

Analysis Report

This report is structured as follows.

Contents

R Code Structure.....	2
Initial Setup:	2
Outlier Inspection:.....	2
Data Preprocessing:.....	2
Data Visualization:.....	2
Descriptive Statistics:.....	2
Normality Check:	2
Linear Mixed Model:	2
Data Export:	3
Normality Assessment	3
Descriptive Statistics.....	4
Mixed Models	12

R Code Structure

This section presents the structure of the R code built for this project.

Initial Setup:

Library & Working Directory: The openxlsx library is loaded, and the working directory is set to a specific folder.

Data Import: Reads an Excel file named "Data.xlsx" into a dataframe called df.

Outlier Inspection:

Setting Dependent Variables: The variable of interest (dvs) is specified.

Outlier Detection Function: A custom function find_outliers is defined to calculate Z-scores and identify outliers.

Execute Function: The function is executed, and its results are stored in two dataframes.

Data Preprocessing:

Check for Duplicate Columns: The code checks for duplicate column names and prints a message if any are found.

Make Column Names Unique: Ensures unique column names.

Factor Conversion: The 'Year' variable is converted to a factor.

Data Visualization:

Boxplots: Uses ggplot2 to create boxplots for each dependent variable.

Line Plots: Line plots are generated to show mean scores across groups and timepoints.

Descriptive Statistics:

Mean Value Calculation: Calculates the mean values for each group and timepoint and stores them in a dataframe.

Normality Check:

Normality Function: A function calc_descriptive_stats_for_dvs is defined to calculate skewness, kurtosis, and Shapiro-Wilk test statistics.

Execute Normality Function: Executes the function and stores the statistics.

Linear Mixed Model:

Library Import: Loads the nlme and dplyr libraries.

Factor Conversion: 'Year' and 'Group' are converted to factors.

Modeling: A loop constructs a linear mixed-effects model for each unique pair of groups.

ANOVA Table: Retrieves the ANOVA table from the model.

Model Summary: Stores the summary statistics of the model in a dataframe.

Data Export:

Export to Excel: Three dataframes (mean values, model summary, and normality assessment) are written to separate sheets in an Excel file.

Normality Assessment

The normality of the dependent variable was assessed using skewness, kurtosis, and the Shapiro-Wilk test (table below). The variables ‘%.imported.floor.material’, ‘%.subsistence.type.of.lighting’ and ‘%.of.Expensive.Type.of.Bathroom’ were highly kurtotic and severely not normally distributed. Shapiro-Wilk’s tests indicated all variables were non-normal. Although the central limit theorem supports the robustness of linear models to non-normality, especially when the sample size is large and the design is balanced, the researcher decided to normalize the variables using log-transformation. This procedure alleviated the Kurtosis and Skewness values of all the variables but ‘%.subsistence.type.of.lighting’. Further inspection revealed the existence of too many zero values in this variable along with 3 outliers. For this variable, no log-transformation was executed but three outliers were removed to normalize kurtosis.

Variable	Skewness	Kurtosis	W	P_Value
%.imported.material.of.roof	-0.960	2.735	0.865	0.000
%.imported.floor.material	3.230	15.792	0.656	0.000
%.imported.material.of.wall	2.571	8.870	0.507	0.000
%.subsistence.type.of.lighting	-2.983	13.875	0.641	0.000
%.Department.Born.not.from.Gracias.a.Dios	1.580	4.275	0.682	0.000
%.of.Expensive.Type.of.Bathroom	2.498	9.241	0.662	0.000
%.Department.lived.5.years.ago.not.from.Gracias.a.Dios	2.203	6.318	0.513	0.000
%.No.Land.Change.Job	1.356	3.051	0.614	0.000

The table below shows the final results for the log-transformed variables and ‘%.subsistence.type.of.lighting’ without the 3 extreme cases (Tusidaksa, Klubki and Benk in 2013).

Variable	Skewness	Kurtosis	W	P_Value
log_%.imported.material.of.roof	-1.742	5.340	0.754	0.000
log_%.imported.floor.material	0.382	2.935	0.985	0.256
log_%.imported.material.of.wall	0.010	1.901	0.949	0.149
log_%.Department.Born.not.from.Gracias.a.Dios	-0.362	1.813	0.920	0.000
log_%.of.Expensive.Type.of.Bathroom	0.122	2.343	0.985	0.273
log_%.Department.lived.5.years.ago.not.from.Gracias.a.Dios	-0.228	1.726	0.919	0.000

log_%.No.Land.Change.Job	-0.309	1.612	0.872	0.000
%.subsistence.type.of.lighting	-1.968	6.685	0.737	0.000

Descriptive Statistics

The data reflects changes in various metrics from the year 2001 to 2013 for Control and Treatment groups. Focusing on higher differences, several key observations can be made.

For the metric "%.imported.material.of.roof," both Control and Treatment groups show significant increases, with the Control group moving from 57.5% to 89.0% and the Treatment group from 49.5% to 87.2%. The difference is more pronounced in the Treatment group, suggesting a greater impact.

The metric "%.Department.Born.not.from.Gracias.a.Dios" also undergoes substantial changes. The Control group sees an increase from 1.5% to 28.7%, whereas the Treatment group experiences an even more dramatic rise from 11.1% to 43.4%.

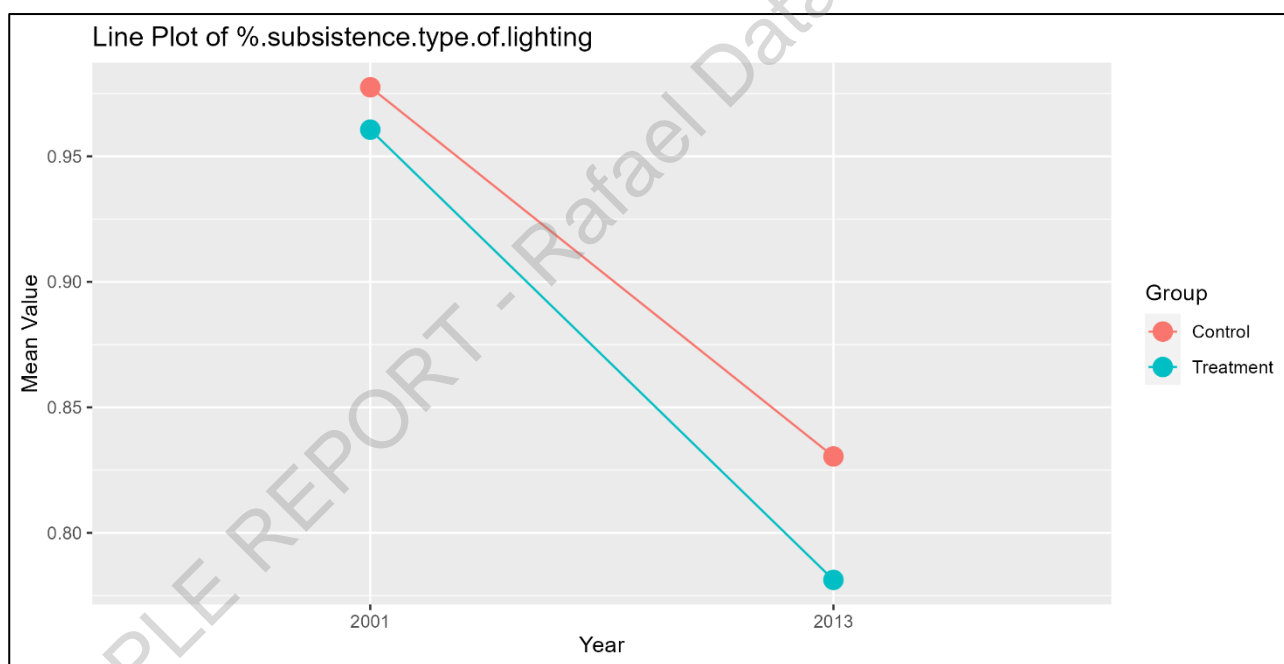
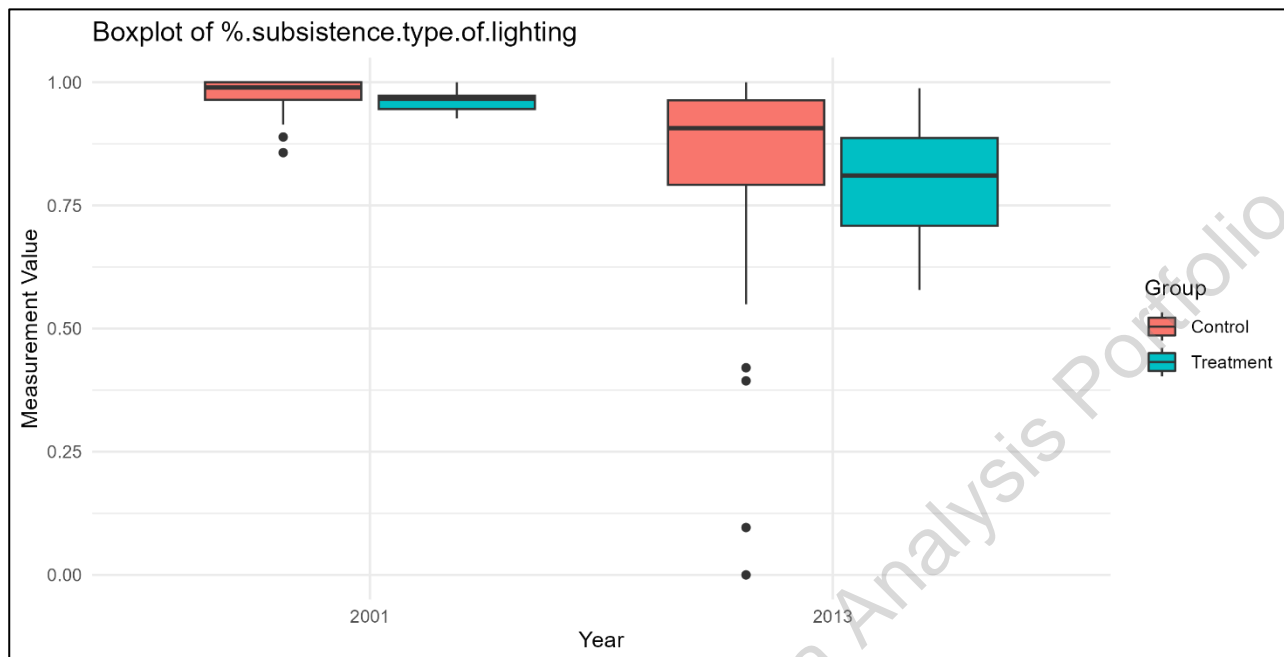
Another noteworthy metric is "%.of.Expensive.Type.of.Bathroom," where the Control group changes from 2.1% to 7.9%. The Treatment group shows a more marked increase, going from 7.9% to 22.5%.

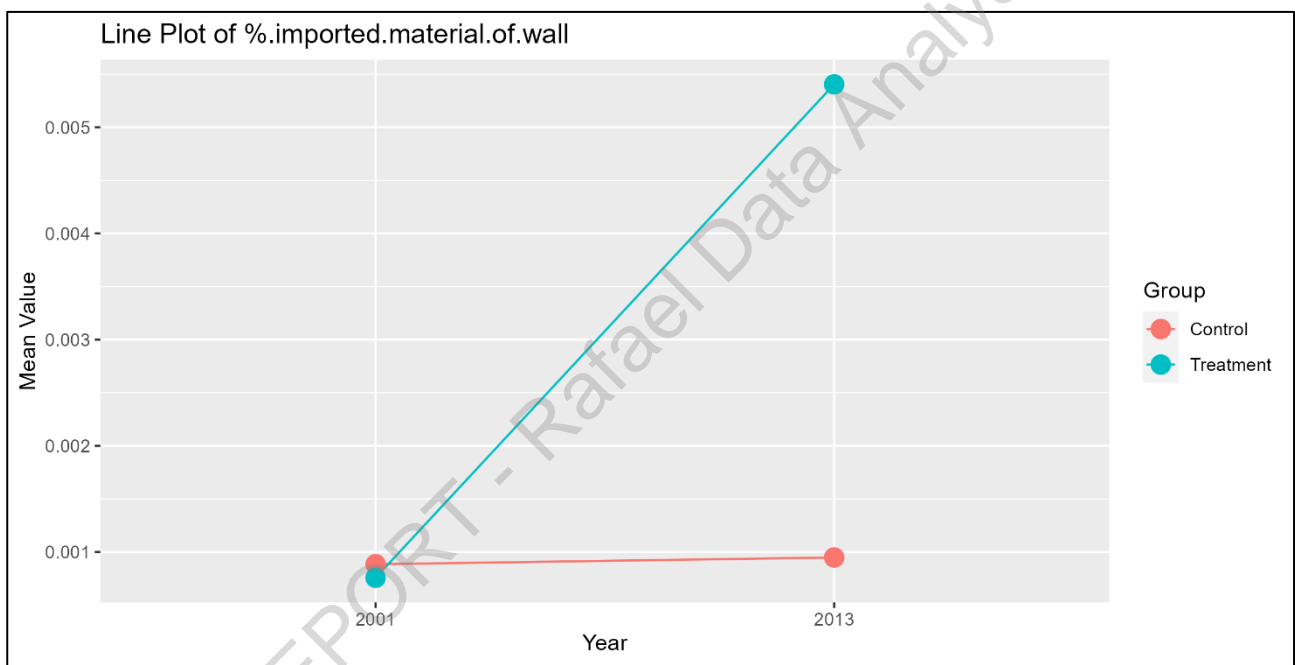
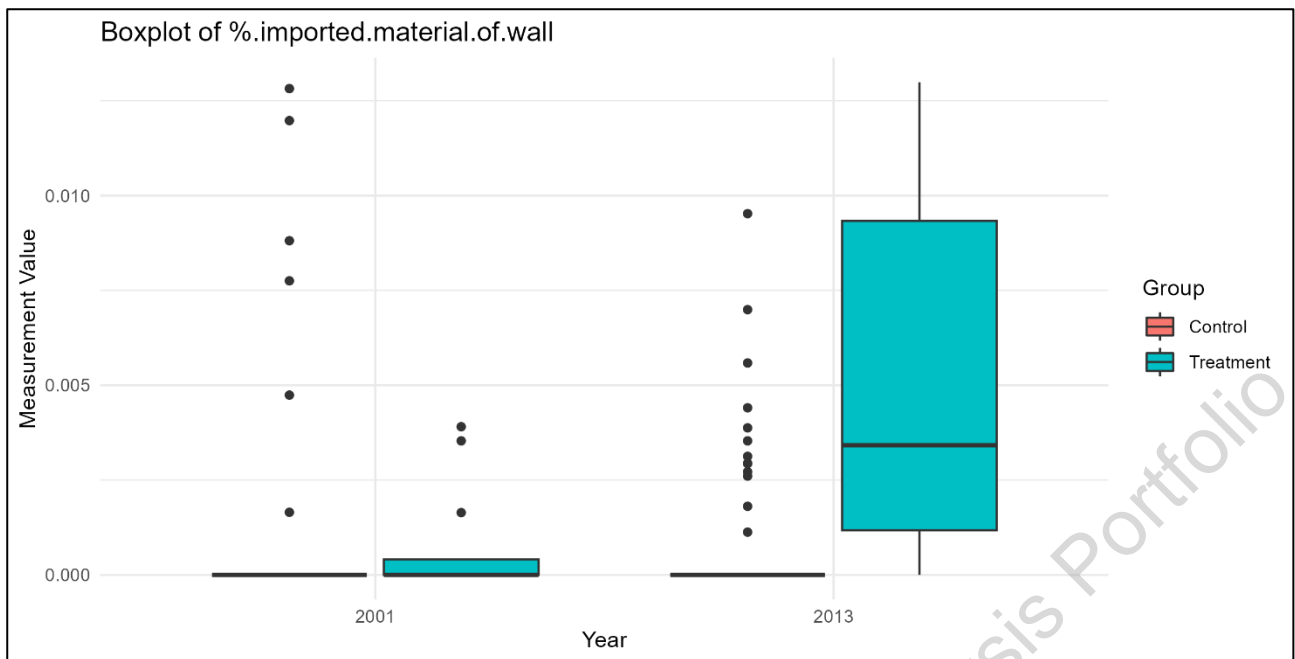
Conversely, "%.No.Land.Change.Job" shows a significant decline, especially for the Control group, which drops from 33.3% to 7.7%. The Treatment group also decreases but to a lesser extent, moving from 53.8% to 19.4%.

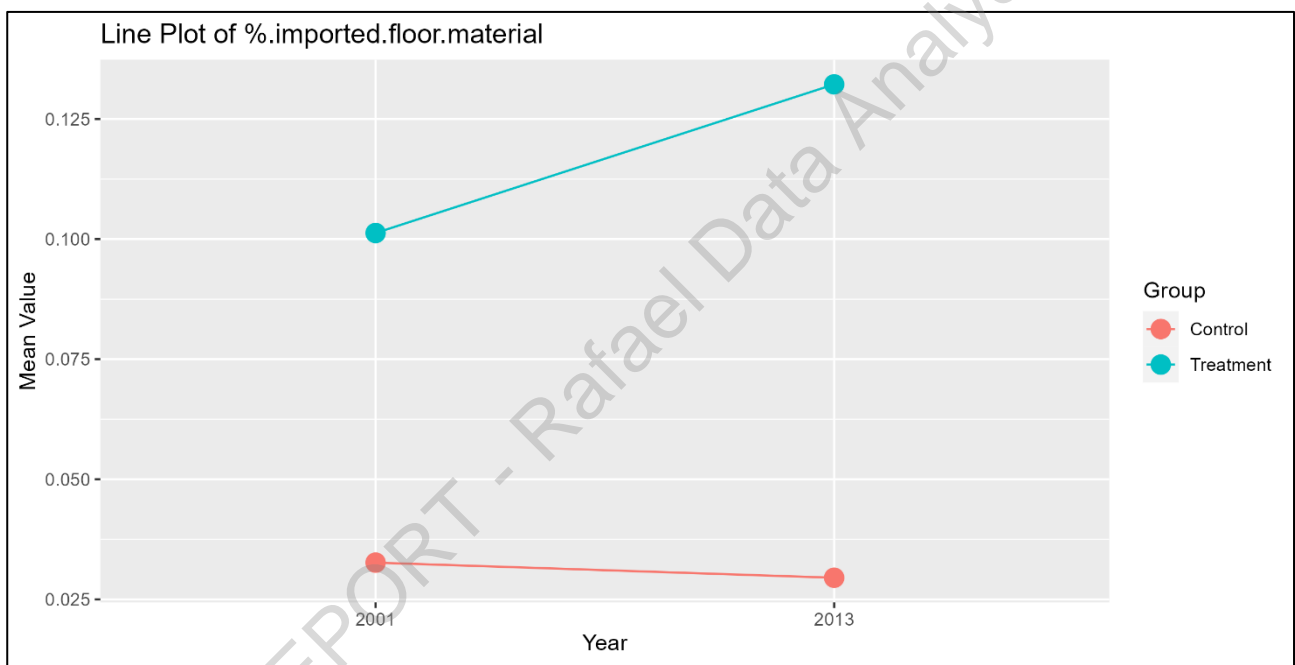
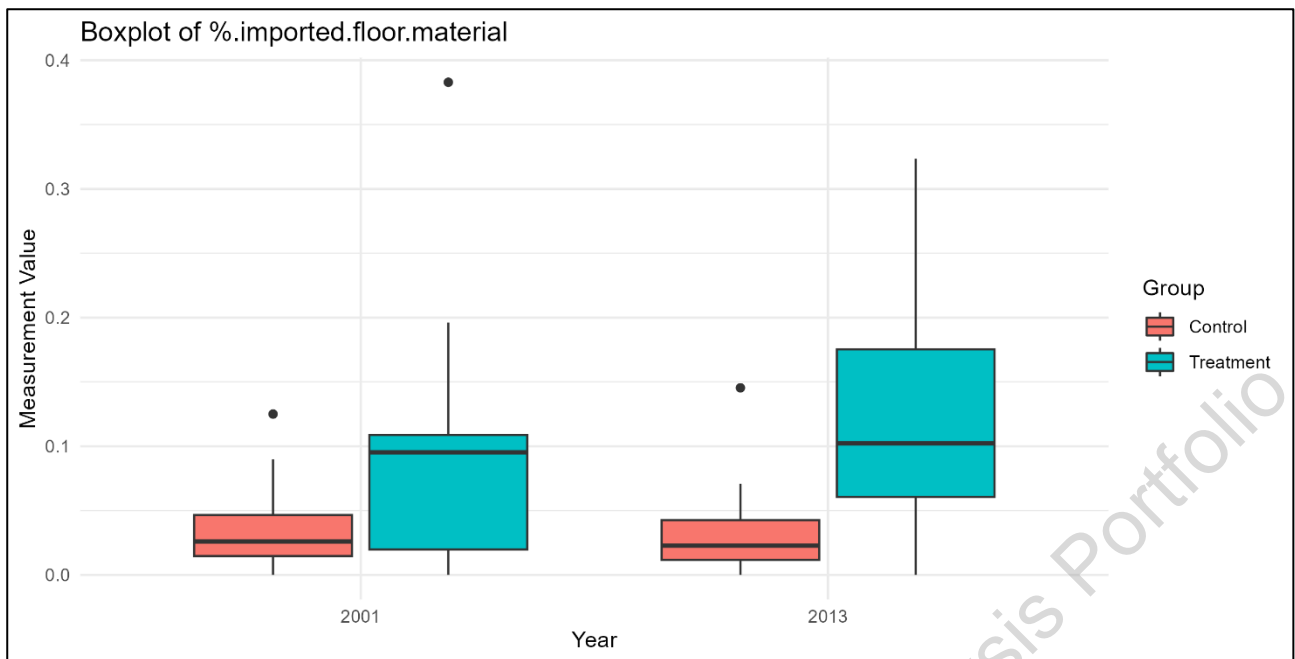
Overall, the Treatment group generally experiences more substantial changes than the Control group across these highlighted metrics. These differences could indicate the impact of an intervention or different conditions affecting the Treatment group more profoundly. The statistical analysis to come will test whether these differences are statistically significant.

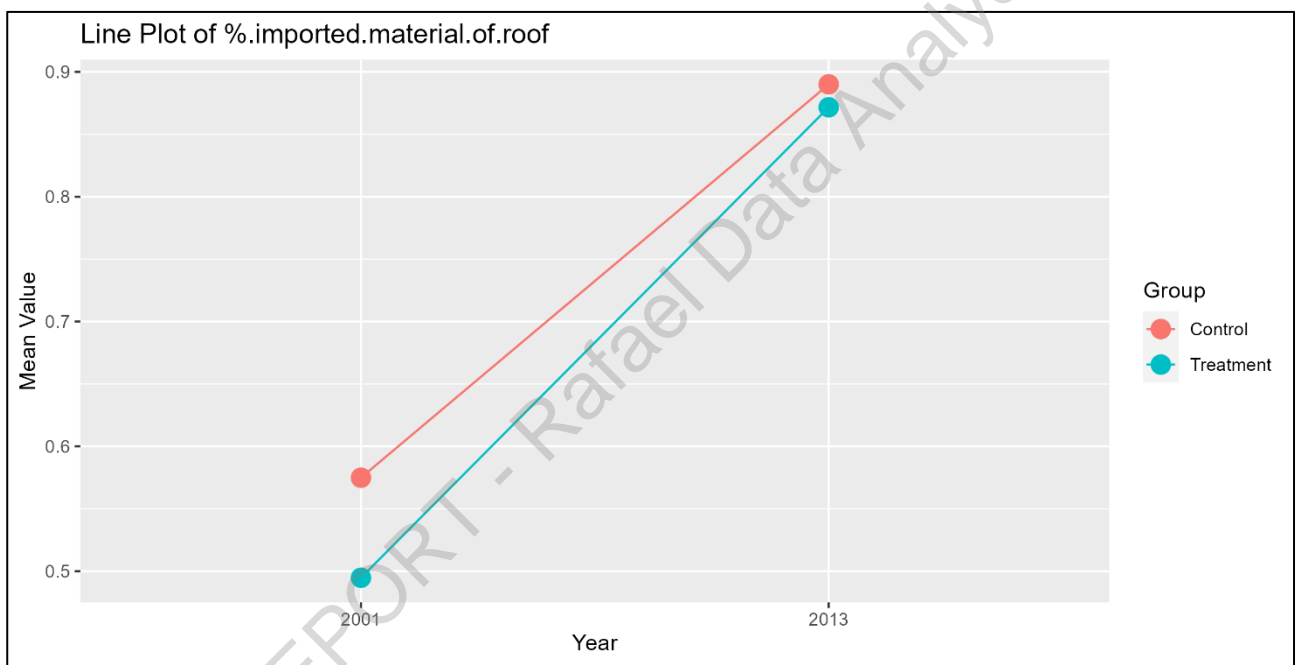
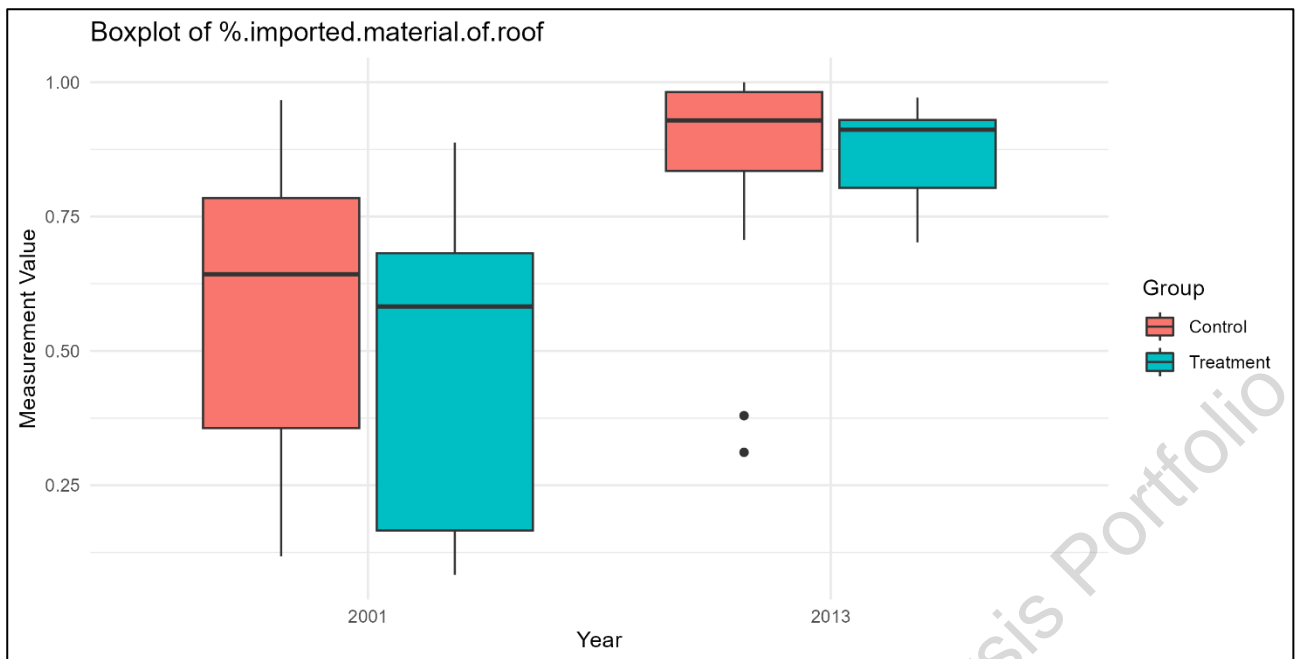
Variable	Group	2001	2013
%.imported.material.of.roof	Control	57.5%	89.0%
%.imported.material.of.roof	Treatment	49.5%	87.2%
%.imported.floor.material	Control	3.3%	2.9%
%.imported.floor.material	Treatment	10.1%	13.2%
%.imported.material.of.wall	Control	0.1%	0.1%
%.imported.material.of.wall	Treatment	0.1%	0.5%
%.subsistence.type.of.lighting	Control	97.8%	83.0%
%.subsistence.type.of.lighting	Treatment	96.1%	78.1%
%.Department.Born.not.from.Gracias.a.Dios	Control	1.5%	28.7%
%.Department.Born.not.from.Gracias.a.Dios	Treatment	11.1%	43.4%
%.of.Expensive.Type.of.Bathroom	Control	2.1%	7.9%
%.of.Expensive.Type.of.Bathroom	Treatment	7.9%	22.5%
%.Department.lived.5.years.ago.not.from.Gracias.a.Dios	Control	0.4%	25.3%
%.Department.lived.5.years.ago.not.from.Gracias.a.Dios	Treatment	3.6%	34.0%
%.No.Land.Change.Job	Control	33.3%	7.7%

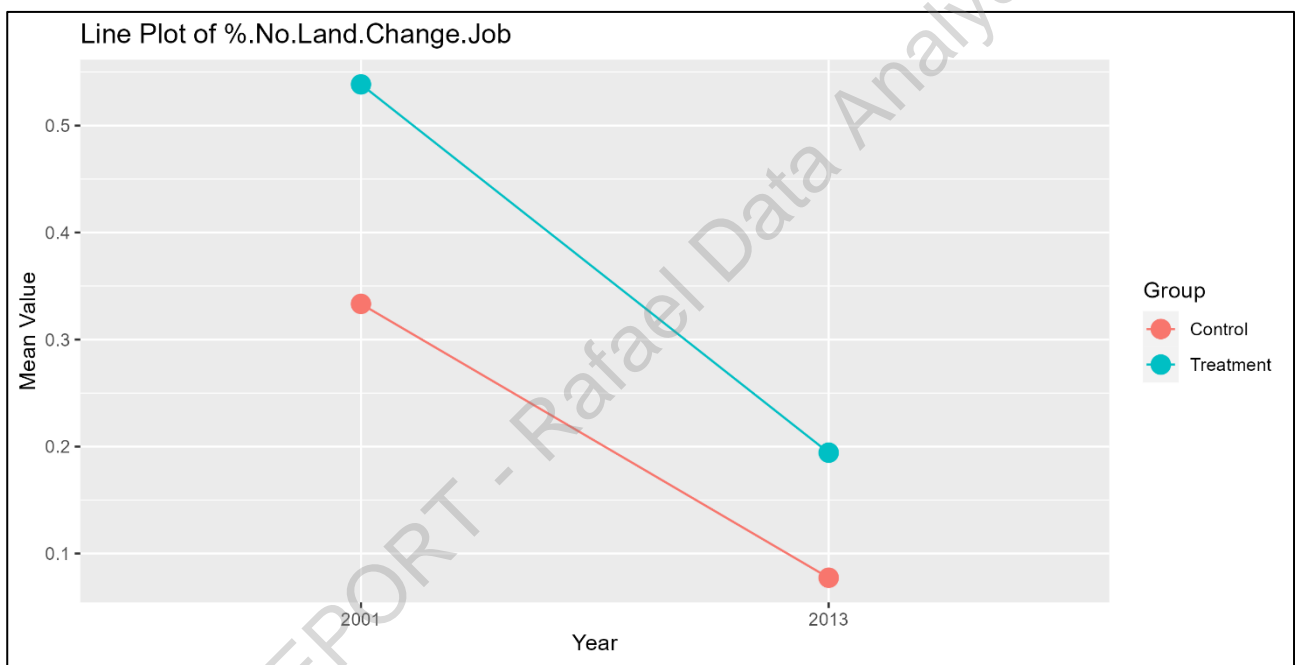
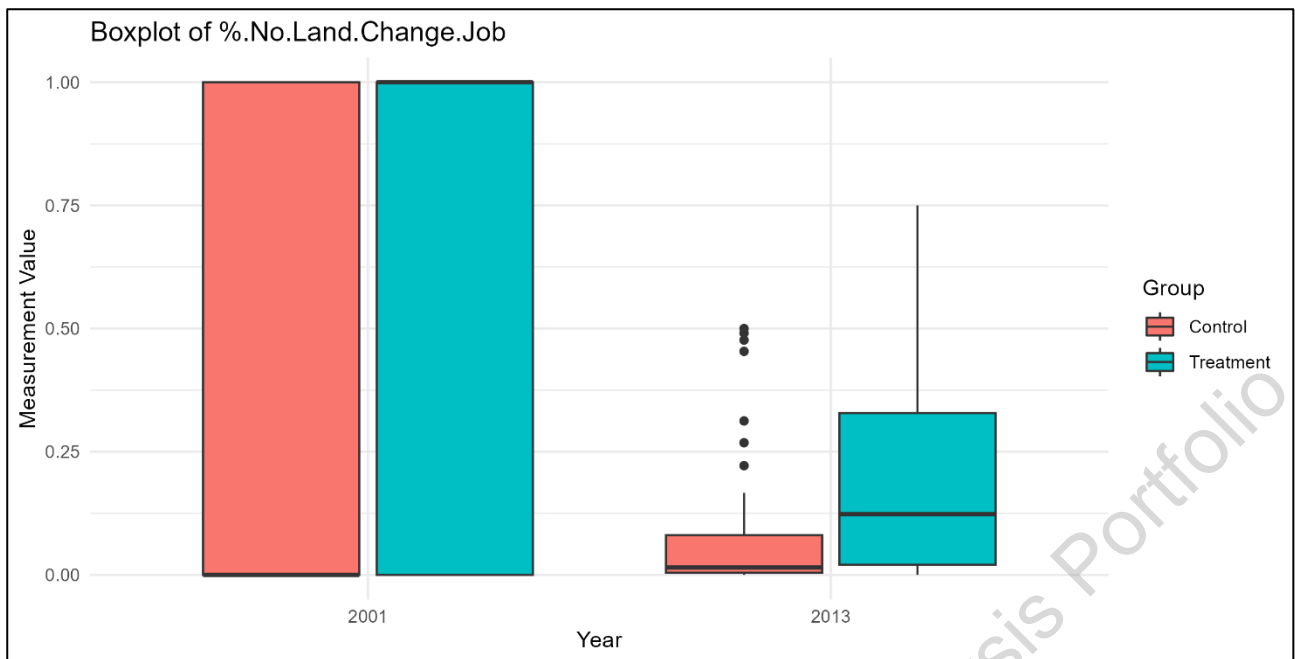
The graphs below illustrate these differences. Boxplots and line plots for the means were created.

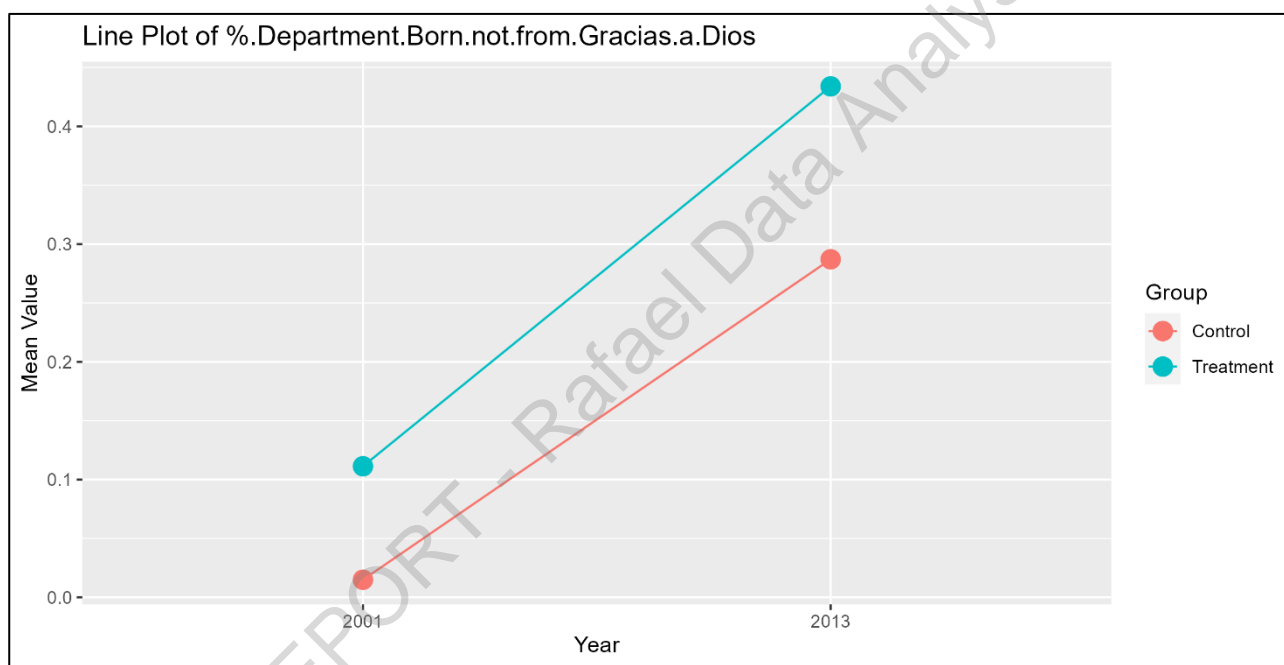
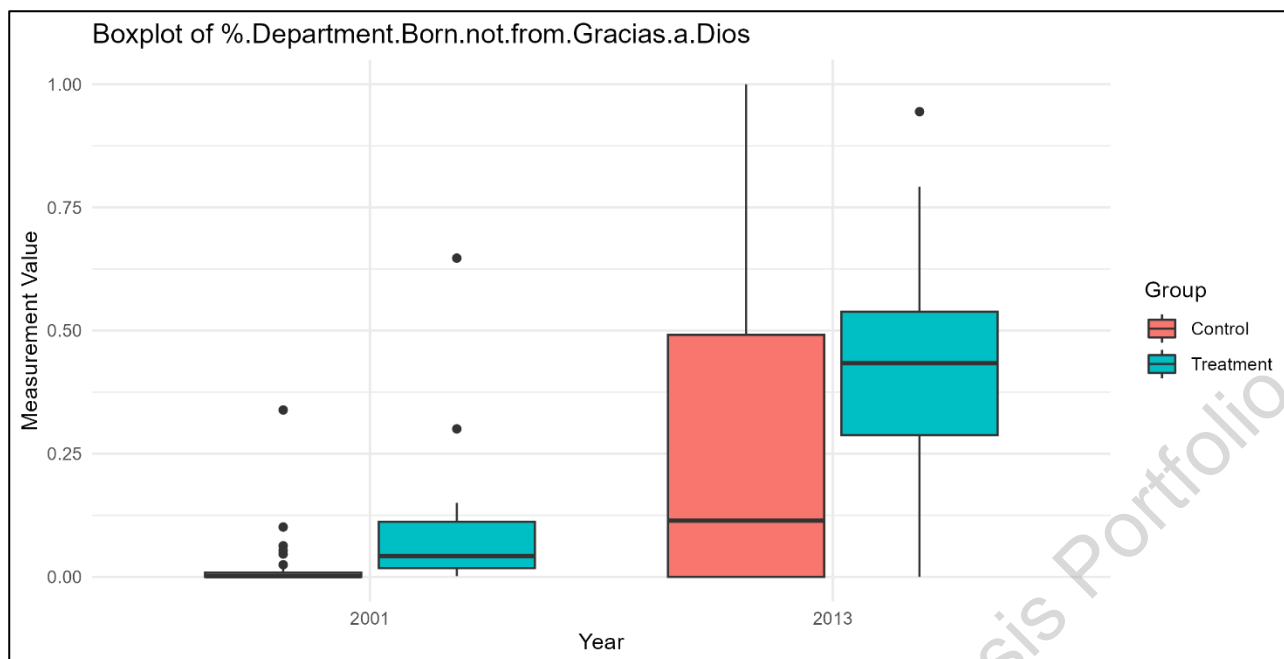


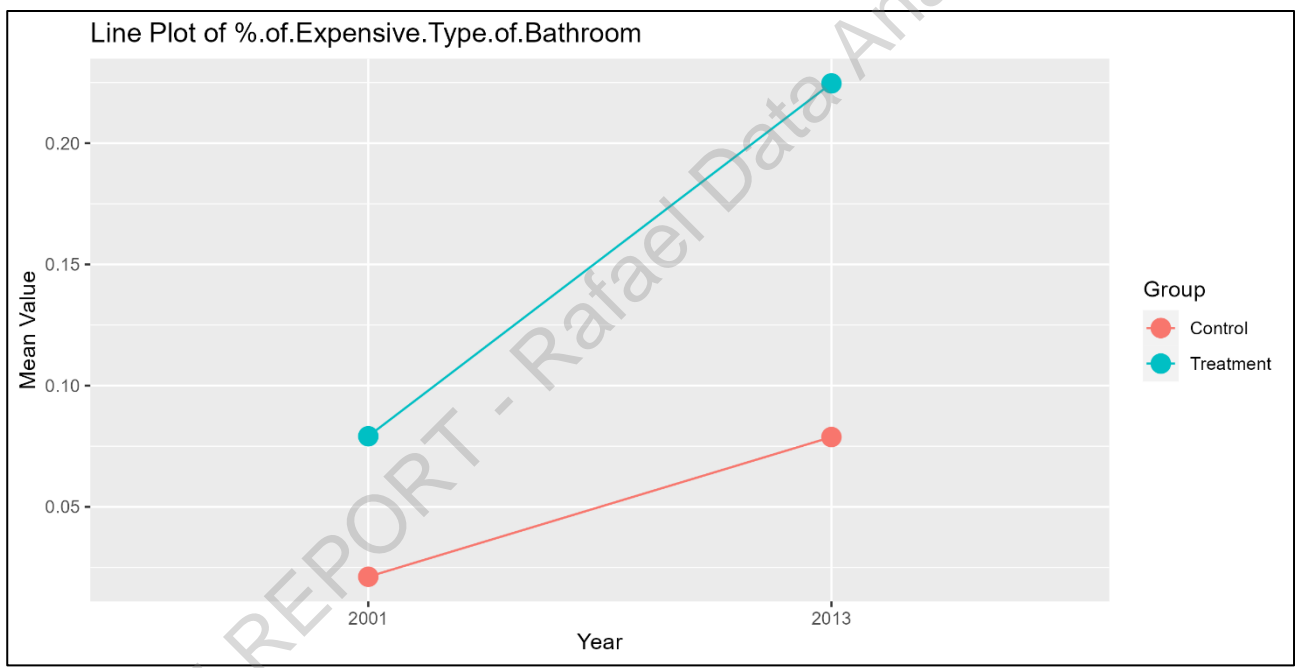
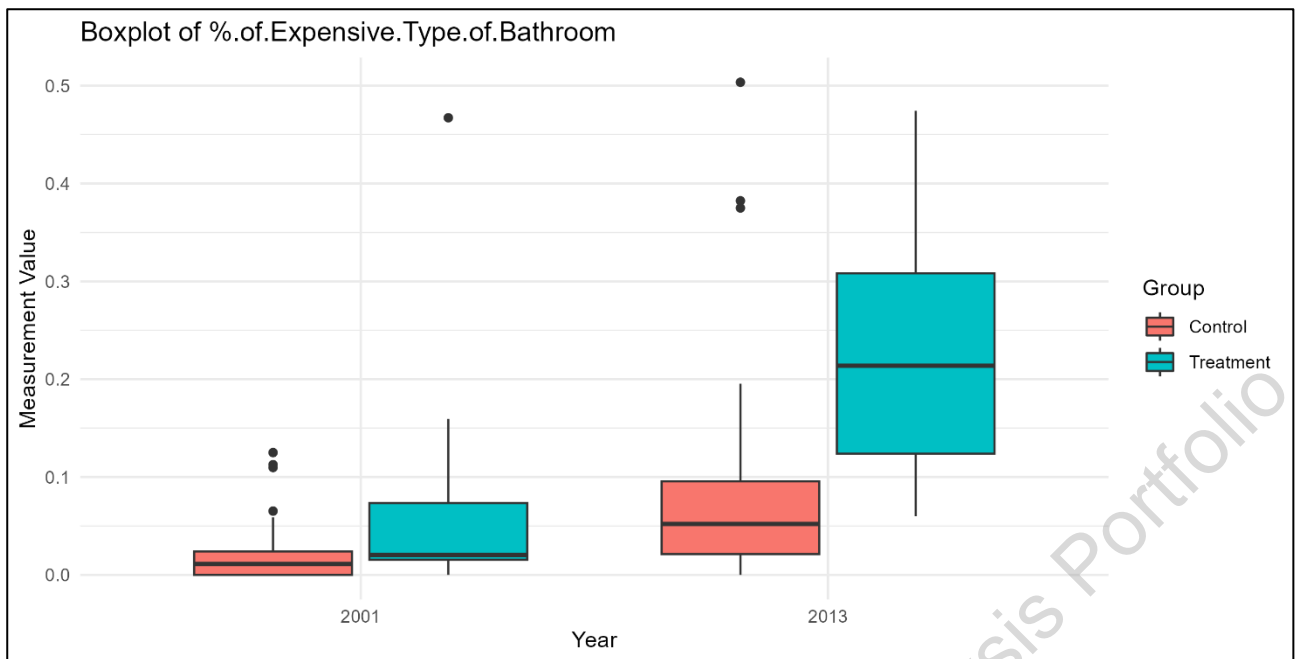


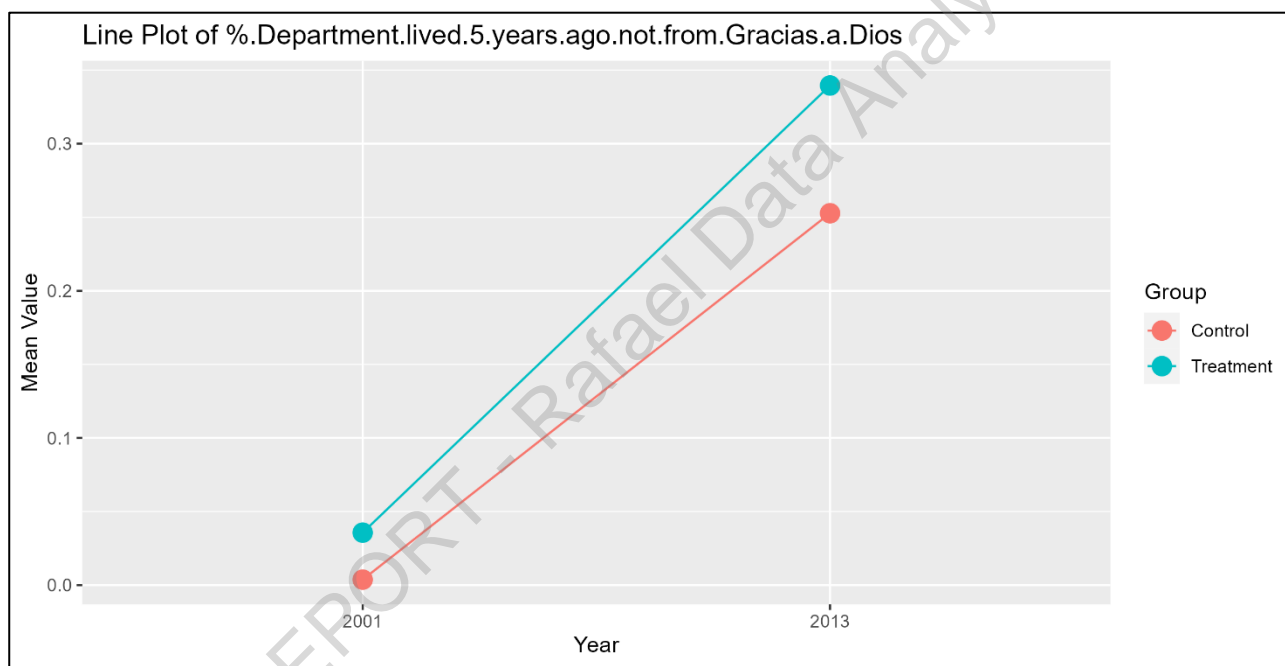
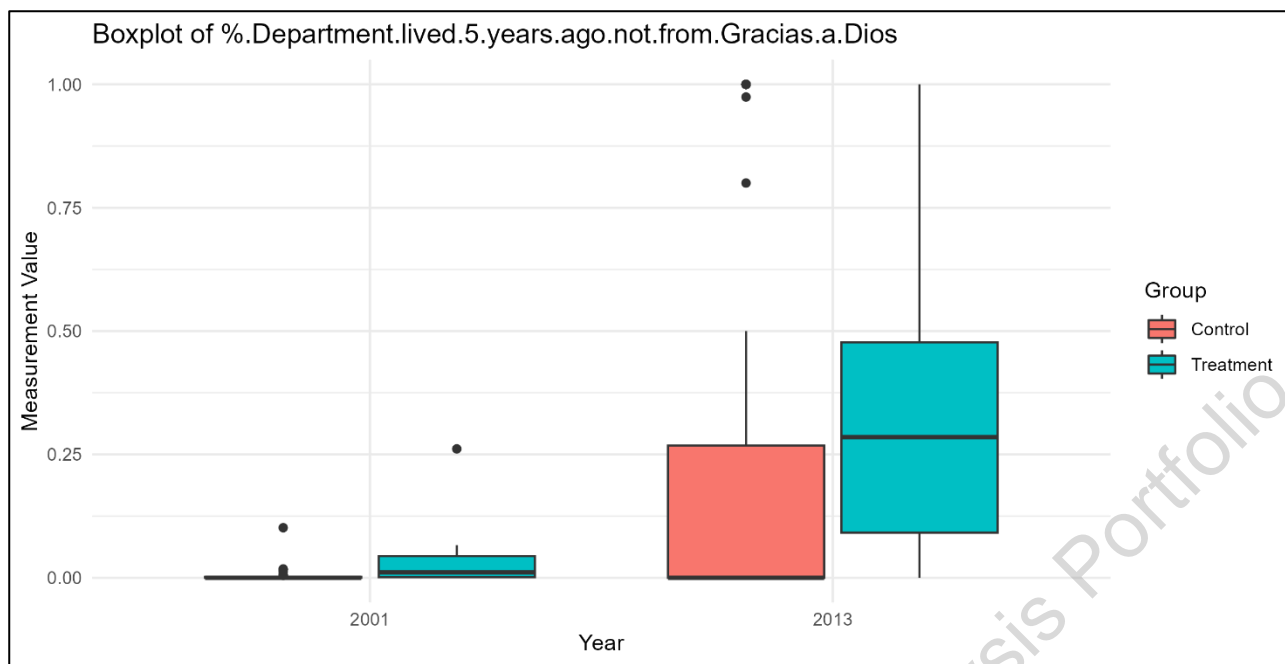












Mixed Models

A linear mixed-effects model analysis was conducted using the 'nlme' package in R. This package is commonly used for analyzing repeated measures or clustered data where both fixed and random effects are present. 'Departamento' and 'Municipio' were included as nested random effects, while 'Group' and 'Year' were included as fixed effects in the model.

the study assessed the effects of two fixed factors: 'Group' and 'Year.' The 'Group' factor compares two levels—Control and Treatment—while the 'Year' factor assesses temporal changes in the dependent variables.

The results are shown in the table below.

Dependent_Variable	Group_Pair	Effect	F_value	pValue
log_%.imported.material.of.roof	Control - Treatment	(Intercept)	99.787	0.000
log_%.imported.material.of.roof	Control - Treatment	Group	0.855	0.359
log_%.imported.material.of.roof	Control - Treatment	Year	65.956	0.000
log_%.imported.material.of.roof	Control - Treatment	Group:Year	1.940	0.169
log_%.imported.floor.material	Control - Treatment	(Intercept)	1535.740	0.000
log_%.imported.floor.material	Control - Treatment	Group	26.372	0.000
log_%.imported.floor.material	Control - Treatment	Year	0.006	0.941
log_%.imported.floor.material	Control - Treatment	Group:Year	7.468	0.009
log_%.imported.material.of.wall	Control - Treatment	(Intercept)	1536.087	0.000
log_%.imported.material.of.wall	Control - Treatment	Group	0.486	0.493
log_%.imported.material.of.wall	Control - Treatment	Year	6.394	0.065
log_%.imported.material.of.wall	Control - Treatment	Group:Year	3.156	0.150
log_%.Department.Born.not.from.Gracias.a.Dios	Control - Treatment	(Intercept)	266.531	0.000
log_%.Department.Born.not.from.Gracias.a.Dios	Control - Treatment	Group	5.084	0.028
log_%.Department.Born.not.from.Gracias.a.Dios	Control - Treatment	Year	201.016	0.000
log_%.Department.Born.not.from.Gracias.a.Dios	Control - Treatment	Group:Year	2.763	0.106
log_%.of.Expensive.Type.of.Bathroom	Control - Treatment	(Intercept)	994.620	0.000
log_%.of.Expensive.Type.of.Bathroom	Control - Treatment	Group	16.287	0.000
log_%.of.Expensive.Type.of.Bathroom	Control - Treatment	Year	41.620	0.000
log_%.of.Expensive.Type.of.Bathroom	Control - Treatment	Group:Year	1.262	0.268
log_%.Department.lived.	Control - Treatment	(Intercept)	237.117	0.000
5.years.ago.not.from.Gracias.a.Dios	Control - Treatment	Group	1.859	0.181
log_%.Department.lived.	Control - Treatment	Year	179.822	0.000
5.years.ago.not.from.Gracias.a.Dios	Control - Treatment	Group:Year	2.833	0.107
log_%.No.Land.Change.Job	Control - Treatment	(Intercept)	220.487	0.000
log_%.No.Land.Change.Job	Control - Treatment	Group	6.120	0.017
log_%.No.Land.Change.Job	Control - Treatment	Year	101.780	0.000
log_%.No.Land.Change.Job	Control - Treatment	Group:Year	2.207	0.152
%.subsistence.type.of.lighting	Control - Treatment	(Intercept)	10600.39	0.000
%.subsistence.type.of.lighting	Control - Treatment	Group	1	0.013
%.subsistence.type.of.lighting	Control - Treatment	Year	6.476	0.000
%.subsistence.type.of.lighting	Control - Treatment	Group:Year	55.691	0.000
%.subsistence.type.of.lighting	Control - Treatment	Group:Year	3.051	0.086

There were several significant differences across Years or across Groups (main effects). The following paragraphs explore the interaction effects.

For "log_%.imported.floor.material," the interaction effect between Group and Year is significant (F-value: 7.468, p-value: 0.009). This indicates that the change in the percentage of imported floor material over the years differs significantly between the Control and Treatment groups.

For "log_%.Department.Born.not.from.Gracias.a.Dios," although the interaction effect is not significant (p-value: 0.106), the F-value of 2.763 suggests that it might be worth investigating further. The change in the percentage of people born outside of Gracias a Dios over the years may not be uniform across the two groups. The same applies for 'log_%.Department.lived.5.years.ago.not.from.Gracias.a.Dios'.

The metric "%.subsistence.type.of.lighting" also shows a borderline interaction effect (F-value: 3.051, p-value: 0.086). Although not statistically significant at the 0.05 level, the result may be practically significant, warranting further examination.

For other variables like "log_%.imported.material.of.roof," "log_%.imported.material.of.wall," "log_%.of.Expensive.Type.of.Bathroom," and "log_%.No.Land.Change.Job," the interaction effects are not significant. This suggests that the changes in these metrics over time are consistent between the Control and Treatment groups.

Overall, the interaction effects are generally not significant, implying that for most metrics, the trends over time are similar for both Control and Treatment groups. However, the variable "log_%.imported.floor.material" stands out as an exception, suggesting different trends in the two groups over time.