Project Report: LENDING CLUB LOAN DATA

Data VISUALIZATION

INDIVIDUAL REPORT:

VISHNU VARDHAN ANNABATTIN

# **Purpose & Data set:**

Goal of this project activity is to plot & visualize “**Non-trivial**, **Interesting & subtle**” features derived out of our data set. We have chosen **Lending club loan dataset** as it has diverse types of variables and massive amount (around 800,000) individual observations.

Lending club loan data set (Ref: <https://www.kaggle.com/wendykan/lending-club-loan-data>) has following Metric variables.

1. *Loan Amount,*
2. *Funded Amount,*
3. *Interest rate,*
4. *Installments (monthly payments),*
5. *Annual Income (Borrower declared annual income during application),*
6. *Issue date (The month which the loan was funded),*
7. *Dti(A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income),*
8. *delinq\_2yrs (The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years),*
9. *inq\_last\_6mths (The number of inquiries in past 6 months excluding auto and mortgage inquiries),*
10. *The number of months since the borrower's last delinquency,*
11. *The number of months since the last public record,*
12. *Number of derogatory public records,*
13. *Revolving line utilization rate the amount of credit the borrower is using relative to all available revolving credit,*
14. *Remaining outstanding principal for total amount funded,*
15. *Payments received to date for total amount funded,*
16. *Last month payment was received,*
17. *loan application type,*

And following Categorical variables,

1. *Term of loan (with two possible values of “36 months”, “60 months”),*
2. *Grade (Loan grade with 7 possible values of “A”, “B”, “C”, “D”, “E”, “F”, “G”),*
3. *Employment Length (Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years),*
4. *home ownership (with possible values of “ANY”, “MORTGAGE”, “NONE”, “OTHER”, “OWN”, “RENT”),*
5. *verification status (Borrower application verification with three possible values of “Not Verified”, “Source Verified”, “Verified”),*
6. *Loan status (Current status of the loan with ten possible values of “Fully Paid”, “Charged Off”, “Current”, “Default”, “Does not meet Credit policy status:charged off”, “Does not meet Credit policy status:fully paid”, “In Grace Period”, “Issued”, “Late(16-30) days”, “Late(31-120) days”),*
7. *payment plan (Indicates if a payment plan has been put in place for the loan with possible values of “y”, “n”),*
8. *Purpose (category provided by the borrower for the loan request with possible values of “Car”, “Credit Card”, “Debt Consolidation”, “Educational”, “Home Improvement”, “House”, “Major purchases”, “Medical”, “Moving”, “Renewable Engery”, “Small Business”, “Vacation”, “Wedding”, “Other”).*

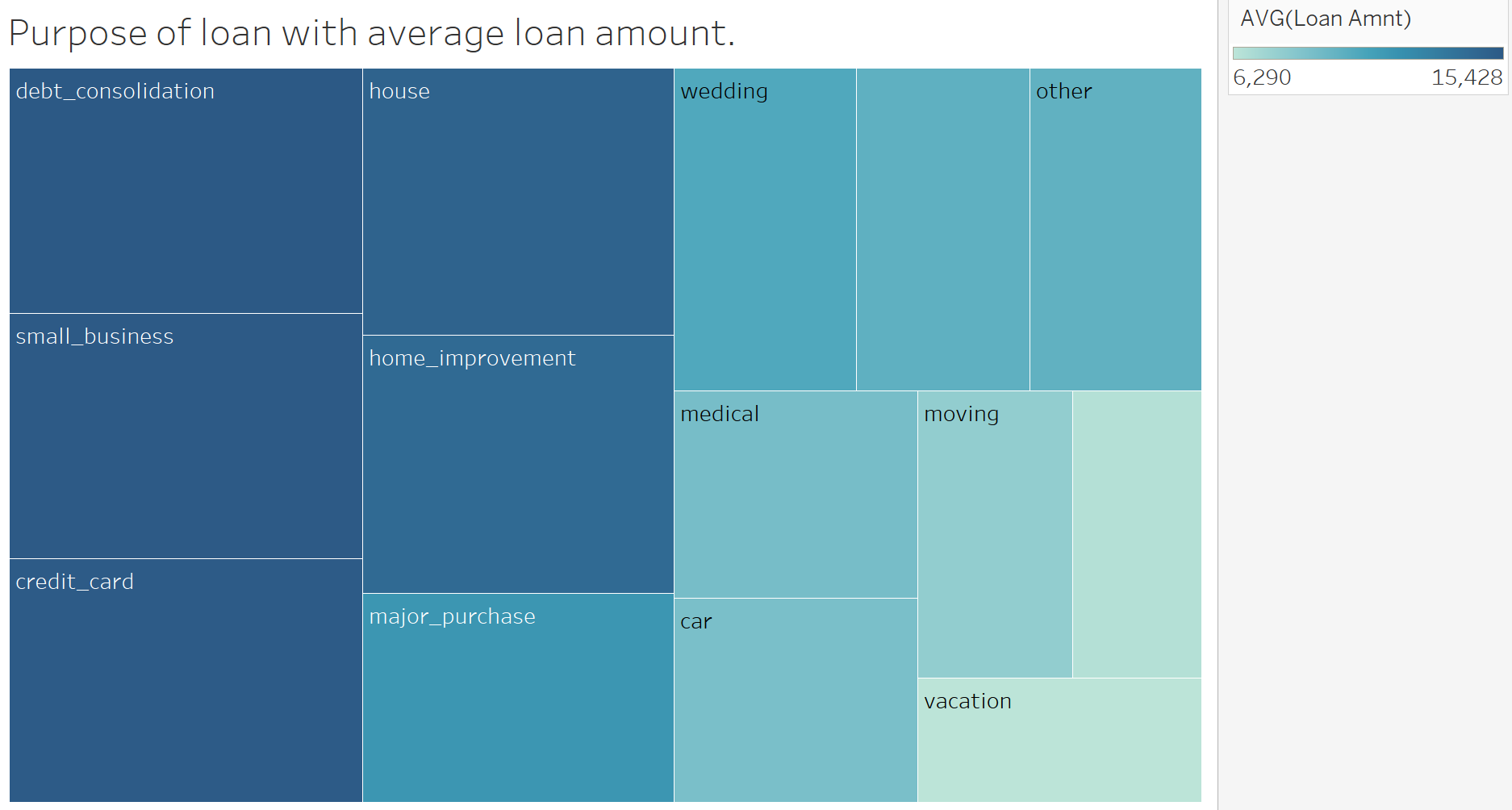
And following non-numeric variables,

1. *Employee Title (Job Title of the borrower),*
2. *Description (Description entered by borrower during loan application),*
3. *Title (The loan title provided by the borrower).*

And Geographical variables such as *Zip code, State of USA.*

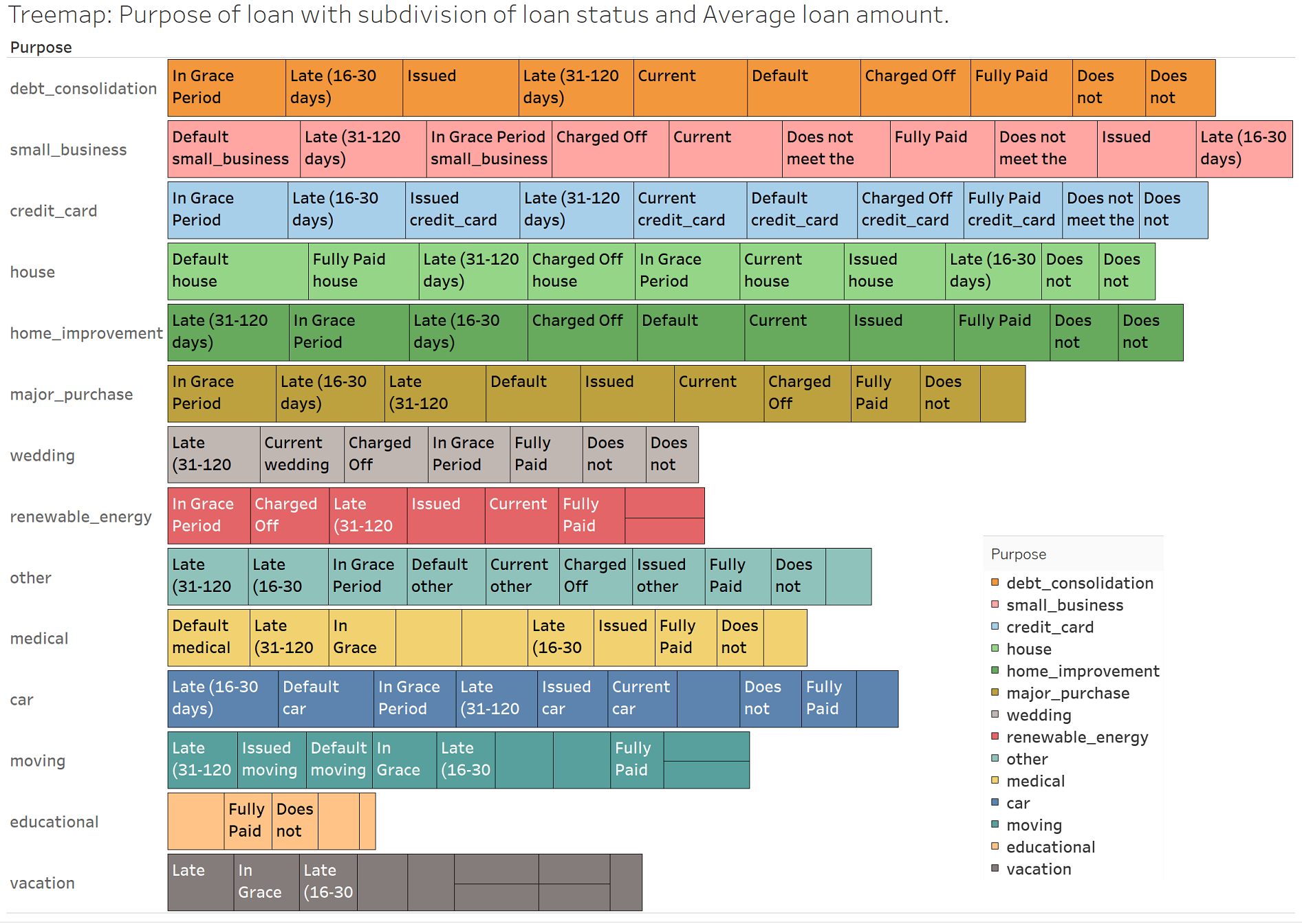
# **Initial Exploratory Visualizations:**

**Tree map:** **Initial plot did on “purpose category” of loan filled with average loan amount in that category to see which purpose category has highest/lowest loan amounts.**



**Observation:** **“Debt\_consolidation” has highest average loan amount and “vacation” being least average loan amount.**

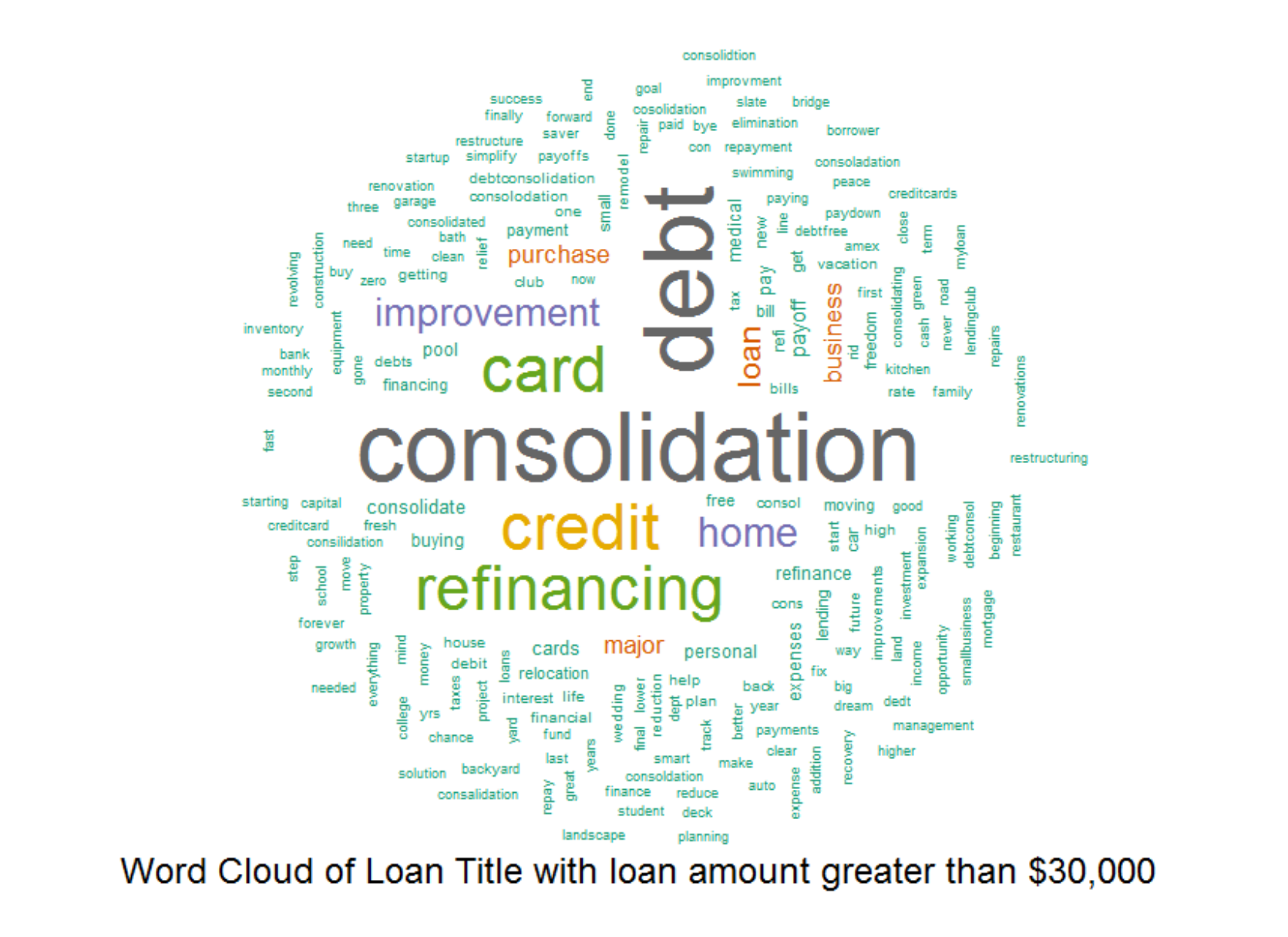
**Tree map:** I**mprovised tree map by adding additional categories to make it more hierarchical. Below is the redone treemap.**

****

**Observation:** **“Debt\_consolidation” has highest average loan amount and “vacation” being least average loan amount.**

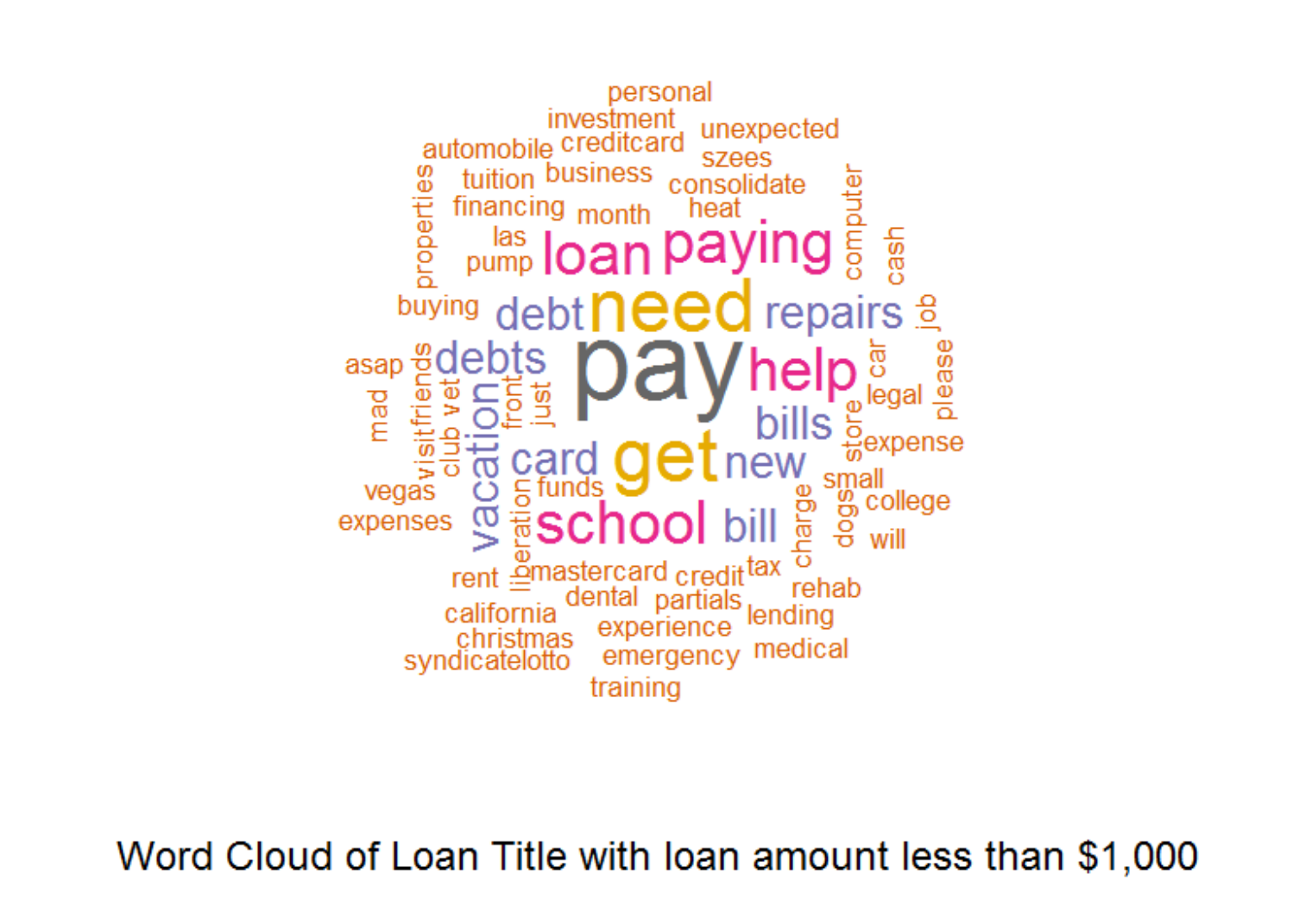
**WordCloud:** **Did wordcloud using R on loan title provided by borrower as part of loan application process with loan amount greater than $30,000.**

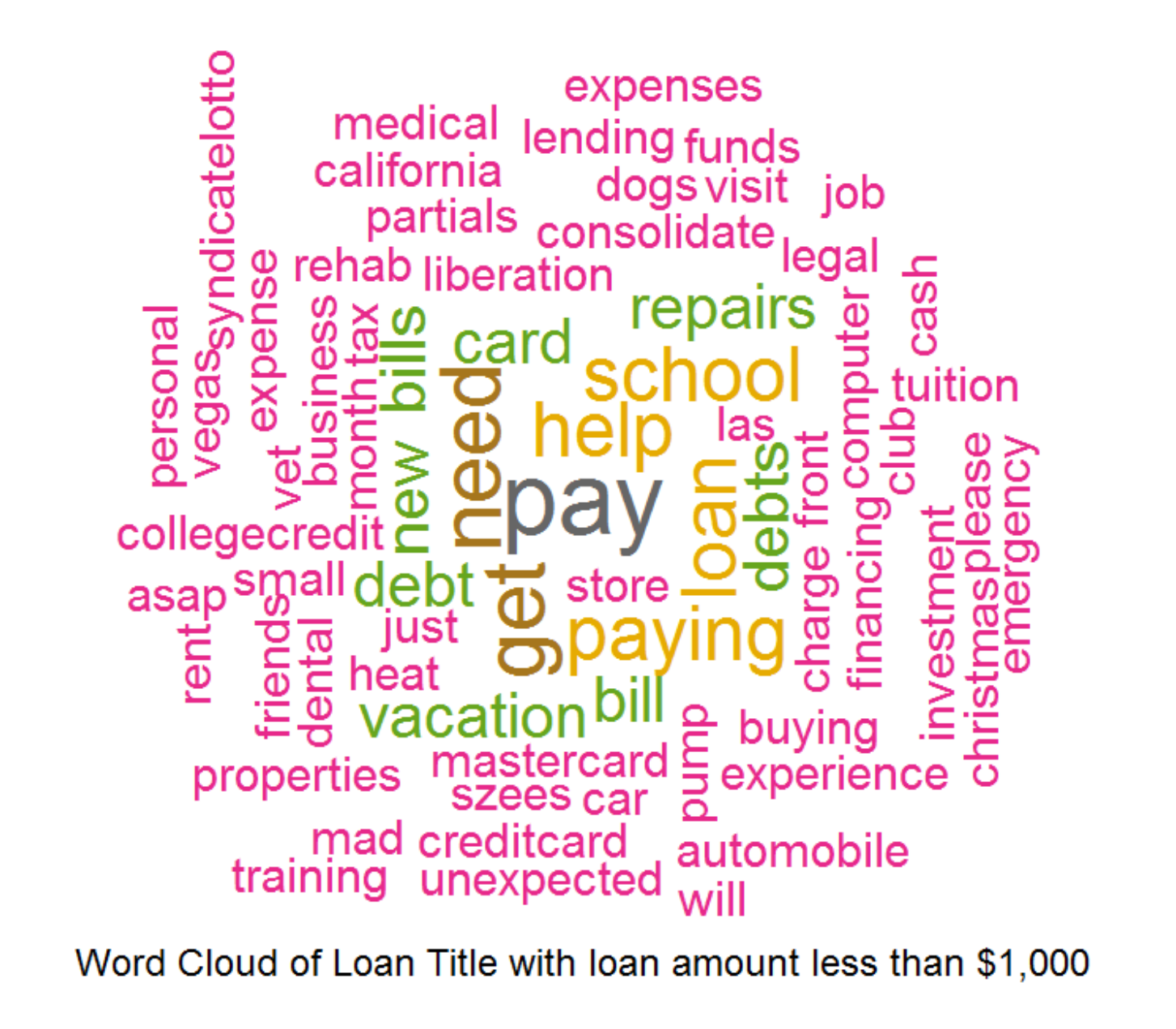




**Observation:** **“Debt”, “Consolidation”, “creditcard” are some of frequent words used by Borrowers while filing loan applications.**

**WordCloud:** **Did wordcloud using R on loan title provided** **by borrower as part of loan application process with loan amount less than $1000.**

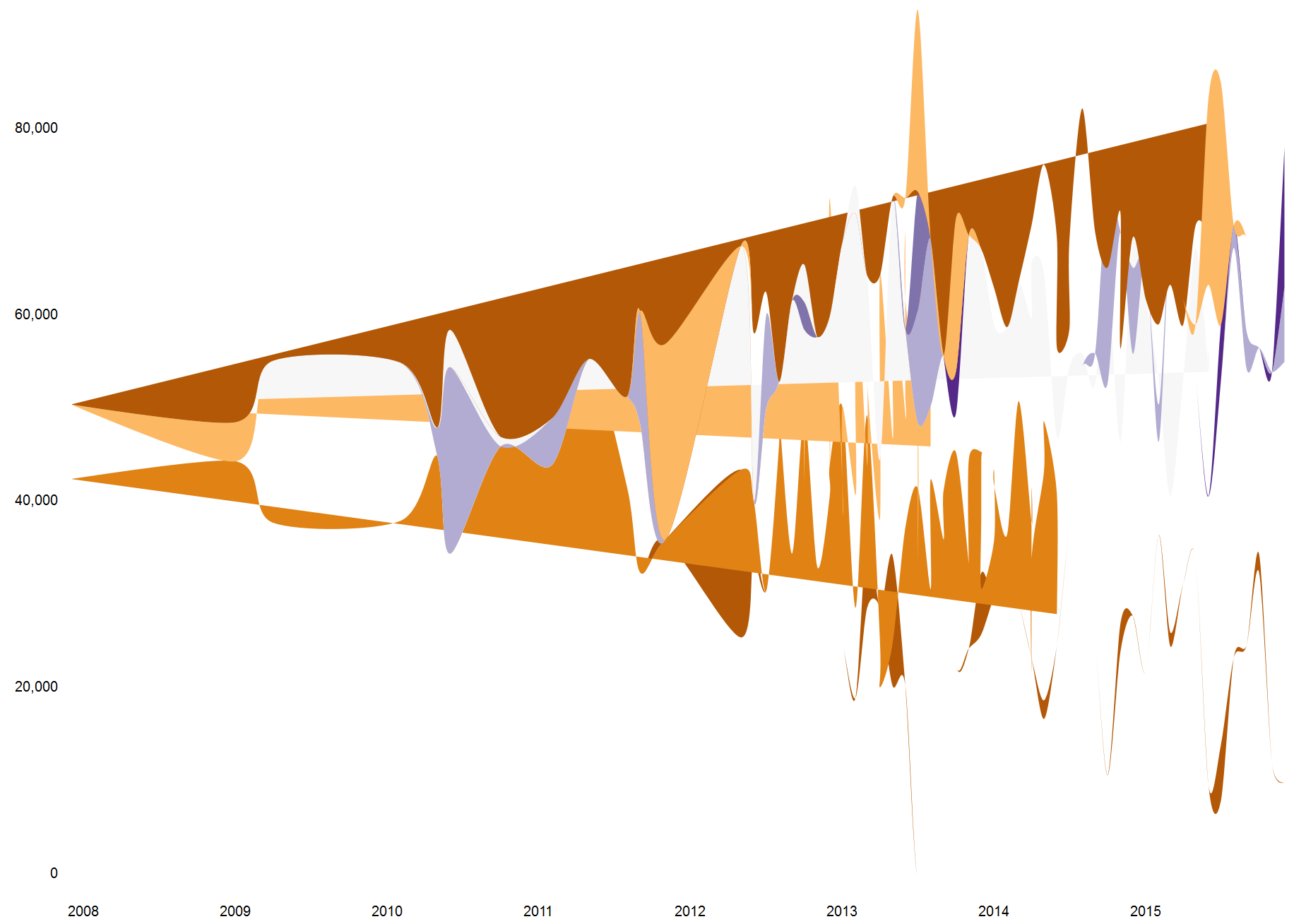




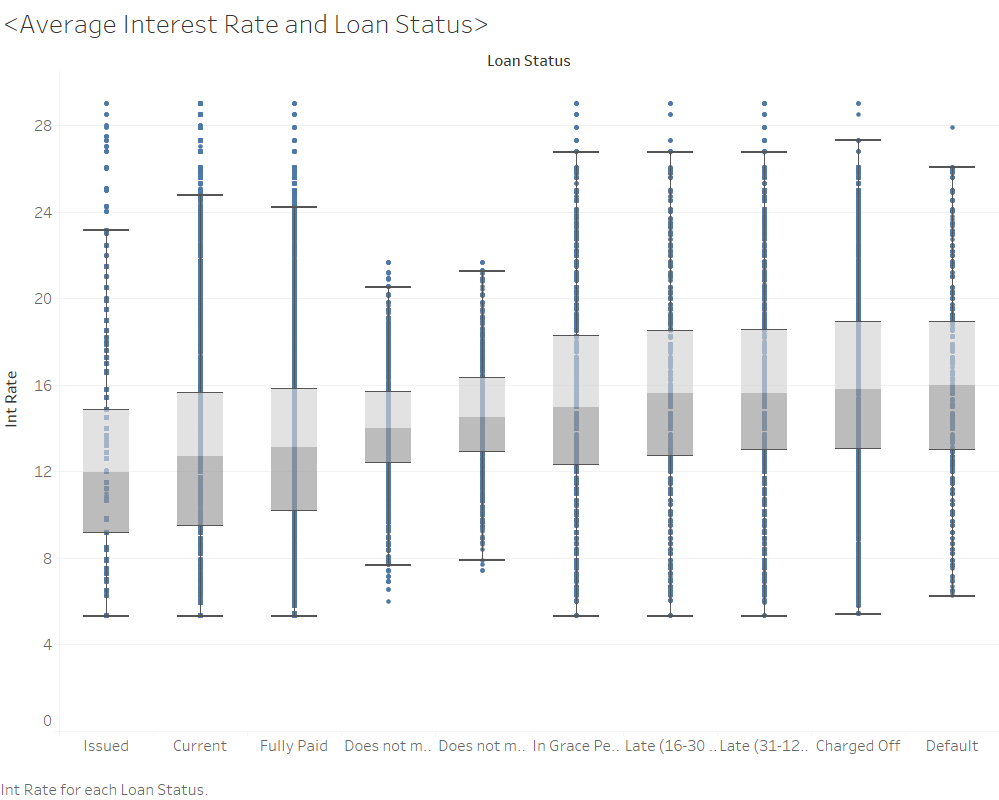
**Observation:** **“repairs”, “vacation”, “debt” are some of frequent words used by Borrowers while filing loan applications.**

**StreamGraph:** **Tried doing StreamGraph(ThemeRiver) to see if loan trends are varying based on month, unfortunately couldn’t get it to work properly to see clear visualization.** (**if still have time before final submission, will definitely retry doing it**)

**Below is the plot produced so far,**

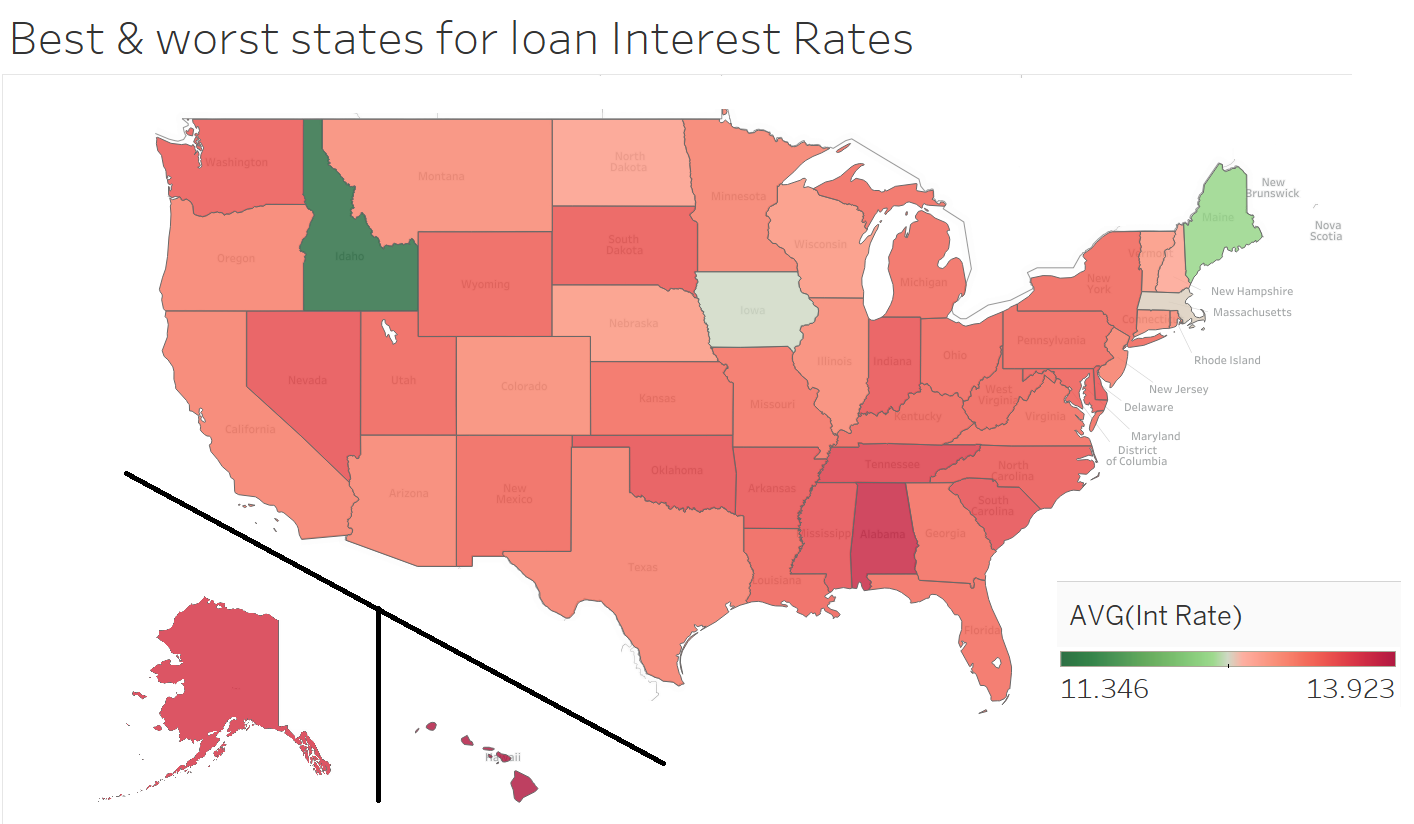


**Boxplot:** **Plotted interest rates with loan status.**

****

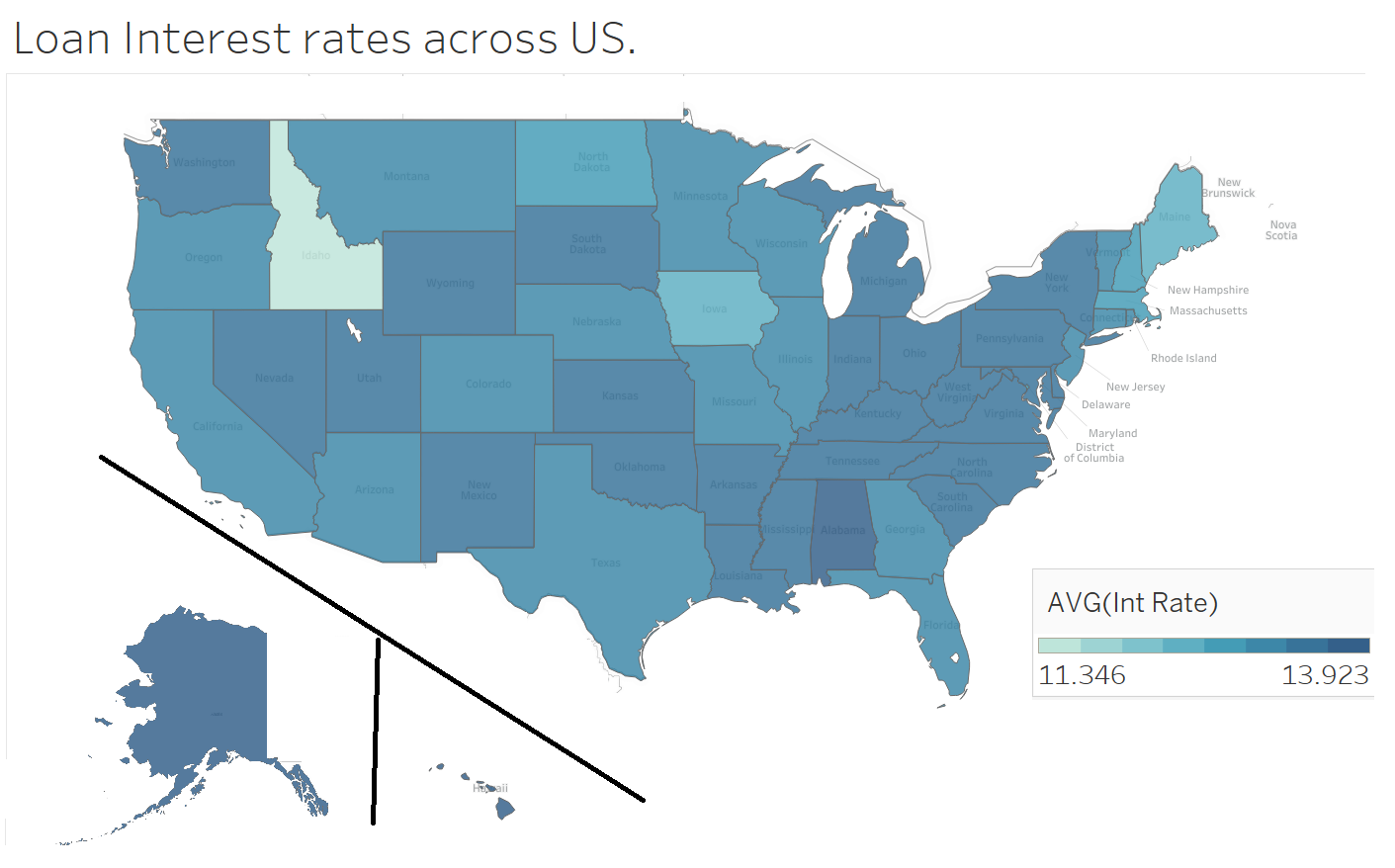
**Observation:** **Borrowers with “Default” status are lending with high interest rates.**

**Choropleth:** **Best & worst states with loan interest rates. (USING Divergent colors to indicate opposite sides)**



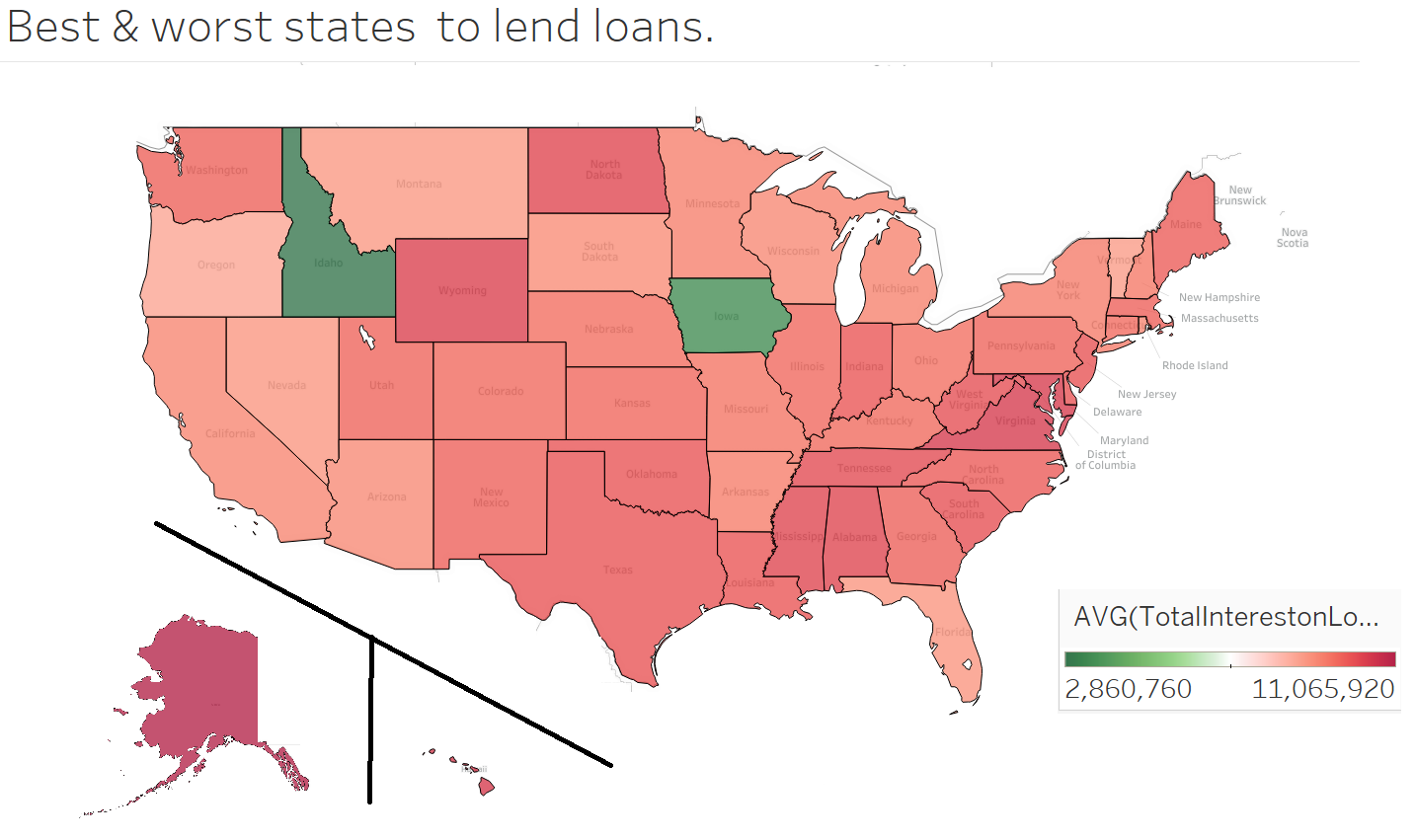
**Observation: “Idaho” & “Maine” have least interest rates.**

**Choropleth:** **loan interest rates across US States (non divergent colors)**



**Observation: “Idaho” & “Maine” have least interest rates.**

**Choropleth:** **Best & worst states for amount of pay back on loans. (USING Divergent colors to indicate opposite sides)**



**Observation: “Idaho” & “Iowa” are best for borrowers.**