

# DAR F21 Project Status 4 Observing Trends through Users' Average Transactions IDEA-Blockchain (DeFi)

Duke Kwon

10/14/21

## Contents

Weekly Work Summary . . . . .	1
Personal Contribution . . . . .	2
Discussion of Primary Findings . . . . .	2
Preliminary Data Loading . . . . .	2
Plotting Weekly Transactions of All Users . . . . .	3
Plotting Weekly Transactions of Users by Cluster . . . . .	4
Plotting Weekly Transactions by Type, Stacked by Cluster . . . . .	6
Side by Side Comparisons of Clusters & Averaged Transaction Types Weekly . . . . .	9
Further Analysis: . . . . .	12

## Weekly Work Summary

- RCS ID: kwond2
- Project Name: IDEA-Blockchain (DeFi)
- General Summary:
  - Further analysis and visualization was done on the behavior of users by their transactions by cluster. Specifically, we look at how user transactions vary through time monthly and weekly. We were able to show some distinct differences between their overall transaction types and amount for the 3 largest clusters.
- Github Commits:
- Branch: dar-kwond2
- Directory: <https://github.rpi.edu/DataINCITE/IDEA-Blockchain/tree/master/DefiResearch/StudentNotebooks/Assignment05>
- kwond2\_assignment5.Rmd
- kwond2\_assignment5.pdf
- kwond2\_assignment5.html
- References:
- N/A
- Shared Code Base:
- Clusters generated from kwond2\_assignment4.Rmd are used.

## Personal Contribution

- All code and explanation in this notebook was done by me.

## Discussion of Primary Findings

- What did you want to know?
  - After using heatmaps and density estimates of the features to compare differences by cluster, a lot of information was not utilized to its fullest extent. Mainly, types, amounts, and frequencies of transaction for each of the users by clusters through time might provide a better understanding of the groupings.
- How did you go about finding it?
  - The visualizations used were stacked barplots through time - averaged weekly transactions by transaction type and for the largest 3 clusters.
- What did you find?
  - By comparing side by side, differences in overall behavior/style of transactions was found, though the differences still need to be better interpreted.

## Preliminary Data Loading

```
library(ggplot2)
library(dplyr)
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

```
library(lubridate)
library(data.table)
library(ggthemes)
```

We start with loading in the data generated from the previous assignments - cluster labels and the original transaction data.

```
#write_csv(df.4, file='unique_users_avg_wClusters_assign4.csv') ## users_clustered
df.4 <- read_csv("unique_users_avg_wClusters_assign4.csv")
#df.4 <- df.4[, -which(names(df.4) %in% c("X"))] ## get rid of extra row indices column
df.users <- read_csv("df_users.csv") ## read users parsed csv
df.users <- df.users[, -which(names(df.users) %in% c("X"))]
df_tr <- read_csv('transactions_2.csv') ## original transactions data (not unique users from Jan to Aug. )
N <- max(df.4$hdb_clusters)
dict <- table(df.4$hdb_clusters) #
df.dict <- as.data.frame(dict) # Unused except to sort by top clusters in the next code block
```

Since we plan on doing analysis through time, we use the lubridate package to make the timestamps interpretable.

```
df_tr$ymd <- as_datetime(df_tr$timestamp) # fix times for the transactions

setDT(df_tr)[, ymd_new := format(as.Date(ymd), '%Y-%m-%V')] ## '%Y-%m' for just month-year
df_tr$ymd_new_d <- as.Date(df_tr$ymd_new, "%Y%m%d")
```

Now, we group the users and cluster specific transactions.

## Plotting Weekly Transactions of All Users

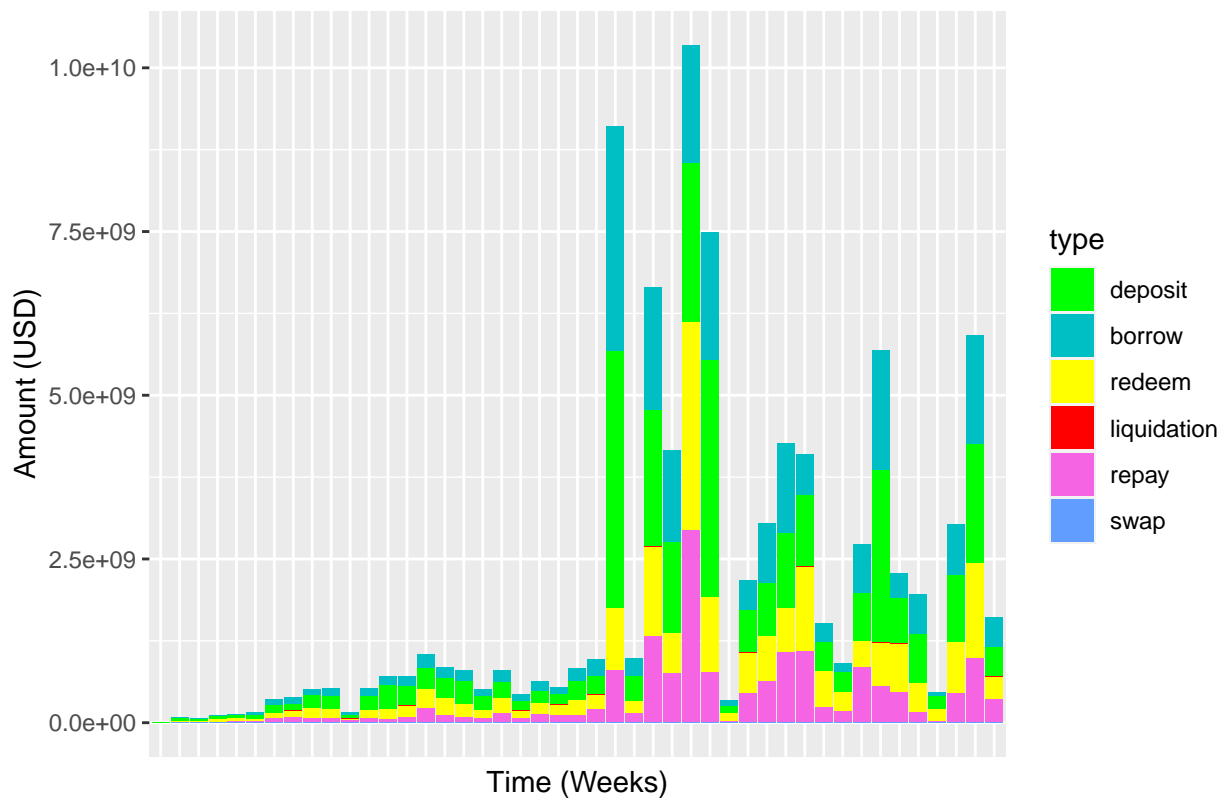
We start by taking a look at the entire user transactions data, to have a baseline result to compare the plots by cluster to.

```
# generate N list of custom colors
gg_color_hue <- function(n) {
  hues = seq(15, 375, length = n + 1)
  hcl(h = hues, l = 65, c = 100)[1:n]
}

# 6-list of ggplot colors explicitly specified
pgg <- gg_color_hue(6)

# Start plotting weekly by time. Start with all Users first
df_tr %>%
  count(ymd_new, type, wt = amount, name = "amount") %>%
  ggplot() +
  geom_bar(aes(x=ymd_new, y= amount, fill = type),
    stat='identity') + ggtitle("ALL Users Transactions Weekly From Jan 2021 to Aug 2021") + labs
  theme(
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank()
  ) +
  scale_fill_manual("type", values = c("deposit"="green", "borrow" = pgg[4], "redeem" = "yellow"
```

ALL Users Transactions Weekly From Jan 2021 to Aug 2021



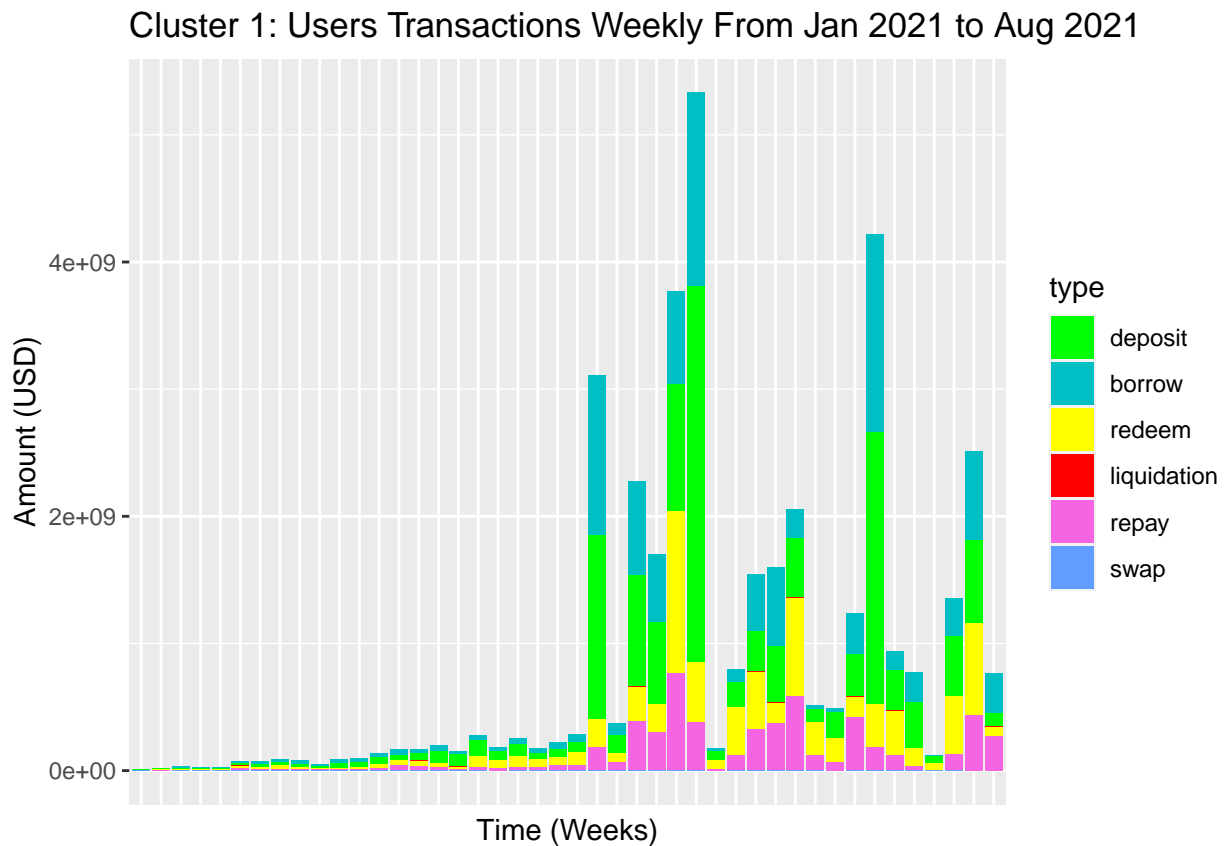
```
# The only important command within geom_bar() is scale_fill_manual() - it's setting custom colors for
# are filled by the list generated by gg_color_hue, which contains a palette of colors
```

## Plotting Weekly Transactions of Users by Cluster

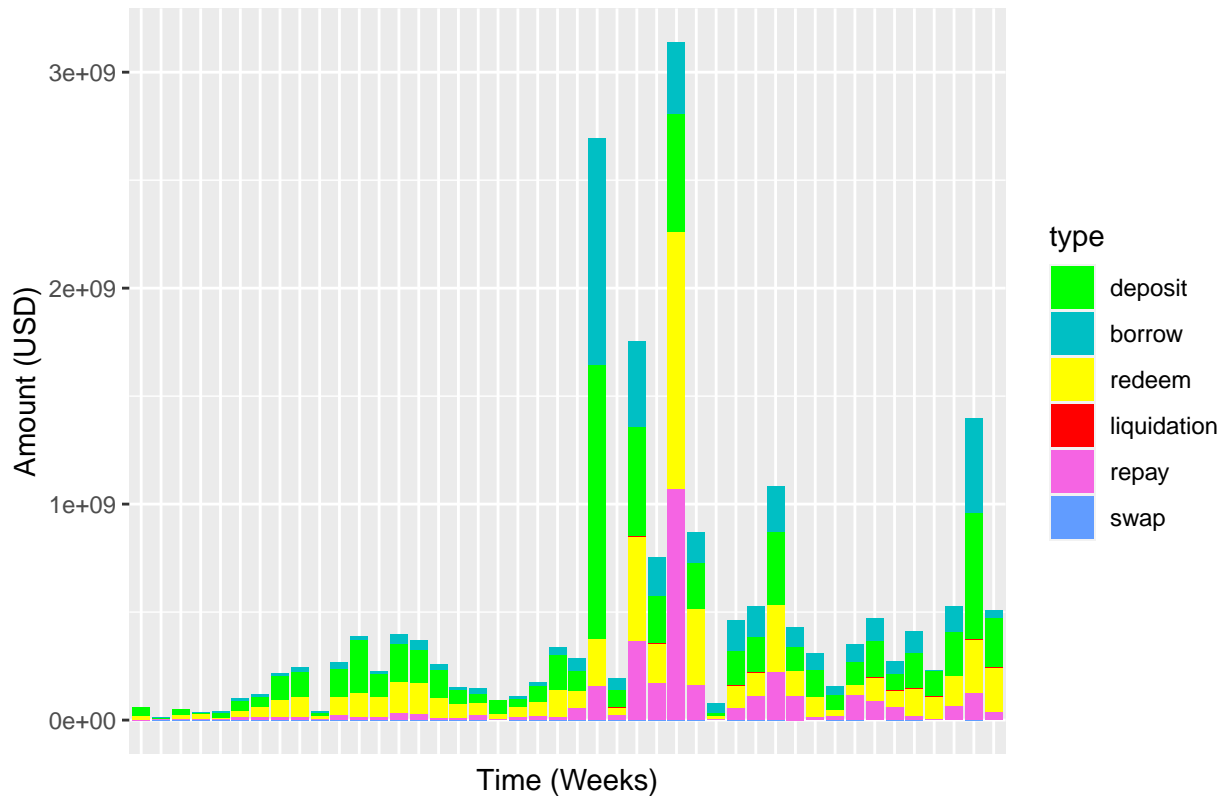
Now, we plot the amount and type of transactions users make from specific clusters. A note has to be made that the data is not averaged by the number of unique users in a specific cluster. However differences are still noticeable comparing the proportion of transactions. The largest cluster, cluster 1, tends to have a large amount of deposits (green), and a moderate number of swaps. Also, it seems like cluster 2 has a generally larger proportion of redeems, and so forth.

Since it's still relatively hard to compare the results via a stacked plot like this one, we then plot the averaged weekly amounts of a specific transaction type next.

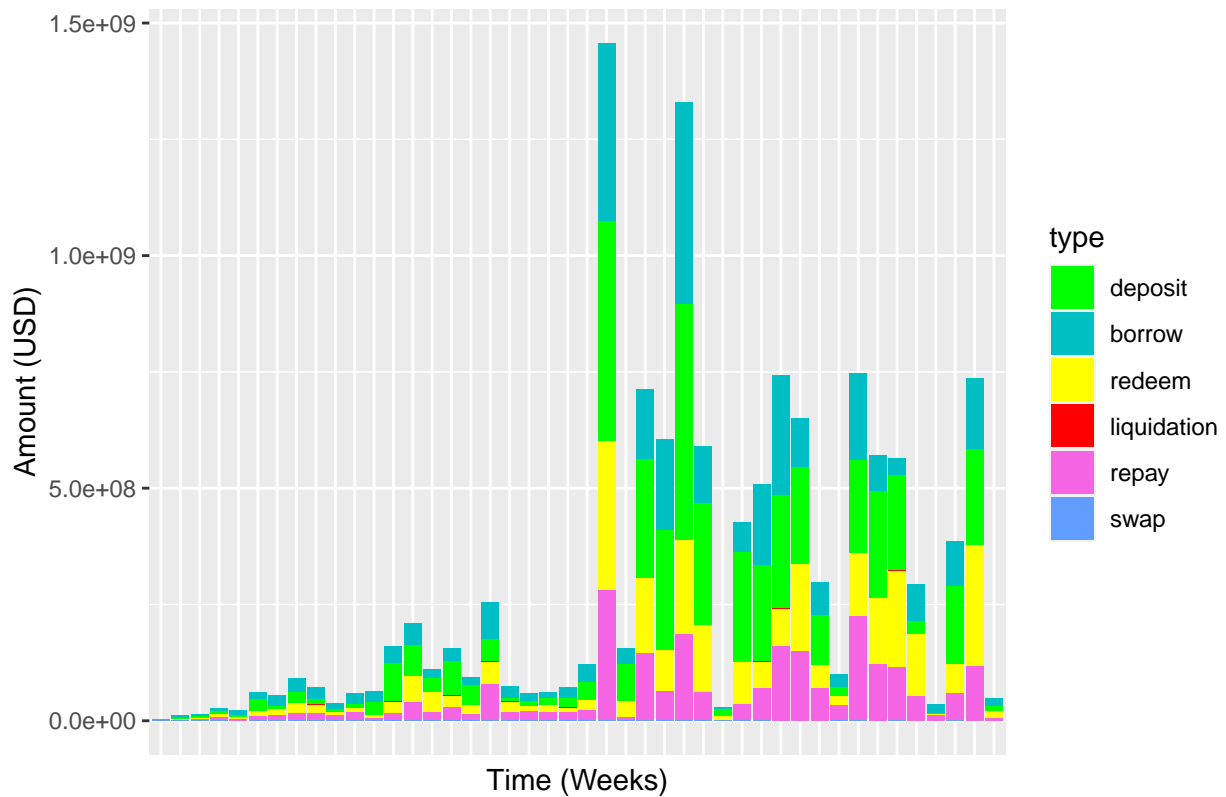
```
clplt = list()
for (m in 1:3) {
  cluster.plt_loop_temp <- ggplot(count(cluster_transactions[[m]] ,ymd_new, type, wt = amount, name = "Cluster ",m," : Users Transactions Weekly From Jan 2021 to Aug 2021")) +
  geom_bar(aes(x=ymd_new,y= amount, fill = type),
    stat='identity') + ggtitle(paste0("Cluster ",m," : Users Transactions Weekly From Jan 2021 to Aug 2021")) +
  theme(
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
    scale_fill_manual("type", values = c("deposit"="green","borrow" = "#00BFC4", "redeem" = "yellow", "liquidation" = "red", "repay" = "magenta", "swap" = "blue"))
  clplt[[m]] <- cluster.plt_loop_temp
  print(clplt[[m]])
}
```



Cluster 2: Users Transactions Weekly From Jan 2021 to Aug 2021



Cluster 3: Users Transactions Weekly From Jan 2021 to Aug 2021



```

deposit_only <- list()
liquidation_only <- list()
repay_only_list <- list()
redeem_only_list <- list()
borrow_o <- list()
for (i in 1:3) {
  dep_only_rows <- cluster_transactions[[i]][cluster_transactions[[i]]$type == "deposit"]
  liq_only_rows <- cluster_transactions[[i]][cluster_transactions[[i]]$type == "liquidation"]
  rep_r <- cluster_transactions[[i]][cluster_transactions[[i]]$type == "repay"]
  red_r <- cluster_transactions[[i]][cluster_transactions[[i]]$type == "redeem"]
  bor_r <- cluster_transactions[[i]][cluster_transactions[[i]]$type == "borrow"]
  dep_only_rows$cluster <- i
  liq_only_rows$cluster <- i
  rep_r$cluster <- i
  red_r$cluster <- i
  bor_r$cluster <- i
  deposit_only[[i]] <- dep_only_rows
  liquidation_only[[i]] <- liq_only_rows
  repay_only_list[[i]] <- rep_r
  redeem_only_list[[i]] <- red_r
  borrow_o[[i]] <- bor_r
  #print(dim(deposit_only[[i]]))
  deposit_only[[i]]$amount <- deposit_only[[i]]$amount/dim(deposit_only[[i]])[1]
  repay_only_list[[i]]$amount <- repay_only_list[[i]]$amount/dim(repay_only_list[[i]])[1]
  redeem_only_list[[i]]$amount <- redeem_only_list[[i]]$amount/dim(redeem_only_list[[i]])[1]
  borrow_o[[i]]$amount <- borrow_o[[i]]$amount/dim(borrow_o[[i]])[1]
}
dep.c<- do.call("rbind", list(deposit_only[[1]], deposit_only[[2]], deposit_only[[3]]))
red.c <-do.call("rbind", list(redeem_only_list[[1]], redeem_only_list[[2]], redeem_only_list[[3]]))
liq.c <- do.call("rbind", list(liquidation_only[[1]], liquidation_only[[2]], liquidation_only[[3]]))
rep.c <- do.call("rbind", list(repay_only_list[[1]], repay_only_list[[2]], repay_only_list[[3]]))
bor.c <- do.call("rbind", list(borrow_o[[1]], borrow_o[[2]], borrow_o[[3]]))

# Plot by transaction type weekly
loop_labels <- c("deposit", "redeem", "repay", "borrow")
cap_loop_labels <- c("Deposits", "Redeems", "Repays", "Borrows")
facet_data_list <- list(dep.c, red.c, rep.c, bor.c)

```

## Plotting Weekly Transactions by Type, Stacked by Cluster

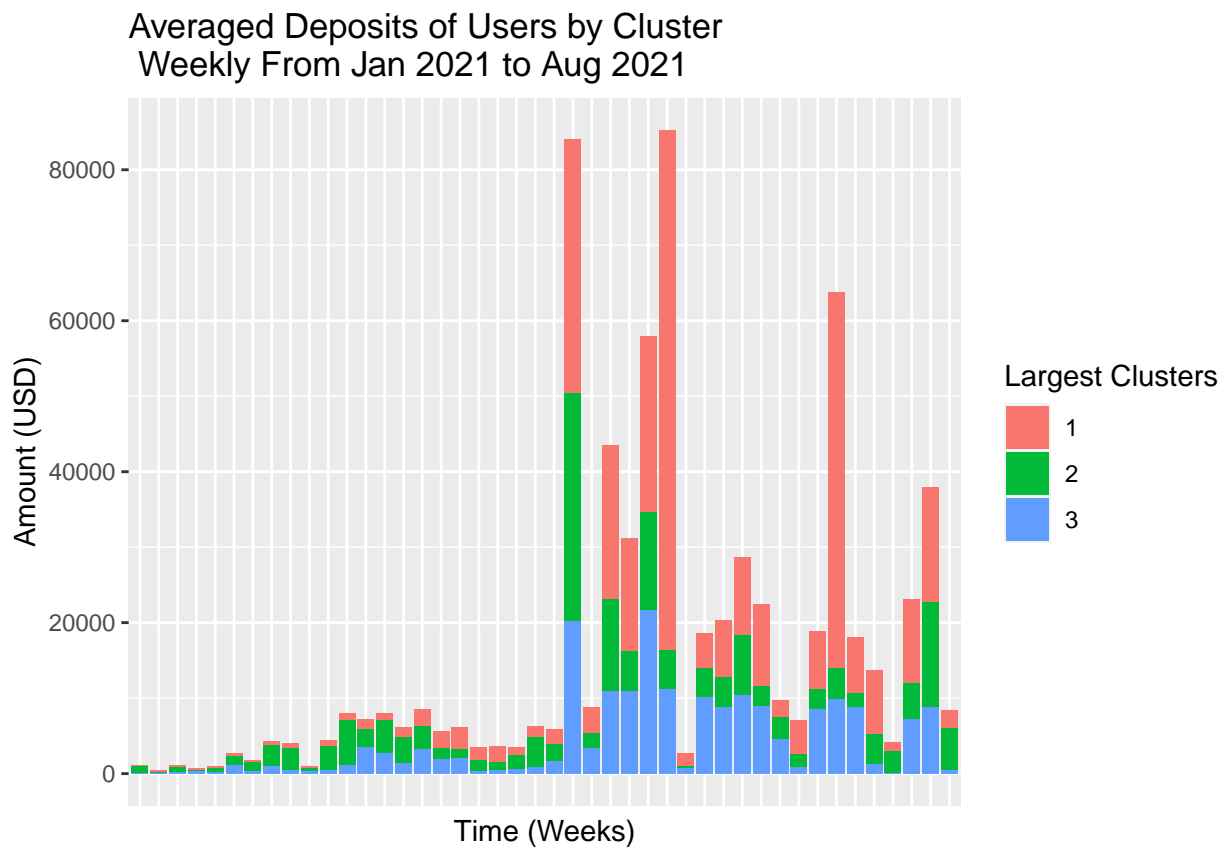
Here, we plot the averaged (by number of unique users per cluster) of amount USD deposits, redeems, repays, and borrows weekly. The stacked values indicate the amount for that specific cluster.

```

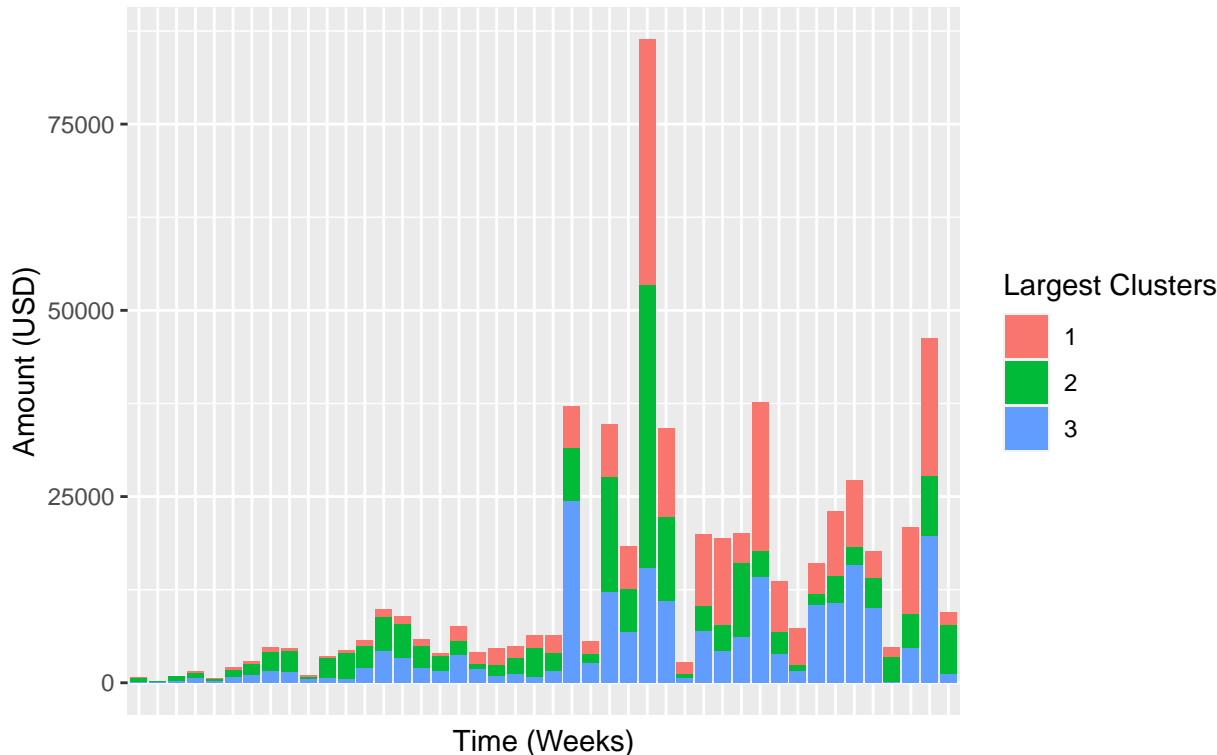
fct_plot_list <- list()
for (m in 1:4) {
  pltobj <- ggplot(count(facet_data_list[[m]], ymd_new, cluster, wt = amount, name = "amount")) +
    geom_bar(aes(x=ymd_new, y= amount, fill = as.factor(cluster)),
             stat='identity') + ggtitle(paste0("Averaged ", cap_loop_labels[m], " of Users by Cluster\n W
    theme(
      axis.text.x=element_blank(),
      axis.ticks.x=element_blank()) + labs(fill="Largest Clusters")
  print(pltobj)
  fct.plt <- pltobj + facet_grid(cols = vars(cluster))
  fct_plot_list[[m]] <-fct.plt

```

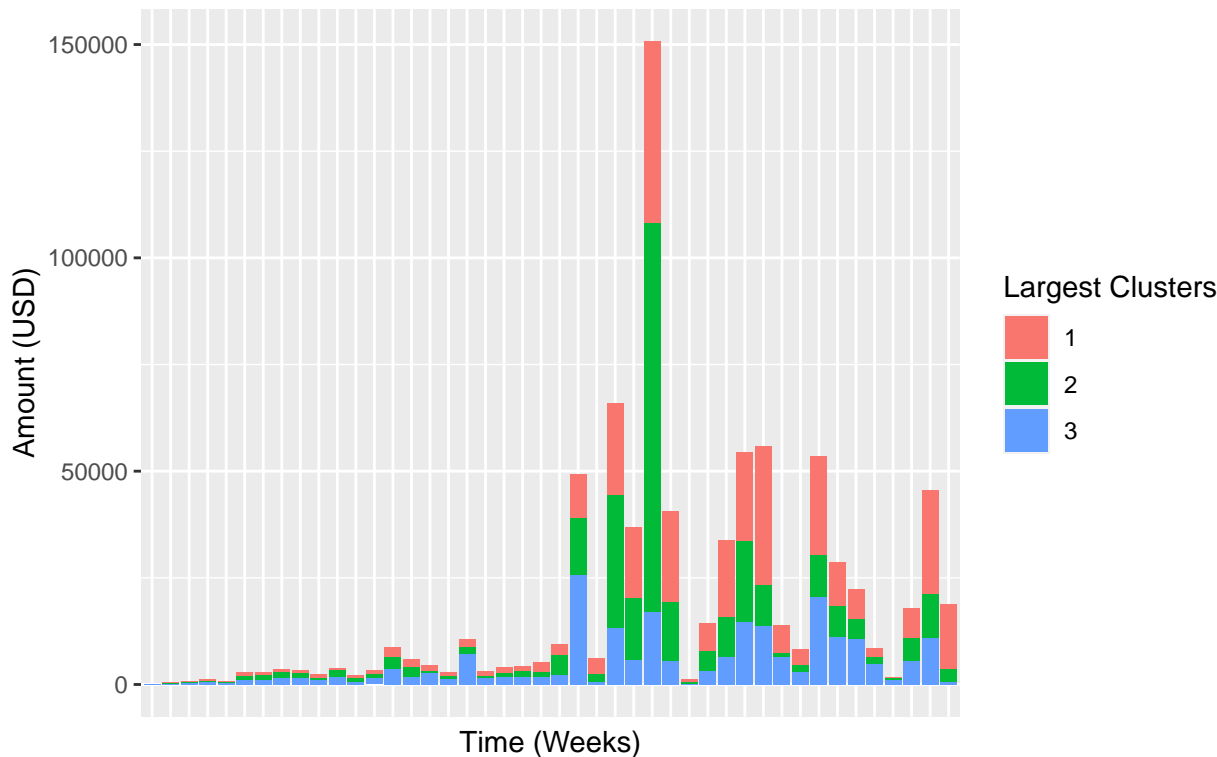
}



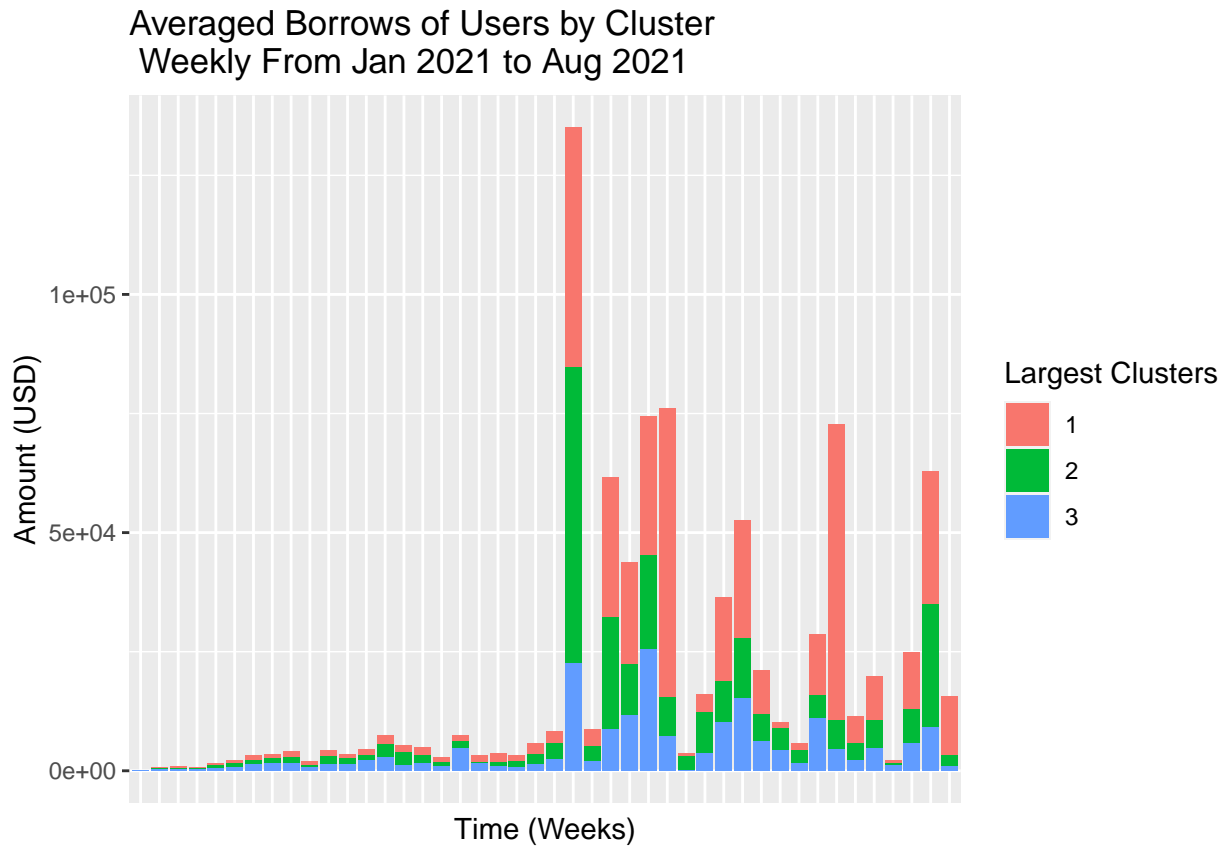
Averaged Redeems of Users by Cluster  
Weekly From Jan 2021 to Aug 2021



Averaged Repays of Users by Cluster  
Weekly From Jan 2021 to Aug 2021





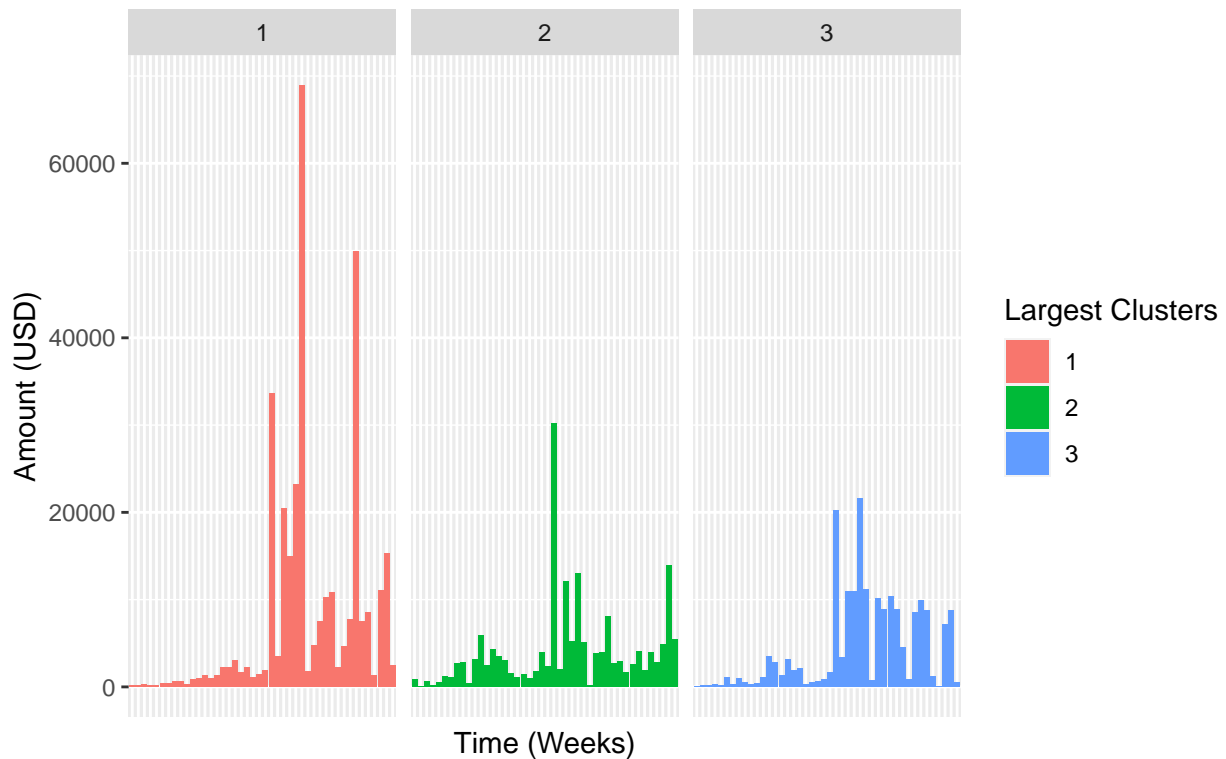


### Side by Side Comparisons of Clusters & Averaged Transaction Types Weekly

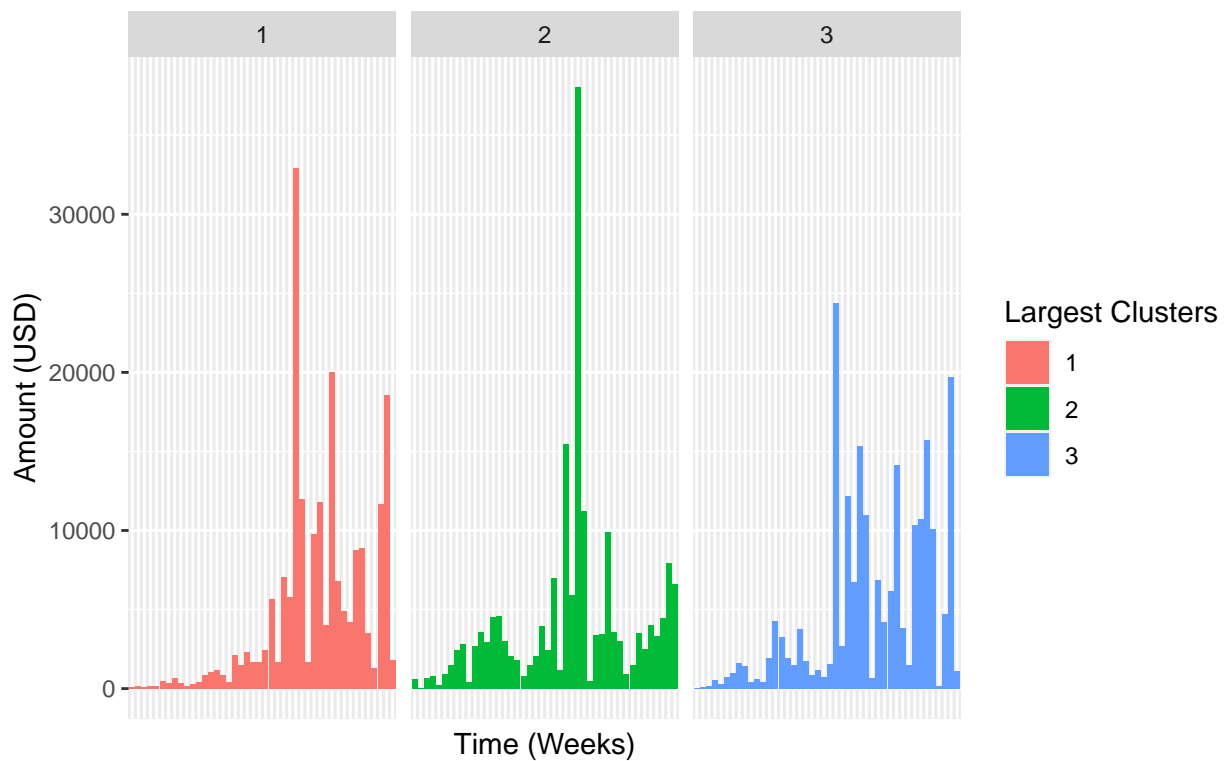
For even further comparisons, we use `facet_grid()`, which helps immensely.

```
## now, facet plot
for (m in 1:4) {
  print(fct_plot_list[[m]])
}
```

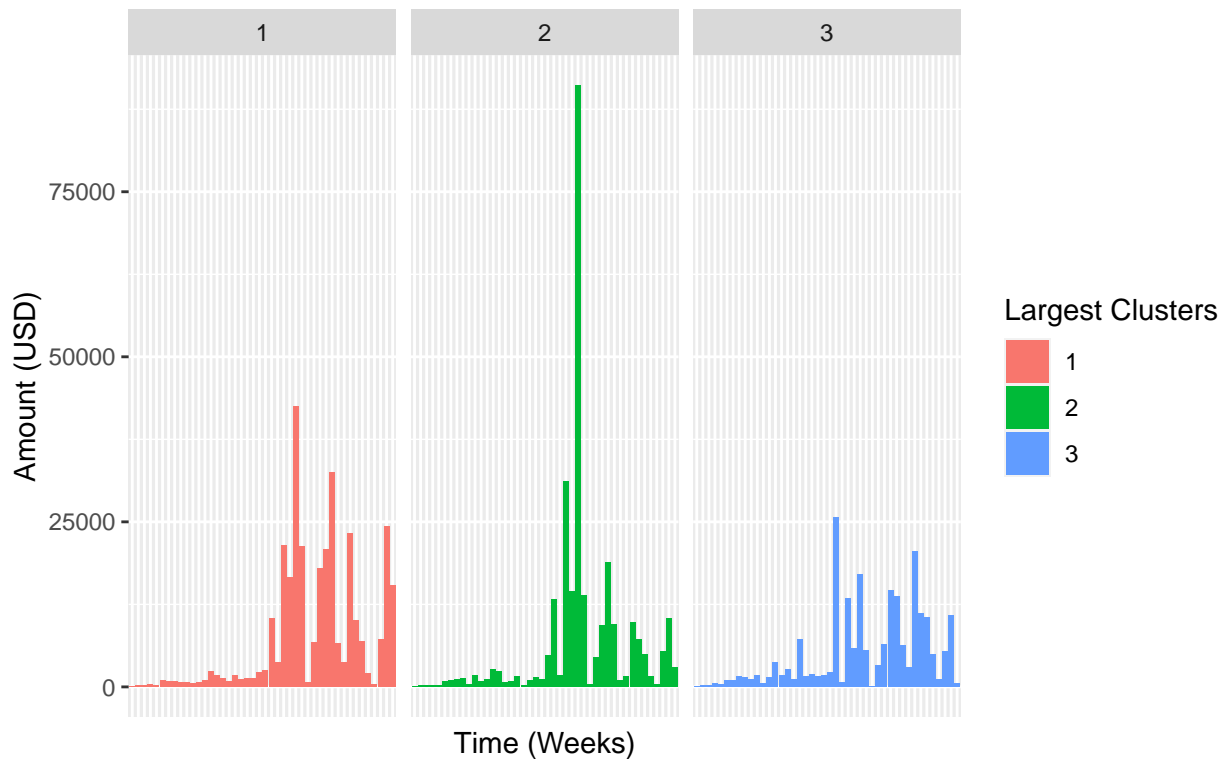
Averaged Deposits of Users by Cluster  
Weekly From Jan 2021 to Aug 2021



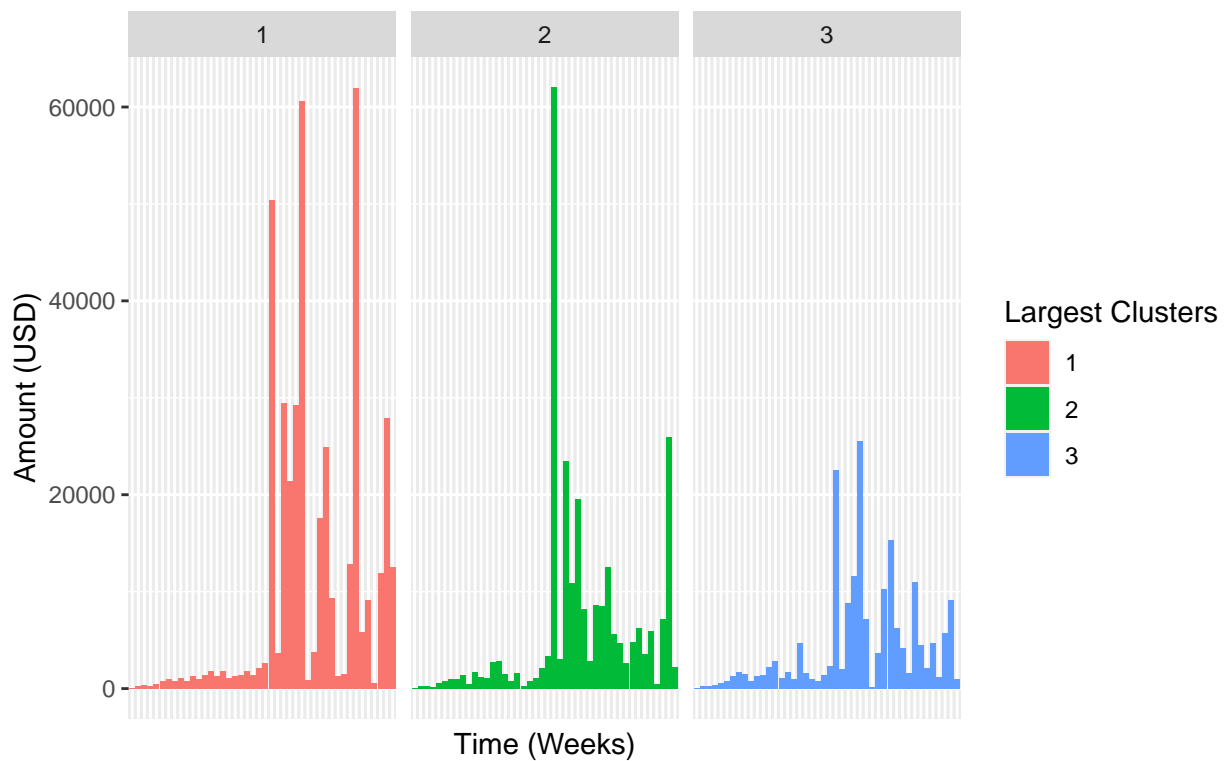
Averaged Redeems of Users by Cluster  
Weekly From Jan 2021 to Aug 2021



Averaged Repays of Users by Cluster  
Weekly From Jan 2021 to Aug 2021



Averaged Borrows of Users by Cluster  
Weekly From Jan 2021 to Aug 2021



As we observed before, cluster 1 seems to have transactions with a significant amount (USD) in their deposits, compared to cluster 2 and 3. Redeems across the clusters seem to be somewhat similar. One particular oddity across all 3 plots is the large spike about ~20-30 weeks into our data set. For repays, cluster 2 seems to have the reoccurring spike, yet no spike in cluster 1 and 3. And finally, we see that clusters 1 and 2 have noticeably higher amounts of USD for averaged weekly borrows.

### **Further Analysis:**

Although we were able to differentiate the clusters to an extent, reasoning or deducing what kinds of groups these are is still quite difficult. Also, more analysis can be done comparing not only the amount of USD weekly for those transactions, but specifically the amount of transactions of users through time as well.

Additionally, exploring the odd spikes/possibly shocks in the data, and aligning them with possible cryptocurrency/DeFi world news within a similar timeframe (i.e. China Bitcoin Ban), would also be interesting to look into.