

MA615__DSR__HW3

Qianhui Rong, Fionnuala Mcpeake, Yifeng Luo, Yaotang Luo

2018/10/1

Libraries

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.0.0      v purrr  0.2.5
## v tibble  1.4.2      v dplyr  0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(scales)

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor

library(ggplot2)
```

Importing csv

```
data <- read.csv("C:\\Users\\xiaoh\\Desktop\\615h3w\\SARStats1500.csv", header = TRUE)
View(data)
```

The data cleaning part is a group work. We've worked together to determine data's selection criteria.

Data cleaning

The main goal of data cleaning is to eliminate observations with string of "[Total]", which is similar to NA.

```
data %>% filter(State!="[Total]"&
                Industry!="[Total]"&
                Suspicious.Activity!="[Total]"&
                Regulator!="[Total]"&
                Relationship!="[Total]"&
                Product!="[Total]"&
```

```
Instrument!="[Total]") -> clean_data
View(clean_data)
```

Check if data still has “[Total]”

```
match("[Total]", clean_data)
```

```
## [1] NA
```

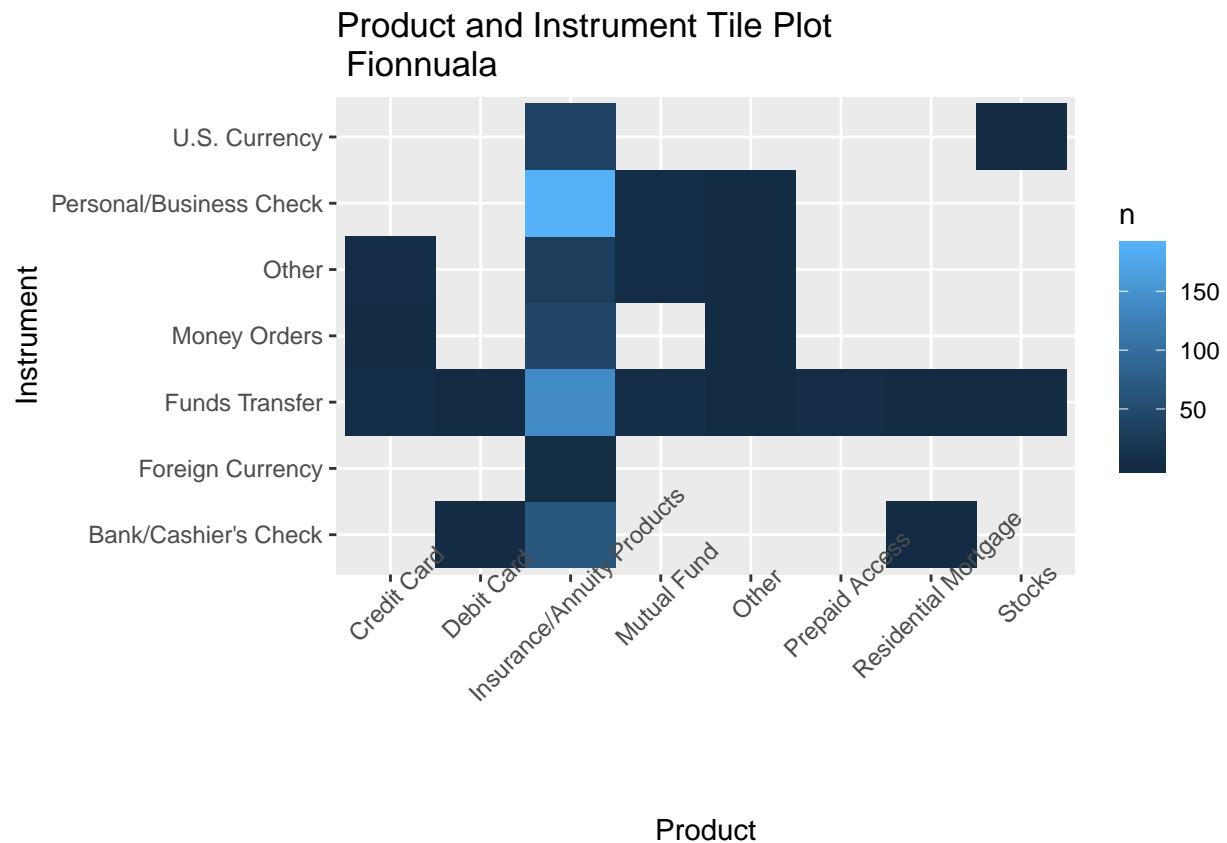
Our work is divided into four parts. Each of us takes charge of one combination of two variables to analyze the co-influence of these two variable on the occurrences of financial crimes.

Fionnuala is in charge of the combination of Product and Intrument. Yifeng is in charge of the combination of Suspicious.Activity and Product. Yaotang is in charge of the combination of Relationship and Suspicious.Activity. Qianhui is in charge of the combination of Relationship and Instrument.

An explanation of all variables we use in this assignment: (Reference: <https://www.fincen.gov/reports/sar-stats>) -Suspicious Acitivity: the criminal activity type; -Instruments: the instrument or mechanism used in suspicious activities; -Products: the product types used in suspicious activities; -Relationship: relationship of the subject to the institution.

Part 1 (Fionnuala Mcpeake)

```
clean_data %>%
  count(Product, Instrument) %>%
  ggplot(mapping = aes(x = Product, y = Instrument))+
  geom_tile(mapping = aes(fill = n))+
  theme(axis.text.x=(element_text(angle = 45)))+
  ggtitle("Product and Instrument Tile Plot \n Fionnuala")
```



```
count(clean_data, Product)
```

```
## # A tibble: 8 x 2
##   Product      n
##   <fct>    <int>
## 1 Credit Card      8
## 2 Debit Card       2
## 3 Insurance/Annuity Products  507
## 4 Mutual Fund     14
## 5 Other           4
## 6 Prepaid Access   3
## 7 Residential Mortgage  2
## 8 Stocks          4
```

```
count(clean_data, Instrument)
```

```
## # A tibble: 7 x 2
##   Instrument      n
##   <fct>    <int>
## 1 Bank/Cashier's Check    69
## 2 Foreign Currency       5
## 3 Funds Transfer    156
## 4 Money Orders       45
## 5 Other            38
## 6 Personal/Business Check 192
## 7 U.S. Currency       39
```

Visualization and Discussion

In order to determine the most frequently used combination of product and instrument used to commit fraud, a tile chart was created. As some combinations of product and instrument are not possible, such as using money orders to create stock fraud, there are blank spaces in the tile chart. As there is not a large number of combinations that can be made, due in part to the facts that both variables are categorical, and both categories contain a reasonably small number of possibilities, it was not necessary to use the `geom_bin2d()` or `geom_hex()` functions in R.

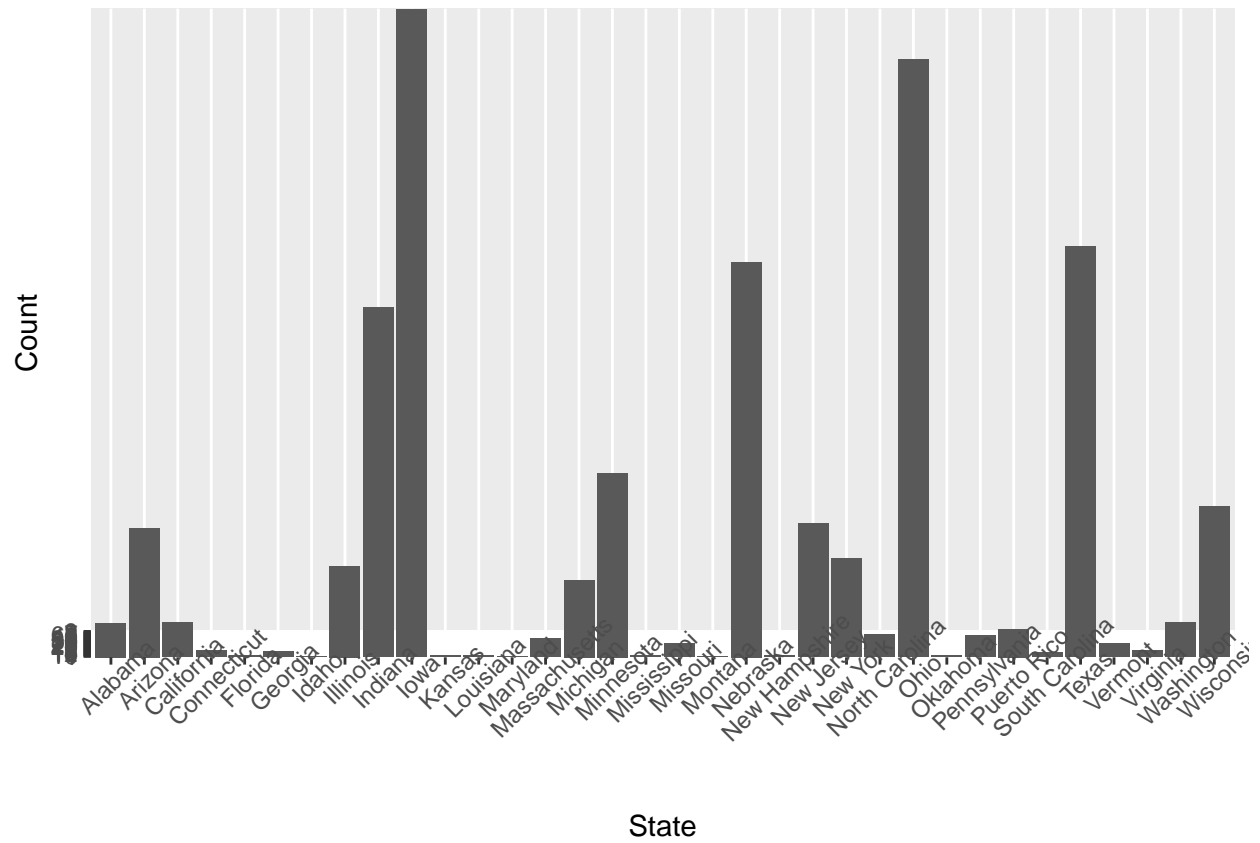
From the tile chart, it appears that the instrument funds transfer could be used with any product considered in the cleaned data set, and that it was most likely to be used with the product insurance/annuity products. However, this is not the most commonly used instrument that was used for the product insurance/annuity products, which is actually personal/business checks, which had 192 instances. Overall, the insurance/annuity products are the most likely to be used out of any, with 507 instances. Despite funds transfer being used in conjunction with every product, it is only the second most common instrument used, with personal/business check having 36 more occurrences. This implies that the use of the combination of personal/business check and insurance/annuity products are much more prevalent, making up for the gap between the top two most common instruments.

Conclusion

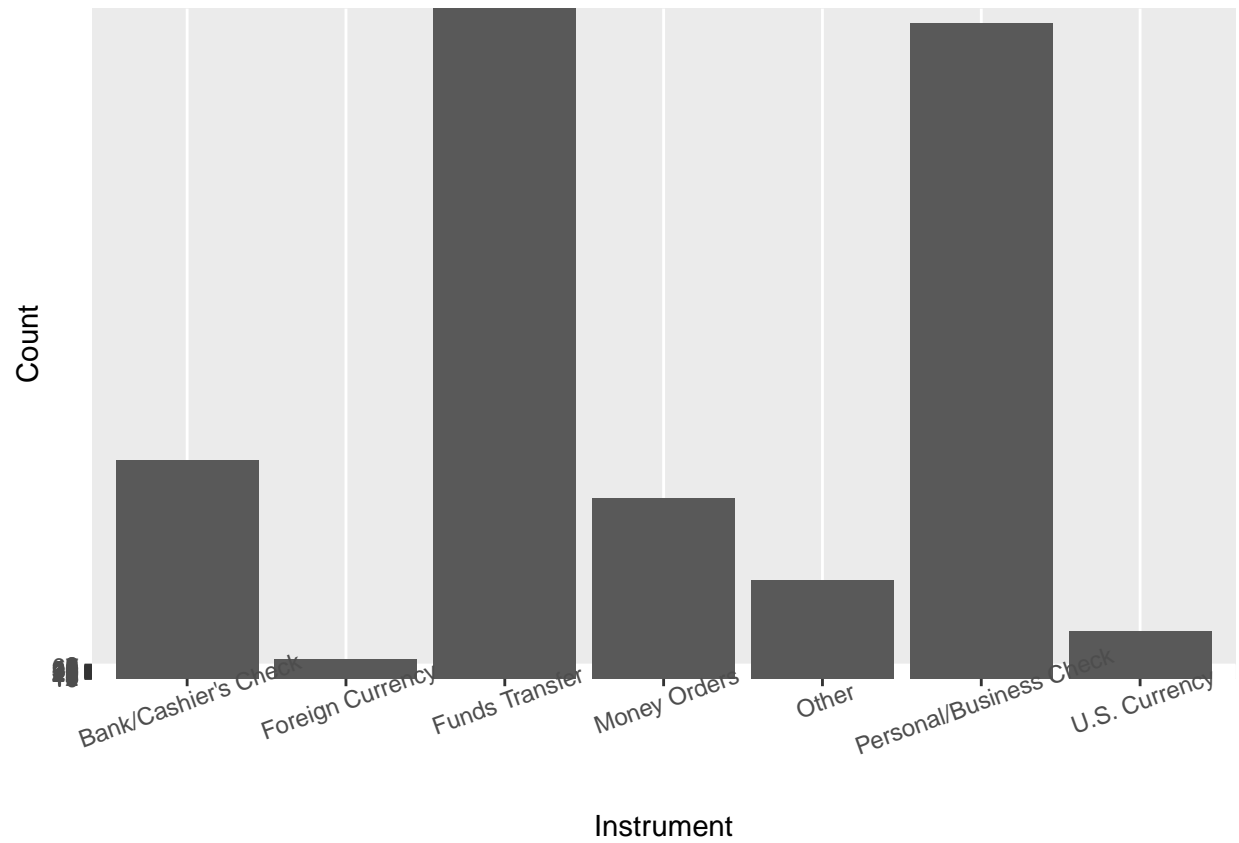
If a person wanted to reduce financial crimes, additional information would be needed in conjunction with the data used in this report. The findings in this paper indicate what the most common crimes are, and what tools are used to commit them. In order to prevent these crimes, more knowledge on exactly how these crimes are usually performed would be needed, in addition to determining signals that indicate if a crime is being committed, so an oversight body would be able to investigate while being efficient with their resources.

Part 2 (Yifeng Luo)

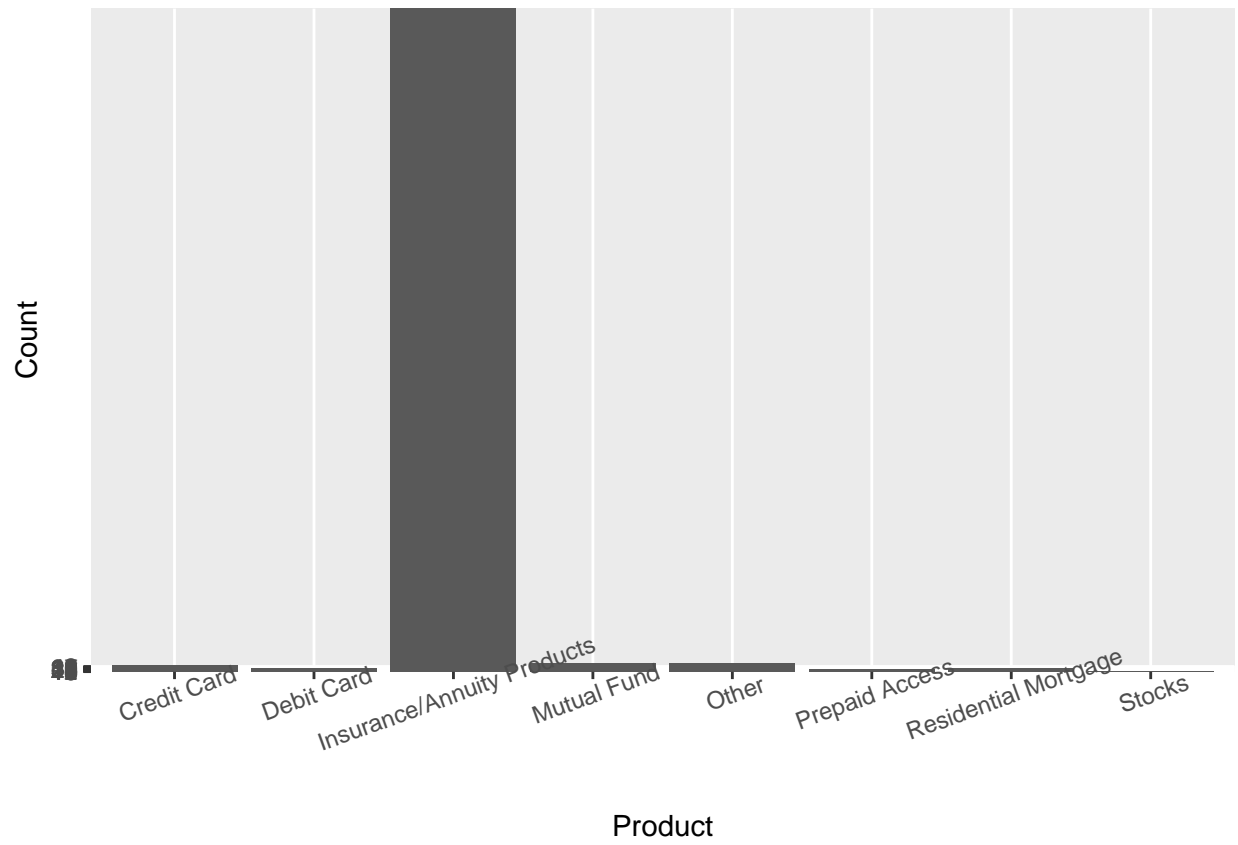
```
#We want to check state, instruments, products and suspicious.activity  
#Bar charts  
par(mfrow=c(2,2))  
ggplot(data = clean_data)+geom_bar(mapping = aes(y=Count,x=State),stat = "identity")+theme(axis.text.x=
```



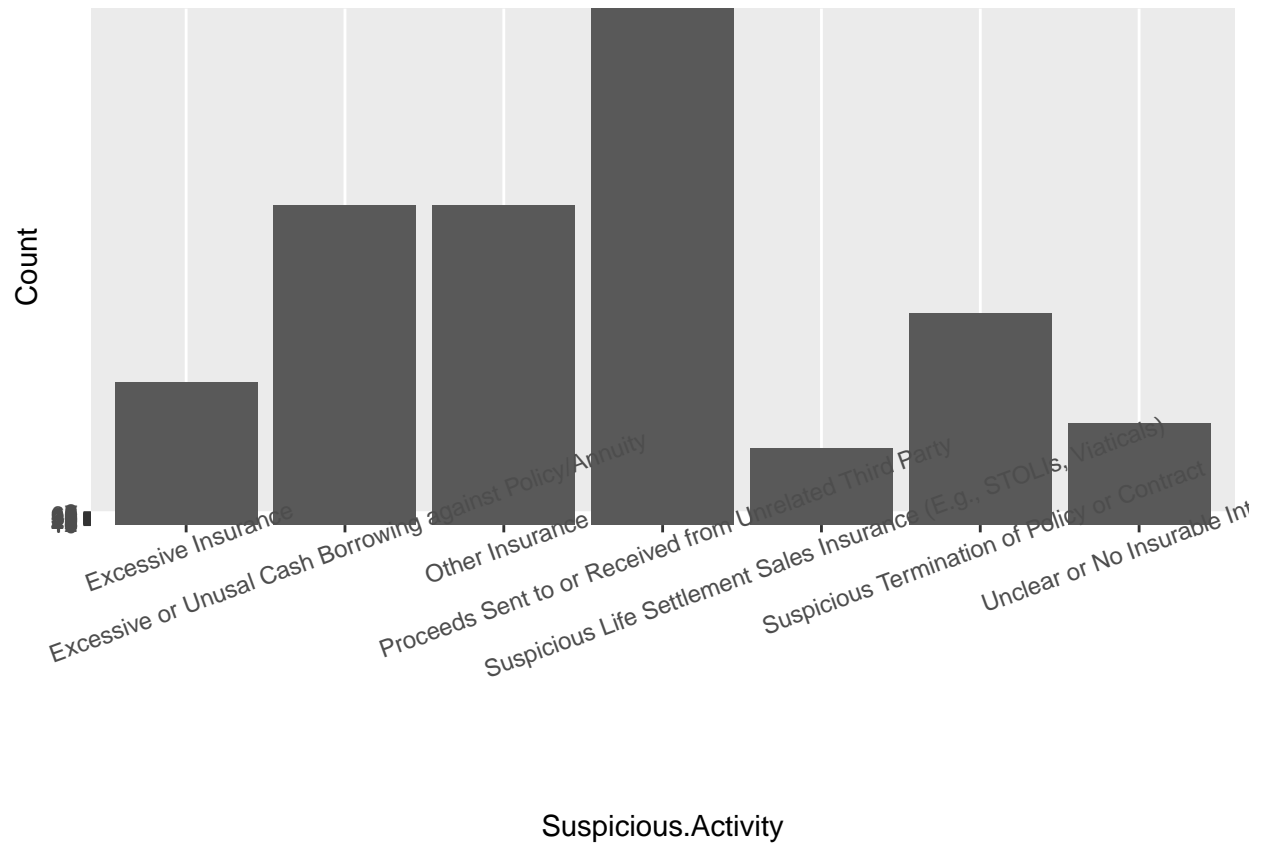
```
ggplot(data = clean_data)+geom_bar(mapping = aes(y=Count,x=Instrument),stat = "identity")+theme(axis.te
```



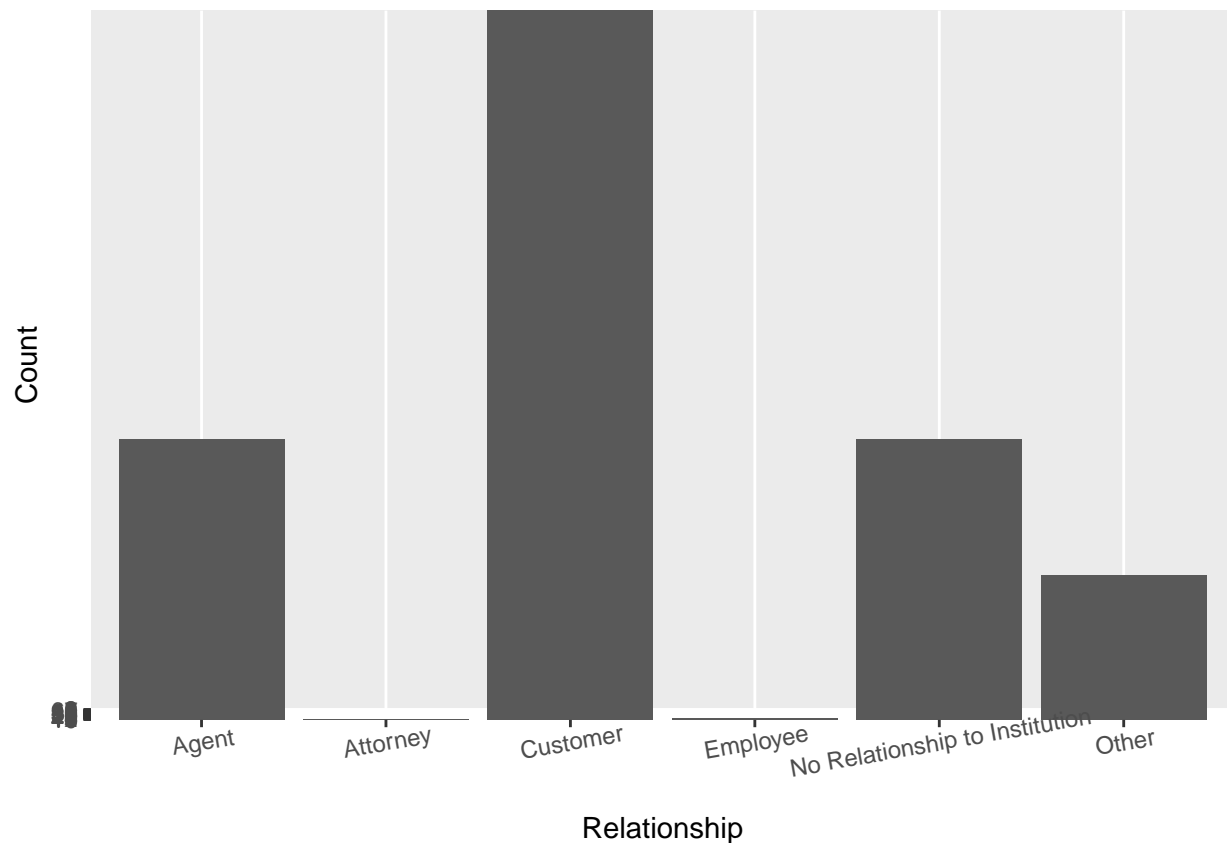
```
ggplot(data = clean_data)+geom_bar(mapping = aes(y=Count,x=Product),stat = "identity")+theme(axis.text.x =
```



```
ggplot(data = clean_data)+geom_bar(mapping = aes(y=Count,x=Suspicious.Activity),stat = "identity")+them
```



```
ggplot(data=clean_data)+geom_bar(mapping = aes(y=Count,x=Relationship),stat='identity')+theme(axis.text
```

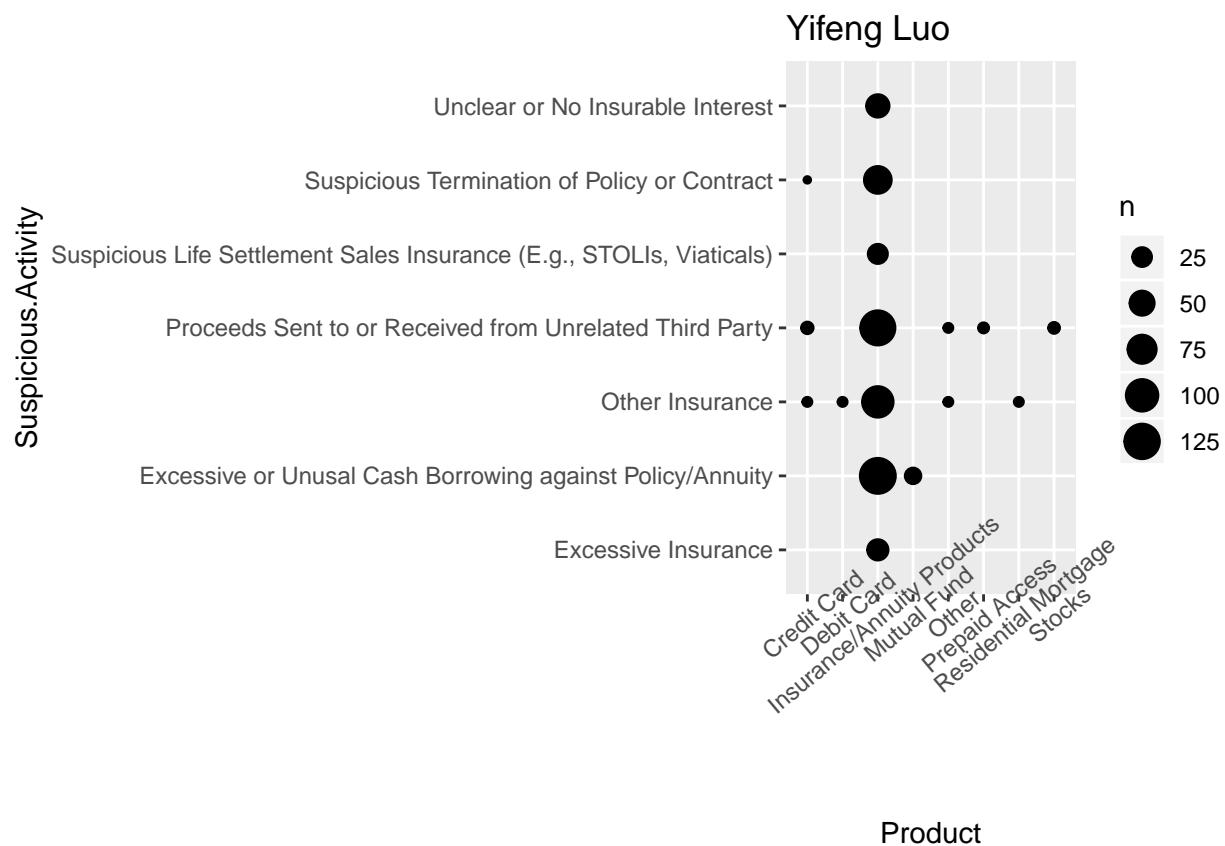



```
#suspicious and product
summary(clean_data)
```

```
##          Year.Month      State      Industry
## 2017          :328    Ohio      : 77    Insurance Company:544
## 2018          :216    Iowa      : 74
## All Years [Total]: 0    Nebraska : 53
##                               Indiana : 40
##                               Texas   : 40
##                               Minnesota: 29
##                               (Other) :231
##
##                               Suspicious.Activity
## Excessive or Unusal Cash Borrowing against Policy/Annuity:143
## Proceeds Sent to or Received from Unrelated Third Party :136
## Other Insurance                                           :100
## Suspicious Termination of Policy or Contract              : 67
## Unclear or No Insurable Interest                          : 41
## Excessive Insurance                                        : 31
## (Other)                                                    : 26
##
## Regulator      Relationship
## [Total]: 0      [Total]      : 0
## FRB      : 2      Agent      : 97
## IRS      :501    Attorney     : 1
## OCC      : 9      Customer     :270
## SEC      : 32    Employee      : 3
##                               No Relationship to Institution:118
```

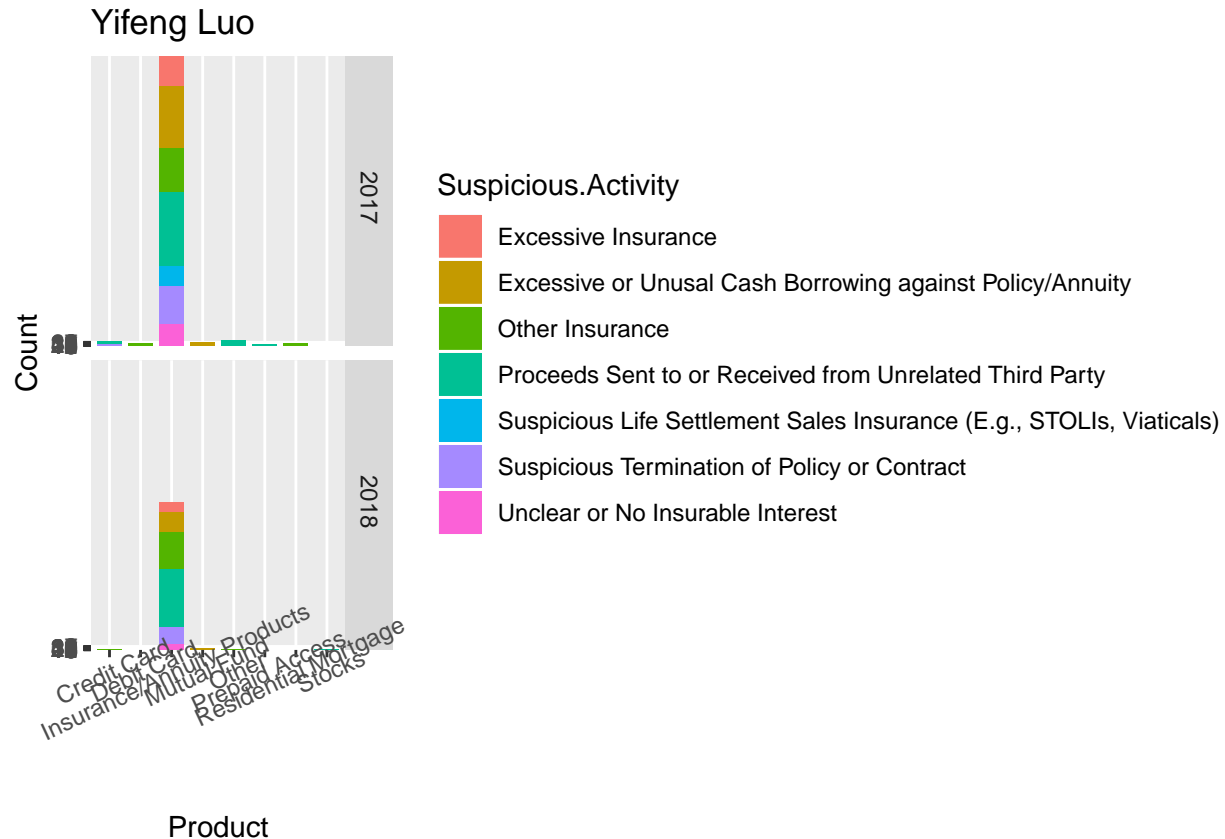
```
##          Other          : 55
##          Product          Instrument
## Insurance/Annuity Products:507 Personal/Business Check:192
## Mutual Fund          : 14 Funds Transfer          :156
## Credit Card          : 8 Bank/Cashier's Check    : 69
## Other                : 4 Money Orders          : 45
## Stocks              : 4 U.S. Currency          : 39
## Prepaid Access      : 3 Other                : 38
## (Other)             : 4 (Other)              : 5
##          Count
## 1          :291
## 2          :120
## 3          : 43
## 4          : 23
## 5          : 14
## 13         : 7
## (Other): 46
```

```
ggplot(data = clean_data) +
  geom_count(mapping = aes(x = Product, y = Suspicious.Activity)) + theme(axis.text.x=(element_text(angle = 45)))
```



From the graphs above, we can see among the 7 types of insurance product, the 'insurance/annuity product' is the most popular, moreover, it always appears with the suspicious activity in 'proceeds sent to or received from unrelated third party' and 'excessive or unusual cash borrowing against policy/annuity'. compare with the other insurance product, like 'prepaid access', 'residential mortgage', 'stock' and 'debit card' rare happend suspicious activity.

```
clean_data %>% select(Product,Suspicious.Activity,Count) -> clean_data1
ggplot(clean_data1,aes(Product,Count))+
  geom_bar(aes(fill=Suspicious.Activity),stat = "identity",width=0.8)+
  facet_grid(clean_data1$Year.Month~.)+
  labs(title="Yifeng Luo")+
  theme(axis.text.x=(element_text(angle = 25)))
```



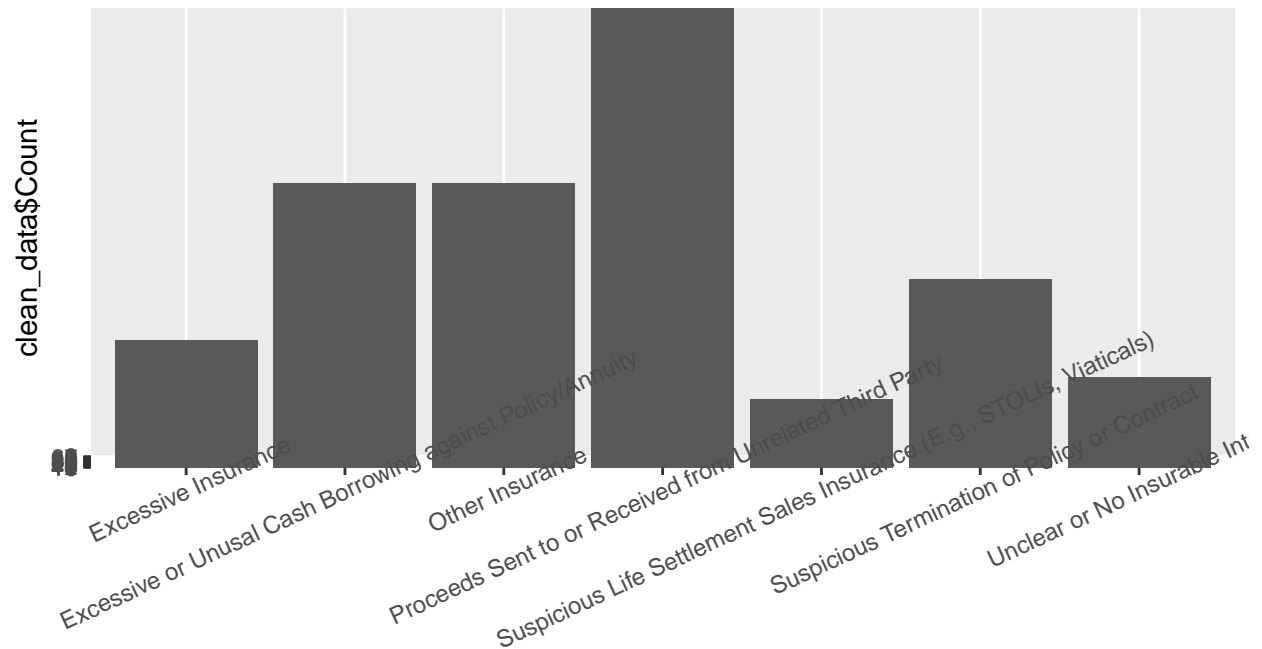
the graph above shows that overall the number of suspicious activity decreased from 2017 to 2018. But the suspicious activity is more likely to occur in insurance/annuity products both of 2017 and 2018 and 'proceeds sent or received from unrelated third party' shared a big proportion of the way of suspicious activity in these two years.

Part 3 (Yaotang Luo)

We are going to find some specific feature of the relationship between suspicious activity and relationship.

We first obtain the bar charts

```
ggplot(data = clean_data) + geom_histogram( mapping = aes(x = Suspicious.Activity, y = clean_data$Count.
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

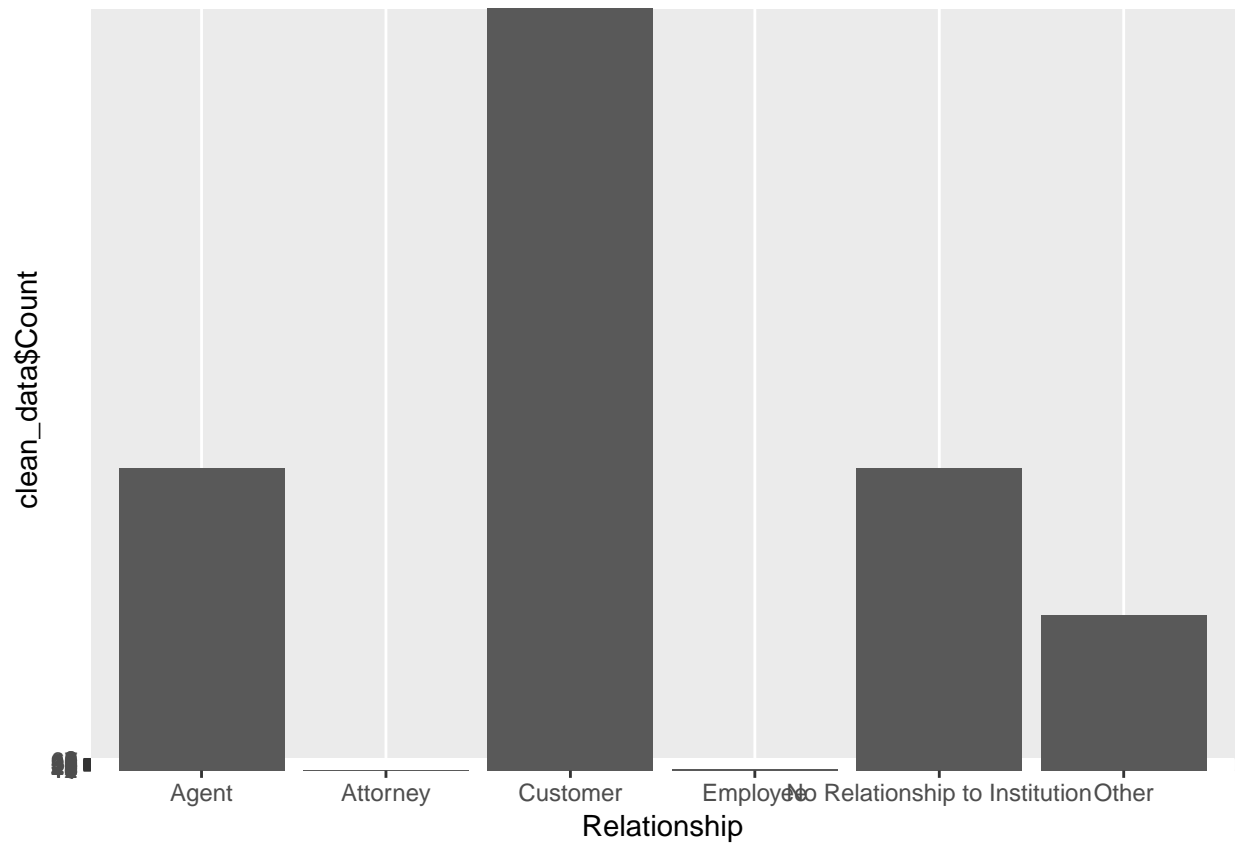


Suspicious.Activity

We can find that the most frequent suspicious activity is that proceeds sent to or received from unrelated third party. The second frequent suspicious activity are from excessive or unusual cash borrowing against policy and from other insurance. The number of suspicious activity comes from life settlement sales insurance is really low.

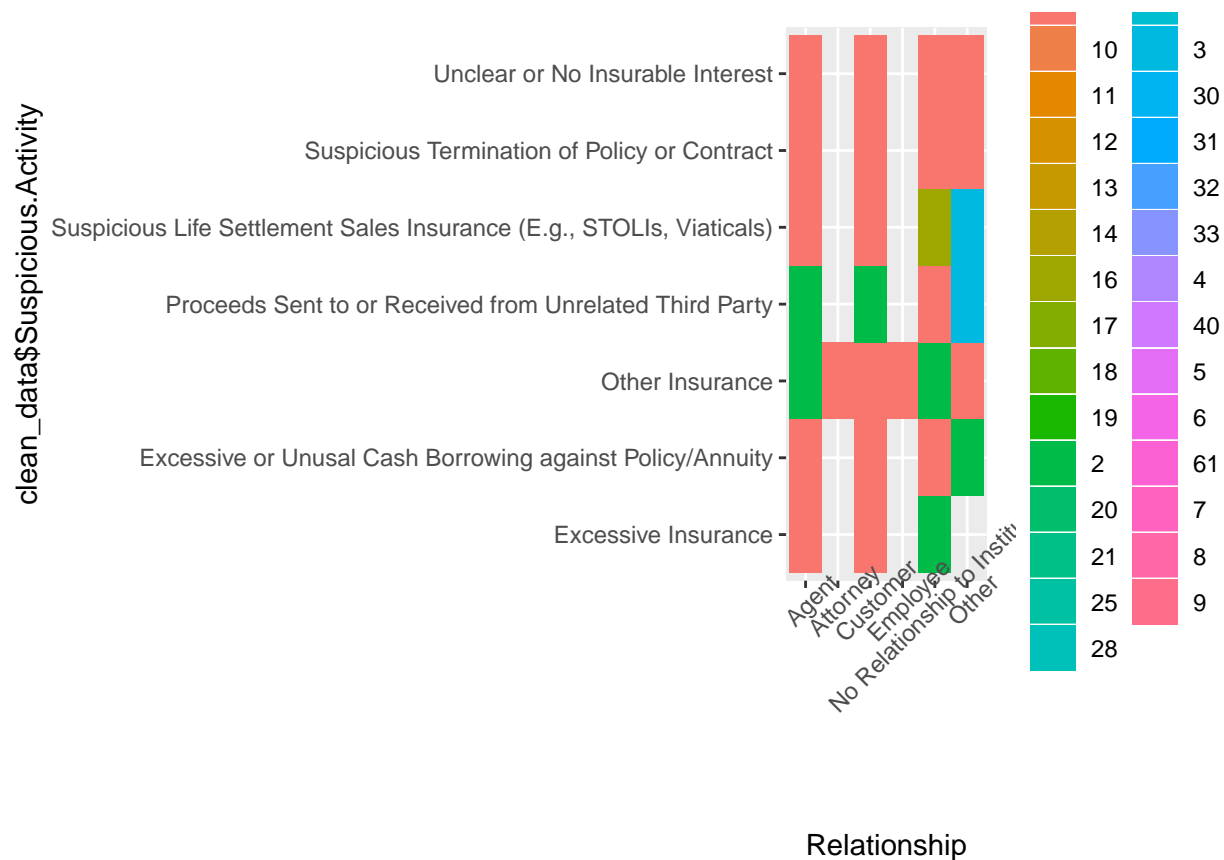
```
ggplot(data = clean_data) + geom_histogram( mapping = aes(x = Relationship, y = clean_data$Count), stat = "count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



We can obviously find that the suspicious activity is rarely happen among attorney and employee, but most of the suspicious activity comes from costumers.

```
ggplot(data = clean_data, mapping = aes(x = Relationship, y = clean_data$Suspicious.Activity, stat= "id
```



When we compare these two variable together, we can find more information about the connection between suspicious activity and relationship. In this figure, we first find that, for the type of relationships that has least suspicious activity such as attorney and employee, their suspicious activity would just come from 'other insurance'. Second, we find that one of the type of suspicious activity, 'other insurance', contains all of the types of relationships including agent, attorney, customer, employee, no relationship and other. Third, although most of suspicious activity comes from customers, both of the type 'no relationship to institution' and 'agent' has more suspicious activity in some areas compare to customers. And the type of relationship, which is called 'other', is responsible for most of 'suspicious life settlement sales insurance' and 'proceeds sent to or received from unrelated third party', instead of customers.

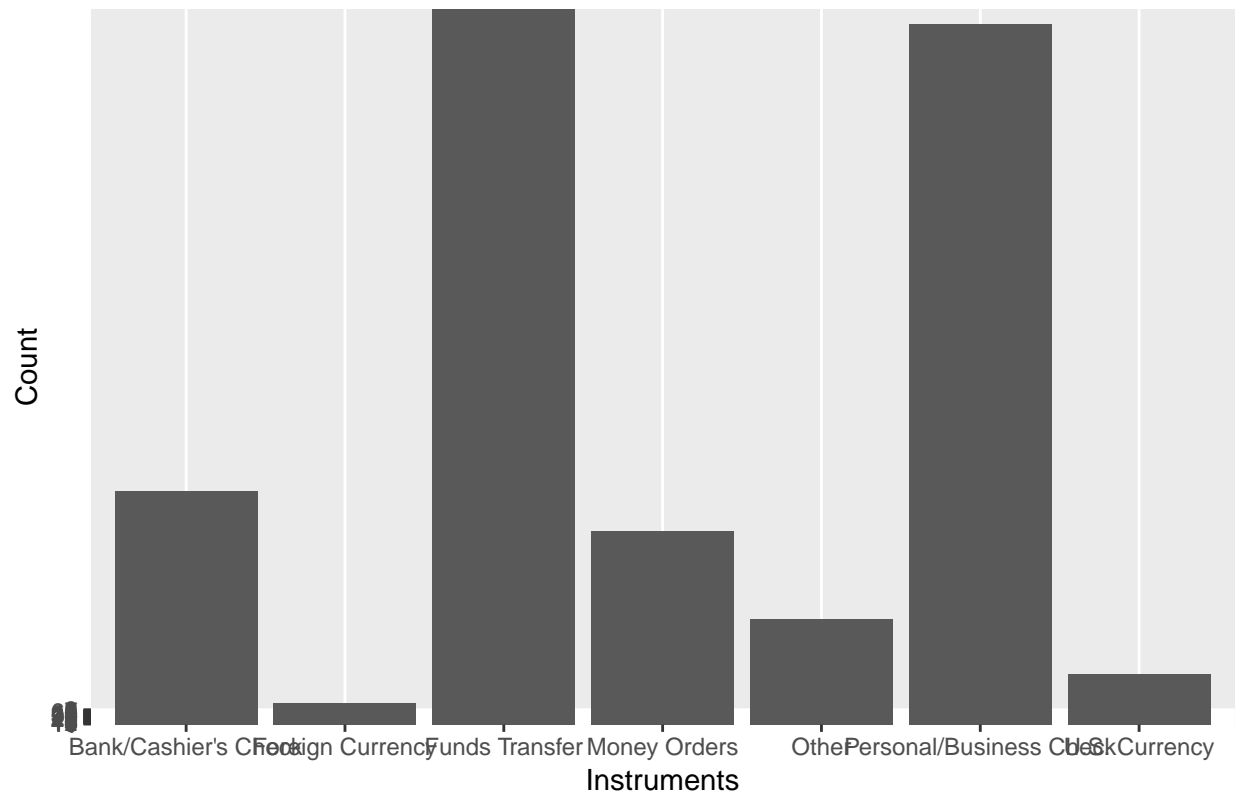
Part 4 (Qianhui Rong)

Objectives

In order to see which instruments are mostly used in financial crimes, which relationships occur the most and which combinations of instruments and relationships are most frequent crime carriers, some plots need to be drawn to make deductions. Because the two concerning variables are categorical, some kinds of plots will not be useful and direct. I'll first plot bar charts to check each variable's separate relation to occurrences, and then use point plot and tile plot to see the combination of my two variables.

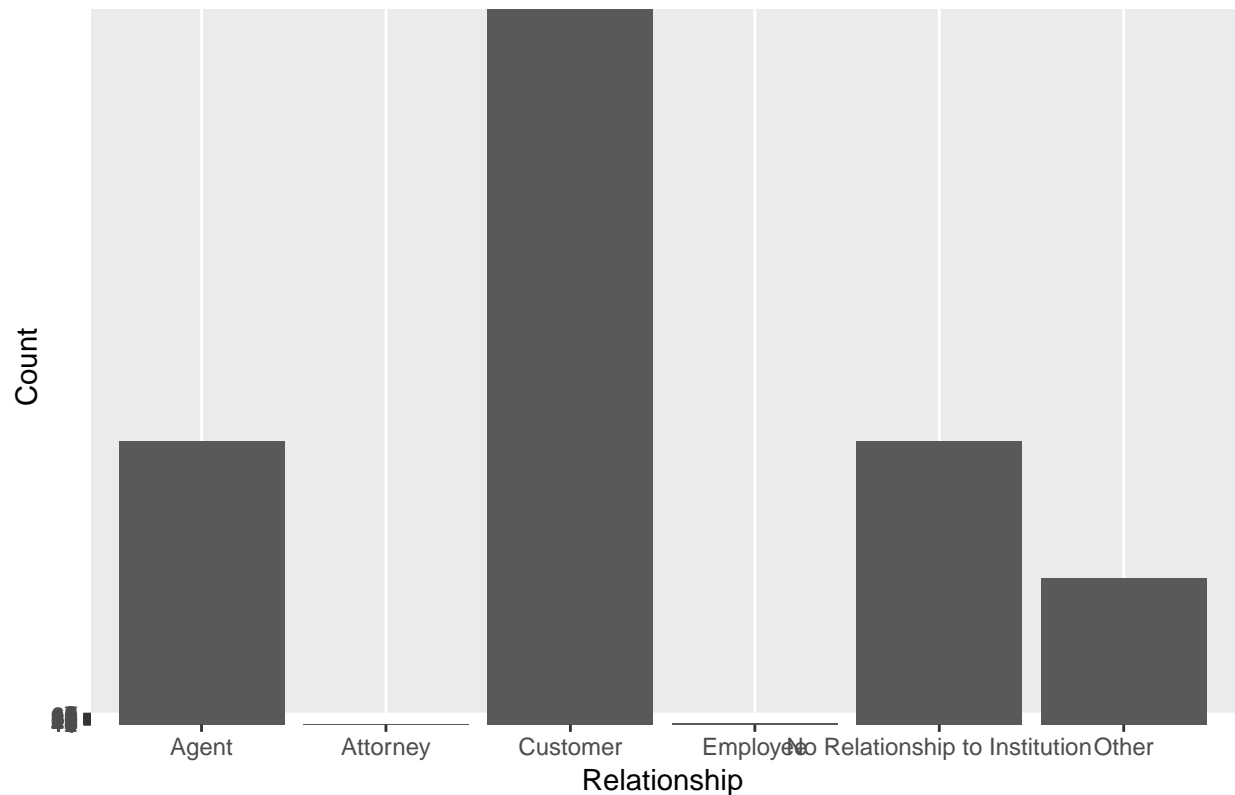
```
ggplot(data = clean_data)+
  geom_bar(mapping = aes(y=Count,x=Instrument),stat = "identity")+
  ggtitle("Crime Occurencies of Different Instruments by Qianhui Rong") +
  labs(x="Instruments",y="Count")
```

Crime Occurencies of Different Instruments by Qianhui Rong



```
ggplot(data = clean_data)+
  geom_bar(mapping = aes(y=Count,x=Relationship),stat = "identity")+
  ggtitle("Crime Occurencies of Different Relationships by Qianhui Rong") +
  labs(x="Relationship",y="Count")
```

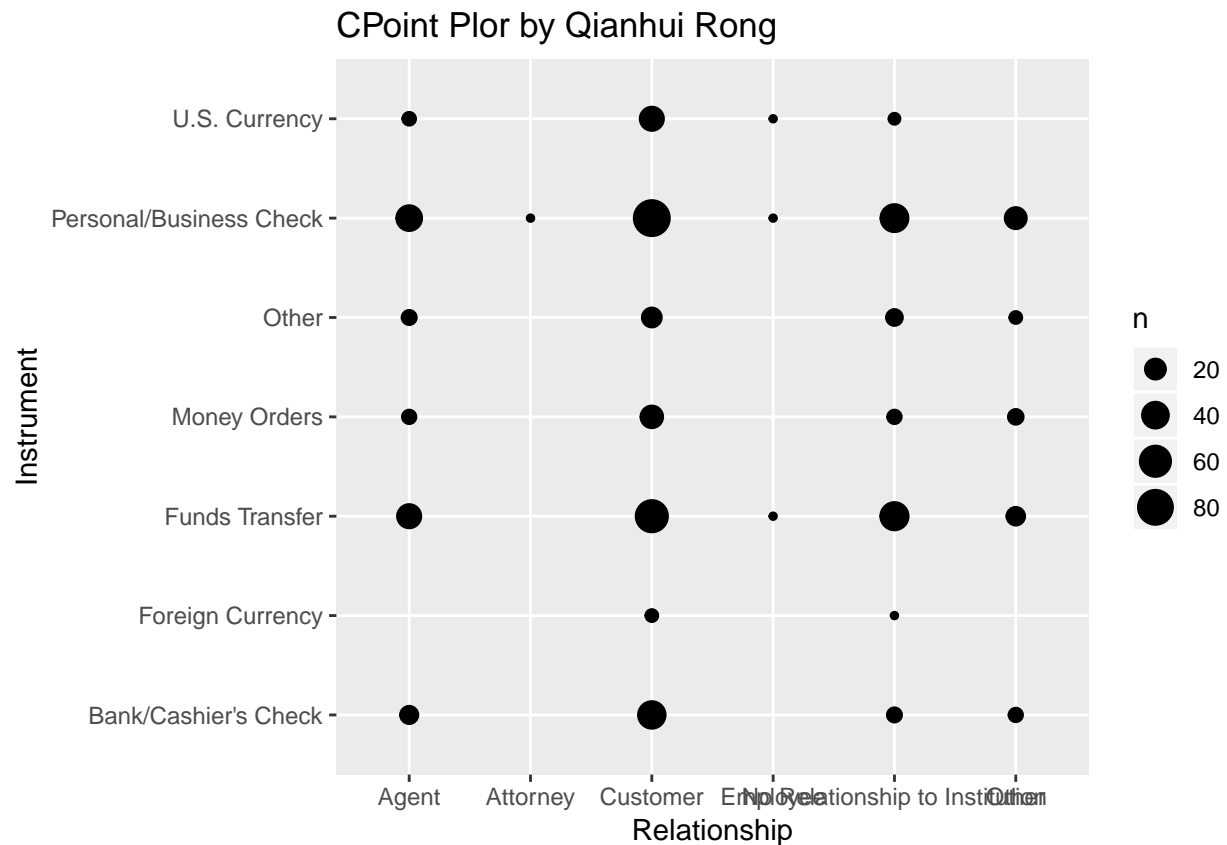
Crime Occurencies of Different Relationships by Qianhui Rong



#Bar Charts From the bar chart of “Crime Occurencies of Different Instruments”, We can see that Funds Transfer and Personal/Business Check are the most frequent instruments used in crimes, and the gaps between these two instruments and the others are large. Foreign Currency, as an instrument, can be neglected.

From the bar chart of “Crime Occurencies of Different Relationships”, We can see that Customer as a relationship type is dominant in financial crimes. Agent and Relationship to Institutioner are the second ones, which are about a half of Customer. Attornet and Employer are negligible.

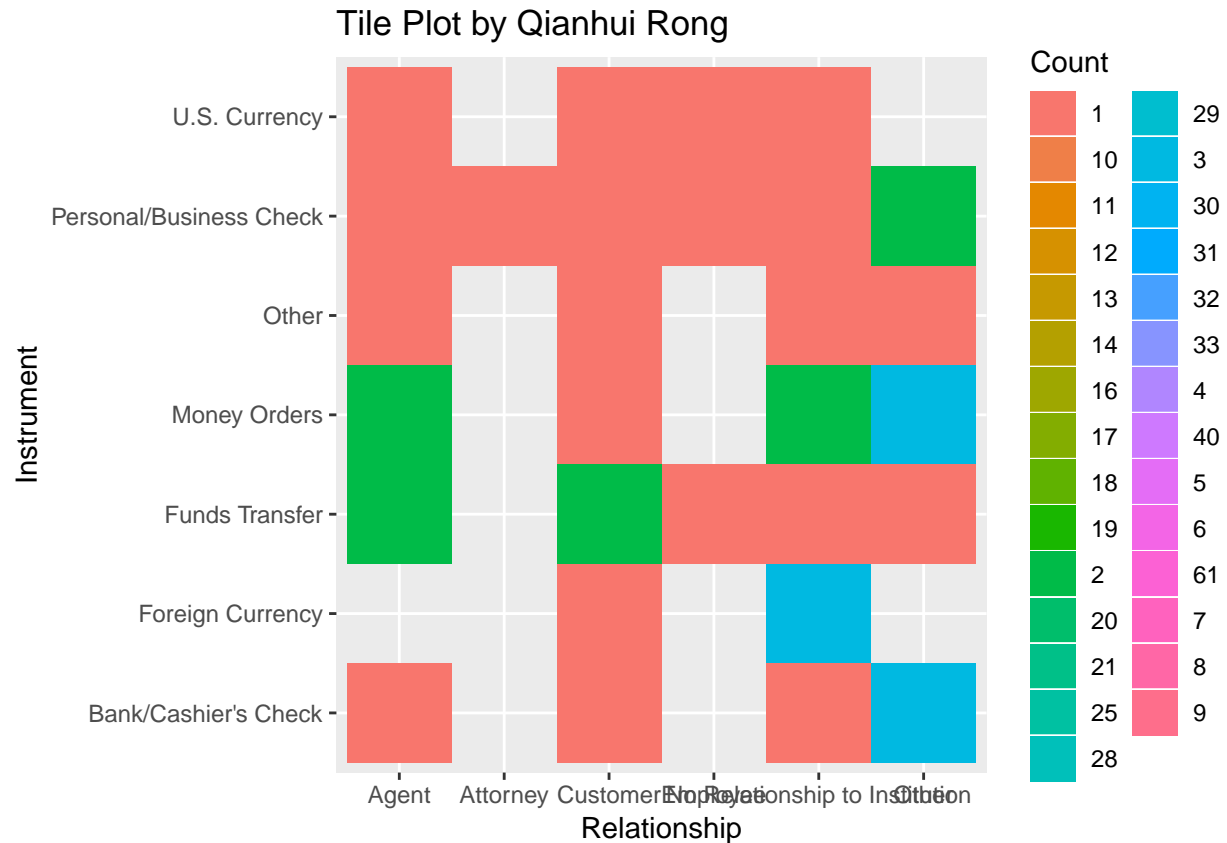
```
ggplot(data = clean_data) +
  geom_count(mapping = aes(x = Relationship, y = Instrument)) +
  ggtitle("CPoint Plor by Qianhui Rong")
```

#Point Plot From this point plot, we can certainly see some combinations have more occurrences than others. The most frequent combination is Customer v.s P/B Check, which is about 80 times. Customer v.s Funds Transfer and Bank/Cashier Check are secondary, which are about 60 times. There are certain cross without circles, that is to say some Instruments and Relationships are not possible.

Tile Plot

```
ggplot(data = clean_data) +
  geom_tile(mapping = aes(x=Relationship,y=Instrument,fill=Count))+
  ggtitle("Tile Plot by Qianhui Rong")
```



From the tile plot, besides the combinations that don't exist, most of the combinations occur one time, and there five combinations occurred 2 times, 3 combinations occurred 3 times(including Foreign Currency v.s. No Relationship, B/C Check v.s Others, and Money Orders v.s Others).

Conclusion

This part of work indicates which kind of relationship is most common and which instrument is most used in crimes. If we want to draw some conclusions about how to reduce financial crime in some certain aspect or by controlling certain instrument and relationship, Foreign Currency, Bank/Cashier's Check and Money Orders should be paid attention to; No relationship to institution and Other relationship should be checked for caution. Additional information will be needed to determine some other cross combinations' influence on financial crime occurrences. We can combine our four parts of work to explore more.