Data Analysis

April 22, 2021

# Install and load libraries

We need to first install all the required packages for the data analysis process. If you have never used the packages below, it is more likely that you have not installed them on your machine either. Please make sure you install each of the packages below using the following command:

install.packages("janitor")  
install.packages("tidyverse")  
install.packages("knitr")  
install.packages("DataExplorer")  
install.packages("SmartEDA")

Then you need to load the following packages:

library(janitor)  
library(tidyverse)  
library(knitr)  
library(here)  
library(DataExplorer)  
library(SmartEDA)  
library(data.table)  
library(summarytools)

# Import clean dataset

There are two ways you can import the clean dataset into this file:

1. By running the R script from clean folder

source(here("scripts/cleaning", "unfccc-emissions-clean.R"))

This will also produce some of the temporary files.

1. By importing the clean dataset

unfccc\_emissions <- read.csv(here("scripts/cleaning",   
 "unfccc-emissions-clean.csv"))

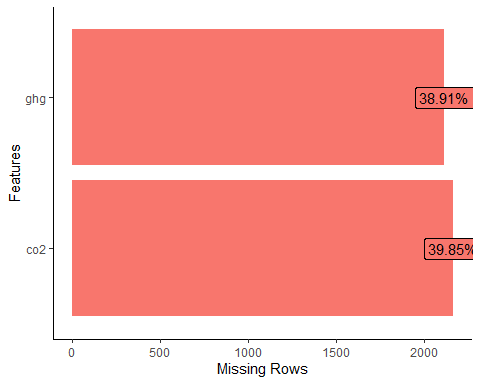
# Describe dataset

So if you are using R, then you are lucky to have at your finger tips a plethora of packages that help you visualize the structure of the dataset, and what it would mean for your data analysis.

Things that you can accomplish during this stage:

* Describe the level of missing data

DataExplorer::plot\_missing(unfccc\_emissions, missing\_only = TRUE,   
 ggtheme = ggplot2::theme\_classic(), theme\_config = ggplot2::theme(legend.position = "none"))



knitr::kable(dfSummary(unfccc\_emissions))

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Variable | Stats / Values | Freqs (% of Valid) | Graph | text.graph | Valid | Missing |
| 1 | X [integer] | Mean (sd) : 2714.5 (1567.1) min < med < max: 1 < 2714.5 < 5428 IQR (CV) : 2713.5 (0.6) | 5428 distinct values (Integer sequence) |  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : | 5428 (100.0%) | 0 (0.0%) |
| 2 | party [character] | 1. United Kingdom of Great B 2. Dominican Republic 3. Australia 4. Austria 5. Belarus 6. Belgium 7. Bulgaria 8. Canada 9. Croatia 10. Cyprus [ 80 others ] | 120 ( 2.2%)  116 ( 2.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%)   60 ( 1.1%) \4712 (86.8%) |  | \ |  |  |

                  IIIIIIIIIIIIIIIII |5428  
(100.0%) |0  
(0.0%) | | 3|type  
[character] |1. Total GHG emissions with  
2. Total GHG emissions witho |\2714 (50.0%)  
\2714 (50.0%) | |IIIIIIIIII   IIIIIIIIII |5428  
(100.0%) |0  
(0.0%) | | 4|year  
[character] |1. 1990  
2. 1991  
3. 1992  
4. 1993  
5. 1994  
6. 1995  
7. 1996  
8. 1997  
9. 1998  
10. 1999  
[ 20 others ] | 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
 184 ( 3.4%)  
\3588 (66.1%) | |                    IIIIIIIIIIIII |5428  
(100.0%) |0  
(0.0%) | | 5|ghg  
[numeric] |Mean (sd) : 547277.9 (1351516)  
min < med < max:  
-1740723 < 72478 < 12300200  
IQR (CV) : 409384.9 (2.5) |3143 distinct values | |  
  :  
  :  
  :  
  :  
  : . |3316  
(61.1%) |2112  
(38.9%) | | 6|group  
[character] |1. Annex I  
2. Non-Annex I |\2760 (50.8%)  
\2668 (49.2%) | |IIIIIIIIII   IIIIIIIII |5428  
(100.0%) |0  
(0.0%) | | 7|co2  
[numeric] |Mean (sd) : 436431.1 (1096493)  
min < med < max:  
-1766401 < 56411.4 < 10274851  
IQR (CV) : 315100.4 (2.5) |3094 distinct values | |  
  :  
  :  
  :  
  :  
  : . |3265  
(60.2%) |2163  
(39.8%) | | 8|region  
[character] |1. Africa  
2. Americas  
3. Asia  
4. Europe  
5. Oceania | 986 (18.2%)  
\1048 (19.3%)  
 820 (15.1%)  
\2396 (44.1%)  
 178 ( 3.3%) | |III   III   III   IIIIIIII   |5428  
(100.0%) |0  
(0.0%) |