YData: An Introduction to Data Science

Lecture 03: Tables

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University

Spring 2021

Credit: data8.org

Announcements

- Assignments 0 and 1 have been posted
- hw00 due Monday; hw01 due Thursday
- YCRC computing environment coming soon
- New office hours added
- Questions?

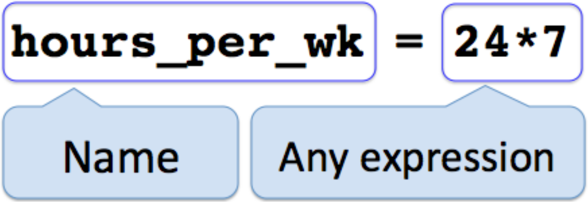
Python

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 watching—like driving a car
- Follow along: mybinder.org/v2/gh/YData123/sds123-sp21/main?filepath=demos/lec03/lec03.ipynb

Names

Assignment Statements



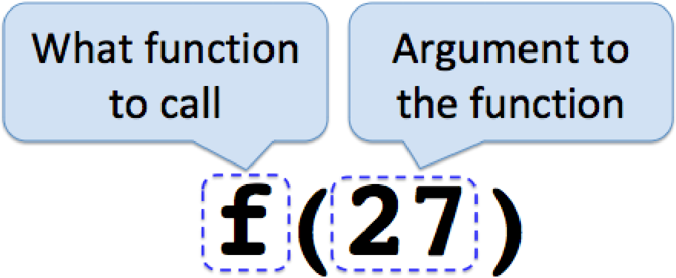
```
hours_per_wk = 24*7
```

- 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



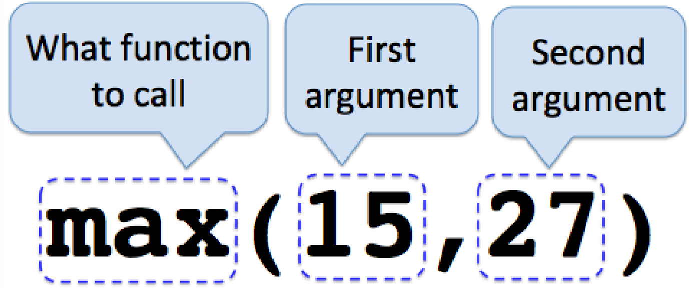
What function to call

Argument to the function

```
f(27)
```

“Call f on 27”

Anatomy of a Call Expression



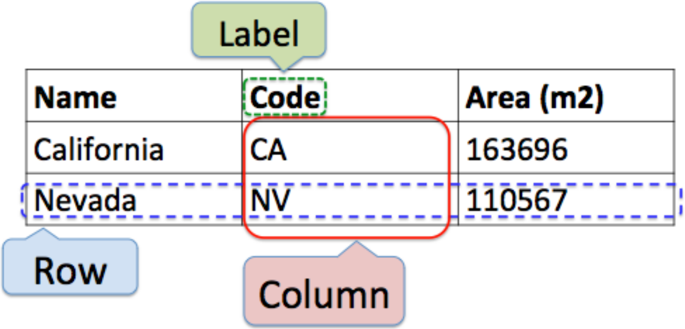
```
max(15,27)
```

(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



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

(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?

Pandas

- FYI: The `datascience` package is a Berkeley product
- It's a light wrapper on top of `pandas`
- Later in the course we'll give an introduction to Pandas

Summary

Today we talked about how to:

- Assign a value to a name
- Call a function
- Build a `Table` (aka dataframe)
- Operate on Tables

Readings

Chapter 3 in "Computational and Inferential Thinking"

https://www.inferentialthinking.com/chapters/03/programming-in-python.html