Week1 Assignment Question 2

December 19, 2016

Question 2

Make a plot (possibly multi-panel) that answers the question: how does the relationship between mean covered charges (Average.Covered.Charges) and mean total payments (Average.Total.Payments) vary by medical condition (DRG.Definition) and the state in which care was received (Provider.State)?

1. Load Data

```
mypath <- "./data/week1_payments.csv"
mydf <- read.csv(mypath)
dim(mydf)
```

```
## [1] 6401   12
```

```
names(mydf)
```

```
##  [1] "DRG.Definition"
##  [2] "Provider.Id"
##  [3] "Provider.Name"
##  [4] "Provider.Street.Address"
##  [5] "Provider.City"
##  [6] "Provider.State"
##  [7] "Provider.Zip.Code"
##  [8] "Hospital.Referral.Region.Description"
##  [9] "Total.Discharges"
## [10] "Average.Covered.Charges"
## [11] "Average.Total.Payments"
## [12] "Average.Medicare.Payments"
```

2. Check Data

```
summary(mydf$DRG.Definition)
```

```
##                                 194 - SIMPLE PNEUMONIA & PLEURISY W CC
##                                                                  1075
##                                      292 - HEART FAILURE & SHOCK W CC
##                                                                  1077
```

```
##             392 - ESOPHAGITIS, GASTROENT & MISC DIGEST DISORDERS W/O MCC
##                                                                    1069
## 641 - MISC DISORDERS OF NUTRITION,METABOLISM,FLUIDS/ELECTROLYTES W/O MCC
##                                                                    1040
##                           690 - KIDNEY & URINARY TRACT INFECTIONS W/O MCC
##                                                                    1087
##                  871 - SEPTICEMIA OR SEVERE SEPSIS W/O MV 96+ HOURS W MCC
##                                                                    1053
```

```
summary(mydf$Provider.State)
```

```
##   CA   FL   IL   NY   PA   TX
## 1514  972  743  933  808 1431
```

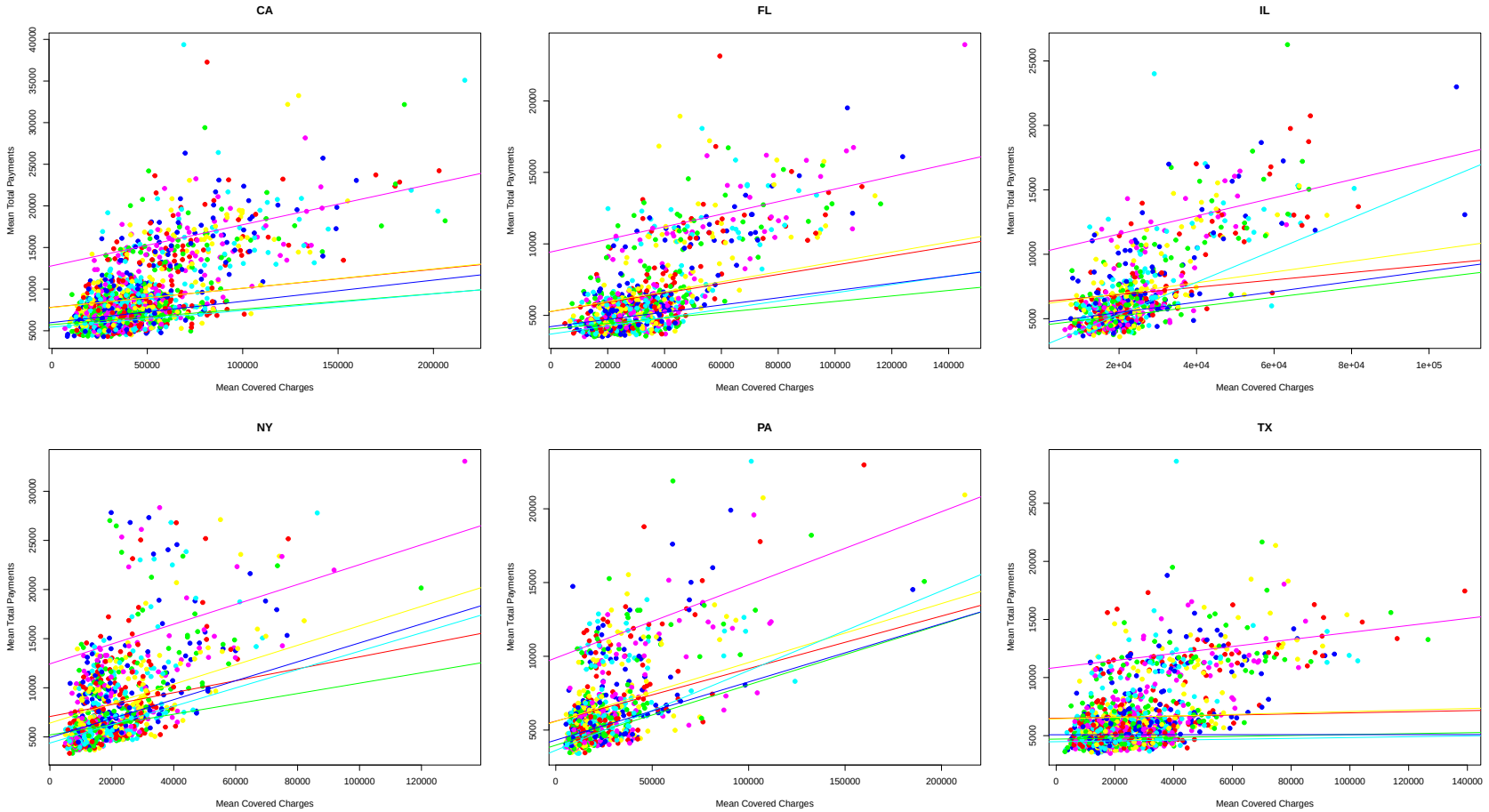
3. Plot

```
state.levels <- levels(mydf$Provider.State)
drg.levels <- levels(mydf$DRG.Definition)
drg.num <- length(drg.levels)
cols <- rainbow(length(drg.levels))

layout(matrix(c(1,2,3,4,5,6,7,7,7), 3, 3, byrow = TRUE))
for (i in state.levels) {
    par(new)
    mydf3 <- mydf[mydf$Provider.State == i, ]
    plot(mydf3$Average.Covered.Charges,
         mydf3$Average.Total.Payments,
         pch = 19,
         col = cols,
         main = i,
         xlab = "Mean Covered Charges",
         ylab = "Mean Total Payments")
    for (j in 1:drg.num) {
        abline(lm(data = mydf3[mydf3$DRG.Definition == drg.levels[j], ],
                  formula = Average.Total.Payments~Average.Covered.Charges),
               col = cols[j])
    }
}

plot.new()
title(main = list("Plot 2: Relationship between Mean Covered Charges and Mean Total Payments
                   Vary by Medical Condition and the State",
                  cex = 4, col = "black", font = 1),
```

```
      line = -4)
legend("center", ncol = 2,
       legend = levels(mydf$DRG.Definition),
       cex = 1.5,
       fill = cols,
       title = "Medical Conditions")
```



Plot 2: Relationship between Mean Covered Charges and Mean Total Payments Vary by Medical Condition and the State