

INFORMATION NUMÉRIQUE

Enjeux et Pratiques

Licence 3 / UE502 – Accès expert à l'information – 2024-2025

Initiation à la bibliométrie - session 2

valerie.bonvallot@inist.fr
lucile.bourguignon@inist.fr
philippe.houdry@inist.fr

Sommaire : initiation à la bibliométrie

6 mars 2025

1. Exemples, définition et usages
2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

20 mars 2025

- 3. Indicateurs bibliométriques**
- 4. Méthodologie de réalisation d'études bibliométriques**

NB : Support inspiré des cours « Initiation à la bibliométrie » de Lucile Bourguignon UL, 2023-2024, « Bibliométrie » de Valérie Bonvallot 2021-2022

3. Indicateurs bibliométriques

3. Indicateurs bibliométriques

Qui ?

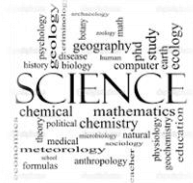
Auteur(s)
Organisme(s)
Pays
Continent(s)



Production scientifique

Fait quoi ?

Mots-clés
Classification
Texte intégral
Résumé



Contenu thématique

Où ?

Organisme(s)
Pays
Continent(s)



Répartition géographique

Quand ?

Date de publication



Evolution temporelle
Tendances

Avec qui ?

Auteur(s)
Organisme(s)
Pays
Continent(s)



Coopération(s)
Relation(s)
scientifique(s)

Comment ?

Type(s) de document
Langue(s)
Revue(s)
Editeur(s)
Financement
OA (open access)



Stratégie d'édition

3. Indicateurs bibliométriques

3.1. Indicateurs sur la production scientifique (Qui ?)

Calcul du nombre de publications au niveau de :

- Chercheur.se
- Organisme de recherche (laboratoire, institut, université)
- Pays
- Continent

⇒ **Interrogation des bases de données** : champs *Auteurs*, *Affiliations* (=adresse d'un organisme de recherche, signature de l'auteur), *Pays...* : penser à toutes les formes d'écriture + historique et évolution du nom + identifiants (ex. nom de jeune fille, un laboratoire, université de Lorraine, France et outremer, liste de pays pour un continent ...)

⇒ **Traitement des données** : il peut exister plusieurs formes d'écriture pour un même auteur, une même affiliation. Il est nécessaire d'homogénéiser les données : une forme préférentielle, un identifiant

3. Indicateurs bibliométriques

3.1. Indicateurs sur la production scientifique (Qui ?)

Calcul du nombre de publications au niveau de :

- Chercheur.se → repérage d'experts
- Organisme de recherche (laboratoire, institut, université)
- Pays
- Continent

⇒ **Interrogation des bases de données** : champs *Auteurs*, *Affiliations* (=adresse d'un organisme de recherche, signature de l'auteur), *Pays...* : penser à toutes les formes d'écriture + historique et évolution du nom + identifiants (ex. nom de jeune fille, un laboratoire, université de Lorraine, France et outremer, liste de pays pour un continent ...)

⇒ **Traitement des données** : il peut exister plusieurs formes d'écriture pour un même auteur, une même affiliation. Il est nécessaire d'homogénéiser les données : une forme préférentielle, un identifiant.

<input type="checkbox"/> Adamatzky, Andrew	49	<input type="checkbox"/> Andy Adamatzky	3
<input type="checkbox"/> Adamatzky, Andrew I.	8	<input type="checkbox"/> Adamatkzy, Andrew	2

3. Indicateurs bibliométriques

3.1. Indicateurs sur la production scientifique (Qui ?)

Zoom sur les identifiants auteurs : code alphanumérique unique pérenne (homonymie, changement de nom ...)

International

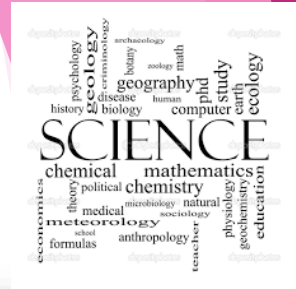
- researcherId (WoS)
- ORCID (Open Researcher and Contributor ID)
- id Ror (Research Organization Registry)

France

- idRef auteur (ABES, liens entre identifiants auteurs : orcid, idRef, idHal)
- idHal (HAL)
- RNSR (Répertoire National des Structures de Recherche)

3. Indicateurs bibliométriques

3.2. Indicateurs de contenu thématique (A propos de quoi ?)



Calcul du nombre de publications au niveau du contenu thématique :

- Indexation : Mots-clés auteurs / Descripteurs - Index. manuelle / automatique
- Classifications : au niveau de l'article ou du périodique ?
- A partir du résumé ou du texte intégral (full text) ?
- ➔ Interdisciplinarité : cooccurrence de termes

Author Keywords: **Physarum polycephalum**; Slime mold; Population dynamics; Microbial ecosystem

Web of Science Categories

Biology; Mathematical & Computational **Biology**

3.3. Indicateurs sur la répartition géographique (Où ?)

Calcul de la répartition géographique des publications à partir de l'affiliation de la publication (le champ « Adresses »).

- Organisme de recherche (laboratoire, institut, université)
- Niveau local, régional, national
- Pays
- Continent

Même logique d'interrogation et de traitement que la question « QUI ? »

Corresponding Address: Harrington, Kyle (corresponding author)

▼ Max Delbrück Ctr Mol Med, D-13125 Berlin, Germany

Addresses :

▼ ¹ Univ Idaho, Moscow, ID 83844 USA

▼ ² Max Delbrück Ctr Mol Med, D-13125 Berlin, Germany

³ Ctr Adv Syst Understanding CASUS, D-02826 Gortitz, Germany

3. Indicateurs bibliométriques

3.4. Indicateurs sur la périodicité (QUAND ?)

Répartition de l'évolution temporelle des publications sur une période donnée (plusieurs années). Utile pour montrer les grandes tendances d'évolution (émergence, disparition de thématiques).

⇒ Traitement possible des données au préalable : bien homogénéiser les dates de publication (ex: « 2 023 », « 2023 » => écritures différentes pour une même année).

!! Changement de politique de la base : modification de la couverture, de la structure

3. Indicateurs bibliométriques

3.4. Indicateurs sur la périodicité (QUAND ?)

Répartition de l'évolution temporelle des publications sur une périodicité donnée (sur plusieurs années par exemple). Utile pour montrer les grandes tendances d'évolution.

⇒ Traitement possible des données au préalable : bien homogénéiser les dates de publication (ex: « 2 023 », « 2023 » => écritures différentes pour une même année).

!! Changement de politique de la base : modification de la couverture, de la structure

3.5. Indicateurs sur les coopérations et relations scientifiques (AVEC QUI ?)

Calcul des coopérations des auteurs, des organismes et des régions, des pays. Mise en évidence des co-publications, des réseaux

Même logique d'interrogation et de traitement que la question « QUI ? »

Enrichir ces données : tag sur les organismes (privés/publics), tag sur les localisations (régions)

3. Indicateurs bibliométriques

3.6. Indicateurs sur les stratégies d'édition (COMMENT ?)

Répartition par stratégie d'édition :

- Type de document (article, ouvrage, chapitre d'ouvrage, acte de conférence...)
- Éditeur (Elsevier, Springer...)
- Source (revue, ouvrage, conférence...)
- Langue de publication (anglais, français, allemand...)
- Sources de financement (projet ANR, subvention européenne...)

3. Indicateurs bibliométriques

3.6. Indicateurs sur les stratégies d'édition (COMMENT ?)

Répartition par stratégie d'édition :

- Type de document (article, ouvrage, chapitre d'ouvrage, acte de conférence...)
- Éditeur (Elsevier, Springer...)
- Source (revue, ouvrage, conférence...)
- Langue de publication (anglais, français, allemand...)
- Sources de financement (projet ANR, subvention européenne...)

3.7. Indicateurs sur la Science ouverte (l'accès ouvert - AO ou Open access - OA)

Répartition des publications en accès ouvert (oui, non...), par type d'accès ouvert (bronze, gold, green, hybrid, closed...)

- ⇒ Travail d'enrichissement des données à partir d'*Unpaywall* (base de données qui recense les publications en libre accès sur le web à partir de leur numéro *DOI*, Digital Object Identifier).
- ⇒ Ces indicateurs sont de plus en plus demandés et deviennent obligatoires dans les études du CNRS.

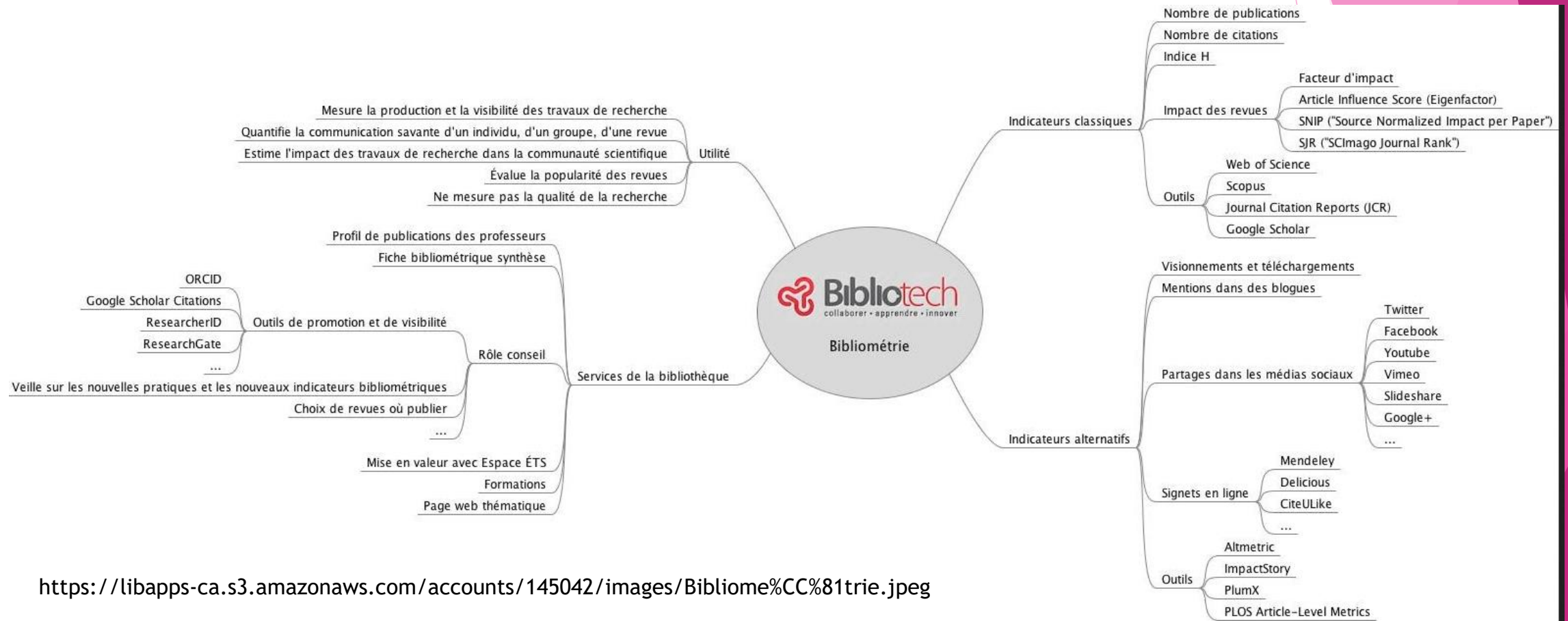


3. Indicateurs bibliométriques

3.8. Indicateurs de notoriété ou de visibilité = impact / citations

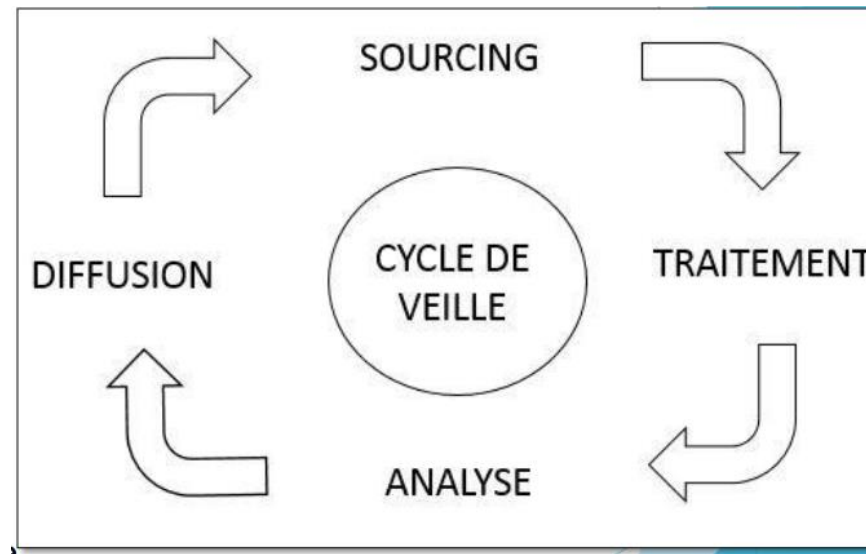
- **Le facteur d'impact (FI ou impact factor, IF)** : indicateur créé par Clarivate Analytics (WoS), qui estime directement la visibilité d'une revue scientifique, calculé au niveau d'une revue pour une année donnée (le FI d'une revue est égal à la moyenne des nombres de citations des articles de cette revue publiés durant les 2 années précédentes).
 - ⇒ Une revue avec un facteur d'impact élevé serait considérée comme plus importante qu'une revue avec un FI faible.
- **Le « h index »** : un.e chercheur.se obtient un indice de h quand il a publié h articles, et que chacun a été cité au moins h fois.
 - ⇒ L'indice ne sera jamais plus élevé que le nombre de publications du chercheur.se (pénalisation des jeunes chercheur.ses). Il assimile productivité et impact.

- ➡ Ces indicateurs de notoriété ou visibilité sont de moins en moins utilisés car décriés (autocitations, qualité ?, différences selon les domaines ...)
- ➡ Internet ➔ statistiques d'usage et métrique alternative (altmetrics)



<https://libapps-ca.s3.amazonaws.com/accounts/145042/images/Bibliome%CC%81trie.jpeg>

4. Méthodologie de réalisation d'une étude bibliométrique



4. Méthodologie de réalisation d'une étude bibliométrique

Méthodologie de réalisation d'une étude bibliométrique ?

➡ 2 grands types d'études les plus couramment réalisés :

- Etude sur la production scientifique d'une entité de recherche (ex : laboratoire, département universitaire, université, institut, organisme de recherche...)
- Etude sur la production scientifique sur une thématique (ex : énergie, blob...)

4. Méthodologie de réalisation d'une étude bibliométrique

4.1. Identification des besoins du commanditaire

- **Périmètre et objet de l'étude** : étude thématique (ex : énergie, blob...) ou étude sur la production scientifique d'un acteur de la recherche (ex : laboratoire, structure de recherche, université...)
- **Période souhaitée** : année ou fourchette d'années
- **Indicateurs souhaités**
- **Bases de données bibliographiques à interroger** : pluridisciplinaire, thématique, propriétaire, en open access... (cf: voir cours sur les bases de données)
- **Définition du choix des mots-clés** (// équation de recherche pour interroger les bases de données bibliographiques)



2018
2019
2020

Web of Science™

astrophysics
data system

OpenAlex

4. Méthodologie de réalisation d'une étude bibliométrique

4.2. Constitution du corpus de notices bibliographiques

- Interrogation de(s) base(s) de données bibliographiques (ou sources d'information)
- Collecte des données plus ou moins « structurées »
- Traitement des données pour préparer un corpus de notices bibliographiques (fichier de données)
- Envoi du corpus au commanditaire pour validation



4. Méthodologie de réalisation d'une étude bibliométrique

4.3. Traitement des données

- Opérations de dédoublonnage
- Opérations d'homogénéisation (ex : nom des revues différents, auteurs écrits sous plusieurs formes d'écriture...)
- Opérations d'enrichissement (ex : ajout des informations liées à l'Open access)



Objectif : rendre les données les plus propres possibles pour garantir leur qualité et améliorer leur utilisation dans l'analyse bibliométrique ou statistique.

Etape la plus longue et technique. Traitement des données à partir d'outils de curation de données comme OpenRefine, ou avec des traitements automatisés sous forme de scripts (sous Python par exemple).



Déf. Dédoublonnage : action de suppression des doublons ou données redondantes d'une liste, d'une base de données ou d'un fichier

Déf. Homogénéisation : action de nettoyage des données afin d'obtenir une valeur unique et cohérente pour des entrées de même type (ex : adresses, auteurs, mot-clés...)

Déf. Enrichissement : action d'ajouter des informations complémentaires à des données existantes à partir d'autres fichiers (ex : ajout du taux d'open access, de catégories thématiques spécifiques...)

4. Méthodologie de réalisation d'une étude bibliométrique (5)

4.4. Valorisation et visualisation des données sous forme d'indicateurs bibliométriques

- Utilisation d'outils de visualisation de données structurées comme VOSviewer ou Lodex (développé par le CNRS-INIST)
- Envoi du site web au commanditaire. L'étude peut être en mode public ou confidentiel (avec usage d'un identifiant / mot de passe).

➡ Création d'un outil web clé en main de type tableau de bord, qui permet l'exploration du corpus, et qui met en valeur les données sous forme d'indicateurs, à l'aide de graphiques.



5. Conclusions

Indicateurs **quantitatifs** pour la recherche scientifique

qui fait quoi où quand avec qui comment ?

**Productivité
scientifique**

**Contenu
thématique**

**Evolution
Tendances**

**Coopération
Relation scientifique**

Stratégie d'édition

Reconnaissance

Visibilité scientifique (citations)

Connaissance des sources

Données structurées, homogènes, enrichies

**Résultats accompagnés d'une méthodologie :
sources, modes de calcul, outils utilisés**

**Ne pas être utilisés seuls + environnement de la science ouverte,
multiplication des documents**

Pour aller plus loin

Calculs et représentations graphiques

Pour aller plus loin : le mode de comptage

Nombre de publications

- ✓ comptage **d'occurrence**
un acteur / une publication
→ Ne pas additionner
- ✓ comptage **fractionnaire**
un acteur / une part de publication
→ On peut additionner

Pour aller plus loin : le mode de comptage

Exemple : une notice bibliographique

Titre : Links between particle surface hardening and rehydration impairment ...

Source : FOOD HYDROCOLLOIDS 2016, Vol 61, P 277-285

5 Auteurs : Burgain, J (Burgain, Jennifer)^[1,2] ; Scher, J (Scher, Joel)^[2] ; Petit, J (Petit, Jeremy)^[2] ; Francius, G (Francius, Gregory)^[3] ; Gaiani, C (Gaiani, Claire)^[2,4]

4 Adresses :

- ▶ [1] **CNIEL**, 42 Rue Chateaudun, F-75314 Paris 9, **France**
- ▶ [2] **Univ Lorraine**, LIBio Lab Ingn Biomol, 2 Ave Foret de Haye, TSA 40602, F-54518 Vandoeuvre Les Nancy, **France**
- ▶ [3] **Univ Lorraine**, LCPME, UMR 7564, F-54600 Villers Les Nancy, **France**
- ▶ [4] **Univ Queensland**, Sch Agr & Food Sci, St Lucia, Qld 4072, **Australia**

Pour aller plus loin : le mode de comptage

5 coauteurs

	Occurrence	Fractionnaire
Burgain, Jennifer	1	0,2
Scher, Joel	1	0,2
Petit, Jeremy	1	0,2
Francius, Gregory	1	0,2
Gaiani, Claire	1	0,2
Total	5	1

Pour aller plus loin : le mode de comptage

3 institutions - 2 pays

		Occurrence	Fractionnaire
Organismes	Univ Lorraine	1	0,33
	Univ Queensland	1	0,33
	CNIEL	1	0,33
	Total	3	1
Pays	France	1	0,5
	Australie	1	0,5
	Total	2	1

Pour aller plus loin : le mode de comptage

Part de publications / Pourcentage pour rapporter,
pour avoir un élément de **comparaison**

Type de document	Nb de publications	Part
ARTICLES	94 223	75,8%
REVIEW ARTICLES	9 103	7,3%
MEETING ABSTRACTS	6 670	5,4%
EARLY ACCES	6 131	4,9%
Autres	15 692	0,17%
Total	131 819 (124 316)	105,5% (100)

Pays coopérant	Nb de copublications avec	Part
USA	21 730	32,7%
GERMANY	14 975	22,58%
ENGLAND	14 729	22,21%
ITALY	13 247	19,97%
...		
Total	(66 316)	(100)

Une publication peut
avoir plusieurs types
de documents

cu=France and py=2021 (WoS)
124 316 publications

Une publication peut
avoir plusieurs pays
coopérants

Pour aller plus loin : les représentations graphiques

Tableaux

Radars

Bâtons

Courbes

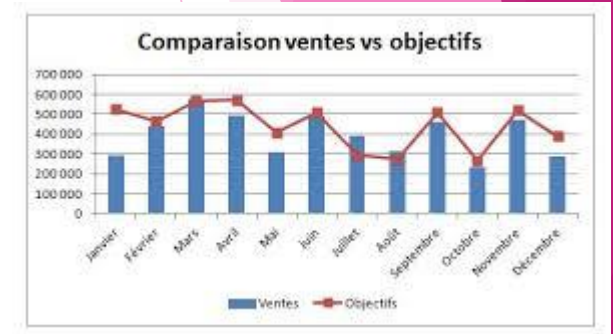
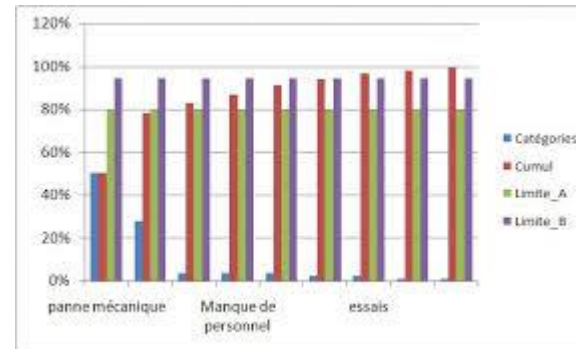
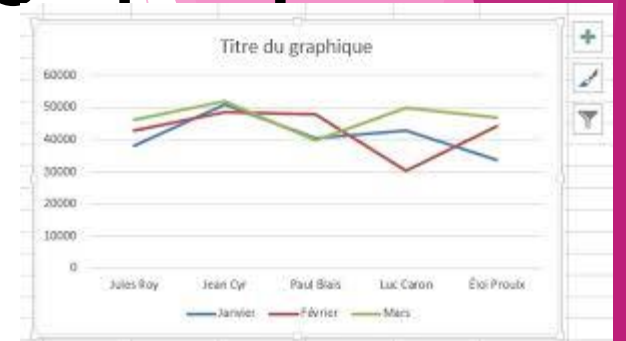
Bulles

Cartographie

Réseaux

....

Excel



Gephi

