



RATTLE-GUI

UN PACKAGE R DE DATAMINING

26/10/2016

Mathilde EVIN
Emilien FROMAGE
Nirasay SISOMBAT
Matthieu ZAMBEAU

SOMMAIRE

- Introduction
- Présentation de Rattle
- Evaluation
- Conclusion
- Démonstration

INTRODUCTION

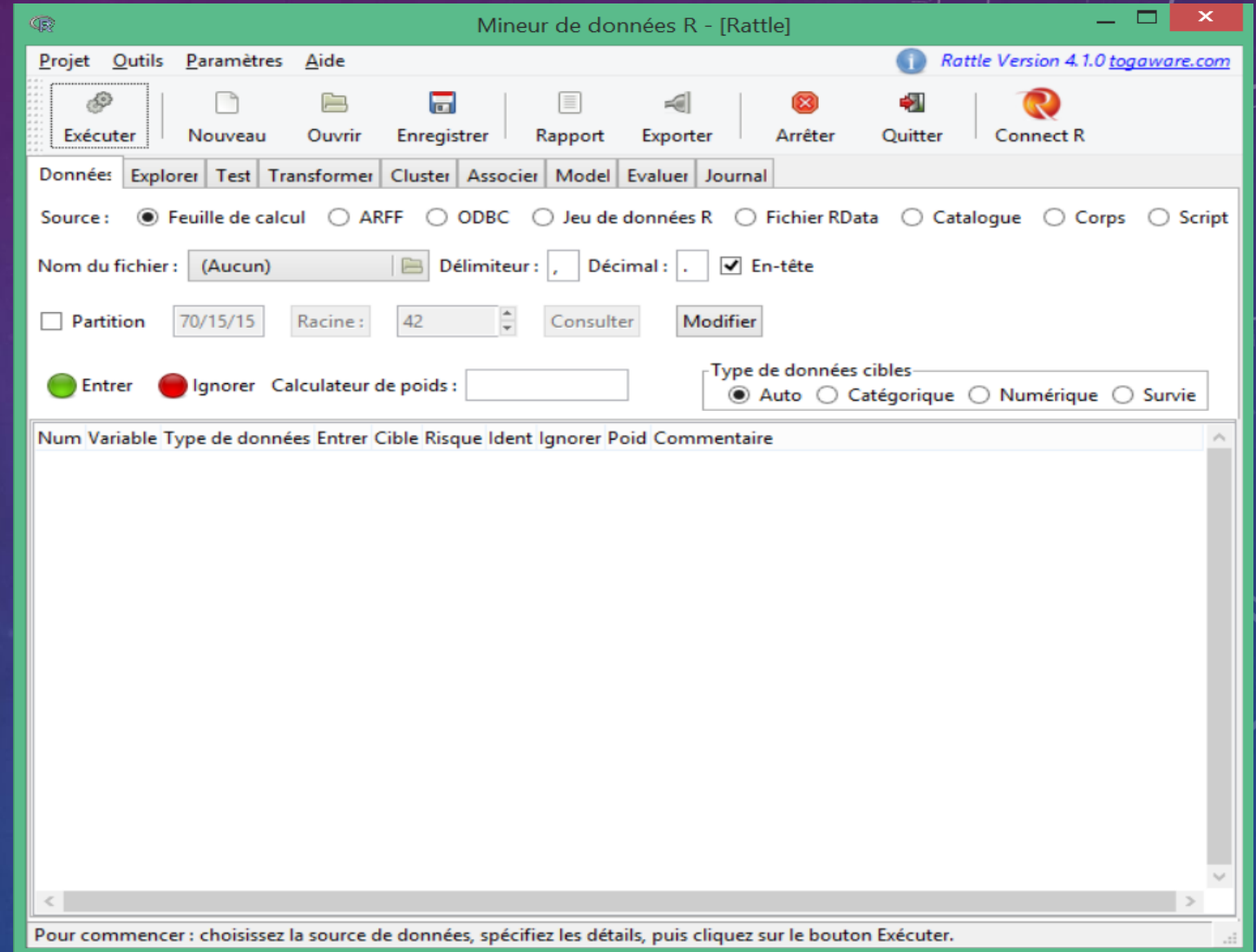
- Logiciel libre de datamining
- RATTLE : R Analytic Tool To Learn Easily
- GUI : Graphical User Interface
- Package accessible sous R et RStudio

HISTORIQUE

- Développé par le Dr. Graham Williams dans les années 2000
- Version finale sortie en janvier 2013
- Régulièrement mis à jour depuis (version actuelle : 5.0.2)

UN PREMIER APERCU

- Logiciel de clic-bouton
- Différents onglets de navigation
- Bouton Exécuter pour lancer la tâche ou pour modifier un paramètre



INSTALLATION DE RATTLE

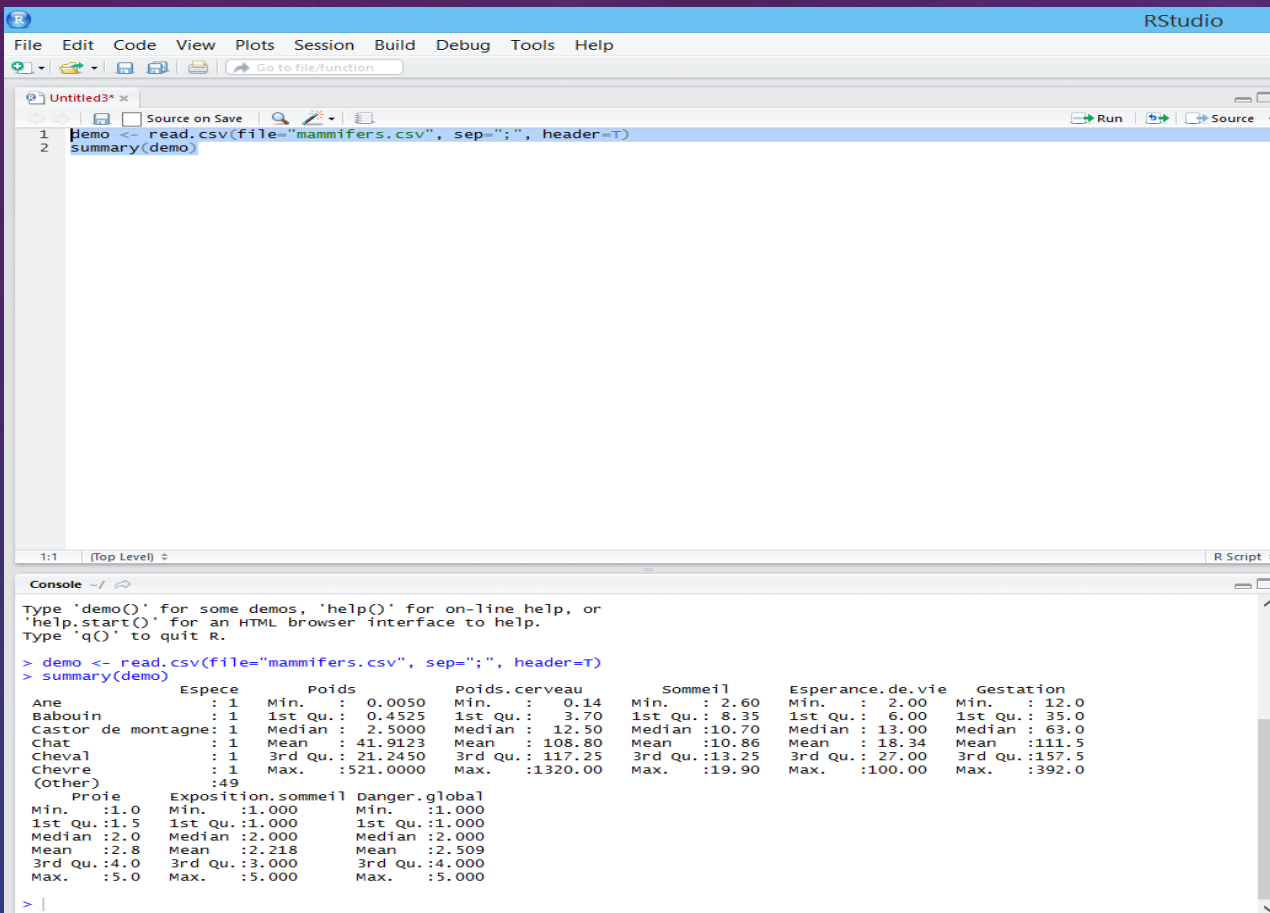
- 3 lignes de commande sous R/Rstudio :
 - `install.packages(« rattle »)`
 - **`library(rattle)`**
 - **`rattle()`**

PRINCIPAUX PACKAGES

- Beaucoup de packages supplémentaires à installer, parmi les principaux :
 - **GTK+** : permet la construction d'interfaces graphiques pour l'utilisateur
 - **RGtk2** : installations supplémentaires pour les interfaces graphiques de programmation
 - **Stringr** : facilite la manipulation de chaîne de caractères
 - **XML** : utile dans la lecture et la création de fichiers XML
 - **CairoDevice** : permet d'afficher des graphiques à l'écran et de les enregistrer dans un fichier ou en mémoire

EXEMPLE DE COMPARAISON R ET RATTLE

Sous R : summary()



```
1 demo <- read.csv(file="mammifere.csv", sep=";", header=T)
2 summary(demo)
```

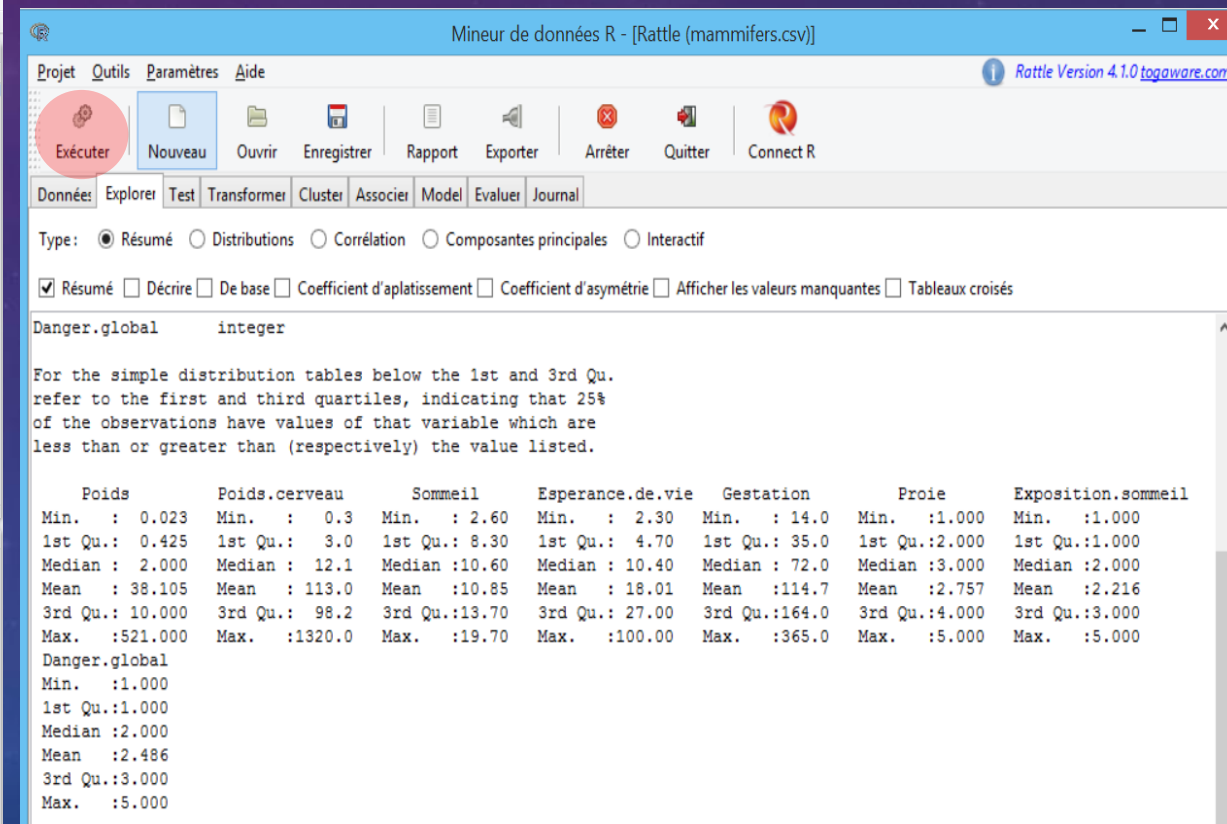
Console

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> demo <- read.csv(file="mammifere.csv", sep=";", header=T)
> summary(demo)

      Espece      Poids      Poids.cerveau      Sommeil      Esperance.de.vie      Gestation
Ane      : 1 Min.      : 0.0050 Min.      : 0.14 Min.      : 2.60 Min.      : 2.00 Min.      : 12.0
Babouin  : 1 1st Qu.: 0.4525 1st Qu.: 3.70 1st Qu.: 8.35 1st Qu.: 6.00 1st Qu.: 35.0
Castor de montagne: 1 Median: 2.5000 Median: 12.50 Median: 10.70 Median: 13.00 Median: 63.0
Chat     : 1 Mean     : 41.9123 Mean     : 108.80 Mean     : 18.34 Mean     : 111.5
Cheval   : 1 3rd Qu.: 21.2450 3rd Qu.: 117.25 3rd Qu.: 13.25 3rd Qu.: 27.00 3rd Qu.: 157.5
Chevre   : 1 Max.     : 521.0000 Max.     : 1320.00 Max.     : 19.90 Max.     : 100.00 Max.     : 392.0
(Other)  : 49
      Proie      Exposition.sommeil      Danger.global
Min.      :1.0 Min.      :1.000 Min.      :1.000
1st Qu.:1.5 1st Qu.:1.000 1st Qu.:1.000
Median :2.0 Median :2.000 Median :2.000
Mean     :2.8 Mean     :2.218 Mean     :2.509
3rd Qu.:4.0 3rd Qu.:3.000 3rd Qu.:4.000
Max.     :5.0 Max.     :5.000 Max.     :5.000
```

Sous Rattle



Mineur de données R - [Rattle (mammifere.csv)]

Rattle Version 4.1.0 togaware.com

Projet Outils Paramètres Aide

Exécuter Nouveau Ouvrir Enregistrer Rapport Exporter Arrêter Quitter Connect R

Données Explorer Test Transformer Cluster Associer Model Evaluer Journal

Type: ☒ Résumé ☐ Distributions ☐ Corrélation ☐ Composantes principales ☐ Interactif

☒ Résumé ☐ Décire ☐ De base ☐ Coefficient d'aplatissement ☐ Coefficient d'asymétrie ☐ Afficher les valeurs manquantes ☐ Tableaux croisés

Danger.global integer

For the simple distribution tables below the 1st and 3rd Qu. refer to the first and third quartiles, indicating that 25% of the observations have values of that variable which are less than or greater than (respectively) the value listed.

Poids	Poids.cerveau	Sommeil	Esperance.de.vie	Gestation	Proie	Exposition.sommeil
Min. : 0.023	Min. : 0.3	Min. : 2.60	Min. : 2.30	Min. : 14.0	Min. : 1.000	Min. : 1.000
1st Qu.: 0.425	1st Qu.: 3.0	1st Qu.: 8.30	1st Qu.: 4.70	1st Qu.: 35.0	1st Qu.: 2.000	1st Qu.: 1.000
Median : 2.000	Median : 12.1	Median : 10.60	Median : 10.40	Median : 72.0	Median : 3.000	Median : 2.000
Mean : 38.105	Mean : 113.0	Mean : 10.85	Mean : 18.01	Mean : 114.7	Mean : 2.757	Mean : 2.216
3rd Qu.: 10.000	3rd Qu.: 98.2	3rd Qu.: 13.70	3rd Qu.: 27.00	3rd Qu.: 164.0	3rd Qu.: 4.000	3rd Qu.: 3.000
Max. : 521.000	Max. : 1320.0	Max. : 19.70	Max. : 100.00	Max. : 365.0	Max. : 5.000	Max. : 5.000

Danger.global

Min. : 1.000

1st Qu.: 1.000

Median : 2.000

Mean : 2.486

3rd Qu.: 3.000

Max. : 5.000

LISTING DES METHODES DE DATAMINING

Analyses factorielles ACP ☒ ACM ☐ AFC ☐ AFDM ☐

➡ Quid des variables catégorielles ?

Régression linéaire ☒ linéaire généralisée ☒ de Poisson ☒
logistique ☒ logistique multinomiale ☒
Analyse de la variance ☐ Analyse de la covariance ☐

Clustering K-means ☒ K-médoides ☐ Classification floue ☐
CAH ☒ CDH ☐

➡ Caractérisation des classes difficile

LISTING DES METHODES DE DATAMINING

Tests statistiques

Student ☒

Fisher ☒

Khi-2 ☐

V de Cramer ☐

Wilcoxon-Mann-Whitney ☒

Rangs signés de Wilcoxon ☒

Apprentissage supervisé

Arbres de décision ☒

Support vector machine (SVM) ☒

Random forest ☒

Analyse discriminante ☐

Règles d'association ☒

Réseaux de neurones ☒

Séries temporelles ☐

Modèles de survie ☒

Boosting ☒

Data visualisation ☐

POINTS FORTS DE RATTLE

- Même capacité à gérer les bases de données volumineuse que R
- Diversité des sources de données (Excel, R, ODBC, etc.)
- Analyse descriptive assez complète (illustrations graphiques, modalités, statistiques descriptives, etc.)
- Interface graphique (facilité d'utilisation pour les dataminers)
- Syntaxe code auto-gérée par Rattle

POINTS FAIBLES DE RATTLE

- Peu de documentation sur Internet
- Interface (présentation des résultats, graphiques, etc.)
- Absence de certaines méthodes statistiques
- Connaissance préalable du langage R souhaitable
- Dépendance d'autres packages (GTK implique l'installation de beaucoup de packages)
- Outil parfois instable

EVALUATION CRITIQUE

- Moins complet que d'autres logiciels gratuits (Knime et Orange par exemple)
- Rendu graphique insatisfaisant
- Types des données traitées : uniquement quantitatives, qualitatives et binaires. Pas de traitement des données textuelles (contrairement à Orange, Weka, RapidMiner, Knime par exemple)



Logiciel limité pour les dataminers, car seules les principales méthodes statistiques sont présentes sous Rattle, donc un **complément de programmation en R** est quasiment obligatoire pour aller plus loin dans les analyses

CONCLUSION

- Utilisation plutôt intuitive et facile
- Réalisation des principales méthodes de datamining
- Limité pour les statisticiens



Merci de votre attention !

Avez-vous des questions ?