

The ABC of Computational Text Analysis

#9 *INTRODUCTION TO PYTHON*

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

May 5, 2023

Recap last Lecture

- from words to embeddings
 - recontextualized word meaning
- data-driven NLP is both powerful and biased
- data is never raw but depends on many decisions

Outline

- enter the shiny world of Python 😎
 - programming basics
 - development editor
- think about mini-project



Python

Python is ...

a programming language that is ...

- general-purpose
not specific to any domain
- interpreted
no compiling
- standard language in data science



Popular programming languages src

How to learn programming?

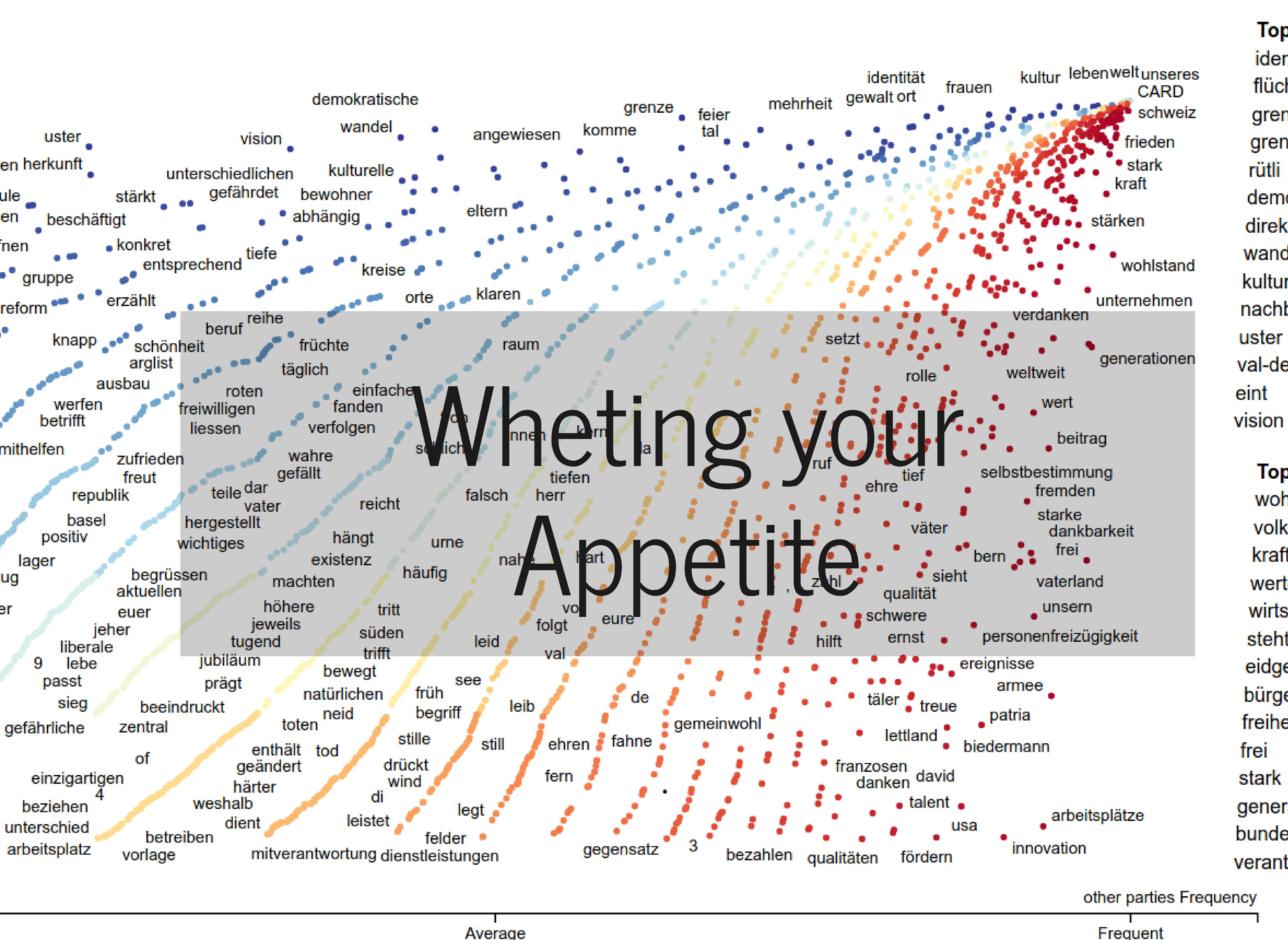
Three inconvenient truths



- programming cannot be learnt in a course
I try to make the start as easy as possible!
- frustration is part of learning
fight your way!
- the Python ecosystem is huge
grow skills by step-by-step

Programming can be absolutely captivating!





Programming Concepts and Python Syntax

Variables

Variables are kind of storage boxes

```
1 # define variables
2 x = "at your service"
3 y = 2
4 z = ", most of the time."
5
6 # combine variables
7 int_combo = y * y      # for numbers any mathematical operation
8 str_combo = x + z      # for text only concatenation with +
9
10 # show content of variable
11 print(str_combo)
```

Data Types

The type defines the object's properties

Name	What for?	Type	Examples
String	Text	str	"Hi!"
Integer, Float	Numbers	int, float	20, 4.5
Boolean	Truth values	bool	True, False
:	:	:	:
List	List of items (ordered, mutable)	list	["Good", "Afternoon", "Everybody"]
Tuple	List of items (ordered, immutable)	tuple	(1, 2)
Dictionary	Relations of items (unordered, mutable)	dict	{"a":1, "b": 2, "c": 3}

Data Type Conversion

Combine variables of the same type only

```
1 # check the type
2 type(YOUR_VARIABLE)
3
4 # convert types (similar for other types)
5 int('100') # convert to integer
6 str(100)   # convert to string
7
8 # easiest way to use a number in a text
9 x = 3
10 mixed = f"x has the value: {x}"
11 print(mixed)
```

Confusing Equal-Sign

= vs. == contradicts the intuition

```
1 # assign a value to a variable
2 x = 1
3 word = "Test"
4
5 # compare two values if they are identical
6 1 == 2          # False
7 word == "Test"  # True
```

Comments

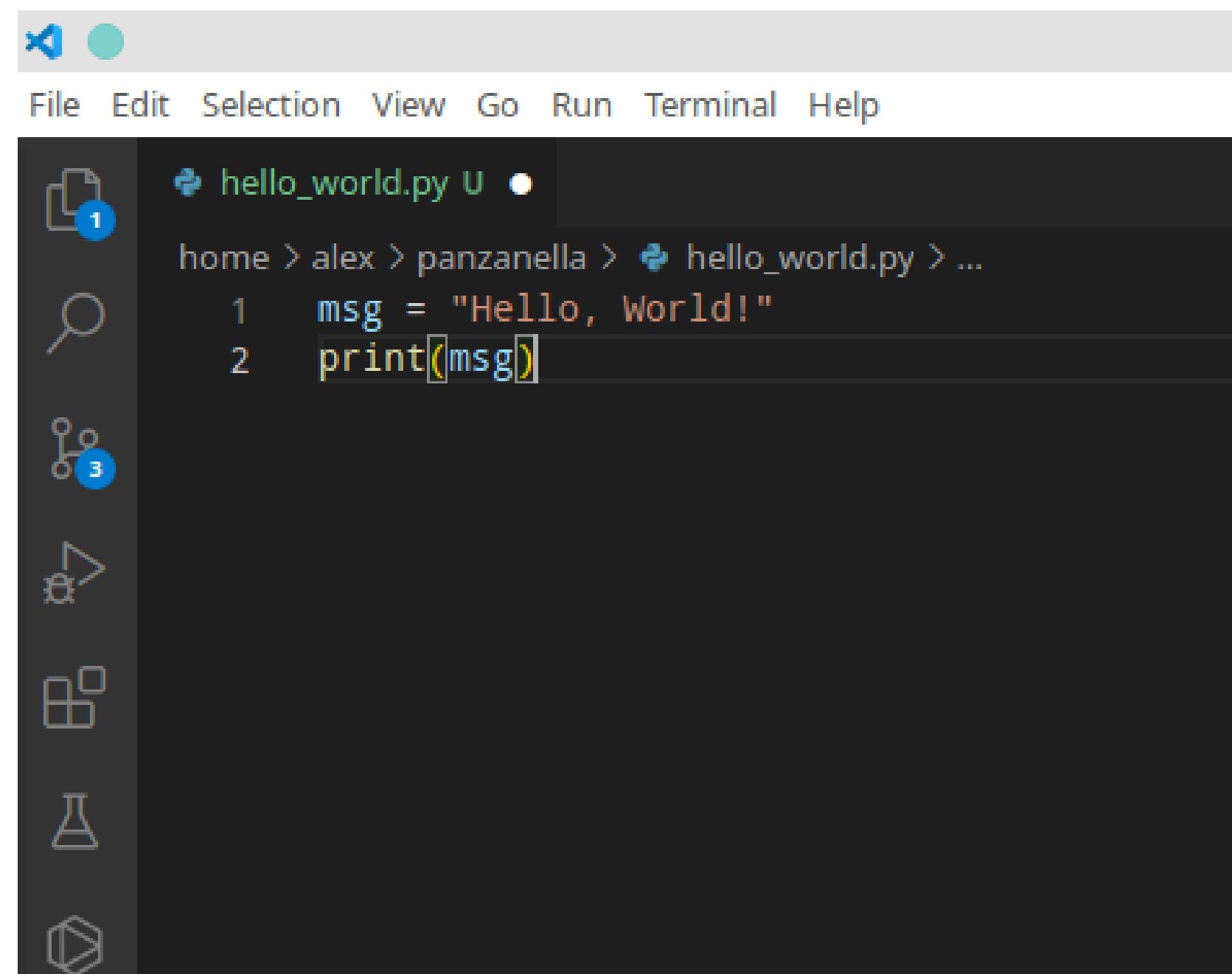
- lines ignored by Python
- write comments, it helps you ...
 - to learn initially
 - to understand later

```
1 # single line comment
2
3 """
4 comment across
5 multiple
6 lines
7 """
```

Visual Studio Code

The (best) editor to program in Python

- integrated development environment (IDE)
 - interactive development
 - similar to RStudio
- 3 views in editor
 - programming (left)
 - output (right)
 - additional information (bottom)
- use **tab** for autocomplete



A screenshot of the Visual Studio Code interface. The top bar shows the menu: File, Edit, Selection, View, Go, Run, Terminal, Help. The title bar indicates the file is 'hello_world.py'. The left sidebar has icons for file (with a '1' notification), search, symbols (with a '3' notification), and other development tools. The main editor area contains the following Python code:

```
msg = "Hello, World!"  
print(msg)
```

Visual Studio Code

In-class: Run your first Python Program I

In-class: Run your first Python Program I

A screenshot of the Visual Studio Code interface. The top menu bar shows 'File', 'Edit', 'Selection', 'View', 'Go' (which is highlighted in blue), 'Run', 'Terminal', and 'Help'. The title bar indicates the file 'hello_world.py - Visual Studio Code'. The left sidebar has icons for 'Get Started', 'python_basics.ipynb', and 'hello_world.py'. The main area shows a terminal window with the following content:

```
C: > Users > ked > Desktop > hello_world.py
1 msg = "Hello World!"
2 print(msg)
```

The terminal also displays the output of the Python code:

```
Started 'Python 3.8.10 64-bit' kernel
Python 3.8.10 (default, Nov 26 2021, 20:14:08)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.30.1 -- An enhanced Interactive Python. Type '?' for help.

✓ msg = "Hello World!" ...
... Hello World!
```

At the bottom, there is a status bar with the text 'WSL: Ubuntu', 'tabnine: Partial support, click to resolve', 'Ln 2, Col 11', 'Spaces: 4', 'UTF-8', 'CRLF', 'Python 3.8.10 64-bit', and icons for refresh and search.

Iterations

for-loop

do something with each element of a collection

```
1 sentence = ['This', 'is', 'a', 'sentence']
2
3 # iterate over each element
4 for token in sentence:
5
6     # do something with the element
7     print(token)
```

Conditionals

if-else statement

condition action on variable content

```
1 sentence = ['This', 'is', 'a', 'sentence']
2
3 if len(sentence) < 3:
4     print('This sentence is shorter than 3 tokens')
5 elif len(sentence) == 3:
6     print('This sentence has exactly 3 tokens')
7 else:
8     print('This sentence is longer than 3 tokens')
```

Indentation matters!

- intend code within code blocks
loops, if-statements etc.
- press **tab** to intend



```
1 if 5 > 2:  
2     print('5 is greater than 2')
```



```
1 if 5 > 2:  
2 print('5 is greater than 2')
```

Methods

Do somethin with an object

```
1 sentence = 'This is a sentence'  
2  
3 # split at whitespace  
4 tokens = sentence.split(' ')  
5  
6 # check the variables  
7 print(sentence, type(sentence), tokens, type(tokens))  
8  
9 # add something to a list  
10 tokens.append('.')
```

11

```
12 # concatenate elements to string  
13 tokens = ' '.join(tokens)  
14 print(tokens, type(tokens))
```

Functions and Arguments

DRY: Don't Repeat Yourself

- functions have a name and optional arguments

function_name(arg1, ..., argn)

```
1 # define a new function
2 def get_word_properties(word):
3     """
4     My first function to print word properties.
5     It takes any string as argument (variable: word).
6     """
7
8     # print(), len() and sorted() work also as functions
9     length = len(word)
10    sorted_letters = sorted(word, reverse=True)
11    print(word, 'length:', length, 'letters:', sorted_letters)
12
13 get_word_properties('computer') # call function with any word
```

Indexing

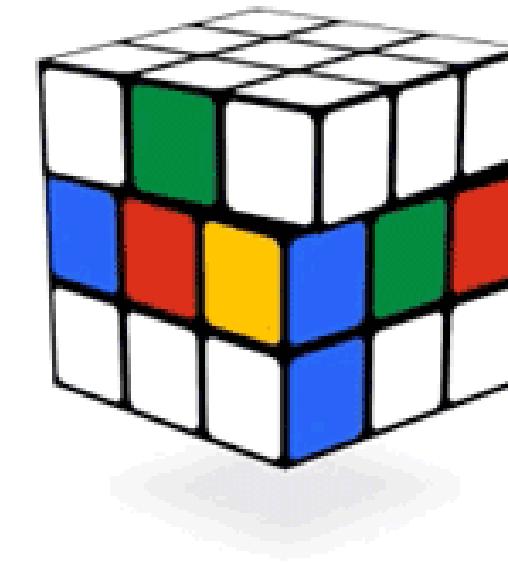
Computers start counting from zero! 😱

```
1 sentence = ['This', 'is', 'a', 'sentence']
2
3 # element at position X
4 first_tok = sentence[0]      # 'This'
5
6 # elements of subsequence [start:end]
7 sub_seq = sentence[0:3]       # ['This', 'is', 'a']
8
9 # elements of subsequence backwards
10 sub_seq_back = sentence[-2:]    # ['a', 'sentence']
```

Errors

A myriad of things can go wrong

1. read the message
2. find the source of the error
script name + line number
3. paste message into Google



Play with more cubes at [Chrome Cube Lab](#)

Learning by doing, doing by googling

Modules/Packages

No programming from scratch



- packages provide specific functionalities
- packages need to be installed first

NLP Packages

- **spaCy**

industrial-strength Natural Language Processing (NLP)

- **textaCy**

NLP, before and after spaCy

- **scattertext**

beautiful visualizations of how language differs across corpora

Mini-Project

present project on 2 June 2023

- analyze any collection of documents
- apply quantitative measures + interpretation
 - compare historically
 - compare between actors
- form groups of 2-4 people

In-class: Exercises I

1. Open the script with the basics of Python in your Visual Studio Editor:
`materials/code/python_basics.ipynb`
2. Try to understand the code in each cell and run them by clicking the play symbol left to them. Check the output. Modify some code as well as data and see how the output changes. Initially, this try-and-error is good strategy to learn. Some ideas:

Combine a string and an integer variable without converting it. What error do you get? How can you avoid it?

Select `is a` from the list using the right index.

In-class: Exercises II

1. Write a Python script that

takes text (a string)

splits it into words (a list)

iterates over all the tokens and print all tokens that are longer than 5 characters

Bonus: wrap your code in a function.

2. Go to the next slide. Start with some of the great interactive exercises out there in the web.

Resources

learn basics interactively

- Python Principles
- LearnPython

official Python introduction

- Python introduction



Questions?

