Data Manipulation

Evangelos Pournaras, Izabela Moise, Dirk Helbing
Data Manipulation

The process of data **transformation, formatting & structuring**.

Examples

updating, adding/removing, sorting, selection, merging, shifting, aggregation, etc.

Tip

*In Data Science, data come with collection & science starts with manipulation!*
A "Dirty Job"
Do you really need it?

- Big Data and Internet of Things result in large amount of unstructured data.
- **New data collection opportunities** require advanced data manipulation techniques.
- Involvement with data manipulation becomes more likely & required nowadays.
Data Format & Manipulation

Select a data manipulation approach based on how data are stored and managed:

1. Text files
2. Databases
3. Big Data
How to manipulate data

Most programming languages and several software tools can manipulate data:

- Java, Python, C, C++, etc.
- Matlab, R, Excel, etc.

Criteria for selection:

- Ease of use
- Library support
- Portability
- Performance
- Data format
What is next?

- Data manipulation with AWK