Mesures de distance

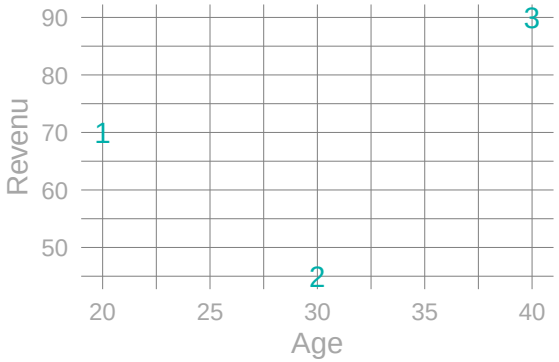
Variables continues

Véronique Tremblay

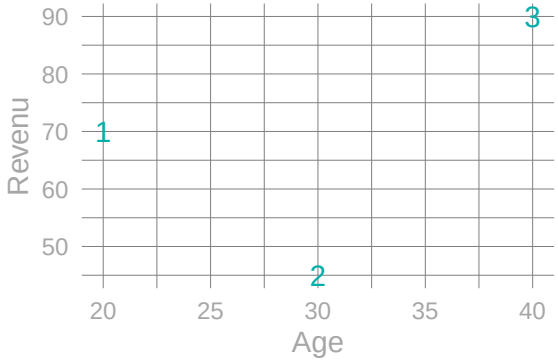
Variables continues

Distance euclidienne

$$d(i,j) = \sqrt{\sum_{k=1}^{K}(x_{ik} - x_{jk})^2}$$



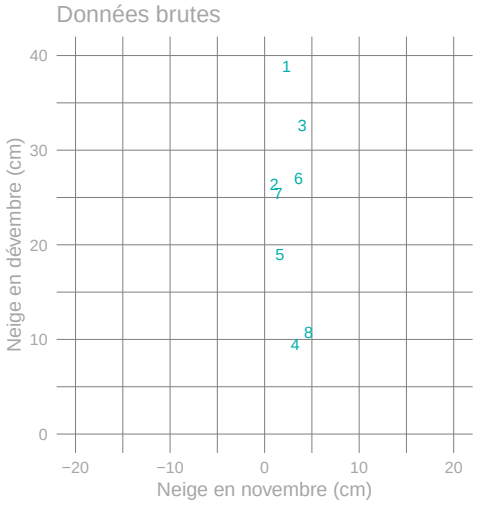
Exemple



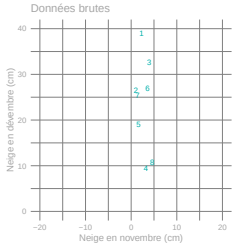
Distance euclidienne de l'individu 1 avec...

	l'individu 2	l'individu 3
Unités en k$	26.93	28.28
Unités en $	2.5×10^4	2×10^4

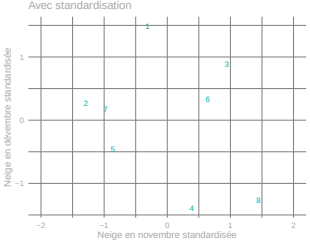
Impact de la standardisation



Effet de la standardisation



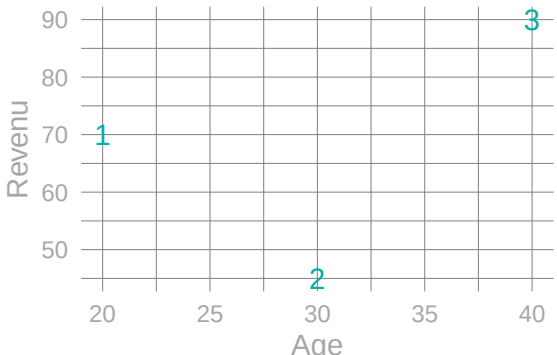
	1	2	3	4	5	6	7	8
1	0.0							
2	12.6	0.0						
3	6.5	6.9	0.0					
4	29.4	17.1	23.2	0.0				
5	19.9	7.4	13.8	9.7	0.0			
6	11.9	2.6	5.6	17.6	8.3	0.0		
7	13.5	1.1	7.7	16.0	6.4	2.7	0.0	
8	28.2	16.1	21.9	1.9	8.8	16.3	15.0	0.0



	1	2	3	4	5	6	7	8
1	0.0							
2	1.6	0.0						
3	1.4	2.3	0.0					
4	3.0	2.4	2.3	0.0				
5	2.0	0.8	2.2	1.6	0.0			
6	1.5	1.9	0.6	1.7	1.7	0.0		
7	1.5	0.3	2.0	2.1	0.6	1.6	0.0	
8	3.3	3.1	2.2	1.1	2.4	1.8	2.8	0.0

Distance de Manhatan

$$d(i,j) = \sum_{k=1}^{K} |x_{ik} - x_{jk}|$$



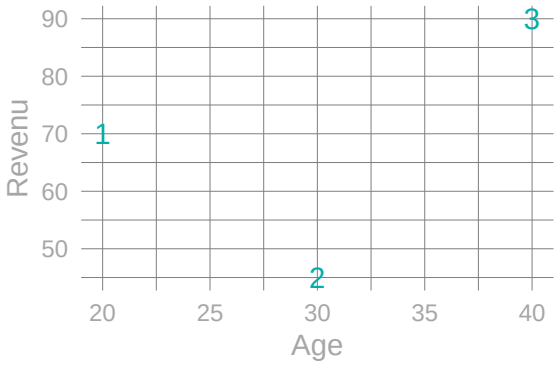
Variante de la distance de Manhatan

$$d(i,f) = \sum_{k=1}^{K} |x_{ik} - x_{jk}|/r_k$$

où r_k est l'étendue de la variable k.

Distance de Chebyshev

$$d(i,j) = \max(|x_{i1} - x_{j1}|, |x_{i2} - x_{j2}|...|x_{iK} - x_{iK}|)$$



Résumé

- Choisir la mesure de distance
- Standardisation?