

# **Unit 1:**

# **What is data science?**

Lesson 3

Derek Ruths

# Overview of unit

- **What is data science?**
- What is data?
- Data collection and curation
- What is data analytics?
  - How does data analytics fit into data science?
- What is machine learning?
  - How does machine learning fit into data science?

# Lesson overview

## Objectives

- Review unit goals
- Define data science
- Define data science process

## Outline

- Overview of unit
- Definition of data science
- Data science project phases
- Example

# Definition of data science

Data science is about the **innovative** use of data to understand something.

Question → insight

✗ Building a classifier

✗ Making a visualization

✗ Collecting data

# Data science project phases



# Data science project phases

*new  
or  
unclear* →

One or more of the phases below should be “messy.” E.g.,

- Question hasn't been asked before (using data)
- Type of data is new
- Unclear how to annotate it
- Analysis statistics & visuals are unclear



*Just  
data/observation*

*tells us  
what the  
data is*

# Lesson wrap-up

## Takeaways

- Data science is about innovation and understanding
- Data science projects have a structure

## Up next

- What is data?