

WIKIDATA & HAL

Présentation, outils et ponts à construire.

Hugo Lopez

Wikimédien en résidence

hugo.lopez@univ-toulouse.org

PLAN

1. Contexte
2. Introduction à Wikidata
3. Utilisation des données
4. Interroger Wikidata
5. Éditer Wikidata
6. Tableau de bord

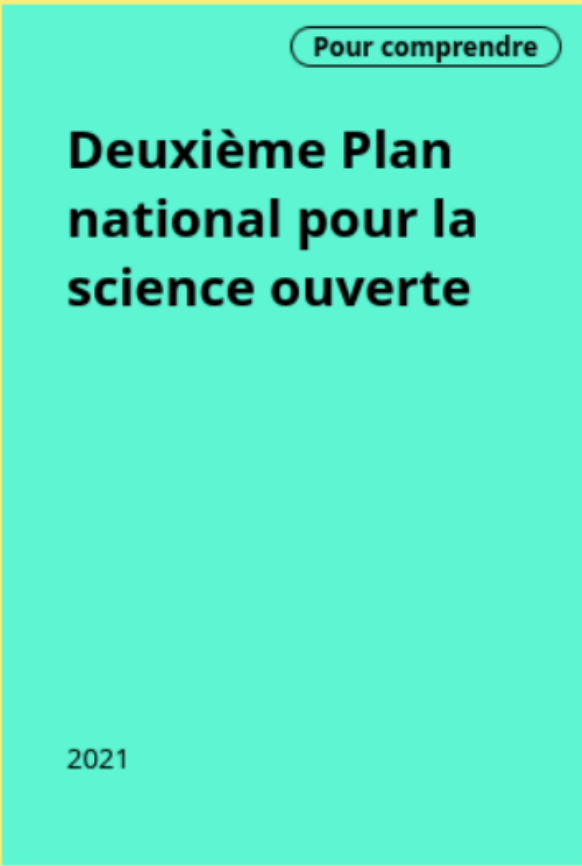
Objectifs

En fin d'atelier, le participant :

- connaîtra les concepts clefs de Wikidata
- sera capable d'éditer, corriger manuellement Wikidata
- connaîtra 4 approches d'édition sur Wikidata
- saura faire une requête SPARQL "croisée" mon institution + HAL id
- pourra réfléchir à un dashboard croisé HAL + Wikidata

CONTEXTE DE FORMATION

- Formateur : 20 ans de Ressources Éducatives Libres
- Formation : Sciences - Science Ouverte (HAL) - Wikidata
- WER : tour d'horizon, connecter les acteurs.

The image shows a document cover with a light blue background. At the top right, there is a small rounded rectangle with the text 'Pour comprendre'. In the center, the title 'Deuxième Plan national pour la science ouverte' is written in bold black font. At the bottom left, the year '2021' is visible.

Pour comprendre

Deuxième Plan national pour la science ouverte

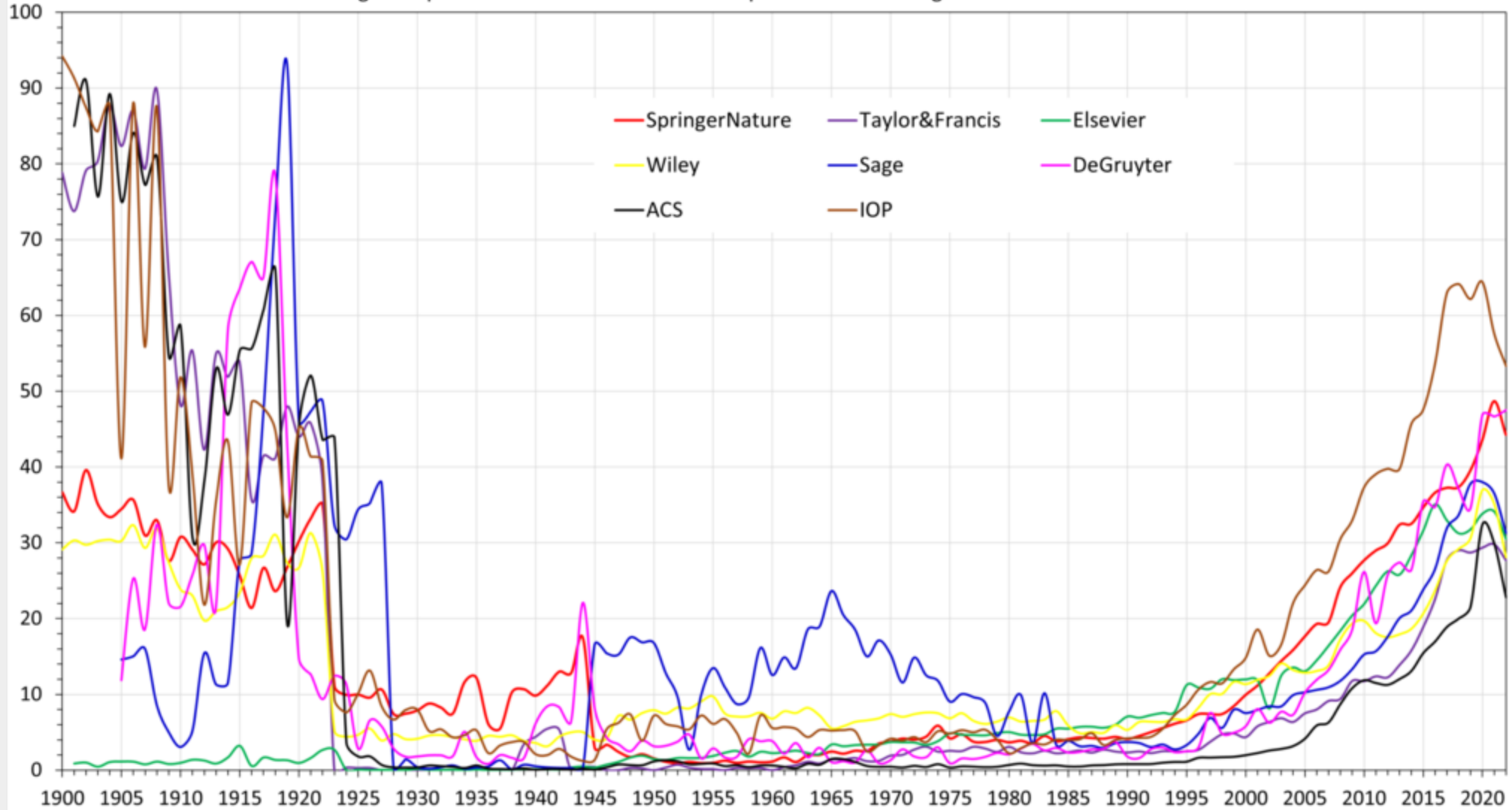
2021



Télécharger

La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche.

Percentage of Open Access articles from different publishers according to Web of Science on 2023-01-30.



Premier axe

Généraliser l'accès ouvert aux publications

8

Mesures

10

- 1 Généraliser l'obligation de publication en accès ouvert des articles et livres issus de recherches financées par appel à projets sur fonds publics
- 2 Soutenir les modèles économiques d'édition en accès ouvert sans frais de publication pour les auteurs (modèle « diamant »)
- 3 Favoriser le multilinguisme et la circulation des savoirs scientifiques par la traduction des publications des chercheurs français

Troisième axe

Ouvrir et promouvoir les codes sources produits par la recherche

16

Mesures

18

- 7 Valoriser et soutenir la diffusion sous licence libre des codes sources issus de recherches financées sur fonds publics
- 8 Mettre en valeur la production des codes sources de l'enseignement supérieur, de la recherche et de l'innovation
- 9 Définir et promouvoir une politique en matière de logiciels libres

Deuxième axe

Structurer, partager et ouvrir les données de la recherche

12

Mesures

14

- 4 Mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics
- 5 Créer Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche
- 6 Promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR)

Quatrième axe

Transformer les pratiques pour faire de la science ouverte le principe par défaut

20

Mesures

22

- 10 Développer et valoriser les compétences de la science ouverte tout au long du parcours des étudiants et des personnels de la recherche
- 11 Valoriser la science ouverte et la diversité des productions scientifiques dans l'évaluation des chercheurs et enseignants-chercheurs, des projets et des établissements de recherche
- 12 Tripler le budget de la science ouverte en s'appuyant sur le Fonds national pour la science ouverte et le Programme d'investissements d'avenir

Formation récente

Gossa, Godefroid et Beaudouin (2023.06.13:p7-27)

Données de la recherche et Wikidata (en ligne)

- Comprendre les enjeux liés au partage et à l'ouverture des données
- Découvrir les bases de Wikidata
- Découvrir les outils de visualisation et d'export de données
- Découvrir des exemples de réutilisations des données de Wikidata

Tableau de bord multi-sources

- TDB des Langues

Languages

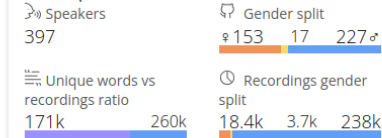
[Log in and record few words](#)
 Search by language name


Languages with over 20k recordings (10)

For activated languages [log in and start recording vocabulary](#), most languages have vocabulary lists ready to record : at Step 3, search "List: {your_ISO}/Unilex".

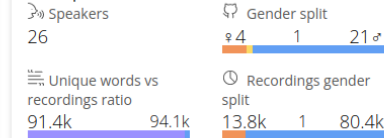
Langue Française

210M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

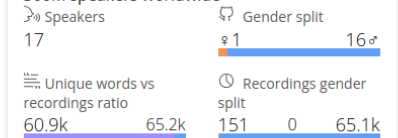
Polszczyzna

40M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

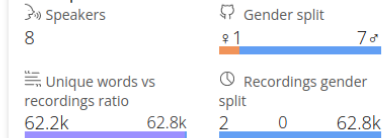
বাংলা

300M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

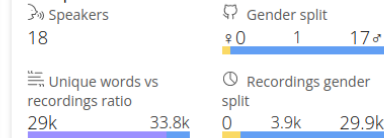
ଓଡ଼ିଆ ଭାଷା

35M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

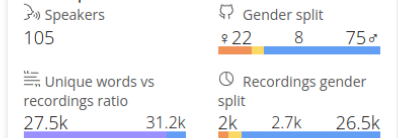
Esperanto

2M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

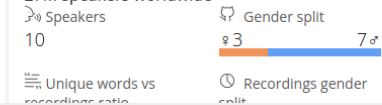
English

750M speakers worldwide


[CONTRIBUTE](#)
[DOWNLOAD](#)

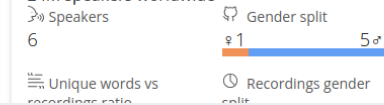
Українська Мова

27M speakers worldwide



Românește

24M speakers worldwide



Català

5.1M speakers worldwide

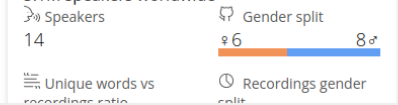


Tableau de bord multi-sources

Grace à une jointure via le Qid Wikidata...

- [Wikidata] Language name
- [Wikidata] Number of native speakers worldwide.
- [Lingualibre] Number of contributing speakers.
- [Lingualibre] Gender split.
- [Lingualibre] Unique words vs recordings ratio.
- [Lingualibre] Recordings gender split.

DÉCOUVRONS NOS USAGES

Vos usages

- A) Combien de fois consultez-vous Wikidata / semaine :
<10 ?
<100 ?
<1000 ?
- B) Consommez-vous Wikidata ?
- C) Qui a déjà édité ?
- D) Qui a lu une page de discussion
- E) Qui à discuté ?

Vos représentations

Fiabilité: sur wikidata, les erreurs et biais sont
























- A) fréquents
- B) occasionels
- C) rares

WIKIDATA

- Quelques éléments fondamentaux. P.Y.Beaudouin (p7-27)

UTILISATION DES DONNÉES

- Quelques éléments fondamentaux. P.Y.Beaudouin (p7-27)

		Utilisation Partage	Adaptation Modification	Utilisation commerciale	Modification de licence	
TRÈS LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
LIBRE						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
NON LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix NON permis
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix NON permis



BY

ATTRIBUTION

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

L'auteur doit être cité



SA

PARTAGE DANS
LES MÊMES CONDITIONS

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Partage sous licence compatible



NC

POUR USAGE
NON COMMERCIAL

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Pour usage non commercial



ND

PAS DE MODIFICATION

Vous pouvez retenir, réutiliser et redistribuer.

Aucune modification permise

INTERROGER WIKIDATA

- Utiliser SPARQL pour interroger Wikidata
- Questionner les éléments HAL

Ouvrir :

- Help:SPARQL: [lingualibre.org/wiki/Help:SPARQL_\(HAL\)](http://lingualibre.org/wiki/Help:SPARQL_(HAL))
- WDQS : <https://query.wikidata.org>

Propriétés notables : HAL

ItemLabel	WD Qid	sur WD	sur HAL
HAL	Q3144107	--	--
HAL author ID	P4450	1,816	127,272?
HAL structure ID	P6773	5,950	35,992
HAL article ID	P7864	716	1,199,139
HAL journal ID	P8903	29,733	113,000+?

- Manque de réconciliation

Propriétés notables : Articles

P7864

ItemLabel	WD Qid
title	P1476
author / author name string	P50 / P2093
publication date	P577
published in	P14233
volume	P478
issue	P433
page(s)	P304
main subject	P921

Propriétés notables : Revues

P8903

ItemLabel	WD Qid
inception	P571
title	P1476
main subject	P921
country of origin	P495

Propriétés notables : Auteurs

P4450

ItemLabel	WD Qid
sex or gender	P21
date of birth	P569
place of birth	P19
country of citizenship	P27
given name	P735
doctoral advisor	P184
doctoral student	P185

Propriétés notables : Structures

P6773

! ItemLabel	WD Qid
inception	P571
title	P1476
main subject	P921
country	P17
coordinate location	P625
parent organization	P749
official website	P856
ROR ID	P6782

ÉDITER WIKIDATA

1. À la main

- Osez la simplicité !

2. User script

- Hacker votre JS, boostez vos contributions !
 - `User:Myname/common.js`
- Semi-massive edits

3. OpenRefine et reconciliation

- Un outil puissant de nettoyage des données
- OpenRefine via votre compte Wikipedia :
<https://hub-paws.wmcloud.org>
- Reconciliation et export vers Wikidata

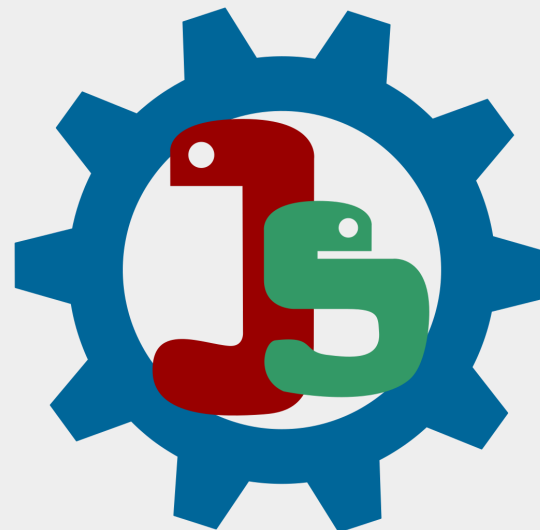
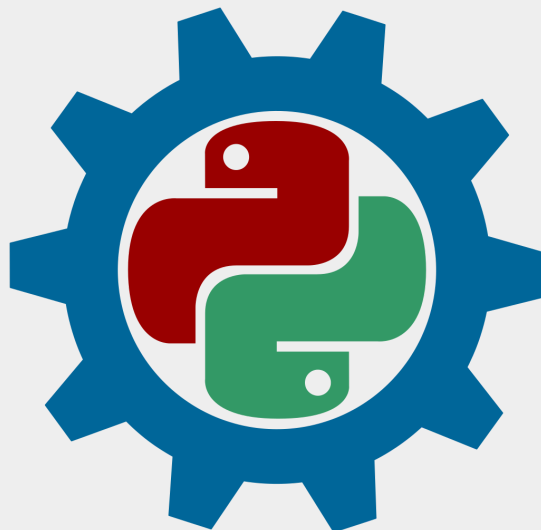
Bots

PyWikiBot

WikiapiJS

Python

Javascript



Voir aussi :mw:Manual:Creating a bot

TABLEAU DE BORD HAL + WIKIDATA

Questions

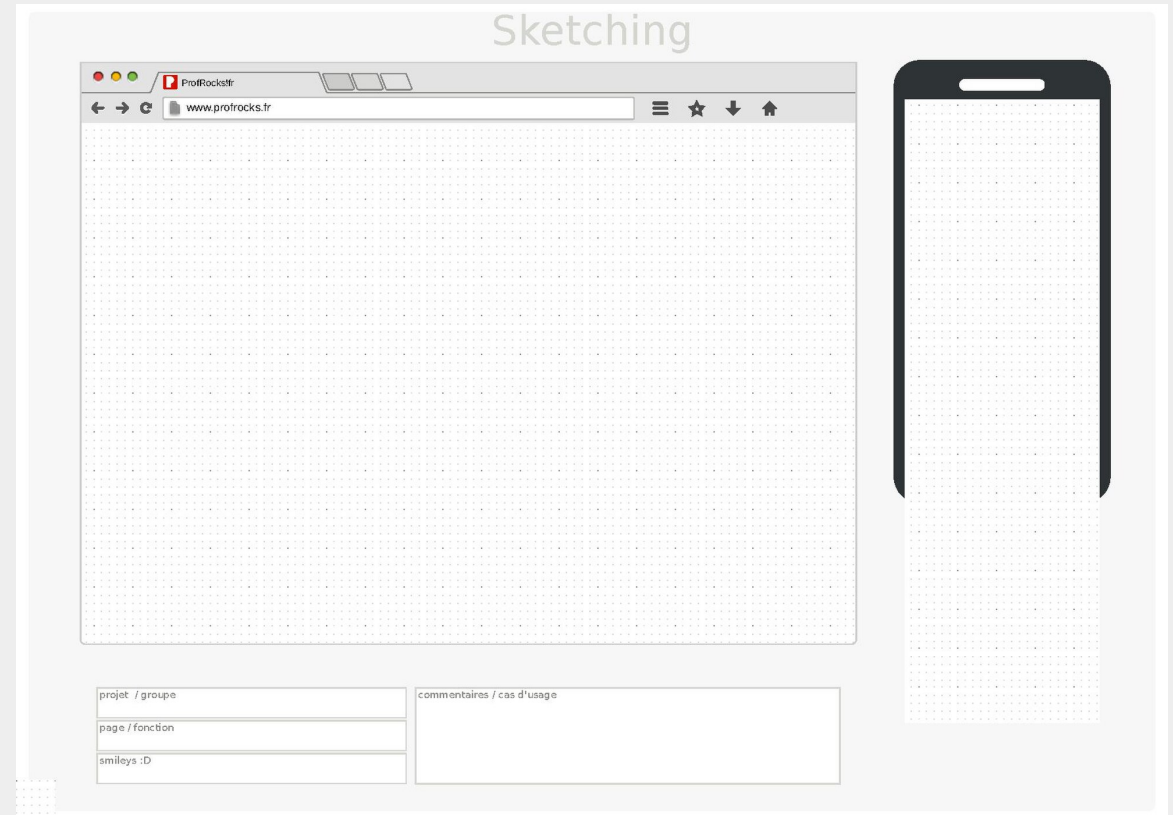
Sur vos productions et créateurs de nouveaux savoirs

- Identifier les informations disponibles: WD + HAL
- Identifier les informations intéressantes à réunir
- Quelles interfaces ?

Ébauche d'interface

Objectif: imaginer un visuel

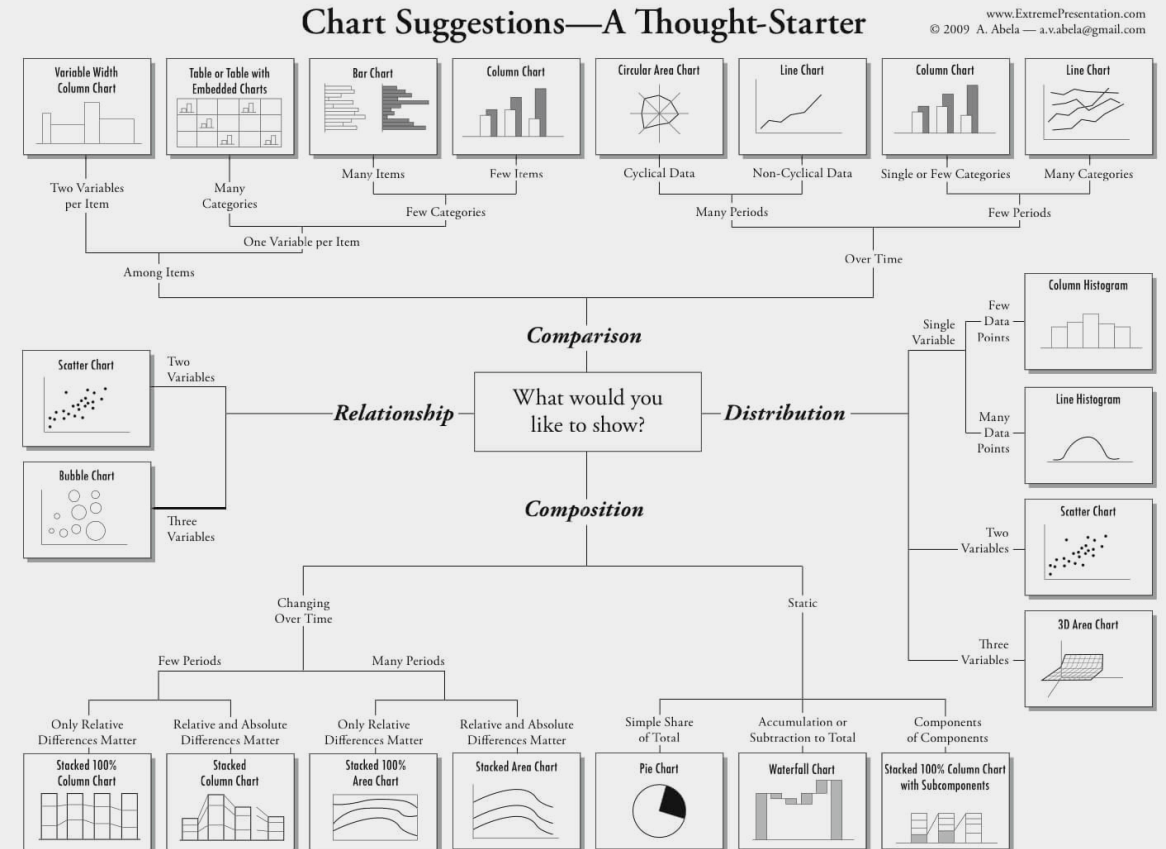
- intéressant
- lisible
- (engageant)



Ébauche d'interface

Objectif: imaginer un visuel

- intéressant
- lisible
- (engageant)



POUR ALLER PLUS LOIN

Métriques par article : URFIST

- Page statistics
- Pageviews

Permet d'identifier les éditeurs actifs.

Restons en contact

Institution	Wikimédien en résidence	Specialité
URFIST Occitanie	User:Hugo en résidence hugo.lopez@univ-toulouse.fr	Dev web, langues
Musée de Bretagne	User:VIGNERON en résidence	Wikidata
URFIST Bretagne+	User:Juliette en résidence	Bibliothèques
URFIST Bordeaux	User:Pyb en résidence	Com/réseau

Formateurs aux communs numériques libres.
Contactez nous

Restons en contact

Projet:Wikifier la
science/Lettre
WikiL@b



**Pierre-Yves
Beaudouin**

Urfist
de Bordeaux

Bordeaux



**Juliette
Halimi**

Urfist de Bretagne
et des Pays de la Loire

Rennes



**Hugo
Lopez**

Urfist
Occitanie

Toulouse

Remerciements

Tous les contenus textuels sont sous licence libre.

Crédit: Hugo Lopez, Amélie Charles, Pierre-Yves Beaudouin CC-BY-SA 4.0.

Les illustrations sont la propriété de leurs auteurs respectifs.