

# Data cubes mit R erstellen

adminR: R in Swiss Official Statistics: First Meetup

Sandro Petrillo

USTAT, Ufficio di statistica del Canton Ticino

Bern, 20 September 2018



Qu'est-ce qu'on entend par cube de données

Créer un cube de données avec R

Que faire avec un cube? Quelques exemples



Qu'est-ce qu'on entend par cube de données



# Caractéristiques d'un cube de données dans la statistique publique

- ▶ Un conteneur de **résultats de statistique publique**
- ▶ Qui comprend toutes les combinaisons des modalités des variables (“croisements”), y compris la modalité “Total” de chaque variable
- ▶ Peut contenir une ou plusieurs statistiques pour chaque croisement (estimation de la numérosité, nombre d'entreprises, salaires médians, etc. . . )
- ▶ Contient uniquement les résultats de statistique publique qui peuvent être publiés (cette partie est facilement “automatisable”)
- ▶ Un cube de données est quelque chose entre des données individuelles (pour la richesse de son contenu) et des tableaux “standard” (parce qu'il peut être donné à quiconque sans problèmes de protection des données)

# Exemple: cubes de STAT-TAB (<https://www.pxweb.bfs.admin.ch/>)



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

## STAT-TAB - interactive tables (FSO)

[Home](#) >> [px-x-0302010000\\_108](#) >> Foreign cross-border commuters by gender, canton of work and age class

Foreign cross-border commuters by gender, canton of work and age class

[Information](#) [Footnotes](#)

For variables marked \* you need to select at least one value

Gender	Canton of work	Group of age	Quarter *
<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Total 3 Selected 0	Total 27 Selected 0	Total 11 Selected 0	Total 78 Selected 0
<div>Gender - Total</div> <div>Man</div> <div>Woman</div>	<div>Switzerland</div> <div>Zürich</div> <div>Bern / Berne</div> <div>Luzern</div> <div>Uri</div> <div>Schwyz</div> <div>Obwalden</div> <div>Nidwalden</div>	<div>Group of age - Total</div> <div>Age 15-19</div> <div>Age 20-24</div> <div>Age 25-29</div> <div>Age 30-34</div> <div>Age 35-39</div> <div>Age 40-44</div> <div>Age 45-49</div>	<div>2018Q2</div> <div>2018Q1</div> <div>2017Q4</div> <div>2017Q3</div> <div>2017Q2</div> <div>2017Q1</div> <div>2016Q4</div> <div>2016Q3</div>
Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>
<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row

Figure 1: Un exemple de cube sur STAT-TAB

Créer un cube de données avec R



## Question

- Est-ce que tout le monde connaît ce genre de code?

```
library(dplyr)
mtcars %>%
  group_by(cyl) %>%
  summarise(n = n(), mean_disp = mean(displacement))
```

```
## # A tibble: 3 x 3
##   cyl      n mean_disp
##   <dbl> <int>     <dbl>
## 1     4    11     105.
## 2     6     7     183.
## 3     8    14     353.
```

## Approche générale avec `distr`

- ▶ Le point de départ est toujours:
  - ▶ Une source de données dont on dispose des données individuelles (ESS, STATENT, ESPA, RS, ...) (argument `.data`, un `data.frame`)
  - ▶ Des variables catégorielles qui définissent les croisements qui nous intéressent (argument `.variables`, un vecteur de type `character`)
- ▶ Deux possibilités pour créer un cube:
  - ▶ avec la fonction `dcc5()`
  - ▶ avec la fonction `dcc6()`





## Un exemple avec le paquet `distr`

- On utilise les données `invented_wages` du paquet `distr`, qui contiennent 1.000 observations de salariés inventés.

```
library(distr)
head(invented_wages, n = 8)
```

```
## # A tibble: 8 x 5
##   gender sector    education wage sample_weights
##   <fct>  <fct>    <fct>    <dbl>         <dbl>
## 1 men    tertiary III      8400          105
## 2 women secondary II       4200           32
## 3 men    tertiary II       5100           36
## 4 women tertiary II       7400           12
## 5 men    secondary II       4300           21
## 6 men    secondary I       4900           46
## 7 men    tertiary III      5400           79
## 8 women secondary I       2900          113
```



## Avec la fonction dcc6()

```
library(distr)
# On crée une liste d'appel à deux fonctions
liste_fonctions <- list(
  salaire_moyen = ~weighted.mean(wage, sample_weights),
  p50 = ~wq(wage, sample_weights)
)

# Et on l'utilise comme argument .funs_list dans la fonction dcc6
dcc6(.data = invented_wages,
     .variables = c("gender", "sector"),
     .funs_list = liste_fonctions
)
```

► (résultat à la page suivante...)

## Résultat avec dcc6()

```
## # A tibble: 9 x 4
##   gender sector    salaire_moyen    p50
## * <fct>   <fct>          <dbl> <dbl>
## 1 Totale Totale          4645.  4300
## 2 Totale secondary      4444.  4100
## 3 Totale tertiary       4872.  4400
## 4 men      Totale          5323.  5100
## 5 men      secondary      5334.  5300
## 6 men      tertiary       5313.  4800
## 7 women    Totale          3614.  3400
## 8 women    secondary      3357.  3100
## 9 women    tertiary       4001.  3800
```

Que faire avec un cube? Quelques exemples

## Autres exemples

- ▶ [page html](#) avec les cubes créés par le secteur économie à l'Ustat (csv, xlsx, px, rda)
- ▶ exemple d'interface shiny pour l'exploitation des cubes  
([shiny::runApp\("examples/ridisponi3"\)](#))
- ▶ exemple d'exploitation d'un cube avec R (graphiques,  
[exemple\\_exploitation\\_cube.R](#))
- ▶ exemple de [document pdf](#) avec tableaux et graphiques ([source R markdown](#))

## Page “README” du paquet distr

<https://cran.r-project.org/web/packages/distr/README.html>

<https://github.com/gibonet/distr>

