Data Visualization in Python: Matplotlib

A Reproducible Research Workshop

Simon Stone
Research Data Services
Dartmouth College

About the Reproducible Research Group

- Joint venture of **Research Computing @ ITC** and **Research Data Services @ Library**
- Consult with **experts** on
  - research data management,
  - data visualization,
  - biomedical research support,
  - spatial data and GIS,
  - high performance and research computing,
  - statistical analysis,
  - economics and social sciences data
- **Meet** the people on campus that support your reproducible research lifecycle
- **Engage** in community discussions to learn from other researchers on campus
- Attend a workshop to **learn** practical tools and tips

About Research Data Services

Research Data Management

Data Management Plans (DMPs) for sponsored projects

Finding and using 3rd party data

Collection and cleaning of data

Organization and documentation

Publishing and Repositories

Data Analysis/Visualization

Textual, numeric, spatial data

Reproducible research workflows

Scripting in R: tidyverse core package (i.e. ggplot, dplyr, tydr, tibble, etc.)

Scripting in Python: NumPy, SciPy, Pandas, Scikit-learn, Matplotlib, Seaborn, (OpenCV, PyTorch, TensorFlow, Tesseract, NLTK, etc.)

Computational Scholarship

Computational project planning

Collections as Data

Storytelling with data and visualizations

Text and data mining

Digital Humanities support

Computational Pedagogy

Work with us

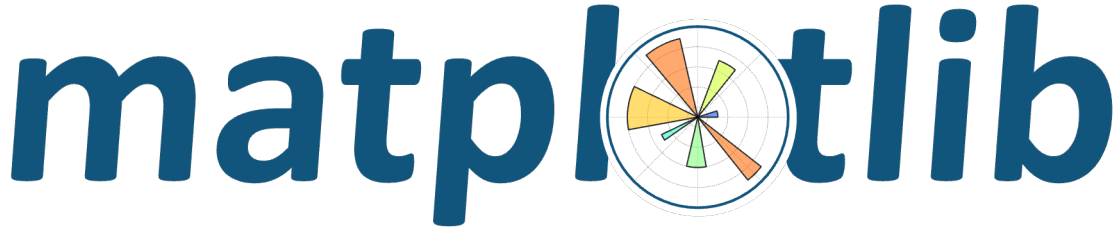
ResearchDataHelp@groups.dartmouth.edu

Jeremy Mikecz
Research Data Science Specialist
jeremy.m.mikecz@dartmouth.edu
dartgo.org/jeremyappts

Simon Stone
Research Data Science Specialist
simon.stone@dartmouth.edu
dartgo.org/meetwithsimon

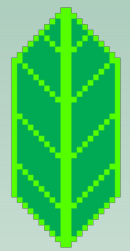
Lora Leligdon
Head of Research Data Services
lora.c.leligdon@dartmouth.edu
dartgo.org/lora

The Data Visualization Ecosystem in Python



Tabular data

Geospatial data



Folium

Dashboards

and many, many more...

Matplotlib is...

📊 ...a Python library for creating static, animated, and interactive visualizations.

 ...designed after MATLAB's plotting interface (hence the name).

🫶 ...an extremely popular choice for scientific, publication-ready visualizations.

💪 ...versatile, efficient, and performant.

 ...used by other visualization libraries "under the hood" (e.g., seaborn).

😬 ...starting to show its age a little bit and can be a bit bland (sometimes verbose interface, not great for dashboarding).

What you will learn in this workshop

- **How** matplotlib is organized
- **How** to get started with matplotlib's interface(s)
- **How** to compose a visualization
- **How** to choose the right kind of plot for your data

What we will work with in this workshop

- Platform: https://jhub.Dartmouth.edu
- Python
- matplotlib
- Materials: **www.dartgo.org/rr-matplotlib**

What we will *not* work with in this workshop

- Matplotlib is ideally used together with NumPy
- Numpy is a Python library for numerical computing (arrays, matrices, lots of fancy math)

- Here, we want to focus on matplotlib itself and will thus only use Python's built-in `list` type as our data structure

Let’s start plotting…

Next steps

Explore matplotlib's many official tutorials covering a vast array of topics from basic to advanced!

Learn more about NumPy and how to use it with matplotlib

Use your new knowledge to learn more about your visualizations in pandas and seaborn (which are both built on matplotlib)

Thank you.