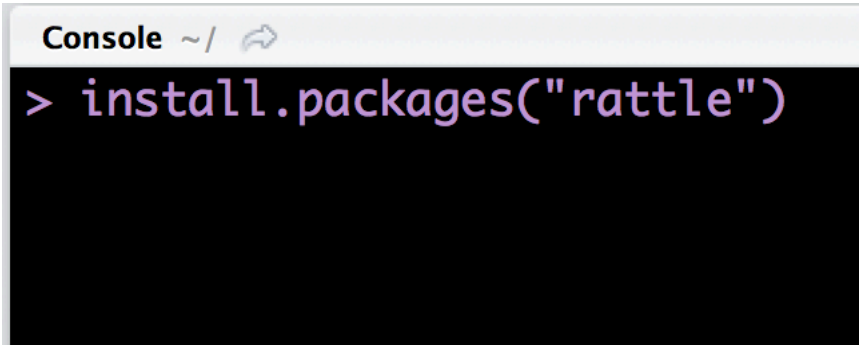
Rattle

A Data Mining GUI for R

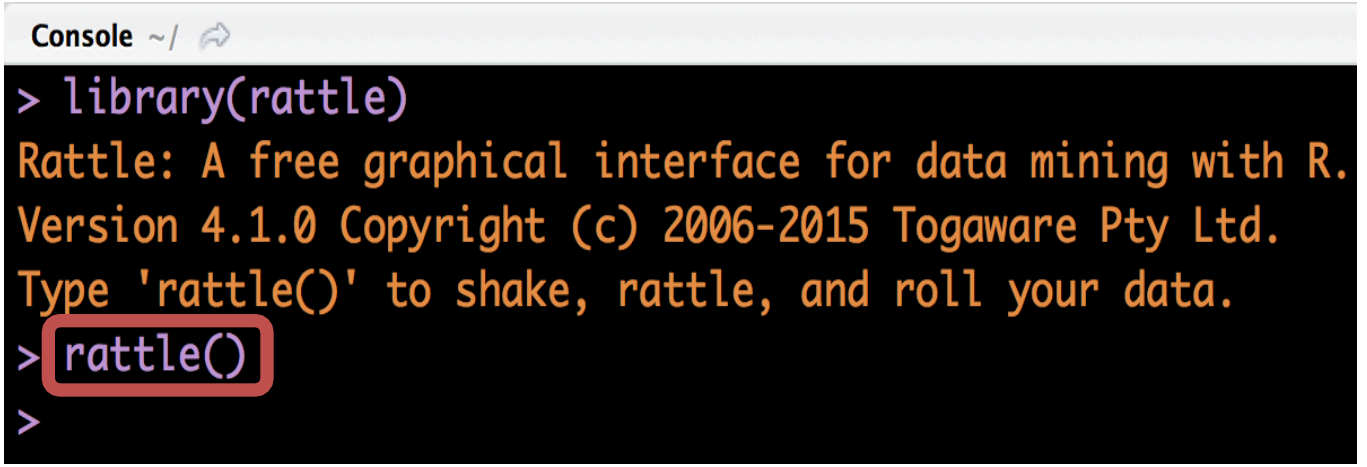
Kristen Chan

安裝、載入 Rattle

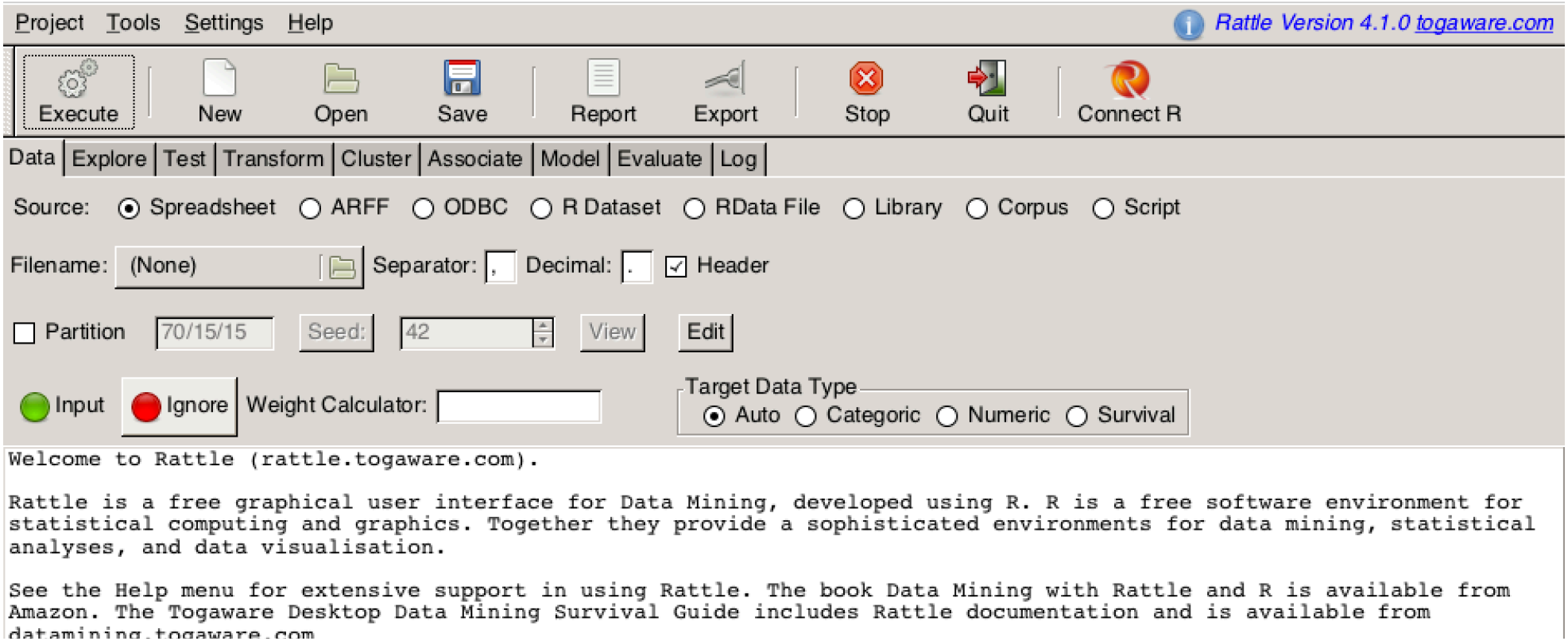
安 裝



載 入

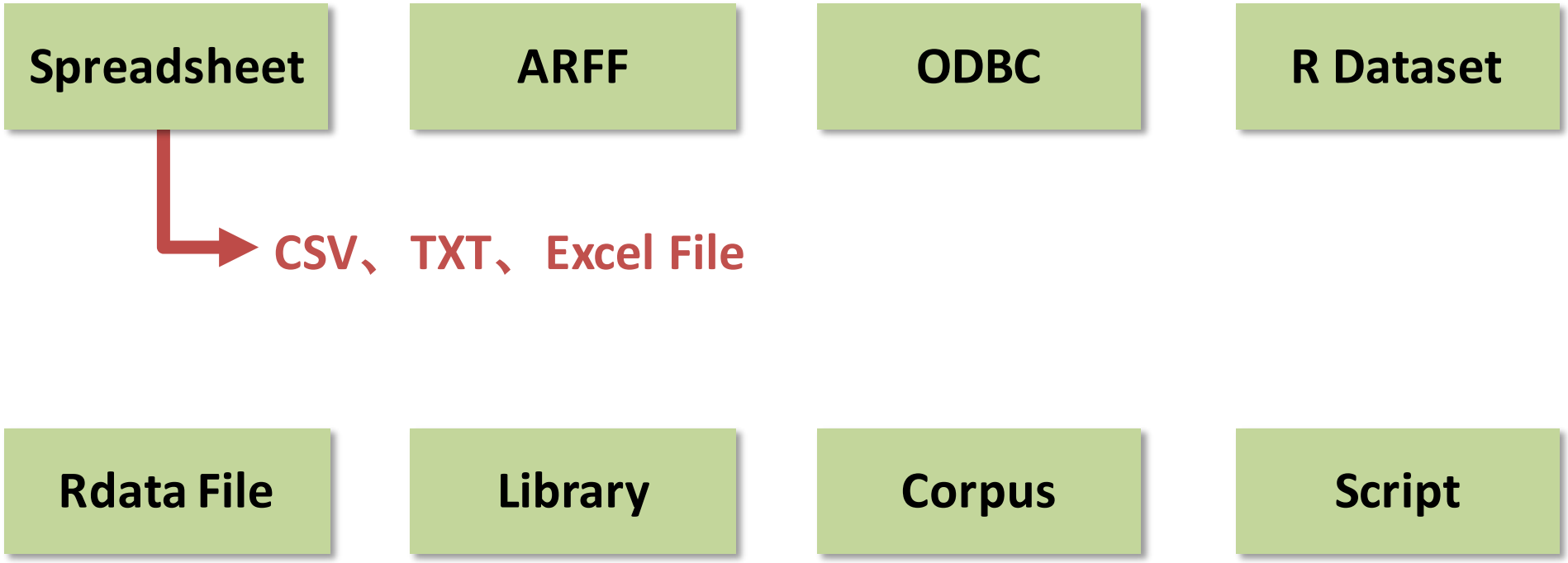


Rattle

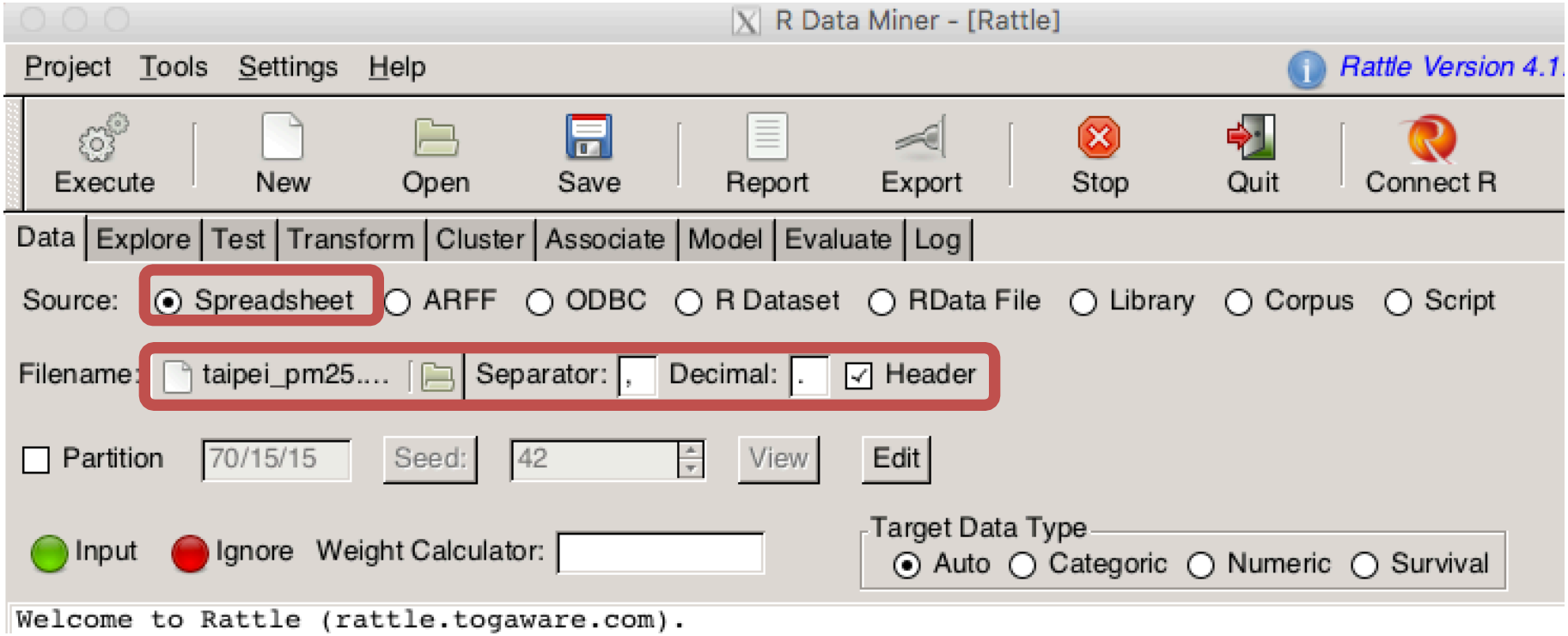


Kristen Chan

Import Data 載入數據

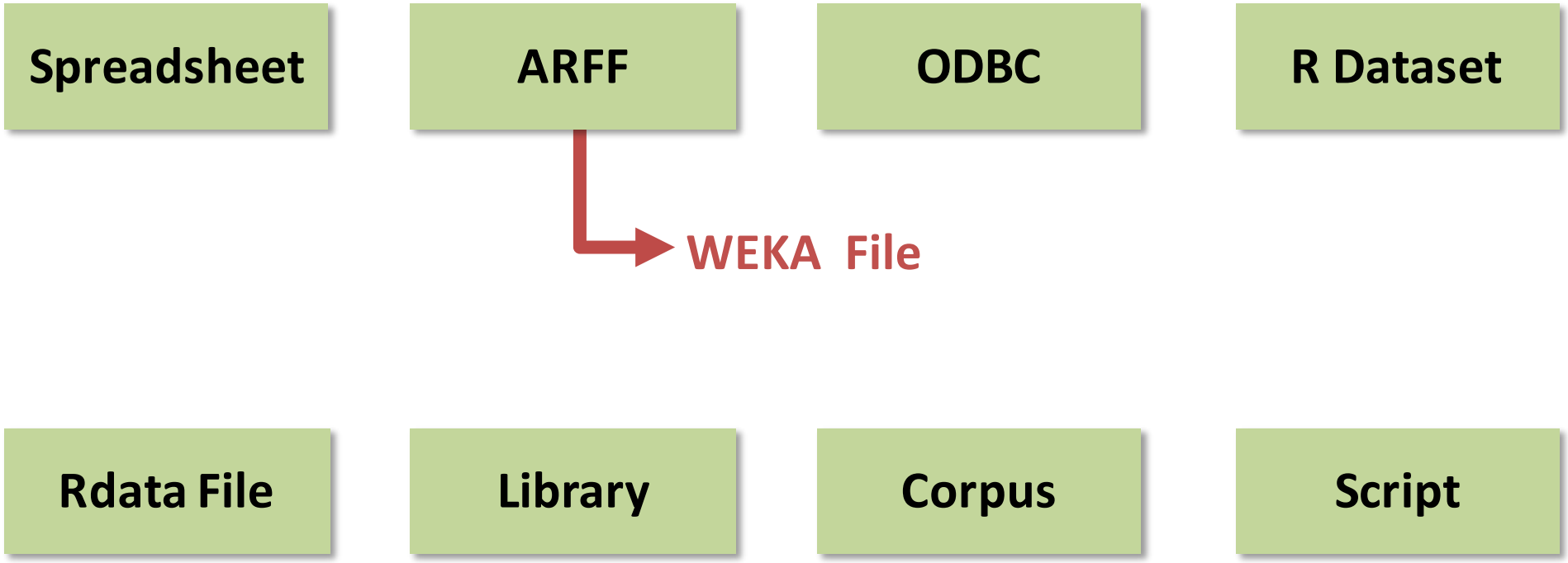


Import Data – Spreadsheet



Kristen Chan

Import Data 載入數據



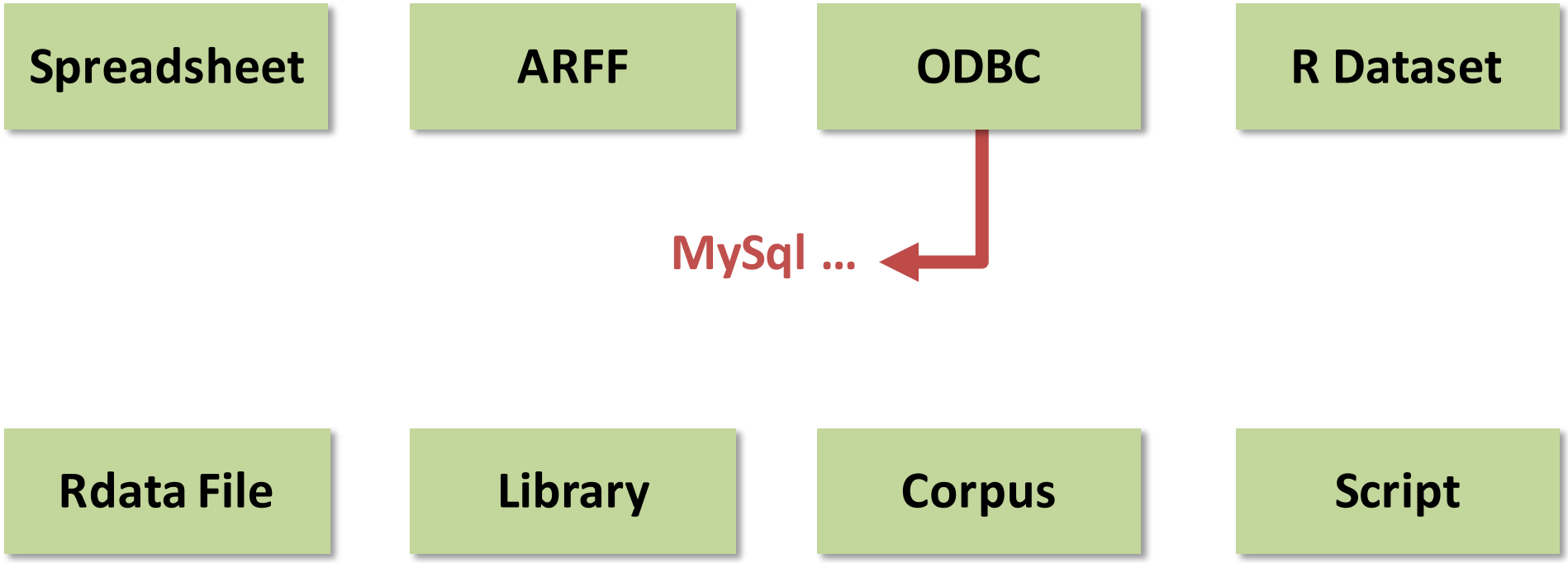
Weka

Weka (Waikato Environment for Knowledge Analysis)

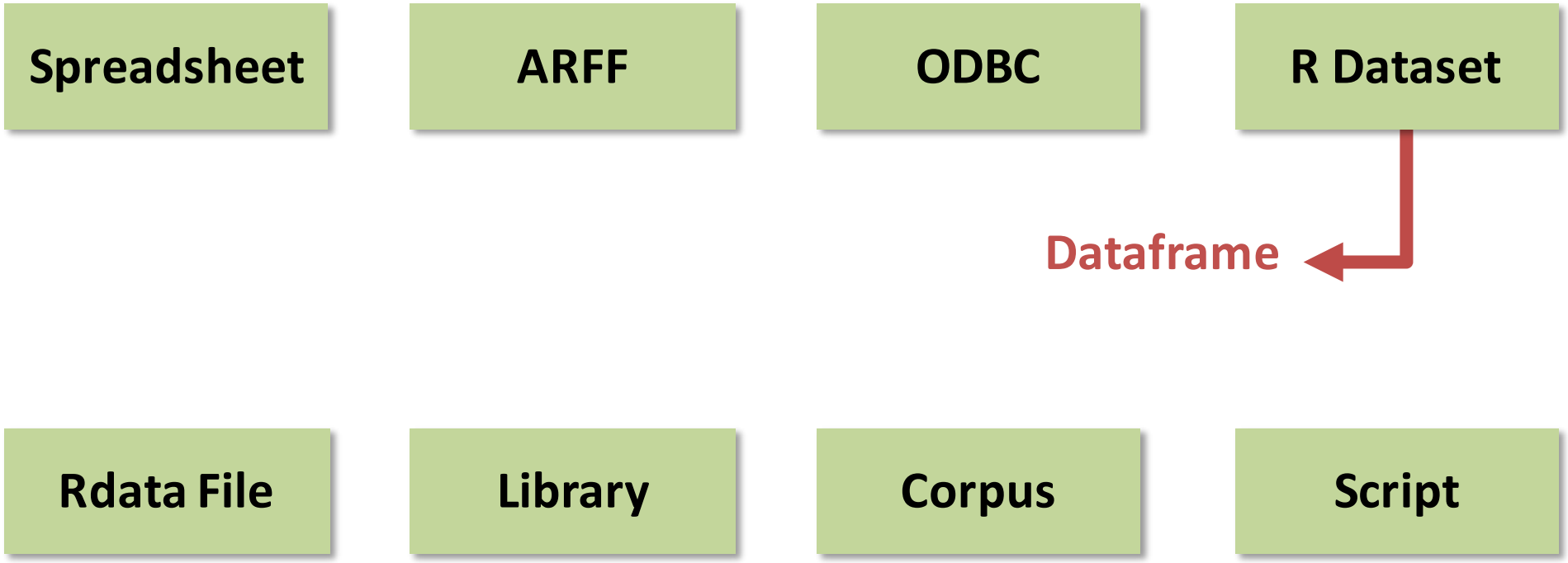
提供 data mining 以及 machine learning 各種工具的一套軟體
包括資料前置處理工具、分類工具、回歸分析等，也能將資料以視覺畫的方式呈現。

Kristen Chan

Import Data 載入數據

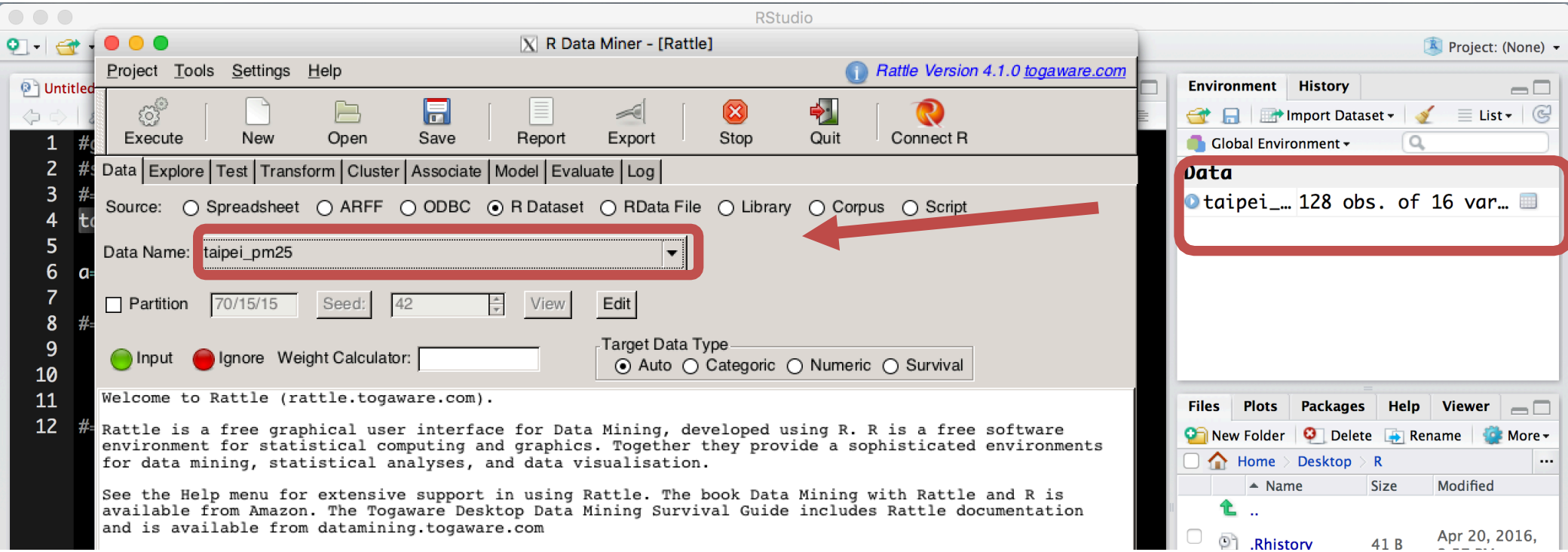


Import Data 載入數據



Kristen Chan

Import Data – R Dataset



Kristen Chan

Import Data 載入數據

Spreadsheet | ARFF | ODBC | R Dataset

Rdata File | Library | Corpus | Script

Kristen Chan

Import Data 載入數據

Spreadsheet

ARFF

ODBC

R Dataset

Rdata File

Library

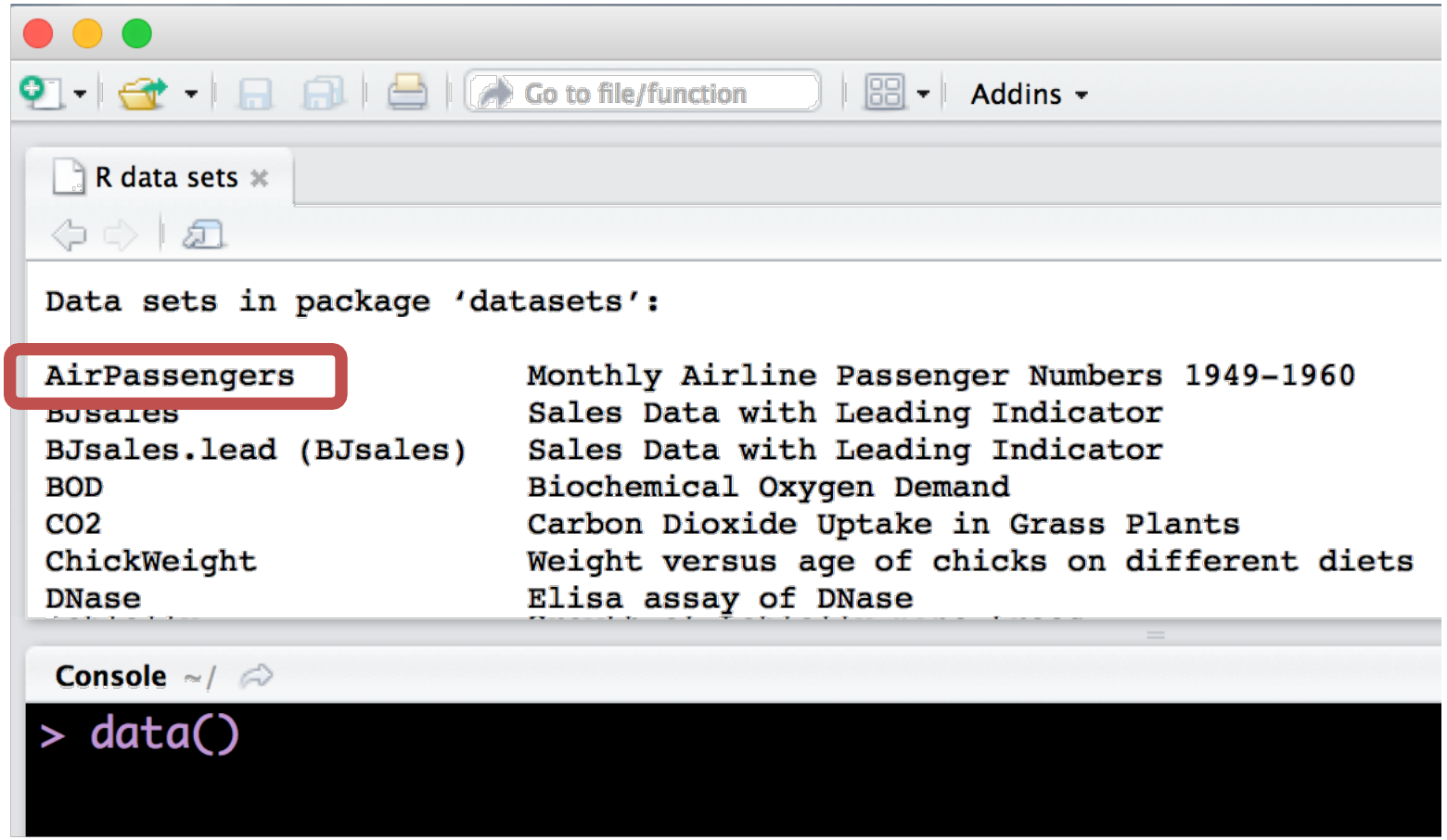
Corpus

Script

套件中的資料

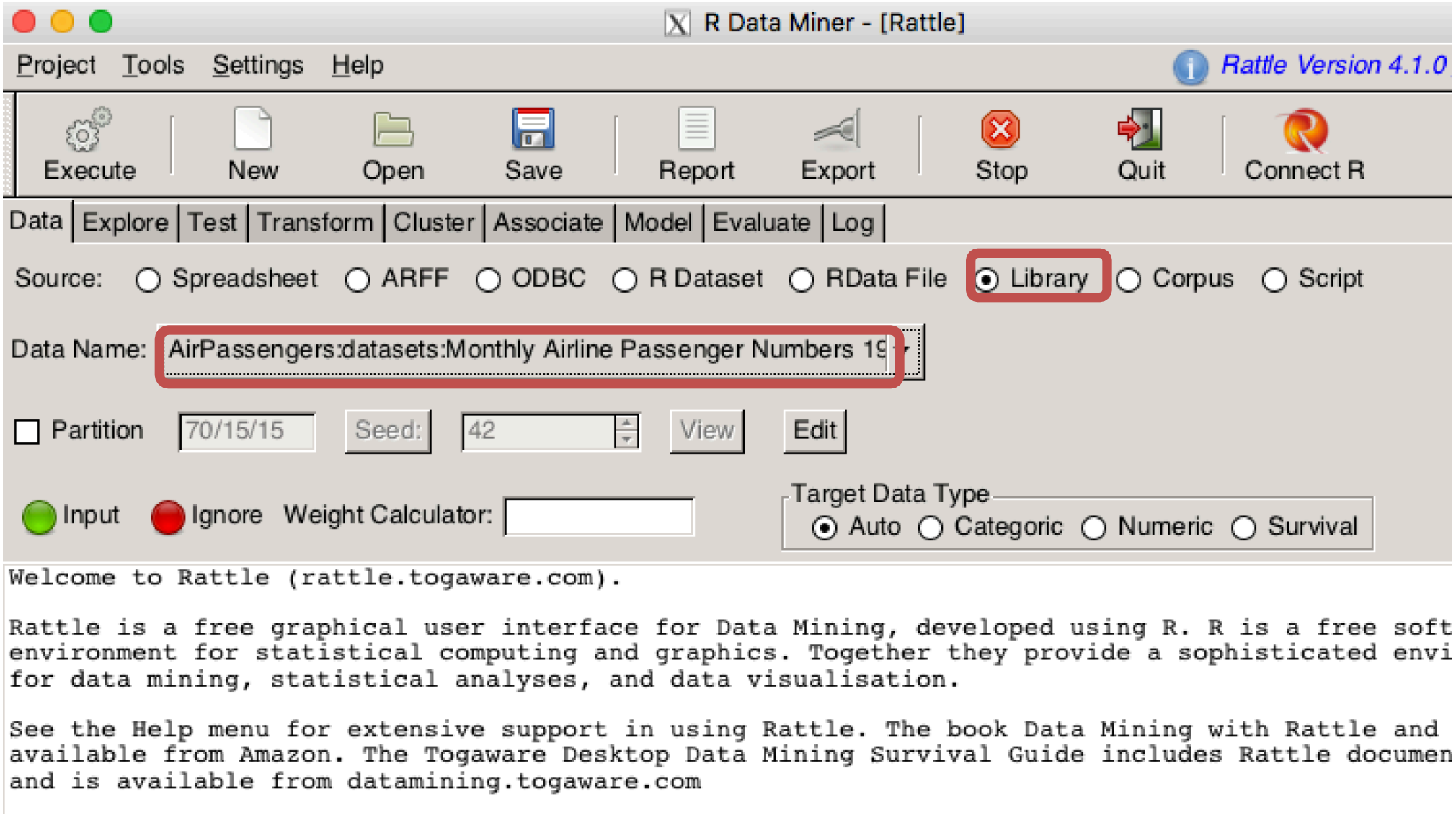
Kristen Chan

看目前可用資料集



Kristen Chan

Import Data – R Dataset



Kristen Chan

臺北市環保局空氣品質監測站－每月監測值統計

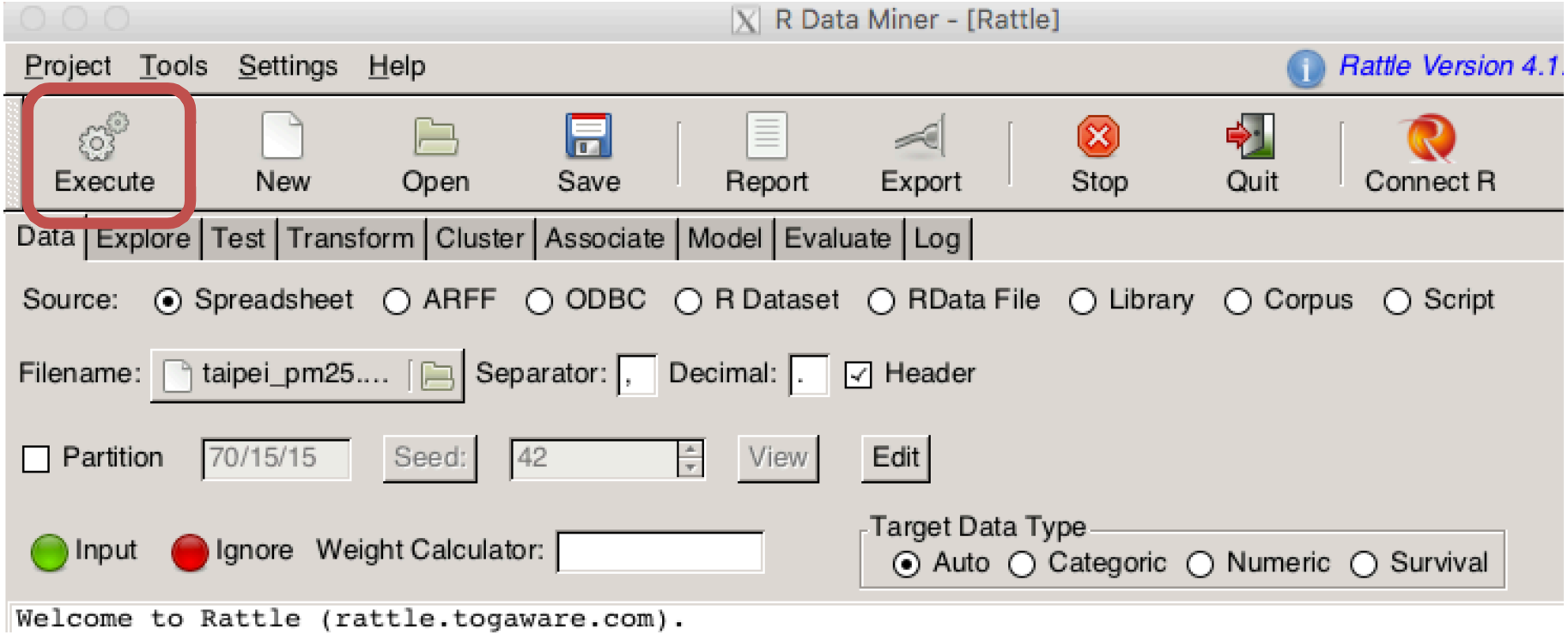
懸浮微粒

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

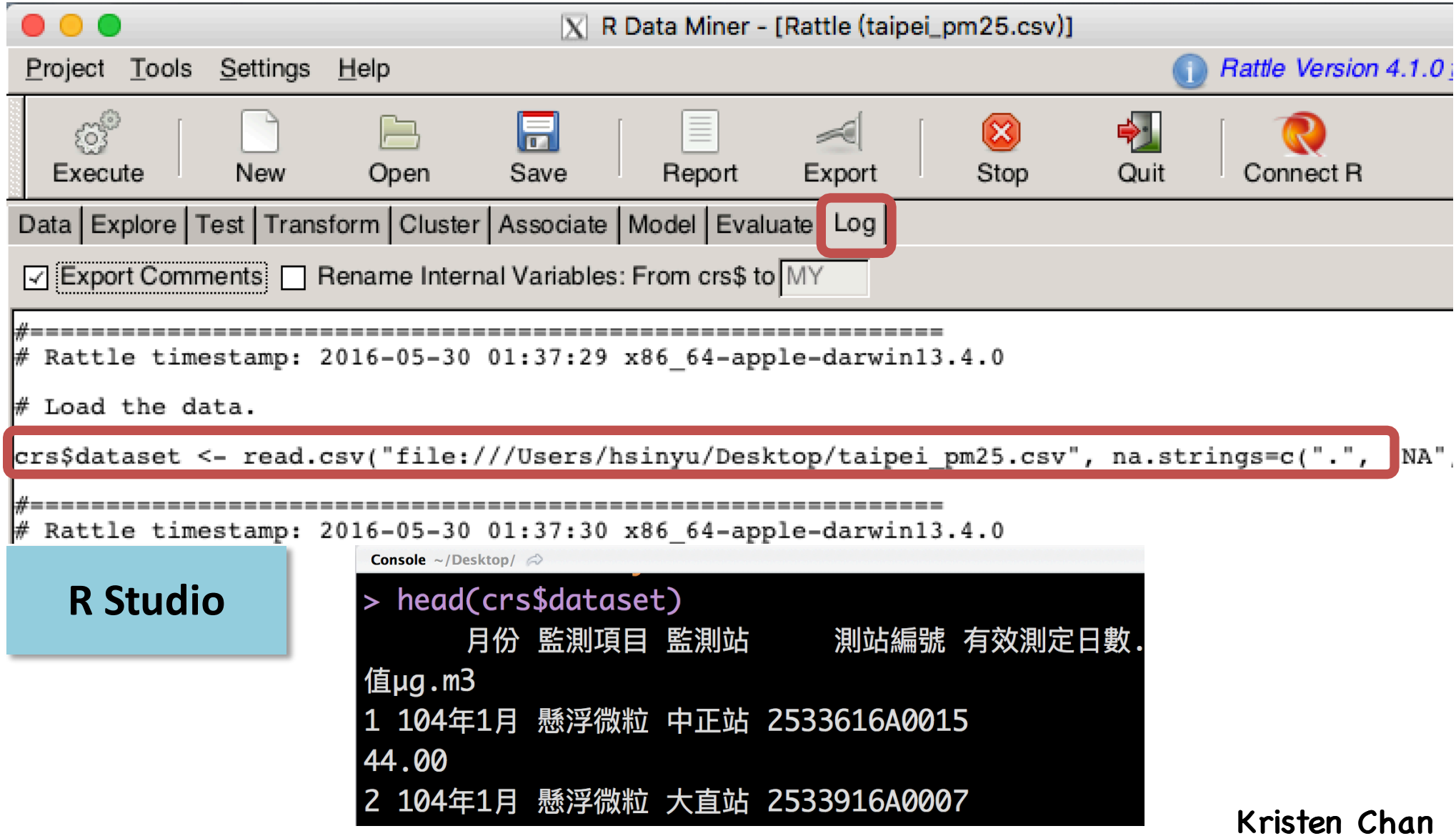
Kristen Chan

Import Data – 懸浮微粒



Kristen Chan

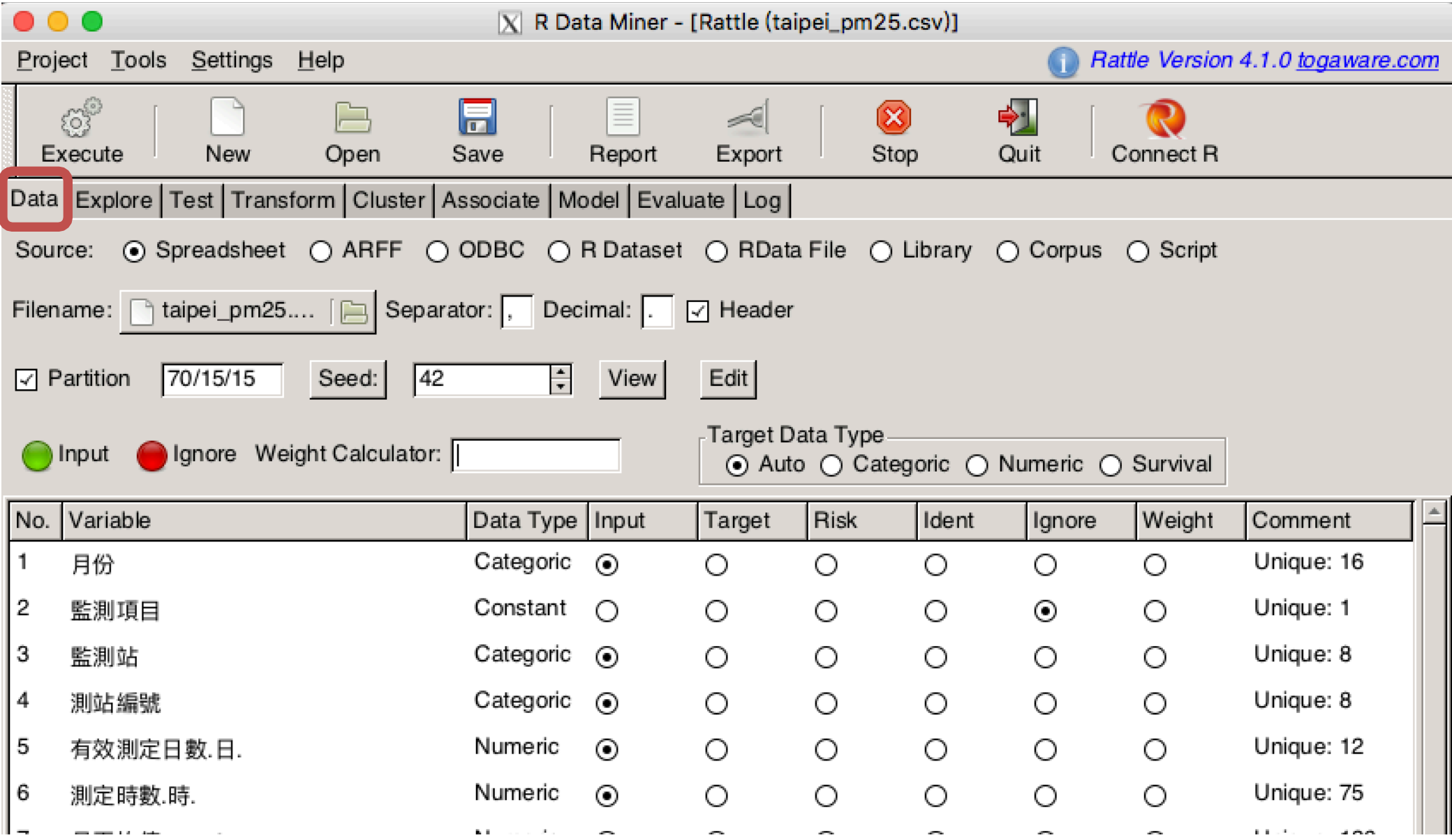
Import Data – 懸浮微粒



R Studio

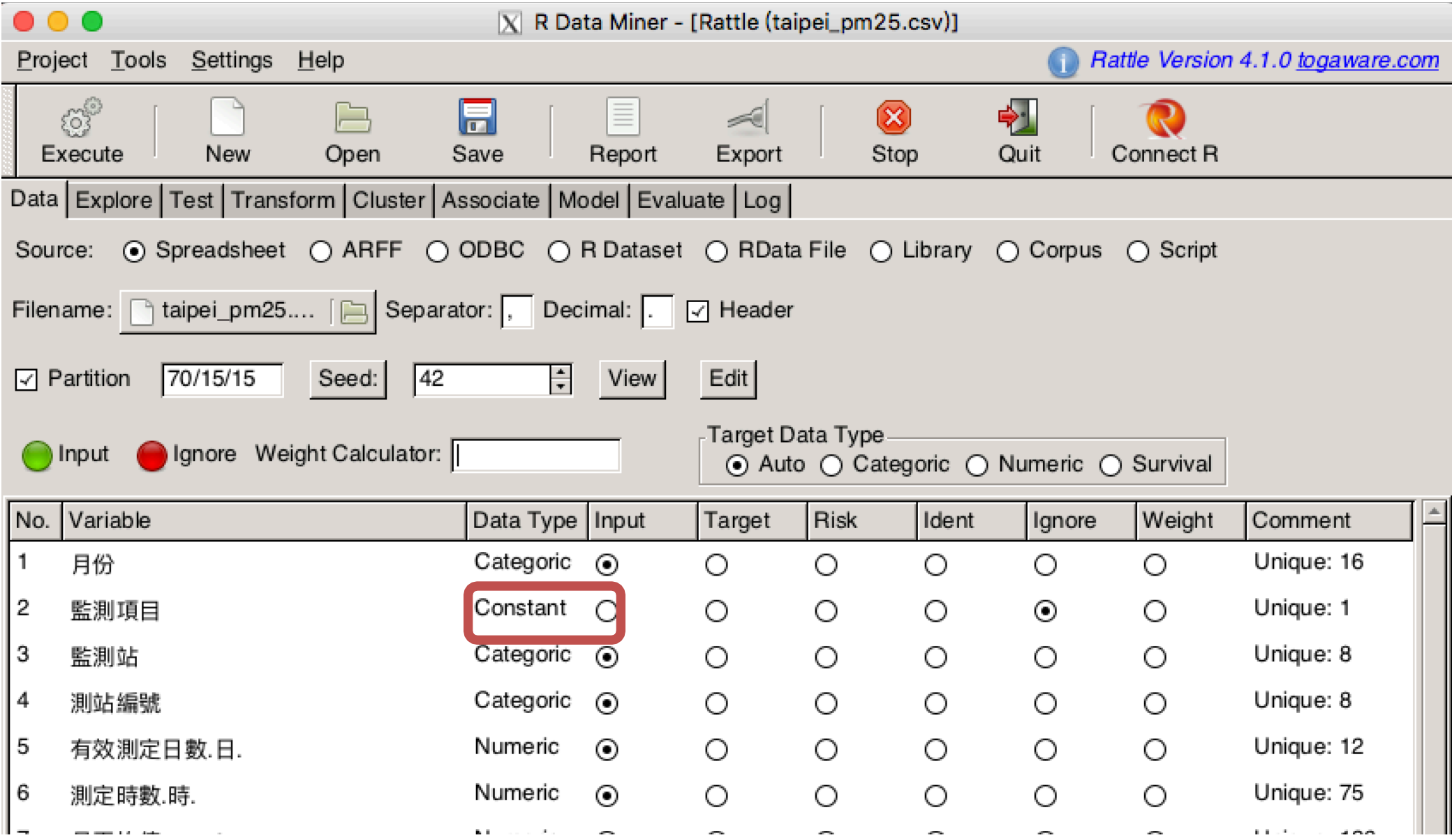
Kristen Chan

Import Data – 懸浮微粒



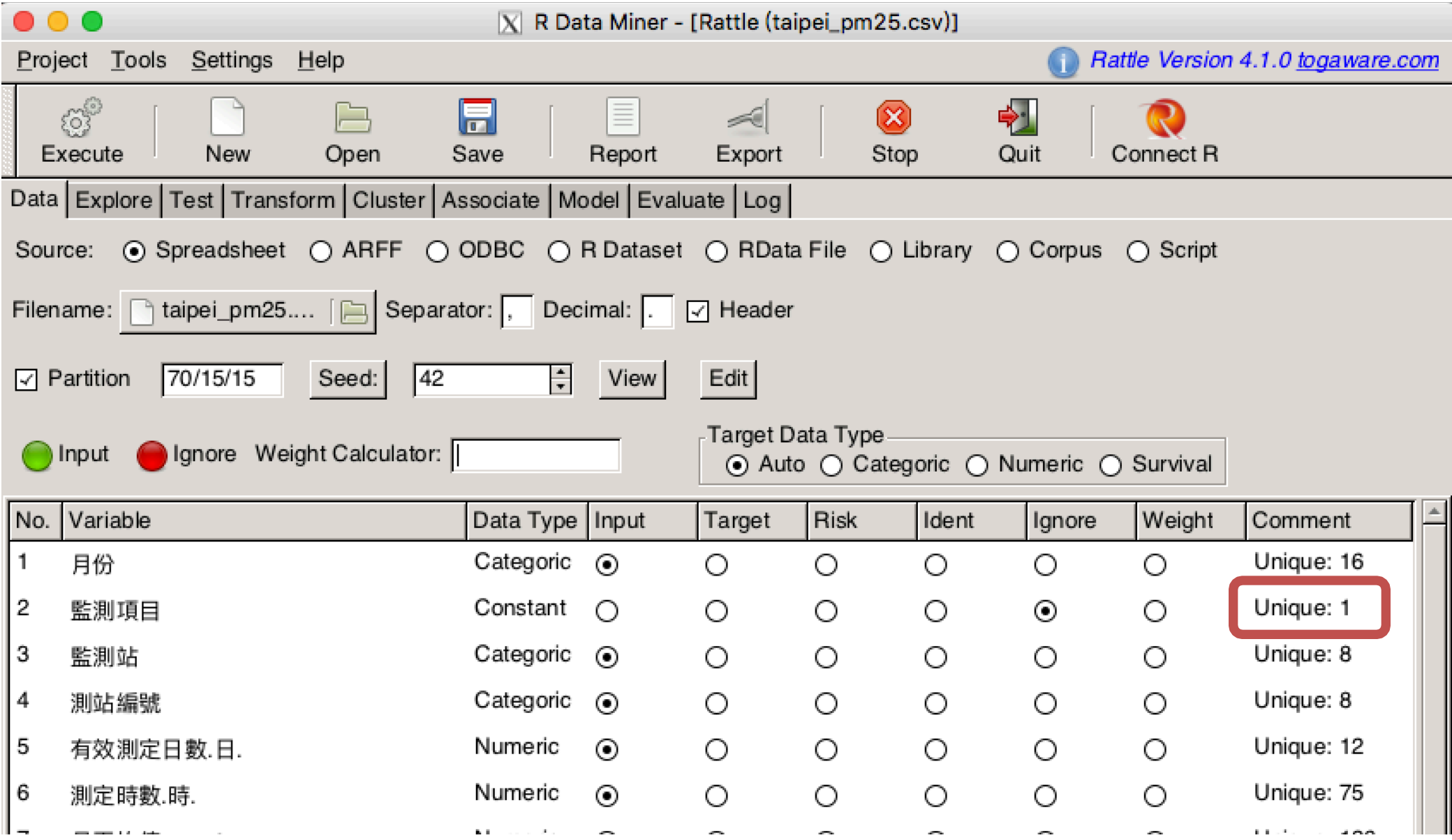
No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	月份	Categoric	◉	○	○	○	○	○	Unique: 16
2	監測項目	Constant	○	○	○	○	◉	○	Unique: 1
3	監測站	Categoric	◉	○	○	○	○	○	Unique: 8
4	測站編號	Categoric	◉	○	○	○	○	○	Unique: 8
5	有效測定日數.日.	Numeric	◉	○	○	○	○	○	Unique: 12
6	測定時數.時.	Numeric	◉	○	○	○	○	○	Unique: 75

Import Data – 懸浮微粒

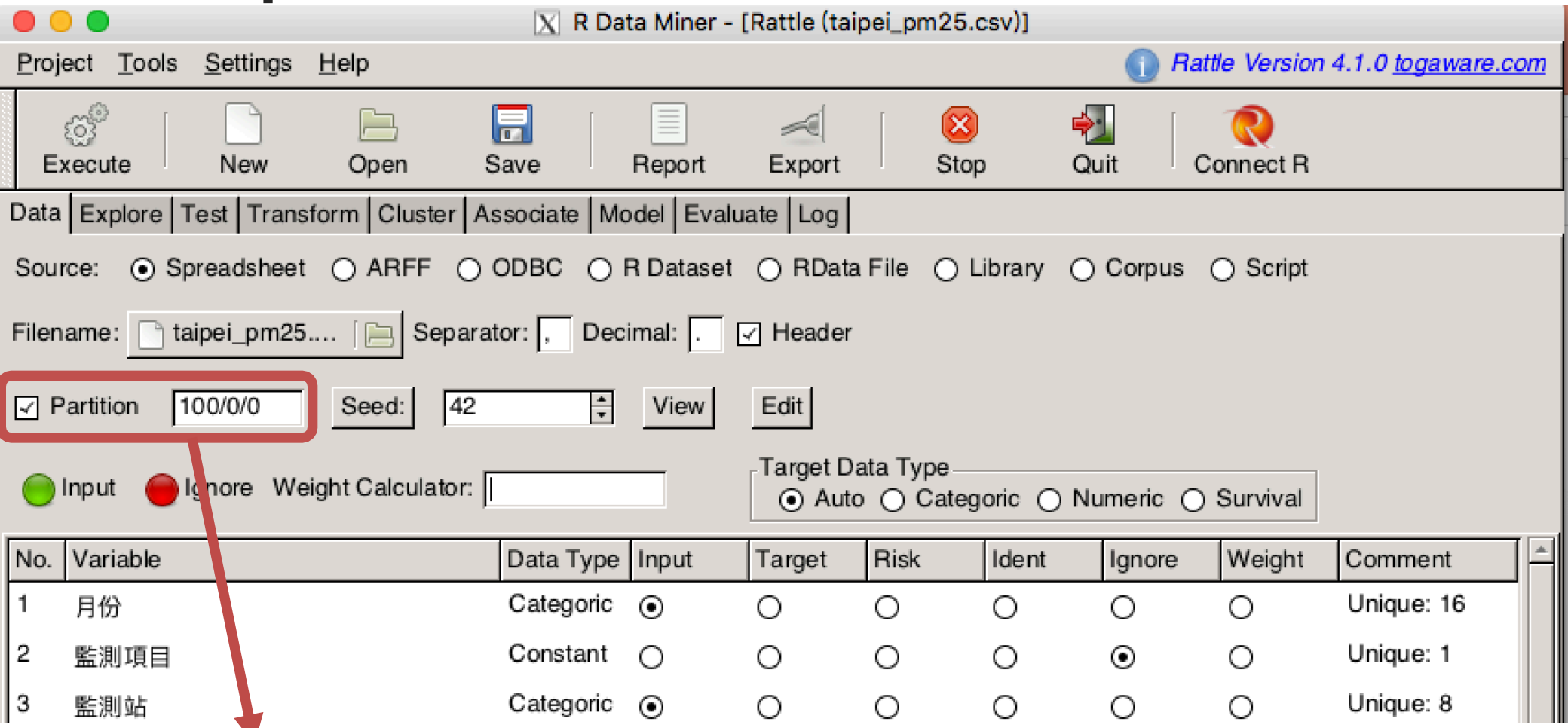


No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	月份	Categoric	◉	○	○	○	○	○	Unique: 16
2	監測項目	Constant	○	○	○	○	◉	○	Unique: 1
3	監測站	Categoric	◉	○	○	○	○	○	Unique: 8
4	測站編號	Categoric	◉	○	○	○	○	○	Unique: 8
5	有效測定日數.日.	Numeric	◉	○	○	○	○	○	Unique: 12
6	測定時數.時.	Numeric	◉	○	○	○	○	○	Unique: 75

Import Data – 懸浮微粒



Import Data – 懸浮微粒



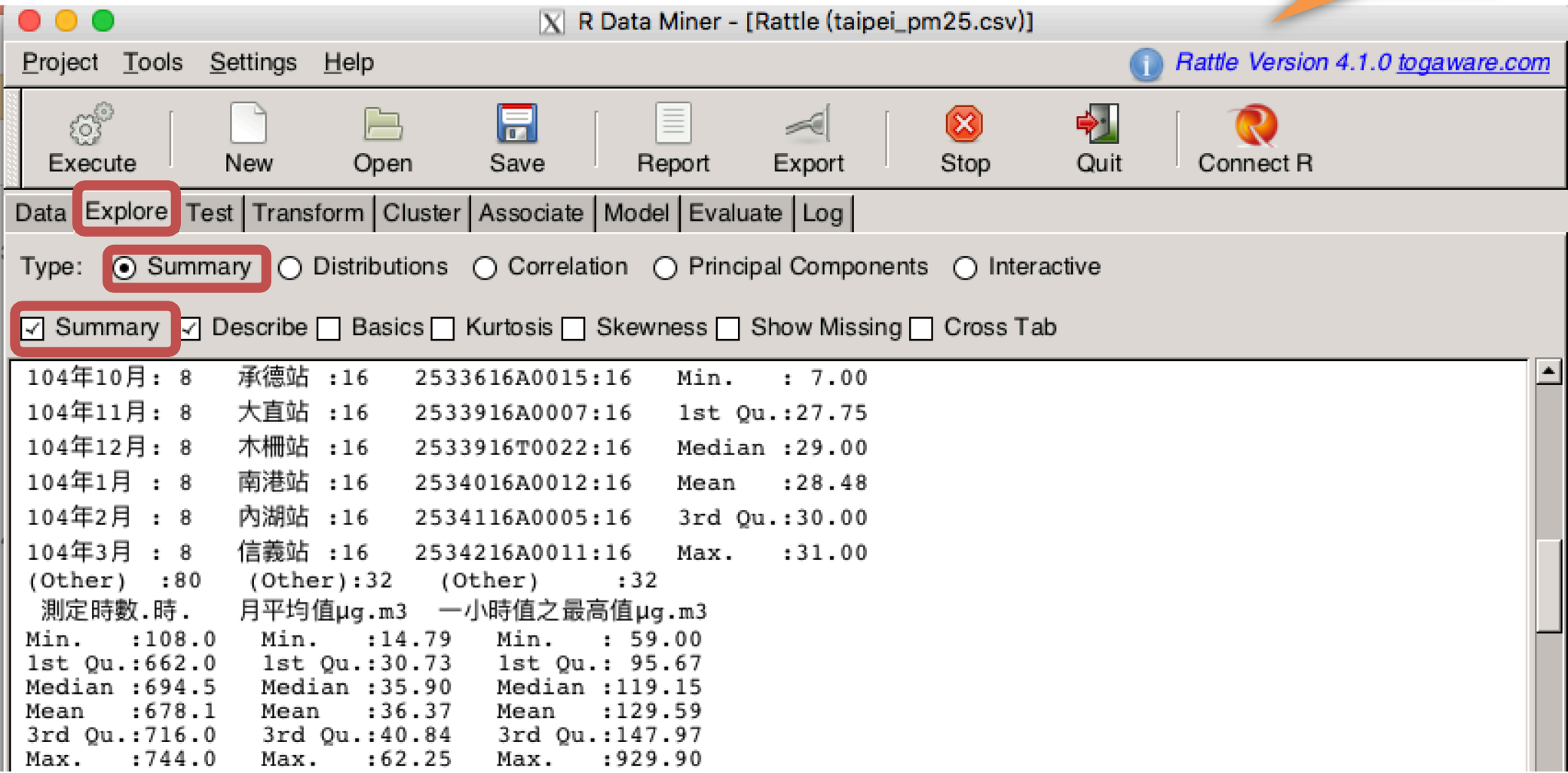
training dataset / validation dataset / test dataset

Kristen Chan

Explore Data – 懸浮微粒

Rattle

Summary



Explore Data – 懸浮微粒

R Studio

Summary

Console ~/

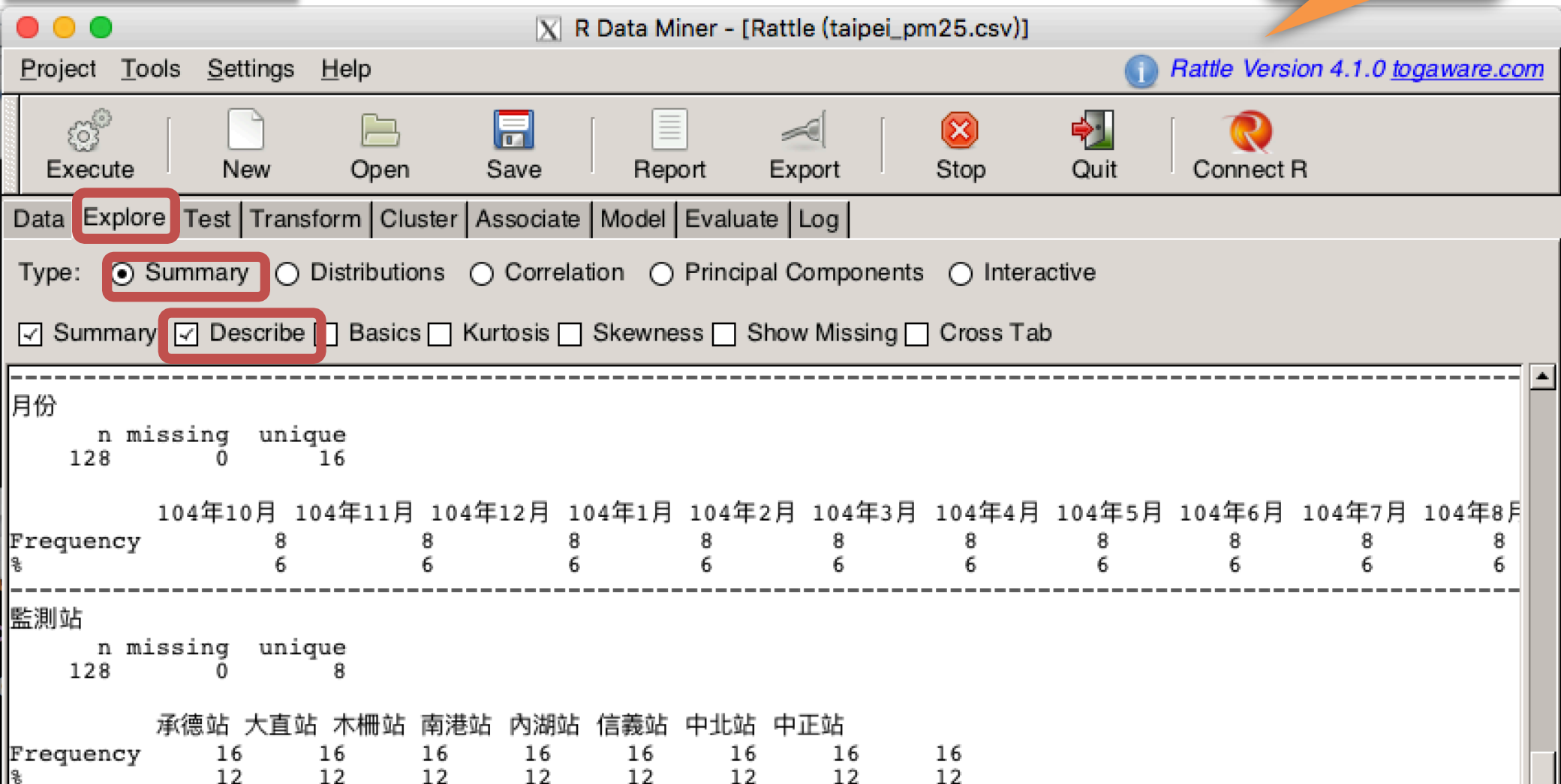
```
> summary(crs$dataset)
        月份              監測項目          監測站                測站編號
 104年10月: 8   懸浮微粒:128   承德站 :16   2533616A0015:16
 104年11月: 8                  大直站 :16   2533916A0007:16
 104年12月: 8                  木柵站 :16   2533916T0022:16
 104年1月 : 8                  南港站 :16   2534016A0012:16
```

Kristen Chan

Explore Data – 懸浮微粒

Rattle

Describe



Explore Data – 懸浮微粒

R Studio

Describe

Console ~/

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

 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
```

Package: Himsc

Explore Data – 懸浮微粒

Rattle

Missing Value

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μg.m3 一小時值之最高值μg.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μg.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200μg.m3之小時時數.時.
[1,]                          1                                  1
[2,]                          0                                  0
```

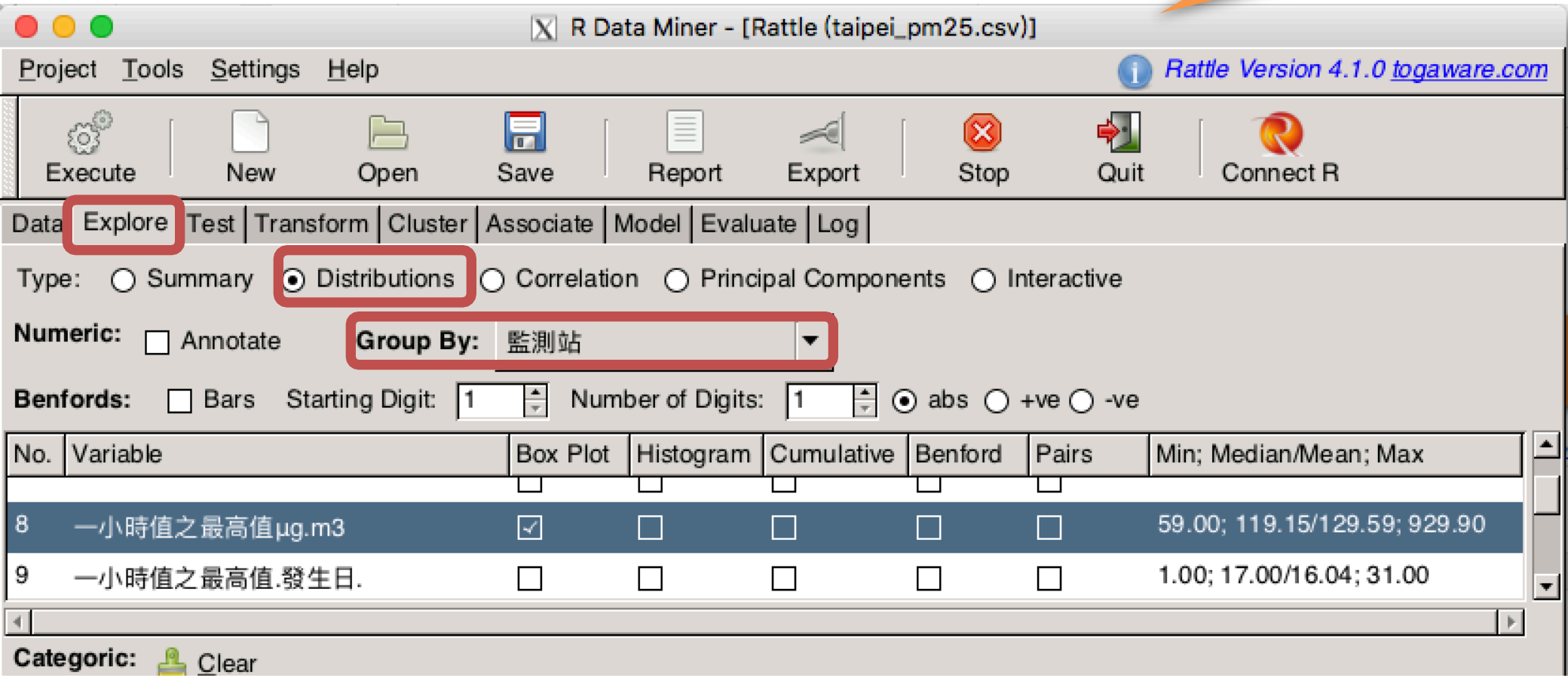
Package: mice

Kristen Chan

Explore Data – 懸浮微粒

Rattle

Boxplot

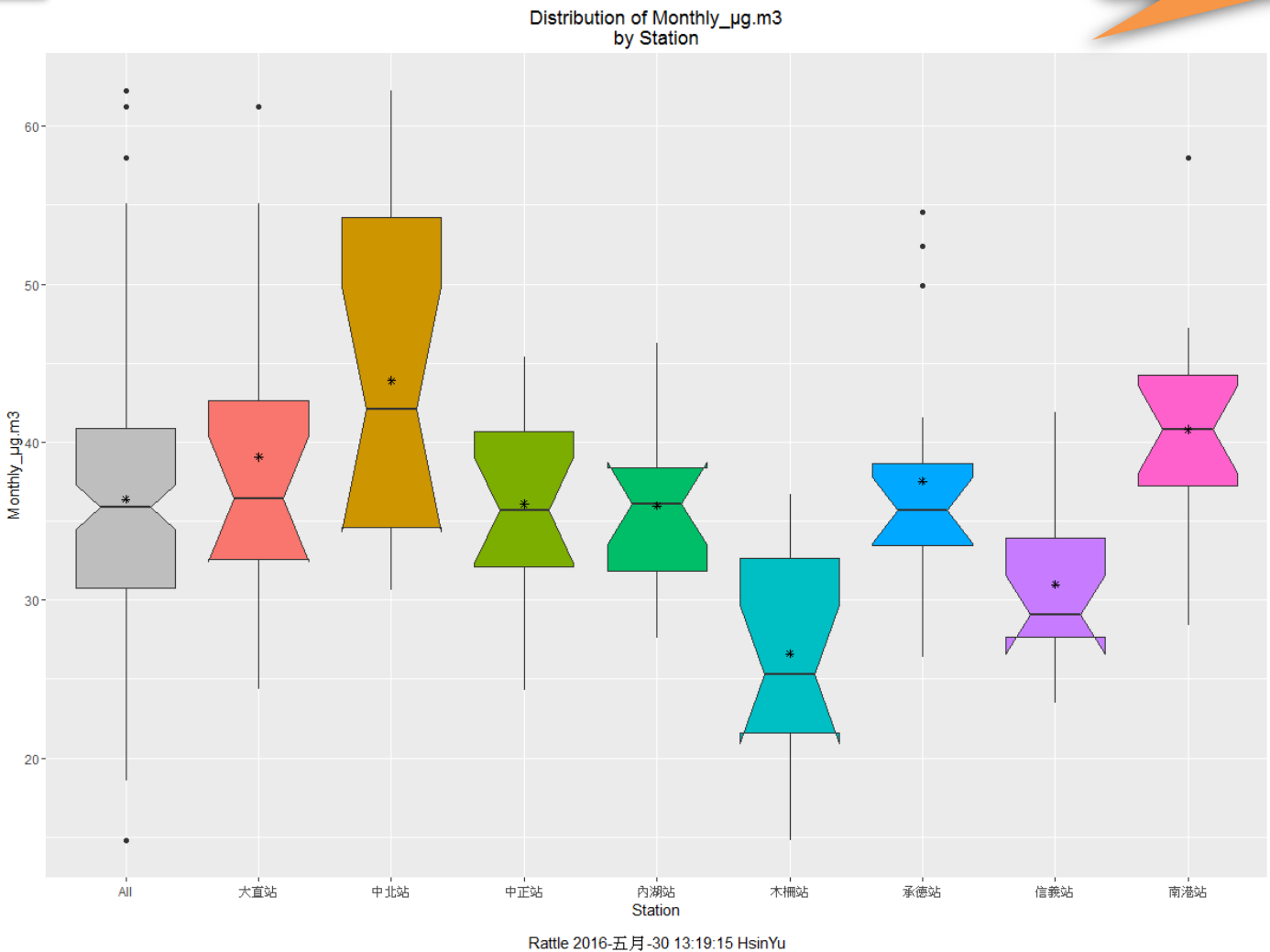


Kristen Chan

Explore Data – 懸浮微粒

Rattle

Boxplot



Kristen Chan

鳶尾花(iris)

資料筆數：150

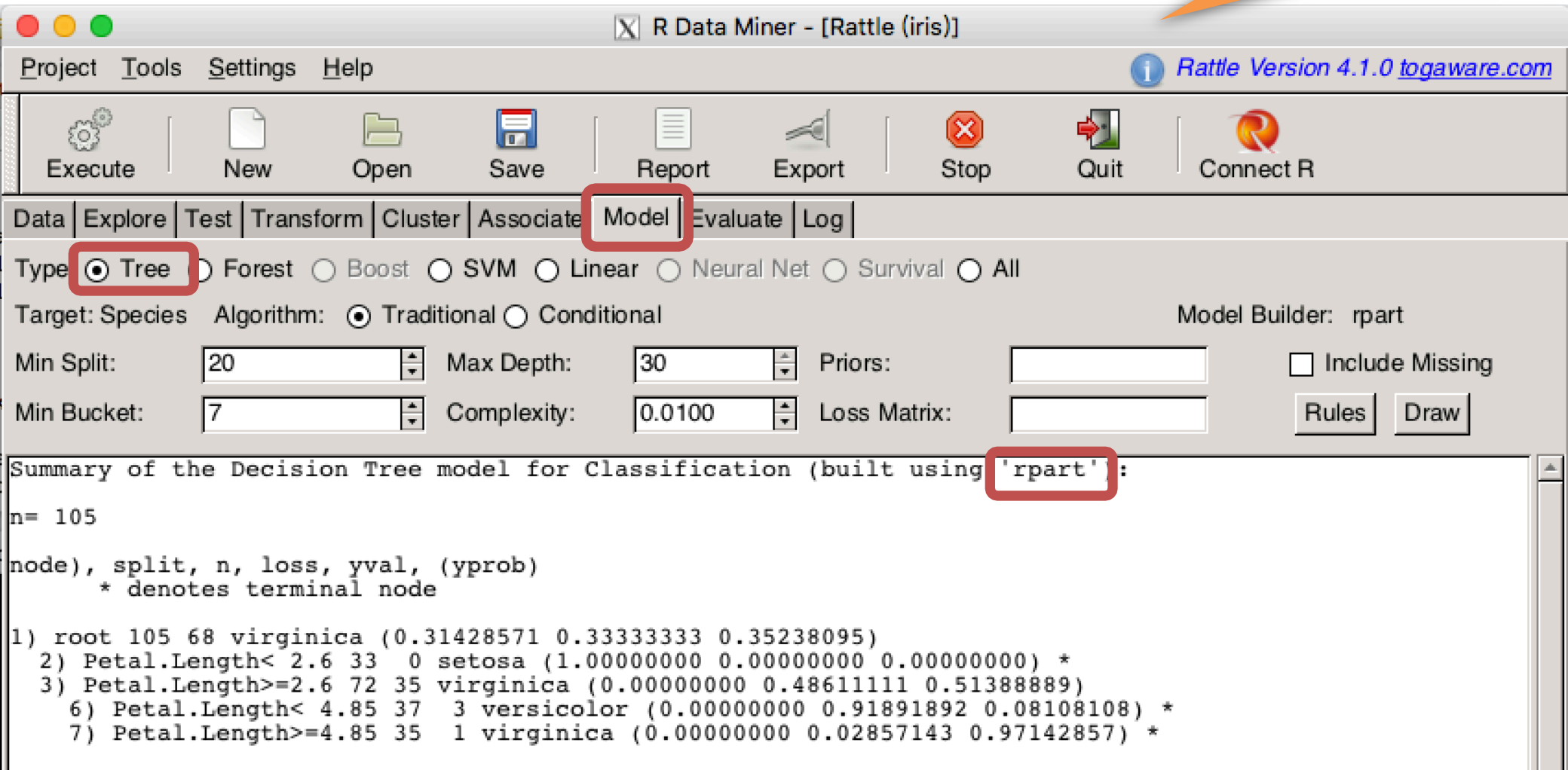
Sepal Length – 花萼長度<cm>
Sepal Width – 花萼寬度<cm>
Petal Length – 花瓣長度<cm>
Petal Width – 花瓣 寬度<cm>
Species – iris 種類< Setosa、Versicolor、Virginica >

Kristen Chan

Model

Rattle

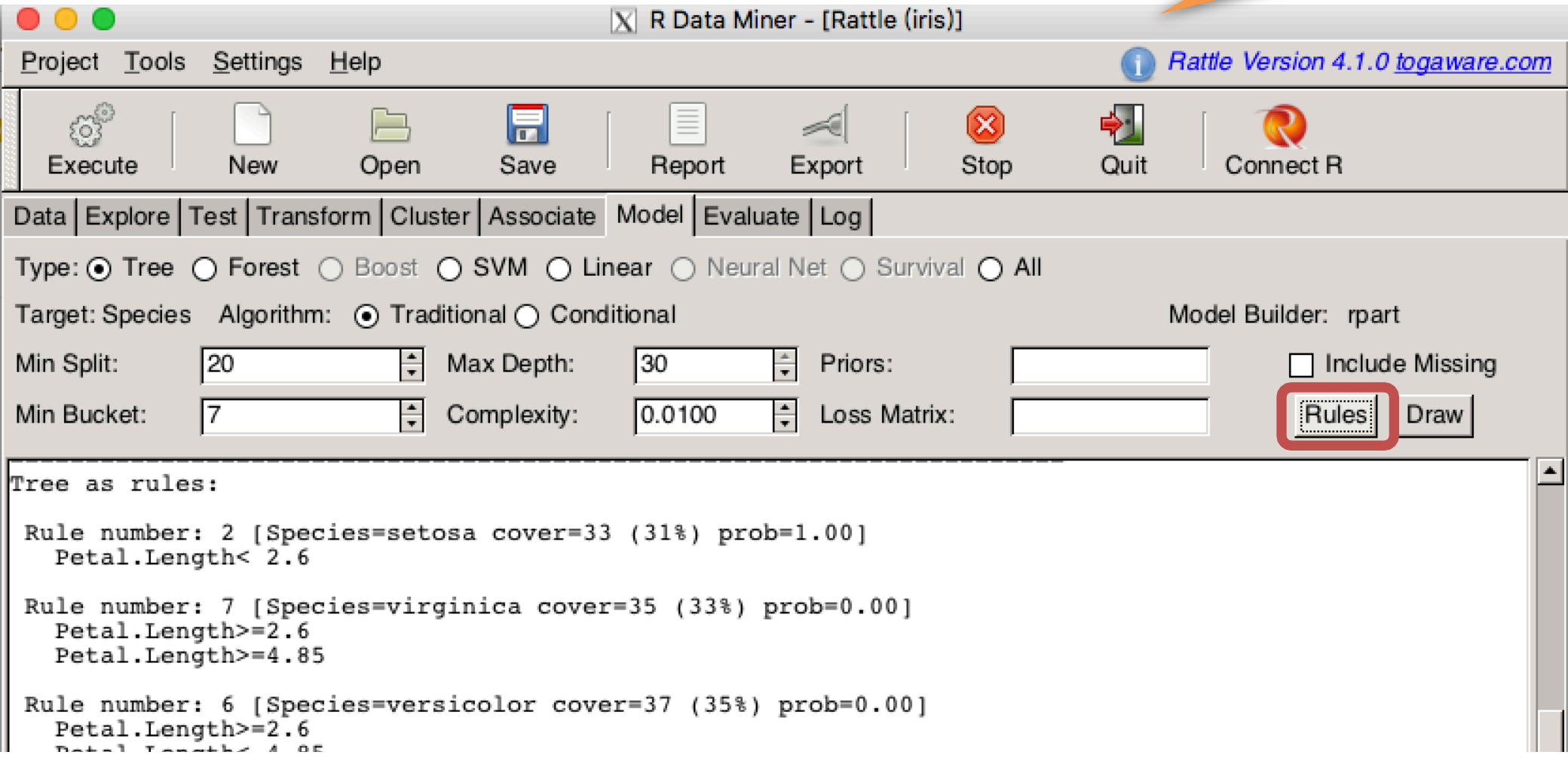
Decision Tree



Model

Rattle

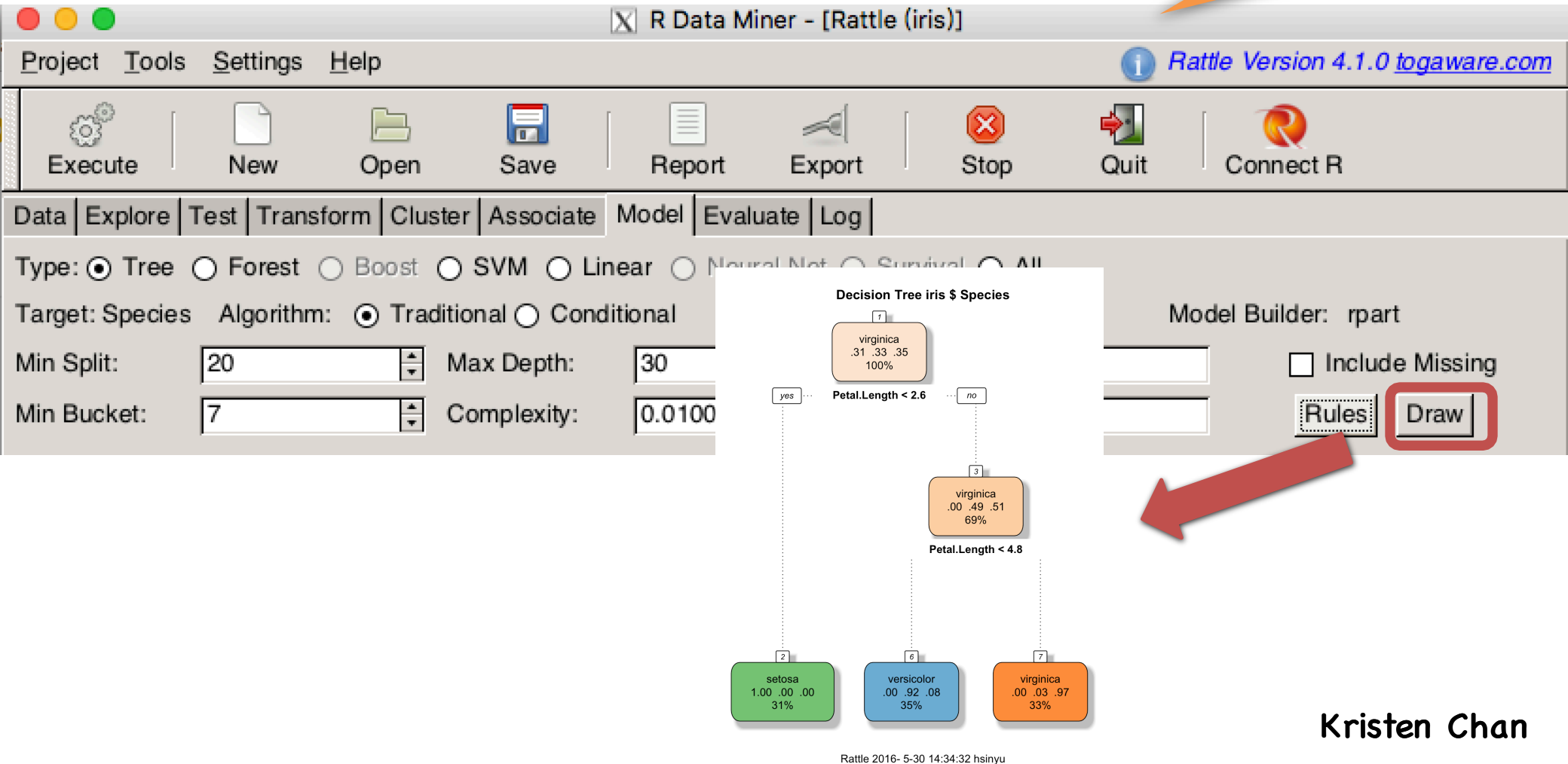
Decision Tree



Model

Rattle

Decision Tree



Kristen Chan