🕸 DAA Operations & Visualizations Categories 🕸

Type	Name	Difficulty	Description
Operation	Mean	Easy 🔴	The average of a set of values.
Operation	Median	Easy 🔴	The middle value in a sorted dataset.
Operation	Mode	Easy 🔴	The most frequently occurring value in a dataset.
Operation	Variance	Easy 🔴	A measure of how data points spread out from the mean.
Operation	Standard deviation	Easy 🔴	The square root of variance, indicating spread around the mean.
Operation	Min/Max	Easy 🔴	The smallest (min) and largest (max) values in a dataset.
Operation	Percentiles / Quantiles	Easy 🔴	Values below which a certain percentage of data falls.
Operation	Correlation (Pearson's r)	Easy 🔴	Strength and direction of linear relationship between two variables.
Operation	Covariance	Easy 🔴	Measures how two variables vary together (sign indicates direction).
Operation	Simple Linear Regression	Easy 🔴	Modeling the relationship between two variables with a best-fit line.

Type	Name	Difficulty	Description
Operation	z-score calculation	Easy 🔴	Standardizing a value by number of standard deviations from the mean.
Operation	t-test (one-sample, two-sample)	Easy 🔴	Statistical test comparing means (one or two groups) relative to variance.
Operation	Chi-square test	Easy 🔴	Statistical test comparing observed categorical frequencies to expected frequencies.
Operation	Shapiro-Wilk normality test	Easy 🔴	Test to assess whether a dataset is drawn from a normal distribution.
Operation	Frequency counts / Value counts	Easy 🔴	Counting occurrences of each unique value or category in a dataset.
Operation	Moving average (time series smoothing)	Easy 🔴	Smoothing a time series by averaging data points in a rolling window.
Operation	Proportion / Probability calculation	Easy 🔴	Calculating the fraction of occurrences or likelihood of an event.
Operation	Confidence interval (basic)	Easy 🔴	Range of values likely to contain the true population parameter at a given confidence.
Operation	Rank / Percent rank	Easy 🔴	Ordering data and determining the percentile position of each value.

Type	Name	Difficulty	Description
Operation	Bayes' Theorem (basic calculation)	Easy 🔴	Rule for updating probabilities based on prior belief and new evidence.
Operation	Binning / Discretization	Easy 🔴	Converting continuous values into categorical bins or intervals.
Operation	F-test (for equality of variances)	Easy 🔴	Test to compare the variances of two or more groups.
Operation	Lift / Gain calculation (for classification models)	Easy 🔴	Measuring model improvement over baseline by comparing cumulative gains.
Operation	Precision/Recall/F1 calculation	Easy 🔴	Computing precision, recall, and F1-score from classification outcomes.
Operation	Confusion matrix analysis	Easy 🔴	Evaluating classifier performance by tabulating true/false positives and negatives.
Operation	p-value interpretation	Easy 🔴	Understanding the significance level: probability of data given the null hypothesis.
Operation	R^2 (Coefficient of Determination)	Easy 🔴	Proportion of variance in the outcome explained by the regression model.
Operation	Range calculation	Easy 🔴	Computing the difference between the maximum and minimum values in data.

Type	Name	Difficulty	Description
Operation	Cumulative sum / product	Easy 🔴	Sequential accumulation of sums or products across a series of numbers.
Operation	Lag / Lead operations	Easy 🔴	Shifting a time series forward (lead) or backward (lag) for analysis.
Operation	Difference / Percent change	Easy 🔴	Calculating change between successive values, in absolute or percentage terms.
Operation	Rolling statistics (sum, mean, std)	Easy 🔴	Calculating statistics over a moving window (e.g., 7-day rolling average).
Operation	Normalization (0-1 scaling)	Easy 🔴	Rescaling data to a 0 to 1 range (min-max normalization).
Operation	Feature scaling (z-score, min-max, robust)	Easy 🔴	Rescaling features using methods like standardization or min-max scaling.
Operation	One-hot encoding	Easy 🔴	Converting categorical values into binary indicator columns.
Operation	Label encoding	Easy 🔴	Assigning each categorical label a unique numeric code.
Operation	Skewness calculation	Easy 🔴	Measuring the asymmetry of a distribution around its mean.

Type	Name	Difficulty	Description
Operation	Kurtosis calculation	Easy 🔴	Measuring the 'tailedness' (heaviness of tails) of a distribution.
Operation	Interquartile range (IQR)	Easy 🔴	The range between the 25th and 75th percentile (middle 50% of data).
Operation	Outlier detection (IQR, z-score)	Easy 🔴	Identifying anomalous points that lie far from the rest of the data.
Operation	Coefficient of variation	Easy 🔴	Ratio of standard deviation to the mean (a normalized measure of spread).
Operation	Cross-tabulation	Easy 🔴	Contingency table showing the frequency distribution of two categorical variables.
Operation	Data type conversion	Easy 🔴	Changing the data type of values (e.g., string to integer).
Operation	Simple exponential smoothing	Easy 🔴	Forecasting method giving exponentially decreasing weights to past observations.
Operation	ANOVA (one-way)	Med 🟡	Analysis of variance to test if the means of multiple groups are equal.
Operation	Logistic regression	Med 🟡	Regression technique for binary outcomes using a logistic function.
Operation	Spearman rank correlation	Med 🟡	Non-parametric correlation for ranked data (monotonic relationship measure).

Type	Name	Difficulty	Description
Operation	Time-series decomposition (trend, season)	Med 🟡	Breaking a time series into trend, seasonal, and residual components.
Operation	PCA (Principal Component Analysis)	Med 🟡	Dimensionality reduction by transforming to a new set of orthogonal variables (principal components).
Operation	Clustering (k-means, hierarchical)	Med 🟡	Unsupervised grouping of data into clusters based on similarity.
Operation	Missing value imputation	Med 🟡	Filling in missing data using methods like mean substitution or predictive models.
Operation	Feature engineering (basic aggregations/transformations)	Med 🟡	Creating or transforming features based on raw data to improve model performance.
Operation	Bootstrapping	Med 🟡	Resampling technique to estimate statistics (and confidence intervals) by sampling with replacement.
Operation	Cross-validation (k-fold)	Med 🟡	Evaluating model performance by training and testing on multiple data splits.
Operation	Hyperparameter tuning (grid search)	Med 🟡	Systematically searching through combinations of model parameters for best performance.

Type	Name	Difficulty	Description
Operation	Kaplan-Meier estimator (survival)	Med 🟡	Non-parametric estimate of the survival function from lifetime data.
Operation	Multiple comparison corrections	Med 🟡	Adjusting p-values or significance criteria when performing multiple statistical tests (e.g., Bonferroni correction).
Operation	Chi-square test for independence	Med 🟡	Chi-square test to determine if two categorical variables are statistically independent.
Operation	Gini coefficient (simple)	Med 🟡	Measure of inequality or dispersion (e.g., income inequality) ranging from 0 (equal) to 1 (max inequality).
Operation	Cohen's d (effect size)	Med 🟡	Standardized measure of mean difference (in standard deviation units).
Operation	ROC AUC calculation	Med 🟡	Computing the area under the ROC curve, summarizing binary classifier performance.
Operation	Predictive modeling (Decision Trees, Random Forest)	Med 🟡	Building models (like tree-based) to predict an outcome from input features.
Operation	Q-Q plot interpretation	Med 🟡	Assessing if data follows a distribution by comparing empirical quantiles to theoretical ones in a Q-Q plot.

Type	Name	Difficulty	Description
Operation	McNemar's test (paired nominal data)	Med 🟡	Test for changes in paired binary outcomes (e.g., before vs after in a 2x2 table).
Operation	Non-parametric tests (Mann-Whitney U, Kruskal-Wallis)	Med 🟡	Rank-based tests for comparing medians across groups without assuming normal distribution.
Operation	Levene's test (homogeneity of variance)	Med 🟡	Test to assess if multiple groups have equal variances.
Operation	ANOVA (two-way, repeated measures)	Med 🟡	ANOVA with two factors or with measurements repeated on the same subjects.
Operation	Polynomial regression	Med 🟡	Regression model that includes polynomial terms to capture non-linear relationships.
Operation	Ridge regression	Med 🟡	Linear regression with L2 regularization to reduce overfitting.
Operation	Lasso regression	Med 🟡	Linear regression with L1 regularization for feature selection/shrinkage.
Operation	Gradient Boosting (XGBoost, LightGBM)	Med 🟡	Ensemble method building strong models by sequentially adding weak learners (with gradient boosting).
Operation	Support Vector Machines (SVM)	Med 🟡	Classification/regression technique finding a

Type	Name	Difficulty	Description
			hyperplane maximizing the margin between classes.
Operation	Naive Bayes classification	Med 🟡	Probabilistic classifier assuming features are independent given the class.
Operation	KNN (K-Nearest Neighbors)	Med 🟡	Instance-based learning method classifying based on closest training examples in feature space.
Operation	DBSCAN clustering	Med 🟡	Density-based clustering that groups points closely packed together and labels sparse regions as noise.
Operation	Gaussian Mixture Models (GMM)	Med 🟡	Probabilistic clustering modeling data as a mixture of Gaussian distributions.
Operation	LDA (Linear Discriminant Analysis)	Med 🟡	Technique to find linear combinations of features that best separate two or more classes.
Operation	Feature selection (filter methods)	Med 🟡	Selecting features based on inherent properties (e.g., correlation with target, without using a model).
Operation	Feature selection (wrapper methods)	Med 🟡	Selecting feature subsets by training models (wrappers) and choosing the subset with best performance.

Type	Name	Difficulty	Description
Operation	Model interpretation (SHAP, LIME for basic models)	Med 🟡	Techniques (like SHAP, LIME) to explain predictions by attributing contributions to features.
Operation	Matthews correlation coefficient	Med 🟡	Balanced measure of binary classification performance (accounts for all confusion matrix categories).
Operation	Cramér's V	Med 🟡	Measure of association between two nominal categorical variables (0 = no association, 1 = perfect association).
Operation	Kendall's tau	Med 🟡	Non-parametric correlation coefficient for ordinal data measuring the similarity of orderings.
Operation	Point-biserial correlation	Med 🟡	Correlation between a continuous variable and a binary categorical variable.
Operation	Phi coefficient	Med 🟡	Correlation coefficient specifically for two binary variables (equivalent to Pearson's r on binary data).
Operation	Fisher's exact test	Med 🟡	Exact test for independence in a 2x2 contingency table (useful for small sample sizes).
Operation	Wilcoxon signed-rank test	Med 🟡	Non-parametric test for comparing two paired samples (tests median differences).

Type	Name	Difficulty	Description
Operation	Friedman test	Med 🟡	Non-parametric test for differences across multiple paired groups (like repeated measures ANOVA).
Operation	Dunn's test	Med 🟡	Post-hoc test following a Kruskal-Wallis to identify which groups differ.
Operation	Anderson-Darling test	Med 🟡	Test for whether a sample comes from a specific distribution (with emphasis on tails).
Operation	Kolmogorov-Smirnov test	Med 🟡	Non-parametric test to compare a sample with a reference distribution or two samples with each other.
Operation	Ljung-Box test (autocorrelation)	Med 🟡	Test to check whether a time series has significant lack-of-fit by examining autocorrelations of residuals.
Operation	Augmented Dickey-Fuller test	Med 🟡	Statistical test for a unit root in a time series (to check stationarity).
Operation	KPSS test (stationarity)	Med 🟡	Statistical test for stationarity in a time series (null hypothesis that series is stationary).
Operation	Granger causality test	Med 🟡	Test to determine if one time series is useful in forecasting another (based on lagged regression).

Type	Name	Difficulty	Description
Operation	ACF/PACF analysis	Med 🟡	Analyzing autocorrelation and partial autocorrelation plots to identify time series patterns.
Operation	Holt-Winters method	Med 🟡	Triple exponential smoothing for time series forecasting with trend and seasonality.
Operation	ARIMA modeling	Med 🟡	Time series forecasting using autoregressive, integrated, moving average components.
Operation	Seasonal decomposition (STL)	Med 🟡	Decomposing a time series into seasonal, trend, and remainder components using STL.
Operation	Experimental design (A/B/n tests, factorial)	Med 🟡	Planning experiments (A/B tests or multi-factor) to reliably test hypotheses.
Operation	Power analysis	Med 🟡	Determining sample size required to detect an effect of a given size with sufficient power.
Operation	Permutation tests	Med 🟡	Computing significance by comparing observed statistics to distribution of statistics from permuted data.
Operation	Jackknife resampling	Med 🟡	Estimating statistics by systematically leaving out one observation at a time from the sample.

Type	Name	Difficulty	Description
Operation	Stratified sampling	Med 🟡	Sampling method that preserves the proportion of subgroups (strata) in the sample.
Operation	SMOTE (oversampling)	Med 🟡	Synthetic Minority Oversampling Technique to generate artificial samples for minority class.
Operation	Tomek links (undersampling)	Med 🟡	Undersampling method removing overlapping majority-minority pairs (Tomek links) to clean data.
Operation	Bayesian inference with MCMC	Hard 🟢	Using Markov Chain Monte Carlo to approximate the posterior distribution in Bayesian models.
Operation	Causal inference (propensity scores, IVs)	Hard 🟢	Estimating causal effects using methods like propensity score matching or instrumental variables.
Operation	Deep learning for structured data	Hard 🟢	Applying neural networks (deep learning) to tabular or relational datasets.
Operation	Text Vectorization (TF-IDF, Bag-of-Words)	Hard 🟢	Converting text into numerical feature representations (e.g., term-frequency vectors).
Operation	Survival analysis (Cox model, hazards)	Hard 🟢	Analyzing time-to-event data, often using Cox proportional hazards models for risk over time.

Type	Name	Difficulty	Description
Operation	ARIMAX / SARIMAX forecasting	Hard 🟢	Advanced time series models (ARIMA) that include exogenous variables or seasonal components.
Operation	Prophet/ETS forecasting models	Hard 🟢	Forecasting time series using specialized models like Facebook Prophet or exponential smoothing (ETS).
Operation	Reinforcement learning methods	Hard 🟢	Training agents to make sequential decisions via reward feedback (exploration and exploitation).
Operation	Dimensionality reduction (t-SNE, UMAP)	Hard 🟢	Non-linear techniques to embed high-dimensional data into low-dimensional space for visualization.
Operation	Regularization (elastic net, L1/L2 mix)	Hard 🟢	Penalization techniques combining L1 and L2 regularization to prevent overfitting.
Operation	Graph-based learning (embeddings, detection)	Hard 🟢	Machine learning methods that operate on graph data structures (e.g., node embeddings, community detection).
Operation	Advanced hierarchical models (mixed effects)	Hard 🟢	Statistical models with both fixed and random effects, handling nested data structures.

Type	Name	Difficulty	Description
Operation	Generative models (GANs, VAEs)	Hard 🟢	Neural network models that learn to generate new data similar to the training data.
Operation	Bayesian hierarchical models	Hard 🟢	Bayesian models with multi-level structures allowing partial pooling across groups.
Operation	Simulation modeling (Monte Carlo, agent-based)	Hard 🟢	Using repeated random sampling or agent simulation to model complex processes or estimate uncertainty.
Operation	Extreme Value Theory (EVT)	Hard 🟢	Statistical framework focusing on the extreme deviations (tails) of distributions.
Operation	Change point detection (e.g., PELT, CUSUM)	Hard 🟢	Methods for identifying points where the statistical properties of a sequence change significantly.
Operation	Optimal Transport / Earth Mover's Distance	Hard 🟢	Computing the minimal cost to transform one distribution into another (measure of distribution difference).
Operation	Functional Data Analysis (FDA)	Hard 🟢	Analyzing data that come as functions or curves (e.g., growth trajectories or time series as single units).
Operation	Bayesian Networks	Hard 🟢	Probabilistic graphical models representing variables and conditional

Type	Name	Difficulty	Description
			dependencies with directed acyclic graphs.
Operation	Hidden Markov Models (HMM)	Hard 🟢	Statistical models for sequence data with hidden states and observed outputs following Markov processes.
Operation	Conditional Random Fields (CRF)	Hard 🟢	Probabilistic models for structured prediction (e.g., sequences) that model conditional probability of label sequences given input sequences.
Operation	Variational inference	Hard 🟢	Approximate inference method that turns inference into an optimization problem using variational approximations.
Operation	Gaussian Processes	Hard 🟢	Non-parametric models defining a distribution over functions, used for regression and classification with uncertainty estimates.
Operation	Quantile regression	Hard 🟢	Regression technique that estimates conditional quantiles (like median or other percentiles) instead of the mean.
Operation	Zero-inflated models	Hard 🟢	Statistical models for count data with excess zeros, combining a point mass at zero and a count distribution.

Type	Name	Difficulty	Description
Operation	Hurdle models	Hard 🟢	Two-part models for count data: one part for zero vs non-zero, and one for positive counts.
Operation	Structural Equation Modeling (SEM)	Hard 🟢	Statistical technique combining factor analysis and multiple regression to model complex relationships, including latent variables.
Operation	Mediation analysis	Hard 🟢	Analyzing whether the effect of an independent variable on an outcome is transmitted through an intermediate (mediator) variable.
Operation	Instrumental variables regression	Hard 🟢	Regression method using instruments (variables correlated with the independent variable but not with the error term) to estimate causal relationships.
Operation	Difference-in-differences	Hard 🟢	Quasi-experimental method using before-and-after differences between treated and control groups to estimate causal effects.
Operation	Regression discontinuity design	Hard 🟢	Causal inference method exploiting a threshold-based assignment to approximate an experiment at the cutoff.

Type	Name	Difficulty	Description
Operation	Synthetic control methods	Hard 🟢	Using a weighted combination of control units to construct a synthetic comparison for a treated unit in observational study.
Operation	Transfer learning	Hard 🟢	Improving learning in a new task by leveraging knowledge (models or features) from a related task.
Operation	Meta-learning	Hard 🟢	Techniques where models learn how to learn, enabling quick adaptation to new tasks with minimal data.
Operation	Federated learning	Hard 🟢	Training models across decentralized devices or servers holding local data, without sharing raw data (privacy-preserving).
Operation	Attention mechanisms / Transformers	Hard 🟢	Neural network components focusing on relevant parts of input (Transformers rely on self-attention for sequence modeling).
Operation	Graph Neural Networks (GNN)	Hard 🟢	Neural networks designed to directly operate on graph-structured data.
Operation	Autoencoders	Hard 🟢	Neural networks trained to compress data into a lower-dimensional representation and reconstruct it back.

Type	Name	Difficulty	Description
Operation	Recurrent Neural Networks (RNN/LSTM/GRU)	Hard 🟢	Neural networks for sequential data that maintain state across sequence steps (LSTM/GRU are advanced RNN variants).
Operation	Capsule Networks	Hard 🟢	Neural network architecture with capsules that aim to preserve spatial hierarchies (an alternative to traditional convnets).
Operation	Neural Architecture Search (NAS)	Hard 🟢	Automated searching of neural network architectures using algorithms to optimize design for a given task.
Operation	Adversarial training	Hard 🟢	Training models with adversarial examples to improve robustness against intentionally perturbed inputs.
Operation	Conformal prediction	Hard 🟢	Framework for providing prediction intervals or sets with a formal coverage guarantee, often model-agnostic.
Operation	Copula modeling	Hard 🟢	Statistical modeling of complex multivariate dependencies by separating marginal distributions from dependency structure.
Operation	Spatial statistics (kriging, variograms)	Hard 🟢	Analyzing spatially correlated data (e.g.,

Type	Name	Difficulty	Description
			geostatistics) including interpolation methods like kriging.
Operation	Point process models	Hard 🟢	Models for random points in time or space (e.g., Poisson processes and their variants for events occurrence).
Operation	Spectral analysis (FFT, wavelets)	Hard 🟢	Decomposing time series or signals into frequency components (FFT for overall spectrum, wavelets for time-frequency).
Operation	Independent Component Analysis (ICA)	Hard 🟢	Unsupervised technique to separate a multivariate signal into additive independent non-Gaussian components.
Operation	Non-negative Matrix Factorization (NMF)	Hard 🟢	Matrix factorization that decomposes data into non-negative factors, useful for parts-based representation.
Operation	Canonical Correlation Analysis (CCA)	Hard 🟢	Finding linear combinations of two sets of variables that are maximally correlated with each other.
Operation	Partial Least Squares (PLS)	Hard 🟢	Regression technique that finds latent variables by maximizing covariance between predictors and response.

Type	Name	Difficulty	Description
Operation	Manifold learning (Isomap, LLE)	Hard 🟢	Non-linear techniques to discover low-dimensional manifold structure embedded in high-dimensional data.
Operation	Topological Data Analysis (TDA)	Hard 🟢	Analyzing the shape of data using techniques from topology (e.g., persistent homology to identify clusters, holes).
Operation	Causal discovery algorithms	Hard 🟢	Algorithms (like PC, FCI) to infer causal relationships or structures from observational data.
Visualization	Histogram	Easy 🔴	A bar chart representing the frequency distribution of a single numeric variable.
Visualization	Bar chart	Easy 🔴	Chart with rectangular bars representing values or counts for different categories.
Visualization	Line plot (time series)	Easy 🔴	Graph of data points connected by lines, often used for trends over time.
Visualization	Scatter plot	Easy 🔴	Chart of paired x and y values as points to show the relationship between two variables.
Visualization	Pie chart	Easy 🔴	Circular chart divided into slices to illustrate proportion of categories in a whole.

Type	Name	Difficulty	Description
Visualization	Area chart	Easy 🔴	Line chart with the area below the line filled in, showing cumulative magnitude over time or categories.
Visualization	Donut chart	Easy 🔴	A variant of the pie chart with a blank center, used to display parts of a whole.
Visualization	Stacked bar chart	Easy 🔴	Bar chart with segments stacked on each bar, showing parts of a total per category.
Visualization	Gauge chart	Easy 🔴	Circular dial (like a speedometer) indicating a value or progress toward a target.
Visualization	Column chart	Easy 🔴	Vertical bar chart (bars extend upwards from the horizontal axis).
Visualization	Step chart	Easy 🔴	Line chart where changes occur as abrupt steps, staying constant between jumps.
Visualization	Sparkline	Easy 🔴	Very small line chart (usually without axes) embedded to show a quick trend.
Visualization	Bullet chart	Easy 🔴	Bar chart variant showing performance against a target, with markers for qualitative ranges.

Type	Name	Difficulty	Description
Visualization	Range plot	Easy 🔴	Chart displaying a range (e.g., high-low values) for each category or time point.
Visualization	Lollipop chart	Easy 🔴	Bar chart variation with a line and a dot at the end, resembling a lollipop.
Visualization	Grouped bar chart	Easy 🔴	Bar chart where categories have multiple sub-bars grouped side by side.
Visualization	Table with conditional formatting	Easy 🔴	Data table where cell formatting (like color) highlights certain values or patterns.
Visualization	Waterfall chart	Med 🟡	Chart showing a running total as values are added or subtracted, highlighting contributions to the final total.
Visualization	Funnel chart	Med 🟡	Chart showing stages in a process (e.g., sales funnel) with progressively smaller bars indicating drop-off at each stage.
Visualization	Heatmap (correlations, confusion matrix)	Med 🟡	Grid of cells colored by value, commonly used to display correlation matrices or frequencies (e.g., confusion matrix).
Visualization	Boxplot	Med 🟡	Chart depicting distribution of a dataset via quartiles and outliers (box-and-whisker plot).

Type	Name	Difficulty	Description
Visualization	Violin plot	Med 🟡	Similar to a boxplot, but shows the full distribution density mirrored on each side.
Visualization	Pairplot / Scatterplot matrix	Med 🟡	Grid of scatter plots for each pair of multiple variables (and often histograms on the diagonal).
Visualization	Bubble chart	Med 🟡	Scatter plot where the size of each point represents an additional variable.
Visualization	Density plot (KDE)	Med 🟡	Smoothed curve representing the distribution of a numeric variable (kernel density estimate).
Visualization	Radar / Spider chart	Med 🟡	Chart with multiple axes from a central point, showing multivariate data as a polygon (one axis per variable).
Visualization	Sankey diagram	Med 🟡	Flow diagram showing how quantities split and merge between categories, with arrow widths proportional to flow volume.
Visualization	Tree map	Med 🟡	Nested rectangles representing a hierarchy, where area of each rectangle is proportional to a quantitative value.
Visualization	Control chart (Shewhart, I-MR)	Med 🟡	Process control chart showing measurements

Type	Name	Difficulty	Description
			over time with control limits (I-MR for individual values and moving range).
Visualization	Marginal distribution plot (with scatter)	Med 🟡	Scatter plot with additional histograms or density plots on the margins to show distributions of each variable.
Visualization	Parallel Coordinates plot	Med 🟡	Plot with multiple vertical axes (one per feature) and lines passing through them, representing observations across dimensions.
Visualization	Ridgeline plot (Joy plot)	Med 🟡	Multiple overlapping density plots, each for a subgroup, typically staggered vertically to compare distributions.
Visualization	Hexbin plot	Med 🟡	Two-dimensional histogram where counts of points are binned into hexagonal cells and color-coded by frequency.
Visualization	Contour plot	Med 🟡	Plot with contour lines to represent levels of a third variable over a 2D plane (like elevation or density).
Visualization	2D density plot	Med 🟡	Color- or shade-based representation of point density in a two-variable plot (a filled contour or heatmap of density).

Type	Name	Difficulty	Description
Visualization	Strip plot	Med 🟡	Plot showing individual data points for a categorical variable (often with jitter) to display distribution.
Visualization	Swarm plot (bee swarm)	Med 🟡	Like a strip plot but with points adjusted to avoid overlap, giving a sense of distribution.
Visualization	Rug plot	Med 🟡	Small tick marks along an axis indicating the positions of individual data points for a distribution.
Visualization	Correlation matrix plot	Med 🟡	Visual depiction of a correlation matrix, often using colored cells to represent correlation coefficients.
Visualization	Clustered heatmap (with dendrograms)	Med 🟡	Heatmap accompanied by dendrograms that show clustering of rows and columns based on similarity.
Visualization	Calendar heatmap	Med 🟡	Heatmap laid out as a calendar, often to visualize values by day (e.g., daily activity over a year).
Visualization	Horizon chart	Med 🟡	Dense time series visualization where data above a threshold is layered and colored to save space (horizon layering).
Visualization	Slope chart	Med 🟡	Chart showing change between two time points for

Type	Name	Difficulty	Description
			multiple categories (lines connecting two values per category).
Visualization	Dumbbell chart	Med 🟡	Shows two values for each category with circles connected by a line, highlighting the change or difference.
Visualization	Marimekko chart	Med 🟡	Mosaic chart with variable-width and variable-height rectangles, showing proportion in two categorical dimensions.
Visualization	Forest plot	Med 🟡	Display of estimates (e.g., effect sizes) and confidence intervals for multiple studies or categories, often used in meta-analysis.
Visualization	Manhattan plot	Med 🟡	Genomic visualization with points (e.g., SNP p-values) along chromosomes; looks like a skyline of significant signals.
Visualization	Volcano plot	Med 🟡	Scatter plot (often in genomics) of significance (p-value) vs effect size (fold-change), highlighting significant points.
Visualization	Bland-Altman plot	Med 🟡	Plot to compare two measurement methods by plotting the difference vs the average of the two methods.

Type	Name	Difficulty	Description
Visualization	ROC curve	Med 🟠	Graph of true positive rate vs false positive rate at various threshold settings for a classifier.
Visualization	Precision-Recall curve	Med 🟠	Plot of precision vs recall for different thresholds of a classifier, focusing on performance for positive class.
Visualization	Calibration plot	Med 🟠	Plot comparing predicted probabilities to actual observed frequencies of an outcome (model calibration).
Visualization	Lift chart	Med 🟠	Plot showing the performance of a classifier by comparing results sorted by model score to baseline random selection.
Visualization	Cumulative gain chart	Med 🟠	Cumulative version of lift chart showing the percentage of total positives captured as more of the population is targeted.
Visualization	Pareto chart	Med 🟠	Bar chart ordered by frequency combined with a line showing cumulative percentage, to identify the most important factors.
Visualization	Gantt chart	Med 🟠	Timeline chart showing project tasks on a calendar, with bars representing task

Type	Name	Difficulty	Description
			durations and start/end times.
Visualization	CUSUM chart	Med 🟡	Cumulative sum control chart for monitoring shifts in process mean over time.
Visualization	Run chart	Med 🟡	Line chart of data over time to show trends or shifts (basic control chart without control limits).
Visualization	Dot plot (Cleveland)	Med 🟡	Chart using dots to display values for different categories, focusing on position rather than bar length.
Visualization	Stem-and-leaf plot	Med 🟡	Textual plot where data values are split into a 'stem' (leading digit(s)) and 'leaf' (last digit), showing distribution.
Visualization	Interactive dashboard	Hard 🟢	A collection of interactive charts and UI elements allowing dynamic exploration of data.
Visualization	Network graph	Hard 🟢	Graph visualization with nodes and edges to illustrate relationships or connections among entities.
Visualization	Chord diagram	Hard 🟢	Circular diagram with arcs connecting different groups, showing flows or connections between categories.

Type	Name	Difficulty	Description
Visualization	Sunburst chart	Hard 🟢	Radial chart showing hierarchical data with nested arcs; similar to a tree map in circular form.
Visualization	3D scatter / surface plot	Hard 🟢	Plotting data in three dimensions, showing either points (scatter) or a continuous surface.
Visualization	Animated time-series plot	Hard 🟢	Time-series visualization that changes or plays out over time, often as an animation.
Visualization	Geospatial choropleth map	Hard 🟢	Map where geographic regions are colored based on data values (e.g., population density by area).
Visualization	Flow map / Streamgraph	Hard 🟢	Visualization of flows or changes: flow map shows movement on a map; streamgraph shows stacked area flows over time.
Visualization	Ternary plot	Hard 🟢	Triangle plot used for three components that sum to a constant (often percentages summing to 100%).
Visualization	Alluvial diagram	Hard 🟢	Flow diagram showing how categories in different dimensions or time periods split and merge (similar to Sankey but often for categorical changes).

Type	Name	Difficulty	Description
Visualization	Word cloud (if used for quantitative analysis)	Hard 🟢	Display of words with sizes proportional to some numeric value (like frequency or importance).
Visualization	Force-directed graph	Hard 🟢	Network layout where nodes repel and edges attract (like springs), resulting in an organic graph arrangement.
Visualization	Arc diagram	Hard 🟢	Network graph layout with nodes placed along a line and edges drawn as semicircles (arcs) between them.
Visualization	Adjacency matrix	Hard 🟢	Matrix representation of graph connections, with cell shading indicating presence or strength of connection between node pairs.
Visualization	Hierarchical edge bundling	Hard 🟢	Visualization of connections in hierarchical data where edges between leaves are drawn as bundled curves to reduce clutter.
Visualization	Dendrogram (complex hierarchical)	Hard 🟢	Tree diagram showing hierarchical relationships (often from clustering algorithms).
Visualization	Phylogenetic tree	Hard 🟢	Tree diagram depicting evolutionary relationships among species or entities (specialized dendrogram).

Type	Name	Difficulty	Description
Visualization	Voronoi diagram	Hard 🟢	Partition of a plane into regions based on distance to a set of seed points (each region contains all points closer to its seed than others).
Visualization	Cartogram	Hard 🟢	Map in which geographic areas are distorted so that their area reflects a data value (e.g., population).
Visualization	Dot density map	Hard 🟢	Map representing quantities by the number of dots placed in an area (each dot represents a certain quantity).
Visualization	Isopleth map	Hard 🟢	Map with contour lines (or color bands) connecting points of equal value (e.g., temperature or precipitation).
Visualization	Flow field visualization	Hard 🟢	Visualizing vector fields (like fluid flow or wind) often with arrows or streamlines showing direction and magnitude.
Visualization	Spiral plot	Hard 🟢	Plot where data (often time series) is arranged along a spiral, useful for finding periodic patterns.
Visualization	Radial tree	Hard 🟢	Tree layout arranged radially (root in center, leaves on outer circle), to visualize hierarchies.

Type	Name	Difficulty	Description
Visualization	Circle packing	Hard 🟢	Visualization of hierarchical data using nested circles, where circle size represents value.
Visualization	Bump chart (ranking over time)	Hard 🟢	Line chart tracking rankings of items over time, where lines bump into each other when ranks change.
Visualization	Connected scatterplot	Hard 🟢	Scatterplot where points are connected in some meaningful order (like time) to show trajectory or progression.
Visualization	3D heatmap / Volume rendering	Hard 🟢	Visualization of three-dimensional volumetric data, often using color/opacity to represent values in a 3D grid.
Visualization	Animated bubble chart (Gapminder-style)	Hard 🟢	Bubble chart that animates over time, showing how values change (e.g., Hans Rosling's Gapminder bubbles).
Visualization	Small multiples (faceted) with interactions	Hard 🟢	Series of similar charts (small multiples) allowing user interaction to filter or highlight data across them.
Visualization	Biplot (PCA visualization)	Hard 🟢	Plot showing both principal component scores of observations and the loading vectors of variables in PCA.

Type	Name	Difficulty	Description
Visualization	Scree plot with parallel analysis	Hard 🟢	Line chart of eigenvalues (variance explained by each principal component) with reference line to decide number of components to keep.
Visualization	Interactive 3D network	Hard 🟢	Network graph visualized in three dimensions, often user-rotatable and zoomable for exploration.
Visualization	Time-varying network animation	Hard 🟢	Animated visualization showing how the connections in a network change over time.
Visualization	Multivariate glyphs (Chernoff faces, star glyphs)	Hard 🟢	Displaying multivariate data using glyphs (like faces or stars) where different features map to parts of the glyph.