

A04Ga_biometry

Public template

generated from [BioDataScience-Course/A05Ga_biometry](#)

1 Branch

1 Tag

Go to file

t

About file

Code

...

GuyliannEngels

minor modification of t...

8bfd901 · 2 years ago

3 Commits

R	Version 2024...	2 years ago
.Rbuild...	Initial commit	2 years ago
.gitign...	Version 2024...	2 years ago
A04Ga...	Version 2024...	2 years ago
DESCR...	Version 2024...	2 years ago
LICENSE	Initial commit	2 years ago
READ...	minor modific...	2 years ago
bio_ex...	Version 2024...	2 years ago

Projet de groupe de SDD I: collecte de données relatives à l'obésité

- Readme
- View license
- Activity
- Custom properties
- 0 stars
- 2 watching
- 0 forks
- Audit log
- Report repository

Releases 1

Version for academic year 20...

Latest

now

README

License

No packages published

[Publish your first package](#)

Collecte de données liées à l'obésité humaine

Ce projet nécessite d'avoir assimilé l'ensemble des notions des cinq premiers modules du cours de science des données biologiques 1. Il correspond au dépôt GitHub https://github.com/BioDataScience-Course/A04Ga_biometry. Il est distribué sous licence [CC BY-NC-SA 4.0](#).

Contributors 2

phgrosjean

Philippe Grosjean

GuyliannEngels

Guyliann Engels

Objectifs

Ce projet est à réaliser par groupe de quatre étudiants, avec la partie définition du protocole et collecte des données commune à toute la classe. Il doit vous permettre de démontrer les compétences suivantes :

Languages

R 100.0%

- définir un protocole d'expérience
- collecter des données et les encoder correctement
- se sensibiliser à la précision des mesures et à l'intercalibration
- rédigier un dictionnaire des données

- respecter la vie privée (RGPD) via la pseudonymisation des données
- remanier un tableau de données
- (en option) réaliser et interpréter des graphiques et des tableaux

Consignes

Partagez-vous le travail et effectuez bien les commits relatifs à la partie que vous avez réalisée : vos enseignants regarderont l'historique des commits pour vérifier la répartition du travail et noteront en fonction de cela. Un étudiant ayant moins contribué n'aura donc pas la même note qu'un autre qui aura fait plus de travail. Décidez d'une répartition du travail au sein de l'équipe dès le début. Par contre à la fin, vérifiez les résultats de tout le monde. L'ensemble de l'équipe est responsable de la totalité du travail. Vous devez vous mettre d'accord entre vous à la fin sur le document dans son intégralité.

Il n'y a pas de tests ici car le projet est libre et aucune réponse stéréotypée n'est donc attendue. Assurez-vous tout de même que le document Quarto compile sans erreurs à la fin, que tous les conflits éventuels dans les commits sont réglés et que la dernière version est bien dans GitHub (vérifiez directement dans GitHub après vos derniers commit-pull-push).

Ce projet est de niveau 3 et sera comptabilisé comme les projets individuels étant donné que l'essentiel du travail est réalisé de manière commune à toute la classe, il n'est pas possible de coter équipe par équipe comme pour les projets de groupes plus classiques.

Il se décompose en plusieurs phases qui s'étalent sur le module 4 et le module 5 du cours de Science des Données I.

Phase 1 (module 4)

Dans un premier temps, vous réfléchissez aux variables qu'il vous faudra employer pour déterminer l'IMC des personnes incluses dans l'étude et pour le mettre en relation avec ses caractéristiques d'une manière générale. À la fin de cette phase, vous aurez défini les mesures à réaliser et chaque étudiant collectera les données relatives à quatre individus.

Phase 2 (modules 4 & 5)

Vous allez définir un protocole permettant de pseudonymiser les données. Ensuite, vous allez encoder vos données respectives dans un fichier commun.

Cliquez sur le lien suivant pour accéder à Google Sheet et complétez le dictionnaire des données. Ce document est collectif pour l'ensemble des étudiants du cours de Science des données I.

https://docs.google.com/spreadsheets/d/1FyvoJtKkdvf5uwSASZhl_EkhzYR5033Yi2vdEnYMgFQ/edit?usp=sharing

Cliquez sur le lien suivant pour accéder à un second Google Sheet et compléter le tableau de données associé à votre expérience. L'encodage des observations n'est pas une tâche à prendre à la légère. Soyez particulièrement attentif à ne pas commettre d'erreur et à respecter le dictionnaire des données défini ci-dessus.

https://docs.google.com/spreadsheets/d/1UbaVJZzkCBacusqWYkIZqYZoi3OSCXc89JXka_vVDhc/edit?usp=sharing

Phase 3 (module 5)

Le dictionnaire et le tableau de données peuvent être directement importés dans R avec `read$csv()` et les liens suivants :

- Dictionnaire des données : https://docs.google.com/spreadsheets/d/e/2PACX-1vQ14kFDtlqxUqJpfKlcZRHA2i3ZnCwSdT_bqcx7BWp3hk_fqEGk9JmBvRsHvdBZrl3KCACV-LHQ-tAv/pub?gid=0&single=true&output=csv
- Tableau de données : https://docs.google.com/spreadsheets/d/e/2PACX-1vTTvgkHmrXPXHJ16cypzK7ooBhcSybscjBMc0obVSt1dvxINL9rtage91IKD8Jec-3n4eX_6O-VdW7f/pub?gid=0&single=true&output=csv

Dans le script `R/import_tidy.R`, importez, remaniez, ajoutez les labels et les unités et enfin, sauvegardez localement vos données nettoyées dans le dossier `data/` et nommez le fichier `biometry.rds` (`biometry_metadata.rds` pour le dictionnaire des données).