@LI Jie

Edited 2019.10.12

Contents

Download and Install R 1
Download and Install Rstudio 1
Download and Install Rattle package 1
Start up rattle 2
Simple Sample: 2
Common Data Mining Workflow 8

Download and Install R

https://cran.rstudio.com/

Download and Install Rstudio

https://rstudio.com/products/rstudio/download/#download

Download and Install Rattle package

Rattle (Williams, 2011) is a package written in R providing a **graphical user interface** to very many other R packages that provide functionality for data mining.

The packages will usually be installed with the following command:

```
> install.packages("RGtk2")
```

```
> install.packages("rattle", dependencies=c("Depends", "Suggests"))
```

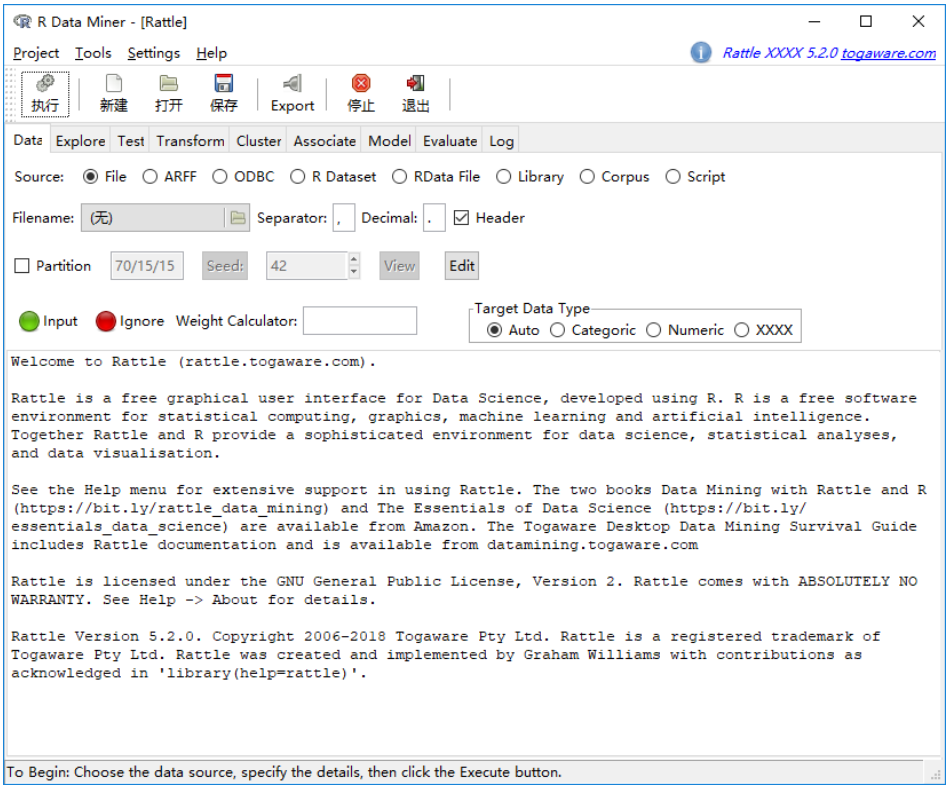
Then wait many minutes. (a very very long time)

Start up rattle

> library(rattle)

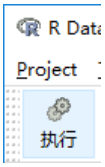
```
> library(rattle)
Rattle: A free graphical interface for data science with R.
XXXX 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.
键入'rattle()'去轻摇、晃动、翻滚你的数据。
>
```

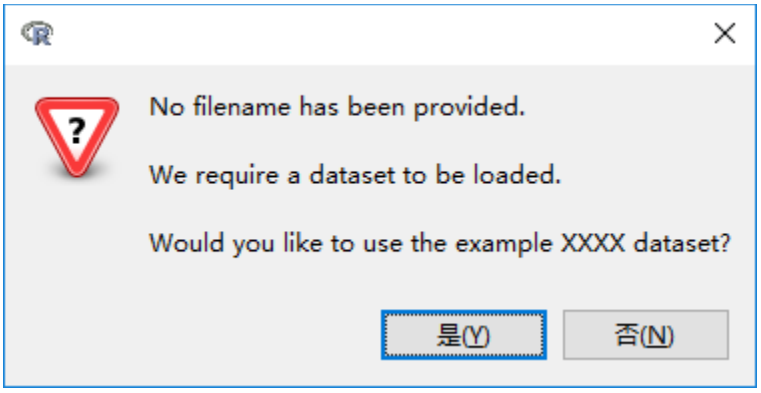
> rattle()



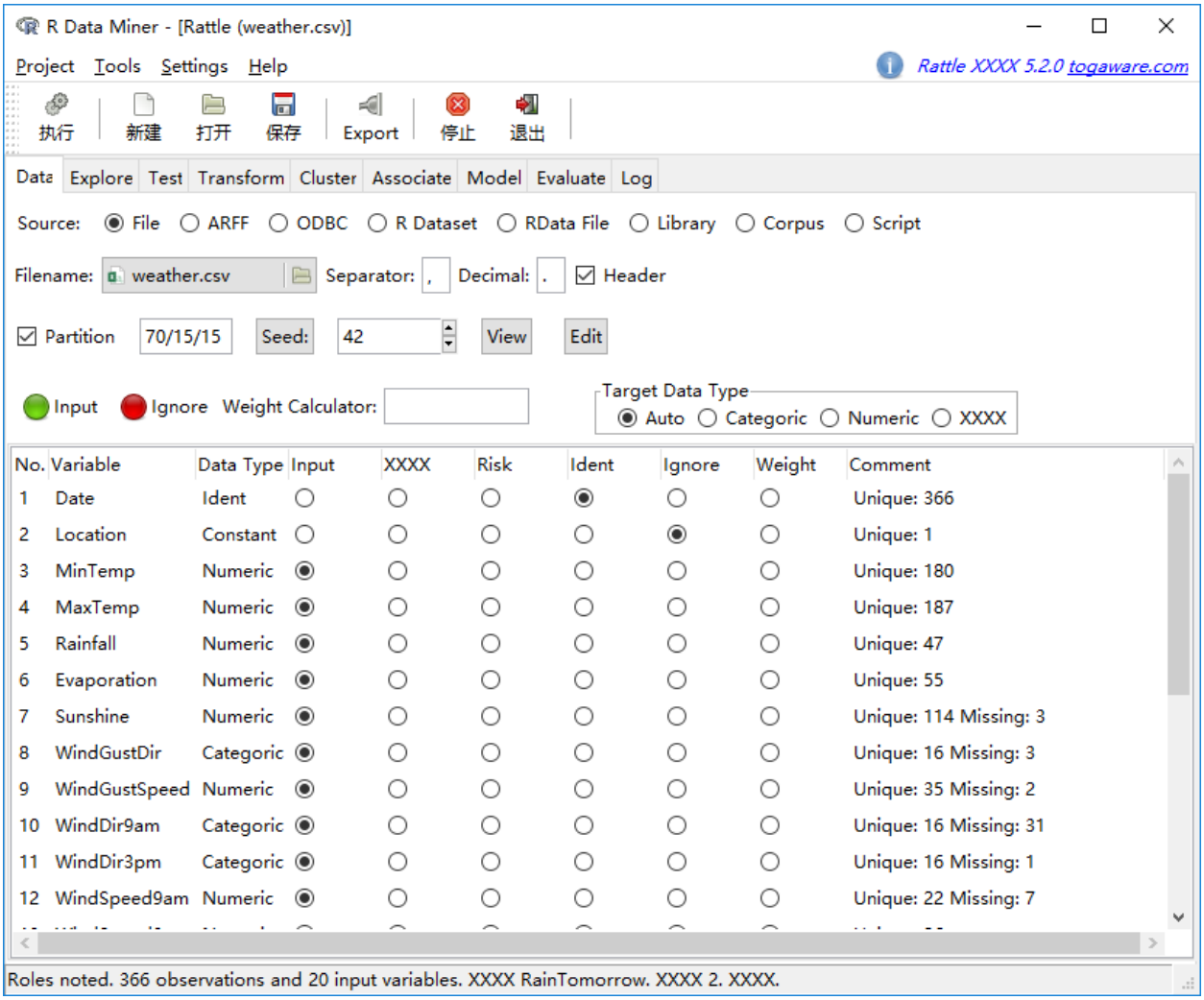
Simple Sample:

1.Click on the **Execute(执行)** button;

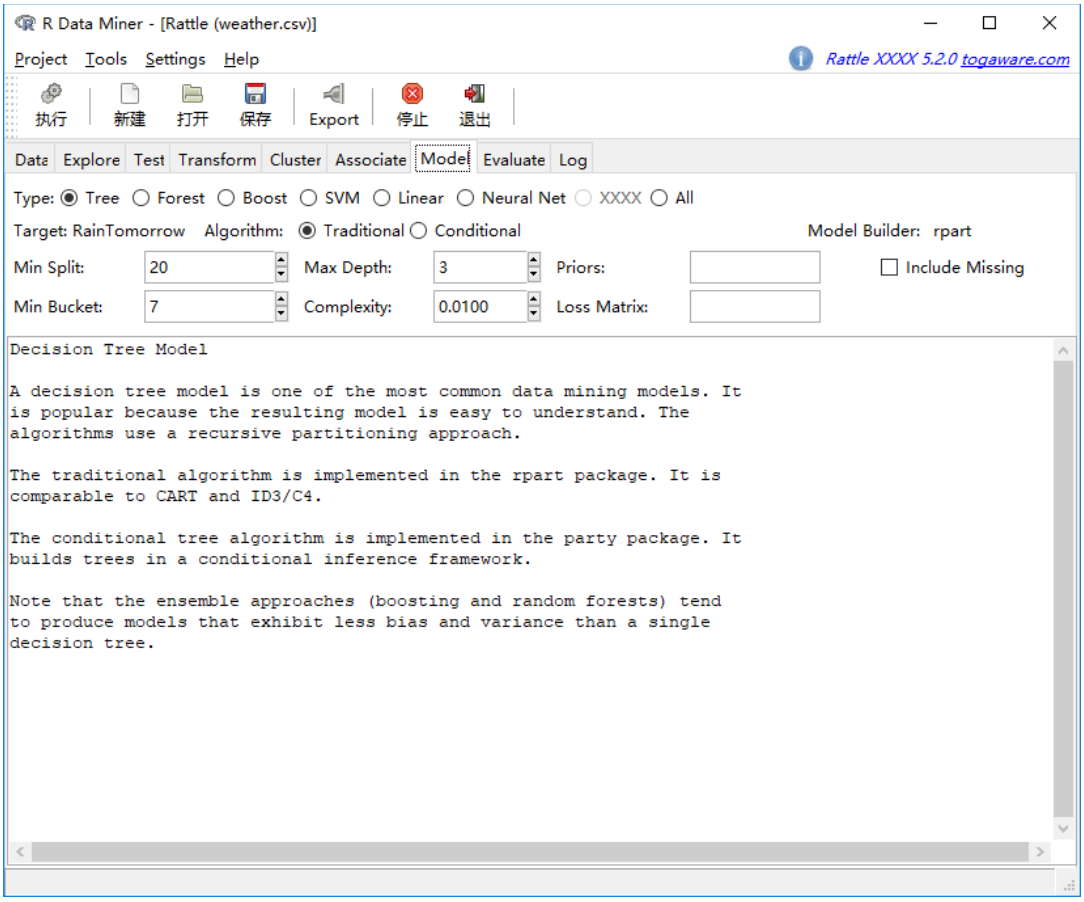




2. Click on **Yes（是）** within the resulting popup; (load the sample weather dataset)



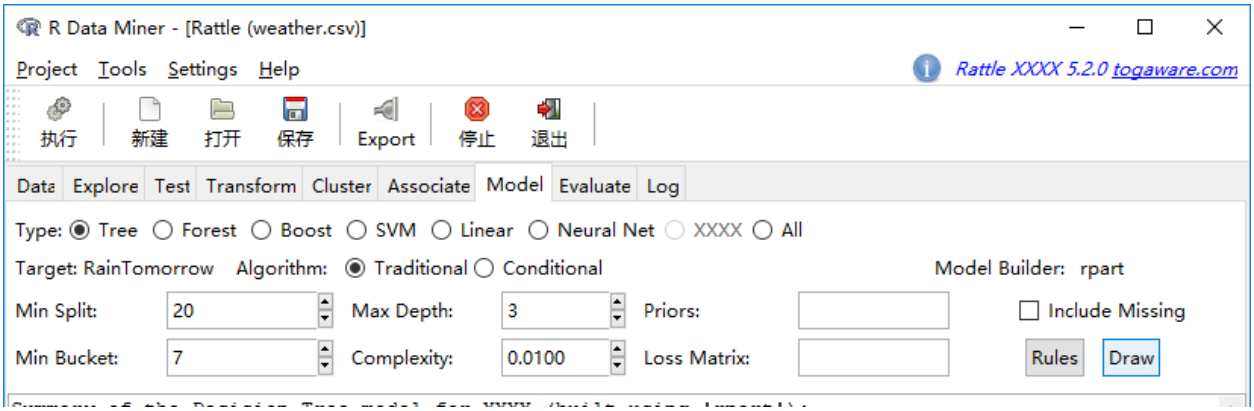
3. Click on the **Model** tab;



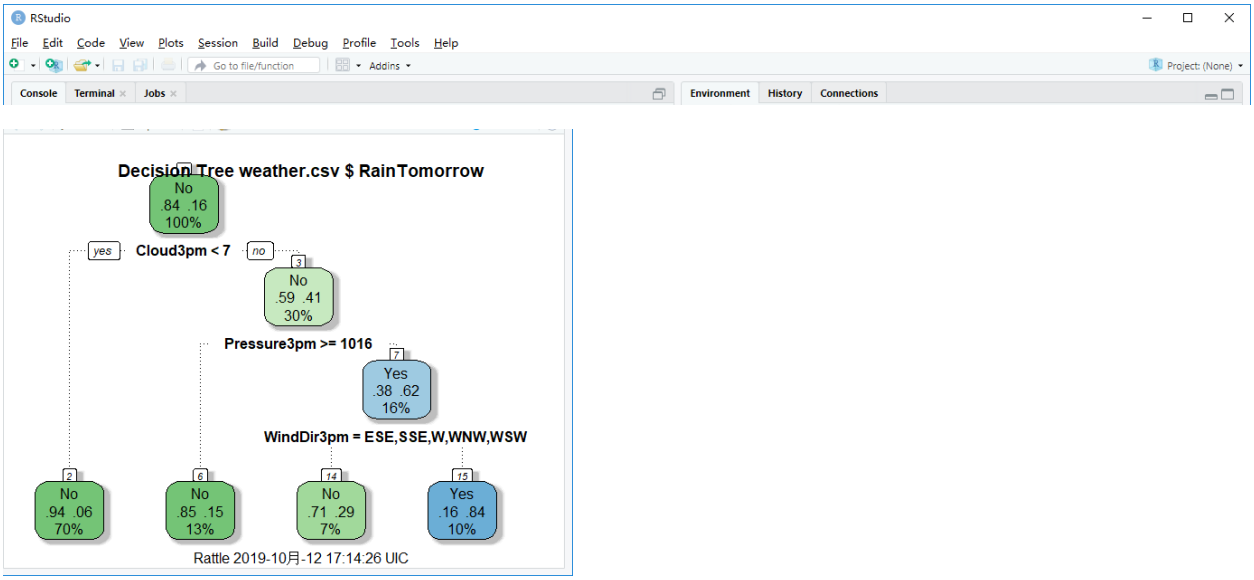
4. Click on the **Execute** button. (to build a decision tree)

Now we have a decision tree built from a sample classification dataset.

5. Click **Draw** to display the decision tree

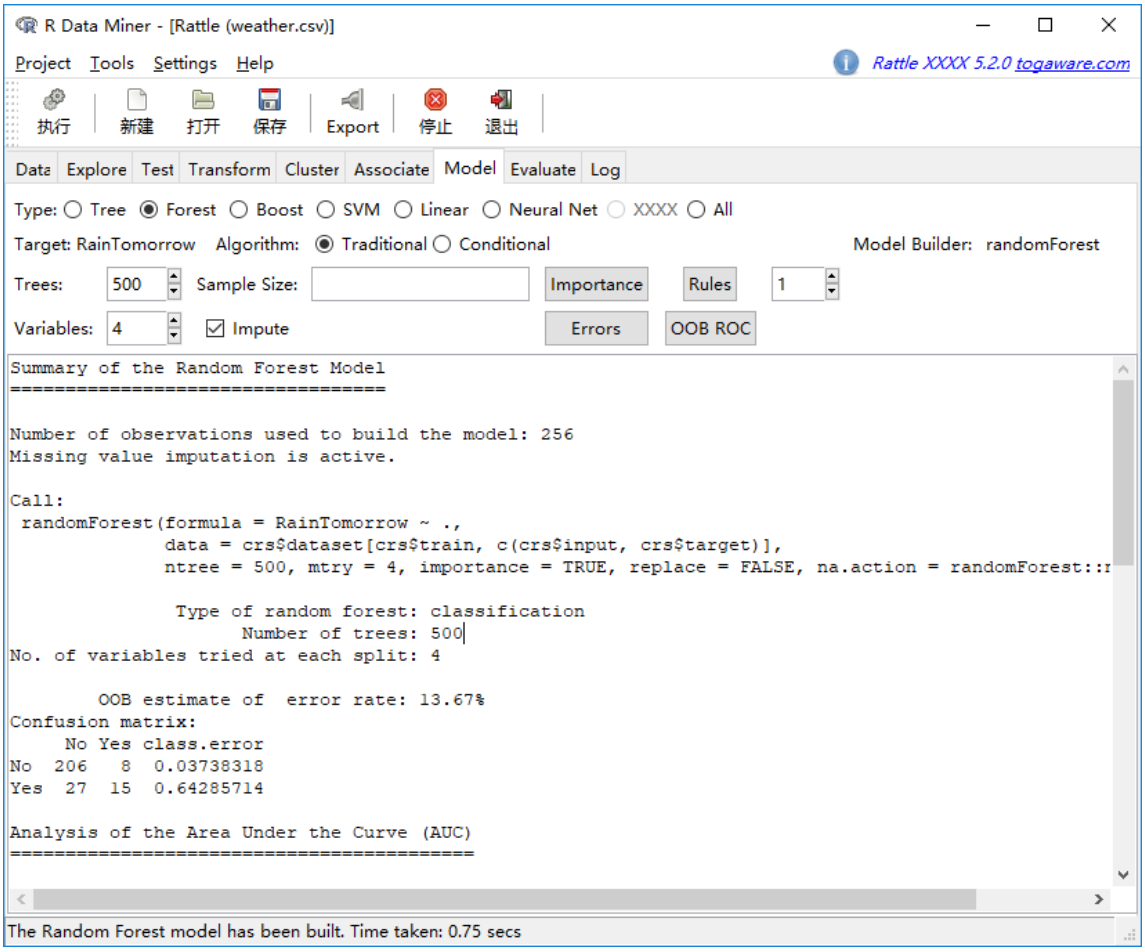


The picture will be showed in the right bottom of the original RStudio Window

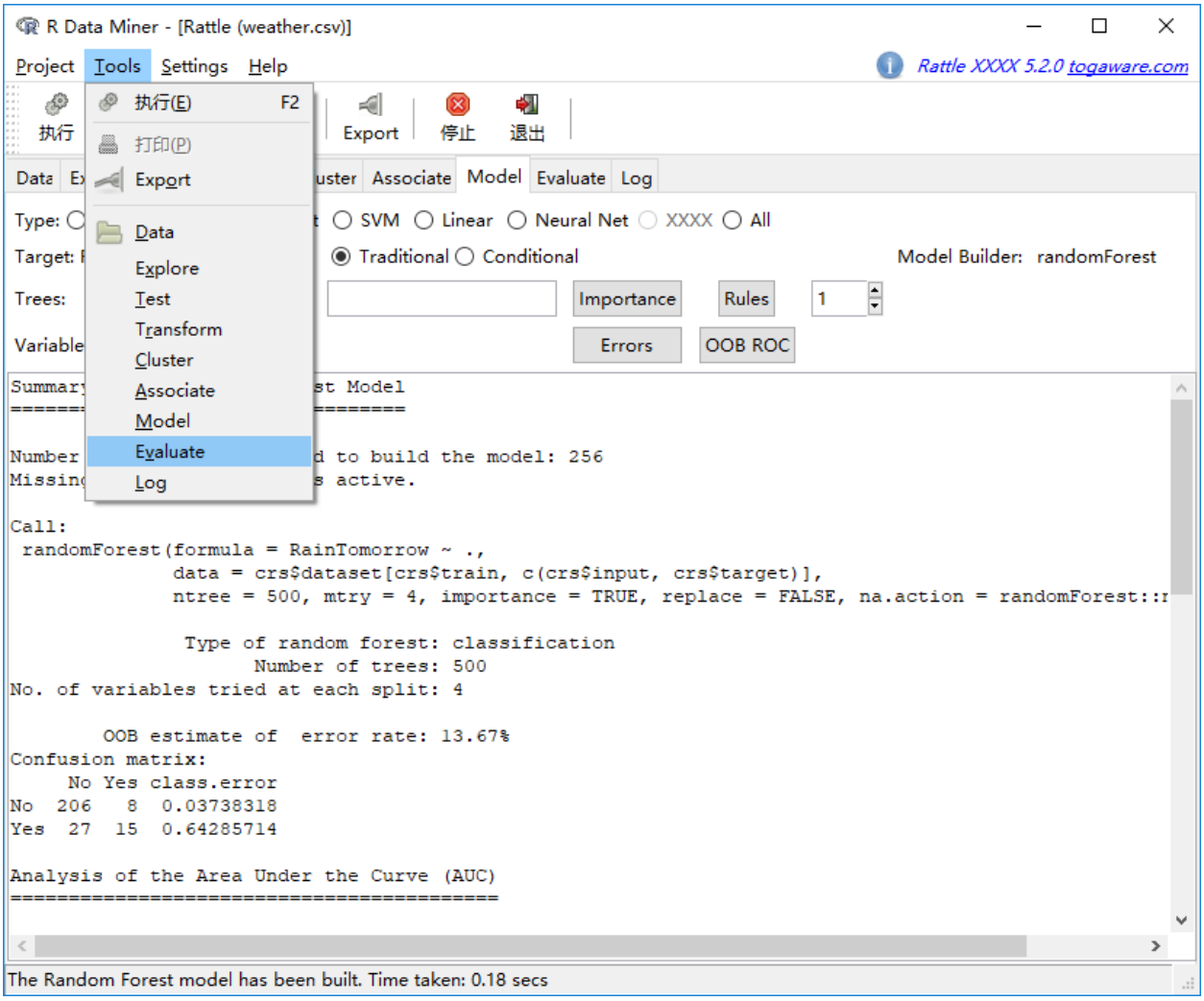


6. Click on the **Forest** radio button

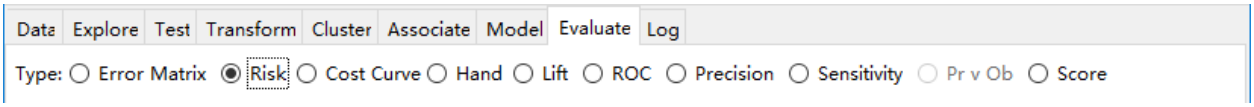
7. Click on **Execute** (to build a random forest)



8. Click on the **Evaluate** tab in **Tools**

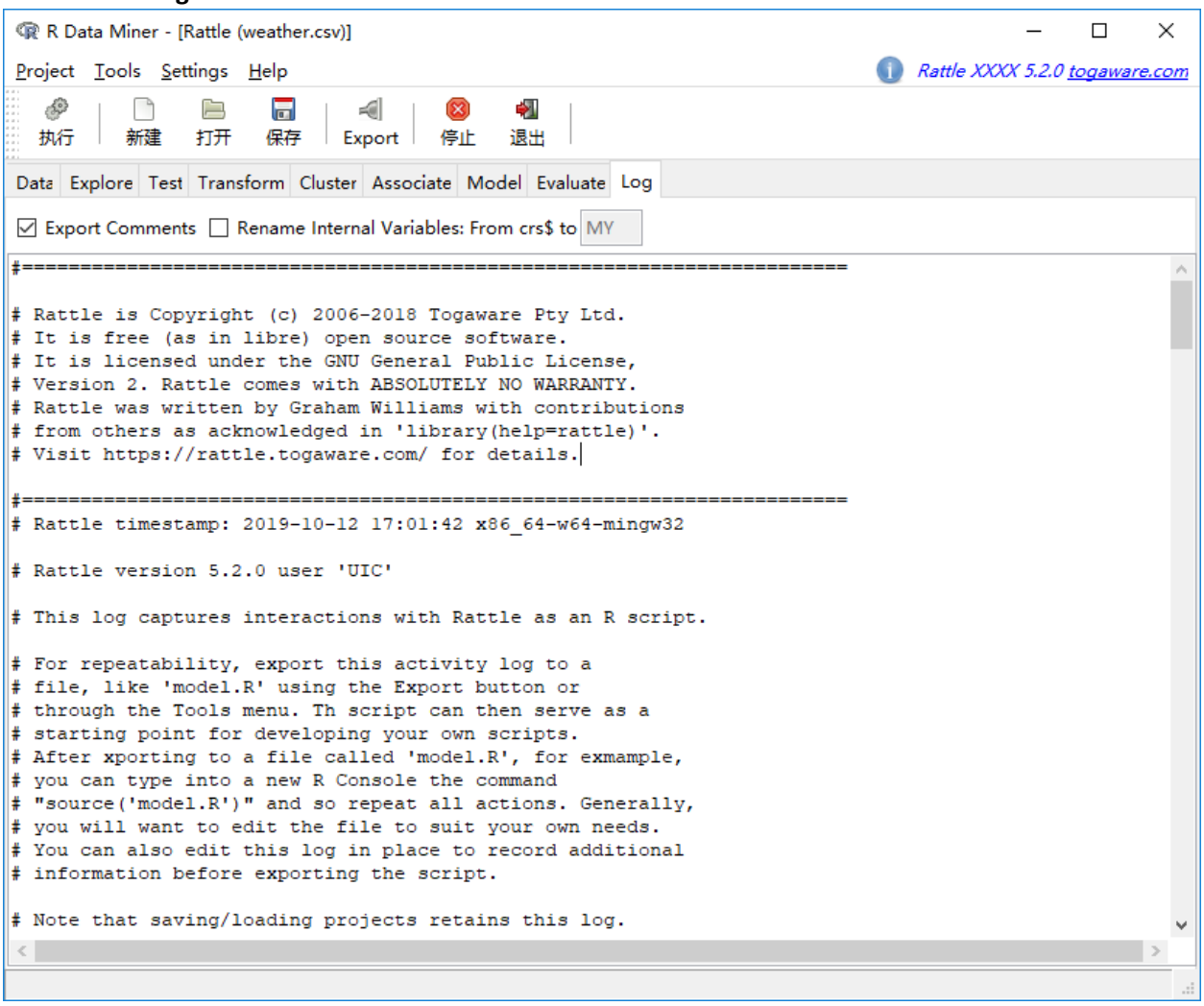


9. Choose the **Risk** button

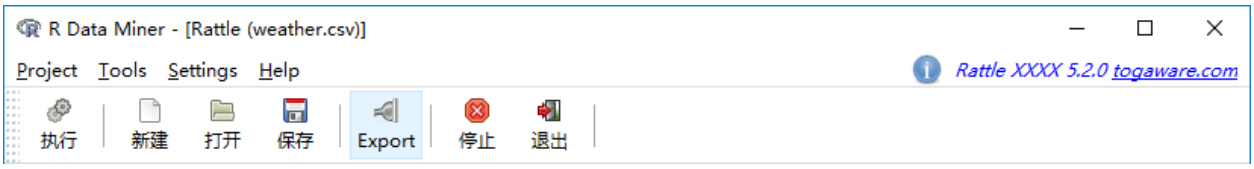


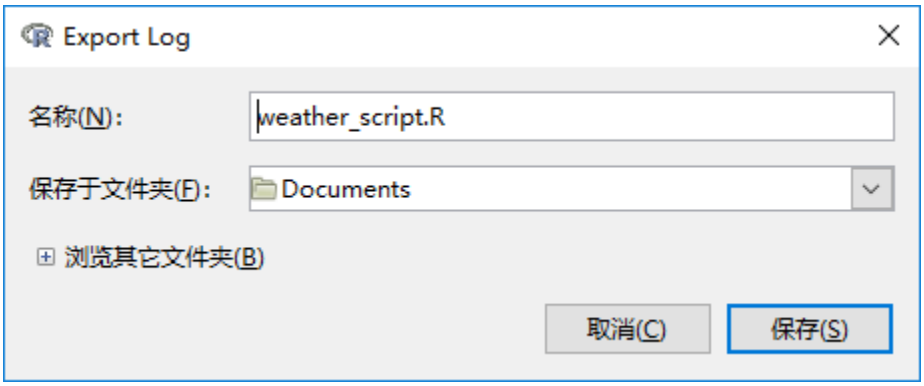
10. Click on **Execute** to display two Risk (Cummulative) performance plots

11. Click the **Log** tab.



Click the **Export** button to save script to file weather script.R to home folder Now exit from R (and rattle) and start R up again.





> source("~/weather_script.R")

This will rerun everything that was done in the GUI session but purely as a script.

Common Data Mining Workflow

The common work flow for a data mining project can be summarised as:

1. Load a Dataset and select variables;
2. Explore the data to understand distributions;
3. Test distributions;
4. Transform the data to suit the modelling;
5. Build Models;
6. Evaluate models and score datasets;
7. Review the Log of the data mining process