



VISUALISATION DE DONNÉES

MANUEL D'UTILISATION DE L'APPLICATION : IMAGINR



SERGE GUILLEMART
QUENTIN MARCOU
Promotion 2020

Responsable du module : LYDIA
BOUDJELOUD-ASSALA

1er avril 2019

“La raison d’être des statistiques, c’est de vous donner raison.”

Abe Burrow

Table des matières

1	Introduction	3
2	Données	4
3	Univarié	4
3.1	Analyse univariée pour une variable quantitative	4
3.2	Analyse univariée pour une variable qualitative	4
4	Bivarié	5
4.1	Analyse bivariée de contingence	5
4.2	Analyse bivariée : regression	5
4.3	Analyse Anova	5
5	Multivarié	5
5.1	Regression multiple	5
5.2	Analyse factorielle	6
5.3	classification	6
6	Cartographie	6

1 Introduction

Le projet Cartelec (Beauguitte et Colange, 2013 ; Colange, Beauguitte et Freire-Diaz, 2013) a constitué une étape majeure dans l'évolution de la géographie et la sociologie électorales contemporaines. Ils ont contribué à installer le bureau de vote comme une échelle de référence dans l'analyse électorale.

Les données utilisées dans ce projet sont un croisement des listes électorales parisiennes avec les résultats du 1er tour des élections présidentielles de 2007.

En effet, les listes électorales sont des documents communicables au public par les communes et les préfectures, aux termes des articles L. 28 et R. 16 du Code électoral, dès lors que la personne requérante possède la qualité d'électeur (sans qu'il soit besoin que ce soit dans la commune dont on demande communication de la liste électorale) et qu'elle s'engage sur l'honneur à ne pas en faire « un usage purement commercial ». Ce qui vaut pour la liste électorale vaut aussi, juridiquement, pour les archives des listes électorales, et ce sans délai. Sur ces listes électorales figurent, outre les noms et prénoms des électeurs, leur adresse et le numéro du bureau de vote auquel ils sont rattachés, ainsi qu'un certain nombre d'informations complémentaires : lieu et date de naissance, généralement le sexe, parfois des renseignements complémentaires utiles à l'administration. On utilise donc comme base de données un découpage des bureaux de vote parisiens selon la carte suivante :

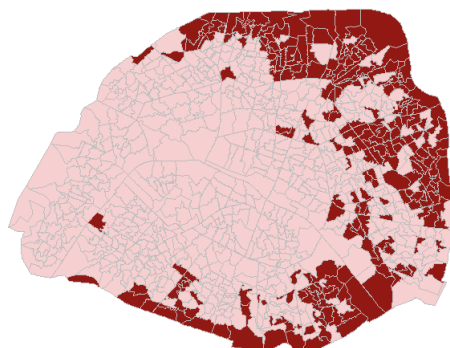


FIGURE 1 – Découpage de Paris par bureaux de vote

Ainsi, chaque enregistrement de notre base de données correspond à un bureau de vote parisien, qui possède un identifiant unique, et qui correspond à une zone géographique délimitée. Pour chacune de ces zones, la base de données contient des données statistiques (taux de chômage...) et les scores obtenus par chaque candidat au premier tour de l'élection présidentielle de 2007.

2 Données

Ce premier onglet fait apparaître l'ensemble du tableau à explorer, tableau qui permet de faire des recherches et des tris. L'utilisateur peut travailler sur les données d'exemple ou bien charger son propre tableau au format CSV. Il est aussi possible d'appliquer un filtre, c'est-à-dire une sélection des lignes du tableau à partir d'un test conditionnel.

D'autres possibilités sont offertes à l'utilisateur au niveau de l'affichage des données, il peut choisir le nombre d'enregistrements affiché sur une page ou encore naviguer entre les différentes pages qui représentent les enregistrements successifs de la base de données.

Un champ de recherche est également implanté, il permet de repérer des mots clés.

L'utilisateur peut également récupérer les données courantes (avec l'application d'un filtre par exemple) au format CSV. Le choix du séparateur qui correspond aux normes anglaise et française est laissé à l'utilisateur.

3 Univarié

3.1 Analyse univariée pour une variable quantitative

Le choix de la variable à analyser est laissée au choix de l'utilisateur à travers un menu déroulant. Une fois celle-ci sélectionnée, l'histogramme de répartition de la variable est affiché et des informations numériques sont apportées. Il est également précisé si des occurrences de la variable étudiées sont manquantes pour certains enregistrements de la base de données.

Des options de personnalisation de l'histogramme sont disponibles, l'utilisateur peut choisir le nombre de classes et d'afficher ou non les 3 premiers quartiles de la série statistique.

L'histogramme peut être récupéré par l'utilisateur en choisissant les dimensions de l'image qui sera enregistrée.

3.2 Analyse univariée pour une variable qualitative

Le choix de la variable à analyser est laissée au choix de l'utilisateur à travers un menu déroulant. Une fois celle-ci sélectionnée, l'histogramme de répartition de la variable est affiché et des informations numériques sont apportées. Ces informations sont la fréquence absolue (nombre d'enregistrements ayant cette valeur de variable qualitative) et la fréquence relative qui correspond au pourcentage d'enregistrements ayant cette valeur de variable qualitative.

Des options de personnalisation de l'histogramme sont disponibles, l'utilisateur peut choisir le nombre de classes.

L'histogramme peut être récupéré par l'utilisateur en choisissant les dimensions de l'image qui sera enregistrée.

4 Bivarié

4.1 Analyse bivariée de contingence

Pour deux variables qualitatives on effectue un tri croisé. Le résultat d'un tri croisé, portant sur deux variables à items, est un tableau croisé ou tableau de contingence, donnant en un tableau rectangulaire les effectifs ou les fréquences des diverses combinaisons de caractères observées.

4.2 Analyse bivariée : regression

On effectue une régression linéaire simple où une variable X est expliquée, modélisée par une fonction affine d'une autre variable y . La finalité d'un tel modèle est multiple et dépend donc du contexte et surtout des questions sous-jacentes. Ce peut-être juste une approche exploratoire ou alors la recherche d'une réponse à une question du type : une variable quantitative X a-t-elle une influence sur la variable quantitative Y ? Ou encore la recherche d'un modèle de prévision de Y en fonction de X .

4.3 Analyse Anova

L'analyse de variance ou analyse factorielle est une technique permettant de savoir si une ou plusieurs variables dépendantes (appelées aussi variables endogènes ou variables à expliquer) disposées dans différentes lignes d'un tableau, sont en relation avec une ou plusieurs variables dites indépendantes (ou variables exogènes ou variables explicatives), disposées dans différentes colonnes d'un tableau.

5 Multivarié

5.1 Regression multiple

En statistiques, la régression linéaire multiple est une méthode de régression mathématique étendant la régression linéaire simple pour décrire les variations d'une variable endogène associée aux variations de plusieurs variables exogènes.

5.2 Analyse factorielle

Cette méthode d'analyse entre plusieurs variables quantitatives permet d'obtenir 4 types d'information :

- la matrice de corrélation entre les variables
- le cercle de corrélation
- la décomposition de l'inertie. On mesure l'inertie totale comme la somme des carrés des distances des points par rapport au centre de gravité. Dans le cas où tous les points ne sont pas affectés du même poids (en mécanique, on aurait dit « masse »), les carrés des distances sont bien sûr pondérés.
- la représentation des coordonnées des individus.

5.3 classification

L'idée de cette partie est de regrouper par classes les différents enregistrements de la base de données. On réalise un dendrogramme qui illustre l'arrangement de groupes générés par un regroupement hiérarchique ou hiérarchisant. On affiche aussi le profil des classes afin de montrer les valeurs ayant permis le regroupement dans une même classe.

6 Cartographie

Une carte choroplèthe est une carte thématique où les régions sont colorées ou remplies d'un motif qui montre une mesure statistique, tels la densité de population ou le revenu par habitant. Ce type de carte facilite la comparaison d'une mesure statistique d'une région à l'autre ou montre la variabilité de celle-ci pour une région donnée.

On a ici chargé un fond de cartes correspondant aux zones de Paris délimitées par les différents bureaux de vote. On choisit alors une variable donc on affiche le taux pour chaque zone délimitée.