Unit 0: Before Coming to the School

August 2021

You have registered for the Summer School "Data Science: Programming with Python", and we look very much forward to seeing you soon! To be able to hit the ground running on the first day of the school, we ask you to complete this "Unit 0" beforehand. In particular, make sure that you have the Anaconda Data Science platform installed and running on your computer (follow the instructions below).

We also include some (optional) computational thinking puzzles. They don't teach you programming or Python directly, but are very suitable to practice the kind of thinking that is needed for successful computer programming. If you enjoy them, you will probably also enjoy programming!

The Python Programming Language

Python (yes, indeed named after the British comedy group Monty Python (https://en.wikipedia.org/wiki/Monty_Python)) is one of the most popular programming languages today. It has been released for the first time already back in 1990, but gained extreme popularity only in the last years, hand in hand with the increasing importance of the world wide web, big data and data science.

Although older versions (in particular the second generation) are still operational, we will use **Python 3** in the course to make full use of the features of the latest generation. As of summer 2021, the latest stable release of Python has number 3.9.1.

There is a lot of free literature about Python available that you can use for the course in addition to the lecture notes that we provide you with. Especially if you have difficulties understanding a particular concept, it is often a good idea to look at alternative explanations. What works good for the one, might just not be the best way to put it for the other. Here are some links to useful Python online books, but please feel free to check out also other sources of information:

- https://python.swaroopch.com (https://python.swaroopch.com) (“A Byte of Python”, especially for beginners)
- http://greenteapress.com/wp/think-python-2e/ (http://greenteapress.com/wp/think-python-2e/) (“Think Python”, also targeted at beginners)
- https://docs.python.org/3/tutorial/index.html (https://docs.python.org/3/tutorial/index.html) (the official Python tutorial)

Anaconda Setup

During the course we will work with the Anaconda Python Data Science Platform (official website: https://www.anaconda.com/ (https://www.anaconda.com/)), in particular we will use the Spyder IDE (Integrated Development Environment) and Jupyter notebooks. Anaconda is a free and open source distribution of Python and some other programming languages used in scientific applications. It runs on all major operating systems and provides a number of very useful tools for Python programming.

Here is what you have to do to get ready for the course:

1. Go to https://docs.anaconda.com/anaconda/install/ (https://docs.anaconda.com/anaconda/install/) and follow the installation instructions for your operating system.

2. Go to https://docs.anaconda.com/anaconda/user-guide/getting-started/#open-nav-lin (https://docs.anaconda.com/anaconda/user-guide/getting-started/#open-nav-lin) and follow the instructions how to start Spyder.
3. Ready to start coding!

If you have problems with installing Anacaonda, please get in touch with the course instructors as soon as possible.

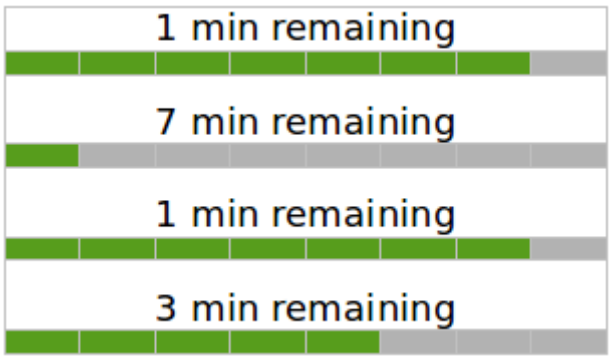
Optional: Computational Thinking Exercises

To get warmed for the summer school, here some puzzles (inspired by and adapted from the GI's Informatik-Biber (bwinf.de/biber)) that require you to use different computational thinking skills. No programming experience needed, just read the instructions carefully and set your brain to work! :)

1. Download Times

When downloading multiple files from the internet at the same time, the downloads share the capacity of the connection between them, each getting the same share. For example, when downloading 10 files at the same time, each of the downloads can use 1/10 of the capacity. Downloading a single file would be ten times as fast.

Now consider the following situation (illustrated below): A user is downloading 4 files at the same time. For each file the remaining download time is displayed. This time is calculated based on the capacity of the internet connection (which stays the same all the time). How many minutes will it take until all 4 files have finished downloading?



2. Maximal Gain

The grid below contains numbers and two special cells S (for start) and G (for goal). The task: Find the path from S to G on which the sum of the numbers on the visited fields is the highest. But: you are only allowed to step upwards and to the right (not down or to the left).

What is the maximal sum that can be gained this way, and through which path?

2	0	1	1	**G**
1	2	0	2	3
2	2	0	2	1
3	1	0	2	0
S	0	1	3	5

3. Number Segments

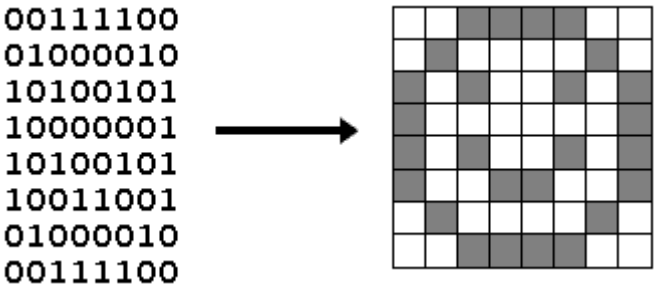
In a house with 10 floors the lift shows the current floor with a 7-segment digital number display. The numbers 0 (for ground floor) to 9 look like this:

Thus, to display a number, at least 2 and at most 7 of the segments are switched on.

Recently one of the segments was broken and did not switch on at all. Nevertheless it was possible to recognize all numbers and tell them apart. Which segment could have been the broken one?

4. Quarters

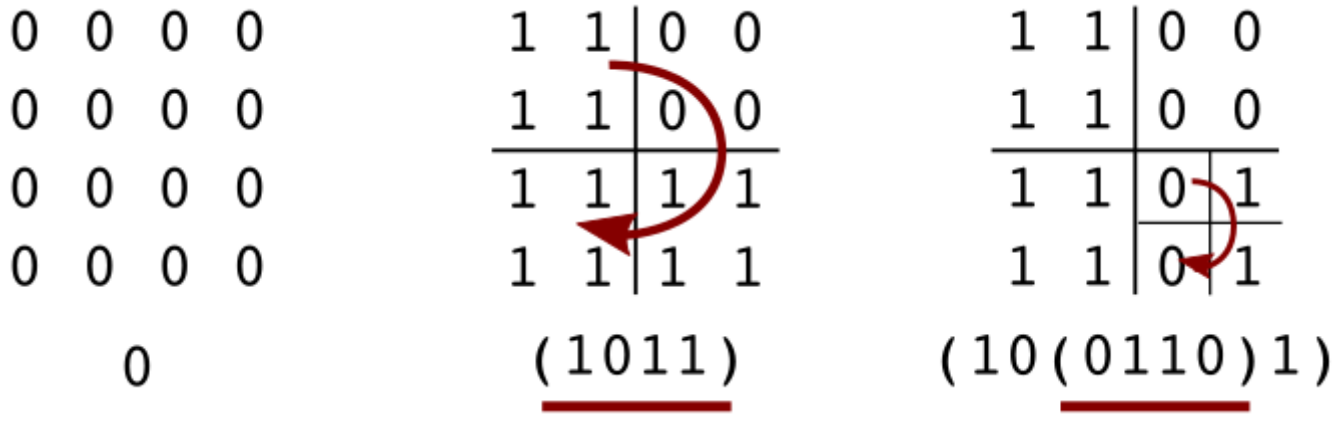
Black-and-white bitmaps can be represented with the binary characters 0 and 1, as illustrated below (picture from http://livecode.byu.edu/images/DigitalImages.php (http://livecode.byu.edu/images/DigitalImages.php)): A 0 stands for a white pixel, and a 1 for a black pixel. A picture with 4 x 4 pixels can thus be encoded with 16 characters, and a picture with 8 x 8 pixels with 64 characters.



Many pictures can however also be represented with less characters. Therefore the characters are arranged in a square raster, and the following procedure is applied:

- If all characters in the raster are 0, the result is the character 0.
- If all characters in the raster are 1, the result is the character 1.
- Otherwise the raster is split into 4 rasters of same size, and the procedure is applied to all of these rasters, starting with the upper left and proceeding in clock-wise direction. The result is then created as follows: The four results are written after one another, between the brackets (and).

Here are three examples of how the procedure works on a 4 x 4 raster:



What is the result for the 8 x 8 raster below?

```
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 0 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
```

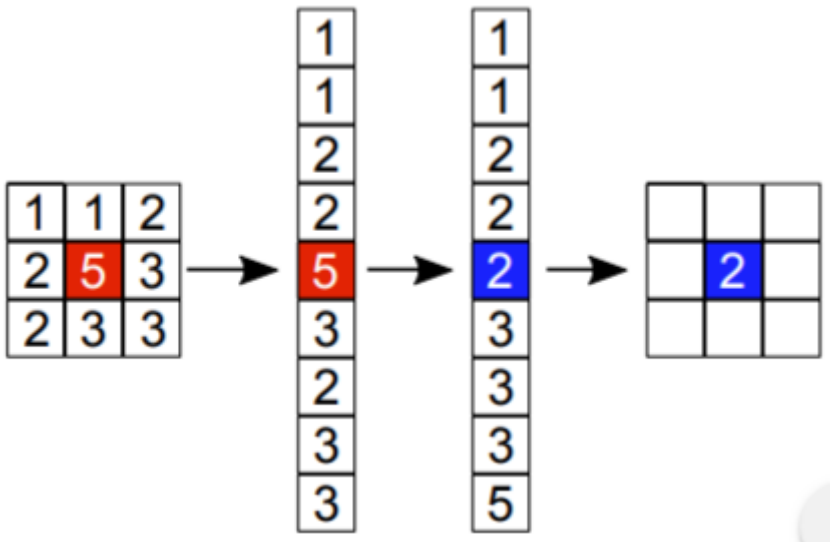
Extra question: How many characters would the procedure save when applied on the smiley bitmap from above?

5. Filter

A gray-scale bitmap can be saved as a table or grid, with values of 1 to 5 for each pixel. The value 1 stands for black, the value 5 stands for white, and the values 2 to 4 stands for the lighter-getting shades of gray between them.

A filtering algorithm computes a picture of same size for a given original according to the following procedure (illustrated on an example below):

For each pixel in the original picture, it takes the value of the pixel and its eight neighboring pixels, sorts them from smallest to largest, and then takes the value in the middle, i.e. the 5th of 9 values, as the new value of the pixel in the filtered picture.



When applied to the leftmost picture below, which of options A-D is the result?

A)

B)

C)

D)