Cheat Sheet - Day 3

Introduction to Data Analysis with Python
https://github.com/janekfleper/Workshop-Konstanz-2024

Basics

`ord("a")`
`len([0, 1, 2])`
	A *function* is called with (optional) arguments

`y.count("a")`
`"abc".upper()`
	A *method* is called on a value or on a variable

`# a comment`	A `#` starts a *comment* that will not be evaluated
`type(x)`	Get the *type* of a value or of a variable
`del x`	Delete a variable

`def square(x):`
`    return x**2`
	Define a function with a *parameter* `x` and a *return* value

Data types

`"hello"`, `'abc'`, `"0.9"`, `str(123)`	A *string* is a sequence of characters in quoation marks
`12`, `-4`, `int("5")`	An *integer* is a number without a decimal part
`0.9`, `-3.1415`, `float("-0.1")`	A *float* is a number with a decimal part
`True`, `False`, `bool(0)`, `x < 1`	A *boolean* can only take the values `True` or `False`

`[0, "abc", 0.1]`
`list("hello")`
	A *list* is a mutable, sorted collection of values

`{"a": 1, "b": 2}`
`dict(a=1, b=2)`
	A *dictionary* is a mutable collection of key-value pairs

`(0, "0.9", True)`
`tuple([0, 1, 2])`
	A *tuple* is an immutable, sorted collection of values

Files

`with open("a.txt") as f:`
`    # do something`
	Open a file for reading

`for line in f:`
`    print(line)`
	Iterate over all lines in a file

`f.read()`	Read a file as a string

Arrays

`import numpy as np`	

`n = [1, 1, 2, 3]`
`a = np.array(n)`
	Initialize an array

`a[0]`, `a[0:-3]`	Get values from an array

`a[0] = 4`
`a[0:-3] = 0`
	Set values in an array

`a[a == 1]`	Filter data in an array
`np.unique(a)`	Get unique values in an array (and their counts)
`np.average(a)`	Compute the (weighted) average of an array
`np.sort(a)`	Get a sorted copy of an array
`np.argsort(a)`	Get the indices that would sort an array

`a.sum()`, `a.prod()`
`a.mean()`, `a.std()`
`a.min()`, `a.max()`
	Run computations on all values in an array

`a.shape`	Get the shape of an array
`a.dtype`	Get the data type of an array
`a.astype(t)`	Get a copy of an array with a specific data type `t`

Data frames

`import pandas as pd`	
`df = pd.DataFrame()`	Initialize a data frame

`df.iloc[1]`
`df.iloc[2:-4]`
	Access a row (or multiple rows) by the position

`df.loc["a"]`
`df.loc["a":"z"]`
	Access a row (or multiple rows) by the index

`df["col"]`
`df[["col1","col2"]]`
	Access a column (or multiple columns)

`df.at["row","col"]`	Read a single value from a data frame
`df.set_index("col")`	Move a column to index
`df.assign(col=data)`	Assign a (new) column
`df.sort_values("col")`	Sort based on a column
`df.query("col > 2")`	Query/filter a data frame
`df.plot("x", "y")`	Create a plot from columns in a data frame

`pd.read_csv()`
`pd.read_excel()`
`etc...`
	Open a file and read the content into a data frame

Jupyter shortcuts

Enter / Esc	Start/exit the edit mode
Shift + Enter	Run cell(s) and select next
A / B	Insert new cell above/below
↑, K / ↓, J	Select cell above/below
X, C, V	Cut, copy or paste cell(s)
D + D	Delete cell(s)
Z / Shift + Z	Undo/redo cell operation
Shift + Tab	Open the documentation

made with https://typst.app and https://typst.app/universe/package/cram-snap