

TP 1 – Exploration de données avec Tableau

Visualisation Interactive de Données

Romain Vuillemot – romain.vuillemot@ec-lyon.fr

Objectifs du TP 1 Tableau

L'objectif est d'explorer visuellement un jeu de données en utilisant Tableau Software. Pour rappel, Tableau permet de créer des visualisations interactives à partir de données. Il est possible de créer des visualisations à partir de données stockées dans des fichiers CSV, Excel, SQL, etc. Tableau permet de créer des visualisations interactives, c'est-à-dire que l'utilisateur peut interagir avec la visualisation pour filtrer les données, changer les représentations, etc. Tableau permet également de partager les visualisations en ligne, de les intégrer dans des sites web, etc.

Question 1 – Charger dans Tableau le jeu de données suivant :

<https://partage.liris.cnrs.fr/index.php/s/o5JgKWy7jqXDR9a>.

Question 1.a À quoi correspondent les icônes dans l'en-tête des colonnes ?

Question 1.b Lister les principaux types proposés.

Question 1.c Quelles colonnes avez vous dû corriger ?

Question 1.d Quelles données sont manquantes ? Le sont-elles tout le temps ?

Question 2 – Visualiser les résultats sous forme de bar chart. Basculer maintenant sur la 1e Feuille (Sheet 1). Commencer par afficher un graphique avec les jours DAY(Jour) en colonne, et les hospitalisation en (Hosp) en ligne. Vous devriez obtenir une courbe des hospitalisations. Transformer la courbe en diagramme en batons (changer la marque graphique).

Question 2.a Dans les données brutes à quoi correspond une ligne d'hospitalisations ?

Question 2.b Comment Tableau a-t-il agrégé les données ?

Basculer de Mesure à Dimension, en cliquant sur Hosp. Passer avec la souris sur les colonnes vous verrez qu'elles sont maintenant composées de nombreux blocs. Utiliser la dimension Région et l'associer à la couleur.

Question 3 – Explorer le graphique.

Question 3.a Dans quel ordre les régions sont elles empilées ?

Ré-organiser les régions (par ordre du plus d'hospitalisations ligne selon la mesure SUM). Sauvegarder et partager. Renommer votre feuille de calcul en double cliquant sur le titre de la visualisation, selon le numéro de l'exercice. Sauvegarder, créer un compte Tableau Public <https://public.tableau.com/s/> si ce n'est déjà fait.

Question 4 – Filtrer de manière basique. Pour des raisons géographiques, sociales, démographiques, etc. chacun des DOM devrait être analysé séparément. Nous allons les filtrer pour nous concentrer sur la France métropolitaine.

Question 4.a Quelles régions/départements avez vous filtré ?

Sauvegardez puis créez une nouvelle feuille dans votre classeur (vous pouvez dupliquer les feuilles pour ne pas repartir de 0).

Question 5 – Identifier les tendances. Afficher de nouveau un graphique avec la courbe des hospitalisations, et celles de la positivité. Pour ces deux valeurs, les lisser sur 7 jours pour réduire la variabilité due à la remontée irrégulière des données. Pour cela ajouter une "quick table calculation", elle est de deux jours en arrière par défaut, modifier cela pour passer à 7 jours en arrière.

Question 5.a Listez les périodes de croissance et de décroissance du virus

Question 5.b quel décalage temporel pouvez vous identifier visuellement entre les différents indicateurs ?

Sauvegardez cette feuille, et passez sur une nouvelle feuille pour la suite.

Question 6 – Cartographier. Créer une nouvelle feuille, ajouter Latitude et Longitude en ligne et colonne. Rajouter la dimension Région comme Repère (Mark) détail, et la Mesure Tx Occupation (moyenne) en couleur.

Clarifier les données au besoin pour expliciter les régions et département français. (1) Faites en sorte que le type de Région soit : Geographic > NUTS Europe. (2) Dans Map > Edit Locations corrigez les données non reconnues. Vous trouverez ici les codes NUTS manquants :

https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_France#NUTS_codes Via Montre-moi, constater la différence entre Carte de symbole (Symbol Map) et Carte pleine (Filled Maps). Rajouter le temps en Pages. Configurez le pour naviguer d'une semaine à l'autre. Corrigez pour que la couleur max soit à 100%.

Question 6.a Entre la Carte de symbole et la Carte pleine laquelle vous semble plus appropriée ? (justifiez en deux phrases)

Question 6.b Quelles sont les régions (donnez en trois ou quatre) qui se sont déteriorées le plus tôt lors de la 2e vague ?

Question 7– Explorer d'autres représentations (BONUS)

Sur une nouvelle feuille, explorez d'autres représentations qui vous semblent intéressantes. Essayez de reproduire une visualisation que vous avez pu voir dans la presse ou en ligne (heatmaps, donnés croisées).

Voici quelques liens vers des représentations intéressantes :

- <https://www.gouvernement.fr/info-coronavirus/carte-et-donnees>
- <https://observablehq.com/@aotten/covidoscope-v5-0>
- <https://observablehq.com/@cfleury/coronavirus-en-bretagne>
- Tableau de bord <https://covidtracker.fr/covidtracker-france/>