Data Visualization Course Project Proposal

Zhuoyue Zhao, Ya Gao
{zyzhao, ygao}@cs.utah.edu

1 Proposal

Basic Info. Visualization for FDIC Failed Bank List
Group members:
Zhuoyue Zhao (u1065756), zyzhao@cs.utah.edu
Ya Gao (u1137971), ygao@cs.utah.edu
https://github.com/zzy7896321/dataviscourse-pr-fdicfailedbanks

Background and Motivation. We looked at the resources on the course web page and found this FDIC failed bank list. FDIC (Federal Deposit Insurance Corporation) was created in 1933 to restore public confidence in banking during the Greate Depression in the 1930s. It insures a depositor for up to $250,000 in a FDIC member bank in the event of a bank failure. FDIC usually find a bank that is willing to acquire a failed bank so that the deposits are simply transferred to the new bank. However, sometimes, it cannot find a acquiring bank and will send the insured funds directly to the customers. It is amazing that no depositor has ever lost FDIC insured funds since its establishment but it is also worrying that a massive wave of bank failures or a large bank failure could fail FDIC. We are curious about how reliable FDIC has been and will be in the future.

Project Objectives. We plan to visualize the spatio-temporal FDIC bank failures as well as the distribution of acquiring banks. 1) The primary goal is to get an idea about how useful and reliable FDIC has been in the past. 2) We might also want to know how likely will FDIC fail if a lot of bank or a large bank fails based on the peak period of failure in history. 3) We want to use d3 and svg to create the visualization. We also want to use brushing and linking to allow exploring the details of the data.

Data. We downloaded a detailed list of FDIC failed banks from 1934. It can be obtained by launching the select-all query at `https://www5.fdic.gov/hsob/SelectRpt.asp?EntryTyp=30&Header=1`. It includes the name, the location of the headquarter, the effective date, the total deposits, the total assets and other data of the banks.
We also have a second list of FDIC failed banks from October 1, 2000, available at `https://www.fdic.gov/bank/individual/failed/banklist.html`. In addition to the data available in the first table, it also lists the acquiring institutions of the failed banks.

Data Processing. The data in both tables are quite clean. There are a few missing values in irrelevant columns and the estimated loss column. To deal with that, we'll ignore those rows with missing values in the estimated loss column. There are 4095 records in the table, which might be too many for our purpose. We are considering restricting the data to starting from October 1, 2000. We also plan to do some aggregations over the data on the numeric columns grouped by either

year or state so that we can have a spatio-temporal summary of the data. We can also compute the statistics of failed banks each acquiring institutions acquired by joining the two table. That's the second reason why we will restrict the data in the first table to the ones after October 1, 2000.

Visualization Design. We proposed the following 3 prototype designs.

Design 1) See Figure 1.

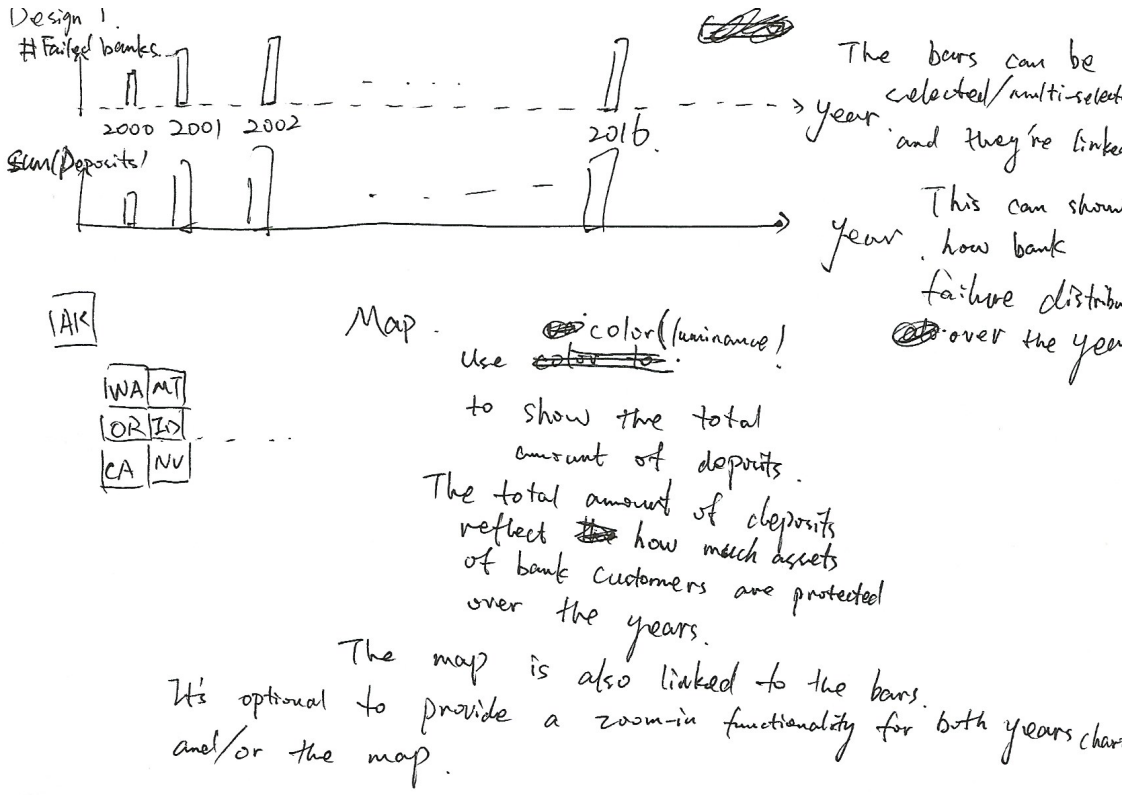

Figure 1: Design 1

Our first design consists of 3 charts: two yearly bar charts and a map.

The first bar chart is on the top, and plot the number of failed banks of each year from 2000 to 2016. The second bar chart is below the first one, and has the same x-axis. The y-axis is instead the summation of the total deposits in the failed banks in each year. These two bar charts show the failure distribution over the years in terms of both numbers and amounts. We anticipate that the two quantities are correlated but can vary a lot. For example, we expect the hike in the recent financial crisis from 2008 might be more obvious in the second bar chart since several large banks went bankcrupt. The map uses color luminance to show the total amount of deposits of the failed banks in each state. It visualizes the spatial distribution of the failed banks.

The bars in the first bar chart can be selected or multi-selected. The second bar chart and the map are linked to the first bar chart. The bars of the same years in the second bar charts are highlighted as in the first bar charts. The map is updated to display the sum of the selected years.

An optional feature of this design is to provide a zoom-in feature for both the year charts and the map to allow exploration of the details at a lower granularity.

The main advantage of this design is giving a summary of the failed banks and clearly shows the trends and distribution in both time and space. However, it lacks the ability to allow user to browse the details even if we have a zoom-in feature that allows the users to zoom-in to monthly charts or state maps.

Design 2) See Figure 2.

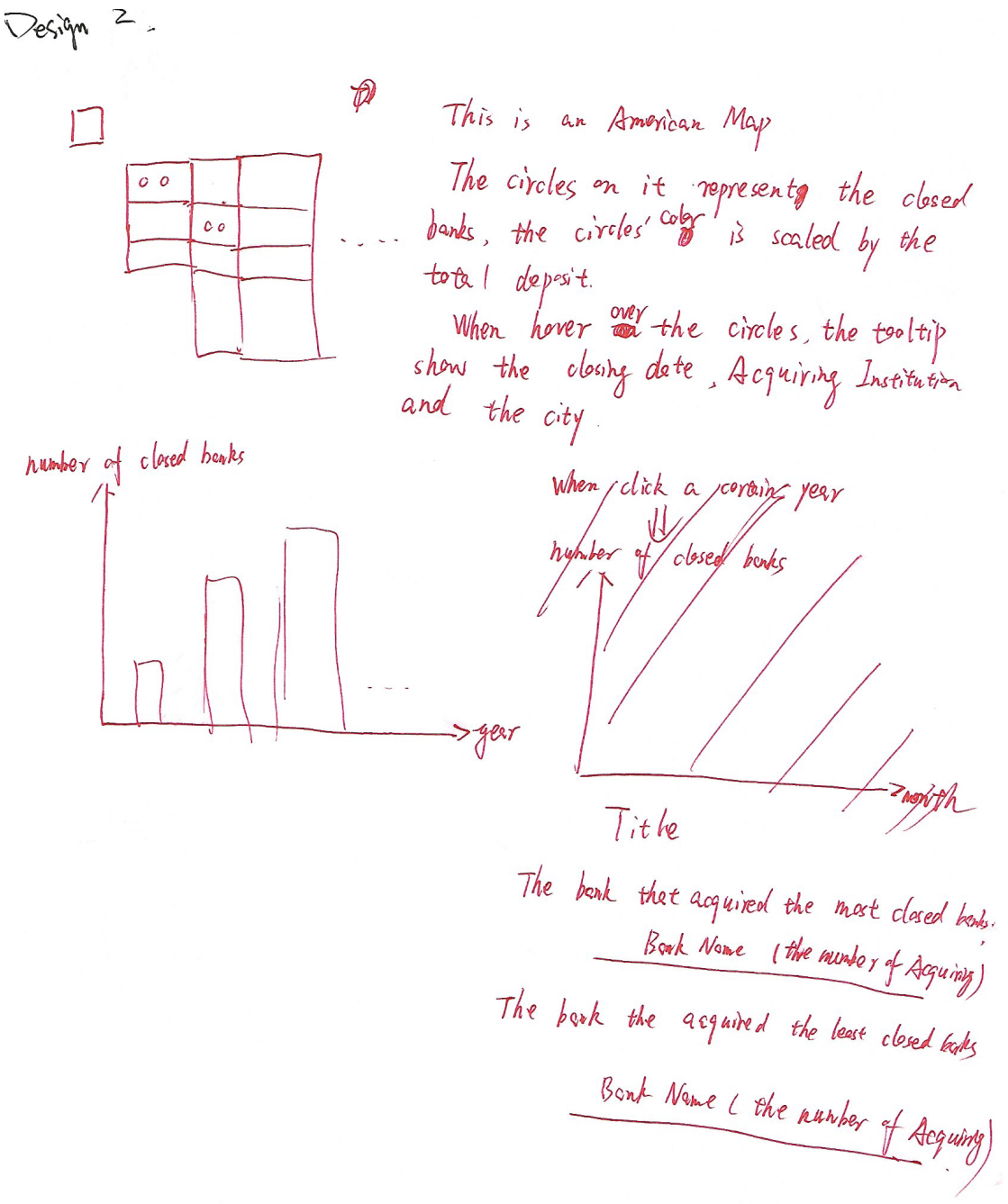

Figure 2: Design 2

The second design features a map, a bar chart and a list of top acquiring institutions.

At the top of the page is a map of the states. In each state, there are circles that represents failed banks, whose color lumination is scaled according to the total deposit. When a user hover the mouse over a circle, a tooltip is shown with the details of the failed bank.

Below the map are the bar chart ant the list. The bar chart on the left shows the number of failed banks over the years while the list on the right shows the top acquiring institions of the failed banks by the total number of failed banks they acquired over the years.

This design enables browsing details of the failed banks. But the three visualization elements are unlinked. It's hard to use it to explore the data.

Design 3) See Figure 3.

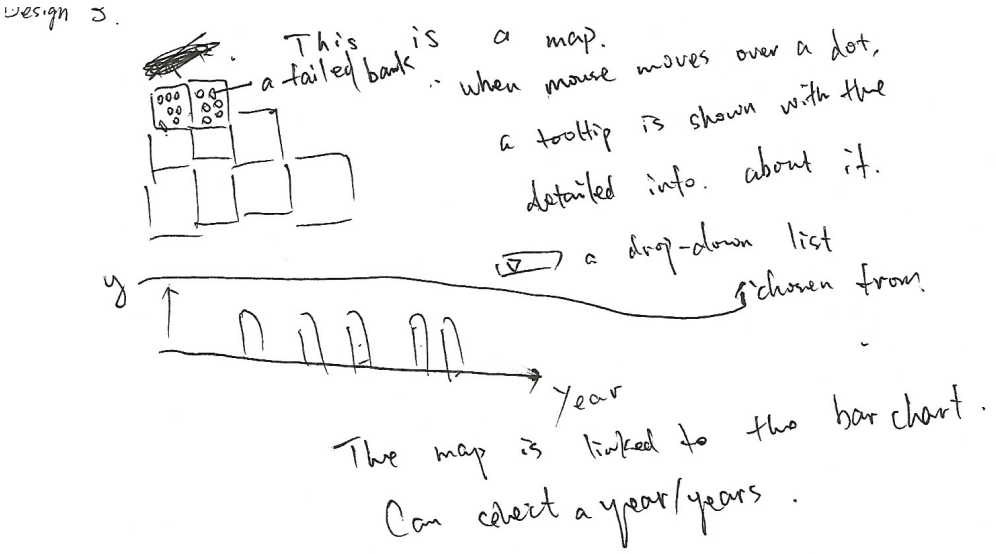

Figure 3: Design 3

The third design features a map, a bar chart with a configurable y-axis (through a drop-down list).

The map, similar to that in design 2, plots the failed banks as circles in the states. The color luminance of the circles is scaled according to the total deposits of the failed banks. When the mouse hovers over a circle, the details of that bank is shown in a tooltip.

The bar chart has year as its x-axis. The y-axis can be chosen in the drop-down list, including total deposits, number, total assets or estimated loss. This is more concise than the previous ones in the first chart and also shows different angles of the data. The bars in the bar chart can be selected as in design 1 and the map is also linked to the selections to reflect the selected years.

This design combines the benefits of the previous two and also provides a cleaner and more intuitive visualization. However, it is not as easy as a line chart would be to see the trend in the bar chart. The placement of circles in the map is also hard because there might be multiple banks in the same state. Finally, it lacks the information of acquiring banks, which can give an idea of the roles of large banks and FDIC in the process of bank failure.

After weighing the benefits and drawbacks of the ideas in each design, we decided on the following final design (see Figure 4).

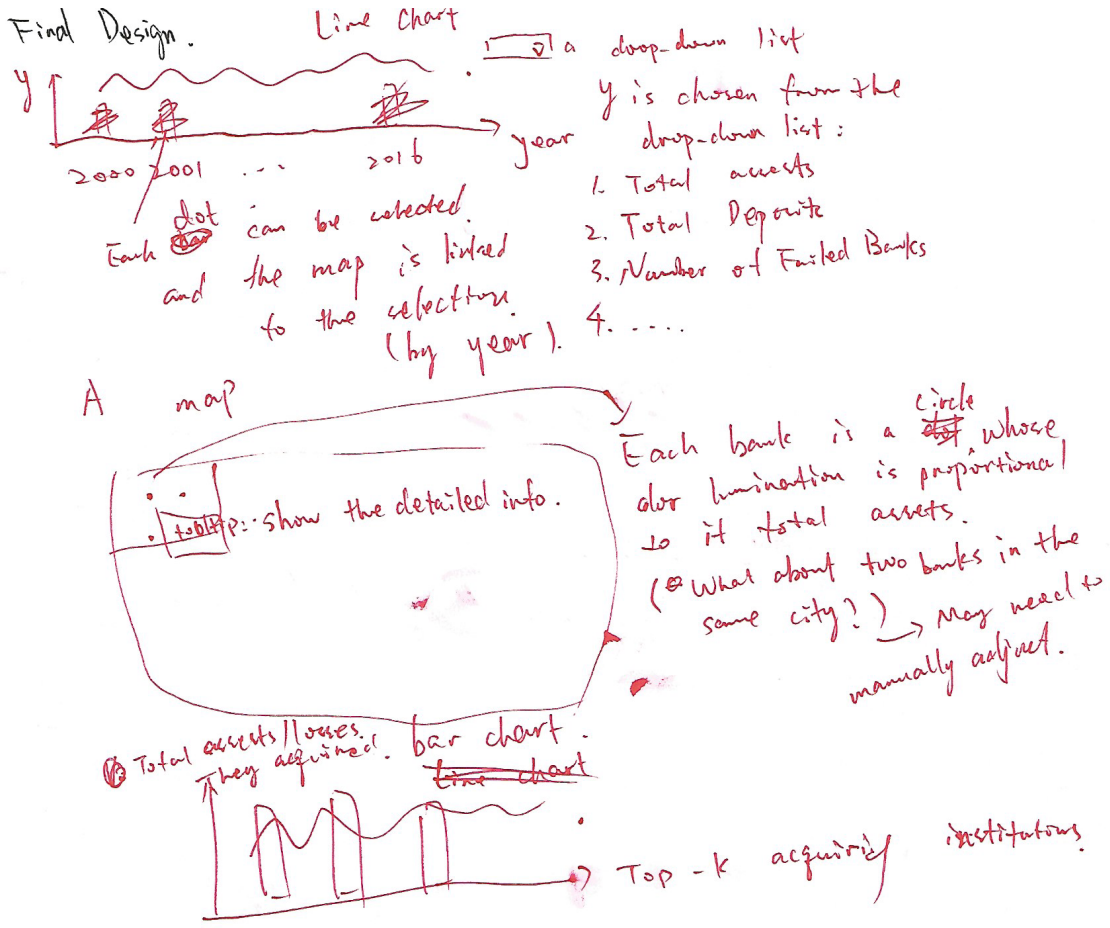

Figure 4: Final Design

The final design features a line chart, a map and a bar chart.

The line chart is similar to the bar chart in design 3. It has years as the x-axis and a configurable y-axis through a drop-down list. The dots in the year chart can be selected and the selection is linked to the other two charts.

The map uses the map with the realistic shapes instead of square grids and each bank is represented by a circle whose luminance is scaled according to the total assets/deposits (TBD). The location of the dot is at the real location of the headquarters of the failed bank. If multiple are in the same city, they will be placed around that city. To achieve this, we can either design an algorithm to automatically decide where to place the dots, or preprocess the data by hand, or a combination of the two. Each dot also has a tooltip that shows the detailed information about the failed bank. The dots shown in the map will reflect the year selections.

The last chart is a bar chart that shows the top-k acquiring institutions and the special case where there's no acquirer if it is not in the chart. The y-axis is the total assets of the failed banks that a particular institution acquired. This can show the ability of large banks and FDIC to acquire failed banks and might be able to show how concentrate the acquirings are.

Must-Have Features. The must have features include

- A line chart that shows the aggregations over the years.
- A map that shows the spatial distribution of the failed banks.
- A bar chart that shows the top-k acquiring institutions.
- They are linked together and years can be selected by clicking the bars in the first bar chart.

Optional Features. The optional features include

- Zoom-in functionality on both year charts (into month chart) and the map (into state-level maps).
- Fancy animations.

Project Schedule.

Week	Task
Oct 30 - Nov 4	Create a static html with empty divs. Start working on the charts. Zhuoyue will be responsible for the line chart and preparing the data. Ya will be responsible for plotting the map and the bar chart.
Nov 6 - Nov 11	Continue working on the charts and at least be able to show something on the webpage. Prepare the Project Milestone.
Nov 13 - Nov 24	Continue working on the charts.
Nov 27 - Dec 1	Finish everything, write the report and create the project screen-cast.

2 Feedback from Peer Review

We discussed our project proposal with Yanqing Peng and Yuwei Wang. While they found our project scope is adequate for a course project and our story make sense, we received the following valuable comments from them.

1. **Though the design gives a global picture of how bank failures are distributed over time and space, it is hard to pinpoint a specific bank of interest. The design does provide detailed information of a bank by hovering over a dot on the map, a bank of interest might not be known by its headquarter location. Rather it's more likely to be the bank name.**

   We will add a search box that allows searching specific banks. Those banks whose names partially match the search string will be highlighted on the map.

2. **The design lacks the usage of colors.**

   We plan to use a color scale on the dots of the map to indicate the size of it (TBD, either using the total deposits or the total assets).

3. **The bar chart is a bit confusing. It's not clear that it is linked to the line chart and the order of x-axis is not specified.**

   We will clarify the linking by very obvious chart titles/axis titles. We will also optimize the layout so that the linking is easy to identify.

We appreciate Yanqing and Yuwei's inspiring and high-quality comments. We both find them essential to make our visualization easier to explore and more self-explanatory. We have added the fixes to our implementation plan.

3 Data Collection

We have collected data from FDIC website with a list of failed banks since Oct 2000, which contains acquiring institutions, and another list of failed banks since the establishment of FDIC.
We joined the two tables based on their CERT number, which is a primary key of the tables. We filtered out any items that has N/A in the estimated loss column, which are all in 2017. Now we have data from Oct 2000 to Dec 2016.
We also want to get the geographic coordinates (longitutde and latitude) of the headquarters of the failed banks. We used a site called www.latlong.net. The site has a input box where one can enter the name of a location and a button "find", which loads the latitude and longitude into the other two input boxes when clicked. This website is awesome except that the api are hidden in a complex and lengthy JavaScript. Instead of trying to decipher the JavaScript code, we wrote a script to emulate input and click to fetch the coordinates for a list of locations. The script is in data/latlong.net.js, where the first part is used to load jQuery and has to be run first separately. After a few seconds, we can paste the remaining part of the script into the console in Chrome. Then we set the city_names global variable to a list of locations we want to get coordinates for and invoke start_first(). After a while, the results global variable is populated with a list of triple of location name, latitude and longitude.
A note to the script: there has to be a wait after both input to the location name box and clicking the find button. I don't know if this is due to the lag of my system or network issue or google map needs more time to load or their bugs. Now I set both waits at 5 seconds. On my own laptop, this allows me to fetch more than 300 locations in a row without any error. In case an error occurs, an alert dialog (which is part of the website's design, not mine), will pop up saying invalid location. When this happens, the current location in the input box has better to be recorded so that we can come back to do it again (manually of course). The remainder of the script will typically run as usual. From my personal experience, when the number of location gets near 400, the error starts to pop up every 1 or 2 locations. I had to manually correct about 40 locations.
We then joined the coordinates with the data. Some of the data are in Puerto Rico, which cannot be correctly displayed by d3.geoAlbersUsa(). We simply filtered out those records.

4 Screenshot before milestone (Nov 10, 2017)

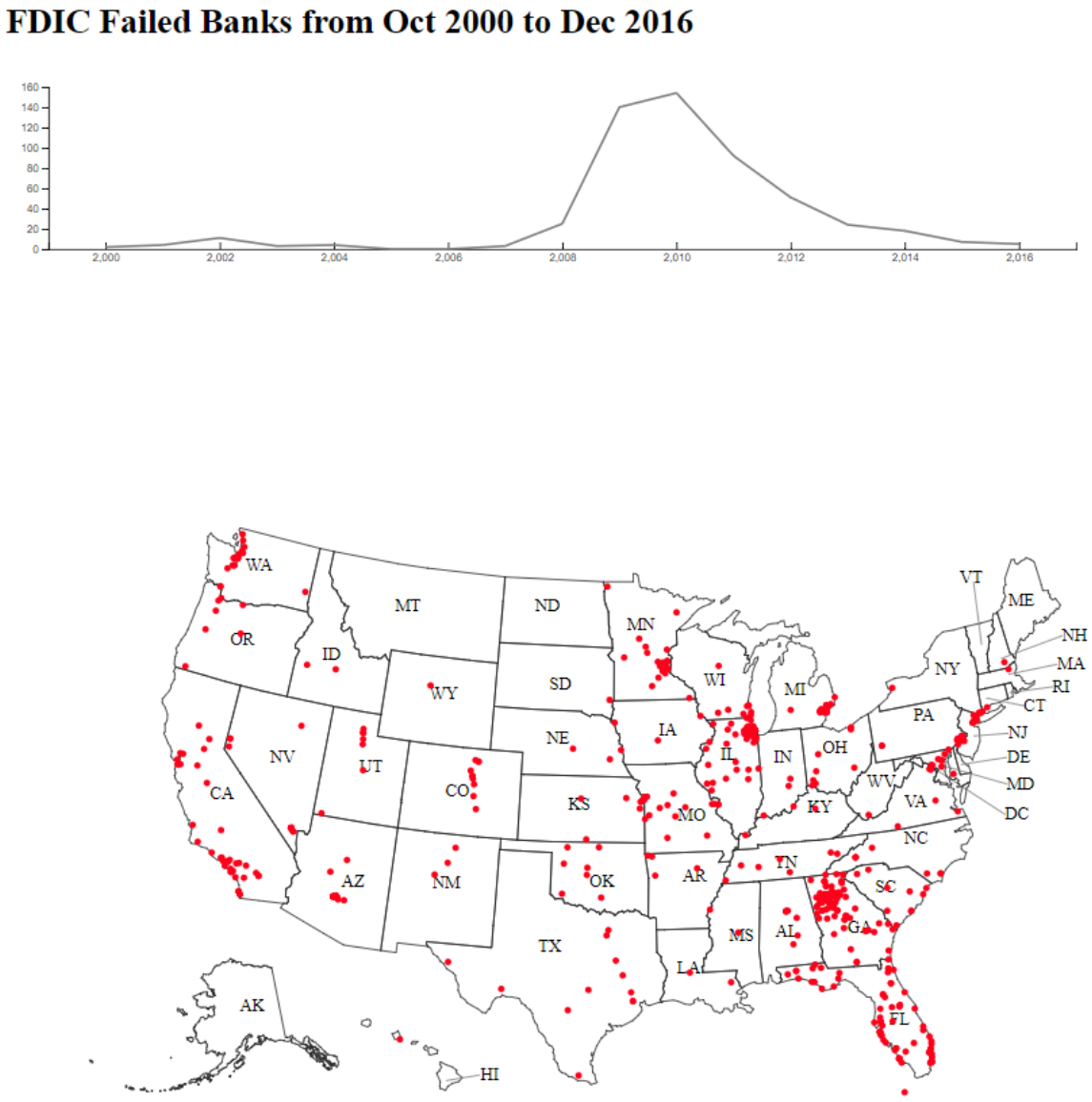

Figure 5: A screenshot for the milestone as of November 10

Now, this is what we have (see Figure 5) as of the milestone (Nov 10, 2017). We have the line chart and the map set up with the bar chart yet to complete.

The line chart still doesn't have a drop-down list to select the y-axis, which is currently the number of failed banks in a year. We plan to add a dots to the line chart so that a user can easily select one or some of the years. We also plan to add tooltips to each dot so that the exact number is easy to read, as some of them are quite small compared to those of 2008-2010.

The map hasn't been linked to the line chart and only plots the full dataset. As shown in the figure, some areas has dense population of failed banks, making it hard to use tooltip to show the details. And some of the banks are in the same city (e.g., surprisingly 19 in Chicago, IL, 10 in Atlanta,

GA, while New York City only has 2), which also makes tooltip infeasible. Instead of tooltips, we plan to use 2D brush to select a range of banks. And we'll show a list of those banks. The user can click on one of them to expand the detailed information about it on a separate box.
Hopefully, our work will have a functioning line chart and map by the next week.

5 Line Chart

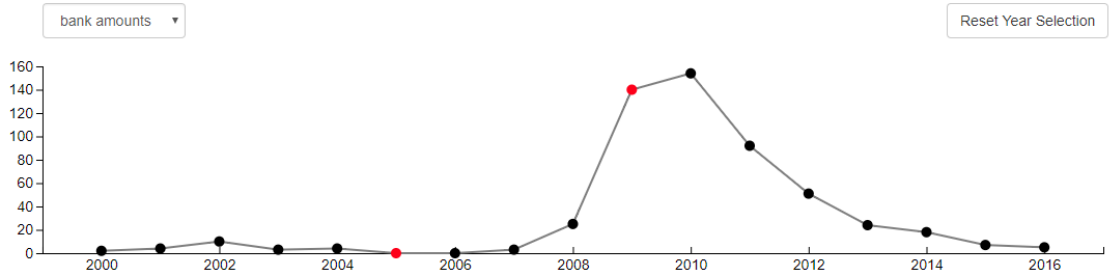

Figure 6: The line chart

The line chart (Figure 6) serves as our primary view which allows the user to have an overview of the temporal aggregates of the failed banks. This part is primarily implemented by Ya. She first implemented a line chart that features two axises with the y-axis configurable via a drop-down box, a line with scatter points that represents each years' data on it. The user can select/de-select one or multiple years to change the data shown in the bar chart as well as the map. If none is selected, all years are assumed to be selected. Compared to the 1-D brushing used in the Brazil world cup homework, it has a much better flexibility in that the selected years are not necessarily consecutive. It, however, is hard for the user to start over with the selection. In order to overcome that problem, we introduced a "reset year selection" button that allows the user the clear all the selections.

6 Map

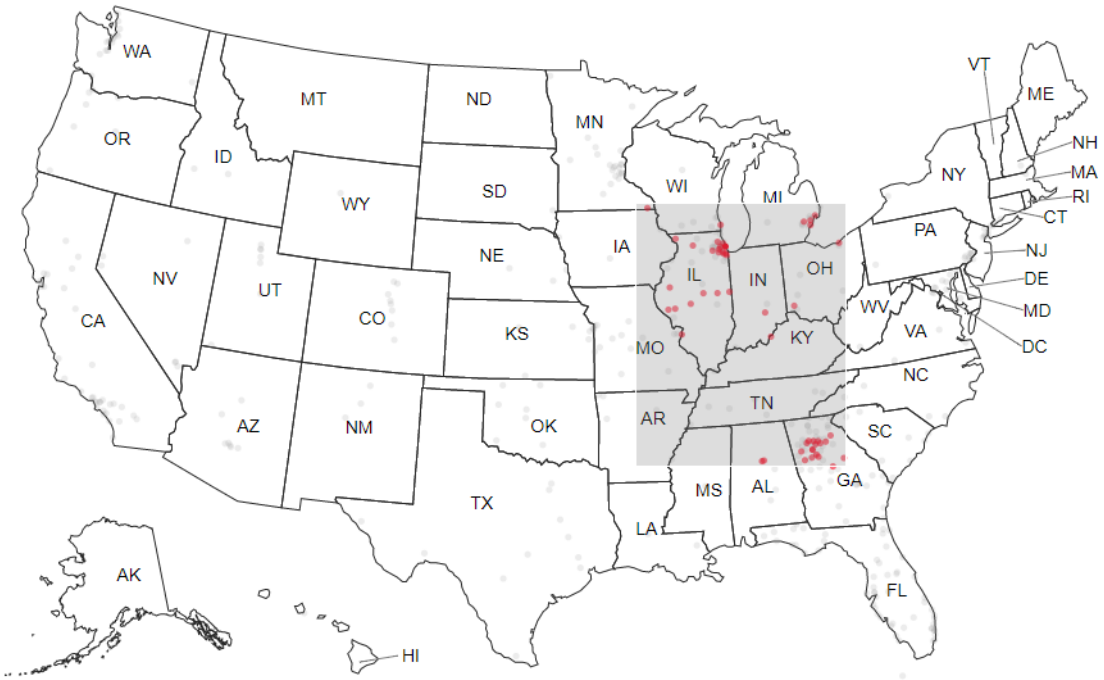

Figure 7: The map

The map (Figure 7) allows the user to freely browse and search for the spatial distribution and the details of the failed banks. This part is primarily implemented by Zhuoyue. We derived the map from the map data in the course material in lecture 9. As we mentioned before, we crawled the geo-coordinates of all the states the banks' cities from www.latlong.net. We immediately identified two challenges. First, the states on the northeast as well as Hawaii have small areas on the map, so we need to use an alternate coordinates for their names and certain states' coordinates are a bit off. We manually fixed the locations of them in the data. Second, certain cities have multiple failed banks (e.g. 18 in Chicago). Instead of plotting a point for each for bank, we grouped the banks by their cities, and plot the cities instead. We have a 2D brushing on the map so a user can select a rectangular region. If he/she is interested in a particular state or several adjacent states, 2D brushing is a good solution for such selection. We did not investigate multi-selection but it might be interesting to try to implement that if time permits. If any of the bank in the city is in the selection (according to the years in the line chart, the 2D brush on the map and the text search we will describe later), the point for that city is shown.

Initially, we simply hide those points that have no banks in the selection while the others are shown in red with no opacity. We later found out that setting the opacity of each point to some range can better represent the density of banks. The points not in the selection can be set to a very low opacity so that they can still be seen but will not attract unnecessary attention from the user.

Finally, we have a search box and a list of banks in the selection (Figure 8). The text in the search box (regardless in lower case or upper case), are searched in the names of the failed banks. A bank is said to be in the selection, if its effective year is selected in the line chart and its name contains the substring in the search box. In addition to being highlighted on the map in red, the banks are listed in the unordered list under the search box, grouped by states, where the states can be

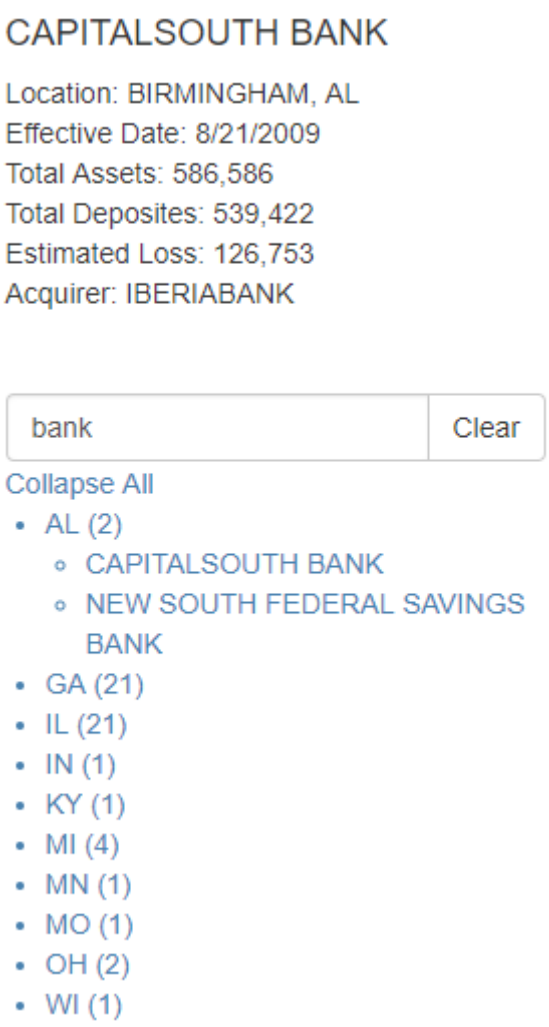

Figure 8: The search box area

expanded or collapsed. If a bank in the list is clicked, its details are shown above the search box.

7 Bar chart

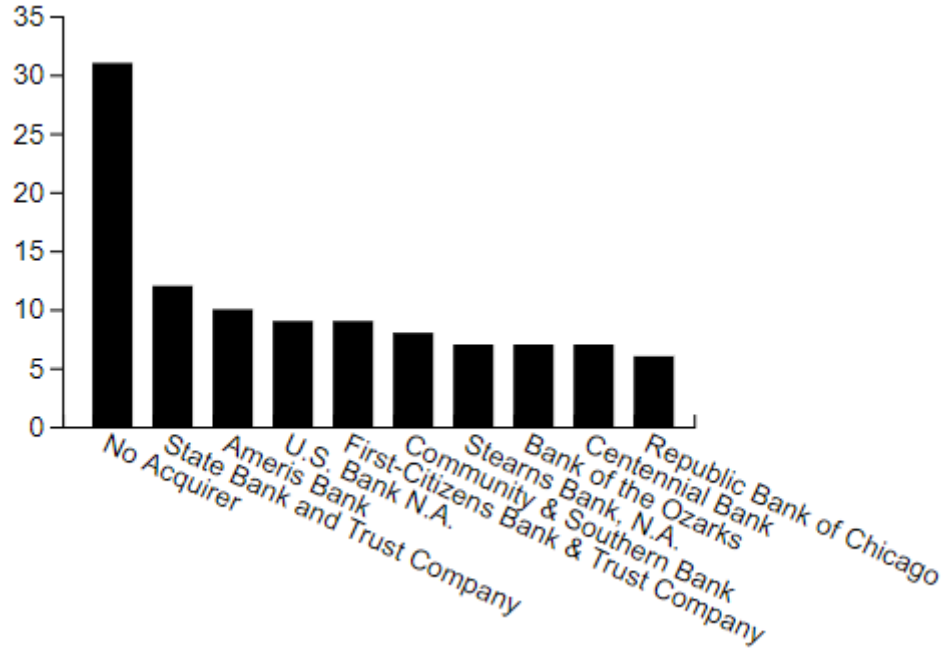

Figure 9: The bar chart

The bar chart (Figure 9) mainly show the top acquirers of the failed banks. This part is primarily handled by Ya and Zhuoyue also have significant changes to it. Since no FDIC insured funds have ever been lost, it is interesting to us to see how are the failed banks are absorbed by different entities. We limited them to at most the largest 10 such entity according to the y-axis as defined in the line chart so that we do not run out of display space. An challenge is to properly display the name of the acquiring institutions without get cut off. We find that rotating 30 degrees and setting the font size to 14px gives us the best result in terms of space compactness and readability. Initially, we designed the bar chart to only reflect the year selection but we later found that it is quite intuitive when a user is also selecting regions on the map as well as filtering the banks using the search box. So we changed the bar chart to honor all the three selections. A nice side effect of that is we can easily find the top acquirers in a specific region.
A down-side of a single view is its inability to allow comparison between different selections. An alternate approach would be allowing creating multiple views according to different selections, which, however, we do not have enoght time to explore.

8 Summary

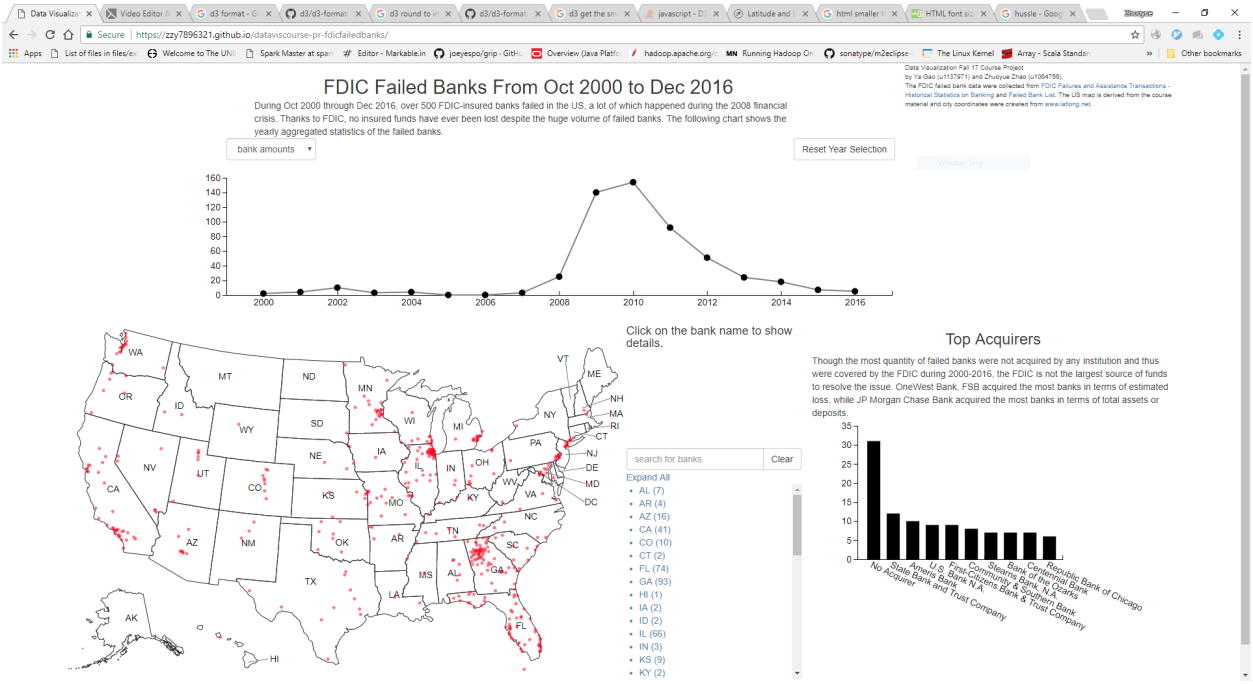

Figure 10: The final visualization

To sum up, Figure 10 shows our final visualization of the FDIC failed banks. In addition to d3 and svg for plotting charts, we used bootstrap to improve the layouts.

The main view, the line chart, is placed right below the title and a short introduction to the project. This is the place that a user might first look at. Then the map and the bar chart are placed side by side below the main view. When the user starts clicking on the years, he/she can immediately notice the linked changes below. When further selections are made on the map and/or the search box, the bar chart continues changes. We also included a short text below the title of the bar chart to give a hint of the story we want to tell.

We learnt a lot from the projects about how to design different visualizations to show different aspects of our data. We also have a better understanding in the d3 library.

As of the time we submit this work, the website is hosted on Github at `https://zzy7896321.github.io/dataviscourse-pr-fdicfailedbanks/`.