MATP-4910 Final Project Notebook

DAR Project- DeFi

Cole Paquin

15 November 2021

Contents

Introduction

- github repository: dar-paquic
- Your github ID: paquic
- Final notebook: *PAQUIC_final_notebook.Rmd*

Github Additions: paquic-Assignment7-f21.Rmd, PAQUIC_final_notebook.Rmd, PAQUIC_final_notebook.pdf

Github Issues: 93, 95, part of 96

All code is mine.

Overview & Problems Tackled

I focused on three overall topics. These topics are user clustering, transaction trends analysis, and coin utilization and uses. For user clustering, I wanted to see if we could identify how users interact in AAVE and if we can see trends of users that may be more likely to liquidate. Then, I wanted to see how AAVE has evolved over time and how users are now handling their coins. Finally, I looked into how each individual coin has been used and if different types of coins are used differently.

Data Description

Our data is made up of 745,612 different transactions in the AAVE v2 lending protocol. These transactions are a total of 7 types (deposit, redeem, borrow, repay, swap, liquidation, collateral). There are a total of 53 different coins in the data. These are divided into stablecoins, non-stable coins, and Amm coins. For each

transaction, we have a total of 33 columns which record everything about the transaction from the date to the user and the amount of money involved.

Results

Problem 1- We are able to cluster users into groups that allow us the see the behavior of users who are often liquidated. Most users are not overly active and prefer to let their coins sit and make whatever interest they can

Problem 2-Although the number of active users in AAVE has not increased, people are using larger amounts of money in their transactions. AAVE;s market is heavily subject to shock events that will briefly disrupt transaction patterns.

Problem 3- Stable coins are borrowed at much higher rates than other coins, and their is a wide variety of uses for these other non-stable coins. Stable coins all seem to high fairly similar patterns in their utilizations rates.

Problem 1

The first problem I looked into was user clustering. This provides us with an understanding of what different people are trying to do to make money. Can we see patterns in different users? Not only do we want to know the clusters, but we also want to be able to see what distinguishes the different clusters.

Methods

I created a new dataframe where each row represented an individual user. Each user has certain statistics for each type of transaction, as well as the amount of time they have been active. The first few rows are shown below.

```
## `summarise()` has grouped output by 'user'. You can override using the `.groups` argument.
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(ntypes)` instead of `ntypes` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

user	user_alias	timefirst	timelast	N	timeactive	total_borrow	nborrow
1.000000e+00	Tracy Nisbett	2021-07-22	2021-07-22	1	0 days	NA	NA
5.700500e+04	Millicent Gandhi	2021-02-22	2021-07-12	2	140 days	NA	NA
2.577533e+33	Steven Ferguson	2021-02-13	2021-10-09	60	238 days	NA	NA
6.663597e+34	Ronald Masella	2021-05-29	2021-06-01	10	3 days	9.447565	2
4.867107e+35	Steven Ahumada	2021-04-25	2021-04-25	4	0 days	NA	NA
8.009427e+35	Melissa Esters	2020-12-05	2020-12-25	4	20 days	NA	NA

From here, I used pca and kmeans clustering to build the actual clusters.

Results

With a bit of work we can select the columns we would like to cluster on, as well as manually scaling the database. To preserve the original users dataframe, we save the scaled data separately. Also, I also removed outlier users who we either know or suspect of being SMART Contracts. They throw off clustering in a way we do not want. From their, we can run PCA.

my.pca
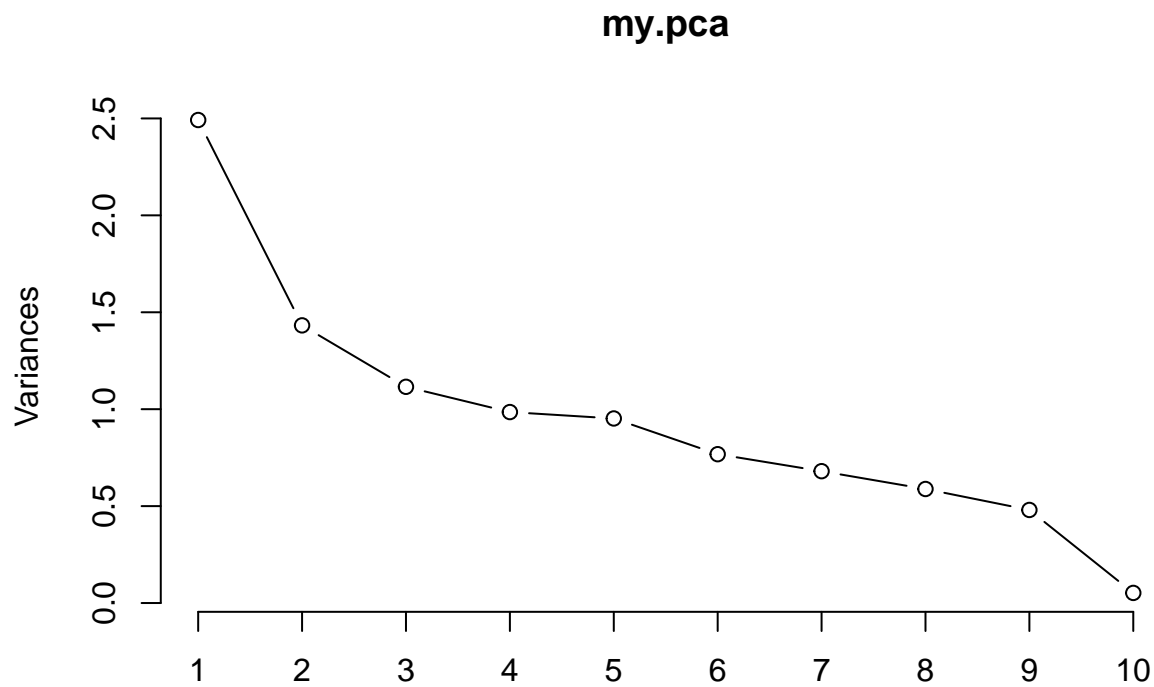
Plotting the PCA allows us to chose the number of cluster we wish to have. In this case, three appears optimal, however I chose to work with 5. If I chose 3, we lose some of the distinctions between clusters and they appear to mean less. Thus let's look at what the clusters are and how users are distributed.

Kmeans Cluster Size

First, we can see that it appears most users get put into cluster 3. This should represent our more average user, as opposed to cluster 2. Until we see the cluster centers we cannot draw any real conclusions.

Color Key and Histogram

Kmeans Cluster Centers
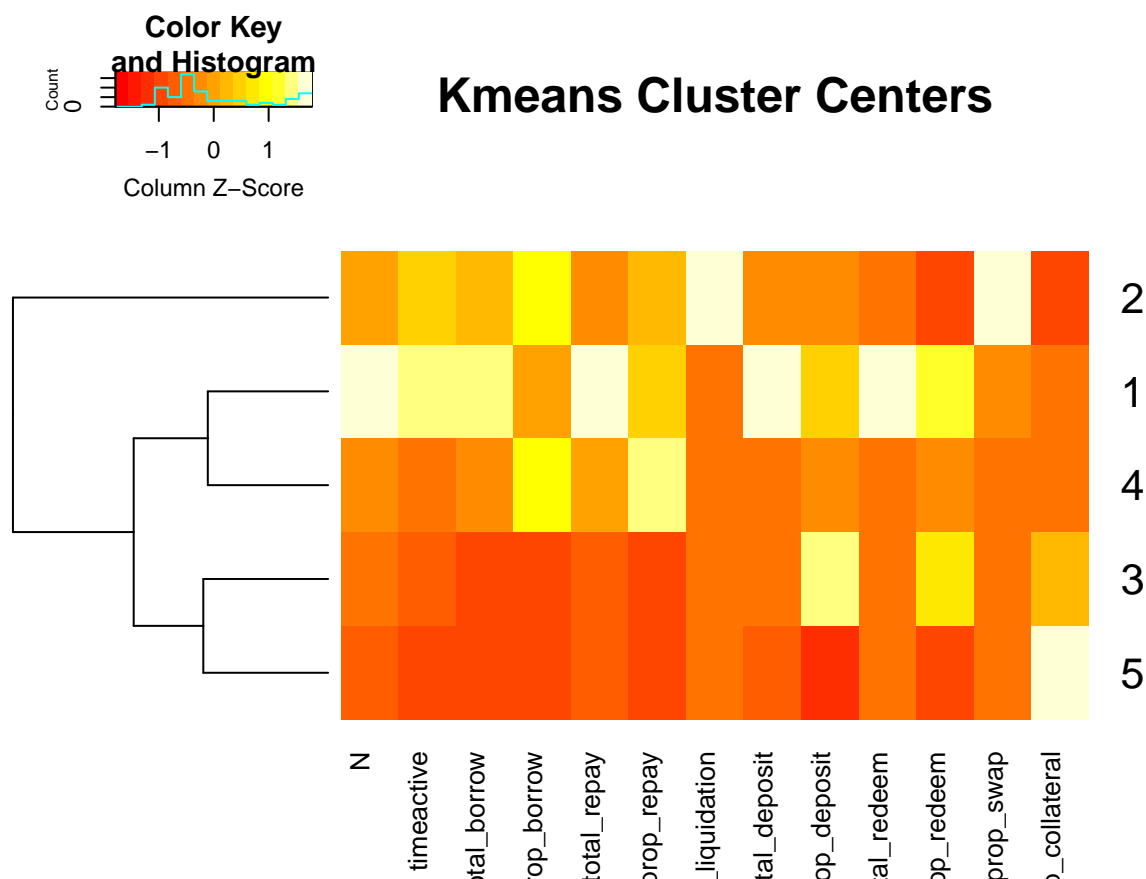
From this we can start to see how users are actually using AAVE. Cluster 3, our largest cluster, is not very active. They mostly deposit and redeem their money, seemingly focused on just earning a return on their cryptocurrencies. Cluster 2, our outlier cluster, contains the users who liquidate the most. What is interesting, is that they also seem to be the ones who swap between variable and stable interest rates. Both of these transaction types are significantly more prevalent in cluster 3 than in any other cluster. Cluster 1 appears to be our "Best" users. They are more active and have been around for a long time. However, they do not liquidate very often, meaning they are good a maintaining a good health factor. Cluster 4 is our main borrowing groups, who take out loans but do a good job of repaying them. Finally, cluster 5 is very inactive other than in collateral switches, which could be further investigated, but I did not. Now we want to see how these transactions work over time.

Transaction Types Over Time by User Cluster
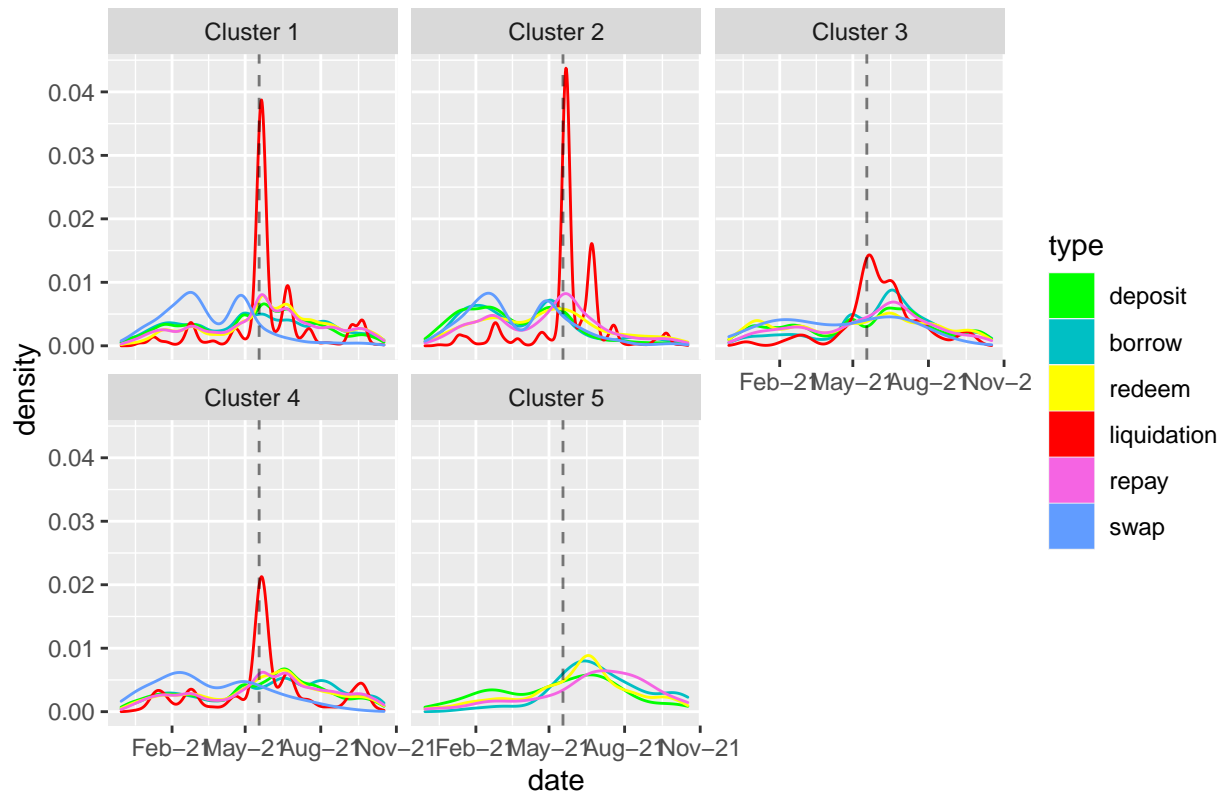
It should be noted, in any graphic if there is a dotted vertical line, it shows the day (May 18th 2021) when China announced their ban on crypto. As we can see in this case, and it others, this caused a large shock to the market. Our borrowing clusters (1, 3, 4) were all heavily impacted and saw increases in liquidations.

Discussion

What we learn from this is that we can cluster users based on previous transactions and get some meaningful results. People borrowing money are far more susceptible to market shocks. On the other hand, we also have seen that many users are not overly active and are simply deposit their money to try and get a return. Also, we are now able to examine users who are more prone to being liquidated (those in cluster 2).

Problem 2

Not only do we want to see how clusters interact in AAVE over time, we also want to see how the overall trends in AAVE work.
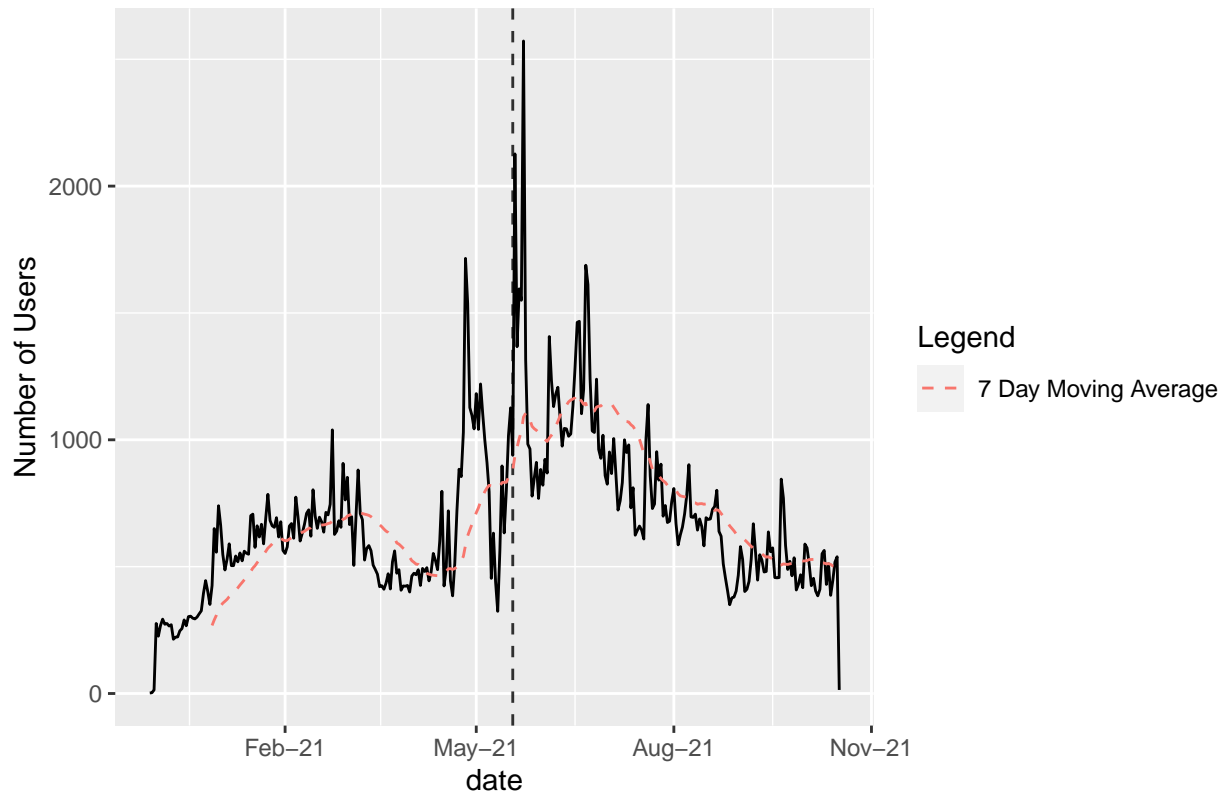
Methods

Most of this section involves a lot of work in ggplot to visualize different actions. It also involved me making a weekly data frame, where users transactions are averaged out to a weekly basis.

Results

Before analyzing trends in AAVE over time, I think it's a good idea to understand the scale of what we are working with.

How many people use AAVE every day?

We can see than recently there have been a little more than 500 people using AAVE v2 every day. This number peaked in May, soon after the China announcement, when the weekly average was over 1,000 users. What is interesting is that there has not really been an increase in the number of people per day, although the cumulative users is growing at a fairly constant rate. This means that many users are not active most days and some may have left the platform altogether.

Number of Transactions per Week
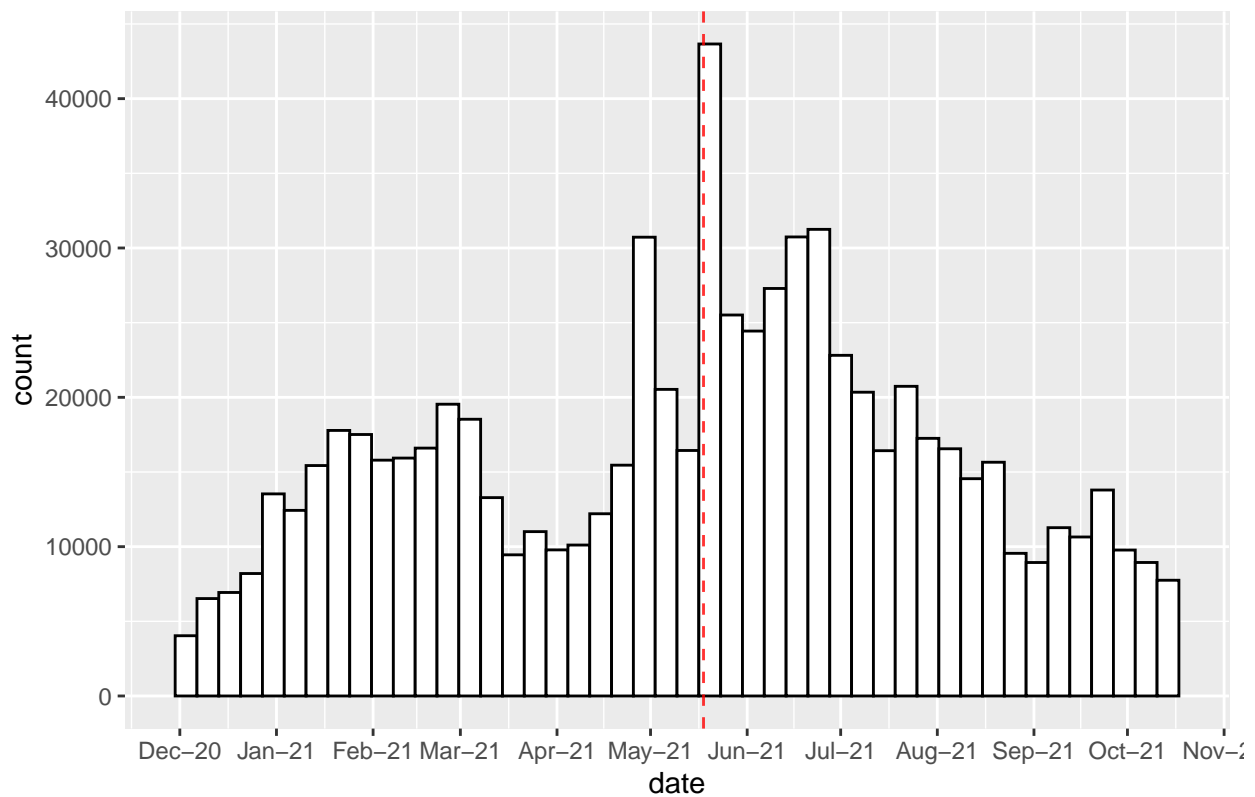
We can see that the chart of total transactions seems to follow pretty closely to the number of users. This is good, since I do not have to look into any weird descepancies. With that being said, we still want to further understand what is happening.

```
## `summarise()` has grouped output by 'user'. You can override using the `.groups` argument.
```

## How often do people use AAVE in a week?
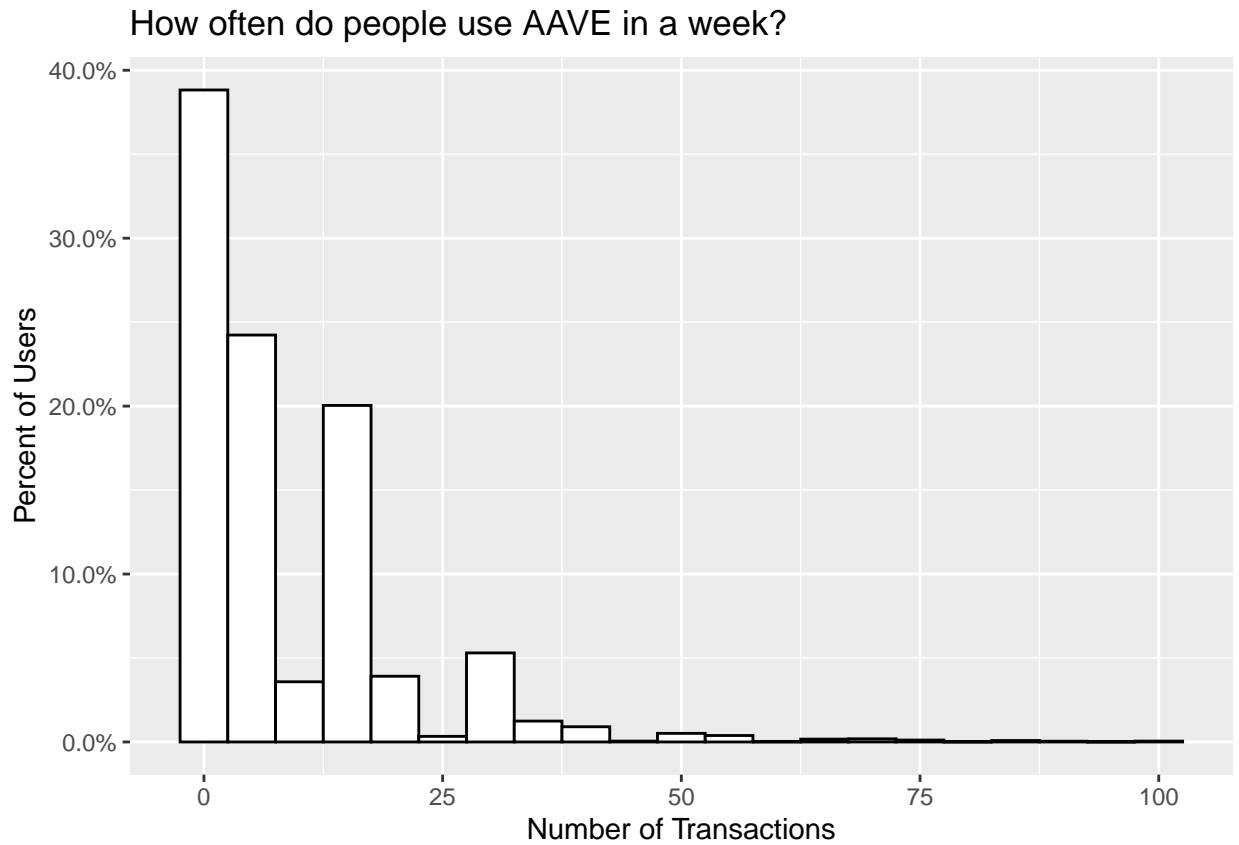
This confirms what I found earlier about many users not being very active. The majority of users have less than 10 transactions per active week. However, there are many outliers that are very active and fall into my first cluster. However, this does do a good job of showing why cluster 3 is the largest.

Number of Transactions per Day
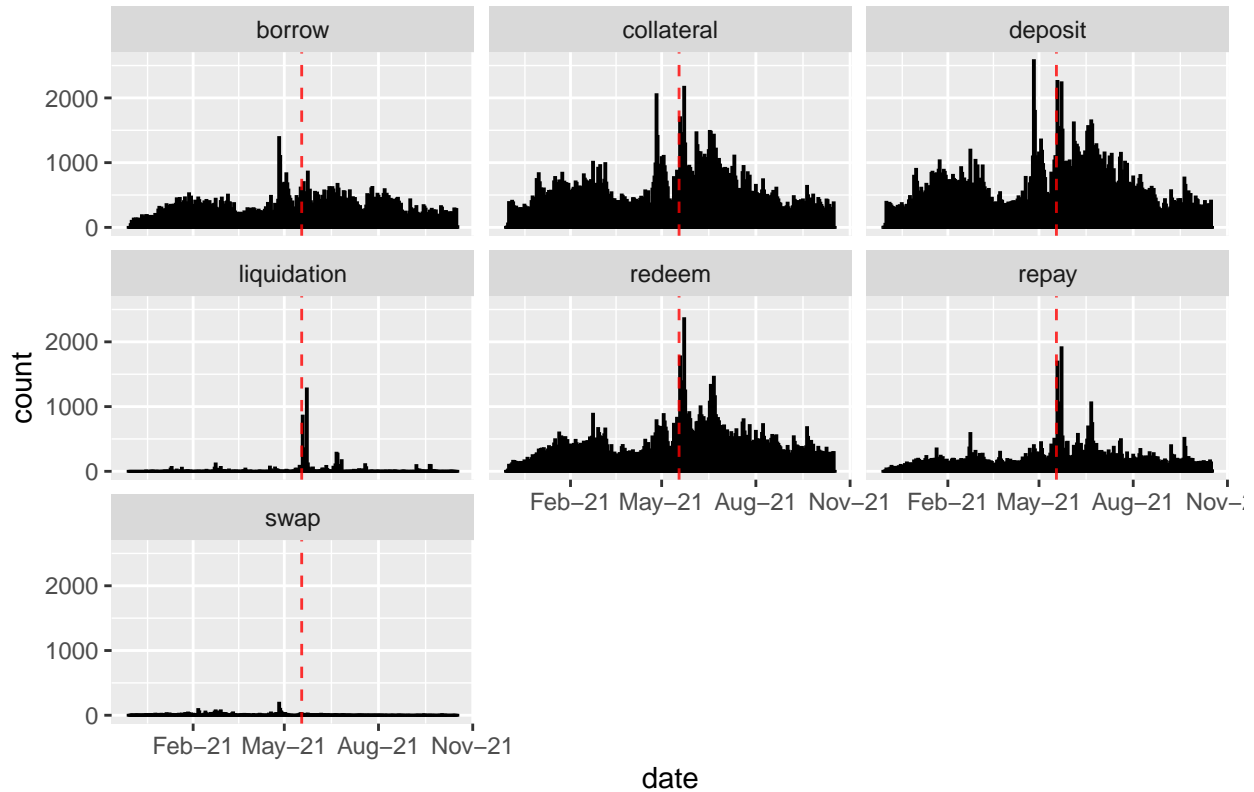
This proves how shock events impact AAVE. We can see peaks in most types of transactions immediately following May 18th. We note that borrows seemed to be unaffected by the news, although a lot of people soon repayed their loans. In the redeem chart, we can see how a lot of people began to take their money out of the protocol, perhaps not trusting the falling prices of cryptocurrencies. However, what these charts lack is a way to see if the number of transactions is always related to the amount of money being moved. To do this, we need to build charts with mutiple scales.

Deposits Summary

Borrows Summary
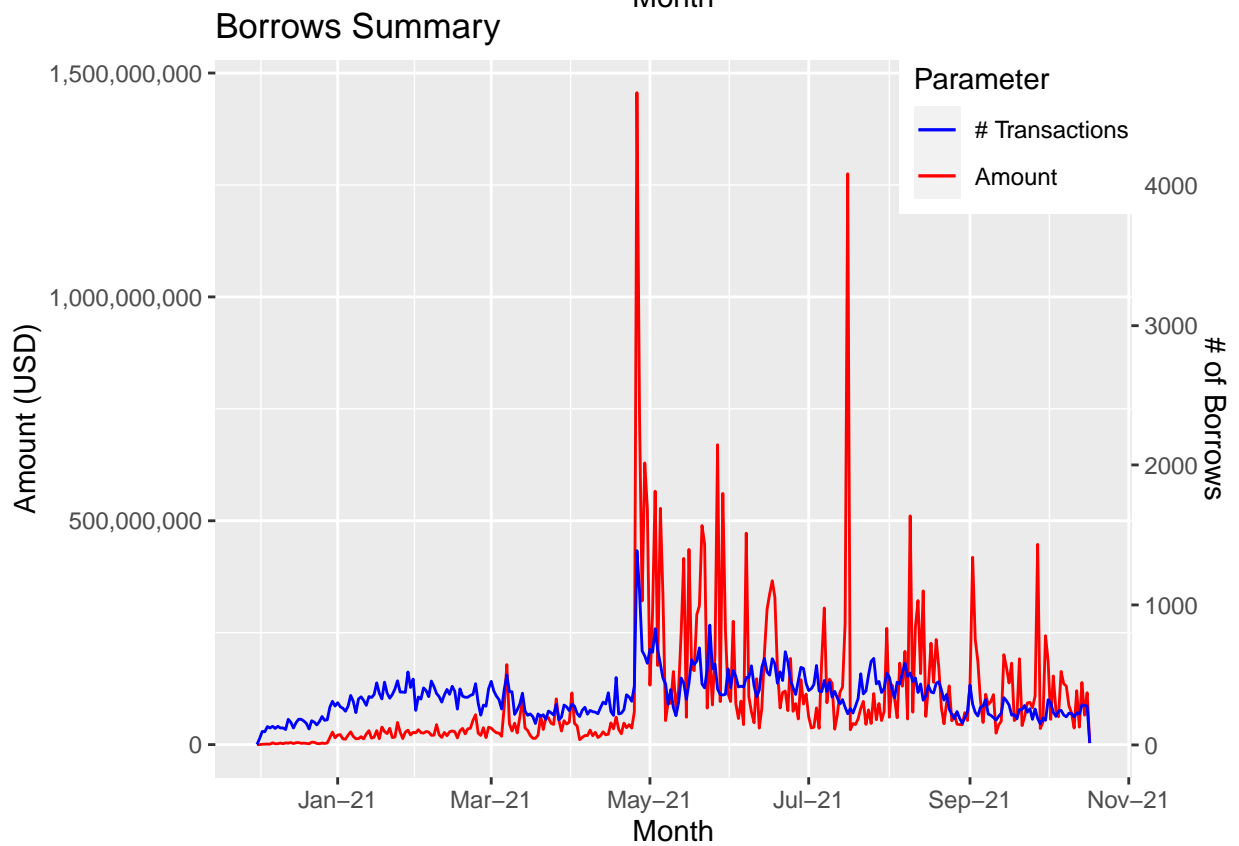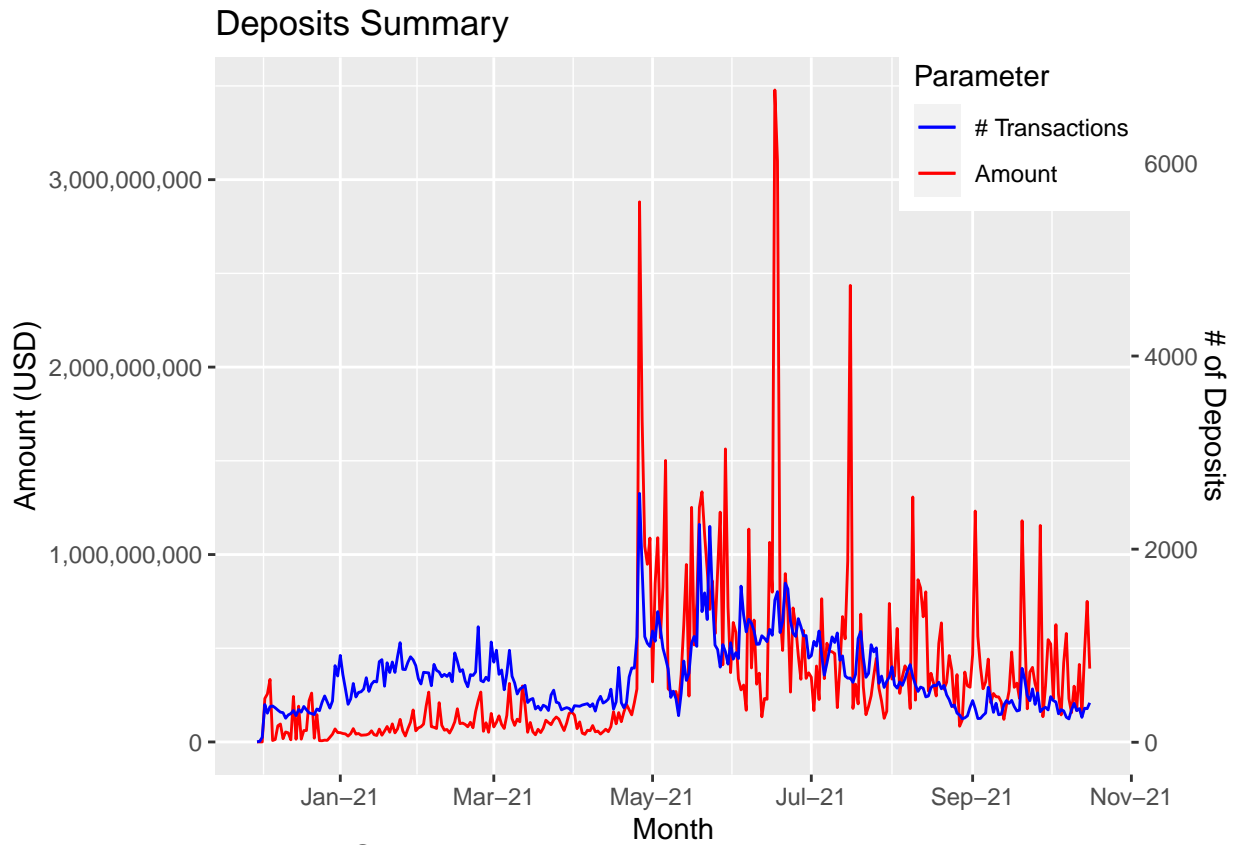
These show examples of how we can take a closer look at each type of transaction. In this case, I only showed

borrows and deposits. We can see similar trends in both of these. Over time, the number of transactions per day has not greatly changed. However, we are now seeing much larger sums of money being used. In recent months, we can also see an increase in "shock" days, where the amount of money drastically increases for one day before returning to its normal level. Could there be shock events on all of these days?

Discussion

We can see several trends that stand out. Throughout time, although the number of transactions are not increasing, we are seeing an increase in the USD amount of transactions. Also, even though the daily number of users is not increasing AAVE is continuing to attract new users. Finally, we saw that shock events can really impact the protocol as a whole and fill actively draw new activity in almost every transaction type.

Problem 3

Coin Utilization (Total Borrows / Total Liquidity) is the main factor influencing the different interest rates. Each coin has its own optimal utilization rate, and a differing actual utility will move the interest rates. In order to make money in AAVE, you must either be able to know how the prices or interest rates of coins will move. For example, AMPL has recently had a borrow interest APY of more than 87,000%. If we are able to better predict how these interests rates will move, we will be able to make money. First we must gain a better understanding of how utilization works and how we can visualize it.

Methods

I used utilization to be the following formula: (Total Borrows) / (Total Deposits - Total Redeems + Total Repays), and then converted it into a percentage. We can get this for every coin over time and then are able to graph them relative to other coins.

Results

First, it is helpful to fully understand what utilization looks like. The following chart will show how USDC is being used over time.

USDC Utilization Summary
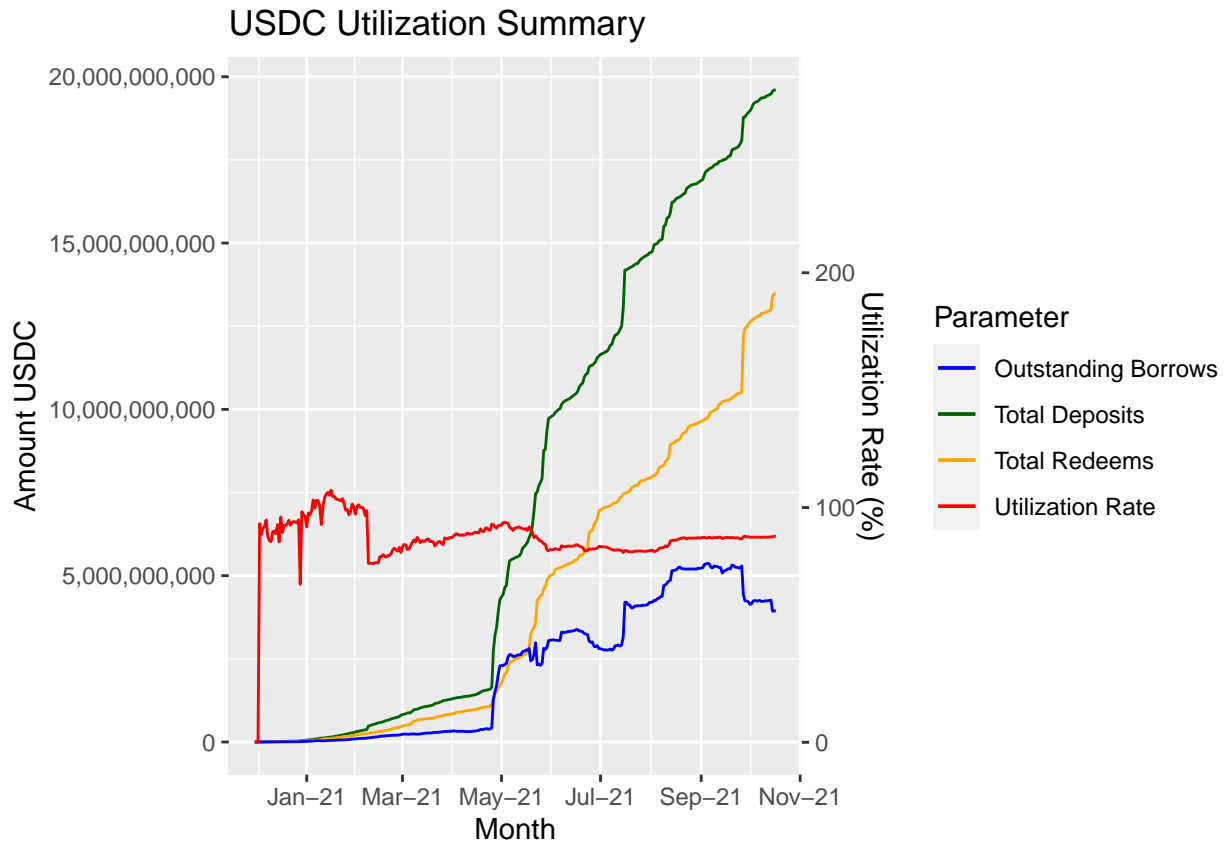
We can see that USDC has a high utilization rate, even approaching 100% at times. This chart let's us see how variations in certain transaction types can affect the coins interest rate. Now we want to make a dataframe storing all of this data for every coin so that we could see this plot for any desired coin.

| date	WETH_dep	WETH_red	WETH_rep	WETH_bor
2020-11-30	0.0002	0.00000	0.0000000	0.000000
2020-12-01	0.0020	0.00200	0.0000000	0.000000
2020-12-02	0.8200	0.26552	0.1050003	0.115000
2020-12-03	374795.1693	372570.72124	0.0050001	5.657405
2020-12-04	435008.2530	432493.32856	3.1354144	58.425536
2020-12-05	555710.3020	552695.65909	10.6275895	198.566212

This dataframe is much larger than what is shown, as it includes 10 columns for every coin in the dataset. However, it stores all of all results and doesn't make us rerun code to chart for different coins when building our app. Now, I divide the coins into their respective groups (stable, non-stable, Amm) and plot all of them to see if they have differences.

Stable Coins Utilization Rate
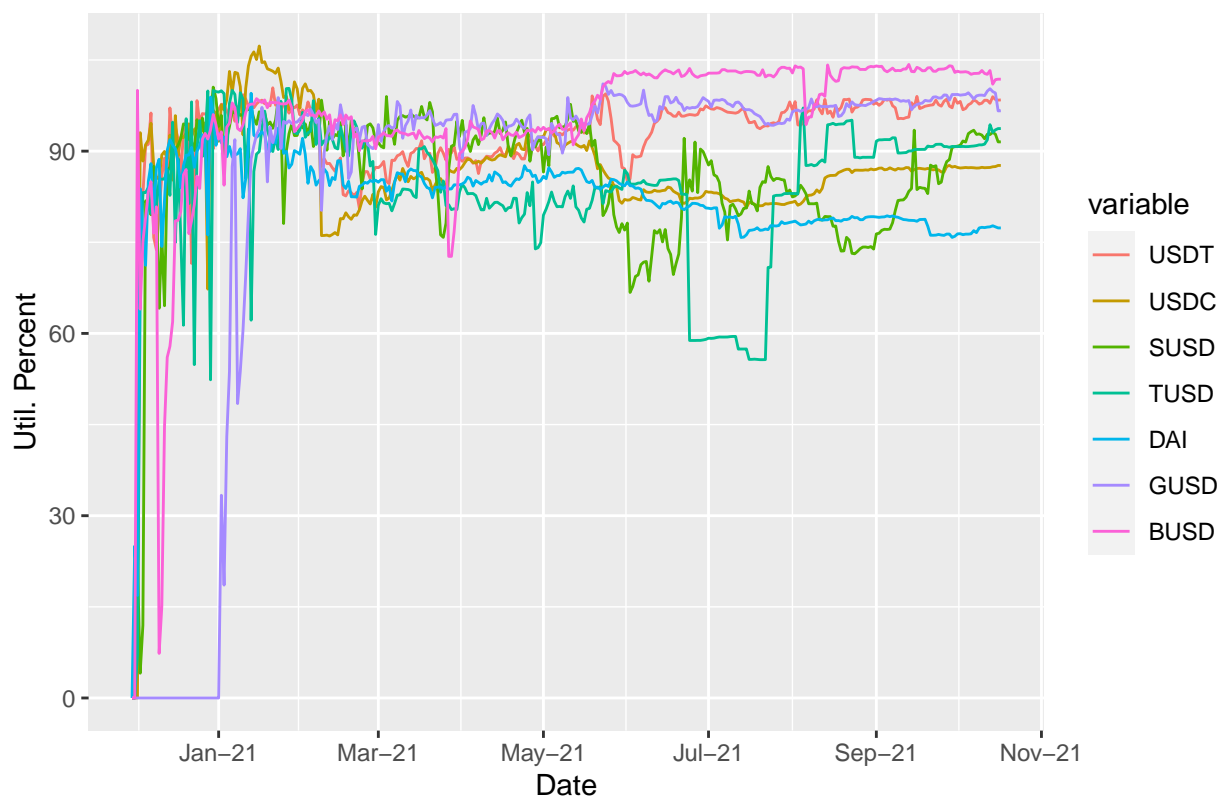
We can see that all of the stable coins have high utilization rates throughout time. However, other than the one brief decrease in TUSD, they also are all fairly bunched together. As we'll see, this is not always the case. Let's look at the non-stable coins.

Non−stable Coins Utilization Rate
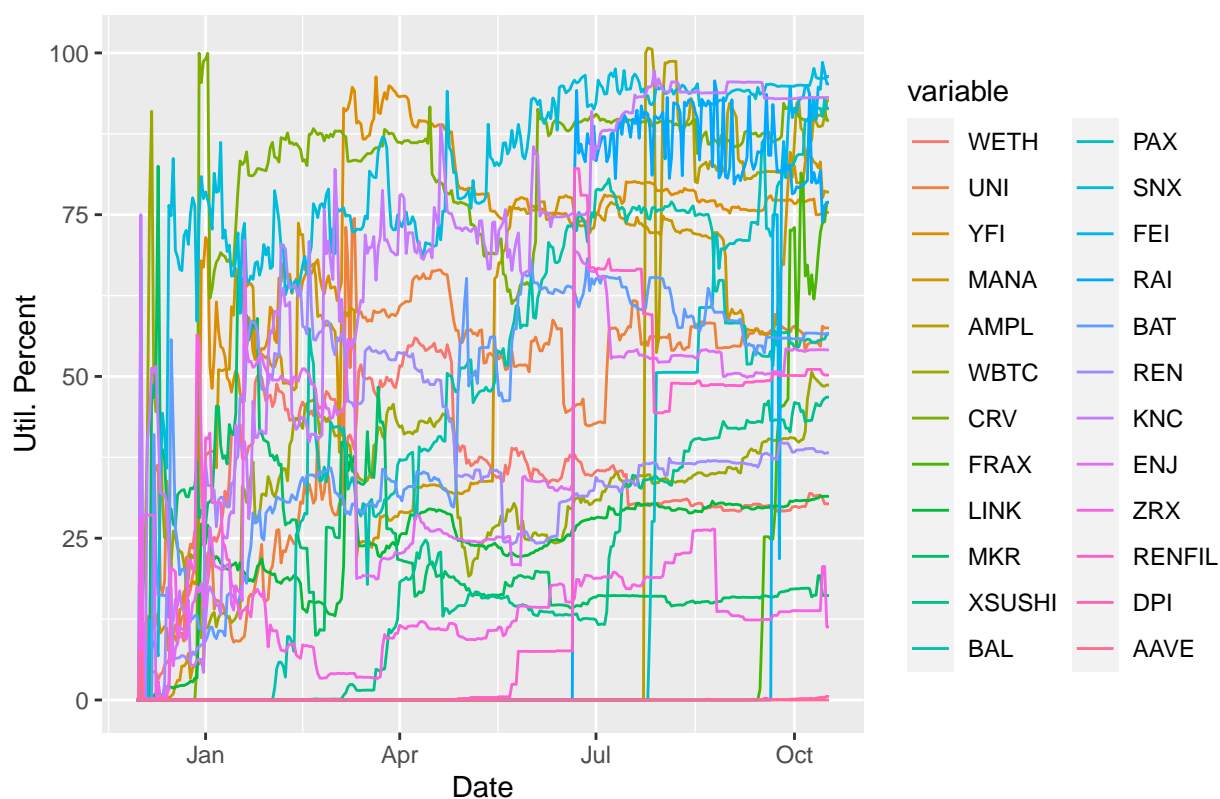
Woah, what a mess. However, this does show the point that we are looking for. Non-stable coins have wide-ranging utilization rates that are much more subject to fluctuations. As mentioned, the deposit rate for AMPL this week was >87,000% because utilization got too high. However, some of these bigger coins will have a more stable utilization rate throughout time. Finally, we look at Amm coins.

Amm Coins Utilization Rate
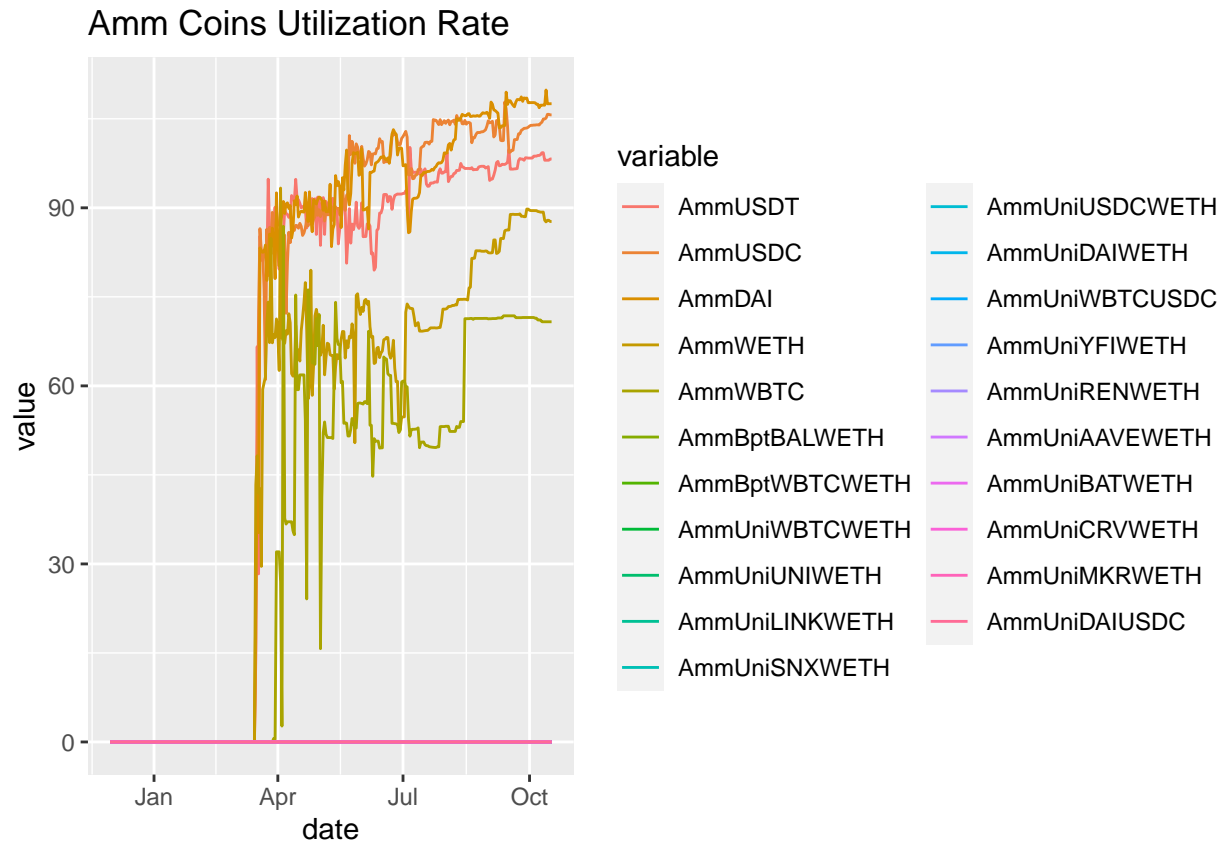
These are different because several of them never get borrowed. This causes a big distinction betwee these coins. We can see that some behave more like the stable coins, although there appears to be more variation in their rates.

```
stable_only <- stable_utils[,stable_coins]
stable_utils$average <- rowMeans(stable_only)
stable_stdev <- rowSds(as.matrix(stable_utils[ , -which(names(stable_utils) %in% c("date"))]))
unstable_only <- unstable_utils[,unstable_coins]
unstable_utils$average <- rowMeans(unstable_only)
unstable_stdev <- rowSds(as.matrix(unstable_utils[ , -which(names(unstable_utils) %in% c("date"))]))

Amm_only <- Amm_utils[,Amm_coins]
Amm_utils$average <- rowMeans(Amm_only)
Amm_stdev <- rowSds(as.matrix(Amm_utils[ , -which(names(Amm_utils) %in% c("date"))]))

averages <- data.frame(stable_utils$average, unstable_utils$average, Amm_utils$average, stable_utils$da

ggplot(averages, aes(x = stable_utils.date))+
  geom_line(aes(y = stable_utils.average, color = "Stablecoins"))+
  geom_errorbar(aes(ymin=stable_utils$average-stable_stdev, ymax=stable_utils$average+stable_stdev), wi
              position=position_dodge(0.05), alpha = 0.2, color = "blue")+
  geom_line(aes(y = unstable_utils.average, color = "Non-stable Coins"))+
  geom_errorbar(aes(ymin=unstable_utils$average-unstable_stdev, ymax=unstable_utils$average+unstable_st
              position=position_dodge(0.05), alpha = 0.2, color = "darkgreen")+
  geom_line(aes(y = Amm_utils.average, color = "Amm Coins"))+
  geom_errorbar(aes(ymin=Amm_utils$average-Amm_stdev, ymax=Amm_utils$average+Amm_stdev), width=.2,
              position=position_dodge(0.05), alpha = 0.2, color = "red")+
  labs(x = "Date", y = "Utilization Rate", color = "Coin Type")+
```

Average Utilization by Coin Type
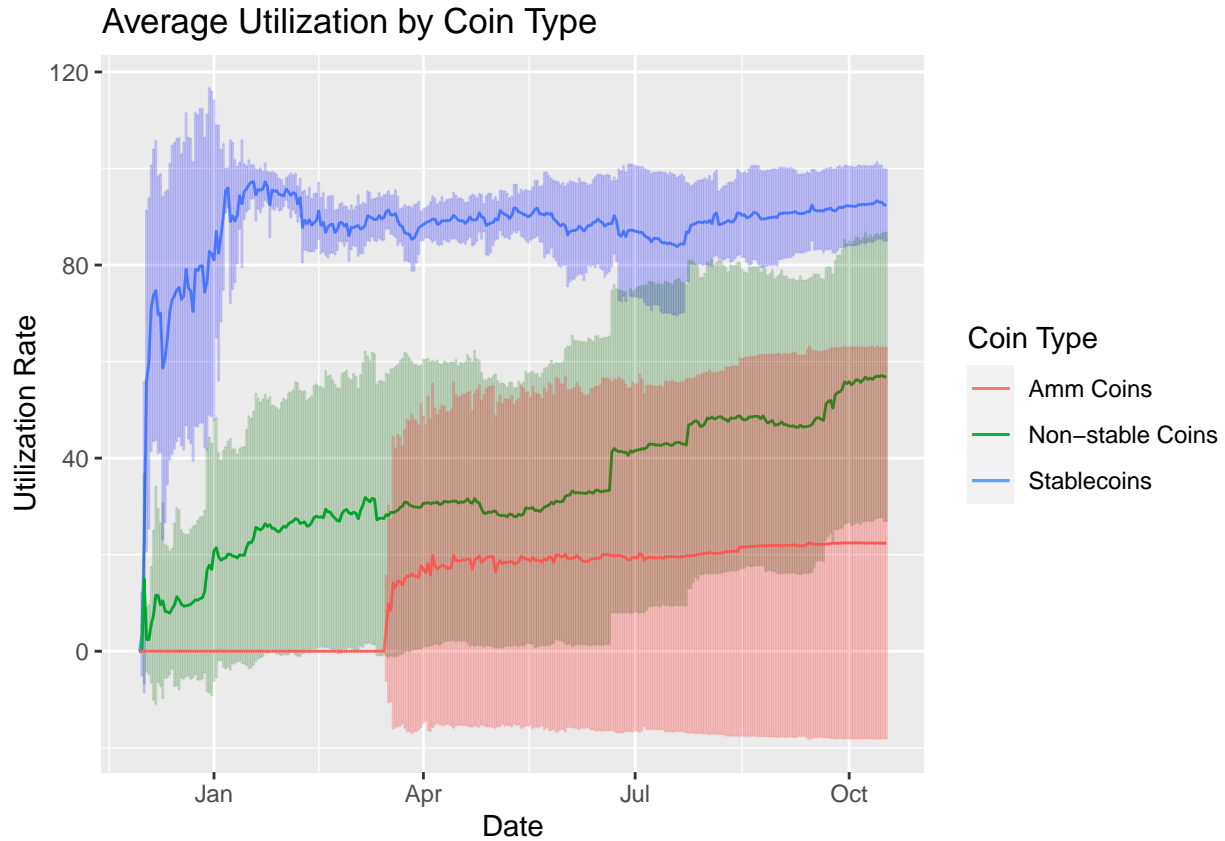
Here we put it all together to see how it comes together. Sable coins also have the most stable utilization rates over time. Non-stable coins are getting borrowed with more frequency, although they have a wide standard deviation. Finally, we do not really need to analyze Amm coins since they seem to act independently of the other two groups. Now let's look at what coins perform the most similarly.

coin_1	coin_2	correlation
AmmDAI	AmmUSDC	0.9953700
AmmDAI	AmmUSDT	0.9944786
AmmUSDC	AmmUSDT	0.9926419
AmmWETH	AmmDAI	0.9833476
AmmWETH	AmmUSDT	0.9804248
AmmWETH	AmmUSDC	0.9777085
RENFIL	RAI	0.9715199
FEI	FRAX	0.9655371
PAX	AMPL	0.9320101
BAT	BAL	0.9159820
AmmWBTC	AmmDAI	0.9092752
AmmWBTC	AmmWETH	0.9007144
AmmWBTC	AmmUSDC	0.9000754
AmmWBTC	AmmUSDT	0.8913831
DPI	FRAX	0.8687376
LINK	YFI	0.8646332
BAL	AmmUSDC	0.8601220
USDC	DAI	0.8575616
CRV	GUSD	0.8552684

coin_1	coin_2	correlation
KNC	AmmWETH	0.8523437

This is helpful in showing how some coins have very similar utilization patterns. Could these then be used to predict future utilization or interest rates? Maybe, but either way I though the number of non-stable pairings was interesting since that group had a lot of variation.

Discussion

We can see big differences between how different types are used. Stable coins are borrowed much more frequently than other types, so they have less liquidity in the lending pool. However, non-stable coins are being borrowed more frequently than they used to. Also, stable coins all tend to act in the same manner, while there are large variations in the other groups.

Summary and Recommendations

I'm making four main conclusions based off of this work:

1- We are able to cluster users into groups that allow us the see the behavior of users who are often liquidated.

2- Although the number of active users in AAVE has not increased, people are using larger amounts of money in their transactions.

3- This market is heavily subject to shock events that will briefly disrupt transaction patterns.

4- Stable coins are borrowed at much higher rates than other coins, and their is a wide variety of uses for these other non-stable coins.

Going forward, I believe that predicting utilization/interest rates would be very helpful in terms of making money. Along with Jason's interests rates, we will be adding utilization to the coin page in our app. Finally, we could try to use clustering on weekly data (which I have) to try to predict which users are at higher risk of being liquidated in the near future. This would be helpful with the addition of Chris's new data.