

Suivi de la collecte des données avec R

ONANENA AMANA Jeanne De La Flèche
SOMA Ben Idriss Diloma

27 mars 2025

Introduction

- Suivi de la collecte de données : contrôle de la qualité et de la complétude des données.
- Importance cruciale dans les enquêtes statistiques pour garantir des analyses fiables.

Guide Méthodologique de Suivi d'Enquête avec R

1. Stratégie de Suivi avec R

1.1 Objectifs

- Comprendre l'importance du suivi des enquêtes.
- Explorer des outils R pour le suivi.
- Étudier un cas pratique : suivi d'une enquête sur la sécurité alimentaire.

Fondements théoriques (1/2)

- **Fiabilité** : cohérence des réponses
- **Validité** : correspondance aux réalités observées
- **Précision** : faible niveau d'erreur d'échantillonnage

Fondements théoriques (2/2)

- **Complétude** : taux de réponse satisfaisant
- **Consistance** : absence de contradictions internes
- **Traçabilité** : contrôle de la provenance des données

Méthodologie de suivi

- Utilisation de R pour :
 - Contrôle qualité automatisé
 - Visualisation géographique des points de collecte
 - Calcul d'indicateurs de complétude

Cas Pratique : Suivi d'une enquête sur la sécurité alimentaire

Contexte

Dans ce cas pratique, nous avons développé une application **R Shiny** pour le suivi en temps réel d'une enquête portant sur la **sécurité alimentaire**, réalisée en collaboration avec **KoboToolbox**. L'enquête visait à collecter des informations essentielles concernant la sécurité alimentaire des ménages, y compris :
- **La disponibilité** des produits alimentaires. - **L'accès** aux produits alimentaires, en particulier pour les ménages vulnérables. - **L'utilisation** de ces produits dans la nutrition des ménages.

Cette enquête est d'une grande importance dans les contextes où la sécurité alimentaire est un enjeu majeur, et elle nous permet de recueillir des informations fiables et actuelles grâce à une collecte de données efficace et bien suivie.

L'application R Shiny est intégrée à l'**API de KoboToolbox**, ce qui permet de récupérer les données collectées sur le terrain en temps réel, assurant ainsi un suivi continu de la qualité des informations fournies par les enquêteurs.

Objectifs du Cas Pratique

Le principal objectif de ce cas pratique était de démontrer comment utiliser **R Shiny** pour : 1. **Contrôler la qualité des données** collectées dans l'enquête. Cela inclut la vérification de la complétude des données, des incohérences potentielles, et des erreurs dans les entrées des enquêteurs. 2. **Suivre géographiquement** la collecte des données, en utilisant des cartes interactives pour visualiser la répartition des points de collecte. 3. **Détecter des anomalies** dans les réponses des enquêteurs, telles que des incohérences géographiques ou des erreurs temporelles.

Méthodes utilisées

L'application **R Shiny** a été conçue pour récupérer les données via l'API de KoboToolbox. Une fois les données récupérées, plusieurs étapes ont été mises en place pour analyser et visualiser les informations : 1. **Contrôle de la qualité des données** : L'application vérifie la complétude des réponses, avec des alertes pour les questionnaires partiellement remplis ou erronés. 2. **Visualisation géographique** : En utilisant le package **leaflet** de R, l'application génère des cartes interactives pour localiser chaque point de collecte. Cela permet aux responsables de visualiser rapidement où des erreurs peuvent survenir ou si certaines zones sont sous-représentées. 3. **Détection d'anomalies** : L'application analyse les données pour repérer des incohérences comme des **valeurs aberrantes** ou des **réponses contradictoires**. Par exemple, si un enquêteur signale une **localisation GPS invalide** ou une **plage horaire incorrecte**, l'application le signale automatiquement.

Application concrète de l'outil

Dans un environnement de terrain, les enquêteurs remplissent des questionnaires sur **KoboToolbox** qui sont ensuite automatiquement envoyés à l'application R Shiny. L'application récupère ces réponses en temps réel grâce à l'**API**. À chaque fois qu'une réponse est reçue, l'application : - Vérifie la validité de la localisation. - Vérifie le taux de complétude de chaque réponse. - Vérifie la cohérence des données, comme les dates de collecte par rapport à la période de l'enquête.

Les résultats sont affichés dans des tableaux résumés et sur des cartes interactives, ce qui permet aux responsables de l'enquête de suivre l'état de la collecte et de résoudre rapidement les problèmes rencontrés sur le terrain.

Résultats attendus

L'objectif final est d'assurer un suivi efficace et transparent des données, ce qui permet de : - **Garantir la qualité des données** : La vérification en temps réel permet de détecter et corriger les erreurs immédiatement, réduisant ainsi les risques d'anomalies dans les résultats finaux. - **Améliorer la couverture géographique** : Les cartes interactives permettent d'identifier les zones où la collecte de données est insuffisante ou incomplète. - **Accélérer la prise de décision** : Grâce à l'automatisation du contrôle de la qualité et à la visualisation géographique, les équipes peuvent intervenir rapidement sur le terrain pour corriger toute anomalie.

Conclusion

Ce cas pratique montre l'intérêt de l'intégration de technologies modernes, comme **R Shiny** et **KoboToolbox**, dans le suivi des enquêtes. L'utilisation de R permet de faciliter non seulement le contrôle de la qualité des données collectées, mais aussi d'assurer une gestion efficace des informations sur une plateforme interactive, ce qui est essentiel pour la réussite des enquêtes complexes, en particulier dans les contextes où la rapidité et la précision sont cruciales.