


Rattle


A Data Mining GUI for R

安裝、載入 Rattle

安 裝

```
Console ~/   
> install.packages("rattle")
```

載 入

```
Console ~/   
> library(rattle)  
Rattle: A free graphical interface for data mining with R.  
Version 4.1.0 Copyright (c) 2006-2015 Togaware Pty Ltd.  
Type 'rattle()' to shake, rattle, and roll your data.  
> rattle()  
>
```

Rattle

The screenshot displays the Rattle software interface. At the top is a menu bar with 'Project', 'Tools', 'Settings', and 'Help'. Below this is a toolbar with icons for 'Execute' (gears), 'New' (document), 'Open' (folder), 'Save' (floppy disk), 'Report' (document with lines), 'Export' (key), 'Stop' (red X), 'Quit' (red arrow), and 'Connect R' (R logo). A status bar on the right indicates 'Rattle Version 4.1.0 togaware.com'. The main workspace has a tabbed interface with 'Data' selected. Below the tabs are radio buttons for data sources: 'Spreadsheet' (selected), 'ARFF', 'ODBC', 'R Dataset', 'RData File', 'Library', 'Corpus', and 'Script'. The 'Filename' field is set to '(None)' with a folder icon. The 'Separator' is a comma, 'Decimal' is a period, and 'Header' is checked. There are checkboxes for 'Partition' and 'Seed' (set to 42), with 'View' and 'Edit' buttons. At the bottom left are 'Input' (green circle) and 'Ignore' (red circle) buttons, followed by a 'Weight Calculator' field. On the right is a 'Target Data Type' section with radio buttons for 'Auto' (selected), 'Categoric', 'Numeric', and 'Survival'. The bottom panel contains a welcome message and introductory text about Rattle and R.

Project Tools Settings Help Rattle Version 4.1.0 togaware.com

Execute New Open Save Report Export Stop Quit Connect R

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source: ☒ Spreadsheet ☐ ARFF ☐ ODBC ☐ R Dataset ☐ RData File ☐ Library ☐ Corpus ☐ Script

Filename: (None) Separator: , Decimal: . ☒ Header

☐ Partition 70/15/15 Seed: 42 View Edit

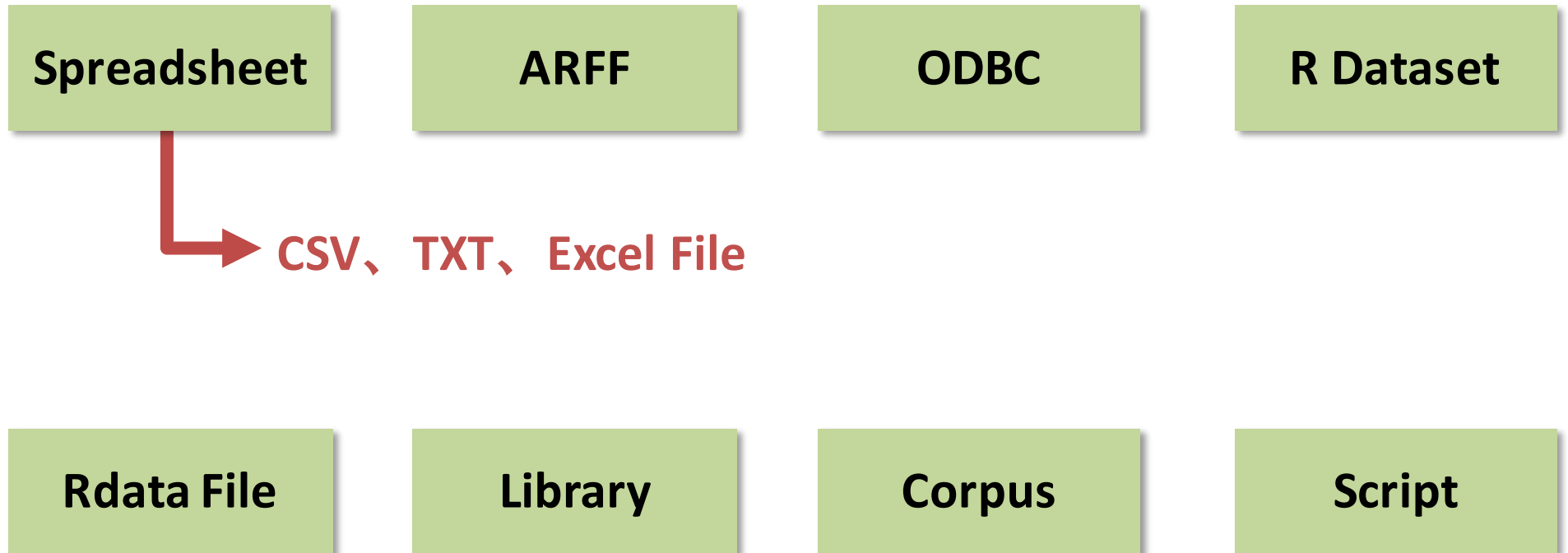
☒ Input ☐ Ignore Weight Calculator: Target Data Type: ☒ Auto ☐ Categoric ☐ Numeric ☐ Survival

Welcome to Rattle (rattle.togaware.com).

Rattle is a free graphical user interface for Data Mining, developed using R. R is a free software environment for statistical computing and graphics. Together they provide a sophisticated environments for data mining, statistical analyses, and data visualisation.

See the Help menu for extensive support in using Rattle. The book Data Mining with Rattle and R is available from Amazon. The Togaware Desktop Data Mining Survival Guide includes Rattle documentation and is available from datamining.togaware.com

Import Data 載入數據



Import Data – Spreadsheet

R Data Miner - [Rattle]

Project Tools Settings Help Rattle Version 4.1.

Execute New Open Save Report Export Stop Quit Connect R

Data | Explore | Test | Transform | Cluster | Associate | Model | Evaluate | Log

Source: ☒ Spreadsheet ☐ ARFF ☐ ODBC ☐ R Dataset ☐ RData File ☐ Library ☐ Corpus ☐ Script

Filename: Separator: Decimal: ☒ Header

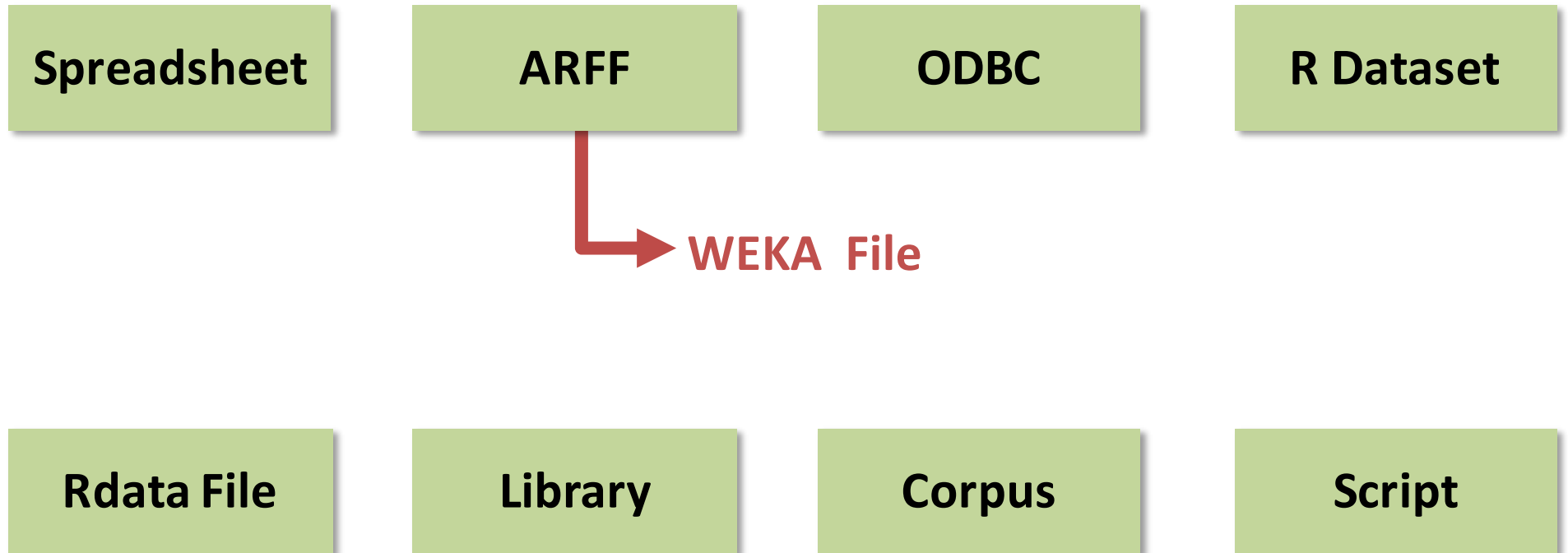
☐ Partition Seed:

☒ Input ☒ Ignore Weight Calculator:

Target Data Type
☒ Auto ☐ Categorical ☐ Numeric ☐ Survival

Welcome to Rattle (rattle.togaware.com).

Import Data 載入數據



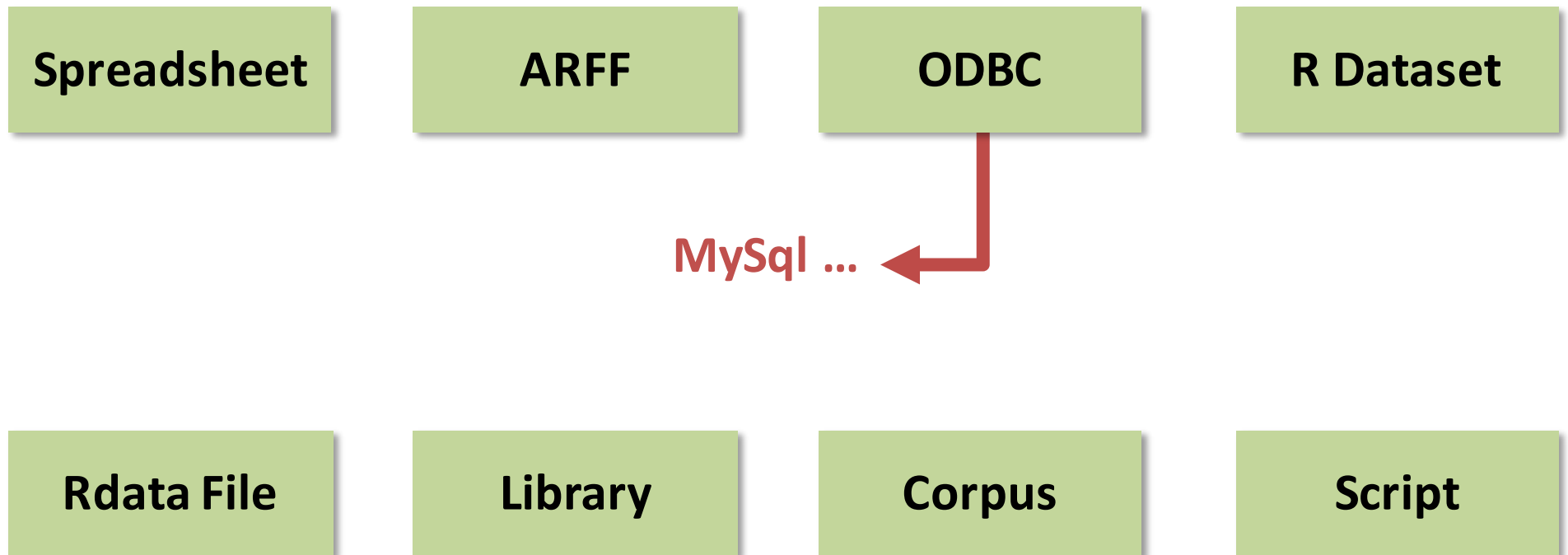
Weka

Weka (Waikato Environment for Knowledge Analysis)

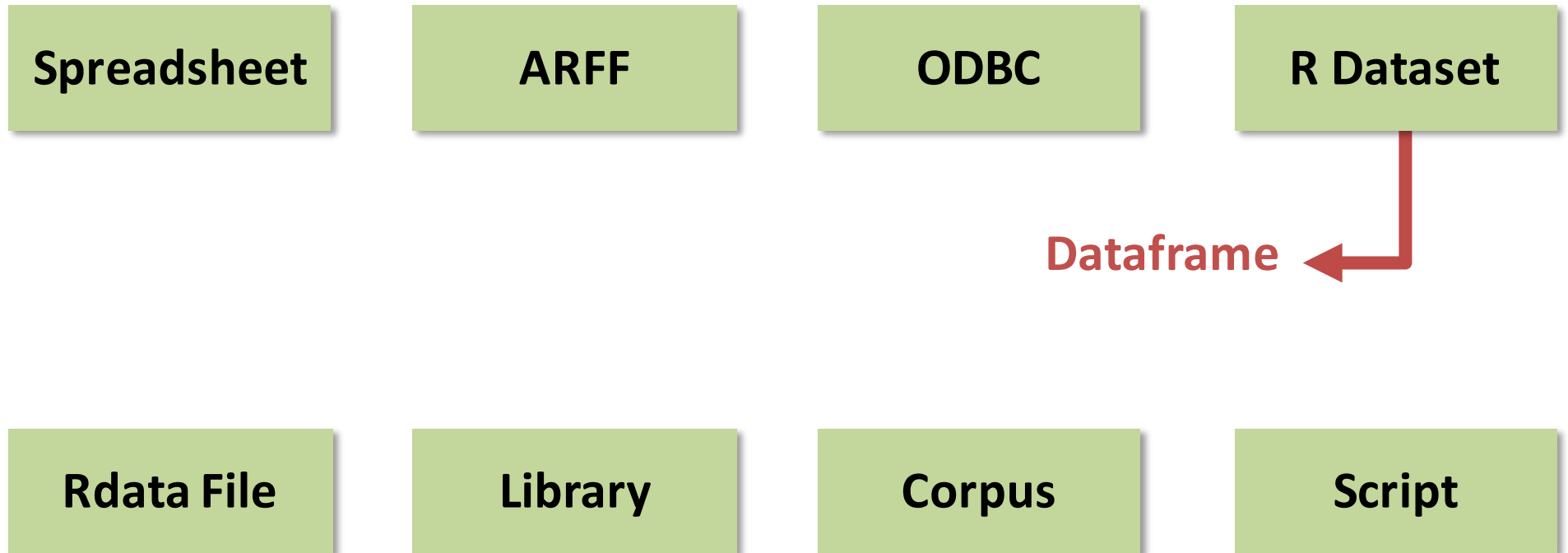
提供 data mining 以及 machine learning 各種工具的一套軟體
包括資料前置處理工具、分類工具、回歸分析等，也能將資料以視覺畫的方式呈現。



Import Data 載入數據



Import Data 載入數據



Import Data – R Dataset

The screenshot displays the RStudio environment with the Rattle application open. The Rattle window, titled 'R Data Miner - [Rattle]', shows the 'Data' tab selected. The 'Source' section has radio buttons for 'Spreadsheet', 'ARFF', 'ODBC', 'R Dataset' (selected), 'RData File', 'Library', 'Corpus', and 'Script'. A red arrow points to the 'R Dataset' option. Below this, the 'Data Name' dropdown menu is set to 'taipei_pm25' and is highlighted with a red box. Other options include 'Partition' (70/15/15), 'Seed' (42), and 'Target Data Type' (Auto, Categorical, Numeric, Survival). The bottom of the Rattle window contains a welcome message and documentation links.

Environment History
Import Dataset ▾ List ▾
Global Environment ▾
Data
taipei_... 128 obs. of 16 var...

Files Plots Packages Help Viewer
New Folder Delete Rename More ▾
Home > Desktop > R
Name Size Modified
..
.Rhistorv 41 B Apr 20, 2016, ...

Import Data 載入數據

Spreadsheet

ARFF

ODBC

R Dataset

Rdata File

Library

Corpus

Script



Import Data 載入數據

Spreadsheet

ARFF

ODBC

R Dataset

Rdata File

Library

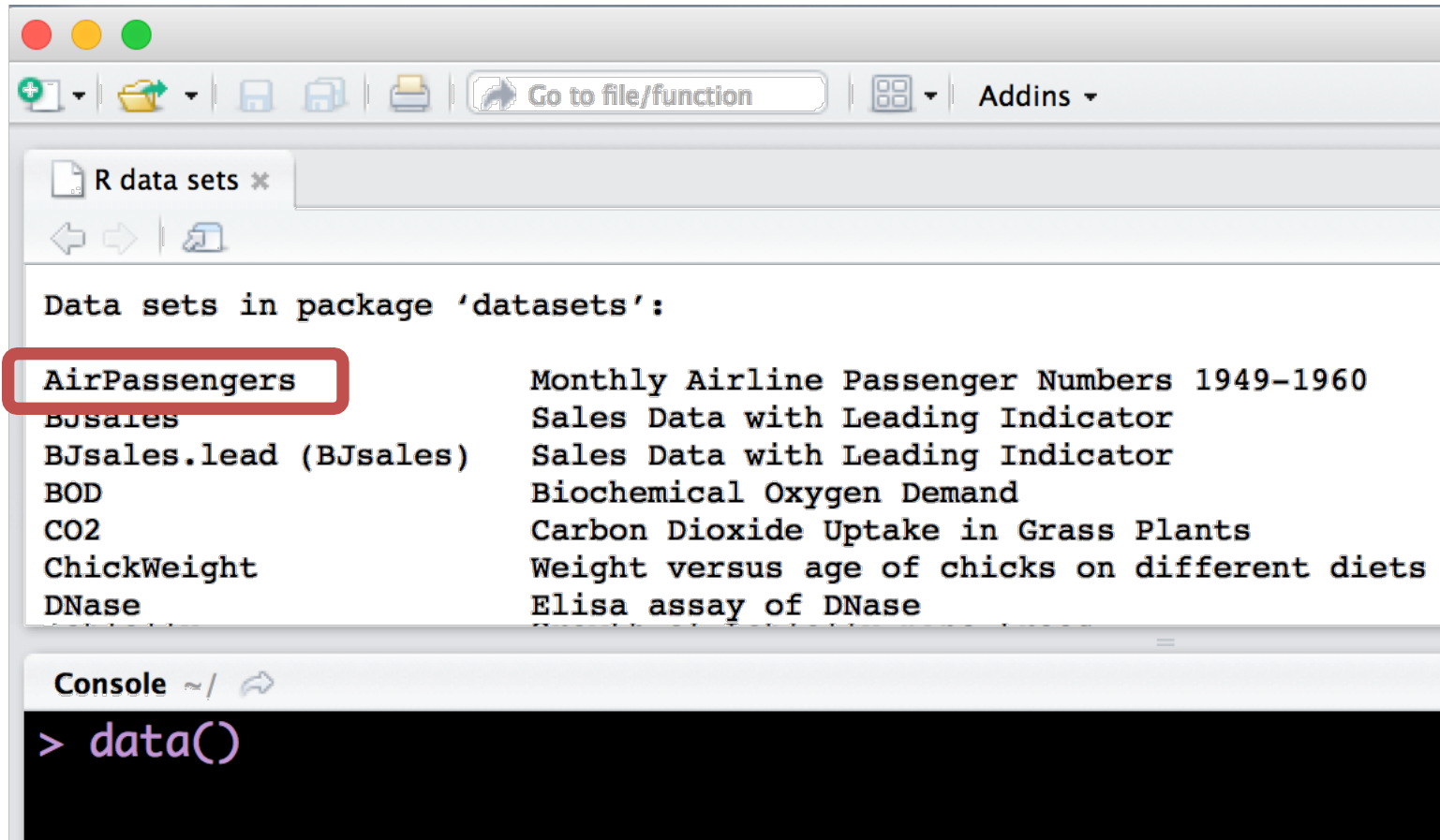
Corpus

Script



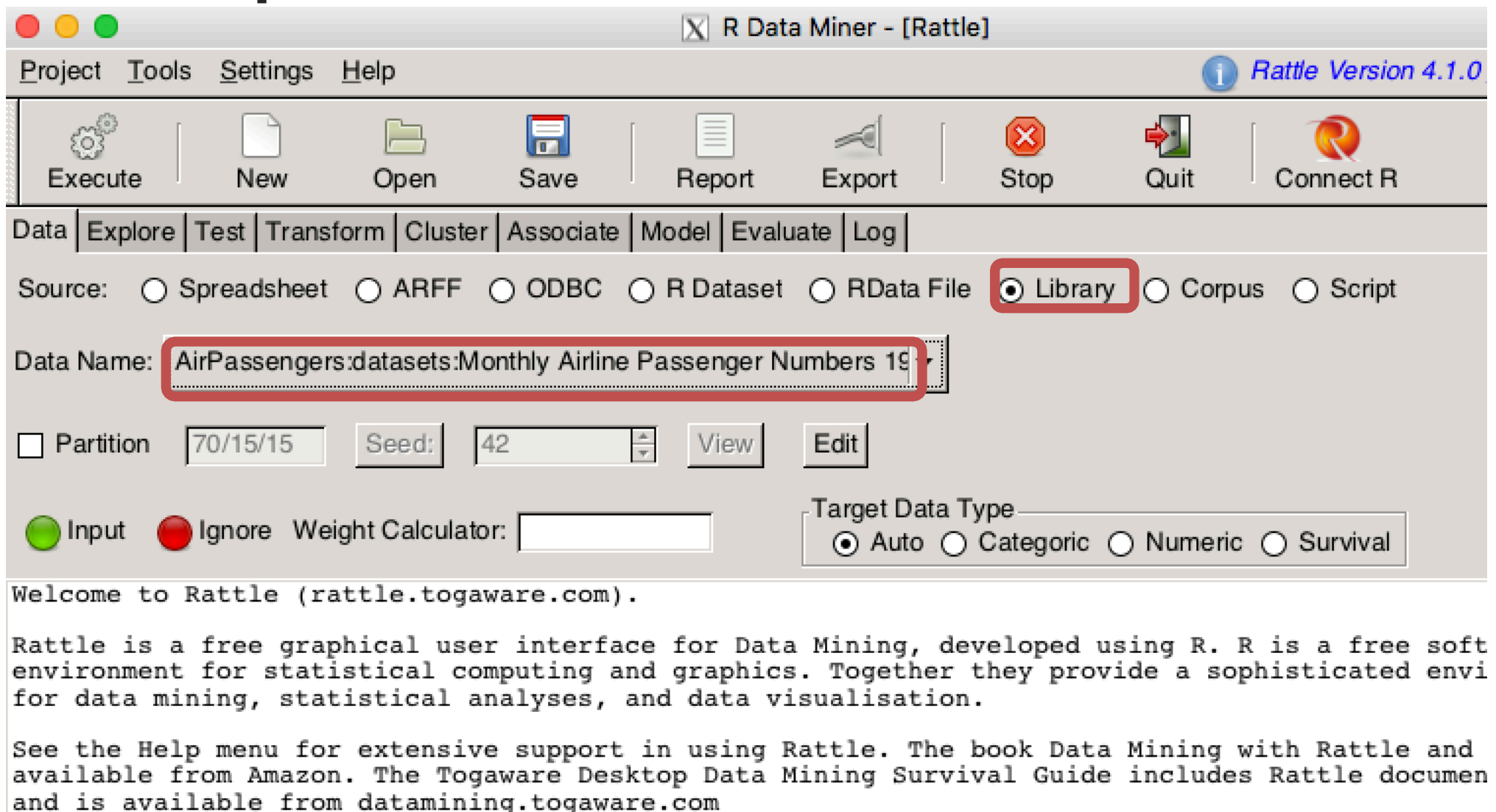
套件中的資料

看目前可用資料集



```
R data sets ✕  
Data sets in package 'datasets':  
AirPassengers      Monthly Airline Passenger Numbers 1949-1960  
BJsales            Sales Data with Leading Indicator  
BJsales.lead (BJsales) Sales Data with Leading Indicator  
BOD                Biochemical Oxygen Demand  
CO2                Carbon Dioxide Uptake in Grass Plants  
ChickWeight        Weight versus age of chicks on different diets  
DNase              Elisa assay of DNase  
> data()
```

Import Data – R Dataset



R Data Miner - [Rattle]

Project Tools Settings Help Rattle Version 4.1.0

Execute New Open Save Report Export Stop Quit Connect R

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source: ☐ Spreadsheet ☐ ARFF ☐ ODBC ☐ R Dataset ☐ RData File ☒ Library ☐ Corpus ☐ Script

Data Name: AirPassengers:datasets:Monthly Airline Passenger Numbers 19

☐ Partition 70/15/15 Seed: 42 View Edit

☒ Input ☐ Ignore Weight Calculator: Target Data Type: ☒ Auto ☐ Categorical ☐ Numeric ☐ Survival

Welcome to Rattle (rattle.togaware.com).

Rattle is a free graphical user interface for Data Mining, developed using R. R is a free soft environment for statistical computing and graphics. Together they provide a sophisticated envi for data mining, statistical analyses, and data visualisation.

See the Help menu for extensive support in using Rattle. The book Data Mining with Rattle and available from Amazon. The Togaware Desktop Data Mining Survival Guide includes Rattle documen and is available from datamining.togaware.com

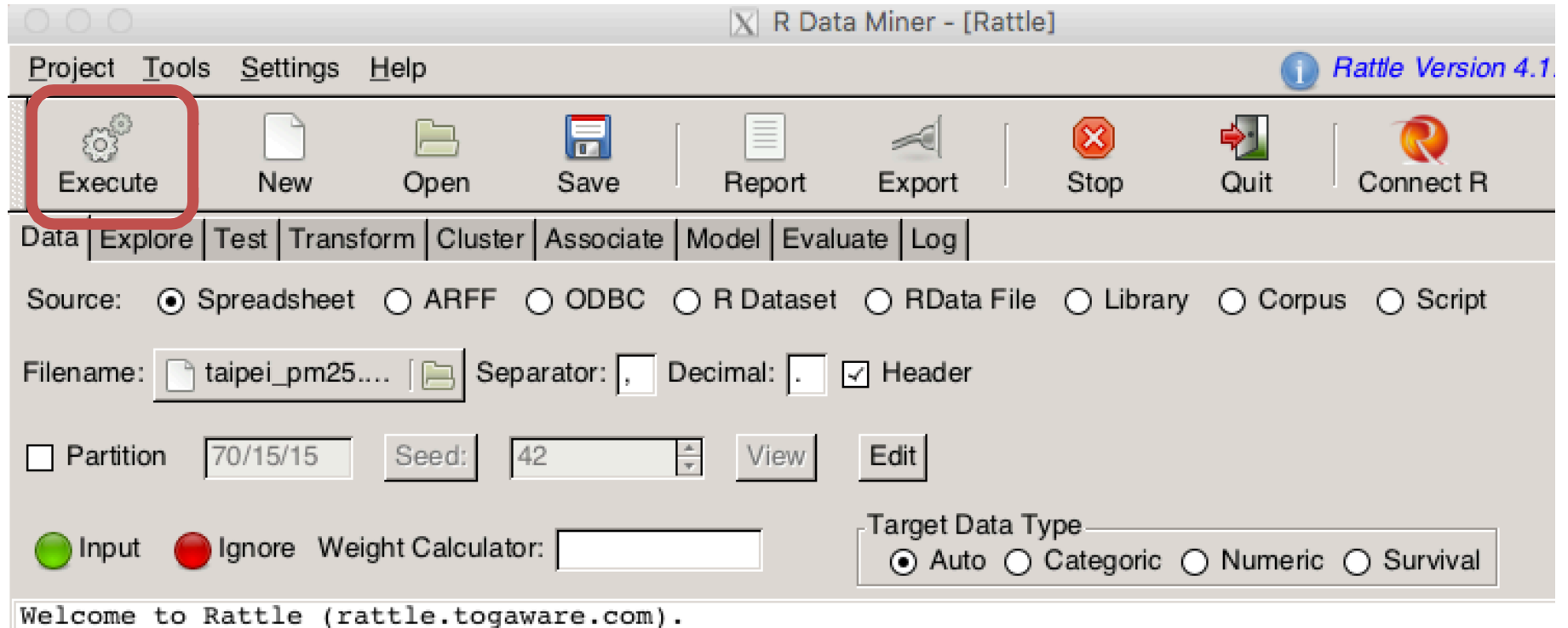
臺北市環保局空氣品質監測站－ 每月監測值統計

懸浮微粒

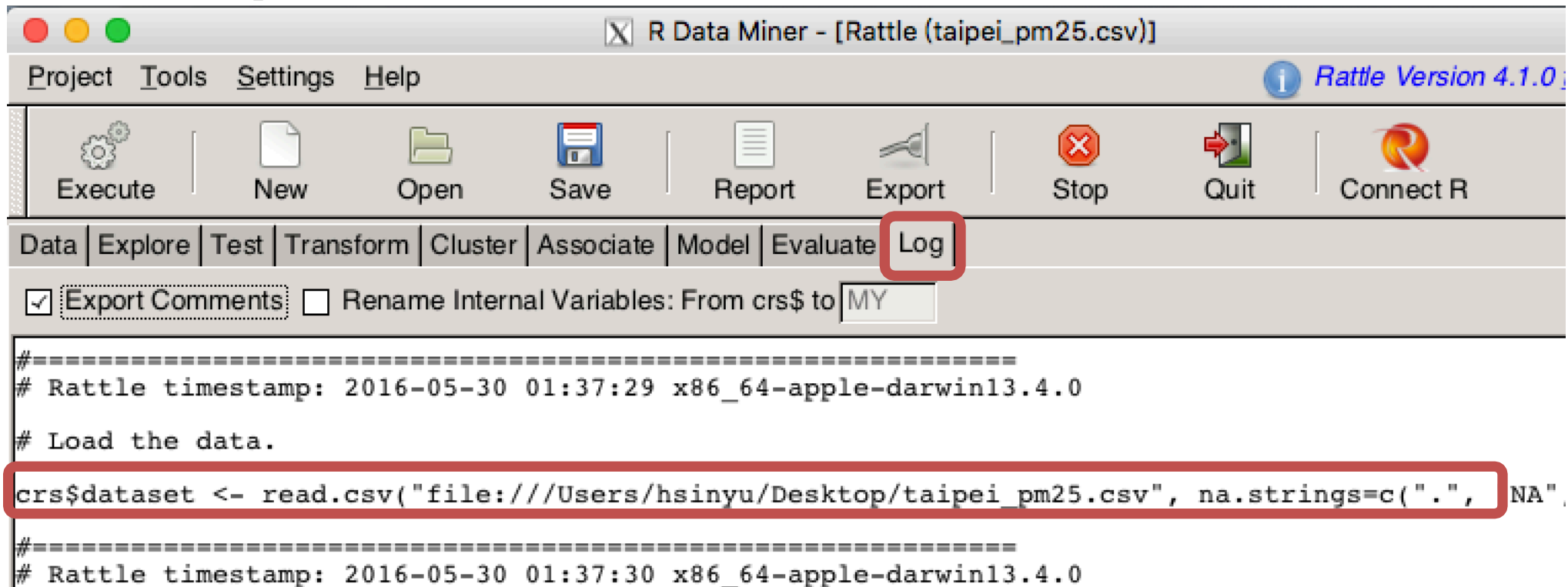
資料區間：104/01 ~ 105/04

測站：中正站、大直站、信義站、南港站、內湖站
木柵站、承德站、中北站

Import Data – 懸浮微粒



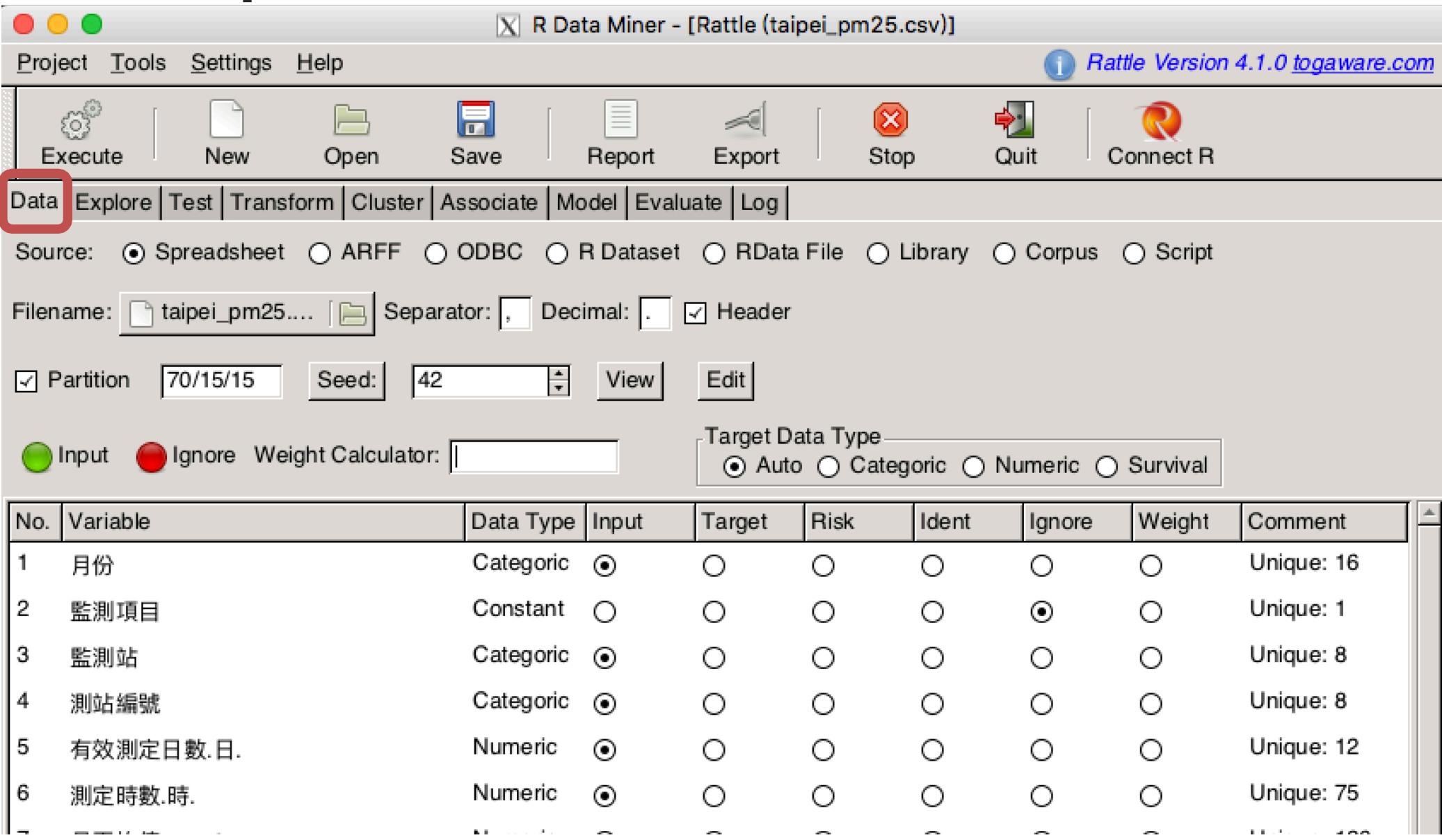
Import Data – 懸浮微粒



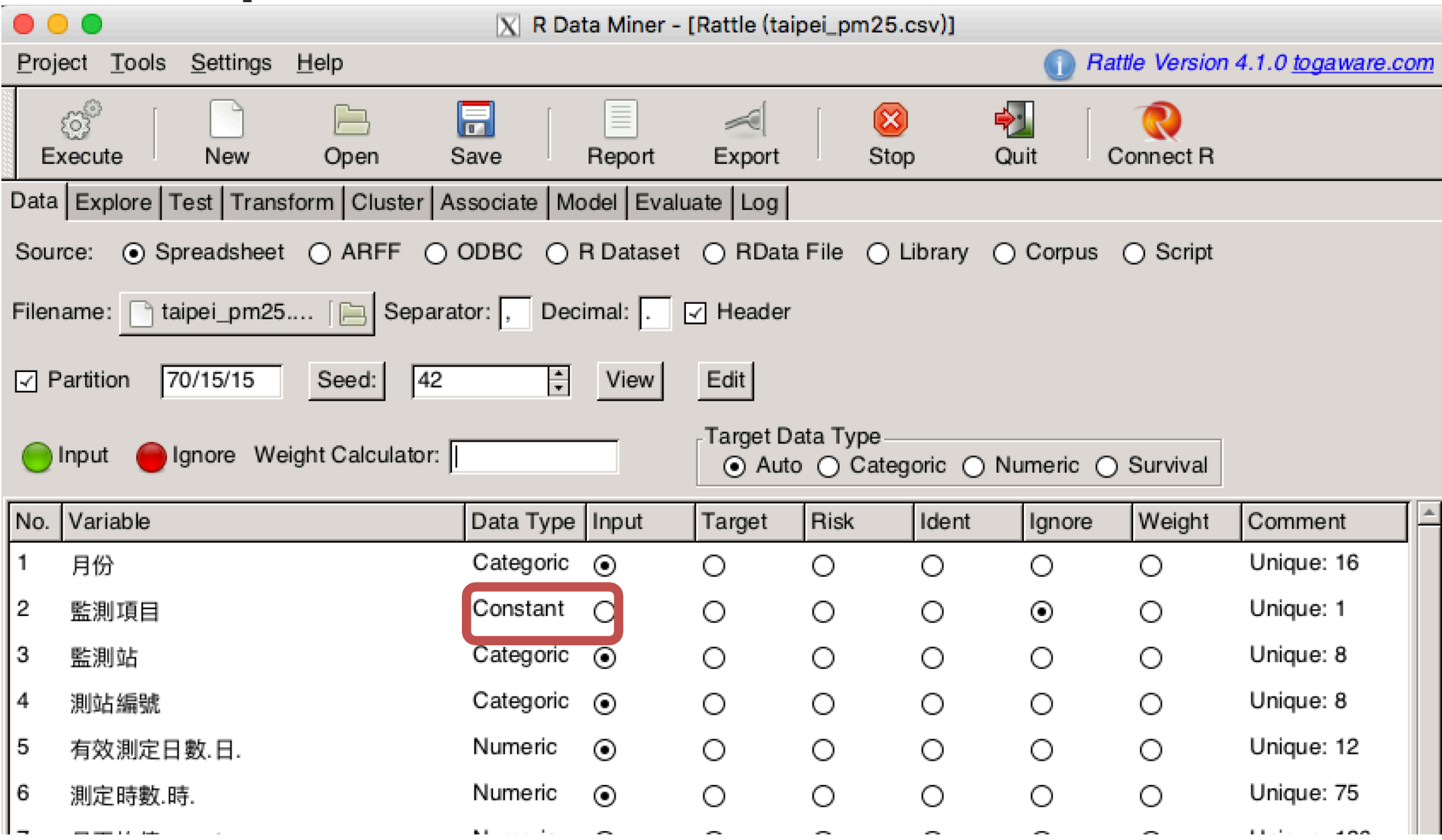
R Studio

```
> head(crs$dataset)
      月份 監測項目 監測站      測站編號 有效測定日數.
值µg.m3
1 104年1月 懸浮微粒 中正站 2533616A0015
44.00
2 104年1月 懸浮微粒 大直站 2533916A0007
```

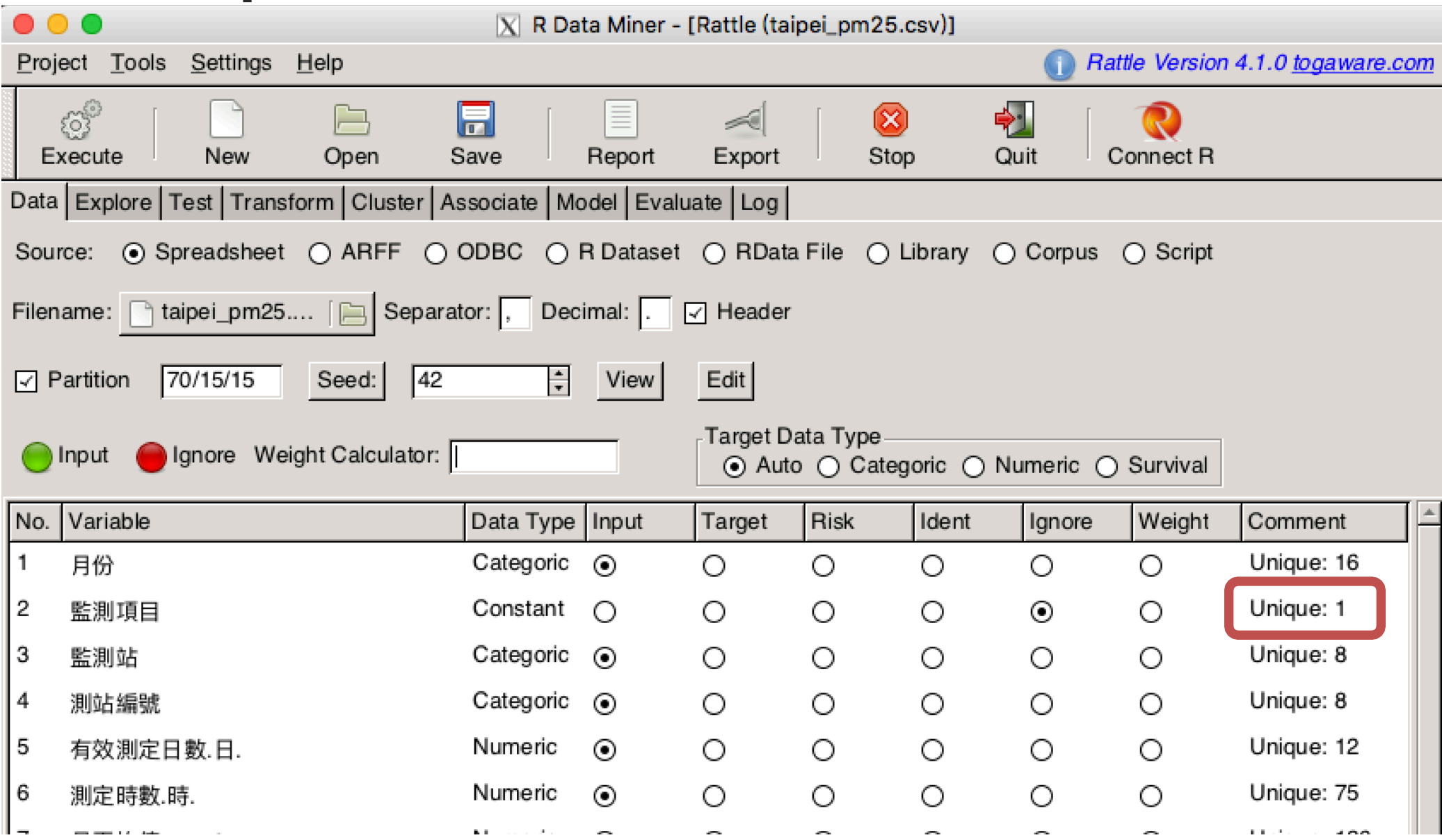
Import Data – 懸浮微粒



Import Data – 懸浮微粒



Import Data – 懸浮微粒



Import Data – 懸浮微粒

R Data Miner - [Rattle (taipei_pm25.csv)]

Project Tools Settings Help

Rattle Version 4.1.0 togaware.com

Execute New Open Save Report Export Stop Quit Connect R

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source: ☒ Spreadsheet ☐ ARFF ☐ ODBC ☐ R Dataset ☐ RData File ☐ Library ☐ Corpus ☐ Script

Filename: taipei_pm25.... Separator: , Decimal: . ☒ Header

☒ Partition 100/0/0 Seed: 42 View Edit

☒ Input ☐ Ignore Weight Calculator: Target Data Type: ☒ Auto ☐ Categorical ☐ Numeric ☐ Survival

| No. | Variable | Data Type | Input | Target | Risk | Ident | Ignore | Weight | Comment |
|-----|----------|-------------|----------------------------------|-----------------------|-----------------------|-----------------------|----------------------------------|-----------------------|------------|
| 1 | 月份 | Categorical | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Unique: 16 |
| 2 | 監測項目 | Constant | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | Unique: 1 |
| 3 | 監測站 | Categorical | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Unique: 8 |

training dataset / validation dataset / test dataset

Explore Data – 懸浮微粒

Rattle

Summary

R Data Miner - [Rattle (taipei_pm25.csv)]

Project Tools Settings Help

Rattle Version 4.1.0 togaware.com

Execute New Open Save Report Export Stop Quit Connect R

Data **Explore** Test Transform Cluster Associate Model Evaluate Log

Type: ☒ Summary ☐ Distributions ☐ Correlation ☐ Principal Components ☐ Interactive

☒ Summary ☒ Describe ☐ Basics ☐ Kurtosis ☐ Skewness ☐ Show Missing ☐ Cross Tab

| | | | |
|---------------|---------------|-----------------|---------------|
| 104年10月: 8 | 承德站 :16 | 2533616A0015:16 | Min. : 7.00 |
| 104年11月: 8 | 大直站 :16 | 2533916A0007:16 | 1st Qu.:27.75 |
| 104年12月: 8 | 木柵站 :16 | 2533916T0022:16 | Median :29.00 |
| 104年1月 : 8 | 南港站 :16 | 2534016A0012:16 | Mean :28.48 |
| 104年2月 : 8 | 內湖站 :16 | 2534116A0005:16 | 3rd Qu.:30.00 |
| 104年3月 : 8 | 信義站 :16 | 2534216A0011:16 | Max. :31.00 |
| (Other) :80 | (Other):32 | (Other) :32 | |
| 測定時數.時. | 月平均值µg.m3 | 一小時值之最高值µg.m3 | |
| Min. :108.0 | Min. :14.79 | Min. : 59.00 | |
| 1st Qu.:662.0 | 1st Qu.:30.73 | 1st Qu.: 95.67 | |
| Median :694.5 | Median :35.90 | Median :119.15 | |
| Mean :678.1 | Mean :36.37 | Mean :129.59 | |
| 3rd Qu.:716.0 | 3rd Qu.:40.84 | 3rd Qu.:147.97 | |
| Max. :744.0 | Max. :62.25 | Max. :929.90 | |

Explore Data – 懸浮微粒

R Studio

Summary

Console ~/ ↻

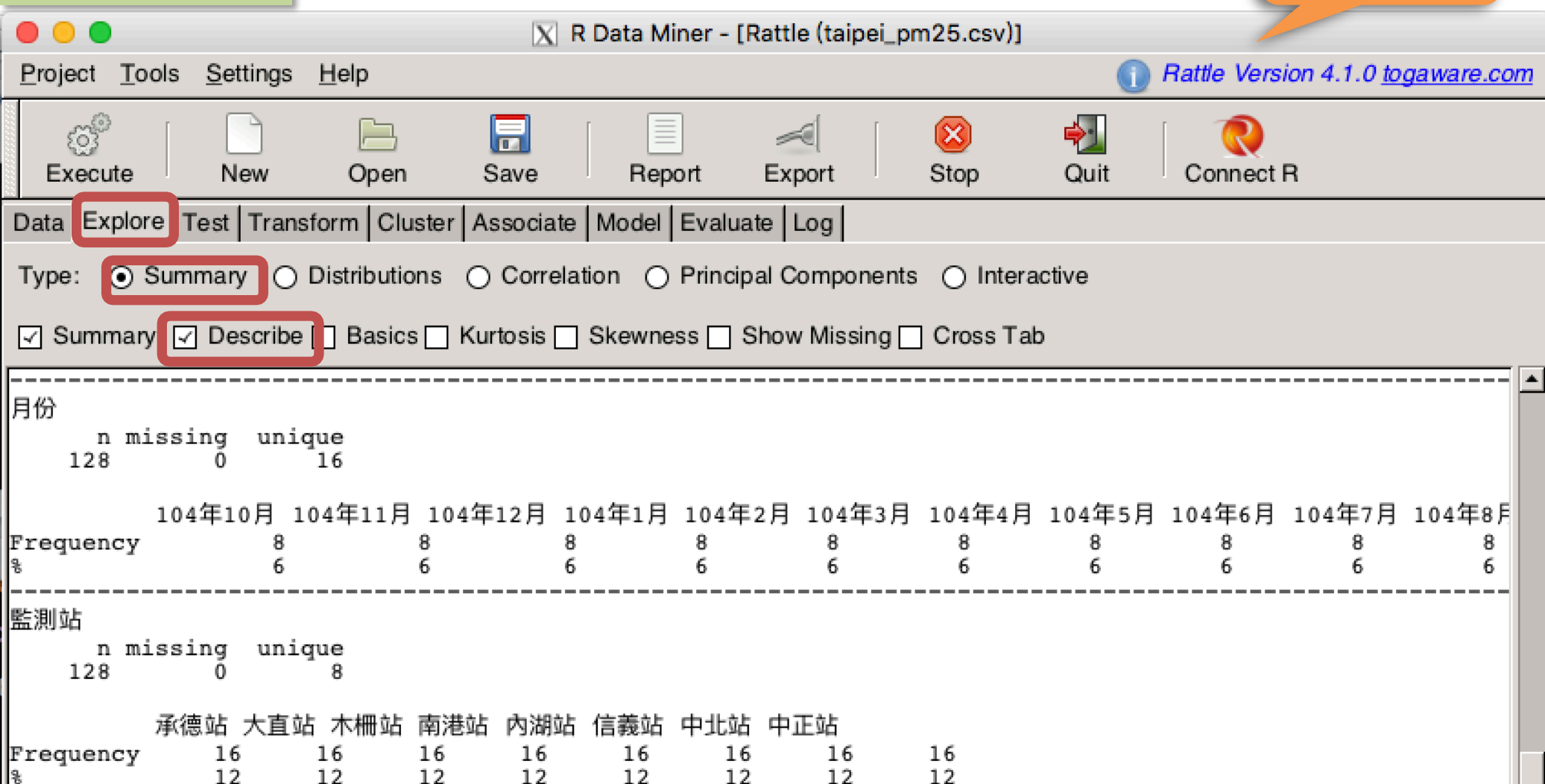
```
> summary(crs$dataset)
```

| 月份 | 監測項目 | 監測站 | 測站編號 |
|------------|----------|---------|-----------------|
| 104年10月: 8 | 懸浮微粒:128 | 承德站 :16 | 2533616A0015:16 |
| 104年11月: 8 | | 大直站 :16 | 2533916A0007:16 |
| 104年12月: 8 | | 木柵站 :16 | 2533916T0022:16 |
| 104年1月 : 8 | | 南港站 :16 | 2534016A0012:16 |

懸浮微粒

Rattle

Describe



Explore Data – 懸浮微粒

R Studio

Describe

Console ~/ ↻

```
> describe(crs$dataset)
```

```
crs$dataset
```

Package: Hmisc

```
16 Variables      128 Observations
```

```
-----  
月份
```

| | n | missing | unique |
|--|-----|---------|--------|
| | 128 | 0 | 16 |

| | 104年10月 | 104年11月 | 104年12月 | 104年1月 | 104年2月 |
|--|---------|---------|---------|--------|--------|
|--|---------|---------|---------|--------|--------|

| | | | | | |
|-----------|---|---|---|---|---|
| Frequency | 8 | 8 | 8 | 8 | 8 |
| % | 6 | 6 | 6 | 6 | 6 |

Explore Data – 懸浮微粒

Rattle

Missing Value

R Data Miner - [Rattle (taipei_pm25.csv)]

Project Tools Settings Help Rattle Version 4.1.0 togaware.com

Execute New Open Save Report Export Stop Quit Connect R

Data **Explore** Test Transform Cluster Associate Model Evaluate Log

Type: ☒ Summary ☐ Distributions ☐ Correlation ☐ Principal Components ☐ Interactive

☐ Summary ☐ Describe ☐ Basics ☐ Kurtosis ☐ Skewness ☒ Show Missing ☐ Cross Tab


Missing Value Summary

| | 月份 | 監測站 | 測站編號 | 有效測定日數.日. | 測定時數.時. | 月平均值 $\mu\text{g.m}^3$ | |
|------|---|-----|------|-----------|---------|------------------------|--|
| [1,] | 1 | 1 | 1 | 1 | 1 | 1 | |
| [2,] | 0 | 0 | 0 | 0 | 0 | 0 | |
| | 一小時值之最高值 $\mu\text{g.m}^3$ 一小時值之最高值.發生日. | | | | | | |
| [1,] | | | 1 | | 1 | | |
| [2,] | | | 0 | | 0 | | |
| | 一小時值之最高值.發生時. 日平均值之最高值 $\mu\text{g.m}^3$ | | | | | | |
| [1,] | | | 1 | | 1 | | |
| [2,] | | | 0 | | 0 | | |
| | 日平均值之最高值.發生日期. 一小時值超過 $200\mu\text{g.m}^3$ 之小時時數.時. | | | | | | |
| [1,] | | | 1 | | | 1 | |
| [2,] | | | 0 | | | 0 | |
| | 一小時值超過 $200\mu\text{g.m}^3$ 之小時時數... 日平均值超過 $125\mu\text{g.m}^3$ 之日數.日. | | | | | | |

Explore Data – 懸浮微粒

R Studio

Missing Value

```
Console ~/ 
> md.pattern(crs$dataset)
```

| | 月份 | 監測項目 | 監測站 | 測站編號 | 有效測定日數.日. | 測定時數.時. |
|------|----|------|-----|------|-----------|---------|
| [1,] | 1 | 1 | 1 | 1 | 1 | 1 |
| [2,] | 0 | 0 | 0 | 0 | 0 | 0 |

月平均值 $\mu\text{g.m3}$ 一小時值之最高值 $\mu\text{g.m3}$ 一小時值之最高值.發生日.

| | | | |
|------|---|---|---|
| [1,] | 1 | 1 | 1 |
| [2,] | 0 | 0 | 0 |

一小時值之最高值.發生時. 日平均值之最高值 $\mu\text{g.m3}$

| | | |
|------|---|---|
| [1,] | 1 | 1 |
| [2,] | 0 | 0 |

日平均值之最高值.發生日期. 一小時值超過 $200\mu\text{g.m3}$ 之小時時數.時.

| | | |
|------|---|---|
| [1,] | 1 | 1 |
| [2,] | 0 | 0 |

Package: mice

Explore Data – 懸浮微粒

Rattle

Boxplot

R Data Miner - [Rattle (taipei_pm25.csv)]

Project Tools Settings Help

Execute New Open Save Report Export Stop Quit Connect R

Data **Explore** Test Transform Cluster Associate Model Evaluate Log

Type: ☐ Summary ☒ **Distributions** ☐ Correlation ☐ Principal Components ☐ Interactive

Numeric: ☐ Annotate **Group By:** 監測站 ▼

Benfords: ☐ Bars Starting Digit: 1 Number of Digits: 1 ☒ abs ☐ +ve ☐ -ve

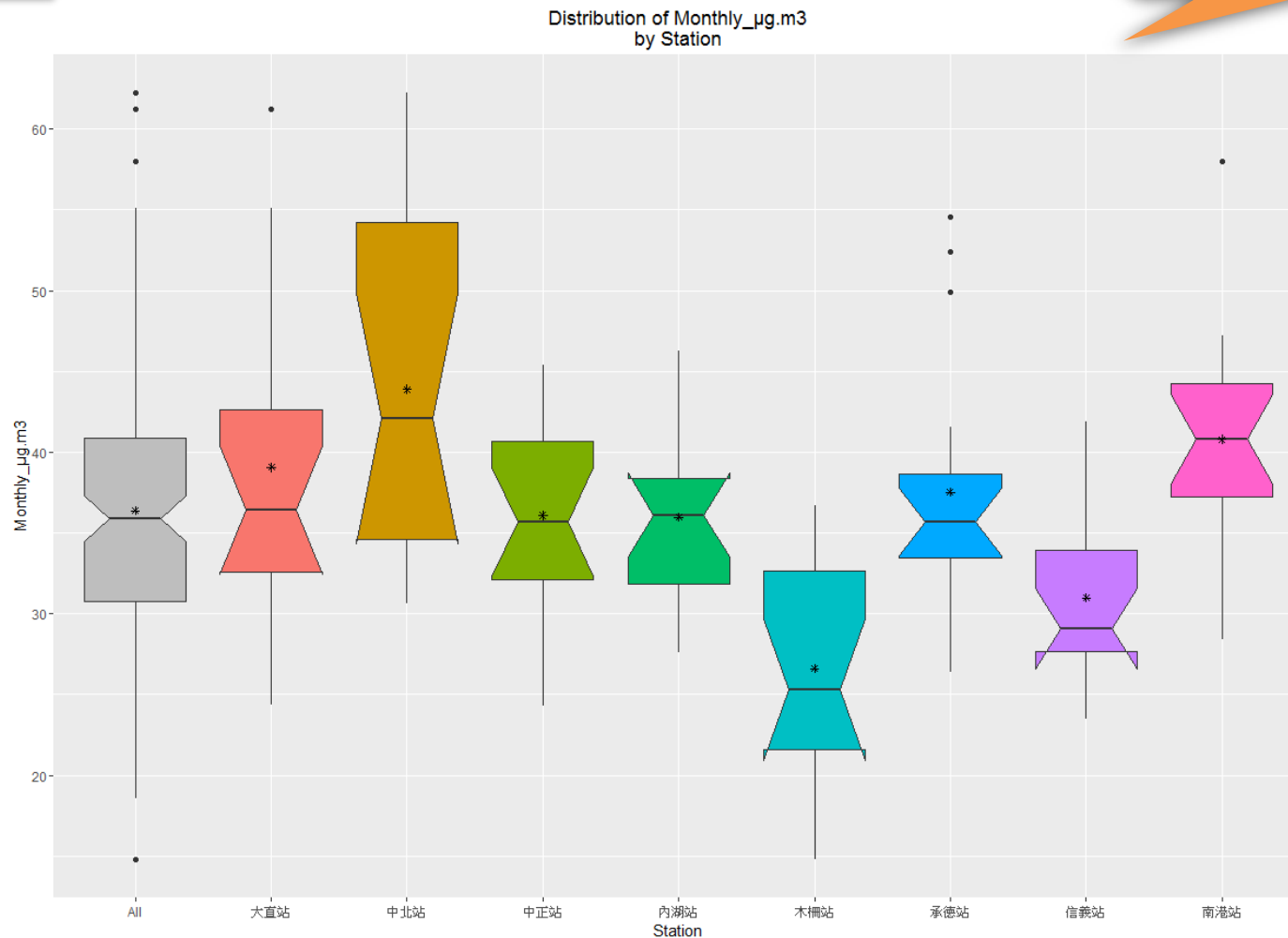
| No. | Variable | Box Plot | Histogram | Cumulative | Benford | Pairs | Min; Median/Mean; Max |
|-----|---------------|-------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------------------------|
| 8 | 一小時值之最高值µg.m3 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 59.00; 119.15/129.59; 929.90 |
| 9 | 一小時值之最高值.發生日. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 1.00; 17.00/16.04; 31.00 |

Categoric: Clear

Explore Data – 懸浮微粒

Rattle

Boxplot



鳶尾花(iris)

資料筆數：150

Sepal Length – 花萼長度<cm>

Sepal Width – 花萼寬度<cm>

Petal Length – 花瓣長度<cm>

Petal Width – 花瓣 寬度<cm>

Species – iris 種類< Setosa、Versicolor、Virginica >

Model

Rattle

Decision Tree

The screenshot shows the Rattle R Data Miner interface. The title bar reads "R Data Miner - [Rattle (iris)]". The menu bar includes "Project", "Tools", "Settings", and "Help". The toolbar contains icons for "Execute", "New", "Open", "Save", "Report", "Export", "Stop", "Quit", and "Connect R". The main menu has tabs for "Data", "Explore", "Test", "Transform", "Cluster", "Associate", "Model", "Evaluate", and "Log". The "Model" tab is selected and highlighted with a red box. Below the tabs, the "Type" section has radio buttons for "Tree", "Forest", "Boost", "SVM", "Linear", "Neural Net", "Survival", and "All". The "Tree" option is selected and highlighted with a red box. The "Algorithm" section has radio buttons for "Traditional" and "Conditional", with "Traditional" selected. The "Model Builder" is set to "rpart". The "Min Split" is 20, "Max Depth" is 30, "Priors" is empty, and "Include Missing" is unchecked. The "Min Bucket" is 7, "Complexity" is 0.0100, "Loss Matrix" is empty, and "Rules" and "Draw" buttons are present. The output window shows a summary of the decision tree model for classification, built using "rpart". The text is as follows:

```
Summary of the Decision Tree model for Classification (built using 'rpart'):  
n= 105  
node), split, n, loss, yval, (yprob)  
* denotes terminal node  
1) root 105 68 virginica (0.31428571 0.33333333 0.35238095)  
 2) Petal.Length< 2.6 33 0 setosa (1.00000000 0.00000000 0.00000000) *  
 3) Petal.Length>=2.6 72 35 virginica (0.00000000 0.48611111 0.51388889)  
    6) Petal.Length< 4.85 37 3 versicolor (0.00000000 0.91891892 0.08108108) *  
    7) Petal.Length>=4.85 35 1 virginica (0.00000000 0.02857143 0.97142857) *
```

Model

Rattle

Decision Tree

The screenshot shows the Rattle R Data Miner interface. The title bar reads "R Data Miner - [Rattle (iris)]". The menu bar includes "Project", "Tools", "Settings", and "Help". The toolbar contains icons for "Execute", "New", "Open", "Save", "Report", "Export", "Stop", "Quit", and "Connect R". The main window has tabs for "Data", "Explore", "Test", "Transform", "Cluster", "Associate", "Model", "Evaluate", and "Log". The "Model" tab is active, showing the "Model Builder" window. The "Type" is set to "Tree" (radio button selected). The "Target" is "Species". The "Algorithm" is "Traditional" (radio button selected). The "Model Builder" is "rpart". The "Min Split" is 20, "Max Depth" is 30, "Priors" is empty, "Include Missing" is unchecked, "Min Bucket" is 7, "Complexity" is 0.0100, "Loss Matrix" is empty, and the "Rules" button is highlighted with a red box. The "Tree as rules:" section displays the following rules:

```
Tree as rules:

Rule number: 2 [Species=setosa cover=33 (31%) prob=1.00]
  Petal.Length < 2.6

Rule number: 7 [Species=virginica cover=35 (33%) prob=0.00]
  Petal.Length >= 2.6
  Petal.Length >= 4.85

Rule number: 6 [Species=versicolor cover=37 (35%) prob=0.00]
  Petal.Length >= 2.6
  Petal.Length < 4.85
```


Model

Rattle

Decision Tree

The screenshot displays the Rattle R Data Miner interface. The title bar reads "R Data Miner - [Rattle (iris)]". The menu bar includes "Project", "Tools", "Settings", and "Help". The toolbar contains icons for "Execute", "New", "Open", "Save", "Report", "Export", "Stop", "Quit", and "Connect R". The main menu bar includes "Data", "Explore", "Test", "Transform", "Cluster", "Associate", "Model", "Evaluate", and "Log".

Under the "Model" menu, the "Type" is set to "Tree" (selected), with other options being "Forest", "Boost", "SVM", "Linear", "Neural Net", "Survival", and "All". The "Target" is "Species", and the "Algorithm" is "Traditional" (selected), with "Conditional" also available. Configuration parameters are set as follows: "Min Split" is 20, "Max Depth" is 30, "Min Bucket" is 7, and "Complexity" is 0.0100.

The "Model Builder" section shows "rpart" as the selected builder. There is an "Include Missing" checkbox which is unchecked. The "Rules" button is highlighted with a red box, and a red arrow points from it towards the decision tree diagram.

The decision tree diagram, titled "Decision Tree iris \$ Species", shows the following structure:

- Root node (1): virginica (.31 .33 .35, 100%). Split on "Petal.Length < 2.6".
 - Yes branch: Leaf node (2): setosa (1.00 .00 .00, 31%).
 - No branch: Node (3): virginica (.00 .49 .51, 69%). Split on "Petal.Length < 4.8".
 - Yes branch: Leaf node (6): versicolor (.00 .92 .08, 35%).
 - No branch: Leaf node (7): virginica (.00 .03 .97, 33%).