# Construction de la base candidats aux élections.

Théo Voldoire, Olivier Godechot, Julie Fournier

Ce document vise à présenter la manière dont a été constituée la base de données « Candidats aux élections ». Cette base vise à permettre l’étude de la carrière politique des élus en France.

**Présentation générale.**

La base prend comme unité de base un individu candidat à une élection : un individu pour une élection par ligne de la base. Du fait de cette contrainte de stricte correspondance entre une candidature individuelle et une ligne de la base, nous avons fait le choix de ne pas la normaliser. Nous avons fait ce choix car il existe une grande diversité des modalités électorales, par exemple entre scrutins à liste et scrutins majoritaires, au gré des transformations du droit électoral, qui rend le schéma normalisé complexe.

La base provient des données du ministère de l’intérieur (<https://www.resultats-elections.interieur.gouv.fr/>). Elle a nécessité un travail considérable de nettoyage et de standardisation pour permettre d’inclure dans une seule archive toutes les informations de ces fichiers. Il subsiste néanmoins des erreurs, n’hésitez pas à nous contacter si vous en trouvez !

**Champ de la base.**

Les candidats de toutes les élections françaises (cantonales/départementales, européennes, législatives, municipales, présidentielles, régionales, sénatoriales) depuis 1995, mis à part les municipales de 1995 et 2001 et les régionales de 1998. Nous n’avons pas trouvé de fichier disponible pour ces élections où figurent le nom des candidats (pas seulement des listes) d’une manière exhaustive. Pour les élections municipales, nous avons inclut une restriction sur la taille des communes, en ne considérant que les villes avec plus de 30'000 inscrits (ce seuil pourrait être rabaissé sous réserve que cela ne dégrade pas l’identification des candidats). Nous avons également renseigné les nominations de ministre, en considérant que leur nomination correspondait à une « date d’élection ».

|  |  |
| --- | --- |
| Election | Nombre de candidats |
| cantonale 1998 | 11159 |
| cantonale 2001 | 11501 |
| cantonale 2004 | 12302 |
| cantonale 2008 | 8520 |
| cantonale 2011 | 10361 |
| departementale 2015 | 18195 |
| departementale 2021 | 7961 |
| europeenne 1999 | 819 |
| europeenne 2004 | 3458 |
| europeenne 2009 | 3287 |
| europeenne 2014 | 3753 |
| europeenne 2019 | 2686 |
| legislative 1993 | 5218 |
| legislative 1997 | 6378 |
| legislative 2002 | 8492 |
| legislative 2007 | 7690 |
| legislative 2012 | 6857 |
| legislative 2017 | 8128 |
| legislative 2022 | 6293 |
| ministere 1995 | 77 |
| ministere 1997 | 26 |
| ministere 1998 | 5 |
| ministere 1999 | 3 |
| ministere 2000 | 41 |
| ministere 2001 | 4 |
| ministere 2002 | 69 |
| ministere 2004 | 83 |
| ministere 2005 | 33 |
| ministere 2007 | 84 |
| ministere 2008 | 38 |
| ministere 2009 | 28 |
| ministere 2010 | 35 |
| ministere 2011 | 15 |
| ministere 2012 | 75 |
| ministere 2014 | 87 |
| ministere 2016 | 38 |
| ministere 2017 | 80 |
| ministere 2020 | 43 |
| municipale 2008 | 41108 |
| municipale 2014 | 47493 |
| municipale 2020 | 57530 |
| presidentielle 1995 | 9 |
| presidentielle 2002 | 16 |
| presidentielle 2007 | 12 |
| presidentielle 2012 | 10 |
| presidentielle 2017 | 11 |
| presidentielle 2022 | 12 |
| regionale 2004 | 17658 |
| regionale 2010 | 20585 |
| regionale 2015 | 21456 |
| regionale 2021 | 19084 |
| senatoriale 1995 | 277 |
| senatoriale 1998 | 457 |
| senatoriale 2001 | 180 |
| senatoriale 2004 | 1240 |
| senatoriale 2008 | 774 |
| senatoriale 2011 | 1275 |
| senatoriale 2014 | 1778 |
| senatoriale 2017 | 1874 |
| senatoriale 2020 | 1476 |

**Schéma de la base.**

Nous organisons les différentes variables de la base en huit grandes catégories. Il y a tout d’abord les variables d’identification : de l’élection ; de la circonscription où le candidat se présente ; de la liste du candidat (si le scrutin est à liste) ; du candidat lui-même. Il y a ensuite des variables de résultat électoral : informations électorales sur la circonscription (participation, abstention, etc.) ; résultat de la liste du candidat (si le scrutin est à liste) ; résultat du candidat seul (si le scrutin n’est pas à liste en plus du résultat, s’il élu ou pas) ; un identifiant du candidat couvrant différentes élections. Les résultats de chaque tour sont précisés sur une seule ligne ; si l’élection n’est qu’à un tour, les résultats sont précisés dans les champs de variable du tour 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Libellé | Modalités | Remarque |
| Identification et informations sur l’élection | | | |
| election | Type d’élection | Municipale, régionale, présidentielle, cantonale, départementale, ministere, senatoriale, legislative, europeenne |  |
| annee | Année de l’élection | Nombre, 1995-2022 |  |
| scrutin | Type de scrutin | Majoritaire, À listes | À corriger |
| election2 | Type et année de l’élection | Concaténation de annee et election |  |
| decoupage | Découpage de l’élection | Circonscriptions européennes, circonscription, municipalité | À compléter |
| Identification de la circonscription[[1]](#footnote-1) | | | |
| code\_circo\_europ | Code de la circonscription européenne |  |  |
| code\_region | Code de région (après réforme 2014) |  |  |
| code\_region\_old | Code de la région (avant réforme 2014) |  |  |
| code\_region\_new | Code de la région (après réforme 2014) |  | À supprimer (redondant) |
| code\_departement | Code de département |  |  |
| code\_commune | Code de commune (code postal) |  |  |
| code\_circo | Code de circonscription |  | Pour les législatives |
| Informations électorales sur la circonscription | | | |
| inscrits\_t1 | Inscrits au premier tour |  |  |
| votants\_t1 | Votants au premier tour |  |  |
| abstentions\_t1 | Abstentions au premier tour |  |  |
| exprimes\_t1 | Votes exprimés au premier tour |  |  |
| inscrits\_t2 | Inscrits au second tour |  |  |
| votants\_t2 | Votants au second tour |  |  |
| abstentions\_t2 | Abstentions au second tour |  |  |
| exprimes\_t2 | Votes exprimés se second tour |  |  |
| Identification de la liste du candidat (pour élections à listes) | | | |
| code\_depot\_liste\_t1 | Code de dépôt (panneau) de liste au t1 |  |  |
| intitule\_liste\_t1 | Intitulé de la liste au t1 |  |  |
| code\_nuance\_liste\_t1 | Code de nuance de la liste au t1 |  |  |
| sexe\_tete\_liste\_t1 | Sexe de la tête de liste au t1 |  |  |
| nom\_tete\_liste\_t1 | Nom de la tête de liste au t1 |  |  |
| prenom\_tete\_liste\_t1 | Prénom de la tête de liste au t1 |  |  |
| code\_depot\_liste\_t2 | Code de dépôt (panneau) de liste au t2 |  |  |
| intitule\_liste\_t2 | Intitulé de la liste au t2 |  |  |
| code\_nuance\_liste\_t2 | Code de nuance de la liste au t2 |  |  |
| sexe\_tete\_liste\_t2 | Sexe de la tête de liste au t2 |  |  |
| nom\_tete\_liste\_t2 | Nom de la tête de liste au t2 |  |  |
| prenom\_tete\_liste\_t2 | Prénom de la tête de liste au t2 |  |  |
| Résultats électoraux de la liste du candidat (pour élections à listes) | | | |
| voix\_liste\_t1 | Nombre de voix de la liste au t1 |  |  |
| sieges\_t1 | Sièges acquis de la liste au t1 |  |  |
| voix\_liste\_t2 | Nombre de voix de la liste au t2 |  |  |
| sieges\_t2 | Sièges acquis de la liste au t2 |  |  |
| Identification du candidat | | | |
| nom | Nom du candidat |  |  |
| prenom | Prénom du candidat |  |  |
| date\_naissance | Date de naissance du candidat |  |  |
| profession | Profession du candidat |  |  |
| code\_csp | Code de catégorie CSP/PCS du candidat |  |  |
| sortant | Le candidat est-il sortant ? |  | À supprimer (redondant) |
| nuance\_candidat | Nuance du candidat |  |  |
| Résultats électoraux du candidat | | | |
| numero\_candidat\_t1 | Place du candidat dans la liste au t1 |  | Pour scrutins à liste |
| voix\_candidat\_t1 | Nombre de voix du candidat |  | Pour scrutins non à liste |
| elu\_candidat\_t1 | Candidat élu au t1 ? |  |  |
| numero\_candidat\_t2 | Place du candidat dans la liste au t2 |  | Pour scrutins à liste |
| voix\_candidat\_t2 | Nombre de voix du candidat |  | Pour scrutins non à liste. |
| elu\_candidat\_t2 | Candidat élu au t2 ? |  |  |
| elu\_candidat | Candidat élu ? |  | Variable agrégée des deux tours. |
| Informations diverses | | | |
| date\_debut | Date de début de mandat |  | Pour les ministre nommés |
| Variables d’identification agrégées | | | |
| id | Identifiant de candidature à une élection |  |  |
| id\_std | Identification d’une personne |  | Problème législatives 2022 |
|  |  |  |  |

Toutes les variables ne sont pas présentes dans toutes les bases. Par exemple, la date de naissance ou la CSP/PCS du candidat ne sont présentes que pour les élections les plus récentes.

**Procédure d’identification des candidats.**

La variable id\_std vise faciliter le repérage de toutes les élections auxquelles une même personne se serait présentée, ainsi que toutes les positions politiques qu’elle aurait occupé. Nous l’avons construite nous-même afin d’éviter les problèmes d’homonymie (nécessairement présents lorsque l’on commence à inclure les élections locales, où le nombre de candidats atteint plusieurs dizaines de milliers). Le problème de relier des entrées de différentes bases est abordé dans la littérature informatique du « record linkage », et notre procédure s’en inspire.

Nous définissons une procédure à partir d’une paire d’élections (A, B). Nous utilisons trois informations comme première identification d’un candidat : le nom, le prénom et la date de naissance. Si la date de naissance n’est pas présente dans un des fichier d’élection, nous ne gardons que le nom et le prénom. Nous comptons le nombre d’occurrence du couple (nom, prénom) dans chacune des bases A et B.

* Si ce couple est présent une (et une seule) fois dans chaque élection A et B, nous réalisons le lien et considérons que les deux candidatures correspondent à la même personne.
* Si le couple est présent plus d’une fois dans une des deux élections, nous ne relions deux candidatures que si elles ont lieu dans la même unité géographique.

Dans le second cas, l’unité géographique est déterminée comme l’unité la plus précise à laquelle il est possible de comparer les deux élections. Si A est une élection départementale et B une élection régionale, le niveau le plus précis auquel il est possible de comparer les élections est la région. Si A est une élection municipale et B une élection départementale, le niveau le plus précis est le département. Nous rajoutons une exception pour les élections municipales, qui représentent un très grand nombre de candidats : nous n’autorisons, dans ce cas, qu’un recoupement avec des élections de niveau départemental ou inférieur.

Nous réalisons la procédure de recoupement sur tous les couples (A, B) d’élections du total. Nous obtenons ensuite une liste de tous les liens entre candidatures, et nous agrégons cette liste par transitivité pour produire un identifiant global.

1. *Dans le cas où une élection est locale (par exemple les municipales), les informations des niveaux supérieurs sont déduites. Par exemple, une candidature aux municipales de Paris, bien qu’elle soit identifiant au niveau municipal, a également les champs de département et de région renseignés. Le code régional principal retenu est celui des nouvelles régions après la réforme de 2014.*  [↑](#footnote-ref-1)