Interactive data visualisation in Python

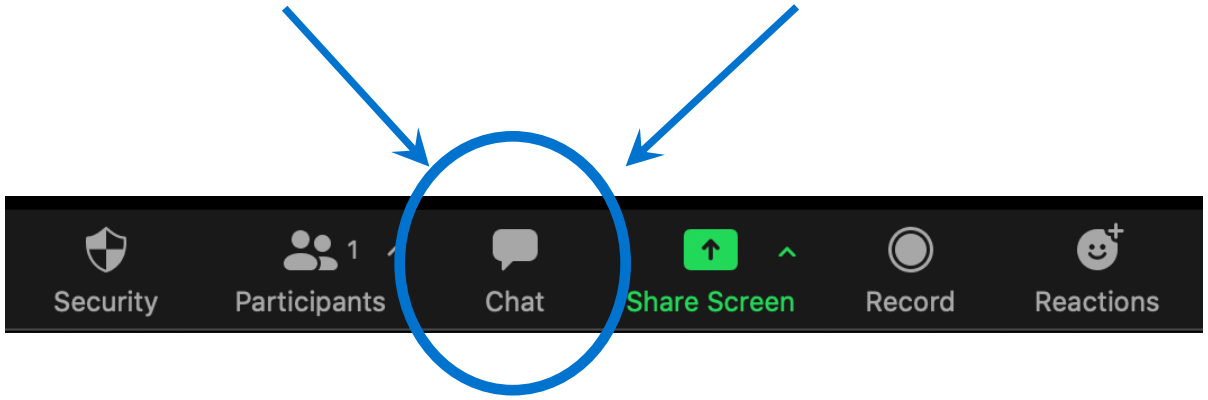
CAMBAM-CRM Mini-workshop
Monday 27 July 2020

Instructor: Thomas Bury[1]

[1]Dept. Physiology, McGill University

Introductions

- Get in the chat!
  - Where are you from?
  - What research do you do?



- Zoom etiquette: please stay muted unless asking a question to avoid background noise.

Checklist for coding throughout this workshop

- Installed the necessary software:
  - Python 3
  - JupyterLab
  - numpy, pandas, plotly, dash, jupyter-dash

- Run the Jupyter notebook `test.ipynb` successfully.
- Download the Folder `cambam_workshop_shared` and its contents from Google Drive (link in email)

Learning objectives

By the end of this workshop, you will be able to:

- Import recent and high impact public datasets
- Conduct an initial data exploration in Pandas
- Rapidly create interactive plots (e.g. scatter, histogram, box) as html files
- Visualise a third dimension using grid, 3D and heat plots
- Create a slider to view a fourth dimension
- Modify a Dash template to create your own web app

Why bother with interactivity?

- Data these days if often ‘big’ : lots of variables (dimensions), lots of entries.
- E.g. physiological output from wearable sensors.
- Can be difficult adequately represent it on a 2D static plot
- Having interactivity allows us to
    1. Navigate through different sections of data
    2. Vary the scale of the axes
    3. View all variable values with a tooltip
    4. Compute statistics on the fly
- Emphasis will be on **exploratory** visualisation

Icentia CardioSTAT

Demonstration of our dashboard using Plotly

Disclaimers

- I am not sponsored by Plotly
- I am not an expert in Plotly

Agenda

9.00am	My introductory waffling
9.20am	Notebook #1 – basic Plotly functionality ○ Demonstration (me) [30 mins] ○ Participation (you) [20 mins]
10.10am	Break [5 mins]
10.15am	Notebook #2 – visualisation with higher dimensions [1hr]
11.15am	Break [15 mins]
11.30am	Notebook #3 - sliders and buttons [45 mins]
12.15pm	Break [5 mins]
12.20pm	Notebook #4 – introduction to Dash [40 mins]
1.00pm	Sharing of visualisations and feedback [20 mins]
1.20pm	Closing remarks
1.30pm	Fin

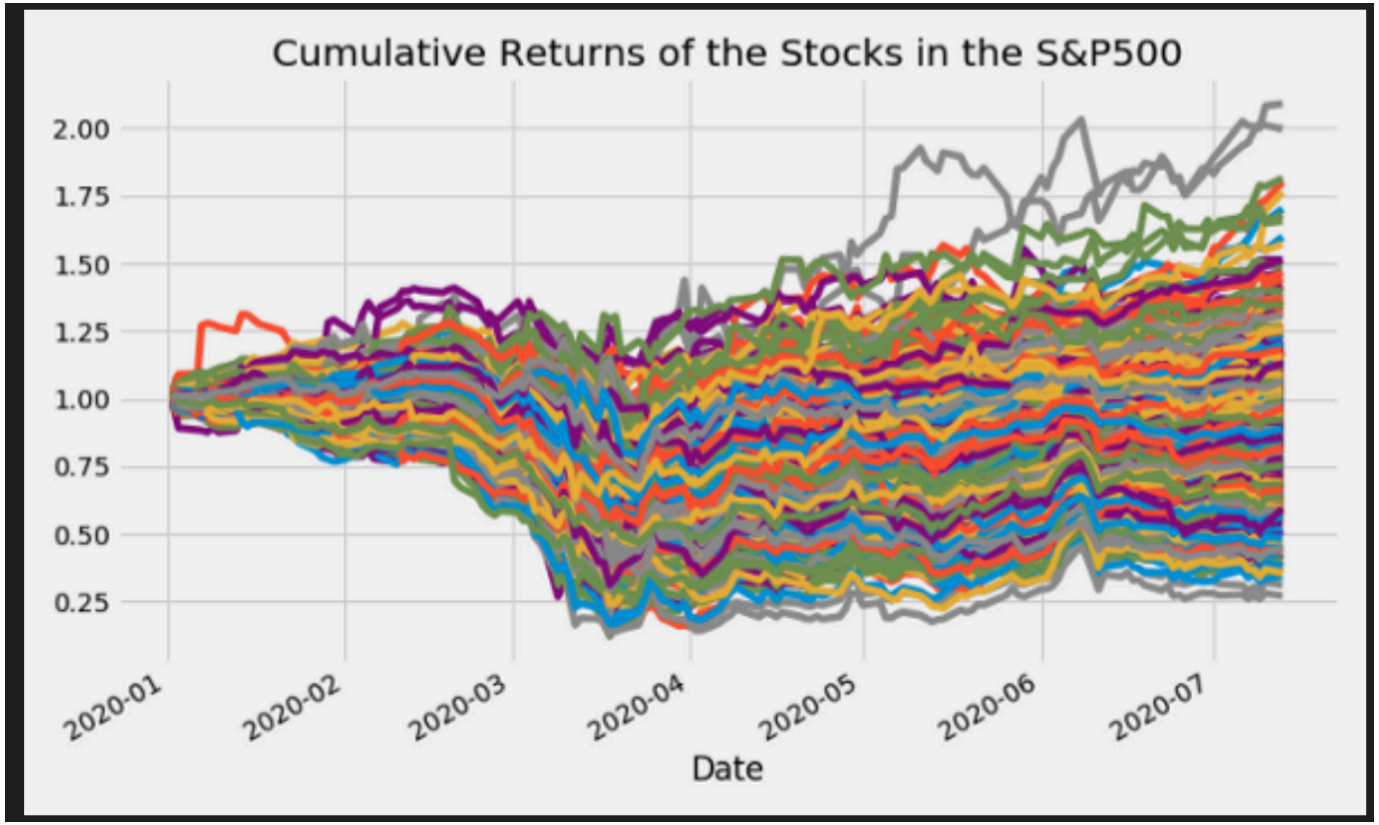
Notebook #1

Plotly fundamentals

Visualising additional dimensions

What attributes can you add to a plot to view a higher dimension?

Brainstorm!



Possible approaches

- Static visualisation methods:
  - Colour of data points
  - Size of datapoints
  - Grid of plots
  - Contour plots / Heat maps
- Interactive visualisation methods
  - Navigation through 3D plots
  - Hover data
  - Buttons, sliders and drop-down boxes

Possible approaches

- Static visualisation methods:
  - Colour of data points
  - Size of datapoints
  - Grid of plots
  - Contour plots / Heat maps
- Interactive visualisation methods
  - Navigation through 3D plots
  - Hover data
  - Buttons, sliders and drop-down boxes

Notebook #2

Visualising higher dimensions

Notebook #3

Sliders and buttons

Notebook #4

An introduction to Plotly Dash

Sharing visualisations

- Any volunteers for sharing the visualisations they have made?
- Complete or incomplete / working or not working (I can provide feedback)

Closing remarks

Learning objectives

By the end of this workshop, you will be able to:

- Import recent and high impact public datasets
- Conduct an initial data exploration in Pandas
- Rapidly create interactive plots (e.g. scatter, histogram, box) as html files
- Visualise a third dimension using grid, 3D and heat plots
- Create a slider to view a fourth dimension
- Modify a Dash template to create your own app.

Resources for Dash app deployment

- Read through the Plotly tutorial on deployment
  https://dash.plotly.com/deployment

- You can deploy several applications with Heroku for free (easiest option, though memory restrictions apply)
  https://devcenter.heroku.com/articles/getting-started-with-python

- Alternatively, your institution may provide cloud services where you can host Dash apps. E.g with Compute Canada.
  https://www.computecanada.ca/research-portal/national-services/compute-canada-cloud/

  "Users of the cloud service should ideally have at least an intermediate technical knowledge in systems management"

Resources for further Plotly/Dash tutorials

- Youtube channel ‘Charming data’
  https://www.youtube.com/channel/UCqBFsuAz41sqWcFjZkqmJqQ
- Plotly tutorial for beginners: Kaggle
  https://www.kaggle.com/kanncaa1/plotly-tutorial-for-beginners

Tips for fast improvement

- Design and create your own visualisations and look up the documentation as required.
  https://plotly.com/python/
- Get inspired by other Dash apps (e.g. Dash app gallery https://dash-gallery.plotly.host/Portal/), get the source code, and play with it.

Final questions