Project01Wan

Wan

January 22, 2016

This is the data of comsumer complaints about finance institution. It is downloaded from <https://catalog.data.gov/dataset/consumer-complaint-database>.

Consumer\_Complaints <- read.csv("/media/wan/64GB/R\_CLASS/R\_Project\_Data/Consumer\_Complaints.csv", stringsAsFactors=FALSE)  
  
#separate into Year and Month  
Consumer\_Complaints$year <- substr(Consumer\_Complaints$Date.received, 7, 10)  
Consumer\_Complaints$month <- substr(Consumer\_Complaints$Date.received, 1, 2)  
Consumer\_Complaints  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

the data format looks like "03/11/2015"

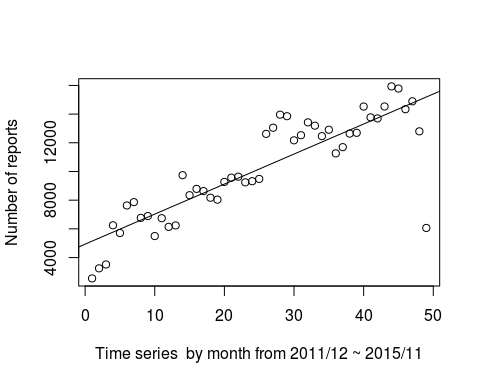
Table 1. Consumer complaints data group by Year.

dateData <- select(Consumer\_Complaints, year, month)  
groupByYear <- group\_by(dateData, year) %>% summarise(., yearNum = n())  
groupByYear

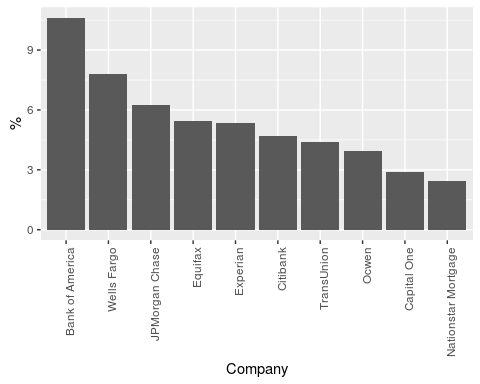
## Source: local data frame [5 x 2]  
##   
## year yearNum  
## (chr) (int)  
## 1 2011 2549  
## 2 2012 72523  
## 3 2013 108272  
## 4 2014 153164  
## 5 2015 161703

The complaints report increase from 2012 to 2015. 2011 has only Decomber. So I checked by month nested in year.

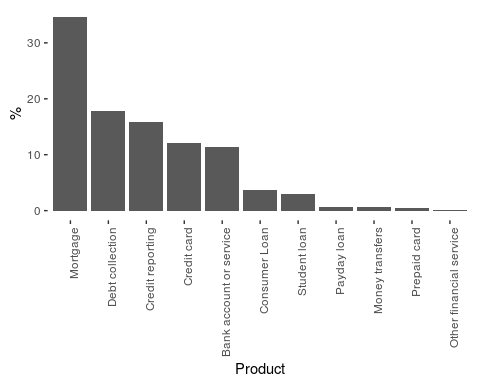
ByYearMonth <- group\_by(dateData, year, month) %>% summarise(., total = n())  
ByYearMonth$seq <- c(1:nrow(ByYearMonth))  
  
regre <- lm(total ~ seq, data=ByYearMonth)  
plot(total ~ seq, data = ByYearMonth, ylab = "Number of reports", xlab = "Time series by month from 2011/12 ~ 2015/11")  
abline(regre)

  
The total complaints number is regressed against time.

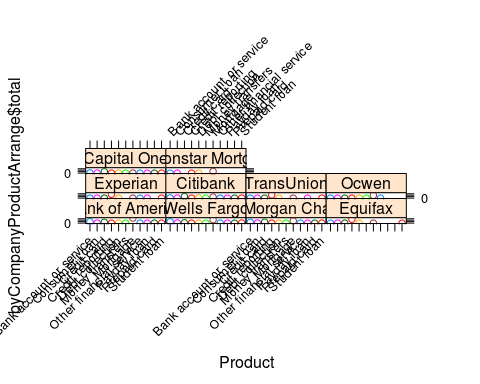
#top10 company  
Consumer\_complainYearMonth <- read.csv("/media/wan/64GB/Documents/DataScience/porject/porject01/Consumer\_complainYearMonth.csv")  
companyRank <- group\_by(Consumer\_complainYearMonth, Company) %>% summarise(., total = n()) %>% arrange(., desc(total) )  
top10Company <- companyRank[1:10,]  
top10Company$Company <- factor(top10Company$Company, levels = top10Company$Company)  
top10Company$percentage <- top10Company$total/nrow(Consumer\_complainYearMonth)\*100  
  
library(ggplot2)  
barCompany <- ggplot(top10Company, aes(x=Company, y=percentage)) + geom\_bar(stat="identity")  
barCompany + theme(axis.text.x = element\_text(angle = 90, hjust = 1)) + ylab("%")



byProduct <- group\_by(Consumer\_complainYearMonth, Product) %>% summarise(., total= n() ) %>% arrange(., desc(total))  
byProduct$Product <- factor(byProduct$Product, levels = byProduct$Product)  
  
byProduct$Percentage <- byProduct$total/nrow(Consumer\_complainYearMonth)\*100  
barPro <- ggplot(byProduct, aes(x=Product, y=Percentage)) + geom\_bar(stat="identity")  
barPro + theme(axis.text.x = element\_text(angle = 90, hjust = 1), panel.background = element\_blank()) + ylab("%")



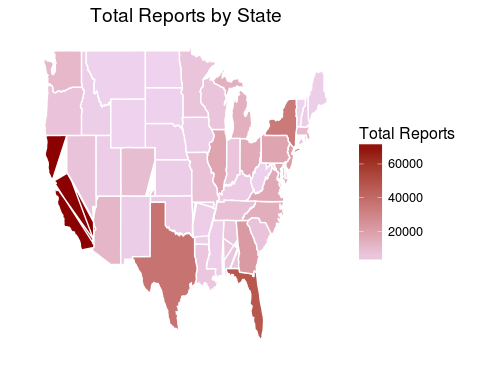
CcTop10Company <- filter(Consumer\_complainYearMonth, Company %in% top10Company$Company)  
CcTop10Company$Company <- factor(CcTop10Company$Company, levels = top10Company$Company)  
byCompanyProduct <- group\_by(CcTop10Company, Company, Product)   
byCompanyProductTotal <- summarise(byCompanyProduct, total = n())  
byCompanyProductArrange <- arrange(byCompanyProductTotal, desc(total))  
  
library(ggplot2)  
library(lattice)  
groupXYplot <- xyplot(byCompanyProductArrange$total ~ byCompanyProductArrange$Product | byCompanyProductArrange$Company, data=byCompanyProductArrange, xlab = list(label = "Product", cex=1), scales=list(x=list(rot=45)), groups = Product)  
groupXYplot



onlyState <- select(Consumer\_complainYearMonth, State)  
byState <- group\_by(onlyState, State) %>% summarise(., total = n())  
byState <- byState[-1,]  
  
  
onlyState <- select(Consumer\_complainYearMonth, State)  
byState <- group\_by(onlyState, State) %>% summarise(., total = n())  
byState <- byState[-1,]  
  
stateAbb <- read.csv("/media/wan/64GB/R\_CLASS/Project/data/stateAbb.csv", stringsAsFactors=FALSE)  
fullName <- c()  
total <- c()  
byStateFullname <- data.frame(fullName, total)  
for (i in 1: nrow(byState)) {  
 posi <- match(byState$State[i], stateAbb$abbreviation)  
 if (!is.na(posi)) {  
 lastPosi <- length(fullName) + 1  
 fullName[lastPosi] <- stateAbb$state.name[posi]  
 total[lastPosi] <- byState$total[i]  
 }   
}  
  
byState <- data.frame(fullName, total)  
  
library(maps)

##   
## # ATTENTION: maps v3.0 has an updated 'world' map. #  
## # Many country borders and names have changed since 1990. #  
## # Type '?world' or 'news(package="maps")'. See README\_v3. #

library(ggplot2)  
all\_state <-map\_data("state")  
byState$region <- byState$fullName  
Total <- merge(all\_state, byState, by="region")  
Total <- Total[Total$region!="district of columbia",]  
p <- ggplot()  
p <- p + geom\_polygon(data=Total, aes(x=long, y=lat, group = group, fill=Total$total),colour="white"  
 ) + scale\_fill\_continuous(low = "thistle2", high = "darkred", guide="colorbar")  
p1 <- p + theme\_bw() + labs(fill = "Total Reports"   
 ,title = "Total Reports by State", x="", y="")  
p1 + scale\_y\_continuous(breaks=c()) + scale\_x\_continuous(breaks=c()) + theme(panel.border = element\_blank())



byMediaYearMonth <- group\_by(Consumer\_complainYearMonth, year, month, Submitted.via) %>% summarise(., total = n())  
byMediaYearMonth <- arrange(byMediaYearMonth, total)  
  
byMediaYearMonth$month <- as.character(byMediaYearMonth$month)  
for (i in 1: nrow(byMediaYearMonth)) {  
 if ( nchar(byMediaYearMonth$month[i], type = "chars", allowNA = FALSE, keepNA = NA) == 1) {  
 byMediaYearMonth$month[i] <- paste(0, byMediaYearMonth$month[i], sep = "" )  
 }  
}  
  
byMediaYearMonth$yymm <- paste(byMediaYearMonth$year, byMediaYearMonth$month, sep = "/")  
byMediaYearMonth <- filter(byMediaYearMonth, yymm != "2015/12")  
  
graph <- qplot(yymm, total, data = byMediaYearMonth, color = Submitted.via)  
graph45 <- graph + theme(axis.text.x = element\_text(angle = 45,size = 14, hjust = 1), panel.background = element\_blank())  
  
yymm <- byMediaYearMonth$yymm  
yymm <- unique(yymm)  
YearMonth <- data.frame(yymm, stringsAsFactors=FALSE)  
  
YearMonth$yymm <- as.character(YearMonth$yymm)  
ticks5 <- c()  
lastRow <- nrow(YearMonth)  
for (i in 1: lastRow) {  
 if (i %% 5 == 0) {  
 ticks5[i] <- YearMonth$yymm[i]  
 cat("yymm", YearMonth$yymm[i])  
 } else {  
 ticks5[i] <- ""  
 }  
}

## yymm 2012/04yymm 2012/09yymm 2013/02yymm 2013/07yymm 2013/12yymm 2014/05yymm 2014/10yymm 2015/03yymm 2015/08

graph45five <- graph45 + scale\_x\_discrete(breaks = ticks5,labels =ticks5)  
graph45five+ theme(axis.text.x = element\_text(angle = 45, hjust = 1))

