Installation

devtools::install\_github("timat35/Rcan", ref = "dev", subdir="Rcan")

WARNING: Rtools is required to build R packages, but is not currently installed.

Please download and install Rtools 3.5 from http://cran.r-project.org/bin/windows/Rtools/.

\* DONE (Rcan)

In R CMD INSTALL

Warning messages:

1: In untar2(tarfile, files, list, exdir) :

skipping pax global extended headers

2: In untar2(tarfile, files, list, exdir) :

skipping pax global extended headers

data\_individual\_file:

* remplacer la variable “grade” par la variable “beh” du fichier de données originales : beaucoup plus parlant et utile pour les conversions ICD éventuellement à venir

ICD\_group\_file:

* remplacer « C44 – OTHER » par « C44 – SKIN »
* remplacer « C51 – OTHER » par « C51 – VULVA »
* remplacer « C52 – OTHER » par « C52 – VAGINA »
* remplacer « C60 – OTHER » par « C60 – PENIS »

Bon, ceci dit, je ne sais pas si on veut un fichier d’exemple… ou si on veut valider un truc passe-partout…

csu\_group\_cases:

* remplacer les age\_group\_labels « 0-4 »et « 5-9 » par « 00-04 » et « 05-09 »
* aucun label n’est alloué au groupe d’age 19 : « Unknown »
* renommer le paramètre « cross\_by » en « var\_age\_group » pour garder la même logique que pour la fonction csu\_asr ; aussi la même conrainte peut être exigée pour plus de clarté dans le Help file de la variable « group\_by », i.e. « Any variable included in the next option var\_age\_group must be also included here ».
* en sortie de fonction utilisée avec le groupement d’ICD, peux-tu stp rajouter ue mise à jour du nombre de niveaux pour le facteur LABEL (avec droplevels() par exemple) ? En effet, sur cet exemple, la commande

csu\_group\_cases(data\_individual\_file, var\_age="age",

cross\_by=c("sex", "regcode"),

group\_by=c("reglabel"),

df\_ICD = ICD\_group\_file,

var\_ICD ="site")

genère une table avec les 29 niveaux de LABEL alors que 22 seulement sont représentés.

data.table:

* les bases de données exemples « data\_individual\_file » et « ICD\_group\_file » sont de classe « data.table » (« data\_individual\_file » a la double nationalité data.frame, mais pas « ICD\_group\_file » qui n’est que de classe data.table) : cela peut jouer des tours lorsqu’on manipule ces fichiers (e.g. avec les fonctions basiques aggregate ou merge notamment, ou on est obligé de les convertir en data.frame) ; ce n’est pas le cas des bases « csu\_registry\_data\_1 » et « csu\_registry\_data\_2 ». Cela est vraisemblablement nécessaire pour les manipulations engendrées par la fonction (et probablement que cela n’embête que moi, donc pas de pb a priori ☺ )

Messages :

* La commande

csu\_group\_cases(data\_individual\_file, var\_age = "age",

cross\_by = c("sex", "regcode"),

group\_by = c("reglabel"),

var\_year = "doi")

provoque le message

Warning message:

In `[.data.table`(dt\_data, , `:=`(age\_group, cut(get(var\_age), c(seq(0, :

Invalid .internal.selfref detected and fixed by taking a (shallow) copy of the data.table so that := can add this new column by reference. At an earlier point, this data.table has been copied by R (or was created manually using structure() or similar). Avoid key<-, names<- and attr<- which in R currently (and oddly) may copy the whole data.table. Use set\* syntax instead to avoid copying: ?set, ?setnames and ?setattr. If this message doesn't help, please report your use case to the data.table issue tracker so the root cause can be fixed or this message improved.

* La fonction soulève un message d’erreur si df\_ICD est utilisée sans var\_ICD ; elle fait quelque chose lorsque var\_ICD est utilisée sans df\_ICD mais je ne sais pas trop quoi>

Help file

|  |  |
| --- | --- |
| csu\_group\_cases {Rcan} | R Documentation |

csu\_group\_cases

**Description**

csu\_group\_case groups individual data into 5 years age-group data and other user defined variable (sex, registry, etc...).  
Optionally: Group cancer based on a standard ICD10 coding; Extract year from custom year format.

**Usage**

csu\_group\_cases(df\_data,

var\_age ,

group\_by=NULL,

var\_age\_group = NULL,

var\_cases = NULL,

df\_ICD = NULL,

var\_ICD=NULL,

var\_year = NULL)

**Arguments**

|  |  |
| --- | --- |
| df\_data | Individual data (need to be R data.frame format, see example to import csv file). |
| var\_age | Age variable. (Numeric). Value > 150 will be consider as missing age. |
| cross\_by | (Optional) A vector of variables to create the all possible categories for each age group (and ICD and Year in option). For instance sex (1, 2) and registry code (regA, regB, regC) would create six categories:   |  |  | | --- | --- | | sex | country\_code | | 1 | regA | | 1 | regB | | 1 | regC | | 2 | regA | | 2 | regB | | 2 | regC | |  |  | |
| group\_by | (Optional) A vector of other relative variables to keep. For instance, sex label, cancer label (to avoid create false category such as "1 male", "1 female", etc..) |
| var\_cases | (Optional) cases variable: If there is already a variable for the number of cases. |
| df\_ICD | (Optional) ICD file for ICD grouping information. Must have 2 fields: "ICD", "LABEL" . 2 formats are possible: Each ICD code separate by ICD group   |  |  | | --- | --- | | ICD | LABEL | | C82 | NHL | | C83 | NHL | | C84 | NHL | | C85 | NHL | | C96 | NHL | |  |  |   ICD code already grouped.   |  |  | | --- | --- | | ICD\_group | LABEL | | C82-85,C96 | NHL | |  |  |   example: [ICD\_group\_file](http://127.0.0.1:44093/help/library/Rcan/help/ICD_group_file) |
| var\_ICD | (Optional) ICD variable: ICD variable in the individual data. Must be filled if df\_ICD parameter used |
| var\_year | (Optional) Year variable: Extract year from custom format ("yyyymmdd","ddmmyyy", "yyyy/mm","dd-mm-yyyy", etc..) and group data by year. |

**Details**

For most analysis, individual cases databases need to be grouped by category.  
This function groups data by 5 years age-group and other user defined variable.  
Next step will be to add 5 years population data.

**Value**

Return a dataframe.

**Author(s)**

Mathieu Laversanne

**See Also**

[csu\_ageSpecific](http://127.0.0.1:44093/help/library/Rcan/help/csu_ageSpecific) [csu\_asr](http://127.0.0.1:44093/help/library/Rcan/help/csu_asr) [csu\_time\_trend](http://127.0.0.1:44093/help/library/Rcan/help/csu_time_trend)[csu\_trendCohortPeriod](http://127.0.0.1:44093/help/library/Rcan/help/csu_trendCohortPeriod) [csu\_eapc](http://127.0.0.1:44093/help/library/Rcan/help/csu_eapc)

**Examples**

# you can import your data from csv file using read.csv:

# mydata <- read.csv("mydata.csv", sep=",")

data(ICD\_group\_file)

data(data\_individual\_file)

#group individual data by

# 5 year age group

df\_data\_age <- csu\_group\_cases(data\_individual\_file,

var\_age="age",

cross\_by=c("sex", "regcode", "site"),

group\_by=c("reglabel"))

#group individual data by

# 5 year age group

# ICD grouping from dataframe ICD\_group\_file

df\_data\_icd <- csu\_group\_cases(data\_individual\_file,

var\_age="age",

cross\_by=c("sex", "regcode"),

group\_by=c("reglabel"),

df\_ICD = ICD\_group\_file,

var\_ICD ="site")

#group individual data by

# 5 year age group

# ICD grouping from dataframe ICD\_group\_file

# year (extract from date of incidence)

df\_data\_year <- csu\_group\_cases(data\_individual\_file,

var\_age="age",

cross\_by=c("sex", "regcode"),

group\_by=c("reglabel"),

df\_ICD = ICD\_group\_file,

var\_ICD ="site",

var\_year = "doi")

# you can export your result as csv file using write.csv:

# write.csv(result, file="result.csv")

[Package *Rcan* version 1.3.65 [Index](http://127.0.0.1:44093/help/library/Rcan/html/00Index.html)]