HW8_Jiayi_Lian

Jiayi Lian

11/5/2019

```
path<- "/Users/lianjiayi/Documents/STAT_5014_2019_-906275604-/Edstats_csv/EdStatsData.csv"
edu <- read.csv(path)
```

P3

```
edu_none_miss<-edu %>% filter(!is.na(X1970)) %>% filter(!is.na(X1971)) %>% filter(!is.na(X1972)) %>% fil
dim(edu)
```

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

```
dim(edu_none_miss)
```

```
## [1] 13294    70
```

Before cleanning all missing values, there are 886930 x 45 observations. After cleanning, only 13294 x 45 observations left.

```
# Number of indicators left for two countries
Number_in_ARB<-edu_none_miss %>% filter(as.character(Country.Code) == 'ARB' ) %>% select(Indicator.Name
Number_in_EAS<-edu_none_miss %>% filter(as.character(Country.Code) == 'EAS' ) %>% select(Indicator.Name
tab<-matrix(c(Number_in_ARB,Number_in_EAS), ncol = 2)
colnames(tab)<- c("ARB", "EAS")
rownames(tab)<-c("number")
tab
```

```
##        ARB EAS
## number 143 86
```

P4

```
# I would extract data of "Enrolement in secondary vocational, both sexes (number)" and "Enrolment in s
X.Y<-edu_none_miss %>% filter(Country.Code == "ARB") %>% filter(Indicator.Code =="SE.SEC.ENRL.VO" | Ind
X.Y<-matrix(as.numeric(t(X.Y)[5:49,]),ncol=2)
X<-X.Y[,1]
Y<-X.Y[,2]
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")
```
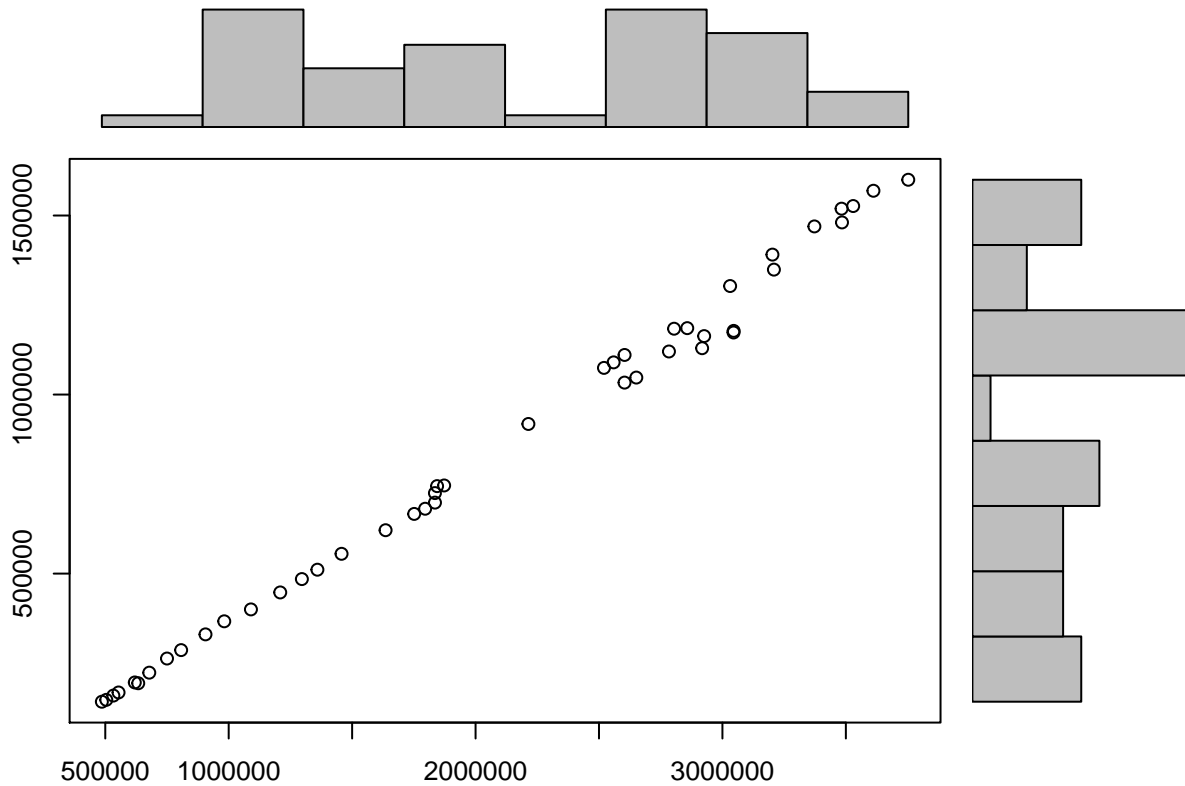

P5

```
X.Y<-data.frame(X=X.Y[,1],X=X.Y[,2])
p<-X.Y %>% ggplot(aes(x=X,y=Y)) +geom_point(alpha = .3)
ggExtra::ggMarginal(p, type = "histogram")
```