

Práctica: Selección del conjunto de datos

Maite Gracia

3 de May, 2021

Contents

1 Introducción

1

1 Introducción

Combinación de las tres bases de datos escogidas para la práctica de Visualización de datos del Máster de Ciencias de datos de la Universitat Oberta de Catalunya. Con esto se pretende crear un dataset final que englobará el dataset “Data on CO2 and Greenhouse Gas Emissions” enlace, “Respiratory diseases death rate” enlace and “Death rate upper respiratory age” enlace, todas ellas extraídas de la web Our World in Data

```
# Carga dataset Our World in Data sobre la emisión de CO2 y los gases de
# efecto invernadero
dfowid <- read.csv('./data/owid-co2-data.csv')
head(dfowid, 2)
```

```
##   iso_code    country year   co2 co2_growth_prct co2_growth_abs
## 1      AFG Afghanistan 1949 0.015                NA              NA
## 2      AFG Afghanistan 1950 0.084                475              0.07
##   consumption_co2 trade_co2 trade_co2_share co2_per_capita
## 1              NA        NA              NA          0.002
## 2              NA        NA              NA          0.011
##   consumption_co2_per_capita share_global_co2 cumulative_co2
## 1                      NA              0.000          0.015
## 2                      NA              0.001          0.099
##   share_global_cumulative_co2 co2_per_gdp consumption_co2_per_gdp
## 1                      0              NA              NA
## 2                      0              0.004              NA
##   co2_per_unit_energy cement_co2 coal_co2 flaring_co2 gas_co2 oil_co2
## 1              NA        NA    0.015              NA    NA    NA
## 2              NA        NA    0.021              NA    NA    0.063
##   other_industry_co2 cement_co2_per_capita coal_co2_per_capita
## 1              NA              NA          0.002
## 2              NA              NA          0.003
##   flaring_co2_per_capita gas_co2_per_capita oil_co2_per_capita
## 1              NA              NA              NA
## 2              NA              NA          0.008
##   other_co2_per_capita share_global_coal_co2 share_global_oil_co2
## 1              NA              0.000              NA
```

```
## 2          NA          0.001          0.004
## share_global_gas_co2 share_global_flaring_co2 share_global_cement_co2
## 1          NA          NA          NA
## 2          NA          NA          NA
## cumulative_coal_co2 cumulative_oil_co2 cumulative_gas_co2
## 1          0.015          NA          NA
## 2          0.036          0.063          NA
## cumulative_flaring_co2 cumulative_cement_co2 share_global_cumulative_coal_co2
## 1          NA          NA          0
## 2          NA          NA          0
## share_global_cumulative_oil_co2 share_global_cumulative_gas_co2
## 1          NA          NA
## 2          0          NA
## share_global_cumulative_flaring_co2 share_global_cumulative_cement_co2
## 1          NA          NA
## 2          NA          NA
## total_ghg ghg_per_capita methane methane_per_capita nitrous_oxide
## 1          NA          NA          NA          NA          NA
## 2          NA          NA          NA          NA          NA
## nitrous_oxide_per_capita primary_energy_consumption energy_per_capita
## 1          NA          NA          NA
## 2          NA          NA          NA
## energy_per_gdp population          gdp
## 1          NA          7663783          NA
## 2          NA          7752000 19494799360
```

```
# Carga dataset United Nations, muertes por causa de muertes por pais
drespiratory <- read.csv('./data/respiratory-disease-death-rate.csv')
head(drespiratory, 2)
```

```
##          Entity Code Year
## 1 Afghanistan AFG 1990
## 2 Afghanistan AFG 1991
## Deaths...Chronic.respiratory.diseases...Sex..Both...Age..Age.standardized..Rate.
## 1          95.27378
## 2          95.27066
```

```
# Cambiamos el nombre de las columnas Code a is_code y Year a year para que coincida
# con los nombres de las columnas del dataset the Our World in data.
```

```
drespiratory$Entity <- NULL
names(drespiratory)[1] <- paste('iso_code')
names(drespiratory)[2] <- paste('year')
names(drespiratory)[3] <- paste('death_rate')
head(drespiratory, 2)
```

```
## iso_code year death_rate
## 1 AFG 1990 95.27378
## 2 AFG 1991 95.27066
```

```
# Carga dataset United Nations, muertes por causa de muertes por pais
drespiratoryage <- read.csv('./data/death-rate-upper-respiratory-age.csv')
head(drespiratoryage, 2)
```

```
##          Entity Code Year
## 1 Afghanistan AFG 1990
## 2 Afghanistan AFG 1991
```

```
## Deaths...Upper.respiratory.infections...Sex..Both...Age..50.69.years..Rate.
## 1 0.05150558
## 2 0.05236421
## Deaths...Upper.respiratory.infections...Sex..Both...Age..Under.5..Rate.
## 1 0.2151638
## 2 0.1990897
## Deaths...Upper.respiratory.infections...Sex..Both...Age..70..years..Rate.
## 1 0.1604764
## 2 0.1623168
## Deaths...Upper.respiratory.infections...Sex..Both...Age..5.14.years..Rate.
## 1 0.01064985
## 2 0.01062747
## Deaths...Upper.respiratory.infections...Sex..Both...Age..15.49.years..Rate.
## 1 0.010105182
## 2 0.009865648
```

```
# Al igual que en el caso anterior, cambiamos el nombre de las columnas Code a
# iso_code y Year a year para que coincida
# con los nombres de las columnas del dataset the Our World in data.
```

```
drespiratoryage$Entity <- NULL
names(drespiratoryage)[1] <- paste('iso_code')
names(drespiratoryage)[2] <- paste('year')
names(drespiratoryage)[3] <- paste('death_age_50_to_69')
names(drespiratoryage)[4] <- paste('death_age_under_5')
names(drespiratoryage)[5] <- paste('death_age_above_70')
names(drespiratoryage)[6] <- paste('death_age_5_to_14')
names(drespiratoryage)[7] <- paste('death_age_15_to_49')
head(drespiratoryage, 2)
```

```
## iso_code year death_age_50_to_69 death_age_under_5 death_age_above_70
## 1 AFG 1990 0.05150558 0.2151638 0.1604764
## 2 AFG 1991 0.05236421 0.1990897 0.1623168
## death_age_5_to_14 death_age_15_to_49
## 1 0.01064985 0.010105182
## 2 0.01062747 0.009865648
```

```
# Se crean las nuevas variables que añadiremos al dataset final extraídas de los
# dos datasets complementarios sobre muertes por enfermedades respiratorias y
# muertes por enfermedades respiratorias repartidos por edad
```

```
drespiratoryNew <- drespiratory
drespiratoryageNew <- drespiratoryage
final_dataset <- dfowid
final_dataset['death_rate'] <- NA
final_dataset['death_age_50_to_69'] <- NA
final_dataset['death_age_under_5'] <- NA
final_dataset['death_age_above_70'] <- NA
final_dataset['death_age_5_to_14'] <- NA
final_dataset['death_age_15_to_49'] <- NA

for (i in 1:length(drespiratoryNew$iso_code)) {
  for (j in 1:length(final_dataset$iso_code)) {
    if (drespiratoryNew$iso_code[i] == final_dataset$iso_code[j] && drespiratoryNew$year[i] == final_da
      final_dataset$death_rate[j] <- drespiratoryNew$death_rate[i]
    }
  }
}
```

```

}

for (i in 1:length(drespiratoryageNew$iso_code)) {
  for (j in 1:length(final_dataset$iso_code)) {
    if (drespiratoryageNew$iso_code[i] == final_dataset$iso_code[j] && drespiratoryageNew$year[i] == final_dataset$year[j]) {
      final_dataset$death_age_50_to_69[j] <- drespiratoryageNew$death_age_50_to_69[i]
      final_dataset$death_age_under_5[j] <- drespiratoryageNew$death_age_under_5[i]
      final_dataset$death_age_above_70[j] <- drespiratoryageNew$death_age_above_70[i]
      final_dataset$death_age_5_to_14[j] <- drespiratoryageNew$death_age_5_to_14[i]
      final_dataset$death_age_15_to_49[j] <- drespiratoryageNew$death_age_15_to_49[i]
    }
  }
}

library(writexl)
# Generamos el dataset final y lo exportamos a csv y excel
write.csv(final_dataset, './data/co2_greenhouse_emissions_deaths.csv')
write_xlsx(final_dataset, './data/co2_greenhouse_emissions_deaths.xlsx')

```