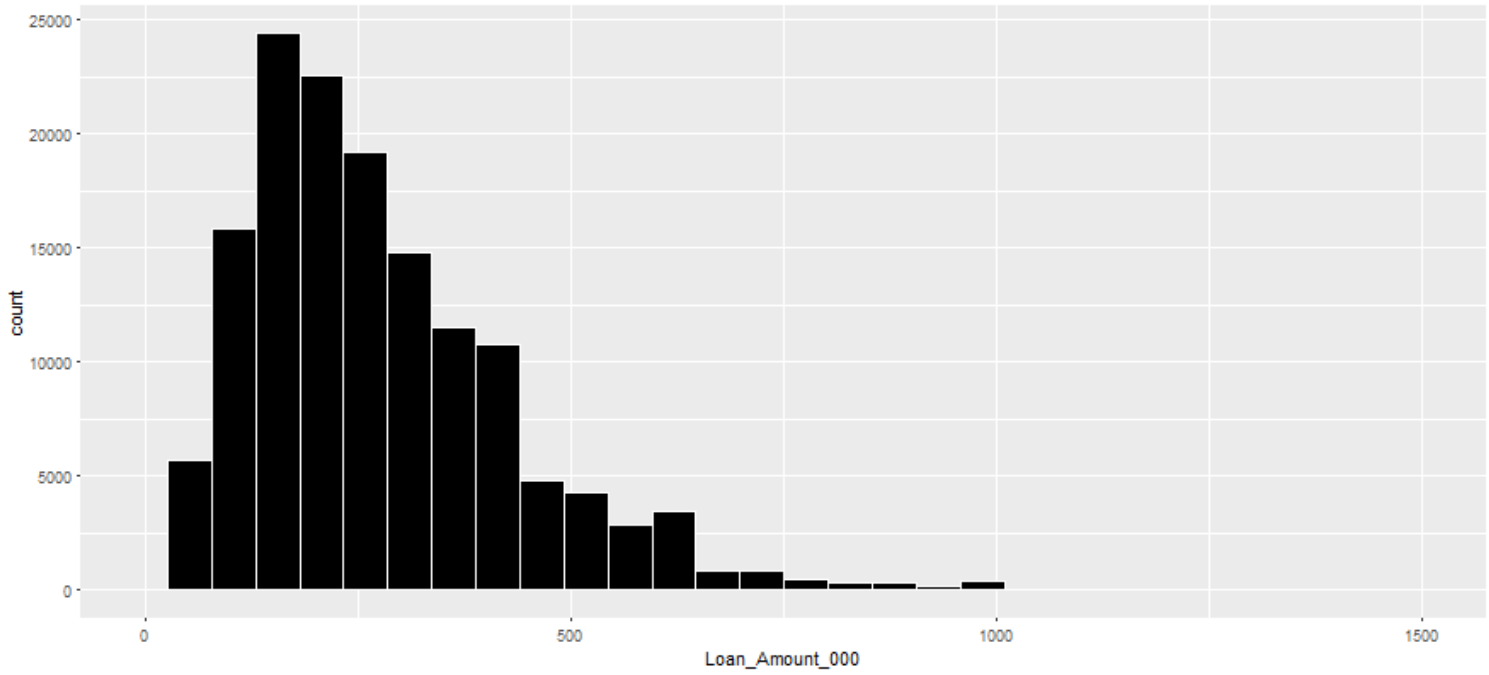
Business Insights

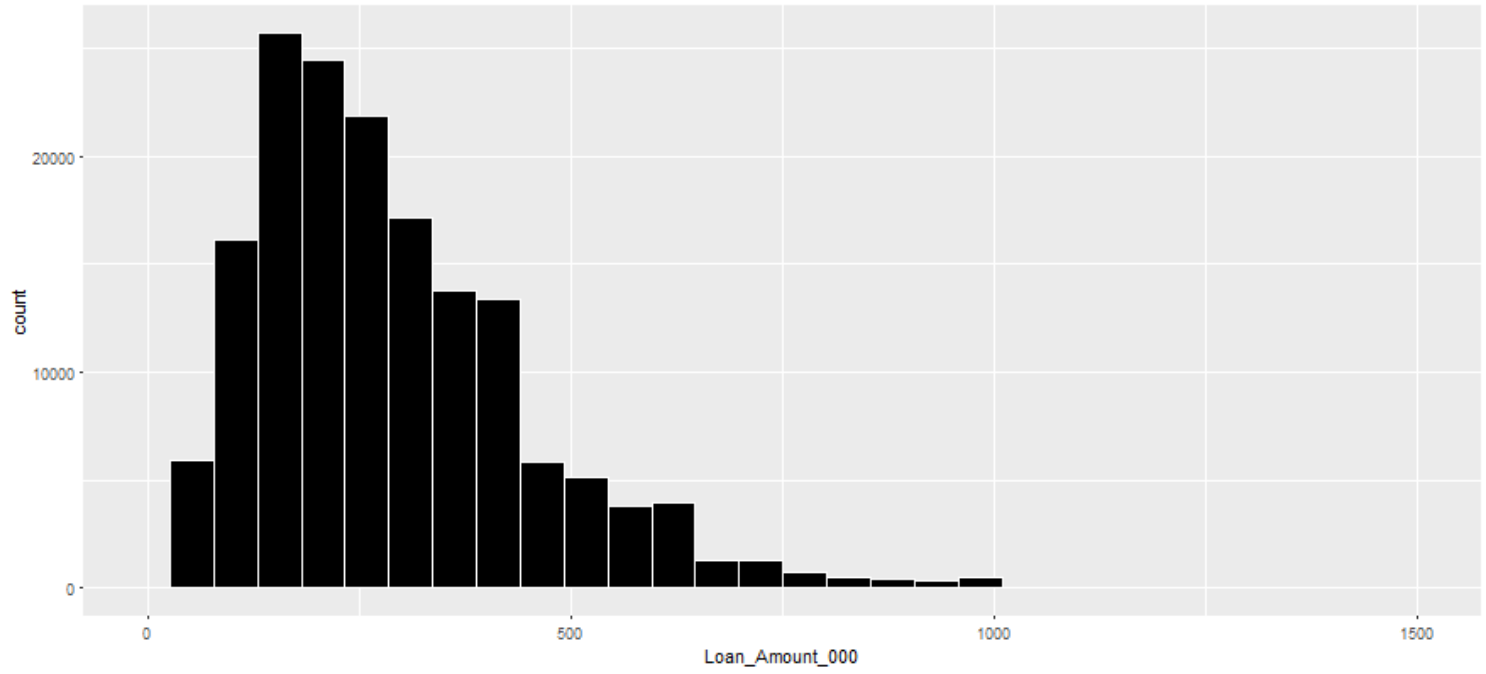
Key Insights

Loan Amount

Year 2012:

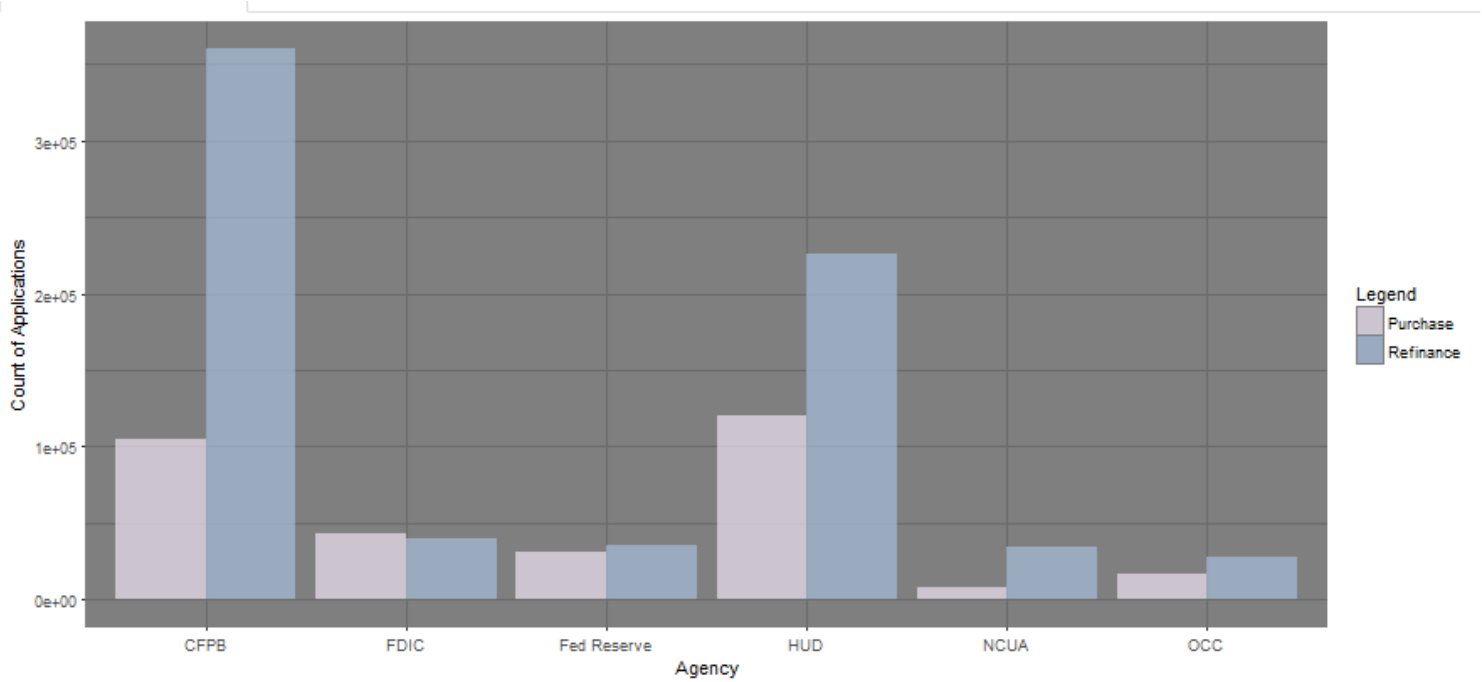


Year 2013:



There is an increase in loan amount from 2012 to 2013 for loan purpose = “Purchase”

Loan Types



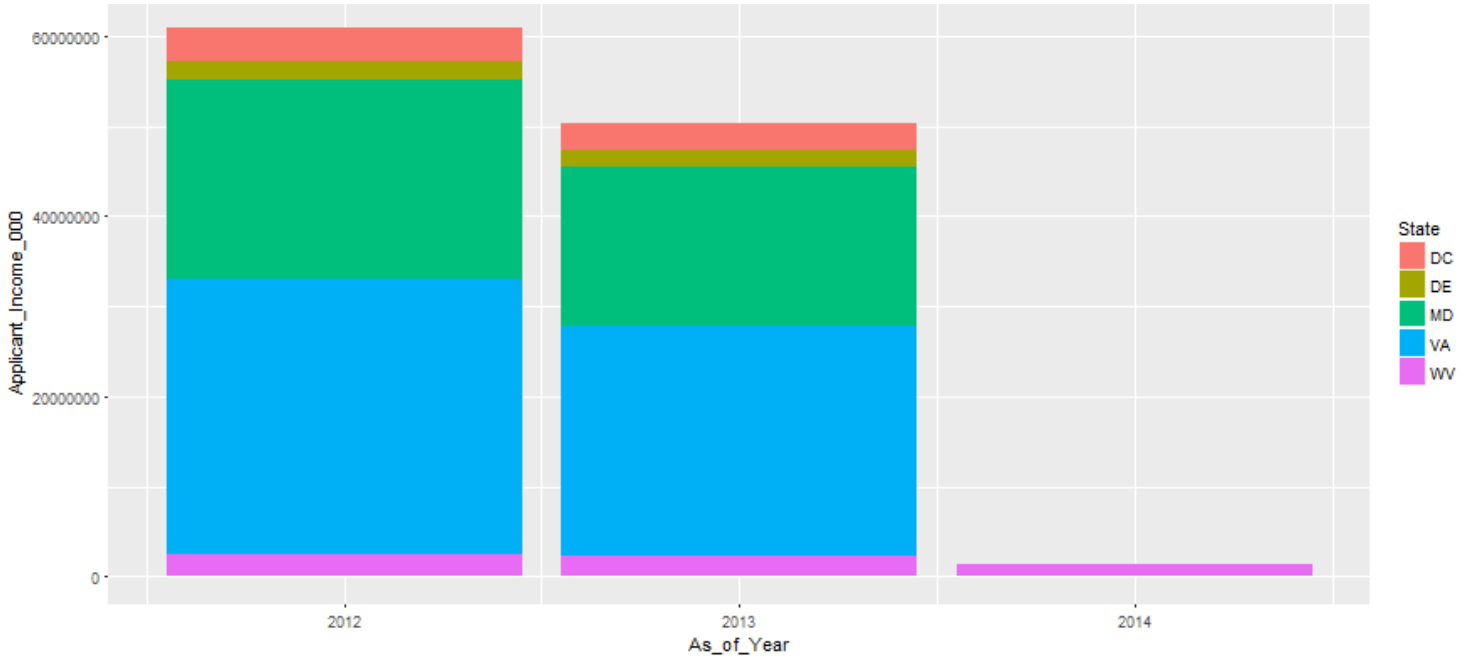
Refinancing loans:

There is overall decrease in loan types across the agencies compared to CFFB

Purchase loans:

There is overall decrease in loan types across the agencies compared to CFFB

Applicant Income



There is a decrease of 16.66% percent of applicant income from 2012 to 2013. There is not enough information about applicant income for the year 2014.

Pattern and Outliers

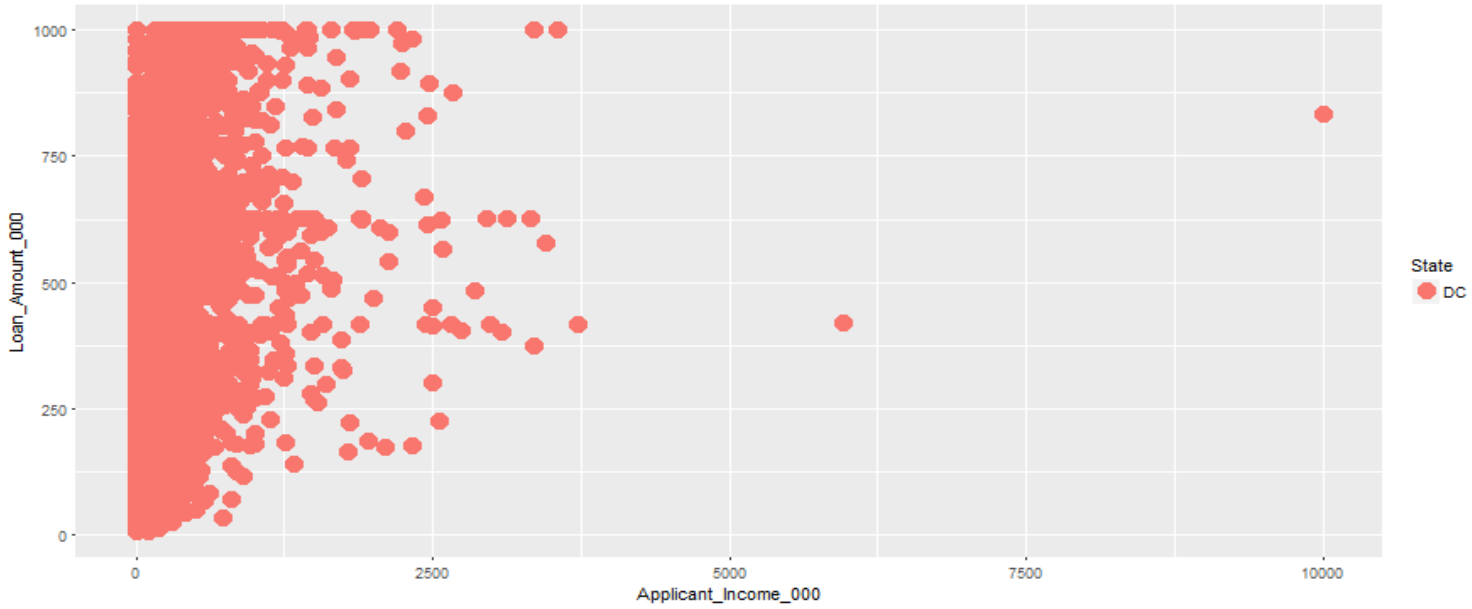
The scatter plot easily identifies concentrations and outliers, therefore, efficiently identifying risks and opportunities

Loan Amount vs. Applicant Income

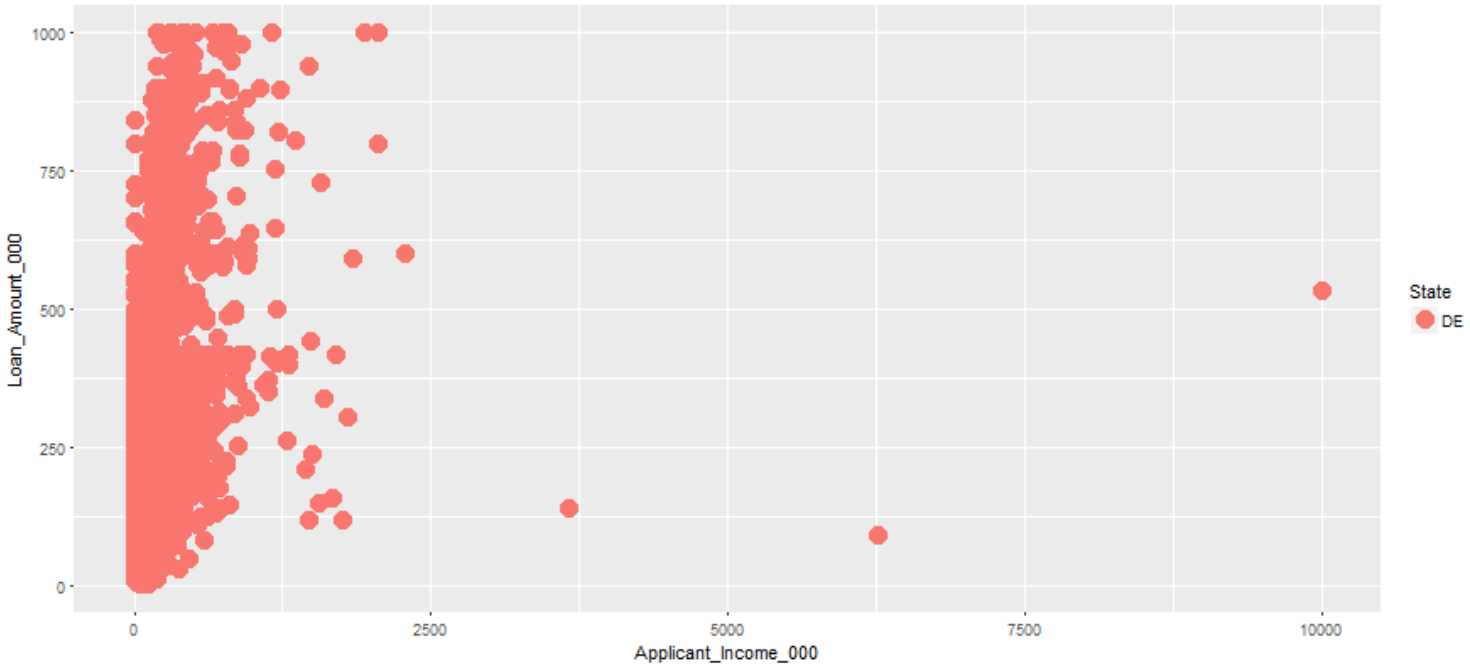
The ratio of loan amount to applicant income of minorities is evenly spread across all states and agencies

For each state:

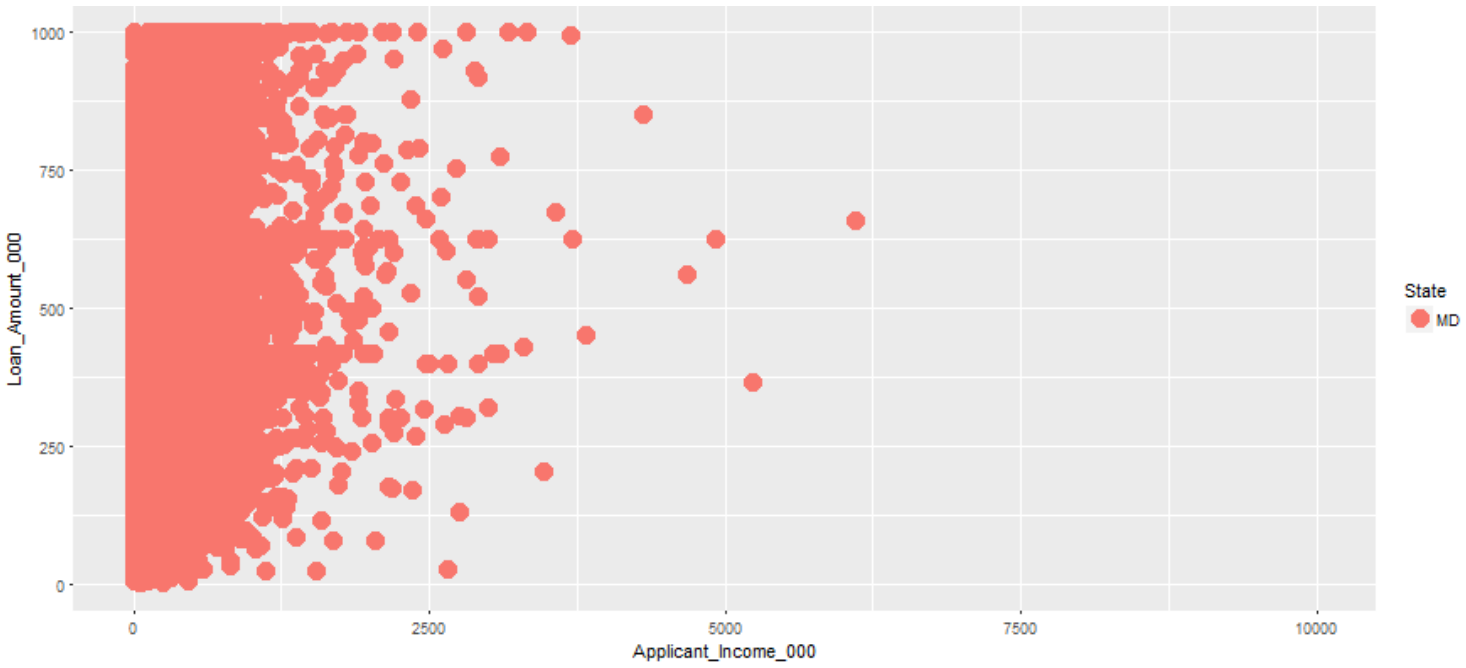
DC:



DE:



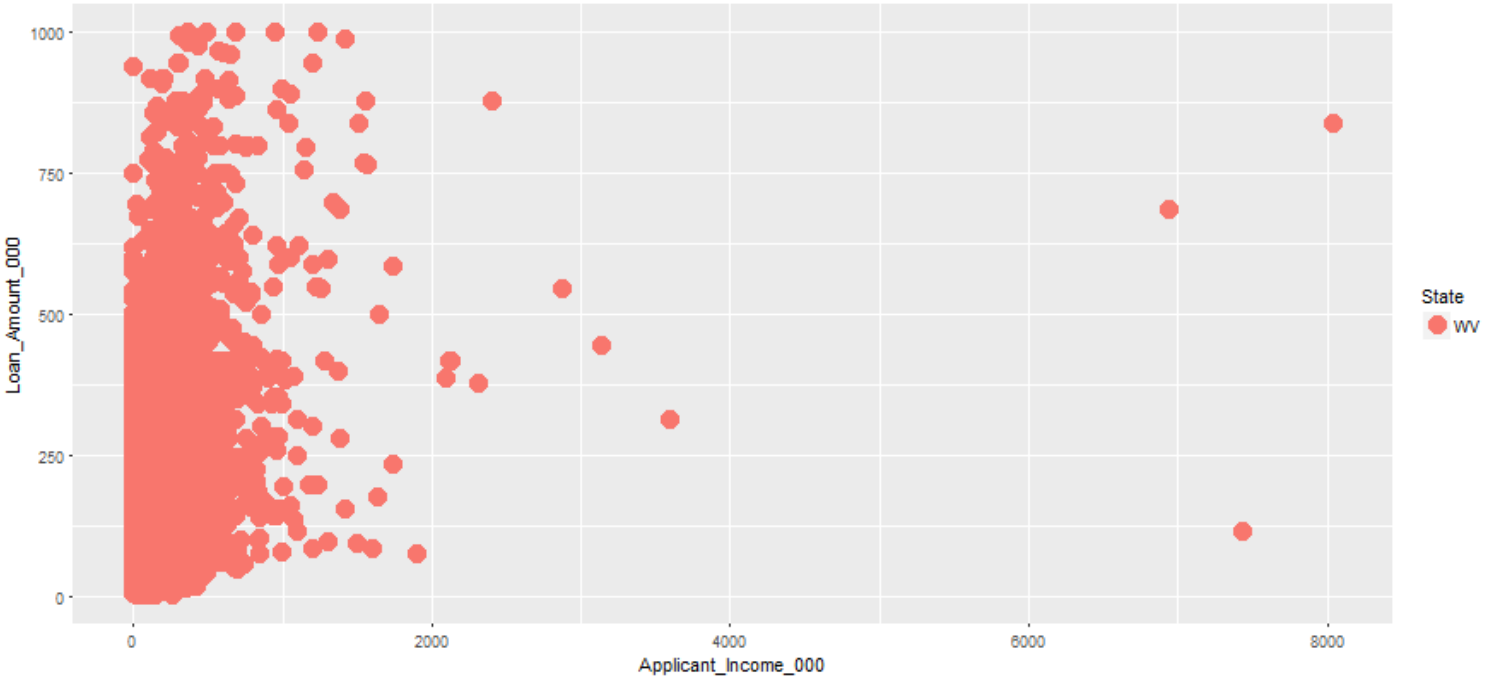
MD:


1000
750
500
250
0
Loan_Amount_000
0
2500
5000
7500
10000
Applicant_Income_000
State
MD

VA:


1000
750
500
250
0
Loan_Amount_000
0
2500
5000
7500
10000
Applicant_Income_000
State
VA

WV:



The state “MD” has more even distribution of loan amount and applicant income as compared to other states.

Data Management

I selected only specific fields for the data analysis. That is the reason I have selected a handful of fields to do an in-depth analysis of the data set.

Data Cleaning

- Filter: The dataset consisted of 50 plus columns from which I selected 24.
- Lookups: The columns were recoded for better usability.
- Factors: The continuous columns were factorized for better readability.
- Filtered rows due to lack of sufficient factors
- Managing Outliers : The outliers are managed by excluding the extreme outliers from the data set.

Data Enrichment

- I have included few filters for better analysis of data. The filters include
    - States: MD, WV, VA, DC, DE
    - Loan Purpose: Purchase, Refinance
    - Year: 2012:2014
- I have imputed NA values to 0 for efficient data interpretation.
- Have used scatter plot, histogram, stacked plot for the more effective representation.