

Introduction à la visualisation de données

Kit de survie en milieu numérique pour l'étudiant en SHS - Mercredi 4 octobre 2017

Antoine Courtin

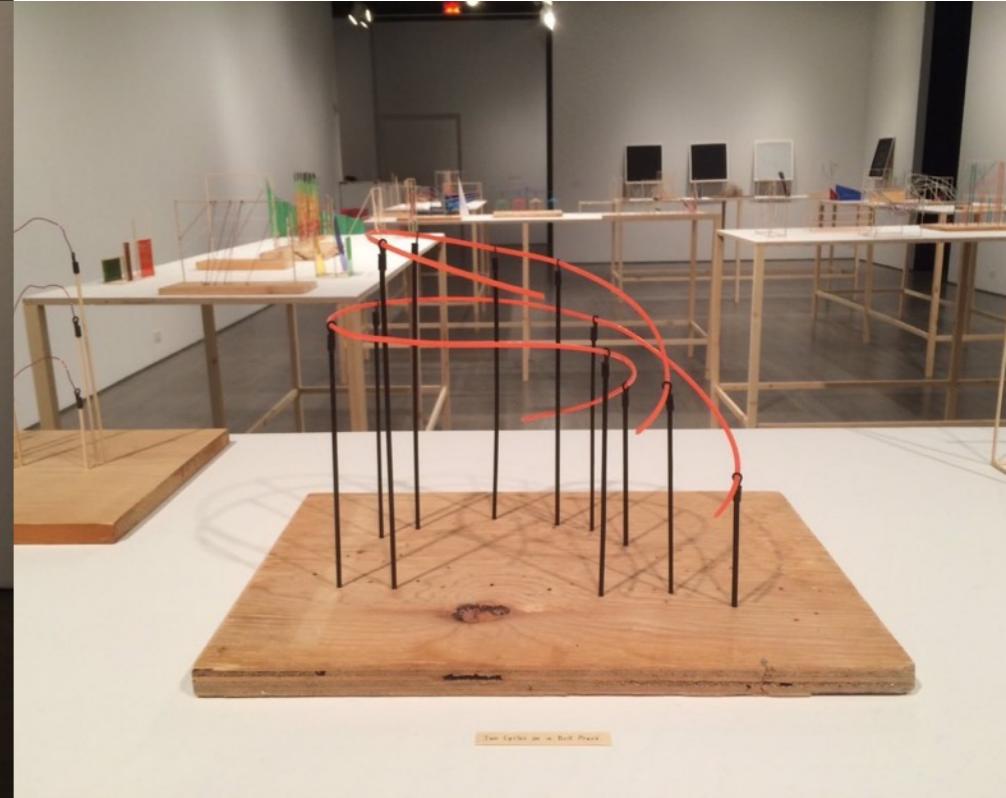
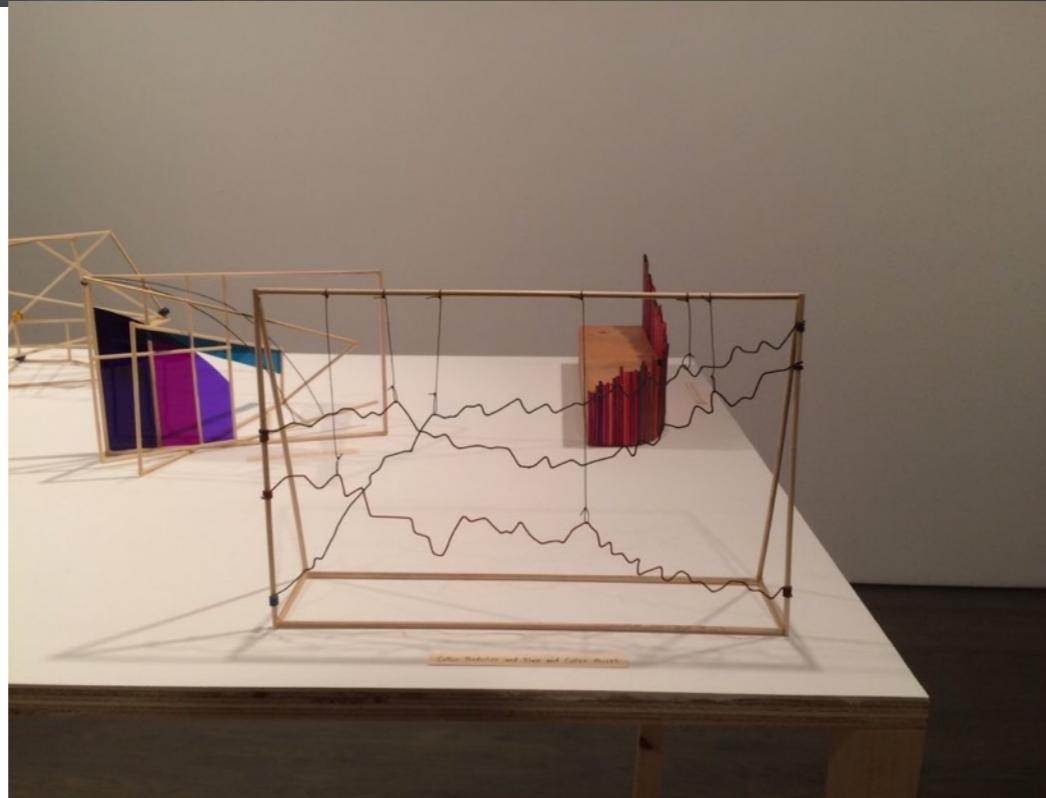
Responsable de la cellule d'ingénierie documentaire du département des études et de la recherche, INHA



Par Sam Fentress, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=215276>



Each Number Equals One Inhalations and One Exhalation, 2016



Introduction

Historique
Concepts-clés
Quelques exemples
Ressources

C'est l'ensemble des techniques de représentation graphique et d'exploration visuelle de données quantitative et/ou statistiques (mais pas uniquement) permettant de traduire un ensemble de données brutes en information décisive

Une/Des terminologies (avec des nuances)

- Dataviz/datavision
- Visualisation de données
- Design d'information
- Infographie
- etc,

Infovisualisation comme...

- ... outil d'aide à la décision / contrôle qualité
- ... outil d'accès et d'exploration de ressources
- ... outil d'aide à l'interprétation

Mettre en image/graphiques des informations/ données chiffrées

- Faire « parler » des données (brutes) ?
- Traduire un important corpus de manière visuelle
- Synthétiser les enjeux essentiels d'un corpus

Visualisation de données ou infographie ?

Une infographie... c'est plutôt :

- favoriser la compréhension d'une information que l'on connaît déjà en la représentant sous forme graphique
- un volume modeste de données
- une approche didactique tournée vers autrui
- une aide à la compréhension

Une dataviz... c'est plutôt :

- faire émerger une information que l'on ignorait en analysant des données présentées sous forme graphique
- un gros volume de données
- d'abord un outil de connaissance pour soi-même
- une aide à la décision

En fait de **calculs** et de proportions, le plus sûr moyen de frapper l'esprit, est de parler aux yeux

William Playfair, 1780

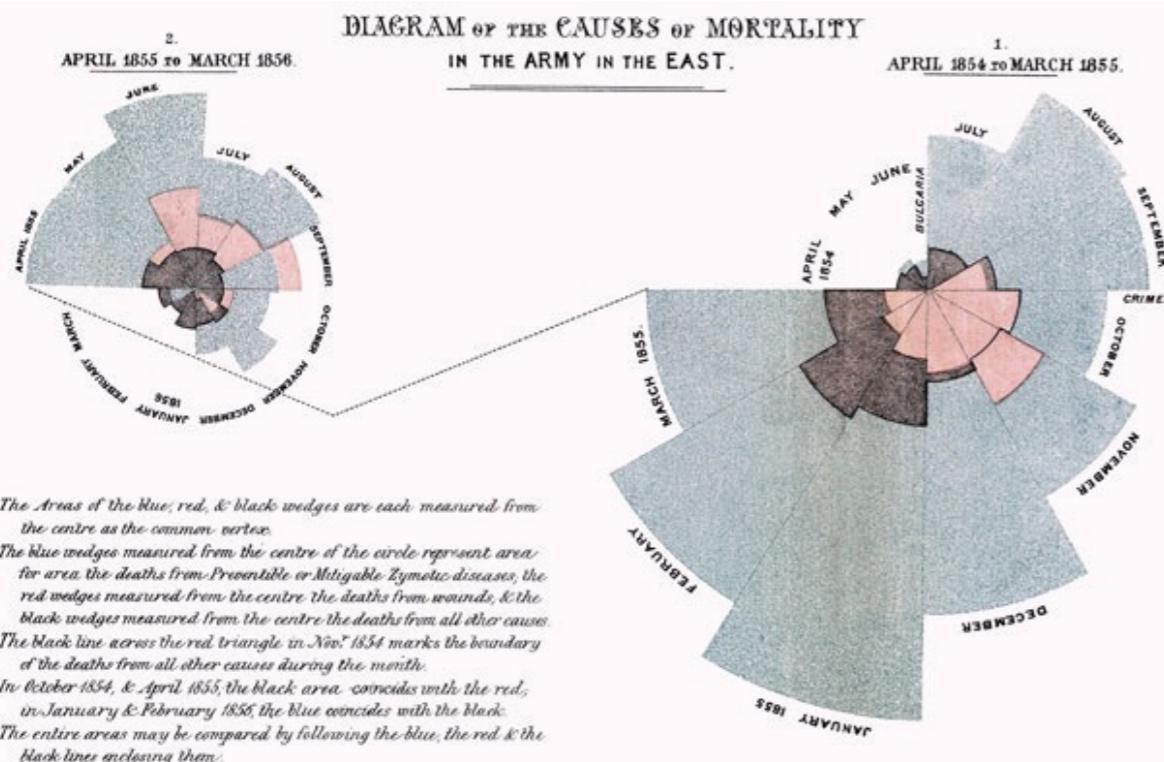
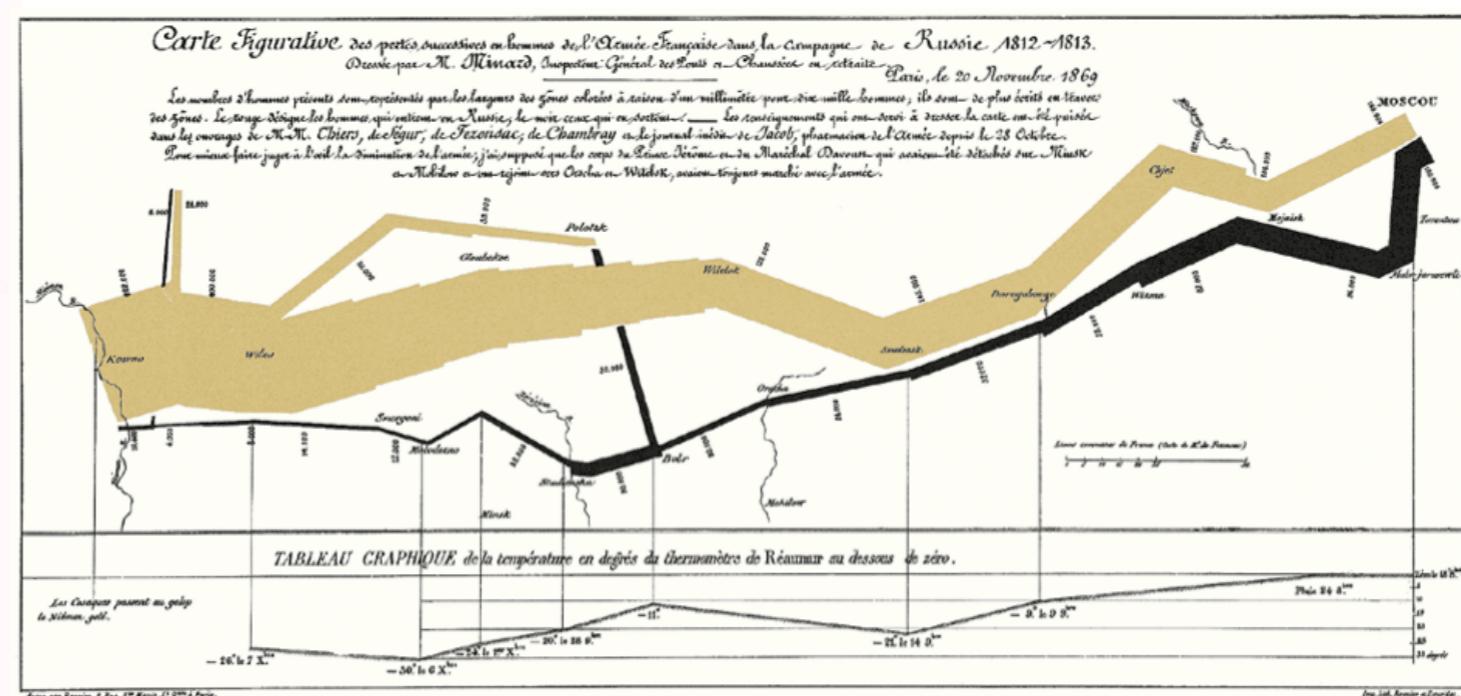


Diagramme des causes de mortalité au sein de l'armée en Orient par Florence Nightingale.



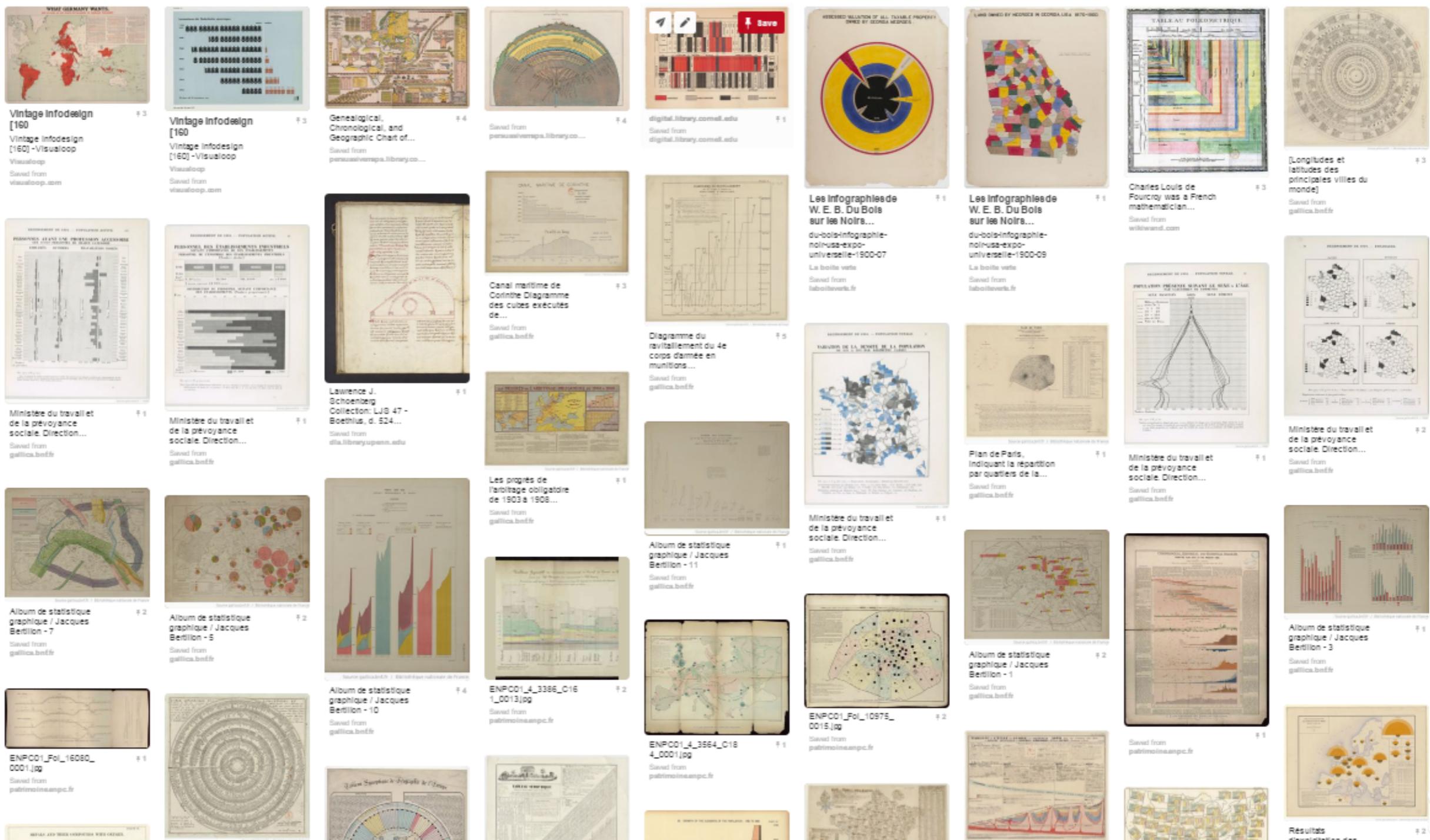
Carte figurative des pertes successives en hommes de l'armée
française dans la campagne de Russie 1812-1813 par Charles-Joseph
Minard.

InfoViz Vintage

88 Pins
108 Followers
Sélection personnelle et subjective d'Infoviz/dataviz datant des siècles derniers, issues des bibliothèques numériques telles que Gallica, DPLA, etc.



Save a Pin



<https://fr.pinterest.com/antoinecourtin/infoviz-vintage>

Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization

An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis

[Home](#) | [introduction](#) | [milestones project](#) | [varieties of data visualization](#) | [related](#) | [references](#) | [keyword index](#)

[Search](#)

[Pre-1600](#) [1600s](#) [1700s](#) [1800+](#) [1850+](#) [1900+](#) [1950+](#) [1975+](#)

Pre-1600: Early maps and diagrams

The earliest seeds of visualization arose in geometric diagrams, in tables of the positions of stars and other celestial bodies, and in the making of maps to aid in navigation and exploration. We list only a few of these here to provide some early context against which later developments can be viewed.

In the 16th century, techniques and instruments for precise observation and measurement of physical quantities were well-developed. As well, we see initial ideas for capturing images directly, and recording mathematical functions in tables. These early steps comprise the beginnings of the husbandry of visualization.



[Jump to Milestone...](#)

6200 BC [Unknown](#) **Oldest known map?**

Added: 2005-10-01

The oldest known map? (There are several claimants for this honor.)

[Town map, with an erupting volcano \(Hasan Daö?\) and the Konya plain](#)

[An extended description of the most ancient maps](#)

References:
[none](#)



Konya town map

Mickaël Friendly, Milestones in the history of thematic cartography, statistical graphics, and data visualization, <http://www.datavis.ca/>

A quoi sert la visualisation de données

Ce qu'en dit Jacques Bertin.

1. A traiter des données pour comprendre et en tirer de l'information

- Idée de la fouille de données
- Traiter des données implique l'exhaustivité.

2. A communiquer cette information

- Il faut quelques heures pour construire un graphique ; quelques secondes pour le lire.
- Communiquer implique (souvent) la simplification.

source Christophe Bontemps

- Les visualisations profitent des caractéristiques du système de traitement cognitif humain. D'après Colin Ware, les visualisations offrent de nombreux avantages dont les principaux sont les suivants :
 - Les visualisations graphiques ont le pouvoir de permettre à l'homme de manipuler des structures bien plus complexes représentées par une visualisation (représentations externes) que dans la mémoire de travail visuelle et verbale (représentations internes).
 - Elles permettent aussi de percevoir l'émergence de propriétés dans les données cartographiées qui ne sont pas anticipées.
 - Les visualisations permettent de mettre en évidence des problèmes dans les données, dans leur collecte. Avec une visualisation appropriée, les erreurs dans les données sont rapidement perceptibles.
 - La visualisation permet de percevoir simultanément des propriétés à grande et à petite échelle sur les données.
 - La visualisation facilite la formation d'hypothèses sur les données.

John W. Tukey

EXPLORATORY DATA ANALYSIS



“Pour faire des découvertes surprenantes dans les données, le meilleur outil dont nous disposons est l’oeil, et sa faculté de lecture des images.

“ L’approche exploratoire des données est un type d’analyse des données pour en résumer les principales caractéristiques d’une façon claire et compréhensible, souvent en utilisant des graphiques, sans utiliser de modèles statistiques, et en n’ayant pas non plus formulé d’hypothèse.”

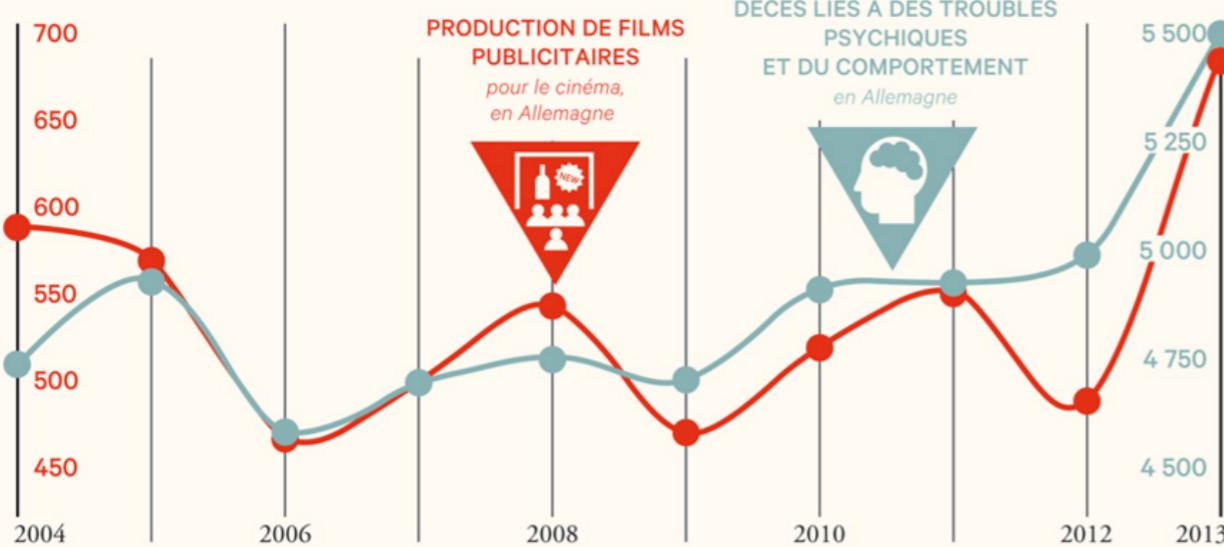
Méthode qualitative / Méthode quantitative



WARNING

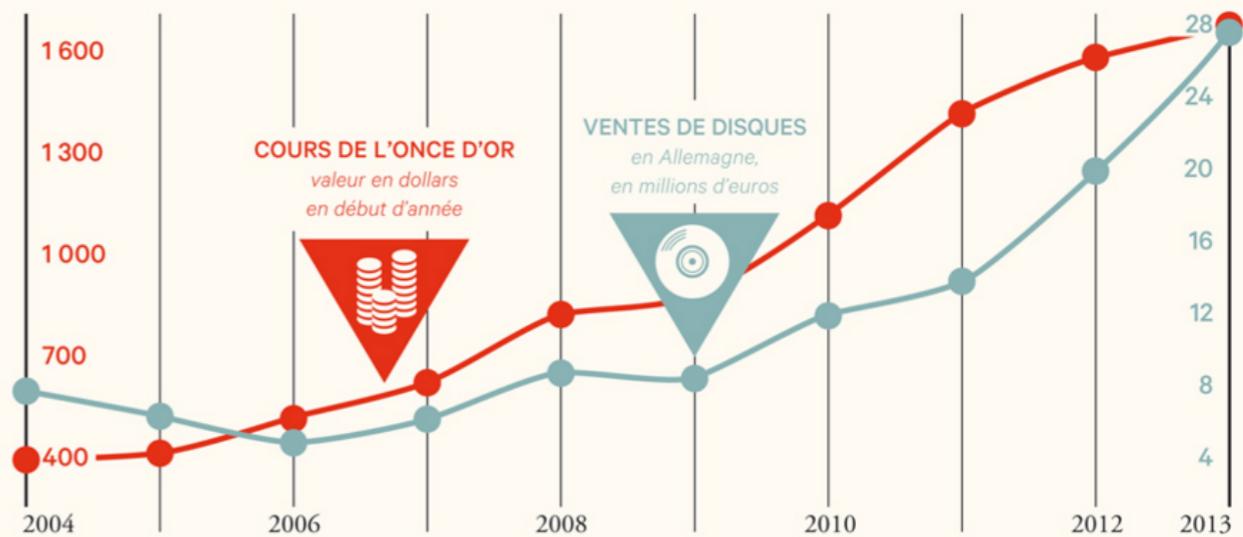
PUBLICITÉ MENSONGÈRE

Les publicités au cinéma ont-elles une incidence sur le nombre de décès dans la population ?
Coefficient de corrélation : 0,792



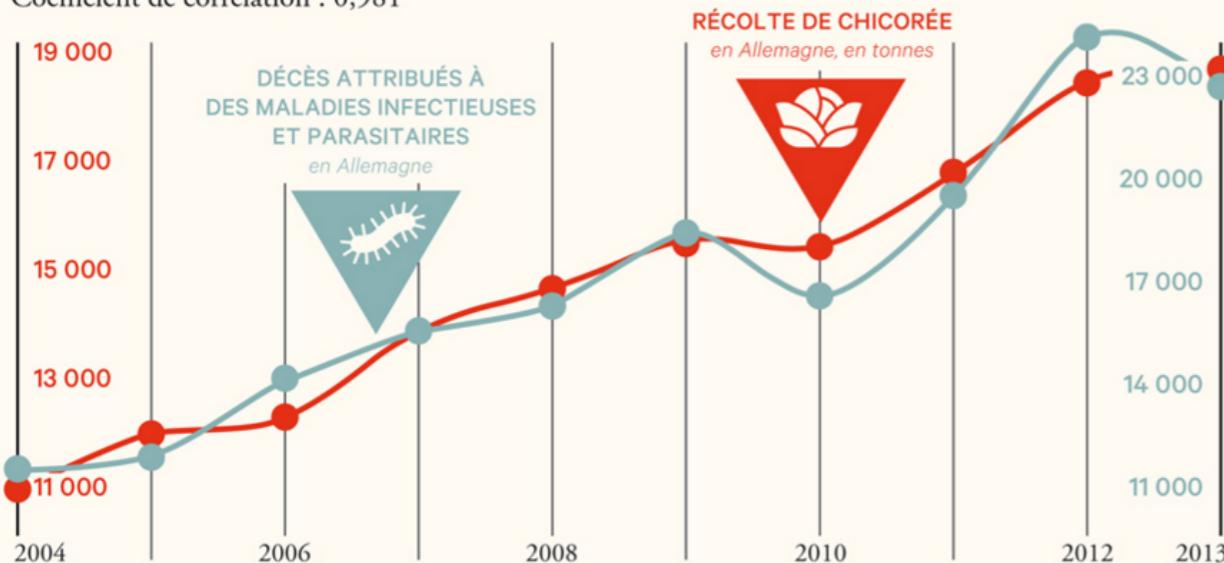
DISQUE D'OR

Les ventes de disques vinyle font-elles monter le cours de l'or ?
Coefficient de corrélation : 0,905



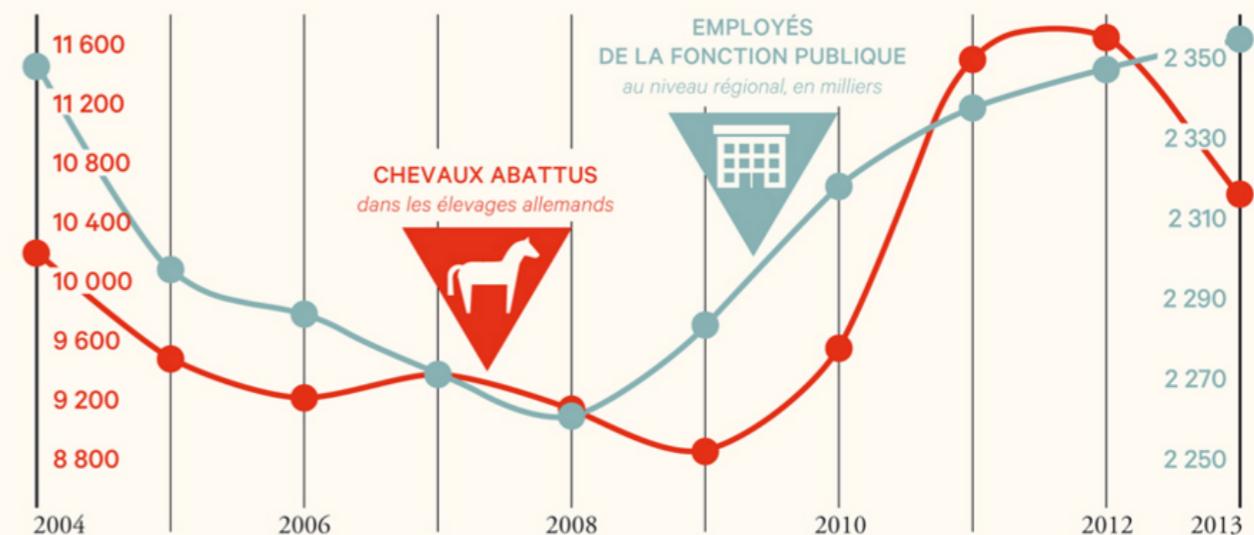
POISON VERT

La chicorée consommée en salade est-elle responsable du nombre de maladies infectieuses ?
Coefficient de corrélation : 0,981

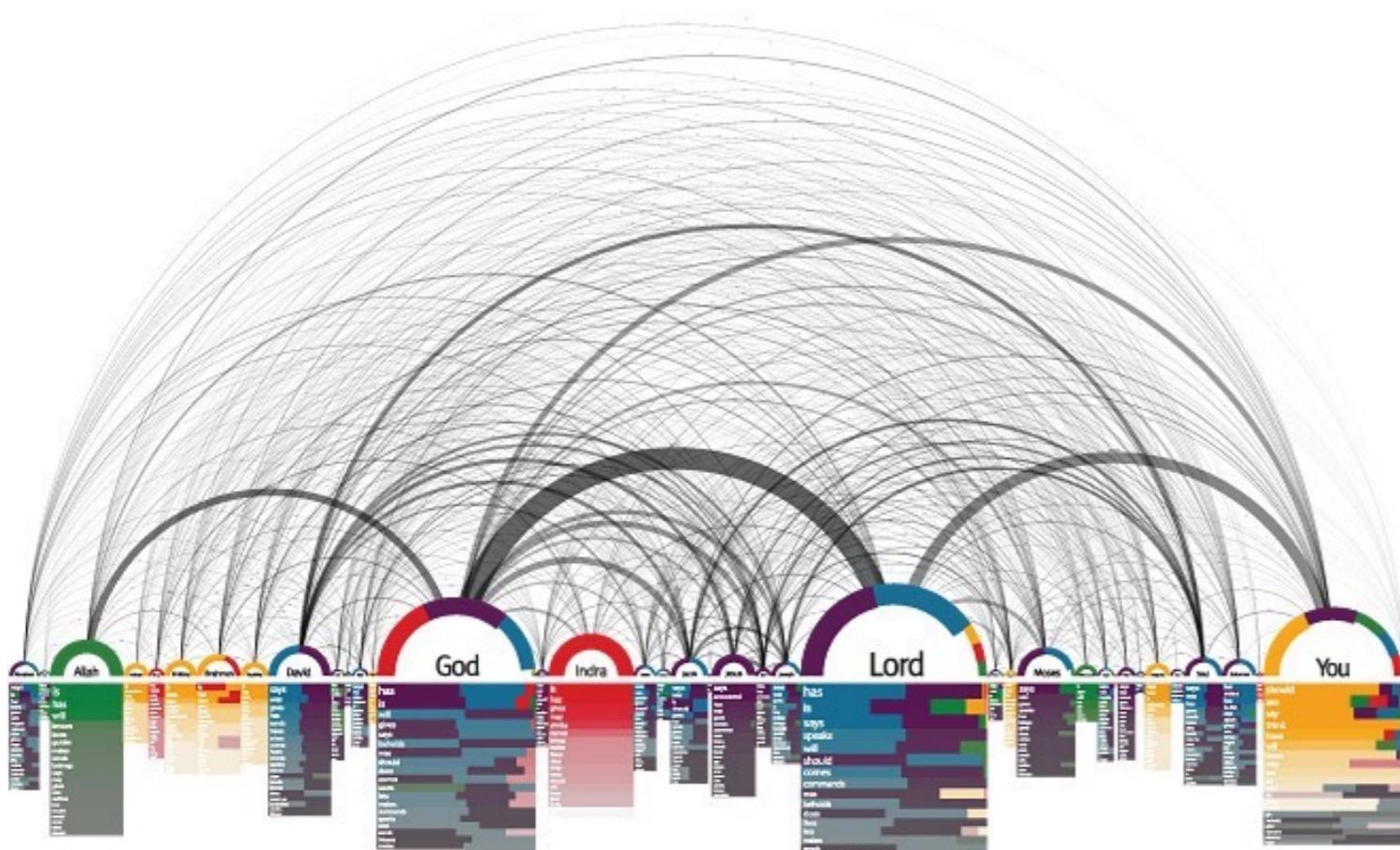
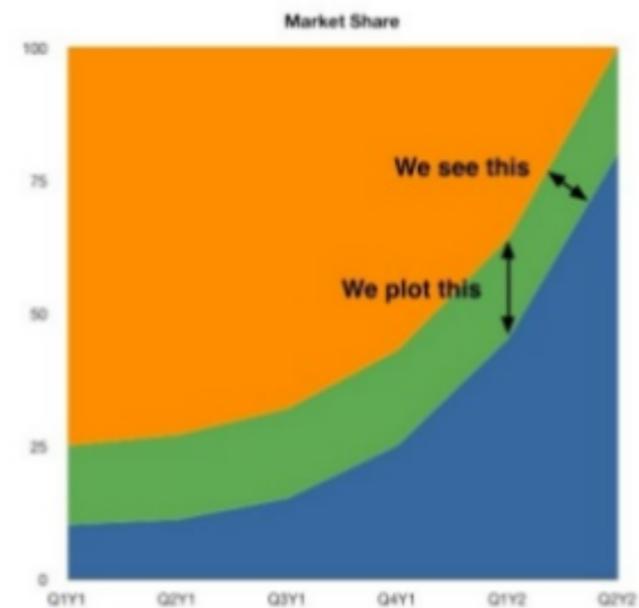
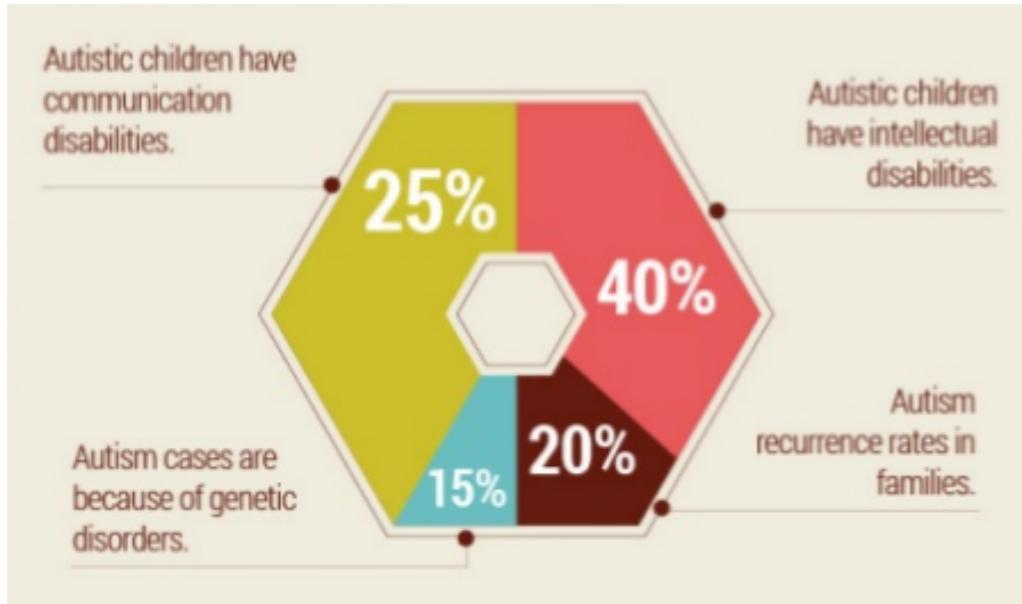


ON ACHÈVE BIEN LES CHEVAUX

Les abattoirs de chevaux augmentent-ils leur activité en fonction du nombre de fonctionnaires ?
Coefficient de corrélation : 0,821



Statistiques. Les corrélations de l'absurde, <http://goo.gl/9mM7zl>



Ce de mots
Nuage mais
est attrayant est-il
très vraiment utile?

Martin Grandjean. Introduction à la visualisation de données : l'analyse de réseau en histoire. Geschichte und Informatik, Chronos, 2015, pp.109-128.



The Rules:

1. Show Your Data
2. Use Graphics
3. Avoid Chartjunk
4. Utilize Data-ink
5. Use Labels
6. Utilize Micro/Macro
7. Separate Layers
8. Use Multiples
9. Utilize Color
10. Understand Narrative

<http://www.sealthreinhold.com/school/tuftes-rules/>

Dans la vision académique d'Edward Tufte, l'excellence graphique peut être estimée au travers de 3 critères fondateurs :

- **data-ink ratio**
 - Le ratio entre informations restituées/encre utilisée () .
- **lie-factor**
 - distorsion de la restitution par rapport à la réalité des données sources).
- **chartjunk**
 - L'absence de scories graphiques (élément décoratif n'apportant aucune plus-value informative)

Visualisation de données temporelles

Visualisation de la relation entre plusieurs variables

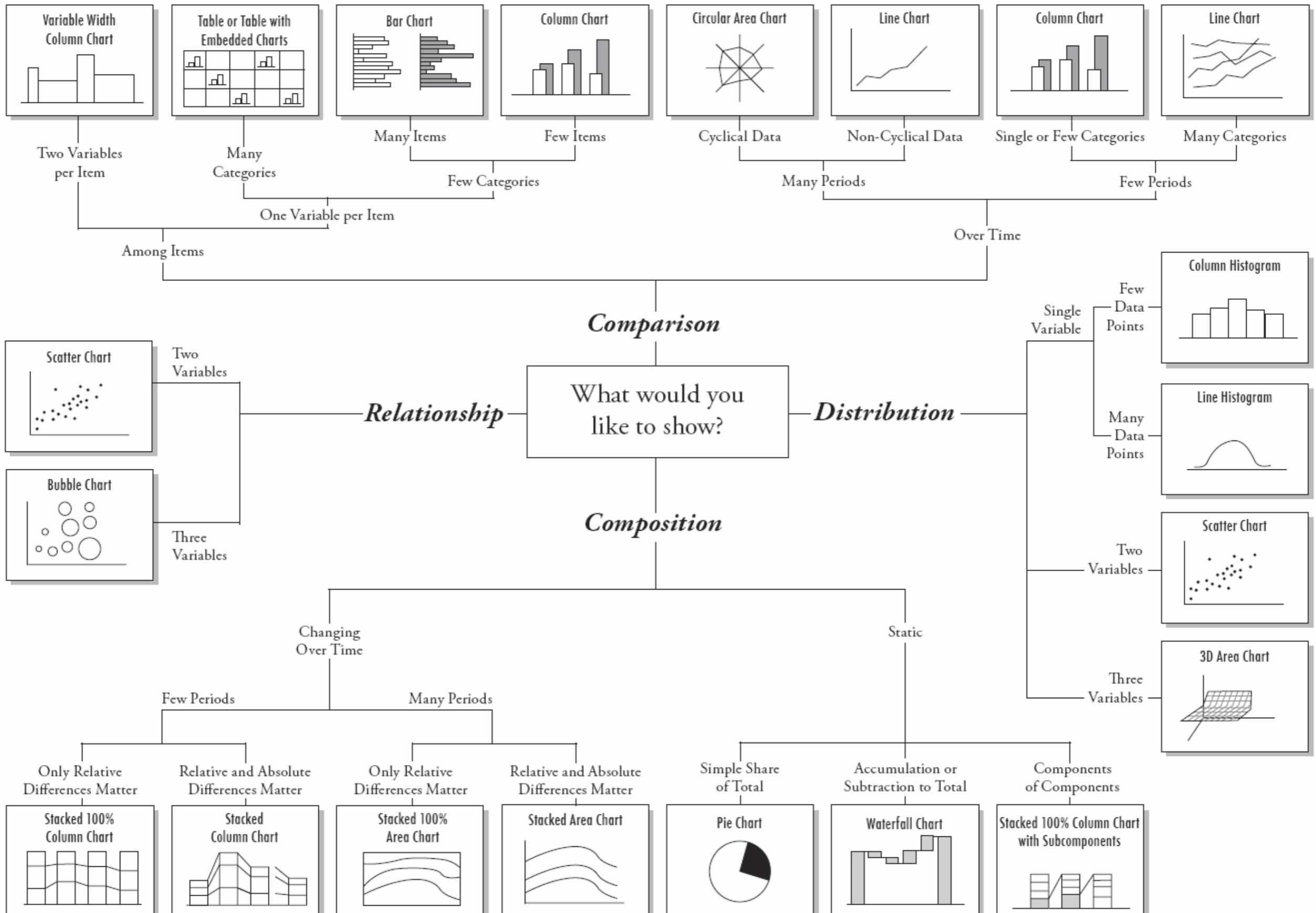
Visualisation d'une arborescence

Visualisation de flux

Visualisation de données géolocalisées

Visualisation des proportions

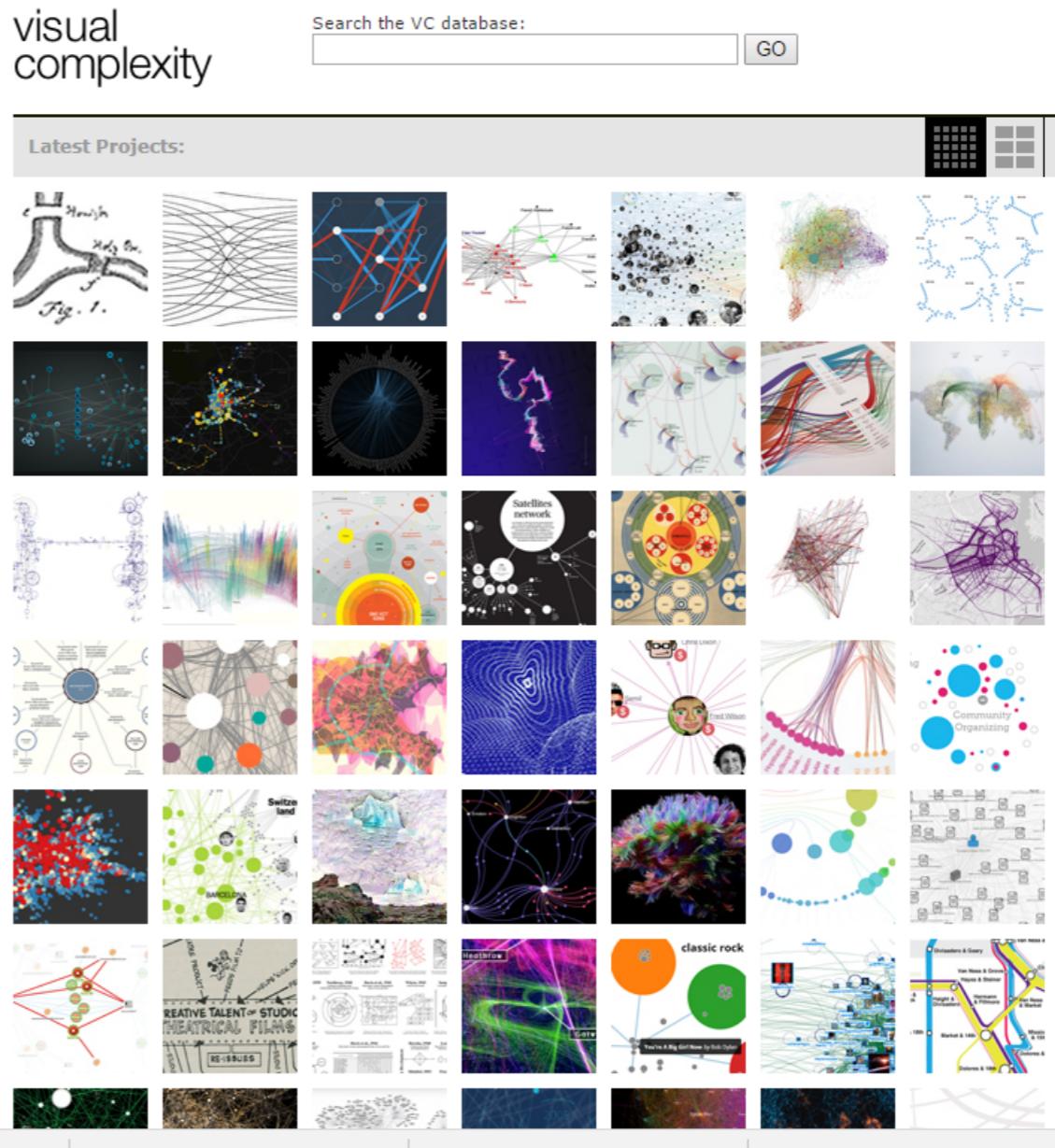
Visualisation de réseaux



Conception et source d'inspiration



<http://www.datavizcatalogue.com>



<http://www.visualcomplexity.com/vc/>

SCHLAGWORTKATALOG
BAY - BEE

SCHLAGWORTKATALOG
BEF - BERF

BERG - BERL

34 SCHLAGWORTKATALOG
BIBLIOGRAPH:N - BIBLIOTHEKEN:DD

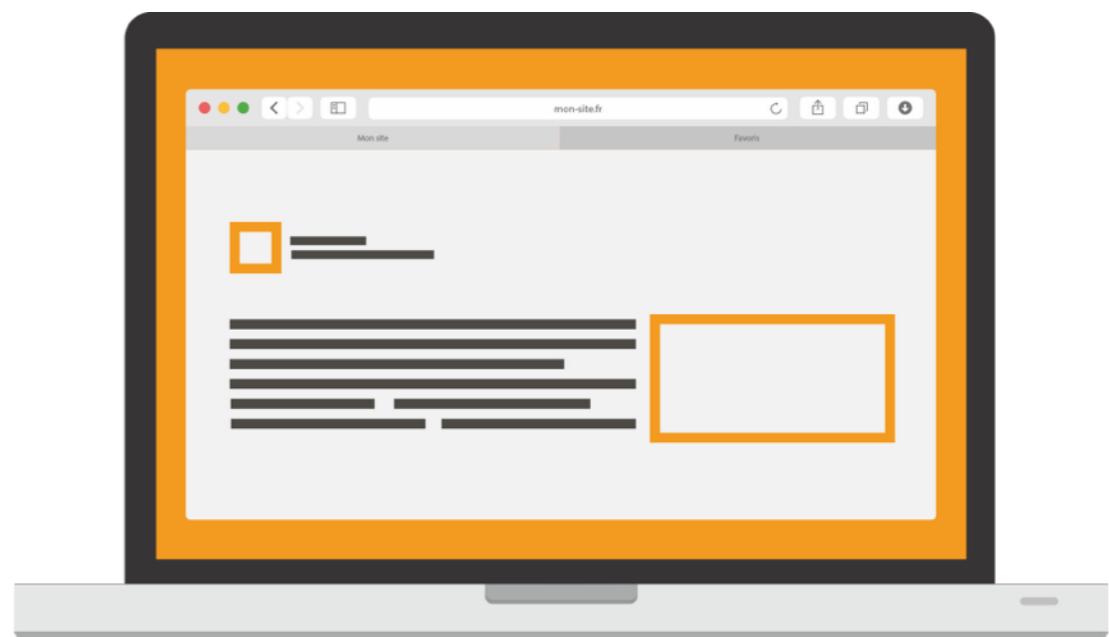
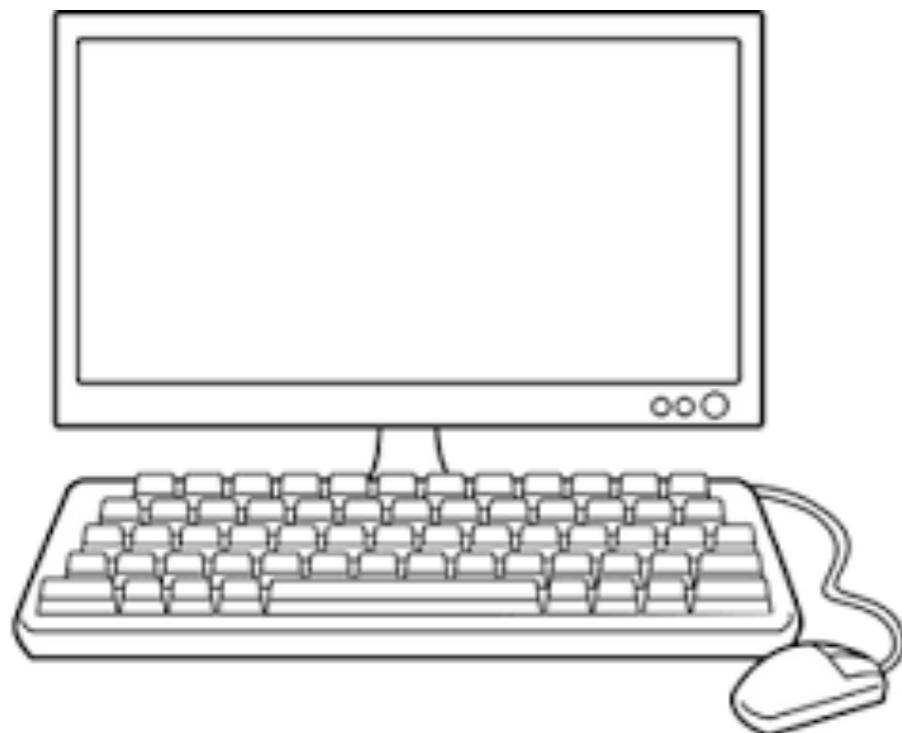
SCHLAGWORTKATALOG

35 SCHLAGWORTKATALOG
BIBLIOTHEKEN:DE - BIBLIOTHEK:N

Bibliotheken:
Kataloge: Schlagwortkatalog
(Schleimer, Hans): Der bibliothekarische Schlag-
wortkatalog. Mit Regeln f. die U.B. in Graz u. eines Anh.:
Systematischer oder alphabetischer Sachkatalog?
(Leipzig: Harrassowitz, 1901).
Aus: Zentralblatt

I
206-919

Où sont les données ?



the terrors of

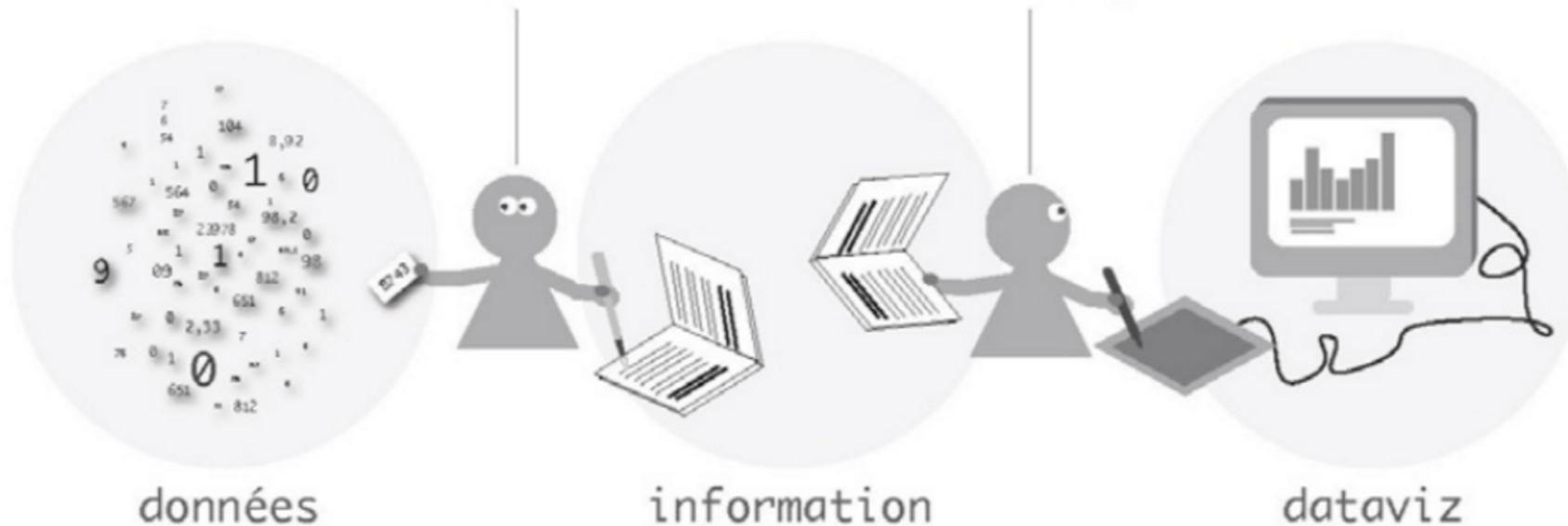
DIRTY DATA



CLEAN ALL THE
DATA!



<https://media.linkedin.com/mpr/mpr/AEAAQAAAAAAAALAAAAJGU3YjAwY2I5LTg1ODMtNDg5Ny04MTg2LTNhMWUwYTdkZDk2Zg.pn>



La datavisualisation vue par © Baptiste Fuchs

Acquisition des données

- saisie des données
- collecte de données
- enrichissement des données tierces

Structurer les données

- nettoyer les données
- optimiser les formats
- compléter les données

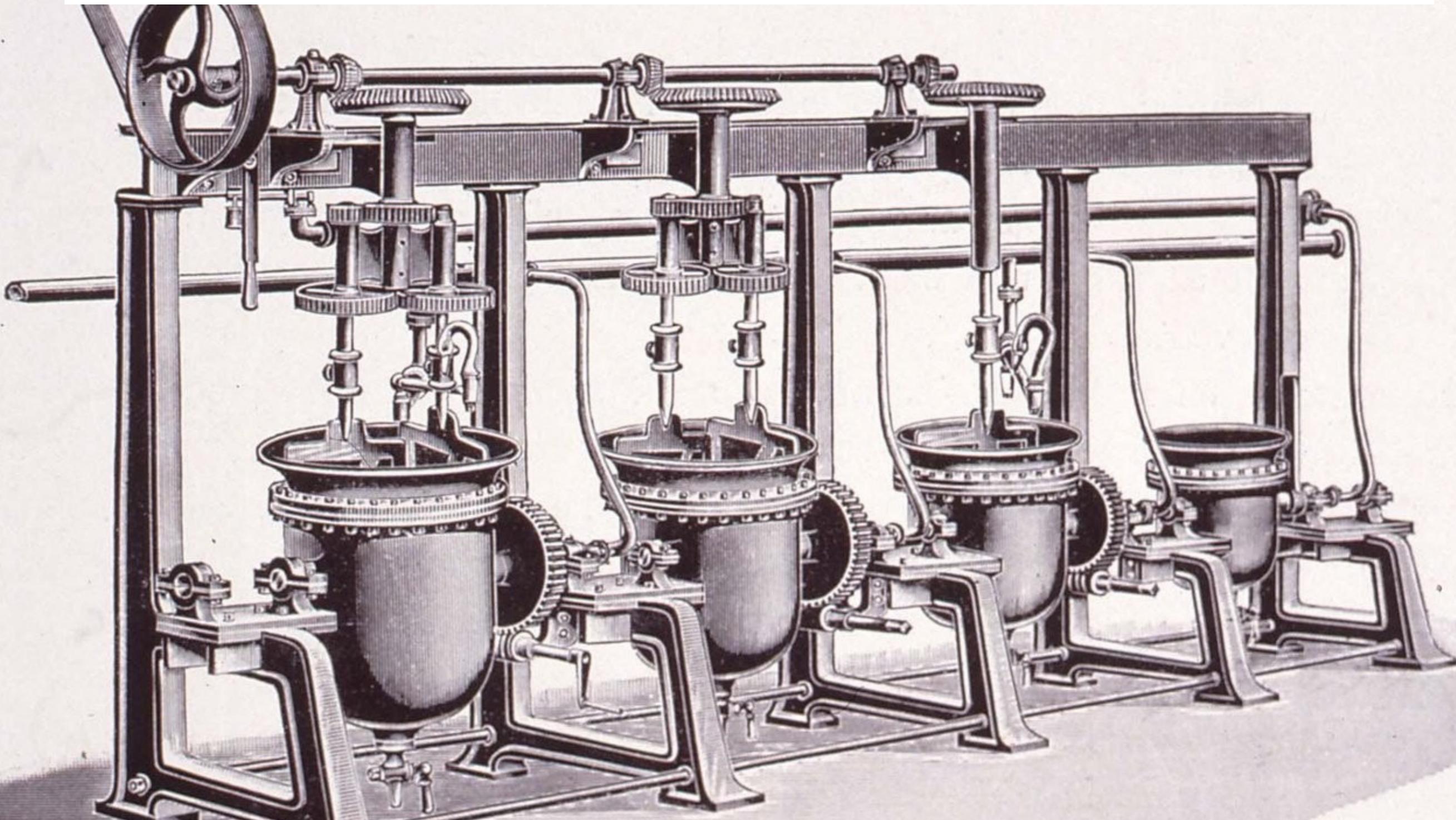
Visualiser l'information

- choisir des informations à révéler
- tester les formes de visualisation

Editorialiser la visualisation de données

- apporter un discours
- légender

logiciels spécialisés de « self service data preparation »



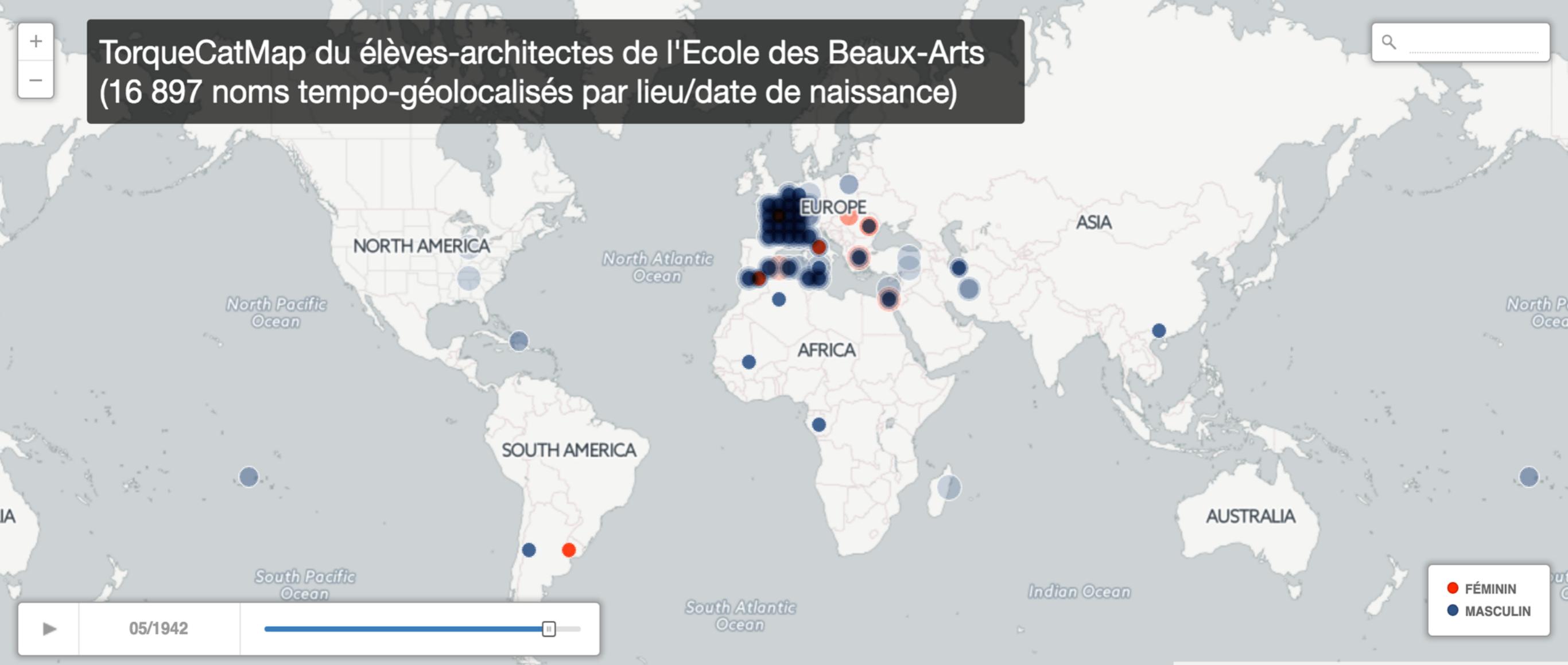
- Trifacta wrangler
- Open Refine (ex Google refine)
- Talend data preparation
- Dataiku DSS

◆ Le dictionnaire comme un cas d'école



66887

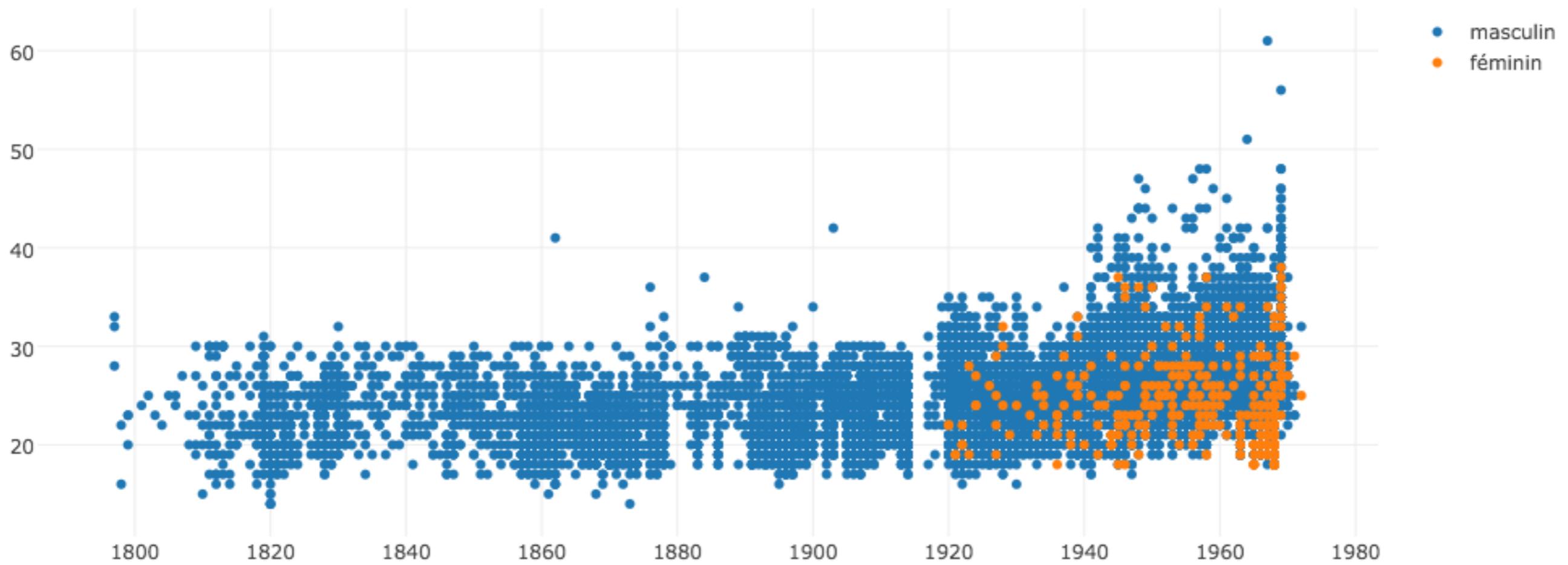
TorqueCatMap du élèves-architectes de l'Ecole des Beaux-Arts (16 897 noms tempo-géolocalisés par lieu/date de naissance)



© OpenStreetMap contributors © CartoDB, CartoDB attribution

https://antoinecourtin.cartodb.com/viz/fcd542aa-df01-11e5-9105-0e98b61680bf/public_map

- Croiser les dimensions
 - temporelle (date de naissance)
 - géospatiale (par ville de naissance des élèves)
 - catégorielle (par sexe (homme/femme))



- Croiser les dimensions
 - temporelle (date d'entrée à l'école)
 - géospatiale (par ville de naissance des élèves)
 - catégorielle (par âge d'entrée à l'école des beaux-arts)

RETIF en Hauts-de-France v.0.9.3

- [A propos](#)
- [Partager](#)
- [Télécharger](#)

Rechercher un tableau

Liste des œuvres

	[Abate, Nicolò dell'] Le Massacre des Triumvirs (Beauvais)	»
	[Albani, Francesco] Sainte Marie-Madeleine en prière (Chantilly)	»
	[Albotto, Francesco] Vue de la place Saint-Marc à Venise (Compiègne)	»
	[Albotto, Francesco] Vue de Venise (Compiègne)	»
	[Allori, Alessandro] Buste de jeune italienne (Fontaine-Chaalis)	»
	[Allori, Alessandro] La Vierge et l'Enfant avec sainte Elisabeth (Chantilly)	»
	[Allori, Alessandro] Tête d'homme (Laon)	»
	[Allori, Alessandro] Un Ange montre à saint François d'Assise le Christ détaché de la Croix (Chantilly)	»
	[Allori, Cristofano] Judith tenant la tête d'Holopherne (Chamant)	»
	[Allori, Cristofano] Sainte Catherine de Sienne en prière (Amiens)	»
	[Allori, Cristofano] Sainte Catherine de Sienne en prière (Compiègne)	»

[Anguissola, Sofonisba] Portrait de femme (Chantilly)

Auteur Anguissola, Sofonisba (1532-1625)

Titre Portrait de femme

Matière et technique bois. peinture à l'huile

Date de création 16e siècle

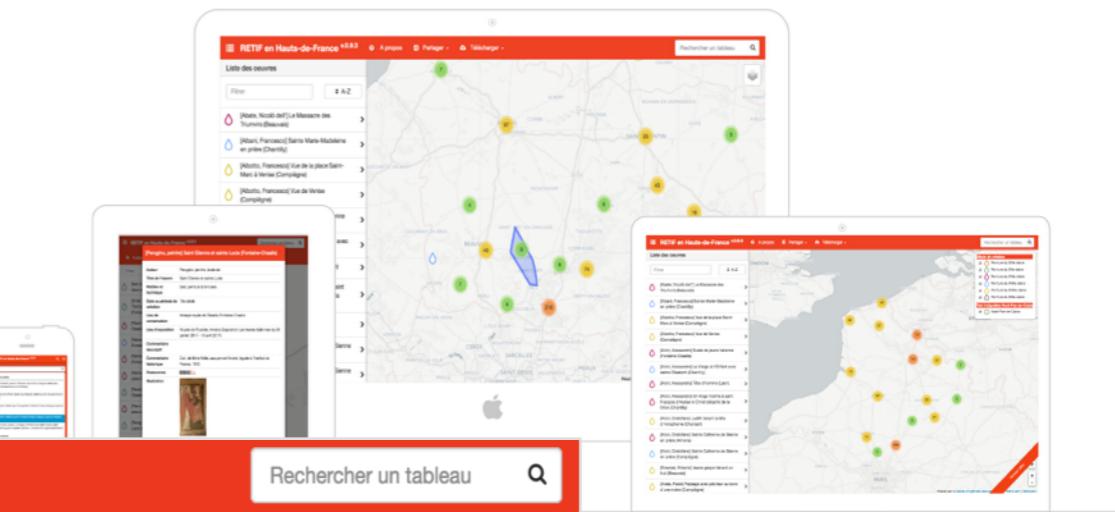
Lieu de conservation Musée Condé, Chantilly

Commentaire historique Provenance inconnue.

Ressources [AGORHA Joconde](#)

Illustration

Suivant Fermer



Elèves-architectes (1800-1900)

Réseau des élèves-architectes de l'école des beaux-arts de Paris nés entre 1800 et 1900. Les ateliers ayant moins de 2 élèves ont été écartés.

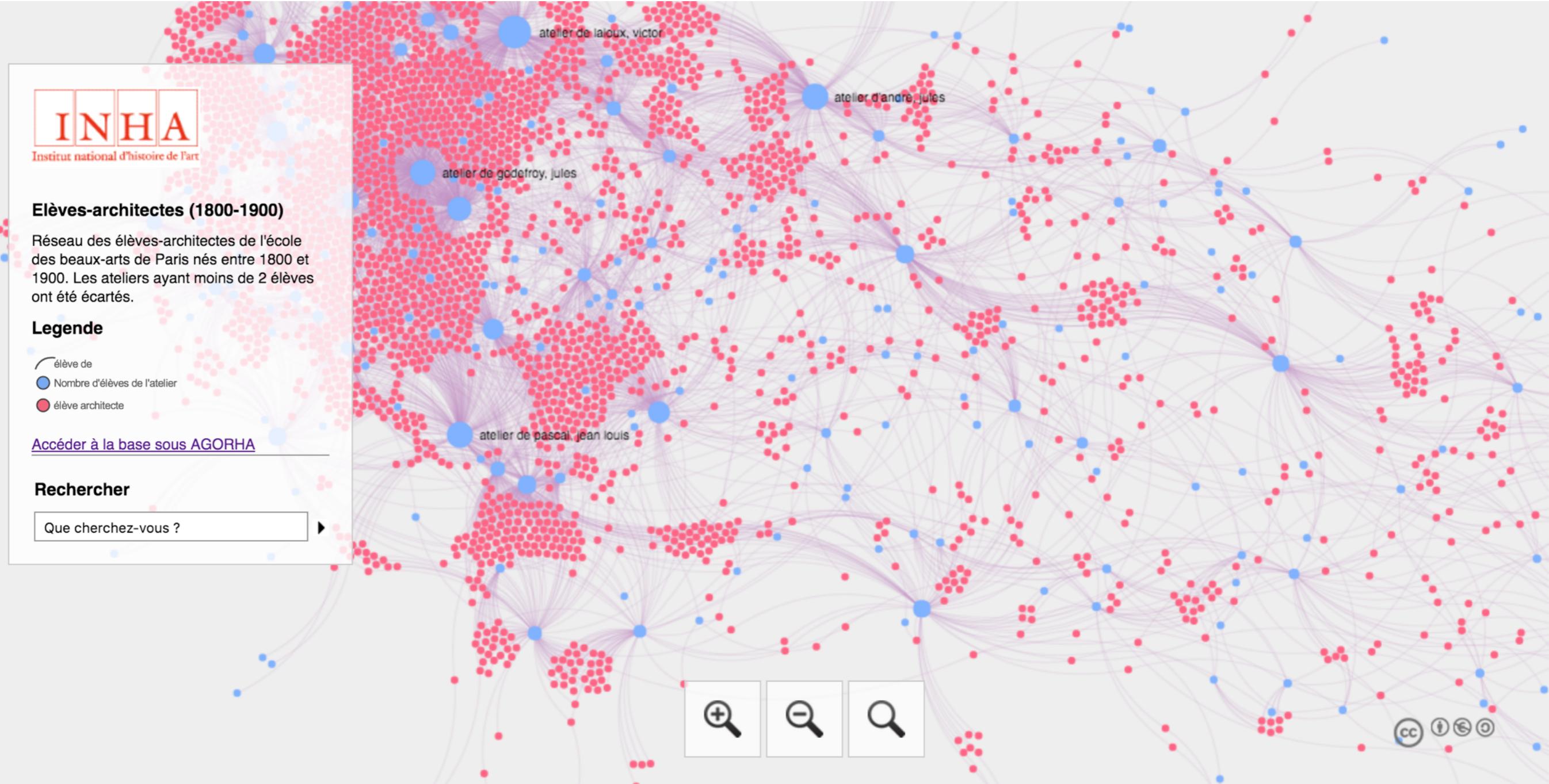
Legende

- élève de
- Nombre d'élèves de l'atelier
- élève architecte

[Accéder à la base sous AGORHA](#)

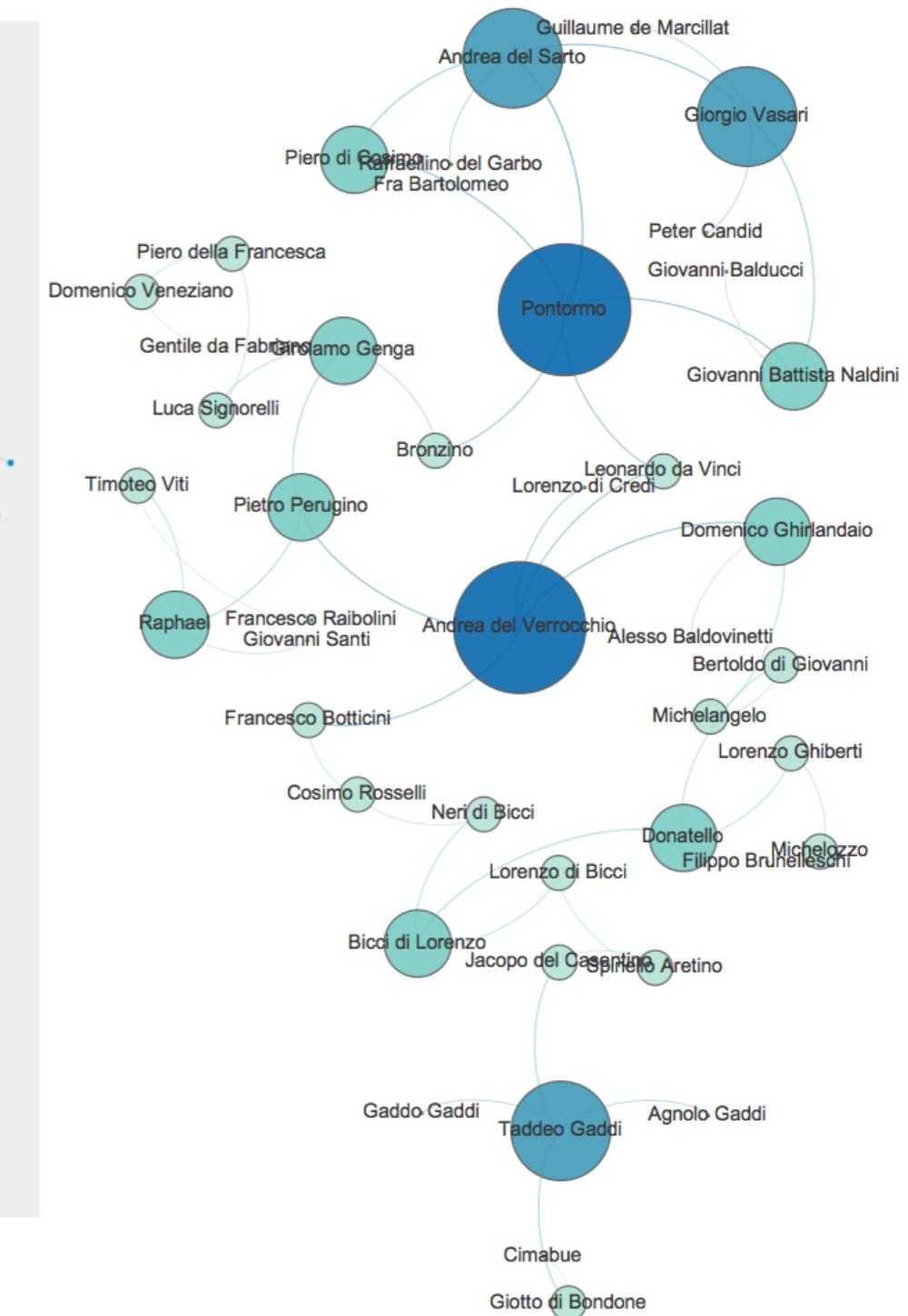
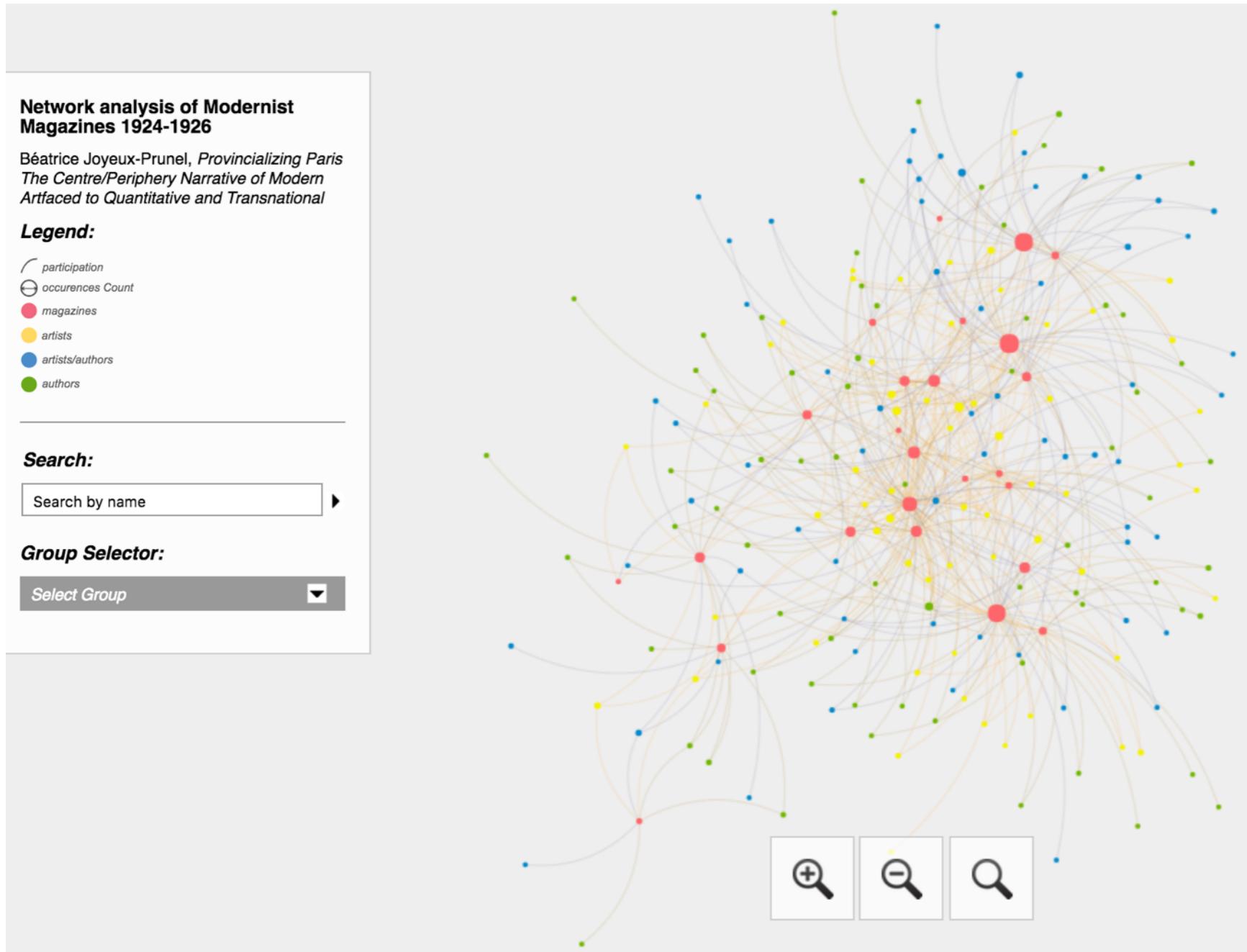
Rechercher

Que cherchez-vous ?



- SNA (Social Network Analysis)
 - Relation entre élèves-architectes et atelier
 - Corpus: les élèves nés entre 1800 et 1900
 - Ne sont affichés que les ateliers ayant au minimum 3 élèves

Expérimenter - s'amuser avec des données que l'on peut récupérer



<https://github.com/Dorialexander/Wikidaesthetics>



N° du recueil

type d'objet

Domaine thématique

lieu conservation

Répartition typo-chronologique

matière

11

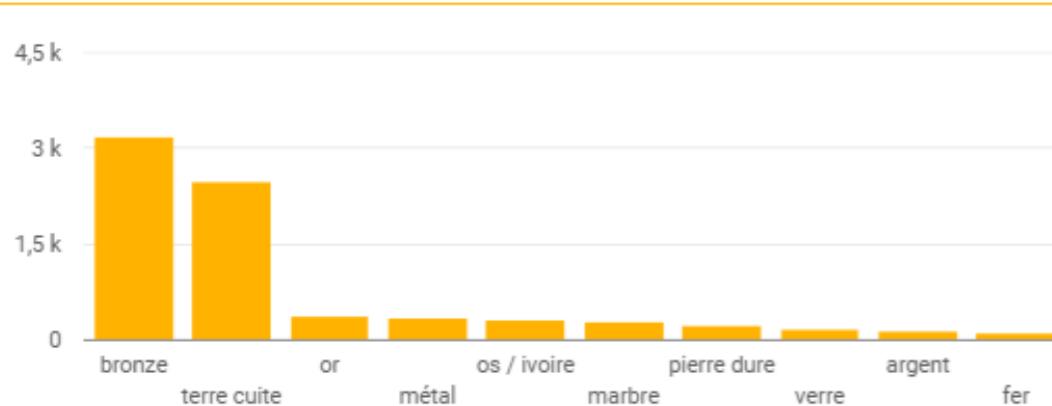
Nombre de recueils

8 035

Nombre d'objets

4 568

Nombre d'objets
légendés par Muret



Les 10 "matières" les plus utilisés

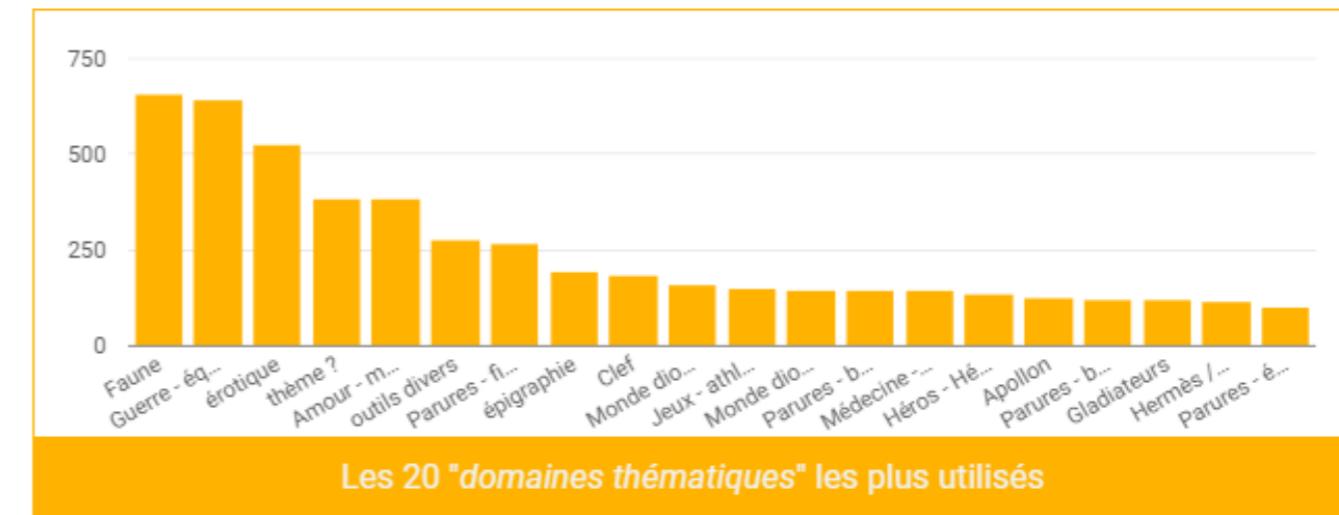
Nombre d'objets restant à passer en revue

7 433

7,5%

92,5%

VU



Dépouillement des 11 recueils de J-B Muret <https://goo.gl/hHhBWz>

L'importance du dynamisme/ interactivité

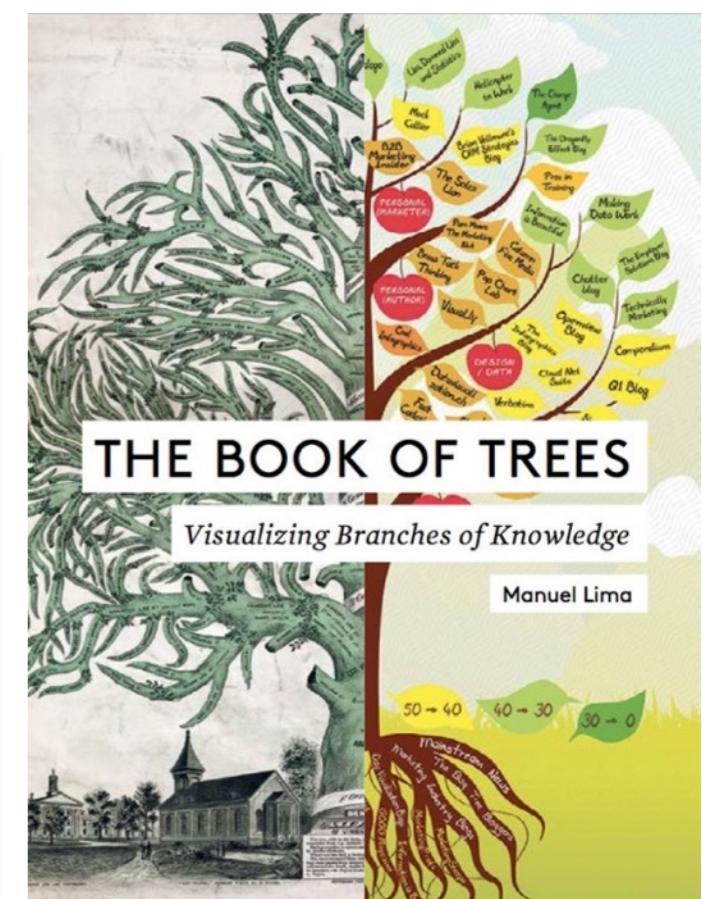
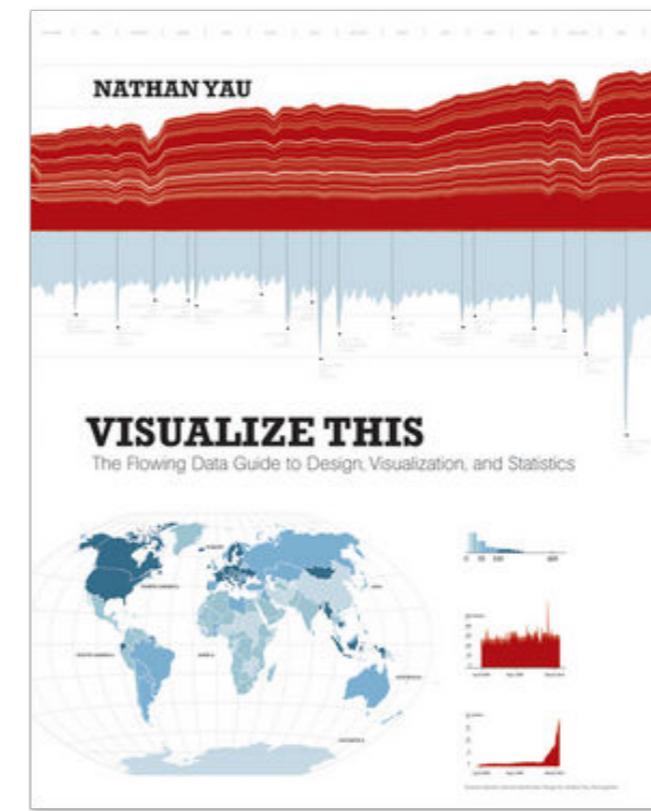
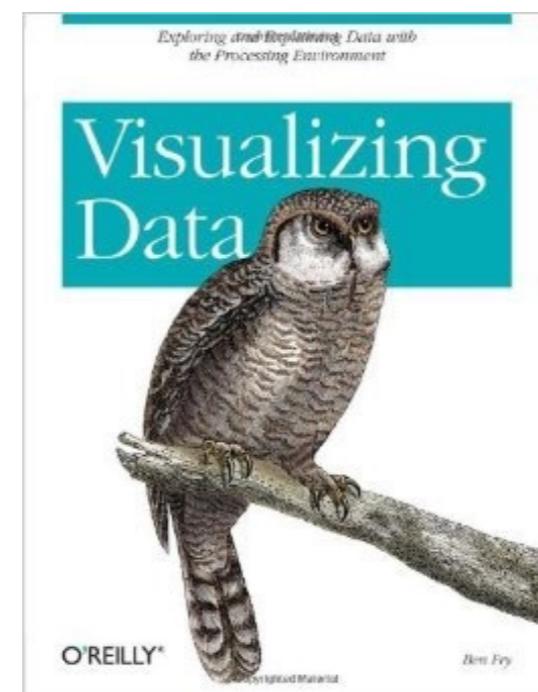
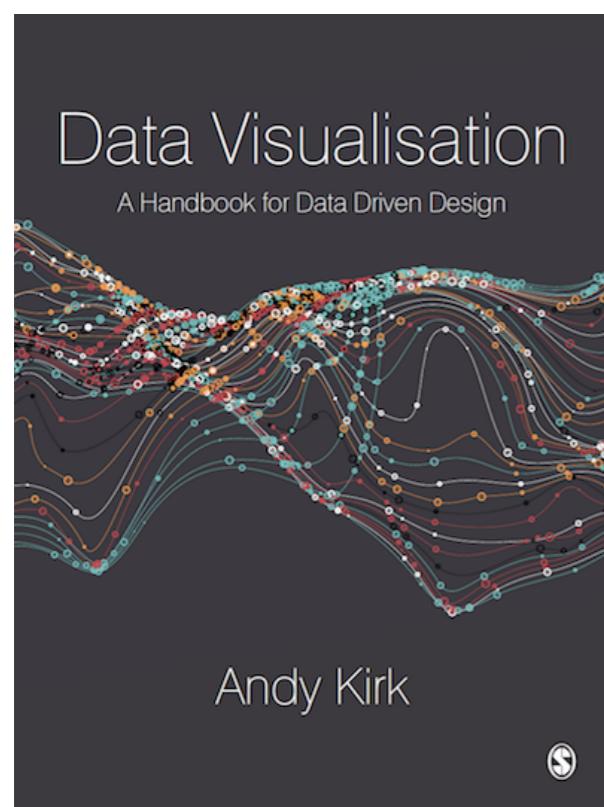
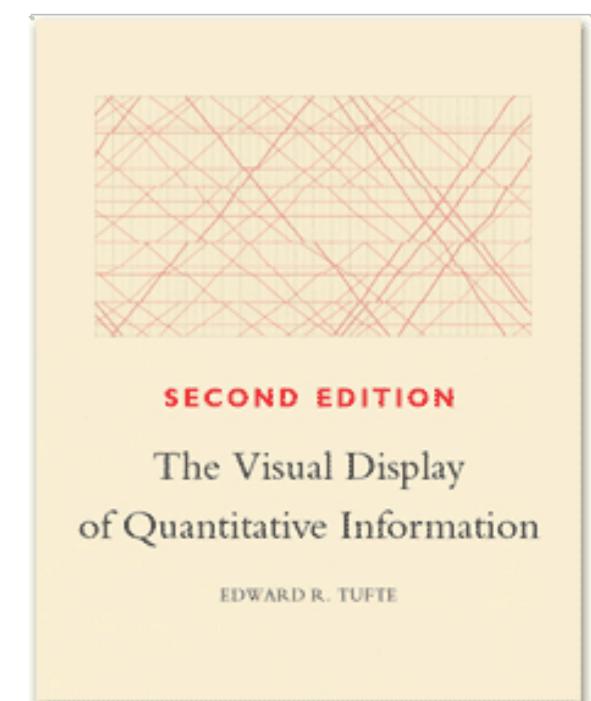
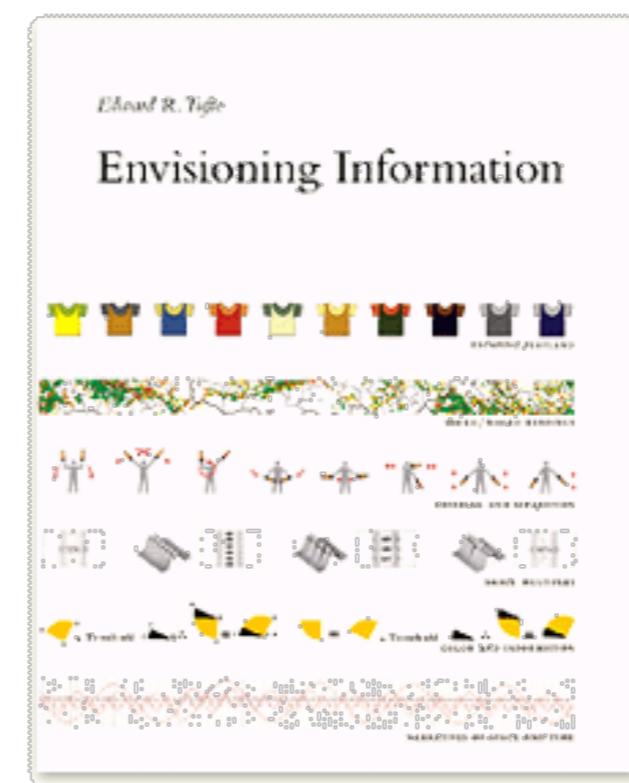
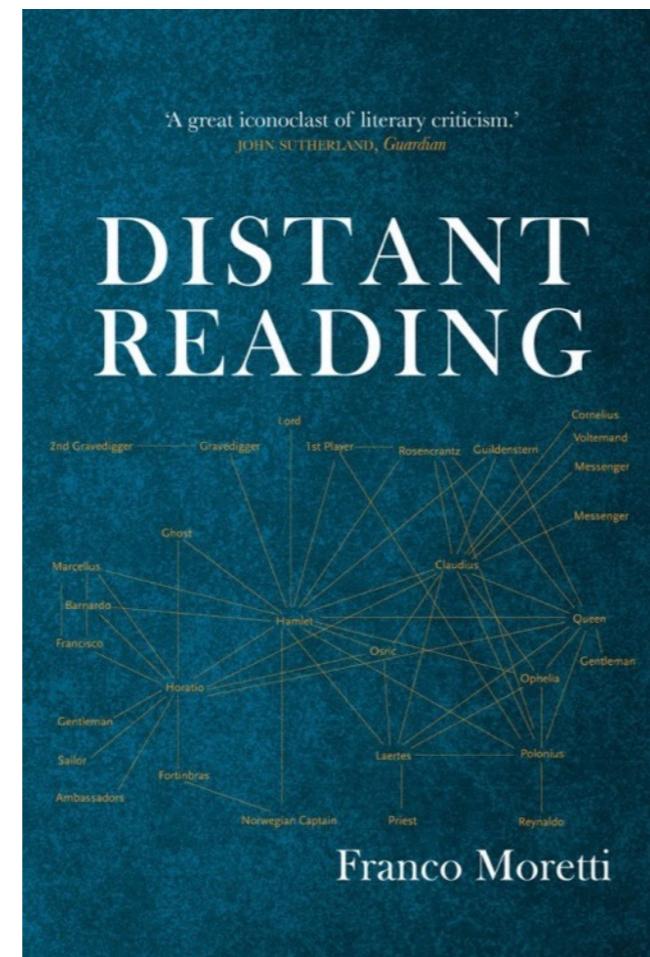
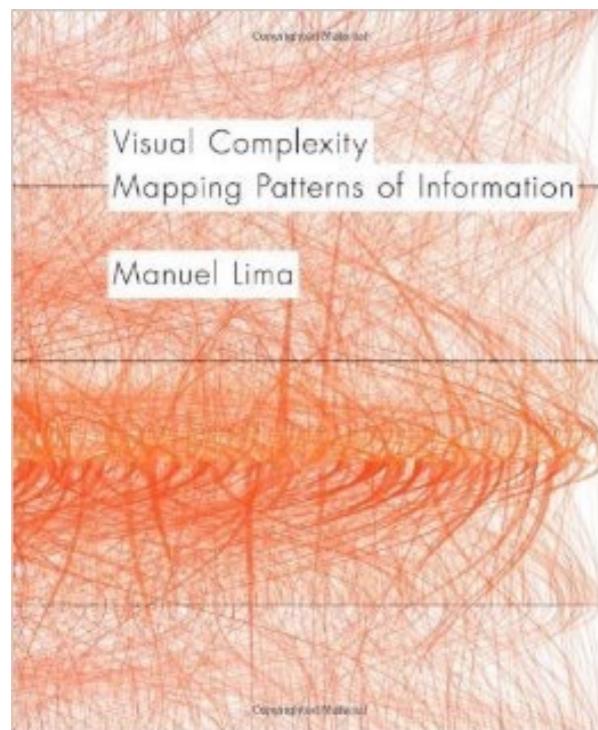


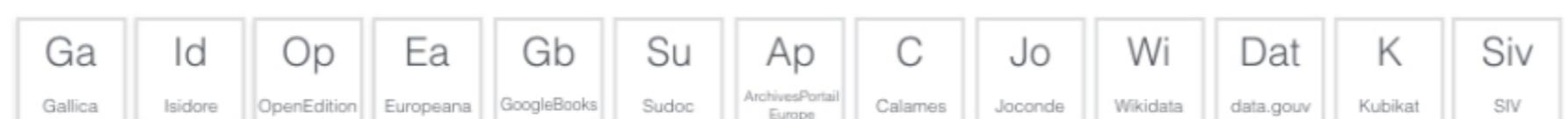
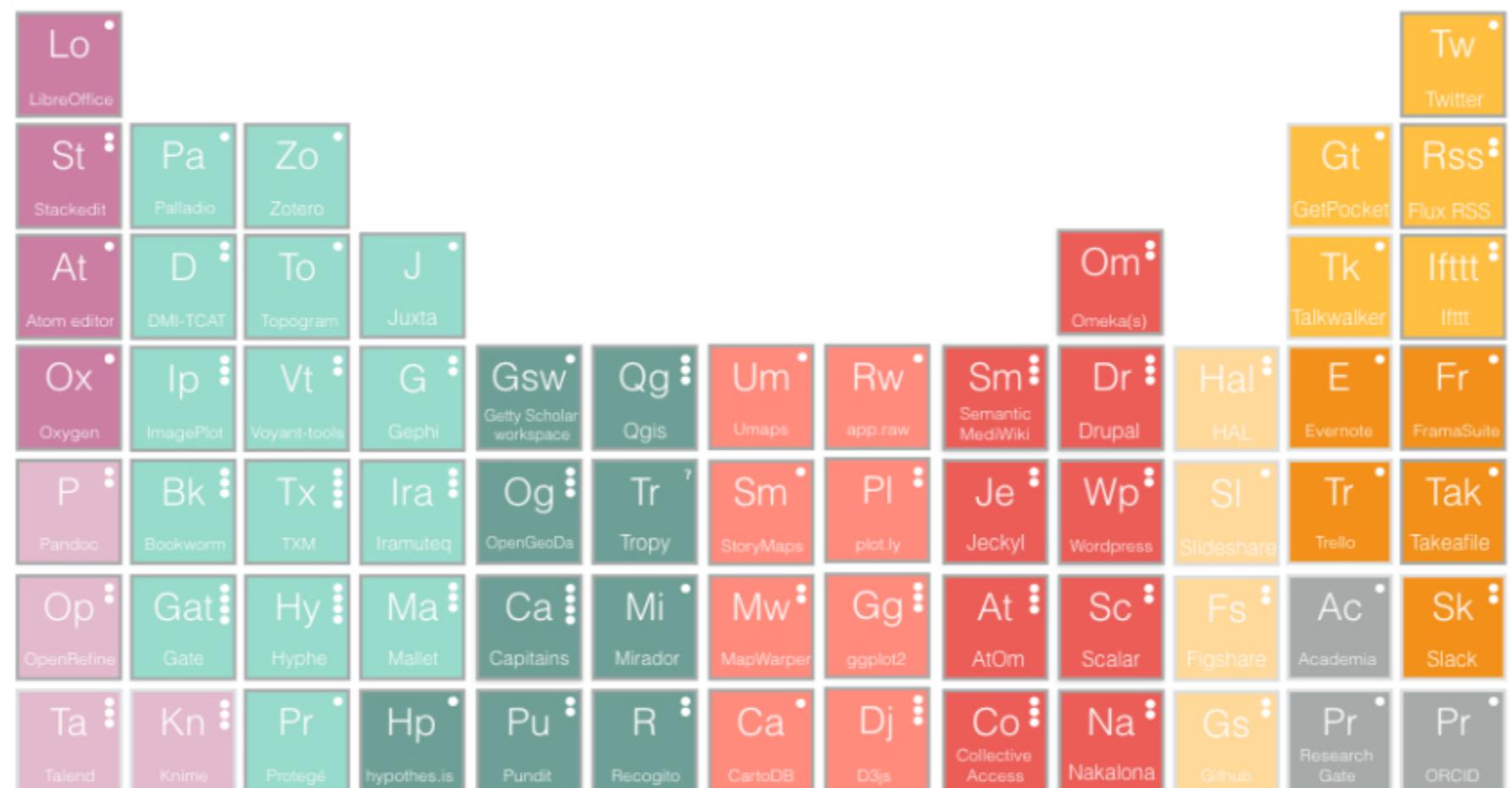
L'apport de la publication web



Boite noire ? / Algorithmes

$$\mu_3 = \frac{12,648 + 7,8}{(7+9+2,5)} \left(\frac{1}{4} + \frac{1}{9} \right) = 0,3709$$
$$\mu_1 = 0,41 \quad \mu_2 = 0,508 \quad \mu_3 = 0,609$$





Exploiter/révéler ses données

Gérer des données/Outils DH

Système de publication

Visualiser sous d'autres formes ces données

Partager sa recherche

Veiller

Gestion de projets

Identité du chercheur

Système de stockage

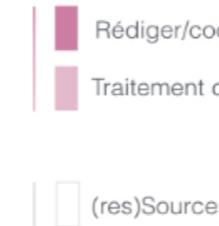
Langages

Interrogation/manipulation

Syntaxe de structuration de l'information

Rédiger/coder

Traitement de données



1

An icon showing four circles arranged in a horizontal row above a thick horizontal arrow pointing to the right.

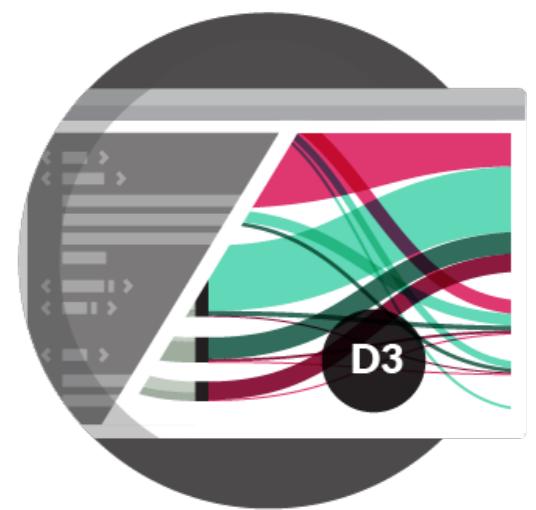


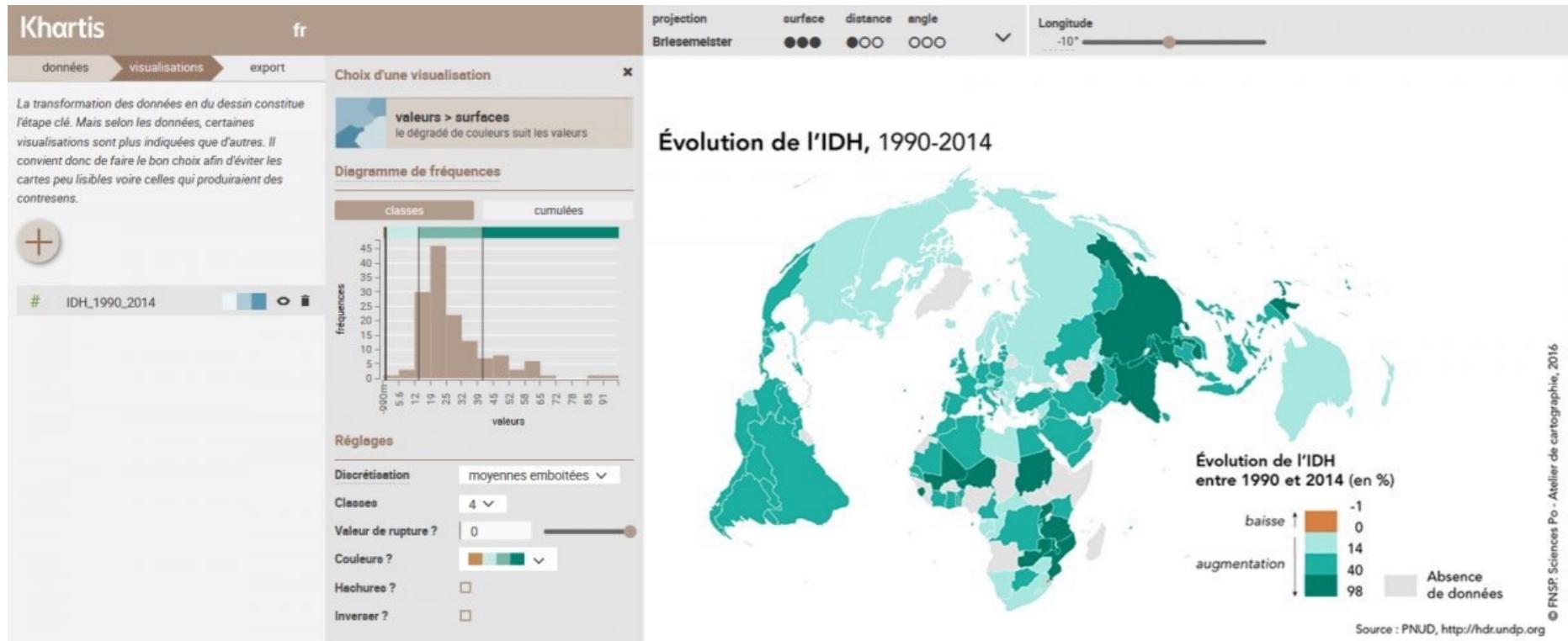
Weviz
Démarche de création collaborative

<https://goo.gl/rwoHNu>

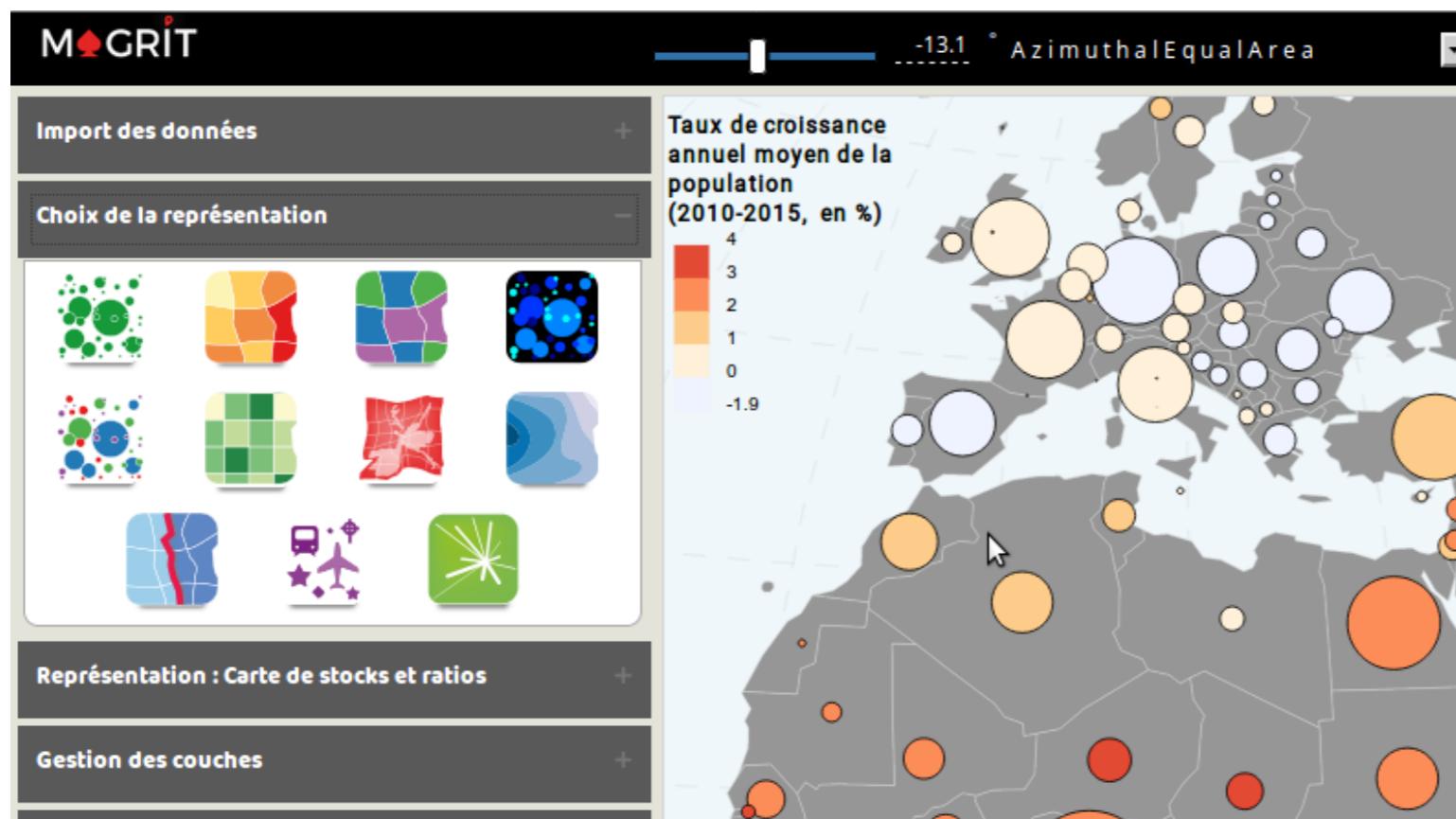


WIKIDATA





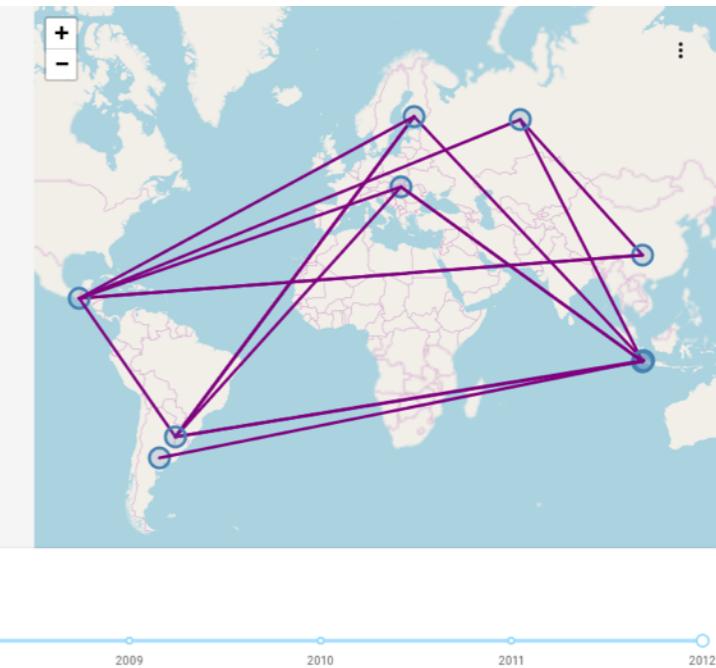
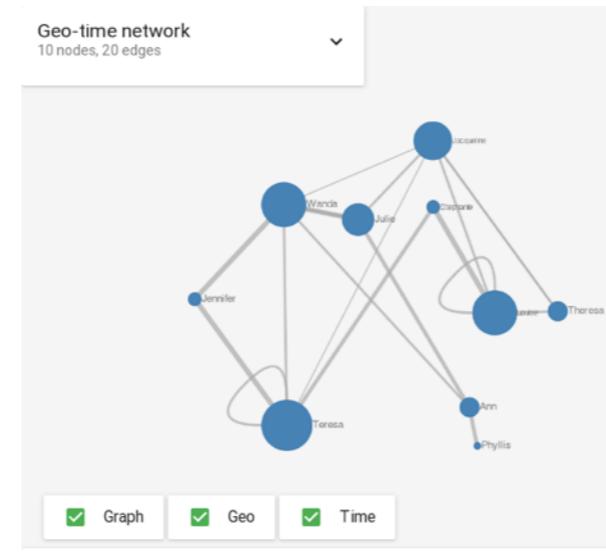
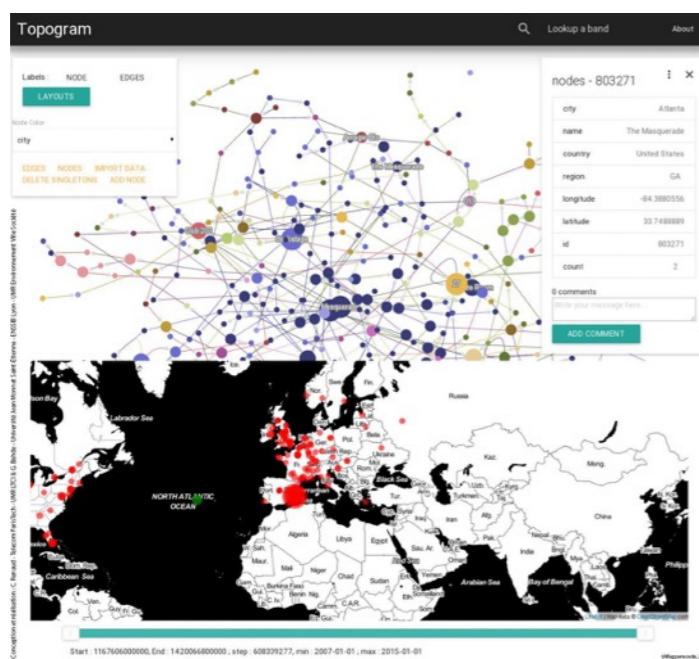
<http://www.sciencespo.fr/cartographie/khartis/>



<http://magrit.cnrs.fr/>



<https://goo.gl/btXWbj>



<http://topogram.io/#/?id=topogram>

MUS_BIEN-export (10) - Microsoft Excel

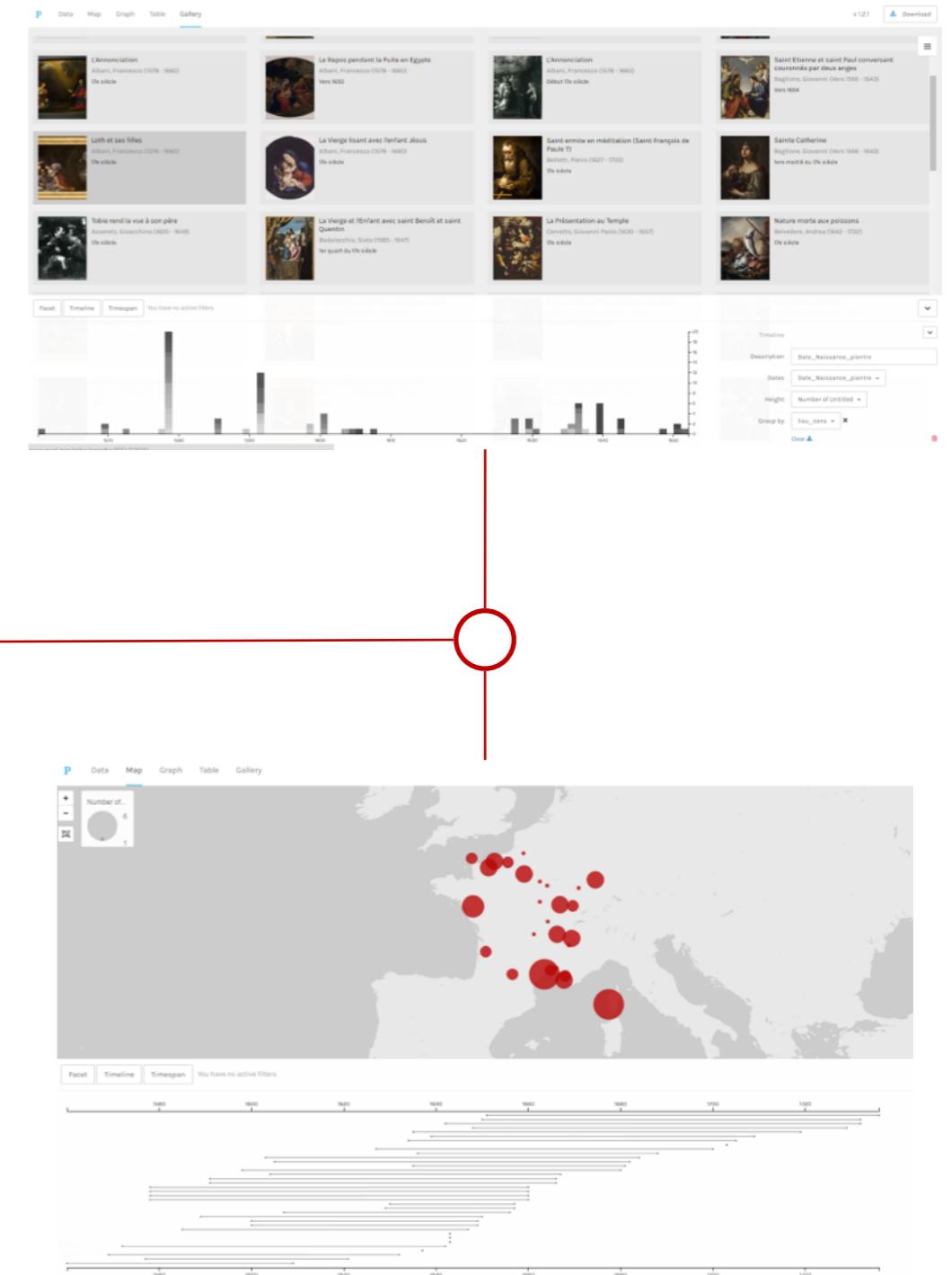
Fichier Accueil Insertion Mise en page Formules Données Révision Affichage

A1 f# NUM_NOTICE

	A	B	C	D	E	F	G	PRECISION_LIE
1	NUM_NOTI	TYPE_OEU	PRECISION_TYPE_O	TITRES	COMMENTAIRE_GEN	COTE_NUM_INV	LIEU_DE_CONSERVATION	
2	228343	peinture;		Adoration des bergers #			Guise	
3	223543	peinture;		Baptême du Christ #			Chapelle Saint-Jean-Baptiste (Petrete-Bicchisano)	
4	26390	peinture;		Bon Pasteur #			Eglise paroissiale Santissima Annunziata (Gavignano)	
5	23204	peinture;		CAMPO DE L'ÉGLISE GIOVANNI ET PAOLO AVEC LA SCUOLA DI SAN MARCO, A VENISE #		PF 2252 (numéro d'inventaire) #	Palais de l'Elysée (Paris)	
6	21021	peinture;		Copie d'après une fresque antique : Le départ de Briseïs #	1859 Cat. I (numéro d'inventaire) #		Musée Gustave Moreau (Paris)	
7	227760	peinture;		Couronnement de la Vierge #			Saverne	Eglise Saint-An
8	104764	peinture;		Crucifixion vernaculaire #			Eglise paroissiale Saint-Michel (Valle-di-Mezzana)	
9	223550	peinture;		Décollation de saint Jean-Baptiste #			Chapelle Saint-Jean-Baptiste (Petrete-Bicchisano)	
10	140966	peinture;		L'Adoration des bergers #			Eglise paroissiale (Saint-Victurnier)	
11	103553	peinture;		L'Annonciation #			Eglise paroissiale de la Sainte-Trinité (Frasseta)	
12	104163	peinture;		L'Apparition de l'Archange à saint Pierre #			Olmeto	
13	104160	peinture;		L'Apparition de l'enfant Jésus à saint Antoine de Padou #			Olmeto	

MUS_BIEN-export (10)

Prêt



palladio

Site web: <http://palladio.designhumanities.org/>

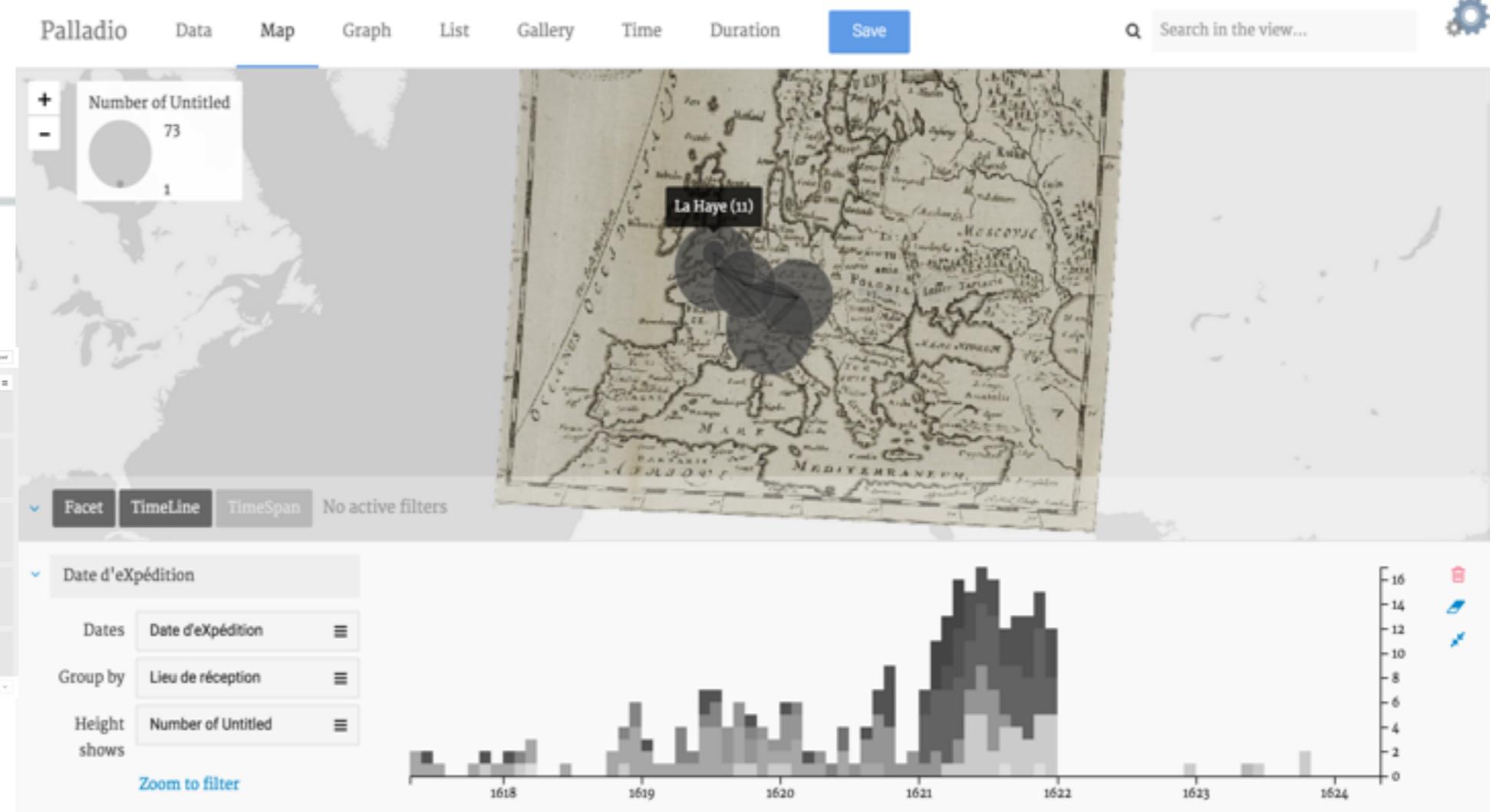
Ressources: <http://goo.gl/9hZmek>

Pour faire quoi ?

- ❖ Explorer ses données sous forme temporelle, géographique, de réseaux, tableaux, etc.

Les avantages

- ❖ Très simple d'utilisation
- ❖ Nombreux filtres dynamiques
- ❖ export



app.raw

Site web: <http://app.raw.densitydesign.org/>

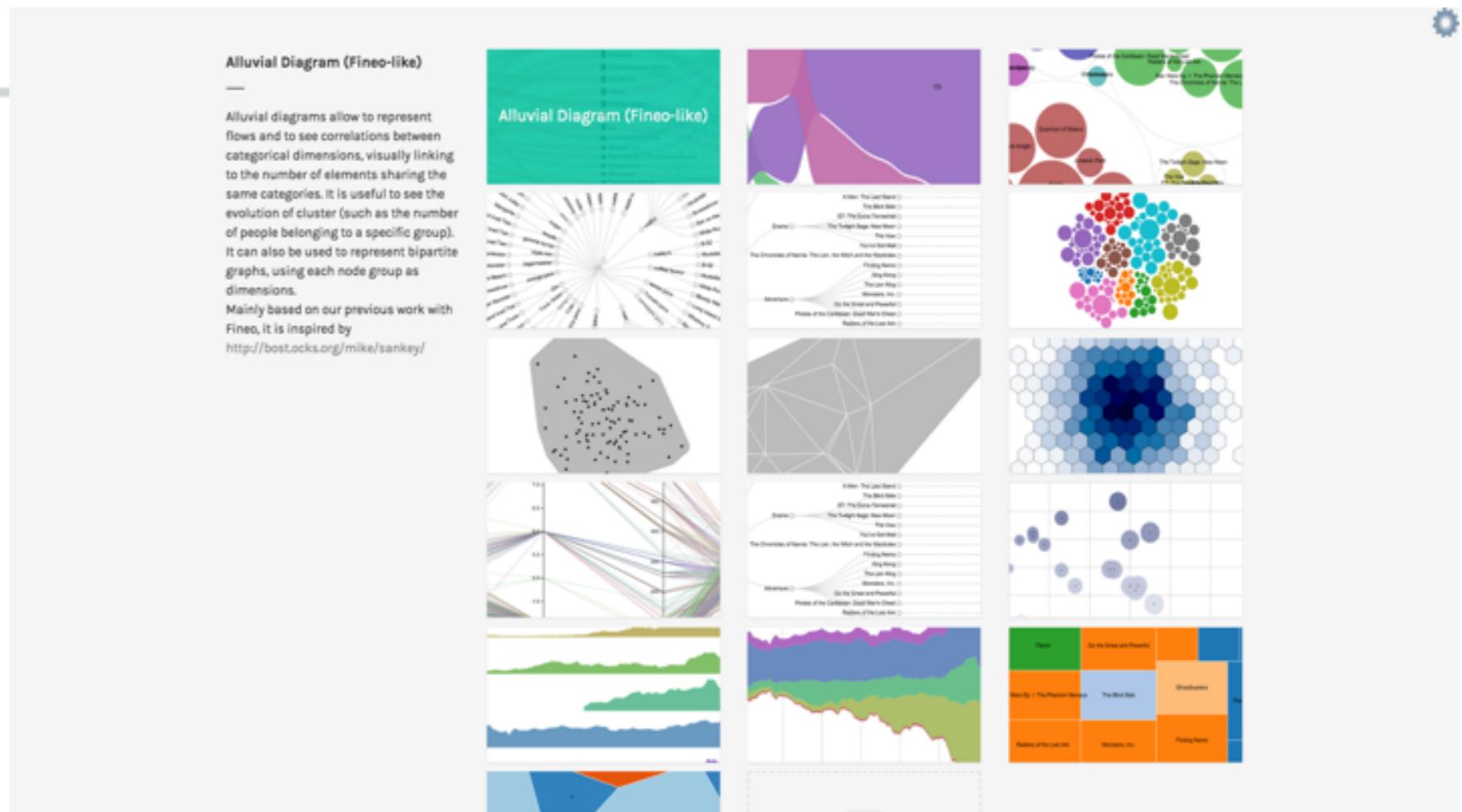
Ressources: <http://goo.gl/qu5rCK>

Pour faire quoi ?

- ❖ Visualiser des données facilement avec données en CSV en utilisant la librairie D3.js

Les avantages

- ❖ Très facile à prendre en main
- ❖ Fonctionne avec du simple csv ou données tabulées
- ❖ Quelques options selon la forme du graphique
- ❖ Exporte les visualisation en SVG, PNG et les données en json (pr ensuite, aller + loin)



gephi

Site web: <http://gephi.github.io/>

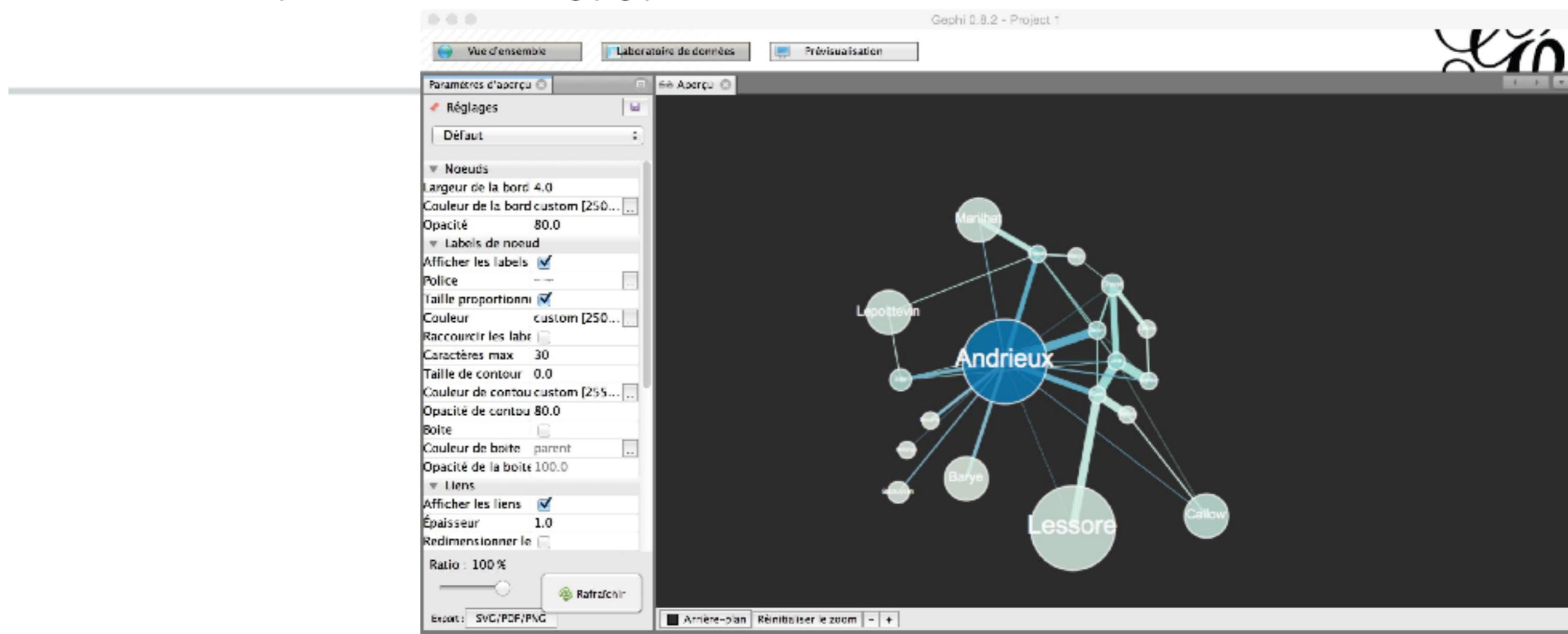
Ressources: <http://goo.gl/i48zJd>

Pour faire quoi ?

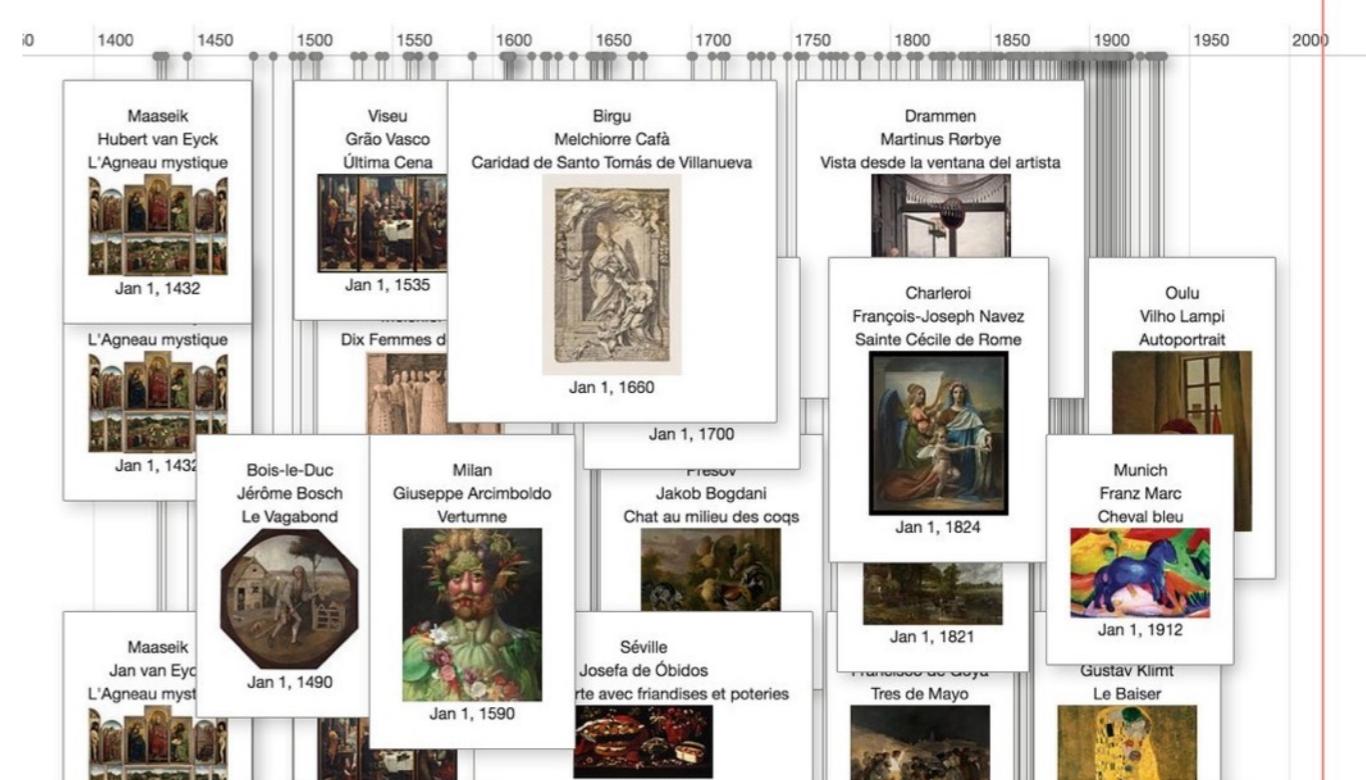
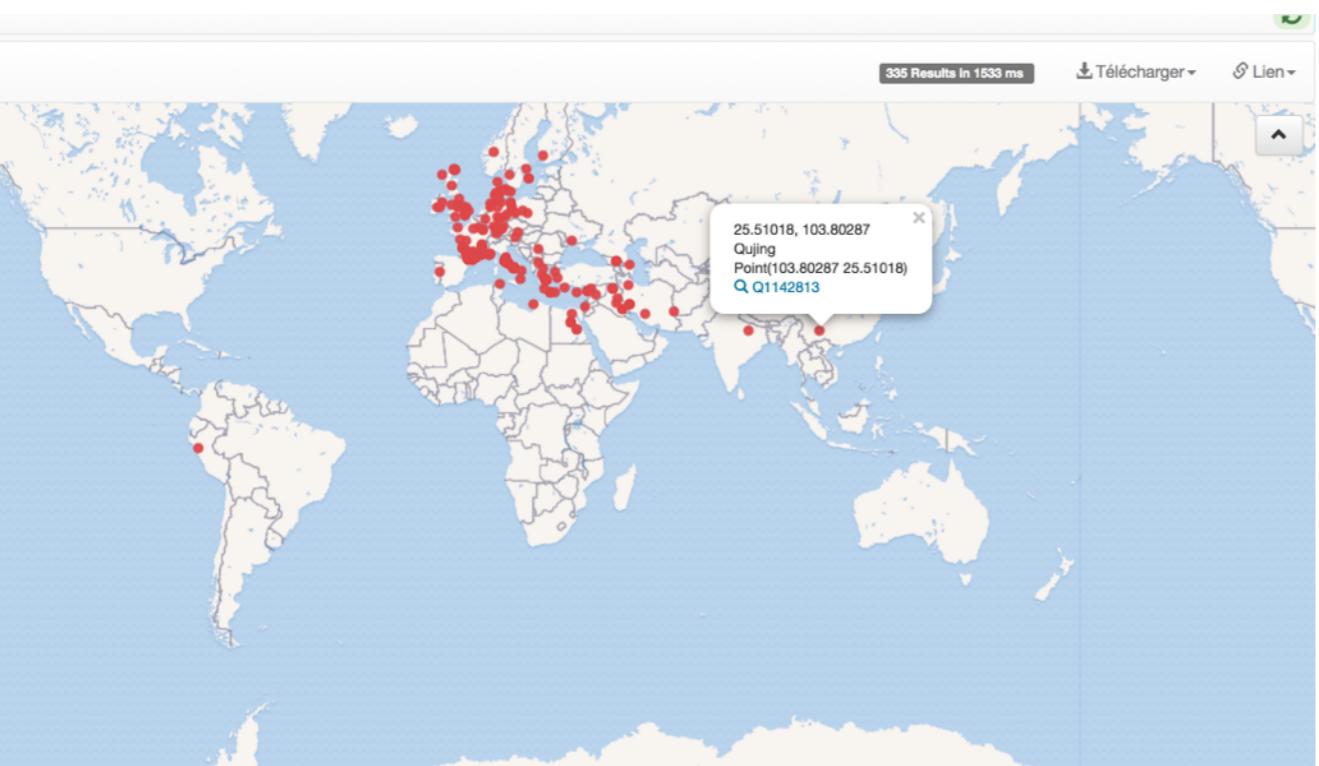
- ❖ Visualiser les données sous forme de réseaux.

Les avantages

- ❖ Logiciel très complet en open source
- ❖ De nombreux plugins
- ❖ Personnalisation des visualisations
- ❖ Nombreuses sorties possibles: format réseaux, svg, png, pdf, etc.



De la visualisation dans les interfaces de requête de données structurées



<http://tinyurl.com/yb4vxgfk>

<https://t.co/NxmadRocRu>

Pour choisir l'outil, se poser des questions :

- Qu'est-ce que je souhaite faire apparaître ?
- Quels sont les formats de mes données ?
- Quel est le type de mes données (temporelle, cartographique, statistiques, etc.) ?
- Est-ce pour explorer un jeu de données par moi-même ou l'objectif est de partager le résultat ?
- Est-ce pour du *print* ou pour une version en ligne ? Et si oui, est-il nécessaire d'avoir une version dynamique ?

Conclusion

- Visualisation très efficace mais à condition de suivre des règles strictes
- Tout ne s'y prête pas forcément
- Garder un esprit critique faces aux visualisations
- Les visualisations efficaces nécessitent un apprentissage

“On ne lit pas un graphique, on lui pose des questions”

Jacques Bertin, 1981