Crime Data Pipeline (USArrests) — with Student Notes

Grad Student (fictional)

2025-10-14

Contents

Purpose & Provenance

Student note — Goal. Build a reproducible data-processing/EDA pipeline on the USArrests dataset (rates of violent crimes per 100,000 residents in 1973, across the 50 U.S. states). I'll load, tidy, standardize, explore, flag outliers, do a quick dimensionality reduction (PCA), and save a clean version.

Input. Built-in R dataset USArrests Output. A clean tibble with a state column and standardized columns, plus a few sanity plots.