ABSTRACT
Currently much politically relevant data such as campaign contributions, election results, demographic information, legislative actions and polls are siloed, making big picture political analysis very difficult. We seek to build an interactive app that can show trends in monetary flow and political developments. The influence of sectors over candidates has been shown only on a candidate-by-candidate basis, but never has there been an effective visualization on a bird’s eye view level. We present in this paper an effective way to see trends and clusters of candidates in response to their campaign contributions. In addition, we use a visualization technique that we demonstrate can be used in other fields.

Author Keywords
Politics; finance; contribution; data visualization; money; influence;

ACM Classification Keywords
General Terms
Human Factors; Design; Algorithms;

INTRODUCTION
During the 2012 United States election, a lot of attention brought to PACs (political action committees), and their influence over campaign contributions [1]. In addition, the 2012 election was the most expensive campaign in the history of the United States. Corporations gave roughly $2 billion dollars to political campaigns [2]. It’s not hard to imagine that the candidates who won and received donations from certain sectors and corporations would vote on legislature in ways that would benefit those corporations, especially if they hope to get money again from those same sources when they seek re-election.

There are tools that exist to the campaign finance breakdown on a candidate level. For example, a pie chart of a candidate’s campaign finances would demonstrate the breakdown of how much each sector (such as Agribusiness, Labor, etc.) has contributed that candidate’s campaign, and if fact these tools do exist [3]. However, the problem with these tools is that it is hard to see the campaign finance on a bird’s eye view level.

It’s not hard to imagine when we put even 10 of these bar charts together, it becomes hard to compare. Also, if users wanted to view the Senate, which has 100 members, users would be overwhelmed to compare 100 bar charts side-by-side. If we aggregate all those bar charts into just one bar chart, then we lose the detail on how each candidate contributed into the making of the bar chart.

The importance of seeing it on a bird’s eye view level comes when we factor in how these politicians operate. For example, in the Senate or House of Representatives, legislature needs a certain number of votes in order to pass. At this point, we can care less about politicians as unique individuals, and look at them as votes. If we believe that money from sectors affect how politicians vote, it would be interesting to see which sectors have the most influence, and hence the higher likelihood of votes going towards directions that benefit them.

With the factors in mind that we wanted to capture the aggregate sum of each sector, while still maintaining the ability to view the specifics of each politician, we designed our visualization in order to be able to achieve these goals.

However, we also wanted to design a general visualization technique that could be used in other fields given the team of comparing a sort of “pull” or “influence”, which we will give examples of later in the paper.

RELATED WORK
In the field of visualizations of campaign finance, Influence Explorer contains visualizations of the finances of each candidate when you select a candidate [3]. However, as we discussed in the introduction, this is adequate on a candidate level, however, these techniques would be horrible when trying to compare many candidates. That being said though, they had the best source of campaign finance data, so we used the Influence Explorer API created by Sunlight Labs that was able to access their data [4].

Figure 1. Breakdown of Obama’s campaign as shown by Influence Explorer
Another visualization that was inspired by the same questions we had had been created by Michael Bommarito and Daniel Martin Katz from the University of Michigan, who worked with data from 2007-2009 [5]. However, the weakness in their visualization is that they used companies instead of simply sectors. This makes it much harder to notice which companies or sectors are their biggest influences because there are too many connections and the graph becomes cluttered [6].

![Figure 2. Visualization of 2007-2009 finance by Bommarito and Katz](image)

It might be hard to read the labels in the image we provide, and that is because the labels are hard to read in the visualization itself. Users have to scroll all the way in till they lose sight of the overall structure in order to read the labels.

Wall Street Journal released a visualization called “Political Moneyball” [7]. However, it’s still hard to decipher the connections between candidates and corporations because the only clue to their relationship is only implicitly defined by location. We wanted to a visualization that explicitly showed the relationships.

![Figure 3. Moneyball by Wall Street Journal](image)

Another problem with these visualizations was that the interactivity didn’t work the way we wanted it to. It was hard to compare the influence of different sectors with each other, because of the lack of interactivity, and how their drill down features worked. Their drill down features were already pre-calculated and didn’t allow users to change parameters in the way calculations were made, so it wasn’t that useful for exploring the data.

**METHODOLOGY**

**Data Set**

We pulled our data from Influence Explorer using the Sunlight Labs API. However, because at the time this paper was written, which was after the general 2012 election, but before the inauguration of the election winners, most of the API’s out there that pull the information of the current senators and congressmen, would pull data for politicians currently in office, but that did not win the election, because the inaugurations had not happened yet. In order to solve this problem, we applied some filters and annotated the data ourselves to get the politicians that won the 2012 general election.

**Main Visual Components**

The visualization that we decided upon has three main visual components that we think best demonstrates the relationships between candidates, and sectors.

**Candidates**

The first component is the candidates that are encoded by the circles inside the biggest circle. Their party affiliation is encoded by color, red for Republicans, blue for Democrats, and green for other parties, as usually done by the media.

Sometimes the media chooses to encode colors based on the specific party like yellow for Libertarian and green for Green party candidates, however, we decided that it would not be worth encode that specific of colors because the number of candidates that win that are affiliated with those parties are a small minority. In addition, we decided to use green for 3rd parties because green is a good complementary color to red and blue based on the color wheel that uses the basic colors, red, green, and blue.

For readers unfamiliar with the terminology, the United States mainly has a de facto two-party system, where a majority of the political candidates come in either one of two parties. In a simple generalization, the Democrats are the liberals and the Republicans are the conservatives. There are other minor parties like the Libertarians, the socially liberal and financially conservative party, which candidates can run from. However, because of the huge voter support and membership of these two parties, most candidates come from these two parties.

![Figure 4. Senators as shown in the Senate View](image)
The size of the circle is proportional to how much total money the candidate has received from the respective sectors. It is important that we purposely chose not to show all the sectors that a candidate receives his money from. The two sectors we left out were Other and Unknown. The reason for this is that a majority of candidates had Other or Unknown as their biggest contributing sector. The problem with this is that Other and Unknown are made up of a lot of constituent groups and it makes it hard to figure out a hypothesis for where this money is coming from and why it is going to a candidate. The formula for calculating the radius of the circle is as follows.

\[
\frac{20 + \text{totalsum}}{20*100000}
\]

totalsum is the total amount of money that a candidate has received from all the sectors that are currently represented in the visualization. We scaled it so that certain candidates that received more money wouldn’t completely overshadow candidates that didn’t receive as much money. We also wanted to encode the total amount they received because we wanted to show if there was a trend in total amount of money and sectors they were being pulled towards.

**Control Points**

The 2nd main component of our visualization is the sectors encoded by rectangular control points on the outside of the biggest circle. Each control point is given a unique color and a label so that discernment will be fast. In addition, each control point has its height encoded based on how much money their corresponding sector has given to all the candidates. The default locations for the control points are decided by the alphanumeric ordering of the sectors found in the Sunlight labs API. However, we give the user the power to move the control points around the circle so that they can have their own custom view.

**Lines**

The 3rd main component is the lines between the circles and the control points. The stroke between each candidate and sector is calculated based on

\[
0.3+0.7*(\text{tensionFraction})*\text{maxStrokeWidth}
\]

where maxStrokeWidth is an upper limit on the stroke width (by default it is 15), and tension fraction is the fraction of the money given to the candidate to that sector over all the money that the sector gave out in total. The length is determined by the candidate placement, which we will explain in the next paragraph.

**Placement of Candidates**

Each candidate is connected to a sector by a vector modeled as a spring. In accordance with Hooke’s Law, this spring exerts a force \(F\) with components \(F_{ix}\) and \(F_{iy}\) on the candidate such that

\[
F_{ix} = k_i(x_{\text{candidate}} - x_i) \quad F_{iy} = k_i(y_{\text{candidate}} - y_i)
\]

where “\(i\)” represents each sector. The spring constant “\(k\)” was taken as the cash received by the candidate from sector “\(i\)”. Each candidate thus has his/her own set of \(k\)-values.

The candidate is placed in coordinates \(x\) and \(y\) such that all forces are in equilibrium. That is:

\[
\sum_{i=0}^{\text{all sectors}} F_{ix} = 0 \quad \sum_{i=0}^{\text{all sectors}} F_{iy} = 0
\]

These equations can be solved for the \(x\)- and \(y\)-coordinates such that

\[
\frac{\sum_{i=0}^{\text{all sectors}} k_i x_i}{\sum_{i=0}^{\text{all sectors}} k_i} = x_{\text{candidate}} \quad \frac{\sum_{i=0}^{\text{all sectors}} k_i y_i}{\sum_{i=0}^{\text{all sectors}} k_i} = y_{\text{candidate}}
\]

**Drill-down Techniques**

We implement a lot of drill-down techniques using interactivity in order to encourage the user to explore the data.
Click Selection
We give users the ability to click on a candidate and highlight that candidate. All other candidates become transparent, and as do their lines. The highlighted candidate then has his corresponding lines visible, making it clear which sectors are the most influential and by how much.

Drag Selection
We implemented this technique so users could examine an outlying cluster they find interesting. Everything that would be applied to a single clicked selection is then applied to all the candidates in this sector. We also provide a menu for a user to see which politicians are selected, and let them cycle through the politicians.

Congressional Committee Selection
We leverage a piece of metadata called congressional committees. Each congressman and senator serve on congressional and senate committees, which are in charge of discussing and passing laws regarding topics that fall under the domain of their committee. For example, the foreign aid committee is in charge of resolving issues with aid getting sent to Africa or Israel. We thought this piece of metadata would be interesting because then we could view if there’s a correlation between serving on a committee and the sectors that a candidate got money from.

We implemented a selection bar on the right side, where if you select a congressional or senate committee, the members of that committee automatically get highlighted, and isolated.

Search
We also have a search bar that has an autocomplete feature that filters the candidates by name so you find a specific politician that you’re looking for, and they would behave as if they were clicked if there was a match on the search query. Users can go through the auto-complete menu using their arrow keys.

Tool-tip
Lastly, we implemented a tool-tip on mouse over that would bring up the politician’s name and a link to their Twitter page if they had one.

RESULTS
It is very hard to quantify the results our system brings up, because it is designed for users to get new insights from the data. We will instead highlight interactions that our users had with the system in this section.

Based on the default arrangement of the sectors, there is an apparent divide between Democrats and Republicans.
Majority of the users typically would start then moving around sectors to see which ones had the greatest influence in furthering this divide. A typical reaction would be that users would notice that the Labor sector had influence on Democrats and none on Republicans and the converse was true for the Agribusiness sector, which confirmed their hypothesis and assumption based the reporting of the media that Republicans were strong in rural sectors of the United States and that Democrats were stronger with people from a lower income demographic. Likewise, they noticed that finance had a pull on a lot of candidates, irrespective of party. Users typically could confirm their hypotheses almost immediately using the click and drag interactivity. It was interesting to note that almost all users started by trying to create more separation based off the default view.

Next, we found that by moving all the industry sectors to one side except one you could more easily tell which politicians are receiving the most from the lone sector on the opposite side. We tried this with the health sector. As expected when highlighting members of the health committee they appeared to be closest to the health sector. Discussion about the System

Our system allows people to explore and analyze the influence of sectors unlike any other visualization. For one, it allows them to realign the sectors that they can view one sector versus every other sector, and it allows them to see which candidates are more susceptible to movements within a sector.

The result is that the audience felt our system allowed them to see the cash flow on a much higher level than they could before, and from that they could explore the data and easily form new hypotheses and confirm some based on their customization of the view.

Another suggestion was that it would be interesting if the system allowed users to see which influences were missing. A possible solution to this problem is discussed in the Further Work section.

A minority of the users was able to use the congressional subcommittees investigate their hypotheses. For example, one user was surprised that the total amount of money given by the Energy sector was small. However, when he filtered by the congressional subcommittee, the republicans on that committee were volatile to movement, while the democrats barely had any movement, which meant that energy corporations targeted specific politicians heavily, like the Republicans who serve on the energy committee.

**DISCUSSION**

**Insights from the Visualization**

One doesn’t have to be an expert in public policy to be able to use ContributionExplorer to gain valuable insights on what kind of influence the industry sectors have on our politicians. The following are several findings we made just from playing around with our tool. The most conspicuous discovery was that dragging around the labor sector significantly pulled almost all of the Democrats, but almost no Republicans. This clearly illustrated a sharp partisan divide with respect to major funding sectors. To a lesser extent, we found that the Energy & Natural Resources sector along with construction were mainly influenced Republicans. In the house, if you highlight only members of the Agriculture Committee you'll see that Republicans in it are much closer to the Agribusiness sector than Democrats. Similarly, in the Energy Committee, Republicans appear to be more influenced by the Energy Sector than the Democrats also serving on the committee.

However, some users struggled with recognizing how much candidates were moving based on the pull of the sector. The only real way to compare magnitude was to drag another sector afterwards, but there was no real tool in comparing magnitudes directly.

Users were able to find specific candidates that they wanted extremely quickly. The search bar with autocomplete made it really fast, and in fact helped one user notice that they were looking for a congressinan in the senate view instead of the house view. This means that sometimes despite having 400 plus candidates in the house view, users rarely had difficulty narrowing the points down to the candidates they were interested in.
Discussion about the technique

There was discussion how the visualization technique we used in this domain could be applied in other domains. Here we will present two examples for brevity.

The criteria for the data set that this technique can be used for must meet the following criteria. First, the data that will be encoded in circles must have at least one quantifiable attribute. Up to non-quantitative attributes can be encoded using color. Up to two attributes can be quantified using size and location, however, the location variable has restrictions, which we will discuss.

The data points encoded in the control points must also meet certain criteria. Each control point should represent some sort of category that the data encoded in the circles has some sort of score or value for.

Lastly, the data that is encoded in the lines should be based on the value that user meets in the categories or data encoded in the control points. Mainly, the line should be the data that represents the relationships between the data in the circles and the data in the control points. Zero values are allowed.

We believe the optimal way to use this method is that if the data encoded by color of the circle deals with a result that is possibly the result of the values in the categories encoded that a data point has. Another way to use it is if the color of the circles deals with data that is tangentially related and instead might be a factor in the scores encoded by the lines.

Here are the two examples of the encoding.

Myers-Briggs and Sales

In this example, we use the same system with different data. The salespeople are encoded as circles, the different components on the Myers-Briggs test as control points, and their aptitude of the different components as the influence on the component on them. For the readers who don’t know, the Myers-Briggs is a popular personality test [7]. There are 4 components to the test that a person can get ranked on. For example, in one component, a person is placed on the Introvert-Extrovert scale, and in another, they are placed on the Judging and Perception scale, and so on and so forth. We can then color code the salespeople by binning them based on the number of sales they have. What the result would be is that we would be able to identify if there is a certain personality that does well in sales, or a certain personality that doesn’t. This means we can visualize the influence of personality over sales.

Predicting the Professional Careers of College Athletes

All the time in evaluations of the chances of a collegiate athlete going pro and succeeding, a lot of focus is put on their physical attributes. The common belief is that physical attributes are a good determining factor if an athlete will succeed in the pros or not, even more so than their collegiate success [8].

In this example, we can encode collegiate athletes as circles, their physical benchmarks as control point, and their performance in that physical benchmark, where closer to the control point means they scored better on a benchmark. This makes them easy to place because their scores in physical benchmarks are quantified and thus easily compared. For example, height is usually a quality sought in players especially in basketball and American football, so we can use that. Another is their bench press, which is a good sign of upper body strength. For running backs in American football, the 40-yard dash is a good indicator of their future success. To normalize scores, we could assign score based on percentile based on their performance.

We can bin athletes then based on their success in the professional leagues. We can bin them by if they lasted less than 3 years, if they ever started, and if they made won a significant award. We might be able to see cluster form, and this will tell us if certain physical attributes are more important than others based on cluster formation.

FUTURE WORK

Ability to share visualizations

Our visualization uses a RESTful URL, and doesn’t pass any parameters into the Uniform Resource Locator (URL). This means that it makes it hard very for people to share visualization outside of taking a screen shot and sending it.

The most obvious way would be to give a method to users to share their visualizations base via URL.

Removing a Control Point

A lot of users asked for a feature to remove the influence of a sector completely so that they could limit it to two to three sectors. This is something that will help in analysis if implemented. For example, it would be a lot faster to examine if Finance or Defense has more pull over overall.

In addition, this allows users to really isolate the sectors that they are curious about, and see interesting cluster form. For example, we could check if we remove the usual sectors of funding for Democrats, are there certain Democrats that behave like Republicans.

Other Filters

Another view we can add is the ability to filter by party. Thus, we could see who the outliers are in a certain party.

Another filter we could do is to filter by state. We could check if politicians in New York get more money from Finance because of Wall Street, and if politicians in California get more money from Electronics, because of Silicon Valley.

Lastly, the ability to randomly remove and throw in politicians would be a great tool in order to compare two politicians or candidates against each other more saliently.
API
Given that we believe that the visualization technique we used could be used in other fields, it would be interesting to create an API for this technique. Example methods that the API would use would be allow users to pass in the data to be encoded in the circles, the data to be encoded in the sector area, and the function to evaluate position in the circle. If an API is created, then we could easily implement the two visualizations that we talked about in the Discussion section of this paper.

Drill-down
Unfortunately we were unable to get accurate data on which contributions came from each sector, so we had to settle on aggregates; it would be fairly simple to extend the visualization such that if a sector were clicked, we could use the same technique and replace the donation sectors with the top contributing companies given the sector that was selected. This would allow a person to see which companies in a sector have the most influence. In addition, we could do this a company and replace the companies using a transition with the employees of the company that was selected and find out which ones gave the most money.

Community Detection K-Means
Right now, we don’t really optimize the placement of the sectors on the outside circle. If we use a k-means in order to optimize the layout to arrange the layout such that clusters appear by default when possible, then users would probably be able to pick up the tool faster, if an API version would be released. Implementing this would also help prevent an edge case where a candidate would gravitate towards a sector that provided little funding, if the adjacent sectors provided a huge amount of funding. This would be fixed by grouping the sectors based on which ones shared a lot of nodes. Note that given the political dataset we used, clusters already appeared in the default view, so this may or may not help all the time [10].

Reverse Reaction
One user pointed out that it would be interesting that if we moved a circle, if the sectors would re-align to match the placement of the circle. This could possibly demonstrate which candidates have a lot of free space, which demonstrates how little influence the sectors have over him, based on much they move. This is a possible solution to see which influences are missing or weak, especially if we perform this on a cluster of nodes.

CONCLUSION
In this paper, we contribute a tool to help the average person with their understanding of the financial campaign situation of politics by giving them the ability of exploring the data. We present it in what we hope was an intuitive method for users by presenting influence as a physical “pull” over the items in consideration. In addition, we contribute a visualization technique that we believe can be used for data sets that meet certain criteria. We believe it holds potential as a tool for detecting clusters that might arise from data with specific relationships. We hope that our visualization can be used as a stepping stone for more contributions in the field of visualization.

ACKNOWLEDGMENTS
We would like to thank Sunlight labs for their API and Influence Explorer for their data.

REFERENCES
3. www.influenceexplorer.com