Introduction to data science

Experiential Data science for Undergraduate Cross-disciplinary Education

Dr. Kim Dill-McFarland, U. of British Columbia

Introduction to data science

Learning objectives

- Define data science
- List common tools used in data science

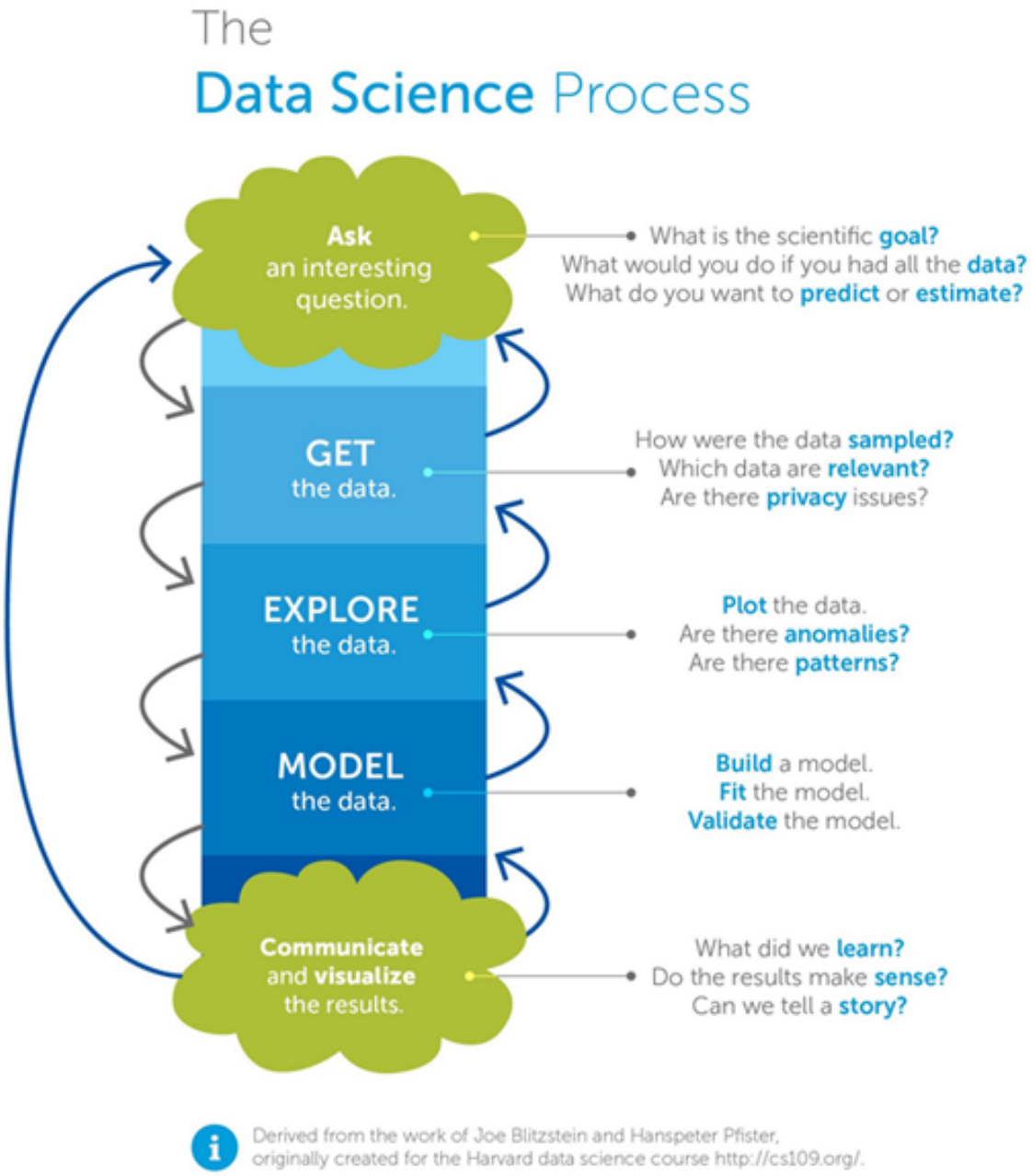
What is 'data science'?

Wikipedia defines 'data science' as

- "an *interdisciplinary* field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from *data* in various forms"
- "a concept to unify *statistics, data analysis, machine learning and their related methods* in order to understand and analyze actual phenomena with data."

So, data science comprises many types of **data** and whatever **tools** you need to **extract meaning** from that data. The process is both cyclic and iterative, similar to the scientific method. The skills you learn here are applicable to a wide range of disciplines and data types. Thus, data science can serve as the foundational to many current and future careers.

Figure 1: (c)Forbes 2012



Fundamental tools in data science

Common tools in data science that are:

- command line (*e.g.* Unix) to remotely access computational resources
- scripting languages (*e.g.* R, python, MATLAB) to manipulation and plot data
- statistics to extract significant results

To learn more about these tools and their applications, please continue to explore the EDUCE curriculum such as the module(s) contained here. There are also many other tools like deep learning, machine learning, and discipline-specific command line programs that are not covered here! So, be sure to explore more with online courses on platforms like:

- The Carpentries
- edX
- Codecademy
- swirl
- Udemy
