RF (RANDOM FOREST)

- Principe de la forêt aléatoire :
  - Faire M arbres CART (une forêt !) puis moyenner les prédictions des M CART pour obtenir la prédiction finale.
  - Pour que les M CART ne soit pas tous identiques et « explorent » diverses combinaisons de X, on introduit de l'aléa dans la construction des M CART :
    - Chaque arbre CART est construit sur un échantillon d'observations **bootstrap** (tirage uniforme avec remise). Combiné avec le principe d'agrégation des différents CART, cela s'appelle le « **bagging** » (bootstrap aggregating)
    - Dans la construction d'un arbre, lors de la recherche de la meilleure coupure on ne teste pas l'ensemble des variables. En chaque nœud on tire un sous-ensemble de variables (paramètre *mtry)* parmi lesquelles on recherche la coupure optimale. En plus de l'exploration, on gagne ici du temps de calcul.
  - Les arbres CART ne sont **pas élagués** (gain de temps) et le critère d'arrêt est en général basé sur le nombre d'individus des feuilles (gain de temps là aussi)
  - On peut donc penser que chaque arbre est beaucoup sur-appris, toutefois la moyennisation permet de réduire le sur-apprentissage