

HELOC Data Visualization

Executive Summary

The scope of this Data Visualization project is to understand whether we are lending the right amount of money to the right people. It is not sensible to check the rationality of loan on a individual basis, considering the manual work it will cost. The logic of my analysis is to start with analyzing outliers and anomalies in the dataset, to dive deeper on why they are so deviated from other data points. After analysis outliers, we will analyze the rest majority of accounts on a group basis, which means we will use clustering method to segment the customers into several different categories. And then by visualizing each segment, we can understand how we are lending money to each customer now.

Exploratory Data Analysis

Data Processing

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.2
```

```
heloc <- read.csv("Data Visualization.csv")
str(heloc)
```

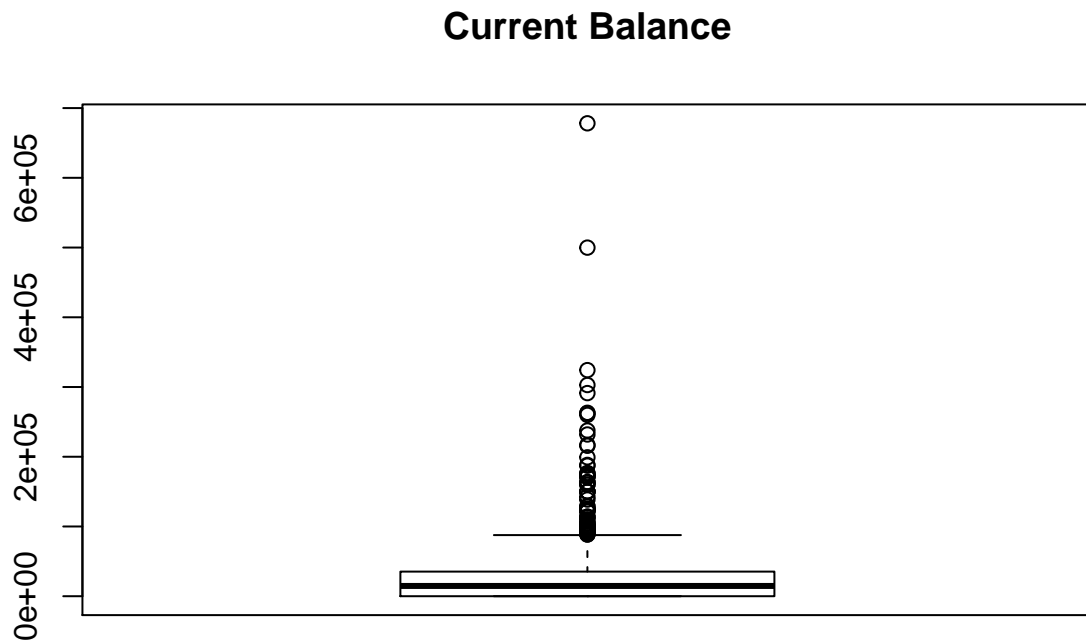
```
## 'data.frame':    1045 obs. of  19 variables:
## $ Product.Type      : Factor w/ 1 level "HELOC": 1 1 1 1 1 1 1 1 1 1 ...
## $ Account.Number    : int  2 3 4 5 6 7 8 9 10 11 ...
## $ Current.Balance    : Factor w/ 762 levels "$0 ", "$0.46 ", ..., 34 281 94 592 410 162 1 1 736 1 ...
## $ Loan.Limit.Amount : Factor w/ 141 levels "$0 ", "$10,000.00 ", ..., 49 58 66 42 101 57 102 85 126 1 ...
## $ Int.Rate           : num  4.25 6.25 5 4 3 4 4 4 4 4 ...
## $ Date.Org.Opened    : Factor w/ 932 levels "1/1/13", "1/10/12", ..., 414 349 288 600 257 636 520 450 ...
## $ Date.Closed        : logi  NA NA NA NA NA NA ...
## $ Term               : int  12 28 24 120 120 120 120 120 120 120 ...
## $ CREDIT.SCORE        : int  726 745 759 785 0 672 768 717 800 805 ...
## $ ORIGINAL.L.V        : num  58 61 29 46 54.9 ...
## $ DEBT.TO.INCOME      : num  0 0 0 29.6 0 ...
## $ OCCUPANCY.CODE.Cd.1 : int  1 2 1 1 1 1 1 1 1 1 ...
## $ PROPERTY.TYPE       : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Score.Range         : Factor w/ 7 levels "1 >=740", "2 720-739", ..., 2 1 1 1 7 4 1 3 1 1 ...
## $ DTI.Range           : Factor w/ 7 levels "1 >50.01", "2 43.01-50.00", ..., 7 7 7 6 7 3 6 4 3 6 ...
## $ Orig.LTV.Range      : Factor w/ 6 levels "2 90% - 99.99% LTV", ..., 5 4 5 5 5 2 5 2 5 5 ...
## $ Yield.Range         : Factor w/ 5 levels "2 3.0% - 3.99%", ..., 3 5 4 2 1 2 2 2 2 2 ...
## $ Delinquency.Range   : Factor w/ 2 levels "1 Current", "2 1 to 29 DPD": 1 1 1 1 1 1 1 1 1 1 ...
## $ Days.Late           : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
# remove the dollar sign and comma
```

```
heloc$Current.Balance = as.numeric(gsub("[,$]", "", heloc$Current.Balance))
heloc$Loan.Limit.Amount = as.numeric(gsub("[,$]", "", heloc$Loan.Limit.Amount))
summary(heloc)
```

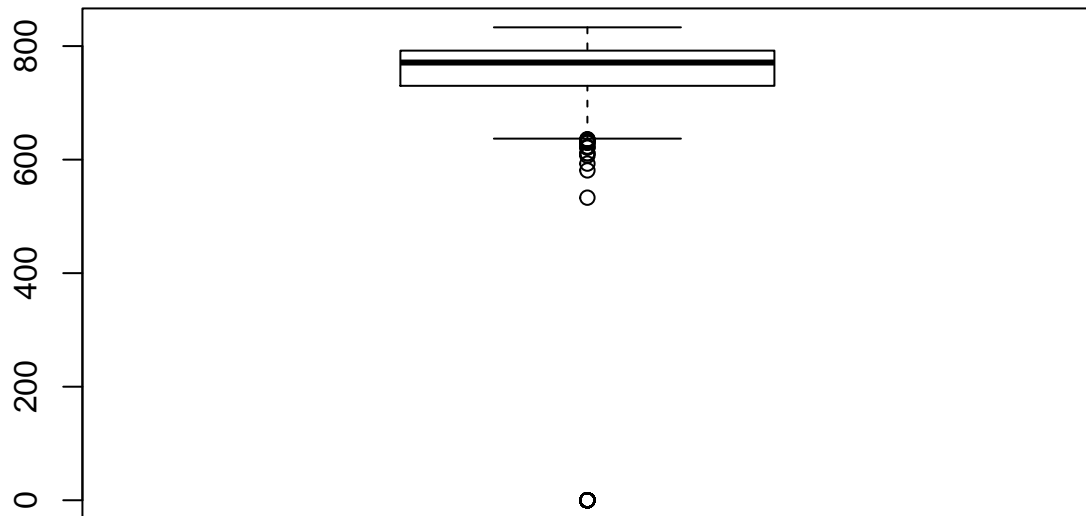
```
## Product.Type Account.Number Current.Balance Loan.Limit.Amount
## HELOC:1045 Min. : 2 Min. : 0 Min. : 0
## 1st Qu.: 263 1st Qu.: 0 1st Qu.: 30000
## Median : 524 Median : 14714 Median : 50000
## Mean : 524 Mean : 28451 Mean : 73915
## 3rd Qu.: 785 3rd Qu.: 35302 3rd Qu.: 100000
## Max. : 1047 Max. : 678161 Max. : 787000
##
## Int.Rate Date.Org.Opened Date.Closed Term
## Min. :3.000 10/21/11: 4 Mode:logical Min. : 6.0
## 1st Qu.:4.000 1/9/02 : 3 NA's:1045 1st Qu.:120.0
## Median :4.000 11/30/18: 3 Median :120.0
## Mean :3.958 3/6/18 : 3 Mean :119.3
## 3rd Qu.:4.000 6/19/20 : 3 3rd Qu.:120.0
## Max. :6.250 8/22/18 : 3 Max. :123.0
## (Other) :1026
## CREDIT.SCORE ORIGINAL.L.V DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1
## Min. : 0.0 Min. : 0.00 Min. : 0.00 Min. :1.000
## 1st Qu.:730.0 1st Qu.:31.00 1st Qu.: 18.55 1st Qu.:1.000
## Median :771.0 Median :58.00 Median : 27.00 Median :1.000
## Mean :723.6 Mean :54.52 Mean : 26.92 Mean :1.006
## 3rd Qu.:792.0 3rd Qu.:78.00 3rd Qu.: 36.00 3rd Qu.:1.000
## Max. :833.0 Max. :93.00 Max. :172.00 Max. :3.000
##
## PROPERTY.TYPE Score.Range DTI.Range
## Min. :0 1 >=740 :737 1 >50.01 : 25
## 1st Qu.:0 2 720-739 : 80 2 43.01-50.00 : 34
## Median :0 3 680-719 :101 3 40.01-43.00 :103
## Mean :0 4 640-679 : 56 4 35.01 - 40.00:120
## 3rd Qu.:0 5 600-639 : 18 5 30.01 - 35.00:144
## Max. :0 6 <= 600 : 3 6 <= 30 :559
## 7 No Score: 50 No DTI : 60
## Orig.LTV.Range Yield.Range
## 2 90% - 99.99% LTV : 20 2 3.0% - 3.99% : 50
## 3 80% - 89.99% LTV :220 3 >= 4.% - 4.25% :990
## 4 >= 70% - 79.99% LTV:142 4 >= 4.25% - 4.49%: 2
## 5 >= 60% - 69.99% LTV:112 7 >= 5% - 5.99% : 2
## 6 <60% LTV :549 8 >=6% LTV : 1
## No LTV Calculated : 2
##
## Delinquency.Range Days.Late
## 1 Current :1005 Min. :0.0000
## 2 1 to 29 DPD: 40 1st Qu.:0.0000
## Median :0.0000
## Mean :0.2632
## 3rd Qu.:0.0000
## Max. :7.0000
##
```

```
# The distributions of Current Balance, Credit Score and Debt to Income  
boxplot(heloc$Current.Balance, main = "Current Balance")
```



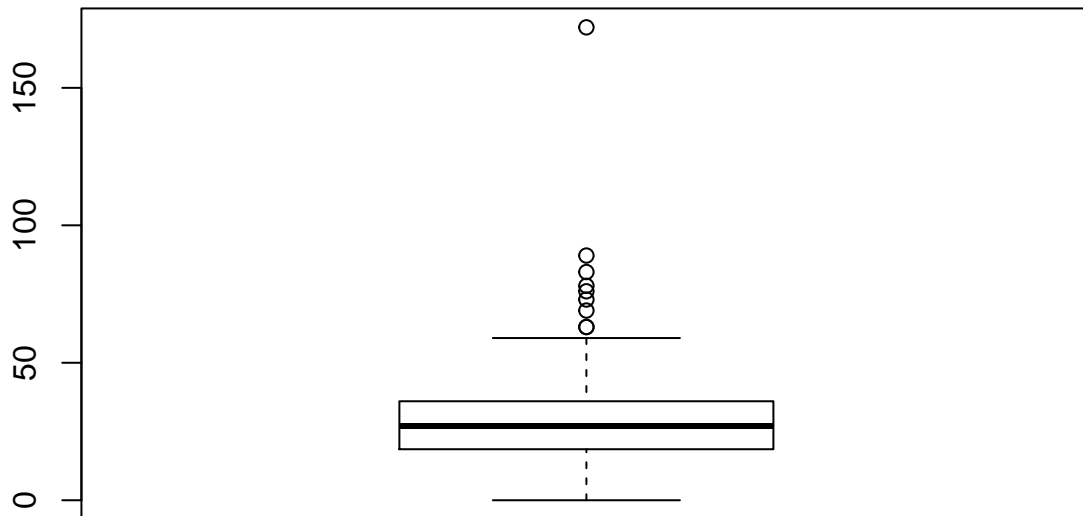
```
boxplot(heloc$CREDIT.SCORE, main = "Credit Score")
```

Credit Score



```
boxplot(heloc$DEBT.TO.INCOME, main = "Debt to Income")
```

Debt to Income



When we are lending customers loan, the Loan Limit Amount, Loan-to-Value (LTV) Ratio and Interest Rate are three important metrics. Intuitively, we should lend more loan with low Interest Rate to a customer who has a relatively high Credit Score, low Debit-to-Income (DTI) Ratio, reasonable Current Balance Amount and ideally better with a long history with us.

In our case, since we don't have very discrete formula of how we came up with the Loan Limit Amount as well as the LTV Ratio and Interest Rate, it's hard to say whether we give the most reasonable number to each of the customers. On the other hand, we still can identify data points which are outliers, regarding to Loan Limit Amount and LTV Ratio and Interest Rate, in the whole dataset.

Therefore, we will start with analyzing those data points which deviate a lot from the general distribution of the dataset. That's being said, the first thing we want to do is to figure out whether those outliers are showing for some specific and logical reasons.

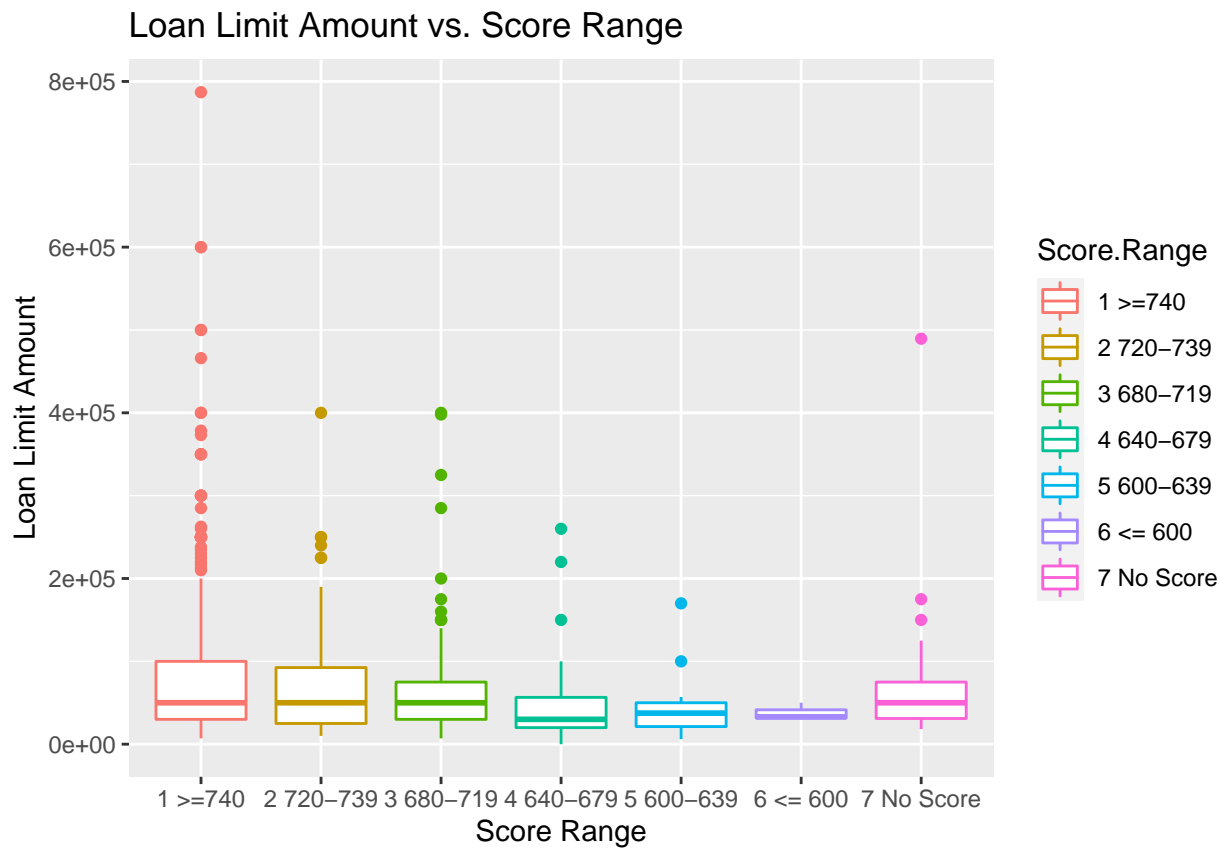
Here we use box plot and scatter plot to locate evident outliers. In the box plots, those data points that are located outside and far away from the boxes are outliers. For the scatter plot, we can tell that there is a reasonable distribution of Loan Limit Amount versus the current balance, so as the LTV Ratio versus the current balance. Most of the data points in "Loan Limit Amount vs. Current Balance" plot and "Loan-to-Ratio vs. Current Balance" plot are limited in a triangle area. Those data points that are located outside the triangle are the outliers.

```
p1 <-ggplot(heloc, aes(x = Score.Range, y = Loan.Limit.Amount, color = Score.Range)) + geom_boxplot() +
p2 <-ggplot(heloc, aes(x = Score.Range, y = ORIGINAL.L.V, color = Score.Range)) + geom_boxplot() + lab
p3 <-ggplot(heloc, aes(x = Score.Range, y = Int.Rate, color = Score.Range)) + geom_boxplot() + labs(ti
p4 <-ggplot(heloc, aes(x = DTI.Range, y = Loan.Limit.Amount, color = DTI.Range)) + geom_boxplot() + lab
```

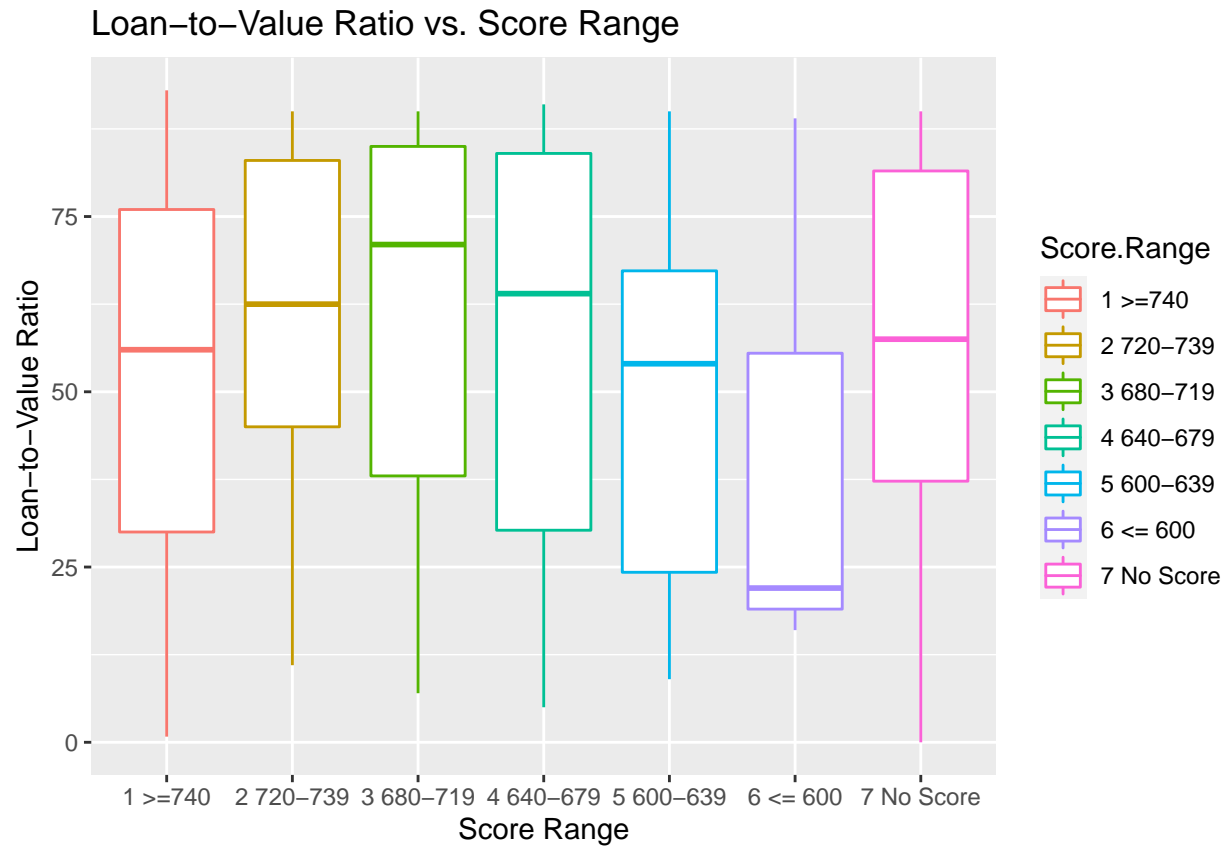
```

p5 <- ggplot(heloc, aes(x = DTI.Range, y = ORIGINAL.L.V, color = DTI.Range)) + geom_boxplot() + labs(title="Loan Limit Amount vs. DTI Range")
p6 <- ggplot(heloc, aes(x = DTI.Range, y = Int.Rate, color = DTI.Range)) + geom_boxplot() + labs(title="Interest Rate vs. DTI Range")
p7 <- ggplot(heloc, aes(x = Current.Balance, y = Loan.Limit.Amount)) + geom_point() + labs(title="Loan Limit Amount vs. Current Balance")
p8 <- ggplot(heloc, aes(x = Current.Balance, y = ORIGINAL.L.V)) + geom_point() + labs(title="Loan-to-Value Ratio vs. Current Balance")
p9 <- ggplot(heloc, aes(x = Current.Balance, y = Int.Rate)) + geom_point() + labs(title="Interest Rate vs. Current Balance")
p1

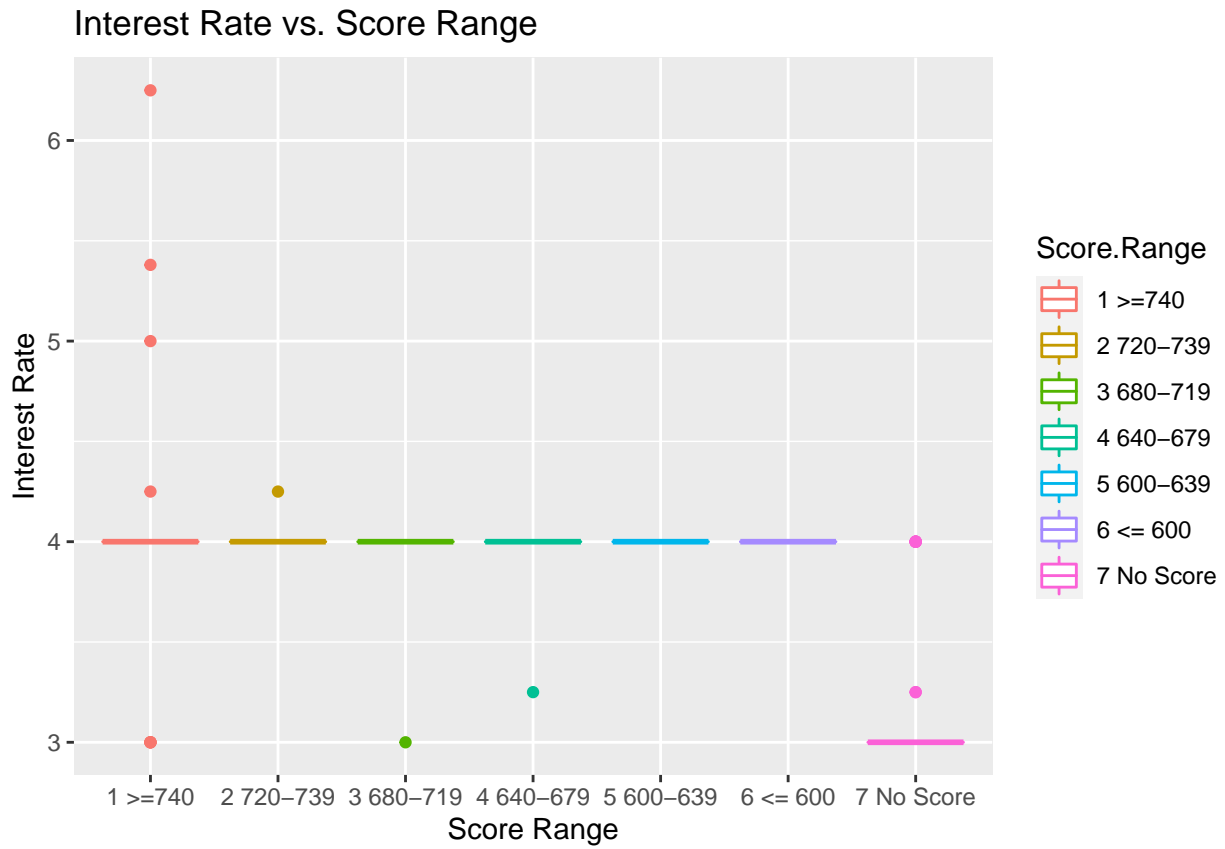
```



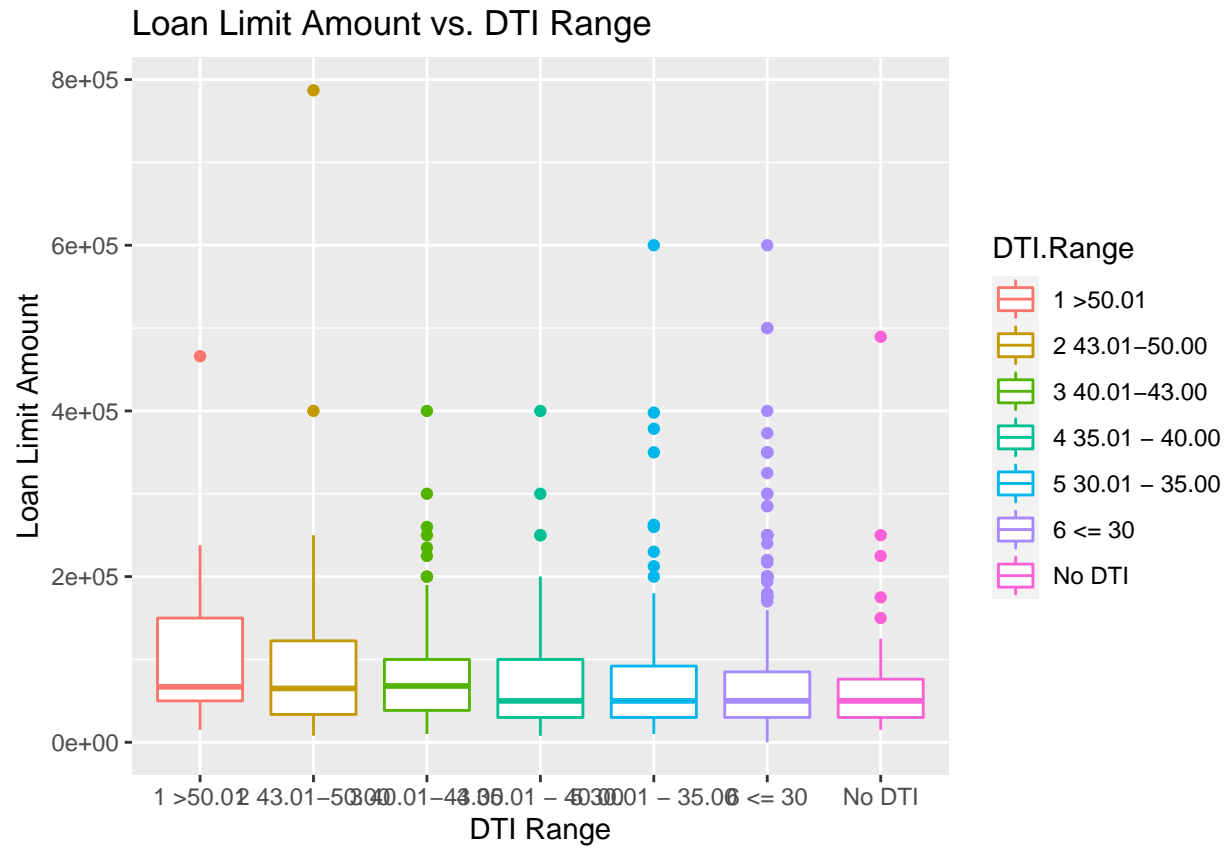
p2



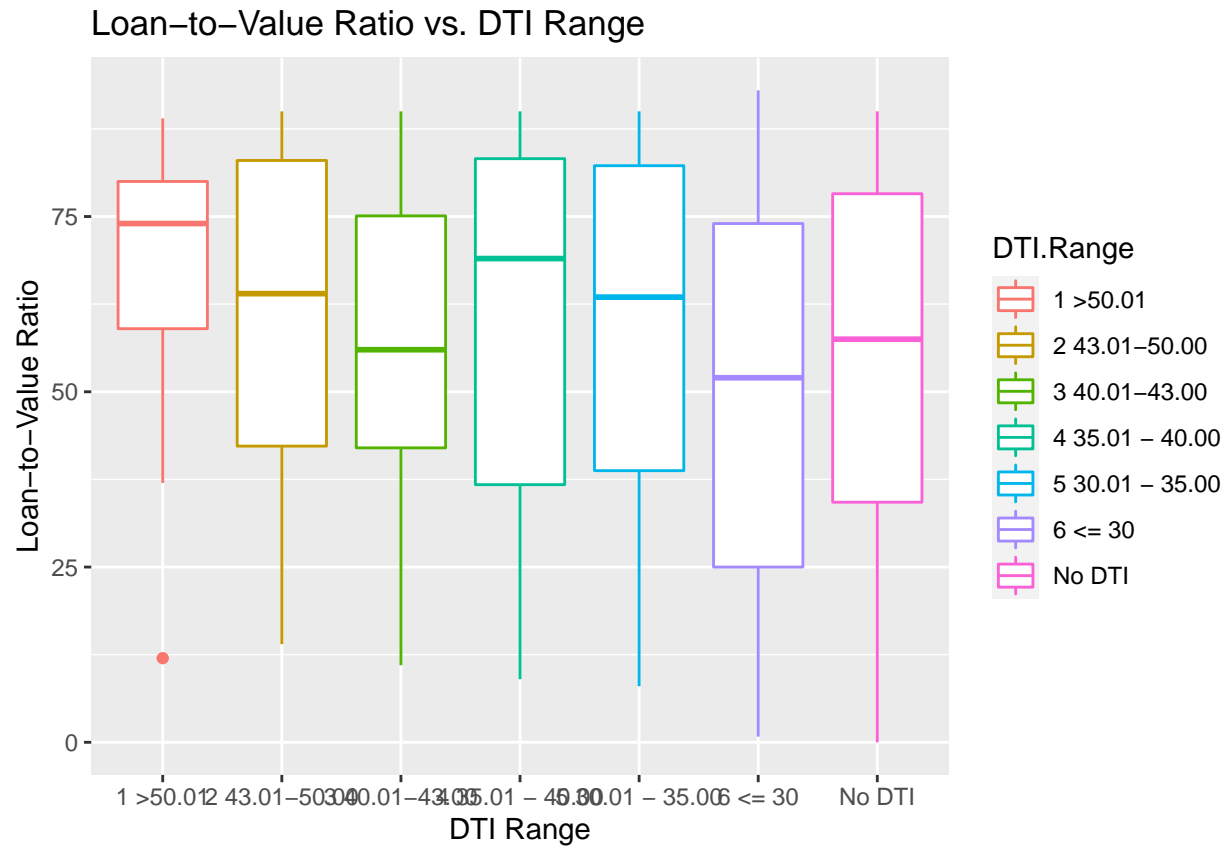
p3



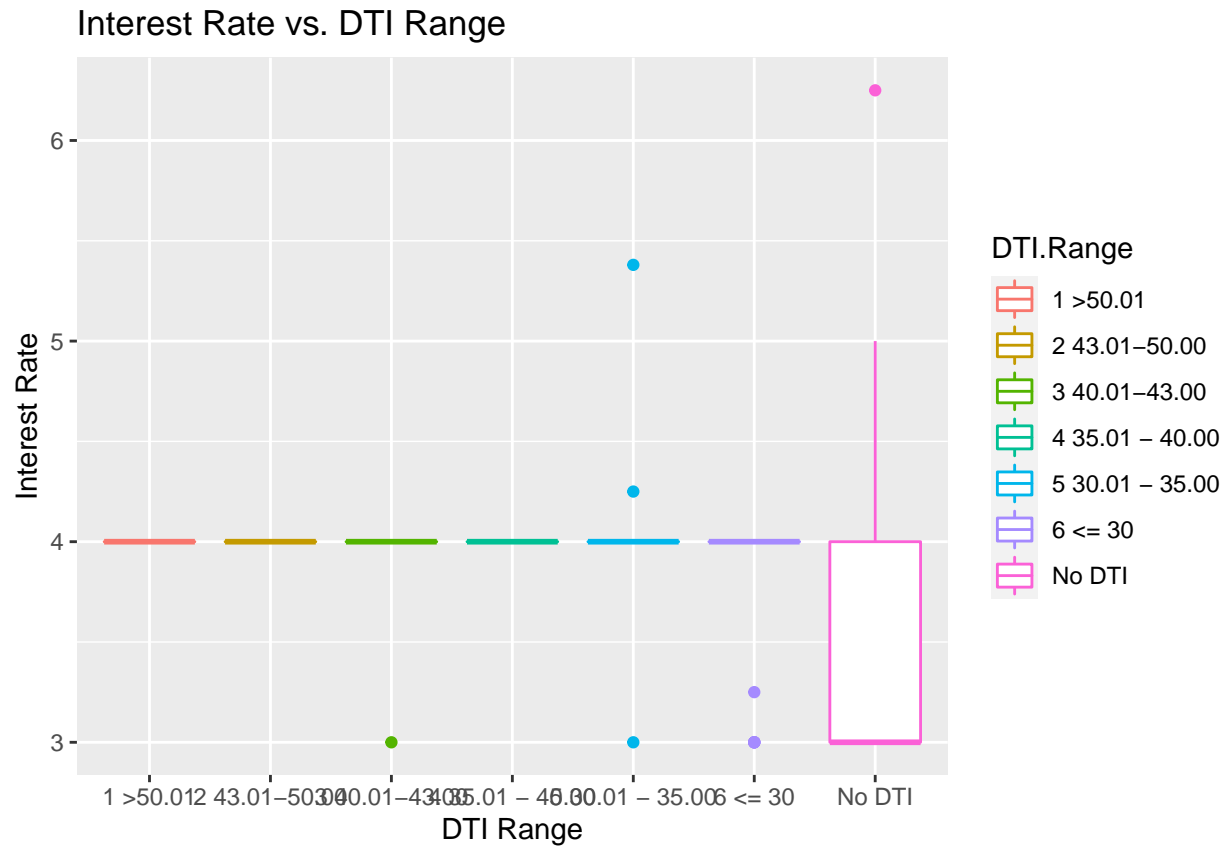
p4



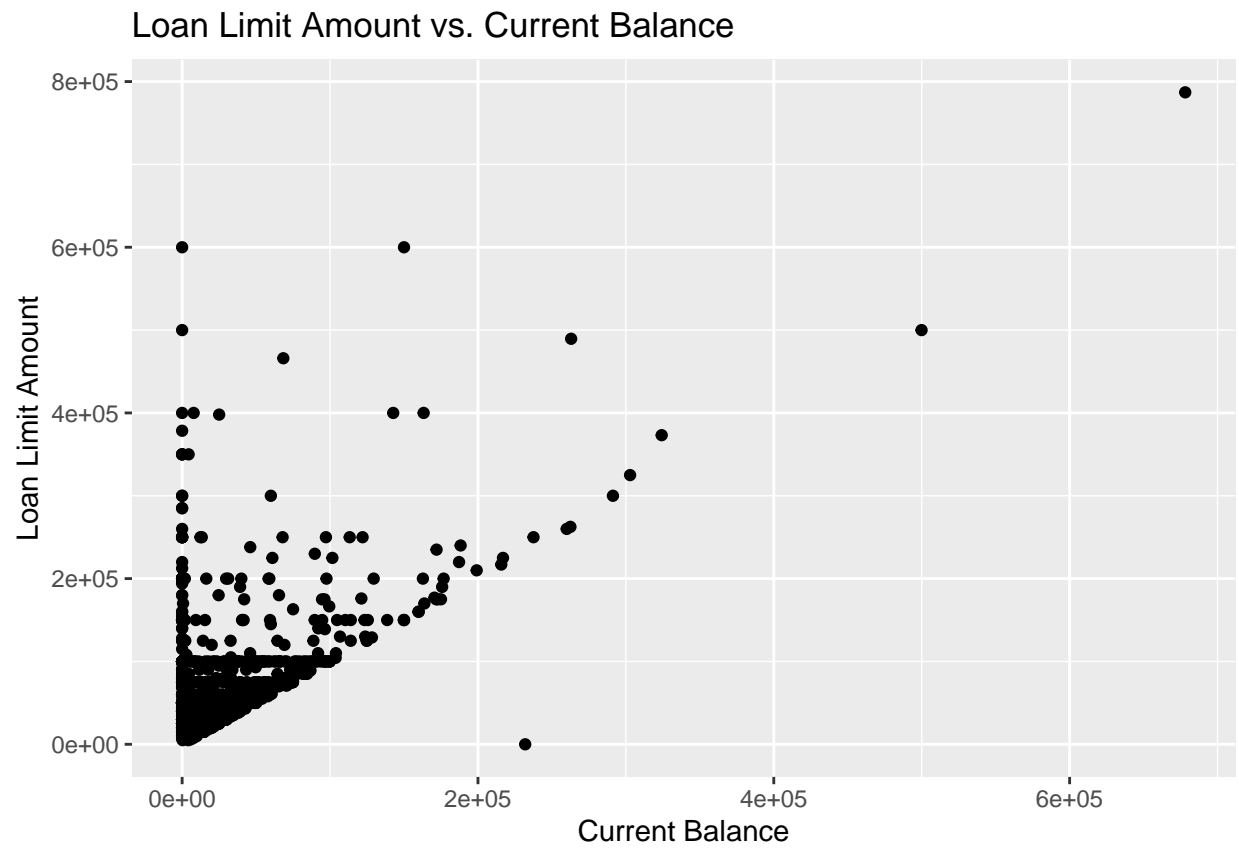
p5



p6

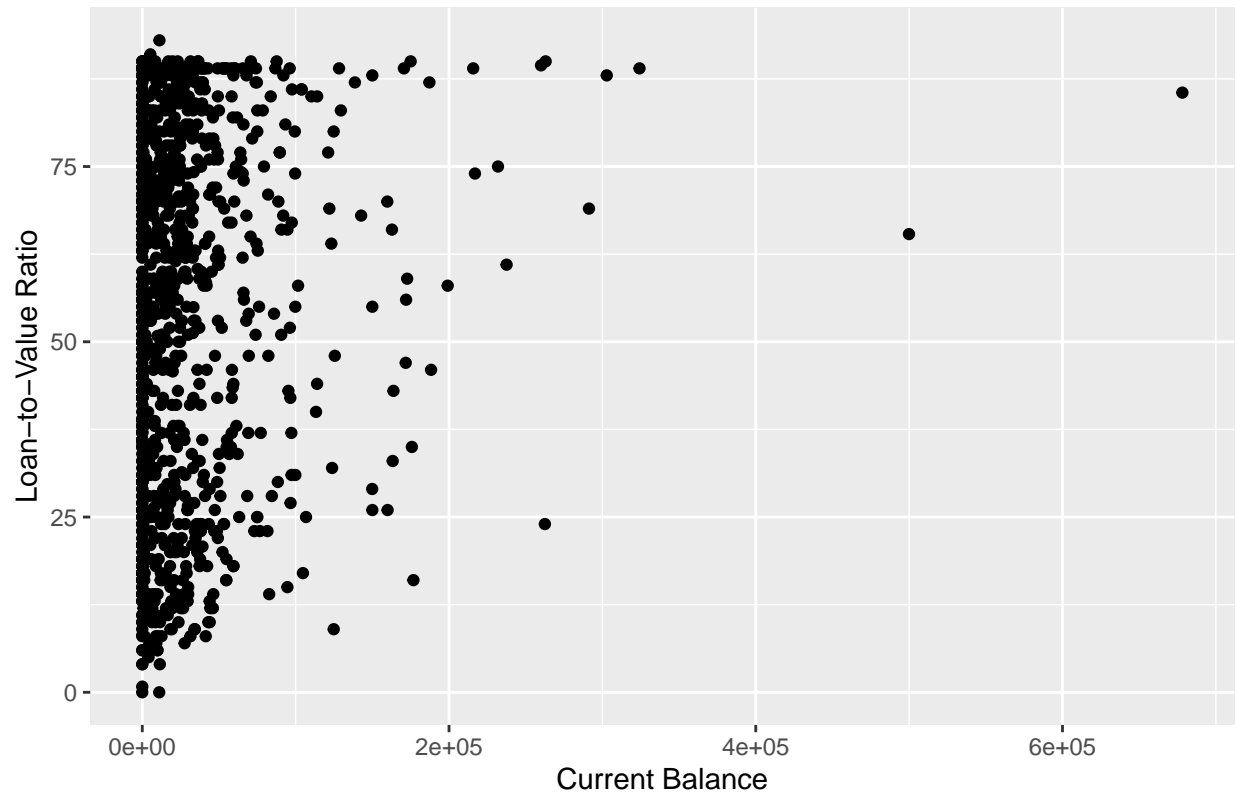


p7

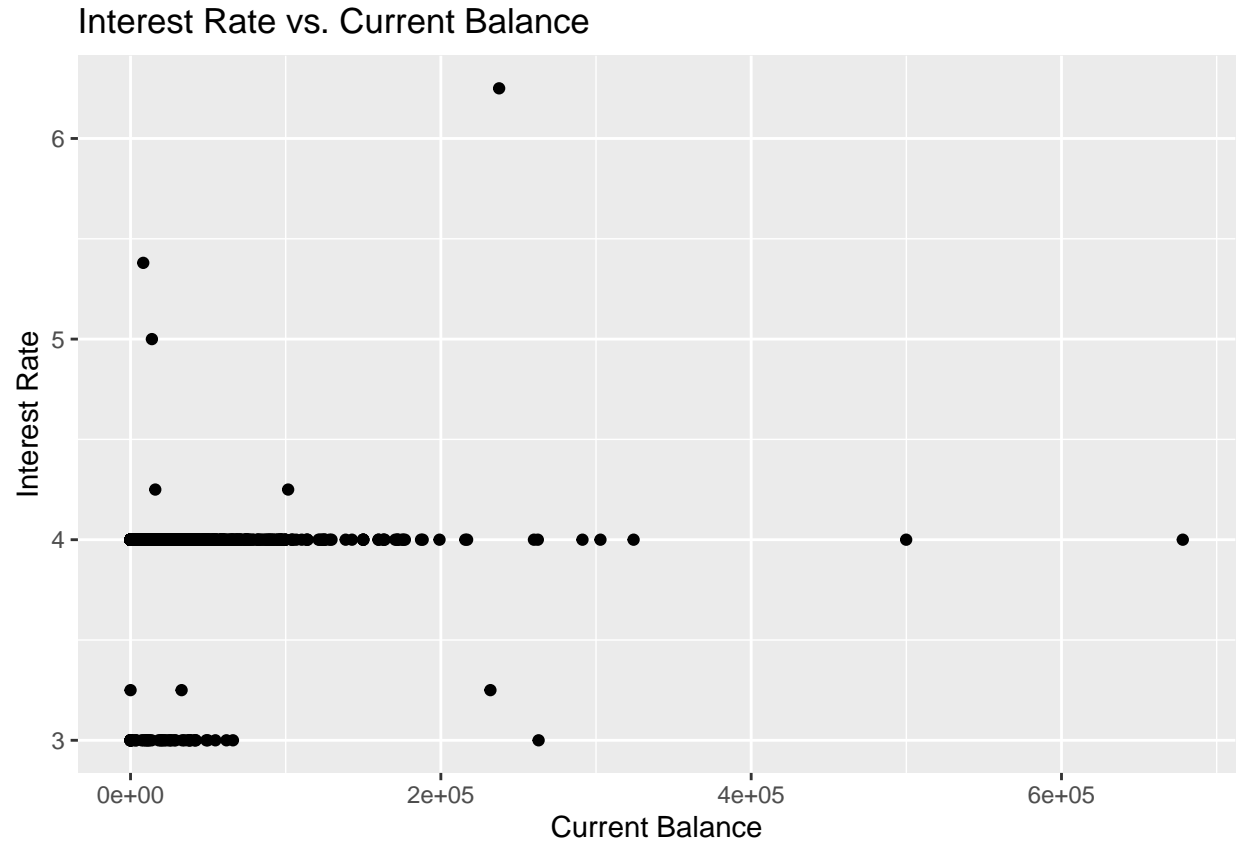


p8

Loan-to-Value Ratio vs. Current Balance



p9



Outliers Analysis

Outliers in Loan Limit Amount vs.Credit Score Range

After finding out those outliers' existence, we need to identify those account record. They are treated as outliers under single dimension, but still we need to check other dimensions simultaneously.

```
lla_outliers_score7 <- heloc[(heloc$Score.Range=="7 No Score") & (heloc$Loan.Limit.Amount>=150000),]
lla_outliers_score7
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 68      HELOC      69      41953.65      175000      3
## 348     HELOC     349     263000.00     489500      3
## 648     HELOC     649        0.00     150000      3
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 68      8/18/09      NA 120      0      58
## 348     9/25/19      NA 120      0      90
## 648     9/12/08      NA 120      0      50
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range DTI.Range
## 68      0      1      0 7 No Score No DTI
## 348      0      1      0 7 No Score No DTI
## 648      0      1      0 7 No Score No DTI
##      Orig.LTV.Range Yield.Range Delinquency.Range Days.Late
## 68      6 <60% LTV 2 3.0% - 3.99%      1 Current      0
```

```
## 348 2 90% - 99.99% LTV 2 3.0% - 3.99%          1 Current          0
## 648          6 <60% LTV 2 3.0% - 3.99%          1 Current          0
```

```
lla_outliers_score5 <- heloc[(heloc$Score.Range=="5 600-639") & (heloc$Loan.Limit.Amount>=100000),]
lla_outliers_score5
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 216      HELOC          217      163774.20      170000          4
## 288      HELOC          289      93306.34      100000          4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 216          2/7/20          NA 120          631          43
## 288          2/11/04          NA 120          621          81
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 216          36.45          1          0 5 600-639
## 288          78.00          1          0 5 600-639
##      DTI.Range      Orig.LTV.Range      Yield.Range Delinquency.Range
## 216 4 35.01 - 40.00      6 <60% LTV 3 >= 4.% - 4.25%          1 Current
## 288      1 >50.01 3 80% - 89.99% LTV 3 >= 4.% - 4.25%          1 Current
##      Days.Late
## 216          0
## 288          0
```

```
lla_outliers_score4 <- heloc[(heloc$Score.Range=="4 640-679") & (heloc$Loan.Limit.Amount>100000),]
lla_outliers_score4
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 65      HELOC          66      260000.0      260000          4
## 305      HELOC          306      150000.0      150000          4
## 894      HELOC          895      187211.2      220000          4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 65          9/20/19          NA 120          679          89.43
## 305          6/23/17          NA 120          652          55.00
## 894          3/8/17          NA 120          669          87.00
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 65          40.70          1          0 4 640-679
## 305          42.80          1          0 4 640-679
## 894          28.11          1          0 4 640-679
##      DTI.Range      Orig.LTV.Range      Yield.Range Delinquency.Range
## 65 3 40.01-43.00 3 80% - 89.99% LTV 3 >= 4.% - 4.25%          1 Current
## 305 3 40.01-43.00      6 <60% LTV 3 >= 4.% - 4.25%          2 1 to 29 DPD
## 894      6 <= 30 3 80% - 89.99% LTV 3 >= 4.% - 4.25%          1 Current
##      Days.Late
## 65          0
## 305          7
## 894          0
```

```
lla_outliers_score3 <- heloc[(heloc$Score.Range=="3 680-719") & (heloc$Loan.Limit.Amount>150000),]
lla_outliers_score3
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount
## 8      HELOC          9          0.0      400000
## 109     HELOC         110          0.0      285000
```

##	276	HELOC	277	162796.2	200000
##	397	HELOC	398	302857.0	325000
##	424	HELOC	425	171770.2	175000
##	885	HELOC	886	0.0	160000
##	1039	HELOC	1041	25038.7	398000
##		Int.Rate	Date.Org.Opened	Date.Closed	Term CREDIT.SCORE ORIGINAL.L.V
##	8	4	4/14/16	NA	120 717 88
##	109	4	2/19/15	NA	120 694 65
##	276	4	1/14/12	NA	120 690 66
##	397	4	5/12/17	NA	120 718 88
##	424	4	6/25/18	NA	120 696 47
##	885	4	1/19/05	NA	120 693 63
##	1039	4	10/24/18	NA	120 693 89
##		DEBT.TO.INCOME	OCCUPANCY.CODE.Cd.1	PROPERTY.TYPE	Score.Range
##	8	39	1	0	3 680-719
##	109	20	1	0	3 680-719
##	276	52	1	0	3 680-719
##	397	26	1	0	3 680-719
##	424	43	1	0	3 680-719
##	885	4	1	0	3 680-719
##	1039	33	1	0	3 680-719
##		DTI.Range	Orig.LTV.Range	Yield.Range	
##	8	4 35.01 - 40.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	
##	109	6 <= 30 5 >= 60%	- 69.99% LTV 3	>= 4.% - 4.25%	
##	276	1 >50.01 5 >= 60%	- 69.99% LTV 3	>= 4.% - 4.25%	
##	397	6 <= 30 3 80%	- 89.99% LTV 3	>= 4.% - 4.25%	
##	424	3 40.01-43.00	6 <60% LTV 3	>= 4.% - 4.25%	
##	885	6 <= 30 5 >= 60%	- 69.99% LTV 3	>= 4.% - 4.25%	
##	1039	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	
##		Delinquency.Range	Days.Late		
##	8	1 Current	0		
##	109	1 Current	0		
##	276	1 Current	0		
##	397	1 Current	0		
##	424	1 Current	0		
##	885	1 Current	0		
##	1039	1 Current	0		

```
lla_outliers_score2 <- heloc[(heloc$Score.Range=="2 720-739") & (heloc$Loan.Limit.Amount>=200000),]
lla_outliers_score2
```

##		Product.Type	Account.Number	Current.Balance	Loan.Limit.Amount	Int.Rate
##	1	HELOC	2	101582.47	225000	4.25
##	249	HELOC	250	188287.16	240000	4.00
##	427	HELOC	428	163280.02	400000	4.00
##	599	HELOC	600	97212.88	250000	4.00
##	630	HELOC	631	0.00	250000	4.00
##	864	HELOC	865	216920.50	225000	4.00
##		Date.Org.Opened	Date.Closed	Term	CREDIT.SCORE	ORIGINAL.L.V
##	1	3/4/05	NA	12	726	58
##	249	7/31/18	NA	120	736	46
##	427	11/19/19	NA	120	731	33
##	599	1/21/20	NA	120	736	37
##	630	1/27/16	NA	120	733	21

## 864	12/21/18	NA 120	737	74
##	DEBT.TO.INCOME	OCCUPANCY.CODE.Cd.1	PROPERTY.TYPE	Score.Range
## 1	0.00	1	0	2 720-739
## 249	20.00	1	0	2 720-739
## 427	29.72	1	0	2 720-739
## 599	36.97	1	0	2 720-739
## 630	16.00	1	0	2 720-739
## 864	43.00	1	0	2 720-739
##	DTI.Range	Orig.LTV.Range	Yield.Range	
## 1	No DTI	6 <60% LTV 4	>= 4.25% - 4.49%	
## 249	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%	
## 427	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%	
## 599	4 35.01 - 40.00	6 <60% LTV 3	>= 4.% - 4.25%	
## 630	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%	
## 864	3 40.01-43.00 4	>= 70% - 79.99% LTV 3	>= 4.% - 4.25%	
##	Delinquency.Range	Days.Late		
## 1	1 Current	0		
## 249	1 Current	0		
## 427	1 Current	0		
## 599	1 Current	0		
## 630	1 Current	0		
## 864	1 Current	0		

```
lla_outliers_score1 <- heloc[(heloc$Score.Range=="1 >=740") & (heloc$Loan.Limit.Amount>400000),]
lla_outliers_score1
```

##	Product.Type	Account.Number	Current.Balance	Loan.Limit.Amount
## 7	HELOC	8	0.00	500000
## 34	HELOC	35	0.00	600000
## 497	HELOC	498	68403.55	466000
## 711	HELOC	712	499899.35	500000
## 986	HELOC	987	150000.00	600000
## 1000	HELOC	1001	678161.36	787000
##	Int.Rate	Date.Org.Opened	Date.Closed	Term CREDIT.SCORE ORIGINAL.L.V
## 7	4	5/10/16	NA 120	768 56.00
## 34	4	4/17/19	NA 120	762 85.00
## 497	4	3/22/16	NA 120	785 89.00
## 711	4	9/30/19	NA 120	790 65.36
## 986	4	3/17/17	NA 120	776 88.00
## 1000	4	10/5/19	NA 120	786 85.54
##	DEBT.TO.INCOME	OCCUPANCY.CODE.Cd.1	PROPERTY.TYPE	Score.Range
## 7	7.00	1	0	1 >=740
## 34	25.00	1	0	1 >=740
## 497	73.00	1	0	1 >=740
## 711	4.22	1	0	1 >=740
## 986	31.00	1	0	1 >=740
## 1000	49.63	1	0	1 >=740
##	DTI.Range	Orig.LTV.Range	Yield.Range	
## 7	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%	
## 34	6 <= 30	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	
## 497	1 >50.01	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	
## 711	6 <= 30 5	>= 60% - 69.99% LTV 3	>= 4.% - 4.25%	
## 986	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	
## 1000	2 43.01-50.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	

```
##      Delinquency.Range Days.Late
## 7      1 Current      0
## 34     1 Current      0
## 497    1 Current      0
## 711    1 Current      0
## 986    1 Current      0
## 1000   1 Current      0
```

When it comes to the outliers in Loan Limit Amount by Credit Score Range category, there are six accounts that need more consideration: 1. 649. This account doesn't have any balance and the Credit Score is 0, yet the Loan Limit Amount is very high in this score range category. 2. 1041. Generally speaking, we have observed that the Loan Limit Amount and Current Balance is somehow very similar, if this specific customer have a decent Credit Score and relative low DTI Ratio. But for this account, the Loan Limit Amount is extremely high considering its Credit Score and DTI level. Neither its Credit Score is good nor its DTI level is very desirable. 3. 428. The account 428 has a relative similar balance level with account 250, but the Credit Score of 428 is lower than account 250 and the DTI of 428 is higher than account 250. Under this situation, account 428 has almost double amount of Loan Limit than account 250. My suggestion is to dig deeper on whether we give a reasonable Loan Limit Amount to account 428. 4. 35. Similar to what is mentioned in bullet point 3, account 35 and account 8 has the same 0 balance. However, the Loan Limit Amount of account 35 is higher than account 8 when account has lower Credit Score and higher DTI Ratio comparing to account 8. So, we need to dive deeper about account 35. 5. 498. Same reason for account 498 when comparing to account 712. 6. 987. Same reason for account 498 when comparing to account 712.

Outliers in Interest Rate vs.Credit Score Range

```
ir_outliers <- heloc[heloc$Int.Rate != 4,]
```

When it comes to the Interest Rate, one thing that I found is that Interest Rate is mainly associated with Term. Even the outliers in the plot also accord with rule of higher Interest Rate corresponding to longer Term. Thus, the Interest Rate makes sense in nearly all the observations.

Outliers in Loan Limit Amount vs.DTI Range

```
lla_outliers_dti1 <- heloc[(heloc$DTI.Range=="1 >50.01") & (heloc$Loan.Limit.Amount>=400000),]
lla_outliers_dti1
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 497      HELOC      498      68403.55      466000      4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 497      3/22/16      NA      120      785      89
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range DTI.Range
## 497      73      1      0      1 >=740      1 >50.01
##      Orig.LTV.Range      Yield.Range Delinquency.Range Days.Late
## 497 3 80% - 89.99% LTV 3 >= 4.% - 4.25%      1 Current      0
```

```
lla_outliers_dti2 <- heloc[(heloc$DTI.Range=="2 43.01-50.00") & (heloc$Loan.Limit.Amount>=400000),]
lla_outliers_dti2
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount
## 346      HELOC      347      7792.59      400000
## 1000     HELOC      1001     678161.36      787000
##      Int.Rate Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 346      4      4/3/20      NA 120      814      43.00
## 1000     4      10/5/19     NA 120      786      85.54
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 346      46.69      1      0      1 >=740
## 1000     49.63      1      0      1 >=740
##      DTI.Range      Orig.LTV.Range      Yield.Range Delinquency.Range
## 346 2 43.01-50.00      6 <60% LTV 3 >= 4.% - 4.25%      1 Current
## 1000 2 43.01-50.00 3 80% - 89.99% LTV 3 >= 4.% - 4.25%      1 Current
##      Days.Late
## 346      0
## 1000     0
```

```
lla_outliers_dti3 <- heloc[(heloc$DTI.Range=="3 40.01-43.00") & (heloc$Loan.Limit.Amount>=200000),]
lla_outliers_dti3
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 38      HELOC      39      0.0      200000      4
## 65      HELOC      66      260000.0      260000      4
## 547     HELOC      548      142711.2      400000      4
## 592     HELOC      593      113291.7      250000      4
## 684     HELOC      685      172000.0      235000      4
## 690     HELOC      691      60000.0      300000      4
## 864     HELOC      865      216920.5      225000      4
## 923     HELOC      924      31119.9      200000      4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 38      8/9/10      NA 120      803      79.00
## 65      9/20/19     NA 120      679      89.43
## 547     10/2/13     NA 120      776      68.00
## 592     12/4/14     NA 120      780      40.00
## 684     9/14/18     NA 120      747      56.00
## 690     7/10/19     NA 120      819      70.00
## 864     12/21/18    NA 120      737      74.00
## 923     4/27/18     NA 120      804      41.00
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 38      42.0      1      0      1 >=740
## 65      40.7      1      0      4 640-679
## 547     42.0      1      0      1 >=740
## 592     43.0      1      0      1 >=740
## 684     42.0      1      0      1 >=740
## 690     43.0      1      0      1 >=740
## 864     43.0      1      0      2 720-739
## 923     42.0      1      0      1 >=740
##      DTI.Range      Orig.LTV.Range      Yield.Range
## 38 3 40.01-43.00 4 >= 70% - 79.99% LTV 3 >= 4.% - 4.25%
## 65 3 40.01-43.00 3 80% - 89.99% LTV 3 >= 4.% - 4.25%
## 547 3 40.01-43.00 5 >= 60% - 69.99% LTV 3 >= 4.% - 4.25%
## 592 3 40.01-43.00      6 <60% LTV 3 >= 4.% - 4.25%
## 684 3 40.01-43.00      6 <60% LTV 3 >= 4.% - 4.25%
## 690 3 40.01-43.00 4 >= 70% - 79.99% LTV 3 >= 4.% - 4.25%
## 864 3 40.01-43.00 4 >= 70% - 79.99% LTV 3 >= 4.% - 4.25%
```

```
## 923 3 40.01-43.00          6 <60% LTV 3 >= 4.% - 4.25%
##      Delinquency.Range Days.Late
## 38      1 Current          0
## 65      1 Current          0
## 547     1 Current          0
## 592     1 Current          0
## 684     1 Current          0
## 690     1 Current          0
## 864     1 Current          0
## 923     1 Current          0
```

```
lla_outliers_dti4 <- heloc[(heloc$DTI.Range=="4 35.01 - 40.00") & (heloc$Loan.Limit.Amount>200000),]
lla_outliers_dti4
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 8      HELOC          9          0.00          400000          4
## 100     HELOC         101          0.00          250000          4
## 585     HELOC         586          0.00          300000          4
## 599     HELOC         600          97212.88          250000          4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 8      4/14/16          NA 120          717          88.00
## 100     8/30/10          NA 120          803          79.00
## 585    10/19/19          NA 120          814          57.14
## 599    1/21/20          NA 120          736          37.00
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 8      39.00          1          0 3 680-719
## 100     36.00          1          0 1 >=740
## 585     38.58          1          0 1 >=740
## 599     36.97          1          0 2 720-739
##      DTI.Range      Orig.LTV.Range      Yield.Range
## 8  4 35.01 - 40.00  3 80% - 89.99% LTV 3 >= 4.% - 4.25%
## 100 4 35.01 - 40.00 4 >= 70% - 79.99% LTV 3 >= 4.% - 4.25%
## 585 4 35.01 - 40.00          6 <60% LTV 3 >= 4.% - 4.25%
## 599 4 35.01 - 40.00          6 <60% LTV 3 >= 4.% - 4.25%
##      Delinquency.Range Days.Late
## 8      1 Current          0
## 100     1 Current          0
## 585     1 Current          0
## 599     1 Current          0
```

```
lla_outliers_dti5 <- heloc[(heloc$DTI.Range=="5 30.01 - 35.00") & (heloc$Loan.Limit.Amount>200000),]
lla_outliers_dti5
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount
## 196     HELOC         197          0.00          212500
## 217     HELOC         218          262500.00          262500
## 224     HELOC         225          0.00          350000
## 250     HELOC         251          89696.65          230000
## 311     HELOC         312          0.00          260000
## 986     HELOC         987          150000.00          600000
## 996     HELOC         997          0.00          378500
## 1039    HELOC        1041          25038.70          398000
##      Int.Rate Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
```

##	196	4	11/9/15	NA	120	796	84		
##	217	4	9/13/17	NA	120	785	24		
##	224	4	3/31/17	NA	120	800	53		
##	250	4	3/4/16	NA	120	775	77		
##	311	4	1/27/17	NA	120	791	83		
##	986	4	3/17/17	NA	120	776	88		
##	996	4	5/10/19	NA	120	791	77		
##	1039	4	10/24/18	NA	120	693	89		
##	DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range								
##	196	35	1	0	1	>=740			
##	217	32	1	0	1	>=740			
##	224	34	1	0	1	>=740			
##	250	32	1	0	1	>=740			
##	311	34	1	0	1	>=740			
##	986	31	1	0	1	>=740			
##	996	34	1	0	1	>=740			
##	1039	33	1	0	3	680-719			
##	DTI.Range Orig.LTV.Range Yield.Range								
##	196	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%					
##	217	5 30.01 - 35.00	6 <60% LTV 3	>= 4.% - 4.25%					
##	224	5 30.01 - 35.00	6 <60% LTV 3	>= 4.% - 4.25%					
##	250	5 30.01 - 35.00	4 >= 70% - 79.99% LTV 3	>= 4.% - 4.25%					
##	311	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%					
##	986	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%					
##	996	5 30.01 - 35.00	4 >= 70% - 79.99% LTV 3	>= 4.% - 4.25%					
##	1039	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%					
##	Delinquency.Range Days.Late								
##	196	1 Current	0						
##	217	1 Current	0						
##	224	1 Current	0						
##	250	1 Current	0						
##	311	1 Current	0						
##	986	1 Current	0						
##	996	1 Current	0						
##	1039	1 Current	0						

```
lla_outliers_dti6 <- heloc[(heloc$DTI.Range=="6 <= 30") & (heloc$Loan.Limit.Amount>=270000),]
lla_outliers_dti6
```

##	Product.Type	Account.Number	Current.Balance	Loan.Limit.Amount	Int.Rate	
##	7	HELOC	8	0.00	500000	4
##	34	HELOC	35	0.00	600000	4
##	90	HELOC	91	4389.45	350000	4
##	109	HELOC	110	0.00	285000	4
##	292	HELOC	293	291264.12	300000	4
##	392	HELOC	393	0.00	285000	4
##	397	HELOC	398	302856.99	325000	4
##	423	HELOC	424	0.00	300000	4
##	427	HELOC	428	163280.02	400000	4
##	711	HELOC	712	499899.35	500000	4
##	777	HELOC	778	0.00	350000	4
##	785	HELOC	786	324213.32	373150	4
##	965	HELOC	966	0.00	350000	4
##	Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V					

## 7	5/10/16	NA	120	768	56.00
## 34	4/17/19	NA	120	762	85.00
## 90	12/6/16	NA	120	786	54.00
## 109	2/19/15	NA	120	694	65.00
## 292	4/13/18	NA	120	797	69.00
## 392	10/8/15	NA	120	767	89.00
## 397	5/12/17	NA	120	718	88.00
## 423	4/24/18	NA	120	792	57.00
## 427	11/19/19	NA	120	731	33.00
## 711	9/30/19	NA	120	790	65.36
## 777	2/28/18	NA	120	805	72.00
## 785	8/21/15	NA	120	751	89.00
## 965	5/14/19	NA	120	805	70.00
##	DEBT.TO.INCOME	OCCUPANCY.CODE	Cd.1	PROPERTY.TYPE	Score.Range DTI.Range
## 7	7.00		1	0	1 >=740 6 <= 30
## 34	25.00		1	0	1 >=740 6 <= 30
## 90	25.00		1	0	1 >=740 6 <= 30
## 109	20.00		1	0	3 680-719 6 <= 30
## 292	24.00		1	0	1 >=740 6 <= 30
## 392	22.00		1	0	1 >=740 6 <= 30
## 397	26.00		1	0	3 680-719 6 <= 30
## 423	29.00		1	0	1 >=740 6 <= 30
## 427	29.72		1	0	2 720-739 6 <= 30
## 711	4.22		1	0	1 >=740 6 <= 30
## 777	20.00		1	0	1 >=740 6 <= 30
## 785	23.00		1	0	1 >=740 6 <= 30
## 965	26.00		1	0	1 >=740 6 <= 30
##	Orig.LTV.Range	Yield.Range	Delinquency.Range	Days.Late	
## 7	6 <60% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 34	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 90	6 <60% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 109	5 >= 60% - 69.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 292	5 >= 60% - 69.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 392	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 397	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 423	6 <60% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 427	6 <60% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 711	5 >= 60% - 69.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 777	4 >= 70% - 79.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 785	3 80% - 89.99% LTV 3	>= 4.% - 4.25%	1 Current	0	
## 965	4 >= 70% - 79.99% LTV 3	>= 4.% - 4.25%	1 Current	0	

```
lla_outliers_dti7 <- heloc[(heloc$DTI.Range=="No DTI") & (heloc$Loan.Limit.Amount>100000),]
lla_outliers_dti7
```

##	Product.Type	Account.Number	Current.Balance	Loan.Limit.Amount	Int.Rate
## 1	HELOC	2	101582.47	225000	4.25
## 2	HELOC	3	237551.34	250000	6.25
## 68	HELOC	69	41953.65	175000	3.00
## 326	HELOC	327	0.00	125000	3.00
## 348	HELOC	349	263000.00	489500	3.00
## 648	HELOC	649	0.00	150000	3.00
##	Date.Org.Opened	Date.Closed	Term	CREDIT.SCORE	ORIGINAL.L.V
## 1	3/4/05	NA	12	726	58

Account Number	Date	Product	Balance	Limit	Score	DTI
2	3/13/13	NA	28	745	61	
68	8/18/09	NA	120	0	58	
326	3/8/17	NA	120	0	35	
348	9/25/19	NA	120	0	90	
648	9/12/08	NA	120	0	50	

Account Number	DEBT.TO.INCOME	OCCUPANCY.CODE	Cd.1	PROPERTY.TYPE	Score.Range	DTI.Range
1	0		1	0	2 720-739	No DTI
2	0		2	0	1 >=740	No DTI
68	0		1	0	7 No Score	No DTI
326	0		1	0	7 No Score	No DTI
348	0		1	0	7 No Score	No DTI
648	0		1	0	7 No Score	No DTI

Account Number	Orig.LTV.Range	Yield.Range	Delinquency.Range	Days.Late
1	6 <60% LTV 4 >= 4.25% - 4.49%		1 Current	0
2	5 >= 60% - 69.99% LTV	8 >=6% LTV	1 Current	0
68	6 <60% LTV	2 3.0% - 3.99%	1 Current	0
326	6 <60% LTV	2 3.0% - 3.99%	1 Current	0
348	2 90% - 99.99% LTV	2 3.0% - 3.99%	1 Current	0
648	6 <60% LTV	2 3.0% - 3.99%	1 Current	0

Here are some accounts that seems to have an abnormal Loan Limit Amount by DTI category: 1. 9. Comparing to account 101 and 586, it has lowest Credit Score and Highest DTI, but still it has the highest Loan Limit Amount among all three. 2. 1041. This account has the highest Loan Limit Amount in the "5" DTI Range while it has the lowest Credit Score and a relatively high DTI value. 3. 35. Comparing to account 8. 4. 428. This account has a very high Loan Limit Amount considering its Balance. It neither has a very high Credit Score nor a low DTI Ratio, but still one of the highest Loan Limit Amount among this group.

Outliers in Loan-to-Value Ratio vs.DTI Range

```
lti_outliers_dti1 <- heloc[(heloc$DTI.Range=="1 >50.01") & (heloc$ORIGINAL.L.V <= 25),]
lti_outliers_dti1
```

Account Number	Product	Current Balance	Loan Limit Amount	Int. Rate
266	HELOC	267	46008.85	50000
266	Date.Org.Opened	Date.Closed	Term	CREDIT.SCORE
266	7/22/08	NA	120	714
266	DEBT.TO.INCOME	OCCUPANCY.CODE	Cd.1	PROPERTY.TYPE
266	57		1	0
266	Orig.LTV.Range	Yield.Range	Delinquency.Range	Days.Late
266	6 <60% LTV 3 >= 4.% - 4.25%		1 Current	0

Outliers in Loan Limit Amount vs.Current Balance

```
lla_outliers_bal <- heloc[(heloc$Loan.Limit.Amount >= 300000) | (heloc$Loan.Limit.Amount == 0),]
lla_outliers_bal
```

Account Number	Product	Current Balance	Loan Limit Amount
7	HELOC	8	0.00
7			500000

## 8	HELOC	9	0.00	400000
## 34	HELOC	35	0.00	600000
## 90	HELOC	91	4389.45	350000
## 224	HELOC	225	0.00	350000
## 292	HELOC	293	291264.12	300000
## 346	HELOC	347	7792.59	400000
## 348	HELOC	349	263000.00	489500
## 397	HELOC	398	302856.99	325000
## 423	HELOC	424	0.00	300000
## 427	HELOC	428	163280.02	400000
## 497	HELOC	498	68403.55	466000
## 547	HELOC	548	142711.23	400000
## 583	HELOC	584	231938.68	0
## 585	HELOC	586	0.00	300000
## 690	HELOC	691	60000.00	300000
## 711	HELOC	712	499899.35	500000
## 777	HELOC	778	0.00	350000
## 785	HELOC	786	324213.32	373150
## 965	HELOC	966	0.00	350000
## 986	HELOC	987	150000.00	600000
## 996	HELOC	997	0.00	378500
## 1000	HELOC	1001	678161.36	787000
## 1039	HELOC	1041	25038.70	398000
##	Int.Rate	Date.Org.Opened	Date.Closed	Term CREDIT.SCORE ORIGINAL.L.V
## 7	4.00	5/10/16	NA 120	768 56.00
## 8	4.00	4/14/16	NA 120	717 88.00
## 34	4.00	4/17/19	NA 120	762 85.00
## 90	4.00	12/6/16	NA 120	786 54.00
## 224	4.00	3/31/17	NA 120	800 53.00
## 292	4.00	4/13/18	NA 120	797 69.00
## 346	4.00	4/3/20	NA 120	814 43.00
## 348	3.00	9/25/19	NA 120	0 90.00
## 397	4.00	5/12/17	NA 120	718 88.00
## 423	4.00	4/24/18	NA 120	792 57.00
## 427	4.00	11/19/19	NA 120	731 33.00
## 497	4.00	3/22/16	NA 120	785 89.00
## 547	4.00	10/2/13	NA 120	776 68.00
## 583	3.25	11/7/07	NA 60	657 75.00
## 585	4.00	10/19/19	NA 120	814 57.14
## 690	4.00	7/10/19	NA 120	819 70.00
## 711	4.00	9/30/19	NA 120	790 65.36
## 777	4.00	2/28/18	NA 120	805 72.00
## 785	4.00	8/21/15	NA 120	751 89.00
## 965	4.00	5/14/19	NA 120	805 70.00
## 986	4.00	3/17/17	NA 120	776 88.00
## 996	4.00	5/10/19	NA 120	791 77.00
## 1000	4.00	10/5/19	NA 120	786 85.54
## 1039	4.00	10/24/18	NA 120	693 89.00
##	DEBT.TO.INCOME	OCCUPANCY.CODE.Cd.1	PROPERTY.TYPE	Score.Range
## 7	7.00	1	0	1 >=740
## 8	39.00	1	0	3 680-719
## 34	25.00	1	0	1 >=740
## 90	25.00	1	0	1 >=740
## 224	34.00	1	0	1 >=740

## 292	24.00	1	0	1	>=740
## 346	46.69	1	0	1	>=740
## 348	0.00	1	0	7	No Score
## 397	26.00	1	0	3	680-719
## 423	29.00	1	0	1	>=740
## 427	29.72	1	0	2	720-739
## 497	73.00	1	0	1	>=740
## 547	42.00	1	0	1	>=740
## 583	20.00	1	0	4	640-679
## 585	38.58	1	0	1	>=740
## 690	43.00	1	0	1	>=740
## 711	4.22	1	0	1	>=740
## 777	20.00	1	0	1	>=740
## 785	23.00	1	0	1	>=740
## 965	26.00	1	0	1	>=740
## 986	31.00	1	0	1	>=740
## 996	34.00	1	0	1	>=740
## 1000	49.63	1	0	1	>=740
## 1039	33.00	1	0	3	680-719
##	DTI.Range	Orig.LTV.Range	Yield.Range		
## 7	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%		
## 8	4 35.01 - 40.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 34	6 <= 30	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 90	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%		
## 224	5 30.01 - 35.00	6 <60% LTV 3	>= 4.% - 4.25%		
## 292	6 <= 30 5	>= 60% - 69.99% LTV 3	>= 4.% - 4.25%		
## 346	2 43.01-50.00	6 <60% LTV 3	>= 4.% - 4.25%		
## 348	No DTI	2 90% - 99.99% LTV 2	3.0% - 3.99%		
## 397	6 <= 30	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 423	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%		
## 427	6 <= 30	6 <60% LTV 3	>= 4.% - 4.25%		
## 497	1 >50.01	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 547	3 40.01-43.00 5	>= 60% - 69.99% LTV 3	>= 4.% - 4.25%		
## 583	6 <= 30 4	>= 70% - 79.99% LTV 2	3.0% - 3.99%		
## 585	4 35.01 - 40.00	6 <60% LTV 3	>= 4.% - 4.25%		
## 690	3 40.01-43.00 4	>= 70% - 79.99% LTV 3	>= 4.% - 4.25%		
## 711	6 <= 30 5	>= 60% - 69.99% LTV 3	>= 4.% - 4.25%		
## 777	6 <= 30 4	>= 70% - 79.99% LTV 3	>= 4.% - 4.25%		
## 785	6 <= 30	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 965	6 <= 30 4	>= 70% - 79.99% LTV 3	>= 4.% - 4.25%		
## 986	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 996	5 30.01 - 35.00 4	>= 70% - 79.99% LTV 3	>= 4.% - 4.25%		
## 1000	2 43.01-50.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
## 1039	5 30.01 - 35.00	3 80% - 89.99% LTV 3	>= 4.% - 4.25%		
##	Delinquency.Range	Days.Late			
## 7	1 Current	0			
## 8	1 Current	0			
## 34	1 Current	0			
## 90	1 Current	0			
## 224	1 Current	0			
## 292	1 Current	0			
## 346	1 Current	0			
## 348	1 Current	0			
## 397	1 Current	0			

```
## 423      1 Current      0
## 427      1 Current      0
## 497      1 Current      0
## 547      1 Current      0
## 583      1 Current      0
## 585      1 Current      0
## 690      1 Current      0
## 711      1 Current      0
## 777      1 Current      0
## 785      1 Current      0
## 965      1 Current      0
## 986      1 Current      0
## 996      1 Current      0
## 1000     1 Current      0
## 1039     1 Current      0
```

1. 35. Comparing to account 8.
2. 584. It has a zero Loan Limit Amount.
- 3.1041. The Credit Score is relatively low, but still it has a high Loan Limit Amount.

Outliers in Loan-to-Value Ratio vs.Current Balance

```
ltv_outliers_bal_1 <- heloc[(heloc$Current.Balance >= 100000) & (heloc$ORIGINAL.L.V <= 12.5),]
ltv_outliers_bal_1
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 142      HELOC      143      124832.6      125000      4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 142      12/16/13      NA 120      688      9
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 142      38      1      0 3 680-719
##      DTI.Range Orig.LTV.Range      Yield.Range Delinquency.Range
## 142 4 35.01 - 40.00      6 <60% LTV 3 >= 4.% - 4.25%      1 Current
##      Days.Late
## 142      0
```

```
ltv_outliers_bal_2 <- heloc[(heloc$Current.Balance >= 150000) & (heloc$ORIGINAL.L.V <= 18.75),]
ltv_outliers_bal_2
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 146      HELOC      147      176805.8      2e+05      4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 146      12/12/02      NA 120      745      16
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range DTI.Range
## 146      7      1      0 1 >=740 6 <= 30
##      Orig.LTV.Range      Yield.Range Delinquency.Range Days.Late
## 146      6 <60% LTV 3 >= 4.% - 4.25%      1 Current      0
```

```
ltv_outliers_bal_3 <- heloc[(heloc$Current.Balance >= 200000) & (heloc$ORIGINAL.L.V <= 25),]
ltv_outliers_bal_3
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 217      HELOC          218          262500          262500          4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 217      9/13/17      NA 120          785          24
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 217          32          1          0          1 >=740
##      DTI.Range Orig.LTV.Range      Yield.Range Delinquency.Range
## 217 5 30.01 - 35.00      6 <60% LTV 3 >= 4.% - 4.25%          1 Current
##      Days.Late
## 217          0
```

```
ltv_outliers_bal_4 <- heloc[(heloc$Current.Balance >= 400000) & (heloc$ORIGINAL.L.V <= 75),]
ltv_outliers_bal_4
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount Int.Rate
## 711      HELOC          712      499899.3          5e+05          4
##      Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 711      9/30/19      NA 120          790          65.36
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range DTI.Range
## 711          4.22          1          0          1 >=740      6 <= 30
##      Orig.LTV.Range      Yield.Range Delinquency.Range Days.Late
## 711 5 >= 60% - 69.99% LTV 3 >= 4.% - 4.25%          1 Current          0
```

```
ltv_outliers_bal_5 <- heloc[(heloc$Current.Balance >= 600000) & (heloc$ORIGINAL.L.V <= 87.5),]
ltv_outliers_bal_5
```

```
##      Product.Type Account.Number Current.Balance Loan.Limit.Amount
## 1000      HELOC          1001      678161.4          787000
##      Int.Rate Date.Org.Opened Date.Closed Term CREDIT.SCORE ORIGINAL.L.V
## 1000      4      10/5/19      NA 120          786          85.54
##      DEBT.TO.INCOME OCCUPANCY.CODE.Cd.1 PROPERTY.TYPE Score.Range
## 1000          49.63          1          0          1 >=740
##      DTI.Range Orig.LTV.Range      Yield.Range Delinquency.Range
## 1000 2 43.01-50.00 3 80% - 89.99% LTV 3 >= 4.% - 4.25%          1 Current
##      Days.Late
## 1000          0
```

The outliers in scatter plots seems to be reasonable, not very extreme.

Customer Segmentation Analysis

```
library(scatterplot3d)
library(RColorBrewer)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## Attaching package: 'dplyr'

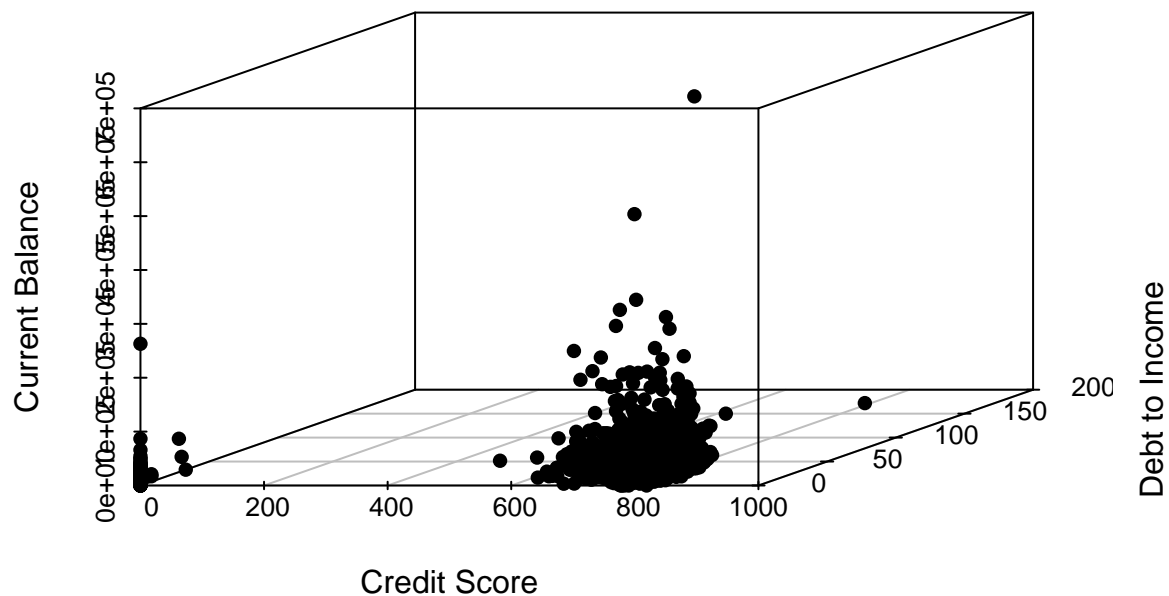
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.6.2
```

```
scatterplot3d(heloc %>% select(CREDIT.SCORE, DEBT.TO.INCOME, Current.Balance),
  xlab = "Credit Score",
  ylab = "Debt to Income",
  zlab = "Current Balance",
  pch = 16)
```



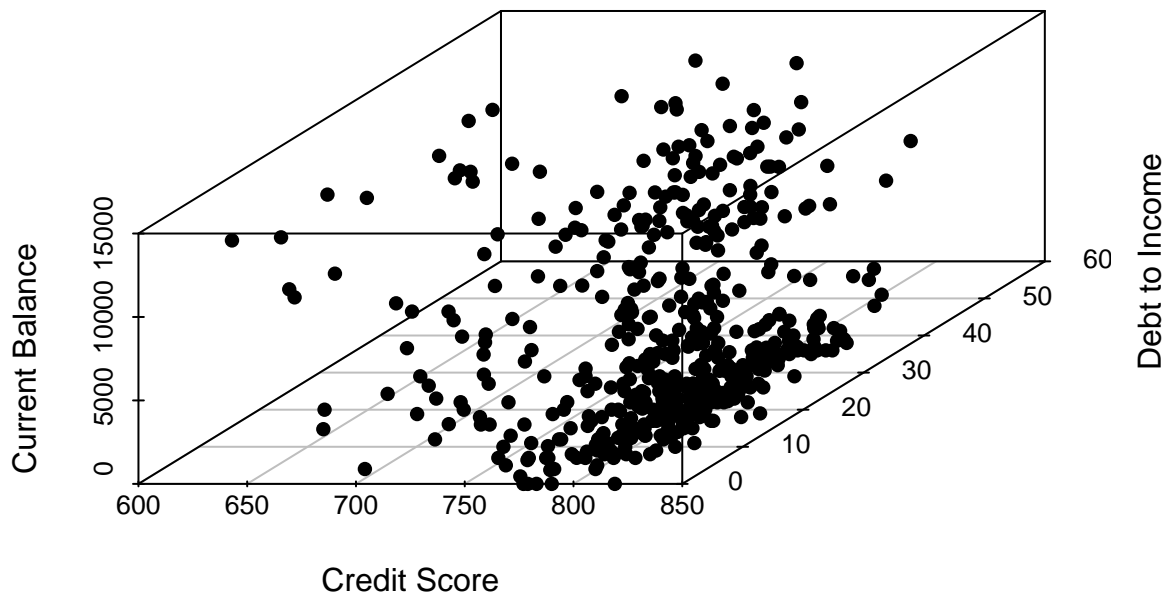
```
# Remove the outliers in scatter plot
```

```
scatterplot3d(heloc %>% select(CREDIT.SCORE, DEBT.TO.INCOME, Current.Balance) %>% filter(CREDIT.SCORE >
```

```

xlab = "Credit Score",
ylab = "Debt to Income",
zlab = "Current Balance",
pch = 16)

```



Clustering

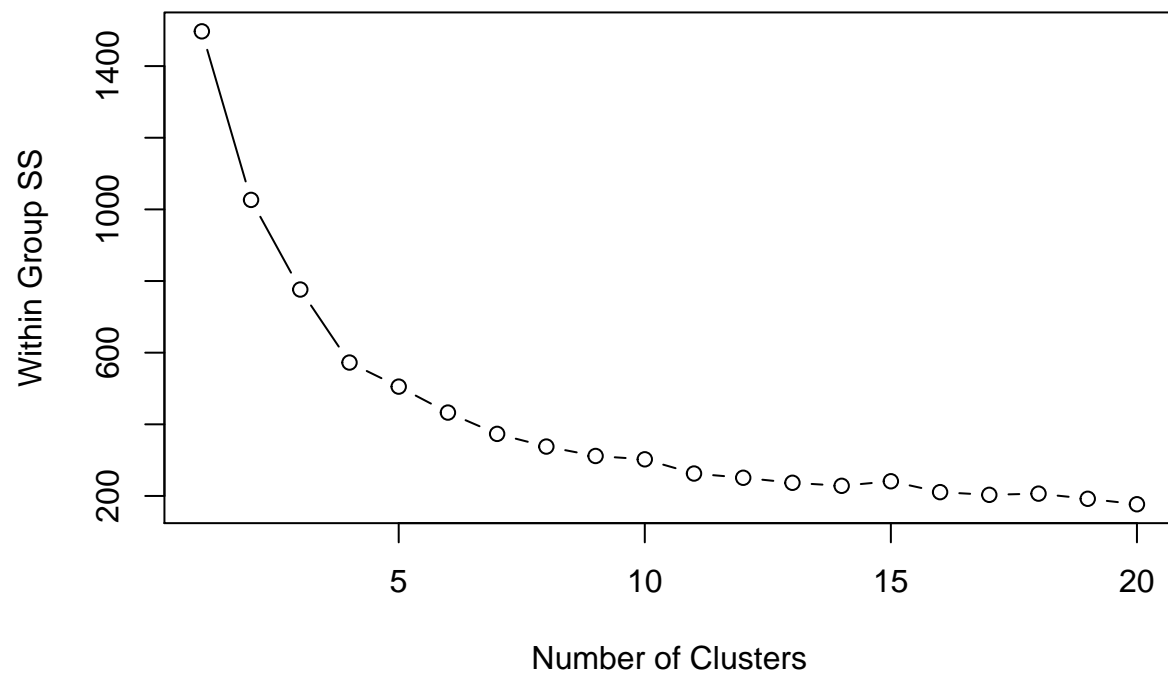
The code below shows a standard process of clustering data. In this dataset, based on Current Balance, Credit Score, Debt-to-Income Ratio variables, it is best to segment into four clusters.

```

# Normalization
bal_sc_dti_main <- heloc%>% filter(CREDIT.SCORE >= 575, DEBT.TO.INCOME <= 59, Current.Balance <= 14800)
m <- apply(bal_sc_dti_main[c("CREDIT.SCORE", "DEBT.TO.INCOME", "Current.Balance")],2,mean)
s <- apply(bal_sc_dti_main[c("CREDIT.SCORE", "DEBT.TO.INCOME", "Current.Balance")],2,sd)
bal_sc_dti_main_nor <- scale(bal_sc_dti_main[c("CREDIT.SCORE", "DEBT.TO.INCOME", "Current.Balance")],m,s)

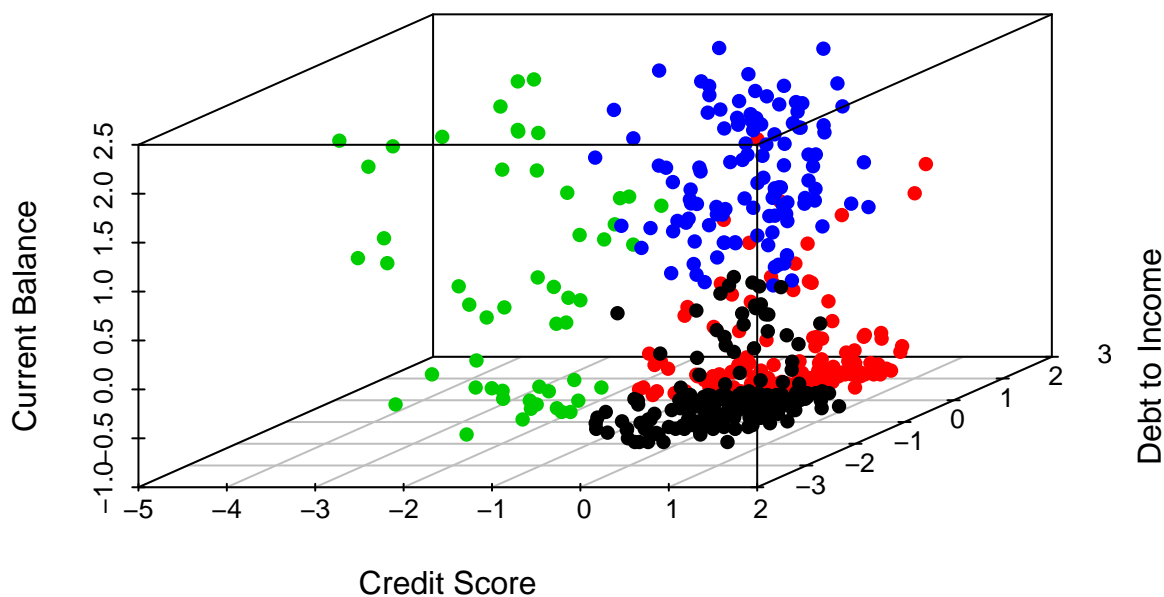
# Scree plot to decide how many clusters
wss <- (nrow(bal_sc_dti_main_nor) - 1) * sum(apply(bal_sc_dti_main_nor,2,var))
for (i in 2:20) wss[i] <- sum(kmeans(bal_sc_dti_main_nor, centers = i)$withinss)
plot(1:20, wss, type = "b", xlab = "Number of Clusters", ylab = "Within Group SS")

```



```
kc <- kmeans(bal_sc_dti_main_nor, 4)
bal_sc_dti_main <- as.data.frame(cbind(bal_sc_dti_main, Cluster = as.factor(kc$cluster)))
bal_sc_dti_main_nor <- as.data.frame(cbind(bal_sc_dti_main_nor, Cluster = as.factor(kc$cluster)))

scatterplot3d(bal_sc_dti_main_nor,
  xlab = "Credit Score",
  ylab = "Debt to Income",
  zlab = "Current Balance",
  pch = 16,
  color = colors[as.numeric(bal_sc_dti_main_nor$Cluster)])
```



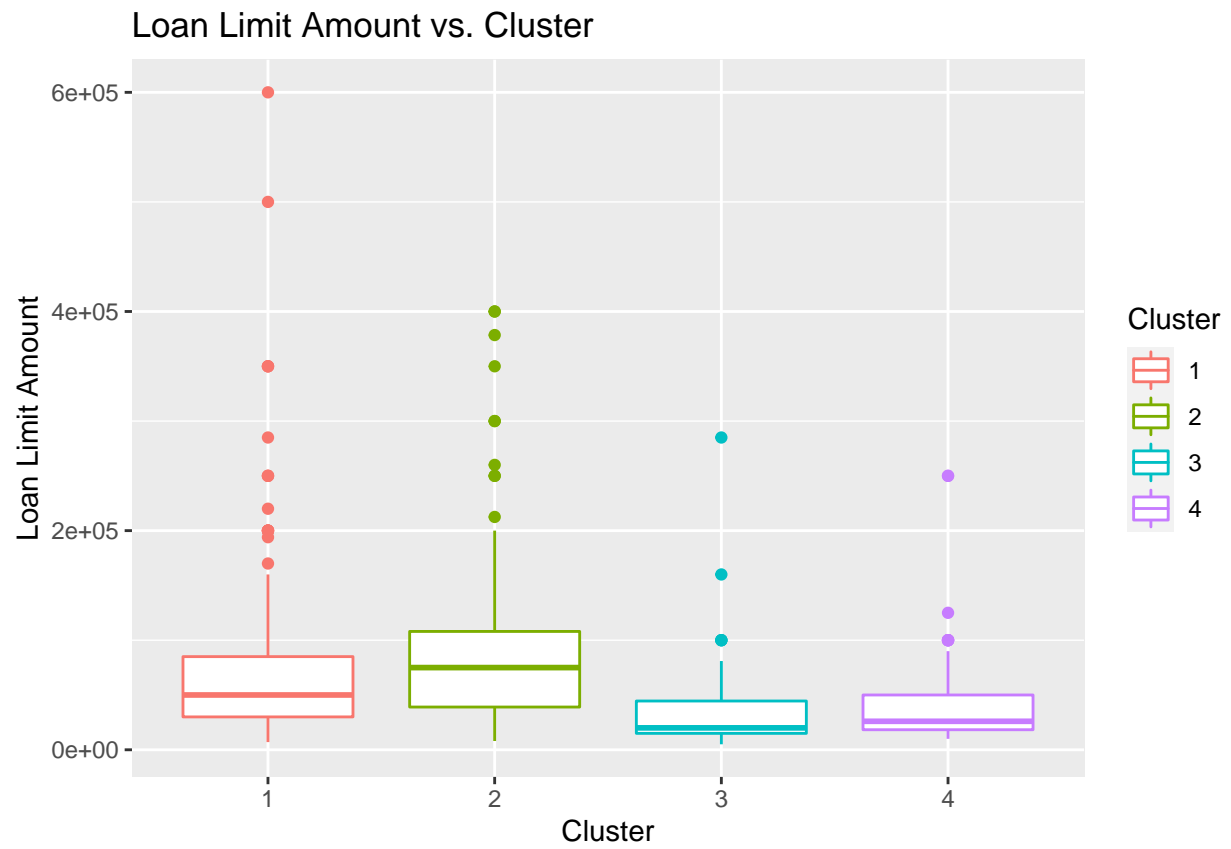
After segment all accounts in this dataset into four different category, we plotted box plots and histogram to check the distributions of Loan Limit Amount, Loan-to-Value Ratio for each segment. The point of clustering these accounts to four different clusters is that it helps to find the most similar customers. Within each customer segment, we can say that the homogeneity is very high. As for different segment, the heterogeneity will be very high. With that being said, to determine whether we have given the right customers with right quote, we need to check the Loan Limit Amount distribution and Loan-to-Value distribution. The Loan limit Amount for each segment shouldn't have a large range. On the other hand, we should see diversity in Loan-to-Ratio, that means a relatively normal distribution of Loan-to-Value Ratio under each segment.

```
p10 <-ggplot(bal_sc_dti_main, aes(x = Cluster, y = Loan.Limit.Amount, color = Cluster)) + geom_boxplot()
p11 <-ggplot(bal_sc_dti_main, aes(x = Cluster, y = ORIGINAL.L.V, color = Cluster)) + geom_boxplot() + labs(
  title = "Boxplot of Loan Limit Amount and Loan-to-Value Ratio by Cluster",
  subtitle = "The plot shows the distribution of Loan Limit Amount and Loan-to-Value Ratio for four different customer segments.",
  xlab = "Cluster",
  ylab = "Loan Limit Amount and Loan-to-Value Ratio",
  caption = "Boxplot of Loan Limit Amount and Loan-to-Value Ratio by Cluster",
  theme = theme_minimal()
)

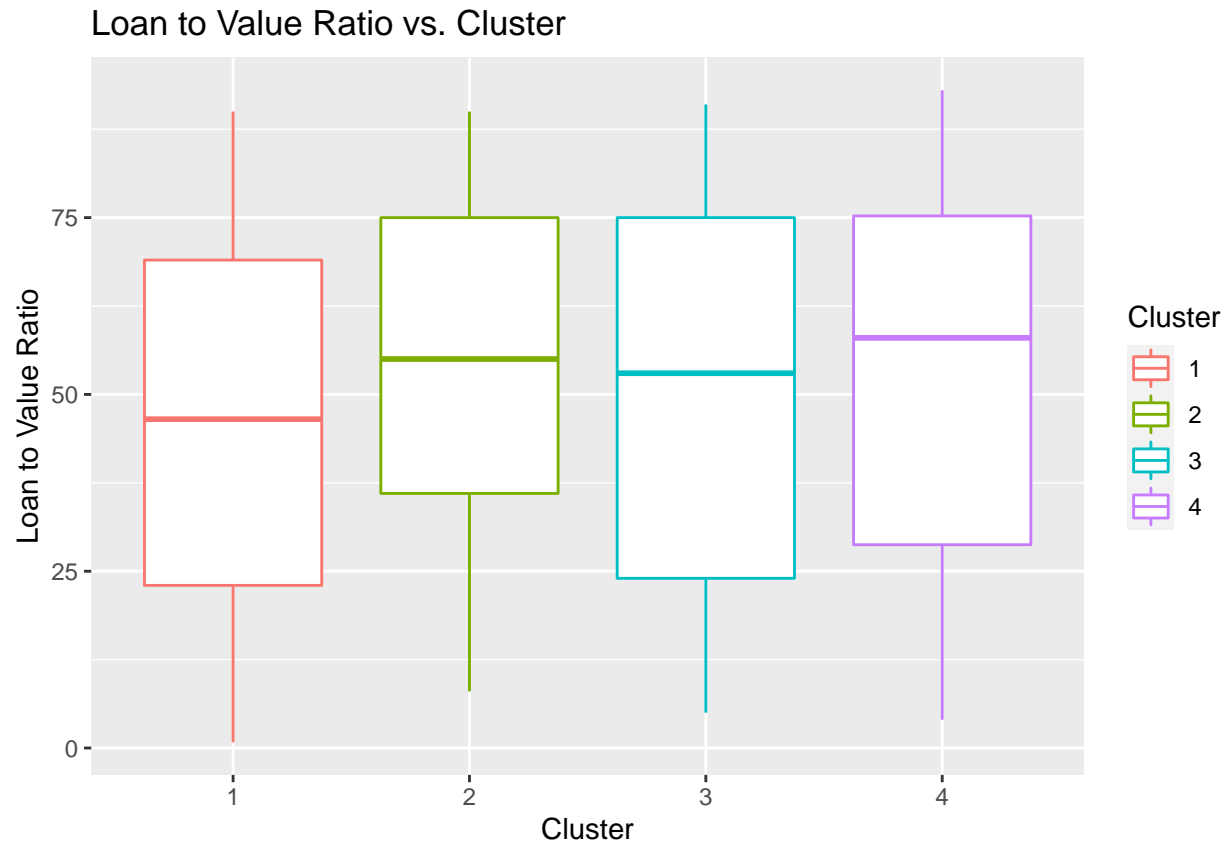
p12 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 1), aes(x = Loan.Limit.Amount)) + geom_histogram(fill = "#90EE90", alpha = 0.5)
p13 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 2), aes(x = Loan.Limit.Amount)) + geom_histogram(fill = "#6495ED", alpha = 0.5)
p14 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 3), aes(x = Loan.Limit.Amount)) + geom_histogram(fill = "#FF6347", alpha = 0.5)
p15 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 4), aes(x = Loan.Limit.Amount)) + geom_histogram(fill = "#000000", alpha = 0.5)

p16 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 1), aes(x = ORIGINAL.L.V)) + geom_histogram(fill = "#90EE90", alpha = 0.5)
p17 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 2), aes(x = ORIGINAL.L.V)) + geom_histogram(fill = "#6495ED", alpha = 0.5)
p18 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 3), aes(x = ORIGINAL.L.V)) + geom_histogram(fill = "#FF6347", alpha = 0.5)
p19 <- ggplot(bal_sc_dti_main %>% filter(Cluster == 4), aes(x = ORIGINAL.L.V)) + geom_histogram(fill = "#000000", alpha = 0.5)

p10
```



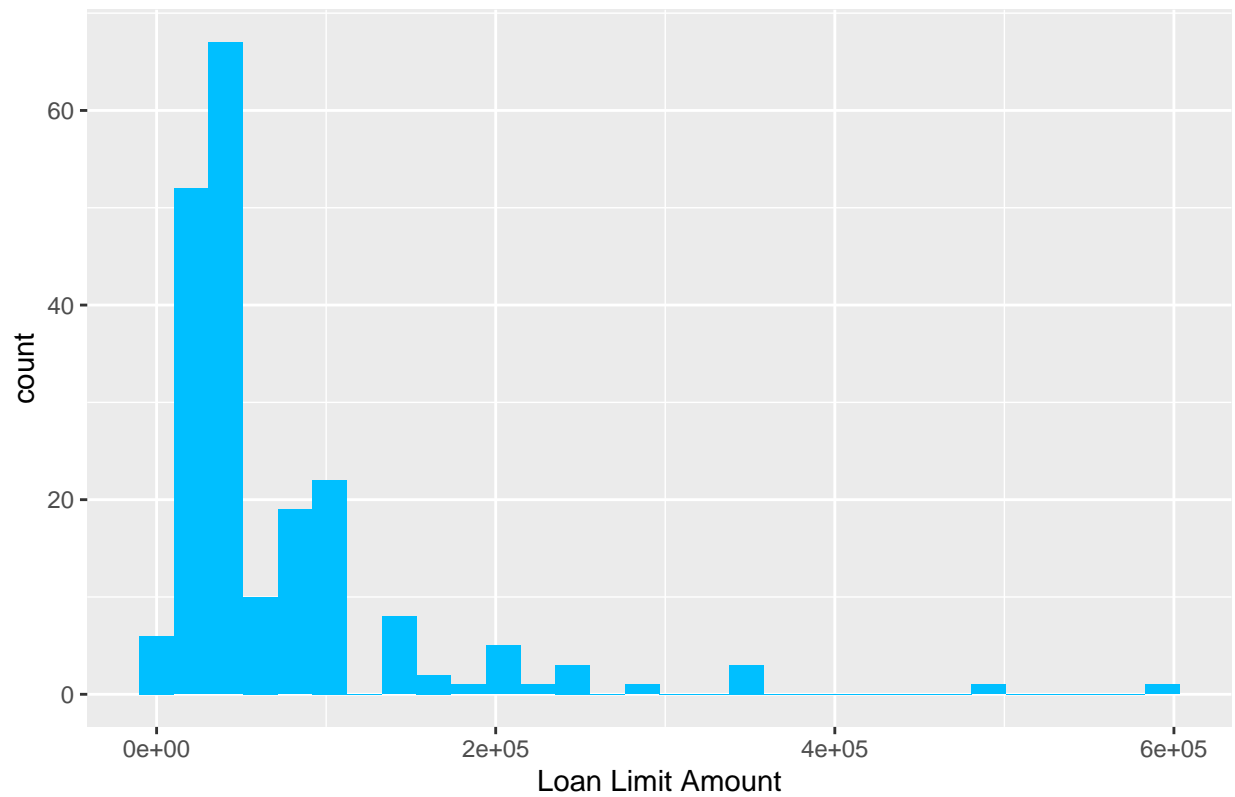
p11



p12

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

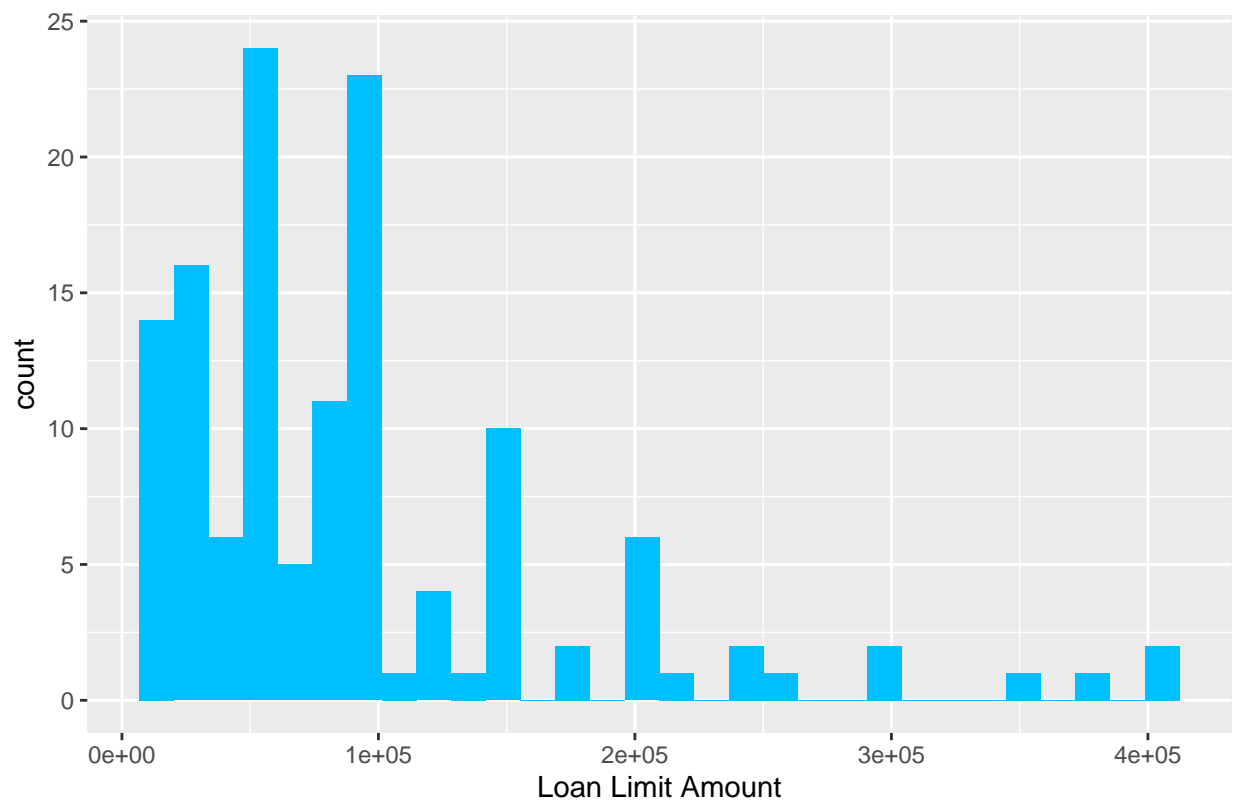
Loan Limit Amount Distribution for Cluster 1



p13

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

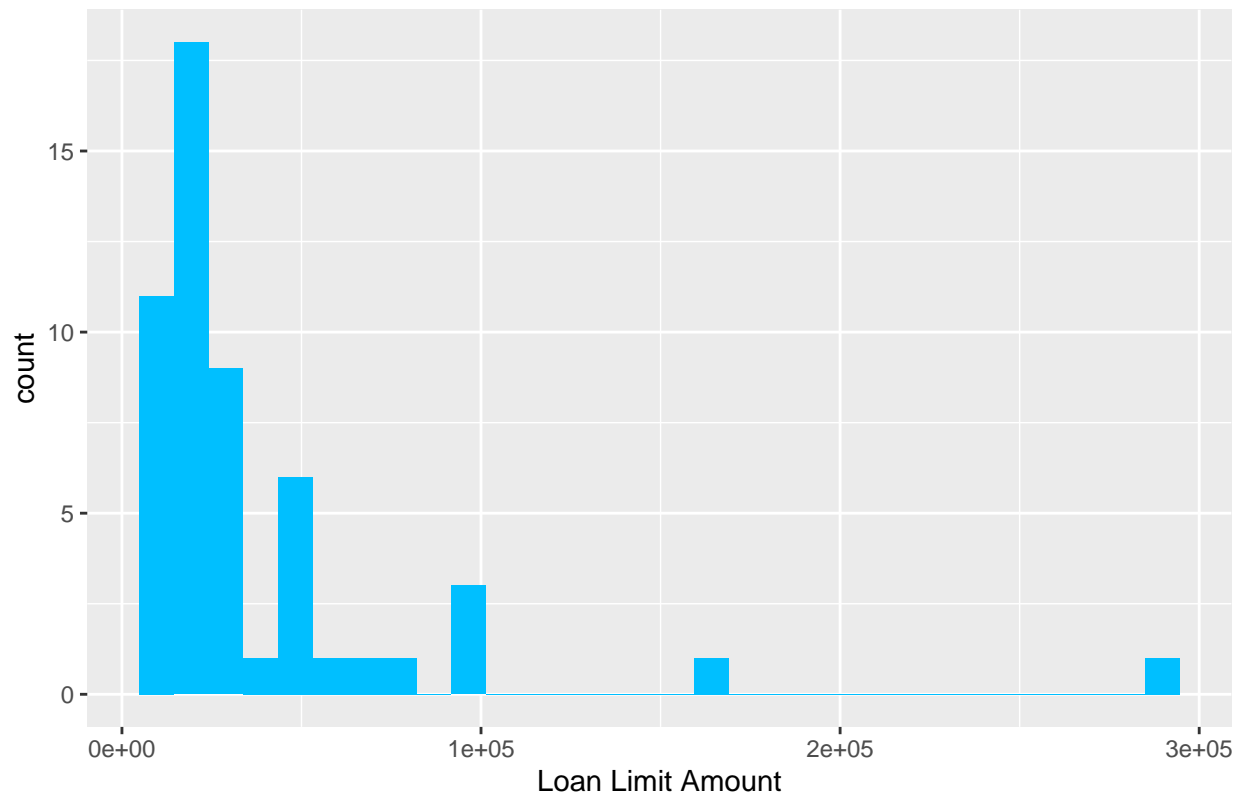
Loan Limit Amount Distribution for Cluster 2



p14

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

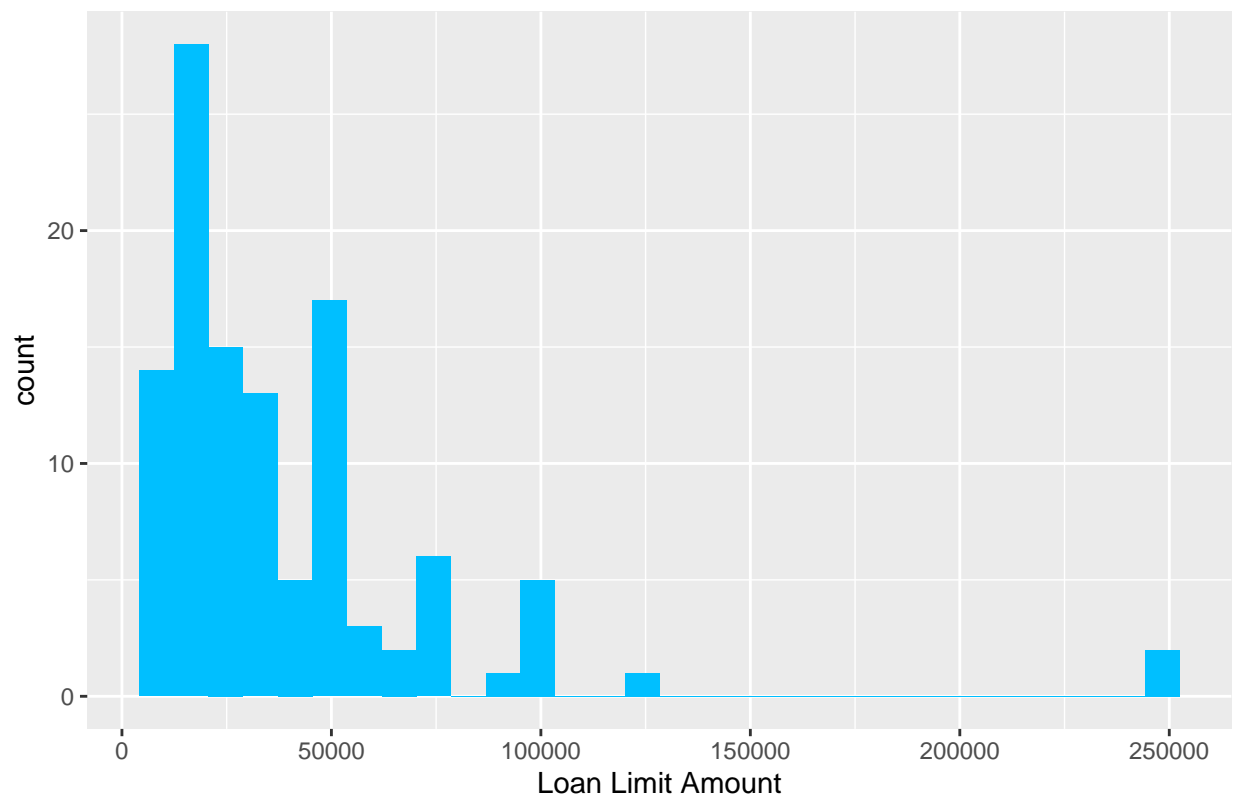
Loan Limit Amount Distribution for Cluster 3



p15

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

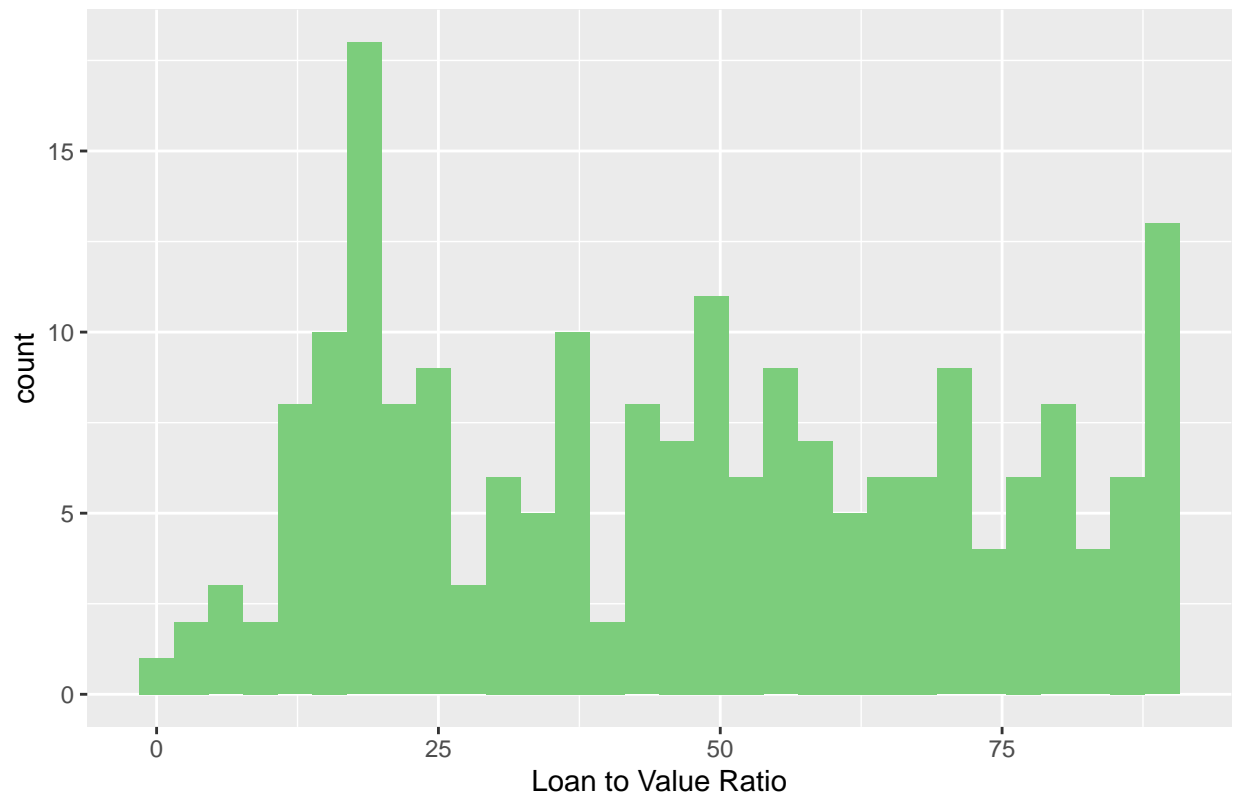
Loan Limit Amount Distribution for Cluster 4



p16

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Loan to Value Ratio Distribution for Cluster 1



p17

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Loan to Value Ratio Distribution for Cluster 2



p18

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

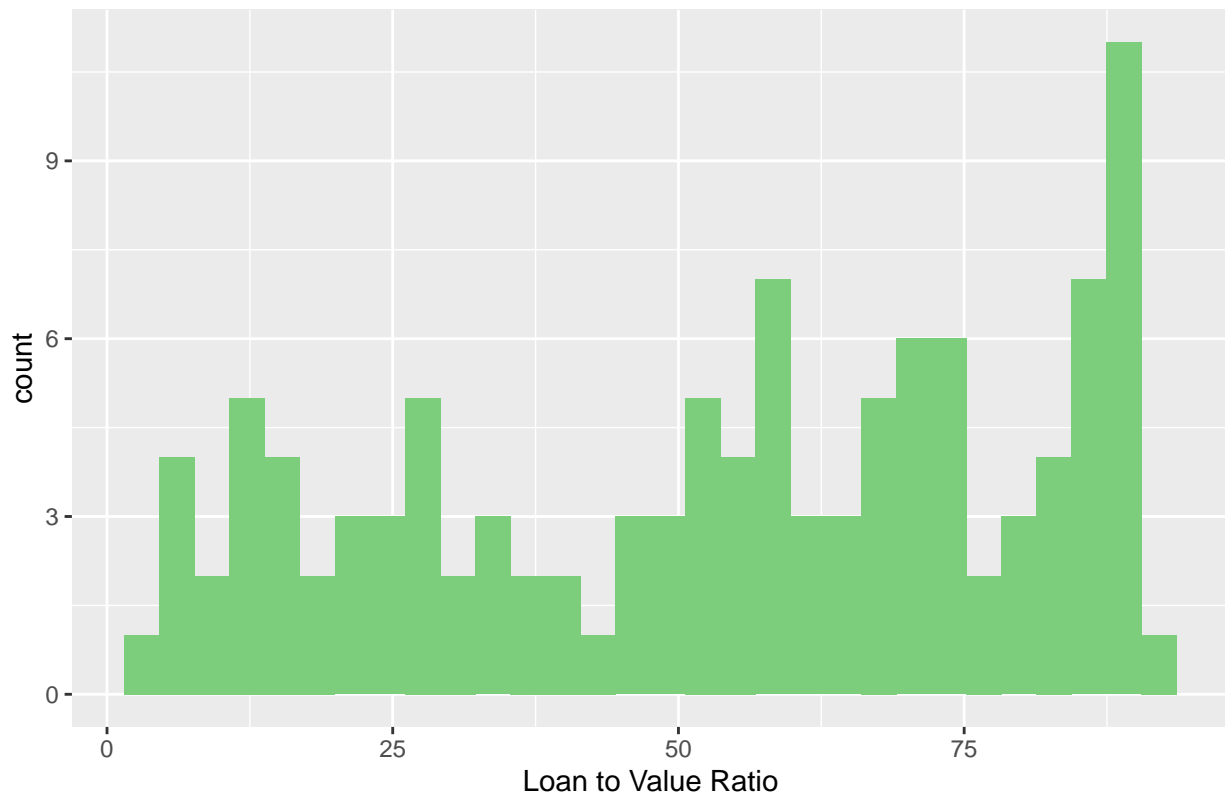
Loan to Value Ratio Distribution for Cluster 3



p19

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Loan to Value Ratio Distribution for Cluster 4



Customer Segment Profile

```
cluster1 <- bal_sc_dti_main %>% filter(Cluster == 1)
cluster2 <- bal_sc_dti_main %>% filter(Cluster == 2)
cluster3 <- bal_sc_dti_main %>% filter(Cluster == 3)
cluster4 <- bal_sc_dti_main %>% filter(Cluster == 4)
#summary(cluster1)
#summary(cluster2)
#summary(cluster3)
summary(cluster4)
```

```
## Product.Type Account.Number Current.Balance Loan.Limit.Amount
## HELOC:112 Min. : 4.0 Min. : 5278 Min. : 10000
## 1st Qu.: 266.2 1st Qu.: 8472 1st Qu.: 18300
## Median : 641.0 Median :10065 Median : 26000
## Mean : 581.7 Mean :10464 Mean : 39791
## 3rd Qu.: 886.0 3rd Qu.:12765 3rd Qu.: 50000
## Max. :1037.0 Max. :14764 Max. :250000
##
## Int.Rate Date.Org.Opened Date.Closed Term
## Min. :3.000 11/30/18: 3 Mode:logical Min. : 24.0
## 1st Qu.:4.000 10/23/15: 2 NA's:112 1st Qu.:120.0
## Median :4.000 6/26/15 : 2 Median :120.0
## Mean :4.012 6/7/19 : 2 Mean :118.6
## 3rd Qu.:4.000 7/15/13 : 2 3rd Qu.:120.0
## Max. :5.380 1/18/12 : 1 Max. :120.0
```

```

##          (Other) :100
## CREDIT.SCORE  ORIGINAL.L.V  DEBT.TO.INCOME  OCCUPANCY.CODE.Cd.1
## Min.      :711.0  Min.      : 4.00  Min.      : 0.00  Min.      :1
## 1st Qu.    :761.5  1st Qu. :28.75  1st Qu. :19.08  1st Qu. :1
## Median    :780.0  Median :58.00  Median :27.00  Median :1
## Mean      :776.9  Mean   :53.52  Mean   :26.75  Mean   :1
## 3rd Qu.    :796.2  3rd Qu. :75.24  3rd Qu. :33.00  3rd Qu. :1
## Max.      :820.0  Max.    :93.00  Max.    :49.00  Max.    :1
##
## PROPERTY.TYPE      Score.Range      DTI.Range
## Min.      :0      1 >=740      :102      1 >50.01      : 0
## 1st Qu.    :0      2 720-739 : 9      2 43.01-50.00 : 5
## Median    :0      3 680-719 : 1      3 40.01-43.00 : 8
## Mean      :0      4 640-679 : 0      4 35.01 - 40.00:10
## 3rd Qu.    :0      5 600-639 : 0      5 30.01 - 35.00:22
## Max.      :0      6 <= 600 : 0      6 <= 30      :65
##          7 No Score: 0      No DTI      : 2
##          Orig.LTV.Range      Yield.Range
## 2 90% - 99.99% LTV : 2      2 3.0% - 3.99% : 1
## 3 80% - 89.99% LTV :22      3 >= 4.% - 4.25% :109
## 4 >= 70% - 79.99% LTV:16      4 >= 4.25% - 4.49%: 0
## 5 >= 60% - 69.99% LTV:11      7 >= 5% - 5.99% : 2
## 6 <60% LTV :61      8 >=6% LTV : 0
## No LTV Calculated : 0
##
## Delinquency.Range  Days.Late      Cluster
## 1 Current :107      Min. :0.0000      1: 0
## 2 1 to 29 DPD: 5      1st Qu.:0.0000      2: 0
##          Median :0.0000      3: 0
##          Mean   :0.2768      4:112
##          3rd Qu.:0.0000
##          Max.   :7.0000
##

```

To utilize this segmentation, we can try to profile the. - Customer Segment 1: A mean Current Balance of 10606, a mean Credit Score of 775, a mean DTI of 28.81. The Mean Loan Limit Amount id 44200, the LTV Ratio range from 4% to 93%. - Customer Segment 2: A mean Current Balance of 5534, a mean Credit Score of 679, a mean DTI of 23.13. The Mean Loan Limit Amount id 36723, the LTV Ratio range from 5% to 91%. - Customer Segment 3: A mean Current Balance of 978, a mean Credit Score of 789.7, a mean DTI of 16.16. The Mean Loan Limit Amount id 69116, the LTV Ratio range from 0.81% to 90%. - Customer Segment 4: A mean Current Balance of 635.6, a mean Credit Score of 776.8, a mean DTI of 35.30. The Mean Loan Limit Amount id 95040, the LTV Ratio range from 8% to 90%.

Conclusion

The accounts that are having abnormal quote are account 649, 1041, 428, 35, 498, 987, 9, 584, 1041.