

MM916 Project 1 (2023): CO₂ emissions by European cities
Sample solutions, reformatted

Results

The GCoM dataset contains data on 7765 cities from 28 countries. Italy and Spain appear to be overrepresented (4037 and 2029 cities, respectively: see Table 1) while a number of countries seem to be underrepresented (for example, only 5 cities to represent Estonia). Populations of these cities range from 28 (Lobera de Onsella, Spain) to 1.2 million (London: note that the City of London is only a small, central portion of greater London, population 9 million). Populations follow a roughly lognormal distribution (Fig. 1) with a median of 4540.

Table 1. Countries with the highest and lowest number of cities included in the GCoM dataset.

Country (code)	Number of cities
<i>Italy (it)</i>	4037
<i>Spain (es)</i>	2029
<i>Belgium (be)</i>	457
...	
<i>Finland (fi)</i>	13
<i>Luxembourg (lu)</i>	9
<i>Estonia (ee)</i>	5

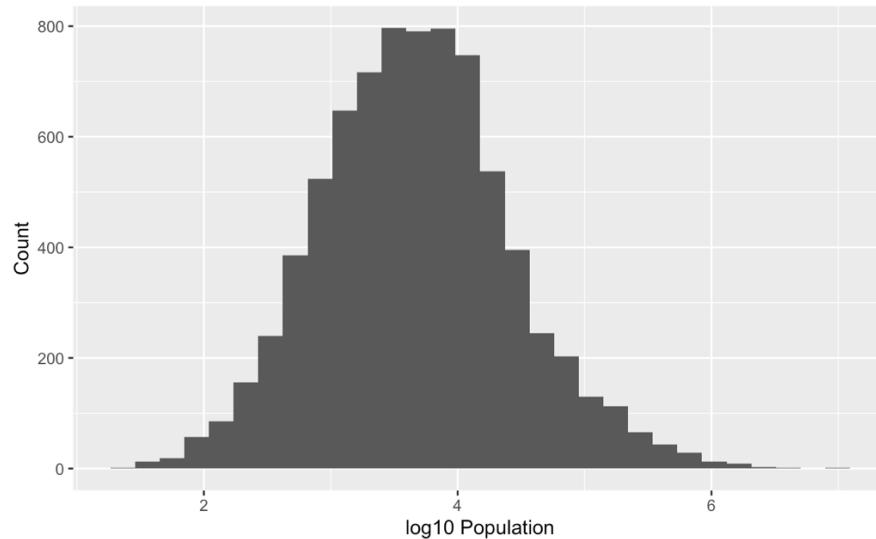


Figure 1. Histogram of city populations in the GCoM dataset.

Estonia, Luxembourg, and Germany appear to have the highest median per-capita emissions, while Latvia, Lithuania, and Romania have the lowest (Fig. 2). However, note that Estonia and Luxembourg, which appear from Figure 2 to be outliers, both are represented by a very small number of cities.

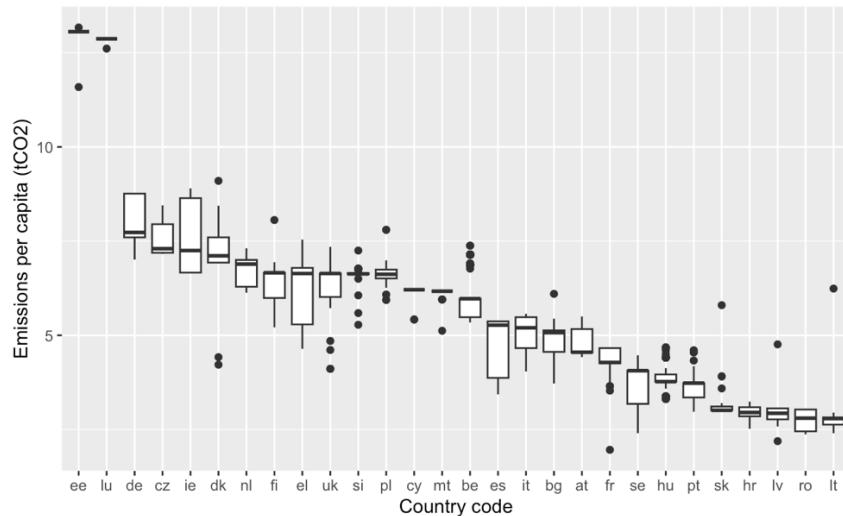


Figure 2. Distribution of city-level emissions per capita, grouped by country. The thick horizontal line gives the median for each country, and the white bars the interquartile range.



Figure 3. Total emissions by economic sector, summed across all cities in the dataset.

Across the full dataset, transportation and residential buildings are responsible for the largest shares of total emissions (Fig. 3), but the relative importance of the six sectors varies substantially by country (Fig. 4). In some countries, residential buildings are responsible for approximately 1/4 of emissions and in others more than half; the importance of transportation-related emissions is even more variable.

Based on visual inspection of a scatter plot relating emissions per capita to Heating Degree Days (Fig. 5), it seems very unlikely that there is a statistically significant correlation between these variables. Many of the cities in Scandinavia do have much higher HDD than the rest of the dataset (Fig. 5, blue points), but their emissions per capita are within the same range as most of the data. A second potentially explanatory scatter plot (Fig. 6) relates emissions per capita to GDP per capita. London (City of London, including the financial centre) is a clear outlier in terms of GDP. There might be a weak correlation between GDP and emissions per capita (Fig. 6): it is hard to tell without fitting a linear model carefully. Colour-coding by country shows a number of horizontal bands: sets of cities with very different GDP levels but very similar emissions levels.

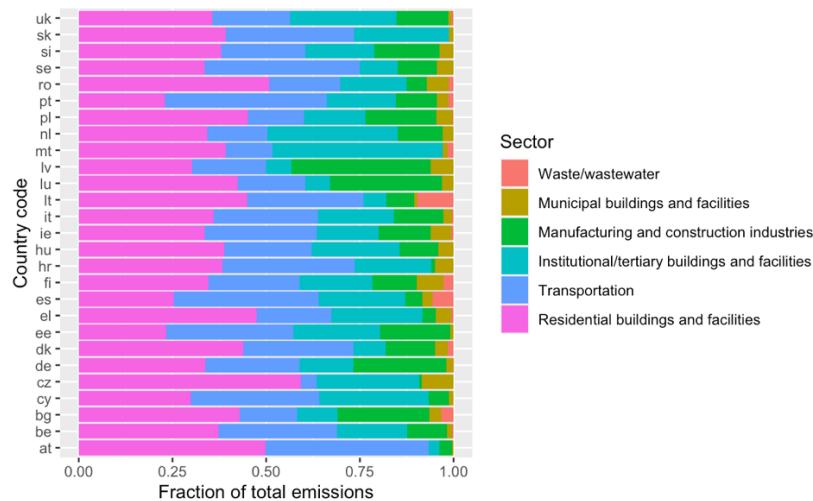


Figure 4. Fraction of total emissions associated with each economic sector, by country.

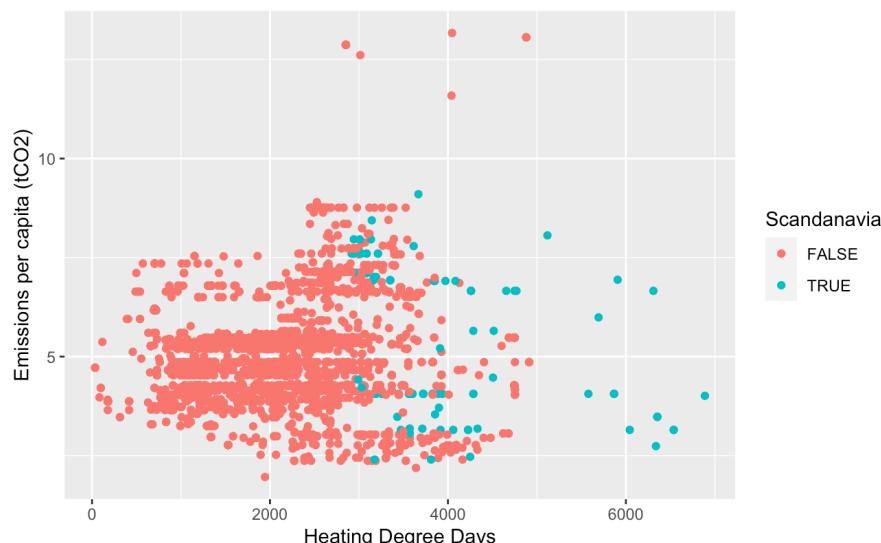


Figure 5. Emissions per capita vs. Heating Degree Days (HDD), for each city in the dataset. Scandinavian cities (Norway, Sweden, Finland, and Denmark) are highlighted in blue.

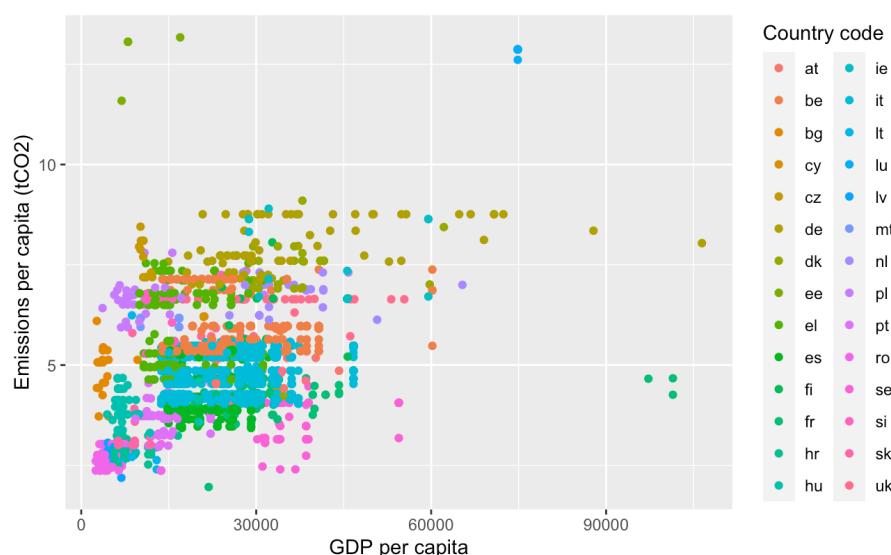


Figure 6. Emissions per capita vs. GDP per capita, for each city in the dataset. One outlier (City of London) has been removed.

In summary, this is a quite comprehensive dataset for evaluating city-level patterns in CO2 emissions: it covers many countries and a very wide range of populations. At the same time, it is quite imbalanced, with some countries represented by far more cities than others: one would have to be careful about the use of calculations that sum emissions (Figs. 3, 4) as opposed to using a median. Heating Degree Days (HDD) does not seem to be a promising hypothesis for explaining emissions variation. This is possibly because of the difference between energy use and emissions: perhaps Scandinavian countries are less reliant on fossil fuels for home heating and transportation? Data on energy use by country and city would be very helpful. There might be a correlation between GDP and emissions, but much of the variation in the last plot (the horizontal bands of points by colour) suggests that country-level factors, such as political decisions about climate action, or maybe national energy systems, may be important.