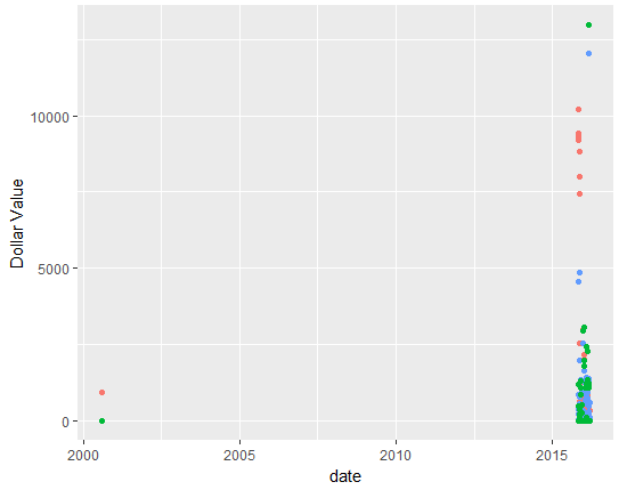
Solutions

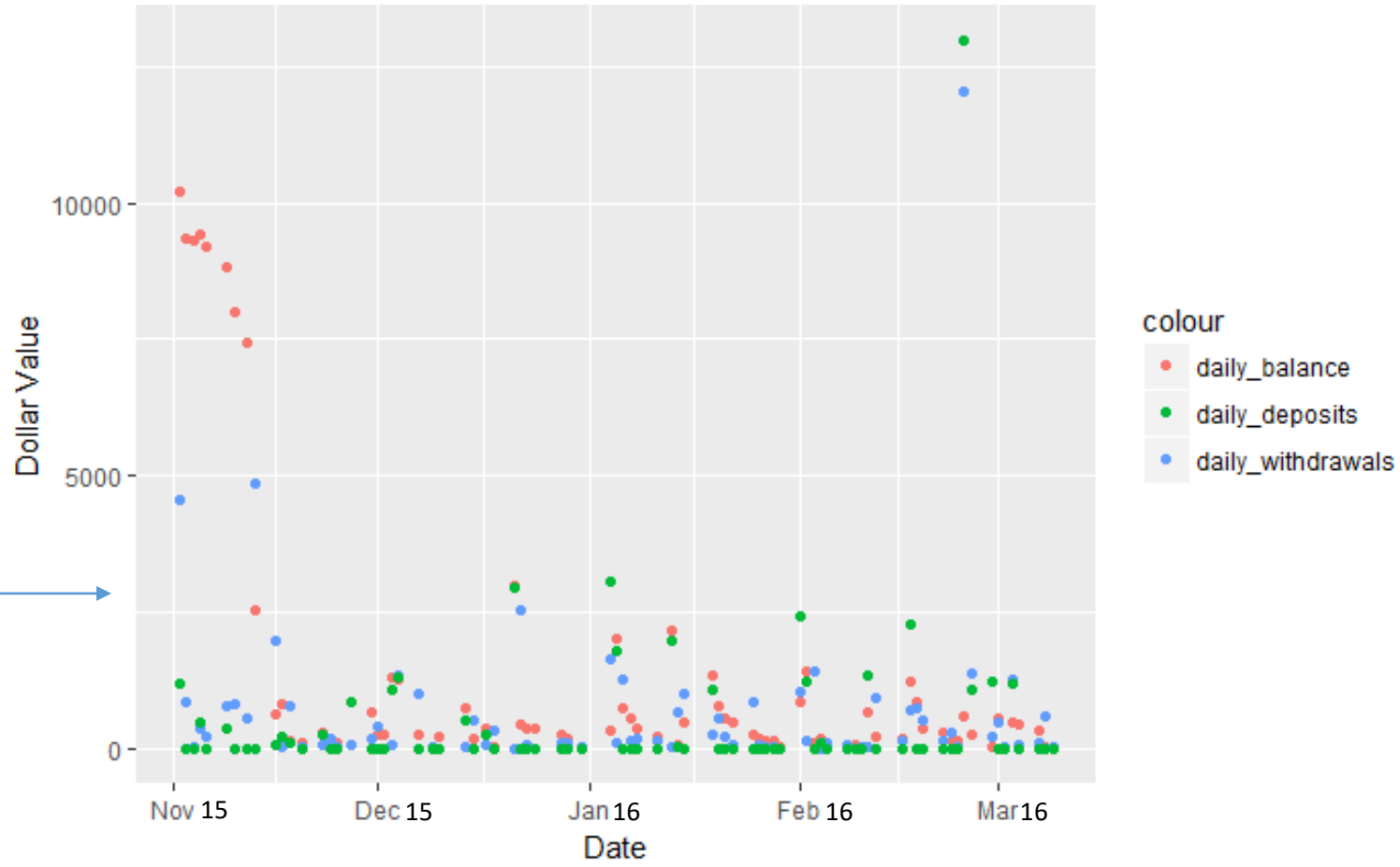
By

Neha Bora

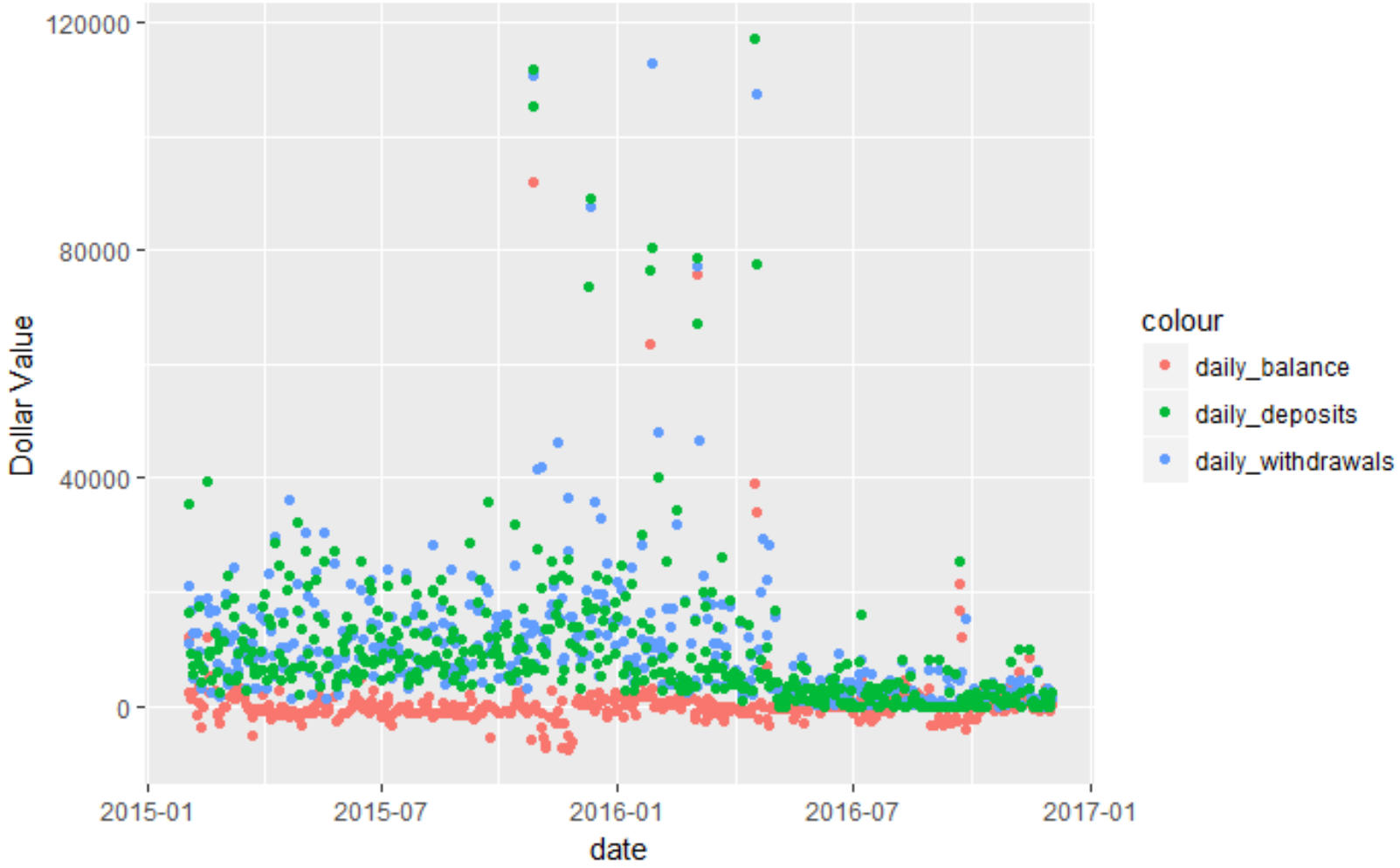
Q2(i) userid: 219154, bank_account_id: 4614 - plot the withdrawals, deposits and daily balance as a daily time series



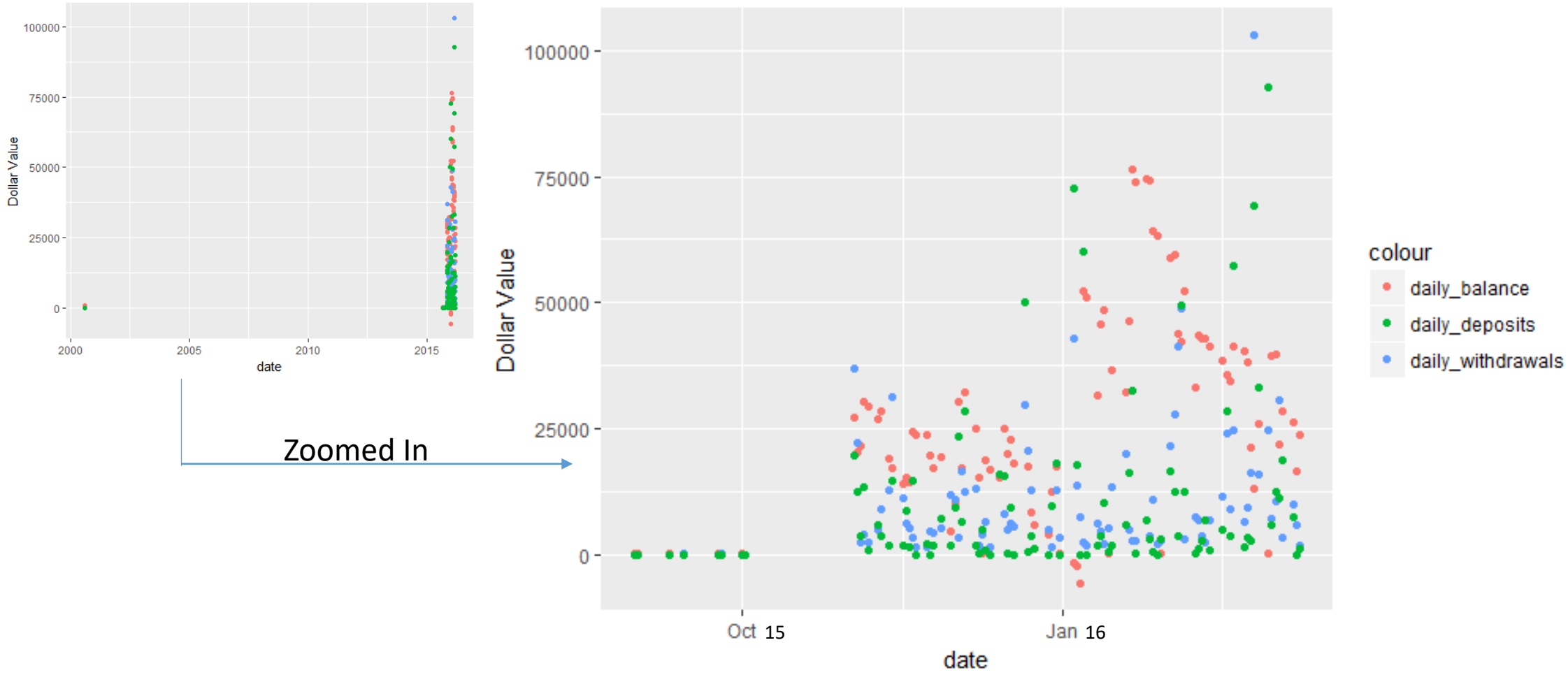
Zoomed In



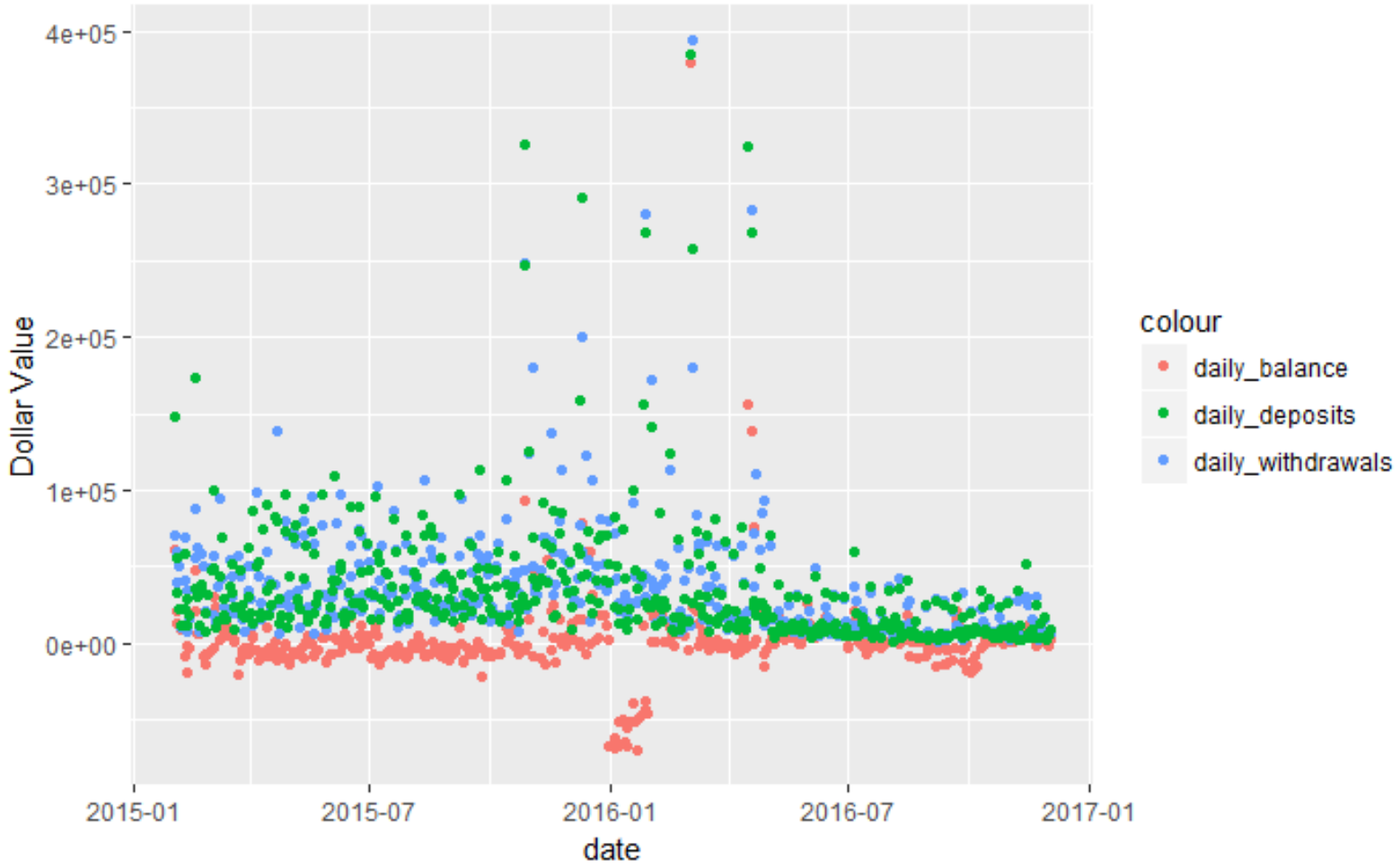
Q2(ii) userid: 217867, bank_account_id : 5420 - plot the withdrawals, deposits and daily balance as a daily time series



3(i) userid: 219154 - plot the withdrawals, deposits and daily balance as a daily time series (aggregate over all their bank accounts)



3(ii) userid: 217867 - plot the withdrawals, deposits and daily balance as a daily time series (aggregate over all their bank accounts)



4. Subcategories based on transaction method

1. ach - Automated Clearing House : Bank to Bank transfers within US
2. cc - Credit/ Debit Card
3. direct – Payment made by depositing/withdrawing cash or check
4. Latefees – Overdraft charges
5. online_transfers - ONLINE or INTERENT transfers
6. other – payment by other methods
7. paymenttechs - payment using third party services like Heartland Payment Systems
8. Wallet – Payment by VENMO/ PAYPAL/ STRIPE similar apps

5. Other cash flow variables can you hypothesize that could be relevant for predicting customer default?

1. Debit/ Credit ratio

Less than one means lesser deposits than withdrawals which could be used to identify defaulters

2. Percentage of total withdrawals made for loans and insurance w.r.t total deposits or withdrawals

This would give us an idea of what percent of a customer’s income is spent on loans and thus will help in predicting customer default.

4. Ideas to classify transactions into subcategories

1. Use classification based of expenditure : Travel /Groceries/Shopping/ Services etc.
2. We could use k-means on amount to cluster transactions