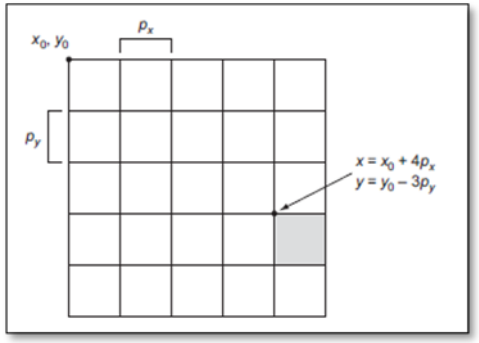
Introduction to Numpy

Chaitanya K S
Deputy Director
National Water Academy
Central Water Commission, Pune

Raster Data

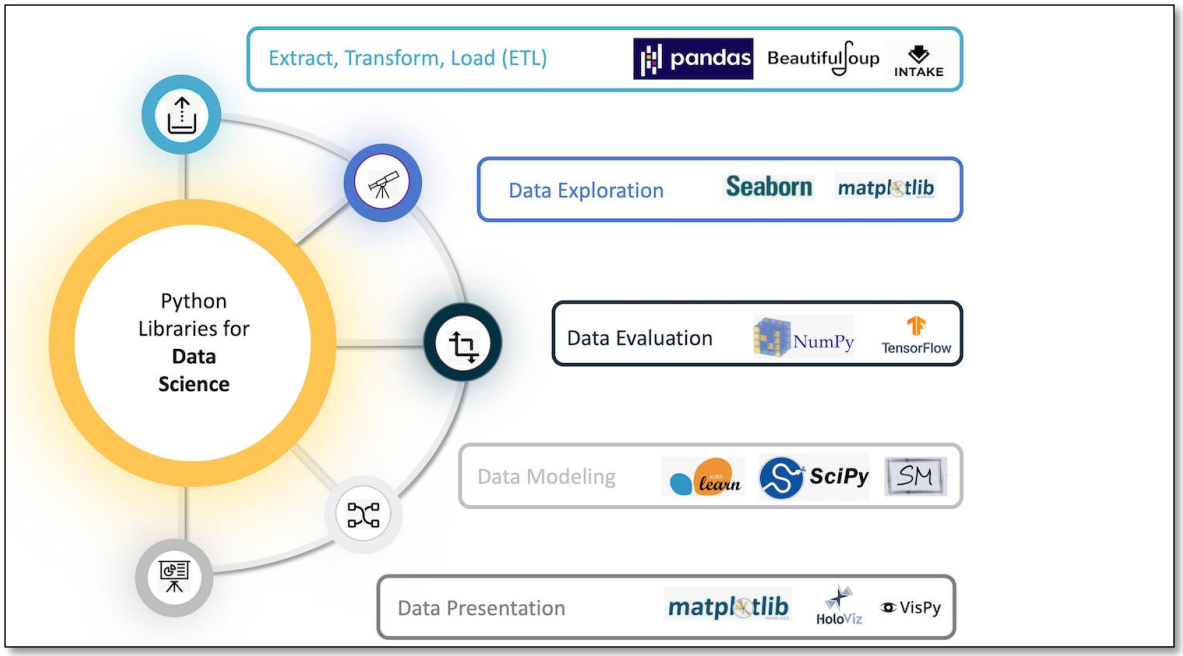
- Hydrological parameters vary spatially in a continuous manner and therefore better represented in raster data format.
- Raster datasets are stored as numerical arrays of rows and columns (2D arrays) with each pixel representing space and the value of the parameter
- Multispectral data – multiple arrays of same size called bands. A collection of bands becomes a 3D array
- While for vector data it is enough to know spatial reference system, for raster data, pixel size, coordinates of origin and the amount by which the dataset is rotated are needed to determine the rest of the coordinates



Source: 'Geoprocessing with Python' by Chris Garrard

Numpy Module

- Numpy is an open-source fundamental package for scientific computing with Python



NumPy

https://numpy.org/

- NumPy lies at the core of a rich ecosystem of data science libraries.
- A typical exploratory data science workflow might look like:
  - **Extract, Transform, Load:** Pandas, Intake, PyJanitor
  - **Exploratory analysis:** Jupyter, Seaborn, Matplotlib, Altair
  - **Model and evaluate:** scikit-learn, statsmodels, PyMC3, spaCy
  - **Report in a dashboard:** Dash, Panel, Voila

Numpy Module

- Numpy is designed for processing large arrays of data – useful for handling raster datasets
- At its core is the ndarray object which encapsulates n-dimensional arrays of homogeneous data types, with many operations being performed in compiled code for performance.
- NumPy arrays Vs standard Python sequences:
  - They have a fixed size at creation
  - Elements in a them are to be of the same data type
  - They facilitate advanced mathematical and other types of operations on large numbers of data
  - A growing plethora of scientific and mathematical Python-based packages are using them
  - With them code is more concise and easier to read and resembles standard mathematical notation
  - Without vectorization, code would be littered with inefficient and difficult to read for loops

https://numpy.org/doc/stable/user/whatisnumpy.html#.

Thank you