

## CAMBAM workshop

# Interactive data visualisation in Python

## Scheduled

Monday 27th July 9.00am-1.30pm

## Summary

Modern scientific methods give rise to vast quantities of data. Creating effective visualisations is essential for both presentation and more importantly exploration of the data. This is no easy task when the data contains dozens of variables and millions of entries. Traditional visualisations are static, that is, what the user sees is what the user gets. Using interactive data visualisations allows the user to vary parameters, honing in on subsections of the data, or switching between different plot types - all without touching the code. This allows for rapid exploration of the data and seamless sharing amongst collaborators, who only require a web browser to open the visualisation.

This workshop will equip participants with the skills required to begin creating interactive visualisations in Python. The format will be highly interactive, with alternation between demonstrations by the instructor and participants working through their own Jupyter notebook (provided in advance). Participants will come away having made several of their own visualisations of either a large public dataset, or their own dataset if they would like to bring one. An example of what can be achieved using these tools can be found at the following link, where data output from a model of a cardiac arrhythmia is interactively viewed and analysed.

<https://modulated-parasystole.herokuapp.com/>

---

## Description of methodology

Basic knowledge of the Pandas library (handling of dataframes) will be assumed prior to the workshop. For those not familiar with Pandas, a (brief) tutorial will be provided in the form of a Jupyter notebook to work through, which will provide sufficient background. We will use the Plotly library for creating interactive plots, and the Dash library to combine these into a dashboard. Participants of the workshop will learn and implement the following techniques:

1. Organisation of Pandas DataFrames to facilitate rapid plotting.
2. Use of basic Plotly functionality including the creation of different plot representations (Bar, Line, Scatter, Histogram) and exportation to html files for interactive viewing.
3. Visualisation of higher-dimensional data with the use of multiple traces, grid plots and 3-dimensional plotting.
4. Integration of sliders and drop-down boxes to interactively switch between different visualisations of the data.
5. (If time allows.) Combining interactive plots into a dashboard with components connected by callback functions.

## Software requirements

The following should be installed on the participant's computer prior to the workshop:

- Python 3 (Anaconda distribution <https://www.anaconda.com/products/individual> recommended)
- Jupyter (comes installed with the Anaconda dist. above)
- Python libraries: Numpy, Pandas, Plotly, Dash

A Jupyter notebook will be sent around prior to the workshop to ensure that software is working in advance. If problems arise, please contact the instructor for help.

## Instructor

Dr. Thomas Bury  
Postdoctoral researcher  
Department of Physiology, McGill University  
Contact: [thomas.bury@mcgill.ca](mailto:thomas.bury@mcgill.ca)