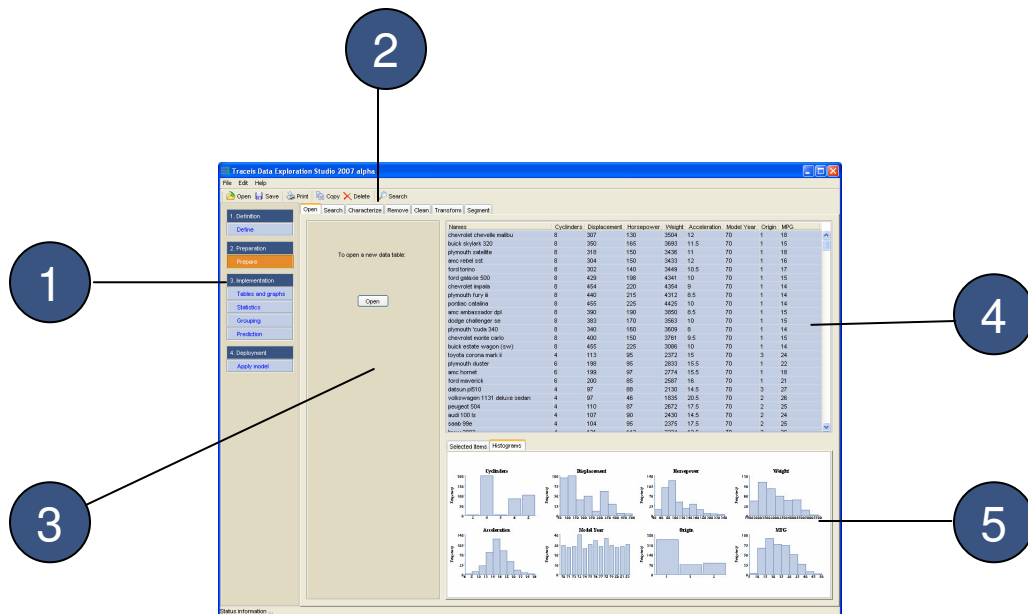


MAKING SENSE OF DATA

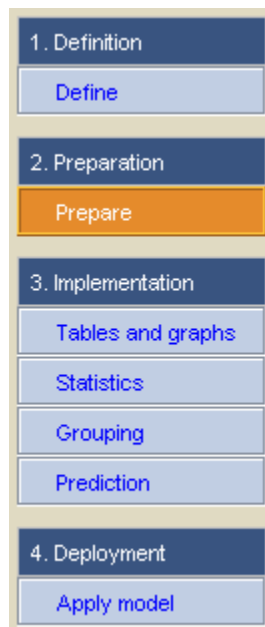
Traceis™ Data Exploration Studio Software Overview



The Traceis™ Data Exploration Studio is organized into five separate areas:

- 1 Process window
- 2 Analysis options
- 3 Tools window
- 4 Display window
- 5 Selected items window

MAKING SENSE OF DATA



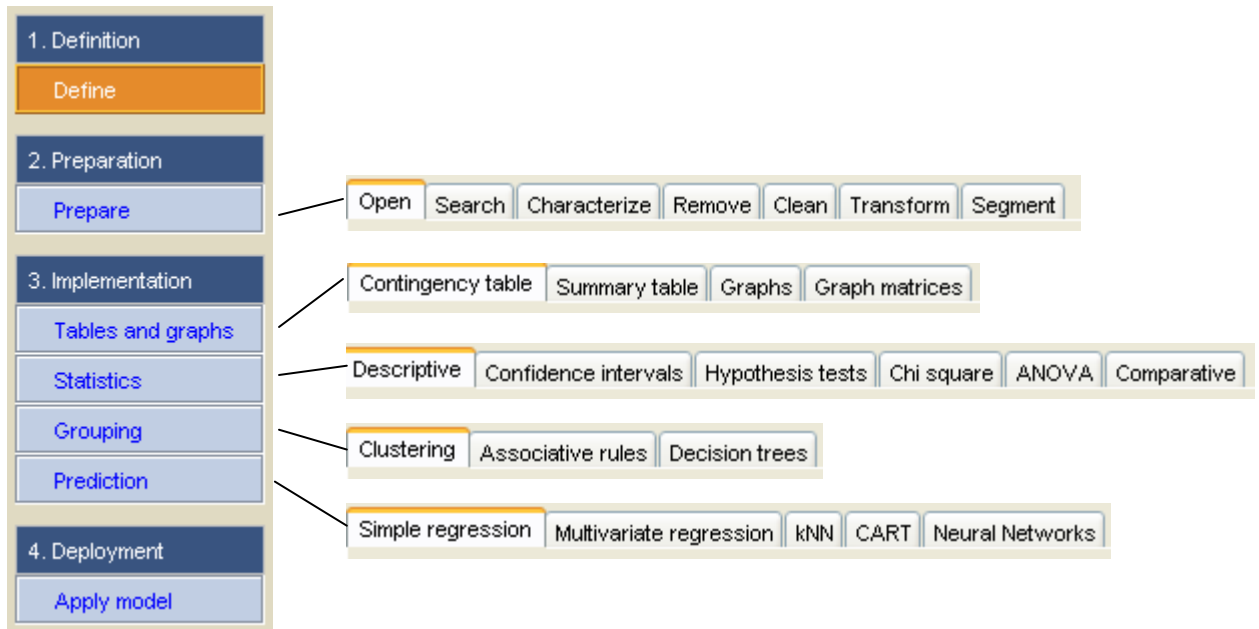
1 Process window

The book *Making Sense of Data* outlines a four step process for accomplishing any data analysis or data mining project. The software is organized according to these steps:

1. **Definition:** The goals should be defined prior to starting the project.
2. **Preparation:** The data should be prepared before implementing any analysis.
3. **Implementation:** Different methods for (1) visualizing the data, (2) generating statistics, (3) grouping and (4) making predictions are available.
4. **Deployment:** The results should be applied to the problem.

Selecting the light blue shaded options will alter the display according to the selected step. Selected items are shown in orange, as illustrated by the *Prepare* selection above.

MAKING SENSE OF DATA



2

Analysis options

Alternative ways of visualizing and manipulating the data can be accomplished according to the selected item in the process window:

- **Prepare:** (1) open a data table, (2) search using Boolean logic, (3) characterize the variables, (3) remove data or variables, (4) clean errors in the data, (5) apply mathematical transformations to the data, and (6) create subsets of the data.
- **Tables and graphs:** (1) contingency tables, (2) summary tables, (3) graphs (e.g. histograms or scatterplots) and (4) graphs matrices (e.g. scatterplot matrices).
- **Statistics:** (1) descriptive (e.g. mean and standard deviations), (2) confidence intervals, (3) hypothesis tests, (4) chi-square test, (5) one-way analysis of variance and (6) comparative statistics (e.g. r^2).
- **Grouping:** (1) hierarchical agglomerative and k-means clustering, (2) associative rules, and (3) decision trees.
- **Prediction:** (1) regression models, (2) k-nearest neighbors (kNN), (3) classification and regression trees (CART), and (4) neural networks.

MAKING SENSE OF DATA

The image displays three screenshots of a data analysis tool interface, likely for statistical software.

Left Screenshot: Select variables (descriptors):

- Names: A list of variables including Cylinders, Displacement, Horsepower, Weight, Acceleration, Model Year, Origin, and MPG. Cylinders is selected.
- Clustering methods: ☒ Agglomerative hierarchical, Linkage: Average, Number of groups: 10.
- Similarity methods: ☒ Euclidean, ☐ Jaccard.
- Display button.

Middle Screenshot: Search criteria

- Number of search terms: 3.
- Boolean operator: AND.
- Search criteria: Displacement > 150 AND Cylinders ≤ 8 AND MPG > 15.
- Save results as a new variable checkbox.
- Search results input field.
- Search button.

Right Screenshot: Display options

- Group on: Model Year.
- Show count checkbox (checked).
- Number of columns: 3.
- Column (1): Statistic: Minimum, Variable: MPG.
- Column (2): Statistic: Maximum, Variable: MPG.
- Column (3): Statistic: Mean.
- Display button.

3

Tools window

The tools window contains options and parameters for implementing and displaying the selected analysis, for example, defining the types of graphs to display or setting parameters for building a neural network. The results are shown in the display window.

MAKING SENSE OF DATA

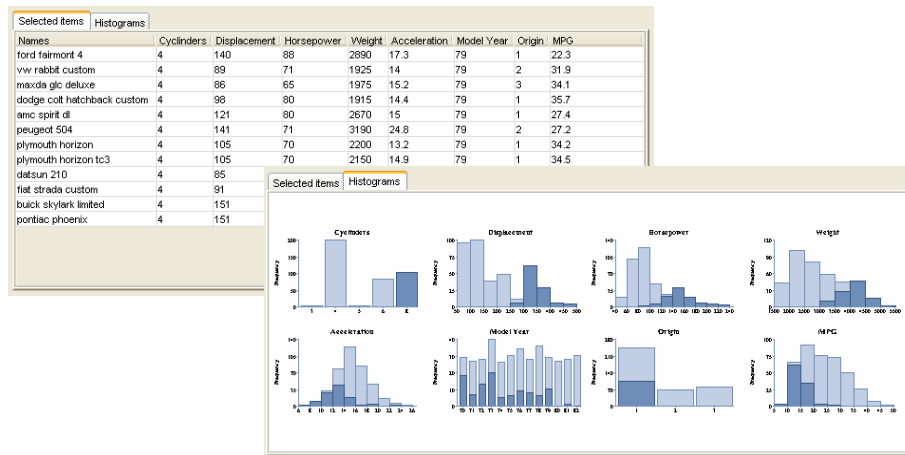


4

Display window

The display window is the where the selected analysis is displayed. This window presents, for example, hierarchical clustering, different graphs and statistical analysis. Often the window is interactive and any group or individual observation can be selected. The resulting selection is highlighted throughout the different analysis options. The items chosen are also presented in the selected items window.

MAKING SENSE OF DATA



5

Selected items window

The observations highlighted in the display window can be seen in the selected items window. There are two views: a table or histograms. The table presents details of the selection. The histograms present a frequency distribution of the variables and the selected observations are highlighted on the graphs.