

INF552

Élection présidentielle 2017

Visualisation de données

Table des matières

<u>TABLE DES MATIERES.....</u>	<u>2</u>
<u>INTRODUCTION.....</u>	<u>3</u>
<u>STRUCTURE DU CODE</u>	<u>3</u>
VUE D'ENSEMBLE.....	ERREUR ! SIGNET NON DEFINI.
DATA PROCESSING	3
HTML – D3.JS	3
<u>FONCTIONNALITES</u>	<u>4</u>
LE MODE	4
LE FILTRE.....	5
LE HIGHLIGHT	5
LE TOUR.....	5
LE NIVEAU	5
<u>RESULTATS DE VISUALISATION</u>	<u>5</u>
FONCTIONNALITES PROPOSEES	6
FORMAT DE REPRESENTATION.....	6
<u>FIGURES.....</u>	<u>7</u>

Introduction

Pour ce projet de visualisation de données, j'ai choisi le jeu de données des élections présidentielles de 2017, disponible sur :

<https://www.data.gouv.fr/fr/pages/donnees-des-elections/>

L'ensemble des données représente un volume d'environ 4-5 Mo. J'ai choisi de représenter ces données sous l'angle géographique. J'ai donc complété le jeu de données avec celles de

<https://github.com/gregoireddavid/france-geojson>

qui contiennent les contours des régions, départements, cantons et communes. Enfin, les données de

<https://www.data.gouv.fr/fr/datasets/contours-detaillies-des-circonscriptions-des-legislatives/>

ajoutent à cela les circonscriptions. L'ensemble des données géographiques représente un volume d'environ 100 Mo.

Le projet est partagé sur un git :

<https://github.com/LouisProffitX/INF552.git>

L'intégralité des données, avant le processing y figure, n'hésitez pas à le consulter.

L'objectif de ce projet est moins de travailler l'implémentation que d'optimiser les choix des méthodes de visualisation. Je présenterai donc brièvement la structure de mon implémentation avant de justifier les choix techniques et les résultats obtenus.

Structure du code

Data processing

Les données de vote (répartitions des voix) apparaissent dans les sources sous une forme brute très peu propice au traitement (.xlsx, tableaux hétérogènes, pas de normalisation des noms...) La partie processing les renormalise. Les outils utilisés sont Make et python. Le Makefile du dossier **data/pipeline** permet d'effectuer tout le traitement.

En sortie de pipeline, les données de tous les niveaux hiérarchique sont sous la même forme : un tableau csv par niveau hiérarchique et par tour, une ligne par candidat sur une unité géographique. Par exemple, une ligne est consacrée au résultat de E. Macron dans la région Nouvelle-Aquitaine, une autre pour son résultat en Ile de France, une autre encore pour le résultat de Mme Le Pen dans la région Nouvelle-Aquitaine.

Cette représentation est préférée à une représentation avec une ligne par unité géographique ou une ligne par candidat, car elle favorise les opérations de « join » exécutées par le code JavaScript.

HTML – d3.js

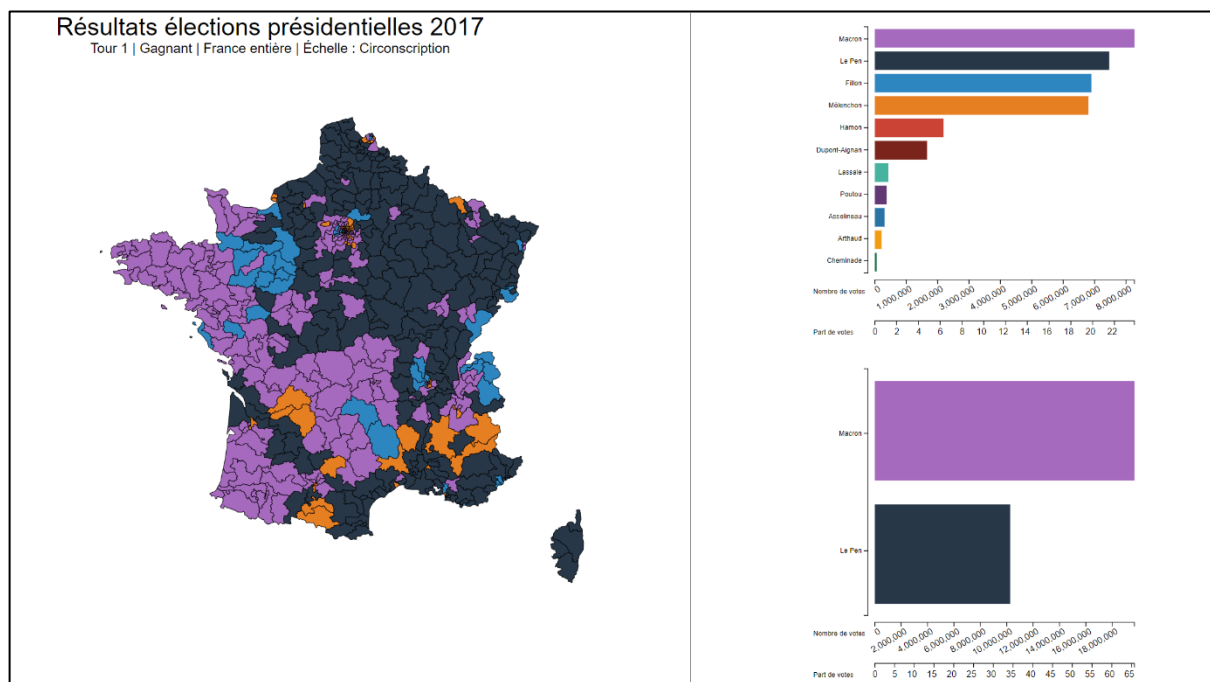
La partie visualisation s'appuie sur la bibliothèque d3.js v.7. Le code spécifique au projet est divisé en deux fichiers : **data.js** et **graphic.js**. Comme leurs noms l'indiquent, le premier contient les fonctions spécifiques au traitement des données, tandis que le deuxième contient les fonctions d'affichage, de visualisation à proprement parler.

Pour les deux fichiers, les constantes apparaissent en haut, puis les fonctions sont écrites selon leur ordre d'importance et d'utilisation (chronologique).

Le volume total de données utilisé rend l'initialisation relativement lente. Pour cette raison que les données géographiques ont été compressées, ce qui permet de gagner en réactivité.

Fonctionnalités

L'écran de visualisation prend la forme suivante :



La carte représente la répartition géographique des résultats. Le graphique en haut à droite représente la répartition par candidats des résultats au premier tour, tandis que le graphique en bas à droite représente cette même répartition au second tour.

L'utilisateur peut utiliser certaines commandes pour faire évoluer l'affichage, les paramètres sont le **mode**, le **filtre**, le **highlight**, le **tour** et le **niveau**.

Le mode

Il existe trois modes : **global**, **candidate** et **conflict**.

- Le mode **global** permet de visualiser sur chaque zone le vainqueur de la zone.
- Le mode **candidate** permet de visualiser un candidat en particulier et de visualiser la répartition géographique de ses voix.
- Le mode **conflict** permet de comparer les résultats de deux candidats, via la différence de leur proportion de voix. Cela utilise une échelle de couleur divergente, avec le blanc (milieu), initialisé à la médiane.

Les transitions entre les modes se font de la manière suivante :

- Pour passer de **global** à **candidate**, cliquer sur la barre d'un candidat (pour M. Macron et Mme Le Pen, les deux barres conviennent).

- Pour passer du mode **candidate** au mode **conflict**, cliquer sur un second candidat (même remarque pour M. Macron et Mme Le Pen).
- En mode **conflict**, cliquer sur un candidat non déjà sélectionné vous fait rester en mode **conflict**, mais change le plus récent des candidats sélectionnés.
- Pour passer du mode **conflict** au mode **candidate**, cliquer sur un des deux candidats sélectionnés, l'autre sera conservé.
- Pour passer du mode **candidate** au mode **global**, cliquer sur le candidat sélectionné.

Le filtre

Le filtre permet de visualiser une partie seulement de la zone géographique. Il est accessible depuis les niveaux **departement** et **region**, à l'aide d'un double clic, lorsque vous filtrez une zone, le niveau hiérarchique descend, afin de ne pas avoir à l'écran qu'une seule unité géographique. Si vous êtes en mode **circonscription** ou **canton**, un double clic appliquera le zoom sur votre zone mais ne la filtrera pas.

Ex : Vous êtes au niveau **region**, et vous faites un double clic sur la région Centre. Le reste de la France disparaît, et le niveau descend à **departement**.

Le filtre met à jour les barres à droite de l'écran, en ne présentant que les résultats de la zone filtrée.

La touche 'f' remet à zéro le filtre.

Le highlight

Il n'est pas possible de filtrer une circonscription ou un canton, car vous auriez alors à l'écran une seule zone géographique. Si vous souhaitez consulter les résultats d'un canton ou d'une circonscription, vous pouvez néanmoins utiliser le **highlight**. Avec un simple clic sur une unité géographique, les autres zones deviennent légèrement transparentes pour concentrer l'attention sur la zone choisie. Le highlight fait évoluer les barres à droite de l'écran en ne présentant que les résultats de la zone sélectionnée. Bien que le highlight serve en particulier pour les niveaux **circonscription** en **canton**, il est également accessible au niveau **departement** et **region**.

Le highlight est compatible avec le filtre (vous pouvez highlight une sous-zone d'une zone que vous avez filtrée)

La touche 'f' remet à zéro le highlight.

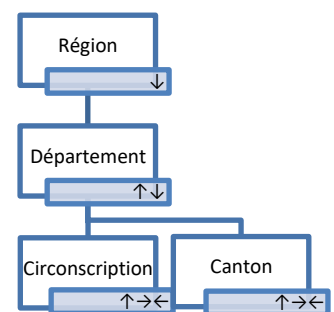
Le tour

Le tour sélectionné par défaut pour la visualisation sur la carte est le premier. En cliquant sur la touche 't', on peut passer de l'un à l'autre. Si vous êtes actuellement au premier tour, en mode conflict ou candidate, et que l'un des candidats sélectionnés n'était pas au second tour, vous ne pouvez pas changer de tour (pour des raisons logiques évidentes).

Le niveau

Il y a quatre niveaux hiérarchiques : **region**, **departement**, **circonscription** et **canton**.

On peut passer d'un niveau hiérarchique à un autre avec les flèches du clavier. Comme il n'y a pas de relation de hiérarchie entre circonscriptions et cantons, on utilise les flèches horizontales pour passer de l'un à l'autre



Résultats de visualisation

Pour ce qui est de la visualisation en elle-même, je me suis fixé trois objectifs :

- Pouvoir visualiser l'intégralité des données disponibles : chaque case des tableaux source doivent être lisibles pour l'utilisateur.
- Proposer un visuel synthétique : deux ou trois blocs doivent suffire.
- Permettre une certaine forme d'analyse : le jeu de données est extrêmement descriptif, et les possibilités de corrélérer ou combiner les variables sont très faibles. La visualisation doit permettre de tirer autant de constats que possible des données.

Fonctionnalités proposées

Le premier objectif est atteint par les fonctionnalités de **focus** et de **highlight**. Grâce à elles, on peut visualiser le nombre et la proportion de voix de chaque candidat dans chaque zone. Chaque ligne des tableaux d'entrée est donc susceptible d'être consultée.

Le troisième objectif est atteint grâce aux modes **conflict** et **candidate** :

Le mode **candidate** permet d'étudier la répartition géographique des votes des candidats. Consulter un candidat indépendamment des autres permet d'avoir autant de pertinence pour M. Macron que pour M. Cheminade, malgré le fait qu'ils se trouvent sur un ordre de grandeur différent. Ce mode de visualisation permet d'avoir une bonne représentation des types de vote également : on constate facilement que les votes de M. Macron sont très citadins, que ceux de M. Lassale sont largement concentrés dans sa région d'origine...

Le mode **conflict** permet d'opposer la répartition géographique des votes de candidats en faisant abstraction de leur différences de résultats totaux. On peut facilement comparer M. Macron et M. Cheminade sans que le résultat soit aberrant. Cette comparaison est d'autant plus pertinente dans le duel Macron / Le Pen, car on peut également observer l'évolution de ce duel entre les deux tours.

Format de représentation

L'utilisation d'une carte pour représenter les données géographiques est inévitable, et la projection n'influence pas vraiment le résultat. Le graphique par barres, quant à lui, comporte deux avantages :

- La simplicité : c'est la manière la plus concise de représenter l'information. La double échelle horizontale permet de visualiser avec une seule barre le nombre de votes et la proportion.
- Grâce aux transitions, il permet de visualiser les évolutions : d'une région à une autre ou d'un tour à l'autre.

Les couleurs choisies sont celles fournies par le ministère de l'Intérieur pour la campagne. Des médias comme France Info en ont choisi d'autres, plus claires, mais qui ne conviennent pas. En effet, les modes **conflict** et **candidate** nécessitent d'établir une échelle entre la couleur du candidat et le blanc, et donc que la couleur soit relativement contrastée.

Figures

Dans l'ordre, de gauche à droite et de bas en haut : mode gagnant et échelle région, mode gagnant et échelle canton, mode candidat (Macron) et échelle région, mode candidat (Le Pen) et échelle canton, mode conflit (Macron et Le Pen) et échelle région, mode conflit (Fillon et Lassalle) et échelle circonscription.

