

Science des données I : module 7



Approche pédagogique du cours de science des données

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Laboratoire d'Écologie numérique



<https://wp.sciviews.org>
sdd@sciviews.org

De quoi allons-nous discuter ?

Suite à l'analyse des avis pédagogiques de l'an dernier, nous nous rendons compte que l'approche en classe inversée n'est pas toujours bien comprise

- Explication de notre approche pédagogique
- Debriefing de vos résultats Q1
- Retour et propositions de votre part

C'est quoi la classe inversée ?

Vidéo expliquant ce qu'est la classe inversée :
<https://www.youtube.com/watch?v=uLKmLDrGyiw>

Notre approche : **pédagogie active en classes inversées** (vous apprenez *d'abord* à la maison, nous appliquons *ensuite* en présentiel).

C'est quoi la pédagogie active ?

*I **hear** and I forget.*
*I **see** and I remember.*
*I **do** and I understand.*
— Confucius

Vidéo expliquant la pédagogie active :
<https://www.youtube.com/watch?v=ygjSle9Pkg4>

Et moi, je fais quoi dant tout cela ?

Vous êtes **acteur de votre apprentissage**, les enseignants sont des **facilitateurs** (plus en retrait par rapport à l'approche classique)

Plus de séparation entre **cours théorique** et **exercices** ; le professeur et le ou les assistants se répartissent le travail de manière plus uniforme

ECTS

European Credits Transfer System, créé en 1988 de manière standardisée par la Commission Européenne comme correspondant à une **charge de travail totale** pour l'étudiant de **25 à 30 heures de travail**.



Comment voulez-vous passer vos 25-30h/ECTS ?

Cours classique *ex cathedra* + séance d'exercices

Écouter passivement
des cours ex cathedra

Faire des exercices avant
d'assimiler la matière

Mettre ses notes
en ordre

Terminer
les exercices

Préparer et passer
un examen

- Le réel apprentissage se déroule **après** les séances de cours et d'exercices, lorsque vous “potassez” la matière
- Un examen est nécessaire pour vérifier que vous avez bien assimilé la matière

Optimisation du temps de travail... comparé à la classe inversée

Cours classique *ex cathedra* + séance d'exercices

Écouter passivement des cours ex cathedra	Faire des exercices avant d'assimiler la matière	Mettre ses notes en ordre	Terminer les exercices	Préparer et passer un examen
---	--	---------------------------	------------------------	------------------------------

Approche en classe inversée

Préparer à la maison	Appliquer en présentiel	Préparer à la maison	Appliquer en présentiel, poser des questions en présentiel	Finaliser le module chez soi à son rythme
----------------------	-------------------------	----------------------	--	---

- *Aucune* séance en présentiel ne se fait sans préparation
- Chaque heure de travail est pleinement consacrée à l'apprentissage et/ou à la compréhension
- Vous êtes actifs **tout le temps** et vous gérez votre apprentissage à votre rythme
- La matière est *comprise* et *assimilée* à la fin du processus : **pas besoin d'un examen à la fin** pour le vérifier. Votre activité certifie de votre progression.

Le professeur est un coach et un facilitateur

- Le professeur (et l'assistant) ne mettent **pas** leur savoir en avant. C'est vous qui construisez votre propre savoir.
- Les enseignants **ne répondent pas directement à vos questions** : ils vous mettent sur une piste et vous font réfléchir pour trouver la réponse *par vous-même*
- Les enseignants se mettent en retrait, mais sont *disponibles* pour vous aider (Discord, forum Moodle, mail, etc.)

Tout cela est normal en classe inversée, mais déroutant : le professeur n'apparaît plus comme l'omniscient qui transmet de manière unilatérale son savoir !

Que pensez-vous de tout ceci ?

Debriefing des notes du Q1

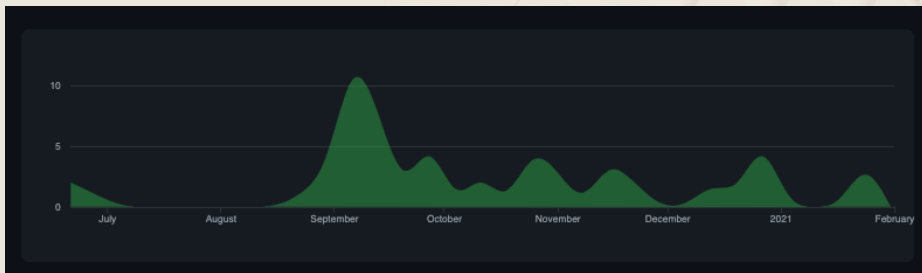
- 9 étudiants sur 10 ont fait à peu près tous les exercices
- Ceux qui ont fait tous les exercices ont une note bonne à excellent
- Travail important, matière complexe et dense... mais **vous avez très bien travaillé dans l'ensemble**

Erreurs récurrentes à améliorer

- Répartissez mieux votre travail : les applications dans les **assignments GitHub** sont les plus importantes
- La **compréhension** et l'**interprétation** sont fondamentales et doivent encore être affinées
- Entre-aidez-vous, nous l'encourageons, mais **citez vos sources**. Si un résultat provient d'un collègue, vous l'indiquez ! Vous pouvez citer comme ceci [Untel, comm. pers.]
- L'étudiant qui aide reçoit des points. Cela qui sait intégrer judicieusement les bonnes idées et qui cite ses sources reçoit aussi des points !
- Par contre, le plagiat doit être fortement sanctionné (pas fait en Q1, mais nous serons plus sévères en Q2)

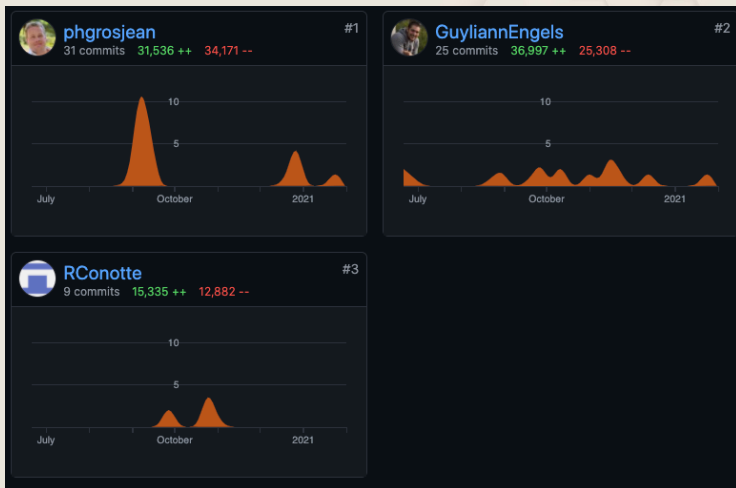
Répartitions du travail

GitHub propose des outils pour visualiser l'activité d'un dépôt (Insight -> Contributors)



L'activité globale peut être **divisée par collaborateur**.

L'activité de chaque membre est **disponible**



Vérifiez dans vos projets que votre activité est prise en compte.

```
pop2020 <- tibble(  
  age = factor(c("Moins de 18 ans", "18 à 64 ans", "65 ans et plus"),  
               levels = c("Moins de 18 ans", "18 à 64 ans", "65 ans et plus"),  
  population = c(2312040, 6976123, 2204478))  
knitr::kable(pop2020, col.names = c("Groupe d'age", "Nombre de personnes"),  
              caption = "Population au premier septembre 2020 en Belgique")
```

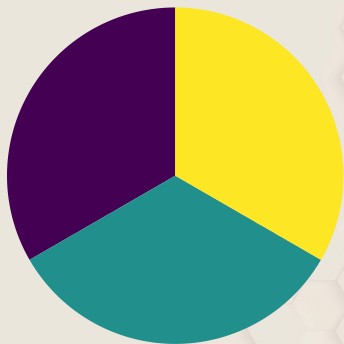
Table 1: Population au premier septembre 2020 en Belgique

Groupe d'age	Nombre de personnes
Moins de 18 ans	2312040
18 à 64 ans	6976123
65 ans et plus	2204478

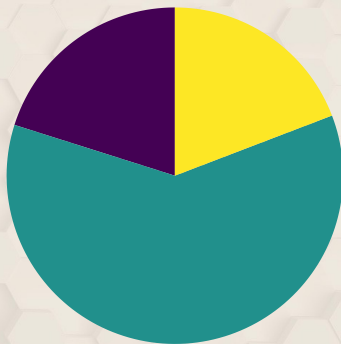
Quel est le graphique le plus pertinent ?

age Moins de 18 ans 18 à 64 ans 65 ans et plus

A

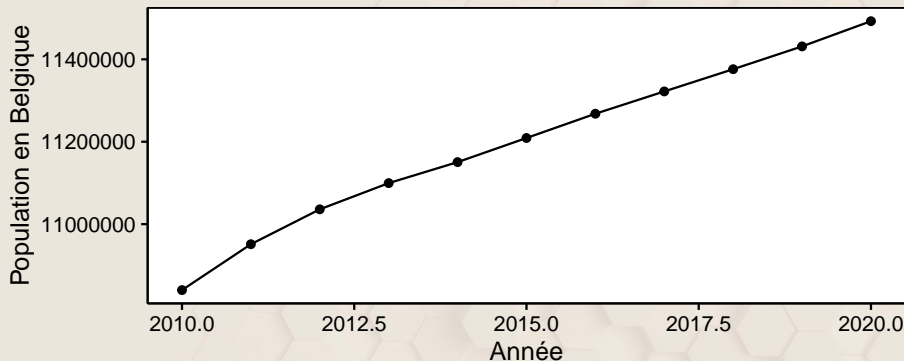


B

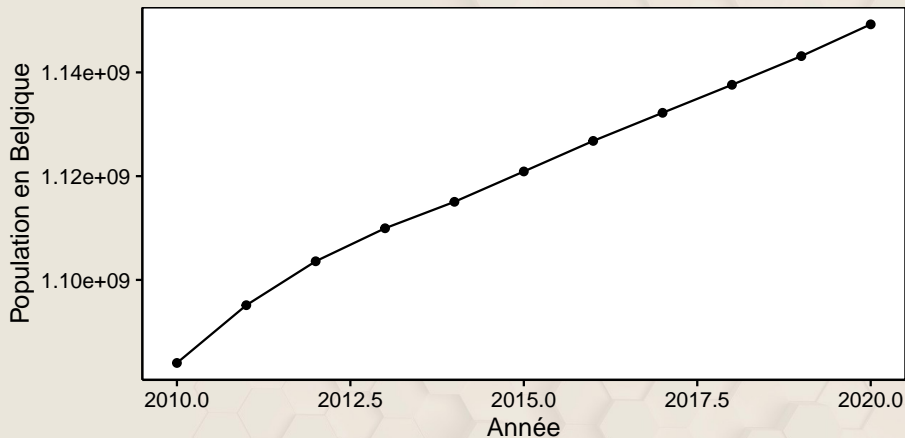


Compréhension et interprétation : les axes

```
chart(pop, density ~ year) +  
  geom_point() +  
  geom_line() +  
  labs(y = "Population en Belgique", x= "Année")
```

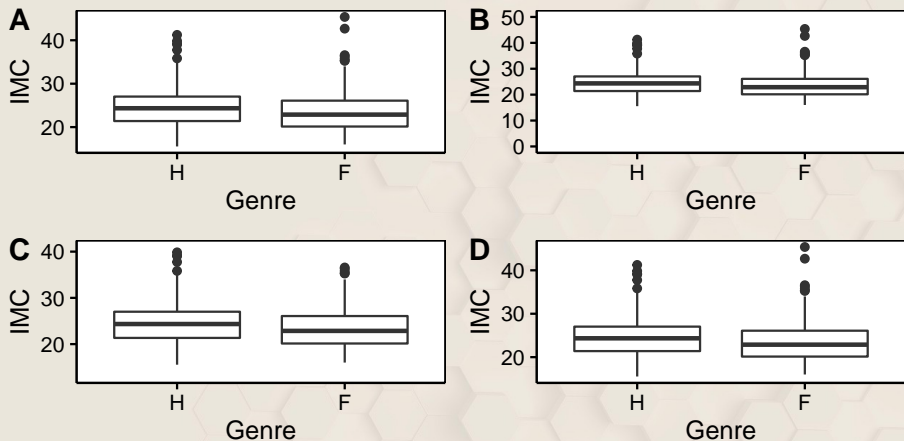


Repérez l'erreur sur le graphique ci-dessous.



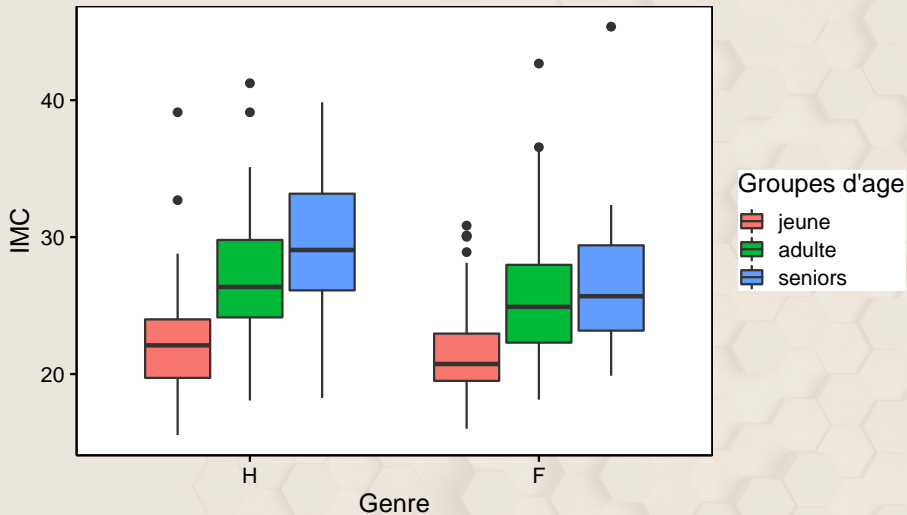
Compréhension et interprétation : les graphiques multiples

Variation de l'indice de masse corporelle en fonction du genre et du groupe d'âges.



Ce graphique manque d'informations pour être compris.

- Ce graphique propose la même information que le graphique multiple précédent.



Un graphique multiple ne doit être utilisé que si les graphiques associés sont indissociables.

Compréhension et interprétation : les valeurs manquantes.

```
drop <- tibble(  
  x = runif(10),  
  y = runif(10),  
  z = c(5, 3, rep(NA, 6), 3, 2)  
)  
knitr::kable(drop[1:3,])
```

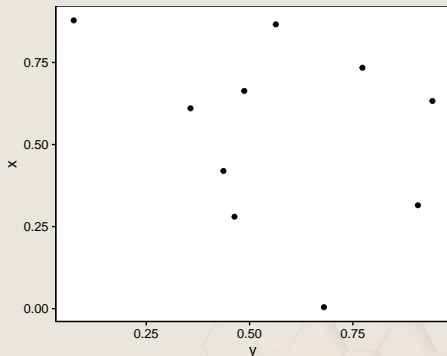
x	y	z
0.7340858	0.7731140	5
0.6634229	0.4871618	3
0.8782159	0.0743409	NA

Elimination naïve des données manquantes.

```
drop1 <- drop_na(drop)
```

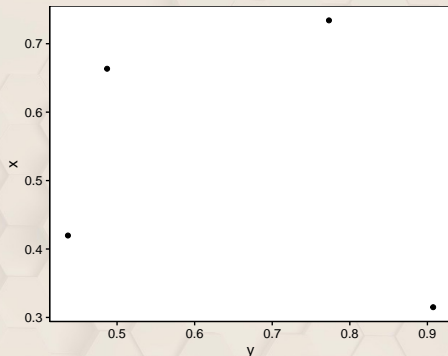
Utilisation du tableau drop

```
chart(drop, x ~ y) +  
  geom_point()
```



Utilisation du tableau drop1

```
chart(drop1, x ~ y) +  
  geom_point()
```



Soyez très vigilant avec les NA. Les variables x et Y ne contenaient pas de valeurs manquantes.

Vous avez la parole !

- Exprimez-vous, suggérez : **votre avis compte énormément...**
- Indiquez **deux points positifs et deux points négatifs** du cours et de la matière dans le formulaire