

ABSTRACT

- I. This project used the following frameworks – **Pandas**, **NumPy**, **Matplotlib**, **DuckDB's SQL** functionality within Python, and **Tableau** – to produce the following, all within 48 hours:
 - A. Performed EDA on greenhouse gas emissions dataset from Climate TRACE¹.
 1. Created 15 figures:
 - a) 5 bar graphs
 - b) 1 pie chart
 - c) 4 global heatmaps
 - d) 5 correlation matrices
 - B. Created a Tableau Public Story² consisting of two dashboards
 - C. Wrote a three-page executive summary of the work, which you can read below.

INTRODUCTION

This project explores datasets from Climate TRACE containing greenhouse gas (GHG) emissions from coal mining, and gleans insights from them. Climate TRACE is an organization that uses independently-verified data, sourced and derived from advanced technologies like satellite imagery, enhanced with machine learning. Climate TRACE and the values and needs of any sustainability-minded person or company are closely aligned; the work described here will show the promise of using Climate TRACE's data to help everyone make informed, sustainable decisions.

METHODS

There were three main tables that were used for this project³ and are listed below:

1. GHG emissions sourced from coal mining operations, grouped by year (2015-2022) and country
2. A table highlighting individual sources for emissions (i.e., individual mines) with other relevant data like country included, and
3. A table listing the corporate owners of those mines along with their revenue.

An extensive exploratory data analysis (EDA) was performed using Python and Tableau to understand the relationships of the data. Figures created from the EDA can be found in the Appendix section below. I used **Pandas**, **NumPy**, and **Matplotlib** to manipulate and plot the data, and used **DuckDB's SQL** functionality within Python to manipulate the tables further before being plotted. The Python notebook shows the steps taken to create the figures.⁴

INSIGHTS

Broader Context

Before we focus on the companies operating the mines, we should take a moment to learn about the global trends in coal mine emissions. Global emissions of methane due to coal mining were on a steady increasing trajectory, going from about 42.2 million metric tons in 2015 to 51.5 million metric tons in 2022, an increase of about 20% (**Figure 1**). In terms of the biggest emitters globally, for all the years in the dataset, China, USA, Australia, and Russia claim the top four spots respectively (**Figure 2**). China, the first on this list, emits about ten times as much methane emissions compared to the number two, USA. China and USA are also number one and two, respectively, on total emissions sources (**Figure 3**). We are starting to see a pattern emerge of which countries emit the most, which is validated in the news stories we read on what seems to be on a near-daily basis.

The countries that have little to no mines include Argentina, Bangladesh, Spain, Nepal, and Egypt. In terms of emissions sources, solely underground- and solely surface-based emissions make up almost 98% of all emissions sources, with the other few percentage points going to mines with

both underground and surface areas, and refuse (**Figure 4**). A global overview of the countries included in the dataset with measurable emissions is found in **Figure 5**.

An analysis of correlations (categorical and continuous data respectively) in the emissions sources dataset yields some minor insights, like a mild positive correlation (0.44) between source type and country (**Figures 6 and 7**).

The distribution of coal mines' emissions roughly follows the broader lithological distribution of coal beds, so we can expect to see coal mines in the eastern half of the US, in parts of central and eastern Asia, South Africa, and eastern Australia, for example (**Figure 8**).

The correlations in the ownership dataset show nothing too revealing beyond the expected correlation between capacity and activity (**Figure 9 and 10**).

Parent Company-Focused Insights

With the overall coal mine emissions picture understood, let us turn our attention towards the main focus of this project: the companies that own these mines. If we want to promote sustainable business practices, it is important that we look at the companies running their mining operations. As stated, this section of the analysis was completed in Tableau. I created two dashboards to focus on two categories: 1) the top twenty companies whose holdings are the most polluting, and 2) the bottom twenty companies whose holdings are the least polluting.

The patterns observed in the global context are largely replicated here, with Chinese and Australian companies being the most emissive. I found it interesting to look at revenue for these heavy emitters, and found that there was a diversity of countries represented — though Chinese companies were still the most common heaviest polluters found within the top twenty revenue group.

Analyzing the bottom twenty companies involved more legwork as I opted to filter out the companies who did not have any emissions. It is not practical to have a measured emission of zero, this instead must represent a lack of data for those companies. In this group, we see various countries here, with Myanmar-based companies holding six of the twenty slots. The revenue for these lower-emitting companies was not always readily available, though we see an extreme outlier of Norway in the form of the Ministry of Trade and Industry being very profitable compared to the others. The mines that comprise the bottom twenty group include those from Indonesia, India, New Zealand, and the Philippines.

Additional Data Source: Iron Mining Emissions Data

Using the mineral extraction data downloaded from the Climate TRACE website,⁵ I opted to focus on iron mining emissions. Although it is not a perfect correlation, I thought that iron deposits could be found in places that are rich in coal. The other minerals included in the download package from Climate TRACE included copper and bauxite. From my geology background, I know that copper and bauxite are not usually found in places with coal beds. I did explore other data sources, like forest clearing, but the constraints that I put this project under necessitated me using data that was more readily manipulatable.

Similar figures to the ones created for the broader context section above can be found in **Figures 11 through 15**. We can see that contrary to methane emissions, CO₂ emissions dropped by about 30% in 2022 compared to the highest year in the dataset, 2019. The usual big emitters are represented in the iron mining emission data as well, though the order of the top five most emitting is India, China, Brazil, Russia, and Australia. USA is sixth.

Was my hypothesis correct? Did iron ore deposits correlate with coal deposits? Although **Figure 14** suggests that it is correct, the true picture may have another explanation: We may be observing a quirk in how minerals are distributed in the earth's crust. There are numerous⁶ articles⁷ discussing⁸ the geopolitics of rare earth minerals' distribution, for example, and China is blessed with

rich deposits of many of these. We might also be able to witness the difference in sustainable priorities among nations, and therefore their companies. While the USA is a big emitter of GHGs, some recent administrations have taken steps to reduce the USA's contribution. The same level of effort may not be there in other countries. Norway, for example, seems to have struck a better balance than most with their revenue and emissions.

DISCUSSION

It goes without saying that this project should in no way be used for investing advice, nor should it be used to inform a company's — let alone a country's — GHG emission reduction plan. What this exercise does show is that ClimateTRACE is able to highlight companies that are at the forefront — or lagging behind their peers — of sustainability. I demonstrated here that no mining company can eliminate all of their GHG emissions, but some appear to do it better than others. The home country where the company is based looks like it has a large influence on the emissions behavior of the company.

Just as with the organic food movement, the finance industry should adopt these same principles to create an “organic finance” system, to allow for transparency and real-economy-based products.⁹ As part of that, especially in a warming world, people want to try and make a difference in any way that they can. Investing in sustainable companies, or entities that are better than their peers, are good steps that people and entities should take.

Questions and Potential Future Work

This exercise brought up many questions and showed the way towards future projects:

- What are the largest-emitting companies doing that causes them to pollute so much?
 - Additionally, what are the companies doing that have lower emissions?
 - Is it simply a measure of scale? Future work could help disentangle this question.
- If there was more time, I would look into the entities on the top twenty list and bottom twenty list.
 - What are the companies doing in terms of their communities that they operate in?
 - Are there other metrics beyond emissions that we can tie to the companies?
- Using more data sources.
 - I have a personal interest in geospatial information. If there was more time, I would attempt a deeper geospatial analysis of the mining companies and their relationship to their surroundings, not just for their community but also for the natural environment, like investigating the case of mountaintop removal mining practices.¹⁰
- A recent article from the New York Times described a study that showed that the 2023 Canadian wildfires, if they were a country, would be fourth on the list of biggest emitters that year.¹¹ I have a previous project¹² where I trained a model to predict whether a location will burn based on its satellite image and topographic information (aspect, elevation, and slope). I imagine that for Climate TRACE it can be difficult to figure out where the emissions are coming from, and so a future project could be focused on increasing the accuracy of where one attributes the source of emissions.

APPENDIX

Figure 1. Global emissions, in metric tons per year, of methane for coal mining. The y-values are on the order of 10,000,000.

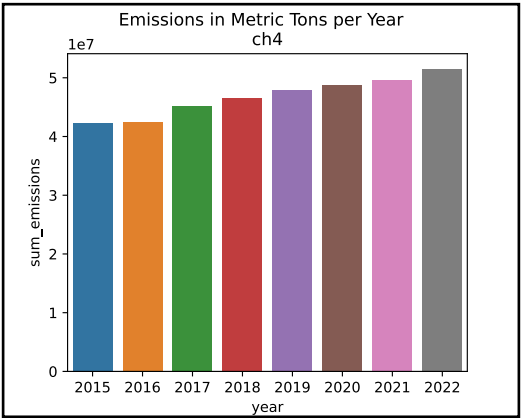


Figure 2. Total methane emissions by country (metric tons, log-scale).

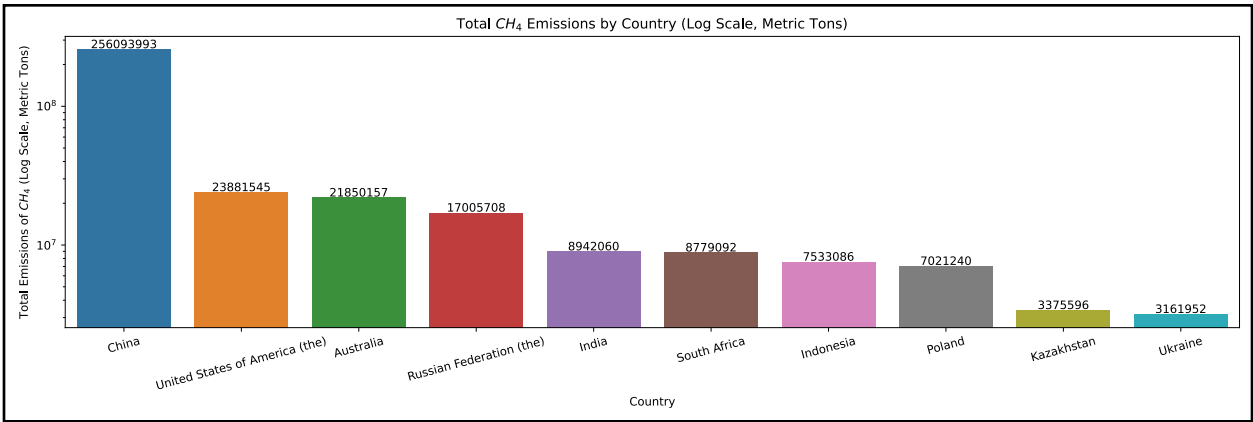


Figure 3. Count of emissions sources by country (metric tons, log scale). These sources are mines.

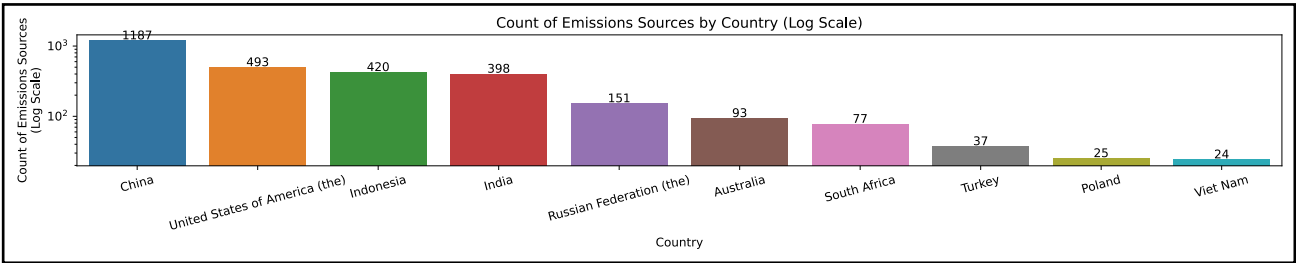


Figure 4. Pie chart of mine types.

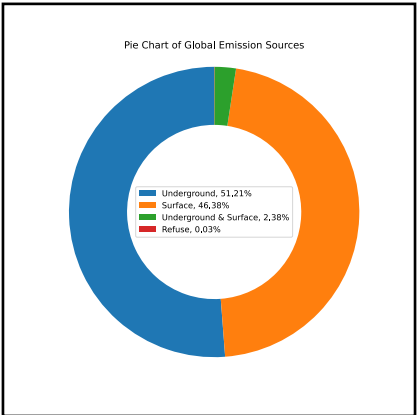


Figure 5. Global heat map of countries with coal mine methane emissions. The scale is relative, not absolute.

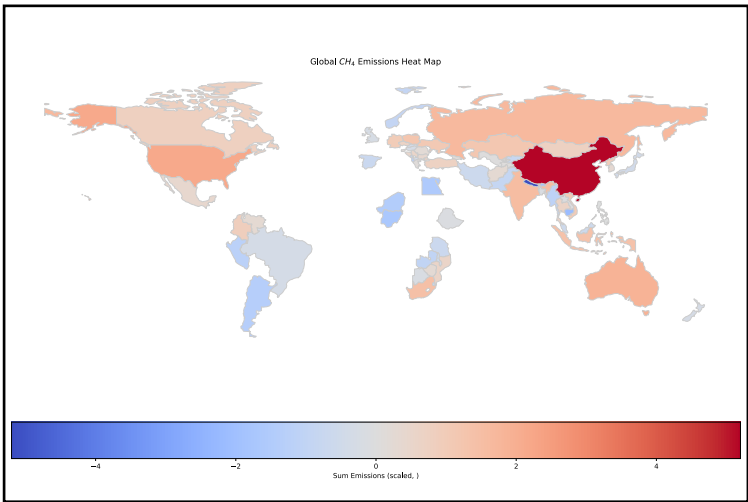


Figure 6 and 7. Categorical and numerical variables correlation table, respectively for the emission sources table.

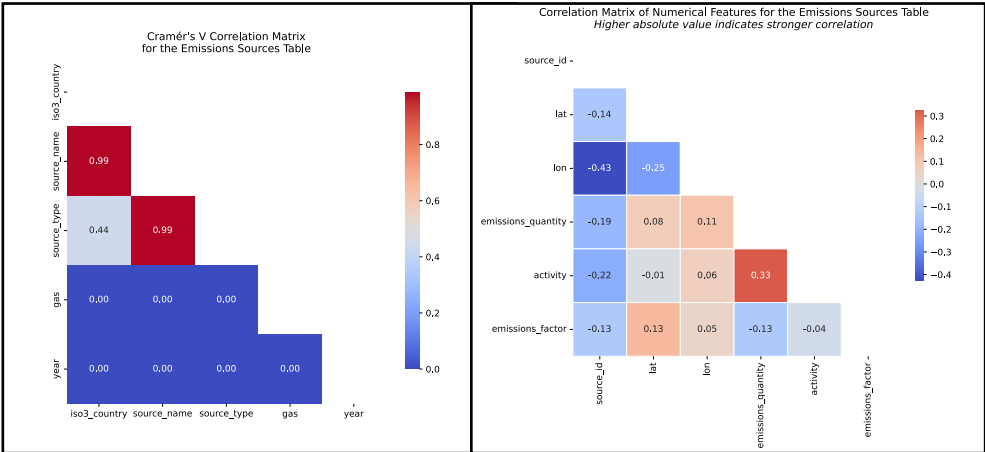


Figure 8. Density map of comparative global methane emission levels from coal mines.

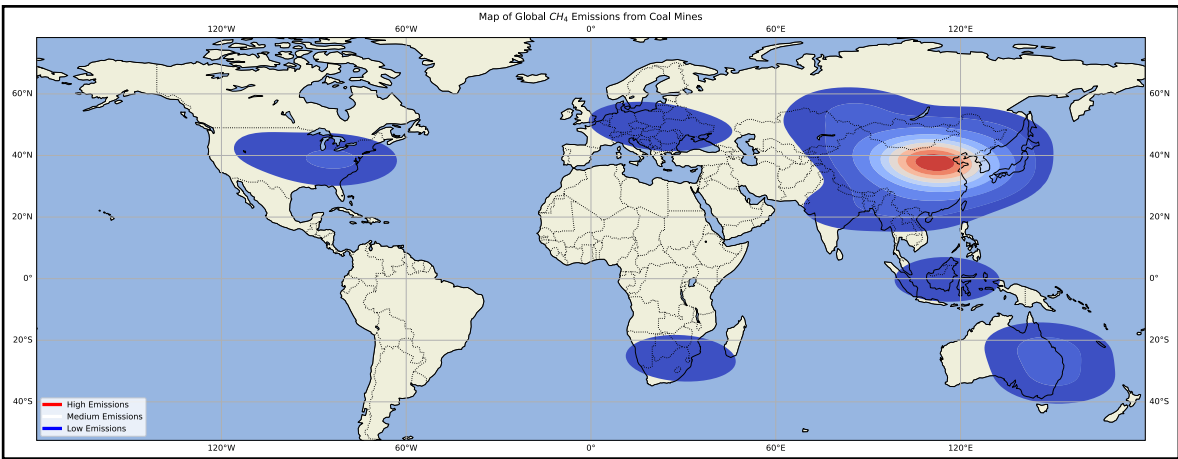


Figure 9 and 10. Categorical and numerical variables correlation table for the mine ownership table.

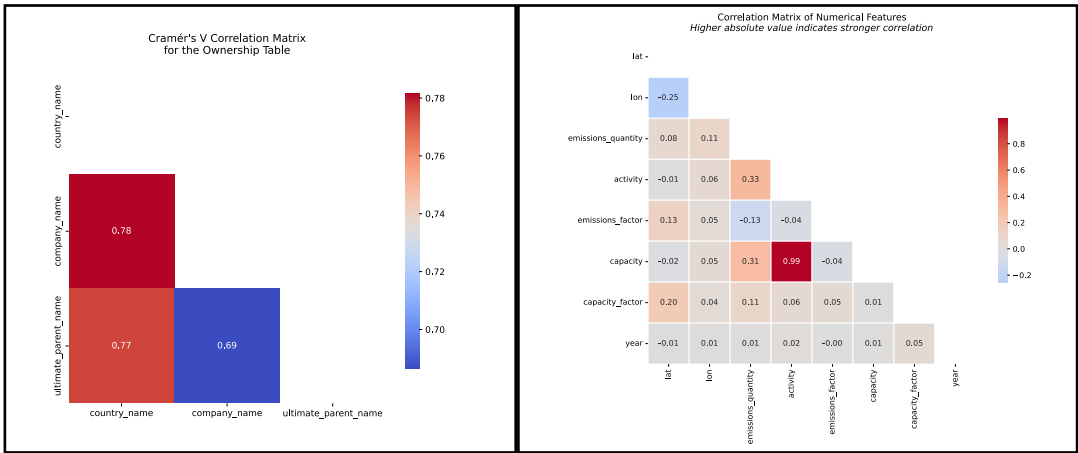


Figure 11. Global emissions, in metric tons per year, of carbon dioxide emissions from iron mining. The scale is on the order of 100 million metric tons.

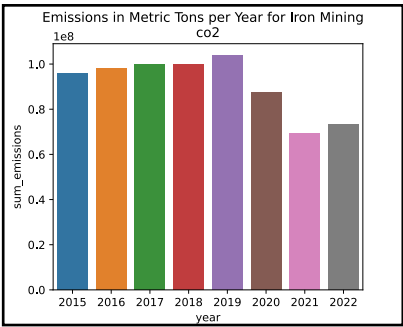


Figure 12. Total iron mining emissions by country (metric tons, log scale)

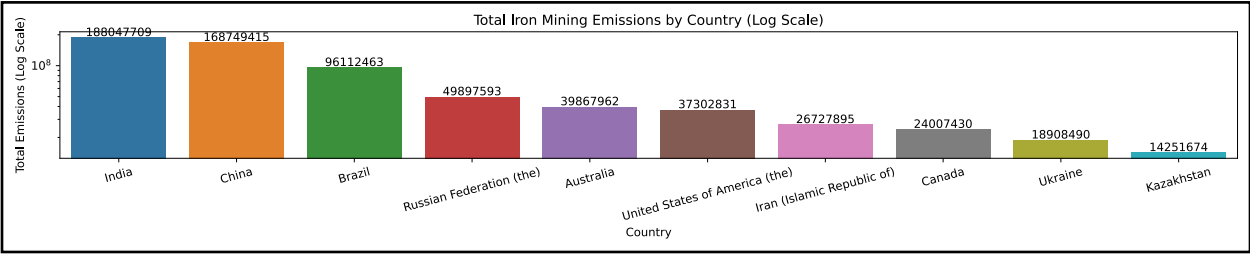


Figure 13. Global heat map of countries with iron mining CO₂ emissions. The scale is relative, not absolute. The iron mine with the most emissions is found in Brazil.

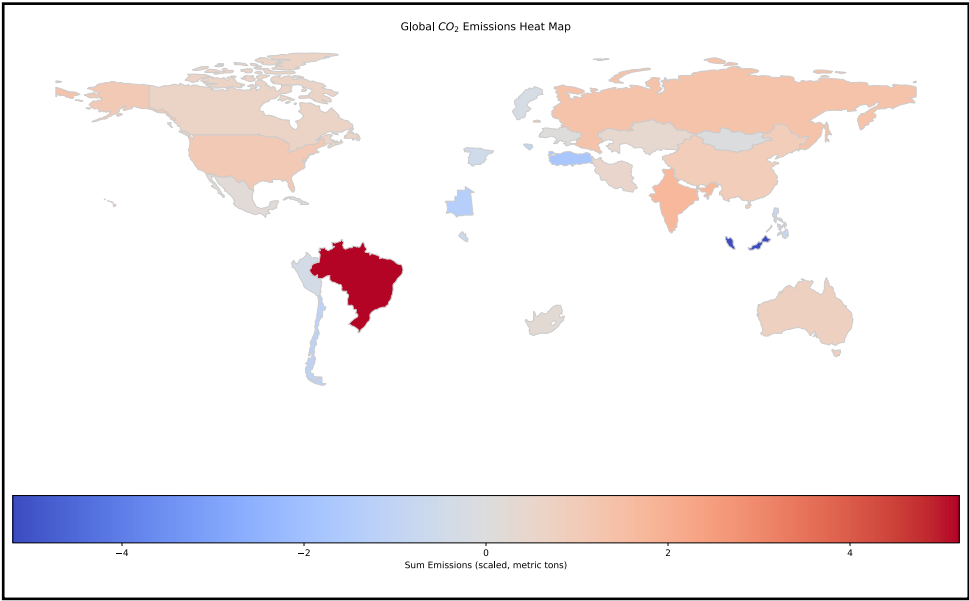


Figure 14. Density map of comparative global carbon dioxide emission levels from iron mines. The area with the highest cumulative emissions is South Asia.

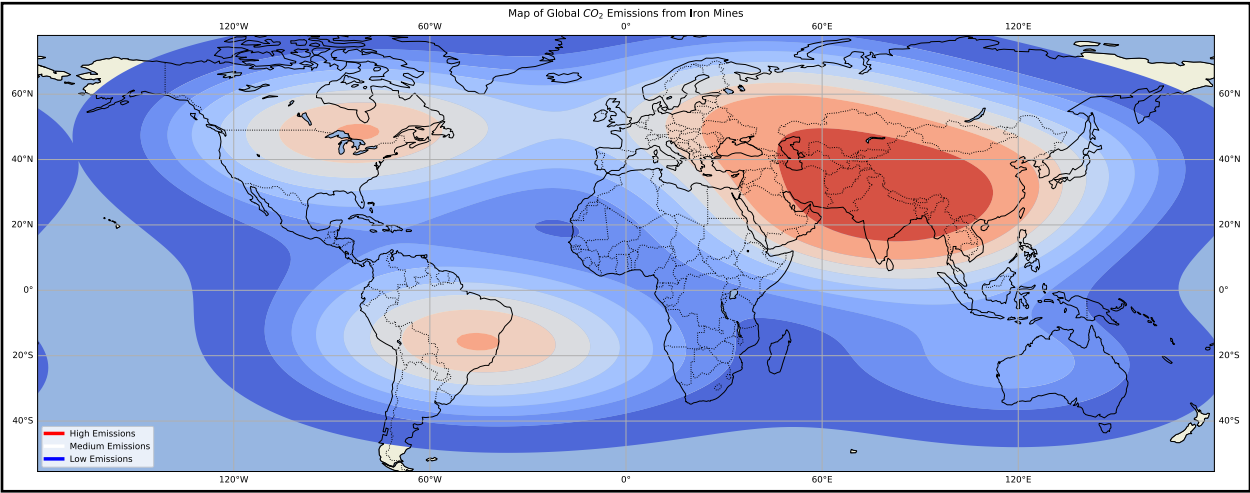
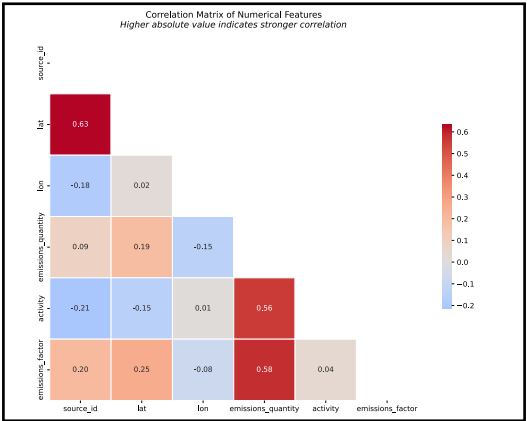


Figure 15. Continuous variable correlation matrix for iron mine emissions of carbon dioxide. The categorical correlation matrix was not instructive.



ENDNOTES

¹ <https://climatetrace.org/>

² https://public.tableau.com/app/profile/samuel.alter/viz/sustainability_analysis/EmissionsAnalysis

³ <https://climatetrace.org/data>

⁴ https://github.com/sralter/sustainability_insights

⁵ https://downloads.climatetrace.org/v02/sector_packages/mineral_extraction.zip

⁶ <https://www.reuters.com/markets/commodities/chinas-rare-earths-dominance-focus-after-mineral-export-curbs-2023-07-05/>

⁷ https://www.jstor.org/stable/resrep00311?searchText=Chinese+rare+earth+geology&searchUri=/action/doBasicSearch?Query=Chinese+rare+earth+geology&so=rel&ab_segments=0/basic_search_gsv2/control&refreqid=fastly-default:bb962a346827d9e1fc6b6a94dea225ef&seq=1

⁸ https://www.jstor.org/stable/resrep23660?searchText=Chinese+rare+earth+geology&searchUri=/action/doBasicSearch?Query=Chinese+rare+earth+geology&so=rel&ab_segments=0/basic_search_gsv2/control&refreqid=fastly-default:bb962a346827d9e1fc6b6a94dea225ef

⁹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2696448

¹⁰ <https://www.jstor.org/stable/41263246>

¹¹ <https://www.nytimes.com/2024/08/28/climate/canada-wildfires-emissions-carbon.html>

¹² https://github.com/sralter/brainstation_2023_ds_capstone