

# **CS171 Final Project**

## The Changing Landscape of the American Economy

---

A Process Book

By Katherine Dai and Patrick Xu

# Table of Contents

---

- I. Initial Proposal
  - II. Design Studio and Milestone I
  - III. Revised Proposal
  - IV. Data
  - V. Components
  - VI. Visual Design
  - VII. Conclusion
-

# I. Initial Proposal

## Background and Motivation

In 1908, Henry Ford began producing and selling the Ford Model T, decimating the formerly robust buggy and whip industry. More recently, Amazon has displaced booksellers and companies in the technology space have dominated the economy in both in terms of overall market cap and growth. We are interested in seeing how the American economy has evolved over the years, both in terms of what the most successful companies do and how big they are.

## Project Objectives

By analyzing trends in the companies in the S&P 500 market index, which is composed of the 500 largest companies by market cap on the NYSE, we hope to not only visualize the overall growth in the American economy but also to break down the components of this growth.

- How big is the American economy each year?
- What are the ten largest companies in the economy for each year?
- What do these companies do?
- Where are these companies located?

There are several benefits to this project. First, we will learn how to geomap data and create a visualization using a new structure. Additionally, we'll be able to strengthen our API manipulation skills. Finally, we'll be able to discern trends in industry movements and get a better understanding of where the American economy has historically been and what it seems to be trending towards.

## Data

We are planning to collect our data primarily from Baker Library, where we can gain access to historical S&P 500 indices and to the Wharton Research Data Services (WRDS) to collect company financial information. Additionally, we plan on using the Yahoo! Finance API to help better categorize each company by industry and sector.

## Data Processing

Although we are getting data from multiple sources, the actual cleanup process should be relatively painless. We plan on grabbing the top 500 companies by year according to the S&P 500, from 1955 (or potentially earlier) to current day, and putting it in CSV format. From Yahoo! Finance, we will retrieve the industry for each of these companies, as well as the market cap during the fiscal year; this can be directly downloaded in CSV format. Finally, WRDS will provide us the headquarters of the companies, from which we will likely only store the state name.

## Visualization

There are a number of features that we are planning on implement to visualize the S&P 500 and its historical growth. First, we want to represent current and historical market cap with a line chart of the S&P 500. This line chart will also include interactivity - users will be able to focus in on a year to better understand growth and market cap, and hovering over any year will reveal a tooltip with numerical data on the S&P 500. Next, we want to focus in on the top 10 companies in the S&P 500 (by market cap), and create a bar chart that will show each individual market cap. Additionally, we want to color the bars based off of their industries. Finally, we want to geomap the companies by state onto the US. We want to see how the headquarters of the S&P 500 companies have changed over time.

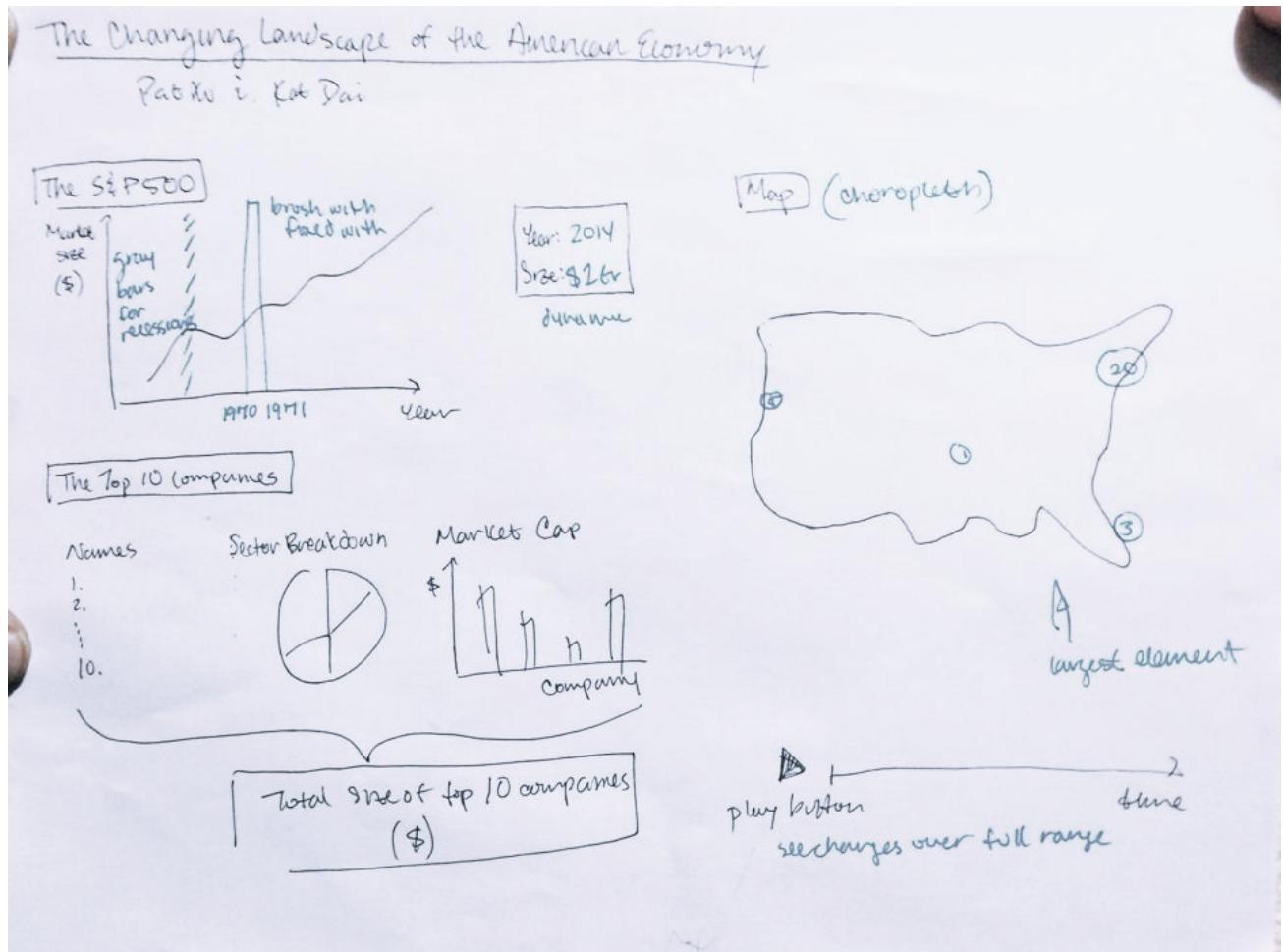
## Must-Have Features

- The line chart with zoom in and tool tip capabilities
- Bar chart that has industry/market cap information
- Map of the US that depicts number of companies by states

## Optional Features

- Features that would be nice to implement:
- Play button that shows changes over time
- Companies mapped by cities rather than states
- Pie chart that shows industry breakdown
- Markers for recessions

## Initial Sketch



# II. Milestone I and Design Studio

## Feedback from: Karine Hsu and Alex Liu

Given that the project is mostly bare bones, the feedback from Karine and Alex consisted mainly of answering the questions of who benefits from the visualization as well as some minor technical issues to look into.

First, the audience of our visualization is a bit nebulous at best. Although we generalize it to “anyone interested in learning about the American economy”, Karine suggested that we include elements of recent interest to pique curiosity about our topic – more specifically, she suggested that we integrate news stories, especially those pertaining to the most significant economic crises and movements in the economy. This could be achieved by using tooltips, similar to the work we have done in previous problem sets, but also by trying to analyze the data at hand further to discover more trends (for example, the rise of technology companies).

Karine and Alex also made the recommendation to better use color and interaction to unify the visualization. Our information is currently encoded in various graphs and charts but don't seem to be part of the same dataset, especially since they each display different facets of the data. Several solutions suggested during the design studio included using one color legend for industries across the entire visualization, and to use a more coherent theme for the project (one complaint is that the white background is not particularly inspiring). Alex also helped me start the search for an inspirational visualization, which is a good direction for us to go.

The feedback was very helpful in two regards. First, Alex and Karine helped me think of more solutions to the problems I had been staring at for several weeks without an outside perspective (Patrick is currently in London). Next, the feedback forced me to think of new directions to take the project, which I have shared with Patrick. I am excited to see how much the visualization can develop because of this design studio.

## Concerns at first milestone

- Data readability
  - There's a high density of graphs on the screen, which makes the overall visualization seems a bit fragmented (though interactivity will help increase the sense of cohesion)
- Emphasizing movement
  - one thought was to implement a visual element that shows the path of a company moving from New York to California

# III. Revised Proposal

## **Fortune 500 rather than S&P500**

Some companies in the Fortune 500 are privately-owned, and thus would not be included in a stock market index. Additionally, changes to the “basket” that makes up the S&P500 are not tracked, and changes are not made on a preset schedule. The Fortune 500 data was much easier to gather via scraping the website (they would have charged us \$600 per year for historical data otherwise).

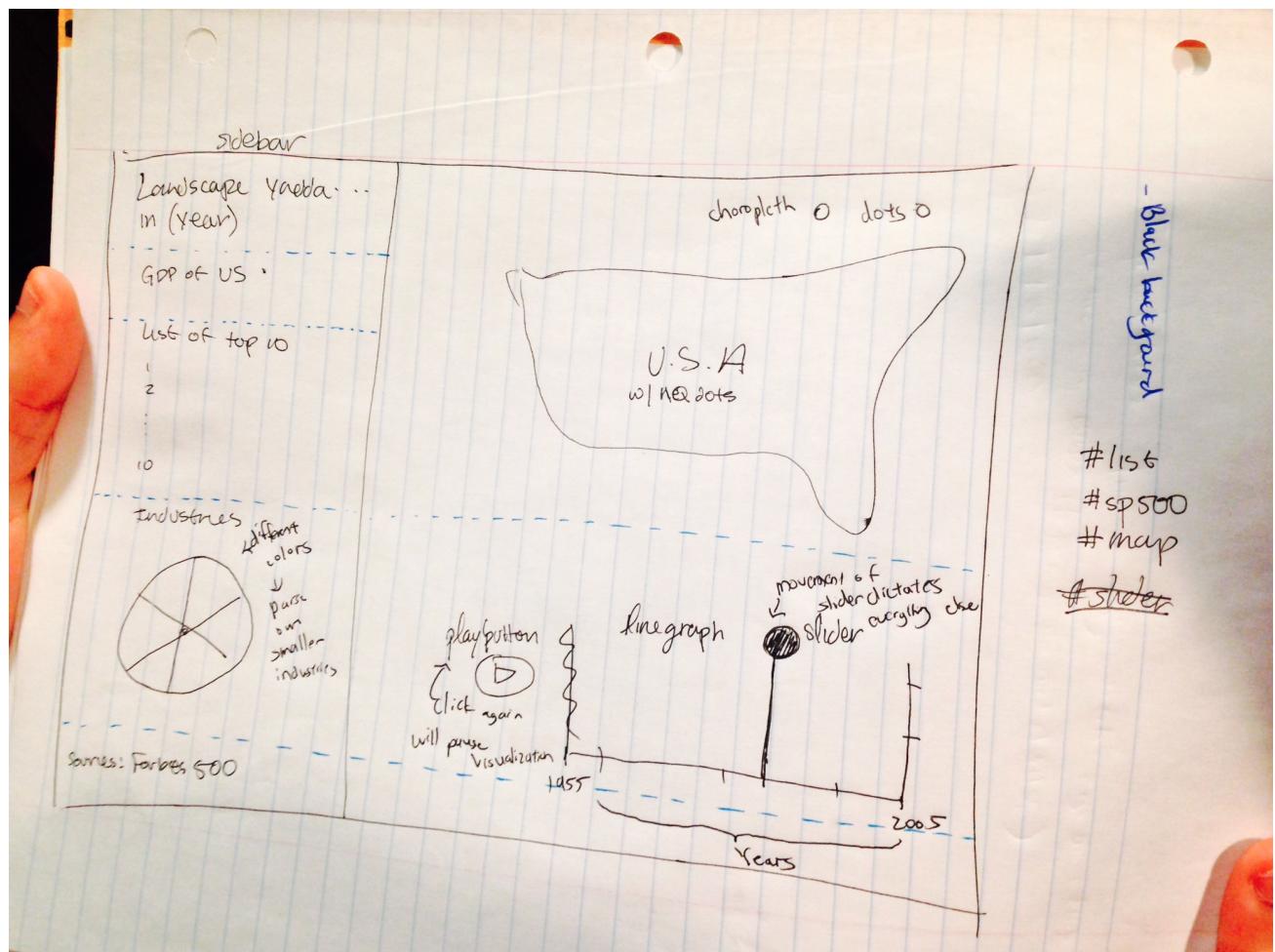
## **Real GDP growth in place of S&P500 to measure economic growth**

Upon further consideration, we decided to use annual real U.S. GDP growth instead of the overall value of the S&P 500. The S&P 500 is only representative of the stock market and thus measures how much investors value companies, which is useful for measuring attitudes towards the economy but is not stable enough to be a metric of real economic growth. Annual real GDP growth, on the other hand actually tells us how much companies are producing, a measure much more relevant to our visualization.

## **Revised method of processing data**

Although all 500 companies on the Forbes list for every year in our visualization was collected, only the top 200 in each year was used. More data points would crowd the visualization, and although the script we wrote to scrape for company addresses completed much of the data set, filling in the blanks for over 25,000 points would not be feasible.

## Updated Sketch



# IV. Data

## Data Collection

### *Basic company profiles*

The company data was collected from Fortune Magazine's online archives of Fortune 500 rankings. Each data point contained the company name, the Fortune 500 ranking, profit, and revenue from a given year.

### *Fleshing out the details*

The location of the headquarters for each company were taken from a combination of LinkedIn (scraped) and Wikipedia (hard-coded) and limited in scope to city and state. Companies with headquarters outside of the U.S. were simply entered as non-U.S. for simplicity's sake, and certain companies with more recognizable abbreviations were reduced to such for ease of understanding (e.g. International Business Machines became IBM). Finally, the longitudinal and latitudinal coordinates were extracted using Google Maps API.

Additionally, LinkedIn provided some minimal information about the industry each company operates in, so many of these points were hard-coded using sources such as Wikipedia, Yahoo! Finance, and Google Finance.

### *The economy*

*The real GDP* growth data from 1955-2005 was provided by the U.S. Bureau of Economic Analysis, and information contained in the introductory "stories" were taken from Encyclopedia Britannica and Wikipedia.

## Data Analysis

There were no significant calculations to be made outside of d3 for our dataset. The biggest change, previously mentioned in the revised proposal, was limiting the data to only the top 200 companies for every year.

# V. Components

## In This Section

Rather than write a linear account of the entire project all at once, I broke down the notes we had for each stage by component to give a better sense of our train of thought for each one. Elements are mentioned in the order that we created them. Additionally, each component's interaction with the remainder of the visualization is documented separately as well, if only to reduce the complexity of the overall explanation.

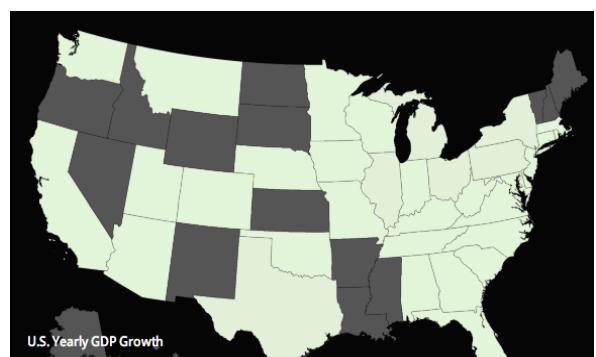
## Map

### *Locations*

We chose to use the familiar d3.projection used in problem set 4 to portray the location of the headquarters of the companies in that year's Fortune 500. We appended circles to the appropriate longitude and latitude and made the radius of the circles correspond to the size of that company's revenue. The opacity of the circles allows for smaller companies to be more visible as well.

### *Choropleth*

We chose to use choropleth because it better portrays company density, especially among areas such as the Eastern Seaboard. We added a toggle button allow both types of maps to be displayed at the same time. The two screenshots below show the difference between the two maps. We eventually changed the choropleth to a blue color scale given our color scheme and the difficulty in seeing white circles on a light green background. Finally, a legend was added in order to give more clarity to the color scale used.



## Toolips

Tooltip functionality was taken from Mike Bostock's bar graph example and edited to display the company name and address upon hovering over the circle.

## GDP growth graph

### Basic structure

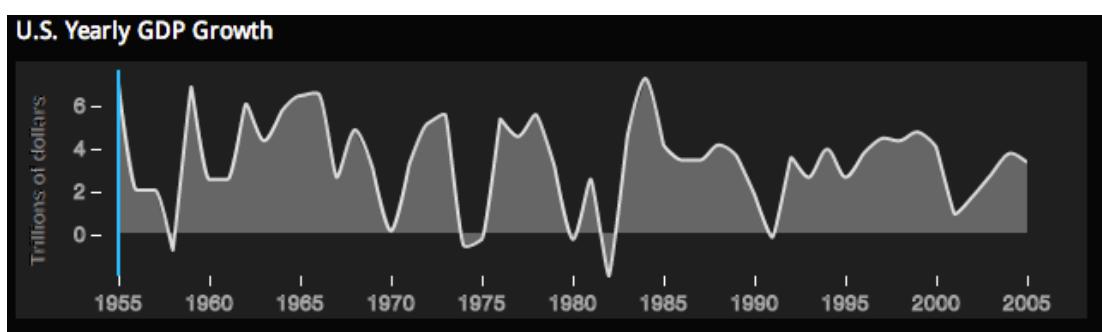
The basic structure of the line graph did not change throughout the course of building the visualization, although we added the gray area to better delineate years that had growth rates below zero, indicating a recession.

### Drawing a path

Throughout the first half of the project, we attempted to draw a path on top of the existing line graph as the visualization played through, similar to the case found at <http://jsfiddle.net/nrabinowitz/XytnD/>. However, we ran into significant problems with this implementation. Because the path depended on a time duration, the draw was not consistent across areas that had more fluctuations (more line length to cover) and those that did not (less line length cover). In the end, the path was not consistent enough with the display of the year and the slider and was scratched.

### Brush → Slider

Rather than have a dropdown list of years or radio buttons, we chose to have the user interact with the graph directly to select the year displayed by the rest of the visualization. This idea was inspired by the brush we implemented in problem set 4, but because our visualizations are based on one point only, the range the brush provides would have extraneous. Instead, we chose to use a slider, which is a line with a d3.drag function. The action updates the global variable, which the entire visualization then responds to. In the following screenshot, the bright blue line is the implementation of the slider.



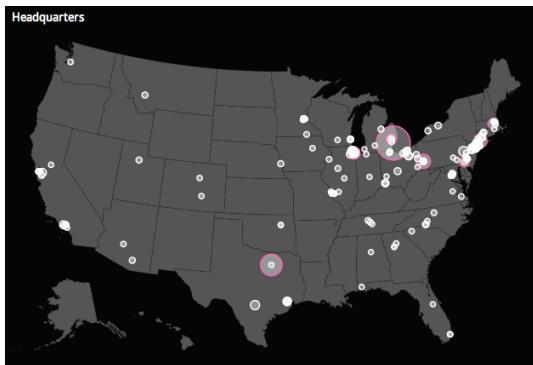
## Top ten table

*Table → bar graph*

In the first iteration of the visualization, we used d3 to append data to table elements in the HTML. This method was not, however, conducive to updates. As a workaround in the interest of time, we created a bar graph with rectangles of all the same length, which served as a platform for the labels that we were in search of.

## Highlighting

In order to give users the ability to learn more about the largest companies of each year, hovering over a specific element in the table will call the tooltip for that company and make the corresponding point much more visible. Later on, we realized it would be useful to know right off the bat where the top 10 companies were and so gave their circles hot pink outlines.



This screenshot shows the hot pink outlines of the top ten companies in that year and the general concept of “highlighting” in our visualization.

## Industry pie chart

*Filtering for top five industries*

The pie chart was implemented to get a better sense of how industries behaved in time frame we were looking at. The chart was encoded using the number of companies in a respective industry rather than the revenue they brought in, since certain industries consistently have much higher revenues than others. Showing all industries among 200 companies, however, resulted in a pie chart that was unreadable. In response, we only show the top five industries and lumped the remainder in a category called ‘other’. We chose to append labels to add more clarity and information to the chart.

## Highlighting

As with most of the other components in this visualization, we wanted to point out which points on the map corresponded with the wedges in the pie chart. In order to do so, we reused the same color scale to outline the relevant points upon hovering over the pie chart.

## Storytelling

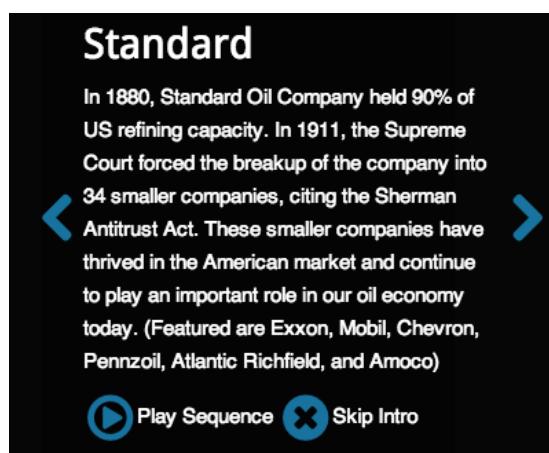
### *Sharing our data exploration*

Given the number of moving parts we eventually decided to place in our visualization, we decided to add a storytelling component to balance the number of data visualizations. Initially, we wanted to include news stories, but our data was too broad in scope to see the impact of something that would be easily accessible in newspaper archives. We eventually targeted two different types of “stories” that we felt showed different dimensions of the data best.

- **Wal-Mart:** From its humble origins in Bentonville, AR, Wal-Mart Stores has grown to be the largest retailer in the world. Wal-Mart pioneered the model of superstore retailing we know today, and it has paid off. Wal-Mart entered the top ten of the domestic Fortune 500 in 1995 and has since skyrocketed to the top of the Fortune Global 500, with 11,000 stores in 27 countries. To this day, though, Wal-Mart is based in Arkansas and managed by the Walton family.
- **Rockefeller's legacy:** In 1911, U.S. anti-trust policy forced Standard Oil to break up, creating a multitude of oil companies in a space that had previously only seen monopolies. We decided to feature Exxon, Mobil, Chevron, Pennzoil, Atlantic Richfield, and Amoco.

### *Interaction*

In addition to displaying these text stories, we gave each one interaction with the rest of the visualization. Each story slide has a button that gives the option of “Play sequence”, which initiates several actions. First, the play function is set to play for a certain interval that best demonstrates the trends we mention in the story. Wal-Mart played from 1992-2005, and Rockefeller's legacy played from 1955-2005. Next, certain ‘mouseover’ functions were triggered. Showing tooltips on the map made it relevant to the story, as did making certain circles larger or of a noticeable color. The user has the option to exit the storytelling at any time by clicking ‘Exit intro’. The following screenshot is of an early iteration.



## Miscellaneous Interaction

### *Play/pause button*

This was by far the most important part of the entire visualization, given how one dimension of the data set was change over time. Initially, we wanted a play button that simply cycled through the years, but soon realized the value of pausing the visualization – this would give users more flexibility and time to absorb data.

### *Dynamic year and total revenue display*

A visualization is not much without some bite-sized piece that can be easily digested, and so we created a column with dynamic facts (specifically revenue and the year) in bold lettering to increase the accessibility of the visualization. In a way, a user does not have to expend as much effort to understand the moving parts but simply reads the summary.

Interesting tidbit: we made the words ‘top ten companies’ in hot pink to reference the hot pink outlines of their respective circles.

### In 1955

the top ten companies in the Forbes Fortune 500 were:

1	General Motors
2	Exxon Mobil
3	US Steel
4	General Electric
5	Esmark
6	Chrysler
7	Armour
8	Gulf Oil
9	Mobil
10	DuPont

and had a total of

**\$33,425B**

in revenue.

 Play

 Show by Industry

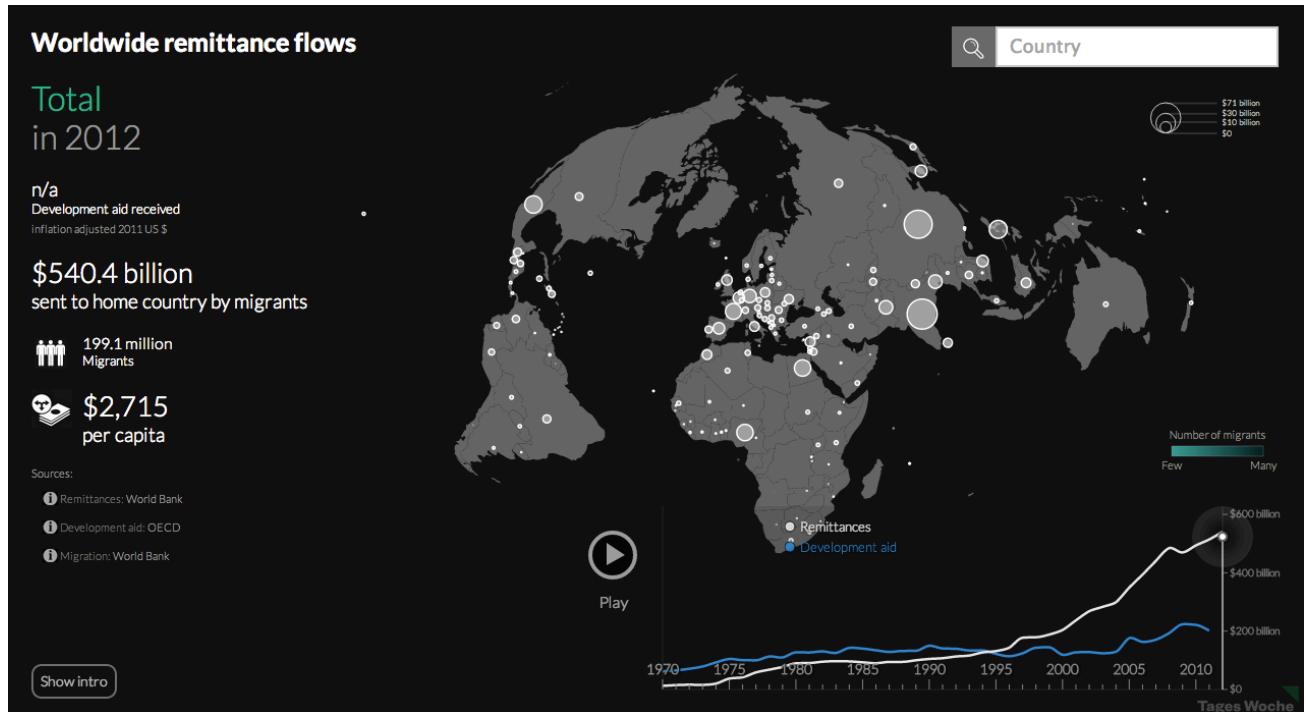
# VI. Visual Design

## Inspiration

Our main visual inspiration came from visualization chronicling worldwide remittance flows created by the German news site TagesWoche (<http://remittances.herokuapp.com/?en>). We saw parallels in the types of information we had (geospatial and quantitative) and in the relationships between the different sets of data; for instance, the graph in the bottom right corner shows trends in developmental aid over the same time period, similar to how we included real domestic GDP growth. As a result, we decided to use a similar layout.

Furthermore, this site just looks incredibly sleek.

## Screenshot



## Implementation

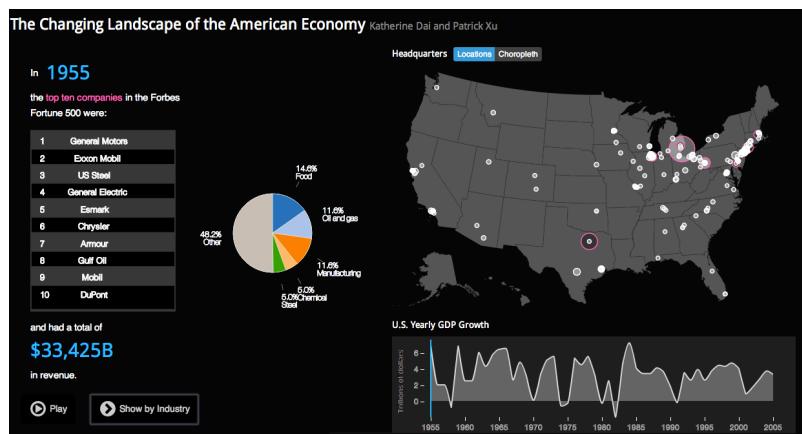
### Bootstrap and Font Awesome

These libraries were used to create a grid system to align our elements and to create buttons that looked polished and sleek.

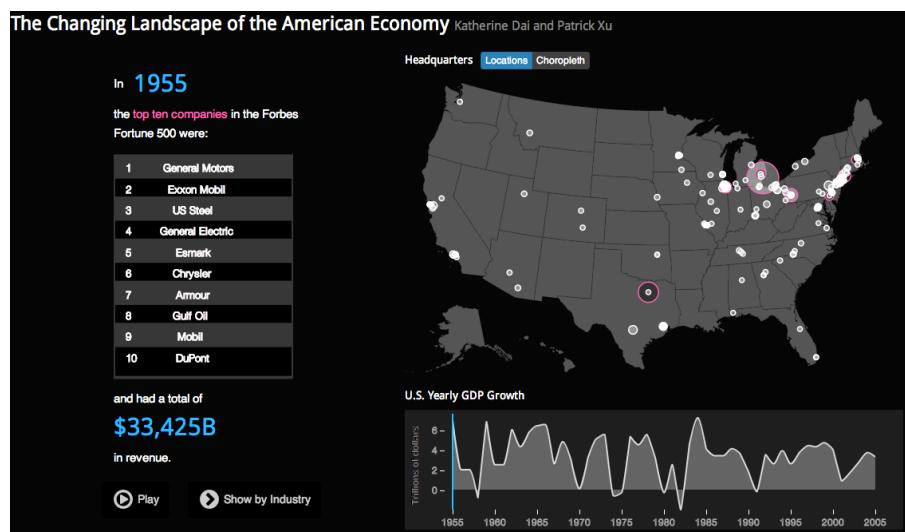
### *Collapsing column*

We soon realized that we had arrived at an unique problem – we had too many visuals, but each one was integrated well with the rest that we couldn't bear to part with it. As a result, we hijacked Bootstrap's grid system to create a 'hidden column' containing the pie chart. To view this chart, users simply have to click the 'Show by Industry' button, which triggers the change. The two following screenshots show the visualization with and without this hidden column in view.

*With the column*



*Without the column*



# VII. Conclusion

## **There were some issues...**

At first, we were surprised by how few trends we saw when making exploratory visuals. There seemed to only a few trends worth mentioning, and the limited time meant that we would have fewer points and dimensions to work with. In the future, I think we'd focus on the data collection stage, given its pivotal role in the rest of the project.

## **...but we had a great time.**

This project was fun to put together, and we think that the end project is relatively bug-free and of a good caliber. There are always ways to improve visualizations, and being able to integrate more stories would make the visualization richer and broaden its scope. Additionally, we would find more recent data, which could have more interesting trends from companies that have only been started in the past decade or so. Along the same lines, using the global Fortune 500 data would also be extremely interesting. In general, it seems that several companies continually dominate the list. Whether or not this is good for society in general, it is certainly capitalism at work.

## **Thank you!**