

`{crosstable}` : décrivez vos datasets en quelques lignes

Dan Chaltiel*

Résumé (max 300 mots)

C'est le quotidien de presque tous les statisticiens et de beaucoup de data-scientists, il faut toujours et encore décrire ses données. Avec `{crosstable}`, il est possible de décrire complètement un dataset entier en quelques lignes de code seulement. Mieux, il existe une myriade d'options pour que le résultat soit le plus personnalisé possible. Le résultat est une dataframe classique, mais tout est pensé pour facilement la transformer en table exportable (`flextable`, `gt`, Excel, Word...). Il est possible de pivoter et de transposer les crosstables, ce qui facilite grandement la description de certaines bases. `{crosstable}` intègre également plusieurs helpers pour utiliser les options, gérer les labels de colonnes, et s'interfacer avec le package `{officer}`. La documentation est très détaillée et disponible sur GitHub. S'il existe déjà plusieurs packages pour l'analyse descriptive, notamment `{gtsummary}` et `{table1}`, `{crosstable}` se pose dans la complémentarité en termes d'interface (mais aussi parfois de performances). Testez-les tous et choisissez votre préféré !

Mots-clefs : Package – Statistique – épidémiologie – Biostatistique – Description – Data – Dataset

*Institut Gustave Roussy, dan.chaltiel@gustaveroussy.fr