Descriptions of the datasets we will use during the course

Heart failure dataset

This dataset contains 918 patients and 12 variables.

Source: Heart Failure Prediction Dataset, Kaggle

Fedesoriano. (September 2021). Heart Failure Prediction Dataset. Retrieved [08-12-2021] from
https://www.kaggle.com/fedesoriano/heart-failure-prediction

Variable	Type		Description
Age	Numerical	Continuous	age of the patient in years
Sex	Character	Categorical M = male F = female	Gender of the patient
ChestPainType	Character	Categorical TA = typical angina ATA = atypical angina NAP = non-anginal pain ASY = asymptomatic	Type of chestpain
RestingBP	Numerical	Continuous	Resting blood pressure [mm Hg]
Cholesterol	Numerical	Continuous	Serum cholesterol [mm/dl]
FastingBS	Numerical	Categorical 0 = otherwise 1 = FastingBS > 120 mg/dl	Fasting blood sugar [mg/dl]
RestingECG	Character	Categorical Normal = normal ECG ST = having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.005 mV) LVH = showing probable or definite left ventricular hypertrophy by Estes' criteria)	Resting electrocardiogram results
MaxHR	Numerical	Continuous	Maximum heart rate achieved [b/m]

ExerciseAngina	Character	Categorical Y = yes N = No	Exercise-induced angina
Oldpeak	Numerical	Continuous	Numerical value measured in depression (ST)
ST_Slope	Character	Categorical Flat = flat slope Up = upsloping Down = downsloping	The slope of the peak exercise ST segment
HeartDisease	Numerical	Categorical 0 = normal 1 = heart disease	Label that classifies whether the patient had truly heart disease or not

Covid dataset

This dataset contains 25 COVID-19 related variables for 89 countries over 2020-10-01 till 2021-10-01.

Source: Covid Dataset, Our world in Data (all missing values were filtered out)

Hasell, J., Mathieu, E., Beltekian, D. et al. A cross-country database of COVID-19 testing. Sci Data 7, 345 (2020). https://doi.org/10.1038/s41597-020-00688-8

Mathieu, E., Ritchie, H., Ortiz-Ospina, E. et al. A global database of COVID-19 vaccinations. Nat Hum Behav (2021). https://doi.org/10.1038/s41562-021-01122-8

Variable	Type		Description
Continent	Character	Categorical	Continent of the geographical location
Location	Character	Categorical	Geographical location
Date	Character	Continuous	Date of observation
Total_cases	Numerical	Continuous	Total confirmed cases of COVID-19
Total_deaths	Numerical	Continuous	Total deaths attributed to COVID-19
Total_cases_per_million	Numerical	Continuous	Total confirmed cases of COVID-19 per 1,000,000 people

New_cases_per_million	Numerical	Continuous	New confirmed cases of COVID-19 per 1,000,000 people
Total_deaths_per_million	Numerical	Continuous	Total deaths attributed to COVID-19 per 1,000,000
New_deaths_per_million	Numerical	Continuous	New deaths attributed to COVID-19 per 1,000,000
Reproduction_rate	Numerical	Continuous	Estimate of the effective reproduction rate (R) of COVID-19
Stringency_index	Numerical	Categorical	Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response)
Population	Numerical	Continuous	Population (latest available values)
Population_density	Numerical	Continuous	Number of people divided by land areas, measured in square kilometers, most recent year available
Median_age	Numerical	Continuous	Media age of the population, UN projection for 2020

Aged_65_older	Numerical	Continuous	Share of the population that is 65 years and older, most recent year available
Aged_70_older	Numerical	Continuous	Share of the population that is 70 years and older in 2015
Gdp_per_capita	Numerical	Continuous	Gross domestic product at purchasing power parity, dollars, most recent year available
Extreme_poverty	Numerical	Continuous	Share of population living in extreme poverty, most recent year available since 2010
Cardiovasc_death_rate	Numerical	Continuous	Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people)
Diabetes_prevalence	Numerical	Continuous	Diabetes prevalence (% of population aged 20 to 79) in 2017
Female_smokers	Numerical	Continuous	Share of women who smoke, most recent year available
Male_smokers	Numerical	Continuous	Share of men who smoke, most recent year available
Hospital_beds_per_thousand	Numerical	Continuous	Hospital beds per 1,000 people, most recent year available since 2010

Life_expectancy	Numerical	Continuous	Life expectancy at birth in 2019
Human_development_index	Numerical	Continuous	A composite index measuring average achievement in three basic dimensions of human development – a long and healthy life, knowledge, and a decent standard of living. Values for 2019.

Chronic pain dataset

This dataset contains 742 patients who presented with chronic pain at the Pain clinic of the Maastricht UMC+, who were asked to complete a number of patient-reported outcome measures (PROMs).

Data are a subset of patients and features from:

Brouwer, B., Waardenburg, S., Jacobs, C. et al. Biopsychosocial baseline values of 15 000 patients suffering from chronic pain: Dutch DataPain study. Reg Anesth Pain Med. 2020 Oct;45(10):774-782. doi: 10.1136/rapm-2020-101476.

Variable	Type		Description
ID	Integer	Categorical	Unique identifier
Sex	Integer	Categorical	1 = male, 2 = female
Age	Integer	Continuous	Age at presentation
Partner	Integer	Categorical	0 = no, 1 = yes
Child	Integer	Categorical	0 = no, 1 = yes
Randsc_ff	Integer	Continuous	RAND-36 score for the physical functioning domain
Randsc_sf	Integer	Continuous	RAND-36 score for the social functioning domain
Randsc_rf	Integer	Continuous	RAND-36 score for the role physical domain

Randsc_re	Integer	Continuous	RAND-36 score for the role emotional domain
Randsc_mg	Integer	Continuous	RAND-36 score for the mental health domain
Randsc_v	Integer	Continuous	RAND-36 score for the vitality domain
Randsc_p	Integer	Continuous	RAND-36 score for the pain domain
Randsc_gv	Integer	Continuous	RAND-36 score for the health change domain
Randsc_ag	Integer	Continuous	RAND-36 score for the general health domain
Pcs_sc	Integer	Continuous	Pain Catastrophizing Scale (PCS) score
Hadssc_a	Integer	Continuous	Hospital Anxiety and Depression Scale anxiety score
Hadssc_d	Integer	Continuous	Hospital Anxiety and Depression Scale depression score
Hadssc_t	Integer	Continuous	Hospital Anxiety and Depression Scale total score