

Projet JADE

Justice algorithmique des élections

Université Grenoble-Alpes, IDEX / IRGA UGA, 2022-2025
IERDJ, 2023-2026

Pré-proposition AAPG ANR 2023-2024 – Consortium UGA /
Université Nice Côte d'Azur – CERDACFF et GREDEG

15 décembre 2023

Semaine « DATA – SHS »

Plan

- 1) Présentation générale du projet JADE (Romain Rambaud)
- 2) L'élaboration technique des bases de données du projet JADE (Caroline Bligny)
- 3) Les problématiques juridiques des bases de données du projet JADE : données de recherche vs. données sensibles (Alya Hafsaoui)
- 4) L'enrichissement des données : la question de l'annotation (Thomas Lebarbé)
- 5) L'utilisation mathématique des données : description voire prédiction (Marie-Jo Martinez, représentée par Caroline Bligny)
- 6) L'apprentissage à partir des données : machine learning et/ou explicabilité (Gilles Bisson et Annette Casagrande)

1) Présentation générale du projet JADE

- **Constat** : montée en puissance de la « justice prédictive », de la « jurimétrie » mais peu de travaux en droit public (ex. Supra Legem)
- **Contentieux des élections politiques** : contentieux ancien, standardisé, normé, chiffré, donc plus réceptif aux outils de la justice algorithmique.... **Problématique** : quelle rationalité = notion de sincérité du scrutin, écart de voix ? (v. R. Rambaud, AJDA)
- **Méthodologie** : hypothèse qu'une « méthodologie » de la justice algorithmique se développe MAIS qu'elle est actuellement surtout développée par des start-ups / Legal tech donc échappent à la recherche publique et au service public

2) Elaboration technique des bases de données JADE

- Les bases de données Décisions : la problématique de la (non)préparation des données / métadonnées par les juridictions (XML – Zip Datagouv)
 - **Conseil constitutionnel** : Les balises sémantiques suivantes permettent de trier informatiquement les décisions selon : <ECLI> (numéro unique de la décision), <URL> (lien internet), <NATURE> (AN, SEN, etc.), <TITRE> (type d'élection et circonscription concernée), <DATE_DEC> (date) et <SOLUTION> (sens de la décision). Il existe en plus sur certains décisions les « Abstracts » en JSON

- **Conseil d'Etat** :

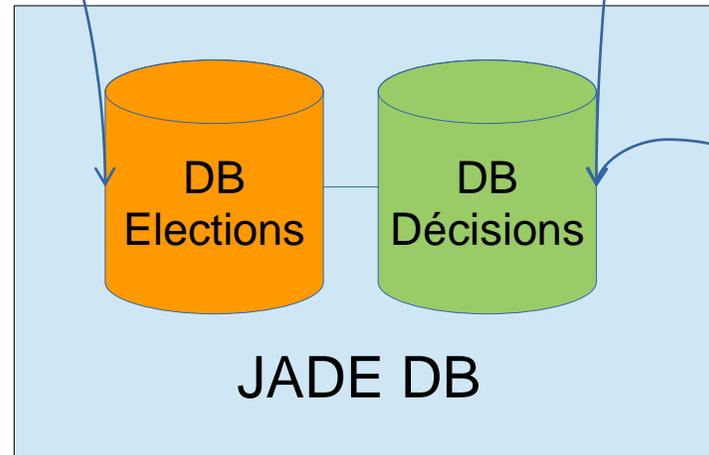
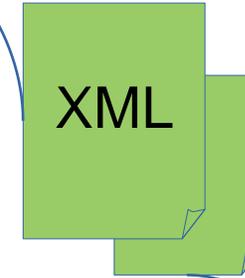
Code PJCA	Texte Code PCJA	Total décisions	Balises solution remplies			Balises solution non remplies	
			Nombre	%	Nb Métadonnées	Nombre	%
28-005	Dispositions générales	240	98	40,8	25	142	59,2
28-01	Présidentielle	13	7	53,8	5	6	46,2
28-02	Législatives	42	23	54,8	7	19	45,2
28-023	Européennes	37	10	27,0	3	27	73,0
28-025 sauf 28-025-04	Régionales	176	29	16,5	10	147	83,5
28-03	Départementales	465	122	26,2	23	343	73,8
28-04	Municipales	1868	470	25,2	61	1398	74,8
28-08	Procédure spéciale	820	359	43,8	104	461	56,2
Total	Elections politiques	3661	1118	30,5	-	2543	69,5

Construction

Résultats des élections,
data.gouv.fr



Décisions du conseil constitutionnel
data.gouv.fr



Annotation des décisions



Experts juridiques du projet

Affichage, visualisation d'une partie de la base Export de données consolidées à des fins d'analyse

Décisions de justice du conseil constitutionnel

- <https://echanges.dila.gouv.fr/OPENDATA/CONSTIT/>

```
<?xml version="1.0" encoding="UTF-8"?>
<TEXTE_JURI_CONSTIT>
  <META>
    <META_COMMUN>
      <ID>CONSTEXT000027146926</ID>
      <ORIGINE>CONSTIT</ORIGINE>
      <NATURE>AN</NATURE>
    </META_COMMUN>
    <META_SPEC>
      <META_JURI>
        <TITRE>A.N., Oise (2ème circ.)</TITRE>
        <DATE_DEC>2013-01-25</DATE_DEC>
        <JURIDICTION>Conseil constitutionnel</JURIDICTION>
        <NUMERO>2012-4594</NUMERO>
        <SOLUTION>Annulation</SOLUTION>
      </META_JURI>
      <META_JURI_CONSTIT>
        <URL_CC>http://www.conseil-constitutionnel.fr/decision/2013/20124594an.htm</URL_CC>
        <ECLI>ECLI:FR:CC:2013:2012.4594.AN</ECLI>
      </META_JURI_CONSTIT>
    </META_SPEC>
  </META>
```

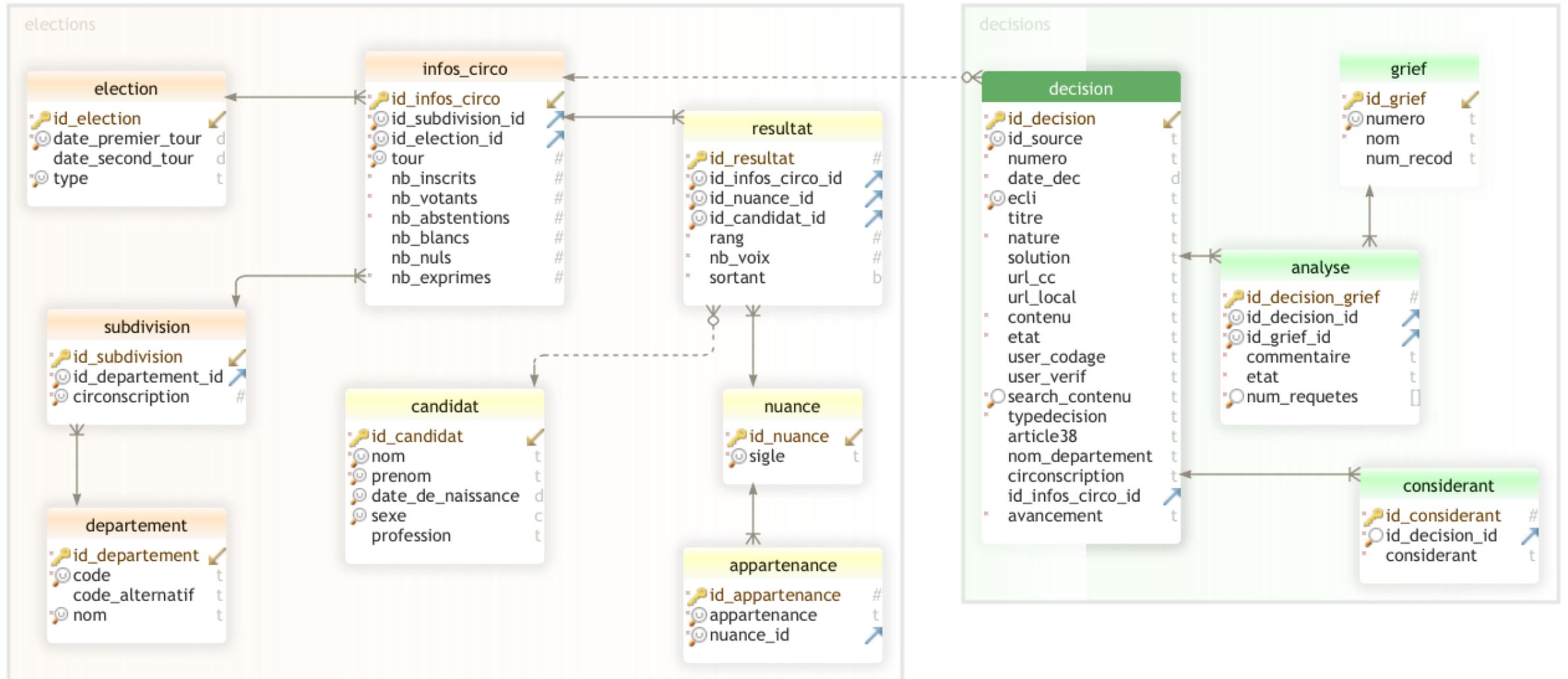
```
<TEXTE>
  <BLOC_TEXTUEL>
    <CONTENU>
      LE CONSEIL CONSTITUTIONNEL,
      <br/>
      <br/>Vu la requête n° 2012-4594 AN présentée pour Mme ...,
      demeurant à Beauvais (Oise),
      ....
      Rendu public le 25 janvier 2013.
    <br/>
    </CONTENU>
  </BLOC_TEXTUEL>
</TEXTE>
</TEXTE_JURI_CONSTIT>
```

Résultats des élections

- Elections législatives et sénatoriales depuis 58
- Récapitulatifs des liens :
 - <https://www.data.gouv.fr/fr/pages/donnees-des-elections/>*
- Fichiers format excel
- 6 liens de téléchargement, 30 fichiers, 13 structures différentes
- Information :
 - Nbr de voix par candidat, département, tour, circonscription
 - Nbr d'inscrits, de votant, de nulls/blancs/exprimés par circonscription
 - Nuance politique, nom et genre candidat après 1988

Moyens

- Construction d'une base de données relationnelle, postgresQL
- 2 stages :
 - Intégration des données des élections
 - Développement d'une interface permettant l'enrichissement juridique (partie Griefs)
- Développement code python / SQL
- Mise en place d'un serveur UGA, infrastructure Winter



Base de données JADE, Modèle Physique des Données

Quelques chiffres

- Au 04 mai dernier :
- **3888** décisions, dont **3365** connectées à une élection
- **37** élections
 - **8 466** résultat par circo
- **87 146** résultats par candidats
- **1442** association grief-décision

72	Annulation
21	Annulation - Inéligibilité
1	Approbation
1	Conformité et rejet
6	Enquête
1880	Inéligibilité
5	Irrecevabilité
223	Non lieu à prononcer l'inéligibilité
47	Non lieu à statuer
2	Non lieu à statuer sur la QPC et annulation
1	Null
8	Rectification d'erreur matérielle
1618	Rejet
1	Rejet - Conformité [QPC]
1	Rejet - non lieu à statuer [QPC]
1	Réformation

Export pour les analyses statistiques

Caractéristiques décisions
(solution, numéro, ..)
+ bilan élection

Au moyen de vues SQL + export
csv

v_bilan	
id_infos_circo_tf	#
id_infos_circo_t1	#
annee_election	#
date_premier_tour	d
date_second_tour	d
type_election	t
nom_dep	t
circonscription	#
tour_final	#
tf_nb_exprimees	#
tf_nb_votant	#
tf_nb_inscrit	#
tf_ecart1	#
tf_r1_nuance	t
tf_r1_nb_voix	#
tf_r1_sortant	b
tf_r1_civilite	c
tf_r2_nuance	t
tf_r2_nb_voix	#
tf_r2_sortant	b
tf_r2_civilite	c
t1_nb_exprimes	#
t1_nb_votant	#
t1_nb_inscrit	#
t1_ecart1	#
t1_ecart2	#
t1_r1_nuance	t
t1_r1_nb_voix	#
t1_r1_sortant	b
t1_r1_civilite	c
t1_r2_nuance	t
t1_r2_nb_voix	#
t1_r2_sortant	b
t1_r2_civilite	c
t1_r3_nuance	t
t1_r3_nb_voix	#
t1_r3_sortant	b
t1_r3_civilite	c

v_resultat	
id_resultat	#
id_infos_circo	#
nuance	t
nom_candidat	t
prenom_candidat	t
civilite_candidat	c
rang	#
rang_calc	#
nb_voix	#
sortant	b
id_candidat	#
id_nuance	#

v_infos_circo	
id_infos_circo	#
nom_dep	t
code_dep	t
tour	#
tour_inverse	#
circonscription	#
date_premier_tour	d
date_second_tour	d
type_election	t
nb_inscrits	#
nb_votants	#
nb_abstentions	#
nb_blancs	#
nb_nuls	#
nb_exprimes	#
id_subdivision	#
id_election	#

Bilan

- Intégration des élections : travail bien avancé, premières analyses statistiques possibles. Des éléments à compléter (ex, législatives partielles)
- Enrichissement des données juridiques : travail d'annotation à réaliser avec des outils dédiés
- Interface web publique d'accès aux données : à faire

3) Les problématiques juridiques des bases de données du projet JADE : données de recherche vs. données sensibles

L'open data des données de justice : un préalable nécessaire

→ Loi dite CADA de 1978 : Droit d'accès aux documents administratifs

→ Ordonnance de 2005 : Droit d'accéder aux informations contenues dans les documents administratifs

→ Loi pour une République numérique de 2016 : Ouverture par défaut des données publiques et dispositif de réutilisation (articles 20 et 21)

→ Horizon 2030 : Politique européenne d'ouverture des données de la recherche



La protection des données à caractère personnel

Définition : toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement

Outils et concepts :

- accountability (analyse d'impact, plan de gestion des données)
- compliance (privacy by design)
- DPO

L'adaptation des grands principes du RGPD “as open as possible, as closed as necessary”

Dérogations reconnues par les article 9 RGPD et 80 LIL au titre de l'expression universitaire (85 RGPD)

- Principe de licéité (intérêt légitime, mission d'intérêt public, consentement, obligation légale)
- Droit à l'information
- Droit à l'opposition
- Minimisation des données
- Principe de loyauté du traitement
- Droit d'accès, de rectification et à la limitation
- Droit à l'effacement

⇒ L'échappement au RGPD : l'anonymisation (rendre impossible toute identification de la personne)

L'adaptation nécessaire de la protection des données personnelles

→ L'illicéité d'une anonymisation générale : l'article L. 112-1 du code de la recherche

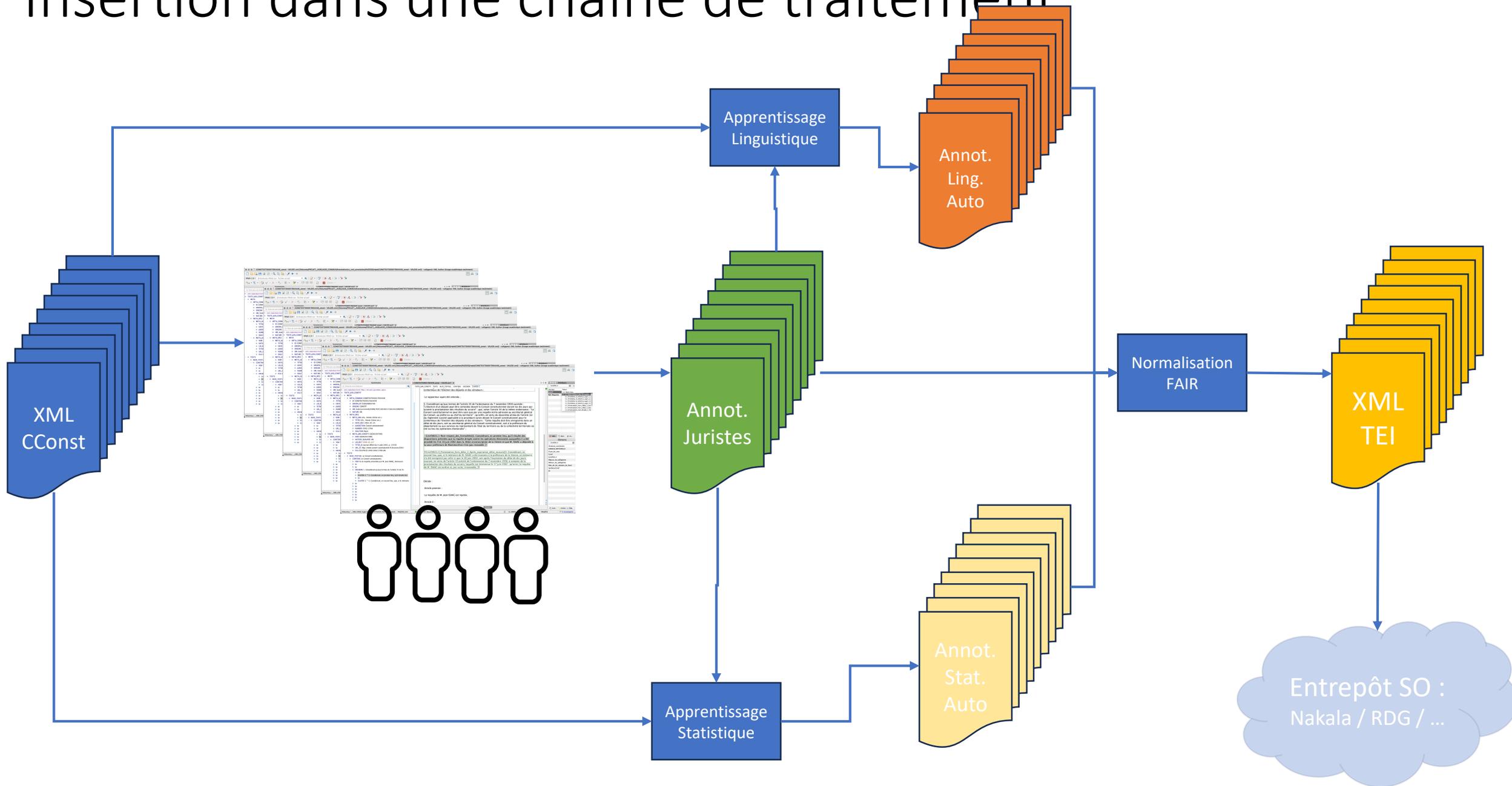
→ L'inconventionnalité de l'anonymisation générale française : la nécessité de mise en balance de la liberté d'expression et la protection des données personnelles :

- CJUE 24 sept. 2019, aff. C-136/17, Google c/ Cnil, Légipresse
- CEDH 28 juin 2018, M. L. et W. W. c/ Allemagne, no 60798/10

4) Enrichissement de la base de données et annotation des décisions

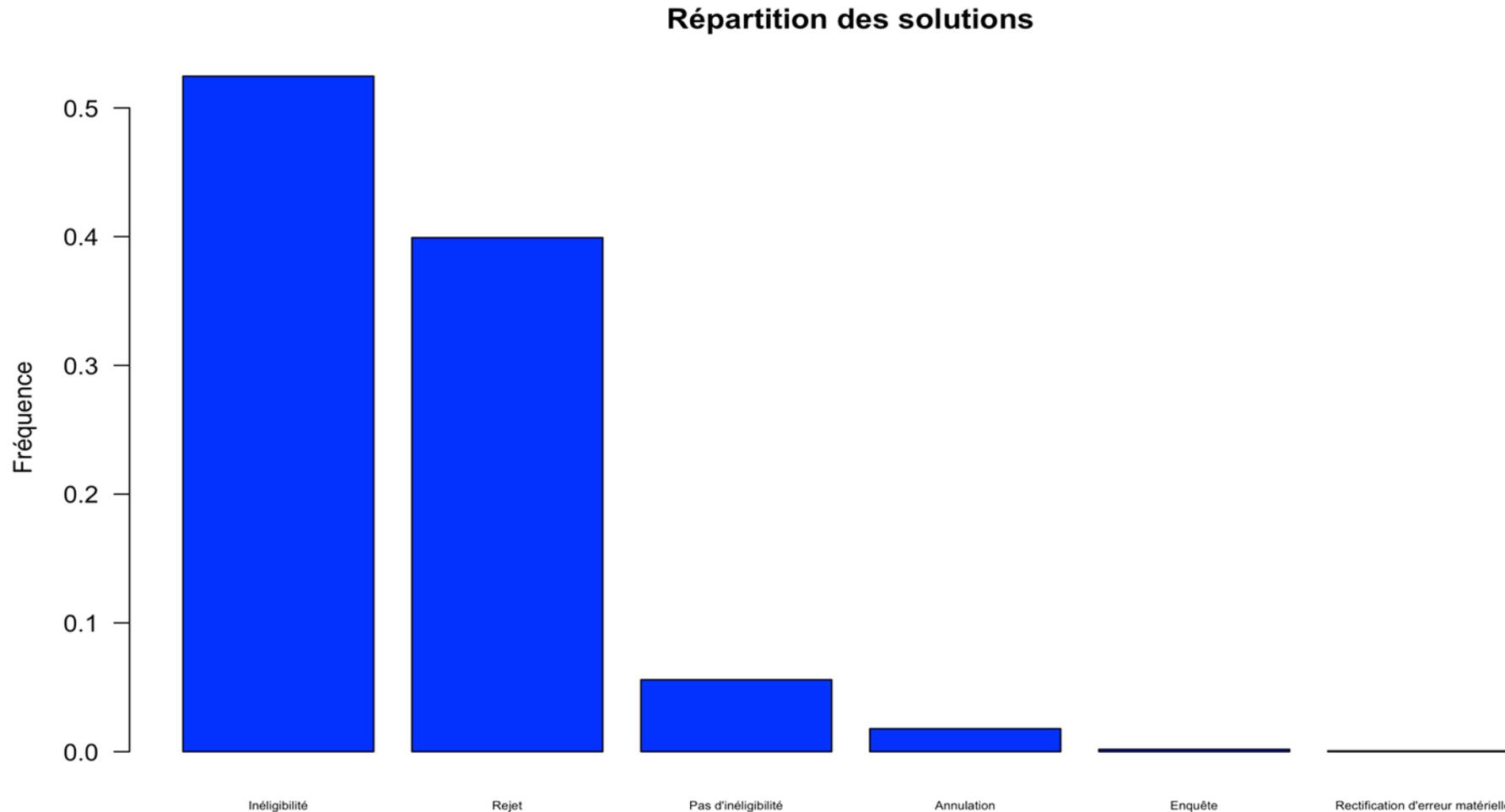
- De la catégorisation des décisions
À la sémantisation de segments de décisions
 - Finalité(s) :
 - Identifier les énoncés pertinents
 - Rechercher les marqueurs linguistiques
 - Disposer d'une donnée structurée et sémantisée pour les apprentissages
 - Moyen(s) :
 - Gestes techniques simples → ne pas encombrer l'annotateur
 - Lexique des détenteurs de savoir → éviter le jargon normatif
 - Coût humain non négligeable → qualité de la donnée
 - Résultat(s) :
 - Une grammaire en cours de stabilisation
 - Une prise en main rapide et efficace par l'équipe
 - Un premier échantillon de données

Insertion dans une chaîne de traitement



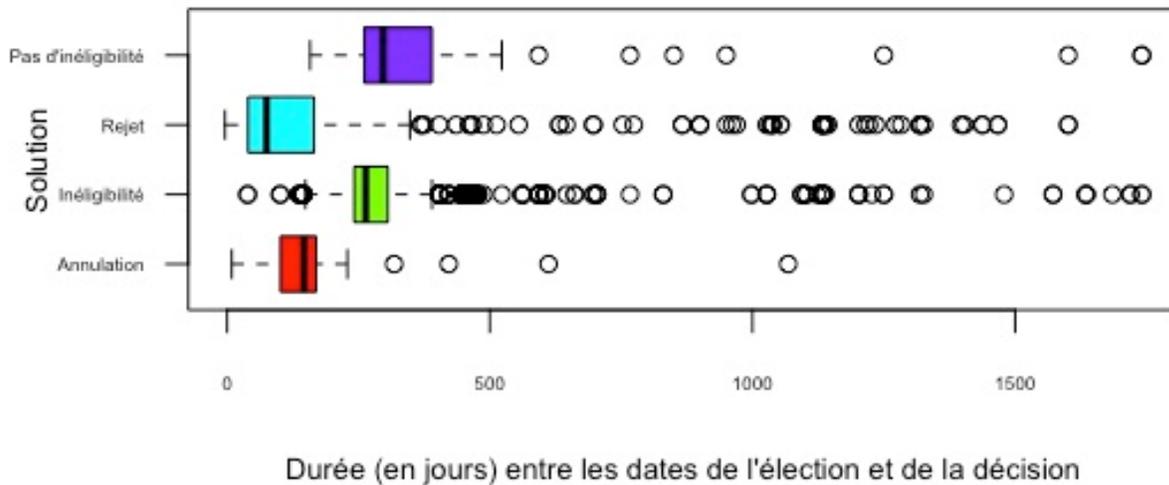
5) Exploitation des données / description

- Techniques statistiques descriptives sur le rapport Décisions / Elections

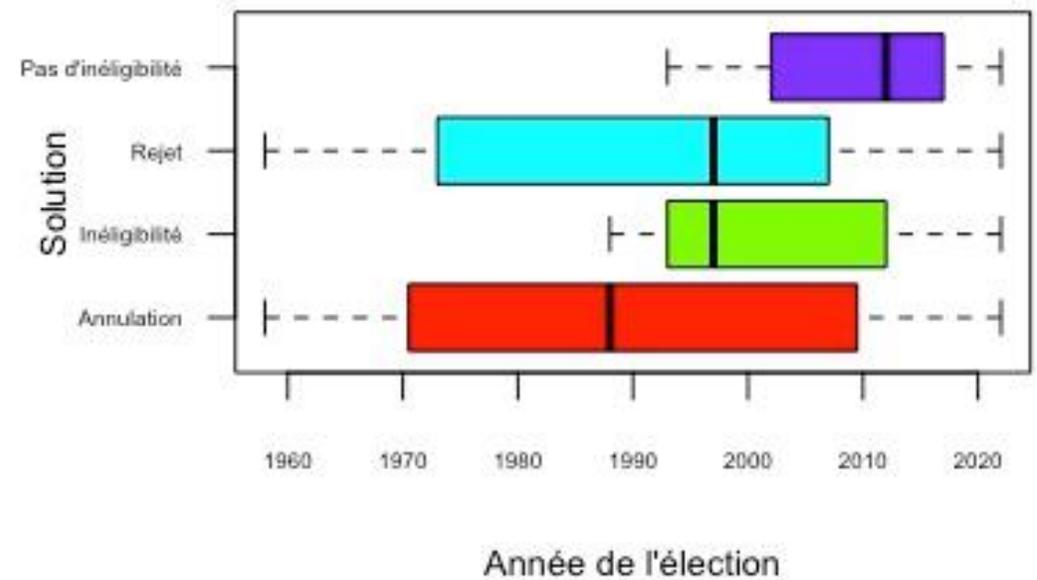


La temporalité du contentieux électoral : : des statistiques confirmatives

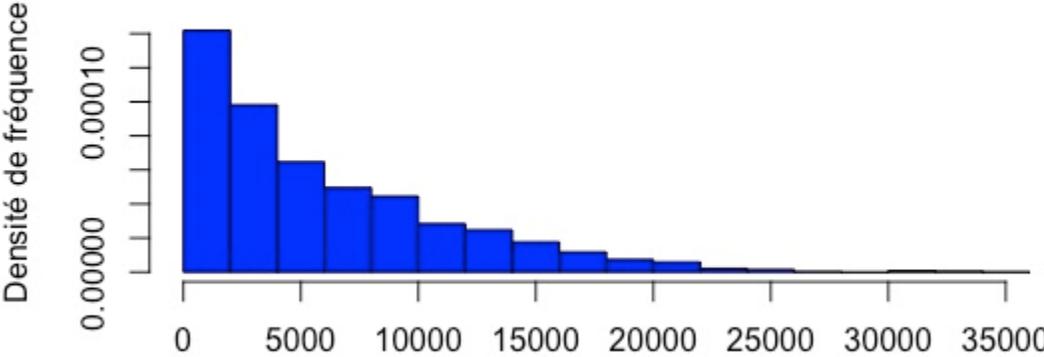
Répartition des durées par solution



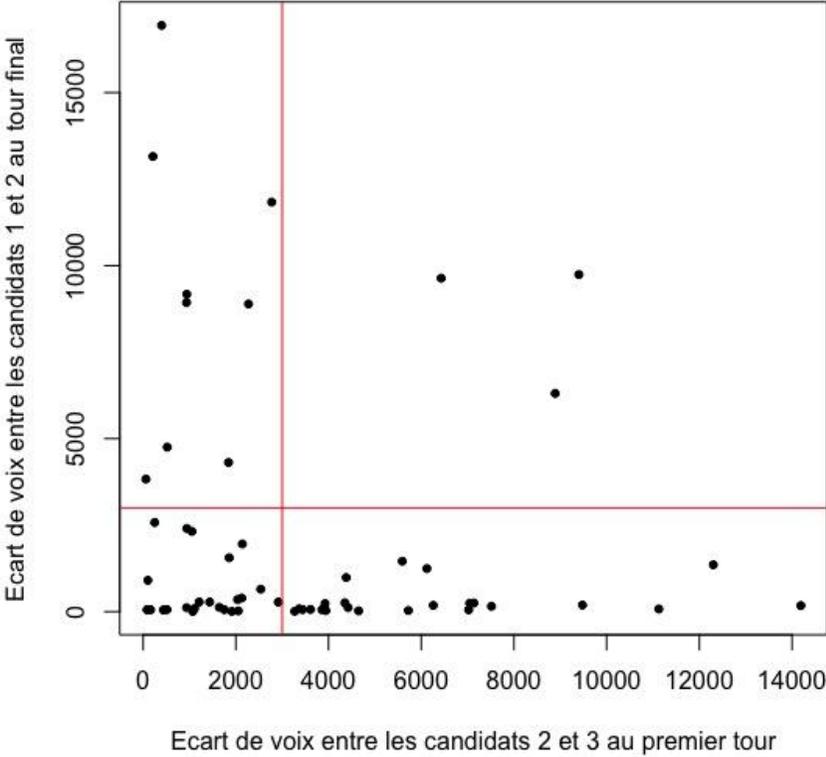
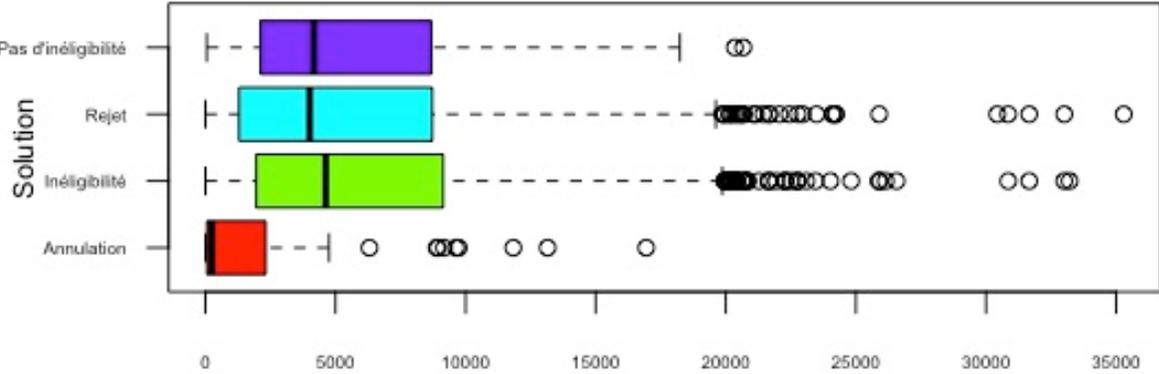
Répartition des années d'élection par solution



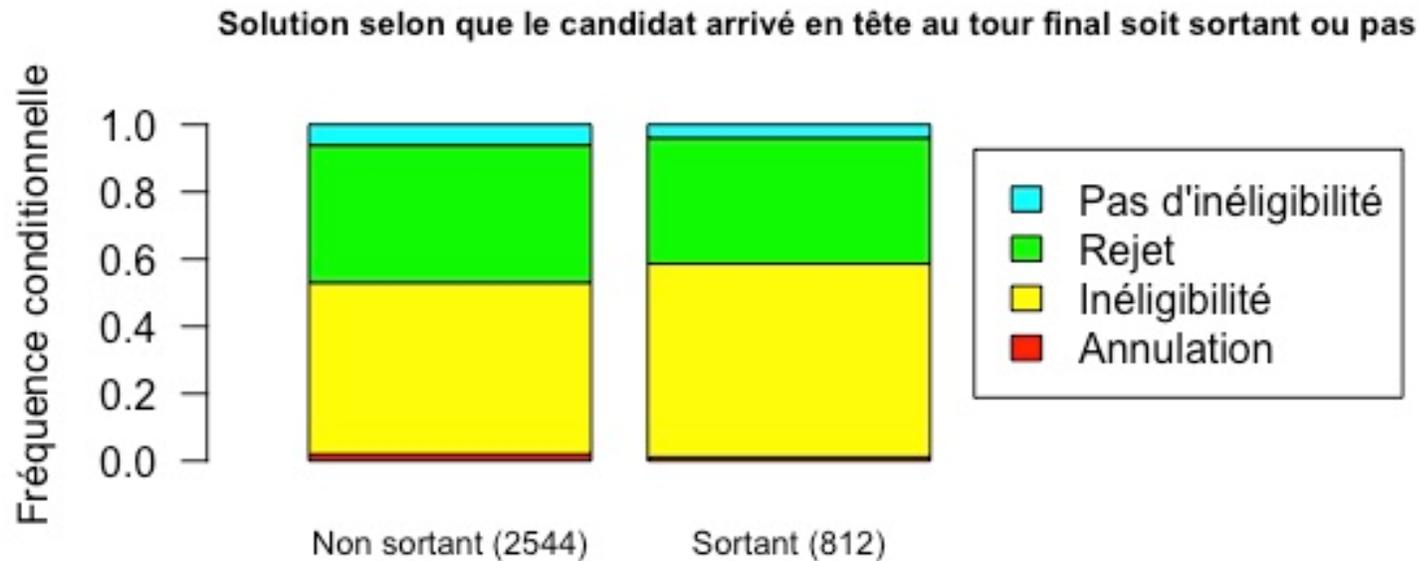
Écarts de voix : des statistiques confirmatives



Répartition des écarts de voix entre les candidats 1 et 2 par solution



Un exemple de résultat nouveau ? La solution contentieuse selon la qualité ou non de sortant...



6) L'apprentissage à partir des données : machine learning et/ou explicabilité (exploratoire)

Présentation par Gilles Bisson et Annette Casagrande