



UNIVERSITÉ DE NANTES



IAE NANTES
ÉCONOMIE & MANAGEMENT

Projet économétrie des variables qualitatives

Madame TRAVERS

ROMAND Kyllien
DEL'CHATEAU Jean-Baptiste

2020/2021

Sommaire

Introduction

Partie économique

Explication du questionnaire

Partie statistique

Partie économétrique

Conclusion

Bibliographie

Annexes

Introduction

“La télévision est là pour créer de l’émotion. [...] En fait on vote pour un homme et pas un parti, une idée et pas une idéologie. La télévision s’utilise donc pour créer de l’affect.” Jacques Séguéla, publicitaire.

En 2018, une affaire a fait grand bruit dans le monde médiatique, il s’agit de l’affaire Cambridge Analytica, affaire nous ayant tout particulièrement intéressé, c’est pour cette raison que nous avons décidé de travailler sur un sujet dans l’axe de cette affaire. Dans ce dossier, nous allons essayer de prédire les votes des différents sondés pour les prochaines élections, à partir d’un questionnaire posé, afin de démontrer le risque de dévoiler sa vie privée sur internet, en particulier les réseaux sociaux. Il existe certaines campagnes de sensibilisation contre le risque de dévoiler sa vie sur les réseaux sociaux mais celle-ci ne sont pas assez nombreuses et impactantes, puisque certaines personnes dévoilent encore tout sur leur personne et leur famille, ces personnes ne se rendent pas compte des risques liés à cette surexposition dans les réseaux.

Le fait de dévoiler sa vie privée sur les réseaux sociaux s’est intensifié avec la démocratisation de l’accès à internet. L’apparition d’internet s’est faite en janvier 1983, la démocratisation a commencé en 1990 avec l’apparition du www (World Wide Web) et le premier navigateur en 1994. Le premier réseau social se fait en 1997, mais il s’agissait plus d’un site de médias sociaux. En 2004 un géant fait son apparition, il s’agit de Facebook, c’est donc à partir de cette années que les adhérents ont commencé à exposer leur vie privée à tout le monde sans savoir que Facebook gardait les données et les analysait. Et depuis ces années, le nombre de réseaux sociaux et autres récupérateurs de données n’ont cessé de croître ce qui a entraîné une croissance de ces entreprises. Ces entreprises n’ont jamais été aussi puissantes, de plus la crise actuelle n’a fait qu’augmenter leur puissance. Les 4 géants d’internet (Les GAFA) ont donc des milliards de données sur tous les utilisateurs des différentes plateformes, c’est avec ces données que l’affaire Cambridge Analytica a eu lieu.

Ce dossier va être divisé en 5 parties, nous allons voir dans un premier temps la partie économique de notre sujet, puis nous passerons dans un second temps à l’explication du questionnaire, dans un troisième temps, nous verrons la partie

statistique, à la suite de cette partie nous verrons la partie économétrique et nous finirons par une conclusion sur ce dossier. Concernant notre base de données, elle a été faite à l'aide d'un sondage que nous avons créé, nous disposons de 189 individus, une variable dépendante avec 4 valeurs et 23 variables explicatives.

Analyse économique du sujet

Nous allons commencer par expliquer ce qu'est le scandale Cambridge Analytica, le scandale commence en mars 2008 avec deux journaux, l'*Observer* de Londres et le *New York Times*, publiant que l'entreprise Cambridge Analytica avait recueilli des données personnelles issues des profils Facebook de plus de 87 millions d'utilisateur sans leur consentement. Cambridge Analytica est une société de conseil stratégique basée à Londres. La collecte des données s'est faite via un quizz élaboré par A. KOGAN, un psychologue à l'université de Cambridge, grâce à ce quizz, ce psychologue a eu accès à plus de 200 millions de données appartenant aux membres qui ont accepté de répondre au quizz, les répondant au quizz ont été rémunérés pour répondre au questionnaire. Ce questionnaire a été constitué de différentes questions liées aux personnalités des sondés tel que leur impulsivité, leur enthousiasme, leur anxiété, leur agréabilité, leur extraversion; il a permis de "créer" artificiellement le type de personnalité de chacun. C'est la raison pour laquelle dans notre questionnaire nous utilisons cette méthode afin de déterminer si elle est probante ou non. Ce questionnaire a agit comme une sorte de "virus" puisque c'est grâce à celui-ci que les données ont pu être récupérées, le nombre important de données récupérées, vient du fait que les amis des personnes ayant répondu au sondage ont eux aussi été piratés. Ce questionnaire permettait de récupérer différents types de données, les contenus consultés, likés et les messages publiées. Une fois ces données récupérées par le questionnaire, elles étaient directement transmises à Cambridge Analytica, puis recruté par les équipes de Donald TRUMP pour la campagne de 2016. Ces données auraient été utilisées pour adresser des messages politiques personnalisés aux utilisateurs, semer le doute dans l'esprit des votants et favoriser la candidature de TRUMP via des pubs, des publications, des vidéos émises sur leur Facebook. Mais le scandale n'a pas éclaté avant 2018 quand le président du parlement européen a déclaré que ce scandale est "une violation inacceptable des droits des citoyens en matière de vie privée". Après ce message posté sur Twitter, plusieurs pays ont commencé à ouvrir une enquête sur ce scandale puis le 2 mai 2018, sous la pression des différents acteurs gouvernementaux et non gouvernementaux, SCL une société britannique privée

fondée en 1990 qui est aussi la maison mère de Cambridge Analytica a annoncé sa fermeture et donc de leur filiale Cambridge Analytica.¹

Bien que cette affaire soit choquante, celle-ci est un exemple mineur de tout ce qui peut se passer en matière de collecte et de revente de données prises à l'insu des utilisateurs. Un exemple tout simple dans la vie de tous les jours, lors d'une connexion à un site internet, nous recevons des cookies, que souvent nous acceptons sans même lire de quoi ça retourne. Un cookies est un témoin de connexion, ces cookies sont là pour récupérer certaines données de type adresse mail, localisation géographique ... Toutes ces données sont ensuite revendues à des annonceurs pour avoir une pub beaucoup plus ciblée, une pub juste pour vous. C'est la raison pour laquelle lorsque l'on va acheter des chaussures sur un site comme Zalando ou Sarenza, lors d'une prochaine connexion, nous aurons des bandeaux avec les chaussures que l'on recherche. Le commerce de la vente de données est énorme et les sommes d'argent sont elles aussi colossales. C'est donc le but de ce dossier de montrer la dangerosité des réseaux, que vous pouvez être influencé sans le savoir.

En France des lois sont là pour éviter la surexposition médiatique au niveau de la télévision. Lors des campagnes présidentielles, il y a un chronométrage des propos liés à l'actualité électorale pour éviter une bipolarisation de la représentation de la vie politique. Pour résumer, c'est le chronométrage des différents candidats politiques, pour que tous aient le même temps de parole dans les médias télévisés, pour éviter des inégalités médiatiques.² Mais cette loi n'existe que dans les médias télévisés et elle n'existe pas sur internet comme sur Youtube ou sur les réseaux sociaux. Sachant que de nos jours les jeunes regardent de moins en moins la télé et de plus en plus les réseaux sociaux. Pour les candidats, ce sont les votes de jeunes qui sont les plus difficiles à avoir.

Pour toucher ces jeunes publiques, le président s'intéresse de plus en plus aux médias non conventionnels, de type réseaux sociaux ou chaîne Youtube. Comme par exemple l'interview Brut en décembre 2020, premier président à se faire interviewer par un média internet (ni journaux, ni télévisuel), média actif sur Youtube

¹ MANOKHA I. (2018), *Le scandale Cambridge Analytica contextualisé: le capital de plateforme, la surveillance et les données comme nouvelle « marchandise fictive »*

² DENIS ML. (2012), *La régulation audiovisuelle et l'élection présidentielle*

et les réseaux sociaux. Pour attirer un public encore plus jeune, le président E.MACRON s'intéresse aussi à des chaînes youtube qui ne sont pas des chaînes d'informations

mais de divertissement. Comme par exemple E. MACRON qui a posé un défi à Mcfly & Carlito (une chaîne youtube de plus de 6 millions d'abonnées) de faire un clip vidéo sur les gestes barrières et si cette vidéo faisait plus de 10 millions de vues, les deux youtubeurs venaient tourner à l'Elysée avec le président. Ce défi est en fait un énorme coup de communication puisqu'en posant ce défi E. MACRON va toucher un public jeunes qui sera en âge de voter en 2022, un public qui ne regarde pas la télévision et s'informe via Youtube ou les réseaux sociaux. Le défi a été réalisé par les deux youtubeurs et leur clip a atteint plus de 10 millions de vues, ils vont donc faire un concept phare de leur chaîne (un concours d'anecdote) avec le président E. MACRON. Ce qui va lui permettre d'attirer un public jeune et donc potentiellement des votes pour l'élection de 2022.

Cette loi sur la modération télévisuelle des candidats a été instaurée par le fait que notre cerveau va voter pour quelqu'un qu'on va le plus voir. Prenons un exemple, si toute la journée sans s'en rendre compte nous voyons le chiffre 7 un peu partout dans tous nos déplacements, si le soir une personne nous demande de choisir un chiffre, inconscient nous allons choisir le 7, notre cerveau s'est "formaté". C'est exactement ce qui s'est passé pour l'affaire Cambridge Analytica, pour les personnes indécises, des publications ont été créées pour eux. Si une personne se sentait en insécurité aux Etats-Unis, une publication s'affichait sur son mur Facebook, en disant que D. TRUMP allait augmenter le nombre de policiers dans les villes pour que les habitants se sentent en meilleure sécurité, ce qui poussait la personne indécise à voter TRUMP. Mais cette personne ne se rendait même pas compte que ces articles n'étaient pas un hasard, ils ont donc réussi à "rentrer dans le cerveau des gens".

C'est dans cet objectif que nous avons créé un questionnaire pour essayer de comprendre si certaines choses pouvaient expliquer le vote de certaines personnes. Savoir si les personnes votant un parti ont toutes les mêmes caractéristiques ou justement savoir si ces personnes sont diamétralement opposées.

Explication du questionnaire

Nous nous sommes donc intéressés aux partis politiques de nos sondés ce qui constitue notre variable dépendante à savoir pour quel parti il souhaite voter en 2022. Dans ce questionnaire, nous allons parler de parti et non pas de candidats, ces derniers ne sont pas encore présentés. Nous allons donc essayer d'expliquer les différents partis en fonction de plusieurs variables explicatives.

Notre variable dépendante va prendre les valeurs :

1. maire

La première valeur que prendra notre variable dépendant est le maire de la ville de résidence des sondées. Dans le questionnaire nous avons demandé l'orientation politique du maire. Nous allons donc le classer avec nos différents partis.

2. parent 1

Pour cette variable, le parti voté par le parent 1 du sondé

3. parent 2

Pour cette variable, le parti voté par le parent 2 du sondé

4. axe du parti politique

Nous allons en fonction des partis définir les principaux axes politiques et en fonction de ce qu'ont répondu les sondés, nous allons les classer dans un parti politique. Les axes que nous avons choisis sont le redressement économique, la réduction des inégalités socio-économiques, l'écologie, la sécurité, l'immigration, le revenu universel, la nouvelle république, l'augmentation du pouvoir d'achat et la souveraineté.

Pour expliquer cette variable dépendante, nous allons avoir 23 variables explicatives que nous allons expliquer. Ces variables explicatives sont liées à l'individu.

Première variable explicative : La confiance des sondés envers les partis politiques

Notre première variable explicative concerne la confiance que les sondées pouvaient avoir dans les différents partis politiques. Pour les personnes ayant répondu qu'ils ne se sentaient pas classer dans un parti politique, la réponse à cette question se fait dans l'hypothèse où ils se classaient dans un parti, elle peut aussi comprendre dans la confiance qu'ils donnent aux politiques en général.

Deuxième variable explicative : Nombre d'habitant

Notre deuxième variable explicative est le nombre d'habitant de la ville de résidence du sondés. Nous pensons que cette variable peut influencer sur le fait de se sentir en sécurité, les personnes habitant dans une grande ville vont peut être mettre la sécurité en axe principal.

Troisième à sixième variables explicatives : Les parents

Pour ces quatre variables explicatives, ce sont des questions qui concernent les parents de nos sondés. Nous leur avons demandé les études et la catégorie socio-professionnelle de chacun de leurs parents, pour les sondés dont les parents sont décédés, nous leur avons demandé de mettre la catégorie socio-professionnelle avant leur décès. Avec cette variable, nous allons voir si les études et catégorie socio-professionnelle influent sur le sondés. A savoir si un fils d'artisan ayant un CAP va faire des études ou va possiblement faire le même parcours professionnel que son père

Septième et huitième variable explicative : les types de médias

Nous avons demandé aux sondés quels sont les types de médias qui les informent tels que les journaux papiers, les journaux TV, les chaînes d'informations en continu ... Suite à cette question, nous leur avons demandé quel est le principal média qui les informe, cette question est une question ouverte, nous avons donc trié tous nos résultats pour faire des catégories. Le but de ces questions est de montrer que chaque médias ont des lignes éditoriales différentes et donc soit on adhère à cette ligne éditoriales ou pas. Par exemple une personne votant RN ne va pas aller lire un média du type libération qui est plus axé sur un axe politique de gauche voir d'extrême gauche.

Neuvième à quatorzième variable explicative : les questions du quizz de KOGAN

Pour les six variables explicatives suivantes, ce sont des échelles en fonction de différents traits de caractères. Les traits de caractères sont l'ouverture d'esprit, la satisfaction de la vie des sondés, l'impulsivité, l'extraversion, l'agréabilité et l'anxiété. Pour tous ces traits de caractères l'échelle allait de 1 à 10, 1 étant pas du tout et 10 étant tout à fait. Ces variables explicatives découlent du test de personnalité fait par

A. KOGAN lors de l'affaire Cambridge Analytica, ces questions étaient posées aux utilisateurs Facebook pour extraire leur données, ce qu'est qu'une partie de toutes les questions posées dans le test A. KOGAN.

Quinzième variable explicative : les publicités

Cette variable explicative concerne les différentes publicités que l'on peut voir sur les réseaux ou autres médias. Nous leur avons demandé s'ils se sentaient influencés par ces publicités. La réponse à cette question s'est faite via une échelle de 1 à 10, 1 étant je ne me sens pas du tout influencé par les pubs et 10 étant je me sens très influencé par les pubs. Cette variable va nous permettre de savoir s'ils se sentent manipulés par les pubs ce sera donc plus facile de les manipuler pour les élections présidentielles.

Seizième variable explicative : adhérent ou militant pour un parti

Cette variable explicative cherche à savoir si le sondé est adhérent ou militant pour un parti politique, celui qu'il soutient. Nous n'avons eu que 3 sondés étant adhérent ou militant pour un parti.

Dix-septième variable explicative : catégorie socio-professionnelle

La seizième variable explicative concerne la catégorie socio-professionnelle des sondés, ce qui va nous permettre de savoir si une catégorie socio-professionnelle est plus dirigée à voter pour un parti politique. Pour les catégories socio-professionnelles, nous avons pris la liste qui se trouve sur le site de l'INSEE à laquelle nous avons ajouté la catégorie étudiants car nous savons que beaucoup d'étudiants vont répondre au sondage.

Dix-huitième variable explicative : le revenu

Cette variable concerne les revenus des sondés, la réponse à cette question est sur les revenus nets par mois en euros. Nous avons donc créé des intervalles pour simplifier le questionnaire. Les intervalles sont : aucun revenu, moins de 500€, entre 500€ et 800€, entre 800€ et 1000€, entre 1500€ et 2000€, entre 2000€ et 3000€, entre 3000€ et 5000€ et plus de 5000€. Cette question va nous permettre

d'analyser si des tranches de revenus vote plus un parti politique. Pour vérifier si les clichés des revenus les plus élevés votent plus à droite.

Dix-neuvième et vingtième variable variables explicatives : le type d'habitation

Ces deux variables sont liées, elle concerne en type d'habitation. La 18^{eme} variable explicative concerne si les sondés sont propriétaire ou locataire, la 19^{eme} concerne elle le type d'habitation (maison, appartement, chambre crous ...). Une personne qui aurait répondu propriétaire dans le questionnaire est normalement d'un certain âge, peu d'étudiants sont propriétaire par un manque de moyen. Mais ce n'est pas parce qu'une personne est locataire toute sa vie qu'elle n'a pas de revenu, elle peut faire ça par choix.

Vingt-et-unième explicative : les études

La vingtième variable explicative concerne les études du sondé. Nous allons pouvoir l'analyser avec le revenu et donc le parti politique. Une personne ayant un bac +5 gagne logiquement mieux sa vie qu'une personne ayant un CAP, et sera peut être plus dirigée vers un parti de droite. Nous allons aussi vérifier les clichés qui existent en fonction du nombre d'études, à savoir un ouvrier qui vote à gauche et un cadre qui vote à droite.

Vingt deuxième variable explicative : le domaine d'étude

Cette variable explicative concerne le domaine d'étude du sondé, cette question était une question ouverte, à savoir, ils notaient leur domaine d'étude respectif, a la suite, nous avons fait des catégories pour regrouper les domaines d'études tel que économie, santé, sciences ... Le domaine d'étude va nous permettre d'analyser avec les questions de A. KOGAN, parce que une personne en psychologie (santé), philosophie va peut être plus ouverte d'esprit qu'une personne étant dans la finance.

Vingt-troisième et vingt-quatrième variables explicatives : âge et genre

Ces deux dernières variables explicatives concerne l'âge et le genre du sondés à savoir si c'est un homme ou une femme. L'âge est intéressant du point de vue politique parce que généralement on commence notre vie politique à gauche pour finir à droite. Ces deux variables sont aussi là pour déterminer notre population, savoir si elle est hétérogène ou homogène.

Nous avons également deux questions du questionnaire à titre d'information pour savoir si les sondés auront 18 ans ou plus en 2022 et savoir s'ils ont eu connaissance de l'affaire Cambridge analytica. En ce qui concerne l'âge en 2022, nous avons dû changer la question en cours de questionnaire, la question était : "En 2022, aurez-vous 18 ans ?" et nous l'avons changé pour "En 2022, aurez-vous plus de 18 ans". Ce changement est dû au fait que certaines personnes mettaient non la question de base parce qu'ils auront plus que 18 ans et pas 18 ans en 2022. Nous nous sommes rendu compte de ce biais en regardant l'âge des sondés.

Tableau 1 : Récapitulatif des Variables

Question	Variable non complet	Variable pour R
<u>Pour notre variable dépendante et ses 4 valeurs</u>		
En matière de politique, on classe habituellement les Français(es) par rapport à des partis politiques. Vous personnellement, quel est votre parti politique ?	Les partis politiques sondé	PP
Pour quel parti politique vote ou votait votre parent 1 (si décédé(e)) ?	Vote du parent 1	PP_Parent1
Pour quel parti politique vote ou votait votre parent 2 (si décédé(e)) ?	Vote du parent 2	PP_Parent2
Quel est l'axe principal politique de votre parti politique qui vous motive ?	L'axe principal du parti	PP_Axe
De quelle orientation politique est le maire de votre ville de votre résidence principale ?	orientation politique du maire	PP_Maire
<u>Pour nos variables explicatives</u>		
Etes-vous adhérent ou militant pour un parti politique ?	Adhérent ou militant	Adherent
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quel niveau de confiance avez-vous dans le parti politique que vous soutenez ?	confiance du parti	Confiance
Quel est le nombre d'habitants de la ville où se situe votre résidence principale ?	Le nombre d'habitants de la ville	Nbre_Hab
Quel est ou était le niveau d'étude de votre parent 1 (si décédé(e)) ?	Le niveau d'étude du parent 1	Etude_Parent1
Quelle est ou était la catégorie socioprofessionnelle de votre parent 1 (si décédé(e)) ?	Catégorie socio-professionnelle parent 1	Categorie_Parent1
Quel est ou était le niveau d'étude de votre parent 2 (si décédé(e)) ?	Niveau d'étude du parent 2	Etude_Parent2

Quelle est ou était la catégorie socioprofessionnelle de votre parent 2 (si décédé(e)) ?	Catégorie socio-professionnelle du parent 2	Categorie_Parent2
À travers quels types de média trouvez-vous l'information ?	Les médias qui informent	Media_Type
Quel est le principal média que vous utilisez pour vous informer (Le Monde, Le Figaro, TF1 ...) ?	Principal médias utilisé	Media_Princ
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quelle est votre ouverture d'esprit ?	Ouverture d'esprit du sondé	Ouverture
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quel est le niveau de satisfaction de votre vie ?	La satisfaction de la vie du sondé	Satisfaction
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quel est votre niveau d'impulsivité ?	Impulsivité du sondé	Impulsivite
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quel est votre niveau d'extraversion ?	Extraversion du sondé	Extraversion
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), quel est votre niveau d'agréabilité ?	Agréabilité du sondé	Agreabilite
Sur une échelle de 1 à 10 (1 étant très anxieux(se) et 10 étant pas du tout), quel est votre niveau d'anxiété ?	Anxiété du sondé	Anxiete
Sur une échelle de 1 à 10 (1 étant pas du tout et 10 étant tout à fait), vous sentez-vous influencé(e) par les pubs ?	Influence des pubs sur le sondé	Pub
Quelle est votre catégorie socioprofessionnelle ?	La catégorie socio-professionnelle du sondé	Categorie
Dans quelle tranche de revenu net appartenez-vous ? (en euros net par mois)	Revenu du Sondé	Revenu
Êtes-vous propriétaire ou locataire ?	Propriétaire ou locataire	Proprietaire
Quel est votre type d'habitation ?	Le type d'habitation du sondé	Type_Habitation

Quel est votre niveau d'étude ?	Le niveau d'étude du sondé	Etude_Nv
Quel est votre domaine d'étude ?	Domaine d'étude du sondé	Etude_Dom
Vous êtes ?	Le genre du sondé	Sexe
Actuellement, vous avez ?	L'âge du sondé	Age
En 2022, aurez-vous plus de 18 ans ?	18 et plus en 2022	Plus18
Avez-vous eu connaissance de l'affaire Cambridge Analytica ?	Connaissance de l'affaire	Cambridge

Pour pouvoir analyser nos variables avec le logiciel R, nous avons recodé nos variables pour remplacer les phrases par des chiffres, nous avons décidé de le faire directement via le tableur, car nous trouvons ça plus simple. Le tableau récapitulatif se situe en annexe (Annexe 1). Nous n'avons pas recodé nos variables dont les réponses se faisaient par une échelle.

Analyse statistique

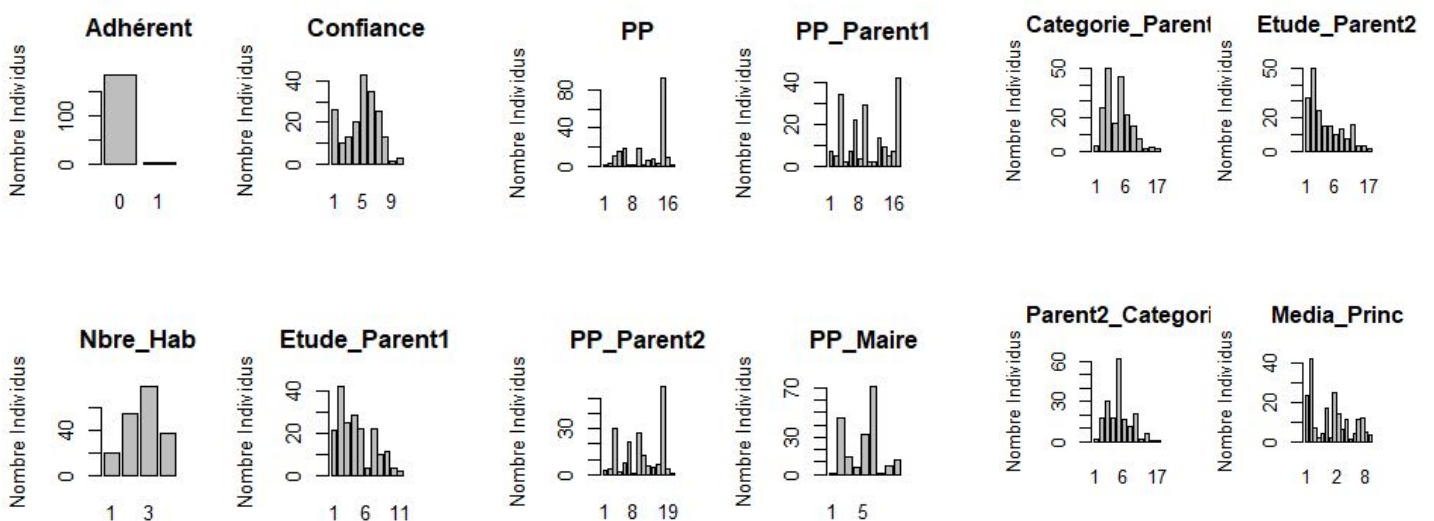
Pour les variables type de média et l'axe des partis politique, dans notre questionnaire, la question était à choix multiples, nous nous sommes donc retrouvés avec plusieurs réponses par sondés pour une même variable. Nous avons donc "split" ces deux variables, pour refaire une base de données où chaque réponse possible correspond à une colonne (une variable), ce qui nous donne des résultats binaire, avec si l'individu utilise ce média 1 lui sera affecté et 0 sinon.

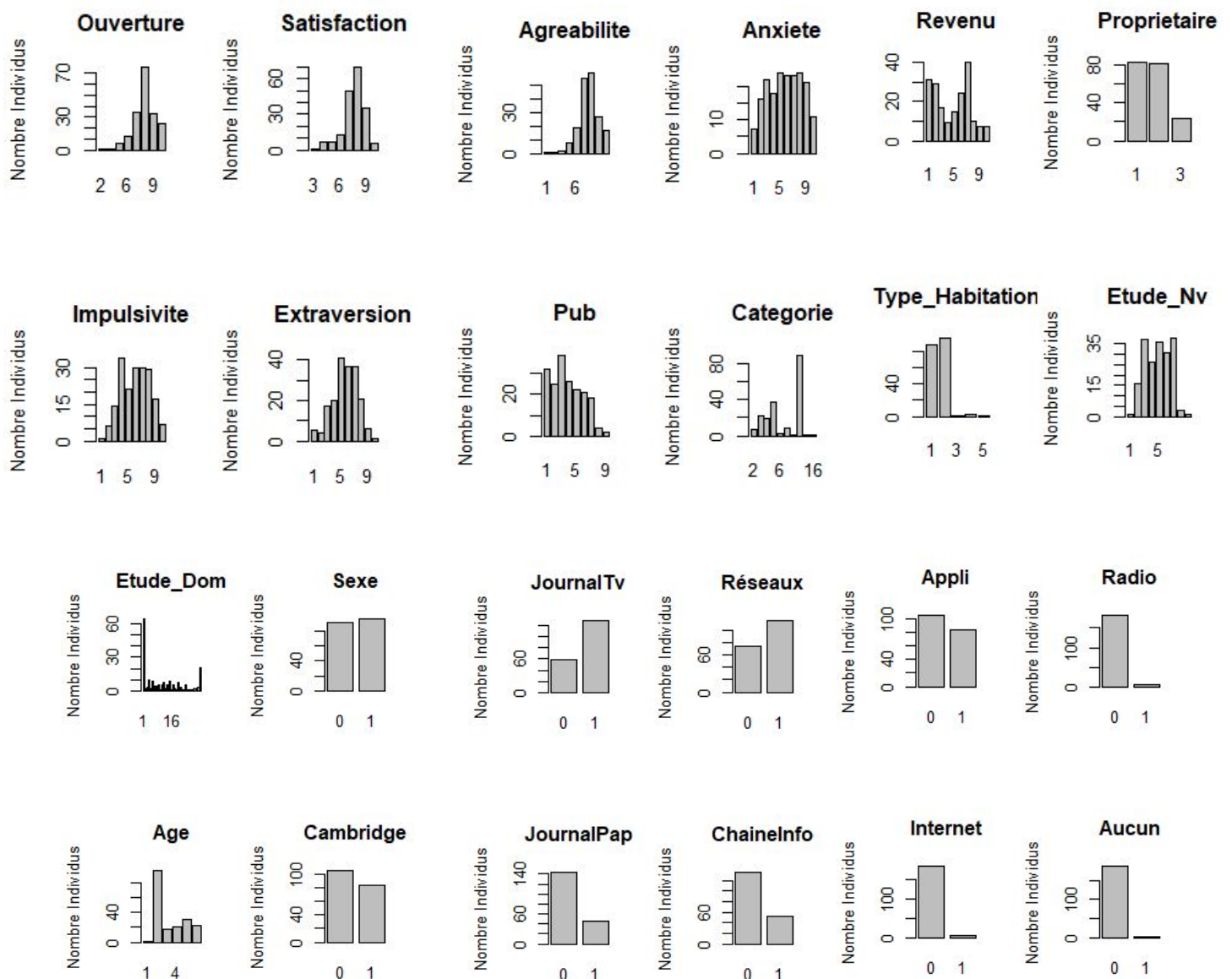
Dans un premier temps, nous allons regarder les distributions de nos variables à l'aide d'histogrammes. Dans notre base de données, nous n'avons que des variables qualitatives, nous ne pouvons donc pas retirer les individus aberrants, du fait que notre base de données est construite à l'aide d'un questionnaire, nous ne devrions pas avoir beaucoup d'individus aberrants et nous faisons confiance aux sondés. Nous allons donc analyser la distribution de nos variables.

Analyse Statistique de l'ensemble des données

Dans cette partie, nous allons analyser les histogrammes de l'ensemble de nos données.

Graphique 1 : Les histogrammes de toutes nos variables





Sur nos différents graphiques nous voyons les différentes modalités qui existent pour chaque variable. Nous avons des variables comme les variables ouverture satisfaction agréabilité et extraversion qui semblent avoir des distribution normale décalée vers la droite pour certaines. Pour d'autres variables, les réponses données sont hétérogènes, nous avons tout type de réponse et au contraire pour d'autre variables la population est homogène comme par exemple la variable pour savoir si les individus étaient adhérent à un parti politique ou non, nous avons 3 personnes adhérentes ou militantes sur 189 sujets. Au niveau des variables du type

de média, il y a aussi un déséquilibre sur les variables radio, internet et aucun média. Sur les autres variables binaires, nous avons plutôt une distribution égale, le fait d'utiliser les applications pour s'informer est assez équilibré entre ceux qui les utilisent et ceux qui ne les utilisent pas.

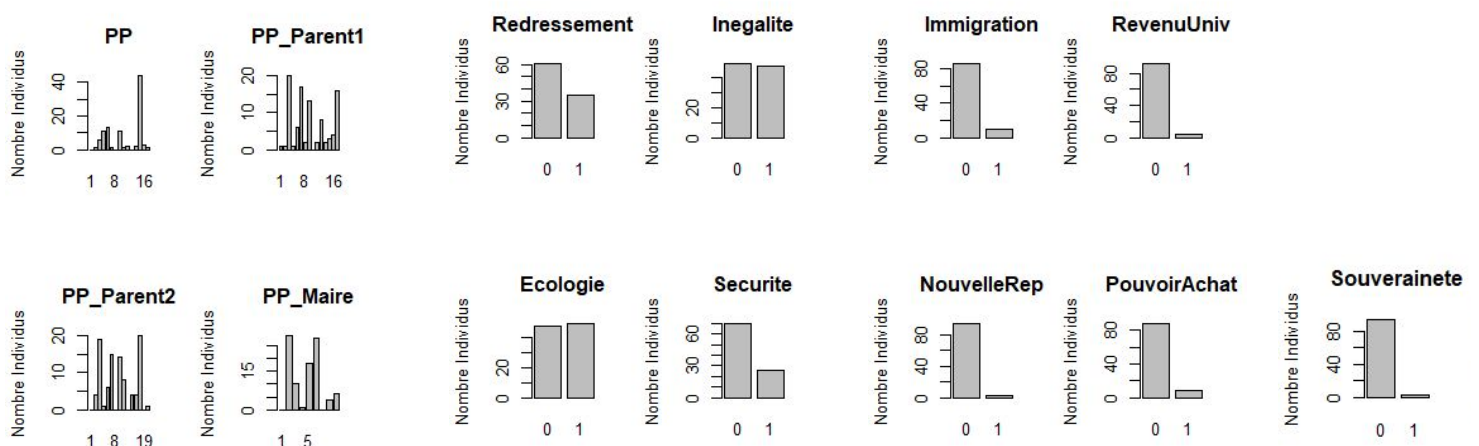
A titre d'information, nous voyons qu'il y a plus de la moitié des sondés qui ne connaissent pas l'affaire Cambridge Analytica.

Dans l'ensemble, nous avons plutôt un échantillon hétérogène c'est donc un point positif pour l'analyse économétrique qui va suivre. Mais nous avons cependant un grand nombre de sondés dans la tranche d'âge 18-25 ans, ce qui pourrait être un désavantage, nous avons eu plus de mal à toucher des populations plus âgées. Les populations plus âgées que nous avons touchées sont grâce à nos parents respectifs et leurs amis. Nous voyons les limites de la liste de diffusion de l'IAE, bien qu'elle soit bénéfique dû au nombre de réponses que nous avons reçues, la population qui y répond est jeune et habite dans les alentours de Nantes.

Analyse statistique par rapport à l'âge des sondés

Nous allons maintenant regarder notre variable dépendante et ses valeurs en fonction de l'âge et du revenu du sondé, ce qui va permettre de voir si une population en fonction de son âge ou de son revenu va plus soutenir un parti politique.

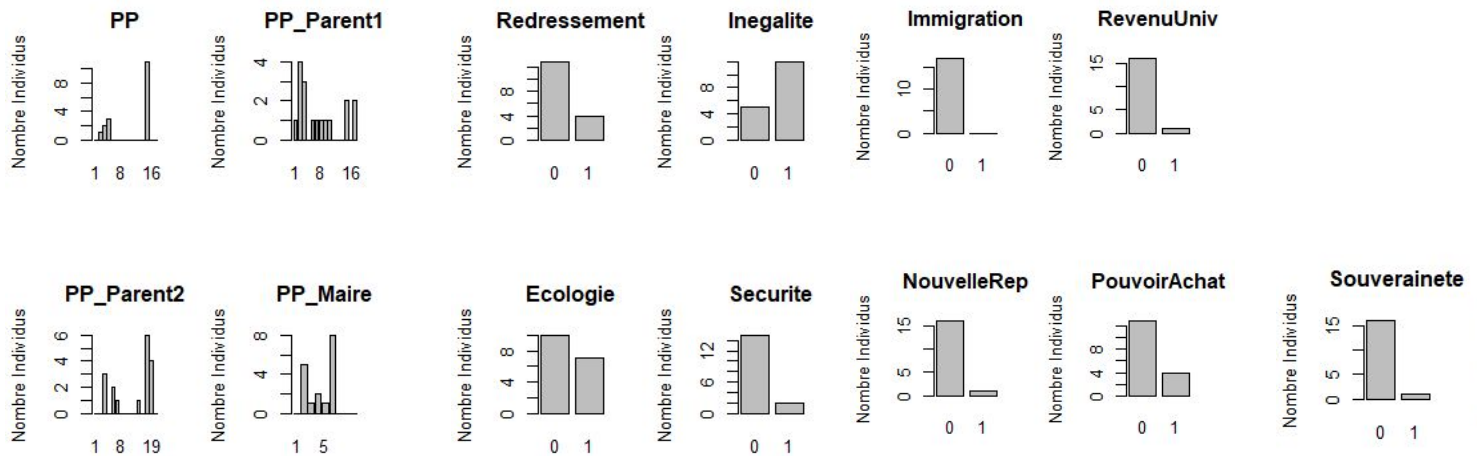
Graphique 2 : Les valeurs de la variable dépendante en fonction de l'âge, tranche 18-25 ans



Pour

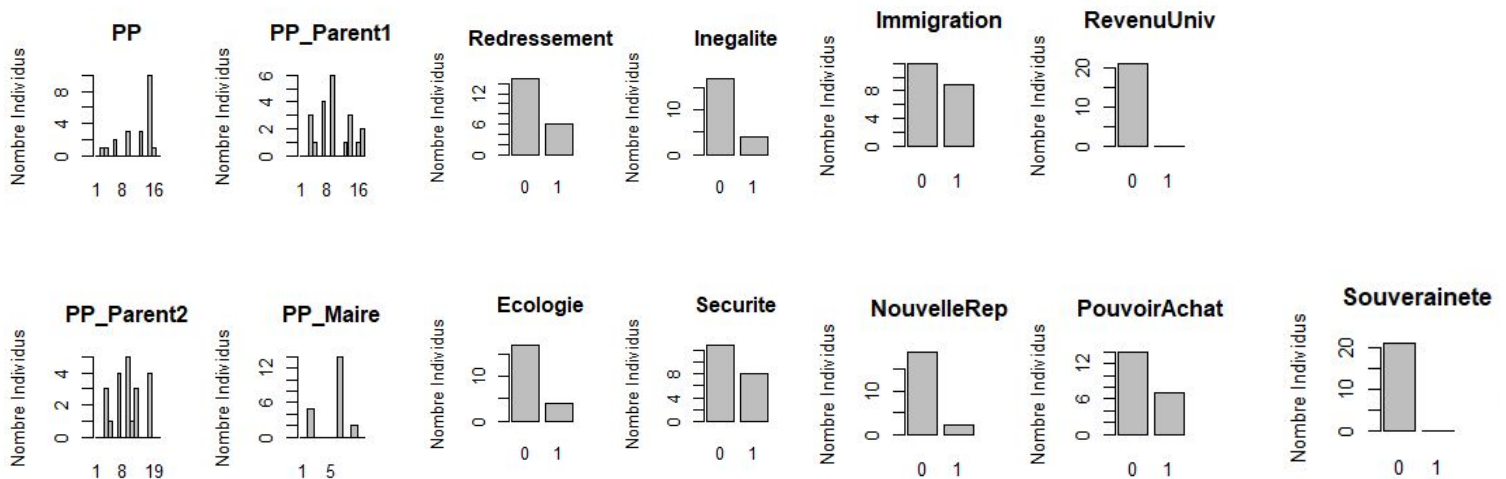
les variables binaires, nous voyons que ce qui n'intéresse pas les jeunes sont le pouvoir d'achat, le revenu universel et la souveraineté, ils sont plus intéressés par l'écologie, réduire les inégalités, le redressement économique et pour une partie d'entre eux la sécurité est aussi un point à améliorer. Mais il n'y a pas d'axes politiques où ils sont tous d'accord certains axes se détachent sans faire l'unanimité. Pour les variables non binaires, nous voyons une grande barre pour le graphique sur les partis politiques, il s'agit des personnes qui se classent dans aucun parti. Nous pouvons donc dire que les jeunes sont perdus en politique et qu'ils ne savent pas ou voter. Il n'y a pas de parti qui les intéresse plus que ça. La distribution des parents des sondés est beaucoup plus hétérogène, ce sont des parents qui se classent et donc qui votent pour un parti. Au niveau du maire des sondés, entre les maires de droite et de gauche c'est assez équilibré.

Graphique 3 : Les valeurs de la variable dépendante en fonction de l'âge, tranche 26-35 ans



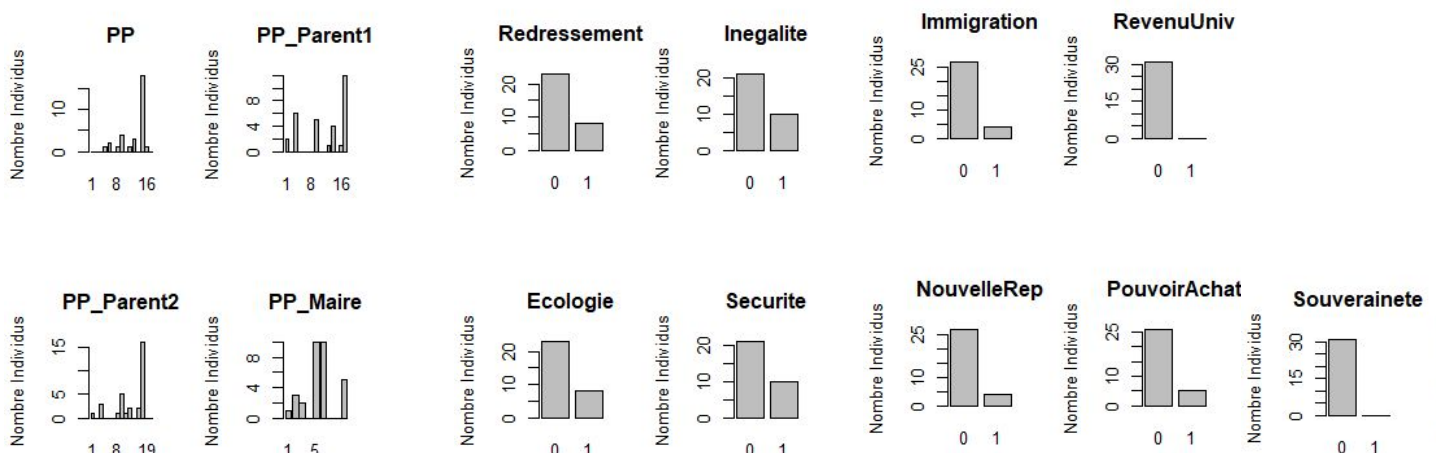
C'est au niveau de cette population que nous avons le moins de sondés. Nous voyons que sur cette population, il se passe le même phénomène que pour la tranche d'âge au-dessus, c'est-à-dire que beaucoup d'entre eux se classent dans aucun parti et c'est même encore plus radical pour cette tranche d'âge. Concernant les parents des sondés, nous avons également une distribution plus hétérogène, avec des parents qui se classent dans un parti, et pour cette tranche d'âge les parents se classent dans un parti politique de gauche ou du centre, des barres entre 1 et 8 nous l'indiquent. Leur intérêt pour les axes politiques est similaire à celui de la classe d'âge 18-24 ans si ce n'est que certains d'entre eux sont pour une augmentation du pouvoir d'achat, cela s'explique qu'à partir de cette âge on rentre sur le marché du travail et on se débrouille tout seul pour toutes les dépenses. Ils sont également plus sensibles à la sécurité et moins sensibles à l'écologie.

Graphique 4 : Les valeurs de la variable dépendante en fonction de l'âge. tranche 36-45 ans



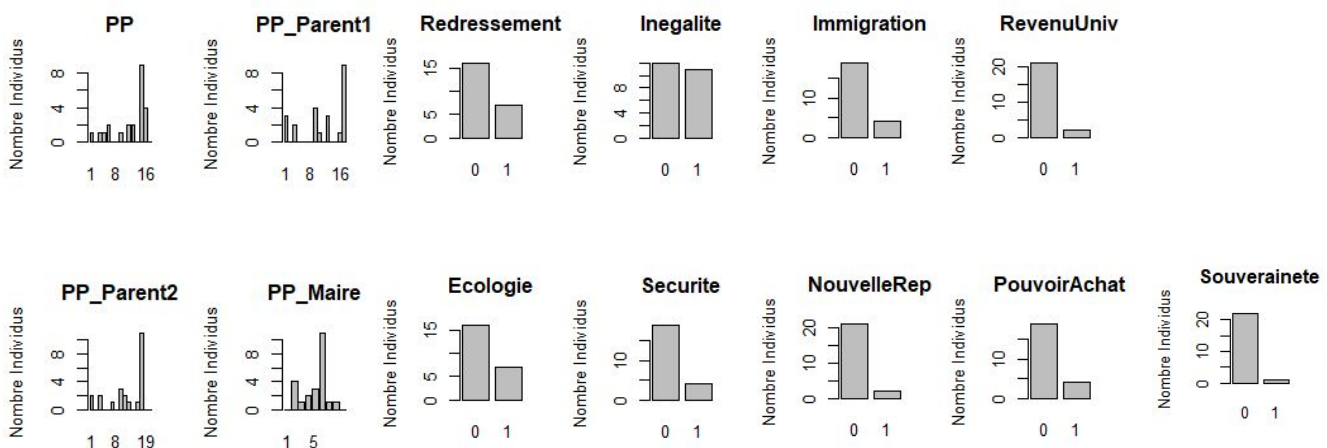
Nous voyons encore une fois que beaucoup d'individus ne se classent dans aucun parti politique. Avec les classe d'âge que nous avons déjà analysé, nous voyons que beaucoup de nos sondés ne se retrouvent dans aucun parti politique. Au niveau des parents de ces sondés, la population est plus hétérogène que pour les autres tranches d'âge. Concernant l'orientation du maire de leur ville, la distribution est ici plus homogène que les autres sondés. Pour les axes politiques de cette tranche, les idées ne sont pas les mêmes que pour les autres tranches d'âges, l'écologie est beaucoup moins présente mais le pouvoir d'achat est plus présent, concernant l'immigration, plus de sondés sont sensibles à ce sujet et pour les inégalités, cette population est moins sensible que les deux autres.

Graphique 4 : Les valeurs de la variable dépendante en fonction de l'âge. tranche 46-55 ans



Il se passe encore une fois le même phénomène que pour les tranches d'âge précédent, que de nombreux sondés ne se classent dans aucun parti politique. Pour les parents les résultats sont différents, beaucoup d'entre eux ne savent pas dans quel parti politique se classe ses parents. Ce phénomène peut s'expliquer par le fait que cette population commence à être âgée, leur parents sont peut être décédés, ces individus ne savaient pas pour quel parti votaient leurs parents. A cette époque il y avait moins de partis politiques et ce n'était pas forcément le nom que le parti porte maintenant. Les principaux axes que soutiennent ces sondés concernent la baisse des inégalités, la sécurité et l'écologie.

Graphique 5 : Les valeurs de la variable dépendante en fonction de l'âge, tranche plus de 56 ans



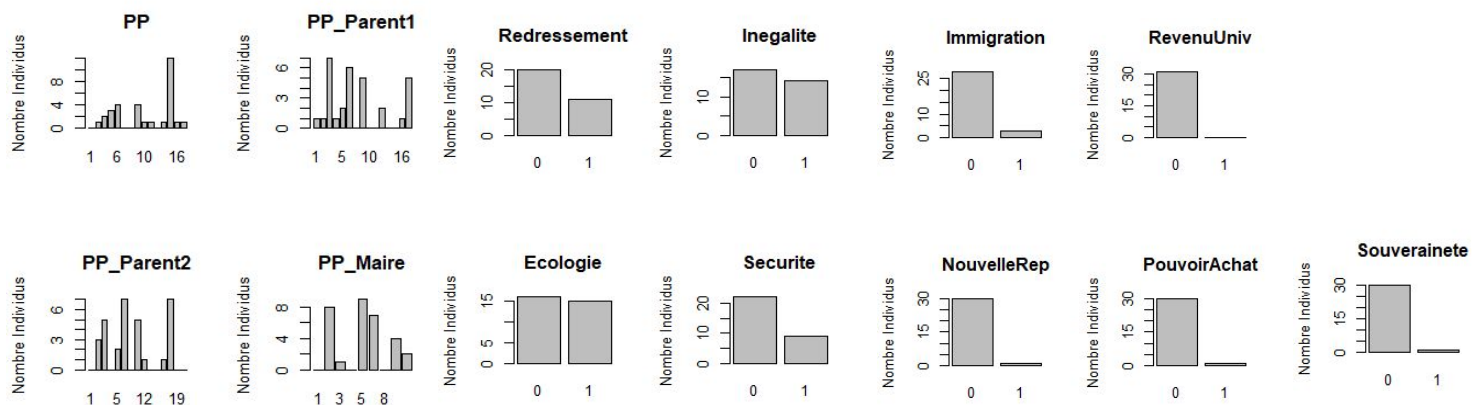
Sur cette dernière tranche d'âge, nous voyons qu'il y a plus de personne qui ne souhaite pas dire dans quel parti politique il se classe, il y a encore également un grand nombre de sondés qui se classe dans aucun parti. Nous voyons le même phénomène pour les parents des sondés, ils ne savent pas pour qui ils votaient. Pour ces personnes, la réduction des inégalités arrive en axe principal, ils sont ensuite attirés par l'écologie et le redressement économique.

Pour conclure sur cette partie, nous voyons que de nombreux sondés ne se classent dans aucun parti politique, cela représente presque 50% de notre sondage. Lors du Cambridge Analytica, ce sont ces personnes qui étaient dans l'hésitation qui étaient influencées pour leur vote. Pour ce faire, dans notre exemple, nous pourrions prendre les principaux axes de chaque population, mettre un candidat regroupant tous ces axes et nous aurions des chances que celui-ci soit élu en faisant des publications Facebook ou des pubs pour lui. Voilà comment vulgairement nous pouvons résumer l'affaire Cambridge Analytica. De plus pour toutes nos populations se sont souvent les mêmes axes principaux, du type écologie, réduire les inégalités, augmenter le pouvoir d'achat. Nous voyons autre chose également le fait qu'avec l'avancée de l'âge de nos sondés, les votes des parents sont inconnus, nous pouvons donc supposer qu'avant on ne parlait pas politique en famille, aujourd'hui beaucoup de jeunes savent ce que leur parents votent et il n'y a pas beaucoup de chances pour que leur vote change avec les années. Au vu des résultats nous pouvons dire que dans nos sondés nous avons très peu d'idées extrêmes que ce soit extrême droite ou extrême gauche. Pour toutes les tranches d'âge, nous en avons aucune qui prône une nouvelle république ou une souveraineté ou encore un revenu universel, ces idées sont plutôt orientées pour les partis politiques d'extrêmes.

Analyse Statistique par rapport aux revenus des sondés

Dans cette partie sera étudiée graphiquement la répartition des individus en fonction des revenus afin de remarquer s'il y a une divergence en fonction de ces derniers.

Graphique 6 : Histogramme de la variable dépendante et ses valeurs pour les sondés ayant aucun revenu

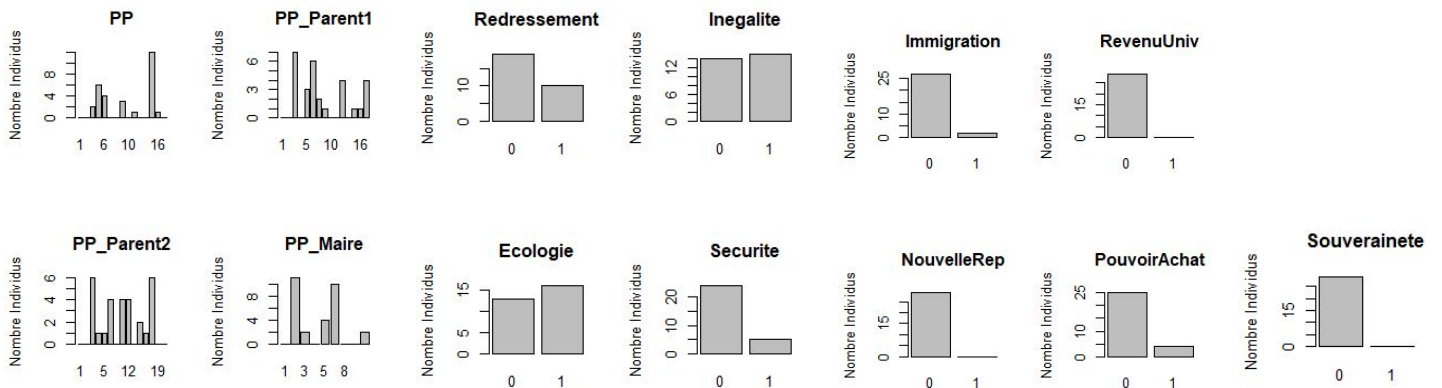


Cette sous-population est constituée de 31 sondés.

Il ressort à partir de ces différents graphiques (graphique 6) que les personnes n'ayant pas de revenu, personne étant étudiante, se considèrent principalement comme n'ayant aucune affinité avec aucun parti et deux autres sous groupes se distinguent : les républicains et la république en marche. Concernant le parti politique de leurs parents ou le parti politique de leur maire, la population est assez hétérogène et rien ne ressort réellement. Cependant pour les axes politiques considérés comme importants à leur yeux, les quatre axes principaux sont, dans l'ordre croissant, l'écologie, la diminution des inégalités socio-économiques, le redressement économique et la sécurité.

Le graphique suivant présente les mêmes variables mais pour la population ayant un revenu inférieur à 500 euros.

Graphique 7 : Histogramme de la variable dépendante et ses valeurs pour les sondés ayant un revenu inférieur à 500 euros

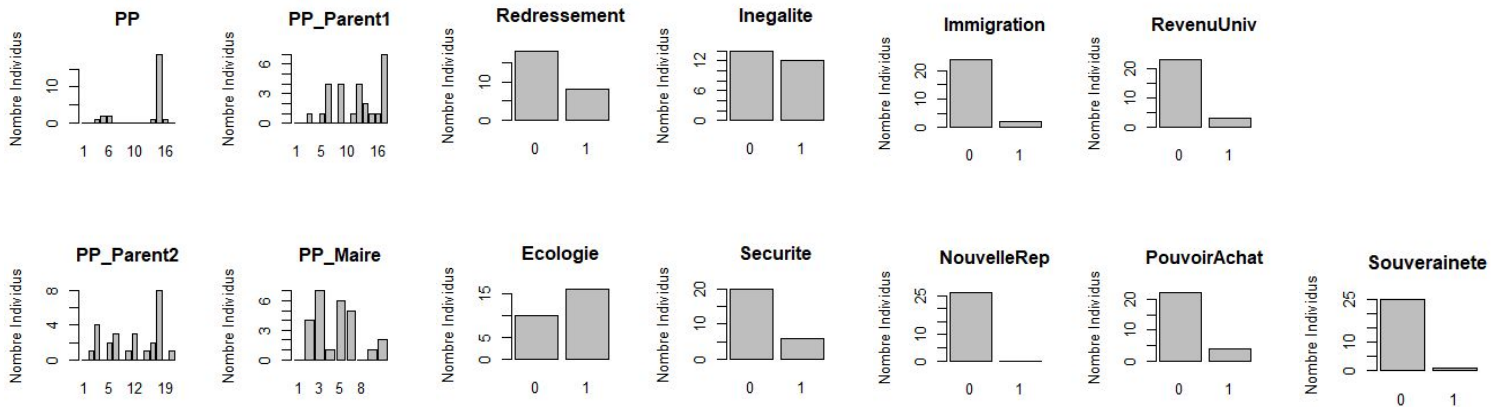


Cette sous-population est constituée de 29 sondés.

D'après les différents graphiques observés ci-dessus, les sondés se considèrent principalement n'étant adepte d'aucun partis politiques, même si la "divers gauche" et "LREM" recueillent quelques votes. Pour les variables dépendantes, parti politique des deux parents, la population est hétérogène alors que pour le parti politique du maire il en ressort deux mouvements : "gauche" et "droite". Les principaux axes politiques qui intéressent cette catégorie de population sont, dans l'ordre croissant, l'écologie, la diminution des inégalités puis le redressement économique.

Dans le graphique qui suit, il est présenté les mêmes variables que précédemment mais pour les sondés ayant un revenu compris entre 500 et 1000 euros (les catégories 3 et 4 ont été regroupés puisque un très petit nombre de sondés appartenant à la quatrième catégorie).

Graphique 8 : Histogramme de la variable dépendante et ses valeurs pour les sondés ayant un revenu compris entre 500 et 1000 euros

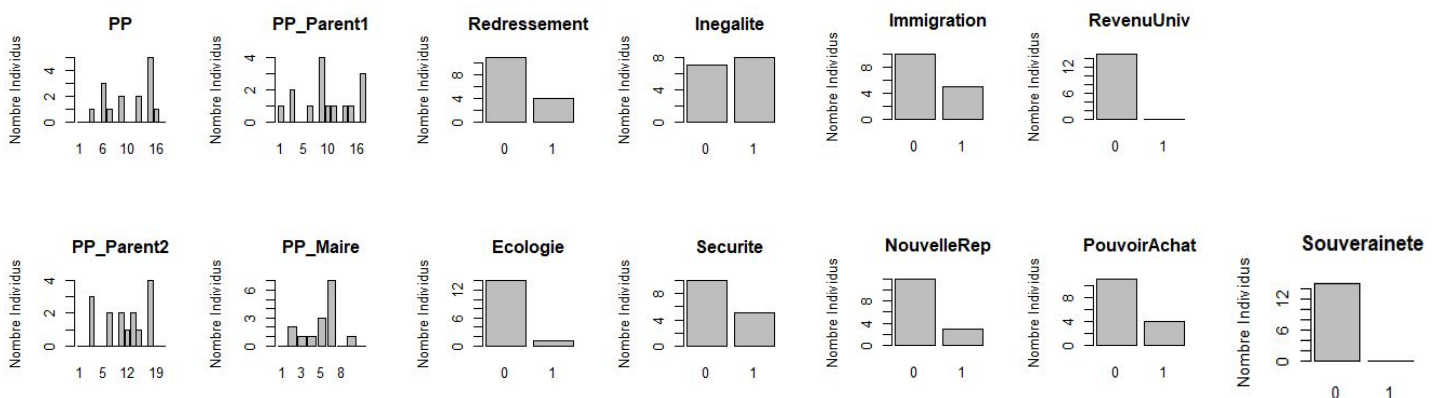


Cette population est constituée de 26 sondés.

Suite au graphique, il en ressort que cette catégorie de sondés est quasi uniquement constituée de personnes n'ayant aucune affinité avec aucun parti politique, alors que le parti politique du maire de leur ville est hétérogène. Malgré un nombre important de "je ne sais pas", pour les partis politiques des parents, le reste des partis politiques est plutôt hétérogène. Pour les axes principaux considérés comme importants à leurs yeux, un petit nombre se détache : l'écologie, la diminution des inégalités socio-économiques, le redressement économique et la sécurité.

Le graphique suivant présente nos variables pour les sondés ayant un salaire compris entre 1000 et 1500 euros.

Graphique 9 : Histogramme de la variable dépendante et de ses valeurs pour les sondés ayant un revenu compris entre 1000 et 1500 euros

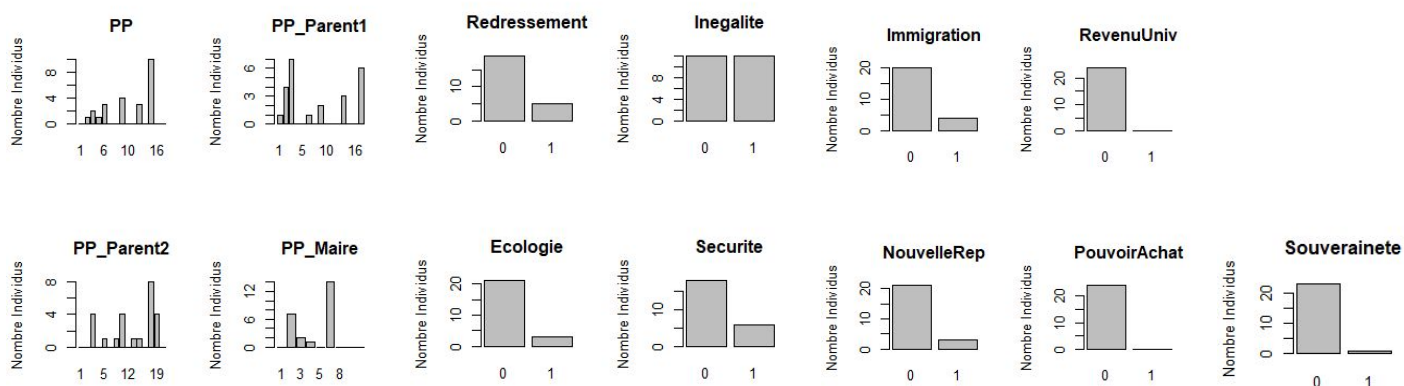


Cette population est constituée de 15 sondés.

Avec ce graphique, il est permis de voir que les personnes étant dans cette tranche de revenus sont sensibles principalement aux inégalités, à l'immigration ou à la sécurité. Un parti politique se détache si on excepte les personnes se classant dans aucun parti, "divers gauche", se sont donc des personnes qui vont voter pour un candidat de gauche lors des prochaines élections. Pour les votes des parents, du parti respectif se détache, pour le parent 1 il s'agit de "Les républicains, et pour le parent 2 il s'agit du "parti Socialiste"; Pour l'orientation politique du maire, nous avons des villes ayant des maires de droite.

Le graphique suivant présente nos variables pour les sondés ayant un salaire compris entre 1500 et 2000 euros.

Graphique 10 : Histogramme de la variable dépendante et de ses valeurs pour les sondés ayant un revenu compris entre 1500 et 2000 euros

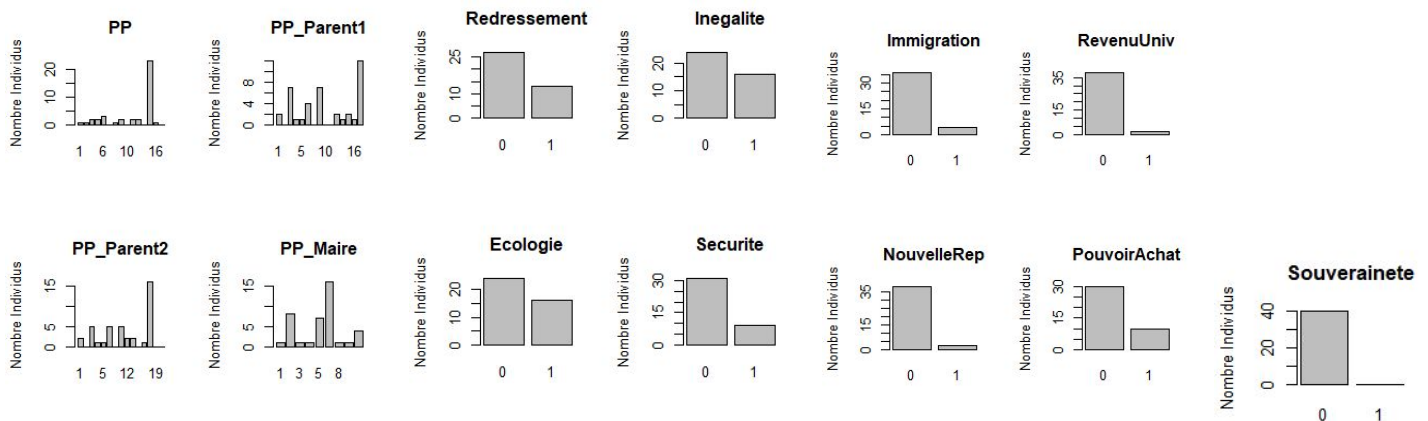


Cette population est constituée de 24 sondés.

Ce graphique nous permet d'analyser que pour cette tranche de revenu, les principaux axes, dans l'ordre croissant sont les inégalités, l'immigration et la sécurité. Dans cette tranche de revenu, nous avons des personnes votant dans des partis d'extrêmes, puisque l'axe d'une nouvelle république est représenté. Pour l'orientation du maire, celle-ci est toujours à droite. Pour le parti politique des parents, le parent 1 semble plus voter pour un parti de gauche, et pour le parent 2, les sondés ne savent pas pour quel parti il vote.

Le graphique suivant présente nos variables pour les sondés ayant un salaire compris entre 2000 et 3000 euros.

Graphique 11 : Histogramme de la variable dépendante et de ses valeurs pour les sondés ayant un revenu compris entre 2000 et 3000 euros

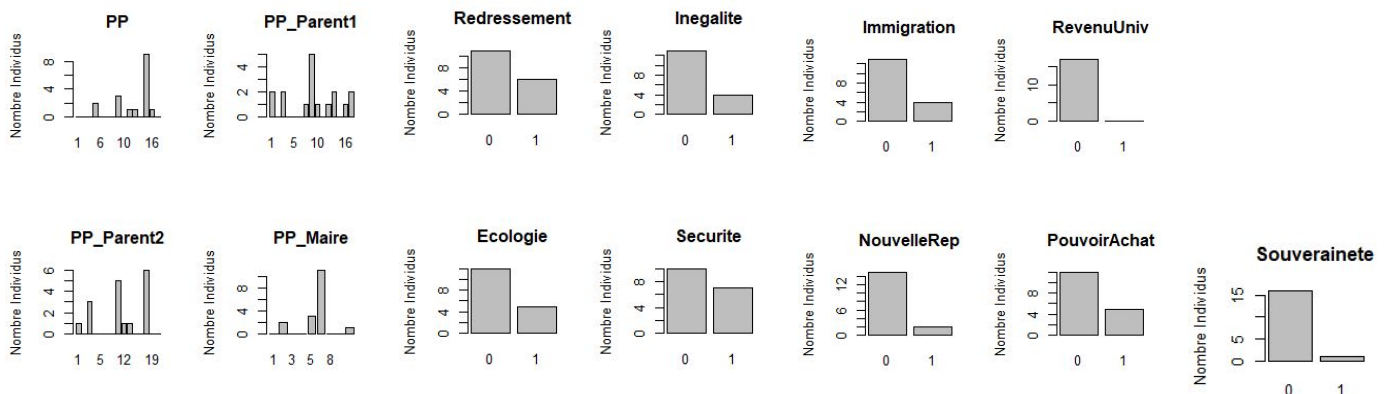


Cette population est constituée de 40 sondés.

c'est sur cette tranche de revenu, que nous avons le plus de sondés. Aucun parti ne se détache, les sondés ne se classent dans aucun parti politique. Pour l'orientation du maire, ce sont principalement des maires de droite même s'il y a quelques maires de gauche. Pour les parents des sondés, une grande partie d'entre eux ne savent pas pour qui ils votent. Les principaux axes de cette tranche de revenu sont les inégalités, l'écologie, le pouvoir d'achat et la sécurité.

Le graphique suivant présente nos variables pour les sondés ayant un salaire supérieur à 3000 euros (les catégories 8 et 9 sont regroupés car leur nombre est faible).

Graphique 12 : Histogramme de la variable dépendante et de ses valeurs pour les sondés ayant un revenu supérieur à 3000 euros



Cette population est constituée de 17 sondés.

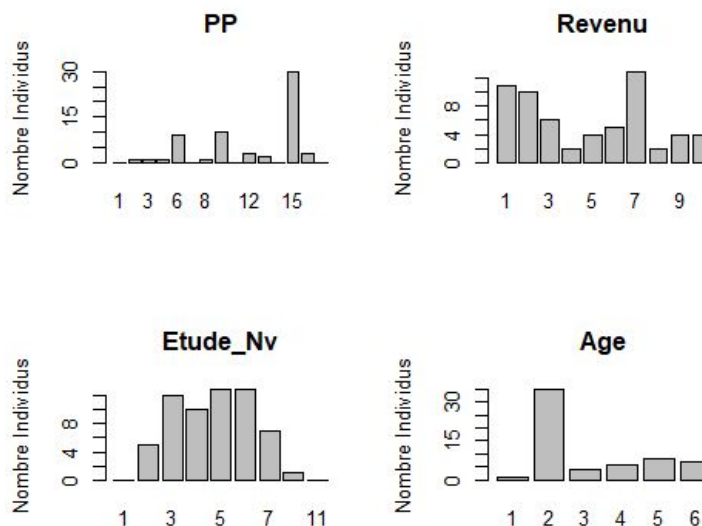
Pour cette tranche de revenu, nous avons réuni la tranche entre 3000 et 5000 euros et plus de 5000 euros. Il y a encore une fois aucun parti politique qui se détache, beaucoup de ces sondés ne se classent pas. L'orientation du maire est de droite. Dans l'ordre croissant, les principaux axes sont la sécurité, le redressement économique, l'écologie, les inégalités et l'immigration. Pour cette tranche de revenu, nous avons plus d'axes principaux que pour les autres tranches. Pour les parents de ces sondés, se sont des parents votant principalement à droite sur l'échiquier politique.

Pour conclure sur les tranches de revenus, nous pouvons dire que les principaux axes politiques sont souvent les mêmes, les inégalités, la sécurité et l'écologie. Certains axes politiques ne sont jamais les principaux, comme la nouvelle république ou la souveraineté. les idées reçues comme quoi les personnes ayant les plus gros revenus votent à droite n'est pas vérifié par notre sondage, nous voyons que les personnes ayant les plus gros revenus ne se classent dans aucun parti politique, alors que les personnes ayant des revenus les plus faibles eux se classent plus dans un parti. Pour nos sondés, se sont souvent des maires de droite ou sinon de gauche. Les parents de nos sondés, eux se classent plus dans un parti politique, se sont souvent les partis politiques à la base de la cinquième république.

Analyse statistique par rapport aux axes politiques des sondés

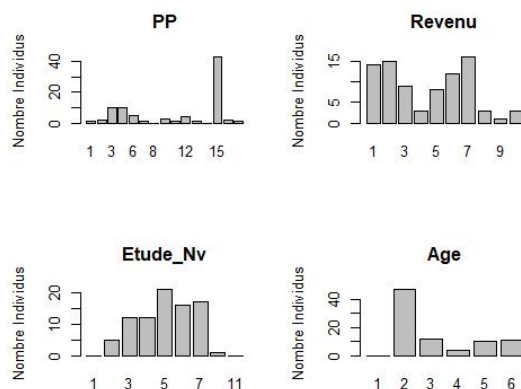
Dans cette partie, nous allons analyser les parti politique du revenu du niveau d'étude et de l'âge en fonction des différents axes politiques

Graphique 13 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction du redressement économique



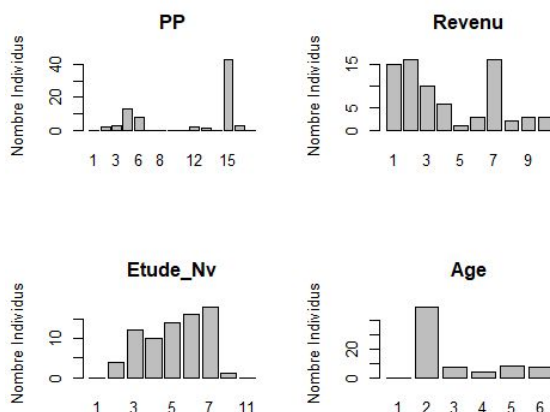
Sur ces différents histogrammes, nous voyons que ce sont les populations les plus jeunes qui se sentent les plus intéresser au redressement économique mais c'est aussi la population dont nous avons le plus de réponse, il y a peut être un biais. Concernant le niveau d'étude se sont plutôt des sondés qui ont fait ou qui font des études longues, après le BAC. Ce sont les personnes qui gagnent entre 2000 et 3000 € par mois qui se sentent le plus concernés par cet axe, ce qui rejoint le nombre d'études pour ceux qui les ont déjà finis. Nous voyons qu'il y a un autre biais avec le parti politique, nous avons la moitié qui ne classe dans aucun parti, il y a donc un risque qu'on obtienne un graphique similaire pour chaque axe politique.

Graphique 14 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction des inégalités



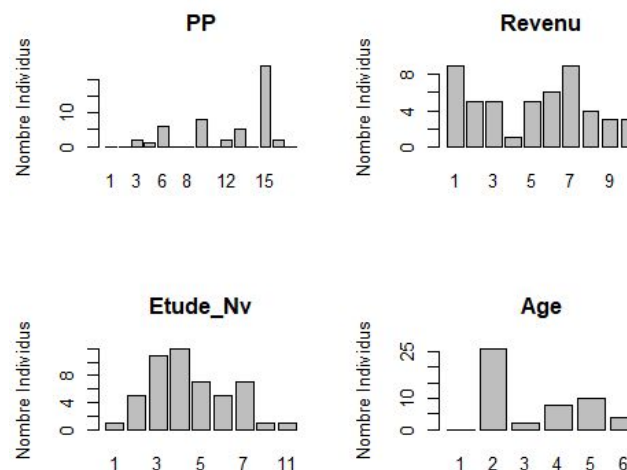
En ce qui concerne les inégalités, nous voyons bien nos deux biais sur les graphiques, au niveau du parti politique et au niveau de l'âge des sondés. Si on ne regarde pas les sondés se classant dans aucun parti, nous voyons que ce sont des gens se classent plus sur un partie de gauche que correspond les inégalités. Au niveau des études des sondés, nous retrouvons des personnes faisant des études supérieures également. Pour le revenu des sondés, les tranches de revenus sont les plus basses et ceux gagnant entre 2000 et 3000€. Ce gap entre les tranches de revenus s'explique par le fait que c'est dans la tranche de 2000 et 3000€ que nous avons le plus de sondés.

Graphique 15 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction de l'écologie



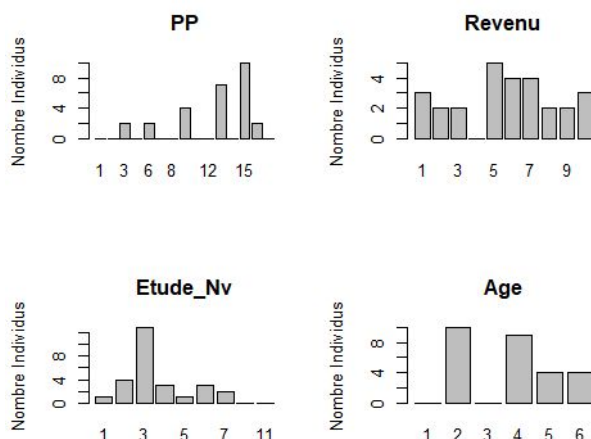
Nous allons maintenant voir l'axe concernant l'écologie, en comparaison aux deux graphiques ci-dessus, nous voyons que pour la variable des revenus, ce sont les personnes ayant les plus bas revenus du sondage qui se sentent les plus intéressés par cet aspect. Au niveau de l'âge nous retrouvons toujours le même biais, avec une population pas assez homogène. C'est également les sondés se classant dans un parti politique de gauche que l'écologie est présente, ce qui est logique sachant que le parti écologiste EELV se classe pour sur la gauche de l'échiquier politique.

Graphique 16 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction de la sécurité



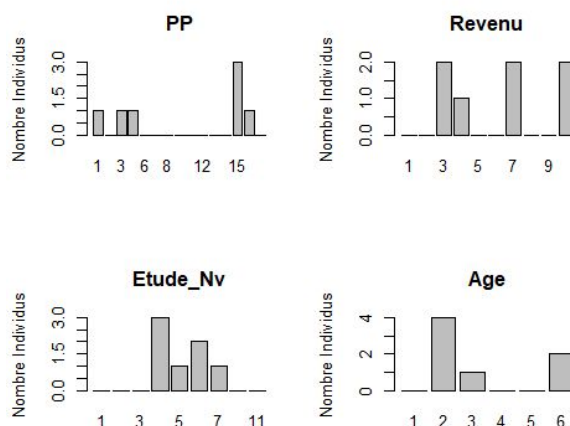
Au niveau de l'axe concernant la sécurité nous retrouvons des résultats similaires aux axes au-dessus. Cela s'explique par le fait que notre nombre d'individus sondés n'est pas assez grand. Il y a quand même certaines différences au niveau de la variable revenu, nous voyons que les revenus les plus sensibles à la sécurité sont les plus bas revenus ainsi que les revenus qui se trouvent dans la moitié de notre sondage. Les plus bas revenus restent les plus sensibles à cet axe, cela peut se comprendre par le fait que dans cette tranche de revenu se sont de nombreux étudiants ce qui signifie des sorties la nuit et nous savons que si nous prenons la ville de Nantes, en pleine nuit il y a beaucoup d'insécurité. Nous le voyons également avec le niveau d'étude, puisque ce sont les personnes en études supérieures qui se sentent le plus concernées par la sécurité.

Graphique 17 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction de l'immigration



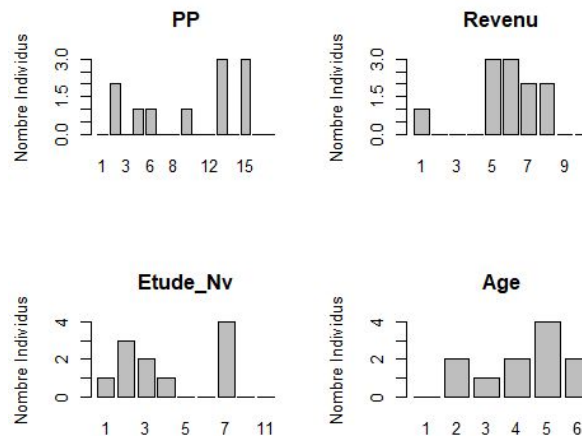
Concernant la variable de l'immigration, nous voyons qu'il y a une tranche de revenu qui ne considère pas du tout l'immigration comme son axe principale, il s'agit de la tranche 800-1000 €. En revanche la tranche la plus investie dans l'immigration est pour les revenus étant entre 1000 et 1500 €. Ce sont également les 18-25 ans qui se sentent le plus concernés, ce qui n'est pas logique puisque nos sondés (beaucoup d'étudiants) ne touchent pas entre 800 et 1000 € par mois. Pour les partis politiques, nous avons toujours notre biais des personnes qui ne se classent dans aucun parti. Mais au-delà de ce biais, ce sont logiquement les gens se classant dans des partis d'extrêmes droites ou de droite. Il se peut que cette question soit mal comprise du fait que nous avons aussi des gens étant à gauche qui sont sensibles à l'immigration. Mais il est possible que ces deux groupes de personnes n'aient pas la même idée de l'immigration, un premier groupe plutôt contre et le deuxième plutôt pour. Ce sont les personnes les plus âgées de notre sondage qui sont sensibles à cet axe.

Graphique 18 : histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction du revenu universel



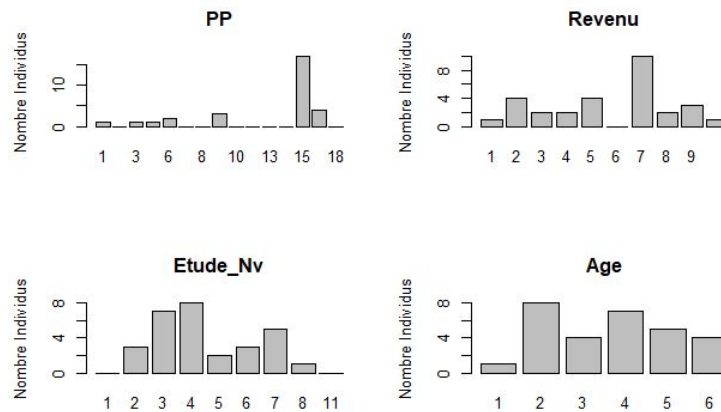
Nous pouvons voir que les personnes étant pour le revenu universel sont les personnes étant à gauche ou extrême gauche sur l'échiquier politique, ce qui semble logique puisque cet axe est plus celui des partis de gauche. Au niveau des revenus, nous voyons également que les gros revenus (plus de 5000€) se classent dans ce partie, ce qui va à l'encontre des clichés comme quoi les personnes les plus riches votent plus pour un parti de droite.. . Sur l'âge de sondés, une tranche qui ne se sentait pas du tout concernée par l'immigration, se sent concernée par le revenu universel, la tranche des 26-35 ans. Au niveau des études, nous sommes sur des personnes qui sont ou étaient en études supérieurs au-delà du BAC. Pour finir, au niveau de l'âge, si on excepte la tranche d'âge 18-25, nous voyons que les personnes de plus de 56 ans arrivent comme étant les plus concernés par cet axe.

Graphique 19: histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction de la nouvelle république



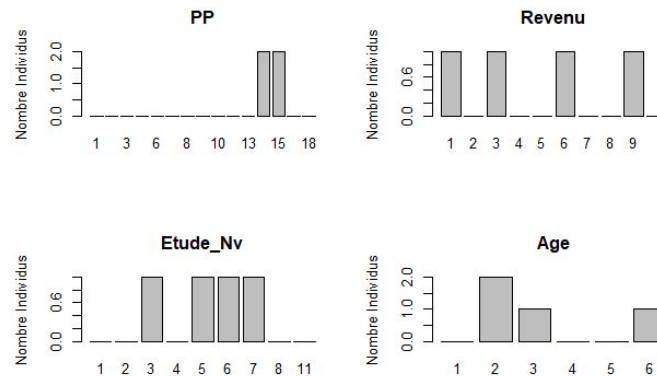
Pour cet axe politique, nous sommes sur une idée d'extrême gauche. Cela correspond au premier pic dans l'histogramme des partis politiques. Mais nous avons un pic pour le parti d'extrême droite RN pour cet axe. Nous sommes donc amenés à nous demander pourquoi, du fait que ce n'est pas du tout une idée d'extrême droite. Nous pouvons supposer que souvent les personnes votant pour les partis d'extrême ont marre des politiques actuelles et veulent du changement, c'est peut être l'hypothèse qui explique ce pic. Nous sommes sur des personnes ayant fait de longues études pour la majorité au moins BAC+5. Et nous voyons malgré notre nombre important de sondés ayant entre 18 et 25 ans, que se sont les personnes au-delà de 45 ans qui prônent cette idée. Nous pouvons constater que les jeunes ne sont pas du tout pour une nouvelle république. Pour les revenus ce sont dans les tranches du milieu de notre sondage.

Graphique 20: histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction du pouvoir d'achat



Cet axe politique concerne tous les partis, de nombreuses personnes de notre sondage ont répondu que l'axe principal est une augmentation du pouvoir d'achat. Se sont les personnes ne se classant dans aucun parti sont les plus sensibles à cet axe. Mais nous voyons qu'il n'y a pas de partis politiques qui se détachent pour cet axe, comme on le voit pour une nouvelle république par exemple. Concernant les revenus, se sont les personnes gagnant entre 2000 et 3000 euros. Ce sont les personnes étant en BAC ou BAC+2 les plus sensibles, ce qui correspond à la tranche de revenu la plus sensible à cet axe.. Pour les âges de nos sondés, c'est assez hétérogène. Nous pouvons donc conclure que pour cet axe, notre population est donc en général pour une augmentation du pouvoir d'achat.

Graphique 21: histogramme du parti politique du revenu du niveau d'étude et de l'âge en fonction de la souveraineté



Pour cet axe politique, nous avons eu que 4 sondés pour la souveraineté, ce sont 4 sondés ayant un revenu différents, ayant également un niveau d'étude différent, se sont des personnes ne se classant dans aucun partis ou ne voulant pas répondre, au niveau de l'âge deux ont entre 18 et 25 ans, un a entre 26 et 35 ans et une personne a plus de 55 ans. Au vu du nombre très faible de sondés, nous ne pouvons pas tirer de conclusion pour cet axe.

Pour conclure sur les axes politiques, nous voyons que certains axes sont plébiscités par les sondés tels que le pouvoir d'achat ou les inégalités. Certains axes n'ont pas du tout intéressé nos sondés, ceux ou les réponses ont été très faibles comme la souveraineté ou le revenu universel (7 réponses). Sachant que ces questions étaient à choix multiples, nous avons néanmoins des réponses hétérogènes.

Partie économétrique

Mise en place de la base de données

Pour ce projet, nous avons deux types de variables, nous avons plusieurs variables explicatives liées au choix, les partis politiques, au nombre de 4, le parti politique des du parent 1 et 2 , l'orientation politique du maire de leur ville et l'axe des partis politiques. Nous avons également des variables directement liées à l'individu, elles sont au nombre de 23, nous avons par exemple les catégories socio-professionnelles du sondés et de ses deux parents ou encore trois variables à titre indicatif à savoir si le sondé aura plus que 18 ans en 2022, savoir si le sondé est adhérent à un parti politique ou si le sondé a déjà eu connaissance de l'affaire Cambridge Analytica. Pour ces variables, nous allons parler de modèle multinomial simple. Pour pouvoir faire des modèles sur notre base nous allons devoir la modifier le format de la base. Nous allons donc effectuer un modèle logit multinomial sur notre base de données, dans ce type de modèle, la valeur des variables explicatives ne changent pas avec les alternatives mais avec les individus.

Nous allons donc dans un premier temps changer la forme de notre base de données, pour avoir nos 4 variables dépendantes. Après avoir changé la forme de notre base, nous allons effectuer un modèle sur cette base.

Les variables concernant le vote des parents seront découpées en variable binaire, afin que les partis soient pris en compte. Pour exemple, un individu ayant son Parent1 qui a voté ou vote pour le PS, verra sa colonne Parent1PP.PS ("Parent1PP.3" dans la base) avec 1 et les autres colonnes concernant Parent1PP, tel que Parent1PP.RN ("Parent1PP.13" dans la base) ou Parent1PP.LREM ("Parent1PP.6" dans la base), seront remplis de 0.

Pour effectuer le changement de base concernant les axes, il est nécessaire d'effectuer une méthode de scoring pour chaque parti politique et chaque axe, c'est-à-dire d'effectuer des probabilités concernant les axes des partis politiques choisis par les sondés pour chaque réponse. Par exemple, les sondés ayant choisis le parti politique PS (Parti Socialiste ou "3" dans la base), 10% ont choisi l'axe

“Redressement Économique” (“1” dans la base), 100% ont choisi “Disparition/Atténuation des Inégalités Socio-Économiques” (“2” dans la base), 30% ont choisi “Ecologie” (“3” dans la base), 20% ont choisi “Sécurité” (“4” dans la base), 20% ont choisi “Immigration” (“5” dans la base), 10% ont choisi “Revenu Universel” (“6” dans la base) et 10% ont choisi “Pouvoir d’Achat” (“8” dans la base). Ainsi le choix “Redressement Économique” aura un poids égale à 0.1 dans la colonne de l’Axe lié au parti politique PS (“AxePP.3” dans la base), tandis que le choix “Disparition/Atténuation des Inégalités Socio-Économiques” aura un poids de 1 dans cette même colonne, et tous les axes non cités précédemment auront un poids égales à 0. Chaque choix d’axe aura donc un poids qui diffère pour chaque parti politique.

Un nouveau biais se présente puisque certains partis, tel que le PCF, n’ont reçu qu’un seul vote, et donc les poids des axes en fonction des sondés sera donc fallacieux, cependant il est impossible de mettre un poids en fonction du programme d’un de ces partis; car il est difficile de mesurer réellement ce que les votants pour ces partis considèrent comme étant important à leur yeux. Ceux ayant reçu 0 vote ne pourront être comptabilisés, par exemple pour le parti Agir, parti européen et pro-libéral, le redressement économique devrait être pour eux un des axes importants cependant on ne peut le quantifier à partir de ce sondage.

Tableau 2 : Récapitulatif des poids de chaque axe en fonction des partis politiques
(se référer au code utilisé pour chaque axe et chaque parti politique : Annexe 1)

AXE PARTI	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	1	0	1	0	0	0
2	1/3	2/3	2/3	0	0	0	2/3	0	0	0	0
3	1/10	1	3/10	1/5	1/5	1/10	0	1/10	0	0	0
4	1/15	3/5	4/5	1/15	0	1/15	1/15	1/15	0	0	0
5	0	0	1	1	0	0	0	0	0	0	0
6	9/19	5/19	8/19	6/19	2/19	0	1/19	2/19	0	1/19	0
7	0	1	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	0	0	0
9	10/19	3/19	0	8/19	4/19	0	1/19	3/19	0	0	0
10	0	1	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0
12	3/5	4/5	2/5	2/5	0	0	0	0	0	0	0
13	1/4	1/8	1/8	5/8	7/8	0	3/8	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0
15	10/31	43/93	43/93	24/93	10/93	1/31	1/31	17/93	2/93	2/31	1/93
16	1/3	2/9	1/3	2/9	2/9	1/9	0	4/9	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0
19	0	1	0	0	0	0	0	0	0	0	0

Afin de mieux comprendre comment les colonnes concernant les Axes en fonction des partis politiques sont calculées, ci-dessous est proposé un exemple pour l'observation numéro 3 qui a choisi pour axes politiques :

“Disparition/Atténuation des inégalités socio-économiques”, “Ecologie” et “Sécurité”
(respectivement axes “2”, “3” et “4”)

Tableau 3 : Exemple de calcul des colonnes AxePP pour l'observation 3

(se référer à l'annexe 2 pour le codage des colonnes)

	AxePP.1	AxePP. 2	AxePP. 3	AxePP. 4	AxePP. 5	AxePP. 6	AxePP. 7	AxePP. 8	AxePP. 9
Calcul	1	$\frac{2}{3} + \frac{2}{3}$	$1 + \frac{3}{10} + \frac{1}{5}$	$\frac{3}{5} + \frac{4}{5} + \frac{1}{15}$	$1 + 1$	$\frac{5}{19} + \frac{8}{19} + \frac{6}{19}$	1	0	$\frac{3}{19} + \frac{8}{19}$
Poids	1	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{22}{15}$	2	1	1	0	$\frac{11}{19}$
	AxePP.10	AxePP.11	AxePP.12	AxePP.13	AxePP.14	AxePP.15	AxePP.16	AxePP.17	AxePP.19
Calcul	1	0	$\frac{4}{5} + \frac{2}{5} + \frac{2}{5}$	$\frac{1}{8} + \frac{1}{8} + \frac{5}{8}$	0	$\frac{43}{93} + \frac{43}{9} + \frac{3+24}{93}$	$\frac{2}{9} + \frac{1}{3} + \frac{2}{9}$	0	1
Poids	1	0	$\frac{8}{5}$	$\frac{7}{8}$	0	$\frac{110}{93}$	$\frac{7}{9}$	0	1

L'AxePP.18 n'est pas à calculer car il correspond aux parents votant dans un autre pays, du moins il est égale à 0 pour n'importe quelle observation.

Analyse des modèles

Avec toutes nos variables

Avant de commencer l'analyse de nos modèles économétriques, nous avons rencontré de nombreux problèmes. Un des problèmes majeurs, Kyllien avait sur 189 individus un qui retournait un FALSE alors que pour Jean-Baptiste il y avait 189 TRUE. Nous avons exactement le même code et la même base de données. Kyllien a alors cherché qu'elle était l'individu et l'a retiré de sa base. Il m'a ensuite donné le code et c'était à mon tour d'avoir un individu avec un FALSE. C'était encore une fois la même base et le même code. Nous allons donc travailler avec la base de Kyllien, nous allons donc avoir 188 individus. Nous avons également un second problème au niveau de nos matrices, celle-ci n'est pas inversible, nous ne pouvons pas mettre certains couples de variables dans la même équation.

Néanmoins, nous avons réussi à tirer quelques modèles de cette base de données même si ceux-ci ne sont pas des modèles de qualité. Nous allons expliquer le meilleur modèle dans ce dossier, il y aura également des résultats de modèle en annexe.

Nous avons donc essayé plusieurs modèles, n'ayant que très peu de variables statistiquement significatives dans nos différents modèles, et n'ayant aucun modèles qui se détachent, nous allons donc comparer nos modèles avec le "Log-Likelihood", ce qui correspond à la fonction de vraisemblance, que l'on va essayer de maximiser. Mais nous avons rencontré un biais, nous nous sommes aperçus que plus nous ajoutons de variables à un modèle et plus la fonction de vraisemblance augmente ainsi plus la significativité de nos variables baissent.

Nous allons donc comparer 4 modèles, chacun des 4 modèles aura la même variable liée à l'individu, la catégorie socioprofessionnelle du parent 1. Nous allons ajouter une variable explicative à chaque modèle. Nous allons commencer avec un modèle sans variables explicatives, dans le deuxième modèle, nous allons avoir l'axe des partis politique en variables explicative. Pour le troisième modèle, nous aurons l'axe des partis politiques ainsi que le parti politique du parent 1 et pour le dernier modèle, nous allons retrouver toutes nos variables explicatives: l'axe des partis politiques, les partis politiques du parent 1 et 2.

Nous avons décidé de comparer nos différents modèles avec le test du rapport de vraisemblance (tableau 4). Nous allons prendre le modèle dont la p-value est la plus basse, c'est donc le modèle 2 que nous allons choisir. En annexe (annexe 3 à 5), nous allons pouvoir trouver les résultats du test de rapport de vraisemblance des trois autres modèles avec le modèle 2. Les résultats de ces comparaisons nous disent que quels que soit le "lretst" que nous effectuons, c'est le second modèle de la ligne qui est toujours le meilleur. Ce biais provient sûrement du manque de puissance du test au vu du peu de significativité de nos modèles. Nous choisissons le modèle au vu du tableau où les 4 modèles sont testés ensemble, et non en comparaison du modèle 2 par rapport aux autres.

Tableau 4 : résultats du Likelihood ratio test pour nos 4 modèles

```
Likelihood ratio test
```

Model 1:	PP ~ 0 Categorie_Parent1
Model 2:	PP ~ AxePP Categorie_Parent1
Model 3:	PP ~ AxePP + Parent1PP Categorie_Parent1
Model 4:	PP ~ AxePP + Parent1PP + Parent2PP Categorie_Parent1

	#Df	LogLik	Df	Chisq	Pr(>Chisq)	
1	34	-333.63				
2	35	-253.26	1	160.736	< 2.2e-16	***
3	36	-227.38	1	51.755	6.286e-13	***
4	37	-220.35	1	14.077	0.0001755	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nous avons ensuite voulu tester l'hétéroscédasticité de nos différents modèles, nous avons eu la même erreur que lors de la formation des modèles, à savoir que la matrice créée par la fonction "mlogit" n'est pas inversible, nous ne pouvons donc pas tester si les erreurs sont hétéroscédastiques. Nous rencontrons de nouveau un biais.

Une fois le modèle "mlogit" fait nous avons fait un modèle avec la library VGAM et la fonction "vglm". La différence entre ces deux méthodes pour faire un modèle, dans le "vglm", nous ne mettons pas de variables explicatives dans le modèle mais juste la variable à expliquer et les variables liées à l'individu. Ce modèle va nous permettre de faire un graphique. Nous allons donc faire un modèle en mettant juste la catégorie socioprofessionnelle du parent 1, ce choix se fait en fonction des modèles précédents, c'est avec cette variable que nous allons avoir le

meilleur modèle. Les résultats de ce modèle nous montre que nous avons 4 variables statistiquement significatives au seuil de 5 %, qui sont des constantes. Nous avons également une variable et une constante qui sont significatives au seuil de 10%, la variable est la sixième catégorie socioprofessionnelle, qui correspond aux personnes exerçant le métier d'ouvrier. En référence pour ce modèle, nous choisissons le septième parti politique, correspondant au MoDem, c'est ce parti qui nous donne le plus de variables statistiquement significatives. Dans notre modèle nous avons 14 constantes, ce qui correspond à 14 partis politiques différents, ce qui signifie que les partis politiques dans lesquels personne ne se classe n'ont pas été pris en compte dans le modèle ainsi que le parti de référence, ce qui est différent par rapport aux modèles d'avant.

Tableau 5 : Summary de la fonction VGLM avec Categorie_Parent1 en variable explicative et le parti "Modem" en référence

```
vglm(formula = PP ~ Categorie_Parent1, family = multinomial(refLevel = 7),
data = baseEco2)

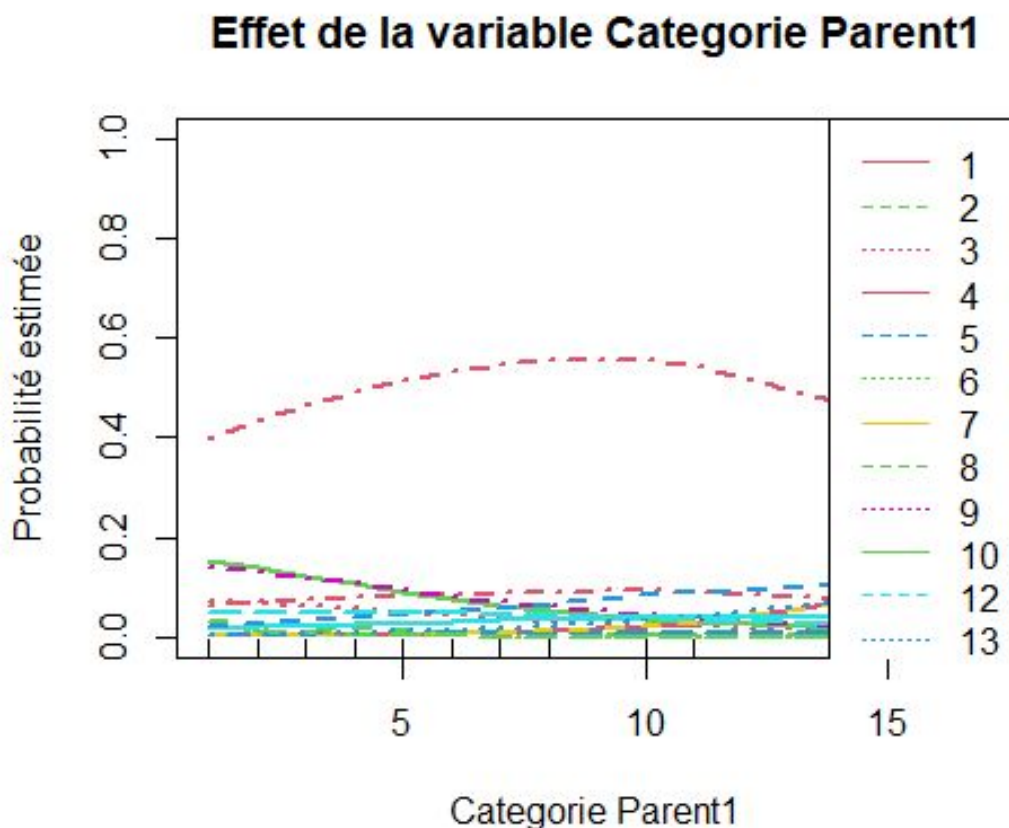
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1    -5.333e-13  2.785e+00  0.000  1.0000
(Intercept):2     3.818e+00  2.483e+00  1.538  0.1241
(Intercept):3     4.566e+00  2.168e+00  2.106  0.0352 *
(Intercept):4     4.252e+00  2.100e+00  2.024  0.0429 *
(Intercept):5     1.154e+00  3.058e+00  0.378  0.7058
(Intercept):6     5.348e+00  2.100e+00  2.547  0.0109 *
(Intercept):7     3.769e+00  3.236e+00  1.165  0.2441
(Intercept):8     5.208e+00  2.098e+00  2.482  0.0131 *
(Intercept):9     3.769e+00  3.236e+00  1.165  0.2441
(Intercept):10    2.949e+00  2.264e+00  1.303  0.1927
(Intercept):11    3.128e+00  2.145e+00  1.458  0.1447
(Intercept):12    1.160e+00  2.454e+00  0.473  0.6363
(Intercept):13     6.097e+00  2.021e+00      NA      NA
(Intercept):14     4.043e+00  2.175e+00  1.859  0.0631 .
Categorie_Parent1:1  5.467e-14  3.428e-01  0.000  1.0000
Categorie_Parent1:2 -5.668e-01  4.428e-01 -1.280  0.2005
Categorie_Parent1:3 -4.478e-01  3.140e-01 -1.426  0.1538
Categorie_Parent1:4 -2.798e-01  2.807e-01 -0.997  0.3188
Categorie_Parent1:5 -1.985e-01  4.865e-01 -0.408  0.6833
Categorie_Parent1:6 -4.833e-01  2.904e-01 -1.664  0.0961 .
Categorie_Parent1:7 -8.842e-01  8.144e-01      NA      NA
Categorie_Parent1:8 -4.478e-01  2.882e-01 -1.554  0.1202
Categorie_Parent1:9 -8.842e-01  8.144e-01      NA      NA
Categorie_Parent1:10 -2.363e-01  3.206e-01 -0.737  0.4612
Categorie_Parent1:11 -1.775e-01  2.842e-01 -0.625  0.5323
Categorie_Parent1:12 -7.227e-02  3.230e-01 -0.224  0.8230
Categorie_Parent1:13 -2.843e-01  2.567e-01 -1.108  0.2679
Categorie_Parent1:14 -3.475e-01  3.083e-01 -1.127  0.2598
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of linear predictors: 14

Residual deviance: 667.2596 on 2604 degrees of freedom

Log-likelihood: -333.6298 on 2604 degrees of freedom
```

Graphique 22 : Effet de la variable Categorie_Parent1 sur le Parti Politique



En retirant certaines variables

En essayant d'autres modèles, nous nous sommes rendus compte que le fait de garder les partis politiques n'ayant pas reçu de votes, cela faussait nos résultats. Nous avons fait des modèles en retirant ces partis politiques.

Nous avons créé deux modèles, ces deux modèles ont les deux même variables liées à l'individu, l'âge et le type d'habitation, et au niveau des variables explicatives, nous avons l'axe des différents partis politiques et les partis politiques du parent 1 et 2. La différence entre ces deux modèles se fait que sur le second, nous avons mis une référence de 15 qui correspond aux personnes ne se classant dans aucun parti, correspondant au plus grand nombre de sondés. Nous allons donc effectuer un "Lrtest" sur nos deux modèles pour choisir le meilleur des deux. Avec le tableau 6, nous voyons que le modèle 2 est le meilleur, celui avec la référence. La même conclusion revient quand nous regardons la significativité des variables ainsi que le R^2 de McFadden, plus grand sur le modèle avec la référence (les résultats du modèle sans référence se trouvent en annexe 6). Concernant l'hétéroscédasticité de nos modèles, nous avons toujours un message d'erreur, nous sommes en incapacité d'effectuer le test. On suppose que les erreurs n'ont pas de problème d'hétéroscédasticité, que donc elles sont homoscédastiques.

Tableau 6 : LrTest entre le modèle sans référence et avec référence = "15"

```
Likelihood ratio test

Model 1: PP ~ AxePP + Parent1PP + Parent2PP | Age + Type_Habitation
Model 2: PP ~ AxePP + Parent1PP + Parent2PP | Age + Type_Habitation
#Df  LogLik Df Chisq Pr(>Chisq)
1   45 -200.28
2   45 -200.28  0      0 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nous allons maintenant analyser les résultats du modèle avec la référence. Nous voyons que 2 variables concernant le type d'habitation sont statistiquement significatives à hauteur de 5% et deux variables concernant l'âge de l'individu sont statistiquement significatives à hauteur de 10%. Le type d'habitation est significatif par rapport aux personnes se classant dans les partis politiques d'Europe Ecologie

Les Verts et de La République En Marche. La variables âge quant à elle est significative pour les sondés se classant dans le parti du Rassemblement National. Nous avons également un R^2 de McFadden égal à 0.408, ce que nous n'avions pas pour les modèles où nous avons pas retiré certains partis politiques. Au vu du résultat du "lrtest", nous pouvons conclure que nous allons préférer le modèle avec la référence 15. Selon les mots de McFadden un modèle ayant un R^2 entre 0.2 et 0.4 est la preuve d'un bon ajustement de modèle, nous allons donc dire que pour notre modèle, l'ajustement est bon.

Tableau 7 : Summary du modèle avec comme variable lié à l'individu : Age et Type_Habitation

-----		Parent1PP	1.093*** (0.380)	Type_Habitation:1	6.860 (5,019.748)
Dependent variable:				Type_Habitation:2	-20.480 (14,760.760)
-----		PP	Parent2PP	1.458*** (0.390)	
(Intercept):1	-72.257 (19,126.390)	Age:1	7.773 (2,712.406)	Type_Habitation:3	-0.208 (0.823)
(Intercept):2	16.100 (14,760.760)	Age:2	-0.054 (0.529)	Type_Habitation:4	-1.915** (0.778)
(Intercept):3	-3.647*** (1.329)	Age:3	-0.230 (0.340)	Type_Habitation:5	-20.061 (11,975.500)
(Intercept):4	-0.644 (1.076)	Age:4	0.098 (0.268)	Type_Habitation:6	-1.201** (0.591)
(Intercept):5	20.552 (11,975.520)	Age:5	-2.983 (11.740)	Type_Habitation:7	2.635 (1.643)
(Intercept):6	-0.420 (0.937)	Age:6	0.088 (0.203)	Type_Habitation:8	0.722 (1.383)
(Intercept):7	-1.749 (55.450)	Age:7	-5.004 (27.627)	Type_Habitation:9	0.318 (0.479)
(Intercept):8	-9.367 (6.018)	Age:8	0.565 (0.846)	Type_Habitation:10	-17.860 (8,183.014)
(Intercept):9	-2.130** (1.028)	Age:9	-0.166 (0.227)	Type_Habitation:12	-0.794 (1.168)
(Intercept):10	20.003 (8,183.021)	Age:10	-3.009 (5.304)	Type_Habitation:13	-0.577 (1.405)
(Intercept):12	-5.561*** (1.742)	Age:12	0.496 (0.343)	Type_Habitation:14	1.040 (1.508)
(Intercept):13	-8.470** (4.211)	Age:13	1.066* (0.630)	Type_Habitation:16	-0.139 (0.892)
(Intercept):14	1.581 (10.719)	Age:14	-3.392 (5.274)	-----	
(Intercept):16	-4.116*** (1.490)	Age:16	0.487* (0.291)	Observations	188
AxePP	5.131*** (0.636)			R2	0.408
				Log Likelihood	-200.275
				LR Test	275.939*** (df = 45)
				=====	
				Note:	*p<0.1; **p<0.05; ***p<0.01

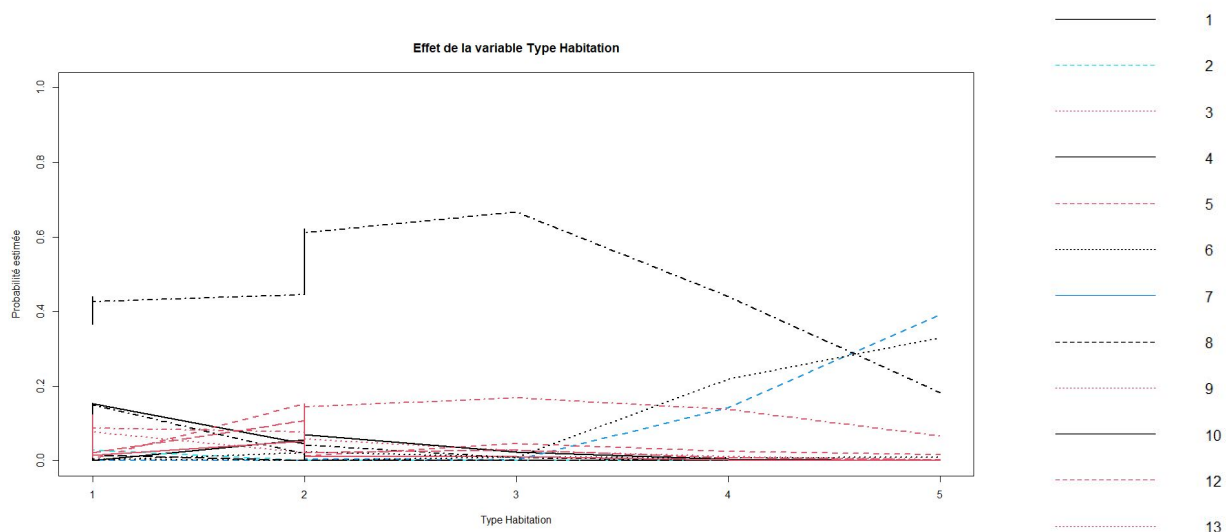
Nous allons maintenant comparer les choix des individus à se classer dans aucun parti par rapport à ceux qui se classent dans un parti en particulier. Pour ce faire, nous allons garder notre équation avec notre variable concernant la catégorie socioprofessionnelle du parent 1, à laquelle nous allons ajouter une référence, celle-ci concerne les personnes ne se classant dans aucun parti, ce qui correspond à la valeur 15. Les partis dans lesquels aucun de nos sondés ne se classaient, ont également été retirés de l'analyse. Il va donc y avoir une comparaison entre la variable de référence et les 14 partis politiques restants. La comparaison via le score test n'est pas possible pour le modèle étudié précédemment, la matrice que nous avons utilisée n'est pas inversible. Nous avons donc fait le choix d'en sortir un avec la catégorie socioprofessionnelle du parent 1, ce qui nous donne qu'au seuil de risque de 5%, une certaine catégorie socioprofessionnelle du parent 1 ne change en rien le choix de se classer dans un parti politique par rapport au personne ne se classant dans aucun parti politique. Ainsi l'hypothèse d'autocorrélation est vérifiée.

Tableau 8 : Vérification de l'hypothèse d'autocorrélation

Dependent variable:		Categorie_Parent1:2	
pp		-0.254 (1.003)	
(Intercept):1	-12.660 (30.133)	Categorie_Parent1:3	-0.283 (0.378)
(Intercept):2	-4.306 (4.739)	Categorie_Parent1:4	0.005 (0.260)
(Intercept):3	-3.566* (1.938)	Categorie_Parent1:5	-0.102 (2.624)
(Intercept):4	-2.753** (1.272)	Categorie_Parent1:6	-0.113 (0.181)
(Intercept):5	-6.874 (13.274)	Categorie_Parent1:7	0.348 (2.122)
(Intercept):6	-1.108 (0.999)	Categorie_Parent1:8	-0.413 (3.705)
(Intercept):7	-7.535 (14.882)	Categorie_Parent1:9	-0.149 (0.225)
(Intercept):8	-3.980 (11.049)	Categorie_Parent1:10	-1.048 (3.959)
(Intercept):9	-1.202 (1.087)	Categorie_Parent1:12	-0.059 (0.334)
(Intercept):10	-1.619 (11.725)	Categorie_Parent1:13	-0.169 (0.398)
(Intercept):12	-4.611*** (1.646)	Categorie_Parent1:14	0.505 (1.325)
(Intercept):13	-3.638 (2.259)	Categorie_Parent1:16	-0.136 (0.284)
(Intercept):14	-6.586 (6.853)	iv:Type1	0.739*** (0.236)
(Intercept):16	-1.489 (1.504)	iv:Type2	0.659*** (0.205)
AxePP	5.476*** (1.040)	Observations	188
Parent1PP	1.271*** (0.407)	R2	0.352
		Log Likelihood	-219.047
		LR Test	238.395*** (df = 33)

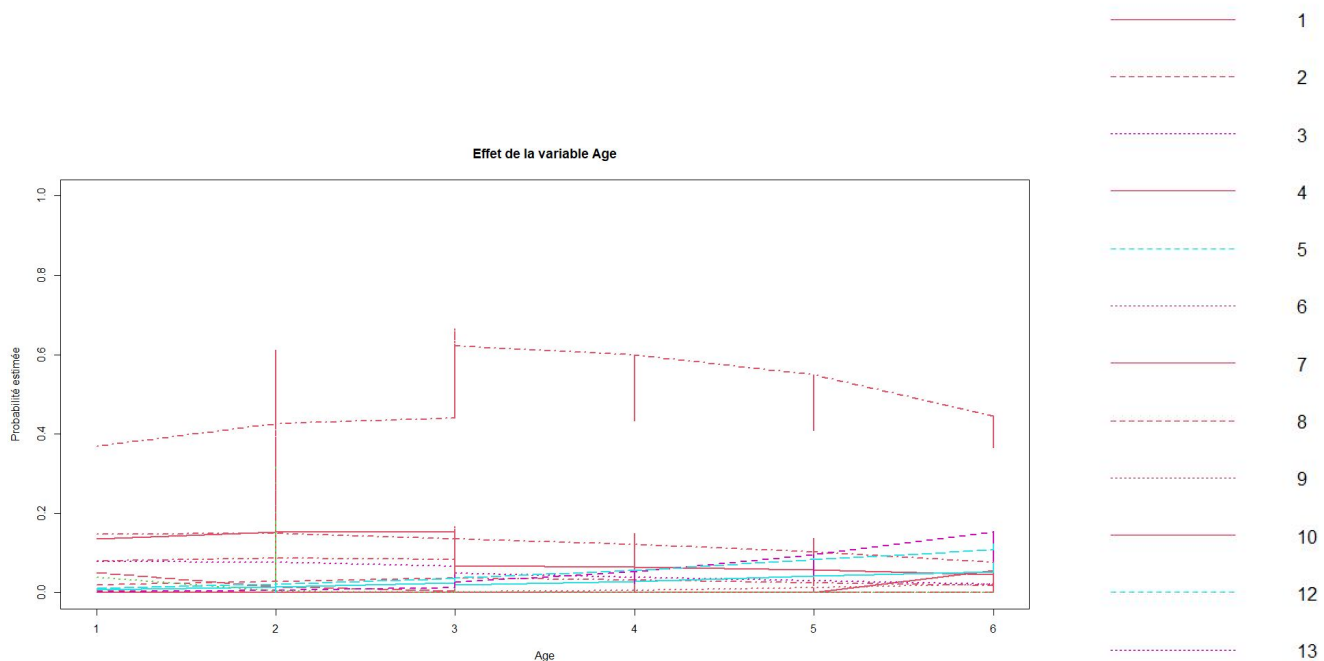
Suite à ces nouveaux modèles, nous avons décidé de refaire deux graphiques, comme ceux fait dans la partie au-dessus. Pour faire ces graphiques, nous avons refait deux modèles avec la fonction “vglm”, ces modèles auront les mêmes variables liés à l’individu que les deux modèles précédents, à savoir la tranche d’âge des individus et le type d’habitation des individus. Avec ces modèles, nous voyons que nous avons plus de variables statistiquement significatives. Pour nos deux graphiques, nous allons prendre les variables du type d’habitation et tranche d’âge. Nous avons changé de variable liée à l’individu, du fait que cette variable ressort comme la variable étant la plus significative. Sur le premier graphique (graphique 23) concernant le type d’habitation, nous voyons que pour le quatrième parti politique (EELV), le type d’habitation 1 2 ou 3 (appartement-maison-HLM) estime entre 40 et 70% de ce parti par rapport à nos sondés.

Graphique 23 : Estimation des Partis par rapport la variable Type_Habitatio



Sur le second graphique (graphique 24) concernant la tranche d'âge de nos individus, nous voyons que les tranches d'âge de nos sondés sont significatives pour expliquer le parti politique numéro 8 (Mouvement Républicains). Cette estimation varie entre 40% et 60% pour la tranche d'âge 3 correspondant aux sondés ayant entre 26 et 35 ans. Nous allons dire que les sondés ayant entre 26 et 35 ans estiment le mieux le parti politique du Mouvement Républicains.

Graphique 24 : Estimations des Partis avec la variable Age



Pour terminer ce dossier, nous allons analyser les effets marginaux et les odds-ratios de notre meilleur modèle, celui ayant le type d'habitation et les tranches d'âge de sondés en variables liées à l'individu ainsi que la référence. Sur le tableau 9, nous voyons les odds-ratios de nos variables liées à l'individu. Pour interpréter ce tableau, nous allons dire que le fait de changer de tranche d'âge, de vieillir, augmente la probabilité de $(1/2.905)$ 34.42% voter pour RN. Nous voyons bien que notre modèle n'est pas représentatif de la France

Tableau 9 : Odds-ratio pour le modèle avec Age et Type_Habitation en variable lié à l'individu

(Intercept):1	(Intercept):2	(Intercept):3	(Intercept):4	(Intercept):5	(Intercept):6	(Intercept):7	(Intercept):8								
0	9,825,401.000	0.026	0.525	842,748,096.000	0.657	0.174	0.0001								

=====															
(Intercept):9	(Intercept):10	(Intercept):12	(Intercept):13	(Intercept):14	(Intercept):16	AxePP	Parent1PP								
0.119	486,386,644.000	0.004	0.0002	4.859	0.016	169.270	2.984								

=====															
Parent2PP	Age:1	Age:2	Age:3	Age:4	Age:5	Age:6	Age:7	Age:8	Age:9	Age:10	Age:12	Age:13			
4.296	2,375.638	0.947	0.795	1.103	0.051	1.092	0.007	1.760	0.847	0.049	1.643	2.905			

=====															
Age:14	Age:16	Type_Habitation:1	Type_Habitation:2	Type_Habitation:3	Type_Habitation:4	Type_Habitation:5	Type_Habitation:6	Type_Habitation:7	Type_Habitation:8	Type_Habitation:9	Type_Habitation:10	Type_Habitation:12	Type_Habitation:13	Type_Habitation:14	Type_Habitation:16
0.034	1.627	953.412	0	0.812	0.147	0	0.301	13.950	2.058	1.375	0.00000	0.452	0.562	2.829	0.870

=====															

Voyons maintenant les effets marginaux de notre modèle avec le tableau 10, pour interpréter ces effets marginaux, nous avons fait une moyenne par rapport à tous les individus, le résultat de cette moyenne se trouve dans le tableau 10. Dans les résultats, la valeur 15 au début du tableau correspond à la référence. Nous allons dire que l'augmentation d'une tranche d'âge de nos sondés, va par exemple réduire le fait de se classer dans le Parti Socialiste de 0.003 mais va augmenter le fait de se classer dans le Rassemblement National de 0.0002. Malgré que nous ne pouvons pas faire d'inférence statistique, nous voyons que le cliché : la vie politique commence à gauche et se termine à droite est vérifié par notre modèle.

Tableau 10 : Les effets marginaux moyen de l'âge sur l'échantillon

15	1	2	3	4	5	6	7	8	9	10	12	13	14	16
-0.017	0	-0	-0.003	0.004	-0	0.004	-0.00000	0.001	-0.011	-0.00000	0.002	0.0002	-0.002	0.022

Conclusion

Pour conclure, après avoir élaboré un questionnaire pour pouvoir faire ce projet, nous avons eu de nombreux problèmes, concernant au début la base de données puis ensuite des erreurs de codes dont nous ne savions pas d'où ça venait. Nous avons réussi à régler beaucoup de problèmes mais certains persistent comme le test d'hétéroscédasticité que nous n'avons pas réussi à passer. Néanmoins nous avons réussi à sortir un modèle avec nos 3 variables explicatives et 2 variables liées à l'individu, de plus ce modèle a certaines variables statistiquement significatives et un R^2 de Mc fadden de 0.4, ce qui signifie que nous avons un ajustement de nos variables qui est bon. Nous aurions pu avoir un meilleur modèle avec un nombre de sondés plus important, une population qui serait plus représentative de la France, avec une situation géographique plus hétérogène et des tranches d'âge plus hétérogène. Notre base n'est pas faite pour faire de l'inférence statistique. De plus, nous voyons qu'avec l'interprétation des odds-ratios que notre base n'est pas du tout représentative. Nous ne pouvons donc pas conclure sur le fait que le BIG DATA peut avoir un effet négatif sur une élection. Toutefois il faut se méfier de ce qu'on publie ou poste sur les réseaux sociaux, une publication ou un post n'est jamais anodin. Il est possible qu'avec une plus grande base de données, des regroupements soit fait et que l'influence via des pubs soit faite sur vos réseaux, la base de données qu'on eu accès à Cambridge Analytica contenait des millions de données.

Nous finirons ce dossier comme nous l'avons commencé, par une citation, celle-ci sera de René RÉMOND

“ les médias ne font pas l'élection, mais ils contribuent grandement à façonner la vie politique et concourent à la formation des électeurs. Leur action sur les élections n'est pas nulle, mais elle s'exerce de façon diffuse et dans le long terme.” René

Rémond.

Annexes

Annexe 1 : Récapitulatif des réponses recodées pour R

Nom de la Variable	Réponses	Réponses recodées
<u>Pour notre variable dépendante et ses 4 valeurs</u>		
PP PP_Parent1 PP_Parent2	PCF	1
	LFI	2
	PS	3
	EELV	4
	DVG	5
	LREM	6
	MoDem	7
	MR	8
	LR	9
	UDI	10
	Agir	11
	DVD	12
	RN	13
	LP	14
	Je me classe dans aucun parti politique	15
	Je ne veux pas répondre	16
	Je ne sais pas	17
	Parent vote dans un autre pays	18
	Ne vote pas	19
PP_Axe	Redressement économique	1

	Disparition/Atténuation des inégalités socio-économiques	2
	Ecologie	3
	Sécurité	4
	Immigration	5
	Revenu Universel	6
	Nouvelle République	7
	Augmentation du Pouvoir d'achat	8
	Souveraineté/Sortie de l'euro	9
	Aucun	10
	Ne veut pas répondre	11
PP_Maire	Extrême Gauche	1
	Gauche	2
	Centre Gauche	3
	Les verts	4
	Centre	5
	Centre Droite	6
	Droite	7
	Extrême Droite	8
	Sans étiquette	9
	Ne sait pas	10
<u>Pour nos variables explicatives</u>		
Adhérent	Non	0

	Oui	1
Nbre_Hab	Moins de 2000 habitants	1
	Entre 2000 et 10 000 habitants	2
	Entre 10 000 et 100 000 habitants	3
	Plus de 100 000 habitants	4
Catégorie socio-professionnelle (Sondé, Parent1 et Parent2)	Agriculteurs Exploitants	1
	Artisans, commerçants et chefs d'entreprise	2
	Cadres et professions intellectuelles supérieures	3
	Professions intermédiaires	4
	Employés	5
	Ouvriers	6
	Retraités	7
	Autres personnes sans activité professionnelle	8
	Étudiants	9
	Ne veux pas répondre	10
	Ne sais pas	11
Niveau d'étude (Sondé, Parent1 et Parent2)	Brevet	1
	CAP/BEP	2
	BAC	3
	BAC +2	4
	BAC +3	5
	BAC +4	6
	BAC +5	7
	BAC +6 et plus	8
	Ne sais pas	9

	Ne veux pas répondre	10
	Autre	11
Media_Type	Journal TV	1
	Les Réseaux Sociaux	2
	Journal Papier	3
	Chaine d'Info en Continu (BFM TV, Cnews)	4
	Application Mobile	5
	Radio	6
	Internet	7
	Aucun	8
Principal Média	TF1	1
	France 2 / ARTE	2
	M6	3
	BFM	4
	CNews	5
	France Info	6
	France Inter / RTL	7
	Bloomberg / Les Echos	8
	Figaro	9
	Le Monde	10
	Ouest France / Dauphine	11
	20 Minutes	12
	Alternatives Eco / MediaPart / Canard Enchaîné	13
	Réseaux Sociaux (Twitter, HugoDecrypte, Brut, Youtube, MediaAvenir)	14
	Courrier International / Le Matin	15
	Aucun / Ne souhaite pas	16

	répondre	
Revenu du sondé par mois	Aucun revenu	1
	Moins de 500	2
	ENTre 500 et 800	3
	Entre 800 et 1000	4
	Entre 1000 et 1500	5
	Entre 1500 et 2000	6
	Entre 2000 et 3000	7
	Entre 3000 et 5000	8
	Plus de 5000	9
	Ne veux pas répondre	10
Propriétaire ou locataire	Locataire	1
	Propriétaire	2
	Ne veux pas répondre	3
Type d'habitation	Appartement	1
	Maison	2
	HLM	3
	Crous	4
	Ne veux pas répondre	5
Domaine d'étude	Economie/ gestion/ management	1
	Art	2
	Sociologie	3
	Santé	4
	Informatique	5
	Sciences	6
	Communication	7
	Administratif	8
	Banque	9

	Psychologie	10
	Marketing	11
	Industrie	12
	Ingénierie	13
	Sociale	14
	Commerce	15
	Mathématiques	16
	BTP	17
	Mécanique	18
	Logistique	19
	Enseignement	20
	Droit	21
	Langues	22
	Tourisme	23
	Electronique	24
	Ménager	25
	Assurance	26
	Vétérinaire	27
	Sport	28
	Ressources humaines	29
	Transport	30
	Energies	31
	Ne souhaite pas répondre	32
Genre	Femme	0
	Homme	1
Âge	Moins de 18 ans	1
	Entre 18 et 25 ans	2
	Entre 26 et 35 ans	3

	Entre 36 et 45 ans	4
	Entre 46 et 55 ans	5
	Plus de 56 ans	6
Plus de 18 ans 2022	Non	0
	Oui	1
Affaire Cambridge Analytica	Non	0
	Oui	1

Annexe 2 : Tableau récap AxePP

Code	AxeP P.1	AxeP P.2	AxeP P.3	AxeP P.4	AxeP P.5	AxeP P.6	AxeP P.7	AxeP P.8	AxeP P.9
Réelle	PCF	LFI	PS	EELV	DVG	LREM	MoDem	MR	LR
Code	AxeP P.10	AxeP P.11	AxeP P.12	AxeP P.13	AxeP P.14	AxeP P.15	AxeP P.16	AxeP P.17	AxeP P.19
Réelle	UDI	Agir	DVD	RN	LP	Aucun Parti	Je ne veux pas répond re	Je sais pas	Ne Vote pas

Annexe 3 : Test du Rapport de Vraisemblance entre le modèle 2 et le modèle 1

```
Likelihood ratio test

Model 1: PP ~ AxePP | Categorie_Parent1
Model 2: PP ~ 0 | Categorie_Parent1
#Df  LogLik Df  Chisq Pr(>Chisq)
1   35 -253.26
2   34 -333.63 -1 160.74 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Annexe 4 : Test du Rapport de Vraisemblance entre le modèle 2 et le modèle 3

```
Likelihood ratio test

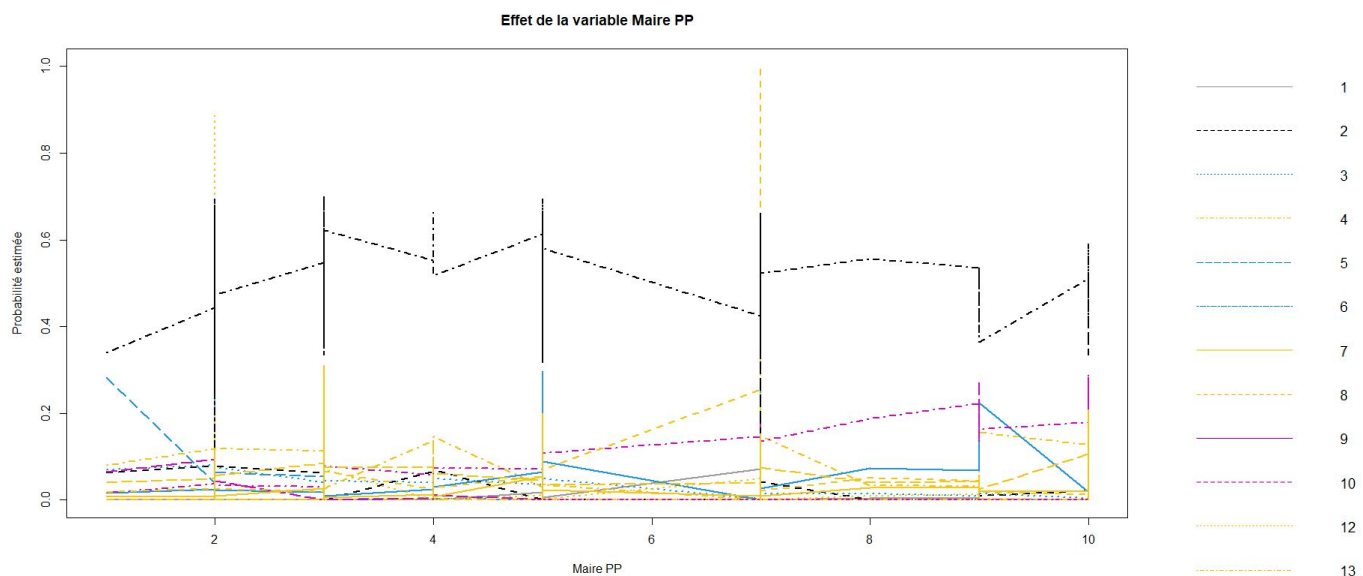
Model 1: PP ~ AxePP | Categorie_Parent1
Model 2: PP ~ AxePP + Parent1PP | Categorie_Parent1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   35 -253.26
2   36 -227.38  1  51.755  6.286e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Annexe 5 : Test du Rapport de Vraisemblance entre le modèle 2 et le modèle 4

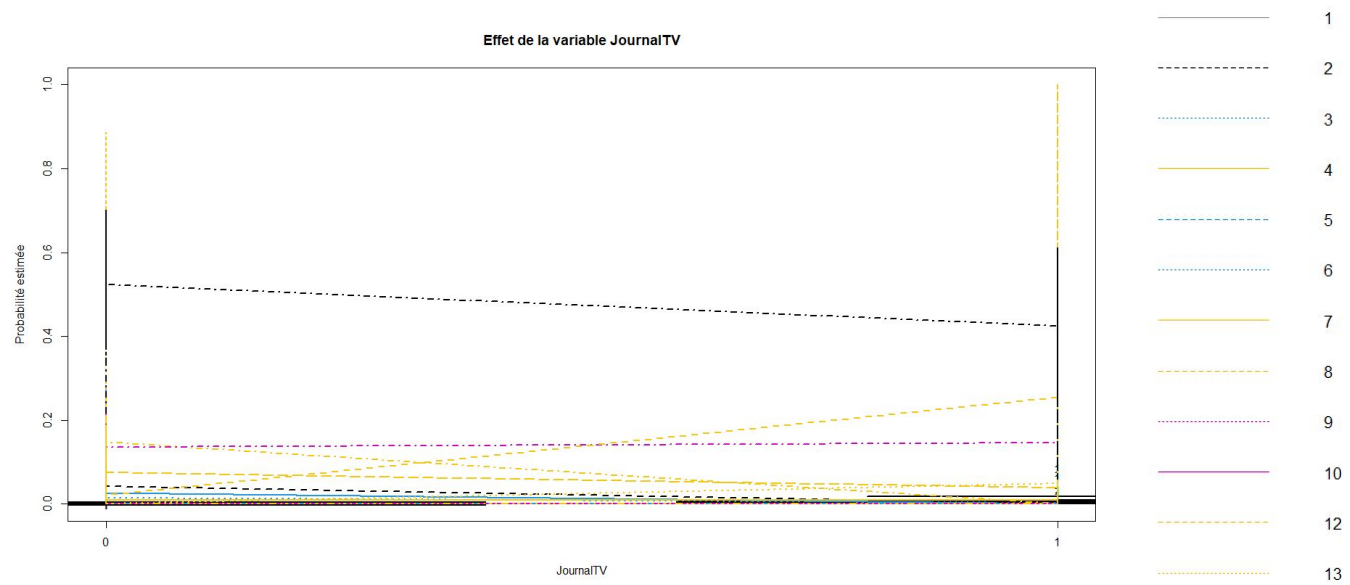
```
Likelihood ratio test

Model 1: PP ~ AxePP | Categorie_Parent1
Model 2: PP ~ AxePP + Parent1PP + Parent2PP | Categorie_Parent1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   35 -253.26
2   37 -220.35  2  65.832  5.066e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Annexe 6 : Graphique de la variable Maire.PP pour estimer la variable PP



Annexe 7 : Graphique de la variable JournalTV pour estimer la variable PP



Annexe 8 et 9 seront envoyés avec le dossier car ces fichiers sont trop long

Bibliographie

Les sites qui nous ont permis de faire ce dossier et pour l'oral de ce projet :

DELPORTE C. & LHERAULT M. (2004), *Les médias font-ils l'élection ? Retour sur une controverse*

[Les médias font-ils l'élection ? Retour sur une controverse | Cairn.info](#)

VEYRAT-MASSON I. (2017), *La télévision au centre de la bataille présidentielle*

[La télévision au centre de la bataille présidentielle | la revue des médias \(ina.fr\)](#)

DENIS ML. (2012), *La régulation audiovisuelle et l'élection présidentielle*

[La régulation audiovisuelle et l'élection présidentielle | Conseil constitutionnel \(conseil-constitutionnel.fr\)](#)

CAYROL R. (2007), *Voter à l'heure des médias et des sondages*

[Voter, à l'heure des médias et des sondages | Cairn.info](#)

SKORNICKI A. & TOURNADRE J. (2015), *La nouvelle histoire des idées politiques*

[V. La mise en politique des idées | Cairn.info \(univ-nantes.fr\)](#)

Table des matières

Sommaire	2
Introduction	3
Analyse économique du sujet	5
Explication du questionnaire	8
Analyse statistique	16
Analyse Statistique de l'ensemble des données	16
Analyse statistique par rapport à l'âge des sondés	19
Analyse Statistique par rapport aux revenus des sondés	24
Analyse statistique par rapport aux axes politiques des sondés	30
Partie économétrique	38
Mise en place de la base de données	38
Analyse des modèles	42
Avec toutes nos variables	42
En retirant certaines variables	46
Conclusion	52
Annexes	53
Bibliographie	61
Table des matières	62