CentraleSupélec

Plastic pollution across different countries:

How is plastic disposed in different countries of the EU, how is the trend evolving over the last years and what is the relationship between these and the generation of plastic?

Davide Rendina
Computer Science Department
CentraleSupélec, Université Paris-Saclay
davide.rendina@student-cs.fr

Margarita Hernández Casas
Computer Science Department
CentraleSupélec, Université Paris-Saclay
margarita.hernandez@student-cs.fr

1 Finding Data

1.1 Selected datasets

Dataset 1: Eurostat - Plastic management operations in EU countries, over the years

This dataset collects data on how plastic waste is disposed in different EU countries. It contains records from 2004 to 2020, every two years. Although it contains some empty information, it is a very comprehensive dataset as it is divided into two categories: disposal (landfill, incineration, other) and recovery (energy recovery, recycling and backfilling).

Additional dataset: Eurostat - Plastic generation

In addition to the previous dataset, we found the total amount of plastic generated in the same countries (European Union) that would help putting into context the proportion of it that is being disposed and recovered.

Collected by: Margarita Hernández Casas

Dataset 2: OECD - Plastic waste by country and end-of-life fate, over the years

Similarly to the previous dataset, it provides information on how plastic waste is dealt with, including different types of disposal methodologies (i.e. incincerated, landfilled, mismanaged, littered and recycled). This datasets contains information about OECD countries, which includes 38 countries from all over the world [OECD, 2022], these are group based on continent.

Collected by: Davide Rendina

Dataset 3: OECD - Particulates PM2.5 by source and country, over the years

This dataset provides information about national emissions of the chosen air pollutant: Particular Matter (PM2.5). The data is broken down into categories regarding the source of pollution: mobile sources, combustion, power stations, waste, etc. The dataset does not include non man-made emissions, international aviation or maritime transports emissions.

Collected by: Margarita Hernández Casas

Dataset 4: Marine waste incidences in the Pacific

This dataset is provided by [of Environment, 2022], it contains information about marine pollution in the Pacific ocean. The type of pollution is divided into two main categories, these being 'waste dumped overboard' and 'oil spillage and leakage'. In addition, it is also provided the material (e.g. metal, plastic etc.). The quantity of each pollutant is also indicated, although using various different measurements. The country affected by the pollution is also provided alongside the coordinates (latitude and longitude).

Collected by: Davide Rendina

1.2 Dataset choice

The main datasets we are going to work with in this project will be the Eurostat - Plastic management operations in EU countries, over the years. Taking into consideration the feedback received in the first assignement regarding the question number 3, we are also going to integrate the Eurostat - Plastic generation. This will help provide insights regarding the relationship between waste treated and waste produced.

Both datasets have been collected by the European-Commission [2022], specifically by Eurostat which is the statistical office of the European Union. The two datasets were downloaded on Saturday 8th October 2022.

The datasets were first discovered by reading the study conducted by Bucea-Manea-Ţoniş and Zecheru [2022], which uses the Eurostat data about plastic management operations to compare how selected european countries manage and treat waste in relation to the Greenn Deal objectives set by the European Union. The second dataset, related to the plastic generation, was found browsing the Eurostat platform.

Originally, the dataset was in a wide format with each year represented in a different column, which would be not ideal for further visualising the data. Moreover, the different plastic management operations were not selected in the default dataset. In order to retrieve the datasets in the desired format, respecting the tidy data principles, the variables had to be rearranged from the Eurostat website using the interface provided, as shown in Fig 2 and 1 in Appendix B: 'Geopolitical entity' and 'time' were set as rows, and 'waste management operations' was added and set as column. Same process was followed for the second dataset from Eurostat that will be merged with this one. These changes were performed so the tidy data principles are met: observations as rows and variables as columns. Fixed variables such as country and year are placed in the first columns.

1.3 Research question

After thoroughly looking for datasets to answer the three proposed research questions, we decided to continue with the question number 3, which shows how plastic is disposed in different countries. The main reason was the amount of data found, as well as the quality and reliability of it. In fact, data regarding marine pollution, needed to answer question number 2, was scarce and often from non-governative sources. Similarly, only one comprehensive dataset was found about air pollution (Dataset 3) but none regarding commuters, which would have helped us answering question number 1. In addition, question 3 was also the one that received more interest from the Assignment 1 feedback.

The originally proposed research question was: "How is plastic disposed in different countries?". Following the dataset selection and the feedback received, this was rephrased to "How is plastic disposed in different countries of the EU and how is the trend evolving over the last years?". In fact, both selected datasets contain information about countries in the EU which led us to narrow the focus on EU countries, in contrast with the original and more broad research question. Moreover, we decided to also consider how the trend evolved throughout the years. In addition, in order to make the question more exploratory and potentially more interesting, we decided to incorporate the Eurostat dataset regarding the plastic generation, which in the end led us to reformulate the final question to: "How is plastic disposed in different countries of the EU, how is the trend evolving over the last years and what is the relationship between these and the generation of plastic?".

The aim of this analysis would be to provide a comparative study of the different plastic disposal methodologies adopted in different countries of the EU across the years, to find how the trend changed and how these relate to the global plastic generated in each country.

2 Data Cleaning

In this section we are going to describe the cleaning process that led to the generation of the attached dataset. This will be documented with screenshots attached in the Appendice of this report.

First, the two datasets downloaded from Eurostat - Plastic management operations in EU countries, over the years and Eurostat - Plastic generation had to be merged into one single dataset. Eurostat offers the option to download the selected data in csv format, but the output is not as clean as one may expect. This is why we created a new sheet in each file, pasting only the desired dataset here and saving only this new added sheet. Also, when inspecting the tables, we discovered that the columns containing data regarding waste generation and disposal were in non-numeric format. In addition, one dataset used commas as thousands separator and the other one used dot. Therefore, we decided to remove any type of thousand separator directly from the Excel files and save them to continue with the cleaning.

Using Python, we merged the two dataset on the columns *TIME* refering to year and *GEO (Labels)* refering to each country.

The merged dataset was loaded on OpenRefine and the following steps were performed to clean the data:

- The column 'Column', generated by the merge, was dropped
- 'Geo (Labels)' column was renamed to 'Country'

- 'plastic wastes' column was renamed to 'plastic_waste_generation, as it is more indicative of what the value represents (i.e. the total plastic waste generated)
- All other columns were renamed to remove the ID included in the name (e.g. D1)
- Missing data in the dataset is marked with a ':'. Therefore in order to better deal with missing values, we remove this by replacing it with an empty string using: value.replace(":","") and transform to number
- Columns 'Disposal - landfill' and 'Disposal - other' were dropped. This because there is already a column called 'Disposal - landfill and other'. In addition, the column 'other' add multiple zero values and it is not specified what 'other represents.
- Two new columns were created, namely 'Disposal - Total' and 'Recovery - Total'. As the name suggests, these contain the some of the values in the correspondent category for each country.
- 'Germany (until 1990 former territory of the FRG)' and 'Kosovo (under United Nations Security Council Resolution 1244/99)' were renamed to 'Germany' and 'Kosovo'.
- Entries about 'European Union - 27 countries (from 2020)' and 'European Union - 28 countries (from 2013-2020)' were removed. We assume this duplicate entries are due to Brexit, which removes United Kingdom in 2020. In addition, the total value of each attribute is not relevant for our purpose as we compare different countries individually.
- Entries (rows) about 'Albania', 'Bosnia', 'Kosovo' and 'Liechtenstein' were removed, as this containing for the most missing values (as it can also be seen in Fig. 3 and 4 in Appendix B) Rows with data before 2010 (2004-2006-2008) were removed as this contained missing values for all the countries. As can be seen in the screenshot in Fig. 5 in Appendix B Rows with missing values, namely 'UK' in 2020, 'Montenegro' 2010 and 'North Macedonia' 2016), were also removed. This can seen in Fig. 6 in Appendix B.

For reproducibility purposes, the .json file containing the history of the operations performed in OpenRefine and the original dataset can be found, alongside the cleaned dataset and all material used to work on this assignment, in the shared OneDrive folder in Appendix A.

References

R. Bucea-Manea-Ţoniş and T. Zecheru. Untapped aspects of waste management versus green deal objectives. *Sustainability*, 14(18), 2022. doi: 10.3390/su141811474. URL https://www.mdpi.com/2071-1050/14/18/11474.

European-Commission. Eurostat, 2022. URL https://tonga-data.sprep.org/dataset/marine-pollution.

OECD. Who we are, 2022. URL https://www.oecd.org/about/.

T. D. of Environment. Marine pollution, 2022. URL https://ec.europa.eu/eurostat.

Appendices

A Cleaned dataset and supporting material

Here is the link to access all the sources related to this report

B Figures

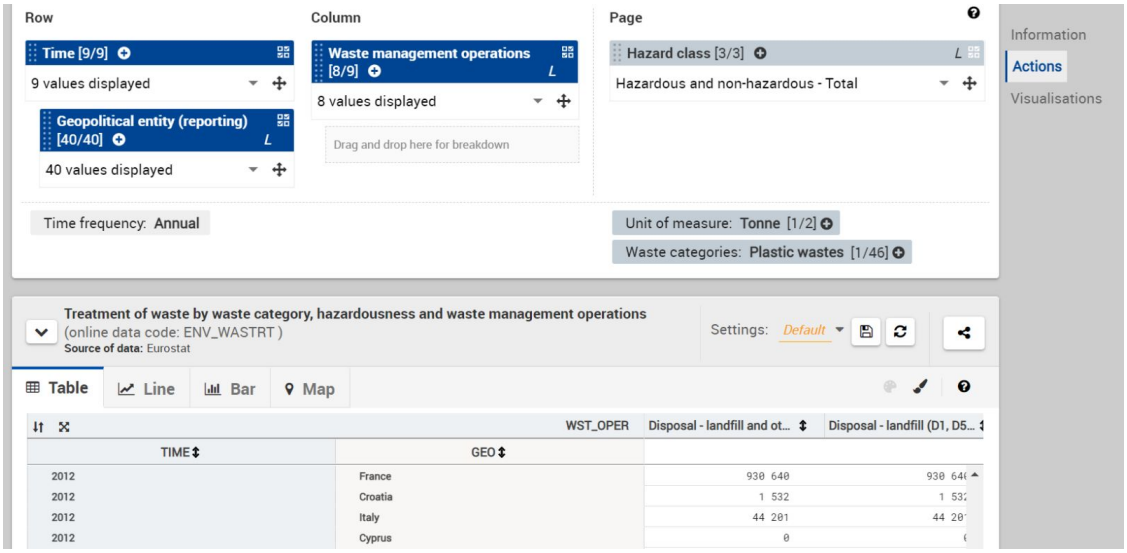

Figure 1: Eurostat plastic management operations

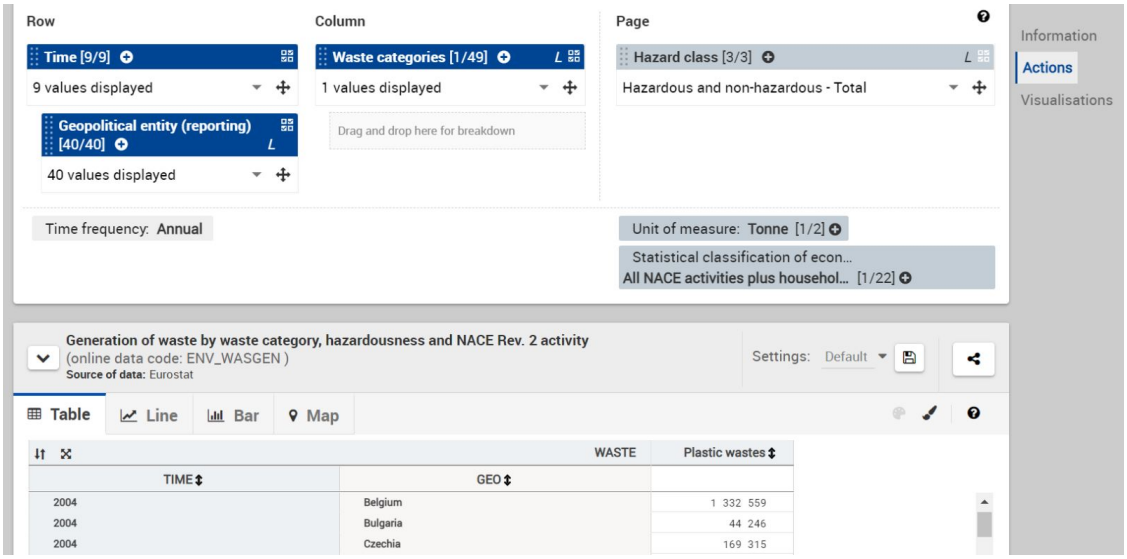

Figure 2: Eurostat plastic generation

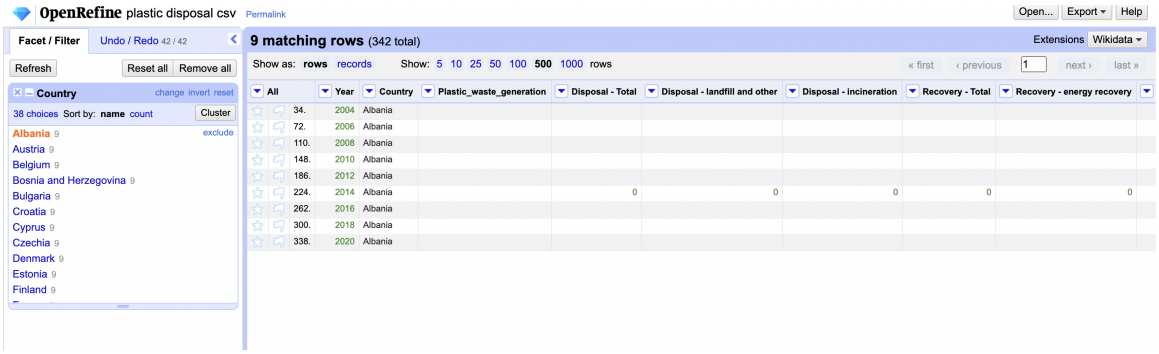

Figure 3: Missing values for 'Albania'

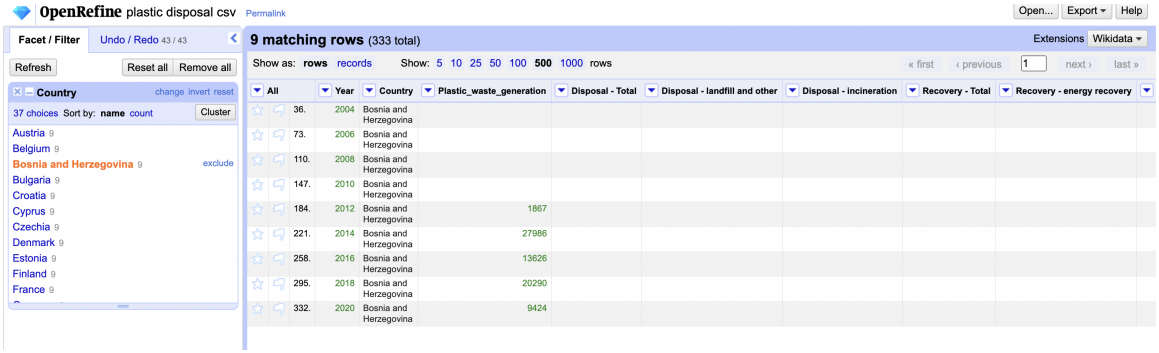

Figure 4: Missing values for 'Bosnia'

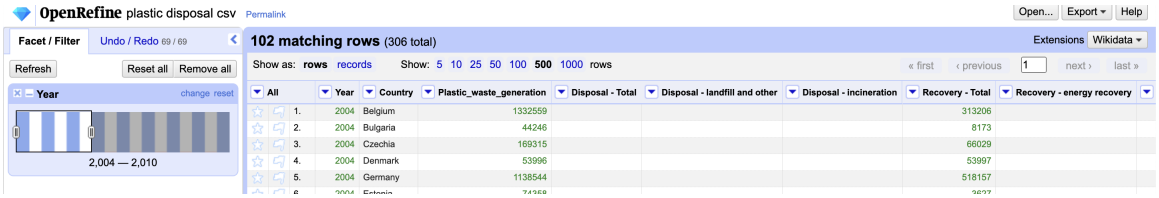

Figure 5: Missing values in 2004 for all countries

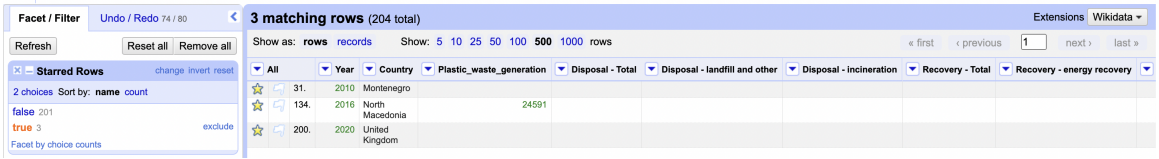

Figure 6: Delete selected rows with missing values