

# *Entretien StGobain*

Etude de cas

Cyril Becot

31.01.2022

# Introduction

- *Objectif : comprendre les ressorts derrière les resultats* - 5 pistes :
  - Distribution géographique des votes ?
  - Balance des *ages & des sexes*
  - Corrélation avec la *dynamique des ménages* : revenus et taille
  - Impact du niveau *d'éducation & des catégories socio-professionnelles*
- Données disponibles (par ville)
  - *Resultats* du referendum
  - *Recensement* : population, classes d'âge et catégories socioprofessionnelles
  - Distribution des *revenus* : revenus medians & ecart de salaire
  - Niveau d'*education* : derniers diplomes atteints
  - Données géographiques : *frontieres* des villes — Correspond a **01/01/2022** !

# Implementation technique

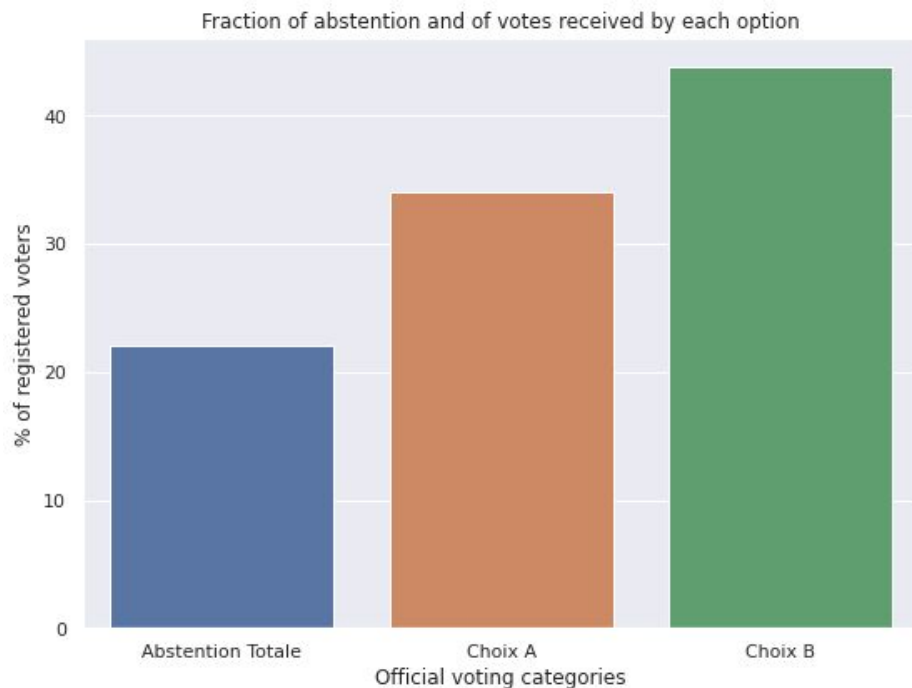
- Lecture et transformation des données : (geo)pandas
  - Classe 'helper' pour faciliter la lecture/nettoyage (dataclass)
- Gestion du code via github
  - Tests unitaires & pipeline docker
- Visualisation avec seaborn/matplotlib
- Analyse exploratoire faite sur Jupyter Lab/Jupyter Notebook
  - *Depuis le container docker cree, qui contient toutes les données*

# Nettoyage & indexation des données

- *Territoires hors metropole* (code dept 97 ou 'Z\*')
  - TOM & Français de l'étranger seulement dans le référendum
  - DOMs absents de certaines BDD
  - Par défaut **garde seulement la métropole**
- *Grandes villes* (Paris/Lyon/Marseille)
  - **Garde comme une entité** - fusion/suppression d'arrondissements nécessaire
- Données vides : seulement dans diplomes (villes apparaissant/disparaissant), drop
  - Revenus peut aussi avoir des cases vides sur des villes trop petites
- Choix de l'index : code INSEE, à construire (code dept + code ville, 2+3 digits)
- Recensement "peu précis" sous 200 habitants, et pour certaines entrées sous 2000.  
Les villes en dessous de ces valeurs seront enlevées

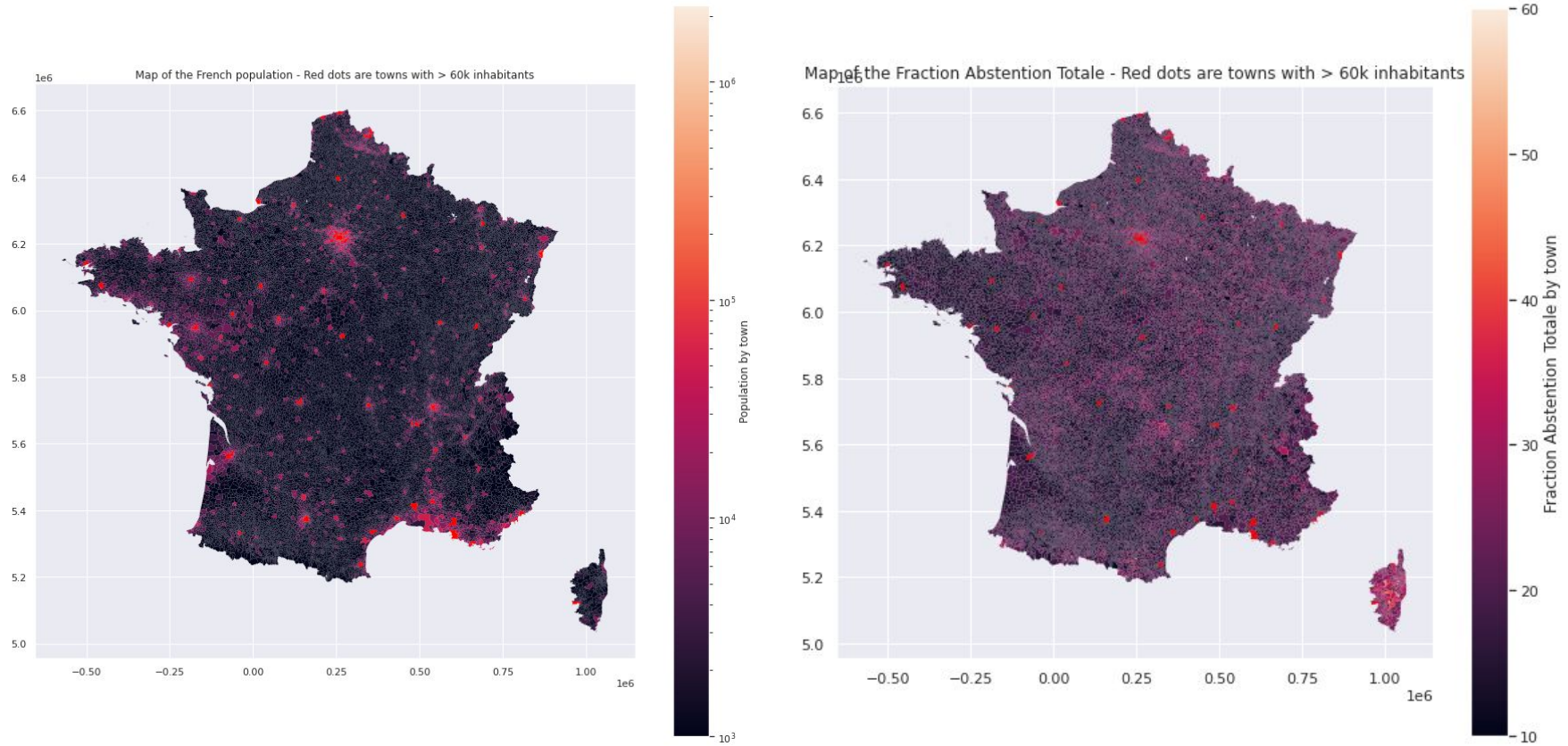
# Proposition gagnante

- Ici (seulement), toutes données du referendum utilisees, inc. DomTom & Etranger et villes sous 200 habitants



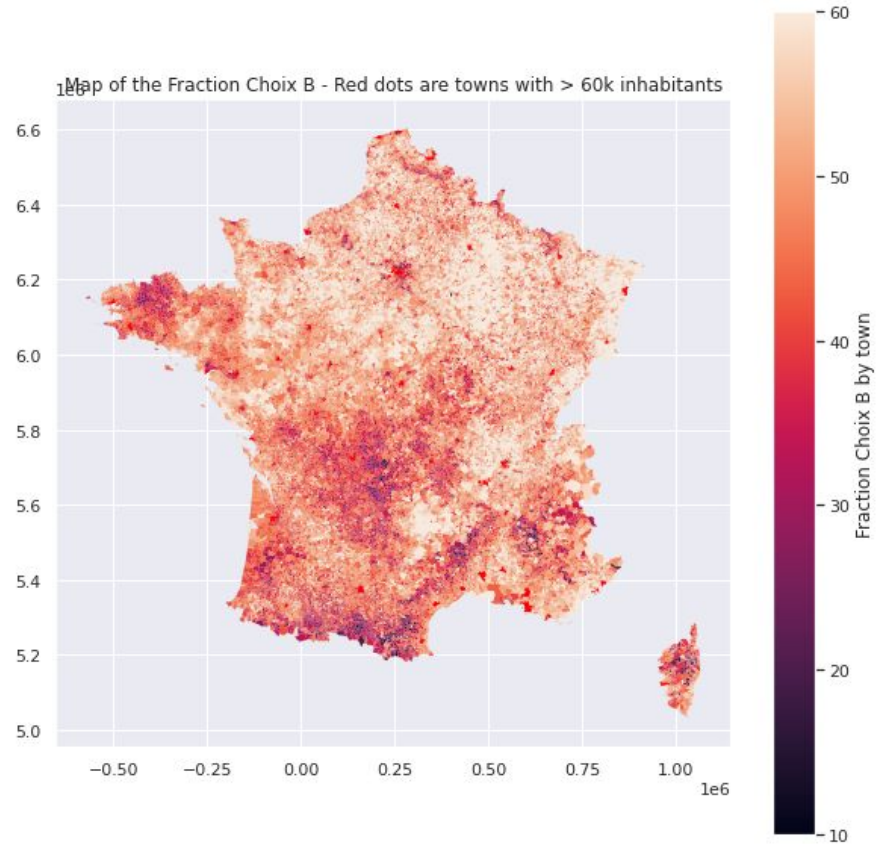
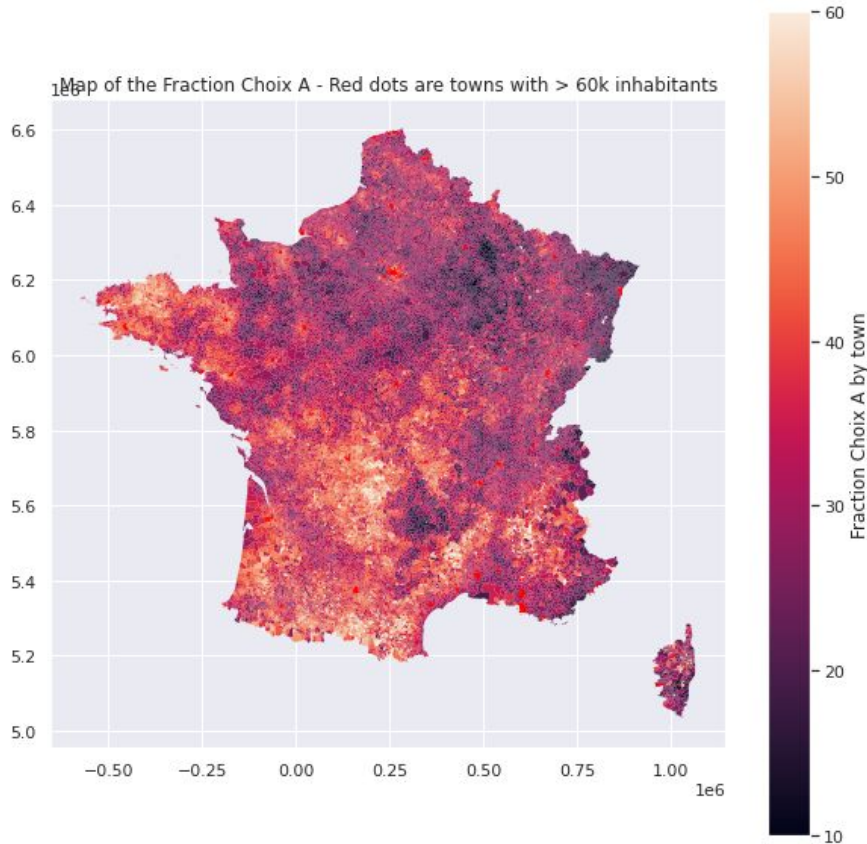
- Donc Choix B est clairement plebiscite

# Géographie de la population et de l'abstention



Abstention importante en region parisienne (carte sans marqueur)

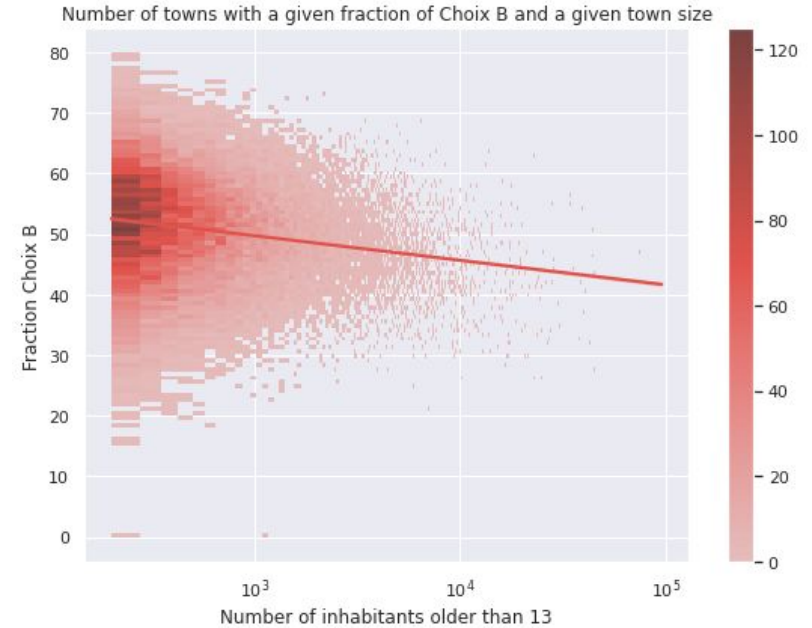
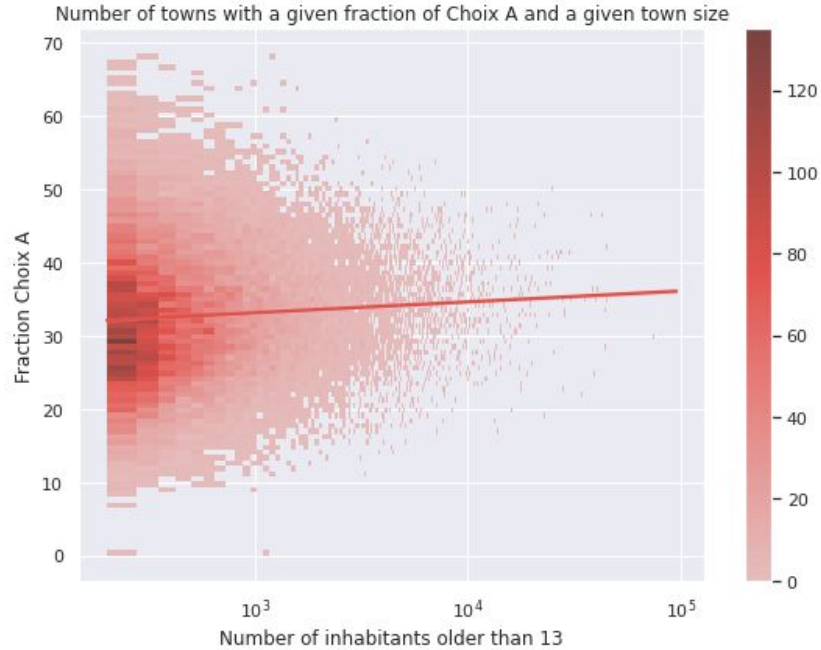
# Géographie des résultats



**Clivage Nord-Est/Sud-Ouest**

Pas de corrélation claire avec les centres urbains

# Résultats vs taille de la ville



Lien clair entre les résultats et la taille de la ville (log)  
L'abstention augmente aussi en log(taille)

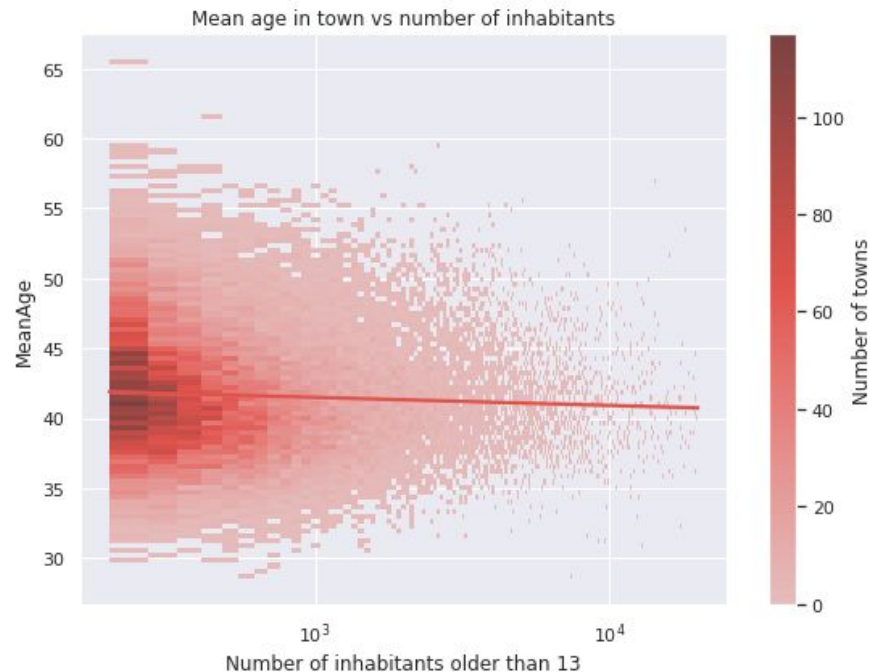


# Traitement du genre, du socio-professionnel, age et diplome

- Données catégorielles : utilisation la classe la plus importante pour étiqueter la ville
  - Et ensuite, *pondérer les données par le nombre d'habitants*
  - Donne un genre/une catégorie socio-pro à chaque ville (dilue l'info)
- Alternative possible : même résultats dans toutes les catégories pour chaque ville
  - Multiplier les effectifs de tranche d'âge puis agréger
  - Probleme : interdit les dynamiques d'âge/socio-pro dans chaque ville...
- Age : calcul de l'âge moyen depuis les classes d'âge
- Niveau d'éducation : crée un "niveau d'étude" = années d'études (depuis naissance) nécessaires pour obtenir ce diplôme (6 à 24 ans)

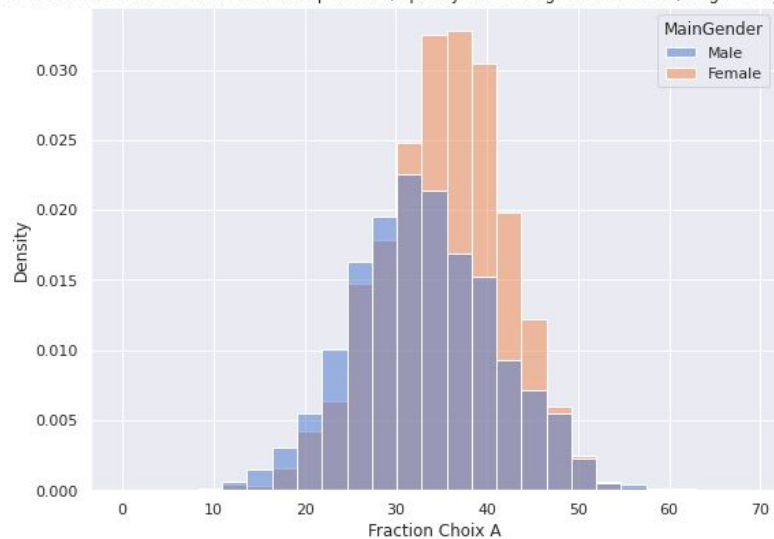
# Illustration importance ponderation

- Corrélation entre âge moyen dans une ville et sa taille
- Et comparativement plus de petites villes
- Sans pondération on augmente l'impact d'une population plus agee

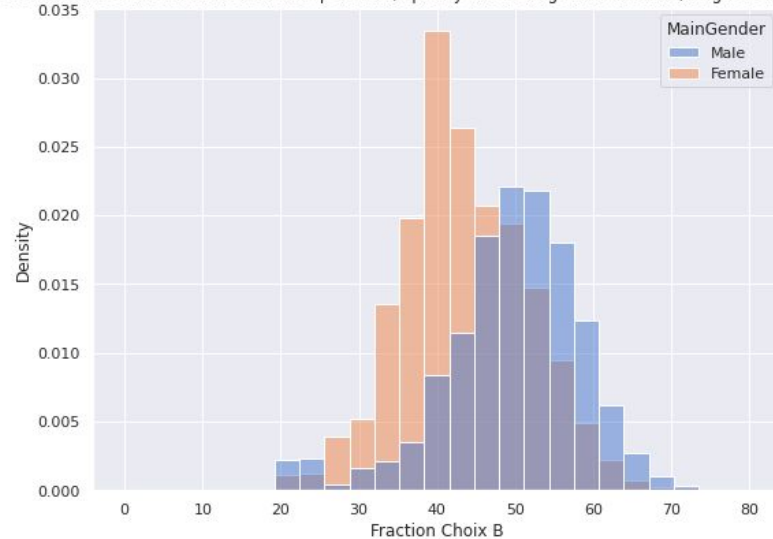


# Résultats vs sexe

Distributions of the fraction of Choix A per town, split by the main gender in town (weighted by town size)

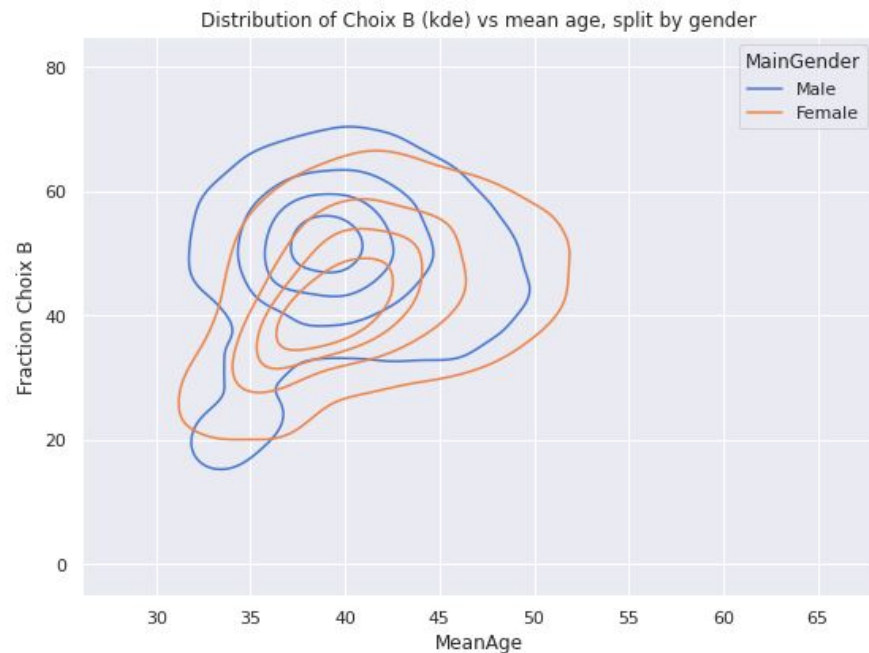
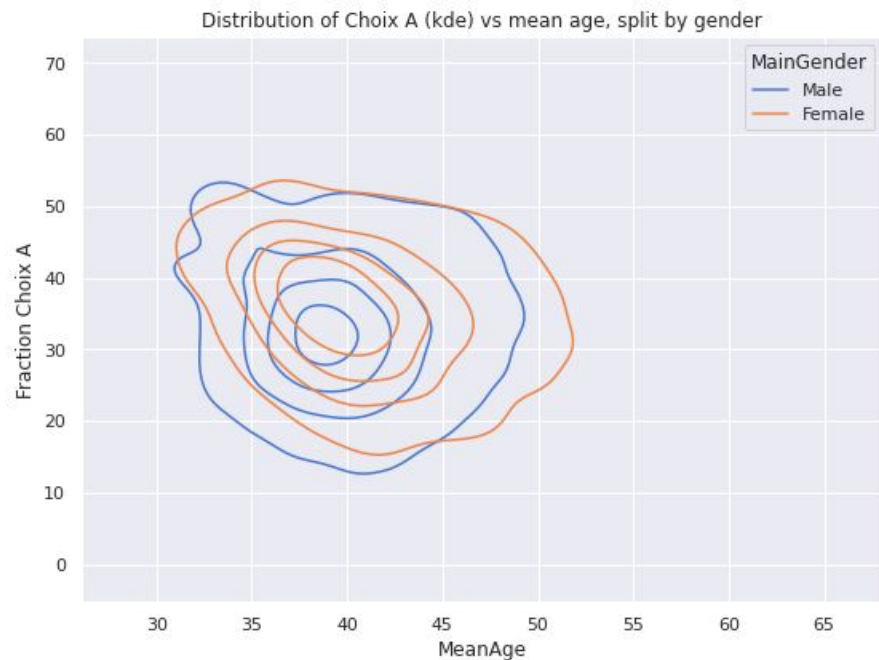


Distributions of the fraction of Choix B per town, split by the main gender in town (weighted by town size)



Préférence des femmes pour B moins marquée  
Elles s'abstiennent aussi plus

# Résultats vs age



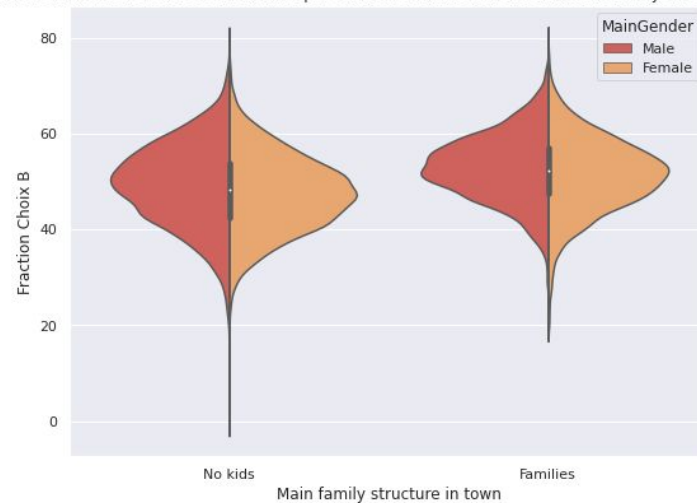
Les femmes âgées ont plus tendance à être en faveur de B  
Correlation age/vote aussi observée en "population générale"

# Résultats vs nombre de personnes/familles

Distribution of the fraction of vote for Choix A per town, as a function of the dominant family structure and gender



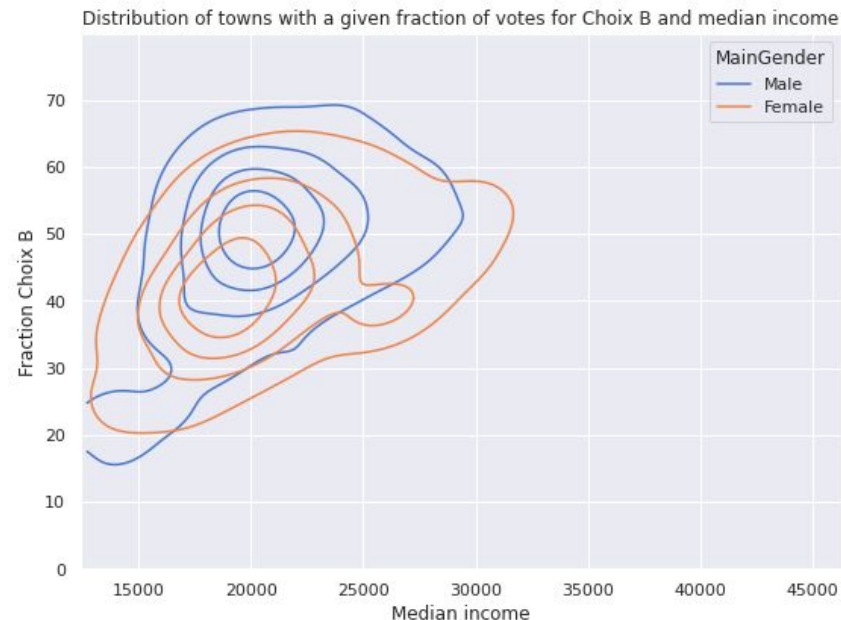
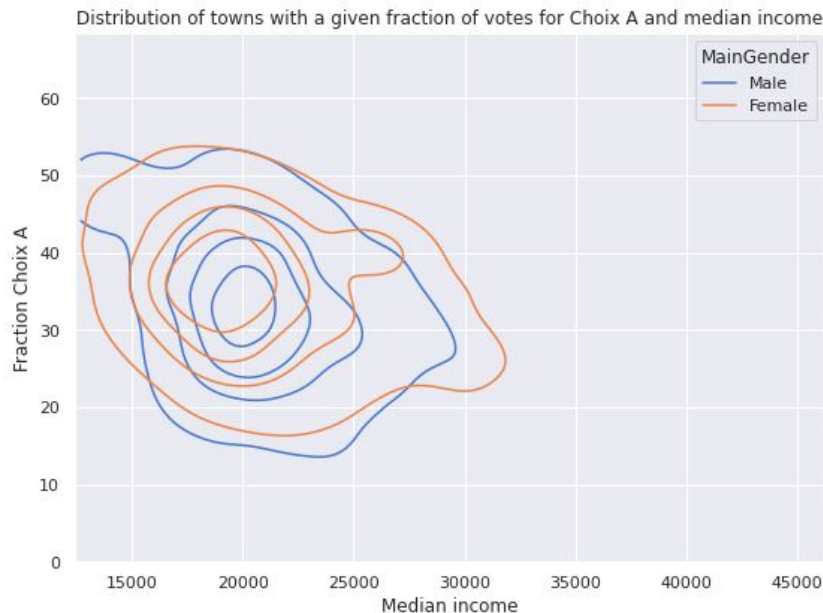
Distribution of the fraction of vote for Choix B per town, as a function of the dominant family structure and gender



Après arrondi le nombre moyen de personne par foyer dans chaque ville a deux valeurs :  
2 (dominé par des couples sans enfants/célibataires) ou 3 (familles)

Préférence pour B plus marquée dans les villes plus familiales

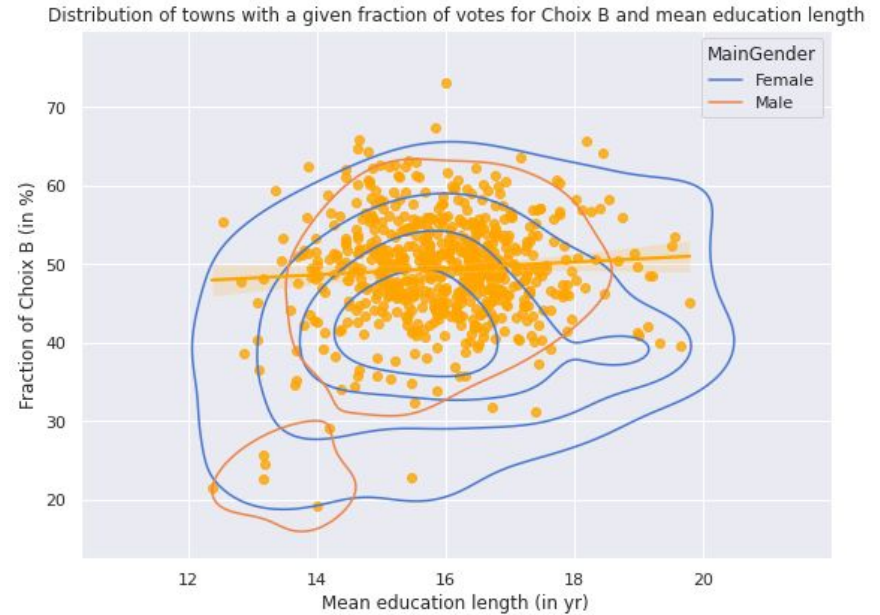
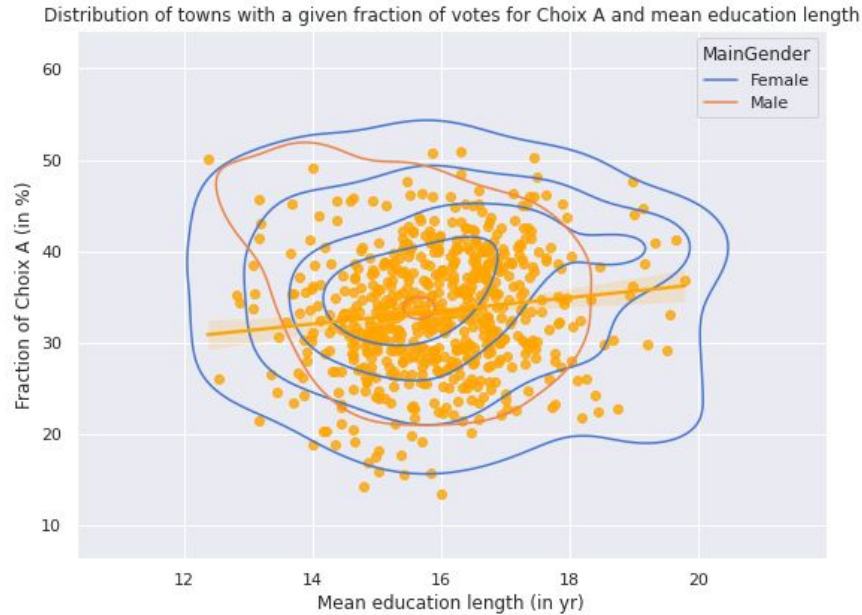
# Résultats vs revenu median



*Correlation entre revenu median et preference pour B plus marquée chez les femmes*  
A verifie : par de correlation evidente entre revenu median et age median

“Wage gap” : principale correlation avec le taux d’abstention (cf notebook)

# Résultats vs niveau d'éducation

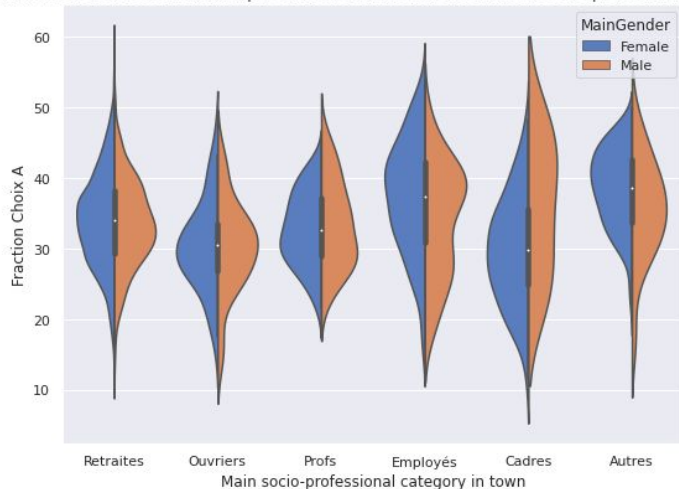


Légère augmentation du vote pour A en fonction du niveau d'éducation  
Ajout d'une distribution des villes/régression pour les hommes, problème avec KDE (stat?)

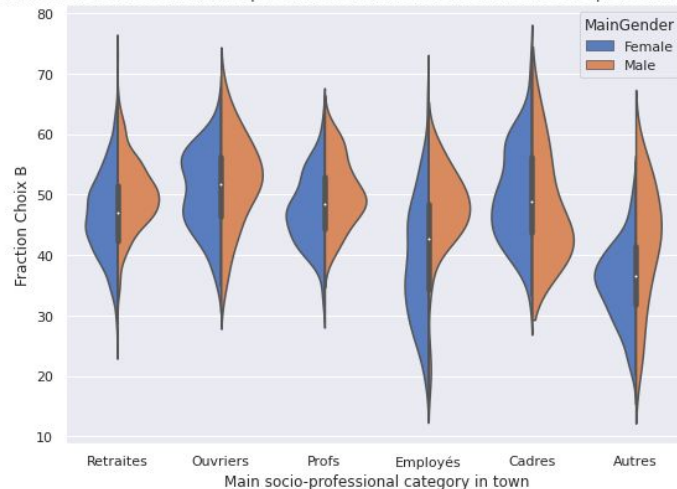
Principal impact du niveau d'éducation : diminution de l'abstention

# Impact des categories socio-culturelles

Distribution of the fraction of vote for Choix A per town, as a function of the dominant socio-professional category and gender



Distribution of the fraction of vote for Choix B per town, as a function of the dominant socio-professional category and gender



Pas toujours suffisamment de stat, une autre approche pourrait etre porteuse  
(peu de villes dominees par des hommes et des cadres)

Toutefois : les hommes semblent dominer le vote pour B chez les employés et autres,  
et les femmes le vote pour A chez les “autres”

Dans la majorite des categories la distribution des taux de vote est relativement symétrique entre les sexes 16



# Conclusions

- La préférence pour B est plus marquée chez les hommes que les femmes
  - Toutefois elle en manque chez les femmes plus vieilles ou plus leurs revenus sont importants
- Les familles semblent légèrement plus en faveur de B
- Plus les villes sont importantes, plus les gens votent pour A au détriment de B
- Il semble y avoir un clivage Sud-Ouest/Nord-Est dans ce vote
- Le niveau d'éducation a le même impact sur les deux votes, en cela qu'il impacte surtout le niveau d'abstention.
- Il ne semble pas y avoir de tendance socio-professionnelle claire dans ce vote
- *Des graphiques additionnels sont présents dans le notebook*