Final Project Report

By Garrett Chang and Steven Fiacco (and Darren Lin and Anders Maraviglia [SD&D])

Motivation

Our motivation for this visualization resulted from a simple problem when viewing election polling data. Often there are a few critical days or weeks for each of the candidates in the campaign. Determining which days these are is very easy. Critical days for that candidate are at the deep falls or high jumps in that candidate's line graph. Understanding why those days were critical is a much harder question.

Our main research question is as follows, "What is the relationship between the news and election results on any day during the election?" Our hypothesis to that research question is, "A map with headlines will be able to show a positive correlation between election results and the news."

The visualization we developed will allow the user to instantly understand the state of the election and newsworthy events on any day in the election. Our main audience is anyone who is familiar with the USA election system and has experience using a web browser.

Previous Works

The main inspiration is http://www.realclearpolitics.com/. RealClearPolitics contains polling results displayed in a line graph for the entire election. The site also features a large list of news articles from a variety of sources for the current day. The while this site is great for identifying critical days as seen in Figure 1, but there is no way to go back and see what the news was like on those days. The polling results are also for an individual state or the entire nation, however there is no electoral map to help the user visualize all the data in a singular place.
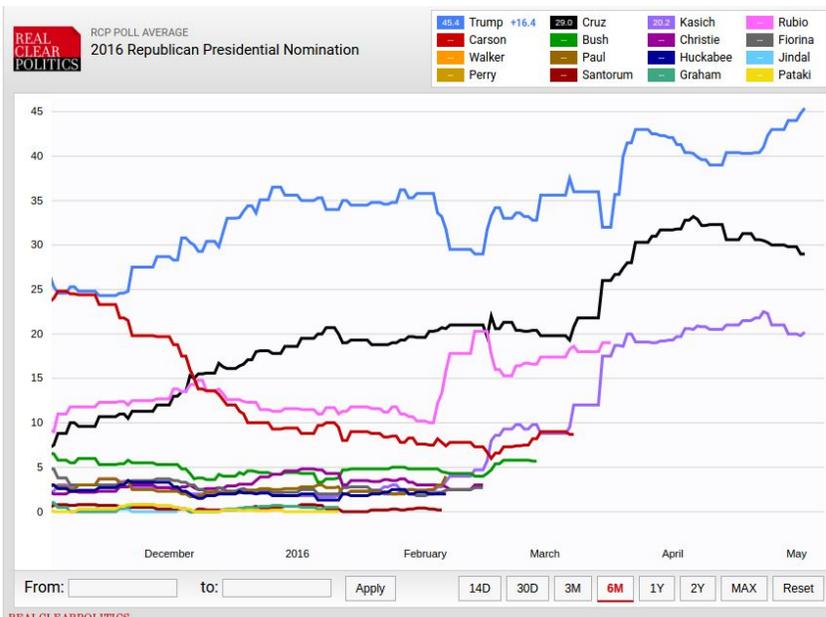

Figure 1: This is a line graph of the GOP primary taken from Realclearpolitics. Large spikes or sudden changes in slope can be easily identified as critical days.

There is a electoral map on the republican primary page of 270towin as shown in Figure 2 that features polling information when you select a state. This map is great at showing the mechanics of the primaries in each state. The map is designed to show off the timing of each state's primary, but this site does not provide the user with an overview of the election or archived news data.
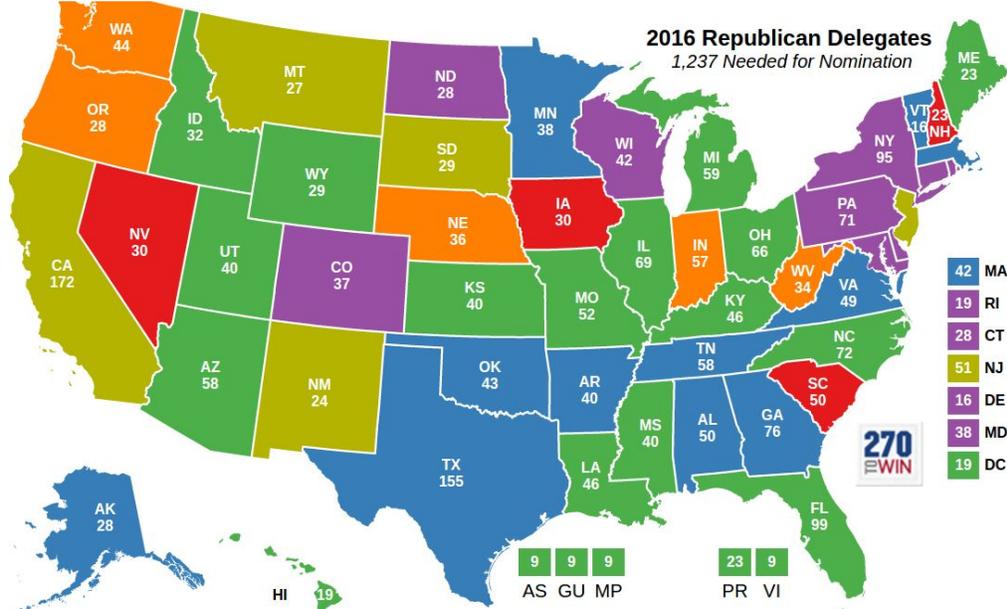

Figure 2: This map is colored based on in which month each state has its primary election for the republican party. The number of delegates each state awards is in the middle of each state. On the website there is a legend for the colors and each state can be clicked on to show information about that election process.

Data Collection

We parse 5 newsites' rss feeds, RealClearPolitics, Fox News, BBC News, CNN News, and Reuters, to obtain headlines and we parse RealClearPolitics for polling data and election results. We took all parsed information and put them into JSON that the front end used to display. Figure 3 shows a sample of the headline JSON, which the backend produces and the frontend displays on the webpage. Figure 4 shows a sample of the polling data JSON and Figure 5 shows a sample of the finished races JSON. The polling data JSON and finished races JSON are in a similar format which allows the map to parse both almost identically.

```
{
  "sub_datestamp": "Wed, 27 Apr 2016",
  "headline_str": "Jane Sanders: No back tax returns until Clinton releases Wall Street transcripts",
  "source": "CNN News",
  "link_str": "http://www.cnn.com/2016/04/26/politics/jane-sanders-tax-returns/index.html?eref=rss_politics",
  "datestamp": "Wed, 27 Apr 2016 08:59:35 EDT",
  "keywords": [
    "Sanders",
    "Clinton"
  ]
},
{
  "sub_datestamp": "Wed, 27 Apr 2016",
  "headline_str": "Rapper Lil Yachty has Lil B to thank for getting into Sanders",
  "source": "CNN News",
  "link_str": "http://www.cnn.com/2016/04/25/politics/lil-yachty-gets-political/index.html?eref=rss_politics",
  "datestamp": "Wed, 27 Apr 2016 11:27:07 EDT",
  "keywords": [
    "Sanders"
  ]
},
{
  "sub_datestamp": "Wed, 27 Apr 2016",
  "headline_str": "The long and winding road to the Republican nomination",
  "source": "CNN News",
  "link_str": "http://www.cnn.com/2016/03/19/politics/trump-delegate-math-contested-convention/index.html?eref=rss_politics",
  "datestamp": "Wed, 27 Apr 2016 15:33:30 EDT",
  "keywords": []
},
{
  "sub_datestamp": "Wed, 27 Apr 2016",
  "headline_str": "Trump's foreign policy speech offers few details but clear message: America first",
  "source": "Reuters",
  "link_str": "http://feeds.reuters.com/~r/Reuters/PoliticsNews/~3/AGbUsHGzuUg/us-usa-election-trump-idUSKCN0XO10R",
  "datestamp": "Wed, 27 Apr 2016 19:15:14 -0400",
  "keywords": [
    "Trump",
    "policy",
    "speech",
    "America"
  ]
},
```

Figure 3: A sample of the JSON containing headlines. The JSON is formatted, {date: keywords: array_of_keywords, headlines: {sub_datestamp: date, headline_str: headline, source: source, link_str: hyperlink, datestamp: exact_datestamp, keywords: array_of_keywords} }. This JSON is used to display headlines below the map and keywords on the bottom right.

```json
"blue_poll_dict_list": [
  {
    "0/04/2016": {
      "Sanders": 45,
      "Clinton": 48
    }
  },
  {
    "0/04/2016": {
      "Sanders": 45,
      "Clinton": 51
    }
  },
  {
    "0/04/2016": {
      "Sanders": 48,
      "Clinton": 46
    }
  },
  {
    "0/04/2016": {
      "Sanders": 25,
      "Clinton": 62
    }
  }
],
"state_name": "Illinois",
"red_poll_dict_list": [
  {
    "0/04/2016": {
      "Cruz": 25,
      "Trump": 34,
      "Rubio": 16,
      "Kasich": 21
    }
  },
  {
    "0/04/2016": {
      "Cruz": 34,
      "Trump": 38,
      "Rubio": 11,
      "Kasich": 16
```

```json
"Oklahoma": {
  "general_poll_dict_list": [],
  "blue_poll_dict_list": [
    {
      "March 1": {
        "Sanders": 21,
        "Clinton": 17
      }
    }
  ],
  "state_name": "Oklahoma",
  "red_poll_dict_list": [
    {
      "March 1": {
        "Cruz": 15,
        "Trump": 13,
        "Rubio": 12,
        "Kasich": 0
      }
    }
  ]
},
"Wyoming": {
  "general_poll_dict_list": [],
  "blue_poll_dict_list": [
    {
      "April 9": {
        "Sanders": 7,
        "Clinton": 7
      }
    }
  ],
  "state_name": "Wyoming",
  "red_poll_dict_list": [
    {
      "March 12": {
        "Cruz": 23,
        "Trump": 1,
        "Rubio": 1,
        "Kasich": 0
      }
    }
  }
```

Figure 4: A sample of the polling data JSON. In the JSON, we only display the latest 4 polls in the last week. The numbers are percentages and the strings are dates organized by week. When displaying polling data, we display an average of the polls.

Figure 5: A sample of the finished races JSON. The number represents the number of delegates that the candidate won. It also has the date of the election which we use so we do not display results when the selected date is before the primary happened.

Visualization for Debugging

        We took our map and put in the same static values, Sanders: 70, Clinton: 30, and Other: 21, in every state. We then created a button that lowered Sanders' percentage by 1 and increased Clinton's and Other's percentage by 1 as shown in Figure 6. We did this to test that the numbers in each state were updating correctly and to test that the gradient was working correctly. Previously, we were using random data and random numbers would change when we clicked a button, so we were never sure if any of numbers or colors were updating properly.
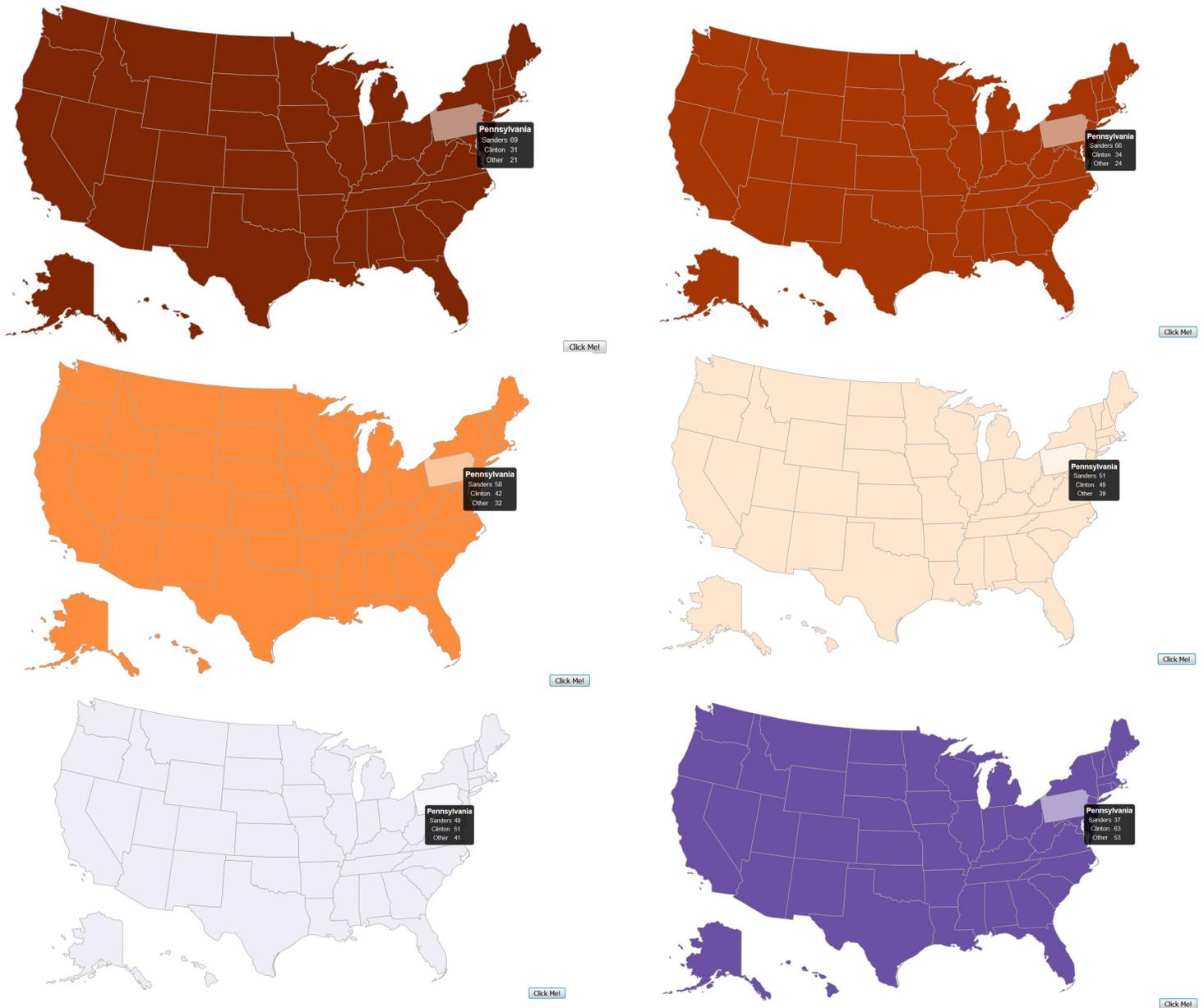

Figure 6: Visual debugging by using set values and colors. The images represent the process of going through the different colors and changing the values.

Visualization Design Evolution

Our initial idea was a map of America showing polling data with headlines below it and keywords to the side of it. The headlines would come from various news sources on the current day, and the keywords would be generated from the headlines. There would also be candidate selection area which would allow the user to switch who was being shown on the map. This means that the user could show the polling results between different candidates from a singular candidate to pairs of candidates. We also wanted the ability to show multiple candidates from the same party as well. Figure 7 shows a simple mockup of our initial design.

Figure 7: Basic mockup of our initial idea containing the map, headlines, keywords, and candidate selection area.

As we progressed, we decided that we wanted to add a progress bar based on the average polling data of all of the states above the map to give the user a quick and easy way of telling how candidates were doing in the race, which is displayed in Figure 8. In the same vein, we wanted the ability to tell the user how close a race in a particular state was. We decided that a gradient would be the best, most intuitive way to show this. The idea is that the lighter the color of the winner, the closer the race was, and the darker, the more of a landslide, the victory was. An early example of the gradient can be seen in Figure 4, in which we were testing both the gradient and how the data changed in the states. We also wanted a date selection tool

which would allow a user to pick the date of the displayed data, so the headlines, keywords, and polling data would change for the selected date. This idea eventually evolved into the calendar that is currently in place. We replaced our initial idea of selecting candidates in favor of party selection because we realized how limited the data we were getting was. It would be impossible to accurately tell how two candidates from different parties were doing in comparison to each other based on the data we were parsing. It also got rid of our problem involving invalid selections of candidates, which is explained in the Class Feedback section.
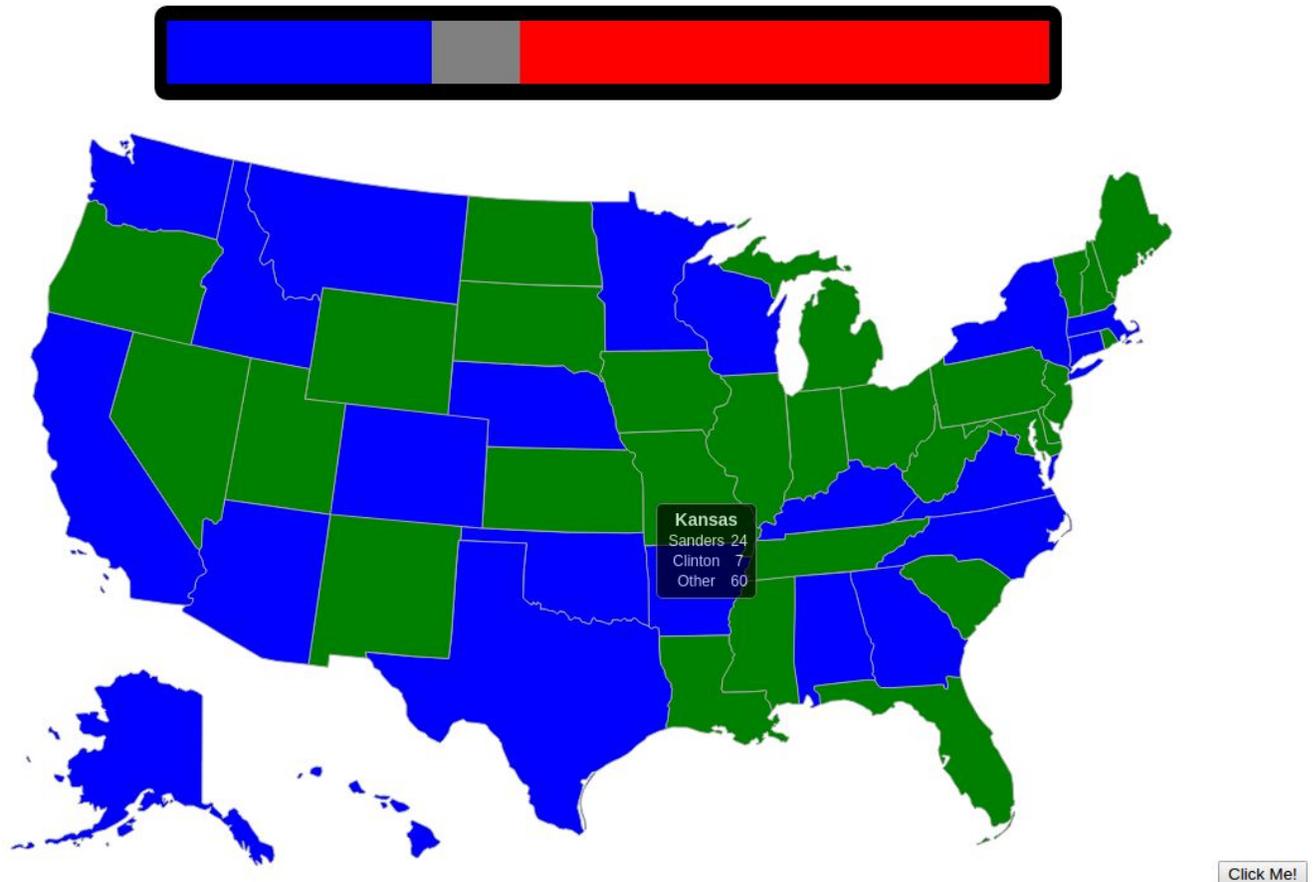

Figure 8: This is early version of the map in D3. All the data used by this map was randomly generated. The bar above the map shows the total percentage of the country for each candidate. The bar and country colors don't match yet. The tool tip currently on Kansas shows the random data for that state.

As we neared the end, the primaries were reaching their conclusion and we realized that we should show number of delegates that each candidate earned in states that had already finished their primaries. For data collection, we were unable to collect polls and headlines that had happened far in the past, and only collect polls that were recently done (in last week). Since it was so late, we did not have nearly enough polling data or headlines such that the map essentially became an electoral map. We also changed the progress bar to display the number of delegates each candidate had from the old idea of polling percentages. We clamped our

gradient so that the lighter colors could be differentiated from white and other light colors, and the darker colors did not seem completely off from lighter ones (dark orange looked brown) as seen in Figure 9. We also added the ability to run through the election week by week to show how the primaries played out as seen in Figure 10.
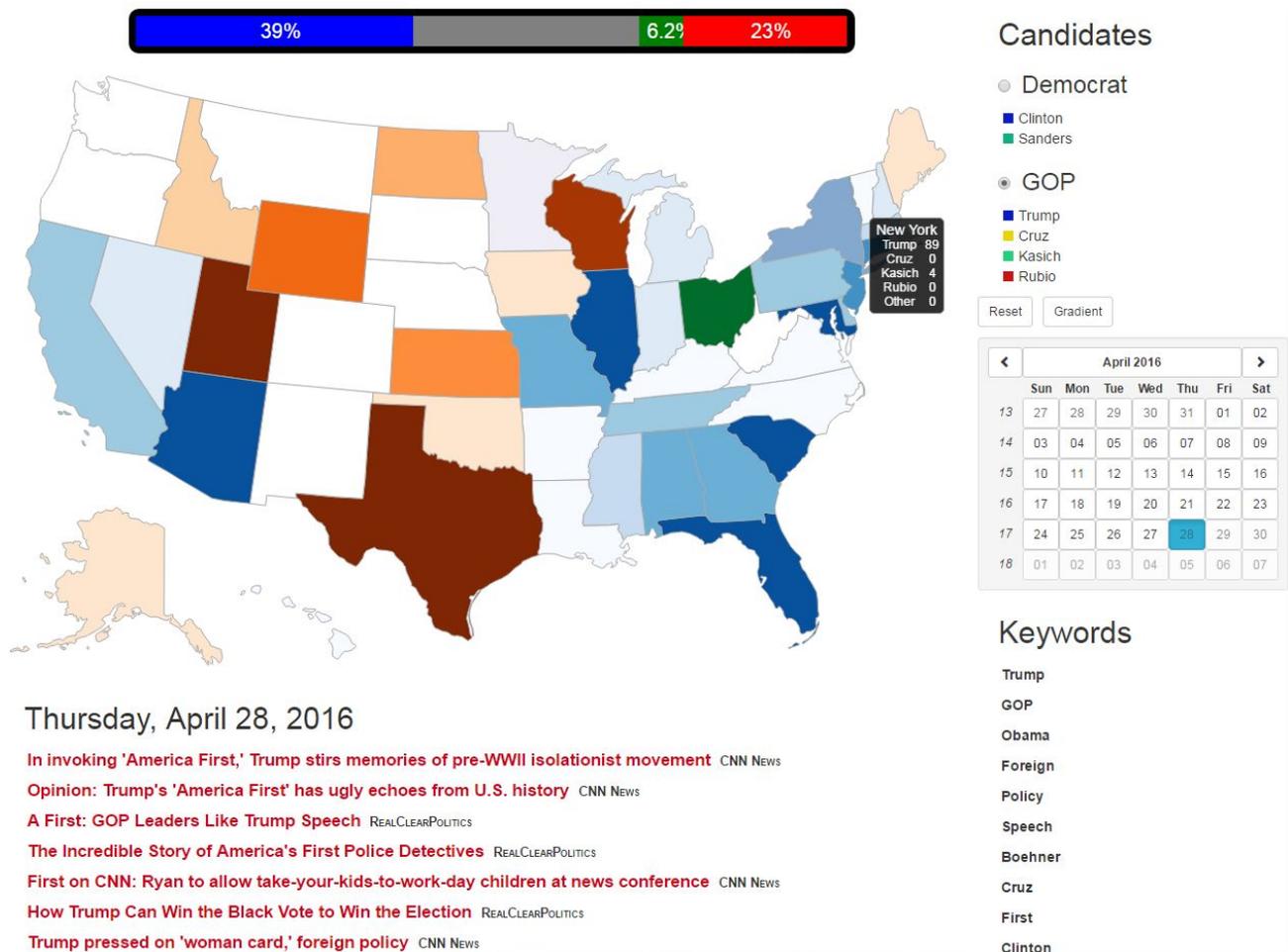

Figure 9: This a later version of the webpage. A list of news articles and their sources appear at the bottom left of the page. A list of keywords based on those articles appear at the bottom right. Above the keywords is the calendar used to change the date. Above the calendar is where the user switched between the primaries. Above the map is a bar which shows the current delegate totals for each candidate currently running. The grey portion of the bar are unallocated delegates and white states on the map don't have information.
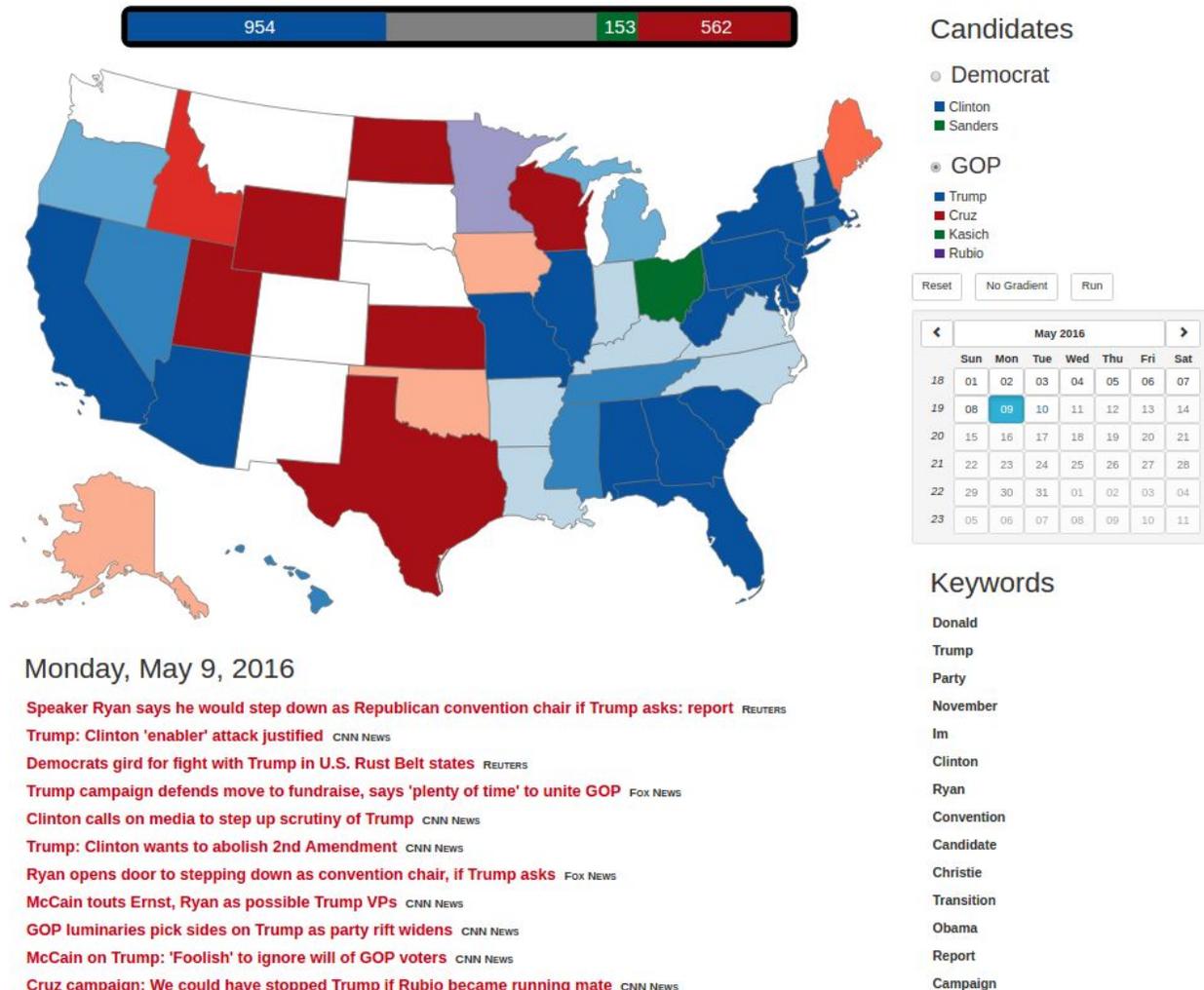
Figure 10: This is final version of the webpage. The layout is the same as Figure 9. There is now a run button above the calendar, the map and bar are colored correctly and the bar shows delegate totals.

Class Feedback

The feedback in class was useful because they were exactly our target audience. The single biggest impact that they had on our project was on our candidate selection section. Originally we were going to implement many possible candidate match-ups. This was a giant mistake. We really underestimated an American's ability to determine for himself or herself which match-ups could not be shown. Clinton vs Sanders is an obvious match-up, but everyone was confused as to why Clinton vs Sanders vs Trump could not be shown. We took the feedback and changed it into a button for the GOP race and the Democrat race.

Another piece of feedback that we were able to implement had to do with our UI. The map was in an iframe, which bothered everyone, but we had gotten used to. The progress bar at the top was also too large as seen in Figure 8. We changed this up by shrinking the progress

bar and adding the map to the html file. Adding the map to the main html took a while but our users convinced us that it was worth it.

There were also some fun features we did not get to implement. Coloring the calendar days would be a good way to better use our space. It was suggested that we color the days based on who is currently winning that day. Another idea would be to color the days based on events that happened in the election. For example, all the super tuesdays could be highlighted a special color. Another person suggested making up data for the less populated states. He thought it was frustrating that we did not just color some of the midwest even though there was no data to color them by.

User Study

The user study will done in person. Participants will receive $25 per hour and this whole process should take no more than 30 minutes. Participants will all use the same Lenovo thinkpad to do the trails. The browser will be Google Chrome and a mouse will be provided.

Participants:

Our participants can be anyone comfortable with a browser. They should have common knowledge of the US election process. Foreign nationals should be avoided (if you're from a parliamentary system this isn't going to make sense for outside reasons). Users should be screened for color blindness before hand and reject if they are color blind. We are looking for 100 participants for overkill on the statistical significance. Each user will answer twenty questions of the format below.

Who/What:

Variations on the following:

Determine which party candidate is winning in state on date?
Who is the winning candidate on date (tests progress bar)?
How many state delegates does candidate have on data?

Time Trials:

Users will be timed on how long it takes for them to find the correct answer.
Use the keyword sort to find 3 Trump headlines. Versus not using the headlines
Use the gradient to find the closest races. Switch gradient off and party and find the closest races. Time both and see if it happens faster with gradient.

Rate (1-5 with 1 being low and 5 being high):

Colors
Gradient (useful/clearness)
Keyword relevance
User friendliness

Core Features

Parse multiple newsites' rss feeds:
> Articles are pulled from a variety of sources. More sources can easily be added by adding the rss feed link to a text file.

Get keywords from articles:
> Words that appear in multiple headlines are used as keywords. Keywords are also connected together. They are ranked on importance in the keyword list based on usage and connectivity.

Display polling data and delegates on the map:
> Polling and delegate data is shown on the map for each of the states. Getting the map to take in the different types of data was a challenge. Getting everything to appear for specific dates was also challenging because it required a lot of consistency throughout the project. Having missing data from our sources was also an issue that we solved by putting the data in a custom JSON formats.

Display number of delegates each candidate has in a progress bar above the map:
> This feature is the only way the user can actually tell who is winning overall. Since the size of the state doesn't have anything to do with the delegates, this is vital.

Change displayed party:
> Users can select between the Democrat or GOP primaries. This allows us to show both races and have a single large map. A race variable is passed into our D3 draw function which uses only that relevant data.

Hover tooltips for states:
> Hovering over each state for a tooltip to popup and show the exact amount a candidate is winning by. This feature is necessary because it lets users objectively compare states.

Reset button:
> Resets the webpage to the current day, the GOP race, no selected keyword, and the gradient is toggled on.

Ability to turn on and off the color gradient on the map:
> The gradient could be a distraction if a user just wanted to figure out who is winning which state. The gradient can be toggled off to a potentially less distracting view which is done by picking the boldest color for each candidate and only using that. All the colors were taken from colorbrewer.

Sorting articles based on keywords:
> Clicking on a keyword will put any headline with that keyword at the top of the headline list.

Change the date and change data for that date
> This was one of the most time consuming features because all of our information needed to be dated and accessible. Going to a date on the calendar will show the articles, headlines and polling for that day.

Running through the election:

>   To get an idea of how the election progressed up to a day the user would have to do a lot of clicking without this feature. Simply hitting the run button will cause webpage to cycle using the monday of each week in the election. This creates a "gif" look on the map that we have gotten a lot of positive feedback on.

Challenges:

>   Missing data was a big issue we had to overcome. Some states did not have any data at all. Data fields in polling data would sometimes not be populated correctly or consistently. Some states also had outdated data with only candidates that were no longer running. Different polls would surprise us with state election data that had nothing to do with the primaries but were still present in the data.

Packages and Leveraged Code

Webpage:
- Angularjs
- Bootstrap
- Ajax

Map:
- D3
- simple US map in D3 http://bl.ocks.org/NPashaP/a74faf20b492ad377312

Backend:
- Beautifulsoup
- NLTK
- Flask

Work Breakdown Structure

Webpage:

>   The UI layout was done largely by our SD&D group member Darren with a slight amount of help from Garrett on the calendar. Garrett and Steven did the coloring and picking between candidates. Garrett implemented the run feature for the webpage which progressed through the election. Garrett also wrote the reset button functionality, while Darren made the button itself.

Map:

>   Garrett and Steven implemented the map and all the data interfacing. Steven did the coloring and the gradient. Steven wrote the progress bar above the map as well. Garrett implemented the map changing data between the two parties and date selection. Garrett also wrote the tooltips such that the GOP's could accommodate the extra candidates.

Backend:
Anders did all the web parsing with RSS feeds Steven found and did most of the backend. Steven wrote the basic Python server and Ajax calls that Anders then modified.

To further separate the classes:

Software Design and Documentation
- Headlines and articles
- Calendar
- Basic Map Functionality
- Data parsing

Interactive Visualization
- Color Gradient and Colors
- Progress Bar
- Reset Button
- Run Through the Election functionality
- Party Tooltips

Future Work

If we were to develop beyond this course we would have to change the UI around to work for a general election. We would also clean up the UI and label our features so people knew what was going on just by the page alone. The article list should be in a small element that you can scroll through so the map is always in view. Being able to click on a state and have it zoom in on a congressional district map would be useful. Our run feature could be improved by having the user select beginning and end days. Our webpage background could be enhanced as it just has a white background, which does not look the best. A background image like the course page would probably be an improvement.

Code

The code that we wrote is mostly in "apps/control/static/web/js/map.js". The code can also be found at:

https://github.com/sirmarcis/Trump-Stats-Mapper

Other

The website is currently being hosted on heroku at:
http://trump-stats-mapper.herokuapp.com/

Bibliography

"2016 Republican Presidential Nomination." 2016 Republican Presidential Nomination. 270 to
    Win, n.d. Web. 03 May 2016.

Brewer, A. Cynthia and Harrower Mark. (2003). ColorBrewer.org: An Online Tool for Selecting
    Colour Schemes for Maps. *The Cartographic Journal*, Vol. 40, No. 1, 27-37.

"RealClearPolitics - Election 2016 - 2016 Republican Presidential Nomination."
    RealClearPolitics - Election 2016 - 2016 Republican Presidential Nomination.
    RealClearPolitics, n.d. Web. 03 May 2016.