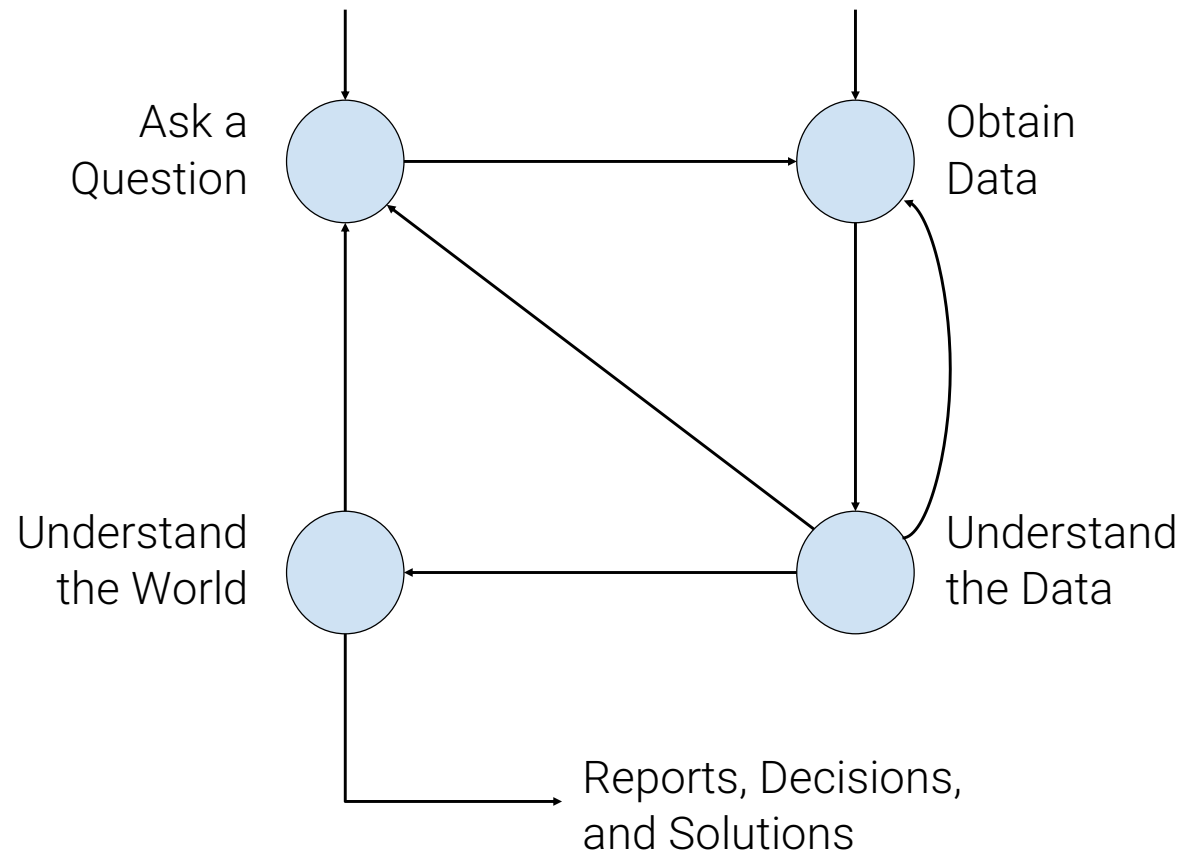


Data Science Lifecycle

The data science lifecycle is a **high-level description** of the data science workflow.

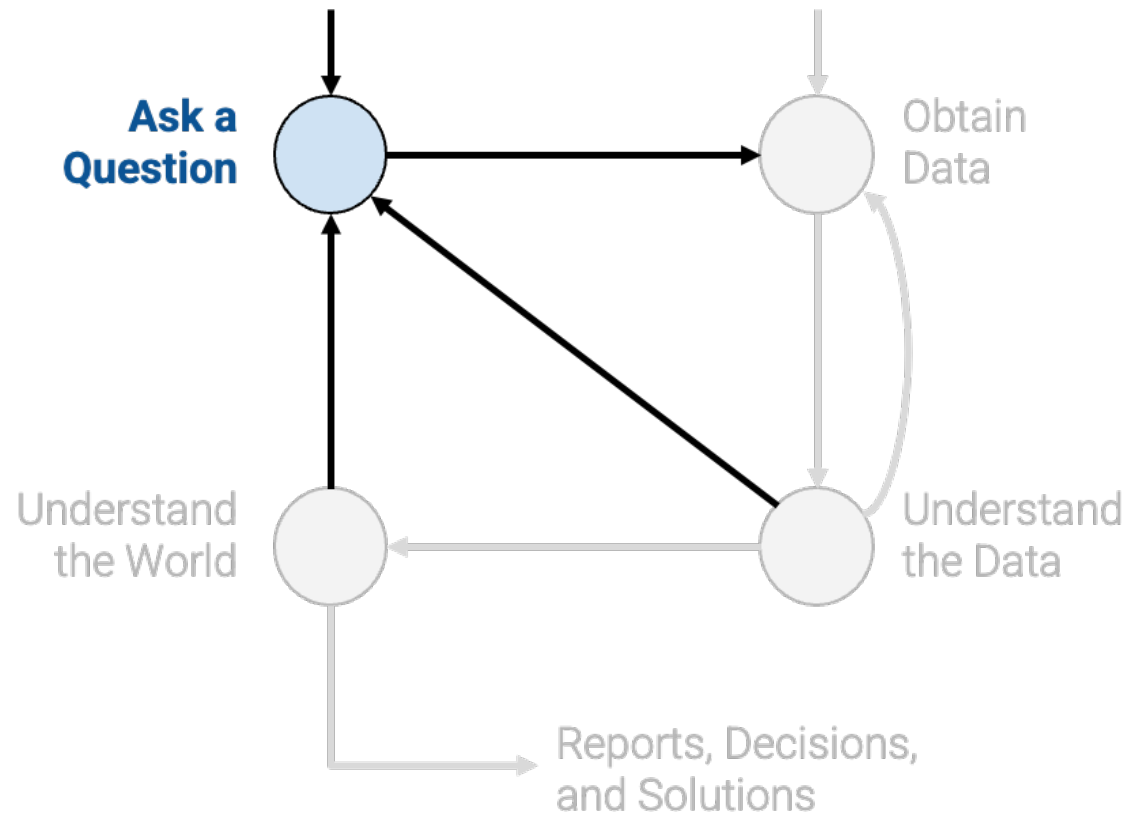
Note the two distinct entry points!



Data Science

Data Science Life Cycle

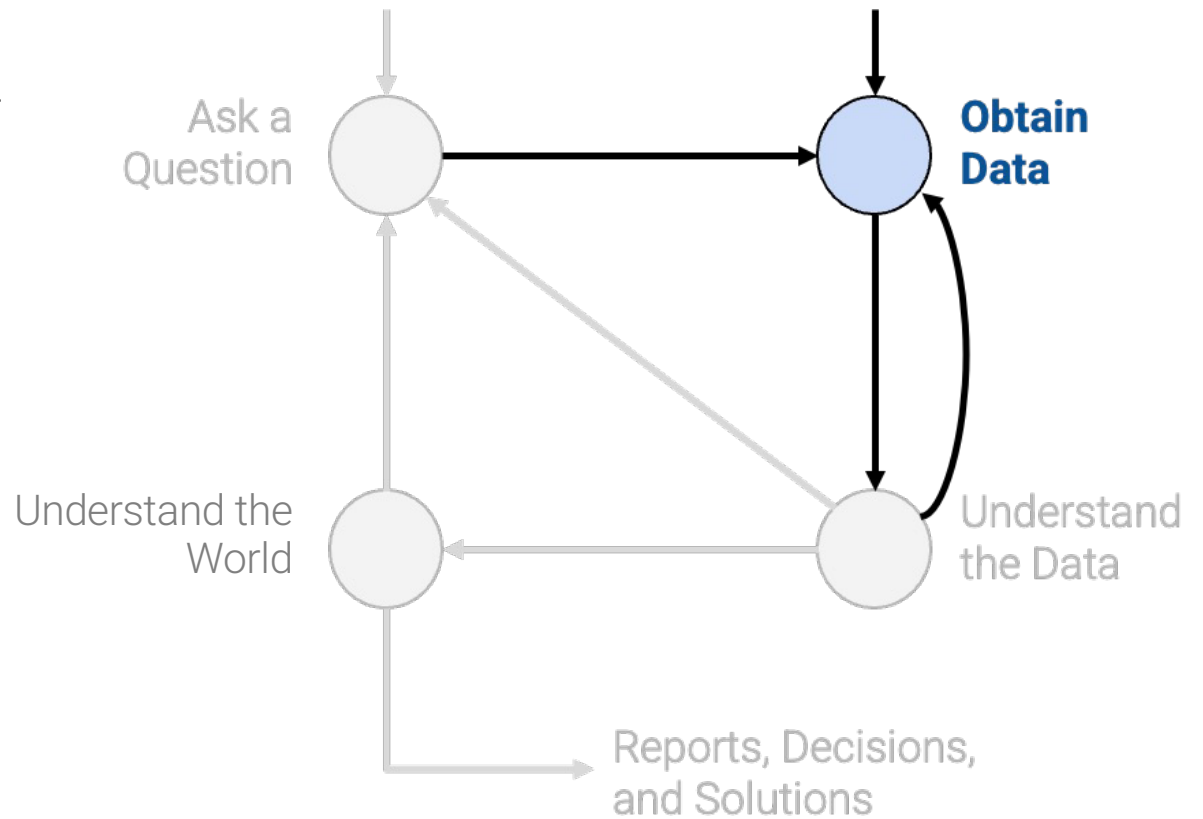
- What do we want to know?
- What problems are we trying to solve?
- What hypotheses do we want to test?
- What are our metrics for success?



Data Science

Data Science Life Cycle

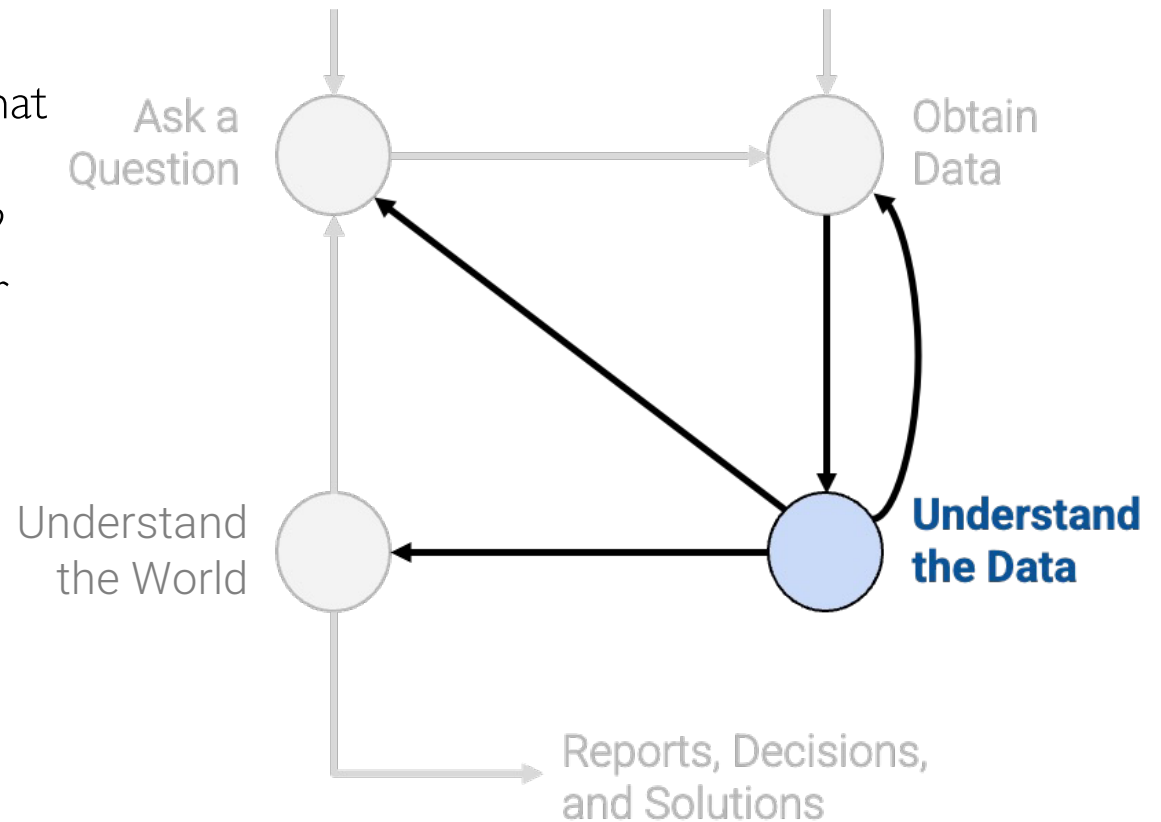
- What data do we have and what data do we need?
- How will we sample more data?
- Is our data representative of the population we want to study?



Data Science

Data Science Life Cycle

- How is our data organized and what does it contain?
- Do we already have relevant data?
- What are the biases, anomalies, or other issues with the data?
- How do we transform the data to enable effective analysis?



Data Science

Data Science Life Cycle

- 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?

