

Perception de R et RStudio par des apprenants dans des cours de science des données biologiques



Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Service d'Écologie numérique

<philippe.grosjean@umons.ac.be>, <phgrosjean@sciviews.org>

Rencontres R 2021, Paris

De quoi allons-nous parler ?

Cours de science des données biologiques à l'Université de Mons
(<https://wp.sciviews.org>), cf. présentation précédente de Guyliann Engels

- Méthode pédagogique, comment **apprendre** ?
- **Perception** des étudiants face à l'environnement logiciel (R, RStudio, R Markdown & GitHub)

Subsection 1

Apprendre

Apprendre, niveau 0.x

Cours en amphis, vidéos, ... = apprenants **passifs**



Apprentissage, niveau 1 ?

Script R de démo => apprenants **exécutent** et **réutilisent** sur d'autres données

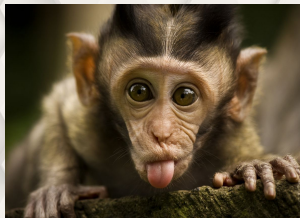
```

1 data(iris)~
2 plot(Sepal.Length~Sepal.Width,data=iris,col=1)~
3 pairs(~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width,~
4   ~data=iris,col=1)~
5 ~
6 iris.pcomp<-pcomp(~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width,~
7   ~data=iris,scale=FALSE)~
8 summary(iris.pcomp,loadings=FALSE,cutoff=0.1)~
9 ~

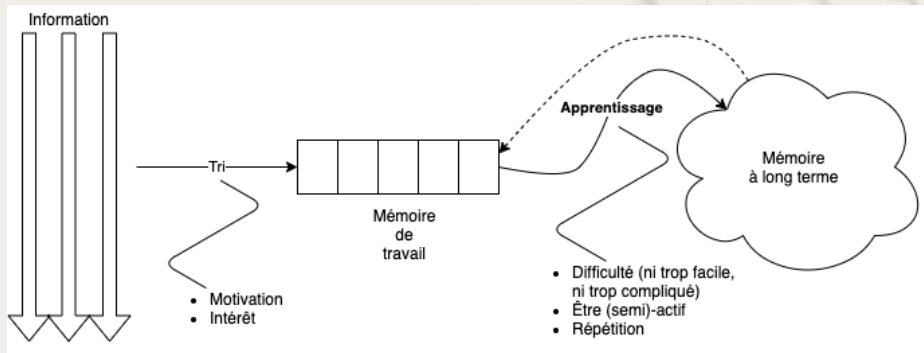
```

Monkey see, monkey do

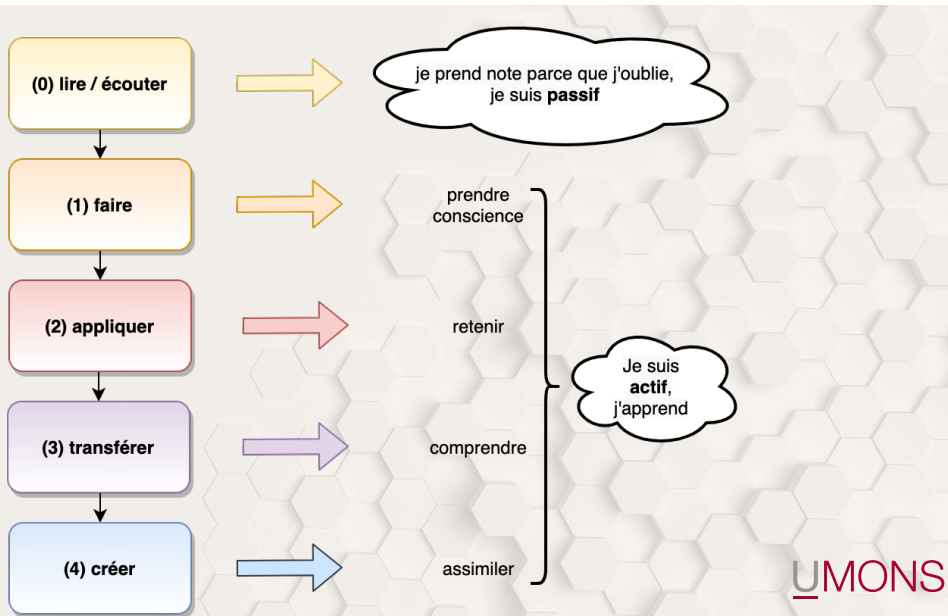
Proverbe en pidgin jamaïcain
Imitation sans réfléchir



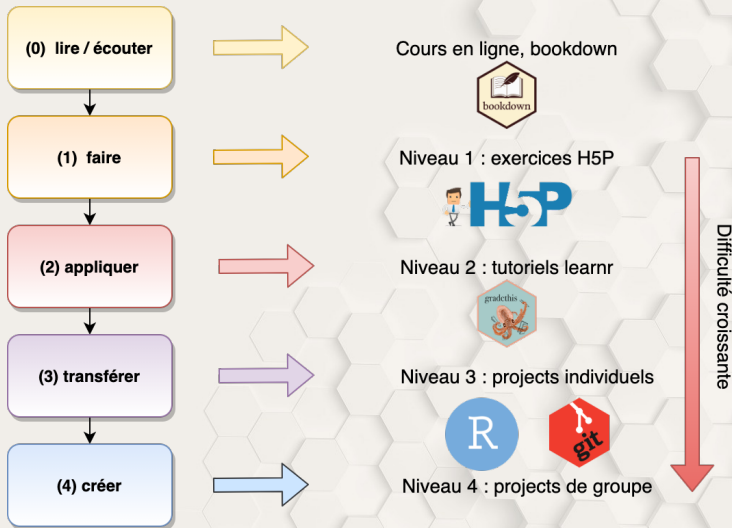
Apprendre... vraiment, un processus complexe



Apprendre, niveaux 1 à 4



Quatre niveaux d'exercices



Exemple : retirer les premiers éléments uniques (vecteur ou tableau)

Jeu de données exemple dans la variable `ex`:

```
head(ex, n = 5)
```

month	value
April	0.6194607
January	1.4951996
January	1.6077667
March	-0.2860269
October	-0.7423510

Je veux récupérer la première ligne pour chaque mois du tableau.

Résolution avec tidyverse ({dplyr}) :

```
library(tidyverse)
ex %>%
  group_by(month) %>%
  slice_head(n = 1)
```

month	value
April	0.6194607
December	-1.2419243
February	-1.3979057
January	1.4951996
March	-0.2860269
October	-0.7423510

Notez ceci :

- L'ordre des éléments dans le tableau final a changé (ordre alphabétique),
- Cela ne fonctionne pas pour un vecteur

Le même... mais en R de base (ordre sans importance)

En R de base, vous pouvez utiliser la fonction `duplicated()`.

`duplicated {base}`

R Documentation

Determine Duplicate Elements

Description

`duplicated()` determines which elements of a vector or data frame are duplicates of elements with smaller subscripts, and returns a logical vector indicating which elements (rows) are duplicates.

Examples

```
x <- c(9:20, 1:5, 3:7, 0:8)
## extract unique elements
(xu <- x[!duplicated(x)])
```

Niveau 1 : faire

Exercice **H5P** (<https://h5p.org/node/1193264>). Facile, ludique, mais un rien plus évolué que le “bête” recopiage.

Construisez une instruction R qui reprend seulement la première occurrence de chaque valeur du vecteur `x <- c("chat", "chat", "chien", "chat", "cheval", "chien")` et place le résultat dans `x2`.

<- [(x)]

!

x2

x

duplicated

✓ Vérifier



UMONS

Niveau 2 : appliquer

Code R exécutable et vérifiable. 5 à 10 exercices dans un **tutoriel learnr**. Exemple disponible à https://go.sciviews.org/learnr_duplicated

(sauver en *duplicated.Rmd*, ouvrir dans *RStudio* et cliquer sur “Run Document”).

duplicated()

Philippe Grosjean & Guyliann Engels

2021-10-07

Utilisation simple de duplicated()

Utilisation de duplicated() sur un

vecteur

Utilisation de duplicated() sur un

tableau

Utilisation créative de duplicated()

Start Over

Utilisation simple de duplicated()

Voici un vecteur `v` :

```
print(v)
```

```
## [1] "chien" "chat" "chien" "chat" "chat" "cheval" "chien"
```

Indiquez à quelle position se situe la première occurrence de chaque mot dans `v` en retournant un vecteur `v1` de même taille contenant des valeurs logiques (`TRUE` ou `FALSE`). Utilisez `duplicated()` pour y arriver.

R Code

Start Over

Hints

Run Code

Submit Answer

```
1 v1 <- ____ (____)
2 v1
3
```

Next Topic

Niveau 3 : transférer

Être capable de transférer le principe vu dans un **contexte différent**.

Dans nos cours : **projets individuels** dans GitHub / RStudio.

Exemple : le jeu de données à analyser implique de récupérer la première occurrence pour chaque type d'évènement.

TODO: retravailler et illustrer cette dia.

Niveau 4 : créer

Inventer des nouvelles utilisations du concept.

Dans nos cours, **projets de groupe** (mobiliser l'**intelligence collective**).

Exemple : dans un projet, devoir extraire le premier évènement pour chaque utilisateur et chaque mois ?

Subsection 2

Perception

Charge cognitive des learners (NASA-TLX, Hart & Staveland 1988)



Les tutoriels learners ont été conçus comme un trait d'union entre la théorie (cours en ligne) et la pratique (assignations GitHub). Par rapport à leur utilisation, notez les six critères suivants.
(1 = faible, 10 = élevée)

Propositions

Entrez les propositions que votre audience pourra noter

Quelle a été l'importance de l'activité mentale et intellectuelle requise ?



Quelle a été l'importance de l'activité physique requise ?



Quelle a été l'importance de la pression temporelle causée par la rapidité nécessitée pour l'accomplissement de la tâche ?



Quelle réussite pensez-vous avoir eu dans l'accomplissement de la tâche ?



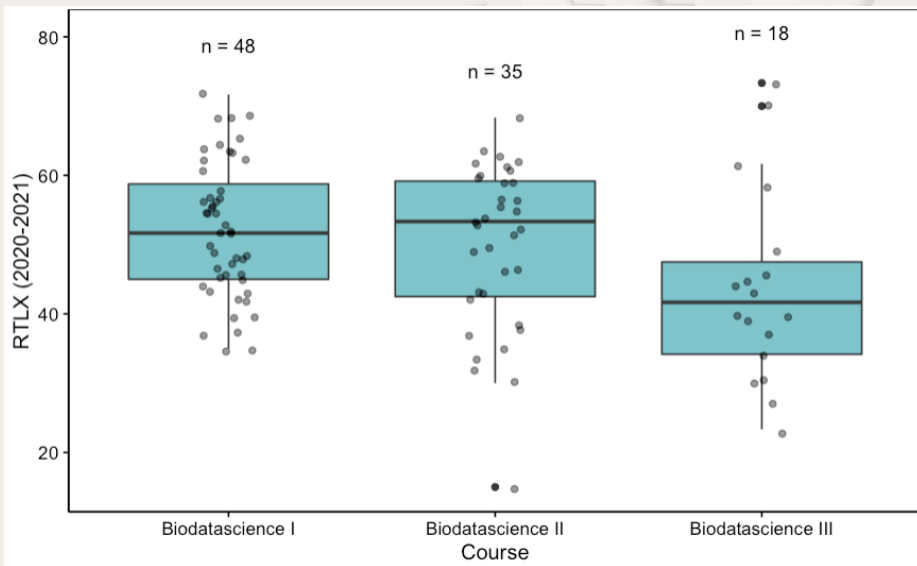
Quel degré d'effort avez-vous dû fournir pour exécuter la tâche demandée ?



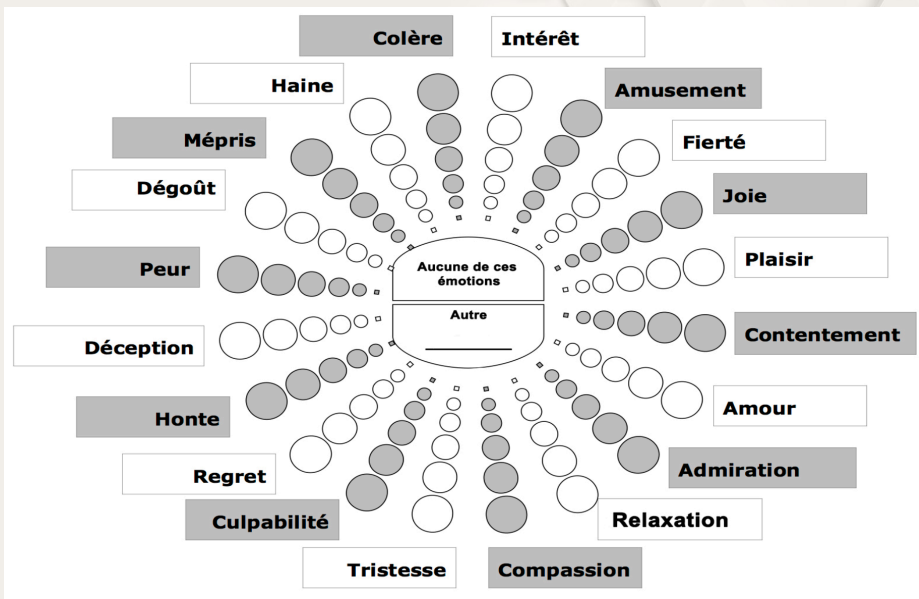
Pendant l'exécution du travail vous êtes-vous senti découragé(e), irrité(e), stressé(e), sans assurance ?



Charge cognitive perçue des learners

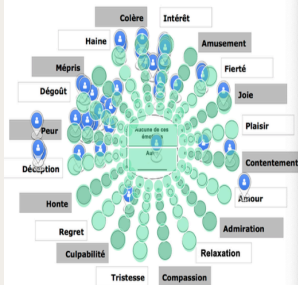


Perception générale (roue des émotions de Genève)

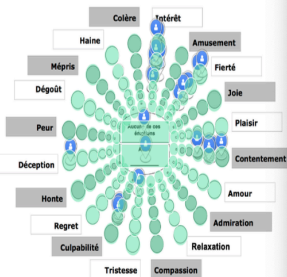


Perception générale - résultats

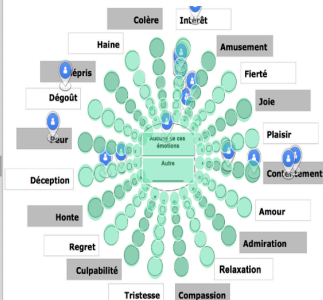
BioDataScience I



BioDataScience II



BioDataScience III



Perception du logiciel (System Usability Scale, Brooke 2013)



Concernant l'environnement logiciel utilisé au cours de Science des Données (R + RStudio + R Markdown), notez ces phrases par rapport à votre maîtrise ACTUELLE des outils.
(1 = tout à fait en désaccord, 2 = en désaccord, 3 = neutre, 4 = d'accord et 5 = tout à fait d'accord)

Propositions

Entrez les propositions que votre audience pourra noter

Je pense que j'aimerais utiliser fréquemment cet outil



Je trouve cet outil inutilement complexe



Je trouve cet outil facile à utiliser



Je pense que j'aurais besoin du support d'un spécialiste pour utiliser cet outil



Je trouve que les différentes fonctions de cet outil sont bien intégrées



Je trouve cet outil trop incohérent



Je pense que cet outil sera facile à apprendre pour beaucoup de personnes



Je trouve cet outil très contraignant à utiliser



Je me suis senti(e) en confiance lorsque j'ai utilisé cet outil



J'ai dû apprendre beaucoup de choses avant de me sentir familiarisé(e) avec cet outil



Interprétation de l'échelle SUS

Sur base de plusieurs centaines d'utilisations, une échelle indicative se dégage (Brooke 2013, SUS: a retrospective).

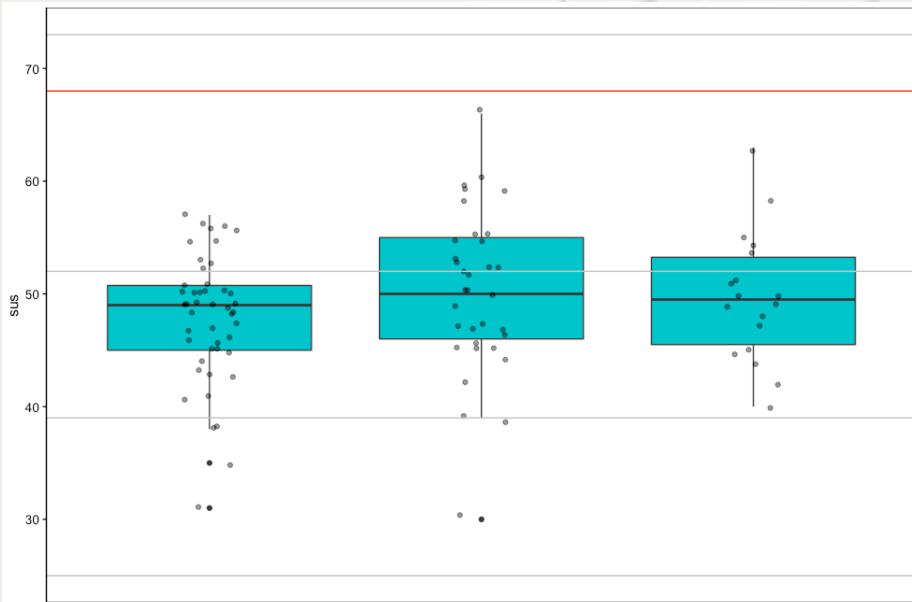
SUS

- 25
- 39
- 52
- **68**
- 73
- 86
- 100

Utilisabilité

- pire imaginable
- mauvais
- acceptable
- **score moyen**
- bon
- excellent
- meilleur imaginable

SUS logiciel R + RStudio - résultats



Conclusions

- Apprentissage **actif** en plusieurs **étapes** de difficulté croissante
- Tutoriels **learnr** efficaces pour *appliquer* (niveau 2) ; Charge cognitive perçue moyenne
- Beaucoup de concepts nouveaux au premier cours : **émotions générales** négatives qui évoluent en positif aux cours suivants
- Projets individuels et en groupe avec **R**, **RStudio** et **GitHub** : utilisabilité perçue plutôt faible

Avez-vous des questions ?



Ressources utiles

- Site web du cours : <https://wp.sciviews.org/>
- {learnitdown} : <https://www.sciviews.org/learnitdown/>
- cette présentation : https://github.com/BioDataScience-Course/sdd_presentations/tree/master/2021_rencontresr_paris