Python for Text Analysis 2018-2019

Lecture 1: Introduction
29-10-2018

Welcome!

Part I: Introduction

- Who is who?
- About this course (Github)

Part II: Getting started

- Installing Anaconda
- Using Jupyter Notebooks
- Chapter 1: Getting started with Variables and Values
- Chapter 2: Basic Data Types (Integers and Floats)
- (Chapter 3: Strings)

Who are we? The Teachers

Pia Sommerauer

Chantal van Son

Who are we? The Teaching Assistants

Lenka Bajčetić

Jenia Kim

Thomas Klein

Who are we? The Contributors

Also many thanks to these fantastic guys for designing this course and contributing to the material in previous years!

Marten Postma

Emiel van Miltenburg

Filip Ilievski

Who are you?

- **BA Students** *(Minor Digital Humanities and Social Analytics)*
  Programming for the Humanities & Social Sciences

- **ReMa Students** *(Research Master Humanities - Linguistics)*
  Python for Text Analysis

- **Canvas:** everyone in right Canvas page? (Python for Text Analysis)

- … we want to know more about you!

SOCRATIVE

Log in at:

b.socrative.com/student

Room:

PYTHON1819VU

About this course

- Practical course geared to those who want to get some **hands-on experience** working with **language data**

- **No knowledge of programming is required or presupposed!**

Goals of this course

You will learn how to:

- work with the **standard library** of Python
- deal with **different file types** (e.g. plain text, CSV/TSV, JSON, XML)
- use some **external libraries** (e.g. to analyse texts)
- **document** and **share** your code and results

Focus on **readability** and **understandability** (for sharing and re-using your code)

Becoming an **independent programmer** (problem solving)

Assignments & Grading: BA Students

Part	weight %		Part	weight %
Assignment 1	9		Total Assignments	60
Assignment 2	17		Exam	40
Assignment 3	17			
Assignment 4	17			
			Total	100
Total Assignments	60			

Assignments & Grading: ReMa Students

Part	weight %		Part	weight %
Assignment 1	5		Total Assignments	35
Assignment 2	10		Exam	20
Assignment 3	10		Final assignment	45
Assignment 4	10			
			Total	100
Total Assignments	35			

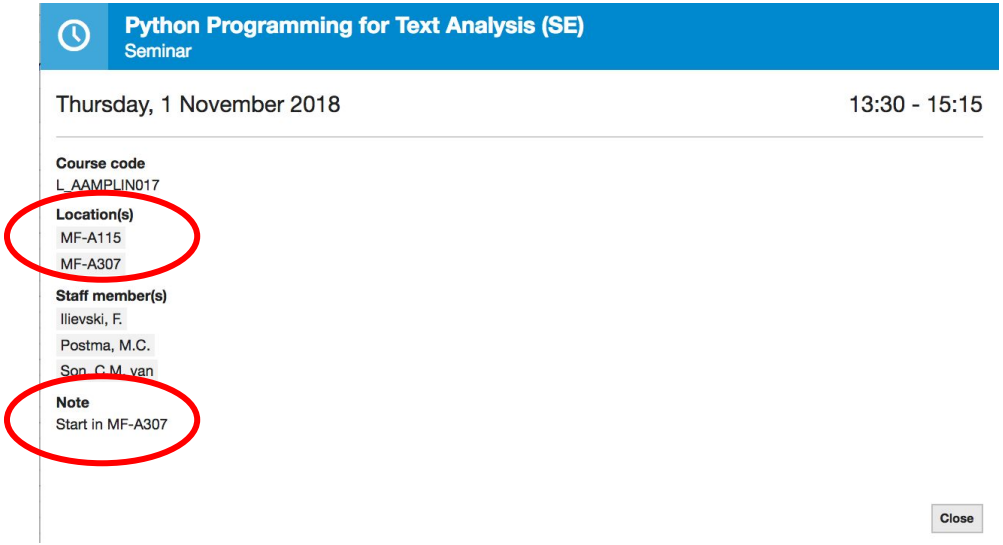
Blocks

- Each block has **three lectures**:
  1. Introduction of theory
  2. Theory + working on assignment
  3. Feedback on assignment

Block	Chapters	Assignment
Block 1	Chapters 1-4	Assignment 1
Block 2	Chapters 5-10	Assignment 2
Block 3	Chapters 11-15	Assignments 3a and 3b
Block 4	Chapters 16-18	Assignments 4a and 4b

Rooms

- We will have many different rooms, so **always check the schedule!**
  - Personal schedule on Canvas/VUnet
  - https://rooster.vu.nl
- Especially the ones on **Thursday**



Important dates

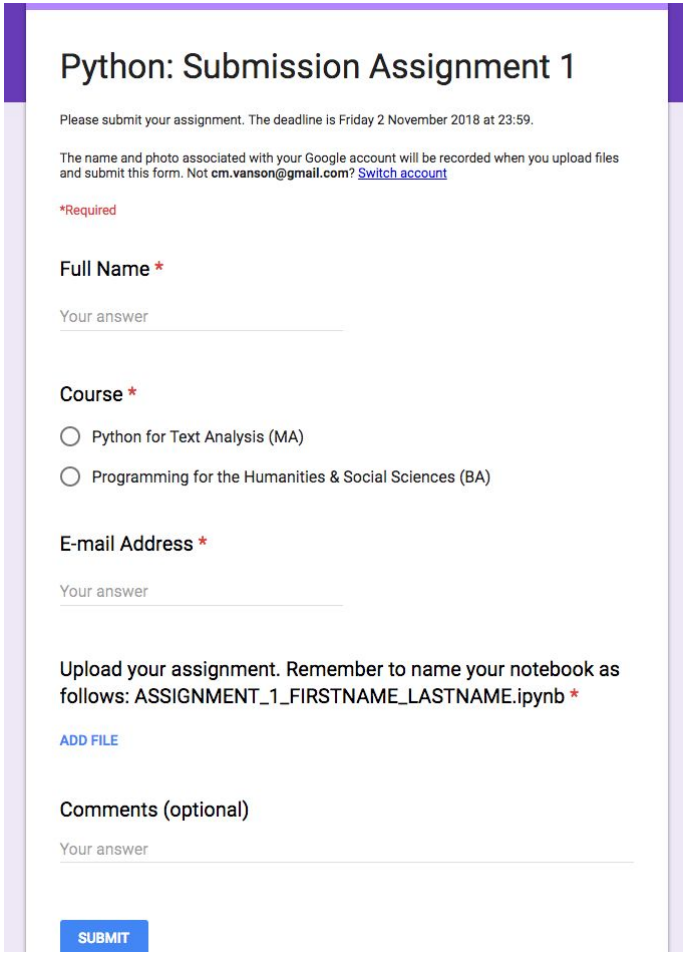
- **Assignment 1:**
  Friday 2 November at 23:59
- **Assignment 2:**
  Tuesday 13 November at 20:00
- **Assignment 3:**
  Friday 23 November at 23:59
- **Assignment 4:**
  Tuesday 4 December at 20:00

- **Exam**
  Monday 17 December at 08:45

- **Final assignment (ReMa):**
  Sunday 3 February at 23:59

Submitting Assignments

- **Google Forms** for submission (see Github)
- Please submit your assignments **on time**!!
  - We are strict about this, because we will need time to grade your assignments and to prepare the feedback sessions
  - **One day late** results in **2 points deduction**
  - After that, the **resulting grade is a 1**
  - If you have any good reason to be late, please let us know ASAP
- **Re-uploading assignments** is possible (we will use the last version), but please try to avoid
- **Plagiarism**
  - Let us know in the comments if you have **worked together** with someone or if you **used code from online sources**



Python: Submission Assignment 1

Please submit your assignment. The deadline is Friday 2 November 2018 at 23:59.

The name and photo associated with your Google account will be recorded when you upload files and submit this form. Not **cm.vanson@gmail.com**? Switch account

*Required

Full Name *

Your answer

Course *

- Python for Text Analysis (MA)
- Programming for the Humanities & Social Sciences (BA)

E-mail Address *

Your answer

Upload your assignment. Remember to name your notebook as follows: ASSIGNMENT_1_FIRSTNAME_LASTNAME.ipynb *

ADD FILE

Comments (optional)

Your answer

SUBMIT

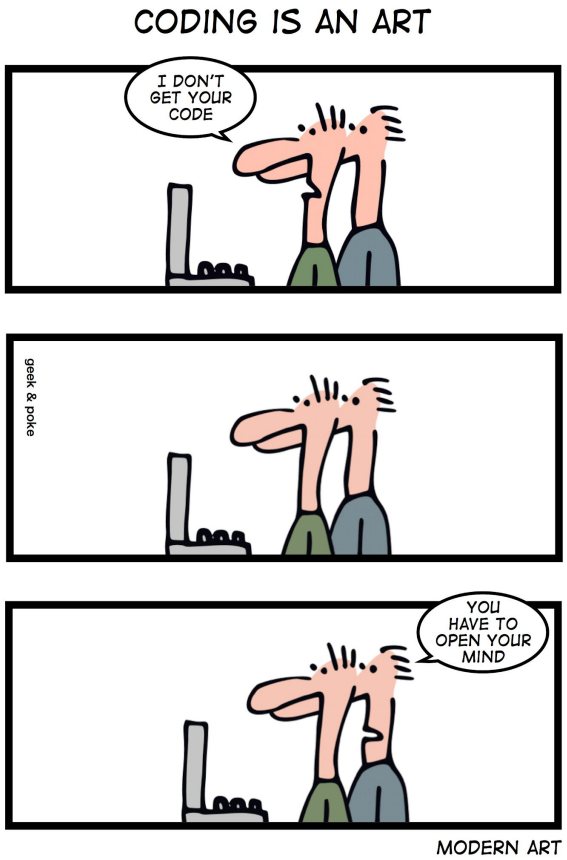
Exam

- Written exam
  - **BA students:** final test
  - **ReMa students:** serves as a check of your knowledge before final assignment
- Test basic knowledge about **syntax** of Python and **its standard library**
- You **cannot** pass the course without a **passing grade** on the exam

Attendance

- We will **not** check attendance; however, we do appreciate it if you can **let us know** when you cannot make it

- Your own responsibility:
  - In our experience, it will be **much harder** to follow the course if you do not attend the lectures



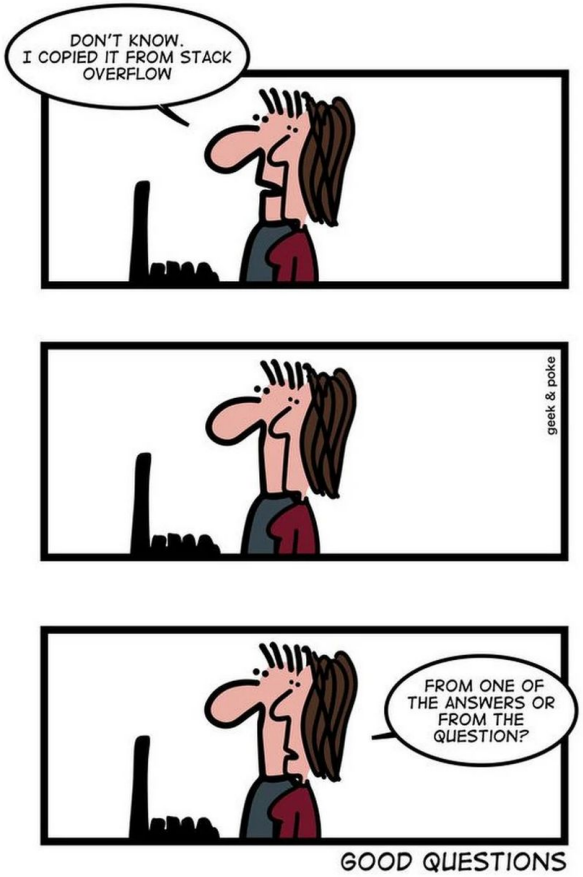
TIP 1:

Pair  Programming

https://en.wikipedia.org/wiki/Rubber_duck_debugging

TIP 2:

Rubber Duck Debugging



TIP 3:

Google & Stack Overflow

If you're stuck for more than 15 minutes: contact us!

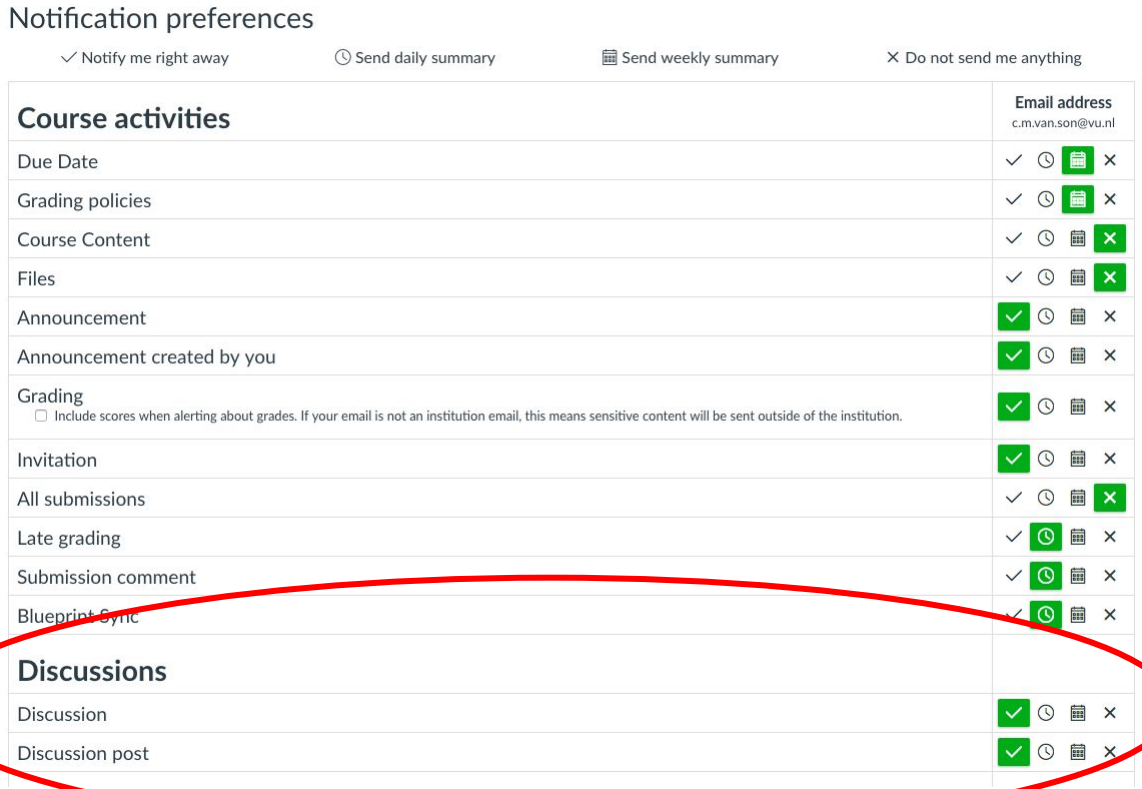
Please use the **Canvas Discussion forum** to ask questions.

If you need to provide a longer code sample to explain your question, send the teachers + TAs a **private message on Canvas**.

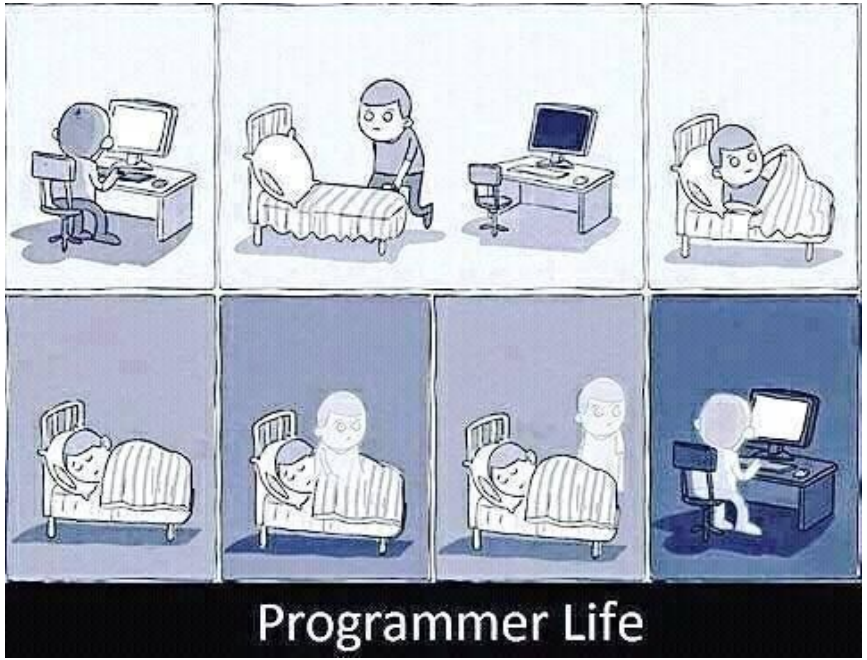
TIP 4:

The Fifteen Minute Rule

Canvas Discussions: Notification settings



Now let’s get started!



Preparation: Anaconda & Github

Installing Anaconda

- Go to: https://www.anaconda.com/download
- Choose version Python 3.7

Download Github repository

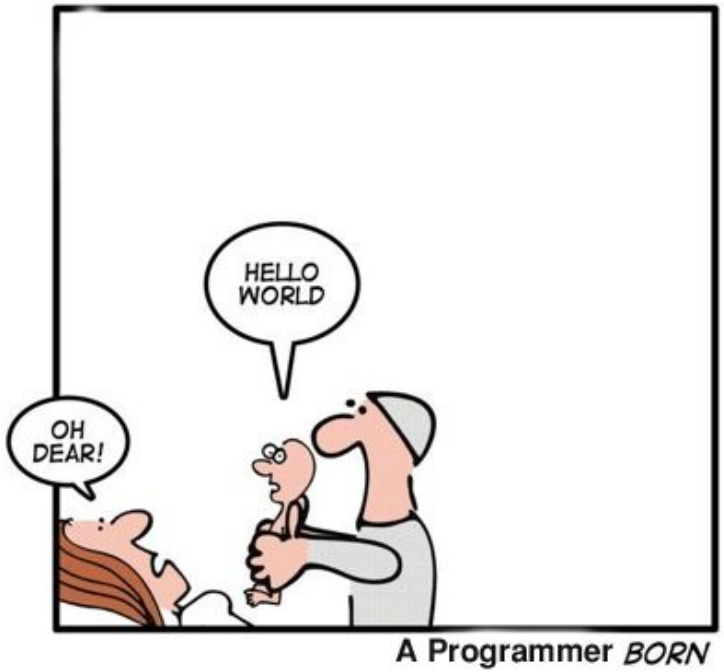
- Go to: https://github.com/cltl/python-for-text-analysis
- Click on ‘Clone or download’ => ‘Download ZIP’

Running Jupyter Notebooks

- To run a Jupyter Notebook, you can either:
  1. Open the **Anaconda Navigator** and click on the **Jupyter Notebook icon**

  2. Open a **terminal** (*cmd* in Windows) and type the following command:
     `jupyter notebook`

- If necessary, you can read the full instructions here



CHAPTER 1:
Getting Started with Variables and Values

Printing “Hello, world!”

```
print(“Hello, world!”)
```

Printing “Hello, world!”

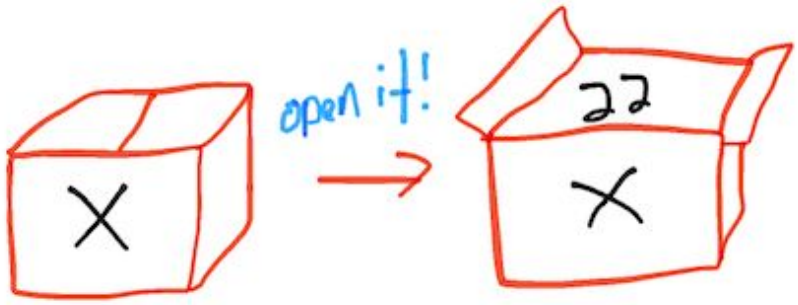
```
print(“Hello, world!”)
```

```
text = “Hello, world!”
print(text)
```

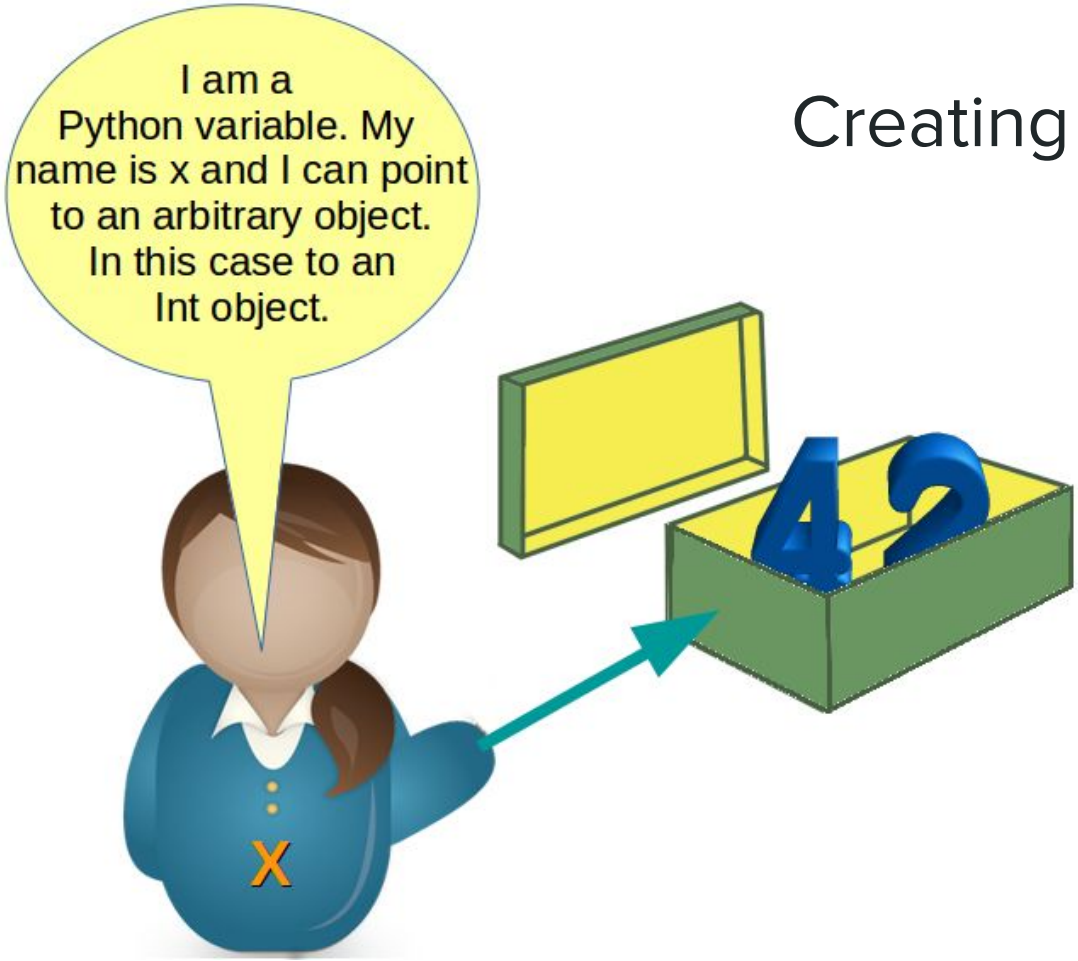
Variables and values

- A **variable** is a place in the computer's memory that holds a temporary **value**
- We can place stuff in the variable by **assigning** it to the variable

```
x = 22
```

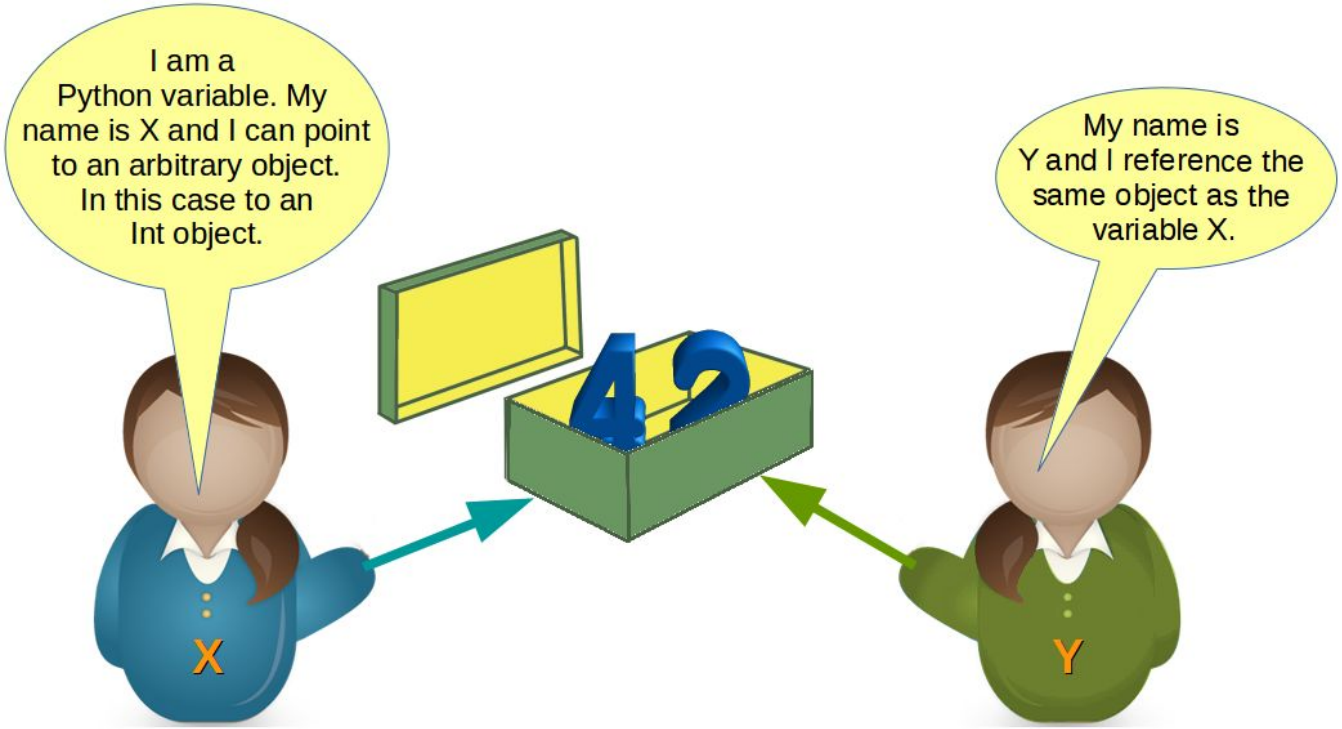


Creating a variable



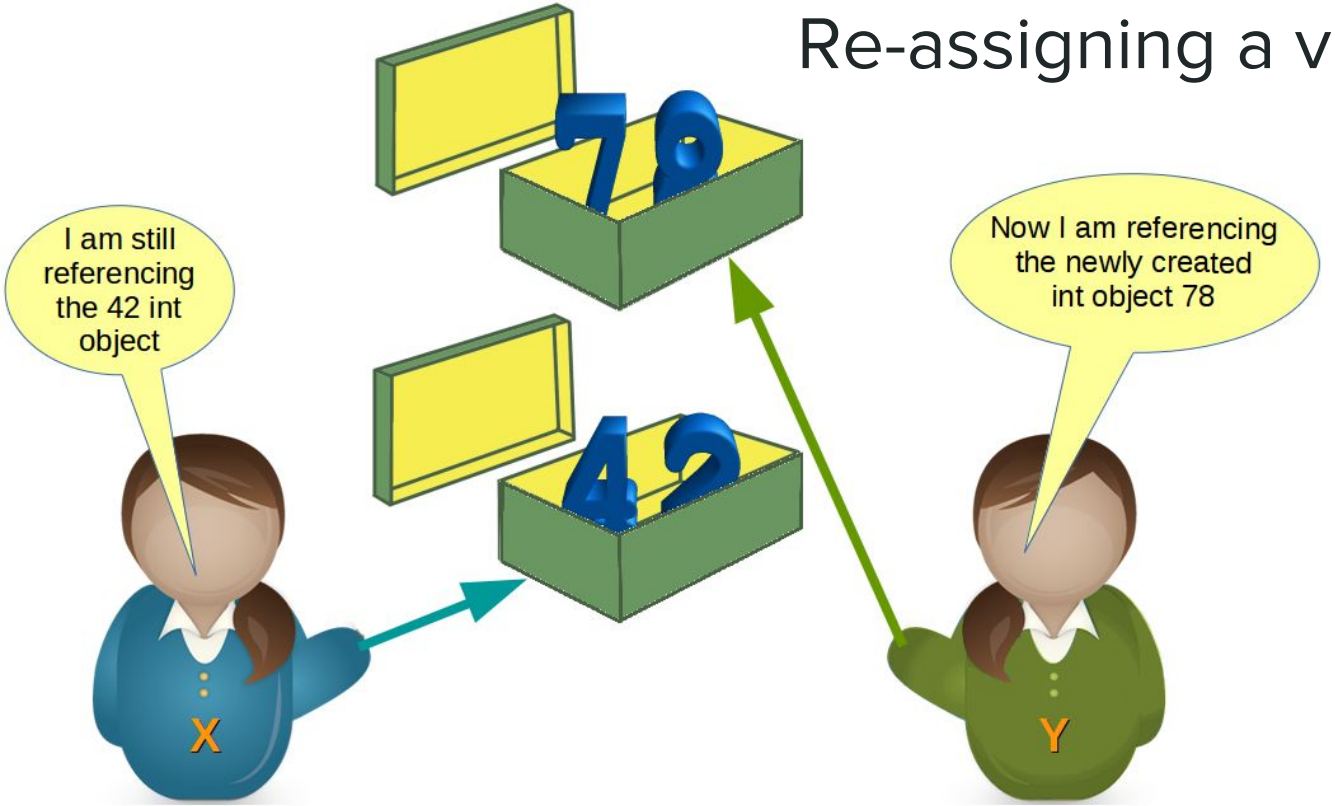

```
x = 42
```

Copying a variable




```
x = 42
y = x
```

Re-assigning a variable




```
x = 42
y = x
y = 78
```

The `input()` function

programmer

```
x = “hello”
```

user

```
x = input()
```

Variables names

- Variable names are **valid** if they:
  - start with a letter or underscore (_)
  - only contain letters, numbers and underscores

- Naming conventions:
  - use **clear, meaningful, descriptive variable names** so that your code will remain understandable
  - use the **lowercase_with_underscores style**, with lowercase characters and underscores for separating words
  - **do NOT use built-in names**, such as print or sum (these will turn green in Jupyter Notebooks)

SOCRATIVE

Log in at:

b.socrative.com/student

Room:

PYTHON1819VU

CHAPTER 2: Basic Data Types (Integers and Floats)

Basic Data Types

Type	Explanation
String	text
Integer	whole numbers
Float	numbers with decimals
Tuple	immutable combinations of values
List	ordered sequences of objects
Set	unordered sets of objects
Dictionary	mappings between objects
Booleans	the truth values True and False
Functions	to manipulate objects, or to produce new objects given some input

Creating Data Types

- Each data type is created with specific characters, such as:
  - double/single quotes
  - round/square/curly brackets

```
a_string      = 'test'
an_integer    = 4
a_float       = 3.14
a_tuple       = (2,5)
a_list        = [1,2,3,1,2,3,'a','b','c']
a_set         = {1,2,3,4,'apple'}
a_dict        = {'milk':2, 'cheese':1, 'pickles':45}
a_function    = print
a_bool        = True
```

Casting Data Types

- Converting one type into another (if possible)

```
x = 42
x = str(x)
```

```
x = “42”
x = int(x)
```

Integers and floats

- Numerical types
  - **Integers (int):** whole numbers
  - **Floats (float):** numbers with decimals

- Mathematical operators (see table)
  - Operator precedence:

```
nr1 = 10-2/4
nr2 = (10-2)/4
nr3 = 10-(2/4)
```

  - Shortcuts:

```
nr_books = 100
nr_books += 1
```

Operation	Result
`x + y`	sum of x and y
`x - y`	difference of x and y
`x * y`	product of x and y
`x / y`	quotient of x and y
`x // y`	floored quotient of x and y
`x % y`	remainder of x / y
`-x`	x negated
`+x`	x unchanged
`x ** y`	x to the power y

SOCRATIVE

Log in at:

b.socrative.com/student

Room:

PYTHON1819VU

Before Thursday:

- Work through Chapters 1-4 (including Exercises)
- Make a start with Assignment 1 (deadline on Friday!!)