

bosson.csv : 209 patients souffrant d'un anévrisme venant de France ou du Vietnam

- country (character) : "Vietnam", "France"
- gender (character) : "F", "M"
- aneurysm (numeric) : taille de l'anévrisme en mm
- bmi (numeric) : indice de masse corporelle en kg/m²
- risk (integer) : nombre de facteurs de risque entre 0 et 5

cardiaque.csv : 462 patients mâles dans une région à risque élevé de maladies cardiaques du Cap-Occidental (Afrique du Sud)

- sujet (integer) : numéro du patient
- systolique (numeric) : pression systolique en mmHg
- tabac (numeric) : tabac cumulé en kg
- cholesterol (numeric) : cholestérol de lipoprotéines de faible densité
- histFam (character) : antécédents familiaux ("Present" ou "Absent")
- hyperActivite (numeric) : comportement type A
- BMI (numeric) : indice de masse corporelle en kg/m²
- alcool (numeric) : consommation courante d'alcool
- age (numeric) : âge au moment de l'enquête en années
- cardiaque (integer) : maladie coronarienne (0=non, 1=oui)
- tabacBinaire (integer) : fumeur ou pas ? (1 si tabac > 0, 0 sinon)
- activiteBinaire (integer) : hyperactif ou pas ? (1 si hyperActivite > 60, 0 sinon)
- obesite (integer) : obésité (0=non, 1=oui)

ferretti.csv : 43 patients avec une tumeur du poulmon

- height (numeric) : taille de la tumeur en mm
- diameter (numeric) : diamètre de la tumeur en mm
- density (character) : différence entre la densité de la tumeur et celle de l'eau ("negative", "null" ou "positive")
- invasive (character) : tumeur invasive ou non ("yes" ou "no").

her.csv : 80 patients américains ayant passé un examen de santé

- iden (character) : identifiant de l'individu
- sex (integer) : 0 pour les hommes, 1 pour les femmes
- age (integer) : âge au moment de l'examen en années
- ht (numeric) : taille en cm
- wt (numeric) : poids en kg
- waist (numeric) : tour de taille en cm
- pulse (integer) : pulsation cardiaque en battements par minute
- sys (numeric) : pression systolique en mmHg
- dias (numeric) : pression diastolique en mmHg
- chol (numeric) : cholestérol en mg
- BMI (numeric) : indice de masse corporelle en kg/m²
- leg (numeric) : longueur de la partie haute de la jambe en cm
- elbow (numeric) : largeur du coude en cm

- wrist (numeric) : largeur du poignet en cm
- arm (numeric) : circonférence du bras en cm
- treat (integer) : 1 si traitement, 0 sinon

proteine.csv : 765 observations issues d'une étude menée par Hunault et Jaspard sur les protéines LEA (Late Embryogenesis Abundant proteins).

- longseq (integer) : longueur de la séquence en nombre d'acides aminés
- regne (character) : règne ("Viridiplantae" pour plante verte, "Fungi" pour champignon, "Metazoa" pour animal ou "Bacteria" pour bactérie)
- repliement (numeric) : indice de repliement
- isoelec (numeric) : point isoélectrique (c'est une valeur de pH)
- poidsmol (numeric) : poids moléculaire en g/mol
- hydro (numeric) : hydropathie moyenne
- genre (character) : genre biologique

titanic.csv : 1046 passagers du Titanic

- pclass (integer) : classe du passager (1, 2 ou 3)
- survived (character) : "yes" si le passager a survécu, "no" sinon
- gender (character) : sexe ("F" ou "M")
- age (numeric) : âge en années