Analysis Report

This report is structured as follows.

Contents

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.
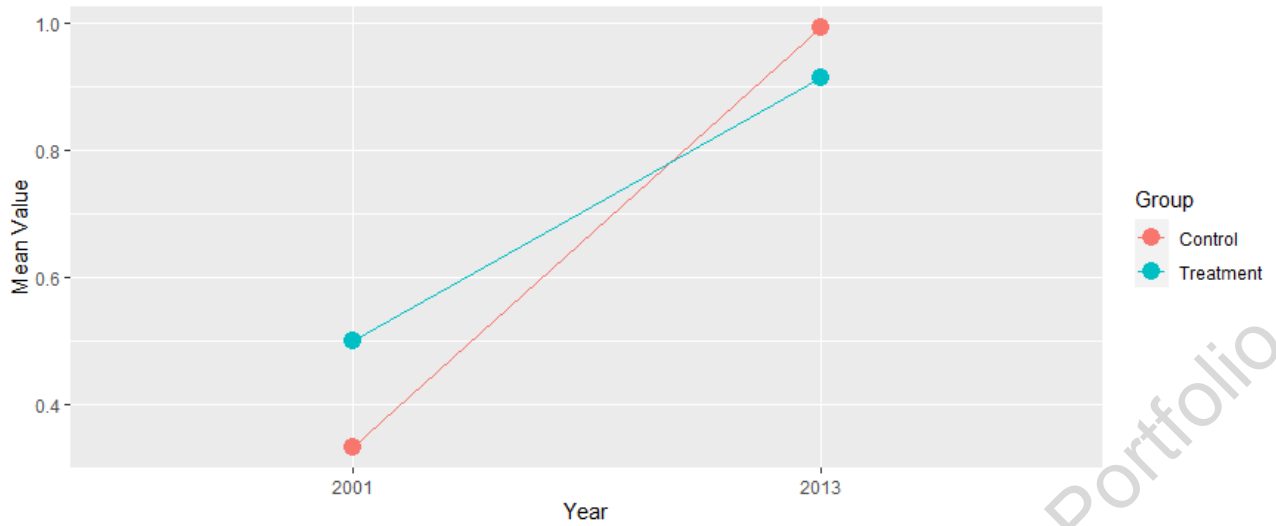
Descriptive Statistics

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.

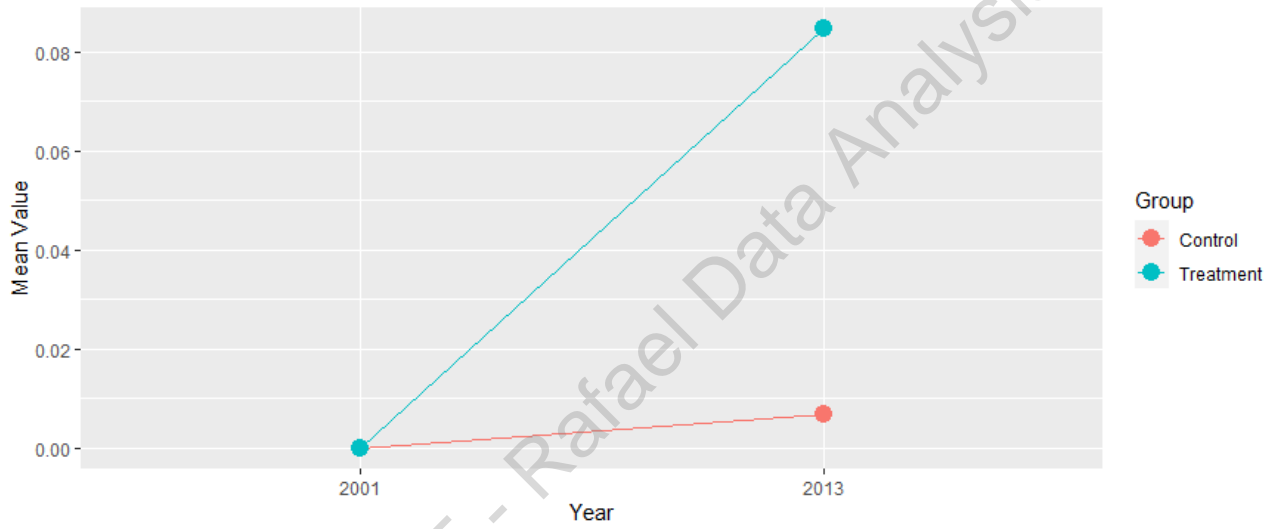| Variable	Group	2001	2013
%_Land_Based_Job__Cattle_Rancher__Agriculture__Forestry_	Control	0.000	0.923
%_Land_Based_Job__Cattle_Rancher__Agriculture__Forestry_	Treatment	0.000	0.806
%_No_Land_Based_Job__Construction_workers__Underwater_Divers__Police_	Control	0.333	0.077
%_No_Land_Based_Job__Construction_workers__Underwater_Divers__Police_	Treatment	0.500	0.194
%_Jobs_in_Cattle	Control	0.000	0.007
%_Jobs_in_Cattle	Treatment	0.000	0.085
%_Non-Cattle_Jobs	Control	0.333	0.993
%_Non-Cattle_Jobs	Treatment	0.500	0.915

The numbers illustrate an increase on the percentage of non-cattle jobs for the control and treatment groups between 2011 and 2013. On the other hand, the percentage of No-Land based jobs decreased on the same period.
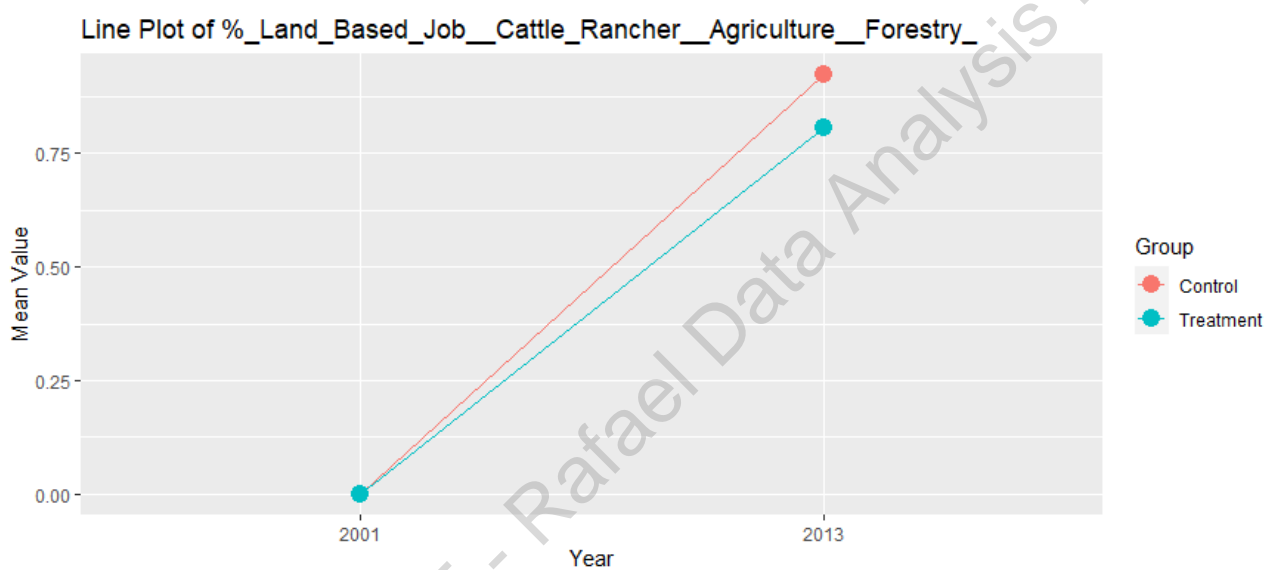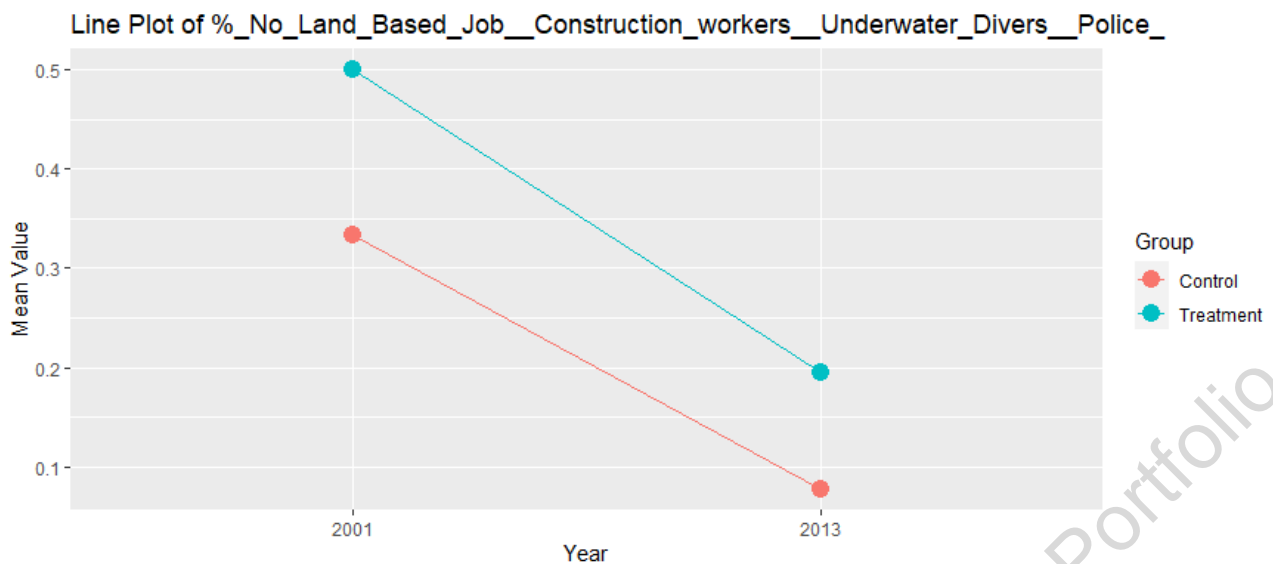
The graphs below illustrate these differences. Boxplots and line plots for the means were created.

Line Plot of %_Non-Cattle_Jobs

Line Plot of %_Jobs_in_Cattle

Line Plot of %_No_Land_Based_Job__Construction_workers__Underwater_Divers__Police_

Line Plot of %_Land_Based_Job__Cattle_Rancher__Agriculture__Forestry_

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	Num DF	Den DF	F	P
% No Land Based Job Construction workers Underwater Divers Police	Control - Treatment	(Intercept)	1	53	237.680	0.000
% No Land Based Job Construction workers Underwater Divers Police	Control - Treatment	Group	1	53	7.056	0.010
% No Land Based Job Construction workers Underwater Divers Police	Control - Treatment	Year	1	23	105.219	0.000
% No Land Based Job Construction workers Underwater Divers Police	Control - Treatment	Group: Year	1	23	2.356	0.138
% Non Cattle Jobs	Control - Treatment	(Intercept)	1	66	3.427	0.069
% Non Cattle Jobs	Control - Treatment	Group	1	66	7.921	0.006
% Non Cattle Jobs	Control - Treatment	Year	1	23	8.088	0.009
% Non Cattle Jobs	Control - Treatment	Group: Year	1	23	14.240	0.001

The group effect shows a statistically significant difference (F = 7.056, p = 0.010) between the control and treatment groups for the percentage of No Land-Based job. This indicates that belonging to either group significantly impacts the percentage of no land-based jobs.

The year effect is also highly significant (F = 105.219, p < 0.001), implying that the year in which the data was collected significantly affects the percentage of no land-based jobs.

However, the interaction effect between group and year is not statistically significant (F = 2.356, p = 0.138), suggesting that the difference between the control and treatment groups over the years is not significant for this variable.
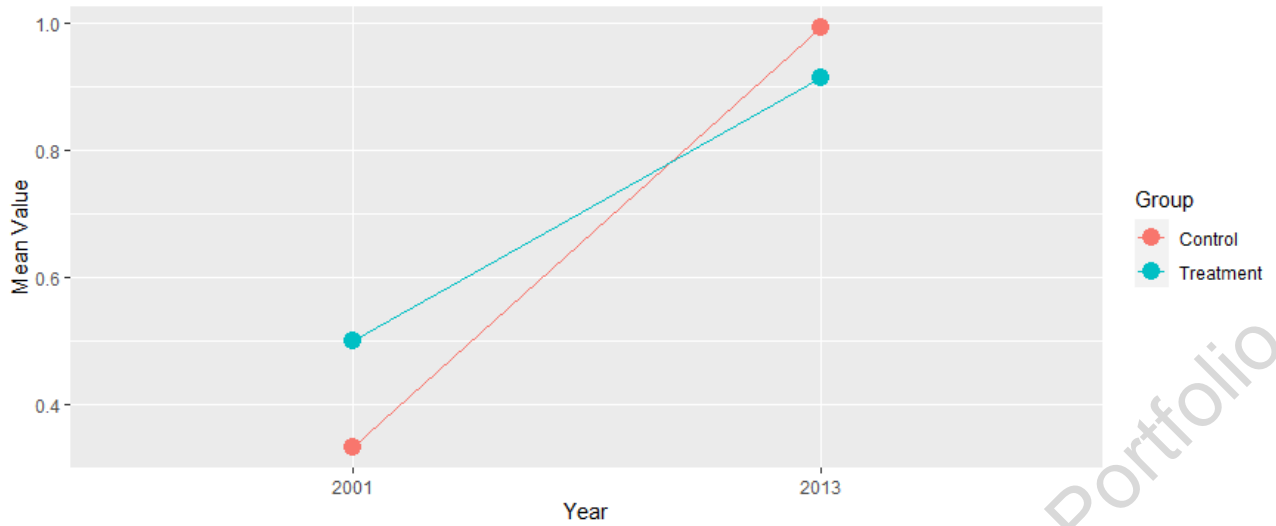
For "% Non Cattle Jobs". the group effect is significant (F = 7.921, p = 0.006), demonstrating that the control and treatment groups differ significantly in their percentage of non-cattle jobs.

The year effect is also significant (F = 8.088, p = 0.009), showing that the year of data collection has a significant impact on the percentage of non-cattle jobs.

Notably, the interaction effect between group and year is highly significant (F = 14.240, p = 0.001). This indicates that the difference between control and treatment groups in terms of non-cattle jobs changes significantly over the years.

When the interaction plot is analyzed (pasted again below), we can see that the pattern indeed diverges, with the increase for the control group being of a higher magnitude compared to the control group from 2001 to 2013.

Line Plot of %_Non-Cattle_Jobs