

Packaging data exploration, cleaning and visualization

This is a collaborative document to:

- Document how to access packaging data collected on Open Food Facts
- List interesting questions that we would like to explore
- Share ideas and instructions on how to answer those questions
- Create data visualizations
- Document our actions (e.g. for data cleaning)

Note: there is no expiration date on the usefulness of this document, but **we have a first deadline to create a first set of learnings / visualizations**, as we will have a webinar to present the results of the French [Plein pot sur les emballages](#) ([Tackling packaging](#)) operation conducted with ADEME in September 2023. It would be fantastic to have interesting results at least 2 weeks before!

If you are interested in helping to explore, clean and visualize the data, you can also join the #packaging channel on the [Open Food Facts Slack](#), and participate in the [weekly packagings community call](#).

This document is in English in order to be useful to as many people as possible. To get results for a particular country or language, you can replace “world” in [world.openfoodfacts.org](#) with your country code (e.g. “fr” for France: [fr.openfoodfacts.org](#)).

Packaging data structure

Packaging components (packagings field)

For each product, we record a list of its packaging components, with data for each component:

- Number of units
- Shape
- Material
- Recycling instruction
- Weight of an empty unit
- Contained quantity

Packaging parent materials (packagings_materials field)

For each product, we aggregate the packaging components by parent material: plastics, glass, metal, paper/cardboard, other materials, and all.

Packaging taxonomies: shapes, materials, recycling

How to access packaging data

Data for one product - Website

Packaging data is displayed at the end of each product page on the Open Food Facts website.

Example: [Cristaline water](#)

The product page only shows data we have. To see the full data structure, including empty fields, you can edit the product page.

Number of units	Shape	Material	Recycling instruction	Weight of one empty unit (g)	Net quantity of product contained per unit
1	Bottle cap x ▾	HDPE 2 -High-density p... x ▾	Recycle x ▾	1.2	
1	Bottle x ▾	PET-Transparent x ▾	Recycle x ▾	20.24	1,5 L
1	Label x ▾	Plastic x ▾	Recycling instruction ▾	0.44	
	Shape ▾	Material ▾	Recycling instruction ▾		

All the packaging parts of the product are listed.

Data for one product - API

Packaging data for one product is accessible through the Open Food Facts API, in the “packagings” field.

Example: <https://world.openfoodfacts.org/api/v3/product/3274080005003?fields=packagings>

```
packagings:
[
  {
    material:
    {
      id: "en:hdpe-2-high-density-polyethylene"
    },
    number_of_units: 1,
```

```

    recycling:
    {
        id: "en:recycle"
    },
    shape:
    {
        id: "en:bottle-cap"
    },
    weight_measured: 1.2
},
{
    material:
    {
        id: "en:pet-transparent"
    },
    number_of_units: 1,
    quantity_per_unit: "1,5 L",
    quantity_per_unit_unit: "l",
    quantity_per_unit_value: 1.5,
    recycling:
    {
        id: "en:recycle"
    },
    shape:
    {
        id: "en:bottle"
    },
    weight_measured: 20.24
},
{
    material:
    {
        id: "en:plastic"
    },
    number_of_units: 1,
    shape:
    {
        id: "en:label"
    },
    weight_measured: 0.44
}
]

```

Example for the `packagings_materials` field:

https://world.openfoodfacts.org/api/v3/product/3274080005003?fields=packagings_materials

```
- packagings_materials: {
  - all: {
    weight: 21.88,
    weight_100g: 1.458666666666667,
    weight_percent: 100
  },
  - en:plastic: {
    weight: 21.88,
    weight_100g: 1.458666666666667,
    weight_percent: 100
  }
}
```

Data for multiple products - Facets

Each product in Open Food Facts can have tags of different types (e.g. categories, labels), that we call facets.

The Open Food Facts website allows to:

- See the list of tags of a given type (e.g. list all labels: <https://world.openfoodfacts.org/labels>), with a count of products for each tag
- See the list of products that have a specific tag (e.g. list all products with the organic label: <https://world.openfoodfacts.org/label/organic>)
- Combine up to 2 tags (e.g. list of categories)

Note: exploring data with facet can timeout, especially with the new packaging facets that are not indexed. It's best to restrict the number of products by using a country prefix (e.g. fr instead of world) and a filter on a category (e.g. /category/yogurt) or another facet (e.g. products with weights: /misc/packagings-with-weights)

We have added several tag types and tags related to packaging data:

Misc tags related to packaging data

Packaging with weights

<https://world.openfoodfacts.org/misc/packagings-with-weights>

Packaging shapes

Use &no_cache=1 in order to get results, as those new facets are not yet fully available.

- https://world.openfoodfacts.org/misc/packagings-with-weights/packaging-shapes&no_cache=1
- https://world.openfoodfacts.org/category/yogurts/packaging-shapes&no_cache=1

Packaging materials

Use

&no_cache=1 in order to get results, as those new facets are not yet fully available.

- https://world.openfoodfacts.org/misc/packagings-with-weights/packaging-materials&no_cache=1
- https://world.openfoodfacts.org/category/yogurts/packaging-materials&no_cache=1

Packaging recycling

Use &no_cache=1 in order to get results, as those new facets are not yet fully available.

- https://world.openfoodfacts.org/misc/packagings-with-weights/packaging-recycling&no_cache=1
- https://world.openfoodfacts.org/category/yogurts/packaging-recycling&no_cache=1

Weighers

People who have added packaging data with weights:

- <http://world.openfoodfacts.org/weighers>

Data for multiple products - CSV export

We have exported packaging data for all products in CSV format, with 1 line for each packaging component of each product:

- All products: <https://world.openfoodfacts.org/data/packagings.all.csv>
- Products that have packaging weights: <https://world.openfoodfacts.org/data/packagings.packagings-with-weights.csv>

Fields:

- code: product barcode
- countries_tags: comma separated list of countries where the product is sold
- categories_tags: comma separated list of categories for the product
- number_of_units
- shape
- material
- recycling
- weight_measured
- weight_specified
- quantity_per_unit

Notes:

- For shape, material and recycling, only the most specific corresponding entry in the shapes, materials and recycling taxonomies is listed.
 - e.g. en:pet-1-polyethylene-terephthalate

Link to CSV for product data (e.g. to get quantity for a specific barcode)

<https://world.openfoodfacts.org/data>

Data for multiple products - JSON aggregation per country, category, and parent material

To help explore packaging data per country, per category, per shape, and/or per material, we have aggregated data stored in JSON files.

Aggregation counts are stored in a structure of the form:

```
{
  countries => {
    "en:world" => ..
    "en:france" => {
      categories => {
        "all" => .. # stats for all categories
        "en:yogourts" => {
          shapes => {
            "en:unknown" => ..
            "all" => .. # stats for all shapes
            ..
            "en:bottle" => {
              materials_parents => .. # stats for parents materials (e.g. PET will also count for
plastic)
              materials => {
                "all" => ..
                "en:plastic" => 12, # number of products sold in France that are yogurts and that
have a plastic bottle packaging component
              }
            },
            ..
          },
          ..
        },
        ..
      }
    },
    ..
  }
}
```

Notes:

- If you don't want to filter by a country, a category, a shape or a material, use the "all" entry

- shapes and materials list the most specific value that the product component has: a component with a PET material will not be counted in the plastics entry
- shapes_parents and materials_parents include parent values: a component with a PET material will also be counted in the plastics entry

Files:

- Small subset (filter on country France + category “fermented dairy desserts” which includes yogurts)
 - Products with weights
 - “Small” file, can be explored with a browser with a JSON viewer extension
 - https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.fr.fermented-dairy-desserts.packagings-with-weights.json
 - All products
 - https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.fr.fermented-dairy-desserts.all.json
- Full data (all countries and categories, too big to explore in a browser)
 - Products with weights
 - https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.packagings-with-weights.json
 - All products
 - https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.all.json

Exploring JSON in a browser:

```

{
  n: "363",
  - shapes: {
    + all: { ... },
    + en:backing: { ... },
    + en:basket: { ... },
    + en:bottle: { ... },
    + en:bottle-cap: { ... },
    + en:box: { ... },
    + en:can: { ... },
    + en:card: { ... },
    + en:container: { ... },
    + en:fastener: { ... },
    + en:film: { ... },
    + en:grouping-package: { ... },
    - en:individual-pot: {
      - materials: {
        + all: { ... },
        - en:cardboard: {
          n: "4",
          - weights: {
            mean: 9.435,
            n: 4,
            sum: 37.74,
          - values: [
              10.39,
              10.19,
              10.44,
              6.72
            ]
          }
        },
        + en:clear-glass: { ... },
        + en:paper: { ... },
        + en:pla-polylactic-acid: { ... },
        - en:plastic: {
          n: "10",
          - weights: {
            mean: 4.46111,
            n: 9,
            sum: 40.15,
          - values: [
              4.1,

```



```

{
  n: "363",
  - shapes: {
    + all: { ... },
    + en:backing: { ... },
    + en:basket: { ... },
    + en:bottle: { ... },
    + en:bottle-cap: { ... },
    + en:box: { ... },
    + en:can: { ... },
    + en:card: { ... },
    + en:container: { ... },
    + en:fastener: { ... },
    + en:film: { ... },
    + en:grouping-package: { ... },
    - en:individual-pot: {
      - materials: {
        + all: { ... },
        - en:cardboard: {
          n: "4",
          - weights: {
            mean: 9.435,
            n: 4,
            sum: 37.74,
            - values: [
              10.39,
              10.19,
              10.44,
              6.72
            ]
          }
        },
        + en:clear-glass: { ... },
        + en:paper: { ... },
        + en:pla-polylactic-acid: { ... },
        - en:plastic: {
          n: "10",
          - weights: {
            mean: 4.46111,
            n: 9,
            sum: 40.15,
            - values: [
              4.1,

```

Available files with packaging data (JSON)

Files can be prefixed with https://world.openfoodfacts.org/data/categories_stats/ to retrieve them from the web.

```

/srv/off/html/data/categories_stats# ls -lrt
total 440896

```

-rw-r--r-- 1 off off 92977347 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.packagings-with-weights.json
-rw-r--r-- 1 off off 14788414 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_materials_stats.packagings-with-weights.json
-rw-r--r-- 1 off off 91743 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.fr.fermented-dairy-desserts.packagings-with-weights.json
-rw-r--r-- 1 off off 4517177 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.packagings-with-weights.popular.json
-rw-r--r-- 1 off off 3537135 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_materials_stats.packagings-with-weights.popular.json
-rw-r--r-- 1 off off 9665 Jul 17 20:21
https://world.openfoodfacts.org/data/categories_stats/categories_packagings_stats.fr.fermented-dairy-desserts.packagings-with-weights.popular.json
-rw-r--r-- 1 off off 288323285 Jul 17 22:34 categories_packagings_stats.all.json
-rw-r--r-- 1 off off 21845843 Jul 17 22:34 categories_packagings_materials_stats.all.json
-rw-r--r-- 1 off off 50777 Jul 17 22:34
categories_packagings_stats.fr.fermented-dairy-desserts.all.json
-rw-r--r-- 1 off off 13986955 Jul 17 22:34 categories_packagings_stats.all.popular.json
-rw-r--r-- 1 off off 11320233 Jul 17 22:34
categories_packagings_materials_stats.all.popular.json
-rw-r--r-- 1 off off 4722 Jul 17 22:34
categories_packagings_stats.fr.fermented-dairy-desserts.all.popular.json

Available files with packaging data (CSV)

Files can be prefixed with <https://world.openfoodfacts.org/data/> to retrieve them from the web.

```
/srv/off/html/data# wc -l packagings.*csv
```

```
448666 https://world.openfoodfacts.org/data/categories\_stats/packagings.all.csv  
26293
```

```
https://world.openfoodfacts.org/data/categories\_stats/packagings.packagings-with-weights.c  
SV
```

Exploration

Interesting questions / ideas of visualisation

- **Vincent Colomb - What are the product categories that have the smallest product weight / packaging weight ratio?**
 - **Plastic vs other materials**
- **Vincent Colomb - What are the product categories where there is a lot of variability?**
 - **Dispersion graphs**
- **Gala - In category (x) the percentages of different types of materials used : bouteille de lait, yaourt, biscuits sec**
- Gala - What does that represent in terms of impact (CO2) ?
- The part of “more ecological” packaging solutions (bonnes pratiques)
- Manon - In category (x) : get the top 3 best and worst product' ratio of material weight per product weight
 - Meaning “10 g of plastic for 125 g of product”
- Manon - Top 5 products with maximalist packaging (the highest number of elements)
- Manon - Part of packaging “to recycle” vs “to discard” relatively to the nb of units
- Manon - Part of packaging “to recycle” vs “to discard” relatively to their weights
- Sarazine - Is there a correlation between the packaging type and the type of food product?
- Sarazine - What is the percentage of food products which do not fit into the “packaging norm” of their category?
- Alex G - Bar graph: how much plastic for 100g of yogurt, pizza, milk, etc.
- Alex G - Same for metal or glass for beer, wine, tuna, etc.
- Alex G - What is the % difference in plastic / 100g for various categories (to see where there is more spread) - could be a boxplot, but not sure it's very readable for a large public
- Alex G - Repartition by matter by categories, (but instead of weight which is misleading for plastic I don't know if we could have an impact)
- Alex G - Repartition of plastics types
- Vincent - Combien de types de plastiques ont été remontés (+ répartition)
- Virginia - Nombre de produits, d'unités, et poids total par catégories et top sous-catégories
- Virginia - Top catégories, formes et consignes de tri par matières
- Virginia - Nombre moyen de matériaux par produit
- Virginia - Poids total, moyen, + lourd et + léger
- Virginia - Zoom sur les top catégories (2 listes, 1 redondante dans les articles, l'autre du Ministère de l'Agriculture, est-ce que ça matche avec vos données ?) :
 - pâtes, riz, pain, fromage, lait, plats préparés, pomme de terre préparées, sodas, cafés, alcools
 - céréales, viandes, lait/yaourts, pommes de terre (frais/préparées), vin, sucre, fromage, tomates (frais/préparées), pommes (frais/préparées), oeufs, riz
- Virginia - Zoom sur les top produits dont on a les données ? (pour commencer à montrer l'impact de la consommation actuelle des français)

- Classement de NielsenIQ au 23/04/23 : 6 x 1,5L d'eau Cristaline, 1L de Ricard, 1,75L de Coca Cola, 20x25cl de Heineken, 6x33 cl de Coca Cola, 6x1,5L Volvic, 6x1L Hépar, 20x25cl en bouteilles Desperados, Whisky William Peel, pot 1kg et 825g de Nutella, Ferrero Rocher, fromage Caprice des dieux 300g, beurre doux Président 250g, café Carte Noire 3x250g
- Vincent - analyser les emballages par type d'usage (nomade, domicile, mixte ..). Si on peut trouver une liste des produits régulièrement retrouvés dans la nature (Romain Tramois) / jeu de données sur collecte déchets des plages.
- Croiser avec les données de ventes pour avoir le tonnage de plastique.
 - Vincent se renseigne

Ideas for displaying data:

- Display the best and worst performers in a category (e.g. water bottles).
- Pie chart with facets
 - Shapes
 - Materials represented
 - Sorting instructions
- Find good performers with materials that have less impact than plastics (lasagne wrapped in glass)

Explorations, experiments, tools etc.

Visualizations by Sarah (End of July)

Data exploration

Datas are from 18/07/2023 :

Total products : 29.293

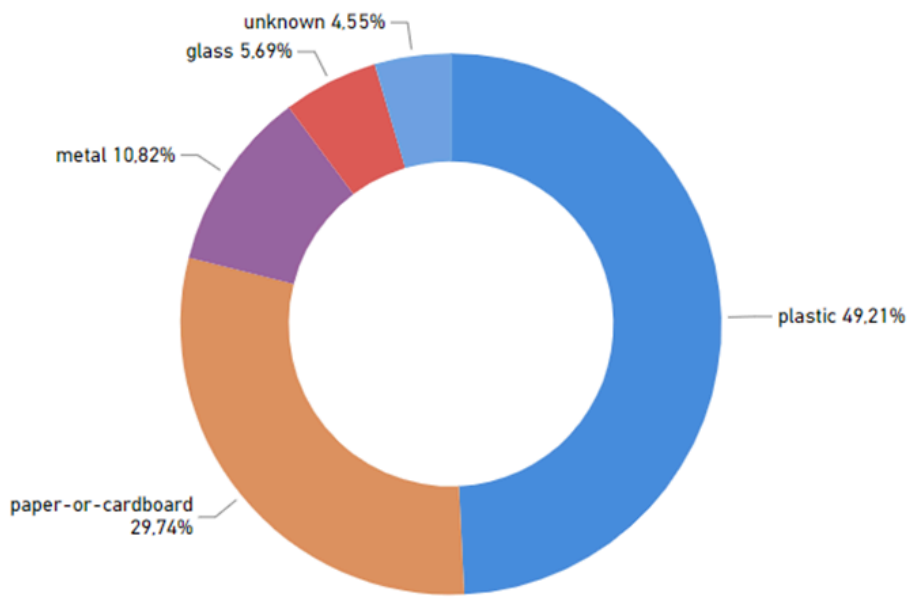
Total units : 49.000

Total weight measured : 175.910g

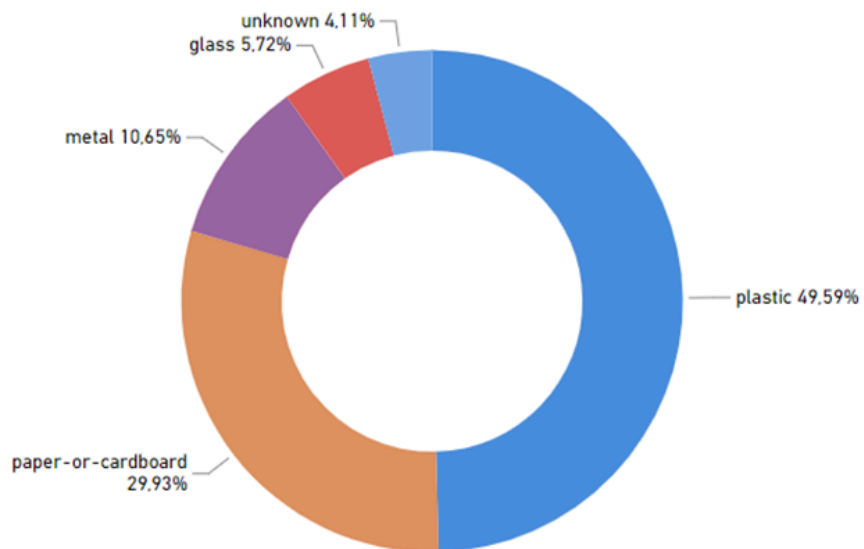
Total weight specified : 546.630g

1) By parent material

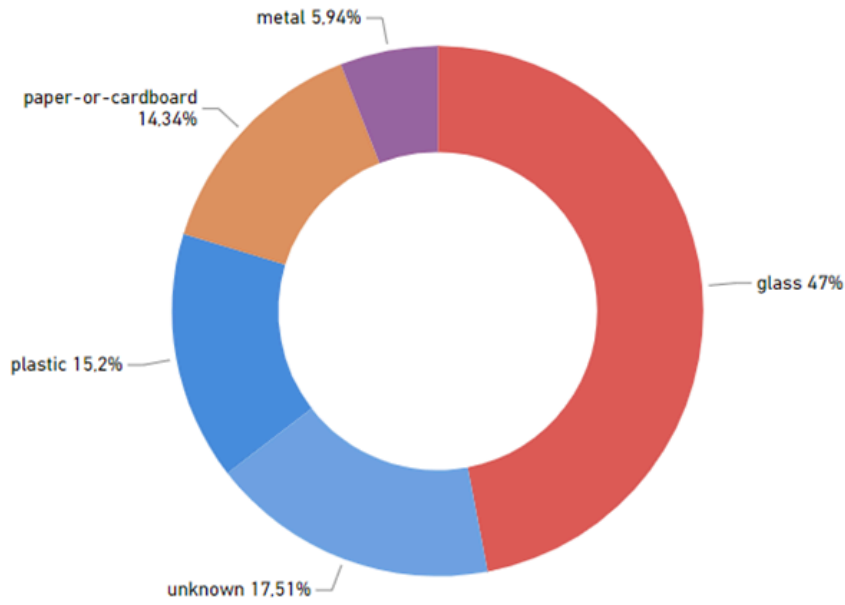
1.1) Products by parent material



1.2) Units by parent material

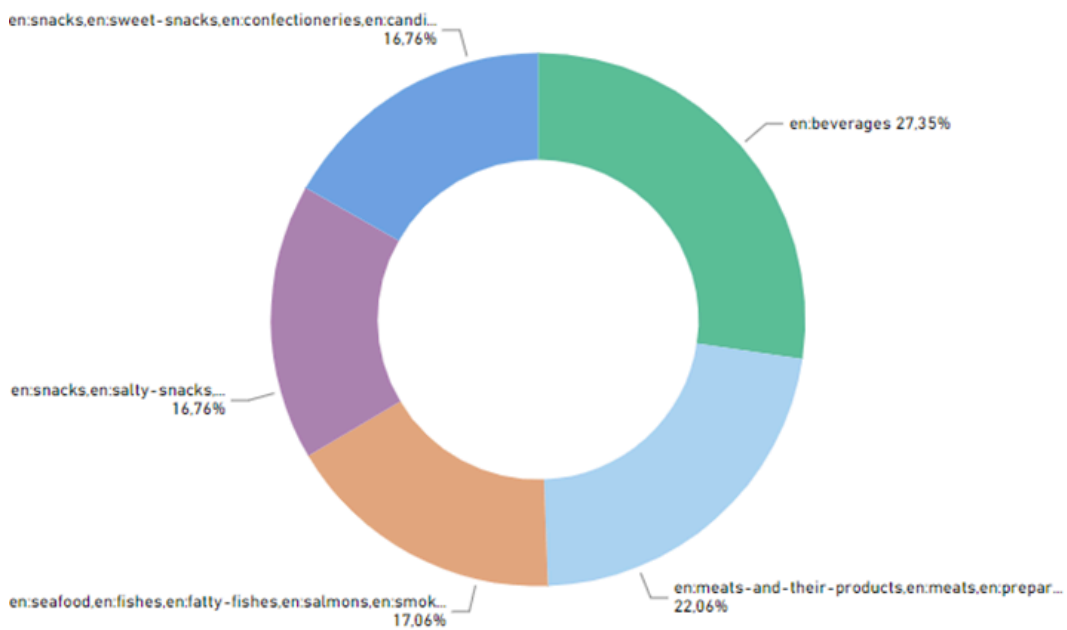


1.3) Total weight measured by parent material

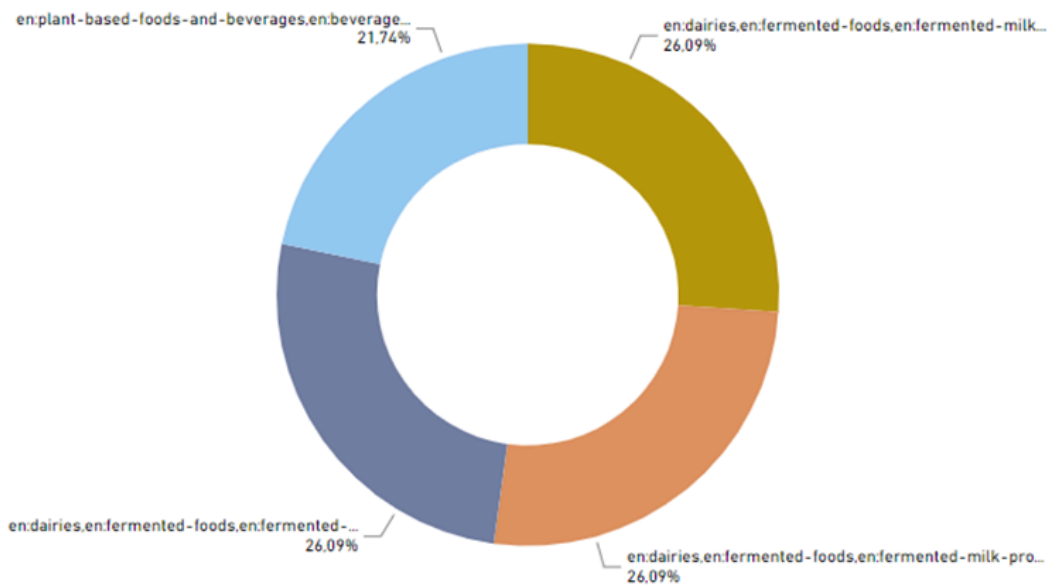


2) By top categories

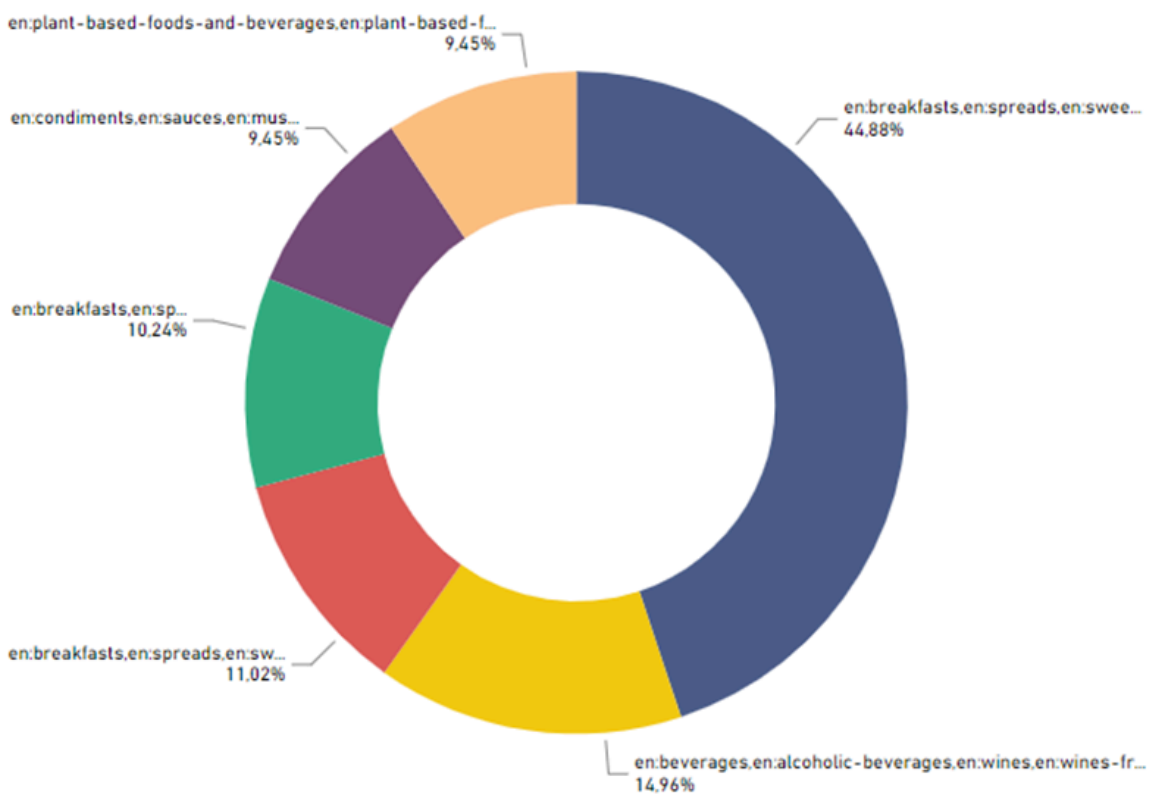
2.1) Plastic products by top categories



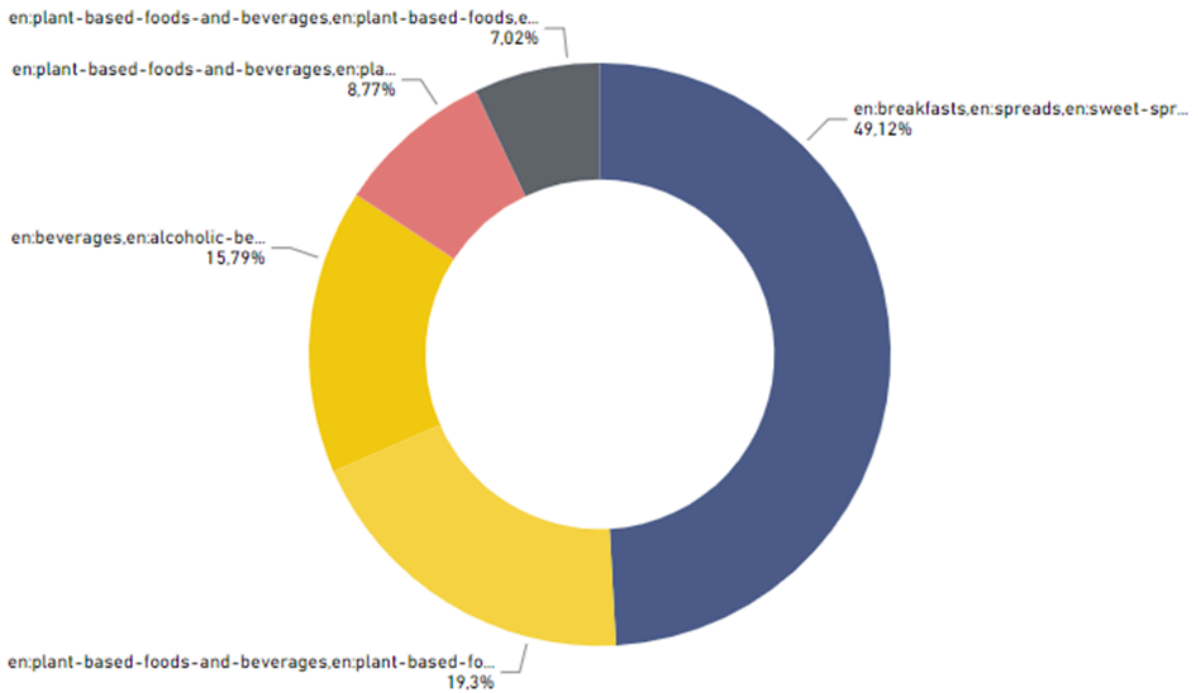
2.2) Paper or Cardboard products by top categories



2.3) Glass products by top categories

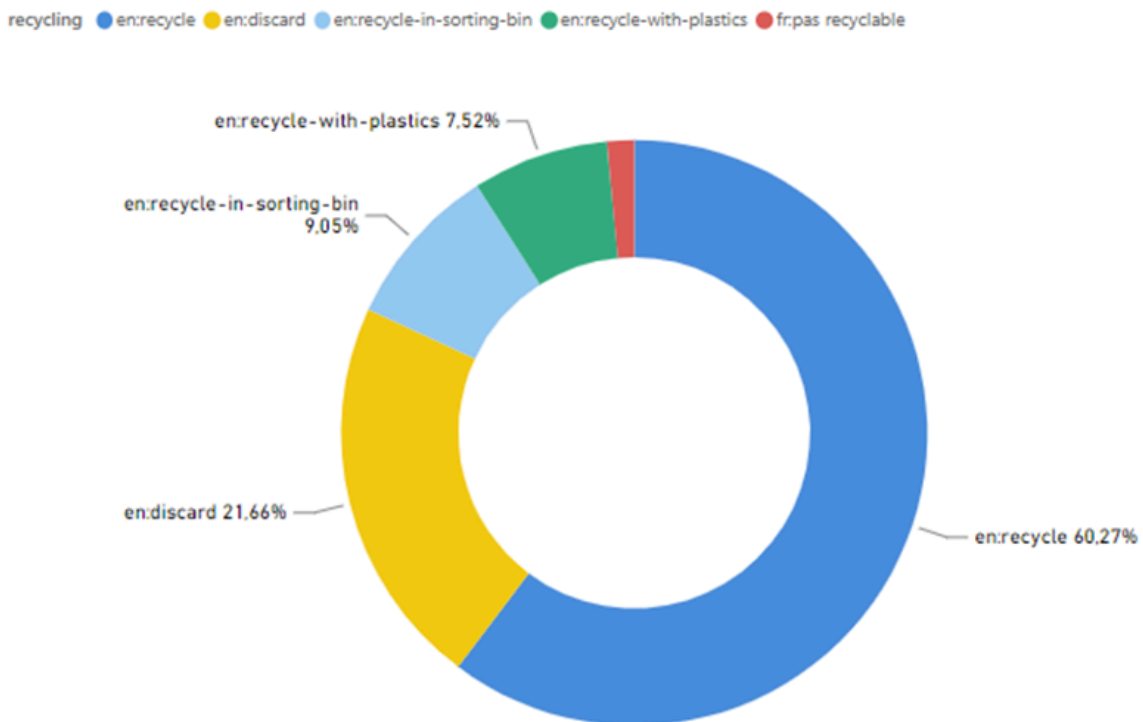


2.4) Metal products by top categories



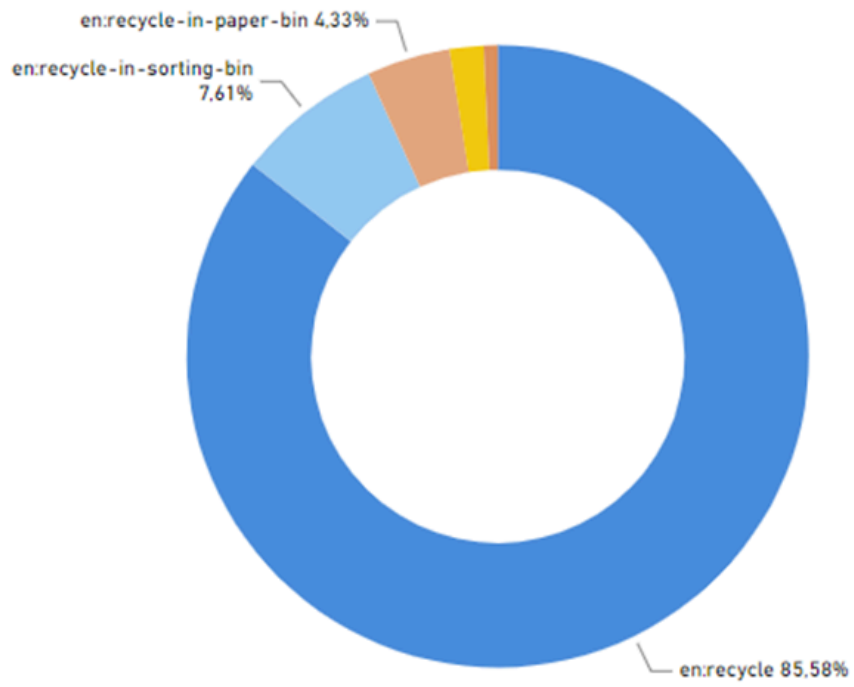
3) By type of recycling

3.1) Plastic products by top type of recycling



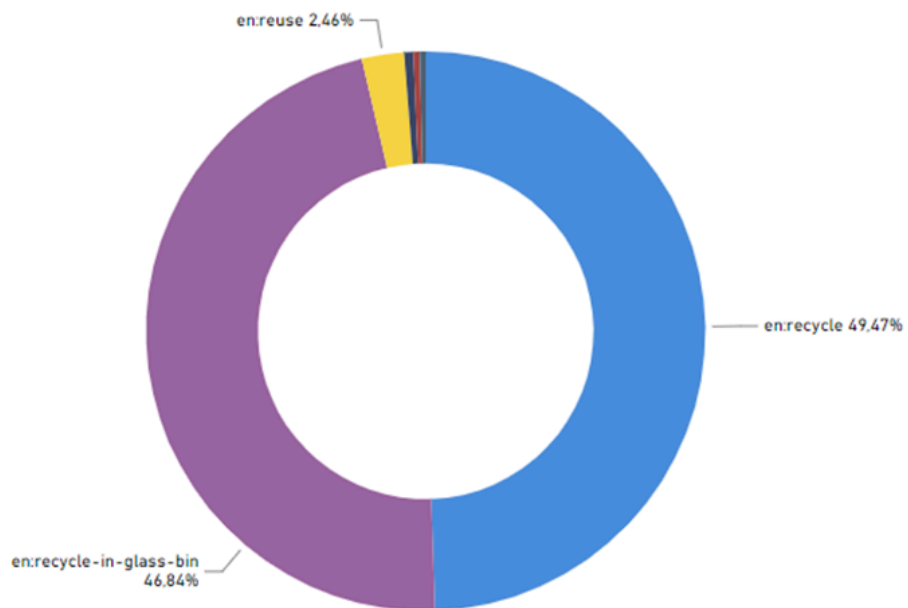
3.2) Paper or Cardboard products by top type of recycling

recycling ● en:recycle ● en:recycle-in-sorting-bin ● en:recycle-in-paper-bin ● en:discard ● fr:100 % recyclable



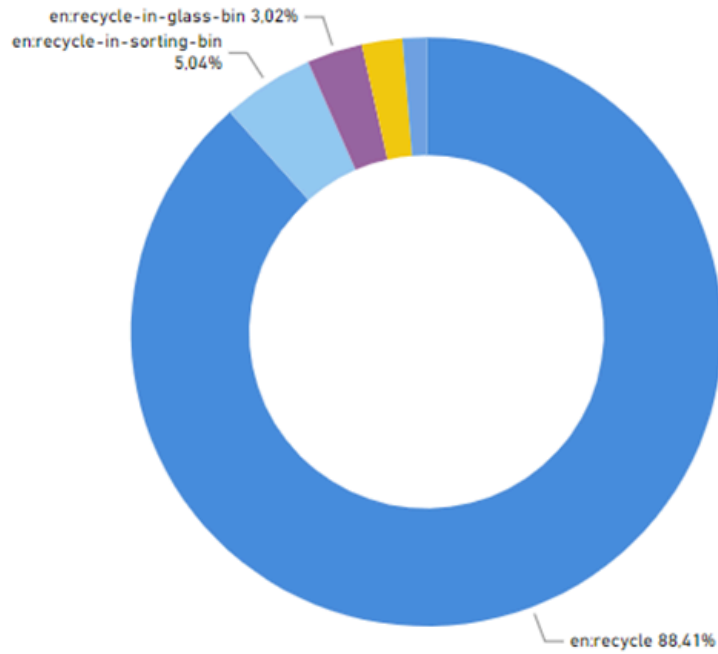
3.3) Glass products by top type of recycling

recycling ● en:recycle ● en:recycle-in-glass-bin ● en:reuse ● en:return-to-store ● en:deposit-refunds ● fr:oui



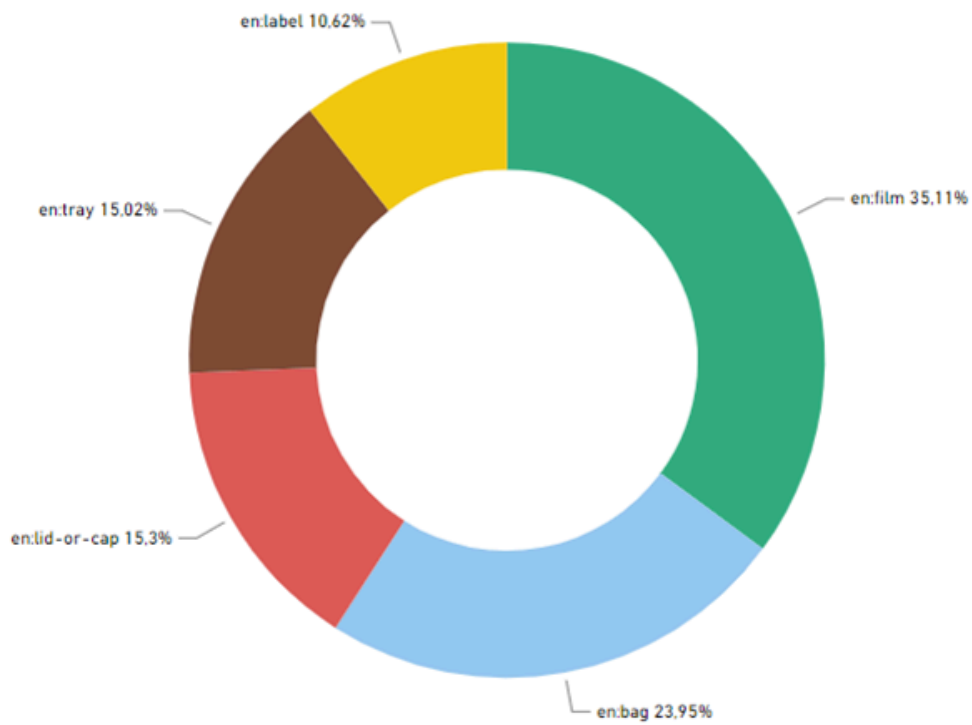
3.4) Metal products by top type of recycling

recycling ● en:recycle ● en:recycle-in-sorting-bin ● en:recycle-in-glass-bin ● en:discard ● en:recycle-with-plastics-metal-and-bricks



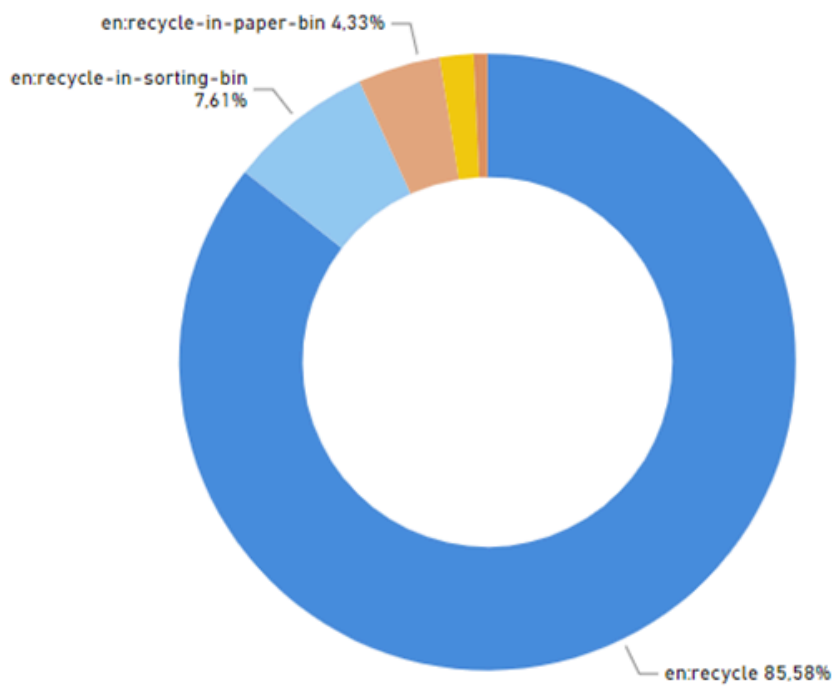
4) By shape

4.1) Plastic products by top shape

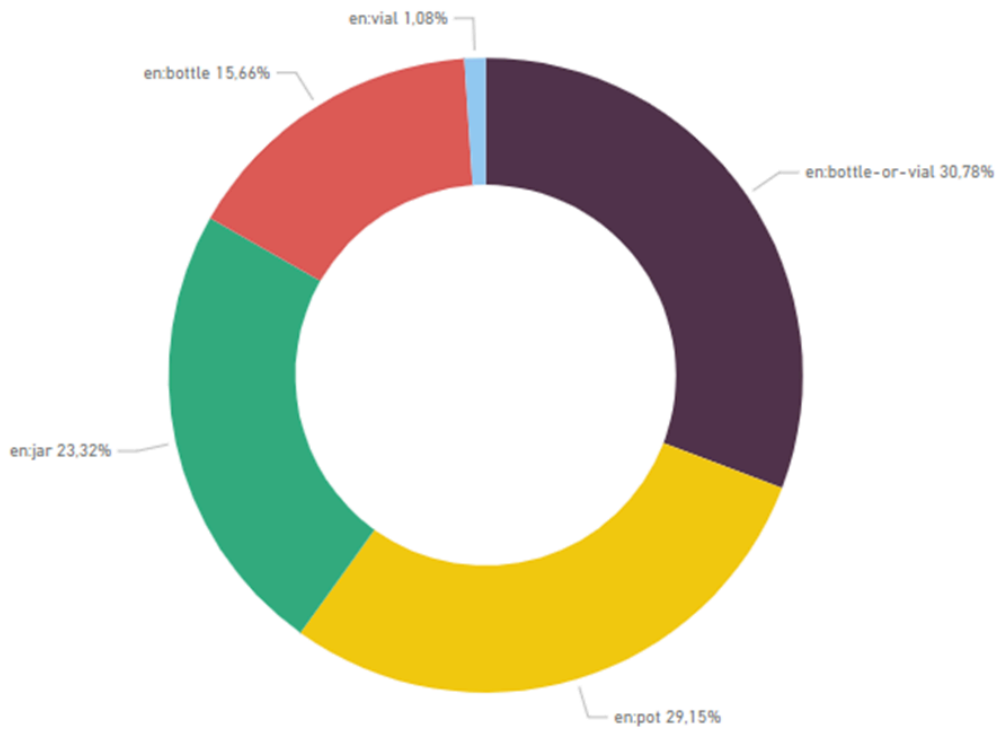


4.2) Paper or Cardboard products by top shape

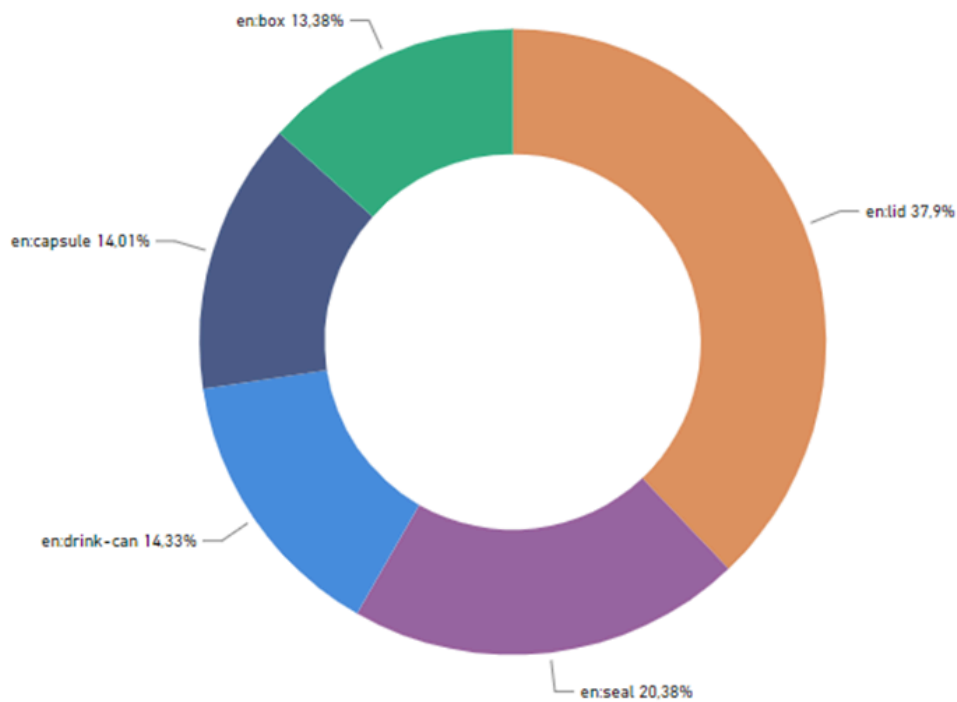
recycling ● en:recycle ● en:recycle-in-sorting-bin ● en:recycle-in-paper-bin ● en:discard ● fr:100 % recyclable



4.3) Glass products by top shape



4.4) Metal products by top shape



Explorations by Virginia

Cleaning

How to identify likely data errors

Using the packaging components CSV export

- List very small values
 - Might be units errors, e.g. kg instead of g
- List very large values
- Same with additional filter on material and or shape

Add data quality checks when saving products

- Compare weights for products of the same category (similar to checks for nutrients)
- Warning when the weight of a packaging is exactly the same as the weight of the product

Visualizations

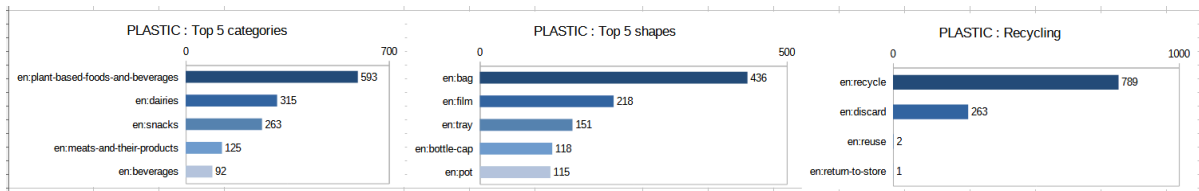
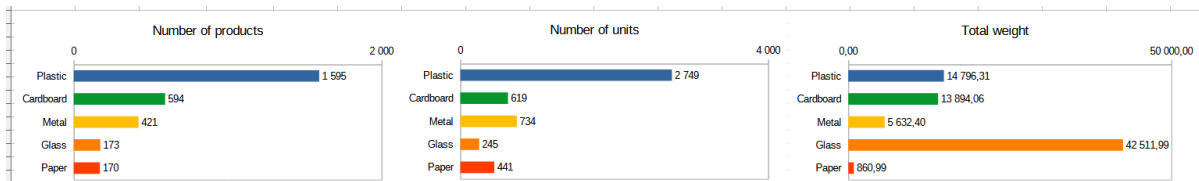
Visualizations by Virginia

Datas are from 24/04 :

3.160 products

5.177 units

for a total weight of 80.182g 



Would it be possible to integrate the visualisations into the existing graphing system?