efl Data Science Course

Introduction

About us: Who we are

efinancelab
at the HOUSE OF FINANCE

efl
the Data Science Institute

Original Mission:

- Investigate and co-shape Digital Finance 2.0
  - Web-based selfservices of customers
- Research was performed in three different Layers:
  - Customers in E-Finance
  - E-Financial Markets and Market Infrastructures
  - IT Infrastructures: Service Systems in E-Finance

New Mission (Since 2019):

- Use expertise in Data Science to deliver cutting edge research in the fields of
  - Financial Services
  - Retail & Marketing
  - Health
  - Law
  - General, cross-sectional research

About us: The elf

Industry - Academic Partnership

Universities

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

TECHNISCHE UNIVERSITÄT DARMSTADT

Sponsors

FACTSET

AMERICAN EXPRESS

Deutsche Leasing

finanz informatik

DEUTSCHE BÖRSE GROUP

DZ BANK Gruppe

About us: Lecturers

M.Sc. Tino Cestonaro

Research Assistant

Tino Cestonaro joined the team in April 2020. His research focuses on Market Microstructure, Financial Machine Learning,...

M.Sc. Micha Bender

Research Assistant

Micha Bender focuses on empirical financial market research. Furthermore, he is interested in the application of machine...

M.Sc. Johannes Chen

Research Assistant

Johannes Chen studied Business Informatics at the Technical University of Darmstadt. During his studies, he focused on...

Dr. Benjamin M. Abdel-Karim

Alumnus

Since March 2018, Benjamin M. Abdel-Karim is a research assistant of Prof. Dr. Oliver Hinz at the Chair of Information...

Agenda

Day 1 (Python Course) (30.10.2023)

9:00 - 10:30 Uhr
Python Basics
Introduction and Primitive Data Types

10:40 - 12:10 Uhr
Data Structures
Lists, Sets, Dictionaries

13:30 - 15:00 Uhr
Control Structures
Loops (for, while), case distinction (if, else)

15:15 - 16:45 Uhr
Functions
Structure of Functions and Application

Day 2 (Python Course) (01.11.2023)

9:00 - 10:30 Uhr
Helpful functions for data processing
Libraries: os, re, csv

10:40 - 12:10 Uhr
Data types and data structures
Libraries: numpy, pandas

13:30 - 15:00 Uhr
Data import and visualization
Libraries: csv (cont'd), matplotlib

15:15 - 16:45 Uhr
Outlook: Data Science
Exemplary implementation of a KDD process

Day 3 (Data Science) (03.11.2023)

9:00 - 10:30 Uhr
Introduction to Data Science
Terminology and basic concepts

10:40 - 12:10 Uhr
Working with Data
Preprocessing, explorative data analysis

13:30 - 15:00 Uhr
Data Analysis I
Classification

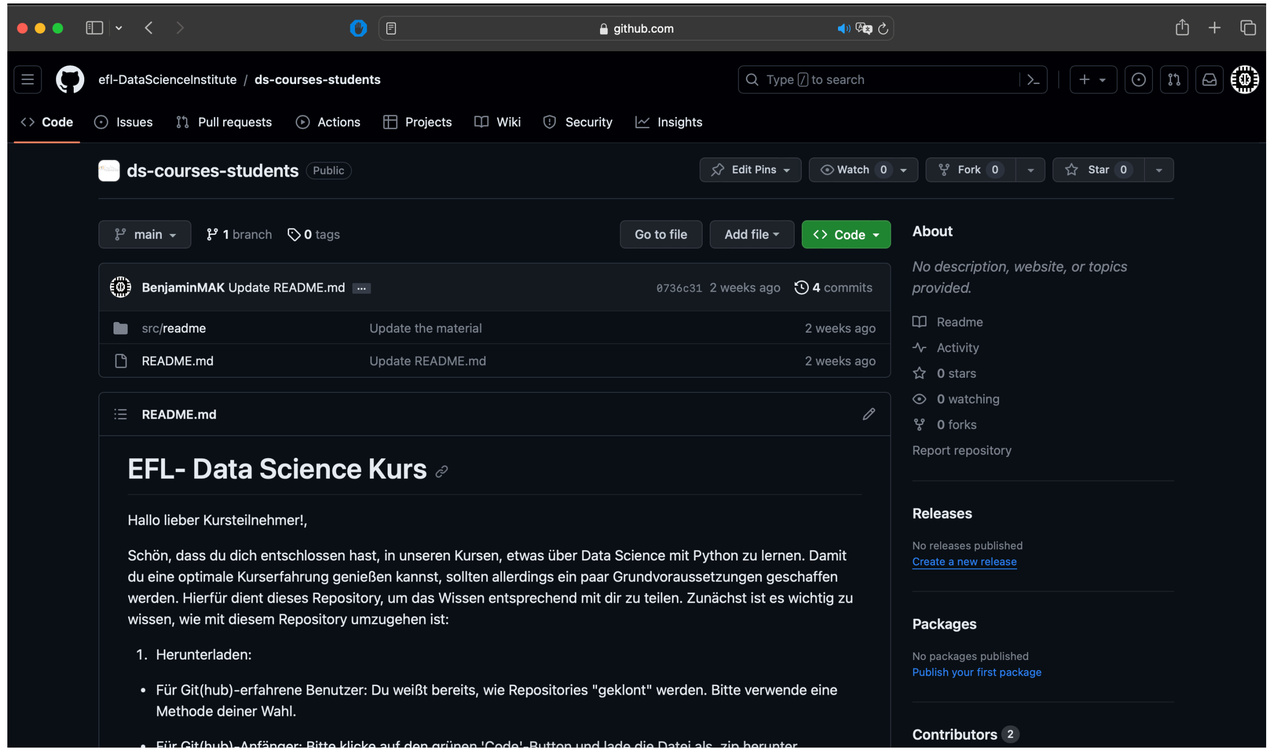
15:15 - 16:45 Uhr
Data Analysis II
Neural Networks

https://www.eflab.de/teaching

Certificates

- Completion of tasks from the last class
- Submission: Description of how the tasks were solved
- Formalities:
  - Min. 2 pages (Arial 11, 1.5 Line spacing, 3 CM Correction margin)
  - Code must be delivered separately (code folder)

Course Material?



https://github.com/efl-DataScienceInstitute/ds-courses-students

Why you are here

Kenntnis in Python und anderen Softwareprogrammen werden auf dem heutigen Arbeits- und Praktikumsmarkt oft vorausgesetzt. Um auch über die in den Vorlesung vermittelten Inhalte hinaus etwas
über statistische Softwareprogramme zulernen, möchte ich an dem Kurs teilnehmen.

Ich möchte mich in diesem Bereich fortbilden, um bei einer Bachelor-Arbeit Daten auszuwerten und zu visualisieren. Darüberhinaus möchte ich wahrscheinlich einen quantitativ-orientierten Master belegen und halte es daher für sinnvoll einen ersten Python-Kurs zu belegen.

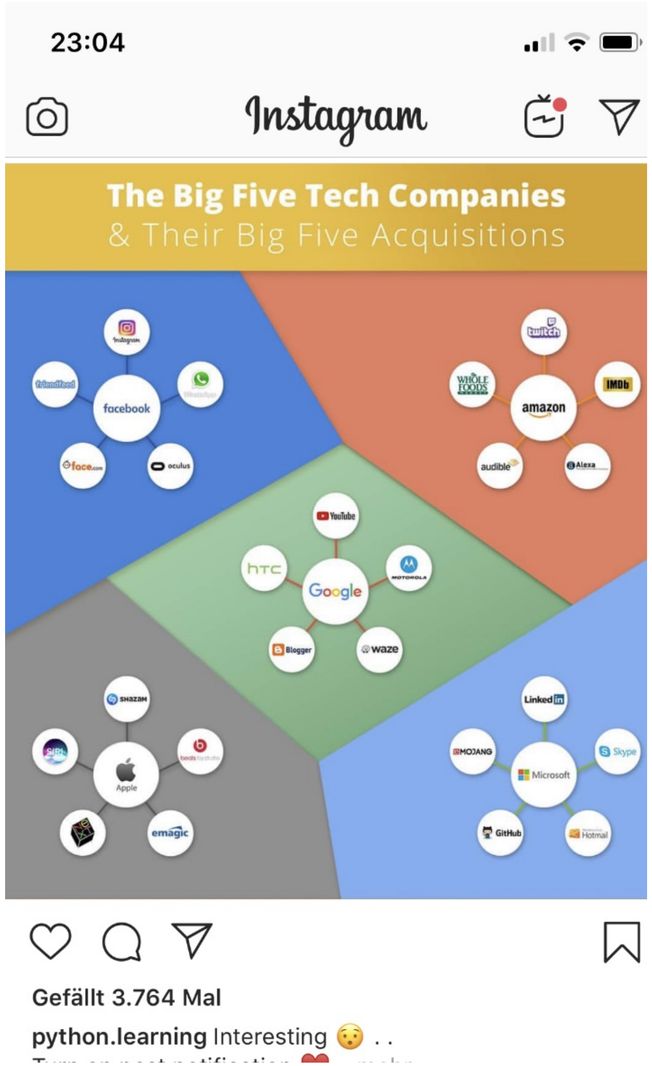
This semester I am taking "Business Informatics" with Professor Dr. Peter Gomber, which brought you to my attention.

Since the importance of Data Science as a tool for business decision making has already been discussed in class, an interest has already been sparked here. Now I hope to gain a deeper understanding through your course.

Auffrischen von Python, neue Einblicke in Software

Bei meiner aktuellen Werksstudentenstelle merke ich immer wieder, wie Entscheidungen sehr oft datengetrieben werden und welche Rolle gut aufgearbeitete Daten spielen. Daher würde ich gerne, um mich beruflich und persönlich weiterzubilden, gerne an dem Kurs teilnehmen.

Why Coding?



NACH ABSCHLUSS IN ...

INFORMATIK, MATHEMATIK, WIRTSCHAFTSINFORMATIK

	TOP BERUFSTITEL*	TOP BRANCHE
1	IT Manager **92.275 €**	Banken **85.067 €**
2	IT-Projektmanager/in **86.695 €**	Chemie- und Erdöl-verarbeitende Industrie **84.033 €**
3	Senior Developer **78.635 €**	Konsum- und Gebrauchsgüter **82.586 €**
4	Analyst **74.469 €**	Unternehmensberatung, Wirtschaftsprüfung und Recht **82.336 €**
5	Software-Architekt/in **74.152 €**	Finanzdienstleister **81.331 €**

*Berufstitel ohne Management- & Personalverantwortung
Bruttodurchschnittsgehalt (inkl. variabler Bezüge), Mittelwert

11 Stepstones - Gehaltsreport

What is coding?

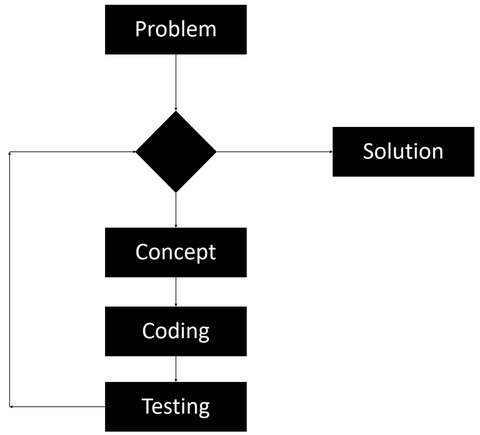
WIKIPEDIA
The Free Encyclopedia

Computer programming

Article Talk

From Wikipedia, the free encyclopedia

Computer programming or **coding** is the composition of sequences of instructions, called programs, that computers can follow to perform tasks.[1][2] It involves designing and implementing algorithms, step-by-step specifications of procedures, by writing code in one or more programming languages. Programmers typically use high-level programming languages that are more easily intelligible to humans than machine code, which is directly executed by the central processing unit. Proficient programming usually requires expertise in several different subjects, including knowledge of the application domain, details of programming languages and generic code libraries, specialized algorithms, and formal logic.



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

Motivation: Why Python?

- Released in 1991 by Guido van Rossum
- With the explosive growth of ‘big data’ in disciplines such as bioinformatics, neuroscience and astronomy, programming know-how is becoming ever more crucial (Perkel 2015, p. 125).

YouTube [2]

[3]

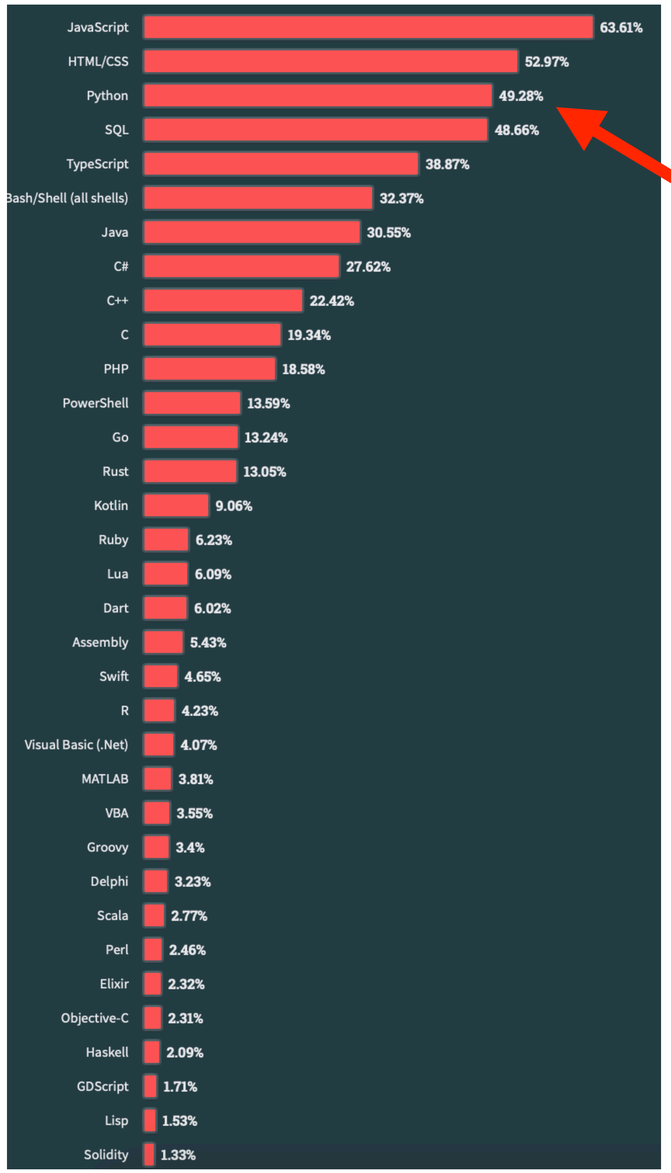
[6]

Dropbox

[8]

Spotify

Guido van Rossum



https://survey.stackoverflow.co/2023/

Motivation: The Elementary Basics in Python

Pros Python:

- General purpose language
- Higher programming language
- It´s simple
- Fast to read
- Structuring by indenting
- No {} or ; => Faster to Code
- Data types are managed dynamically. There is no static type check like in java
- Widespread in science
- Extensive Support Libraries (important data science, math and many more)
- Integration Feature
- Productivity (Many Frameworks such as unit testing)

Ozgur, C., Colliau, T., Rogers, G., & Hughes, Z. (2017). MatLab vs. Python vs. R. Journal of data Science, 15(3), 355-371.
Abdel-Karim, B. M. (2022). Data Science. In Data Science: Best Practices mit Python (pp. 57-62). Wiesbaden: Springer Fachmedien Wiesbaden.

Cons of Python

Some Points:

- Python is an interpreted language, this is slow as compared to C/C++
- See Interpreted vs. Compiled programming languages
- Python can have runtime errors (dynamical typing feature)
- Consumes a lot of memory space
- Not easy to test

Small Satellite*

*) https://www.nbcnews.com/id/wbna43313086

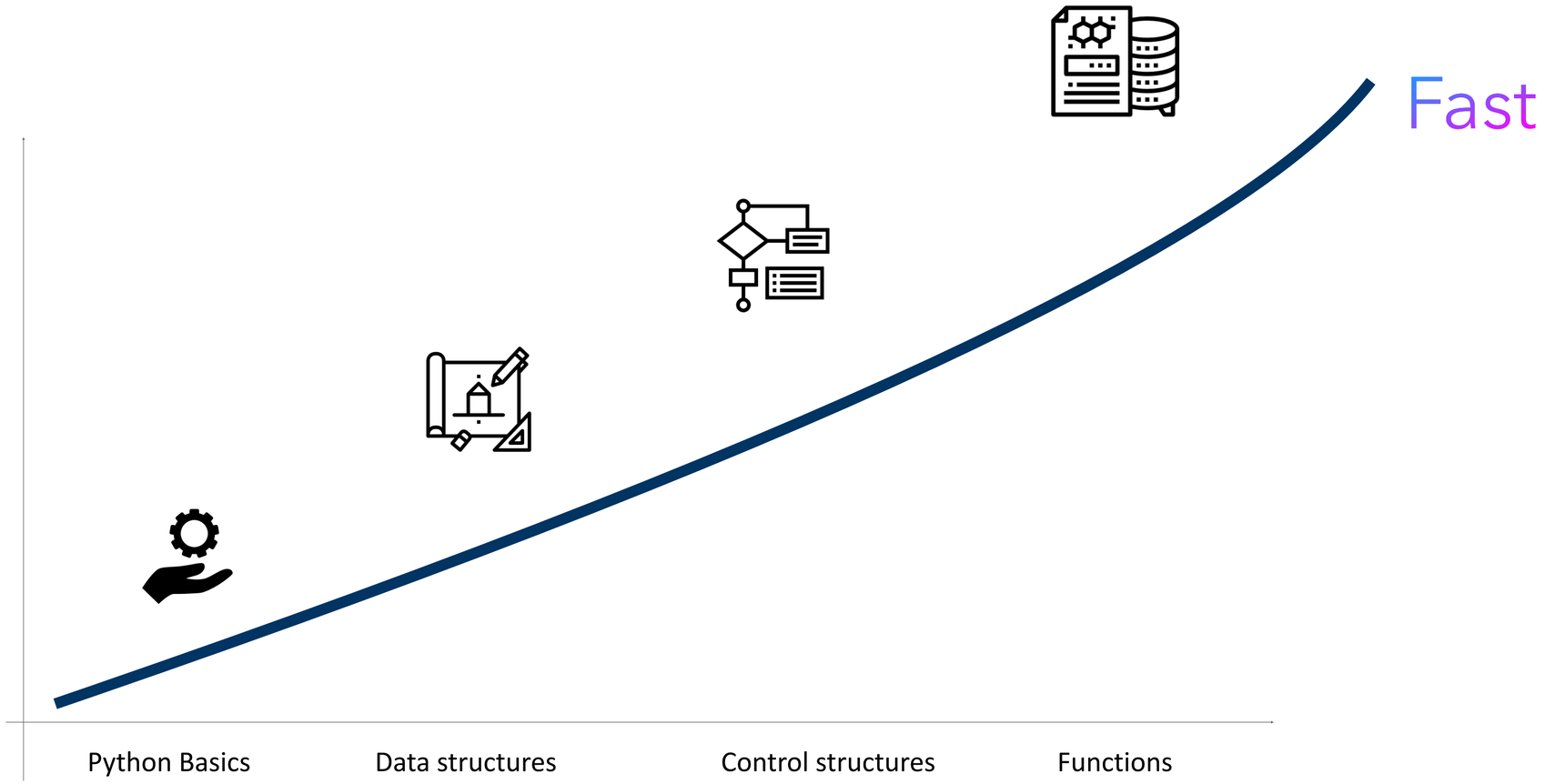
Not in Python …

Prechelt, L. (2000). An empirical comparison of c, c++, java, perl, python, rexx and tcl. *IEEE Computer*, *33*(10), 23-29.
Newhall, T. K. (1999). Performance Measurement of Interpreted, Just-in-Time compiled, and Dynamically Compiled Executions. The University of Wisconsin-Madison.

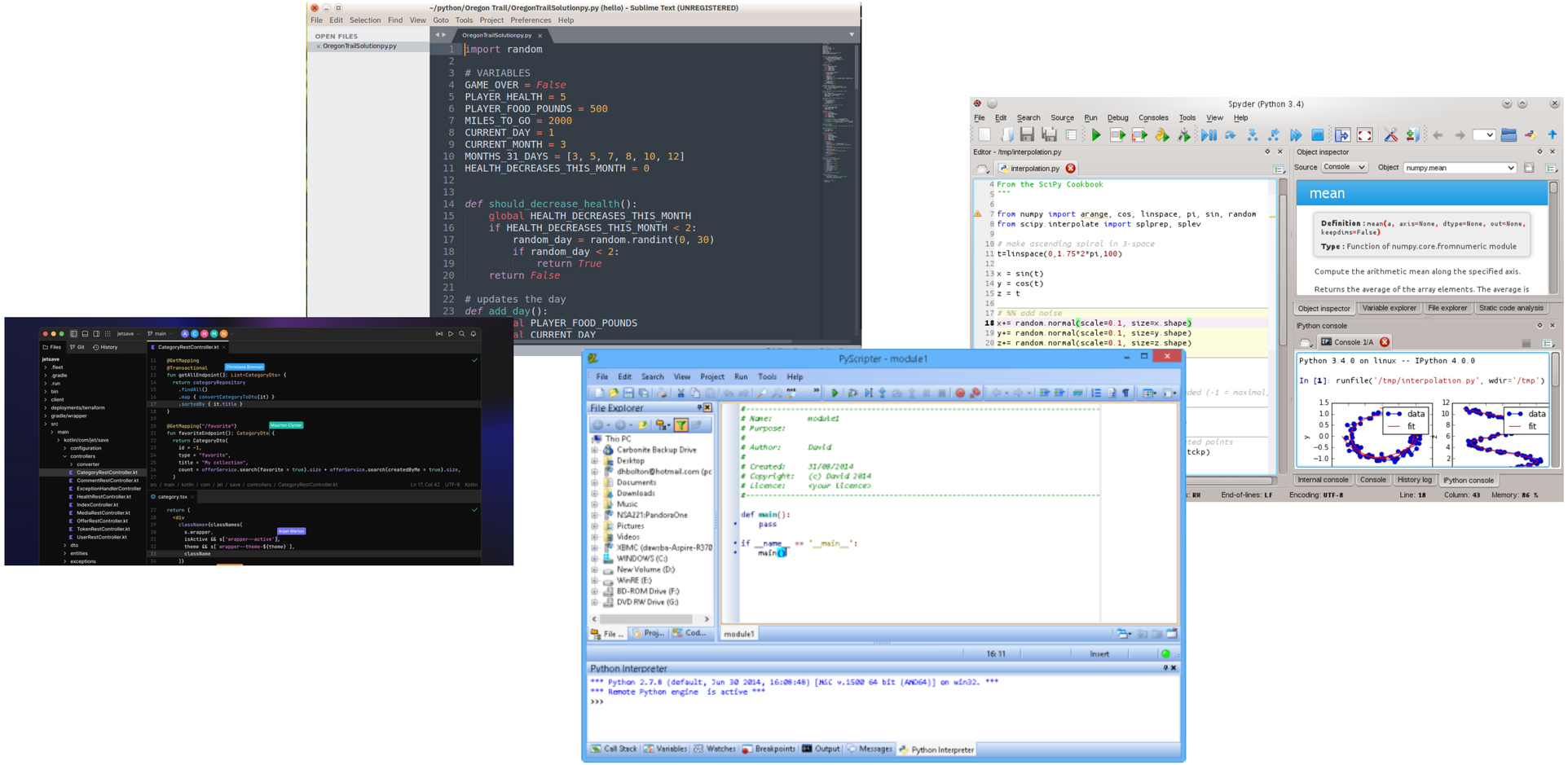
Agreement

- Introduction to programming => All Questions are allowed!!!
- For beginners
- The module is interactive => Use your computer
- **We develop the solutions together!**
- Please be on time
- There are no dumb questions
- Nobody knows everything
- Copying solutions is plagiarism

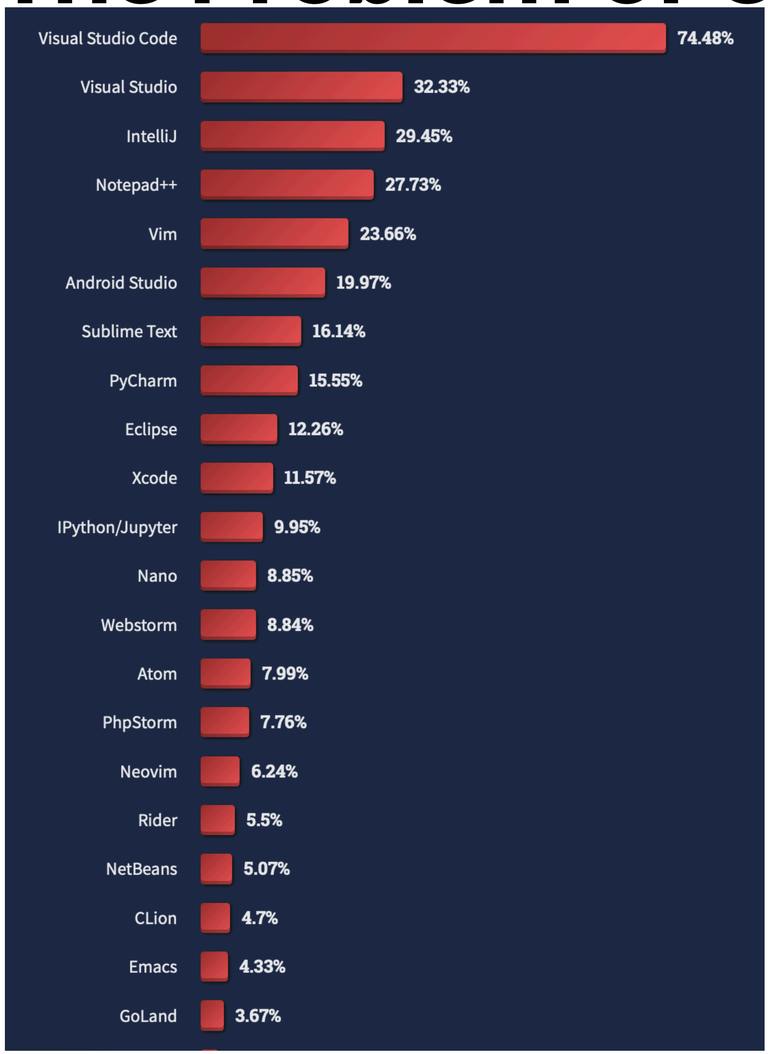
Note: Learning Curve for Today



Integrated Development Environment
IDE a Deeper look? The Problem of Choice

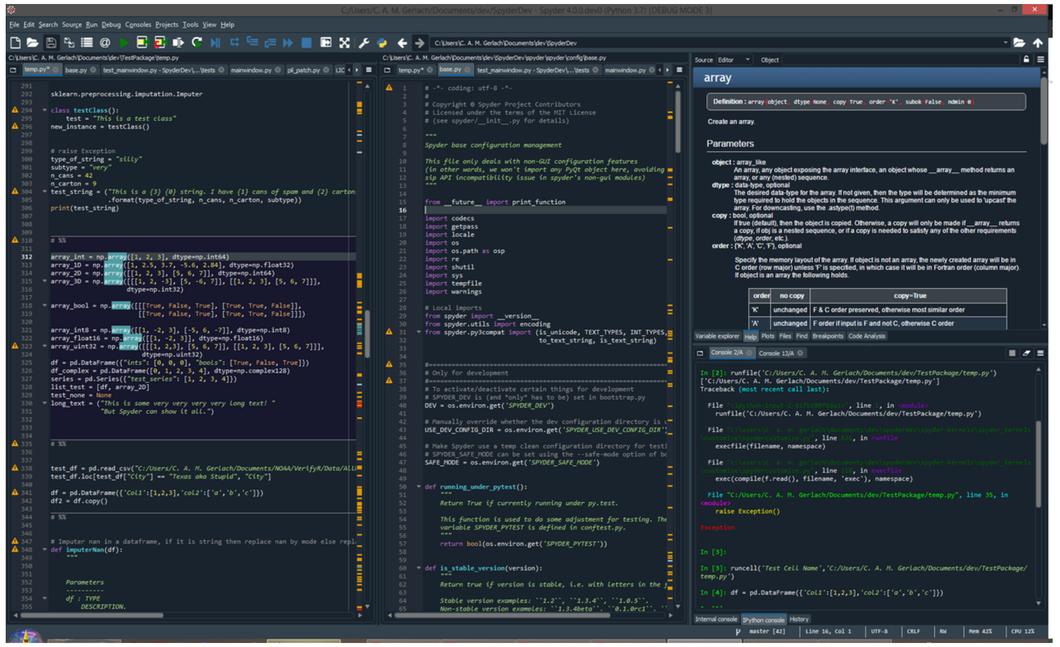


The Problem of Choice



https://survey.stackoverflow.co/2022/#most-popular-technologies-new-collab-tools-prof

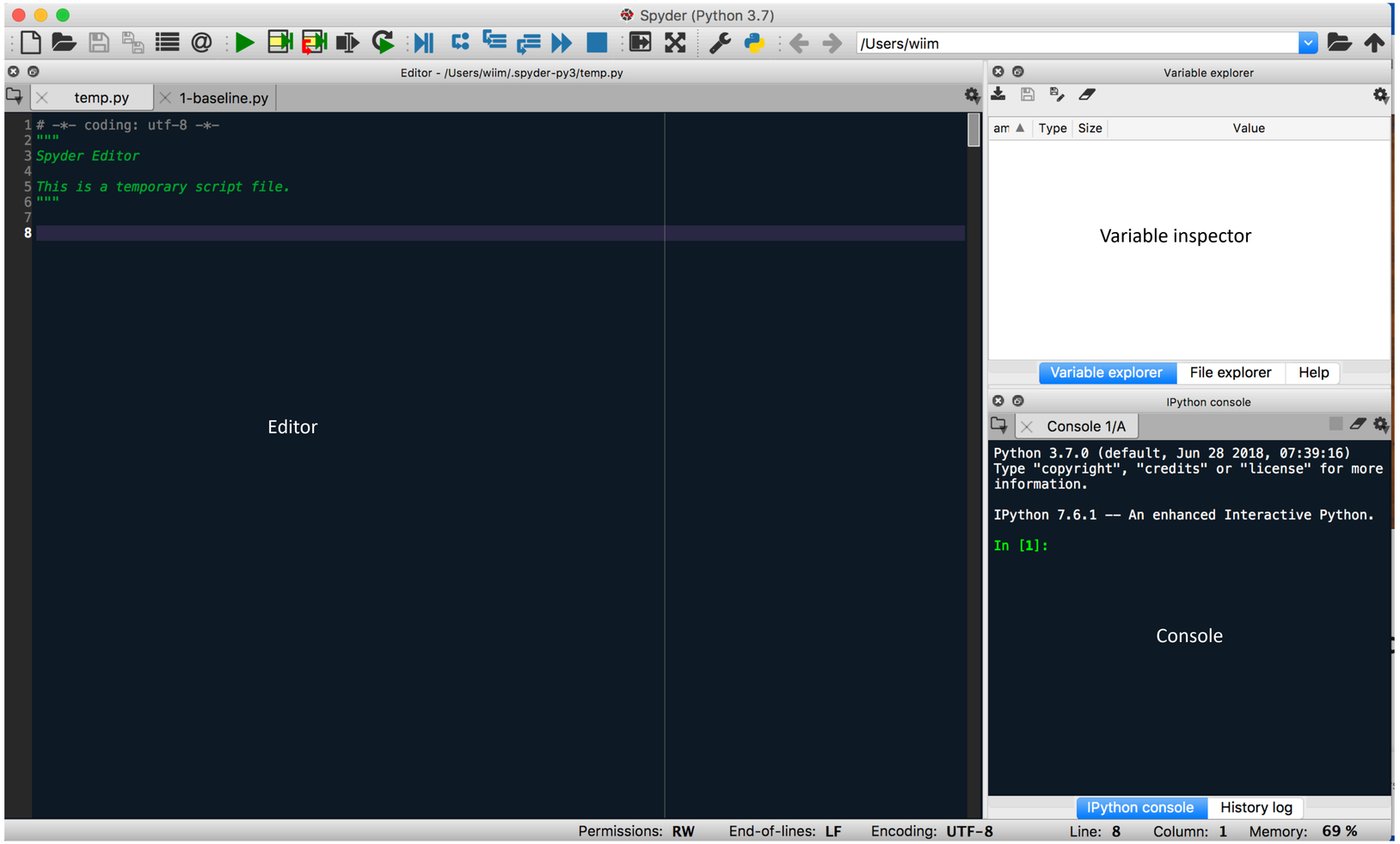
What We Use for This Course



Practice and Questions

- Take your computer and let's get started!
- Create a Folder ‘Day 1’ on our Desktop
- Launch Spyder and select the folder

In a Nutshell:
Integrated Development Environment



In a Nutshell: Hello World

```
print('Hello World')
```

Abstraction (Computer Science)
The first step - The concept of variable

Variable

Specific value

```
int_one = 1
```

Allocation

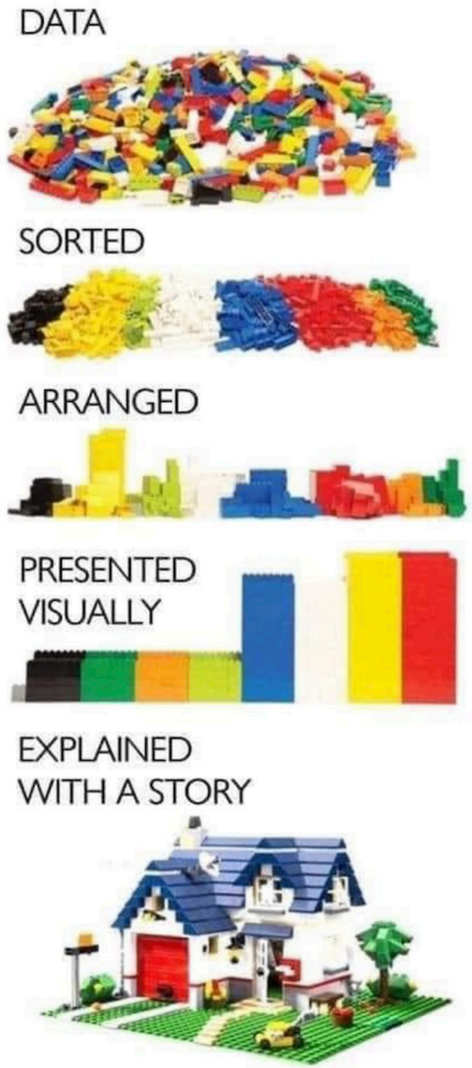
We assign the value 1 to the variable 'one'.
Now we can continue working with the variable 'one'.
Advantage: We are independent of concrete value

The essence of abstractions is preserving information that is relevant in a given context, and forgetting information that is irrelevant in that context.

– John V. Guttag

Different Building Blocks

- Integer: int (signed integers) – They are often just called integers or ints, are positive or negative whole numbers with no decimal point.
- Float: Float (floating point real values, double) – Also called floats, they represent real numbers and are written with a decimal point dividing the integer and fractional parts. Floats may also be in scientific notation, with E or e indicating the power of 10.
- Springs are a sequence of chars. We can create them simply by enclosing characters in quotes. “Hello World” is a String! Therefore strings in Python are bytes representing Unicode characters. In Detail: Python does not have a character data type, a single character is simply a string with a length of 1.
- Boolean: We can perform logical operations with True and False in combination with AND and OR.



Primitive Types

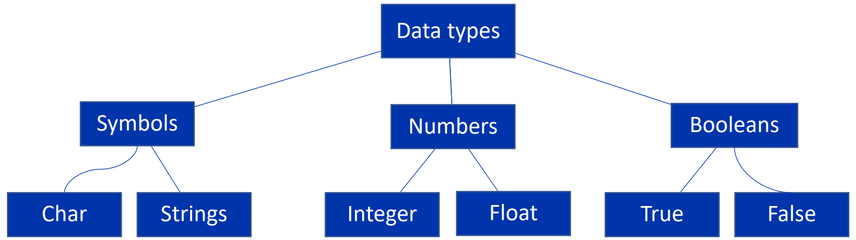
- Integer: int (signed integers) – They are often just called integers or ints, are positive or negative whole numbers with no decimal point.

- Float: Float (floating point real values, double) – Also called floats, they represent real numbers and are written with a decimal point dividing the integer and fractional parts. Floats may also be in scientific notation, with E or e indicating the power of 10.

- Springs are a sequence of chars. We can create them simply by enclosing characters in quotes. “Hello World” is a String! Therefore strings in Python are bytes representing Unicode characters. In Detail: Python does not have a character data type, a single character is simply a string with a length of 1.

- Boolean: We can perform logical operations with True and False in combination with AND and OR.

Primitive Data Types: A first overview



* Simplified Illustration

Python Documentation: https://docs.python.org/3/library/stdtypes.html
Other classification: Chun, W. (2001). Core python programming (Vol. 1). Prentice Hall Professional.

Documentation

Source Code Documentation

@author: name

@since: first implementation date

@version: date of last update

@source: if you using links etc.

@code: special code note

@param: if special parameter is used or you have to describe.

```
# I am a comment
@author: My Name
@since: 2022-10-03
@update: 2023-10-26
@version: v.0.0.2
print('Hello World')
```

More Information: PEP 8 – Style Guide for Python Code: https://peps.python.org/pep-0008/

Naming convention

- Names of attributes, variables, methods start with a small letter
  - may use letters without ß or similar
  - which points to the data type like i, s or l
- This is standard in professional software development.
- Camel Case: Compound words are written in programming language. Every new word is capitalized.

(More PeP8 Style)

- name = Is the name of…
- bscript_main = Simple code file that does something
- CName = Class (later more)
- float_name = Variable that saves a floating point (double)
- int_name = Variable that saves an integer value
- str_name = Variable that saves a string value
- b_name = Boolean for true or false values
- list_name = Object from type list
- method_name = Self-written function

(Camel Case Style)

- Name = Is the name of…
- bScriptName = Simple code file that does something
- CName = Class (later more)
- dName = Variable that saves a floating point (double)
- iName = Variable that saves an integer value
- sName = Variable that saves a string value
- bName = Boolean for true or false values
- LName = Object from type list
- fName = Self-written function

Code is some kind of Art. Therefore...

Jackson Pollock Art

These are all approaches. Find your own style!

Thank You

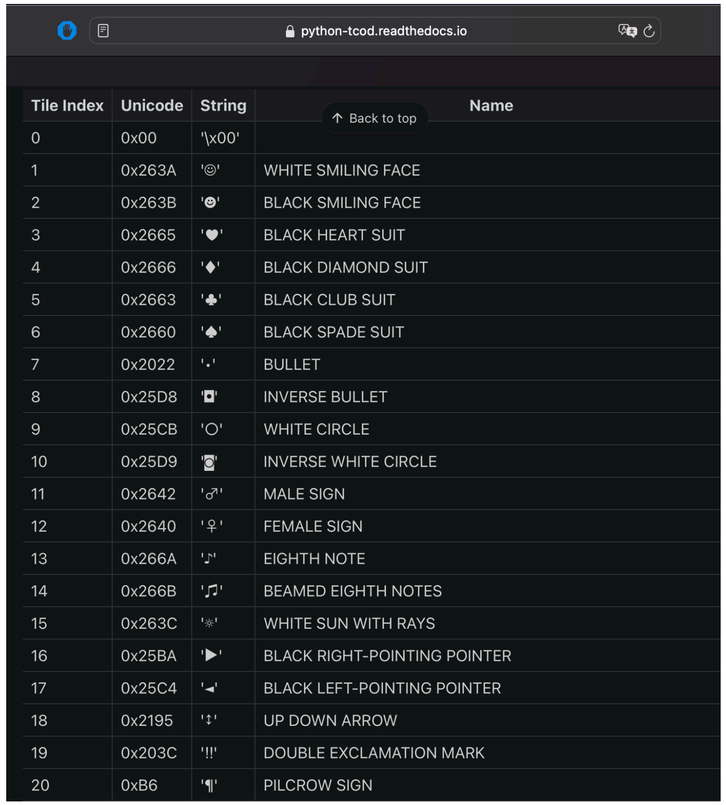

YouTube
Suchen
Data Science in a Nutshell
Dr. Benjamin M. Abdel-Karim
Data Science in a Nutshell - Einleitung
Keine Aufrufe · 17.08.2021
TEILEN
SPEICHERN


Benjamin M. Abdel-Karim Hrsg.
Data Science
Best practices mit Python
EBOOK INSIDE
Springer Vieweg

Appendix

Unicode Table

```
ord(c) inverse chr()
```



Tile Index	Unicode	String	Name
0	0x00	'\x00'	
1	0x263A	'☺'	WHITE SMILING FACE
2	0x263B	'☻'	BLACK SMILING FACE
3	0x2665	'♥'	BLACK HEART SUIT
4	0x2666	'♦'	BLACK DIAMOND SUIT
5	0x2663	'♣'	BLACK CLUB SUIT
6	0x2660	'♠'	BLACK SPADE SUIT
7	0x2022	'•'	BULLET
8	0x25D8	'◘'	INVERSE BULLET
9	0x25CB	'○'	WHITE CIRCLE
10	0x25D9	'◙'	INVERSE WHITE CIRCLE
11	0x2642	'♂'	MALE SIGN
12	0x2640	'♀'	FEMALE SIGN
13	0x266A	'♪'	EIGHTH NOTE
14	0x266B	'♫'	BEAMED EIGHTH NOTES
15	0x263C	'☼'	WHITE SUN WITH RAYS
16	0x25BA	'►'	BLACK RIGHT-POINTING POINTER
17	0x25C4	'◄'	BLACK LEFT-POINTING POINTER
18	0x2195	'↕'	UP DOWN ARROW
19	0x203C	'‼'	DOUBLE EXCLAMATION MARK
20	0xB6	'¶'	PILCROW SIGN

https://python-tcod.readthedocs.io/en/latest/tcod/charmap-reference.html