STAT506 Project Report

Kelsey Luo

December 2025

This project demonstrates the R in statistical analysis using data from the National Survey of Nonprofit Trends and Impacts. The dataset contains responses from 2,306 nonprofit organizations and includes 359 survey items describing organizational characteristics, service delivery, staffing, financial conditions, and operating environments in 2021. These variables allow me to examine how nonprofits adapted their programs and finances during the COVID-19 period and how these adjustments relate to organizational mission and community needs.

Based on the categorization of organizational characteristics and behavioral variables, three analytic dimensions emerge and are linked for analysis. To address these connections, I pose the following research question:

> How are organizational actions associated with funding structures, and how do these relationships shape service scope?

The analysis proceeds in three steps. First, I conduct exploratory data analysis to characterize nonprofits' program changes, financial shifts, and service distributions. Second, I examine statistical relationships among key variables using appropriate statistical tests, focusing on associations between organizational actions and financial conditions. Finally, I provide interpretive reasoning to explain observed patterns and assess how programmatic decisions may influence the scale of services provided.

It is important to note that the tables and examples reported here represent a subset of the full analysis; additional models, robustness checks, and extended results are available in the project's GitHub repository [1].

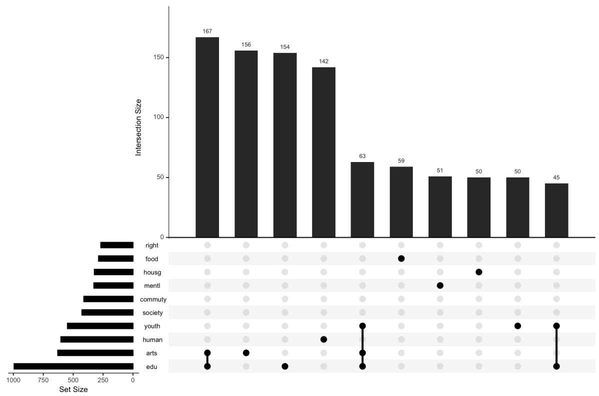

Figure 1: Category Combination

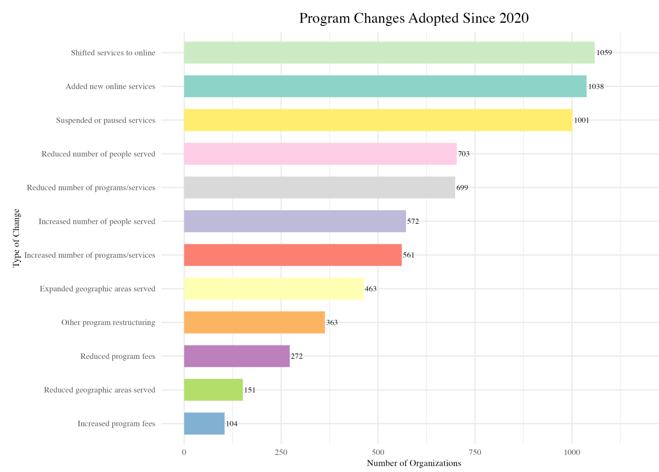

Figure 2: Program Change Actions

Exploratory Analysis

This section provides an exploratory overview of the key variables used in the subsequent analysis. The variables selected here may differ slightly from the original dataset.

To begin, Figure 1 summarizes the distribution of organizational focus areas. The most common categories are *arts*, *education*, and the combined category *arts + education*. These groups are used in later analysis because they contain the largest number of complete observations.

Next, I examine changes in organizational programming during the pandemic (Figure 2). The majority of program adjustments involve shifting services online, adding new online services, or pausing in-person activities. These actions reflect common protective strategies adopted in response to COVID constraints.

Financial changes (Finances_) are measured on a five-level scale, where 5 indicates increased revenue and 1 indicates decreased revenue. Figure 3 shows distribution of these changes. A clear downward shift in revenue is visible, suggesting substantial financial strain during the pandemic. To confirm this pattern, I conducted a McNemar's Chi-squared Test, which indicates a statistically significant negatively difference between pre- and post-COVID financial conditions.

After COVID, declines appear consistently across all types of donation sources (Figure 4), reinforcing the conclusion that COVID-19 has negatively affected nonprofit revenue streams.

Figures 5 and 6 further describe geographic and demographic distributions within target group. These descriptive patterns help contextualize heterogeneity across orga-

[1] `https://github.com/OrangeLyx/stats506`

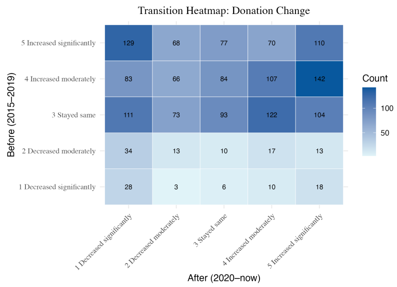

Figure 3: Donation Change

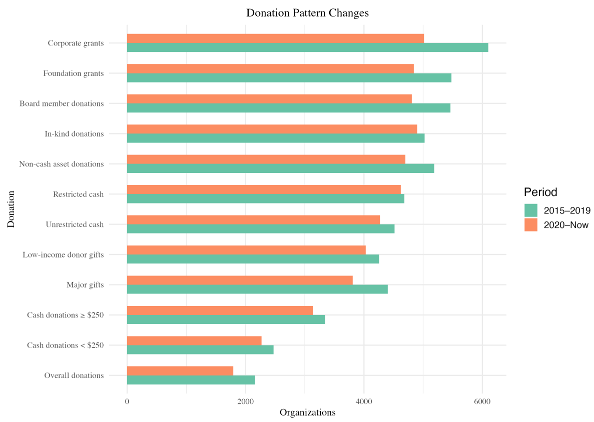

Figure 4: Finance Donation Comparison

nizations.

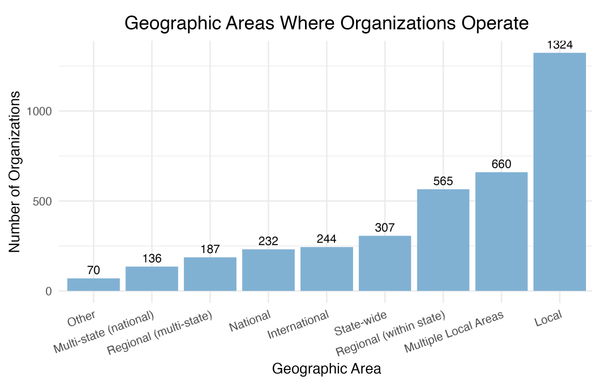

Figure 5: Geographic Distribution

Method

To investigate how program changes and financial conditions evolved across nonprofit organizations, I employ a combination of descriptive inference and causal modeling. Our methods are selected to handle the categorical nature of predictors, the panel structure of the data (2019–2021), and the high dimensionality of financial variables. Specifically, I use Chi-square tests for associations among categorical variables, Difference-in-Differences (DID) for causal comparisons before and after the pandemic shock,

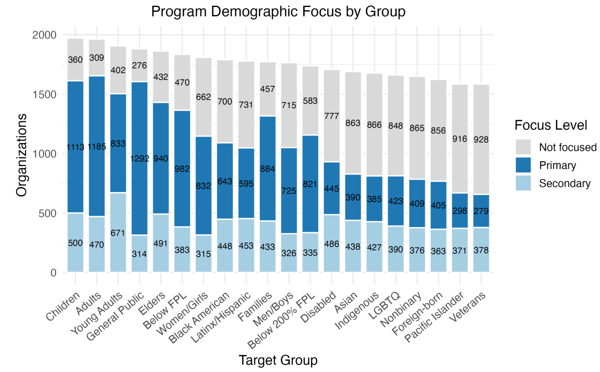

Figure 6: Demographic Distribution

and LASSO regression for variable selection within a high-dimensional financial outcome model.

Chi-square Tests

Many variables in the survey are categorical, including binary indicators and multi-level categorical measures. I use Chi-square tests to examine whether changes in specific donation categories are significantly associated with programmatic changes (e.g., expansion, contraction, service suspension).

The Chi-square test evaluates the null hypothesis:

$$H_0 : \text{Program changes are independent of donation changes.}$$

Using chi-squared tests, I find clear bidirectional associations between program changes and post-pandemic Cash donations (from individuals) below $250 shifts.

In the arts group, this mirrored pattern is strong: action of increasing the number of programs or services is significantly associated with revenue increases, whereas reducing programs or services is significantly associated with revenue declines in the same category. A similar pattern appears for increasing vs. reducing the number of people served.

Overall, expansive actions correspond to higher revenues and contractive actions to lower revenues, indicating a coherent and non-random relationship.

Difference-in-Differences

Since dataset contains repeated annual measures from 2019, 2020, and 2021, I apply a Difference-in-Differences model to estimate how financial outcomes changed after the COVID-19 shock and whether organizations with different program strategies experienced different financial trajectories.

Let Y_{it} denote the financial outcome (e.g., total donations from all sources on each year) for organization i at time t, and let Treat_i indicate whether the organization implemented a particular program change. The DID specification is:

$$Y_{it} = \alpha + \beta \cdot \text{Treat}_i + \gamma \cdot \text{Post}_t + \delta \cdot (\text{Treat}_i \times \text{Post}_t) + \varepsilon_{it}.$$

The coefficient of interest is:

$$\delta = \text{DID estimate},$$

which captures how financial outcomes diverged between treated and untreated groups after 2019.

Given the large number of financial variables and program-change indicators, conducting a full Difference-in-Differences analysis for every combination is impractical. Therefore, I select the source from individual donations and apply the DID specification to illustrate. This example provides a clear demonstration of how program changes actions interact with financial revenues over time.

According to the DID results, most program changes do not exhibit statistically significant effects on this funding stream in either 2020 or 2021. The estimated interaction terms are generally small and imprecise, indicating that operational adjustments—such as expanding, reducing, or shifting programs—did not meaningfully alter financial outcomes for this example. Only reducing the geographic areas served shows a substantial negative effect, with funding decreasing by approximately \$338,000 in 2020 and \$361,000 in 2021, suggesting a substantial decline in funding associated with this particular program action ($p < 0.01$) .

LASSO Regression

I use LASSO regression due to its ability to perform variable selection in high-dimensional settings. For the LASSO regression, the outcome variable is constructed as the total change in financial resources during the post-COVID period.

The predictor matrix X includes:

GeoAreas_*: geographic scope of service delivery.

ProgDem_*: demographic groups targeted by the organization.

FinanceChanges_*: financial changes during the pandemic.

Funding_*: composition of funding sources.

FRChanges_*: pre-pandemic revenue changes (2015–2019) across donor types and formats.

ProgChanges_*: programmatic adjustments made since March 2020, including expansions, reductions, suspensions, and shifts to online service delivery.

Reserves_1_*: indicators of financial reserves and liquidity capacity.

The response variable is:

$$Y = \text{Total finance from all sources.}$$

Given the large number of financial indicators, LASSO solves:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\},$$

where the tuning parameter λ controls sparsity and forces small coefficients toward zero.

LASSO provides a concise set of variables that best explain differences in nonprofits' financial resilience during and after the pandemic.

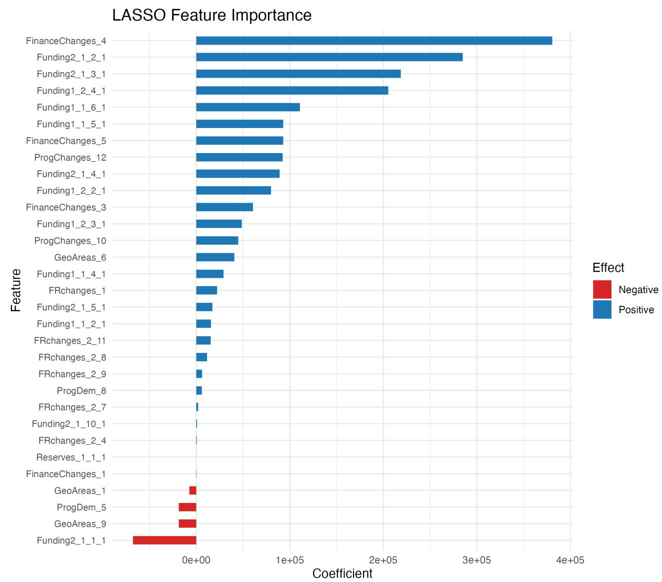

Figure 7: LASSO Feature Importance

From Figure 7, it shows financial shifts during the pandemic and the composition of funding sources are the strongest predictors of overall revenue outcomes. Program changes contribute modestly, while geographic and demographic variables have relatively minor effects. A few funding-related predictors show negative coefficients, indicating potential vulnerabilities in specific revenue streams.

Conclusion

This analysis shows that nonprofit program actions, funding structures, and service scope are closely interconnected. Chi-square tests reveal clear bidirectional associations between program adjustments and changes in small individual donations, with program expansions linked to revenue increases and contractions linked to declines. The DID example indicates that most program actions did not significantly alter funding trajectories, except for reductions in geographic service areas, which led to substantial revenue losses. Finally, the LASSO results highlight that pandemic-related financial changes and underlying funding structures are the strongest predictors of overall financial outcomes, whereas program, geographic, and demographic factors play comparatively minor roles. Together with more interesting data result from code repository, these findings suggest that financial resilience during COVID-19 was shaped more by funding composition and financial shocks than by shifts in service activities.