

We analyzed instructor-provided definitions of data science from the MASDER-administered SOMADS survey using qualitative content analysis. Definitions were coded into seven conceptual themes: workflow, knowledge extraction, interdisciplinary thinking, problem solving, sensemaking, data characteristics, and communication. The heatmap summarizes the percentage of courses within each prerequisite group whose definitions included each theme, illustrating how instructors conceptualize data science for students with different prerequisite backgrounds.

Figure 1 summarizes the prevalence of each conceptual theme across prerequisite groups. While all groups described data science as an interdisciplinary field to some extent, the relative emphasis on workflow, knowledge extraction, sensemaking, and data characteristics differed by prerequisite structure.

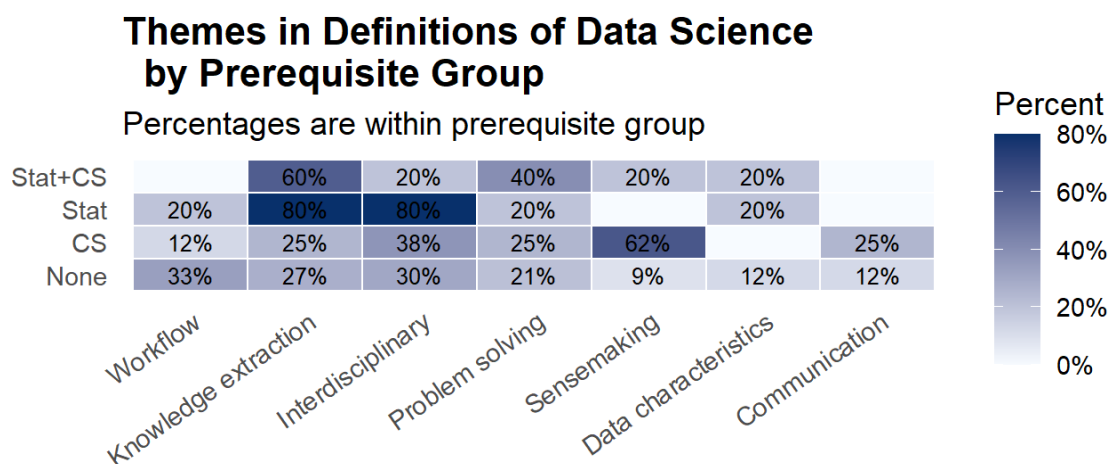


Figure 1: Themes in Definitions of Data Science by Prerequisite Group

Code	Description	Example Quote
Workflow	Defines data science as a process of collecting, cleaning, analyzing, and visualizing data.	"The process of collecting, storing, tidying, analyzing, and visualizing electronically stored information."
Knowledge Extraction	Emphasizes extracting information or knowledge from data.	"The study of the generalizable extraction of knowledge from data."
Interdisciplinary	Defines data science as the intersection of multiple disciplines.	"The intersection of statistics, computer science, and a field of application."

Problem Solving	Focuses on solving problems, answering questions, or supporting decisions.	"Data science is the study of and use of data to make decisions."
Sensemaking	Emphasizes understanding the world or telling stories with data.	"Telling a story with data."
Data Characteristics	Defines data science through the analysis of large or complex datasets.	"Large amounts of complex data."
Communication	Emphasizes communicating findings, audiences, or ethics.	"Conveying that meaning effectively and honestly."

Themes By Prerequisite

No Prerequisite (or Statistics Recommended)

Courses with no formal prerequisite ($n = 32$) most frequently defined data science as a workflow involving collecting, managing, analyzing, and visualizing data. Many definitions also emphasized knowledge extraction, interdisciplinary perspectives, and, to a lesser extent, problem solving and sensemaking through data.

Like the CS prerequisite courses, data science was also frequently described as the intersection of statistics, computer science, and domain expertise (Venn Diagram).

For example, in courses with a no prerequisite, instructors defined data science with the phrases – “Data science is ...”

- “...methods to work with and analyze large amounts of data and extract the underlying the story in the data.”
- Data science is the study of and use of data to make decisions.
- “...a dynamic and universal discipline. It is an umbrella term that covers many individual skills: data management, data wrangling, data workflow, data visualization, programming, communication, collaboration, and data ethics, but also including applied statistics and machine learning. In our increasingly data-driven world, this course will provide you with skills needed to solve problems, make predictions, and provide answers to research questions in the discipline of your choice.”

CS Prerequisite (only)

Courses with a computer science prerequisite ($n = 8$) most commonly framed data science as an interdisciplinary field integrating computer science, statistics, and domain expertise. Compared with other prerequisite groups, these definitions more frequently emphasized sensemaking,

describing data science as a means of understanding the world or communicating stories from data.

For example, in courses with a CS prerequisite, instructors defined data science with the phrases – “Data science is ...”

- “using data to understand our world. Therefore, it requires a combination of computer science, math/statistics, and domain knowledge.
- “[t]he intersection of statistics, computer science, and a field, or fields, of application. applied predictive and descriptive statistics for the computationally inclined. Stats for those who hate design and testing.”

Statistics Prerequisite (only)

Courses with a statistics prerequisite ($n = 5$) primarily emphasized knowledge extraction, describing data science as the process of extracting meaningful information or insights from data. Many definitions also characterized the field as interdisciplinary, combining statistics, mathematics, and computer science.

For example, in courses with a statistics prerequisite (only), instructors defined data science with the phrases – “Data science is ...”

- “...a[n] interdisciplinary field that utilizes scientific methods and technology to analyze and interpret complex data, extracting valuable insights and informing decision-making processes across diverse industries.
- the “...[s]cience of extracting meaningful information from data. It combines the elements of computer science, mathematics, and statistics.”

Statistics+CS Prerequisite

Courses requiring both statistics and computer science prerequisites ($n = 5$) most often emphasized knowledge extraction, while also distinguishing data science by its focus on large or complex datasets. Compared with the other prerequisite groups, these courses most explicitly associated data science with the characteristics and scale of modern data.

For example, in courses with both statistics and CS prerequisites instructors defined data science with the phrases – “Data science is ...”

- “... the application of data centric, computational, and inferential thinking to: Understand the world (science) and Solve problems (engineering).”
- “...a field of bigdata geared towards providing meaningful information based on large amounts of complex data.”