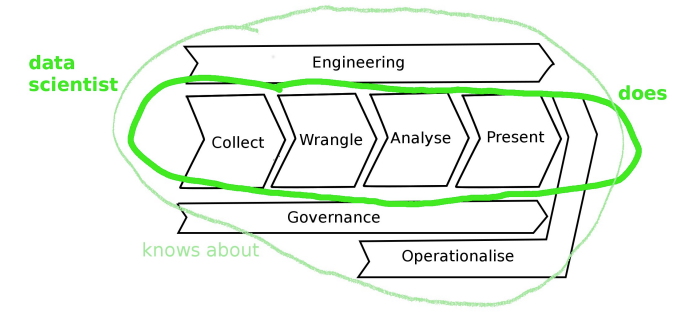
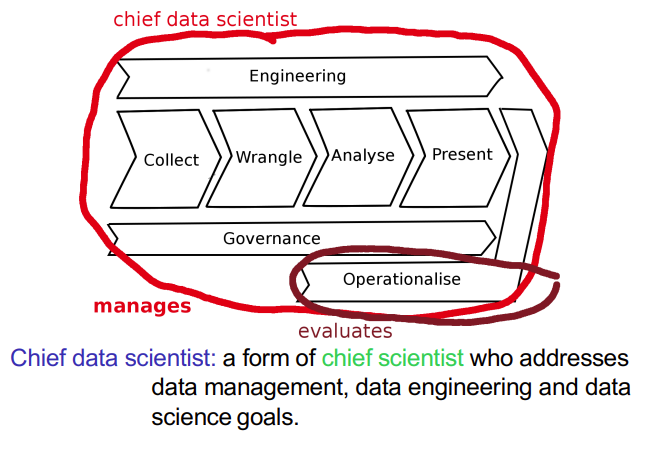
**Data Science** is the extraction of knowledge from data, which is a continuation of the field data mining and predictive analytics.

**Big data** is a broad term for data sets so large or complex that traditional data processing applications are inadequate.

**Machine Learning** is concerned with the development of algorithms and techniques that allow computers to learn.

* concerned with building programs that can learn, oftentimes with computational output
* underlying theory is statistics
* is useful when:
  + Human expertise is not available
  + Many solutions need to be adapted automatically
  + Humans are expensive for the work
  + Situation changes overtime
  + large amounts of data

Data Science Process

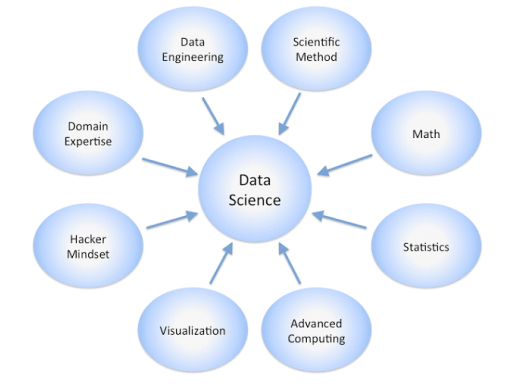
 

Related:

* Data Engineering
  + building scalable systems for storage, processing data
  + Hadoop
  + databases, distributed processing, datalakes, cloud computing, GPUs, wrangling
* Data Analysis
  + performing analysis and understanding results
  + R and Microsoft Azure Machine Learning
  + machine learning, computational statistics, visualisation
* Data Management
  + managing data through its lifecycle
  + ethics, privacy, curation, backup, governance

Standard Value Chain:

* Collection: getting the data
* Engineering: storage and computational resources across full lifecycle
* Governance: overall management of data across full lifecycle
* Wrangling: data preprocessing, cleaning
* Analysis: discovery (learning, visualisation, etc.)
* Presentation: arguing the case that the results are significant and useful
* Operationalisation: putting the results to work, so as to gain benefits or value



Data Science Venn Diagram

