YData: An Introduction to Data Science

Lecture 03: Tables

Jessi Cisewski-Kehe and John Lafferty
Statistics & Data Science, Yale University
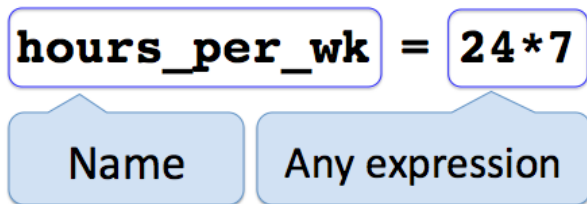Spring 2019

Announcements

Python

Programming Languages

- Python is popular both for data science & general software development

- Mastering the language fundamentals is critical

- Learn through practice, not by reading or listening

- Follow along: hub.ydata123.org

(DEMO)

Names

Assignment Statements

$$\text{hours\_per\_wk} = 24*7$$

Name — Any expression

- Statements don't have a value; they perform an action
- An assignment statement changes the meaning of the name to the left of the $=$ symbol
- The name is bound to a value (not an equation)

(DEMO)

Call Expressions

Anatomy of a Call Expression

What function to call

Argument to the function

$$f(27)$$

"Call f on 27"

(DEMO)

Tables

Table structure

- A Table is a sequence of labeled columns
- Each row represents one individual
- Data within a column represents one attribute of the individuals

Label

Name	Code	Area (m2)
California	CA	163696
Nevada	NV	110567

Row

Column

(DEMO)

Some Table Operations

- `t.select(label)` - constructs a new table with just the specified columns
- `t.drop(label)` - constructs a new table in which the specified columns are omitted
- `t.sort(label)` - constructs a new table with rows sorted by the specified column
- `t.where(label, condition)` - constructs a new table with just the rows that match the condition

Discussion question

`nba` table:

How to display just the row corresponding to the player who had the highest salary?