Analysis of Bank Marketing Data

Deposit Analysis

Mohammed Hasnat
Course: CS544
Boston University
Date: December 14 2019

Data Source

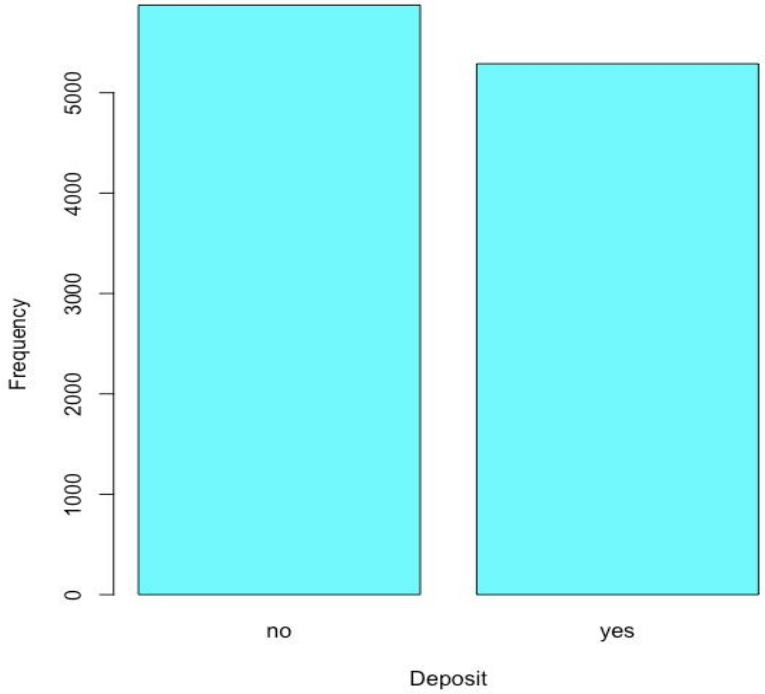
- Kaggle: https://www.kaggle.com/janiobachmann/bank-marketing-dataset
- Bank Marketing Data Set
- 17 attributes, 11162 rows of data
- Columns used: deposit, balance, martial, age, education, duration(talk duration)

Wrangling

- Made sure for no null values on the data set
- Used Filtered data for deposit value “yes” on most analysis
- Used Filtered data for age < 60
- Used Filtered data for balance < 5000

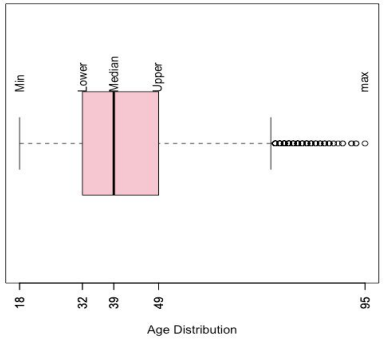
Deposit Analysis

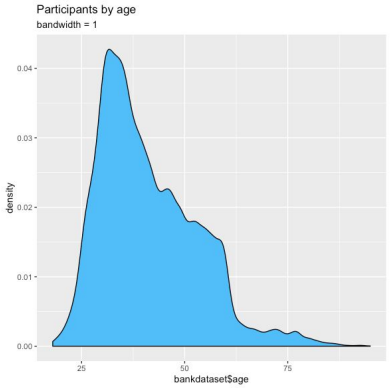
- Barplot indicates more people 5289 people deposited
- 5873 did not deposited
- The percentage of people deposited and not deposited are almost 50% respectively



Deposit and Age Analysis

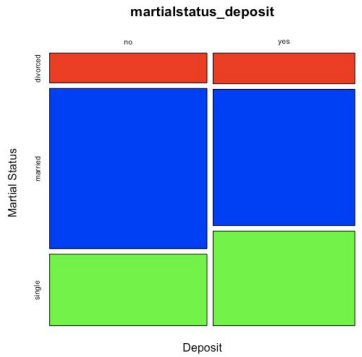
- Bank marketing data set consist of people mainly between age 32 to 49
- The max age group is 95 and min is 18
- The density plot indicates larger portion of the age lies between 32 to 49

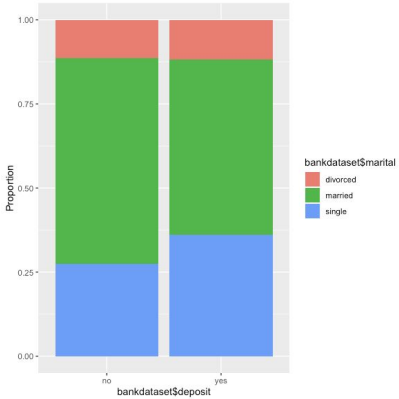




Deposit and Marital Status

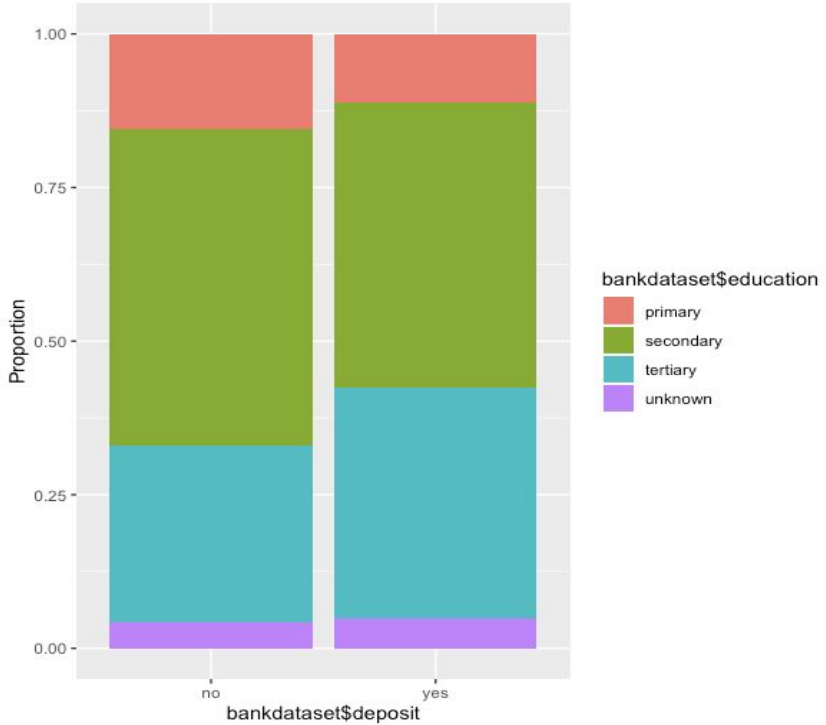
- Among people who deposited, martial has higher proportion
- Among people who deposited and did not deposit, single has higher proportion of depositing
- Divorced couples in general are less likely to deposit





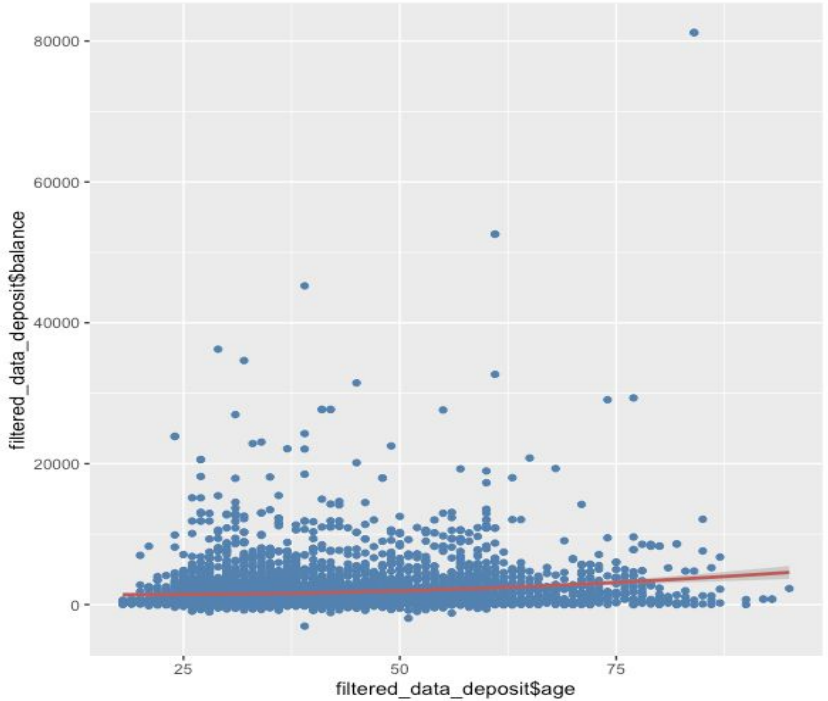
Deposit and Education

- Among people who deposited, people are mostly holding a secondary and tertiary degree
- Among people who deposited, people who are holding primary degree are less



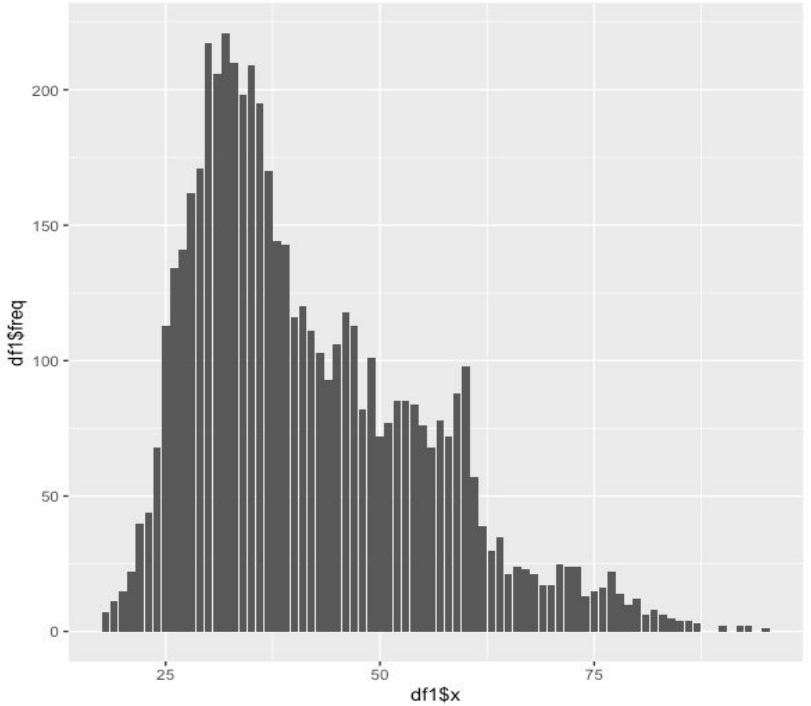
Deposit “Yes”, Balance and age analysis

- Among people who deposited, there are little to no correlation between balance and age
- The trend line is flat



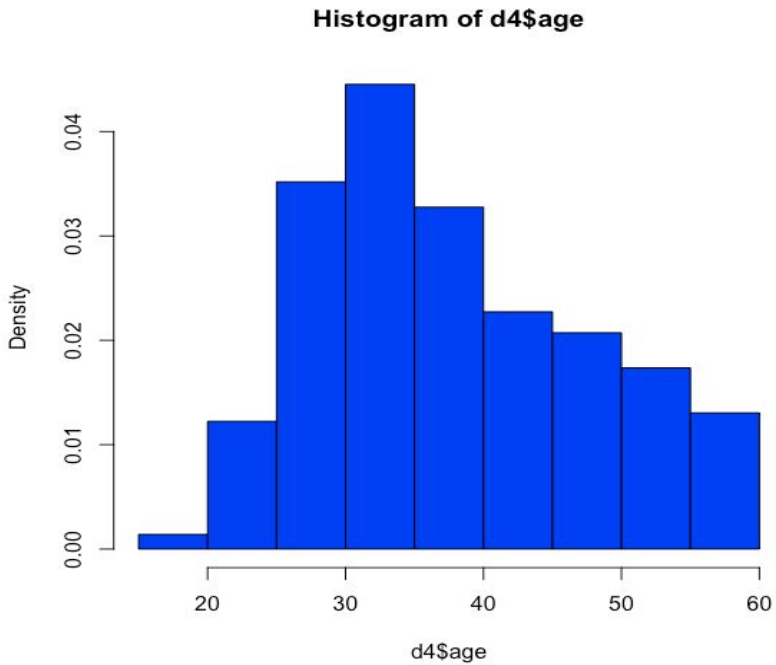
Deposit “Yes”, age data distribution analysis

- People with age 32 has the highest frequency of 221 out of 5289 people who deposited
- The proportion of age 32 depositing is 4.17%
- The proportion of age 30 depositing is 4.10%
- The proportion of age 33 depositing is 3.97%



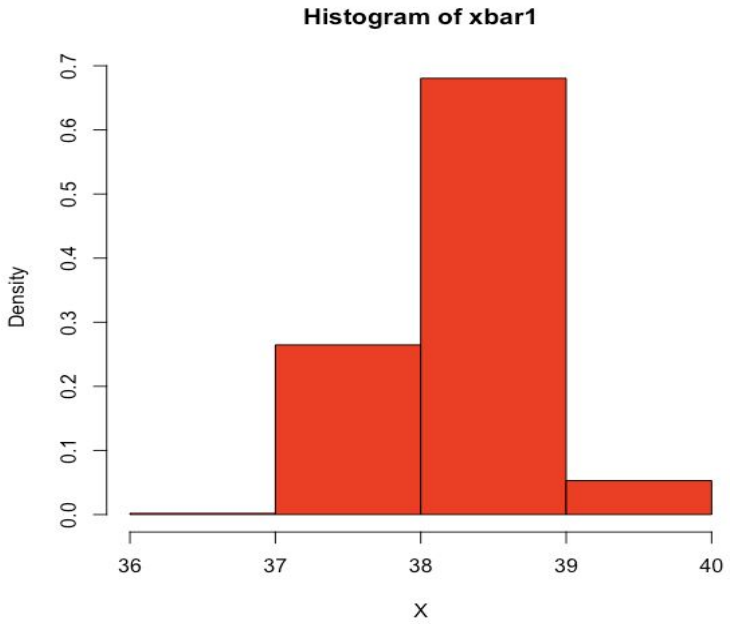
Central Limit Theorem | Deposited “yes” and age < 60

- Dropped value over age 60 for people who deposited
- The mean age is 38 for people who depositing
- SD is extremely high



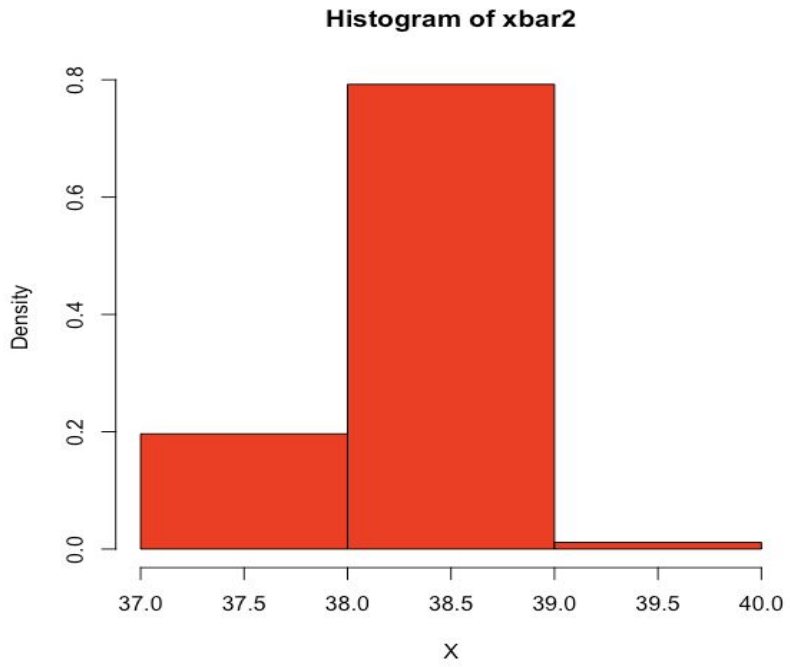
Central Limit Theorem | Deposited “yes” and age < 60

- Sample 500
- Mean age is approximate 38
- SD is lower



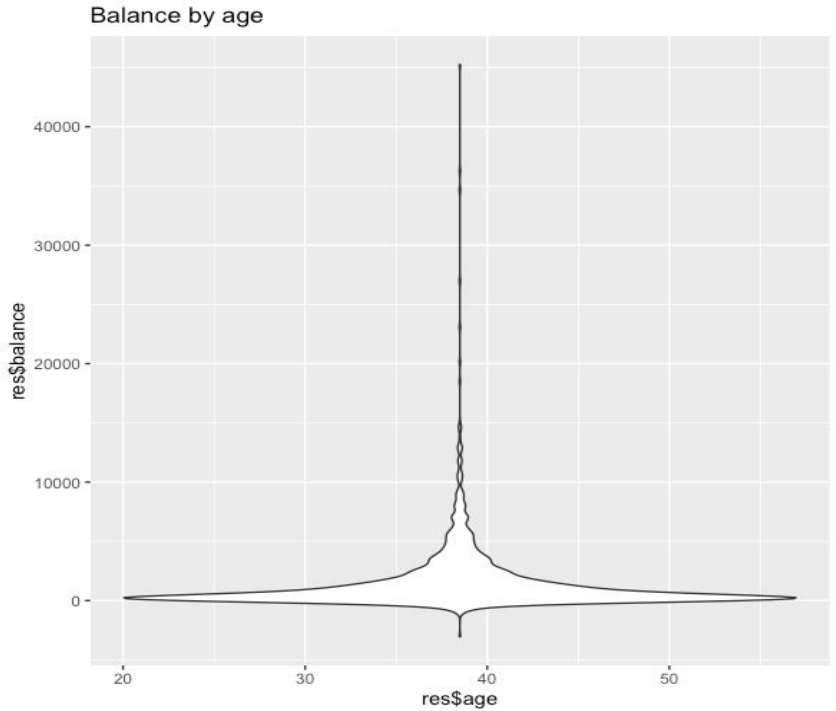
Central Limit Theorem | Deposited “yes” and age < 60

- Sample 1000
- Mean age is 38
- SD is lower than that of the sample size 500



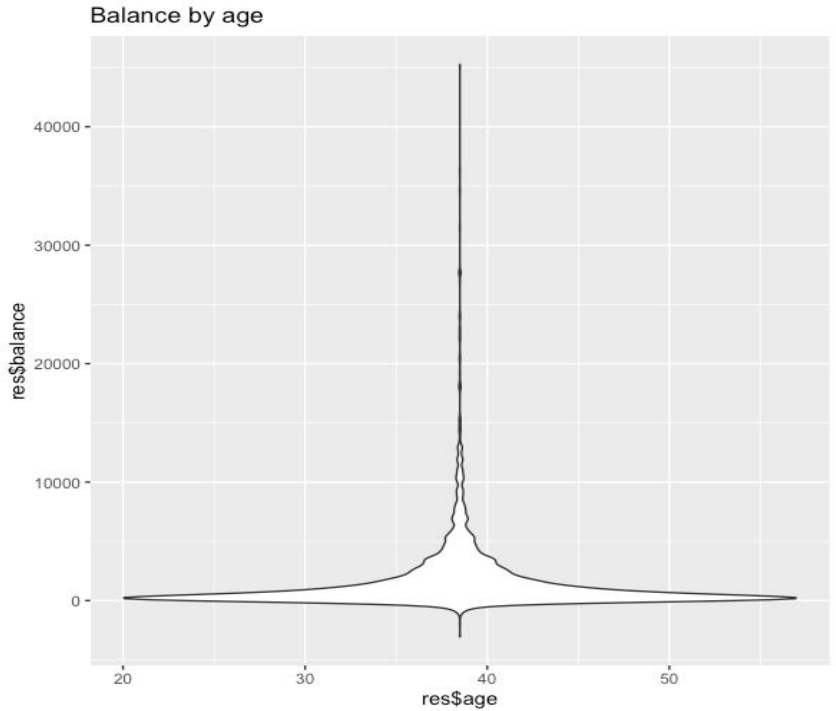
Simple Random Sampling | Deposit and Balance

- Age < 60
- Have deposit
- Mean Balance: $1770.02
- Balances are highly frequent between age 30-50 given they have deposit



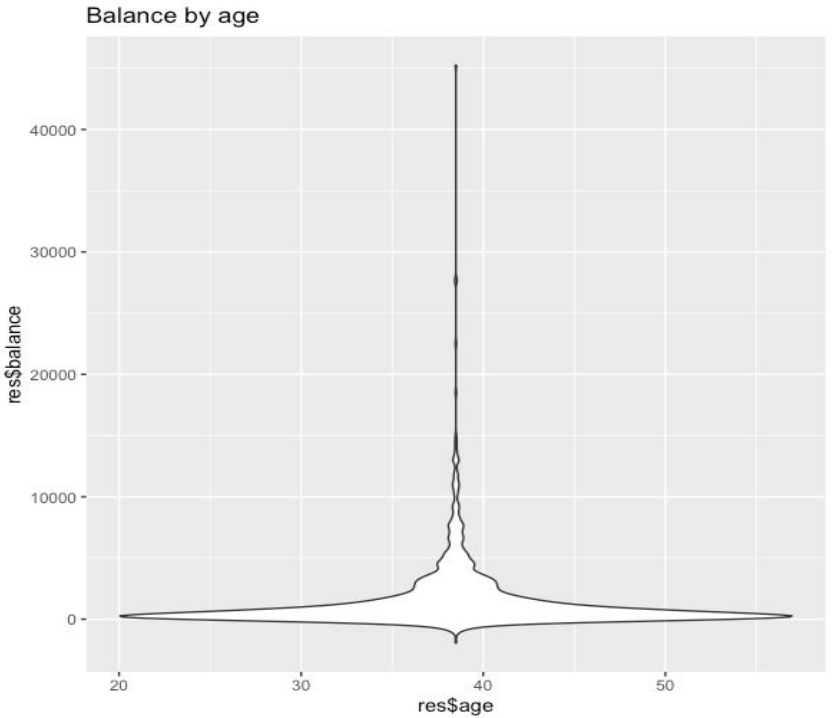
Systematic Sampling | Deposit and Balance

- Age < 60
- Have deposit
- Mean Balance: $1675.181



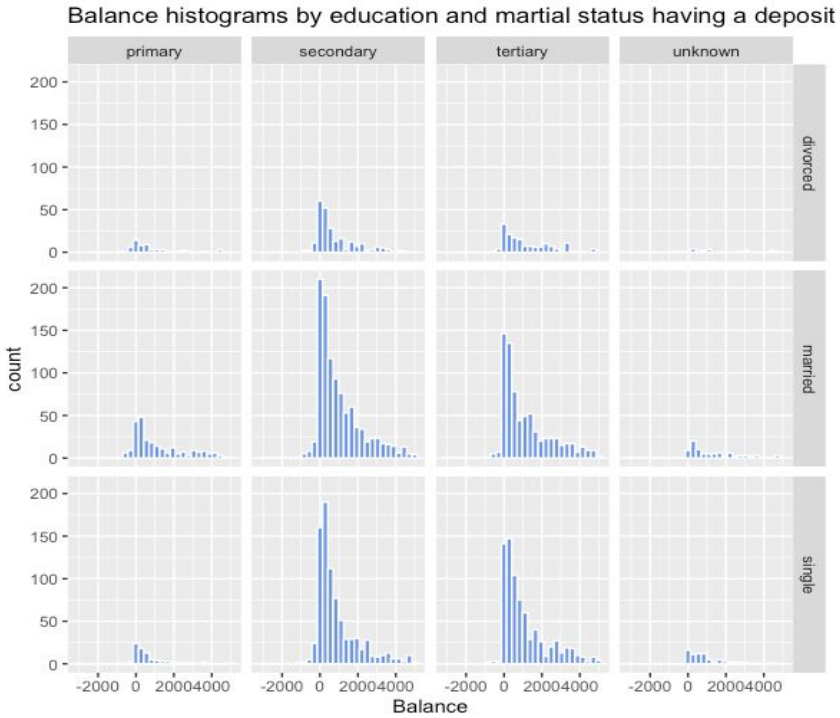
Systematic with Inclusion Probability | Deposit and balance

- Age < 60
- Have deposit
- Mean Balance: $1574.21
- Sample Violin Plot shape for three sampling types are almost same



Frequency Analysis | Deposit “yes”

- Age < 60
- Balance < 5000
- Deposit “yes”
- Single and Married people who have secondary degree and tertiary degree have more likely to have balance than people with primary degree
- Divorced people in general have less balance



Conclusion

1. Age between 30-40 should be targeted for marketing campaign because these age group deposit the most
2. For bigger marketing campaign, age 30-50 should be included
3. Married/Single with secondary and tertiary degree are more likely to have variation of balance with deposit
4. People in general with balance $1500-$1800 under the age of 60 are likely to have a deposit