Final assignment: Project report

Yuka Tatsumi

2022/01/17

Introduction

Climate change is one of the biggest concerns in the world and is very urgent issues to be solved. CO2emission is one of the indicators for the result of actions for climate changes.

In this report, the following aims to be investigated and answered.

Question 1: Does GDP per capita and % of GDP by sectors(industry, agriculture, services) have a relation with CO2 emission per person by countries in 2018?

Question 2: Are there any differences in CO2 emissions per person between Sweden and Denmark between the late 1900s and early 2000s?

Question 3: Are there any differences in CO2 emissions per person among Nordic countries between the late 1900s and early 2000s? Are there any interactions between countries and time period(the late 1900s and early 2000s) that have an effect on CO2 emissions per person?

Questions 4: Are there any associations in attitude/awareness toward environment (intentional purchase of items with eco-friendly/less packaging) and working styles (work from home and people who work away from home) as well as age group, Generation Z, Young Millenials (23-26 years old), Core Millenials(27-32 years old), Mature Millenials(33-36 years old), Generation X, Baby Boomers ?

Method

The data is extracted from gapminder web site to investigate CO2 emissions-related questions (Question 1, 2, 3). Linear regression is used to analyze the relationship of GDP and % of GDP by sector with CO2 emission per country in 2008 because linear regression is appropriate when relationships between more than one variable are investigated(Salkind and Shaw, 2020). In order to investigate differences in CO2emission per person between Sweden and Denmark (Question2) from the late 1900s until early 2000s, t-test or non-parametric alternatives (Wilcoxon's rank-sum test) is conducted after checking if the assumption of normal distribution is met. Two-way anova is used in order to investigate differences in CO2emissions (tonnes per person) among Nordic countries and between the late 1900s and early 2000s. Two way ANOVA is conducted after dividing year in the dataset into two categories, "Late1900s" and "Early2000s" in order to analyze the difference among Nordic countries as well as differences among year category("Late1900s" and "Early2000s") It also aims to analyse if there are any interactions between countries and year_category. For the questions relating to attitude/awareness toward the environment between the current remote workers and non-remote workers as well as age groups(Question 4), the data is based on the result of PwC's June 2021 Global Consumer Insights Pulse Survey and Chi-square test is conducted to investigate the associations.

Result

Question 1: Does GDP per capita and % of GDP by sectors have a relation with CO2 emission per person by countries in 2018?

The data of 178 countries in 2018 are analyzed with no missing data.

Table 1 Descriptive Statistics

Variable	N	Missing	Mean	Variance	SD	Min	Max	Median (Q1,Q3)
Country	178	0				0.02	38	2.54 (0.77, 6.01)
CO2emission	178	0	4.54	32.53	5.7	762	114000	12,800.00 (4,887.50, 28,600.00)
GDP	178	0	20297.84	427871244.63	20685.05	4.87	63.2	24.75 (18.27, 31.48)
Industry_GDP	178	0	25.85	125.47	11.2	0.03	58.9	6.75 (2.35, 16.00)
Agriculture_GDP	178	0	10.43	105.73	10.28	16.8	88.7	54.80 (48.82, 63.22)
Services_GDP	178	0	55.33	132.94	11.53	0.02	38	2.54 (0.77, 6.01)

Scatterplots below (Figure 1) shows that GDP per capita, industry(% of GDP) has a positive correlation with CO2 emissions per person. Agriculture (% of GDP) has a negative correlation with CO2 emissions per person. Services(% of GDP) seems to show a weak correlation with CO2 emissions per person. Up to 50% of Services(% of GDP), Services(% of GDP) seems to have a relatively positive correlation with CO2emissions per person while it shows relatively negative correlation at about 70%.

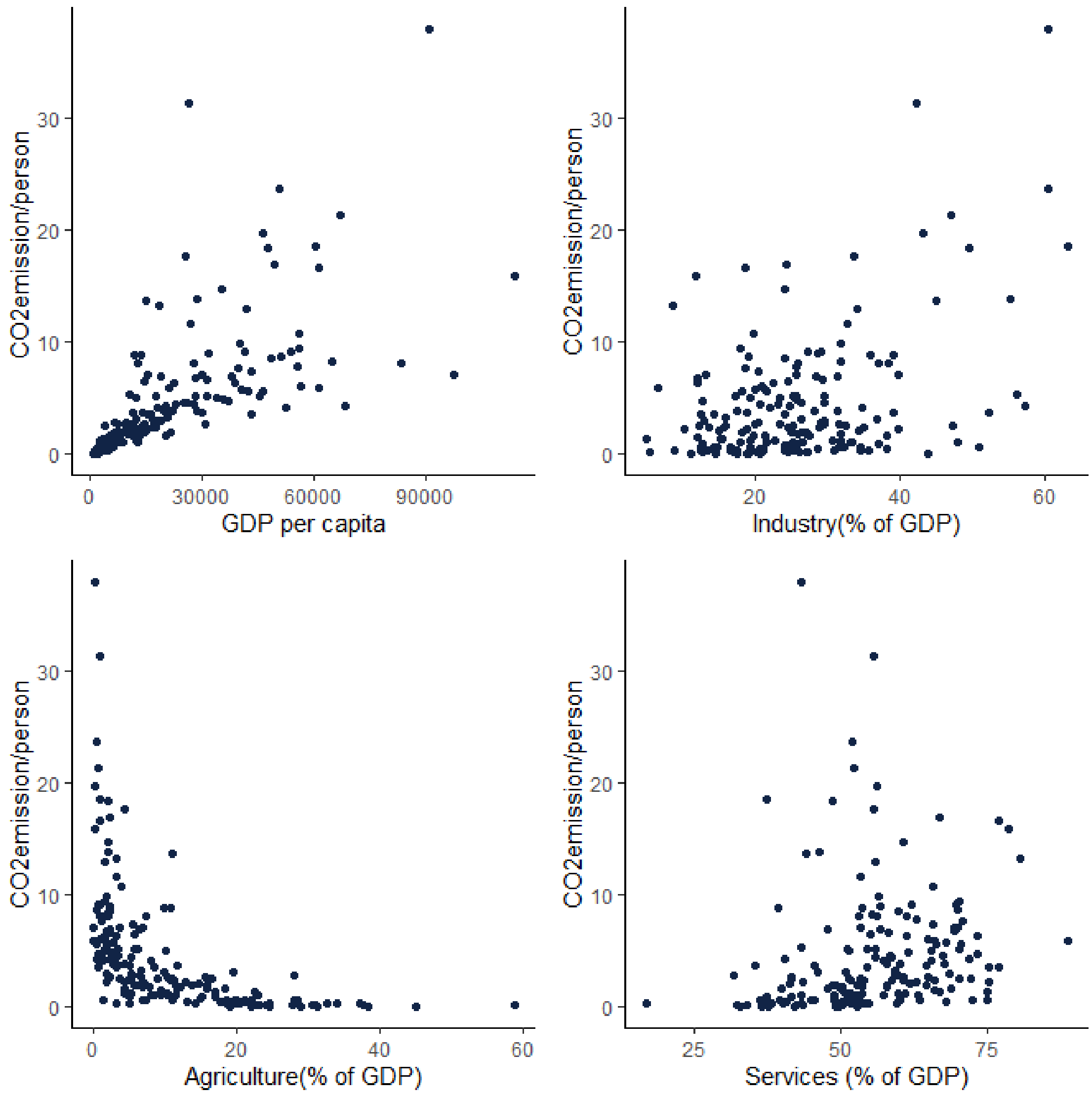

Figure 1 Scatterplots (CO2 emission and GDP, % of GDP by sectors)

Correlation matrix (Figure 2) also indicates that GDP and industry(% of GDP) have a positive correlation with CO2 emissions per person. Agricultural(% of GDP) has a negative correlation with CO2 emissions per person. Services(% of GDP) seems to show a weak correlation with CO2 emissions per person.

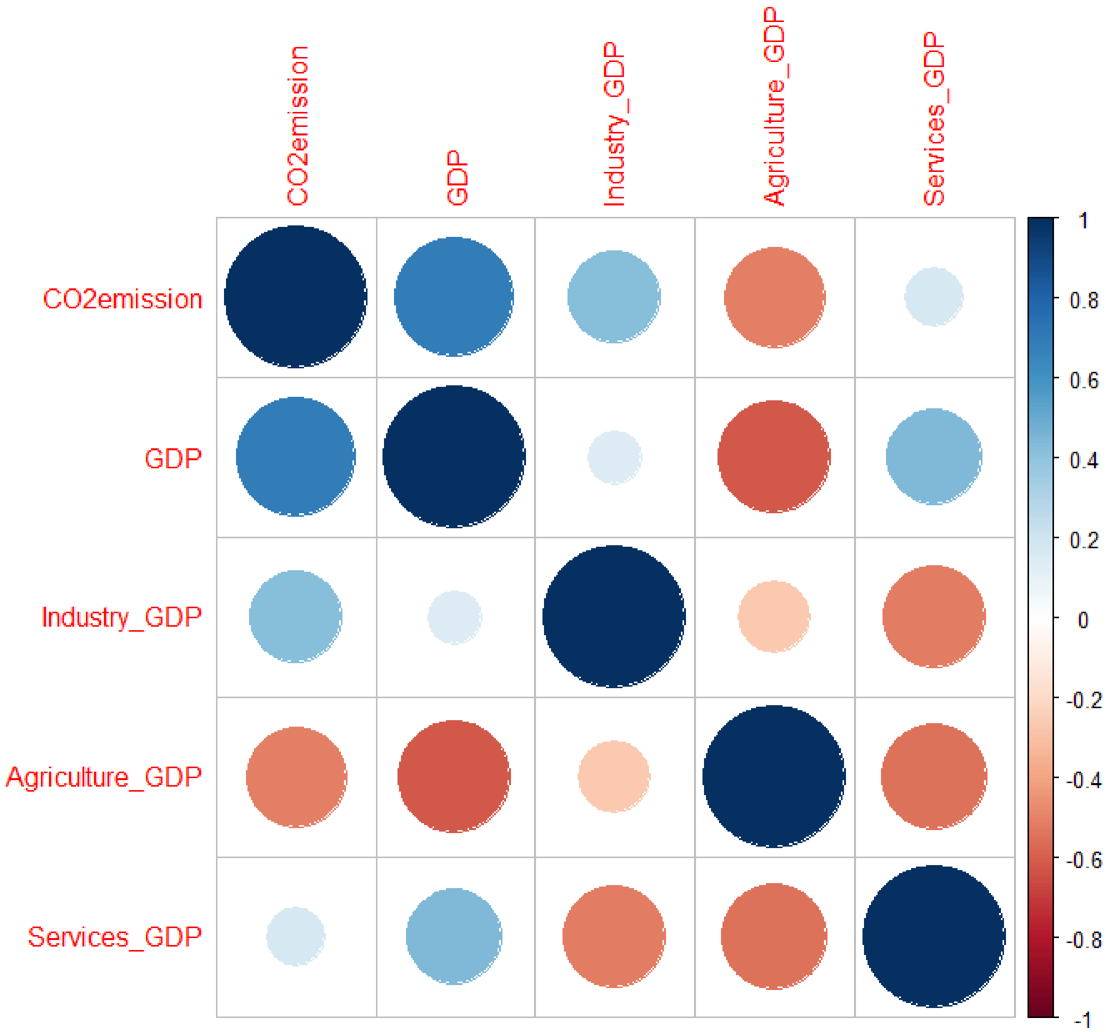

Figure 2 Correlation Plot (CO2 emissions and GDP or percentage of GDP by Sectors)

Linear regression model with all variables (GDP per capitals, industry(% of GDP), agriculture(% of GDP), services(% of GDP)) in Table 2 below indicates low multiple R-squared: 0.5966 and Adjusted R-squared: 0.5873. (F-statistic: 63.98 on 173 and 4 DF, p-value<0.0001).

Table 2 Linear Regression Analysis: Coefficeints Table
CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	-10.537	4.153	-2.537	0.0121	*
GDP	0.000	0.000	9.435	0.0000	***

	Estimate	Standard Error	t value	Pr(>|t|)	
Industry_GDP	0.233	0.045	5.246	0.0000	***
Agriculture_GDP	0.051	0.050	1.007	0.3151	
Services_GDP	0.093	0.051	1.846	0.0666	.

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 3.664 on 173 degrees of freedom

Multiple R-squared: 0.5966, Adjusted R-squared: 0.5873

F-statistic: 63.98 on 173 and 4 DF, p-value: 0.0000

Basic diagnostic plots: CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

The graph for "Residuals vs Fitted" in the basic diagnostic plots(Figure 3) below shows a little gentle curve. In "Normal Q-Q plots", there are some deviations from reference line, which means the residual are not normally distributed. "Scale-location plots" do not show a horizontal line and plots spread unequally, which indicates heteroscedasticity. "Residuals vs Leverage" chart shows that there are outliers that may affect the model as two plots are outside Cook's distance line (0.5).

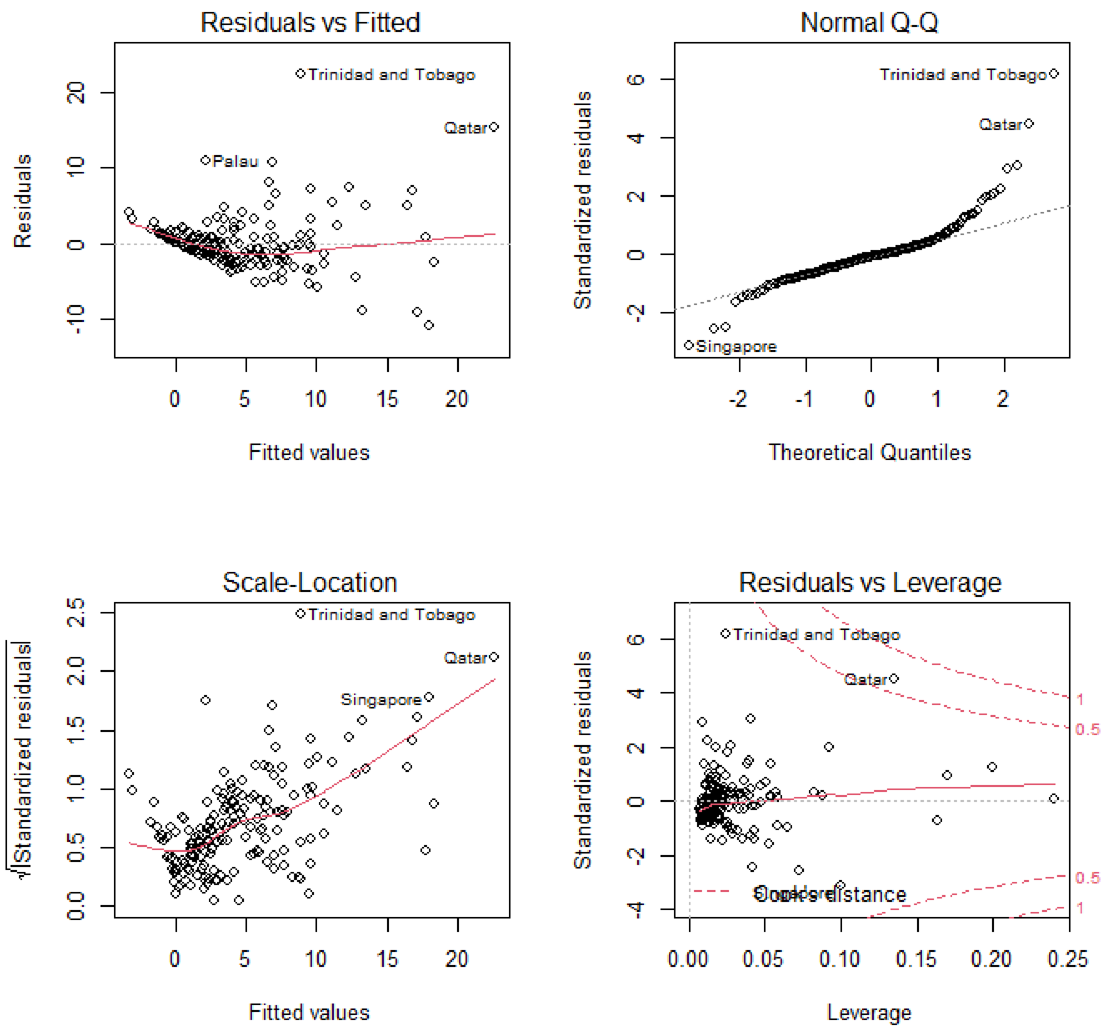

Figure 3 Basic diagnostic plots: CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

In order to meet assumptions of normal distribution and homoscedasticity, log transformation is used.

Basic diagnostic plots:log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

Graph for "Residuals vs Fitted" in the basic diagnostic plots(Figure 4) below shows roughly straight horizontal line and the spread roughly evenly without strong specific pattern. In "Normal Q-Q shows plots", there are less deviations from reference line, which means the residual are roughly normally distributed. "Scale-location plots" show a roughly horizontal line and plots spread equally, which indicates homoscedasticity. "Residuals vs Leverage" chart shows that there are no outliers that may affect the model as all plots are within Cook's distance line. These indicate that this model with log transformation meets the assumption better than the first model without log transformation.

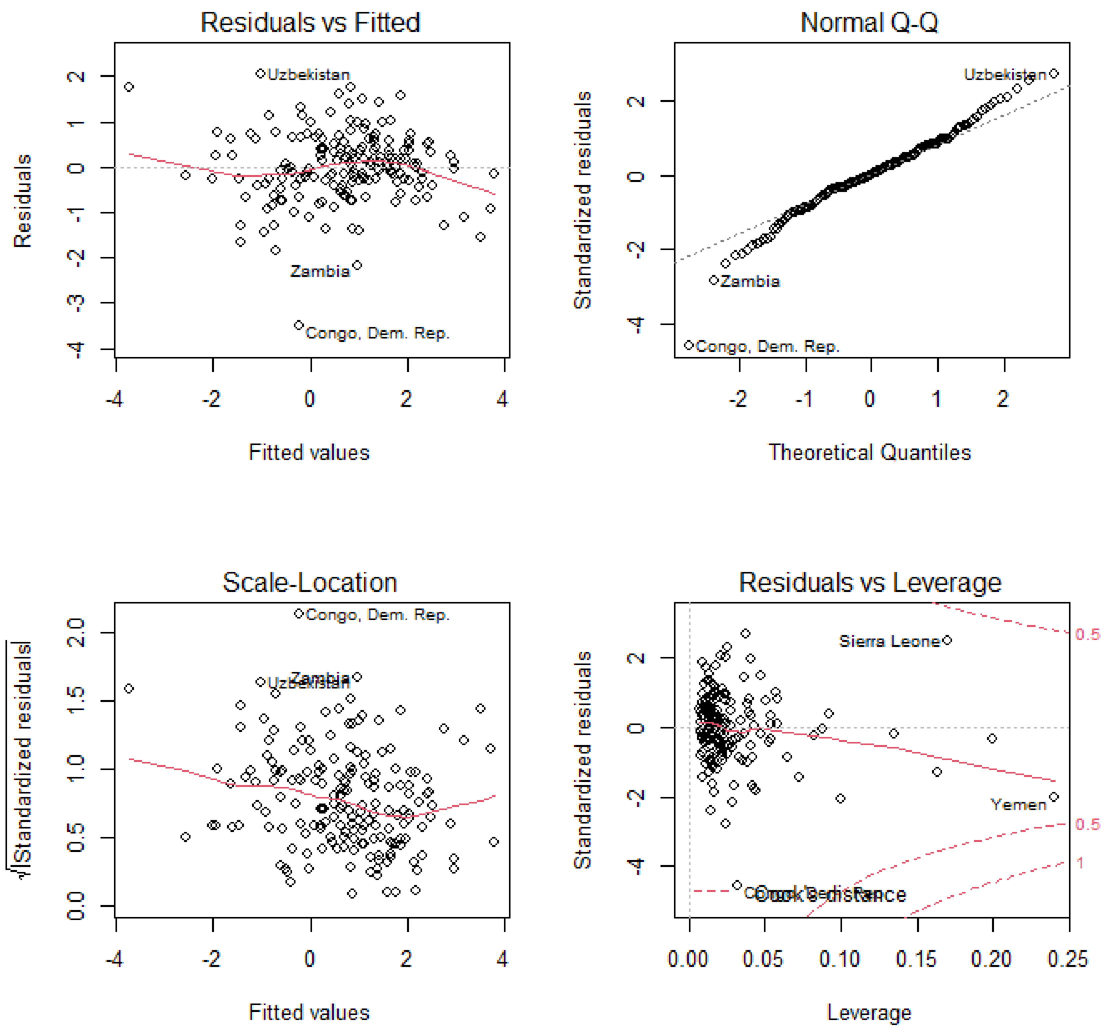

Figure 4 Basic diagnostic plots(after log transformation):log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

According to the result of the linear regression model after log 10 transformation below(F-statistic: 111.7, degree of freedom 173 and 4, p-value<0.001), Adjusted R-squared has improved from 0.5873 to 0.7143, which also indicates that this model with log transformation is better fit than the first model without log transformation.

The coefficients table below (Table 3) indicates that GDP, Industry(% of GDP) and Agriculture(% of GDP) have a statistically significant influence on log transformed CO2 emissions. (GDP:t-value 6.652, p-value<0.001, Industry: t-value 2.538. p-value 0.012, Agriculture: t-value -6.149, p-value<0.001) However, Services(% of GDP) does not have a significant influence on log transformed co2 emissions (t-value 1.320, p-value 0.1884). The coefficients for GDP and Industry(% of GDP) are positive, which means that they have a positive impact on CO2 emissions, which in turn means higher GDP and Industry(% of GDP) relates to higher CO2 emissions. On the other hand, the coefficients for Agriculture(%

of GDP) is negative which implies that it has a negative impact on CO2 emissions, which means that higher % of Agriculture (% of GDP) relates to lower CO2 emissions.

Table 3 Linear Regression Analysis(long transformation) Coefficeints Table :log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	-0.499	0.881	-0.567	0.5718	
GDP	0.000	0.000	6.652	0.0000	***
Industry_GDP	0.024	0.009	2.538	0.0120	*
Agriculture_GDP	-0.066	0.011	-6.149	0.0000	***
Services_GDP	0.014	0.011	1.320	0.1884	

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 0.7775 on 173 degrees of freedom

Multiple R-squared: 0.7208, Adjusted R-squared: 0.7143

F-statistic: 111.7 on 173 and 4 DF, p-value: 0.0000

Multicollinearity

All VIF scores for the predictors in the model is less than 5, which means that multicollinearaity is not identified. Generally speaking, a VIF score that is higher than 5 or 10 indicates that there is a high correlation that can cause problem.(James et al. 2014, cited in Kassambara,2018).

Table 4 Multicollinearity VIF: log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

GDP	Industry_GDP	Agriculture_GDP	Services_GDP
1.7	3.3	3.5	4.5

A scatterplot for log transformed CO2 emissions and agriculture(% of GDP) in Figure 5 shows negative linear correlation. Scatter plots for log transformed CO2 emissions and GDP per capita shows positive correlation but with logarithmic curve. Industry(% of GDP) and Services(% 0f GDP) have a weak positive correlation.

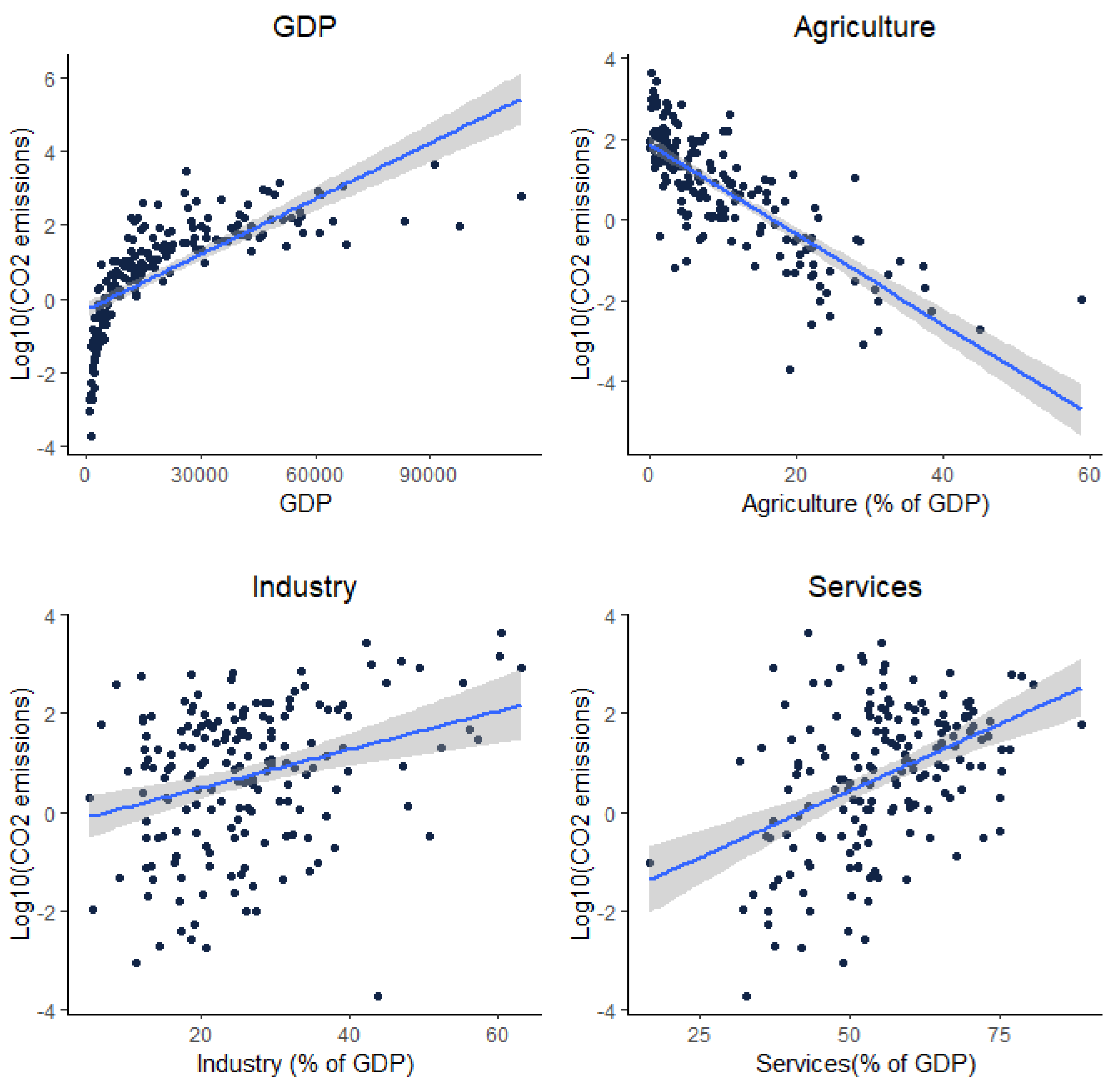

Figure 5 Scatterplot:log(CO2emission) vs GDP, Industry_GDP, Agriculture_GDP

Blue lines in the Added-Variable Plots below (Figure 6) indicates that Services(% of GDP) does not affect on the model as the blue line in Services(% of GDP) is close to horizontal.

```
avPlots(lm_co2_ind_gdp_2018_log_all)
```

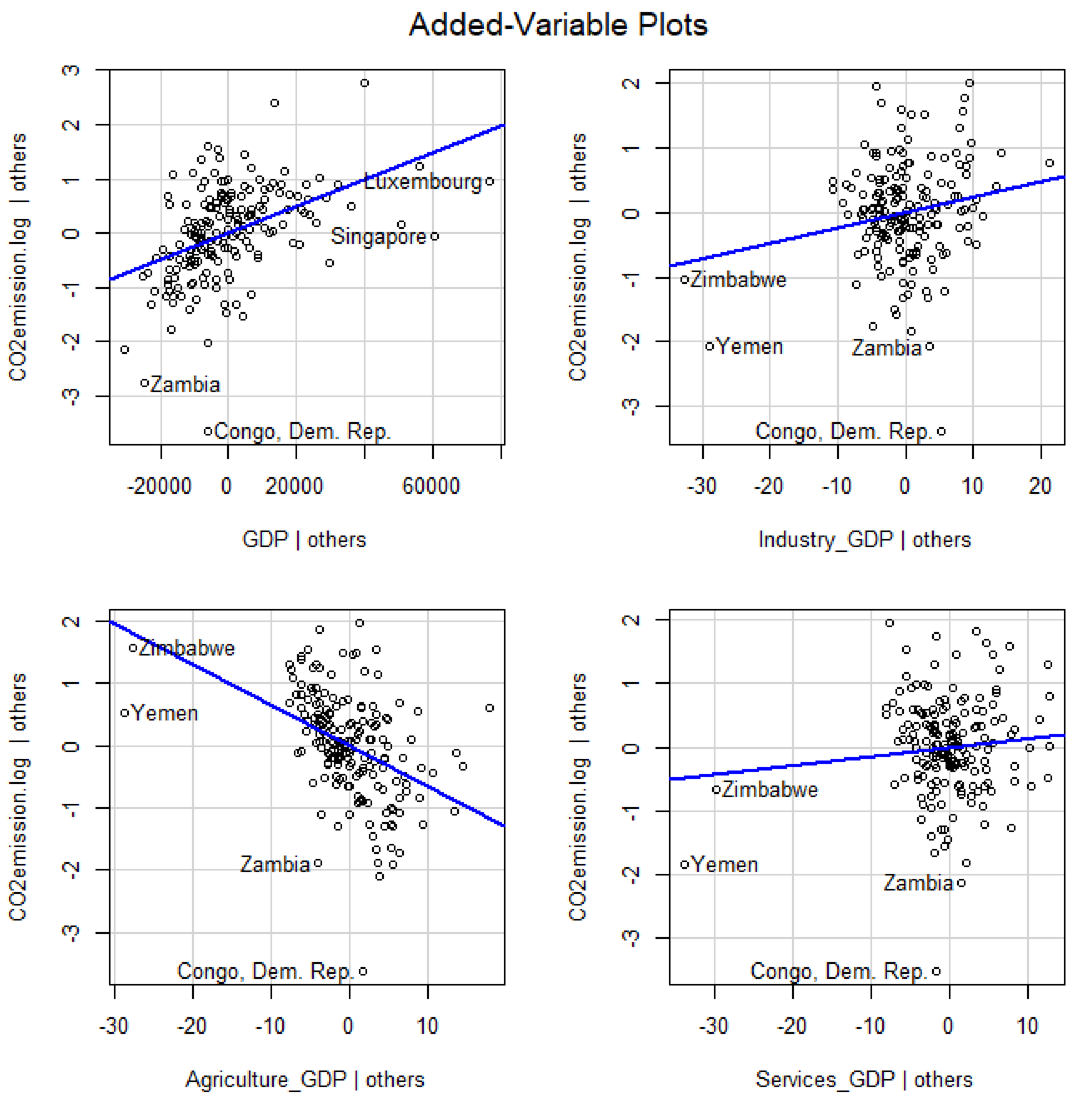

Figure 6 Added variable Plots:log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

Question 2: Are there any differences in CO2 emissions per person between Sweden and Denmark between the late 1900s and early 2000s?

There are 49 observations and no missing values both for Sweden and Denmark. There is not a large difference in variance/SD.

Table 5 Descriptive Statistics

Variable	N	Missing	Mean	Variance	SD	Min	Max	Median (Q1,Q3)

Denmark	49	0	10.37	3.8	1.95	6.06	14.3	10.90 (9.52, 11.80)
Sweden	49	0	7.05	3.87	1.97	4.12	11.5	6.62 (5.90, 7.48)

Histogram and appropriate test to use

According to the histogram below (Figure 7), the data is not normally distributed. Therefore, Wilcoxon's rank-sum test is more appropriate to use.

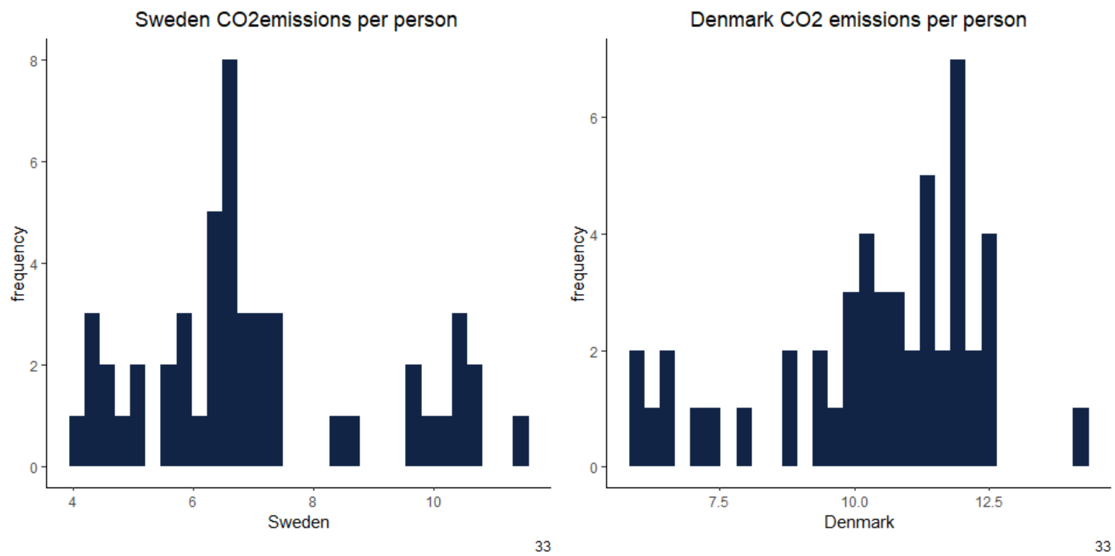

Figure 7 Histogram:Sweden and Denmark CO2emissions (tonnes per person)

Run Wilcoxon's rank-sum test

The result of Wilcoxon's rank-sum test ($v = 1225$, $p\text{-value} < 0.0001$) indicates that there are statistically significant differences in CO2 emission per person between Sweden and Denmark.

Median, mean, SD

Median

There is a difference in median of CO2 emissions between in Sweden(6,62) and Denmark(10,90) in the dataset. The median in Denmark is higher than the one in Sweden.

MEAN

There is a difference in mean of CO2 emissions between Denmark(10.37)and Sweden(7.05) in the dataset.The mean in Denmark is higher than the one in Sweden. But the difference in mean is a little smaller than in median between Sweden and Denmark.

SD

There is a difference in standard deviation of CO2emissions between Denmark(1.95) and Sweden(1.97) in the dataset. However, difference is not large but Denmark has a larger standard deviation than in Sweden.

Box plots to visualize findings

The boxplot below (Figure 8) shows that the median of CO2 is higher in Denmark than in Sweden. The dataset for Denmark is left skewed in Denmark and larger spread(variation) than in Sweden. For Sweden there are outliers in higher CO2 emissions while there is an outlier for Denmark in lower CO2 emissions.

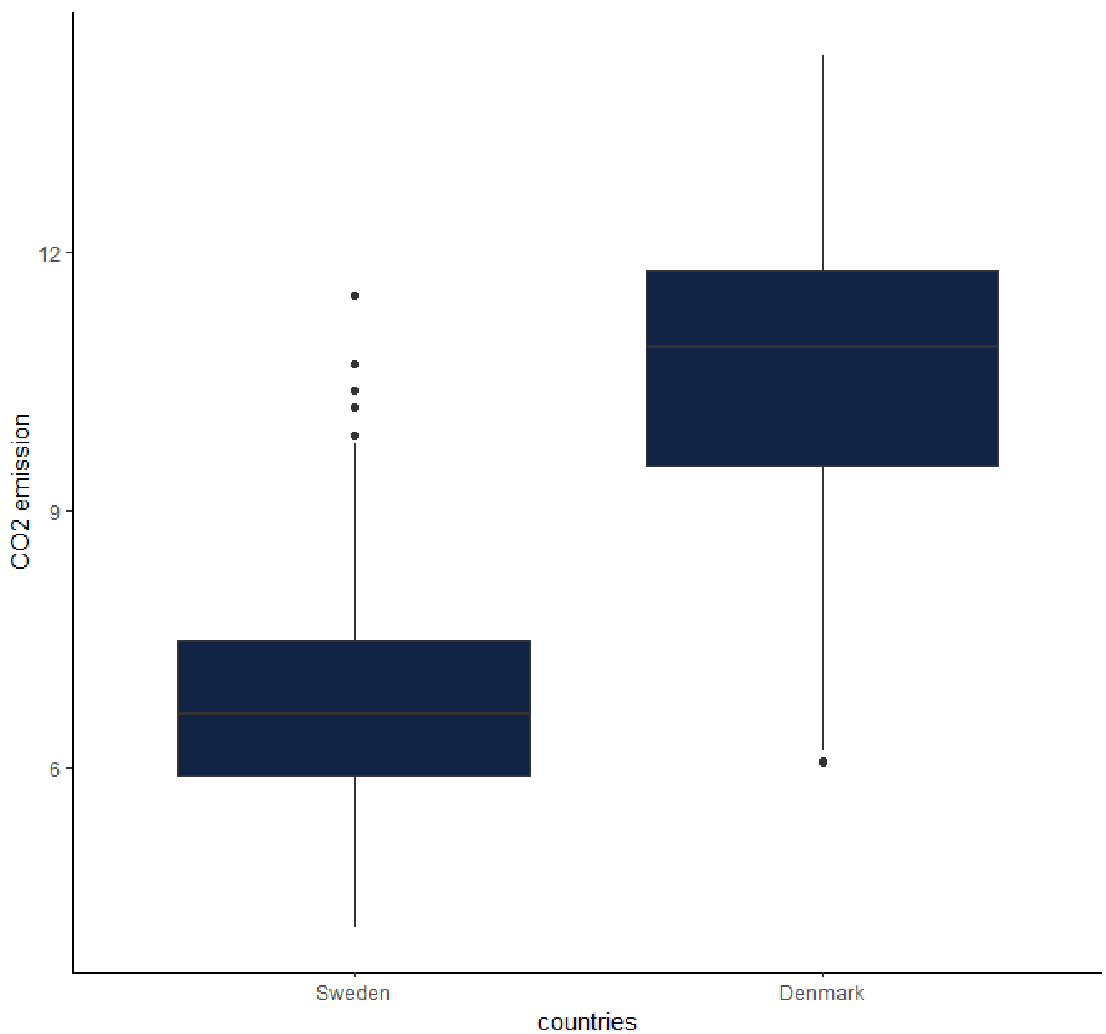

Figure 8 Boxplot:Sweden and Denmark CO2emissions (tonnes per person)

Question 3: Are there any differences in CO_2 emissions per person among Nordic countries as well as between the late 1900s and early 2000s? Are there any interactions between countries and time period(the late 1900s and early 2000s) that have an effect on CO_2 emissions per person?

There are 49 observations and no missing values for every 4 countries. (Sweden, Denmark, Norway, Finland)

Table 6 Descriptive Statistics

Variable	N	Missing	Mean	Variance	SD	Min	Max	Median (Q1,Q3)
Denmark	49	0	10.37	3.8	1.95	6.06	14.3	10.90 (9.52, 11.80)
Finland	49	0	10.67	2.05	1.43	8.06	13.9	10.80 (9.52, 11.50)
Norway	49	0	8.58	0.63	0.8	6.96	9.81	8.47 (8.00, 9.46)
Sweden	49	0	7.05	3.87	1.97	4.12	11.5	6.62 (5.90, 7.48)

Basic diagnostic plots

"Residual vs Fitted values" in diagnostic diagrams (Figure 9) indicates that the biggest spread of the residuals is about three times as large as the smallest spread, which is acceptable level. Normal Q-Q plot in diagnostic diagrams (Figure) shows that there is no large deviation from reference line. This indicates that the residual is roughly normally distributed.

Therefore, Two ways ANOVA is ok to conduct without transformation etc.

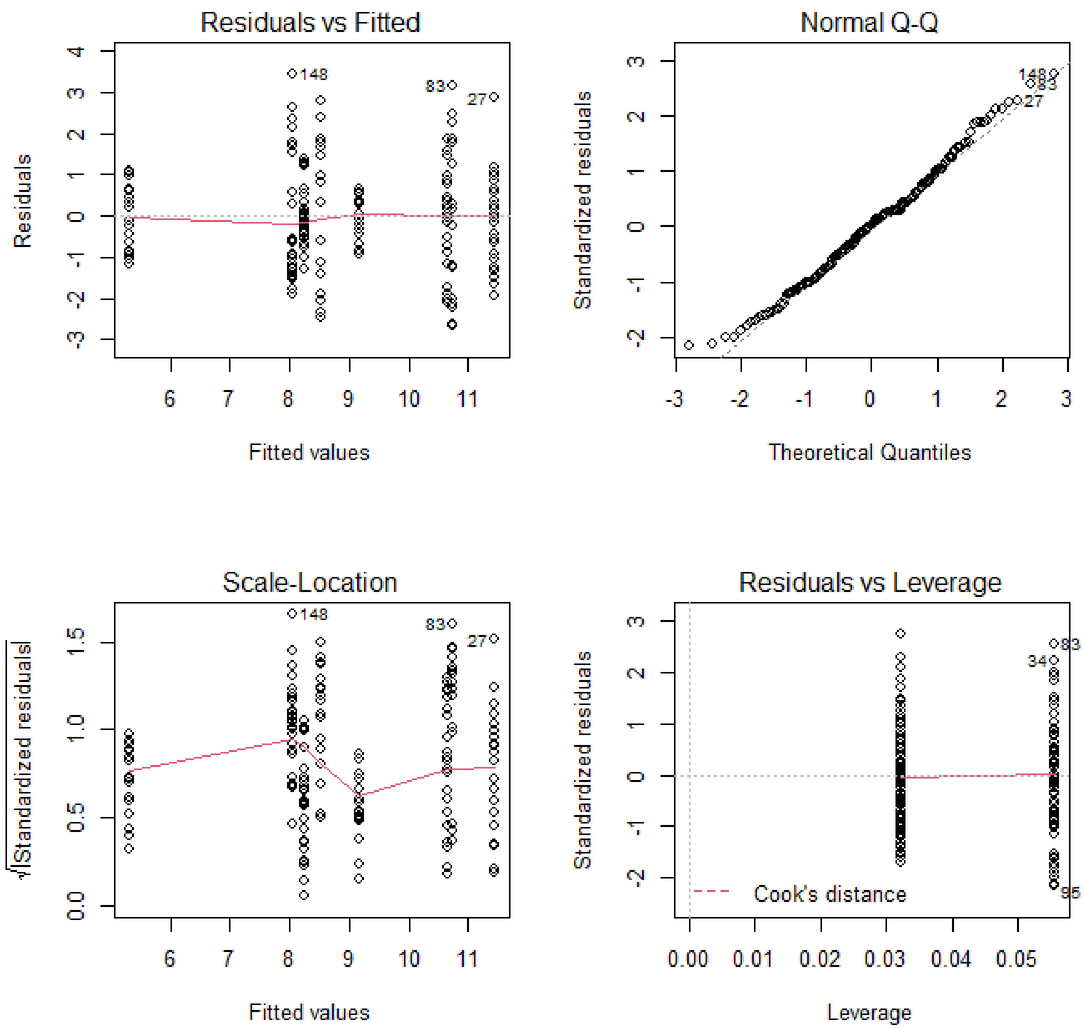

*Figure 9 Basic diagnostic plots:CO2emissions~Country*Year_cat*

The result of ANOVA test below(Table 7) indicates that there is a statistically significant interaction effect of country and year_category(Late 1900s/Early2000s) on CO2 emission per person. (F value:27.35, P value<0.0001)

Table 7 Two way Anova: CO2emissions and Country/Year catergory(Late1900s and Early 2000s)

Anova Table (Type II tests) (continued below)

	Sum Sq	Df	F value
Country	419.6	3	86.74
Year_cat	61.75	1	38.29
Country:Year_cat	132.3	3	27.35

Residuals	303.2	188	NA

	Pr(>F)
Country	0.000000000000000000000000000000000002859
Year_cat	0.000000003732
Country:Year_cat	0.00000000000001007
Residuals	NA

Table 8 Coefficient table: CO2emissions and Country/Year catergory(Late1900s and Early 2000s)

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	11.445	0.228	50.177	0.0000	***
CountryFinland	-0.804	0.323	-2.492	0.0136	*
CountryNorway	-3.210	0.323	-9.953	0.0000	***
CountrySweden	-3.381	0.323	-10.481	0.0000	***
Year_catEarly2000s	-2.915	0.376	-7.746	0.0000	***
CountryFinland:Year_catEarly2000s	3.003	0.532	5.643	0.0000	***
CountryNorway:Year_catEarly2000s	3.852	0.532	7.238	0.0000	***
CountrySweden:Year_catEarly2000s	0.148	0.532	0.278	0.7814	

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1

Residual standard error: 1.27 on 188 degrees of freedom

Multiple R-squared: 0.6693, Adjusted R-squared: 0.657

F-statistic: 54.36 on 188 and 7 DF, p-value: 0.0000

Post-hoc test for differences between Year_category within country

According to the pairwise tests between the late 1900s and early 2000s within countries below (Table 9), there are statistically significant differences between the late 1900s and early 2000s (Denmark F test:60.0063,p-value<0.001, Norway F test: 6.1987, p-value=0.0273, Sweden F test: 54.1, p-value<0.001) within country except for in Finland where there is not statistically significant between the late 1900s and early 2000s (F-test: 0.05499, p-value:0.8149).

Table 9 Pairwise tests between the late 1900s and early 2000s within countries, adjustment = Holms test

Year_category	Value	Df	Sum of Sq	F	Pr(>F)
Late1900s-Early2000s: Denmark	2.91507	1.00000	96.76908	60.00635	0.00000
Late1900s-Early2000s: Finland	-0.08824	1.00000	0.08868	0.05499	0.81486
Late1900s-Early2000s: Norway	-0.93692	1.00000	9.99634	6.19871	0.02730
Late1900s-Early2000s: Sweden	2.76720	1.00000	87.20082	54.07309	0.00000
Residuals		188.00000	303.17770		

Post-hoc test for differences between countries within year category

Pairwise tests between countries within the late 1900s and early 2000s (Table 10) shows that there are statistically significant differences among countries both within the late 1900s and early 2000s in most cases. However, there are no statistically significant differences between Norway and Sweden in the late 1900s (F value:0.2788, P value: 0.598093) and between Denmark and Norway in the early 2000s(F value: 2.2979, p-value:0.262465)

Table 10 Pairwise tests between countries within the late 1900s and early 2000s, adjustment = Holms test

Country	Value	Df	Sum of Sq	F	Pr(>F)
Denmark-Finland : Late 1900s	0.80387	1.00000	10.01623	6.21105	0.04068
Denmark-Norway : Late 1900s	3.21032	1.00000	159.74565	99.05802	0.00000
Denmark-Sweden : Late 1900s	3.38065	1.00000	177.14581	109.84783	0.00000
Finland-Norway : Late 1900s	2.40645	1.00000	89.76065	55.66043	0.00000
Finland-Sweden : Late 1900s	2.57677	1.00000	102.91636	63.81827	0.00000
Norway-Sweden : Late 1900s	0.17032	1.00000	0.44965	0.27883	0.59809
Denmark-Finland : Early 2000s	-2.19944	1.00000	43.53800	26.99785	0.00000
Denmark-Norway : Early 2000s	-0.64167	1.00000	3.70563	2.29785	0.26247
Denmark-Sweden : Early 2000s	3.23278	1.00000	94.05767	58.32501	0.00000
Finland-Norway : Early 2000s	1.55778	1.00000	21.84004	13.54298	0.00122
Finland-Sweden : Early 2000s	5.43222	1.00000	265.58134	164.68656	0.00000

Norway-Sweden : Early 2000s	3.87444	1.00000	135.10188	83.77646	0.00000
Residuals		188.00000	303.17770		

An effect plot below (Figure 10) shows that there are differences between Denmark/Finland and Norway/Sweden in the late 1900s. Norway and Sweden has lower co2 emissions than Denmark and Finland. However, there are not big differences between Denmark and Finland as well as Norway and Sweden in the late 1900s.

In the early 2000s, Sweden has significantly lower CO2 emissions than the other 3 Nordics countries according to the graph. However, there are no big differences between Norway and Finland or between Norway and Denmark while there are some differences between Finland and Denmark. Finland has higher CO2 emissions than the one in Denmark.

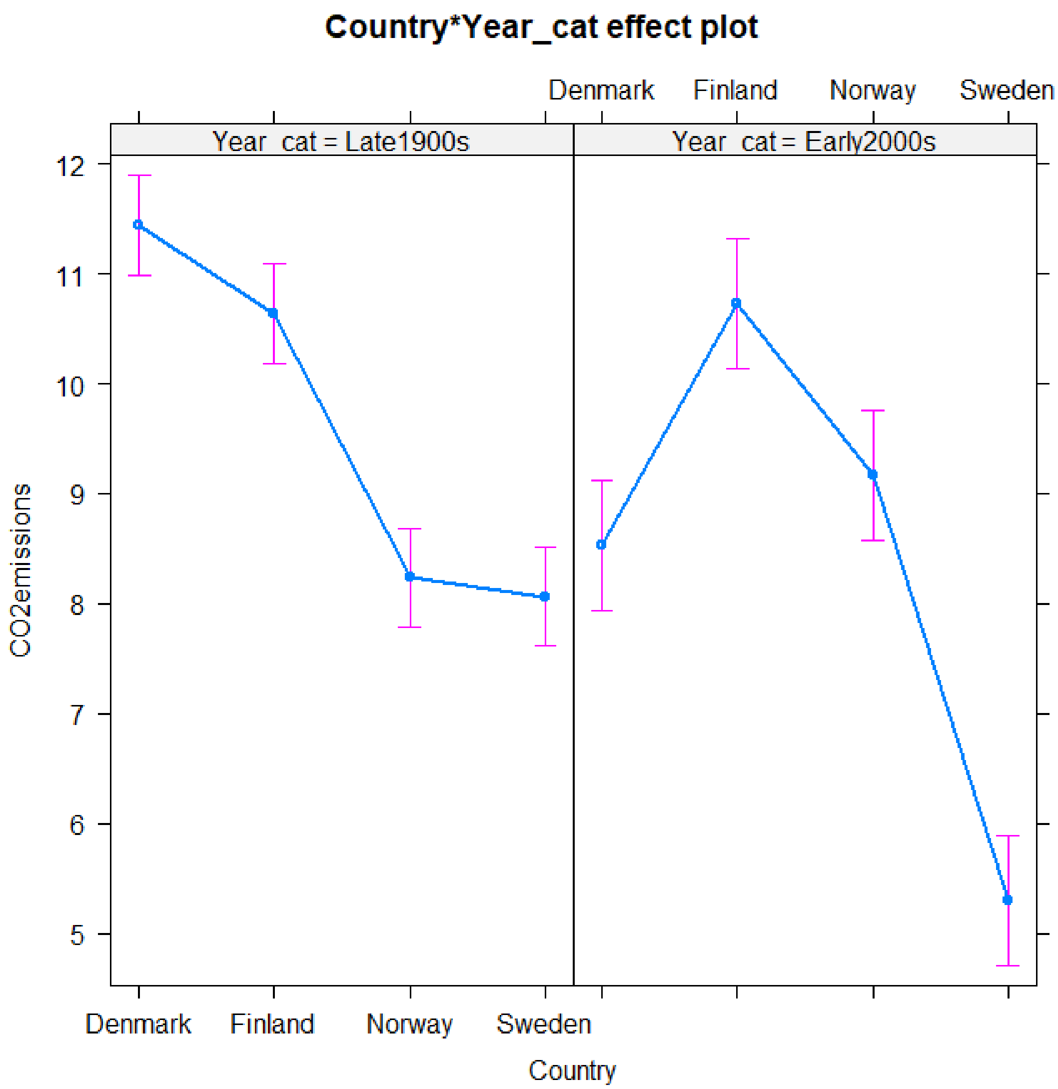

Figure 10 Effect Plot: country and year_category(Late 1900s and Early 2000s)

Effect plot below (Figure 11) indicates that in Denmark and Sweden CO2 emissions per person in the early 2000s is much lower than in the late 1900s while in Norway it is higher in the early 2000s than in the late 1900s. In Finland there is not much difference between the late 1900s and early 2000s.

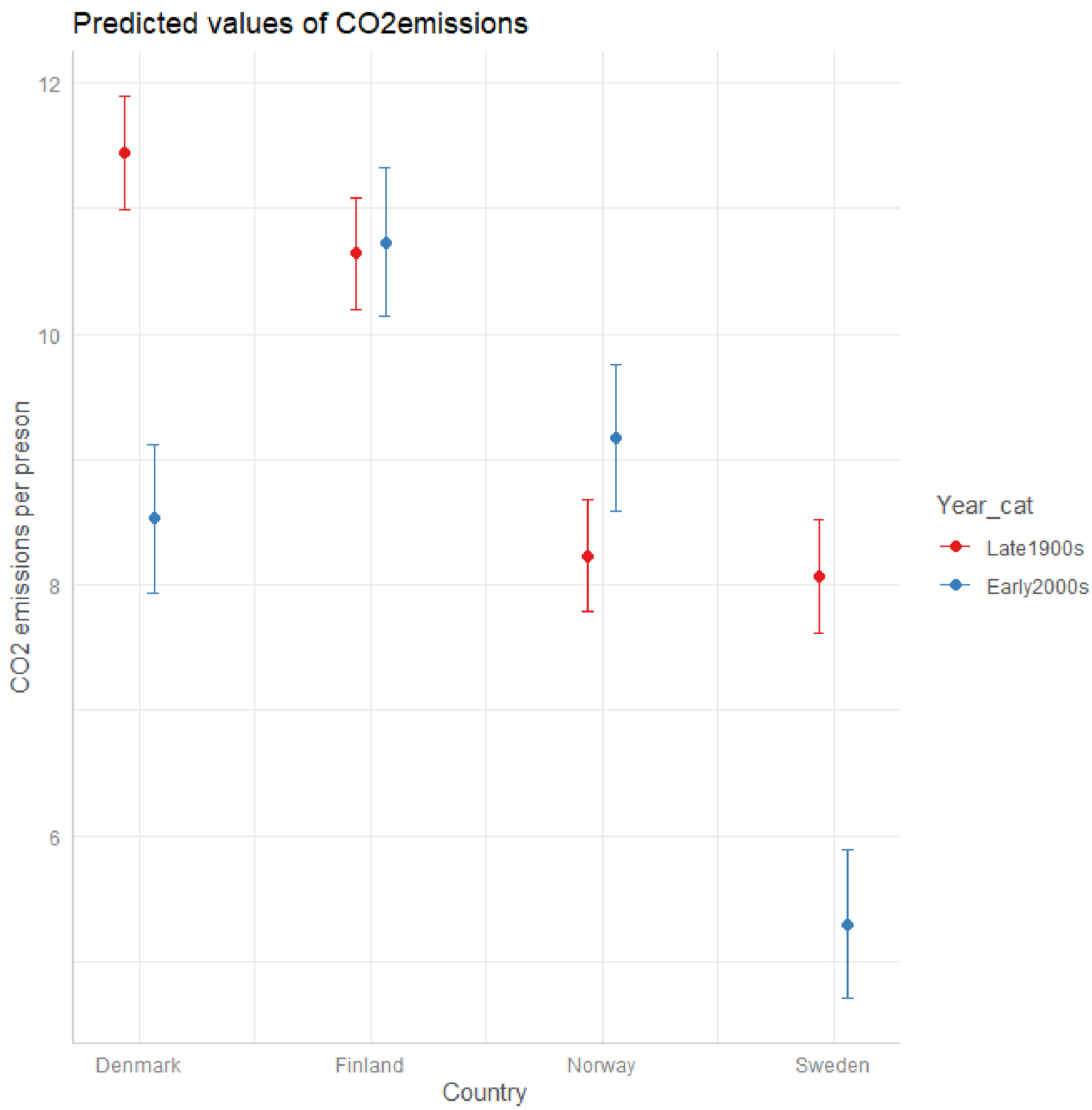

Figure 11 Effect Plot: country and year_category(Late 1900s)

Question 4 Are there any associations in attitude/awareness toward environment (intentional purchase of items with eco-friendly/less packaging) and working styles (work from home and people who work away from home) as well as age group, Generation Z,Young Millenials(23-26 years old), Core Millenials(27-32 years old), Mature Millenials(33-36 years old), Generation X, Baby Boomers ?

The result of the Chi-squared test (X-squared = 47.876, df = 1, p-value < 0.000000000004542) suggests that there is a statistically significant association between working style(work from home/work away from home) and intentionally buying items with eco-friendly/less packaging.

Table 11 Observerd Frequency (remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

Reply	Work from home	Work away from home
I intentionally buy items with eco-friendly/less packaging	2084	1307
I do not intentionally buy items with eco-friendly/less packaging	1332	1206

Table 12 Expected Frequency(remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

Reply	Work from home	Work away from home
I intentionally buy items with eco-friendly packaging or less packaging	1954	1437
I do not intentionally buy items with eco-friendly packaging or less packaging	1462	1076

The graph below (Figure 12) also indicates that a larger portion of remote worker(= workers who work from home) respondents answered that they intentionally buy items with eco-friendly packaging or less packaging while just over half of non remote worker(=workers who work away from home) respondents answered that they intentionally buy items with eco-friendly packaging or less packaging.

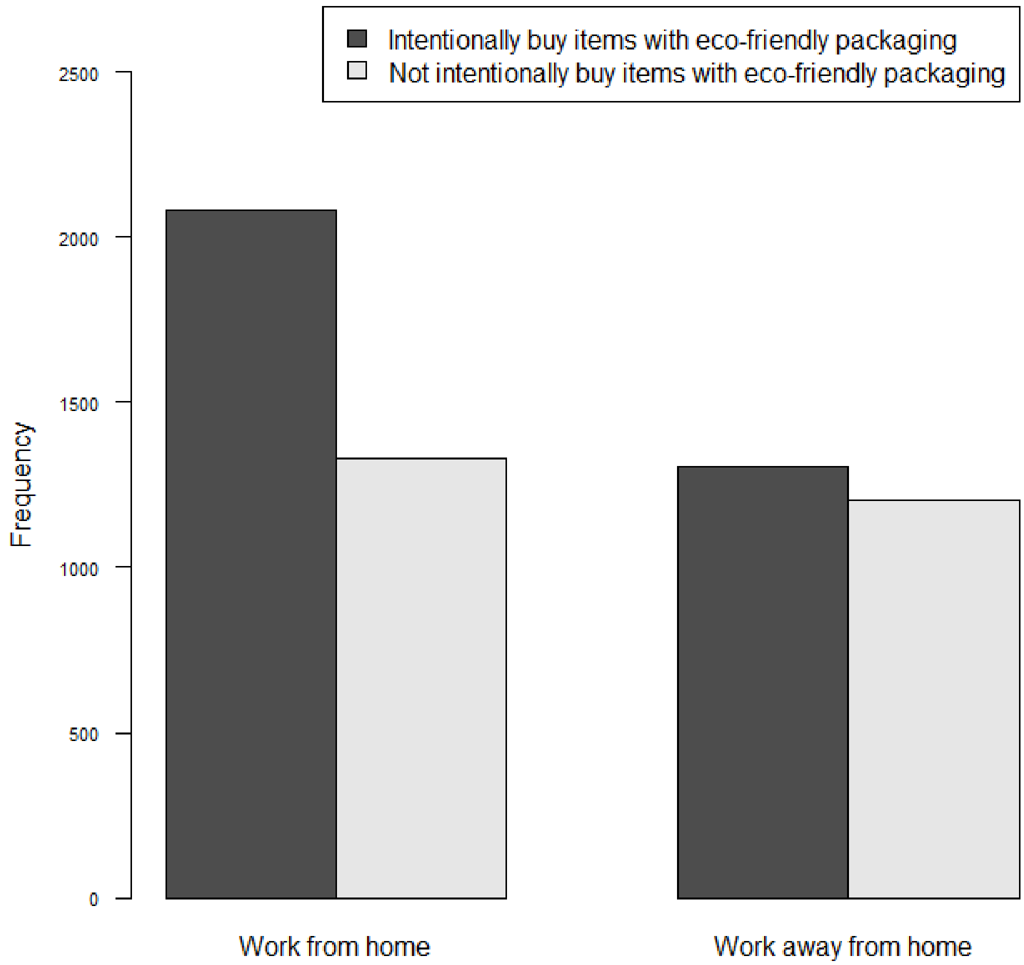

Figure 12 Bar chart (Remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

Pearson's residuals (Table 13) and correlation matrix (Figure 13) suggest that remote workers have a positive association with the answer "I do not intentionally buy items with eco-friendly packaging or less packaging" while non-remote worker have a negative association with the answer. This implies that remote worker are more likely to intentionally choose items with eco-friendly packaging or less packaging.

Table 13 Pearson's residuals (remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

Reply	**Work from home**	**Work away from home**

I intentionally buy items with eco-friendly packaging or less packaging	3	-3
I do not intentionally buy items with eco-friendly packaging or less packaging	-3	4

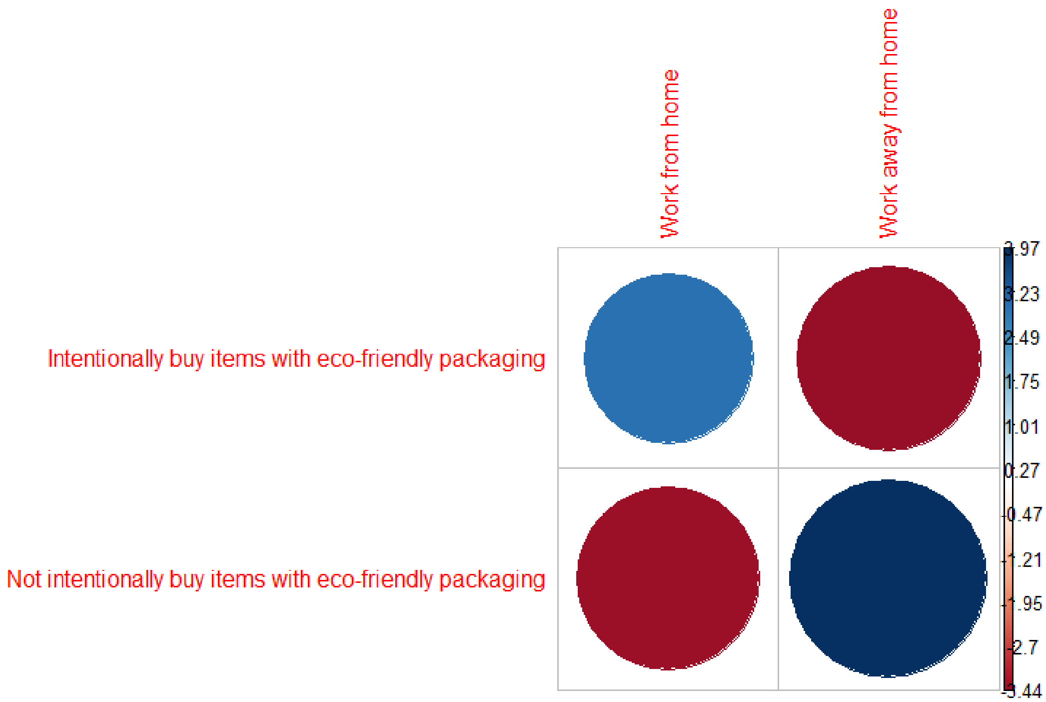

Figure 13 Correlation Plot (Residual: remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

Regarding the association between age groups (“Generation Z”,“Young Millenials”, “Core Millenials”, “Mature Millenials”, “Generation X”, “Baby Boomers”) and choosing items with eco-friendly packaging, the result of the Chi-squared test (X-squared = 47.78, df = 5, p-value = 0.000000003938, X-squared test for association) suggests that there is a statistically significant association between age groups and choosing eco-friendly packaging.

Table 14 Observed Frequency (Age groups and choosing items with eco-friendly packaging)

Reply	**Generation Z**	**Young Millenials**	**Core Millenials**	**Mature Millenials**	**Generation X**	**Baby Boomers**

I intentionally buy items with eco-friendly packaging or less packaging	653	513	953	505	1566	497
I do not intentionally buy items with eco-friendly packaging or less packaging	707	420	635	414	1281	478

Table 15 Expected Frequency (Age groups and choosing eco-friendly packaging)

Reply	Generation Z	Young Millenials	Core Millenials	Mature Millenials	Generation X	Baby Boomers
Intentionally buy items with eco-friendly packaging	739	507	863	500	1548	530
Not intentionally buy items with eco-friendly packaging	621	426	725	419	1299	445

The graph below (Figure 13) indicates that a larger portion of Millenials and Generation X respondents answered that they buy intentionally items with eco-friendly packaging while a larger portion (more than 50%) of Generation Z answered that they do not buy intentionally items with eco-friendly packaging.

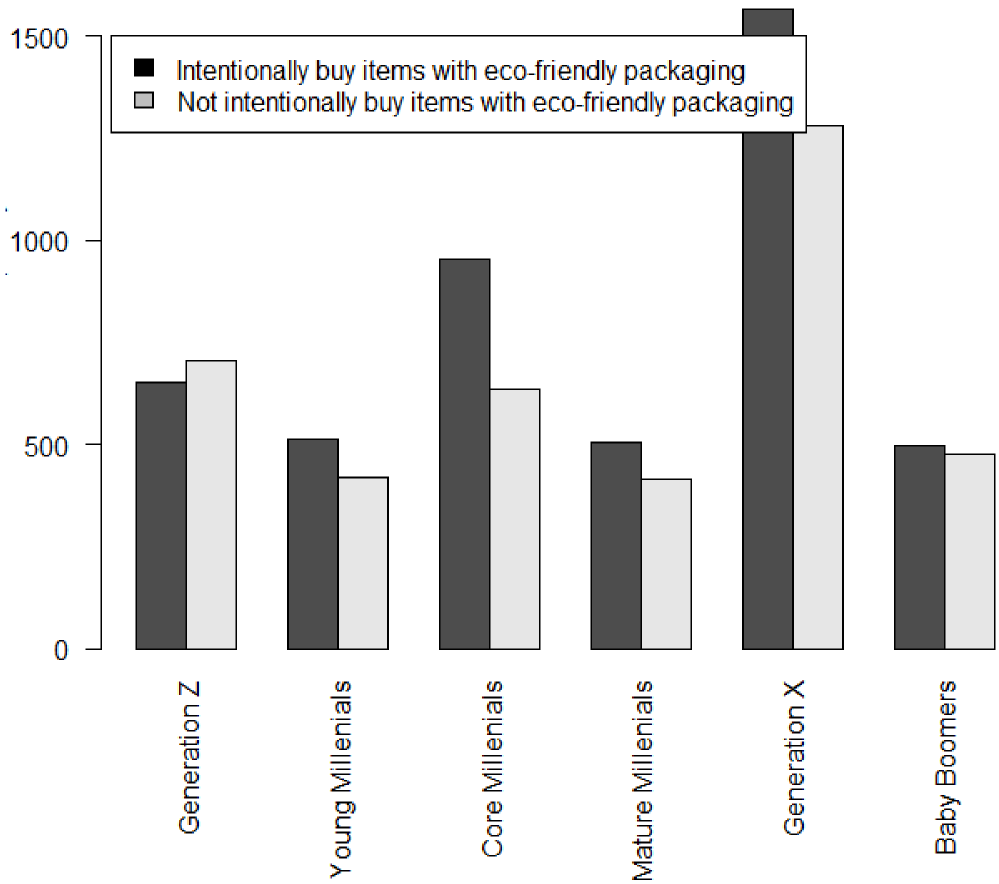

Figure 14 Bar chart (Age groups and choosing items with eco-friendly packaging)

Pearson's residuals (Table 16) and correlation matrix (Figure 15) suggest that Core-Millenials have a positive association with buying intentionally items with eco-friendly packaging while Generation Z have a negative association with buying intentionally items with eco-friendly packaging. It means that Core-Millenials are more likely to buy intentionally items with eco-friendly packaging while Generation Z are less likely to do that.

Table 16 Pearson's residuals (Age groups and choosing items with eco-friendly packaging)

Reply	Generation Z	Young Millenials	Core Millenials	Mature Millenials	Generation X	Baby Boomers

Intentionally buy items with eco-friendly packaging	-3	0	3	0	0	-1
Not intentionally buy items with eco-friendly packaging	3	-0	-3	-0	-1	2

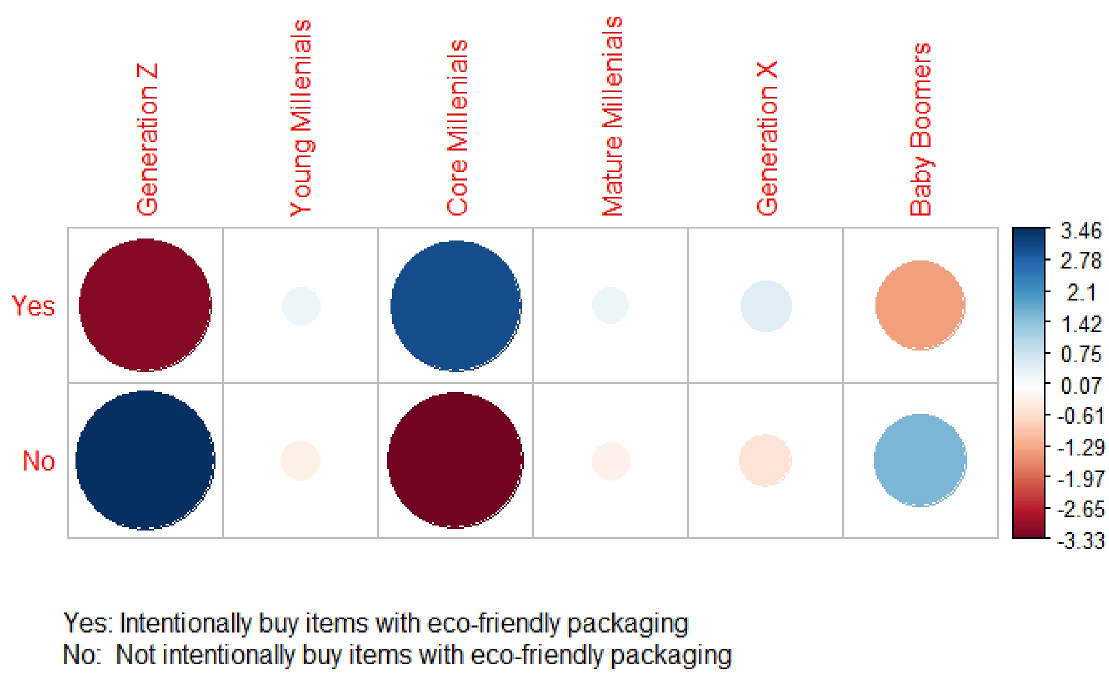

Figure 15 Correlation Plot (residual) for Age groups and choosing items with eco-friendly packaging

Discussion

Question 1: Does GDP per capita and % of GDP by sectors have a relation with CO2 emission per person by countries in 2018?

Linear regression analysis shows that GDP, Industry(% of GDP) have a statistically significant positive relationships with CO2 emissions while Agriculture(% of GDP) has a statistically significant negative relationships with log transformed CO2 emissions. However, it does not indicate that Services(% of GDP) have a significant influence on log transformed CO2 emissions in the model.

Scatterplots show that GDP have a positive linear correlation with CO2 emissions in 2018 while GDP and log transformed CO2 emissions show positive correlation with logarithmic curve. These indicate that GDP per capita has a linear correlation with CO2 emissions. Industry(% of GDP) has a weak linear correlation with log transformed CO2 emissions. Additionally, Agriculture (% of GDP) has a negative linear correlation with log transformed CO2 emissions. It implies that agriculture (% of GDP) has a negative exponential correlation with CO2 emissions, which means that the countries that have higher % of GDP in agriculture have significantly lower CO2 emissions than the countries with lower % of GDP in agriculture in 2018.

Added variable Plots indicates that Services(% of GDP) does not have a enough influence on the model with log transformed CO2 emissions, which implies that removing Services (% of GDP) should be considered for removal from the model for better prediction.

Question 2: Are there any differences in CO2 emissions per person between Sweden and Denmark between the late 1900s and early 2000s?

The result of Wilcoxon's rank-sum test shows that there are statistically significant differences in CO2 between Sweden and Denmark. Denmark's CO2 emissions is higher than Sweden.

Question 3: Are there any differences in CO2 emissions per person among Nordic countries as well as between the late 1900s and early 2000s? Are there any interactions between countries and time period(the late 1900s and early 2000s) that have an effect on CO2 emissions per person?

The result of Anova two-way test suggests that there are statistically significant interaction effect between Country and year_category.

The pairwise tests between the late 1900s and early 2000s within countries indicates that there are statistically significant differences between the late 1900s and early 2000s but there is one exception for in Finland, where statistically significant difference is not identified.

The pairwise tests between countries within the late 1900s and early 2000s shows that there are statistically significant differences between Denmark/Finland and Norway/Sweden in the late 1900s. Norway and Sweden has lower co2 emissions than

Denmark and Finland in the late 1900s. There are also statistically significant difference between Denmark and Finland in 1900s although the difference seems relatively small. However, there is no statistically significant difference between Noway and Sweden in the late 1900s.

In the early 2000s, there are statistically significant difference between Sweden and the other 3 Nordics countries. Sweden has significantly lower CO2 emissions than the other 3 Nordics countries. On the other hand, there are no statistically significant differences between Norway and Finland or Denmark. But there are statistically significant difference between Finland and Denmark. Finland has higher CO2 emissions than the one in Denmark.

Question 4: Are there any associations in attitude/awareness toward environment (intentional purchase of items with eco-friendly/less packaging) and working styles (work from home and people who work away from home) as well as age group, Generation Z,Young Millenials(23-26 years old), Core Millenials(27-32 years old), Mature Millenials(33-36 years old), Generation X, Baby Boomers ?

The result of the Chi-squared test indicates that there is a statistically significant association between remote working and intentionally buying items with eco-friendly/less packaging. Remote workers have a positive association with intentionally buying items with eco-friendly packaging or less packaging while non-remote worker have a negative association with the answer. This implies that remote worker are more likely to intentionally choose items with eco-friendly packaging or less packaging.

Regarding the association between age and choosing eco-friendly packaging, there is a statistically significant association between age groups ("Generation Z","Young Millenials", "Core Millenials", "Mature Millenials", "Generation X", "Baby Boomers").Core-Millenials have a positive association with buying intentionally items with eco-friendly packaging while Generation Z have a negative association with buying intentionally items with eco-friendly packaging. It implies that core-Millenials are more likely to buy intentionally items with eco-friendly packaging while Generation Z are less likely to do that.

References

Gapminder(no date) *Download the data* Availables att:https://www.gapminder.org/data/ (Accessed: 1 Dec 2021).

Kassambara,A.(2018). *Regression Model Diagnostics Multicollinearity Essentials and VIF in R* Availabale at: http://www.sthda.com/english/articles/39-regression-model-diagnostics/160-multicollinearity-essentials-and-vif-in-r/(Accessed: 1 Dec 2021).

Salkind N.J. and Shaw L.A.(2020).*Statistics for People Who (think they) Hate Statistics Using R*, SAGE Publications, Inc.

PWC.(2011).*PwC's June 2021 Global Consumer Insights Pulse Survey.* Retrieved from https://www.pwc.com/gx/en/consumer-markets/consumer-insights-survey/2021/gcis-june-2021.pdf

Appendix

```
library(Rcmdr)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(gridExtra)
library(patchwork)
library(reshape2)
library(plyr)
library(rapport)
library(Hmisc)
library(reshape)
library(readxl)
library(car)
library(huxtable)
library(corrplot)
library(tibble)
library(phia)
library(multcomp)
library(Rmisc)
library(PMCMRplus)
library(flextable)
library(kableExtra)
library(pander)
library(qwraps2)
library(ggeffects)
library(RColorBrewer)
```

Question 1: Does GDP per capita and % of GDP by sectors have an influence on/correaltion with CO2 emission per person by countries in 2018?

```
co2_emissions_all<-read.csv("co2_emissions_tonnes_per_person.csv", sep=",", r
ow.names=1, dec=".",header=FALSE)
rownames(co2_emissions_all)[1]<-"year"
co2_emissions_all<-co2_emissions_all[-c(1),]
colnames(co2_emissions_all)<-1800:2018
co2_emissions_2018<-co2_emissions_all[colnames(co2_emissions_all)==2018]
colnames(co2_emissions_2018)<-2018


gdp_all<-read.csv("income_per_person_gdppercapita_ppp_inflation_adjusted.csv"
, sep=",",row.names=1,header=FALSE)
gdp_all<-gdp_all[-c(1),]
colnames(gdp_all)<-1800:2050
gdp_all_2018<-gdp_all[colnames(gdp_all)==2018]
colnames(gdp_all_2018)<-2018


for(i in 1:nrow(gdp_all_2018))
{
```

```
  var<- gdp_all_2018[i,1]
  if(grepl("k",var)){
    var<- gsub("k","",var)
    var<-as.numeric(var)*1000}

  var<-as.numeric(var)
  gdp_all_2018[i,1]<- var
}

industry_gdp_all<- read.csv("industry_percent_of_gdp.csv", sep=",", row.names
=1)
colnames(industry_gdp_all)<-1960:2020
industry_gdp_all_2018<-industry_gdp_all[names(industry_gdp_all)==2018]

gdp_co2_2018 <- merge(co2_emissions_2018,gdp_all_2018,by=0)
rownames(gdp_co2_2018)<-gdp_co2_2018[,1]
gdp_co2_2018 <-gdp_co2_2018[,-1]
colnames(gdp_co2_2018) <- c("CO2emission", "GDP")
industry_gdp_co2_2018 <- merge(industry_gdp_all_2018,gdp_co2_2018,by=0)
rownames(industry_gdp_co2_2018)<-industry_gdp_co2_2018[,1]
colnames(industry_gdp_co2_2018)[1:2]<-c("Country","Industry_GDP")



agriculture_gdp_all<-read.csv("agriculture_percent_of_gdp.csv", sep=",", row.
names=1)
services_gdp_all<-read.csv("services_percent_of_gdp.csv", sep=",", row.names=
1)

merg_fun <-function(newdf,maindf, year, title){
  colnames(newdf)<-sub("X","", colnames(newdf))
  colnames(newdf)<-as.numeric(colnames(newdf))

  var<-year
  newdf <- newdf[colnames(newdf)==year]
  maindf <- merge(newdf,maindf,by=0)
  names(maindf)[names(maindf)==year]<-title
  rownames(maindf)<-maindf[,"Country"]
  maindf<- subset(maindf,select=-c(Row.names))
  return(maindf)
}

industry_gdp_co2_2018<-merg_fun(agriculture_gdp_all,industry_gdp_co2_2018,"20
18","Agriculture_GDP")
industry_gdp_co2_2018<-merg_fun(services_gdp_all,industry_gdp_co2_2018,"2018"
,"Services_GDP")

industry_gdp_co2_2018 <- industry_gdp_co2_2018[,c("Country","CO2emission","GD
```

```
P","Industry_GDP","Agriculture_GDP","Services_GDP")]
industry_gdp_co2_2018<-na.omit(industry_gdp_co2_2018)
```

Table 1 Descriptive Statistics

```
isnum <- ifelse(sapply(industry_gdp_co2_2018[-1], class) == "numeric", 1, NA)
mean<-sapply(industry_gdp_co2_2018,function(x) round(mean(x),digits=2))
variance<-sapply(industry_gdp_co2_2018,function(x) round(var(x),digits=2))
SD<-sapply(industry_gdp_co2_2018,function(x) round(sd(x),digits=2))
min<-sapply(industry_gdp_co2_2018[-1],function(x) round(min(x),digits=2))
max<-sapply(industry_gdp_co2_2018[-1],function(x) round(max(x),digits=2))
missingvalue<-sapply(industry_gdp_co2_2018[-1],function(x) sum(is.na(x)))
N<-sapply(industry_gdp_co2_2018[-1],function(x) sum(!is.na(x)))
median<-sapply(industry_gdp_co2_2018[-1],function(x) median_iqr(x,na_rm = TRU
E))

summary1<-cbind(Variable=colnames(industry_gdp_co2_2018),N=N, Missing =missin
gvalue, Mean=mean,Variance=variance,SD=SD,Min=min,Max=max,"Median(Q1,Q3)"=med
ian)
summary1<- as.data.frame(summary1)


h_summary1<-as_hux(summary1)
h_summary1  %>%
  set_align("center")%>%
  set_width(1) %>%
  theme_basic()
```

Figure 1 Scatterplots (CO2 emission and GDP, % of GDP by sectors)

```
g1 <- ggplot(industry_gdp_co2_2018)+
  aes(x=GDP, y = CO2emission) +
  labs(x= "GDP per capita", y="CO2emission/person ")+
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  theme_classic()

g2 <- ggplot(industry_gdp_co2_2018)+
  aes(x=Industry_GDP, y = CO2emission) +
  labs(x= "Industry(% of GDP)", y="CO2emission/person")+
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  theme_classic()


g3<- ggplot(industry_gdp_co2_2018)+
  aes(x=Agriculture_GDP, y = CO2emission) +
  labs(x= "Agriculture(% of GDP)", y="CO2emission/person")+
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  theme_classic()


g4 <- ggplot(industry_gdp_co2_2018)+
```

```
  aes(x=Services_GDP, y = CO2emission) +
  labs(x= "Services (% of GDP)", y="CO2emission/person")+
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  theme_classic()



grid.arrange(g1, g2, g3,g4, nrow = 2, ncol=2)
```

Figure 2 Correlation Plot (CO2 emissions and GDP or percentage of GDP by Sectors)

```
cor_data1 <- industry_gdp_co2_2018[,-c(1)]
corr <-cor(cor_data1)
corrplot(corr)
```

Table 2 Linear Regression Analysis: Coefficeints Table
CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
lm_co2_ind_gdp_2018 <- lm(CO2emission~GDP+Industry_GDP+Agriculture_GDP+Servic
es_GDP,industry_gdp_co2_2018)
lm_co2_ind_gdp_2018 %>% as_flextable()
```

Figure 3 Basic diagnostic plots:
CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
par(mfrow=c(2,2))
plot(lm_co2_ind_gdp_2018)
par(mfrow=c(1,1))
```

Linear regression analysis after log transformation:
log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
industry_gdp_co2_2018$CO2emission.log<-log(industry_gdp_co2_2018$CO2emission)
```

Figure 4 Basic diagnostic plots(long
transformation):log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
par(mfrow=c(2,2))
plot(lm_co2_ind_gdp_2018_log_all)
par(mfrow=c(1,1))
```

Table 3 Linear Regression Analysis(long transformation): Coefficeints Table :
CO2emission~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
lm_co2_ind_gdp_2018_log_all <- lm(CO2emission.log~GDP+Industry_GDP+ Agricultu
re_GDP+Services_GDP,industry_gdp_co2_2018)
lm_co2_ind_gdp_2018_log_all%>% as_flextable()
```

Table 4 Multicollinearity VIF:
log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
vif1 <- t(as.data.frame(vif(lm_co2_ind_gdp_2018_log_all)))

vif1%>%
  as_hux() %>%
  add_colnames() %>%
```

```
  set_number_format(1)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  set_width(1) %>%
  theme_basic()
```

Figure 5 Scatterplot:log(CO2emission vs GDP, Industry_GDP,Agriculture_GDP)

```
g7 <- ggplot(industry_gdp_co2_2018) +
  aes(x = GDP, y =CO2emission.log ) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  labs(x = "GDP", y = "Log10(CO2 emissions)", title = "GDP", caption = "Data
source: Gapminder(https://www.gapminder.org/data/)")+
  theme_classic()+
  stat_smooth(method="lm", formula=y~x, geom="smooth")+
  theme(plot.title = element_text(hjust = 0.5))

g8 <- ggplot(industry_gdp_co2_2018) +
  aes(x = Agriculture_GDP, y =CO2emission.log ) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  labs(x = "Agriculture (% of GDP)", y = "Log10(CO2 emissions)", title = "Agr
iculture", caption = "Data source: Gapminder(https://www.gapminder.org/data/)
")+
  theme_classic()+
  stat_smooth(method="lm", formula=y~x, geom="smooth")+
  theme(plot.title = element_text(hjust = 0.5))

g9 <- ggplot(industry_gdp_co2_2018) +
  aes(x = Industry_GDP, y =CO2emission.log ) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  labs(x = "Industry (% of GDP)", y = "Log10(CO2 emissions)", title = "Indust
ry", caption = "Data source: Gapminder(https://www.gapminder.org/data/)")+
  theme_classic()+
  stat_smooth(method="lm", formula=y~x, geom="smooth")+
  theme(plot.title = element_text(hjust = 0.5))


g10 <- ggplot(industry_gdp_co2_2018) +
  aes(x = Services_GDP, y =CO2emission.log ) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  labs(x = "Services(% of GDP)", y = "Log10(CO2 emissions)", title = "Service
s", caption = "Data source: Gapminder(https://www.gapminder.org/data/)")+
  theme_classic()+
  stat_smooth(method="lm", formula=y~x, geom="smooth")+
  theme(plot.title = element_text(hjust = 0.5))


grid.arrange(g7, g8, g9, g10, nrow =2, ncol=2)
```

Figure 6 Added variable Plots:log(CO2emission)~GDP+Industry_GDP+Agriculture_GDP+Services_GDP

```
avPlots(lm_co2_ind_gdp_2018_log_all)
```

Question 2: Are there any differences in CO2 emissions per person between Sweden and Denmark between the late 1900s and early 2000s?

```
co2_emissions_all<-read.csv("co2_emissions_tonnes_per_person.csv", sep=",",row.names=1)
colnames(co2_emissions_all)<-sub("X","", colnames(co2_emissions_all))
co2_emissions<-co2_emissions_all %>%
  filter(row.names(co2_emissions_all) %in% c("Sweden","Denmark","Finland","Norway"))


co2_emissions<- as.data.frame(t(co2_emissions))
co2_emissions <- as.data.frame(sapply(co2_emissions, as.numeric))
co2_emissions$year<-c(seq(1800,2018))
co2_emissions<-subset(co2_emissions,year>=1970)
```

Table 5 Descriptive Statistics

```
mean<-sapply(co2_emissions,function(x) round(mean(x),digits=2))
variance<-sapply(co2_emissions,function(x) round(var(x),digits=2))
SD<-sapply(co2_emissions,function(x) round(sd(x),digits=2))
min<-sapply(co2_emissions,function(x) round(min(x),digits=2))
max<-sapply(co2_emissions,function(x) round(max(x),digits=2))
missingvalue<-sapply(co2_emissions,function(x) sum(is.na(x)))
N<-sapply(co2_emissions,function(x) sum(!is.na(x)))
median<-sapply(co2_emissions,function(x) median_iqr(x,na_rm = TRUE))

summary<-cbind(Variable=colnames(co2_emissions),N=N, Missing =missingvalue, Mean=mean,Variance=variance,SD=SD,Min=min,Max=max,"Median(Q1,Q3)"=median)
summary <- as.data.frame(summary)
summary_ds<-summary %>%
  filter(row.names(summary) %in% c("Sweden","Denmark"))


h_summary<-as_hux(summary_ds)
h_summary  %>%
  set_align("center")%>%
  set_width(1) %>%
  theme_basic()
```

Figure 7 Histogram:Sweden and Denmark CO2emissions (tonnes per person)

```
sco<- ggplot(co2_emissions)+
  aes(x=Sweden)+
  geom_histogram(bins = 30L, fill = "#112446") +
  labs(x = "Sweden", y="frequency", title="Sweden CO2emissions per person " , caption="33") +
  theme_classic()+
```

```
  theme(plot.title = element_text(hjust = 0.5))

dco <- ggplot(co2_emissions)+
  aes(x=Denmark)+
  geom_histogram(bins = 30L, fill = "#112446") +
  labs(x = "Denmark", y="frequency", title="Denmark CO2 emissions per person"
,caption="33") +
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))

grid.arrange(sco, dco, nrow = 1, ncol=2)
```

Run Wilcoxon's rank-sum test

```
with(co2_emissions, wilcox.test(Denmark, Sweden, alternative='two.sided', pai
red=TRUE))
```

Graphs to visualize findings

Figure 8 Boxplot:Sweden and Denmark CO2emissions (tonnes per person)

```
options(scipen=999)
ggplot(co2.m) +
  aes(x=variable, y = value) +
  geom_boxplot(shape = "circle", fill = "#112446") +
  labs(x = "countries", y = "CO2 emission", title="" ,caption="") +
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))
```

Question 3: Are there any differences in CO2 emissions per person among Nordic countries as well as between the late 1900s and early 2000s? Are there any interactions between countries and time period(the late 1900s and early 2000s) that have an effect on CO2 emissions per person?

```
co2_emissions_gather<- gather(co2_emissions, key="Country", value="CO2emissio
ns", 1:4)
co2_emissions_gather$Country<-as.factor(co2_emissions_gather$Country)
co2_emissions_gather$year <- as.numeric(co2_emissions_gather$year)
co2_emissions_gather$Year_cat <-cut(co2_emissions_gather$year,c(1900,1950,200
0,Inf), labels=c("Early1900s","Late1900s","Early2000s"))
```

Table 6 Descriptive Statistics

```
summary_nordic<-summary %>%
  filter(row.names(summary)%in% c("Sweden","Denmark","Finland","Norway"))



h_summary_nordic<-as_hux(summary_nordic)
h_summary_nordic  %>%
  set_align("center")%>%
  set_width(1) %>%
  theme_basic()
```

Two way ANOVA test

```
lm_co2_nordic_all <- lm(CO2emissions~Country*Year_cat,co2_emissions_gather)

Anova_co2_nordic_all<- Anova(lm_co2_nordic_all)
pander(Anova_co2_nordic_all)
```

Basic diagnostic plots

Figure 9 Basic diagnostic plots:CO2emissions~Country*Year_cat

```
par(mfrow=c(2,2))
plot(lm_co2_nordic_all)
par(mfrow=c(1,1))
```

Table 7 Two way Anova: CO2emissions and Country/Year catergory(Late1900s and Early 2000s)

```
Anova_co2_nordic_all<- Anova(lm_co2_nordic_all)
options(scipen=999)
pander(Anova_co2_nordic_all)
```

Table 8 Coefficient table: CO2emissions and Country/Year catergory(Late1900s and Early 2000s)

```
lm_co2_nordic_all <- lm(CO2emissions~Country*Year_cat,co2_emissions_gather)
lm_co2_nordic_all %>% as_flextable()
```

Table 9 Pairwise tests between the late 1900s and early 2000s within countries, adjustment = Holms test*

```
t3<-testInteractions(lm_co2_nordic_all, pairwise="Year_cat", fixed="Country",
adjustment="holm")
options(scipen=999)
t3 <-as_hux(t3)
number_format(t3) <- '%.5f'

t3 %>%
  set_align("center")%>%
  insert_column(c("Year_category","Late1900s-Early2000s: Denmark","Late1900s-
Early2000s : Finland","Late1900s-Early2000s :  Norway", "Late1900s-Early2000s
:  Sweden", "Residuals"))%>%
  theme_basic()
```

Table 10 Pairwise tests between countries within the late 1900s and early 2000s, adjustment = Holms test

```
t4<-testInteractions(lm_co2_nordic_all, pairwise="Country", fixed="Year_cat",
adjustment="holm")
options(scipen=999)
t4 <-as_hux(t4)
number_format(t4) <- '%.5f'

t4 %>%
  set_align("center")%>%
  insert_column(c("Country","Denmark-Finland :  Late 1900s","Denmark-Norway :
Late 1900s","Denmark-Sweden :  Late 1900s","Finland-Norway :  Late 1900s","Fi
nland-Sweden :  Late 1900s", "Norway-Sweden :  Late 1900s","Denmark-Finland :
```

```
Early 2000s","Denmark-Norway :  Early 2000s","Denmark-Sweden : Early 2000s","
Finland-Norway :  Early 2000s","Finland-Sweden :  Early 2000s", "Norway-Swede
n :  Early 2000s","Residuals"))%>%
  theme_basic()
```

Figure 10 Effect Plot: country and year_category(Late 1900s and Early 2000s)

```
plot(allEffects(lm_co2_nordic_all), x.var="Country")
```

Figure 11 Effect Plot: country and year_category(Late 1900s and Early 2000s)

```
data_nordic <- ggeffect(lm_co2_nordic_all, terms = c("Country", "Year_cat"))
plot(data_nordic)+
  scale_x_continuous(labels = c("Denmark", "Finland","Norway" ,"Sweden"), bre
aks = c(1, 2, 3,4))+
  labs(y = "CO2 emissions per preson")
```

Question 4 Are there any associations in attitude/awareness toward environment (intentional purchase of items with eco-friendly/less packaging) and working styles (work from home and people who work away from home) as well as age group, Generation Z,Young Millenials(23-26 years old), Core Millenials(27-32 years old), Mature Millenials(33-36 years old), Generation X, Baby Boomers ?

Table 11 Observerd Frequency (remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
eco = matrix(c(2084,1332,1307,1206),nrow=2)

dimnames(eco)<-list(reply=c("Intentionally buy items with eco-friendly packag
ing","Not intentionally buy items with eco-friendly packaging"), work=c("Work
from home","Work away from home "))

eco%>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","I intentionally buy items with eco-friendly/less p
ackaging", "I do not intentionally buy items with eco-friendly/less packaging
"))%>%
  set_width(1) %>%
  theme_basic()
```

Chisquare test (remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
result_eco<-chisq.test(eco,correct=FALSE)
result_eco
```

Table 12 Expected Frequency(remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
result_eco$expected %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  #set_caption("Expected frequency") %>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","I intentionally buy items with eco-friendly packag
ing or less packaging","I do not intentionally buy items with eco-friendly pa
ckaging or less packaging
"))%>%
  set_width(1) %>%
  theme_basic()
```

Figure 12 Bar chart (Remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
options(scipen = 999)
barplot(eco, beside=T, xlab="", ylab="Frequency",
        legend.text=T, args.legend = list(x=6, y=2700),
        ylim=c(0,1.2*max(eco)),col=gray.colors(2),las=1, cex.axis=0.7)
```

Table 13 Pearson’s residuals (remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
result_eco$residuals %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","I intentionally buy items with eco-friendly packag
ing or less packaging","I do not intentionally buy items with eco-friendly pa
ckaging or less packaging"))%>%
  set_width(1) %>%
  theme_basic()
```

Figure 13 Correlation Plot (Residual: remote/non-remote working and buying items with eco-friendly/less packaging intentionally)

```
corrplot(result_eco$residuals, is.cor = FALSE)
```

Table 14 Observed Frequency (Age and choosing items with eco-friendly packaging)

```
eco2 = matrix(c(653,707,513,420,953,635,505,414,1566,1281,497,478),nrow=2)

dimnames(eco2)<-list(reply=c("Yes", "No"),generation=c("Generation Z","Young
Millenials", "Core Millenials",    "Mature Millenials", "Generation X",    "B
aby Boomers" ))
#eco2<- t(eco2)
```

```
eco2 %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","I intentionally buy items with eco-friendly packag
ing or less packaging","I do not intentionally buy items with eco-friendly pa
ckaging or less packaging"))%>%
  set_width(1) %>%
  set_width(1) %>%
  theme_basic()
```

Chisquare test (Age groups and choosing eco-friendly packaging)

```
result_eco2<-chisq.test(eco2,correct=FALSE)
result_eco2
```

Table 15 Expected Frequency (Age groups and choosing eco-friendly packaging)

```
#result_eco2$expected_t <- t(result_eco2$expected)
result_eco2$expected%>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","Intentionally buy items with eco-friendly packagin
g","Not intentionally buy items with eco-friendly packaging"))%>%
  set_width(1) %>%
  theme_basic()
```

Figure 14 Bar chart (Age and choosing items with eco-friendly packaging)

```
options(scipen = 999)
par(mar=c(8,3,3,3))

barplot(eco2, beside=T, xlab="", ylab="Frequency",
        ylim=c(0,1.2*max(eco2)),col=gray.colors(2),las=2,cex.axis=1)
legend(0.5,1500,legend=c("Intentionally buy items with eco-friendly packaging
","Not intentionally buy items with eco-friendly packaging"), fill=c("black",
"grey"))
```

Table 16 Pearson’s residuals (Age groups and choosing items with eco-friendly packaging)

```
result_eco2$residuals %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Reply","Intentionally buy items with eco-friendly packagin
g","Not intentionally buy items with eco-friendly packaging"))%>%
```

```
  set_width(1) %>%
  theme_basic()
```

Figure 15 Correlation Plot (residual) for Age groups and choosing items with eco-friendly packaging

```
corrplot(result_eco2$residuals, is.cor = FALSE)

mtext ("Yes: Intentionally buy items with eco-friendly packaging
      No:  Not intentionally buy items with eco-friendly packaging", side=1,l
ine=1, at=2.5)
```