HW8_GUO

Qing Guo

10/29/2019

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

Problem 2

```
data<-read.csv("EdStatsData.csv",header=T)
dim(data)
```

```
## [1] 886930     70
```

```
data0<-data[,1:49]
newdata<-na.omit(data0)
dim(newdata)
```

```
## [1] 13294    49
```

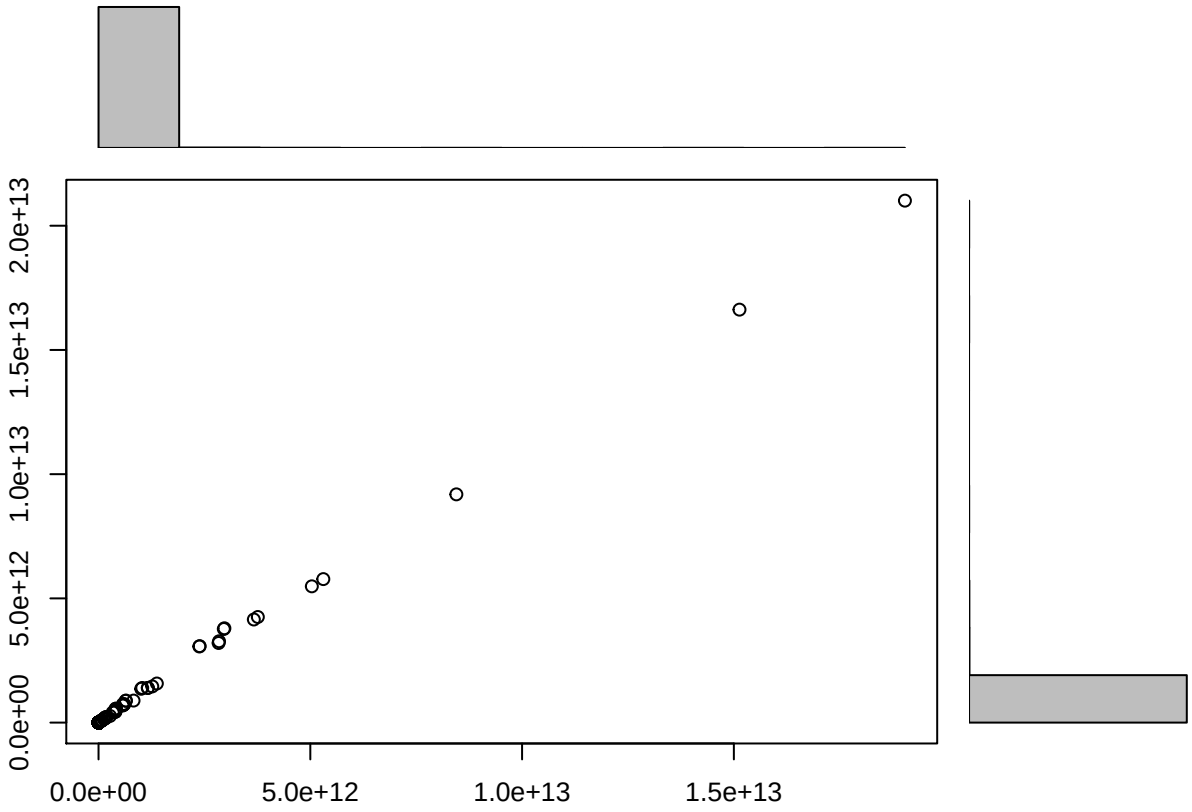
```
data1<-data[which(data$Country.Name == "Arab World"), ]
data2<-data[which(data$Country.Name == "East Asia & Pacific"), ]
data11<-newdata[which(data$Country.Name == "Arab World"), ]
data22<-newdata[which(data$Country.Name == "East Asia & Pacific"), ]
Number_in_Arab<-length(na.omit(data11$Indicator.Name))
Number_in_Pacific<-length(na.omit(data22$Indicator.Name))
tab<-matrix(c(Number_in_Arab,Number_in_Pacific), ncol = 2)
colnames(tab)<- c("Arab", "Pacific")
rownames(tab)<-c("number")
tab
```

```
##        Arab Pacific
## number 3665    3665
```

Problem 3

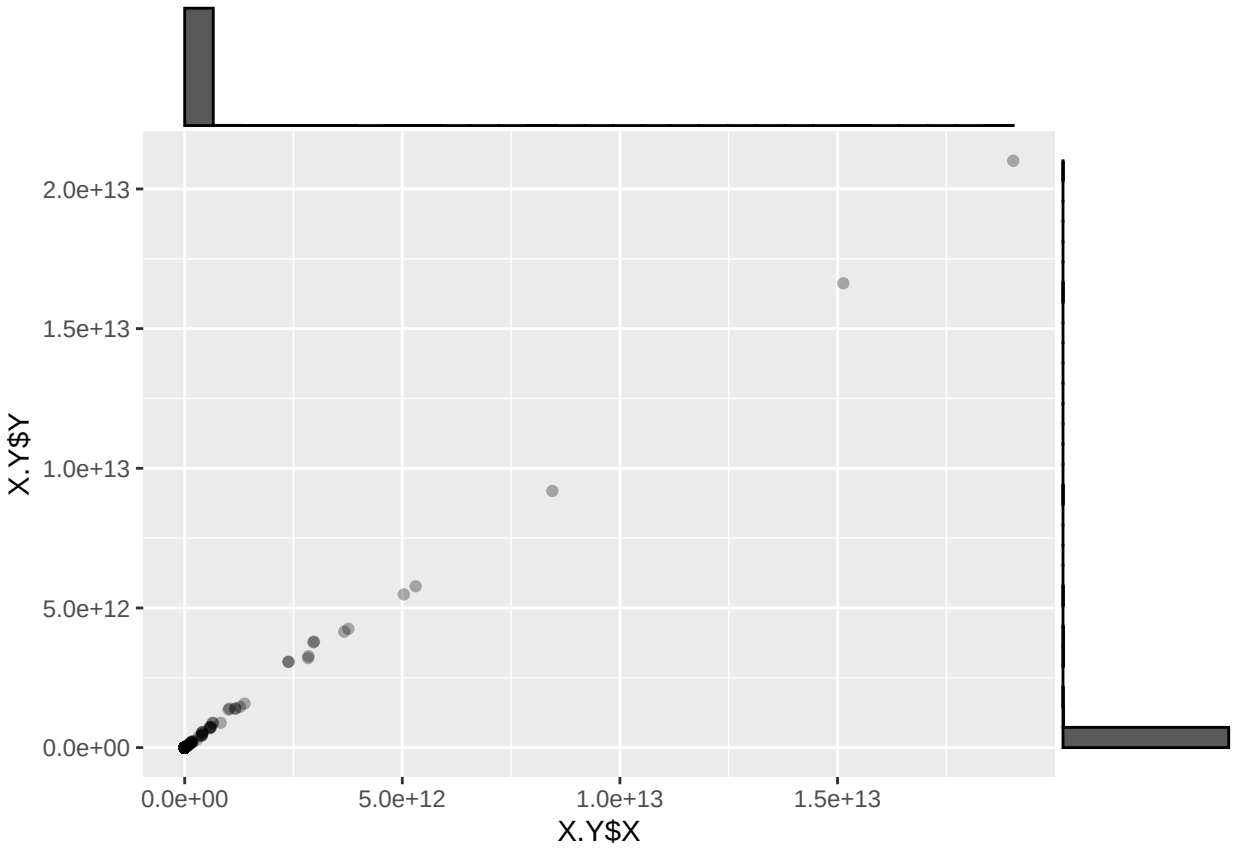
```
##Problem 3
X<-data11$X1970
Y<-data11$X1972
scatterhist = function(x, y, xlab="", ylab=""){
  zones=matrix(c(2,0,1,3), ncol=2, byrow=TRUE)
  layout(zones, widths=c(4/5,1/5), heights=c(1/5,4/5))
  xhist = hist(x, plot=FALSE)
  yhist = hist(y, plot=FALSE)
  top = max(c(xhist$counts, yhist$counts))
  par(mar=c(3,3,1,1))
  plot(x,y)
  par(mar=c(0,3,1,1))
```

```
  barplot(xhist$counts, axes=FALSE, ylim=c(0, top), space=0)
  par(mar=c(3,0,1,1))
  barplot(yhist$counts, axes=FALSE, xlim=c(0, top), space=0, horiz=TRUE)
  par(mar=c(3,0,0,0))
  mtext(xlab, side=1, line=1, outer=TRUE, adj=0,
        at=.8 * (mean(x) - min(x))/(max(x)-min(x)))
  mtext(ylab, side=2, line=1, outer=TRUE, adj=0,
        at=(.8 * (mean(y) - min(y))/(max(y) - min(y))))
}
scatterhist(X,Y, xlab = "X", ylab = "Y")
```



Problem 4

```
X.Y<-data.frame(X,Y)
p<-ggplot(aes(x=X.Y$X,y=X.Y$Y),data=X.Y) +geom_point(alpha = .3)
ggExtra::ggMarginal(p, type = "histogram")
```


2.0e+13
1.5e+13
1.0e+13
5.0e+12
0.0e+00
X.Y$Y
0.0e+00
5.0e+12
1.0e+13
1.5e+13
X.Y$X