European Environment Agency
Polution Reporting

Name	E-mail	PoC
Anders Høgh Sornboe	ansor18@student.sdu.dk	100
Anton Toft Jensen	antje12@student.sdu.dk	100
Marcus Adam Sørensen	marcs19@student.sdu.dk	100
Tjelle Kragelund Strøm	tjstr19@student.sdu.dk	100

The Maersk Mc-Kinney Moeller Institute

University of Southern Denmark

Abstract

Environmental pollution is a serious subject with heavy implications on our daily lives. To combat the environmental changes resulting from emitted pollution, politicians have attempted to remedy the ever-changing environment with restrictions and quotas for emission of pollutants. In this report an investigation will be conducted on a data-set collected on European countries and their biggest industries, provided by the European Environment Agency (EEA). From this data-set a multitude of graphs are created illustrating pollution being emitted on the levels of Country, European Region, Industry Sector, and Pollutant. These graphs serve to illustrate how pollution emissions have evolved throughout the last 14 years (2007 - 2020), and whether any positive or negative trends emerge. It was discovered that the main contributor to pollution was the Energy sector, which had been on a steady rise until 2015 - same time when the Paris agreement was introduced. From this year forward, a steady fall was noticed in the graphs. The other sectors were not as affected by this political decision, and some were even still rising. In general, the emission levels of the top polluting countries and their regions have also been on the decline since the Paris agreement were introduced.

In conclusion, the most polluting countries and sectors have experienced big shifts during the last 14 years, which correlate with the introduction of political changes. The graphs were created to visualise these changes in the data-set in an easily digestible way. Through animation, interactive sliders, and plotly graph functionality, the user is empowered to drill and gain a deeper insight into specific parts of the data-set. Further enhancements such as colour-blind friendly themes and other usability features could be introduced in a future version.

Contents

Shiny App https://antje12.shinyapps.io/Data-Vis/

1 Background and Motivation

The environment is threatened like never before due to pollution being let out into the atmosphere. This sentiment has become more prevalent during the last several decades due to environmental changes starting to affect our day-to-day lives. Due to this, countries around the world have implemented several barriers by not only introducing CO2 quotas, but ways to limit the amount of pollutants released by different industrial sectors [2, 4].

Whether this has actually made an impact on the pollutants released into the environment, remains to be determined. Different studies have been conducted in an effort to investigate the negative evolution observed in relation to the current environmental changes. This includes the collection of a wide variety of data from different industrial sectors. During the execution of the project, data will be collected and act as the foundation for the investigation into industrial emissions in the last 15 years. Based on the results, an effort will be made to try and conclude whether or not political intervention has had any positive effects environment-wise, with the main focus being a reduction in industrial emissions.

Using data collected on European countries and their biggest industries, it will be possible to illustrate the evolution of pollutant emission for each country and sector, identifying potential positive or negative trends. Based on these observations, it should be possible to conclude which countries or sectors have the biggest impact on global air pollution as well as which are actively trying to reduce their pollutant emissions. The results will consist of visual representations of selected data where only air pollution will be the main focus. Different data-sets exist in which other types of pollutants, such as water and ground pollution have been collected and documented, however these will not part of this investigation.

The results and analysis obtained by the execution of this project could potentially serve a wide variety of users; specifically, European politicians to advance their focus and research on emissions created by European countries, resulting in further efforts in order to minimise air pollution of the aforementioned countries. To make a difference, it is imperative to know the scope and severity of a given problem, which is what this project aims to illustrate with our results and report.

2 Project Objectives

Based on the project objectives given and the data-set selected by the group, a list of questions have been established. The questions each serve the purpose of illustrating parts of the problem as described in the the beginning of the report. These include an overview of the air pollutants being emitted by the different European countries, as well as an investigation of the development of air pollution created by industries during the 15 years described in the data-set. Answers to these questions will be illustrated through multiple visual representations consisting of different both static and animated graphs, enhanced through possible user interaction. As a supplement to these graphs, a textual explanation of the findings and conclusions will be elaborated upon during the final chapters of the report.

The questions are defined as follows:

1. Which countries are responsible for the majority of pollution?
2. How has pollution levels developed over time pr. country?
3. Is the capital the most polluted of the country?
4. Did the UN Paris agreement have an impact on the amount of emissions? Has the emissions increased, decreased or stayed the same?
5. How has pollution levels developed over time pr. industry?
6. Which industry sectors are responsible for the majority of pollution?
7. Which pollutants are mostly released?

3 Data

The data set *"F1_4_Detailed releases at facility level with E-PRTR Sector and Annex I Activity detail into Air"* [1] is provided by the European Environment Agency (EEA) and is a subset of a larger database which covers waste transfers, energy inputs and emissions in Europe including Iceland, Liechtenstein, Norway, Serbia, Switzerland and the United Kingdom. The original data set contains a wide variety of emissions; however, it is, in this case, limited to air emissions [1].

As of writing, the data set contains approximately 300,000 records spanning from year 2007 through 2020 and is distributed across 14 variables; however, only nine variables are used in this project. Refer to Table 3.1 for a summary of the variables being used. The other five discarded variables are mainly key or code fields for the given country or industry which do not serve or provide any meaningful information considering the project objectives.

The data set is provided by EEA in a user friendly CSV format, and as such, does not require any substantial data processing. The unused variables are easily discarded using `dplyr::select`. Two of the questions pertain to capitals and regions which are not distinguished by default in the data set. Therefore, two custom made data sets containing capital and region information is left joined to the original data set, introducing the following four new variables:

- countryName (key used to left join)
- capital
- capitalInLocal
- region

For some countries in the original data set, both the English name and the native name are used interchangeably for the capital, which is the reason why the capitals data set contains two names. Both of these names are matched against the city to add to the original data a new boolean variable called "isCapital". Once processing is finished the temporary "capital" and "capitalInLocal" variables are dropped using a negative sign $(-)$ in front of the variables with `dplyr::select`.

The regions are based on the United Nations Geoscheme, which divides the world into regions and sub-regions [3]. The 32 countries were hereby divided into the four regions: Eastern, Northern, Southern, and Western Europe.

Since the pollutant emissions are reported in Kg, the numbers are exceptionally large and therefore hard to represent in a visually pleasing way. For this reason the reported "emissions" variable has been divided by 10^6 to get rid of the scientific number notation in the plots, instead reporting in 1000's of tonnes.

The chosen variables, their types, and descriptions are defined in the following table:

#	**Variable**	**Type**	**Description**
1	countryName	Categorical	Reporting country where the facility is located
2	eprtrSectorName	Categorical	E-PRTR sector describing the industrial sector of the facility
3	facilityName	Categorical	Facility emitting the pollutant
4	Longitude	Numerical	Longitude of the facility
5	Latitude	Numerical	Latitude of the facility
6	City	Categorical	City where the facility is located
6	isCapital	Categorical	Boolean field that describes whether the city is the capital
7	pollutant	Categorical	Substance emitted to the air
8	emissions	Numerical	How much is emitted by the facility in kg pr. year
9	reportingYear	Categorical	The reporting year of the emission
10	region	Categorical	The European region as defined in the United Nations Geoscheme

Table 3.1: Variables

4 Visualization

In order to successfully address the project objectives and answer the questions based on the selected data, a collection of varying visual representations will be made. The final product will be the delivery of a shiny app containing all graphical representations making the solutions easily accessible and enables the user(s) to jump between the different visual representations depending on what question they wish answered. The graphs each represent a solution to one or multiple questions.

4.1 Features

Must have features
Since a multitude of different questions must be answered, it will be necessary to implement a variety of different graphs. As discussed in the Chapter 2, seven questions must be answered which will result in the creation of at least seven graphs, each illustrating the points specific to the related question. To ensure heterogeneity in the way the data is visualized, and create a focus on the specific points of the questions, at least three different types graphs will be employed.

To illustrate how data has evolved and transformed with respect to time an automatic time-slider will be implemented for some of the models. This enables the user to both view the model evolution throughout the years but also the selection of a specific year. An animation will also be implemented using the RStudio extension gganimate.

The report and application will share a two-way link to each other. The report will include a link to the publicly hosted shiny app. And the shiny app will contain a button allowing users to download the report as a manual from the dashboard.

Could have features
To ensure that people with visual impairments would still be able to use the created graphs, a few improvements could be implemented. Firstly, the selected colour scheme could be changed to ensure that people with different kinds of colour blindness could still use the graphs. An setting could be made, to allow the user to set their preferred colour scheme related to their specific blindness. Secondly, a dark-mode could be implemented, ensuring that people with low vision and light sensitivity could still use the graph without causing eye strain or fatigue. This could also be made as a setting, enabling the user to chose their level of darkness.

4.2 Questions Visualised

Question 1: *"Which countries are responsible for the majority of pollution?"*

To visualise how countries are responsible for emitting pollution, a map chart was chosen in combination with a bar chart. Each country in the data-set is represented as a geographical mass on a longitude / latitude graph coloured to indicate its pollution level. Certain countries do not have pollution data for all years, which is handled by simply colouring the country white in the case of missing data. To support this view of how pollution emissions are distributed across Europe, a bar chart showing the top-10 countries ranked by emission is added. To enhance the user experience and serve a better overview of the data points, a slider is added making it possible for the users to select a specific year and compare the pollution levels. This creates an interactive graph, encouraging the user to further investigate the data set. An automatic animation throughout the years can also be started by the user, by pressing the play button. To make the graph easier to interact with, the mouse-over functionality was implemented so users can simply hover over a country to see its specific pollution value.

By using both a map plot as well as a narrowly scoped bar chart help the user to see exactly how the countries compare to each other. In addition, the map plot illustrates a sense of pollution levels of Europe as a whole, while the bar chart focuses on the top-10 most polluting European countries. This combination of charts makes it possible to see where pollution is coming from and to what extent. An alternative would be to simply create a mean for each country, covering the entire period. This was not chosen since it would cover up the fact that some countries are lacking data points, which would skew the mean.

Question 2: *"How has pollution levels developed over time pr. country?"*

To visualise how a country's pollution levels have developed over time a time-series plot was chosen. Since visualising all the countries in the same graph could be incomprehensible for the user, an alternative region-based view has also been implemented. This view divides the 32 countries into the four European regions as defined by the UN GeoCode [3]. The graph shows the pollution levels as they have evolved from 2007 to 2020. If the user hovers over a point in the graph, a more detailed description of the values are presented. Via this graph it is possible to compare the countries on a yearly basis, as well as their individual pollution evolution. By inspecting at the rise and fall of the curves for each country it can be evaluated whether there are any trends linked to specific years in the data-set. These years could be explained by political decisions having an effect on the polluting industries.

Question 3: *"Is the capital the most polluted of the countries?"*

There are a couple of different noteworthy solutions to this question. If the data set is grouped by city type (capital or non-capital) then a bar chart could display the differences in pollution for each country rather easily. The only issue with this type of chart is the number of countries. There are 32 countries, which would yield 64 columns; hence, resulting in a cluttered graph. In an effort to reduce clutter, converting the graph to use facets with sub plots for each country

would significantly reduce said issue; however, displaying the sub plots in a grid style manner introduces new, minor issues when comparing almost identical numbers across rows. In other words comparing two bars, one of which is displayed in a sub plot on row one and the other on row three, might be difficult, especially if the neighbouring columns are of different heights as it creates an illusion that values may not be the same. Labels can eliminate this issue although.

A spine chart can achieve a similar result as a grouped bar chart in addition to reducing clutter. It splits the groups into two sub groups and plots them from both sides of a shared axis. This way highlights the contrast between capitals and non-capitals while simultaneously allowing for conveniently distinguishing values within its own group. Different colours for each measure can enhance differentiating between them and labels for the values help reduce confusion where the measures are of equal size. Another similar chart, the dumbbell chart, does the same job by using a combination of a bar or spine chart and a dot plot. The dot in this chart emphasises the actual value and the thinner columns help illustrate the difference. All in all, a dumbbell chart should result in a cleaner, focused looking chart and will be the driver for answering this question.

Question 4: *"Did the UN Paris agreement have an impact on the amount of emissions? Has the increased, decreased or stayed the same?"*

As a solution for this question an animated line-plot was created in order to visualise the development in pollution with respect to time. Hence the x-axis represents the categorical variable "reportingYear" which is the different years (from 2007-2020) the pollution has been occurring. The y-axis is the numerical variable representing the overall amount of pollutant for each country. Furthermore a dividing line has been added representing the year in which a major effort in minimising the amount of pollution was made. In this case the Paris Agreement from 2015 constitutes the dividing line (a vertical dashed line) hence changes observed after this line may be able to give some clarification in relation to the question proposed.

Question 5: *"How has pollution levels developed over time pr. industry?"*

To visualise how an industry's pollution levels have developed over time a time-series plot was chosen. The graph shows the pollution levels as they have evolved from 2007 to 2020. This enables easy comparison of the levels and evolution of the different sectors. By inspecting the time plot it can be seen how the pollution of the industry sectors across the countries have evolved during the registered years. Since the sectors are split up it can further be investigated whether political decisions have affected certain sectors more than others, or if some sectors are completely unaffected. Much of the same graphical initiatives from Question 2 applies to the visualisation of this question.

Question 6: *"Which industry sectors are responsible for the majority of the pollution?"*

Before each plot was made, the data had to be summarised to be able to see the bigger picture over the data and the sector emissions. The N/A sector were discarded as it was deemed irrelevant by the group even though some emissions could have been plotted by that sector.

However, as that sector is called N/A, it was difficult to determine the cause of the emissions thus being left out. Each sector was represented as a big block or bar respectively in the tree-map plot and bar chart. By doing so, the group could determine and get a sense of which sector was responsible for the majority of the pollution.

The bar chart was made more interactive and user friendly by allowing the user to drag and zoom into a specific part of the chart, hide sectors and even focus sectors by double clicking on a specific sector. There were other potential solutions that could have been used to gain an even deeper perspective in which industry sector was the most polluting and why; however, that would indirectly answer other questions set by the group. A sunburst plot was one of the potential solutions, where the user could hover over a specific sector and see which kind of pollutant was being released by which sector. This would cause an overlap between question 2 and question 5, which the group wanted to avoid.

Question 7: *"Which pollutants are mostly released?"*

The release of pollutants change continuously from year to year especially considering the world's focus on lowering CO2 emissions. For that reason the most suitable illustration is either a line or a slope chart. These types of graphs allow for displaying data as it evolves over time and, depending on the data set, enable a quick overview of the different groups in the data.
Alternative solutions include animated bar charts which show the yearly change or use a filter-widget to allow the user to change the year displayed.

Two line charts are used to answer this question which quite clearly displays a major gap between the top pollutants and the lower ones, therefore, using a single chart will render it impossible to properly read the lower end. Consequently, one of the charts will display the top pollutants, while the other display the rest.

4.3 Visual Styling

A custom colour pallet, consisting of a list of colours, was created and used for the majority of the plots (especially the line plots). The reason for this was mainly the fact that the default colour selector in RStudio did not ensure a high enough interval in colour variation making lines and areas hard to distinguish form each other. Various default colour pallets were tested however many of these did not contain enough colours creating the same problems as prior. However some of the custom RStudio colour schemes were selected, but this was more in relation to the saturation of a specific colour rather than the selection of a multitude of different colours.

Another important consideration was to ensure that the vast majority of potential users would be able to interpret the model(s) correctly. More specifically people with colourblindness were the source of discussion(s) regarding the selection of less red-green colours and potentially implementing a feature that would make it possible for the user to enable a colourblind friendly colour palette on all the plots.

5 Results

Question 1: *"Which countries are responsible for the majority of pollution?"*

During the years, the following conclusions could be drawn by looking at the generated pollution map combined with the top-10 most polluting countries.

The top polluters from the data-set are Estonia, Netherlands, and Greece. These 3 countries are in the top-3 for the majority of the data points. The Netherlands is the top polluter for 5 years and Estonia is the top polluter for 9 years, making these the most polluting countries during the time period of the data-set.

Since the bar-chart with the top 10 polluting countries is only able to visualise the most polluting countries, the lack of an overview of the entire data-set is apparent. To alleviate this, the geographic map comes into play. By analysing this map, it becomes clear that many countries have a much lower degree of pollution compared to the top countries, placing them at 100x1000 tons and below. The shifts in colour tells the story of which countries change their pollution levels, and which are stable.

Furthermore it can be inspected how bordering countries relate to each other. As an example, Germany, Czechia, and Austria seem to follow the same trends indicated by being in the same colour region of the scale throughout the years.

In combination these graphs creates an overview of which countries are main contributors to the global emission of air pollutants, while also showing the general trend of evolution.

Question 2: *"How has pollution levels developed over time pr. country?"* The time series plot visualises how the emission of air pollutants has developed for every country within the data-set. Three different views are relevant in order to solve this question which include a line plot for All countries and countries divided into four EU regions. The third line-plot displaying the mean emission for each EU region serves as a solution mainly for question 4 but is also a relevant solution for this question.

When analysing the evolution of pollution it becomes clear that a lot of the countries are relatively stable, or have a small downward going trend. Most of these countries already had a relatively low pollution level, which could indicate that they were never a target for international pollution changes. Few countries experienced major rises and falls throughout the timeline: Estonia, Netherlands, Greece, Czechia, and Malta. Most of these were also identified in the top-10 polluting countries, indicating that these have been main targets for implementing pollution limits.

This fits very well with what was expected by the group: The countries responsible for the majority of air pollution, would also be the countries mostly affected by global limitation politics. As can be seen these countries either experience a huge downward trend like Malta, or many up-and-down fluctuations like Estonia, struggling to limit their pollutant emissions.

When looking only at the regions, a general downward trend becomes apparent from 2015 and onward in the Southern, Eastern, and Western regions. The Northern region on the other hand struggles to reduce air pollutants. When comparing this to the graphs of all countries divided by region, it becomes apparent that this is mainly due to Estonia driving up the mean. This makes sense seeing as Estonia was the top contender in the top-10 of polluting countries.

Question 3: *"Is the capital the most polluted of the countries?"*

The general trend displayed in the dumbbell chart shows that for most countries the capital is the most polluted. The raw data includes every city; however, this visualisation only distinguishes between what is released in the capital and what is released outside the capital. One could think that such a grouping would result in non-capital pollution being significantly higher, although that is only the case with Estonia.

For Belgium, Germany, and Poland it is almost a tie, although it can be hard to determine solely on the visual representation it is possible to inspect the actual values. Amsterdam, the capital of Netherlands is the most polluted city according to the data. Similarly, the plot shows that pollution outside the capital of Netherlands is greater than the majority of every other capital, which raises questions as to whether this is the actual or if the Netherlands is simply much better at reporting their emissions.

By looking at the graph, it is not possible to tell if the capitals are more polluted than the rest; however, the following statistics were observed: the capital is more polluted in 14 of the countries, the capital is not the most polluted in 12 countries, and the remaining 6 countries contained a NA value in one of the categories. To conclude, capitals are the most polluted according to the data; however, it is with great uncertainty. The 6 countries with unknown status could go either way.

Question 4: *"Did the UN Paris agreement have an impact on the amount of emissions? Has the increased, decreased or stayed the same?"*

A general observation can be made in which an overall decline in pollution levels can be observed in many of the plots created. Specifically the line plots each show a small or significant decline depending on what they each represent. For instance in the plots displaying pollution levels for each country many of the lines can be observed as having a declining tendency. It is however not very significant visually. The line plot *"Mean emission for each European region"* does however display and emphasise a clear decline in emissions after 2015 proposing that the approval and implementation of the UN Paris agreement had a positive impact by minimising overall air pollution in Europe.

Question 5: *"How has pollution levels developed over time pr. industry?"*

The time series plot visualises how the emission of air pollutants has developed for every industry sector within the data-set. When analysing the graph it becomes apparent that the main sector of pollution is the Energy sector, having pollution levels three times as high as the as Paper and Wood, which comes in second place. This fit very well with some of the expectations the group had which was that the biggest source of pollution would be the Energy sector considering its substantial importance in a wide variety of different sectors.

The general trends of the sectors is that everything is either stable or slightly rising, except for the Energy sector which experiences a huge decline from 2017 and onward. This is not surprising, since the general discourse of climate change has historically been based on pollution from the Energy sector. What was surprising is that some other sectors have been allowed to grow in the meantime. Sectors such as Paper and Wood is on a general rise, not seeming affected by the political climate.

Question 6: *"Which industry sectors are responsible for the majority of the pollution?"*

By analysing the tree-map and bar chart, the group is able to determine which industry sector is responsible for the majority of the pollution in the EU.

Looking at the visualisations, it becomes evident which sector is the biggest polluter - the energy sector. It is the clear outlier compared to the other industry sectors, and there can be several causes as to why. This fits well with the narrative and does not come as a surprise to the group as, compared to other industries such Paper and Wood, the needs of the people on the planet are becoming bigger, thus requiring a bigger effort in maintaining that need, resulting in more waste and more pollution caused.

According to the European Environment Agency, energy processes in Europe are responsible for 78% of total EU emissions. These processes are tightly linked to the burning of fossil fuels for heating, electricity, transport in industry. Despite fossil fuel burning being the majority of the pollution released, there are multiple ways the energy sector can pollute.

Question 7: *"Which pollutants are mostly released?"*

Initially CO2 was expected to be among the top pollutants released; however, the visualisation has proven that not only is CO2 the top released pollutant, every other pollutant, excluding the additional CO2 measure, is almost negligible compared to it. While CO2 shows a decrease over the span of the data going from 2.2 billion tons in 2007 to 0.74 billion tons in 2020, the next in line is CO with 0.0012 billion tons; a mere 2000 times smaller than that of CO2.

In spite of the fact that some countries have yet to report their emissions, it is, judging by the chart, plausible that the majority of the missing data pertains to CO2. Moreover, the missing data might be the reason behind the massive drop seen in CO2 from 2017 to 2020.

6 Discussion

During the execution of this project, there were several challenges and considerations that needed to be addressed in order to achieve an acceptable answer for each of the questions. Finding the right balance between what to show and what to hide was a reoccurring challenge; especially with data heavy plots. That said, an acceptable balance was reached that showed the information the user would need to answer the questions, while abstracting from unnecessary details.

6.1 Graph Challenges

In question 1, the bar chart based on the top 10 most polluting countries improved the general overview and gave a clear answer to the question; however left out several countries that may have had an impact on the visualisation. Thus, the map plot was created in order to support the bar chart and give a more nuanced picture.

During the creation of the line plot(s) for question 2, different versions were discussed. The first line plot created represented the pollution of each country as an individual line with its own unique colour. The only disadvantage of this representation was that the plot became cluttered which made it hard to distinguish the countries from each other. Additionally, another consideration was to represent each company as a line, and colour code them based on country. However, this proved troublesome as the user would drown be overflown with information.

Therefore, a line diagram grouping all the countries based on their European location was considered. All 32 countries would be divided into 4 separate European regions based on the definition provided by the UN geocode [3]. Based on this graphical foundation, two representations could be made. Either each line representing each country would be colour coded based on what European region it belongs to, or 4 lines summarising the overall pollution of all countries in the given region could be made. Representing each country would create challenges like the ones in the first diagram. However instead of collecting all countries in one plot, 4 different line plots were created based on each region. This made the diagrams less cluttered and made it easier to observe the evolution of each line (country).

In both plots, the evolution of the pollution in each country is quite clear; however the group discussed the possibility of making the evolution of pollution even more obvious, especially when taking question 4 into consideration. Another line plot was created displaying the summarised air pollution for each European region. The creation of this plot made the variations in pollution levels, visually, more significant. Also, as an animated plot the amount of visualised data was small enough for a user to follow all 4 moving lines at the same time without the disturbance of a multitude of different lines.

The prior challenges mentioned for question 2 and 4 were not apparent in question 5 as the number of lines were minimal and an acceptable overview could be achieved. A potential improvement could be to use two different scales as in the line chart for question 7.

In question 7, the line chart proves to be exceptional at displaying data over time. It is impossible to properly illustrate data as the data is distributed over two different scales. As a result, the data is split into two parts based on the scale and plotted in separate line charts accordingly. Resulting in comparing smaller values against each other more efficiently.

The highlight function is implemented in questions 2, 5, and 7. It works by hovering over a line. To reset the view, the user must double click. There is no built-in way to remove the highlight from the last focus once the user leaves the chart, so the users must explicitly double click the chart to remove the highlight before leaving which is not the optimal functionality the group had aimed for.

The dumbbell chart from question 3 proves slightly inefficient at displaying the differences for each country (the x-axis), especially when the dots are far apart or when they are almost a tie. When it is ordered, however, it is effective at displaying for one part of the category (capital or non-capital) the differences over the y-axis. A grouped bar chart would significantly improve the first issue; however, was scrapped as 64 columns would be too extensive.

In question 6, several charts were considered and tested. In the beginning, a violin plot was chosen in order to visualise the distribution of numerical data between industry sectors. With the help of a violin plot, the group could have determined the primary pollutant sector. However, due to the scale and spread of the data, it was difficult to plot the data. Thus other chart techniques were chosen such as a traditional bar plot and a treemap.

6.2 General Styling

The colour blind feature was not implemented since the number of distinct colours available in this spectrum was too limited compared to what was needed in the graphs. A trial of adding more custom colours to the spectrum was tested, which introduced a trade-off of making the different data points less distinguishable if they had to follow a more colour friendly tone. For this reason colour blind features was pushed to a future version of the app.

7 Conclusion

During the execution of this project, several varying representations were created from the data-set, each serving the purpose of exploring the evolution of air pollution in Europe. The goal was to answer the research questions presented in the motivation for the project. The main benefit of the report would be an indication of whether the pollution has been rising or falling in tune with the political climate in Europe.

With the completion of the project, a total of ten varying visual solutions have been created each serving as the answer to one or multiple questions. The group managed to answer all of the questions proposed while at the same time giving a clear overview of the data set ranging from the abstraction of "Country" level all the way down to "Emissions" level. These graphs served as a powerful tool to investigate and answer the proposed questions.

In conclusion, it was discovered that the major pollution released was caused mainly due to CO2 being emitted from the Energy sector. This sector had been slowly growing from the start of the data-set (2007), until the implementation of the Paris agreement in 2015. From this point onward, a variety of changes were made in the European industry resulting in the decline of emissions from the energy sector. Since the political focus had only been on this energy sector, the other sectors were either unaffected or on the rise. In summary, the most polluting countries and sectors have experienced big shifts during the last 14 years, which correlate with the introduction of political changes.

As an improvement of our investigation the logical step would be to widen the emissions data-set to include other emission such as water and ground pollution and include this in the Shiny app. By doing so, a better indication of the evolution of pollution throughout the European industry could be achieved.

Additional quality-of-life improvements could be to assist people suffering from visual handicaps such as colour blindness, light sensitivity, or other visual impairments, allowing the graphs to be read and understood by a wide variety of users. These personal aids could have been implemented as on-demand services, which could be enabled globally via the settings in the application.

Bibliography

[1] European Environment Agency. Industrial reporting under the industrial emissions directive 2010/75/eu and european pollutant release and transfer register regulation (ec) no 166/2006. https://www.eea.europa.eu/data-and-maps/data/industrial-reporting-under-the-industrial-6, 2022. Accessed: 2022-09-20.

[2] CPCC. Global warming of 1.5°C. https://www.ipcc.ch/sr15/, 2018. Accessed: 2022-10-26.

[3] United Nations Statistics Division. Standard country or area codes for statistical use (m49). https://unstats.un.org/unsd/methodology/m49/#geo-regions, 2022. Accessed: 2022-11-02.

[4] Sam Meredith. 'it's now or never': World's top climate scientists issue ultimatum on critical temperature limit. https://www.cnbc.com/2022/04/04/ipcc-report-climate-scientists-issue-ultimatum-on-1point5-degrees-goal.html. Accessed: 2022-10-26.