

Aphantasia Project

Maël Delem, Colin Fourment, Thomas Junoy, Guillaume Leal de Almeida

2023-02-11

Contents

1	Introduction	1
2	Simulation des données	1
2.1	Mesures (ou variables) et groupes	2
2.2	Variables et capacités cognitives	2
2.3	Interactions entre capacités cognitives	3
2.4	Fonction de simulation	3
3	Analyse de données	3
3.1	Corrélations	3
	Références	7
	Annexe : code entier du dossier	8

1 Introduction

Testons cette histoire de citations. Nous avons un premier article, Cavedon-Taylor (2021), un deuxième article qui nous informe de ce fait là (Fazekas et al., 2021), et cela est vérifié par nombre de leurs collègues (Crowder, 2018; Keogh & Pearson, 2018; Nalborczyk, 2019). Il faut insister néanmoins sur Dawes (2020), dont le papier est réellement central.(Aarsland et al., 2005).

Donc nous avons un moyen de partage et d'édition via trackdown. Testons le download cette fois-ci !

2 Simulation des données

Nous allons simuler des données correspondant à des hypothèses possibles sur nos problématiques de recherche. Pour cela, il faut penser “à l'envers” et repartir de ce qu'on évalue à la base : deux types (selon la littérature) de personnes, les aphantasiques et les non-aphantasiques. Donc nous allons simuler deux groupes correspondants, “Aph” et “Non_A”.

Mais quoi simuler ? Nous allons leur faire passer des tests. Un certain nombre d'ailleurs :

- des tests d'*imagerie visuelle* “*objet*” (la visualisation des formes, des couleurs, de la luminosité) - l'**OSIQ-Objet** et le **VVIQ**
- des tests d'*imagerie spatiale* (la visualisation des positions et mouvements dans l'espace) - le **OSIQ-Spatial**, le **MRT**, le **SRI**
- des tests de raisonnement - les matrices de **Raven**, le test des **similitudes**, la tâche de **Wason**, et des tests de compréhension en **lecture**
- et enfin des tâches évaluant les fonctions exécutives (la mémoire à court terme, la flexibilité mentale) - l'**empan** mnésique et les blocs de **Corsi** pour la mémoire, et le **WCST** pour la flexibilité.

2.1 Mesures (ou variables) et groupes

Nous allons donc créer un dataframe pour chaque groupe (**Non_A** et **Aph**) contenant leurs moyennes et écarts-types à ces tâches, le nom de leur groupe et le nombre de participants testés. Nous fusionnons ces groupes dans un dataframe **variables**:

```
##      name mean  sd group n_subjects
## 1  OSIQ_0 32.50 8.45  Aph         200
## 2  OSIQ_S 58.90 9.54  Aph         200
## 3   VVIQ 30.20 9.67  Aph         200
## 4   Raven 23.60 4.24  Aph         200
## 5  Simili 42.20 6.15  Aph         200
## 6   Wason 36.10 3.47  Aph         200
## 7 Empan_MDT 7.53 1.45  Aph         200
## 8   WCST 33.80 2.62  Aph         200
## 9  Lecture 48.40 9.67  Aph         200
## 10  Corsi  6.80 1.65  Aph         200
## 11   MRT 18.20 5.78  Aph         200
## 12   SRI 38.50 8.21  Aph         200
## 13 OSIQ_0 54.60 8.45 Non_A        200
## 14 OSIQ_S 46.20 9.54 Non_A        200
## 15  VVIQ 63.80 9.67 Non_A        200
## 16  Raven 20.90 5.34 Non_A        200
## 17  Simili 37.80 4.25 Non_A        200
## 18  Wason 32.20 3.78 Non_A        200
## 19 Empan_MDT 6.43 2.12 Non_A        200
## 20  WCST 32.10 5.32 Non_A        200
## 21  Lecture 50.20 8.89 Non_A        200
## 22  Corsi  5.81 1.87 Non_A        200
## 23   MRT 16.50 3.54 Non_A        200
## 24   SRI 35.70 6.23 Non_A        200
```

2.2 Variables et capacités cognitives

Nous devons maintenant simuler des données, mais sans ajout supplémentaire ici les scores seraient parfaitement aléatoires et leur analyse ne reflèteraient pas grand-chose à part les différences entre moyennes établies ici. Nous devons donc postuler des liens entre ces douze variables. Comme dit plus haut, ces variables sont censées évaluer au moins cinq capacités cognitives :

- l'imagerie visuelle-objet
- l'imagerie visuo-spatiale
- le raisonnement fluide
- la mémoire de travail/à court-terme
- la flexibilité mentale

Nous allons donc créer un “*modèle de mesures*”, une matrice **fmodel** de 12*5 définissant les liens entre nos 12 variables et nos 5 capacités par des coefficients :

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.8 0.00  0.0  0.0 0.00
## [2,]  0.0 0.90  0.0  0.0 0.00
## [3,]  0.9 0.00  0.0  0.0 0.00
## [4,]  0.1 0.30  0.8  0.0 0.05
## [5,] -0.2 0.00  0.6  0.0 0.10
## [6,] -0.1 0.00  0.3  0.0 0.00
## [7,]  0.0 0.00  0.0  0.8 0.00
## [8,] -0.1 0.00  0.2  0.0 0.60
```

```
## [9,] 0.4 0.00 0.6 0.0 0.00
## [10,] 0.1 0.70 0.0 0.8 0.00
## [11,] 0.2 0.85 0.0 0.0 0.00
## [12,] 0.1 0.90 0.0 0.0 0.00
```

2.3 Interactions entre capacités cognitives

Un dernier point n'est pas réaliste : jusque-là les cinq capacités cognitives définies sont parfaitement indépendantes. Cela n'est vraisemblablement pas le cas en réalité. Nous allons donc créer une matrice **effects** de 5*5 définissant les effets réciproques des fonctions entre elles, sur la base de nos hypothèses et de la littérature :

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 1.0 -0.1 -0.1 0.2 0.0
## [2,] -0.1 1.0 0.3 0.2 0.0
## [3,] -0.1 0.3 1.0 0.0 0.2
## [4,] 0.2 0.2 0.0 1.0 0.0
## [5,] 0.0 0.0 0.2 0.0 1.0
```

2.4 Fonction de simulation

Nous avons désormais tous les éléments pour simuler des données intéressantes, avec des liens sous-jacents à la fois entre les variables mesurées mais aussi entre les capacités sous-jacentes qu'évaluent ces variables.

Notre fonction **simulation** prendra donc pour argument notre dataframe **variables** pour générer les scores de chaque participant en fonction des moyennes de son groupe, et les dataframes **fmodel** et **effects** pour pondérer les scores de sorte à avoir des relations entre ceux-ci. Il n'y a plus qu'à simuler nos données aléatoires ! Enfin, du moins celles liées à la `seed(14051998)`.

```
#simulation du dataset
data <- simulation(variables,fmodel,effect)
```

3 Analyse de données

Nous avons enfin des données à traiter ! Maintenant, comme première étape de l'analyse de données - et avant la modélisation des données - nous allons réaliser beaucoup de visualisation des données, et dans un second temps les re-partitionner (*clustering*).

Pour l'analyse de données nous allons nous débarrasser des différences d'échelle potentiellement grandes entre les scores en standardisant ceux-ci avec des z-scores (on peut remarquer qu'on vient ici "défaire" ce qu'on a fait plus haut en transformant les scores aléatoires pour leur donner les vraies moyennes, illustrant bien le "reverse engineering" utilisé pour simuler les données) :

```
data_scale <- data %>%
  select(OSIQ_0:SRI) %>%
  mutate(across(everything(), ~ scale(.x)))
```

3.1 Corrélations

Nous allons pouvoir effectuer nos premières analyses, dans un premier temps sur les corrélations entre nos variables.

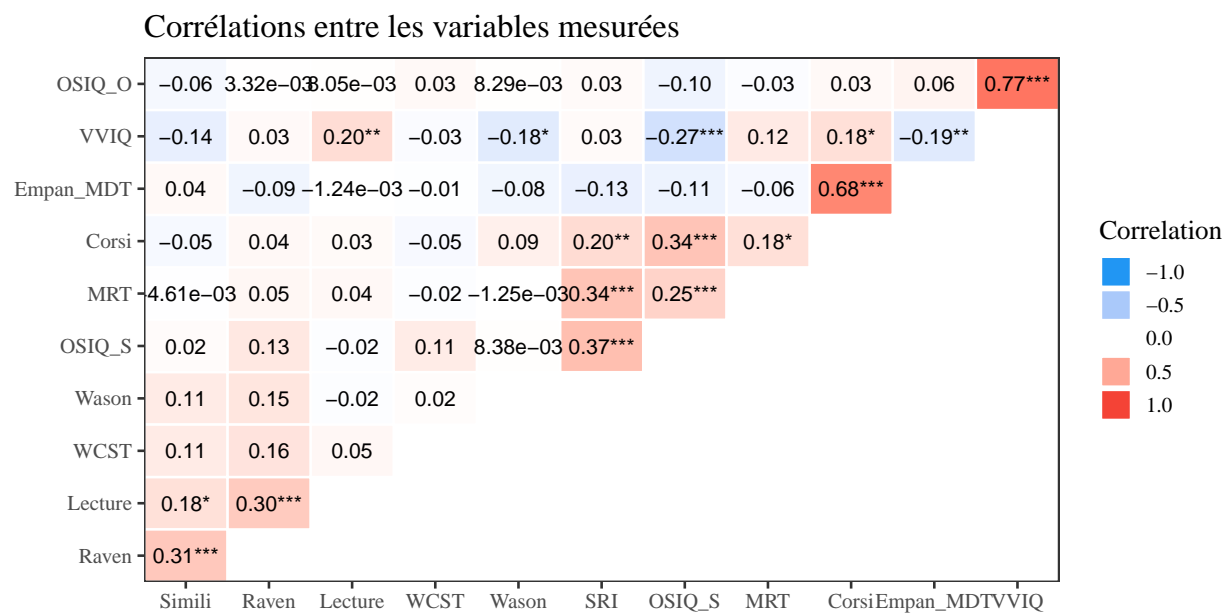


Figure 1: Les lunettes à Matrix

Les graphiques en réseau sont aussi des visualisations assez esthétiques et intuitives des relations entre des variables, c'est une visualisation complémentaire à la matrice, qui elle est plus "factuelle" :

Corrélations entre les variables mesurées

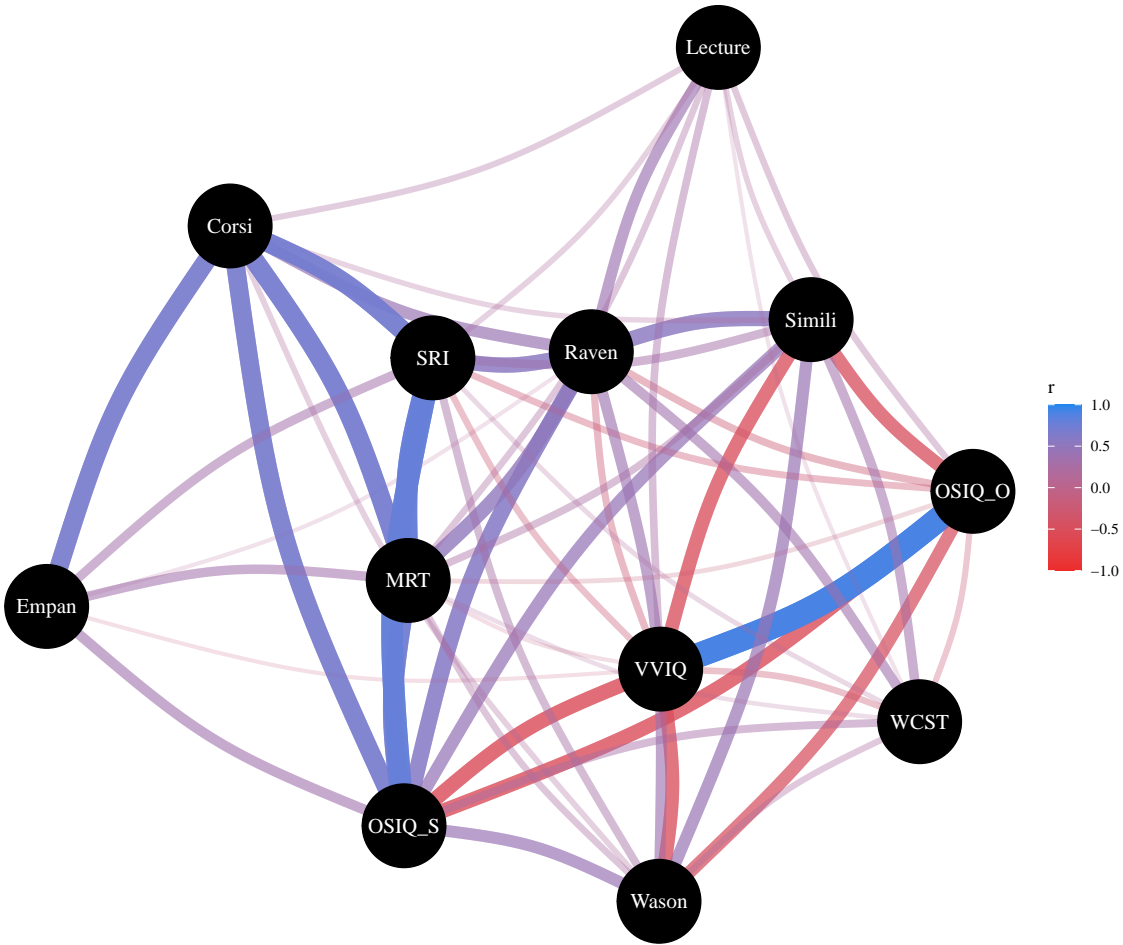


Figure 2: Network

Lalalalalala

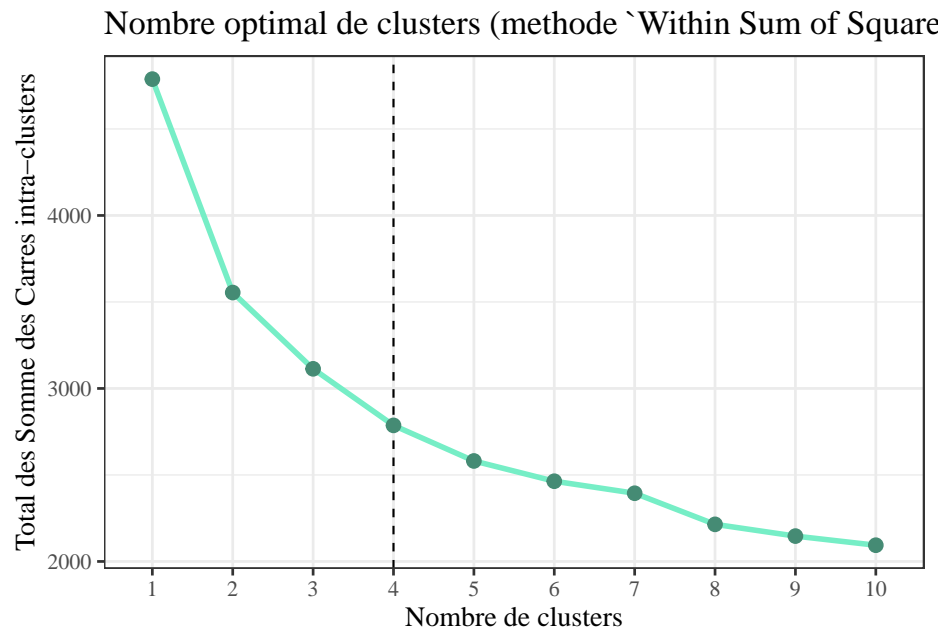


Figure 3: Clusters

Références

- Aarsland, D., Zaccai, J., & Brayne, C. (2005). A systematic review of prevalence studies of dementia in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 20(10), 1255–1263.
- Cavedon-Taylor, D. (2021). *Predictive Processing and Perception: What does Imagining have to do with it?* 15.
- Crowder, A. (2018). *Differences in Spatial Visualization Ability and Vividness of Spatial Imagery Between People With and Without Aphantasia*.
- Dawes, A. J., Keogh, R., Andrillon, T., & Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports*, 10(1, 1), 10022. <https://doi.org/10.1038/s41598-020-65705-7>
- Fazekas, P., Nanay, B., & Pearson, J. (2021). Offline perception: An introduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1817), 20190686. <https://doi.org/10.1098/rstb.2019.0686>
- Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, 105, 53–60. <https://doi.org/10.1016/j.cortex.2017.10.012>
- Nalborczyk, L. (2019). *Understanding rumination as a form of inner speech: Probing the role of motor processes* [Preprint]. Thesis Commons. <https://doi.org/10.31237/osf.io/p6dct>

Annexe : code entier du dossier

```
knitr::opts_chunk$set(message=FALSE,
                        warning=FALSE,
                        echo=FALSE,
                        fig.pos = "H",
                        fig.align = "center",
                        out.extra = "")

# ---- package loading with librarian (loaded on startup) ----
shelf(
  MASS,          # functions and data frame ecosystem
  easystats,     # modelling, visualization and reporting ecosystem
  ez,            # analysis and visualization of factorial exp
  rstatix,       # pipe friendly statistical functions
  scales,        # scale functions
  corrr,         # correlations
  lme4,          # mixed models
  lmerTest,      # tests in lmer
  cluster,       # cluster analysis
  factoextra,    # multivariate data analysis visualization
  GGally,        # more ggplot2 plots
  ggpubr,        # publication plots
  ggradar,       # radar charts
  ggraph,        # auto graph layout
  igraph,        # network graphs
  crsh/citr,     # citations
)

# global theme
theme_set(theme_bw(base_size = 14, base_family = "serif"))

# fixed random seed
set.seed(14051998)
Non_A <- data.frame(
  name = c("OSIQ_0", "OSIQ_S", "VVIQ",
           "Raven", "Simili", "Wason",
           "Empan_MDT", "WCST", "Lecture",
           "Corsi", "MRT", "SRI"),

  mean = c(54.6, 46.2, 63.8,
           20.9, 37.8, 32.2,
           6.43, 32.1, 50.2,
           5.81, 16.5, 35.7),

  sd = c(8.45, 9.54, 9.67,
         5.34, 4.25, 3.78,
         2.12, 5.32, 8.89,
         1.87, 3.54, 6.23),

  group = ("Non_A") %>% factor(),
  n_subjects = 200
)
```



```

Aph <- data.frame(
  name = c("OSIQ_0", "OSIQ_S", "VVIQ",
            "Raven", "Simili", "Wason",
            "Empan_MDT", "WCST", "Lecture",
            "Corsi", "MRT", "SRI"),

  mean = c(32.5, 58.9, 30.2,
            23.6, 42.2, 36.1,
            7.53, 33.8, 48.4,
            6.8, 18.2, 38.5),

  sd = c(8.45, 9.54, 9.67,
          4.24, 6.15, 3.47,
          1.45, 2.62, 9.67,
          1.65, 5.78, 8.21),

  group = ("Aph") %>% factor(),
  n_subjects = 200
)

variables <- bind_rows(Aph, Non_A)
rm(Aph, Non_A)

variables
fmodel <- matrix(c ( .8, 0, 0, 0, 0, # OSIQ-0 = img objet
                     0, .9, 0, 0, 0, # OSIQ-S = img spatiale
                     .9, 0, 0, 0, 0, # VVIQ = img objet
                     .1, .3, .8, 0, .05, # Raven = raisonnement > img s/o > flex
                     -.2, 0, .6, 0, .1, # Simili = raisonnement > flex
                     -.1, 0, .3, 0, 0, # Wason = raisonnement
                     0, 0, 0, .8, 0, # Empan = MDT
                     -.1, 0, .2, 0, .6, # WCST = Flex > raisonnement
                     .4, 0, .6, 0, 0, # Lecture = img objet > raisonnement
                     .1, .7, 0, .8, 0, # Corsi = MDT > img s
                     .2, .85, 0, 0, 0, # MRT = img s
                     .1, .9, 0, 0, 0 # SRI = img s
                     ),
                 nrow=12, ncol=5, byrow=TRUE)

fmodel
effect <- matrix(c ( 1, -.1, -.1, .2, 0, # img o
                    -.1, 1, .3, .2, 0, # img s
                    -.1, .3, 1, 0, .2, # raisonnement
                    .2, .2, 0, 1, 0, # MDT
                    0, 0, .2, 0, 1 # flex
                    ),
                 nrow=5, ncol=5, byrow=TRUE)

effect
simulation <- function(variables, fmodel, effect) {

  ### preparatifs ###
  n_variables <- dim(fmodel)[1] # notre nb de mesures/variables (rows)
  n_skills <- dim(fmodel)[2]    # les capacités sous-jacentes évaluées (columns)

  # matrice de poids des erreurs

```

```

errorweight <- (1 - diag(fmodel %*% t(fmodel))) %>%
  abs() %>% # necessaire pour la racine carree
  sqrt() %>% # doit avoir des arguments positifs
  diag() # recree une matrice diagonalisee

# initialisation d'un dataframe vide
data <- data.frame()

### simulation ###
for (i in levels(variables$group)){ # on simule separement chaque groupe

  var_group = variables %>% filter(group == i) # donnees du groupe isolees
  n_subjects = var_group$n_subjects[1] # nb de sujets dans le groupe
  group = i # nom du groupe

  # generation de scores aleatoires normaux pour chaque capacite cognitive
  randomscores <- matrix(rnorm(n_subjects * (n_skills)),
    nrow = n_subjects,
    ncol = n_skills)
  # ponderation par la matrice d'effets = les scores sont desormais correles
  # entre eux
  skillscores <- randomscores %*% effect

  # genere les valeurs des mesures/variables grace a fmodel
  observedscores <- skillscores %*% t(fmodel)

  # generation d'erreurs normales pour chaque mesure/variable
  randomerror <- matrix(rnorm(n_subjects * (n_variables)),
    nrow = n_subjects,
    ncol = n_variables)
  # ponderation par notre matrice de poids des erreurs
  error <- randomerror %*% errorweight

  # nos mesures effectives = les valeurs reeles + une erreur standard
  measures <- observedscores + error

  # on cree un dataframe avec le nom du groupe
  data_group <- data.frame(measures) %>%
    mutate(Group = group %>% factor())

  # ajout des valeurs reeles de moyenne et d'ecart-type pour chaque variable
  # et renommage
  for (i in 1:length(var_group$name)){
    data_group[,i] = data_group[,i]*var_group$sd[i] + var_group$mean[i]
    colnames(data_group)[i] = var_group$name[i]
  }

  # fusion avec le dataframe complet
  data <- bind_rows(data,data_group)
}

# ajout d'id individuels et stats demographiques
n = length(data[,1]) # nombre total de participants

```

```

data <- data %>%
  mutate(Subject_nr = row_number() %>% as.character(),
         Sex = (c("H", "F") %>% rep(times = n/2) %>% factor()),
         Age = seq(from = 16, to = 55, by = 1) %>% sample(size = n,
                                                         replace = TRUE)
         ) %>%
  relocate(Subject_nr)

# mission accomplished!
return(data)
}

#simulation du dataset
data <- simulation(variables,fmodel,effect)
data_scale <- data %>%
  select(OSIQ_0:SRI) %>%
  mutate(across(everything(), ~ scale(.x)))
# matrice de correlation entre les mesures
data_scale %>%
  correlation(partial = TRUE) %>%
  cor_sort() %>%
  summary() %>%
  visualisation_recipe(
    labs = list(title = "Corrélations entre les variables mesurées")) %>%
  plot() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
data_scale %>%
  correlate() %>%
  stretch %>%
  filter(abs(r) >= .1) %>%
  mutate(x = replace(x, x == "Empan_MDT", "Empan"),
         y = replace(y, y == "Empan_MDT", "Empan")) %>%
  graph_from_data_frame(directed = FALSE) %>%
  ggraph() +
  geom_edge_arc(strength = 0.1,
               aes(edge_alpha = abs(r),
                  edge_width = abs(r),
                  colour = r)) +
  guides(edge_alpha = "none", edge_width = "none") +
  scale_edge_colour_gradientn(limits = c(-1, 1),
                             colors = c("firebrick2", "dodgerblue2")) +
  geom_node_point(size = 20) +
  geom_node_text(aes(label = name,
                    family = "serif"),
               colour = "white",
               repel = FALSE) +
  theme_graph(base_family = "serif", base_size = 10) +
  #theme(legend.position = "none") +
  labs(title = "Corrélations entre les variables mesurées")
data_scale %>%
  fviz_nbclust(kmeans, method = "wss",
              linecolor = "white") +

```

```

geom_vline(xintercept = 4, linetype = 2) +
theme_bw(base_size = 14, base_family = "serif") +
geom_line(aes(group = 1), color = "aquamarine2",size = 1.3) +
geom_point(group = 1, size = 3, color = "aquamarine4") +
labs(title = "Nombre optimal de clusters (methode `Within Sum of Squares`)",
      x = "Nombre de clusters",
      y = "Total des Somme des Carres intra-clusters")

```