Learning Objectives in Introductory Data Science Courses

Learning objectives from $n = 29$ introductory data science syllabi were analyzed using qualitative content analysis and coded into 11 instructional themes. These syllabi were optionally submitted in conjunction with the MASDER-administered SOMADS survey. Multiple themes could be assigned to each syllabus using binary coding (0 = absent, 1 = present). Courses were grouped by prerequisite structure (None, Computer Science, and Statistics) to examine how prior preparation influences the emphasis of introductory data science instruction. Note syllabi from the STAT+CS group were not submitted. Figure 1 below summarizes the percentage of syllabi within each prerequisite group that included each instructional theme.

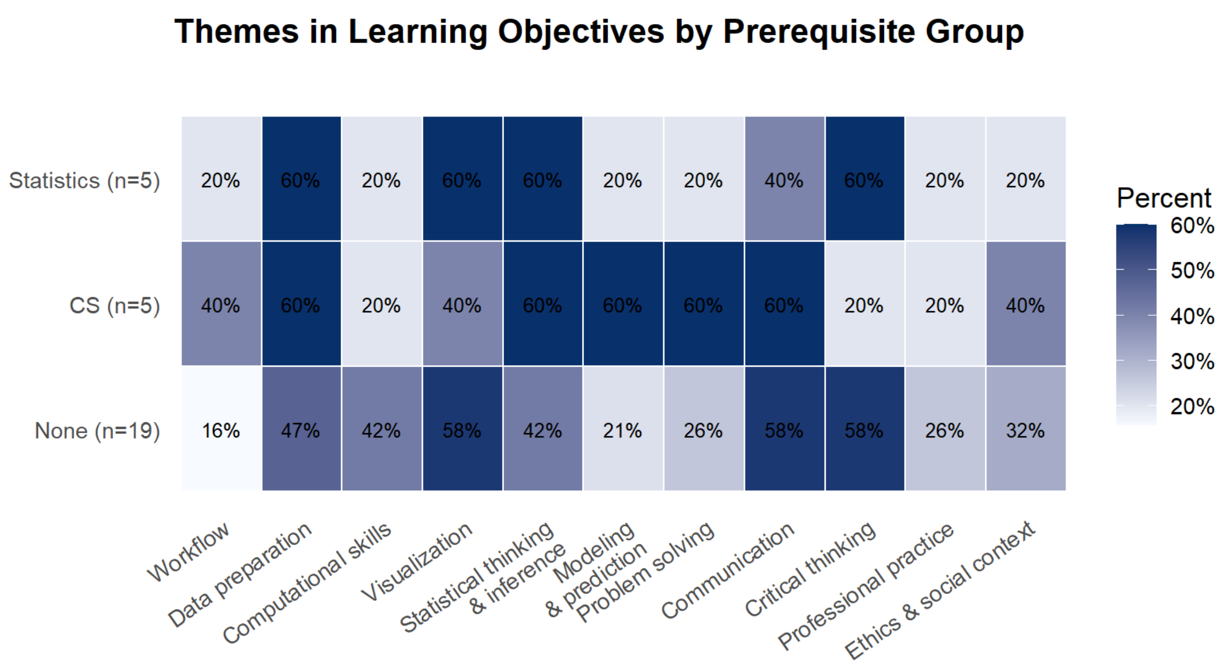

Figure 1: Identified themes within stated learning objectives by prerequisite group. The percentages represent percentage of syllabi within each group that includes the pedagogical theme.

Key Findings

Across all prerequisite groups, introductory data science courses emphasized a shared instructional core centered on data preparation, visualization, statistical thinking, communication, and critical thinking. Differences across prerequisite groups reflected the depth and emphasis of instruction rather than entirely different curricular content.

Courses with No Prerequisites

Courses without prerequisites emphasized broad data literacy and foundational analytic skills. Learning objectives most frequently focused on visualization (58%), communication (58%), critical thinking (58%), and data preparation (47%), with comparatively less emphasis on modeling and professional software practices. These courses introduced students to the complete data science process while developing confidence working with unfamiliar data.

Computer Science Prerequisite

Courses requiring a computer science prerequisite placed greater emphasis on modeling and prediction (60%), problem solving (60%), and communication (60%), while maintaining strong coverage of data preparation (60%) and statistical thinking (60%). Computational skills were introduced when programming itself was an instructional objective rather than simply a tool for completing other tasks. These courses emphasized implementing scalable computational workflows and applying predictive models.

Statistics Prerequisite

Courses requiring statistics prerequisites emphasized statistical thinking and inference (60%), visualization (60%), data preparation (60%), and critical thinking (60%). Learning objectives focused on inference, uncertainty, methodological justification, and interpreting evidence. Modeling was included but generally framed as an extension of statistical reasoning rather than purely predictive computation.

Conclusion

Prerequisite structure influenced how introductory data science was taught more than what was taught. Regardless of prerequisite structure, nearly all courses introduced students to the data science workflow through data preparation, visualization, statistical reasoning, communication, and critical thinking. Computer science prerequisites shifted emphasis toward computational implementation and predictive modeling, whereas statistics prerequisites emphasized inference, evaluation, and methodological reasoning. Courses without prerequisites balanced technical and analytical competencies with communication, providing an accessible introduction to data science.

Theme	Description	Representative Learning Objectives
Workflow	Understanding and applying the end-to-end data science process.	• Build an understanding of the data analytics lifecycle. • Apply the data analytics lifecycle to analyze a real-world dataset. • Design and implement end-to-end data analysis workflows.
Data Preparation	Acquiring, cleaning, transforming, integrating, and managing data for analysis.	• Import, clean, and transform data into a usable format for analysis. • Select or collect and clean data appropriate for analysis. • Programmatically acquire, clean, and transform data from diverse sources, including files, APIs, and databases.
Computational Skills	Developing computational thinking through programming concepts such as functions, loops, algorithms, debugging, and efficient code.	• Design and implement end-to-end data analysis workflows using a high-level programming language. • Manage data and code using modern software practices, including version control, modular design, and reproducibility tools. • Implement ideas using statistical software and computational tools.
Visualization	Creating, interpreting, and evaluating visual representations of data.	• Create and interpret high-quality data graphics. • Construct interactive graphics to facilitate inquiry. • Create data visualizations and assess their effectiveness for analytic purposes.
Statistical Thinking & Inference	Applying statistical reasoning to draw conclusions from data while accounting for uncertainty.	• Select and apply appropriate statistical models and hypothesis tests based on data type and research question. • Quantify and interpret uncertainty using confidence intervals, resampling techniques, and measures of variability. • Assess model assumptions, limitations, and the generalizability of statistical conclusions.

Modeling & Prediction	Building, applying, validating, and evaluating predictive or explanatory models.	• Apply computational and statistical methods to explore data and build predictive models. • Evaluate model performance using appropriate validation techniques and performance metrics. • Apply basic statistical and predictive methods to explore data and generate preliminary conclusions.
Problem Solving	Applying data science methods to investigate authentic questions and support decision making.	• Explain the value of data in scientific, social, and professional decision-making contexts. • Use computational and statistical methods to solve authentic problems. • Apply predictive methods to generate conclusions from data.
Communication	Communicating findings through written, oral, or visual forms for technical and nontechnical audiences.	• Interpret analytic results and communicate findings through clear written, visual, and oral data stories. • Communicate statistical findings clearly to both technical and nontechnical audiences, with attention to uncertainty and limitations. • Communicate data-based findings using effective data visualizations.
Critical Thinking	Evaluating evidence, comparing approaches, and justifying methodological decisions.	• Compare alternative analytic approaches and justify methodological choices using statistical reasoning. • Evaluate model performance using appropriate validation techniques and performance metrics. • Assess model assumptions, limitations, and the generalizability of statistical conclusions.
Professional Practice	Applying reproducible and professional software practices.	• Manage data and code using modern software practices, including version control, modular design, and reproducibility tools.

		• Integrate data storage and processing techniques suitable for larger or more complex datasets. • Document computational analyses using professional software practices.
Ethics & Social Context	Recognizing ethical, societal, and fairness issues associated with data and algorithms.	• Recognize ethical and social issues related to data collection, analysis, and use, including bias and fairness. • Evaluate ethical considerations in statistical analysis, including bias, inference validity, and responsible interpretation of results. • Analyze ethical risks associated with data-driven systems, including data misuse, automation, and infrastructure-level consequences.