

Mashup Base Joconde

Mathilde Liteau, Claire Jahan et Juliette Janes

Mai 2021

1 Introduction

Le dataset de travail, extrait de la Base Joconde¹, est composé de 60 000 lignes et 34 colonnes. Cette imposante taille a rendu le travail de nettoyage à réaliser particulièrement compliqué, notamment pour nos ordinateurs. La base a été remplie à la main, par chacun des musées, et, les colonnes pouvant contenir plusieurs informations différentes, les cellules ne sont pas du tout normalisées. Ainsi, face à la quantité d'informations difficiles à réorganiser et dans l'idée de réduire une étape chronophage, nous avons voulu récupérer des informations plus propres sur Wikidata².

Nous avons donc obtenu deux fichiers csv. Le premier, sur les oeuvres de la Base Joconde contenues dans Wikidata, soit 15 759 éléments, présente à la fois des colonnes qui nous semblaient très compliquées à nettoyer dans notre dataset d'origine et des colonnes d'informations qu'il ne contenait pas, à l'instar des images des oeuvres. Le second, sur les auteurs des oeuvres de la Base Joconde contenues dans Wikidata, est essentiellement composé d'éléments qui n'étaient dans la base. La plupart des images récupérées dans Wikidata étant des images de peinture, nous avons décidé de nous focaliser sur celles-ci, dans l'idée de réaliser une visualisation mettant l'accent sur une galerie d'images.

Il aurait été également intéressant d'associer notre extrait de la Base Joconde, à la base Muséofile³, qui répertorie les musées de France. Cette idée est d'autant plus intéressante que le set de données possède une colonne comprenant les identifiants Muséofile des musées, ce qui aurait permis de lier les deux fichiers csv facilement. Cependant, la base Muséofile est directement liée à la Base Joconde et cette jointure n'aurait créé que des doublons.

2 Préparation des données

Le nettoyage des données a été exclusivement réalisé sur Dataiku⁴.

2.1 Extrait de la Base Joconde

La Base Joconde, dans sa globalité, est composée de 650 000 notices et décrit les oeuvres des institutions de conservation de France. L'extrait sur lequel nous avons tra-

1. <https://github.com/Juliettejns/baseJocondeMashup/tree/main/dataset/extraitJoconde/base-joconde-extrait.tar.xz>

2. https://www.wikidata.org/wiki/Wikidata:Main_Page

3. <https://data.culture.gouv.fr/explore/dataset/musees-de-france-base-museofile/information/>

4. <https://fr.wikipedia.org/wiki/Dataiku>

vaillé est d'environ 60 000 oeuvres, soit un peu moins de 10 % du set total. Nous avons conscience dès le début que le corpus de travail n'est pas du tout représentatif de la globalité de la base. Dans Dataiku, nous avons eu la possibilité de choisir quelle part de ce set nous allions pouvoir utiliser : nous avons sélectionné un panel de 50 % aléatoire.

Les csv manipulés étant trop importants pour GitHub, où nous avons entreposé tout notre devoir, nous avons utilisé une extension appelée GitHub LFS⁵ pour nous permettre de les ajouter au dépôt.

2.1.1 Nettoyage des données

Suppression des colonnes inutiles Sur les 34 colonnes de base, nous avons choisi de supprimer toutes celles ayant moins d'un tiers de données et qui nous semblaient facultatives, soit 11 colonnes. Nous n'avons donc pas conservé par exemple *Date de mise à jour* ou *Période de l'original copié*.

Normalisation des valeurs Sur les colonnes restantes, nous les avons avant tout renommées afin de les rendre toutes compréhensibles. Puis il a fallu nettoyer les données plus directement.

Normalisation des chiffres et nombres Afin de faciliter toute visualisation par la suite et étant donné que ces dates sont très disparates, nous avons retiré toute mention de mois ou de jour. Cela a été le cas par exemple sur *Date d'acquisition*. En ce qui concerne les mesures des oeuvres de la base, hauteurs et longueurs se trouvaient dans la même colonne : nous les avons donc séparées dans deux différentes.

Normalisation des termes Nous avons également dû travailler sur la normalisation des terminologies, notamment en ce qui concerne les colonnes *Ville* et *Région*. Cela est passé par une mise en majuscule de toutes les premières lettres de tous les termes, ainsi que leur uniformisation.

Colonnes à multiples informations Nous avons dû travailler à réduire les données sur de nombreuses colonnes de texte dans lesquelles avaient été agglomérées des informations qui étaient redondantes ou qui n'étaient pas équivalentes selon les lignes. Cela a été le cas pour des colonnes comme *géolocalisation*, *Domaine* ou *Statut juridique*. Par exemple, cette dernière comprenait aléatoirement le nom de l'institution qui conserve l'oeuvre, la ville et/ou la région de l'institution, le statut juridique, ainsi que la manière dont l'oeuvre avait été acquise. Pour ce qui était des deux premières informations, elles existaient par ailleurs dans d'autres colonnes beaucoup plus propres. Nous avons ensuite nettoyé ce qui restait (remplaçant ce qui n'existait pas par la mention "None") et séparé cela en deux colonnes distinctes. Ce qui a été vraiment problématique dans les colonnes de ce type était que les informations n'étaient jamais dans le même ordre, même lorsque cela concernait des oeuvres provenant de la même institution de conservation.

Dans l'idée de rendre plus aisée toute lecture de nos visualisations, nous nous sommes exclusivement penché sur des oeuvres qui avaient été créées par une seule personne : nous avons donc retiré toute oeuvre liée à plusieurs artistes. Nous avons par la suite nettoyé la colonne *Auteurs* en retirant tout ce qui n'était pas le nom et le prénom de ces derniers.

5. <https://git-lfs.github.com/>

Dédoublage de colonnes Toujours pour des raisons de visualisation finale, nous avons dédoublé sciemment certaines colonnes afin de créer des colonnes au contenu simplifié. Par exemple, cela a été le cas pour la colonne *Mode d'acquisition*, où nous avons de nombreuses choses différentes, ce qui, nous le savions, rendrait la lisibilité de tout graphe difficile. C'est pour cette raison que nous avons une colonne *Mode d'acquisition* et *Mode d'acquisition simplifiée*, où pour la première nous avons une entrée "dation" qu'on retrouvera sous "don" dans la seconde.

Dernières étapes avant la jointure Ces dernières modifications du dataset ont eu lieu lorsque toutes les précédentes avaient été réalisées et que les fichiers csv provenant des requêtes SPARQL avaient été également nettoyés. En effet, afin de faciliter la jointure avec les deux autres datasets, nous avons pris le parti de supprimer en amont toutes les lignes dont le domaine n'était pas la peinture, suite au constat de leur plus grande représentativité dans les images récupérées sur Wikidata. Dans la même lancée, nous avons également retiré plusieurs colonnes dont nous savions que les données n'étaient pas aussi complètes et/ou propres que celles récupérées dans Wikidata, comme pour les matériaux et les techniques.

2.1.2 Résultat

Élément supprimé	Élément conservé	Élément ajouté	Liste finale
Ancien dépôt Appellation Date d'acquisition Date de dépôt Découverte / Collecte Utilisation Dépôt Mesures Date de mise à jour Titre Epoque / style / mouvement Millésime Millésime d'utilisation Onomastique Précisions sur l'auteur Période de l'original copié Période Période d'utilisation Matériaux et techniques	Auteur Domaine Ecole Géographie historique Numéro musée de France Lieu de création / utilisation Lieu de conservation Identifiant base Museofile Référence Sujet Source de la représentation Statut juridique Dénomination Date d'import	Date_acquisition Mode_acquisition_simplifié Mode_acquisition Latitude Longitude	Référence Auteurs Date_acquisition Dénomination Entrée_base Domaine Ecole_pays Date_création Identifiant_Museofile Nom_officiel_musée Sujet_représenté Statut_juridique Mode_acquisition_simplifié Mode_acquisition Ville Région Latitude Longitude

TABLE 1 – Différents types de colonnes du dataset extrait de la Base Joconde

2.2 Wikidata

2.2.1 Récupération des données

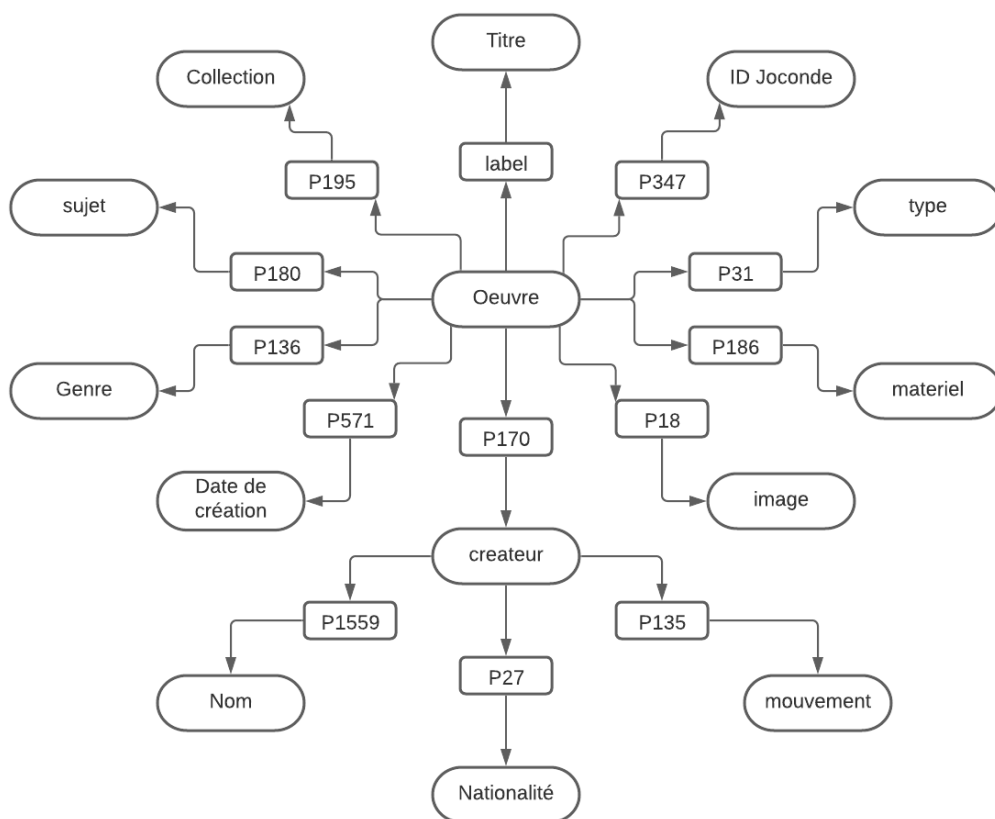


FIGURE 1 – Représentation de la structure de nos requêtes SPARQL

Ce schéma donne une vision globale des différents éléments qu'il est possible de récupérer sur Wikidata à propos de la Base Joconde. Réalisé par nos soins, il nous a notamment permis de bien comprendre la structure des triplets RDF et donc d'élaborer plus facilement des requêtes SPARQL fonctionnelles afin d'obtenir les différents éléments évoqués.

Dans un premier temps, on a ainsi pu récupérer des données qui n'étaient pas dans l'extrait de la base, à l'instar des liens vers des images des oeuvres et de leur genre artistique ainsi que des données sur les artistes, telles que les dates et lieux de naissance et de mort et les mouvements caractéristiques de ceux-ci.

Dans un second temps, face aux différentes données peu manipulables dans l'extrait de la Base Joconde à nettoyer, dû à leur importance et la variété de structure des cellules, nous avons décidé de récupérer certaines données qui, grâce aux requêtes SPARQL, se sont avérées plus normalisées, donc mieux structurées et plus faciles à uniformiser. C'est le cas par exemple des mesures, hauteur et largeur des oeuvres, qui ont toutes été récupérées séparément et en centimètres, des matériaux, qui ont pu être divisés entre techniques utilisées et support ou encore les types.

Dans un dernier temps, nous avons récupéré des données déjà contenues dans l'extrait de la Base Joconde, telles que les titres, dans un souci de certitude de ces informations. En effet, un certain nombre de titres était manquant, voire ne coïncidait pas.

2.2.2 Nettoyage des données

Concernant les oeuvres Le fichier csv contenant les oeuvres de la Base Joconde accessibles sur Wikidata est composé d'un peu plus de 15 000 lignes et de 10 colonnes, présentées ci-dessous :

Élément supprimé	Élément conservé	Élément ajouté	Liste finale
	identifiant titre nomcreateur datecreation largeur hauteur type matériel genre image		identifiant titre nomcreateur anneecreation largeur (cm) hauteur (cm) type techniques support genre image

TABLE 2 – Travail sur les colonnes de la requête sur les oeuvres

Si le résultat de cette requête reste particulièrement propre, il a tout de même fallu le nettoyer un peu. Nous n'avons d'abord conservé que ce qui ressemble à des peintures, ce qui a été fait en filtrant avec la colonne *type*. En ce qui concerne les dates, il a fallu nettoyer pour ne garder que l'année, dans l'idée de réaliser une visualisation à partir de cela. La colonne *matériel* a été un peu plus longue à restructurer, puisqu'il a fallu la diviser en deux colonnes : une première contenant les techniques de peinture et une seconde les supports. Les genres ont été nettoyés pour normaliser les différents termes et une colonne *genre simplifié* a été créée.

Concernant les auteurs Le fichier csv contenant les auteurs ayant réalisé des oeuvres de la Base Joconde accessibles sur Wikidata est composé de 500 lignes et de 6 colonnes, présentées ci-dessous :

Élément supprimé	Élément conservé	Élément ajouté	Liste finale
	nomcreateur datenaissance datemort nationalité lieunaissance mouvement lieumort	anneenaissance	nomcreateur datenaissance datemort anneenaissance lieunaissance mouvement lieumort nationalité

TABLE 3 – Travail sur les colonnes de la requête sur les auteurs

Là aussi, le fait d’avoir requêté assez précisément dans Wikidata a permis d’obtenir des données assez propres et ne nécessitant pas trop de travail de nettoyage. Les dates de naissance et de mort ont été nettoyées pour ne garder également que l’année. Les lieux de naissance et de mort ont été nettoyés dans le but de conserver uniquement les noms de villes et non plus les quartiers. Les mouvements ont été normalisés.

2.3 Association des données

Une fois les trois sets de données nettoyés, il a fallu les associer. Pour ce faire, nous avons tout d’abord projeté de réaliser un *left join*. Cela permettait de conserver toutes les données de l’extrait de la Base Joconde correspondant aux peintures, et de leur associer, quand cela était possible, les informations récupérées sur Wikidata. Les données de l’extrait de la base ont été jointes aux oeuvres de Wikidata grâce à leurs identifiants. Par la suite, les données sur les auteurs ont été associées au dataset obtenu, grâce aux prénoms et noms que nous avons également récupérés, de la même façon, dans le fichier csv des oeuvres de Wikidata. Nous avons alors obtenu un csv de 17 000 lignes, dont seulement quelques petites centaines d’entre elles correspondaient aux données récupérées sur Wikidata.

Nous avons alors réalisé que les données de l’extrait de la Base Joconde sur lesquelles nous avons travaillé et les données sur la base disponibles sur Wikidata ne coïncidaient absolument pas, ou alors très peu. Malgré cela, nous avons choisi de continuer à travailler dessus, en préférant faire un *inner-join* des données, soit uniquement les données présentes à la fois dans notre extrait de la Base Joconde et dans Wikidata. Cela nous a donné un dataset d’environ 700 lignes. Cela nous a permis de récupérer un set de données dans lequel les données de Wikidata ne sont pas à peine visibles, sachant que celles-ci correspondent des colonnes assez importantes dans nos idées de visualisation.

En dépit de la réduction drastique du dataset, qui n’est donc plus vraiment représentatif de la base Joconde, nous en avons finalement obtenu un nettoyé qui reste assez important pour réaliser des visualisations. Nous avons donc travaillé à la mise en graphiques et en images d’un dataset précis, correspondant à l’intersection entre un extrait de la Base Joconde et des données complétant celle-ci rendues accessibles sur Wikidata.

3 Visualisations

Avec cette base de données, nous avons de nombreuses idées de visualisations. La première et la plus évidente était la présentation d’images des oeuvres, même si au début l’idée de ne travailler que sur des peintures n’était pas encore toute tracée. Deux autres axes sont apparus : l’étude de l’oeuvre et celle de l’artiste selon les données récoltées.

En ce qui concerne les oeuvres, nous avons pour objectif de travailler avec les dates de manière importante afin de pouvoir analyser les évolutions au cours du temps. A côté de cela, il nous paraissait important d’étudier les types de ces oeuvres, c’est-à-dire leur genre, le mouvement dans lequel elles s’inscrivent, leur support, la technique utilisée pour leur création...

Pour les artistes, il nous paraissait nécessaire de les décrire. Nous avons pour idée d’utiliser les dates de naissance et de mort. Nous souhaitons également étudier leur migration géographique en comparant les lieux de mort et les lieux de naissance. Il nous fallait également étudier les nationalités, ainsi que les influences qu’ils avaient subi.

Nous savions ce que nous avions envie de montrer, mais lorsqu'il a fallu réaliser la visualisation, les questions de forme, mais aussi sur la donnée et ses possibilités de rendu sont devenues prégnantes. Certaines des visualisations auxquelles nous avons pensé ne fonctionnaient pas, faute de données suffisantes ou de visualisation assez claire. A l'inverse, le nettoyage nous a donné de nouvelles possibilités. Nous avons pu les mettre en oeuvre dans différents cadres : Tableau Public⁶ et Palladio⁷.

3.1 Tableau Public

Tableau Public nous a permis de réaliser des trois tableaux de bord décrivant nos données :

- Une présentation générale⁸ des données récupérées, permettant d'avoir un premier aperçu assez global de notre set, notamment concernant les musées, peintres et périodes représentés dans notre dataset.
- Un focus sur les aspects juridiques⁹ de notre dataset, tels que les types et dates d'acquisition et les propriétaires des oeuvres.
- Une étude des genres et matérialité¹⁰ de nos données, s'attachant plus aux détails et s'intéressant aux genres et dimensions des oeuvres, aux nationalités et écoles des peintres, ainsi qu'aux supports et techniques utilisées.

3.1.1 Tableau de bord 1 : Présentation générale du dataset

Ce premier tableau de bord permet de donner un aperçu rapide des données de ce dataset, en mettant l'emphasis sur les institutions de conservation, auteurs et périodes présents.

Carte des musées La carte des musées était un incontournable [3]. Il était nécessaire de représenter la géographie de ces oeuvres par le biais de leur institution de conservation. Nous pouvons donc voir les différents clusters, les différents lieux "vides". Il nous est apparu important également de visualiser la quantité de ces oeuvres au sein de chaque endroit. Cela donne à voir l'importance de la représentativité des musées au sein de notre dataset. Pour ce faire, nous avons utilisé les colonnes *latitude* et *longitude*, afin de localiser les musées sur une carte.

Nombres d'oeuvres par artiste Ce graphique [4] donne à voir de la même manière que le précédent l'importance de chaque artiste dans notre dataset par rapport au nombre d'oeuvres qui lui correspondent. Pour cette visualisation, nous avons mobilisé la colonne *nom des auteurs*.

6. <https://public.tableau.com/s/>

7. <https://hdlab.stanford.edu/palladio/>

8. <https://public.tableau.com/profile/jahan3241#!/vizhome/EtudesurdesdonneesrcupresdelaBaseJoconde/PrsentationgnraledesdonneesrcupressurlaBaseJoconde>

9. <https://public.tableau.com/profile/jahan3241#!/vizhome/EtudesurdesdonneesrcupresdelaBaseJoconde/AspectjuridiquedesdonneesrcupressurlaBaseJocondes>

10. <https://public.tableau.com/profile/jahan3241#!/vizhome/EtudesurdesdonneesrcupresdelaBaseJoconde/Etudesurlegenreetlamaterialit>

Date de création des oeuvres et de leur insertion dans la Base Joconde Ces deux graphiques [5] [6] ont été ajoutés côte à côte dans le tableau de bord. L'idée est de montrer cette double temporalité qui a son importance, que pour savoir sur quel type d'oeuvres on peut avoir affaire, tant pour rendre compte de la temporalité de la constitution de la Base Joconde, ou du moins de ce que nous en avons extrait. Nous avons utilisé, respectivement pour chacun des graphiques, les colonnes *date de création* et *date d'entrée dans la base* auquel nous avons associé le nombre total d'oeuvres de notre dataset.

Nous avons tout dans un premier temps, eu l'idée d'insérer ces deux types de données dans un même graphique, ce qui a donné la figure 2.

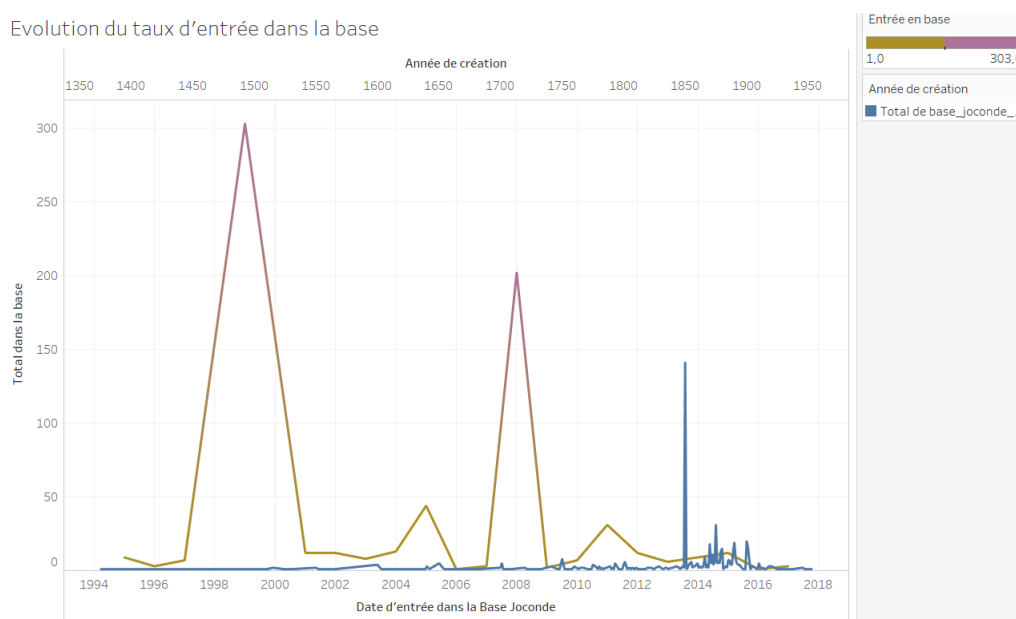


FIGURE 2 – Double graphique date d'entrée dans la base et date de création des oeuvres

Les deux plages de dates ne correspondant pas, nous n'avons pas réussi à n'avoir qu'un seul axe des abscisses pour notre graphique. Ainsi, cette double axe des abscisses ne nous convenant pas, nous avons préféré, dans un souci de meilleure compréhension du graphique pour les personnes extérieures au projet, de nous en tenir à deux graphiques différents.

Pour conclure Les deux premières visualisations ont pour but non seulement de permettre d'analyser les données présentées, mais également de rendre compte de ce que comprend notre dataset. Cela paraissait des plus importants afin de rendre compte à toute personne extérieure de ce à quoi il avait affaire. Cette dernière idée est légèrement moins présente concernant les deux derniers graphiques, qui donnent tout de même la fourchette temporelle des oeuvres de ce dataset, ainsi que celle de leur insertion dans la Base Joconde.

3.1.2 Tableau de bord 2 : Aspects juridiques

Ce second tableau de bord présente les différentes informations juridiques concernant les oeuvres, telles que les dates d'acquisition par le musée, les types de propriétaires et la méthode d'acquisition.

Dates d'acquisition Cette visualisation [7] a pour simple but de montrer la temporalité de l'acquisition des oeuvres de notre dataset et son évolution. Elle a été construite en utilisant la colonne *année d'acquisition* à laquelle on a associé le nombre total d'oeuvres du dataset par *date d'acquisition*.

Statut juridique des oeuvres Cette visualisation [9] a pour but de rendre les différentes possibilités de statut juridique des oeuvres. Il est intéressant de noter qu'il en existait d'autres, mais qu'elles ont disparu en deux temps : la première fois, lors du choix du domaine de la peinture et une seconde fois lors de la jonction avec les deux base de Wikidata. Pour ce faire, nous avons utilisé la colonne *statut_juridique* associée au nombre total d'oeuvres par type de statut dans le dataset.

Mode d'acquisition En ce qui concerne cette visualisation [8], il nous paraissait important de la rendre la plus simple possible. C'est pour cette raison que nous avons décidé de bien séparer les différents modes d'acquisition en les regroupant en catégories. C'est pour cette raison que l'on retrouve "achat", "don", "mode d'acquisition inconnu" et "saisie révolutionnaire" sur la partie haute de la visualisation, tandis que sur la partie base, les chiffres sont explicités.

Pour conclure Il nous paraissait important d'avoir cette vision juridique de nos données étant donné l'importance que cela avait dans la Base Joconde originelle.

3.1.3 Tableau de bord 3 : Étude du genre et de la matérialité des oeuvres

Ce dernier tableau de bord donne un aperçu des possibilités d'exploitation des données de notre dataset plus poussé, notamment en s'intéressant à la relation entre les dimensions et le genre de l'oeuvre, l'évolution des supports et techniques dans le temps ou encore en comparant les nationalités des auteurs et leur école.

Évolution des supports et techniques de peinture dans le temps Par cette visualisation, nous avons l'ambition de montrer au mieux la diversité des supports et des techniques qui leur sont associées. Ce qui était intéressant également était de montrer laquelle de ces associations était la plus représentée, et ce, au cours du temps [10].

Relation entre les dimensions d'une oeuvre et son genre Nous ne pensions pas du tout à réaliser cette association, mais nous souhaitons travailler au départ avec les dimensions des oeuvres au cours du temps. Cette représentation donne à voir le lien entre les dimensions et le genre de l'oeuvre. Cela permet également de voir si un portrait est bien en format portrait, ou une nature morte en format paysage [11].

Comparaison entre les nationalités des peintres et leur école Il paraissait des plus intéressant de faire le lien entre ces deux informations, afin d'avoir une idée de la mobilité des influences artistiques [12].

3.2 Palladio

Grâce à Palladio, nous avons également pu présenter les images des oeuvres de notre dataset que nous avons récupérées sur Wikidata. Cette galerie d'images est accessible en récupérant le fichier json¹¹ disponible dans notre dépôt et en le lançant dans Palladio.

Nous voulions également pouvoir rendre compte de ce que nous avons effectivement fait dans Palladio. Ci-dessous se trouvent quelques images qui rendent compte de la visualisation.

Comme on peut le voir sur cette première image [13], nous avons pu lier les images, ainsi que leur titre, leur sujet et leur artiste. Cela crée de cette manière des cartels assez intéressants et plutôt lisibles.

Cela correspond à quatre colonnes dans la base de données. Ces données ont été choisies dans l'optique de rappeler ce que l'on trouve dans les musées, bien que l'on soit limité en quantité. Comme on peut le voir dans cet encart en haut à droite de cette image, il n'y a pas beaucoup de champs disponibles. Il n'en reste pas moins que s'il on avait eu la possibilité d'ajouter plus d'informations, cela aurait pu rendre une visualisation déjà lourde, encore plus conséquente et difficile à charger. La visualisation par année de création donne une vision de l'évolution de la peinture dans le temps, en ce qui concerne les sujets, les genres, les couleurs [14]...

En outre, n'ayant pas des images pour toutes les oeuvres, nous avons décidé de filtrer toute oeuvre qui n'aurait pas le lien vers une image. Cela a été possible grâce à la colonne de type boolean créée en amont selon ce critère. Ce filtre a pu être mis en place grâce à l'encart qui se trouve en bas de la page de visualisation de Palladio, comme montré sur cette image [15].

4 Conclusion

Au cours de notre travail, il nous a été possible d'observer quelques grands phénomènes à partir de l'étude de notre dataset. Cependant, il est important de garder à l'esprit que celui-ci n'est pas représentatif de l'ensemble de la Base Joconde ni même de l'échantillon que nous avons récupéré. D'abord, le ressèment du travail autour des peintures uniquement a induit une sureprésentation de l'Œuvre de Gustave Moreau et donc du musée éponyme ainsi que de la nationalité française de l'essentiel des artistes. De facto, le panel étudié est essentiellement composé d'oeuvres datant de la deuxième moitié du XIXe siècle même si le pic de création pour l'année 1851 reste remarquable. Cette chronologie explique la prédominance des genres représentés ainsi que l'appétance pour la technique de la peinture à l'huile qui domine la création artistique de l'époque. Enfin, c'est encore le musée Gustave Moreau et son historique qui éclairent la proportion des legs ainsi que le pic des acquisitions sous la forme de legs, au moment où le peintre a cédé le contenu de sa maison dans l'optique de sa conservation en 1897, volonté confirmée à son décès l'année suivante. Ainsi, notre travail offre une représentation assez fidèle de ce que peut constituer une collection picturale de la fin du XIXe siècle et de ses caractéristiques techniques et juridiques.

Outre toutes les informations que cela peut nous apprendre sur les données récoltées, nous avons également beaucoup appris sur la création d'un mashup de données et sa

11. https://github.com/Juliettejns/baseJocondeMashup/blob/documentation_nettoye/datavisualisation/4_Gallerie_Paladio/Datavisualisation%20des%20peintures%20communes%20%C3%A0%20la%20Base%20Joconde%20et%20Wikidata.palladio.1.2.9.json

transformation en datavisualisation. En effet, les données peuvent être très longues à nettoyer afin de s'assurer un rendu et une étude intéressante. De plus, l'ajout de données en provenance d'autres bases n'est pas toujours le plus aisé et peut parfois réduire drastiquement la quantité de celles-ci. A l'inverse, si cette réduction de données peut surprendre, l'amélioration de la qualité des données a été largement appréciable. Concernant la datavisualisation, même si tout n'a pas pu être fait dans le même cadre, nous avons pu voir avec intérêt toutes les possibilités que nous donnaient des données propres.

A Tableau de bord 1

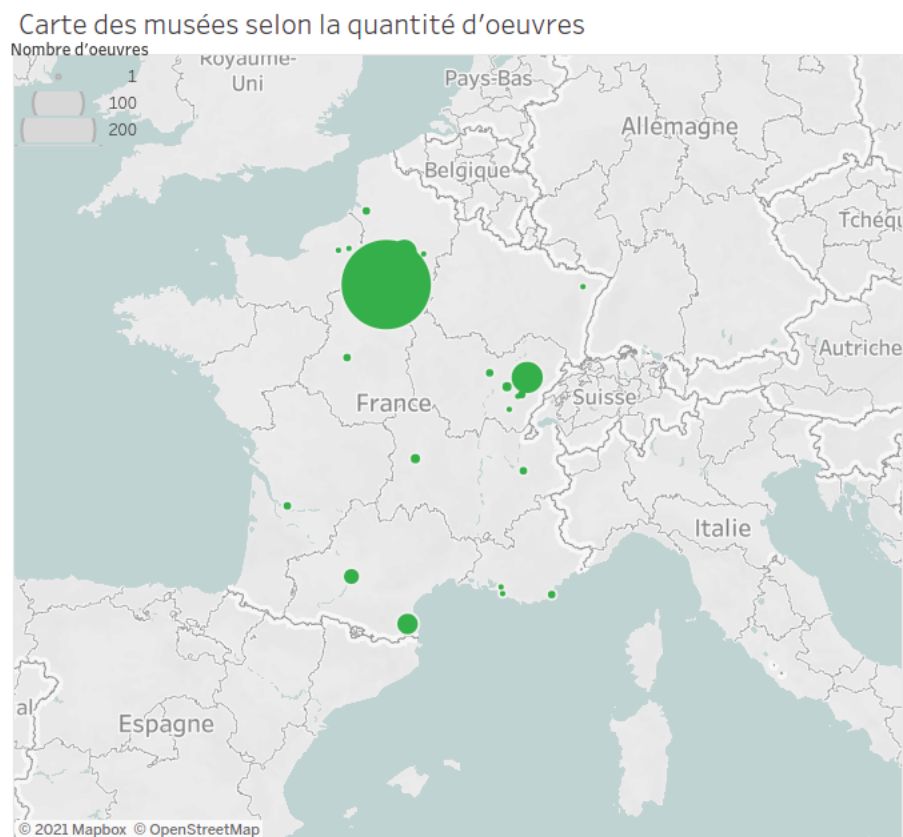


FIGURE 3 – Carte des musées

Nombre d'oeuvres par artiste

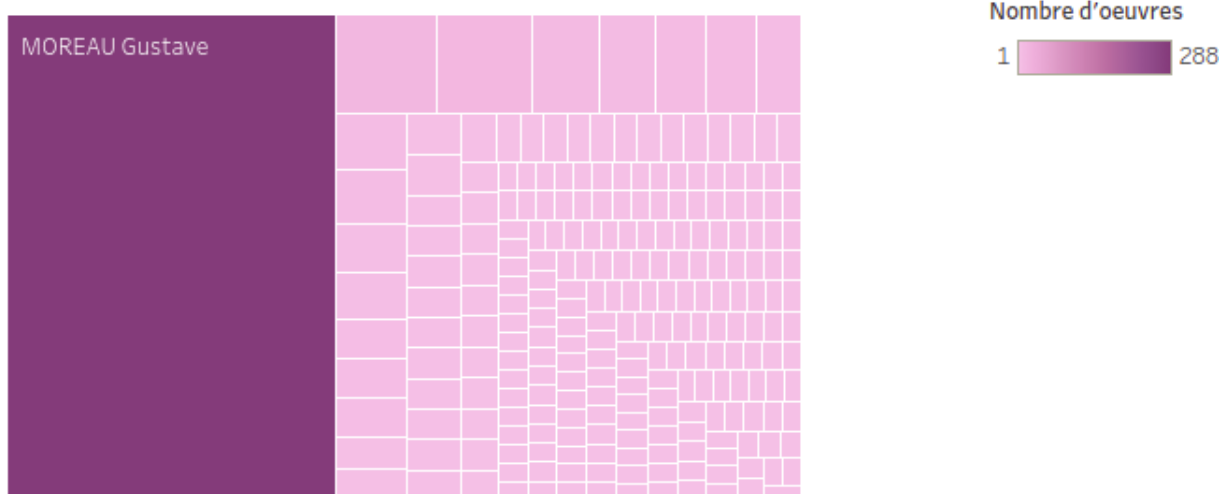


FIGURE 4 – Importance des Artistes dans notre dataset

Evolution du nombre de création d'oeuvres par an

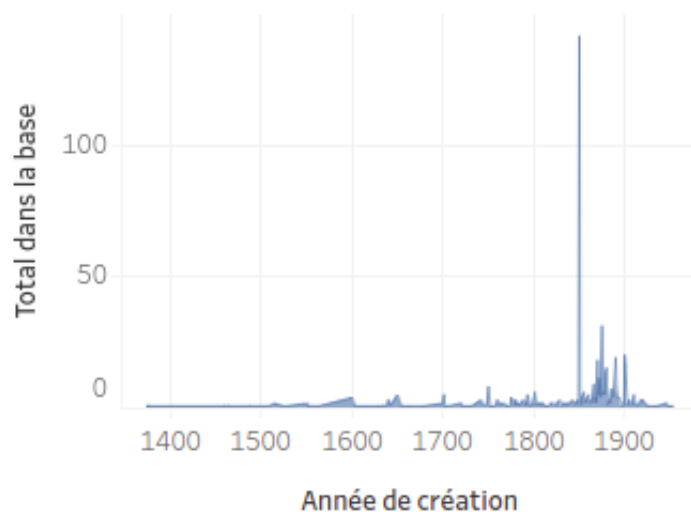


FIGURE 5 – Graphique du nombre d'oeuvres par rapport à leur date de création

Evolution du taux d'entrée dans la base

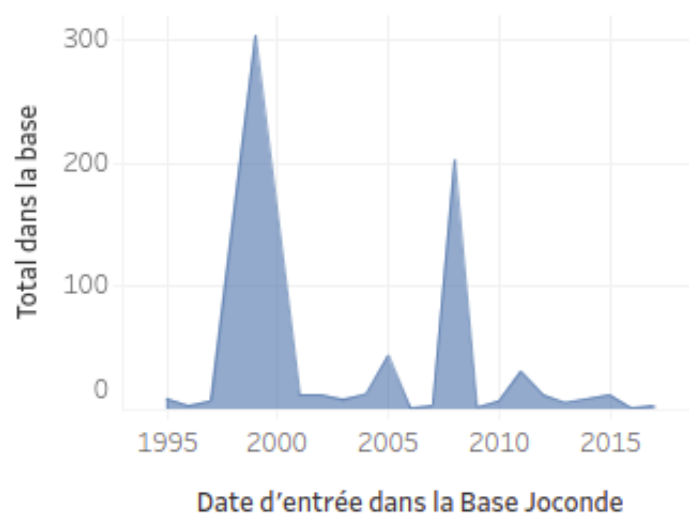


FIGURE 6 – Graphique du nombre d'oeuvres par rapport à leur date d'entrée dans la base

B Tableau de bord 2

Evolution du nombre d'acquisitions par les musées

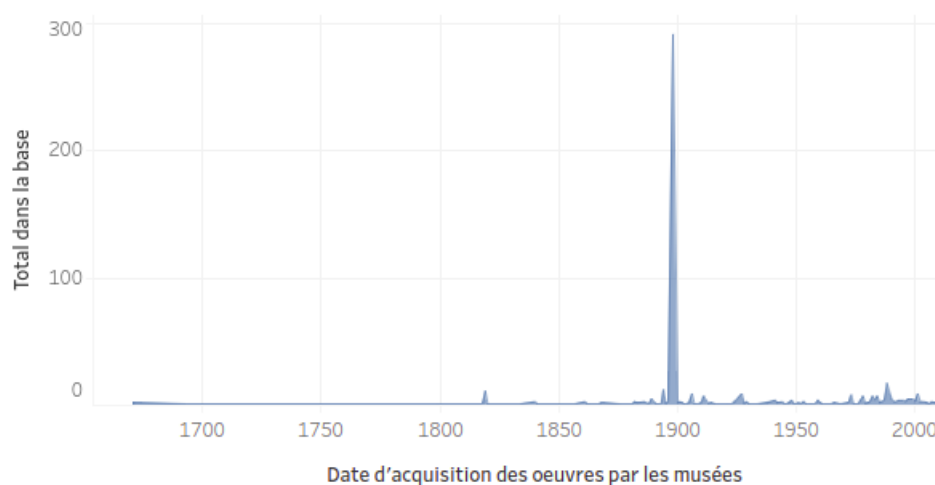


FIGURE 7 – Graphique du nombre d'oeuvres par rapport à leur date d'acquisition par les musées

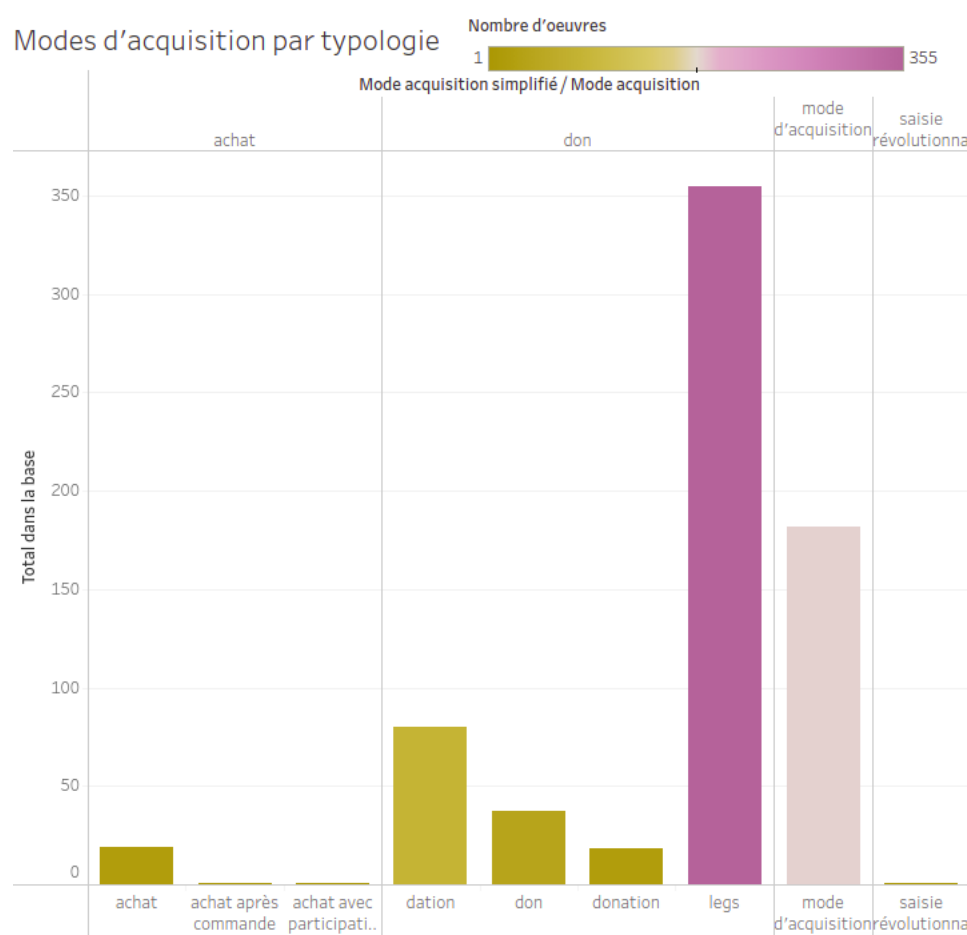


FIGURE 8 – Mode d'acquisition des oeuvres du dataset

Statut juridique des oeuvres

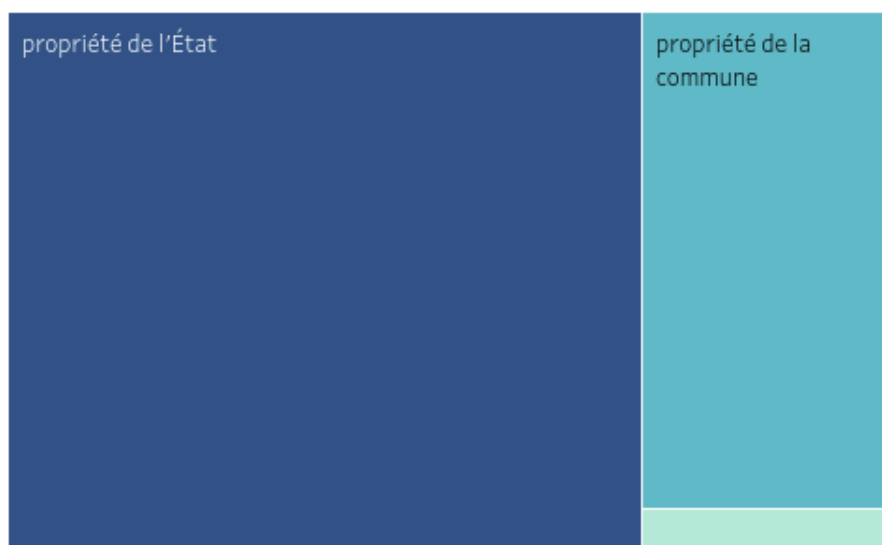


FIGURE 9 – Statut juridique des oeuvres du dataset

C Tableau de bord 3

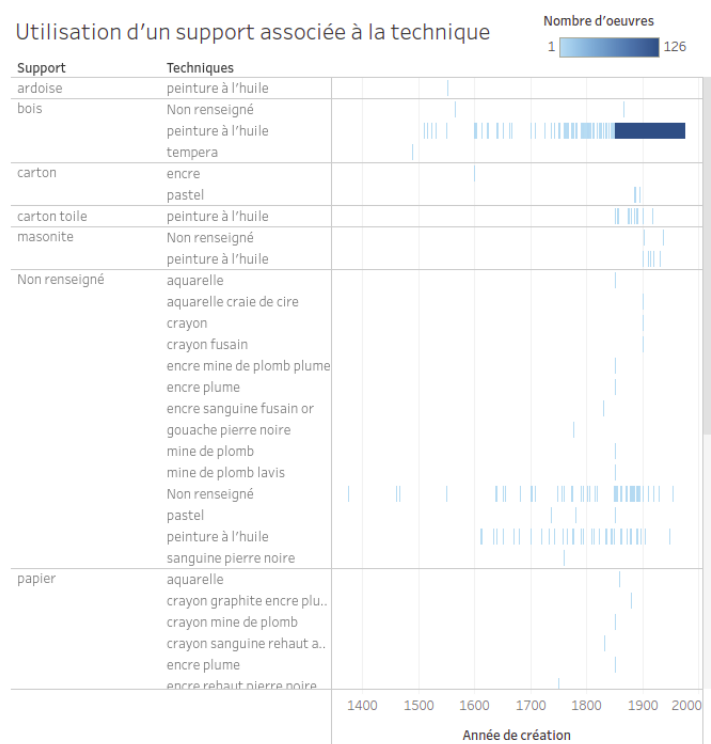


FIGURE 10 – Utilisation d'un support associée à une technique

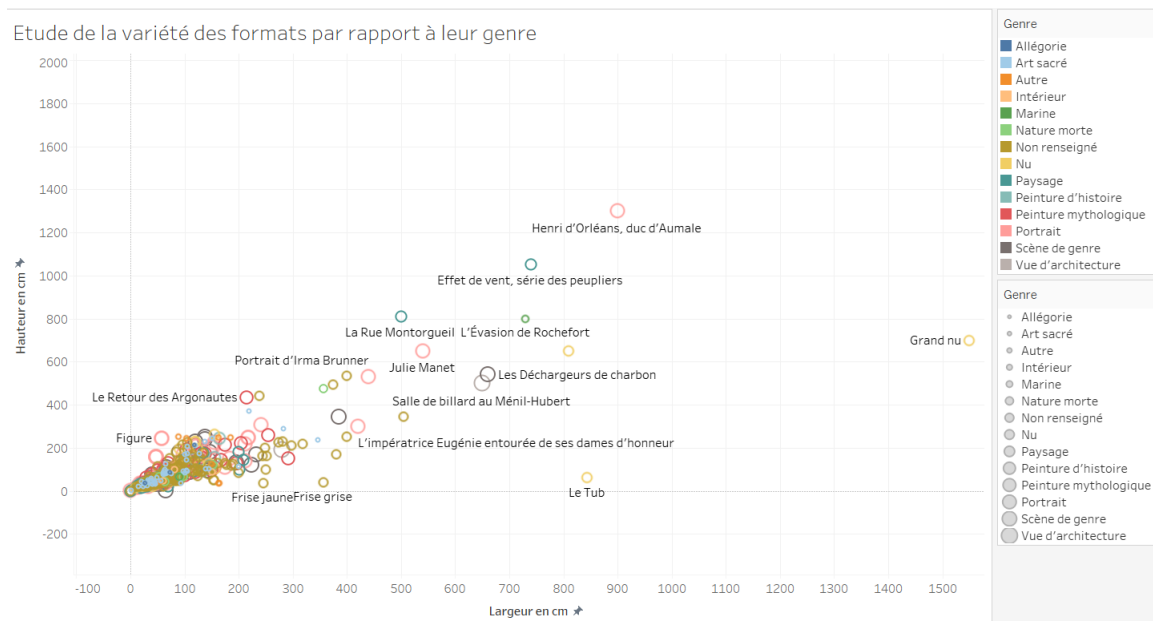


FIGURE 11 – Etude de la variété des genres par rapport à leur format

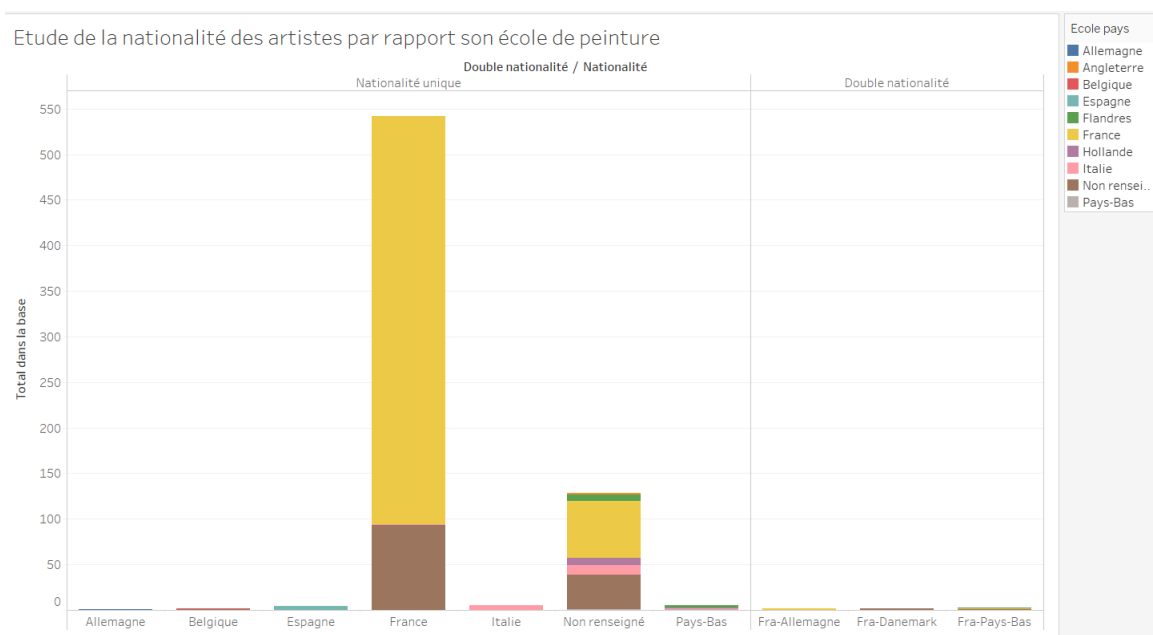


FIGURE 12 – Etude de la nationalité des artistes par rapport à son école de peinture

D Palladio

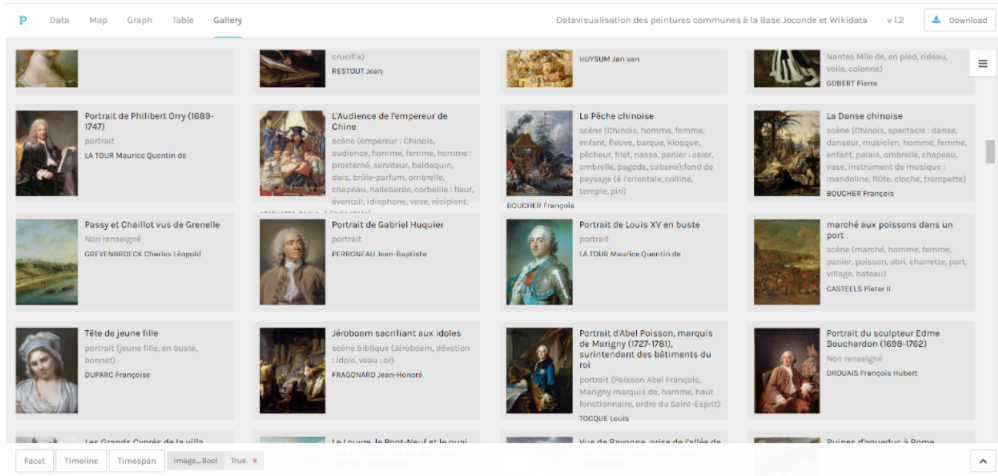


FIGURE 13 – Visualisation sur Palladio (1)

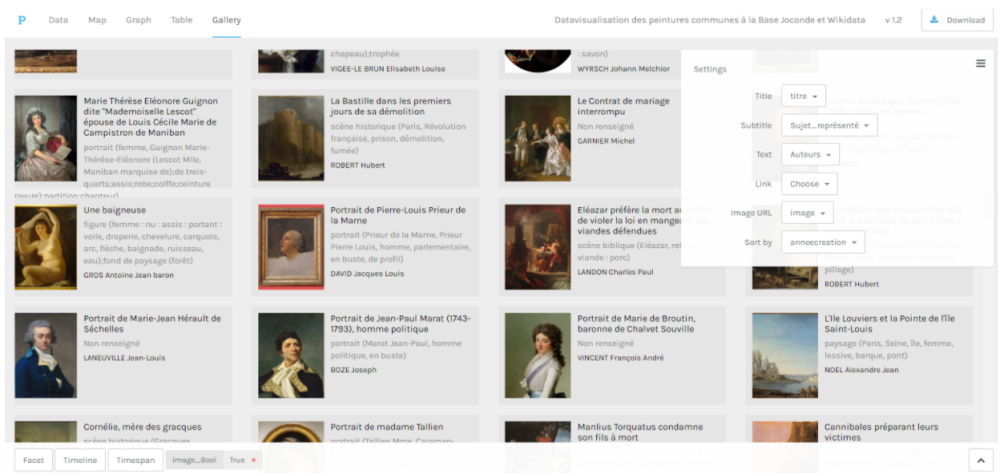


FIGURE 14 – Visualisation sur Palladio (2)

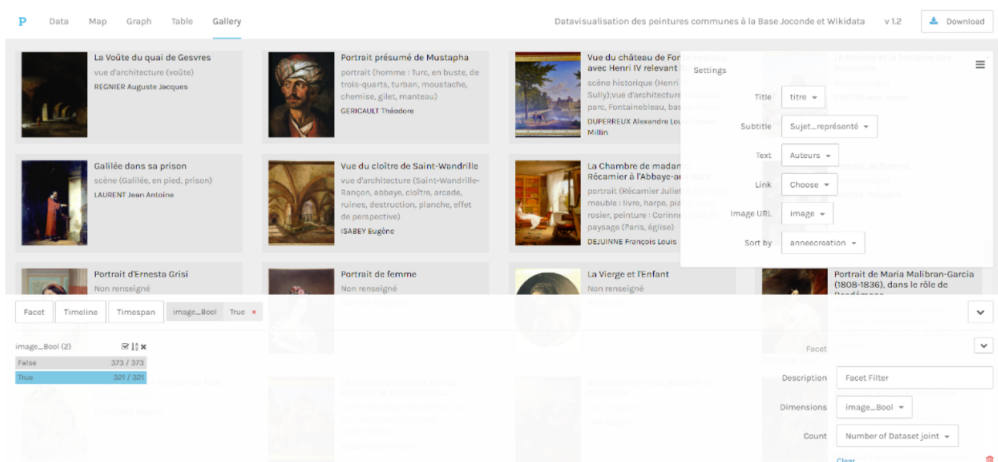


FIGURE 15 – Visualisation sur Palladio (3)