

# PS6\_Wardwell

lwardwell

March 11th, 2025

## 1 Question 5 - Visualization #1

For my first visualization, I chose to practice utilizing the FDIC SDI API and pull more bank information. I am learning that the R loop which has to be used to pull information for more than one financial institution at a time is not terribly fast, so I chose to limit the sample to only active institutions for a single year-end (December 31st, 2020) and for only one variable at a time, in this case "ASSETS" (in hindsight I will probably do more variables at a time, I was just being extra cautious with my code). ASSETS is the FDIC API name for the Bank's total assets at the report date. I added a log of ASSETS to my data in order to transform the large numbers into a smaller measure and utilized a histogram to check the distribution. As expected, the log of total bank assets histogram approaches a normal distribution. This histogram can help us see that most of the observations exist in the expected shape, but there appear to be some outliers in the tails.

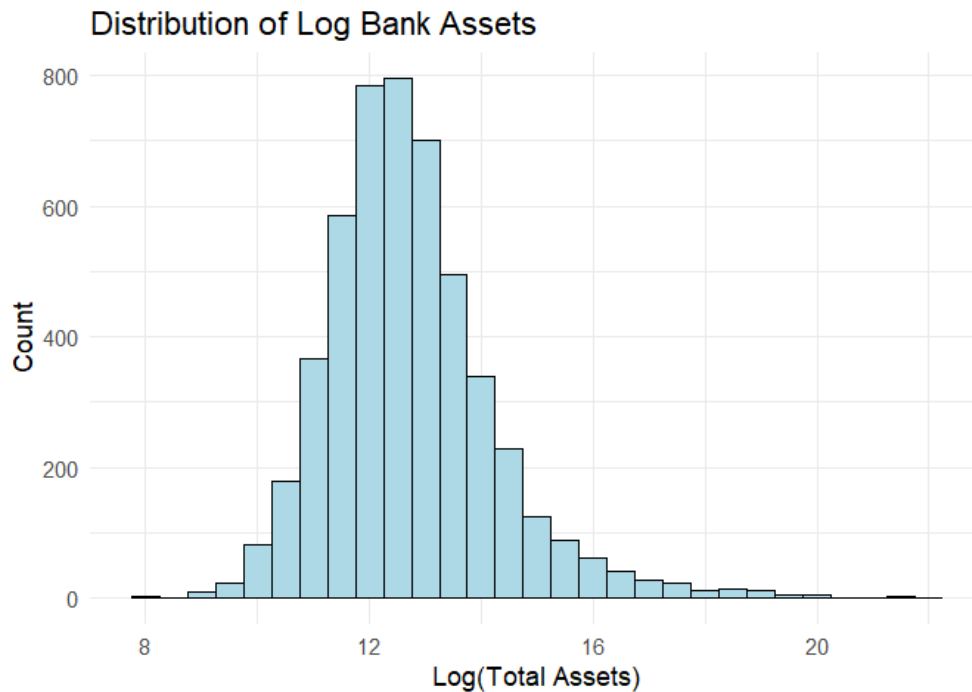


Figure 1: Histogram of the log of total bank assets for all active institutions as of December 31st, 2020.

## 2 Question 5 - Visualization #2

For my second visualization, I chose to utilize the API again and this time pulled down total deposits information ("DEP") for the same banks at the same report date. Again, this process took quite some time to run the R loop. If you are going to run my code to test it, I recommend just running one of the processes and not both. If I use this code again, I will definitely optimize it or see if ChatGPT can recreate it Python or SAS if one of those programs would run faster. With this data I chose to create a quick boxplot, which shows a similar result to our histogram. While most of the observations are clustered in the middle of the distribution, there are still frequent outliers, moreso on the upper end of the distribution. This distribution makes sense to me considering the vast range of financial institution sizes in the United States. The smallest institution I ever audited was \$20 million in assets (Freedom State Bank in Freedom, OK) and the largest bank in the US at that time was JPMorgan Chase at \$2.57 trillion in total assets. Because deposit balances are closely correlated with the bank's asset size, it would make sense logically that the deposit distribution should mimic the assets distribution. This is an incredible size differential, and could result in quite a large range across the boxplot, even after taking a log transformation of the data.

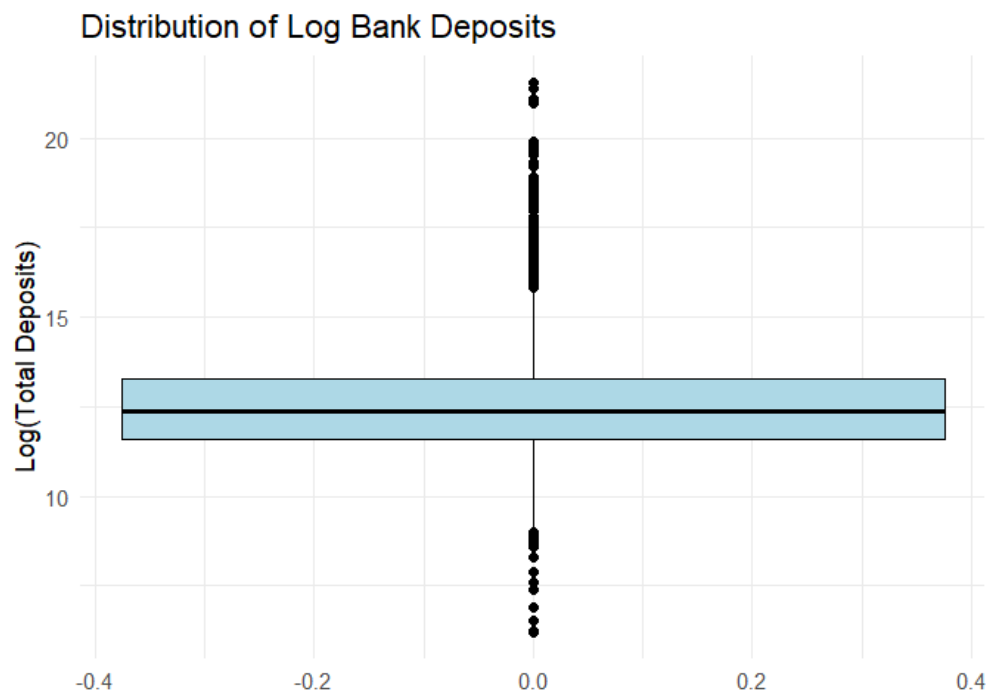


Figure 2: Boxplot of the log of total bank deposits as of December 31st, 2020.

### 3 Question 5 - Visualization #3

For the final visualization, I decided to create a scatterplot asset and deposit sizes between community banks on this date. A community bank is generally defined as an institution with less than \$5 billion in assets and domestic US operations only. I combined my two datasets and filtered to only institutions below the asset size threshold. This scatterplot shows a mostly linear relationship between the two variables, and again, the relationship makes logical sense to me. The outliers at the bottom of the plot are showing a significantly larger log\_asset than log\_deposits. This would probably be indicative of banks that are funding their loan portfolio using wholesale funding from the Federal Home Loan Bank, Federal Reserve, or other correspondent banking relationships to fill the gap left by lower deposit balances. In 2020, it's entirely possible that some of these banks also had significant borrowings on their books from the PPP Liquidity Facility after issuing PPP loans.

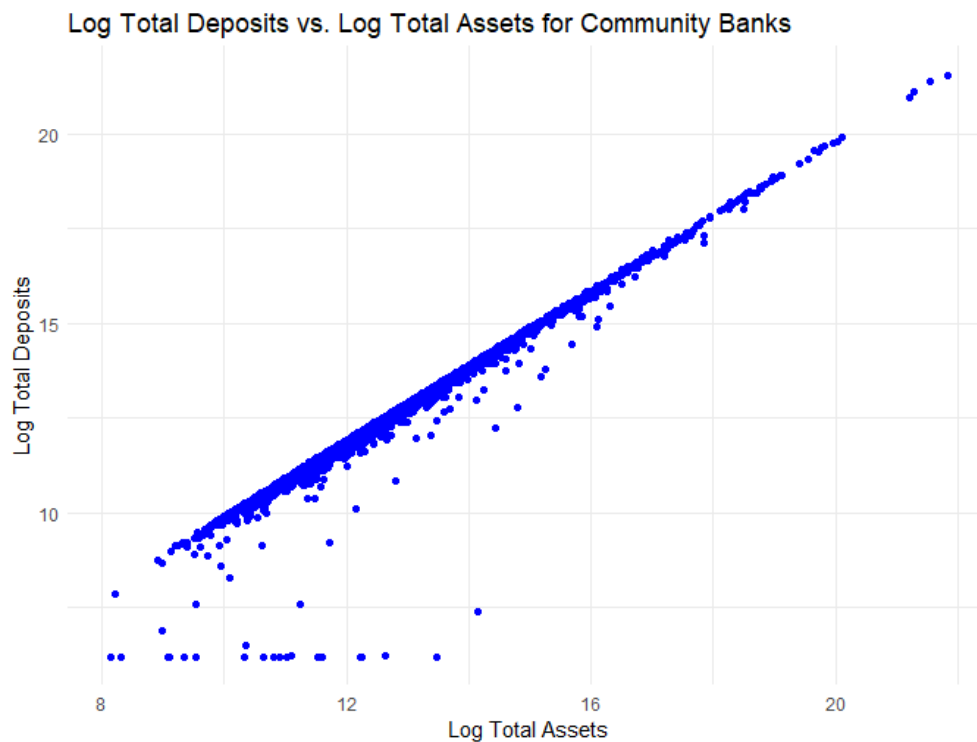


Figure 3: Scatterplot visualizing the relationship of log\_assets to log\_deposits.