Visualizing countries' efforts in reducing global warming

Eindhoven University of Technology

B.J. Grooten
0885158

A.C.E. Husson
0956648

Abstract

We present an interactive visualization in the form of a webpage about global warming data from the World Bank. The problem of comparing the efforts of countries easily is addressed. Variables such as greenhouse gas emissions per person and renewable energy usage are analyzed together with the national income level, access to electricity, and access to internet. The encodings that are used are a parallel coordinate plot, a scatter plot, and a line plot. Years between 1995 and 2012 can be chosen by the user with a slider. Interesting findings have been made, such as the inverse relation between access to internet and renewable energy usage, or the fact that Canada generally emits more CO_2 equivalent per person than the USA.

1 Introduction

According to the World Bank, most global warming has been caused by human activities. [2] We have changed the chemical composition of the atmosphere through a buildup of greenhouse gases, primarily carbon dioxide, methane, and nitrous oxide. The rising global temperatures can be dangerous, as they will alter local climate conditions and cause sea level rise, which affects water supplies, crop yields and forests. It may affect animals, human health, and many types of ecosystems.

Since the Paris Agreement of 2015 many countries promise to perform more serious actions in order to decrease their greenhouse gas emissions and use more renewable forms of energy. The World Bank has a lot of interesting data available about this. [2] Unfortunately, they do not yet have emission data for the years 2013 and onwards. However, with the data until 2012 we can check what the actual results are of the actions that countries contributed before Paris. We made an interactive visualization such that these insights can be gained easily.

2 Problem description

The given data set gives rise to many viable research questions about countries' efforts to reduce global warming. Are countries contributing to the reduction to global warming? How much? Are countries suiting there action to their words? And are there other factors that might influence the emissions of a country? These are general questions that are of interest of analysts with respect to the given data set. In particular the following questions are of specific interest:

- Are countries using more renewable energy?
- Which countries show the largest growth?
- Which countries use the most renewable energy now?
- What are the proportions between the different countries and can we compare these countries?
- What influence does attribute x have on the total greenhouse gas emissions?

Attribute x is a property of which researchers might expect to be of influence on the emissions or energy consumption of a country. The properties that we chose are:

- national income per capita,
- access to electricity,
- access to internet.

The national income per capita is adjusted for inflation to the current value of the US dollar. The data about access to electricity and internet are given as a percentage of the total population.

2.1 Task analysis

In order to answer the research questions, a set of tasks will have to be constructed. The data set obtained from the World Bank has an enormous amount of information about each country in a time span from several years. Investigating this data set with respect to the effects of climate change means that we are only interested in some of the countries' properties. The most important attributes that are of interest are the total greenhouse gas emissions per capita and the percentage of renewable energy consumption, for each country. These variables represent the real life results of the efforts that countries have put in to reduce global warming. The emissions per capita and the renewable energy percentage will also be used as an indicator of how a country negatively contributes to climate change. We want to compare these two variables to the other attributes in order to explore whether the amount of access to internet, electricity or the national income influences the emissions and renewables of a country.

The tasks we want to be able to perform with our visualization program are the following:

- compare the total greenhouse gas emissions per person (in CO_2 equivalent), by the different countries.
- compare the percentage of renewable energy usage by country.
- see the development of all these statistics over the years.
- investigate whether certain attributes affect greenhouse gas emission or renewable energy consumption.
- discover trends between the different variables and maybe discover new attributes that influence climate change.

It would be interesting to investigate how effects like access to internet, access to electricity or income affect climate change. Maybe they do not have any effect, but perhaps new discoveries could be made. Also, the variable of emissions per capita is

something we do not often see in literature or media. Usually the total emissions are given [1].

For countries themselves it would be interesting to see how well they are doing compared to the world and to see whether they should improve. Or maybe to get a confirmation that they are going in the right direction. Are there countries that should receive a penalty for negatively influencing the climate?

For scientists and domain experts this visualization can show the development of the emission rates around the world. It would be interesting to see whether the emissions rates are really as bad as the media portrays. Maybe it is even worse, and the world as we know it today might not exist in the future. Seeing the development of these variables, could lead to drastic measures that will have to be taken in order to make the world a better place for future generations.

3 Visualization design

In order to get more insight in the data set and in order to answer the research questions we made an interactive design. There are several different ways to visually encode our data, so in this section we will motivate our choices for choosing a specific design. In particular three different visualization techniques are integrated into the design. These three different graphs are linked, so that the user can interact with all of them.

In this visualization we want to show the development of the variables and countries over the years. Therefore there needs to be some kind of time element will have to be implemented. We have chosen to integrate a slider into our design. In this way we can show visualizations for every year and by using the slider the user can see its progression. Also we have chosen for one visualization to show the development over all the years at once regardless of the slider, such that this can be seen more clearly.

A scatterplot can show the correlation between two variables when taking a look at the pattern of the points in the plot. In our case, an interesting aspect is to compare the variables like income, electricity and internet to the variables that are known to be an indicator of climate change: renewable energy consumption and greenhouse gas emissions. Because with a scatterplot you can see patterns and correlations between two variables and we are interested in seeing these correspondences, we have chosen to use a scatterplot. Also when using a scatterplot you can look at each country individually and indicate which countries can be classified as outliers. These outlier countries are the interesting countries in our analysis. In the specific scatterplot we used, different colors will indicate the different regions indicated by the World Bank. By making all countries of the same color, it can be investigated if countries in the same region show the same pattern. When the user hovers over a specific country in the scatterplot, a gray line will indicate the path of this country over time. In this way the user cannot only see the progression over time with the use of the slider, but also in the scatterplot itself.

Using this dataset we want to see how the different variables relate to each other. We can see this in the scatterplot for the comparison of two variables, but it would also be interesting to see all the variables together and have some kind of overview. The researchers also want to identify trends between the different variables. A graph that could indicate all of these aspects would be a parallel coordinate plot. This plot could serve as an overview plot. In this plot we could see which country has the biggest values and which has the lowest values. We could also see for each country all the attribute values together in one plot. A parallel coordinate plot can also be used to see which countries kind of link together and show similar patterns. In order to show the paths in the graph for individual countries, the line will light up and the country name will be visible on hover.

Another interesting aspect is seeing the progression of the emission rates and renewable energy rates of time. The slider allows us to see the different plots of the different years, however it would also be interesting in seeing all different years in one plot, in this way we can see whether a country emits more or less over time, and compare it with other countries. This can be done when using a stream chart where all countries are of a different color and the x-axis represents the years. However, we are specifically interested in how much a country emits. When the countries are stacked on top of each other in a stream graph it is difficult for the user to compare two countries if their emissions do not differ that much. Therefore a stream graph would not be a good idea.

We have chosen to use a line chart to visualize the greenhouse gas emissions and renewable energy consumption over time. The x-axis represents the years and the y-axis represents the amount of emission or the amount of renewable energy consumption. A feature will be added in which the user can specify themselves, which countries will be seen in the line graph. In this way not all countries are visualized at the same time, which would make the graph unclear.

Of course one could also use a bar chart or a pie chart to show the values of variables for different countries and find out which country uses the most renewable energy. However, using a pie chart it is difficult for the user to compare the different angles and this would be easier using a line or a bar in which we can see its height. However, when using a bar chart one would need one bar chart for each different year. An option would be to use a stacked bar chart for the different variables or for different years, but then it would again be difficult to compare values as because they are stacked the height cannot be seen clearly.

For the linked views we implemented the slider, if the user interacts with this slider the parallel coordinate plot and the scatterplot will change. The line chart interacts with a menu. On this menu every country is shown, divided in the different regions as stated by the World Bank. This makes it easier for the user to see which countries participate in the dataset. When the user clicks on a country this country will show up in the line chart. The user could add or delete countries in the line chart by clicking on them in the menu.

4 Use case

4.1 Data preparation

Because we are only interested in a specific part of the World Bank data set, we have adapted it accordingly so that our data only contains information relevant to this subject. As stated in the introduction, we only have data about greenhouse gas emissions until the year 2012. And before the year 1995 there is almost no data about internet access (which is quite logical since the internet only just existed back then). Therefore, we focused our data on this time interval of 18 years. If there was a missing value for particular countries and attribute, we interpolated it using the values of the previous and the next year. We did the same when there were 2 or 3 consecutive values missing. If a country had 4 or more consecutive values missing, we decided to delete the entire country (so with its all attributes) from our data set. This left us with 139 countries and 43 regions (such as “Arab World” or “South Asia”, defined by the World Bank).

4.2 Realization

The visualization design has been discussed in Section 3, in this section we will show how to use the visualization application and

which results can be obtained. First we will describe the implementation and realization of our application. Figure 1 below shows an overview of the application.

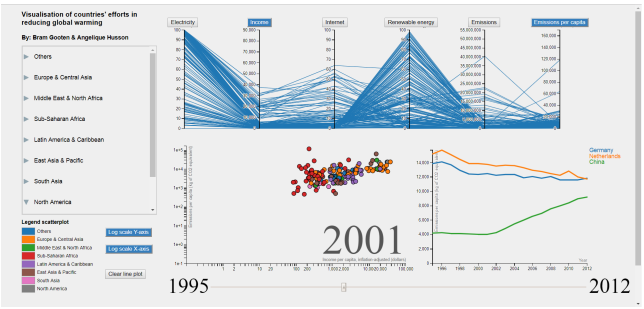


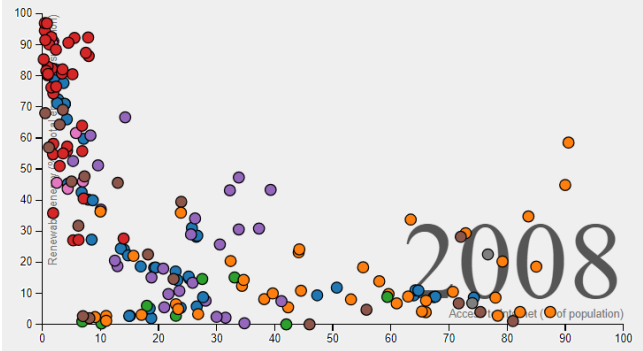
Figure 1: Overview application

On the left top side of Figure 1 the menu can be seen, in which all of countries per region are displayed. These regions are defined by the world bank and are: ”Others”, ”Europe Central Asia”, ”Middle East & North Africa”, ”Sub-Saharan Africa”, ”Latin America Caribbean”, ”East-Asia & Pacific”, ”South Asia” and ”North America”. Below this menu a legend of the scatterplot can be seen. The different colors in the scatterplot represent the different regions. Next to the legend on the right, there are three buttons with which the user can specify if logarithmic or linear axis will be preferred with which the user can clear all lines on the line plot. On the top right, a parallel coordinate plot can be seen, with on top 6 buttons for each variable. When these buttons are clicked the bottom two plots will be shown with the specified variables on the axis. Below the parallel coordinate plot, the scatterplot can be seen on the left and the line plot on the right. With the slider below these plots, the user can change the years of the scatter and parallel coordinate plot simultaneously. In this application, renewable energy and income where selected. In the scatterplot the log scale was used and the plots are shown for the year 2001.

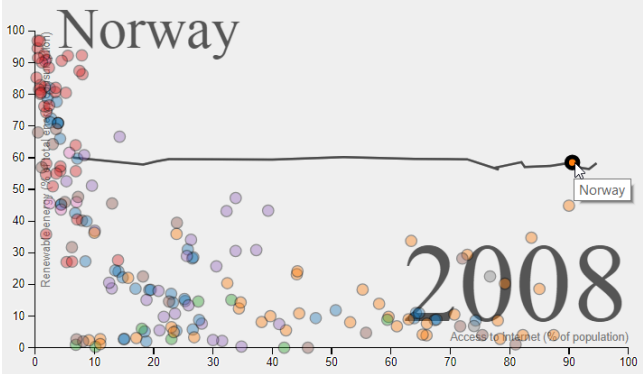
We will now zoom in, to the three separate visualizations to take a closer look. For the scatterplot the log or linear scale can be chosen for the axis. We have chosen that for the variables that are represented in percentages, the linear scale is appropriate and that for the other variables the logarithmic scale is appropriate.

In Figure 2 the scatterplot with and without mouse interaction can be seen for variables renewable energy consumption and income per capita. This Figure shows the year 2008. In the second figure the user interacts with the figure by clicking or hovering over a country. While hovering, the name of the country appears on the left top part of the plot and the other countries are less visible. In this case this country is Norway. The user can also see the path of this country over the years. So in this case we can see that over the years Moldova increased in income per capita and also in renewable energy. However this path is not a monotone increasing path as can be seen in the figure.

In Figure 3 an example can be seen of the line plots with and without mouse interaction. When the user hovers on one of the lines, the country that that line represents can be seen on the top left corner. The first figure shows the Renewable energy consumption and the second one shows the Emissions per capita over the years.

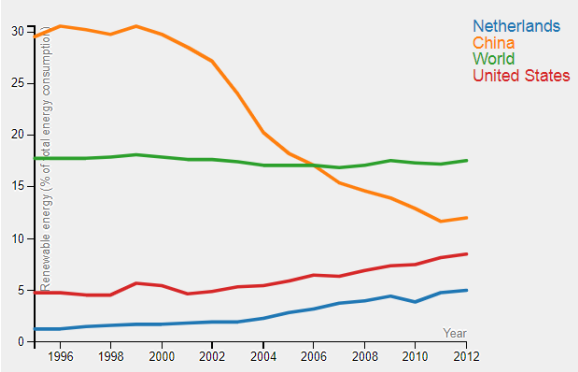


(a) Scatterplot

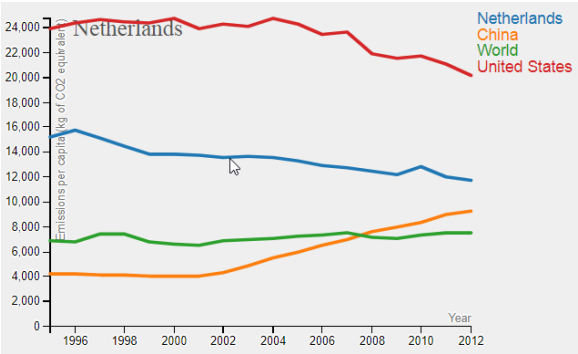


(b) Scatterplot with mouse interaction

Figure 2: Scatterplot of percentage of internet access against renewable energy consumption in the year 2008.



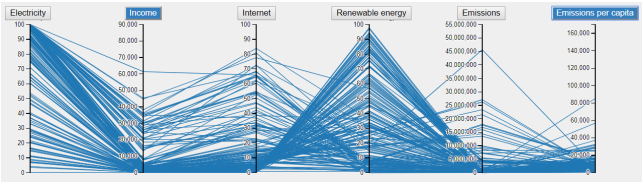
(a) Line plot renewable energy consumption



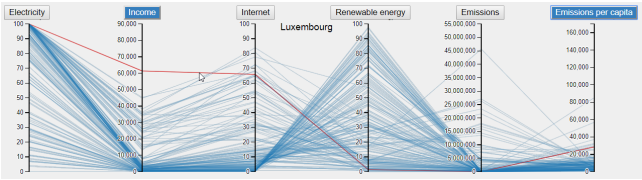
(b) Line plot emissions per capita with mouse interaction

Figure 3: Line plot of countries Netherlands, China, United States and the World.

In Figure 4 the parallel coordinate plot of the year 2004, with and without mouse interaction can be seen. When the user hovers over a line, the name of the country will be displayed. The line will turn red and the other lines will be set to a lower opacity value.



(a) Parallel coordinate plot without mouse interaction



(b) Parallel coordinate plot with mouse interaction

Figure 4: Parallel coordinate plot of the year 2004

4.3 Data exploration

Now that the application and the dataset are properly explained, the exploration of the data can begin. From the plots seen in the previous section we can already see some interesting aspects.

4.3.1 Scatterplot exploration

Internet

In the previous section the scatterplot of internet access against renewable energy consumption was shown in Figure 2. From this figure we can see that certain clusters of countries are formed that correspond to the colors in the legend. The legend can be seen up close in Figure 5 below.

Figure 5: Legend of the scatterplot

Using this legend we can see in Figure 2 the red dots or the Sub-Saharan Africa countries cluster together. Also the purple dots which are Latin America and Caribbean countries and the orange dots which are countries from Europe and Central Asia clearly cluster together. Hence, we can see some kind of pattern that countries from the same region form a specific cluster. Countries from Sub-Saharan Africa invest a significantly higher amount of money into Renewable energy than countries from other regions, whereas these countries have only limited internet access. Countries from Europe and Central Asia have the highest percentage of internet access but score low on renewable energy investment. We can also see some interesting outliers. For example Noway is one of the few countries with a high percentage of internet access that does invest in renewable energy consumption. Hence, Norway kind of sets an example to the rest of the world.

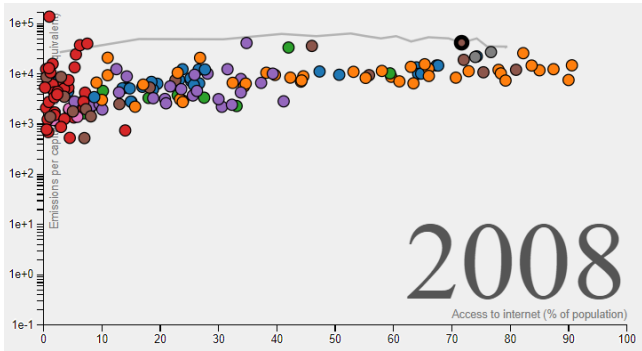
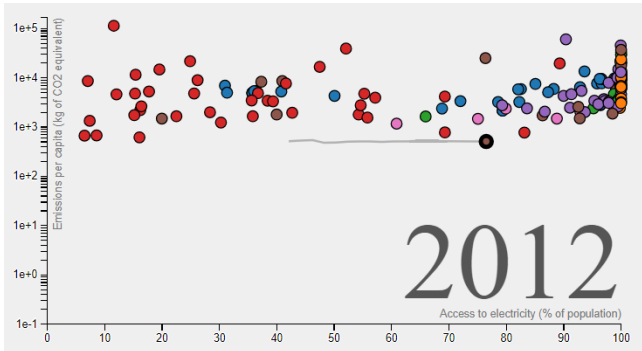


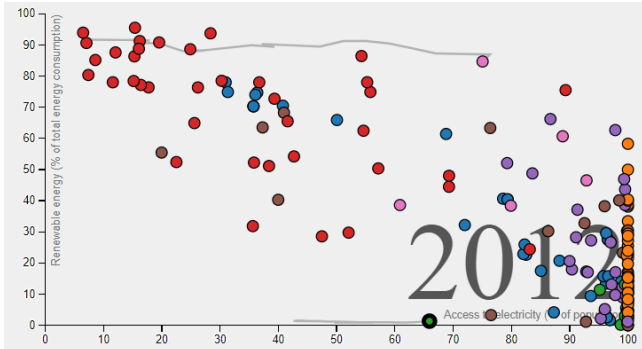
Figure 6: Scatterplot Internet against Emissions per capita in the year 2008

Figure 6 also shows internet access but now against the emissions per capita of a country. In this graph we again can see the clustering of the different regions. And the dots at the far right are the countries with the highest internet percentage such as Norway, Denmark and the Netherlands. The black circled country is Australia, we can see this countries path. This path progresses relatively horizontal, which means that internet percentage increases drastically over the years, whereas the emissions per capita stay roughly the same. Most of the lines progress horizontally. We can also see that the lower dots on this scale perform better than countries higher on the graph, and we can compare these countries to each other.

Electricity



(a) Scatterplot Emissions per capita of the year 2012

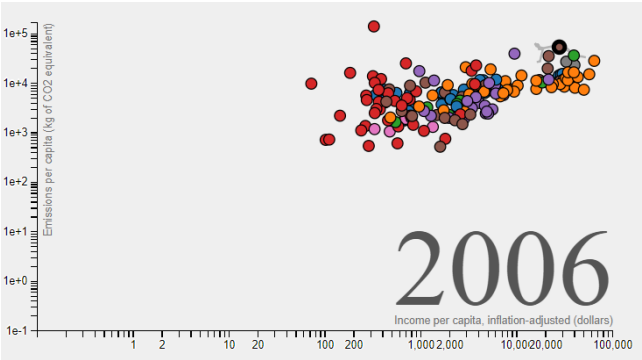


(b) Scatterplot Renewable energy consumption of the year 2002

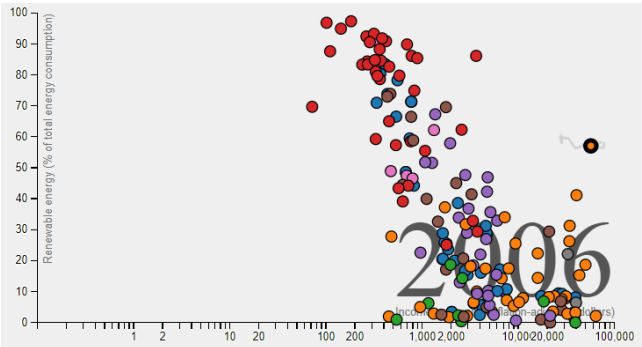
Figure 7: Scatterplots of percentage of access to electricity.

In Figure 7 the percentage of electricity with respect to the emissions per capita in 7a and to the renewable energy consumption in 7b can be seen. We again see the same pattern that countries from the same region form clusters. Also, note that the renewable energy consumption graph shows more severe outliers than for the graph of emissions per capita and also shows a different pattern. The graph of emissions shows that the amount of emissions stays similar even if the amount of internet access is increased. However, when the amount of internet access grows the amount of a country spend in renewable energy consumption decreases. So in the second graph we can see kind of a downwards trend. We can also see some outliers. For example, the bold green dot on the bottom of the second graph is Yemen, this is a well-known relatively unstable country and hence it makes sense that it does not behave as the other countries.

Income



(a) Scatterplot Emissions per capita



(b) Scatterplot Renewable energy consumption

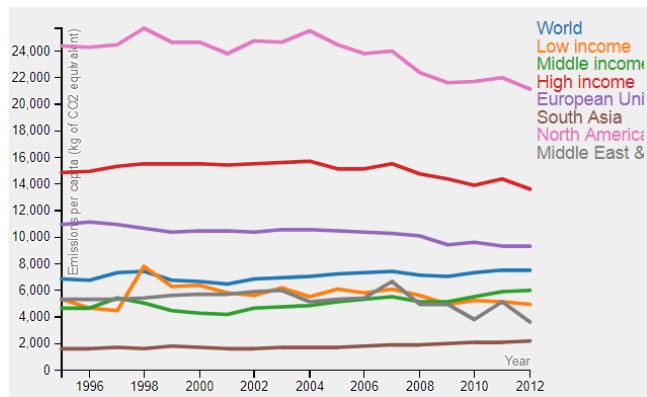
Figure 8: Scatterplots of the amount of income per capita in the year 2006.

In Figure 8 the effect of variable income on renewable energy and emission can be seen of the year 2006. From these graphs we can again see the clusters per region and also some interesting outliers. For example, in graph 7a the country with the highest emission rate is Central African Republic. Compared to the other countries this country also has quite low income rates. One a positive note we can see the outlier Norway in graph 7b. From all of the higher income countries, Norway spends the most into renewable energy.

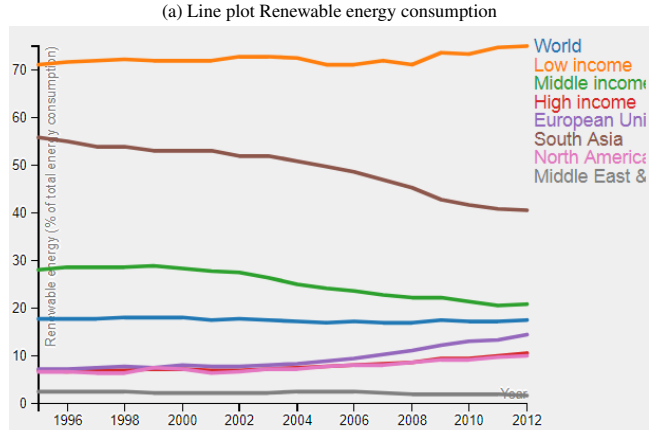
4.3.2 Line chart exploration

From the line plots in Figure 3 the emissions per capita and the renewable energy consumption as a percentage of a countries total energy consumption can be seen for four countries. As the government or researchers in the Netherlands it would be interesting to see how well the Netherlands is doing compared to a few other big countries and compared to the world. The first Figure 3a shows the Renewable energy consumption as a percentage of its total consumption. We can see that the Netherlands scores really low compared to the other countries shown in the graph. This could be seen as a surprise as the Netherlands is known for its large amount of windmills. In the United States and the Netherlands we can see a growing trend in Energy consumption, as expected. As over the years renewable energy became a more important energy source. However, we see that the world shows some kind of a stable trend. It would be interesting to see further than the year 2012, because we would expect that after the Paris agreement this line of the world would show an increasing trend. Also the most remarkable in this figure is that China is decreasing its renewable energy consumption. Figure 3b shows the emissions of a country per capita, so the amount of citizens of a country does not have an effect. This causes us to be able to compare bigger and smaller countries. From this Figure it stands out that the United States has by far the highest emission rate. The other thing that stands out is that while all lines are decreasing, the emissions of China are increasing over the years. This is probably due to the fact that China until 2015 was the largest-growing major economy.

Instead of looking at separate countries we can also take a look at groups of countries linked together, such as the European Union, Middle income countries etc.. In Figure 9 combined plots can be seen.



(a) Line plot Renewable energy consumption



(b) Line plot Emissions per capita

Figure 9: Line plot of combined country groups

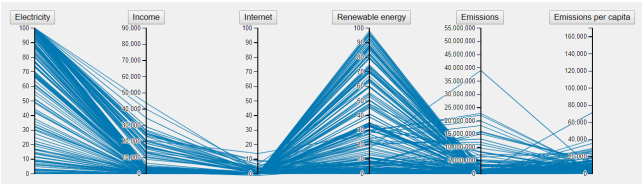
In Figure 9 we can the pattern that the higher the income of a country the lower its percentage in renewable energy consumption. This was

not what we expected as higher income countries often portray in the media that they care more about clean energy. We can also see that higher income countries have significantly higher emission rates than other countries. Also note that the Middle East which is the gray line in the second graph has the lowest percentage of investment in renewable energy. This is as expected as the Middle East produces vasts amount of fossil fuels needed by countries all over the world. One line that stands out in the first graph is North America. North America has the highest amount of emissions per capita. Whereas South Asia has the lowest emissions per capita. South Asia consists of countries such as India and Bangladesh. From India you would not directly expect that its emission rates are so low with respect to the rest of the world. We expect that this may be the case, because there was a relatively low amount of industrialisation. When further experimenting with the line plots we found that Canada generally emits more CO_2 equivalent per person than the USA, which was quite surprising to us. However, when looking at renewable energy percentages, Canada was doing a much better job.

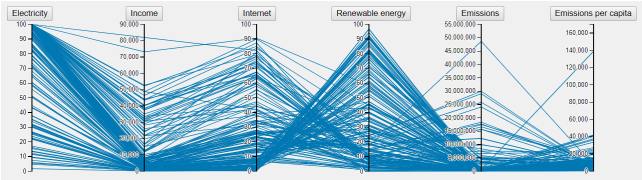
4.3.3 Parallel coordinate plot exploration

In the parallel coordinate plot we can see the general development of the specific attributes for all the countries. For example, when comparing Figure 10a to Figure 10b, one sees immediately that the access to internet has greatly increased all over the world. Although, there is still a big majority of countries with less than 10% of the population with access. This majority is getting smaller and smaller, as can been seen when moving the slider of time further.

It is striking to see in Figure 10b that many countries with a high amount of access to internet use a low amount of renewable energy, while for countries that do not have much internet it is the other way around. This seems to some kind of inverse relation. Many of the countries that have a renewable energy percentage over 80% are African, which might be a bit surprising.



(a) Parallel coordinate plot 1995



(b) Parallel coordinate plot 2008

Figure 10: Parallel coordinate plot changes

The vertical lines of Emissions and Emissions per capita (5th and 6th line) are a bit distorted by one outlier. For Emissions per capita, this is the country "Central African Republic", which apparently emitted more than 160,000 kt of CO_2 equivalent in 2007. Way more than any other country, which is a little unexplainable. For Emissions the line with the highest value is "World", which makes sense because it should be a total. We did not delete the region since we wanted to be able to compare countries with the World average in the other plots. However, for further research a method of filtering the parallel coordinate plot and then changing the maximum and minimum values accordingly would be a good next step.

Other trends that are noticeable are the fact that almost all countries have high access to electricity, no matter what there income level is. A high income does have effect on the access to internet, although there are a couple of exceptions such as Kuwait, which has a low access rate compared to its income, and South Korea, which has relatively high internet access for its income level.

5 Discussion

We started our research approach with choosing the World Bank data set. This was interesting to both of us. Since their data contains so much information, we needed to really dive into the domain situation and figure out what we wanted to focus on. We chose climate change since we feel a bit of responsibility as (beginner) visualization experts to bring more awareness and clearity, as it is big in the news of the Netherlands lately. An article in the Dutch newspaper NRC [1] used the same line plot as we did, but over a bigger time span. (After checking the conversion of their units (G kg) to ours (kt) and the y-axes, it was indeed the same.) We found it an exciting project to work on, and intend to put the results on a real website later.

The next step after choosing our domain was to define the tasks. This harder than expected, especially getting them concrete and clear. We needed to make adjustments later on, as we discovered that some of our data attributes had many more missing values than we thought. Since we could not interpolate these values in an accountable way, we had to make some decisions and choose attributes (and years) that did have enough data.

The visual encodings we chose for the design were discussed in detail by us during the project. We had many ideas that were possible, which had arguments for and against their use. These arguments were presented in Section 3.

The eventual implementation has become quite pretty, which we are proud of. It was a hassle to learn Javascript and D3 in such a short time period, but with a lot of help from the instructors it worked. We prioritized the functionalities that we really needed to show and made those first. Of course there are further improvements which we would have liked to do. These include: having more data attributes, having a bigger and more recent time span, using the same color for the countries in the three different plots, having better links in the interaction such that when user clicks a country it lights up in every plot, turning some buttons off automatically when others are pressed, et cetera.

6 Conclusion

The main contribution we made with this visualization project is that the community that brings awareness to global warming can use this interactive tool to compare the World Bank data in a way that invites people to explore. During our own exploration of the data we found some interesting facts, such as Canada emitting more per person than the USA, or the inverse relation between access to internet and renewable energy usage. An important finding for ourselves is that it is difficult, but very useful when interacting, to have links between different visualizations of a page.

References

[1] E. van der Walle. De politieke wil ontbrak om milieudoelen te halen. *NRC*, page A1, 25 January 2019.

[2] World Development Indicators, *The World Bank*. `https://datacatalog.worldbank.org/dataset/`

world-development-indicators Accessed: 18 December 2018.

Appendix

A WORK DIVISION

Angelique: HTML lay-out, D3 parallel coordinate plot, D3 scatter plot, data exploration, screencast.

Bram: data pre-processing in Python, D3 scatter plot, D3 line plot, data exploration.