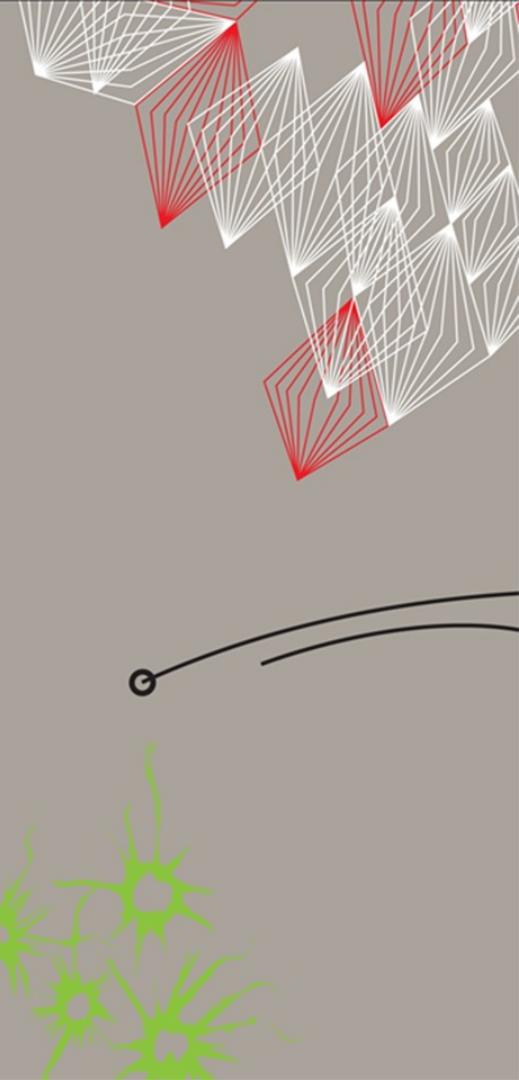


GEOSPATIAL DATA ANALYSIS



Mahdi KHODADADZADEH
October 2021

WHAT IS A GIS?

- **Geographic Information System (GIS):** A computer-based system for the collection, storage, organization, maintenance, and analysis of spatially-referenced data, and the output of spatially-referenced information.
- **Data:** Any collection of related facts; the basic elements of information.
- **Information:** Data that have been processed to be useful

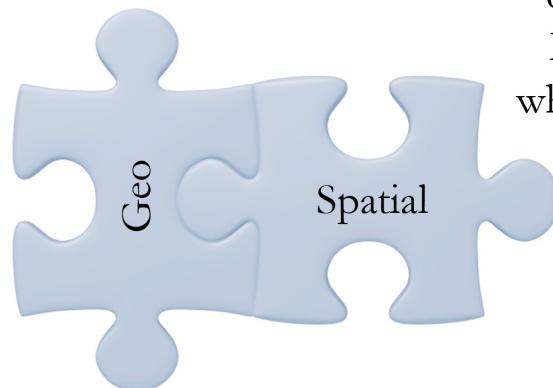
WHAT TYPES OF QUESTIONS?

- GIS provides answers to "who", "what", "where", and "when" questions
 - Where are particular features found?
 - What geographic patterns exist?
 - Where have changes occurred over a specified time period?
 - Where do certain conditions apply?
 - When is the spring onset?

WHAT IS GEOSPATIAL DATA?

- NASA says geospatial means the distribution of something in a geographic sense; it refers to entities that can be located by some co-ordinate system.
- Geospatial data are related to Earth and give geographically referenced information.

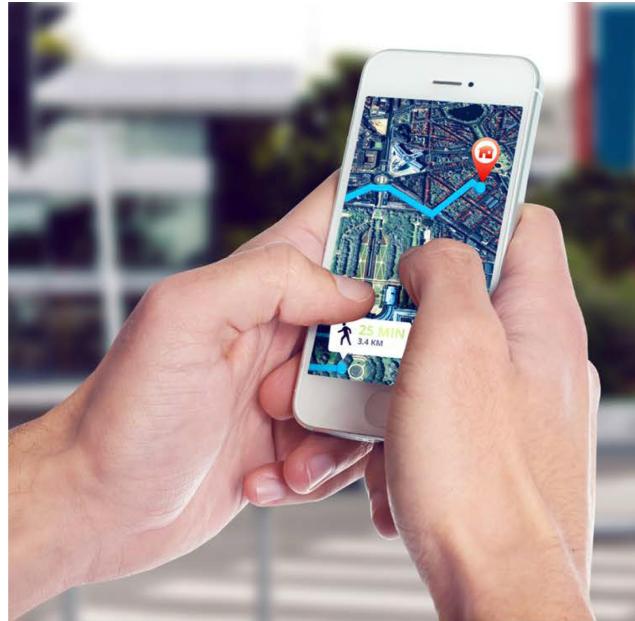
from Greek 'gaya'
meaning **Earth**



originated from
Latin 'spatium',
which means **space**

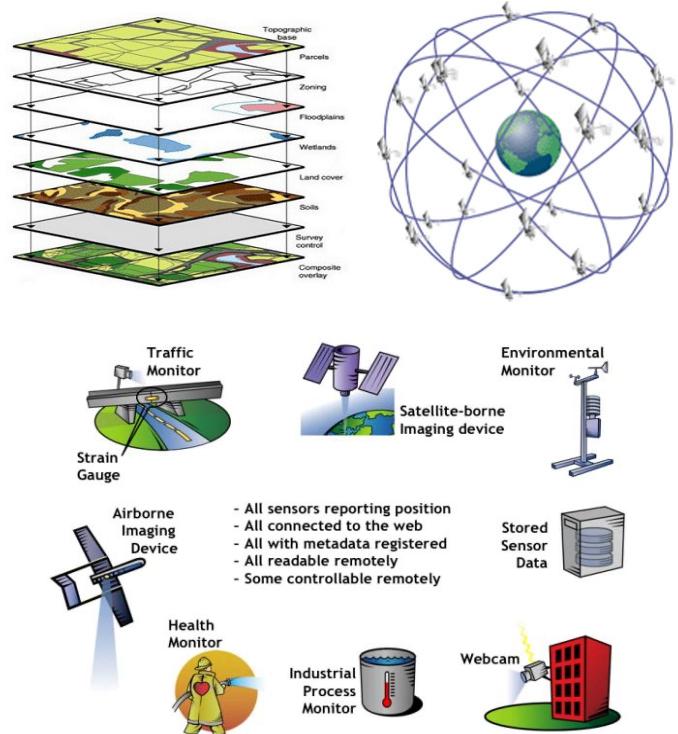
WHERE?

- Geospatial Data are everywhere!



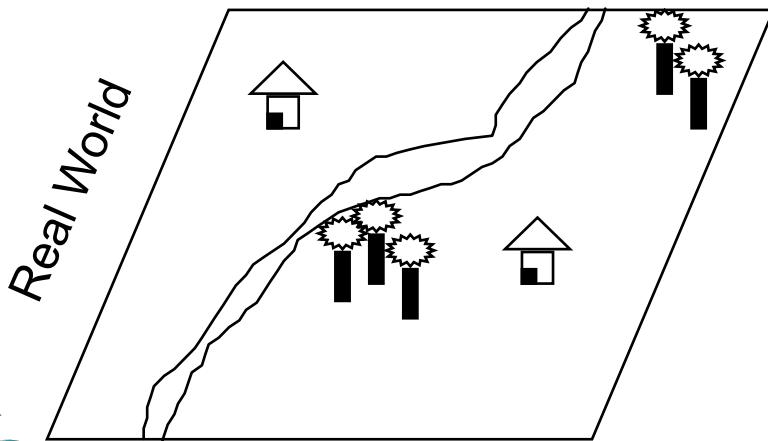
DATA AND DATASETS

- GIS, GPS and sensor data
 - Raster: satellite images, model outcomes (e.g. climate simulations), drone images, etc.
 - Vector: trajectory data (e.g. Uber data), geo-located twitter data, etc.



RASTER DATA

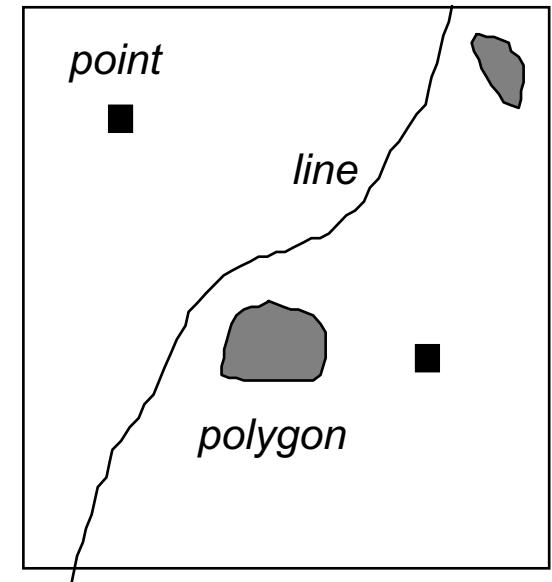
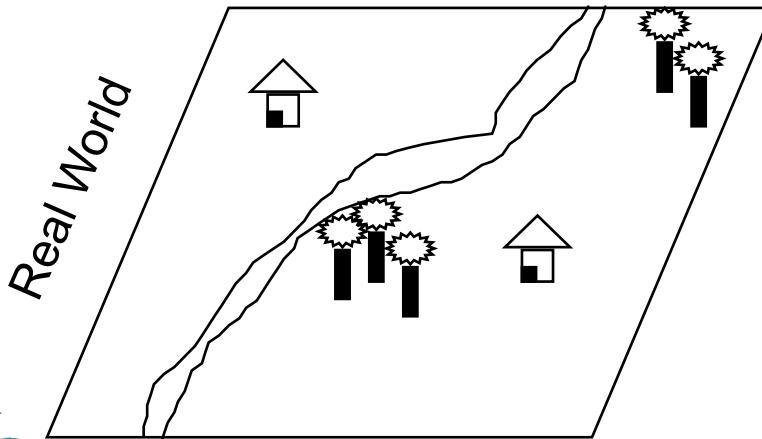
- Location is referenced by a grid cell in a rectangular array (matrix)
- Attribute is represented as a single value for that cell



	0	1	2	3	4	5	6	7	8	9
0						R	T			
1					R			T		
2	H				R					
3					R					
4			R	R						
5			R							
6	R		T	T	H					
7	R		T	T						
8	R									
9	R									

VECTOR DATA

- point (node): 0-dimension
 - single x,y coordinate pair
- line (arc): 1-dimension
 - two (or more) connected x,y coordinates
- polygon : 2-dimensions
 - four or more ordered and connected x,y coordinates



COPERNICUS -- THE SENTINELS



Sentinel 1 (A/B/C/D)
SAR Imaging

All weather, day/night applications,
interferometry



Sentinel 2 (A/B/C/D)
Multispectral Imaging

Land applications: urban, forest, agriculture, ...
Continuity of Landsat, SPOT



Sentinel 3 (A/B/C/D)
Ocean & Global Land Monitoring

Wide-swath ocean colour, vegetation, sea/land
surface temperature, altimetry



Sentinel 4 (A/B)
Geostationary Atmospheric

Atmospheric composition monitoring, pollution;
instrument on MTG satellites



Sentinel 5 (A/B/C) & Precursor
Low-Orbit Atmospheric

Atmospheric composition monitoring;
instrument on MetOp-SG satellites



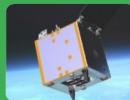
Sentinel 6
Jason CS (A/B)

Altimetry reference mission

COPERNICUS – CONTRIBUTING MISSIONS

Optical High & Very High Resolution

DMC



Pléiades



RapidEye



Deimos-2



SPOT (HRS)



Optical Medium & Low Resolution

SPOT



PROBA-V



and many more ...

Synthetic Aperture Radar

Cosmo SkyMed



Radarsat



TerraSAR-X
Tandem-X



Altimetry

Cryosat



Jason



Atmosphere

MetOp



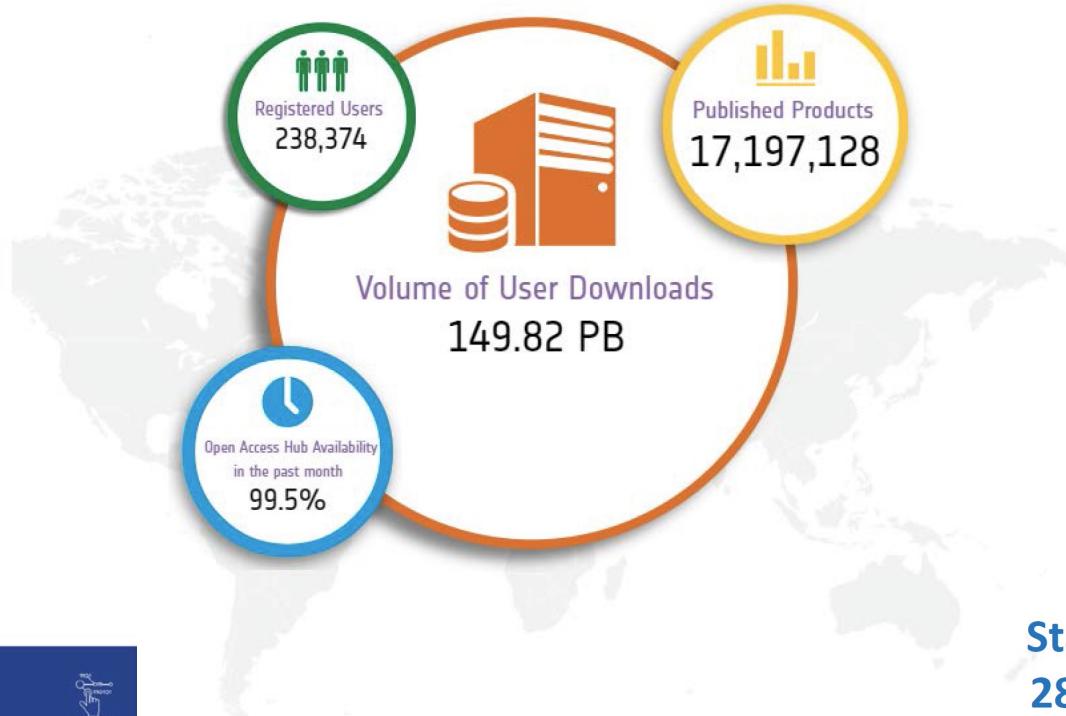
MSG



FREE EO DATA AND INFORMATION SERVICES

Registered Sentinel Users

The real number of users is much higher but unknown due to the free, full & open data policy.



Statistics on
28 Jun 2019



UNIVERSITY OF TWENTE.

CITIZEN SCIENCE



OpenStreetMap

budburst
a project of the Chicago Botanic Garden

About Budburst Get Started Plants

Welcome to Budburst

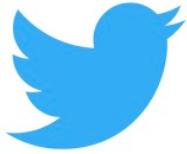
Budburst brings together researchers, educators, gardeners, and citizen scientists on a shared journey to uncover the stories of plants and animals affected by human impacts on the environment. We hope that sharing these stories will increase appreciation of plants and the natural world and inspire conservation action.

[GET STARTED >](#)



UNIVERSITY OF TWENTE.

Millions of
spatio-temporal
observations!



de NATUURKALENDER

Beuk, Bladontploeling, 2011

- eerste waarneming: 2-4, 38 waarnemingen
- 2-4 t/m 8-4
- 9-4 t/m 15-4
- 15-4 t/m 22-4
- 23-4 t/m 29-4
- 30-4 t/m 6-5
- 7-5 t/m 13-5
- 14-5 t/m 20-5
- 21-5 t/m 27-5
- 26-5 t/m 3-6
- 4-6 t/m 14-6
- laatste waarneming

Aantal waarnemingen

Gen. min. temperatuur

uitleg

Keuzemogelijkheden

Kaart zonder details

2011

Bladontploeling

de natuurkalender

Floron

Geo-Wiki: Earth Observation & Citizen Science

Launch Geo-Wiki

Register, login or enter Geo-Wiki as guest

Get involved

Our results

News Team About Projects Tools Data Publications Contact

The global land use crowdsourcing tool

Waarneming.nl

Meld een tekenbeet

Bent u gebeten door een teek of heeft u een erythema migrans ^(?)? Klik dan op onderstaande buttons en doe mee aan onderzoek naar tekenbitten en de ziekte van Lyme.

Tekenbeet
melden ▶

wat is een tekenbeet?

Erythema
migrans melden ▶

wat is een erythema
migrans?

Als u een recente tekenbeet meldt kunt u uitgenodigd worden de teek op te sturen. U vult dan in anderhalf jaar een aantal korte vragenlijsten in. Binnen 10 maanden hoort u of de teek besmet was met de bacterie die de ziekte van Lyme kan veroorzaken.

Meer informatie:

- Het onderzoek
- Waarom meedoen

Teken verzameld via tekenradar.nl in 2012

- geen besmetting
- besmet met *Borrelia burgdorferi*
- besmet met *Borrelia miyamotoi*
- besmet met *Borrelia miyamotoi* en *Borrelia burgdorferi*



wachting

verwachting van
sniveau van het
uur.

ng:

+10 dgn

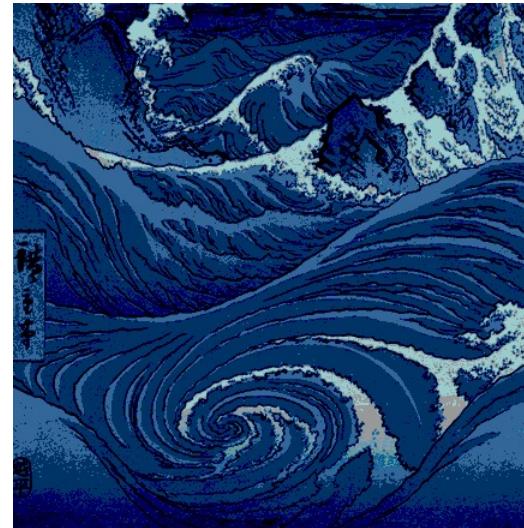
itleg tekenverwachting

kenverwachting

- hoog
- zeer hoog

DATA AND DATASETS

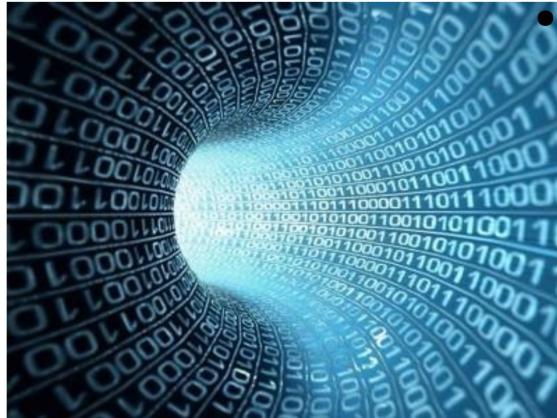
- Although lack of data in certain regions or on certain topics might still be an issue
...
▪ ... the real challenge in normal circumstances is to develop **innovative** ways to deal with all the available data and datasets.



The tsunami of data

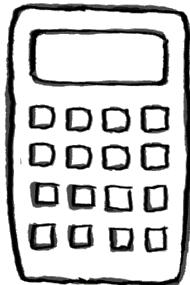
BIG GEODATA

- Every day we create 2.5 trillion (10^{18}) bytes of data. 80 % of these are already georeferenced or can be.
- It's a huge dataset, equal to a DVD tower that goes from the Earth to the Moon every day.



COMPUTER POWER

Huge development of computer power ...



~ 10 flops

Floating point Operations Per Second



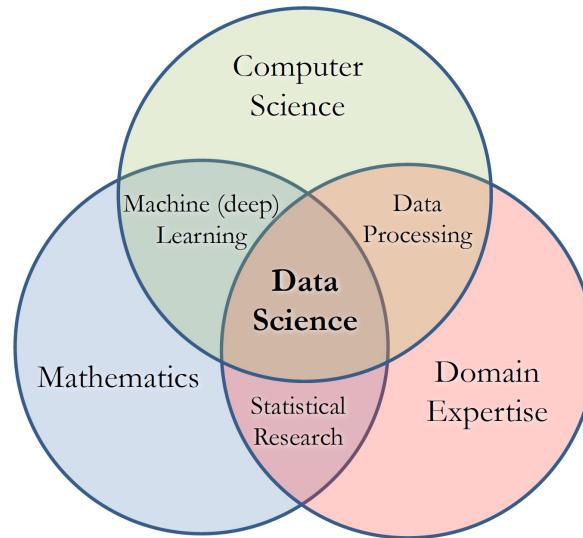
Jaguar -- Oak Ridge National Laboratory
Tennessee (USA)

1.75 petaflops

peta = 10^{15}

WHAT IS DATA SCIENCE?

- Data science is the field of study that combines domain expertise, programming skills, and knowledge of math and statistics to extract meaningful insights from data.



WHAT DOES GEOSPATIAL DATA SCIENTIST DO?

- Geospatial data scientist
 - analyses vast geospatial datasets to extract useful insights about the events happening on our planet.
 - develops information about features, objects, and classes on Earth's surface and/or near Earth's surface.
 - finds interesting and previously unknown spatial patterns.
 - visualizes large image and location datasets and query them.
 - uses geospatial datasets for making predictions.

GEODATA – APPLICATIONS

Urban Planning, Management & Policy

- Zoning, subdivision planning
- Land acquisition
- Economic development
- Code enforcement
- Housing renovation programs
- Emergency response
- Crime analysis
- Tax assessment

Environmental Sciences

- Monitoring environmental risk
- Modeling storm water runoff
- Management of watersheds, floodplains, wetlands, forests, aquifers
- Environmental Impact Analysis
- Hazardous or toxic facility siting
- Groundwater modeling and contamination tracking

Political Science

- Redistricting
- Analysis of election results
- Predictive modeling

Civil Engineering/Utility

- Locating underground facilities
- Designing alignment for freeways, transit
- Coordination of infrastructure maintenance

Business

- Demographic Analysis
- Market Penetration/ Share Analysis
- Site Selection

Education Administration

- Attendance Area Maintenance
- Enrollment Projections
- School Bus Routing

Real Estate

- Neighborhood land prices
- Traffic Impact Analysis
- Determination of Highest and Best Use

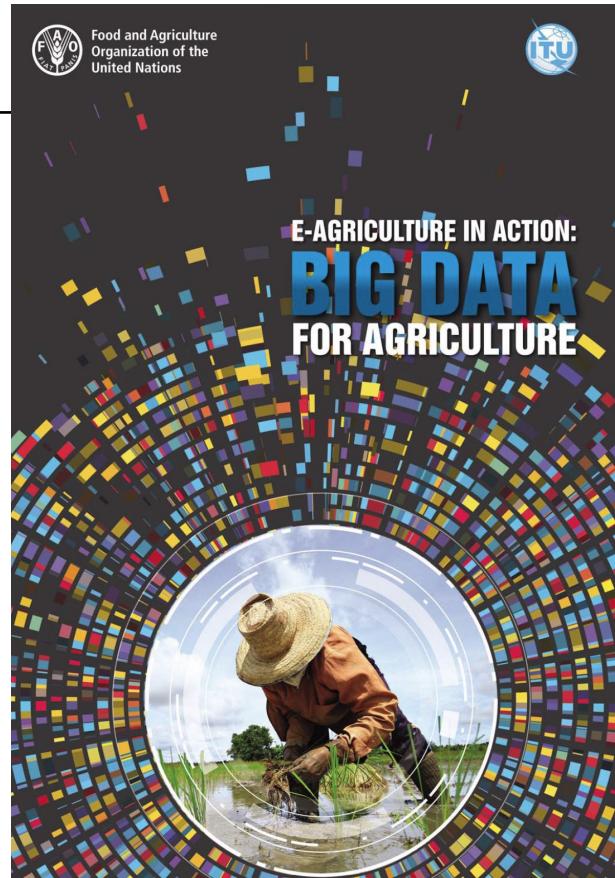
CHOLERA MAP OF DR. JOHN SNOW (UK 1850S)



GEODATA – APPLICATIONS



Source: FAO, ITU.



GEOSPATIAL DATA & PYTHON

- Jupyter Notebook
- *an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.*



The screenshot shows a Jupyter Notebook interface. On the left is a file browser with files like '1_EDA_DASK.ipynb' and '1_EDA PYTHON.ipynb'. The main area has two code cells and a data preview:

```
[1]: #matplotlib inline
import matplotlib.pyplot as plt
import pandas as pd
# Turn interactive plotting off
plt.ioff()
## pd.options.display.mpl_style='default'

[2]: <matplotlib.pyplot._Figure at 0x7f7c5b8160>

[2]: #read a csv file to a dataframe: here we need to assign column names, we use species name for the
df=pd.read_csv(r'/data/shared/Big-Geodata-Processing-2021-1a/distributed_computing_machine_learning'
names=['sepal.length','sepal.width','petal.length','petal.width','species'])
df..
```

	sepal length	sepal width	petal length	petal width
species				
Iris-setosa	5.1	3.5	1.4	0.2
Iris-setosa	4.9	3.0	1.4	0.2
Iris-setosa	4.7	3.2	1.3	0.2
Iris-setosa	4.6	3.1	1.5	0.2
Iris-setosa	5.0	3.6	1.4	0.2



Geospatial Computing Platform

Center of Expertise in Big Geodata Science (CRIB)



UNIVERSITY OF TWENTE.

GEOSPATIAL DATA & PYTHON

- Numpy
- *a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.*



GEOSPATIAL DATA & PYTHON

- Pandas
 - *a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series*
- Geopandas
 - *an open source project to make working with geospatial data in python easier. GeoPandas extends the datatypes used by pandas to allow spatial operations on geometric types*



GEOSPATIAL DATA & PYTHON

- Matplotlib
 - *a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.*
- Seaborn
 - *a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics*



GEOSPATIAL DATA & PYTHON

- Scikit-learn
- *a free software machine learning library for the Python programming language*

