## **Interviewing Your Data: Getting Data**

Miklós Koren





# Data vs story

## Data vs document



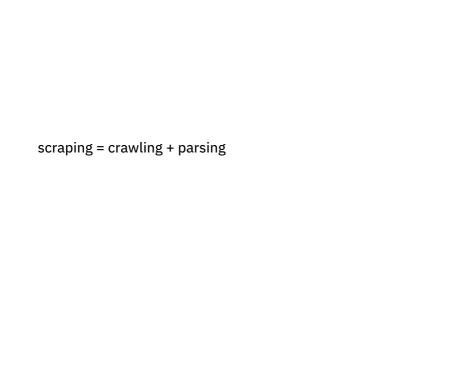








# Scraping



## Four steps of a scraping project

- 1 Recon
- 2 Crawl
- 3 Parse
- 4 Store

#### Recon

- 1 Locate the interesting documents and tables
- 2 Note the structure of URLs and tables
- Explore robots.txt and terms of use
- 4 Explore robot protection

### Exercise