Final assignment – CO2 Emissions

For Winc Academy – Data Analytics with Python

Hugo Maat – 14/3/2022

Preface

Emission of greenhouse gasses, like CO_2, is a global concern because of the environmental effects. Most governments, national and otherwise, take this problem very seriously. There are however no simple solutions, as is usually the case with worldwide issues. In order to reduce emissions and quite literally turn the rising tide, we need to better understand the problems of carbon emissions.

This report examines three statistical questions pertaining to carbon emissions. The first question is: what is the best predictor for CO_2 emission per capita in a country? We are not looking for a cause here, that would be a very difficult question to answer. Secondly: which countries have made the biggest strides in reducing emissions? Knowing which countries have reduced emissions already may after all show the path for others to follow. The third and final question is about alternatives: which non-fossil fuel energy technology is going to have the best price in the future?

It turns out that *price* is a key when we consider the results of these questions. In all but the most recent history, carbon emissions have been the result of economic prosperity. Accordingly, reduction in emissions is as often a result of economic devastation as it is the product of environmental policy. It is difficult to persuade countries to lower carbon emissions if the same countries that pollute are the countries that prosper.

Contents

The sources and methods used for each of these questions are discussed together per question in turn. Different datasets and approaches were used in each instance. The results, discussion and conclusion at the end are also divided into three parts.

The data used in this report was retrieved from the OurWorldInData website. These datasets in turn were based on various academic and government publications. All images used were published under creative commons license.

Question 1: Predictors for per capita emissions

Data

To determine the best predictors for CO_2 emissions per capita I have made use of several different datasets. CO_2 per capita is the dependent variable, and this dataset was sourced by OurWorldInData from the Global Carbon Project. The data goes back as far as 1750 for certain regions, but for this question only the data from 1970 onwards was used. Many of the independent variables considered did not have as much historical data, and a longer timeline would have undervalued countries that were late to industrialize.

Several potential predictors were selected using domain knowledge. The initial list included: Gini coefficient; GDP per capita; meat production/consumption; dairy production/ consumption; motor vehicle ownership; per capita electricity consumption; primary energy source; urbanization. These datasets were used as the independent variables.

The emissions per capita dataset was filtered to exclude some geographical entities that did not qualify as countries. It was not necessary to apply the same kind of filter on the independent variable datasets. These sets were merged with the emissions per capita dataset on matches in country and year, so mismatches were excluded.

Method

Two methods were used to determine predictors. The first was to calculate Pearson correlations for combinations of emissions per capita and one of the independent variables. For each potential predictor the calculation was made for the entire dataset, for the dataset grouped by year, and for the dataset grouped by country. The latter two were averaged out. This produced three correlation coefficients, labeled 1, 2 and 3. Coefficients 2 and 3, because these are based on a set of measurements, have other statistical information as well. These distributions have a standard deviation, for instance. Coefficient 2 is the result of a comparison of the dependent and independent variables in countries with each other in the same time frame. Coefficient 3 compares these variables within the same country over time using historical data.

The second method was a linear regression of the dependent variable and one of the independent variables. The regression was trained on $1/10^{th}$ of the combined dataset. The order of rows in the dataset was randomized beforehand so that the fit would not be biased towards countries that started with ‘A’. The other $9/10^{th}$ was predicted using the training data. For each of the independent variables the R^2 was calculated between the predicted values and the actual distribution of data points. The higher this value would be, the more accurate a predictor the independent variable would be. Figure 1 contains some visual representations of these linear regressions, where the scatter plot is the actual data points and the red line represents the prediction.

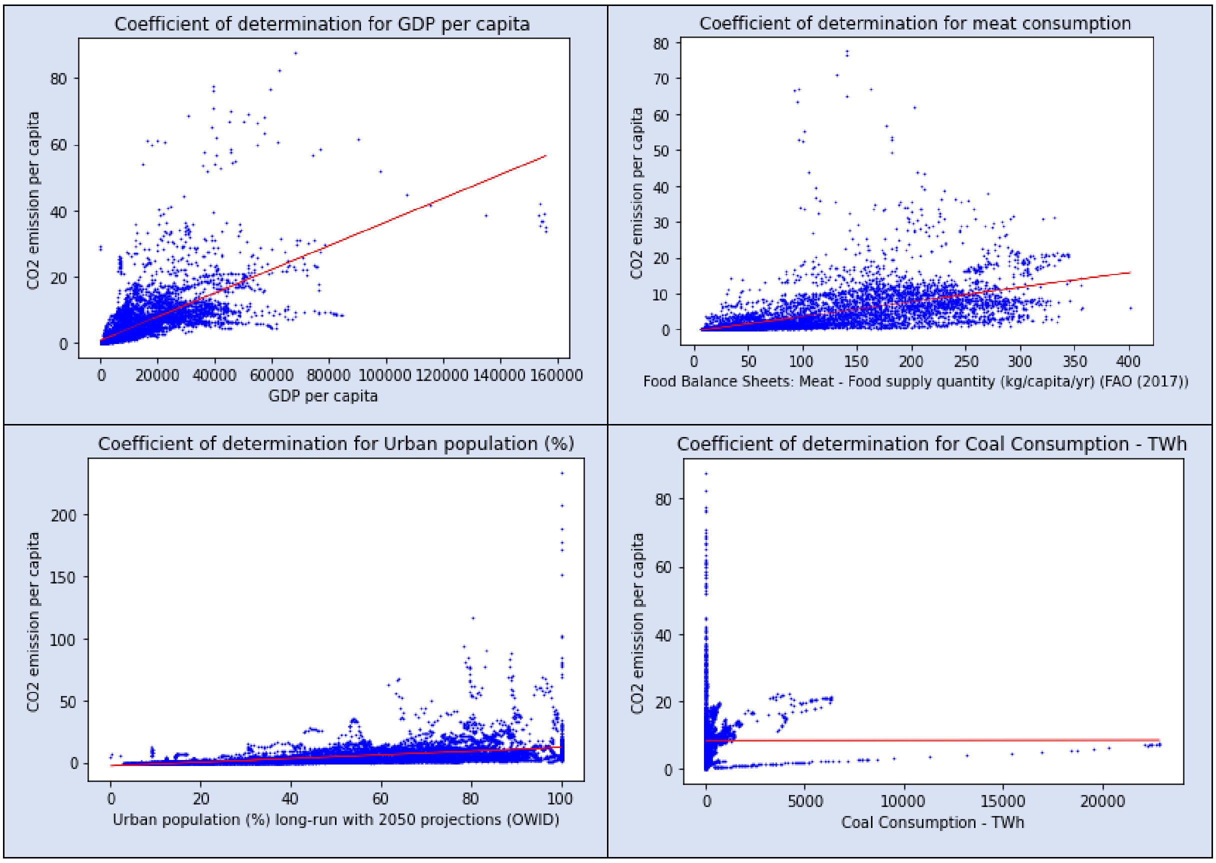

Figure 1: examples of linear regressions using different independent variables

The weakness of the first method, using the Pearson correlations, is that the correlations are not scaled. If for instance the meat consumption per capita in country A has increased over the past fifty years and the emission per capita as well, the correlation coefficient is going to be high, as long as the change is consistent. It does not matter very much whether the change in either variable has been very small or very large. This has the risk of producing misrepresentations. On the other hand the first method has been a quick and easy way to test different potential independent variables. Coefficient number 3 is a good metric to check if a correlation implies a causal relations between the variables, because it calculates how the variables change respective to each other over time.

The second method has the virtue of having an element of prediction. This produces a very direct answer to the question of how good a predictor each independent variable is. Despite the training data portion of the dataset being randomized every time, the values of the coefficient of determination does not change much on repeat calculations.

Question 2: Biggest reductions in carbon emissions

Data

Two datasets were used for question 2. The first was the data for annual CO_2 emissions and the second was a dataset of population estimates, both by year and region. The carbon emissions dataset was filtered to only include countries. The rows for emission totals per continent or for dependent territories were removed. Some values for CO_2 emissions were so low that they were noted in the dataset as 0. These were treated as null-values instead.

Secondly it was important to establish a timeline. Both the dataset for carbon emissions and for population totals contain a lot of historical data that stretches back centuries for some regions. When we think about reducing carbon emissions it is important to keep in mind that the general awareness of the dangers of greenhouse gasses is a recent development. Ideally we would look for intentional reductions in carbon emissions by countries as the result of policy. Only data after 1972, the founding of the environmental think tank the Club of Rome, was used. The starting year was set as a variable to compare long and short term views.

Method

Two approaches were used to determine the countries with the biggest reductions in CO_2 emissions. For the first I used the value of each country's carbon emissions relative to the total amount of CO_2 produced worldwide and calculated the difference over time. The results of this method are listed in table 1.

Country	Relative reduction 1972-2020 (%)
United States	-15.14
Russia	-5.54
Germany	-4.71
United Kingdom	-3.14
Ukraine	-2.60

Table 1: Top 5 reducers CO_2 emissions relative to worldwide total

The problem with this approach is that it favors early industrialized countries. The United States comes out on top in this table because it was by far the largest producer of CO_2 in the world. The fact that its share is reduced by such an extent is mostly because of the increase in carbon emissions by other countries in the decades since. In the same calculation China increased its share by 25 percent.

Because of this, the second approach calculates each country's increase and decrease in emissions over time. It uses a benchmark set on a specific year and compares those values to the values for 2020, which is the most recent year in the dataset. This results in a ratio expressing the increase or decrease in total carbon emissions by that country. Next, change in population was factored in. The population of countries can change over time, and carbon emissions often correlate with population. The same benchmark year that was used to calculate relative change in emissions was used to calculate relative change in total population. The population change ratio was combined with the carbon emissions change ratio for the adjusted carbon emissions change ratio.

Finally, the carbon emissions change ratio was calculated with several different benchmark years, from 1970 to 2000 in steps of 5 years. Both the ratio adjusted for population changes and the ratio not adjusted for population changes were calculated. Some of the initial results were very small countries, so only countries with a population of at least 50.000 in 2020 were included.

Country	**CO2 change**	**Adjusted for population**
Moldova	0.18	0.16
Curacao	0.29	0.26
North Korea	0.32	0.19
Bahamas	0.36	0.16
Latvia	0.38	0.49

Table 2: Top 5 reducers CO2 emissions since 1972

Moldova, for instance, went from 28.576.432 tonnes of CO2 in 1972 to 5.146.876 tonnes in 2020, while the population grew slightly.

Country	**CO2 change**	**Adjusted for population**
Bahamas	0.36	0.16
Moldova	0.18	0.17
North Korea	0.33	0.19
Democratic Republic of Congo	0.82	0.19
Liberia	0.67	0.20

Table 3: Top 5 reducers CO2 emissions since 1972 (adjusted for population)

If we calculate these values for different benchmark years and take the averages we get the following (Table 4):

Entity	CO2 change	Adjusted for population
North Korea	0.31	0.23
Democratic Republic of Congo	0.81	0.27
Somalia	0.84	0.32
Sierra Leone	1.02	0.35
Tajikistan	0.88	0.36

Table 4: Top 5 reducers CO2 emissions averaged (adjusted for population)

If we look at all the countries that at least make the top 10 for decrease in adjusted carbon emissions before averaging out the results, using benchmarks between 1970 and 2000 as starting years, these are the big reducers (Figure 2):

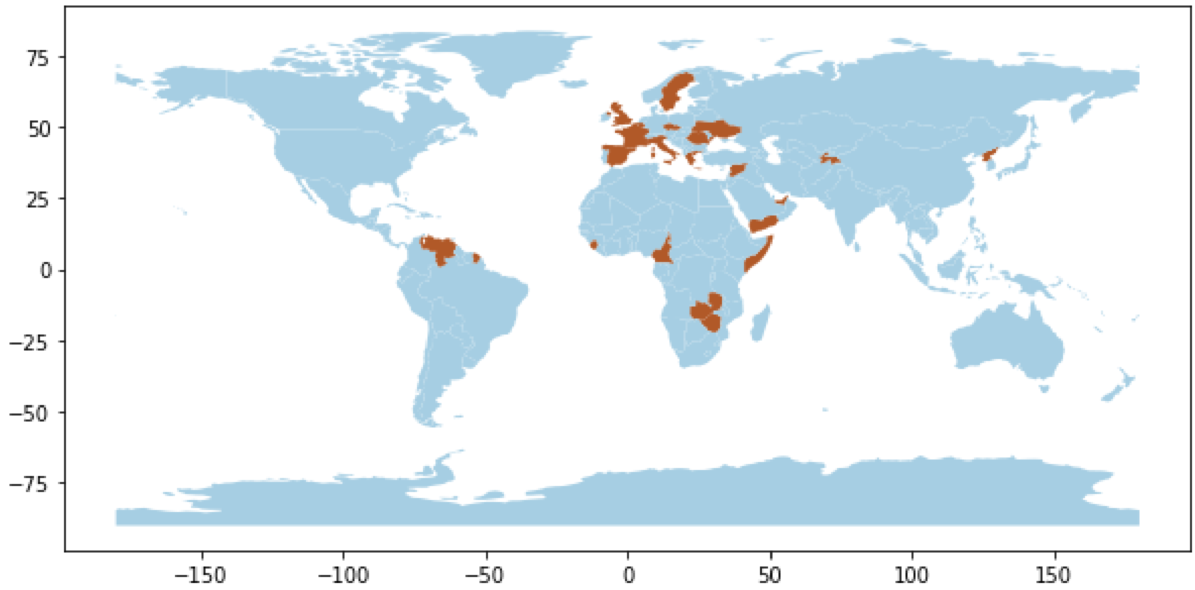

Figure 2: Biggest reducers in carbon emissions, measured between 1970 and 2020

Question 3: Best future price for non-fossil fuel energy

Data

One dataset was used for this question, the Levelized Cost of Energy (LCOE) by various non-fossil fuel energy technologies. Levelized Cost of Energy is determined through a formula that takes several factors into account. On the one hand it involves estimated costs for constructing and maintaining a new power plant (using the specific technology) against any financial benefits such as sustainable energy subsidies from governments. On the other hand it takes into account the life expectancy of a new power plant and the energy output. Figure 3 shows the formula as explained on the University of Calgary website.[1]

$$\textbf{Total Lifetime Cost} = \sum_{t=1}^{n} \frac{I_t + M_t + F_t}{(1+r)^t}$$

$$\textbf{Total Lifetime Output} = \sum_{t=1}^{n} \frac{E_t}{(1+r)^t}$$

$$\textbf{LCOE} = \frac{Total\ Lifetime\ Cost}{Total\ Lifetime\ Output}$$

$$\textbf{LCOE} = \frac{\sum_{t=1}^{n} \frac{I_t + M_t + F_t}{(1+r)^t}}{\sum_{t=1}^{n} \frac{E_t}{(1+r)^t}}$$

- **I_t = Investment and expenditures for the year (t)**
- **M_t = Operational and maintenance expenditures for the year (t)**
- **F_t = Fuel expenditures for the year (t)**
- **E_t = Electrical output for the year (t)**
- **r = The discount rate**
- **n = The (expected) lifetime of the power system**

Figure 3: LCOE formula

The LCOE dataset from OurWorldInData includes seven non-fossil fuel energy sources. There is no LCOE data for nuclear energy on the same site, so nuclear energy has not been taken into account. The dataset contains LCOE for different years and a select number of individual countries as well as the worldwide data. However, only onshore wind and solar power are available for the individual countries. For some countries the dataset only contains LCOE for one of these energy technologies, and for several countries only a few years are available. Because of this the analysis in question 3 was mostly based on the worldwide estimates.

[1] https://energyeducation.ca/encyclopedia/Levelized_cost_of_energy

Of the worldwide LCOE values the earliest point for which all seven technologies are listed is the year 2010. The dataset goes on to 2019, which gives us ten years with which to work. There is no LCOE for geothermal energy in 2011, that value was interpolated.

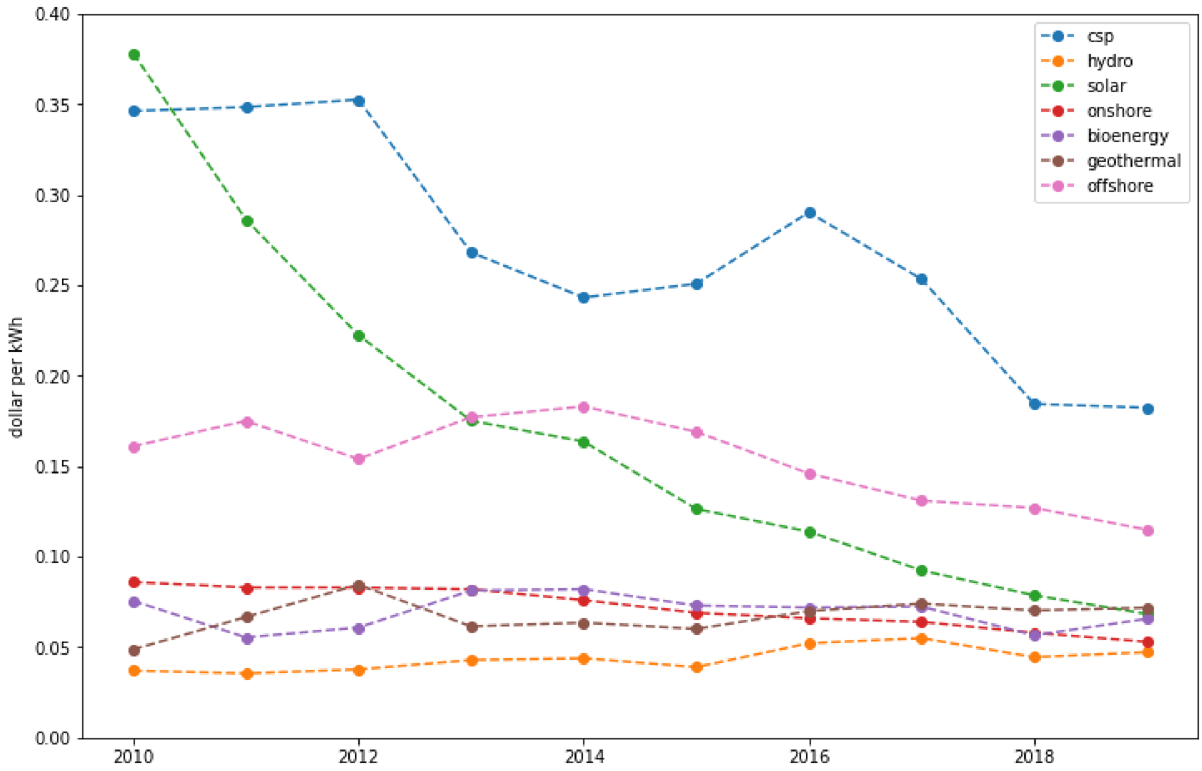

Figure 4: Available data for LCOE (2010-2019) per energy technology

Method

Two methods were used to predict future prices for non-fossil fuel energy. Because the available data spanned ten years the predictions were limited to ten years after the most recent data point, or 2029. The first method used was a linear regression. However, this model ended up predicting a negative value for LCOE in the case of solar energy, starting in 2022. Referring back to Figure 3, with the formula for LCOE, a negative value is only possible if the sum of Investment, Maintenance and Fuel becomes negative. That is practically impossible. The way LCOE is defined it must always be a value greater than zero.

Instead, the second method used non-linear regression. Rather than a linear function, which can and will result in negative values, this method used a function for exponential decay. Such a model assumes that LCOE will decrease as a factor of time but not become negative, and that the rate of decrease will gradually become smaller. Such a model is more appropriate to simulate changes in energy technologies than a linear function.

For the second method different exponential decay functions were fitted through machine learning to the available data. Future values were plotted based on the parameters of the best possible fit for each energy technology. In Figure 5 the projected values are represented by dotted lines. The available data points, up to 2020, are represented as dots.

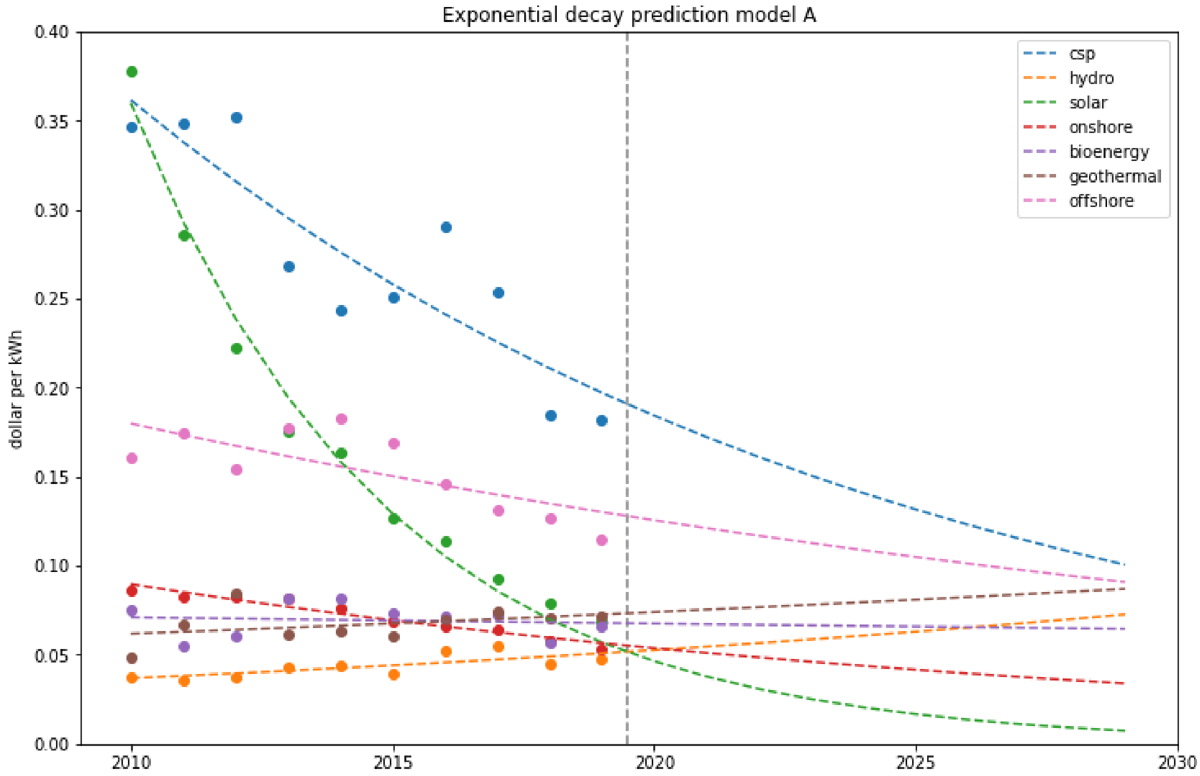

Figure 5: Projected worldwide LCOE per energy technology

The LCOE dataset contains data for China, France, Germany, India, the United Kingdom and the United States, though only for onshore wind and solar power. Using the LCOE values for these specific countries and energy technologies we end up with comparable results as depicted in Figure 6.

Finally, because the LCOE values for onshore wind energy go back further than the data on other energy technologies, it is possible to compare the models based on long and short term available data. Figure 7 compares the line graphs for exponential decay models fitted to both sets to examine the impact of the short term view on the projections.

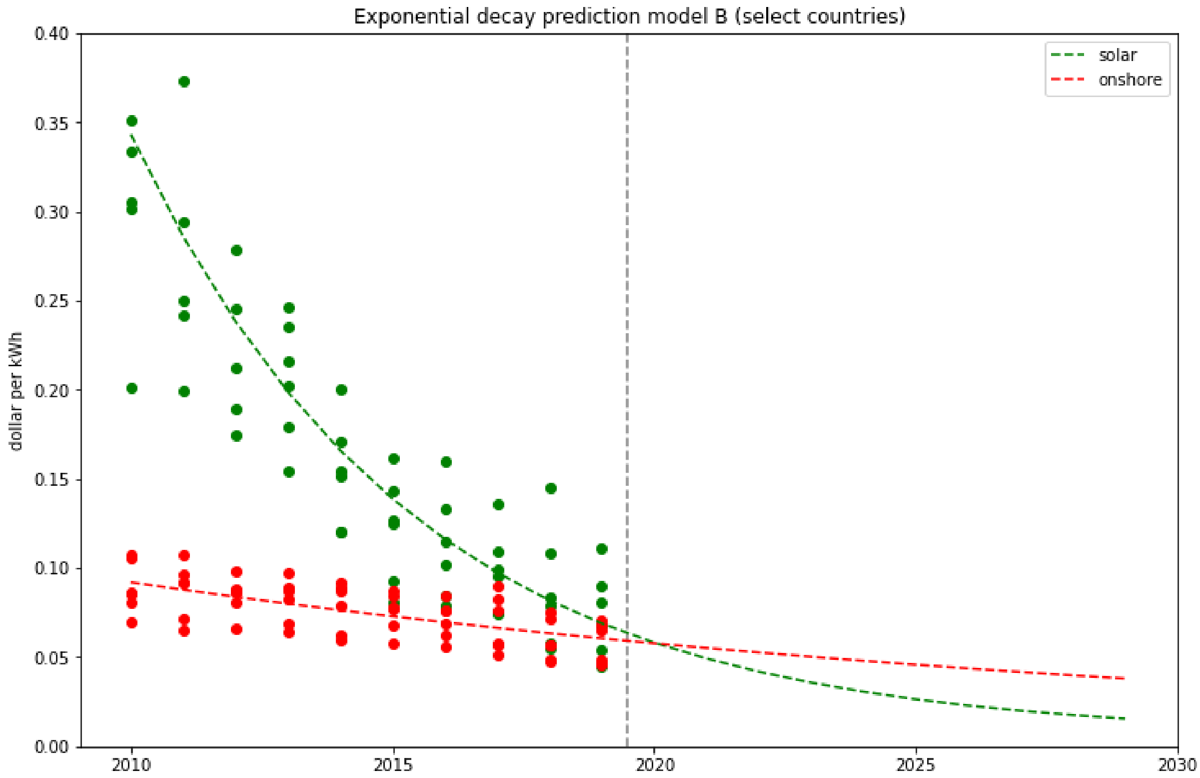

Figure 6: Projected LCOE for solar and onshore wind power in selected countries

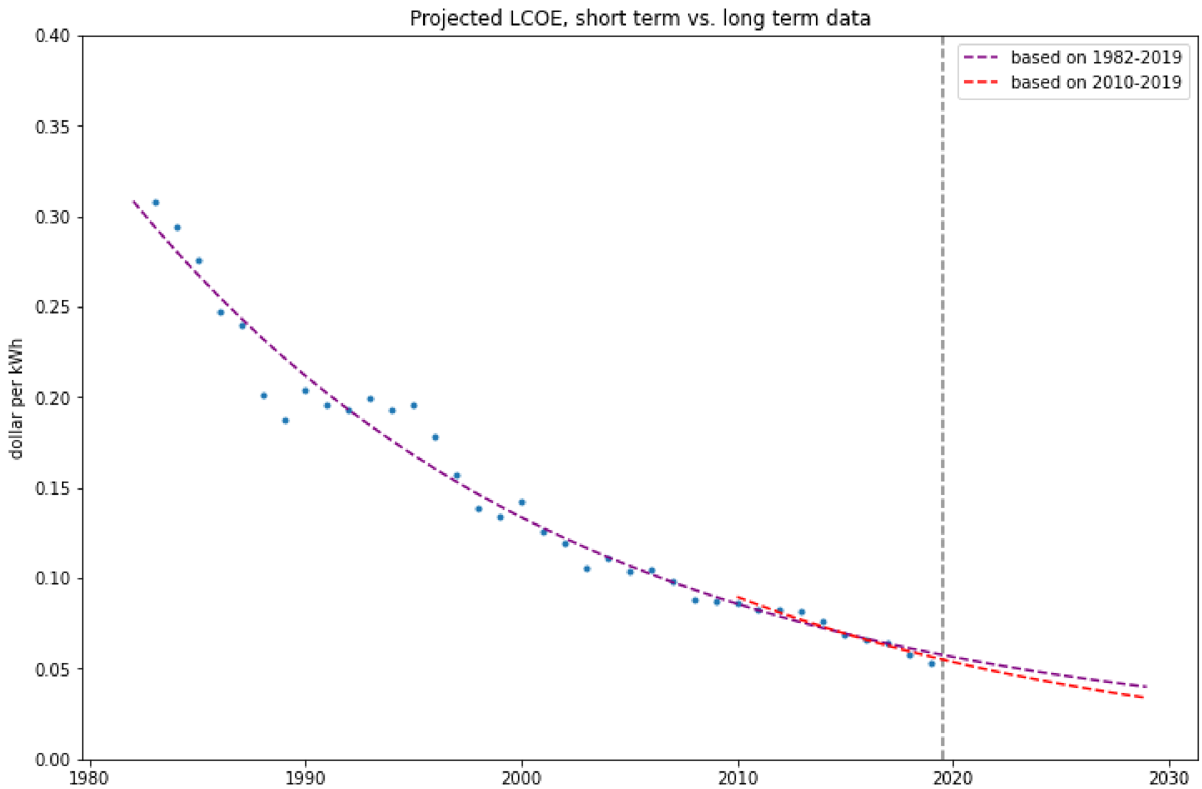

Figure 7: Projected worldwide LCOE for onshore wind energy

Conclusion

Question 1:

There are many factors contributing to the level of carbon emissions in a country. Every human being and most things we do in life have a carbon footprint in some sense. This report examined correlations between the values for emissions per capita and independent variables in various countries and years. Unsurprisingly, several factors correlated with a country's carbon emissions per capita, factors like car ownership, access to electricity, and urbanization.

The variable tested as most accurate for predicting a country's carbon emissions per capita was a country's GDP per capita. While there were other independent variables that performed well for correlations or predictions, GDP was the strongest predictor. What's more, the other factors are connected to the economic welfare of a country and its population. Repeating the tests showed that GDP per capita can serve as a predictor for these other factors as well. In other words, the statistics imply that the populations with high GDP own more cars, have more access to electricity, have more meat in their diet, and also generally produce more CO_2.

Variable	**Pearson coefficient**	**Coefficient of determination**
GDP	0.67	0.43
Motor vehicles	0.62	0.35
Urbanization	0.45	0.20
Electricity access	0.55	0.30
Meat supply	0.56	0.33

Table 5: Predictors for emissions per capita

Question 2:

What can be done to decrease carbon emissions worldwide? One approach may be to examine countries in which carbon emissions have decreased recently. I compared changes in carbon emissions per country in the past 50 years, adjusted for increasing or decreasing population figures. The biggest reductions are often not in environmentally conscious countries, but in countries devastated by war or economic crisis. Carbon emissions in North Korea have undergone relative decrease in the past fifty years, and the Democratic Republic of Congo, while maintaining stable levels of carbon emissions, has seen a large population increase since the 1970s. The same applies to Liberia: adjusted for population growth the carbon emissions in 2020 were only 20% of what they were in 1972.

There are two problems when trying to find an answer to the question which countries have made the biggest strides in reducing carbon emissions. The first problem is that it is not possible to tell whether change in CO2 emissions is the result of environ-mental policy or unintentional devastation.

The second problem is that measuring reduction is biased towards the countries that industrialized relatively early, which coincides with economic prosperity. The United States was the greatest contributor to worldwide carbon emissions in the 1970s. That share has decreased significantly since, in large part because countries like China and India increased carbon emissions since. This becomes more problematic if we consider the relationship between a country's GDP per capita and carbon emissions discussed in question 1. The biggest reducers in carbon emissions are not the countries one would treat as an example for future policy.

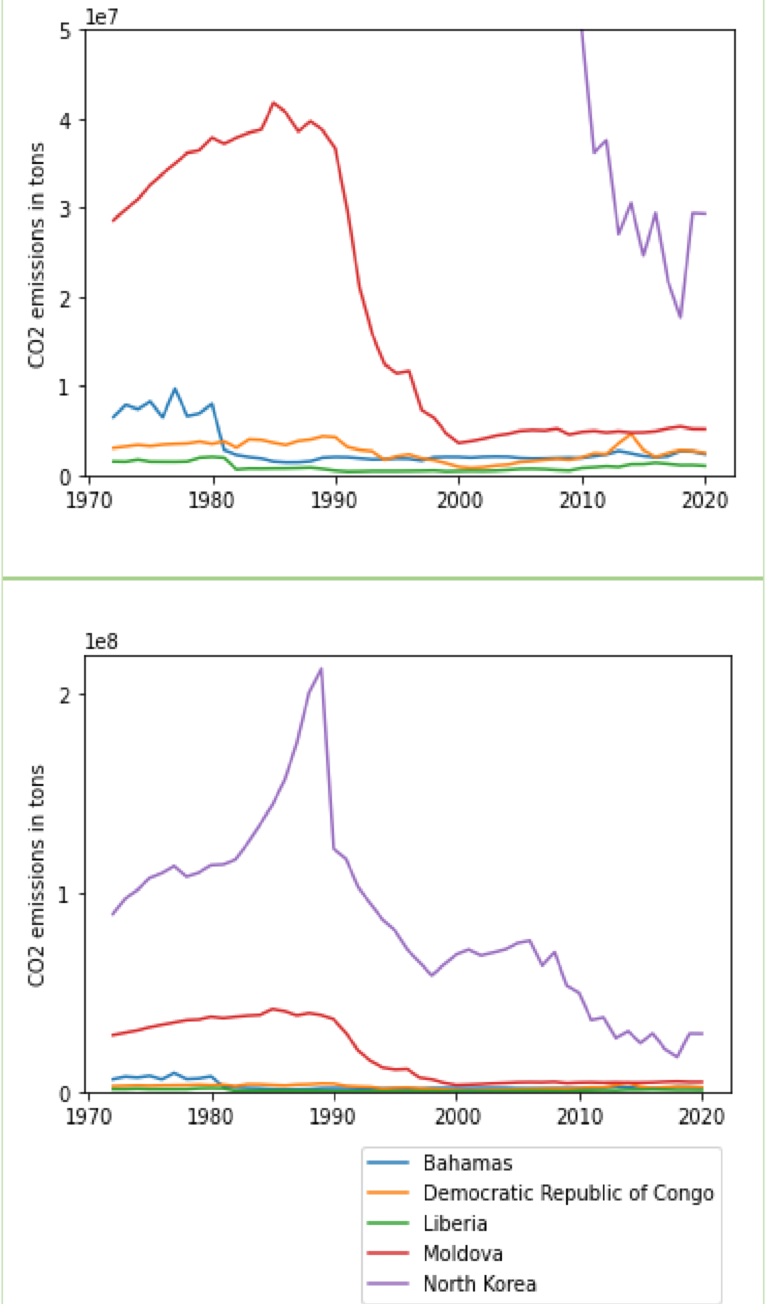

Table 6: Top 5 reducers in carbon emissions, adjusted for changes in population (1970-2020)

Question 3:

If the increase in carbon emissions is a byproduct of economic development, we should look for a solution in economic incentives. One of the paths to reducing global emissions is the energy transition from fossil fuel to sustainable power sources. Future prices for non-fossil fuel energy were estimated based on the Levelized Cost of Energy (LCOE) for different energy technologies. Based on the developments of the past decade, the projections indicate that solar energy and onshore wind power will be the most cost-effective non-fossil fuel sources of energy in the near future. In the long term this may change to concentrated solar power (CSP) or offshore wind.

Predictions about LCOE should be used carefully. Scientific and technological innovations can reduce costs and increase output of power plants, and therefore decrease LCOE. The prediction model used here assumes that this is true for the future of non-fossil fuel energy. The LCOE value is also influenced by government incentives for sustainable energy sources. This means that countries have the ability to shape the future of energy.