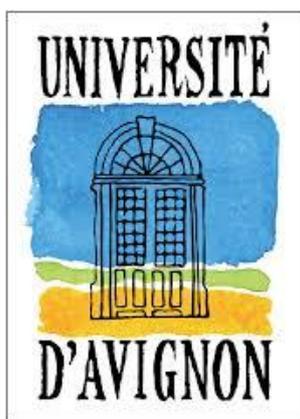




Mairie d'Avignon



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

Analyse statistique des prénoms sur la commune d'Avignon pour les tranches d'années de 1989 à 2017

Travail réalisé par Emile COUTIER au sein de la mairie d'Avignon à l'occasion
d'un stage, dirigé par Monsieur François GANZ.

coutieremile@gmail.com



Le 06/07/18

Table des matières :

I- Introduction	1
II- Les prénoms masculins et féminins les plus donnés par tranche d'années	2
a) Tranche 1990-1999	3
b) Tranche 2000-2009	4
c) Tranche 2010-2017	5
III- Régression linéaire simple sur le prénom le plus donné pour chacune des tranches	6

I- Introduction

On étudiera donc la commune d'Avignon qui regroupe les registres d'état-civil de Montfavet et Avignon. Ville de Provence dans le sud-est de la France, qui est traversée par le Rhône. Aussi préfecture du département du Vaucluse, Avignon occupe une superficie de 6478 hectares pour une population de 91 451 habitants. A la fois capitale spirituelle, capitale politique, capitale économique et capitale culturelle, Avignon se prévaut d'un patrimoine architectural et artistique exceptionnel qui en fait aujourd'hui encore le plus grand et le plus bel ensemble gothique d'Europe.

Notre fichier d'entrée est l'extraction des prénoms de 1989 à 2017 faite par le service de l'Etat Civil et de la collectivité, et fourni sous forme de fichier XLS. Nous allons devoir normaliser notre fichier selon les spécifications d'open data locale, lien vers celui-ci : <http://opendatalocale.net/> ainsi que le lien vers les spécifications des prénoms https://docs.google.com/document/d/1Vv0kpBw3MIocai9JgovLK2HxcUA_3QHnZicqxuOpcQ8/edit.

Une fois cette normalisation effectuée nous allons pouvoir publier notre fichier sur la plateforme DataSud, lien vers l'emplacement du fichier : <https://trouver.datasud.fr/dataset/avignon-prenoms-cumulatif>. Ces données seront en open data et il faut savoir que l'open data sont des données auxquelles l'accès est totalement public et libre de droit, au même titre que l'exploitation et la réutilisation. Ces données offrent de nombreuses opportunités pour étendre le savoir humain et créer de nouveaux produits et services de qualité.

Dans notre fichier, nous avons donc 3 variables. La première est la variable PRENOMS où il y a les prénoms des nouveaux nés sur notre commune. Pour les valeurs de cette variable nous pouvons avoir qu'un seul prénom mais aussi plusieurs si l'individu possède un deuxième ou/et un troisième prénom. Comme séparateur de champs entre chaque prénom il peut y avoir une virgule, un espace, une virgule suivie d'un espace, une double virgule... c'est pour dire que le séparateur de champs n'est pas le même pour chaque ligne, et c'est un paramètre qu'il a fallu prendre en compte lors de la normalisation. A noter que depuis le 28/10/2011, une circulaire s'intitulant le bulletin officiel du ministère de la justice et des libertés, à la page 36 a précisé que le caractère "espace" n'est plus un séparateur de champ mais un caractère faisant partie du prénom. Ainsi "Jean Marie" est un seul prénom et non deux. Lien vers cette circulaire : www.textes.justice.gouv.fr/art_pix/JUSC1119808C.pdf#page=36. Notre deuxième variable est le SEXE qui est défini par : Masculin, Féminin ou Indéterminé. Et notre dernière variable est l'ANNEE DE NAISSANCE qui est comprise entre 1989 et 2017.

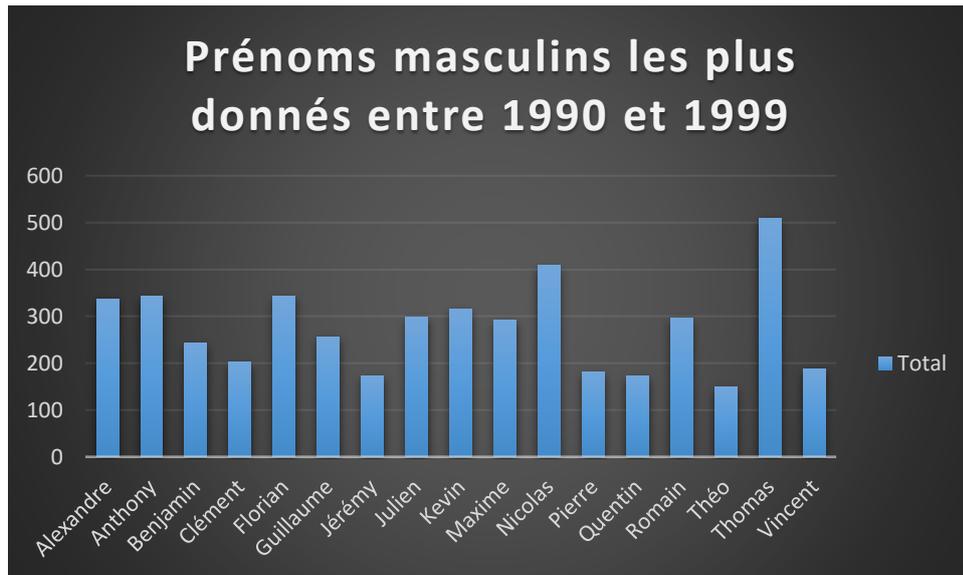
Ci-dessous un extrait de la spécification d'open data locale qui nous permet de voir à quoi ressemble notre fichier normalisé :

Champ	Objet
COMMUNE_NOM	Nom officiel de la commune
COLL_INSEE	Code INSEE de la commune
ENFANT_SEXE	Sexe correspondant au prénom : M ou F ou I, respectivement pour masculin, féminin ou intersexué/indéterminé.
ENFANT_PRENOM	Prénom de nouveau(x)-né(s) constaté comme premier prénom dans les actes d'état-civil de l'année correspondante
NOMBRE_OCCURRENCES	Nombre d'occurrences du prénom pour l'année correspondante.
ANNEE	Année de relevé, sur quatre chiffres.

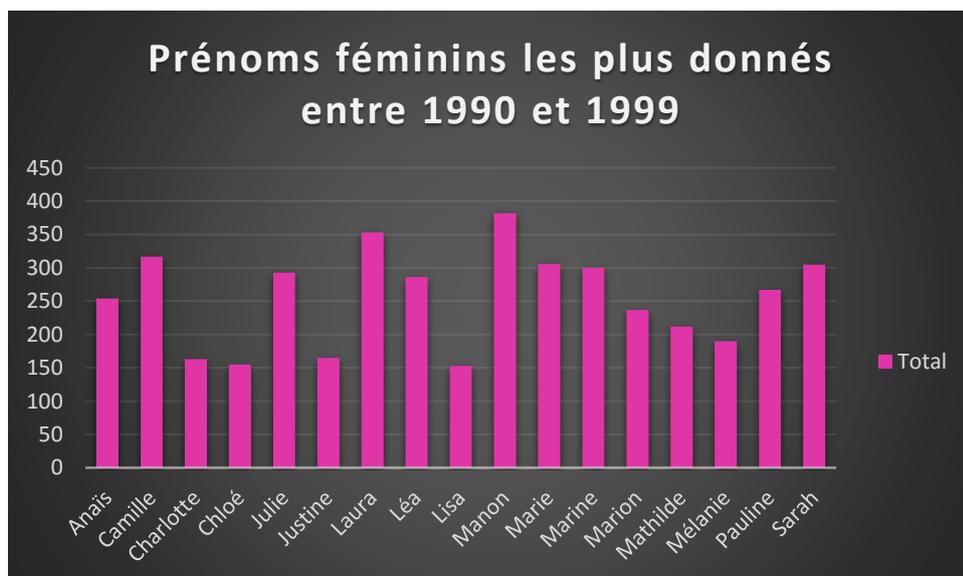
De plus si l'on souhaite vérifier si notre fichier est bien conforme aux spécifications d'open data locale on peut le tester sur un site de vérification s'intitulant validata : <https://dev.validata.fr/>

II- Les prénoms masculins et féminins les plus donnés par tranche d'années

a) Tranche 1990-1999

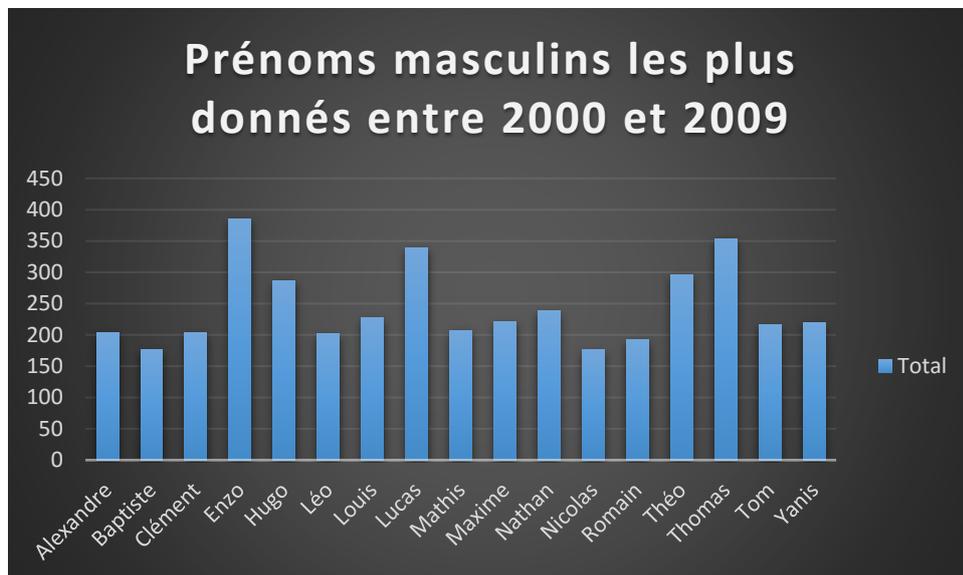


- 1 : Thomas
- 2 : Nicolas
- 3 : Florian

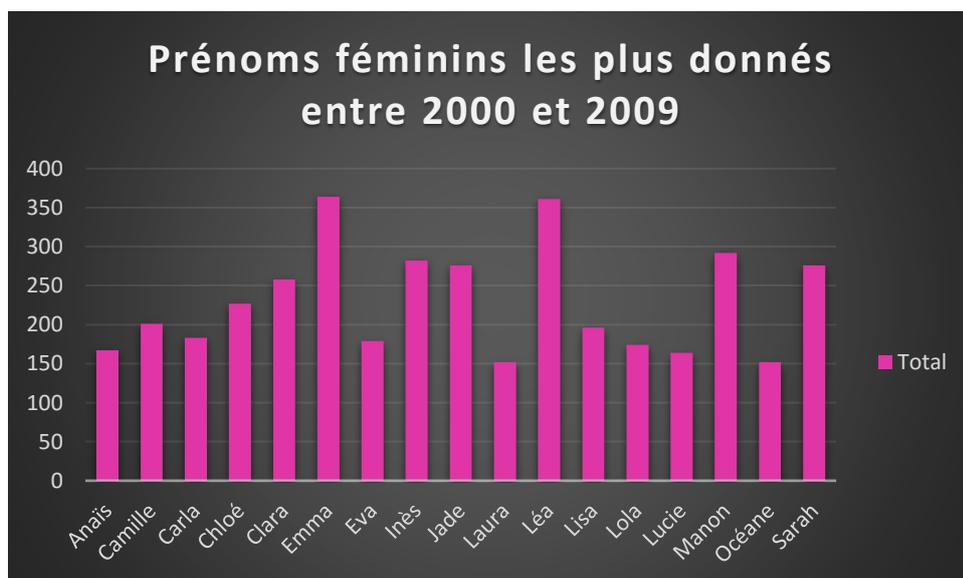


- 1 : Manon
- 2 : Laura
- 3 : Camille

b) Tranche 2000-2009

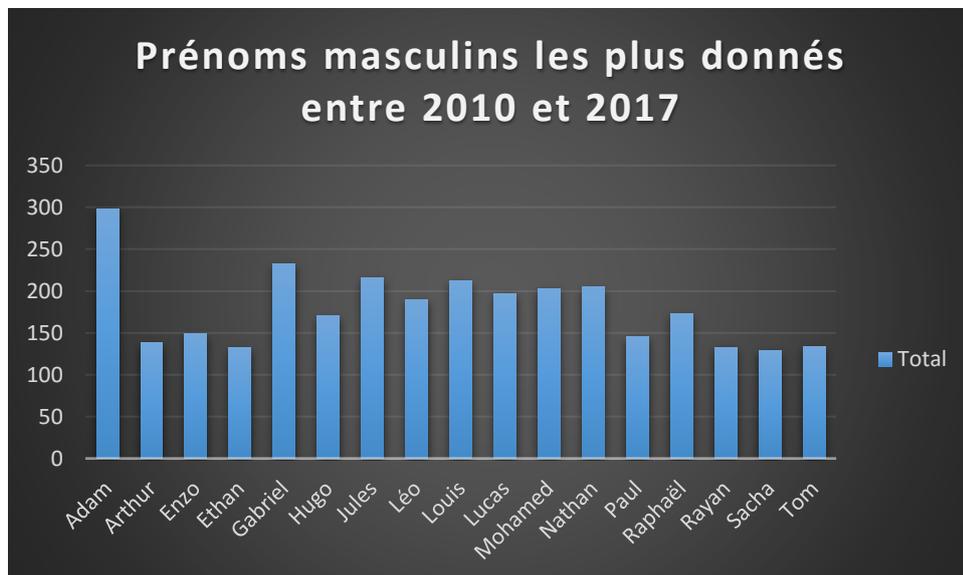


- 1 : Enzo
- 2 : Thomas
- 3 : Lucas

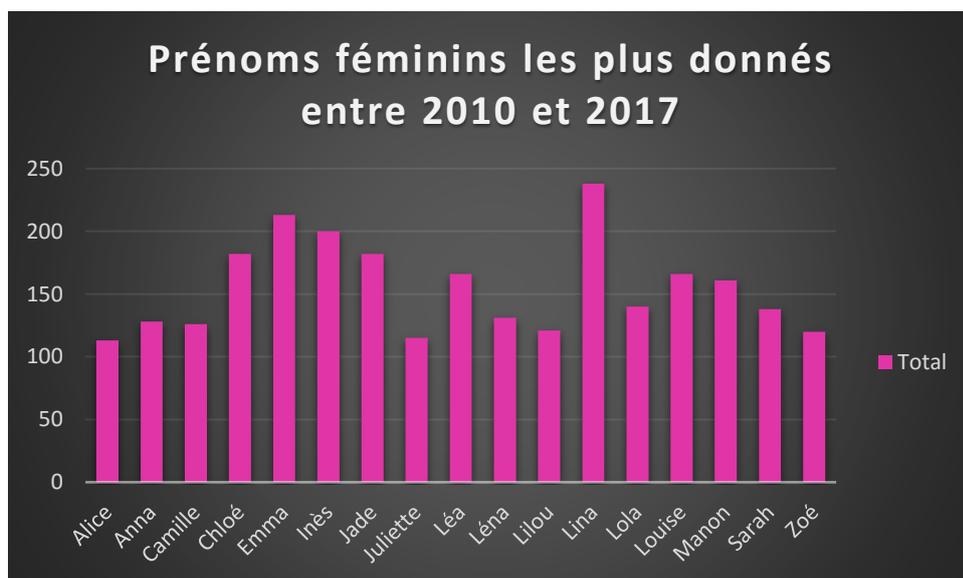


- 1 : Emma
- 2 : Léa
- 3 : Manon

c) Tranche 2010-2017



- 1 : Adam
- 2 : Gabriel
- 3 : Jules



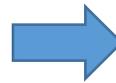
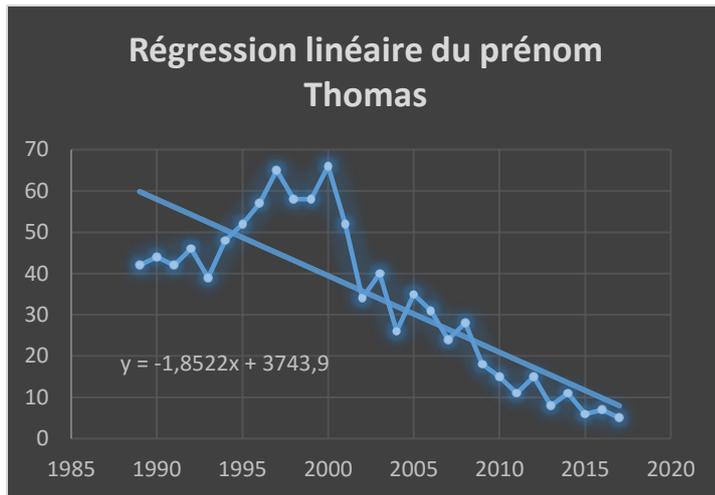
- 1 : Lina
- 2 : Emma
- 3 : Inès

Pour les prénoms masculins, on remarque au niveau des années 1990→2009 que le prénom Thomas fut un prénom très populaire, car il arrive dans le top des prénoms pour les deux tranches alors que pour Nicolas et Kévin, on peut voir qu'ils ne ressortent pas pour la période 2000 → 2017. On en déduit que le prénom Thomas a été plus résistant et plus à la mode sur cette période. On peut remarquer que sur la dernière tranche que de nouveaux prénoms ont commencé à apparaître comme Adam, Gabriel ou Jules et ces prénoms n'étaient pas à la mode sur la tranche 1990 → 2009.

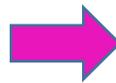
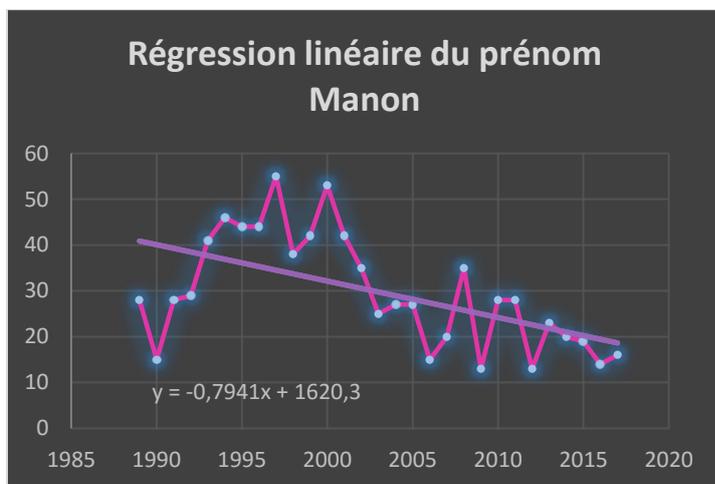
Pour les prénoms féminins, on remarque que le prénom Manon a été populaire sur les 3 périodes (1990 → 2017), mais une part bien plus effective sur la première et la deuxième période. Pour le prénom Camille même observation, mais un peu moins populaire que le prénom Manon. D'autre part, on remarque que le prénom Lina apparaît sur la dernière période alors qu'il n'était pas présent sur les deux premières tranches.

III-Régression linéaire simple sur les prénoms les plus donnés

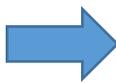
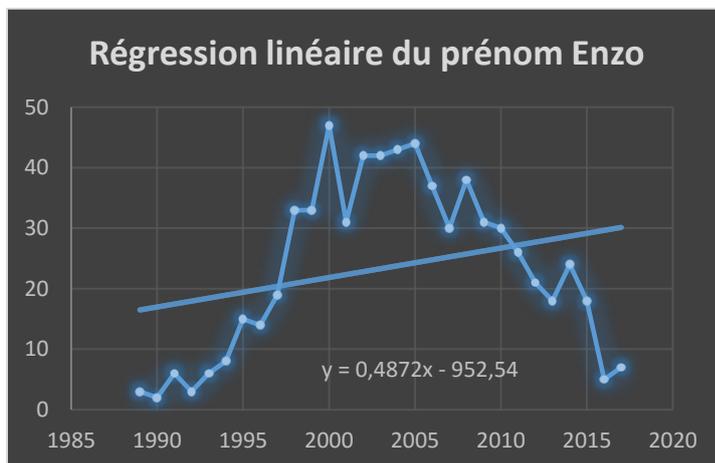
Ci-dessous, nous allons effectuer des régressions linéaires sur six prénoms différents. Cette opération va permettre de voir les tendances de ces prénoms, à quelles périodes, ils étaient populaires ou à contrario en chute. De plus grâce à l'équation des droites de régression nous allons avoir une prévision et savoir quand ce prénom disparaîtra et ne sera plus donné. Par exemple pour notre premier prénom Thomas, nous avons l'équation suivante : $y = -1,8522x + 3743,9$, il suffit de remplacer y par 0 et de calculer en fonction de x. Ce qui nous donne $x = 3743,9 / 1,8522 = 2021$, on peut donc estimer que le prénom Thomas aura disparu en 2021.



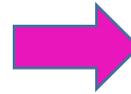
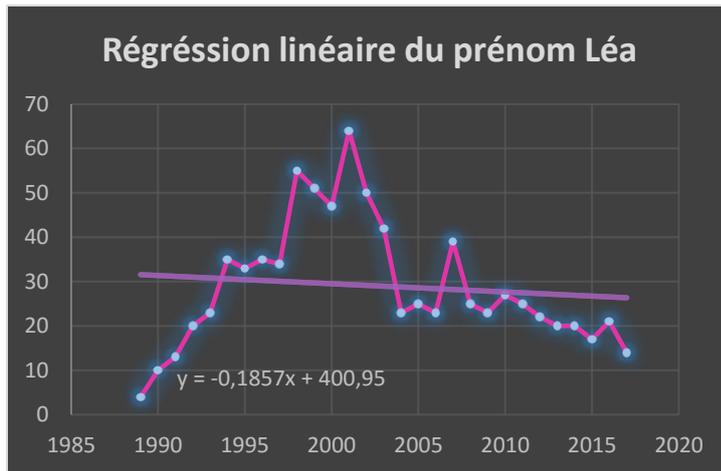
-Nous pouvons voir que le prénom Thomas était beaucoup choisi avant les années 90 et a eu un pic dans les années 2000. Mais ensuite il a subi une décroissance au fil du temps et on remarque grâce à la droite de régression que le prénom est devenu beaucoup moins à la mode.



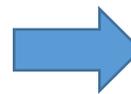
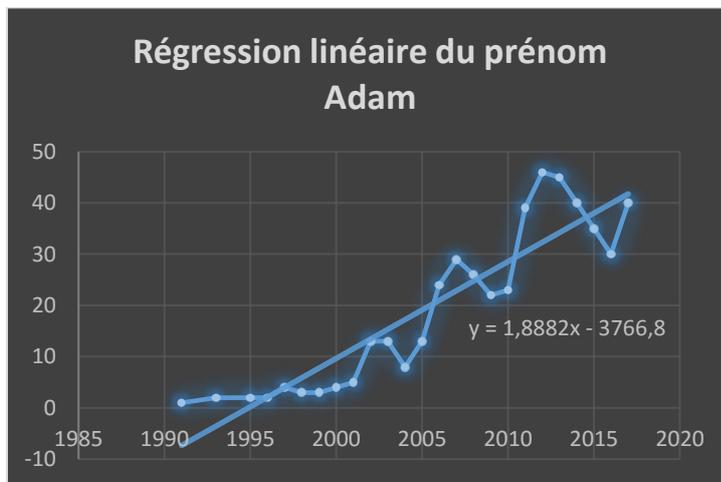
-Pour le prénom Manon on peut voir qu'il est resté assez populaire tout le long de la période et a subi un pic pour les années 97 et 2000.



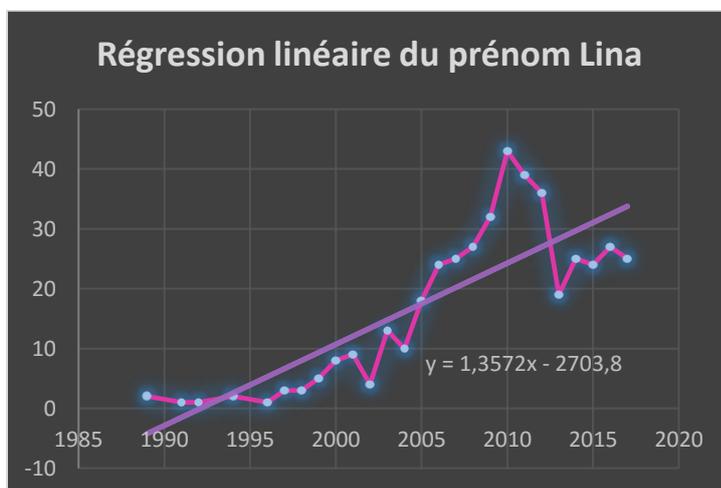
-On remarque qu'au début des années 90 le prénom Enzo n'était pas trop connu de la population et ce n'est qu'à la fin des années 90 que le prénom est devenu à la mode, jusqu'à l'année 2008. Ensuite le prénom a subi une forte baisse de popularité. La courbe s'apparente à une loi normale ou Gaussienne.



-Pour le prénom Léa, nous avons les mêmes caractéristiques que le prénom Enzo, c'est-à-dire que le prénom commence à se faire connaître au début des années 90. Puis il a eu un pic en 2001 et après a une baisse de popularité mais a fini par stagner à 20 occurrences par an.



-Des années 90 jusqu'à l'année 2000 le prénom Adam n'était pas du tout connu puis il a subi une forte croissance au fil du temps et de nos jours il est devenu un prénom très apprécié. On le remarque par sa droite de régression croissante.



Pour le prénom Lina on remarque qu'il est devenu populaire à partir de 2003 et a subi un fort pic en 2010 puis a légèrement diminué pour enfin stagner et reste un prénom apprécié du grand public.