association des archivistes français



Genrer les données patrimoniales

Atelier

Intervenant : Maïwenn Bourdic
Lieu : Rennes
Date : 26 et 27 mars 2025

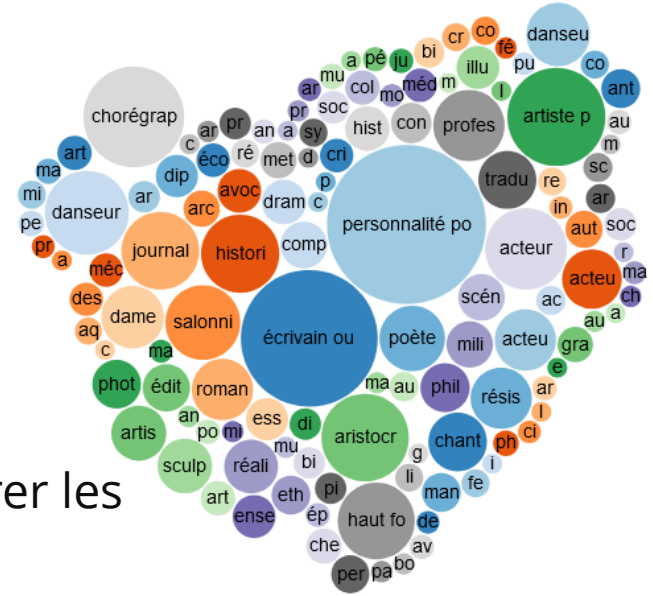
avec attention
archives, archivistes & sociétés

xelians SAINT-GOBAIN naoned IRON MOUNTAIN Prisme vision Archinoé Évolution

Source de l'image : tweet de J. Benedetti AD13

Objectifs

- Enrichir nos métadonnées archivistiques (sans retourner au carton)
- Mettre en lumière des thématiques, améliorer les réponses aux recherches
- Révéler l'implicite de nos métadonnées

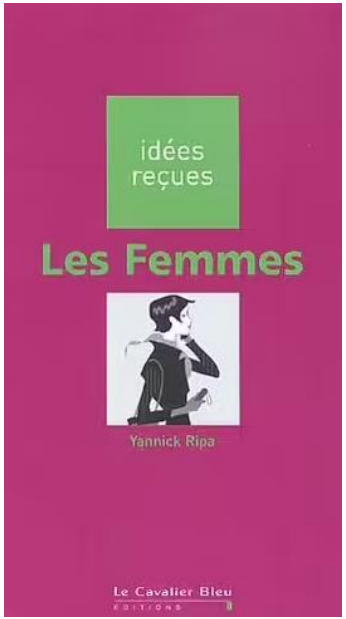

danseu
artiste p
chorégrap
danseur
personnalité po
journal
histori
écrivain ou
acteur
salonni
dame
poète
profes
tradu
roman
aristocr
haut fo
chant
résis
phil
acteu
scén
comp
dram
avoc
dip
édit
phot
artis
sculp
réali
mili
gra
aut
soc
ant
co
pu
illu
hist
con
met
cri
éco
arc
ess
eth
ense
per
pa
bo
che
man
ma
au
ar
pr
soc
sy
an
ré
ac
ch
ph
ci
ma
mi
pe
aq
des
méc
art
dip
ma
mi
pe
pr
a
c
m
sc
re
in
r
e
l
i
g
li
fe
de
av
ép
pi
bi
mu
po
di
ju
pê
fê
cr
bi
m
mo
méd
col
ar
pr
soc
an
a
sy
hist
con
profes
artiste p
danseu
co
ant
au
m
sc
ar
aut
soc
r
ma
ch
a
au
gra
e
ar
l
ci
ph
i
fe
de
av
bo
pa
per
che
ép
pi
eth
ense
haut fo
man
ma
au
phil
résis
chant
aristocr
ess
di
mu
bi
réali
sculp
artis
an
po
mi
phot
édit
roman
écrivain ou
poète
mili
acteu
scén
ac
acteu
personnalité po
tradu
re
in

Pourquoi (si besoin d'arguments…)

Il est évident qu'on trouvera des hommes dans les archives.

Mais quid des femmes ? Et en quelles proportions ?

- Avoir conscience des biais et implicites des descriptions archivistiques pour les éroder un peu
- Enrichir nos données (inventaires, index) pour permettre aux chercheurs d'aller au-delà des seules recherches unitaires et des personnes « célèbres »



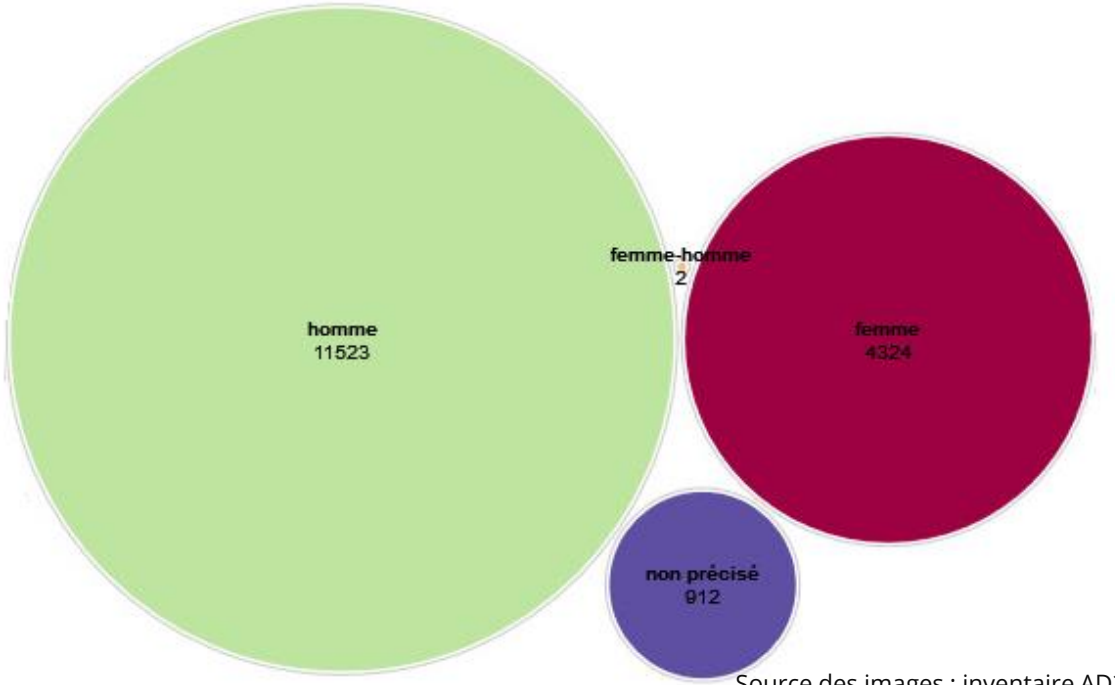
- JEAN IV
- JEAN V
- JEAN-EMILIEN
- JEAN-LOUIS 2000
- JEANNE CHABOT
- JEANNEAU (M.)
- JEANNENEY (Jean Marcel)
- JEANTIL (Sophie)
- JEGADO (Hélène)

A

Aage (de l'). Voir L'Aage.
Abalain (Jean), 227.
Abalan (François), 264.
Abbayes de filles, 3.
Abbé (Guy-Pierre), 267.
Abéguillé (Françoise), *62*.
— (Jean), *62*.
Abgrall (Alain), 233, 241.
— (Jean), 187, 259, 267.
— (Nicolas), 213.
Abhamon (Jean-Abgrall) dit Callac, 262.
Abhervé (Allain), *61*.
— (Yves), 15.
Abiven de Feunteun-Leas (François), 295.
Abjean (Nicolas), 265.
Abraham (Michel), 222.
— (Yves), 6.
Accouchements clandestins, 279.

Index du tome 1 des inventaires sommaires des séries A et B (1933) – AD du Finistère.

[> En savoir plus sur cet exemple]



Source des images : inventaire AD29 ; visualisations Maïwenn Bourdic

Quelques recettes possibles (parmi d'autres)

- Excel / Calc (filtres)
- OpenRefine*
- Wikidata (alignement avec des référentiels externes / via OpenRefine, pour les VIP)
- Python (script ANMT)

Source des images : logos sur Commons

OpenRefine

- Import de données CSV / XML / TXT…
- Filtres et facettes puissantes
- Modélisation libre de l'export des données libres (dont CSV / XML / TXT / Wikicode…)
- Logiciel open source, évolutif, en français, installable sans droits administrateur

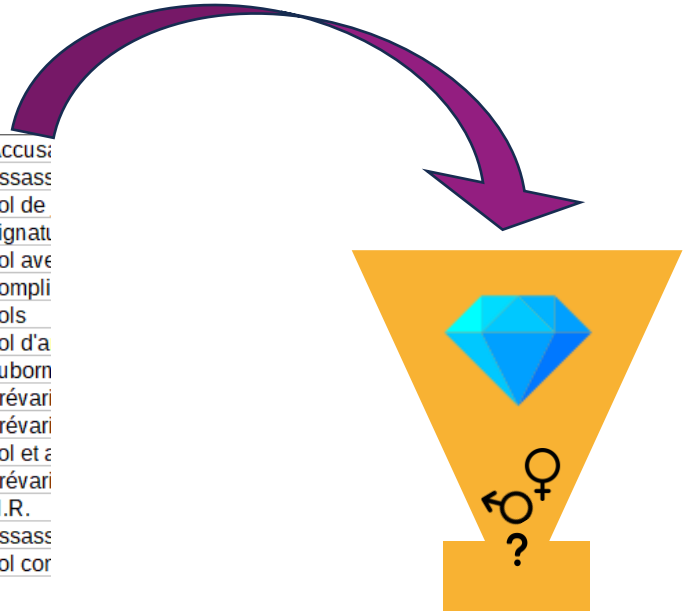
Tutoriels OpenRefine

« Reconnaître » les femmes

- Suffixe de prénoms
  - …-ie
  - …-ine (sauf Antoine)
  - …-elle
- Qualificatifs
  - veuve, femme, épouse, divorcée, demoiselle, baronne…
- Métiers
  - …-euse
  - …-ière
  - …-ère
  - …-onne
  - …-ande
  - …-ante

Recette OpenRefine

référence f	Noms des accusés	Prénoms	Ages	Qualités	Demeures	Juridiction	Accusa
76-77	Begnier	Michel	24	scieur de long		Châtelet	assass
76-77	Begon	Denise	26	couturière	Aurillac	Aurillac	vol de
76-77	Begue	André	40	manouvrier		Troyes	signatu
76-77	Begue	Claude		Journalier		Verneuil [V	vol ave
76-77	Beguier	Joseph	37	serger		Saumur	compli
76-77	Beguignon	René	24	tisserand		Chartres	vols
76-77	Beguignot	François	28	charretier	Chaumont	Chaumont	vol d'a
76-77	Beguin	Claude	31	pêcheur	L'Ile-Saint-	baillage du	suborn
76-77	Beguin	François	40	huissier des mo	faubourg S	en la Cour	prévari
76-77	Beguin	François		huissier des mo	faubourg S	en la Cour	prévari
76-77	Beguin	Jacques		manouvrier		Saint-Pierr	vol et a
76-77	Beguin	Jean Jacq	36	exempt		Châtelet	prévari
76-77	Beguin, veuve Raimbau	Marie				marquisat	N.R.
76-77	Beguin, femme Delafoss	Madeleine	58			Senlis	assass
76-77	Behère	Catherine	23	lingère	rue Jean d	Châtelet	vol cor



Un fichier d'indexation CSV (1750-1780) – 22 588 entrées

Arrêts criminels du Parlement de Paris (18e s.)

Source des images : flaticon.com ; logo OpenRefine sur Commons

Interface d'OpenRefine

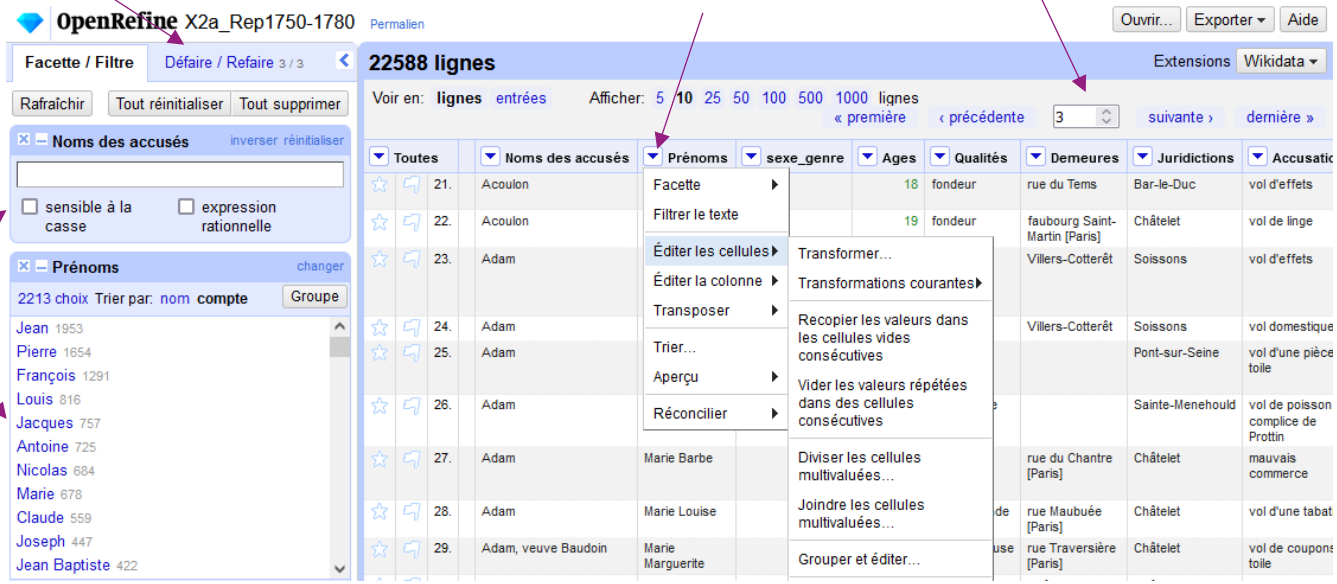
Pagination

Export / extension des données

Enregistrement de toutes les modifications

Menu déroulant pour manipuler les données

Filtres* et facettes** cumulables pour créer des ensembles de données sur lesquelles on va agir



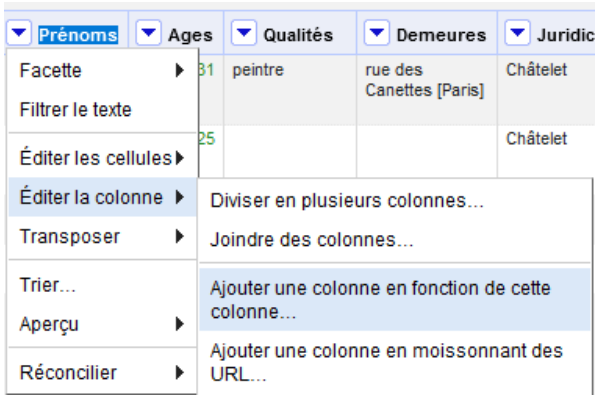
Filtres* : recherche d'une chaîne de caractère précise dans la cellule

VS Facettes** : recherche d'occurrences totalement identiques

Avant tout chose…

Insérer une nouvelle colonne *sexe_genre* :

éditer la colonne / ajouter une colonne… / ""



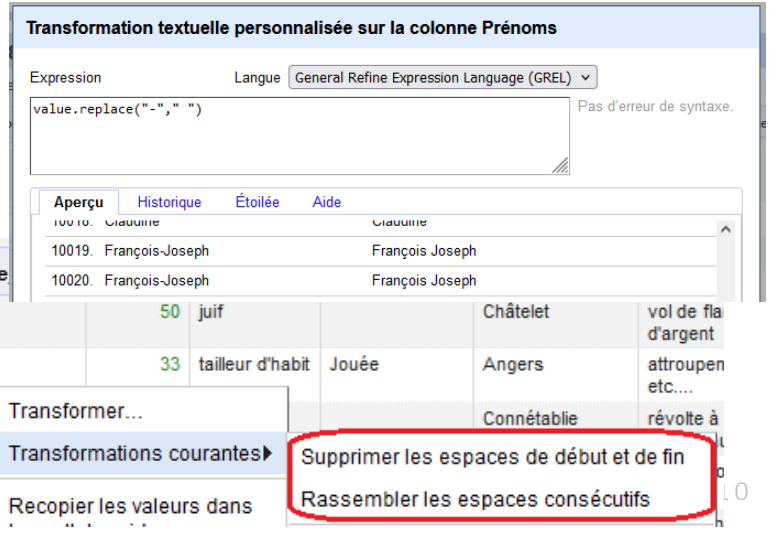
Optimiser rapidement les données* (surtout les prénoms) :

- Remplacer les - et , par des espaces :

  éditer les cellules / transformer : value.replace("-"," ")

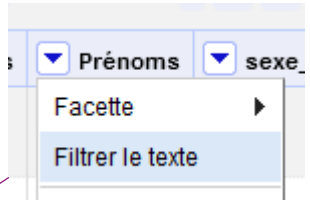
- Supprimer les espaces invisibles :

éditer les cellules / transformations courantes



*Si besoin, dupliquer les colonnes d'origine pour les conserver

Filtre de texte



OpenRefine X2a_Rep1750-1780 Permalien

Facette / Filtre | Défaire / Refaire 3 / 3

Rafraîchir | Tout réinitialiser | Tout supprimer

Noms des accusés inverser réinitialiser

veuve 1

sensible à la casse | expression rationnelle

sexe_genre changer

0 choix Trier par: nom compte | Groupe

(blank) 405

Facette par nombre de choix

Facette textuelle

405 matching lignes (22588 total)

Voir en: lignes entrées Afficher: 5 10 25 50 100 500 1000 lignes « premiè

Toutes		Noms des accusés	Prénoms	sexe_genre	Ages	Qualités	Demeures	Juric
2.		Abéllard, veuve Gigault	Louise		33	tailleur d'habit	Jouée	Angers
29.		Adam, veuve Baudoin	Marie Marguerite		27	blanchisseuse	rue Traversière [Paris]	Châtelet
35.		Adiot, veuve Mauberger	Marthe		53	couturière		Loudun
49.		Affequin, veuve Touze	Marie Anne					
56.		Agogne, veuve Abrué (morte)	Anne					
86.		Albé, veuve l'Epinasse	Catherine		65	blanchisseuse		Orléans

edit 2

Type de donnée : texte

femme 3

4

Appliquer | Appliquer à toutes les cellules identiques

Veuve / Femme / Epouse / Sœur : 1729 lignes genrées

Ajout de l'information en masse

Filtre de texte – jouer sur les suffixes

« Identifier les prénoms terminant par -ine

métiers -euse

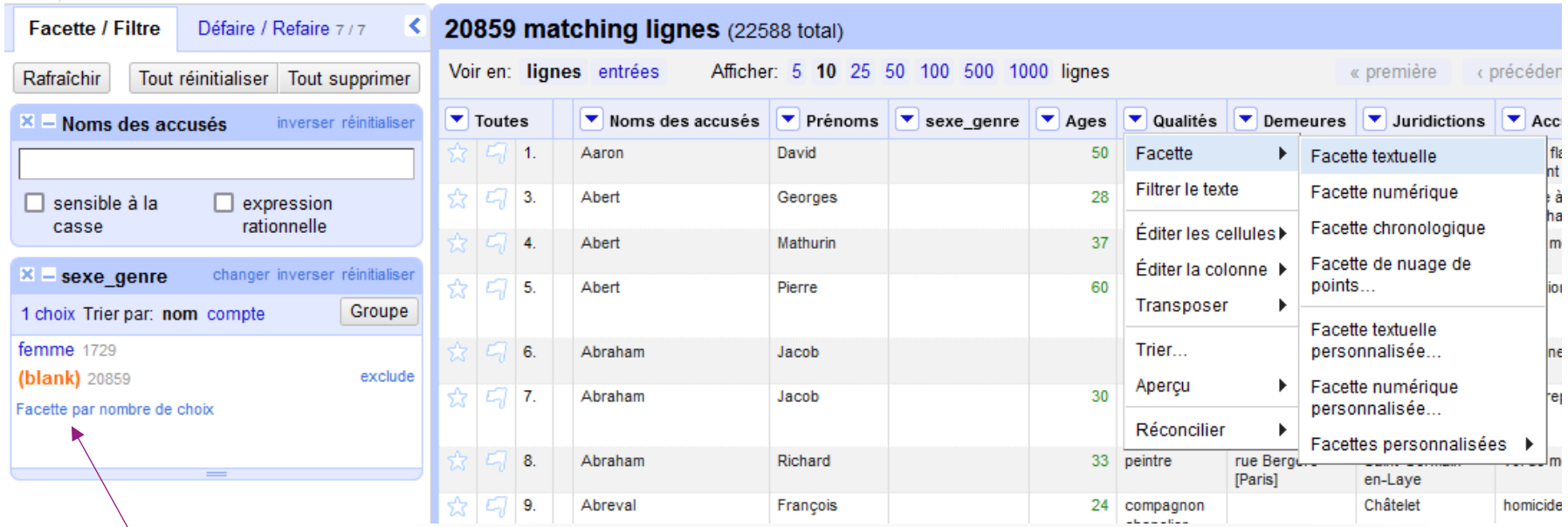
…

sauf … ! »

Garder un œil de contrôle avec une facette par mot



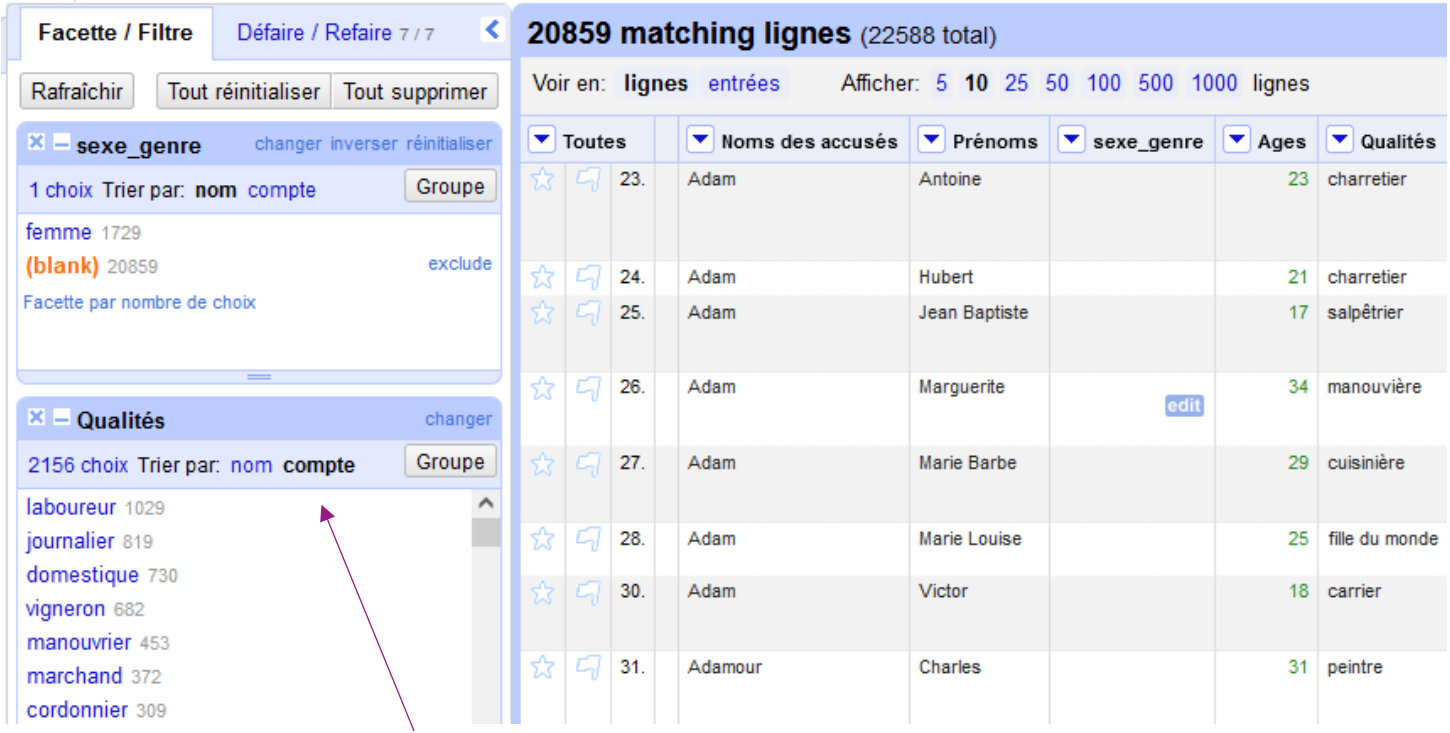
Jouer avec les facettes



Uniquement les lignes dont la colonne sexe_genre n'est pas renseignée

But du *jeu* : faire diminuer le nombre *blank* en un minimum de clic

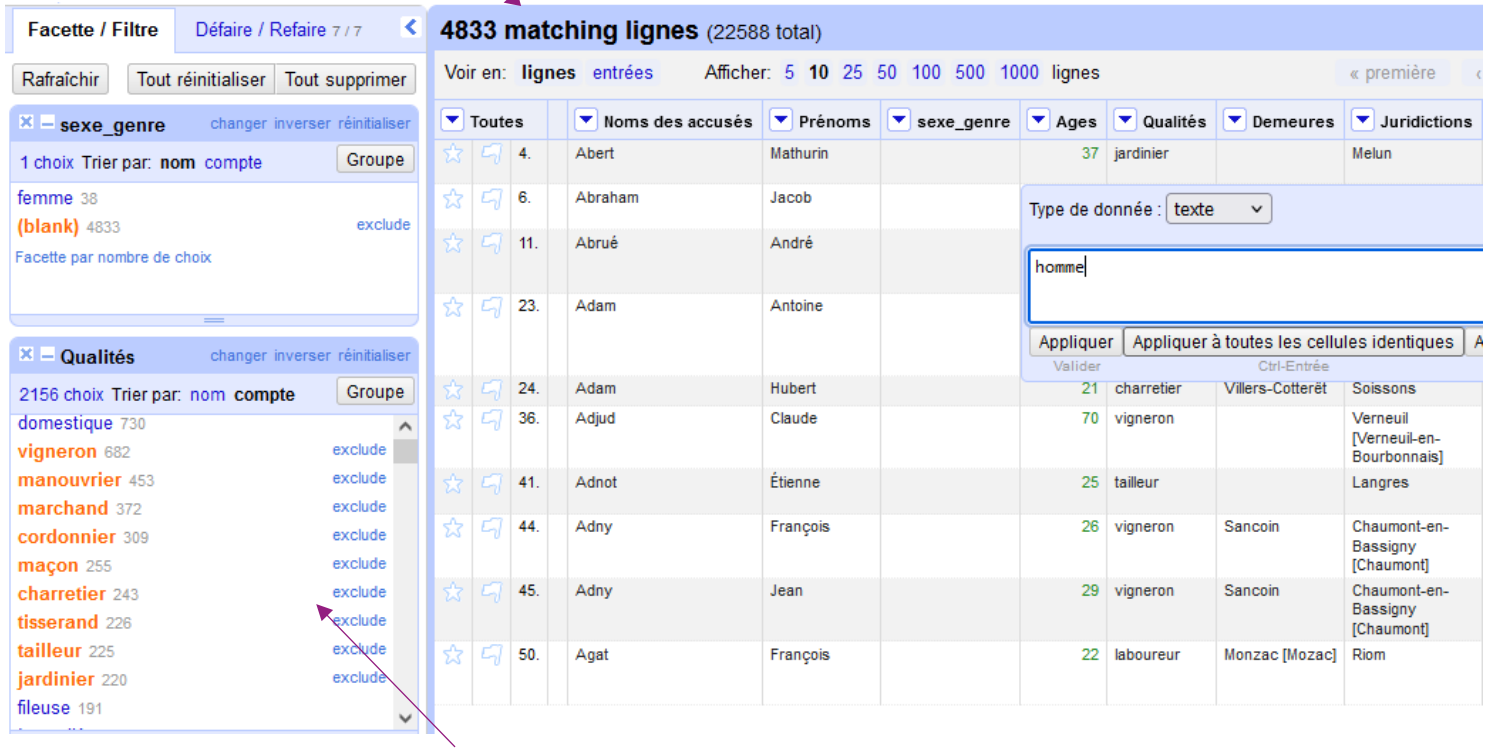
Facette simple



Tri décroissant par nombre d'occurrence

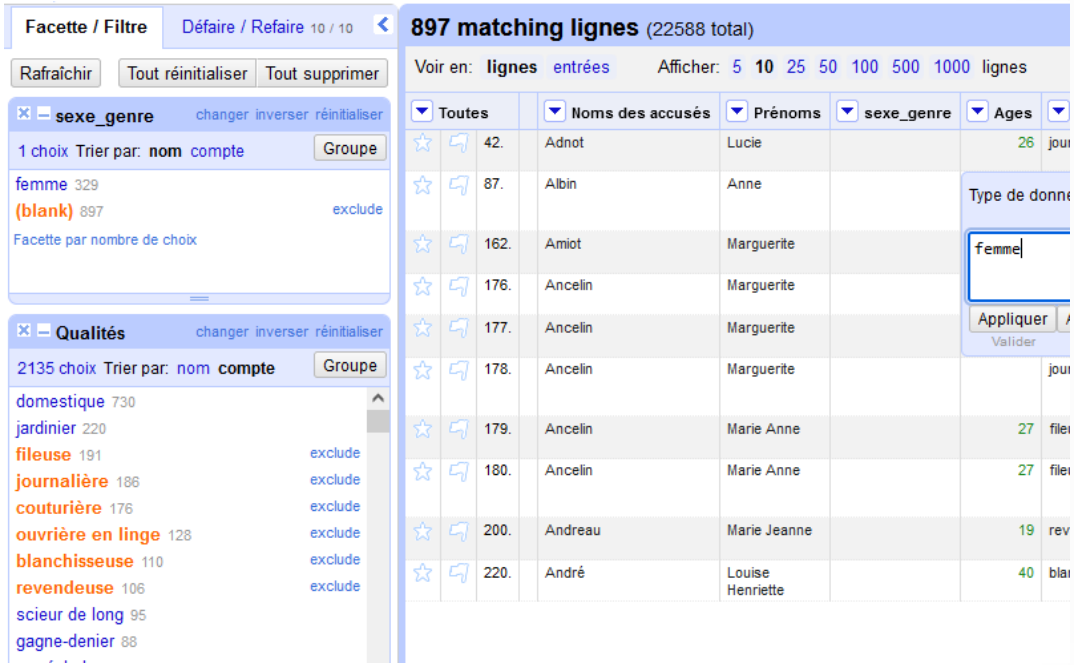
Nombre de lignes qui vont être renseignées en masse

Facette simple



Cumul d'ensembles de données avec une qualité genrée au masculin

Facette simple

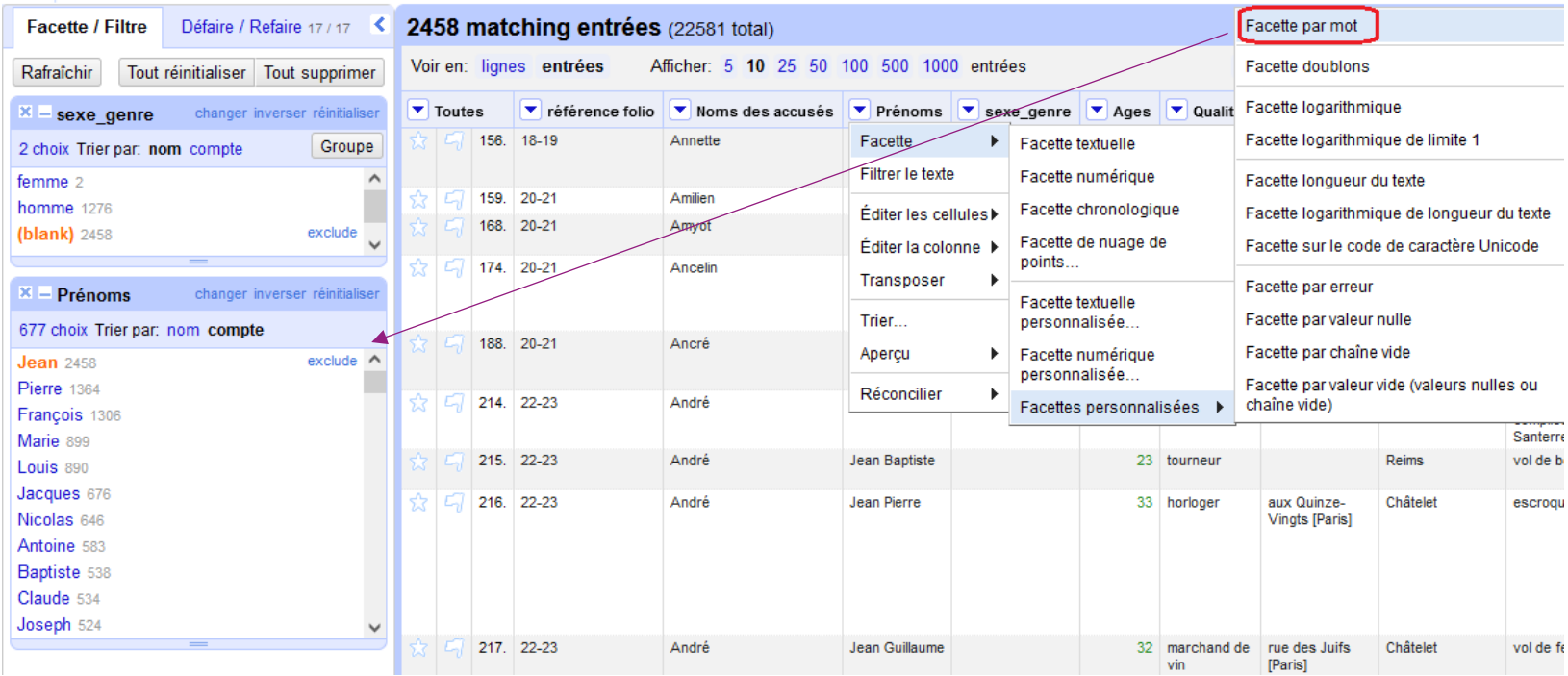


En 8 manipulations de données, 66% du corpus genré



Facette par mot

Idéal pour la colonne *Prénoms*



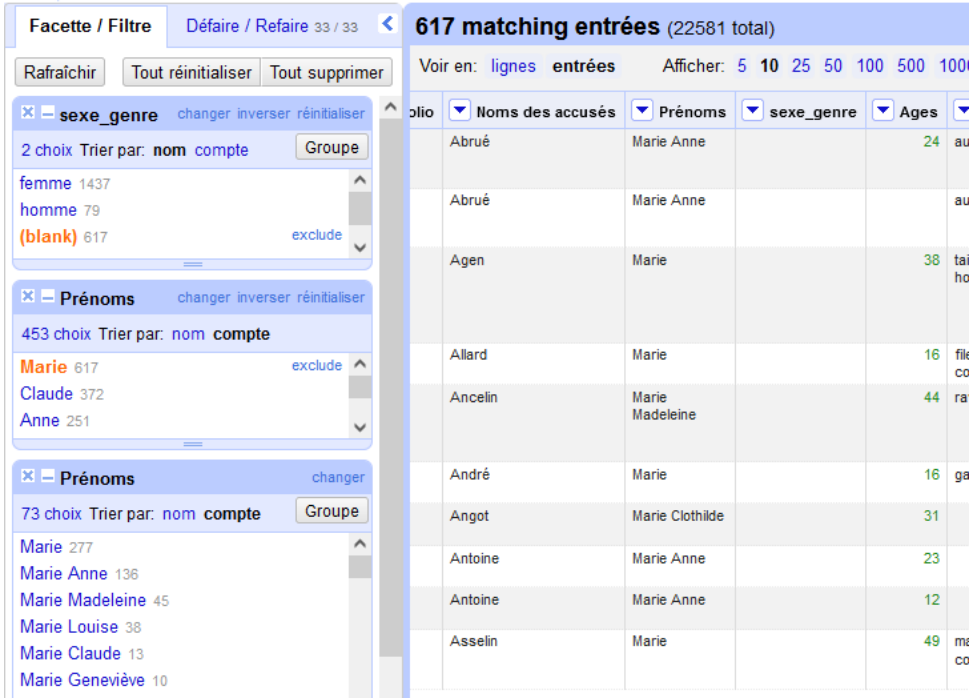
Facette par mot (colonne Prénoms)

Combiner les facettes pour gérer les prénoms épicènes :

Facette par mot « Marie »

+

Facette textuelle



Fignoler à la main

Nombre d'occurrences réduites, métiers ou prénoms mixtes, …



23 matching entrées (22581 total)

Voir en: lignes **entrées** Afficher: 5 10 **25** 50 100 500 1000 entrées «

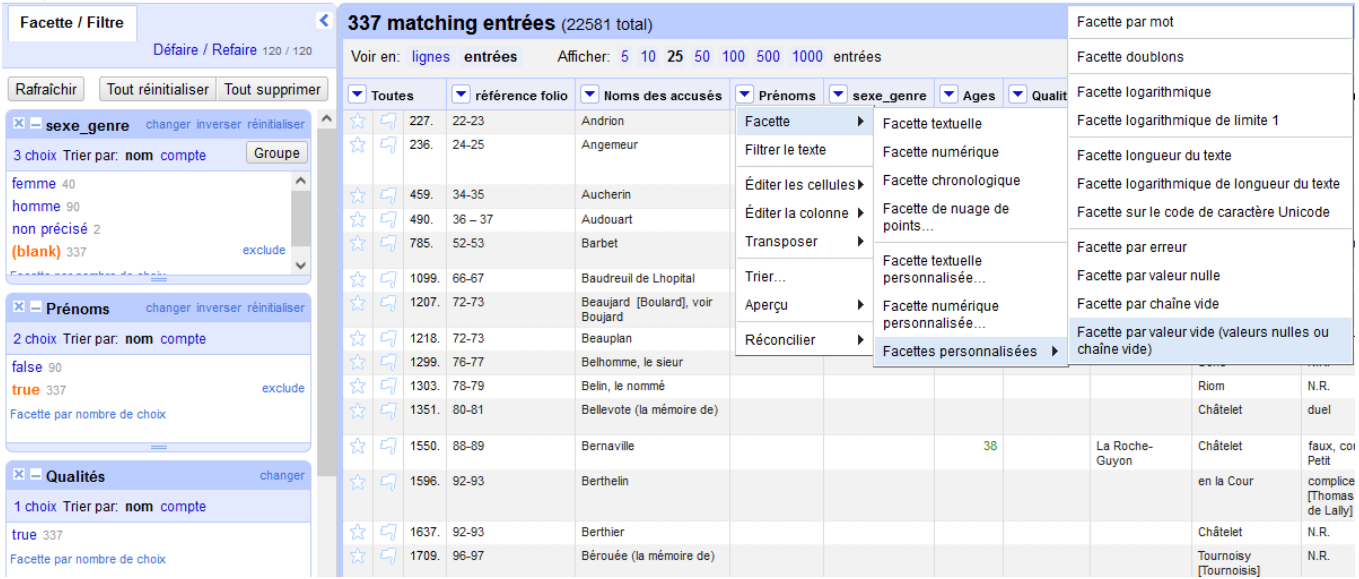
Toutes	référence folio	Noms des accusés	Prénoms	sexe_genre	Ages	Qualités
2639.	123	Boullanger	Aimé		33	domestique
3648.	149	Capoin, le nommé		edit	17	domestique
7104.	241	De Voisine	Marie Denise	femme	22	domestique
8486.	280	Favier	Gilberte	femme	20	domestique
8782.	288	Florent	Basile		20	domestique
9239.	300	Froc			23	domestique
10769.	342	Guerard	Victoire		25	domestique

Éditer cette cellule

S’autoriser à ne pas choisir (« non précisé »)

- Absence de prénom et d’informations identifiantes
- Prénoms épicènes

Facettes par valeur vide (*blank*)

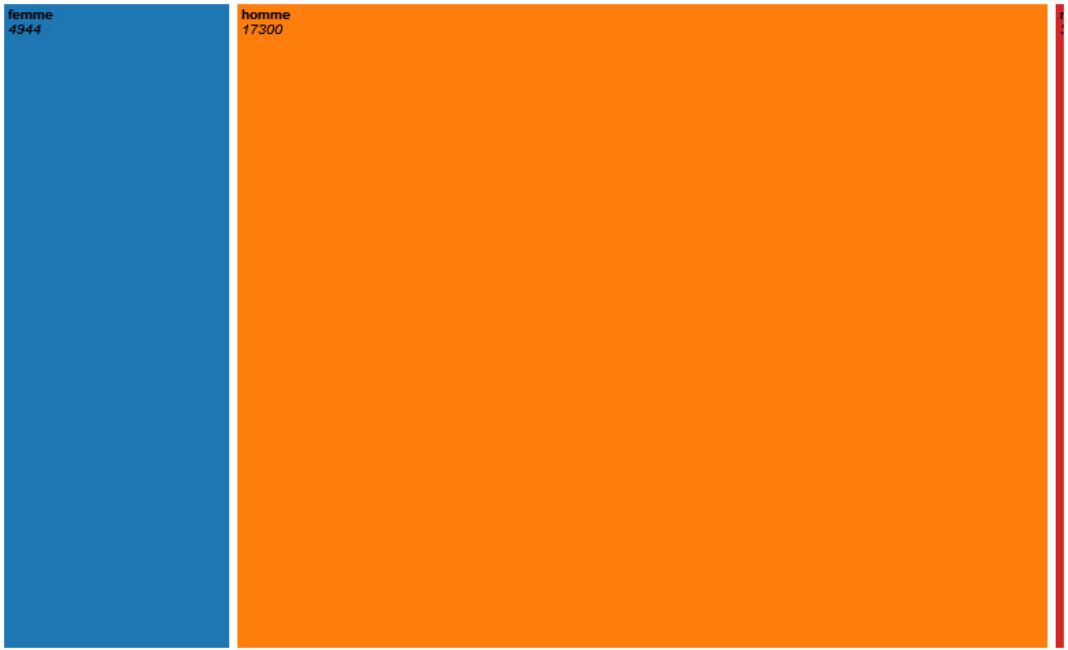


En moins de 2 heures…



Plus de 22500 lignes genrées

22% de femmes



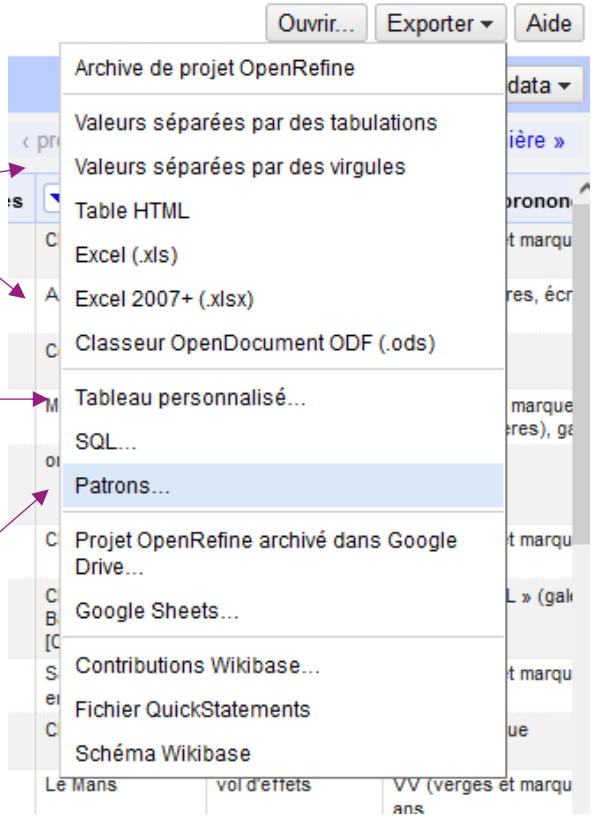
Source de l'image : graph avec RawGraph – Maïwenn Bourdic

Récupérer les données

Formats d'export classiques (CSV, XLS, ODS)

Tableau personnalisé : paramétrage des champs exportés, de leur ordre et du format d'export

Patrons : personnalisation de l'export



Récupérer les données

Export du modèle

Modèle de ligne

```
"Noms des accusés" : {{jsonize(cells["Noms des accusés"].value
"Prénoms" : {{jsonize(cells["Prénoms"].value)}},
"sexe_genre" : {{jsonize(cells["sexe_genre"].value)}}
```

Modélisation des données

Aperçu de l'export

```
"Noms des accusés" : "Aaron",
"Prénoms" : "David",
"sexe_genre" : "homme"

"Noms des accusés" : "Abéllard, veuve Gigault",
"Prénoms" : "Louise",
"sexe_genre" : "femme"
```

Modélisation libre de l'encodage…
y compris en XML (EAD)

```
<c><did>
<unittitle><persname>
{{jsonize(cells["Noms des accusés"].value).replace('"','')}},
{{jsonize(cells["Prénoms"].value).replace('"',“)}}
</persname><unittitle>
</did>
<bioghist><p>Sexe ou genre :
{{jsonize(cells["sexe_genre"].value).replace('"','')}}</p>
</bioghist></c>
```

Ressources en ligne

Patrimoine et numérique, « Où sont les femmes ? Ou comment genrer nos données patrimoniales »

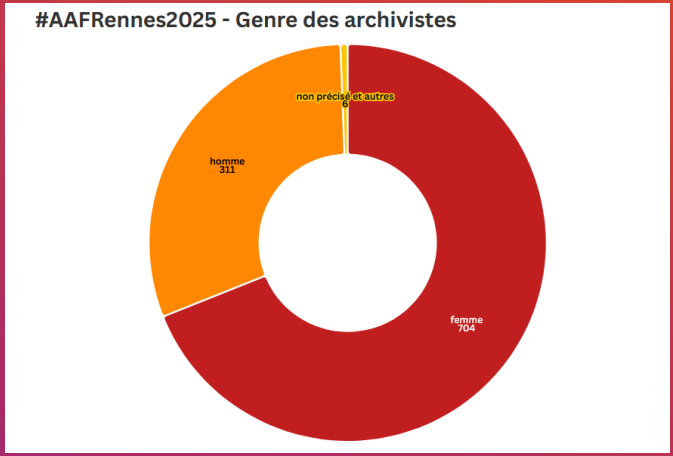
Patrimoine et numérique, Tutoriels et cas d'usage OpenRefine

Patrimoine et numérique, « Des gens dans les archives (datavisualisation des femmes avec identifiant FranceArchives Agent) »

d'Aïeux et d'ailleurs, « Sur les traces de Cartouche - visualisation de données de l'inventaire X2a (1700-1720) »

ANMT, script Python

**Divette : pourrait être Denysis (en tournée à Marseille)*



Merci

de votre attention

Source de l'image : Portraits d'archivistes
#AAFRennes2025- Maïwenn Bourdic

association des archivistes français

avec attention
archives, archivistes & sociétés

Le Forum se déroule avec le soutien de :

ancestry

julhiet sterwen

PROMUSEUM
Une équipe à vos côtés au service des collections

mipih SIB

La captation du Forum des archivistes est réalisée par Ikonokast

Ikonokast