

INTRODUCTION TO DATA ANALYSIS

Instructor: Patrick Boily

Patrick is a professor in the Department of Mathematics and Statistics at the University of Ottawa and a founding member of the Data Action Lab. Since 1999, he has taught more than 50 courses at Ottawa-area universities and conducted more than 150 days of workshops for federal government officials on data science, data visualization, and statistical analysis. His interests include the use of mathematics and statistics for decision support; he has managed and participated in over 35 consulting projects throughout his career.

Comments

- Overview course - programming and statistical skills are neither required, nor essential.
- Exercises will be completed with the help of free online software (no installation required).

Module 1 – Data Insight Fundamentals

- Analysis Planning
- Data 101: Basic Data Concepts
- Some Practical Definitions
- Workflows and Pipelines
- Models and Systems Thinking
- Ethical Considerations and Best Practices
- Exercises

Module 4 – Basic Data Analysis

- Insights and # Crunching: Core Concepts
- Insights and # Crunching: Core Techniques
- Data Descriptions (in-depth)
- Correlation
- Regression Analysis
- Time Series and Control Charts
- Exercises

Module 2 – Data Collection and Data Management

- Data Sources
- Sampling Theory and Study Design
- Web Scraping and Automated Data Collection
- Data and Knowledge Modeling
- Exercises

Module 5 – Statistical Learning

- Learning Context
- Association Rules
- Classification
- Clustering
- Issues and Challenges
- Exercises

Module 3 – Data Processing and Data Cleaning

- Data Cleaning
- Data Reduction and Transformations
- Data Quality and Data Validation
- Exercises

Module 6 – Data Visualization Basics

- Pre-Analysis Data Visualization
- Post-Analysis Data Visualization
- Visualization Catalogue
- Basic Rules of Design and Layout
- Dashboards
- Exercises

INTRODUCTION À L'ANALYSE DES DONNÉES

Formateur : Patrick Boily

Patrick est professeur au département de mathématiques et de statistique de l'Université d'Ottawa et membre fondateur du Data Action Lab. Depuis 1999, il a donné plus de 50 cours dans des universités de la région d'Ottawa et a animé plus de 150 jours d'ateliers à l'intention des fonctionnaires du Gouvernement fédéral, portant sur la science des données, la visualisation des données, et l'analyse statistique. Il s'intéresse à l'appui décisionnel que peut apporter l'utilisation des mathématiques et des statistiques, et a géré et participé à plus de 35 projets de consultation au cours de sa carrière.

Notes:

- Cours de survol – les compétences en programmation et en statistique ne sont pas essentielles
- Les exercices se feront à l'aide d'un logiciel en ligne disponible gratuitement (aucune installation requise)

1. Principes fondamentaux de l'analyse des données

- Planification des analyses
- Données 101: notions de données de base
- Quelques définitions pratiques
- Flux de travail et sources
- Modèles et pensée systémique
- Considérations éthiques et meilleures pratiques
- Exercices

2. Collecte et gestion des données

- Sources des données
- Théorie de l'échantillonnage et planification d'étude
- Moissonnage du web et collecte automatisée de données
- Modélisation des données et des connaissances
- Exercices

3. Traitement des données

- Nettoyage des données
- Réduction et transformation des données
- Qualité de données et validation des données
- Exercices

4. Techniques de base d'analyse des données

- Aperçus et calculs : concepts de base
- Aperçus et calculs : techniques de base
- Descriptions de données
- Corrélation
- Analyse de régression
- Séries chronologiques et cartes de contrôle
- Exercices

5. Apprentissage statistique

- Contexte d'apprentissage
- Règles d'association
- Classification
- Regroupement
- Enjeux et défis
- Exercices

6. Principes de la visualisation des données

- Visualisation des données avant l'analyse
- Visualisation des données post-analyse
- Catalogue des visualisations
- Règles de base de la conception et de la mise en page
- Tableau de bord
- Exercices