**DANIELE TORREGIANI – S1107037 – ECONOMICS OF ICT – EXAM PROJECT: PATENT LANDSCAPE ANALYSIS**

In questo report si illustrano i risultati della *Patent landscape analysis* relativa alle tecnologie *green*, con un riferimento specifico ai brevetti che contengono tecnologie per il trattamento delle acque di scarico o per la gestione dei rifiuti (codice CPC: Y02W 🡪 *CLIMATE CHANGE MITIGATION TECHNOLOGIES RELATED TO WASTEWATER TREATMENT OR WASTE MANAGEMENT*).

Le classi Y02 si distinguono dalle altre perché il codice di classificazione viene assegnato in modo automatico da algoritmi ottimizzati per ridurre gli errori (anziché analizzando singolarmente ogni documento brevettuale).

Il codice Y02 viene così assegnato a tutte le tecnologie o applicazioni per la mitigazione o l’adattamento contro il cambiamento climatico (ovvero alle tecnologie che riducono le emissioni di gas responsabili dell’effetto serra), nell’ottica di rispettare gli accordi presi durante la Conferenza di Parigi COP21 del 2015.

Esistono 8 sottoclassi rispetto alla codifica Y02:

Immagine che contiene testo

Descrizione generata automaticamente

Ora verranno mostrati i risultati ottenuti e il relativo commento.

**DATASET**

Foglio excel di partenza: “green\_patents\_secondo\_appello.xlsx”

Prima di svolgere l’analisi brevettuale vera e propria è stata necessaria una pulizia iniziale dei dati, in modo da ottenere una tabella di dati che si concentrasse unicamente sui brevetti contenenti tecnologie Y02W (dataframe “data” nello script R).

Successivamente, per costruire le statistiche descrittive richieste per l’analisi, ho optato per la creazione di altri dataframe esclusivi per ogni punto dell’analisi (dataframe “one”, “two”, “three”, … nello script), facendo sempre riferimento al df “data” iniziale.

**DATA ELABORATION: MACRO LEVEL**

**1) NUMERO DI BREVETTI PER PERIODO (Tabella)**

Per la prima statistica descrittiva ho filtrato il dataframe iniziale in modo da lasciare solo gli identificativi relativi ai brevetti contenenti tecnologie Y02W, aggiungendo una ulteriore colonna che indicasse il decennio nel quale il singolo brevetto è stato pubblicato.

Contando il numero di brevetti prodotti in ogni decennio, la tabella che si ottiene è la seguente:

Immagine che contiene testo, armadietto, arredamento, tabellonesegnapunti

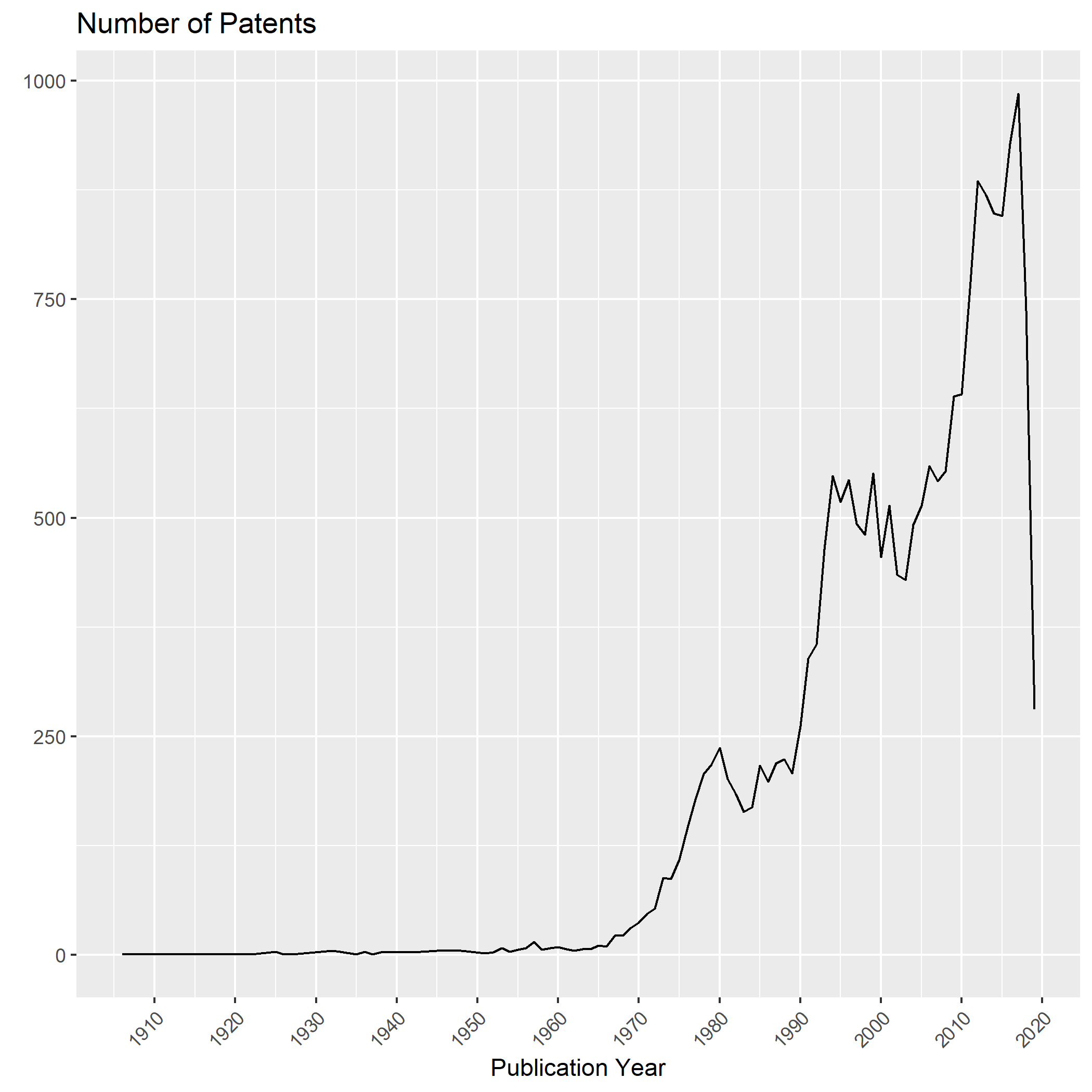
Descrizione generata automaticamenteSi può notare come il numero di brevetti contenenti tecnologie “attente” al riciclo ed al riutilizzo dei materiali sia cresciuto esponenzialmente dagli anni 2000.

Non è un risultato casuale, in quanto a partire da questi anni aumenta la consapevolezza e la necessità di adottare un modello economico circolare rispettoso dell’ambiente e della sostenibilità.

Tuttavia, ci sono due punti da sottolineare.

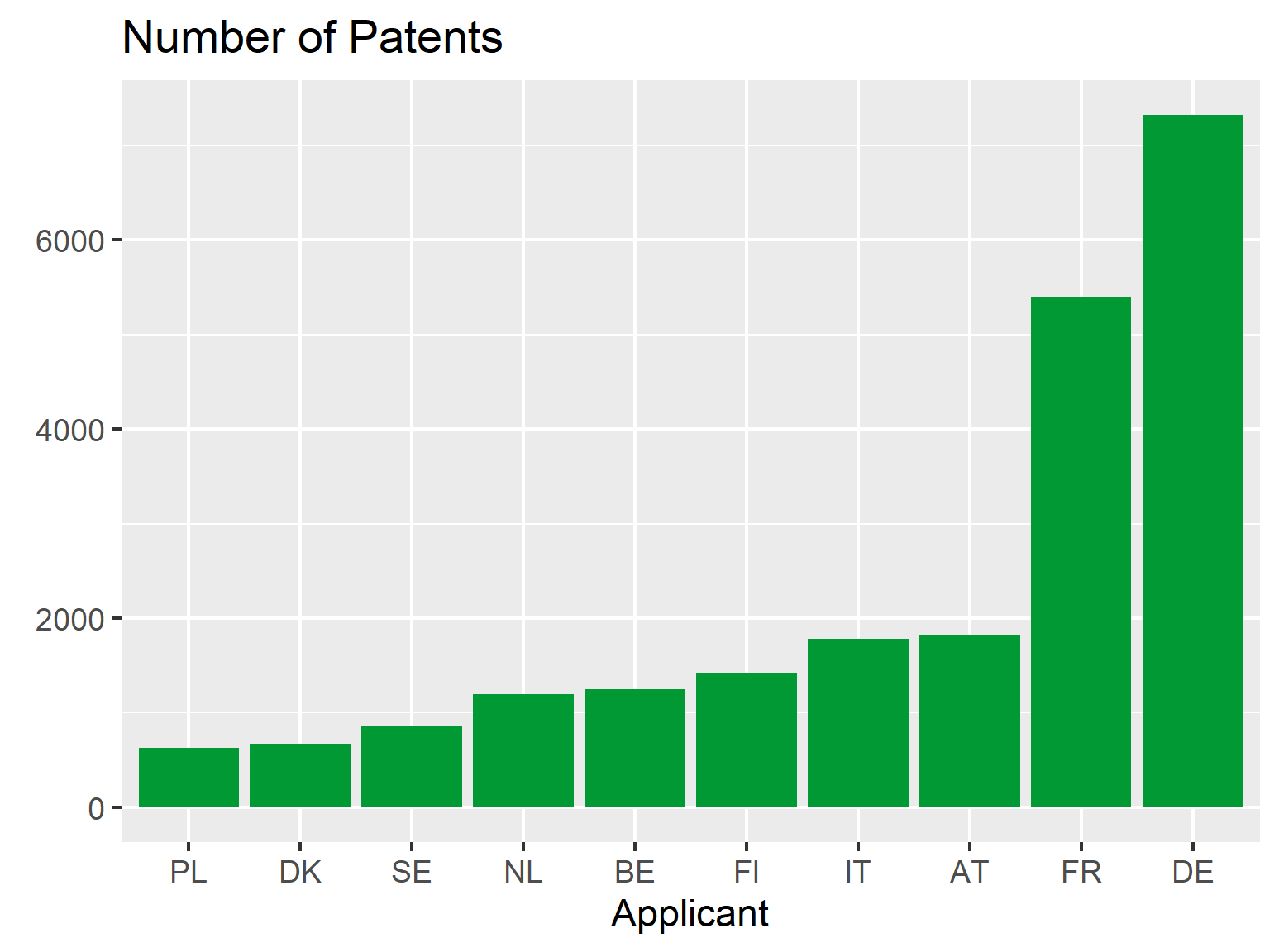
Il primo è la presenza di tecnologie contrassegnate dall’algoritmo automatico come Y02W già a partire dall’inizio del secolo scorso (ad esempio il brevetto FR361300 pubblicato nel 1906 col titolo tradotto dal francese “Metodo per l'utilizzo dei rifiuti di gomma” e il brevetto FR448616 del 1913, anch’esso di origine francese, relativo ad un metodo di riciclo della carta bianca).

Il secondo punto riguarda invece la notevole flessione nell’ultimo decennio rispetto al precedente (passaggio da 11784 a 3897 brevetti). Questo risultato potrebbe essere spiegato dal fatto che i dati non sono aggiornati. Infatti, c’è un lag temporale dal momento in cui l’impresa fa domanda all’ufficio brevettuale per ottenere un brevetto al momento in cui l’ufficio pubblica il documento brevettuale. Senza questo ritardo, magari, anche i valori dell’ultimo decennio risulterebbero in linea con il trend positivo.

**2) NUMERO DI BREVETTI PER ANNO (Grafico in serie storica)**

Dal secondo grafico sembra confermarsi l’ipotesi del punto precedente, ovvero che la flessione sia dovuta ad un ritardo temporale tra la domanda iniziale all’effettiva pubblicazione da parte dell’ufficio brevettuale. Infatti, il trend di crescita è visibile fino al 2017, mentre cala notevolmente solo nel 2018 e nel 2019 (intervallo di tempo che potrebbe corrispondere al ritardo tra la domanda e l’effettiva pubblicazione).

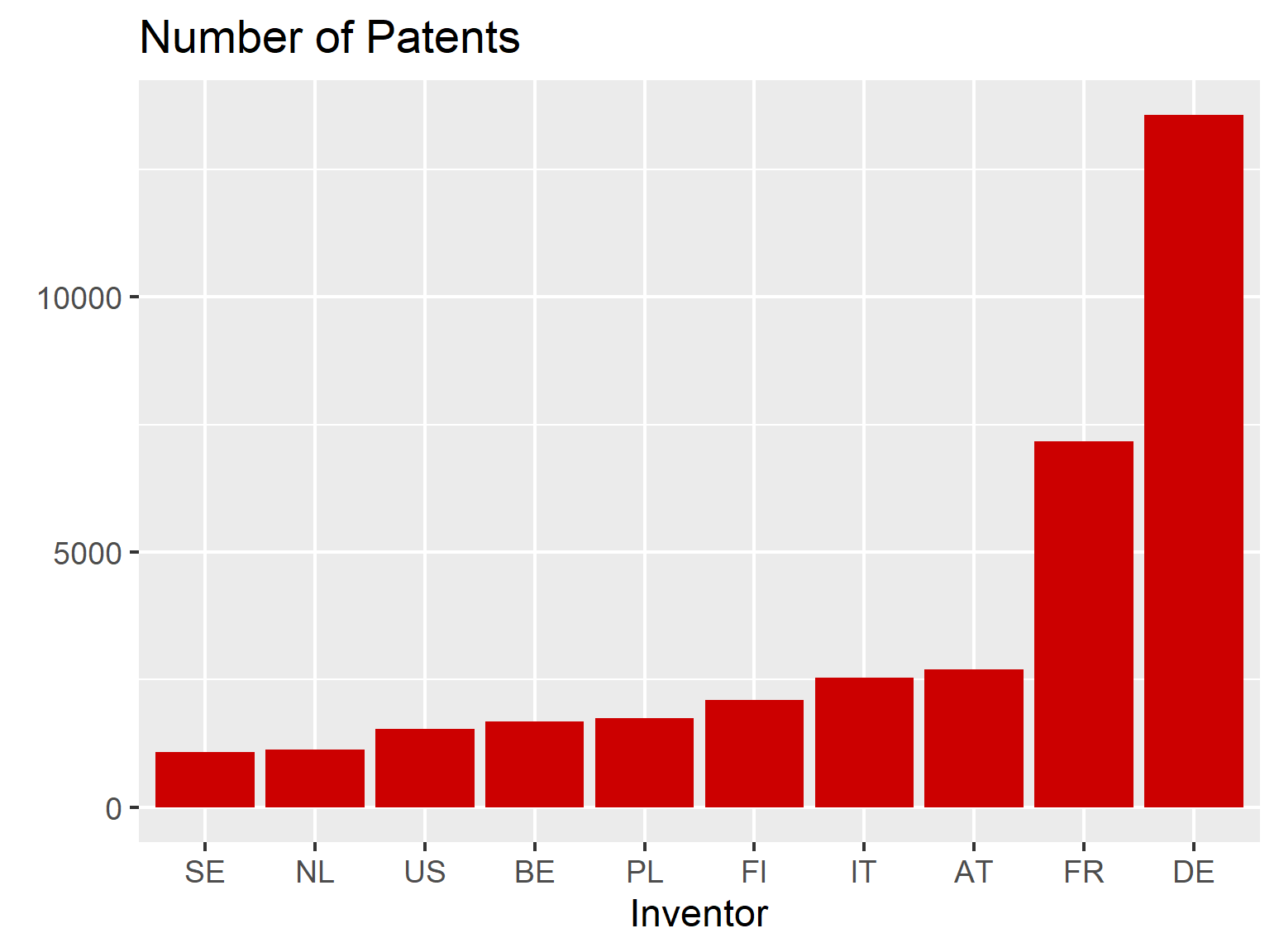
In ogni caso da questa descrittiva si può notare come il ciclo di vita delle tecnologie relative al riciclo dei materiali YO2W segua una forma ad S. In realtà, potremmo trovarci ancora in uno stato di crescita che purtroppo non è visibile dal grafico per via del possibile problema dei dati non aggiornati.

**3) NUMERO DI BREVETTI PER PAESE DI ORIGINE DELL’APPLICANT (Grafico a barre)**

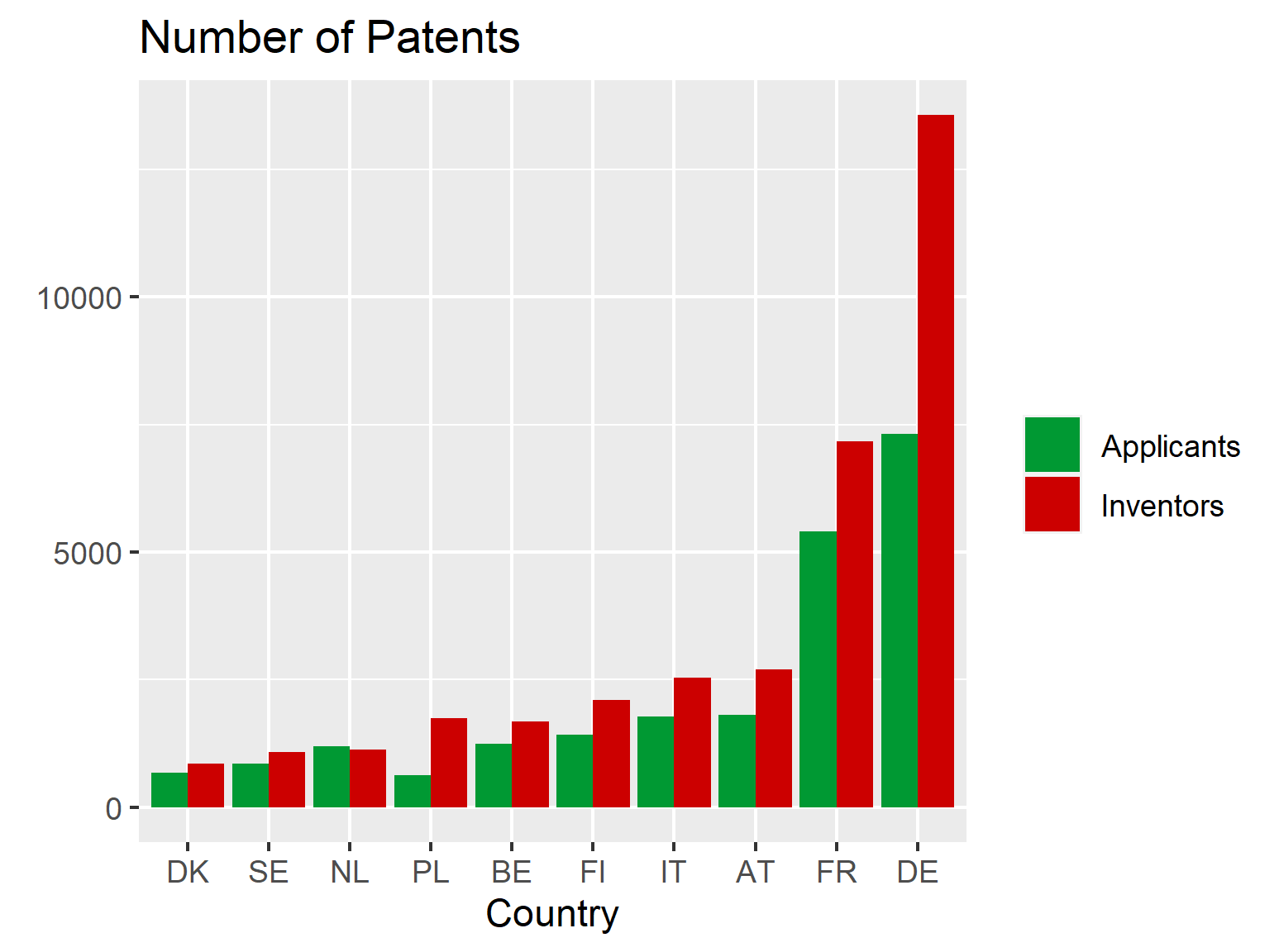
Nella terza descrittiva vengono riportati i primi 10 Paesi europei con il numero maggiore di brevetti rispetto al Paese di origine del proprietario della tecnologia (*Applicant*).

Emerge il ruolo della Germania come leader tecnologico per l’economia circolare in Europa, seguita dalla Francia. Nelle posizioni successive si trovano Austria e Italia, anche se il numero di brevetti cala notevolmente (meno di un terzo rispetto alla Germania).

**4) NUMERO DI BREVETTI PER PAESE DI ORIGINE DELL’INVENTORE (Grafico a barre)**

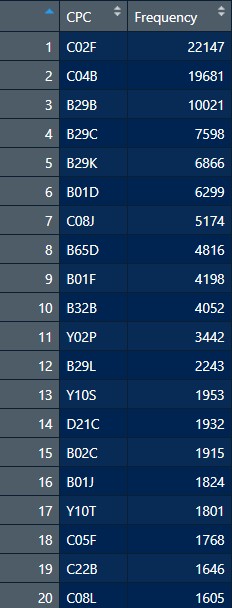
Per dare maggiore robustezza al risultato al punto 3, è opportuno confrontarlo con i Paesi con più brevetti rispetto al Paese di origine dell’inventore della tecnologia.

Questo perché il proprietario o applicant è solitamente l’organizzazione/impresa privata che ha il diritto di fare domanda di brevetto, mentre i veri inventori sono la persona o il team che creano la nuova tecnologia e che solitamente fanno capo all’applicant. Per questo è importante verificare se ci sono differenze tra i due risultati. Dal grafico a lato non sembrano emergere differenze importanti nelle prime posizioni.

Per visualizzazione meglio questo confronto riporto anche un altro grafico a barre con entrambi i valori (Numero di brevetti per Stato dell’applicant e dell’inventore): i risultati sembrano confermare la leadership tedesca, seguita da Francia, Austria e Italia.

Una menzione va però alla Polonia. Infatti, essa può vantare un’alta presenza di inventori specializzati nelle tematiche ambientali e di riciclo; tuttavia, essendo un Paese in via di sviluppo, la partecipazione delle imprese e delle organizzazioni attente a questo tema è più contenuta.

**5) NUMERO E FREQUENZA DELLE CLASSI CPC PER BREVETTI RELATIVI ALLA TECNOLOGIA Y02W (Tabella)**

****Riporto la tabella con le frequenze delle altre tecnologie CPC contenute nei brevetti che hanno anche la classe Y02W. Per motivi di sintesi mostro soltanto le prime 20 tecnologie in ordine di frequenza decrescente (ottenuta conteggiando i codici per ogni brevetto). Quindi, nella tabella a lato, non vengono mostrati tutti gli altri 317 codici CPC meno numerosi.

Questa descrittiva è molto utile per cercare di visualizzare la portata della tecnologia Y02W, ovvero il grado di associazione della tecnologia Y02W con altre innovazioni. Essendo associata con oltre 300 altri codici, si potrebbe affermare che l’attenzione verso il riciclo si estende su molti campi. Andando a vedere le tecnologie maggiormente collegate con la Y02W emerge che essa viene applicata soprattutto nel settore della chimica (codice C) per il trattamento delle acque, acque di scarico e delle fogne (C02F). La tecnologia Y02W è molto impiegata anche nella lavorazione della plastica (codici B29).

Sembrano risultati ragionevoli: l’applicazione delle tecnologie per la gestione dei rifiuti è legata in particolare al corretto riutilizzo delle sostanze plastiche e al trattamento delle acque reflue.

**6) NUMERO DI COLLABORAZIONI SUI BREVETTI (Tabella)**

Per costruire la seguente tabella ho creato una funzione (collaboration() nello script) in modo tale da riuscire a contare, sul totale dei brevetti che avessero dati disponibili sugli applicant (no NA), il numero di brevetti nei quali compaiono più proprietari diversi (più organizzazioni co-proprietarie dello stesso brevetto).

I risultati basati sui dati disponibili mostrano una buona tendenza alla collaborazione tra le organizzazioni (41% sul totale dei brevetti).

La collaborazione e un modello di innovazione di tipo “aperto” (*Open Innovation*) possono portare notevoli benefici, tra cui la possibilità di condividere i costi per la Ricerca e Sviluppo, di ridurre i tempi per l’innovazione e di ottenere benefici economici e strategici stipulando accordi di licenza.

Ho ritenuto interessante vedere se la collaborazione sui brevetti per la tecnologia in questione (Y02W) fosse un fenomeno relativamente recente. Per questo ho filtrato i dati tenendo in considerazione solo i brevetti pubblicati a partire dal 2010.

Immagine che contiene testo

Descrizione generata automaticamenteCome si può vedere da questa seconda tabella, il numero di collaborazioni in % sul totale dei brevetti per la tecnologia Y02W, calcolato con i soli dati dal 2010 in poi, è cresciuto (52%). È sintomo che il fenomeno di Open Innovation per il riciclo dei materiali è recente.

**7) NUMERO DI COLLABORAZIONI NAZIONALI ED INTERNAZIONALI SUI BREVETTI (Tabella)**

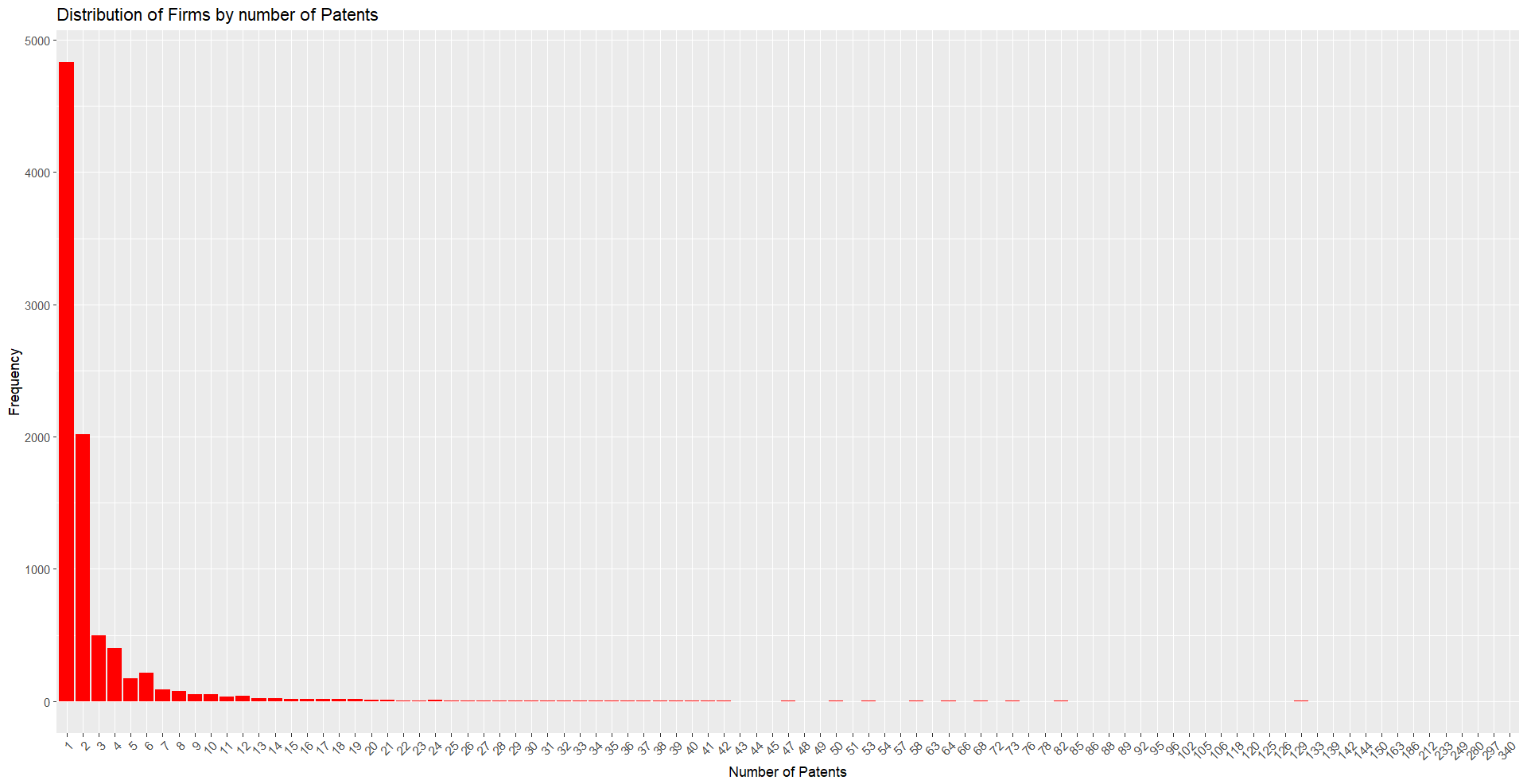
Ho aggiunto un’altra funzione ad hoc (int\_nat() nello script) per contare le collaborazioni tra proprietari di Paesi diversi (collaborazioni internazionali) o tra proprietari dello stesso Paese (collaborazioni nazionali).

Questa descrittiva è interessante anche per verificare la presenza o meno di un flusso di conoscenza tra diversi Stati in tema di riciclo dei materiali.

Immagine che contiene testo

Descrizione generata automaticamenteDalla tabella emerge che la maggior parte delle collaborazioni è confinata tra proprietari dello stesso Paese (80% sul totale delle collaborazioni); tuttavia, non è da escludere l’esistenza di un flusso di conoscenza delle tecnologie per la gestione di rifiuti anche tra Stati diversi (20%).

**8) DISTRIBUZIONE DELLE IMPRESE PER NUMERO DI BREVETTI (Grafico a barre)**



Nel grafico a barre si riportano le frequenze (conteggio sull’asse y) del numero di brevetti prodotti dalle singole organizzazioni (asse x).

La distribuzione è di tipo *Power Law*: la maggior parte delle organizzazioni nel dataset detiene soltanto 1 o 2 brevetti relativi alla tecnologia Y02W. Tuttavia, anche se in minoranza, non mancano le organizzazioni proprietarie di un numero ingente di brevetti per la gestione dei rifiuti (6 organizzazioni dispongono di oltre 200 brevetti sul tema, con un picco per la società francese “*Degremont*”, id = 2097, specializzata nella produzione di acqua potabile e nel trattamento delle acque di scarico).

**DATA ELABORATION: FIRM LEVEL**

**9) DISTRIBUZIONE DEI BREVETTI PER APPLICANT (Tabella) E CONCENTRATION RATIO INDEX (CR4)**

Ho costruito altre 2 funzioni specifiche per calcolare il percentuale di brevetti detenuti dalle singole organizzazioni proprietarie rispetto al totale dei brevetti e per ottenere il Concentration Ratio Index (CR4 è un indicatore utile per capire la dimensione delle singole imprese rispetto all’intera industria).

Immagine che contiene testo

Descrizione generata automaticamenteRiportando le prime 20 organizzazioni per numero di brevetti relativi alla tecnologia Y02W troviamo appunto la società Degremont con 340 brevetti, seguita da Lafarge, Erema Engineering e Bayer AG.

Guardando alle percentuali delle singole imprese rispetto al totale dei brevetti per i quali sono disponibili i dati anche sugli applicant (oltre 28.000 brevetti) non emerge un leader monopolista dominante del settore, in quanto Degremont, al primo posto, detiene appena l’1.20% dei brevetti connessi alla gestione dei rifiuti.

Non a caso, il CR calcolato per le prime 4 organizzazioni è relativamente basso (CR4 = 4.10 %).

**10) DISTRIBUZIONE DEI BREVETTI PER APPLICANT SU PERIODI DIVERSI (Tabella)**

Immagine che contiene testo, screenshot, schermo, computer

Descrizione generata automaticamenteÈ lecito chiedersi ora se la leadership tecnologica in tema di riciclo sia sempre stata la stessa nel tempo.

**Brevetti pubblicati prima del 2000**

CR4 = 6.63 %

Immagine che contiene testo, screenshot, schermo, computer

Descrizione generata automaticamente

**Brevetti pubblicati tra il 2001 e il 2009**

CR4 = 5.35 %

Immagine che contiene testo, screenshot, schermo, computer

Descrizione generata automaticamente

**Brevetti pubblicati dopo il 2009**

CR4 = 6.90 %

Come si può vedere dalle 3 tabelle, le leadership cambiano nel corso del tempo: solo Degremont resta sempre tra le prime 10 nel corso del tempo; alcune organizzazioni scompaiono dalla lista col passare degli anni, altre invece emergono solo nell’ultimo decennio. Tra i motivi di questo cambiamento della leadership potrebbe rientrare, ad esempio, l’impatto della crisi finanziaria del 2007/2008, che ha avuto riflessi importanti nella finanza e negli investimenti di tutto il mondo.

In generale, il Concentration Ratio Index relativo solo ai tre periodi è maggiore rispetto al CR4 visto nel punto precedente (calcolato su tutti i dati disponibili). Questo potrebbe essere motivato dal fatto che mantenere una quota alta di brevetti altamente specializzati costante nei decenni è molto complicato, e in ogni caso non emerge mai un leader monopolista dominante per la tecnologia Y02W. Tuttavia, dalla seconda alla terza tabella l’indicatore cresce dal 5.35% al 6.90%. Sarebbe interessante scoprire se questa leggera polarizzazione sarà una nuova tendenza confermata anche nel futuro dell’innovazione nella gestione dei rifiuti oppure se si tratta solo di un fenomeno transitorio.

**SCRIPT R:**

## EXAM PROJECT ##

rm(list=ls())

library(rstudioapi)

current\_path<-getActiveDocumentContext()$path

setwd(dirname(current\_path))

library(readxl)

library(tidyverse)

# importo dataset

data1 = read\_excel("green\_patents\_secondo\_appello.xlsx")

View(data1)

# 1) Publication\_number: Id of the patent

# 2) Publication\_year: Publication Year

# 3) Currentowners: Applicant's name

# 4) id\_Currentowners: ID of the applicant

# 5) Currentownerscountrycode: Applicant's country

# 6) Inventors: Inventor's name

# 7) Inventorscountrycode: Inventor's country

# 8) CooperativeClassification: CPC code

# 9) CooperativeClassificationlabel: CPC label

## CPC - COOPERATIVE PATENT CLASSIFICATION

## Y02W - CLIMATE CHANGE MITIGATION TECHNOLOGIES RELATED TO WASTEWATER TREATMENT OR WASTE MANAGEMENT

# focus su brevetti con tecnologie CPC: Y02W

data2 = data1 %>% filter(grepl("Y02W", CooperativeClassification)) %>% select(Publication\_number)

data = inner\_join(data1, data2, by = "Publication\_number") %>% distinct() # elimino duplicati con distinct

## DATA ELABORATION: MACRO LEVEL

# 1° - NUMBER OF PATENTS BY PERIOD - TABLE

# seleziono solo brevetti univoci

one1 = data %>% distinct(Publication\_number, Publication\_year)

# aggiungo colonna per il decennio relativo alla data di pubblicazione del brevetto

for (i in 1:nrow(one1)){

if (between(one1$Publication\_year[i], 1900, 1909)){

one1$Period[i] = "1900-1909"

}

else if (between(one1$Publication\_year[i], 1910, 1919)){

one1$Period[i] = "1910-1919"

}

else if (between(one1$Publication\_year[i], 1920, 1929)){

one1$Period[i] = "1920-1929"

}

else if (between(one1$Publication\_year[i], 1930, 1939)){

one1$Period[i] = "1930-1939"

}

else if (between(one1$Publication\_year[i], 1940, 1949)){

one1$Period[i] = "1940-1949"

}

else if (between(one1$Publication\_year[i], 1950, 1959)){

one1$Period[i] = "1950-1959"

}

else if (between(one1$Publication\_year[i], 1960, 1969)){

one1$Period[i] = "1960-1969"

}

else if (between(one1$Publication\_year[i], 1970, 1979)){

one1$Period[i] = "1970-1979"

}

else if (between(one1$Publication\_year[i], 1980, 1989)){

one1$Period[i] = "1980-1989"

}

else if (between(one1$Publication\_year[i], 1990, 1999)){

one1$Period[i] = "1990-1999"

}

else if (between(one1$Publication\_year[i], 2000, 2009)){

one1$Period[i] = "2000-2009"

}

else if (between(one1$Publication\_year[i], 2010, 2019)){

one1$Period[i] = "2010-2019"

}

else {

one1$Period[i] = "2020-2029"

}

}

# conteggio numero di brevetti per decennio e aggiungo riga per il totale

one = one1 %>% count(Period, name = "Number of Patents")

one$`Number of Patents` = as.numeric(one$`Number of Patents`)

total = c("Total", sum(one$`Number of Patents`))

one = rbind(one, total)

View(one)

# 2° - NUMBER OF PATENTS - TIME SERIES

two = one1 %>% count(Publication\_year, name = "Number of Patents")

colnames(two)[1] = c("Publication Year")

two %>%

ggplot(aes(`Publication Year`, `Number of Patents`)) +

geom\_line() +

scale\_x\_continuous(breaks = c(1910,1920,1930,1940,1950,1960,1970,1980,1990,2000,2010,2020)) +

theme(axis.text.x = element\_text(angle=45, hjust = 1)) +

labs(title = "Number of Patents", y = element\_blank()) +

ggsave(filename = "Num\_patents2.png")

# 3° - NUMBER OF PATENTS BY APPLICANT COUNTRY (TOP 10)

# dataframe solo con i 10 Paesi con più brevetti

three1 = data %>%

drop\_na(Currentownerscountrycode) %>%

count(Currentownerscountrycode, name = "Patents", sort = TRUE)

colnames(three1)[1] = "Applicant"

three1 = three1 %>%

mutate(Applicant = fct\_reorder(Applicant, Patents))

three = three1[1:10,]

three %>%

mutate(Applicant = fct\_reorder(Applicant, Patents)) %>%

ggplot(aes(Applicant, Patents)) +

geom\_bar(fill = "#009933", stat = "identity") +

labs(title = "Number of Patents", y = element\_blank()) +

ggsave("Patents\_by\_Applicant3.png")

# 4° - NUMBER OF PATENTS BY INVENTOR COUNTRY (TOP 10)

# dataframe solo con i 10 Paesi con più brevetti

four1 = data %>%

drop\_na(Inventorscountrycode) %>%

count(Inventorscountrycode, name = "Patents", sort = TRUE)

colnames(four1)[1] = "Inventor"

four1 = four1 %>%

mutate(Inventor = fct\_reorder(Inventor, Patents))

four = four1[1:10,]

four %>%

mutate(Inventor = fct\_reorder(Inventor, Patents)) %>%

ggplot(aes(Inventor, Patents)) +

geom\_bar(fill = "#CC0000", stat = "identity") +

labs(title = "Number of Patents", y = element\_blank()) +

ggsave("Patents\_by\_Inventor4.png")

# confronto Paesi Applicants - Inventors

colnames(four1)[1] = "Country"

colnames(three1)[1] = "Country"

full = full\_join(three1, four1, by = "Country")

full = full %>% pivot\_longer(cols = c("Patents.x", "Patents.y"))

full %>%

mutate(Country = fct\_reorder(Country, value)) %>%

drop\_na(value) %>%

head(20) %>%

ggplot(aes(Country, value, fill = name)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Number of Patents", y = element\_blank(), fill = element\_blank()) +

scale\_fill\_manual(values = c("#009933", "#CC0000"),

labels = c('Applicants', 'Inventors')) +

ggsave("Applicant\_vs\_Inventor.png")

# 5° - THE NUMBER OF CPC CLASSES OF PATENTS RELATED TO Y02 TECHNOLOGY

# estraggo solo le prime 4 lettere del codice CPC

five = data %>% distinct() %>% select(c(1,2,8))

five$CPC = substr(five$CooperativeClassification, 1, 4)

# costruisco una tabella con il conteggio delle altre tecnologie contenute nei brevetti con Y02W

five %>% drop\_na() %>% filter(CPC != "Y02W") %>%

count(CPC, name = "Frequency", sort = TRUE) %>% head(20) %>%

View()

# 6° - NUMBER OF PATENT COLLABORATION

# no NA per Stato e Id dell'applicant

six\_1 = data %>% drop\_na(id\_Currentowners, Currentownerscountrycode)

six = six\_1 %>% count(Publication\_number, name = "CPC\_for\_Patent") # numero totale brevetti = 17364

six = six %>% filter(CPC\_for\_Patent > 1) # brevetti con più di una tecnologia al suo interno

# collaborazione se brevetto con più di una tecnologia prodotte da applicant diversi

coll = inner\_join(six\_1, six, by = "Publication\_number")

# funzione per contare collaborazioni nei brevetti

x = 0

collaboration = function(dataframe, x){

for (i in 1:(nrow(dataframe) - 1)){

if (dataframe$Publication\_number[i] == dataframe$Publication\_number[i+1]) {

if (dataframe$id\_Currentowners[i] != dataframe$id\_Currentowners[i+1]) {

x = x + 1

}

}

}

return(x)

}

# Numero di collaborazioni

n\_coll = collaboration(coll, x)

n\_coll

# Numero totale di brevetti senza NA

tot = 17364

tab = data.frame(rbind(c("Number of patent with multiple applicants", n\_coll, round(n\_coll/tot,2)),

c("Number of total patent", tot, tot/tot)))

colnames(tab) = c("Variable", "Number", "%")

tab

# ma la collaborazione è un fenomeno recente?

# filtro i dati dal 2010 in poi e costruisco la stessa tabella

six\_2 = six\_1 %>% filter(Publication\_year > 2009)

six\_b = six\_2 %>% count(Publication\_number, name = "CPC\_for\_Patent") # numero totale brevetti = 7014

six\_b = six\_b %>% filter(CPC\_for\_Patent > 1) # brevetti con più di una tecnologia al suo interno

# collaborazione se brevetto con più di una tecnologia prodotte da applicant diversi

coll2 = inner\_join(six\_2, six\_b, by = "Publication\_number")

# numero di collaborazioni dal 2010 in poi

n\_coll2 = collaboration(coll2, x)

n\_coll2

# Numero totale di brevetti senza NA dal 2010 in poi

tot1 = 7014

tab1 = data.frame(rbind(c("Number of patent with multiple applicants", n\_coll2, round(n\_coll2/tot1,2)),

c("Number of total patent", tot1, tot1/tot1)))

colnames(tab1) = c("Variable (Year > 2010)", "Number", "%")

tab1 # le collaborazioni sono un fenomeno più recente

# 7° - NUMBER OF INTERNATIONAL AND NATIONAL PATENT COLLABORATION

# funzione per contare collaborazioni nazionali ed internazionali nei brevetti

y = 0

z = 0

int\_nat = function(dataframe, y, z){

for (i in 1:(nrow(dataframe) - 1)){

if (dataframe$Publication\_number[i] == dataframe$Publication\_number[i+1]) {

if (dataframe$id\_Currentowners[i] != dataframe$id\_Currentowners[i+1]) {

if (dataframe$Currentownerscountrycode[i] == dataframe$Currentownerscountrycode[i+1]) {

y = y + 1

} else {z = z + 1}

}

}

}

return(a = c(y,z))

}

# Numero di collaborazioni nazionali e internazionali

n\_int\_nat = int\_nat(coll, y, z )

n\_int\_nat

tab2 = data.frame(rbind(c("National collaborations", n\_int\_nat[1], round(n\_int\_nat[1]/n\_coll, 2)),

c("International collaborations", n\_int\_nat[2], round(n\_int\_nat[2]/n\_coll, 2)),

c("Total number of collaborations", n\_coll, n\_coll/n\_coll)))

colnames(tab2) = c("Variable", "Number", "%")

tab2

# 8° - DISTRIBUTION OF FIRMS BY NUMBER OF PATENT

# creo grafico a barre in cui:

# ascisse = numero di brevetti

# ordinate = numero di imprese con quei brevetti

eight1 = data %>% filter(!is.na(id\_Currentowners)) %>%

distinct(Publication\_number, id\_Currentowners, CooperativeClassification, .keep\_all = TRUE) %>%

count(id\_Currentowners, name = "Patents", sort = TRUE)

eight = data.frame(table(eight1$Patents))

eight %>%

ggplot(aes(Var1, Freq)) +

geom\_col(fill = "red") +

labs(title = "Distribution of Firms by number of Patents",

x = "Number of Patents",

y = "Frequency") +

theme(axis.text.x = element\_text(angle=45, hjust = 1)) +

ggsave("Firms\_npatents8.png")

## DATA ELABORATION: FIRM LEVEL

# 9 ° - DISTRIBUTION OF PATENT BY APPLICANT

nine = data %>% drop\_na(Currentowners) %>%

distinct(Publication\_number, Currentowners, CooperativeClassification, .keep\_all = TRUE) %>%

count(Currentowners, name = "Number\_of\_Patents", sort = TRUE)

# funzione per aggiungere una colonna con il numero totale di brevetti ed una con le percentuali

add <- function(dataframe) {

dataframe = dataframe %>%

mutate(Tot = sum(Number\_of\_Patents)) %>%

mutate(Perc = Number\_of\_Patents/Tot)

return(dataframe)

}

# funzione per calcolare il Concentration Ratio Index

CR <- function(dataframe) {

dataframe <- head(dataframe, n=4)

cr <- sum(dataframe$Perc)

return(cr)

}

# 20 imprese con maggior numero di brevetti Y02W

nine = add(nine)

head(nine, 20)

# Concentration Ratio Index

Conc\_ind = CR(nine)

Conc\_ind

# 10° - DISTRIBUTION OF PATENT BY APPLICANT (DIFFERENT PERIODS)

# patents before 2000

ten\_1 = data %>% filter(Publication\_year < 2000) %>%

drop\_na(Currentowners) %>%

distinct(Publication\_number, Currentowners, CooperativeClassification, .keep\_all = TRUE) %>%

count(Currentowners, name = "Number\_of\_Patents", sort = TRUE)

ten\_1 = add(ten\_1)

# patents between 2000 and 2009

ten\_2 = data %>% filter(between(Publication\_year, 2000, 2009)) %>%

drop\_na(Currentowners) %>%

distinct(Publication\_number, Currentowners, CooperativeClassification, .keep\_all = TRUE) %>%

count(Currentowners, name = "Number\_of\_Patents", sort = TRUE)

ten\_2 = add(ten\_2)

# patents after 2009

ten\_3 = data %>% filter(Publication\_year > 2009) %>%

drop\_na(Currentowners) %>%

distinct(Publication\_number, Currentowners, CooperativeClassification, .keep\_all = TRUE) %>%

count(Currentowners, name = "Number\_of\_Patents", sort = TRUE)

ten\_3 = add(ten\_3)

# confronto leadership nel tempo

ten = cbind(head(ten\_1, 10), head(ten\_2, 10), head(ten\_3, 10))

View(ten)

# CR Index nel tempo

CR\_before = CR(ten\_1)

CR\_between = CR(ten\_2)

CR\_after = CR(ten\_3)

print(c(CR\_before, CR\_between, CR\_after))

# cambia anche l'indice di concentrazione nel tempo