David Ingraham
Andrew Hughen
Final Project Proposal: Data Analysis and Visualization
March 3rd, 2023

Link to dataset:

https://www.kaggle.com/datasets/saravananselvamohan/freddie-mac-singlefamily-loanlevel-dataset

This data set contains information on single family loans purchased by Freddie Mac from 1999 to 2018. There are 27 million rows in this dataset, and that could be too much for R to handle so I will just take the first 3,000 mortgages in 1999 for the analysis.

Uploading this to kaggle three years ago, Freddie Mac is a government sponsored liquidity provider for the US residential mortgage market. Founded in 1970, it was a part of a push by the US government to turn as many Americans into homeowners as possible.

The csv file has 27 variables and are listed below, although the loan sequence number won't serve any analytical purpose. A potential source or error in the data may be the debt to income ratio, because in the first half of this time interval, income verification by the originator was notoriously lax.

The common perception today is that lending standards have tightened since the 2008 financial crash, but is this true? Did delinquent loans in the early 2000's have similar traits to loans in the 2010s? The data stops at 2018, but are we due for another housing crash?

```
 [1] "CREDIT_SCORE"
 [2] "FIRST_PAYMENT_DATE"
 [3] "FIRST_TIME_HOMEBUYER_FLAG"
 [4] "MATURITY_DATE"
 [5] "METROPOLITAN_STATISTICAL_AREA"
 [6] "MORTGAGE_INSURANCE_PERCENTAGE"
 [7] "NUMBER_OF_UNITS"
 [8] "OCCUPANCY_STATUS"
 [9] "ORIGINAL_COMBINED_LOAN_TO_VALUE"
[10] "ORIGINAL_DEBT_TO_INCOME_RATIO"
[11] "ORIGINAL_UPB"
[12] "ORIGINAL_LOAN_TO_VALUE"
[13] "ORIGINAL_INTEREST_RATE"
[14] "CHANNEL"
[15] "PREPAYMENT_PENALTY_MORTGAGE_FLAG"
[16] "PRODUCT_TYPE"
[17] "PROPERTY_STATE"
[18] "PROPERTY_TYPE"
[19] "POSTAL_CODE"
[20] "LOAN_SEQUENCE_NUMBER"
[21] "LOAN_PURPOSE"
[22] "ORIGINAL_LOAN_TERM"
[23] "NUMBER_OF_BORROWERS"
[24] "SELLER_NAME"
[25] "SERVICER_NAME"
[26] "PREPAID"
[27] "DELINQUENT"
```