Data Importing and Wrangling

April 05, 2021

# Install and load libraries

library(openxlsx)  
library(tidyverse)  
library(janitor)  
library(knitr)

* openxlsx:
* tidyverse:
* janitor:
* knitr:

# Data Import and Wrangling

## Import data

### Annex I parties

Annex\_I\_emissions <- openxlsx::read.xlsx("Annual-Net-emissions-removals-Annex-I-raw.xlsx",   
 startRow = 5, fillMergedCells = TRUE, rows = 5:96) %>%   
 janitor::clean\_names()

### Non-Annex I parties

Non\_Annex\_I\_emissions <- openxlsx::read.xlsx("Annual-Net-emissions-removals-Non-Annex-I-raw.xlsx",   
 startRow = 5, fillMergedCells = TRUE, rows = 5:96) %>%   
 janitor::clean\_names()

## Clean variables

### Annex I parties

Annex\_I\_emissions <- Annex\_I\_emissions[-c(1), ]  
  
colnames(Annex\_I\_emissions)[1] <- "party"  
  
Annex\_I\_emissions <- Annex\_I\_emissions %>% dplyr::rename\_all(funs(str\_replace\_all(.,   
 "gh\_gs", "ghg")))

### Non-Annex I parties

Non\_Annex\_I\_emissions <- Non\_Annex\_I\_emissions[-c(1),   
 ]  
  
colnames(Non\_Annex\_I\_emissions)[1] <- "party"  
  
Non\_Annex\_I\_emissions <- Non\_Annex\_I\_emissions %>%   
 dplyr::rename\_all(funs(str\_replace\_all(., "gh\_gs",   
 "ghg")))

## Disaggreagate data by gas type

### Annex I parties

Annex\_I\_emissions\_ch4 <- Annex\_I\_emissions %>% dplyr::select("party",   
 "gas", contains("ch4")) %>% dplyr::rename(base\_year = "ch4")  
  
colnames(Annex\_I\_emissions\_ch4)[4:32] <- 1990:2018  
  
Annex\_I\_emissions\_co2 <- Annex\_I\_emissions %>% dplyr::select("party",   
 "gas", contains("co2")) %>% dplyr::rename(base\_year = "co2")  
  
colnames(Annex\_I\_emissions\_co2)[4:32] <- 1990:2018  
  
Annex\_I\_emissions\_ghg <- Annex\_I\_emissions %>% dplyr::select("party",   
 "gas", contains("ghg")) %>% dplyr::rename(base\_year = "aggregate\_ghg")  
  
colnames(Annex\_I\_emissions\_ghg)[4:32] <- 1990:2018

### Non-Annex I parties

Non\_Annex\_I\_emissions\_ch4 <- Non\_Annex\_I\_emissions %>%   
 dplyr::select("party", "gas", contains("ch4"))  
  
colnames(Non\_Annex\_I\_emissions\_ch4)[3:31] <- 1990:2018  
  
Non\_Annex\_I\_emissions\_co2 <- Non\_Annex\_I\_emissions %>%   
 dplyr::select("party", "gas", contains("co2"))  
  
colnames(Non\_Annex\_I\_emissions\_co2)[3:31] <- 1990:2018  
  
Non\_Annex\_I\_emissions\_ghg <- Non\_Annex\_I\_emissions %>%   
 dplyr::select("party", "gas", contains("ghg"))  
  
colnames(Non\_Annex\_I\_emissions\_ghg)[3:31] <- 1990:2018

## Convert datasets to long

### Annex I parties

Annex\_I\_emissions\_ch4 <- Annex\_I\_emissions\_ch4 %>%   
 tidyr::gather(year, ch4, base\_year:`2018`)  
  
Annex\_I\_emissions\_co2 <- Annex\_I\_emissions\_co2 %>%   
 tidyr::gather(year, co2, base\_year:`2018`)  
  
Annex\_I\_emissions\_ghg <- Annex\_I\_emissions\_ghg %>%   
 tidyr::gather(year, ghg, base\_year:`2018`)

### Non-Annex I parties

Non\_Annex\_I\_emissions\_ch4 <- Non\_Annex\_I\_emissions\_ch4 %>%   
 tidyr::gather(year, ch4, `1990`:`2018`)  
  
Non\_Annex\_I\_emissions\_co2 <- Non\_Annex\_I\_emissions\_co2 %>%   
 tidyr::gather(year, co2, `1990`:`2018`)  
  
Non\_Annex\_I\_emissions\_ghg <- Non\_Annex\_I\_emissions\_ghg %>%   
 tidyr::gather(year, ghg, `1990`:`2018`)

## Merge datasets back

### Annex I parties

Annex\_I\_emissions <- Annex\_I\_emissions\_ghg %>% dplyr::full\_join(Annex\_I\_emissions\_co2,   
 by = c("party", "gas", "year")) %>% dplyr::full\_join(Annex\_I\_emissions\_ch4,   
 by = c("party", "gas", "year"))  
  
rm(Annex\_I\_emissions\_ch4, Annex\_I\_emissions\_co2, Annex\_I\_emissions\_ghg)

### Non-Annex I parties

Non\_Annex\_I\_emissions <- Non\_Annex\_I\_emissions\_ghg %>%   
 dplyr::full\_join(Non\_Annex\_I\_emissions\_co2, by = c("party",   
 "gas", "year")) %>% dplyr::full\_join(Non\_Annex\_I\_emissions\_ch4,   
 by = c("party", "gas", "year"))  
  
rm(Non\_Annex\_I\_emissions\_ch4, Non\_Annex\_I\_emissions\_co2,   
 Non\_Annex\_I\_emissions\_ghg)

## Merge Annex I and Non-Annex I

UNFCCC\_Emissions <- dplyr::bind\_rows(Annex\_I\_emissions,   
 Non\_Annex\_I\_emissions) %>% dplyr::rename(type = "gas") %>%   
 dplyr::mutate(type = stringr::str\_replace(type,   
 "Total GHG emissions excluding LULUCF/LUCF",   
 "Total GHG emissions without LULUCF")) %>%   
 dplyr::mutate(type = stringr::str\_replace(type,   
 "Total GHG emissions including LULUCF/LUCF",   
 "Total GHG emissions with LULUCF")) %>% readr::type\_convert() %>%   
 dplyr::mutate\_if(is.character, as.factor)  
  
rm(Annex\_I\_emissions, Non\_Annex\_I\_emissions)

## Export as a CSV file

write.csv(UNFCCC\_Emissions, "UNFCCC-Emissions-clean.csv")

## Export as an R script for future use

Only run this chunk manually once within the .Rmd file. It produces an error when knitting it as a whole because of chunk label duplicates. As of April 05, 2021, there hasn’t been a viable solution to run the code below when as part of the knitting process.

knitr::purl("UNFCCC-Emissions.Rmd", "UNFCCC-Emissions-clean.R")