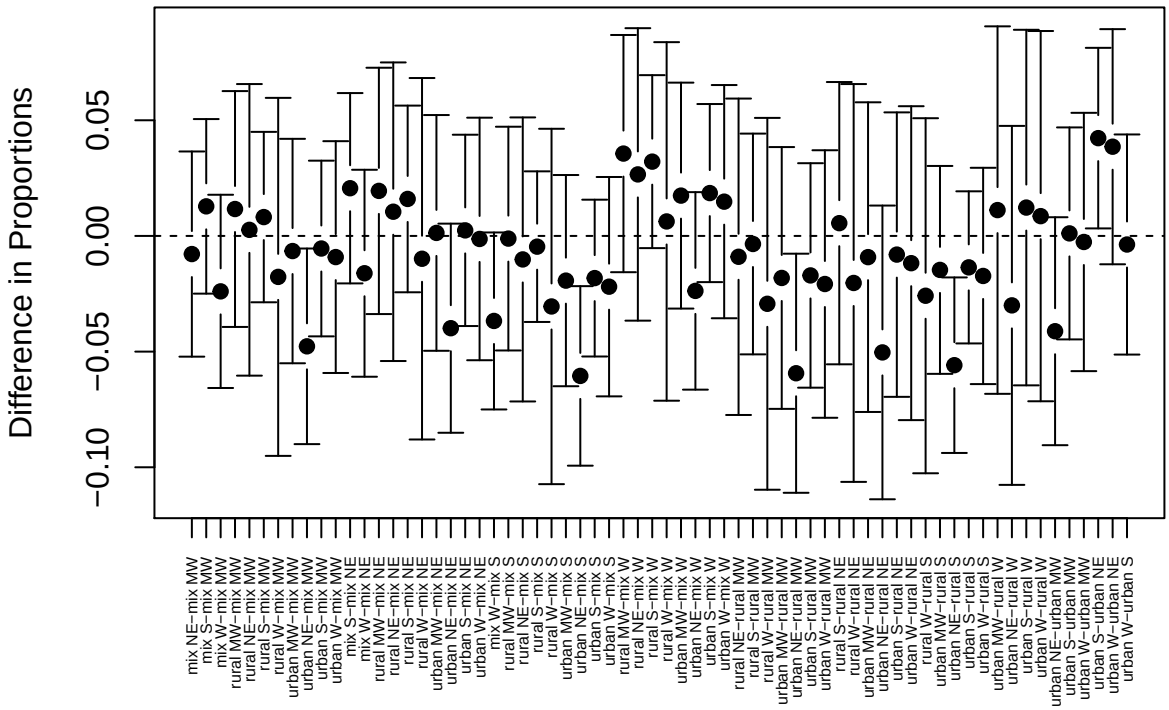

```
#Repeat the same for PatientPays
ttestCI<-function(x,variableName){
    tout<-t.test(excluded$PatientPays[excluded[,variableName] == x[1]],excluded$PatientPays[excluded[,va
    unlist(tout[c("estimate","conf.int")])
    #unlist makes it a vector rather than list
}
#Create pairs and run ttestCI on each pair
factors<-levels(as.factor(excluded$smfacvar))
pairsoffactors<-combn(x=factors,m=2)
t.CIPairs<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestCI,variableName="smfacvar")
colnames(t.CIPairs)<-paste(pairsoffactors[2,],pairsoffactors[1,],sep="-")
t.CIPairs<-rbind(t.CIPairs,diff=t.CIPairs["estimate.mean of x",]-t.CIPairs["estimate.mean of y",])
```

```
require(gplots)
plotCI(x=t.CIPairs["diff",],li=t.CIPairs["conf.int1",],ui=t.CIPairs["conf.int2",],ylab="Difference in Pa
axis(1,at=1:ncol(t.CIPairs),labels=colnames(t.CIPairs),las=2,cex.axis=0.6)
abline(h=0,lty=2)
```



With the confidence intervals for pctPatientPays and PatientPays we can see the same significant values that we saw in the pairwise p-values above.

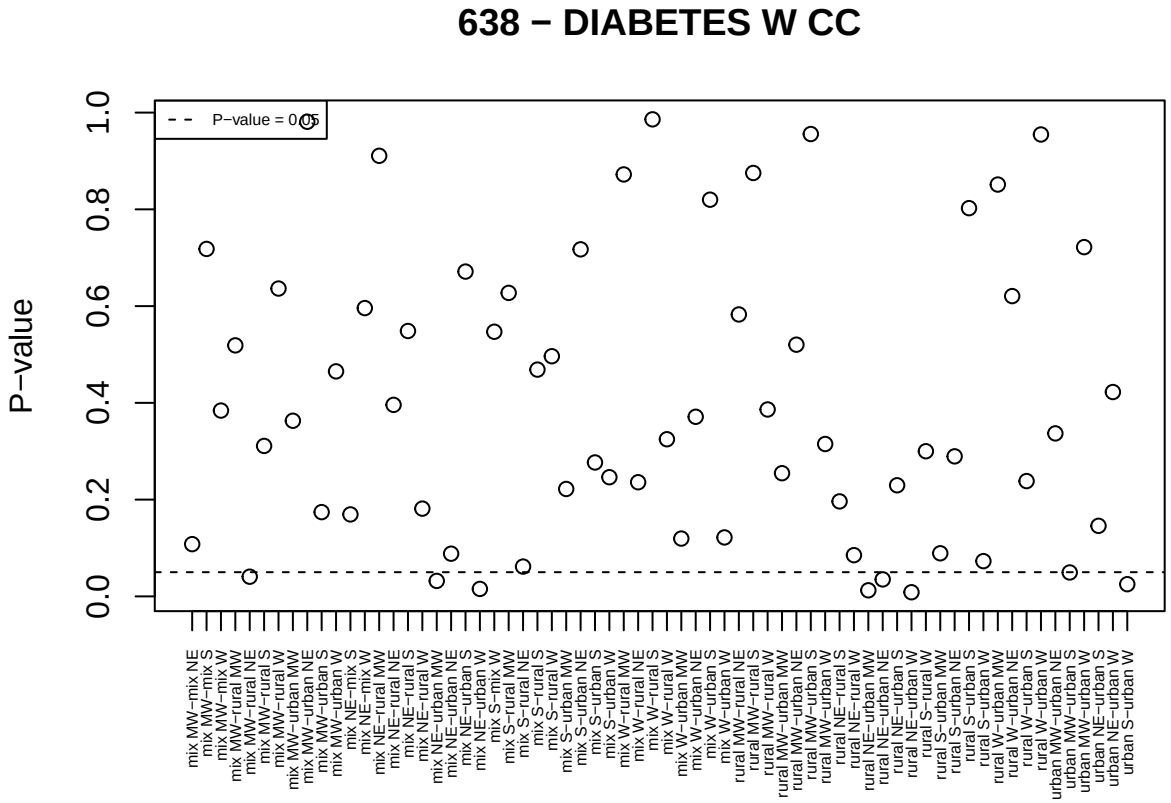
```
#Run code for bonferroni adjusted CI for pctPatientPays
ttestCIAdj<-function(x,variableName){
    tout<-t.test(excluded$PctPatientPays[excluded[,variableName] == x[2]],excluded$PctPatientPays[excluc
    unlist(tout[c("estimate","conf.int")]) #unlist makes it a vector rather than list
}
t.CIPairsAdj<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestCIAdj,variableName="smfacvar")
colnames(t.CIPairsAdj)<-paste(pairsoffactors[2,],pairsoffactors[1,],sep="-")
t.CIPairsAdj<-rbind(t.CIPairsAdj,diff=t.CIPairsAdj["estimate.mean of x",]-t.CIPairsAdj["estimate.mean o
plotCI(x=t.CIPairsAdj["diff",],li=t.CIPairsAdj["conf.int1",],ui=t.CIPairsAdj["conf.int2",],ylab="Differ
axis(1,at=1:ncol(t.CIPairsAdj),labels=colnames(t.CIPairsAdj),las=2,cex.axis=0.5)
abline(h=0,lty=2)
```

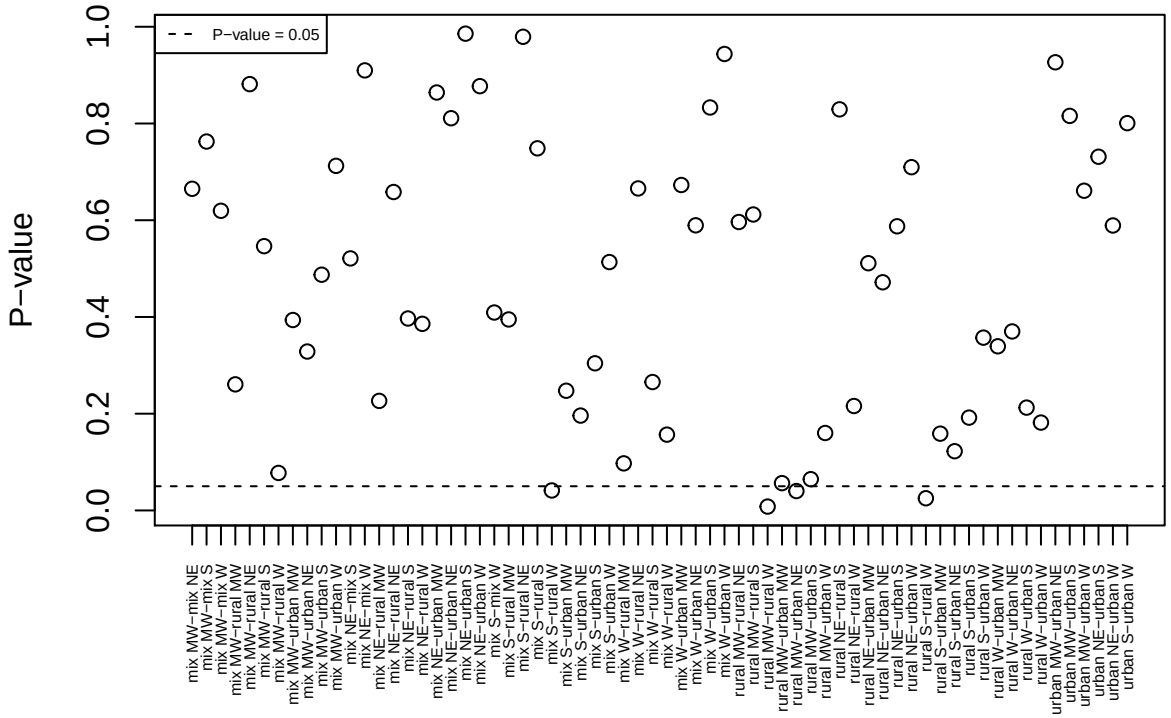


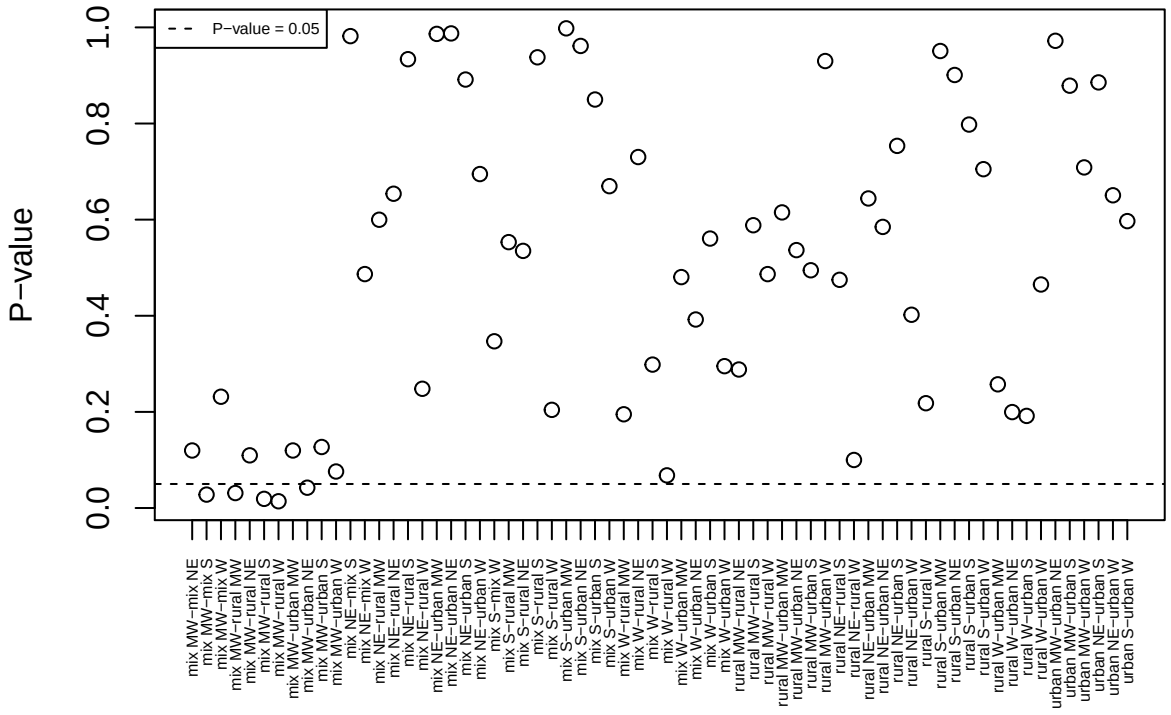
After running through the bonferonni adjusted CIs we see that the urban NE factor is a lower percentage because the first four signficant CIs are negative intervals since the urban NE is less than the other regions. The last pair urban S-urban NE is a positive interval because the urban NE value is much smaller than the urban S value.

While there are not many significant differences, the urban NE group does appear to pay a smaller percentage of their bill than other groups and regions.

```
#Run the pairwise ttest for each of the medical conditions
for (i in unique(excluded$DRG.Definition)[-4]){
  ttestFun<-function(x,variableName){
    tout<-t.test(excluded$PatientPays[excluded[excluded$DRG.Definition==i,variableName] == x[1]],exclud
    unlist(tout[c("statistic","p.value")])
    #unlist makes it a vector rather than list
  }
  #Create pairs and run ttestFun on each pair
factors<-levels(as.factor(excluded$smfacvar))
pairsoffactors<-combn(x=factors,m=2)
t.testPairs<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestFun,variableName="smfacvar")

#Plot p.values
plot(t.testPairs["p.value",],main=i,xaxt="n",xlab="",ylab = "P-value")
abline(h=0.05,lty=2)
legend("topleft",legend="P-value = 0.05",lty=2,cex=0.5)
colnames(t.testPairs)<-paste(pairsoffactors[1,],pairsoffactors[2,],sep="-")
axis(1,at=1:ncol(t.testPairs),labels=colnames(t.testPairs),las=2,cex.axis=0.5)
}
```


638 − DIABETES W CC
P−value
P−value = 0.05
0.0
0.2
0.4
0.6
0.8
1.0
mix MW−mix NE
mix MW−mix S
mix MW−mix W
mix MW−rural MW
mix MW−rural NE
mix MW−rural S
mix MW−rural W
mix MW−urban MW
mix MW−urban NE
mix MW−urban S
mix MW−urban W
mix NE−mix S
mix NE−mix W
mix NE−rural MW
mix NE−rural NE
mix NE−rural S
mix NE−rural W
mix NE−urban MW
mix NE−urban NE
mix NE−urban S
mix NE−urban W
mix S−mix W
mix S−rural MW
mix S−rural NE
mix S−rural S
mix S−rural W
mix S−urban MW
mix S−urban NE
mix S−urban S
mix S−urban W
mix W−rural MW
mix W−rural NE
mix W−rural S
mix W−rural W
mix W−urban MW
mix W−urban NE
mix W−urban S
mix W−urban W
rural MW−rural NE
rural MW−rural S
rural MW−rural W
rural MW−urban NE
rural MW−urban S
rural MW−urban W
rural NE−rural S
rural NE−rural W
rural NE−urban MW
rural NE−urban NE
rural NE−urban S
rural NE−urban W
rural S−rural W
rural S−urban MW
rural S−urban NE
rural S−urban S
rural S−urban W
rural W−urban MW
rural W−urban NE
rural W−urban S
rural W−urban W
urban MW−urban NE
urban MW−urban S
urban MW−urban W
urban NE−urban S
urban NE−urban W
urban S−urban W


293 − HEART FAILURE & SHOCK W/O CC/MCC
P−value
P−value = 0.05
0.0
0.2
0.4
0.6
0.8
1.0
mix MW−mix NE
mix MW−mix S
mix MW−mix W
mix MW−rural MW
mix MW−rural NE
mix MW−rural S
mix MW−rural W
mix MW−urban MW
mix MW−urban NE
mix MW−urban S
mix MW−urban W
mix NE−mix S
mix NE−mix W
mix NE−rural MW
mix NE−rural NE
mix NE−rural S
mix NE−rural W
mix NE−urban MW
mix NE−urban NE
mix NE−urban S
mix NE−urban W
mix S−mix W
mix S−rural MW
mix S−rural NE
mix S−rural S
mix S−rural W
mix S−urban MW
mix S−urban NE
mix S−urban S
mix S−urban W
mix W−rural MW
mix W−rural NE
mix W−rural S
mix W−rural W
mix W−urban MW
mix W−urban NE
mix W−urban S
mix W−urban W
rural MW−rural NE
rural MW−rural S
rural MW−rural W
rural MW−urban NE
rural MW−urban S
rural MW−urban W
rural NE−rural S
rural NE−rural W
rural NE−urban MW
rural NE−urban NE
rural NE−urban S
rural NE−urban W
rural S−rural W
rural S−urban MW
rural S−urban NE
rural S−urban S
rural S−urban W
rural W−urban MW
rural W−urban NE
rural W−urban S
rural W−urban W
urban MW−urban NE
urban MW−urban S
urban MW−urban W
urban NE−urban S
urban NE−urban W
urban S−urban W

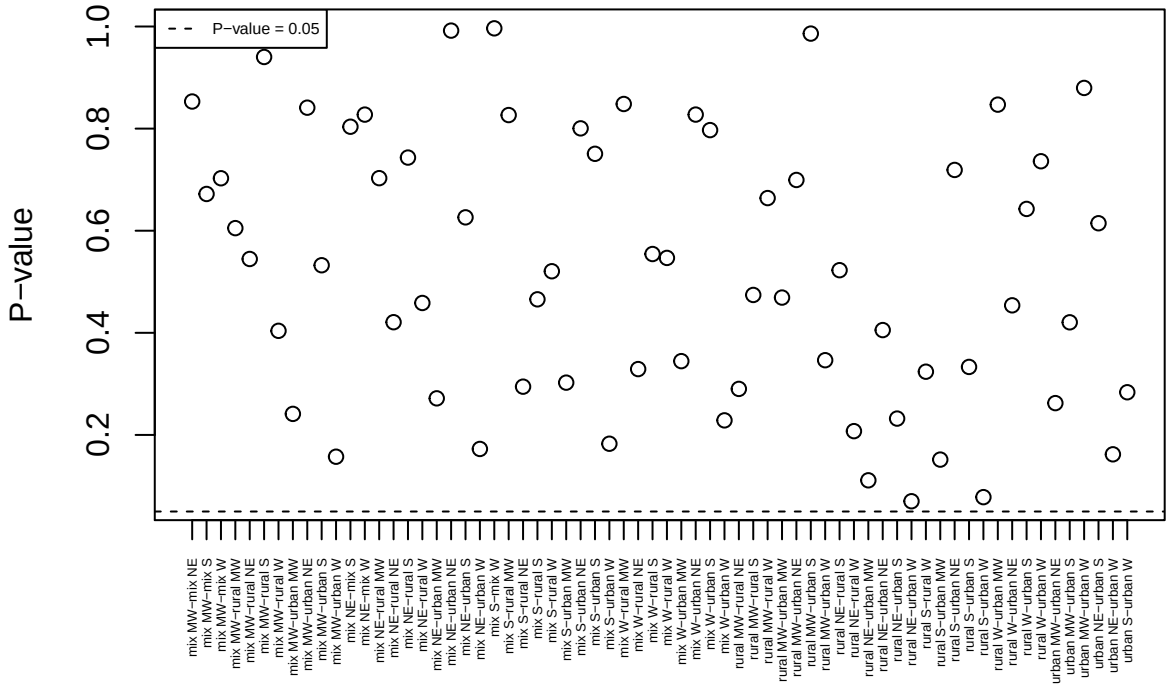


(Was unable to run the function for the values in FRACTURES OF THE HIP group.)

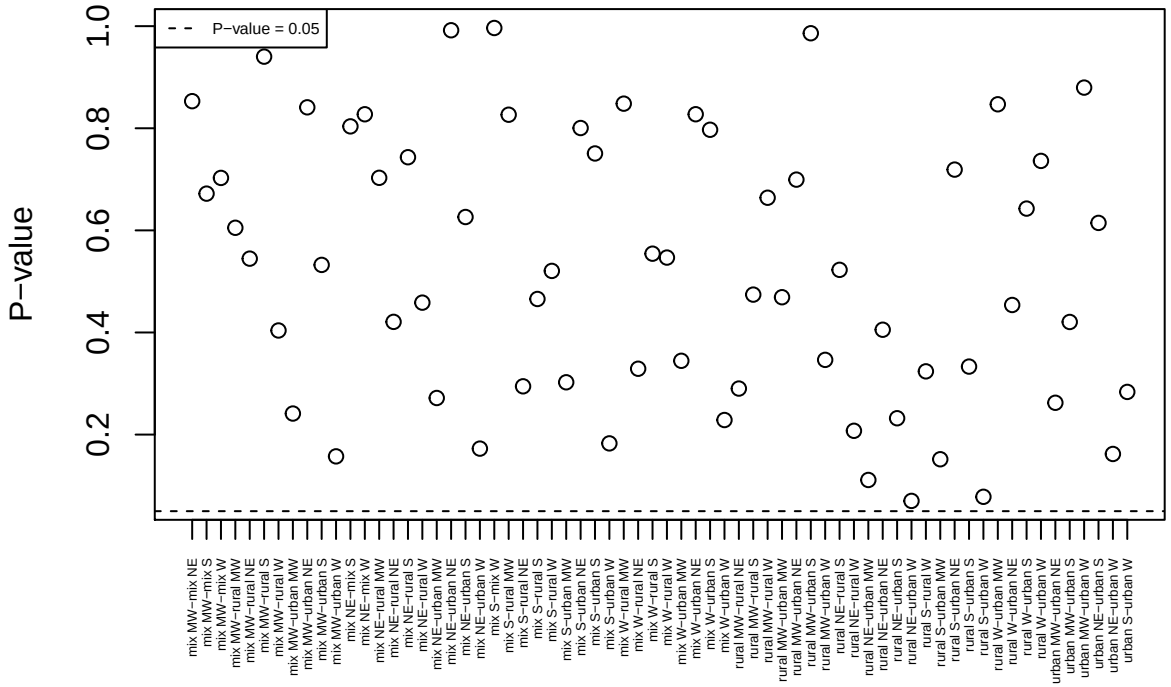
When observing the p-values of the different urbanByRegions for each of the treatments, there are clumps of significant p-values that suggest a serious regional difference in the percent people pay for their treatment. Most notably the mix MW groups are all or nearly all significantly different in the COPD group. Likewise, there are clumps of significant or nearly significant values for the rural MW group in the heart failure patients and diabetes patients.

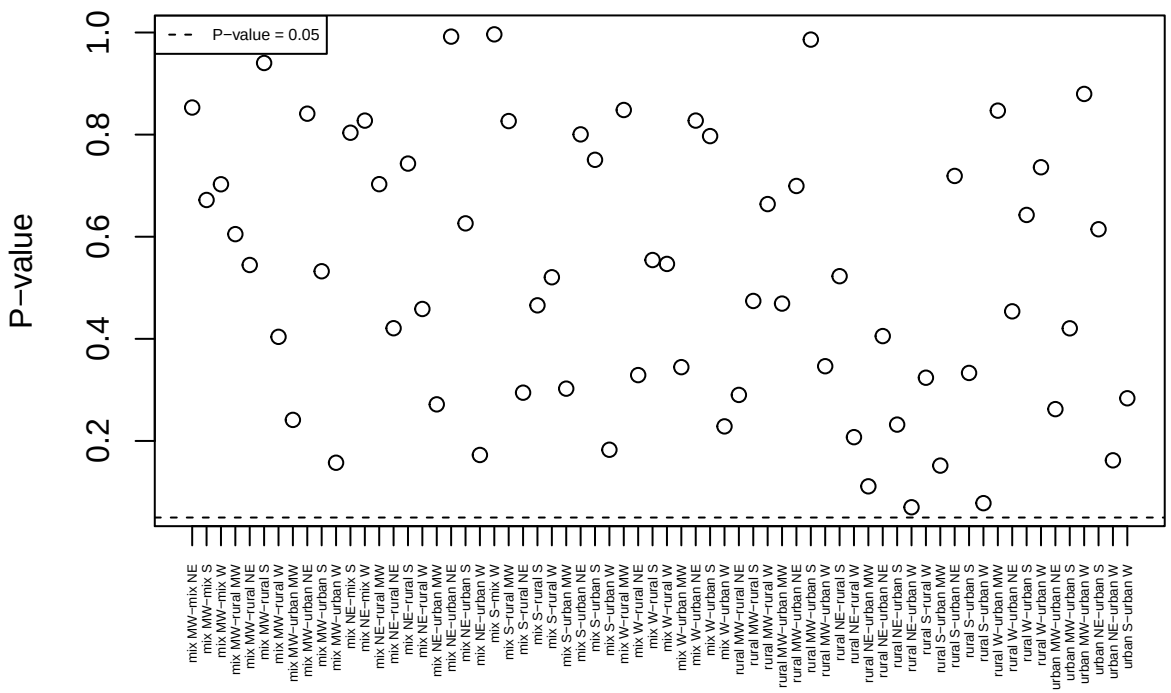
```
for (i in unique(excluded$DRG.Definition)[-4]){
ttestFunPctPP<-function(x,variableName){
    tout<-t.test(excluded$PctPatientPays[excluded[excluded$DRG.Definition==i,variableName] == x[1]],excl
    unlist(tout[c("statistic","p.value")])

}

factors<-levels(as.factor(excluded$smfacvar))
pairsoffactors<-combn(x=factors,m=2)
t.testPairsPctPP<-apply(X=pairsoffactors,MARGIN=2,FUN=ttestFunPP,variableName="smfacvar")

plot(t.testPairsPctPP["p.value",],main=i,xaxt="n",xlab="",ylab = "P-value")
abline(h=0.05,lty=2)
legend("topleft",legend="P-value = 0.05",lty=2,cex=0.5)
colnames(t.testPairsPP)<-paste(pairsoffactors[1,],pairsoffactors[2,],sep="-")
axis(1,at=1:ncol(t.testPairsPP),labels=colnames(t.testPairs),las=2,cex.axis=0.45)
}
```

638 − DIABETES W CC



293 − HEART FAILURE & SHOCK W/O CC/MCC





(Was unable to run function for the values in HIP FRACTURES.)

When comparing these results with the Patient pays group just above, there is little overlap in the significant groups. So it is interesting to note that there are groups that have significant differences in what the patients pay but not in the percent that they pay.

Question 5

After analysis of the amounts patients are paying and the percent patients are paying, there are few significant differences across a large number of different groups. The only significant differences after correcting for the multiple testing were related to the percentage people paid in the urban NE group. People in this group paid a smaller percentage than those in the other groups. As a whole there were few differences in how much those who are receiving medicare aid are having to pay on their treatments. Although treatments do vary in their price, the dollar amount and percent of the total bill people are paying saw little significant difference across the various regions and populations.

For the data itself it would be greatly useful to know more information on the hospitals themselves. Whether or not they are a large institution or not, how they work their charges, how many people does the cummulative hospital cover, or exactly how many people they serve in their general areas. I am also curious about why those in the urban NE would pay less and why those in other areas would not have to make equal payments. It must be something that has to do with medicare law but I am curious as to why this difference exists.

It would also be helpful to change the geographic categorizaiton to see if the grouping is representative or whether or not it would even be useful to eliminate the region factor and just compare by population. There could be something that arises with just urban areas themselves or there are issues with how the different regions are drawn up.

For any future analysis I would first want to know more about the Medicare laws and how they apply for individual people as a basis behind how many of the charges are covered. After that point it would be to look at how the different hospitals determine their charges and whether or not this disproportionately affects less wealthy people in the various areas. I'd also be curious as to whether or not people from lower incomes go under treatment more often and how much their payments hurt them especially considering the hospital charges that vary from different regions.