

Valeurs extrêmes et valeurs aberrantes

Véronique Tremblay

Valeur extrême

Valeur éloignée des autres valeurs dans la population.

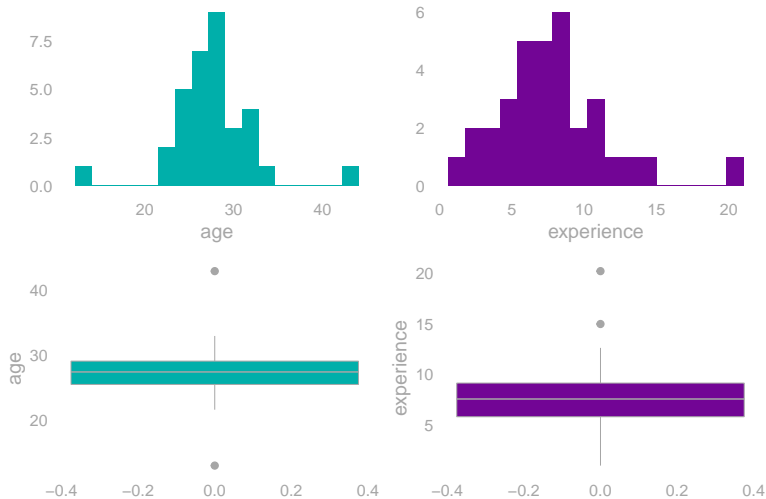
Peut correspondre à un profil particulier et plus rare dans la population.

Valeur aberrante

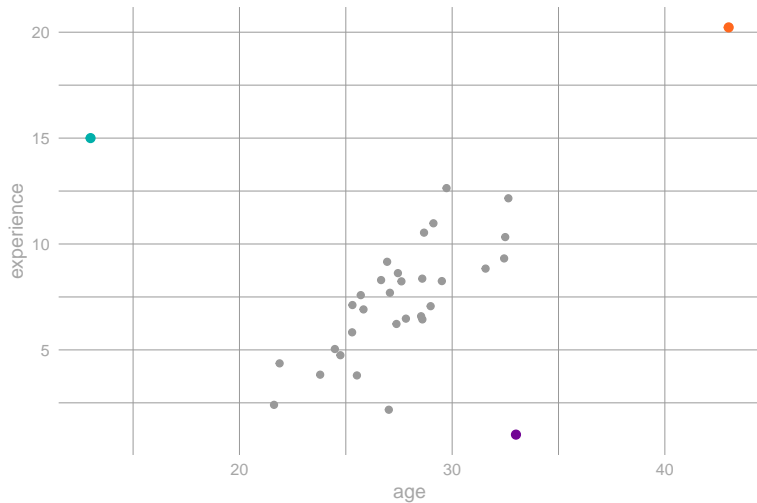
Valeur erronée causée par une erreur de saisie, une erreur de calcul, une mauvaise mesure ou une fausse déclaration.

- Exemples :
 - 8 chambres et 4 pièces
 - Premier achat en 2002 pour un client né en 2015
 - Retraité de 17 ans

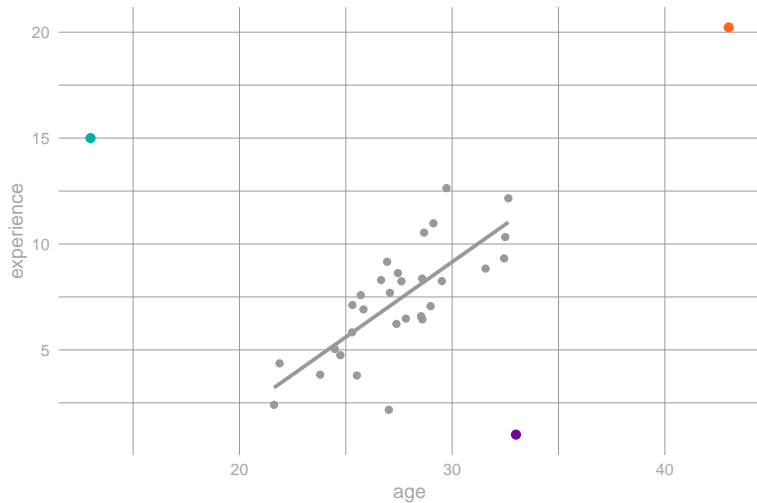
Exemple - visualisation unidimensionnelle



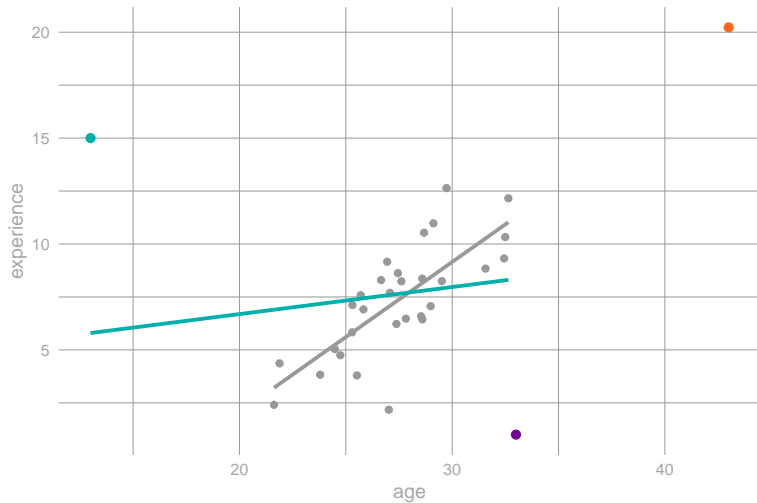
Exemple - visualisation bidimensionnelle



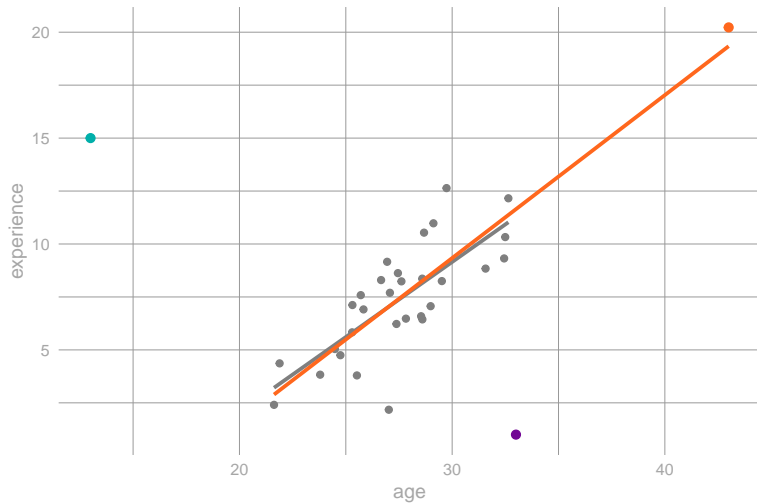
Qu'est ce que ça change?



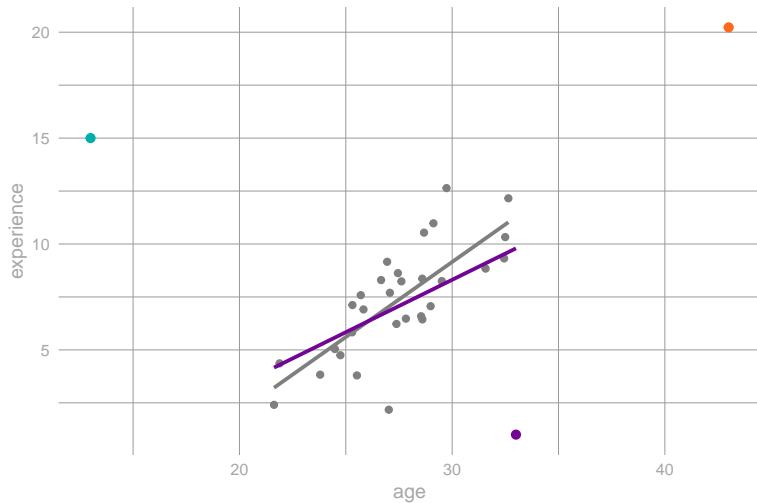
Valeur... aberrante? extrême? influente?



Valeur... aberrante? extrême? influente?



Valeur... aberrante? extrême? influente?



Quoi faire avec les valeurs aberrantes?

- Les remplacer par la bonne valeur
- Les traiter comme des valeurs manquantes
- Retirer l'individu

Quoi faire avec les valeurs extrêmes?

- Retirer les observations
- Utiliser des méthodes robustes
- Transformer la variable ($\log(X)$)
- Imputer par une valeur plus réaliste

Distinguer une valeur **extrême**, une valeur **aberrante** et
une valeur **influyente**.