Eexploratory Data Analysis

Enqun Wang (EW), Yiyan Zhou (YZ)

April 25, 2016

Preprocess data

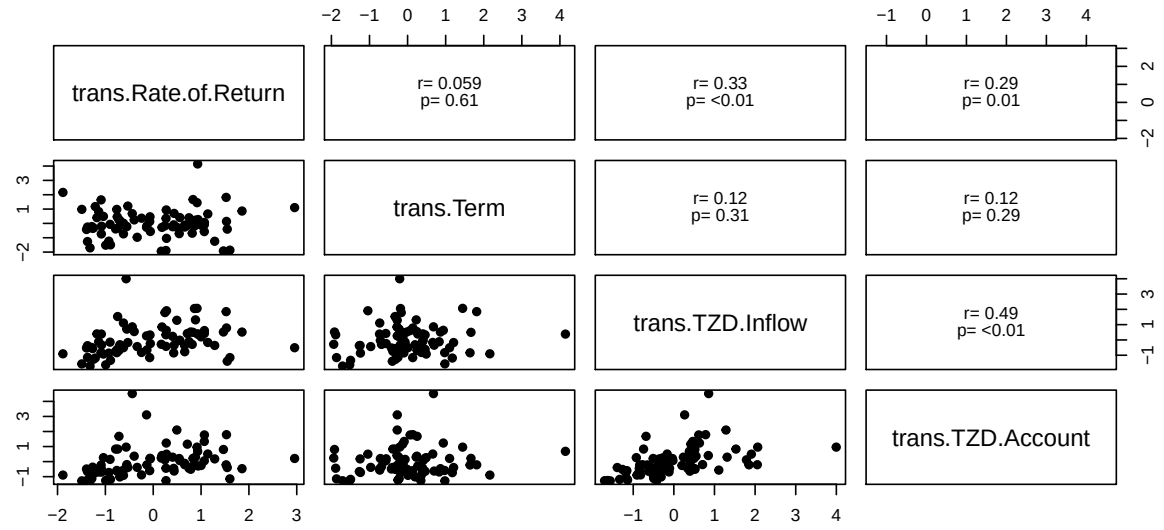
- Based on cause and effect relationship, we divide the variales into four causes: Product Factor, Promotion Factor, Platform Factor, and Market Factor. According to voice of customer (VOC), we would analyze the influence of these factors independently, for that each one represents a different aspect.

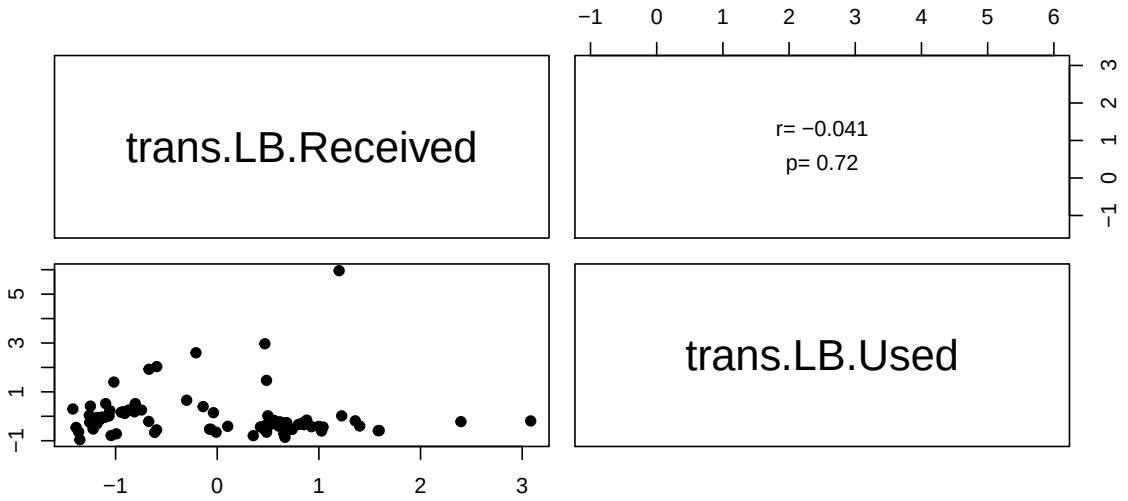
Product Factor	Promotion Factor	Platform Factor	Market Factor
Rate of Return	LB Received	Balance	R.007
Term	LB Used	Capital Inflow	Inerbank Rate
TZD Account			SHIBOR
TZD Inflow			SHA
			GEM

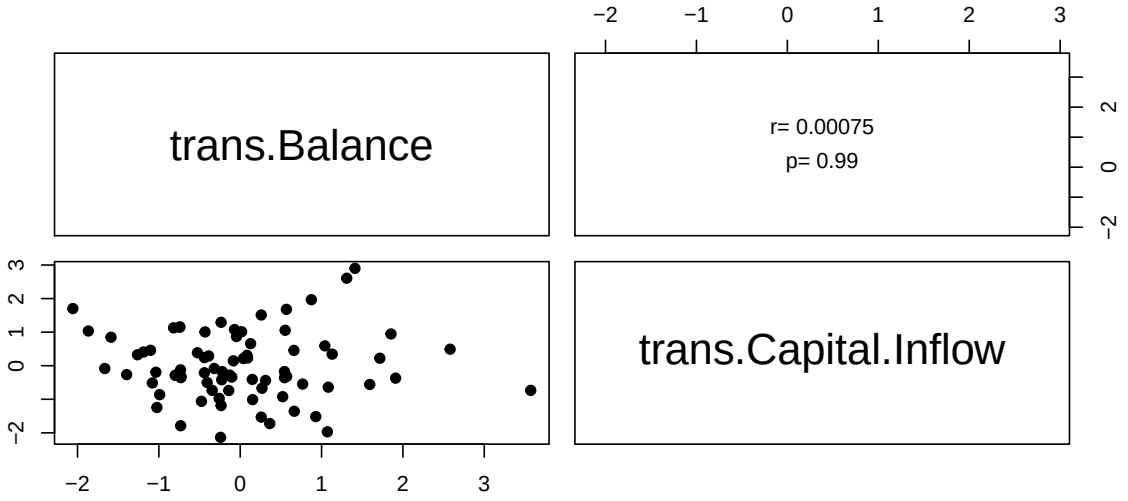
- In order to remove effects of different units, we centered and rescaled the data.

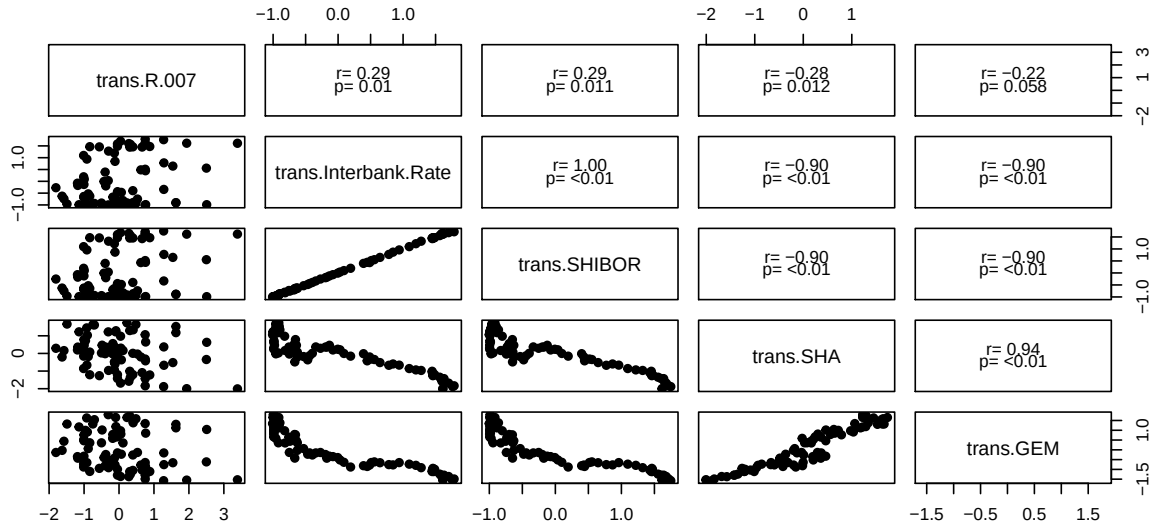
Detect dependent variables

- According to the correlation plot above, we find that `interbank.Rate` , `SHIBOR`, `SHA`, and `GEM` are highly correlated, and that `TZD.Inflow` and `TZD.Account` are highly correlated. So we consider if we could remove some of them.
- Based on the voice of costumers (VOC), we decided to remove `interbank.Rate`, which can be represented by `SHIBOR`; remove `GEM`, which can be represented by `SHA`; and remove `TZD.Inflow`, which can be reflected from `TZD.Account`.



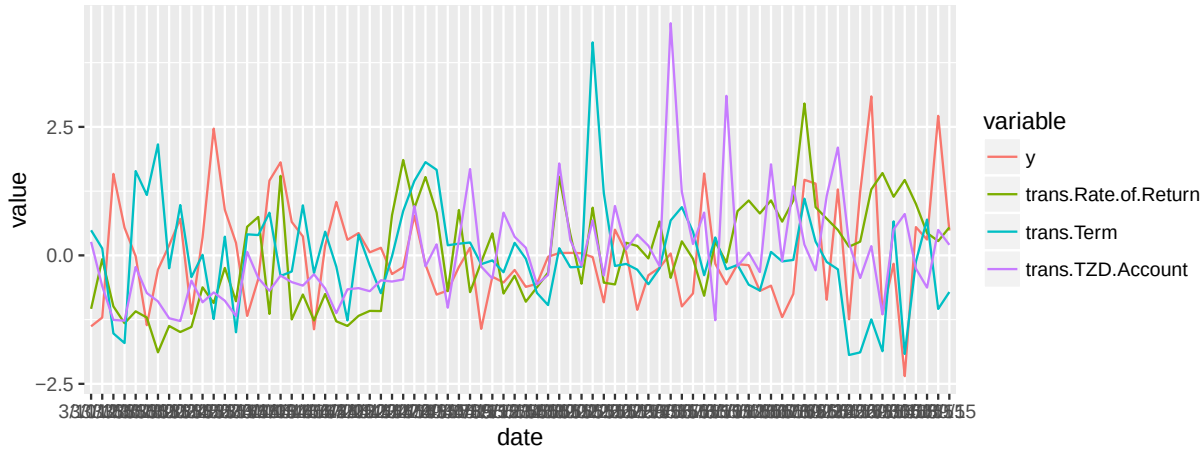

trans.LB.Received
r= −0.041
p= 0.72
trans.LB.Used

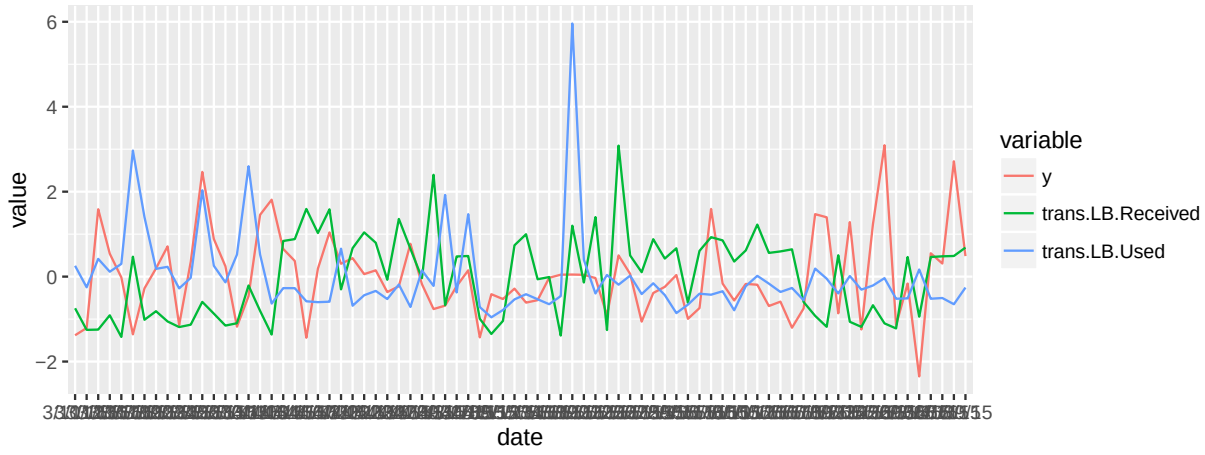

trans.Balance
r= 0.00075
p= 0.99
trans.Capital.Inflow

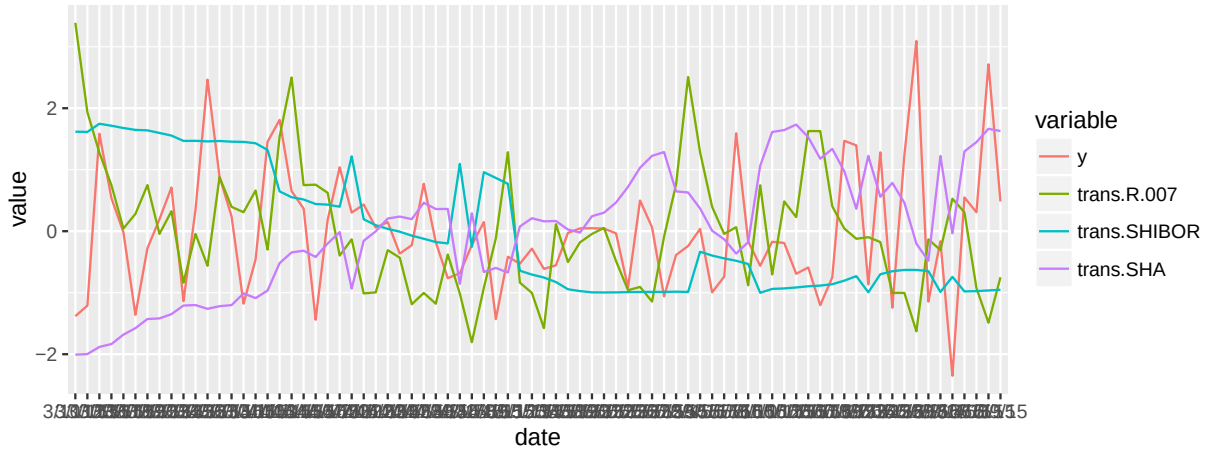


Plot multiple time series

- According to the plots above and VOC, we would remove the variable `SHIBOR` and `SHA`.

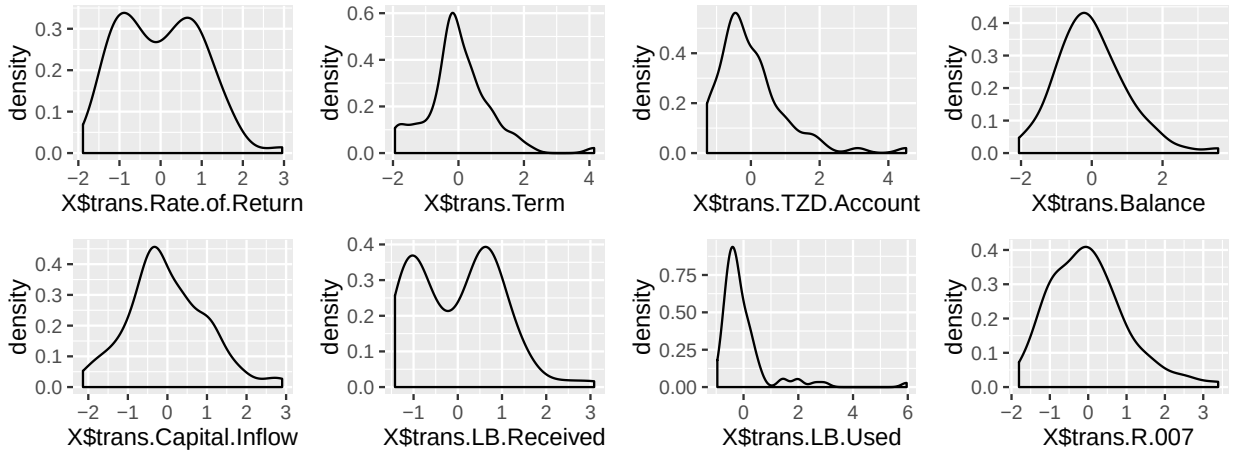






Density plots

- To explore the distributions of the variables, we ploted density plots as follows. It indicates that `Rate of Retrun` and `LB Recevied` are not normally distributed, while others are basicly normal.



First selection of variables

According to the exploratory data analysis, we decide to first elect variables as follows.

Product Factor	Promotion Factor	Platform Factor	Market Factor
1. Rate of Return	1. LB Received	1. Balance	1. R.007
2. Term	2. LB Used	2. Capital Inflow	
3. TZD Account			