

# Data Viz Final Portfolio

Bejamin Roth

12/15/2019

## Introduction

For my data visualization, I focus on campaign data from the 2020 Democratic Presidential Primary. Most of the data I use is from the Federal Elections Commission (FEC), and consists of campaign contributions and campaign spending. I also used polling data aggregated by 538. Initially, I had included data ranging from January 1, 2017 to September 20, 2019 (the day before I downloaded the data). After looking at the data, however, I decided to change the date range to January 1, 2019 through June 30, 2019. Not many observations occurred before 2019, and the data appeared incomplete past June. I break up my 11 graphs into 4 sections:

1. Disbursements
2. Contributions
3. Disbursement and Contribution Comparison
4. Polling

As I believe will be highlighted in the first section, the disbursement data from the FEC seemed of low quality. The disbursement data broke each observation into one of two factor levels: broader disbursement categories, and more granular disbursement descriptions. It seemed that there was an issue matching categories to descriptions, as ~90% of the data appeared to be labeled as 'Other'. While I had tried resorting the descriptions into broader categories, I had difficulty matching the roughly 1,800 disbursement types into categories in a way that made sense.

Overall, I enjoyed the opportunity to expand my visualization abilities and explore campaign data. I also found that manipulating the data into the proper format to graph was challenging at times. The aspect of the project that I struggled with most was choosing colors of certain graphs. With two graphs that had 10+ candidates, it took some time to find the right combination of colors so that each candidate was discernable enough from all of the others. I used ggplot2 to make initial versions of the graphs, and then Adobe Illustrator to add finer details. There were a few issues with some of the graphs when exporting them from Illustrator; I could not fix certain distorted text boxes.

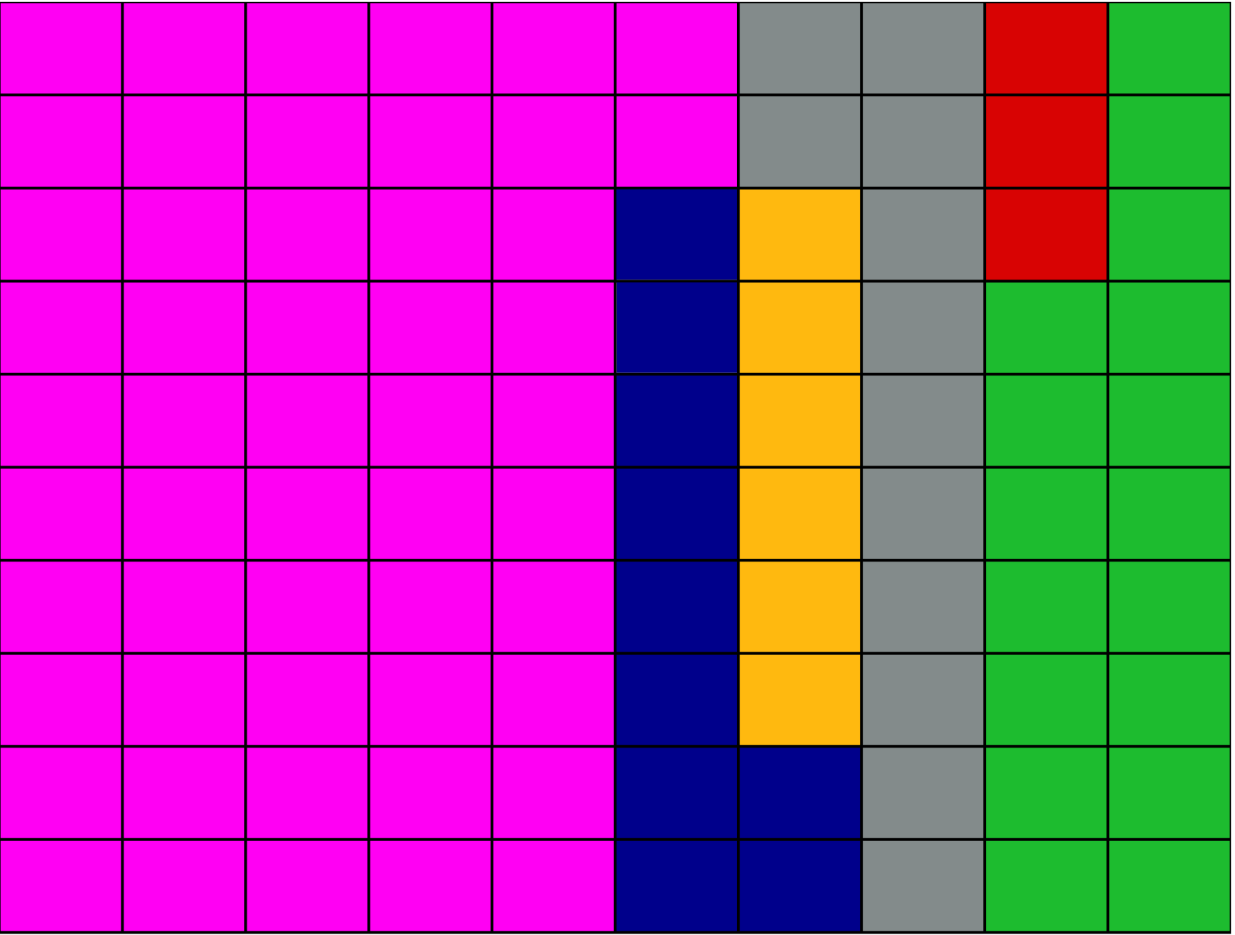
## Disbursements

### Graph 1

The first graph is a waffle graph that compares how often candidates spend money on a certain disbursement category. Initially, I was surprised to see administrative expenses account for so many spending items. However, after thinking about how much campaigns spend on salaries, supplies, rent, etc., I was able to accept that result a little more. Additionally, I was surprised to see how little campaigns spent on events and advertising, though I believe that may have been more of a data quality issue

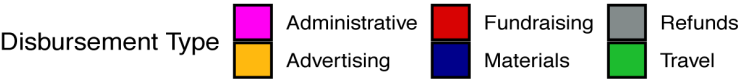
# Adminstrative Expenses Account for More Than 50% of Campaign Spending Items

Composition of Disbursement Types in Democratic 2020 Presidential Primary



Each square represents approximately 1% of the total number of spending items.

Per the FEC, disbursements labeled as Loan-Repayments, Transfers, Events and Contributions each account for less than 1% of total spending items.

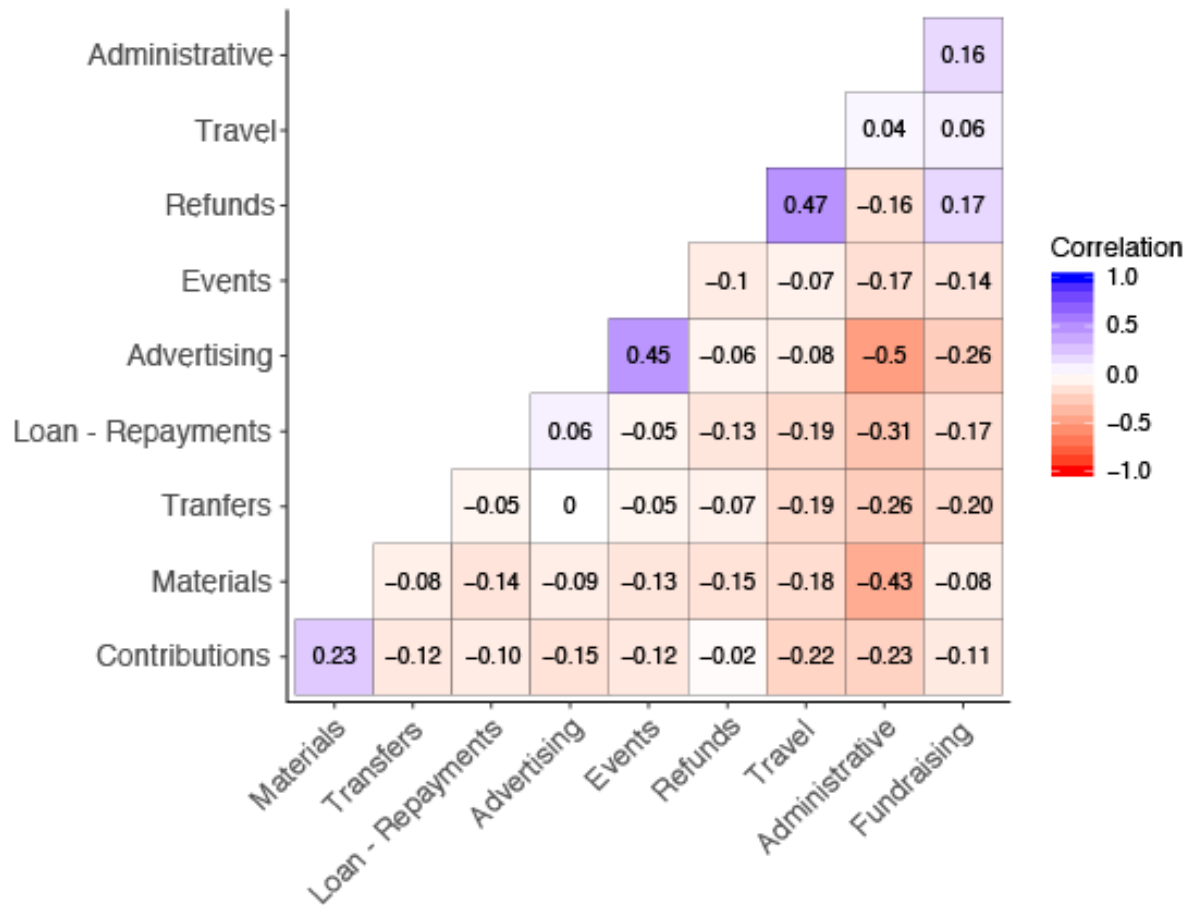


Source: FEC Data, Date Range: 1/1/2019–6/30/2019, downloaded 9/21/2019

## Graph 2

The second disbursement graph seeks to understand correlations between disbursement categories. I would have expected more correlation between categories, as no correlation is stronger than 0.5; for example, I would expect something like a negative correlation between events and advetising.

## No Strong Correlations Between Disbursement Spending Categories For 2020 Democratic Presidential Primary Candidates



*Note: Disbursements labeled as "Other" not included in the calculations*

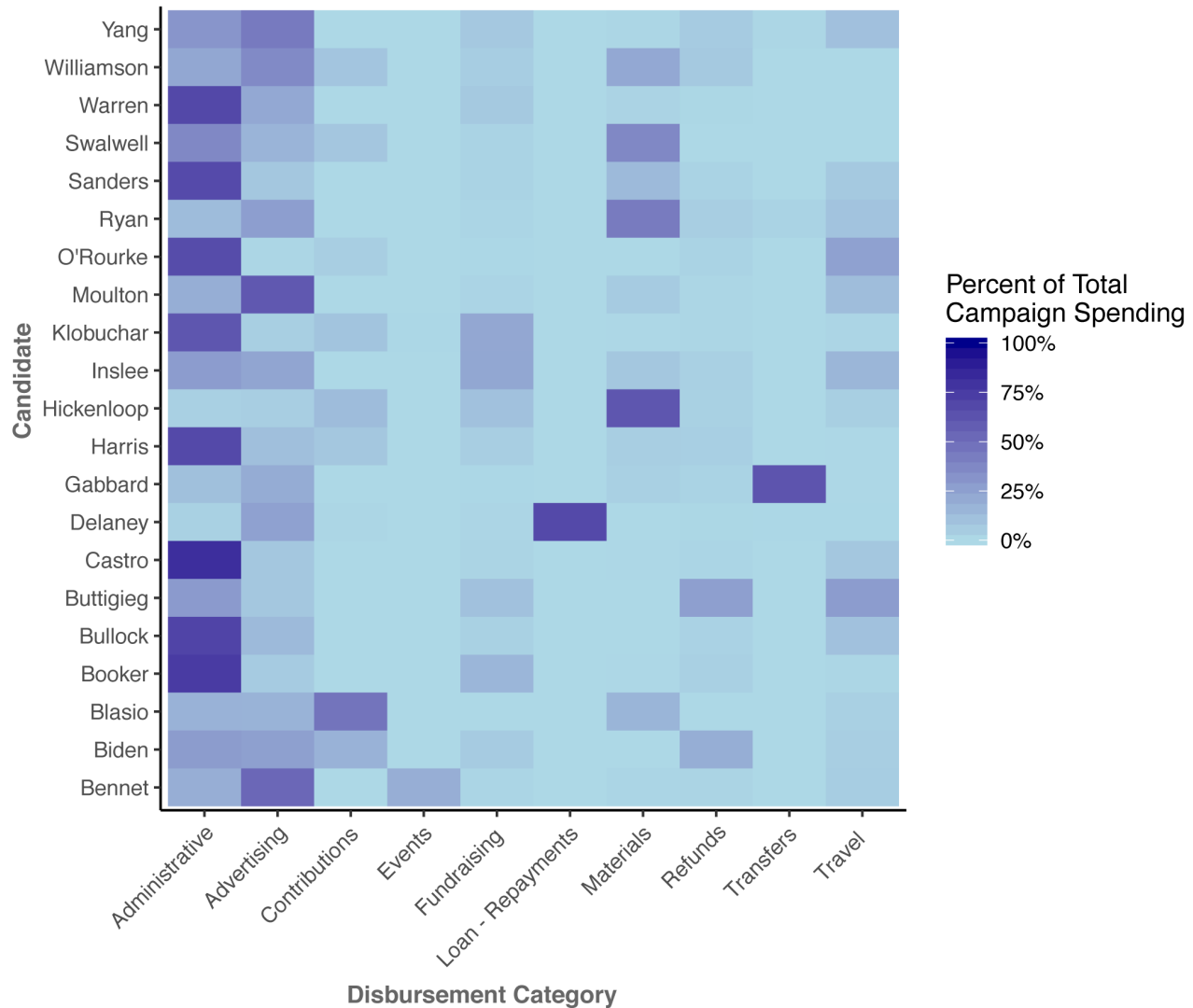
Source: FEC Data, Date Range: 1/1/2019-6/30/2019, downloaded 9/21/2019

### Graph 3

Next, I created a heatmap to compare what disbursement categories campaigns spend their money on. Although it's similar to the waffle map in that campaigns spend the largest portion of their total budget on administrative related expenses, it differs in that campaigns appear to spend more on advertising than the waffle map would suggest. I think that this visualization is necessary and different from the waffle chart, as costs of disbursements could reasonably differ from the number of times that an item is purchased.

## Campaigns Spending Majority of Funds on Administrative Expenses and Ads

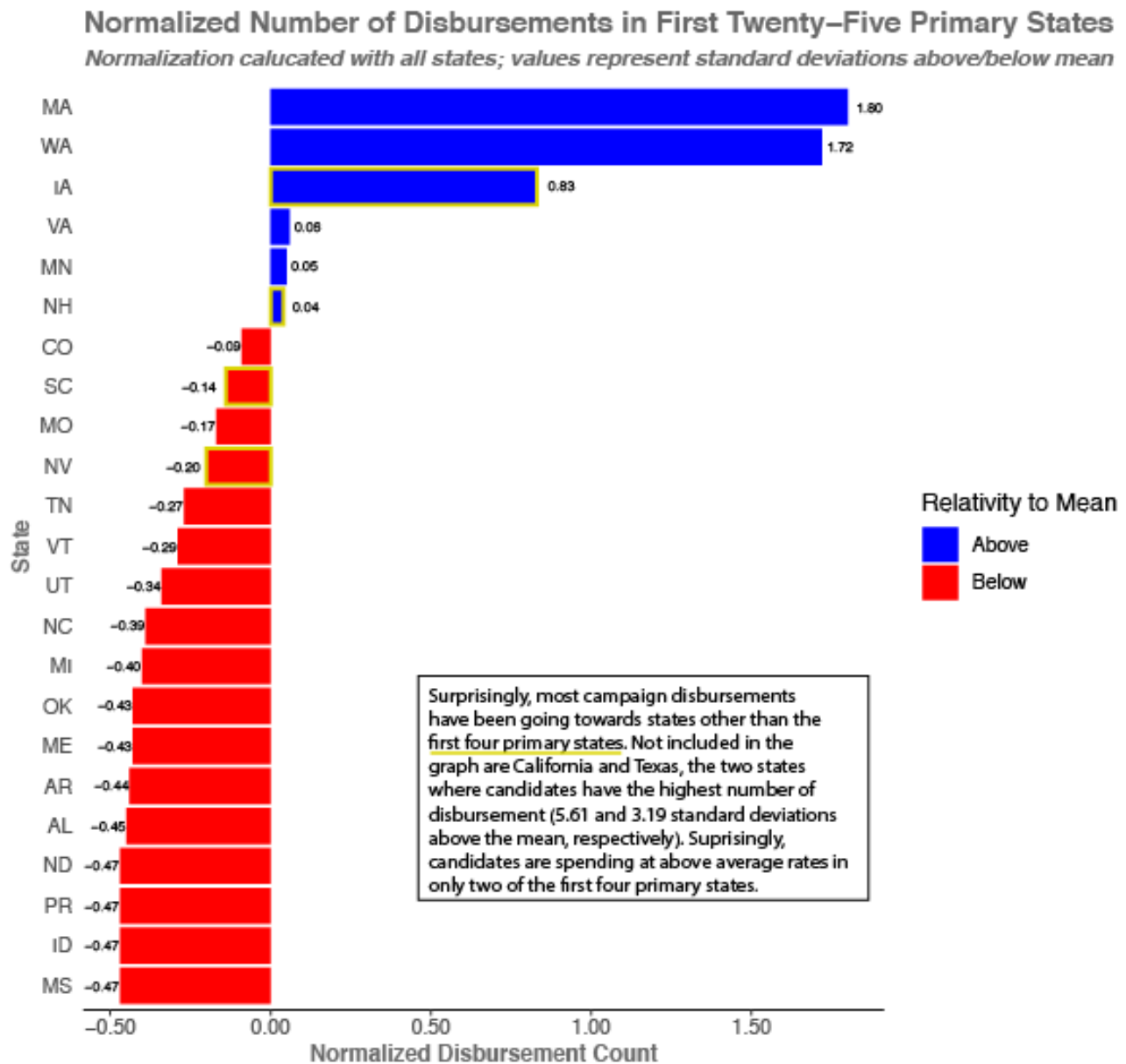
*Heatmap of Campaign Spending in Major FEC Categories, 2020 Presidential Election Democratic Primary Candidates*



Source: FEC Data, Date Range: 1/1/2019–6/30/2019, downloaded 9/21/2019

## Graph 4

The final disbursement graph seeks to understand how much money campaigns are spending in ~20 of the first primary states. The total amount of spending in each state was normalized, so that state spending amounts could be compared relative to the average amount that candidates are spending in each state. I would have expected campaigns to spend the most in the first four primary states, but as the graph shows, that is not the case. Texas and California were removed from the data because they were outliers that skewed the normalized results.



Source: FEC Data, Date Range: 1/1/2019–6/30/2019, downloaded 9/21/2019

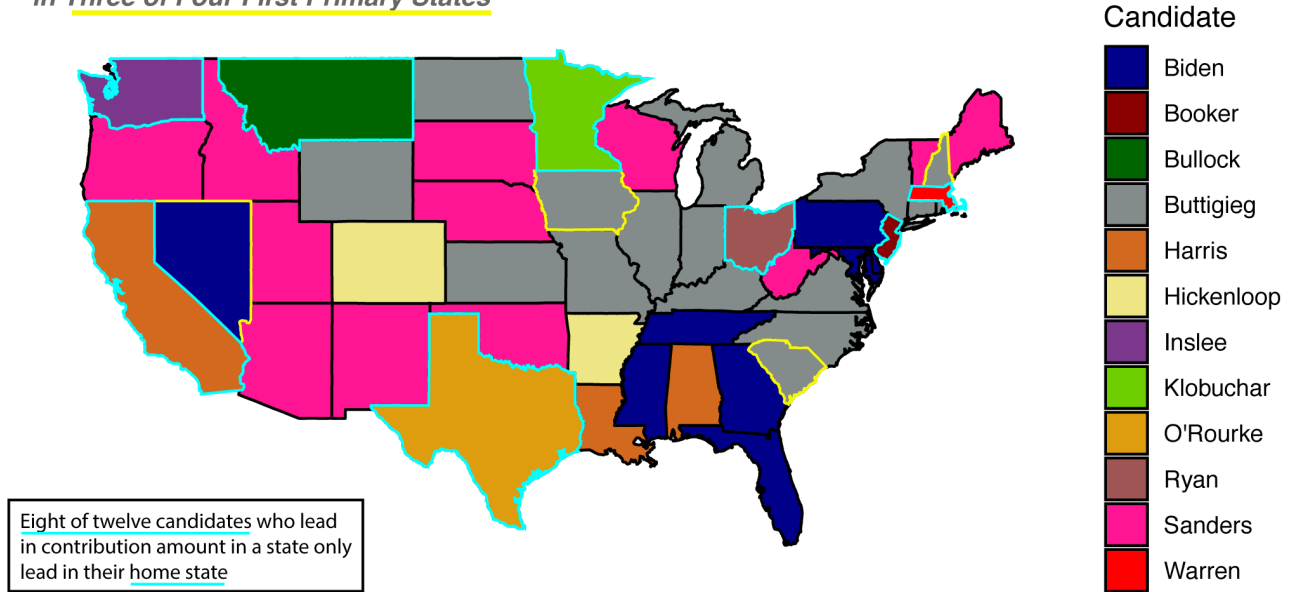
## Contributions

### Graph 5

The first contribution graph seeks to visualize which candidates are raising the most money in each state. There are quite a few candidates that have leads in different states. Although the graph itself offers more analysis, I would like to point out how inexplicable it is for Hickenlooper to have the lead in Arkansas, given his lack of polling popularity.

## Democratic 2020 Presidential Primary: Highest Campaign Contribution Amount, By State

*Through First Six Months, Pete Buttigieg Leads in Total Fundraising Dollars in Three of Four First Primary States*



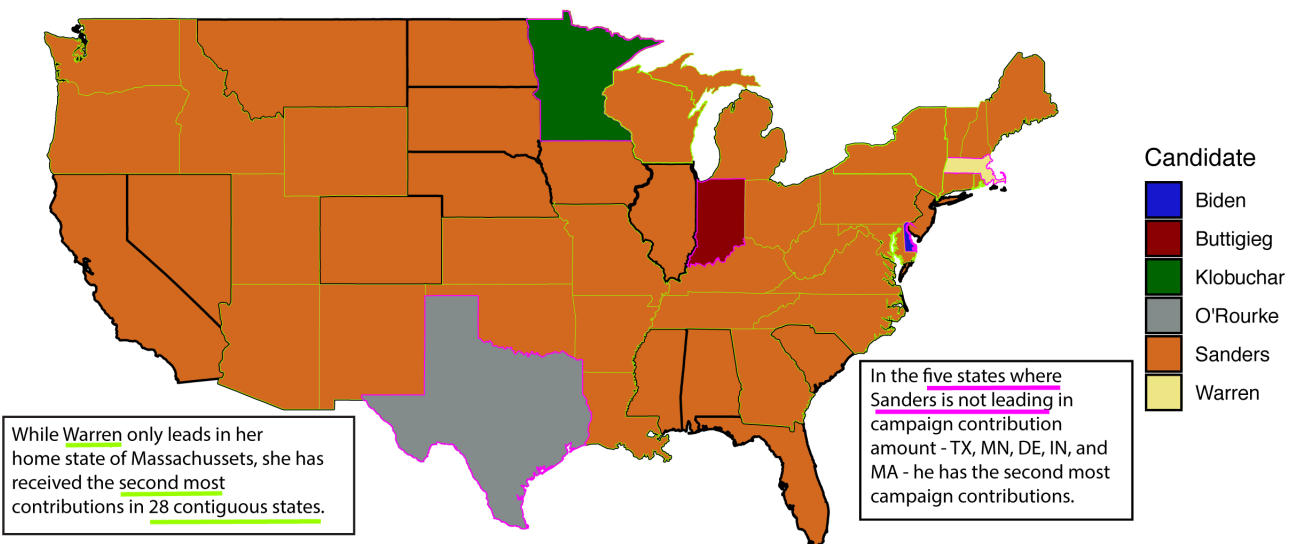
Source: FEC Data, 1/1/2019 – 6/30/2019

### Graph 6

This next map highlights the difference in which candidates are receiving the most money in each state, and which candidates are receiving the most contributions in each state. There is an incredibly stark difference between how many candidates appear in this graph and the previous graph. The popularity of Sanders is truly striking.

## Democratic 2020 Presidential Primary: Most Campaign Contributions, By State

*Contributions made between 1/1/2019 and 6/30/2019*

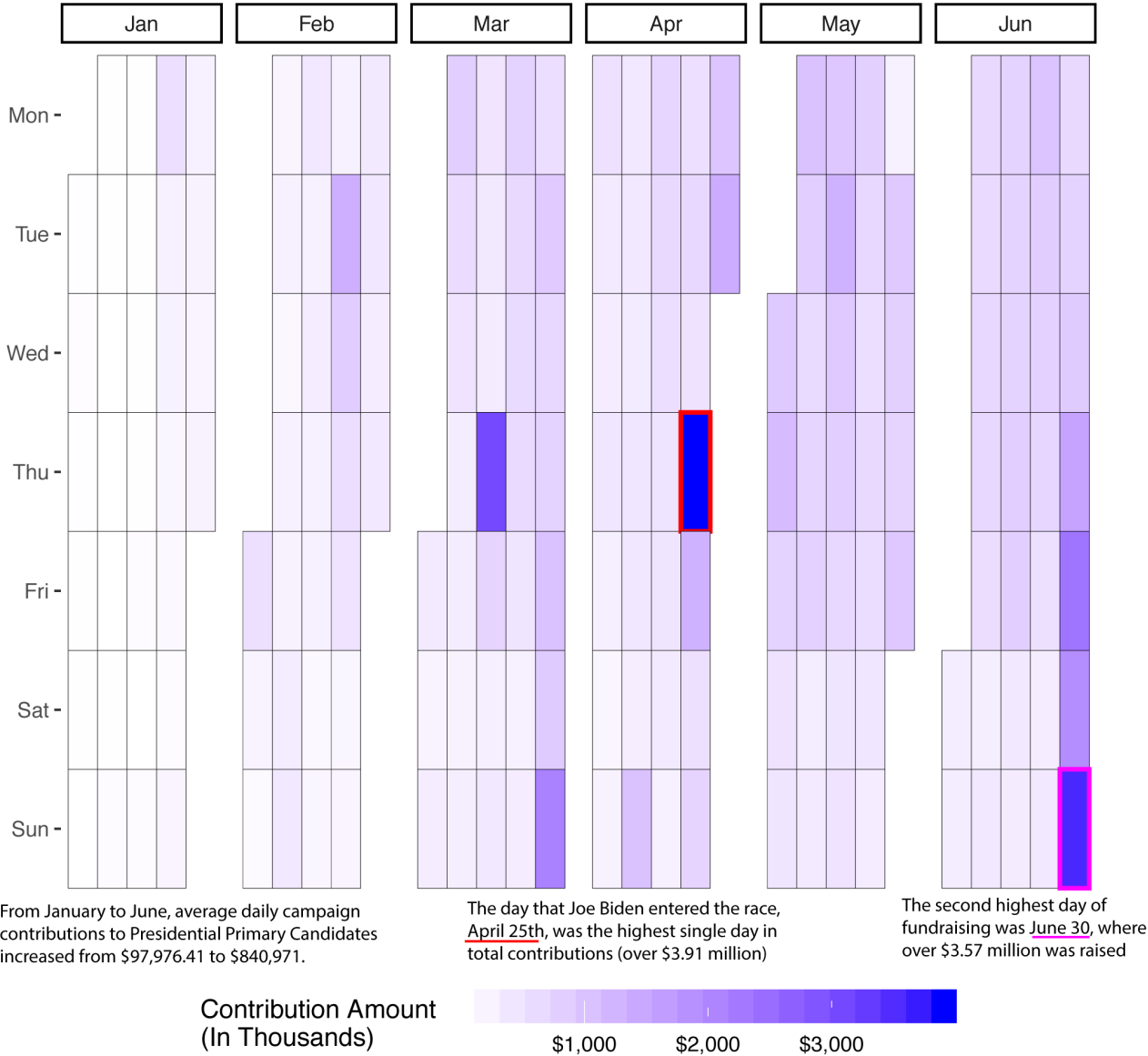


Source: FEC

### Graph 7

# Biden and O'Rourke Entries, Ends of Quarters Stand Out Early in 2020 Democratic Presidential Primary Race as Highest Fundraising Days

Source: FEC Data, downloaded 9/21/2019



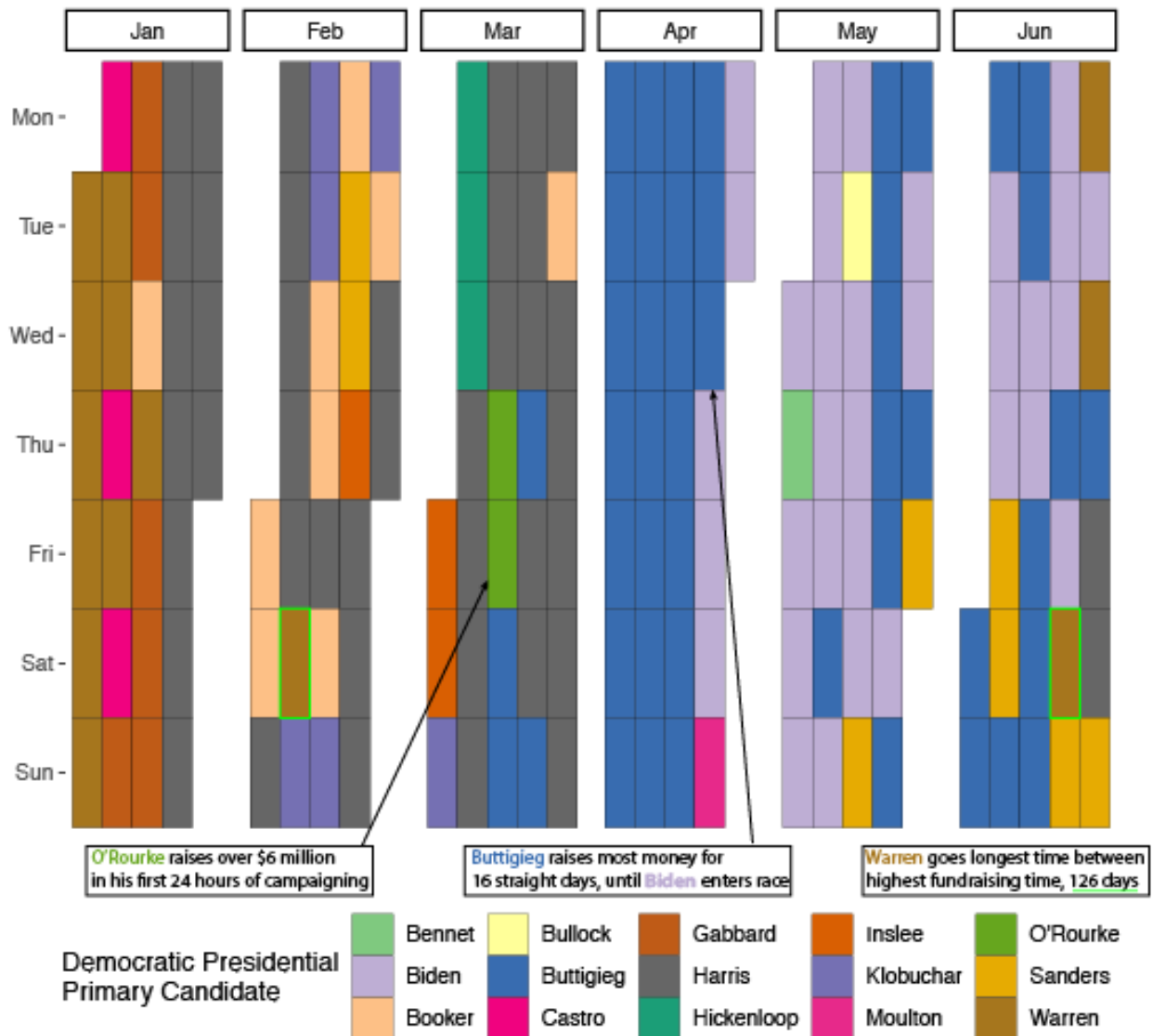
Code: "Time-Series Calendar Heatmap" by Sarang Gupta, [http://www.columbia.edu/~sg3637/blog/Time\\_Series\\_Heatmaps.html](http://www.columbia.edu/~sg3637/blog/Time_Series_Heatmaps.html)

## Graph 8

The second calendar visualization examines which candidates received the most money each day for the first six months of 2019. I think that it does a good job of telling the story about which candidates were relevant during what time periods of the campaign

## Through First Six Months of 2019, Fundraising Days Show Frequent Shifts in Candidate with Highest Fundraising Day

Source: FEC Data, downloaded 9/21/2019



Code: "Time-Series Calendar Heatmap" by Sarang Gupta, [http://www.columbia.edu/~sg3637/blog/Time\\_Series\\_Heatmaps.html](http://www.columbia.edu/~sg3637/blog/Time_Series_Heatmaps.html)

## Disbursement and Contribution Comparison

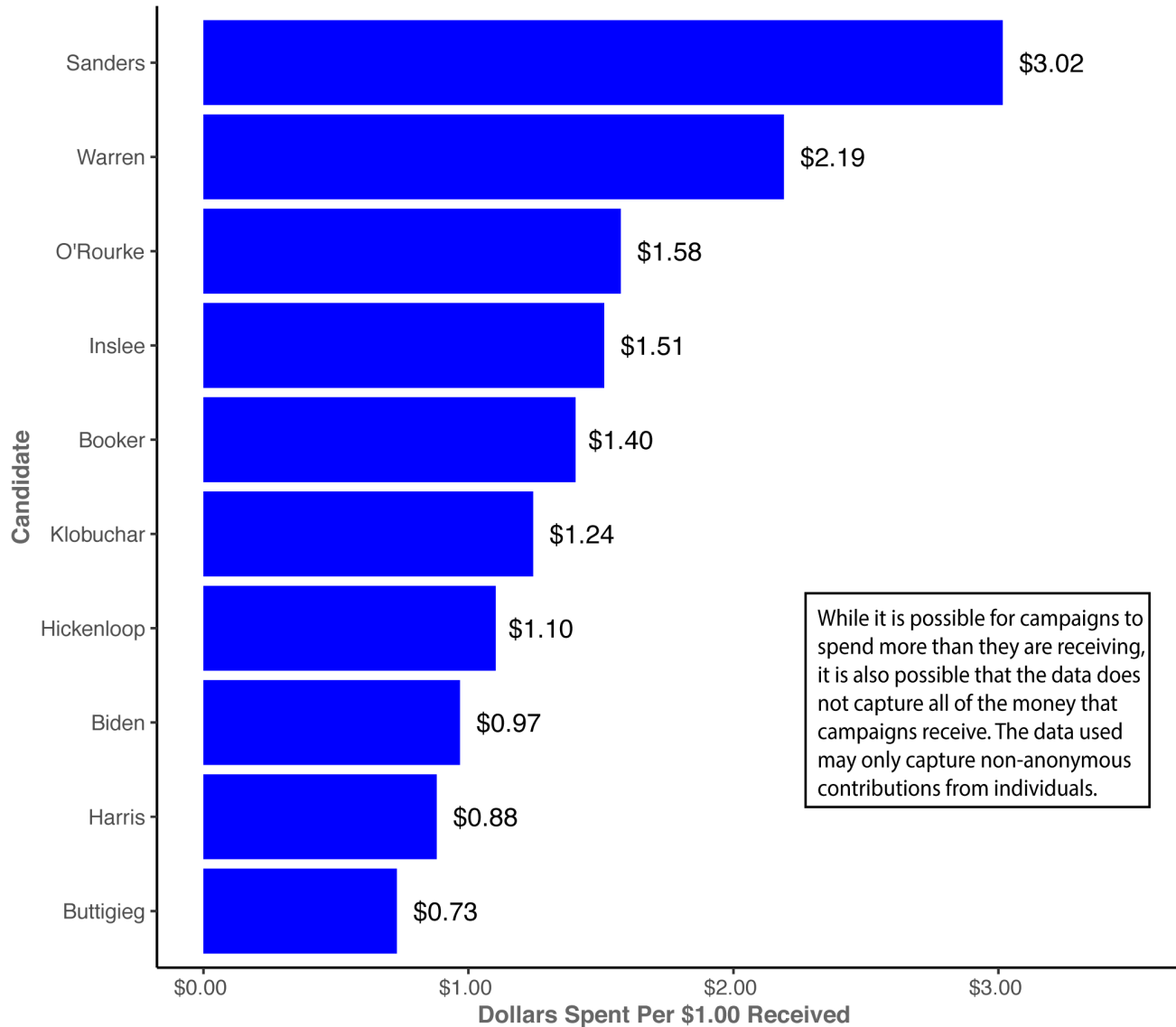
### Graph 9

The first comparison between contributions and disbursements focuses on how many money campaigns are spending compared to how much they are receiving. I found it interesting that the two candidates whose messaging most highlights grassroots donations, Warren and Sanders, are also the candidates with the highest ratios of dollars spent to dollars received.



## Dollars Spent per \$1.00 Received: Democratic 2020 Presidential Primary Candidates (Ten Candidates Who Have Received the Most Contributions)

*Per FEC Data, A Majority of Top Candidates are Spending More Than They Receive*



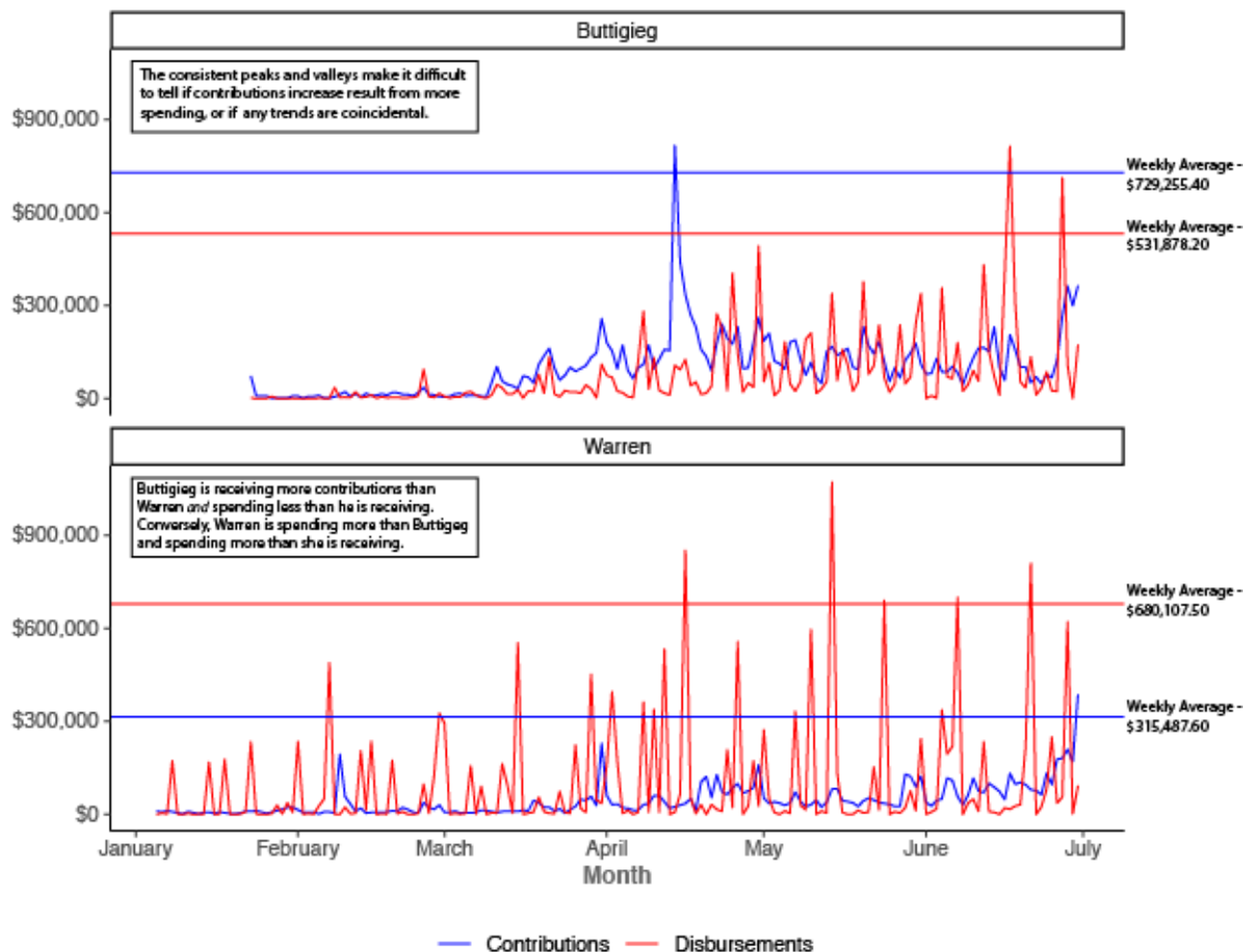
Source: FEC Data, Date Range: 1/1/2019–6/30/2019, downloaded 9/21/2019

## Graph 10

Next, I wanted to both determine if there was a relationship between spending and receiving and see how two of the top candidates compared in terms of their spending and contributions trends over time. My hypothesis would be that candidates who spend more money on staff, events, advertising, etc. would see an increase in contributions. Additionally, I would have guessed that top candidates would display similar contribution and disbursement trends over time. As the graph suggests, neither of those hypotheses were correct.

## Comparing Dollars Spent and Dollars Received During 2020 Democratic Presidential Primary: Buttigieg and Warren

*Highlighting Contributions and Spending of Two Top Candidates displays uncertainty in relationship and differing trends between two campaigns*



Source: FEC Data, Date Range: 1/1/2019–6/30/2019, downloaded 9/21/2019

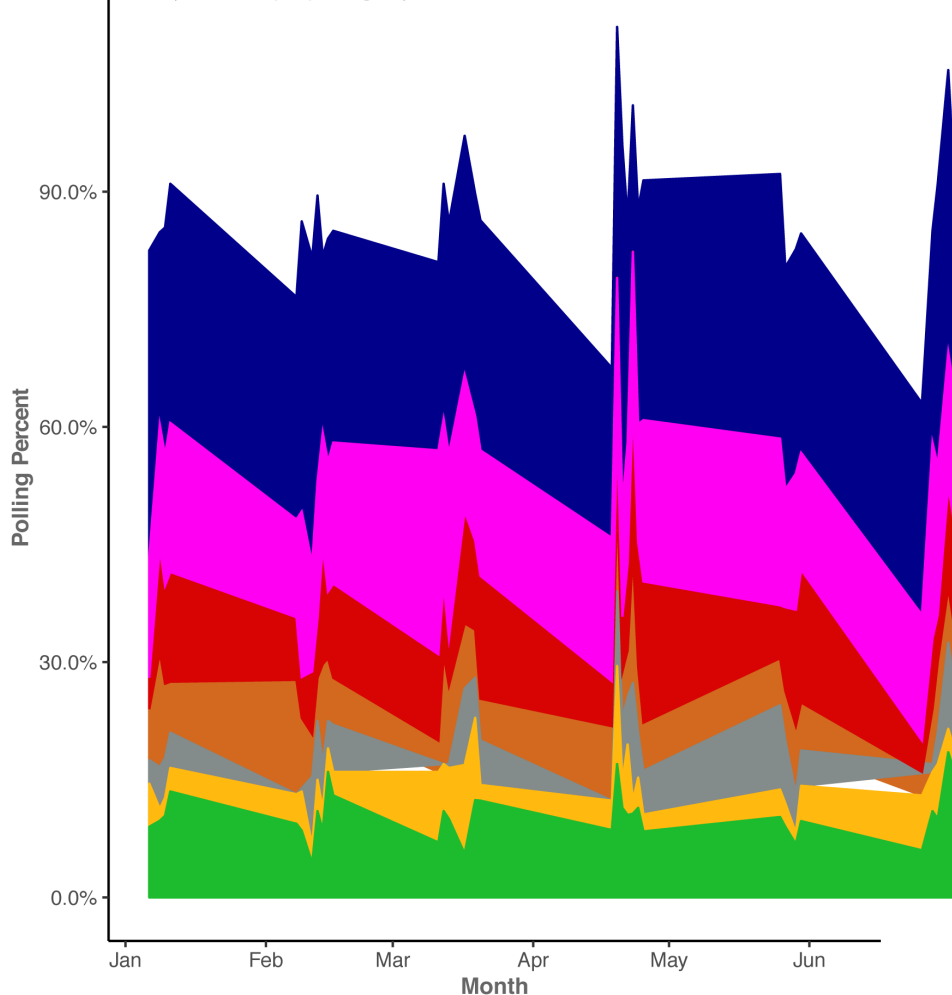
## Polling

### Graph 11

For my final graph, I wanted to see how national polling has changed over time for each of the top candidates.

## Joe Biden Maintains High Daily Polling Average Throughout First Six Months of 2019, Rest of Field Highly Variable

Thirty-four unique polling days



Biden's daily polling from January through June has averaged slightly under 30%

Sanders has also remained relatively consistent in polling, with his daily average ranging from 12.5% to 26.0%.

Warren and Harris have seen their polling averages vary at a greater level, ranging from 4.0% to 23.2% and 3.0% to 16.67%, respectively.

While Buttigieg and O'Rourke have had slightly lower similar polling averages of 0.4% to 15.0% and 1.4% to 12.5% respectively, the two men had only four days polling in the double digits combined, whereas Warren and Harris have 25 double digit polling days between them

Other candidates not in the top have, combined, averaged less than 2.5% in polling throughout the first two quarters of the campaign. Taking the max daily polling average across all other candidates, a daily polling average of 5.0% has only been broken five times.

Source: 538 Aggregated Polling Data

## Data Sources

- FEC Contributions
  - <https://www.fec.gov/data/receipts/individual-contributions/>  
(<https://www.fec.gov/data/receipts/individual-contributions/>)
- FEC Disbursements
  - [https://www.fec.gov/data/disbursements/?data\\_type=processed](https://www.fec.gov/data/disbursements/?data_type=processed)  
([https://www.fec.gov/data/disbursements/?data\\_type=processed](https://www.fec.gov/data/disbursements/?data_type=processed))
- 538 Polling Data
  - [https://projects.fivethirtyeight.com/polls-page/president\\_primary\\_polls.csv](https://projects.fivethirtyeight.com/polls-page/president_primary_polls.csv)  
([https://projects.fivethirtyeight.com/polls-page/president\\_primary\\_polls.csv](https://projects.fivethirtyeight.com/polls-page/president_primary_polls.csv))