

Unit 1: What is data science?

Data analysis

Lesson 6

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

- Distinguish analysis from data preparation
- Understand why data science and analytics are not the same thing

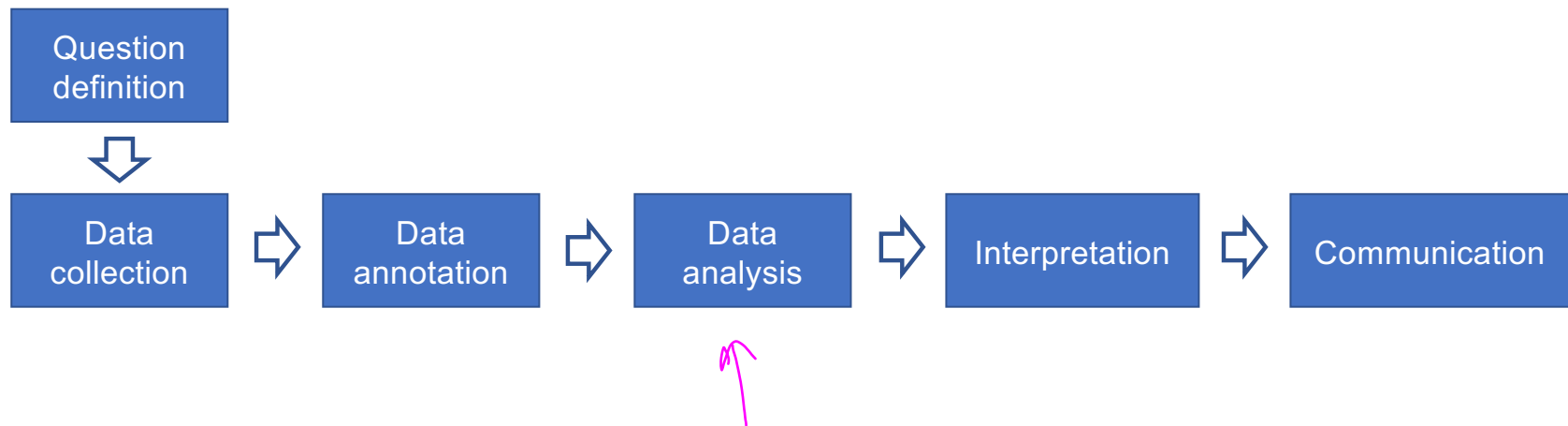
Outline

- Define data analysis
- Examples
- Data analysts vs. data scientists

Which basketball player is the highest scoring player this year?

Probably not
a data science
problem
(there's a lot of
clean data
already out there)

Data science project phases



What is data analysis?

The creation of statistical models and visualizations using existing methods and existing, (relatively) clean quantitative data.

Examples of data analytics

- analytics is on left*
- Market analysis using stock data
 - Bus route analysis using real-time tracking data
 - Assessing team performance using established KPIs
 - Identifying the genes most expressed in a particular experiment

Not data analysis alone

- Analyzing the gender-bias in Reddit memes
- Assessing pollution levels using satellite footage

Data analysts vs. scientists

Data analysts are focused on measuring things to gain a concrete new insight.

Data scientists are focused on measuring things that haven't been measured before.

new insights
new measurements

What happens when you want to measure something there isn't a measure for?

Lesson wrap-up

Takeaways

- Data analytics is a critical part of insight-driven work
- Data science can contribute new techniques in data analytics

Up next

- What is machine learning?