Crossing pedagogical boundaries with open data

Santeri Koivula, Peitsa Veteli and Veera Juntunen
Helsinki Institute of Physics HIP
The agenda

- What is this about?
- Exercise: What is the Jupyter Notebook?
- Going through an advanced example
- Discussion exercises
- Website
- Feedback
Open data is everywhere — CERN, NASA, The World Bank, etc.

Our mission is to help teachers to create research-based exercises for their courses using open data. Multidisciplinarity; curiosity; and skills for data analysis are at the heart of what we do.

We have created study materials, organized workshops for teachers, and helped teachers to use open data in their classes.
Data:
Factual information collected to be used for reasoning, discussion, or decision-making

Open data:
Data that can be freely used, modified, and shared by anyone for any purpose (according to Open Definition)
HOW WOULD THIS BE USEFUL FOR YOU?

- You are planning to teach at some point
- You are planning to go into research
  - Communicating about your results
- Critical thinking
  - Assessing the credibility of data
  - Analyzing information in different contexts
WHY?
WHY?

Future citizenship - understanding, analyzing information and assessing its credibility

The amount of data is increasing

Communication skills

Understanding large phenomena and different contexts

Ability to understand and question

Getting acquainted with the tools of science

Multidisciplinary learning
In its simplest form, the student is provided with a link of an exercise that uses open data. An exercise can be saved either as a notebook or a PDF file.

The exercises can also work as a tool for a student to reflect on their learning, in which case other programs or platforms are not even needed.

Easy!
How much work for teachers?

- What kind of materials?
- How much coding?
- How much can students influence on the exercises?
- Do you want to make materials by yourself?
- What is the group size?
- Returning the exercises?

Beginner – Intermediate – Expert
WHAT OUR EXERCISES LOOK LIKE

Plotting the invariant mass histogram

In this exercise, we learn how to plot the histogram of invariants masses with Python. Let us use the data collected by the CMS detector in 2011 [1]. Events with specific criteria [2] have been selected in the CSV file `Ymumu_Run2011A.csv`, which we are using.

Explore the different code cells below and run the code. Note that normally the code would not be commented as much as this. Here, the reason for these comments is to explain in detail what the code is doing.


1) Start

```python
# Import the needed modules. Pandas is for the data-analysis
# and matplotlib.pyplot for making plots. Modules are named as pd and plt.
import pandas as pd
import matplotlib.pyplot as plt
# Jupyter Notebook uses "magic functions", with this function it is possible to plot
# the histogram straight to notebook.
%matplotlib inline
```

2) Getting the data

```python
# Create a new DataFrame structure from the file "Ymumu_Run2011A.csv"
dataset = pd.read_csv("./Data/Ymumu_Run2011A.csv")
# Create a series structure (basically a list) and name it "invariant_mass".
# Use the column "r" from the "dataset" to the variable "invariant_mass".
invariant_mass = dataset[r]
```

3) Plotting the histogram

Now we can create and plot the histogram of the values of the invarient masses. The histogram shows how many events the invariant mass of the muon pair is in a certain value range. Note that we will use total 500 bins in the histogram, so you will not split the separate bins because there are so many of them.

```python
# Plot the histogram with the function hist() of the matplotlib.pyplot module.
plt.hist(dataset[y], bins=500, color='blue')
plt.xlabel('Invariant mass [GeV/c^2]')
plt.ylabel('Number of events')
plt.title('The histogram of the invariant masses of two muons $\mu^+\mu^-$')
plt.show()
```

4) Analysis

- What does the histogram tell us?
- What happens around the mass 95 GeV?

```
```

By HEP Education and Open Data Team

Development of this material is made possible by a grant from Finnish National Agency for Education (Ouno). The material on this website is licensed under CC BY-4.0 licence.
WHY LEARN A NEW PLATFORM?

Versatility - text, code, images, videos, animations

Everything from instructions to exercises are in the same place

Students can return one document only

Teacher can easily run the commands again while going over the results

... What about Excel?
DIFFERENT SUBJECTS

Open data can be used in many different subjects.

In Finnish we have materials on physics, biology, text analysis, geography, and mathematics.

Similar exercises could be used in other fields as well, such as history, economics, and psychology.
Kerro, kerro, kuvaaja

Tehtävä koostaa uusien lähdimmässä ja kirkkaana mukaan, siten että kirkkaammin tähän on paneena abstraktioon mainosti. Aina on ollut ihmisten halus, että lähdimmässä mainostaa. Parhaillaan on ollut ihmisten halus, että lähdimmässä mainostaa. Parhaillaan on ollut ihmisten halus, että lähdimmässä mainostaa. Parhaillaan on ollut ihmisten halus, että lähdimmässä mainostaa. Parhaillaan on ollut ihmen...
Cumulative number of people who have been to space, 1961 to 2021
For individuals who have been to space multiple times, only the date of their first visit is shown.

Source: Our World In Data

Source: Gapminder


PROJECTS

2021
- HEALTH EDUCATION Autumn 2021
- PROGRAMMING Autumn 2021
- BIOLOGY Autumn 2021

Spring 2022
- SCIENCE COURSE Spring 2022

Autumn 2022
- SCIENCE COURSE Autumn 2022
- CLIMATE THEME DAY Autumn 2022

Teacher training       Workshops       Classes       Theme days
individual work – groupwork – laboratory work – essays – analyses – research work
WEBSITE

and materials


You can find everything you need to get started from our website, such as materials that are ready to use or to modify, and links to websites that publish open data. Currently we have materials in English on particle physics and text analysis.
CONTACT US

Peitsa Veteli
peitsa.veteli@helsinki.fi

Veera Juntunen
veera.juntunen@helsinki.fi

Santeri Koivula
santeri.jan.viliam.koivula@cern.ch
1. Jupyter Notebook - exercise
2. Going through an advanced example
3. Our website
DATA RESOURCES

- CERN Open Data Portal
- Our World In Data
- Figshare
- Zenodo
- World Bank
- WHO: Global Health Observatory