

# Metadados



Software e Fluxos de Trabalho



### 3. Software e Fluxos de Trabalho

3.1 Folha de cálculo, **OpenRefine**, Open Data Editor

3.2 Visual Studio, Notepad++

3.3 Exiftool

3.4 MarcEdit, Metadata++, PDF Metadata Editor

3.5 Tropy, Digikam

3.6 Serviços web

### 3.1 OpenRefine

O OpenRefine (antigo Google Refine) é uma ferramenta open-source para limpar, transformar e explorar dados.

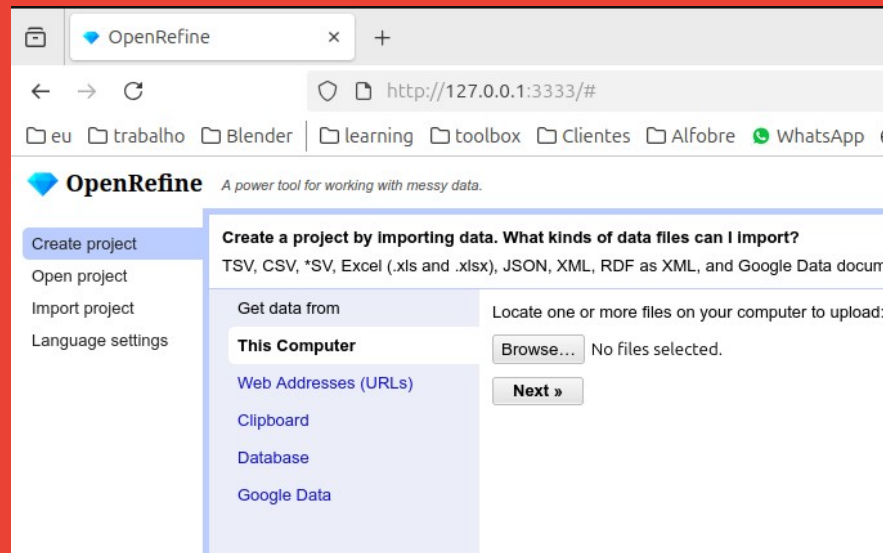
- + Focado na limpeza e transformação de dados
- + Grandes volumes de dados
- + Reprodutibilidade (histórico)

1. Limpar dados: unificar variações, remover espaços em branco, etc.
2. Transformar dados: dividir dados em múltiplas colunas, reorganizar tabelas, padronizar formatos, etc.
3. Explorar dados: Filtrar e agrupar dados rapidamente
4. Integrar com fontes externas: conectar a APIs (dados de Wikidata, VIAF, etc)
5. Exportar para múltiplos formatos

## 3.1 OpenRefine

1) Inicie o OpenRefine clicando 2 vezes no respetivo ícone. Irá abrir um browser com a aplicação.

Se não arrancar, digite **localhost:3333** (ou 127.0.0.1:3333) na barra de endereços do browser



## 3.1 OpenRefine

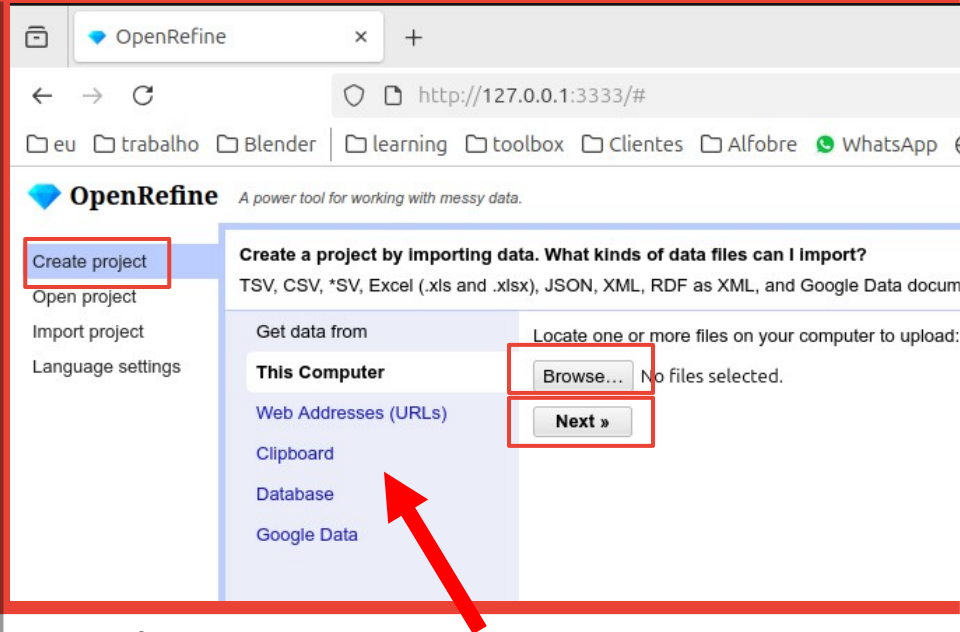
### 2) Clique na aba **Create Project**

Depois no botão **Browse**/Explorar e localize ficheiro Naufragios.csv

Depois no **Next** para importar os dados

A janela seguinte mostra uma prévisualização dos dados e pode alterar algumas configurações de importação caso necessário.

Por fim clique no botão **Create project**



É possível criar um projeto com outras fontes de dados.

Create project

« start over

Configure parsing options

Project name Naufragios csv

Tags

Create project

Open project

Import project

Language settings

Extensions

|     | SHIP'S NAME        | AKA          | SHIP'S OWNER            | VESSEL TYPE      | YEAR BUILT | WHERE BUILT      | DATE LOST  | YEAR | MONTH | DAY | LOCATION LOST                      | ITUDE LOST  | LONGITUDE   |
|-----|--------------------|--------------|-------------------------|------------------|------------|------------------|------------|------|-------|-----|------------------------------------|-------------|-------------|
| 1.  | 74 H               | 93015        |                         | Scow             |            |                  | 7/28/1916  | 1916 | 7     | 28  | Sandy Hook SE 5 mi                 |             |             |
| 2.  | A B Crosby         |              |                         | Schooner         | 1884       |                  | 3/12/1888  | 1888 | 3     | 12  | Atlantic Highlands                 |             |             |
| 3.  | A B Sherman        |              |                         | Schooner         | 1883       |                  | 12/8/1886  | 1886 | 12    | 8   | Five Fathoms Bank Light            |             |             |
| 4.  | A B Thompson       |              |                         | Yacht            |            |                  | 7/18/1874  | 1874 | 7     | 18  | Absecon                            |             |             |
| 5.  | A C Austin         |              |                         | Schooner         |            |                  | 2/1868     | 1868 | 2     |     | Sandy Hook Romer Shoal             |             |             |
| 6.  | A C Wescoat (1989) |              | A C Wescoat Co          | Deck barge       |            |                  | 10/11/1989 | 1989 | 10    | 11  | Atlantic City Artificial Reef      | 39-15-540 N | 74-14-691 W |
| 7.  | A C Wescoat (2006) |              |                         | Deck barge       |            |                  | 7/26/2006  | 2006 | 7     | 25  | Ocean City Artificial Reef         | 39-09-819 N | 74-34-310 W |
| 8.  | A D Scull          |              |                         | Schooner         | 1864       |                  | 4/18/1880  | 1880 | 4     | 18  | Townsend's Inlet                   |             |             |
| 9.  | A E Douglass       |              |                         | Schooner         | 1855       |                  | 6/4/1862   | 1862 | 6     | 4   | Barneget N 8 mi                    |             |             |
| 10. | A F Baillie        |              |                         | Schooner         | 1872       |                  | 11/17/1875 | 1875 | 11    | 17  | Carters Bar, VA                    |             |             |
| 11. | A F Crockett       |              |                         | Schooner         | 1881       | Bath, ME         | 5/21/1894  | 1894 | 5     | 21  | Harvey Cedars NE .5 mi             |             |             |
| 12. | A Field            | A D Field    |                         | Schooner         |            |                  | 2/10/1852  | 1852 | 2     | 10  | Deal Beach                         |             |             |
| 13. | A G Ropes          |              | Luckenbach SS Co        | Schooner - Barge | 1884       | Bath, ME         | 12/26/1913 | 1913 | 12    | 26  | Island Beach                       |             |             |
| 14. | A H Bowman         |              |                         | Steamship        |            |                  | 12/4/1857  | 1857 | 12    | 4   | Deal Lake N .25 mi                 |             |             |
| 15. | A H DuMont         |              | Spentonbush Red Star Co | Tanker           | 1940       | Camden, NJ       | 6/10/1986  | 1986 | 6     | 10  | Garden State North Artificial Reef | 39-37-727 N | 74-01-029 W |
| 16. | A H Edward         |              |                         | Schooner         |            |                  | 1820       | 1820 |       |     | Nine Mile Beach                    |             |             |
| 17. | A H Hurlburt       |              |                         | Schooner         | 1871       | Middletown, CT   | 6/26/1884  | 1884 | 6     | 26  | Cape May LSS SW 4 mi               |             |             |
| 18. | A J Bentley        |              | Bentley & Gildersleeve  | Schooner         | 1865       | Hartford, CT     | 5/27/1873  | 1873 | 5     | 27  | Barneget                           |             |             |
| 19. | A J Hoole          |              |                         | Schooner         |            |                  | 12/19/1902 | 1902 | 12    | 19  | Spermaceti Cove LSS S 1 mi         |             |             |
| 20. | A L & M Townsend   | George White |                         | Schooner         | 1871       | Barneget, NJ     | 12/30/1888 | 1888 | 12    | 30  | Little Egg LSS S 2 mi              |             |             |
| 21. | A L Lee (1897)     |              |                         | Schooner         | 1884       | Mays Landing, NJ | 3/25/1897  | 1897 | 3     | 25  | Ocean City LSS E .5 mi             |             |             |
| 22. | A L Lee (1898)     |              |                         | Schooner         | 1884       | Mays Landing, NJ | 8/26/1898  | 1898 | 8     | 26  | Absecon bar                        |             |             |
| 23. | A L Lee (1902)     |              |                         | Schooner         | 1884       | Mays Landing, NJ | 12/10/1902 | 1902 | 12    | 10  | Absecon bar                        |             |             |
| 24. | A Lincoln          |              |                         | Schooner         | 1865       | Essex, MA        | 10/12/1891 | 1891 | 10    | 12  | Barneget                           |             |             |
| 25. | A M C Smith        |              |                         | Schooner         |            |                  | 10/13/1876 | 1876 | 10    | 13  | Green Island LSS N .25 mi          | 39-36-00 N  | 74-12-06 W  |
| 26. | A M Lee            |              | G H Newhall             | Schooner         | 1864       | Tuckahoe, NJ     | 10/12/1885 | 1885 | 10    | 12  | Highlands SE .5 mi                 |             |             |
| 27. | A P Nowell         |              | Thomas Winsmore         | Schooner         | 1874       | Milton, DE       | 8/3/1886   | 1886 | 8     | 3   | Cape May                           |             |             |

Parse data as

Character encoding

UTF-8-BOM

Update preview

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

JSON-LD files

RDF/N3 files

RDF/N-Triples files

RDF/Turtle files

Columns are separated by

☒ commas (CSV)☐ tabs (TSV)☐ custom ,☒ Use character " " to enclose cells containing column separators☐ Trim leading & trailing whitespace from strings

Escape special characters with \

☐ Ignore first 0 line(s) at beginning of file☒ Parse next 1 line(s) as column headers☐ Column names (comma separated)☐ Discard initial 0 row(s) of data☐ Load at most 0 row(s) of data☐ Disable auto preview☐ Attempt to parse cell text into numbers☒ Store blank rows☒ Store blank columns☒ Store blank cells as nulls☐ Store file source☐ Store archive file

Version 3.9.3 [TRUNK]

Preferences

Help

About

Facet / Filter Undo / Redo 22 / 22

4,721 rows

Extensions Wikibase ▾

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first ‹ previous 1 - 10 next › last »

## Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

| ▼ DATE LOST | ▼ YEAR | ▼ MNTH | ▼ DAY | ▼ LOCATION LOST               | ▼ LATITUDE LOST | ▼ LONGITUDE LOST | ▼ CAUSE OF LOSS       | ▼ CONSTRUCTION | ▼  |
|-------------|--------|--------|-------|-------------------------------|-----------------|------------------|-----------------------|----------------|----|
| 7/28/1916   | 1916   | 7      | 28    | Sandy Hook SE 5 mi            |                 |                  | Grounded              |                |    |
| 3/12/1888   | 1888   | 3      | 12    | Atlantic Highlands            |                 |                  | Grounded in snowstorm |                | En |
| 12/8/1886   | 1886   | 12     | 8     | Five Fathoms Bank Light       |                 |                  | Storm                 | Wood           | US |
| 7/18/1874   | 1874   | 7      | 18    | Absecon                       |                 |                  | Capsized              |                | US |
| 2/1868      | 1868   | 2      |       | Sandy Hook Romer Shoal        |                 |                  | Grounded              | Wood           | US |
| 10/11/1989  | 1989   | 10     | 11    | Atlantic City Artificial Reef | 39-15-540 N     | 74-14-691 W      | Opened hull           | Steel          | US |
| 7/26/2006   | 2006   | 7      | 25    | Ocean City Artificial Reef    | 39-09-819 N     | 74-34-310 W      | Scuttled              | Steel          | US |
| 4/18/1880   | 1880   | 4      | 18    | Townsend's Inlet              |                 |                  | Grounded              | Wood           | US |
| 6/4/1862    | 1862   | 6      | 4     | Barnegat N 8 mi               |                 |                  | Foundered             | Wood           | US |
| 11/17/1875  | 1875   | 11     | 17    | Carters Bar, VA               |                 |                  | Storm                 | Wood           | US |

**OpenRefine** A power tool for working with messy data.

Create project « start over Configure parsing options

Open project

Import project

Language settings

Extensions

|     | SHIP'S NAME        | AKA       | SHIP'S OWNER          |
|-----|--------------------|-----------|-----------------------|
| 1.  | 74 H               | 93015     |                       |
| 2.  | A B Crosby         |           |                       |
| 3.  | A B Sherman        |           |                       |
| 4.  | A B Thompson       |           |                       |
| 5.  | A C Wescoat        |           |                       |
| 6.  | A C Wescoat (1989) |           | A C Wescoat Co        |
| 7.  | A C Wescoat (2006) |           |                       |
| 8.  | A D Soull          |           |                       |
| 9.  | A E Douglass       |           |                       |
| 10. | A F Ballie         |           |                       |
| 11. | A F Crockett       |           |                       |
| 12. | A Field            | A D Field |                       |
| 13. | A G Ropes          |           | Luckenbach SS Co      |
| 14. | A H Bowman         |           |                       |
| 15. | A H DuMont         |           | Spentonbush Red Star  |
| 16. | A H Edward         |           |                       |
| 17. | A H Hurlburt       |           |                       |
| 18. | A J Bentley        |           | Bentley & Gildecleave |

Pode alterar o idioma da interface a qualquer momento. Basta clicar no ícone do programa para aceder ao menu principal.

É também aí que pode importar projetos, abrir projetos existentes ou criar novos projetos.

**OpenRefine** Naufragios csv Perma

**Facet / Filter** Undo / Redo 22 / 22

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)



### 3.1 OpenRefine

Estamos na interface principal de trabalho. Podemos alterar a quantidade de linhas exibidas mas este software não é uma folha de cálculo...

3) Ative para mostrar **25**

Naufragios csv [Permalink](#)

Redo 6 / 6

4,721 rows

Show as: rows records Show: 5 10 **25** 50 100

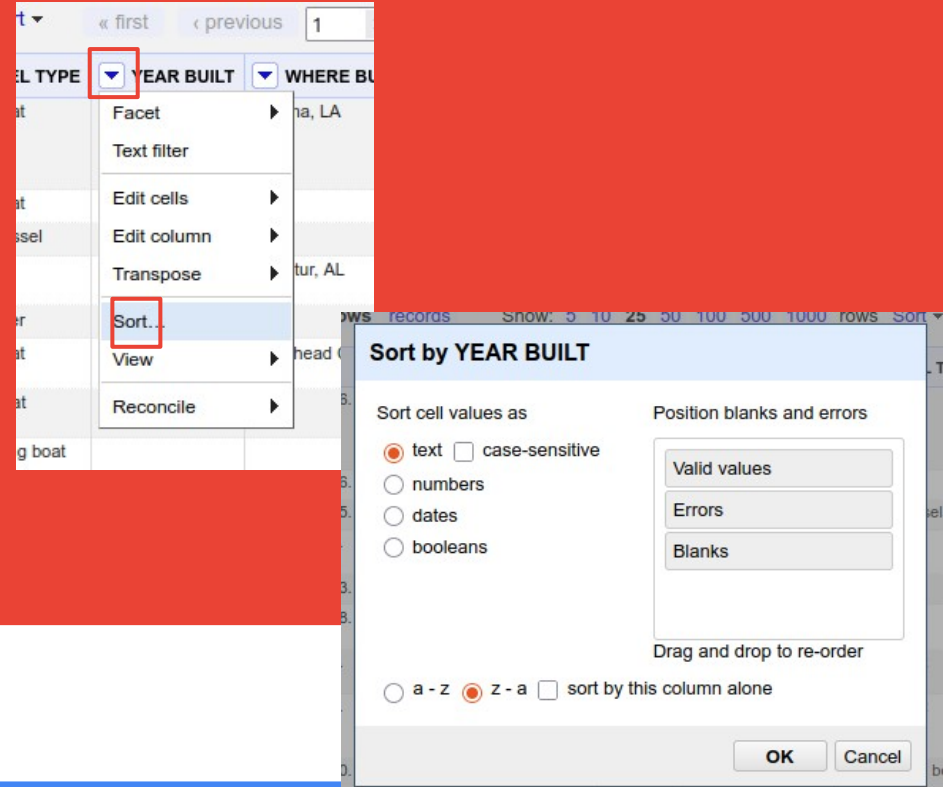
| <input type="checkbox"/> All | <input type="checkbox"/> SHIP'S NAME | <input type="checkbox"/> AKA | <input type="checkbox"/> SHIP'S OWNER |
|------------------------------|--------------------------------------|------------------------------|---------------------------------------|
| ☆                            | 1. 74 H                              | 93015                        |                                       |
| ☆                            | 2. A B Crosby                        |                              |                                       |
| ☆                            | 3. A B Sherman                       |                              |                                       |
| ☆                            | 4. A B Thompson                      |                              |                                       |
| ☆                            | 5. A C Austin                        |                              |                                       |
| ☆                            | 6. A C Wescoat (1989)                | \                            | A C Wescoat Co                        |
| ☆                            | 7. A C Wescoat (1989)                |                              |                                       |

### 3.1 OpenRefine – Sort / Ordenar

Para ordenar os dados usamos a ferramenta **Sort/Ordenar**

Esta ferramenta é configurada de acordo com o tipo de dados. Apesar de serem números, a coluna YEAR BUILT não está reconhecida como dados numéricos...

Este Sort não irá funcionar como desejado...



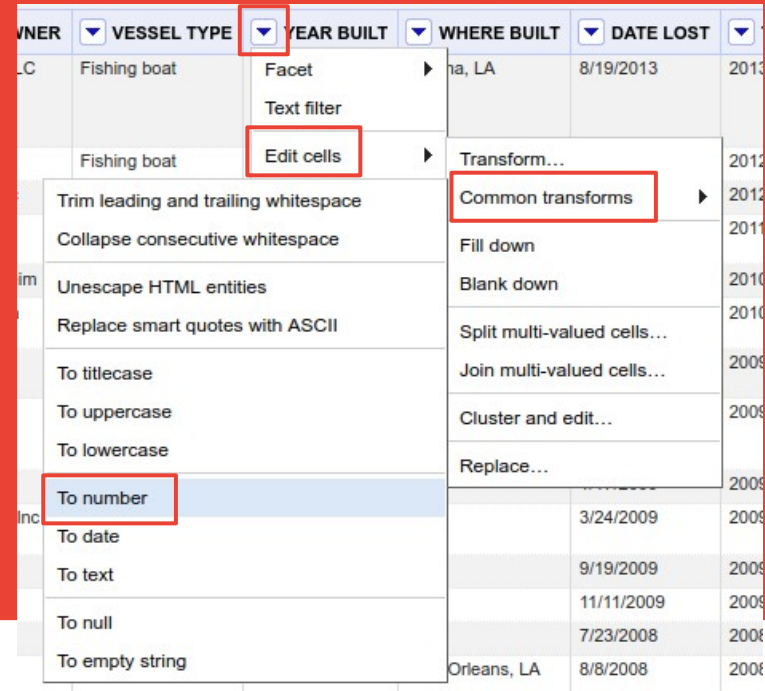
### 3.1 OpenRefine – Sort / Ordenar

4) Se criou entretanto, remova os Sorts criados

**Transforme os dados** da coluna para dados numéricos e depois ative a ferramenta Sort.

Quando alterar, os número passam a estar com a cor verde.

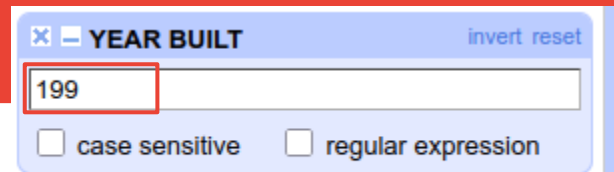
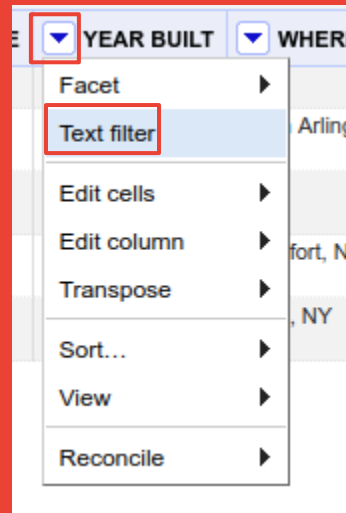
Ative o **Sort**



## 3.1 OpenRefine – Filtrar

### 5) Pode usar um filtro de texto

Ative o filtro de texto e escreva a chave com que pretende filtrar

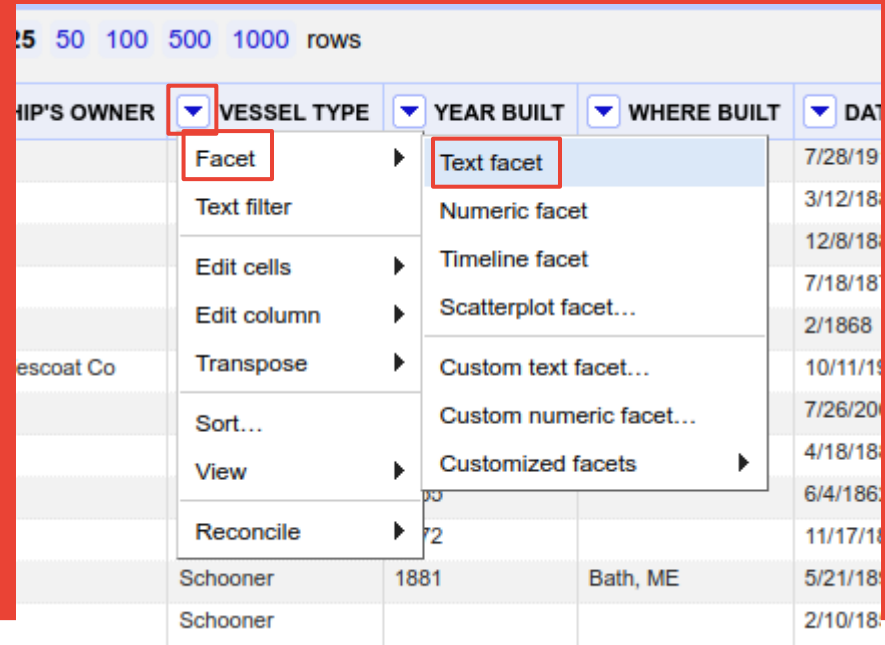


### 3.1 OpenRefine – Facetas de texto

Vamos criar uma faceta de texto

6) Clique no **menu** de VESSEL TYPE, escolha **Facet** e depois **Text facet**

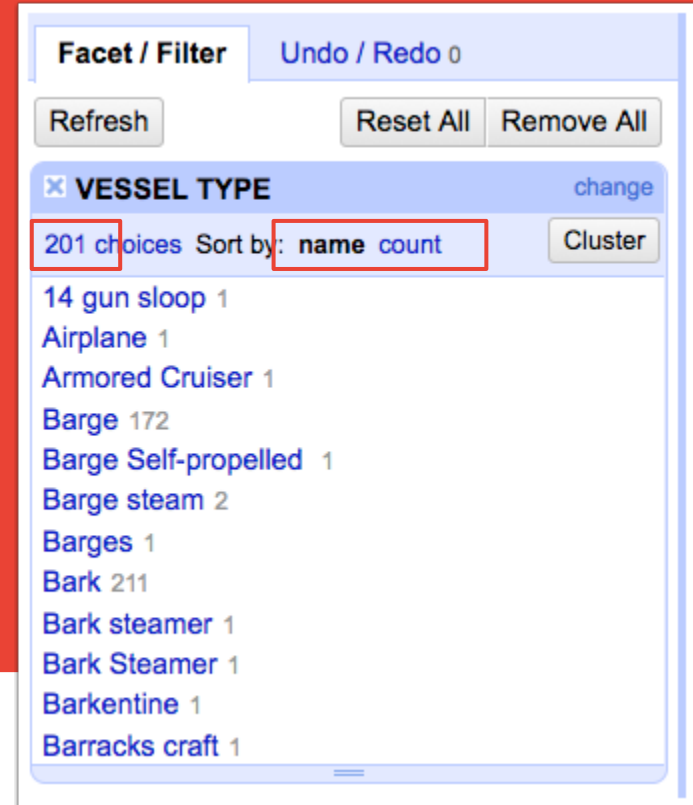
Surgiu agora uma janela-faceta que permite selecionar os termos que existem na coluna VESSEL TYPE.



### 3.1 OpenRefine – Facetas de texto

Temos 201 termos por ordem alfabética

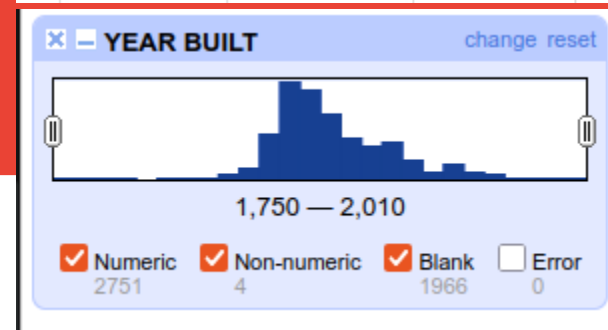
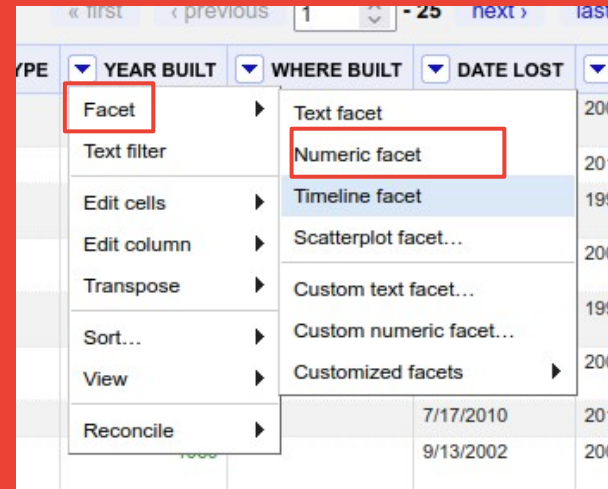
- Experimente ver os termos mais utilizados
- Experimente clicar em cima de um termo mostrar apenas as linhas com esse termo
- Experimente clicar no **include** (quando passa rato por cima)



### 3.1 OpenRefine – Facetas numéricas

7) Também pode criar uma faceta numérica na coluna YEAR BUILT (porque os dados são numéricos!)

- Experimente usar a faceta para filtrar os dados.



### 3.1 OpenRefine – Facetas com datas

Também pode criar uma faceta com datas mas os dados têm que ser datas e...

OpenRefine converte para formato...

YYYY-MM-DDTHH:MM:SSZ



A screenshot of an OpenRefine facet titled 'DATE LOST'. The facet displays a list of ISO 8601 timestamps. The first row is highlighted in light blue. The timestamps are: 1916-07-28T00:00:00Z, 1888-03-12T00:00:00Z, 1886-12-08T00:00:00Z, 1874-07-18T00:00:00Z, 2/1868, 1989-10-11T00:00:00Z, 2006-07-26T00:00:00Z, 1880-04-18T00:00:00Z, 1862-06-04T00:00:00Z, and 1875-11-17T00:00:00Z.

| DATE LOST            |
|----------------------|
| 1916-07-28T00:00:00Z |
| 1888-03-12T00:00:00Z |
| 1886-12-08T00:00:00Z |
| 1874-07-18T00:00:00Z |
| 2/1868               |
| 1989-10-11T00:00:00Z |
| 2006-07-26T00:00:00Z |
| 1880-04-18T00:00:00Z |
| 1862-06-04T00:00:00Z |
| 1875-11-17T00:00:00Z |



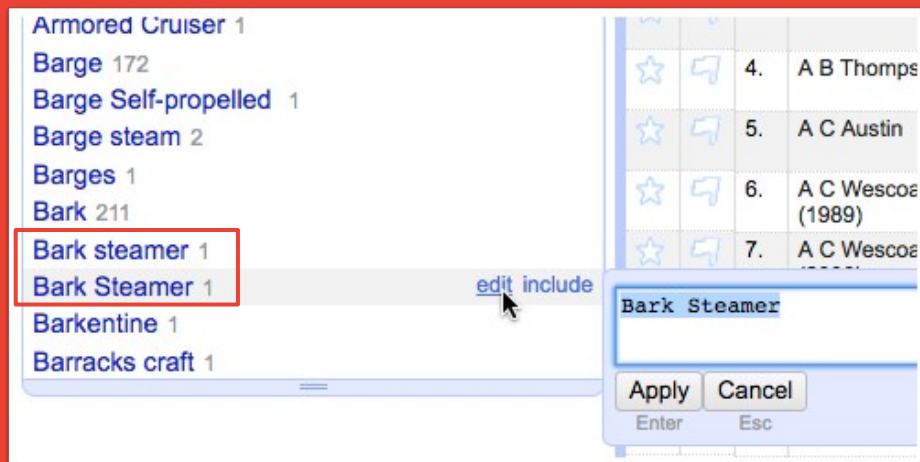
### 3.1 OpenRefine – Editar

Vamos limpar dados!

Na lista da faceta de texto existe **Bark steamer** e **Bark Steamer**

8) Quando passa o rato por cima do termo **Bark Steamer** surge a opção **edit**. Edite e altere para **Bark steamer**. Clique em **Apply** no final.

Os dois termos fundiram-se, agora existem 200 termos.



## 3.1 OpenRefine - Cluster

9) No topo da janela de faceta clique em **Cluster**. Depois clique novamente no botão **Cluster** no centro da janela. Surgem os termos que o OpenRefine sugere que sejam fundidos.

Experimente outros algoritmos (method/keying).

**Merge & re-cluster** para continuar e **close** para fechar.

| Merge?                              | Values In cluster   | New cell value     | Cluster size | Row count | # Rows in cluster |
|-------------------------------------|---|--------------------|--------------|-----------|-------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Privateer Schooner<br><input type="checkbox"/> Schooner - privateer            | Privateer Schooner | 2            | 2         |                   |
| <input type="checkbox"/>            | <input type="checkbox"/> Fishing Vessel<br><input type="checkbox"/> Fishing vessel                      | Fishing Vessel     | 2            | 2         |                   |
| <input type="checkbox"/>            | <input type="checkbox"/> Freighter steam (17 rows)<br><input type="checkbox"/> Steam freighter (3 rows) | Freighter steam    | 2            | 20        |                   |
| <input type="checkbox"/>            | <input type="checkbox"/> Pilot schooner (30 rows)<br><input type="checkbox"/> Pilot Schooner (2 rows)   | Pilot schooner     | 2            | 32        |                   |
| <input type="checkbox"/>            | <input type="checkbox"/> Sailing ship (24 rows)<br><input type="checkbox"/> Sailing Ship                | Sailing ship       | 2            | 25        |                   |
| <input type="checkbox"/>            | <input type="checkbox"/> Brig Privateer (3 rows)<br><input type="checkbox"/> Privateer Brig             | Brig Privateer     | 2            | 4         |                   |

Select all Deselect all Export clusters Merge selected & re-cluster

### 3.1 OpenRefine – Transformações comuns

Existem vários termos com minúsculas/maiúsculas inconsistentes (p. ex. **Armored Cruiser** e **Barge steam**)

10) Vamos tornar tudo minúsculas com a **transformação comum To lowercase**

Existem mais transformações comuns úteis!

The screenshot shows the OpenRefine interface with a table of data. The 'VESSEL T' column is selected, and the 'Edit cells' menu is open. The 'To lowercase' option is highlighted. Red circles with numbers 1 through 4 indicate the steps to apply the transformation: 1. Click the dropdown arrow for 'VESSEL T'. 2. Click 'Edit cells'. 3. Click 'Transform...'. 4. Click 'To lowercase'.

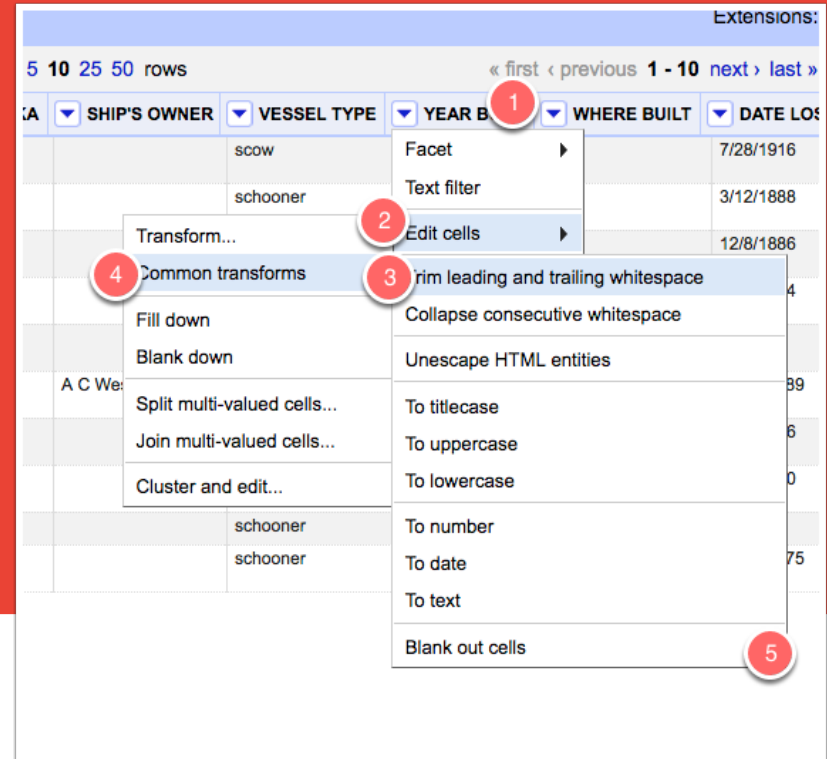
| S NAME | AKA   | SHIP'S OWNER | VESSEL T | YEAR BUILT | WHERE BUILT | DATE      |
|--------|-------|--------------|----------|------------|-------------|-----------|
|        | 93015 |              |          |            |             | 7/28/1... |
| oy     |       |              |          | 84         |             | 3/12/1... |
| nan    |       |              |          |            |             | 12/8/1... |
| ypson  |       |              |          |            |             | 7/18/1... |
| n      |       |              |          |            |             | 2/186E    |
| coat   | \     |              |          |            |             | 10/11/... |
| coat   |       |              |          |            |             | 7/26/2... |
|        |       |              |          |            |             | 4/18/1... |
| lass   |       |              |          | 55         |             | 6/4/18... |
|        |       |              |          | 72         |             | 11/17/... |

### 3.1 OpenRefine – Remover espaços em branco a mais

Existem células que têm espaços em branco antes/depois dos valores...

11) Vamos limpar tudo com a **transformação comum Trim leading and trailing whitespace**

Aproveite e faça Collapse consecutive whitespace para remover espaços em branco consecutivos



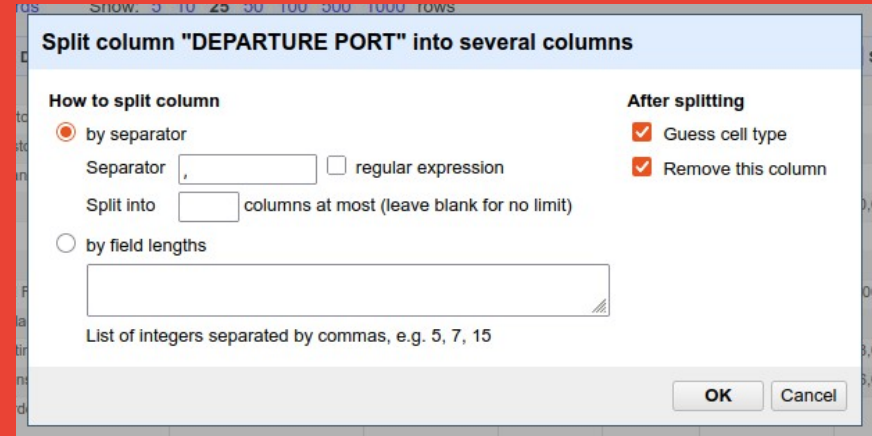
### 3.1 OpenRefine – Dividir em colunas

A coluna DEPARTURE POINT tem uma localização com vários elementos (cidade, estado, país). Vamos separar a coluna...

12) Escolha **Edit columns** e **Split into multiple columns**

Defina o separador. No exemplo utilizámos “,” (vírgula e espaço em branco!)

No mesmo menu escolha **Rename**



### 3.1 OpenRefine – Undo/Redo e Histórico

A janela Undo/Redo guarda um histórico das operações (**que pode ser exportado e reutilizado-aplicado!**)

13) Experimente fazer **Undo** e depois **Redo**

The screenshot shows the 'Undo / Redo' panel in OpenRefine. At the top, there are tabs for 'Facet / Filter' and 'Undo / Redo 19 / 19'. Below the tabs are 'Extract...' and 'Apply...' buttons. A 'Filter' input field is present. The main area is a list of 19 actions, numbered 0 to 19. Action 19 is highlighted in blue. The actions include creating a project, starring/unstarring rows, flagging/unflagging rows, editing single cells, mass editing cells in columns, adding rows, removing rows, mass editing cells in columns, text transforms, and splitting cells.

| Action  |
|---|
| 0. Create project   |
| 1. Star row 9   |
| 2. Unstar row 9   |
| 3. Star row 9   |
| 4. Unstar row 9   |
| 5. Flag row 9   |
| 6. Unflag row 9   |
| 7. Edit single cell on row 1, column SHIP'S NAME                                    |
| 8. Mass edit 1 cells in column VESSEL TYPE  |
| 9. Mass edit 4 cells in column VESSEL TYPE  |
| 10. Add 1 row   |
| 11. Flag row 1  |
| 12. Edit single cell on row 1, column VESSEL TYPE                                   |
| 13. Unflag row 1  |
| 14. Remove 1 rows   |
| 15. Mass edit 85 cells in column VESSEL TYPE  |
| 16. Mass edit 26 cells in column VESSEL TYPE  |
| 17. Text transform on 12 cells in column VESSEL TYPE: value.trim()                  |
| 18. Text transform on 0 cells in column VESSEL TYPE: value.replace(/[p{Zs}s]+/, '') |
| 19. Split 3072 cell(s) in column DEPARTURE PORT into several columns by separator   |

## 3.1 OpenRefine – Undo/Redo e Histórico

Facet / Filter **Undo / Redo 14 / 14**

**Extract...** Apply...

Filter:

0. Create project
1. Text transform on 7 cells in column `¿Facility: value.trim()`
2. Text transform on 0 cells in column `Program: value.trim()`

### Extract Operation History

Extract and save parts of your operation history as JSON that you can apply to this or other projects in the future.

- ☒ Text transform on cells in column `¿Facility` using expression `value.trim()`
- ☒ Text transform on cells in column `Program` using expression `value.trim()`
- ☒ Text transform on cells in column `Attendance` using expression `value.trim()`
- ☒ Text transform on cells in column `Date` using expression `value.trim()`
- ☒ Text transform on cells in column `Quarter` using expression `value.trim()`
- ☒ Text transform on cells in column `Period_Type` using expression `value.trim()`
- ☒ Text transform on cells in column `ObjectId` using expression `value.trim()`
- ☒ Text transform on cells in column `Attendance` using expression `value.toNumber()`
- ☒ Text transform on cells in column `Date` using expression `value.toDate()`

Select All Unselect All

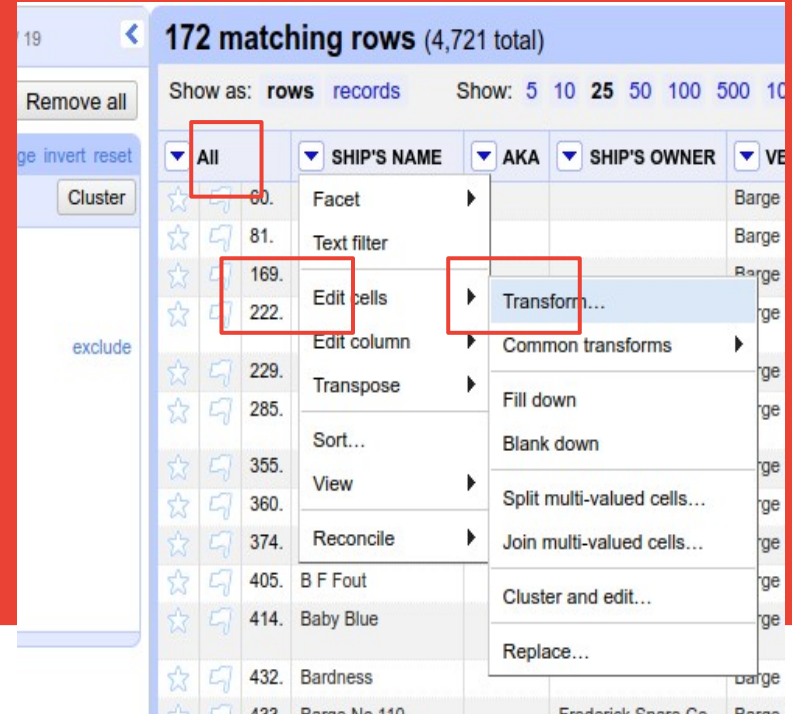
```
[
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "¿Facility",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column ¿Facility using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Program",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Program using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Attendance",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Attendance using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Date",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Date using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Quarter",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Quarter using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Period_Type",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Period_Type using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "ObjectId",
    "expression": "value.trim()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column ObjectId using expression value.trim()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Attendance",
    "expression": "value.toNumber()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Attendance using expression value.toNumber()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Date",
    "expression": "value.toDate()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column Date using expression value.toDate()"
  }
]
```

### 3.1 OpenRefine - Transformações

Vamos adicionar um prefixo a todos os barcos do tipo **barge**.

14) Utilize a faceta para escolher os barcos de tipo **Barge**

15) No menu SHIP'S NAME escolha **Edit cells** e **Transform**





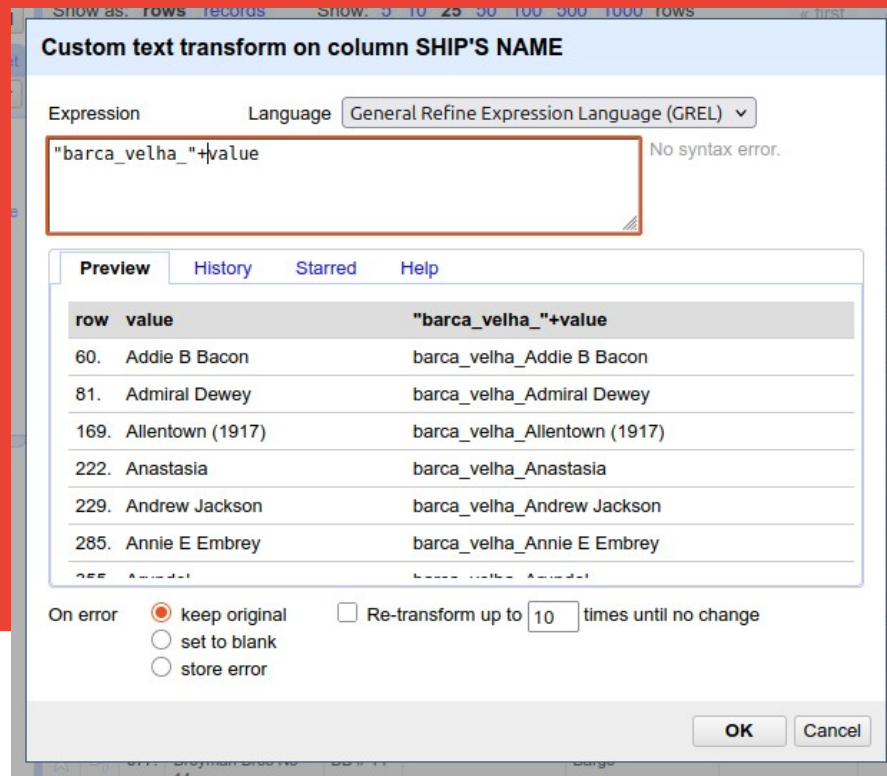
## 3.1 OpenRefine

Na janela vamos escrever a expressão que permite inserir um prefixo. O *value* representa o valor existente na célula.

Utilizámos esta expressão

*"barca\_velha\_" + value*

Em baixo pode pré-visualizar a alteração.



### 3.1 OpenRefine – Outras transformações

Para substituir caracteres (um conjunto ou vários conjuntos)

`value.replace("barca_velha", "111")`

**`value.replaceEach(["barca",  
"velha"], ["111", "222"])`**

The screenshot shows the 'Custom text transform on column SHIP'S NAME' dialog box. The 'Expression' field contains the code `value.replaceEach(["barca", "velha"], ["111", "222"])`. The 'Language' dropdown is set to 'General Refine Expression Language (GREL)'. A 'No syntax error.' message is displayed. Below the expression field, there are tabs for 'Preview', 'History', 'Starred', and 'Help'. The 'Preview' tab is active, showing a table with three columns: 'row', 'value', and 'value.replaceEach(["barca", "v ...'. The table contains several rows of data, including '4115. barca\_velha\_Southwick's' which is transformed to '111\_222\_Southwick's'. At the bottom, there are options for 'On error' (keep original, set to blank, store error) and a checkbox for 'Re-transform up to 10 times until no change'. 'OK' and 'Cancel' buttons are at the bottom right.

Custom text transform on column SHIP'S NAME

Expression Language General Refine Expression Language (GREL) ▾

`value.replaceEach(["barca", "velha"], ["111", "222"])` No syntax error.

**Preview** History Starred Help

| row   | value                         | value.replaceEach(["barca", "v ... |
|-------|-------------------------------|------------------------------------|
| 4115. | barca_velha_Southwick's       | 111_222_Southwick's                |
| 2696. | barca_velha_Libra             | 111_222_Libra                      |
| 2077. | barca_velha_Hughes 137        | 111_222_Hughes 137                 |
| 3122. | barca_velha_Michelle S        | 111_222_Michelle S                 |
| 1189. | barca_velha_Eastern 2         | 111_222_Eastern 2                  |
| 4295. | barca_velha_Thomas J Lanigan  | 111_222_Thomas J Lanigan           |
| 4704. | barca_velha_George A. Kellner | 111_222_George A. Kellner          |

On error ☒ keep original ☐ Re-transform up to  times until no change  
☐ set to blank  
☐ store error

OK Cancel

### 3.1 OpenRefine – Outras transformações

Para substituir data completa

YYYY-MM-DDTHH:MM:SSZ

só pelos anos

```
value.datePart("years")
```

Mas não resolve a faceta de tempo (dados deixam de ser data)!

| DATE LOST            | DATE LOST |
|----------------------|-----------|
| 2013-01-01T00:00:00Z | 1916      |
|                      | 1888      |
|                      | 1886      |
| 2012-01-01T00:00:00Z | 1874      |
| 2012-01-01T00:00:00Z | 1868      |
| 2011-01-01T00:00:00Z | 1989      |
| 2010-01-01T00:00:00Z | 2006      |
| 2010-01-01T00:00:00Z | 1880      |
|                      | 1862      |
| 2009-01-01T00:00:00Z | 1875      |
| 2009-01-01T00:00:00Z | 1894      |
| 2009-01-01T00:00:00Z | 1852      |
|                      | 1913      |
| 2009-01-01T00:00:00Z | 1857      |
| 2009-01-01T00:00:00Z | 1986      |
|                      | 1820      |

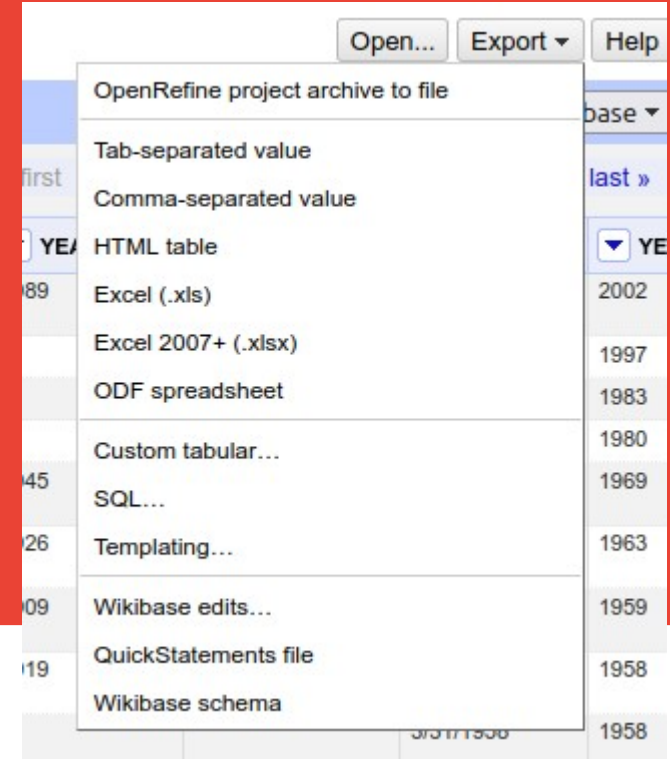
### 3.1 OpenRefine – Exportar e Fechar

**Para exportar** dados (visíveis!) utilizamos o botão **Export**.

**Para partilhar** exportar projeto.

**Para encerrar,**

- a) feche as abas/browser
- b) na janela de terminal clique em CTRL+C (garante gravação)



### 3.1 OpenRefine – Utilidades nas Colunas

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

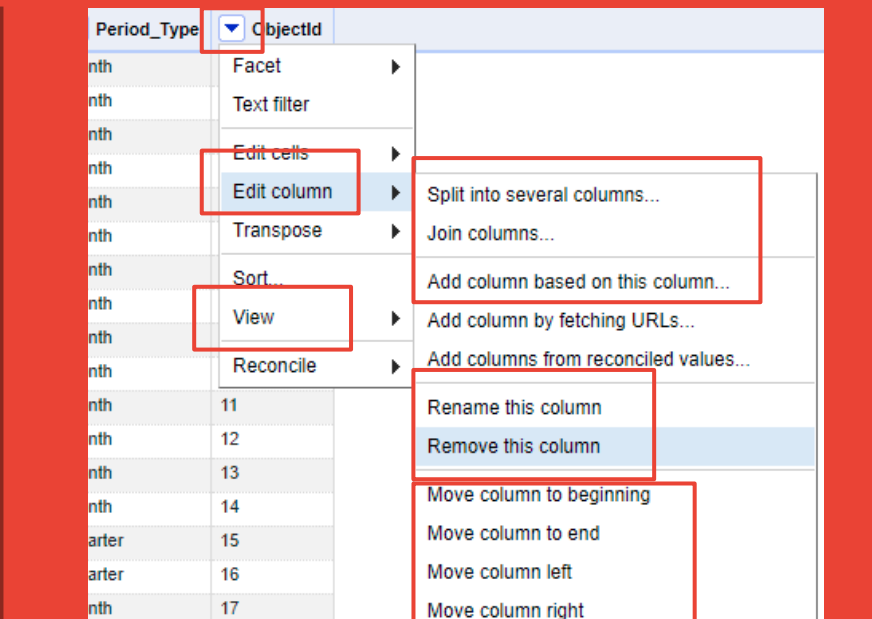
- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View

- Dividir em colunas
- Juntar colunas
- Duplicar colunas
- Renomear colunas
- Remover colunas
- Reordenar/mover colunas
- Esconder (colapsar) colunas no menu View



### 3.1 OpenRefine – Utilidades gerais

# Estrelas e Flags

## Facetas das estrelas e Flags

### Adicionar linhas em branco

**Apagar** = Marcar com Estrela ou Flag + Ativar Facet by Star ou Flag + Edit Rows + Remove Matching

## Estrelas e Flags

Facetas das estrelas e Flags

## Adicionar linhas em branco

**Apagar** = Marcar com Estrela ou Flag + Ativar Facet by Star ou Flag + Edit Rows + Remove Matching

[illegible]

### 3.1 OpenRefine – Exercícios

Usar facets de texto e numéricos

Usar estrelas e flags (e facets)

Filtrar e Ordenar

Adicionar linha (inserir dados)

Duplicar coluna

Juntar e dividir em colunas

Reordenar colunas

Cluster

Remover whitespaces

Transformar para minúsculas

Adicionar prefixo e sufixo

Substituir texto

Exportar tudo

Exportar uma parte