MM916 Project 1 (2023)

CO2 emissions by European cities

Eitika Sharma

CONTENT

LIST OF FIGURES AND TABLES

1. A dataset overview

The “GCoM_emissions.csv” dataset contains data on emissions and population for cities and countries.

The rows with missing values (NA) either in “Emission” or” Population” columns are filtered out, and a new dataset is created and saved as “Cleaned_DFbE”. Using this new dataset, we further created two new datasets namely, CountryDF which contains the number of cities associated with each unique country code and CityDF which contains the number of times each city name appears in the “DFbE” dataset. Next, I have grouped the Cities on the basis of the Country Code and arranged them in deceasing order.

	Country_Code	**City_Count**
1	it	4037
2	es	2029
3	be	457
4	el	163
5	hu	143
6	pt	140
7	fr	103
8	ro	95
9	de	74
10	hr	69
11	se	58
12	pl	42
13	dk	38
14	uk	38
15	si	30
16	sk	29
17	at	26
18	nl	26
19	bg	25
20	cy	24
21	mt	24
22	lv	21
23	cz	17
24	lt	16
25	ie	14
26	fi	13
27	lu	9
28	ee	5

Figure 1: Decreasing Order of Countries on Basis on City count

By doing the above step, it became easy for us to find the Country with maximum and minimum number of cities namely in Table 1.

Table 1 : Max and Min Cities in Country

	Country_Code	**City_Count**
Max number of Cities	it	4037
Min number of Cities	eee	5

There seem to be imbalances as some countries are overrepresented. Figure 1 shows a significant contrast in the number of cities across countries, Italy with 4037 cities while Estonia has just 5. This huge difference between the count of cities has triggered doubts about data accuracy and inconsistent data.

2. Range and distribution of city populations

Histogram below represents the range and distribution of city population.

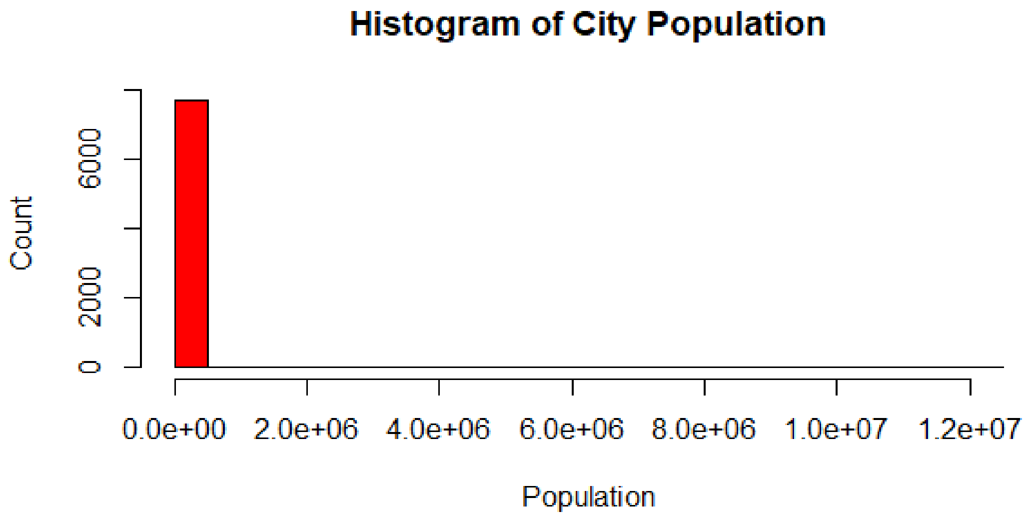

Figure 2: Histogram of City Population

The maximum city population is London in United Kingdom (UK) with 12051223 population.

	Country_Code	City	Population
1	uk	London	12051223

The minimum city population is Lobera de Onsella in Spain (ES) with 28 population.

	Country_Code	City	Population
1	es	Lobera de Onsella	28

The median city population is 4540.

Median_Population	4540

3. Emissions by country

The Boxplot of Emissions per Capita by Country is visualized in Figure 3.

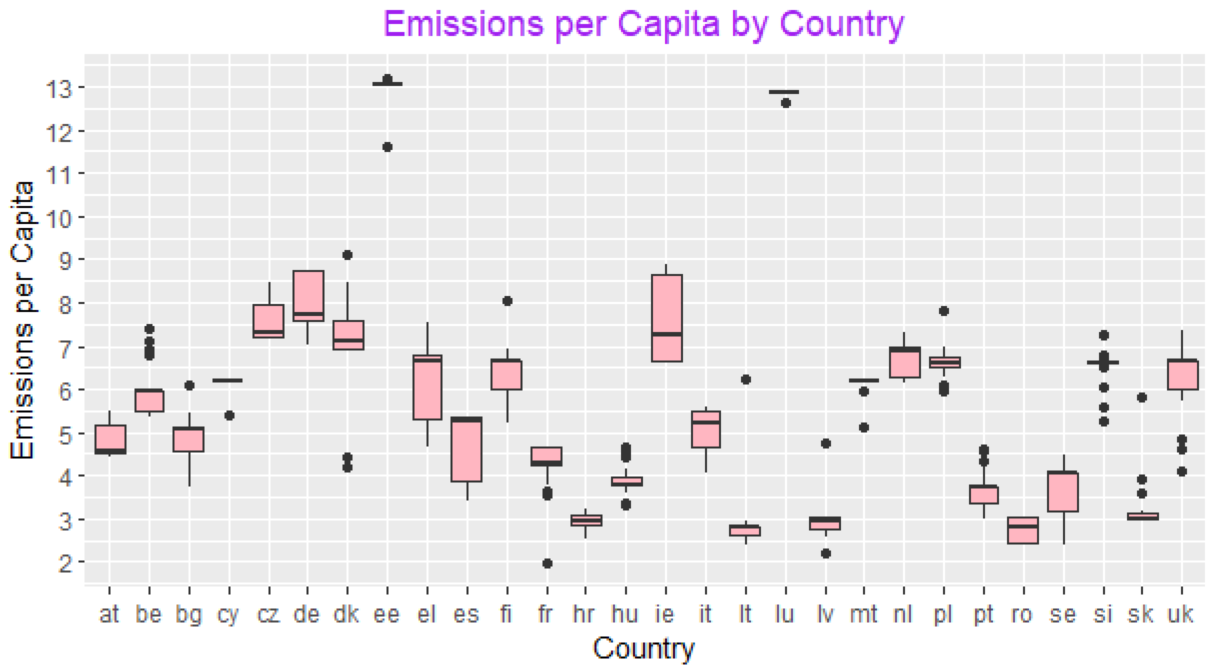

Figure 3: Emissions per Capita by Country Boxplot

After plotting the Boxplot for Emissions per Capita by Country, we have calculated the emission median for all the countries and then arranged them in ascending order. With this we were able to extract the following:

Top 3 Countries with highest median emissions per capita

	Country_Code	Median_Emissions
1	de	7.73
2	lu	12.87
3	ee	13.06

Bottom 3 Countries with lowest median emissions per capita

	Country_Code	Median_Emissions
1	lt	2.79
2	ro	2.80
3	lv	2.93

4. Emissions by sector

Using "GCoM_emissions_by_sector.csv" a dataframe “DFbS” was created and to visualise the total emissions for each of the 6 sectors a ggplot was used. This plot provides a clear overview of emissions distribution across these sectors.

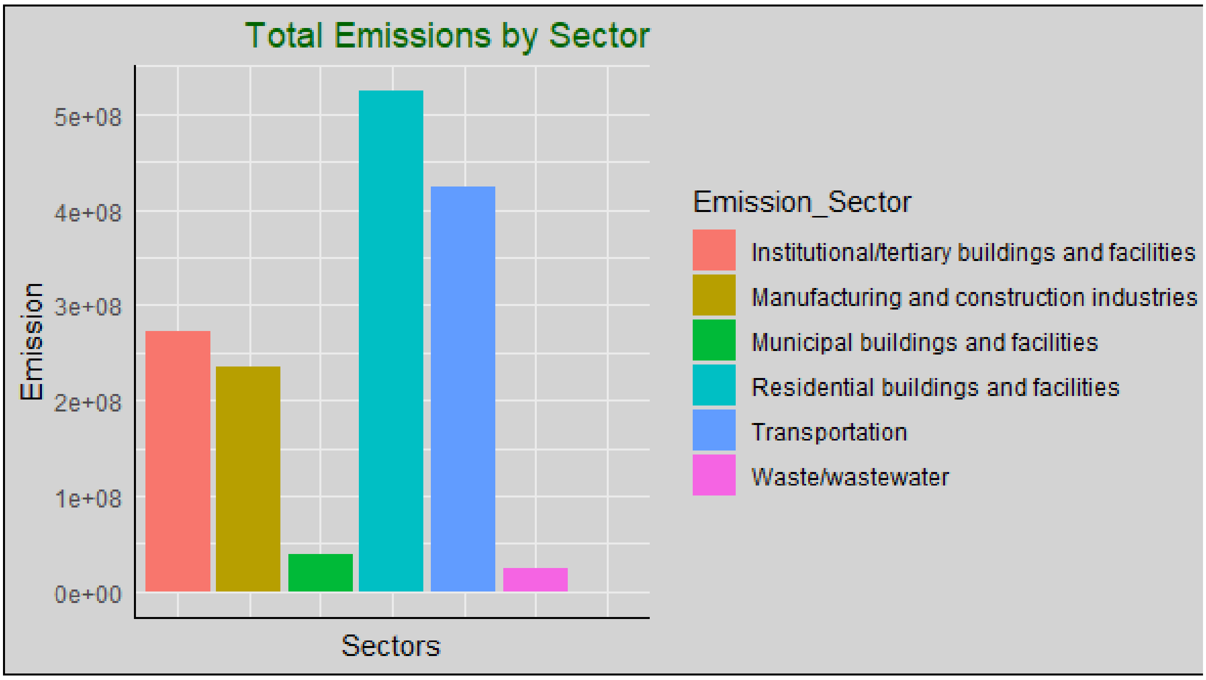

Figure 4: Total Emissions by Sector

From Figure 4, we can clearly note that the sector with the highest total emission is Residential buildings and facilities.

5. Emissions by sector and country

A stacked bar plot(Figure 5) presents a visual overview of how different sectors contribute to a country's total emissions. It starts by joining data to associate city names, countries, and sector-specific emissions. On the y-axis, each country is listed, while the x-axis displays the emissions.

This visualization offers a straightforward way to understand how emissions are distributed across sectors for various countries, making it easy to compare and gain insights into each sector's specific impact on a country's emissions.

Figure 5: Emission by Sector and Country

6. Connecting emissions to heating demand

The plot between Heating Degree Days (HDD) and emissions per capita is shown in Figure 6. The plot's x-axis represents Heating Degree-Days (HDD) and the y-axis represents emissions. It uses different colours to highlight Scandinavian countries (Sweden, Norway, Finland, Denmark) on the plot.

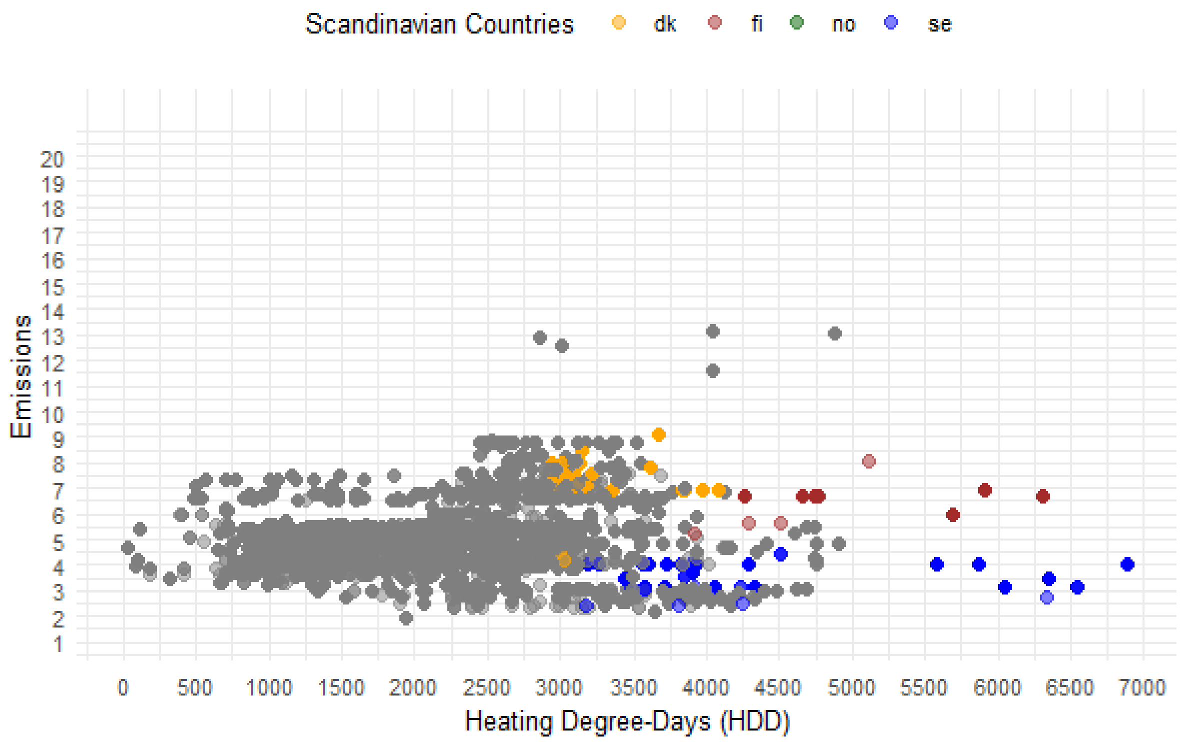

Figure 6: HDD vs Emissions per Capita

7. Connecting emissions to wealth

From Figure 7 we can depict the connection between a country's wealth, measured as GDP per capita, and its CO2 emissions per capita. Initially we removed one city having highest GDP per capita.

London, UK has the highest GDP per Capita.

After that a ggplot is plotted with GDP per capita on the horizontal axis, emissions on the vertical axis, and the points are coloured orange, sized at 2, and made slightly transparent for better clarity.

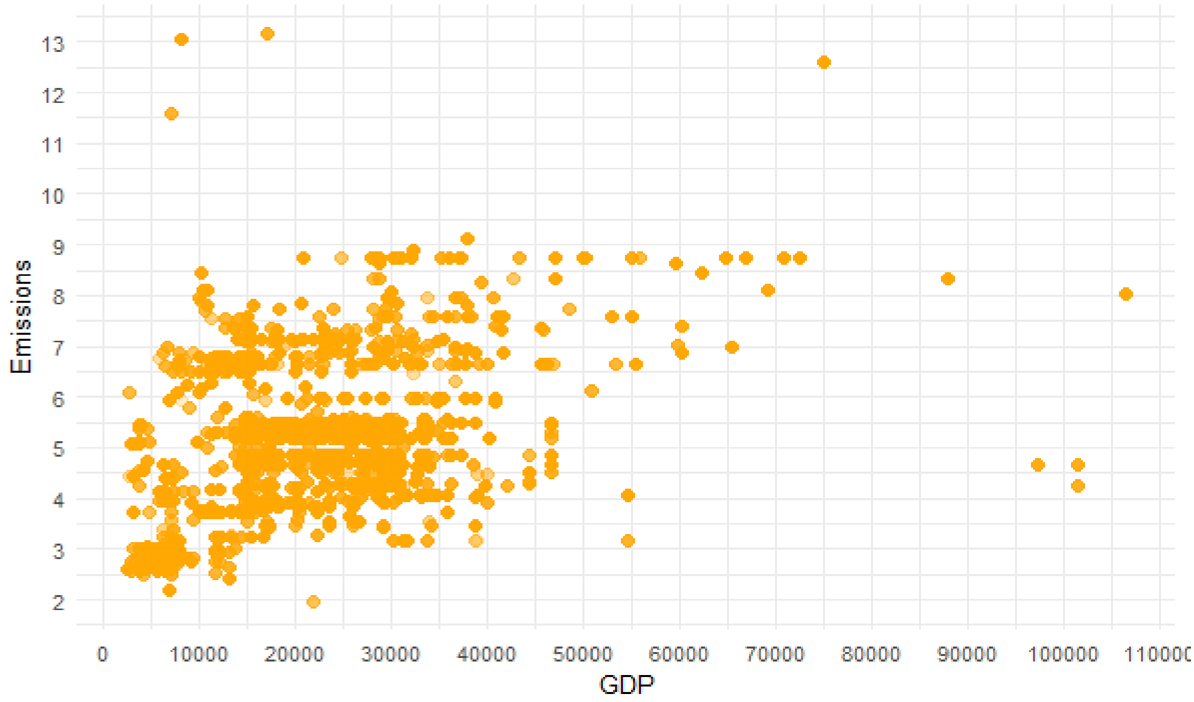

Figure 7: GDP vs Emissions

8. Summary and recommendations

The analysis revealed that there are significant differences in how much CO2 emissions each person produces in various cities and countries. It's clear that when a country's economy is stronger, indicated by its GDP, its people tend to create more CO2 emission. Also, the data highlights that the energy sector plays a big role in these emissions, showing that choices about energy sources matter.

To better understand these variations, we should look into other factors. For instance, we could investigate energy policies, social and economic conditions, government rules, geographic area and how people live. These additional factors might help us get a more complete picture of why some places have higher or lower emissions.

Appendix

```
# Loading the tidyverse package
require(tidyverse)

# Reading data from 'GCoM_emissions.csv'
DFbE = read_csv("C:/Users/LENOVO/Documents/R Assignment/GCoM_emissions.csv") %>%
  rename(Id = 'GCoM_ID',
         City = 'signatory name',
         Country_Code = 'country code',
         HDD = 'Heating Degree-Days (HDD)',
         GDP = 'GDP per capita at NUTS3 [Euro per inhabitant]',
         Emission = 'GHG emissions per capita in GCoM sectors_EDGAR [tCO2-eq/year]',
         Population = 'population in 2018') %>%
  select(Id, City, Country_Code, HDD, GDP, Emission, Population)

# Reading data from 'GCoM_emissions_by_sector.csv'
DFbS = read_csv("C:/Users/LENOVO/Documents/R Assignment/GCoM_emissions_by_sector.csv") %>%
  rename(Id = 'GCoM_ID',
         Emission_Sector = 'emission_inventory_sector',
         EmissionbS='emissions') %>%
  select(Id, Emission_Sector,EmissionbS)
```

1. A dataset overview Code

```
##Data Filter
# Filter rows with missing 'Population'
Pop_filter <- DFbE %>% filter(is.na(Population))

#Missing value count for Emission
print(sum(is.na(DFbE$Emission)))

#Missing value count for Population
print(sum(is.na(DFbE$Population)))

#NA values dropped for Emission and Population
Cleaned_DFbE <- DFbE %>% filter(!is.na(Emission) & !is.na(Population))


##Country Count
CountryDF <- Cleaned_DFbE$Country_Code
Country_Count <- as.data.frame(table(CountryDF))
print(Country_Count)


## City Count
CityDF <- DFbE$City
City_Count <- as.data.frame(table(CityDF))
print(City_Count)


##Ordering the Cleaned_DFbE
Ordering_Cleaned_DFbE<- Cleaned_DFbE %>% group_by(Country_Code) %>% summarize(City_Count = n()) %>% arrange(desc(City_Count))

#Country with maximum cities
Max_Cities <- Ordering_Cleaned_DFbE %>% slice(1)

#Country with minimum cities
Min_Cities <- Ordering_Cleaned_DFbE %>% arrange(City_Count) %>% slice(1)
```

2. Range and distribution of city populations Code

```
Cleaned_DFbE %>% select(Country_Code, City , Population) -> Country_Population

# Histogram
hist(Cleaned_DFbE$Population, breaks = 20, col = "red", border = "black",
     main = "Histogram of City Population", xlab = "Population", ylab = "Count")

##Maximum populated cities
print(Max_Popluated_City <- Country_Population %>% arrange(Population) %>% slice(n()))

##Minimum populated cities
print(Min_Popluated_City <- Country_Population %>% arrange(Population) %>% slice(1))

##median
print(Median_Popluation <- median(Cleaned_DFbE$Population))
```

3.Emissions by country Code

```
Cleaned_DFbE %>% arrange(Country_Code) %>% group_by(Country_Code,
                                          Emission) %>% ggplot(aes(x = Country_Code, y = Emission)) + geom_boxplot(fill = 'light pink') +
  labs(title = "Emissions per Capita by Country", x = "Country", y = "Emissions per Capita") +
  theme(plot.title = element_text(color = "purple", hjust = 0.5))+
  scale_y_continuous(breaks = seq(0, 20, by = 1))


# Median emissions per capita and arranging in ascending order
Median_Emissions <- Cleaned_DFbE %>%
  group_by(Country_Code) %>%
  summarize(Median_Emissions = median(Emission)) %>%
  arrange(Median_Emissions)

# Top 3 countries by median emissions per capita
print(Top3Countries <- tail(Median_Emissions, 3))

# Bottom 3 countries by median emissions per capita
print(Bottom3Countries <- head(Median_Emissions, 3))
```

4. Emissions by sector Code

```
ggplot(DFbS, aes(x = Emission_Sector, y = EmissionbS, fill = Emission_Sector))+
  geom_bar(stat = "identity") +
  labs(title = "Total Emissions by Sector",
       x = "Sectors",
       y = "Emission") +
  theme_minimal()+
  theme(axis.text.x = element_blank(),
        axis.line = element_line(size = 0.7),
        plot.title = element_text(color = "darkgreen", hjust = 1),
        plot.background = element_rect(fill = "lightgray"))
```

5. Emissions by sector and country Code

```
Joined_DF <- DFbE %>% left_join(DFbS, by = "Id")

ggplot(Joined_DF, aes(x = EmissionbS, y = reorder(Country_Code, EmissionbS), fill = Emission_Sector)) +
  geom_bar(position = position_fill(reverse = TRUE), stat = "identity", width = 0.7) +
  labs(title = "Emissions by Sector and Country",
       x = "Emission",
       y = "Country") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1")+
  theme(plot.title = element_text(color = "darkgreen"),
        axis.text.y = element_text(size = 6),
        axis.text.x = element_text(hjust = 1, size = 8))+
  scale_x_continuous(breaks = seq(0, 1, by = 0.1))
```

6. Connecting emissions to heating demand Code

```
scandinavian_colors <- c('se' = 'blue', 'fi' = 'brown', 'dk' = 'orange', 'no' = 'darkgreen')

ggplot(Joined_DF, aes(x = HDD, y = Emission, color = Country_Code)) +
  geom_point(size = 2, alpha = 0.5) +
  labs(x = 'Heating Degree-Days (HDD)', y = 'Emissions') +
  scale_color_manual(values = scandinavian_colors) +
  guides(color = guide_legend(title = 'Scandinavian Countries')) +
  theme_minimal() +
  theme(
    legend.position = 'top',
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 8),
    axis.text = element_text(size = 8),
    axis.title = element_text(size = 10)
  )+
  scale_x_continuous(breaks = seq(0, 7000, by = 500))+
  scale_y_continuous(breaks = seq(0, 20, by = 1))
```

7. Connecting emissions to wealth Code

```
DFbE7 <- na.omit(Joined_DF)

CityMaxGDP <- DFbE7 %>% filter(GDP == max(GDP))
NewDF<- DFbE7 %>% filter(City != CityMaxGDP$City)
print(CityMaxGDP)
ggplot(data = NewDF, aes(x = GDP, y = Emission)) +
  geom_point(color = "orange", size = 2, alpha = 0.2) +
  labs(
    x = "GDP",
    y = "Emissions",
  ) +
  theme_minimal() +
  theme(
    axis.text = element_text(size = 8),
    axis.title = element_text(size = 10)
  ) +
  scale_x_continuous(breaks = seq(0, 140000, by = 10000)) +
  scale_y_continuous(breaks = seq(0, 15, by = 1))
```