

```
In [ ]: library(readxl)
library(dplyr)
library(zoo)
library("stringr")
library(reshape)
library(haven)
library("countrycode")
```

Dataset INPACT

Importo il dataset.

```
In [2]: start_data <- read_dta("INPACT-S.dta")
head(start_data[,1:4])
head(start_data[,5:9])
```

A tibble: 6 × 4

appln_auth	appln_auth_iso3	appln_auth_name	person_etry_code
<chr>	<chr>	<chr>	<chr>
AM	ARM	ARMENIA	AM
AM	ARM	ARMENIA	AM
AM	ARM	ARMENIA	AM
AR	ARG	ARGENTINA	AR
AR	ARG	ARGENTINA	DE
AR	ARG	ARGENTINA	DK

A tibble: 6 × 5

person_etry_code_iso3	person_etry_code_name	appln_filing_year	isic_rev3_2	patents_applt
<chr>	<chr>	<dbl>	<dbl>	<dbl>
ARM	ARMENIA	2000	15	0.11579378
ARM	ARMENIA	2000	24	1.74325302
ARM	ARMENIA	2000	41	0.03337383
ARG	ARGENTINA	1980	10	0.25436512
DEU	GERMANY(3)	1980	10	0.03811470
DNK	DENMARK	1980	10	0.02182874

Seleziono il paese dove risiede l'inventore indicato con codice ISO3c

(person_etry_code_iso3), il paese dove risiede la "autorità dei brevetti" che richiede il rilascio del brevetto (appl_auth_iso3), l'anno (appln_filing_year) e la misura del numero di brevetti

rilasciati (patents_applt). Il motivo per cui il valore non è un intero è il peso che gli autori di questo dataset hanno attribuito allo scambio del brevetto a seconda di quanti paesi hanno rilasciato quello stesso brevetto, a seconda di quanti e quali paesi compongono l'autorità dei brevetti che richiede il rilascio, a seconda di quanti ambiti industriali quella nuova tecnologia va ad impattare.

```
In [5]: new_data <- start_data %>%
  select(person_ctry_code_iso3,appln_auth_iso3,appln_filing_year,patents_applt)
colnames(new_data) <- c("origin","destination","year","patents")
new_data <- new_data[-c(which(new_data$origin == new_data$destination)),]

#raggruppo quelle righe che hanno le stesse colonne di origin, destination e year
inpact <- new_data %>%
  group_by(origin,destination,year) %>%
  summarise(patents = sum(patents))

inpact <- inpact[order(inpact$year),]
```

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

Controllo il codice ISO3c utilizzato.

```
In [7]: paesi_dataset <- c(unique(inpact$origin),
  setdiff(unique(inpact$destination),unique(inpact$origin)))
paesi_giusti <- countrycode(paesi_dataset,origin="iso3c",
  destination="iso3c")

#ok niente da dichiarare
```

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

```
In [9]: p <- paesi_dataset
y <- unique(inpact$year)
tot <- as.data.frame(cbind(rep(p, each = length(p)*length(y)),
  rep(rep(p,times=length(p)),each=length(y)),
  rep(y, times = length(p)^2)))
colnames(tot) <- c("origin","destination","year")
tot <- tot[-c(which(tot$origin == tot$destination)),]
tot <- cbind(tot,rep(0,nrow(tot)))
colnames(tot) <- c("origin","destination","year","patents")
tot$year <- as.numeric(tot$year)
merge <- rbind(inpact,tot)
inpact_tot <- merge %>%
  group_by(origin,destination,year) %>%
  summarise(patents_inpact = sum(patents))
```

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

```
In [11]: inpact_tot <- inpact_tot[order(inpact_tot$year),]
write.csv(inpact_tot,"inpact_tot.csv", row.names = FALSE)
```