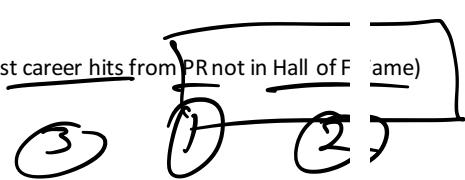
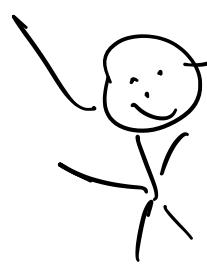


Lesson Plan 2/15

Wednesday, February 14, 2018 9:34 PM

- Admin
 - o HW1 Graded
- HDSC:
 - o Another podcast I like (for those with an entrepreneurial bent): <https://a16z.com/2018/02/13/data-insights-customer-experience-summit/>
- HW2 (due 2/20)
 - o Description
 - o Questions
 - o Example SQL Query (Find 10 players with most career hits from PR not in Hall of Fame)

- Structured data import
 - o `read_csv`
 - o Parsing "challenge.csv"
- Scraping HTML
 - o `html_table`
 - o Low-level scraping: The billboard challenge

Q) Master innerjoin HOF on playerID ?
players from PR in the HOF



Master left join HOF on playerID
where HOF.playerID is NULL



Rows
Entities
or Relationships / Columns (Attributes)

Data Wrangling & Cleaning

Raw



Tidy Data

CSV
XLSX
database
HTML

Normal
Form

||

①

each row corresponds to
a single entity (or relationship)

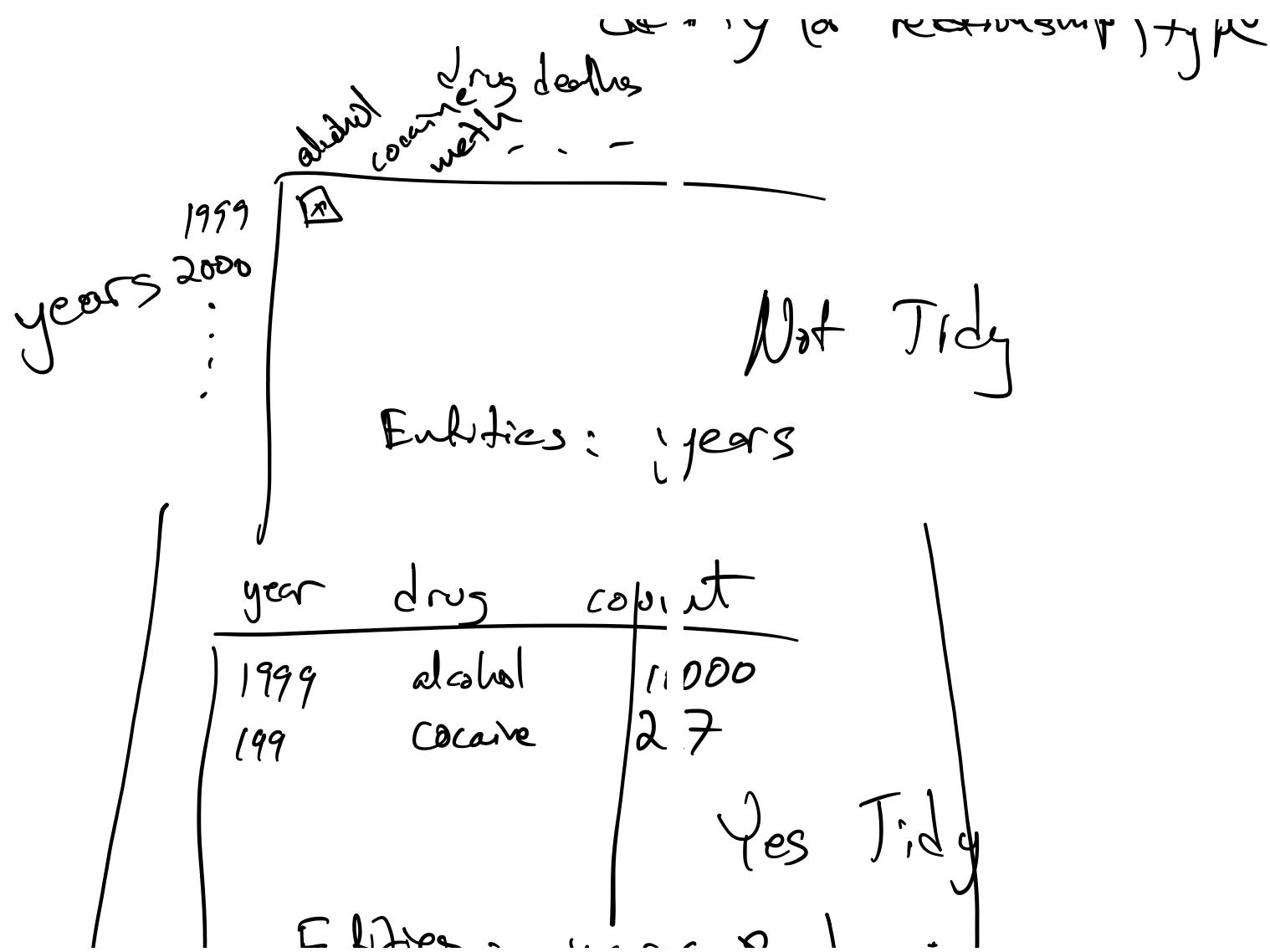
②

each column corresponds to
a single attribute

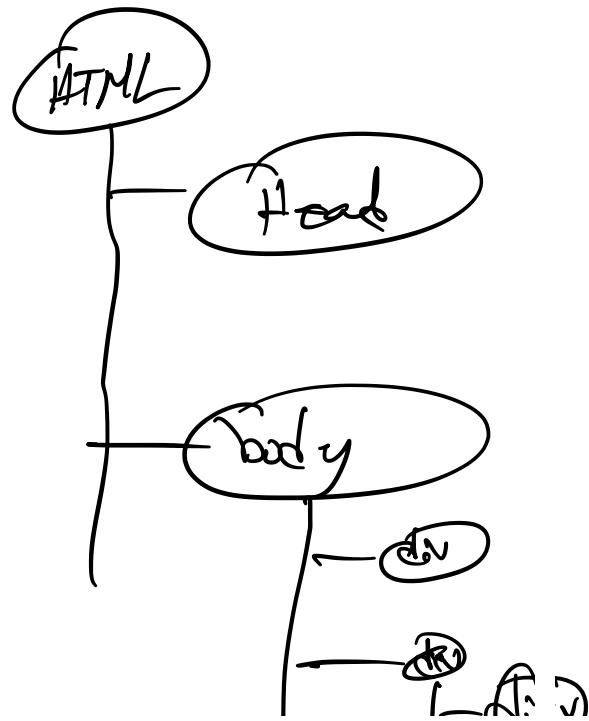
③

each table (relation)

corresponds to a single
entity - column 1:n



|| | ~~years & drug~~ | .



nest
→ 10 .

python:
beautifulsoup



table {
 types
 classes
 attributes

CSS selector syntax

types : "table"
class : ".class"
attribute: "[attribute]"
id : "#id"