

Utilisation du package `{flexdashboard}` pour le contrôle des données de biologie dans un entrepôt de données de santé

Morgane Pierre-Jean 1*

Guillaume Bouzillé 2†

Résumé

Les résultats des laboratoires hospitaliers constituent une source de données importante dans les entrepôts de données de santé. Pour garantir la comparabilité entre les établissements et pour pouvoir être utilisés dans des études de recherche, les résultats doivent être interopérables. Des événements pouvant perturber la distribution des valeurs pour un même dosage biologique peuvent se produire au cours du temps. Par exemple, de nouveaux équipements peuvent être ajoutés au pipeline d'analyse, un automate peut être remplacé, les formules peuvent évoluer en raison de nouvelles connaissances scientifiques et des terminologies existantes peuvent être adoptées. Nous avons automatisé la production de tableaux de bord pour surveiller ces événements et la qualité des données.

Nous avons utilisé le package `flexdashboard` pour produire des pages html pour chacun des dosages considérés, des méthodes de détection de ruptures automatiques telles que PELT et KernSeg ont été utilisées pour la détection d'événements. Pour un code LOINC donné, nous créons un tableau de bord qui synthétise le nombre de codes locaux, et le nombre de patients (par sexe, âge et service hospitalier) associés au code. Enfin, le tableau de bord permet de visualiser les événements temporels qui perturbent la distribution du signal. Malgré la quantité de données dans l'entrepôt du CHU de Rennes, la génération d'un dashboard est simple et rapide. Les dashboards permettent de vérifier que les valeurs d'un dosage biologique au cours du temps sont stable. Ils permettent également de vérifier si les données de l'hôpital ont bien été remontées et de détecter facilement les problèmes techniques. Un exemple de dashboard produit pour un dosage est disponible à cette adresse : <https://mpierrejean.github.io/MIE2024/>.

Mots-clefs : Data – Santé – Reporting – Monitoring – `flexdahboard`

1 Contexte

L'entrepôt de données de santé Madec et al. [2019] du CHU de Rennes permet de réaliser un grand nombre d'études de recherche à l'aide des données produites dans les unités de soin. Les données de biologie constituent une grande source de données pour les entrepôts de données de santé (EDS). Cependant, les données sont dépendantes des technologies de production comme les automates, les nouvelles formules, les changements de réactifs... Ces changements n'affectent pas les soins délivrés aux patients, mais lorsque ces données sont réutilisées en recherche, ils peuvent avoir un impact significatif. Le but de ce projet est de produire très rapidement un monitoring de la qualité des données biologiques disponibles dans l'entrepôt.

2 Méthodes

Nous avons utilisé le package `flexdashboard` Aden-Buie et al. [2023] qui nous a permis de générer des pages html interactives pour parcourir différents onglets qui correspondent à différents types d'indicateurs de qualité. La génération est automatisée à l'aide de paramètres : paramètre du code biologique à explorer et connexion à notre base de données. Nous avons également utilisé la librairie `plotly` Sievert [2020] afin de rendre certains graphiques interactifs.

*Univ Rennes, CHU Rennes, INSERM, LTSI-UM R 1099, morgane.pierre-jean@chu-rennes.fr

†Univ Rennes, CHU Rennes, INSERM, LTSI-UM R 1099, guillaume.bouzille@chu-rennes.fr

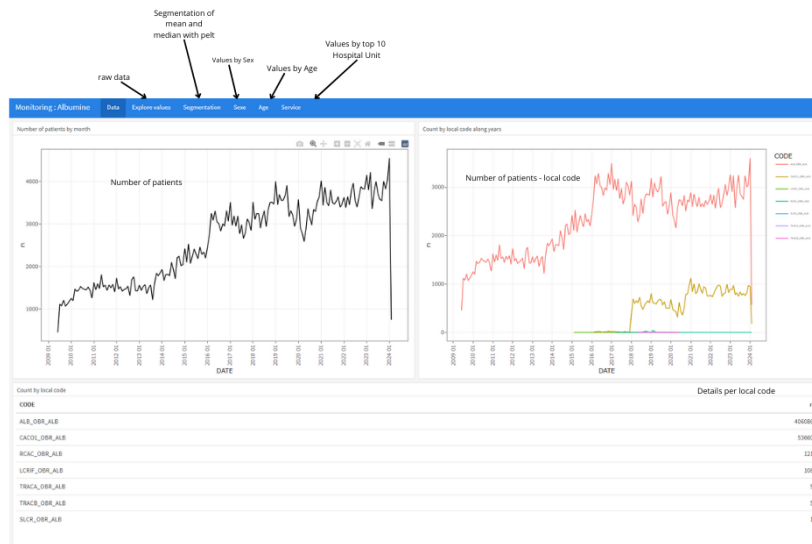


FIGURE 1 – Dashboard annoté

3 Résultats

Le workflow de génération contient, un fichier template Rmd pour le dashboard et un fichier R qui permet de générer autant de dashboards que de dosages biologiques que l’on veut explorer.

Le dashboard contient donc plusieurs onglets afin de pouvoir explorer le dosage à différents niveaux et détecter de potentielles anomalies. On peut donc explorer le nombre de valeurs produites au cours du temps, les valeurs brutes, la distribution des valeurs par année, l’évolution de la moyenne et de la médiane, une exploration est possible par sexe, par âge et par unité de soin 1. Un exemple de dashboard produit pour un dosage est disponible à cette adresse : <https://mpierrejean.github.io/MIE2024/>.

Grâce à ces dashboards, nous avons pu facilement détecter plusieurs types d’anomalies. Les anomalies au niveau des données brutes générées sont souvent des changements de technique de dosage ou des changements de machine. Les anomalies au niveau du nombre de données générées peuvent être dues à un problème technique ou un changement de pratique (le dosage n’est plus prescrit automatiquement).

4 Conclusion

La génération de dashboards avec les libraries R disponibles est simple et rapide. Nous avons pour projet de déployer ces dashboards en production afin de détecter des problèmes de qualité dans les données au fil de l’eau.

Références

Garrick Aden-Buie, Carson Sievert, Richard Iannone, JJ Allaire, and Barbara Borges. *flexdashboard : R Markdown Format for Flexible Dashboards*, 2023. <https://pkgs.rstudio.com/flexdashboard/>, <https://github.com/rstudio/flexdashboard/>.

Julia Madec, Guillaume Bouzillé, Christine Riou, Pascal Van Hille, Christian Merour, Marie-Lisen Artigny, Denis Delamarre, Veronique Raimbert, Pierre Lemordant, and Marc Cuggia. ehop clinical data warehouse : From a prototype to the creation of an inter-regional clinical data centers network. *Studies in health technology and informatics*, 264 :1536–1537, 2019.

Carson Sievert. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC, 2020. ISBN 9781138331457. URL <https://plotly-r.com>.