

# From published tables to rtauargus input: an automated approach

C. Baudry and J. Jamme

03-05-24

# Formal description of a table

table\_name : indicator  $\otimes$  {grouping\_var\_1 x grouping\_var\_2}

Example :

T1: turnover\_pizzas  $\otimes$  {nuts2 x size}

	BE10	BE21	...	Total
wf1	10	8	...	50
wf2	10	12	...	50
Total	20	20	60	100

# List of published tables

- T1: to\_pizzas  $\otimes$  {nuts2 x size}
- T2: to\_pizzas  $\otimes$  {nuts3 x size}
- T3: to\_pizzas  $\otimes$  {a10 x nuts2}
- T4: to\_pizzas  $\otimes$  {a10 x nuts3}
- T5: to\_pizzas  $\otimes$  {a21 x nuts2}
- T6: to\_pizzas  $\otimes$  {a21 x nuts3}
- T7: to\_pizzas  $\otimes$  {a88 x nuts2}
- T8: to\_pizzas  $\otimes$  {a88 x nuts3}
- T9: to\_batavia  $\otimes$  {a10 x size}
- T10: to\_batavia  $\otimes$  {a10 x diversity}
- T11: to\_batavia  $\otimes$  {a21 x size}
- T12: to\_batavia  $\otimes$  {a21 x diversity}
- T13: to\_batavia  $\otimes$  {a88 x size}
- T14: to\_batavia  $\otimes$  {a88 x diversity}
- T15: to\_arugula  $\otimes$  {a10 x size}
- T16: to\_arugula  $\otimes$  {a10 x diversity}
- T17: to\_arugula  $\otimes$  {a21 x size}
- T18: to\_arugula  $\otimes$  {a21 x diversity}
- T19: to\_arugula  $\otimes$  {a88 x size}
- T20: to\_arugula  $\otimes$  {a88 x diversity}
- T21: to\_lettuce  $\otimes$  {a10 x size}
- T22: to\_lettuce  $\otimes$  {a10 x diversity}
- T23: to\_lettuce  $\otimes$  {a21 x size}
- T24: to\_lettuce  $\otimes$  {a21 x diversity}
- T25: to\_lettuce  $\otimes$  {a88 x size}
- T26: to\_lettuce  $\otimes$  {a88 x diversity}

# List of published tables

- T1: **to\_pizzas**  $\otimes$  {nuts2 x size}
- T2: **to\_pizzas**  $\otimes$  {nuts3 x size}
- T3: **to\_pizzas**  $\otimes$  {a10 x nuts2}
- T4: **to\_pizzas**  $\otimes$  {a10 x nuts3}
- T5: **to\_pizzas**  $\otimes$  {a21 x nuts2}
- T6: **to\_pizzas**  $\otimes$  {a21 x nuts3}
- T7: **to\_pizzas**  $\otimes$  {a88 x nuts2}
- T8: **to\_pizzas**  $\otimes$  {a88 x nuts3}
- T9: **to\_batavia**  $\otimes$  {a10 x size}
- T10: **to\_batavia**  $\otimes$  {a10 x diversity}
- T11: **to\_batavia**  $\otimes$  {a21 x size}
- T12: **to\_batavia**  $\otimes$  {a21 x diversity}
- T13: **to\_batavia**  $\otimes$  {a88 x size}
- T14: **to\_batavia**  $\otimes$  {a88 x diversity}
- T15: **to\_arugula**  $\otimes$  {a10 x size}
- T16: **to\_arugula**  $\otimes$  {a10 x diversity}
- T17: **to\_arugula**  $\otimes$  {a21 x size}
- T18: **to\_arugula**  $\otimes$  {a21 x diversity}
- T19: **to\_arugula**  $\otimes$  {a88 x size}
- T20: **to\_arugula**  $\otimes$  {a88 x diversity}
- T21: **to\_lettuce**  $\otimes$  {a10 x size}
- T22: **to\_lettuce**  $\otimes$  {a10 x diversity}
- T23: **to\_lettuce**  $\otimes$  {a21 x size}
- T24: **to\_lettuce**  $\otimes$  {a21 x diversity}
- T25: **to\_lettuce**  $\otimes$  {a88 x size}
- T26: **to\_lettuce**  $\otimes$  {a88 x diversity}

# List of published tables

- T1: to\_pizzas  $\otimes$  { nuts2 x size}
- T2: to\_pizzas  $\otimes$  { nuts3 x size}
- T3: to\_pizzas  $\otimes$  { a10 x nuts2}
- T4: to\_pizzas  $\otimes$  { a10 x nuts3}
- T5: to\_pizzas  $\otimes$  { a21 x nuts2}
- T6: to\_pizzas  $\otimes$  { a21 x nuts3}
- T7: to\_pizzas  $\otimes$  { a88 x nuts2}
- T8: to\_pizzas  $\otimes$  { a88 x nuts3}
- T9: to\_batavia  $\otimes$  { a10 x size}
- T10: to\_batavia  $\otimes$  { a10 x diversity}
- T11: to\_batavia  $\otimes$  { a21 x size}
- T12: to\_batavia  $\otimes$  { a21 x diversity}
- T13: to\_batavia  $\otimes$  { a88 x size}
- T14: to\_batavia  $\otimes$  { a88 x diversity}
- T15: to\_arugula  $\otimes$  { a10 x size}
- T16: to\_arugula  $\otimes$  { a10 x diversity}
- T17: to\_arugula  $\otimes$  { a21 x size}
- T18: to\_arugula  $\otimes$  { a21 x diversity}
- T19: to\_arugula  $\otimes$  { a88 x size}
- T20: to\_arugula  $\otimes$  { a88 x diversity}
- T21: to\_lettuce  $\otimes$  { a10 x size}
- T22: to\_lettuce  $\otimes$  { a10 x diversity}
- T23: to\_lettuce  $\otimes$  { a21 x size}
- T24: to\_lettuce  $\otimes$  { a21 x diversity}
- T25: to\_lettuce  $\otimes$  { a88 x size}
- T26: to\_lettuce  $\otimes$  { a88 x diversity}

# List of tables to protect

- T1\_T3:  $\text{to\_pizzas} \otimes \{ \text{HRC\_NUTS} \times \text{size} \}$
- T\_rest:  $\text{to\_pizzas} \otimes \{ \text{HRC\_NAF} \times \text{HRC\_NUTS} \}$
- T\_odd:  $\text{to\_lettuce} \otimes \{ \text{HRC\_NAF} \times \text{size} \times \text{HRC\_lettuce}^h \}$
- T\_even:  $\text{to\_lettuce} \otimes \{ \text{HRC\_NAF} \times \text{diversity} \times \text{HRC\_lettuce}^h \}$

With  $\text{HRC\_lettuce}^h$  a holding variable.

# Analysis automation steps

1. The user enters the metadata for the tables to be published in the required format.

Then the program:

1. Identifies hierarchies and renames variables accordingly
2. Breaks down the request into independent sub-requests (clusters)
3. Detects overlapping tables
4. Creates inclusion graphs
5. Groups tables included in each other into a single table
6. Creates a summary of the tables needing protection

# Metadata file

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	tableau	champ.population	hrc_champ.population	champ.date	hrc_champ.date	champ.lieu	hrc_champ.lieu	variable_interet	metrique	hrc_indicateur	croisement	hrc_croisement	totcode
2	T1	entreprises		2023		france		pizzas	ca		nuts2	hrc_nuts	total
3	T1	entreprises		2023		france		pizzas	ca		treff		total
4	T2	entreprises		2023		france		pizzas	ca		nuts3	hrc_nuts	total
5	T2	entreprises		2023		france		pizzas	ca		treff		total
6	T3	entreprises		2023		france		pizzas	ca		a10	hrc_naf	total
7	T3	entreprises		2023		france		pizzas	ca		nuts2	hrc_nuts	total
8	T4	entreprises		2023		france		pizzas	ca		a10	hrc_naf	total
9	T4	entreprises		2023		france		pizzas	ca		nuts3	hrc_nuts	total

tableau
T1
T1
T2
T2
T9
T9

variable_interet	metrique	hrc_indicateur
pizzas	ca	
pizzas	ca	
pizzas	ca	
pizzas	ca	
batavia	ca	hrc_salades
batavia	ca	hrc_salades

croisement	hrc_croisement
nuts2	hrc_nuts
treff	
nuts3	hrc_nuts
treff	
a10	hrc_naf
treff	



# Metadata

```
dfMetadatalettuce <- read_ods(file.path(
  rep_metadata,"metadata_salades_for_nested_totcode_champs.ods"))
str(dfMetadatalettuce)
```

```
## tibble [50 × 13] (S3: tbl_df/tbl/data.frame)
## $ tableau          : chr [1:50] "T1" "T1" "T2" "T2" ...
## $ champ.population : chr [1:50] "entreprises" "entreprises" "entreprises" "entreprises" ...
## $ hrc_champ.population: logi [1:50] NA NA NA NA NA NA ...
## $ champ.date        : num [1:50] 2023 2023 2023 2023 2023 ...
## $ hrc_champ.date    : logi [1:50] NA NA NA NA NA NA ...
## $ champ.lieu        : chr [1:50] "france" "france" "france" "france" ...
## $ hrc_champ.lieu    : logi [1:50] NA NA NA NA NA NA ...
## $ variable_interet  : chr [1:50] "pizzas" "pizzas" "pizzas" "pizzas" ...
## $ metrique          : chr [1:50] "to" "to" "to" "to" ...
## $ hrc_indicateur    : chr [1:50] NA NA NA NA ...
## $ croisement        : chr [1:50] "nuts2" "size" "nuts3" "size" ...
## $ hrc_croisement    : chr [1:50] "hrc_nuts" NA "hrc_nuts" NA ...
## $ totcode           : chr [1:50] "total" "total" "total" "total" ...
```

# Metadata

```
dfMetadata$lettuce %>%  
  filter(tableau %in% c("T1", "T2", "T9")) %>%  
  select(tableau, variable_interet, metrique, hrc_indicateur, croisement, hrc_croisement)
```

```
## # A tibble: 6 × 6
```

```
##   tableau variable_interet metrique hrc_indicateur croisement hrc_croisement  
##   <chr>    <chr>          <chr>    <chr>          <chr>    <chr>  
## 1 T1      pizzas        to      <NA>          nuts2    hrc_nuts  
## 2 T1      pizzas        to      <NA>          size     <NA>  
## 3 T2      pizzas        to      <NA>          nuts3    hrc_nuts  
## 4 T2      pizzas        to      <NA>          size     <NA>  
## 5 T9      batavia       to      hrc_salades    a10      hrc_naf  
## 6 T9      batavia       to      hrc_salades    size     <NA>
```

# Identify hierarchies

```
dfHrcIdentlettuce <- identify_hrc(dfMetadatalettuce)
dfHrcIdentlettuce %>%
  filter(tableau %in% c("T1", "T2", "T9")) %>%
  select(tableau, indicateur, hrc_indicateur, croisement, hrc_croisement)
```

```
## # A tibble: 7 × 5
```

```
## # Rowwise:
```

```
##   tableau indicateur hrc_indicateur croisement hrc_croisement
##   <chr>    <chr>      <lgl>          <chr>      <chr>
## 1 T1      to_pizzas  NA              HRC_NUTS   hrc_nuts
## 2 T1      to_pizzas  NA              size       <NA>
## 3 T2      to_pizzas  NA              HRC_NUTS   hrc_nuts
## 4 T2      to_pizzas  NA              size       <NA>
## 5 T9      to_salades NA              HRC_NAF     hrc_naf
## 6 T9      to_salades NA              size       <NA>
## 7 T9      to_salades NA              HRC_SALADES hrc_salades
```

# Changes generated by hierarchy identification

- T1: to\_pizzas  $\otimes$  {nuts2 x size}
- T2: to\_pizzas  $\otimes$  {nuts3 x size}
- T9: to\_batavia  $\otimes$  {a10 x size}
- ...

- T1: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T2: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T9: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- ...

# Split in clusters

Split the list of tables in independant clusters, i.e. linked tables clusters.

Independant tables do not need to be treated together. Call Tau-Argus multiple times independently.

```
lSplitlettuce <- split_in_clusters(dfHrcIdentlettuce)
```

```
names(lSplitlettuce)
```

```
## [1] "to_pizzas" "to_salades"
```

# Which tables are inside each cluster?

- to\_pizzas : T1, T2, T3, T4, T5, T6, T7, T8
- to\_lettuce : T9, T10, T11, T12, T13, T14, T15, T16, T17, T18, T19, T20, T21, T22, T23, T24, T25, T26

# Detection of tables included in others

```
lDescLienlettuce <- create_edges(lSplitlettuce)
```

Cluster to\_pizzas :

```
## NULL
```

Cluster to\_lettuce :

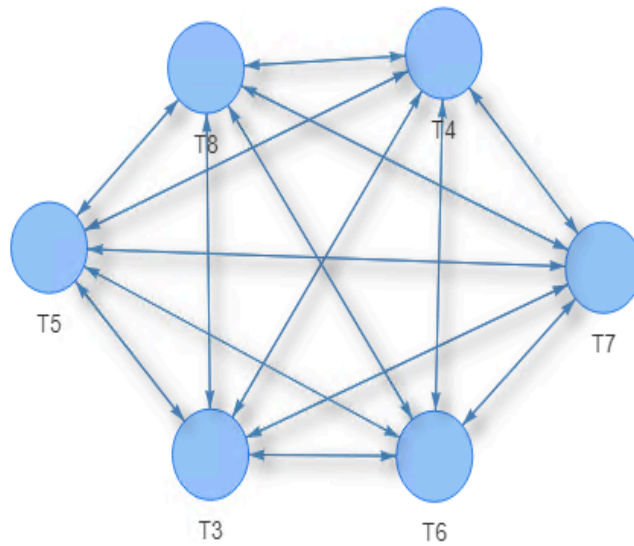
```
## NULL
```

# Inclusion graphs

graph\_links\_tab(lDescLienlettuce)

Select by id ▼

**pizzas**



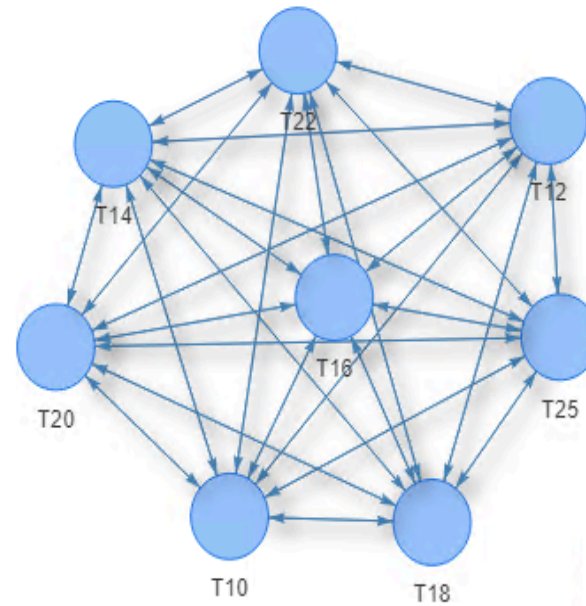
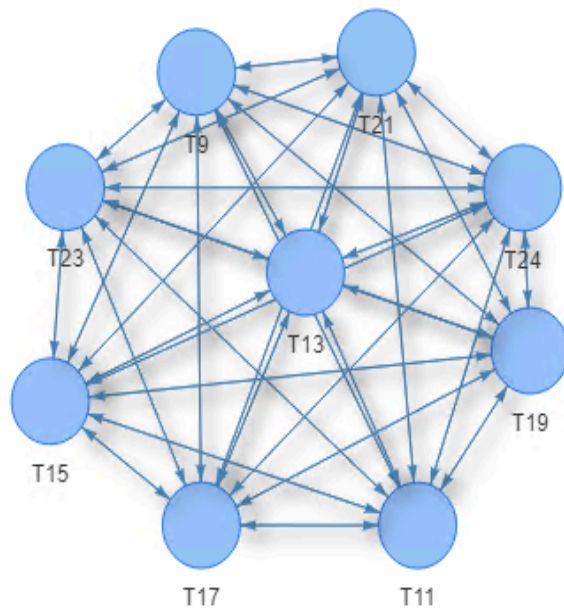


# Inclusion graphs

graph\_links\_tab(lDescLienlettuce)

Select by id ▼

**hrc\_salades**



# Group tables based on links

lTabIndep\$ca\_pizzas

## NULL

lTabIndep\$ca\_salades

## NULL

# Result of table grouping

- to\_pizzas : T1, T2, T3, T4, T5, T6, T7, T8
- to\_lettuce : T9, T10, T11, T12, T13, T14, T15, T16, T17, T18, T19, T20, T21, T22, T23, T24, T25, T26

- to\_pizzas : T1\_T2, T3\_T4\_T5\_T6\_T7\_T8
- to\_lettuce : T10\_T12\_T14\_T16\_T18\_T20\_T22\_T24\_T26,  
T9\_T11\_T13\_T15\_T17\_T19\_T21\_T23\_T25

# R output creation

```
lTabTreat <- tab_to_treat(lTabIndep)
```

```
## $to_pizzas
```

```
## # A tibble: 2 × 7
```

```
## # Groups:   tableau [2]
```

```
##   tableau      champ      indicateur croisement_1 croisement_2 hrc_croisement_1 hrc_croisement_2
```

```
##   <chr>      <chr>      <chr>      <chr>      <chr>      <chr>      <chr>
```

```
## 1 T1_T2      entreprises2023france to_pizzas HRC_NUTS      size      hrc_nuts      <NA>
```

```
## 2 T3_T4_T5_T6_T7_T8 entreprises2023france to_pizzas HRC_NAF      HRC_NUTS      hrc_naf      hrc_nuts
```

```
##
```

```
## $to_salades
```

```
## # A tibble: 2 × 9
```

```
## # Groups:   tableau [2]
```

```
##   tableau      champ      indicateur croisement_1 croisement_2 croisement_3 hrc_croisement_1
```

```
##   <chr>      <chr> <chr>      <chr>      <chr>      <chr>      <chr>
```

```
## 1 T10_T12_T14_T16_T18_T20_T22... entr... to_salades HRC_NAF      diversity HRC_SALADES hrc_naf
```

```
## 2 T11_T13_T15_T17_T19_T21_T23... entr... to_salades HRC_NAF      size      HRC_SALADES hrc_naf
```

# .csv output creation

One .csv file for each cluster:

```
output_csv(lTabTreat, "resultats/totcode/res_analyse_lettuce")
```

# Summary

From the **26 published tables** defined in the metadata file, the program suggests to **protect 4 rearranged tables**.

For this particular example `rtauargus::tab_multi_manager()` would be called twice:

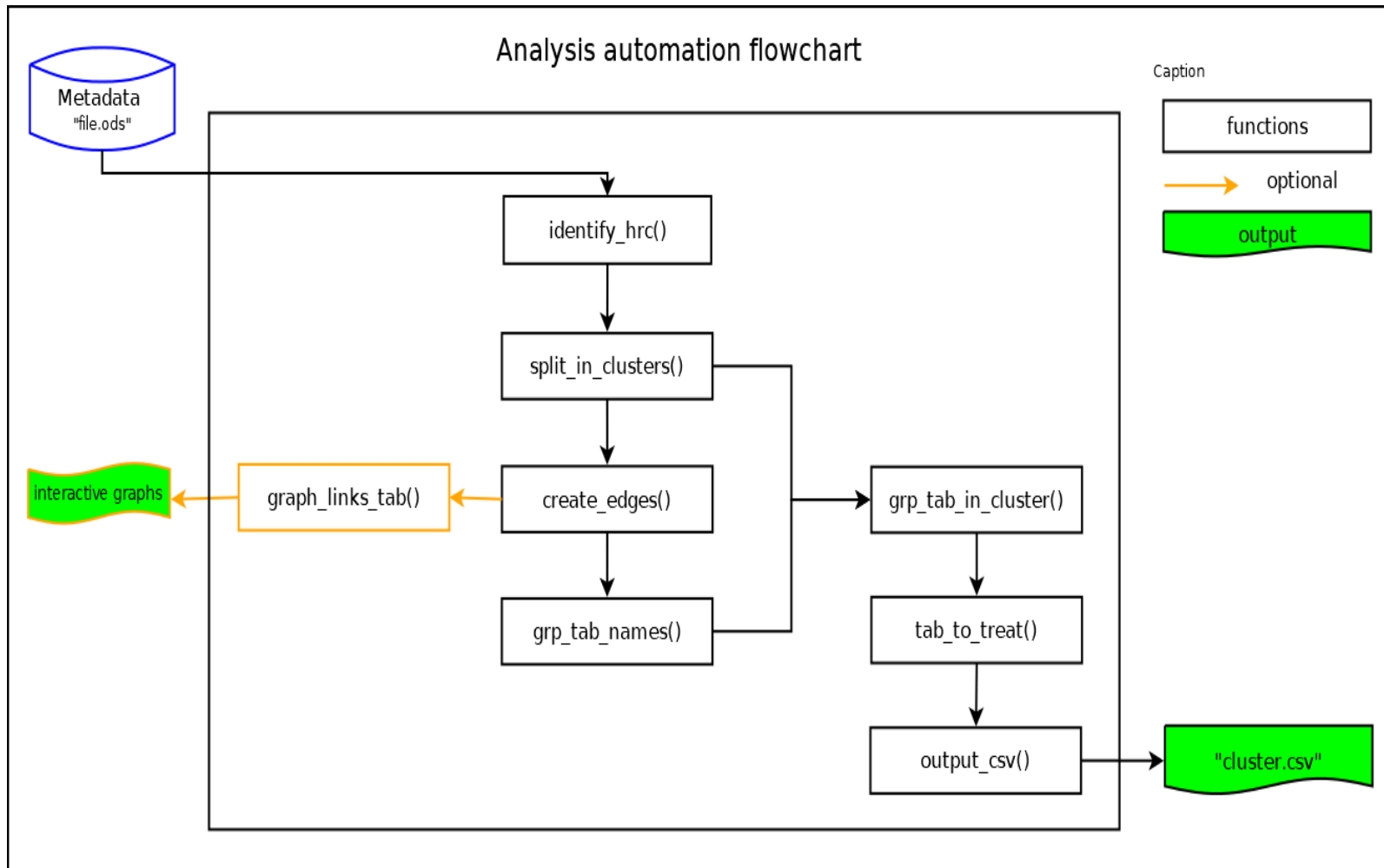
- Once for the pizza turnover tables
- Once for the salad turnover tables

# Further work

- Keep on testing the program on different lists of tables, especially for the non-nested hierarchies option recently added
- Check the hierarchies provided (nested, non-nested)
- Automatically generate the input tables for `rtauargus` functions `tab_rtauargus()` and `tab_multi_manager()` using the output of this analysis program

Code on gitlab : [repo git](#)

# Appendix 1 : program flowchart





# Appendix 2 : step-by-step table changes induced by the program

- T1: to\_pizzas  $\otimes$  {nuts2 x size}
- T2: to\_pizzas  $\otimes$  {nuts3 x size}
- T3: to\_pizzas  $\otimes$  {a10 x nuts2}
- T4: to\_pizzas  $\otimes$  {a10 x nuts3}
- T5: to\_pizzas  $\otimes$  {a21 x nuts2}
- T6: to\_pizzas  $\otimes$  {a21 x nuts3}
- T7: to\_pizzas  $\otimes$  {a88 x nuts2}
- T8: to\_pizzas  $\otimes$  {a88 x nuts3}
- T9: to\_batavia  $\otimes$  {a10 x size}
- T10: to\_batavia  $\otimes$  {a10 x diversity}
- T11: to\_batavia  $\otimes$  {a21 x size}
- T12: to\_batavia  $\otimes$  {a21 x diversity}
- T13: to\_batavia  $\otimes$  {a88 x size}
- T14: to\_batavia  $\otimes$  {a88 x diversity}
- T15: to\_arugula  $\otimes$  {a10 x size}
- T16: to\_arugula  $\otimes$  {a10 x diversity}
- T17: to\_arugula  $\otimes$  {a21 x size}
- T18: to\_arugula  $\otimes$  {a21 x diversity}
- T19: to\_arugula  $\otimes$  {a88 x size}
- T20: to\_arugula  $\otimes$  {a88 x diversity}
- T21: to\_lettuce  $\otimes$  {a10 x size}
- T22: to\_lettuce  $\otimes$  {a10 x diversity}
- T23: to\_lettuce  $\otimes$  {a21 x size}
- T24: to\_lettuce  $\otimes$  {a21 x diversity}
- T25: to\_lettuce  $\otimes$  {a88 x size}
- T26: to\_lettuce  $\otimes$  {a88 x diversity}

# Appendix 2 : step-by-step table changes induced by the program

- T1: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T2: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T3: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T4: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T5: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T6: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T7: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T8: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T9: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T10: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T11: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T12: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T13: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T14: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T15: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T16: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T17: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T18: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T19: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T20: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T21: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T22: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T23: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T24: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T25: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T26: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}

# Appendix 2 : step-by-step table changes induced by the program

## \$to\_pizzas

- T1: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T2: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T3: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T4: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T5: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T6: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T7: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}
- T8: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}

## \$to\_lettuce

- T9: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T10: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T11: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T12: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T13: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T14: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T15: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T16: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T17: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T18: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T19: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T20: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T21: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T22: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T23: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T24: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}
- T25: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T26: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}

# Appendix 2 : step-by-step table changes induced by the program

Extrait des 32 lignes :

```
## NULL
```

Extrait des 128 lignes :

```
## NULL
```

# Appendix 2 : step-by-step table changes induced by the program

## \$to\_pizzas

- T1\_T2: to\_pizzas  $\otimes$  {HRC\_NUTS x size}
- T3\_T4\_T5\_T6\_T7\_T8: to\_pizzas  $\otimes$  {HRC\_NAF x HRC\_NUTS}

## \$to\_lettuce

- T9\_T11\_T13\_T15\_T17\_T19\_T21\_T23\_T25: to\_lettuce  $\otimes$  {HRC\_NAF x size x HRC\_lettuce}
- T10\_T12\_T14\_T16\_T18\_T20\_T22\_T24\_T26: to\_lettuce  $\otimes$  {HRC\_NAF x diversity x HRC\_lettuce}