

DAR F21 Project Status Notebook:

DeFi

Roman Vakhrushev (vakhrr)

10/31/2021

Contents

Biweekly Work Summary	1
Personal Contribution	1
References	1
GitHub Issues	1
GitHub Commits	1
Discussion of Primary Findings	2
What did you want to know?	2
How did you go about finding it?	2
What did you find?	2

Biweekly Work Summary

Personal Contribution

Some of the graphs were inspired by two papers listed below

Some modification of code by Cole Paquin was used to add second axis on graphs.

All other work was completed by me.

References

Two papers I read and built upon:

<https://arxiv.org/pdf/2106.06389.pdf> <https://arxiv.org/pdf/2009.13235.pdf>

GitHub Issues

Issue #92 self-assigned and completed

GitHub Commits

Branch name: dar-vakhrr

Discussion of Primary Findings

What did you want to know?

I was focusing mostly on two aspects of liquidation: deficient liquidators and good liquidators. I wanted to know what differences can be observed between deficient liquidators, good liquidators, and regular users.

Additionally, I wanted to find good graphs on liquidation to include for the app.

How did you go about finding it?

I was doing deficient and good liquidator sections separately. I analyzed all groups by creating new data frames and data visualizations.

Additionally, I read two papers on liquidations in order to explain some data patterns and find interesting graphs for the app.

What did you find?

```
#data collection as always
df<-read_rds('../Data/transactions2.rds')
# Use deplyr to drop NA reserves, add the counts and then kep only the top 20
reservecoins <- df %>% drop_na(reserve) %>%
count(reserve) %>%
arrange(-n) %>%
head(20)
```

```
#function to mark stable and non-stable coins
coinType <- function(coin) {
  #stable_coins <- list("USDC","USDT","DAI","BUSD","SUSD","GUSD","TUSD")
  if(str_contains(coin,"USD",ignore.case = TRUE))
  {
    result = "stable"
  }
  else if(str_contains(coin,"DAI",ignore.case = TRUE))
  {
    result = "stable"
  }
  else
  {
    result = "non-stable"
  }
  return(result)
}
```

```
defLiquid <- function(principal, collateral) {
  if(collateral < principal)
  {
    result = TRUE
  }
  else
  {
    result = FALSE
  }
}
```

```

}
  return(result)
}

```

Deficient Liquidators

Let's start by building a dataframe for deficient liquidations and computing the percentage of deficient liquidations over all liquidations.

```

#Show transactions, where collateral<principal (exclude WETH and AmmWETH for now).
dfst <- df %>% filter(type == "liquidation") %>% filter(amountUSDCollateral<amountUSDPincipal)

dfst$collateralType <- mapply(coinType, dfst$collateralReserve)
dfst$principalType <- mapply(coinType, dfst$principalReserve)

dfst <- dfst %>% group_by(user) %>% summarize(num_def = n(),total_collateral = sum(amountUSDCollateral))

head(dfst,10)

```

```

## # A tibble: 10 x 5
##       user num_def total_collateral total_principal percent
##   <dbl>   <int>         <dbl>         <dbl>   <dbl>
## 1 9.58e45     1         68163.         93000.    73.3
## 2 1.39e46     1          1333.          1368.    97.5
## 3 1.81e46     1           366.           379.    96.7
## 4 2.56e46     1        12674.         42057.    30.1
## 5 4.28e46     2           599.           625.    95.9
## 6 4.50e46     1        17769.         20465.    86.8
## 7 5.03e46     4          4016.          8520.    47.1
## 8 5.46e46     1        41045.         67071.    61.2
## 9 5.48e46     1          3222.          4111.    78.4
## 10 5.68e46     3          9007.         44602.    20.2

```

```

#plot <- ggplot(dfst,aes(y = total_collateral,x = total_principal,color = as.factor(n))) + geom_point()

plot <- ggplot(dfst,aes(y = total_collateral,x = total_principal,color = as.factor(num_def))) + geom_point()

dfst <- dfst[order(dfst$percent),]

dfst <- dfst %>% rownames_to_column('user_number')

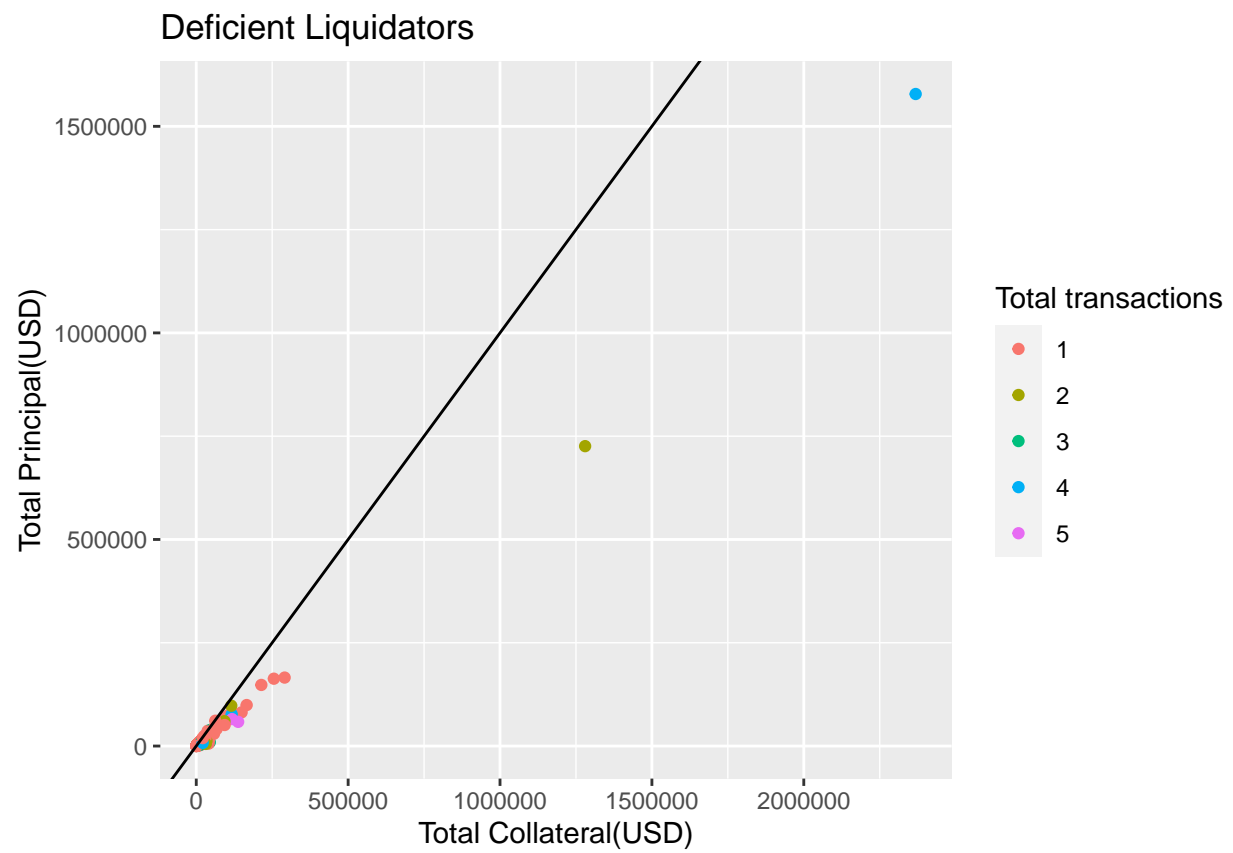
dfst$user_number <- as.integer(dfst$user_number)

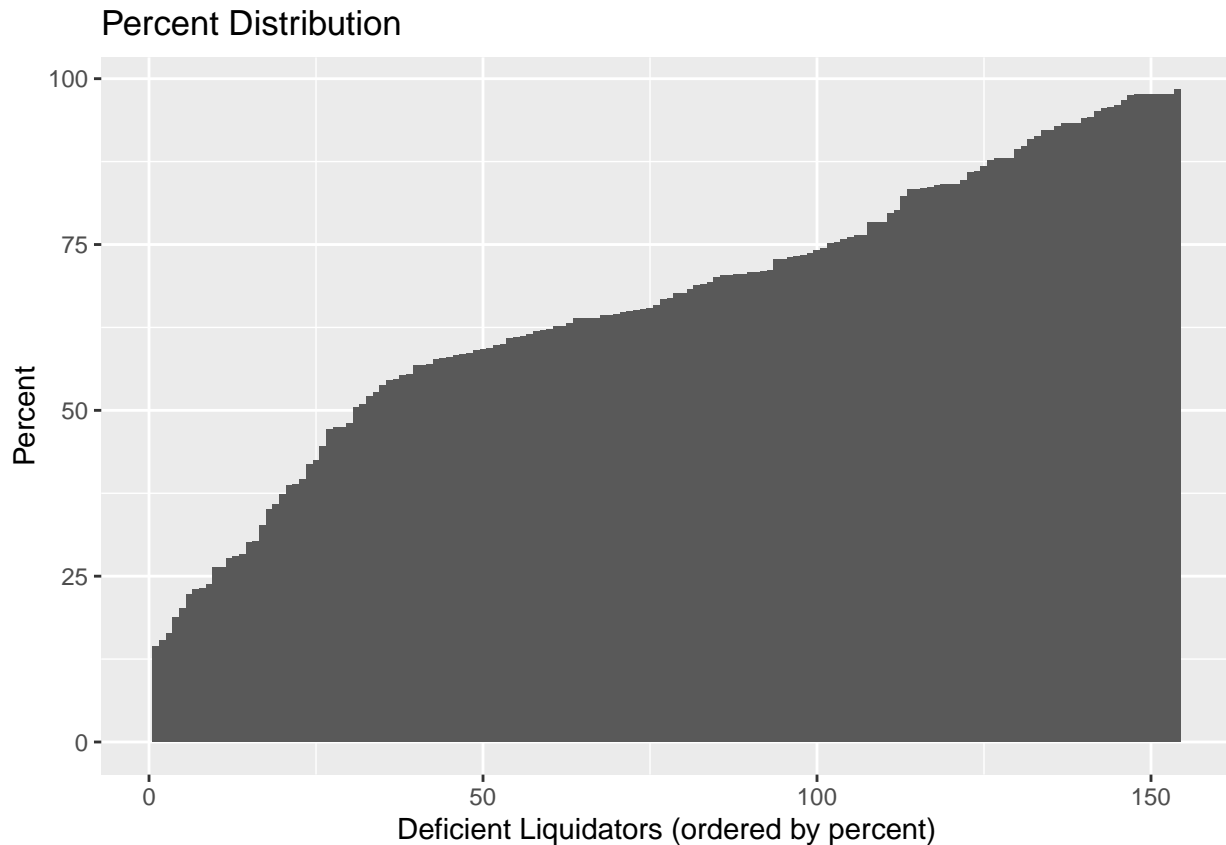
plot2 <- ggplot(dfst,aes(y = percent, x = user_number)) + geom_bar(stat="identity", position=position_dodge())

#ggplotly(plot)

plot

```





The table (part of it) and graph above show the data on deficient liquidators. There are 154 deficient liquidators in our data (people who made at least 1 deficient liquidation). All of the figures above were built around total numbers for collateral and principal used in all deficient liquidations for different users. Percent in the table (and barplot) means $\text{total_collateral} / \text{total_principal} * 100$. As we can see from the table, both the number of deficient liquidations and percent can be different. Number of deficient liquidations is skewed towards low numbers, which makes sense since we do not expect users to make a lot of deficient liquidations. At the same time, percent varies a lot from low numbers to high numbers. Additionally the barplot above was created just to illustrate the distribution of percent variable over deficient liquidators. It looks like some deficient liquidations cannot be explained by just small fluctuations in price: the percent for many users is way below 50, which means the difference between principal and collateral values was very significant. From the graph, we can tell that there are two outliers: users that did several liquidations with very big trading volume, and probably lost a lot of money due to these liquidations. Other than that, the rest of the users are concentrated near (0,0) and never exceed 300000 USD in collateral and 170000 USD in principal.

```
#deficient users data
defusers <- df %>% filter(user %in% dfst$user)

defusers_liq <- defusers %>% filter(type == "liquidation")

defusers_liq$deficiency <- mapply(defLiquid,defusers_liq$amountUSDPrincipal,defusers_liq$amountUSDCollateral)

defusers_liq <- defusers_liq %>% group_by(user) %>% summarize(num = n(), num_def = sum(ifelse(deficiency < 0, 1, 0)))

head(defusers_liq,10)

## # A tibble: 10 x 4
##   user      num num_def percent_def
##   <dbl> <int>   <dbl>       <dbl>
```

```
## 1 9.58e45    2      1      50
## 2 1.39e46    1      1     100
## 3 1.81e46    1      1     100
## 4 2.56e46    1      1     100
## 5 4.28e46    2      2     100
## 6 4.50e46    4      1      25
## 7 5.03e46    6      4     66.7
## 8 5.46e46    3      1     33.3
## 9 5.48e46    1      1     100
## 10 5.68e46   11      3     27.3
```

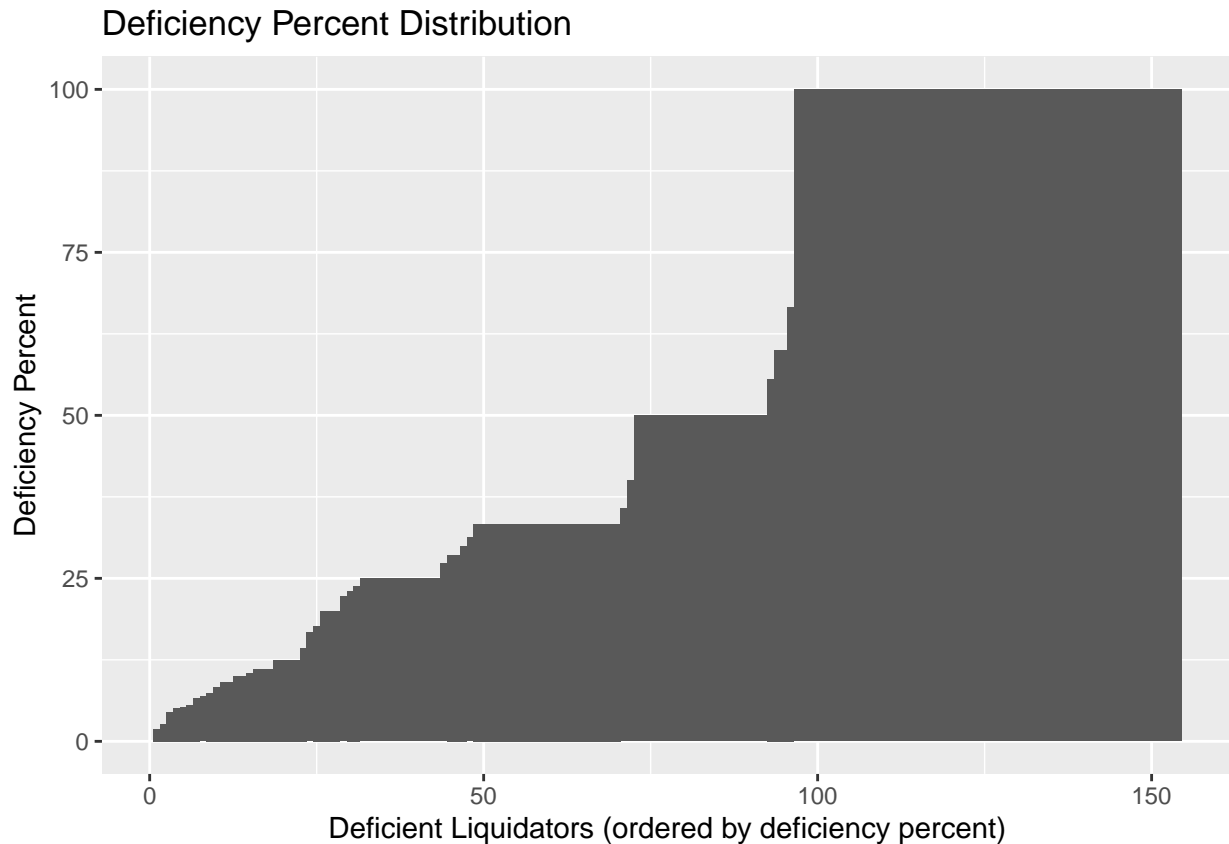
```
defusers_liq <- defusers_liq[order(defusers_liq$percent_def),]

defusers_liq <- defusers_liq %>% rownames_to_column('user_number')

defusers_liq$user_number <- as.integer(defusers_liq$user_number)

plot3 <- ggplot(defusers_liq, aes(y = percent_def, x = user_number)) + geom_bar(stat="identity", position="stack")

plot3
```



Let's take a look into these deficient liquidators more closely. First let's check how regular vs deficient liquidations they make. The table (part is shown) gives us some insights. First, we observe a lot of deficient liquidators made only a few liquidations overall, sometimes even one liquidation in total (and it was deficient). However, there are some users that made a lot of liquidations, sometimes even more than 40 (and up to 5 deficient ones). We can also take a look into deficiency percent (ratio of number of deficient liquidations over all liquidations done by particular user) of the deficient liquidators. The barplot shows the distribution of this parameter over users. Its easy to see that there are some flat parts in the barplot. This is because a lot

of users made a few liquidations and many of them got the same deficiency percent (i.e. 100% is usually 1 deficient liquidation/ 1 liquidation total). Besides this group of few-time liquidators, we can see that some users have their deficiency percent really close to 0, which means they make a lot of profitable liquidations and a few deficient ones may be just a coincidence.

```
#create groups: deficient and regular users
bd_def <- defusers %>% group_by(type) %>% summarize(n=n())%>% mutate(percent = n/sum(n)*100)

bd_reg <- df %>% group_by(type) %>% summarize(n=n())%>% mutate(percent = n/sum(n)*100)

bd_def$group <- 'deficient_liquidators'

bd_reg$group <- 'regular_users'

bd_joint <- rbind(bd_def, bd_reg)

plot <- ggplot(data=bd_joint, aes(x=type, y=percent, fill=group)) +
  geom_bar(stat="identity", position=position_dodge()) + ggtitle("All Users vs Deficient-Liquidation Users")

ggplotly(plot)
```

We can also try to analyze this population of deficient liquidators not only from liquidation perspective, but also from various other points of view. The barplot above shows the breakdown of different types of transaction for deficient liquidators and regular users. We can see that the two groups look quite differently from this point of view. First, deficient liquidators make a lot more liquidations compared to regular users. This is probably due to the fact that there are just a lot more liquidators in the corresponding group, so they are more active with this type of transaction. Secondly, deficient liquidators make more borrows and repays than regular users. This could also be explained from the liquidation perspective: liquidators need to borrow corresponding cryptocurrency to do liquidations, and after they get the collateral from liquidation, they repay their loans.

```
#Deficiency coins

defusers_lc_coins <- defusers %>% filter(type == "liquidation") %>% group_by(collateralReserve) %>% summarize(n=n())
defusers_lp_coins <- defusers %>% filter(type == "liquidation") %>% group_by(principalReserve) %>% summarize(n=n())

#head(defusers_lc_coins[order(-defusers_lc_coins$n),],10)
#head(defusers_lp_coins[order(-defusers_lp_coins$n),],10)

#Regular coins

regusers_lc_coins <- df %>% filter(type == "liquidation") %>% group_by(collateralReserve) %>% summarize(n=n())
regusers_lp_coins <- df %>% filter(type == "liquidation") %>% group_by(principalReserve) %>% summarize(n=n())

#head(regusers_lc_coins[order(-regusers_lc_coins$n),],10)
#head(regusers_lp_coins[order(-regusers_lp_coins$n),],10)

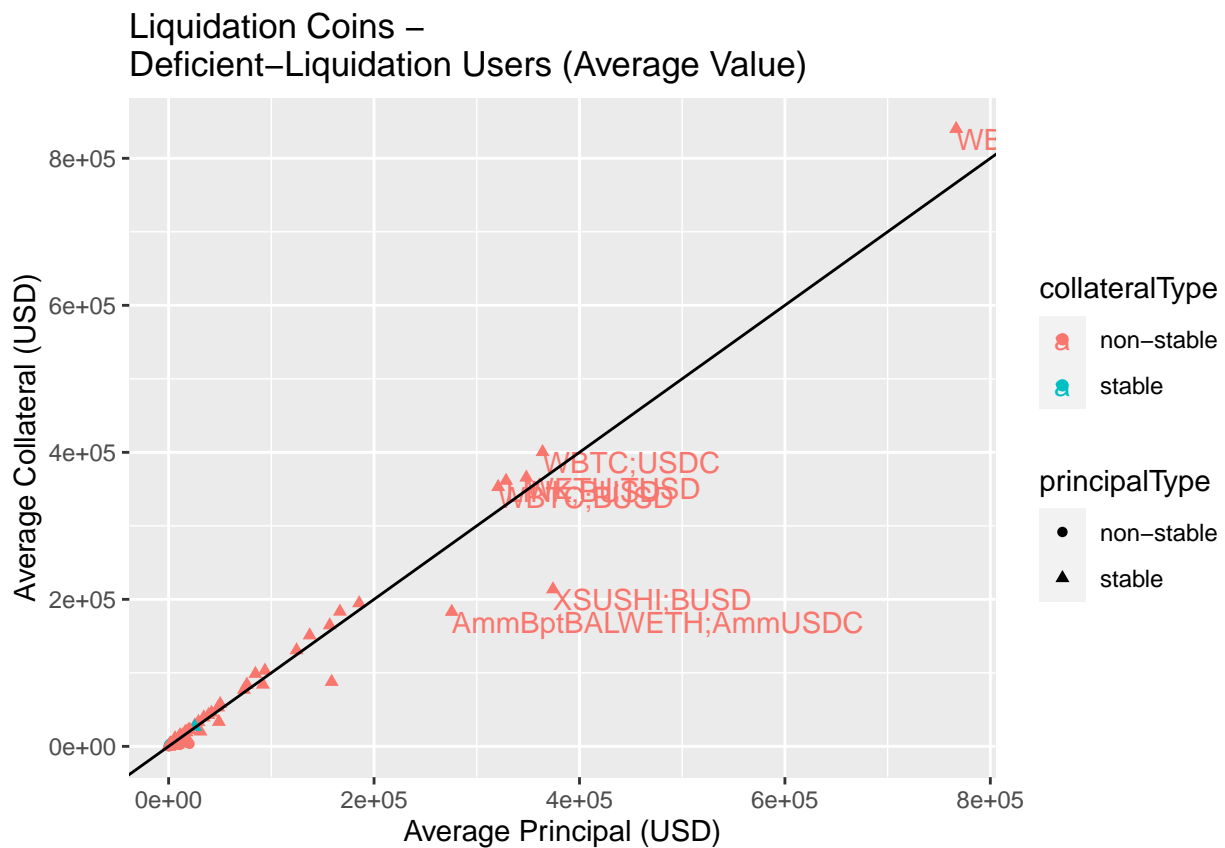
defusers$group <- 'deficient'

df$group <- 'regular'
```

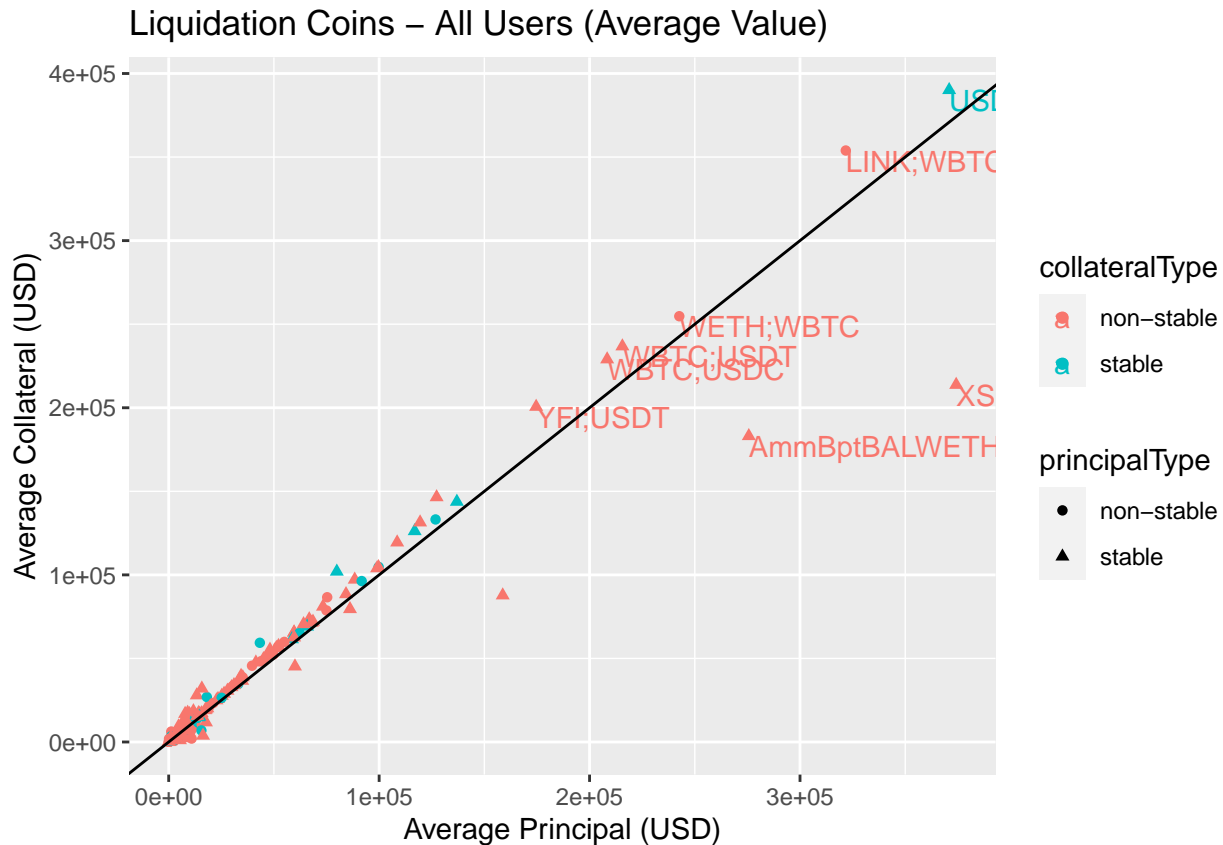
```

LiquidSummaryDef <- defusers %>% filter(type == "liquidation") %>% group_by(collateralReserve, principalReserve)
## `summarise()` has grouped output by 'collateralReserve'. You can override using the `.groups` argument
LiquidSummaryReg <- df %>% filter(type == "liquidation") %>% group_by(collateralReserve, principalReserve)
## `summarise()` has grouped output by 'collateralReserve'. You can override using the `.groups` argument
p_avg_def <- ggplot(LiquidSummaryDef, aes(avg_usd_princ, avg_usd_collat, color = collateralType, shape = principalType))
p_avg_reg <- ggplot(LiquidSummaryReg, aes(avg_usd_princ, avg_usd_collat, color = collateralType, shape = principalType))
p_avg_def

```



p_avg_reg



I also wanted to take a look into different kinds of coins that were popular in deficient vs profitable liquidations. The two graphs above show the most popular coin combinations for liquidations of deficient liquidators and all users correspondingly. We can observe a number of interesting differences, although for the most part (the most popular coin combinations) the graphs look similar. For example, deficient liquidators almost never use stable coins for transactions, but this might just be due to smaller data set for deficient liquidators (liquidations involving stable coins are rare in regular data set as well).

Graphs from Papers

```
#graphs from paper of liquidations
dfst <- df %>% filter(type == "liquidation") %>% filter(amountUSDCollateral < amountUSDPrincipal)

dfst$collateralType <- mapply(coinType, dfst$collateralReserve)
dfst$principalType <- mapply(coinType, dfst$principalReserve)

#dfst

dfstsushi <- dfst %>% filter(collateralReserve == "XSUSHI")

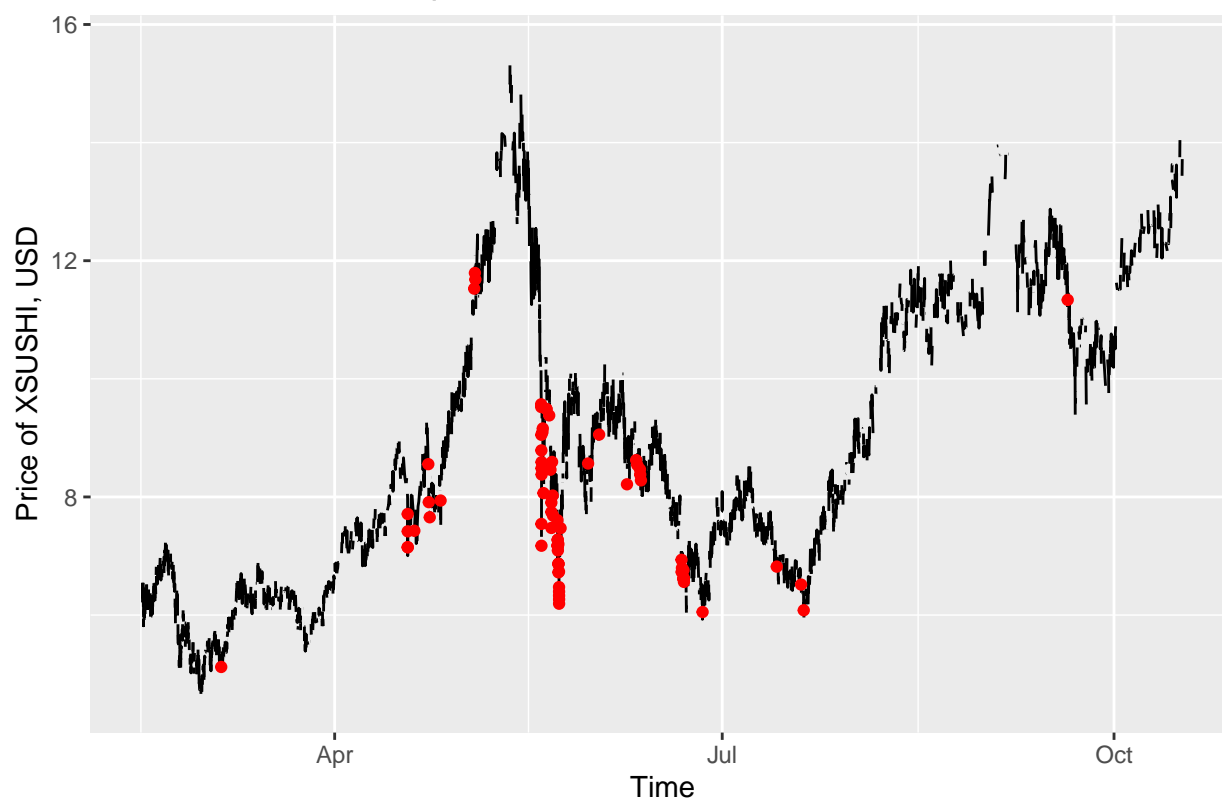
dfsushi <- df %>% filter(reserve == "XSUSHI" )

sushi.plot <- ggplot(dfsushi, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD)) + geom_

#ggplotly(sushi.plot)

sushi.plot
```

Price and Deficient Liquidations of XSUSHI



```
#graphs from paper of liquidations
dfst <- df %>% filter(type == "liquidation") %>% filter(amountUSDCollateral<amountUSDPrincipal)

dfst$collateralType <- mapply(coinType, dfst$collateralReserve)
dfst$principalType <- mapply(coinType, dfst$principalReserve)

dfstenj <- dfst %>% filter(collateralReserve == "ENJ")

dfenj <- df %>% filter(reserve == "ENJ" )

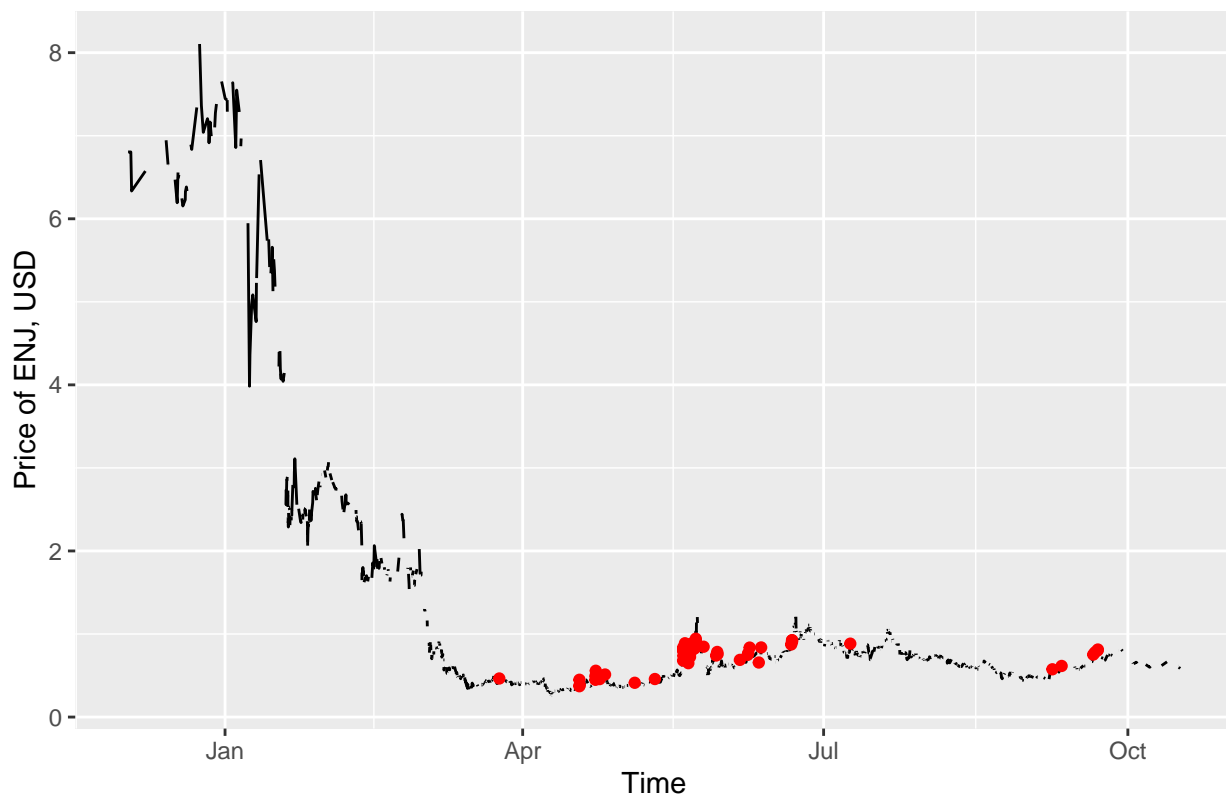
enj.plot <- ggplot(dfenj, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD)) + geom_line

#ggplotly(enj.plot)

enj.plot

## Warning: Removed 1 row(s) containing missing values (geom_path).
```

Price and Deficient Liquidations of ENJ



```
#dfst <- df %>% filter(type == "liquidation") %>% #filter(amountUSDCollateral<amountUSDPrincipal)

#dfst$collateralType <- mapply(coinType, dfst$collateralReserve)
#dfst$principalType <- mapply(coinType, dfst$principalReserve)

#dfstamm <- dfst %>% filter(collateralReserve == "AmmBptBALWETH")

#dfamm <- df %>% filter(reserve == "AmmBptBALWETH" )

#amm.plot <- ggplot(dfamm, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD)) + #geom_li

#ggplotly(amm.plot)

#amm.plot
```

The two graphs above (the third one is commented out for technical reasons) were inspired by the following paper: <https://arxiv.org/pdf/2106.06389.pdf>. The paper, among many other things, proposes the hypothesis for why we could observe non-profitable(deficient) liquidations. According to their explanation, the reason might be just small fluctuations in price of collateral reserves over time. In order to check this hypothesis, I decided to build graphs of price for different coins and put red points at places where deficient liquidations occurred. The reserves were selected based on popularity of collateral in deficient liquidations.

We can somewhat verify the hypothesis from the paper: it looks like a lot of liquidations occur when price drops (at least for XSUSHI). However, we definitely cannot say that the price of collateral explains all deficient liquidations. The price curve for ENJ is much more flat and it looks like most of the liquidations for this reserve did not occur when price dropped.

It is quite hard to study these curves in more details as we do not know prices of reserves at all times: the

graphs were built from limited available information. So, we cannot always clearly say whether price really dropped or went up (AmmBptBALWETH graph was commented out since it was even more discontinuous). For the same reason, I do not include graphs with collateral-principal price ratio, which would otherwise be very useful.

#Other interesting graphs from that paper

```
df1 <- df %>% filter(type == "liquidation")
df1 <- df1 %>% mutate(timestamp = as_datetime(timestamp, tz = "UTC"))
```

#df1

```
dft1 <- df1 %>% group_by(month = lubridate::floor_date(timestamp, "month")) %>% summarize(total_profit =
```

```
dfweek <- df1 %>% group_by(week = lubridate::floor_date(timestamp, "week")) %>% summarize(total_profit =
```

#dfweek

```
dfcum <- df1 %>% arrange(timestamp)
```

```
dfcum <- dfcum %>% mutate(cum_collateral = cumsum(amountUSDCollateral))
```

```
library(scales)
```

```
##
```

```
## Attaching package: 'scales'
```

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

```
##
```

```
##      discard
```

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

```
##
```

```
##      col_factor
```

```
plot_t1 <- ggplot(dft1, aes(x = month, y = total_profit)) + geom_bar(stat="identity", position=position_
```

```
plot_tlw <- ggplot(dfweek, aes(x = week, y = total_profit)) + geom_bar(stat="identity", position=positi
```

```
plot_t2 <- ggplot(dfcum, aes(x = timestamp, y = cum_collateral)) + geom_line() + ggtitle("Collateral So
```

```
df_bor_avg = df %>% filter(type == "borrow") %>% mutate(timestamp = as_datetime(timestamp, tz = "UTC"))
```

#df_bor_avg

```
df_bor_liq = cbind(dft1,df_bor_avg)
```

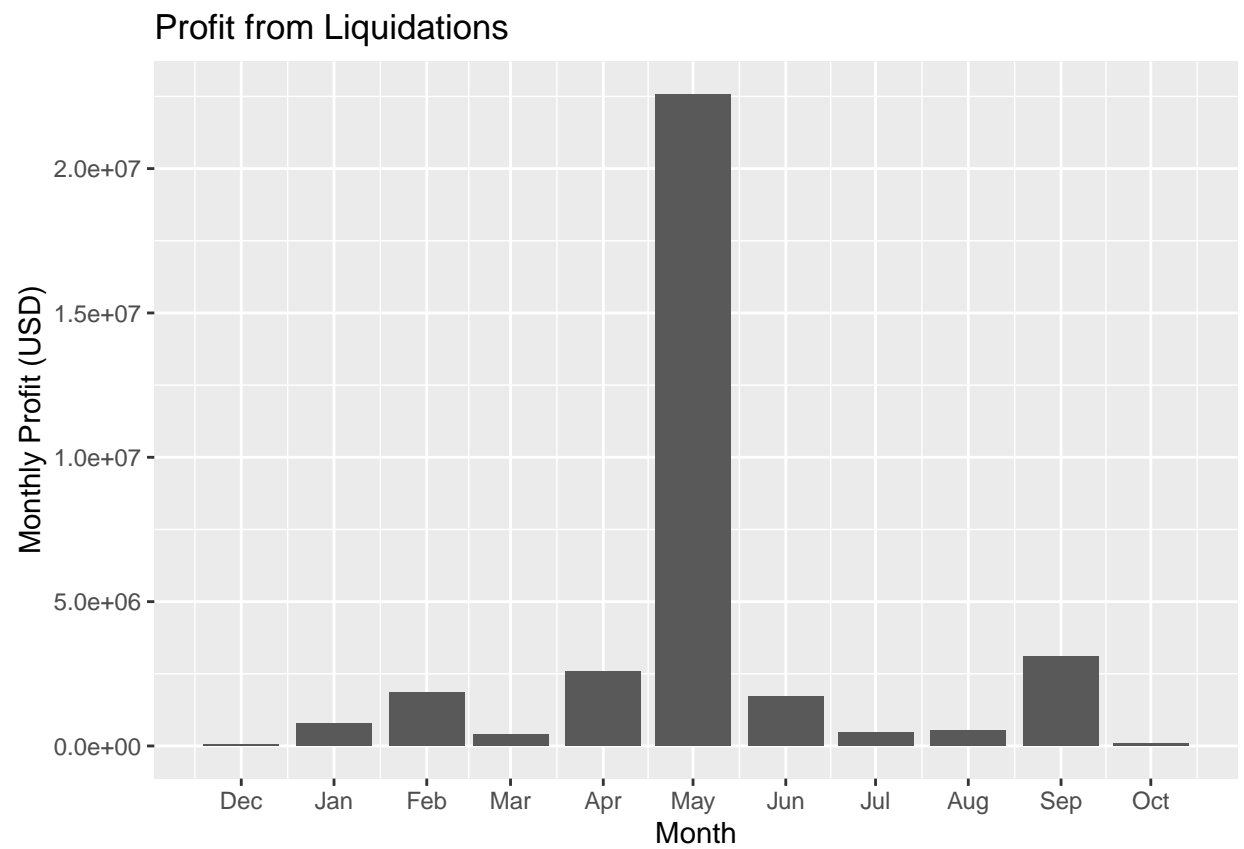
```
df_bor_liq <- df_bor_liq[, !duplicated(colnames(df_bor_liq), fromLast = TRUE)]
```

```
df_bor_liq <- df_bor_liq %>% mutate(ratio = total_profit/total_borrow)
```

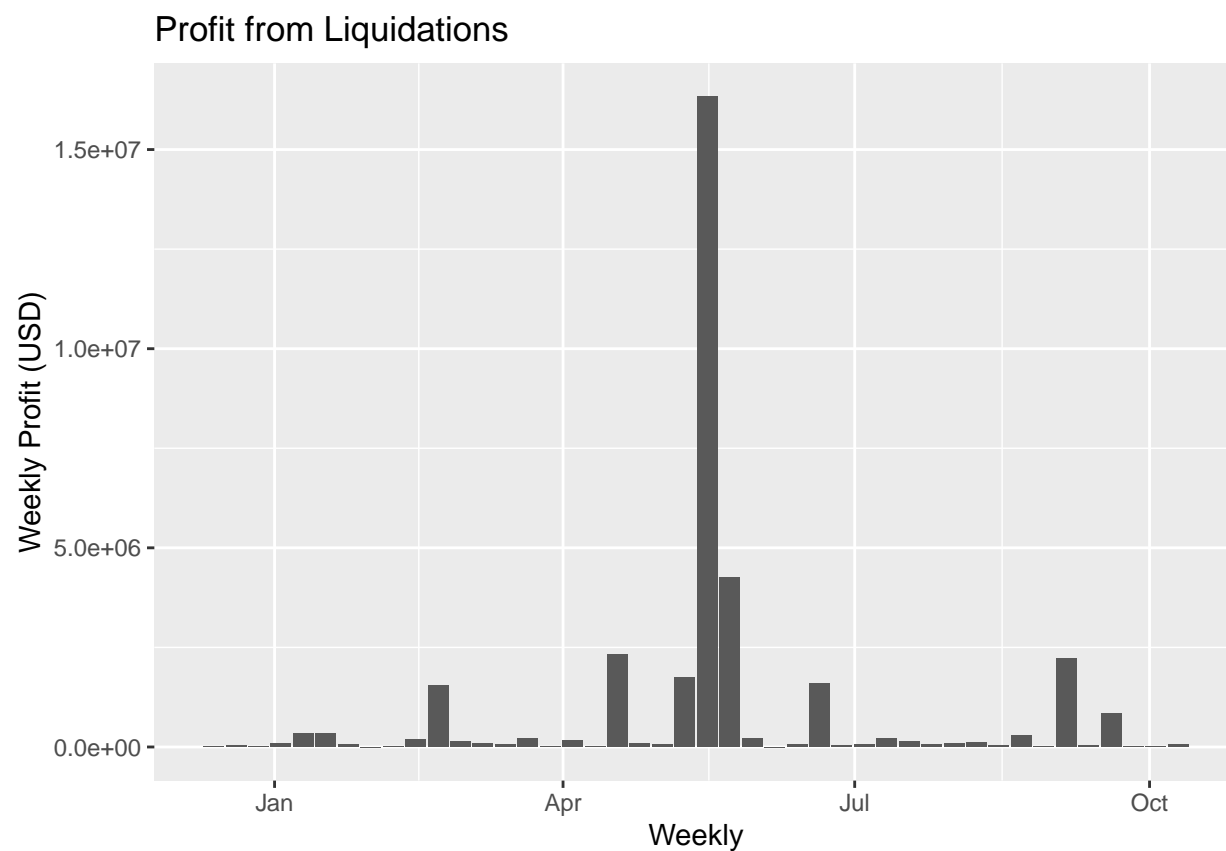
```
library(scales)
```

```
plot_t3 <- ggplot(df_bor_liq, aes(x = month, y = ratio)) + geom_line() + ggtitle("Comparison Between Mo
```

```
plot_t1
```

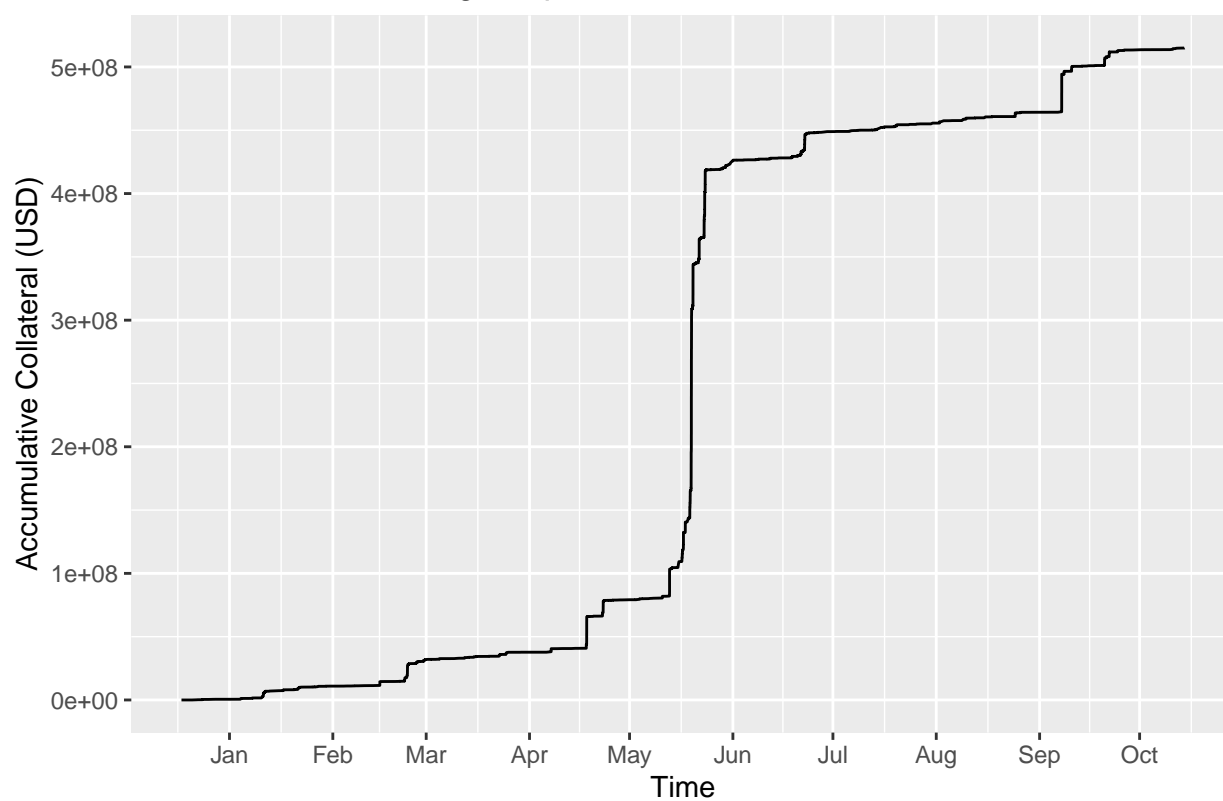


plot_tlw



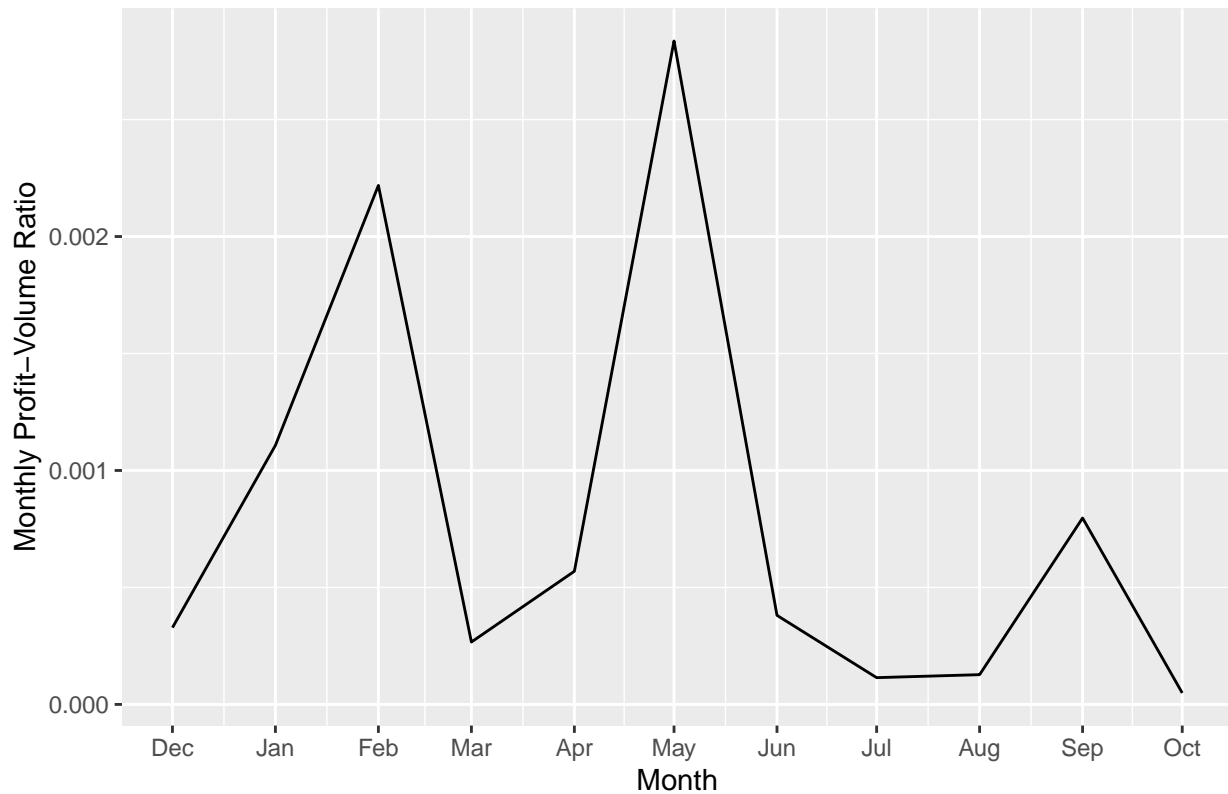
plot_t2

Collateral Sold Through Liquidations



plot_t3

Comparison Between Monthly Collateral Profit and Monthly Borrow Volume



#dfcum

The graphs above were inspired by this paper: <https://arxiv.org/pdf/2106.06389.pdf>. The idea was to try to see what graphs might be useful for the app (for liquidations section). It turned out that a lot of graphs from paper cannot really be built with the data we currently have. However, I was able to build several graphs, which look very interesting. I do not have a lot to say about the contents of the graphs: we have seen most of the trends these graphs show before. I, personally, like the cumulative graph and find it very informative not only for liquidations, but for other transactions as well.

```
dfcumUSDC <- df1 %>% arrange(timestamp) %>% filter(collateralReserve == "USDC")
dfcumUSDC <- dfcum %>% mutate(cum_collateral = cumsum(amountUSDCollateral))
```

```
dfcumDAI <- df1 %>% arrange(timestamp) %>% filter(collateralReserve == "DAI")
dfcumDAI <- dfcumDAI %>% mutate(cum_collateral = cumsum(amountUSDCollateral))
```

```
dfcumWETH <- df1 %>% arrange(timestamp) %>% filter(collateralReserve == "WETH")
dfcumWETH <- dfcumWETH %>% mutate(cum_collateral = cumsum(amountUSDCollateral))
```

```
plot_t4 <- ggplot(dfcumUSDC, aes(x = timestamp, y = cum_collateral)) + geom_line() + geom_line(data = d
```

```
df_bor_avg = df %>% filter(type == "borrow") %>% mutate(timestamp = as_datetime(timestamp, tz = "UTC"))
```

#df_bor_avg

#plot_t4

```
dfst <- df %>% filter(type == "liquidation") %>% filter(collateralReserve == "WETH")
```



```

dfst$collateralType <- mapply(coinType, dfst$collateralReserve)
dfst$principalType <- mapply(coinType, dfst$principalReserve)

dfstweth <- dfst %>% filter(collateralReserve == "WETH")

dfweth <- df %>% filter(reserve == "WETH")

#dfstweth

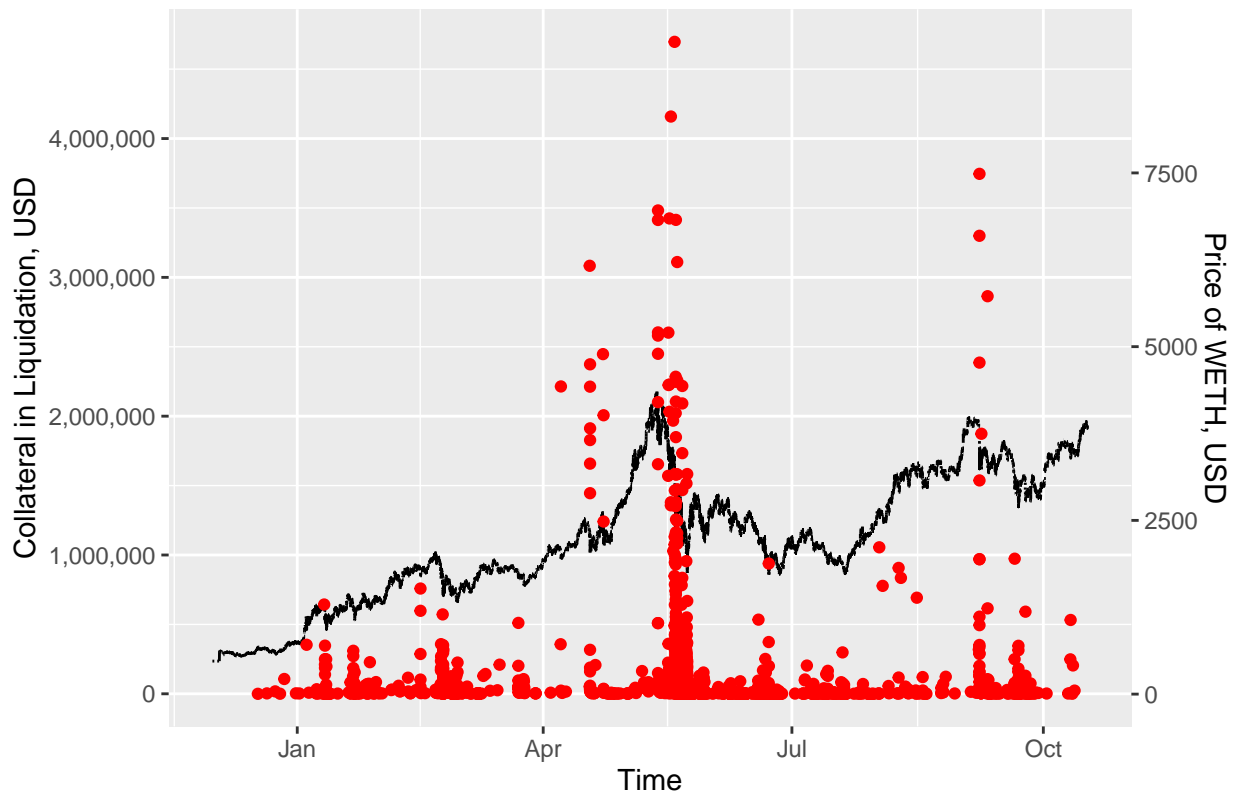
r <- 500
weth.plot <- ggplot(dfweth, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD*r)) + geom_

#ggplotly(weth.plot)

weth.plot

```

Price and Liquidations of WETH as Collateral



```

dfstweth <- dfstweth %>% mutate(timestamp = as_datetime(timestamp, tz = "UTC"))

dfstweth <- dfstweth %>% mutate(week = lubridate::floor_date(timestamp, "week"), day = lubridate::floor_date(timestamp, "day"))

dfplotday <- dfstweth %>% group_by(day) %>% summarize(daily_liquid = sum(amountUSDCollateral))

dfplotweek <- dfstweth %>% group_by(week) %>% summarize(weekly_liquid = sum(amountUSDCollateral))

r <- 2000
weth.plotday <- ggplot(dfweth, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD*r)) + geom_

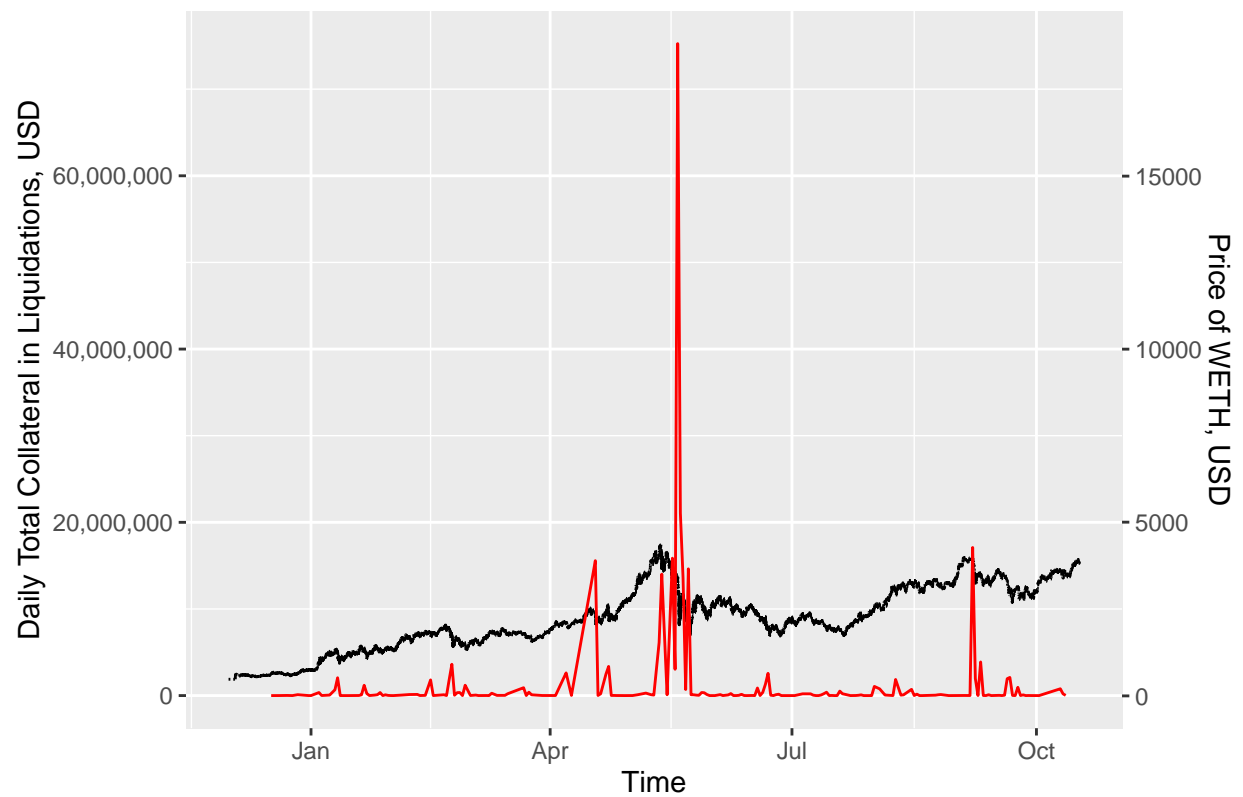
```

```
r <- 4000
```

```
weth.plotweek <- ggplot(dfweth, aes(x = as_datetime(timestamp, tz = "UTC"), y = reservePriceUSD*r)) +
```

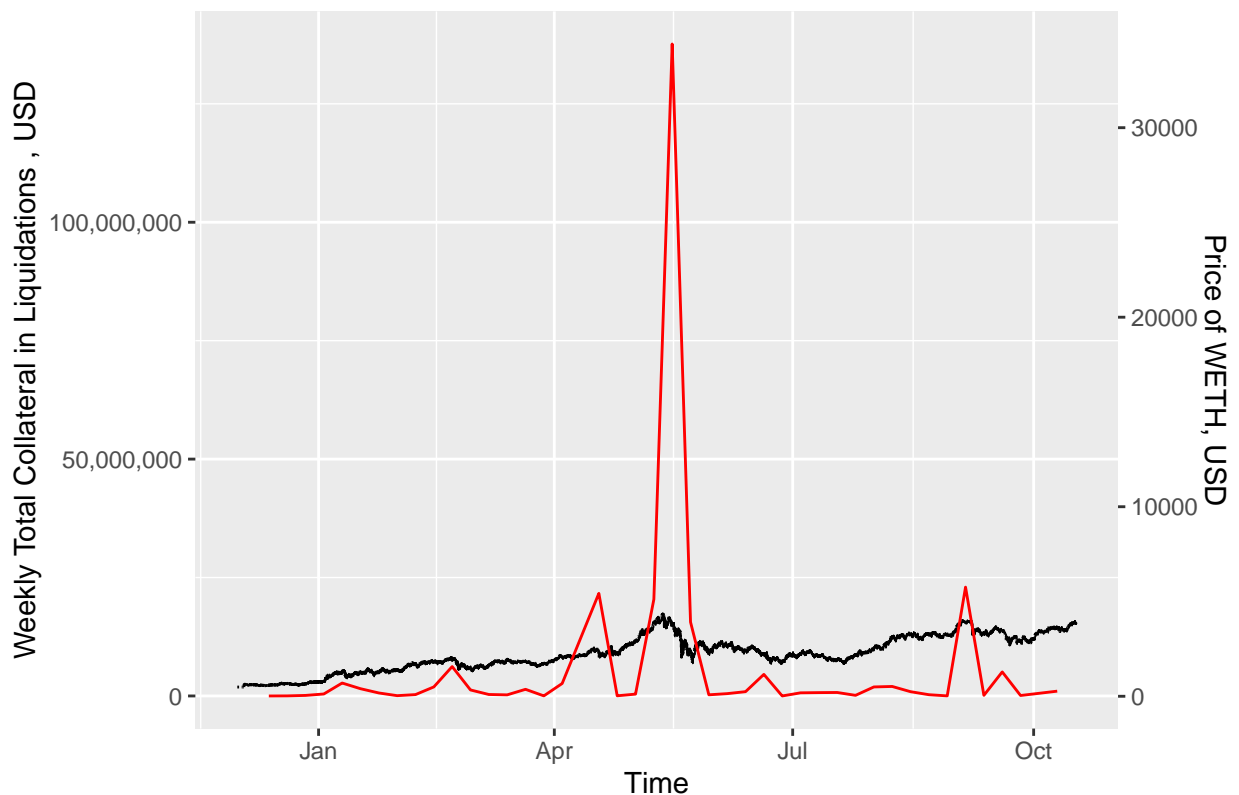
```
weth.plotday
```

Price and Liquidations of WETH as Collateral



```
weth.plotweek
```

Price and Liquidations of WETH as Collateral



These three graphs were inspired by different paper: <https://arxiv.org/pdf/2009.13235.pdf>. The idea was, again, to see if we can use these kinds of graphs in our app. I used some variation of Cole Pacquin's code for the second axis. The black line shows the price of WETH and red points or lines show amount of collateral in liquidations.

Speaking of the second graph (daily liquidations), it looks like we can say something about relationship between price of collateral reserve and liquidations. From the graph, we can see that there are spikes in daily total collateral amounts usually at times when collateral price drops. However, it is not always true: the spike in April seems to be one counterexample.

Good Liquidators

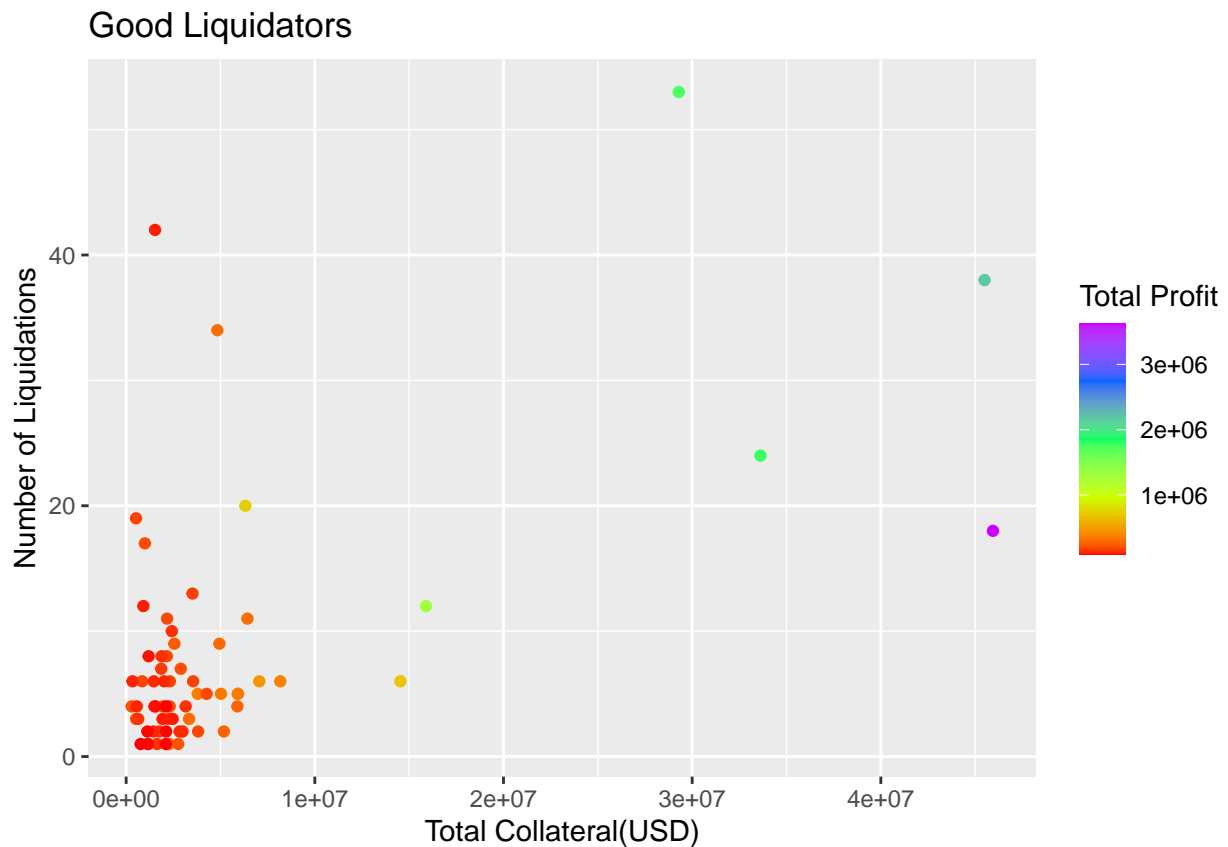
```
#Good liquidators: those that make profit from liquidations
#Extension from previous work

df1 <- df %>% filter(type == "liquidation") %>% group_by(user_alias) %>% summarize(total_collateral = s
df1 <- df1[order(-df1$total_profit),]

df1 <- df1 %>% filter(total_profit > 100000 )

#df1

plot_liq <- df1 %>% ggplot(aes(x = total_collateral, y = num_transact, color = total_profit)) + geom_po
plot_liq
```



```

dfl_all <- df %>% filter(df$user_alias %in% dfl$user_alias)

#dfl_all

bd_liq <- dfl_all %>% group_by(type) %>% summarize(n=n()) %>% mutate(percent = n/sum(n)*100)

bd_liq$group <- 'good_liquidators'

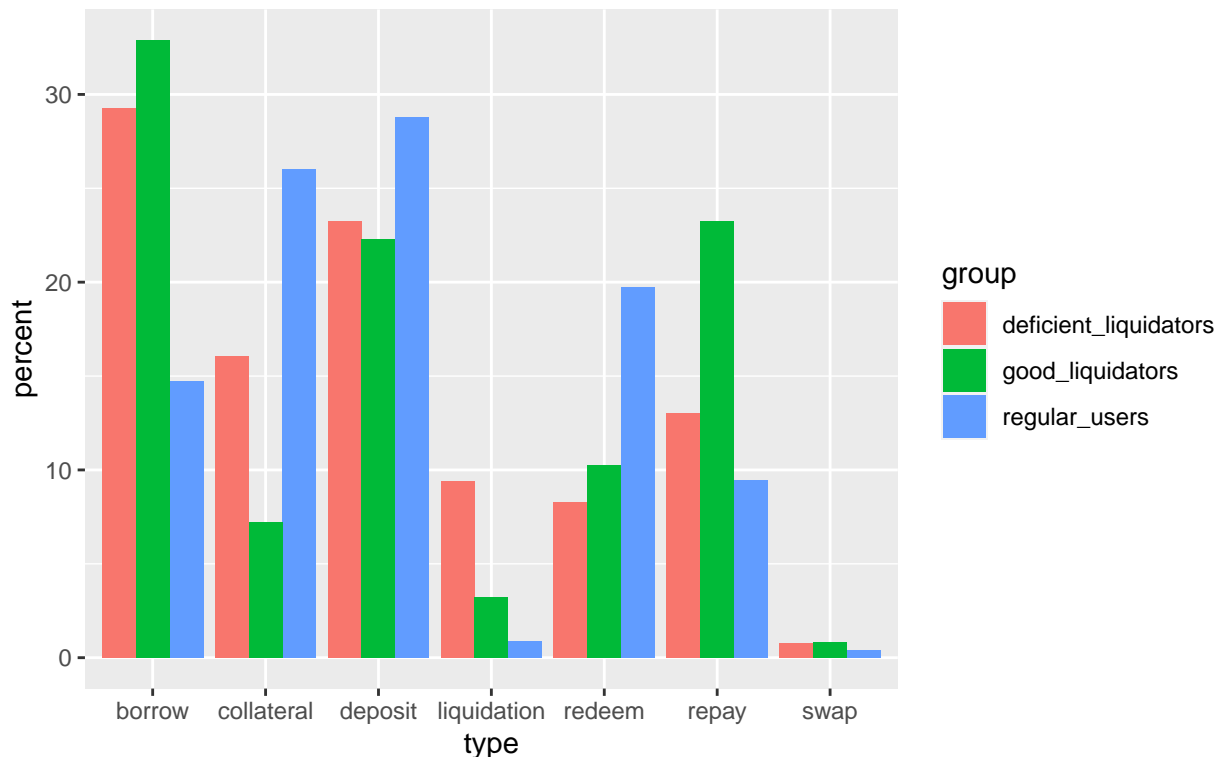
bd_joint <- rbind(bd_joint, bd_liq)

plot <- ggplot(data=bd_joint, aes(x=type, y=percent, fill=group)) +
  geom_bar(stat="identity", position=position_dodge()) + ggtitle("Deficient Liquidators, Best Liquidators")

plot

```

Deficient Liquidators, Best Liquidators, and All Users by Types of Transactions



The graph and barplot above show the information on best liquidators (liquidators that made at least 100000 USD profit from liquidations). The plot compares number of liquidations and total collateral claimed in all of liquidations for best liquidators. Colors correspond to amount of profit (collateral - principal) gained in these transactions.

One trend that we observe is that the most profitable liquidators (green, blue, purple points on graph) all have more than 10 liquidations and, most importantly, also claimed a lot of collateral. So, it looks like liquidators have to liquidate loans of high volume in order to make most of the profit. The strategy of just doing a lot of low-volume liquidations does not show to be profitable.

The barplot extends the barplot from the beginning of this notebook by adding good_liquidators column. So, we can now compare all three groups: deficient liquidators, good(best) liquidators, and regular users. As we can see from the barplot, all three groups look very different in terms of types of transactions they do. Speaking of good liquidators in particular (the two other groups were summarized above), they seem to do less liquidations compared to deficient liquidators. We can guess that they probably aim for high-volume liquidations more rather than number of liquidations and also they probably do a lot of other transactions just to collect more profit. High numbers for borrow and repay could have the following explanation: good liquidators have to borrow a lot in order to collect money for liquidation, once a loan is liquidated they pay back the loans.

```
#Time-series plot good vs normal liquidations
dfl_all_time <- dfl_all %>% filter(type == "liquidation")

count(dfl_all_time)
```

```
##      n
## 1 547
```

```

dfliq <- df %>% filter(type == "liquidation")

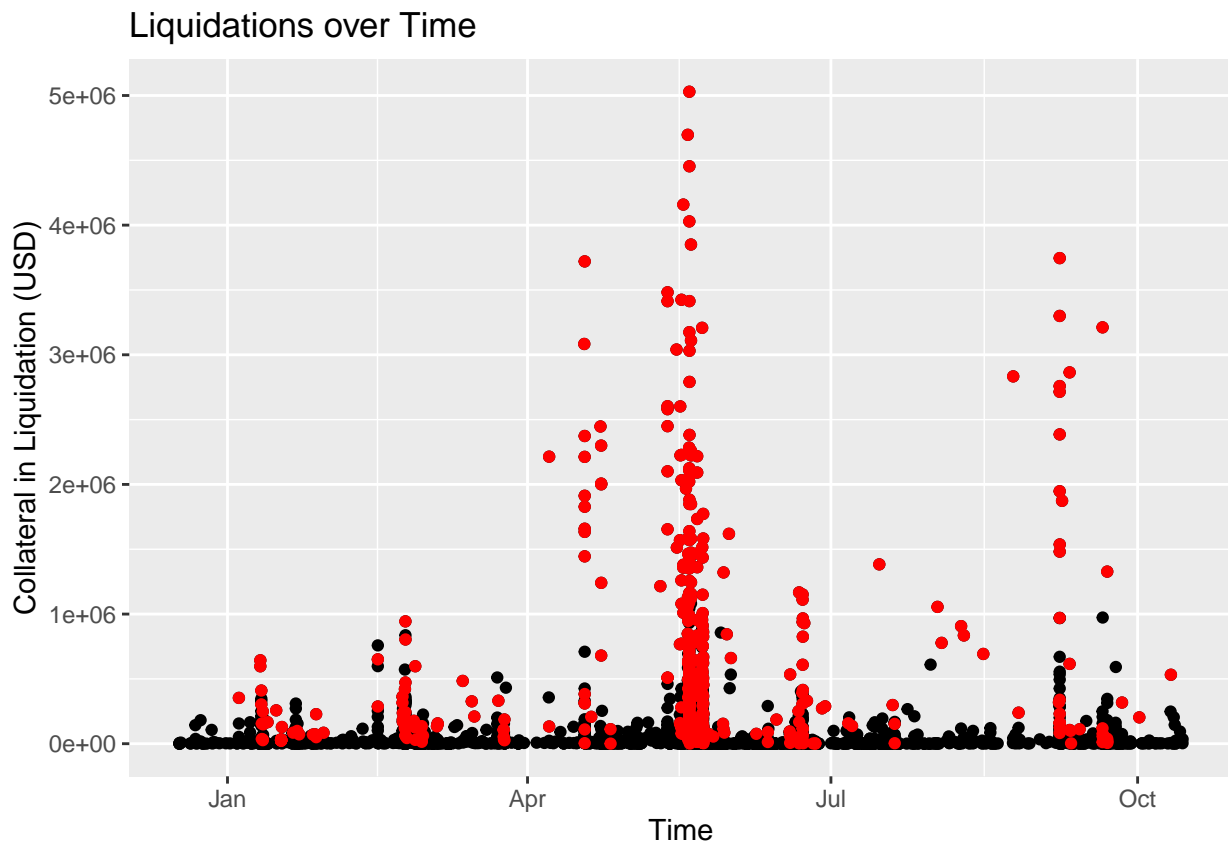
count(dfliq)

##           n
## 1 6731

#print(dfliq)

dfl_time_plot <- ggplot(data = dfliq, aes(x = as_datetime(timestamp, tz = "UTC"), y = amountUSDCollateral))
dfl_time_plot

```



Let's now compare how many liquidations are due to good liquidators. First, let's take a look at the graph of liquidations over time. The black points are liquidations by all users and red points are liquidations by the good liquidator group. Visually, it looks like good liquidators make about 50% of liquidations, which is not true. Actually, there are only 547 liquidations from good liquidators, whereas all users made 6731 liquidations total (so about 8% are from good liquidators). Another interesting observation is that all points that have a lot of collateral are red, which means good liquidators focus more on those types of liquidations (and make their profit due to doing these liquidations). The high-volume liquidations are possible only at times of instability (i.e. Chinese Crypto Ban), which is very good for liquidators.