DS-UA 112
Introduction to Data Science

Lecture 2

Agenda

- Review
- Lesson
- Demo

What is Data Science?

- Drawing useful conclusions from data using computation
  - **Exploration**
  - **Prediction**
  - **Inference**

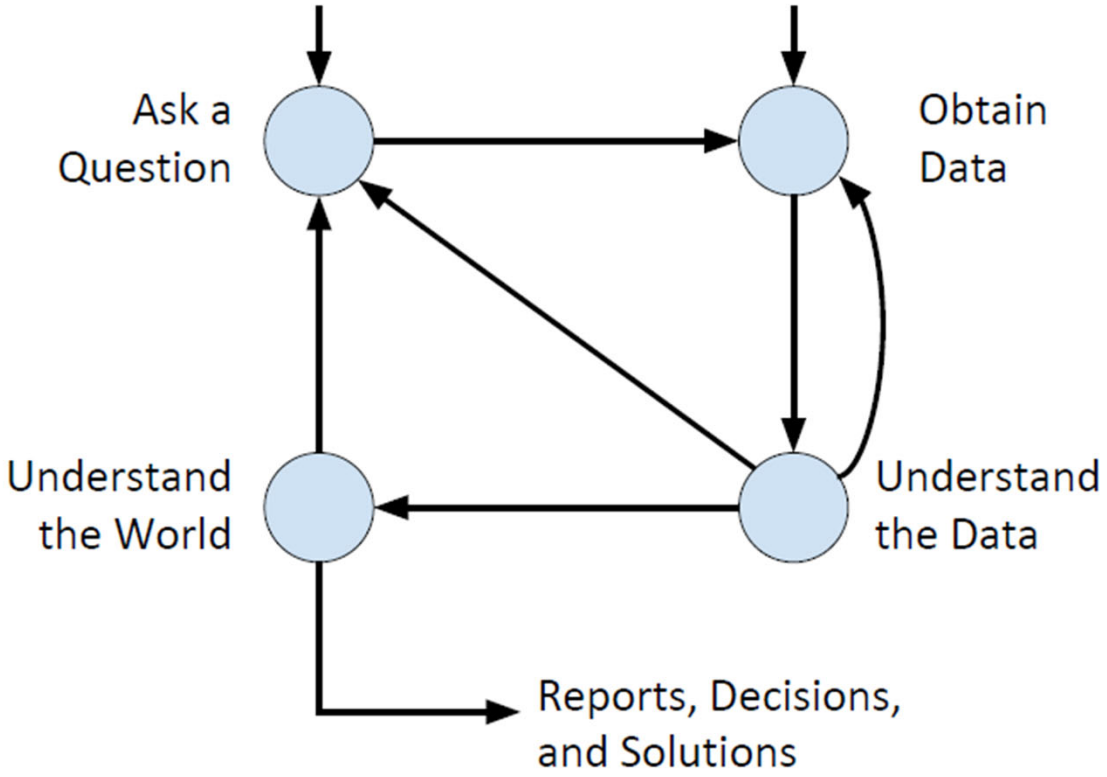
Who uses Data Science?

- Academia
- Industry
- Government

Why Study Data Science?

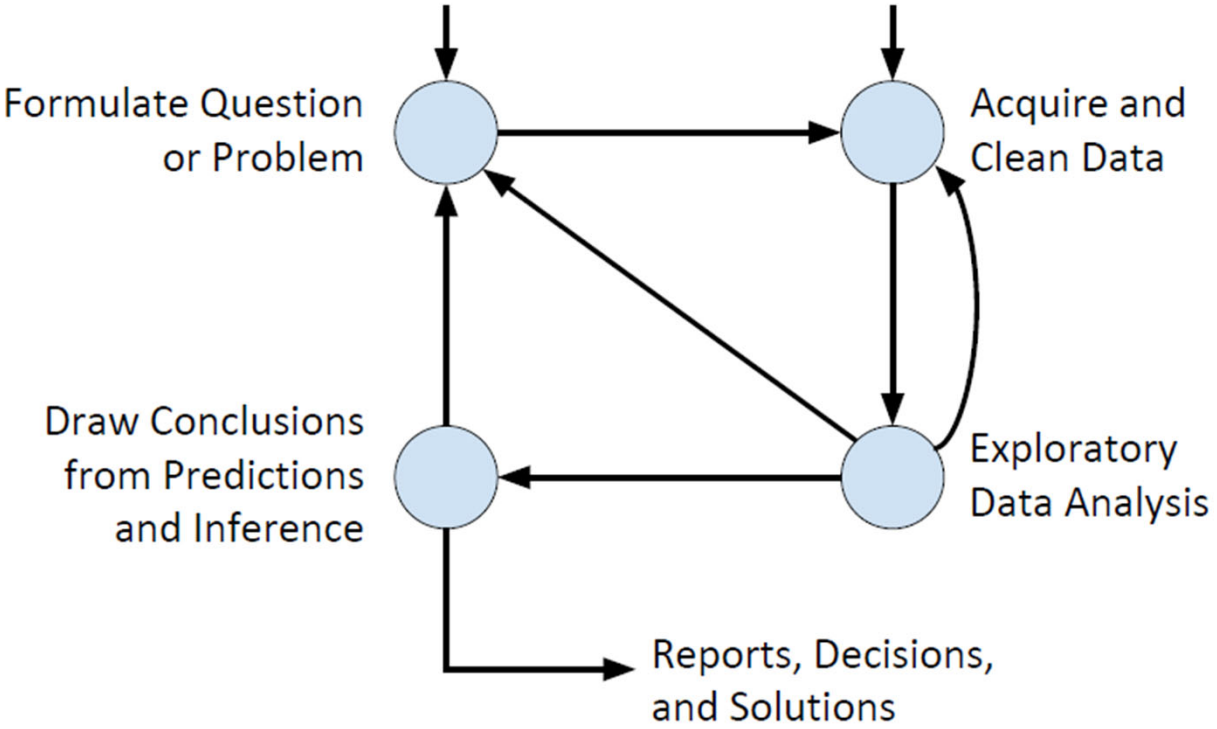
- Data is not just numbers, it includes
  - Text
  - Images
  - Videos
- Technology has made data ubiquitous
  - Communication
  - Storage
  - Digitization

Steps



Note the two distinct entry points!

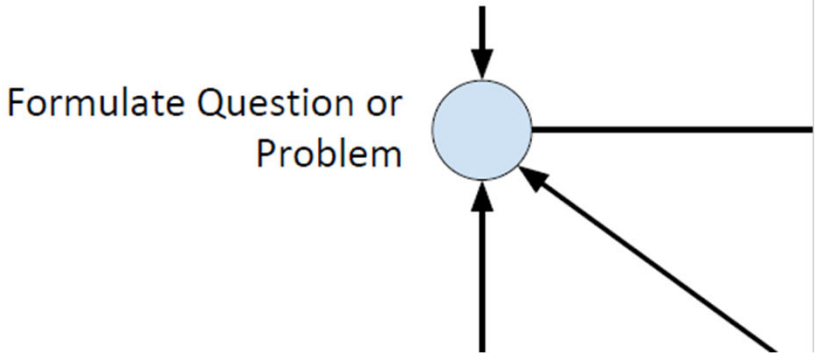
Steps



Note the two distinct entry points!

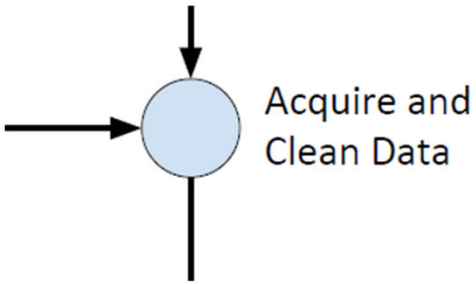
Step 1

- 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?

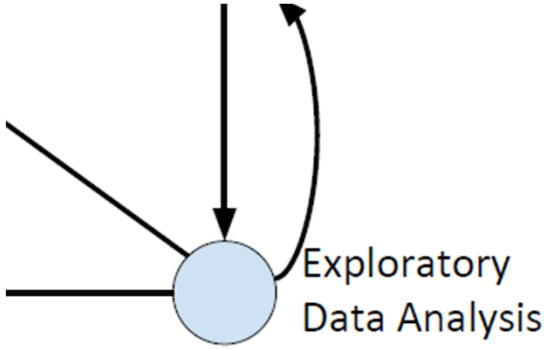


Step 2

- 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?

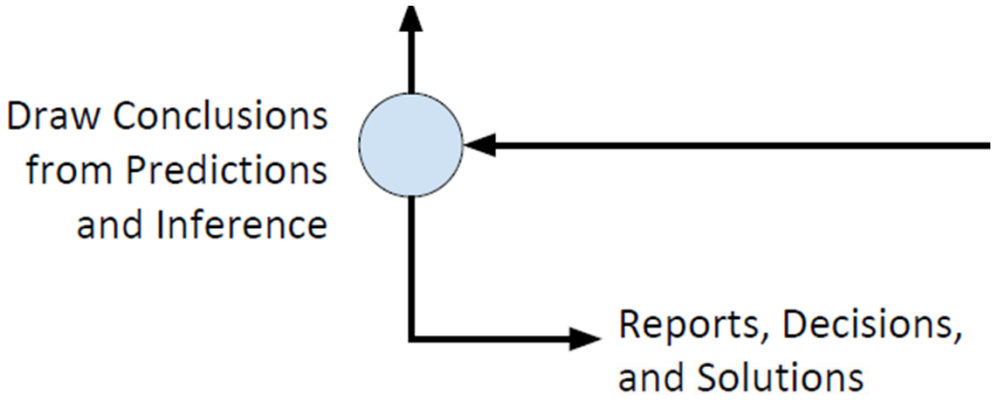


Step 3



- 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?

Step 4



- What does the data say about the world?
- Does it answer our questions or accurately solve the problem?
- How robust are our conclusions and can we trust the predictions?

Demo

- Programming
  - Learning through examples
  - Focus on applications
  - Many tools for research

Lesson

- Determining the Chance of Different Outcomes
  - Addition Rule
  - Multiplication Rule
  - Trials Rule

- Examples
  - Dice
  - Marbles
  - Coins

When an Event Happens?

$$0 \leq P(\text{an event happens}) \leq 1$$

When an Event Happens?

$$P(\text{an event doesn't happen}) = 1 - P(\text{an event happens})$$

When Events Have the Same Chances?

$$\frac{\text{Number of Even Faces}}{\text{Number of Odd Faces}} = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}}$$

When Events Have the Same Chances?

$$P(\text{event}) = \frac{\#\text{ of outcomes in event}}{\#\text{ of all possible outcomes}}$$
$$= \text{proportion of outcomes in event}$$

When Two Events Occur in Order?

$$P(\text{green first, then red}) = \frac{\#\{\text{GR}\}}{\#\{\text{RB, BR, RG, GR, BG, GB}\}} = \frac{1}{6}$$

When Two Events Occur In Order?

$$P(\text{two events both happen}) = P(\text{one event event happens}) \times P(\text{the other event happens, given that the first one happend})$$

When Two Events Occur In Order?

$$P(\text{two events both happen}) = P(\text{one event event happens}) \times P(\text{the other event happens} \mid \text{the first one happend})$$

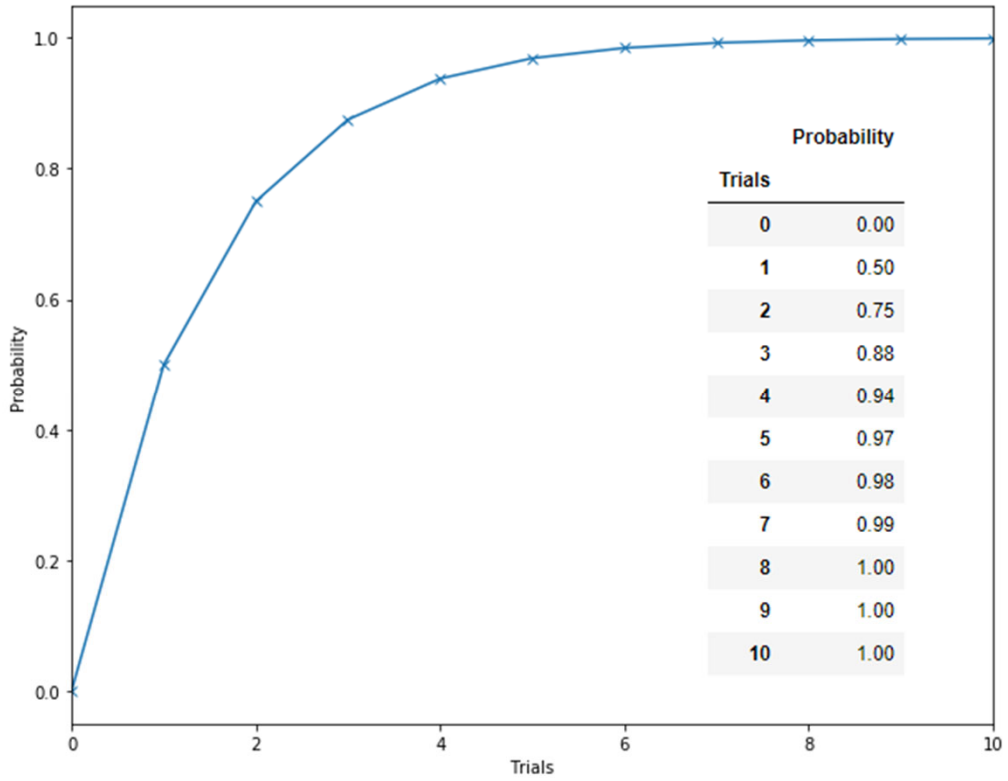
Independence

$$P(\text{the other event happens} \mid \text{the first one happend}) = P(\text{the other event happens})$$

When Multiple Events Occur in Order?

$$P(\text{at least one head in 17 tosses}) = 1 - P(\text{all 17 are tails}) = 1 - \left(\frac{1}{2}\right)^{17}$$

When Multiple Events Occur in Order?



Trials	Probability
0	0.00
1	0.50
2	0.75
3	0.88
4	0.94
5	0.97
6	0.98
7	0.99
8	1.00
9	1.00
10	1.00

When Two Events Occur?

$$P(\text{one green and one red}) = P(\text{GR}) + P(\text{RG}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

When Two Events Occur?

$$P(\text{an event happens}) = P(\text{first way it can happen}) + P(\text{second way it can happen})$$

Demo

- Goals
  - How can frequencies indicate the chance of events?
  - How can probability allow us to incorporate prior knowledge about a problem?

Reminders

- Section
  - Section 002 Tuesdays 3:30-4:20pm (60th 5th Avenue, Room C04)
  - Section 003 Thursdays 3:30-4:20pm (60th 5th Avenue, Room 161)
  - Section 005 11:00-11:50pm (60th 5th Avenue, Room C04)
  - Section 006 12:55-1:45pm (60th 5th Avenue, Room C04)
- Homework 1
  - Please access under Resources
- Surveys
  - Students in Section 002 should complete the follow-up survey about rescheduling.