



DEEP  
LEARNING  
INSTITUTE



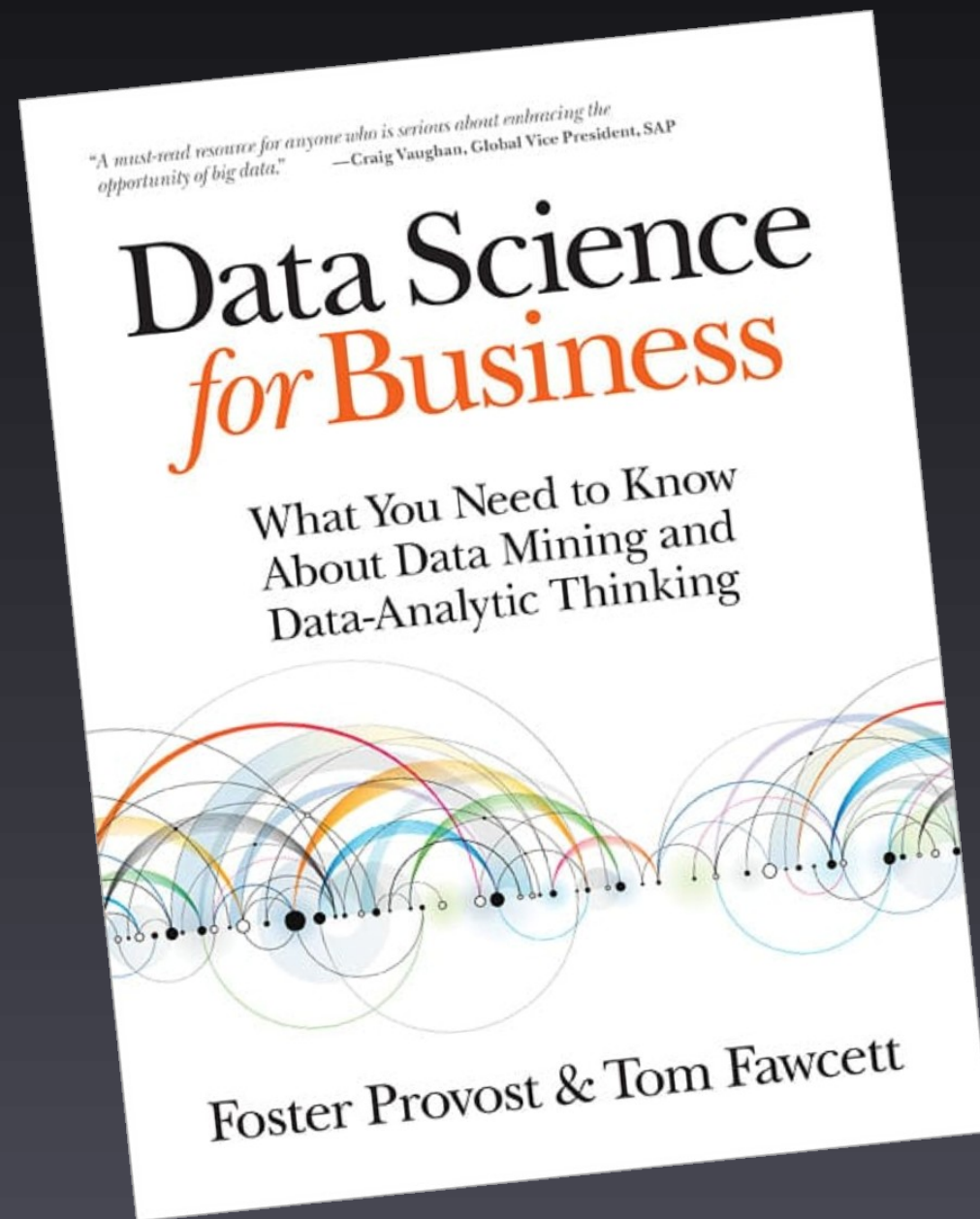
DLI Accelerated Data Science Teaching Kit

# Lecture 1.2 - What is Data Science?



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

# What is Data Science?



“At a high level, data science is a set of **fundamental principles** that guide the extraction of **knowledge from data**.”

Principles can be statistical, computational, algorithmic, visual, etc.

# What is **Data Science**?

In more detail...

- Relatively young field
- The term often broadly used, due to its **hype**
- “Definition” (e.g., what it is, what it includes) often hotly debated; a **controversial** topic!  
([https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science))
  - Related to statistics, computer science, visualization, data mining, machine learning, human-computer interaction, big data, distributed computing, and more!

# What is **Data Science**?

Why so complex?

- Data science is **interdisciplinary**, due to its goal to aid discoveries, decision making, etc.
  - Applicable to **many domains** (e.g., sciences, finance, healthcare, etc.)
  - Often requires analysis of **large amount of data** (helpful to use scalable algorithms, distributed computing, etc.)
  - **Statistics** is a pillar of data science
  - **Users** may need to explore data, and understand and present analysis results (benefits from visualization, good user interface design, etc.)
  - And more!



DEEP  
LEARNING  
INSTITUTE



DLI Accelerated Data Science Teaching Kit

# Thank You