Working With Data On The Web
Online Course Report

Course link : http://swcarpentry.github.io/web-data-python/

Etienne JEAN

September 2018

This course deals with getting and publishing data on the web, and general handling of those data, using Python. The lesson is supposed to take approximately 1h30. Reading and testing the code provided, I took a little bit longer than that, around 2h in total. It requires only basic Python skills to be understood.

The course is divided in 6 parts :

— Getting data
— Handling CSV data
— Generalizing and handling erroros
— Visualization
— Publishing data
— Making data findable

1   What I learned

Since the course gives an overview of getting and publishing data on the web, I learned about the `request` package of Python to make requests to a server and get data from urls. I also learned about the `csv` package to handle csv files, although there is not much value added by this package in my opinion, since csv files are very simple to parse and write.

I did not know about the status codes, that gives standardized information about the response of the server. An good point is highlighted in the course : these codes are not always respected by the databases themselves, so we have to be careful about errors, and sometimes handling errors yourself is necessary.

I learned about the `assertion` key word in Python, which is used to prevent silent errors. It returns an assertion error in case the condition given as an argument is not respected.

Finally, I learned some of the standard practices when publishing data on the web, like the `public_html` directory on servers, and the index file that gives informations about data available on the database.

2 Judgement about this course

2.1 Negative points

Firstly, I think this course needs to be updated. It looks unfinished (it is written "FIXME" next to each chapter on the home page), yet abandonned (the URL of the World Bank's climate data in the first chapter is actually an old format. It still worked with this format though).

But more importantly, it is not really informative about working with data *on the web*, which is the main topic. It sometimes looks like a standard python course, which has not much to do with handling data on the web (general error handling, visualization of data with pyplot, writing files).

Finally, I think one of the solution is wrongly implemented in the last chapter : the function `update_index` opens and memorizes and entire file only to append a line at the end of the file. The argument `'a'` of the function `open()` does the job.

2.2 Positive points

This course gives a good and rather quick overview of how to deal with data on the web. It is also clearly organized and clearly explained. The chapters are well balanced in time to spend and things to learn. Moreover, the links to the glossary make it easy to have a quick look at a concept without losing track of the course.

3 In conclusion

The idea of the course is good, but it seems rather badly thought, and incomplete. I did not learn much, but it stills gives a quick overview of how to get and publish data on the web.