

Group#9

Xuan, Megha, Jianhao, Guangyan

September 29, 2018

Readme

We are using the data from:

```
head(read.csv("SARStats.csv"))
```

```
##   Year.Month      State      Countym      Industry
## 1      2018 California Alameda County, CA Depository Institution
## 2      2018 California Alameda County, CA Depository Institution
## 3      2018 California Alameda County, CA Depository Institution
## 4      2018 California Alameda County, CA Depository Institution
## 5      2018 California Alameda County, CA Depository Institution
## 6      2018 California Alameda County, CA Depository Institution
##              Suspicious.Activity      Product      Instrument Count
## 1              ACH Credit Card U.S. Currency      6
## 2              Check Credit Card U.S. Currency     10
## 3 Consumer Loan (see instructions) Credit Card U.S. Currency      2
## 4              Credit/Debit Card Credit Card U.S. Currency     27
## 5              Mail Credit Card U.S. Currency      2
## 6              Mass-Marketing Credit Card U.S. Currency      6
```

This is how we generated the dataset:

Filter

Jianhao's Plot

Xuan's plot

As we get a lot of regions here, it is hard to put all the information on the same plot. So I randomly picked 4 counties: ("Los Angeles County, CA", "Santa Clara County, CA", "Orange County, CA", "Santa Barbara County, CA").

I don't want the text to be overlapped with each other, so I switch the direction of x-axis and y-axis. Now the x-axis becomes vertical and the y-axis becomes horizontal.

Position = "dodge" places overlapping objects directly beside one another. This makes it easier to compare individual values.

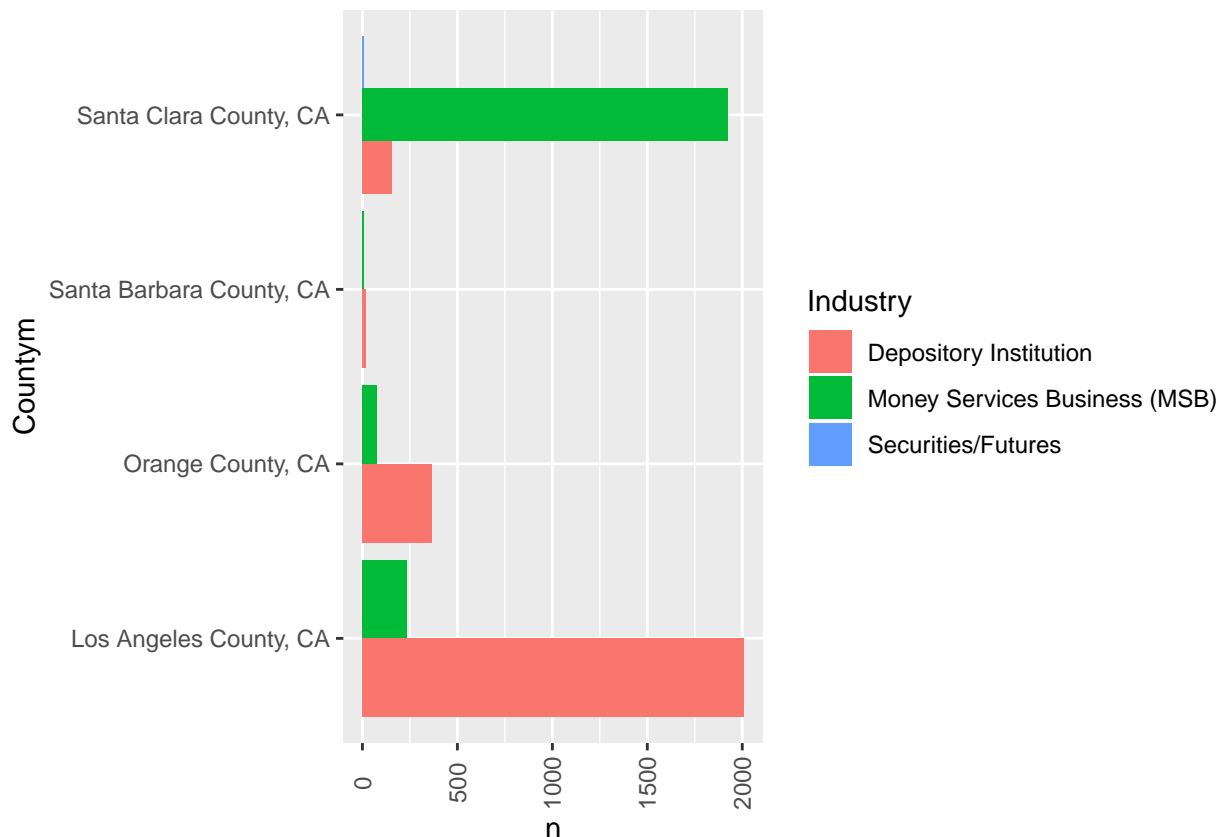
```
SARStats <- read_csv("~/Desktop/SARStats.csv")
```

```
## Parsed with column specification:
## cols(
##   `Year Month` = col_character(),
##   State = col_character(),
##   Countym = col_character(),
##   Industry = col_character(),
##   `Suspicious Activity` = col_character(),
##   Product = col_character(),

```

```
## Instrument = col_character(),
## Count = col_number()
## )

SARStats$Count <- as.numeric(SARStats$Count)
xuan <-SARStats %>%
  group_by(Industry,Countym)%>%
  summarise(n=sum(Count))%>%
  filter(Industry != '[Total]' & Countym %in% c("Los Angeles County, CA","Santa Clara County, CA","Orange County, CA"))
  arrange(desc(n))
ggplot(data=xuan,mapping=aes(x=Countym,y=n),group=factor(1),xlab(Countym))+
  geom_bar(position = "dodge",aes(fill=Industry),stat = "Identity")+
  theme(axis.text.x = element_text(angle = 90, hjust = 0.5, vjust = 0.5))+
  coord_flip()
```

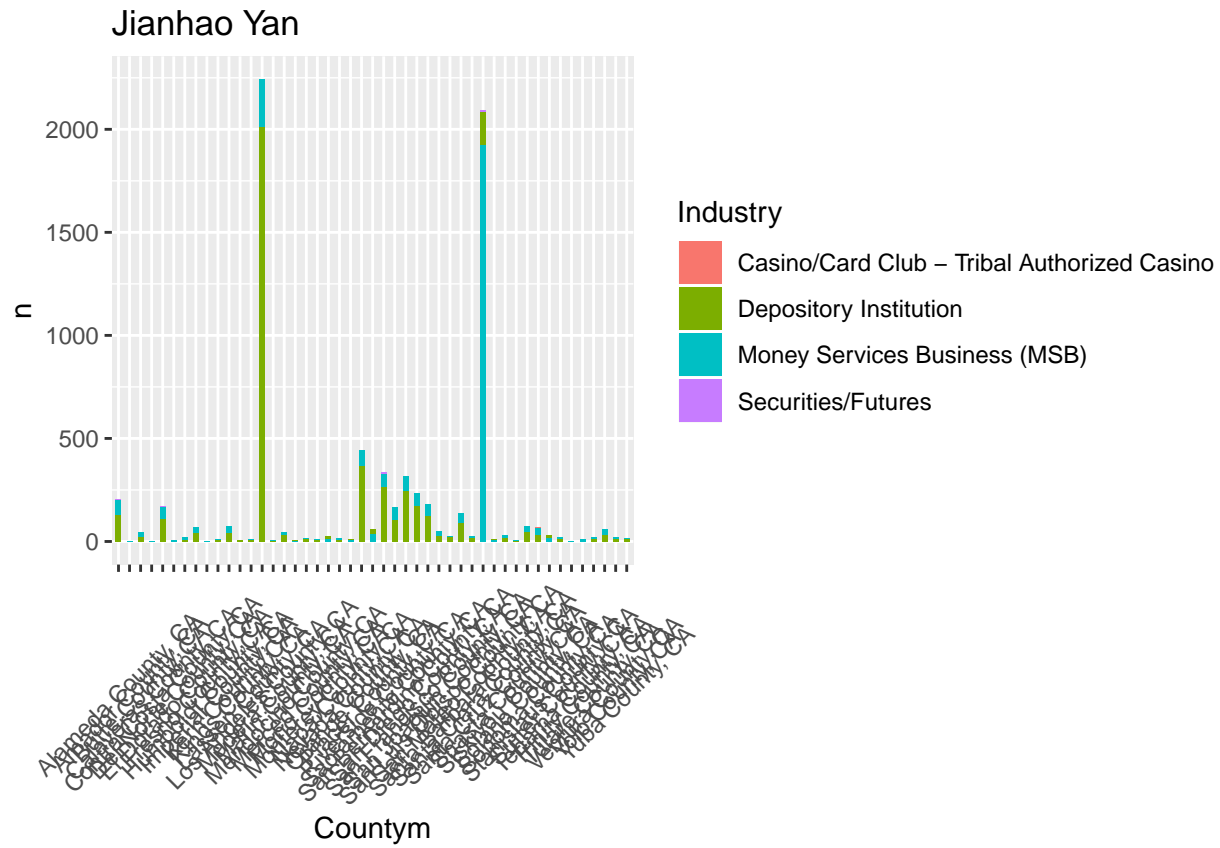


Discussion & Conclusion

Jianhao's plot

```
data_new<-SARStats %>%
  group_by(Countym,Industry) %>%
  summarise(n=sum(Count))%>%
  filter(Countym!='[Total]')%>%
  filter(Industry!='[Total]')%>%
  arrange(desc(n))
ggplot(data_new, aes(x = Countym,y =n, group = factor(1))) +
```

```
geom_bar(stat = "identity", width = 0.5, aes(fill=Industry)) + theme(axis.text.x = element_text(angle = 45))
```



Discussion & Conclusion

##From this graph, we can find that the Los Angeles County, CA has the most financial frauds, and these