Exploration of Climate change effects on poverty

# by

# Alain Antonio

Abstract

The United Nation (2020) states that without drastic action today, adapting to the impact caused by the climate change in the future will be more difficult and costly. Climate change and poverty constitute major challenges facing mankind since the industrial revolution. There is no consensus among scholars on the link between poverty and climate change because it is hard to evaluate the damage and the cost caused by climate change. Tol (2009) argues that the amount and the intensity of research effort on the economic effects of climate change seems incommensurate with the perceived size of the climate problem, the expected costs of the solution, and the size of the existing research gaps. Politicians ﻿are proposing to spend hundreds of billions of dollars on green- house gas emission reduction, and at present, economists cannot say with confidence whether this investment is too much or too little. Further research and investigations are more than ever needed in order to serve as additional grounds for policy makers to prevent the worse.

Contents

[Abstract 2](#_Toc100422229)

[Project Specification 11](#_Toc100422230)

[Introduction 12](#_Toc100422231)

[Data Collection & Preparation 13](#_Toc100422232)

[Analysis and production of analytics 16](#_Toc100422233)

[Further Research 21](#_Toc100422234)

# Project Specification

The data science research project’s goal and objectives included;

1. Explore and explain the link between climate change events, agricultural production, , food security and poverty as well as
2. Show the relationship between climate change events, health and poverty

# Introduction

A data science research and analytics generation methodology was outlined and formulated. Amongst other things, it linked the data analytics work with the research goals and objectives and proposed a method for developing the components for producing analytics.

For more details see the document : Methodology Data Science and Analytics.doc

# Data Collection & Preparation

**The sample size was all 192 countries of the world.**

The data was collected electronically by visiting the relevant websites, selecting and downloading the relevant data.

The data pertained to indicators of economic performance for all countries over a period of 40 years up to 2019, indices and rankings for climate change events for all countries of the world were also obtained for 20 years up to 2017.

The key data sources were;

Data Source: [Global economy, world economy | TheGlobalEconomy.com](https://www.theglobaleconomy.com/)

We used this source mainly for data on economic indicators.

Data Source :

<https://www.germanwatch.org/en/cri>

and the from the World Health Organization on

<https://apps.who.int/nha/database/Select/Indicators/en>

After basic data preparation the data were 8100 rows with 26 variables each. Data preparation and feature engineering produced a data set of 20 000 rows of data with 49 variables.

The Data Dictionary in the Data Engineering report and the R Notebook provides more details on the details of the data used.

Data Integration

When we integrate poverty indicators data with climate events impact data

The logic of the left\_join function requires that country indicator data be duplicated in cases where there are

2 lines of climate events data.

**So after the integration each country will have m x n lines of data,**

**Where n = the number of distinct climate events lines, and m is the number of distinct lines of data for a country.**

Therefore during analysis, data must be filtered by the events year, to ensure that we explore indicators data that

goes with specific climate events data. If this filtering is over-looked there will be double counting and the statistics would be inflated and distorted.

Feature Engineering

Impact – it measures the level of impact from climate change events that a country experienced. Its is generated using the Climate Risk Index rank provided by the German watch website.

The formula used was;

Impact is Severe if the ranking was 20 or less

Impact is Bad if the ranking was between 21 and 40

Impact is Mild if the ranking was between 41 and 60

Impact is Low if the ranking was between 61 and 192

Food Security – it quantifies the food security level of a country.

The formula used was;

Food Security = Crop Production + Food Production + Livestock Production

Well Being – it measures the degree of poverty or the level of well being of a country

Well.Being = Food Security + Household Consuption + Health Expenditure per Capita + Savings

# Analysis and production of analytics

An analysis approach was outlined as follows;

1.Explore Poverty and food security levels

* Explore the global economy data to reveal levels of poverty
* Explore links between gdp and food security

2. Investigate poverty by sub-dividing using income groups, the goal was to see if high income countries could be performing better presumably because of better economics. We used the income groupings provided by WHO.

3. Investigate poverty and food security using categories of climate change events,

We wanted to see if some countries could be performing badly because of the impact of climate change events.

We engineered new features or variables from the climate change risk data provided by

<https://www.germanwatch.org/en/cri>

4. Explore Impact of climate change events

5. Investigate the relationship between climate events and poverty indicators

3.1 Aggregated poverty indicators in the search for predictive features

3.2 Integrate indicators with climate events and then

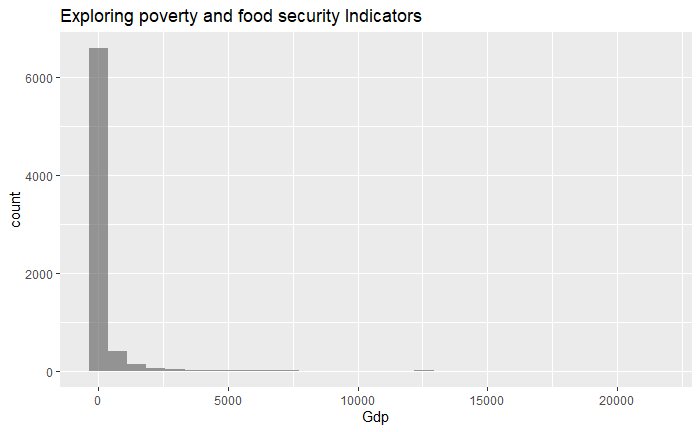
3.3 Explored using visualizations, correlations and charts

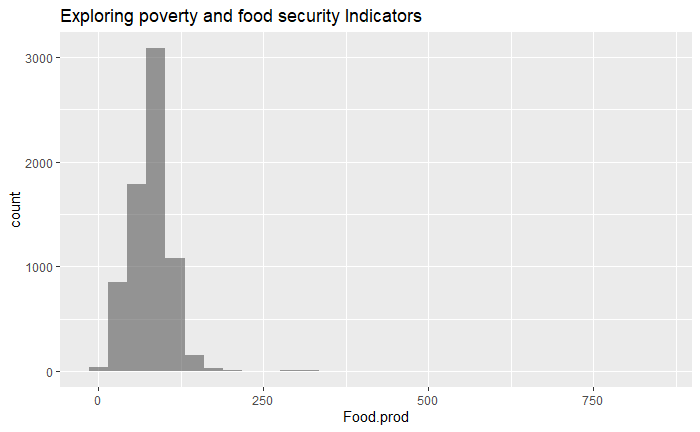
How to use the Climate Risk Index rank and score

The Global Climate Risk Index (CRI) developed by Germanwatch analyses quantified impacts of extreme weather events7 – both in terms of fatalities as well as economic losses that occurred – based on data from the Munich Re NatCatSERVICE, which is worldwide one of the most reliable and complete databases on this matter. The CRI examines both absolute and relative impacts to create an average ranking of countries in four indicating categories, with a stronger emphasis on the relative indicators (see chapter “Methodological Remarks” for further details on the calculation). The countries ranking highest (figuring in the “Bottom 10”8 ) are the ones most impacted and should consider the CRI as a warning sign that they are at risk of either frequent events or rare, but extraordinary catastrophes. The CRI does not provide an all-encompassing analysis of the risks of anthropogenic climate change, but should be seen as just one analysis explaining countries' exposure and vulnerability to climate-related risks based on the most reliable quantified data – along with other analyses. 9 It is based on the current and past climate variability and – to the extent that climate change has already left its footprint on climate variability over the last 20 years – also on climate change.

<https://germanwatch.org/sites/default/files/Global%20Climate%20Risk%20Index%202021_1.pdf>

Key Observations and Findings





Food Security & Poverty

The food security situation is depicted by crop production, food production, livestock production, health expenditure

percapita and even savings.

As you will see from the histograms they are all right skewed, suggesting that some countries spend very little on health, do very little on crop, food and livestock production.

But again, is that most countries have poor food security or is that some countries are doing so badly on poverty eradication and food security such that they drag the global mode and average down?

We need more analytics to explore and shed some light on those and other issues.

Globalization

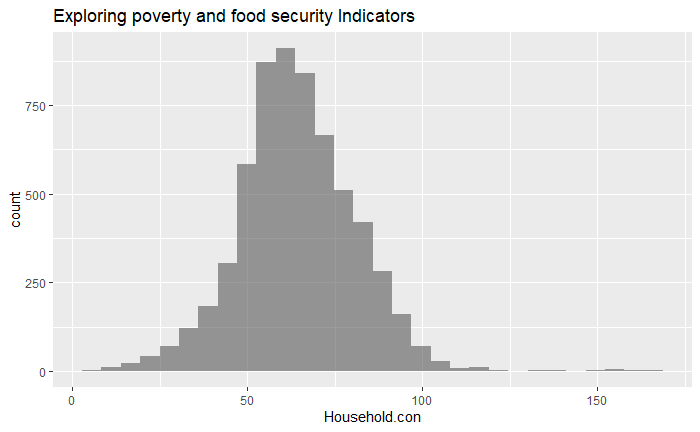
The data distribution for political and general globalization are multi-modal suggesting that there are different clusters of organizations with different globalization indices.

What are these clusters of countries and what globalization exposure do they experience?

Household Consumption

The data distribution suggests that there is little spread.

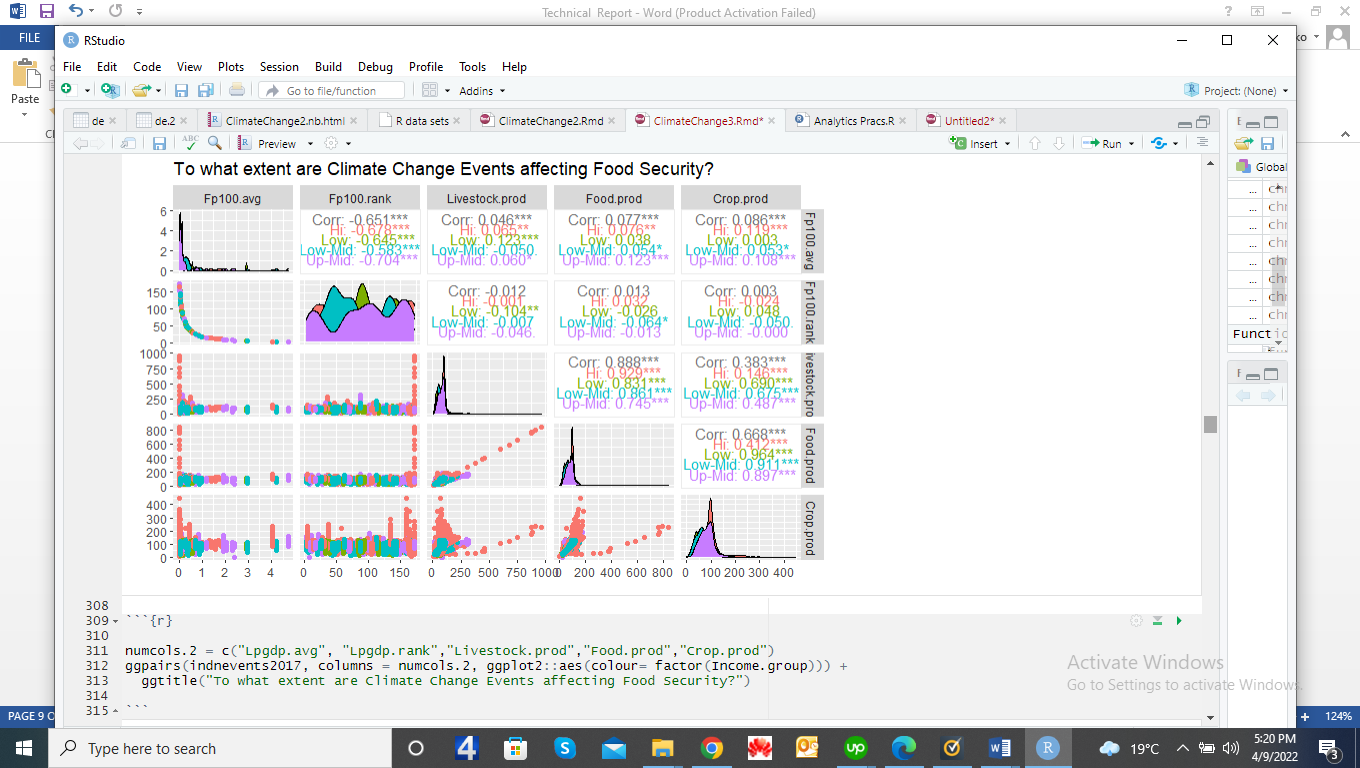
This could be because the data came from similar countries, like developed economies that have comprehensive data collection systems and procedures. Those kind of countries could experience similar household consumption levels.



Amongst other things the scatter plots show that

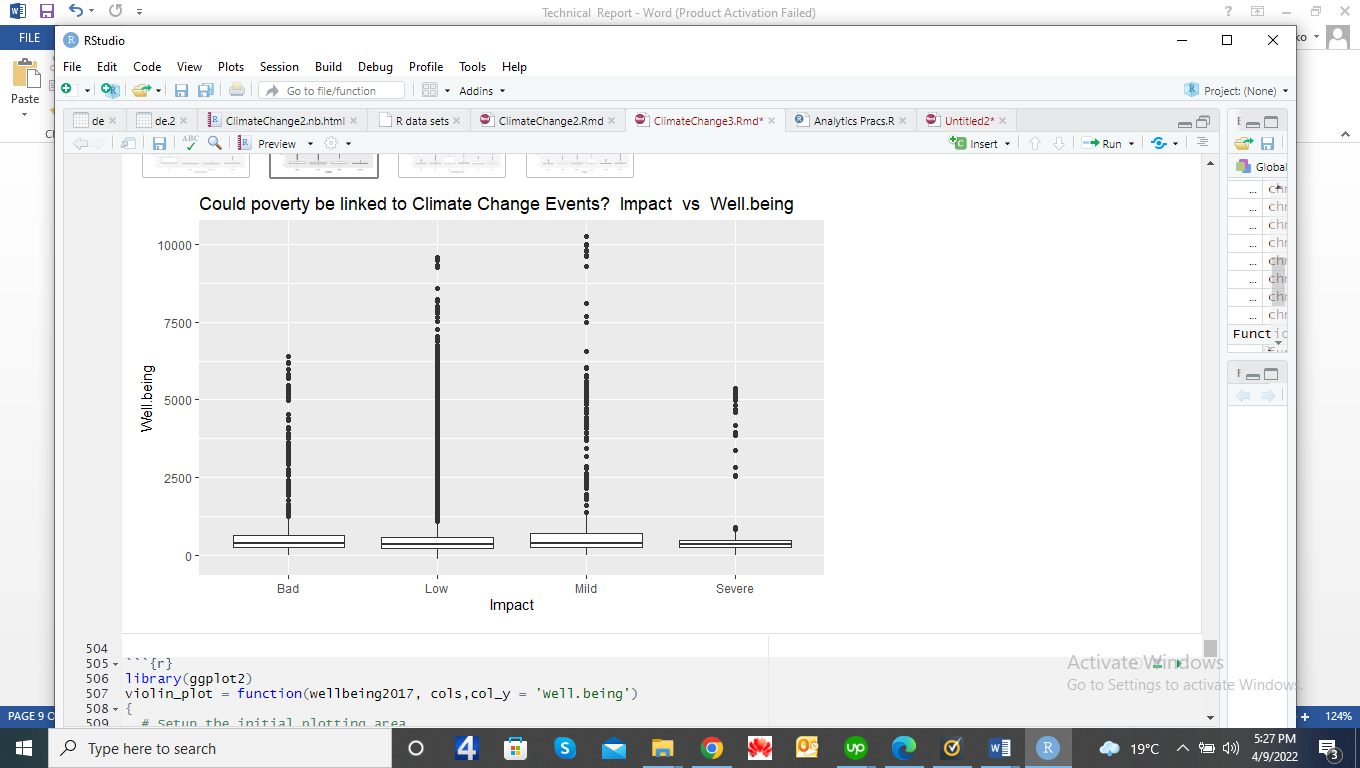
* Food, crop and livestock production are are being pursued more by spread in high and middle income countries.
* Even globalization is more prominent in high and middle income countries

To what extend are climate change events affecting poverty and food security?



The significant correlations above suggest a strong link between some poverty indicators and climate change events indices.

Even when we look at poverty or well-being according to the extent to which countries were impacted by climate change events, those recorded as having experienced a severe impact had highest poverty levels or the lowest well-being measurement.



# 

# Regression Modeling

# As shown in the scatter plots the predictors pf Co2Emissions are;

# Household Consumption

# Food Security

# Well being or poverty

# Gdp

# Livestock production

# Crop production

# Food Production

# Globalization

# Health.capita

# Population

# Urban Population

# Rural population

# The visualizations also show a collinear relationship between the following variables;

# Crop production

# Livestock production

# Food production

# For purposes of regression modeling these three variables have been summed to form a new feature, food security.

Regression Model for all countries

lm(formula = Co2Emissions ~ Household.con + Food.security + Well.being +

G.globalization + P.globalization + Gdp + Urban.pop + Rural.pop,

data = wellbeing2017)

Residuals:

Min 1Q Median 3Q Max

-2286273 -57825 -15041 35913 4680273

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.631e+08 2.228e+08 -1.630 0.1032

Household.con -1.246e+03 1.794e+02 -6.949 4.11e-12 \*\*\*

Food.security 1.677e+01 4.605e+01 0.364 0.7157

Well.being -8.863e+01 4.527e+00 -19.579 < 2e-16 \*\*\*

G.globalization -1.606e+03 6.023e+02 -2.666 0.0077 \*\*

P.globalization 2.462e+03 3.500e+02 7.036 2.21e-12 \*\*\*

Gdp 6.171e+02 7.002e+00 88.125 < 2e-16 \*\*\*

Urban.pop 3.631e+06 2.228e+06 1.630 0.1032

Rural.pop 3.632e+06 2.228e+06 1.630 0.1031

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 320400 on 5565 degrees of freedom

(822 observations deleted due to missingness)

Multiple R-squared: 0.6165, Adjusted R-squared: 0.6159

F-statistic: 1118 on 8 and 5565 DF, p-value: < 2.2e-16vs developped countries and the a quartile regression for the 100 most climate change effected.

Predictive Variables for all countries

As shown in the model output above, the most predictive variables are;

* Household Consumption
* Well being
* General Globalization
* Political Globalization and
* Gdp

Improving the model

When we remove the less significant features from the variable the model slightly weakens as shown by adjusted R-square which reduces from 61.59 % to 61.50 %

# Regression model for Developing Countries

# Call:

# lm(formula = Co2Emissions ~ Household.con + Well.being + G.globalization +

# P.globalization + Gdp, data = developing.countries)

# Residuals:

# Min 1Q Median 3Q Max

# -422825 -12319 -4332 5540 510697

# Coefficients:

# Estimate Std. Error t value Pr(>|t|)

# (Intercept) 13667.167 5359.305 2.550 0.0108 \*

# Household.con 86.495 49.209 1.758 0.0789 .

# Well.being -103.146 15.049 -6.854 9.29e-12 \*\*\*

# G.globalization 304.416 230.033 1.323 0.1859

# P.globalization 8.018 110.315 0.073 0.9421

# Gdp 1010.993 8.189 123.453 < 2e-16 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Residual standard error: 56670 on 2187 degrees of freedom

# (267 observations deleted due to missingness)

# Multiple R-squared: 0.8875, Adjusted R-squared: 0.8872

# F-statistic: 3451 on 5 and 2187 DF, p-value: < 2.2e-16

# The significant predictors are Well being and Gdp and the model improves over the model for all 192 countries with an Adjusted R-Squared of 0.8872

# Regression model for developed countries

# Call:

# lm(formula = Co2Emissions ~ Household.con + Well.being + G.globalization +

# P.globalization + Gdp, data = developed.countries)

# Residuals:

# Min 1Q Median 3Q Max

# -2236870 -96184 -16318 72690 4677095

# Coefficients:

# Estimate Std. Error t value Pr(>|t|)

# (Intercept) 367855.002 36639.681 10.040 < 2e-16 \*\*\*

# Household.con -2384.265 336.093 -7.094 1.58e-12 \*\*\*

# Well.being -75.026 6.029 -12.445 < 2e-16 \*\*\*

# G.globalization -8287.415 906.401 -9.143 < 2e-16 \*\*\*

# P.globalization 5397.078 585.694 9.215 < 2e-16 \*\*\*

# Gdp 594.622 9.071 65.551 < 2e-16 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Residual standard error: 401800 on 3375 degrees of freedom

# (555 observations deleted due to missingness)

# Multiple R-squared: 0.6121, Adjusted R-squared: 0.6115

# F-statistic: 1065 on 5 and 3375 DF, p-value: < 2.2e-16

# For the developed countries all variables are significant but the model performance is lowest with an Adjusted R-squared of 0.6115

# Hundred Countries most hit by Climate Change Events

# They have Climate Risk Index rank of 100 or less as 1 represents the most affected country and 192 the least affected.

# Call:

# lm(formula = Co2Emissions ~ Household.con + Well.being + G.globalization +

# P.globalization + Gdp, data = mostaffected.countries)

# Residuals:

# Min 1Q Median 3Q Max

# -2466090 -85976 -24261 55146 4538470

# Coefficients:

# Estimate Std. Error t value Pr(>|t|)

# (Intercept) 174866.054 31586.304 5.536 3.35e-08 \*\*\*

# Household.con -1911.207 318.151 -6.007 2.10e-09 \*\*\*

# Well.being -152.063 8.093 -18.788 < 2e-16 \*\*\*

# G.globalization -5425.646 851.172 -6.374 .11e-10 \*\*\*

# P.globalization 5397.957 605.567 8.914 < 2e-16 \*\*\*

# Gdp 641.155 9.490 67.564 < 2e-16 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Residual standard error: 410100 on 3137 degrees of freedom

# (424 observations deleted due to missingness)

# Multiple R-squared: 0.6359, Adjusted R-squared: 0.6353

# F-statistic: 1096 on 5 and 3137 DF, p-value: < 2.2e-16

# Panel Regression - All Countries

# Oneway (individual) effect Within Model

# Call:

# plm(formula = Co2Emissions ~ Gdp + Household.con + Well.being +

# G.globalization + P.globalization, data = allpaneldata, effect = "individual",

# model = "within", index = c("Year.x"))

# Unbalanced Panel: n = 39, T = 354-486, N = 17532

# Residuals:

# Min. 1st Qu. Median 3rd Qu. Max.

# -5196914 -67968 -21207 27344 5034133

# Coefficients:

# Estimate Std. Error t-value Pr(>|t|)

# Gdp 542.1517 3.2770 165.4403 < 2.2e-16 \*\*\*

# Household.con -1355.2134 103.1933 -13.1328 < 2.2e-16 \*\*\*

# Well.being -101.3997 2.7655 -36.6662 < 2.2e-16 \*\*\*

# G.globalization -2392.6117 285.7248 -8.3738 < 2.2e-16 \*\*\*

# P.globalization 3208.0281 193.5838 16.5718 < 2.2e-16 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Total Sum of Squares: 5.4459e+15

# Residual Sum of Squares: 1.9643e+15

# R-Squared: 0.63931

# Adj. R-Squared: 0.63842

# F-statistic: 6199.28 on 5 and 17488 DF, p-value: < 2.22e-16

# Panel Regression – Developing Countries

# Oneway (individual) effect Within Model

# Call:

# plm(formula = Co2Emissions ~ Gdp + Household.con + Well.being +

# G.globalization + P.globalization, data = developingcountries.panel,

# effect = "individual", model = "within", index = c("Year.x"))

# Unbalanced Panel: n = 39, T = 145-190, N = 6927

# Residuals:

# Min. 1st Qu. Median 3rd Qu. Max.

# -414448.9 -13546.8 -5426.9 6869.9 483074.3

# Coefficients:

# Estimate Std. Error t-value Pr(>|t|)

# Gdp 1013.3344 4.5435 223.0309 < 2e-16 \*\*\*

# Household.con -18.9422 26.9572 -0.7027 0.48228

# Well.being -2.4701 11.2584 -0.2194 0.82635

# G.globalization 218.4688 123.1688 1.7737 0.07615 .

# P.globalization 108.5265 59.0616 1.8375 0.06618 .

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Total Sum of Squares: 1.8565e+14

# Residual Sum of Squares: 2.0399e+13

# R-Squared: 0.89012

# Adj. R-Squared: 0.88944

# F-statistic: 11152 on 5 and 6883 DF, p-value: < 2.22e-16

# Panel Regression – Developed Countries

# Oneway (individual) effect Within Model

# Call:

# plm(formula = Co2Emissions ~ Gdp + Household.con + Well.being +

# G.globalization + P.globalization, data = developedcountries.panel,

# effect = "individual", model = "within", index = c("Year.x"))

# Unbalanced Panel: n = 39, T = 209-296, N = 10605

# Residuals:

# Min. 1st Qu. Median Mean 3rd Qu. Max.

# -4866416 -105492 -25459 0 65084 4993021

# Coefficients:

# Estimate Std. Error t-value Pr(>|t|)

# Gdp 524.3238 4.1823 125.367 < 2.2e-16 \*\*\*

# Household.con -2733.7258 193.9612 -14.094 < 2.2e-16 \*\*\*

# Well.being -98.3948 3.7526 -26.220 < 2.2e-16 \*\*\*

# G.globalization -9537.3755 550.5989 -17.322 < 2.2e-16 \*\*\*

# P.globalization 7027.8126 337.3436 20.833 < 2.2e-16 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Total Sum of Squares: 5.1664e+15

# Residual Sum of Squares: 1.8426e+15

# R-Squared: 0.64336

# Adj. R-Squared: 0.6419

# F-statistic: 3810.22 on 5 and 10561 DF, p-value: < 2.22e-16

# Quantile Regression – using All Countries Panel data

# 100/192 = 52.08 %

# Therefore the tau parameter was configured as 53th percentile so that we can capture the 100 most climate change affected countries

# P.globalization + Gdp, tau = 0.53, data = allpaneldata)

# tau: [1] 0.53

# Coefficients:

# Value Std. Error t value Pr(>|t|)

# (Intercept) -8366.58358 272.33637 -30.72151 0.00000

# Household.con -12.48425 2.27531 -5.48683 0.00000

# Well.being -17.99043 0.72326 -24.87400 0.00000

# G.globalization 360.09198 17.29460 20.82106 0.00000

# P.globalization 75.56353 14.82699 5.09635 0.00000

# Gdp 378.50988 12.46938 30.35514 0.00000

Panel data regression models – The panel data referred to all valid data available, and the panel regression models have higher Adjusted R-squared values than their corresponding Linear regression models.

Data Sets Created

|  |  |  |  |
| --- | --- | --- | --- |
| Name of Data Set | Content | Data frame produced | Use it for |
| ClimateChangeFileA.csv | Basic economic indicators data obtained from website | Climatechange |  |
| ClimateChangeFileB.csv | Additional economic indicators data obtained from website | Climatechangeupdates | Update basic indicators |
| Data frame  updclimatechange | All economic indicators merged for each country. |  |  |
| Climatedata.csv | A disk data set of the above data frame |  | To update with climate events data |
| ClimateEvents.csv | Climate Risk Indices | events | Integrate with Climatedata.csv |
|  |  |  |  |
| povertyncevents (a data frame) | Integrated economic indicators with climate risk events data |  |  |
| indnevents2017 | Economic indicators data that was integrated with climate risk events but only where the climate risk indices covered the 20 years upto 2017. |  |  |
| EindicatorsnEvents.csv | A disk version of the data frame indnevents2017 |  | To generate data frame wellbeing2017 |
| Wellbeing2017 | File EindicatorsnEvents.csv, with the Food Security and Poverty features generated during feature analysis and engineering |  | Whenever you restart your R, you can regenerate this data frame starting from the chucnk which reads in file EindicatorsnEvents.csv |
| Allpaneldata.csv | Economic indicators and climate change events data organized as Panel data. |  | Derive the other panel data from this file. |

Files attached with this report

1. ClimateChange3.Rmd – an R Notebook file with all the R code and analysis outputs

Use this file for details of visualizations, other analysis and the code used.

1. ClimateChangeFileA.csv – not changed from last time
2. ClimateChangeFileB.csv – not changed from last time
3. Climatedata.csv - changed
4. EindicatorsnEvents.csv – new file
5. Data Engineering Report.doc - because it has been updated
6. Methodology Data Science Methodology.doc – repeated for your convenience
7. ClimateEvents.csv
8. Allpaneldata.csv - Economic indicators and climate change events data organized as Panel data for all 192 countries.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*