

Corrigé

P1C2 - Explorez vos données et choisissez vos graphiques



1. **Distribution.** On veut voir comment une variable numérique unique (l'âge) se répartit sur l'ensemble des 918 patients : un histogramme est l'outil naturel. Sur ce dataset, l'âge moyen est d'environ 54 ans, avec un minimum de 28 ans et un maximum de 77 ans : un histogramme révélera si la population est concentrée autour de cette moyenne ou plus étalée.
2. **Relation.** On cherche à savoir si deux variables numériques, l'âge et la fréquence cardiaque maximale, évoluent ensemble. Un nuage de points est l'outil de référence pour révéler une éventuelle tendance, par exemple une fréquence cardiaque qui diminuerait avec l'âge.
3. **Comparaison.** On veut comparer la fréquence de 4 catégories de douleur thoracique entre elles. Un graphique en barres est adapté : sur nos données, la catégorie ASY (asymptomatique) domine très largement avec 496 patients sur 918, suivie par NAP (203), ATA (173) et TA (46).
4. **Composition.** On veut montrer comment l'ensemble des patients se répartit entre les 3 résultats possibles d'ECG au repos. Un graphique en barres empilées ou un donut chart convient : Normal représente la majorité (552 patients), suivi de LVH (188) et ST (178).
5. **Multi-variables.** Le profil d'un patient à risque ne se résume pas à une seule variable : il faut croiser plusieurs dimensions, par exemple l'âge, la fréquence cardiaque maximale et la présence ou non d'une maladie cardiaque (**HeartDisease**), pour faire émerger un profil. Un nuage de points colorié selon **HeartDisease** (508 patients malades contre 410 sains) permettrait de visualiser si les patients à risque se regroupent dans une zone particulière du graphique.

Erreurs fréquentes à éviter : utiliser un camembert pour comparer 4 catégories ou plus rend la lecture difficile, un graphique en barres triées est presque toujours préférable. De même, utiliser un graphique en barres pour représenter une relation entre deux variables numériques masque la tendance réelle : un nuage de points reste l'outil adapté.