Coursera - Reproducible Research - Plotting practice

Jean-Luc BELLIER

23 juillet 2016

Instructions

The goal of this project is to practice the plotting methods and relationships between variables of datasets.

Question 1

Make a plot that answers the question : *what is the relationship between mean covered charges (Average.Covered.Charges) and mean total payments (Average.Total.Payments) in New-York ?

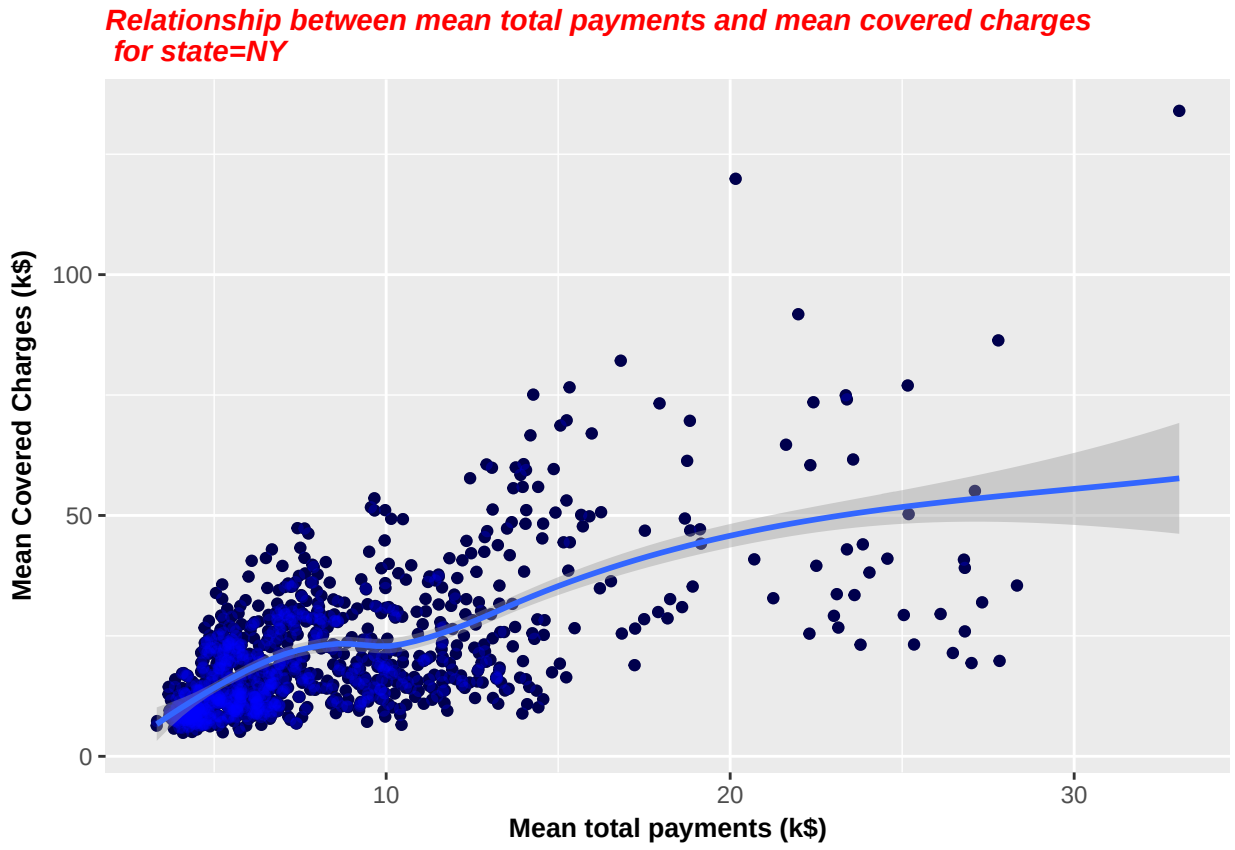
```
paymentsDF <-  read.csv(file="payments.csv",header=TRUE, sep=",")
```

```
library(ggplot2)
```

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

```
paymentsDF_NY <- subset(paymentsDF,Provider.State == "NY")
g <- qplot(Average.Total.Payments/1000, Average.Covered.Charges/1000, data=paymentsDF_NY)
mygraph <- g + geom_point(alpha=0.3, color="blue") + geom_smooth()  + labs(x="Mean total payments (k$)"
mygraph <- mygraph + theme(plot.title = element_text(color="red", size=11, face="bold.italic"), axis.ti
            axis.title.y = element_text(size=10, face="bold"))
print(mygraph)
```

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




```
ggsave("plot1.pdf")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using method = 'loess'
```

Answer : *lots of points are concentrated in the bottom-left corner of the graphic, i.e. for low values. For the mean total payments, most of the values are under 8 k$ whereas for the mean covered charges, most of the values are under 25 k$.I also represented the curve which approches at the best the set of points. The relation between these two values is nearly linear..*

Question 2

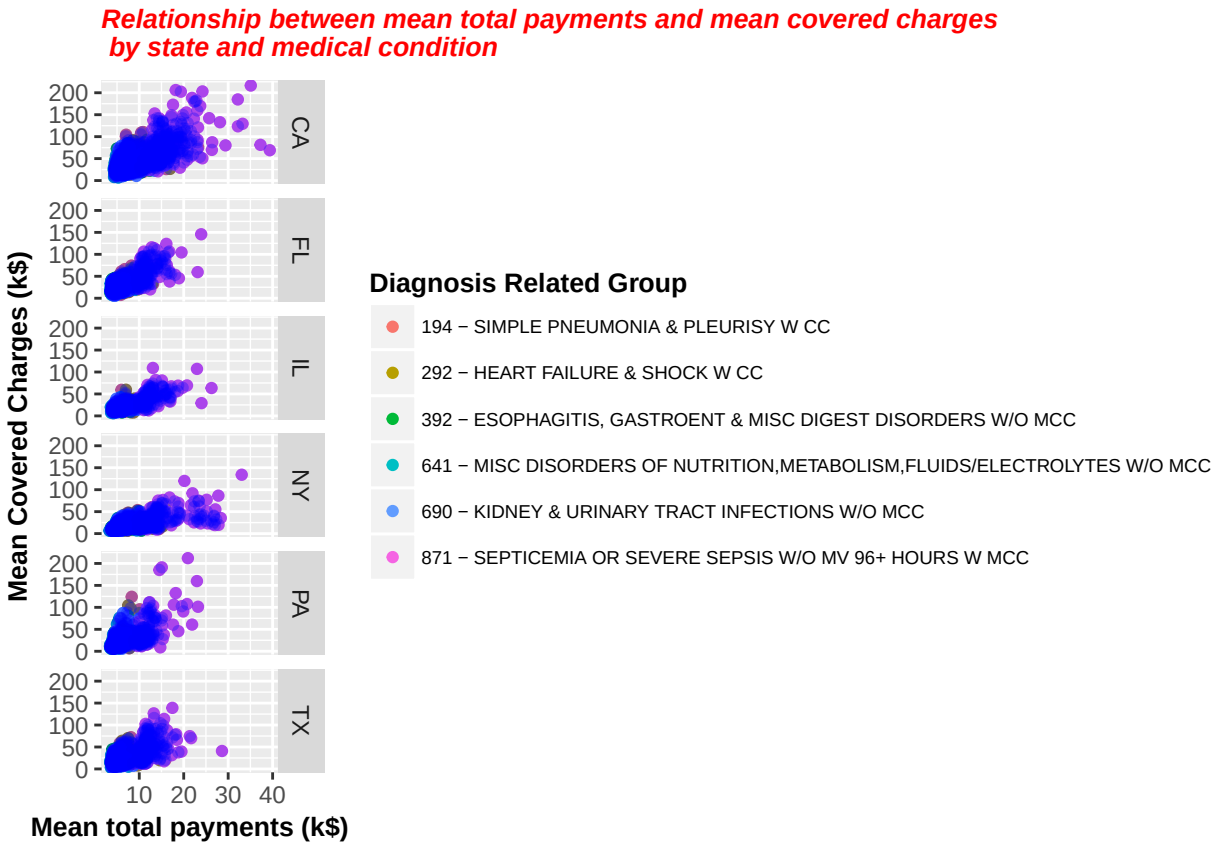
Make a plot (possibly multi-panel) that answers the question : how does the relationship between mean covered charges (Average.Covered.Charges) and mean total payments (Average.Total.Payments) vary by medical condition (DRG.Definition) and the state in which care was received (Provider.State) ?

```
# Creation of a new device with specific size
dev2 = dev.new(width=15, height=8)
# Representation of the variations by DRG Definition. Then I drew one panel by provider state code.
# I preferred this to representing data by state code and put panels on DRG definition because I found
#I also reduced the size of legend items and put appropriate texts on the axes and legend title.
g <- qplot(Average.Total.Payments/1000, Average.Covered.Charges/1000, data=paymentsDF,color=DRG.Definit
mygraph <- g + geom_point(alpha=0.3, color="blue")+ facet_grid(Provider.State~.) + labs(x="Mean total pa
mygraph <- mygraph + theme(plot.title = element_text(color="red", size=10, face="bold.italic"), axis.ti
```

```
            axis.title.y = element_text(size=10, face="bold"), legend.title=element_text(size=10,face="
mygraph <- mygraph + scale_colour_discrete(name  ="Diagnosis Related Group")
dev.set(which = 2)
```

```
## pdf
##   2
```

```
print(mygraph)
```




```
# Save the graph into PDF format
ggsave("plot2.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

```
dev.off(which=2)
```

```
## pdf
##   3
```

Answer : *The full computer code for doing the data analysis is made publicly available.*