Be Well Program: EOP High Participation Analysis

The Data

Variable list:

1. HbA1C: hba1c
2. Fasting Blood Glucose: fast_blood_gluc
3. Blood Pressure (Systolic): blood_press_sys
4. Weight
5. BMI
6. LDL
7. Triglycerides (indicators 13, 14, 17, 18, 20, 22, 25)
8. Vegetable Dervings : pa_21
9. Fruit Servings :pa_22
10. Soda Consumption: pa_23
11. Sugar-Sweetened Beverages Consumption: pa_24
12. Attitudes Toward Healthy Foods : healthy_eating_important

We will be renaming the time points in the 'time' variable as follows only for more convenience when analyzing the data:

'1' for baseline measurement

'2' for the end of the program assessment

'3' for the 6-month mark after the program

'4' for the 12-month mark after the program

'5' for the 24-month mark after the program

In this analysis, we will only be focusing on the baseline and EOP measurements.

Quick Inventory Check

The data has been filtered to include only: - People who have data for both baseline and EOP - Observations, for each health metric, that were recorded in both baseline and EOP. (Eg: if someone had only baseline triglycerides levels recorded, that observation is disregarded in this analysis by being turned into NA)

Total sample size:

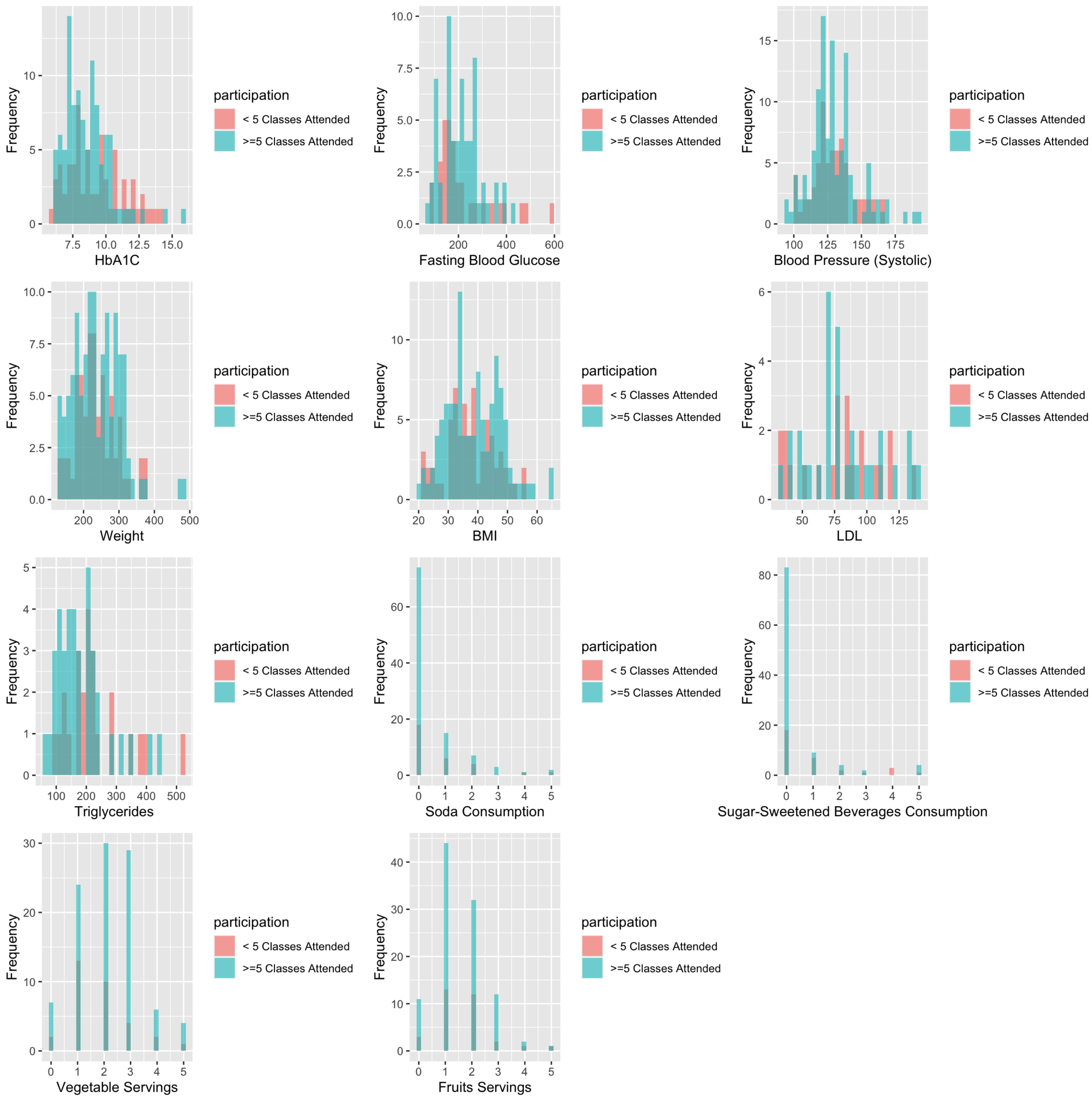
[1] 194 So the data set has 194 rows for 97 people.

Sample size for high participation: [1] 122 Out of the 97 people we have in our data, 61 had high participation. That is a bit under 2/3 of the population.

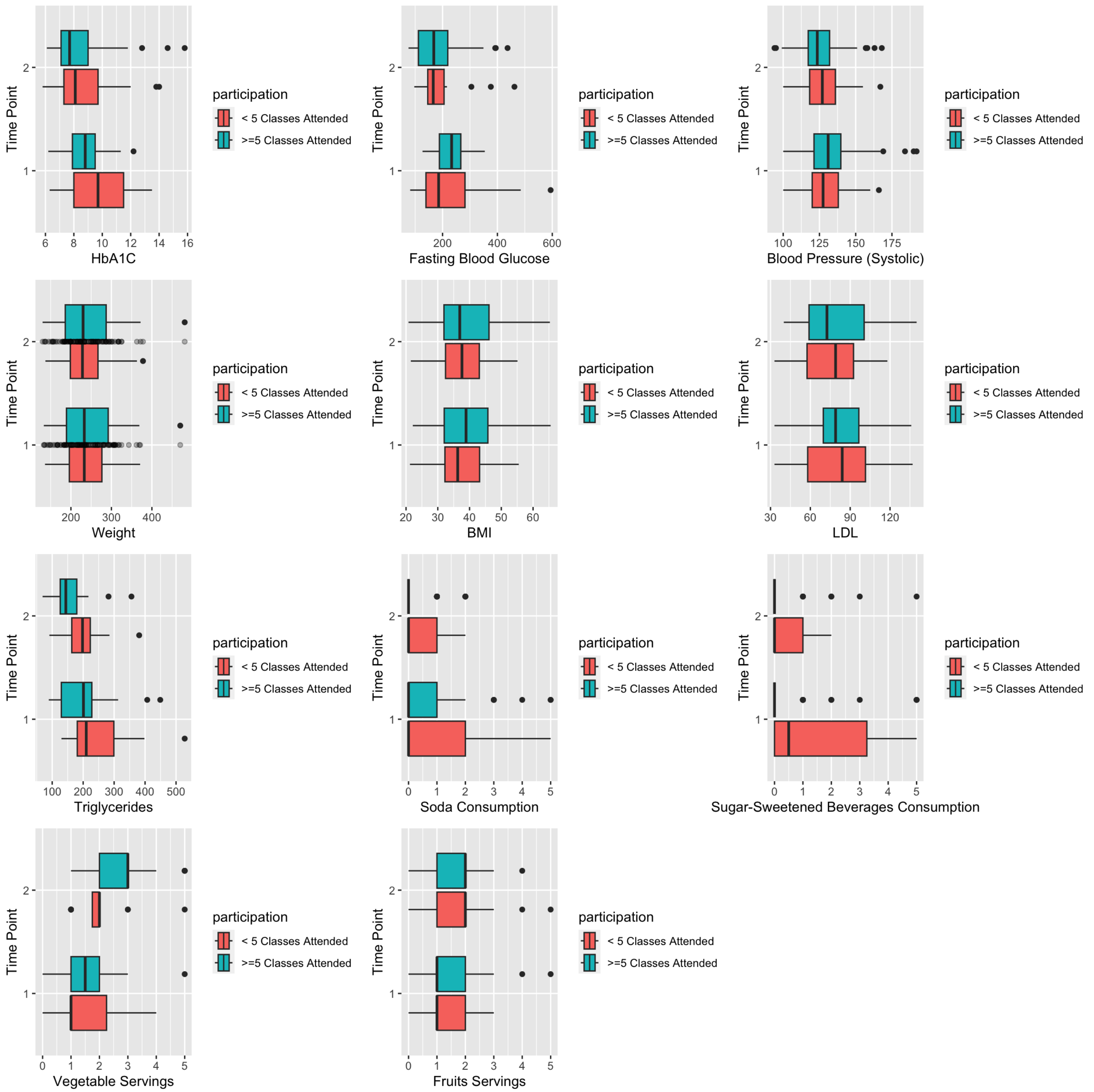
Plots and Graphs

Distributions of The Variables

In assessing the health program's impact, distribution plots were generated for key variables, combining data from both Time 1 (baseline) and Time 2 (EOP). This approach offers a comprehensive view of participant characteristics, facilitating a nuanced understanding of the program's potential influence. The distribution plots can aid in visually narrating the dataset's overall evolution, laying the groundwork for further detailed explorations of specific variables across the program's timeline.



Box Plots: The Distribution of Each Variable at Each Time Point:

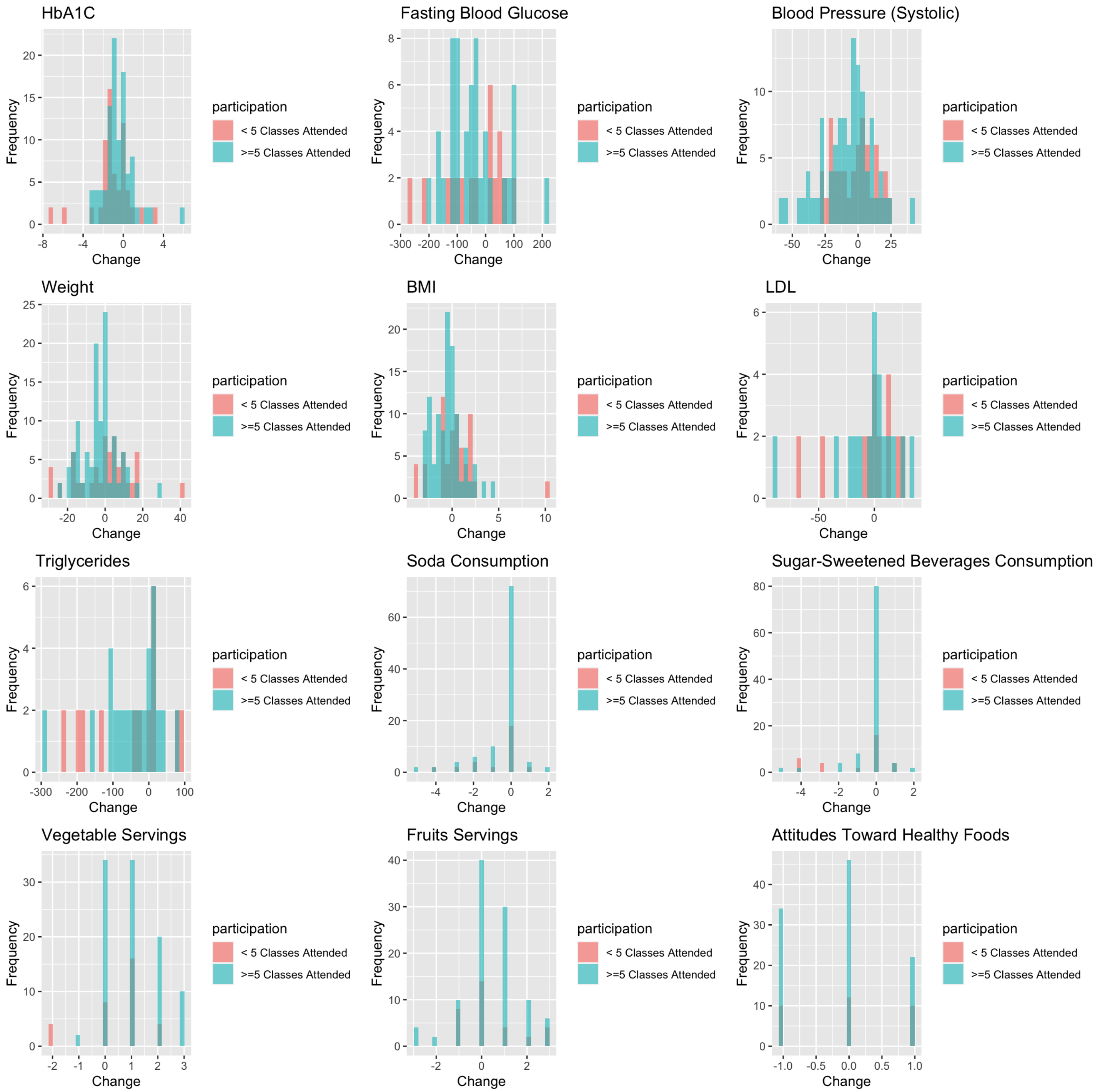


The box plots reaffirm the trends observed in mean values, with the addition of quartiles trends, offering a detailed portrayal of the data distribution.

One noteworthy observation from the box plots is the reduction in extreme outliers at the end of the program (EOP) compared to the baseline. These outliers, indicative of poorer health conditions, appear to diminish post-program participation. This reduction underscores a potential mitigation of severe health issues among participants, aligning with the overarching goal of the health program.

This visual confirmation strengthens our earlier findings and reinforces the notion that the health program may be instrumental in fostering healthier conditions among participants.

The Distribution of Rates of Change in the Variables Before and After the Program

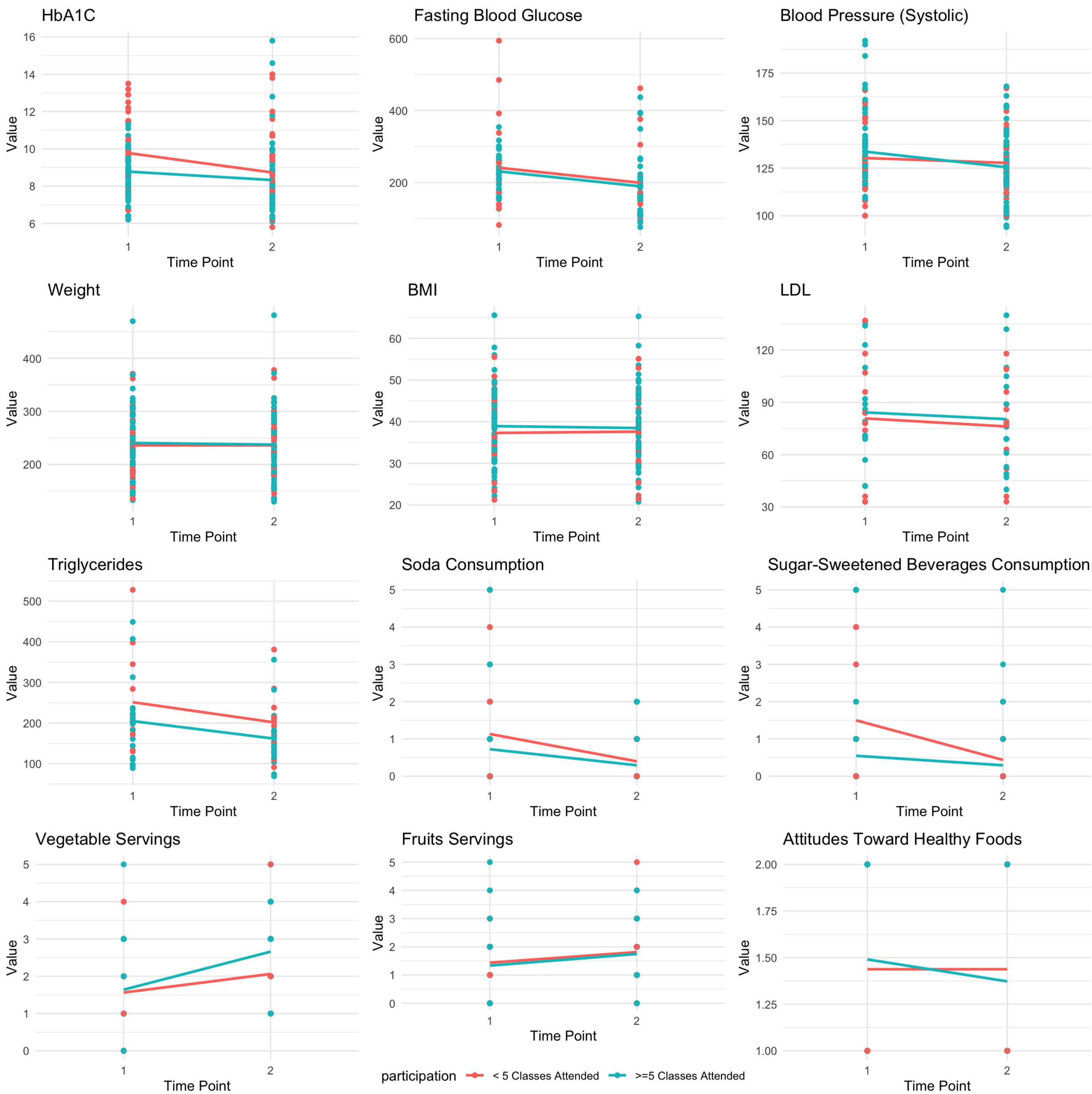


This analysis reveals a favorable trend, with the majority of variables showcasing decreases rather than increases.

For health attributes that we aim to decrease, such as HbA1c and blood sugar levels, the prevalence and extremity of decreases in the rates of change are particularly encouraging. Although there are instances of participants recording higher measurements in certain health attributes, there are also even more notable decreases observed in the majority of variables.

This asymmetry in the distribution suggests that, on the whole, participants are experiencing improvements in key health indicators rather than deteriorations.

Mean Line Graphs



Linear Mixed Effects Models

The models aim to provide rates of change in the response variable from the baseline to the target time point.

We have chosen to treat the ‘time’ variable as a categorical factor rather than a numeric variable, even though it has been transformed into sequential numbers (1, 2, 3, 4, 5), for several reasons. First, our decision is rooted in the recognition that the rates of change in the response variables over time may not be linear and continuous. Representing ‘time’ as a numeric variable could potentially mislead by implying a linear relationship between time points and response changes that might not accurately reflect the underlying dynamics.

Secondly, treating ‘time’ categorically helps avoid extrapolation or interpolation beyond the provided time points. Using ‘time’ as a factor acknowledges that the observations at our specific time points are the only reliable data we have, and we should refrain from assuming relationships outside these intervals. This approach can be helpful in maintaining the integrity of our findings and prevent unwarranted assumptions about the program’s effects at unobserved time points.

In our model, we will have time points 1 and 2 for baseline and EOP. The summaries of our models will be presented in forms of tables that include the following columns that we will be analyzing further with each variable: - effect: this column indicates whether a variable is a fixed or random effect. In our analysis, the investigated variables will be the fixed effect and the participant id will be the random effect, accounting for the relationship between different observations of the same participant. - term: the coefficient, or in this case, the time point (baseline and EOP as Intercept and time2) - estimate: the average difference in the investigated variable between baseline and EOP - p.value: suggests whether there is evidence that the observed change is based on pure chance or potential effects from the program.

P-Value Scale:

For convenience, we can refer to this P-Value scale when determining if an observed trend statistically has strong evidence or is possibly due to random chance.

$p < 0.001$: Extremely strong evidence.

$0.001 \leq p < 0.01$: Very strong evidence.

$0.01 \leq p < 0.05$: Strong evidence.

$0.05 \leq p < 0.1$: Moderate evidence.

$p \geq 0.1$: Not considered statistically significant.

For High Participation

The data has been filtered to only contain people who have high participation/attended 5 or more classes.

HbA1C

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	8.7784314	0.2375459	36.954670	73.00071	0.000000
fixed	NA	time2	-0.4529412	0.2102912	-2.153876	50.00000	0.036095
random	record_id	sd__(Intercept)	1.3229365	NA	NA	NA	NA

On average, there is a decrease in HbA1c levels from baseline to the end of the program by approximately 0.45 units. This finding is supported by a statistically significant p-value of 0.036, falling within the range of 0.01 to 0.05, which provides strong evidence against the possibility that these results occurred due to chance alone.

Fasting Blood Glucose

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	230.97419	13.98363	16.517465	55.45124	0.000000
fixed	NA	time2	-41.55484	16.70548	-2.487498	30.00000	0.018649

effect	group	term	estimate	std.error	t-value	df	p.value
random	record_id	sd__(Intercept)	41.66739	NA	NA	NA	NA

On average, fasting blood glucose levels decrease by approximately 41.55 units from baseline to the end of the program. This inference is bolstered by a significant p-value of 0.019, within the range of 0.01 to 0.05, offering strong evidence against the hypothesis that the observed changes are merely random.

Blood Pressure (Systolic)

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	133.689655	2.360984	56.624560	96.84435	0.000000
fixed	NA	time2	-8.172414	2.540910	-3.216333	57.00000	0.002141
random	record_id	sd__(Intercept)	11.665146	NA	NA	NA	NA

Systolic blood pressure exhibited a reduction over time, with an estimated decrease of approximately 8.17 mmHg from baseline to the end of the program ($p = 0.002$). There is strong evidence provided by the p-value falling below the 0.01 threshold, indicating that the observed change is unlikely to be due to chance alone.

Weight

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	240.596491	8.779118	27.405543	56.60466	0.000000
fixed	NA	time2	-3.031579	1.286673	-2.356138	56.00000	0.021991
random	record_id	sd__(Intercept)	65.923997	NA	NA	NA	NA

The analysis reveals a decrease in weight following the program, with an estimated reduction of approximately 3.03 pounds from baseline to the program's conclusion ($p = 0.022$). The p-value suggests that the observed decrease is unlikely to occur by chance.

BMI

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	38.934182	1.2657754	30.759156	54.78705	0.000000
fixed	NA	time2	-0.464000	0.2153337	-2.154795	54.00000	0.035655
random	record_id	sd__(Intercept)	9.319075	NA	NA	NA	NA

The BMI analysis indicates a modest decline, with an estimated decrease of approximately 0.46 units from baseline to the end of the program ($p = 0.036$). The p-value falls within the range of $0.01 \leq p < 0.05$, indicating strong evidence supporting the observed BMI decrease.

LDL

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	84.25000	7.451894	11.3058501	22.98527	0.000000

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	time2	-3.87500	7.050340	-0.5496189	15.00000	0.590671
random	record_id	sd__(Intercept)	22.15476	NA	NA	NA	NA

The LDL analysis suggests a slight reduction, with an estimated decrease of approximately 3.88 units from baseline to the end of the program (p = 0.591). This finding indicates that the observed change in LDL levels may occur due to chance. The p-value, falling within the range of p ≥ 0.1, implies that there is insufficient evidence to support a meaningful association.

Triglycerides

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	204.84211	19.48477	10.512933	28.51638	0.000000
fixed	NA	time2	-43.21053	19.24398	-2.245404	18.00000	0.037540
random	record_id	sd__(Intercept)	60.78921	NA	NA	NA	NA

The analysis of triglycerides indicates a notable decrease, with an estimated reduction of approximately 43.21 units from baseline to the end of the program (p = 0.038). The p-value falls within the range of 0.01 ≤ p < 0.05, indicating strong evidence against the null hypothesis.

Soda Consumption

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	0.7254902	0.1433117	5.062322	91.34839	0.000002
fixed	NA	time2	-0.4313725	0.1686275	-2.558139	50.00000	0.013600
random	record_id	sd__(Intercept)	0.5677613	NA	NA	NA	NA

The analysis of soda consumption reveals an estimated reduction of approximately 0.43 units from baseline to the end of the program (p = 0.014). The p-value falls within the range of 0.01 ≤ p < 0.05, indicating strong evidence that the observed changes are unlikely to be due to chance.

Sugar-Sweetened Beverages Consumption

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	0.5490196	0.1554347	3.532157	76.9013	0.000701
fixed	NA	time2	-0.2549020	0.1477760	-1.724921	50.0000	0.090719
random	record_id	sd__(Intercept)	0.8217628	NA	NA	NA	NA

The analysis of sugar-sweetened beverages consumption indicates an estimated decrease of approximately 0.25 units from baseline to the end of the program (p = 0.091). The p-value falls within the range of 0.05 ≤ p < 0.1, indicating moderate evidence that the observed change is not random. Therefore, while there appears to be a trend towards reduced consumption, further investigation is warranted to draw definitive conclusions.

Vegetable Servings

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.6400000	0.1516831	10.81201	75.38281	0.000000
fixed	NA	time2	1.0200000	0.1442526	7.07093	48.99892	0.000000
random	record_id	sd__(Intercept)	0.7938313	NA	NA	NA	NA

The analysis suggests an increase in vegetables servings over time, as indicated by the coefficient estimate of 1.02. This result is supported by a p-value of near 0, indicating extremely strong evidence that the observed increase is unlikely to be due to chance and merits further exploration.

Fruits Servings

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.3333333	0.1335926	9.980593	98.64524	0.000000
fixed	NA	time2	0.4117647	0.1775133	2.319628	49.99999	0.024485
random	record_id	sd__(Intercept)	0.3265986	NA	NA	NA	NA

The analysis reveals an increase of about 0.41 servings in fruit servings over time. This observation is reinforced by a p-value of 0.025 which falls within the range $0.01 \leq p < 0.05$, indicating strong evidence that the observed rise in fruit consumption is unlikely to be attributed to random variation alone.

Low Participation

I thought including the analysis with filtered data containing only participants who attended fewer than 5 classes in this file as well could help with possible comparisons between participation levels. For this part, I didn't include the results interpretations since they are, in statistical terms, not much different from the other interpretations provided above. The more useful parts of this would be the coefficients and their p-values.

HbA1C

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	9.775758	0.3631266	26.921071	47.69222	0.000000
fixed	NA	time2	-1.045454	0.3309219	-3.159219	32.00000	0.003445
random	record_id	sd__(Intercept)	1.595151	NA	NA	NA	NA

Fasting Blood Glucose

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	241.14375	30.76124	7.839207	21.28852	0.000000
fixed	NA	time2	-41.88750	26.11282	-1.604097	15.00000	0.129536
random	record_id	sd__(Intercept)	98.41255	NA	NA	NA	NA

Blood Pressure (Systolic)

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	130.323529	2.684673	48.5435421	51.11142	0.000000
fixed	NA	time2	-2.558823	2.575834	-0.9933961	33.00000	0.327746
random	record_id	sd__(Intercept)	11.500446	NA	NA	NA	NA

Weight

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	236.0588235	9.611713	24.5594961	34.13088	0.000000
fixed	NA	time2	0.1235294	2.495401	0.0495028	33.00000	0.960817
random	record_id	sd__(Intercept)	55.0929341	NA	NA	NA	NA

BMI

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	37.2969697	1.4471125	25.773372	33.38378	0.000000
fixed	NA	time2	0.3057576	0.4211438	0.726017	32.00000	0.473105
random	record_id	sd__(Intercept)	8.1351071	NA	NA	NA	NA

LDL

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	80.818182	9.307200	8.6834046	14.84187	0.000000
fixed	NA	time2	-4.636364	8.432915	-0.5497937	10.00000	0.594526
random	record_id	sd__(Intercept)	23.700979	NA	NA	NA	NA

Triglycerides

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	251.33333	28.91949	8.690795	19.12674	0.000000
fixed	NA	time2	-49.58333	32.00579	-1.549199	11.00000	0.149608
random	record_id	sd__(Intercept)	62.36841	NA	NA	NA	NA

Soda Consumption

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.1333333	0.3137080	3.612702	24.96938	0.001331
fixed	NA	time2	-0.7333333	0.3581256	-2.047699	14.00000	0.059832

effect	group	term	estimate	std.error	t-value	df	p.value
random	record_id	sd__(Intercept)	0.7171372	NA	NA	NA	NA

Sugar-Sweetened Beverages Consumption

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.5000000	0.3474236	4.317497	29.56838	0.000162
fixed	NA	time2	-1.0625000	0.4606947	-2.306300	15.00000	0.035786
random	record_id	sd__(Intercept)	0.4830459	NA	NA	NA	NA

Vegetable Servings

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.5625000	0.2845867	5.490418	24.27695	0.000012
fixed	NA	time2	0.5000000	0.2886751	1.732051	15.00000	0.103771
random	record_id	sd__(Intercept)	0.7932003	NA	NA	NA	NA

Fruits Servings

effect	group	term	estimate	std.error	t-value	df	p.value
fixed	NA	(Intercept)	1.4375000	0.2676090	5.371643	28.2232	0.000010
fixed	NA	time2	0.3750000	0.3275541	1.144849	15.0000	0.270211
random	record_id	sd__(Intercept)	0.5361903	NA	NA	NA	NA