TMM026 Introductory Data Science

Learning Materials Overview

Learning Outcomes	*Modern Dive* (R)	*Statistical and Computational Thinking* (Python)	*Data Science in a Box* (R)	*Statistics with Randomization and Simulation* (R)	*Data Computing 2e* (R)	*Modern Data Science with R* (R)	*Data Science: A First Introduction* (R)
1.1 Types and Sources of Data	Chapter 4 – Importing Data and "Tidy" Data	Chapter 4 – Data Types	Chapter 7 – Exploring Data	Chapter 4 – Data Types	Chapter 1 – Tidy Data	Chapter 9 Statistical Foundations	Chapter 1 R, Jupyter and tidyverse
1.2 Collecting Data	Chapter 4 – Importing Data and "Tidy" Data		Chapter 7 – Exploring Data		Chapter 16 – Data Cleaning	Chapter 4 Data wrangling on one table	Chapter 2 Reading In Data locally and from the web
1.3 Organizing and Standardizing Data	Chapter 4 – Importing Data and "Tidy" Data	Chapter 4 – Data Types	Chapter 19 - Pedagogy	Chapter 6 - Tables	Chapter 1 – Tidy Data	Chapter 5 Data wrangling on multiple tables	Chapter 5 Collaboration and version control
1.4 Cleaning Data	Chapter 4 – Importing Data and "Tidy" Data		Chapter 19 - Pedagogy		Chapter 16 – Data Cleaning	Chapter 6 Tidy data	Chapter 3 Cleaning and Wrangling Data
2.1 Traditional Plots	Chapter 2 – Data Visualization	Chapter 7 – Visualization	Chapter 19 - Pedagogy	Chapter 7 - Visualization	Chapter 5 – Intro to Data Graphics	Chapter 2 Data visualization	Chapter 4 Effective data visualization
2.2 Single vs. Two or More Variables	Chapter 2 – Data Visualization	Chapter 7 – Visualization	Chapter 19 - Pedagogy	Chapter 7 - Visualization	Chapter 10 – More Data Verbs	Chapter 3 A grammar for graphs	

2.3 Dynamic Visualization							
2.4 Distribution Maps					Chapter 5 – Intro to Data Graphics		
2.5 Time Series Plots	Chapter 2 – Data Visualization	Chapter 7 – Visualization		Chapter 7 - Visualization	Chapter 5 – Intro to Data Graphics		
2.6 Misleading Graphs							
3.1 Statistical Model	Chapter 7 - Sampling	Chapter 9 - Randomness				Section 13 Randomizing functions	
3.2 Estimator and sampling distribution	Chapter 7 - Sampling	Chapter 10 – Sampling and Empirical Distributions					
3.3 Simulation-Based Estimation and Prediction	Chapter 8 – Bootstrapping and Confidence Intervals	Chapter 9 - Randomness	Chapter 8 – Making Rigorous Conclusions	Chapter 9 - Randomness		Chapter 13 Simulation	Chapter 11: Introduction to statistical inference
3.4 Loss Function and Decision				Chapter 11 – Testing Hypotheses			
3.5 Model Diagnostics		Chapter 16 – Inference for Regression				Chapter 10 Predictive Modeling	Chapter 8: Regression I: K-nearest neighbors
3.6 Estimation (mathematical approaches)							

4.1 Machine Learning		Chapter 11 – Testing Hypotheses	Chapter 8 – Making Rigorous Conclusions	Chapter 11 – Testing Hypotheses		Appendix E Regression Modeling	Chapter 9: Regression II: Linear regression
4.2 Supervised Learning	Chapter 10 – Inference for Regression	Chapter 17 – Classification	Chapter 8 – Making Rigorous Conclusions	Chapter 17 – Classification	Chapter 18 – Machine Learning	Chapter 11 Supervised Learning	
4.3 Unsupervised Learning		Chapter 18 – Updating Predictions	Chapter 8 – Making Rigorous Conclusions	Chapter 17 – Classification	Chapter 18 – Machine Learning	Chapter 12 Unsupervised Learning	Classification II: Evaluating and tuning
4.4 Sentiment Analysis		Chapter 17 – Classification	Chapter 8 – Looking Forward	Chapter 17 – Classification	Chapter 18 – Machine Learning	Chapter 19 Text as Data	Chapter 10: Clustering
5.1 Laws and Regulations			Chapter 8 – Looking Forward			Chapter 8 Data Science Ethics	
5.2 Unfair discrimination and social bias			Chapter 8 – Looking Forward			Section 8.3 The role of data science in society	
5.3 Transparency of methods and analysis			Chapter 8 – Looking Forward			Section 8.4 Some settings for professional ethics	
5.4 Sampling bias			Chapter 8 – Looking Forward			Section 8.6 Algorithmic bias	

Additional Textbook Recommendations

Textbook	Author(s)
Statistics for Data Science: Leverage the power of statistics for Data Analysis, Classification, Regression, Machine Learning, and Neural Networks	James D. Miller
R for Everyone: Advanced Analytics and Graphics	Jared P. Lander
Moderndive	Chester Ismay and Albert Y. Kim
Statistics Using Technology (R)	Kate Kozak
Introduction to Data Science	Jeffrey Stanton
Data Science from Scratch: First Principles with Python, Second Edition	Joel Grus
Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data 1st Edition	EMC Education Services
Modern Data Science with R, Second Edition	Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton
Practical Data Science with R, Second Edition	Nina Zumal and John Mount

Additional Resource Recommendations

Introduction to Data Science: High School Statistics Curriculum
Curriculum Framework for Introductory Data Science