

CluZzy : enquête aux frontières du flou

Alisson Baron*

Uranie Jean-Louis†

Résumé (max 300 mots)

Le Mystère du Domaine Moutarde est un jeu pédagogique interactif, développé avec le package `{learnr}`, qui permet d'apprendre les jointures floues en R avec le package `{fuzzyjoin}`, sans nécessiter d'ouvrir R. Dans une approche inspirée du Cluedo, les participants incarnent des agents du Bureau d'Investigation des Données Anormales, chargés de résoudre la disparition mystérieuse du Dr Moutarde. Les données proposées sont volontairement brouillées : doublons, fautes d'orthographe, dates approximatives et entrées contradictoires. Pour progresser, les joueurs doivent recouper ces fichiers avec des bases de référence afin d'identifier le coupable, le lieu et le moment exact de l'incident. Le scénario narratif immersif transforme un exercice technique en une expérience ludique et engageante, tout en renforçant la compréhension de concepts clés de manipulation de données. Ce jeu peut être utilisé en présentiel ou à distance et combine analyse de données imparfaites et esprit critique. Les joueurs développent simultanément leurs compétences en R, en logique de croisement d'informations et en résolution de problèmes. Cette démarche démontre comment rendre l'apprentissage de concepts techniques à la fois interactif et motivant, tout en favorisant la mémorisation et l'application pratique.

Mots-clefs (3 à 5) : Data - Package - Enseignement

Développement

Le jeu est implémenté sous la forme d'un module `learnr` structuré en quatre chapitres, chacun correspondant à un type de jointure floue proposé par le package `{fuzzyjoin}`. L'objectif est de permettre aux participants de manipuler et comparer différents types de données imparfaites, dans un cadre narratif unique et cohérent.

Après un rappel des principes fondamentaux des jointures classiques (clé exacte, correspondance stricte), le module introduit la notion de jointure floue : une méthode de rapprochement de données reposant non plus sur l'égalité parfaite, mais sur des critères de similarité, de distance ou de tolérance à l'erreur. Cette approche reflète des situations fréquentes en analyse de données réelles, où les informations sont incomplètes, bruitées ou hétérogènes.

Chaque chapitre met en œuvre une fonction spécifique du package `{fuzzyjoin}`, associée à un type d'indice narratif :

- `fuzzy_join()` : jointures basées sur des fonctions logiques décrivant si deux vecteurs peuvent être appariés (dates approximatives, intervalles, seuils) ;
- `stringdist_join()` : jointures floues sur des chaînes de caractères (noms mal orthographiés, variations lexicales) ;
- `distance_join()` : jointures fondées sur une métrique de distance numérique ;
- `geo_join()` : jointures reposant sur des distances géographiques à partir de coordonnées latitude/longitude.

Les chapitres peuvent être explorés dans un ordre libre, chaque section apportant un indice distinct (lieu, date, objet, protagoniste). Contrairement au Cluedo classique, il ne s'agit toutefois pas de scénarios

*Institut de l'élevage, alisson.baron@idele.fr

†Institut de l'élevage, uranie.jean-louis@idele.fr

multiples. En effet, l'enquête repose sur une seule et même histoire, dont la résolution finale n'est possible qu'en recoupant l'ensemble des informations.

Afin de renforcer l'engagement, le jeu intègre également une contrainte temporelle : les participants disposent de 1h30 pour résoudre l'enquête avant la disparition annoncée des données. Cette échéance relève toutefois du cadre narratif et sert de moteur ludique ; le module reste accessible au-delà du temps imparti. Cette mécanique, inspirée de la dynamique de progression observée dans certains jeux d'investigation populaires, introduit une tension maîtrisée qui favorise la concentration, la prise de décision et l'apprentissage actif.

Le Mystère du Domaine Moutarde favorise l'autonomie et l'esprit critique, tout en ancrant des concepts techniques dans une narration engageante et progressive.