# Rapport EI Exposome

Groupe 5

2024-06-10

## Introduction

La phrase du philosophe José Ortega y Gasset : « Je suis moi et mes circonstances », n'est pas loin de ce que la science a prouvé par la suite : nous sommes le résultat de notre génome et de nos interactions avec notre environnement. C'est pourquoi le développement de l'être humain pendant l'enfance peut être fortement altéré par de nombreux facteurs, depuis l'alimentation, le tabagisme et la pollution pendant la période de grosses jusqu'à l'activité physique de l'enfant.

Le projet HELIX, Human Early-life Exposome, vise à prendre en compte le maximum de ces facteurs, appelés exposomes, qui peuvent conditionner la croissance des enfants. En particulier, l'objet d'étude de cet article est le développement neurologique des enfants jusqu'à l'âge de 11 ans.

## Description des données

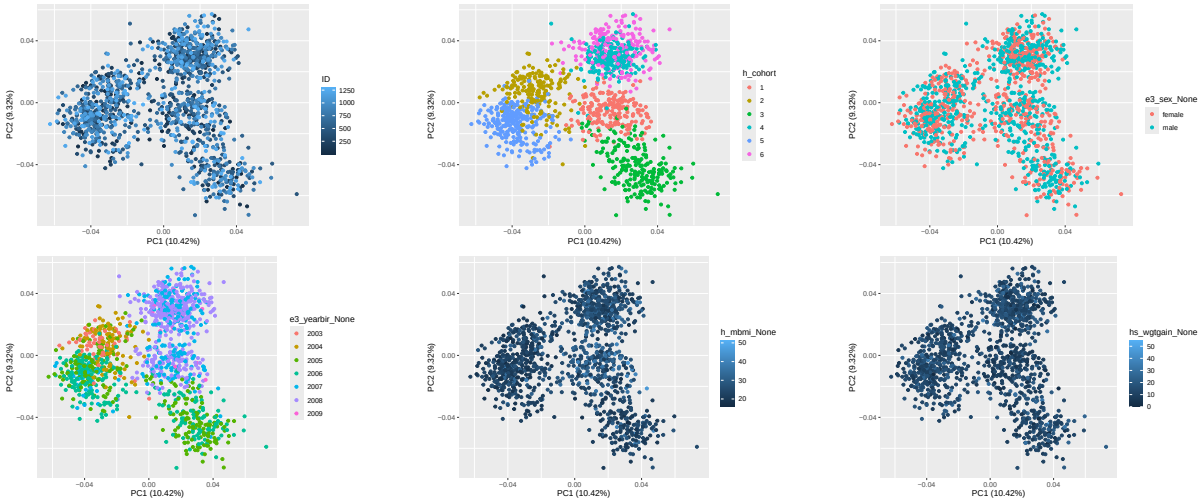
The projet Athlete nous a fourni avec les données de 1301 personnes collectés dans 6 pays différentes: France, Espagne, Grèce, Royaume-Uni, Norvège et Lituanie.

Les donnés ont été collectées de trois manières différentes, par questionnaires, pour recolecter les habitudes de vie et facteurs socio-écologiques; capteurs personnels, pour recolecter composition de l'air, et modélisation des exposomes externes en utilisant Système d'information géographique (SIG) et Géolocalisation des domiciles des participants pendant la grossesse et l'enfance et des écoles des enfants.

## Détection des covariantes principales influant le modèle

Dans un premier pas, on s'est concentré sur les valeurs numériques de l'exposome.

Ceci a été fait pour pouvoir effectuer une analyse de composantes principales sur les données obtenues et identifier les "covariates" influant sur les différentes expositions afin de déterminer lesquelles sont à prendre en compte les analyses suivantes.

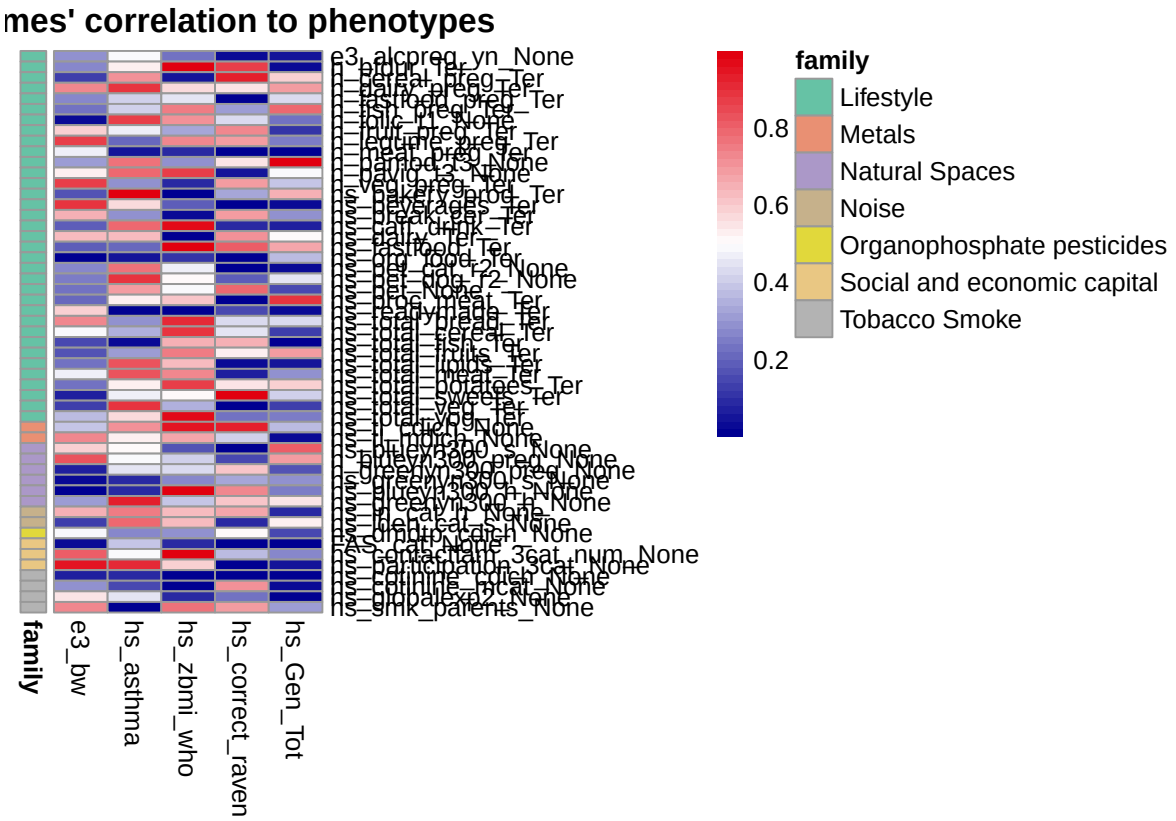



L'analyse en composantes principales nous permet d'identifier une forte corrélation avec la variable h_cohort, qui représente le pays dans lequel les données ont été collectées et à e3_yearbir_None qui représente l'année de naissance des enfants. Dans la suite du projet, nous ne prendrons en compte que la variable h_cohort. En effet, il semblerait que les cohortes ne soient pas créées en même temps car l'année de naissance est visiblement liée à la cohorte (chaque cohorte correspond à certaines années de naissance).

Les familles d'exposition qui vont être considérés sont : Organochlorines, Air pollution, Phthalates, Lifestyle and PFAS.

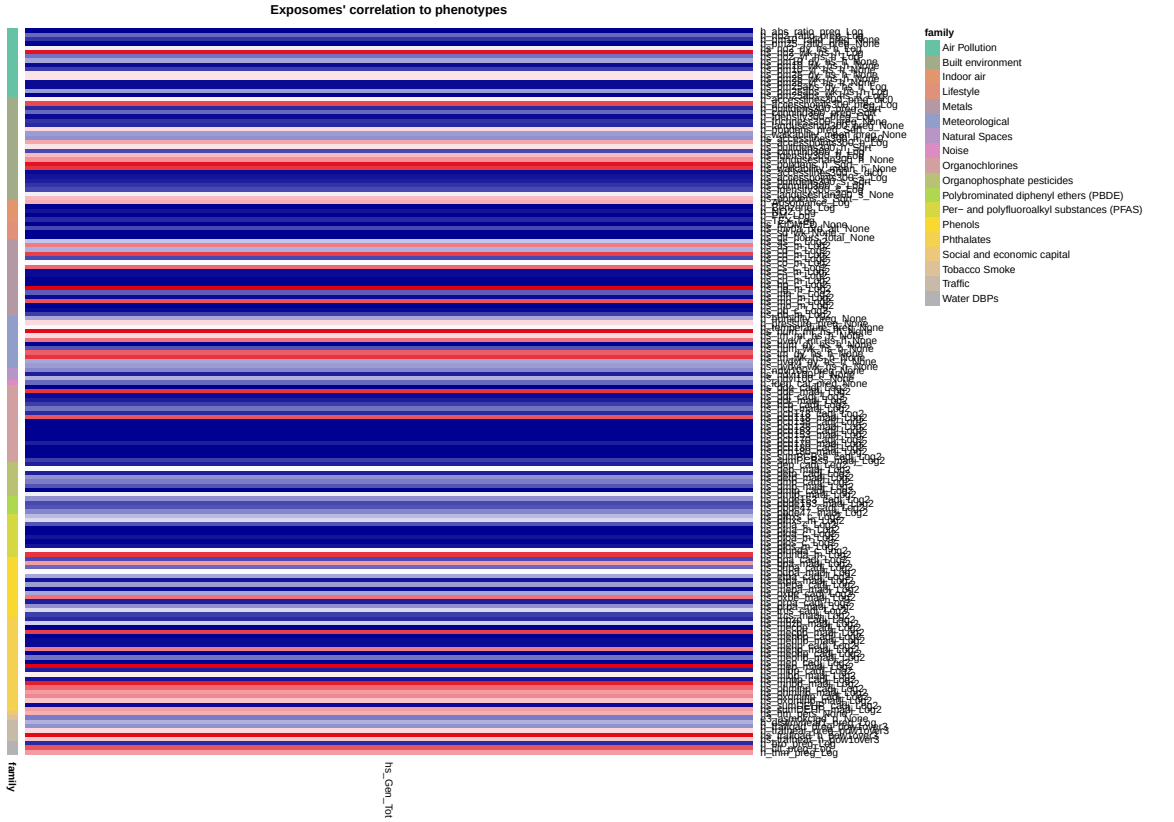
# Analyse des Exposomes de type Facteur

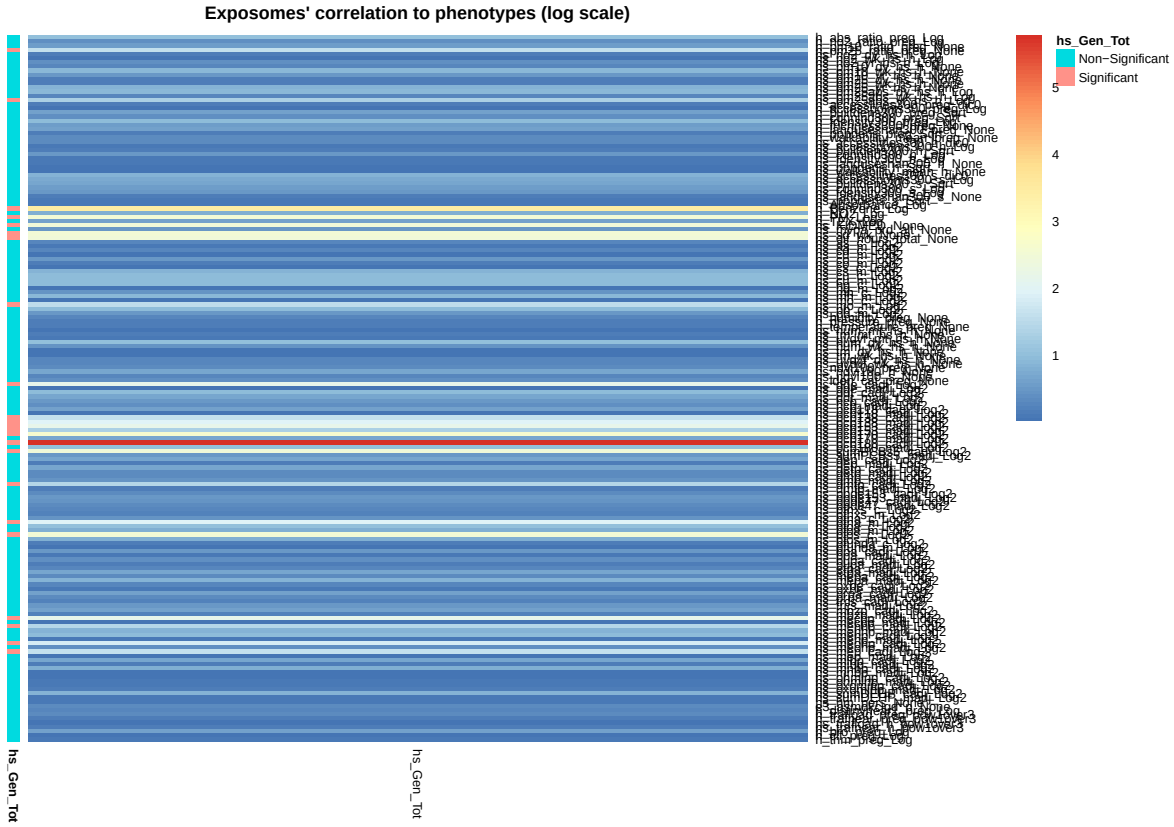
On prend maintenant les exposomes qualitatifs pour analyser leur impact sur le phénotype choisi, hs_GenTot. Pour ce faire, nous transformons les facteurs en données numériques pour les traiter en obtenant leur p-value dans le modèle linéaire généralisé. On ajuste les p-values calculés et on les représentent sur une heatmap. Puis, pour pouvoir identifier si elles sont pertinentes ou non pour notre phénotype, on utilise le logarithme des valeurs. On constate qu'il n'y a pas de facteurs significatifs dans le comportement neuronal. C'est pour cette raison que le reste de l'analyse ne portera que sur les données numériques.




Nous nous intéressons ensuite aux exposomes numériques afin de déterminer lesquels sont corrélés avec l'exposome que nous avons choisi, et ont ainsi une influence non négligeable. Nous utilisons à nouveau un modèle de régression linéaire généralisé pour cela et nous étudions les p-values, que nous aurons ajustées au préalable à l'aide de la méthode "False Discovery Rate" (FDR).

# Matrice de “corrélations” (p_values) entre les expositions numériques et les différents phénotypes



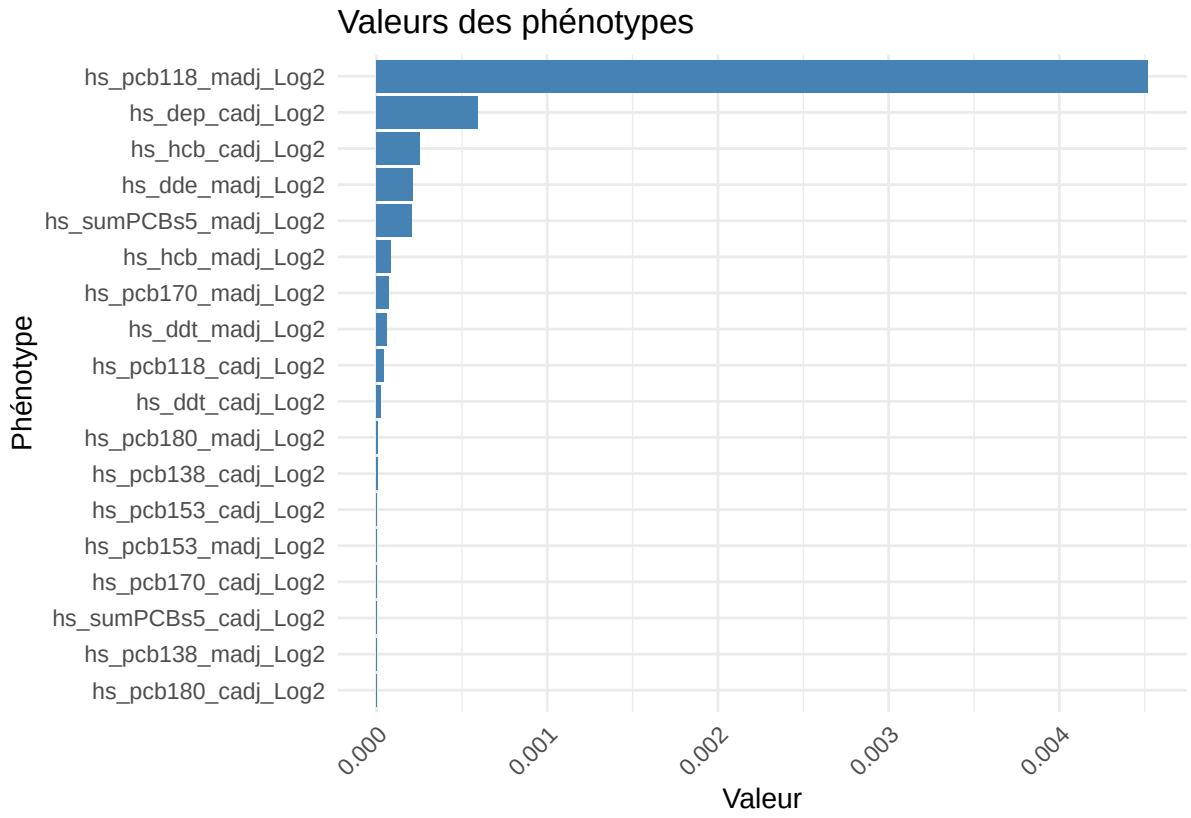

```
##                                pvalue Adjusted pvalue
```

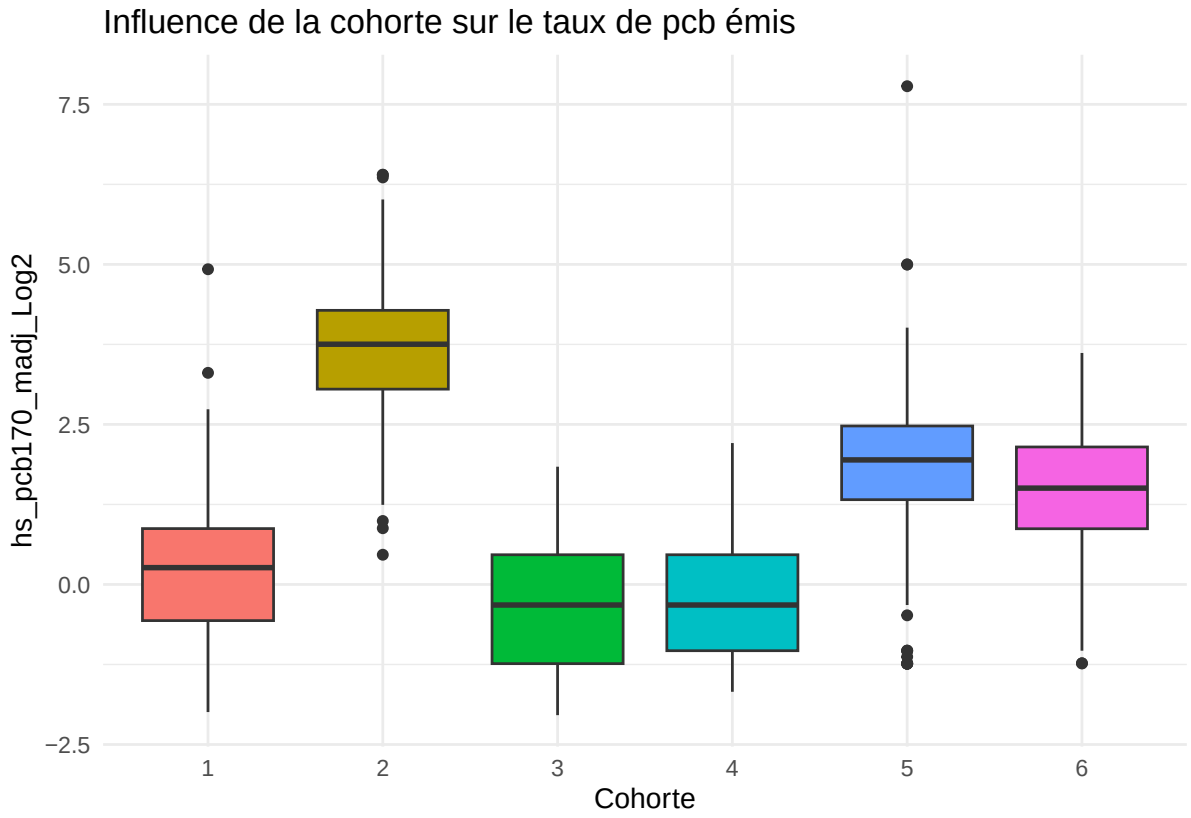
```
## h_pm25_ratio_preg_None  2.603395e-03    2.588081e-02
## hs_pm25abs_yr_hs_h_Log  6.311886e-03    4.848676e-02
## h_Benzene_Log           4.202194e-06    3.550854e-04
## h_PM_Log                1.052631e-04    2.964912e-03
## hs_KIDMED_None          1.750946e-04    3.287887e-03
## hs_sd_wk_None           9.977089e-05    2.964912e-03
## hs_dif_hours_total_None 1.595049e-04    3.287887e-03
## hs_mo_m_Log2            3.143187e-03    2.795782e-02
## hs_dde_cadj_Log2        4.924506e-04    6.991549e-03
## hs_pcb138_cadj_Log2     2.451419e-03    2.588081e-02
## hs_pcb138_madj_Log2     1.006068e-03    1.214468e-02
## hs_pcb153_cadj_Log2     4.964413e-04    6.991549e-03
## hs_pcb153_madj_Log2     6.673700e-03    4.903718e-02
## hs_pcb170_cadj_Log2     4.224408e-05    2.379750e-03
## hs_pcb180_cadj_Log2     1.009890e-08    1.706714e-06
## hs_sumPCBs5_cadj_Log2   1.428027e-04    3.287887e-03
## hs_dmtp_cadj_Log2       4.183211e-03    3.534814e-02
## hs_pfna_m_Log2          9.059914e-04    1.177789e-02
## hs_pfos_c_Log2          9.641230e-05    2.964912e-03
## hs_mecpp_cadj_Log2      4.678396e-04    6.991549e-03
## hs_mehhp_cadj_Log2      4.938423e-03    3.974254e-02
## hs_meohp_cadj_Log2      1.319179e-03    1.486275e-02
## hs_mep_cadj_Log2        2.785889e-03    2.615640e-02
##                                                                  Family
## h_pm25_ratio_preg_None                                      Air Pollution
## hs_pm25abs_yr_hs_h_Log                                      Air Pollution
## h_Benzene_Log                                                  Indoor air
## h_PM_Log                                                       Indoor air
## hs_KIDMED_None                                                  Lifestyle
## hs_sd_wk_None                                                   Lifestyle
## hs_dif_hours_total_None                                         Lifestyle
## hs_mo_m_Log2                                                       Metals
## hs_dde_cadj_Log2                                          Organochlorines
## hs_pcb138_cadj_Log2                                       Organochlorines
## hs_pcb138_madj_Log2                                       Organochlorines
## hs_pcb153_cadj_Log2                                       Organochlorines
## hs_pcb153_madj_Log2                                       Organochlorines
## hs_pcb170_cadj_Log2                                       Organochlorines
## hs_pcb180_cadj_Log2                                       Organochlorines
## hs_sumPCBs5_cadj_Log2                                     Organochlorines
## hs_dmtp_cadj_Log2                            Organophosphate pesticides
## hs_pfna_m_Log2          Per- and polyfluoroalkyl substances (PFAS)
## hs_pfos_c_Log2          Per- and polyfluoroalkyl substances (PFAS)
## hs_mecpp_cadj_Log2                                          Phthalates
## hs_mehhp_cadj_Log2                                          Phthalates
## hs_meohp_cadj_Log2                                          Phthalates
## hs_mep_cadj_Log2                                            Phthalates
```

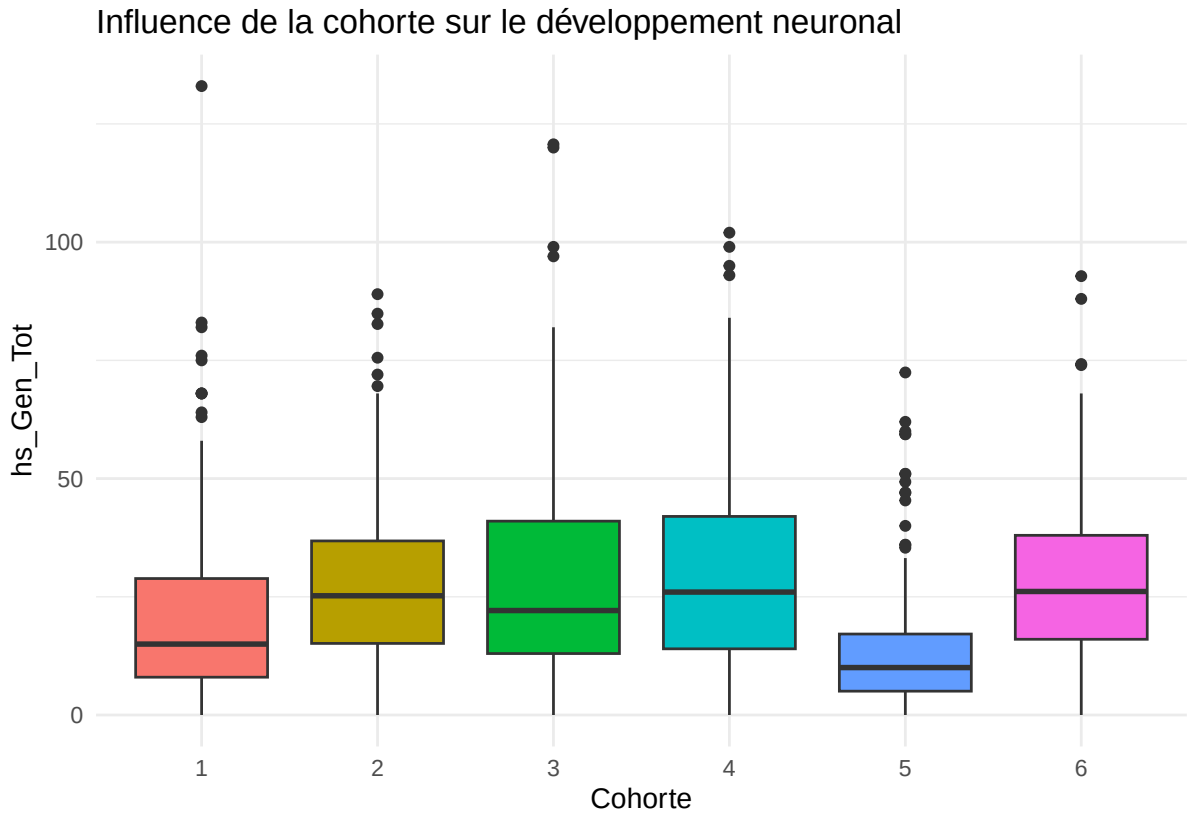
Nous modifions les p_values pour les avoir en échelle logarithmique afin de pouvoir identifier visuellement quels exposomes sont significatifs, avant d'extraire la matrice des exposomes significatifs ainsi que leur p-value.

## Analyse univarié: Effet des expositions sur le phénotype d'intérêt

Nous nous concentrons maintenant sur l'effet de la famille "Organochlorines" sur le phénotype choisi.


Valeurs des phénotypes
hs_pcb118_madj_Log2
hs_dep_cadj_Log2
hs_hcb_cadj_Log2
hs_dde_madj_Log2
hs_sumPCBs5_madj_Log2
hs_hcb_madj_Log2
hs_pcb170_madj_Log2
hs_ddt_madj_Log2
hs_pcb118_cadj_Log2
hs_ddt_cadj_Log2
hs_pcb180_madj_Log2
hs_pcb138_cadj_Log2
hs_pcb153_cadj_Log2
hs_pcb153_madj_Log2
hs_pcb170_cadj_Log2
hs_sumPCBs5_cadj_Log2
hs_pcb138_madj_Log2
hs_pcb180_cadj_Log2
Phénotype
0.000
0.001
0.002
0.003
0.004
Valeur


Influence de la cohorte sur le taux de pcb émis
7.5
5.0
2.5
0.0
−2.5
hs_pcb170_madj_Log2
1
2
3
4
5
6
Cohorte




Pour mieux étudier les influences des uniques expositions dans la famille organochlorines sur le phénotype de développement neuronal, on a identifié les exposomes les plus significatifs selon leur p-value dans cette famille à partir d'une régression linéaire généralisée. L'influence significative du taux de polychlorinated biphenyl PCB (138 et 170 ) a été remarqué, ceux-ci étant des composés chimiques utilisés dans diverses applications industrielles et commerciales et dont les études ont montré la nocivité sur la santé de l'être humain. On se pose alors la question de l'influence de la cohorte sur le taux du pcb consommé par la mère et l'enfant, vu qu'il y a des pays plus industriels que d'autres.

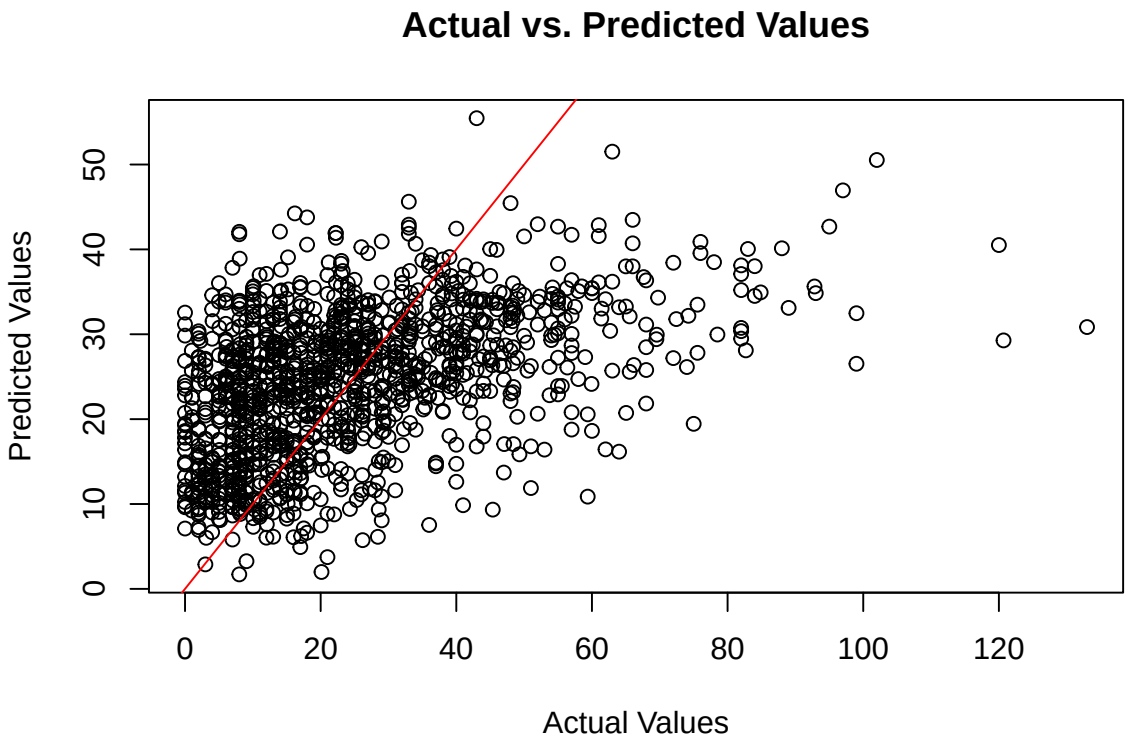
Effectivement le taux de gen_tot le plus bas (développment neuronal retardé) a été mesuré dans les pays avec le taux de pcb le plus élevé. Ceci a été mesuré dans le cas où ca soit la mère qui a consommé ces produits, mais leur consommation par l'enfant ne semble pas avoir d'influence significative.

# Prédiction: Methode LASSO

L'objectif à présent est de proposer un modèle multi-expositions afin de tenter de mettre en évidence un lien entre les exposomes et le développement neuronal de l'enfant, tout en prenant en compte les "covariates" dont on a identifié l'effet précédemment.

```
## Optimal lambda: 0.07428214
```

```
## Corrélation :  0.4640262
```




Le modèle proposé ici n'est pas très satisfaisant, ne proposant qu'une corrélation d'environ 46%. Le graphe suivant illustre l'aspect insatisfaisant du modèle, car les valeurs prédites peuvent être très différentes des valeurs réelles.

## Conclusion