

TP1 : Manipulation de données avec R

Yannig Goude : yannig.goude@math.u-psud.fr

Projet Machine Learning & prévision, M2 MDA & StatML

Exercice 1 :

Charger les packages tidyverse et weathermetrics.

1. Importer les données de température au parc montsouris <https://www.math.u-psud.fr/~goude/Materials/Data/montsouris.csv> et de consommation française d'électricité https://www.math.u-psud.fr/~goude/Materials/Data/cdc_conso.csv.
2. Représenter la consommation française au pas demi-horaire en fonction du temps sur toute la période, pour le mois de février 2012.
3. Calculer le max de consommation. Quand a t-il été atteint ?
4. Calculer la moyenne de la consommation électrique pour chaque jour de la semaine puis pour chaque demi-heure de la journée. Représenter ces informations sur un graphique adéquate.
5. Convertir les températures en degrés celsius.
6. Représenter la température moyenne au parc montsouris au pas journalier en fonction du temps sur toute la période. Superposer la température min et max de la journée. Faire le même graphique pour le mois de février 2012.
7. Calculer le min de température. Quand a t-il été atteint ?
8. Construire une table contenant la date, les variables météo et la consommation électrique au pas journalier.
9. Représenter le nuage de point de la consommation en fonction de la température.

Exercice 2 : web scraping

Charger les packages Rcurl et XML

1. Importer les données des sondages effectués en france pour le premier tour des élections à l'adresse : https://en.wikipedia.org/wiki/Opinion_polling_for_the_French_presidential_election,_2017

2. nettoyer ces données de manière à obtenir une table comprenant en colonnes la date des sondages et les candidats, en ligne les intentions de vote obtenues.
3. A l'aide des fonctions `as.POSIXct` et `strptime` convertir la colonne `Date` en un format `date`.
4. Représenter sur un même graphique les intentions de vote de chaque candidat en fonction du temps. Représenter chaque candidat par une couleur distinct. Superposer le nom des candidats.
5. Proposer une méthode pour estimer la "tendance" des intentions de vote de chaque candidat et la superposer au graphique.