



# DS-UA 112

## Introduction to Data Science

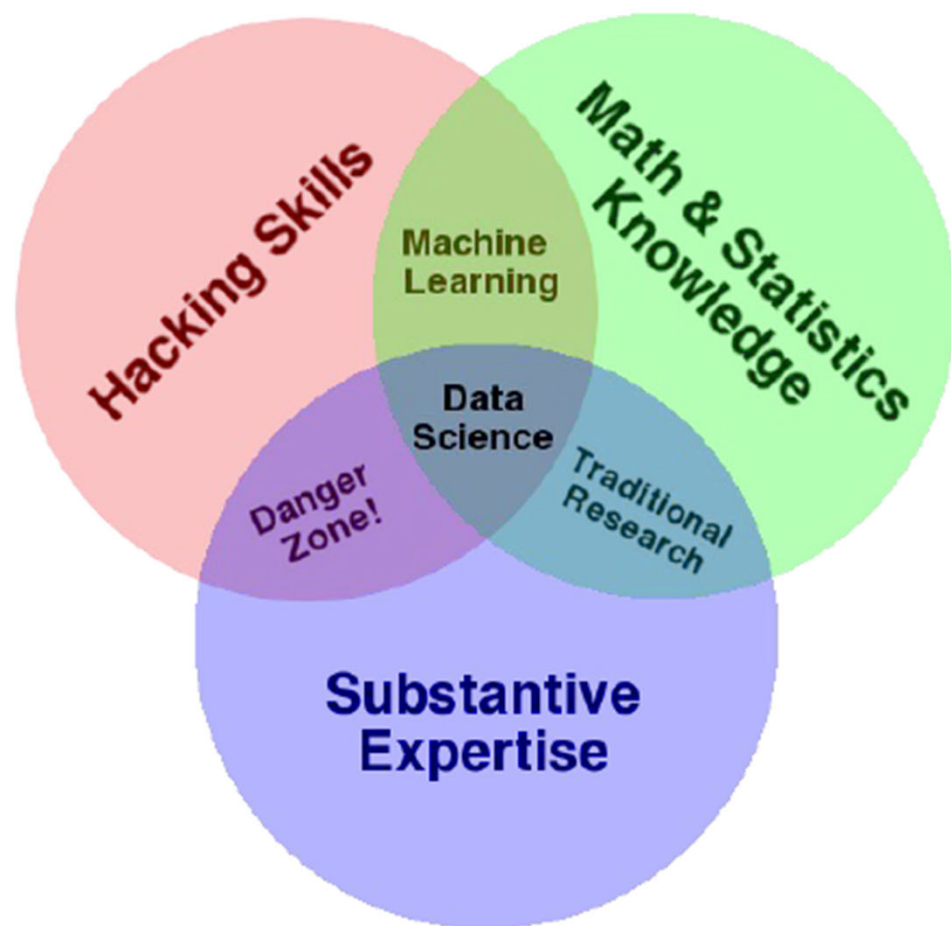
Lecture 1

# Agenda

- ▶ Overview
  - ▶ Data Science
  - ▶ DS-UA 112
- ▶ Lesson
- ▶ Demo

# What is Data Science?

- ▶ Drawing useful conclusions from data using computation
- ▶ **Exploration**
  - ▶ Identifying patterns in information
  - ▶ Uses charts and tables
- ▶ **Prediction**
  - ▶ Making informed guesses
  - ▶ Uses procedures to determine appropriate hypotheses
- ▶ **Inference**
  - ▶ Quantifying whether those patterns are reliable
  - ▶ Uses randomization

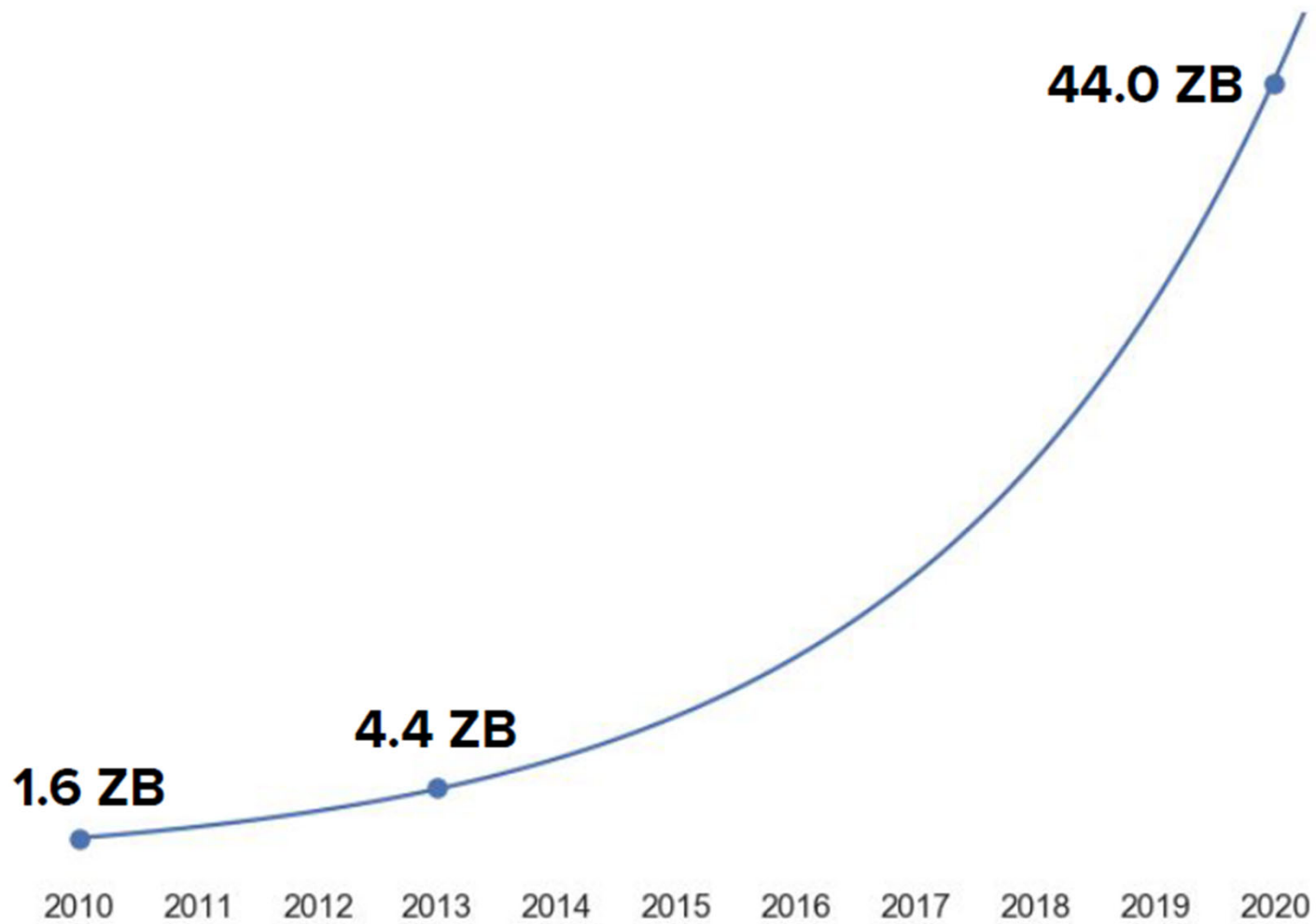


# Who uses Data Science?

- ▶ Academia
  - ▶ Event Horizon Telescope
- ▶ Industry
  - ▶ UPS
- ▶ Government
  - ▶ Contractors and Social Media

# Why Study Data Science?

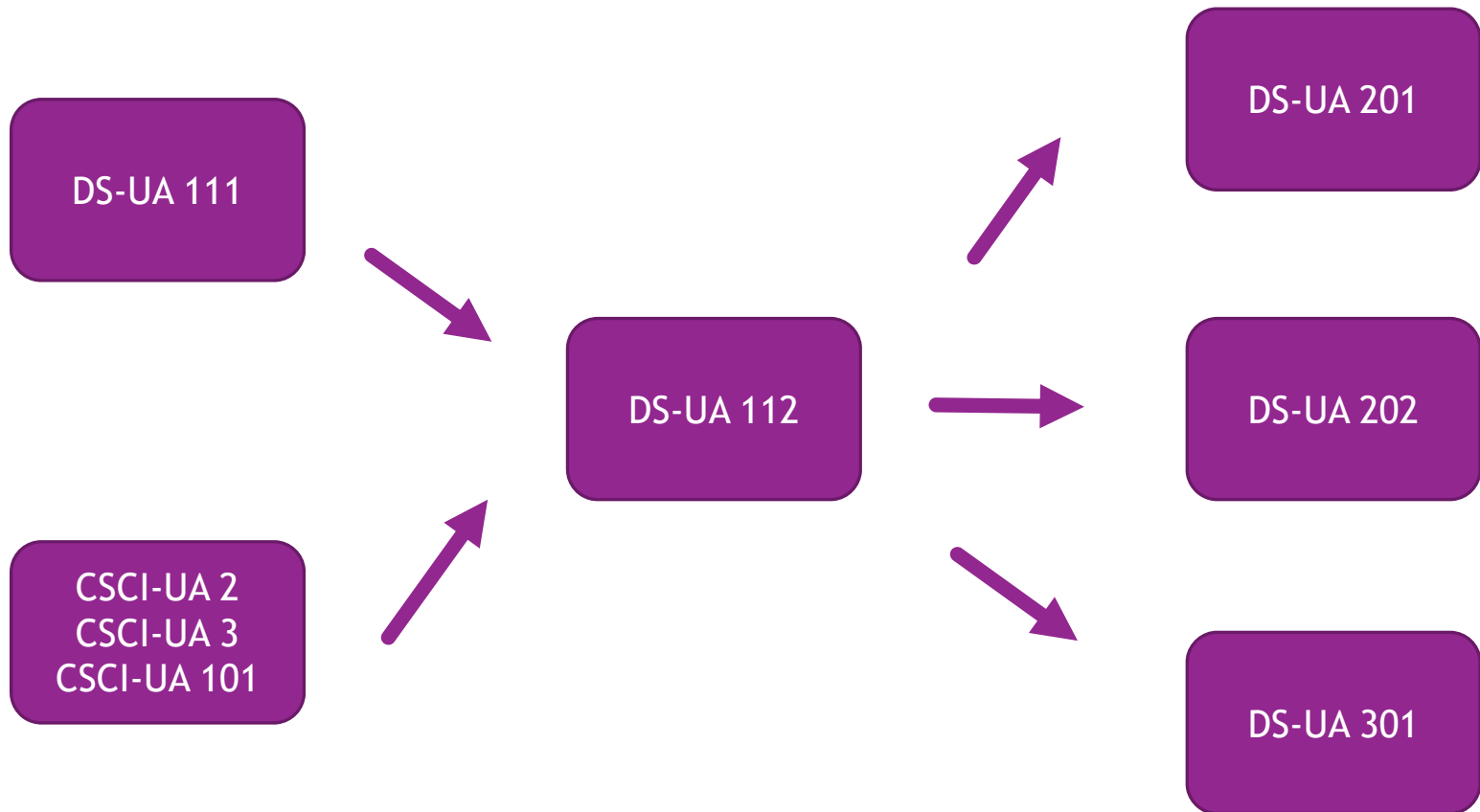
- ▶ Data is not just numbers, it includes
  - ▶ Text
  - ▶ Images
  - ▶ Videos
- ▶ Technology has made data ubiquitous
  - ▶ Storage
  - ▶ Communication
  - ▶ Digitization



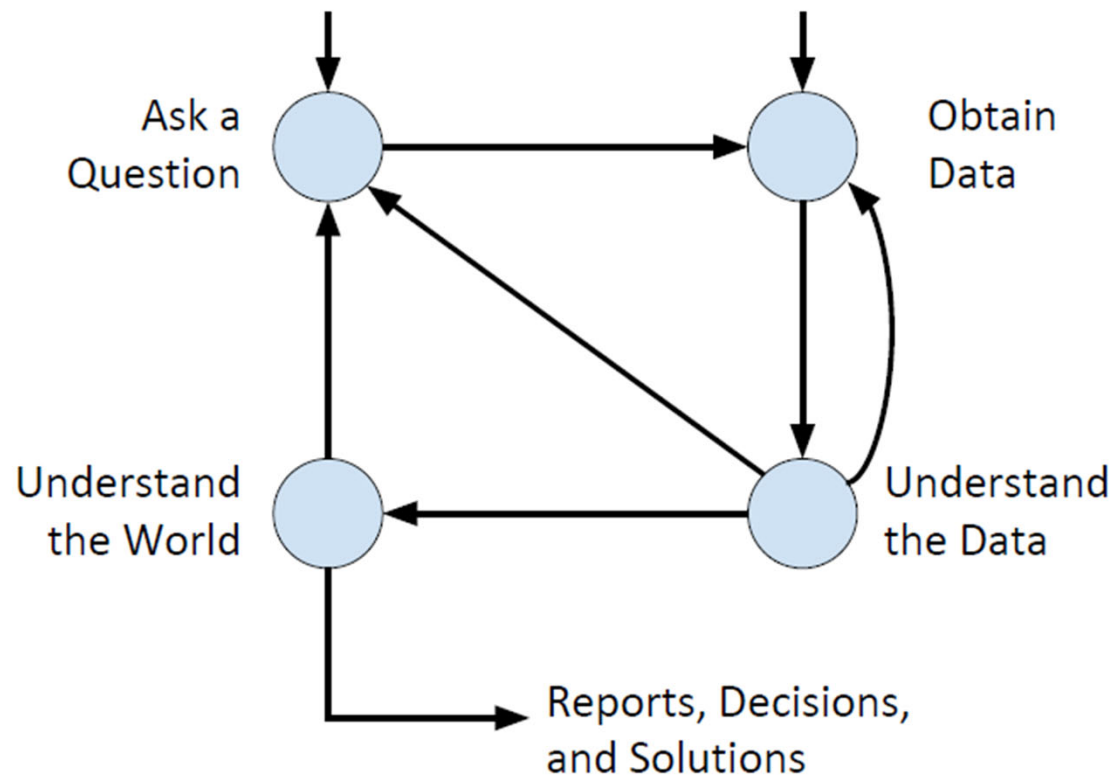
# DS-UA 112 Introduction to Data Science

- ▶ Interdisciplinary course for students in the sciences and humanities
- ▶ Part of the Major and Minor in Data Science
- ▶ Goals
  - ▶ Prepare
  - ▶ Empower
  - ▶ Enable



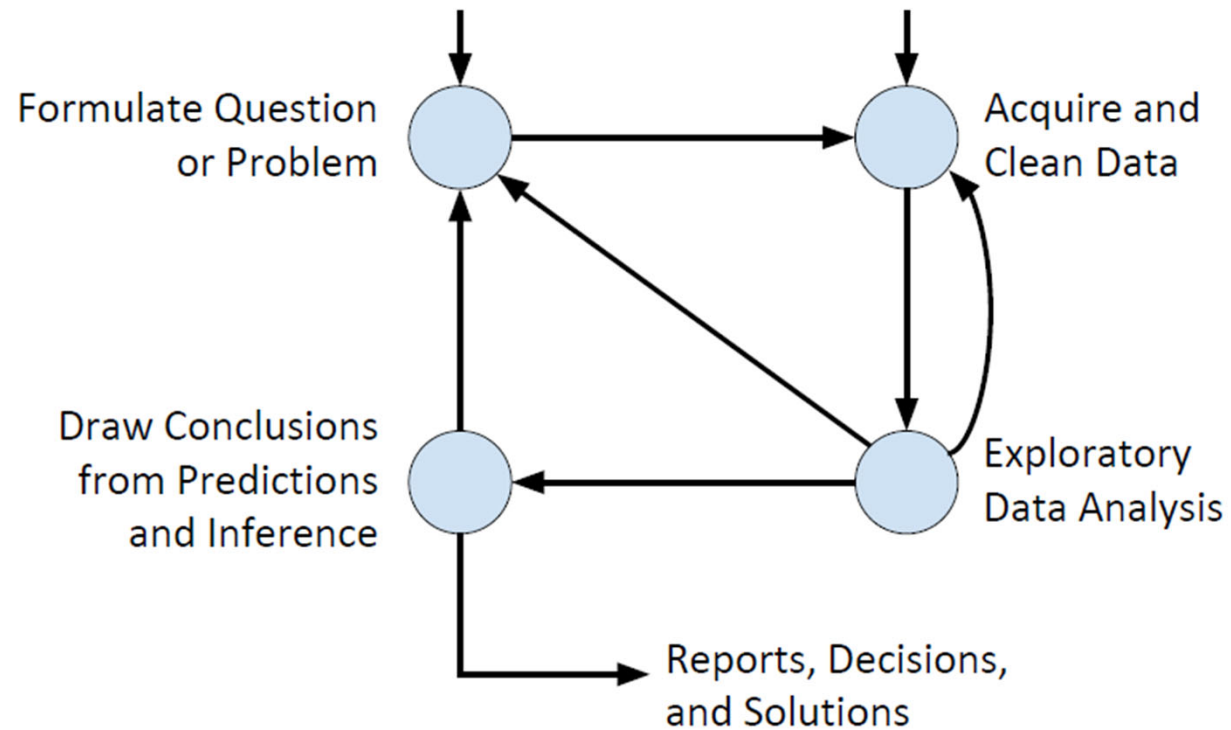


# Lesson



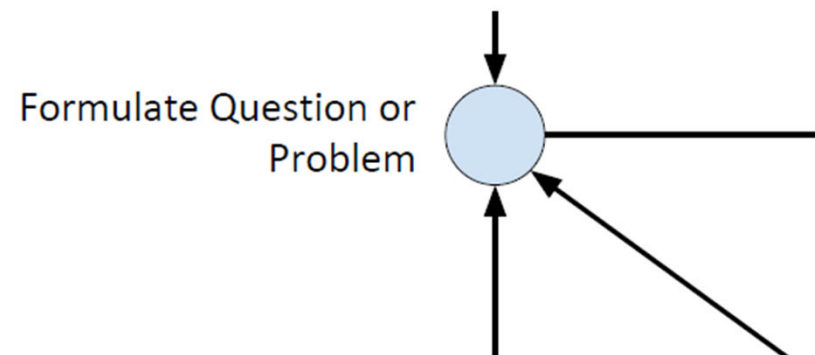
Note the two distinct entry points!

# Lesson



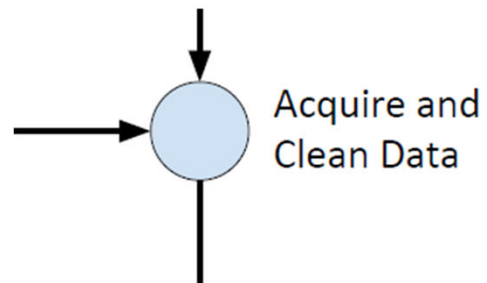
# Lesson

- What do we want to know?
- What problems are we trying to solve?
- What are the hypotheses we want to test?
- What are our metrics for success?

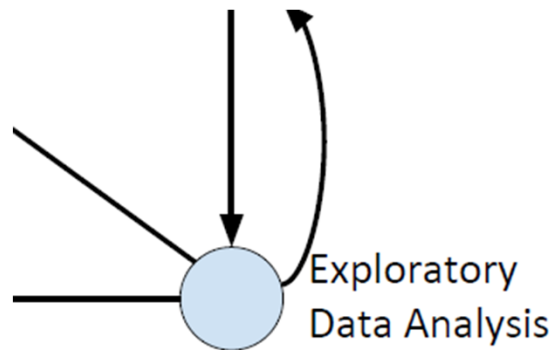


# Lesson

- What data do we have and what data do we need?
- How will we collect more data?
- How do we organize the data for analysis?



# Lesson



- Do we already have relevant data?
- What are the biases, anomalies, or other issues with our data?
- How do we transform the data to enable effective analysis?

# Demo

- ▶ Understanding information from novels
  - ▶ Step 1: Determine a question
  - ▶ Step 2: Gather Data
  - ▶ Step 3: Explore the Data
- ▶ Examples of coding in Python on JupyterHub
- ▶ Goal
  - ▶ How to summarize information?
  - ▶ How can text be data?

# Reminders

## ▶ Section

- ▶ Section 002 Tuesdays 3:30-4:20pm (60<sup>th</sup> 5<sup>th</sup> Avenue, Room C04)
- ▶ Section 003 Thursdays 3:30-4:20pm (60<sup>th</sup> 5<sup>th</sup> Avenue, Room 161)
- ▶ Section 005 11:00-11:50pm (60<sup>th</sup> 5<sup>th</sup> Avenue, Room C04)
- ▶ Section 006 12:55-1:45pm (60<sup>th</sup> 5<sup>th</sup> Avenue, Room C04)

## ▶ Syllabus

- ▶ Please review the course policies about assignments and grading.

## ▶ Surveys

- ▶ Please complete Survey 1
- ▶ Students in Section 002 should complete the survey about rescheduling