4-H National Mentoring Program Analysis

Joanna Skluzacek, Ji Hoon Woo

Abstract

4-H National Mentoring Program conducted both pre and post surveys to observe how the perceptions of participants have been changed through the curriculum. The program positively affected participants in 2017 and it is discovered that the mean difference of questionnaire between pre and post surveys was statistically significant to say that the program indeed was effective from most participants ($p < .001$). On the other hand, Year 2018 was not successful as the prior year. It is observed that the efficacy was not large, and even the program worsened off for those particular groups (e.g. Hispanic ethnicity or participants from Starbuck area). The program needs enhancements and additional development to bring more success in the following year.

Approach

Individual participants were asked to fill out the survey before they enter the program. A year after the end of the program, they again did identical survey to see how the program impacted forming relationships with family or people around them. The questionnaire is made of 7 points Liker-Scale. 1 indicates "Very Strongly Disagree", 2 for "Strongly Disagree", 3 for "Disagree", 4 for "Neither Agree nor Disagree" and vice versa.

After collecting data, four total datasets were given for the analysis. It contained information from year 2017 & 2018 and each year conducted both pre and post surveys. To investigate statistical analysis, data cleansing was performed from each single excel file using Microsoft Word. After removing unnecessary questions or information, the names of the columns are abbreviated for convenience.

Data compilation has been performed in the next process to match participants from both pre and post surveys. Individual IDs that do not correspond to both surveys were dropped out. That is, the test was performed only with those matching pairs sharing identical ID Number.

Method

The test was conducted in two methods: One sample Paired T-test and Mann-Whitney U Test. Since the Paired T-test assumes for normal distribution within sample, the given datasets could not satisfy the assumption due to their small sample sizes. Instead, we conducted one additional test called Mann-Whitney U Test which does not assume for normality and robust with small sample sizes as a nonparametric test. Then, the outcome between two different tests were compared to ensure the analysis.

There were 45 exact matches ($N = 45$) in 2017 while there were only 26 matches ($N = 26$) in 2018 due to large dropouts in post 2018 survey. We approached the analysis by delving into the overall structure then focusing into small categories: Gender, Ethnicity and Place.

Data Analysis & Result

Section 1.1 - Overall Distribution of Year 2017

The Figure 1 displays the overall distribution of the answers from 12 questions that are based on Likert-Scale. While the histogram of Pre 17 is slightly skewed to the right, the Post 17 is heavily right-skewed that we can infer that lar number of participants positively answered questions than Pre 17.

The Figure 1 is shown in gradient to represent the proportion in each question. It aids in recognizing the approximate change from Pre 17.

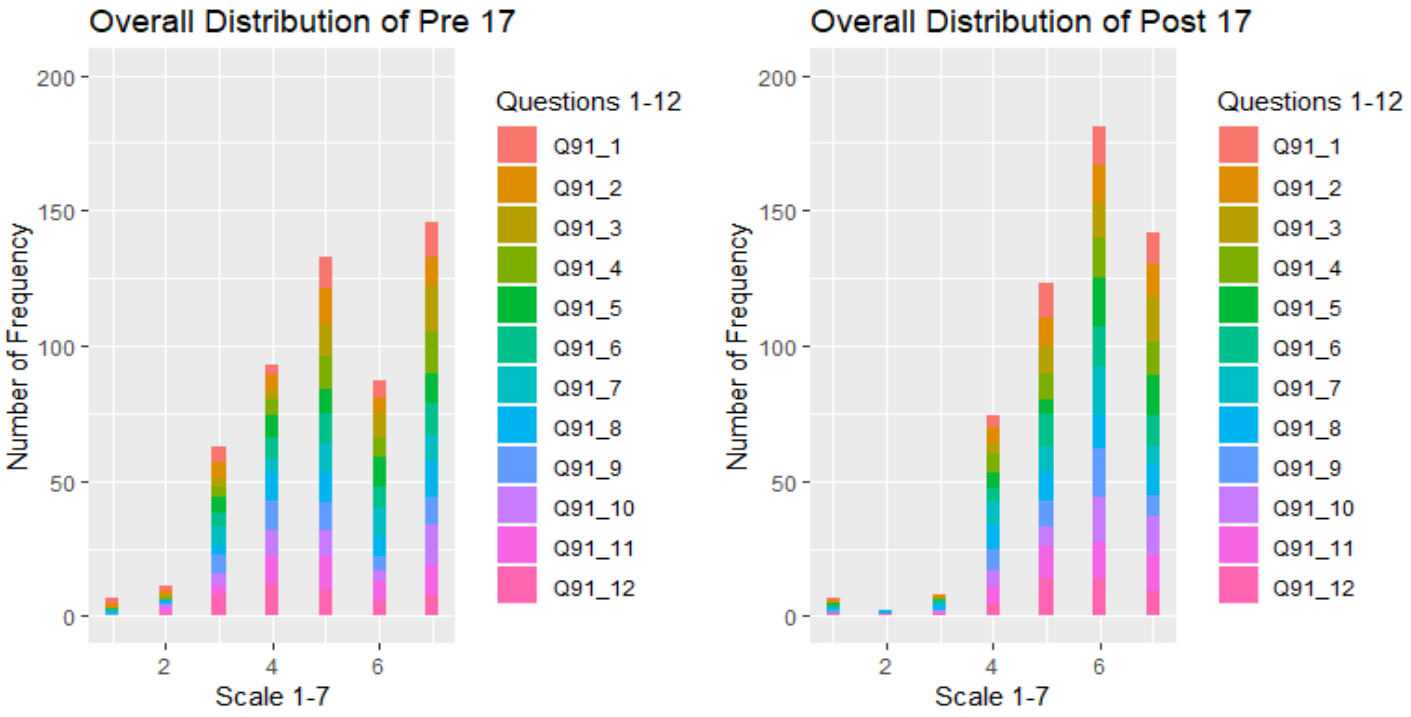

Figure 1

Section 1.2 - Year 17 Categorization by Question

This section concentrates on the distribution of each question in depth. Each boxplot represents each question in order along with its quartiles. Bolded line within colored boxplots represent the median of the distribution.

When comparing the distribution of questions in pre17 with those of post 17, it is clear that pre 17 is widespread that the boundary of each box is longer than those of post 17. In other words, there were some respondents who answered questions with low value of scale. Afterwards, post 17 is describing more condense answers that are highly distributed within high value of scale considering low value as outliers (dotted points).

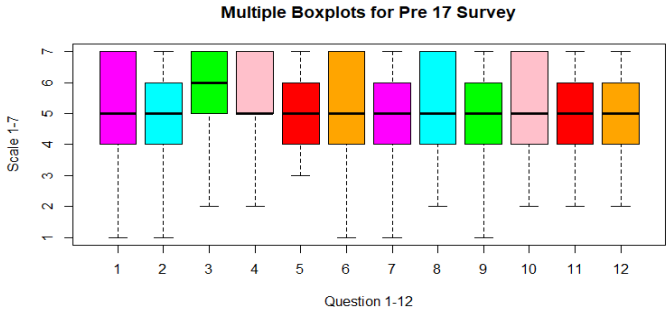

Figure 2.1

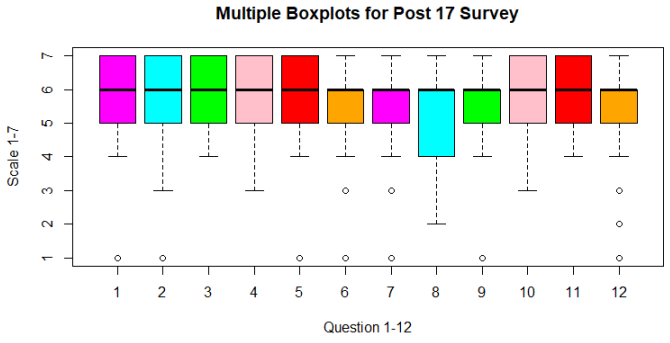

Figure 2.2

Section 1.3 - Mean Difference in Pre and Post 2017

The mean for pre survey was 5.2 while the post was 5.6. We aggregated every single value of answers earned in each twelve questions, then experimented with one sample Paired T-test identify the mean difference between pre and post survey. The result was strongly significant that p-value was less than 0.001 and 0.003 for Mann-Whitney U-Test.

Section 1.4 - Year 17 Categorization by Gender

The distribution of the survey is illustrated in Figure 3.1 that the number of male participants were larger than the number of females. Since there was a missing out data, the total number accounted for N = 44 including one response which answered "Female" in pre 17 but changed to "Other" in Post 17. We first divided the dataset into 2 major subsets: Female and Male to identify overall influence of the program in each gender.

The result was quite interesting that the mean difference from the overall answers was significantly large in male (4.974 vs. 5.621) while female (5.662 vs. 5.654) have not had any change in the mean. The Figure 3.2 is displaying the distribution based on each gender. There is almost no change in the left plot while huge change is shown in male participants. Then, we focused on male in detail to observe how many questions were significantly changed from the pre 17.

After conducting Mann-Whitney U-Test for each question with only male, the outcome showed that question 1, 2, 5, 9, 11 and 12 had p-value less than 0.05 meaning that male students were positively affected by the program and answered with high value.

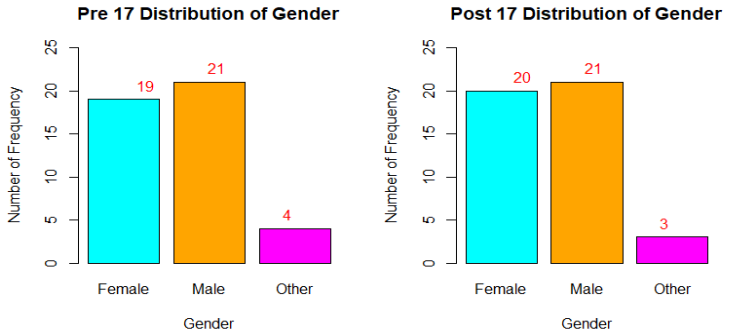

Figure 3.1

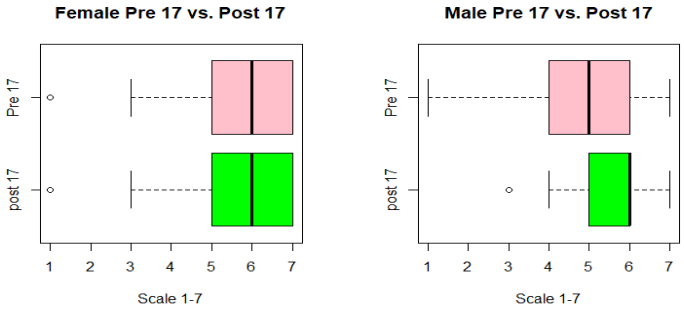

Figure 3.2

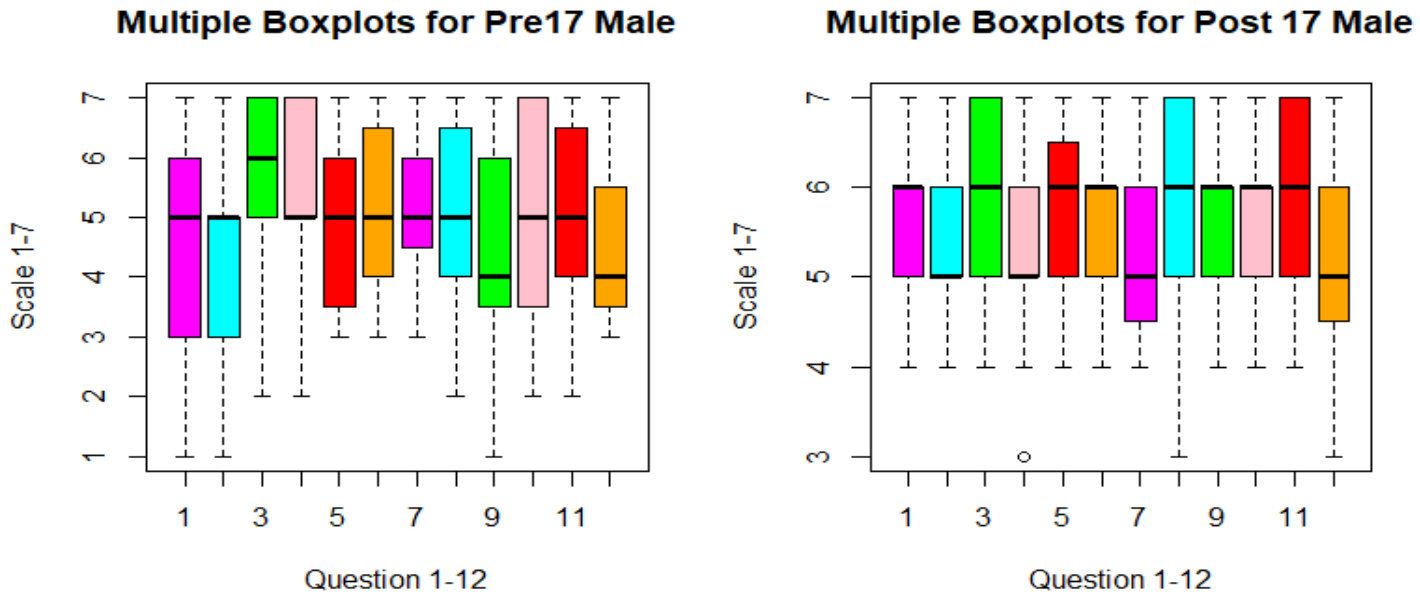

Figure 3.3

Figure 3.4 through 3.6 illustrate each single question that has significant changes in its mean value. The percentage represent the proportion of answers in scale of 5 – 7. We tried to describe the overall positiveness and in fact male led to significant changes in their answers.

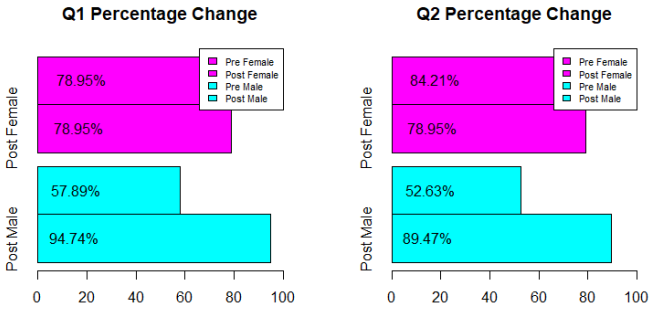

Figure 3.4

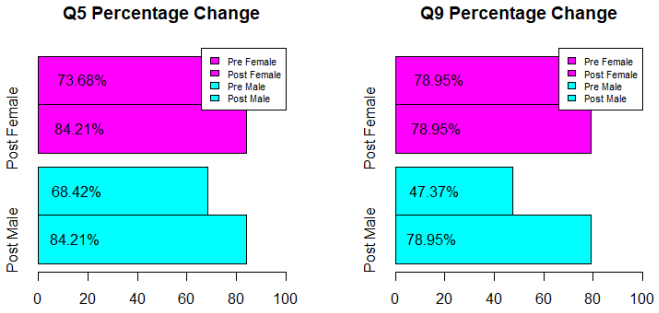

Figure 3.5

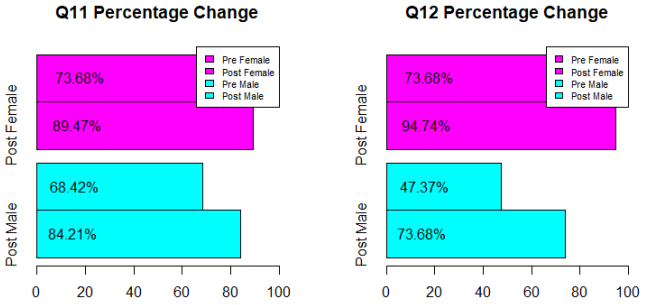

Figure 3.6

Section 1.5 - Year 17 Categorization by Race

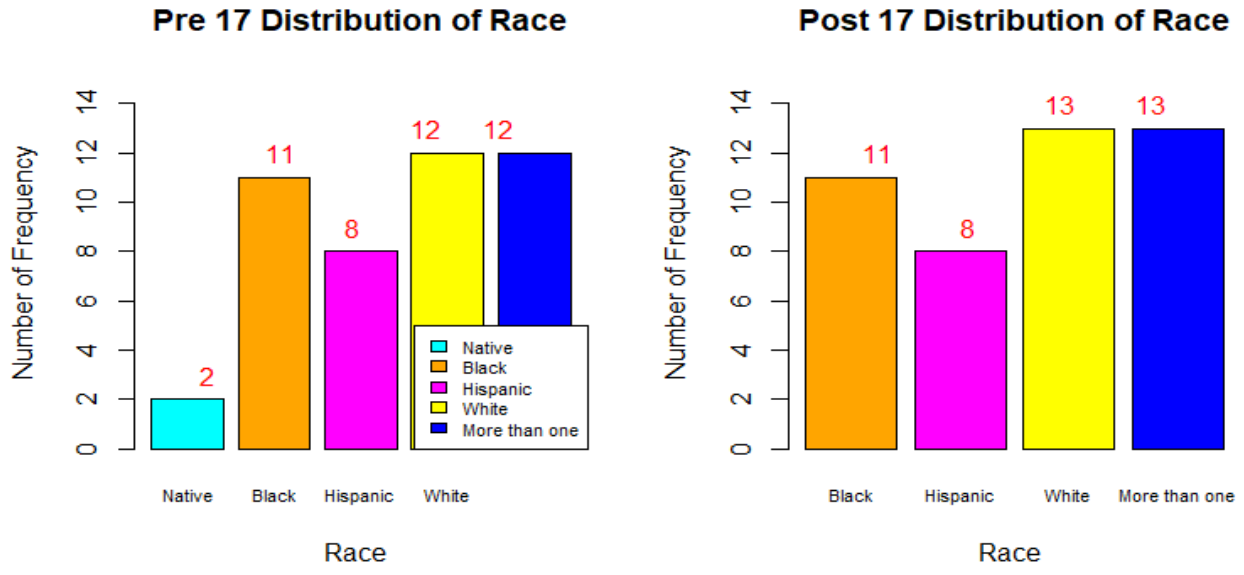

Figure 4.1

Figure 4.1 illustrates the number of people in each ethnicity. Each ethnicity except for Native American is proportionately distributed. Likewise, we only conducted the analysis with matching pairs, the number of samples were actually small than the figure shown above.

As we conducted gender analysis, race analysis has been executed in the process and further analysis has been performed with only significant features. According to Mann-Whitney U-Test, both Hispanic (n = 6, p-value < .002) and Mixed race (n = 12, p-value < .002) have very low p-values than other races. For Hispanic, the overall mean from the answers in pre 17 was 4.94, however it increased to 5.96. when it comes to Mixed, the mean of 5.05 rose up to 5.49. Figure 4.2 shows how answers were combined to high value.

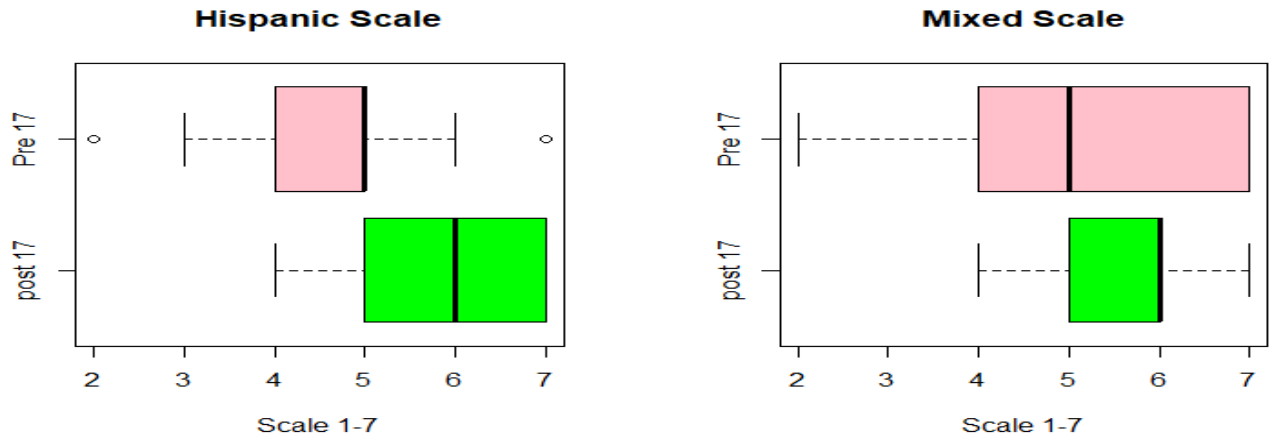

Figure 4.2

We conducted Mann-Whitney U-Test by combining Hispanic with Mixed race to increase the power of significance between those largely impacted by program and vice versa. Question 6 and 12 were shown to be statistically significant that it is confident to say that they were positively impacted by the program in 2 questions.

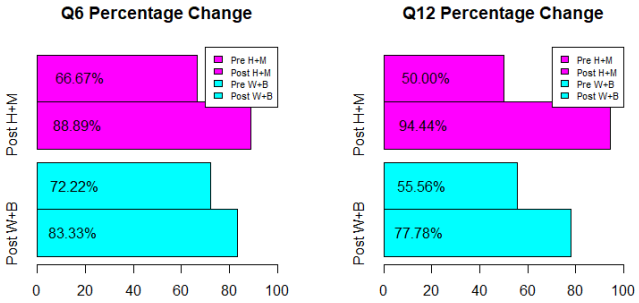

Figure 4.3

Section 1.5 – Year 17 Categorization by Place

Figure 5.1 shows the number of students at that moment of survey. Highland had the most students followed by Dr. Jones, Starbuck and Longfellow.

The outcome for each place is surprising that Starbuck actually had a significance of mean decrease from pre 17 survey (p-value = 0.005). Dr. Jones (p-value = .002) and Longfellow (0.2) both had significance of positive impact of the program. Even though we found interesting outcome in terms of the place, this is suspicious since the sample size is too small to prove its validity.

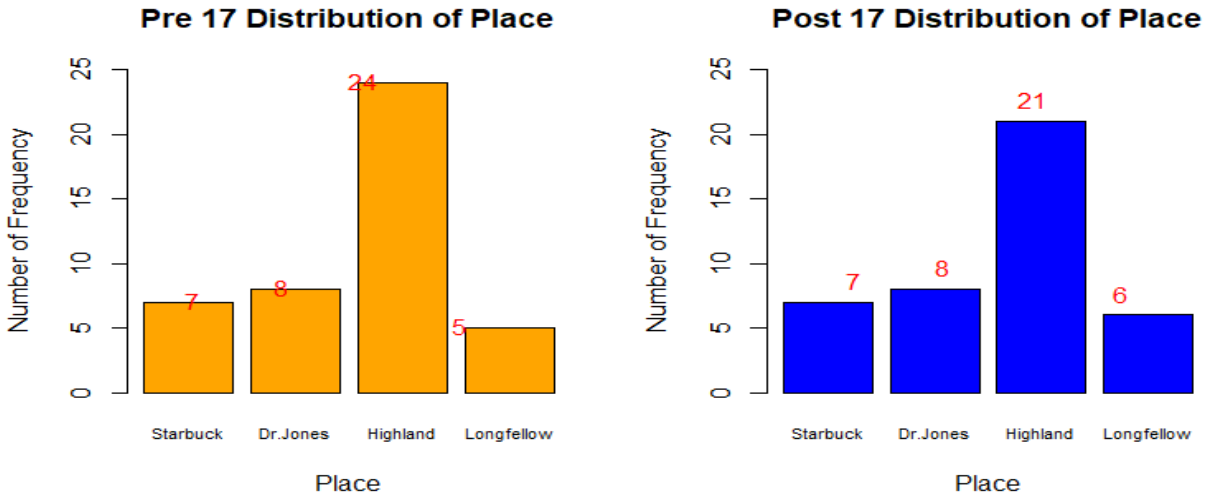

Figure 5.1

By looking at the distribution of each district, the result was interesting that both Starbuck and Longfellow showed decrease of mean values after the program. Figure 5.2 and 5.3 show that Starbuck had significant amount of decrease and followed by Longfellow. The other two districts: Dr. Jones and Highland on the other hand improved from the program.

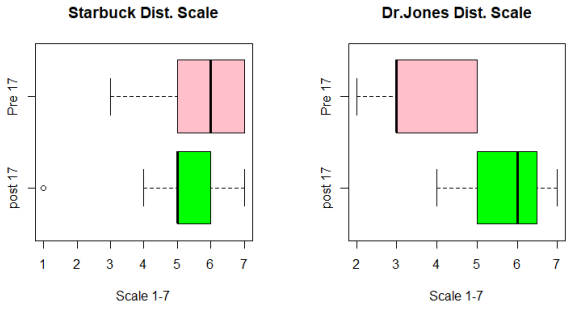

Figure 5.2

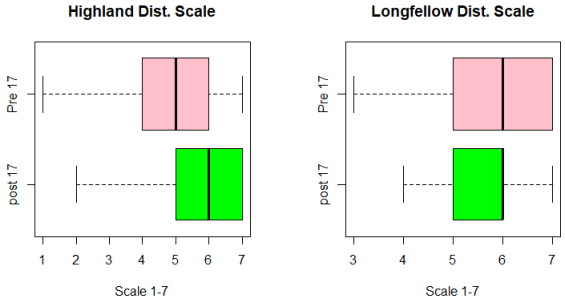

Figure 5.3

Question 3 & 6 were shown to be significantly decreased when Starbuck and Longfellow are combined together and measured for mean difference. Figure 5.4 is showing the percentage change from two groups: Group 1 for Starbuck & Longfellow and Group 2 for Dr. Jones and Highland.

As we observe, the percentage decreased in Group 1 while Group 2 increased that the effect of education differ by districts.

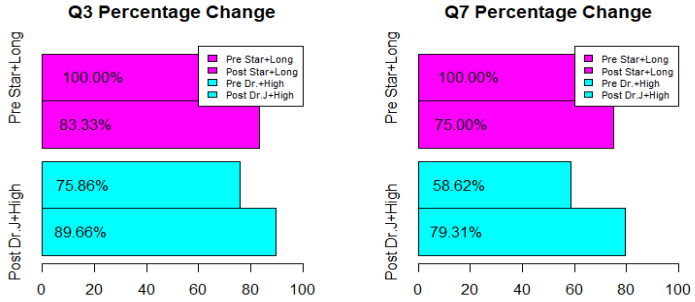

Figure 5.4

Section 2.1 - Overall Distribution of Year 2018

Likewise, same procedure was conducted on Year 2018. Figure 6.1 shows that the distribution is not normally concentrated. There were only 26 (N=26) matched pairs in this experiment. When we conducted hypothesis testing of whether the mean different was statistically significant between pre and post, the result was not (p-value < .2). 2018 Pre survey mean was 5.01 while post survey was only 5.12. The increase was not large, and we could say the program was not as effective as the prior year.

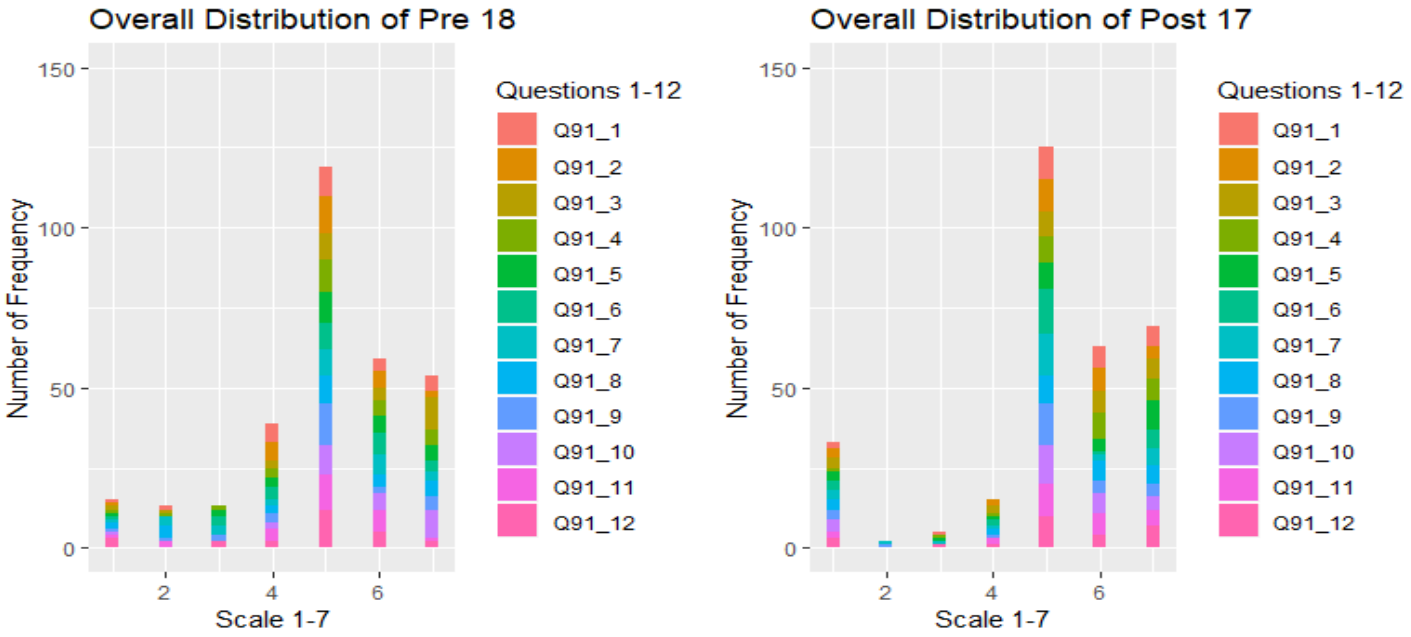

Figure 6.1

Section 2.2 - Year 18 Categorization by Question

In Figure 6.2, we can see that the distribution of each question is more concentrated and scale of 1 is considered as outliers in the post survey. Still, the height of median from both surveys is not much increased indicating that impact was not significant.

In fact, none of the questions were significant after experimenting with Mann-Whitney U-Test.

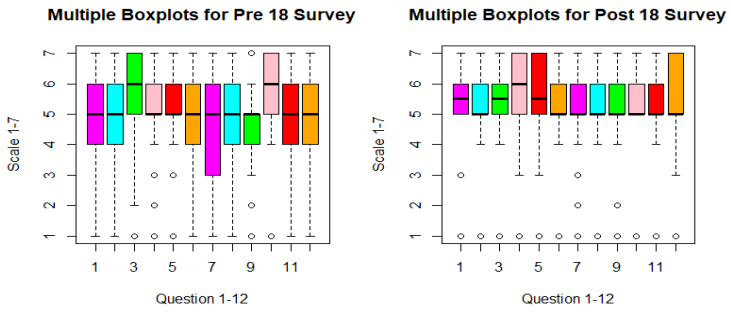

Figure 6.2

Section 2.3 - Year 18 Categorization by Gender

The distribution of gender was not equal. So, we experimented with only matching pairs again. Interestingly, the mean decreased from female participants. In the pre survey, it was 5.4 but it went down to 5.2. On the other hand, the mean for male slightly increased from 4.7 to 4.8. Then, Mann-Whitney U-Test was performed for each questions of female answered, but there was no significant decrease in answers. It was only a slight decrease that are not statistically significant.

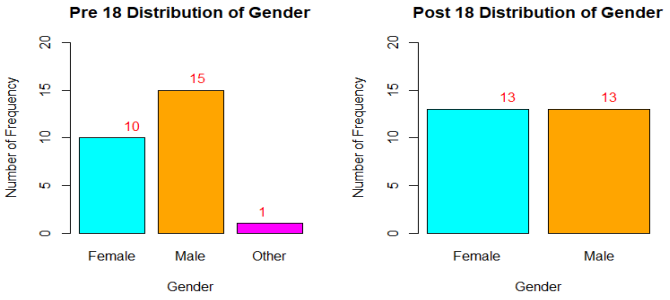

Figure 7.1

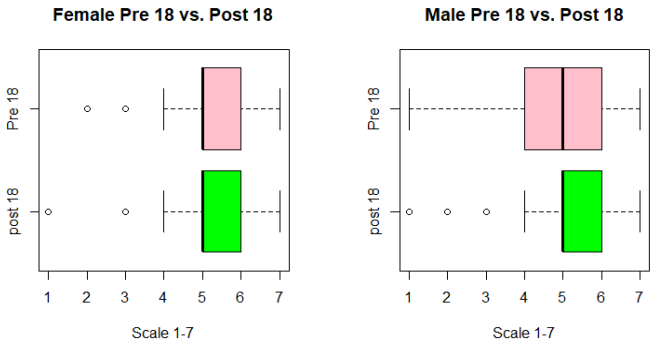

Figure 7.2

Section 2.4 - Year 18 Categorization by Race

Figure 8.1 illustrates the distribution of participants by ethnicity. Since there were unmatched pairs and small samples for each ethnicity, we did not conduct further analysis because the results may be less conclusive.

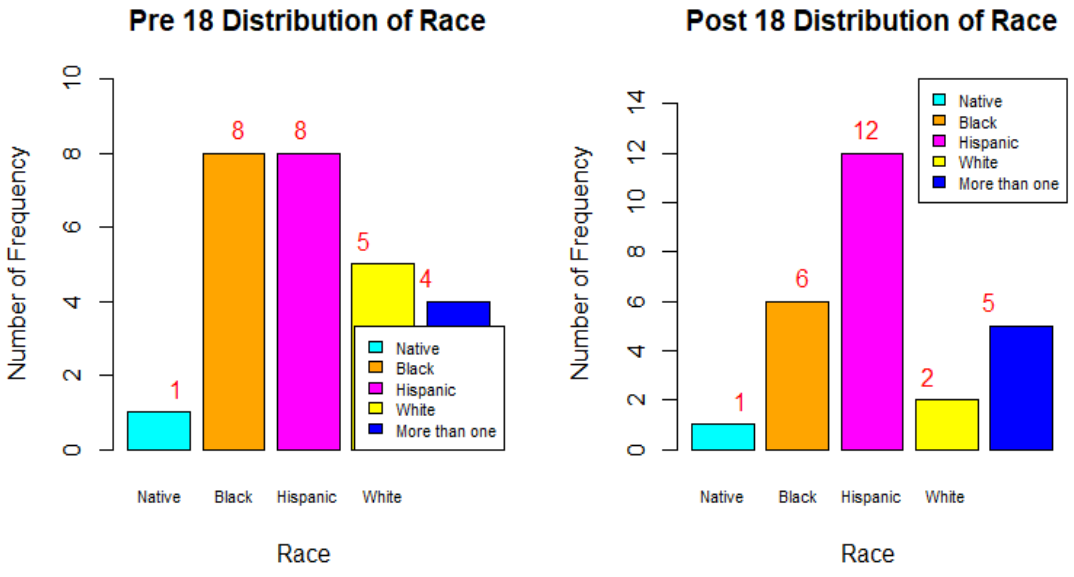

Figure 8.1

Section 2.5 – Year 18 Categorization by Place

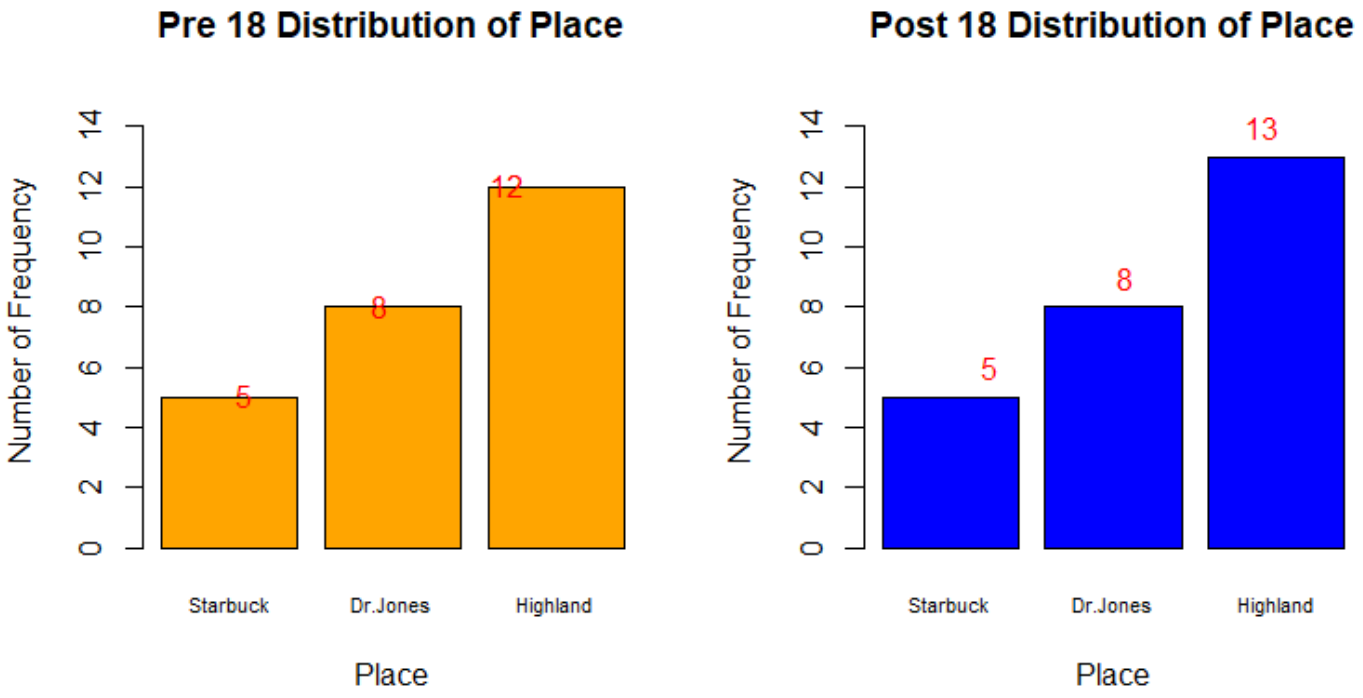

Figure 8.2

Again, the mean decreased in district of Starbuck while two other districts: Dr. Jones and Highland slightly increased their means. In fact, the program is not effective in Starbuck. The only question that the mean of Dr Jones & Highland combined significantly increased was Q1.

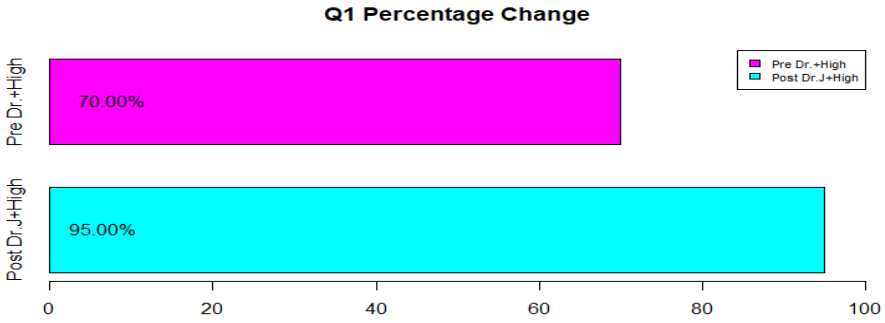

Figure 8.3

Summary

In summary, year 2017 was more successful than 2018. The curriculum in the district of Longfellow was not satisfactory that the statistics in 2017 data implies low numbers compared to other regions. Eventually, Longfellow dropped out the program in the following year in 2018. The program was also successful in male participants than female. We could assume that the male actively engaged in the program and made positive impacts. The other assumption is that since the pre survey answers from the female was already higher than the

male, the marginal increase of scale was slower than the male. The program had both positive and negative impacts based on each category. It should find ways to enhance the strengths while rectifying the problems.

References

Meek, G., Ozgur, C., and Dunning, K., (2007). Comparison of the t vs. Wilcoxon Signed-Rank Test for Likert Scale Data and Small Samples. *Journal of Modern Applied Statistical Methods*: Vol. 6(1), Article 10. DOI: 10.22237/jmasm/1177992540

Boone, H., and Boone, D., (2012 Apr.). Analyzing Likert Data. Journal of Extension: 2TOT2. http://www.joe.org/joe/2012april/tt2p.shtml