Homework6

Yining Song

Problem 3 (a)

```
sp=function(x){
  n=length(x)
  s=sum(x)
  p=s/n
  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,sp)
```

```
##  [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,sp)
```

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

The rowprobabilities are all 1 or 0, and the column probabilities are all 0.6. This is because the columns of P4b_data are all identical.

(d)

```
pvector=function(p){
  x=rbinom(10,1,prob = p)
  return(x)
}
set.seed(12345)
pbty=as.array((30:40)/100)
P4b_data2=apply(pbty,1,pvector)
apply(P4b_data2,2,sp)
```

```
##  [1] 0.6 0.2 0.3 0.4 0.3 0.4 0.6 0.3 0.3 0.5 0.6
```

```
apply(P4b_data2,1,sp)
```

```
##  [1] 0.72727273 0.36363636 0.54545455 0.54545455 0.27272727 0.09090909
##  [7] 0.72727273 0.36363636 0.18181818 0.27272727
```

Problem 4

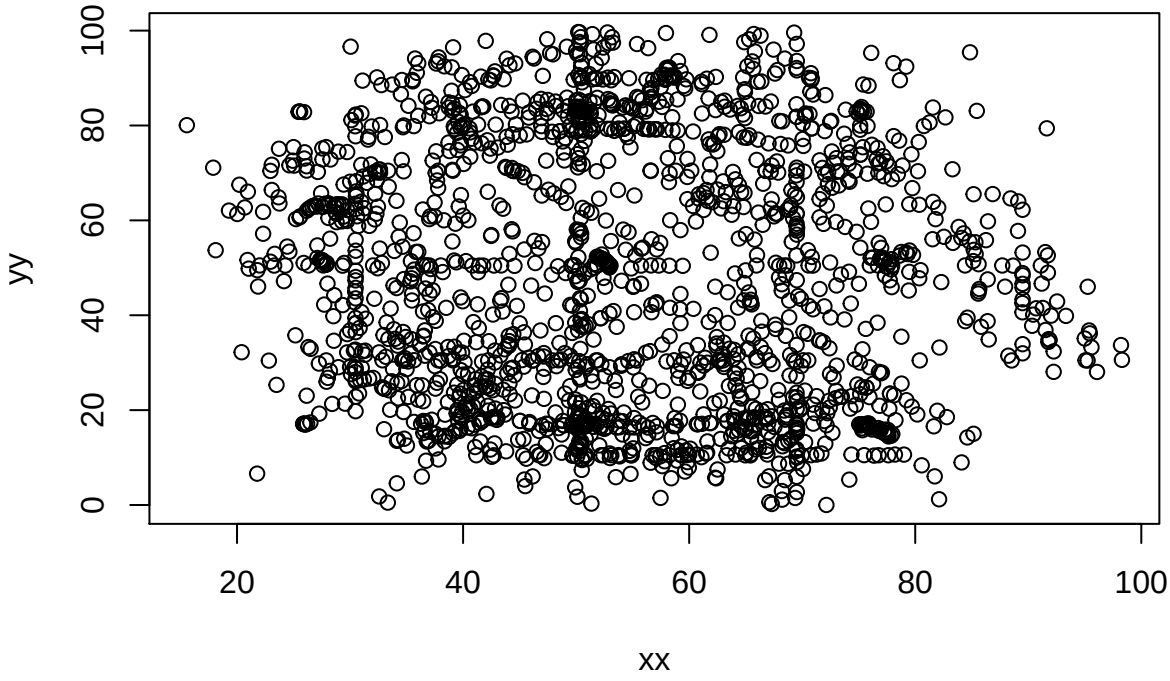
```
A=readRDS("HW4_data.rds")
colnames(A)
```

```
## [1] "Observer" "dev1"     "dev2"
```

```
colnames(A)[2]="x"
colnames(A)[3]="y"
```

(a)

```
SCP=function(T){
  xx=T[,1]
  yy=T[,2]
  plot(xx,yy)
}
SCP(A[,-1])
```



Problem 5

(a)

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.4
```

```
states=fread(input = "states.sql",skip = 23,sep = "'", sep2 = ",", header = F, select = c(2,4))
cities=fread(input = "cities.sql",sep = "'", sep2 = ",", header = F, select = c(2,4))
```

(b)

```
tapply(cities$V2,cities$V4,length)
```

```
##   AK   AL   AR   AZ   CA   CO   CT   DC   DE   FL   GA   HI   IA   ID   IL
##  229  579  605  264 1239  400  269    3   57  524  629   92  937  266 1287
##   IN   KS   KY   LA   MA   MD   ME   MI   MN   MO   MS   MT   NC   ND   NE
##  738  634  803  479  511  430  461  885  810  942  440  360  762  373  528
##   NH   NJ   NM   NV   NY   OH   OK   OR   PA   PR   RI   SC   SD   TN   TX
##  255  579  346   99 1612 1069  585  379 1802   99   70  377  364  548 1466
##   UT   VA   VT   WA   WI   WV   WY
```

```
##  250  839  288  493  753  753  176
```

(c)

```
letter_count=data.frame(matrix(NA,nrow=51, ncol=26))
getCount=function(letter,state_name){
  temp=strsplit(state_name,'')
  count=sum(unlist(temp)==letter)
  return(count)
}
alphab=c('a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p'
         ,'q','r','s','t','u','v','w','x','y','z')
Alphab=c('A','B','C','D','E','F','G','H','I','J','K','L','M','N','O','P'
         ,'Q','R','S','T','U','V','W','X','Y','Z')
for(i in 1:51){
  for(j in 1:26){
    letter_count[i,j]=getCount(alphab[j],states$V2[i])+getCount(Alphab[j],states$V2[i])
  }
}
colnames(letter_count)=alphab
row.names(letter_count)=states$V2
```

(d) package ???fiftystater??? is not available for my R.