Data Visualisation

Associate Professor Goh Wooi Boon

College of Computing and Data Science
Nanyang Technological University

email: *aswbgoh@ntu.edu.sg*

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1

1

Chapter 4 –Visualisation Tools

Contents

- Types Of Visualisation Tools
- Choosing The Visualisation Tools
- Data Visualisation Tools Used In This Course
- Other Python Data Visualisation Libraries

2

2

Types Of Visualisation Tools

"Like a Child in a Candy Store"

- There is a huge variety of tools available to refine and query your data, and to create, edit, and display your visualisations.

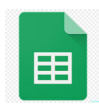- The types of tools for building visualisations fall into four major categories[1]:

- Basic productivity applications

  Microsoft Excel   LibreOffice   Google Sheet

- Visualisation software

  PROTEOPEDIA – LIFE IN 3D   ArcGIS   3DSlicer

- Business intelligence tools

  tableau   Power BI   QilkView

- Developer based packages

  D3.js - Javascript   bokeh seaborn matplotlib Python-based

[1] Kristen Sosulski, Data Visualization Software: Tips for picking the right tools - http://www.kristensosulski.com/2018/11/criteria-for-evaluating-data-visualization-tools/

3

3

Choosing The Visualisation Tool

How do I Choose the Right One?

- When given a free choice, it is important to select the tool that is most suitable to support the nature and purpose of your data visualisation work and workflow.

- Consider the following aspects in the evaluation[1]:

  - **Sharing**
  - **Output**
  - **Display types**
  - **Data Exploration**
  - **Interoperability**
  - **Simplicity**
  - **Persistence**

1. Sharing
2. Output
3. Interoperability
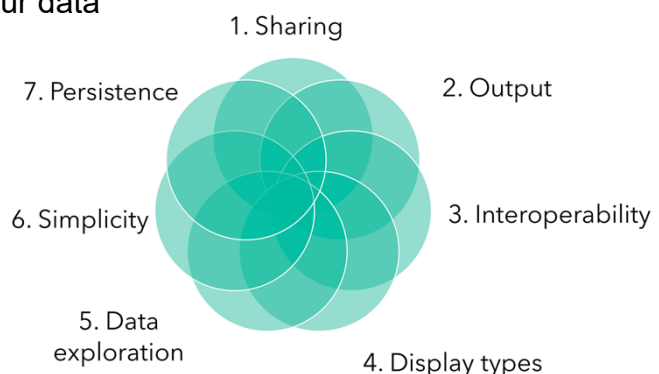4. Display types
5. Data exploration
6. Simplicity
7. Persistence

Image taken from [1]

[1] Kristen Sosulski, Data Visualization Software: Tips for picking the right tools - http://www.kristensosulski.com/2018/11/criteria-for-evaluating-data-visualization-tools/

4

4

Choosing The Visualisation Tool

How do I Choose the Right One?

- **Output** – Are you publishing the visualizations to the web, PDF documents or embedding them into other applications? The destination of your visualization will dictate tool choice. e.g. not all tools have good web interactive support[1].

- **Display types** – What types of visualizations do you intend to create? e.g. maps, networks, and text-based visualizations may not be supported by some tool[1].

- **Interoperability** – How easily can you connect to other data sources? For instance, does the software allow you to import diverse file types, such as .xls, CSV, .txt, or allow you to link to your databases?[1].

- **Data exploration** – Do you need a tool to explore and present your data visually, or just to present your data visualization visuals? e.g. features such as visual querying may not be supported by some tools[1].

[1] Kristen Sosulski, Data Visualization Software: Tips for picking the right tools - http://www.kristensosulski.com/2018/11/criteria-for-evaluating-data-visualization-tools/
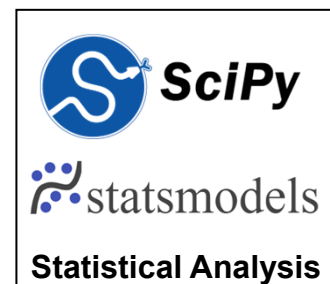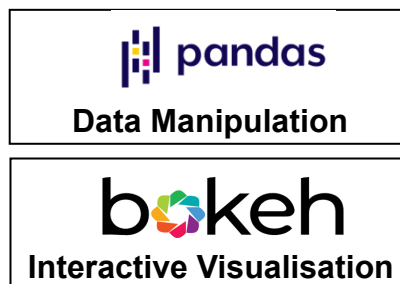
5

5

Choosing The Visualisation Tool

How do I Choose the Right One?

- **Simplicity** – Are you looking to create charts and graphs quickly? How much time can you afford to learn how to use the tool as some may be powerful but require a steep learning curve, even to build a simple bar chart[1].

- **Sharing** –Can others view and edit your visualization and analysis? If your workflow is essentially collaborative, then choose a tool that enables collaboration during the design, prototyping and creation process[1].

- **Persistence** – Do you think that you will have to go back and revise the visualizations you create? If so, it is important to choose a tool that is well supported and will be around for a while[1].

[1] Kristen Sosulski, Data Visualization Software: Tips for picking the right tools - http://www.kristensosulski.com/2018/11/criteria-for-evaluating-data-visualization-tools/

6

6

Data Visualisation Tools Used In This Course

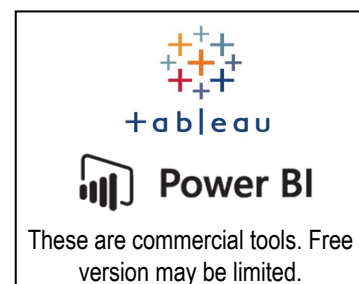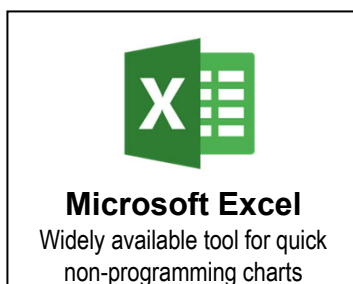## Data Visualisation Using Python

- This data visualization course is specifically designed for students with basic computer science background, as such programming skills are assumed.

- The Python-based tools are supported by numerous **visualisation** and **statistical** libraries that will facilitate both **explanatory** visualisation and **exploratory** analysis.

matplotlib / seaborn — Static Visualisation

pandas — Data Manipulation

bokeh — Interactive Visualisation

SciPy / statsmodels — Statistical Analysis

7

7

Data Visualisation Tools Used In This Course

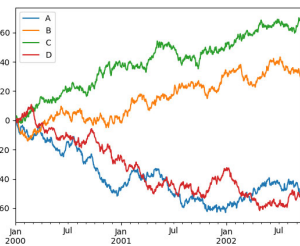## Other Data Visualisation Tools

- This course focuses on the principles and practice of good **human-centric** data visualisation **design** and not the visualisation tool per se.

- As such, students are free to use any other data visualisation tools they are familiar with, as long as it can support the required deliverables in the course.


Microsoft Excel
Widely available tool for quick non-programming charts

D3.js
Works well with web standards like HTML, CSS and SVG

tableau / Power BI
These are commercial tools. Free version may be limited.

8

8
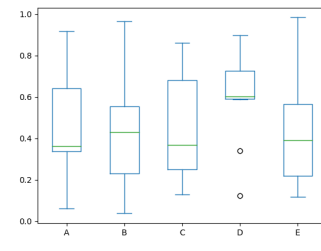
Data Visualisation Tools Used In This Course

Pandas

- Pandas is the Python data analysis library[2].
- It is a powerful, flexible and easy to use open source **data analysis** and **manipulation** tool, built on top of the Python programming language.
- Pandas can also support a variety of basic charts and plots[3] such as:
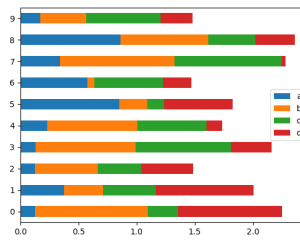
Line plots   **Stacked bar plots**   **Box plots**

[2] Pandas – Python Data Anaysis Library - https://pandas.pydata.org/
[3] Pandas – Chart Visualization - https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html

9

9

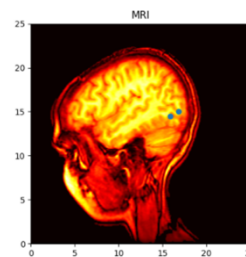# Data Visualisation Tools Used In This Course

Matplotlib

- Matplotlib[4] is a comprehensive library for creating static, animated, and interactive visualisations in Python[4] and it provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, etc.
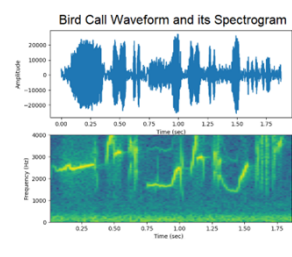- Some examples[5] of plots from Matplotlib are shown here:
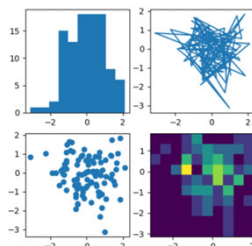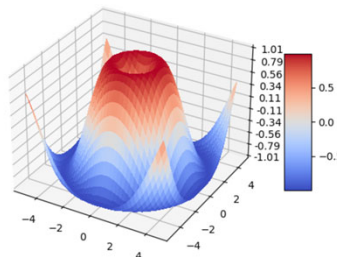
Gridded sub-plots   **3D surface plot**   **Coloured images**   **Sound & spectrogram**

[4] Matplotlib: Visualization with Python - https://matplotlib.org/
[5] Matplotlib – Sample plots in Matplotlib - https://matplotlib.org/stable/tutorials/introductory/sample_plots.html
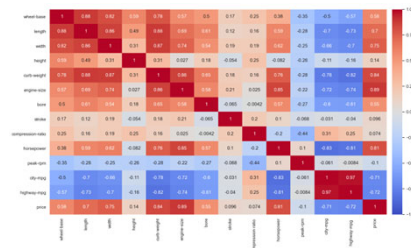
10

10

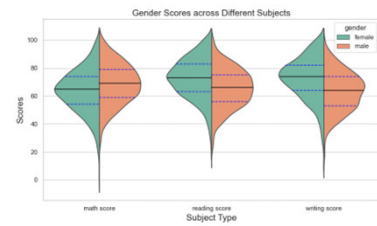# Data Visualisation Tools Used In This Course

Seaborn

- Seaborn[6] is a Python library predominantly used for making **statistical graphics**.
- It is built on top of **Matplotlib** and is closely integrated with **Pandas**'s dataframe structure, which allows sophisticated plots to be done within a single command line.
- Seaborn supports a large variety of charts and plots[7] as shown:
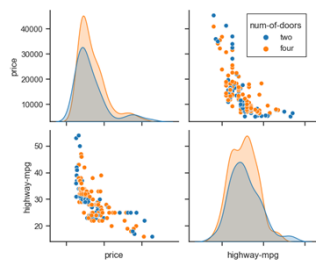
Facet plots (scatter & KDE)   **Heatmaps (Correlations)**   **Violin plots**

[6]   Seaborn – Statistical data visualization - https://seaborn.pydata.org/
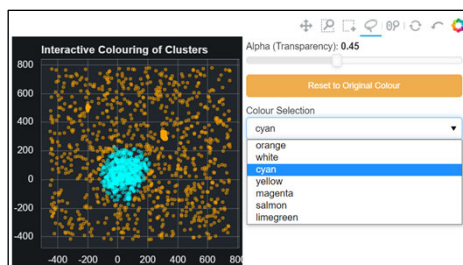[7]   Seaborn – Example Gallery - https://seaborn.pydata.org/examples/index.html
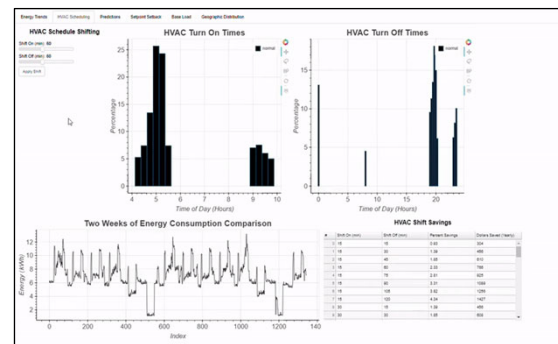
11

11

Data Visualisation Tools Used In This Course

Bokeh

- Bokeh[8] is a Python library for creating **interactive** visualisations for web browsers.
- Bokeh can produce elegant visualisation (including dashboards) that are interactive and works with large or streaming datasets.

Embedding widgets like slider & menus with Bokeh

Example Dashboard create using Bokeh [9]

[8]   Bokeh - https://bokeh.org/

[9]   Will Koehrsen – Data Visualization with Bokeh in Python –
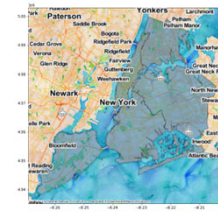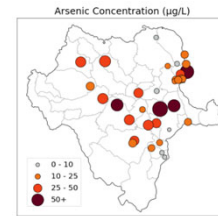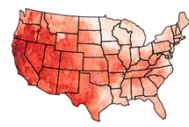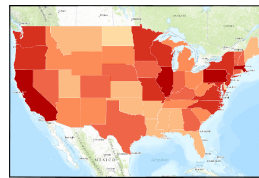https://towardsdatascience.com/data-visualization-with-bokeh-in-python-part-one-getting-started-a11655a467d4

12

12

Other Python Data Visualisation Libraries

Visualising Geospatial Data

- The are many useful visualisation libraries in Python that allow us to visualise geographical and location-based information.

- **GeoPandas**[10] - extends the datatypes used by Pandas to allow spatial operations on geometric types. Its goal is to make working with geospatial data in Python easier.

- **Geoplot**[11] – is a high-level Python geospatial plotting library. It is built on top of matplotlib with the specific purpose of doing cartographic plots.

[10] GeoPandas - https://geopandas.org/
[11] Geoplot: geospatial data visualization – https://residentmario.github.io/geoplot/index.html

13

13

Other Python Data Visualisation Libraries

Visualising Textual Data

- The are other visualisation libraries in Python that allow us to visualise various properties in text-based data.

- **WordCloud**[12] – is a text visualization library that can analyse the frequency of words and correlates the size and opacity of a word to its frequency within a body of text.

- It has an interesting feature that allows an image mask to define the shape over which the word cloud is populated. This can serve as a visual context to the document being analysed[13].
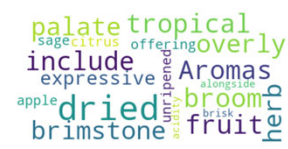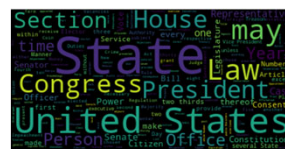
[12] WordCloud for Python documentation – https://amueller.github.io/word_cloud/
[13] Duong Vu, Generating WordClouds in Python – https://www.datacamp.com/community/tutorials/wordcloud-python

14

14

Summary

Visualisation Tools

- There are numerous types of visualisation tools available, both the free open-source variety and sophisticated commercial tools.

- Choosing a suitable tool depends on the purpose of the visualisation. Factors related to the **design** (e.g. collaborative, fast turnaround) and **deployment** environments (e.g. interactive, web-based) should also be considered when making the choice.

- This course has adopted open source tools based on the **Python programming** environment as it supports many different visualisation libraries with good support for data wrangling and statistical analysis.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

15

15

References for Visualisation Tools

[1]   Kristen Sosulski, Data Visualization Software: Tips for picking the right tools – http://www.kristensosulski.com/2018/11/criteria-for-evaluating-data-visualization-tools/

[2]   Pandas – Python Data Anaysis Library - https://pandas.pydata.org/

[3]   Pandas – Chart Visualization - https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html

[4]   Matplotlib: Visualization with Python - https://matplotlib.org/

[5]   Matplotlib – Sample plots in Matplotlib - https://matplotlib.org/stable/tutorials/introductory/sample_plots.html

[6]   Seaborn – Statistical data visualization - https://seaborn.pydata.org/

[7]   Seaborn – Example Gallery - https://seaborn.pydata.org/examples/index.html

[8]   Bokeh - https://bokeh.org/

[9]   Will Koehrsen – Data Visualization with Bokeh in Python – https://towardsdatascience.com/data-visualization-with-bokeh-in-python-part-one-getting-started-a11655a467d4

[10] GeoPandas - https://geopandas.org/

[11] Geoplot: geospatial data visualization – https://residentmario.github.io/geoplot/index.html

[12] WordCloud for Python documentation – https://amueller.github.io/word_cloud/

[13] Duong Vu, Generating WordClouds in Python – https://www.datacamp.com/community/tutorials/wordcloud-python

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Note: All online articles were accessed between Jul – Aug 2021

16

16