

Data gathering and transformation

Contents

| | | |
|----------|---|----------|
| 1 | Data Pre-Processing | 2 |
| 1.1 | Gathering the data and merging | 2 |
| 1.2 | Greenhouse Gas emission | 2 |
| 1.2.1 | Greenhouse gas emissions by source sector (source: EEA) (sdg_13_10) | 2 |
| 1.2.2 | Greenhouse gas emissions intensity of energy consumption (sdg_13_20) | 2 |
| 1.2.3 | Average CO2 emission per km from new passengers cars (sdg_12_30) | 3 |
| 1.2.4 | Greenhouse gas emission from agriculture (tai08) | 4 |
| 1.2.5 | Merge all greenhouse gas emissions data into one dataframe | 5 |
| 1.3 | Drivers - Transport | 5 |
| 1.3.1 | Air transport of passengers by country (yearly data) (ttr00012) | 5 |
| 1.3.2 | Air transport of goods by country (yearly data) (ttr00011) | 6 |
| 1.3.3 | Sea transport of goods (ttr00009) | 6 |
| 1.3.4 | Goods transport by road (ttr00005) | 7 |
| 1.3.5 | Volume of freight transport relative to GDP (tran_hv_frtra) | 8 |
| 1.3.6 | Modal split of passenger transport (tran_hv_psmo) | 9 |
| 1.3.7 | Volume of passenger transport relative to GDP, index = 2010 (tran_hv_pstra) | 10 |
| 1.3.8 | Merge all transportation data into one dataframe | 10 |
| 1.4 | Drivers - Energy | 11 |
| 1.4.1 | Primary energy consumption (sdg_07_10) | 11 |
| 1.4.2 | Final energy consumption in households (t2020_rk200) | 12 |
| 1.4.3 | Final energy consumption by product (ten00123) | 13 |
| 1.4.4 | Merge all transportation data into one dataframe | 14 |
| 1.5 | Drivers - Waste | 14 |
| 1.5.1 | Municipal waste by waste management operations (env_wasmun) | 14 |
| 1.6 | Drivers Industrial process | 16 |
| 1.6.1 | Production in industry - annual data (sts_inpr_a) | 16 |
| 1.6.2 | Production in construction - annual data (sts_copr_a) | 54 |
| 1.6.3 | Merge all industrial process data into one dataset | 55 |
| 1.7 | Merge all possible drivers | 55 |
| 1.8 | Save data | 56 |

1 Data Pre-Processing

1.1 Gathering the data and merging

1.2 Greenhouse Gas emission

1.2.1 Greenhouse gas emissions by source sector (source: EEA) (sdg_13_10)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df<- "sdg_13_10_"
# Read data set
library(data.table)
df = fread("data/Eurostat/Greenhouse/sdg_13_10.tsv.gz", header = TRUE, sep2 = c(",","\t","|",";",";",":",".",
"  

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names
  

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]
  

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))
  

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sapply(df, function(x) gsub("b", "", x)))
df <- data.frame(sapply(df, function(x) gsub("ep", "", x)))
  

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))
  

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names
  

df_sdg_13_10 <- df
```

1.2.2 Greenhouse gas emissions intensity of energy consumption (sdg_13_20)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df<- "sdg_13_20_"
# Read data set
```

```

library(data.table)
df = fread("data/Eurostat/Greenhouse/sdg_13_20.tsv", header = TRUE, sep2 = c("","\t","|",";",";",":",""))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_sdg_13_20 <- df

```

1.2.3 Average CO2 emission per km from new passengers cars (sdg_12_30)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU27_2020")
name_df <- "sdg_12_30_"
# Read data set
library(data.table)
df = fread("data/Eurostat/Greenhouse/sdg_12_30.tsv", header = TRUE, sep2 = c("","\t","|",";",";",":",""))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# EXTRA FOR THIS DATASET

```

```

# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sapply(df, function(x) gsub("e", "", x)))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- "sdg_12_30_CO2KM_EU28"
rownames(df) <- year_names

df_sdg_12_30 <- df

```

1.2.4 Greenhouse gas emission from agriculture (tai08)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "tai08_"
# Read data set
library(data.table)
df = fread("data/Eurostat/Greenhouse/tai08.tsv", header = TRUE, sep2 = c("","\t","|",";",";",":",""))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_tai08 <- df

```

1.2.5 Merge all greenhouse gas emissions data into one dataframe

```
df_gas <- merge(df_sdg_13_10, df_sdg_13_20, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_gas[, 1]
df_gas[, 1] <- NULL
rownames(df_gas) <- year_names

df_gas <- merge(df_gas, df_sdg_12_30, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_gas[, 1]
df_gas[, 1] <- NULL
rownames(df_gas) <- year_names

df_gas <- merge(df_gas, df_tai08, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_gas[, 1]
df_gas[, 1] <- NULL
rownames(df_gas) <- year_names
```

1.3 Drivers - Transport

1.3.1 Air transport of passengers by country (yearly data) (ttr00012)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df<- "ttr00012_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/ttr00012.tsv.gz", header = TRUE, sep2 = c(",","\t","|",";",".",":",""))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_ttr00012 <- df
```

1.3.2 Air transport of goods by country (yearly data) (ttr00011)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "ttr00011_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/ttr00011.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":", ""))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_ttr00011 <- df
```

1.3.3 Sea transport of goods (ttr00009)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
```

```

name_df<- "ttr00009_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/ttr00009.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":", ""))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_ttr00009 <- df

```

1.3.4 Goods transport by road (ttr00005)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df<- "ttr00005_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/ttr00005.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":", ""))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

```

```

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_ttr00005 <- df

```

1.3.5 Volume of freight transport relative to GDP (tran_hv_frtra)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "tran_hv_frtra_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/tran_hv_frtra.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":"))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)

# Transpose column and rows, we want year in the row and variables in the columns
df <- data.frame(t(df))

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sapply(df, function(x) gsub("e", "", x)))
df <- data.frame(sapply(df, function(x) gsub("ep", "", x)))
df <- data.frame(sapply(df, function(x) gsub("p", "", x)))

# Transform everything to numeric

```



```
df <- data.frame(sapply(df, as.numeric))

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_tran_hv_frtra <- df
```

1.3.6 Modal split of passenger transport (tran_hv_psmod)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "tran_hv_psmod_"

# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/tran_hv_psmod.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":"))

# Assign first column to row names and delete it
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)

# Transpose column and rows, we want year in the row and variables in the columns
df <- data.frame(t(df))

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sapply(df, function(x) gsub("e", "", x)))
df <- data.frame(sapply(df, function(x) gsub("ep", "", x)))
df <- data.frame(sapply(df, function(x) gsub("p", "", x)))

# Transform everything to numeric
df <- data.frame(sapply(df, as.numeric))
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_tran_hv_psmod <- df

```

1.3.7 Volume of passenger transport relative to GDP, index = 2010 (tran_hv_pstra)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df<- "tran_hv_pstra_"
# Read data set
library(data.table)
df = fread("data/Eurostat/Transport/tran_hv_pstra.tsv.gz", header = TRUE, sep2 = c(",","\t","|",";",":"))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# FOR THE FUNCTION df <- data.frame(t(df))
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- t(df)

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sub("e", "", df, fixed = TRUE))
df <- data.frame(sapply(df, as.numeric))

```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_tran_hv_pstra <- df

```

1.3.8 Merge all transportation data into one dataframe

```
df_tran <- merge(df_tran_hv_frtra, df_tran_hv_psmod, by="row.names", all=TRUE)
```

```

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

df_tran <- merge(df_tran, df_tran_hv_pstra, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

df_tran <- merge(df_tran, df_ttr00005, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

df_tran <- merge(df_tran, df_ttr00009, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

df_tran <- merge(df_tran, df_ttr00011, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

df_tran <- merge(df_tran, df_ttr00012, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_tran[, 1]
df_tran[, 1] <- NULL
rownames(df_tran) <- year_names

```

1.4 Drivers - Energy

1.4.1 Primary energy consumption (sdg_07_10)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "sdg_07_10_"
# Read data set
library(data.table)
df = fread("Data/Eurostat/Energy/sdg_07_10.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":", ""))

```

```

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- data.frame(t(df))

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sapply(df, function(x) gsub("e", "", x)))
df <- data.frame(sapply(df, function(x) gsub("b", "", x)))
df <- data.frame(sapply(df, function(x) gsub("p", "", x)))

#Change from character to numeric
df <- data.frame(sapply(df, as.numeric))

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_sdg_07_10 <- df

```

1.4.2 Final energy consumption in households (t2020_rk200)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "t2020_rk200_"
# Read data set
library(data.table)
df = fread("Data/Eurostat/Energy/t2020_rk200.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|", ";", ":", ""))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row

```

```

year_names <- colnames(df)
df <- data.frame(t(df))

#Change from character to numeric
df <- data.frame(sapply(df, as.numeric))

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_t2020_rk200 <- df

```

1.4.3 Final energy consumption by product (ten00123)

```

# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df <- "ten00123_"
# Read data set
library(data.table)
df = fread("Data/Eurostat/Energy/ten00123.tsv.gz", header = TRUE, sep2 = c(",","\t","|",";",":",""))

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% to_keep
df <- df[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# FOR THE FUNCTION df <- data.frame(t(df))
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)
df <- t(df)

# EXTRA FOR THIS DATASET
# Eliminate character e, which stand for estimated. We trust Eurostat!
df <- data.frame(sub("e", "", df, fixed = TRUE))
df <- data.frame(sapply(df, as.numeric))

```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

df_ten00123 <- df
```

1.4.4 Merge all transportation data into one dataframe

```
df_energy <- merge(df_sdg_07_10, df_t2020_rk200, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_energy[, 1]
df_energy[, 1] <- NULL
rownames(df_energy) <- year_names

df_energy <- merge(df_energy, df_ten00123, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_energy[, 1]
df_energy[, 1] <- NULL
rownames(df_energy) <- year_names
```

1.5 Drivers - Waste

1.5.1 Municipal waste by waste management operations (env_wasmun)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28", "KG_HAB")
name_df <- "env_wasmun_"
# Read data set
library(data.table)
df = fread("data/Eurostat/Waste/env_wasmun.tsv.gz", header = TRUE, sep2 = c(",","\t","|",";",";",":",""))
```

```

# Assign as rows name the content of the first column
var_names <- t(df[, 1])
df[, 1] <- NULL
rownames(df) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df) %like% "EU28"
df <- df[mask, ]

# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df)

# Transpose column and rows, we want year in the row and variables in the columns
df <- data.frame(t(df))

# EXTRA FOR THIS DATASET
# Eliminate character s
df <- data.frame(sapply(df, function(x) gsub("s", "", x)))

#Change from character to numeric
df <- data.frame(sapply(df, as.numeric))

```

```

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df) <- paste(name_df, var_names)
rownames(df) <- year_names

# Only keep data that has kg per habitants as unit of measure (remove tonne and thousand of tonne --> s
mask <- colnames(df) %like% "KG"
df <- df[, mask]

df_waste <- df
```

1.6 Drivers Industrial process

1.6.1 Production in industry - annual data (sts_inpr_a)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df_sts_inpr_a <- "sts_inpr_a_"
# Read data set
library(data.table)
df_sts_inpr_a = fread("data/Eurostat/Industrial/sts_inpr_a.tsv.gz", header = TRUE, sep2 = c(",", "\t", "|")

# Assign as rows name the content of the first column
var_names <- t(df_sts_inpr_a[, 1])
df_sts_inpr_a[, 1] <- NULL
rownames(df_sts_inpr_a) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df_sts_inpr_a) %like% to_keep
df_sts_inpr_a <- df_sts_inpr_a[mask, ]
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df_sts_inpr_a) <- paste(name_df_sts_inpr_a, var_names)
rownames(df_sts_inpr_a) <- year_names

# Only keep interesting production data - intermediate goods, capital goods, consumer goods, manufactur
# Only keep index data (I10) and calendar adjusted data (CA)
to_keep <- c("sts_inpr_a_PROD,MIG_ING,CA,I10,EU28", "sts_inpr_a_PROD,MIG_CAG,CA,I10,EU28",
             "sts_inpr_a_PROD,MIG_COG,CA,I10,EU28", "sts_inpr_a_PROD,C,CA,I10,EU28",
             "sts_inpr_a_PROD,C10,CA,I10,EU28", "sts_inpr_a_PROD,B,CA,I10,EU28", "sts_inpr_a_PROD,D
df_sts_inpr_a <- subset(df_sts_inpr_a, select = to_keep)
```

1.6.2 Production in construction - annual data (sts_copr_a)

```
# Keep data regarding all 28 countries in Europe
to_keep <- c("EU28")
name_df_sts_copr_a <- "sts_copr_a_"
# Read data set
library(data.table)
df_sts_copr_a = fread("data/Eurostat/Industrial/sts_copr_a.tsv.gz", header = TRUE, sep2 = c(",","\t","|")

# Assign as rows name the content of the first column
var_names <- t(df_sts_copr_a[, 1])
df_sts_copr_a[, 1] <- NULL
rownames(df_sts_copr_a) <- var_names

# Only keep data of all the 28 European country, use data aggregation calculated from Eurostat
mask <- rownames(df_sts_copr_a) %like% to_keep
df_sts_copr_a <- df_sts_copr_a[mask, ]

# Transpose column and rows, we want year in the row and variables in the columns
# Keep name of the columns, which are the year that will be set as row
year_names <- colnames(df_sts_copr_a)
df_sts_copr_a <- data.frame(t(df_sts_copr_a))
```

```

# Eliminate character s, which stand for estimated. We trust Eurostat!
df_sts_copr_a <- data.frame(sapply(df_sts_copr_a, function(X) gsub("s", "", X)))

#Change from character to numeric
df_sts_copr_a <- data.frame(sapply(df_sts_copr_a, as.numeric))

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# Set correct name to the variable (column) in consideration and add the name of the dataset used
var_names <- var_names[, mask]
colnames(df_sts_copr_a) <- paste(name_df_sts_copr_a, var_names)
rownames(df_sts_copr_a) <- year_names

# Only keep data expressed as index of 2010 (I10)
mask <- colnames(df_sts_copr_a) %like% "I10"
df_sts_copr_a <- df_sts_copr_a[, mask]

```

1.6.3 Merge all industrial process data into one dataset

```

df_industrial <- merge(df_sts_copr_a, df_sts_inpr_a, by="row.names", all=TRUE)

# Set the year as rownames and delete from first colum
year_names <- df_industrial[, 1]
df_industrial[, 1] <- NULL
rownames(df_industrial) <- year_names

```

1.7 Merge all possible drivers

```

# Include transport and energy data
df_climate <- merge(df_tran, df_energy, by="row.names", all=TRUE)

# Set the year as rownames and delete from first colum

```

```

year_names <- df_climate[, 1]
df_climate[, 1] <- NULL
rownames(df_climate) <- year_names

# Include greenhouse gas emissions data
df_climate <- merge(df_gas, df_climate, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_climate[, 1]
df_climate[, 1] <- NULL
rownames(df_climate) <- year_names

# Include waste data
df_climate <- merge(df_waste, df_climate, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_climate[, 1]
df_climate[, 1] <- NULL
rownames(df_climate) <- year_names

# Include industrial data
df_climate <- merge(df_industrial, df_climate, by="row.names", all=TRUE)

# Set the year as rownames and delete from first column
year_names <- df_climate[, 1]
df_climate[, 1] <- NULL
rownames(df_climate) <- year_names

# Substitute , with _ in the columns names otherwise errors show up
col <- colnames(df_climate)
col <- gsub(",", "_", col)
col <- gsub(" ", "", col)
colnames(df_climate) <- col

```

1.8 Save data

```

write.csv(df_climate, "data/Eurostat/df_climate_raw.csv", row.names = TRUE)

```