DATA 8

Spring 2022

Lecture 3

Tables

Announcements

- HW 1 is due Thursday 1/27 @ 11:59pm
  - Submit on Wednesday 1/26 for a bonus!
- Office hours start this week: http://data8.org/sp22/office-hours.html
  - Swupnil: Mondays 3-4pm berkeley.zoom.us/my/swupnil
  - John: Tuesdays 11am-12pm bit.ly/denerozoom
- Small-group tutoring section signups released early this week

This Week: Python & Tables

- Today:
  - Python basics
  - Tables
- Wednesday:
  - Types of data
  - Arrays
- Friday:
  - Creating new tables
  - Manipulating columns of tables

Python

Python

- Python is popular both for data science & general software development
- Mastering the language fundamentals is critical
- Learn through practice:
  - See some examples & learn the rules
  - Try out variants of those examples yourself
  - Write new code that solves new problems

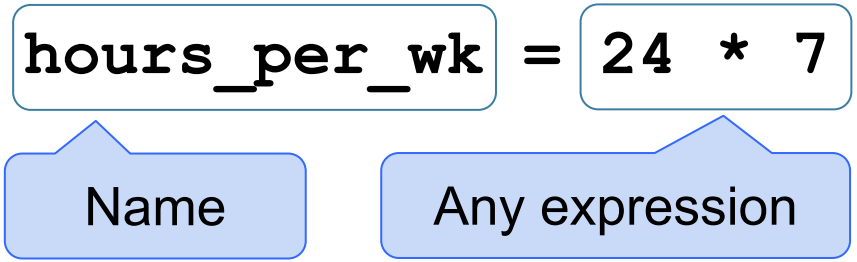
(Demo)

Review of Python Concepts

- An expression evaluates to a value
- Values can be numbers or strings (text); we'll see lots of other kinds of values soon
- The syntax (format) of the language is very rigid — even an extra space can cause a syntax error
- There is particular behavior associated with built-in operators that you need to learn
  (e.g., dividing produces 8.0 instead of 8)

Names

Assignment Statements

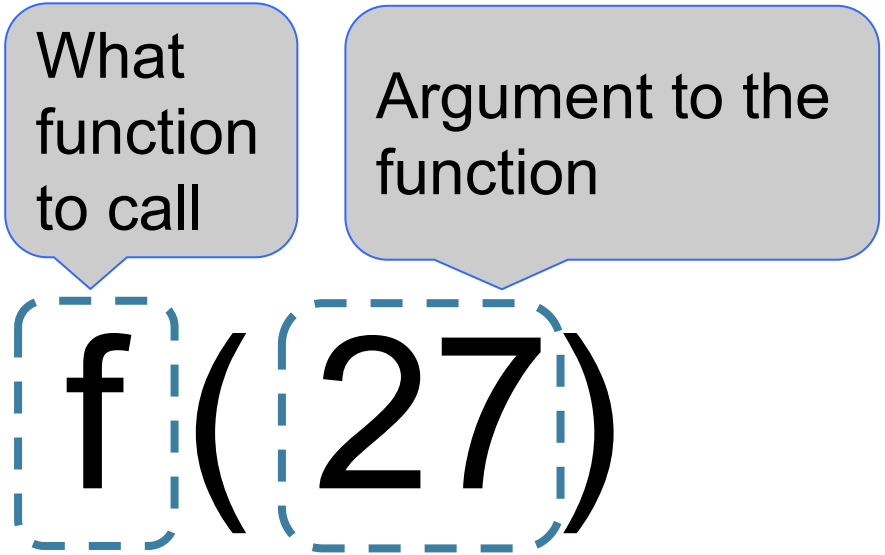


- An assignment statement changes the meaning of the name to the left of the = symbol
- The name is bound to the value of the expression to the right of the = symbol (its current value; not the equation)

(Demo)

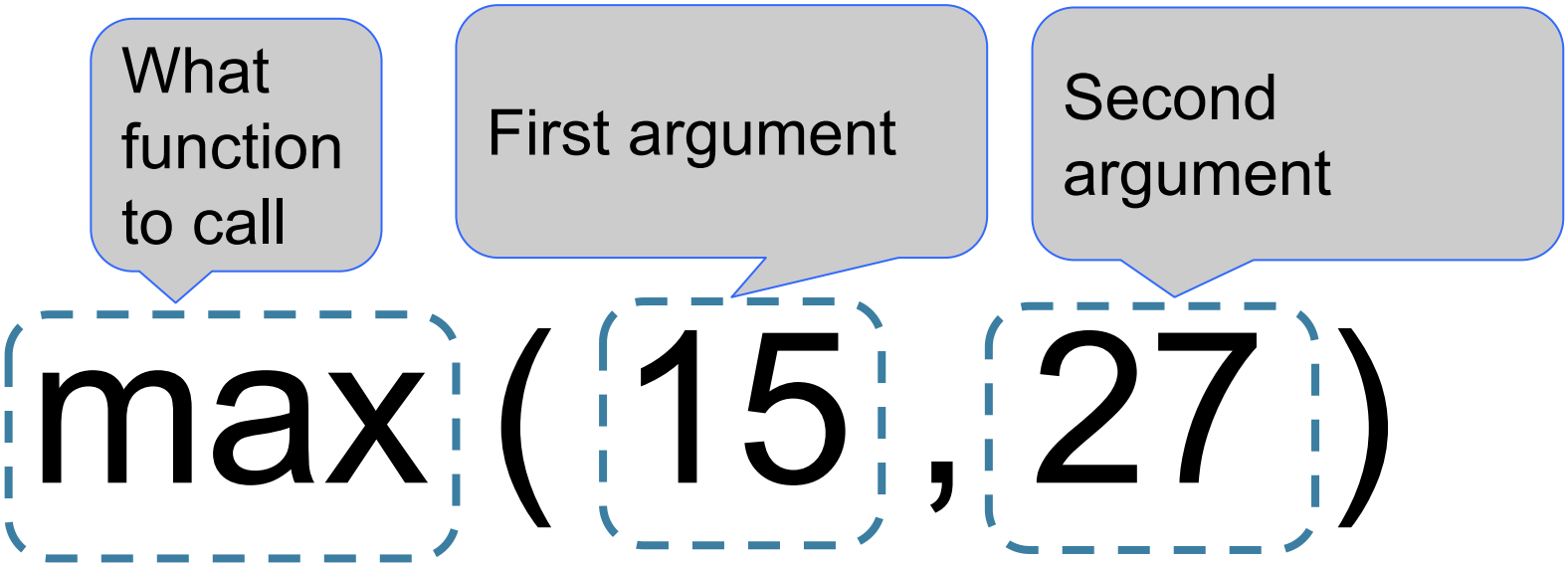
Functions

Anatomy of a Call Expression



"Call f on 27."

Anatomy of a Call Expression



(Demo)

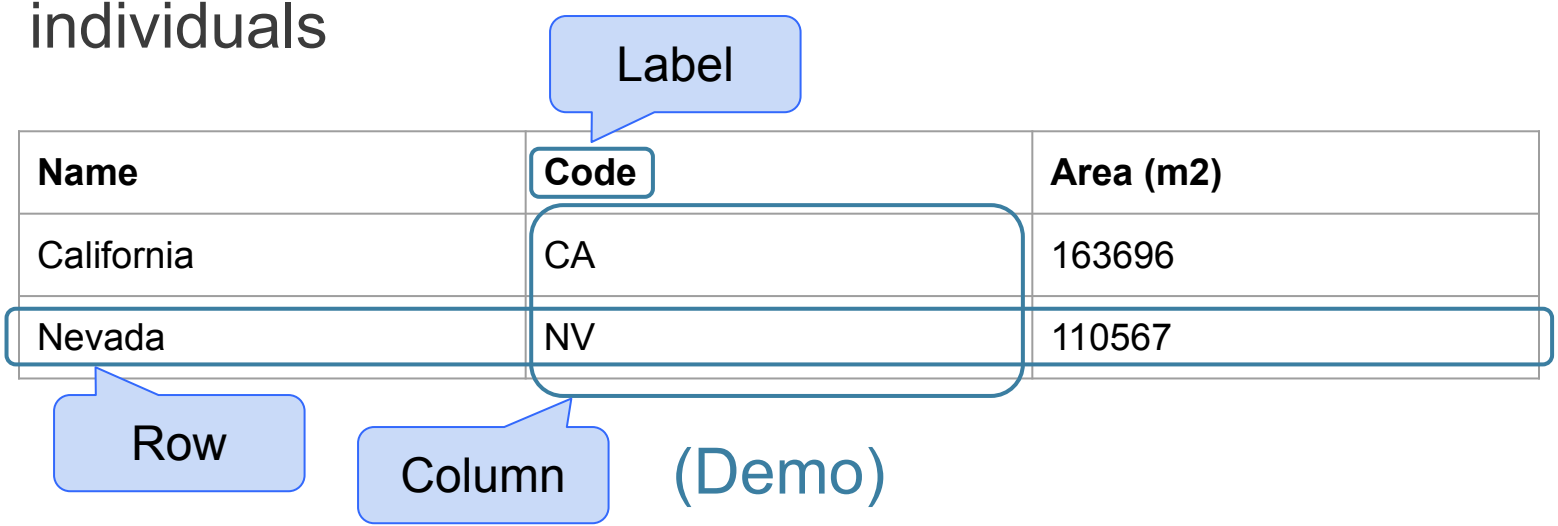
Review of Function Concepts

- Some functions require a particular number of arguments (e.g., **abs** must be called on one value)
- Arguments can be named in the call expression:
  round(number=12.34)
  But the names must match the documentation
- Type a ? after a function name to see its documentation

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



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