



HAL
open science

PyMixmod, le composant pour Python de MIXMOD Classification des données par modèles de mélanges

Classification des données

Florent Langrognet, Christian Poli

► To cite this version:

Florent Langrognet, Christian Poli. PyMixmod, le composant pour Python de MIXMOD Classification des données par modèles de mélanges Classification des données. JDEV (Journées du DEVeloppement logiciel) 2015, 2015, Bordeaux, France. , 2015. hal-01355421

HAL Id: hal-01355421

<https://hal.science/hal-01355421>

Submitted on 23 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PyMixmod, le composant pour Python de MIXMOD

Classification des données par modèles de mélanges

www.mixmod.org



Florent Langrognat^a, Christian Poli^b

^aLaboratoire de Mathématiques de Besançon (Université de Franche-Comté - CNRS)

^bInria Saclay - Île-de-France (Service Expérimentation et développement)



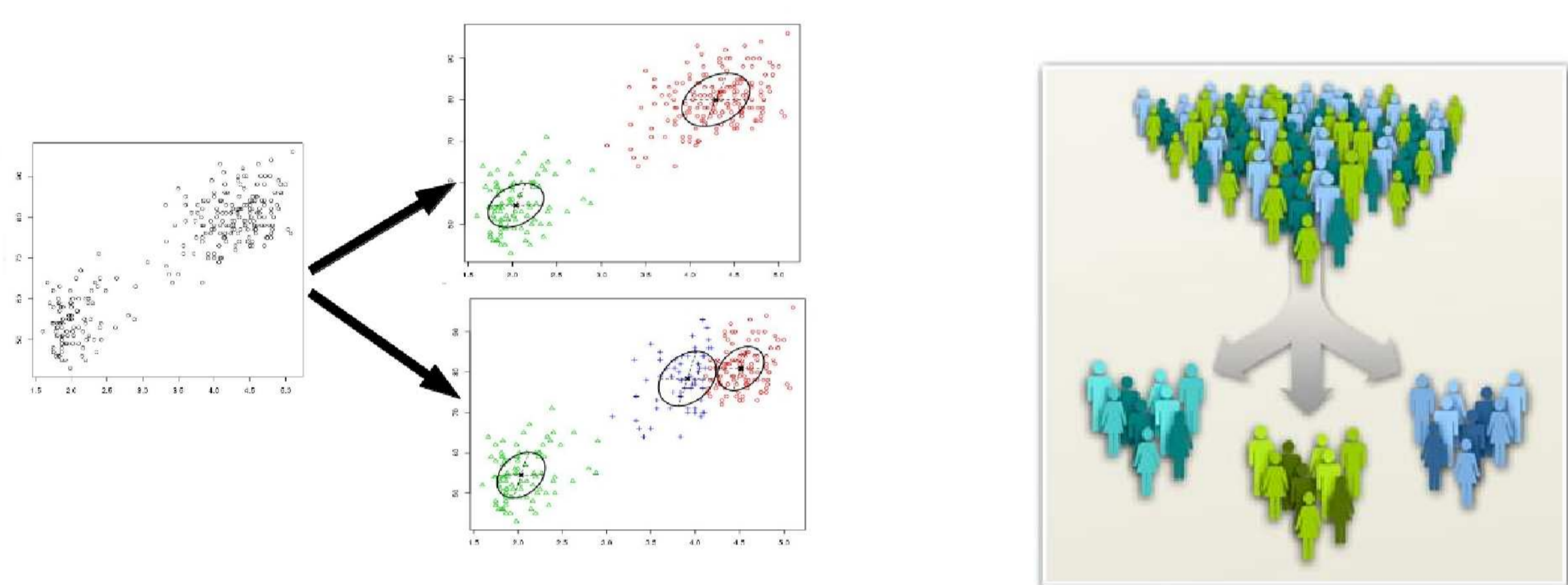
Classification des données

Regrouper des **individus** dans des **classes homogènes**

- **Classification non supervisée**

- ▶ Trouver le 'bon' nombre de classes
- ▶ **Classer** les individus
- ▶ Caractériser les classes

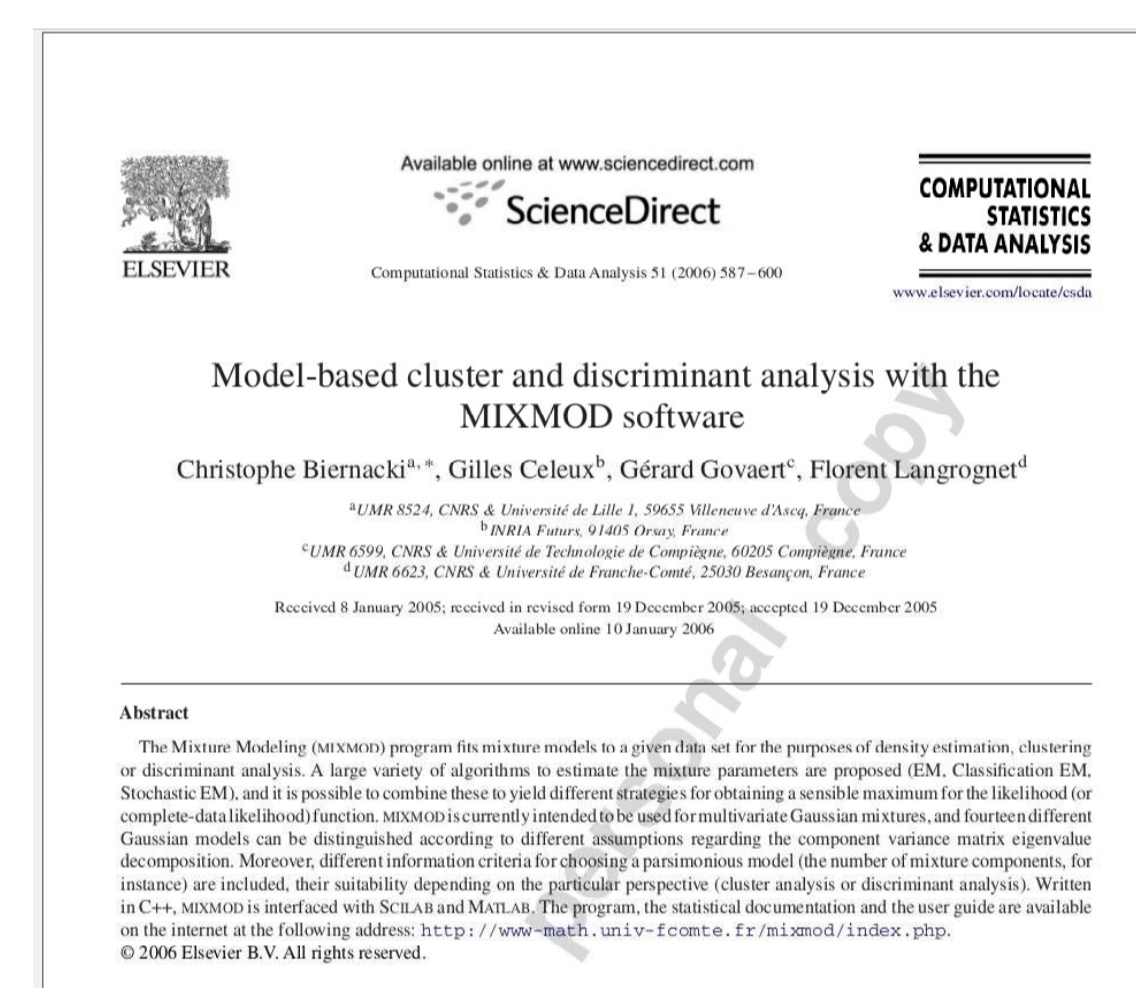
- **Classification supervisée** : apprentissage puis classement



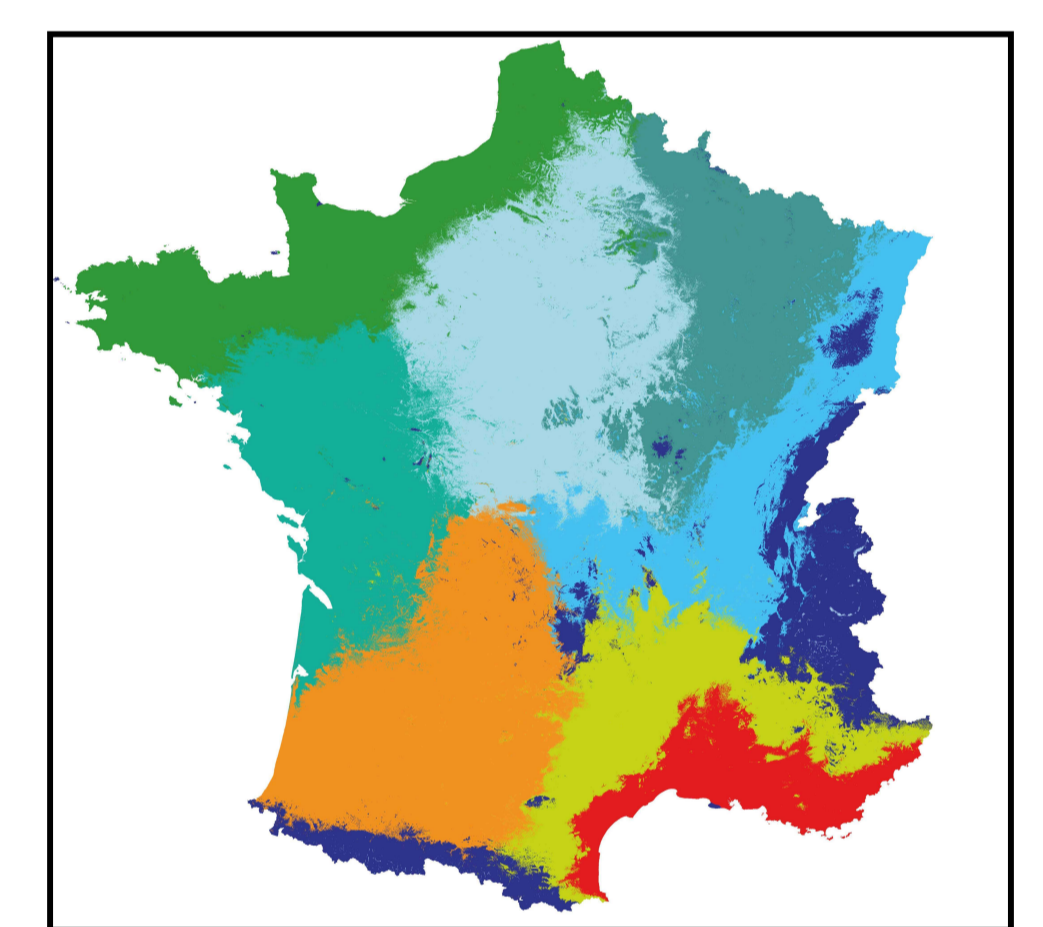
Diffusion, valorisation, utilisations de MIXMOD

- **Domaines très variés** : Banque, marketing, environnement, sociologie, épidémiologie, génomique, biochimie, ...
- **5 composants logiciels** sous **Licence GNU GPL** : www.mixmod.org
- **Utilisateurs** : environ 200 téléchargements par mois
- **Enrichissement et redistribution** - Ex : SelVarClust, RMixmodCombi
- **Intégration** de la bibliothèque mixmodLib dans d'autres logiciels
- **Collaborations** sur des problématiques complexes
- **Conférences, articles**

Article de recherche



Classification des climats en France



MIXMOD et les modèles de mélanges

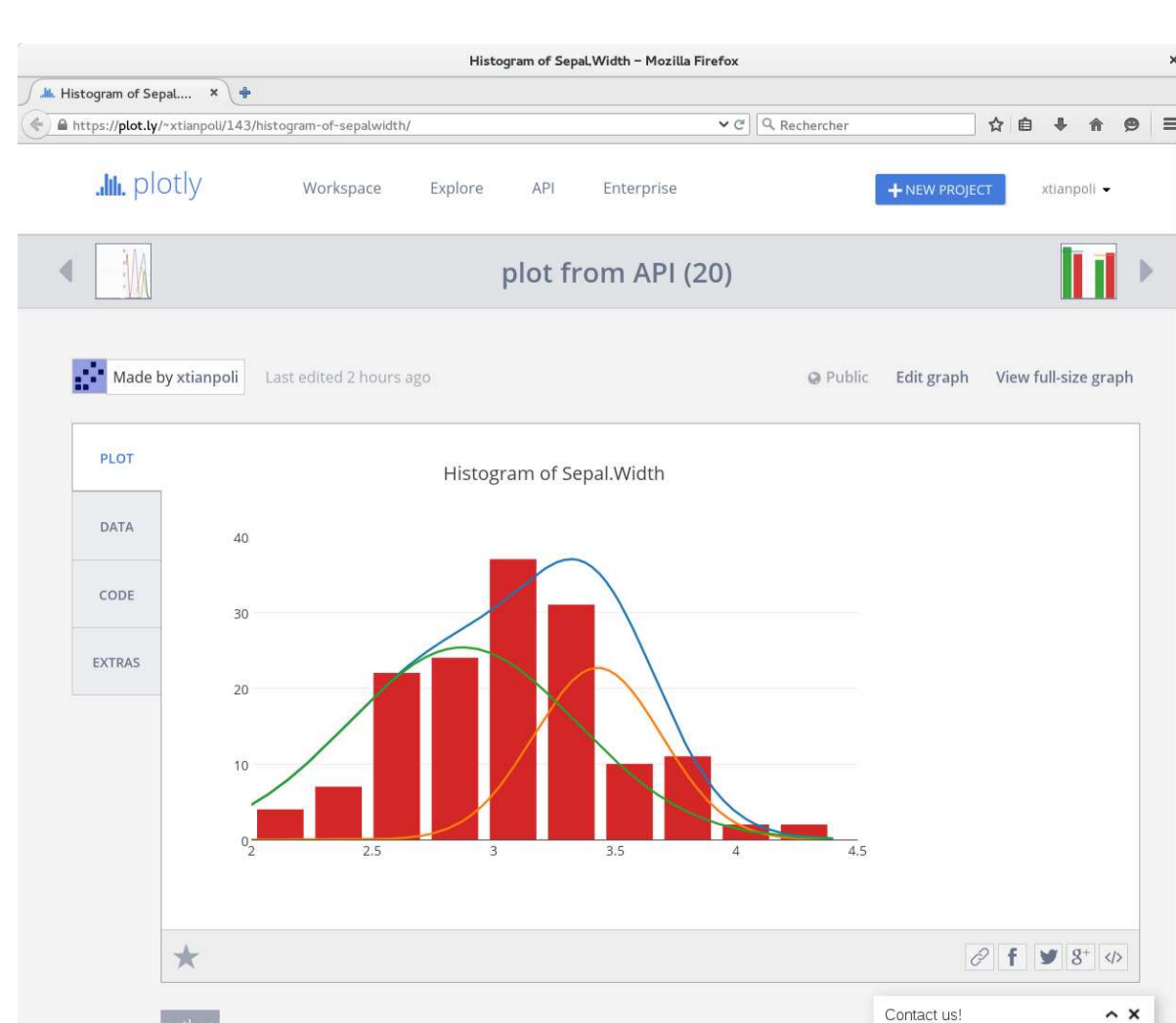
- Approche **probabiliste** souple
- Données **quantitatives, qualitatives, ou mixtes**
- **Big data** (nombre d'individus et/ou nombre de caractéristiques)
- **Principales fonctionnalités**
 - ▶ **3 Algorithmes** : EM, CEM, SEM
 - ▶ **6 Méthodes d'initialisation**
 - ▶ **4 Critères de sélection** : BIC, ICL, NEC, CV
 - ▶ **14 modèles gaussiens** pour les données quantitatives
 - ▶ **5 modèles multinomiaux** pour les données qualitatives
 - ▶ **8 modèles** pour les données quantitatives en **Grande Dimension**
 - ▶ **20 modèles** pour les **données mixtes** (quantitatives/qualitatives)

PyMixmod

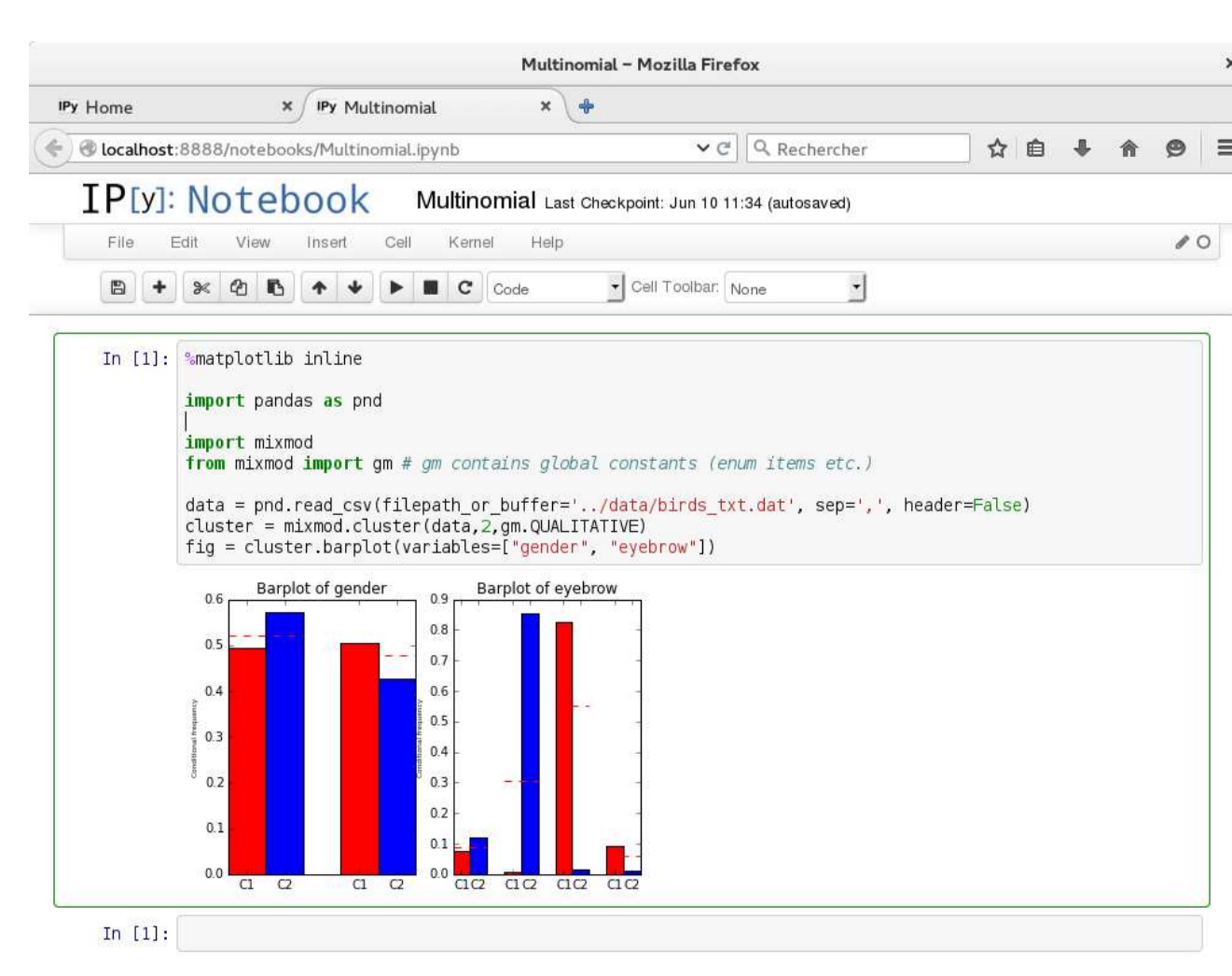
- Destiné à rendre la bibliothèque **mixmodLib** accessible à la communauté scientifique utilisant **Python**, en particulier dans les domaines de la manipulation, analyse et visualisation des données
- Intégré avec l'écosystème Python dédié au calcul scientifique et au **Big data** : **NumPy, SciPy, Pandas** et **Matplotlib**
- Développé en **Cython** pour de meilleures performances grâce à un code 100% natif
- Interopérable avec des outils de visualisation de données en ligne (Plotly, IPython Notebook)
- Compatible avec **Python 2.7.x** et **Python 3.4.x**

Partager en ligne avec PyMixmod

PyMixmod en interaction avec Plotly

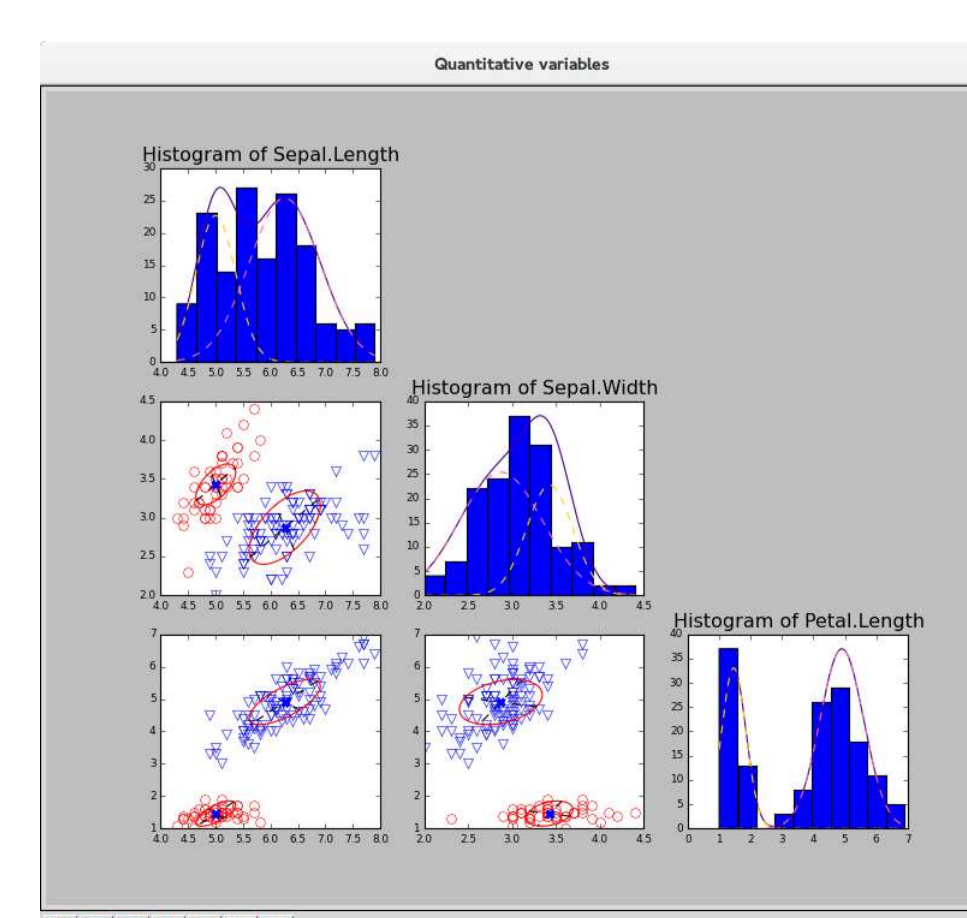


PyMixmod et IPython Notebook



Visualiser avec PyMixmod

Classification (données quantitatives)



Classification (données qualitatives)

