

La science des données pour l'ère moderne

CRKLVS 2021

Diapositives du tutoriel RStudio



Aperçu de la présentation

1. La vue d'ensemble
2. Aperçu de l'interface de R Studio
3. Certaines de mes expériences et motivations
4. Activité interactive (installation R et RStudio)
5. Langue de base
6. Fonctions principales
7. Fonctions Tidyverse
8. Exploration de l'ensemble des données de l'échantillon



La vue d'ensemble

- Systèmes d'acquisition
- Types de données
- Volume

Gestion des données

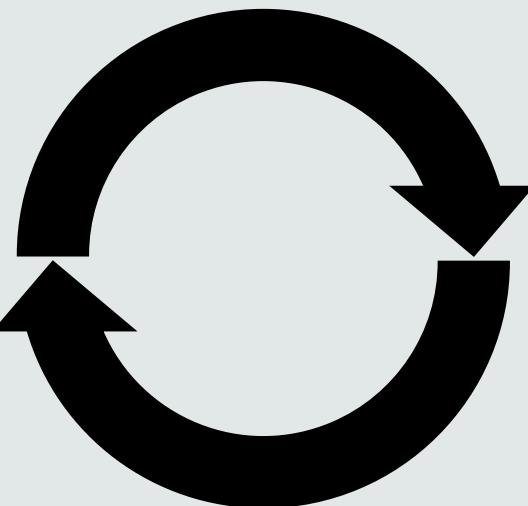
- Fiches techniques
- “Data wrangling”
- Analyse

Visualisation

- Exploration des données
- Chiffres
- Tableaux

Partage

- Dépôts
- Partage des données



Interface R Studio

The screenshot displays the RStudio interface with several red annotations overlaid:

- Environment pane (top right):** Shows the "Environment" tab selected. A red box highlights the text: "Environnement : Où sont enregistrés vos variables, fonctions et ensembles de données."
- Files pane (bottom right):** Shows the "Files" tab selected. A red box highlights the file list:

Name	Size	Modified
..		
CSEP2021.Rproj	218 B	Jul 30, 2021, 2:50 PM
CSEP2021_Tutorialcode.R	1.5 KB	Jul 30, 2021, 3:54 PM
- Console pane (bottom left):** Shows the "Console" tab selected. A red box highlights the text: "Console : Zone de test pour les lignes de code + retour immédiat sur les scripts".
- Script Editor pane (center-left):** Shows a script named "CSEP2021_Tutorialcode.R" with the following content:

```
Scripts : Comment
vous sauvegardez
votre travail et le
partagez avec des
collaborateurs
```

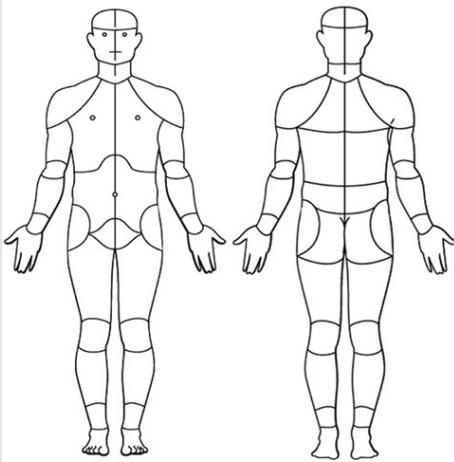
Pourquoi se donner la peine ?

Gratter la surface sur ce qui est possible...

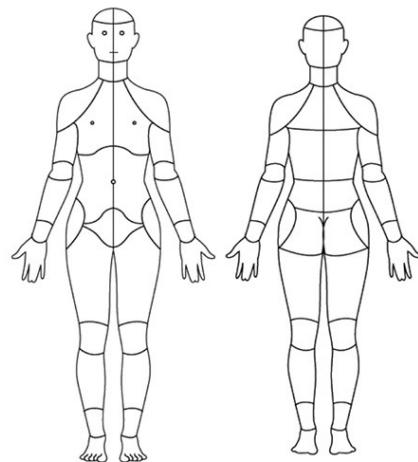


**UNE ÉTUDE DE FAISABILITÉ D'UN ESSAI CONTRÔLÉ RANDOMISÉ ET UN
PROTOCOLE DE RECHERCHE DANS L'ÉVALUATION D'UNE INTERVENTION EN
PHYSIOTHÉRAPIE PRÉOPÉRATOIRE ET POSTOPÉRATOIRE CHEZ DES PATIENTS
BARIATRIQUES DU CHU DGLD**

Select the areas where you are experiencing pain.



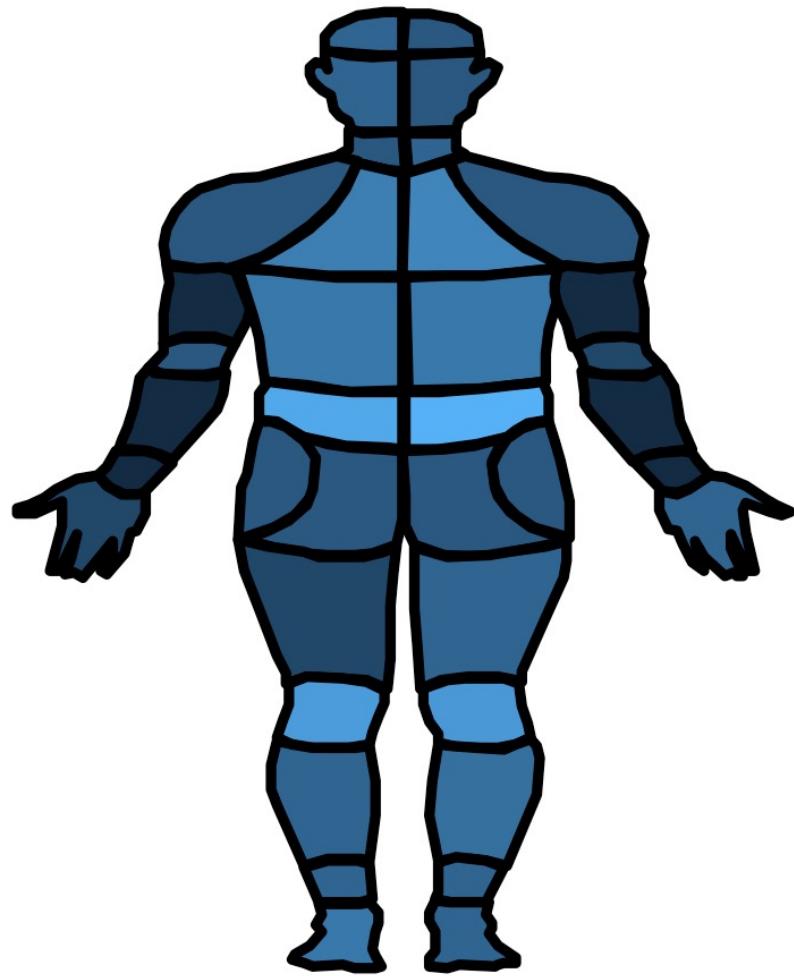
I have no pain



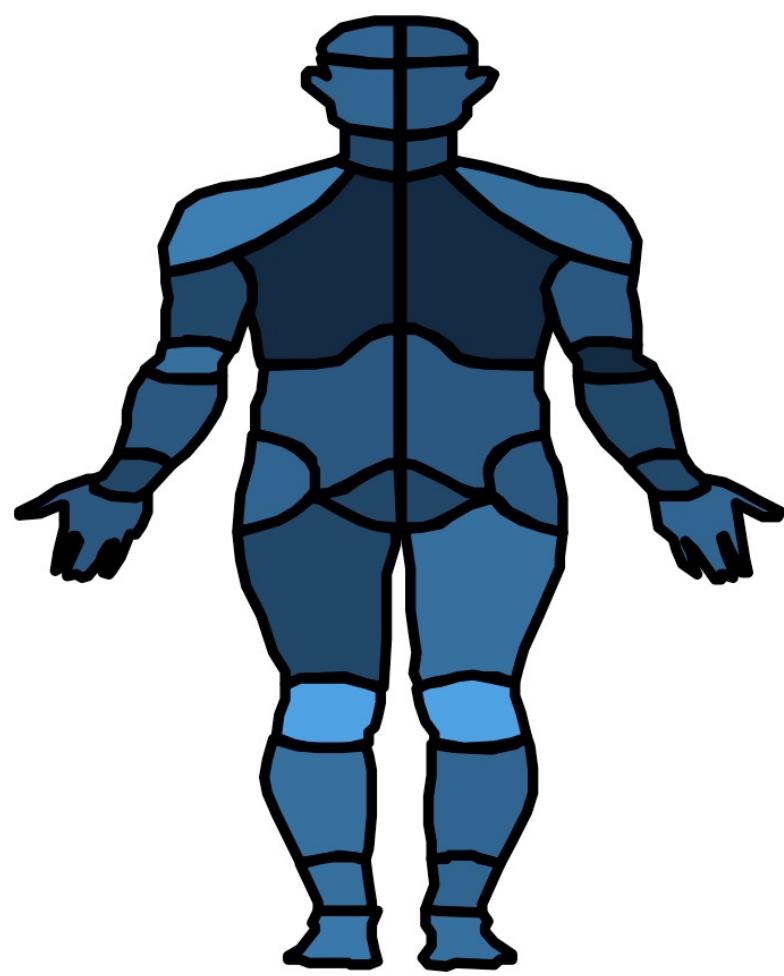
chevilles,épauleD, bas doD	133,107,219
lower and upper back	218,219,208,209
Gtricep, lower backD	112,219
lower back, back neckG	218,219,205,206
both knees	129,130
avant-brasD, bas abdoG	117,215
Ghip, Gknee,mid and lower back	123,130,212,213,218,219
front and back neck	218,225,239
piedD, back hipD, arrière piedG	135,225,237
Gknee, fesseG, mid lower back, upper back G	130,223,212,213,208
head, both knees, lower back	101,102,103,104,201,202,203,204
Glower back	218
Both shoulders and legs, back neck, lower back	107,110,207,210,127,128,129,130,131,132,133,134,229,230,231,232,233,234,235,236205,206,218,219
Both knees and lower back	129,130,231,232,218,219
both wrist and knees, Dthigh, both elbows, upper and lower back	119,124,129,130,231,232,127,113,116,215,216,208,209,218,219
both kneeD, foot, lower back	129,130,231,232,115,130,237,238
both knees, feet, both forearm, DG hip and upper and lower back	129,130,231,232,135,136,237,238,117,118,120,123,222,225,208,209
Dshoulder, upper mid and lower back	107,208,209,212,213,218,219
both biceps shoulder thigh, lower back and Dinguinal	111,112,107,110,127,128,218,219,121,122
both knees, lower back and D front elbow	129,130,231,232,218,219,113
Dknee	129,231
Dshoulder, lower back and both back thigh	107,218,219,229,230
front and back neck,both front elbow hands knees feet, Dgluteus and D ankle	105,106,205,206,113,116,125,126,129,130,135,136,227,228,231,232,237,238224,236,234
bas fesse D	224
Dknee	129,231
Gshoulder,Dhip,both knees and mid lower back	110,120,222,129,130,231,232,212,213,218,219
Dtête(avant et arrière), lower abdo D, scapulaD, coudeD, lower backD, calfD,arrière genouD	101,103,201,203,115,209,113,215,219,234,232
Devant chevilleD et G, ameve chevilleD et G, bas du dos	133,134,235,236,218,219
lower back	218,219
Avant et arrière 2 genoux, hanche D et G, fessier D et G	129,130,131,132
côté jambe D, scapulaD,milleu bas dos,dessus main D	131,233,209,218,219,125
lower back	218,219
Both shoulders and knees and lower back	107,110,207,210,129,130,231,232
head, upper back	101,102,103,104,201,202,203,204,208,209
Lower mid back, both glutes and go down	212,213,223,224,127,128,229,230,129,130,231,232,131,132,233,234,133,134,235,236,135,136,237,238



Back



Front



UNE ÉTUDE DE FAISABILITÉ D'UN ESSAI CONTRÔLÉ RANDOMISÉ ET UN PROTOCOLE DE RECHERCHE DANS L'ÉVALUATION D'UNE INTERVENTION EN PHYSIOTHÉRAPIE PRÉOPÉRATOIRE ET POSTOPÉRATOIRE CHEZ DES PATIENTS BARIATRIQUES DU CHUDGLD

QUESTIONNAIRE DE LAVAL

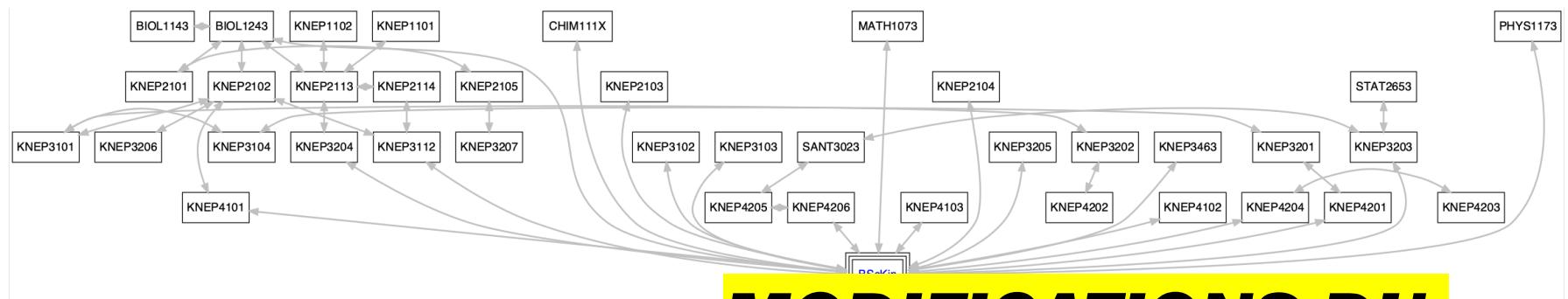
Ce questionnaire a été formulé pour nous faire savoir comment votre obésité a affecté votre vie au cours des 4 dernières semaines. Vous serez questionné à propos de l'impact que votre surplus de poids a pu avoir sur vos activités de tous les jours, vos émotions, votre vie sociale, votre vie sexuelle, votre hygiène corporelle, ainsi qu'à propos des symptômes que l'obésité a pu causer.

Au cours des 4 dernières semaines:	Tout le temps	La plupart du temps	Une bonne partie du temps	Une certaine partie du temps	Une petite partie du temps	Presque jamais	Jamais
1. Vous est-il arrivé d'être essoufflé rapidement au moindre effort?	1	2	3	4	5	6	7
2. Vous êtes-vous senti fatigué, épuisé?	1	2	3	4	5	6	7
3. Vous est-il arrivé de vous réveiller à plusieurs reprises durant la nuit?	1	2	3	4	5	6	7
4. Avez-vous eu l'impression d'avoir toujours chaud (transpiration excessive)?	1	2	3	4	5	6	7
5. Vous êtes-vous senti inquiet et/ou stressé?	1	2	3	4	5	6	7



Questionnaire_Level03	Questionnaire_Level02	Questionnaire_Level03	Questionnaire_Level04	Questionnaire_Level05	Questionnaire_Level06	Questionnaire_Level07	Questionnaire_Level08	Questionnaire_Level09	Questionnaire_Level10	Questionnaire_Level11	Questionnaire_Level12	Questionnaire_Level13	Questionnaire_Level14	Questionnaire_Level15	Questionnaire_Level16	Questionnaire_Level17	Questionnaire_Level18	Questionnaire_Level19	Questionnaire_Level20	Questionnaire_Level21	Questionnaire_Level22	Questionnaire_Level23	Questionnaire_Level24	Questionnaire_Level25	Questionnaire_Level26	Questionnaire_Level27	Questionnaire_Level28	Questionnaire_Level29	Questionnaire_Level30	Questionnaire_Level31	Questionnaire_Level32	Questionnaire_Level33	Questionnaire_Level34	Questionnaire_Level35	Questionnaire_Level36	Questionnaire_Level37																																																																							
6	5	3	4	2	3	4	5	6	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9																																																																		
6	5	3	4	2	3	4	5	6	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9	7	8	9																																																																		
4	3	2	3	4	5	6	7	8	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11	9	10	11																																																									
7	6	4	5	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100
3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100				
5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100		
2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100					
1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100						
5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100		
2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100					
2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100					
5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100		
3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100				
6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100	
5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70</																																





MODIFICATIONS DU PROGRAMME DE PREMIER CYCLE - CRÉATION D'UN PLAN





Marchez avec moi les mercredis !

[Accueil](#) [Balados](#) [Search](#)

KNEP-2105 - Biomécanique en Activité Physique

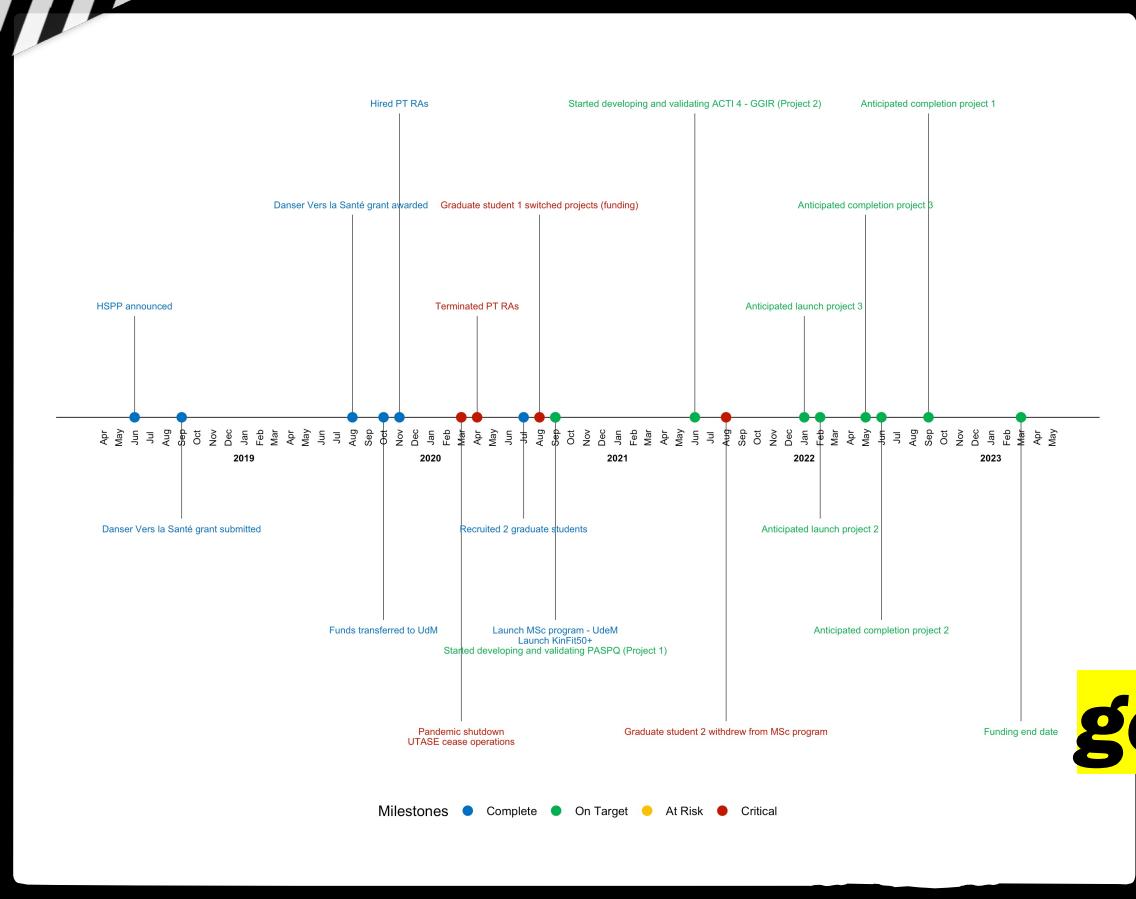
Activité asynchrone

Ceci est du contenu pour le cours 'KNEP 2105 - Biomécanique en AP' à l'Université de Moncton. Ce cours a été développé par Grant Handrigan. Profitez-en, et soyez attentifs à votre environnement lorsque vous écoutez !

KNEP-2105 –

**Biomécanique en
Activité Physique**





Planification stratégique – gestion du projet

Au travail !



**La semaine prochaine, nous ferons des
activités interactives. Soyez prêts à
programmer !**





Les diapositives suivantes
constituent un matériel de
référence pour après les
webinaires



Définitions

Console : Exécute immédiatement le code ; sert de banc d'essai pour les fonctions.

Scripts : Blocs de code exécutable qui peuvent être sauvegardés et améliorés au fil du temps - l'essentiel de votre travail

R Notebooks : Permet de passer naturellement du texte écrit au code exécutable, le tout dans le même fichier.

- Excellent moyen de communiquer les résultats avec les collègues

Projets : Système de tri des dossiers pour organiser vos études

- Les projets se chargent avec leur propre espace de travail et les derniers scripts sur lesquels vous travailliez.



Symboles de base

Syntaxe

<- (dénomination des variables)

= (arguments)

() (fonctions)

[] (indexation [row,col])

{ } (if, else, boucles)

Tout ce qui suit un # est du texte et n'exécutera pas de code.

? [fonction] (Aide ; LE PLUS IMPORTANT !!)

Logiques (VRAI = 1, FAUX = 0)

== exactement égal

!= n'est pas égal

&& et

< > >= <= supérieur à / inférieur à



Terminologie de base

Types variables :

- **Caractère** : Variables de type chaîne de caractères
- **Numérique** : Tout type de nombre
- **Entier** : Nombres entiers
- **Logiques** : VRAI ou FAUX
- **Facteur** : Un classificateur de caractères (par exemple, sédentaire, actif, en forme).

values	
a	"words"
b	3.14
c	4
d	TRUE

Types de structures de données

- **Vecteurs** : Groupe de variables de même type (par exemple, tous les mots, tous les nombres).
- **Cadres de données** : Tableaux de données dont les lignes représentent



Paquets et bibliothèques

R est un logiciel libre : Tout le monde peut créer un paquet et le rendre disponible en ligne.

- Pour accéder à ces fonctions spécialisées, vous devez installer les paquets et charger la bibliothèque spécifique.
- Vous ne devez installer le paquet qu'une seule fois, mais vous devez accéder à la bibliothèque chaque fois que vous démarrez une session RStudio.

EXEMPLES

```
install.packages("tidyverse") install.packages("gameofthrones")
```

```
library(tidyverse) library(gameofthrones)
```

Notez quand vous utilisez et n'utilisez pas de "citations" !



Fonctions

Fonction : Une instruction préprogrammée qui prend des entrées (arguments) et fournit une sortie.

- Vous êtes probablement déjà familier avec les fonctions d'Excel !
- Les fonctions fonctionnent de manière très similaire en R, mais peuvent être compliquées

ID	VO2max
S01	52
S02	36
S03	34
S04	43
S05	38
S06	39
S07	32
S08	33
S09	42
S10	37
MEAN	=AVERAGE(C3:C12)
SD	AVERAGE(number1, [number2], ...)

```
read.csv(file, header = TRUE, sep = ",", quote = "\"",  
        dec = ".", fill = TRUE, comment.char = "", ...)
```

Par exemple, `read.csv()` est la fonction, et tout ce qui se trouve entre les crochets sont les "arguments" que vous devez fournir pour lui donner des instructions.

Fonctions principales (exemples)

Fonction	Description	Exemple
seq()	Créez une séquence de chiffres	seq(0, 100, by = 10)
c()	Concaténation d'un groupe d'éléments	c(2, 4, 6, 8, 10) c("R ", "is ", "fun !")
coller()	Concatène deux chaînes de caractères ensemble	paste(1:3, c("st", "nd", "rd")) - “1 st ”, “2 nd ”, “3 ^{rd”}
rnorm()	Crée une distribution normale aléatoire	rnorm(n=100, moyenne = 250, sd=20) - 100 observations avec une moyenne de 250 et un écart-type de 20)
échantillon()	Réorganisation aléatoire d'un vecteur	échantillon(c("A", "B", "C")) [1] "B", "C", "A"
as.integer() as.character() facteur()	Convertit en nombre entier Convertit en caractère Convertit en un facteur	
data.frame()	Combine les vecteurs dans un objet de type cadre de données	data.frame(v1,v2)



Fonctions Tidyverse

Le "Tidyverse" est une collection de fonctions très utiles et largement utilisées dans le domaine de la science des données.

Qu'est-ce qu'une donnée "ordonnée" ?

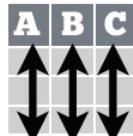
- Chaque variable est dans une colonne
- Chaque observation est une ligne
- Chaque valeur est une cellule



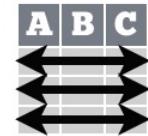
Tidy Data with `tidyverse`

Tidy data is a way to organize tabular data. It provides a consistent data structure across packages.

A table is tidy if:



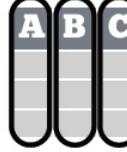
&



Each **variable** is in its own **column**

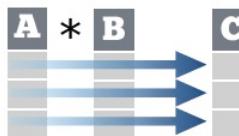
Each **observation**, or **case**, is in its own **row**

Tidy data:



Makes variables easy to access as vectors

$A * B \rightarrow C$



Preserves cases during vectorized operations

Readr : Fonctions d'importation de données

Read Tabular Data

- These functions share the common arguments:

```
read_*(file, col_names = TRUE, col_types = NULL, locale = default_locale(), na = c("", "NA"),
quoted_na = TRUE, comment = "", trim_ws = TRUE, skip = 0, n_max = Inf, guess_max = min(1000,
n_max), progress = interactive())
```

a,b,c		A	B	C
1,2,3		1	2	3
4,5,NA		4	5	NA

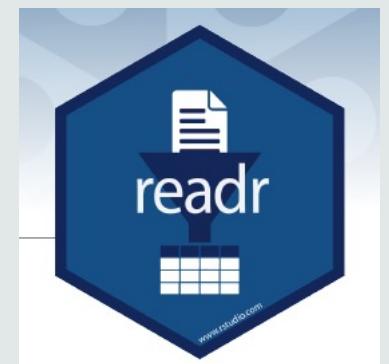


Comma Delimited Files

`read_csv("file.csv")`

To make file.csv run:

```
write_file(x = "a,b,c\n1,2,3\n4,5,NA", path = "file.csv")
```



USEFUL ARGUMENTS

a,b,c	
1,2,3	
4,5,NA	

Example file

```
write_file("a,b,c\n1,2,3\n4,5,NA","file.csv")
f<- "file.csv"
```

1	2	3
4	5	NA

Skip lines

`read_csv(f, skip = 1)`

A	B	C
1	2	3
4	5	NA

No header

`read_csv(f, col_names = FALSE)`

A	B	C
1	2	3
4	5	NA

Read in a subset

`read_csv(f, n_max = 1)`

x	y	z
A	B	C
1	2	3
4	5	NA

Provide header

`read_csv(f, col_names = c("x", "y", "z"))`

A	B	C
NA	2	3
4	5	NA

Missing Values

`read_csv(f, na = c("1", "."))`



Tidyr : Fonctions pour le rangement des données

Reshape Data - change the layout of values in a table

Use **gather()** and **spread()** to reorganize the values of a table into a new layout.

**gather(data, key, value, ..., na.rm = FALSE,
convert = FALSE, factor_key = FALSE)**

gather() moves column names into a **key** column, gathering the column values into a single **value** column.

table4a

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K

→

country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

key value

```
gather(table4a, `1999`, `2000`,  
key = "year", value = "cases")
```

**spread(data, key, value, fill = NA, convert = FALSE,
drop = TRUE, sep = NULL)**

spread() moves the unique values of a **key** column into the column names, spreading the values of a **value** column across the new columns.

table2

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T

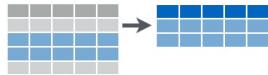
key value

```
spread(table2, type, count)
```



Dplyr : Fonctions pour le traitement des données

Subset Observations (Rows)



`dplyr::filter(iris, Sepal.Length > 7)`

Extract rows that meet logical criteria.

`dplyr::distinct(iris)`

Remove duplicate rows.

`dplyr::sample_frac(iris, 0.5, replace = TRUE)`

Randomly select fraction of rows.

`dplyr::sample_n(iris, 10, replace = TRUE)`

Randomly select n rows.

`dplyr::slice(iris, 10:15)`

Select rows by position.

`dplyr::top_n(storms, 2, date)`

Select and order top n entries (by group if grouped data).

Summarise Data



`dplyr::summarise(iris, avg = mean(Sepal.Length))`

Summarise data into single row of values.

`dplyr::summarise_each(iris, funs(mean))`

Apply summary function to each column.

`dplyr::count(iris, Species, wt = Sepal.Length)`

Count number of rows with each unique value of variable (with or without weights).



Summarise uses **summary functions**, functions that take a vector of values and return a single value, such as:

`dplyr::first`

First value of a vector.

`min`

Minimum value in a vector.

`dplyr::last`

`max`

Maximum value in a vector.

`dplyr::nth`

`mean`

Mean value of a vector.

`dplyr::n`

`median`

Median value of a vector.

`dplyr::n_distinct`

`var`

Variance of a vector.

`IQR`

`sd`

Standard deviation of a vector.

Subset Variables (Columns)



`dplyr::select(iris, Sepal.Width, Petal.Length, Species)`

Select columns by name or helper function.



Dplyr : Fonctions pour le traitement des données

dplyr::%>%

Passes object on left hand side as first argument (or . argument) of function on righthand side.

`x %>% f(y)` is the same as `f(x, y)`
`y %>% f(x, ., z)` is the same as `f(x, y, z)`

"Piping" with `%>%` makes code more readable, e.g.

```
iris %>%
  group_by(Species) %>%
  summarise(avg = mean(Sepal.Width)) %>%
  arrange(avg)
```

Make New Variables



`dplyr::mutate(iris, sepal = Sepal.Length + Sepal.Width)`

Compute and append one or more new columns.

`dplyr::mutate_each(iris, funs(min_rank))`

Apply window function to each column.

`dplyr::transmute(iris, sepal = Sepal.Length + Sepal.Width)`

Compute one or more new columns. Drop original columns.



Mutate uses **window functions**, functions that take a vector of values and return another vector of values, such as:

Group Data

`dplyr::group_by(iris, Species)`

Group data into rows with the same value of Species.

`dplyr::ungroup(iris)`

Remove grouping information from data frame.

`iris %>% group_by(Species) %>% summarise(...)`

Compute separate summary row for each group.



ggplot2 : Grammaire des graphes

Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and geoms—visual marks that represent data points.



To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.



Complete the template below to build a graph.

```
ggplot (data = <DATA>) +  
  <GEO_FUNCTION> (mapping = aes(<MAPPIGS>),  
    stat = <STAT>, position = <POSITION>) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

required
Not required, sensible defaults supplied

`ggplot(data = mpg, aes(x = cty, y = hwy))` Begins a plot that you finish by adding layers to. Add one geom function per layer.

discrete x , continuous y
`f <- ggplot(mpg, aes(class, hwy))`

`f + geom_col()`, x, y, alpha, color, fill, group, linetype, size

`f + geom_boxplot()`, x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight

`f + geom_dotplot(binaxis = "y", stackdir = "center")`, x, y, alpha, color, fill, group

`f + geom_violin(scale = "area")`, x, y, alpha, color, fill, group, linetype, size, weight

