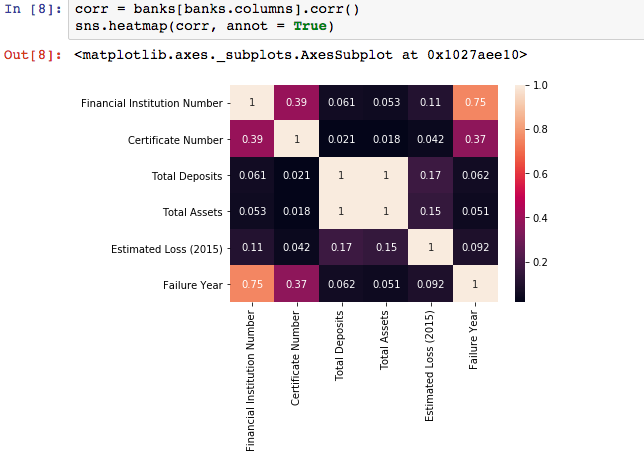
Exploring the data of failed banks from 1934-2017

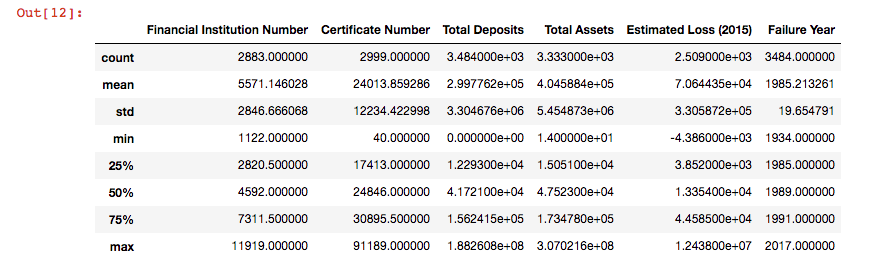
For my Capstone project, I wanted to explore a dataset that not only had appropriate depth and breadth for critical analysis and visualization, but also one whose contents aligned with my particular interests. It took a few days to sift through the various open datasets linked to the course program, but eventually I stumbled across an interesting dataset on Kaggle that contained detailed bank failure information from 1934 to 2017.

This report lists each failure of a commercial bank, savings association, and savings bank since the establishment of the FDIC in 1933. Each record includes the institution name and FIN number, institution and charter types, location of headquarters (city and state), effective date, insurance fund and certificate number, failure transaction type, total deposits and total assets last reported prior to failure (in thousands of dollars), and the estimated cost of resolution. Data on estimated losses are not available for FDIC insured failures prior to 1986 or for FSLIC insured failures from 1934-88. It should be noted that this bank failure report dataset was downloaded from the [FDIC website](https://catalog.data.gov/dataset/fdic-failed-bank-list).

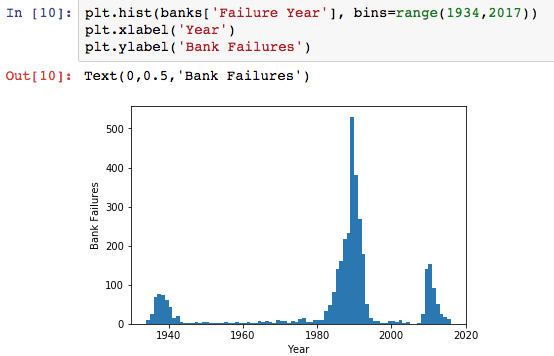
As stated in my thesis, the inspiration for this project comes kaggle but more so from my lifelong curiosity about recurring financial and economic problems. In that regard, the lifespan of banks; their state and national influence peaks my interest. And I will utilize the techniques gleaned from this course to answer to the following questions to the best of my ability. How have bank failure rates changed over time? What type of banking institution is the most likely to fail? What commercial bank failure cost the federal government the most to resolve?

To get started, I decided it would be helpful in deciding where to go next by visualizing the correlations of the entitled headers for various columns and rows of the banks data into my own rows and columns using a heat map from the seaborne library. The Summary Statistics show us a basic roadmap of what normal and what’s not in the data and visuals to come. In the heat map I found a high correlation between “Failure Year” and “Financial Institution Number”. Therefore I focused on these data and data like it such as “Institution Name” which was just as significant as the value for FIN, which is basically the account number for a bank a way to account for banks if you will.  So using this name rather than a number will allow people to grasp more easily with more significance.





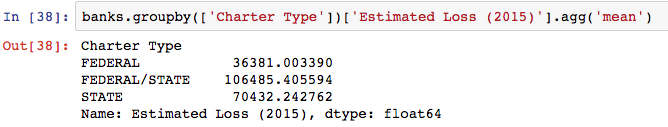
Analytic question #1

How have bank failures changed over time? Which bank is more likely to fail State or National. I decided to group the banks data by Failure Year. I recommend seeing line 9 in the project, to contrast with the histogram below. From this data frame I noticed the numbers arranged themselves into three distinct groupings around separate periods like tailored bell curves without any extra visual aid. And so when I did add the histogram the results were striking! hadn’t considered the 1980’s to be a period of such high bank failures.

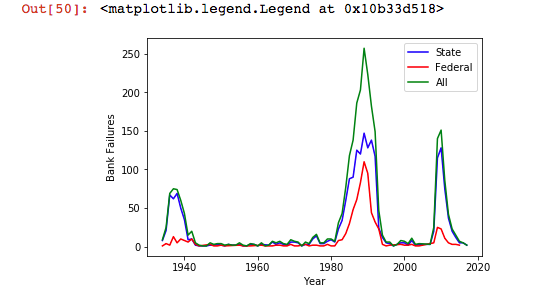
The most significant changes over time have revealed themselves to me in three distinct periods of time. Period 1. 1930-1940; The Great Depression, 2. 1980-1990 The recession of the 1980s, and 3. 2008-2013 the great recession. Judging by this graph I believe it’s worth rethinking the 1980s as the period when the actual ‘great recession’ occurred.

Analytic question #2

What type of bank is more likely to fail Federal or State? I separated this data into three values: State, Federal, and All(a summation of the two). The statistics below show that state bank failures are almost double those of Federal bank failures.

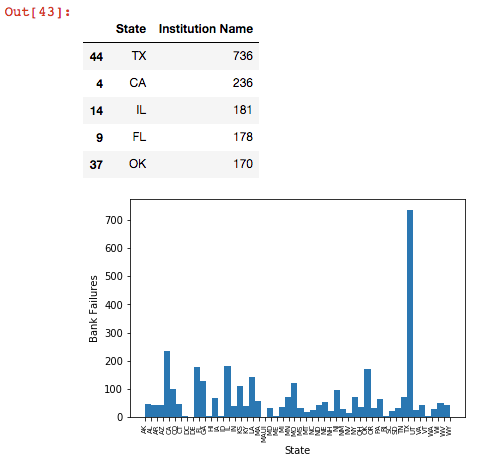


The graph below makes it clear to us how much more state chartered banks tend to fail. Additionally the difference is double overall, the difference is triple in the 1930’s, so and so in the 1980’s and quadruple in the 2000’s.

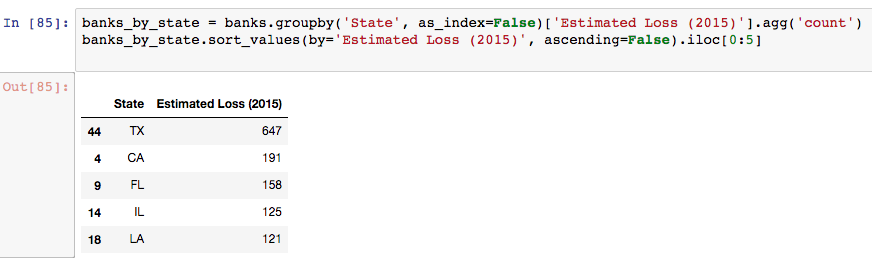


Analytic question #3

Which states had the highest failures; state chartered and federal chartered combined? Which state cost the federal government the most to resolve?



Texas had triple the bank failures of California, the largest state in the country. The state that cost the government the most to resolve was Texas again.



Proposition for further research

I believe that the total assets and the total deposits for each bank at the date of failure are significant values worth understanding more for further research. The total deposits consist of money placed into the banking institution for safekeeping. These deposits are made to deposit accounts such as savings, checking, and money market accounts. Total assets are the total worth of the bank or better yet its value to investors. The asset portion of a bank’s assets includes the cash that depositors let the bank hold, government securities, and interest-earning loans like mortgages, letters of credit and inter-bank loans. Looking more thoroughly into both values and their relationship to each other is worth spending time on and can be built upon with the work that has been done for this project (see heat map in beginning).