ProblemSet2_1

Name: Zhen Qin, Uniqname: qinzhen, UMID: 48800866, Dept: Statistics

a.

First, I stored the data in `recs`.

```
## Store data in recs
library('readr')
library('dplyr')
library('tidyr')
setwd('E:\\UM academy\\stat 506\\ProblemSet2')
recs=read_delim('./recs2009_public.csv', delim=',',col_names=TRUE)
```

Set variables. `Sweight` is sum of nweight grouped by `State` and `Roof`, `SSweight` is sum of nweight grouped by `State`. Generate the Rate matrix.

```
## Calculate matrix including rate as Ratemat
select(recs, State=REPORTABLE_DOMAIN, Roof=ROOFTYPE, Weight = NWEIGHT) %>%
  group_by(State, Roof) %>%
  summarise(Sweight=sum(Weight)) %>%
  group_by(State) %>%
  mutate(SSweight=sum(Sweight),Rate=Sweight/SSweight) %>%
  select(Roof, State, Rate) %>%
  filter(Roof==2)->Ratemat
Ratemat=Ratemat[,-1]
```

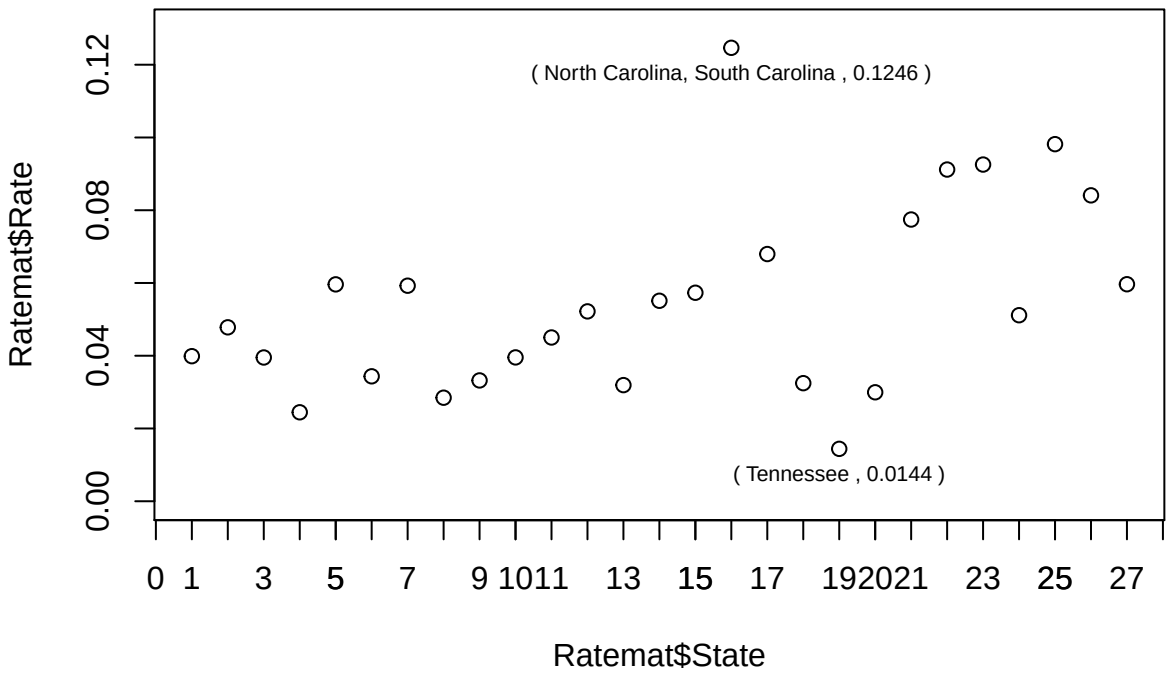
Then showed `Ratemat` and calculate max value and min value of `Rate`.

```
## Calculate max and min value
Statestr="Connecticut, Maine, New Hampshire, Rhode Island, Vermont
Massachusetts
New York
New Jersey
Pennsylvania
Illinois
Indiana, Ohio
Michigan
Wisconsin
Iowa, Minnesota, North Dakota, South Dakota
Kansas, Nebraska
Missouri
Virginia
Delaware, District of Columbia, Maryland, West Virginia
Georgia
North Carolina, South Carolina
Florida
Alabama, Kentucky, Mississippi
Tennessee
Arkansas, Louisiana, Oklahoma
Texas
Colorado
Idaho, Montana, Utah, Wyoming
Arizona
```

```
Nevada, New Mexico
California
Alaska, Hawaii, Oregon, Washington"
Statename=unlist(strsplit(Statestr,split = "\n"))
tmpmat=Ratemat
tmpmat$State=Statename
knitr::kable(tmpmat)
```

State	Rate
Connecticut, Maine, New Hampshire, Rhode Island, Vermont	0.0398771
Massachusetts	0.0478129
New York	0.0395206
New Jersey	0.0244547
Pennsylvania	0.0596170
Illinois	0.0343415
Indiana, Ohio	0.0592537
Michigan	0.0284647
Wisconsin	0.0332125
Iowa, Minnesota, North Dakota, South Dakota	0.0395430
Kansas, Nebraska	0.0450165
Missouri	0.0521929
Virginia	0.0319221
Delaware, District of Columbia, Maryland, West Virginia	0.0551071
Georgia	0.0573164
North Carolina, South Carolina	0.1246495
Florida	0.0679545
Alabama, Kentucky, Mississippi	0.0324584
Tennessee	0.0144062
Arkansas, Louisiana, Oklahoma	0.0299494
Texas	0.0774461
Colorado	0.0911870
Idaho, Montana, Utah, Wyoming	0.0925590
Arizona	0.0511356
Nevada, New Mexico	0.0981669
California	0.0840876
Alaska, Hawaii, Oregon, Washington	0.0596553

```
maxp=which.max(Ratemat$Rate)
minp=which.min(Ratemat$Rate)
plot(Ratemat$State,Ratemat$Rate,ylim = c(0,0.13))
axis(at=1:30,side = 1)
text(Ratemat$State[c(maxp,minp)],Ratemat$Rate[c(maxp,minp)],
     paste("(",Statename[c(maxp,minp)],",",round(Ratemat$Rate[c(maxp,minp)],4),")")
     ,pos=1,cex=0.7)
```



The max value is .1246495 and `State` is 16, the min value is .0144062 and `State` is 19. The answer 1 is North Carolina, South Carolina, the answer 2 is Tennessee.

b.

Compute the proportion of each roof type for all houses constructed in each decade.

```
## Compute the proportion
Roofstr="Not Applicable,Ceramic or Clay Tiles,Wood Shingles/Shakes,Metal,\
Slate or Synthetic Slate,Composition Shingles,Asphalt,Concrete Tiles,Other"
Roofname=unlist(strsplit(Roofstr,split = ","))
tmpstr=rep(Roofname[1],7)
for(i in 2:9){
  tmpstr=c(tmpstr,rep(Roofname[i],7))
}
Drecs=recs
Drecs$YEARMADERANGE[Drecs$YEARMADERANGE==8] = 7
select(Drecs, Roof=ROOFTYPE, Decade = YEARMADERANGE, Weight = NWEIGHT) %>%
  group_by(Roof, Decade) %>%
  summarise(Sweight=sum(Weight))->Smat
group_by(Smat,Decade) %>%
  mutate(SSweight=sum(Sweight),Rate=Sweight/SSweight) %>%
  select(Decade, Roof, Rate)->tmpmat
tmpmat$Decade=rep(c("before 1950","1950s",
                    "1960s","1970s","1980s","1990s","2000s"),9)
tmpmat$Roof=tmpstr
```

```
knitr::kable(tmpmat)
```

Decade	Roof	Rate
before 1950	Not Applicable	0.1387202
1950s	Not Applicable	0.0944621
1960s	Not Applicable	0.1933055
1970s	Not Applicable	0.2248517
1980s	Not Applicable	0.2059078
1990s	Not Applicable	0.1528900
2000s	Not Applicable	0.1565083
before 1950	Ceramic or Clay Tiles	0.0099519
1950s	Ceramic or Clay Tiles	0.0095485
1960s	Ceramic or Clay Tiles	0.0163117
1970s	Ceramic or Clay Tiles	0.0245931
1980s	Ceramic or Clay Tiles	0.0398188
1990s	Ceramic or Clay Tiles	0.0561358
2000s	Ceramic or Clay Tiles	0.0484759
before 1950	Wood Shingles/Shakes	0.0503573
1950s	Wood Shingles/Shakes	0.0726859
1960s	Wood Shingles/Shakes	0.0663760
1970s	Wood Shingles/Shakes	0.0488869
1980s	Wood Shingles/Shakes	0.0666404
1990s	Wood Shingles/Shakes	0.0518347
2000s	Wood Shingles/Shakes	0.0541310
before 1950	Metal	0.0591081
1950s	Metal	0.0410991
1960s	Metal	0.0506997
1970s	Metal	0.1012533
1980s	Metal	0.1094275
1990s	Metal	0.0979082
2000s	Metal	0.0379680
before 1950	Slate or Synthetic Slate	0.0233793
1950s	Slate or Synthetic Slate	0.0143400
1960s	Slate or Synthetic Slate	0.0108053
1970s	Slate or Synthetic Slate	0.0059041
1980s	Slate or Synthetic Slate	0.0094907
1990s	Slate or Synthetic Slate	0.0082757
2000s	Slate or Synthetic Slate	0.0037785
before 1950	Composition Shingles	0.4646863
1950s	Composition Shingles	0.5542037
1960s	Composition Shingles	0.4753213
1970s	Composition Shingles	0.4342896
1980s	Composition Shingles	0.4138830
1990s	Composition Shingles	0.4738540
2000s	Composition Shingles	0.5402771
before 1950	Asphalt	0.2300020
1950s	Asphalt	0.1938381
1960s	Asphalt	0.1706935
1970s	Asphalt	0.1370480
1980s	Asphalt	0.1381545
1990s	Asphalt	0.1382170
2000s	Asphalt	0.1309489
before 1950	Concrete Tiles	0.0033874

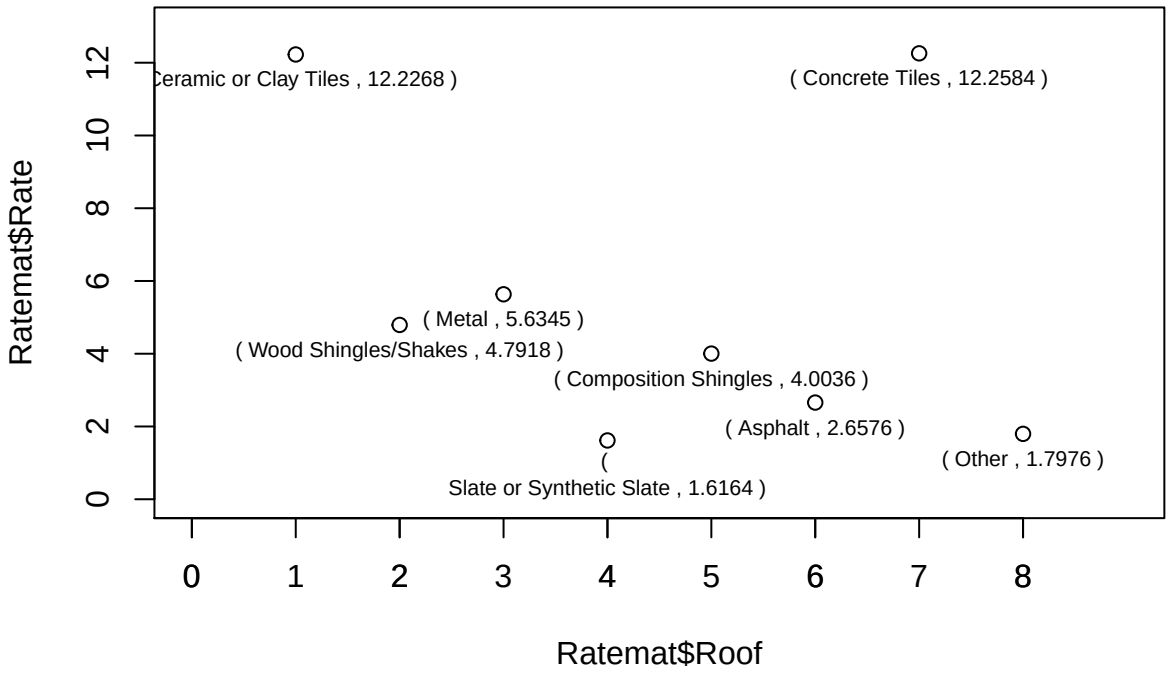
Decade	Roof	Rate
1950s	Concrete Tiles	0.0075852
1960s	Concrete Tiles	0.0056330
1970s	Concrete Tiles	0.0099774
1980s	Concrete Tiles	0.0114342
1990s	Concrete Tiles	0.0159820
2000s	Concrete Tiles	0.0249567
before 1950	Other	0.0204076
1950s	Other	0.0122376
1960s	Other	0.0108540
1970s	Other	0.0131960
1980s	Other	0.0052431
1990s	Other	0.0049025
2000s	Other	0.0029557

Calculate the max rate and generate a plot, and then use the string "Ceramic or Clay Tiles,Wood Shingles/Shakes,Metal,Slate or Synthetic Slate,Composition Shingles,Asphalt,Concrete Tiles,Other,Not Applicable" to get the names.

```
## Calculate the max rate and generate a plot
spread(Smat, Decade,Sweight) %>%
  transmute(Rate=(`2`+`3`+`4`+`5`+`6`)/`1`)->Ratemat
Roofname[which.max(Ratemat$Rate)]
```

```
## [1] "Concrete Tiles"
```

```
plot(Ratemat$Roof,Ratemat$Rate,ylim = c(0,13),xlim = c(0,9))
axis(at=-2:8,side = 1)
text(Ratemat$Roof,Ratemat$Rate,paste("(",Roofname,",",round(Ratemat$Rate,4),")")
     ,pos=1,cex=0.7)
```



Use this subset by decades to calculate, Concrete Tiles roof saw the largest relative rise.