HWXX_GUO_QING

Qing Guo

10/2/2019

Problem 3

```
#a.
sucess_p<-function(x){
  n<-sum(x == 1)
  p<-n/length(x)
  return(p)
}
#b.
    set.seed(12345)
    P4b_data <- matrix(rbinom(10, 1, prob = (30:40)/100), nrow = 10, ncol = 10, byrow = FALSE)
#c.
apply(P4b_data,2,FUN=mean)
```

```
##  [1] 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6
```

```
apply(P4b_data,1,FUN=mean)
```

```
##  [1] 1 1 1 1 0 0 0 0 1 1
```

I find the mean of column are all the same and the mean of row are 0 or 1

d.

```
output<-function(p,n){
  n1<-10
  success<-n1*p
  n2<-n1-success
  out1<-c(rep(1,success),rep(0,n2))
  out<-rep(out1,n/n1)
  return(out)
}
data<-matrix(output(0.6,100),nrow=10,ncol=10,byrow=FALSE)
apply(data,2,mean)
```

```
##  [1] 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6
```

```
apply(data,1,mean)
```

```
##  [1] 1 1 1 1 1 1 0 0 0 0
```

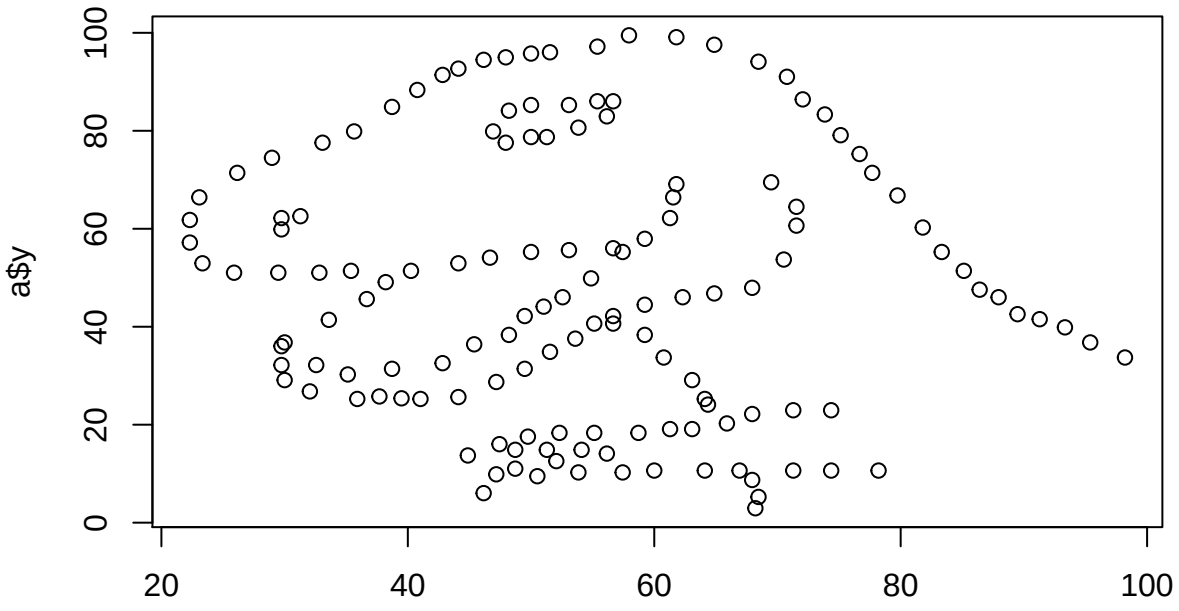
##Problem 4 1.

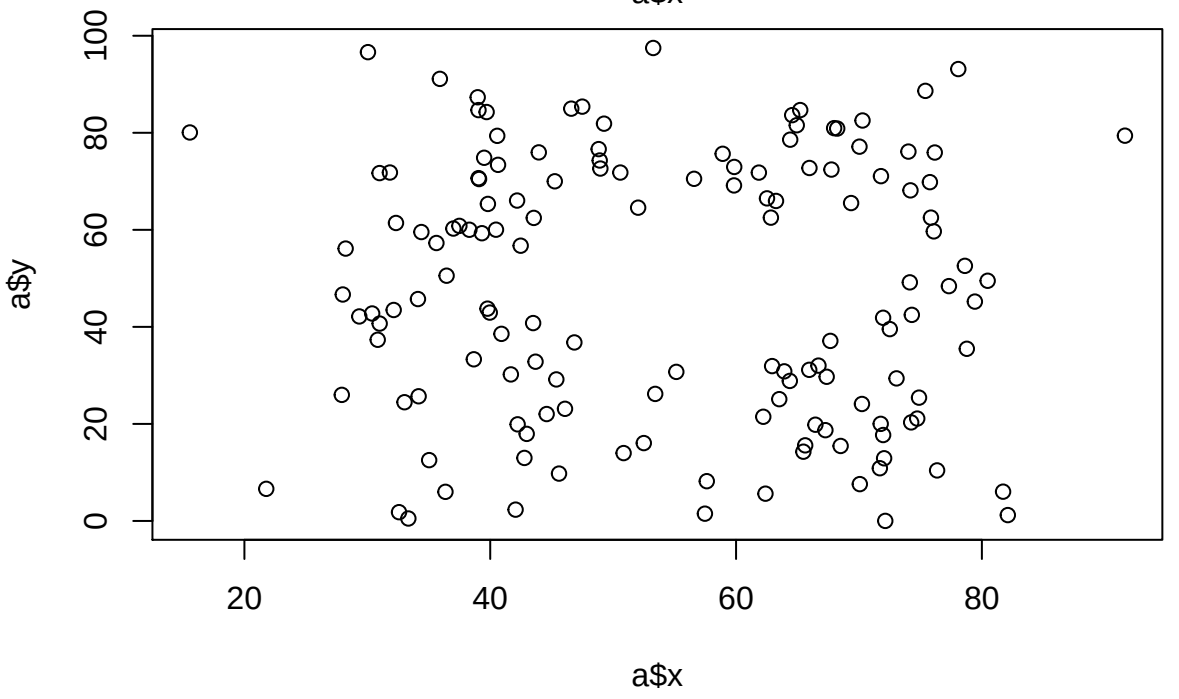
```
b<-readRDS("HW4_data.rds")
b$x<-b$dev1
b$y<-b$dev2
b<-b[,-(2:3)]
```

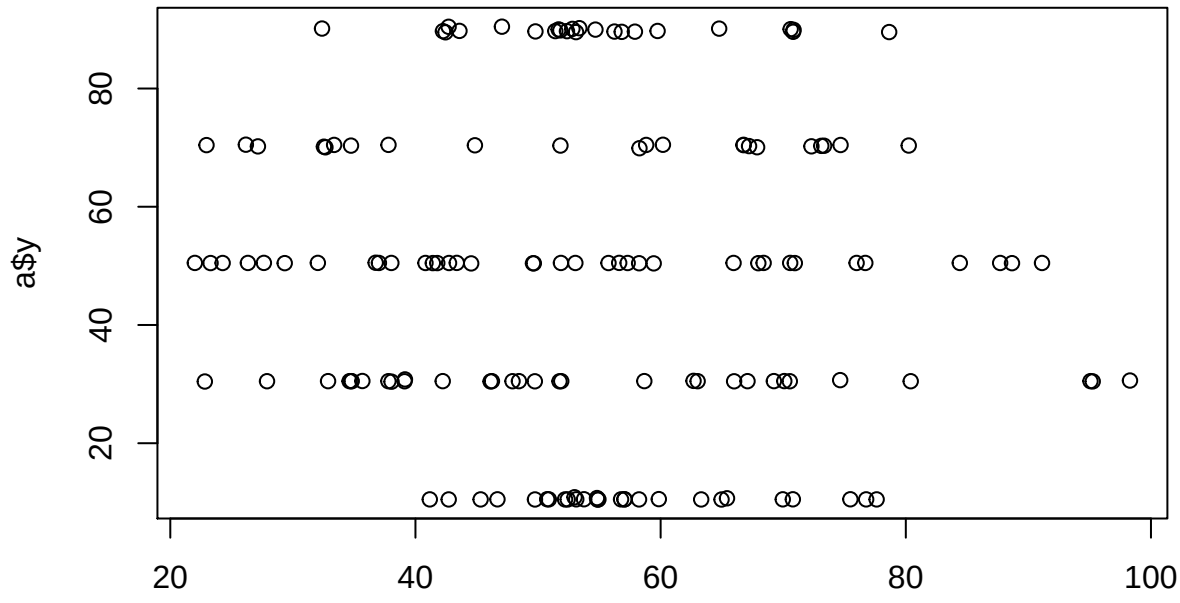
2.

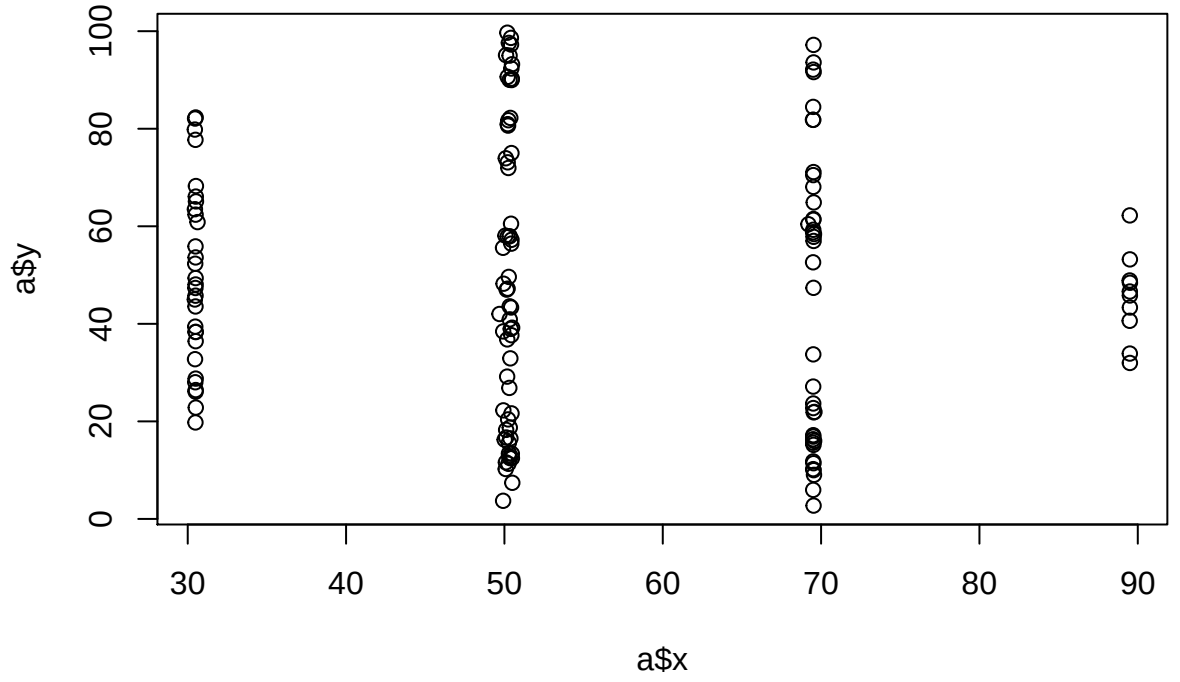
```
scatterp<-function(b){
  plot(b$y~b$x)
```

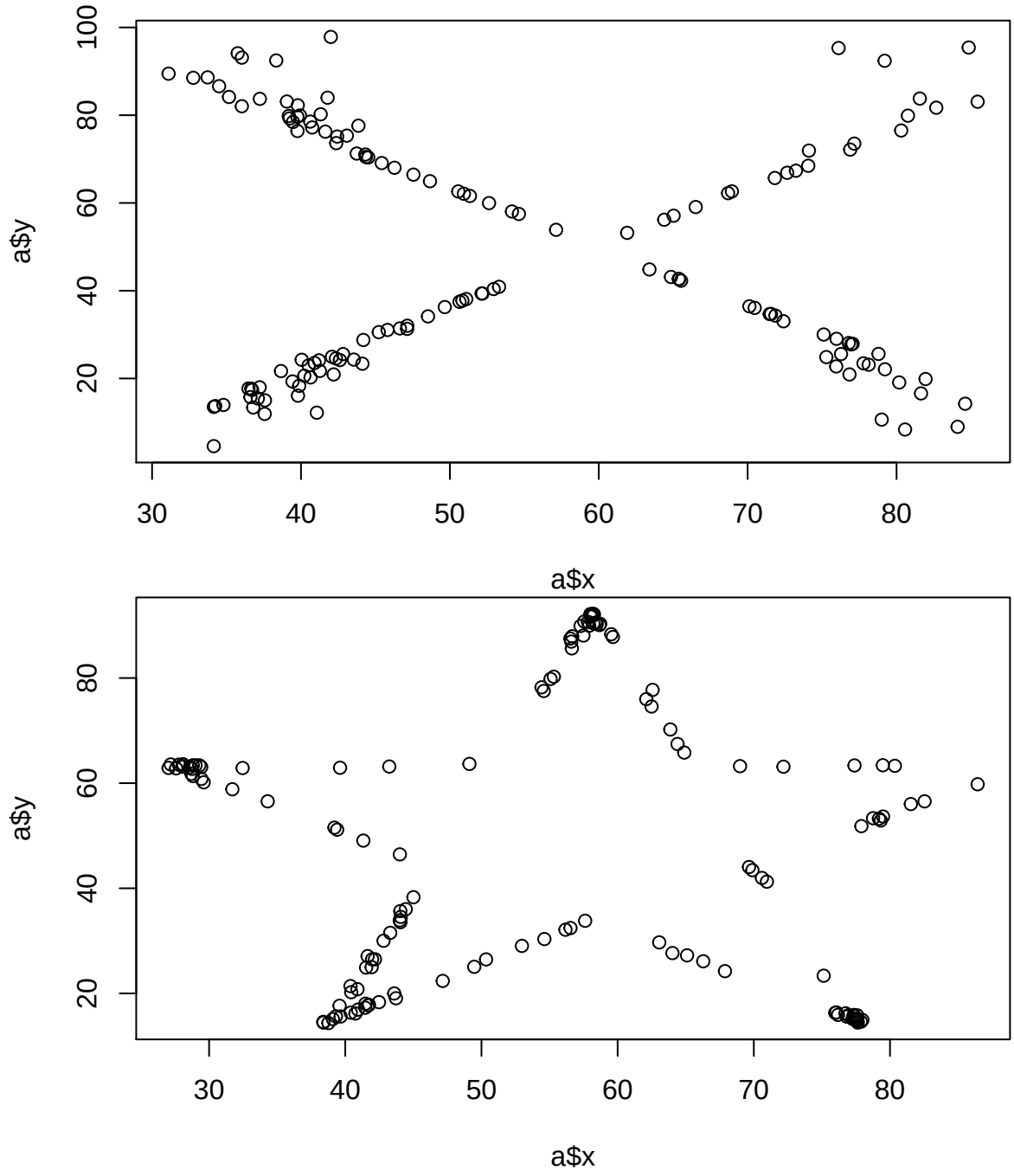
```
}
sscatterp<-function(b,observer){
  for (i in observer){
    a<-b[which(b$Observer==i),]
    plot(a$y~a$x)
  }}
sscatterp(b,unique(b$Observer))
```

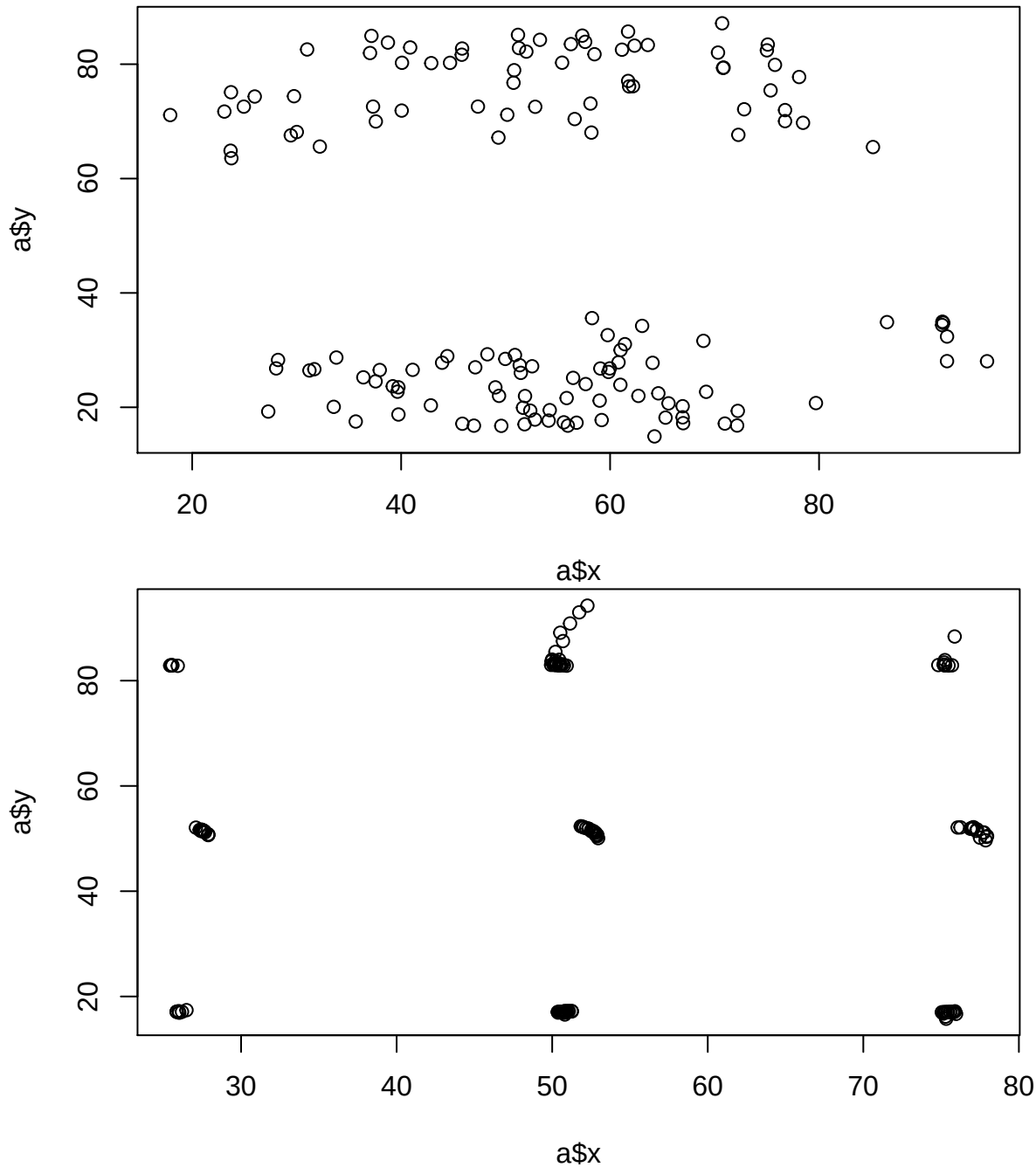


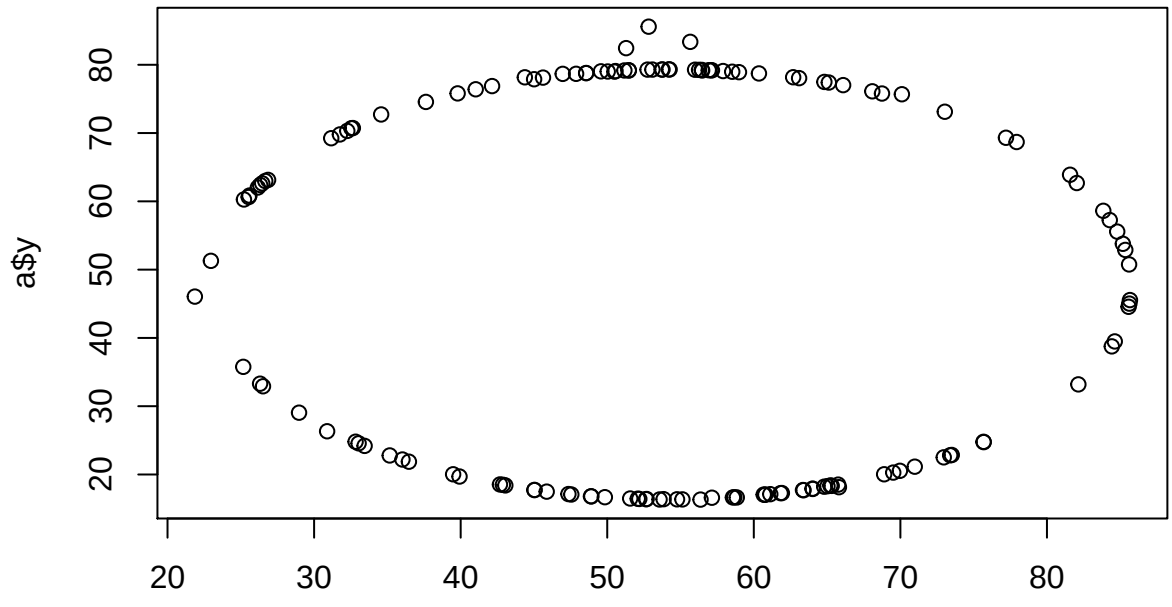


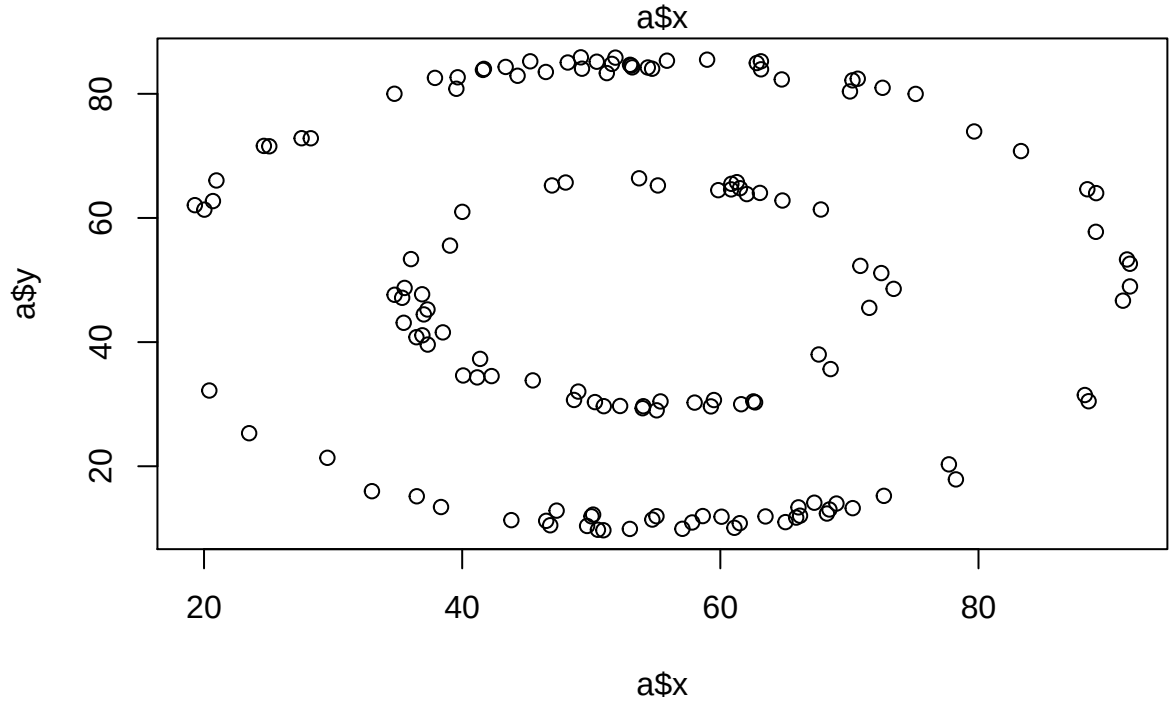

a$y
20
40
60
80
20
40
60
80
100
a$x

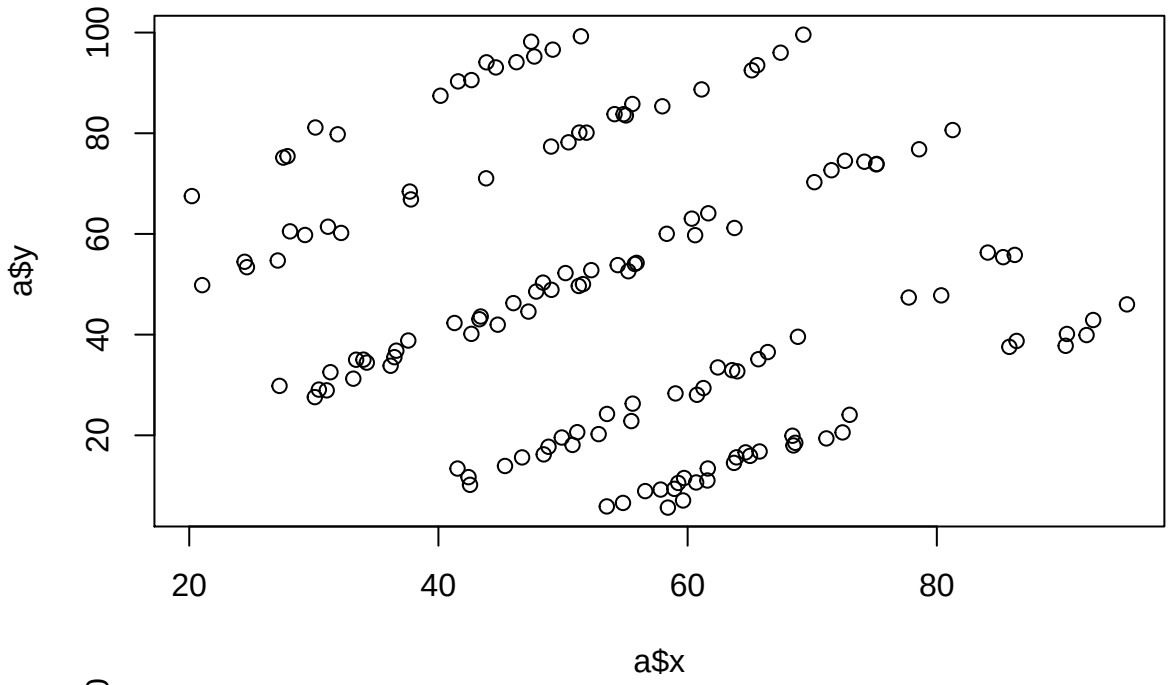

a$y
0
20
40
60
80
100
30
40
50
60
70
80
90
a$x

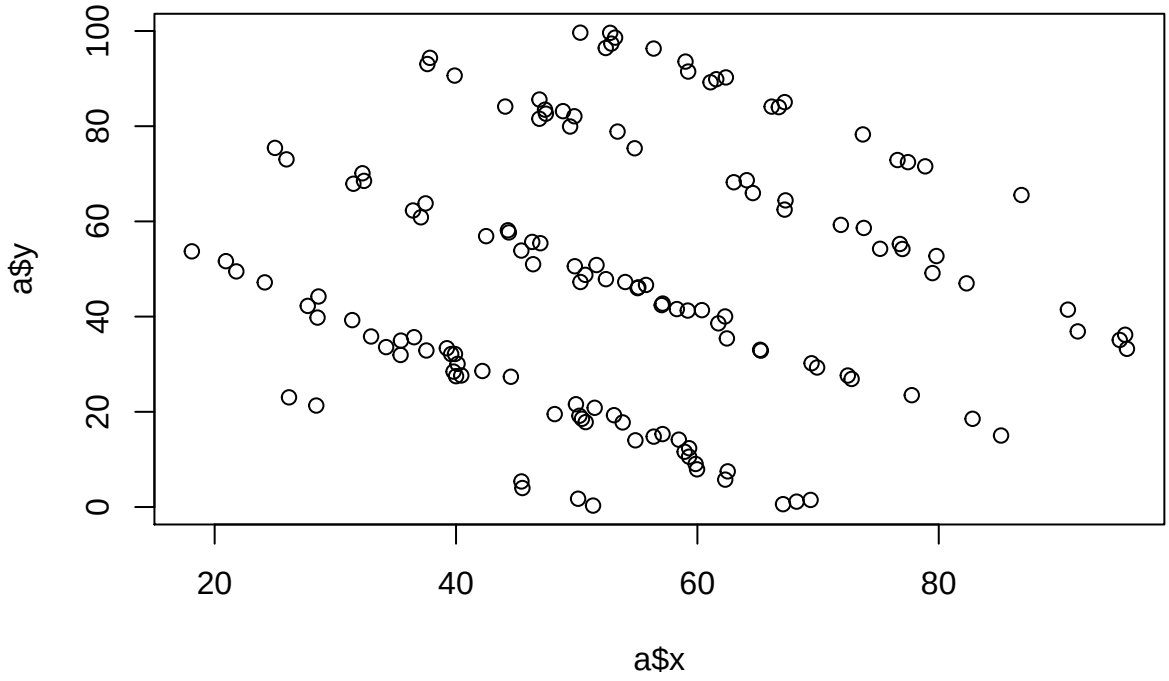

a$y
20
40
60
80
100
30
40
50
60
70
80
a$x
a$y
20
40
60
80
30
40
50
60
70
80
a$x

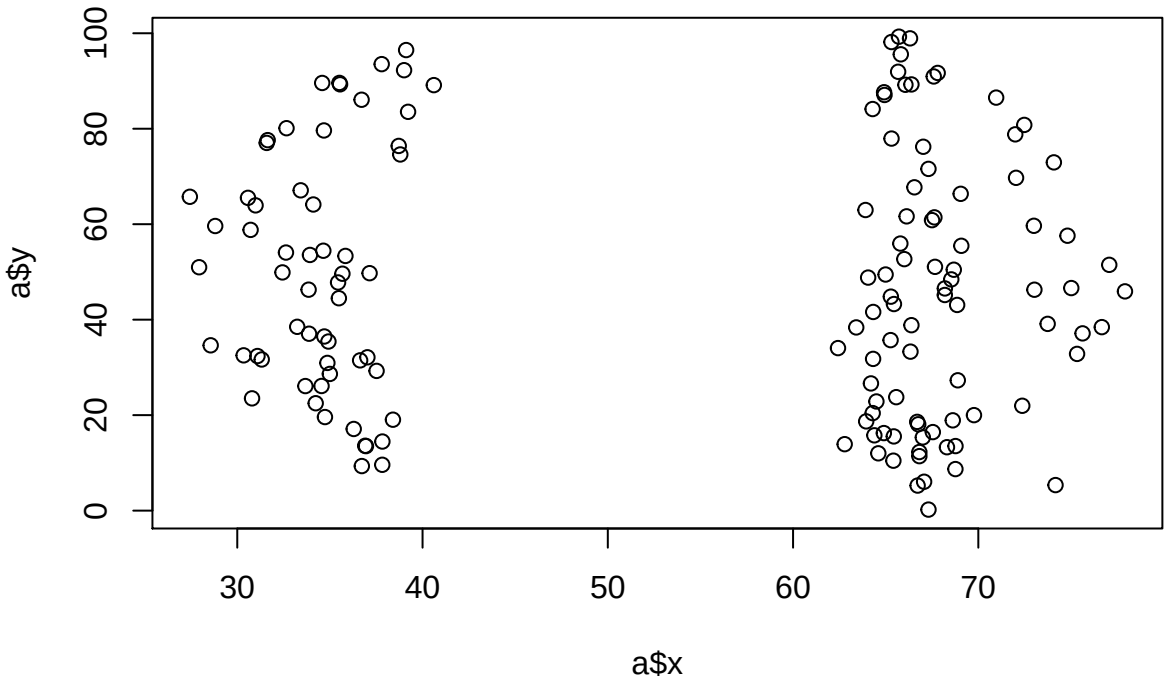

a$y
20
40
60
80
20
40
60
80
a$x
a$y
20
40
60
80
30
40
50
60
70
80
a$x


a$y
20
30
40
50
60
70
80
a$x


a$y
20
40
60
80
a$x


a$y
100
80
60
40
20
20
40
60
80
a$x


a$y
100
80
60
40
20
0
20
40
60
80
a$x



Problem 5

part a

```
library(downloader)
    download("http://www.farinspace.com/wp-content/uploads/us_cities_and_states.zip",dest="us_cities_sta
    unzip("us_cities_states.zip", exdir="./")

    #read in data, looks like sql dump, blah
    library(data.table)
    states <- fread(input = "./us_cities_and_states/states.sql",skip = 23,sep = "'", sep2 = ",", header
    cities <- fread(input = "./us_cities_and_states/cities_extended.sql",skip = 23,sep = "'", sep2 = ",
```

part b

```
citiesnumber<-c()
statesname<-states$V4
for (i in 1:length(statesname)){
  citiesnumber<-c(citiesnumber,sum(cities$V4==statesname[i]))
}
summary<-cbind(statesname,citiesnumber)
```

part c

```
countletter<-function(letter,state_name){
  number<-sum(state_name==letter)
  return(number)
}
##pseudo code
letter_count <- data.frame(matrix(NA,nrow=50, ncol=26))
getCount <- function(args){
    temp <- strsplit(state_name)
        # how to count??
    return(count)
```

```
    }
for(i in 1:50){
    letter_count[i,] <- apply(args)
    }
```

part d

```
    #https://cran.r-project.org/web/packages/fiftystater/vignettes/fiftystater.html
    library(ggplot2)
    #library(fiftystater)

    data("fifty_states") # this line is optional due to lazy data loading
    crimes <- data.frame(state = tolower(rownames(USArrests)), USArrests)
    # map_id creates the aesthetic mapping to the state name column in your data
    p <- ggplot(crimes, aes(map_id = state)) +
      # map points to the fifty_states shape data
      geom_map(aes(fill = Assault), map = fifty_states) +
      expand_limits(x = fifty_states$long, y = fifty_states$lat) +
      coord_map() +
      scale_x_continuous(breaks = NULL) +
      scale_y_continuous(breaks = NULL) +
      labs(x = "", y = "") +
      theme(legend.position = "bottom",
            panel.background = element_blank())

    p
    #ggsave(plot = p, file = "HW6_Problem6_Plot_Settlage.pdf")
```