Cheat Sheet: Geospatial Analytics Extension for KNIME Analytics Platform

Open for Innovation KNIME

Center for Geographic Analysis Harvard University

SPATIAL DATA I/O

GeoFile Reader: Reads single-layer spatial data from a local file path or a URL and supports various file formats such as Shapefile (.shp), zipped Shapefile (.zip), GeoJSON (.geojson), GeoPackage (.gpkg), and more.

GeoPackage Reader: Reads multi-layer spatial data from a local file path or a URL and primarily supports GeoPackage (.gpkg) and GeoDatabase (.GDB) formats but can also read zip files compressed from a GeoDatabase folder. The layer to read can be specified in the node configuration.

GeoFile Writer: Writes spatial data and its ancillary data to a local file, supporting Shapefile (.shp), GeoJson (.geojson), and GeoParquet (.parquet) formats.

GeoPackage Writer: Writes spatial data to a new file or as a layer into an existing file and supports GeoPackge (.gpkg) and GeoDatabase (.GDB) formats.

OPEN DATASETS

OSM Boundary Map: Retrieves place boundaries from *OpenStreetMap* by geocoding the place name. The query must match the locations in the *Nominatim database*.

OSM Road Network: Retrieves a geospatial network and its associated attributes from *OpenStreetMap*, including networks for driving, cycling, walking, etc. Similarly, the **OSM POIs node** can be used to retrieve geometries and attributes for geospatial entities.

US ACS 5-Year Estimates: Retrieves American Community Survey 5-Year Data (2009-2020). The ACS covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population

US2020 Census Data: Retrieves US 2020 Census Redistricting Data (Decennial Census P.L. 94-171 Redistricting Data) which provides variables such as population and household information.

US2020 TIGER Map: Retrieves the specific geospatial boundaries for one specific state of the United States. The popular TIGER/ Line levels are Block group, Roads, Blocks, Tracts.

SPATIAL MANIPULATION

Buffer: Generates a buffer with the given distance for each geometric object.

Clip: Clip target geometries to the mask extent. The geometries will be clipped to the full extent of the clip object.

Overlay: Perform spatial overlay between two geometries based on various modes such as Union, Intersection, etc.

Spatial Join: Merge the left (top) and the right (bottom) table based on their spatial relationship of the two selected columns to one another.

Nearest Join: Merge the left (top) and the right (bottom) table based on the distance between their geometries of the two selected columns.

Dissolve: Aggregate geometries based on group id and keeps only the two columns group id and dissolved geometries.

Euclidean Distance: Calculates the Euclidean distance in the selected unit between each origin and destination path. To take into account the curvature of the earth, use the **Haversine Distance** node.

Voronoi (Thiessen) Polygons: Creates Voronoi (Thiessen) polygons from the input point data according to the reference boundary.

SPATIAL CALCULATION

Area, Bounds, Length: Calculate the geometric attributes - **Area, Length** and **Bounds** - for each spatial object based on the object's coordinate system.

Bounding Box, Unary Union, Convex Hull: Generate new geospatial objects through calculations. **Bounding Box** and **Convex Hull** create minimum bounding geometries for each row to encapsulate the object using a box or polygon. **Unary Union** merge all objects into a single geometry row.

SPATIAL TRANSFORMATION

Polygon To Line returns the boundaries of each polygon; **Line To MultiPoint** returns points from the lines; **Geometry To Point** returns centroids or representative points guaranteed to be within each geometry.

Points To Line: Generates lines from points according to group id and serial label.

Multipart To Single Part: Explodes multi-part geometries into multiple single geometries.

Create Random Points: Generates random points from a uniform distribution within/along each input geometry. It can be used for polygons and lines, other geometry types are ignored.

Projection: Transforms the Coordinate Reference System (CRS) of a geometry column into a new CRS. It can be used to unify the rows with various CRS.

SPATIAL CONVERSION

Geocoding retrieves spatial location points based on the given addresses or vice versa with the **Reverse Geocoding** node.

Convert the **Well-known Text (WKT)** column into a geometry column, or vice versa with the **Geometry to WKT** node.

Convert the **GeoJSON** column into a geometry column, or vice versa with the **Geometry to GeoJSON** node.

Convert the given **latitude and longitude** column into a point geometry column, or vice versa with the **Geometry to Lat/Lon** node.

IP To Geometry: Retrieves the spatial location of the provided IP addresses by choosing the desired service provider and API key (optional).

Geometry To Metadata: Extracts metadata for each geometry within the selected column. It includes the CRS, the geometry type, and a flag indicating the presence of z-coordinates.

SPATIAL VISUALIZATION

Geospatial View: Creates an interactive map view based on the geometric elements of the input table. It supports several modifications such as changing the base map or shape color and size.

Geospatial View Static: Creates a static visualization of the geometric elements and creates an image at its output port (supports .svg or .png files). It is useful for creating Choropleth Maps and supports various settings to customize the legend.

Kepler.gl Geoview: Creates an interactive map view using the *kepler.gl* framework. It allows changing various aspects of the view within the visualization, e.g., adding *layers* and *filters*.

Spatial Heatmap: Visualizes spatial data on an interactive heatmap using a weight column to represent the intensity at each element.

EXPLORATORY SPATIAL DATA ANALYSIS

Spatial Weights: Constructs a contiguity spatial weights matrix, considering various spatial relations such as Queen, Rook, Binary, Distance, etc., and allows the upload of a custom matrix.

Global Spatial Statistics (Spatial weight for second input port)

Global Geary's G: Measures overall spatial autocorrelation, examining whether values tend to cluster or disperse globally. Valuable for understanding how similar/dissimilar values are.

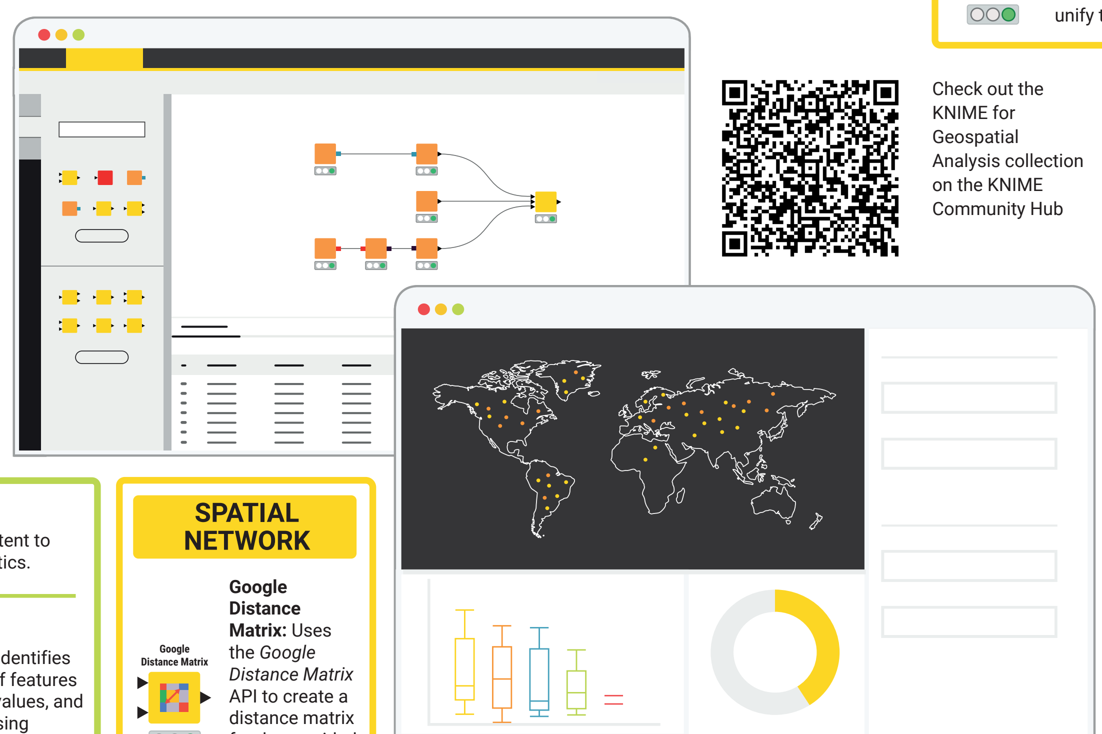
Global Getis-Ord G: Measures spatial clustering patterns on a global scale, helping to detect areas of concentration or dispersion across the study area.

Global Moran's I: Measures overall spatial autocorrelation within a dataset, providing a single value that summarizes the extent to which nearby data points exhibit similar or dissimilar characteristics.

Local Spatial Statistics (Spatial weight for second input port)

Local Getis-Ord G: Identifies statistically significant hot spots and cold spots using the Getis-Ord Gi* statistic based on a set of weighted features.

Local Moran's I: Identifies spatial clusters of features with high or low values, and spatial outliers using Moran's I statistics.



Check out the KNIME for Geospatial Analysis collection on the KNIME Community Hub

SPATIAL NETWORK

Google Distance Matrix: Uses the *Google Distance Matrix* API to create a distance matrix for the provided origins and destinations. It returns the travel distance (in meter) and duration (in minutes).

OSRM Distance Matrix: Uses the *Open Source Routing Machine (OSRM)* to create a distance matrix for the provided origins and destinations. It returns the driving travel distance and time as well as the route.

Road Network Distance Matrix: Creates a distance matrix for the provided origins and destinations using the given road network. It snaps each pair to the closest point of the network and returns the shortest path.

Road Network Isochrone Map: Calculates the *isochrone map* for the input point based on the given road network and its travel cost column. It snaps the input points to the road network and returns weighted time or distance.

SPATIAL CLUSTERING

SCHC: A special form of constrained clustering, where the constraint is based on contiguity (common borders) and agglomerative hierarchical clustering methods.

SKATER: A clustering method based on the optimal pruning of a minimum spanning tree that reflects the contiguity structure among the observations. It provides an optimized algorithm to prune the tree into several clusters so that the values of selected variables are as similar as possible.

REDCAP: A clustering method that starts from building a spanning three in four different ways (like SKATER).

AZP: A greedy clustering algorithm for automatic zoning procedure. It tries to find the best set of combinations of contiguous spatial units into p regions, minimizing the within sum of squares as a criterion of homogeneity.

MaxP: A greedy clustering algorithm to solve the max-p-region problem. It consists of a search process that starts with an initial feasible solution and iteratively improves upon it while maintaining contiguity among the elements of each cluster.

Mean Center: Computes the mean center of a set of features and assess the compactness of their spatial distribution using Standard Distance (or standard deviation of distances). This measure is typically represented as a circle with a radius of the circle equal to the standard distance.

Standard Deviational Ellipse: Calculate parameters of standard deviational ellipse for a point pattern. These measures define the axes of an ellipse (or ellipsoid) encompassing the distribution of features.

SPATIAL MODELLING

OLS with Spatial Test: Performs Ordinary Least Squares (OLS) regression analysis and offers spatial statistical tests to assess model validity and spatial relationships.

2SLS with Spatial Test: Performs Spatial Two-Stage Least Squares (S2SLS) analysis, providing results and diagnostics, while also performing spatial statistical tests.

Spatial Lag Model: Performs Spatial Lag Model analysis, which examines how variables are influenced by spatially neighboring observations, providing insights into spatial dependencies.

Spatial Error Model: Performs Spatial Error Model analysis, accounting for spatial autocorrelation and addressing potential biases in regression analysis.

(Spatial weight for second input port)

GWR Model: Performs Geographically Weighted Regression (GWR), allowing for the exploration of spatially varying relationships in the data, considering spatial heterogeneity.

MGWR Model: Performs Multiscale Geographically Weighted Regression (MGWR) analysis, a technique that examines spatial relationships at various scales, providing insights into spatial variations.

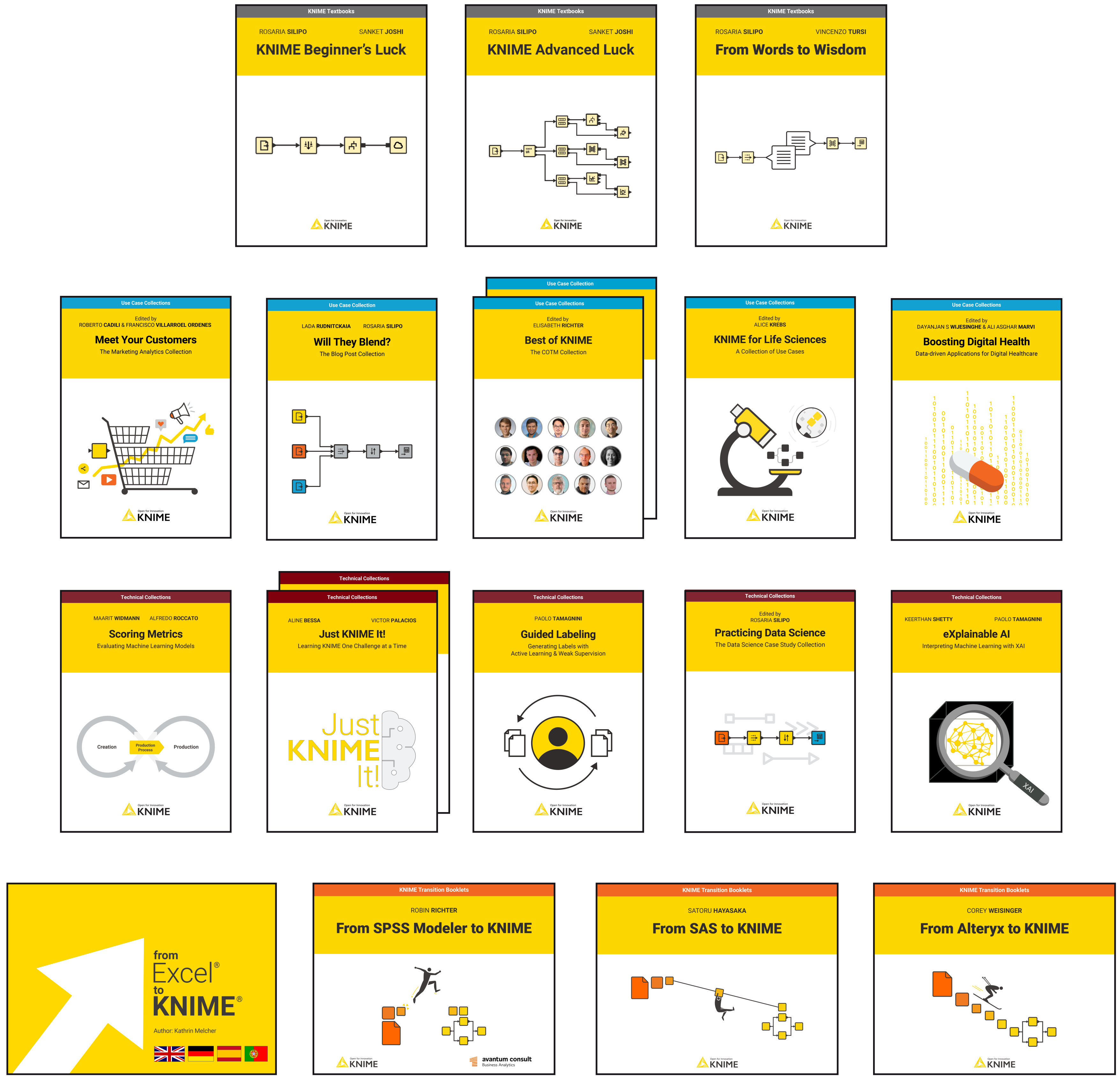
Reference

The 1.2 release of the Geospatial Extension which now supports the experiments detailed in the first KNIME workbook for GIS: *"Computational Methods and GIS Applications in Social Science - Lab Manual"* by Lingbo Liu and Fahui Wang. For a deeper understanding and broader applications, users are encouraged to refer to the main text *"Computational Methods and GIS Applications in Social Science, 3rd Edition"* authored by Fahui Wang and Lingbo Liu. The workflows for the Lab Manual are available in the *Geospatial Space* in the KNIME Hub.

Resources

- **E-Books**: KNIME Advanced Luck covers advanced features & more. Practicing Data Science is a collection of data science case studies from past projects. Both available at ***knime.com/knimepress***
- **KNIME Blog**: Engaging topics, challenges, industry news, & knowledge nuggets at ***knime.com/blog***
- **E-Learning Courses**: Take our free online self-paced courses to learn about the different steps in a data science project (with exercises & solutions to test your knowledge) at ***knime.com/knime-self-paced-courses***
- **KNIME Community Hub**: Browse and share workflows, nodes, and components. Add ratings, or comments to other workflows at ***hub.knime.com***
- **KNIME Forum**: Join our global community & engage in conversations at ***forum.knime.com***
- **KNIME Business Hub** : For team-based collaboration, automation, management, & deployment check out KNIME Business Hub at ***knime.com/knime-business-hub***

Extend your KNIME knowledge with our collection of books from KNIME Press. For beginner and advanced users, through to those interested in specialty topics such as topic detection, data blending, and classic solutions to common use cases using KNIME Analytics Platform - there's something for everyone. Available for download at www.knime.com/knimepress.



Need help?
Contact us!