

# CS 6630 Final Project: Company Links

Alec Nelson

Yu Zhu

Yuanyuan Tong

u0741446  
u0741446@utah.edu

u1211862 yu.zhu@utah.edu

u1142943  
yuanyuan.tong@utah.edu

## Motivation

Trade and investment liberalization of the open markets could create considerable benefits for both the individuals and the societies. Public companies play an important role on the open markets, and the board is a key component of a company. Board transition happens, a board member could join a different company, and a person might be on several different boards simultaneously. We believe that the connections between companies and board members (including their demographic data) underlie rich information to be explored.

## Overview

Based on our team members' interests in financial markets and US businesses, we would like to visualize a map-based network to show how US public companies are linked from the perspective of board member transfer. Specifically, we want to populate the top 2500 public companies in the US on map and draw links between companies if there was once board member transferring between them. Besides, we are interested in uncovering how the public companies and board members' demographics affect each other. We want to give people the freedom to explore a company's basic information and its board composition thus finding interesting patterns. Furthermore, we want to group the companies into industry sectors to see how the links exist between industries.

## Related Work

Our basic visualization idea resembles migration in many senses. A board member "immigrates" from a university to a company or from one company to another. We want to visualize the links between institutions (i.e., public companies and universities) by showing a migration-like relationship diagram. So, we surveyed a lot about the migration visualization.

- <https://migrationdataportal.org/blog/10-coolest-visualizations-migration-data> These are ten interesting examples for visualizing migration data.
- <http://metrocosm.com/global-migration-map.html> Much like our final decision on our main view, a map-based migration visualization.
- <https://www.peoplemov.in/#> A Sankey diagram to show the migration corridors between countries.

## Questions

- What is the geospatial distribution of public companies in the US like?
- What kinds of connections exist between public companies in terms of board member transferring?
  - Are these connections correlated to the scale of companies (in terms of market cap, number of employees and revenue) ?
  - Are these connections correlated to the industry sectors of them?
- Are connections (in terms of board member transitioning) between companies regional?
- If board members leave their positions, do they tend to stick around and accept a new job in the same region? Do they prefer to stay in the same industry?
- Is the past education (or lack thereof) of big company board members indicative or predictive of their respective companies or industrial sectors? In other words, does a certain university contribute more to a certain company or industry?
- Is a company (or an industry) are inclined to recruit board members with certain education background?

## Data

**Data collection:** board member, education, employment history, company & university profiles

(Yu, 11/1/2019 to 11/8/2019)

In our project, we're going to show two types of links: (1) company-company, (2) company-university. These links are mainly built based on four data sets: (1) individual-employment, (2) individual-education, (3) company-profile, (4) individual-profile. Here's how I process the data.

First, we need to filter out "dead" companies and companies that are "too small." (we don't want our network to be a "giant hairball") Currently we keep the 1000 largest (in terms of market capitalization) "alive" companies as our sample (that is, companies that are still being traded).

Then we extract the coordinates of these companies through the Google Map API. The "headoffice-address" variable in the "company-profile" dataset makes it possible. This step is done by Yuanyuan Tong.

Next, we'll start to build the company-university link. A link between a company and a university is established when a board member of this company is graduated from the university. We also take into consideration the number of graduates. For example, if there is only one board member from the focal university, then the link intensity is one, if there are two graduates, then intensity will be two. We achieve

this by first merging the `individual-education` table and `individual-employment` table, and then aggregate for each unique (company, university) pair. The final output is as follows, where “lat” means latitude and “lng” means longitude.

company	university	n_grad	company_lat	company_lng	university_lat	university_lng
L3HARRIS TECHNOLOGIES INC (Harris Corp prior to 07/2019)	University of New Hampshire	1	28.09078	-80.63855	43.13895	-70.93703

Finally let’s build “company-company” links. The main logic is the same: if one person has stayed in the board of both company A and company B (not necessarily at the same time), then a link is established between these two companies through this person. The link intensity is also computed in a similar fashion as in `company-university`. We achieve this by self-joining the “individual-employment” table. We’ve not finished the final merging, but a preview is as follows.

from_company	from_company_id	to_company	to_company_id	name	name_id	role_in_frc
GREENHILL & CO INC	40757	1 800 CONTACTS INC (De- listed 09/2007)	3	Steve Key	59796	Indepen

(Yu, 11/28)

In this update, I add some statistics for each company: the gender distribution (male/female), the degree distribution (bachelor/master/PhD), the age distribution (below 40, 40 to 50, 50 to 60, etc.), and nationality distribution (US/non-US/unknown). The picture below shows some of the variables we created.

To create these variables, I first link the “company\_profile” table to the “individual\_employment” table, and then group by company id. By doing so, I’m able to get the all the board members that have ever worked in a company. Then I add individual information (age/gender/nationality) by merging the current table to the “individual profile” table. At last, for every company id, I calculate its statistics, like age/gender distribution.

degree_phd_pct	degree_others_pct	male_pct	female_pct	us_pct	non_us_pct	na_nationality_pct	age_below_40
0.11111111	0.1666667	0.7777778	0.2222222	0.4444444	0	0.5555556	0.11111111
0.17647059	0.5294118	0.8235294	0.1764706	0.5294118	0	0.4705882	0.00000000

na_nationality_pct	age_below_40	age_40_to_50	age_50_to_60	age_60_to_70	age_above_70	age_unknown
0.5555556	0.11111111	0.33333333	0.2222222	0.27777778	0.05555556	0
0.4705882	0.00000000	0.05882353	0.2941176	0.47058824	0.17647059	0

(Yuanyuan, 11/4/2019)

Since we want to build a map-based visualization, the very important attributes we need to populate the nodes on the map are their coordinates (i.e., latitude and longitude). We could acquire the addresses of the public company headquarters from our datasets, but the important thing is to convert these textual addresses into numeric latitudes and longitudes. We ended up using Google's Geocoding API to complete this conversion.

Geocoding is the process of converting addresses to geographic coordinates. This is just what we need but it is not totally free. However, considering the scale of our dataset (about 8000 companies and 30,000 universities), the free trial will likely cover our requests.

I implemented an html page and a js file to do such conversion. Though the platform also provides SDKs for development in different languages, I stick to JavaScript because of our familiarity with it in this course. However, I noticed a weird pattern when my website did the geocoding requests, Google will restrict the request frequency and somehow reduce your request speed. That means even if you set a decent speed to request the geocoding services within the usage limits, the actual speed you will get valid responses becomes slower and slower by time anyway. Because of this limitation, we only succeeded in getting the coordinates of about 1000 companies and 400 universities by far. But we think this is well enough for our demo at this point.

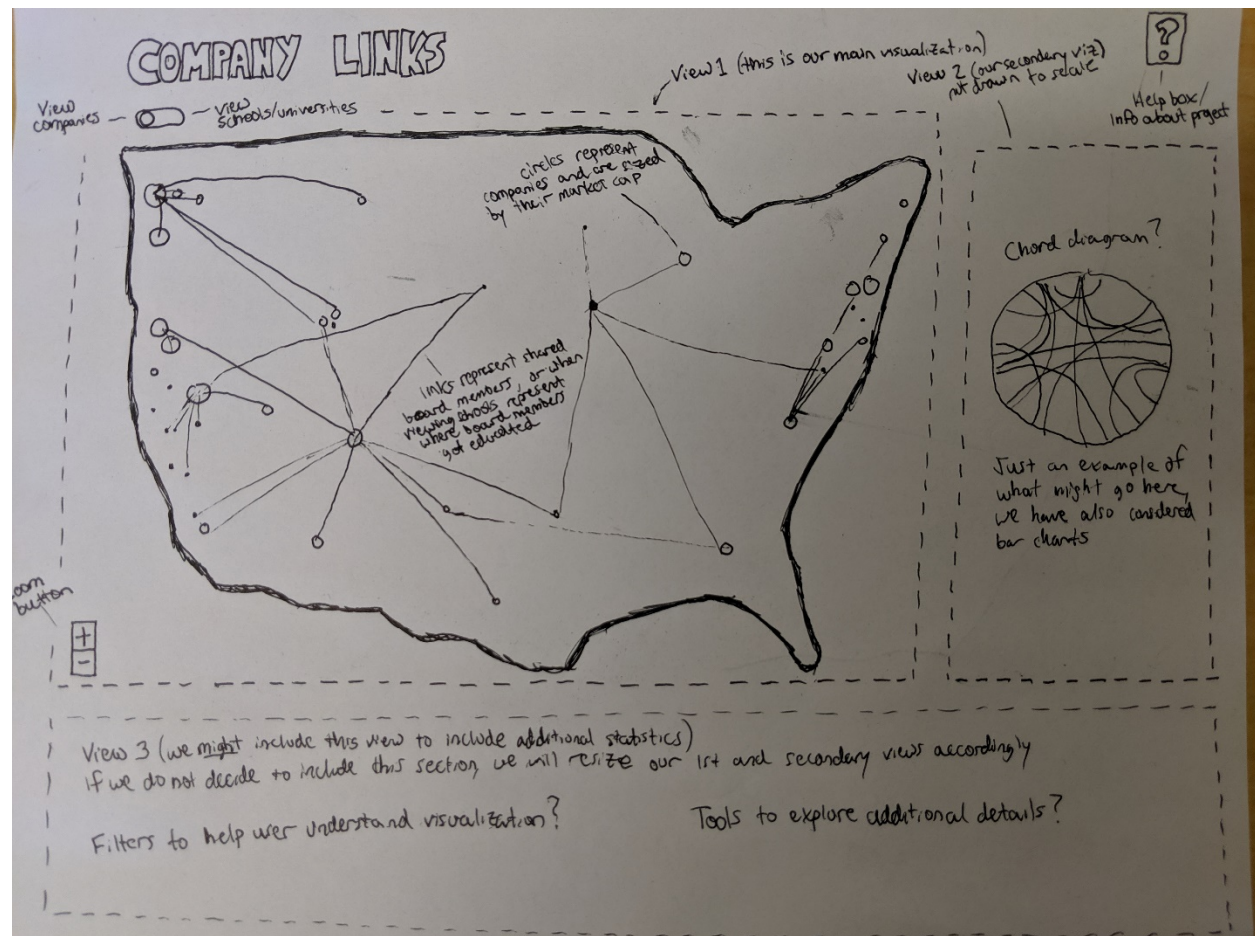
## Preliminary Data Analysis

Honestly, due to the scale and complexity of our raw dataset, it is very hard to do any quick data analysis at first. Most of the things we did were viewing the data fields manually. Especially when the data was distributed by geo location, it was too hard to notice anything unless we populated them properly.

However, after combining information from other dataset, we noticed that there might be interesting patterns about the demographics of the board members. After we sorted the data by different demographic attribute, there are quite a few extremes of different companies.

## Initial Design

Basically, our final project stick with our initial idea in the sketches.



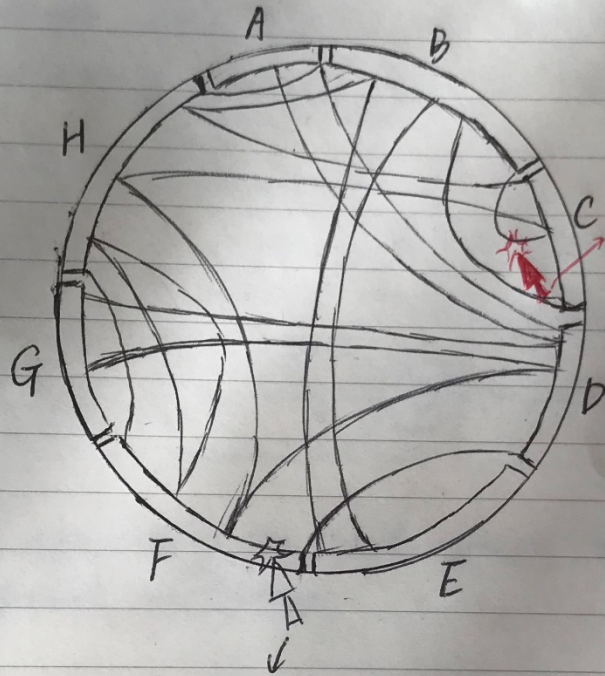
Above is our main view where the map dominates! We want to explore geospatial things and relationships between companies.

And below is our second view, which is mainly a chord diagram. We want to show the relationships between industries with the chord.



## View 2 in detail

This is basically to show information less relevant to geographical elements, yet can be linked with the main view.



clicking on arc:  
highlight corresponding flows in View 1.

upper case letter  
↓  
industry

clicking on node F:

works like a filter, will filter companies of industry F in View 1

This chord<sup>diagram</sup> could show the personnel flows between industries (instead of between companies).

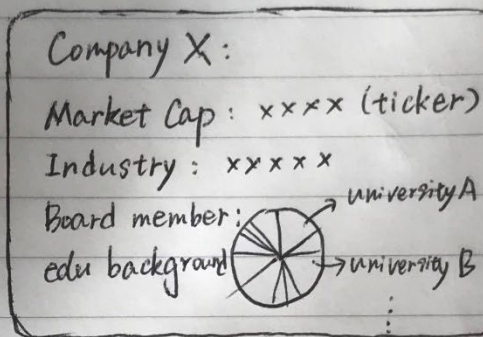
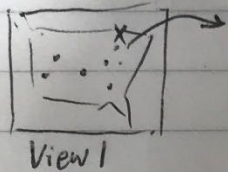
As a whole this diagram would answer whether a board member would stick on a fixed set of industries or not.

Finally, it is our idea about showing the information of a selected company.

### View 3 in Detail

We haven't gone deeply into our datasets yet, but we believe there are many interesting stats to explore

For example, when clicking on a company node X in View 1, besides showing a company profile, we could show a pie



(onclick: panel)

chart representing the board members' ~~big~~ educational background constitution.

## Reflection on TA(Ilkin)'s Feedback

I am not following how the motivation and the objectives part are related. Moreover, I don't think that your current design addresses all of the questions raised in your objectives section (such as question two).

- Yes, we noticed the gap between our motivations and the objectives because we were mentioning too much on stock prices. We changed our wording focus more on the scale of a company, like "market cap", "revenue" and "number of employees".

I think once you create your dataset, there will be lots of nodes and edges. And when you try to visualize it on the map, it will be really cluttered. You can add some kind of filtering or show a limited number of companies on the map. Or show the links only for a selected company.

- Yes, eight thousand companies are too many to show. We finally tried to limit to the top 2500 public companies.
- We now can filter companies by industries. Other filtering could be considered in the future.
- We will only show the links for a selected company.

The chord plot will be good for revealing the cross-industry relationships. You can add interactive highlighting between nodes and links in the map and chord plot and vice versa.

- Will do.

Your data has a lot of information related to people. You can visualize some/most of these attributes in the third view (average age, nationality, time in company, distribution of roles and etc). It would also be interesting to visualize some basic network statistics such as total number of incoming/outgoing.

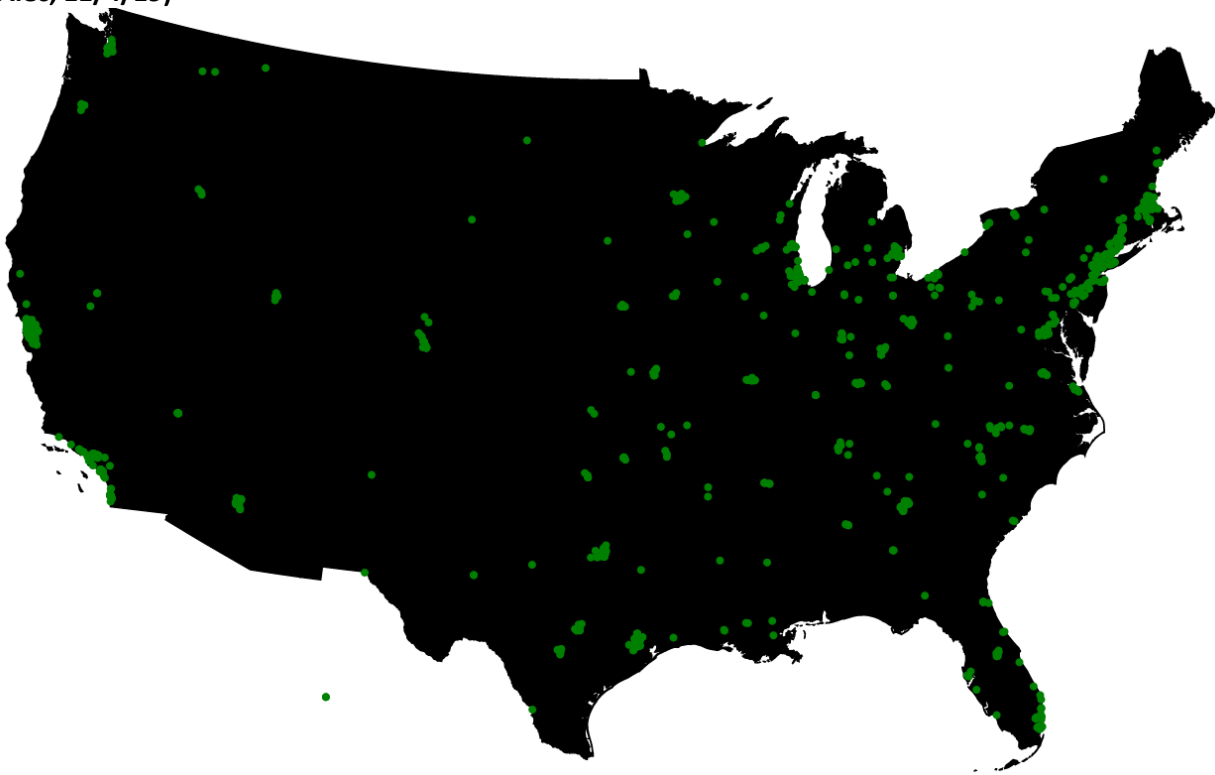
- Good point, still considering what extra could be included, we need to discuss it with Ilkin in person.



## Implementation

**Map is up and running!**

**(Alec, 11/4/19)**

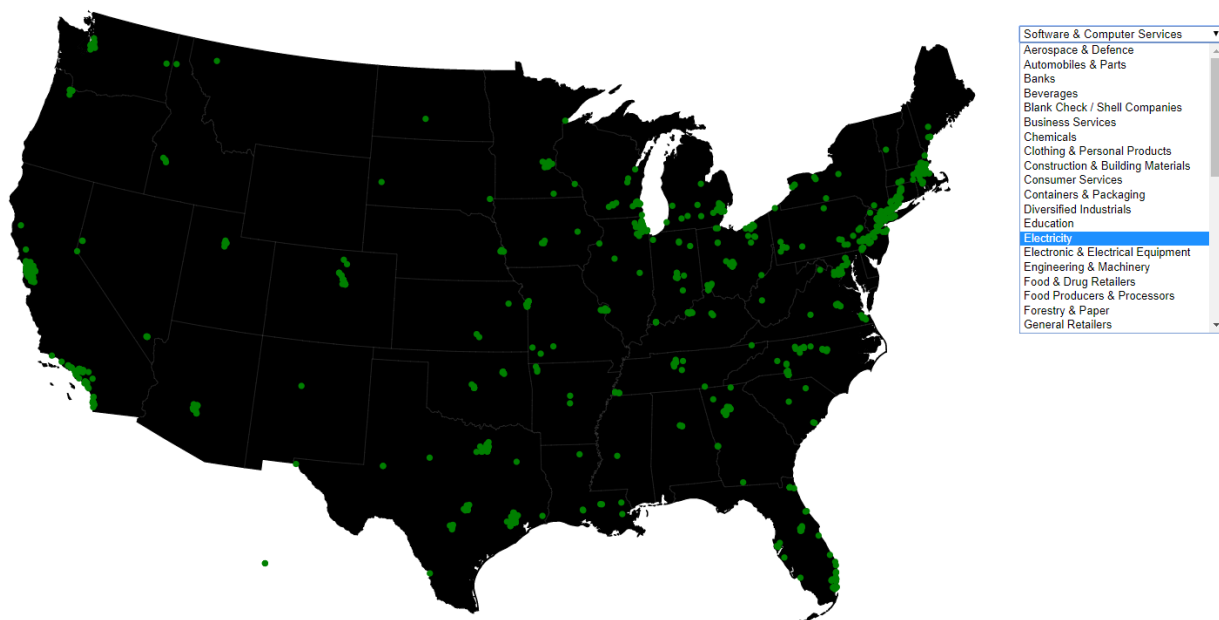


We now have an interactive map! It is exciting to be able to see some of our data now, before the map was up and running. I was a little unsure if there would be any obvious patterns that emerged. I had a hunch there would be quite a few companies in the San Francisco area- a suspicion that was confirmed. I did not realize just how many companies there would be in New England! It is becoming apparent that occlusion will be an issue. I am still not totally sure how we plan to tackle this, we have discussed force directed nodes, but that would change the location of the node, potentially to a large degree in more dense regions.

I sketched out an idea that we talked about the other day above the map. We need to implement some sort of toggle or switch to be able to alternate between viewing universities and companies on the map, as we suspect viewing both at the same time will cause significant crowding and may be too cluttered.

**Added a dropdown menu**

**(Alec 11/5/2019)**



Yu was able to successfully add categorical/business sector attributes to our dataset. This is a big step forward because we can use this information to filter companies, since we plan on displaying a large number of them on the map at once. One thing that is kind of interesting is the narrowness of the company categories (they are incredibly specific). Currently the dropdown doesn't do anything- I actually am a little unsure if I like it because you have to scroll through categories, rather than being able to see them all at once and compare them against one another. This may be something that I try and tackle tomorrow.

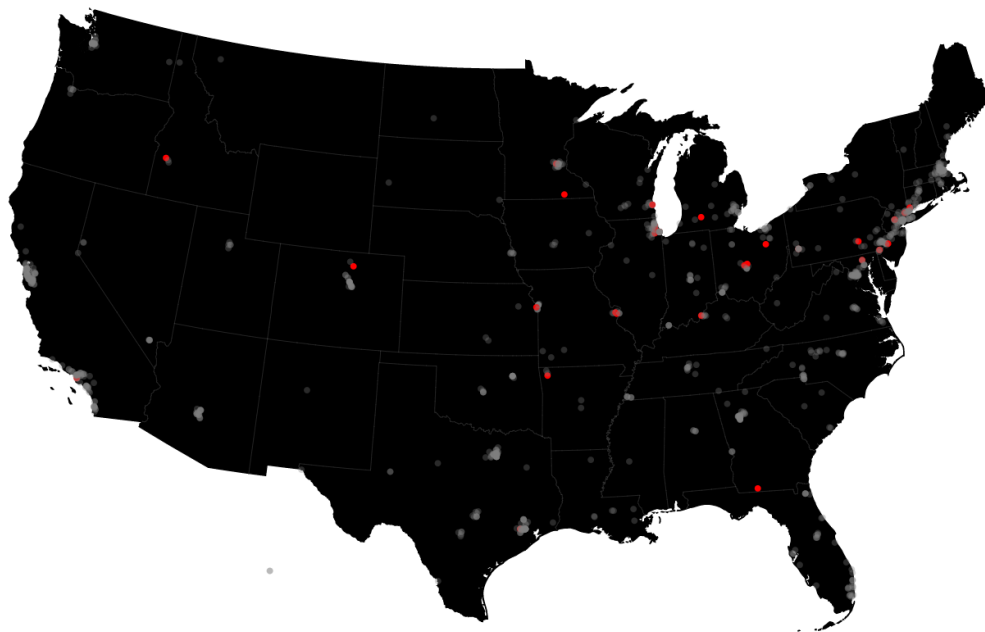
Also, sooner rather than later, we will have to start making some design decisions regarding the display of the nodes on the map. They are currently green because I spontaneously colored them that way, not for any particular reason. Do we need color at all? I remember Alex saying that he tries to minimize colors in his own visualizations, saving their use for highlighting. I will probably try playing around with this in my next iteration.

### Changed dropdown to table

### Colored nodes gray and dropped opacity

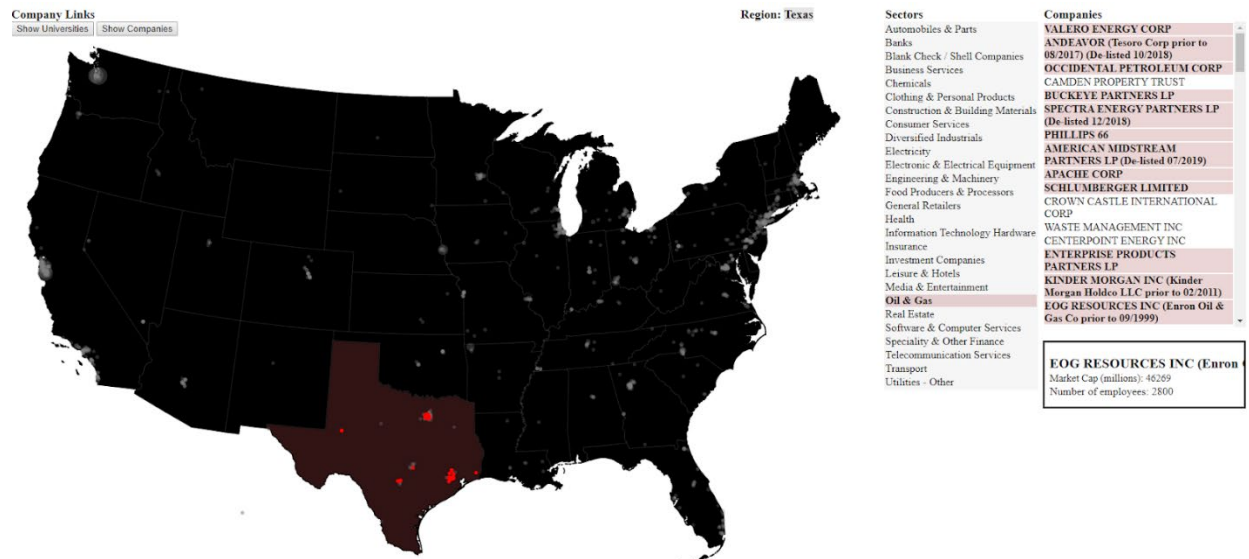
(Alec 11/6/19)

I changed the dropdown menu to a single-column table that is fully visible (you won't have to scroll it). Now when you hover over an item in the table, it highlights those companies in that sector on the map. ***What if when you highlight a node on the map, it highlights that word in the table, and then vertices expand from that word, linking it to another table of data to the right? Continuing to drop node opacity as more elements are added will allow map to scale really large (not much visual clutter) and essentially create a detailed heatmap.***

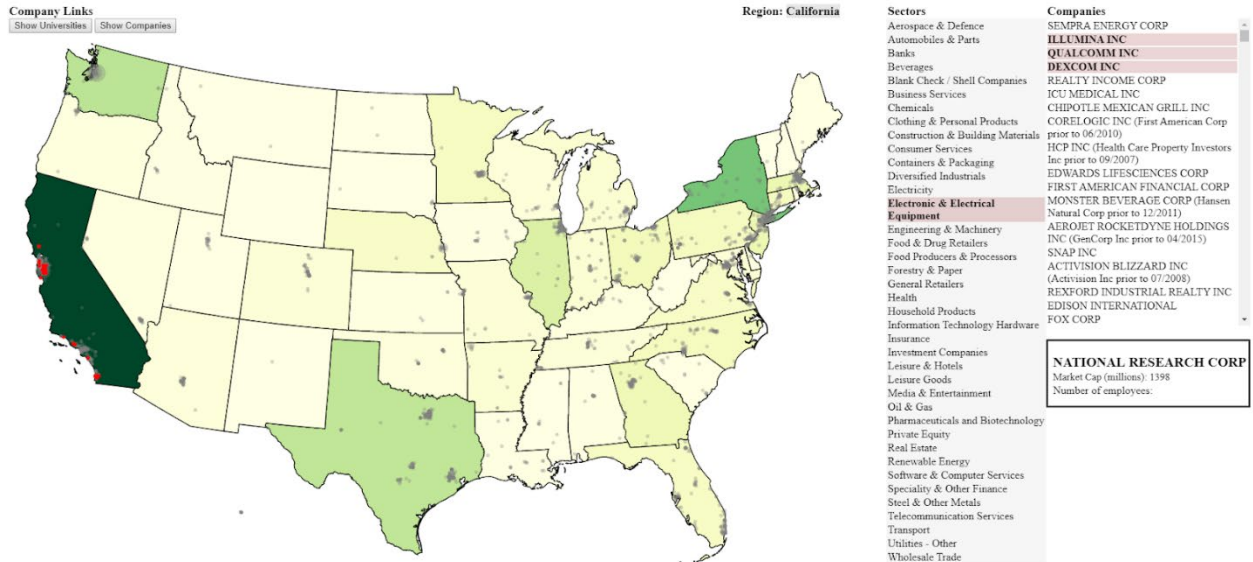


- Aerospace & Defence
- Automobiles & Parts
- Banks
- Beverages
- Blank Check / Shell Companies
- Business Services
- Chemicals
- Clothing & Personal Products
- Construction & Building Materials
- Consumer Services
- Containers & Packaging
- Diversified Industrials
- Education
- Electricity
- Electronic & Electrical Equipment
- Engineering & Machinery
- Food & Drug Retailers
- Food Producers & Processors**
- Forestry & Paper
- General Retailers
- Health
- Household Products
- Information Technology Hardware
- Insurance
- Investment Companies
- Leisure & Hotels
- Leisure Goods
- Life Assurance
- Media & Entertainment
- Mining
- Oil & Gas
- Pharmaceuticals and Biotechnology
- Private Equity
- Publishing
- Real Estate
- Renewable Energy
- Software & Computer Services
- Speciality & Other Finance
- Steel & Other Metals
- Telecommunication Services

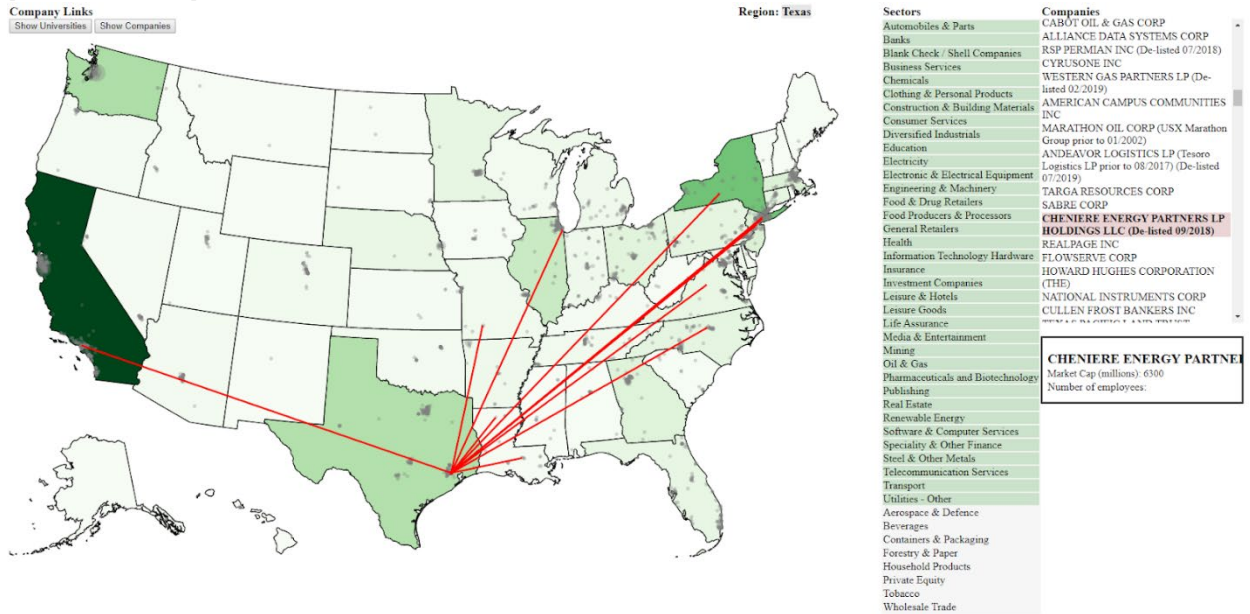
## Added Company Table (Alec 11/11/19)



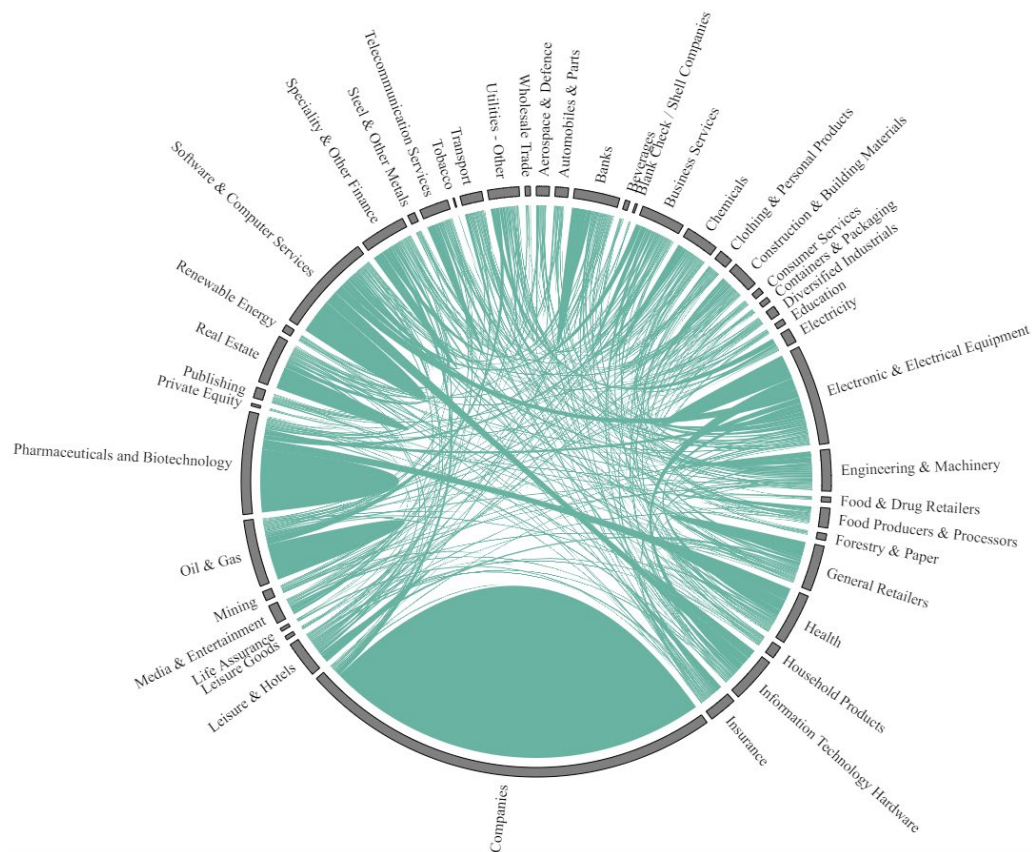
## Colored states by total market cap [Alec 11/13/19]



## Show links to universities on company hover, Highlight sectors that are present in selected state [Alec 11/18/19]



## Chord diagram showing migration between sectors [Yu, Yuanyuan 11/21/19]



## Interactive Zooming (Yuanyuan, 11/20/19)

To make the map interactive enough, I implemented the zooming function of the map. People now could zoom and pan the map!

Later on I realized, the only reason why people might want to zoom in an area of the map is they want to see some details of the companies in an area. And that area is most likely to be a state. So I implemented the zoom function on clicking an state. The view would zoom into a clicked state and you can step out of the state and go back to the whole US if you click on it again.

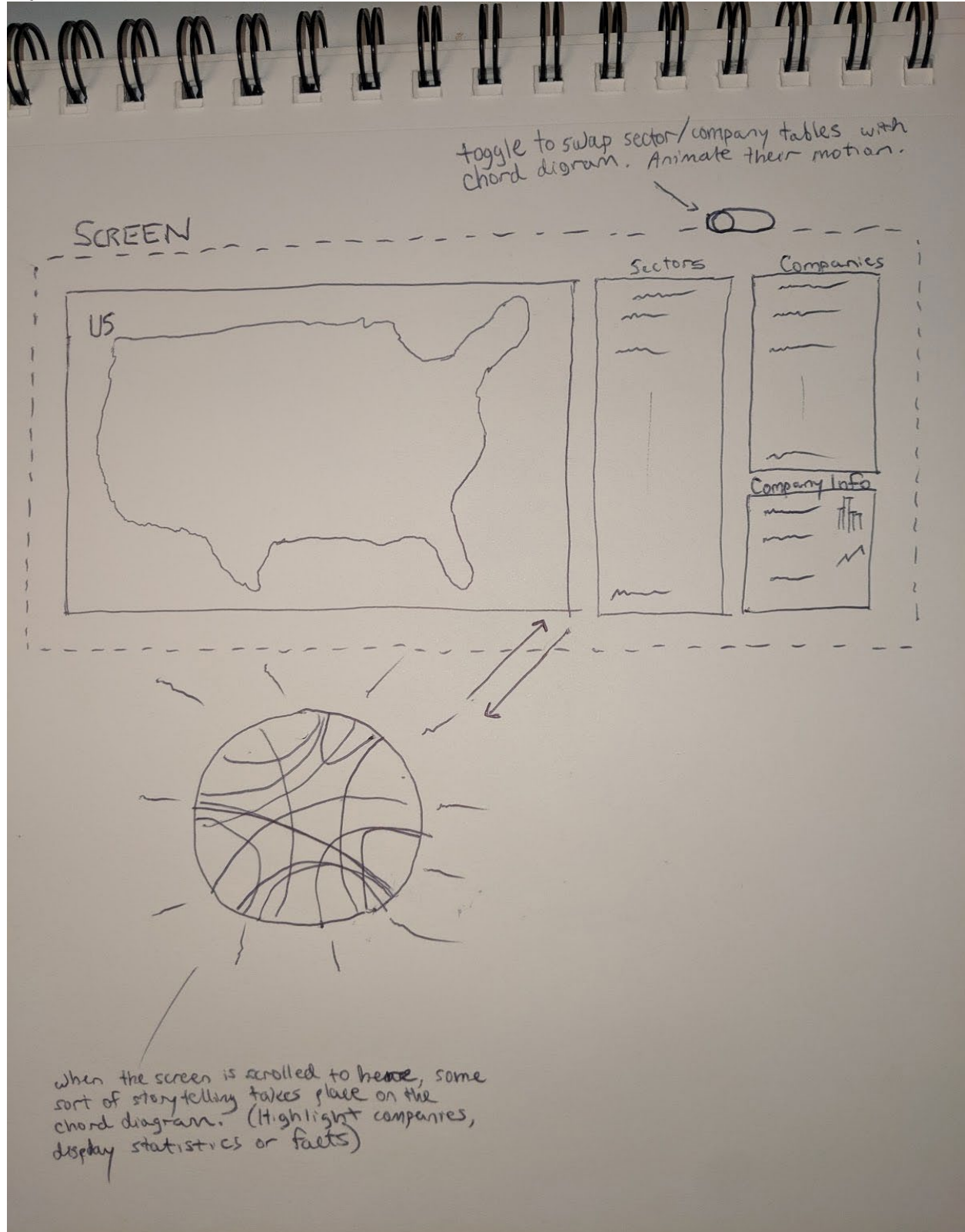
## Hovering function (Yuanyuan, 11/24/19)

The chord diagram is somehow too clustered. People might have no clue about the detailed information. So, I tried to implement some highlighting on hovering. At last I chose the d3.filter to make the selected ribbon “glow” a little bit, because I think it is the best styling way to not mess up with the key element.

From there, I also implemented tooltips for both the companies on map and ribbons on chord.

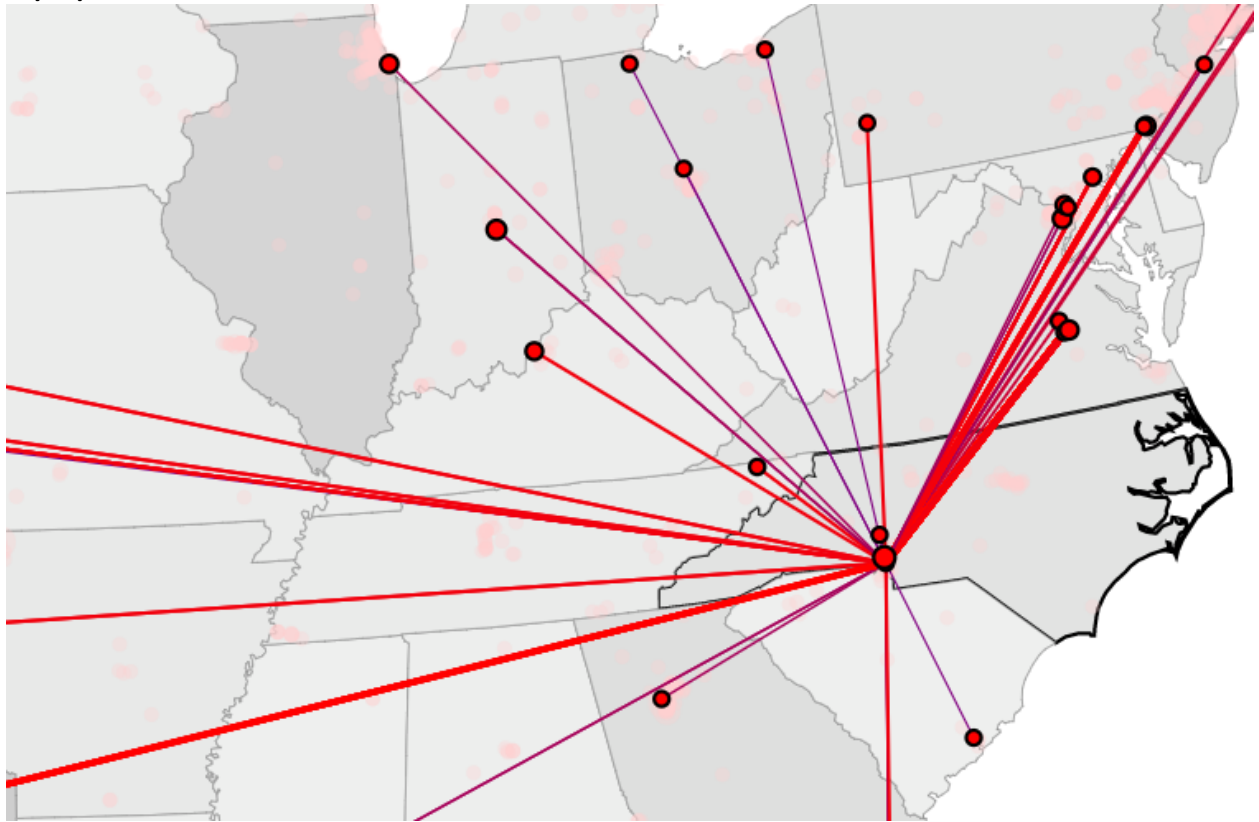


Layout idea [Alec 11/22/19]





11/26/19



Welcome Screen  
[Yu, 12/1/19]

### InfoBox with pie charts (Yuanyuan, 12/1/19)

Yes, I finally finished the information box for a company. When you hovering on a company either on the map or on the table, the InfoBox will update the information of that company.

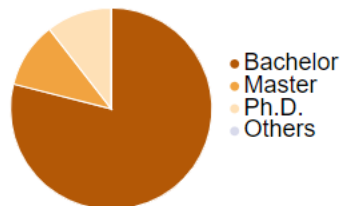
#### AARONS INC

Industry Sector: Consumer Services  
Market Cap (millions): 4281  
Revenue (millions): 3829  
Number of Employees: 11800  
Male vs Female (of the board): 89.47% vs 10.53%

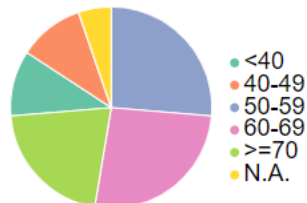
#### Board Composition

MBA

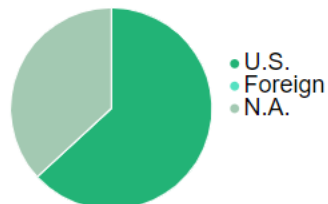
5.26% of the board members hold MBA degrees  
Education



Age



Nationality



## Evaluation and Improvements

Overall we are pleased with the dynamics of our visualization. The dataset was quite large (contained over 2500 companies), so we are pleased that we found a way to be able to explore every individual company on a map. Specifically, we feel the linked table is clever solution to overcome the problem of occlusion, allowing you to 'click' on a company, even though it may be obstructed by another one.

We were surprised to notice that the dominating place of California in terms of both the number of public companies and the aggregate market cap. While the Midwest seems like to be much more "barren".

Banks sector has the most public companies. And the distribution of Bank companies is more even across the country than other industries.

From the chord diagram, we also found board members do like to stick on their own industry sectors if they want to seek opportunities in a different company. Board members from Investment sector seem to be most "adventurous" because there are most people ever tried to go to a new board, though they still preferred to stay in the same industry.

Due to the time limitation, there are many possible improvements of our project. Improvements are not limited to:

- The layout and styling could definitely be done better.
- The loading of the data causes a time delay on our page now. We hope to learn more about the async function and make some loading pages before the datasets are fully loaded.
- We did not get the time to explore the relationships between companies and universities.