WMCS16001.2019-2020.1B
Information Systems

Association Rule Analysis (Market Basket Analysis) Assignment

Dr. George Azzopardi

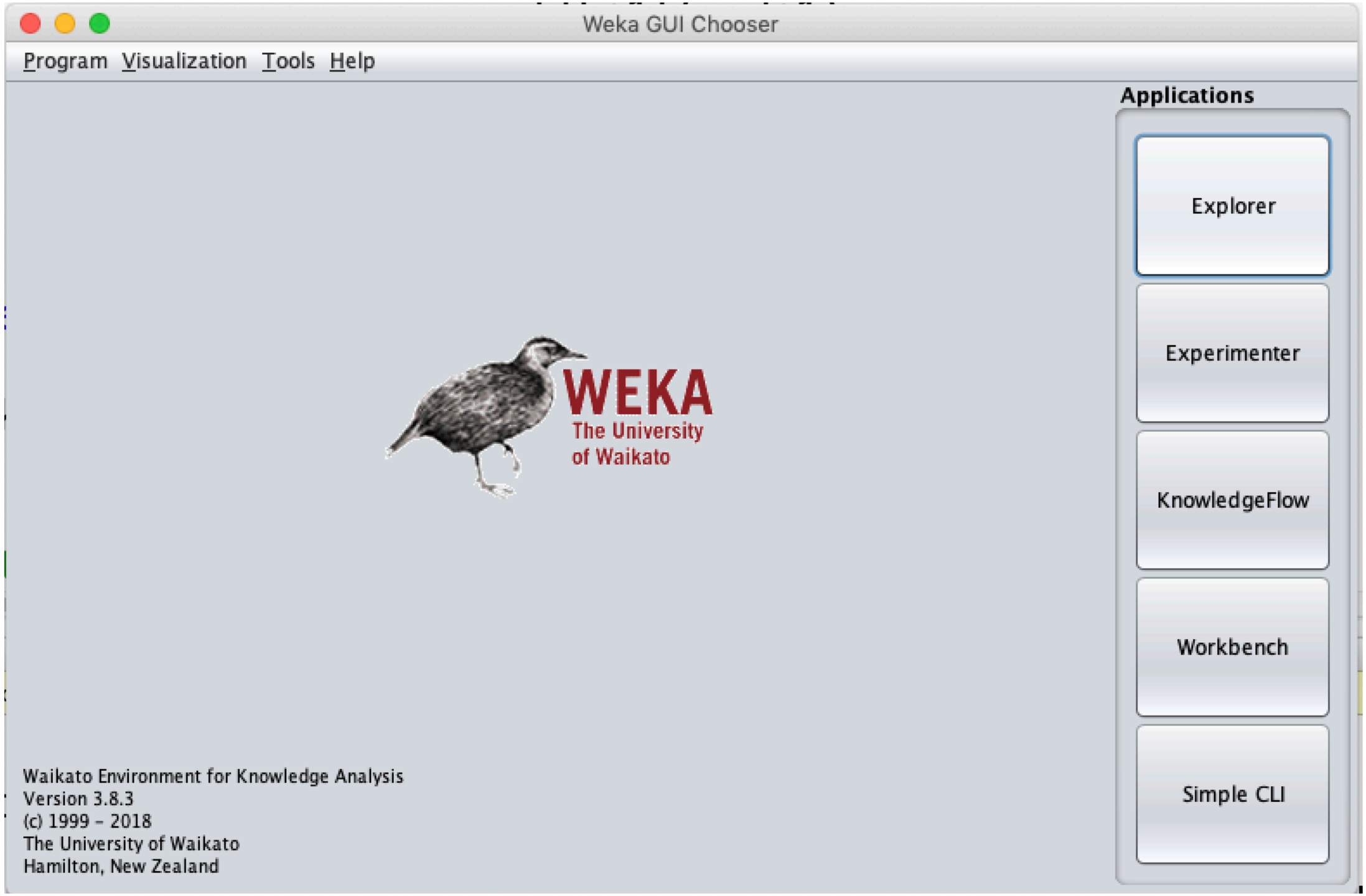
Groceries data set

- The file groceries.txt contains 9835 rows with comma separated items
- Each row can be considered as a receipt with a transaction of multiple grocery items

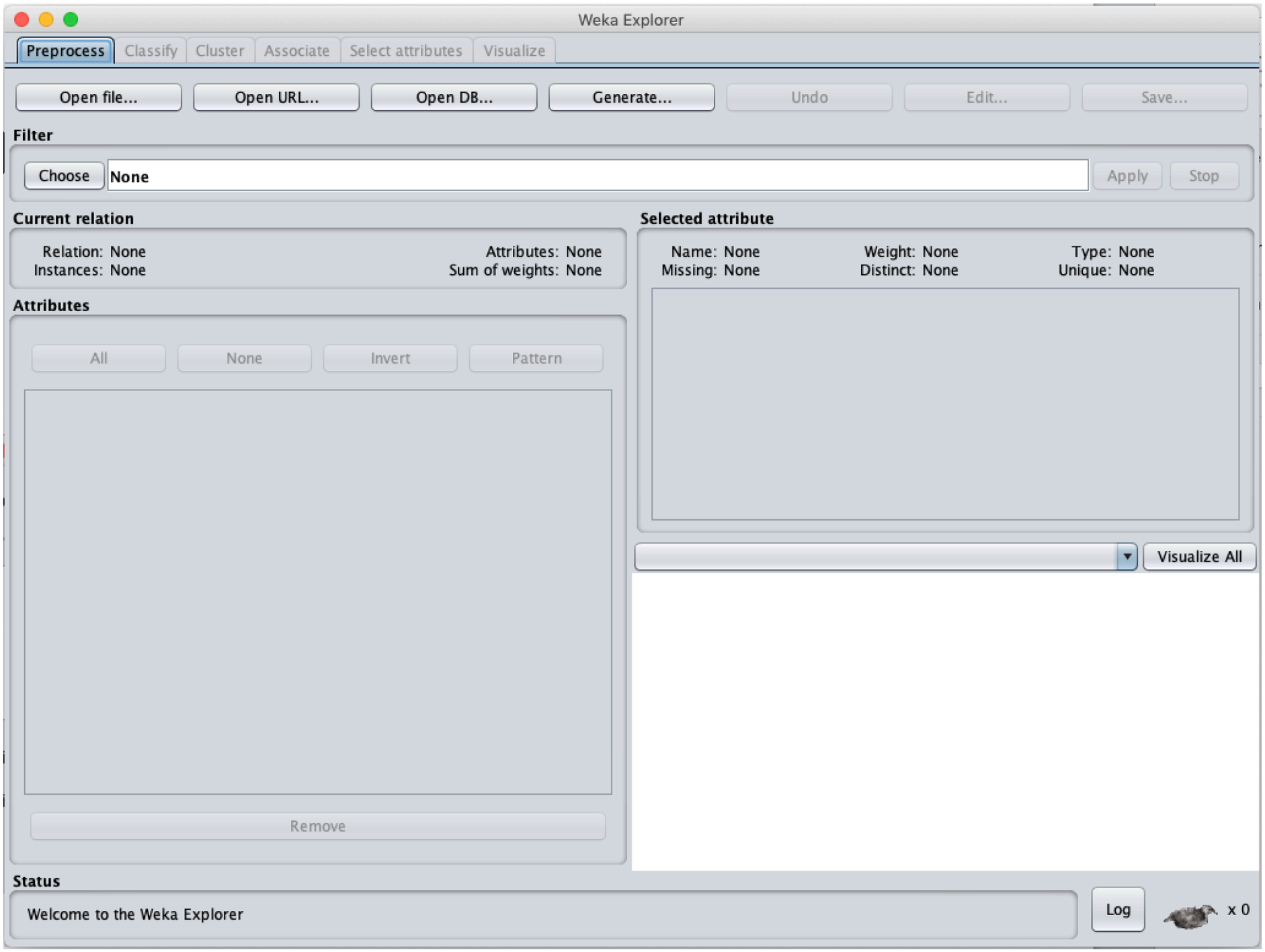
Tasks

1. Show the histogram of all items in the 9835 transactions **[0.5 pts]**
2. Use the given Matlab implementation (associationRules.m) and add to it the code where you determine the association rules from the determined frequent item sets. Your implementation must include the anti-monotone property in slide 44. **[7 pts]**
3. Compare the time needed to generate all association rules when you implement the anti-monotone property in slide 44 versus when such property is ignored. **[1 pt]**
4. How many rules satisfy the following input parameter values? ***support = 0.001, confidence = 0.8.*** Sort the association rules by confidence in descending order and list the top 30 rules. **[0.5 pt]**
5. You can use the weka tools (https://www.cs.waikato.ac.nz/ml/weka/) to check the results that you obtain. See next slides on how to use Weka.
6. Deliverable:
    1. Complete Matlab script that include the code requested in point 2 above.
    2. A report that include the answers to the rest of the questions. **[1 pt for clarity]**

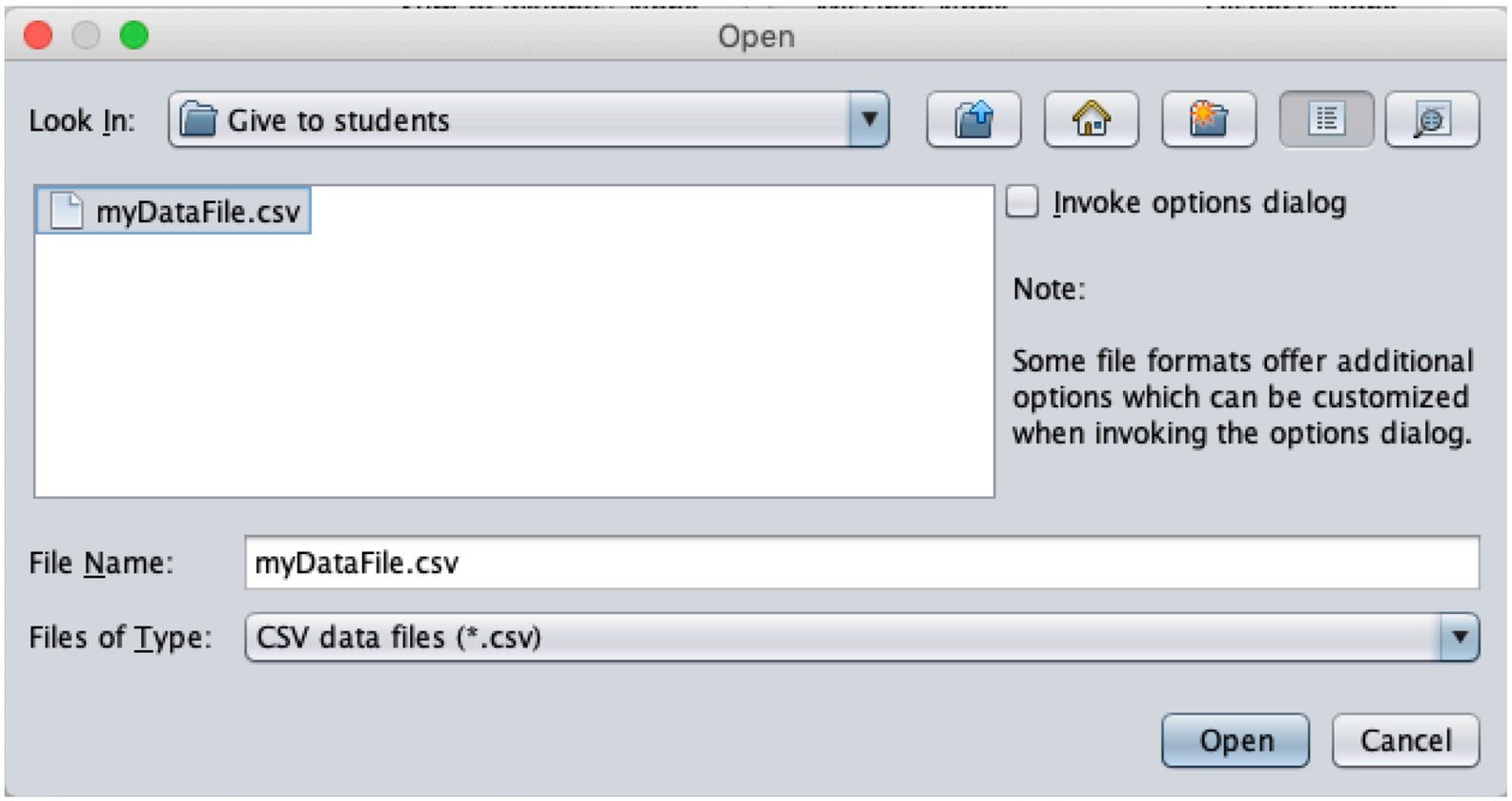
Click on Explorer



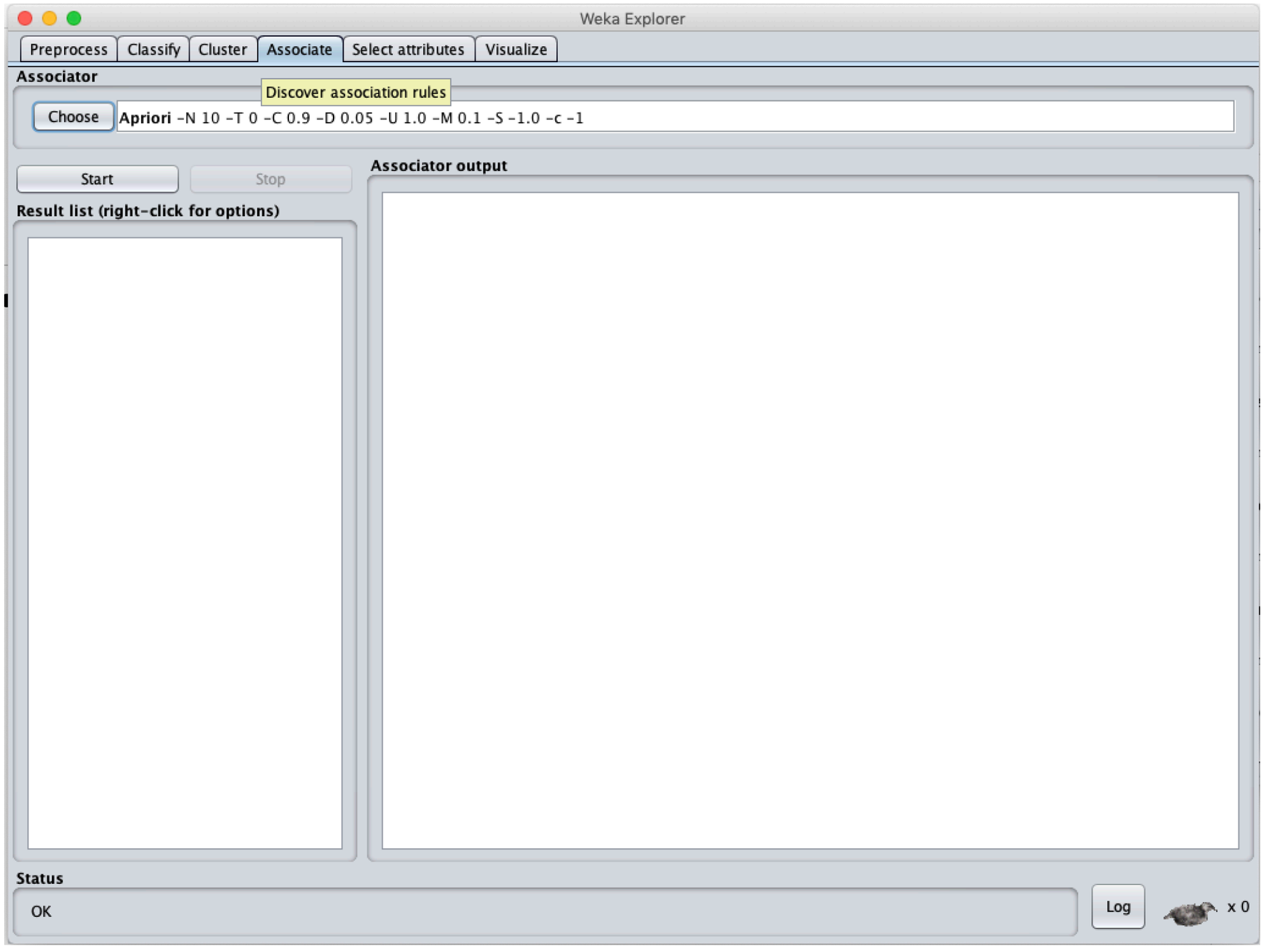
Click on Open File



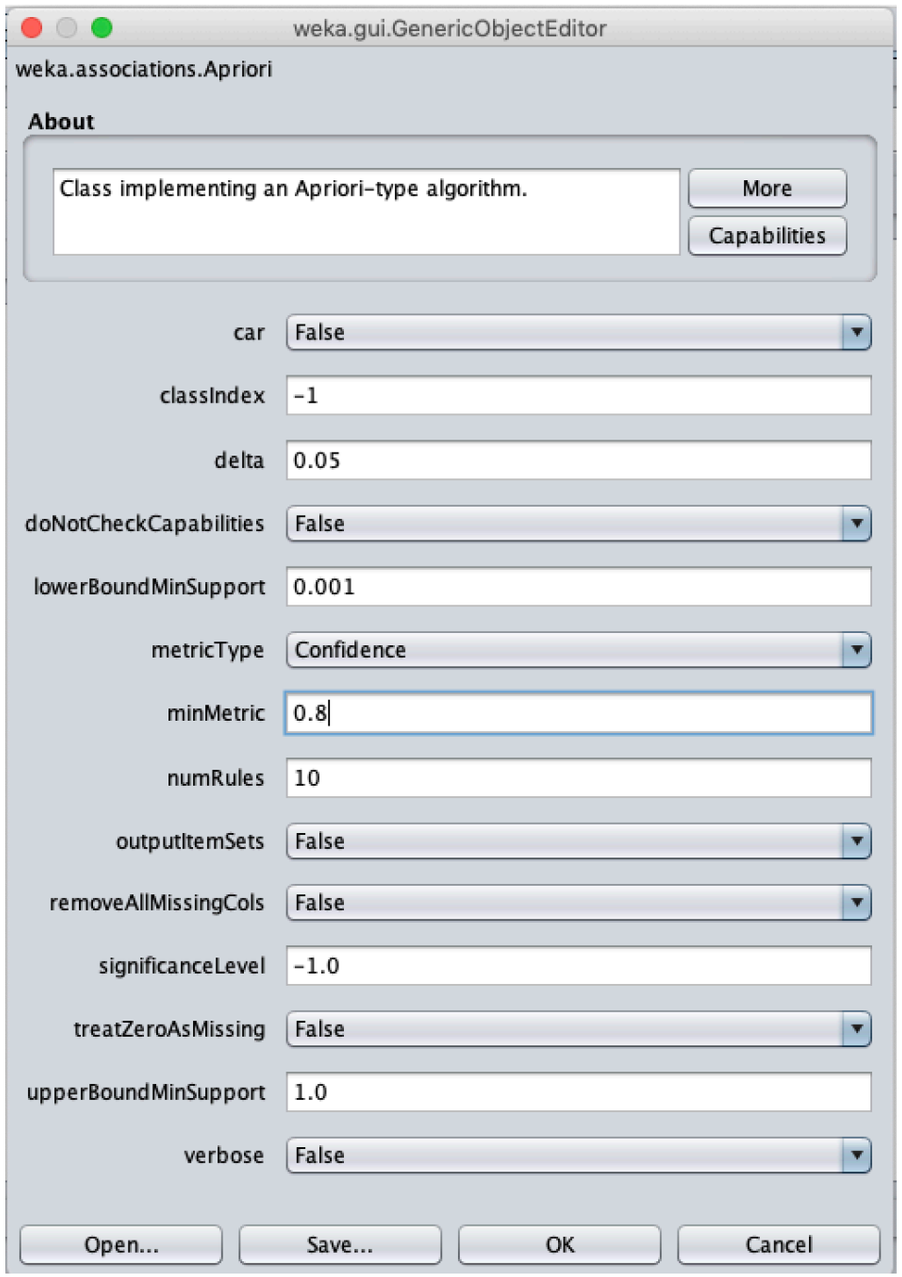
Open the provided myDataFile.csv



Click on the Associate tab, and click in the field next to the Choose button



Set the minimum support and minimum confidence



Finally, press start