



Visualisation de données

Professeur Romain VUILLEMOT

Auteurs Mathieu PERRIN
Thibault POURBAIX
Lancelot PREGNIARD

30 janvier 2020

Source des données

Les données proviennent de : <https://www.data.gouv.fr>. Ce jeu de données a été mis à disposition par RTE, l'entreprise responsable de réseau électrique français. Chaque quart d'heure depuis le premier décembre, RTE mesure la puissance moyenne consommée en France, la puissance moyenne apportée par différentes sources d'énergie, et la puissance moyenne exportée vers plusieurs pays.

Forme des données

Les données sont disponibles sur internet sous forme d'un tableau Excel, dans lequel chaque ligne correspond à un quart d'heure et chaque colonne recense une feature (par exemple la consommation moyenne ou bien l'apport du nucléaire).

Transformations des données

Pour utiliser les données dans d3, on transforme ce tableau excel en un fichier `.csv`. On pourra éventuellement réduire la taille du jeu de données du point de vue du nombre d'échantillons ou du point de vue du nombre de features pour certaines applications.

La date et l'heure étant spécifiées sur 2 colonnes, il sera nécessaire d'adapter le Parsing de la date pour que l'ensemble fonctionne. Enfin, on pourra appliquer des transformations mathématiques de type ACP pour obtenir de nouveaux descripteurs, de sorte à visualiser une représentation plus parcimonieuse de la série temporelle.

Outils utilisés

Le premier outil est Excel, car c'est lui qui permet la visualisation préalable des données. On utilisera probablement python et la librairie scikit-learn pour l'ACP. Enfin, on utilisera `blockbuilder.org` pour la visualisation des données avec d3.js.

Maquette

Voici une maquette contenant l'idée principale de graphique que nous souhaiterions réaliser lors du projet :

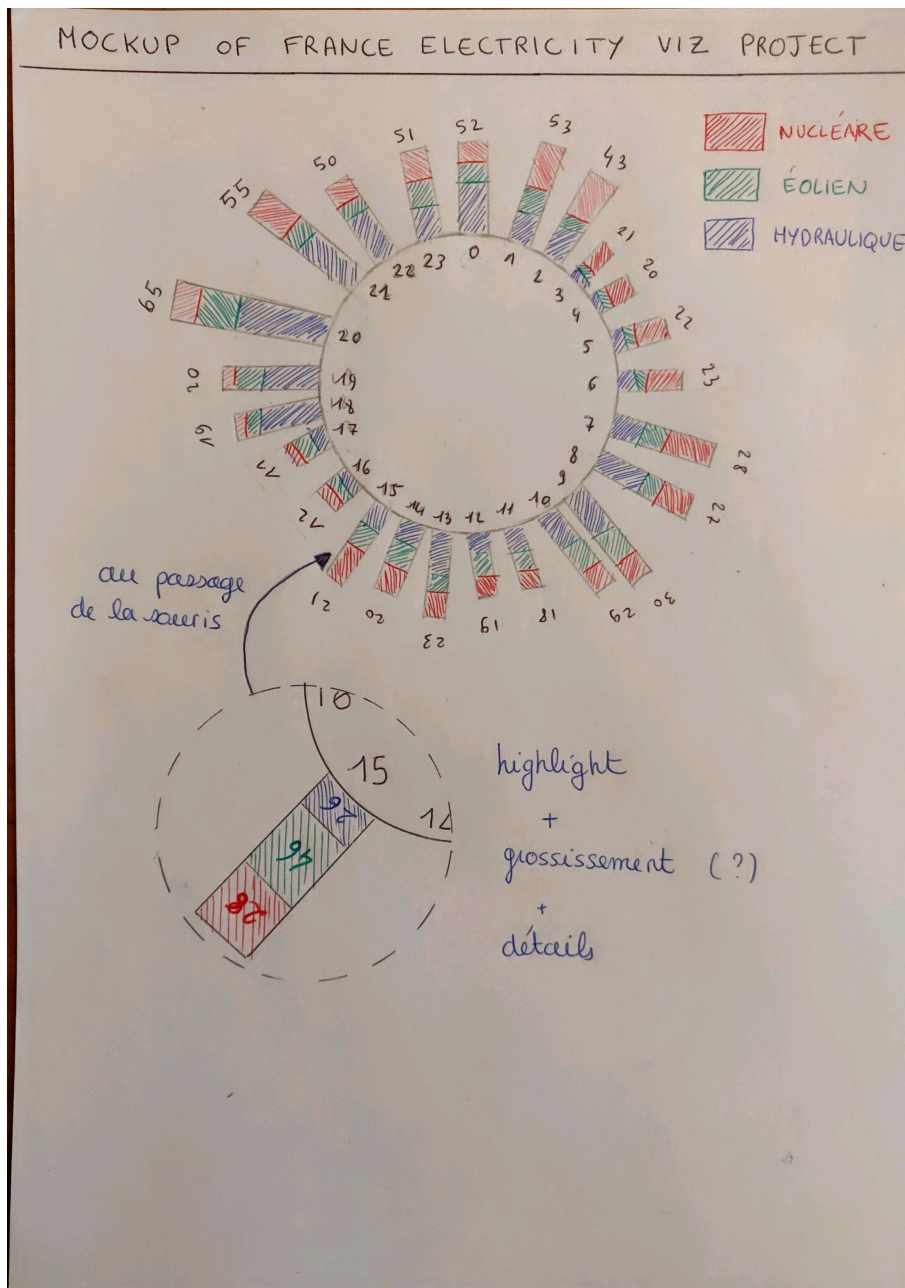


FIGURE 1 – Maquette