

Consigne examen R

Paolo Crosetto

10/21/2020

Quoi?

1. Rapport statistique en ligne

Vous devez faire un **rapport statistique** sur un jeu de données. Le rapport va être fait en Rmarkdown (`file -> New file -> Rmarkdown...`). Je m'attends à ce que vous produisez deux fichiers: le code source (.Rmd) et le rendu (.html). Le deuxième est créé à partir du premier en appuyant sur **Knit**.

Un rapport de statistique est un rapport qui **présente** les données. Il est en large partie **déscriptif** mais il peut, si vous voulez, avoir aussi un volet d'exploration des liens de causalité, s'il y en a.

Il s'agit de montrer le jeu de données au lecteur, en illustrant ses caractéristiques; il s'agit aussi de répondre à des questions sur le jeu de données. Ces questions peuvent être des questions simples, même banales (combien d'avion volent-ils de NY?), ou des questions complexes (quelle est la relation entre l'espérance de vie et le PIB d'un pays au fil du temps?). C'est à vous de poser des questions, et d'y répondre grâce aux données.

Vous avez deux semaines pour préparer le rapport; vous aurez 7 minutes au cours de la dernière séance (le **18 novembre**) pour le présenter à la classe. Pendant la présentation, je vais vous poser des questions de clarification ou je vous demanderai de justifier vos choix.

2. Questions additionnelles

Le 19 novembre, le lendemain de votre présentation, vous allez recevoir sur votre espace git **deux** questions personnalisées. Ces questions viseront à vous pousser à faire un peu plus avec vos données, et seront différentes pour chacun d'entre vous. Vous aurez **3 jours** (jusqu'au mardi 23 novembre, minuit) pour changer votre rapport de stat afin de répondre à ces deux questions additionnelles.

Dois-je présenter?

Oui, vous allez devoir présenter votre travail face à la classe le 26 octobre.

Vous aurez **7 minutes** par personne (tout le monde doit présenter pendant une séance de 3h).

Vous présenterez votre .Rmd transformé en .html, pas besoin de faire des slides en plus.

Sur quelles données?

- vous êtes libres d'utiliser le jeu de données que plus vous aimez parmi ceux présents sur le site: <https://github.com/rfordatascience/tidytuesday>

Comment est-ce que je me coordonne avec les autres pour les jeux de données?

J'écris le jeu de données choisi ainsi que le lien pour récupérer les données ici:

https://docs.google.com/spreadsheets/d/1_UjRx7ZOv4757CCB4dKhWQ9qSwMGJhw8evOfllPtIPo/edit?usp=sharing

Pour quand?

Vous faites cela quand vous voulez, mais il doit être chargé sur github, dans le dossier **Exam**, dans un fichier nommé après votre nom (**VotreNom.Rmd**) **avant** (et non pas *pendant*) le cours du 26 octobre à 9h. Donc la date limite est le 26/10 à 9h. Tout fichier compilé et/ou pushé après 9h ce matin ne sera pas pris en compte, et vous devrez remonter à la dernière version présente sur le repo avant 9h ce matin là. Autrement dit, la date limite **est une vraie limite**.

Où?

- vous travaillez chacun dans votre espace.
- le moment venu, vous copiez votre travail fini dans le dossier **Exam**
- et nommez le fichier **<votre_nom>.Rmd** où vous insérez votre nom
- vous vous assurez que cela compile et qu'un fichier .html est créé après le **kint**
- les questions additionnelles apparaîtront aussi sur ce même fichier, et vous ferez vos changements directement sur ce même fichier.

Comment vais-je être évalué?

Vous allez être évalué sur les critères suivants.

- vous avez su charger le .Rmd et le .html au bon endroit utilisant git
- le .Rmd compile (je vais le compiler moi-même) sans erreurs. [essayez de le compiler sur l'ordi d'un copain pour être sûr, et évitez les parcours absolus pour les données]
- la présentation des données est complète et compréhensible
- vous vous êtes posés des bonnes questions sur les données
- vous avez pu répondre à vos questions (tableaux, graphiques)
- je vais vous poser des questions, du type 'et si on voulait faire X?'; 'comment auriez-vous changé le plot pour montrer Z?', etc. Le degré de confiance dans l'outil montré lors des réponses sera évalué.
- vous aurez bien répondu avec un usage précis des données aux questions additionnelles.

Où est-ce que je trouve de l'aide?

- pour vous inspirer et trouver du code pour les graphiques et des idées d'analyse vous pouvez regarder l'application tidytuesday.rocks, qui vous montre des créations graphiques pour chaque dataset, et très souvent du code aussi, ici: <https://nsgrantham.shinyapps.io/tidytuesdayrocks/>
- pour vous inspirer davantage, il y a du code disponible pour la plupart des plots que vous pouvez concevoir ici: <https://www.r-graph-gallery.com/>
- et encore d'autres idées ici: <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>

Bon courage!