MA 615 assignment 3

Xiangliang Liu(Ian), Zhaobin Liu, Yifudong, Jinfei Xue

1.Packages we need.

```r
library(readr)
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
```

```r
library(knitr)
library(readxl)
library(knitr)
library(ggplot2)
```

2.Data

Description: We collected the data from "https://www.fincen.gov/reports/sar-stats" with Year, State, Industry,Suspicious_Activity, Instrument, Count of financial crimes as predictors. We specifically collected insurance industry data by selecting all the catagories with those variables.

3.Import data and data cleaning.

```r
#There is a subtotal term in the table. We used filter() to eliminate all the rows with "[Total]" term.
SARStats <- read_csv("~/Downloads/SARStats.csv")
```

```
## Parsed with column specification:
## cols(
##   `Year Month` = col_integer(),
##   State = col_character(),
##   Industry = col_character(),
##   `Suspicious Activity` = col_character(),
##   Instrument = col_character(),
##   Count = col_number()
## )
```

```r
df = filter(SARStats, Instrument != "[Total]")
names(df) = c("Year", "State", "Industry","Suspicious_Activity", "Instrument","Count")
```

4.Discussion and visualization.

The count of financial crime in different States in the US. (Xiangliang)

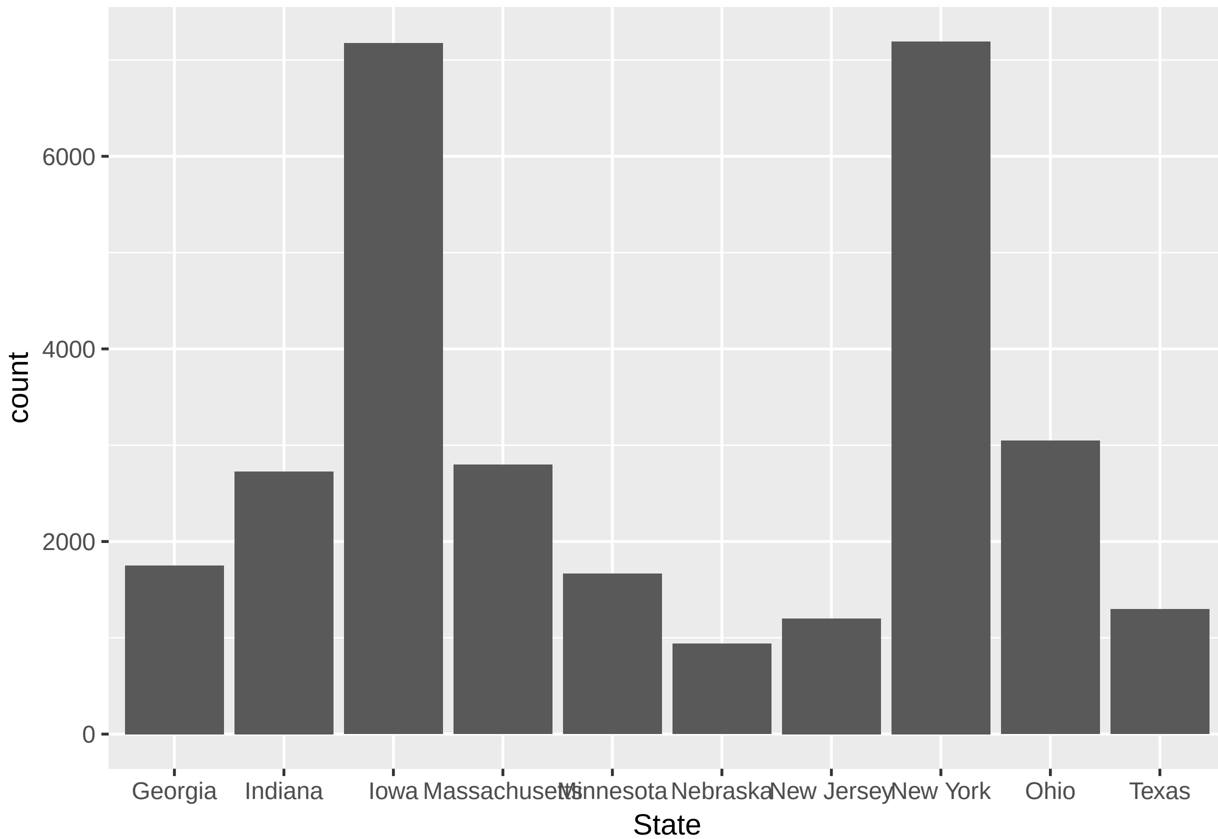
```
df %>% group_by(State)  %>%
  summarize(Count = sum(as.integer(Count))) %>%
  arrange(desc(Count)) %>%
  mutate(Percent = Count/sum(Count))%>%
  slice(1:10) -> x

kable(x, caption = "The number of suspicious activities in insurance company industry Reports by State")
```

Table 1: The number of suspicious activities in insurance company industry Reports by State

State	Count	Percent
New York	7193	0.1975122
Iowa	7174	0.1969905
Ohio	3046	0.0836400
Massachusetts	2797	0.0768027
Indiana	2728	0.0749080
Georgia	1751	0.0480806
Minnesota	1665	0.0457191
Texas	1297	0.0356143
New Jersey	1200	0.0329507
Nebraska	937	0.0257290

```
ggplot(x, aes(x = State, y = Count)) + ylab("count") +
     geom_col(aes(x = State))
```



The count of financial crime in the US by Instrument (Xiangliang)

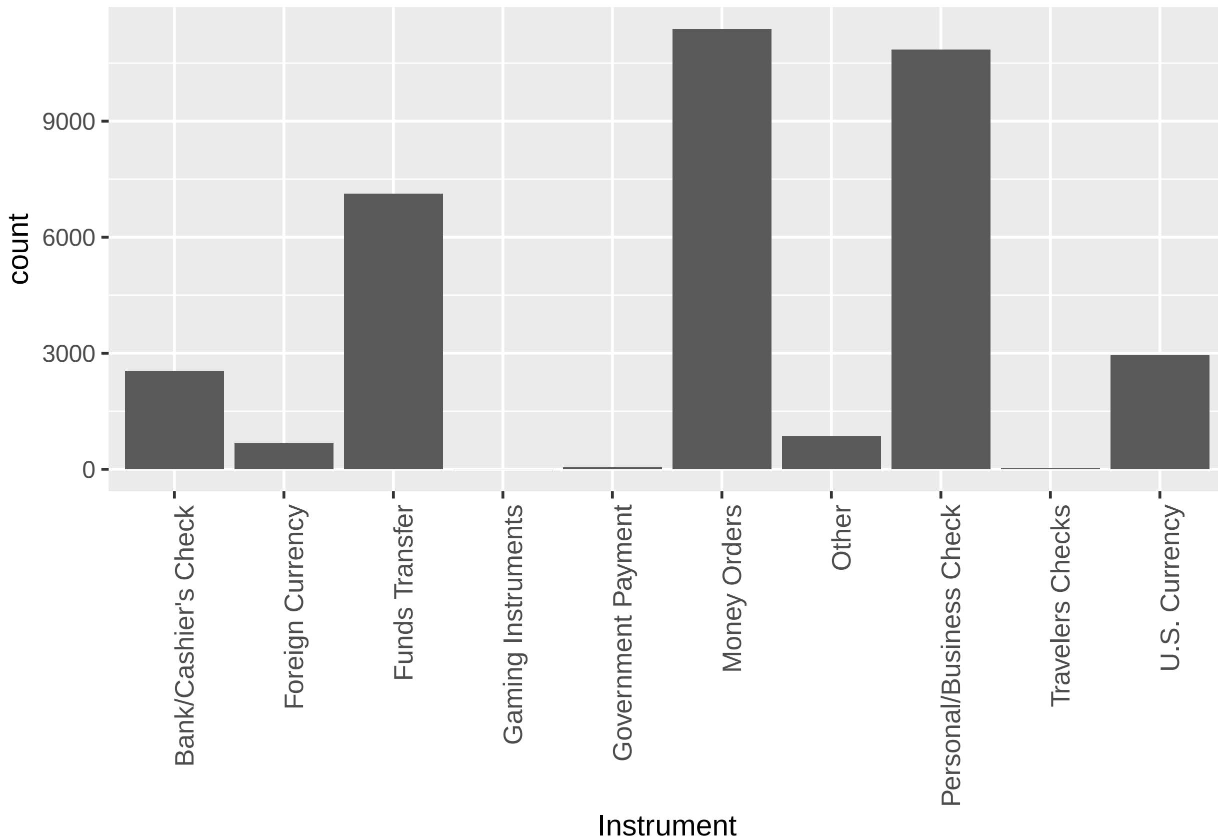
```
(table3 = df %>% group_by(Instrument)  %>%
  summarize(Count = sum(as.integer(Count))) %>%
  arrange(desc(Count))) %>%
mutate(Percent = Count/sum(Count))->z

kable(z, caption = "The number of suspicious activities in insurance company industry Reports by Instru
```

Table 2: The number of suspicious activities in insurance company industry Reports by Instrument

Instrument	Count	Percent
Money Orders	11378	0.3124279
Personal/Business Check	10852	0.2979845
Funds Transfer	7118	0.1954528
U.S. Currency	2954	0.0811137
Bank/Cashier's Check	2534	0.0695810
Other	853	0.0234225
Foreign Currency	663	0.0182053
Government Payment	42	0.0011533
Travelers Checks	17	0.0004668
Gaming Instruments	7	0.0001922

```
ggplot(z, aes(x = Instrument, y = Count)) + ylab("count") +
     geom_col(aes(x = Instrument))+theme( axis.text.x = element_text(size = 10,angle = 90,hjust = 1))
```



The count of financial crime in the US by year(Jinfei)

In the Suspicious Activities Report database, making a table of total number of reports by year to start.*

```
df1<-select(df,Year, Count)
df1 %>% group_by(Year)  %>%
  summarize(Count=sum(Count)) %>%
  arrange(desc(Count)) -> df1.s
kable(df1.s, caption = " Table 1: Suspicious Activity Report by Year")
```

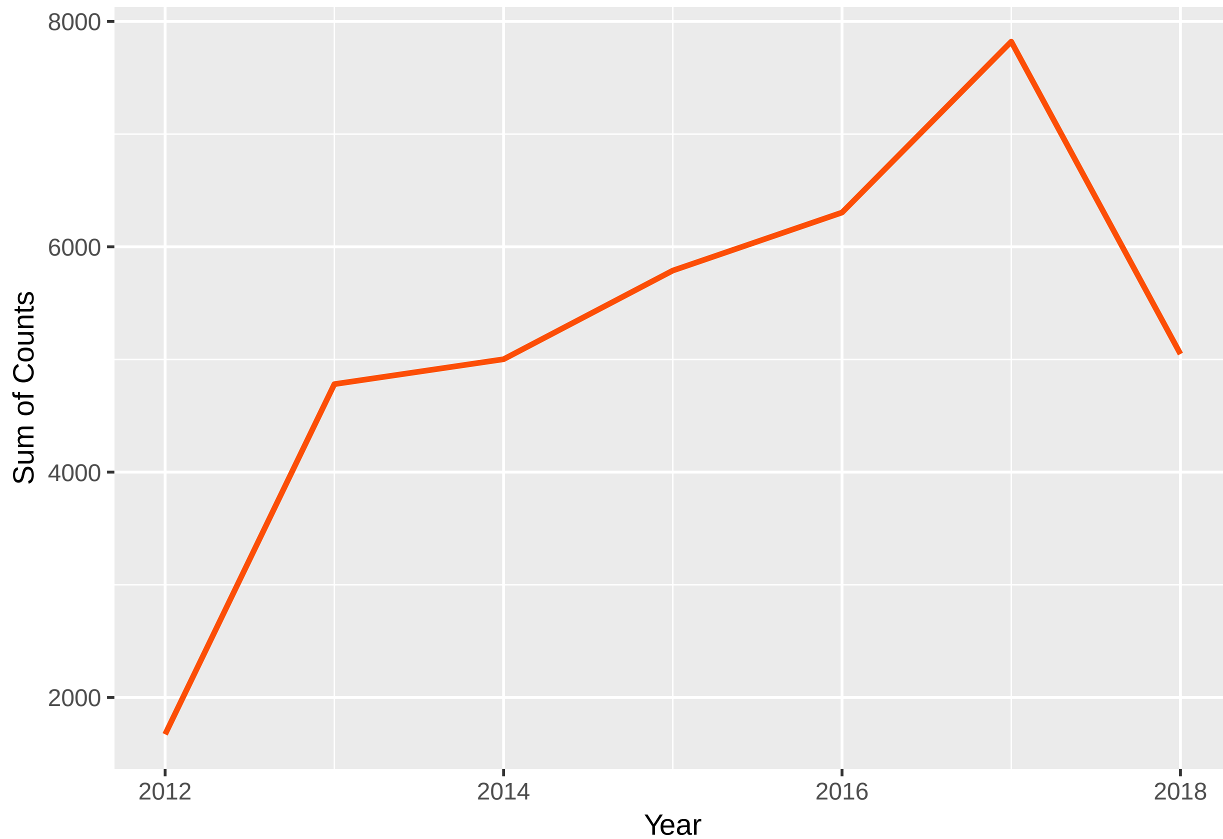
Table 3: Table 1: Suspicious Activity Report by Year

Year	Count
2017	7821
2016	6304
2015	5789
2018	5048
2014	5002
2013	4781
2012	1673

*From the table above, we can see it's in 2017 that the total number of financial crime reports by year is largest from 2012 to 2018.

In order to track the tendency further, we make the following time series plot using ggplot.

```
df1.ss<-arrange(df1.s,Year)
ggplot(df1.ss,aes(x = Year, y = Count,group = 1)) +
  geom_line(color = "#FC4E07",size=1) +
  xlab("Year") + ylab("Sum of Counts")
```



Analyze the financial crisis in Massachusetts by Suspicious Acticity (yifu)

```
#Use filter() and select() function to set up new dataset suspicious with predictors Year, Suspicious,
suspicious <- filter(df,State=="Massachusetts")

suspicious <- select(suspicious,"Year","Suspicious_Activity", "Count")
```

*Now I get the ideal data for analysis. The first thing I want to know is which kind of insurance suspicious activities are the most common in Massachusetts. So I made a table to display the result:

```
#group by suspicious activity
suspicious %>%
  group_by(Suspicious_Activity,Year)%>%
  summarise(Count=sum(Count))%>%
  arrange(desc(Year))%>%
  mutate(percentage=Count/sum(Count))->suspicious1

suspicious1%>%
  group_by(Suspicious_Activity)%>%
```

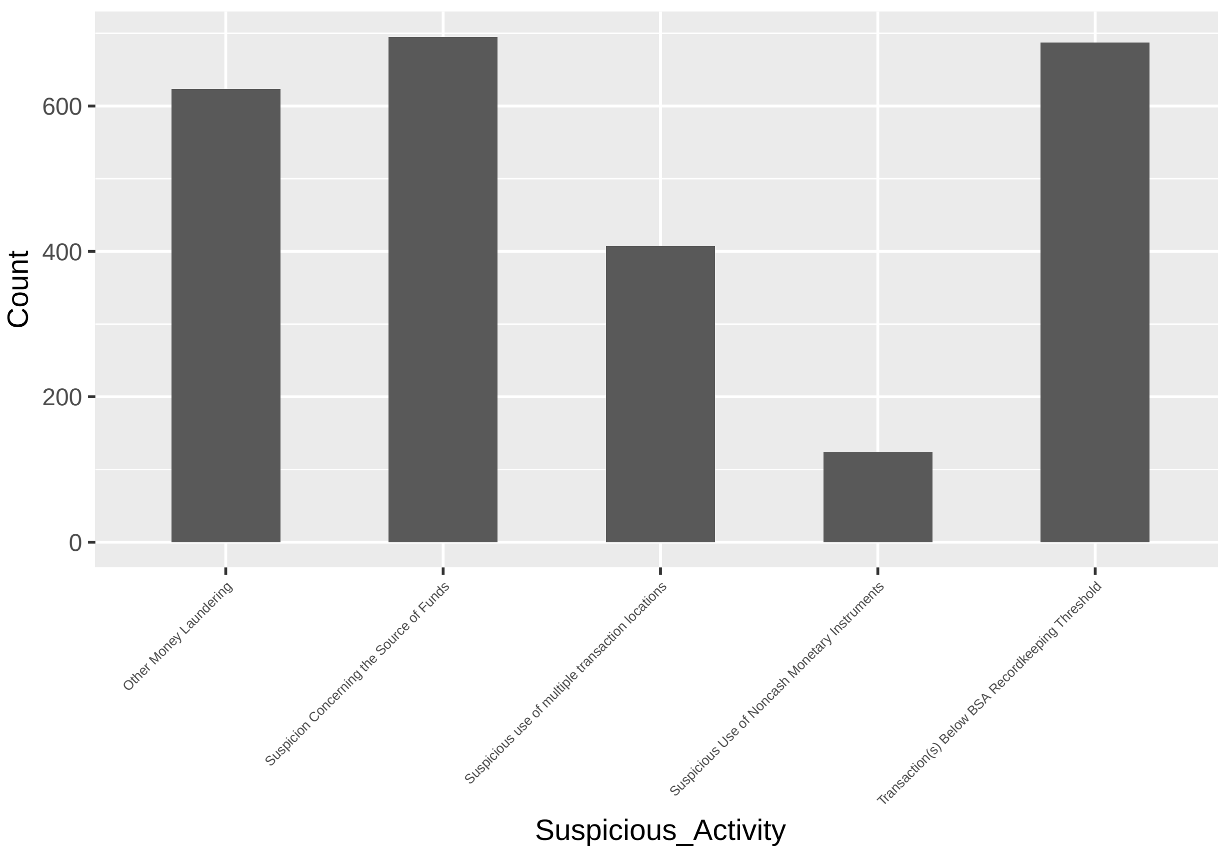
```
  summarise(Count=sum(Count))%>%
  arrange(desc(Count))%>%
  mutate(percentage=Count/sum(Count))%>%
  slice(1:5)->suspicious2

kable(suspicious2, caption = "Insurance Company Suspicious Activity")
```

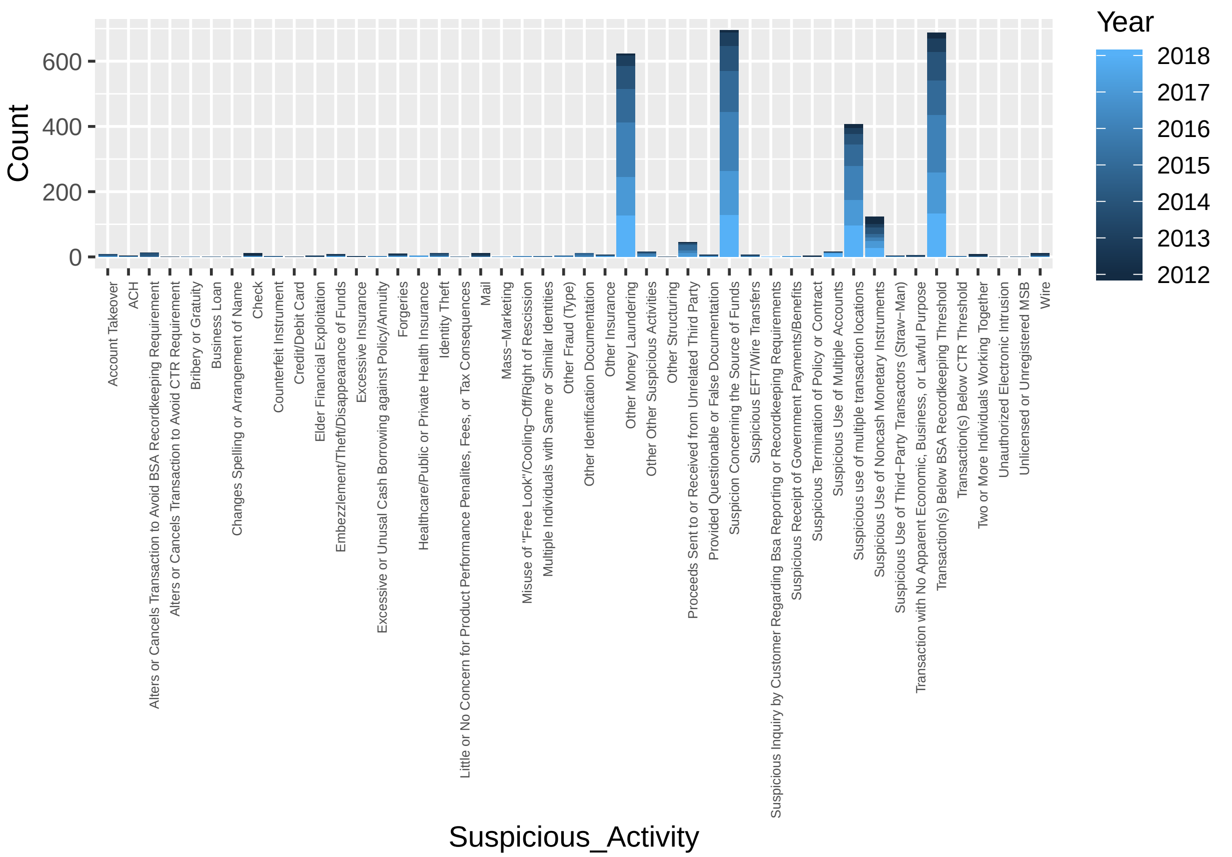
Suspicious_Activity	Count	
Suspicion Concerning the Source of Funds	695	
Transaction(s) Below BSA Recordkeeping Threshold	687	
Other Money Laundering	623	
Suspicious use of multiple transaction locations	407	
Suspicious Use of Noncash Monetary Instruments	124	
discussion		
The result shows that the most common 4 kinds of in	surance	suspicious activities are Suspicion Concerning the Source

Then I want to know the trend of suspicious activities in MA, and whether those criminal activities are reduced effectively. So I draw a plot which contains both the data of suspicious activities and years.

```
ggplot(data=suspicious2)+
  geom_bar(width=0.5,stat="identity",mapping = aes(x=Suspicious_Activity, y=Count))+
  theme( axis.text.x = element_text(size = 5,angle = 45,hjust = 1))
```



```r
#bar plot
ggplot(suspicious1)+
  geom_bar(stat="identity",mapping = aes(x=Suspicious_Activity, y=Count,fill=Year))+
  theme( axis.text.x = element_text(size = 5,angle = 90,hjust = 1))
```



For this plot, the deeper the color is, the more remote the year is. So the top of the bar represents the suspicious activity data of 2012. From the plot above, it's obvious that from 2012 to 2018, for the most 4 common activities shown above, the number of insurance crime in Massachusetts is generally increasing. It may be due to the increasing scale of insurance industry, as well as money inflation. So here we need more research on what factor mainly influence the increasing of insurance criminal activities. Maybe we should pay attention to this 4 kind of criminal activities in the future.

The count of financial crime in the Massachusetts by Instrument (Zhaobin)

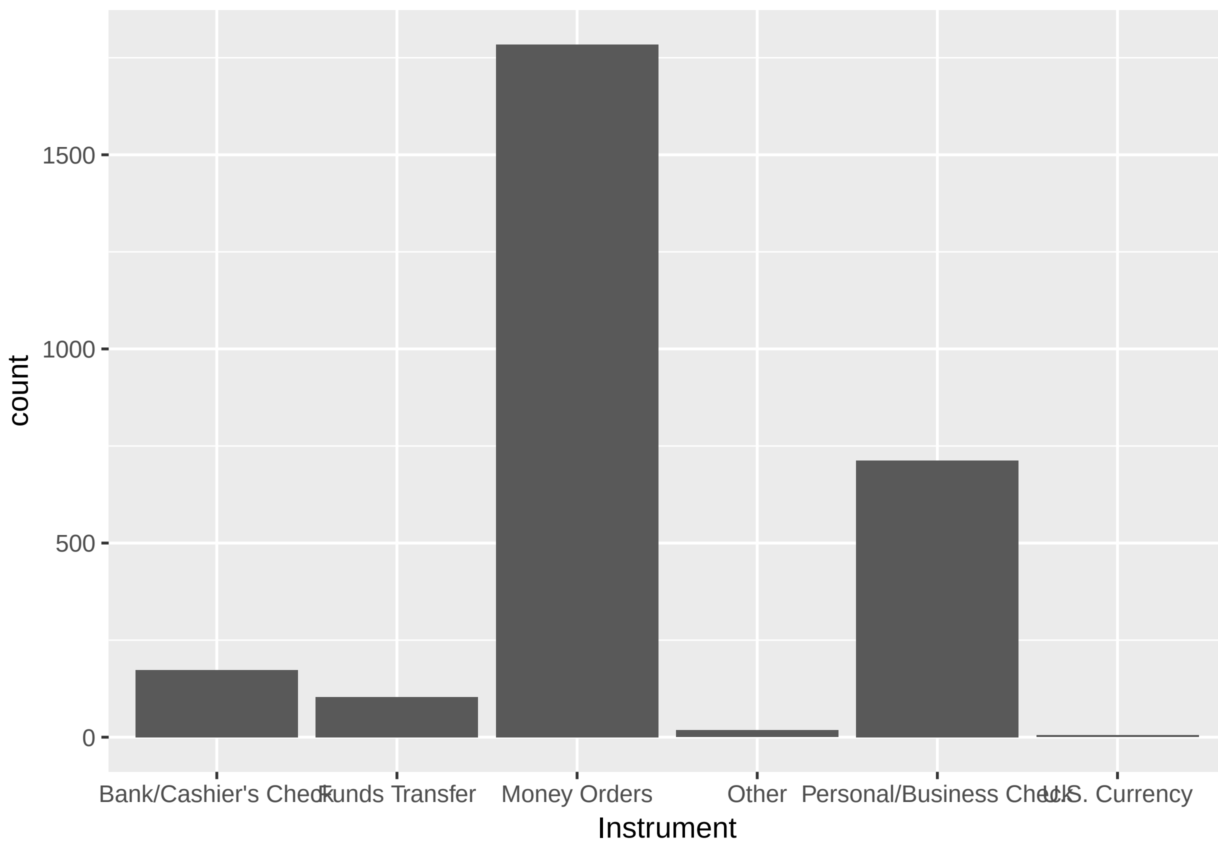
```r
Instrument_of_MA = filter(df,State == "Massachusetts")
Instrument_of_MA %>% group_by(Instrument) %>%
  summarize(Count = sum(Count)) %>%
  arrange(desc(Count)) -> c
kable(c,caption = "Massachusetts instrument types and/or payment mechanisms used
      in the suspicious activity")
```

Instrument	Count

Table 5: Massachusetts instrument types and/or payment mechanisms used in the suspicious activity

Instrument	Count
Money Orders	1784
Personal/Business Check	713
Bank/Cashier's Check	173
Funds Transfer	104
Other	18
U.S. Currency	5

```
ggplot(c, aes(x = Instrument, y = Count)) + ylab("count") +
    geom_col(aes(x = Instrument))
```



By the table and histogram, we can see the order and comparision of the Massachusetts instruments types or payment mechanisms used in the suspicious activities as shown above.

Conclusion

In the US: By analyzing the praphes and charts in "Discussion", we found New York and Iowa have more financial crime counts. The most common financial crime is conducted by money order. From plot shows from 2012 to 2015, the total number of reports increases, and then decreases in 2016, but increases in 2017 followed by decreasing in 2018. It can be concluded that the total number of financial crimes reports fluctuates from 2012 to 2018.

In Massachusetts: In the graph of 2012 to 2018, for the most 4 common activities shown above, the number of insurance crime in Massachusetts is generally increasing. Also, money order is the instrument that leads to financial crime.