**Creating Other Graphs in Python**

This planning document is intended to support teachers who are delivering the NPA/PDA Data Science or for students who are learning independently. It also aligns with the Data Skills for Work framework.

Contents

[Version Control 1](#_Toc103770831)

[Lesson Description 2](#_Toc103770832)

[Lesson Contents 2](#_Toc103770833)

[Learning Intentions 3](#_Toc103770834)

[Success Criteria 3](#_Toc103770835)

[Knowledge Prerequisites 3](#_Toc103770836)

[Lesson Requirements 4](#_Toc103770837)

[Jupyter Notebook 5](#_Toc103770838)

[Datasets 7](#_Toc103770839)

[How you can use this lesson 8](#_Toc103770840)

[Alternative format 8](#_Toc103770841)

# Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Version number | Purpose/Change | By | Date |
| 1.0 | Published by Effini | John Bell | 18 May 2022 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Lesson Description

|  |  |
| --- | --- |
| **Lesson Overview** | Creation and modification of histograms, line graphs and scatter plots in Python using the seaborn package.  This lesson follows **Creating** **Bar Charts in Python**. |
| **Topic** | Visualisation and Storytelling |
| **Book Chapter(s)** | Interpreting data |

|  |  |
| --- | --- |
| **NPA level** | 5, 6 |
| **PDA level** | 7, 8 |
| **Data skills for work level** | Core, Analysis |

# Lesson Contents

This lesson consists of:

* A lesson plan (this document)
* A PowerPoint presentation, ‘Creating Other Graphs in Python’
* 2 Jupyter notebooks:
  + ‘creating\_other\_graphs.ipynb’ (for learners)
  + ‘creating\_other\_graphs\_with\_answers.ipynb’ (for teachers)

# Learning Intentions

We will be learning about **creating graphs in Python,** specifically the process of creating

* **histograms**
* **line graphs**, and
* **scatterplots**

# Success Criteria

I can *create* and *modify* histograms in Python.

I can *create* and *modify* line graphs in Python.

I can *create* and *modify* scatter plots in Python.

# Knowledge Prerequisites

Learners should know:

* Python programming to at least the level defined in SQA Computer Programming Level 5 (HY2C 45)
* How to use a Jupyter notebook to write, edit and run Python code
* That identifying patterns is part of the analysis steps

# Lesson Requirements

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PDA** | **NPA** | **Data Skills for work** |
| **Qualification** | Yes | Yes | Yes |
| **Outcome ID(s)** | CD7.3e, CD8.2h | DC5.3c, DC6.3d | c2.1, a2.1, a2.2, a3.1, a3.2 |
| **Outcome description(s)** | CD7.3e Creating visualisations using software  CD8.2h Creating visualisations using software | DC5.3c Create appropriate visualisations from data.  DC6.3d Create appropriate visualisations from data. | c2.1 Vocabulary used in data science and analytics  a2.1 Use of tools to analyse data  a2.2 Selecting suitable visualisations  a3.1 Visualisation of data to provide insight  a3.2 Visualisation of data to tell stories |
| **Level** | 7, 8 | 5, 6 | Core, Analysis |
| **Software language** | Python | Python | Python |
| **Required equipment /software for student** | Lesson: PowerPoint  Python notebook: Jupyter notebook environment | Lesson: PowerPoint  Python notebook: Jupyter notebook environment | Lesson: PowerPoint  Python notebook: Jupyter notebook environment |

# 

# Jupyter Notebook

There is a Jupyter notebook for this lesson that provides examples and programming tasks for learners, drawn from the examples in the lesson PowerPoint.

The notebook uses Python 3.x and the following packages:

* [numpy](https://numpy.org/) – for scientific computing
* [pandas](https://pandas.pydata.org/) - for data manipulation
* [s3fs](https://pypi.org/project/s3fs/) - an API to AWS S3 (Simple Storage Service), used to import datasets
* [seaborn](https://seaborn.pydata.org/) – for data visualisation
* [matplotlib](https://matplotlib.org/) – for data visualisation

The tasks are described in the table below.

|  |  |  |
| --- | --- | --- |
| **Notebook section** | **Task** | **Description** |
| Histograms | Task 1 - Women's heights | Create a simple histogram to show the distribution of women's average heights across a selection of countries. |
| Task 2 - Average temperature in UK cities between 1994 and 2013 | Create a histogram that shows the distribution of average temperature values across the years 1994 to 2013. |
| Task 3 - Women's and men's heights | Create two histograms on a single plot (and, optionally, adjust the transparency of the bars). |
| Task 4 - male and female life expectancy | Create two histograms on a single plot and add a legend. |
| Line Graphs | Task 5 - Scotland's population | Create a line graph that shows how the population of Scotland has changed over time. Give your graph a title and axis titles. |
| Task 6 - Despacito streams | Create a line graph using showing the number of streams of the song "Despacito" on Spotify over time. |
| Extension Task 1 - Life expectancy trends for girls and boys | Create two line graphs on a single plot.  Give your graph a title and give the axes titles.  Optionally, adjust the colour, line width and line style.  Add a legend.  Modify the tick mark interval on an axis. |
| Scatter Plots | Task 7 - Comparing the cost and earnings of Marvel movies | Create a scatter plot which plots the the budget of the Marvel movies against their earnings. |

# Datasets

The following datasets are used in this lesson.

|  |  |  |
| --- | --- | --- |
| **Dataset name** | **Description** | **Link** |
| height\_of\_male\_and\_female\_by\_country\_2022 | The average heights of men and women across various countries in 2022. | <https://datasets.learn-data.science/height_of_male_and_female_by_country_2022.csv> |
| climate\_data\_per\_city\_uk | The average yearly temperature for cities in the UK from 1743 to 2013. | <https://datasets.learn-data.science/climate_data_per_city_uk.csv> |
| life\_expectancy\_by\_sex\_scotland\_1980\_2018 | For babies born in Scotland between 1980 and 2018, what is the average length (in years) that they can expect to live? | <https://datasets.learn-data.science/life_expectancy_by_sex_scotland_1980_2018.csv> |
| scotland\_population.csv | The approximate population of Scotland from 1900 to 2020. | <https://datasets.learn-data.science/scotland_population.csv> |
| rainfall | A toy dataset showing the rainfall at hourly intervals over the course of a day. | <https://datasets.learn-data.science/rainfall.csv> |
| spotify\_streams | The number of streams of five popular songs in 2017 and 2018 on [Spotify](https://www.spotify.com/uk/). | <https://datasets.learn-data.science/spotify_streams.csv> |
| marvel\_movies | Budget and revenue figures for recent [Marvel movies](https://www.marvel.com/movies). | <https://datasets.learn-data.science/marvel_movies.csv> |

# How you can use this lesson

This lesson has been created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

© 2022. This work is licensed under a [*CC BY-NC-SA 4.0 license*](https://creativecommons.org/licenses/by-nc/4.0/legalcode)*.*

A picture containing text, clipart

Description automatically generated

You are free to:

* **Share** – copy and redistribute the material in any medium or format
* **Adapt** – remix, transform and build upon the material

Under the following terms:

* **Attribution** — You must give [appropriate credit](https://creativecommons.org/licenses/by-nc/4.0/), provide a link to the license, and [indicate if changes were made](https://creativecommons.org/licenses/by-nc/4.0/). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
* **NonCommercial** — You may not use the material for [commercial purposes](https://creativecommons.org/licenses/by-nc/4.0/).
* **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](https://creativecommons.org/licenses/by-nc-sa/4.0/) as the original.

# Alternative format

**If you require this document in an alternative format, such as large print or a coloured background, please contact**

**hello@effini.com**

**or**

**4th Floor, The Bayes Centre**

**47 Potterrow**

**Edinburgh**

**EH8 9BT**