

Big Data Analytics

Geospatial Perspectives

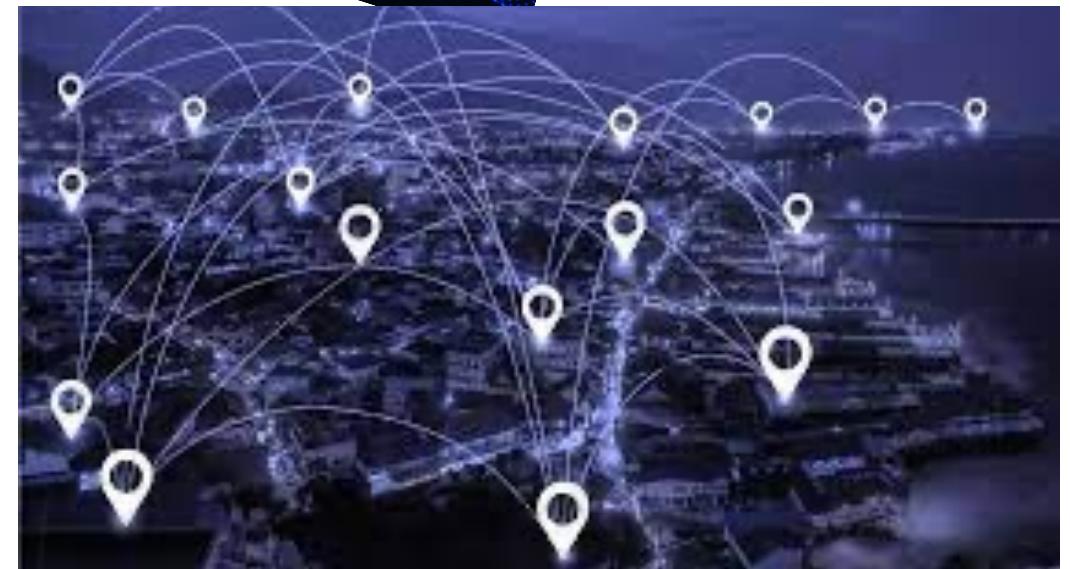
Samson Warkaye

IGSSA Geomatics & Geodessy: RS and GIS PhD student

November 3, 2023

Outline

1. Big Data Analytics (BDA)
2. Geospatial BDA
3. Health use case



Big Data are generated



1. Big. Data Analytics



Big Data Management

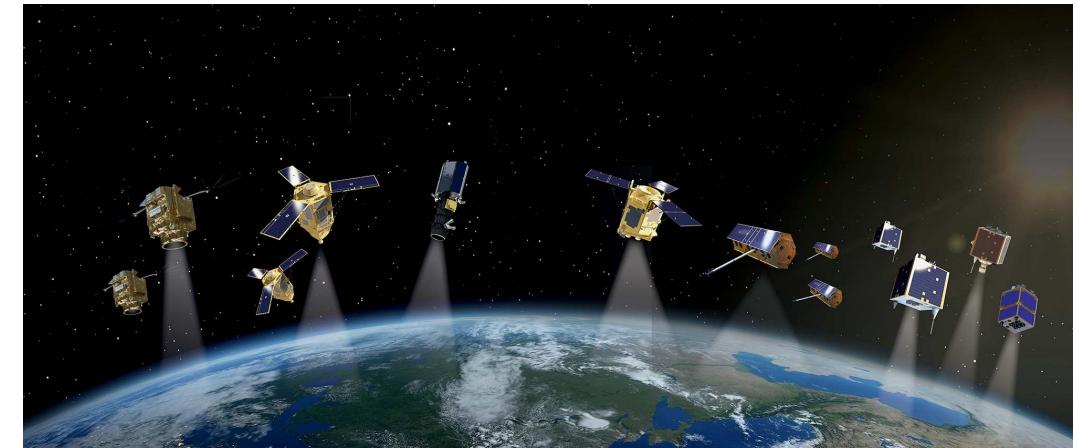
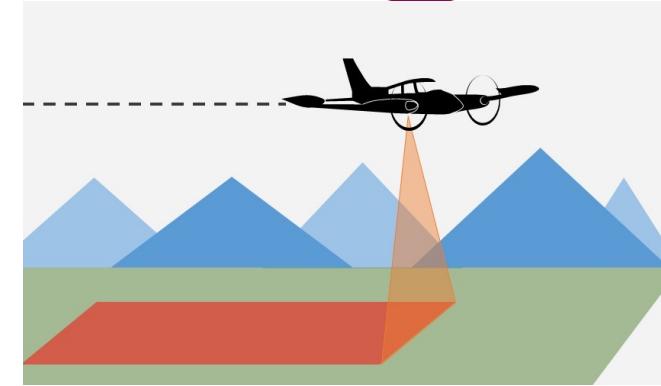


Geospatial Data

Geographic location: Place



Data
Sources



Cont...

❖ **Spatial data science:** methods focused on processing, managing, analyzing, and visualizing spatial big data

❖ **The science and technology is expanding:** Voluntered Geographic Information (VGI) and remote sensing

VGI: Open Street Map



Examples

Investigate the effects of risk factors such as air pollution on **cardiovascular disease**

- ❖ **Geographic variables:** road, LU, OSM
- ❖ **Environmental variables**

Methodology

- Data mining: ML - RF
- Predict particulate matter in air pollution

Advs. Improvement compared to using IDW

Handles

- ✓ Spatial & Temporal resolutions
- ✓ Computational efficiency

Technologies

- Spark
- Hadoop

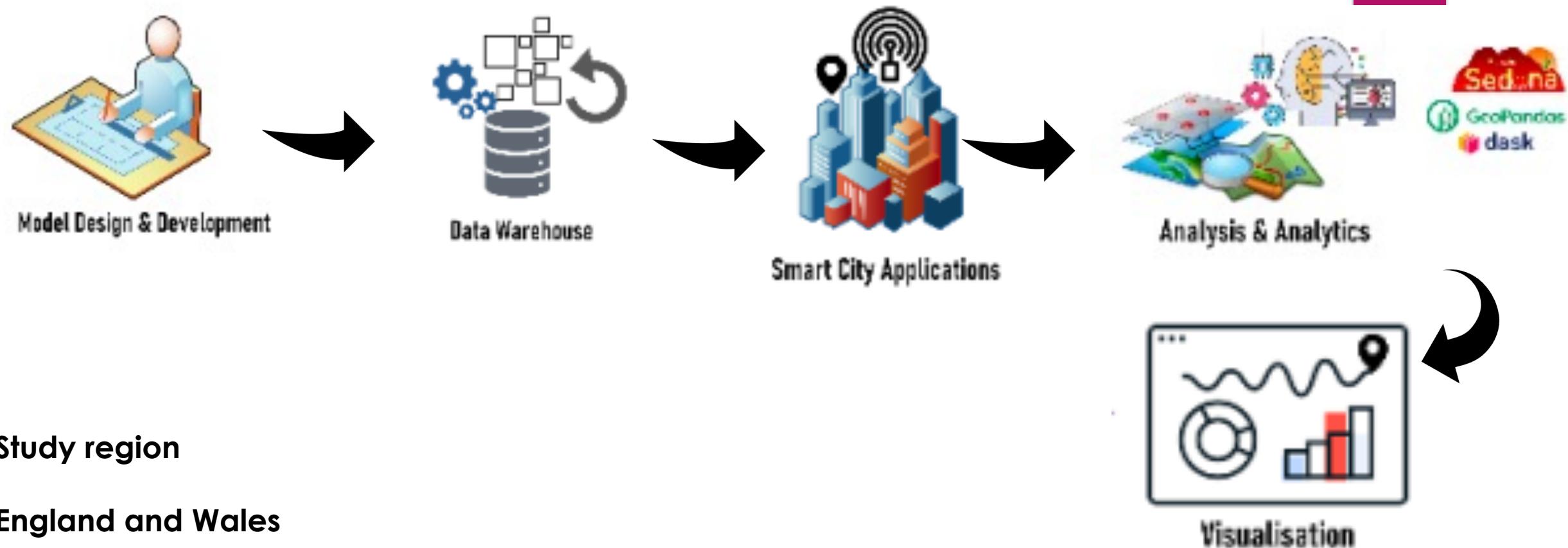
Reference

VoPham T., Hart J.E., Laden F., and Chiang Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*, 17:40

- ❖ Urbanization: despite the benefits cause environmental problems – ecological degradation
- ❖ Smart city paradigms: smart infrastructure and smart environment
- ❖ Smart city services are rely on location-based data
- ❖ Smart city: is a modern city that utilizes information, technologies to improve its services and management and solve problems affecting the city.



Geospatial Big Data Administration Model



Apache Sedona: cluster computing for processing large-scale spatial data
Dask: parallel computing, scales the python libraries

Data

- Energy performance certificates
- Ordnance survey: Unique property reference number
- Building

Data processing

- ❖ Data converted to Geoparquet table format
- ❖ Spatial clustering
- ❖ Spatial join

SQL statements are used to filter energy rates

Datashader used to represent the result

