Our Planet Is on Fire: Who Do We Look to for Guidance?

260720303

MGSC 401

June 27, 2020

Executive Summary

Our planet is warming at an alarming rate. You've probably heard it a million times by now, but it has gained more traction among all demographics in recent years as scientists have set deadlines for change. Most recently, it has been announced that we have until roughly 2030 before anthropogenic climate change becomes irreversible (UN, 2019). We are collectively facing the most dangerous form of the tragedy of the commons, threatening our very existence. There is no time for nations to shift the blame of climate change and avoid the battle against it. So, how can we create a global unified strategy to combat our impending doom? We need to hold nations accountable for their contributions to climate change, and collectively build solutions that can be employed at a governmental level.

In this paper, I find several nations in particular that are contributing a disproportionate amount to climate change, and what factors contribute to this, focusing more on economic factors. To measure a country's contributions to climate change I collected data on greenhouse gas emissions (GHGs) for each country that was publicly available from 1990-2018. To normalize for comparisons, I divided total GHG emissions of a country by their population to create a per capita measure. I aim to create a foundation from which nations can look to for guidance on limiting their impact on climate change. This foundation will focus on providing nations a point of reference to better themselves; Country A can look to a similar Country B to see what systems, policies, and economic structure they have in place that enable them to emit a lower level of GHGs.

First, I create a collection of visual representations of countries based on these chosen factors to identify simple similarities between countries - and to analyze their differences. Next, I build a model that can predict a country's greenhouse gas emissions for a given year, using a country and a year as a composite unique identifier for each data point (observation). This will provide insight on the amount of a country's greenhouse gas emissions that can be explained by the factors discussed herein. Following a simplistic representation of similar countries and a predictive model, I perform a cluster analysis using these same factors. The cluster analysis reduces the scope of where countries should seek guidance in creating environmentally friendly systems of change.

Further, I assess whether there is any correlation between the quality-of-life a country's residents enjoy and its GHG emissions. I aim to disparage any excuse that a country's high level of GHG emissions can be excused by the high quality of life its residents enjoy because of it. Quality of life is a hotly debated concept to quantify, but I decided on two basic forms which can be used as a proxy for it: GDP per capita and life expectancy. Though there are drawbacks to each of these as measures of quality-of-life, when used together they mitigate many of the disadvantages each holds individually.

Data Description

General Data Collection

I used the United Nations general data website to collect all of my data. Initially, I gathered data from each country from the years 1990 to the present day. Each of the variables that I had chosen had a separate dataset, so I merged and cleaned the datasets into one final set to be used in the subsequent analyses. Several countries were completely missing from some datasets, and some countries had data from a few years, while others had data for each year during that period. Through the process of merging and dumping observations with empty fields, thousands of incomplete observations were eliminated. Nonetheless, the final dataset had a competent 445 observations to be used in my analyses. Notably, the final dataset contains only countries in the northern hemisphere, from the years 2000 to 2015 - largely developed nations. There was not enough data available for countries in Africa, South America, and most of Asia. Fortunately, I planned to focus my analysis on developed nations regardless, as they are well known to emit more GHGs per capita than developing nations. Further, developed nations have a duty to lower GHG emissions more so than developing nations, which are typically more ill-equipped in the battle against climate change.

Greenhouse Gas Emissions

In this dataset, GHGs are measured in kilotonnes CO2 equivalents. CO2 equivalents is a measure that normalizes all greenhouse gasses into one metric using CO2 as the reference; this accounts for the different climate warming potential of each gas. Methane, for example, has a climate warming potential roughly 30 times greater than CO2, so, methane emissions must be multiplied by this factor in the calculation of a country's GHG emissions. It is important to note that this GHG data does not include GHG emissions from land use, land-use change, and forestry. To normalize between countries, I analyzed GHG Emissions per capita (Figure 1). The data was relatively normally distributed with a slight right skew.

Top 10% Income Share, Top 20% Income Share, Bottom 10% Income Share, and Bottom 20% Income Share

This data was collected to be used as a measure of economic well-being/equality. It is presented as a number representing the percentage of a nation's income that the top and bottom X% holds. I hypothesized that a country with a larger share of income in the top 10 and 20% would also be a country with higher emissions. Conversely, I hypothesized that a country with a relatively large share of income held by the bottom 10 and 20% would have relatively low GHG emissions. I believe unfettered capitalism results in ecological damage being treated as an externality, and its underlying ideology assumes capabilities for infinite growth, relying on the exploitation of earth's resources as much as possible. Minimally-regulated capitalism has also shown us that money usually accumulates at the top; you need more money to make more money, so the top 10 and 20% of the population should hold a relatively high share of the nation's income. Based on the aforementioned principles, I hypothesized that there would be a positive correlation between higher income share of the top X% and GHG emissions per capita, as well as a negative correlation between the bottom X% and GHG emissions. Interestingly, there was no correlation between

GHG emissions and any of these measures (Figure 3). The top X% and the bottom X% Income share were highly correlated with their respective counterparts. These variable's data are all relatively normally distributed (Figure 1). The summary statistics can be seen in Figure 2.

Population

The distribution of population is right skewed, with a majority of observations holding populations below 50 Million. (Figure 1). This data was mainly collected to create per capita fields where necessary.

Electricity Usage

Electricity usage of a nation is measured by the total kilowatt hours of electricity consumed within a specific year, in millions. To normalize for differences in population, I transformed this into a per capita metric to be used in my analyses. I hypothesized that the electricity usage of a nation would have a significant impact on its GHG emissions. At a glance, we see that there is a weak correlation (0.37) between the two (Figure 3). Electricity usage per capita has a right skewed distribution (Figure 1).

Life Expectancy

Life expectancy was collected to be used as one measure of quality-of-life in a country; the longer their life expectancy, the better their quality of life. This was also gathered to be used in my predictive model, so long as it increased the predictive power. This data was normally distributed (Figure 1).

Gross Domestic Product (GDP)

GDP of a country was collected as an indicator of both economic well-being and quality-of-life, meant to be used in tandem with life expectancy when judging quality-of-life. It is measured in constant 2011 USD (adjusted for inflation). I hypothesized that GDP would have a strong positive correlation with GHG emissions, as I believe economies worldwide still tend to be reliant on exploitative industries. To normalize across countries with different populations, I used GDP per capita in my analyses. At a glance, we see that there is a moderate positive correlation between GDP per capita and GHG emissions per capita, at 0.55 (Figure 3). GDP per capita is relatively normally distributed among observations, with a slight right skew (Figure 1).

So, Where Does Each Country Rank?

Before I build a more complex model, I created several visual interpretations of a country's GHG emissions alongside several other factors, and how they stack up next to their global counterparts. Because my initial dataset included a country-year combination as a unique identifier for an observation, it would be too convoluted to visually interpret on a plot - so, I created another data frame that took the average of each country's emissions over the number of years of data each country had available.

First, I created a global heat map based on GHG emissions per capita, to offer a simple visualization of which countries are the highest and the lowest GHG emitters (Figure 4). At a glance, we can see that European countries tend to emit much less than North America and Australia, while keeping relatively consistent with Japan. Within Europe, we can see that Turkey and Latvia are doing considerably well in this regard, and Luxembourg is doing much worse than its European neighbours (Figure 4).

Next, I created some visually intuitive plots to understand where each country stands with regards to quality-of-life and GHG emissions per capita. In Figure 5, we can see a 2-dimensional plot with per capita GHG emissions on the X-axis, GDP per capita on the Y-axis, and population represented as the size of the data points (flag size), for basic context on the population that each nation must provide for. GDP per capita presents a useful metric for economic well-being of citizens, though it is definitely not a catch-all statistic for a nation's well-being. A good way to interpret this model is by focusing on the upper left area relative to any one country you choose to analyze at a time - these countries serve as guidance to the country in question. These nations are making more money per citizen than said country, while also better mitigating their impact on the climate. Notably, living costs vary greatly across different nations, so it is important to complement this statistic with the life-expectancy of a nation to better encapsulate the quality-of-life in a nation.

In Figure 6, I include both life expectancy and GDP per capita in the plot with GHG emissions per capita. In this case, I use life expectancy as the Y-axis, and base the size of the data point (flag size) on the GDP per capita of a nation. To interpret this, I suggest you once again take a look at one country at a time and look at the (general) upper left section of the plot relative to that country. This will offer insight on which fellow countries provide a lengthy life for their citizens, while not being reliant on high-GHG-emitting industries and lifestyles to achieve this. While looking into the general upper-left hand section relative to a country, one can complement this insight with the flag size: if it is of similar size, their citizens (on average) enjoy a similar, or better, quality of life than the citizens of the country in question.

Predictive Model

My analysis in this paper is not heavily reliant on the predictive model. Rather, the model is used to supplement our analysis and give us insight on the importance of each of the aforementioned factors with regards to the GHG emissions of a country. Thus, I wanted to build a model with the strongest predictive power, rather than a model with a high degree of interpretability. With these criteria in mind, I decided to employ a random forest model for my predictions. With the model choice decided, I developed a new set of criteria for which to base my predictor inclusion on. I would only include variables in the model that would increase the percentage of variance explained by the model by 1% or more. I also had one exception to variable inclusion: I would not use population in my model, because when I provide my later recommendations, I do not want population size to be a possible justification for a country's high per capita emissions.

After testing several combinations and incrementally adding one predictor after another, I had arrived at my final model. The random forest model, predicting GHG emissions per capita for each observation, includes: GDP per Capita, Life Expectancy, Top 10% Income Share, Top 20% Income Share, Bottom 10% Income Share, Bottom 20% Income Share, and Electricity Usage per Capita.

I had tested the predictive difference between a model with 500 trees and one with 10,000, and found that the difference was negligible, so I ended up staying with the much less computationally complex number of trees. The model managed to explain 88 – 89% of the variance in the observations, while achieving a minute mean error percentage of 6.9%. To be clear, the mean error percentage is calculated by taking each residual error as a percentage of the actual value, squaring it, then taking the square root. I believe that this is more easily understood than simply including the MSE, because the values of GHG emissions per capita are very low decimals.

I then analyzed the importance of each predictor in this random forest (Figure 7). We can see that the most important predictor is electricity usage per capita, followed by life expectancy, and GDP per capita. This indicates that these three factors are most highly impactful or impacted by GHG emissions per capita. The top and bottom X% income share variables hold a similar level of importance, though the bottom X% Variables are more important than the top X% Variables. Despite these findings, it is still difficult to identify a causal relationship between these factors and GHG emissions, as there may in fact be reverse-causality.

After recording all GHG emissions predictions in the data frame, I created another world map to visualize predicted GHG emissions (once again taking the average prediction for each country in the set), giving us the ability to compare and contrast with the initial map I created using actual emissions values (Figure 4 & Figure 8). At a glance, we see that, as the low mean prediction error indicates, the random forest model provided accurate predictions on each country's GHG emissions per capita. To better understand which countries were predicted more accurately, I created a bar chart with the average error over the years available for each country (Figure 9). The countries with red bars were under-predicted, indicating that, relative to the other countries studied, they are doing a poor job with GHG emissions, based on the factors used in the model. Conversely, the countries with green bars were over-predicted by the model. This means that, relative to the other countries being studied, they are doing a good job with their GHG emissions, and they were actually expected to emit at a higher rate based on the data from all of the countries in the set.

How Do the High-GHG-Emitting Countries Better Themselves?

This paper is not designed to give an in-depth analysis on the specific environmental policies, economic systems, or political systems that work best for lowering GHG emissions. Rather, the goal of this paper was to point country A in the direction of countries B and C, leading the charge in low GHG emissions, to learn from in the battle against climate change. Humans have always thrived under collective intelligence and cooperation, and there should be no shame in one country seeking guidance or help from another. Countries like the US, Australia, and Canada need to look to European countries and learn from their social, political, and environmental policies to start employing their own strategies to reduce GHG emissions. This paper also aims to put to rest the common misconception that quality-of-life is a trade-off in the fight against climate change. These often come up in countries with economies that are heavily reliant on non-renewable energy, such as oil and gas. If there are several countries providing a great quality-of-life to their citizens while maintaining a relatively low level of GHG emissions, they should be used as the gold standard for developed nations.

In the first visual analyses (Figures 4, 5, & 6), I provided a basis of reference for countries to refer to one another in lowering their GHG emissions. Another way to refer to countries doing "better" than you is by geographic location. Though in a small sample size, it seems that countries in close geographic proximity to one another tend to have similar levels of GHG emissions. Likely due to economic interdependence, political interdependence, or the natural surroundings that these locations utilize. From this perspective, a country can use nearby, lower-GHG-emitting countries for reference when employing a strategy. However, this ignores the many nuances that neighbouring countries have between each other, which motivated me to provide countries with a better understanding of which other countries are most similar to them. This will give nations a good stepping-stone so the shift to lower emissions is easier to make - as opposed to using a country with a completely unique economic/political system for reference.

To find these groups of similar countries, I employed a hierarchical cluster analysis using the same factors I had previously used in my random forest model, without GHG emissions per capita. That is, I used top and bottom X% income share, life expectancy, and GDP per capita. With any luck, this would group together countries that are economically and politically similar, and therefore the stepping stones in progress are made easier for countries trying to lower their emissions; they have a better idea of which countries to look at for guidance.

I tested anywhere between 3 and 5 clusters, and found that with only 3 clusters, nearly all of the countries were grouped into one cluster, while with 4, there were two similarly large clusters, and two clusters with one or two countries each. With 5 clusters, the 5th cluster also only had two countries. Taking all of this into account, I opted to use 4 clusters. The colour-coded dendrogram and mean statistics for each cluster can be seen in Figure 10. The first cluster can be seen on the right of the dendrogram in blue. This cluster holds a relatively high GDP per capita on average, with a high life expectancy, and a medium level of emissions. In this cluster, countries to look at for guidance when lowering emissions include Sweden, France, and Italy, all boasting a high quality-of-life for their citizens while keeping GHG emissions relatively low. The second cluster can be seen on the left of the dendrogram in purple.

This cluster is made up of lower income countries with a lower life expectancy and very low electricity usage per capita. In this cluster, countries to use as referential examples for low GHG emissions while maintaining a decent quality-of-life are Portugal, Malta, and Croatia. The third cluster can be seen in black in the middle of dendrogram and is only made up of Luxembourg. Luxembourg is an outlier in this analysis, so this is not very surprising. As a very wealthy country with a very high level of GHG emissions, Luxembourg can refer to the same countries as cluster 1 and 4 when looking for guidance to lower emissions. Cluster 4 can be seen in Red in the dendrogram and is only made up of Norway and Switzerland. These two countries are both doing well in terms of quality-of-life and GHG emissions, however Switzerland is doing exceptionally well, and should be used as a benchmark for Norway to reach.

Key Takeaways

This analysis can be used as a basic foundation for countries to refer to when looking for guidance in the fight against climate change. This paper also finds that the relatively basic factors used in this analysis were able to explain a majority of the variance in observations between countries, indicating that a country's policies relating to these factors (electricity usage, income equality, GDP per capita, and life expectancy) are important to lower GHG emissions. Despite the capabilities of this model in explaining GHG emissions, there is still a great deal of nuance in a country's actions and its effect on GHG emissions. As such, these analyses are better used to pinpoint which countries should be used as a guide for the rest of the world during this crisis. The political, social, and economic systems in countries like Switzerland, Sweden, France, Spain, and Italy can teach us a great deal about how to handle this crisis; on average, these countries boast a great quality-of-life for their citizens while maintaining relatively low levels of GHG emissions.

How Do We Build on This?

More in-depth analysis is needed to build applicable governmental strategies to solve the climate crisis. Further analysis on the distribution of energy in a country between renewables and non-renewables can provide much needed insight. Analyzing a country's various economic systems would also help create an actionable strategy. Building a more complex model with more factors to encapsulate and quantify these systems, while making each factor interpretable, would be of great value. For example, how reliant is a country's economy on oil and gas or animal agriculture (all known contributors to climate change)? Once this has been assessed, looking into how a country can slowly restructure away from these industries to cleaner ones, once again using other countries who have gone/are going through similar situations as a reference point, are much needed continuations to this analysis.

Citations

"Only 11 Years Left to Prevent Irreversible Damage from Climate Change, Speakers Warn during General Assembly High-Level Meeting — Meetings Coverage and Press Releases." United Nations, United Nations, 28 Mar. 2019, www.un.org/press/en/2019/ga12131.doc.htm.

United Nations data: A world of information http://data.un.org/Default.aspx

Appendix

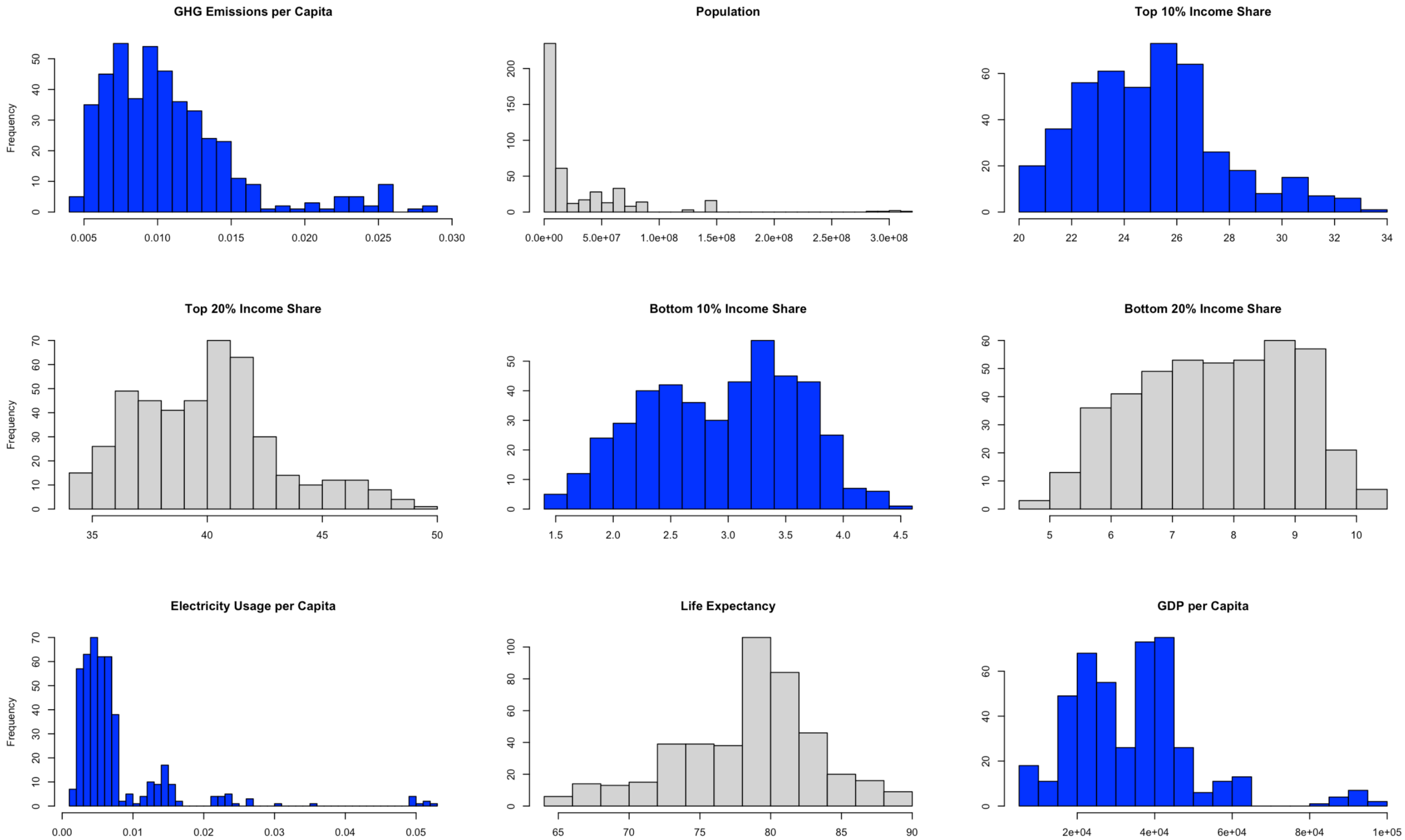


Figure 1: Variable Histograms

Summary Statistics

Statistic	N	Mean	Standard Deviation	Minimum	25th Percentile	75th Percentile	Maximum
Top 10% Income Share	445	25.1	2.7	20.1	23.0	26.4	33.5
Top 20% Income Share	445	40.0	3.1	34.0	37.5	41.7	49.2
Bottom 10% Income Share	445	3.0	0.7	1.5	2.5	3.5	4.5
Bottom 20% Income Share	445	7.8	1.3	4.9	6.8	8.7	10.5
Population	445	28,746,928	44,642,561	287,952	4,415,872	45,429,076	316,400,538
GHG Emissions per Capita	445	0.0107	0.0046	0.0044	0.0074	0.0126	0.0284
GDP per Capita	445	34,122	16,305	5,651	22,544	42,412	98,911
Life Expectancy	445	78.6	5.1	64.6	75.0	81.8	89.0
Electricity Usage per Capita	445	0.0076	0.0077	0.0016	0.0036	0.0072	0.0529

Figure 2: Summary Statistics

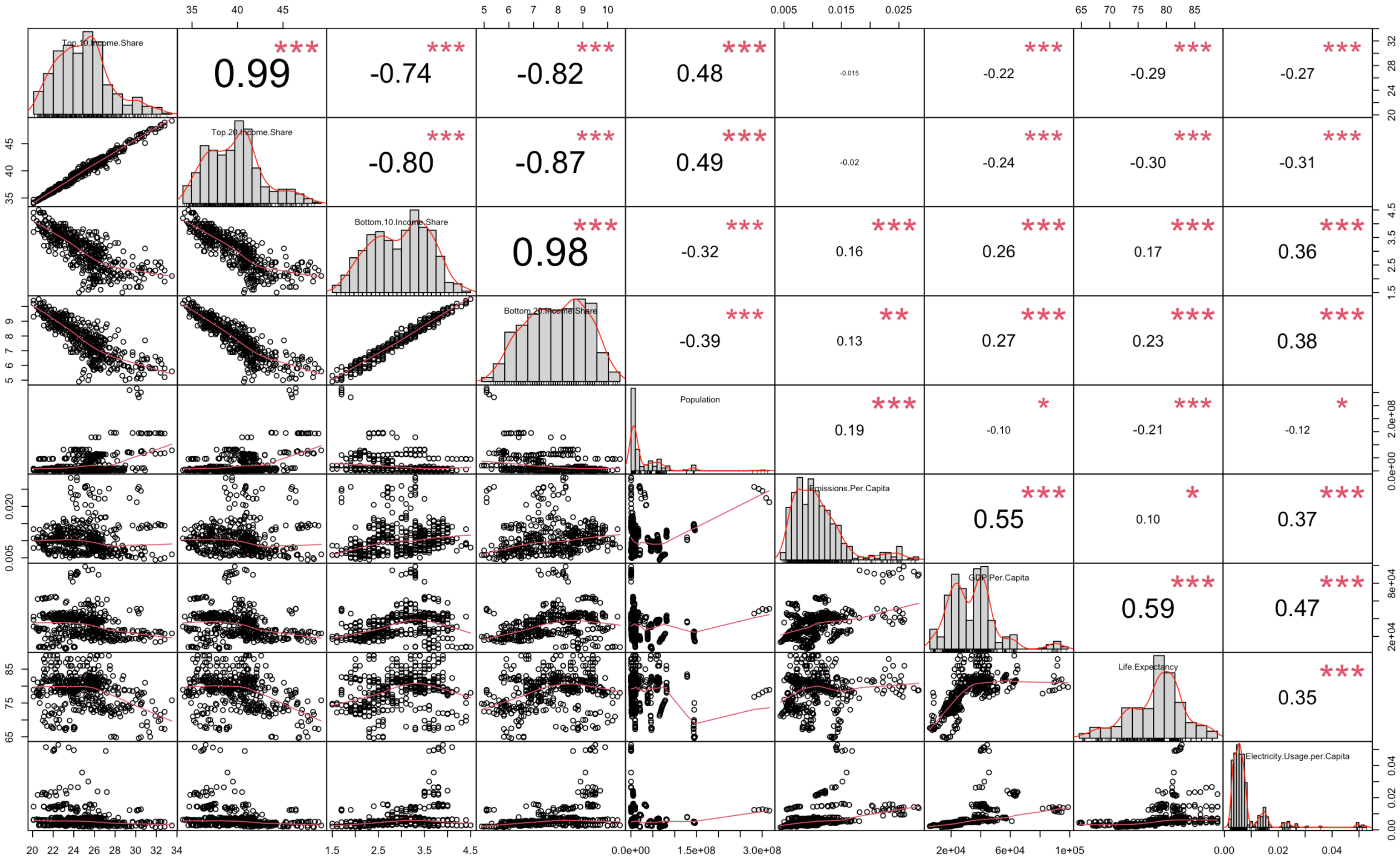


Figure 3: Correlation Matrix

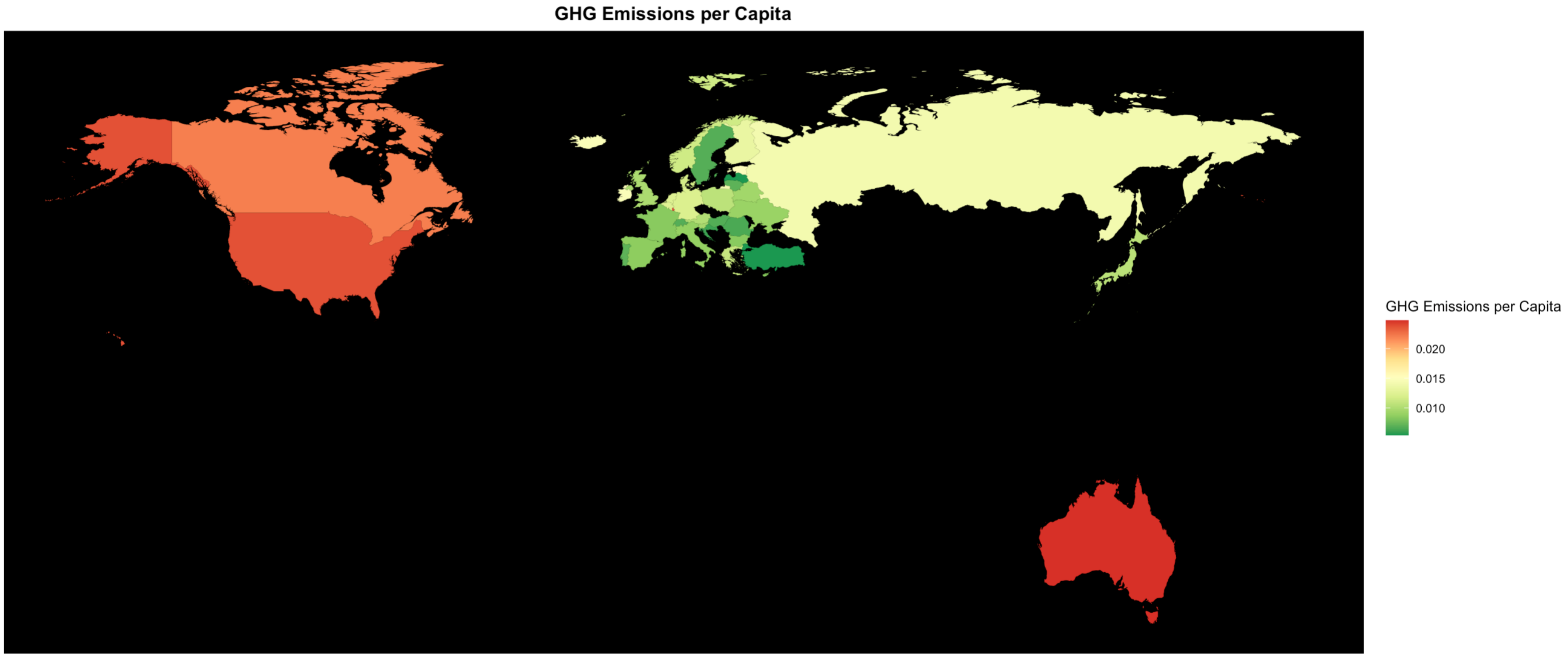


Figure 4: GHG Emissions per Capita

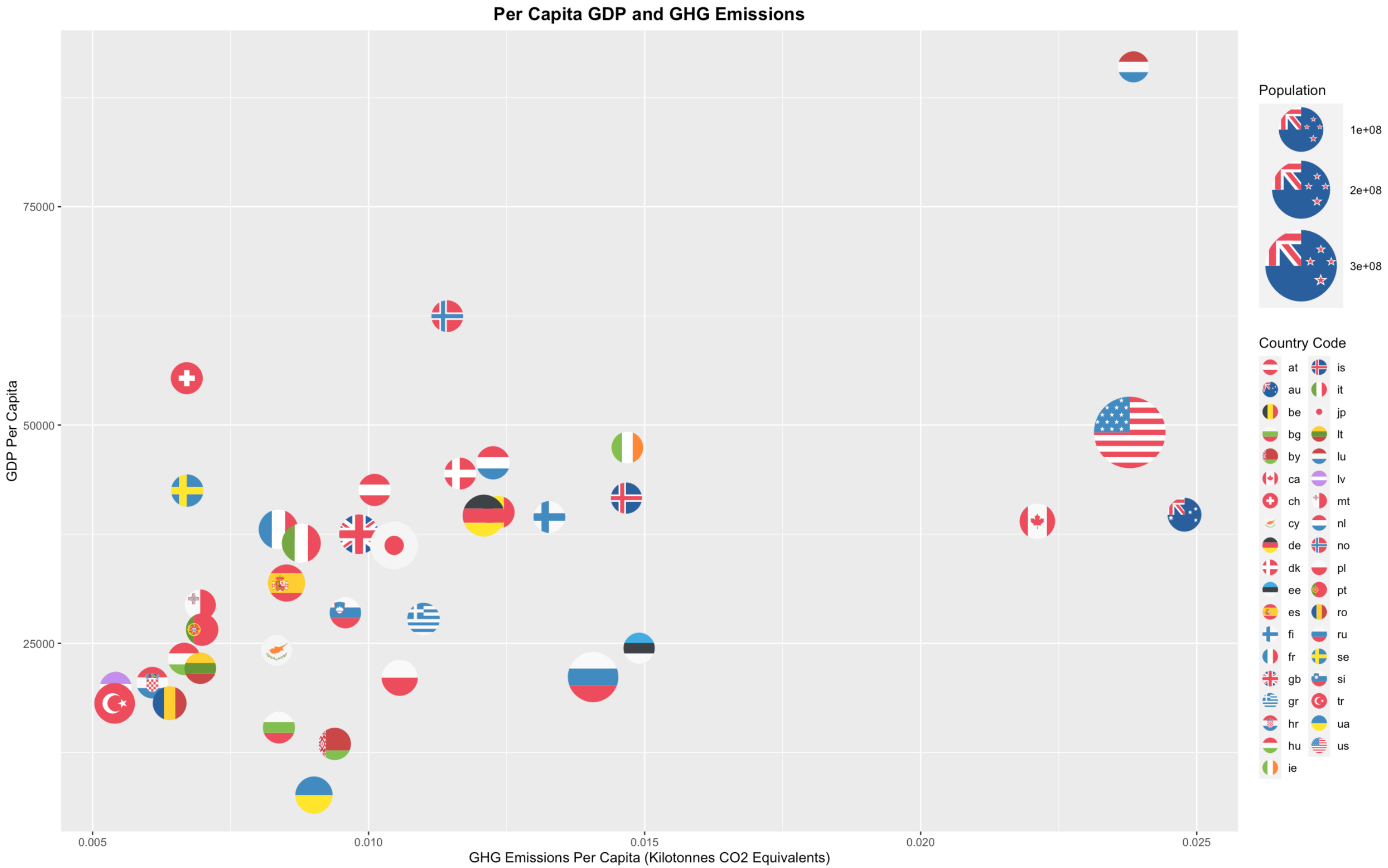


Figure 5: GDP per Capita and GHG Emissions: Population as Size

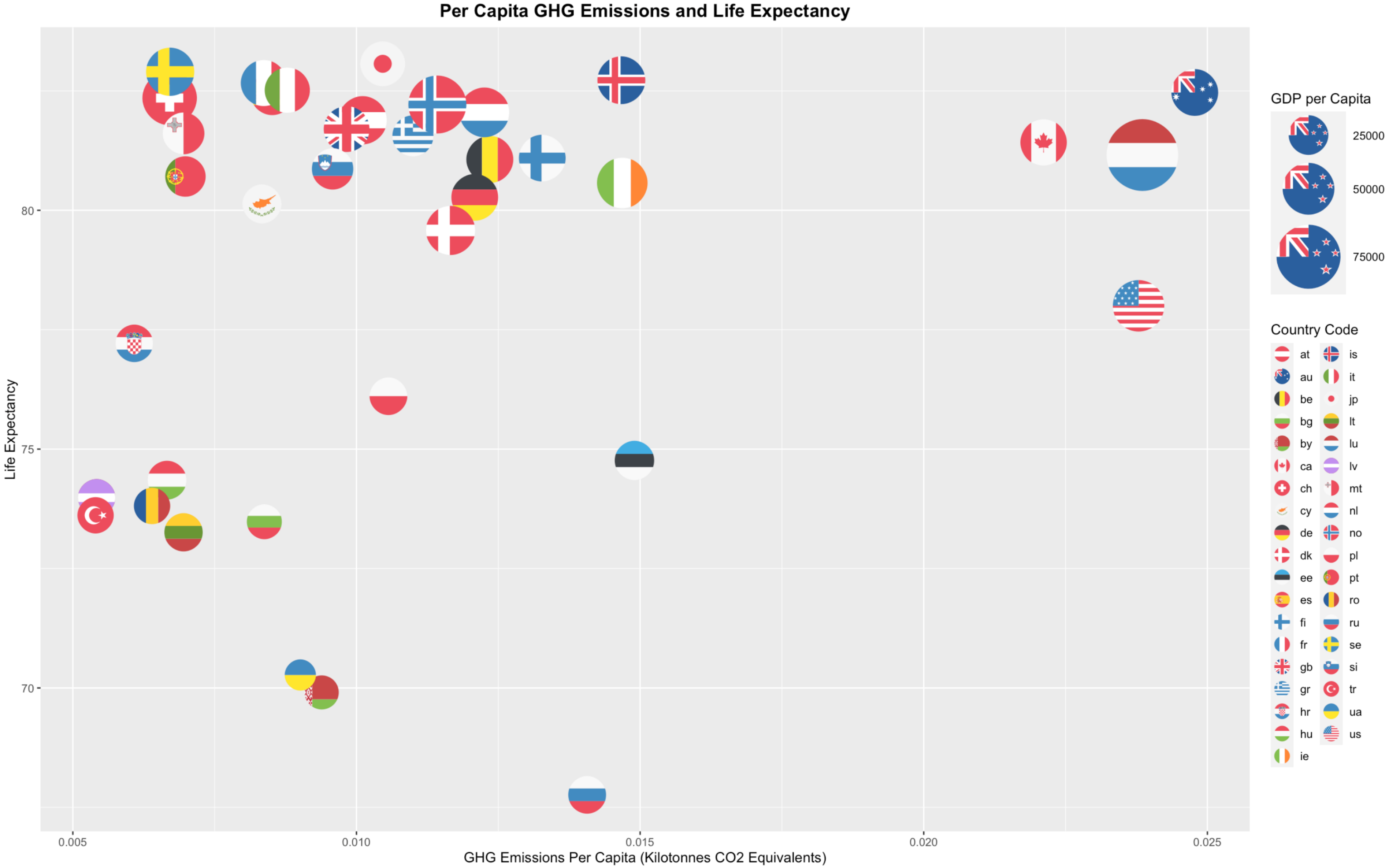


Figure 6: Life Expectancy and GHG Emissions: GDP per Capita as Size

Figure 7: Variance Importance Plot

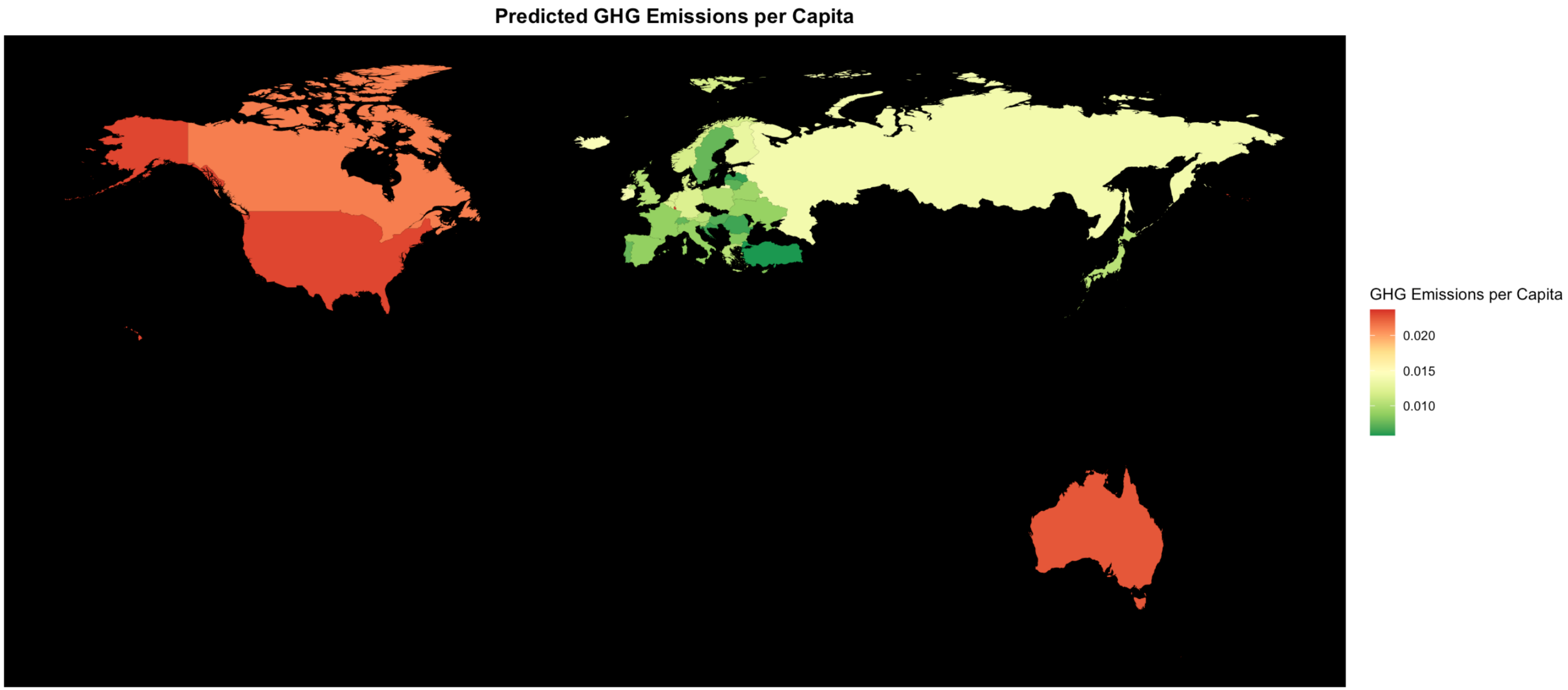


Figure 8: Predicted GHG Emissions per Capita

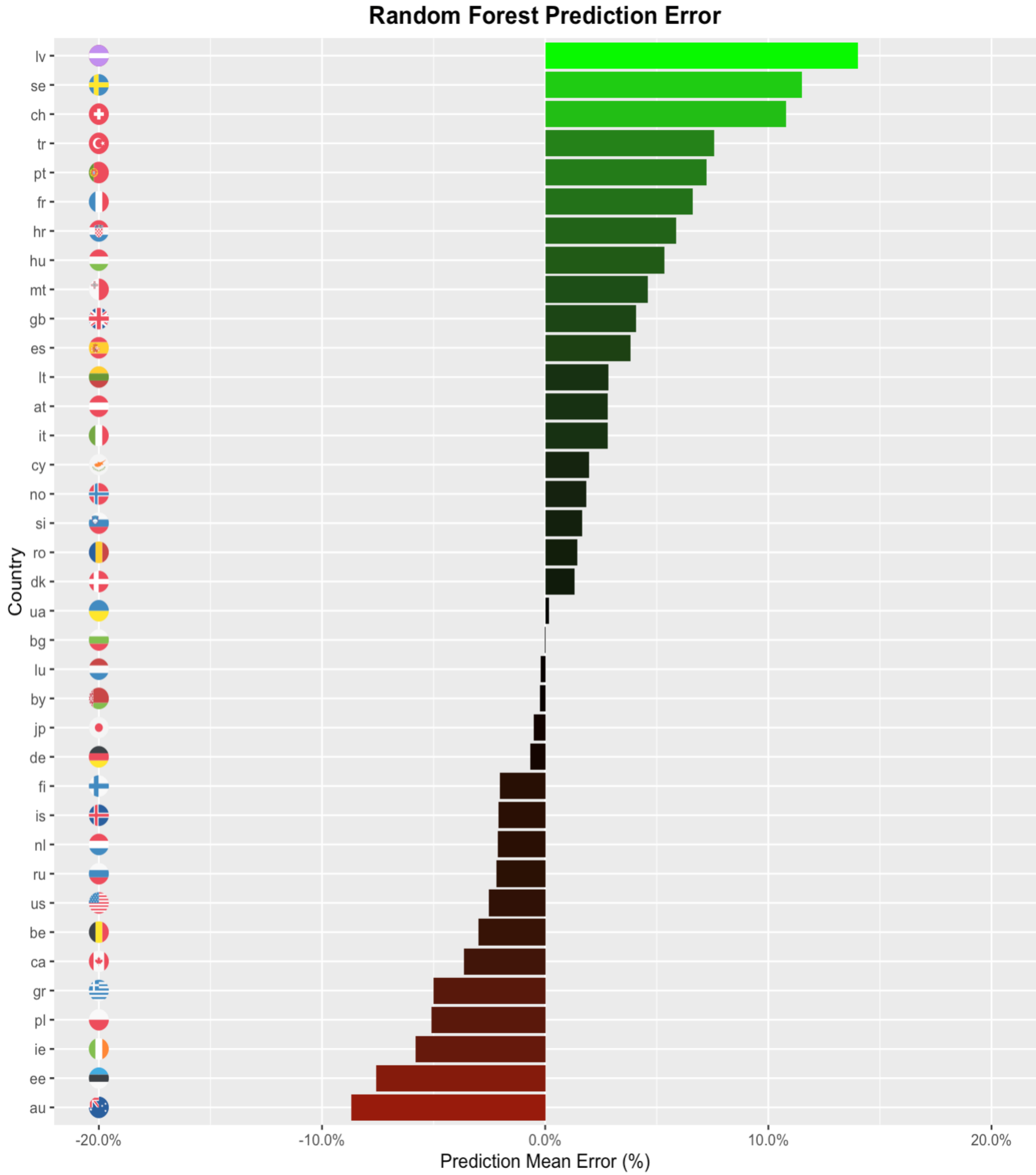
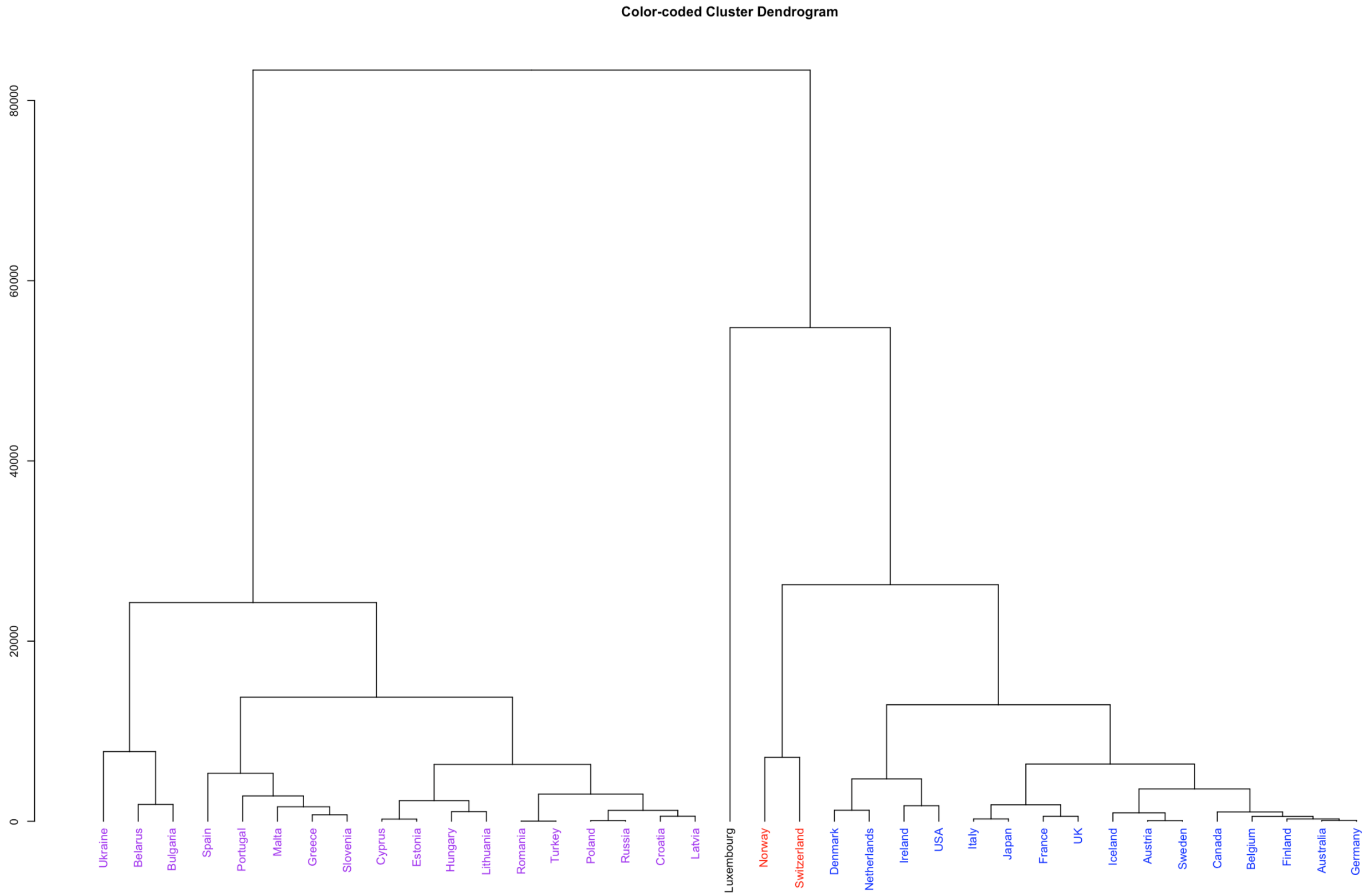


Figure 9: Mean Prediction Error by Country: As a %



Mean by Cluster

Statistic	1 (Blue)	2 (Purple)	3 (Black)	4 (Red)
GHG Emissions per Capita	0.0135	0.0086	0.0239	0.0091
GDP per Capita	41,217	21,869	91,012	58,917
Life Expectancy	81.5	75.9	81.2	82.3
Electricity Usage per Capita	0.0107	0.0038	0.0128	0.0151

Figure 10: Cluster Statistics & Dendrogram