

# TD/TP 1 : Mécanismes de données manquantes

M2 Stats de la SD, 2025-2026

Hadrien Lorenzo

## ⚠ Avertissement

Le travail est à réaliser en trinômes et chaque réponse doit être consciencieusement justifiée.  
Un fichier compilé (pdf) est à envoyer en fin de cours à l'adresse :

**hadrien.lorenzo@univ-amu.fr**

Le nom du fichier doit être : TD\_TP1\_NOM1\_NOM2\_NOM3.pdf

Bonne chance !

Nous allons nous concentrer sur des variations d'un jeu de données, **ozone**, décrit par les 13 variables suivantes. Les 13 variables recueillies sont :

- Max03 : Maximum de concentration d'ozone observé sur la journée
- T9, T12, T15 : Température observée à 9, 12 et 15h
- Ne9, Ne12, Ne15 : Nébulosité observée à 9, 12 et 15h
- Vx9, Vx12, Vx15 : Composante E-O du vent à 9, 12 et 15h
- Max03v : Teneur maximum en ozone observée la veille
- vent : orientation du vent à 12h
- pluie : occurrence ou non de précipitations

Dans chacun des cas, proposer des visualisations, via le package VIM par exemple, afin d'étayer votre propos. Regardez aussi les fonctions **aggr**, **marginplot** et **matrixplot**.

Ouvrir le fichier **ozone\_1.csv**, conjecturer une source éventuelle de données manquantes et la classer parmi MCAR, MAR ou MNAR. Faire de même avec les fichiers **ozone\_2.csv**, **ozone\_3.csv** et **ozone\_4.csv**.