# What Should We Use?

# A Classification Tree

# Root Node

Do you know what you are looking for? No Yes

# Unsupervised Methods

Looking to substantial groups? Yes

Or for anomalies? Yes

# Searching for Groups

Use multidimensional scaling for $n$ up to a few hundred.

Otherwise use Kohonen's SOM.

Check out exploratory projection pursuit, optimization methods of clustering.

# Looking for Outliers

Use projection pursuit with a suitable index.

For up to a few hundred points, cross-check with non-metric MDS.

# Supervised Methods

Do you have to explain yourself? Yes

Or do you just want low error rates? Yes

# Need an Explanation

Consider linear methods (especially logistic regression)

or

classification trees.

## Issues

Nature of problem?

# Low-error-rate methods

Use Bayes' risk estimate to find out how hard the problem is.

Try linear methods to see how close they come to the lower bound.

Then see how much better non-linear methods can do:

Neural networks
k-nearest-neighbour methods
. . . .

## Issues

How competent is the user?

Speed, memory?

Nature of problem?