# Exercices d’exploration et visualisation des données

### Objectifs :

* L’objectif de cet exercice est d’évaluer :
  + Vos compétences de développement
  + Votre capacité à comprendre les données que vous exploitez

### Introduction :

Le Résumé de Passage aux Urgences (RPU) est un recueil standardisé en France des données médicales issues des Services d'Accueil des Urgences, en milieu hospitalier. Le SESAN est responsable de cette base de données pour toutes les structures hospitalières de la région Île-De-France. Lors d’un passage aux urgences, un patient se voit créer un RPU possédant un identifiant unique. Cet RPU contient plusieurs éléments et parmi eux :

* L’id du RPU
* le code postal du patient,
* la date de naissance,
* le mode d'entrée : mutation, transfert ou domicile
* la provenance : unité de soin de provenance du patient
* le transport : transport utilisé pour la venue (personnel, hélicoptère, ambulance, SMUR, etc.)
* le motif : motif de recours aux urgences pour la prise en charge du patient
* la CCMU : état du patient et pronostic médical (Classification Clinique des Malades aux Urgences)
* le diagnostic principal : diagnostic principal de la CIM (classification internationale des maladies)
* la date de sortie : date et heure de sortie du patient
* le mode de sortie : mutation, transfert, domicile ou décès avec la Classification du Groupe d’Etude Multicentrique des Services d’Accueil (GEMSA)
* la destination : unité de soin de destination du patient
* l’orientation (orient) :précise le devenir ou les circonstances associée

Dans la base de données SESAN, la table contenant les ID des RPU est la table **TF\_URG\_RPU**. Vous verrez dans le schéma relationnel (annexe 1), plusieurs variables finissant par \_HIS ou \_QUA, ces variables sont à ignorer dans l’exercice 1. Ainsi par exemple, si vous devez utiliser la variable sexe, vous utiliserez COD\_SEX et non pas COD\_SEX\_HIS ou COD\_SEX\_QUA. Enfin, lorsque l’on parlera de date, c’est à la variable **DAT\_ENT** que l’on fera référence, les autres variables dates sont à ignorer.

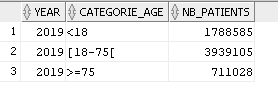
### Exercice 1 : Analyse d’un schéma de base relationnel

Comme indiqué précédemment, le schéma en annexe 1, représente un extrait de la base de données relationnelles des RPU. La description des colonnes de chaque table est fournie dans le fichier Excel *Description\_tables.xlsx.* L’objectif de cet exercice est d’analyser le schéma afin de fournir les requêtes adéquates.

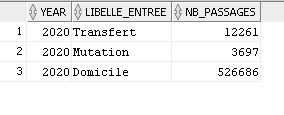
1. En fonction des colonnes présentes dans chacune des tables, merci de donner la liste des clés primaires associées à leur nom de table.

|  |  |
| --- | --- |
| Nom de table | Clé primaire |
|  |  |

1. En fonction du schéma, de la description des colonnes de chaque table, quelle serait selon vous la différence entre la table en rouge et celles en bleues ?
2. Pouvez-vous écrire la requête SQL permettant, de donner le nombre de patients étant passés aux urgences en 2019 selon les catégories d’âge suivantes : <18ans, [18-75[ ans et ≥75 ans ? La requête doit générer la réponse suivante :



1. Pouvez-vous écrire la requête SQL permettant de détailler les différents modes d’entrées des patients de plus de 75 ans étant allés aux urgences en 2020 (nombre de passages par Libellé d’entrée) ?

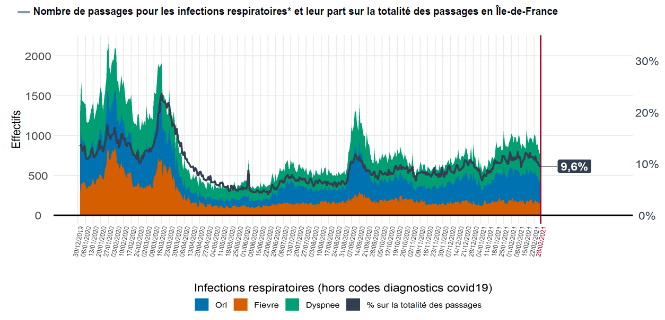


1. Sauriez-vous traduire la requête SQL de la question précédente avec les éléments des packages dplyr et dbplyr ,comme si vous lanciez la requête depuis rstudio (vous pouvez proposer la version python si vous le souhaitez) ? On assumera que le nom du schéma de la base de données est A.

### Exercice 2 : Visualisation des données COVID

L’objectif de cet exercice est d’évaluer votre capacité à générer des graphiques clairs et compréhensibles. Cette compétence est essentielle, pour faciliter au mieux la lecture des données par le personnel médical. La manière dont vous coderez sera évaluée. Ainsi, il ne s’agit pas juste de reproduire le graphique, codez de manière intelligente, claire et précise. Il est préférable d’effectuer cet exercice en R mais vous pouvez également le faire en Python, à vous de choisir. Enfin les données nécessaires se trouvent dans le fichier *extract\_data\_exercice.xlsx* .

1 – En fonction des données fournies, veuillez reproduire le graphique suivant :



*Note1: La courbe noir représente les données de la variable type\_maladie ayant pour valeur total infections. Elle est calculée selon la formule : pct\*max(total\_passages)\*ajustement.  
On prendra un ajustement égal à 6*.

*Note2 : les couleurs du graphique ne sont pas fixées, vous pouvez choisir celles que vous voulez, sauf pour la barre rouge en fin de graphique.*

2 – Créez un graphique permettant de visualiser le nombre de patients hospitalisés pour un jour donné selon les différentes catégories d’âge. On prendra la date du 2 avril 2020.

### Exercice 3: Création d’une interface Shiny ou Dash

1. Incorporez les graphiques précédemment créés dans une interface Shiny (ou Dash).
2. Ajoutez un bouton permettant de sélectionner une date et qui actualise les données du graphique 2 de l’exercice précédent.
3. Ajoutez un titre pour l’interface qui vous semble adéquat par rapport aux graphiques et données affichées
4. Selon vous, quel autre graphique/table de données pourrions-nous afficher dans cette interface, dans l’optique de visualiser l’évolution des passages pour covid19 ? (Décrivez juste le graphique, pas de besoin de le coder).

### Exercice 4 : (bonus) :

Selon vous, quelles sont les grandes étapes de la création d’un tableau de bords de type Shiny ou Dash (des données brutes à la visualisation finale) ? Quels sont les avantages et inconvénients d’une connexion en temps réel à la base de données vs une connexion asynchrone ?

BON COURAGE ! QUE LA FORCE SOIT AVEC VOUS !

### Annexe 1 : Schéma relationnel