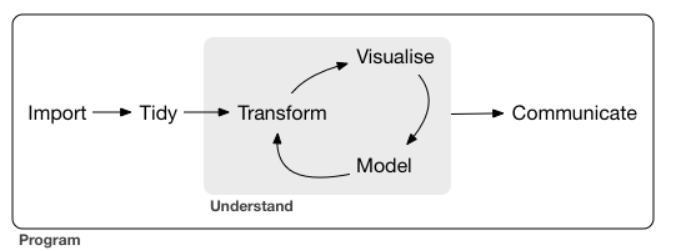
DATA SCIENCE

È una disciplina che si occupa di interpretare i dati come contenuto informativo.



Questo è un semplice schema logico del processo di analisi di un dato.

# IMPORT

* Importare i dati su R
* I dati possono essere di qualsiasi formato: [varietà](#_TRANSFORM)

# WRANGLING = TIDY + TRANSFORM

## TIDY (=ordinare)

* Per ordinare i **dati** è opportuno **individuarne** la **semantica**
* **Identificare** una **struttura** **fisica** che possa **rispettare** la **semantica** definita
* Creando un **dataset** ben **dettagliato** è possibile **comprendere** al meglio il **significato** dei **dati**, per **poterli** ulteriormente **modificare**.
* **DATASET = TABELLA riga x colonna**
  + COLONNA = **variabile**/**proprietà** della **classe** di **oggetti** da analizzare
  + RIGA = **elemento** **appartenente** alla **classe** definita.
    - **Ogni** **elemento** avrà un valore **associato** a ogni **variabile**

Questo dataset riporta i dati di ogni macchina personale, con relative proprietà.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **MODELLO** | **PROPRIETARIO** | **TARGA** | **ANNO IMMAT.** | **CILINDRATA** | **POTENZA** |
| FIAT PANDA | IO | AB111FE | 2002 | 1200CC | 54Kw |
| HYUNDAI I10 | IO | ER334BE | 2008 | 1100CC | 45Kw |
| … | … | … | … | … | … |
| … | … | … | … | … | … |
| IX35 | IO | BT367YT | 2004 | 1700CC | 80kW |

## TRANSFORM

* Osservare e capire il contenuto informativo globale
* Creare nuove variabili derivate da quelle già definite
* Stilare un sommario con i dati rilevati dopo le modifiche

# PRESENTAZIONE

## VISUALIZE

* **Rappresentare** i dati in formato **grafico**, **comprensibile** al **meglio** dagli **umani**
* Una **completa** **visualizzazione** permette di **identificare** **nuovi** aspetti o **caratteristiche** non captate durante il wrangling
* **Oppure** permette di **trovare** degli **errori** di **interpretazione** dei dati **precedenti**
* **NON** rappresentano fedelmente il contenuto informativo, a causa delle approssimazioni grafiche

## MODEL

* I modelli sono complementari per ottenere una completa presentazione
* Definisce un modello grafico di ridotte dimensioni che permette di riassumere completamente il significato globale del dataset.
* Sono degli strumenti tipicamente matematici, quindi sono in grado di rappresentare fedelmente le varie dimensioni globali

# COMUNICAZIONE FINALE

Dopo **aver** **ripetuto** il **ciclo** **wrangling-visualize** è opportuno **comunicare** i **dati** ottenuti. Affinché si ottenga una completa comprensione globale è opportuno **definire** un **filo** **logico** che **permette** a un **estraneo** di **percepire** correttamente **l’informazione**.