Big Data and More Vs

- The extra Vs needed for Big Data, are:
  - Veracity
  - Value
  - Variability
  - Visualization
  - Validity
  - Vulnerability
  - Volatility

Veracity

- How clean of data
  - Preprocessed, clean data is required for successful analysis.
  - Especially when the volume is massive

Value

- How valuable of data
  - Business value drives the research and practice.
  - Domain knowledge plays the crucial role of determining the value of data.

Variability

- How consistent of data

  - If data is acquired from multiple source, or from a long time span, the format, or the meaning, might be inconsistent.

  - Data inconsistency is a common scenario.

    - Different scale for the same feature: inch vs cm

    - Different format: 1000 vs 1,000 or $1e^3$

Visualization

- The visual of data
  - Millions numbers will overwhelm people, but a well-prepared simple plot will not.
  - Choosing the right tool, for the right purpose, is important.
  - Communication is getting more and more crucial, for highly specialized fields nowadays.

Validity

- How valid of data

  - If the source of data is not reliable or not valid, we can not find valid results.

  - Is the data correct for the purpose

Challenges of Big Data

- Managing large volumes

- Dealing with data integration

- Cleaning massive data

- Scaling with efficiency and effectiveness

- Evaluating and selecting technologies

- Ensuring data context

- Securing data