Stats 130 Project

December 2, 2014

Members: XUAN LIN (204278880)

XIAOYU WANG(604327727)

CHENYAO LI()

```
sec201205 <- read.csv("~/Desktop/SEC Registered Adviser Report 5-1-2012.csv", stringsAsFactors=FALSE)[,
sec201305 <- read.csv("~/Desktop/SEC Registered Investment Adviser Report 2013-5-1.csv", stringsAsFacto
sec201405 <- read.csv("~/Desktop/SEC Registered Investment Adviser Report 2014-5-1.csv", stringsAsFacto
```

recode the variable SEC.Region from 201205

```
sec201205$GEOG <- "OTHER"
sec201205$GEOG[sec201205$SEC.Region %in% c("NYRO")] <- "NYRO"
sec201205$GEOG[sec201205$SEC.Region %in% c("BRO")] <- "BRO"
sec201205$GEOG[sec201205$SEC.Region %in% c("CHRO")] <- "CHRO"
sec201205$GEOG[sec201205$SEC.Region %in% c("LARO","SFRO")] <- "CARO"

# recode the variable SEC.Region from 201305
sec201305$GEOG <- "OTHER"
sec201305$GEOG[sec201305$SEC.Region %in% c("NYRO")] <- "NYRO"
sec201305$GEOG[sec201305$SEC.Region %in% c("BRO")] <- "BRO"
sec201305$GEOG[sec201305$SEC.Region %in% c("CHRO")] <- "CHRO"
sec201305$GEOG[sec201305$SEC.Region %in% c("LARO","SFRO")] <- "CARO"

# recode the variable SEC.Region from 201405
sec201405$GEOG <- "OTHER"
sec201405$GEOG[sec201405$SEC.Region %in% c("NYRO")] <- "NYRO"
sec201405$GEOG[sec201405$SEC.Region %in% c("BRO")] <- "BRO"
sec201405$GEOG[sec201405$SEC.Region %in% c("CHRO")] <- "CHRO"
sec201405$GEOG[sec201405$SEC.Region %in% c("LARO","SFRO")] <- "CARO"
```

1. For each of the regions (or by region), please show us in tabular form and by using a graphic how question 7B (possibly called X7B) changed from 2012 to 2014.

```
# For 2012
## clean up missing values
sec201205$X7B[sec201205$X7B==""] <- NA
sec201205$X7B <- as.factor(sec201205$X7B)
prop.table(table(sec201205$X7B))
```

```
##
##      N      Y
## 0.6696 0.3304
```

```
A1 <- table(sec201205$GEOG,sec201205$X7B)
prop.table(A1,1) ## row.proportions
```

```
##
##               N      Y
##   BRO   0.5949 0.4051
##   CARO  0.7159 0.2841
##   CHRO  0.7922 0.2078
##   NYRO  0.4380 0.5620
##   OTHER 0.7451 0.2549
```

```
chisq.test(A1)
```

```
##
##  Pearson's Chi-squared test
##
## data:  A1
## X-squared = 897.7, df = 4, p-value < 2.2e-16
```

```
# For 2013
sec201305$X7B[sec201305$X7B ==""] <- NA
sec201305$X7B <- as.factor(sec201305$X7B)
prop.table(table(sec201305$X7B))
```

```
##
##      N      Y
## 0.6209 0.3791
```

```
A2 <- table(sec201305$GEOG,sec201305$X7B)
prop.table(A2,1) ## row.proportions
```

```
##
##               N      Y
##   BRO   0.5321 0.4679
##   CARO  0.6580 0.3420
##   CHRO  0.7555 0.2445
##   NYRO  0.4353 0.5647
##   OTHER 0.6968 0.3032
```

```
chisq.test(A2)
```

```
##
##  Pearson's Chi-squared test
##
## data:  A2
## X-squared = 623.5, df = 4, p-value < 2.2e-16
```

```r
# For 2014
sec201405$X7B[sec201405$X7B==""] <- NA
sec201405$X7B <- as.factor(sec201405$X7B)
prop.table(table(sec201405$X7B))
```

```
##
##      N      Y
## 0.6212 0.3788
```

```r
A3 <- table(sec201405$GEOG,sec201405$X7B)
prop.table(A3,1) ## row.proportions
```

```
##
##               N      Y
##   BRO    0.5378 0.4622
##   CARO   0.6640 0.3360
##   CHRO   0.7519 0.2481
##   NYRO   0.4407 0.5593
##   OTHER 0.6901 0.3099
```

```r
chisq.test(A3)
```

```
##
##  Pearson's Chi-squared test
##
## data:  A3
## X-squared = 605.8, df = 4, p-value < 2.2e-16
```

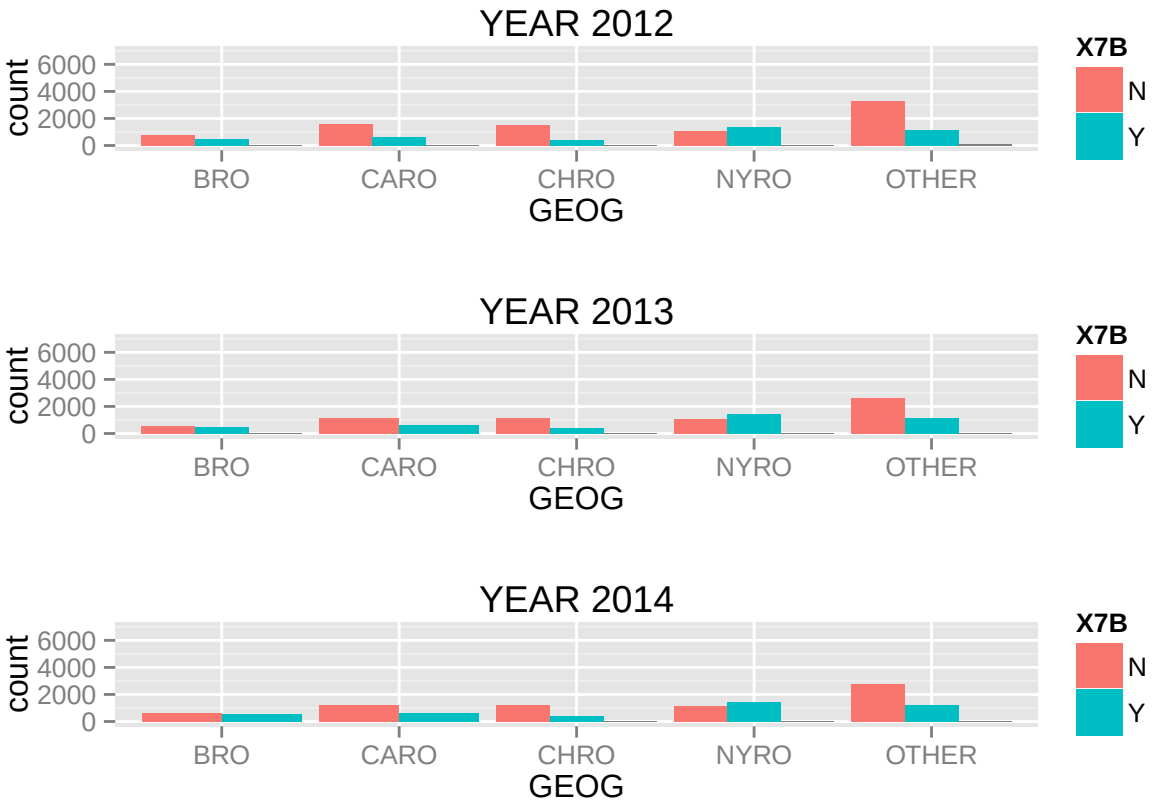
```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
require(grid)
```

```
## Loading required package: grid
```

```r
vplayout <- function(x, y) viewport(layout.pos.row = x, layout.pos.col = y)

plot1 <- ggplot(sec201205, aes(GEOG, fill = X7B)) + geom_bar(position = "dodge") + ggtitle("YEAR 2012")

plot2 <- ggplot(sec201305, aes(GEOG, fill = X7B)) + geom_bar(position = "dodge") + ggtitle("YEAR 2013")

plot3 <- ggplot(sec201405, aes(GEOG, fill = X7B)) + geom_bar(position = "dodge") + ggtitle("YEAR 2014")

grid.newpage()
pushViewport(viewport(layout = grid.layout(3, 1)))
print(plot1, vp = vplayout(1, 1))
print(plot2, vp = vplayout(2, 1))
print(plot3, vp = vplayout(3, 1))
```



2. For each of the regions (or by region), please describe the differences between firms who answered "Y" to 7B and those who answered "N" in 2012 and compare them with firms to answered in 2014.

```
prop.table(A1,1)[,1]
```

```
##    BRO   CARO   CHRO   NYRO  OTHER
## 0.5949 0.7159 0.7922 0.4380 0.7451
```

```
prop.table(A1,1)[,2]
```

```
##    BRO   CARO   CHRO   NYRO  OTHER
## 0.4051 0.2841 0.2078 0.5620 0.2549
```

```
#Difference between who answer "NO" and "YES" in 2012
dif2012 <- c(prop.table(A1,1)[,1] - prop.table(A1,1)[,2])
dif2012
```

```
##     BRO    CARO    CHRO    NYRO   OTHER
##  0.1898  0.4318  0.5843 -0.1240  0.4901
```

```r
#Difference between who answer "NO" and "YES" in 2014
dif2014 <- c(prop.table(A3,1)[,1] - prop.table(A3,1)[,2])
dif2014
```

```
##      BRO     CARO     CHRO     NYRO    OTHER
##  0.07569  0.32790  0.50370 -0.11862  0.38013
```

```r
diff <- rbind(dif2012,dif2014)
diff
```

```
##              BRO   CARO   CHRO    NYRO  OTHER
## dif2012 0.18975 0.4318 0.5843 -0.1240 0.4901
## dif2014 0.07569 0.3279 0.5037 -0.1186 0.3801
```

Based on our table, we see that the differences between firms who answered "Y" to 7B and those who answered "N" in 2012 are higher than 2014.

We can conlude that for BR, CA, CH, AND OTHER regions, more people become to an adviser to private fund in 2014, but for NY region, the number of people become to an adviser is decresed.

3. For each of the regions (or by region), please show us in tabular form and by using a graphic how question 3A (possibly called X3A) changed from 2012 to 2014.

```r
sec201205$X3A[sec201205$X3A==""] <- NA
sec201305$X3A[sec201305$X3A==""] <- NA
sec201405$X3A[sec201405$X3A==""] <- NA

table(sec201205$X3A,sec201205$GEOG)
```

```
##
##                                BRO CARO CHRO NYRO OTHER
##   Corporation                  436  904  885  688  2133
##   Limited Liability Company    691 1176  958 1488  1915
##   Limited Liability Partnership  11    7    6    5    90
##   Limited Partnership            64   57   30  252   145
##   Other (specify)                13   23   17   30   158
##   Partnership                    19   28   11   51    65
##   Sole Proprietorship            34   75   24   25    84
```

```r
table(sec201305$X3A,sec201305$GEOG)
```

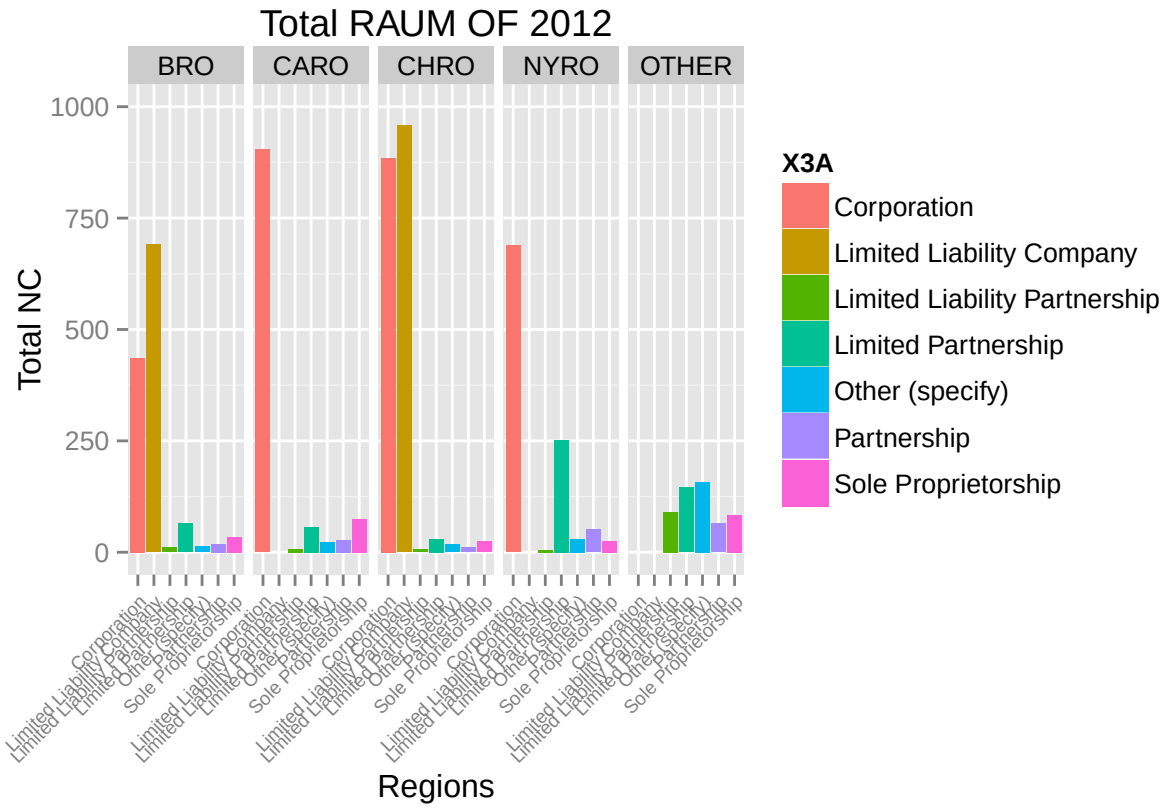
```
##
##                                BRO CARO CHRO NYRO OTHER
##   Corporation                  345  658  689  632  1663
##   Limited Liability Company    630  976  805 1483  1611
##   Limited Liability Partnership  10    9    7    4    96
##   Limited Partnership            70   71   31  294   149
```

```
##    Other (specify)                 11    11    14    20    155
##    Partnership                     13    19     8    39     44
##    Sole Proprietorship             12    34     3    24     40
```

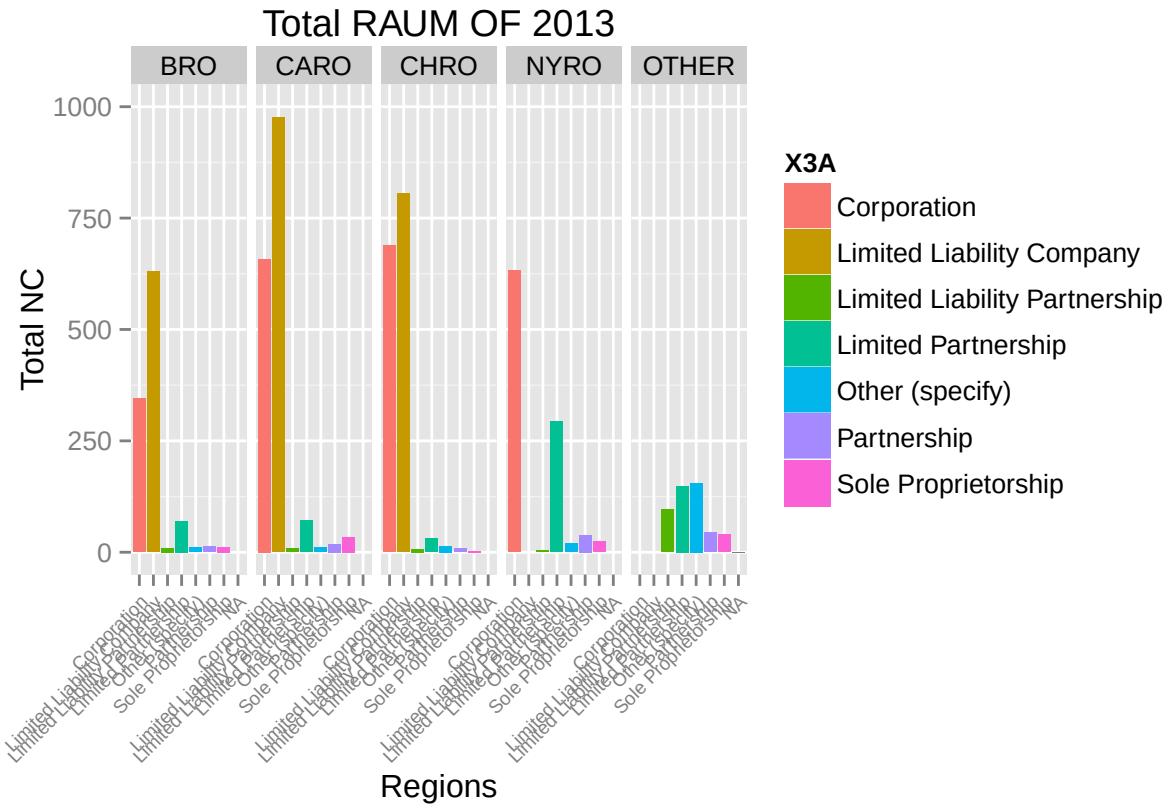
```
table(sec201405$X3A,sec201405$GEOG)
```

```
##
##                                 BRO CARO CHRO NYRO OTHER
##    Corporation                  341  656  696  624  1677
##    Limited Liability Company    658 1036  863 1568  1778
##    Limited Liability Partnership 11    7    4    6   108
##    Limited Partnership          78   84   38  338   160
##    Other (specify)              12   11   13   20   182
##    Partnership                  11   15    5   29    33
##    Sole Proprietorship          12   33    2   21    40
```

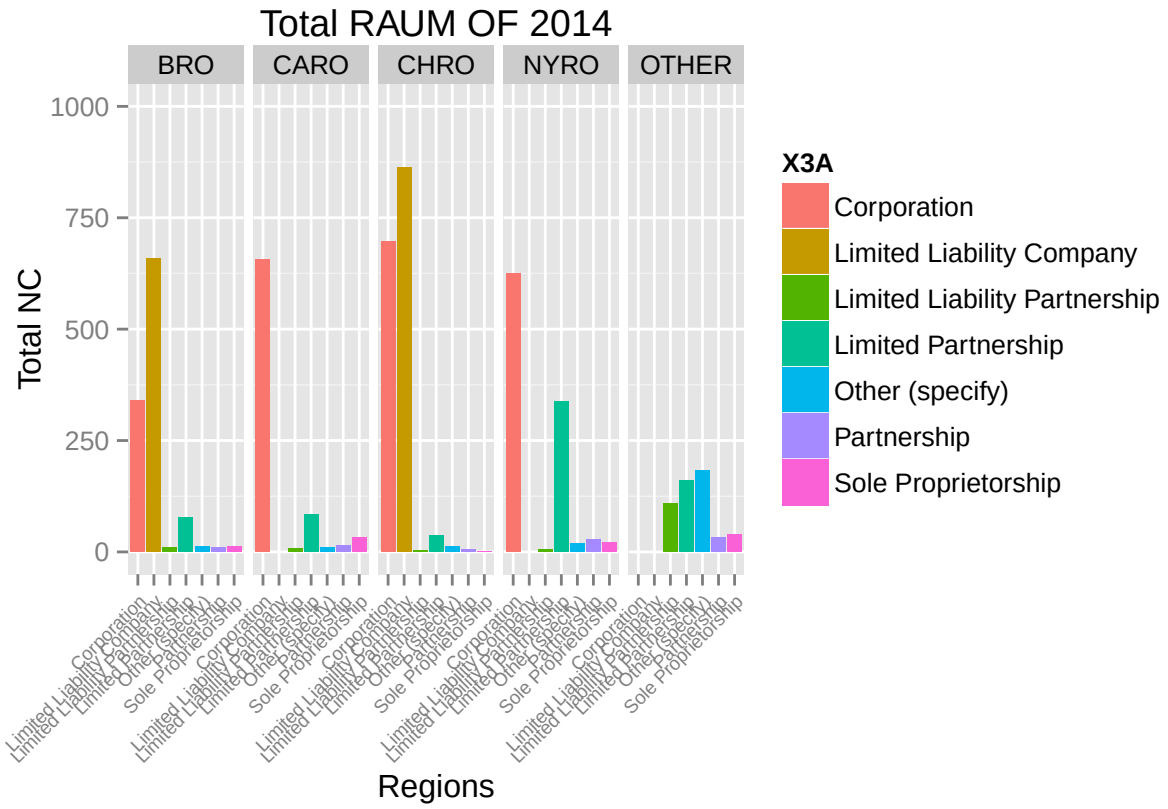
```
ggplot(sec201205, aes(x= X3A, fill= X3A)) + geom_bar(position = "dodge")  + facet_grid(. ~ GEOG) + xlab
```




```
ggplot(sec201305, aes(x= X3A, fill= X3A)) + geom_bar(position = "dodge")  + facet_grid(. ~ GEOG) + xlab
```




```
ggplot(sec201405, aes(x= X3A, fill= X3A)) + geom_bar(position = "dodge")  + facet_grid(. ~ GEOG) + xlab
```



```
# Percentage of each form of organization by Region
# 2012-2014
round(prop.table(table(sec201205$X3A,sec201205$GEOG),1), digits = 2)
```

```
##
##                                 BRO CARO CHRO NYRO OTHER
##   Corporation                  0.09 0.18 0.18 0.14  0.42
##   Limited Liability Company    0.11 0.19 0.15 0.24  0.31
##   Limited Liability Partnership 0.09 0.06 0.05 0.04  0.76
##   Limited Partnership          0.12 0.10 0.05 0.46  0.26
##   Other (specify)              0.05 0.10 0.07 0.12  0.66
##   Partnership                  0.11 0.16 0.06 0.29  0.37
##   Sole Proprietorship          0.14 0.31 0.10 0.10  0.35
```

```
round(prop.table(table(sec201405$X3A,sec201405$GEOG),1), digits = 2)
```

```
##
##                                 BRO CARO CHRO NYRO OTHER
##   Corporation                  0.09 0.16 0.17 0.16  0.42
##   Limited Liability Company    0.11 0.18 0.15 0.27  0.30
##   Limited Liability Partnership 0.08 0.05 0.03 0.04  0.79
##   Limited Partnership          0.11 0.12 0.05 0.48  0.23
##   Other (specify)              0.05 0.05 0.05 0.08  0.76
##   Partnership                  0.12 0.16 0.05 0.31  0.35
##   Sole Proprietorship          0.11 0.31 0.02 0.19  0.37
```

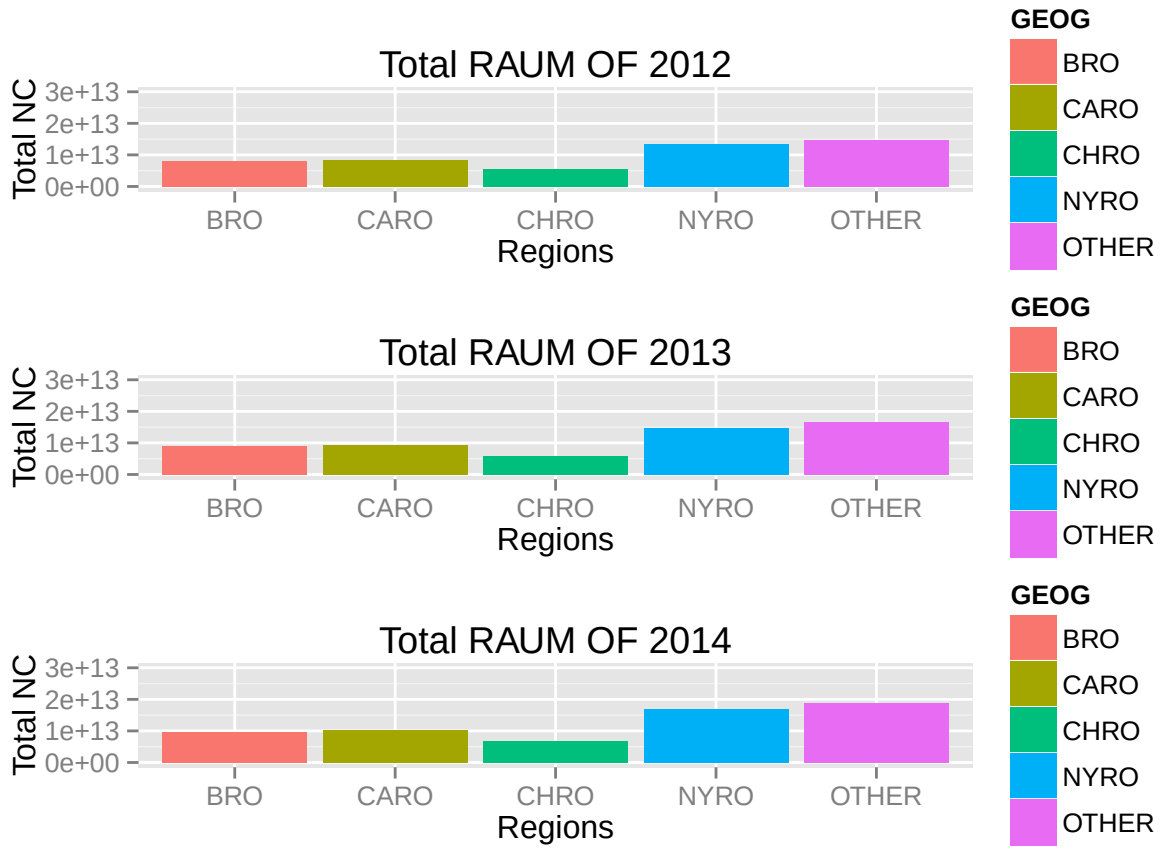
5. For each of the regions (or by region), please show us in tabular form and by using a graphic how question 5F2c (possibly called X5F.2..c) changed from 2012 to 2014.

```
require(scales)
```

```
## Loading required package: scales
```

```
sec201205[,7] <- as.numeric(gsub(",","",sec201205[,7]))

sec201305[,7] <- as.numeric(gsub(",","",sec201305[,7]))

sec201405[,7] <- as.numeric(gsub(",","",sec201405[,7]))

vplayout <-function(x, y) viewport(layout.pos.row = x, layout.pos.col = y)

plot1 <- ggplot(sec201205, aes(x = GEOG, y = X5F.2..c., fill = GEOG)) + geom_bar(stat = "identity") + xl

plot2 <- ggplot(sec201305, aes(x = GEOG, y = X5F.2..c., fill = GEOG)) + geom_bar(stat = "identity") + xl

plot3 <- ggplot(sec201405, aes(x = GEOG, y = X5F.2..c., fill= GEOG)) + geom_bar(stat = "identity") + xla
```

```
grid.newpage()
pushViewport(viewport(layout = grid.layout(3, 1)))
print(plot1, vp = vplayout(1, 1))
print(plot2, vp = vplayout(2, 1))
print(plot3, vp = vplayout(3, 1))
```




```
t1 <- tapply(sec201205$X5F.2..c.,sec201205$GEOG, sum)
t2 <- tapply(sec201305$X5F.2..c.,sec201305$GEOG, sum)
t3 <- tapply(sec201405$X5F.2..c.,sec201405$GEOG, sum)
total <- c(t1,t2,t3)

# total RAUM in 2012-2014
RAUM_total <- matrix(dollar_format()(total), nrow = 3, ncol= 5, byrow =T)
colnames(RAUM_total) <- c("CA", "IL", "MA", "NY", "OTHER")
rownames(RAUM_total) <- c("2012", "2013" ,"2014")
print(RAUM_total)
```

```
##      CA                   IL                   MA
## 2012 "$7,891,412,644,329" "$8,220,656,487,063"  "$5,308,037,130,528"
## 2013 "$8,757,231,214,866" "$9,118,990,026,788"  "$5,832,807,597,007"
## 2014 "$9,604,456,599,436" "$10,064,574,613,442" "$6,762,010,185,528"
##      NY                    OTHER
## 2012 "$13,159,502,864,322" "$14,628,526,667,115"
## 2013 "$14,670,899,447,235" "$16,473,863,887,811"
## 2014 "$16,692,675,190,653" "$18,617,699,695,874"
```

```r
# percentage of total RAUM in 2012-2014
rate_increase <- t3/t1-1
rate_increase
```

```
##    BRO   CARO   CHRO   NYRO  OTHER
## 0.2171 0.2243 0.2739 0.2685 0.2727
```

Based on the plot and table,NY and CA aloways have the most toalt RAUM from 2012 to 2014, alos their RAUM gradually increased from 2012 t0 2014, CA increased 22%, and NY increased 26%.