

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

19 mars, 2023

Contents

1	Introduction	2
1.1	Pourquoi avons-nous choisi de travailler sur les naissances ?	2
1.2	À quoi sert notre application ?	2
1.3	Présentation des différentes bases de données et de leurs usages	2
2	Etude des naissances	3
2.1	À l'échelle du monde	3
2.2	À l'échelle de la France	3
2.3	À l'échelle d'une maternité	4
3	Travail en binôme	4
3.1	Utilisation de git	4
3.2	Répartition des tâches	5
3.3	Notre ressenti	5
3.4	Difficultés rencontrées	5
4	Conclusion	6

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 veut rendre la compréhension des naissances dans le monde de manière accessible et à travers différentes échelles. C'est pour cela que nous avons rendu l'application interactive au maximum, en nous permettant de visualiser et d'analyser des données. Cela va être rendu possible grâce à l'utilisation de RShiny, qui permet de créer des tableaux de bord interactifs, des graphiques interactifs et des visualisations de données dynamiques, permettant aux utilisateurs de filtrer, trier, explorer et analyser les données en temps réel. Elle se veut également simple d'utilisation avec différents onglets traitant chacun d'une échelle différente. L'application veut répondre à la problématique suivante : comment les naissances évoluent-elles à différentes échelles ?

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 données que nous avons utilisé dans le cadre du cours de "Logiciel avancé" enseigné par Nicolas Jegou durant notre licence. La base présente plusieurs variables quantitatives, nous avons donc relevé des variables pertinentes en ce qui concerne le poids, la taille de la mère ou encore celle du bébé par exemple.

Tri de la base de données : Valeurs manquantes

Nous avons décidé de faire un tri sur les individus non totalement renseignés et d'utiliser une base de données sans NA.

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 du taux de fécondité dans les différents pays du monde en fonction des années (de 1960 à 2020).

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. Nous utilisons cette librairie pour pouvoir montrer à l'utilisateur le taux de fertilité dans le monde en fonction des années.

De plus, en important la librairie, nous nous sommes limitées aux années 2017, 2018 et 2019. Il s'agit d'une librairie qui est lourde et qui met du temps à charger. Si nous avions choisi un plus grand nombre d'années, cela aurait considérablement ralenti 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. Les seules années sélectionnables sont de 2017 à 2019, car comme expliqué précédemment, augmenter la plage de la période aurait considérablement ralenti notre application.

2.1.2 Taux de fécondité des pays du monde

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. Le graphique est interactif et nous permet de savoir la valeur du taux de fécondité pour chaque pays et pour chaque année.

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. Cet onglet permet donc de rendre compte de l'évolution d'un seul pays, en observant des hausses ou des baisses que nous ne pouvions pas visualiser dans le graphique précédent.

2.2 À l'échelle de la France

2.2.1 Wordcloud

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.

Nous avons décidé 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 celui qui amène le plus l'utilisateur à interagir avec l'application. 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 en fonction des années s'affichera. Il ne faut pas hésiter à essayer les prénoms de personnes que nous connaissons.

Nous avons décidé de mettre en place un bouton qui permet à l'utilisateur de l'exporter en tant que png image.

2.3 À l'échelle d'une maternité

2.3.1 Visualisation

Cet onglet nous permet de visualiser rapidement les statistiques descriptives de la base de données bébé avec l'utilisation du package RamCharts. Nous avons intégré différents types de représentation (boxplot, jauge..). Cela nous a permis d'utiliser un grand panel des fonctionnalités de ce package.

Nous avons également mis en place un graphique représentant le poids du bébé en fonction de la tranche d'âge de la maman.

2.3.2 Régression simple

Nous allons étudier le poids de naissance des bébés (en grammes).

Les variables explicatives sont toutes les variables quantitatives de notre base de données.

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

$$Y_{PoidsBB} = \beta_0 + \beta_1 X_{ChoixUtilisateur} + \epsilon$$

Nous représentons en clic bouton le choix des variables. L'utilisateur peut ensuite visualiser la variable sélectionnée sous forme d'histogramme, puis la visualiser avec la courbe de régression en fonction de la variable PoidsBB.

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 :

- La 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 r^2 : le coefficient de détermination est un indicateur utilisé en statistiques pour juger de la qualité d'une régression linéaire. Ici, l'utilisateur peut voir avec un code couleur la qualité du modèle. Plus la régression linéaire est en adéquation avec les données collectées.
 - Vert : R^2 supérieur ou égale à 0.6.
 - Orange : R^2 est compris entre 0.4 et 0.6
 - Rouge : R^2 est inférieur à 0.4
- La statistique de Fisher du modèle
- Le summary du modèle de régression en sortie R

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.

3.2 Répartition des tâches

Nous nous sommes réparties les tâches au fur et à mesure de l'avancé de l'application. Nous avons tout d'abord décidé de la structure de l'application ensemble, ainsi que les onglets et les sous-onglets qui seront présents. 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 avons travaillé en nous tenant au courant au fur et à mesure pour continuer la répartition des tâches mais sur des tâches beaucoup plus précises. Bien sûr, en cas de difficulté, l'entraide était présente.

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 utilisé 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. Nous voulions faire un bouton logout mais était pour shiny server pro
5. Difficultés à publier notre application :

Méthode de résolution des problèmes

Pour trouver les erreurs nous avons essayé 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 il faut vérifier étape par étape que notre application fonctionne en local et la publier ensuite. Nous avons utilisé la fonction ci-dessous pour déterminer nos erreurs de façon précise.

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

Problèmes non résolus

- Problème rencontré au niveau de la visualisation de notre base de données taux de fécondité.
- Problème rencontré lorsque que l'on fait appel à un serveur externe type API qui nous permet de réaliser 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

Résolution de la publication

Ces difficultés nous ont permis de voir les limites de notre application et de nous remettre sans cesse en question. Concernant la publication de l'application, nous avons donc trouvé les lignes de code qui nous empêchaient de la publier. Nous avons alors mis en # ces lignes, publié l'application, et enlevé les # afin que l'application "entière" puisse être visualisée si nous la lançons en local. Nous n'avons pas publié notre carte qui fait appel au serveur externe WDI la table de notre base de donnée taux de fécondité en affichage DT ainsi que son summary. L'application publiée n'est donc pas complète, mais elle est tout de même publiée.

4 Conclusion

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... Pour répondre à notre problématique, nous vous laissons le loisir d'aller parcourir notre application.