Les Naissances

Projet de visualisation de données
R Shiny
Le contexte, les données, ce à quoi sert notre appli
Université de Rennes II : Master Mathématiques Appliquées, Statistiques

Margaux Bailleul
Oriane Duclos

18 mars, 2023

Contents

Lien de l'application : https://oxsb16-oriane-duclos.shinyapps.io/BAILLEUL_DUCLOS/
Lien du github : https://github.com/orianeduclos/Projet_Rshiny.git

1 Introduction

1.1 Pourquoi avons-nous choisi de travailler sur les naissances ?

Les données sur les naissances permettent de comprendre les comportements de reproduction des populations et de planifier les services de santé en conséquence. Elles sont importantes pour les études démographiques, telles que la projection de la croissance de la population et la compréhension de la répartition

géographique des naissances. De plus, ces données sont utilisées dans la recherche en santé publique pour mieux comprendre les facteurs qui influencent la santé maternelle et infantile, et pour développer de nouvelles interventions pour améliorer la santé des mères et des enfants.

Cela en fait donc un sujet très intéressant et qui peut présenter énormément de possibilités de traitements visuels : cartes, régression linéaire, graphiques...

1.2 À quoi sert notre application ?

Notre application web se veut tout d'abord intercative. Celle ci nous permet de visualiser et d'analyser des données à l'aide de R. Shiny permet de créer des tableaux de bord interactifs, des graphiques interactifs et des visualisations de données dynamiques, qui permettent aux utilisateurs de filtrer, trier, explorer et analyser les données en temps réel. Elle se veut également simple d'utilisation. Nous voulions traiter les naissances à différentes échelles. En effet, à travers nos différentes bases de données, les traitement varient énormément.

1.3 Présentation des différentes bases de données et de leurs usages

1.3.1 Base de données bebe

Il s'agit d'une base de donnée que nous avons utilisé dans le cadre du cours de "Logiciel avancé" avec Nicolas Jegou. Nous avons des variables pertinantes en ce qui concerne le poids, la taille de la mère et du bébé par exemple.

Tri de bebe : Valeur manquante Nous avons decider de faire un tri sur les individus non totalement renseignés et d'utiliser une base de données sans NA.

Nous avons decider d'utiliser 2 bases de données pour présenter la partie monde car nous voulions à la fois travailler avec notre propre base de donnée "taux_fécondité" et essayer l'utiliser un serveur externe à l'aide d'un package "WDI". Cette démarche nous a été guider après avoir parcouru la documentation sur les cartes. Effectivement, nous utilisons seulement le taux de fertilité pour réaliser la carte.

1.3.2 Base de données taux_fecondité

La base de données taux de fécondité nous indique la localisation, l'indicateur, la fréquence, le temps et la valeur.

1.3.3 Base de données WDI

La bibliothèque WDI (World Development Indicators) est un package R qui permet de télécharger et d'explorer les indicateurs de développement économique et social du monde entier. Cette bibliothèque est basée sur la base de données de la Banque mondiale, qui comprend une grande quantité de données sur les pays du monde entier. Nous utilisons cette library pour pouvoir montrer à l'utilisateur le taux de fertilité dans le monde en fonction des années. De plus, nous avons que 2019, 2018 et 2017 en choix pour la variable année. Il s'agit d'une library qui est lourde et met du temps à charger or, si nous laissons un plus garnd choix d'année, cela ralentit considerablement le chargement de l'application.

1.3.4 Base de données dpt2021

Cette base de données comporte les prénoms donnés à des bébés de 1900 à 2021 en France. Nous avons tout de suite vu la possibilité de faire de la visualisation avec cette base, étant donné la grande période sur laquelle elle s'étend, et étant en lien direct avec les naissances.

2 Etude des naissances

2.1 À l'échelle du monde

2.1.1 Carte

La carte nous permet de mettre en évidence les différents pays dont le taux de fertilité est le plus élevé. Sans surprise, c'est l'Afrique qui se retrouve avec la plupart des pays avec un haut taux de fertilité.

2.1.2 Graphique sur tous les pays

Le graphique présentant tous les pays est pertinent car il nous permet de comparer directement les pays entre eux et d'avoir une idée de la tendance globale de l'évolution du taux de fécondité au fur et à mesure des années.

2.1.3 Graphique sur un seul pays

Le graphique présentant tous les pays n'est cependant pas suffisant. En effet, nous avons du mal à avoir une idée précise de l'évolution d'un seul pays. C'est pour cela que cest onglet a été créé.

2.2 À l'échelle de la France

2.2.1 Worldcoud

Le nuage de mots ou « wordcloud » en anglais est un outil de visualisation qui permet au travers d'une image de percevoir très rapidement quels sont les mots qui sont les plus fréquents au sein d'un texte ou un corpus de texte. L'utilisateur peut en un clic sélectionner une année et observer quels sont les prénoms qui ont été le plus fréquemment donnés sur cette année, en lui laissant le choix de la fréquence d'apparition du prénom ainsi que le nombre de prénoms qui seront présents dans le wordcloud.

Wordcloud est un widget html. Cela signifie que votre wordcloud sera sorti dans un HTMLformat. Nous avons décider de mettre en place un bouton qui permet à l'utilisateur de l'exporter en tant que png image.

2.2.2 Courbe du prénom au fur et à mesure des années

Cet onglet est sûrement le plus interactif avec l'utilisateur. Nous sommes en effet amenés à écrire un prénom, et en fonction de celui-ci, la courbe du nombre de bébés ayant reçu ce prénom s'affichera.

2.3 À l'échelle d'une maternité

2.3.1 Visualisation

Cet onglet nous permet de visualiser rapidement les statistiques descriptives intéractives de base avec l'utilisation du package RamCharts. Nous avons integrer différents types de répresentation (boxplot, jauge..). Cela nous as permis d'utiliser un grand panel des fonctionnalités de ce package.

2.3.2 Regression simple

Nous allons etudier le poids de naissance des bébés (en grammes).
Les variables explicatives sont toutes les variables quantitatifs de notre base de données.

Nous considérons le modèle suivant en fonction de ce que l'utilisateur choisira:

$Y_{PoidsBB} = \beta 0 + \beta 1 X1_{ChoixUtilisateur} + \epsilon$

Nous représentons en clic bouton le choix des variables. Puis, en fonction de la variable choisit, l'utilisateur peut visualiser en histogramme sa variable et le graphique de régression.

2.3.3 Régression multiple

L'utilisateur a le choix de faire une régression multiple en choisissant les différentes variables explicatives. Il pourra, dès que le choix est fait, visualiser :

- Matrice de corrélation : la matrice de corrélation indique les valeurs de corrélation, qui mesurent le degré de relation linéaire entre chaque paire de variables. Les valeurs de corrélation peuvent être comprises entre -1 et +1. Si les deux variables ont tendance à augmenter et à diminuer en même temps, la valeur de corrélation est positive.
- Le r² : le coefficient de détermination est un indicateur utilisé en statistiques pour juger de la qualité d'une régression linéaire. Ici, l'utilisateur pour voir avec un code couleur la qualité du modele. Plus la régression linéaire est en adéquation avec les données collectées.
  -Vert : R^2 supperieur ou égale à 0.6. -Orange : R^2 est compris entre 0.4 et 0.6
  -Rouge : R^2 est inferieur à 0.4
- Statsique de Fisher

3 Travail en binôme

3.1 Utilisation de git

Afin de faciliter notre travail nous avons décidé d'utiliser git. En effet, nous avons eu un cours en début de semestre et cela nous a semblé pertinent de travailler avec cet outil. Nous avons pour ce projet utilisé le terminal et non les boutons de R. Nous avons donc travaillé en effectuant des git commit, des git push, des git pull et même un git remote.

3.2 Répartition des tâches

Nous nous sommes réparties les tâches au fur et à mesure de l'avancée de l'application. Nous avons tout d'abord décidé de la structure de l'application ensemble, ainsi que des onglets et des sous-onglets. Nous avons ensuite établi les grandes idées : les graphiques, où nous voulions faire des cartes, quels packages nous pouvions utiliser qui pouvaient rendre l'application intéressante etc. Après cela, nous nous tenions au courant au fur et à mesure pour savoir qui faisait quoi. Bien sûr, si l'une d'entre nous avait commencé une partie mais n'arrivait pas à la finir ou ne trouvait pas ses erreurs, l'autre l'aidait ou prenait le relai.

3.3 Notre ressenti

Nous avons beaucoup aimé travailler sur le thème des naissances. Avoir différentes échelles à étudier nous a permis de ne pas nous répéter, d'autant plus que nous avons utiliser plusieurs outils de visualisation différents. Nous avons également beaucoup aimé travailler ensemble : nous sommes un binôme qui fonctionne bien car nous avons la même méthode de travail.

3.4 Difficultés rencontrées

Nous avons chacune rencontré des difficultés en travaillant sur le projet. Nous allons ici vous les présenter :

1. Utilisation du package formattable, association avec le package DT qui marchait très bien dans ma console mais impossible de le faire marcher dans l'application à cause d'une erreur html
2. Base de données fertility, impossible de sélectionner trop d'années car trop lourd
3. NA dans le choix des années avec la carte, on ne sait pas comment l'enlever
4. Voulait faire un bouton logout mais il fallait shiny server pro
5. Difficultés à publier notre application :

Methode de resolution des problèmes Pour trouver les erreurs nous avons essayer de publier une application plus basique et nous avons rajouter le code au fur et à mesure. Nous avons mis du temps à réaliser cette étape car cela met du temps de verifier étape par étape que notre application fonctionne en local et de publier. Nous avons utiliser le code

```
rsconnect::showLogs() #fonction pour afficher les messages du journal d'une application déployée
```

Problème non resolu * Problème racontrer au niveau de la visualisation des bases de données au format DT sur nos grosses bases de données (prenom et taux de fecondité). L'une d'entre elle fait 76Mo. * Problème rencontré lorsque que l'on fait appel à un serveur externe type API qui nous permet de réalisé une carte du taux de fertilité avec le package WDI

Les hypothèses du problème * Version shiny.io n'a pas les memes versions de packages que celles qu'on utilise en local * Shiny.io protege son serveur et ne peut pas publier des applications qui font appel à des serveurs externes

Ces difficultés nous ont permis de voir les limites de notre application et de nous remettre sans cesse en question.

4 Conlusion

Pour conclure, cela a été un vrai plaisir de travailler sur ce projet. Nous espérons que l'application vous plaira. Nous avons beaucoup appris : gérer nos erreurs, gérer notre temps (le projet demandant une certaine investigation et ayant une date limite de rendu), répartir les tâches…