

Cadre R chez IMPACT Initiatives

Yann Say*

Résumé (max 300 mots)

IMPACT Initiatives est une O.N.G. qui travaille dans plus de trente pays dans le domaine de la recherche appliquée au secteur humanitaire et de développement. Afin d'améliorer l'efficacité et la transparence du nettoyage et analyse des données, un cadre d'analyse avec R a été créé. Ce cadre permet une harmonisation des processus d'analyse et de garder une modularité et flexibilité qui permet d'adapter les processus à chaque contexte et à chaque équipe de travail. Ce cadre est composé de 4 étapes (nettoyage, composition, analyses et présentation), 4 verbes (vérifier, ajouter, créer et revoir) et 2 adjectifs (indépendant et "pipe-able"). Quatre packages R ont été créés pour accompagner le cadre et sont dans une version beta. Le déploiement du cadre et des outils est en cours au sein de l'organisation.

Mots-clefs (3 à 5) : Data - Humanitaire - Process

Développement

IMPACT Initiatives est une O.N.G. qui travaille dans plus de trente pays dans le domaine de la recherche appliquée au secteur humanitaire et de développement. Les projets de recherche sont mis en oeuvre principalement par les équipes basées sur le terrain, de l'élaboration des questions de recherche à la dissémination des résultats. Les équipes basées à Genève ont pour mission de fournir un soutien et des conseils (y compris la revue et la validation de tous les résultats pertinents) aux équipes pays sur tous les aspects de la mise en oeuvre du cycle de recherche. En 2021, 80% du nettoyage et 50% des analyses utilisaient des scripts personnels aux analystes. Afin d'améliorer l'efficacité et la transparence du nettoyage et de l'analyse des données, un cadre d'analyse avec R a été créé. Le déploiement de ce cadre et ses outils a commencé au début de l'année 2024.

Le cadre d'analyse R chez IMPACT est un cadre modulaire à deux dimensions :

- une dimension horizontale qui se concentre sur le résultat d'une étape, et
- une dimension verticale qui se concentre sur le contenu d'une étape.

Le cadre est construit autour de :

- 4 étapes
 - Nettoyage : Toute manipulation pour passer des données brutes aux données propres
 - Composition : Toute manipulation avant l'analyse, par ex. ajout d'indicateurs, combiner deux bases de données (avant et après une distribution). Les indicateurs sont créés avant l'analyse, et non avec l'analyse.
 - Analyse: Toute manipulation concernant uniquement l'analyse. L'analyse s'arrête à la table des résultats. La table de résultat est format long : statistique + clé d'analyse.
 - Présentation : Toute manipulation pour formater la table des résultats.
- 4 verbes :

*IMPACT Initiatives, yann.say@impact-initiatives.org

- Vérifier (`check_*`): Une fonction qui va signaler certaines valeurs (par exemple, valeurs autres à recoder, valeurs aberrantes, etc.). Ces valeurs sont cataloguées dans un journal. Un `check_*` renverra une liste : le jeu de données vérifié et le journal de nettoyage.
 - Ajouter (`add_*`): Une fonction qui ajoute une variable (colonne) au jeu de données. Par exemple, pour ajouter la durée d’une enquête, pour ajouter la catégorie de score de consommation alimentaire, etc.
 - Créer (`create_*`): Une fonction qui crée, transforme quelque chose, par exemple créer un journal de nettoyage avec des actions à faire, créer une table de résultats d’analyse, créer un produit pour une présentation.
 - Revoir (`review_*`): Une fonction qui examinera un objet en le comparant à des normes ou à un autre objet et signalera les différences, par exemple en examinant le nettoyage en comparant le jeu de données brut, le jeu de données propre et le journal de nettoyage, l’analyse le comparant à une autre analyse.
- 2 adjectifs :
 - « pipeable » : Chaque étape, chaque famille de fonctions (verbes) peuvent être utilisées une après l’autre. Cela permet à l’utilisateur.ice de pouvoir adapter chaque étape avec les fonctions spécifiques à son contexte.
 - Indépendant : A chaque étape, le cadre définit l’entrée et la sortie, le “comment” est libre, chaque utilisateur.ice peut donc choisir l’outil qu’il/elle préfère.

La présentation montrera la structure et l’articulation du cadre d’analyse grâce à quatre packages spécifiques : {`cleaningtools`}, {`addindicators`}, {`analysistools`} et {`presentresults`}. Elle présentera, à travers différentes fonctions, comment les verbes et adjectifs sont intégrés dans chacun des packages ainsi que les définitions de format à chaque étape. Un exemple pratique illustrera l’utilisation de ce cadre, en se concentrant sur les aspects suivants :

- L’étape de nettoyage présentera comment créer un journal de nettoyage avec différentes valeurs à vérifier, comment l’exporter, et créer une base de données nettoyer.
- L’étape de composition montrera comment ajouter différents indicateurs.
- L’étape d’analyse présentera comment créer une table de résultats d’analyses descriptives.
- L’étape de présentation montrera comment passer d’une table de résultats à un produit à partager.

Les packages sont disponibles sur GitHub et peut être installé dans R via la commande :

```
devtools::install_github("impact-initiatives/cleaningtools")
devtools::install_github("impact-initiatives/addindicators")
devtools::install_github("impact-initiatives/analysistools")
devtools::install_github("impact-initiatives/presentresults")
```

Le cadre d’analyse R chez IMPACT Initiatives représente une étape significative dans l’efficacité et la transparence du nettoyage et de l’analyse des données au sein de l’organisation. Ce cadre offre une approche modulaire et flexible qui permet de s’adapter aux différents contextes et analyses. Bien qu’il soit encore trop tôt pour évaluer de ce cadre, ces premiers résultats prometteurs soulignent le potentiel de ce cadre pour renforcer les pratiques de recherche et d’analyse au sein de l’organisation, en favorisant la collaboration entre les équipes.