In [3]:
```
library(magrittr)
library(dplyr)
library(countrycode)
library(reshape)
```

# Dataset COMTRADE

Importo dataset.

In [1]:
```
full_data <- read.csv("comtrade_1996_2020_small.csv",
                      sep=",", header=TRUE)
head(full_data[,1:5])
head(full_data[,5:10])
```

A data.frame: 6 × 5

| | year | five_year | exporter_code | exporter | exporter_iso |
|---|---|---|---|---|---|
| | <int> | <int> | <int> | <chr> | <chr> |
| **1** | 1997 | 2000 | 4 | Afghanistan | AFG |
| **2** | 1998 | 2000 | 4 | Afghanistan | AFG |
| **3** | 1999 | 2000 | 4 | Afghanistan | AFG |
| **4** | 2000 | 2000 | 4 | Afghanistan | AFG |
| **5** | 1996 | 2000 | 4 | Afghanistan | AFG |
| **6** | 1996 | 2000 | 4 | Afghanistan | AFG |

A data.frame: 6 × 6

| | exporter_iso | importer_code | importer | importer_iso | imports | five_yrs_avg_imports |
|---|---|---|---|---|---|---|
| | <chr> | <int> | <chr> | <chr> | <dbl> | <dbl> |
| **1** | AFG | 12 | Algeria | DZA | 213595 | 269108.4 |
| **2** | AFG | 12 | Algeria | DZA | 453745 | 269108.4 |
| **3** | AFG | 12 | Algeria | DZA | 409023 | 269108.4 |
| **4** | AFG | 12 | Algeria | DZA | 24812 | 269108.4 |
| **5** | AFG | 12 | Algeria | DZA | 244367 | 269108.4 |
| **6** | AFG | 20 | Andorra | AND | 25003 | 29412.6 |

Limito il dataset all'anno, al nome del paese esportatore e importatore e al valore in USD dei beni scambiati.

In [5]:
```
new_data<-full_data %>%
  select(year,exporter,importer,imports)
```

```
colnames(new_data)<-c("year","origin","destination","value_comtrade")

#ordino così le colonne
comtrade <- new_data %>%
  select(origin,destination,year,value_comtrade)
```

Costruisco il dataset dei paesi presenti sia tra gli esportatori che tra gli importatori. Verifico che la trasformazione del nome dei paesi in ISO3c sia coerente con quella degli altri dataset appartenenti alla rete utilizzando i codici ISO3c contenuti nella libreria di R "countrycode". Ci sono diversi nomi che raggruppano aree geografiche più ampie: lascio che l'ISO3c corrispondente sia uguale a NA; altri sono scritti in modo diverso (non vengono riconosciuti dalla funzione): inserisco manualmente l'ISO3c.

In [6]:
```
paesi_dataset <- c(unique(comtrade$origin),setdiff(unique(comtrade$destination),
                                                   unique(comtrade$origin)))
paesi_iso3c <- countrycode(paesi_dataset,origin="country.name",destination="iso3c")
paesi <- as.data.frame(cbind(paesi_dataset,paesi_iso3c))

paesi$paesi_iso3c[33] <- "IOT"
paesi$paesi_iso3c[34] <- "VGB"
paesi$paesi_iso3c[81] <- "FSM"
paesi$paesi_iso3c[253] <- "CUW"
paesi$paesi_iso3c[256] <- "BES"
paesi$paesi_iso3c[257] <- "BLM"
```

```
Warning message:
"Some values were not matched unambiguously: Areas, nes, Belgium-Luxembourg, Bonair
e, Br. Antarctic Terr., Br. Indian Ocean Terr., Br. Virgin Isds, Bunkers, CACM, nes,
Caribbean, nes, CuraÃƒÂ§ao, Europe EU, nes, Fr. South Antarctic Terr., Free Zones, F
S Micronesia, LAIA, nes, Neth. Antilles, Neutral Zone, North America and Central Ame
rica, nes, Northern Africa, nes, Oceania, nes, Other Africa, nes, Other Asia, nes, O
ther Europe, nes, Rest of America, nes, Saint BarthÃƒÂ©lemy, Serbia and Montenegro,
So. African Customs Union, Special Categories, Western Asia, nes, World
"
```

Ora sostituisco i nomi dei paesi presenti nel dataset con i rispettivi ISO3c così determinati.

In [8]:
```
comtrade$origin <- paesi$paesi_iso3c[match(as.factor(comtrade$origin),
                                           as.factor(paesi$paesi_dataset))]
comtrade$destination <- paesi$paesi_iso3c[match(as.factor(comtrade$destination),
                                                as.factor(paesi$paesi_dataset))]
#rimuovo le righe con i NA dei paesi non considerati
comtrade <- na.omit(comtrade)

#rimuovo i link interni, ossia che hanno l'importatore uguale all'esportatore
comtrade <- comtrade[-c(which(comtrade$origin == comtrade$destination)),]
```

Creo un dataset con tutte le possibili combinazioni tra *origin, destination e year* (tra quelli presenti nel dataset originale).

In [9]:
```
p <- c(unique(comtrade$destination), setdiff(unique(comtrade$origin),
                                             unique(comtrade$destination)))
```

```
y <- unique(comtrade$year)
tot <- as.data.frame(cbind(rep(p, each = length(p)*length(y)),
                           rep(rep(p,times=length(p)),each=length(y)),
                           rep(y, times = length(p)^2)))
colnames(tot) <- c("origin","destination","year")
tot <- tot[-c(which(tot$origin == tot$destination)),]
tot <- cbind(tot,rep(0,nrow(tot)))
colnames(tot) <- c("origin","destination","year","value_comtrade")
tot$year <- as.numeric(tot$year)
merge <- rbind(comtrade,tot)
comtrade_tot <- merge %>%
  group_by(origin,destination,year) %>%
  summarise(value_comtrade = sum(value_comtrade))
```

```
`summarise()` has grouped output by 'origin', 'destination'. You can override using
the `.groups` argument.
```

In [10]:
```
comtrade_tot <- comtrade_tot[order(comtrade_tot$year),]
write.csv(comtrade_tot,"comtrade_tot.csv", row.names = FALSE)
```