RA Data Exercise Report

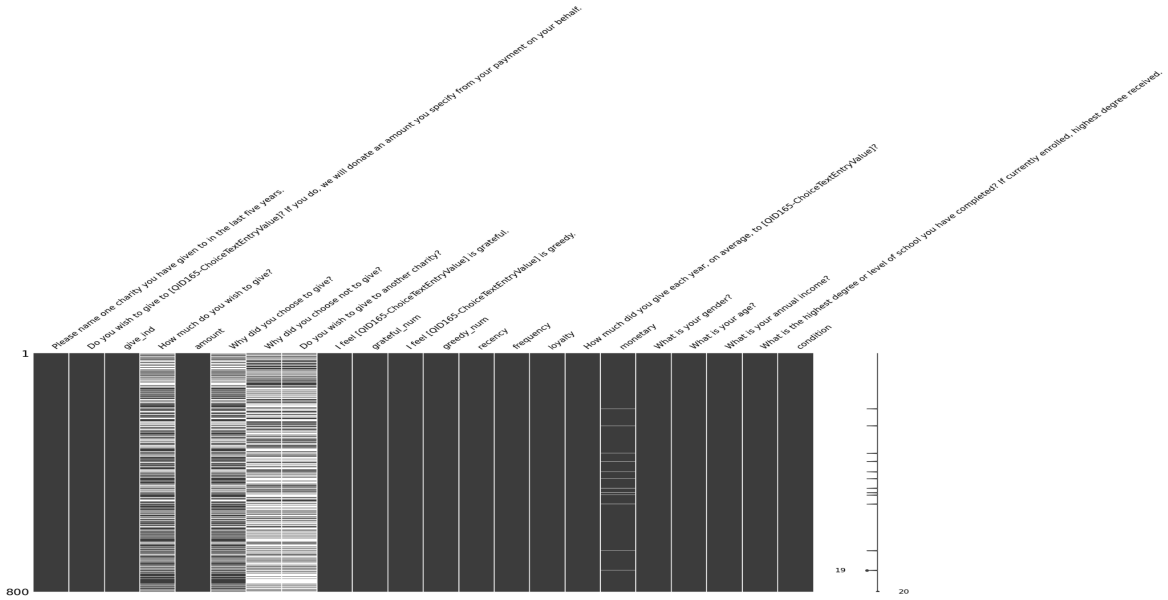
Author: Ruoxuan Li

I. Introduction

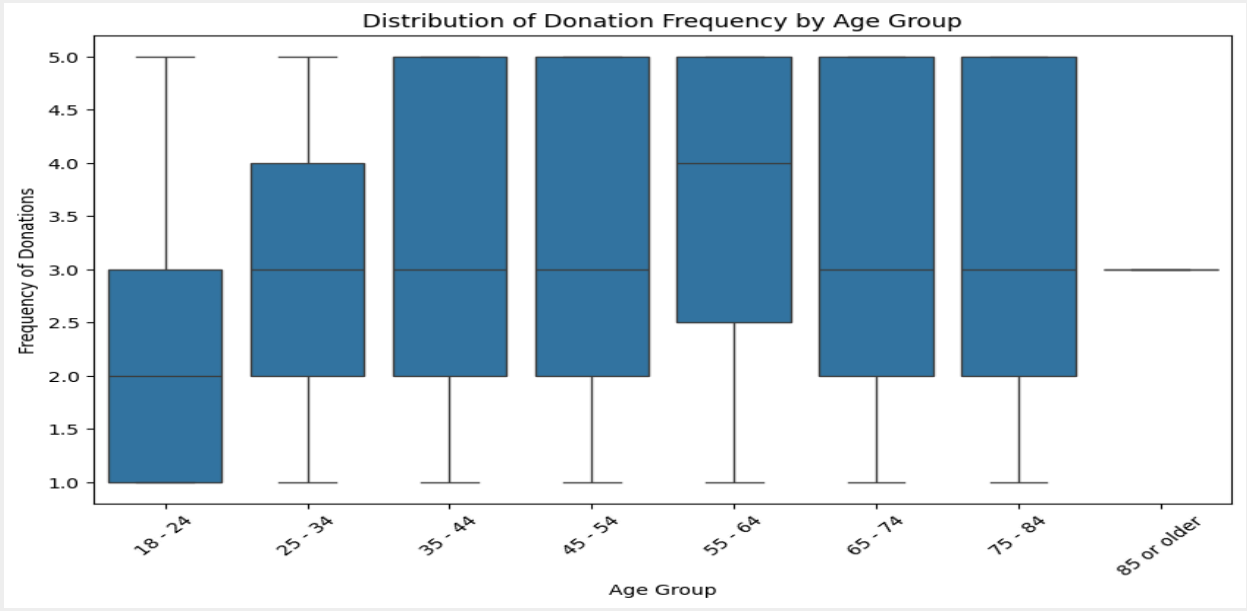
In this project, we are interested in understanding the effectiveness of different messages sent in donation letters on people's donation behavior. Our goal is to explore whether the treatment influences donor behavior and to identify any underlying patterns through detailed visualizations and statistical analysis.

II. Describe the data available in the file.

The dataset comprises **800 observations** across **22 variables**. A preliminary review of the data reveals several key insights that require attention. Notably, the responses to the questions "Why did you choose not to give?" and "Do you wish to give to another charity?" serve as complements to the queries "How much do you wish to give?" and "Why did you choose to give?". Except for that, the column "**monetary**" has **12 missing entries**. It is worth noting when examining the data, a challenge arises from **inconsistencies** in the dataset's encoding of **organizational names**. An example involves the Red Cross, which appears under various guises including "American Red Cross", "Red Cross", "american red cross", "red cross", and "American National Red Cross". This variability in the organization naming suggests encoding issues and underscores the need for data cleaning to unify these references.



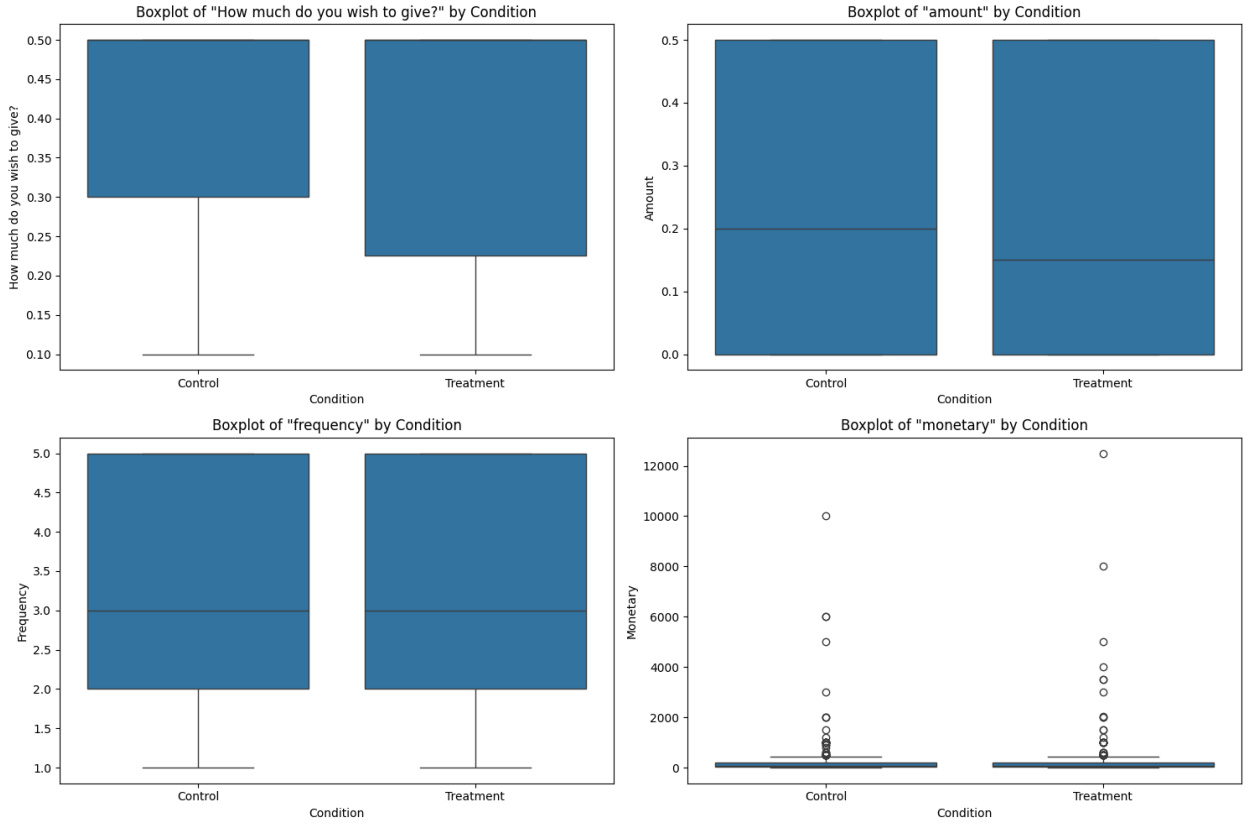
From the boxplots depicting the distribution of donation frequency by age group, we notice that the **55-64 age group** exhibits the **highest frequency of donations** among all age groups, while the **18-24 age group** has the **lowest frequency**. This observation aligns with the expectation that younger individuals, who typically have less stable income, might not donate as frequently. Explaining why the 55-64 age group donates more frequently compared to other older adult groups can be more challenging. One plausible reason could be that individuals in the **55-64 age range are nearing or at the age of retirement**, possibly giving them more disposable income or making them more inclined to contribute to charitable causes as they reflect on their societal impact and legacy.



To examine the motivations behind why individuals choose to donate, I want to use **word clouds** to analyze the frequency of specific terms within the responses to the question 'Why did you choose to give?'. The word clouds showed that "**help**" and "**need**" emerged as central motifs across both control and treatment groups, underscoring a universal interest from both groups to assist and address necessities.

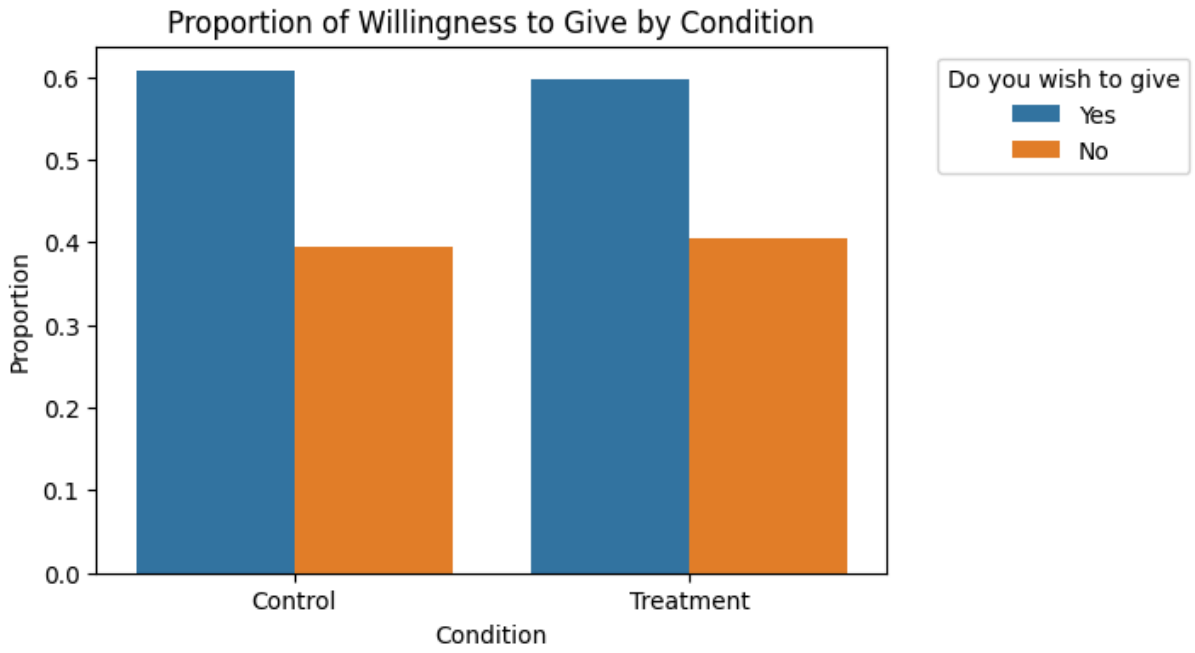
The difference between the word clouds should also be noted. In the control group, we could observe a higher frequency of words like "**believe**" and "**work**", indicating a motivation rooted in personal convictions. However, in the treatment group, the terms "**good**" and "**charity**" were more prevalent, suggesting an emphasis on the positive impact of charity itself.

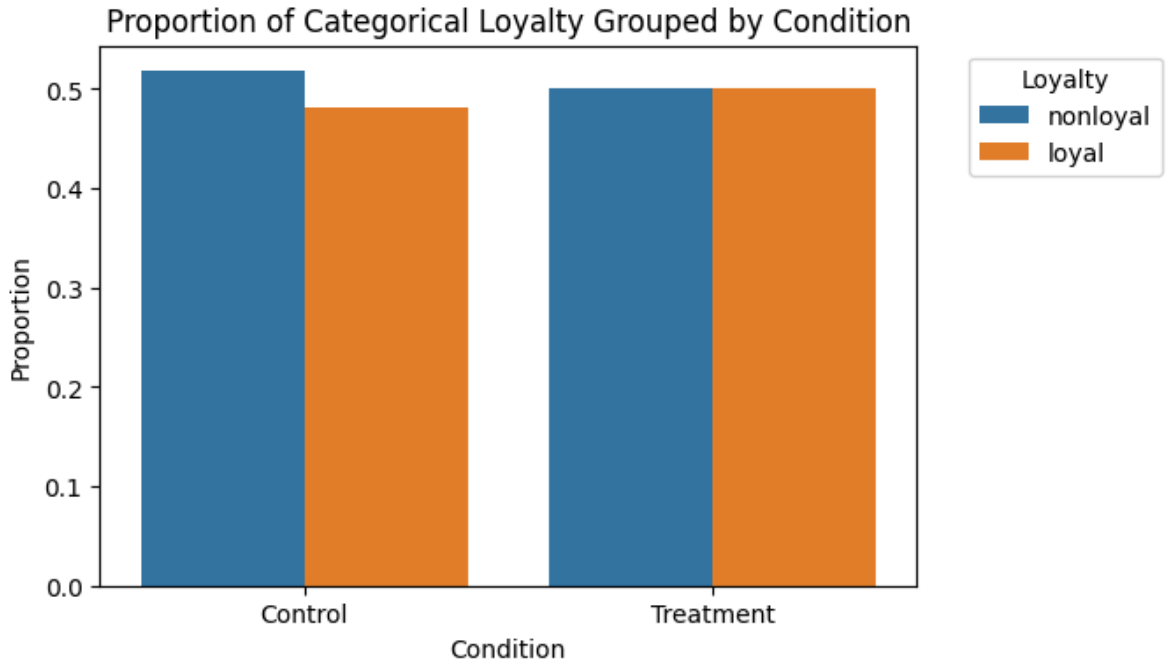
III. Analyze the effect of the experimental treatment on giving (donation) behavior.



Beginning with an analysis of the numerical data within this dataset, it is observed that the median donation amount pledged differs between the two groups. Aside from this difference, the distributions of other variables appear quite **similar** across both groups. This raises the question of whether the variation in donation amounts between the treatment and control groups is statistically significant. Upon conducting t-tests for both the actual amount donated and the amount participants expressed willingness to donate, no significant differences were found between the groups. Consequently, we **fail to reject the null hypothesis, indicating that the treatment does not significantly affect donation behavior.**

By plotting the proportion of **willingness to give** by condition, we can observe that the distribution of the control group and treatment group **resemble** each other. However, the distribution is different when we examine the **loyalty** distribution for both groups. We can observe from the plot below that the treatment group is **slightly more loyal** compared to the control group as more proportion of people claim that they are loyal.

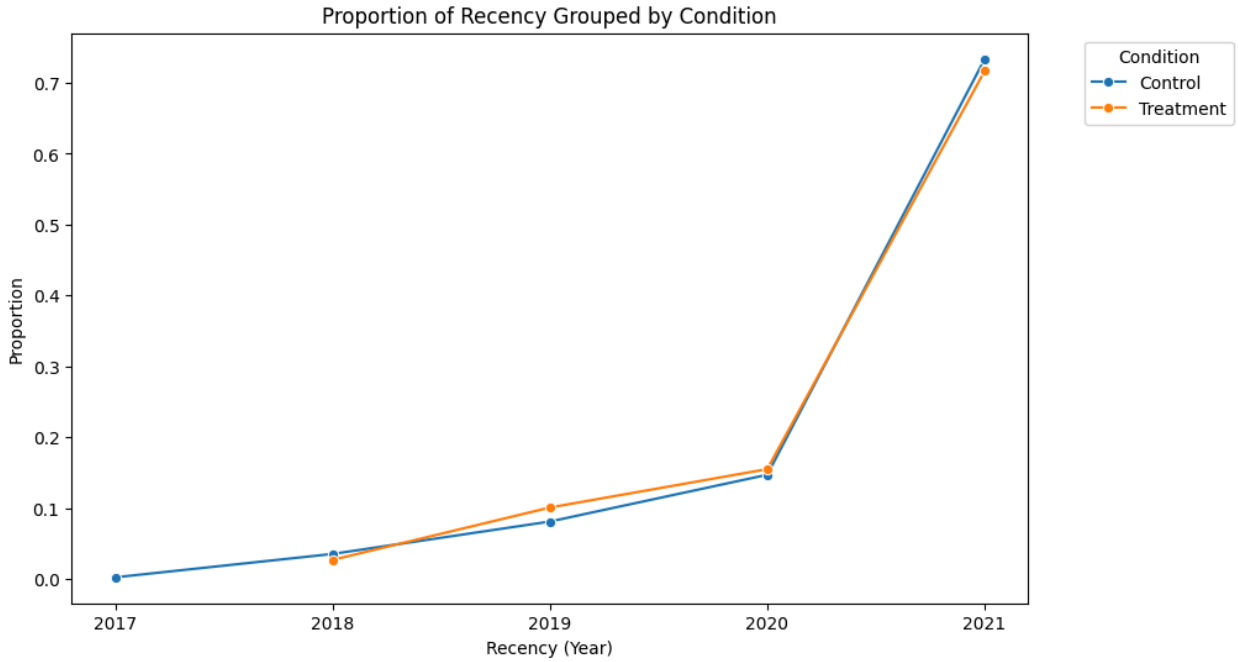




To determine the statistical significance of this difference, I selected an alpha level of 0.05 and conducted a chi-square test on the data. The resulting p-value is 0.66, which leads us to **fail to reject the null hypothesis**. This suggests that the observed difference between the control group and the treatment group could merely be attributed to **random variation**.

Next, we aim to examine another donation indicator - recency, and compare its distribution between the two groups. From the graph, we observe that the control group

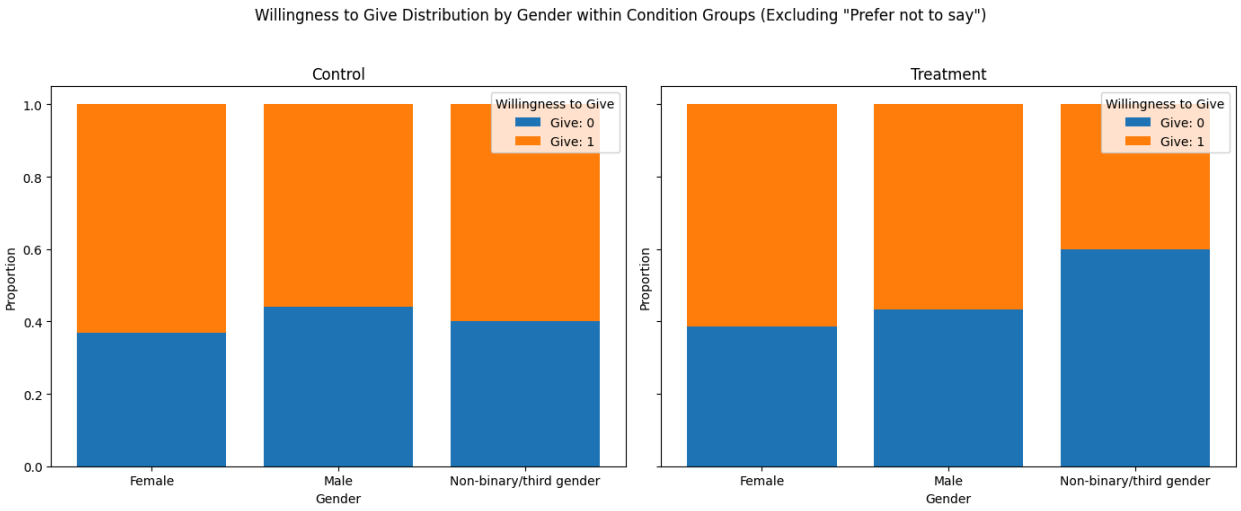
reports charity activity in the year 2017, though only a very small percentage. Starting in 2018, we observed that the proportion of people who reported the year the donor last gave overlaps between the groups. Except in **2019**, where the treatment group has a slightly higher proportion, and in **2021,** where the control group has a marginally higher proportion of people. However, since differences are minor enough, they could all be explained by random variation and **fail to indicate** any real correlation between the treatment and recency of the donation.

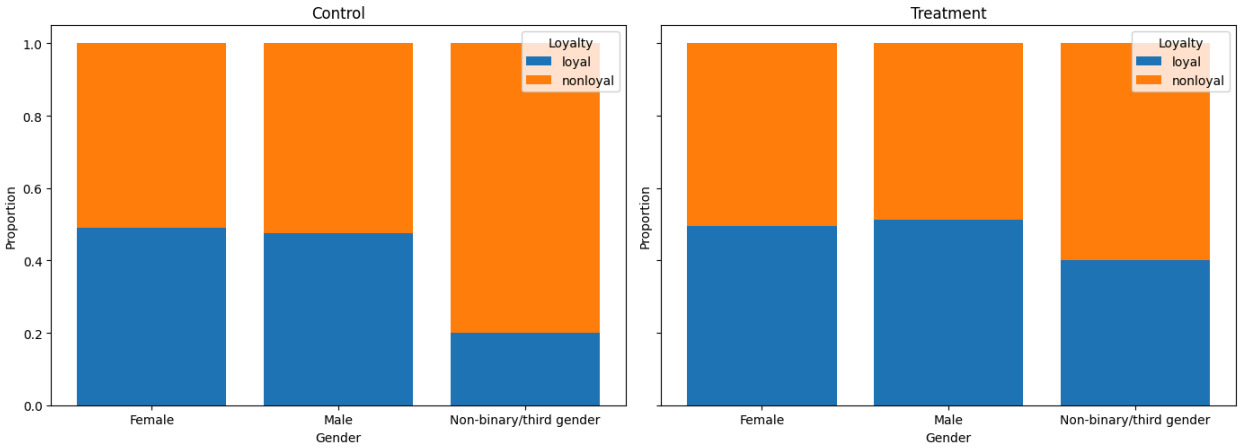


IV. Analyze how the effect differs across different individuals

Now, we aim to explore how the treatment's effect varies among individuals with different demographic characteristics, starting with gender. We are particularly interested in examining how the treatment influences the willingness to give and loyalty across different gender categories. In the control group, some individuals did not report their gender; therefore, for visualization and comparison, we have focused exclusively on the categories of female, male, and non-binary/third gender. From the graphs, it is apparent that the treatment does not significantly differentiate between male and female counterparts in terms of both willingness to give and loyalty. However, for the **non-binary** group, we observe a tendency towards **decreased willingness to give** but **increased loyalty**. It is worth noting that due to the small sample size in the non-binary

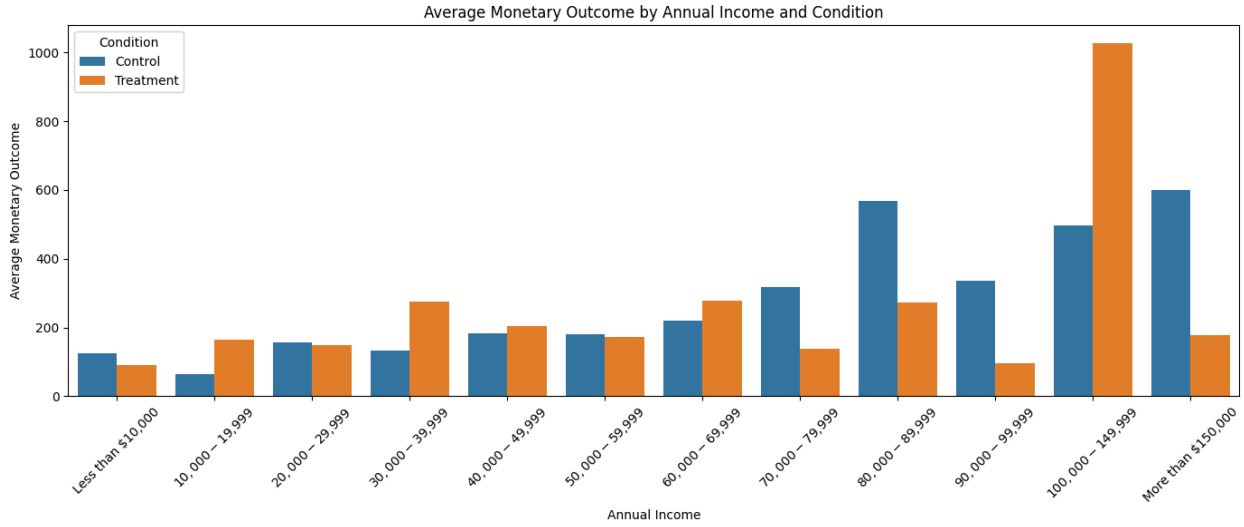
group (5 people from each group), this difference is likely attributable to **systematic variation** rather than an actual effect from the treatment group.



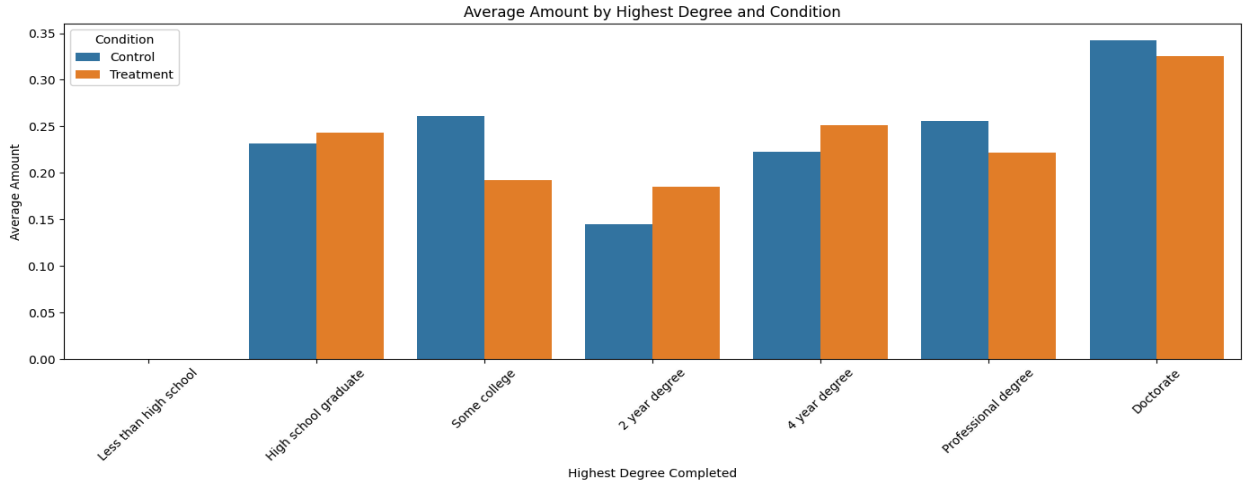


Another interesting pattern we can analyze is the amount donors decided to give each year on average over the last five years, based on their annual income, and how the distribution differs between the control and treatment groups. In the graph, we observe several major differences: for individuals whose annual salary ranges fall within **$30,000-$39,999** and **$100,000-$149,999**, the amount donors decide to give is **higher** in the **treatment** group. Particularly, for the income range of **$100,000-$149,999**, we observe the amount **almost doubled**. For the income ranges of **$70,000-$99,999**, and **more than $150,000**, we observe that the **control group** reports a **higher** amount of

donation they are willing to give.

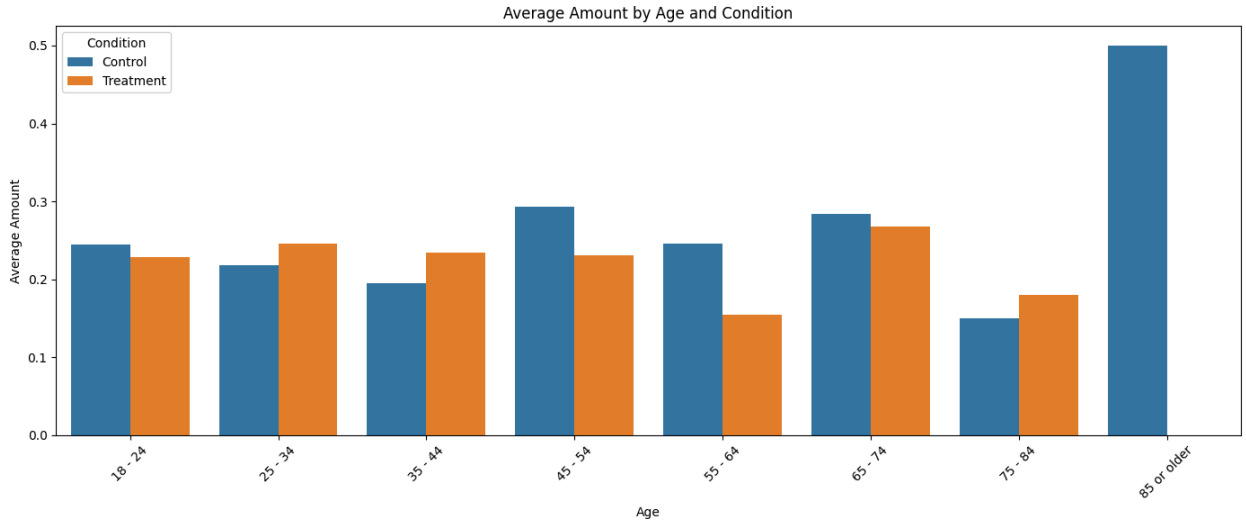


We can also assess how the highest educational degree attained by the donor affects the amount given, specifically focusing on those who decide to give amounts ranging between $0.10 and $0.50, and observing the variation between the control and treatment groups. It has been noted that individuals who have obtained a **high school diploma, 2-year degree, or 4-year degree** tend to **donate more** when exposed to the **treatment**. Conversely, those who have **some college education, a professional degree, or a doctorate** tend to **contribute higher amounts in the control group** compared to the treatment group.



Lastly, we examine the impact of age groups on the amount donated, and how this differs between the treatment and control groups. The plot reveals that, while most age groups exhibit a similar pattern, individuals within the **45-54** and **55-64 age**

brackets tend to donate **lower amounts in the treatment group** compared to the control group.



V. Conclusion

Based on our graphs and statistical analyses, we can reasonably conclude that the **treatment group likely does not have an effect on individuals' willingness to donate to the charity**. Although some differences were observed in the treatment, these differences are predominantly attributable to **random variation** or the **influence of heterogeneous variables**, rather than the effectiveness of the treatment group.

VI. Limitation

Due to time constraints, not every potential correlation or pattern among the variables was examined using statistical methods. Additionally, there are multivariate impacts that could warrant further investigation. However, based on the preliminary results, it seems reasonable to conclude that the treatment does not have a significant impact on individuals' donation behavior.