

R est-il un langage adapté à la production ?

Lino Galiana¹

Résumé (max 300 mots)

R est un langage initialement fait par et pour une communauté scientifique de statisticiens amenés à appliquer leur analyse sur des jeux de données relativement structurés. Les choix structurants faits pour répondre à ce besoin, tels que la place centrale donnée au *dataframe* ou l'usage interactif permettant un dialogue avec des jeux de données sont particulièrement intéressants dans un contexte où l'analyse de données s'est propagée bien au-delà de cette communauté initiale.

Néanmoins, ces principes qui font le coeur du langage R et qui ont permis d'étendre sa communauté d'utilisateurs ne deviennent-ils pas des freins lorsqu'on bascule dans une optique de production où l'intervention humaine est limitée, où le code doit fonctionner dans un environnement minimal et où le livrable attendu ne se résume plus forcément à une figure ou un tableau construit sur mesure ?

La reproductibilité est l'une des valeurs cardinales de la communauté R, à l'origine d'innovations majeures comme R Markdown. L'ouvrage récent de H. Wickham et K. Masiello [1] illustre l'évolution des enjeux spécifiques au passage de l'expérimentation à la production dans un contexte de chaînes de production valorisant des données.

Cette présentation abordera le défi du passage de l'interactif à la production et interrogera l'adéquation des pratiques et outils de l'écosystème R à ces contraintes. Elle proposera enfin une mise en perspective avec Python, très présent dans l'industrie sur les questions de production et dont l'écosystème très dynamique permet l'émergence régulière d'outils ou de pratiques favorisant la mise en production.

Mots-clefs (3 à 5) : Production - Data - Reproductibilité

Bibliographie

[1] H. Wickham et K. Masiello, « R in Production ».

¹Insee, lino.galiana@insee.fr