The field of data mining lies at the confluence of predictive analytics, statistical analysis, and business intelligence. Due to the ever-increasing complexity and size of data sets and the wide range of applications in computer science, business, and health care, the process of discovering knowledge in data is more relevant than ever before.

This book provides the tools needed to thrive in today’s big data world. The author demonstrates how to leverage a company’s existing databases to increase profits and market share, and carefully explains the most current data science methods and techniques. The reader will “learn data mining by doing data mining”. By adding chapters on data modelling preparation, imputation of missing data, and multivariate statistical analysis, Discovering Knowledge in Data, Second Edition remains the eminent reference on data mining.

- The second edition of a highly praised, successful reference on data mining, with thorough coverage of big data applications, predictive analytics, and statistical analysis.

- Includes new chapters on Multivariate Statistics, Preparing to Model the Data, and Imputation of Missing Data, and an Appendix on Data Summarization and Visualization

- Offers extensive coverage of the R statistical programming language

- Contains 280 end-of-chapter exercises
• Includes a companion website for university instructors who adopt the book

## ABOUT THE AUTHOR

**Daniel T. Larose** earned his PhD in Statistics at the University of Connecticut. He is Professor of Mathematical Sciences and Director of the Data Mining programs at Central Connecticut State University. His consulting clients have included Microsoft, *Forbes* Magazine, the CIT Group, KPMG International, Computer Associates, and Deloitte, Inc. This is Larose’s fourth book for Wiley.

**Chantal D. Larose** is an Assistant Professor of Statistics & Data Science at Eastern Connecticut State University (ECSU). She has co-authored three books on data science and predictive analytics. She helped develop data science programs at ECSU and at SUNY New Paltz. She received her PhD in Statistics from the University of Connecticut, Storrs in 2015 (dissertation title: Model-based Clustering of Incomplete Data).

## RELATED RESOURCES

**Instructor**

View Instructor Companion Site

## SERIES

Wiley Series on Methods and Applications in Data Mining

To purchase this product, please visit [https://www.wiley.com/en-us/9780470908747](https://www.wiley.com/en-us/9780470908747)