Delivering Data Science
In Resources & Energy

Special Data Types: Spatial Data

Day 13

Jeremy Mitchell & Ying Yap

Data Mettle

jeremy@datamettle.com, ying@datamettle.com

A CORE partnership

Schedule

AWST	AEST	Agenda	Facilitator
07:30	**09:30**	**Q&A, Issues & Announcements**	
07:45	09:45	Introduction to spatial data	Ying
09:00	11:00	*Morning Tea*	
09:15	11:15	**Project Share (RTA)**	Georgia Hamilton
10:00	12:00	Using spatial data	Ying
11:00	13:00	*Lunch*	
11:45	13:45	**How to pitch a data science project**	Jeremy
13:15	15:15	*Afternoon Tea*	
13:30	15:30	**Project & pitch development**	All
14:45	16:45	Closeout	

Program Timeline

1 day Leading Data Scientists Executive Program	1/2-1 Day*	Enabling your people's data science capability in context of the 15 day program
Data Science Springboard Technical Preparatory	Day 1	Introduction to the program tools & set up
	Day 2	Introduction to the program tools & set up
Introduction to Data Projects	Day 3	Zero to Data Science in a day
Data Analysis	Day 4	Getting to know the program tools - data munging and exploratory data analysis
	Day 5	Simple predictions - regression and statistical model building
	Day 6	Multivariate analysis and model building
Data & Communication Sandbox	Day7	Effective data storytelling - communicating results to non-technical audiences
Data Fusion and Machine Learning	Day 8	Pros and cons of commonly used statistical and machine learning techniques I
	Day 9	Pros and cons of commonly used statistical and machine learning techniques II
Data Fusion Sandbox	Day 10	Sandbox - Consolidate approaches covered and test on datasets
Special Data Types - Time-series Data	Day 11	The 4th dimension and predictions
Special Data Types - Natural Language Processing and Text Mining	Day 12	Finding needles in wordstacks
Special Data Types - Spatial Data	Day 13	Spatial analytics and predictions
Capstone Project Development & Presentation	Day 14	Pitching Capstone Projects
Capstone Propeller	Day 15	Project Review Day

Q&A, Issues & Announcements

Before we Get Started

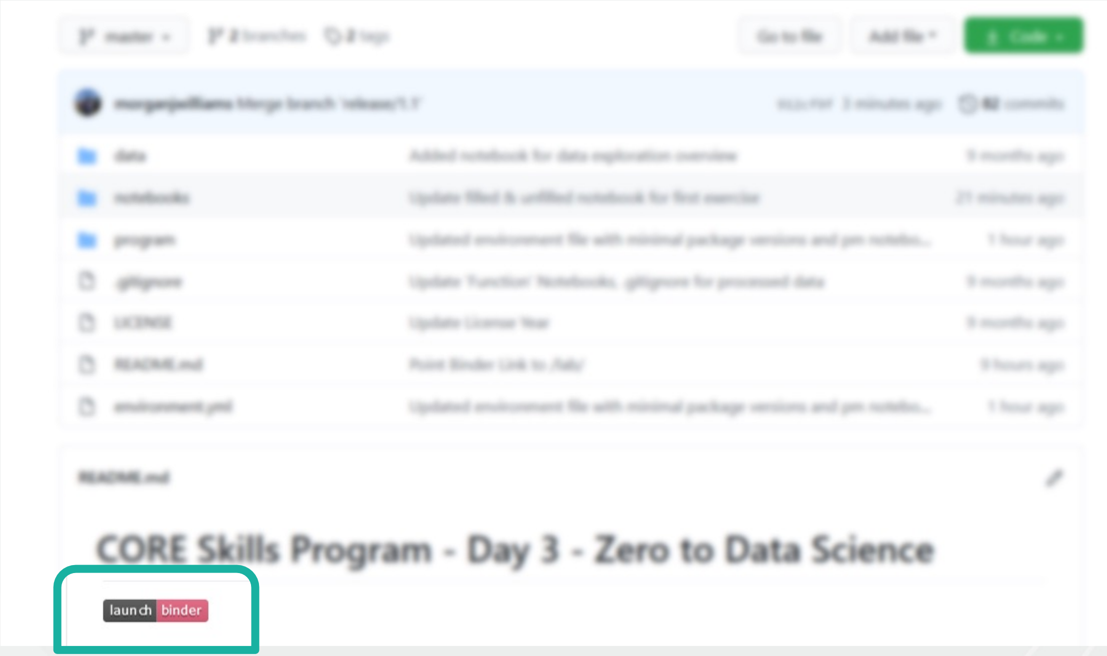
- **Resources & Tasks**

- Reflections, questions and Issues: **Teams**

- The afternoon is time to work on your **project pitch** and any questions you might have

GitHub Content for Today using Binder

[github.com / core-skills / 13-spatial-data](github.com/core-skills/13-spatial-data)



GitHub – Program Notes

[github.com / core-skills / 13-spatial-data / program / 00_overview.md](github.com/core-skills/13-spatial-data/program/00_overview.md)

Overview

Overview | Data Culture | From Here to There | Data Projects | Data Exploration | Closeout

Aim

Provide an overview of a 'typical' data science workflow.

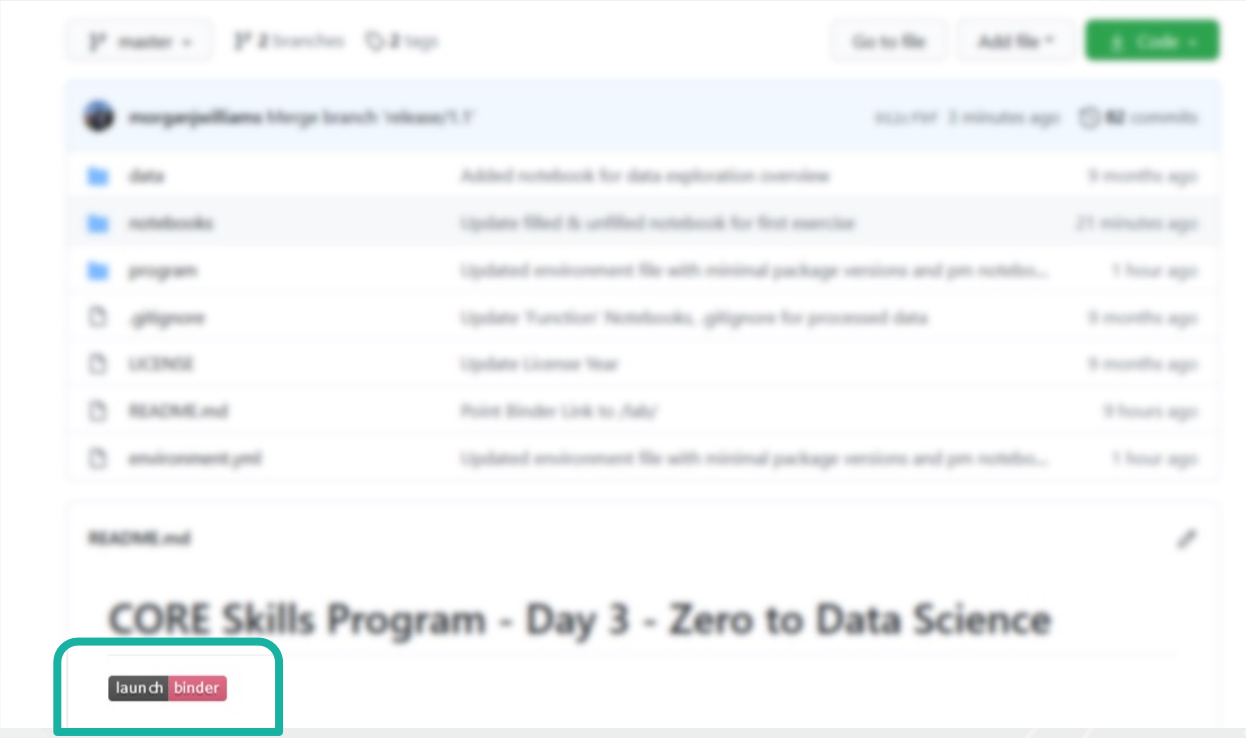
Learning Outcomes

1. To appreciate what data science is
2. To appreciate the fields data science spans
3. Understand the stages of a data science project and define a mental model of it
4. Analyse the opportunity and potential value of data science in your organisation

Schedule

AWST	AEST	Agenda
07:30 - 07:45	09:30 - 09:45	Q&A, Issues & Announcements
07:45 - 09:15	09:45 - 11:15	Creating a Data Culture
09:15 - 09:30	11:15 - 11:30	*Morning Tea*
09:30 - 11:00	11:30 - 13:00	Getting From Here to There

Binder Backup



Aims & Learning Outcomes

Day 13

Aims

- Extract geospatial features for data science applications.
- Present the practical strategies for using geospatial data.

Learning Outcomes

- Understand vector and raster geospatial formats with feature extraction.
- Understand projections and geocoding.
- Understand how to manipulate spatial data sets.

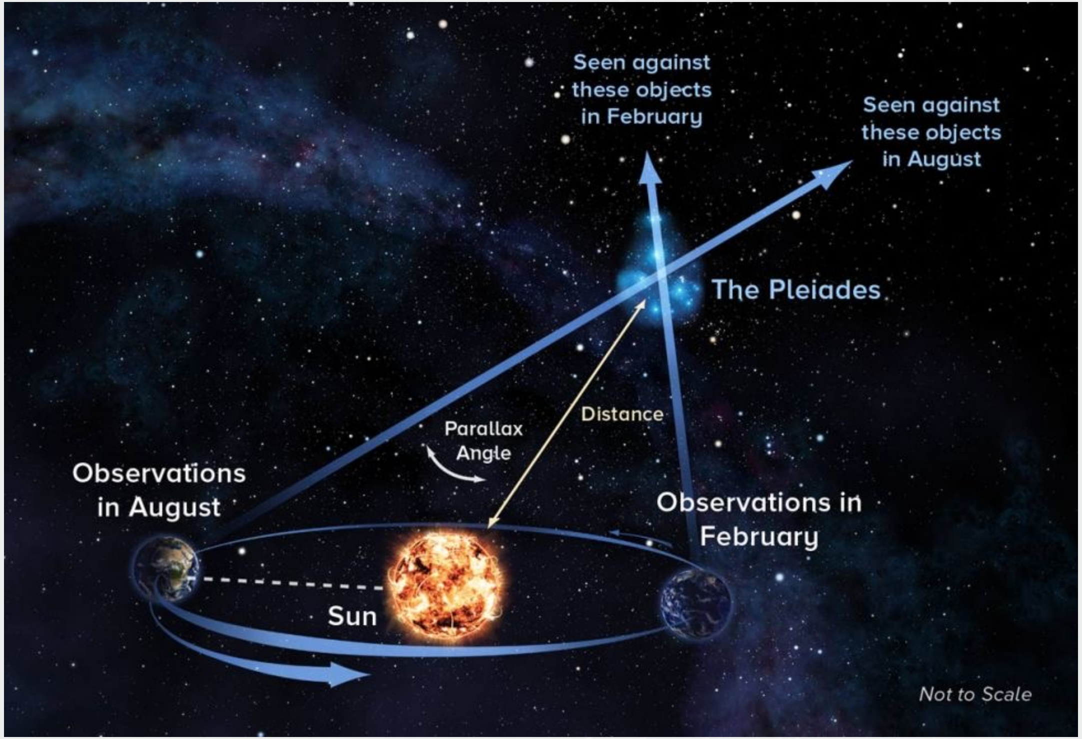
Introduction to spatial data

Fundamental concepts for spatial analysis

Key concepts

- Space
- Location
- Distance

- Each can be described as:
  - Absolute
  - Relative
  - Cognitive
  - Nominal



From: https://www.space.com/30417-parallax.html

Fundamental concepts for spatial analysis

Space

- Absolute space
  - mathematical space, things are given an (x, y, z) tuple and positions are unambiguous.

- Relative space
  - Topological space, can be thought of as relative space. Used to represent connectivity between features of the world

- Cognitive space
  - reflects people's beliefs, experiences and perceptions about places.

Fundamental concepts for spatial analysis

Location

- Absolute
  - an unambiguous descriptor of location, expressed as a coordinate. Can't be confused with any other location.
  - Example: Latitude and longitude.
- Relative
  - Site (physical attributes of location) and situation (location relative to other places).
  - Example: Perth is located far from other cities, near raw materials etc. Or 191 St Georges Terrace
- Cognitive
  - Compiled from personal knowledge, experiences, and impressions
  - Example: traditional knowledge of waterholes as a resource, then a stock route, then a 4WD track. Same physical location but different perception.
- Nominal
  - *Where were you when…?* Linking space and time.
  - Example: Where were you when you heard about 9/11?

Fundamental concepts for spatial analysis

Distance

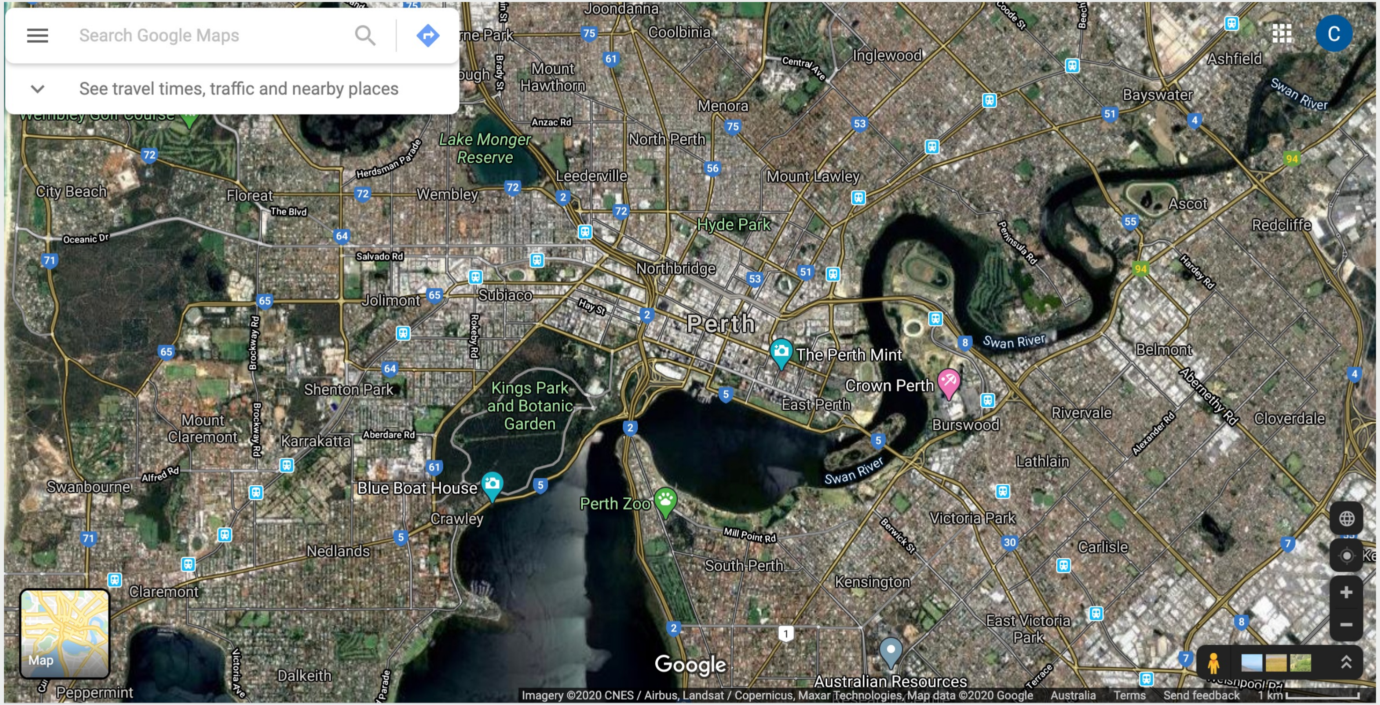
- Absolute
  - a physical unit of measure Example: kilometers between two assets

- Relative
  - Generally calculated using time, effort or cost (will differ from absolute – what happens when you go over a mountain range?). Example: voting maps sized by number of seats vs area.
- Cognitive
  - Individual perception of distance. Example: driving in NZ vs driving in Australia

Geospatial data types

Objects and attributes

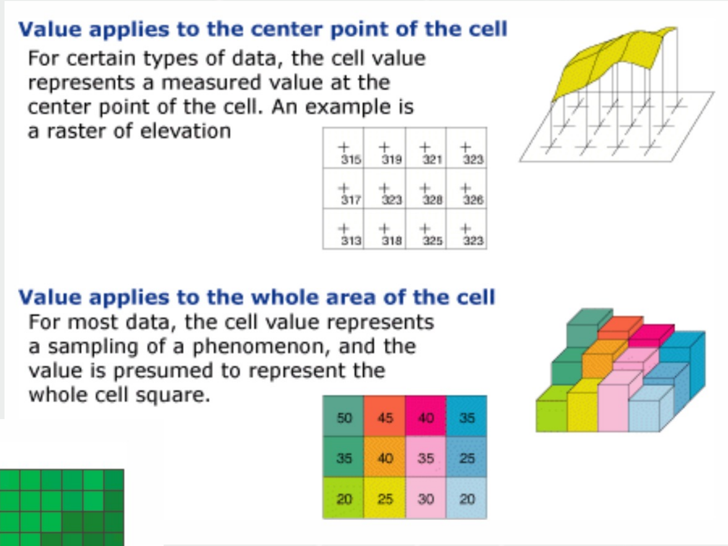
- Raster
- Vector
- Point clouds

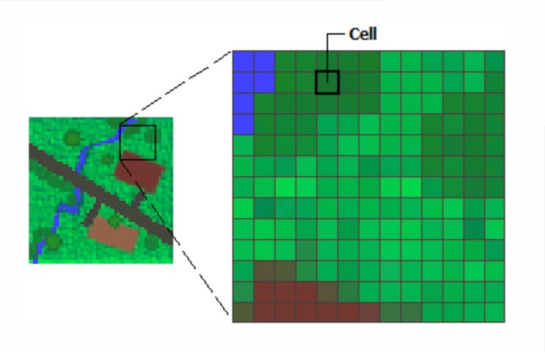


Geospatial data types

Raster

- Regularly defined geographic space, often used for representing continuous phenomena. Example: orthophotos, remote sensing
- Example Attribute: elevation, temperature



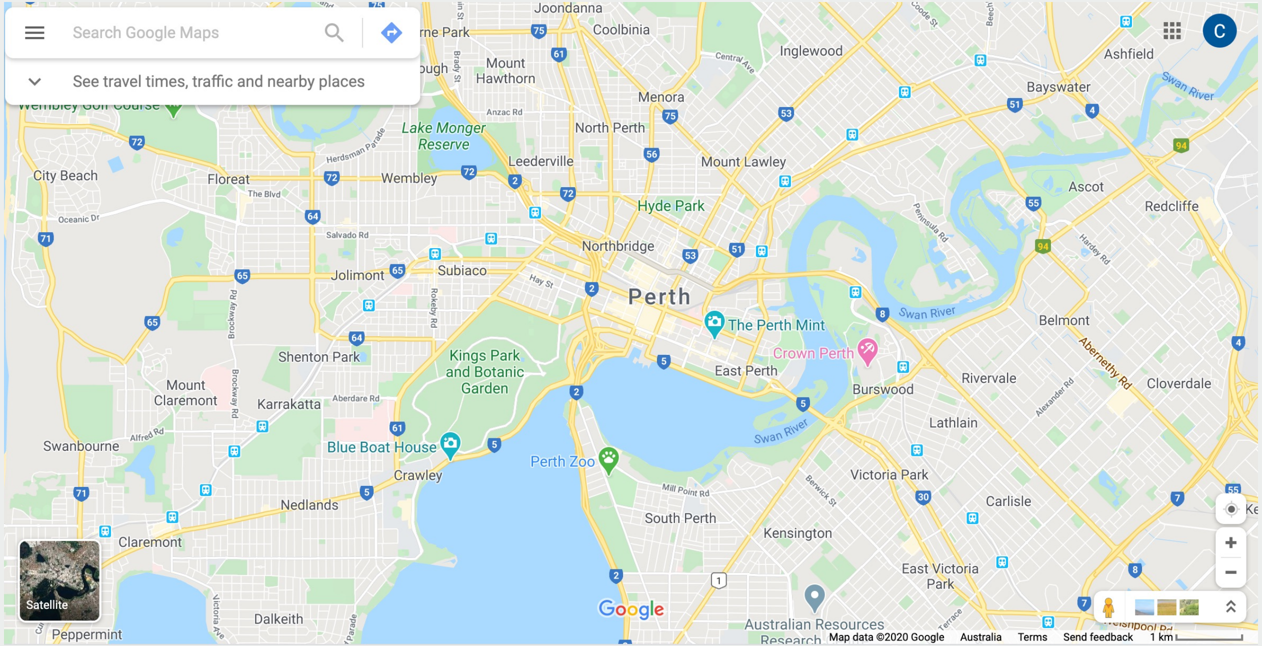


From: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm

Geospatial data types

Vector

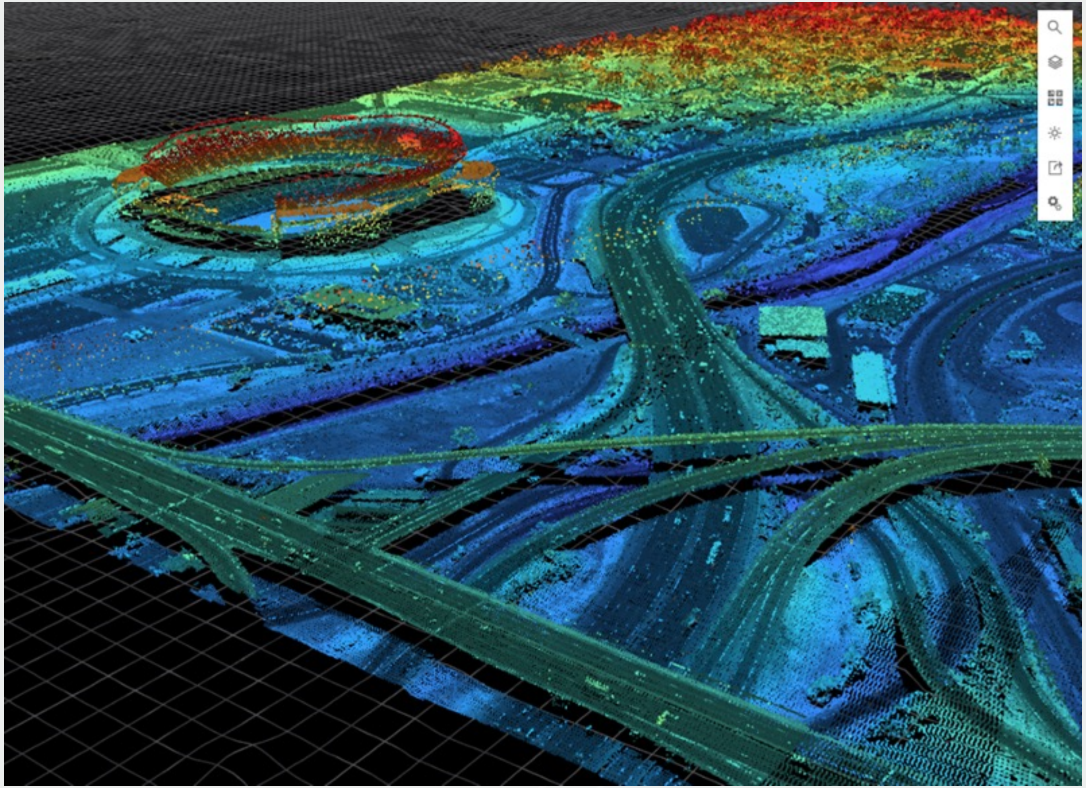
- Defined as point, line, or polygon. Often used for representing discrete phenomena. Example: River locations
- Example attribute: Name of river, pH measurement, date of pH measurement



Geospatial data types

Point clouds

- Vector space representation of a 3D surface or object. Example: lidar data
- Example attribute: color

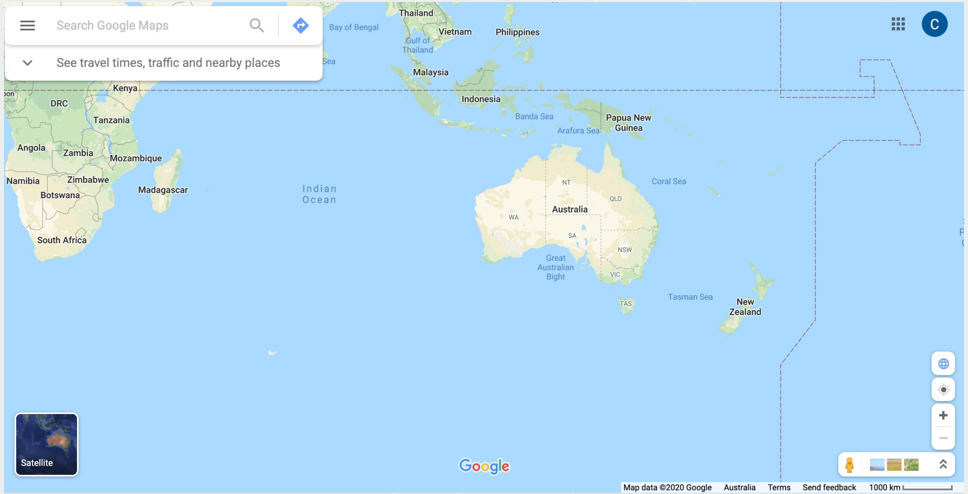


From https://www.esri.com/arcgis-blog/wp-content/uploads/2017/12/blogoverview_smal_cr.png

Spatial reference frame

A static reference surface for identifying locations on the earth.

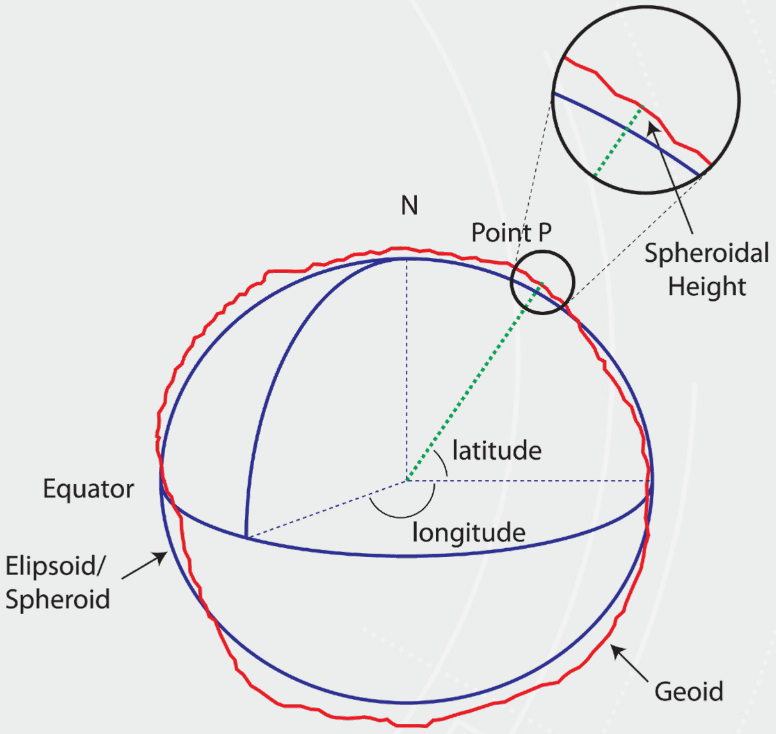
- Datum
- Projection



Spatial reference frame

Datum

- A datum identifies how locations are specified on earth's surface.
- A spheroid (or ellipsoid) is an approximation for the earth's surface, and acts as the reference for the datum.

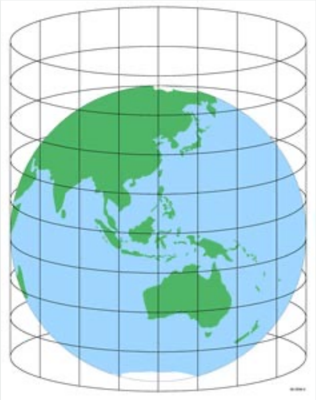


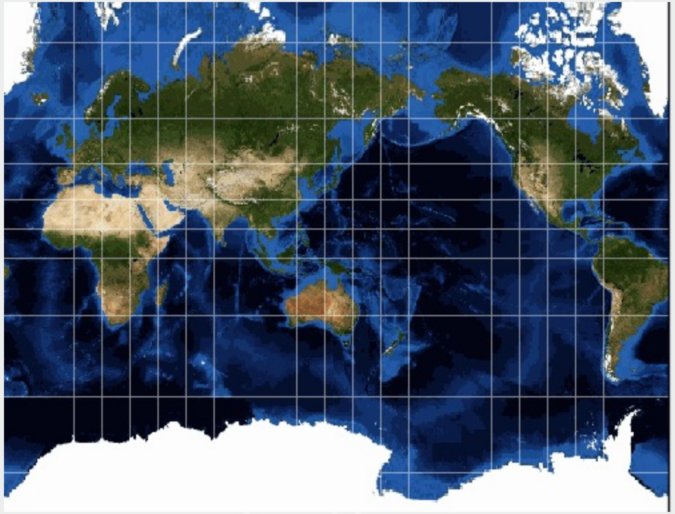
From: https://www.icsm.gov.au/education/fundamentals-mapping/datums/datums-explained-more-detail

Spatial reference frame

Projection

- A projection specifies how to represent the features from a spheroidal surface on a flat plane. Types of projection techniques include:
  - azimuthal
  - cylindrical
  - conical





From: https://www.icsm.gov.au/education/fundamentals-mapping/datums/datums-explained-more-detail

Excellent reference: https://www.icsm.gov.au/education/fundamentals-mapping/projections

Spatial reference frame

The actual location values (latitude, longitude) will vary depending on the reference frame.

It is important to make sure all spatial data have the appropriate spatial reference information in the metadata.

Open am1_intro_to_spatial_data.ipynb

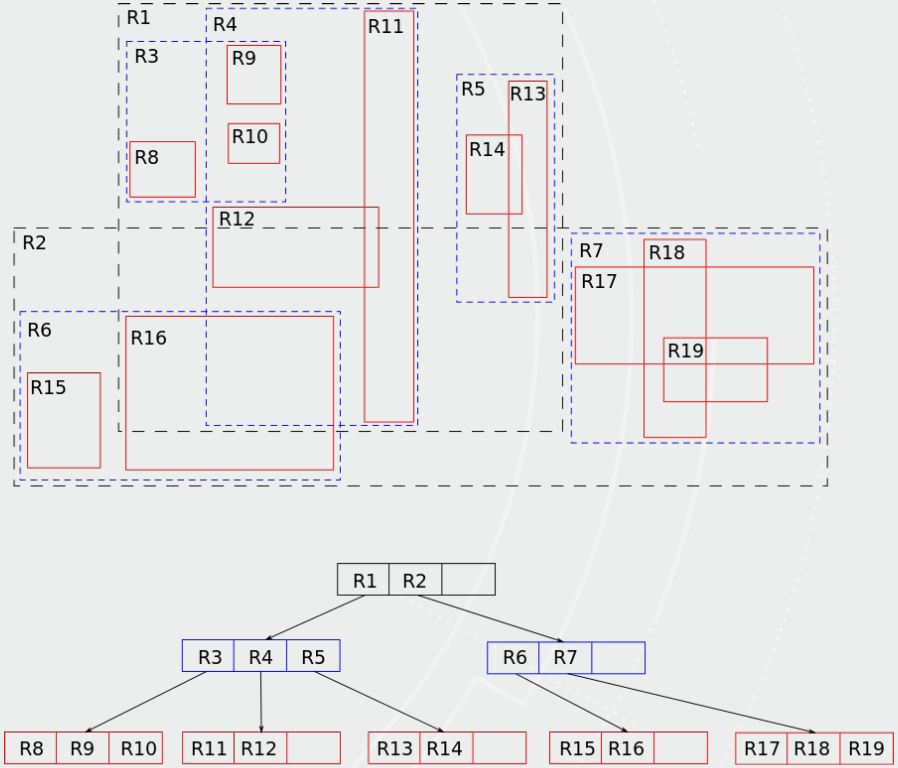
Using spatial data

Standard spatial operations

Source: Wikipedia

Spatial index and selections

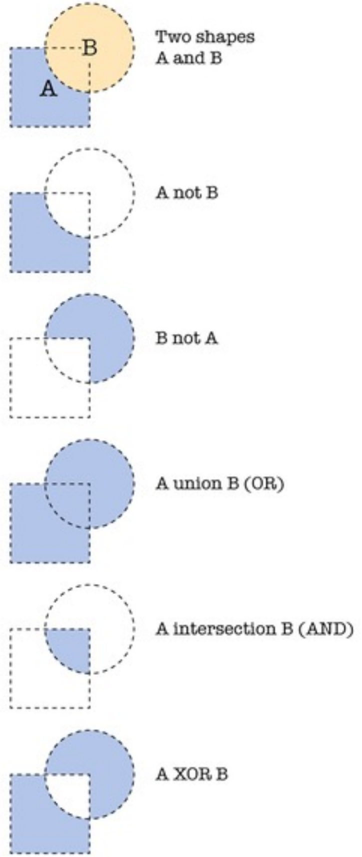
- **R-tree:** hierarchical data structure in which nearby objects are grouped and represented by their minimal bounding rectangle
- Used in spatial data bases for fast querying, e.g., nearest neighbours



Standard spatial operations

Geometry operations

- Splitting using overlay operations (intersect, union, join)
- Joining/dissolving
- Buffering/ eroding

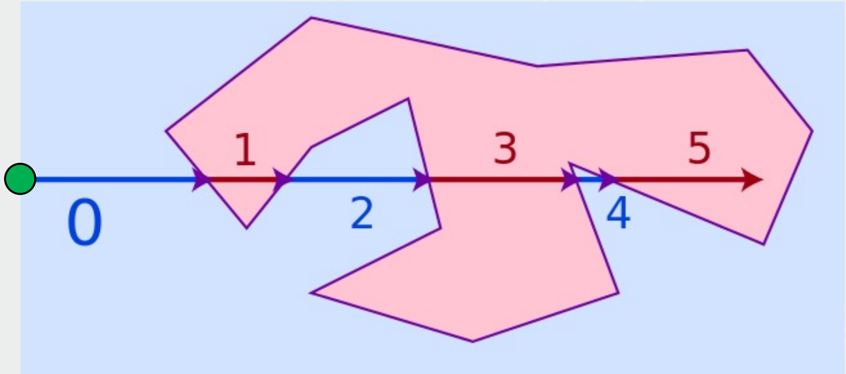


Source: Wikipedia

Standard spatial operations

Point in polygon

- = Identifying whether a point lies inside or outside of a polygon
- Several algorithms exist, the simplest being the Ray casting algorithm
- Check how many times a ray starting from the point intersects a polygon's edge:
  - Even = outside
  - Odd = inside



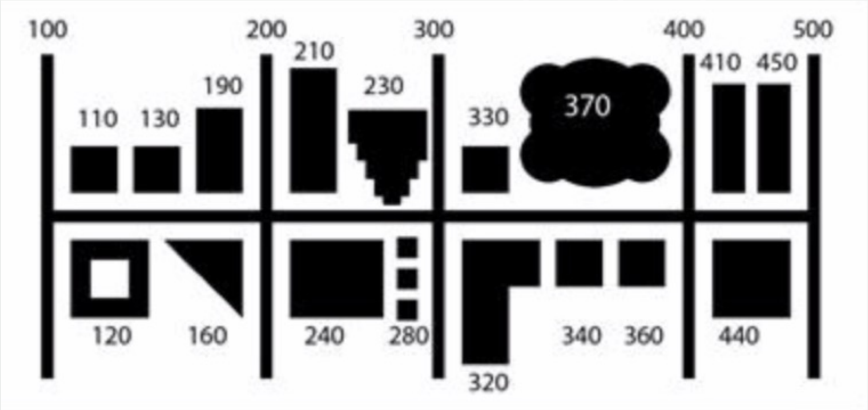
Source: Wikipedia

Standard spatial operations

Geocoding

= Turn a physical address (e.g., a street name) into geographical coordinates (e.g., latitude and longitude):

- Linear geocoding (find a point along a line, e.g. 191 St Georges) – estimate from start and end points of the block. Not as accurate but easy to implement and data is commonly available
- Area geocoding (find a point in a polygon) – much more accurate but requires lots more data

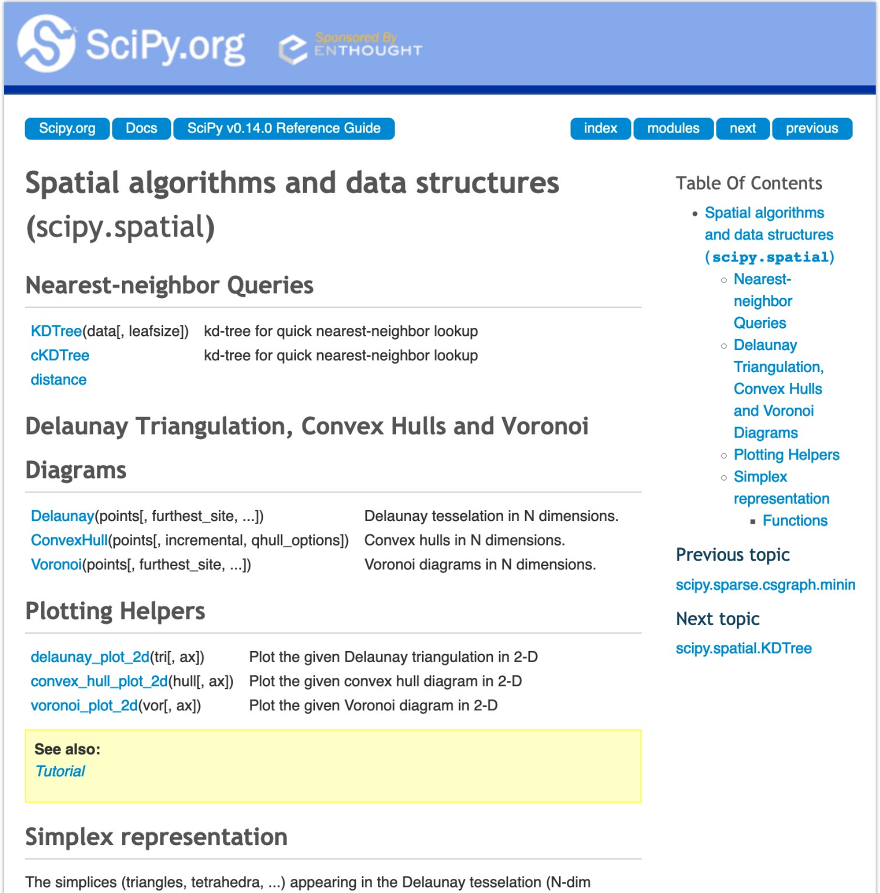


Standard spatial operations

Actual use

Generally, we don't just want to do this ourselves – use an API!

See for instance scipy.spatial, shapely, or geopy in Python



SciPy.org Sponsored By ENTHOUGHT

Scipy.org | Docs | SciPy v0.14.0 Reference Guide | index | modules | next | previous

Spatial algorithms and data structures (scipy.spatial)

Nearest-neighbor Queries

KDTree(data[, leafsize])	kd-tree for quick nearest-neighbor lookup
cKDTree	kd-tree for quick nearest-neighbor lookup
distance	

Delaunay Triangulation, Convex Hulls and Voronoi Diagrams

Delaunay(points[, furthest_site, ...])	Delaunay tesselation in N dimensions.
ConvexHull(points[, incremental, qhull_options])	Convex hulls in N dimensions.
Voronoi(points[, furthest_site, ...])	Voronoi diagrams in N dimensions.

Plotting Helpers

delaunay_plot_2d(tri[, ax])	Plot the given Delaunay triangulation in 2-D
convex_hull_plot_2d(hull[, ax])	Plot the given convex hull diagram in 2-D
voronoi_plot_2d(vor[, ax])	Plot the given Voronoi diagram in 2-D

See also:
Tutorial

Simplex representation

The simplices (triangles, tetrahedra, ...) appearing in the Delaunay tesselation (N-dim

Table Of Contents

Previous topic

scipy.sparse.csgraph.minin

Next topic

scipy.spatial.KDTree

Standard spatial operations

Exercise

Open **am2_geospatial_data_processing.ipynb** and go through the notebook

Spatial predictions

Why do we worry about space at all?

- **Tobler's first law of geography:**

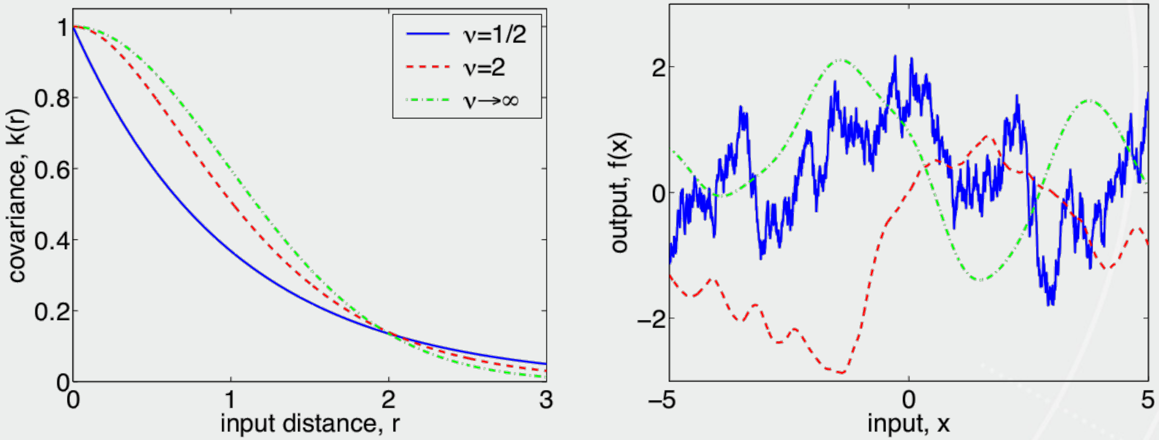
  “Everything is related to everything else,
  but near things are more related than distant things.”

- To predict in space, each location is treated as a variable (same principle in time), and those variables are dependent.

Spatial predictions

Covariance function

- Mathematical formulation of the first law of geography
- Defines how the correlation between two locations changes with distance



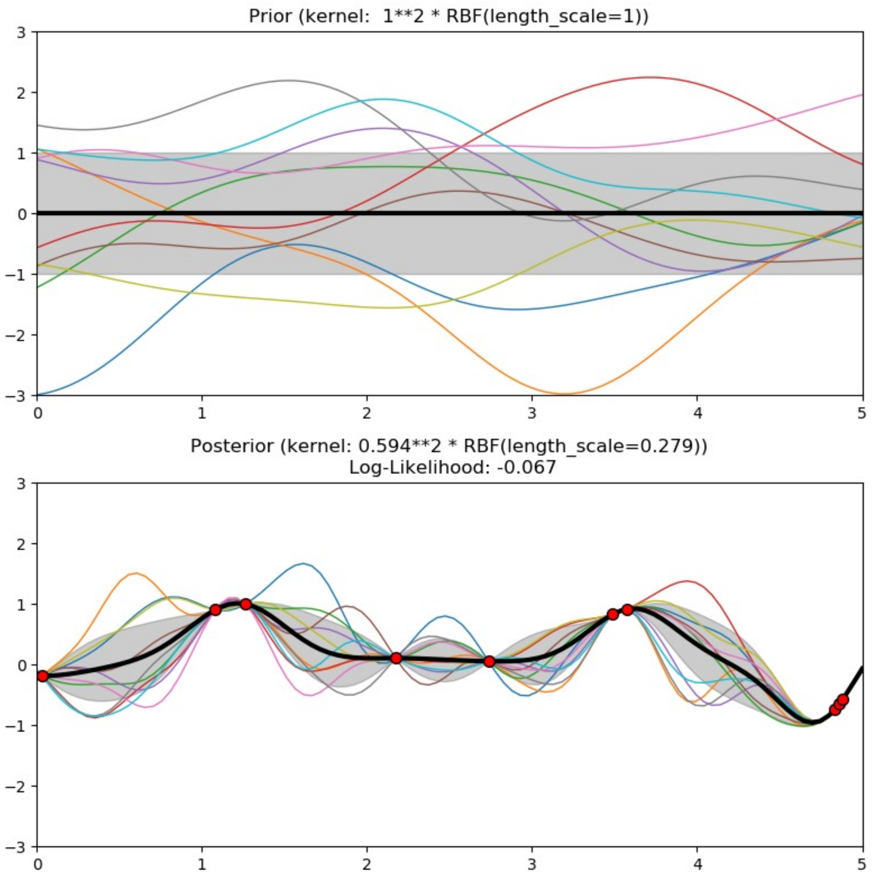
Source: Rasmussen and Williams (2006)

- Variogram in geostatistics, kernel in machine learning

Spatial predictions

Gaussian processes

- Supervised method for regression and classification that uses a covariance function to reproduce spatial continuity
- Can fit data and provide prediction uncertainty
- But computationally demanding
- Equivalent to kriging in geostatistics

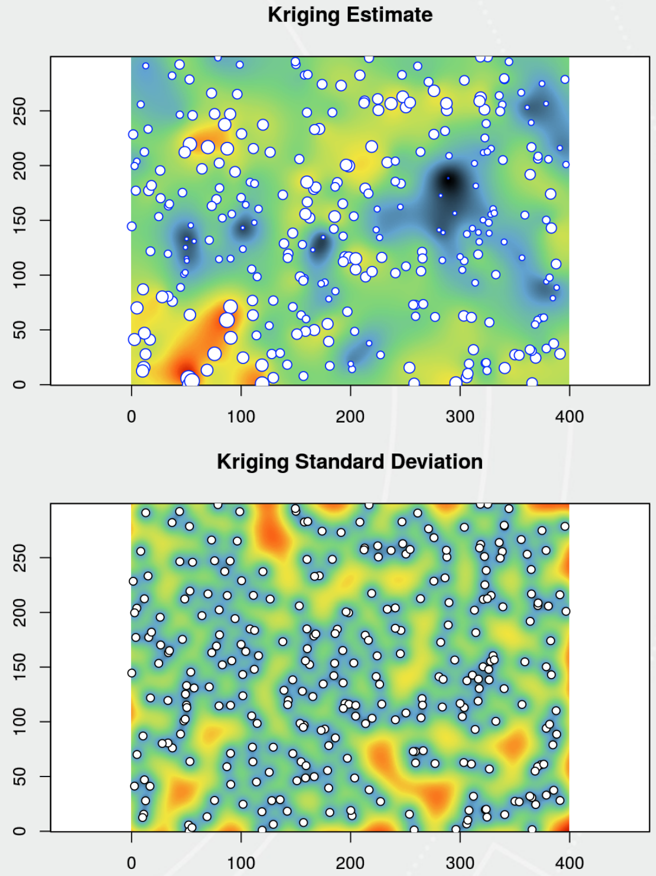


Source: Scikit-Learn documentation

Spatial predictions

Gaussian processes

- Supervised method for regression and classification that uses a covariance function to reproduce spatial continuity
- Can fit data and provide prediction uncertainty
- But computationally demanding
- Equivalent to kriging in geostatistics



Source: Rgeostats documentation

Takeaways

High-level Takeaways From Today

Today we've:

- covered the basic concepts of spatial data: space, location and distance, and the different methods we use to talk about these concepts, including absolute, relative, topological and cognitive measures.
- covered the basics of spatial data analysis, including geometry operations, spatial indices, and geocoding.
- covered how spatial processing can form part of a larger machine learning or data science pipeline, particularly as a feature generator for machine learning algorithms.
- covered how to design an effective pitch that considers the audience, communication styles and content structure.

CORE Skills

A CORE partnership

COREHUB.COM.AU/SKILLS