

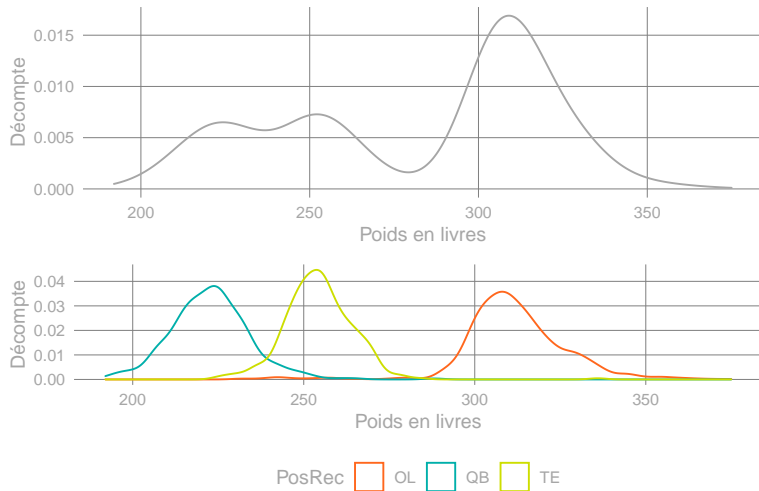
# Modèles de mélange

Nombre de paramètres et taille de l'échantillon

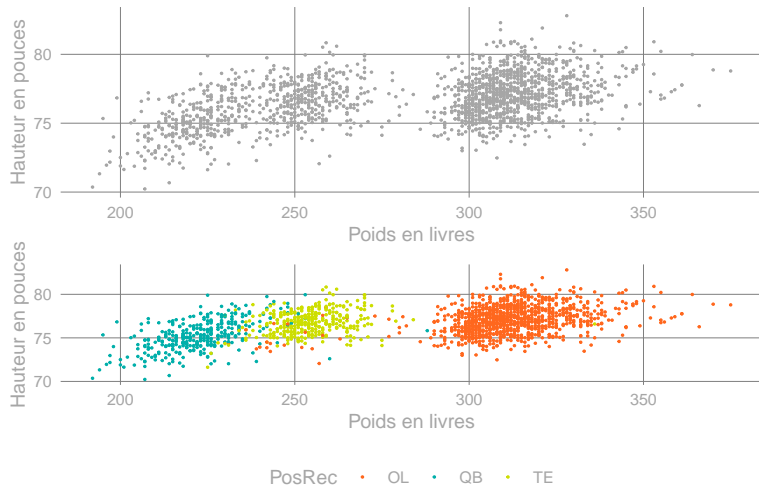
---

Véronique Tremblay

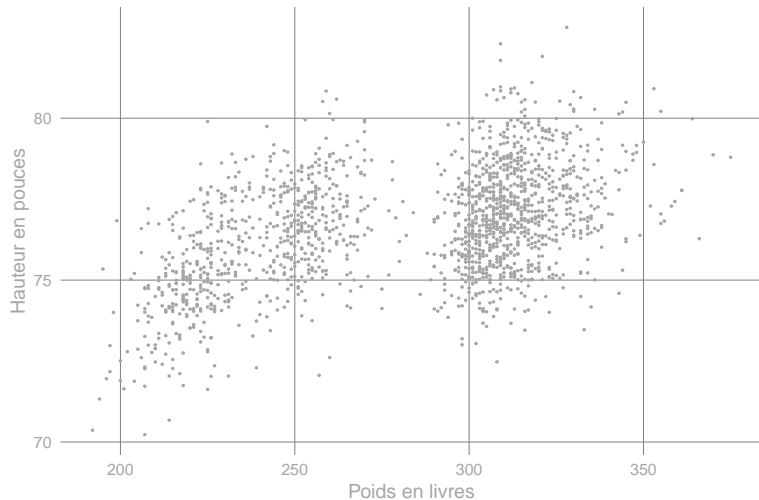
# Nombre de paramètres



# Nombre de paramètres



# Réduire le nombre de paramètres



## Reparamétrisation de $\Sigma$ dans mclust (Scrucca et al. (2016))

---

$$\Sigma_k = \lambda_k D_k A_k D_k^\top$$

où

$\lambda$  est un scalaire qui contrôle le volume

$A$  est une matrice diagonale qui contrôle la forme  
(ellipse)

$D$  est une matrice orthogonale qui contrôle l'orientation

# Système à 3 lettres

---

## La première lettre est liée au volume ( $\lambda$ )

- E : égal
- V : variable

## La deuxième lettre est liée à la forme ( $A$ )

- I : cercle
- E : ellipse égale dans chaque groupe
- V : ellipse différente pour chaque groupe

## La troisième lettre est liée à l'orientation ( $D$ )

- I : horizontale
- E : l'ellipse est inclinée, mais l'inclinaison est égale dans chaque groupe
- V : l'inclinaison de l'ellipse est différente pour chaque groupe

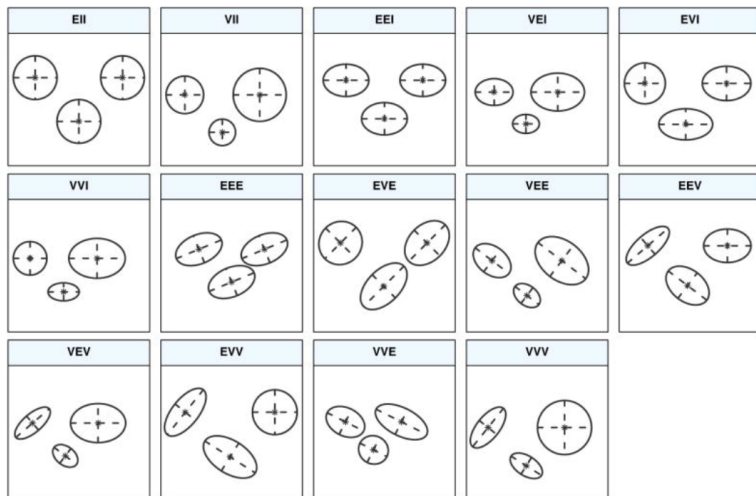


Image tirée de Scrucca et al. (2016)

## Choisir le bon modèle

---



- $\text{AIC} = 2Q - 2\ell(Y, \hat{\theta}),$
- $\text{BIC} = \log(n)Q - 2\ell(Y, \hat{\theta}).$

où  $Q$  est le nombre total de paramètres.

## Taille d'échantillon

---

*“ A common question asked by researchers is, “What sample size do I need for my study?” Over the years, several rules of thumb have been proposed, such as 5 to 10 observations per parameter, 50 observations per variable, no less than 100, and so on. In reality, there is no rule of thumb that applies to all situations. ” (Muthén & Muthén, 2002)*

*"It all depends on how well the classes are separated. I have done successful mixture modeling with only 30 subjects [...]. General rules of thumbs are not worth much for mixtures because results depend so much on the specifics of your situation." (Muthén, 2013)*

- Le nombre de paramètres à estimer augmente rapidement avec le nombre de groupes
- Le nombre d'observations requises est lié au nombre de paramètres

Scrucca, Luca, Michael Fop, Thomas Brendan Murphy, and Adrian E. Raftery. 2016. “mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.” *The R Journal* 8 (1): 205–33. <https://journal.r-project.org/archive/2016-1/scrucca-fop-murphy-et-al.pdf>.