

HW6

Shuangshuang Xu

2019/10/4

Prob 3

```
#a
proportionOfSuccess <- function(x) {
  # x is a vector, calculate proportionn of success
  p <- sum(x>0)/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
#calculate by column
apply(P4b_data, 2, proportionOfSuccess)
```

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

```
#calculate by row
apply(P4b_data, 1, proportionOfSuccess)
```

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

Each column of P4b_data is the same. That means when we create the matrix, we do not generate data of random variable with different probability.

```
#d
set.seed(12345)

prob <- matrix((30:40)/100, nrow = 11)
rbinom10 <- function(x) {
  #generate data from binomial with p = x
  binRandom <- rbinom(10, 1, x)
  return(binRandom)
}
```

```
P4b_data <- apply(prob, 1, rbinom10)
print(P4b_data)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
## [1,]    1    0    0    1    1    1    1    1    1    1    0
## [2,]    1    0    0    0    0    1    0    0    0    1    1
## [3,]    1    1    1    0    1    0    1    0    0    0    1
## [4,]    1    0    1    1    1    0    1    0    0    0    1
## [5,]    0    0    0    0    0    1    1    0    0    1    0
## [6,]    0    0    0    0    0    0    0    0    0    0    1
## [7,]    0    0    1    1    0    1    1    1    1    1    1
```

```
## [8,] 0 0 0 1 0 0 0 1 0 1 1
## [9,] 1 0 0 0 0 0 0 0 0 1 0
## [10,] 1 1 0 0 0 0 1 0 0 0 0
```

```
#calculate by column
```

```
print("marginalSuccess:")
```

```
## [1] "marginalSuccess:"
```

```
apply(P4b_data, 2, proportionOfSuccess)
```

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

problem 4

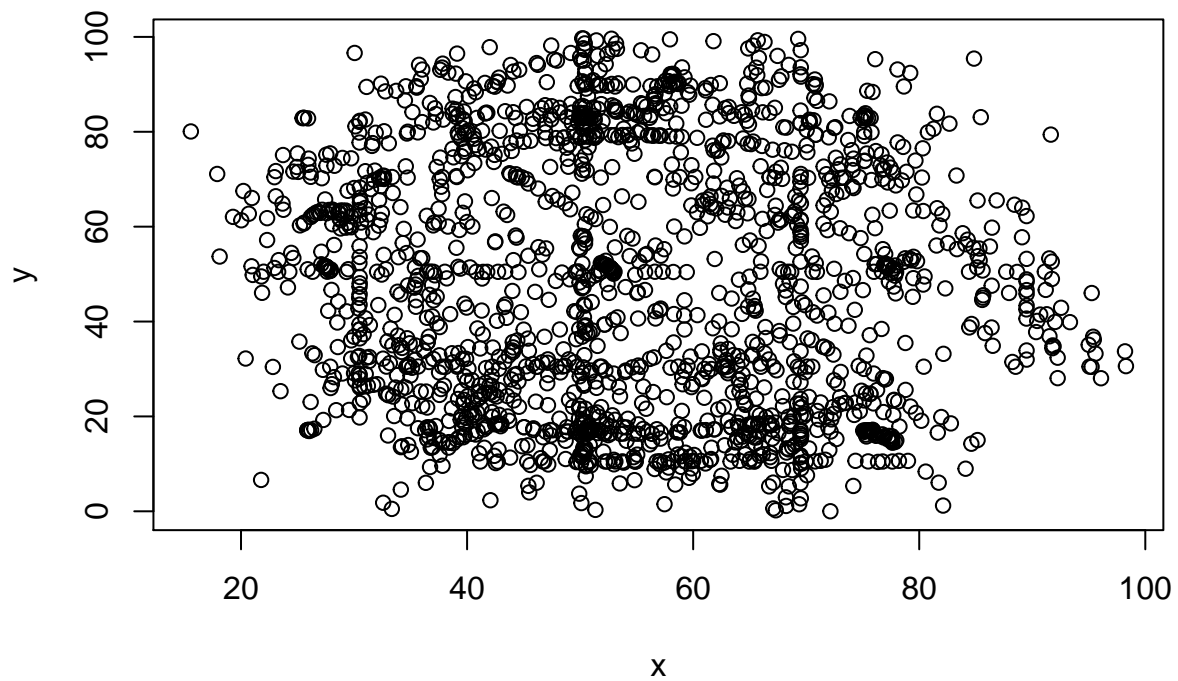
(1)

```
data <- readRDS("C:/Users/44653/Desktop/gitfile/HW4_data.rds")
colnames(data) <- c("Observer", "x", "y")
plotfunc <- function(z){
  #create the plot of dataframe, z <- list(observer, data)
  if (z[[1]] == 0){
    plot(z[[2]]$x, z[[2]]$y, main = "plot of entire dataset", xlab = "x", ylab = "y")
  } else {
    plot(z[[2]]$x, z[[2]]$y, main = paste("plot for observer", z[[1]]), xlab = "x", ylab = "y")
  }
}
```

(2)

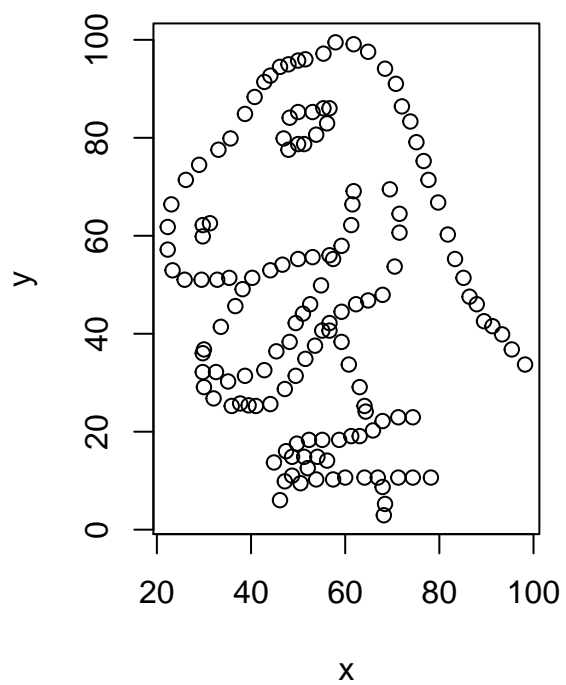
```
plotfunc(list(0,data))
```

plot of entire dataset

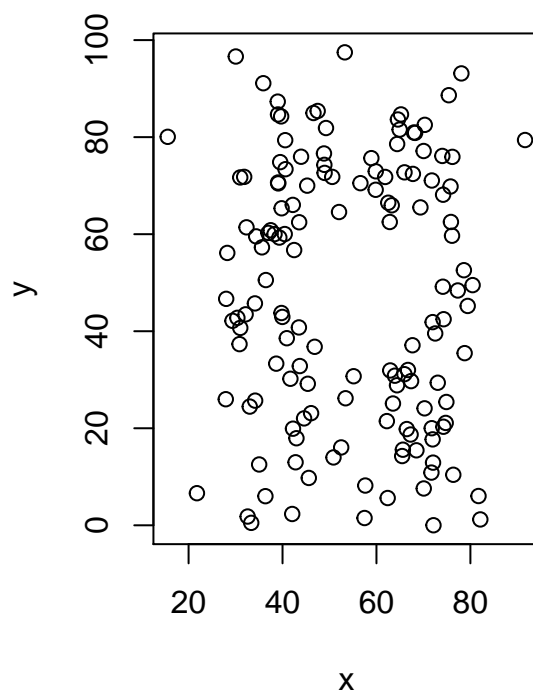


```
observer <- unique(data$Observer)
#list of each observer's data
q <- sapply(observer, function(x) {list(x,data[data$Observer == x, ])}, simplify = FALSE)
par(mfrow=c(1,2))
lapply(q, plotfunc)
```

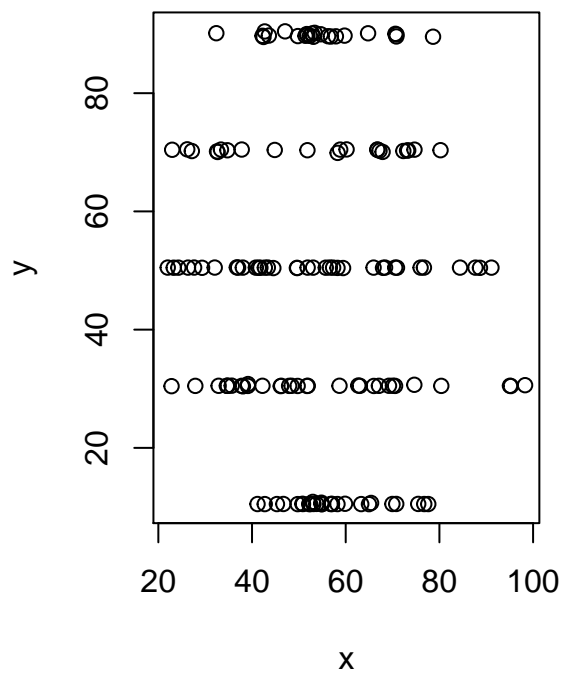
plot for observer 4



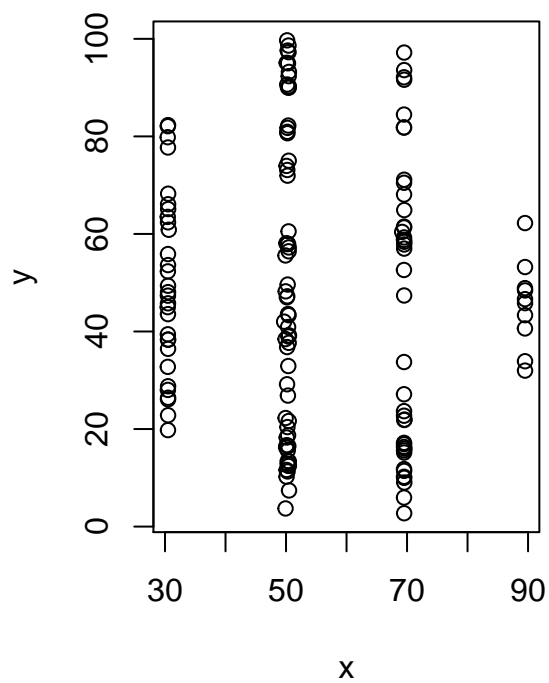
plot for observer 1



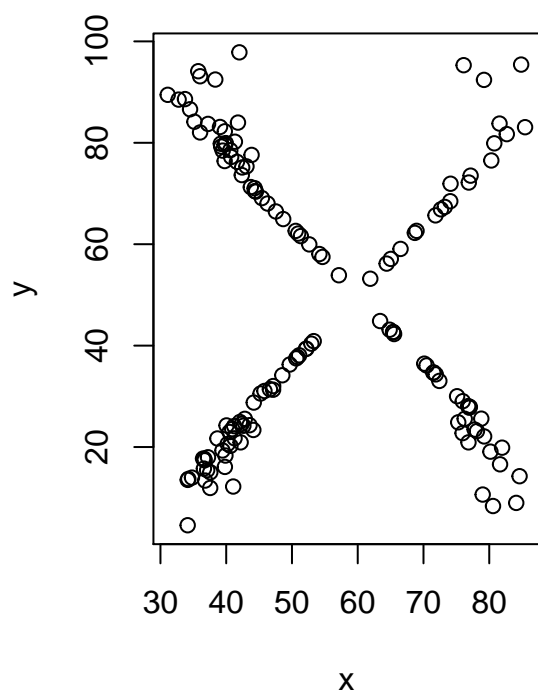
plot for observer 6



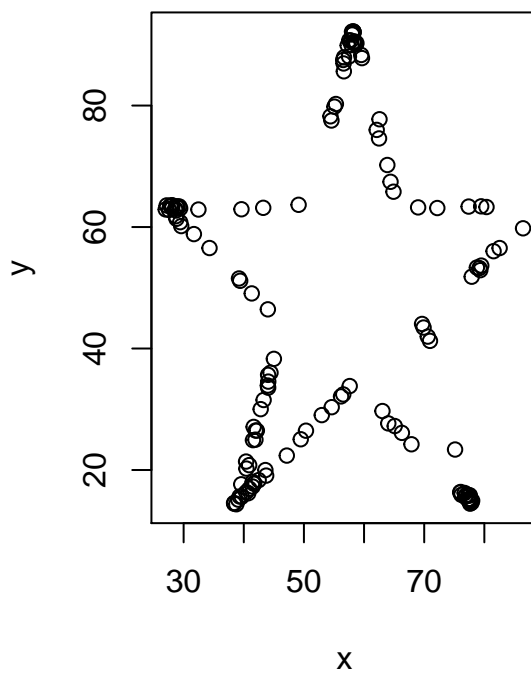
plot for observer 11



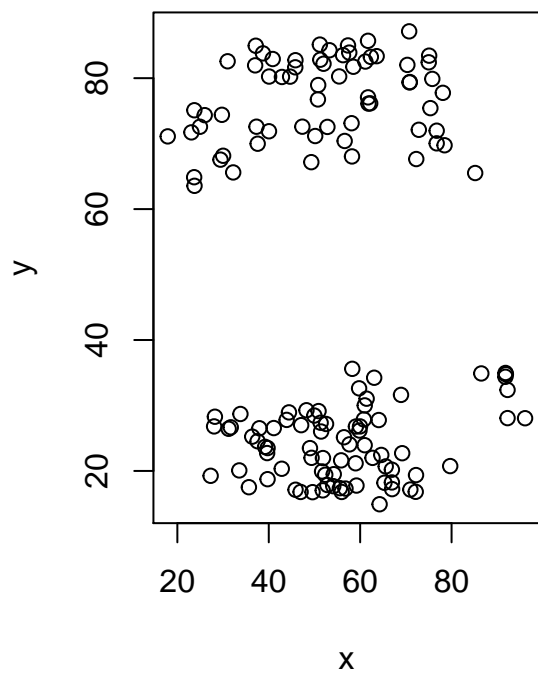
plot for observer 13



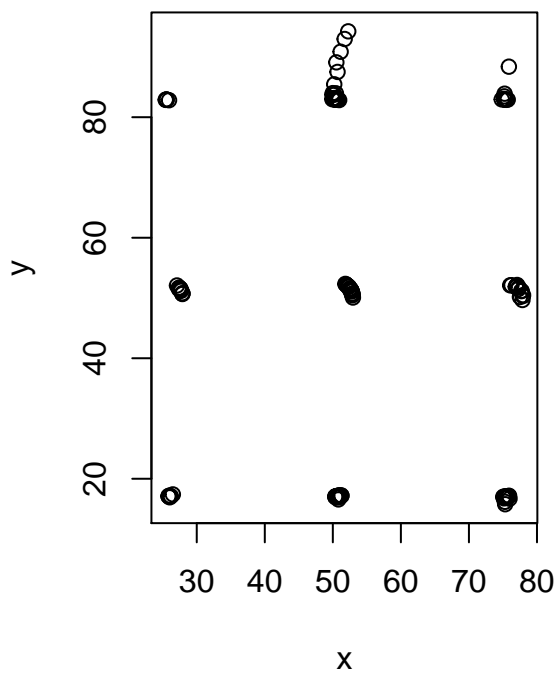
plot for observer 10



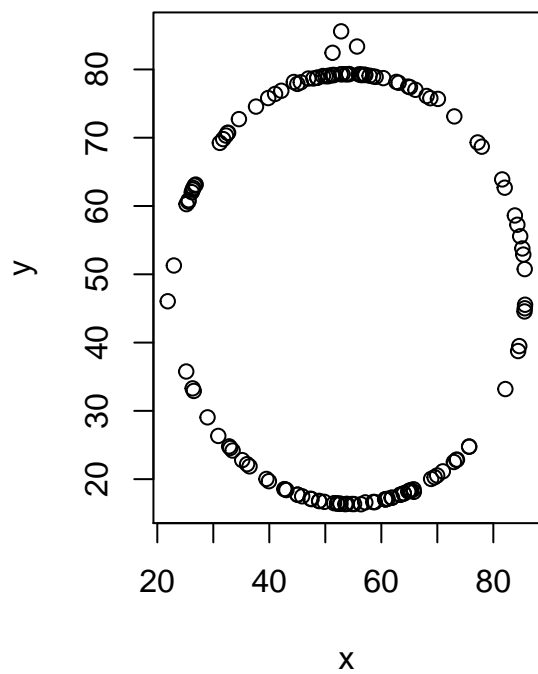
plot for observer 7



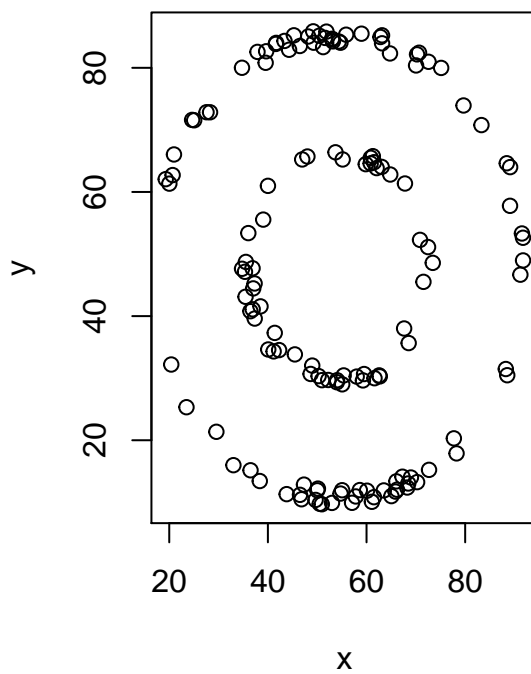
plot for observer 5



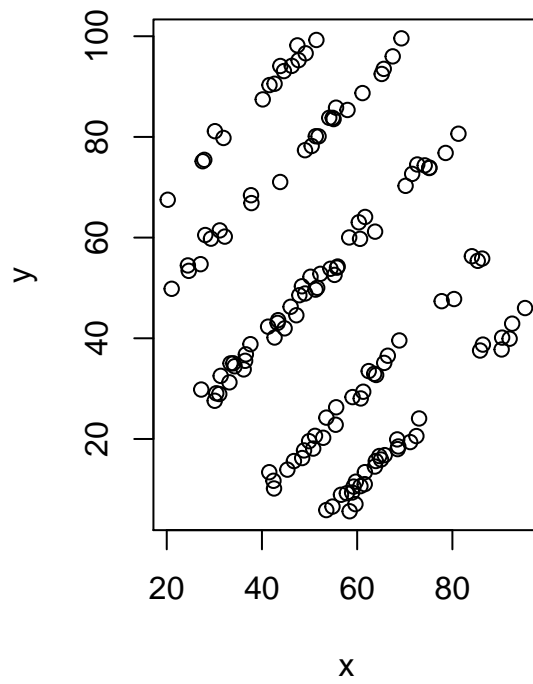
plot for observer 3



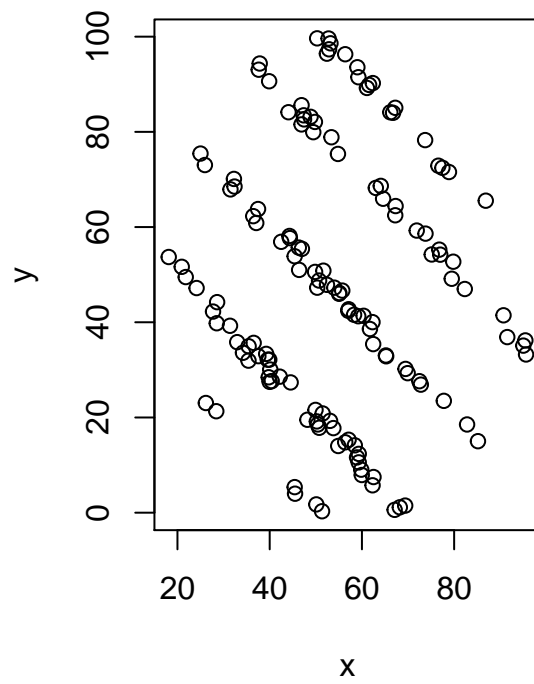
plot for observer 2



plot for observer 9



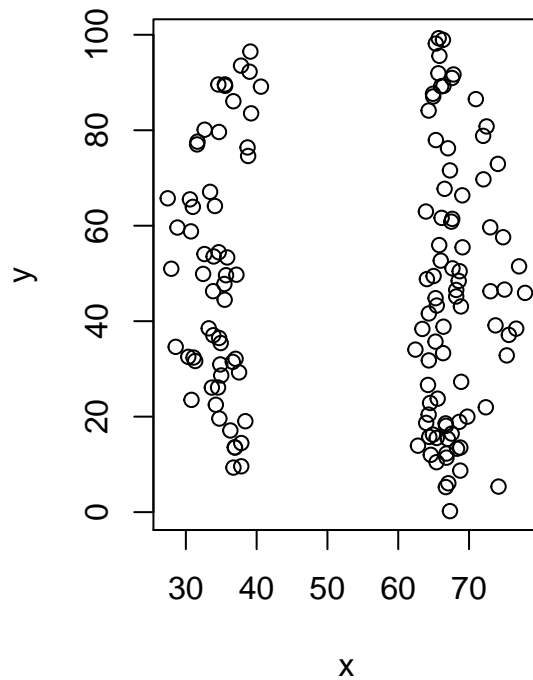
plot for observer 8



```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
##
## [[7]]
## NULL
##
## [[8]]
## NULL
##
## [[9]]
## NULL
```

```
##
## [[10]]
## NULL
##
## [[11]]
## NULL
##
## [[12]]
## NULL
##
## [[13]]
## NULL
```

plot for observer 12



problem 5

(a)

```
## Warning: package 'downloader' was built under R version 3.5.3
## Warning: package 'data.table' was built under R version 3.5.3
##
##           V2 V4
## 1:   Holtsville NY
## 2:    Adjuntas PR
## 3:    Aguada PR
```

```

## 4:    Aguadilla PR
## 5:    Maricao PR
## 6:    Anasco PR
## 7:    Angeles PR
## 8:    Arecibo PR
## 9:    Bajadero PR
## 10:   Barceloneta PR
## 11:   Boqueron PR
## 12:   Cabo Rojo PR
## 13:   Penuelas PR
## 14:   Camuy PR
## 15:   Castaner PR
## 16:   Rosario PR
## 17:   Sabana Grande PR
## 18:   Ciales PR
## 19:   Utuado PR
## 20:   Dorado PR
## 21:   Ensenada PR
## 22:   Florida PR
## 23:   Garrochales PR
## 24:   Guanica PR
## 25:   Guayanilla PR
## 26:   Hatillo PR
## 27:   Hormigueros PR
## 28:   Isabela PR
## 29:   Jayuya PR
## 30:   Lajas PR
## 31:   Lares PR
## 32:   Las Marias PR
## 33:   Manati PR
## 34:   Moca PR
## 35:   Rincon PR
## 36:   Quebradillas PR
## 37:   Mayaguez PR
## 38:   San German PR
## 39:   San Sebastian PR
## 40:   Morovis PR
## 41:   Sabana Hoyos PR
## 42:   San Antonio PR
## 43:   Vega Alta PR
## 44:   Vega Baja PR
## 45:   Yauco PR
## 46:   Aguas Buenas PR
## 47:   Aguirre PR
## 48:   Aibonito PR
## 49:   Maunabo PR
## 50:   Arroyo PR
##      V2 V4

```

(b)

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.5.3
```

```
countCity <- data.frame(table(cities$V4))  
colnames(countCity) <- c("state", "num of cities")  
kable(countCity)
```

state	num of cities
AK	229
AL	579
AR	605
AZ	264
CA	1239
CO	400
CT	269
DC	3
DE	57
FL	524
GA	629
HI	92
IA	937
ID	266
IL	1287
IN	738
KS	634
KY	803
LA	479
MA	511
MD	430
ME	461
MI	885
MN	810
MO	942
MS	440
MT	360
NC	762
ND	373
NE	528
NH	255
NJ	579
NM	346
NV	99
NY	1612
OH	1069
OK	585
OR	379
PA	1802
PR	99
RI	70
SC	377
SD	364
TN	548
TX	1466
UT	250
VA	839

state	num of cities
VT	288
WA	493
WI	753
WV	753
WY	176

(c)

```
letter_count <- data.frame(matrix(NA, nrow = 51, ncol = 26))

getCount <- function(letter) {
  #count the number of letter in the string
  temp <- unlist(strsplit(state, ""))
  count <- table(temp)[letter]
  return(count)
}

# count each state's letter
j <- 1
for (i in tolower(states$V2)){
  state <- i
  letter_count[j, ] <- apply(matrix(letters, nrow = 1), 2, getCount)
  j <- j+1
}

colnames(letter_count) <- letters
rownames(letter_count) <- states$V2
letter_count[is.na(letter_count)] <- 0
kable(letter_count, caption = "count the number of letters in state name")
```

Table 2: count the number of letters in state name

	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
Alaska	3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Alabama	4	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Arkansas	3	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	2	0	0	0	0	0	0	0
Arizona	2	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1
California	2	0	1	0	0	1	0	0	2	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0
Colorado	1	0	1	1	0	0	0	0	0	0	0	1	0	0	3	0	0	1	0	0	0	0	0	0	0	0
Connecticut	0	0	3	0	1	0	0	0	1	0	0	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0
District of Columbia	1	1	2	1	0	1	0	0	3	0	0	1	1	0	2	0	0	1	1	2	1	0	0	0	0	0
Delaware	2	0	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
Florida	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Georgia	1	0	0	0	1	0	2	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Hawaii	2	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Iowa	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
Idaho	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Illinois	0	0	0	0	0	0	0	0	3	0	0	2	0	1	1	0	0	0	1	0	0	0	0	0	0	0
Indiana	2	0	0	1	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Kansas	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0

	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
Kentucky	0	0	1	0	1	0	0	0	0	0	2	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0
Louisiana	2	0	0	0	0	0	0	0	2	0	0	1	0	1	1	0	0	0	1	0	1	0	0	0	0	0
Massachusetts	2	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	4	2	1	0	0	0	0	0
Maryland	2	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0
Maine	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Michigan	1	0	1	0	0	0	1	1	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Minnesota	1	0	0	0	1	0	0	0	1	0	0	0	1	2	1	0	0	0	1	1	0	0	0	0	0	0
Missouri	0	0	0	0	0	0	0	0	2	0	0	0	1	0	1	0	0	1	2	0	1	0	0	0	0	0
Mississippi	0	0	0	0	0	0	0	0	4	0	0	0	1	0	0	2	0	0	4	0	0	0	0	0	0	0
Montana	2	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0	0	0	0
North Carolina	2	0	1	0	0	0	0	1	1	0	0	1	0	2	2	0	0	2	0	1	0	0	0	0	0	0
North Dakota	2	0	0	1	0	0	0	1	0	0	1	0	0	1	2	0	0	1	0	2	0	0	0	0	0	0
Nebraska	2	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0
New Hampshire	1	0	0	0	2	0	0	2	1	0	0	0	1	1	0	1	0	1	1	0	0	0	1	0	0	0
New Jersey	0	0	0	0	3	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	1	0	1	0
New Mexico	0	0	1	0	2	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	0	0
Nevada	2	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
New York	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	0	1	0
Ohio	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Oklahoma	2	0	0	0	0	0	0	1	0	0	1	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0
Oregon	0	0	0	0	1	0	1	0	0	0	0	0	0	1	2	0	0	1	0	0	0	0	0	0	0	0
Pennsylvania	2	0	0	0	1	0	0	0	1	0	0	1	0	3	0	1	0	0	1	0	0	1	0	0	1	0
Rhode Island	1	0	0	2	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0
South Carolina	2	0	1	0	0	0	0	1	1	0	0	1	0	1	2	0	0	1	1	1	1	0	0	0	0	0
South Dakota	2	0	0	1	0	0	0	1	0	0	1	0	0	0	2	0	0	0	1	2	1	0	0	0	0	0
Tennessee	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	0	0	0	2	1	0	0	0	0	0	0
Texas	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
Utah	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Virginia	1	0	0	0	0	0	1	0	3	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0
Vermont	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1	0	1	0	0	0	0
Washington	1	0	0	0	0	0	1	1	1	0	0	0	0	2	1	0	0	0	1	1	0	0	1	0	0	0
Wisconsin	0	0	1	0	0	0	0	0	2	0	0	0	0	2	1	0	0	0	2	0	0	0	1	0	0	0
West Virginia	1	0	0	0	1	0	1	0	3	0	0	0	0	1	0	0	0	1	1	1	0	1	1	0	0	0
Wyoming	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0

(d)

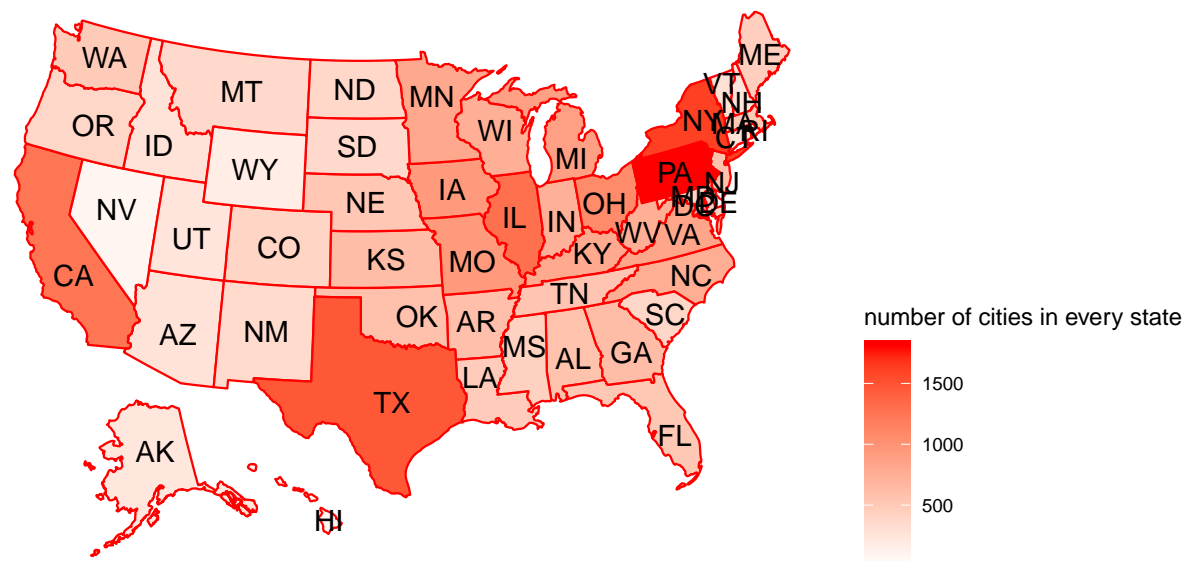
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
library(usmap)
```

```
## Warning: package 'usmap' was built under R version 3.5.3
```

```
plot_usmap(data = countCity, values = "num of cities", color = "red", labels = TRUE, label_color = "black",
  scale_fill_continuous(low = "white", high = "red", name = "number of cities in every state")+
  theme(legend.position = "right")
```



```
#count number of letter which appears more than three in state name
largerThanThree <- data.frame(num=apply(letter_count>3, 1, sum))
largerThanThree$state <- states$V4

plot_usmap(data = largerThanThree, values = "num", color = "blue", labels = TRUE, label_color = "black",
  scale_fill_continuous(low = "white", high = "red", name = "state have > 3 occurrences of ANY letter")+
  theme(legend.position = "right")
```

