Rapport sur le modèle de mélange

Projet tutoré

Océane Baboulaz, Alice Chauveau
Charline Champ et Morgane Maulet

M2 SSD
Année universitaire 2023/2024

Contents

Introduction

Dans cette étude, nous explorons les nuances psychologiques des coureurs de trail en utilisant des techniques de clustering pour identifier des groupes homogènes au sein de cette communauté. L'objectif principal est d'analyser les profils distincts émergents au sein de cette population hétérogène, en mettant l'accent sur les caractéristiques individuelles telles que la motivation, l'anxiété, les habitudes de vie et d'autres aspects psychologiques.

À travers une approche méthodologique basée sur le clustering gaussien implémentée avec le package *mclust* en *R*, cette étude cherche à déterminer le nombre optimal de groupes représentatifs des coureurs de trail. Une fois ces groupes identifiés, nous nous pencherons sur les différences significatives entre eux, afin de mieux comprendre les dynamiques psychologiques et comportementales qui sous-tendent la passion commune du trail running.

Modélisation par clustering (mclust)

Nous avons élaboré divers modèles de mélanges afin d'explorer nos données. Pour simplifier, un modèle de mélange est un outil qui nous guide pour déceler différents groupes ou composants au sein de nos informations, ce qui peut être précieux pour dévoiler des schémas ou des tendances dissimulées. Pour choisir le modèle le plus adapté, nous avons ajusté la variable "G" de 1 à 5, en observant la meilleure valeur du BIC, nous permettant ainsi de sélectionner notre modèle optimal.

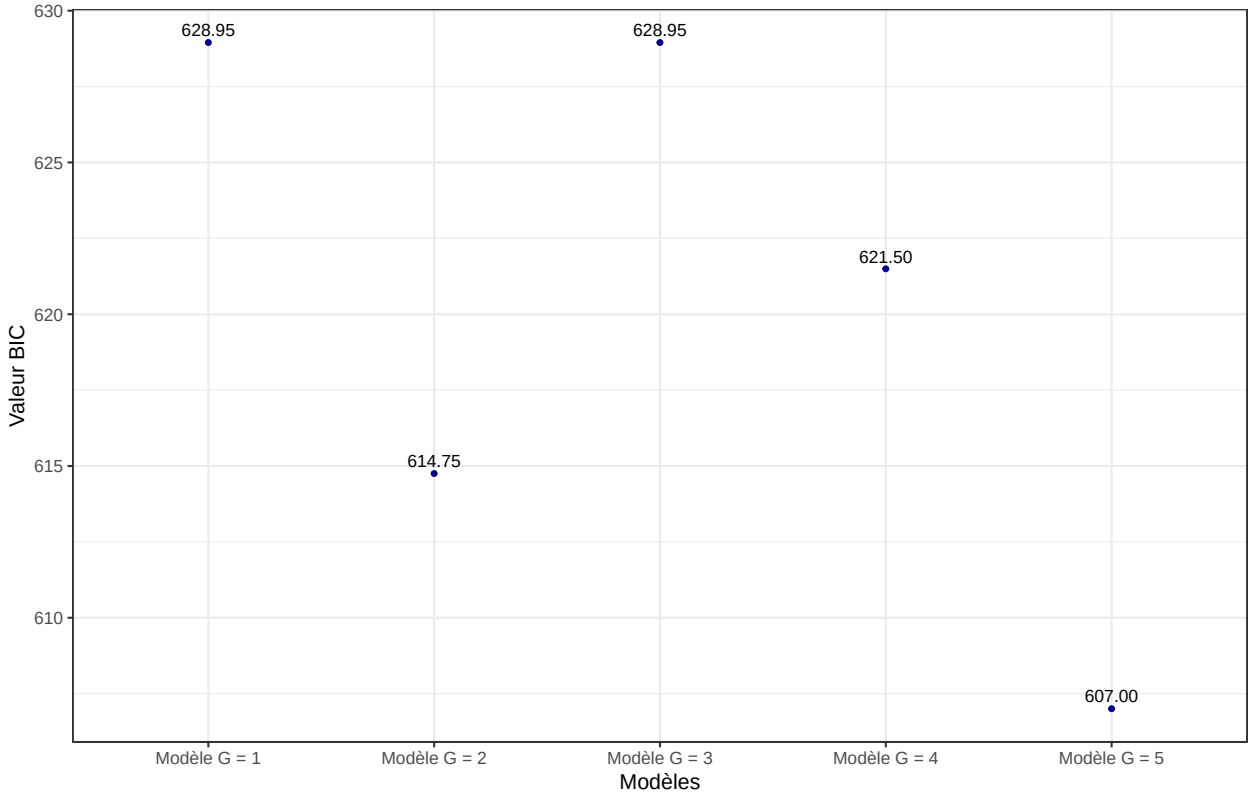


Figure 1: Comparaison des valeurs BIC pour différents modèles

Nous avons choisi le modèle de clustering avec deux groupes, car il a la meilleure valeur du BIC (Bayesian Information Criterion). Cela suggère qu'il correspond le mieux à la structure sous-jacente des données des coureurs de trail.

Attribution des clusters aux données

Voici la sortie générée par R, nous permettant d'examiner notre modèle choisi. Le code suivant a été utilisé pour créer le modèle et afficher ses caractéristiques :

```
> ----------------------------------------------------
> Gaussian finite mixture model fitted by EM algorithm
> ----------------------------------------------------
>
> Mclust EVI (diagonal, equal volume, varying shape) model with 2 components:
>
>  log-likelihood   n df      BIC      ICL
>         330.474 322  8 614.7516 497.0813
>
> Clustering table:
>   1   2
> 230  92
```

Ce morceau de code ajuste un modèle de mélange gaussien avec deux groupes (G=2) en utilisant les données spécifiées. Ensuite, la fonction summary() nous fournit un aperçu détaillé des résultats du modèle. Cela inclut des informations telles que les paramètres estimés pour chaque groupe, les proportions de chaque groupe, et d'autres mesures importantes liées à la qualité du modèle. L'analyse de cette sortie nous permet de mieux comprendre la structure de nos données et les caractéristiques de chaque groupe identifié.

Incorporons les classifications des clusters dans notre dataframe initial. Pour ensuite afficher les premières lignes du dataframe ajusté fournissant un aperçu rapide des modifications.

file_number	harmonious_hobbies	obsessive_hobbies	cluster
ZWT22697828	0.6944444	0.4722222	1
PWV43016656	0.7222222	0.5555556	1
QIQ45303215	0.9444444	0.2777778	2
KNE59187482	0.9444444	0.6944444	2
RON31606396	0.7777778	0.4444444	1
FOZ58315106	0.7500000	0.3888889	1

```
>
>   1   2
> 230  92
```

Analyse des groupes

Données basiques

Sexe

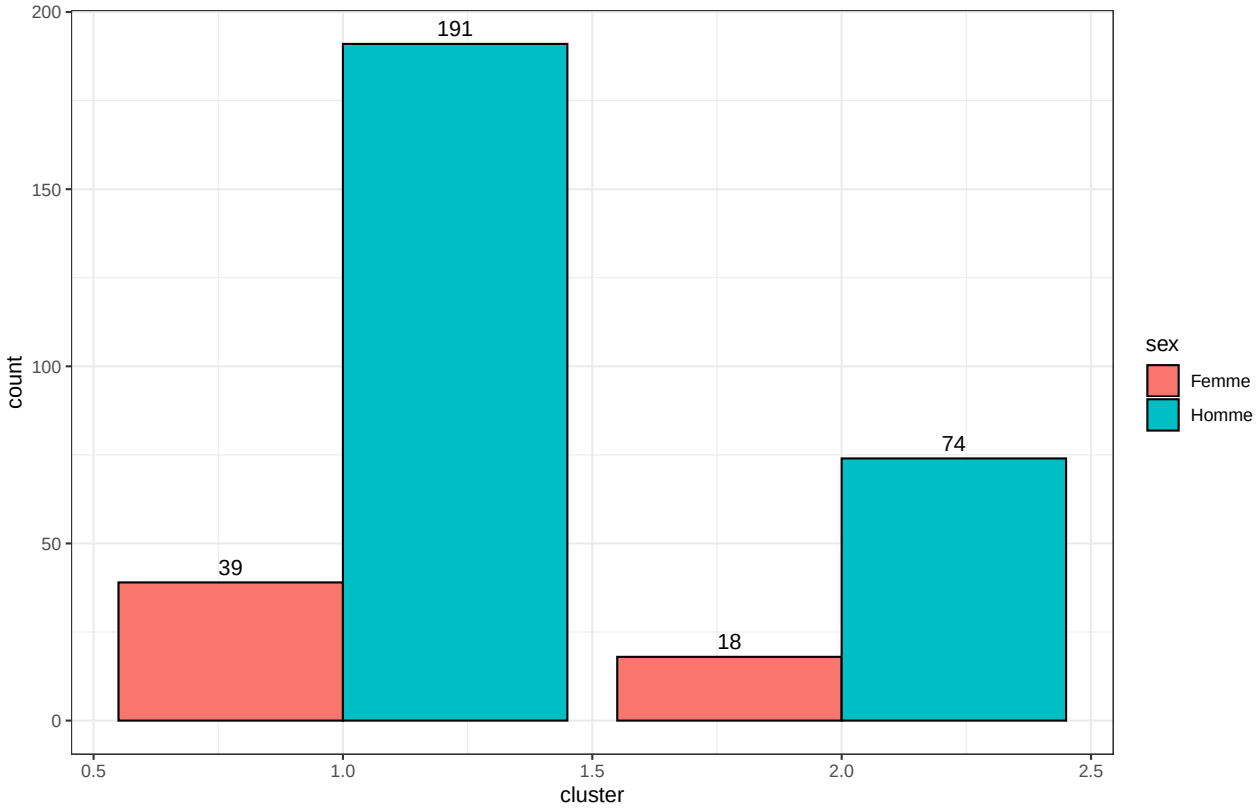


Figure 2: Répartition des sexes par groupe

```
>
>   Pearson's Chi-squared test with Yates' continuity correction
>
> data:  sex_table
> X-squared = 0.15402, df = 1, p-value = 0.6947
```

Age

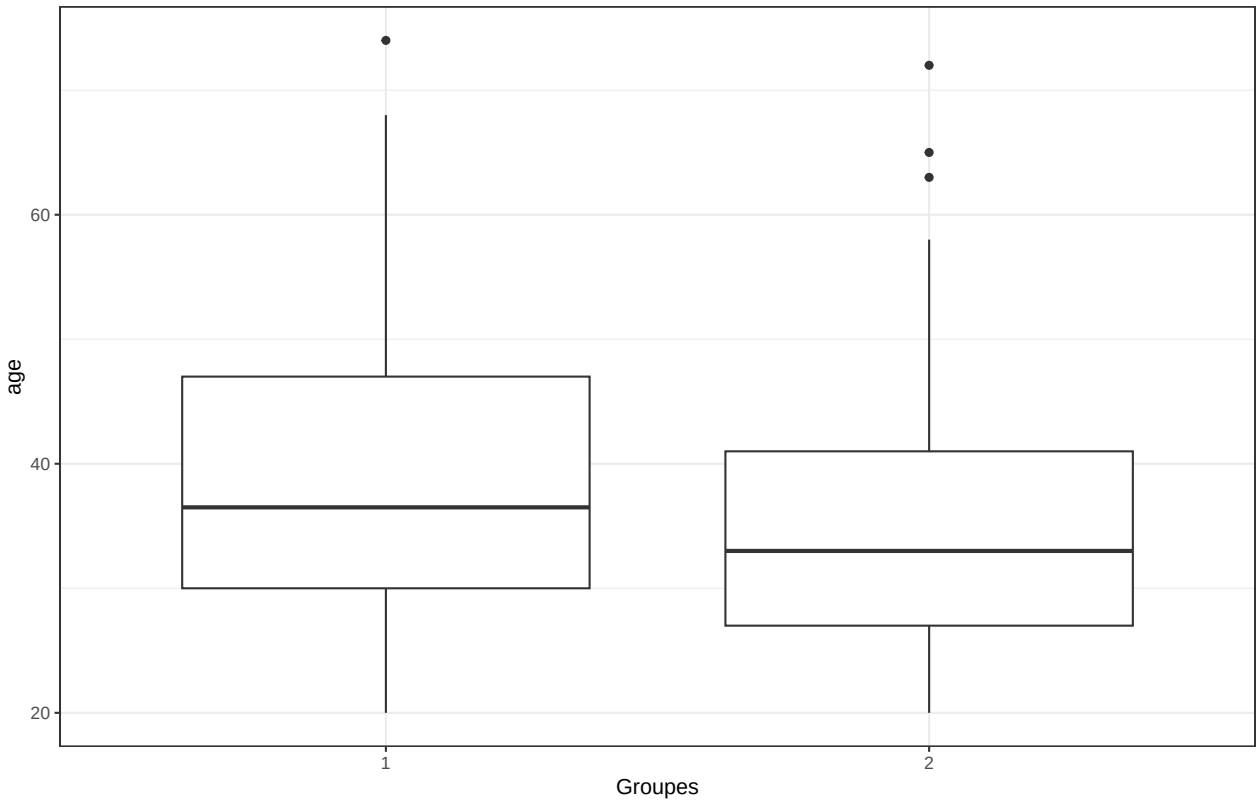


Figure 3: Distribution de l'âge par groupe

```
>
>   Welch Two Sample t-test
>
> data:  age by cluster
> t = 2.5097, df = 168.31, p-value = 0.01303
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  0.7334561 6.1404569
> sample estimates:
> mean in group 1 mean in group 2
>        38.73043        35.29348
```

Poids

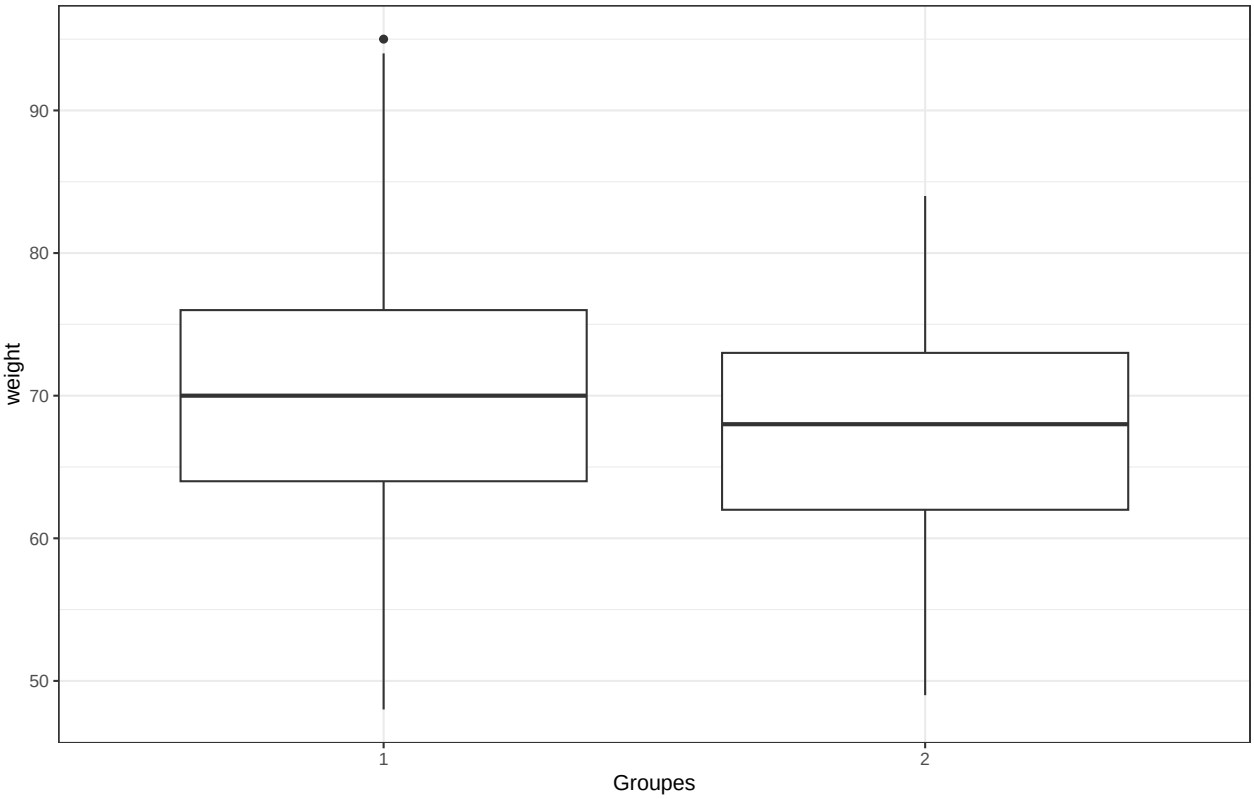


Figure 4: Distribution du poids par groupe

```
>
>   Welch Two Sample t-test
>
> data:  weight by cluster
> t = 2.8344, df = 183.86, p-value = 0.005104
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  0.9606359 5.3606685
> sample estimates:
> mean in group 1 mean in group 2
>        70.15739        66.99674
```

Coach

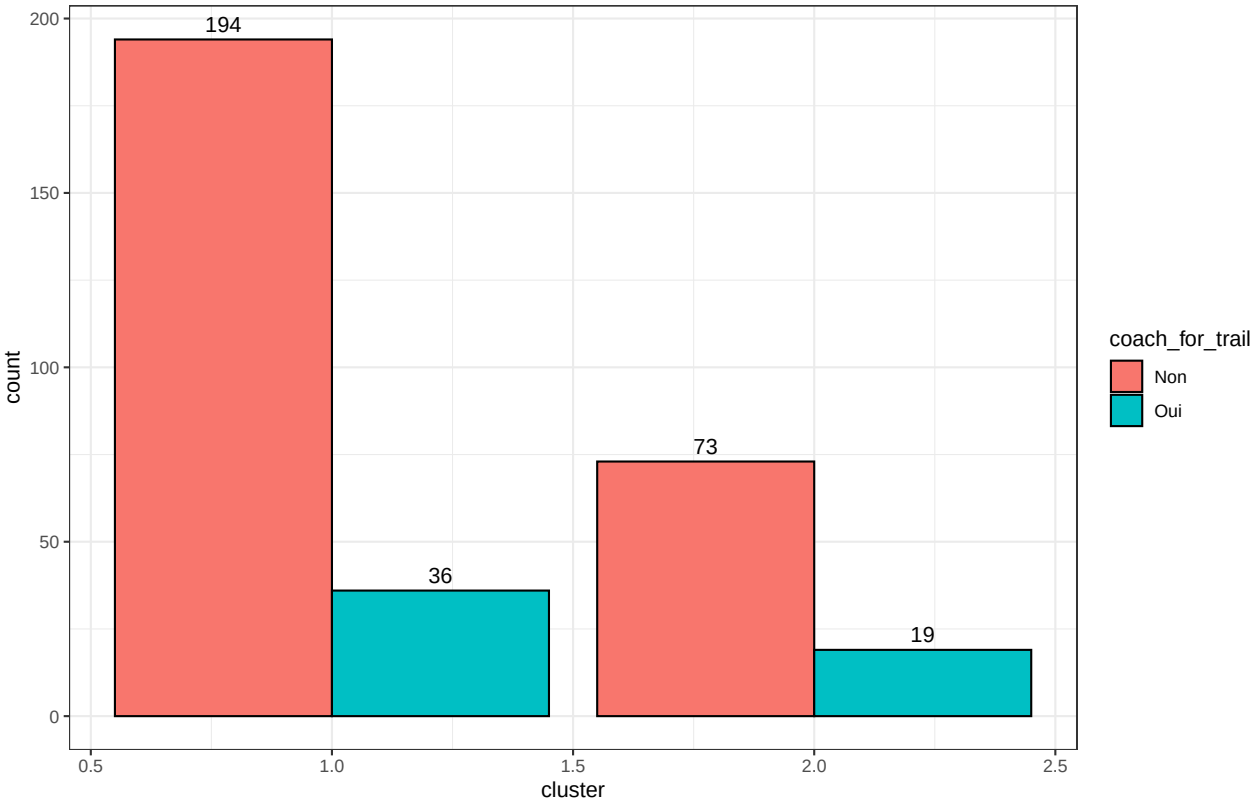


Figure 5: Répartition des coureurs ayant des coachs par groupe

```
>
>   Pearson's Chi-squared test with Yates' continuity correction
>
> data:  coach_table
> X-squared = 0.83378, df = 1, p-value = 0.3612
```

Blessure

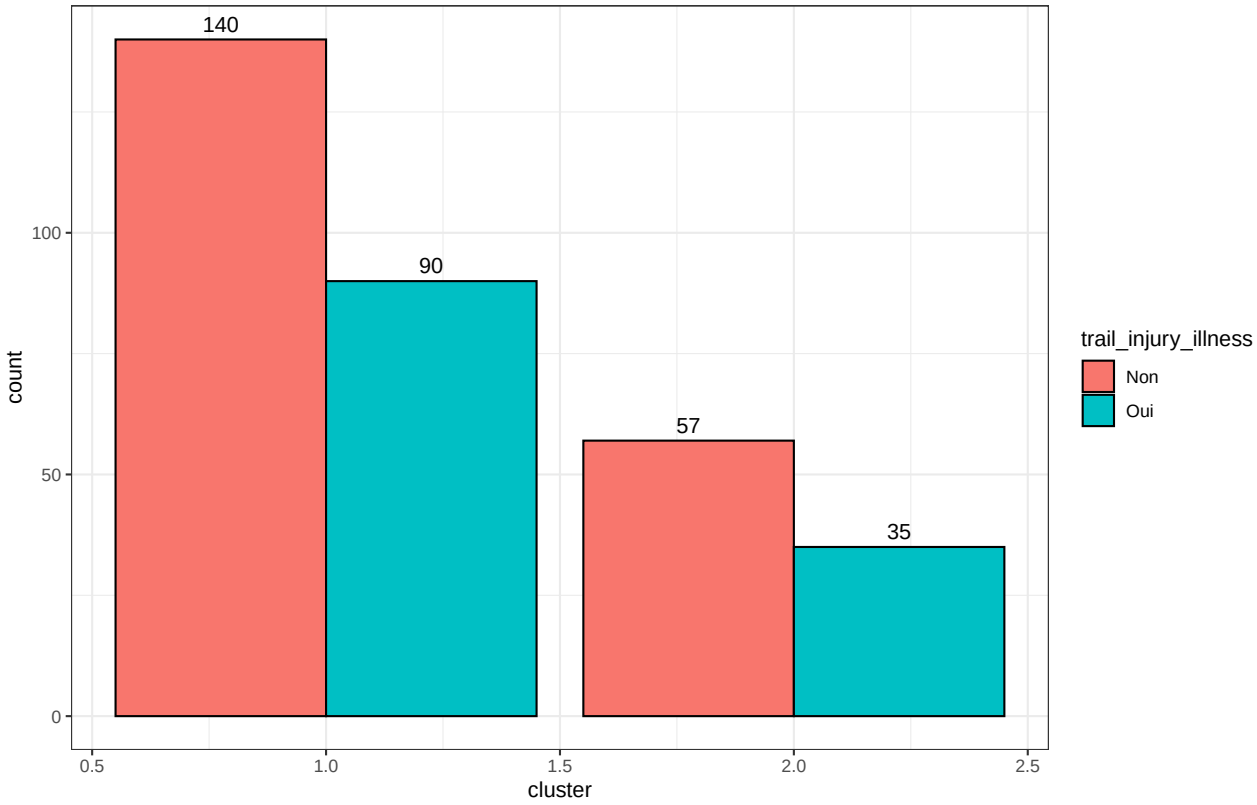


Figure 6: Répartition des coureurs ayant eu une blessure par groupe

```
>
>   Pearson's Chi-squared test with Yates' continuity correction
>
> data:  blessure_table
> X-squared = 0.0029421, df = 1, p-value = 0.9567
```

Côte ITRA

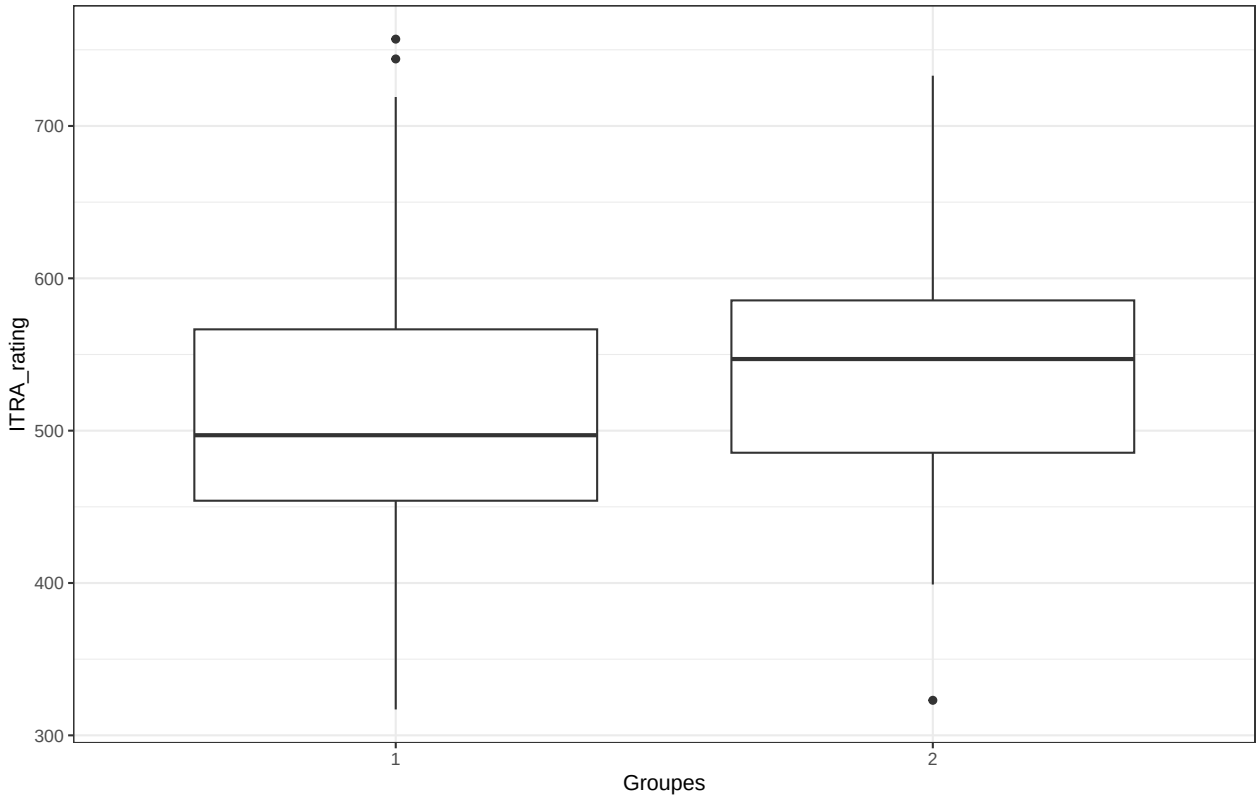


Figure 7: Répartition des côtes ITRA des coureurs par groupe

```
>
>   Welch Two Sample t-test
>
> data:  ITRA_rating by cluster
> t = -2.2372, df = 139.21, p-value = 0.02686
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -54.098116  -3.338593
> sample estimates:
> mean in group 1 mean in group 2
>        515.5802        544.2985
```

Courses

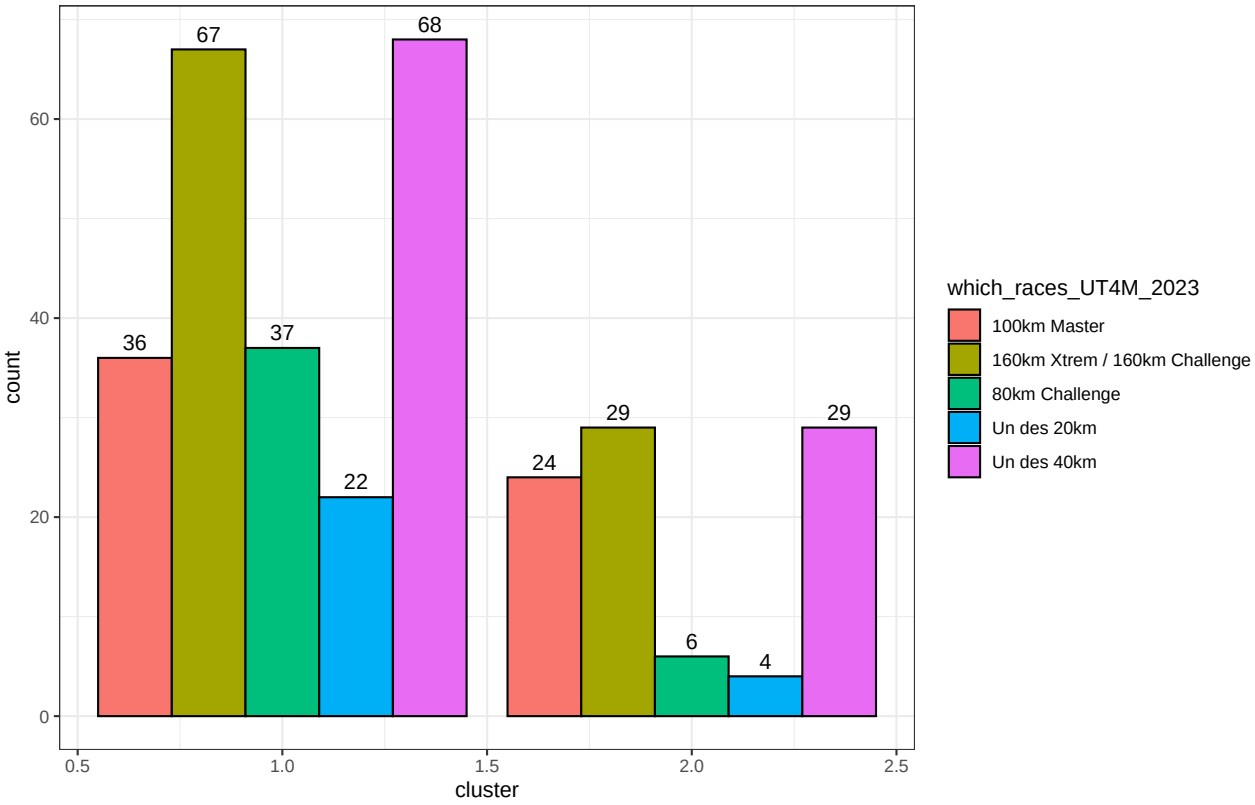


Figure 8: Répartition des courses par groupe

```
>
>   Pearson's Chi-squared test
>
> data:  course_table
> X-squared = 10.767, df = 4, p-value = 0.02931
```

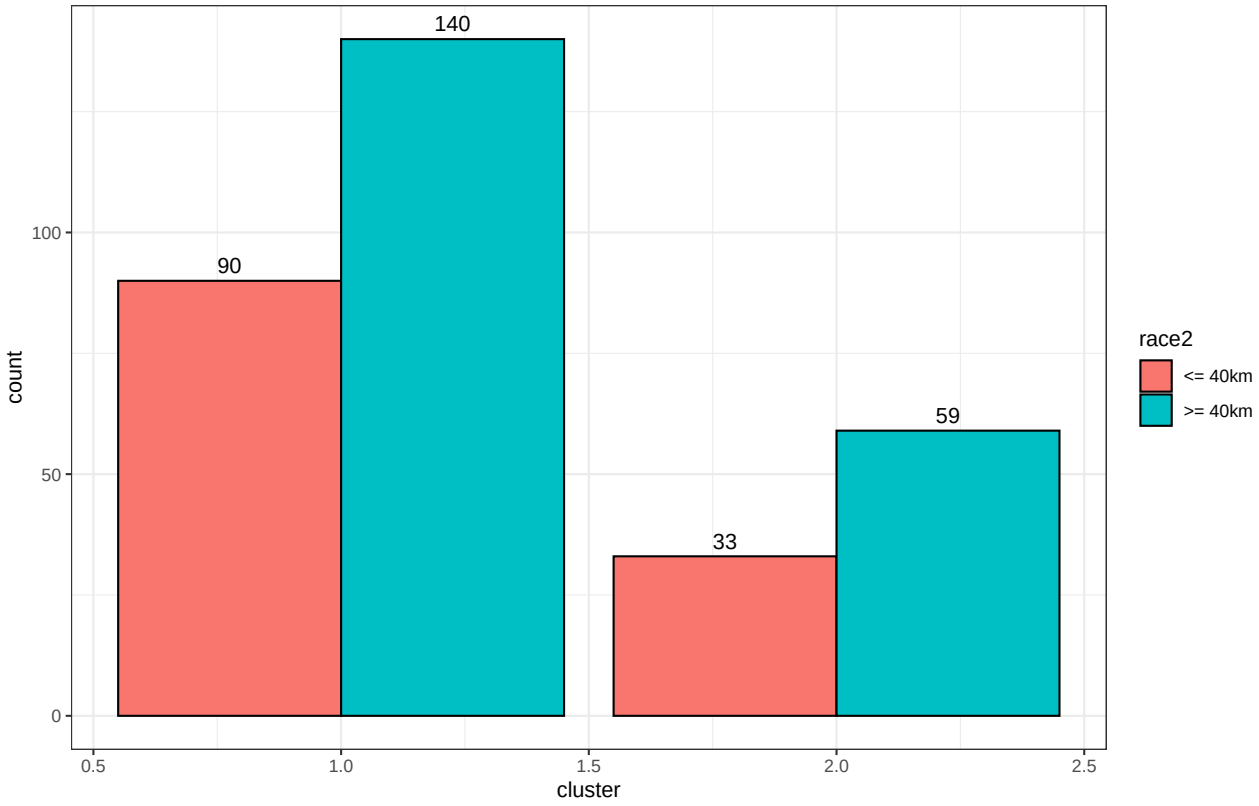


Figure 9: Répartition des courses regroupées

```
>
>   Pearson's Chi-squared test with Yates' continuity correction
>
> data:  course2_table
> X-squared = 0.17398, df = 1, p-value = 0.6766
```

Temps d'expérience en trail

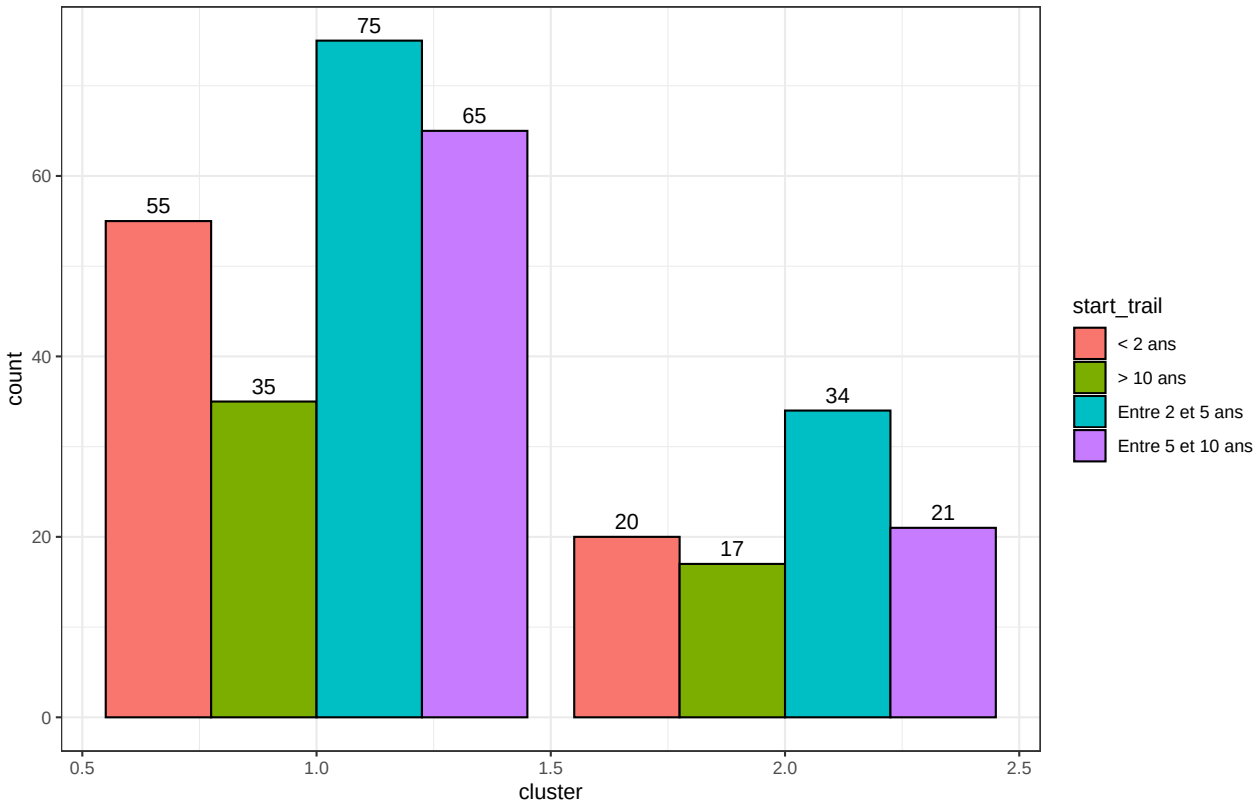


Figure 10: Répartition du temps d'expérience en trail

```
>
>   Pearson's Chi-squared test
>
> data:  experience_table
> X-squared = 1.6597, df = 3, p-value = 0.6459
```

Volume hebdomadaire consacré à la profession

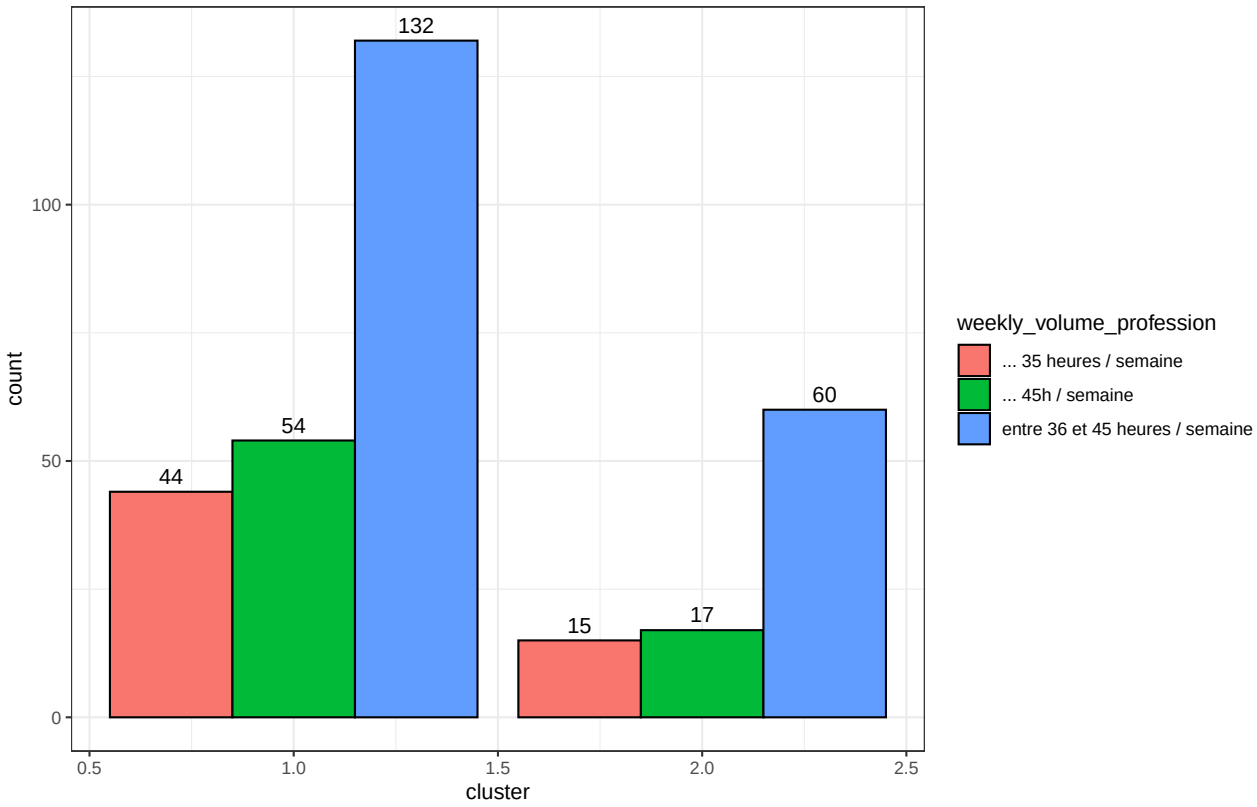


Figure 11: Répartition du volume hebdomadaire consacré à la profession chez les courreurs

```
>
>   Pearson's Chi-squared test
>
> data:  profession_table
> X-squared = 1.7065, df = 2, p-value = 0.426
```

Proportion des autres disciplines

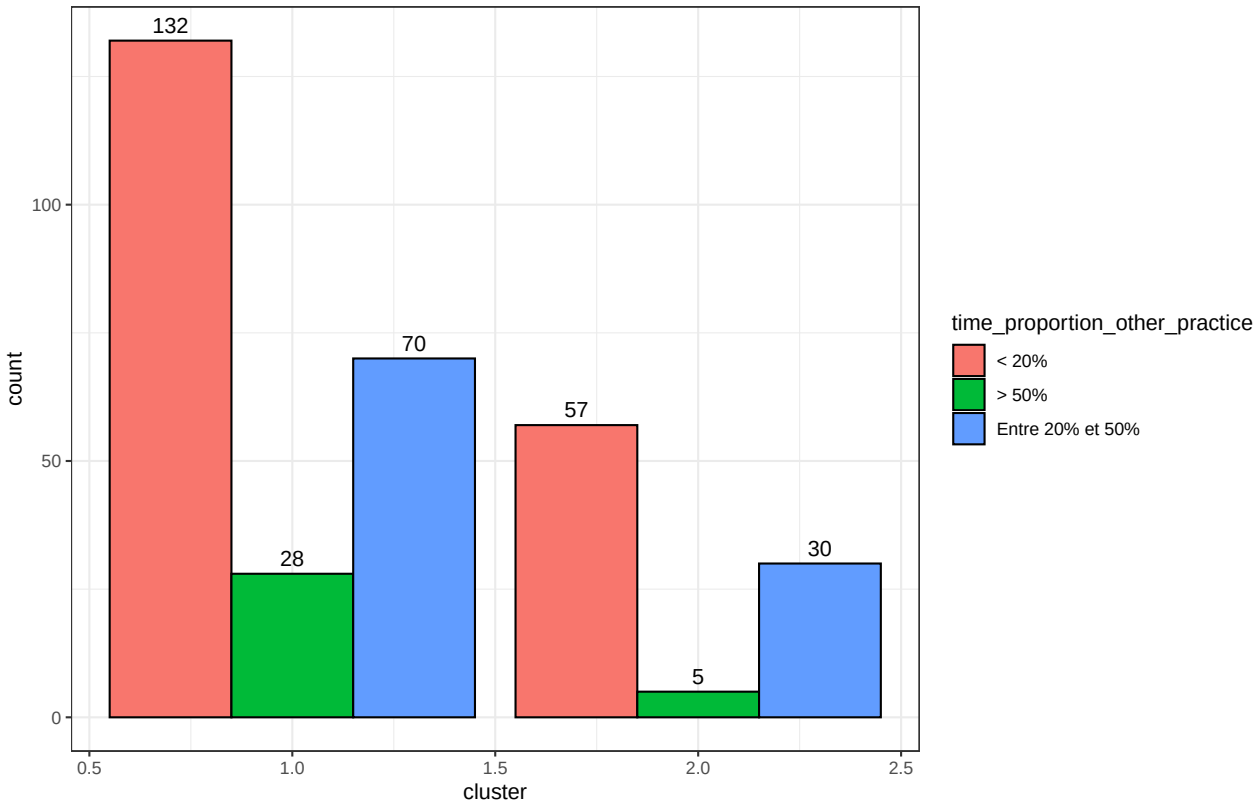


Figure 12: Proportion de ces autres disciplines dans le volume total d'entraînement

```
>
>   Pearson's Chi-squared test
>
> data:  other_table
> X-squared = 3.2455, df = 2, p-value = 0.1974
```

Données sur les scores

Epanouissement

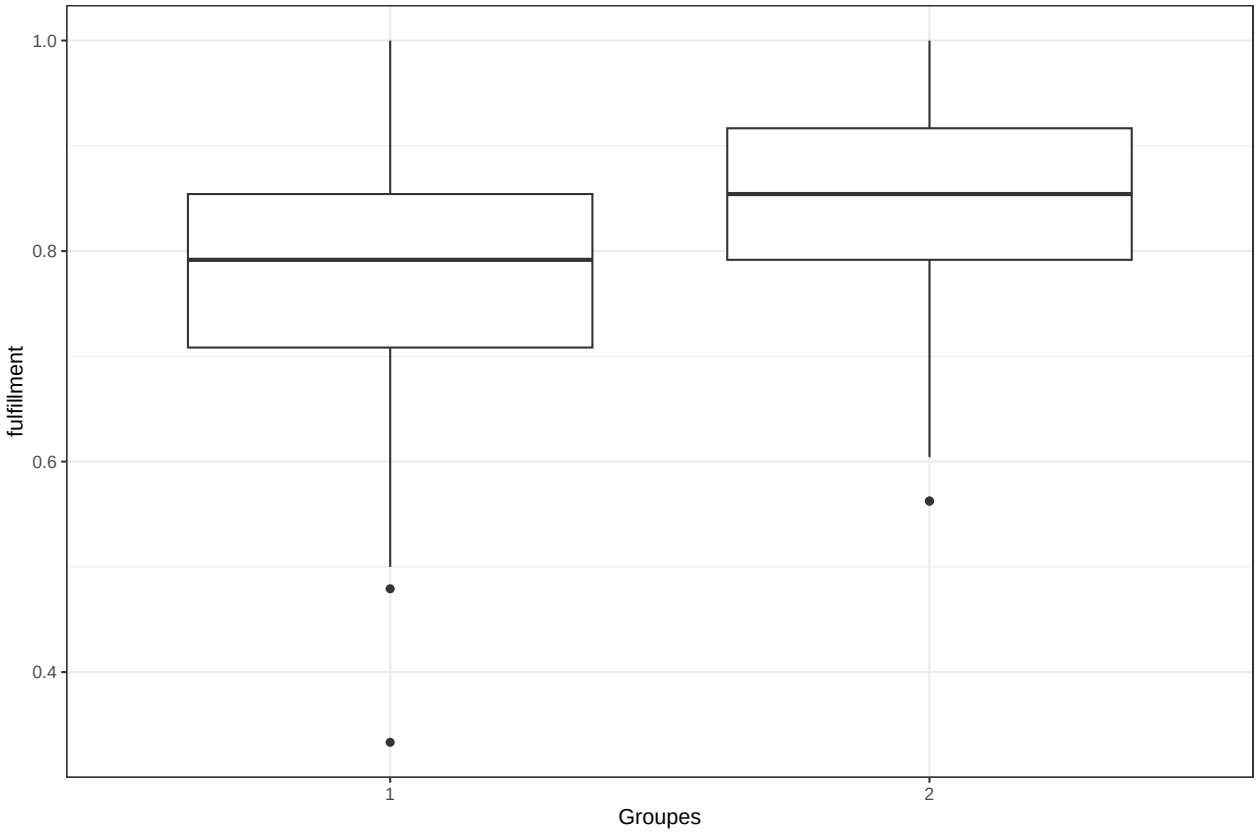


Figure 13: Distribution de l'épanouissement des coureurs par groupe

```
>
>   Welch Two Sample t-test
>
> data:  fulfillment by cluster
> t = -4.8183, df = 185.86, p-value = 2.997e-06
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.09128151 -0.03824747
> sample estimates:
> mean in group 1 mean in group 2
>       0.7803442       0.8451087
```

Conflits entre vie personnelle et pratique de l'activité sportive

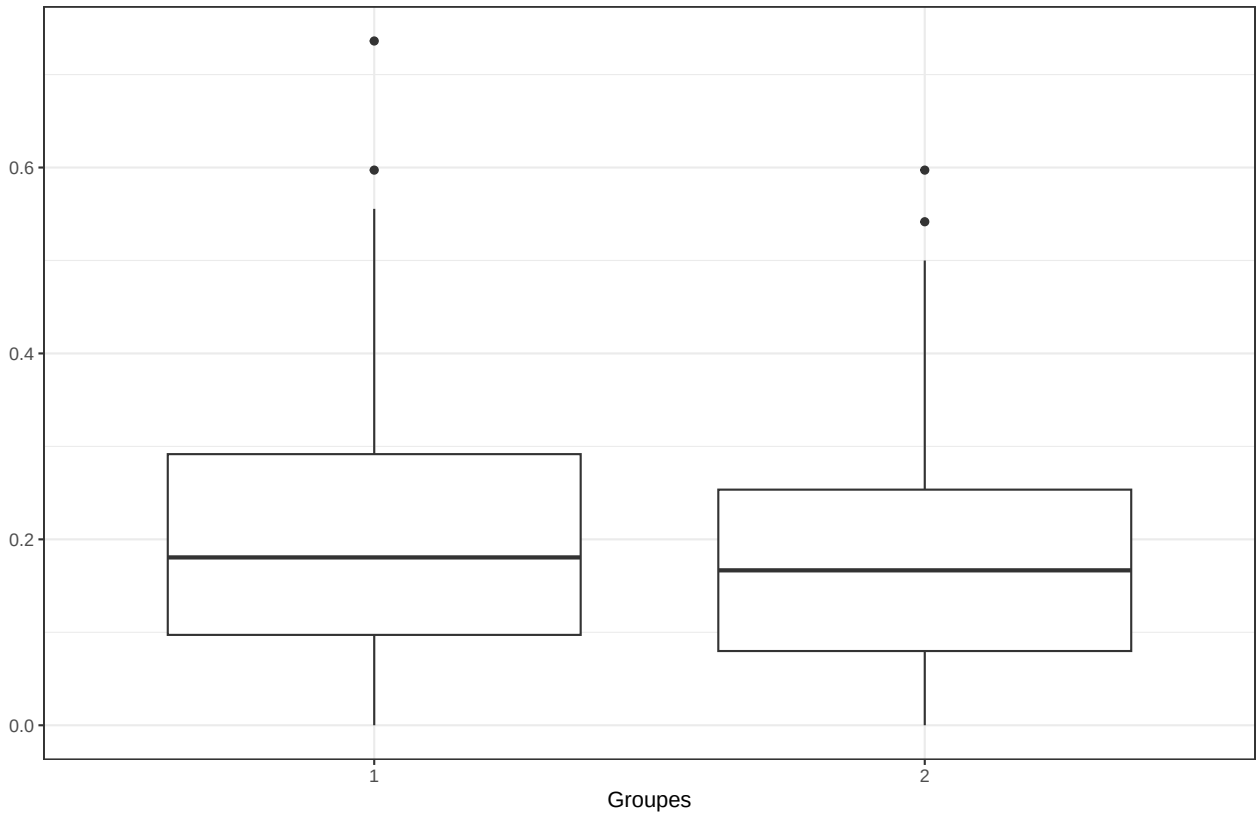
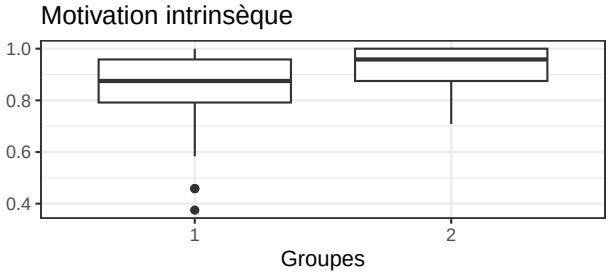
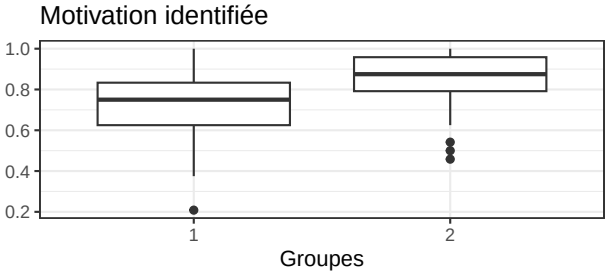


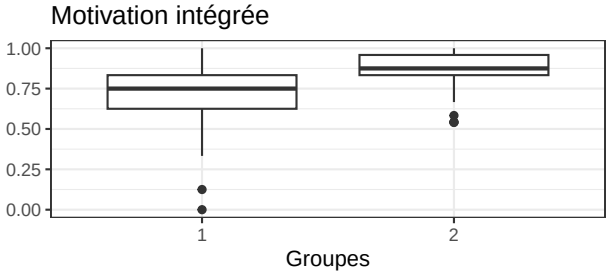
Figure 14: Distribution des conflits entre vie personnelle et pratique de l'activité sportive

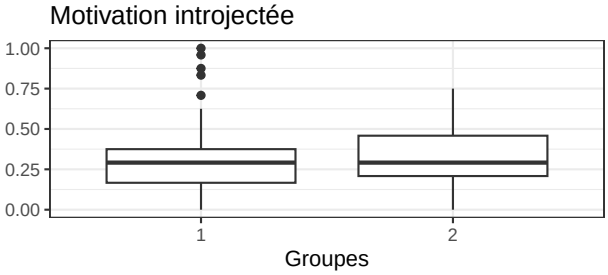
```
>
>   Welch Two Sample t-test
>
> data:  priority by cluster
> t = 1.2425, df = 179.56, p-value = 0.2157
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.01220012  0.05368562
> sample estimates:
> mean in group 1 mean in group 2
>       0.1987319       0.1779891
```

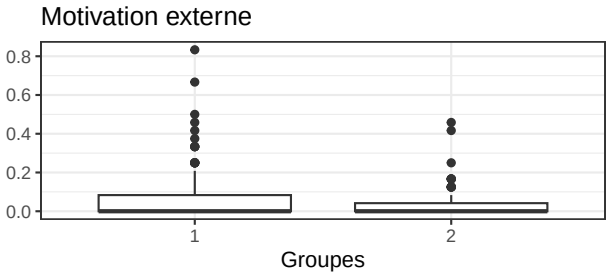
Motivations







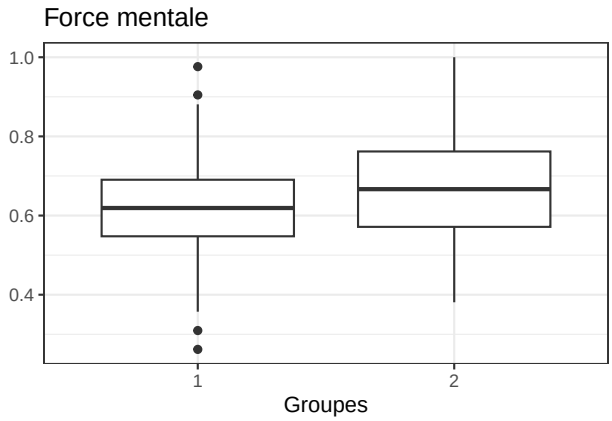


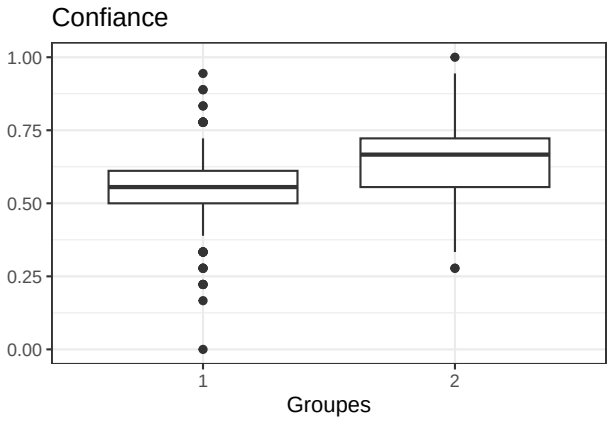


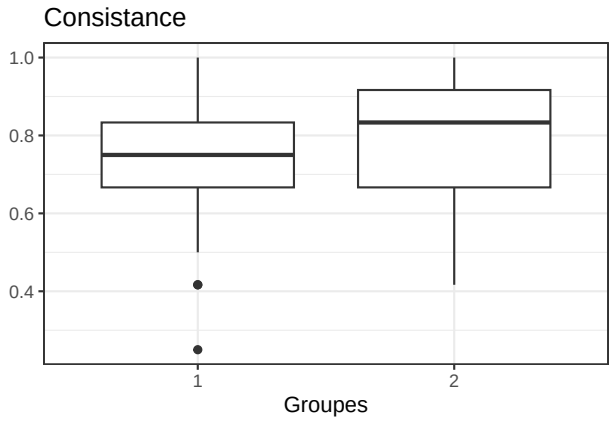

```
>
>   Welch Two Sample t-test
>
> data:  intrinsic_motivation by cluster
> t = -7.3333, df = 253.18, p-value = 3.03e-12
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.10329986 -0.05956246
> sample estimates:
> mean in group 1 mean in group 2
>       0.8574275       0.9388587
>
>   Welch Two Sample t-test
>
> data:  identified_motivation by cluster
> t = -7.6438, df = 188.73, p-value = 1.038e-12
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.15395350 -0.09079288
> sample estimates:
> mean in group 1 mean in group 2
>       0.7349638       0.8573370
>
```

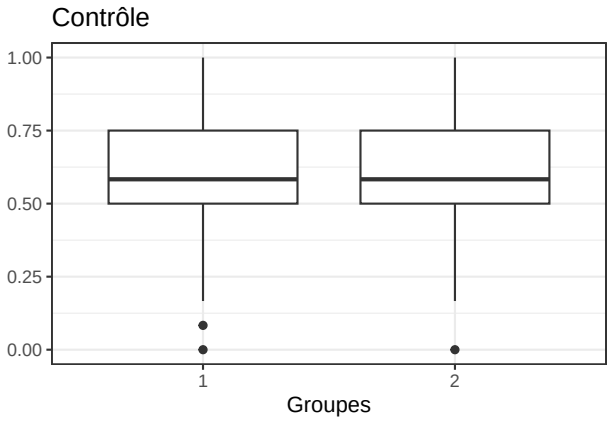
```
>   Welch Two Sample t-test
>
> data:  integrated_motivation by cluster
> t = -8.581, df = 232.9, p-value = 1.344e-15
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.1676222 -0.1050227
> sample estimates:
> mean in group 1 mean in group 2
>       0.7309783       0.8673007
>
>   Welch Two Sample t-test
>
> data:  introjected_motivation by cluster
> t = -1.3477, df = 153.4, p-value = 0.1797
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.07281449  0.01375652
> sample estimates:
> mean in group 1 mean in group 2
>       0.2938406       0.3233696
>
>   Welch Two Sample t-test
>
> data:  external_motivation by cluster
> t = 2.2981, df = 238.21, p-value = 0.02243
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  0.003646854 0.047440103
> sample estimates:
> mean in group 1 mean in group 2
>      0.06268116      0.03713768
```

Force mentale





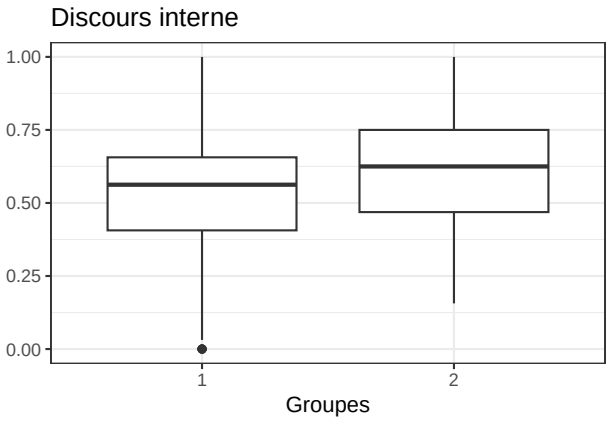


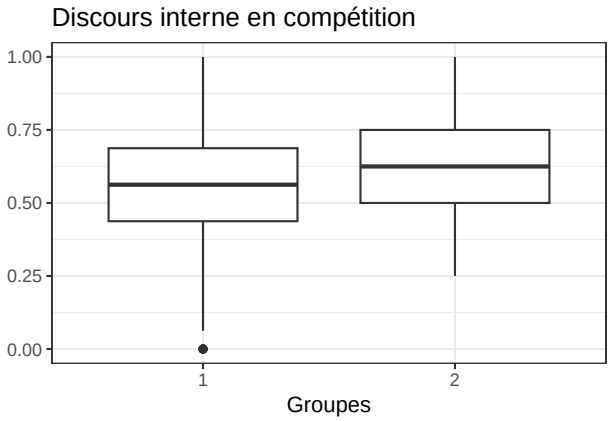


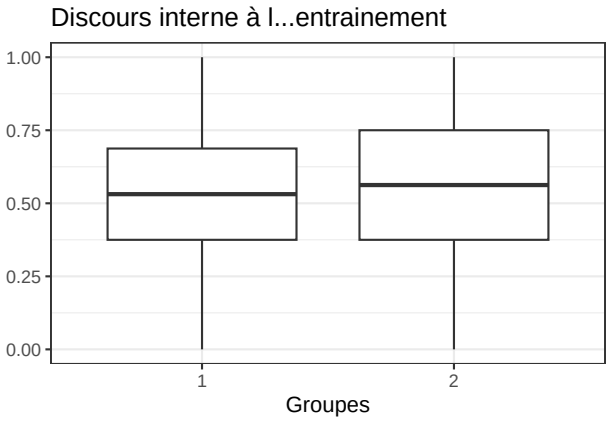

```
>
>   Welch Two Sample t-test
>
> data:  global_mental_force by cluster
> t = -3.2023, df = 152.37, p-value = 0.00166
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.07490503 -0.01774507
> sample estimates:
> mean in group 1 mean in group 2
>       0.6169772       0.6633023
>
>   Welch Two Sample t-test
>
> data:  confidence by cluster
> t = -3.9347, df = 156.4, p-value = 0.000125
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.10339905 -0.03428211
> sample estimates:
> mean in group 1 mean in group 2
>       0.5519324       0.6207729
>
```

```
>   Welch Two Sample t-test
>
> data:  consistency by cluster
> t = -4.8653, df = 152.06, p-value = 2.829e-06
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.11029541 -0.04658865
> sample estimates:
> mean in group 1 mean in group 2
>       0.7231884       0.8016304
>
>   Welch Two Sample t-test
>
> data:  control by cluster
> t = 0.7635, df = 158.51, p-value = 0.4463
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.03104633  0.07017677
> sample estimates:
> mean in group 1 mean in group 2
>       0.6083333       0.5887681
```

Discours interne



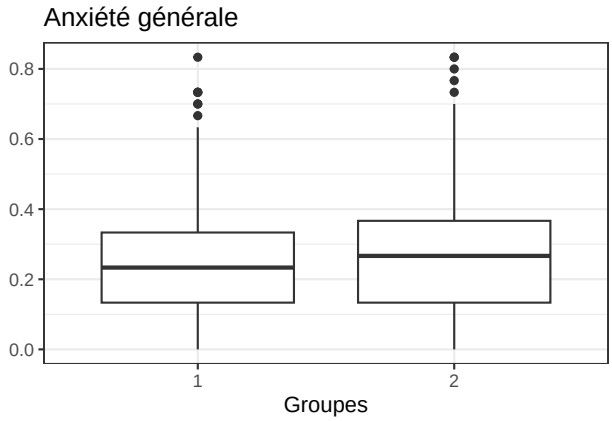


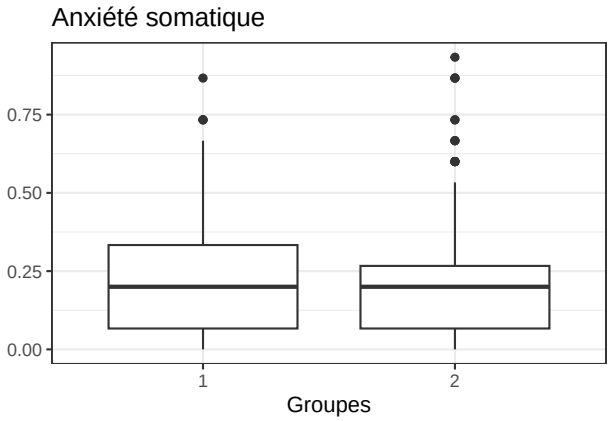


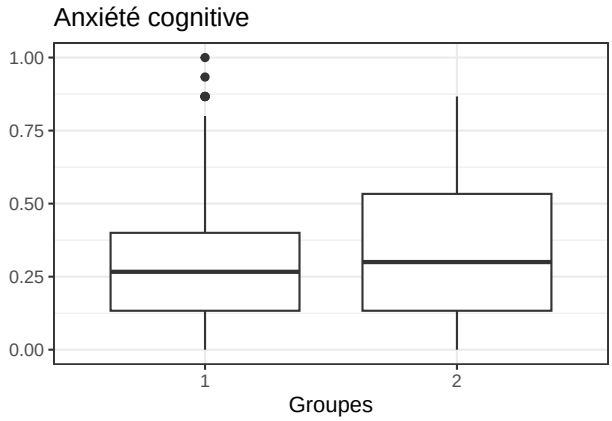

```
>
```

```
>   Welch Two Sample t-test
>
> data:  general_internal_speech by cluster
> t = -3.0944, df = 176.56, p-value = 0.002294
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.11949469 -0.02642922
> sample estimates:
> mean in group 1 mean in group 2
>       0.5323370       0.6052989
>
>   Welch Two Sample t-test
>
> data:  competition_internal_speech by cluster
> t = -3.5144, df = 180.03, p-value = 0.0005576
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.13302242 -0.03735802
> sample estimates:
> mean in group 1 mean in group 2
>       0.5527174       0.6379076
>
>   Welch Two Sample t-test
>
> data:  training_internal_speech by cluster
> t = -2.2095, df = 167.73, p-value = 0.02849
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.114999920 -0.006467471
> sample estimates:
> mean in group 1 mean in group 2
>       0.5119565       0.5726902
```

Anxieté



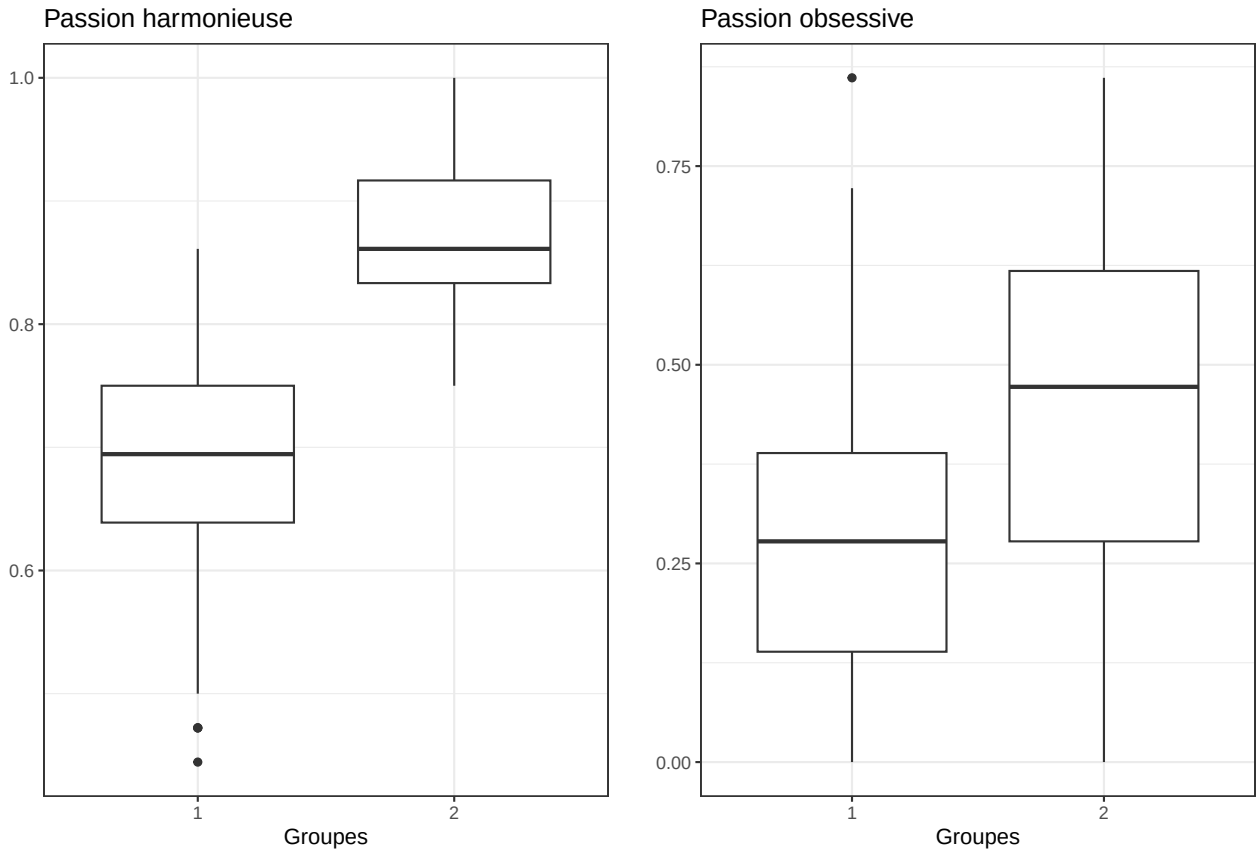





```
>
>   Welch Two Sample t-test
>
> data:  general_precompetitive_anxiety by cluster
> t = -1.0293, df = 143.89, p-value = 0.3051
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.07427844  0.02340888
> sample estimates:
> mean in group 1 mean in group 2
>       0.2597101       0.2851449
>
>   Welch Two Sample t-test
>
> data:  somatic_anxiety by cluster
> t = -0.91101, df = 132.04, p-value = 0.364
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.07951290  0.02936797
> sample estimates:
> mean in group 1 mean in group 2
>       0.2162319       0.2413043
>
```

```
>   Welch Two Sample t-test
>
> data:  cognitive_anxiety by cluster
> t = -0.86207, df = 152.82, p-value = 0.39
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.08491632  0.03332212
> sample estimates:
> mean in group 1 mean in group 2
>       0.3031884       0.3289855
```

Passions




```
>
>   Welch Two Sample t-test
>
> data:  harmonious_hobbies by cluster
> t = -21.913, df = 230.78, p-value < 2.2e-16
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.2007389 -0.1676185
> sample estimates:
> mean in group 1 mean in group 2
>       0.6914251       0.8756039
>
```

```
>   Welch Two Sample t-test
>
> data:  obsessive_hobbies by cluster
> t = -6.5835, df = 132.75, p-value = 9.774e-10
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.2129721 -0.1145641
> sample estimates:
> mean in group 1 mean in group 2
>       0.2752415       0.4390097
```

Addiction

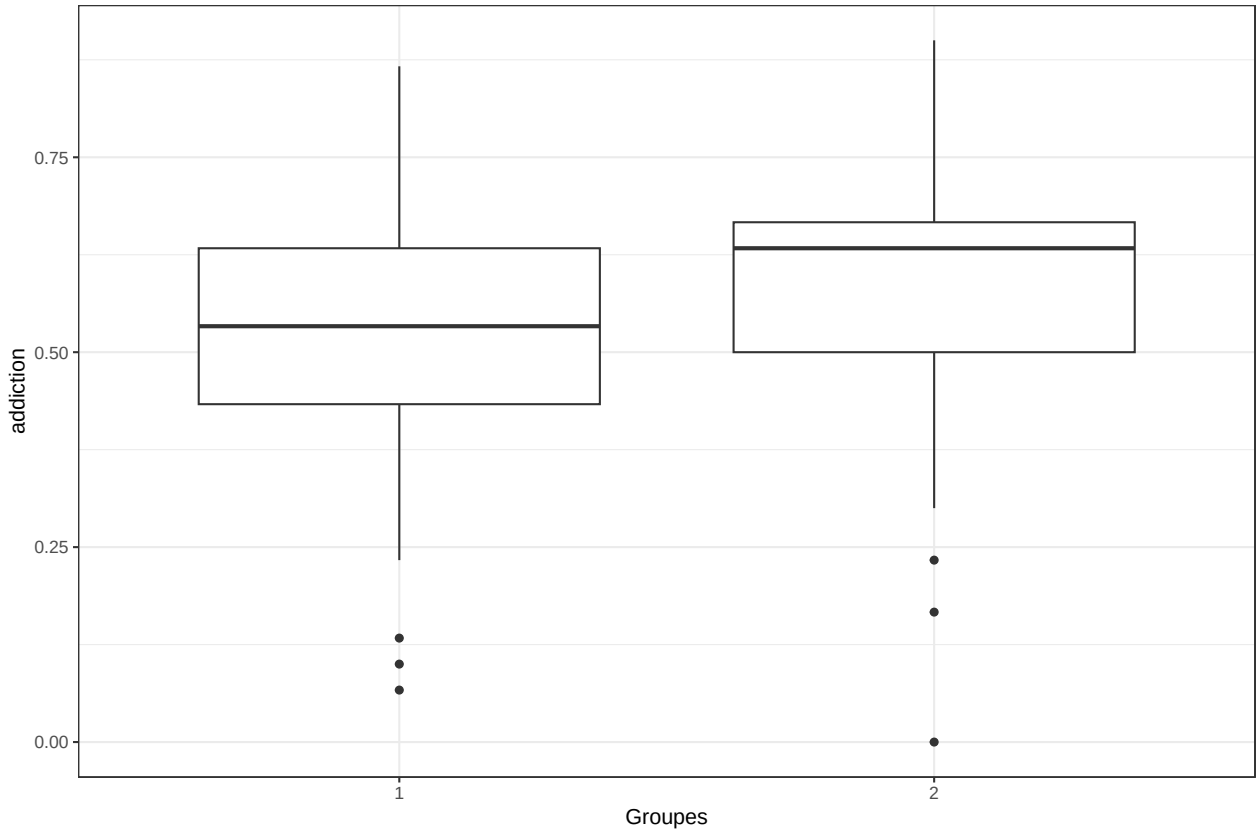


Figure 15: Distribution de l'addiction par groupe

```
>
>   Welch Two Sample t-test
>
> data:  addiction by cluster
> t = -3.1344, df = 164.9, p-value = 0.002039
> alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
> 95 percent confidence interval:
>  -0.09433318 -0.02141815
> sample estimates:
```

```
> mean in group 1 mean in group 2
>       0.5359649       0.5938406
```

Conclusion