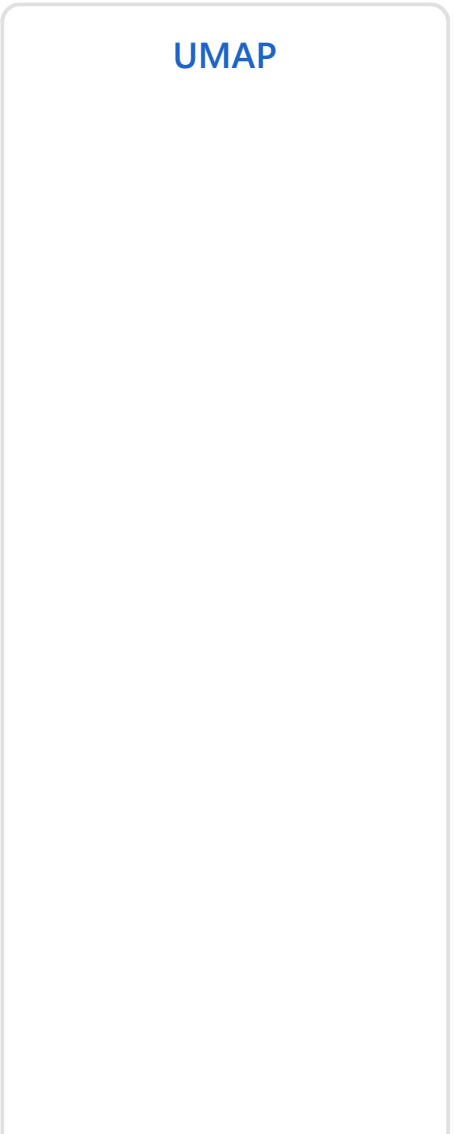
Dimensionality Reduction: PCA, t-SNE, UMAP

Projecting High-Dimensional Data to 2D

PCA

t-SNE

UMAP



Comparison

PCA

- Linear, interpretable
- Fast, deterministic

t-SNE

- Non-linear, local focus
- Perplexity: 5-50

UMAP

- Faster than t-SNE
- Local + global

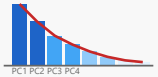
Cautions

- 2D projections lose info
- Can create false patterns
- Validate with multiple methods

✓ Best Practices

- ✓ Color by class labels
- ✓ Try continuous variables
- ✓ Combine with clustering
- ✓ Use scree plot for PCA

PCA Scree Plot: Explained Variance Ratio



Linear Projection

Non-linear

Balanced

Maximizes variance

Preserves local structure

Local + global structure