

Skillsoft-Data Visualization with Python

Overview:

Students will develop a powerful toolkit to visualize data with static graphs and begin exploring interactive charts. This course will help students become expert data communicators to make their analyses shine. By the end of the course, attendees will be able to manipulate and summarize a variety of file formats and build intuitive data visualizations with Python.

Prerequisites:

Attendees must be comfortable using Python to manipulate data and perform basic operations, including data wrangling and must have the correct libraries of cufflinks and plotly installed.

(instructions included)

Objectives:

- Discuss the concepts and use cases of data visualization
- Build static data visualizations in Python
- Build basic interactive data visualizations in Python
- Select the appropriate data visualization for each scenario

Topics covered

- Day One
- Introduction to the course
- Introduce, load and clean dataset
- Reshape data using pandas
- Define use cases of Exploratory Data Analysis (EDA)

Day two

- Visualizing data with matplotlib
- Create histograms, boxplots, and bar charts
- Create scatterplots
- Customize graphs
- Create violin plots
- Create compound visualizations in grid format

Day Three

- Create layered plots
- Save your plots and your data
- Best practices of data visualization

Day Four

- Describe uses and strengths of plotly and cufflinks packages
- Transform dataset for visualizations
- Create basic interactive visualizations using cufflinks
- Visualize multiple metrics using cufflinks
- Create interactive visualizations using plotly
- Generate interactive visualizations with transformed summary data

Software and package requirements

Python and Anaconda

Packages:

- `import pandas as pd`
- `import numpy as np`
- `import pickle`
- `import os`
- `import matplotlib.pyplot as plt`
- `pip install cufflinks==0.16.0`
- `pip install plotly==3.10.0`

Biography



Martin Skarzynski, a Cancer Prevention Fellow since 2017, is passionate about Bioinformatics, Data Science, Epidemiology, and Statistical Computing. Martin uses the Python and R programming languages and command line tools to explore, analyze, visualize and present data and has a strong interest in reproducibility, scientific publishing workflows, and open data/science best practices. Martin is excited to apply his computational skills in combination with his Genomics and Immunology background to the study and prevention of cancer.

Outside of the Cancer Prevention Fellowship Program, Martin is co-chair of the Bioinformatics and Data Science Department at the Foundation for the Advancement of Education in the Sciences (FAES), where he has been an instructor since 2015 and currently teaches Introduction to Python (BIOF309) and Applied Machine Learning (BIOF509). Martin is also an instructor for Software and Data Carpentry, non-profit organizations that teach computational skills.

Martin holds an MPH in Epidemiologic and Biostatistical Methods for Public Health and Clinical Research and a certificate in Data Science from Johns Hopkins University, a PhD in Tumor Biology from Georgetown University, an MS in Biotechnology from Jagiellonian University in Poland, and a BA in Biology from St. Mary's College of Maryland.