

BUAD 312: Statistics and Data Science
Final Project: Claire's Place Foundation

Executive Summary

*George Danzelaud, Mimi Adams, Sarang Ashwin,
Anthony Nasser, Mahum Syed, Jainee Shah*

Part (0) Introduction

The broad goal of this report is to aid Claire's Place in understanding their applicants and implement data-driven policy changes that maximize impact and efficiency. Our data-driven analysis assessed the following three core questions:

1. Based on the EHS (Extended Hospital Stay Grant) data provided, can you investigate the impact of our most recent policy change beginning in June 2023?
2. For our updated Work Proudly Program, can you please provide suggestions on data to collect and potential metrics to apply in order to evaluate the program's impact?
3. Recent trends in the philanthropy sector show a national decline. Given this context, can you analyze our organization's donation data to determine if we are experiencing a similar trend?

Although our approach for each of the designated questions are unique, all are multi-faceted in the way that they incorporate a blend of data visualization, statistical modeling, and strategic analysis based on data collected from the past to evaluate trends and tailor recommendations.

Part (1) Business Description and Data Summary

Question 1: EHS

The Extended Hospital Stay Grant (EHS) program serves families affected by cystic fibrosis by offering financial aid in the form of rent relief. This foundation underwent a policy alteration in June 2023. The change involved capping grants at one month's rent or mortgage, with a goal of widening the distribution of funds in order to reach a higher volume of grantees.

The data set covers each grant application from February 2020 to November 2023, including the date applied, amount granted if any, state of residence, whether the applicant was new or returning, and whether the patient was a minor or adult. The questions we aim to investigate in this report include: Did the policy change lead to more first-time grantees? Did the policy allow the foundation to reach and provide financial assistance to a greater number of people on a per-month basis? Finally, did the policy disproportionately affect grantees based on geographic location?

Question 2: WPP

The Work Proudly Program (WPP) is meant to help individuals affected by cystic fibrosis find employment or transition to a better employment situation. About \$2,000 - \$2,075 has been historically spent on applicants, with only nearly 30% (29/97) applicants completing the program. There was a policy change on 09/19/2023 that required applicants to complete a couple more steps, but since only 10 out of 97 entries (10.4%) in the WPP data set entered the program with the change, we decided to leave out any conclusions founded on whether or not it was effective. We neglected conclusions involving applicant genders because only 37% (36/97) of the data contained gender. We also neglected conclusions related to ethnicity because Cystic fibrosis is known to disproportionately affect people of European descent ([source](#)), which is reflected in the dataset as 52% (50/97) applicants reported to be Caucasian/White.

However, the provided dataset can help us come up with ideas for further helping applicants and what kind of data the program should collect in the future in order to better understand what is working.

Question 3: Donations

We analyzed donation data from Claire's Place from 2019-2023 to assess whether the organization is experiencing the same downward trend in philanthropy seen nationally. In terms of our project goals, we first looked into Annual Donation Trends (is there a year-over-year decline in the total donation amount, average donation amount, and total number of donations received by the foundation? Next, we focused on monthly analysis: do certain months of each year consistently show a higher number of donations or higher donation amounts on average? Third, we measured impactful campaigns: which campaigns have been most successful in attracting donations? And finally, we examined regional analysis: are certain regions contributing more significantly than others, and is there a noticeable change over time?

During our preliminary analysis, we observed that the donation dataset from the Foundation includes information on all 9,502 donations made to Claire's Place from August 2019 through November 2023. For each donation, the dataset includes the date, location, amount, associated campaign, completion status, and details for dedications or recurring amounts. First, we split the complete Donations dataset by region (US and Non-US donations). Next, we filtered the US Donations dataset into Complete and Incomplete Subsets, since only Complete Donations benefit Claire's Place. We found that 88.4% of US Donations were complete, with a mean donation amount of \$129.15 (nearly \$30 above the mean for all regions donations) and median of \$27.15. Based on the IQR Method for identifying outliers, we found an upper bound for outliers of \$184.50 donations and above. However, these are not typical outliers that we should exclude, because every donation contributes positively to the Foundation and should be considered when evaluating trends.

Part (2) Data Visualization

Question 1: EHSG

We consider the start date of the EHSG policy change to be June 1, 2023, and we have created the variable "Period" to signify whether a grant was given before or after this change. The foundation is not reaching a greater number of patients on a per-month basis. The average number of grantees per month decreased from about 6.72 to 5.5 (figure 4). Our simulation-based hypothesis test confirms this decrease, and didn't show significant difference in projected averages.

The graphs depicting grant frequency before and after the policy change were enhanced for readability by categorizing the 50 states into five regions: Atlantic, Central, Mountain, New England, and Pacific. Pacific and Atlantic states generally had more donations than other regions. Analyzing grant data specifically from 2022 to 2023 offered insight into the immediate impacts of the policy shift amid significant macro environmental changes [Appendix 8]. Initially, the top three states in terms of grants were CA, CO, and OH respectively. However, post-policy change, the most frequent grant recipients changed: instead, Southern states like FL, VA, and GA consistently emerged as recipients of a substantial number of grants. CA dropped the most drastically. Notably, CA no longer held the position of the most frequent grant-receiving state after the policy change highlights a shift in grant distribution patterns. When accounting for the all time data, CA, FL, and OH held the record for the most amount of grants with GA in a close 4th place.

Question 2: WPP

First, we created visualizations to analyze the applicant pool for the WPP program. We found that most applicants were low income, especially among older adults. We attempted to look at gender distributions, and although there's more female applicants than male applicants, there was a lot of missing data. We also looked at social services and found that most young middle aged adults use social services, and that they were more common in unemployed applicants. This is important because 60% of applicants were unemployed. Next, we analyzed how the data changed over time over different factors. Doing this allowed us to find that the number of applicants decreased over time and further differences when grouping by gender. As for completion rates, we visualized rates by gender, ethnicity, age, employment, and social services (including each individual service), and saw a trend where having a job, being of older age, or being low income were barriers to completing the program [Appendix 1]. From this vein, we also tried to analyze the impact of the policy change, but it was difficult to claim anything significant since there were only ten applicants after the policy change.

Question 3: Donations

We created a number of graphs to visualize the donation data based on our areas of focus for the project goals. First, when analyzing Annual Donation Trends, we graphed Total Donation Amount by Year, Total Number of Donations by Year, and Average Donation Amount by Year. These graphs revealed a spike from 2020 to 2021 in Total Donation Amount and Average Donation Amount, but a steady decline in Total Donations, Total Number of Donations, and Average Donations from 2021 to 2023. However, we took into account that the data was only measured from August 2019 to November 2023, so we calculated averages in each of our three metrics for the month of December from 2019 to 2022 to predict the final three graphs for 2023. These predicted charts still showed a decline in Total Donation Amount and Total Number of Donations, however our predicted Average Donation amount for 2023 exceeded any of the other years, which suggests greater donor generosity on average. For Monthly Analysis, we visualized Total Amount of Donations by Month for each year, as well as Number of Donations by Month for each year. This showed a high dollar amount and frequency of donations from the later months of each year, especially August through December, correlating with impactful campaigns like Glow Ride in mid August. Next, we focused on top campaigns and regions, which helped identify areas of fundraising success. Finally, we categorized donations into low (\$0-\$50), medium (\$50-\$200), and high (above \$200) levels, alongside our analysis of recurring donations, which offered insights into donor behaviors. Overall, these visualizations were key in understanding donation dynamics over time, measuring successful initiatives, and served as a basis for our modeling and strategic recommendations for future fundraising efforts.

Part (3) Modeling

Question 1: EHSg

First, we built a simulation-based bootstrap distribution in order to investigate if there was a statistically significant increase in the proportion of new patients, from the period before the policy to the period after. We can conclude with 95% confidence that the difference in proportions is within the range of about 2 to 37% of an increase, as you can see in figure 5. This supports the claim that a higher proportion of new patients are indeed more likely to be

granted financial assistance if it is after the introduction of the policy change. An additional simulation-based hypothesis test confirms this conclusion, in which we found a significant p-value of 0.024.

Next, we built a confidence interval illustrating the distribution of the disparities in average amount granted between new grantees and returning grantees before the policy change. We decided to build this model before the policy change as the goal of the policy change was to distribute funds more evenly. The lower bound of the CI was -354 while the upper bound was 168. The plot was a left-skewed distribution which suggests a potential scenario where the average amount granted for returning grantees could be lower than that for new grantees. This implies that new grantees might require a larger amount of financial assistance compared to returning grantees.

Finally, we built a simulation-based bootstrap distribution in order to investigate what the mean number of grants a state could expect to receive before the policy. We can conclude with 95% confidence that states could expect between 4.47 and 9.74 grants before the policy change [Appendix 7]. This helps us understand the variation between states and the amount of grants received. We compared our bootstrap distribution after the policy where the lower bound was 1.26 and the upper bound was 2.16 [Appendix 9]. This indicates a relatively more even distribution between states after the policy.

Question 2: WPP

For the analysis of the WPP data, we conducted a bootstrap analysis to measure whether or not there is a statistically significant difference between employment rates in adults (aged 26-35 or 36-45) who use social services and those who don't. Since the confidence interval (ran with 1000 bootstrap samples) includes zero, it suggests that there is not a significant difference between the groups. This test was also run for the subset of low-income adults only (specified by \$0-25,000 range), and a similar result was found.

However, we did see the visualization of the bootstrap distribution get more narrow, suggesting less variability in the estimates from the sample. Due to this, we decided to investigate with a hypothesis to be absolutely sure there was no significant difference between employment between applicants that use social services and applicants who don't, for both low income adults and the broader applicant pool. In both cases, we found p-values over 0.05 that allowed us to safely fail to reject the null hypothesis that applicants are equally likely to have a job independent of whether or not they use social services.

We also visualized withdrawal rates by gender, and noticed that withdrawal rates seemed higher for females, so we performed a hypothesis test to test this. The hypothesis test yielded a p-value of about 6%, which although not statistically significant, may prove to be helpful because it is within 90% confidence.

To produce a linear model that could aid in predictions, we tried many different combinations of columns. In order to reduce a bias in the outcome measurement, we removed the applicants with 'Withdrawn' status and categorized the remaining by number (1: New, 2: In Progress, 3: Completed). With this new column, we created an interaction model to test against applicant age and their range of annual household income. With this model, there is a p-value of ~0.007 for applicants age 55+, which may indicate that age and annual household income are a reliable predictor for this group. However, it should be noted that since this group only takes up 2% of the dataset, this is likely not a safe conclusion.

Question 3: Donations

When identifying modeling opportunities for Claire's Place Foundation's donation data, we first attempted regression analysis, however the dataset lacks a significant correlation between any two numeric variables. We focused on the trend of donation amounts changing over time to address our core question, yet the correlation between "Amount" and "Year" is only 0.03. Hence, we determined that Regression Testing is not an effective approach to answer this question with the given dataset, instead focusing on bootstrapping and hypothesis testing.

We used bootstrapping techniques to construct confidence intervals and conduct hypothesis testing on the mean donation amounts across different years. Specifically, the bootstrapping method was used to estimate the difference in means between the years 2021-2022 and 2020-2021, resulting in 95% confidence intervals using the Percentile Method. Our null hypotheses were that the mean donation amounts were equal for each pair of years. When analyzing 2021-2022, our alternative hypothesis was that 2021's mean donation exceeded that of 2022, which we gathered from our observed statistic. Similarly for 2020-2021, we observed for our alternate hypothesis that the mean donation of 2020 was less than that of 2021.

For the years 2021-2022, the confidence interval crossed zero (Lower CI = -76.6 and Upper CI = 72.7), indicating no significant difference in average donation amounts between the two years. The first hypothesis test confirmed this with a p-value of 0.451, leading to a failure to reject the null hypothesis. Conversely, our analysis for 2020-2021 showed a confidence interval that did not include zero (Lower CI = -104 and Upper CI = -6.66), and the p-value of 0.016 suggested rejecting the null hypothesis in favor of the alternative, indicating a meaningful increase in average donation amounts from 2020 to 2021. This suggests that Claire's Place experienced growth in donor generosity from 2020-2021, which did not fluctuate significantly in 2022. The confidence intervals and hypothesis tests together painted a picture of growth in average donation for the foundation from 2020-2021 and relative stability for 2021-2022, suggesting resilience in their donor base despite the national downturn in philanthropy. This is consistent with our prediction that the Average Donation Amount in 2023 (\$153.75) will be the highest the foundation has seen in the past five years.

Part (4) Final Recommendations

Question 1: EHSB

Based on our analysis of the EHSB policy change in June 2023, we recommend that the foundation keep this policy change in place because of the predicted higher proportion of new patients being financially assisted. Due to the fact that the policy was implemented relatively recently, and that our statistical findings and analysis are based on predictions, we recommend the foundation should continue collecting data and reassess the impact of the policy in the future. Additional data that would assist in investigating this policy's impact includes more qualitative information about the impact of one-month financial aid; for instance, what proportion of patients who were given one month of rent were able to get back to a state of financial stability? Furthermore, information about health insurance can help in determining which patients are in dire need of assistance, especially if those patients are underinsured. Finally, we recommend the foundation increase outreach efforts to re-engage patient populations in Central, Mountain, and Pacific regions of the US.

Question 2: WPP

Based on the analysis of the WPP data set, we recommend that gender and ethnicity be mandatory fields, as empty data fields make it difficult to draw conclusive insights. Furthermore, due to the difficulty in measuring applicant success (especially for applicants who are already employed, which take up about 40% of the current dataset), we recommend collecting data on income increases on a recurring basis as the foundational metric for success in this group. We also recommend that for all applicants, there is a “before” and “after” metric associated with them that reflect their intentions and goals joining the program, and whether they were able to achieve them after a designated amount of time. Lastly, due to the high level of applicants who are not using social services across all age groups and income levels (appdx. 1), Claire’s should encourage (especially) those who use 0-2 social services to apply for more social services.

Question 3: Donations

Our bootstrapping and hypothesis tests suggested growth followed by stability in Claire’s Place’s Average Donation Amount from 2020-2022. Unfortunately, through our visualizations for Total Annual Donation Amounts, Total Number of Annual Donations, and Average Donation Amount—adjusted with predictions for December 2023—we determined that the foundation is indeed facing a decline in Total Donation Amounts and Number of Donations from 2021 to 2023, consistent with the national downturn in philanthropy. However, we also predict that the Average Donation Amount in 2023 will surpass that of any previous year, which leads us to our business recommendations.

First, we recommend enhancing seasonal campaigns. Given the noticeable peaks in donations during certain months, particularly around major events like the Glow Ride (mid August) and Clairity Ball (mid October), intensify marketing and engagement efforts during these periods to maximize fundraising. Next, we would target high-dollar donors. With a predicted increase in average donation amounts despite fewer donations, focus on strategies to attract and retain high-dollar donors, such as personalized outreach or exclusive events. Third, we believe in expanding regional outreach. Address regional disparities in donations by targeting regions with lower contribution levels, such as the Northeast, through localized campaigns or partnerships. Fourth, we would diversify monthly fundraising efforts. Investigate the reasons behind lower donations in the first half of the year and develop strategies to boost donations during these months, especially through new events and seasonal campaigns. Finally, we recommend enhancing donor communication by providing regular updates on the impact of their donations to increase donor engagement and retention.

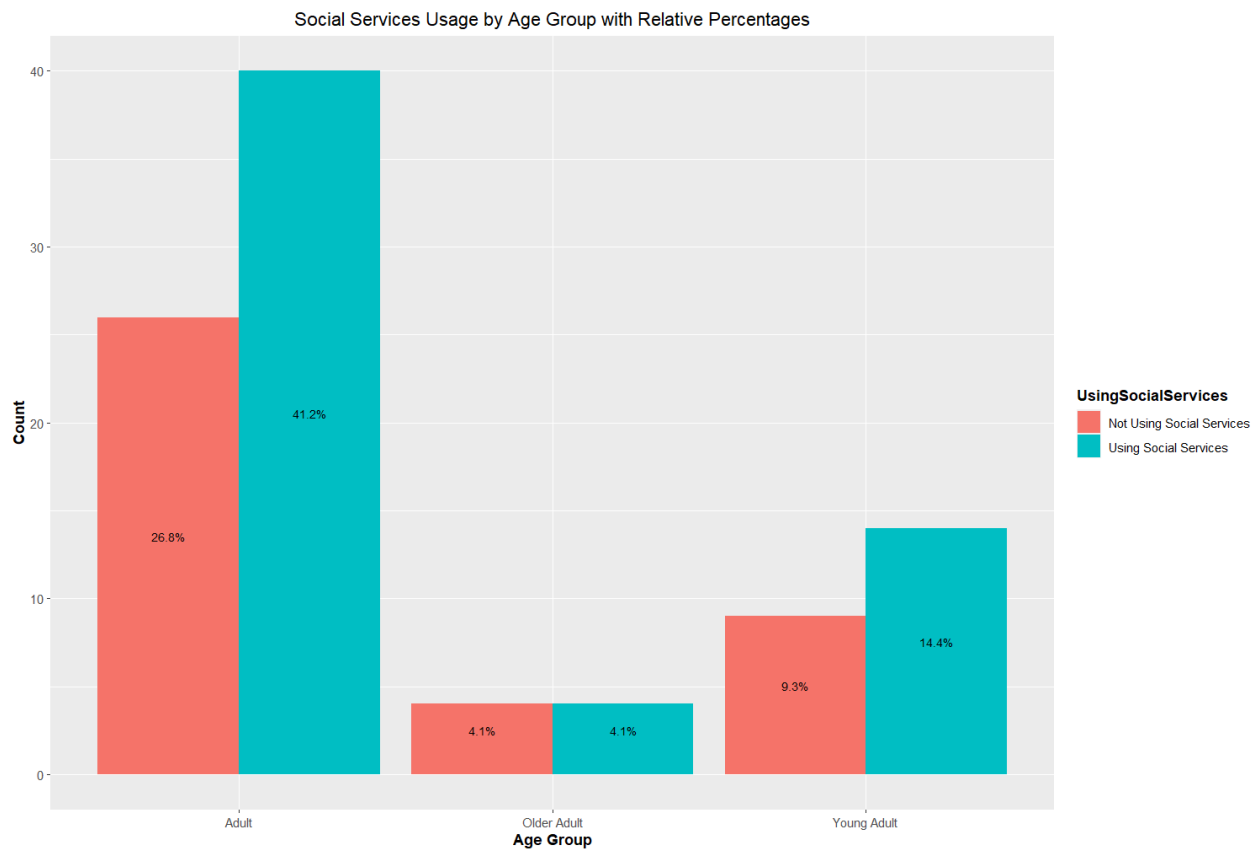
In terms of further considerations, we see potential in conducting donor surveys to understand reasons for discontinuing support; analyzing the effectiveness of various marketing channels to correlate these efforts with donation data; and collecting detailed donor demographics through surveys to launch more targeted and personalized campaigns.

Part (5) Conclusion

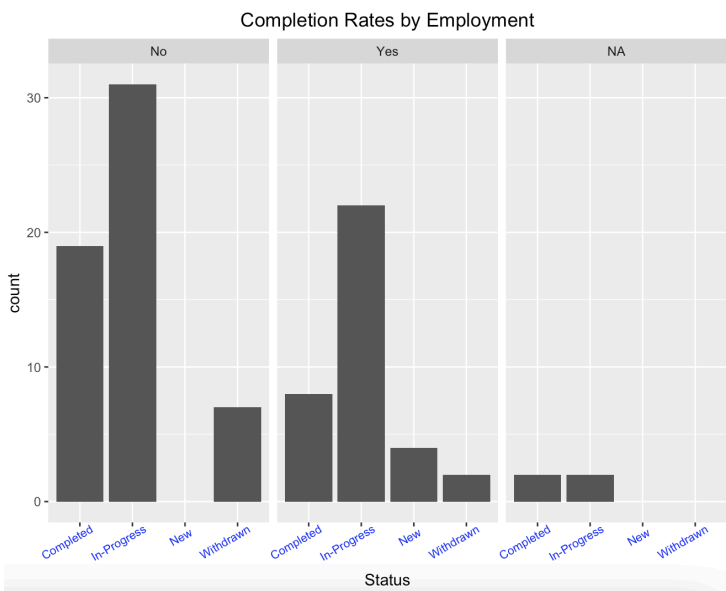
Our evaluation of Claire's Place Foundation's initiatives and donation data has provided a multifaceted view of its operations and impact. The Extended Hospital Stay Grant's policy change, aimed at aiding more families, shows promise in increasing support to new grantees, though its recent implementation calls for ongoing monitoring to fully gauge its effectiveness. The Work Proudly Program, while revealing some barriers to completion, presents opportunities for refined data collection to better tailor support services. Amidst a national philanthropic decline, Claire's Place's donation trends show a similar pattern yet reflect resilience with higher average donations, indicating a strong donor

commitment. These findings underscore the importance of adaptive strategies and robust data analytics to enhance the effectiveness of non-profit programs and maintain vital support for communities, especially during challenging times.

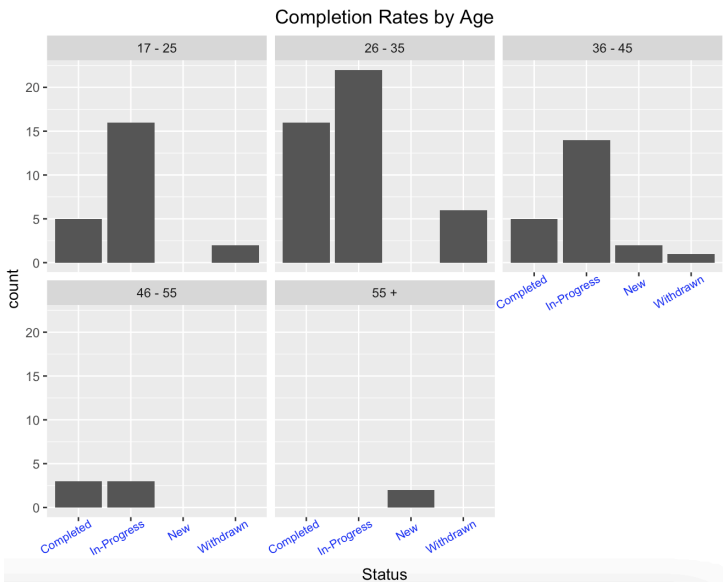
Appendix



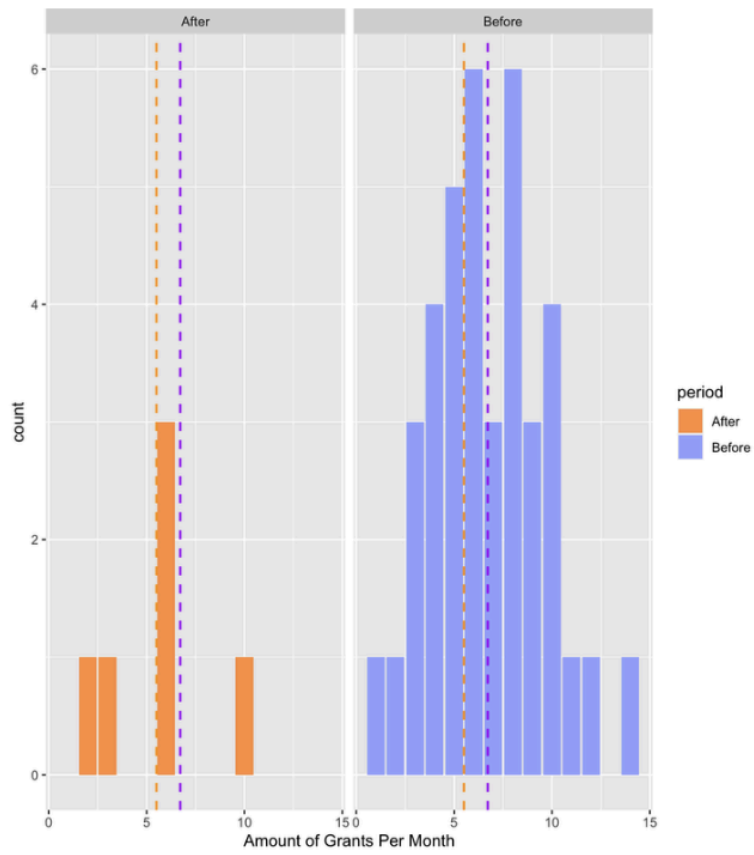
(1)



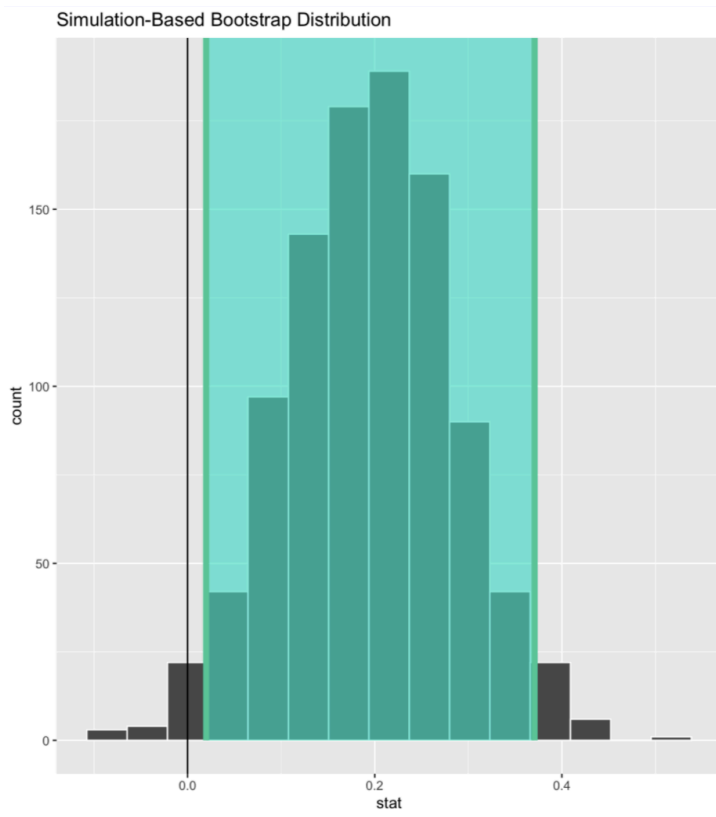
(2)



(3)

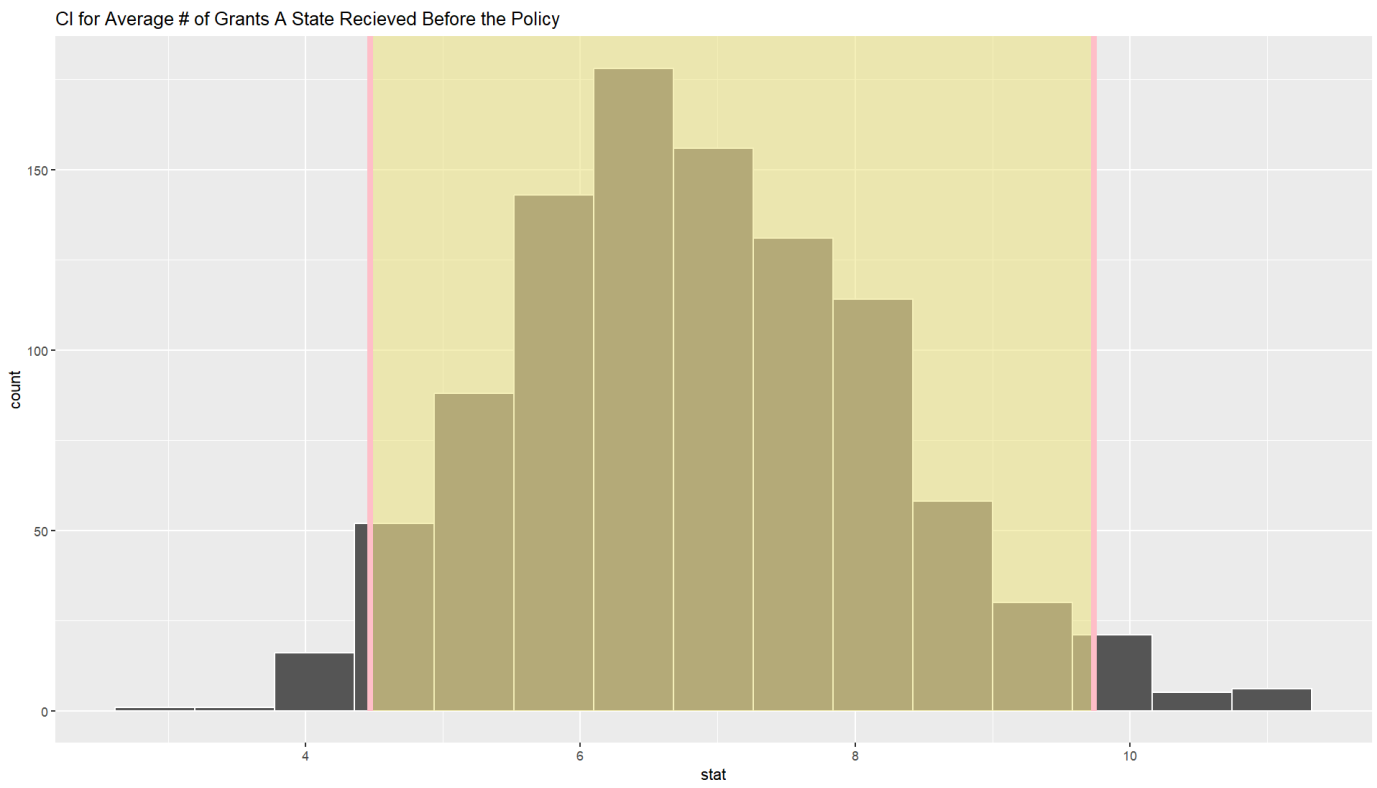


(4)

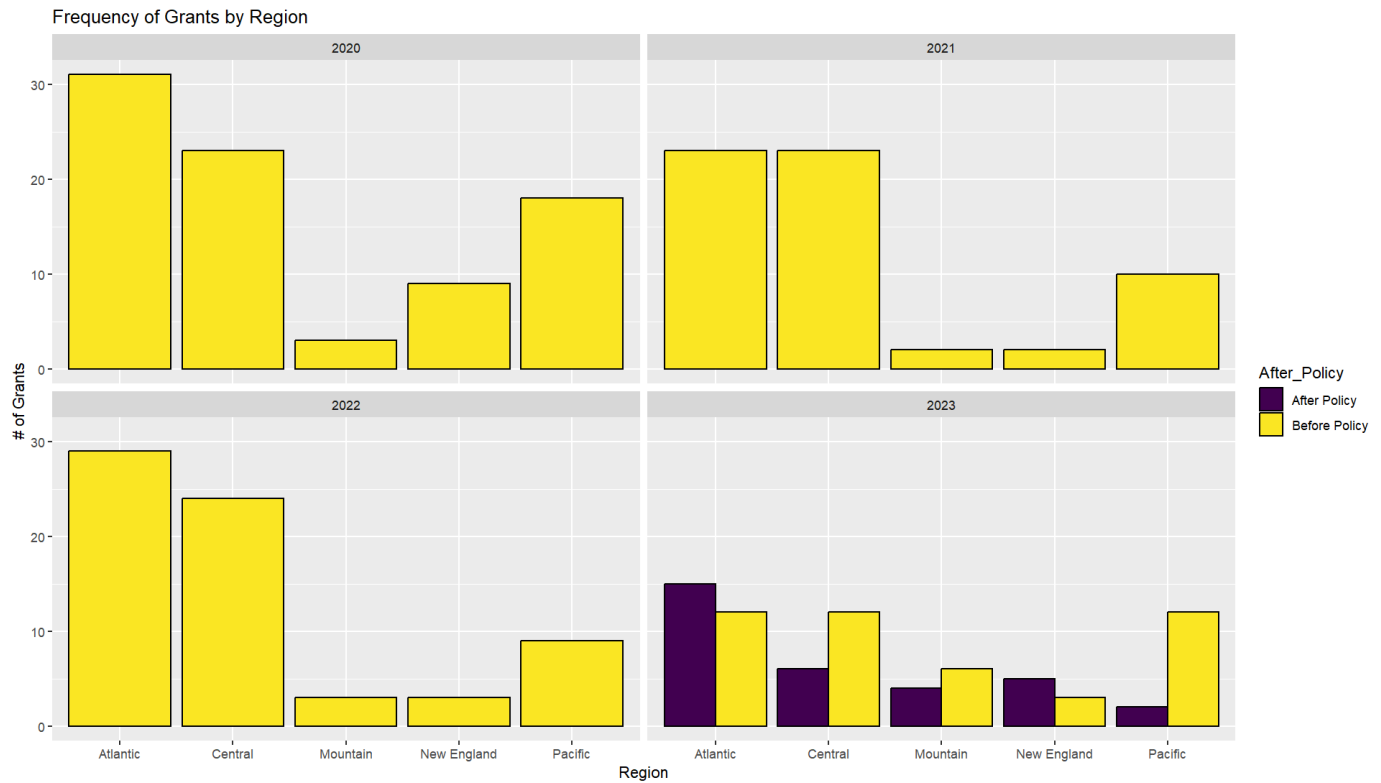


(5)

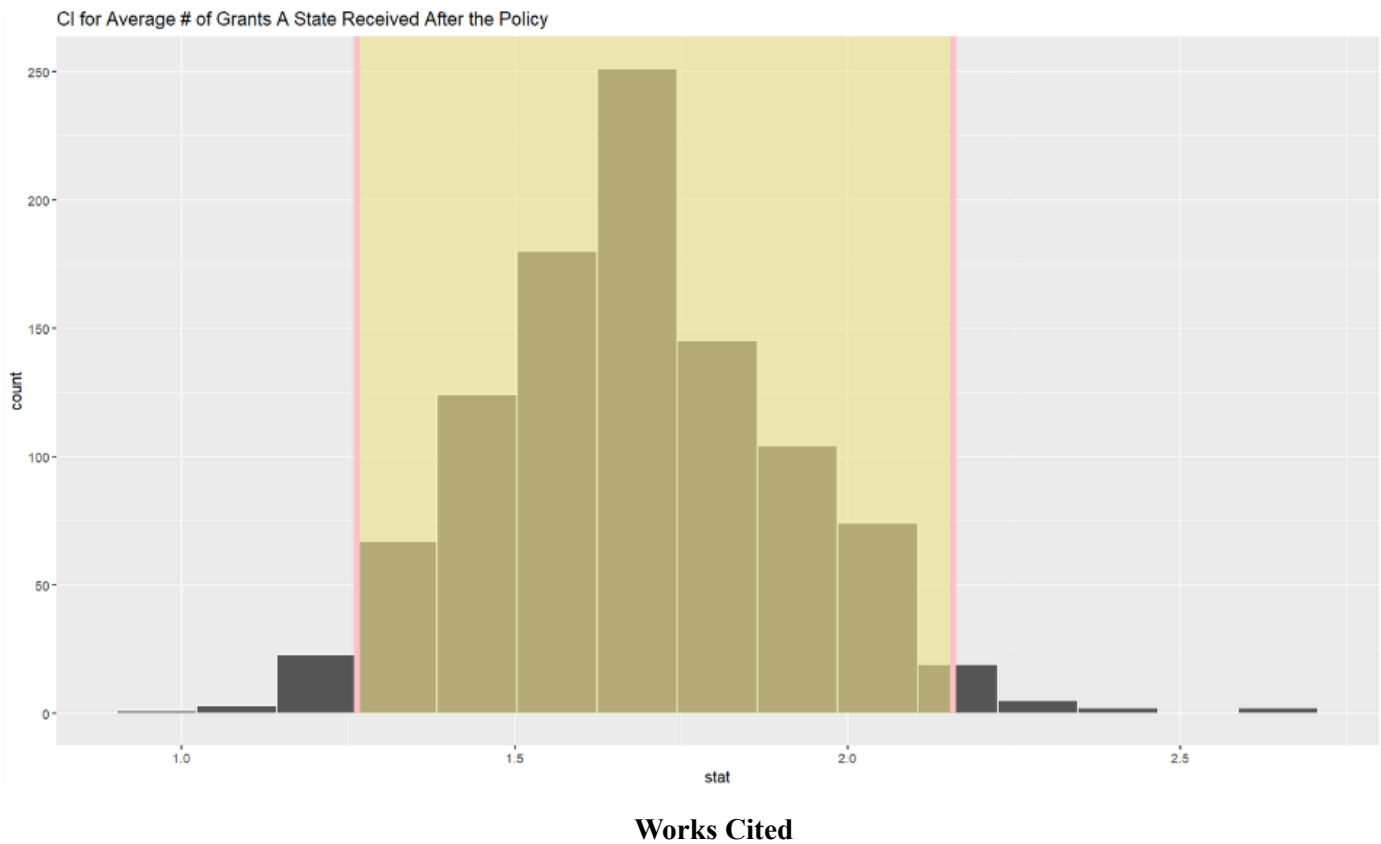
(7)



(8)



(9)



(1) "Cystic Fibrosis - Centers for Disease Control and Prevention." *Cystic Fibrosis*, U.S. Centers for Disease Control and Prevention, Nov. 1995, www.cdc.gov/careerpaths/scienceambassador/documents/cystic-fibrosis-fact-sheet.pdf.