**Software Carpentry**              **Hosted by ELIXIR Belgium**

**Wed-Thu, Oct 24-25, 2018                                                 Location: ICC Ghent**

**Workshop website**[**https://orchid00.github.io/2018-10-24-ugent/**](https://orchid00.github.io/2018-10-24-ugent/)

**Welcome to Software Carpentry**

Teaching computer skills to researchers. Software Carpentry is part of The Carpentries <https://software-carpentry.org/testimonials/>

Testimonials <https://software-carpentry.org/testimonials/>

Yor workshop is part of the many workshops around the globe <https://software-carpentry.org/workshops/>

**Friendly, respectful and active participants**

Participants are expected to follow our **code of conduct:** <https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html>

**Etherpad**

We will use this Etherpad to share links, snippets of code, take notes, ask and answer questions, and whatever else comes to mind.

This pad is synchronized as you type, so that everyone viewing this page sees the same text.

Remember with big power come big responsibilities, do not delete parts of this etherpad.

The Etherpad has three major parts:

* The left side holds today's notes: please edit these as we go along.
* The top right side shows the names of users who are logged in: please add your name and pick the color that best reflects your mood and personality.
* The bottom right is a real time chat window for asking questions of the instructor and your fellow learners.

**Public content page**

All content is publicly available under the Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>

Use of this service is restricted to members of the Software Carpentry and Data Carpentry community;

 this is not for general purpose use (for that, try etherpad.wikimedia.org).

**Pre-survey**

Make you sure you have filled the Software Carpentry workshop pre-survey so we know what is your pre-knowledge before the workshop

<https://www.surveymonkey.com/r/swc_pre_workshop_v1?workshop_id=2018-10-24-ugent>

**Post-survey**

<https://www.surveymonkey.com/r/swc_post_workshop_v1?workshop_id=2018-10-24-ugent>

**Sticky Notes:**Use these on the back of your laptop to signal us

Green is to celebrate your small victories, to indicate that you're happy and everything is fine

Red if we're going too fast, if you're having difficulty keeping up or that you need someone to stop by to help you.

**Twitter**

~~#~~**ELIXIR\_training** @ELIXIRnodeBE @thecarpentries @orchid00 @SVanHoey @sciensano @vibbioinfocore @PSB\_VIB

**Hands-on Training**

* This is a hands-on training :)
* It is interactive, which means your interaction and awareness will improve your learning
* Questions are always welcome
* Let’s start with *short* introductions - Stand up, say your name, one thing/place/food that makes you happy

**Motivation**

“Learning to code is a never ending journey with a set of challenges and delights unique to each person"

<https://twitter.com/aprilwensel/status/953387594783506433>

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Instructors:**

* Christof De Bo, VIB
* Paula Andrea Martinez, ELIXIR-BE
* Stijn Van Hoey, INBO Lifewatch

**Helpers:**

* Mauricio Macossay Castillo, VUB-VIB
* Raf Winand, Sciensano
* Lukasz Kreft, VIB

**Learners:**

**PAN**

Peter

Assia

 Jennifer Jörissen

 Laura Van Poelvoorde

 Bas Berbers

Sarah VW

 Amanda

  hilda

Lisa

Sari CogneauJonna Bouwknegtzongcheng

Florence BuytaersStéphanie Nouws

Ji

jasmijn

Attila Manel

Pieter-Jan Kerkhof

Asli Kucukosmanoglu

Emilie Froussart

gwen

Monah

Sinaeda Anderssen

Gaurav

Floris

Rafael

Mohamed

Sarah

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Project organization**

We recommend you to prepare the following folders to organize the contents of this workshop

2018\_10\_24\_SC

├── shell-novice

├── python-novice

└── git-novice

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Day 1 Shell**

data:    <http://swcarpentry.github.io/shell-novice/data/data-shell.zip>

command -flag -flag parameter

Feedback <https://goo.gl/XLQZEb>

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Day 1 Python**

Folder sctructure

├── python-novice

     ├── data   <http://swcarpentry.github.io/python-novice-gapminder/files/python-novice-gapminder-data.zip>

     Download a file from the internet

     # Linux

     wget <http://swcarpentry.github.io/python-novice-gapminder/files/python-novice-gapminder-data.zip>

     # for gitbash or Mac

     curl -O <http://swcarpentry.github.io/python-novice-gapminder/files/python-novice-gapminder-data.zip>

     # to unzip the compress file

     unzip python-novice-gapminder-data.zip

     └── scripts

**Notebooks**

**Open Jupyter**

On the shell from linux and Mac

jupyter notebook

If you have windows open your installed programs from the search and fnd jupyter notebook

Exercise

- Create a new notebook called `00\_basics` in your `scripts` folder

- Click on the `Help` option from the top menu and follow the tour `User interface Tour`

This tour will show you how everything is laid out

- What happends when you press `p`?

Variables

Values can be fixed, calculated or a result of a transformation. Variables follow naming conventions

* no spaces
* no hyphens
* no numbers at the start

Data types

* integer
* float
* string
* bool

# load the library

import pandas

Exercise - Reading in files (3 min)

- Use `pandas.read\_csv()`  to import data into Python

you will store three variables named `africa`, `europe`, `anz` from

`gapminder\_gdp\_africa.csv`, `gapminder\_gdp\_europe.csv`, `gapminder\_gdp\_oceania.csv`

use `index\_col='country'`

africa = pandas.read\_csv('../data/gapminder\_gdp\_africa.csv' , index\_col = "country")

europe = pandas.read\_csv('../data/gapminder\_gdp\_europe.csv' , index\_col = "country")

- It is recommended that you always use the help to find out more about the new

functions before using them

- Check the function `.info()` with your new variables

DataFramesDataFrames atributes

No parentesis

* df.columns # Column names or titles
* df.index      # Row names if you specify index\_col it will that value, if not it will be numbered from 0
* df.axes       # Rows and column names
* df.dtypes    # data types
* df.shape     # number of rows and columns carefull with the index
* df.ndim       # number of dimensions DataFrames should be 2
* df.size        # total number of cells in your dataFrame
* df.values    # array representation of all values
* df.T           # transpose reverse rows and columns

Data Frames methods

* df.info()
* df.describe()
* df.head()
* df.tail()

What variables do we have so far?

Exercise - Americas (2 min)

- Read the dataset `gapminder\_gdp\_americas.csv` (which should be in the same

directory as `gapminder\_gdp\_oceania.csv`) and index by `country` into a

variable called `americas`

- After reading the data for the Americas, use `help(americas.head)` and

`help(americas.tail)` to find out what `DataFrame.head` and `DataFrame.tail` do.

  - What method call will display the first three rows of this data?

  - What method call will display the last three columns of this data?

  (Hint: you may need to change your view of the data.)

- Display its summary statistics of `americas

americas.head(n = 3)

americas.tail(n = 3)

amr.describe()

Subsetting I

Exercise - optional

- Using `americas` transpose the DataFrame, get the last three rows and then transpose it again

- Save this DataFrame into a variable called `processed`

- use processed.columns.format you should get this

`['gdpPercap\_1997', 'gdpPercap\_2002', 'gdpPercap\_2007']`

americas.T.tail(3).T

Exercise - optional

- subset your favorite country

 Exercise - optional

- Using the `europe` DataFrame get the GDP value from 2007 in `Greece`

- The GDP value from 1952 in `Iceland`. Use two lines of code

- Optional, try getting the result as a DataFrame 2x2. Use one line of code

Subsetting information from a DataFrame

europe.iloc # subset a part of the dataframe using index

europe.loc # subset a part of the dataframe by name

favoritecountry = europe.loc['Belgium'] # by row name

europe.gdpPercap\_1997 picks the column gdpPercap\_1997

europe.loc['Greece','gdpPercap\_2007']

europe.loc[('Iceland','Greece'),('gdpPercap\_1952','gdpPercap\_2007')]

help(processed.to\_excel)

Exercise

- Let's make a folder results at the same level of data and scripts

- Save `processed` (or any other variable) into the `results` as a excel file with `sheet\_name='lastyearsGDP'`  extension .xlsx

- Open the file to check how it looks

- Let's also save the `processed` as a `csv`

Plotting

%matplotlib inline

import matplotlib.pyplot as plt

europe.plot()

---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Day 2 Python**

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Day 2 Git and Github**

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Resources**

* Python for Biologists: <http://www.programmingforbiologists.org/about/why-python/>
* Python for psychologists: <http://blog.efpsa.org/2016/07/12/python-programming-in-psychology-from-data-collection-to-analysis/>
* Biopython: <https://biopython.org/>
* Python Graph gallery: <https://python-graph-gallery.com/matplotlib/>
* <https://www.amazon.com/Python-Biologists-complete-programming-beginners/dp/1492346136>
* <https://www.amazon.com/Python-Data-Analysis-Wrangling-IPython/dp/1491957662>]
* Jupyter install <https://jupyter.org/install>
* Jupyter 2 or 3 <https://stackoverflow.com/questions/28831854/how-do-i-add-python3-kernel-to-jupyter-ipython>
* python 3.7 <https://stackoverflow.com/questions/51279791/how-to-upgrade-python-version-to-3-7>
* python 3.7 <https://superuser.com/questions/241865/updating-python-on-ubuntu-system>
* Reproducibility mybinder <https://mybinder.org/>
* Reproducibility codeocean <https://codeocean.com/>
* Mybinder example <https://hub.mybinder.org/user/binder-examples-nda_environment-6kif41cf/notebooks/index.ipynb>
* Python DataFrame <https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html>
* Pandas cheatsheet <https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf>
* Pandas cheatsheet dataquest <https://www.dataquest.io/blog/large_files/pandas-cheat-sheet.pdf>
* Pandas cheatsheet datacamp <https://s3.amazonaws.com/assets.datacamp.com/blog_assets/PandasPythonForDataScience+(1).pdf>
* Python matplotlib <https://matplotlib.org/users/dflt_style_changes.html>
* Python Style guide <https://www.python.org/dev/peps/pep-0008/>

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

有道词典

subset a part o ...

详细X

子集的一部分dataframe使用索引

有道词典

subset a part o ...

详细X

子集的一部分dataframe的名字

 help  [help]  详细X

基本翻译

n. 帮助；补救办法；帮忙者；有益的东西

vt. 帮助；促进；治疗；补救

vi. 帮助；有用；招待

n. (Help)人名；(芬)海尔普

网络释义

Help: 帮助

help oneself: 自取所需

Help out: 帮帮忙