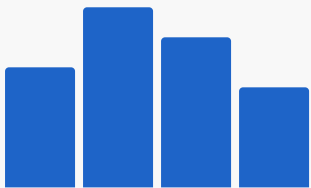


1D Data: Histogram & KDE

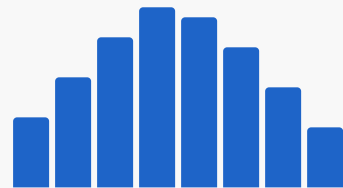
Understanding Distribution Through Binning and Smoothing

Few Bins



⚠ Loses detail

Optimal Bins



✓ Clear pattern

KDE Curve



✓ Smooth & continuous



Key Concepts

- Histogram: Binned frequency distribution
- KDE: Smooth density curve
- Bandwidth affects smoothness
- Identify: normal, skewed, bimodal
- Overlay multiple distributions



Bin Size Rules

Sturges' Rule: $\text{bins} \approx \log_2(n) + 1$

Freedman-Diaconis: Based on IQR & sample size

When to Use

- ✓ Exploring distribution
- ✓ Detecting outliers
- ✓ Understanding data spread
- ✓ Comparing distributions



How KDE (Kernel Density Estimation) Works

1

Place a kernel at each data point

2

Sum all kernels

3

Normalize the curve

$$\text{KDE}(\mathbf{x}) = \frac{1}{nh} \times \sum K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

n = number of data points

- A small bell curve (Gaussian) is centered at every observation

- Add up the contributions from all kernels at each position

- Scale so the total area under the curve equals 1

h = bandwidth (smoothing parameter)

K = kernel function (usually Gaussian)