Fichier des prénoms - Édition 2016

lil-1096

La source

Le fichier des prénoms est établi à partir du répertoire de l'état civil des bulletins de naissances des personnes nées en France et dans les départements d'Outre-Mer (DOM).

L'exhaustivité n'est pas garantie sur toute la période, notamment pour les années antérieures à 1946. Des écarts peuvent donc être constatés avec le nombre annuel des naissances évalué par l'INSEE. Ces écarts, importants en début de période, vont en s'amenuisant. Après 1946, ils sont peu significatifs.

Le jeu de données contient deux fichiers

Un fichier national (NAT2015)
Sans localisation géographique de naissance,
Couvrant toutes les années de naissance de 1900 à 2015
Comporte 15 067 prénoms attribués à des personnes de sexe masculin
et 17 969 prénoms attribués à des personnes de sexe féminin.

Un fichier départemental (DPT2015)
Distinguant les départements de naissance,
Couvrant (comme depuis 2014) les années de naissance de 1900 à 2015.
Comporte 15 067 prénoms attribués à des personnes de sexe masculin
et 17 969 prénoms attribués à des personnes de sexe féminin.

Les champs retenus

Les personnes
L'ensemble des personnes nées en France (Métropole et départements d'Outre-mer)
Et enregistrées à l'état civil sur les bulletins de naissance.
Les personnes nées à l'étranger sont donc exclues.

Les prénoms
Ils sont issus des bulletins de naissance de l'état civil
Les différents prénoms sont séparés par un espace (ou blanc).
2 prénoms séparés par un tiret constituent un seul prénom composé (par ex. : Anne-Laure).
Le 1er prénom, simple ou composé, qui figure en début de liste, est retenu après traitement pour la protection de l'anonymat.

La protection de l'anonymat

Dans le fichier national
2 conditions doivent être respectées :

1) D'abord pour masquer les prénoms les plus rares
   - Pour un sexe donné, on définit les prénoms les plus rares comme étant ceux qui ont moins de 20 occurrences.
   - Leur intitulé est masqué par le texte «_PRENOMS_RARES_».
   - Les occurrences de l'ensemble des prénoms rares sont sommées par sexes et années de naissances.

2) Ensuite pour masquer les années les plus rares
   - Cette condition ne s'impose qu'aux prénoms qui ne sont pas rares.
   - Pour un prénom et un sexe donnés, on définit les années les plus rares comme étant celles qui ont moins de 3 occurrences.
   - Leur intitulé est masqué par le texte «XXXX».
   - Les occurrences de l'ensemble des années rares sont sommées par prénoms et sexes.

Dans le fichier départemental

2 conditions doivent être respectées :

1) D'abord pour masquer les prénoms les plus rares
   - Pour un sexe donné, on définit les prénoms les plus rares comme étant ceux qui ont moins de 20 occurrences sur l'ensemble de la période.
   - Leur intitulé est masqué par le texte «_PRENOMS_RARES_».
   - Les occurrences de l'ensemble des prénoms rares sont sommées par sexes et années de naissance

2) Ensuite pour masquer les combinaisons années/départements les plus rares
   - Cette condition ne s'impose qu'aux prénoms qui ne sont pas rares.
   - Pour un prénom et un sexe donnés, on définit les combinaisons années/départements les plus rares comme étant celles qui ont moins de 3 occurrences.
   - Leurs intitulés sont masqué par le texte «XXXX» pour l'année et le texte «XX» pour le département.
   - Les occurrences de l'ensemble des combinaisons années/départements rares sont sommées par prénoms et sexes.

Description des 2 fichiers

Le corpus est composé de 2 tables :

Table	Nombre d'observations	Nombre de variables
NAT2015	589 411	4
DPT2015	3 405 311	5

Les variables :

Variables	Attributs	Signification	Table		Valeurs masquées
			Nat	Dpt	
PREUSUEL	Texte 25	Prénom usuel	X	X	«_PRENOMS_RARES_»
SEXE	Texte 1	Sexe	X	X	
ANNAIS	Texte 4	Année de naissance	X	X	«XXXX»
DPT	Texte 3	Département de naissance		X	«XX»
NOMBRE	Numérique	Nombre de naissances	X	X	

Les départements

En plus de 100 ans, les noms des départements et le découpage en départements ont changé. C'est ce dernier point qui est le plus lourd de conséquences.

Sur toute la période, le département de naissance (variable DPT) est celui existant au moment de la naissance. Ainsi une personne née à Issy-les-Moulineaux en 1949 sera codée en 75 Seine, et une personne née en 1971 à Issy-les-Moulineaux sera codée en 92 Hauts-de-Seine.

En effet, de par la loi n° 64-707 du 10/07/1964, il y a eu création des départements 75 (Paris), 78 (Yvelines), 91 (Essonne), 92 (Hauts-de-Seine), 93 (Seine-Saint-Denis), 94 (Val-de-Marne) et 95 (Val-d'Oise), à partir des anciens départements 75 (Seine) et 78 (Seine-et-Oise), avec une date d'effet dans le fichier au 1er janvier 1968.

Par ailleurs, les deux départements 2A Corse-du-Sud et 2B Haute-Corse sont regroupés (DPT=20, Corse), et il en est de même des départements d'Outre-mer qui sont regroupés (DPT=97). En 2013 cela concerne les naissances de la Guadeloupe, de la Martinique, de la Guyane, de La Réunion et de Mayotte (devenue un DOM en 2011).

Décrets et lois relatifs aux départements métropolitains
- décret du 18/01/1955 : la Seine-Inférieure devient Seine-Maritime
- décret du 09/03/1957 : la Loire-Inférieure devient Loire-Atlantique
- loi n° 64-707 du 10/07/1964 : création des départements 75 (Paris), 78 (Yvelines), 91 (Essonne), 92 (Hauts-de-Seine), 93 (Seine-Saint-Denis), 94 (Val-de-Marne) et 95 (Val-d'Oise)
- loi du 29/12/1967 : rattachement de 6 communes de l'Ain et de 24 communes de l'Isère au département du Rhône
- decret du 10/10/1969 : les Basses-Pyrénées deviennent Pyrénées-Atlantiques
- décret du 13/04/1970 : les Basses-Alpes deviennent Alpes-de-Haute-Provence
- loi n° 75-356 du 15/05/1975 : création des départements 2A (Corse-du-Sud) et 2B (Haute-Corse)
- décret du 27/02/1990 : les Côtes-du-Nord deviennent Côtes-d'Armor

Liste des départements retenus pour le fichier départemental

Code	Département	Code	Département
01	Ain	48	Lozère
02	Aisne	49	Maine-et-Loire
03	Allier	50	Manche
04	Alpes-de-Haute-Provence	51	Marne
05	Hautes-Alpes	52	Haute-Marne
06	Alpes-Maritimes	53	Mayenne
07	Ardèche	54	Meurthe-et-Moselle
08	Ardennes	55	Meuse
09	Ariège	56	Morbihan
10	Aube	57	Moselle
11	Aude	58	Nièvre

12	Aveyron	59	Nord
13	Bouches-du-Rhône	60	Oise
14	Calvados	61	Orne
15	Cantal	62	Pas-de-Calais
16	Charente	63	Puy-de-Dôme
17	Charente-Maritime	64	Pyrénées-Atlantiques
18	Cher	65	Hautes-Pyrénées
19	Corrèze	66	Pyrénées-Orientales
2A	Corse-du-Sud **Fusionnés en 20 Corse**	67	Bas-Rhin
2B	Haute-Corse	68	Haut-Rhin
21	Côte-d'Or	69	Rhône
22	Côtes-d'Armor	70	Haute-Saône
23	Creuse	71	Saône-et-Loire
24	Dordogne	72	Sarthe
25	Doubs	73	Savoie
26	Drôme	74	Haute-Savoie
27	Eure	75	Paris **(Seine avant 1968)**
28	Eure-et-Loir	76	Seine-Maritime
29	Finistère	77	Seine-et-Marne
30	Gard	78	Yvelines **(Seine-et-Oise avant 1968)**
31	Haute-Garonne	79	Deux-Sèvres
32	Gers	80	Somme
33	Gironde	81	Tarn
34	Hérault	82	Tarn-et-Garonne
35	Ille-et-Vilaine	83	Var
36	Indre	84	Vaucluse
37	Indre-et-Loire	85	Vendée
38	Isère	86	Vienne
39	Jura	87	Haute-Vienne
40	Landes	88	Vosges
41	Loir-et-Cher	89	Yonne
42	Loire	90	Territoire de Belfort
43	Haute-Loire	91	Essonne **(à partir de 1968)**
44	Loire-Atlantique	92	Hauts-de-Seine **(à partir de 1968)**
45	Loiret	93	Seine-Saint-Denis **(à partir de 1968)**
46	Lot	94	Val-de-Marne **(à partir de 1968)**
47	Lot-et-Garonne	95	Val-d'Oise **(à partir de 1968)**
		97	DOM

Informations reprises et adaptées du cédérom d'origine de l'INSEE.