

Chapter 1

Overview

Overview

- **The global picture**
- **What is data visualization/visual analytics?**
- **Applications**

The global picture

- **Sci Vis, information vis**
- **The quest of data analysis/Big data analysis**
- **The 5 Vs**
 - **Volume**
 - **Refers to the vast amounts of data generated every second**
 - **On Facebook alone we send 10 billion messages per day, click the "like" button 4.5 billion times and upload 350 million new pictures each and every day.**
 - **Velocity**
 - **Refers to the speed at which new data is generated and the speed at which data moves around**
 - **Allows us now to analyse the data while it is being generated, without ever putting it into databases**

The global picture

- **The 5 Vs**

- **Variety**

- **Refers to the different types of data we can now use**
 - **In the past we focused on structured data that neatly fits into tables or relational databases, such as financial data (e.g. sales by product or region).**
 - **In fact, 80% of the world's data is now unstructured, and therefore can't easily be put into tables (think of photos, video sequences or social media updates).**

- **Veracity**

- **Refers to the messiness or trustworthiness of the data. With many forms of big data, quality and accuracy are less controllable (just think of Twitter posts with hash tags, abbreviations, typos and colloquial speech as well as the reliability and accuracy of content)**

The global picture

- **The 5 Vs**

- **Value**

- **It is all well and good having access to big data but unless we can turn it into value it is useless**



<https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>

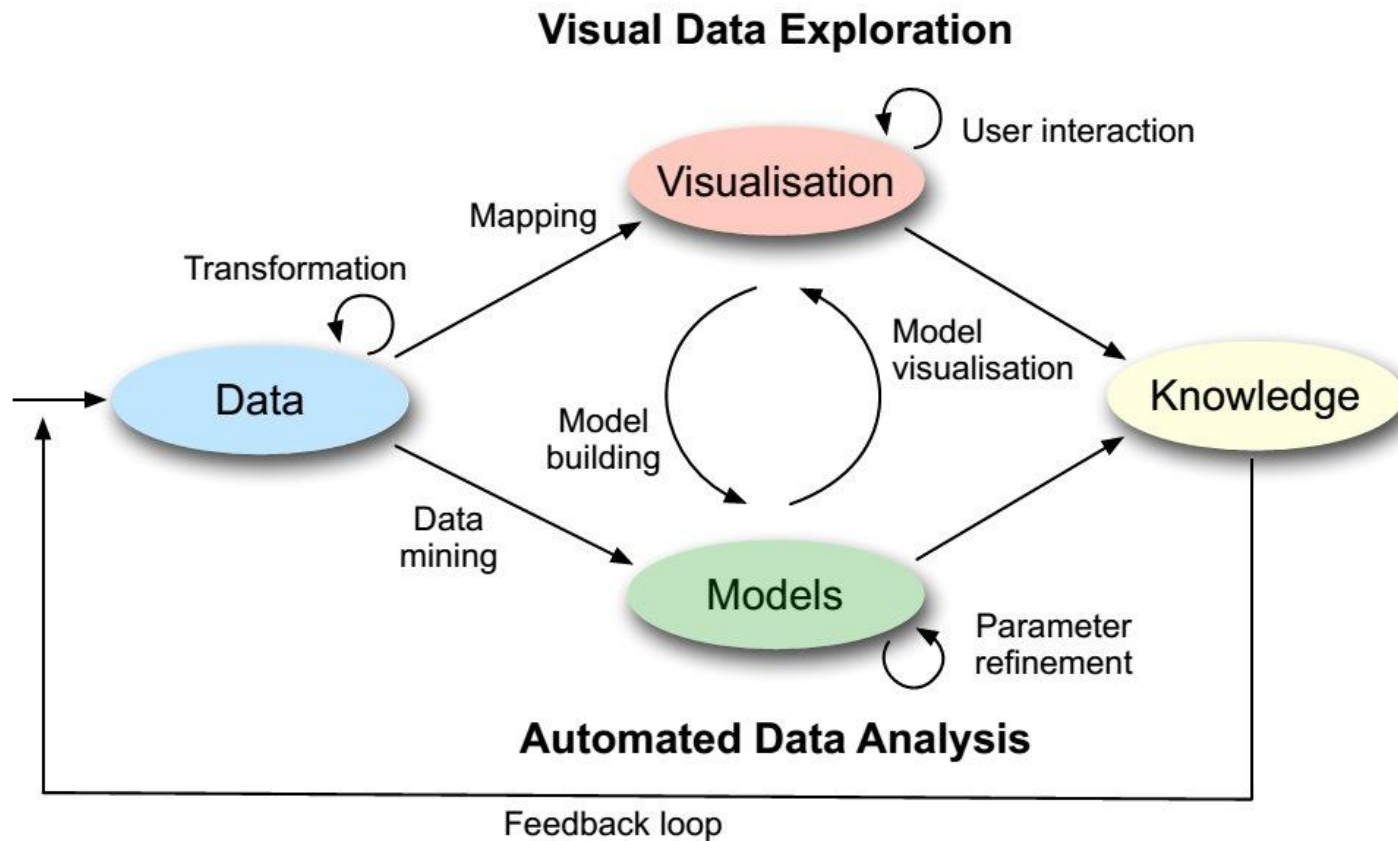
Automated data analysis

- **Data mining, machine learning have proven their usefulness, they still face significant challenges**
 - **Algorithm scalabilities**
 - **Increasing data dimensions**
 - **Data heterogeneity**
 - **Others**
 - **May not be perfect under all analysis scenarios, requiring user's knowledge to iteratively refine the methods**
 - **For complex, interesting patterns discovered, usually difficult to understand and interpret the findings intuitively and meaningfully.**

Visual analytics

- **Information visualization mantra**
 - **Overview first, zoom/filter, details on demand**
- **Visual analytics mantra**
 - **Combines automated analysis with interactive visualizations for effective understanding, reasoning and decision making for a very large and complex dataset.**
 - **Visual analytics mantra**
 - **Analyze first, show the important, zoom/filter, analyze further, details on demand**

Visual analytics process



Visual analytics process

Key steps

- **Data transformation**
 - **Filtering and sampling for further exploration**
- **Visual mapping/layout**
- **Model-based analysis**
- **User interaction**

Visual analytics process

Application categories

- **Space and time**
 - **Geospatial, temporal, spatial-temporal**
- **Multivariate (high-dimensional)**
- **Text**
- **Graph and network**
- **Other applications**
 - **Videos, images**