What's New in Maple 2016

Maple™

Statistics

Linear Regression

All linear regression commands have been updated in Maple 2016 with a new option, summarize, that allows for the display of a summary for the given regression model.

```
with(Statistics) :
X := Vector([1, 2, 3, 4, 5, 6], datatype = float) :
Y := Vector([2, 3, 4, 3.5, 5.8, 7], datatype = float) :
```

By default, the Fit command returns the resulting regression model for the given model function:

```
Fit(a + b * t + c * t^2, X, Y, t);
```

$$1.96000000000000 + 0.164999999999999\, t + 0.110714285714286\, t^2$$

The summarize option includes a full summary for each of the regression coefficients, as well as values for the r-squared and adjusted r-squared for the model. Also, the solution module for regression commands has been extended with the ability to return values for r-squared, adjusted r-squared, and the value of the t-statistic for testing whether the corresponding regression coefficient is different than 0 and its corresponding probability.

```
Fit(a + b * t + c * t^2, X, Y, t, summarize = true) :
```

```
Summary:
----------------
Model: 1.9600000+.16500000*t+.11071429*t^2
----------------
Coefficients:
    Estimate  Std. Error  t-value  P(>|t|)
a    1.9600    1.1720      1.6724   0.1930
```

```
b    0.1650    0.7667      0.2152   0.8434
c    0.1107    0.1072      1.0325   0.3778
----------------
R-squared: 0.9252, Adjusted R-squared: 0.8753
```

The summarize option can also be used to return an embedded table, which contains more details on the residuals:

$\mathit{Fit}(a + b * t + c * t{}^\wedge 2, X, Y, t, \boldsymbol{\mathit{summarize}} = \mathit{embed})$:

Summary

Model: $1.9600000 + 0.16500000\,t + 0.11071429\,t^2$

Coefficients	Estimate	Standard Error	t-value	P(>|t|)
a	1.96000	1.17199	1.67237	0.193045
b	0.165000	0.766748	0.215194	0.843415
c	0.110714	0.107226	1.03253	0.377769

R-squared: 0.925169

Adjusted R-squared: 0.875282

▼ Residuals

Residual Sum of Squares	Residual Mean Square	Residual Standard Error	Degrees of Freedom
1.28771	0.429238	0.655163	3

Five Point Summary

Minimum	First Quartile	Median	Third Quartile	Maximum
-0.891429	-0.290357	0.155714	0.290595	0.548571

▼ Hypothesis Testing

The summarize option has also been added to all hypothesis testing commands. Previously, the infolevel command would have been required to print the results of a hypothesis test as a report.

```
with(Statistics) :
X := Array([9, 10, 8, 4, 8, 3, 0, 10, 15, 9]) :
```

$OneSampleChiSquareTest(X, 7, confidence = .95, summarize = embed)$:

Chi-Square Test on One Sample

<table>
<tr><td>Null Hypothesis:</td><td>Sample drawn from population with standard deviation equal to 7</td></tr>
<tr><td>Alternative Hypothesis:</td><td>Sample drawn from population with standard deviation not equal to 7</td></tr>
<tr><td colspan="2">
<table>
<tr><th>Sample Size</th><th>Sample Standard Deviation</th><th>Distribution</th><th>Computed Statistic</th><th>Computed p-value</th><th>Confidence Interval</th></tr>
<tr><td>10.</td><td>4.24788</td><td><i>ChiSquare</i>(9)</td><td>3.31429</td><td>0.0989571</td><td>2.92184 ..7.75496</td></tr>
</table>
</td></tr>
<tr><td>Result:</td><td><b>Accepted</b>: This statistical test does not provide enough evidence to conclude that the null hypothesis is false.</td></tr>
</table>

Summary and Tabulation

The DataSummary, FivePointSummary, and FrequencyTable commands can also accept a summarize option as well as be used to return summary statistics for DataFrames:

$with(Statistics)$:

$X := DataFrame(\langle Sample(Uniform(0, 1), [50, 2]) | LinearAlgebra\text{:-}RandomVector(50, generator = rand(0 ..3))\rangle)$

$$\left[\begin{array}{cccc} & 1 & 2 & 3 \\ 1 & 0.814723686393179 & 0.276025076998578 & 3 \\ 2 & 0.905791937075619 & 0.679702676853675 & 2 \\ 3 & 0.126986816293506 & 0.655098003973841 & 1 \\ 4 & 0.913375856139019 & 0.162611735194631 & 2 \\ 5 & 0.632359246225410 & 0.118997681558377 & 1 \\ 6 & 0.0975404049994095 & 0.498364051982143 & 0 \\ 7 & 0.278498218867048 & 0.959743958516081 & 1 \\ 8 & 0.546881519204984 & 0.340385726666133 & 0 \\ \ldots & \ldots & \ldots & \ldots \end{array}\right]$$

$DataSummary(X, summarize = embed)$:

	1	2	3
mean	0.5661101110386353	0.48987882283825	1.22
standarddeviation	0.3125414653315035	0.2785520677947388\4	1.1830434635864795
skewness	-0.371651925679829\9	0.1158245251557512\3	0.3106598115832435
kurtosis	1.729650721370971	1.7837876411157099	1.5501229720154914
minimum	0.0318328463774206\76	0.0119020695012413\97	0.0
maximum	0.9705927817606157	0.9597439585160811	3.0
cumulativeweight	50.0	50.0	50.0

Visualizations

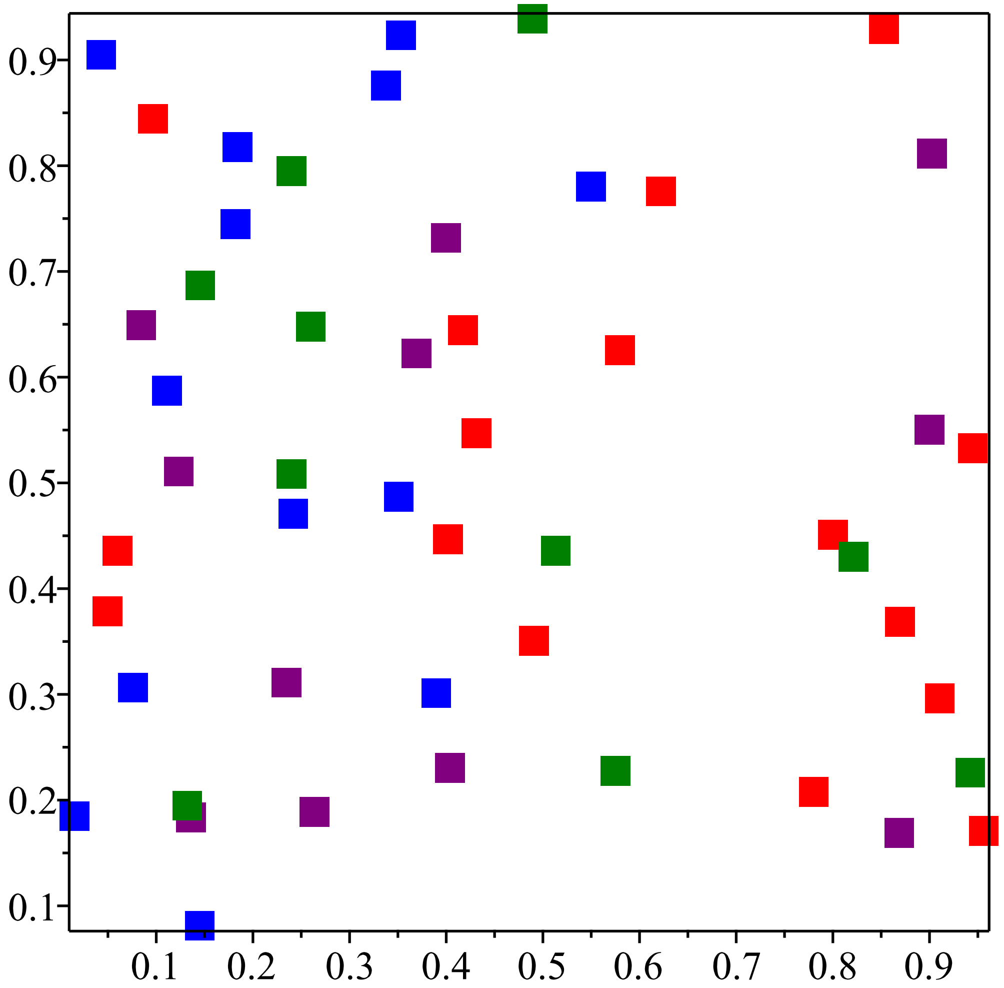
There are many new visualizations in Maple 2016 for statistics and data analysis, including new options for creating colorschemes using point values:

```
with(Statistics) :

data := Matrix(⟨Sample(Uniform(0, 1), [50, 2]) | LinearAlgebra:-RandomVector(50, generator
      = rand(0 ..3))⟩)
```

```
50 x 3 Matrix
Data Type: anything
Storage: rectangular
Order: Fortran_order
```

```
ScatterPlot(data[ .., 1], data[ .., 2], symbolsize = 20, symbol = solidbox,
            colorscheme = ["valuesplit", data[ .., 3], [0 = "Red", 1 = "Blue", 2 = "Green", 3
      = "Purple"]])
```



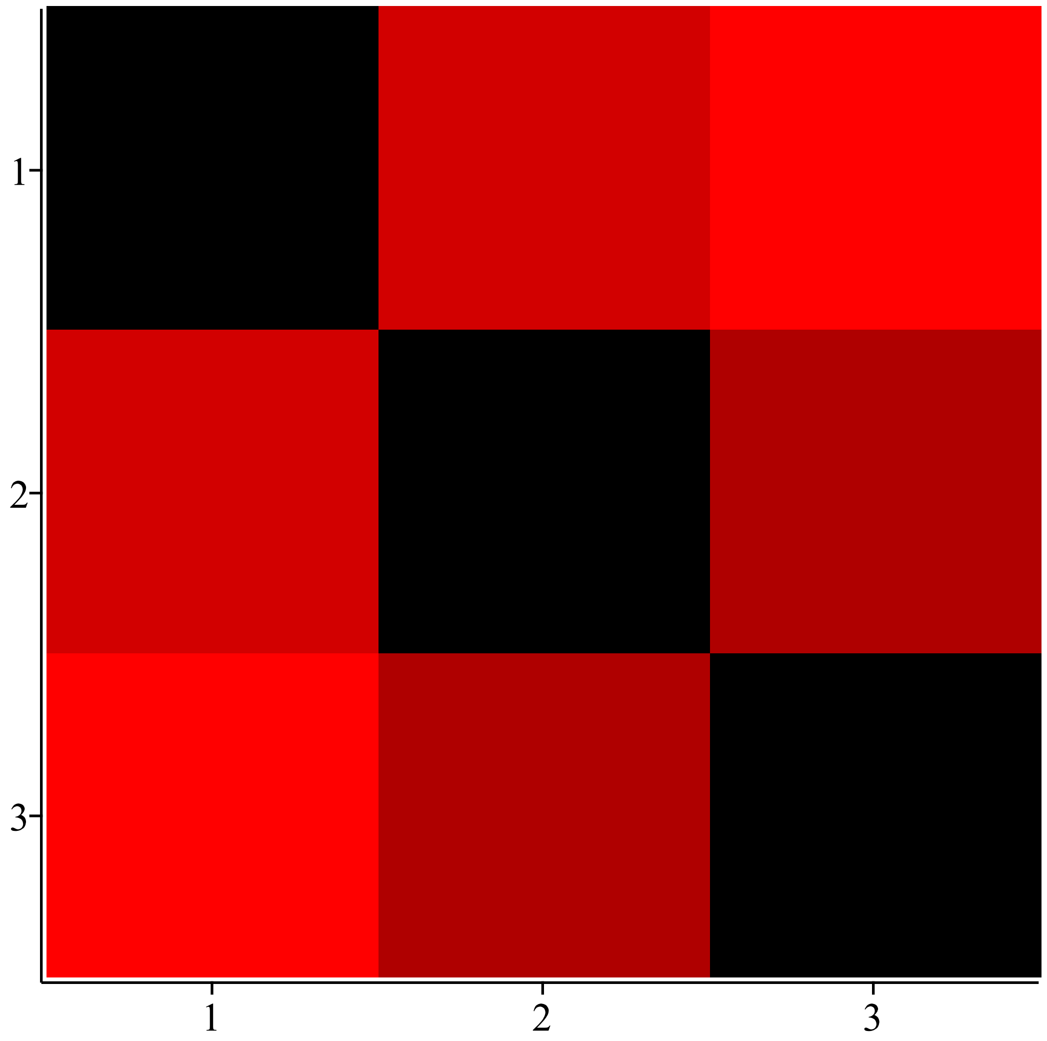
Maple 2016 also introduces a new visualization in Statistics for generating heat maps. A heat map is a visualization method that represents the magnitude of the included data as a discrete density plot.

$U := \langle seq(0..10)\rangle :$
$V := \langle seq(\sin(i), i=0..10)\rangle :$
$W := \langle seq(\cos(i), i=0..10)\rangle :$

$CM := CorrelationMatrix(Matrix([U, V, W]), ignore);$

$$\begin{bmatrix} 1. & -0.116741765101327 & -0.354242435116850 \\ -0.116741765101327 & 1. & 0.0701662833954325 \\ -0.354242435116850 & 0.0701662833954325 & 1. \end{bmatrix}$$

$HeatMap(CM, color = [\text{"Red"}, \text{"Black"}]);$

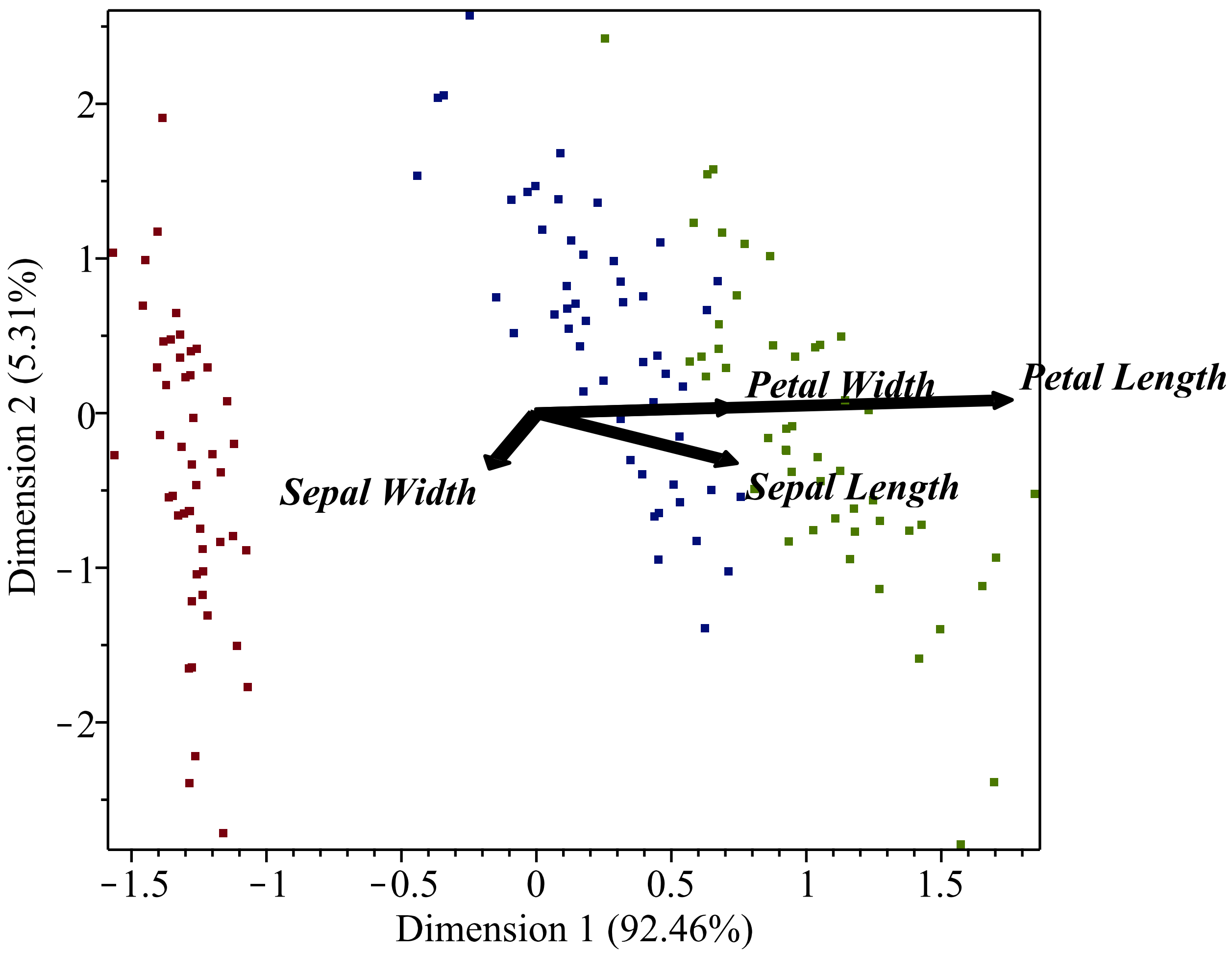


There are also two new visualizations related to Principal Component Analysis: Biplot, and ScreePlot.

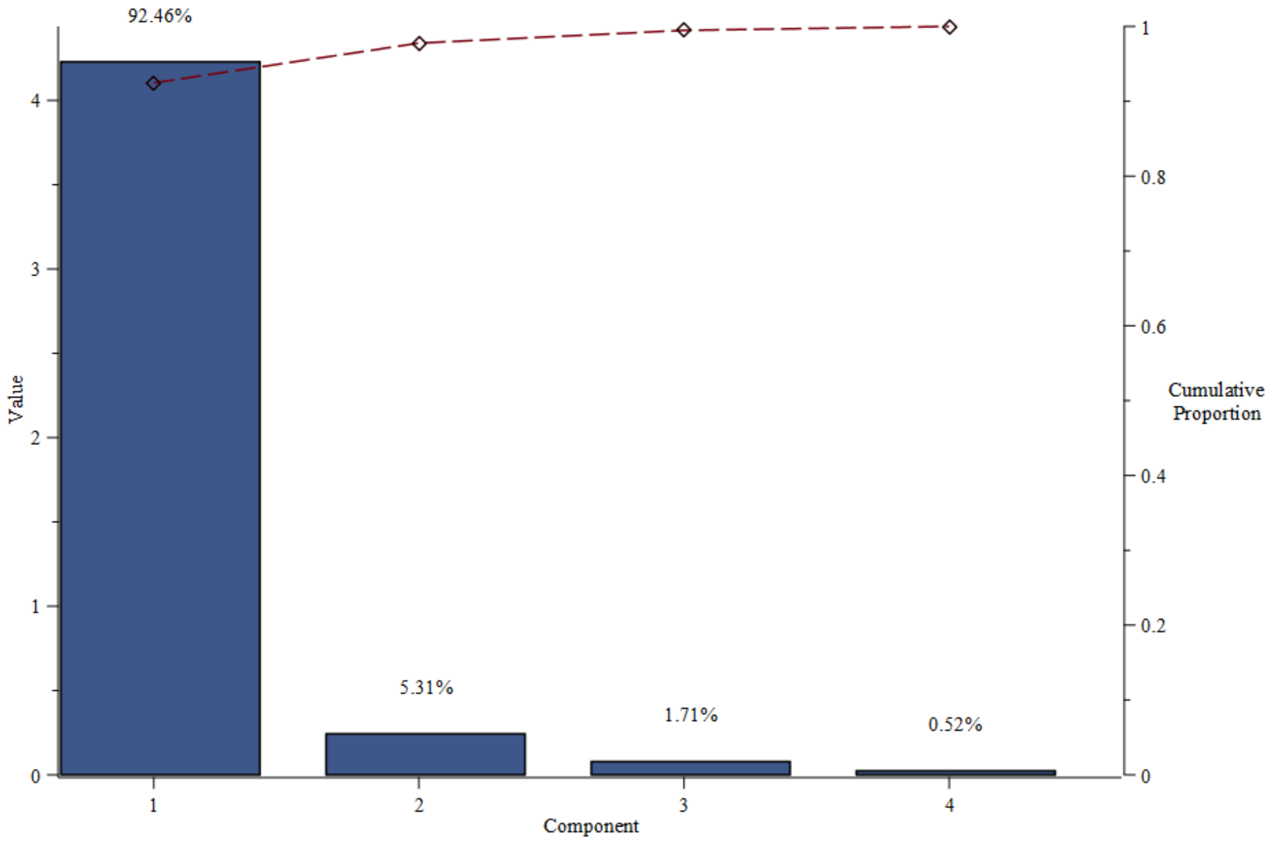
```
IrisDF := Import(FileTools:-JoinPath(["datasets", "iris.csv"], base = datadir))
```

	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
1	5.1	3.5	1.4	0.2	"setosa"
2	4.9	3	1.4	0.2	"setosa"
3	4.7	3.2	1.3	0.2	"setosa"
4	4.6	3.1	1.5	0.2	"setosa"
5	5	3.6	1.4	0.2	"setosa"
6	5.4	3.9	1.7	0.4	"setosa"
7	4.6	3.4	1.4	0.3	"setosa"
8	5	3.4	1.5	0.2	"setosa"
...	...	...	...	...	...

```
Biplot(IrisDF[[`Sepal Length`, `Sepal Width`, `Petal Length`, `Petal Width`]],
colorscheme = ["valuesplit", IrisDF[`Species`]], size = [600, "golden"])
```

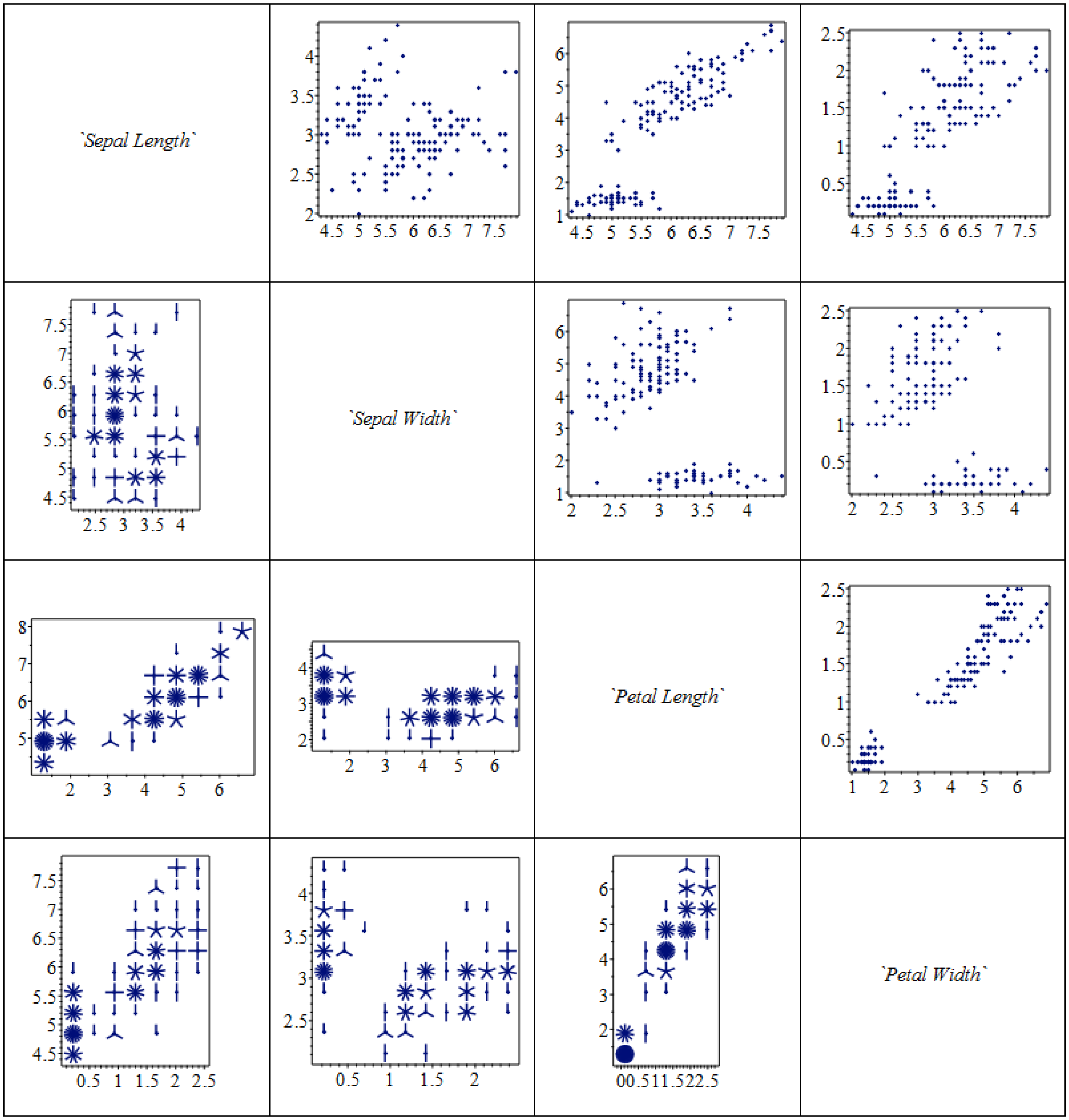


```
ScreePlot(IrisDF[[`Sepal Length`, `Sepal Width`, `Petal Length`, `Petal Width`]])
```



The new GridPlot command is useful for visualizing multidimensional datasets. GridPlot generates a matrix of plots corresponding to the columns of a dataset.

```
GridPlot(IrisDF[[`Sepal Length`, `Sepal Width`, `Petal Length`, `Petal Width`]], upper
    = ScatterPlot, lower = SunflowerPlot, width = 600, widthmode = pixels);
```


`Sepal Length`
`Sepal Width`
`Petal Length`
`Petal Width`
2 2.5 3 3.5 4
4.5 5 5.5 6 6.5 7 7.5
1 2 3 4 5 6
0.5 1 1.5 2 2.5
8 7 6 5
2 3 4 5 6
00.511.522.5