3 Data Analysis & Interpretation

3.1 Correlation and Scatterplots

(a)

In [2]:

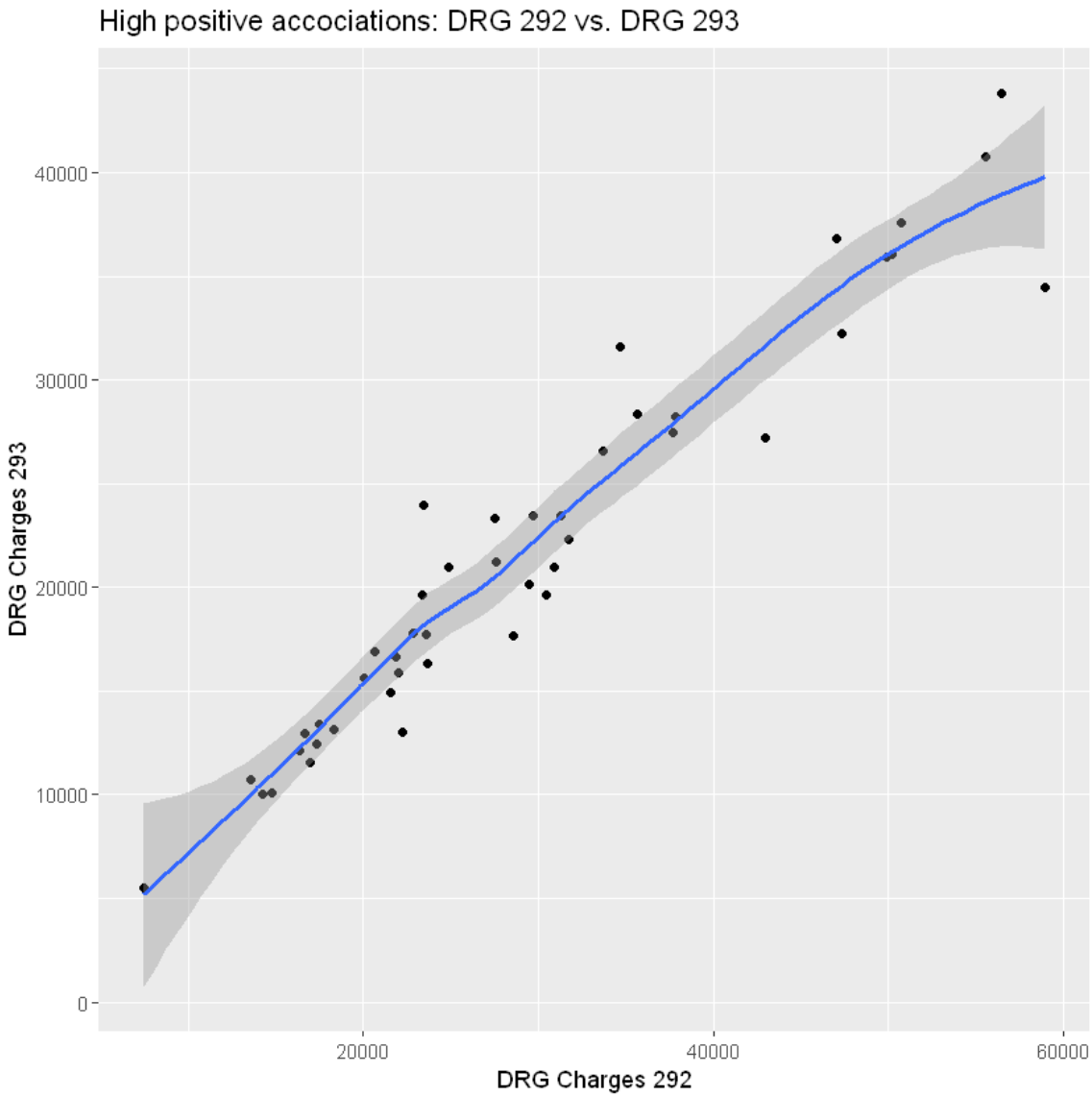
```
# install.packages("ggplot2")
library("ggplot2")
```

In [3]:

```
df = read.csv("100DRG.csv", sep = ',', header = TRUE)
df = na.omit(df, cols = c("df$DRG.Charges.292","DRG.Charges.293","DRG.Charges.481","DRG.Charges.
482",
                          "DRG.Charges.269","DRG.Charges.371","DRG.Charges.315","DRG.Charges.460"
))
```

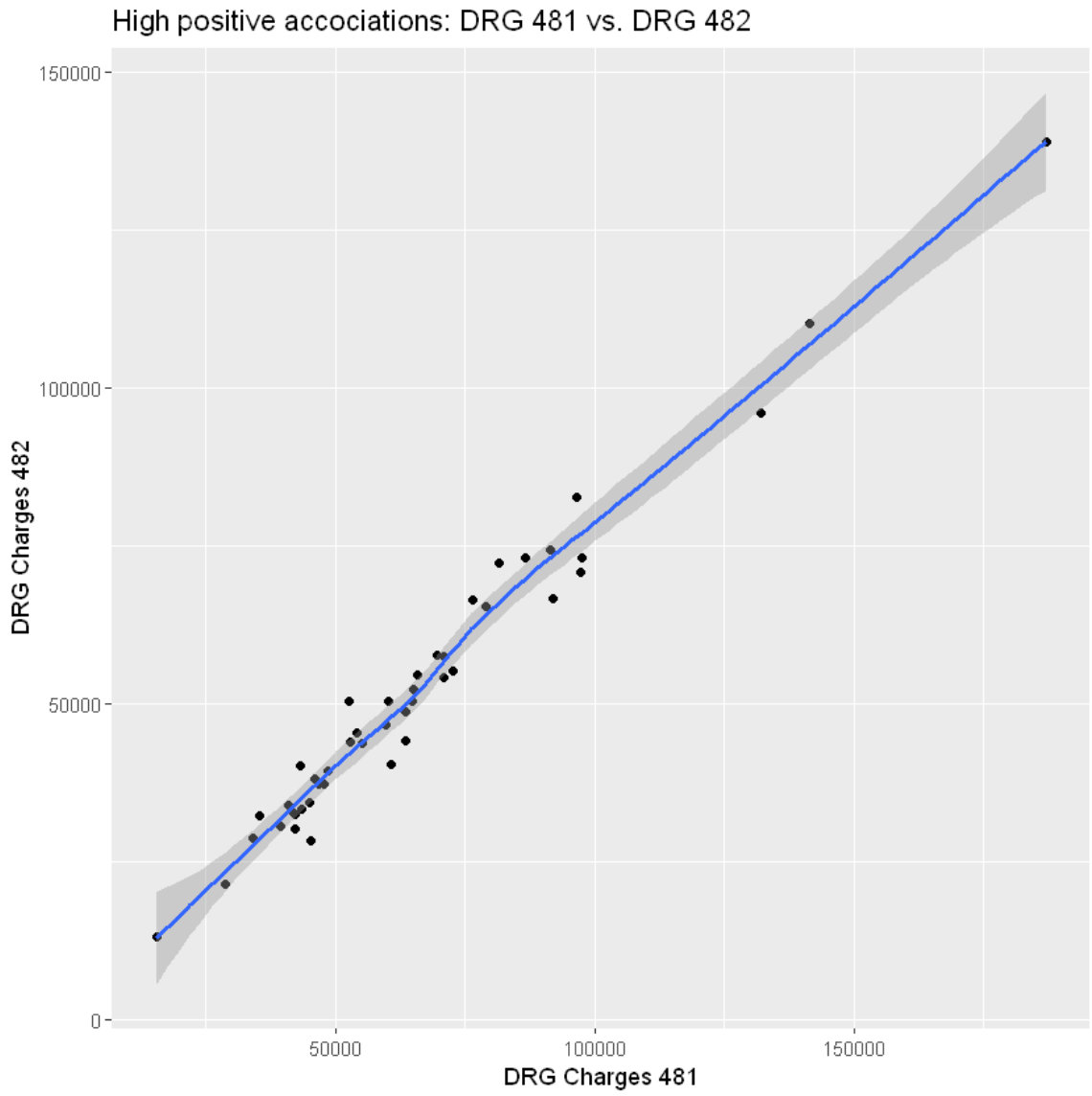
In [4]:

```
# max 1
ggplot(df, aes(x = df$DRG.Charges.292, y = df$DRG.Charges.293)) + geom_point() + geom_smooth(met
hod = 'loess')+
    labs(title="High positive accociations: DRG 292 vs. DRG 293",x="DRG Charges 292", y = "DRG C
harges 293")
```


High positive accociations: DRG 292 vs. DRG 293
DRG Charges 293
DRG Charges 292
0
10000
20000
30000
40000
20000
40000
60000

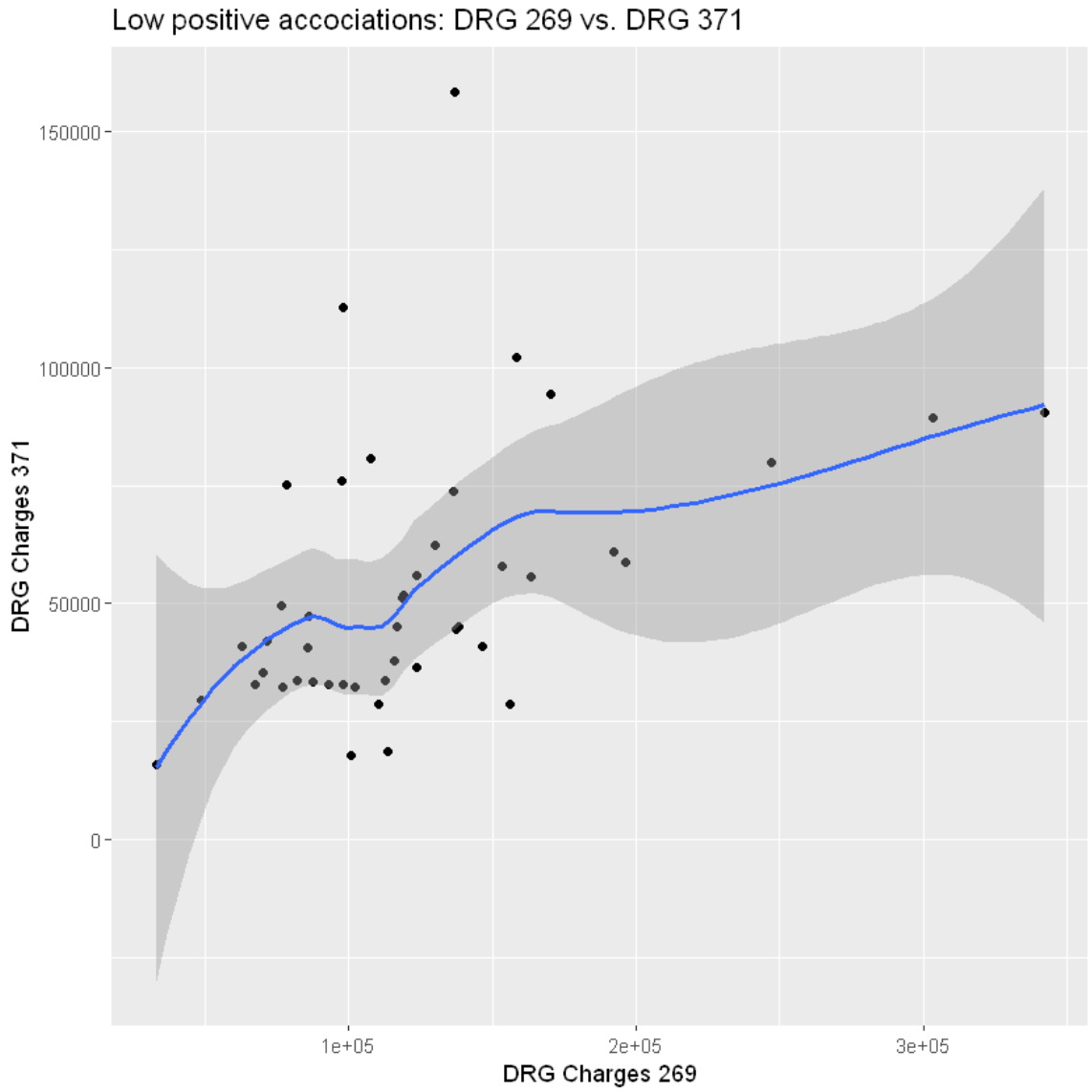
In [5]:

```
# max 2
ggplot(df, aes(x = df$DRG.Charges.481, y = df$DRG.Charges.482)) + geom_point() + geom_smooth(method = 'loess')+
    labs(title="High positive accociations: DRG 481 vs. DRG 482",x="DRG Charges 481", y = "DRG Charges 482")
```



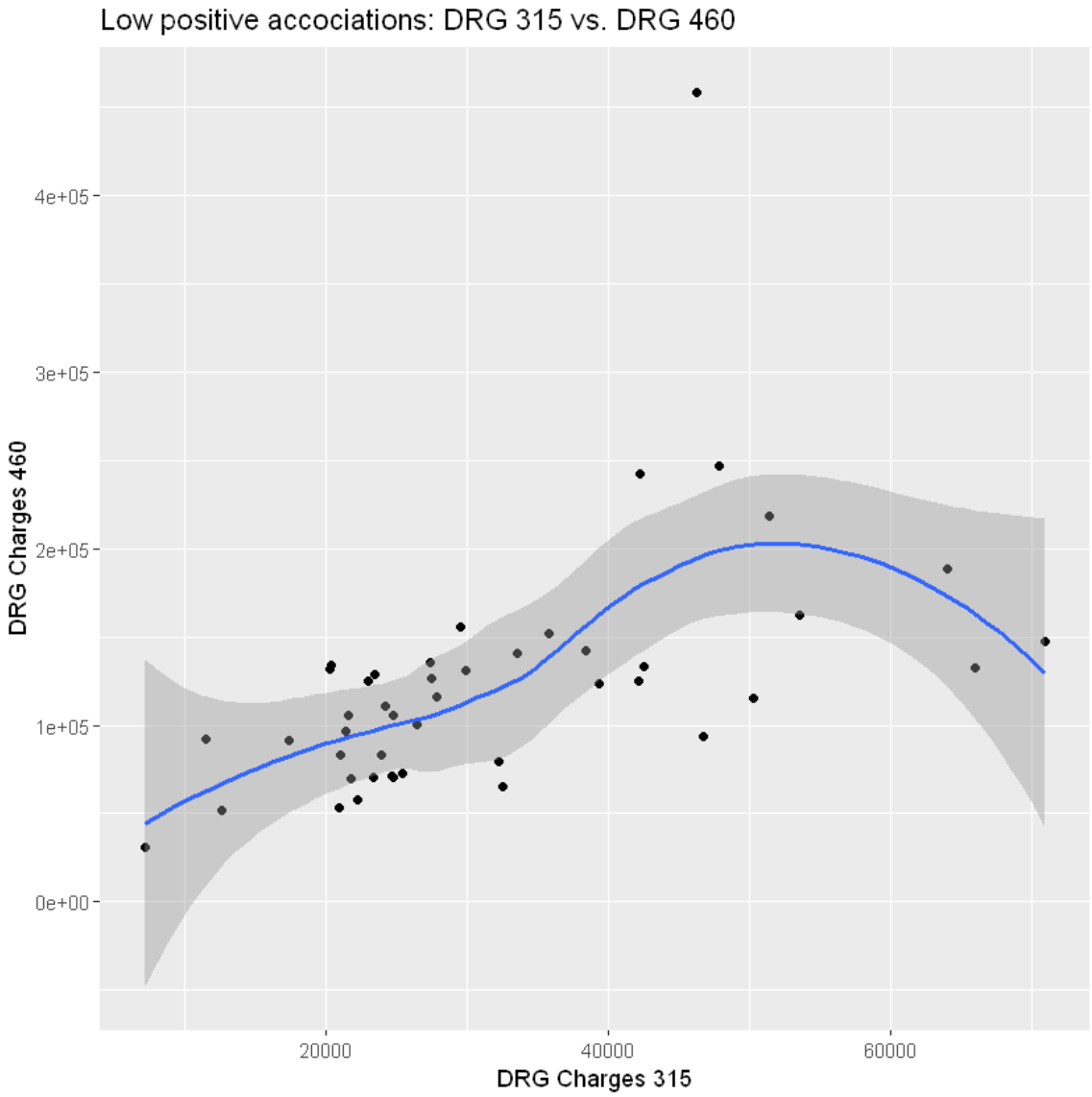
In [7]:

```
# min 1
ggplot(df, aes(x = df$DRG.Charges.269, y = df$DRG.Charges.371)) + geom_point() + geom_smooth(method = 'loess')+
    labs(title="Low positive accociations: DRG 269 vs. DRG 371",x="DRG Charges 269", y = "DRG Charges 371")
```



In [8]:

```
# min2
ggplot(df, aes(x = df$DRG.Charges.315, y = df$DRG.Charges.460)) + geom_point() + geom_smooth(method = 'loess')+
    labs(title="Low positive accociations: DRG 315 vs. DRG 460",x="DRG Charges 315", y = "DRG Charges 460")
```



The observed relations are expeceted, given the DRG category names.
We can easily find out that pairs with high positive associations are basiclly with coherent DRG category numbers, such like 292 and 293. The coherent DRG category numbers means these pairs are more likely to come from a same region or have the same provider. While hospital pairs with low associations are usually come from different district and have different provider. Hence, So their DRG numbers are usually very far apart.