



Data visualization using Python

Prasun Kumar Gupta

Scientist/Engineer-'SE', Geoinformatics Department
Geospatial Technology and Outreach Program Group
Indian Institute of Remote Sensing (ISRO), Dehradun
prasun@iirs.gov.in

9/21/2022

1

Sources: Multiple sources (mostly specified at the bottom of the slide)



Data visualization

- Very contemporary subject
- Making static or interactive plots (2D, 3D)
- Primarily for exploring data
- For example, helping identify outliers, needed data transformations, or coming up with ideas for models.
- Grammer of Graphics
 - [Wilkinson's Grammar of Graphics](#)
 - [Bertin's Visual Variables](#) etc.
- Several options: Tableau, PowerBI, D3, Vega, ggplot2 etc.



Python vizualization ecosystem

- Standard charts (e.g. line chart, bar chart, scatter plot):
 - Matplotlib, Pandas, Seaborn, ggplot, Altair, ...
- Thematic maps
 - Folium, Basemap, Cartopy, Iris, ...
- Other visualisations
 - Vega / Vegalite, Bokeh (interactive plots), Mayavi (3D), plotly, paraview, ...



Matplotlib

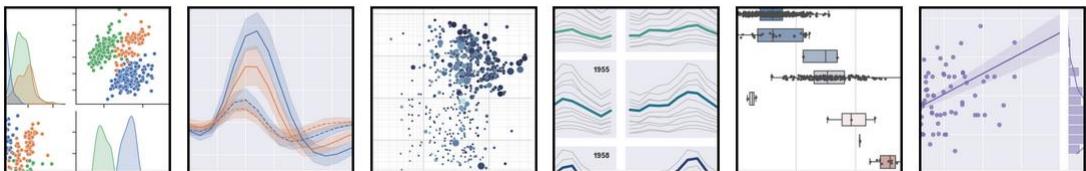
- First released by John D. Hunter, an American neurobiologist in 2003
- Python '**most popular**' 2D plotting library which produces publication quality figures in a variety of hardcopy formats
- Line plots, scatter plots, barcharts, histograms, pie charts etc.
- Relatively low-level; some effort needed to create advanced visualization
- Number of add-on toolkits, such as **mplot3d** for 3D plots and **basemap** for mapping and projections



Seaborn

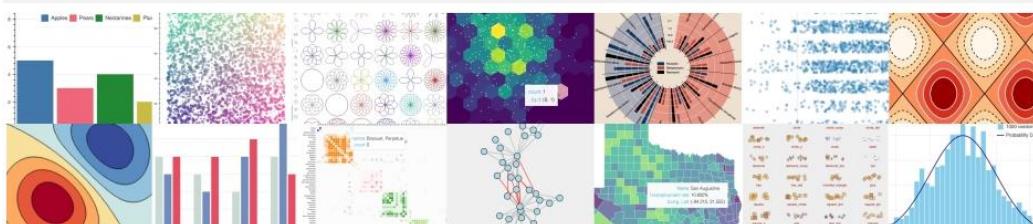
- Based on matplotlib
- Provides high level interface for drawing attractive statistical graphics
- Similar (in style) to the popular ggplot2 library in R

seaborn: statistical data visualization



Bokeh

- Bokeh (pronounced like bouquet) library: browser-based
 - visualizations
 - Interactive plots
- Requires understanding of HTML and Javascript
- Python is used to generate the visualization, HTML is used to display it.



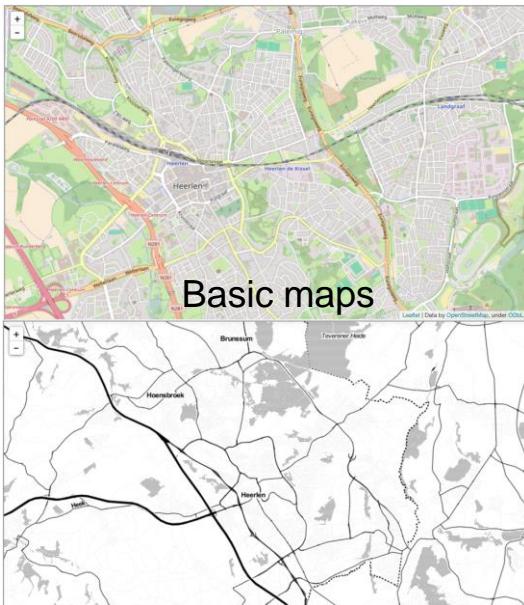


Folium - *for thematic maps*

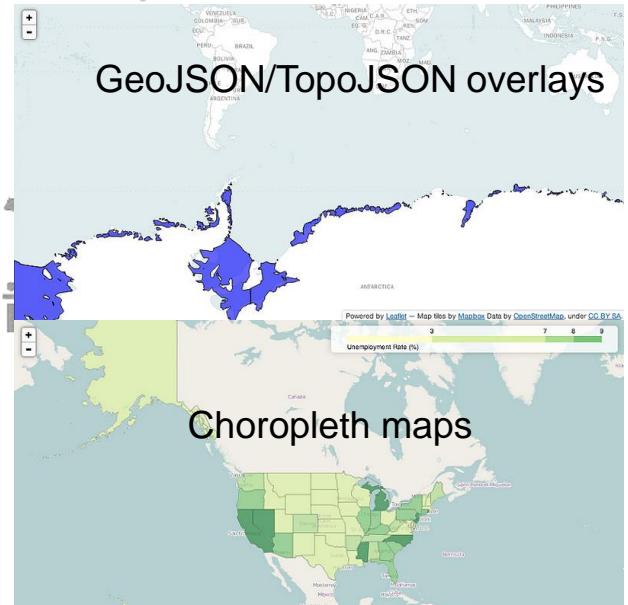
- Visualization where statistical information with a spatial component is shown.
- Other libraries are: Basemap, Cartopy, Iris
- Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library.
- Manipulate your data in Python, then visualize it in on a Leaflet map via Folium.
- Built-in tilesets from OpenStreetMap, Mapbox, and Stamen
- Supports GeoJSON and TopoJSON overlays



Folium examples



Basic maps



Choropleth maps

GeoJSON/TopoJSON overlays



Vega Lite

- Facilitate exploratory data analysis with an expressive yet concise language to specify interactive multi-view graphics
- Try Online - <https://vega.github.io/editor/>
- or install and use in Python
- conda install -c conda-forge vega
- Good Documentation - <https://vega.github.io/vega-lite/docs/>
<https://youtu.be/9uaHRWj04D4>



Many others

- Matplotlib may come across as having less functionality as compared to the excellent **ggplot2** and trellis packages
- There are many other packages such as Chaco, PyQwt, Veusz, gnuplot-py, biggles
- Many of these libraries may not be under active development
- Some of them are maintained only as part of much larger applications
- General trend toward web-based technologies and away from desktop graphics



Time for some hands-on
using Python



Thank You

prasun@iirs.gov.in

