

Traitement de données issues d'un fichier CSV à l'aide d'un tableur

O. CHAUMETTE
à partir d'une idée de **P. LUCAS-FRADIN**
Lycée Jean-Paul Sartre – 69 BRON

Informations générales concernant la séance

*Nous allons travailler avec un tableur pour traiter des **données ouvertes*** portant sur les prénoms donnés en France depuis 1900*

Il est conseillé d'afficher ce document à droite de l'écran et les autres logiciels (Excel, Firefox) à gauche. Rappel: touche  +  ou 

Les instructions pour manipuler les logiciels sont écrites dans des cadres bleus

Vous répondrez, sur votre cahier, aux questions des cadres marron



A la fin de la séance, vous vous connecterez sur Pronote pour pouvoir répondre à un QCM. Les réponses au QCM se trouvent parmi celles que vous aurez écrites sur votre cahier.

* **Données ouvertes** (ou « **open datas** »): données dont l'accès et l'usage sont accessibles à tout le monde

Recherche sur le site data.gouv.fr

Ouvrez un navigateur web (icône « Qwant ») et rendez-vous sur la page Web des données ouvertes du gouvernement français:

<https://www.data.gouv.fr/fr/>



Télécharger le fichier CSV de données

Un fichier de données nationales qui contient les prénoms attribués aux enfants nés en France hors Mayotte entre 1900 et 2018 et les effectifs par sexe associés à chaque prénom ;
Un fichier de données départementales qui contient les mêmes données par département de naissance.

Ressources

Fichier principal

Fichier France hors Mayotte

Fichier de données nationales qui contient les prénoms attribués aux enfants nés en France hors Mayotte entre 1900 et 2018 et les effectifs par sexe associés à chaque prénom

csv 160 Disponible

TÉLÉCHARGER

Fichier France hors Mayotte

Fichier de données nationales qui contient les prénoms attribués aux enfants nés en France hors Mayotte entre 1900 et 2018 et les effectifs par sexe associés à chaque prénom

dbase 1 Disponible

TÉLÉCHARGER

Fichiers par départements de naissance

Fichier d'ouverture de nat2018_csv.zip au niveau département de naissance.

Vous avez choisi d'ouvrir :
nat2018_csv.zip
qui est un fichier de type : Compressed (zipped) Folder
à partir de : https://www.insee.fr

Que doit faire Firefox avec ce fichier ?

- Ouvrir avec Explorateur Windows (par défaut)
- Enregistrer le fichier
- Toujours effectuer cette action pour ce type de fichier.

OK

Annuler

Télécharger le fichier CSV

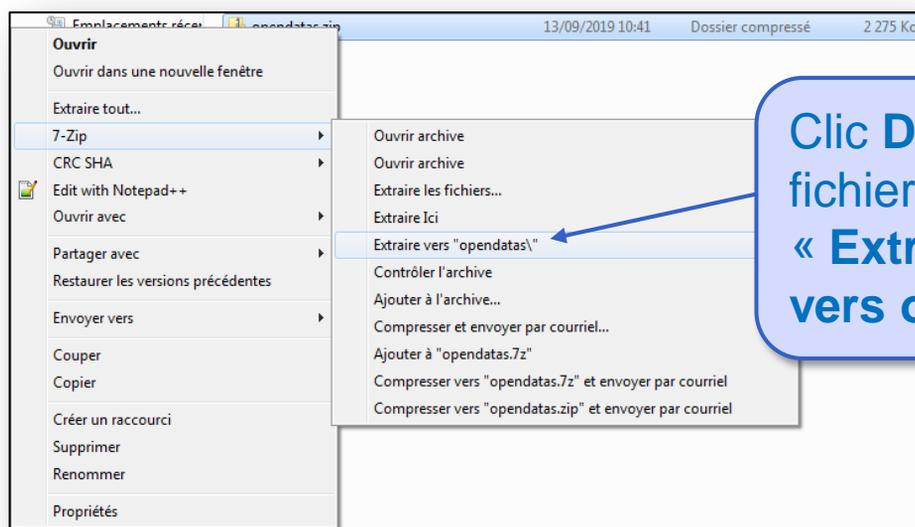
Sauvegarder le fichier « nat2018_csv.zip » dans le dossier SNT de votre dossier personnel

Renommer et dézipper le fichier

Ouvrez l'explorateur de fichier et allez dans le dossier SNT de votre dossier personnel

Renommer le fichier `nat2018_csv.zip` en « `opendatas.zip` »

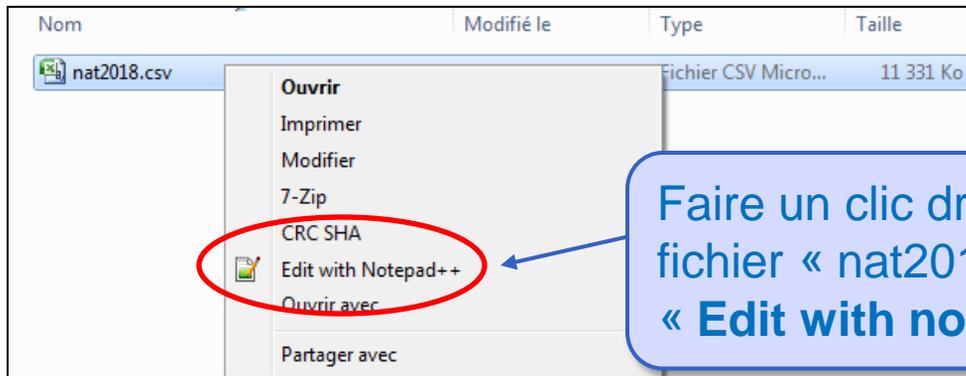
Extraire le fichier CSV de "opendatas.zip":



Clic DROIT sur le fichier puis 7-zip puis « Extraire vers opendatas »

Aller dans le dossier « opendatas ». Le fichier de données « `nat2018.csv` » doit s'y trouver.

Observation du fichier CSV



```

1 sexe;preusuel;annais;nombre
2 1;A;1980;3
3 1;A;1998;3
4 1;A;XXXX;21
5 1;AADAM;2009;4
6 1;AADAM;2014;3
7 1;AADAM;2016;4
8 1;AADAM;2017;4
9 1;AADAM;2018;3
10 1;AADAM;XXXX;9
11 1;AADEL;1976;5
12 1;AADEL;1978;3
  
```

Le contenu brut du fichier CSV apparaît

1. A quoi correspond un objet dans ce fichier ?
2. Citer les 2 premiers descripteurs
3. Par quoi sont séparées les données ?
4. Quel est le type des données du 3nd descripteur ?



Ouverture du fichier CSV avec Excel

**Fermer notepad++ et double cliquer sur « nat2018.csv » :
Le tableur « Excel » s'ouvre en affichant les données.**

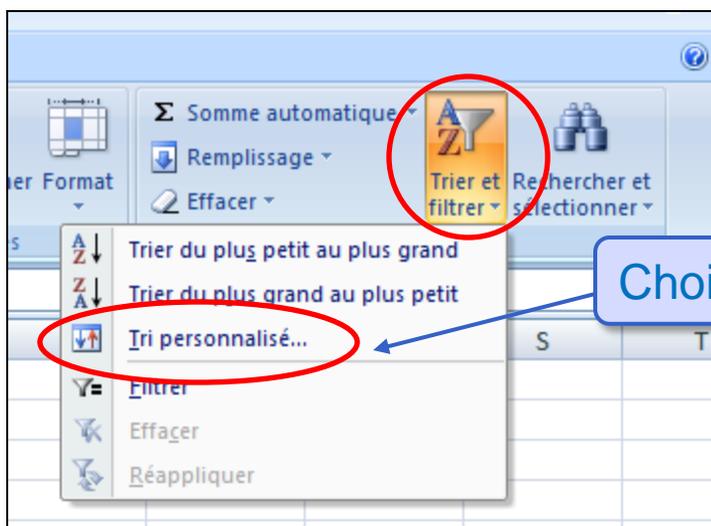
	A	B	C	D
1	sexe	preusuel	annais	nombre
2	1	A	1980	3
3	1	A	1998	3
4	1	A	XXXX	
5	1	AADAM	2009	
6	1	AADAM	2014	
7	1	AADAM	2016	
8	1	AADAM	2017	
9	1	AADAM	2018	
10	1	AADAM	XXXX	

Certaines données n'ont pas de sens. Nous allons donc les supprimer. Pour cela, il faut d'abord trier les données.

Sélectionner toutes les cellules avec la combinaison de touches:

Ctrl

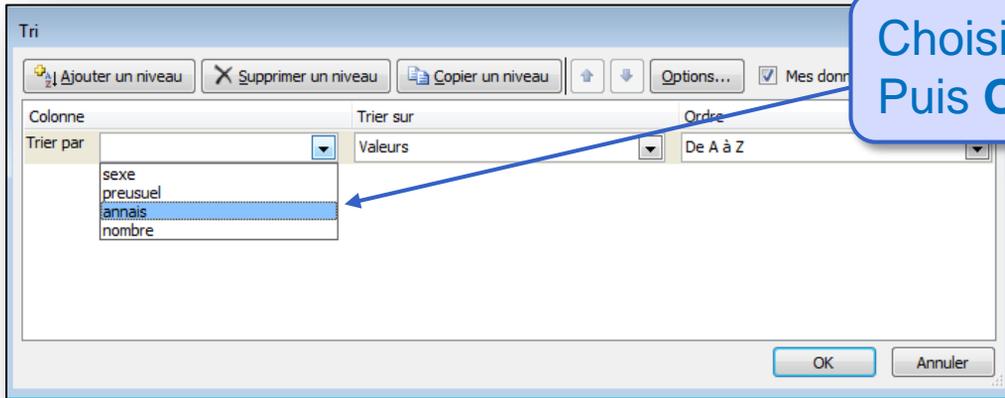
A



Choisir « Tri personnalisé »

Suppression des données inutiles

MÉTHODE "TRIER" à CONNAÎTRE



Choisir « annais »
Puis OK

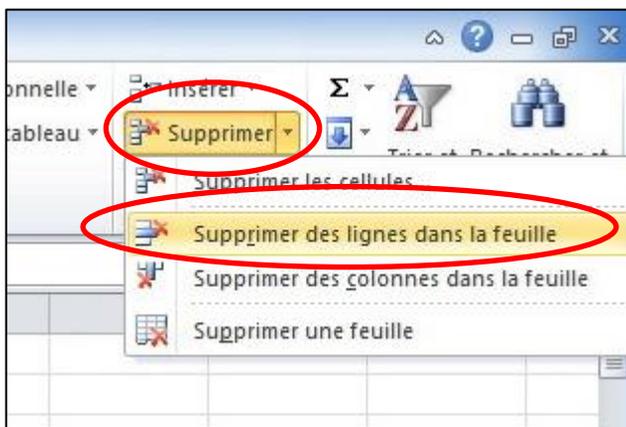
Le
descripteur
"annais"
représente
l'année de
naissance

Cela va permettre de "trier" (donc classer" les objets par année)



601460	2	ÉVY	2018	32
601461	2	ÉYA	2018	3
601462	2	ĪNAYA	2018	5
601463	1	A	XXXX	27
601464	1	AADAM	XXXX	9
601465	1	AADEL	XXXX	27
601466	1	AADIL	XXXX	31
601467	1	AAHIL	XXXX	5

Sélectionner à la
souris toutes les
cellules dont
l'année de
naissance est
XXXX (il y en a
beaucoup !!)



Cliquer sur "Supprimer"
puis "Supprimer des
lignes dans la feuille"

Filtrer des données

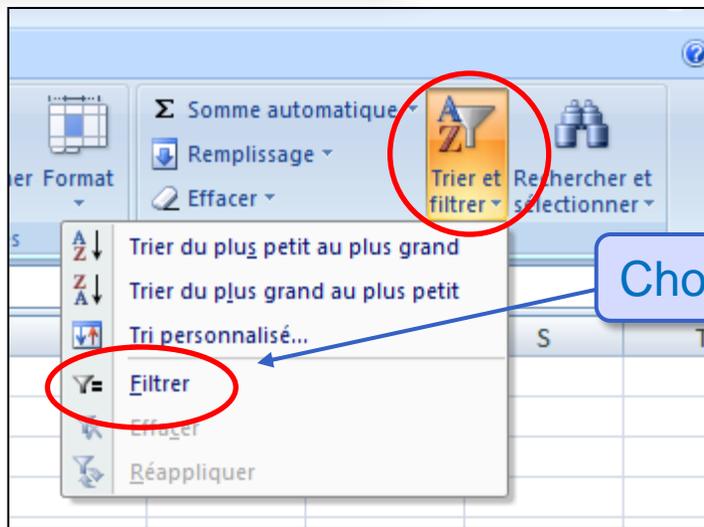
MÉTHODE à CONNAÎTRE



Nous allons afficher que certaines données.

Avant de filtrer, il faut sélectionner un descripteur ("annais" par exemple)

	A	B	C	D
1	sexe	preusuel	annais	nombre
2		1 A	1980	3
3		1 A	1998	3



	A	B	C	D
1	sexe	preusuel	annais	nombre
2		1 A	1980	3
3		1 A	1998	3

Cliquer sur le triangle du descripteur pour faire apparaître la fenêtre permettant de filtrer des données

Filtrer des données

MÉTHODE à CONNAÎTRE



The screenshot shows the Excel interface with the 'Filtres numériques' menu open. The 'Est égal à...' option is highlighted with a red circle. The background shows a table with columns 'sexe', 'preusuel', 'annais', and 'nombre'.

Choisir "**Filtres numériques**" puis "**est égal à...**"

Remarque:
d'autres opérateurs de filtrage sont disponibles

Entrez l'année à filtrer ("2004" par exemple) puis **OK**

The dialog box 'Filtre automatique personnalisé' is shown. The 'Afficher les lignes dans lesquelles : annais' section has 'est égal à' selected in the dropdown and '2004' entered in the text box. The 'Et' radio button is selected.

	A	B	C	D
1	sexe	preusuel	annais	nombre
92		1 AARON	2004	335
203		1 ABASS	2004	4
307		1 ABBES	2004	4
		1 ABD	2004	16
		1 ABD-ALLAH	2004	7
		1 ABD EL RAU	2004	2

Seules les naissances de 2004 s'affichent

Additionner des données triées

Nous allons calculer le nombre de prénoms déclarés en 2004

Noter sur votre cahier le n° de la 1^{ère} ligne contenant une donnée (ici **409398**)

	A	B	C	D
1	se	preusuel	anna	nombre
409398	1	AARON	2004	335
409399	1	ABASS	2004	4
409400	1	ABBES	2004	4
409401	1	ABD	2004	16

Descendre au niveau de la dernière ligne et noter sur votre cahier son n° (ici: **420364**)

420361	2	ZUZANNA	2004	3
420362	2	ZYA	2004	4
420363	2	ZYNEB	2004	4
420364	2	_PRENOMS_	2004	18108
601463				
601464				

Aller dans la cellule correspondante à la fréquence des prénoms juste après la dernière ligne

Taper:

=SOUS.TOTAL(9;D409398:D420364)

"409398" et "420364" sont les n° des 1^{ère} et dernière lignes, "D" la colonne

420363	2	ZYNEB	2004	4
420364	2	_PRENOMS_	2004	18108
601463				787870

Appuyez sur la touche **Entrée**
La somme se calcule

EXERCICE 1

5. Combien d'enfants sont nées en France en 2004 ?

**EXERCICE 2**

Utiliser les méthodes vues précédemment pour répondre aux questions suivantes

6. Combien d'enfants sont nées en France en 1970 ?

**EXERCICE 3**

Utiliser en même temps "Filtrer" et "Trier"

7. Quel est le prénom le plus donné en 2004 ?

PRENOM regroupe les prénoms rares.
Il n'est pas à prendre en compte.

8. Quel est le prénom féminin le plus donné en 2018 ?



Pour information seulement:

un prénom est considéré comme « rare » s'il a été donné moins de 20 fois entre 1900 et 1945 ou bien entre 1945 et 2018

Préparation des données pour tracer un graphe

Filter:

- *Les garçons*
- *Le prénom "Pierre"*

Trier (c'est-à-dire classer):

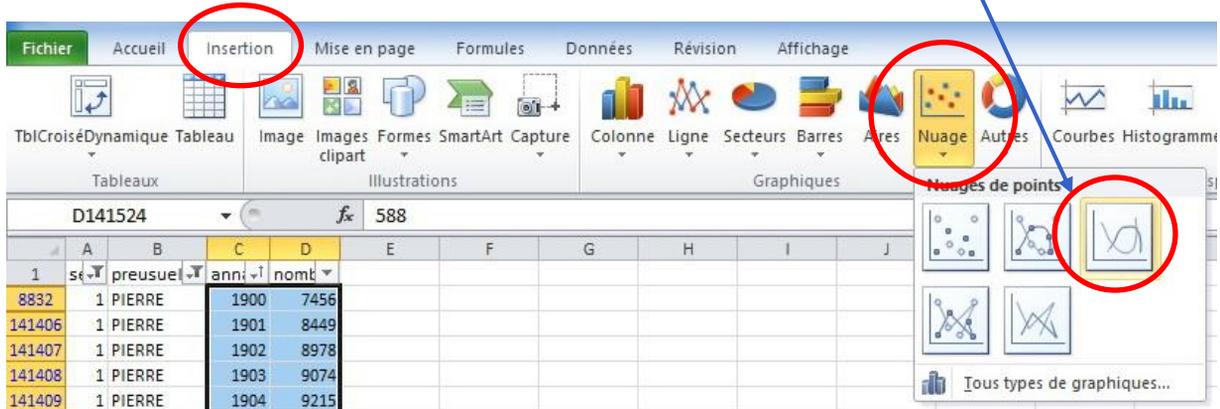
- *Les années de naissance de la plus petite à la plus grande*

	A	B	C	D
1	se	preusuel	ann	nomt
8832	1	PIERRE	1900	7456
141406	1	PIERRE	1901	8449
141407	1	PIERRE	1902	8978
141408	1	PIERRE	1903	9074
141409	1	PIERRE	1904	9215
141410	1	PIERRE	1905	9429
141411	1	PIERRE	1906	9855
141412	1	PIERRE	1907	10074
141413	1	PIERRE	1908	10769
141414	1	PIERRE	1909	10893
141415	1	PIERRE	1910	11156
141416	1	PIERRE	1911	10986
141417	1	PIERRE	1912	11910
141418	1	PIERRE	1913	12208
141419	1	PIERRE	1914	12295
141420	1	PIERRE	1915	7767
141421	1	PIERRE	1916	6060
141422	1	PIERRE	1917	6746

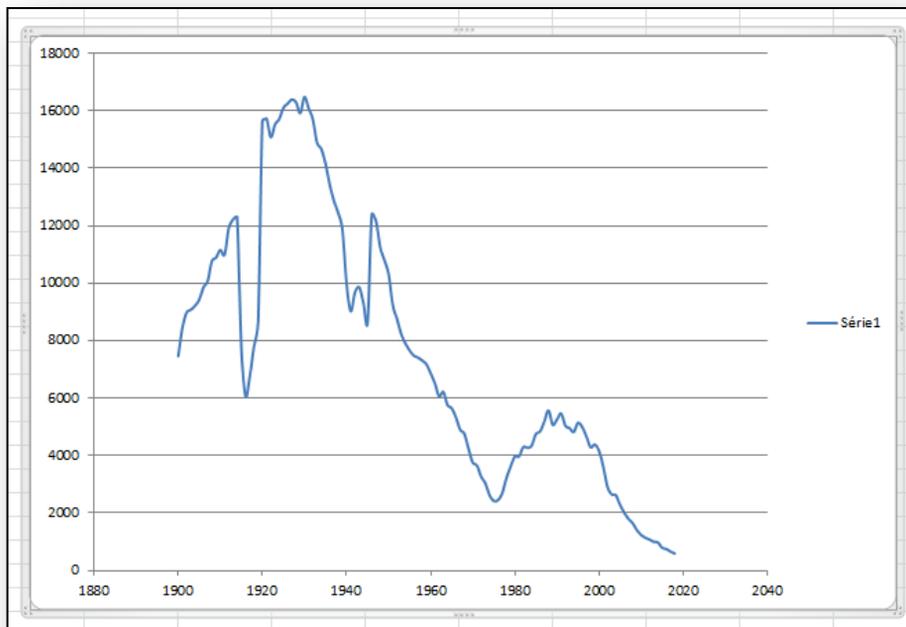
Sélectionner ensuite toutes les cellules "annais" et "nombre"

Tracé du graphe

Onglet "Insertion"
Puis "Nuage"
Puis le tracé continu



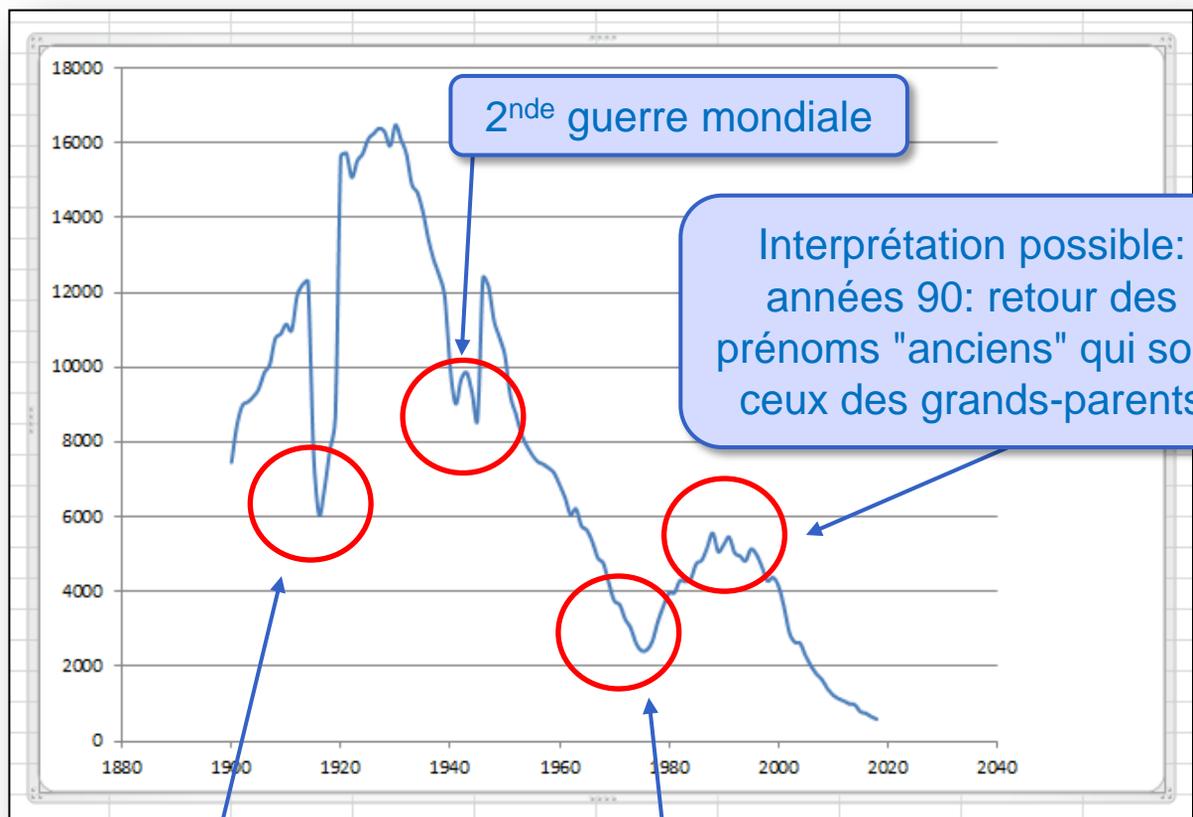
On obtient le graphe représentant la fréquence du prénom "Pierre" en fonction de l'année de naissance:



Analyse du graphe du graphe

Voici un exemple d'analyse du graphe obtenu

Globalement, le prénom Pierre était très utilisé avant le 2nde guerre mondiale puis en nette baisse après. Cela peut-être l'effet de le déchristianisation (Pierre étant un prénom important de la religion catholique).



Très nette baisse:
1^{ère} guerre mondiale:
peu de naissances

Interprétation possible: années 70:
rejet des prénoms "anciens" qui sont
ceux des parents suite à mai 68

EXERCICE 4

Utiliser les méthodes vues précédemment pour tracer la courbe de fréquence en fonction de l'année pour le prénom: MARIE (filtrer: femmes uniquement)

15. La courbe présente-t-elle les mêmes caractéristiques que celles de Pierre ? Pourquoi ?

**EXERCICE 5**

Utiliser les méthodes vues précédemment pour tracer la courbe de fréquence en fonction de l'année pour le prénom: BRIGITTE

17. Pourquoi a-t-on un tel pic vers 1960 ? Faire une recherche internet si vous n'avez pas d'idée...



EXERCICE 6*Tracer la courbe pour votre prénom*

16. Comment évolue la courbe de fréquence pour votre prénom ? Voyez-vous les caractéristiques sur cette courbe qui peuvent être analysées (guerres, film célèbre, acteur/actrice célèbre...)

**EXERCICE 7***Tracer la courbe pour ZINEDINE*

18. Expliquer l'évolution de la courbe

**EXERCICE 8***Tracer la courbe pour ADOLPHE*

19. Expliquer l'évolution de la courbe



POUR LES PLUS RAPIDES**EXERCICE 9**

*Tracer la courbe pour JOSÉ
(prénom d'origine portugaise)*

20. Pourquoi a-t-on une forte augmentation à la fin de la 2nde guerre mondiale?

**EXERCICE 10**

*Tracer la courbe pour SAMIR
(prénom d'origine maghrébine)*

21. Expliquer l'évolution de la courbe

**EXERCICE 11**

Tracer la courbe pour KYLIAN

21. Expliquer l'évolution de la courbe.
Peut-on prévoir son évolution ?



EXERCICE 12

Utiliser en même temps "Filtrer" et "Trier"

12. Quel est le prénom le plus donné en 1970 ?
13. Est-il aussi fréquent en 1980 ?
14. Même question en 2004

**EXERCICE 13**

8. Combien de garçons sont nés en 1980 ?
(indication: il est possible de filtrer deux descripteurs)
9. Combien de filles sont nées entre 2000 et 2004 (inclus)?
(indication: regarder les opérateurs de "filtre numérique")