



VISUALISATION DE DONNÉES ISSUES DE BASES MONGODB

Romain Tavenard

1. Organisation

- Groupes de 2 étudiant(e)s
- Rendu final : au plus tard le vendredi 29/04/2022, 23h59
- Exemple de notebook `R` hébergé sur Google Colab pour celles et ceux qui auraient des soucis pour interroger les base de données hébergées sur MongoDB Atlas depuis leur machine (sous R) : [ici](#)

2. Contenu du projet

Dans ce projet, vous serez amenés à manipuler des données issues de bases MongoDB pour générer des visualisations. Trois visualisations sont demandées et vous devrez en réaliser au moins une en R et au moins une en Python. Le rendu se fera sous la forme d'une page web hébergée sur GitHub Pages. ⚠ Attention, tout *commit* dont la date est postérieure à la date limite de rendu ne sera pas pris en compte dans l'évaluation de votre travail. ⚠

2.1. Centres de vaccination Covid19

La base `doctolib` contient des informations (récupérées en janvier 2022) relatives aux centres de vaccination contre la Covid19 en Bretagne. Vous devrez pour cette première visualisation générer une carte des centres de vaccination situés à moins de 50km de Rennes. L'icône associée à chaque centre sera de couleur rouge, orange ou vert selon le nombre de créneaux de vaccination ouverts sur la période du 26 au 29 janvier 2022 (inclus).

Exemple de bonus. Vous pourrez ajouter une seconde visualisation qui se concentrera sur les vaccinations pour première dose (voire même, vous pourrez ajouter un bouton qui bascule d'une visualisation à l'autre lorsque l'utilisateur clique dessus). **Mise à jour :** Pour ce bonus, vous vous concentrerez sur la période 1er janvier - 1er juin au lieu de la période citée plus haut.

2.2. Réseau de publications scientifiques

La base `publications` contient les informations relatives aux publications de scientifiques du laboratoire IRISA pour l'année 2021 (extraites du service HAL). Visualisez les liens entre les auteurs de ces publications, en utilisant un code couleur qui permette de distinguer les auteurs par leurs nombres de publications et en représentant les liens (co-publications) existant entre les auteurs. Le nombre d'auteurs présents dans la base étant très grand, vous vous focaliserez sur les 20 auteurs les plus prolifiques (*i.e.* qui ont participé à l'écriture du plus grand nombre d'articles) et votre requête MongoDB devra récupérer pour chacun de ces auteurs la liste de ses publications (la suite du traitement qui consiste à calculer le nombre d'articles commun par paire d'auteurs pourra être réalisée dans votre script).

Exemple de bonus. Vous pourrez faire en sorte que l'épaisseur des traits joignant les auteurs soit proportionnelle au nombre de publications communes.

2.3. Visualisation libre de données issues de NYfood

La base `NYfood` contient des informations relatives à des restaurants de New-York. Pour ce troisième exercice, vous êtes libres de visualiser les données contenues dans cette base comme bon vous semble.

3. Évaluation

Pour l'évaluation, il sera porté une attention particulière à ce que :

1. votre code minimise la quantité de calculs à faire par votre machine (en Python ou en R) et que, au contraire, la majeure partie des calculs, sélections de données, *etc.*, soit intégrée dans vos requêtes MongoDB ;
2. votre code minimise la quantité de données à faire transiter du serveur de bases de données à votre machine (vos requêtes MongoDB devront donc synthétiser l'information de manière à ne récupérer que les informations nécessaires à vos visualisation).

De plus, les éléments notés "Bonus" plus haut n'en sont pas vraiment : il n'est pas envisageable d'obtenir la note maximale en se contentant d'implémenter le "but principal". Typiquement, sur les 20 points de la note finale, 6 seront dédiés à la qualité et l'originalité des bonus proposés.