

# Interviewing Your Data: Getting Data

Miklós Koren



What is data?

# Data vs story

# Data vs document

# Tidy data

Where to find data?

How to get it?

# HTTP and HTML



API, CSV, XML and JSON

# Scraping

scraping = crawling + parsing

# 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
- 3 Explore robots.txt and terms of use
- 4 Explore robot protection

## Exercise