

# The ABC of Computational Text Analysis

## #3 SETTING UP YOUR DEVELOPMENT ENVIRONMENT

Alex Flückiger

Faculty of Humanities and Social Sciences  
University of Lucerne

March 17, 2023

# Outline

- research workflow 
- interacting with computers 
- installation of programs 

# Recap last Lecture

- textual data is challenging
- machine-readable data

A large pile of colorful puzzle pieces, mostly grey, yellow, and orange, scattered across the entire background.

# The Zen of Organizing

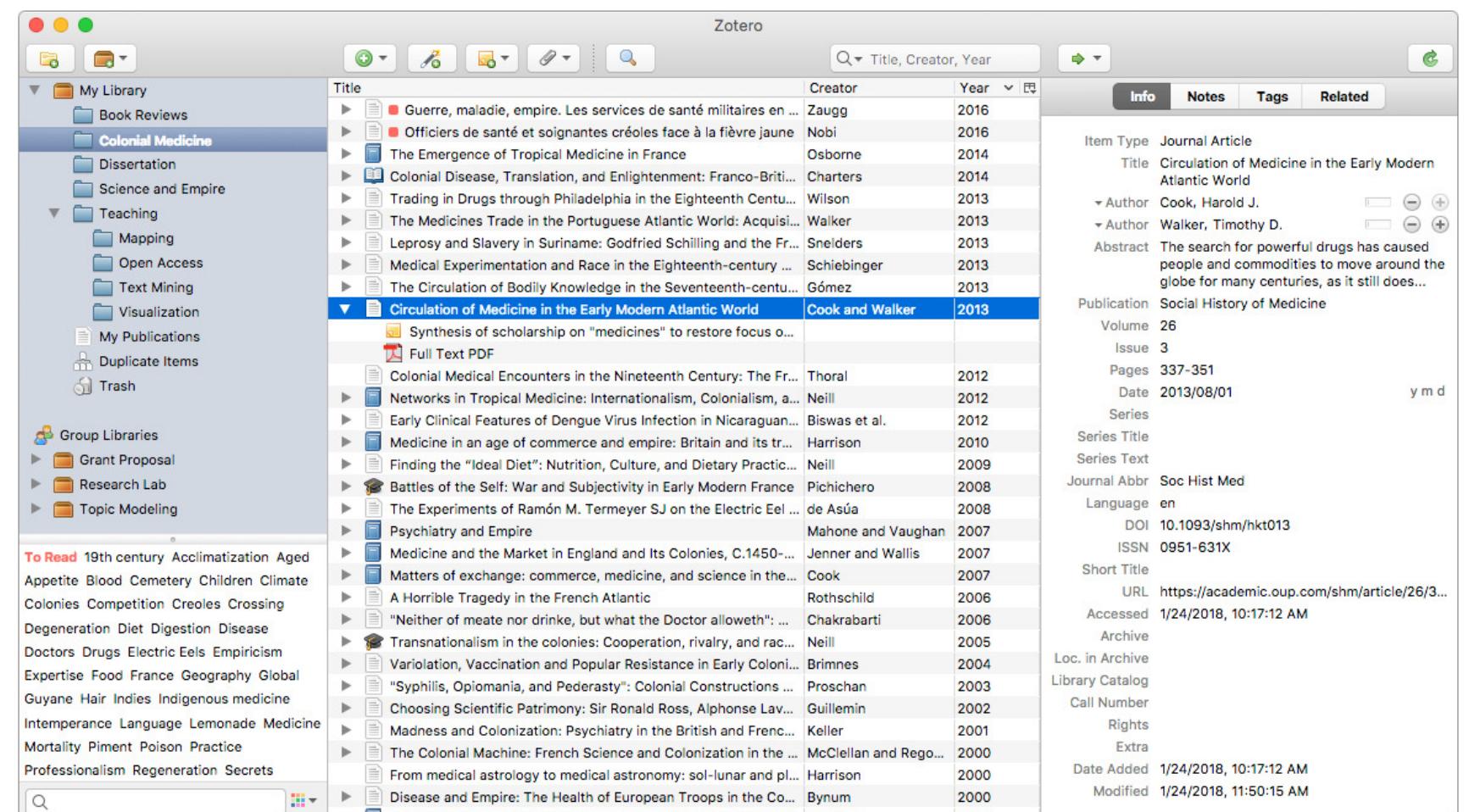
# Research means Organizing

## How a computational approach helps

- scripts as documentation 
- efficient automation   
“don’t repeat yourself”
- less error-prone 
- reproducible 

# Organizing Literature

- manage literature in one place
- generating bibliographies
- collect with a click in your browser



*Zotero is a free, open-source software for managing scientific literature*

# Computers

# Two Trends in Computing

- everyday, mobile, cloud-centered, touch/click-based approach
  - single-purpose application
  - smartphones + tablets
- engineering, open-source, machine-oriented approach
  - powerful due to modularity
  - workflow for data science

(see also Healy 2019)



*Apple Original  
iPhone (2007)*

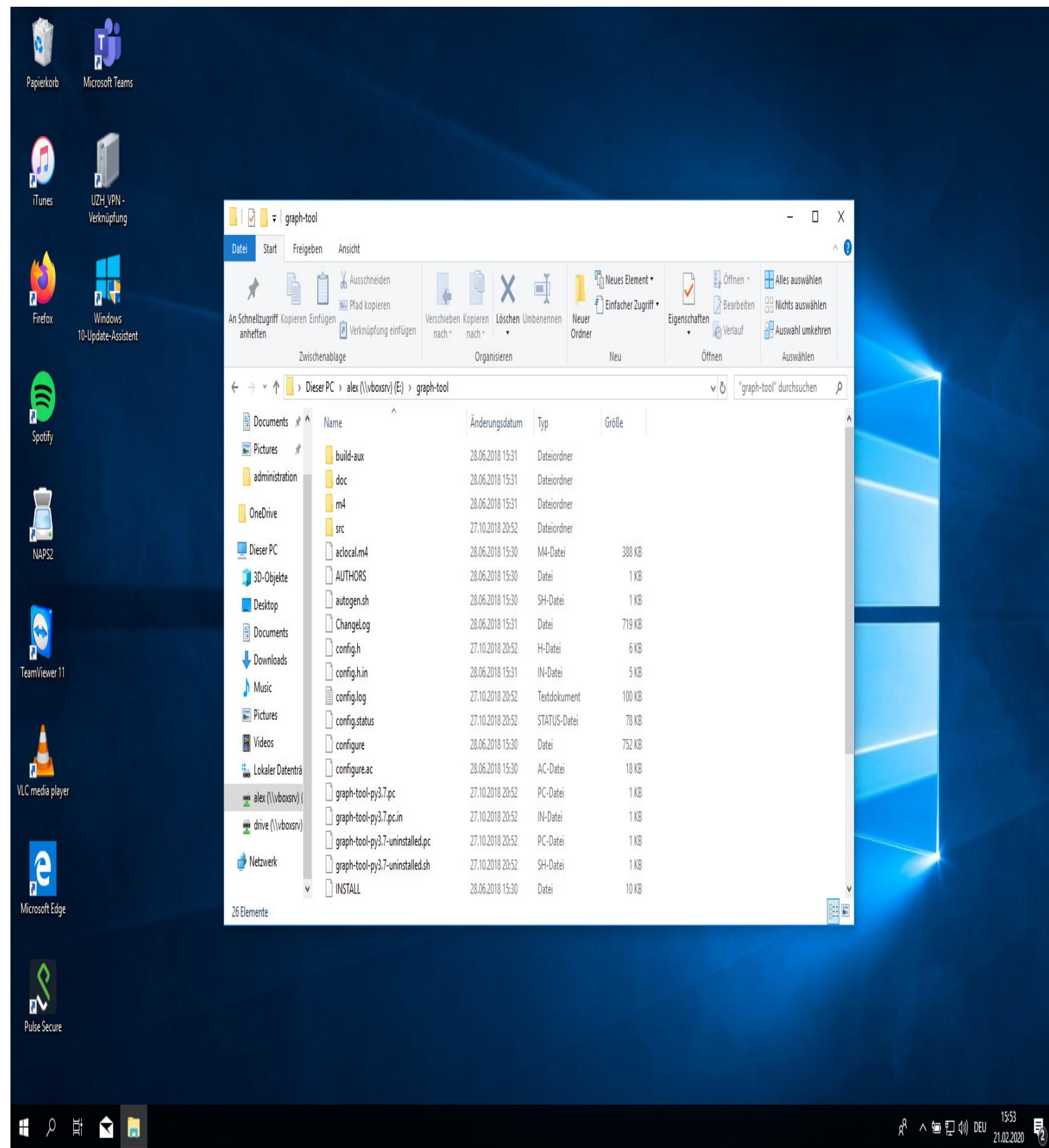
# Operating Systems (OS)

Between hardware and programs

- Microsoft Windows
- macOS
- Linux-based systems 😎

# User Interfaces

## graphical user interface (GUI)



## command line interface (CLI)

```
File Edit View Bookmarks Settings Help  
[alex@alex-pc graph-tool]$ ls -lh  
total 2.6M  
-rw-r--r-- 1 alex alex 388K 28. Jun 2018 aclocal.m4  
-rw-r--r-- 1 alex alex 73 28. Jun 2018 AUTHORS  
-rwxr-xr-x 1 alex alex 963 28. Jun 2018 autogen.sh  
drwxr-xr-x 2 alex alex 4.0K 28. Jun 2018 build-aux  
-rw-r--r-- 1 alex alex 719K 28. Jun 2018 ChangeLog  
-rw-r--r-- 1 alex alex 5.4K 27. Okt 2018 config.h  
-rw-r--r-- 1 alex alex 4.4K 28. Jun 2018 config.h.in  
-rw-r--r-- 1 alex alex 100K 27. Okt 2018 config.log  
-rwxr-xr-x 1 alex alex 78K 27. Okt 2018 config.status  
-rwxr-xr-x 1 alex alex 752K 28. Jun 2018 configure  
-rw-r--r-- 1 alex alex 18K 28. Jun 2018 configure.ac  
drwxr-xr-x 6 alex alex 4.0K 28. Jun 2018 doc  
-rw-r--r-- 1 alex alex 692 27. Okt 2018 graph-tool-py3.7.pc  
-rw-r--r-- 1 alex alex 451 27. Okt 2018 graph-tool-py3.7.pc.in  
-rw-r--r-- 1 alex alex 711 27. Okt 2018 graph-tool-py3.7-uninstalled.pc  
-rw-r--r-- 1 alex alex 1017 27. Okt 2018 graph-tool-py3.7-uninstalled.sh  
-rw-r--r-- 1 alex alex 9.3K 28. Jun 2018 INSTALL  
-rwxr-xr-x 1 alex alex 344K 27. Okt 2018 libtool  
-rw-r--r-- 1 alex alex 35K 28. Jun 2018 LICENSE  
drwxr-xr-x 2 alex alex 4.0K 28. Jun 2018 m4  
-rw-r--r-- 1 alex alex 34K 27. Okt 2018 Makefile  
-rw-r--r-- 1 alex alex 1.5K 28. Jun 2018 Makefile.am  
-rw-r--r-- 1 alex alex 33K 28. Jun 2018 Makefile.in  
-rw-r--r-- 1 alex alex 2.5K 28. Jun 2018 README.md  
drwxr-xr-x 5 alex alex 4.0K 27. Okt 2018 src  
-rw-r--r-- 1 alex alex 23 27. Okt 2018 stamp-h1  
[alex@alex-pc graph-tool]$
```

1

“An awesome Programmer saves the  
World”

8

# Human-Machine Interaction

## How to tell the computer efficiently

- text commands instead of clicks
- similar to human languages
  - syntax (form)
  - semantics (meaning)
- computers are literalists

# Programming

- command languages
  - Bourne-Again shell (Bash), since 1989
- programming languages
  - Python, R, Java, C++ etc.
- instructions → program  $\equiv$  algorithm
  - cooking

# Package Manager

- automation of software installation
  - system: apt, Homebrew
  - Python: pip, conda
- software dependencies
- similar to app stores

# Open-Source is a Mindset

## Standing on the shoulders of giants

- free + open software
- cross-platform portability
- collaboration
  - share + reuse
- helpful community

# Resources everyone is using

- Stack Overflow
  - where you get answers to technical questions
- GitHub
  - where you find open software

Learning by doing, doing by Googleing.



# **Set up your System**

# Backup



You don't need it until you desperately need it!

# Setting up your Development Environment

## Installation

- Python 3.8
  - packages for NLP + visualization
- Tesseract
- various Bash tools
- VS Code Editor

👉 Follow the [installation guide](#) for your OS.

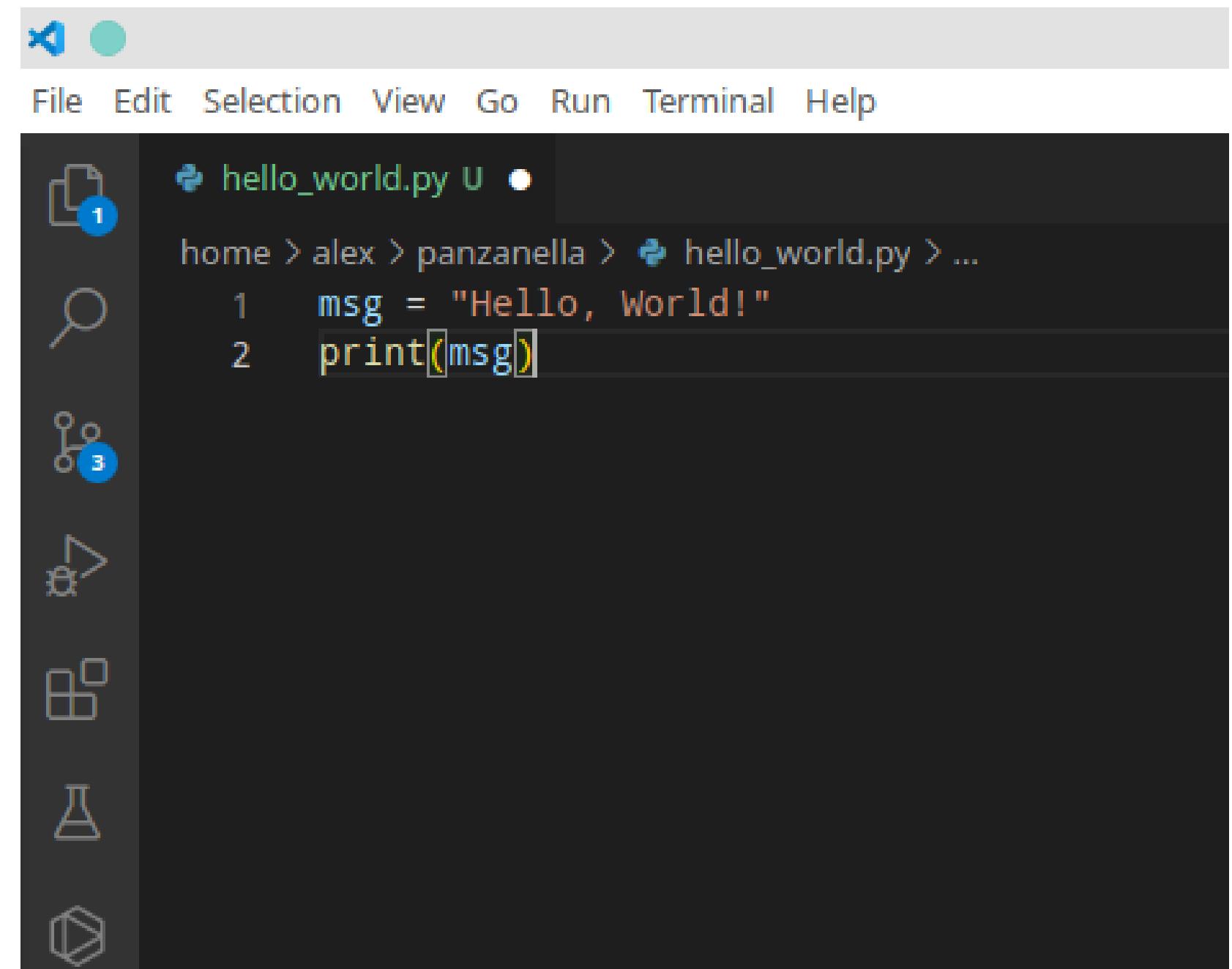


**Relax.  
It takes a while.**

# VS Code Editor

## The Microsoft Word for coding

- powerful integrated development environment (IDE)
- cross-platform
- interactive analysis



A screenshot of the Visual Studio Code (VS Code) interface. The window title is "VS Code". The menu bar includes File, Edit, Selection, View, Go, Run, Terminal, and Help. The left sidebar has icons for file explorer (with 1), search (with 0), symbols (with 3), and other developer tools. The main editor area shows a Python script named "hello\_world.py":

```
home > alex > panzanella > hello_world.py > ...
1 msg = "Hello, World!"
2 print(msg)
```

*Write your first Python script in VS Code*

# First Steps in Python

Learn Python with interactive tutorials

- [Python Principles](#)
- [LearnPython](#)

# Readings

**optional: pimp your workflow**

Healy, Kieran. 2019. “The Plain Person’s Guide to Plain Text Social Science.”  
online.



Questions?

# References

Healy, Kieran. 2019. "The Plain Person's Guide to Plain Text Social Science."  
<https://kieranhealy.org/publications/plain-person-text/>.