



Département de mathématiques et de génie industriel

**Plan du cours MTH8302**

**hiver 2024**

**Modèles de régression et d'analyse de la variance**

Site Internet du cours: <https://cours.polymtl.ca/mth6301/MTH8302.htm>

remarque: **PAS** sur Moodle accès : sans identification et sans mot de passe.

Professeur : Bernard CLÉMENT, PhD

Courriel : [bernard.clement@polymtl.ca](mailto:bernard.clement@polymtl.ca)

**Horaire** : 16h00-17h20 cours / 17h20-17h30 pause / 17h30-18h45 cours / 18h45 Q&R.

**Cours**: présentiel & à distance (ZOOM) **Consultation**: après chaque cours et par courriel

**Calendrier**

**salle : L-3816**

JANVIER 2024							FÉVRIER 2024							MARS 2024						
Dim	Lun	Mar	Mer	Jeu	Ven	Sam	Dim	Lun	Mar	Mer	Jeu	Ven	Sam	Dim	Lun	Mar	Mer	Jeu	Ven	Sam
	1	2	3	4	5	6					1	2	3						1	2
7	8	9	10	11	12	13	4	5	6	7	8	9	10	3	4	5	6	7	8	9
14	15	16	17	18	19	20	11	12	13	14	15	16	17	10	11	12	13	14	15	16
21	22	23	24	25	26	27	18	19	20	21	22	23	24	17	18	19	20	21	22	23
28	29	30	31				25	26	27	28	29			24	25	26	27	28	29	30
														31						

JANVIER mercredi (10-17-24-31) (4x3=12 h) - samedi (13-20-27) (3x2=6 h) total=18 h

FÉVRIER mercredi (07-14-21-28) (4x3=12 h) - samedi (03-10-17-24) (4x2=8 h) total=20 h + 2 h total = 40 h

**Cours 10 janvier: en présentiel - TOUS les autres cours: en ligne via ZOOM - enregistrement vidéo**

## PROGRAMME

### Analyse supervisée

**Régression** : simple, non-linéaire, logistique, multiple, MARS, Lasso, en composantes principales, ridge, Bayes, Partial Least Square (PLS)

**Analyse variance (ANOVA)** : uni factoriel, multifactoriel, covariance, mesures répétées (longitudinales), linéaire généralisé

### Apprentissage machine (Data mining, Machine Learning (ML))

réseaux de neurones (1 ou 2 couches), arbres de classification, forêts aléatoires, ...

**Remarque** : ce cours a un contenu considérable. Il est axé sur l'objectif des méthodes, leurs mises en œuvre avec un logiciel de votre choix (Statistica, JMP Pro, Python, R, Julia, Weka,...).

Il vise surtout l'interprétation des résultats. Nous ne n'avons pas le temps de développer les équations mathématiques. Des documents complémentaires sont disponibles dans JMP.

**Évaluation** basée sur 4 devoirs. Il n'y a pas d'examens dans ce cours.

### Devoirs

DEVOIRS	No1	No2	No3	No4
% note finale	20%	20%	25%	35%
date dépôt	23-janv	12-févr	04-mars	25-mars
date remise	12-févr	04-mars	25-mars	29-avr
énoncé (pdf)	pdf1	pdf2	pdf3	pdf4
page1 (docx)	docx1	docx2	docx3	docx4

## BUT DU COURS

L'analyse de régression et l'analyse de variance sont parmi les plus employées pour analyser des données d'observationnelles (big data) et des données d'expérimentales (small data).

Dans le langage d'aujourd'hui, elles sont dans la catégorie de **l'analyse supervisée** (input-output). Le but principal de ces méthodes est de :

- (a) développer des équations de prédiction entre des variables continues X et une variable de réponse Y continue ou catégorique  $Y = f(X) + \epsilon$  : **modèles de RÉGRESSION, de PRÉDICTION**  
**domaine d'application privilégié** : données observationnelles, données historiques
- (b) tester si des variables catégoriques X (appelés facteurs) et des variables continues ont une influence sur une variable de réponse Y  $Y = f(x) + \epsilon$  : **modèles d'ANALYSE de VARIANCE**  
**domaine d'application privilégié** : données d'expérimentation (cours MTH8301)

## Caractéristiques du cours

- Vise à présenter les 2 principales méthodes importantes de modélisation et d'analyse statistique afin **d'interpréter et donner un sens aux données**.
- Ne développe pas de manière détaillée la partie mathématique / statistique.
- Est orienté vers les applications et la mise en œuvre des analyses avec le logiciel Statistica et JMP Pro (SAS) illustré avec plusieurs dizaines d'exemples en ingénierie et autres sujets.
- En plus de (a) et (b) ci-haut, le cours contient des méthodes du **Machine Learning (Data Mining)** : régression spline (MARS) **réseau de neurones (ANN)** **arbres classification (CRT)** et autres

## QUALITÉS BCAPG : Bureau Canadien Aptitude Programme Génie

1 Connaissance en génie	2 Analyse de problèmes	3 Investigation	4 Conception	5 Utilisation d'outils d'ingénierie	6 Travail individuel et en équipe
oui	oui	oui	oui	oui	oui
7 Communication	8 Professionalisme	9 Impacts environnement	10 déontologie	11 économie et gestion de projets	12 apprentissage continu
oui					

## Exigences concernant les rapports de devoirs

Le but de ces exigences est de limiter le nombre de pages et de faciliter la correction.

- 1) Le rapport doit être fait avec un logiciel de traitement de texte.
- 2) Utiliser la police de caractères **Arial 10** ou **Arial 9** avec des **caractères gras**.
- 3) Numéroté les pages **en haut à droite**.
- 4) Commencer **chaque numéro sur une nouvelle page**.
- 5) La **première page** du rapport à remettre doit être une copie de celle de l'énoncé du devoir. Une copie sera disponible sur le site du cours  
<http://www.cours.polymtl.ca/mth6301/MTH8302.htm>
- 6) **Ne pas répéter l'énoncé des questions** dans le rapport.
- 7) Insérer vos graphiques dans des **zones de texte** dont la taille approximative vertical X horizontal ne dépasse pas 9 X 9 (cm) approximativement
- 8) Viser à contenir le rapport principal à l'intérieur d'un **maximum de 30 pages**. Utiliser des pages annexes au besoin. Ne pas briser le fil conducteur de la solution de chaque question.
- 9) 10 points sur 100 seront attribués à la **qualité générale du rapport**.
- 10) Le rapport n'est **pas seulement un assemblage (copier-coller) des sorties** (tableaux et graphiques) du logiciel **STATISTICA, JMP** ou autre.
- 11) Chaque question doit se terminer avec une **conclusion sommaire sur l'interprétation** des données. Soyez bref et aller droit au but. S'il y a lieu, ajouter des **commentaires que vous jugez pertinents**.
- 12) **Si vous utiliser Python, R, Julia, Perl, Weka, ou un autre logiciel pour faire les analyses**, mettez une copie de votre code en annexe.
- 13) Envoyez votre rapport au professeur : [bernard.clement@polymtl.ca](mailto:bernard.clement@polymtl.ca) avec votre courriel : nom.prenom@polymtl.ca
- 14) Votre rapport doit **OBLIGATOIREMENT** avoir le nom **FFFF\_mmm\_2024\_MTH8302\_devoirN.pdf**  
**FFFF** = nom de famille      **mmm** = matricule      **N** = numéro du devoir (1, 2, ..., 5)

## RÉFÉRENCE et DOCUMENTATION

Notes de cours du professeur disponibles sur le site WEB du cours

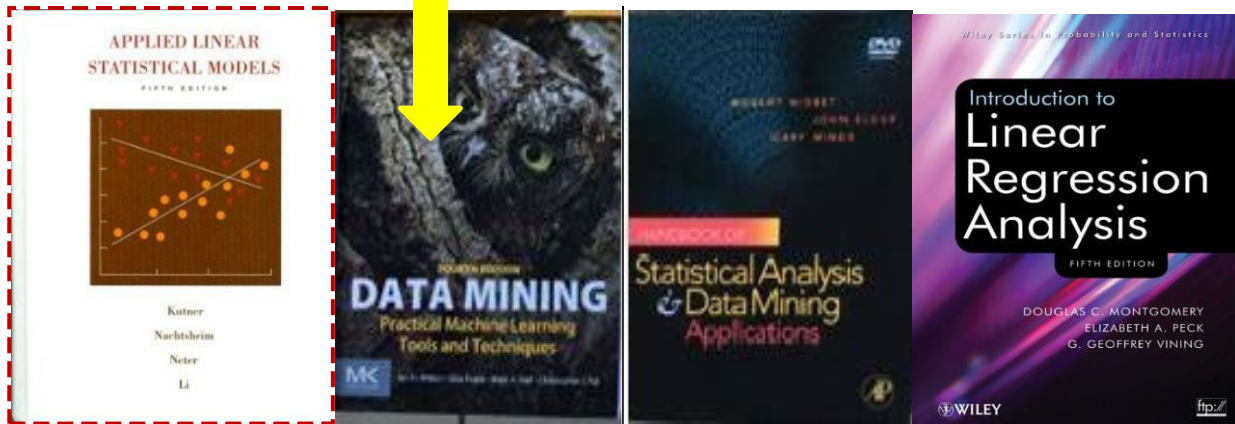
Régression

ANOVA

Data Mining (Machine Learning)

## LIVRES de RÉFÉRENCE

C. PALL, professeur au département GIGL de POLYTECHNIQUE Montréal



**Kutner&all**: Applied Linear Statistical Models, 5<sup>th</sup> ed., (2005), McGraw-Hill, **(référence principale)**

Kutner M., Nachtsheim C., Neter J., Li W.,

**(1400 pages !)**

Version PDF disponible pour téléchargement sur le site WEB du cours.

**RÉGRESSION** : simple (chap. 1 à 5) / multiple (chap. 6 à 12) / non linéaire (chap. 13) / logistique (chap. 14)

**ANALYSE DE VARIANCE** : base (chap. 15 à 21), covariance (chap. 22)

multifactorielle (chap. 23 à 25), avancée (chap. 26 à 28)

**PLANS D'EXPÉRIENCE** : hors programme, (chap. 29, 30) fait partie du cours MTH8301

**Little&all** : SAS for Mixed Models, 2<sup>nd</sup> ed. (2006), SAS Institute

Littell R. C., Milliken W. W., Stroup R. D., Wolfinger R. D., Schabenberger O.

**Montgomery&all** : Introduction to Linear Regression Analysis, 4<sup>th</sup> ed., (2006), John Wiley

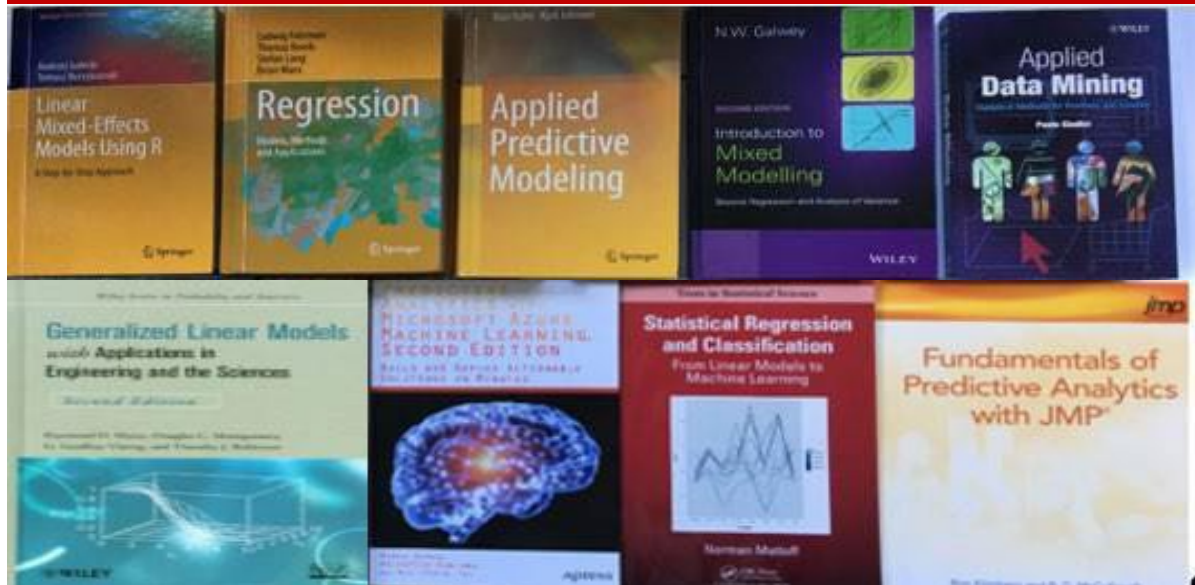
Montgomery D. C., Peck E. A., Vining G. G.

**Myers&all** : Generalized Linear Models, 2<sup>nd</sup> ed, (2010), John Wiley

Myers R. H., Montgomery D. C., Vining G. G., Robinson T. J.

Le site WEB du cours contient d'autres références et documents.

## AUTRES LIVRES



## OBJECTIFS D'APPRENTISSAGE

- construire des modèles de régression : estimation, tests, résidus, méthodes de sélection de variables, régression logistique, nouvelles méthodes de modélisation;
- maîtriser les concepts des modèles d'analyse de variance;
- comprendre les différences entre les modèles de régression et les modèles d'analyse de variance;
- reconnaître la structure des données pour choisir la méthode d'analyse appropriée ;
- maîtriser l'utilisation de 2 logiciels commerciaux : STATISTICA, JMP Pro  
+ au moins 1 logiciel Open source : Python, R, Weka, Julia, ....

## PRÉALABLES essentiels = connaissances de base en analyse statistique

- statistique descriptive : moyenne, variance, écart-type, histogramme, etc.
- distributions: gaussienne (normale) Z, Student T, Fisher F
- tests d'hypothèses sur des moyenne(s) et des variance(s) avec un ou deux échantillons

révision si besoin : voir <http://cours.polymtl.ca/mth6301/MTH2302B.htm>

**ÉVALUATION** 5 devoirs / le devoir5 est une **étude de cas** remplaçant l'examen final

**SITE WEB de MTH8302** <https://cours.polymtl.ca/mth6301/MTH8302.htm>

MTH8302 *Modèles de régression et d'analyse de la variance*

courriel : : [bernard.clement@polymtl.ca](mailto:bernard.clement@polymtl.ca) [cv](#) [prix Méritas \(pdf\)](#)

**PLAN** [\(pdf-00\)](#) lien vers page accueil de **Bernard Clément** [\(pdf\)](#)

**NOTES** **cours en format PDF** : liste [\(pdf\)](#)

[\(pdf-01\)](#) [\(pdf-02\)](#) [\(pdf-03\)](#) [\(pdf-04\)](#) [\(pdf-05\)](#) [\(pdf-06\)](#) [\(pdf-07\)](#) [\(pdf-08\)](#)  
[\(pdf-09\)](#) [\(pdf-10\)](#) [\(pdf-11\)](#) [\(pdf-12\)](#) [\(pdf-13\)](#) [\(pdf-14\)](#) [\(pdf-15\)](#)

**VIDÉOS 2022** **cours en format MP4** : liste [\(pdf\)](#)

[\(mp4-01\)](#) [\(mp4-02\)](#) [\(mp4-03\)](#) [\(mp4-04\)](#) [\(mp4-05\)](#) [\(mp4-06\)](#) [\(mp4-07\)](#)  
[\(mp4-08\)](#) [\(mp4-09\)](#) [\(mp4-10\)](#) [\(mp4-11\)](#) [\(mp4-12\)](#) [\(mp4-13\)](#) [\(mp4-14\)](#)

**VIDÉOS 2023** **cours en format MP4** : liste [\(pdf\)](#)

[\(mp4-00\)](#) [\(mp4-15\)](#) [\(mp4-16\)](#) [\(mp4-17\)](#) [\(mp4-18\)](#) [\(mp4-19\)](#) [\(mp4-20\)](#)  
[\(mp4-21\)](#) [\(mp4-22\)](#) [\(mp4-23\)](#) [\(mp4-24\)](#)

**DATA EXEMPLES** Regression [\(stw\)](#) [\(xlsx\)](#) DataMining [\(stw\)](#) [\(xlsx\)](#)

DataMinerRecipes [\(stw\)](#) [\(sta\)](#) [\(xlsx\)](#) ANOVA [\(stw\)](#) [\(xlsx\)](#)

**EXERCICES** énoncés [\(pdf\)](#) data [\(stw\)](#) [\(xlsx\)](#)

**DEVOIRS** devoir1 [\(pdf\)](#) [\(docx\)](#) devoir2 [\(pdf\)](#) [\(docx\)](#) devoir3 [\(pdf\)](#) [\(docx\)](#)

## OCUMENTS-RÉFÉRENCES-COMPLÉMENTS

- Kutner&all-Applied Linear Statistical Models [\(pdf\)](#)
- Statistique, science des données, intelligence artificielle [\(pdf\)](#)
- Corrélation partielle et paradoxe de Simpson [\(pdf\)](#)
- Modèles Non Linéaires [\(pdf\)](#)
- Critères comparaison modèles [\(pdf\)](#)
- Critères Aikaide (AIK) et Bayes (BIC) [\(pdf\)](#)
- Modélisation et analyse statistique [\(pdf\)](#)
- 4-bar linkage [\(pdf\)](#)
- Codage des variables dans les modèles statistiques [\(pdf\)](#)
- Études statistiques : observationnelles et expérimentales [\(pdf\)](#)
- Introduction au contrôle statistique des processus (SPC) [\(pdf\)](#)
- Courbes ROC et Courbes Lift [\(pdf\)](#)
- Méthodes pour évaluer des modèles prédictifs [\(pdf\)](#)
- Le rôle de la statistique dans l'industrie et les affaires [\(pdf\)](#)
- Introduction à STATISTICA [\(pdf\)](#)
- Besse-Introduction à R [\(pdf\)](#)
- Why R in Data Science ? [\(pdf\)](#)
- Machine Learning by Analogy [\(pdf\)](#)
- Articles plus téléchargés dans DataScienceCentral [\(xlsx\)](#)

# Machine Learning Algorithms You Should Know in 2024

<https://www.kdnuggets.com/2021/01/machine-learning-algorithms-2021.html>

Decision Tree Gradient Boosting Machine Learning Naive Bayes  
Regression SVM K-nearest neighbors

**Statistica : DATA MINING (maintenant Machine Learning)**

**ARBRES CLASSIFICATION**

**Random Forests**

**GAM**

**MARS**

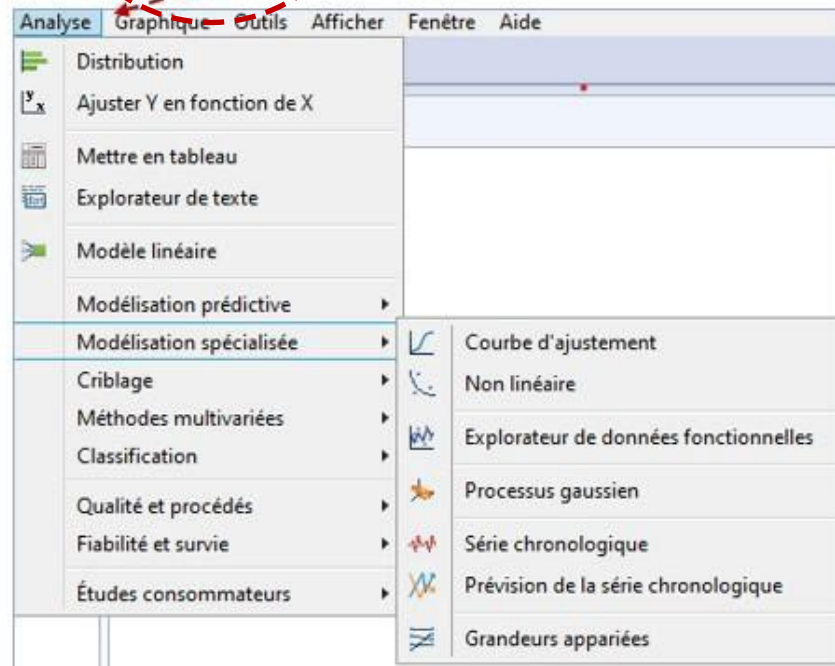
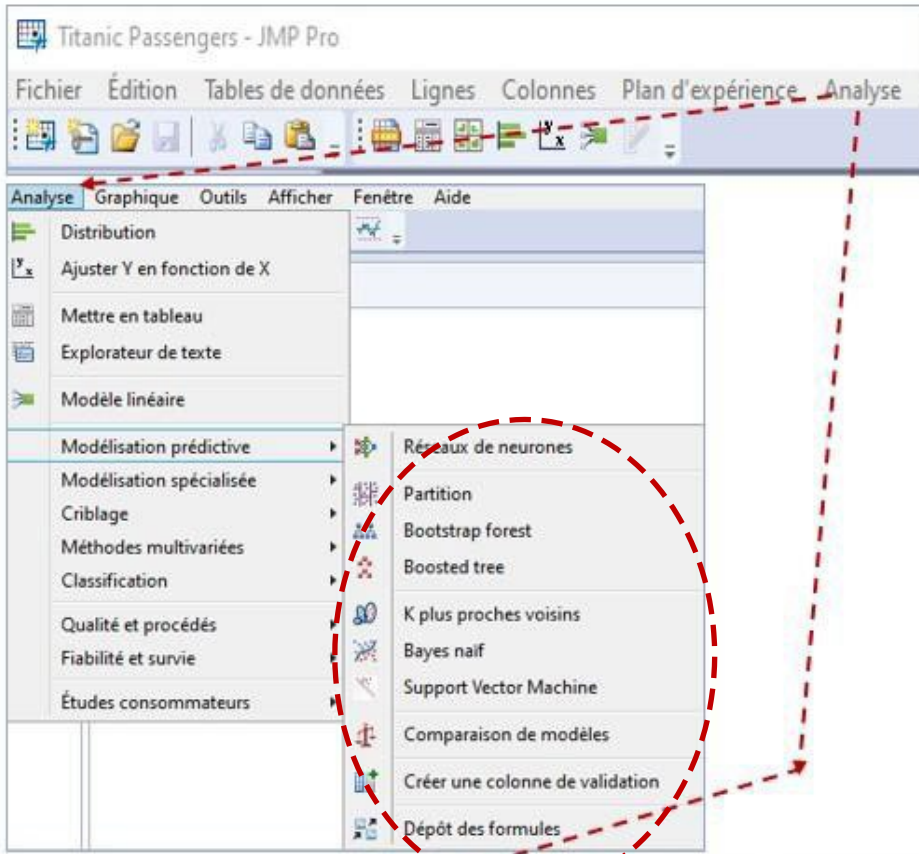
**ANN : Réseaux Neurones**

**SVM : Support Vector Machines**

**Data Mining**

- Data Miner Recipes
- General Classification/Regression Tree Models
- General CHAID Models
- Interactive Trees (C&RT, CHAID)
- Boosted Tree Classifiers and Regression
- Random Forests for Regression and Classification
- Generalized Additive Models
- MARSplines (Multivariate Adaptive Regression Splines)
- Cluster Analysis (Generalized EM, k-Means & Tree)
- Automated Neural Networks
- Machine Learning (Bayesian, Support Vectors, K-Nearest)
- Independent Components Analysis
- Text & Document Mining
- Web Crawling, Document Retrieval
- Association Rules
- Sequence, Association, and Link Analysis
- Rapid Deployment of Predictive Models (PMML)
- Model Converter
- Goodness of Fit, Classification, Prediction
- Feature Selection
- Optimal Binning for Predictive Data Mining
- Weight of Evidence
- Stepwise Model Builder
- Interactive Drill Down
- Process Optimization

## JMP Pro : procédures Machine Learning



## RESSOURCES ADDITIONNELLES (autres sites WEB)

Site Wikistat Université Toulouse <http://Wikistat.fr> et <https://github.com/wikistat>

**Philippe BESSE Wikistat** <http://wikistat.fr/> <https://github.com/wikistat/>



Université de Toulouse

# Philippe Besse

Professeur à l'INSA, Département de Génie Mathématique et Modélisation



INSTITUT de MATHÉMATIQUES de TOULOUSE

Accueil
Enseignement
Publications

- Membre de l'équipe Statistique et Probabilités de L'Institut de Mathématiques (UMR CNRS 5219)
- **Courriel** : Philippe.Besse-at-math.univ-toulouse.fr
- **Dépôt Github**

- **Adresse Postale** : INSA de Toulouse  
Dpt. Génie Mathématique et Modélisation  
F-31077 Toulouse cedex 04
- **Bureau** : 135, avenue de Ranguel  
INSA, Dpt. GMM, porte 04





## Statistique & Machine Learning de Statisticien à Data Scientist



**Attention** Le site est en cours de migration vers [Github.com/wikistat](https://github.com/wikistat) dépôt de tous les scénarios / tutoriels sous la forme de calepins (*jupyter notebook*) en R ou Python.

Tous les documents sont mis à disposition selon les termes de la licence [Creative Commons Paternity 3.0 France](https://creativecommons.org/licenses/by/3.0/) Pas d'Utilisation Commerciale - Partage avec des Conditions Initiales à l'Identique.

**Citer une vignette** selon le format de l'exemple ci-dessous:

```
@misc{ wikistat-dustering,
author = "Wikistat",
title = "Classification non-supervisée -- Wikistat",
year = "2016",
url = "http://wikistat.fr/pdf/st-m-explo-classif.pdf",
note = "[En ligne; Page disponible le 21-janvier-2016]"
}
```

Généralités	Logiciel R	Logiciel SAS
<ul style="list-style-type: none"> <li>• Présentation de Wikistat</li> <li>• De la Statistique à la Science des Données</li> <li>• Statistique et déontologie scientifique</li> <li>• Rédiger un rapport de statistique</li> </ul>	<ul style="list-style-type: none"> <li>• Démarrer en R</li> <li>• Initiation à R</li> <li>• Graphes avec R</li> <li>• Programmer en R</li> </ul>	<ul style="list-style-type: none"> <li>• Tutoriel 1 prise en main</li> <li>• Tutoriel 2 SAS de base</li> <li>• Tutoriel 3 SAS/Graph</li> <li>• Tutoriel 4 macros commandes</li> <li>• Tutoriel 5 SQL</li> </ul>

## Statistique niveau L

Description	Inférence	Modélisation
<ul style="list-style-type: none"> <li>• Statistique élémentaire: Introduction</li> <li>• Statistique descriptive unidimensionnelle</li> <li>• Statistique descriptive bidimensionnelle</li> </ul>	<ul style="list-style-type: none"> <li>• Probabilités élémentaires</li> <li>• Estimation statistique</li> <li>• Tests statistiques</li> </ul>	<ul style="list-style-type: none"> <li>• Introduction à l'analyse en composantes principales</li> <li>• Régression linéaire simple</li> <li>• Introduction régression multiple</li> </ul>

## Statistique niveau M1

### Statistique et apprentissage automatique (*machine learning*) niveau M2

#### Modèle mixte

- Introduction à la modélisation statistique
- Généralités sur le modèle linéaire
- Analyse de variance univariée
- Plans d'expérience incomplets
- Analyse de variance multivariée
- Modèle à effets aléatoires et modèle mixte
- Modèles pour mesures répétées
- A propos de la méthode Bonferroni
- Différents types de sommes de carrés
- Tests multidimensionnels pour données répétées
- Bibliographie

#### Apprentissage Statistique ...

- Apprentissage machine: Introduction
- Qualité de prévision et risque
- Sélection de modèle dans le cas gaussien
- Modèles non paramétriques
- Régression logistique
- Composantes principales et régression PLS parcimonieuses
- Analyse discriminante décisionnelle
- Réseaux de neurones
- Arbres binaires de décision
- Agrégation de modèles
- Machines à vecteurs supports
- Imputation de données manquantes
- Détection d'anomalies
- Introduction au bootstrap
- Conclusion sur l'apprentissage

#### & ... Data Science

- Convex optimisation and application
- High dimensional regression
- Collaborative filtering
- Sequential and reinforcement learning, Stochastic Optimization 1
- Sequential and reinforcement learning, Stochastic Optimization 2
- An introduction to network inference and mining

#### High Dimensional Statistics

- Linear Models for Regression and Model Selection
- High dimensional regression
- Nonparametric Learning and Regularization
- High Dimensional Classification
- Neural Networks and Introduction to Deep Learning

<https://github.com/wikistat/>

**Wikistat 2.0**  
Tutoriels (carnaps jupyter) d'Auto-Apprentissage en Science des Données & Intelligence artificielle  
Toulouse - Site compagnon de <http://wikistat.fr/>

Repositories 11 | People 2 | Projects 0

**Grow your team on GitHub**  
GitHub is home to over 35 million developers working together. Join them to grow your own development teams, manage permissions, and collaborate on projects.  
[Sign up](#)

**Pinned repositories**

- Exploration**  
Science des Données Saison 2: Exploration statistique multidimensionnelle, ACP, AFC, AFD, Classification non supervisée  
• Jupyter Notebook ★ 11 🍴 17
- Apprentissage**  
Science des Données Saison 3: Apprentissage Automatique / Statistique pour l'Intelligence Artificielle  
• Jupyter Notebook ★ 12 🍴 12
- Intro-Python**  
Python pour Statistique et Science des Données -- Syntaxe, Trafic de Données, Graphes, Programmation, Apprentissage  
• Jupyter Notebook ★ 9 🍴 22
- High-Dimensional-Deep-Learning**  
Science des données Saison 4 : Apprentissage en grande dimension, Données fonctionnelles, Détection d'anomalies, Introduction au Deep Learning  
• Jupyter Notebook ★ 3 🍴 1
- AI-Frameworks**  
Science des Données Saison 5: Technologies pour l'apprentissage automatique / statistique de données massives et l'Intelligence Artificielle  
• Jupyter Notebook ★ 10 🍴 6
- Intro-PySpark**  
Python pour Statistique et Science des Données Avec SparkML et MLlib -- Mining, Stats Élémentaire, ML, SQL  
• Jupyter Notebook ★ 3 🍴 3



**AUTRES SITES WEB**

Site KDnuggets : un des meilleurs sites <https://www.kdnuggets.com/>

The screenshot shows the KDnuggets website homepage. At the top, there's a navigation bar with social media icons (Twitter, Facebook, LinkedIn) and a search bar. Below the navigation bar, there's a horizontal menu with categories like SOFTWARE, News/Blog, Top stories, Opinions, Tutorials, JOBS, Companies, Courses, Datasets, EDUCATION, Certificates, Meetings, and Webinars.

The main content area is divided into several sections:

- Machine Learning, Data Science, Data Mining, Big Data, Analytics, AI**: A sidebar menu listing various categories like Software, Jobs - Industry, Academic, Meetings, Conferences, Companies, Courses in Big Data, Data Science, Datasets, Education, Webcasts, and Podcasts.
- Most Recent**: A list of recent articles, including "Which Deep Learning Framework is Growing Fastest?" and "Build Your First Chatbot Using Python & NLTK".
- Latest News | Tutorials**: A list of news items, such as "Build Your First Chatbot Using Python & ..." and "Powerful like your local notebook. Sharable I...".
- Opinions, Interviews, Reports**: A list of opinion pieces, including "Which Deep Learning Framework is Growing Fast..." and "Interview Questions for Data Science - ...".
- Right Sidebar (Tweets)**: A section titled "Tweets by @kdnuggets" showing several tweets from the account, including one about "Faster #ScikitLearn with @intsidevtools Distribution for #Python buff.ly/2vBgtM9" and another about "Seven Guidelines to Ensure #Ethical #AI".

This screenshot shows a detailed view of the KDnuggets website content:

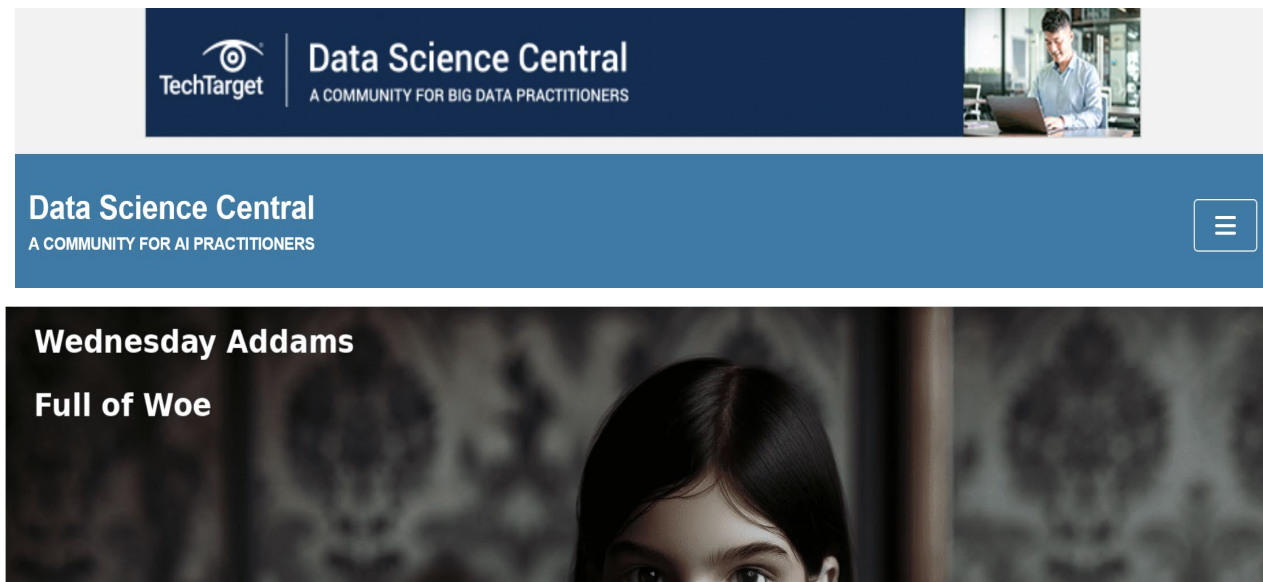
- Courses | Meetings | Webcasts**: A list of events, including "MS in Health Analytics Online", "On Stage at PAW Industry 4.0: Bayer, Contin...", "Delivering Trusted AI with DataRobot and Micr...", "Meet the World's Leading AI & Deep...", and "Join the new generation of AI technologists".
- Jobs**: A list of job openings, such as "Monash University: Lecturer/Senior Lecturer (...)", "Fors Marsh Group: Lead Data Scientist [Arling...", "NASDAQ 100: Leader of Data Science [East Coast]", and "Bottomline Technologies, Inc: Data Scientist ...".
- Top Stories Past 30 Days**: A table with two columns: "Most Popular" and "Most Shared".
 

Most Popular	Most Shared
1. <b>The most desired skill in data science</b>	1. <b>Another 10 Free Must-See Courses for Machine Learning and Data Science</b>
2. <b>Top 10 Coding Mistakes Made by Data Scientists</b>	2. <b>Top 10 Coding Mistakes Made by Data Scientists</b>
3. <b>How to Recognize a Good Data Scientist Job From a Bad One</b>	3. <b>Data Visualization in Python: Matplotlib vs Seaborn</b>
4. <b>Data Visualization in Python: Matplotlib vs Seaborn</b>	4. <b>An Introduction on Time Series Forecasting with Simple Neural Networks &amp; LSTM</b>
5. <b>An Introduction on Time Series Forecasting with Simple Neural</b>	
- More Recent Stories**: A list of recent articles, including "Learn About Data Science & the Future of Investing from H...", "Interview Questions for Data Science - Three Case Interv...", "Seeking KDnuggets Editors, part-time, work remotely", "Normalization vs Standardization—Quantitative analysis", "MS in Health Analytics Online", "On Stage at PAW Industry 4.0: Bayer, Continental, HP, Vodafone...", "Strata SF day 1 Highlights: from Edge to AI, scoring AI projec...", "Top Data Science and Machine Learning Methods Used in 2018, 2019", "Pandas DataFrame Indexing", "Monash University: Lecturer/Senior Lecturer (Machine Learning ...)", "Delivering Trusted AI with DataRobot and Microsoft", "AI and the data production landscape", and "The most desired skill in...".

## AUTRES SITES WEB

Site Data Science Central

<https://www.datasciencecentral.com/>



### Eight Techniques for Powering ChatGPT Content

Kurt Cagle | December 22, 2023 at 6:05 am



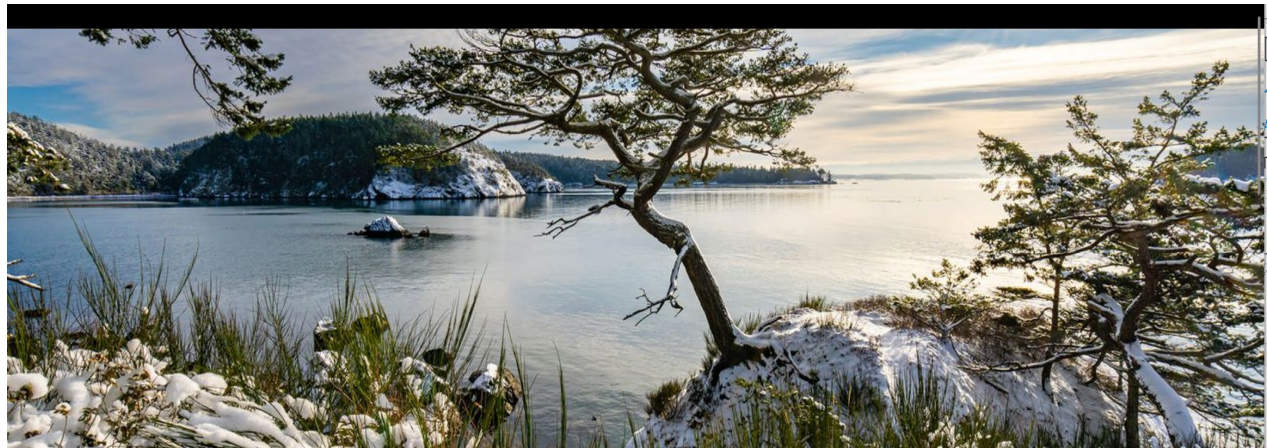
### Video: Introduction to Machine Learning

Vincent Granville · October 5, 2022 at 2:21 pm

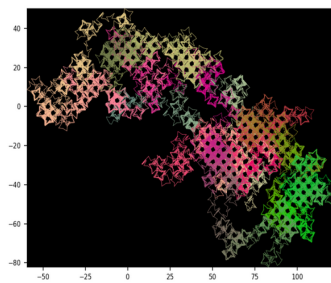
This 30 minutes video features my interview about the upcoming course "Intuitive Machine Learning", based on my new book with the same title. Hosted by...

## AUTRES SITES WEB

Site de Vincent Granville : <https://mltechniques.com/>



 Menu



Books Featured Posts Generative AI Machine Learning Python Synthetic Data

### My Top 10 GenAI Articles of the Year



VINCENT GRANVILLE DECEMBER 22, 2023

Here is some good reading for the holiday season. More than just reading as the material includes full Python implementations and datasets. The most up-to-date versions are in my new book *Statistical Optimization for GenAI and Machine Learning*, available here. As a courtesy, if you buy it by December 31, you are entitled to a [...]

READ MORE

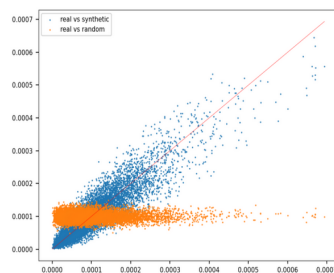


Search ...

Search

#### Recent Posts

- [My Top 10 GenAI Articles of the Year](#)
- [Genome: Synthesizing DNA Sequences with LLM Techniques](#)
- [10 GenAI Notebooks: OpenAI, LLM, RAG, GPT, and More](#)
- [Easy Trick to Debias GenAI Models: Quantile Convolution](#)
- [New Book: Understanding Deep Learning](#)



Data Sets Explainable AI Featured Posts Generative AI Natural Language Processing

### Genome: Synthesizing DNA Sequences with LLM Techniques



VINCENT GRANVILLE DECEMBER 8, 2023

This methodology is not focused on genome data alone. The purpose is to design a generic solution that may also work in other contexts, such as synthesizing molecules. The problem involves dealing with a large amount of "text". Indeed, the sequences discussed here consist of letter arrangements, from an alphabet that has 5 symbols: A, [...]

#### Recent Comments

messerh5467 om

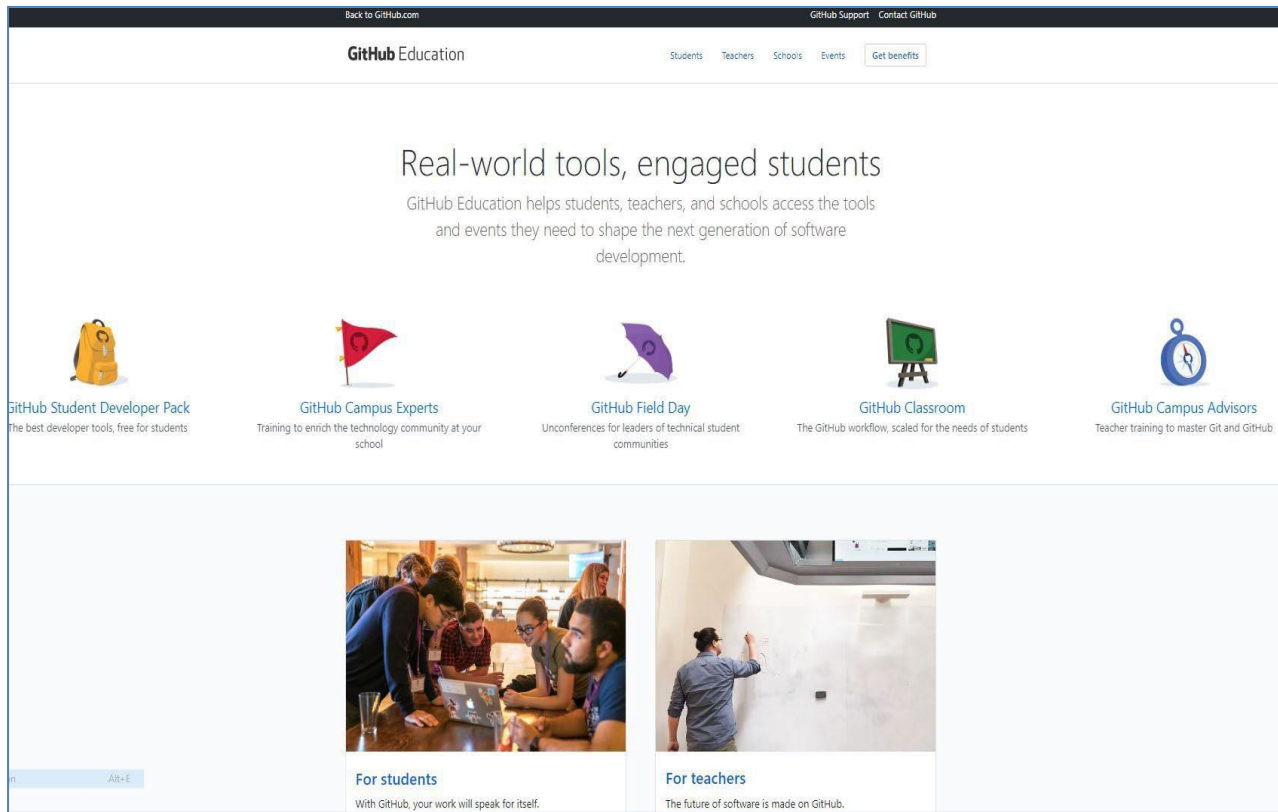
Liste de **90 articles les plus lus** des 10 dernières années

Consulter le fichier Excel sur le site du cours MTH8302

[https://cours.polymtl.ca/mth6301/mth8302/2019-DataScienceCentral-Selection\\_Articles.xlsx](https://cours.polymtl.ca/mth6301/mth8302/2019-DataScienceCentral-Selection_Articles.xlsx)

**AUTRES SITES WEB**

Site GitHub <https://education.github.com/>



Back to GitHub.com GitHub Support Contact GitHub

GitHub Education Students Teachers Schools Events Get benefits

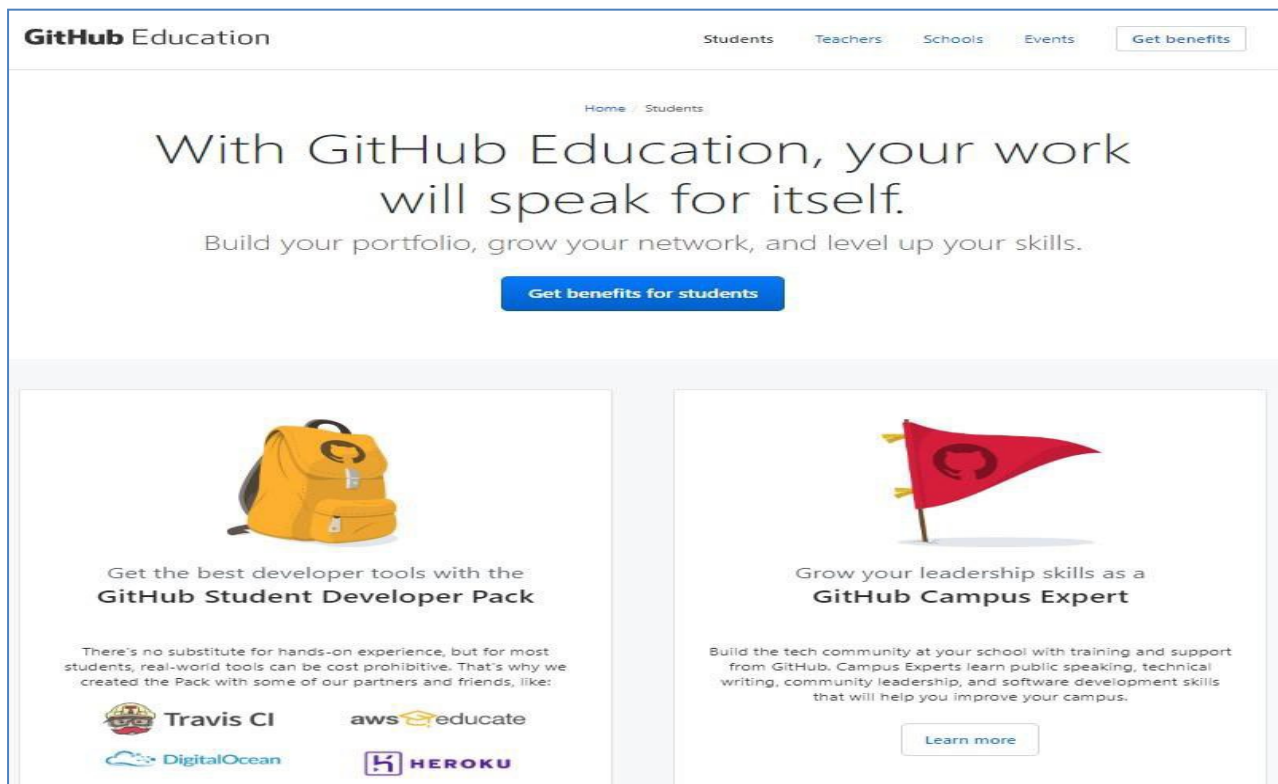
## Real-world tools, engaged students

GitHub Education helps students, teachers, and schools access the tools and events they need to shape the next generation of software development.

- GitHub Student Developer Pack**  
The best developer tools, free for students
- GitHub Campus Experts**  
Training to enrich the technology community at your school
- GitHub Field Day**  
Unconferences for leaders of technical student communities
- GitHub Classroom**  
The GitHub workflow, scaled for the needs of students
- GitHub Campus Advisors**  
Teacher training to master Git and GitHub

**For students**  
With GitHub, your work will speak for itself.

**For teachers**  
The future of software is made on GitHub.




GitHub Education Students Teachers Schools Events Get benefits

Home Students

## With GitHub Education, your work will speak for itself.

Build your portfolio, grow your network, and level up your skills.


[Get benefits for students](#)



### Get the best developer tools with the GitHub Student Developer Pack

There's no substitute for hands-on experience, but for most students, real-world tools can be cost prohibitive. That's why we created the Pack with some of our partners and friends, like:

- Travis CI
- DigitalOcean
- aws educate
- HEROKU



### Grow your leadership skills as a GitHub Campus Expert

Build the tech community at your school with training and support from GitHub. Campus Experts learn public speaking, technical writing, community leadership, and software development skills that will help you improve your campus.

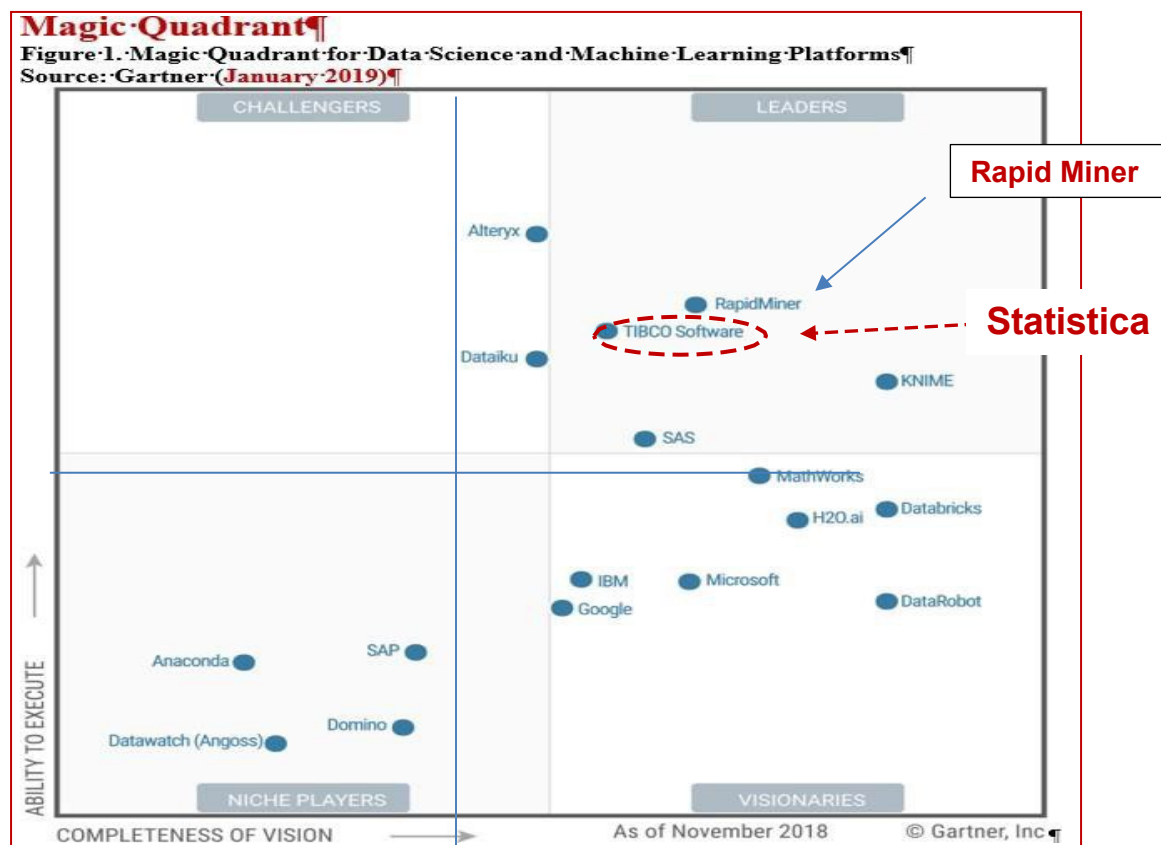
[Learn more](#)

## LOGICIELS STATISTIQUES / MACHINE LEARNING

<https://www.gartner.com/doc/reprints?id=1-65OPDHH&ct=190125&st=sb&submissionGuid=e9e519ac-077f-4a78-9523-7a2399a9133d>

### Magic Quadrant for Data Science and Machine Learning Platforms

published 28 January 2019 - ID G00354456



..... **Successful consolidation:** On a single platform, **TIBCO** brings together powerful visualization capabilities, strong descriptive analytics and visionary predictive analytics features (**from Statistica and Alpine Data, now rebranded as Spotfire Data Science**). At the same time, TIBCO has maintained its platform's necessary **extensibility to open-source environments**. Open-source code, for example, can be developed within the platform or in an outside environment and then seamlessly integrated into the data science pipeline's workflow. .... **page 23**

**Statistica (TIBCO Data Science)** <https://www.tibco.com/fr/products/data-science>

**JMP Pro (division de SAS)** [https://www.jmp.com/fr\\_fr/home.html](https://www.jmp.com/fr_fr/home.html)

**Téléchargement de Statistica et JMP Pro :** consulter lien

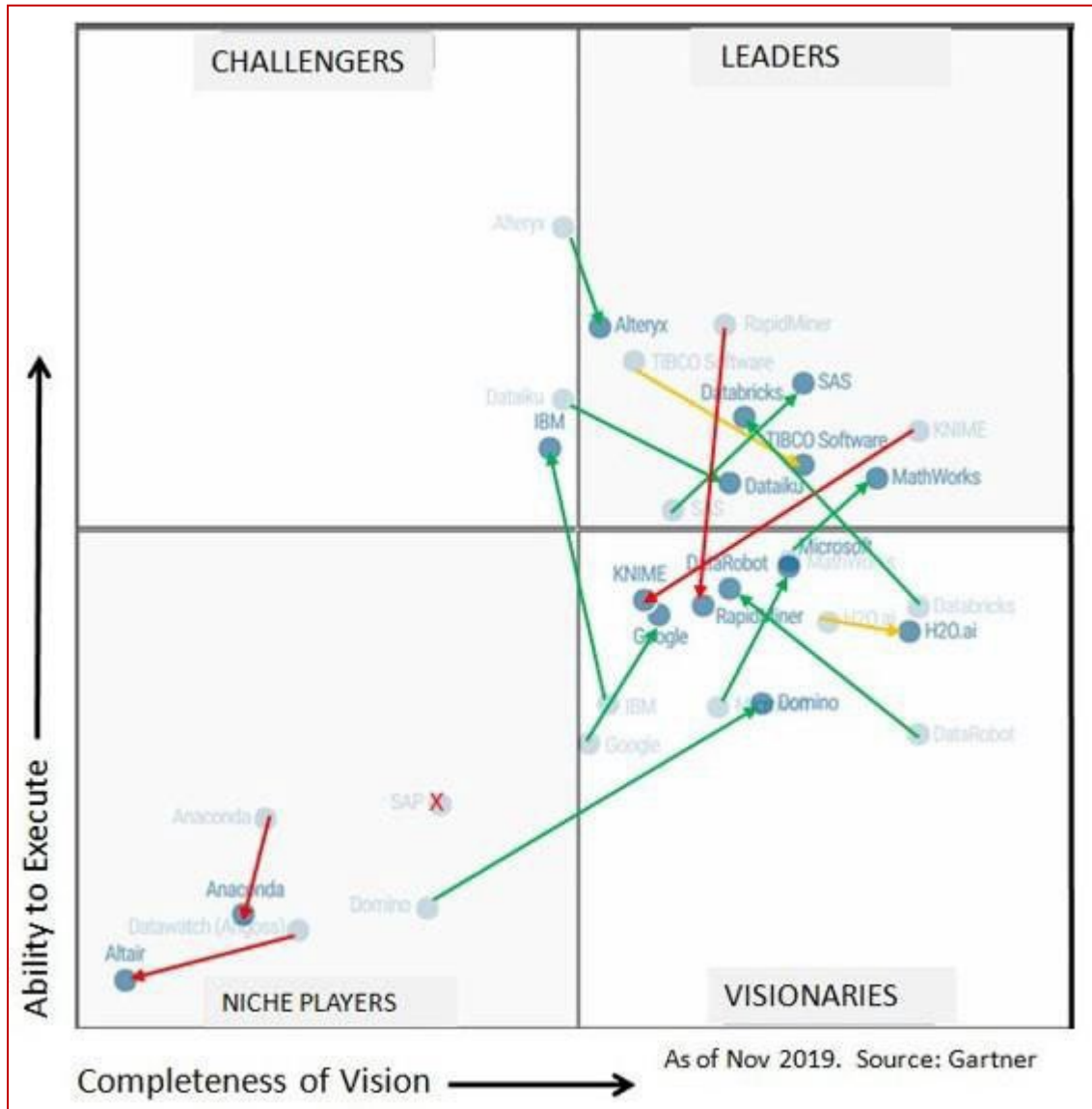
<https://www.polymtl.ca/si/licences-institutionnelles>

**AUTRES** **RapidMiner** <https://rapidminer.com/>

**KNIME** <https://www.knime.com/>

**H2ai** [H2O.ai | AI Cloud Platform](https://h2o.ai/)

**SAS** [https://www.sas.com/en\\_ca/home.html](https://www.sas.com/en_ca/home.html)

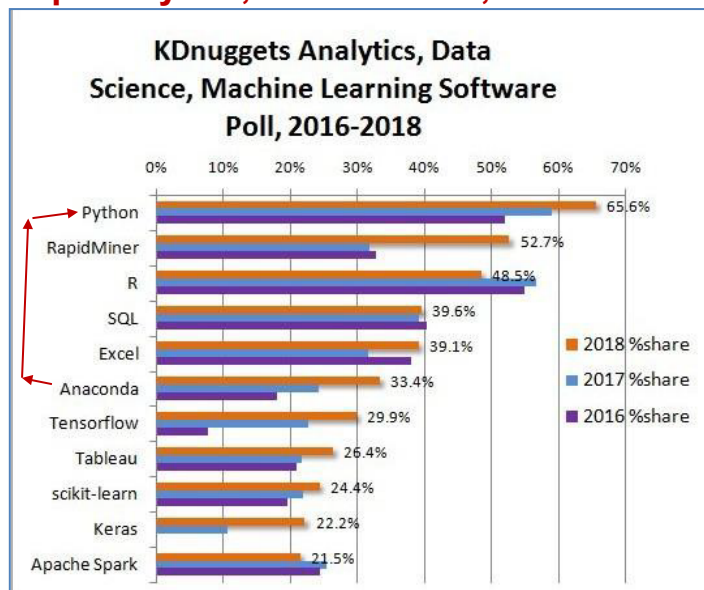


- **Leaders** : Alteryx, Dataiku, Databricks, MathWorks, SAS, **TIBCO (Statistica)**
- **Challengers**: IBM
- **Visionaries** : DataRobot, Domino, Google, H2O.ai, KNIME, Microsoft, RapidMiner
- **Niche players**: Anaconda, Altair (former DataWatch / Angoss)

## LOGICIELS STATISTIQUES

<https://www.kdnuggets.com/2018/05/poll-tools-analytics-data-science-machine-learning-results.html>

### Top Analytics, Data Science, Machine Learning Tools



**Python eats away at R:**

**Top Software for Analytics,**

**Data Science, Machine Learning**

**in 2018: Trends and Analysis**

**Why R in Data Science?**

<https://dimensionless.in/why-r-programming-in-data-science/>

**Question : R ou Python ?**

**Réponse de Bernard Clément : les 2 + ....**

**La tendance se maintient : Python prend plus le dessus sur R depuis 2018**

#### Recommandation de Besse

R pour modéliser et interpréter des modèles statistiques  
 Python pour des modélisations efficaces à seule fin prédictive  
 au détriment de l'interprétation.

**Construction de tableaux de bords (important dans les organisations)**

Les deux approches pouvant d'ailleurs être traitées de façon complémentaire.

**Autre recommandation : Rapid Miner - The Full Data Science Lifecycle**

**RapidMiner** stands out due to its comprehensive nature. It's your true companion throughout your complete data science lifecycle. From data modeling and analysis to data deployment and monitoring, this tool covers it all. It offers a visual workflow design, eliminating the need for intricate coding. This tool can also be used to build custom data science workflows and algorithms from scratch. The extensive data preparation features in RapidMiner enable you to deliver the most refined version of data for modeling. Here are some of the key features:

- It simplifies the data science process by providing a visual and intuitive interface.
- RapidMiner's connectors make data integration effortless, regardless of size or format.

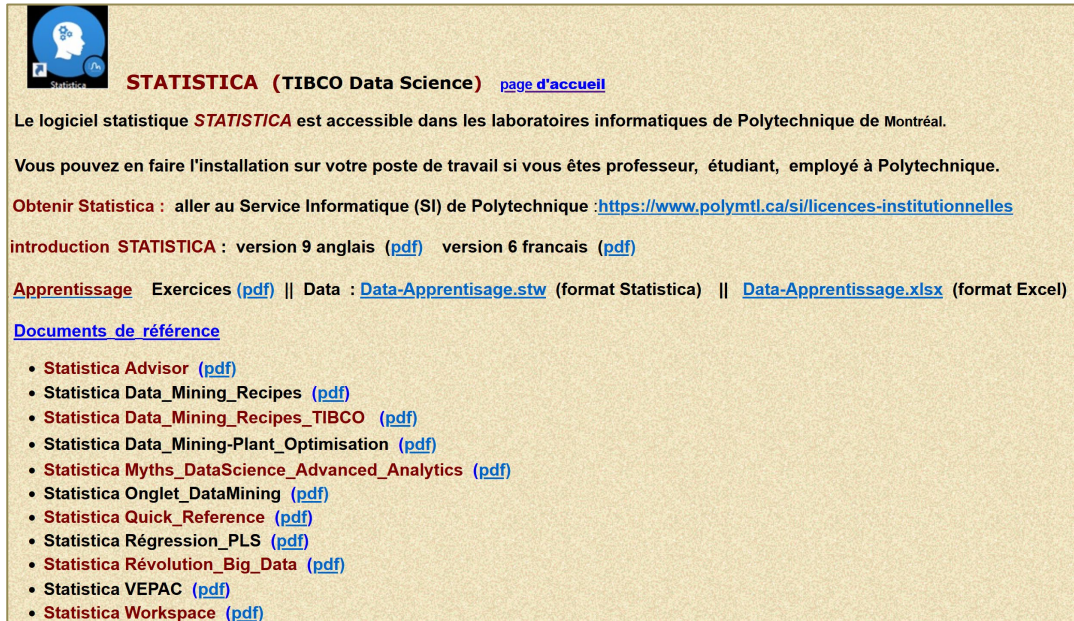
<https://www.kdnuggets.com/a-comparative-overview-of-the-top-10-open-source-data-science-tools-in-2023>

<https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1505653-rapidminer/>

## LOGICIEL STATISTICA

<http://www.cours.polymtl.ca/mth6301/STATISTICA.htm>

disponible en Windows seulement.



**STATISTICA (TIBCO Data Science)** [page d'accueil](#)

Le logiciel statistique **STATISTICA** est accessible dans les laboratoires informatiques de Polytechnique de Montréal.

Vous pouvez en faire l'installation sur votre poste de travail si vous êtes professeur, étudiant, employé à Polytechnique.

**Obtenir Statistica** : aller au Service Informatique (SI) de Polytechnique : <https://www.polymtl.ca/si/licences-institutionnelles>

**introduction STATISTICA** : version 9 anglais ([pdf](#)) version 6 français ([pdf](#))

**Apprentissage** Exercices ([pdf](#)) || Data : [Data-Apprentissage.stw](#) (format Statistica) || [Data-Apprentissage.xlsx](#) (format Excel)

**Documents de référence**

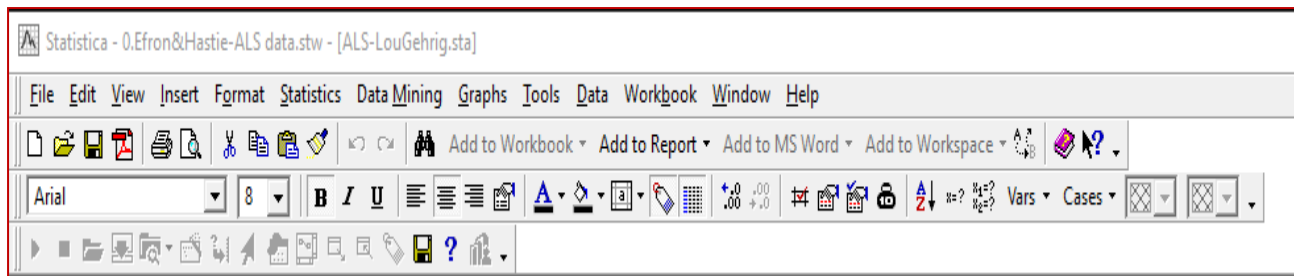
- [Statistica Advisor \(pdf\)](#)
- [Statistica Data\\_Mining\\_Recipes \(pdf\)](#)
- [Statistica Data\\_Mining\\_Recipes\\_TIBCO \(pdf\)](#)
- [Statistica Data\\_Mining-Plant\\_Optimisation \(pdf\)](#)
- [Statistica Myths\\_DataScience\\_Advanced\\_Analytics \(pdf\)](#)
- [Statistica Onglet\\_DataMining \(pdf\)](#)
- [Statistica Quick\\_Reference \(pdf\)](#)
- [Statistica Régression\\_PLS \(pdf\)](#)
- [Statistica Révolution\\_Big\\_Data \(pdf\)](#)
- [Statistica VEPAC \(pdf\)](#)
- [Statistica Workspace \(pdf\)](#)

**Procédure d'installation de STATISTICA sur votre poste de travail :**

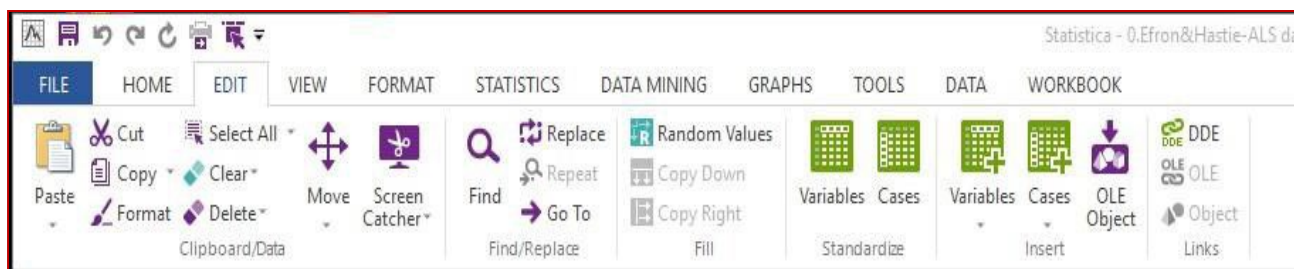
<https://www.polymtl.ca/si/licences-institutionnelles>

## INTERFACE

**mode standard**



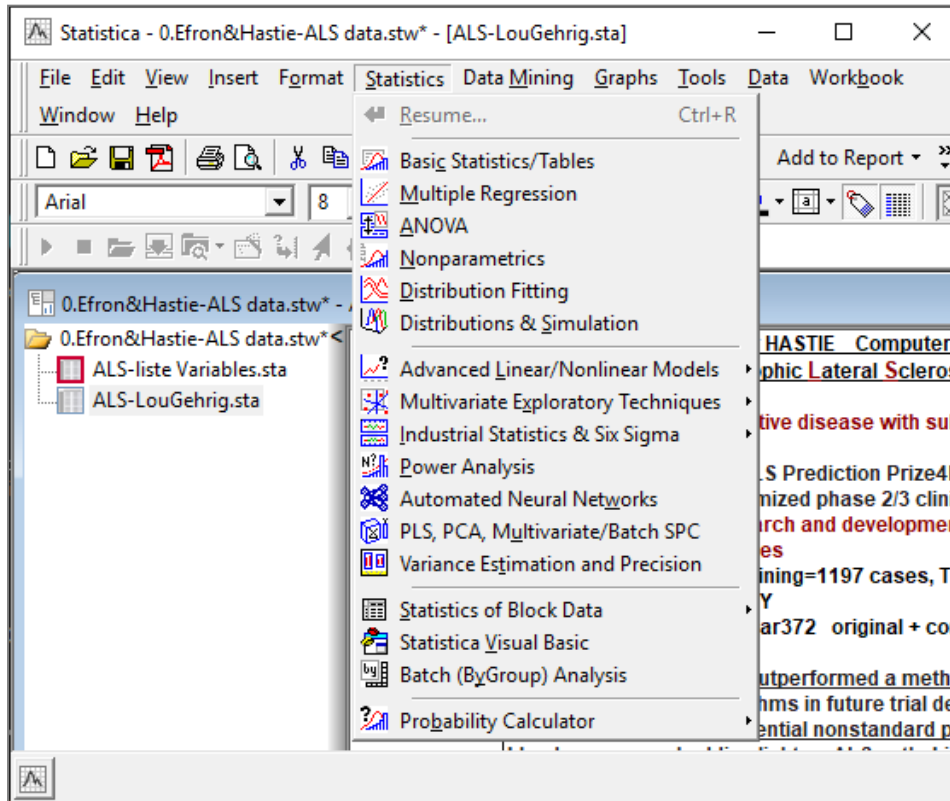
**mode Ribbon Bar**



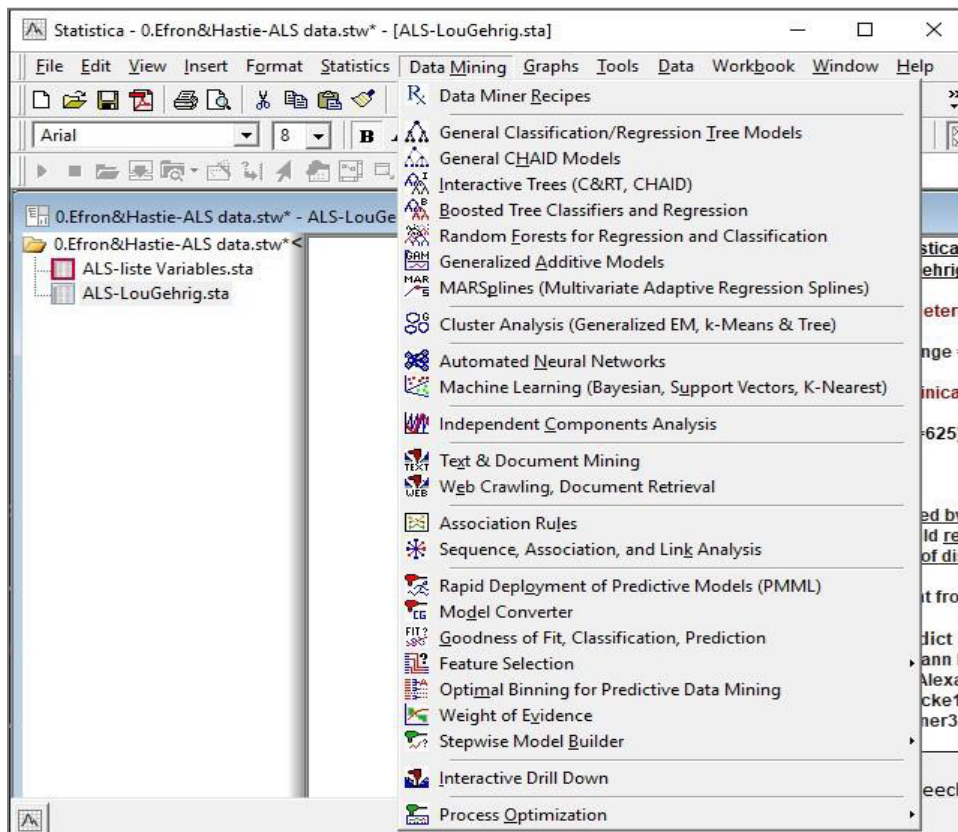
**Possibilité d'intégration de R et Python à l'intérieur de Statistica**



## Menu Statistics



## Menu Data Mining



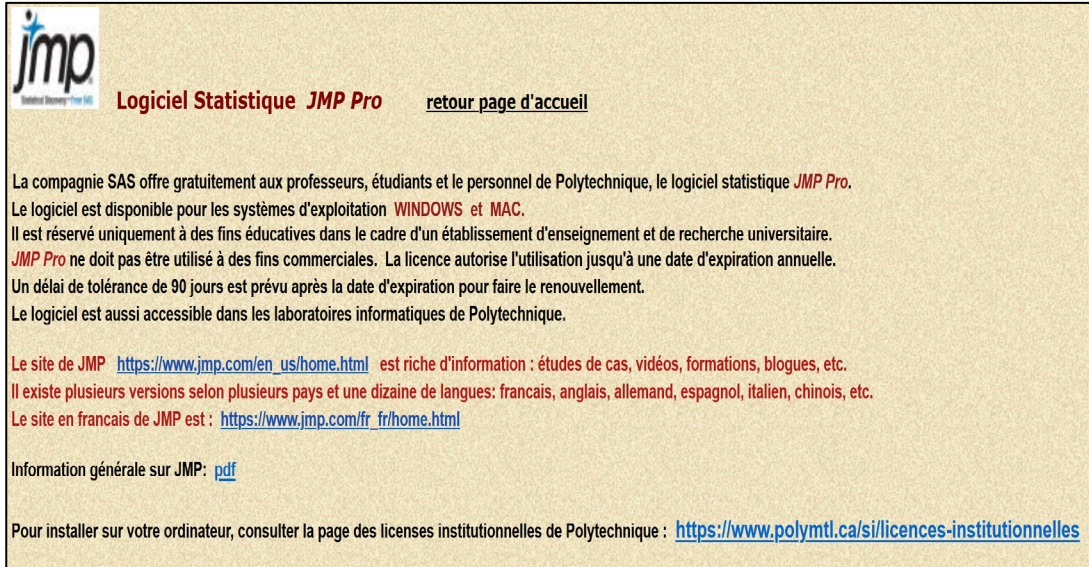
## LOGICIEL JMP (version Pro)

<http://cours.polymtl.ca/mth6301/JMP.htm>

disponible en version Windows et en version MAC

Demande de JMP : <https://www.polymtl.ca/si/licences-institutionnelles>

### Possibilité d'intégration de R et Python à JMP



Logiciel Statistique **JMP Pro** [retour page d'accueil](#)

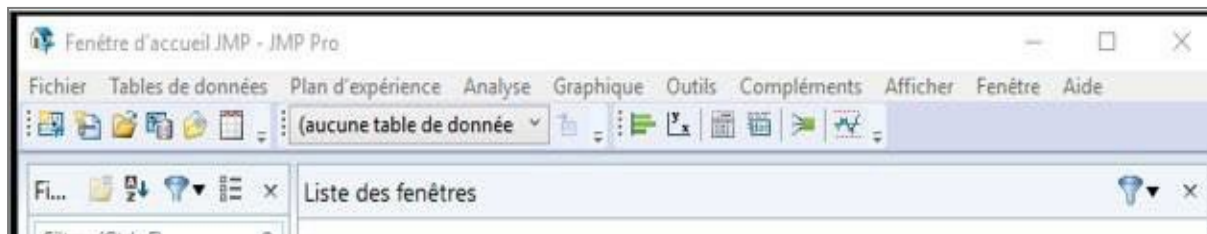
La compagnie SAS offre gratuitement aux professeurs, étudiants et le personnel de Polytechnique, le logiciel statistique **JMP Pro**. Le logiciel est disponible pour les systèmes d'exploitation **WINDOWS** et **MAC**. Il est réservé uniquement à des fins éducatives dans le cadre d'un établissement d'enseignement et de recherche universitaire. **JMP Pro** ne doit pas être utilisé à des fins commerciales. La licence autorise l'utilisation jusqu'à une date d'expiration annuelle. Un délai de tolérance de 90 jours est prévu après la date d'expiration pour faire le renouvellement. Le logiciel est aussi accessible dans les laboratoires informatiques de Polytechnique.

Le site de JMP [https://www.jmp.com/en\\_us/home.html](https://www.jmp.com/en_us/home.html) est riche d'information : études de cas, vidéos, formations, blogues, etc. Il existe plusieurs versions selon plusieurs pays et une dizaine de langues: français, anglais, allemand, espagnol, italien, chinois, etc. Le site en français de JMP est : [https://www.jmp.com/fr\\_fr/home.html](https://www.jmp.com/fr_fr/home.html)

Information générale sur JMP: [pdf](#)

Pour installer sur votre ordinateur, consulter la page des licences institutionnelles de Polytechnique : <https://www.polymtl.ca/si/licences-institutionnelles>

### INTERFACE



### Menu Analyse

