DSC 10, Spring 2018
Lecture 2

Expressions

sites.google.com/eng.ucsd.edu/dsc-10-spring-2018

Credit: Anindita Adhikari and John DeNero

Announcements

- Lab 1 due tonight at 11:59pm.
- HW 1 now available, due Sunday 11:59pm.
- Confirm participation points on TritonEd.
- Waitlisted students, please listen to podcast and leave seats in the classroom for enrolled students.

Confounding

Key to establishing causality

If the treatment and control groups are *similar apart from the treatment,* then differences between the outcomes in the two groups can be ascribed to the treatment.

Trouble

If the treatment and control groups have systematic differences other than the treatment, then it might be difficult to identify causality.

Such differences are often present in **observational studies.**

When they lead researchers astray, they are called confounding factors.

Randomize!

- If you assign individuals to treatment and control **at random,** then the two groups are likely to be similar apart from the treatment.
- You can account – mathematically – for variability in the assignment.
- **Randomized Controlled Experiment**

Randomized Controlled Experiments

- Assign individuals to treatment and control **at random**

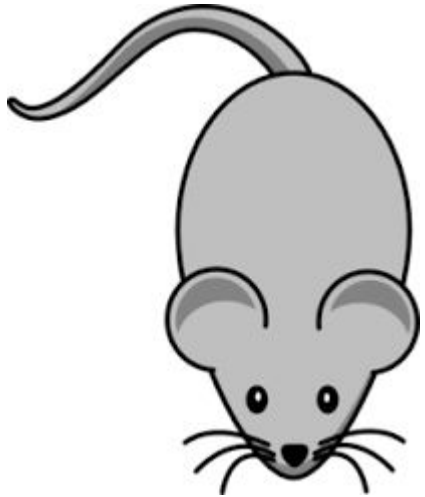
Which of these questions cannot be answered by running a randomized controlled experiment?

A. Does daily meditation reduce anxiety?
B. Does playing video games increase aggressive behavior?
C. Does smoking cigarettes cause weight loss?
D. Does early exposure to classical increase a child's IQ?

Careful ...

Regardless of what the dictionary says,
in probability theory

Random ≠ Haphazard



Summary: Cause & Effect

Comparison

- Group by some *treatment* and measure some *outcome*
- Simplest setting: a *treatment group* and a *control group*
- If the *outcome* differs between these two groups, that's evidence of an *association* (or *relation*)
  - E.g., the top-tier chocolate eaters died of heart disease at a lower rate (12%) than chocolate abstainers (17%)
- If the two groups are similar in all ways but the *treatment*, a difference in the *outcome* is also evidence of *causality*

Confounding

- If the treatment and control groups have systematic differences other than the treatment itself, then it might be difficult to identify a causal link
- When these systematic differences lead researchers astray, they are called *confounding factors*
- Such differences are often present in observational studies
  - *Observational study*: the researcher does not choose which subjects receive the treatment
  - *Controlled experiment*: the researcher designs a procedure for selecting the treatment and control groups

Randomize!

- When subjects are split up *randomly*, it's unlikely that there will be systematic differences between the groups
- And it's possible to account for the chance of a difference
- Therefore, *randomized controlled experiments* are the most reliable way to establish causal relations

Expressions

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: jupyterhub.ucsd.edu

(Demo)

Arithmetic Operators

Operation	Operator	Example	Value
Addition	+	2 + 3	5
Subtraction	-	2 - 3	-1
Multiplication	*	2 * 3	6
Division	/	7 / 3	2.66667
Remainder	%	7 % 3	1
Exponentiation	**	2 ** 0.5	1.41421

Numbers

Ints and Floats

Python has two real number types

- `int`: an integer of any size
- `float`: a number with an optional fractional part

Ints and Floats

Python has two real number types

- **`int`**: an integer of any size
- **`float`**: a number with an optional fractional part

An **`int`** never has a decimal point; a **float** always does

A **`float`** might be printed using scientific notation

Ints and Floats

Python has two real number types

- `int`: an integer of any size
- `float`: a number with an optional fractional part

An `int` never has a decimal point; a **float** always does

A `float` might be printed using scientific notation

Three limitations of float values:

- They have limited size (but the limit is huge)
- They have limited precision of 15-16 decimal places
- After arithmetic, the final decimal few places can be wrong

Rank from least to greatest

Rank the results of the following expressions in order from least to greatest.

A. `3 * 10 ** 10`

B. `10 * 3 ** 10`

C. `(10 * 3) ** 10`

D. `10 / 3 / 10`

E. `10 / (3 / 10)`

iClicker:

Select the expression whose value is exactly in the middle of the five values (the third largest).

Rank from least to greatest

Rank the results of the following expressions in order from least to greatest.

A. `3 * 10 ** 10`

B. `10 * 3 ** 10`

C. `(10 * 3) ** 10`

D. `10 / 3 / 10`

E. `10 / (3 / 10)`

A. `30000000000`

B. `590490`

C. `590490000000000`

D. `0.3333333333333337`

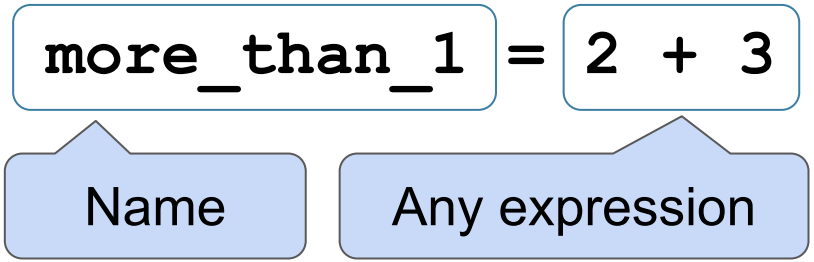
E. `33.333333333333336`

Names (Variables)

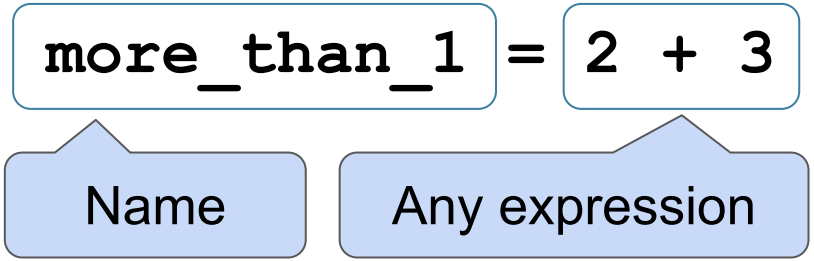
Assignment Statements

```
more_than_1 = 2 + 3
```

Assignment Statements

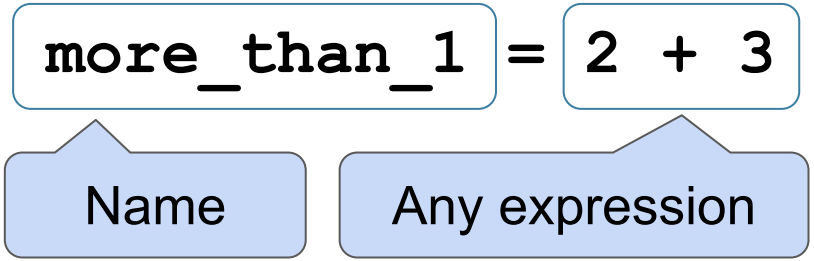


Assignment Statements



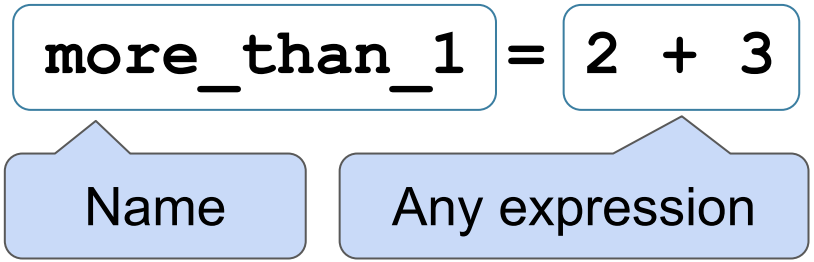
- Statements don't have a value; they perform an action

Assignment Statements



- 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

Assignment Statements

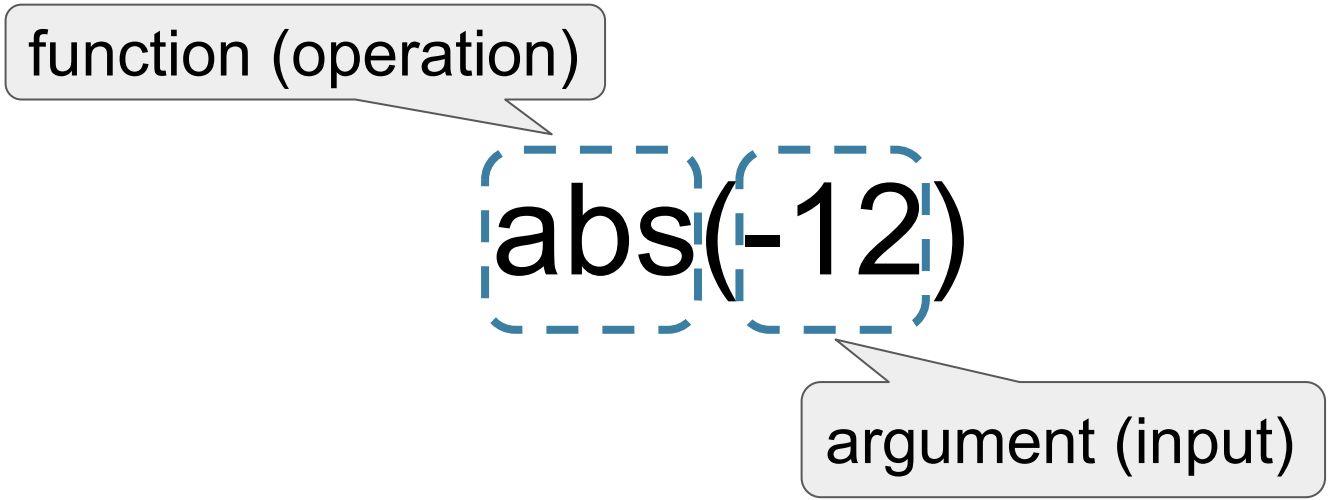


- Statements don't have a value; they perform an action
- An assignment statement changes the meaning of the name to the left of the = symbol
- The name is bound to a value (not an equation)

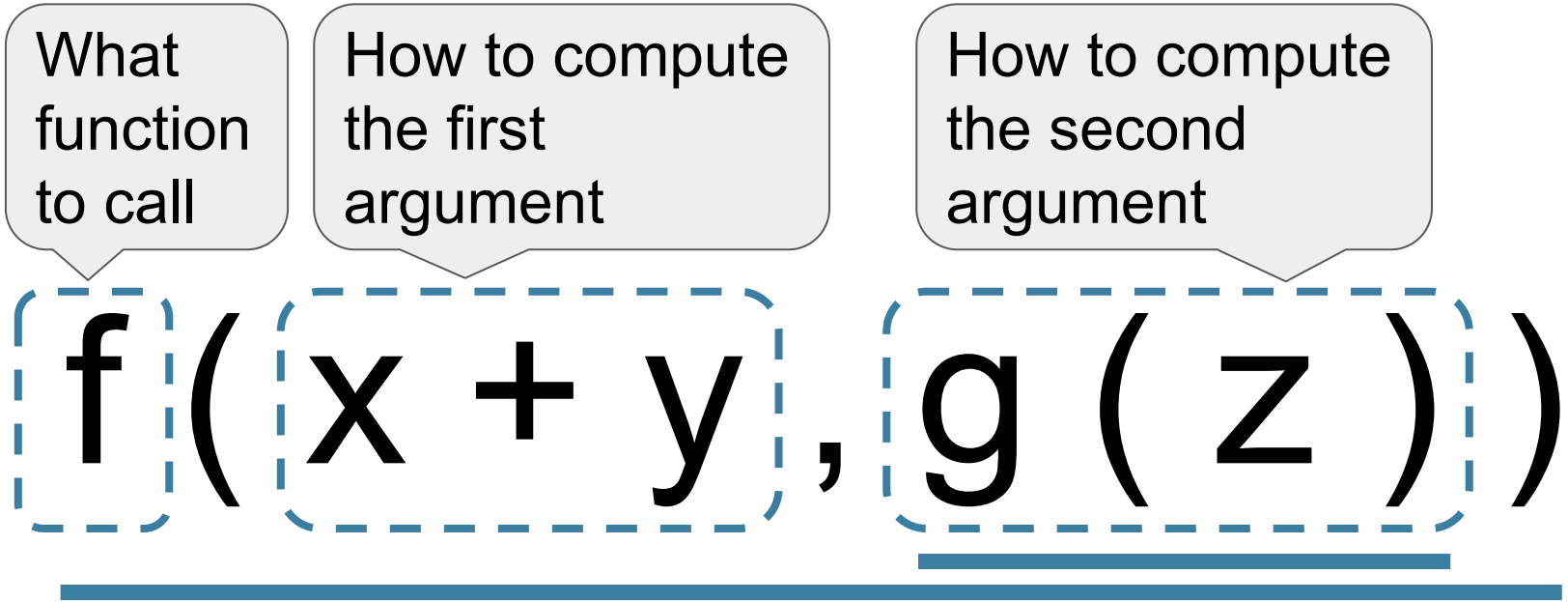
(Demo)

Call Expressions

Anatomy of a Call Expression



Anatomy of a Call Expression



"Call f on the result of adding x + y and the return value of calling g on z."

Discussion Question

Assume you have run the following statements:

```
x = 3
y = -2
```

Which of these examples results in an error?

A. `abs(x, y)`
B. `math.pow(x, abs(y))`
C. `round(x, max(abs(y**2)))`
D. `math.pow(x, math.pow(y, x))`
E. More than one of the above