- Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

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.

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 members. Besides, we are interested in uncovering how the companies and board members' degree offering universities affect each other. Furthermore, we want to group the companies into industry sectors to see how the links exist between industries.

- Related Work: Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.

(Yuanyuan)

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 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 questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?
    - What is the geospatial distribution of public companies in the US like?
    - What kinds of connections exist between public companies, and can these connections be correlated to stock prices?
    - Are connections 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 industrial sectors?
    - Does a company (or an industry) have an inclination to recruit board members with certain education background?

- Data: Source, scraping method, cleanup, etc.

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 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 form 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 CHNOLOGIES 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_fr
GREENHILL & CO INC	40757	1 800 CONTACTS INC (De-listed 09/2007)	3	Steve Key	59796	Indepen

(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 a 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.

- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
- Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?
- Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.
- Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Initial feedback from our TA (Ilkin). We have not got the chance to meet with him in person, but we got the feedback on Canvas.

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

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 and thirty thousand universities are too many to show. We try to limit the number of them, but have not decided on the specific number.
- We now are able to filter companies by industries. Other companies could be considered in the future.
- We haven't considered to select companies only, because we are afraid the links between several companies would be too sparse.

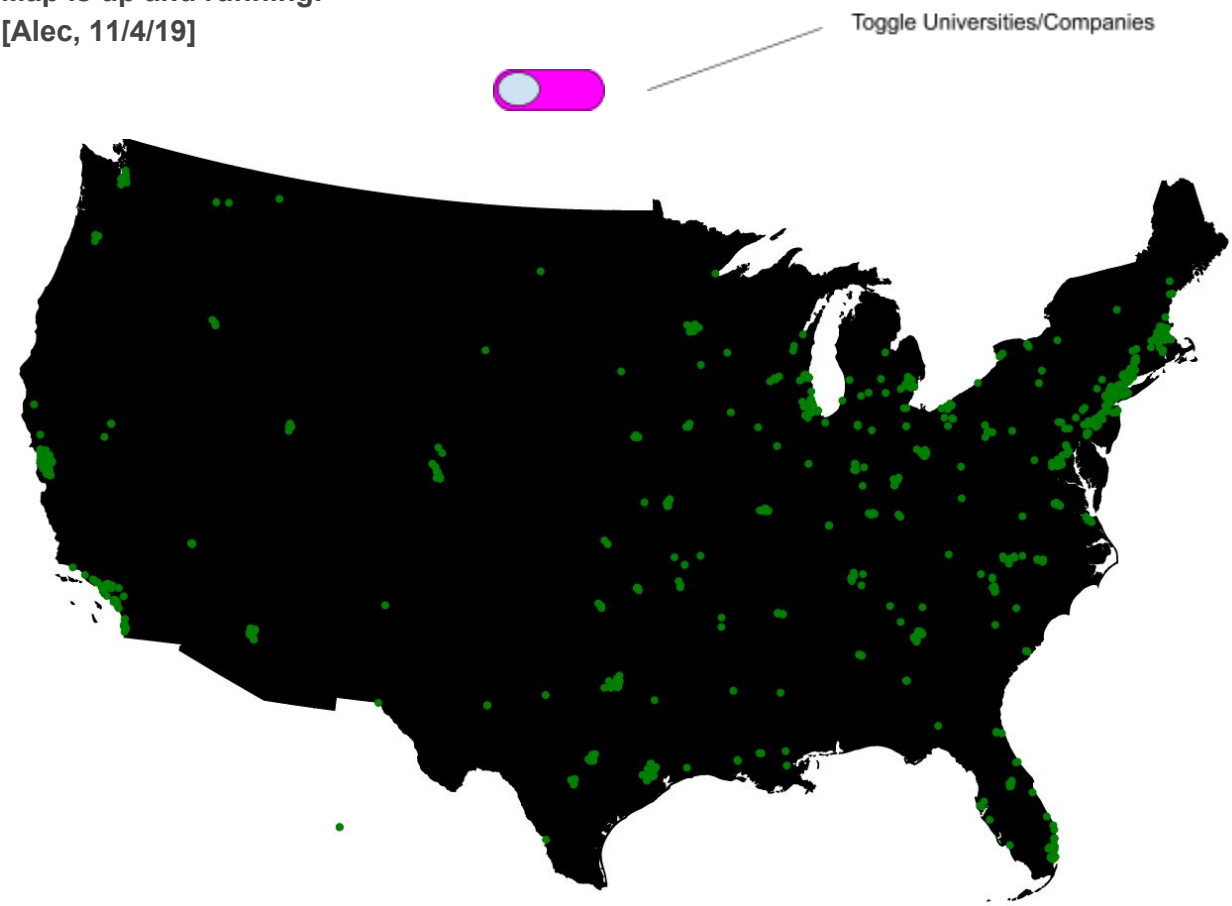
The chord plot will be good for revealing the cross industry relationships. You can add interactive highlighting between nodes and links in the map and chord plot and vice versa.

- 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.

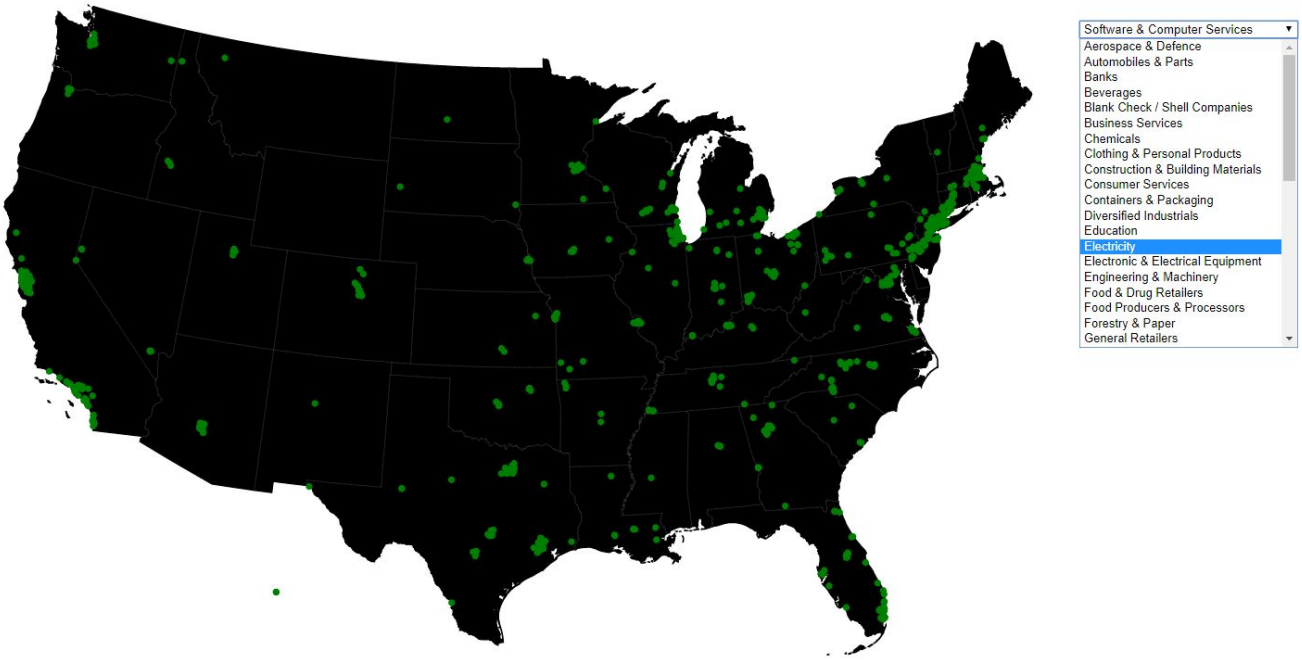
Map is up and running!
[Alec, 11/4/19]



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]

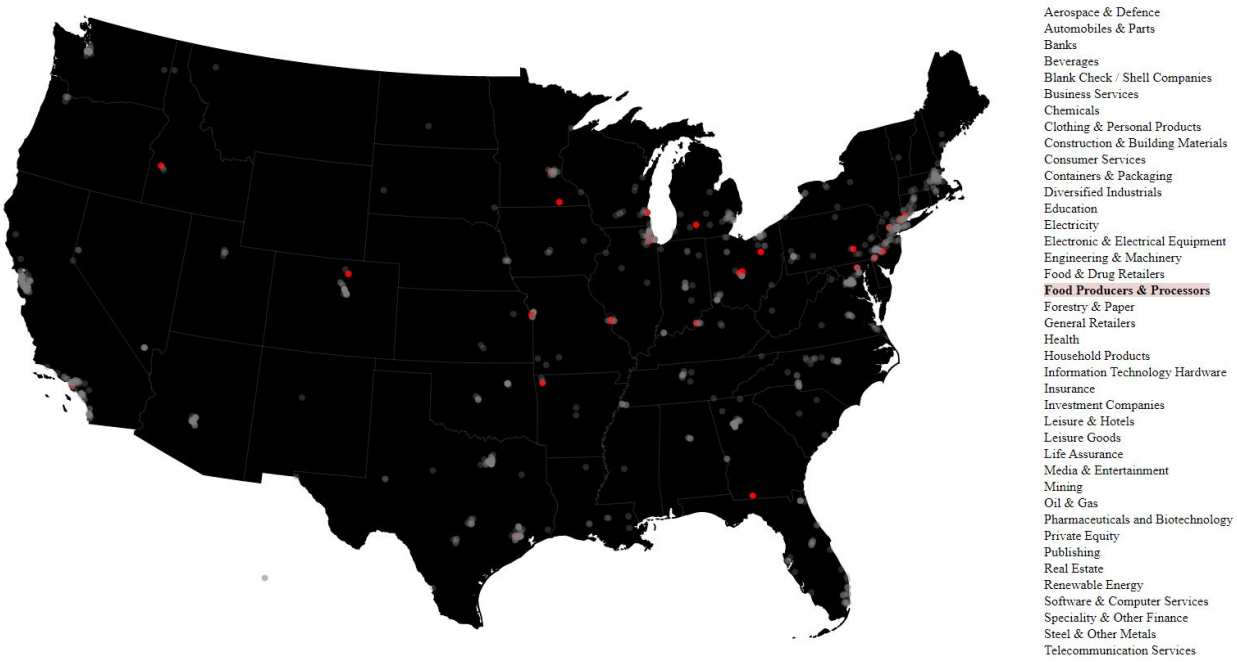


Yu was able to successfully add categorical/business sector attributes to our dataset. This is a big step forward because we can use this information to filter companies, since we plan on displaying a large number of them on the map at once. One thing that is kind of interesting is the narrowness of 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.**