Mapping Donor Engagement: Strategic Insights for Fundraising Success

Compiled and Analyzed By Patrick Norton

February 27, 2025

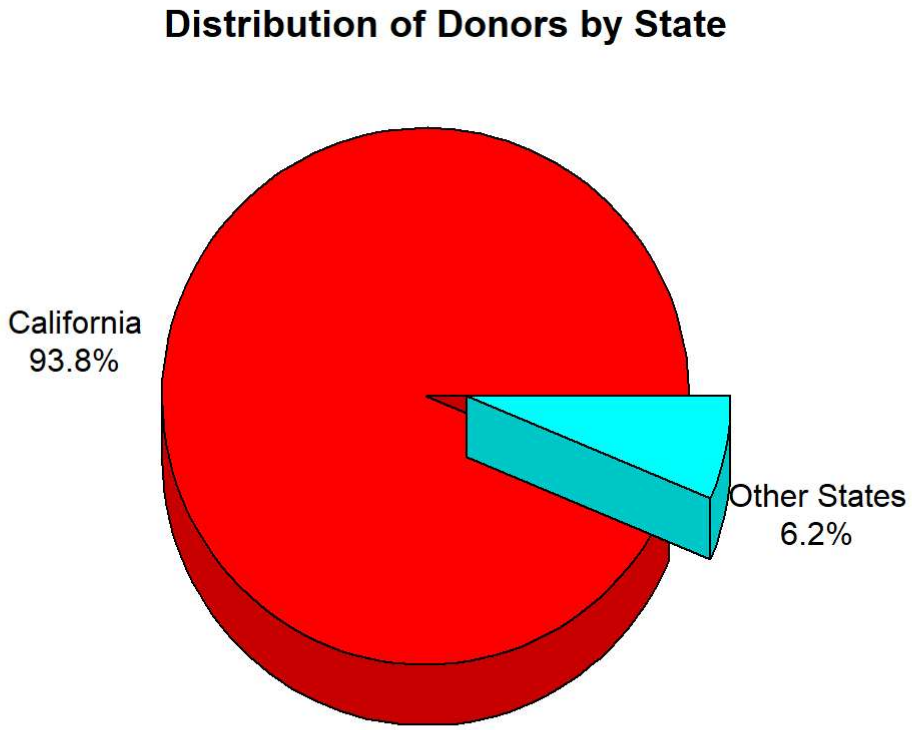
Introduction

First off, thank you for taking the time to read my insights and what I have gathered so far on donors within NewOrg's system. Though I do not have the means to fully excavate all the data and draw every insight at the time of writing this report, my hope is that this report will give a solid understanding and visualization of some key characteristics of our donors and how we have introduced new donors to the message and mission of PARCA.

The dataset that I chose to disect and look into is a comprehensive dataset that has pooled both new donor contacts into our donor list from October 2023 until January 2025 as well as imported contact information from the previous system pre-September 2023. I cleaned and processed the data to create some general visualizations as well as predictions using predictive linear regression models that give us a glimpse into a predictive window for 2025.

Basic Demographic Visualizations

The first and most basic representation that I want to point out is the geographical location of our 4,000 or so contacts, individual and corporation, within the United States. Here is a simple pie chart that separates all donors based on if they're from California or another state.



Within the state of California, the 93.8% represents approximately 3,700 individuals, households, and businesses that have contributed to PARCA on record. Out of the 3,700, about **3,484 were labeled households/individuals while 217 were from foundations or organizations**. Since the majority of states within the US were somewhere in between 0.1% to 1%, I grouped them all together into one category. Although the percentages are seemingly obvious considering the scale of our operations is primarily within the Bay Area, I created a table that will give some interesting insight into where our donors are coming from which will answer:

Where are our donors coming from?

To answer this, I have created a table that lists the top five cities from the top five states our donors are from regardless of whether the donor is an organization or household. The list is organized by `state` in alphabetical order starting with each state's highest `donor_count` to its lowest .

State	City	donor_count
CA	San Francisco	497
CA	San Mateo	267
CA	Pacifica	250
CA	San Jose	167
CA	Burlingame	154
CO	Denver	5
CO	Boulder	3
CO	Carbondale	1
CO	Castle Rock	1
CO	Highlands Ranch	1
NJ	Princeton	13
NJ	Cedar Grove	2
NJ	Florham Park	1
NJ	Hamilton	1
NJ	Hawthorne	1
NV	Reno	5
NV	Henderson	2
NV	Sparks	2
NV	Gardnerville	1
NV	Genoa	1

State	City	donor_count
TX	Austin	3
TX	Plano	2
TX	Bellaire	1
TX	Cedar Park	1
TX	Coppell	1

This table was particularly interesting to me due to a few outstanding insights:

1. Pacifica has a `donor_count` that is only *17 donors less* than San Mateo, which is significant given San Mateo is ***3x the population size*** of Pacifica. This suggests that Pacifica has a disproportionately engaged donor base relative to its size, deeming it **an outlier** in terms of per capita giving.

2. When I expanded the table to the top 10 cities in each `state`, South San Francisco and Hillsborough had essentially the **same amount of donors** despite their vastly different demographics and community structures.

3. ***Not one town*** outside the Peninsula (except San Jose) made it onto either the top five or top 10.

4. Princeton, NJ has the highest non-Californian donor contacts of any `city` with 13 donors, which are **all from organizations.**

5. Denver and Reno had the **highest non-Californian household donors** with 5.

6. All the aforementioned cities with donors outside of California were imported into the system on the same date in 2023, meaning they have been **donors for at least over a couple years** and **no new donors have emerged** from said places in 2024.

Before I get into the meat and potatoes of this report, I want to give some context for my endeavors into predictive modeling. Initially, I had set out to make some visualizations that compared a couple quantitative variables from the system before NewOrg labeled `Financial.Score` and `Social.Score` with the geographical ones mentioned before. These scores were indexed with a number between 0-99 and awarded to donors based on certain criteria unbeknownst to me. Unfortunately, I have not been able to find any resources that could help me understand the nature of the scores and their significance. Without the proper understanding needed to be able to convey insights honestly and properly, I decided to pull on my background in working with regression models and see if there were any predictions that were worth discussing.

After a couple weeks of vigorously constructing and deconstructing different models, testing and retesting them, I discovered something worth discussing.

Time Series Analysis: Donor Predictions 2025

For those unfamiliar with regression approaches, please read the paragraphs below to get a better understanding of what models I used and why their differences are important to know:

The first model uses an XGBoost regression approach; a machine learning method that leverages decision trees to capture potentially complex, non-linear relationships in the data. It converts dates into a numeric format and normalizes this predictor to forecast donation counts. In contrast, the second approach employs two traditional time series models: ARIMA and ETS. ARIMA (AutoRegressive Integrated Moving Average) is designed to handle trends, autocorrelation, and non-stationarity through differencing, while ETS (Error, Trend, Seasonality) focuses on smoothing components such as trend and seasonal variations. These differences matter because each model type has its own strengths: XGBoost can capture non-linear patterns without assuming a strict temporal structure, whereas ARIMA/ETS are specifically tailored to modeling time-dependent structures, providing interpretable components (like trends and seasonality) along with prediction intervals that quantify uncertainty.

From the XGBoost model, we gain an understanding of the broader trends in donation activity based on historical data. This model can forecast future donation counts and highlight non-linear shifts or sudden changes that might not be immediately apparent with classical time series methods. On the other hand, the ARIMA and ETS models offer insights into the underlying time series dynamics. They not only forecast future values but also help identify seasonal patterns, trend behavior, and provide confidence intervals around predictions. Together, these models can be used to cross-validate forecasts: ***if both approaches signal an upward trend or similar patterns, there is***

higher confidence in the forecasted direction. Ultimately, these insights can help guide our decisions on donor acquisition efforts, resource allocation, and campaign planning by revealing both the magnitude and uncertainty of future donation trends.

XGBoost Model Prediction on 2025 New Donor Entries

This analysis focuses on forecasting monthly donation counts using a time series approach. The initial step involves converting donation dates into a proper date format and filtering the data to include only records from October 1, 2023, onward. The donations are then aggregated by month, ensuring that every month in the period is represented—even those with zero donations.

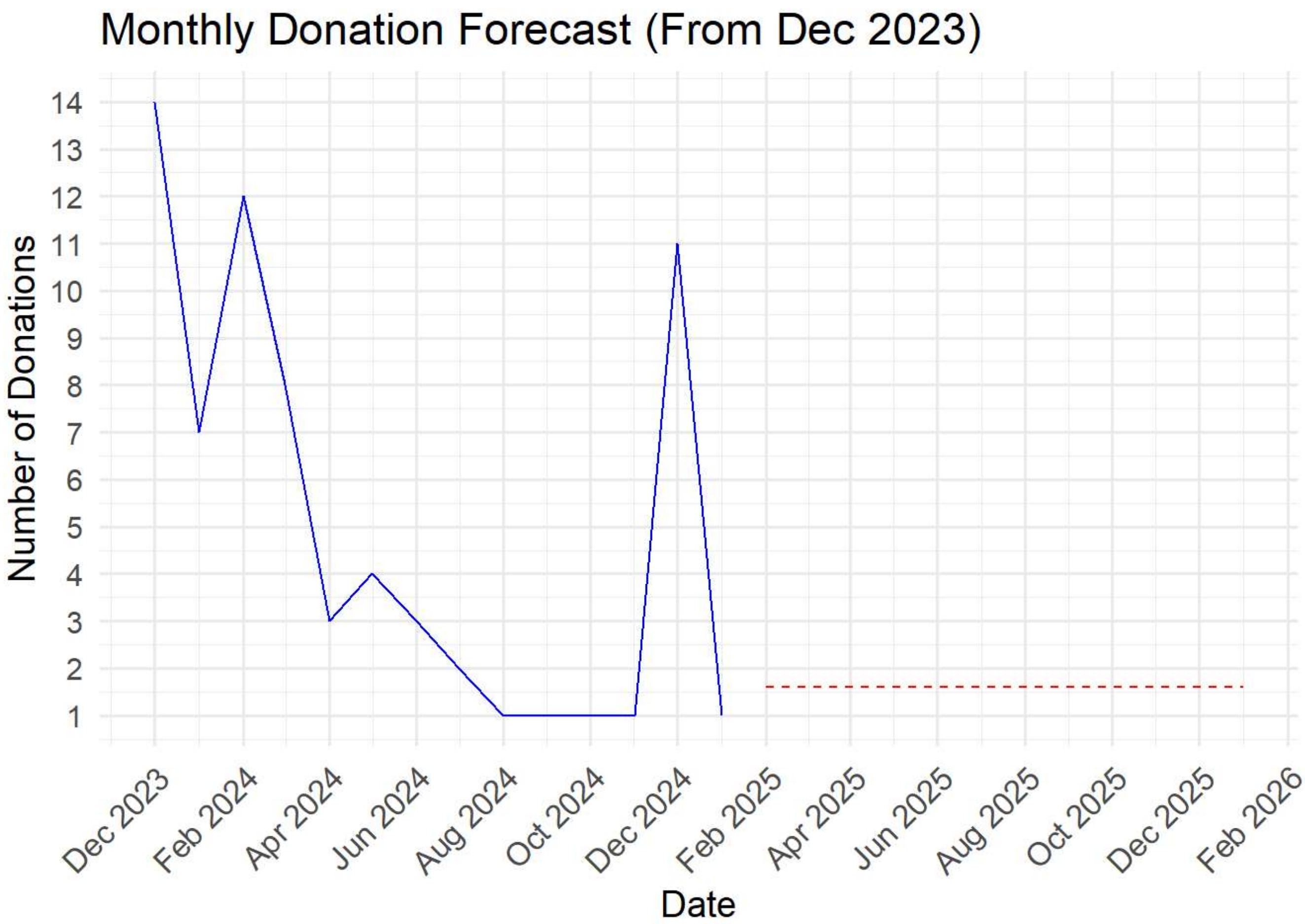
The time variable is converted into a numeric form called `month_num`, which serves as the predictor in the forecasting model. By representing each month as a number (for example, days since a reference date), the model is able to capture the temporal progression in the donation data.

The dataset is split into training and testing sets, with 80% of the data used to train the model. A recipe is defined with the monthly donation count (`n`) as the target variable and `month_num` as the predictor. The predictor is normalized to help improve the performance of the subsequent modeling process.

An XGBoost regression model is specified, employing 500 trees and a minimum of 5 observations per terminal node. This model is then integrated into a workflow with the defined recipe and is trained on the historical donation data. The purpose of this setup is to capture underlying trends and patterns in the donation counts over time.

Forecasts for the next 12 months are generated using the trained model. A new set of future dates is created, and the corresponding numeric values are computed. The model then produces predicted donation counts for these future months, represented in the output as `.pred`. In essence***, these predictions indicate the model's expected monthly donations based on historical trends.***

Finally, the results are visualized by plotting the historical monthly donation counts alongside the forecasted values. ***The blue line represents the actual donation counts from the historical data, while the red dashed line shows the predicted counts***. This visual comparison helps to highlight any expected increases or decreases in future donations.



Looking at the blue line, the majority of our donor entries into NewOrg in 2024 came in around the holidays and remained strong into early spring, with a continual decrease in new donors added each month well into summer and autumn and a massive spike in new donors this past December. The red dotted line is predicting that over the course of 2025, with all events/activities held constant from 2024, ***we should conservatively expect to see an average of 1.6 new donors each month over the course of 2025***. This comes out to be about 19 **new** donations, meaning those that are added into NewOrg this year.

Though this model does take a more conservative stance since it is only evaluating trends based off one year's worth of data and looking at the data with a broader scope, the ARIMA model will give us seasonal insight into the beginning of 2025 and more volatile predictions.

ARIMA Seasonal Forecast Model

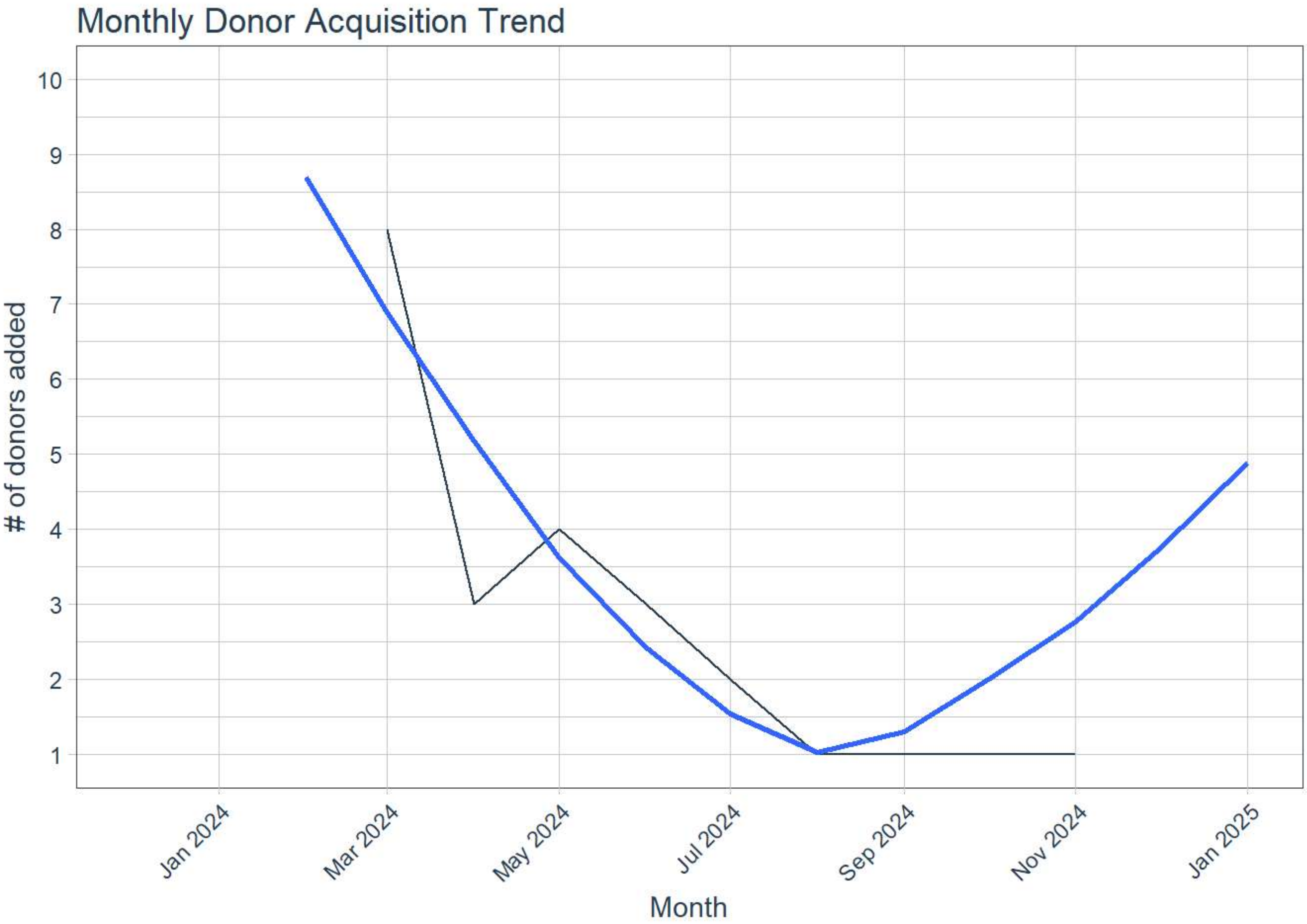
This code begins by preparing the time series data. The donation dates are converted into a standard date format and then aggregated by month using the `floor_date` function. The `complete` function ensures that every month between the minimum and maximum dates is included, filling in any missing months with zero donations. This sets the stage for a continuous monthly time series of donor counts.

Next, the code identifies the first month where there is at least one donation. This is done by selecting the minimum month with a non-zero donation count, ensuring that subsequent visualizations and analyses start from the first meaningful data point.

For visualization, the data is filtered to include only months starting from the first non-zero month. A time series plot is generated to show the monthly donor acquisition trend, with the y-axis limited between 1 and 10 and the x-axis formatted to display every other month. This helps to clearly illustrate the trend over time while managing the scale and labeling for readability.

After modeling, the code generates forecasts for the next 12 months. The `forecast` function produces these predictions along with a 95% prediction interval, which provides a range of likely values (using the `hilo` function). This interval helps to assess the uncertainty in the forecasts.

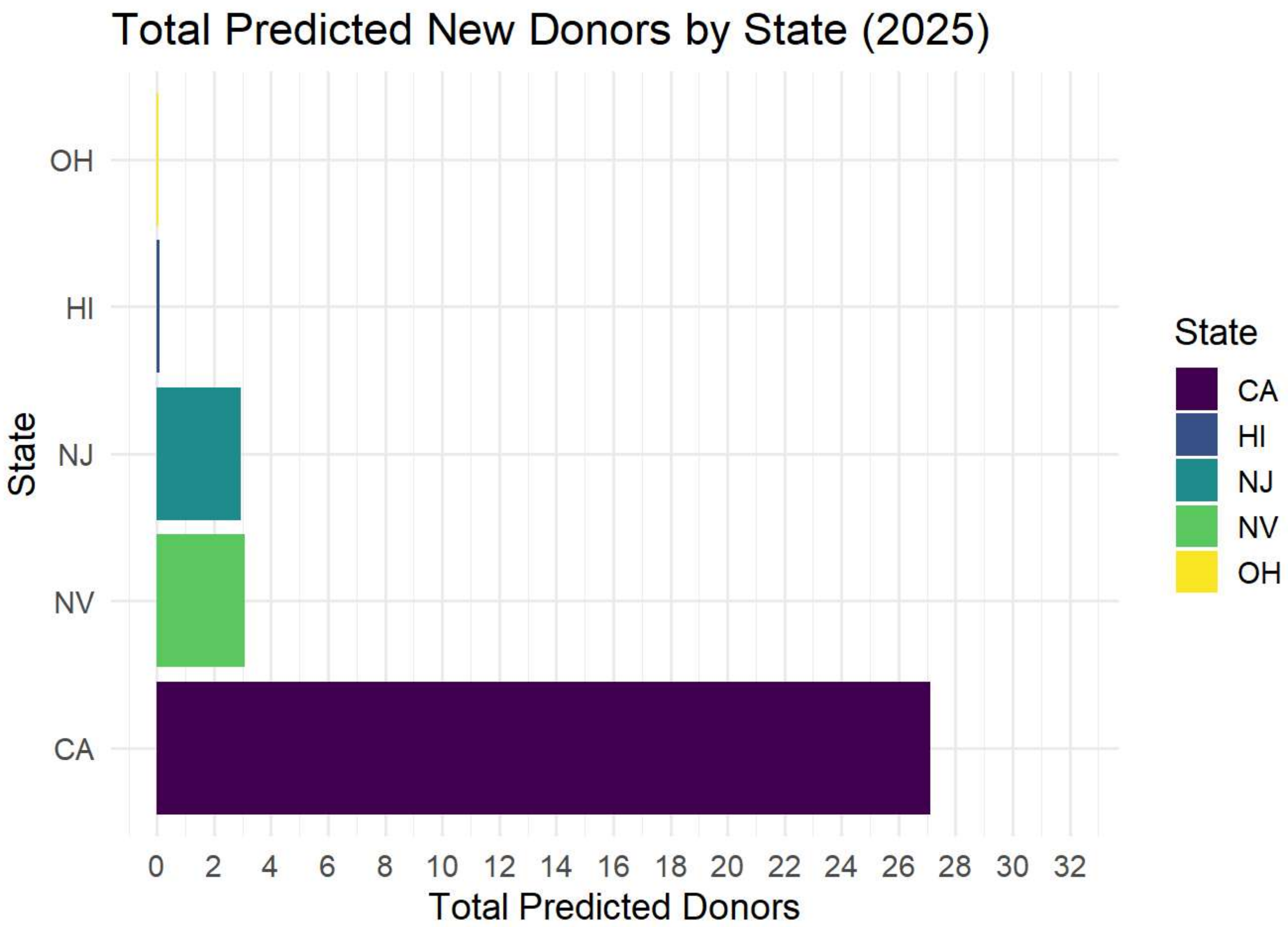
Finally, the forecasted data is visualized alongside the historical data using `autoplot`. The plot is formatted similarly to the earlier visualization, with clear axis labels and date formatting. The resulting graph displays both the past donor acquisition trends and the 12-month forecast, offering a comprehensive view of expected future performance.



In essence, the model's blue trend line takes the sample mean of donations through each month (indicated on the red line) and is telling us that based on the data for 2024, we should have ***expected around 5 donors added this past January***. Since I extracted the dataset in January, I did not have enough for the model to make a prediction on February solely based on the number of new donors. However, based on my own analysis working with the models and seeing the data, ***I believe it would be confidently sensible if the number of new donors added in January is around 8 to 10 with a similar turnout in February and March***.

Time Series Bar Chart for 2025

Much like the first model, this visualization is using the XGBoost method to look at an anticipated macro view of where donors will come from and how many over the course of 2025. These `predictions` were then put into a tibble called `state_predictions` that grouped all 50 U.S states and gave a ranking for the top five highest predicted donor states for 2025.



Evidently, the prediction across all of 2025 is ***about 34 new donors all variables held constant*** (meaning same turnout at fundraising events, steady donation amounts/frequency, etc). This is on par with turnout in 2024, with 31 new donors added into our database over the course of 2024. Again, these new donors take into account any donor regardless of any organization or individual classification. By far and away our top states for donor participation outside of California are **Nevada and New Jersey** with the other 47 individually contributing about the same relative donor rate making up sub-1% of our donor base.

Conclusion

I hope these visualizations I have made as well as the experiments I ran on the data will provide a deeper understanding of where PARCA stands in relation to our donor base. I want to reiterate that the predictions made in this report should be interpreted as the anticipated floor for new contributions made over the course of the year and are subject to change if we decide to do new or different things this year. In other words, if we were to engage communities with the same level of intent as last year, we should come to expect turnout to resemble the graphs. Based on what I have gathered, we should consider doing these three things:

- Exploring potential avenues to increase donors that come from communities in the East Bay, especially those closely connected to San Mateo County (Hayward, Alameda, San Lorenzo)

- Looking into getting Hillsborough more involved with PARCA since the amount of donors from there are equivalent to San Bruno and yet are ***almost 100 donors*** less than nearby Burlingame and San Mateo.

- Consider looking into donors that have either relocated to the Reno/Tahoe area or potentially establishing a network that could help attract new donors from that region since almost all donations from that region were contributed by households and they are our strongest out-of-state contributors in terms of participation.

If there are any questions about the data or more needed interpretation of what I found, I can be reached at patrickn@parca.org. Thank you!