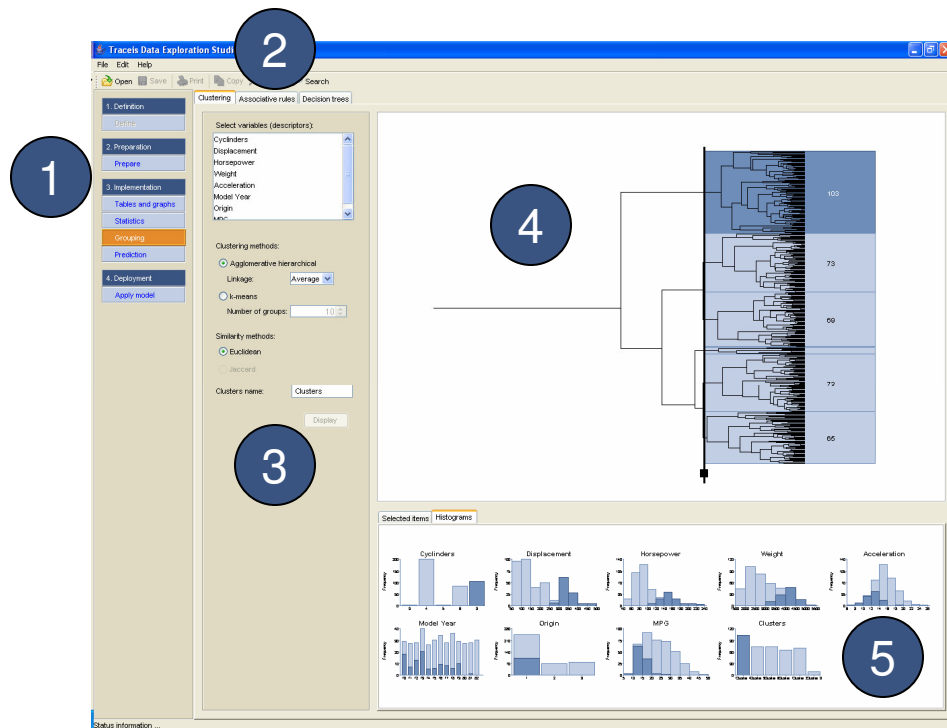


Traceis™ Data Exploration Studio Software Overview



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

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.

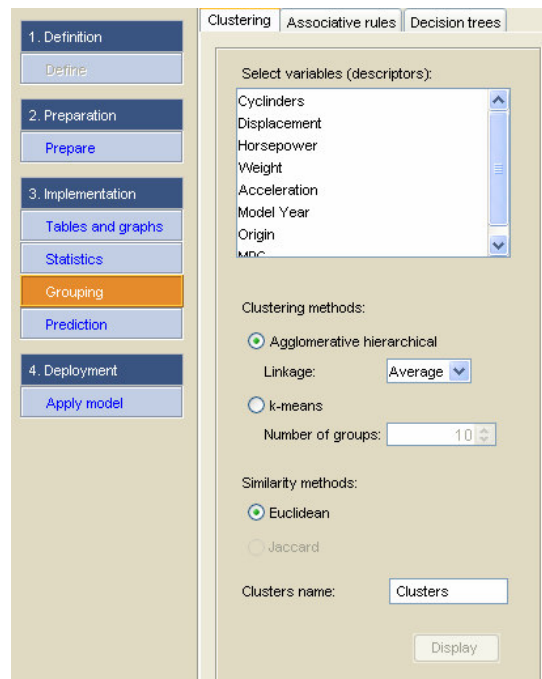


2

Analysis options

Alternative ways of visualizing and manipulating the data can be accomplished according 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.



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.

Number of search terms:

Boolean operator:

Cylinders	=	<input type="text" value="8"/>	AND
Weight	<	<input type="text" value="4000"/>	AND
Acceleration	=	<input type="text" value="11"/>	

Save results as a new variable

Search results:

Select variables (descriptors):

- Cylinders
- Displacement
- Horsepower
- Weight
- Acceleration
- Model Year
- Origin
- MPG

Clustering methods:

Agglomerative hierarchical

Linkage:

k-means

Number of groups:

Similarity methods:

Euclidean

Jaccard

Clusters name:

Select variables (descriptors):

- Cylinders
- Displacement
- Horsepower
- Weight
- Acceleration
- Model Year
- Origin
- MPG

Select variables (response):

Number of cycles (epochs):

Number of hidden layers:

Learning rate:

Cross-validation percentage:

Prediction name:

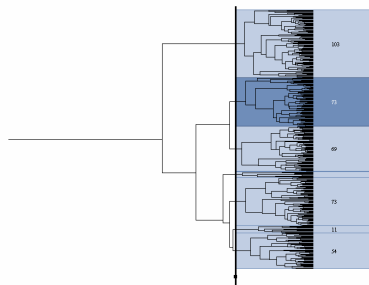
Residual name:

MAKING SENSE OF DATA

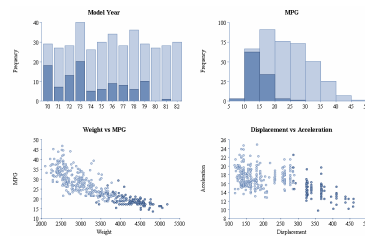
4 Display window

The display window is 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 windows. The items chosen are also presented in the selected items window.

All observations	
Number of observations	392
Central tendency	
Mode	13.0
Median	22.8
Mean	23.4
Variation	
Minimum	9
Maximum	46.8
Quartiles	Q1: 17.02, 22.8, Q3: 29
Variance	60.9
Standard deviation	7.8
Shape	
Skewness	0.5
Kurtosis	-0.5



Name	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model Year	Origin	MPG
chevrolet chevelle malibu	8	307	130	3504	12	70	1	18
buick wildcat 350	8	290	165	3690	11.5	70	1	15
plymouth volvo	8	318	150	3436	11	70	1	18
amc rebel sst	8	302	140	3441	10.5	70	1	17
ford torino 302	8	304	160	3426	12	70	1	16
chevrolet monte carlo	8	454	225	4354	9	70	1	14
jeep van 4	8	440	225	4212	9.5	70	1	14
ford torino	8	405	225	4425	10	70	1	14
amc Ambassador	8	360	190	3665	8	70	1	16
volvo 460 gle	8	303	170	3562	10	70	1	15
plymouth coupe 160	8	340	160	3600	8	70	1	14
chevrolet corvair 4	4	400	150	3761	9.5	70	1	15
ford torino 350	8	450	225	4366	10	70	1	14
toyota corolla 4	4	113	95	2932	15	70	3	24
toyota corolla 3	4	112	95	2773	15.5	70	3	27
toyota corolla 2	4	112	95	2702	15.5	70	3	27
datsun 280z	4	275	150	3111	15.5	70	2	26
datsun 280sl	4	275	150	3111	15.5	70	2	27
volvo 460 gcr	4	81	45	1825	20.5	70	2	28
volvo 460 gl	4	110	87	3072	17.5	70	2	25
saab 900	4	137	90	2426	14.5	70	2	24
saab 900le	4	104	95	2231	17.5	70	2	25
volvo 460	4	104	95	2231	17.5	70	2	25



Cylinders	Model Year												Totals	
	70	71	72	73	74	75	76	77	78	79	80	81		82
3	0	0	1	1	0	0	0	1	0	0	1	0	0	4
4	7	12	14	11	15	12	15	14	17	12	23	20	27	198
5	0	0	0	0	0	0	0	0	1	1	1	1	0	3
6	4	8	0	8	6	12	10	5	12	6	2	7	3	83
8	18	7	13	20	5	8	9	8	6	10	0	1	0	103
Totals	29	27	28	40	26	30	34	28	36	29	27	28	30	392

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 in dark blue.

Names	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model Year	Origin	MPG
plymouth 'cuda 340	8	340	160	3609	8	70	1	14
buick estate wagon (sw)	8	455	225	3086	10	70	1	14
amc hornet	6	199	97	2774	15.5	70	1	18
dodge d200	8	318	210	4382	13.5	70	1	11
plymouth satellite custom	6	225	105	3439	15.5	71	1	16
ford galaxie 500	8	351	153	4154	13.5	71	1	14
dodge monaco (sw)	8	383	180	4955	11.5	71	1	12
ford country squire (sw)	8	400	170	4746	12	71	1	13
chevrolet vega (sw)	4	140	72	2408	19	71	1	22
plymouth cricket	4	91	70	1955	20.5	71	1	26
toyota corona hardtop	4	113	95	2278	15.5	72	3	24
chevrolet vega	4	140	90	2408	19.5	72	1	20
chevrolet impala	8	350	165	4274	12	72	1	13
pontiac catalina	8	400	175	4385	12	72	1	14
buick lesabre custom	8	350	155	4502	13.5	72	1	13
oldsmobile delta 88 royale	8	350	160	4456	13.5	72	1	12

