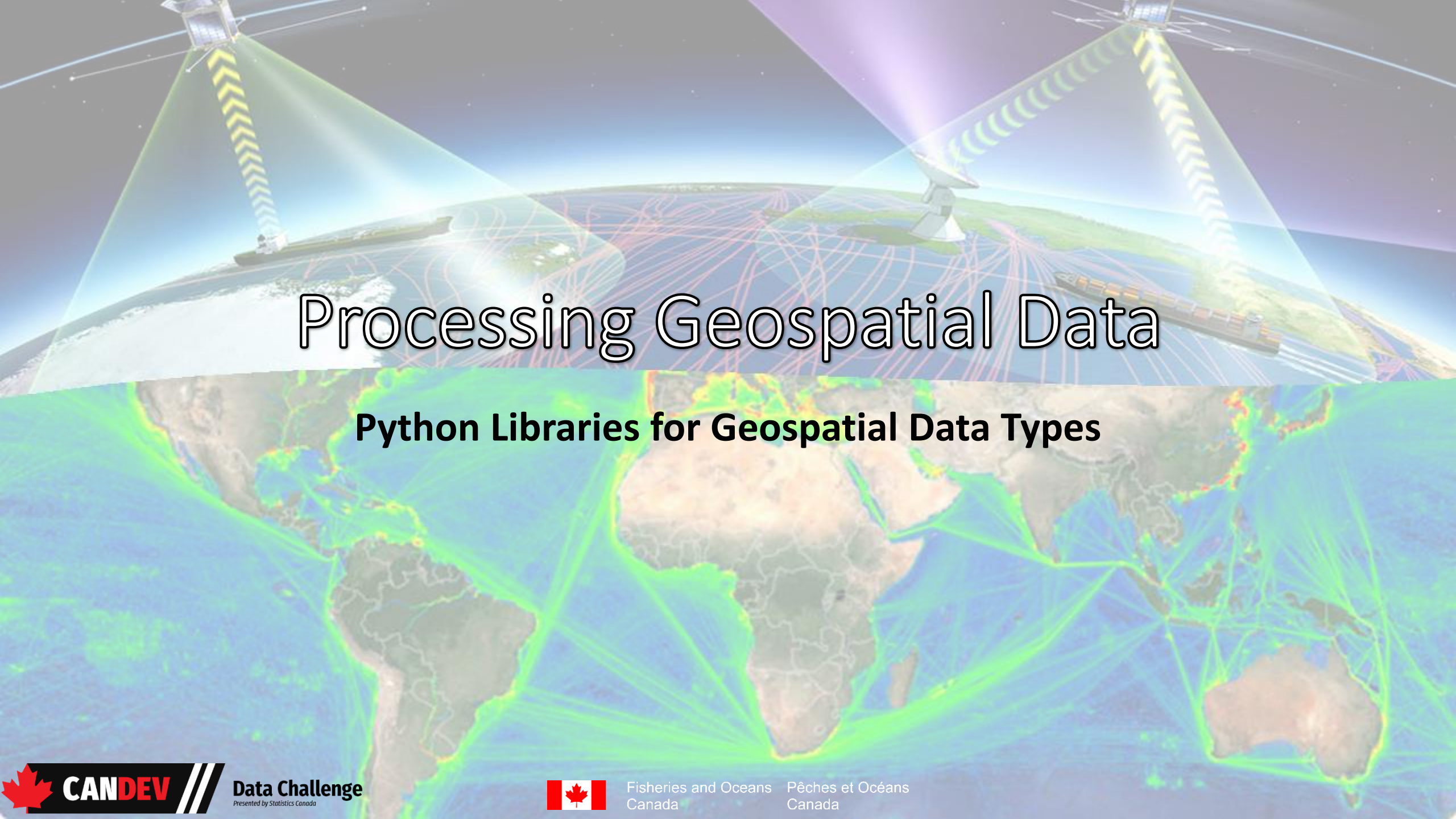


Processing Geospatial Data

Python Libraries for Geospatial Data Types

CANDEV Data Challenge
Presented by Statistics Canada

Fisheries and Oceans Canada
Pêches et Océans Canada

Geospatial Data in the Wild

Geospatial data: Any form of geographically-referenced data. In addition to its geographic component, the data may contain non-geographic information.

- Many real-world problems involve geospatial data
  - Ex: GPS, mapping, transportation planning, collision avoidance, terrain assessment, etc.
  - Often, various geographically-referenced data sources must be processed, combined and analyzed

- **DFO Use Case**: Teams are tasked with the design of a model to capture the risk of harmful interactions between ships and whales

Data Source	Format
Historical ship track data	Tabular (geo-referenced)
Whale habitat suitability data	Raster
Designated slow zones	Vector
Marine protected areas	Vector

Workshop Topics

- The purpose of the workshop is to cover libraries that will allow for quick and easy processing of geospatial data sources
  - All content will use Python as the programming language

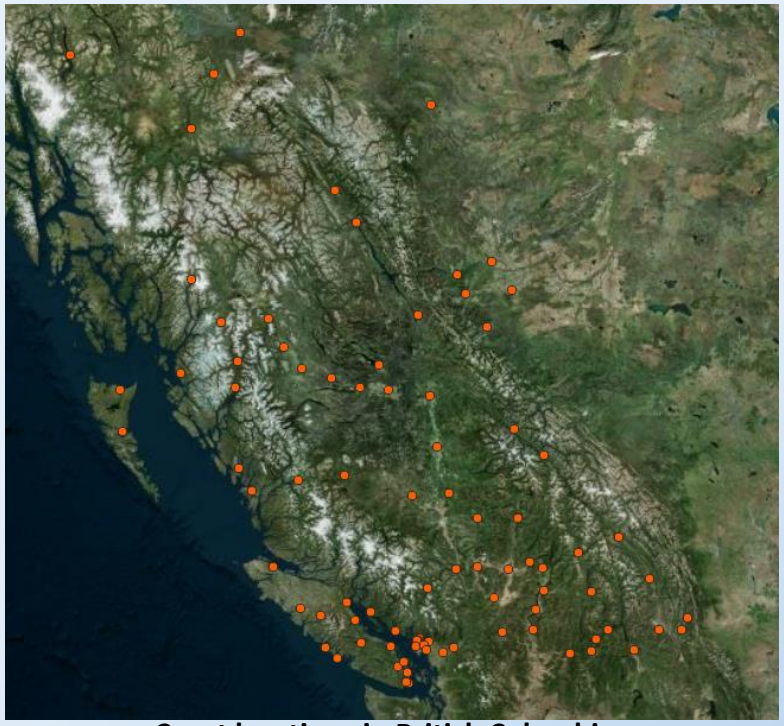
Topic	Library
Processing tabular data	Pandas
Processing vector data	Shapely
Processing raster data	Rasterio
Visualizing geospatial data	HoloViews

Processing Tabular Data – Pandas

Tabular data: Structured data over a fixed set of fields (columns) where each record (row) contains a value (possibly null) for each field.

Placemark_name	Latitude	Longitude
CAMPBELL RIVER LAW COURTS	50.027912	-125.247125
COURTENAY LAW COURTS	49.689531	-124.999417
DUNCAN LAW COURTS	48.777968	-123.711027
GANGES	48.859991	-123.509222
GOLD RIVER	49.781045	-126.047610

1. Load a CSV file into a DataFrame
2. View the structure of the data
3. Access the data by columns or by rows
4. Access the data based on query conditions



Court locations in British Columbia

Alternative Libraries:

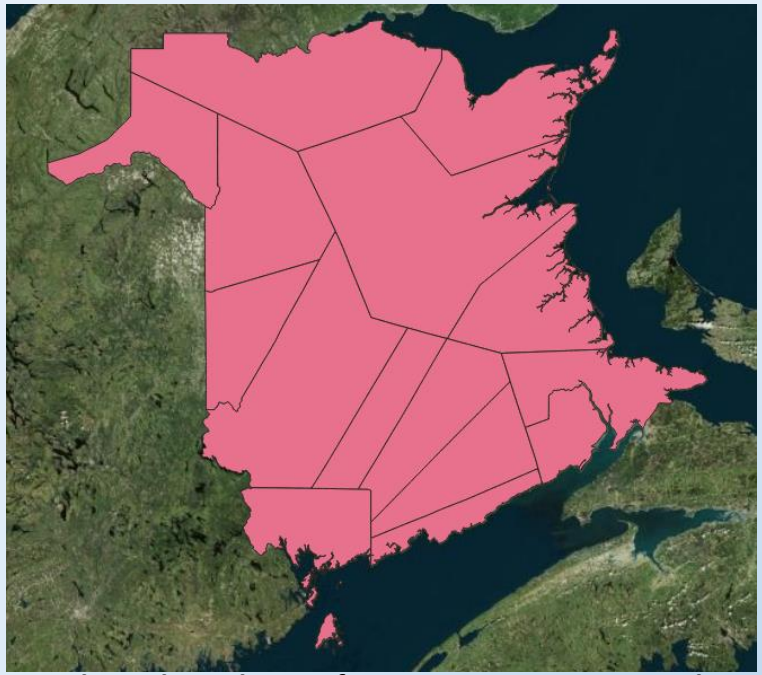
- NumPy

Processing Vector Data – Shapely

Vector data: Geographic data represented as collections of points, lines and polygons, each of which are broken down into ordered lists of vertices.


```
Out[2]: {'type': 'FeatureCollection',
 'features': [{'type': 'Feature',
   'properties': {'usernum': '5001.0',
    'area': '12222346789.7',
    'dispx': '2497477.1',
    'keyword': '500100009',
    'key': '500100009',
    'fre_name': 'Northumberland',
    'date': '2004_01_01',
    'sourceid': 'AAVDADTIAA04',
    'fcode': 'DLBNCO_C',
    'eng_name': 'Northumberland',
    'dispy': '7551277.35'},
   'geometry': {'type': 'MultiPolygon',
    'coordinates': [[[[-67.04959821852964, 47.361774783365014],
       [-67.03619645680641, 47.364684733286936],
       [-66.99997383594139, 47.37254745056295],
       [-66.9739655329058, 47.37818005350879],    ...
```


1. Load shapefiles and GeoJSON files
2. Access geographic features (points, lines, polygons)
3. Check for relations between of geographic features
4. Calculate the distance between geographic features



Polygon boundaries of counties in New Brunswick

Alternative Libraries:

- Fiona
- GeoPandas

Processing Raster Data – Rasterio

Raster data: Geographic data represented as a grid of equal-sized cells in which each cell is assigned a value pertaining to the area it covers.

```
[[ 2.97   2.617  8.903 ... 26.107  0.    14.143]
 [ 3.333  6.297  8.903 ...  8.6   21.717 10.   ]
 [30.317  3.877  0.    ...  0.     0.    21.717]
 ...
 [ 2.573 12.953  9.2   ...  0.     8.903 19.15 ]
 [ 0.     7.57  21.943 ... 11.817  1.877 15.57 ]
 [ 0.     0.    12.953 ... 14.903 14.903  6.667]]
```

1. Load raster files
2. Look up raster indices for coordinates
3. Query raster values
4. Perform raster arithmetic

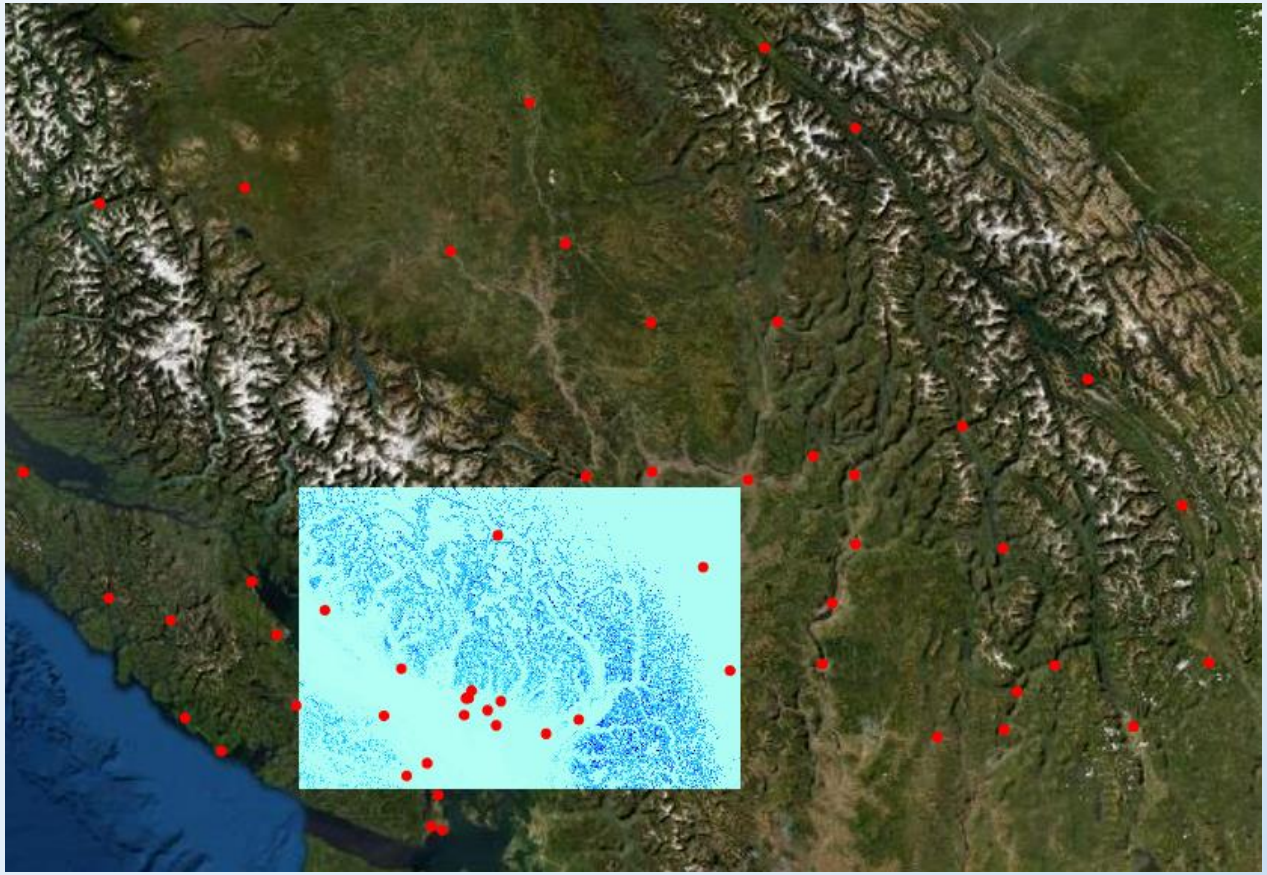
Raster of tree density in British Columbia

Alternative Libraries:

- GDAL

Visualizing Geospatial Data – HoloViews

1. Load and view basemaps
2. Re-project and plot vector data
3. Re-project and plot raster data
4. Apply basic styling to plotted data



Visualization of court locations and tree density on the same map

Alternative Libraries:
- Cartopy
- Power BI