# Spatio-temporal Data Analytics with KNIME

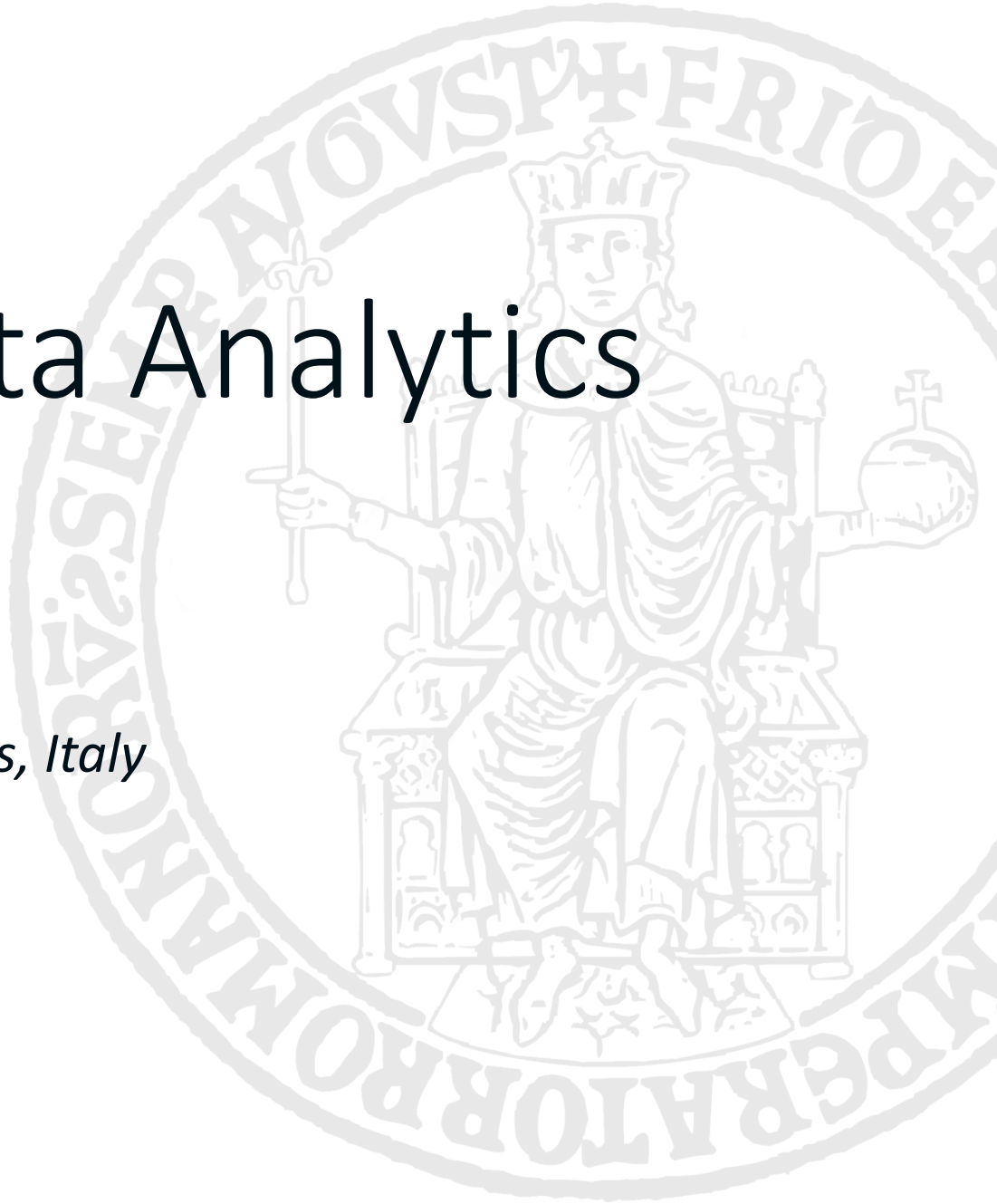
Luigi Libero Lucio Starace, Ph.D.

*Università degli Studi di Napoli Federico II, Naples, Italy*

luigiliberolucio.starace@unina.it

https://luistar.github.io

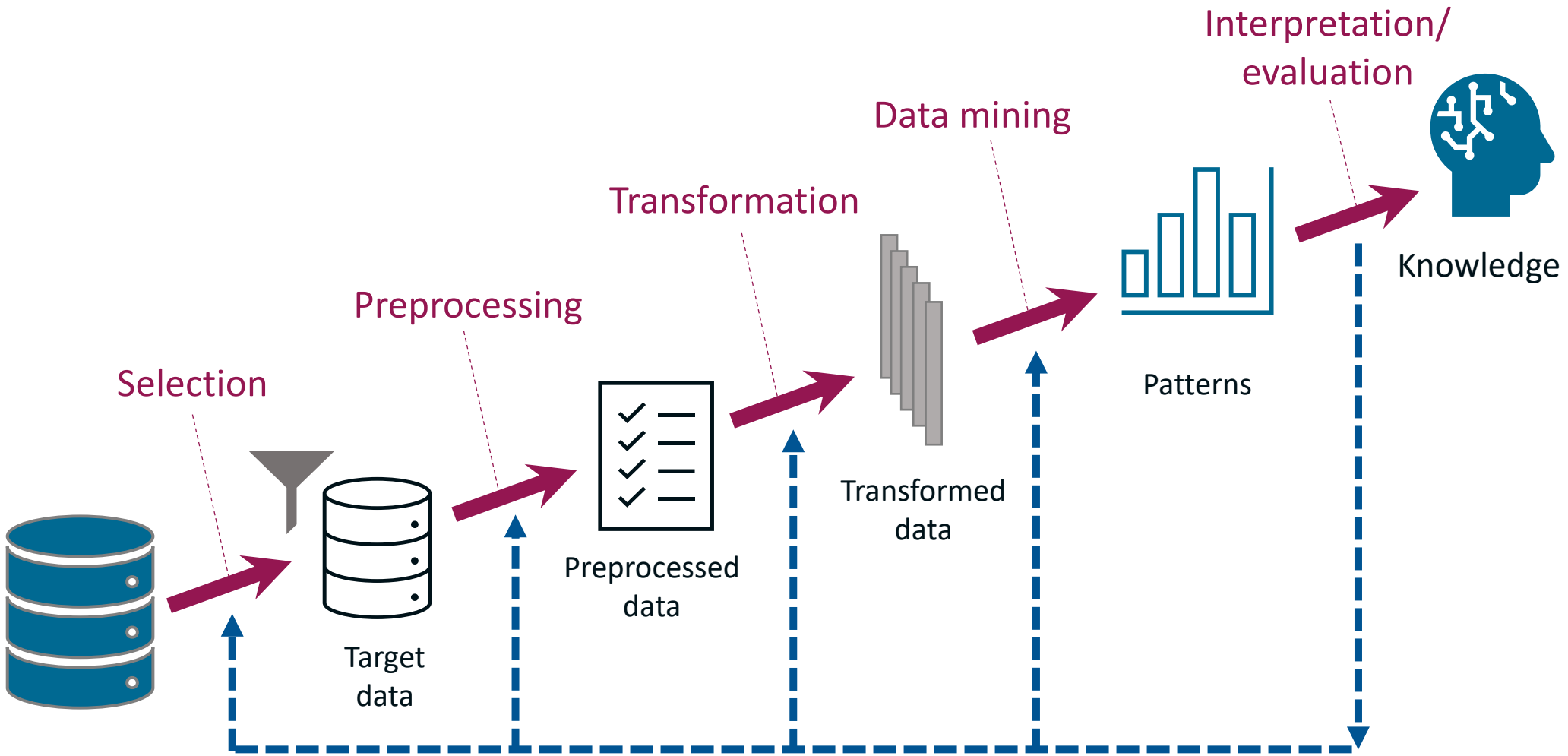
April 18, 2023 – University College Dublin



# Overview

- Knowledge Discovery from Data (KDD)
- Implementing KDD pipelines
  - Visual platforms (KNIME)
- KDD pipelines on Spatio-temporal (Mobility) Data
  - Peculiarities and Challenges
- A proposed solution: the KNIME Mobility Toolkit
- Examples from real-world Data Analytics scenarios

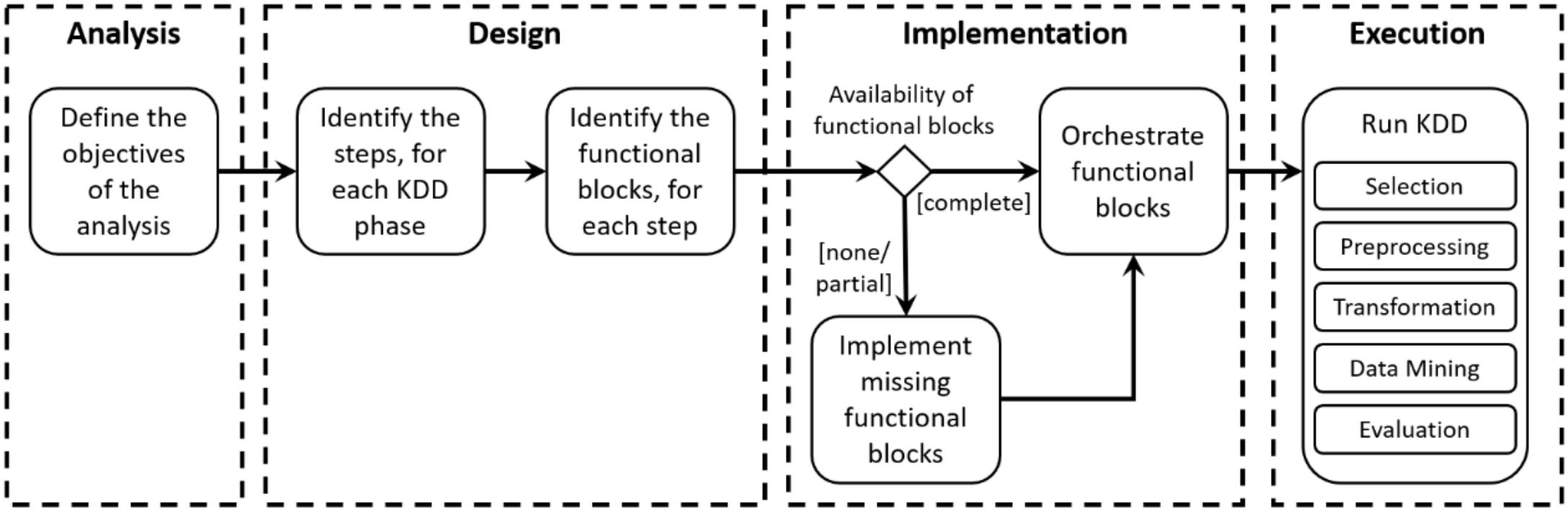
# Knowledge Discovery from Data



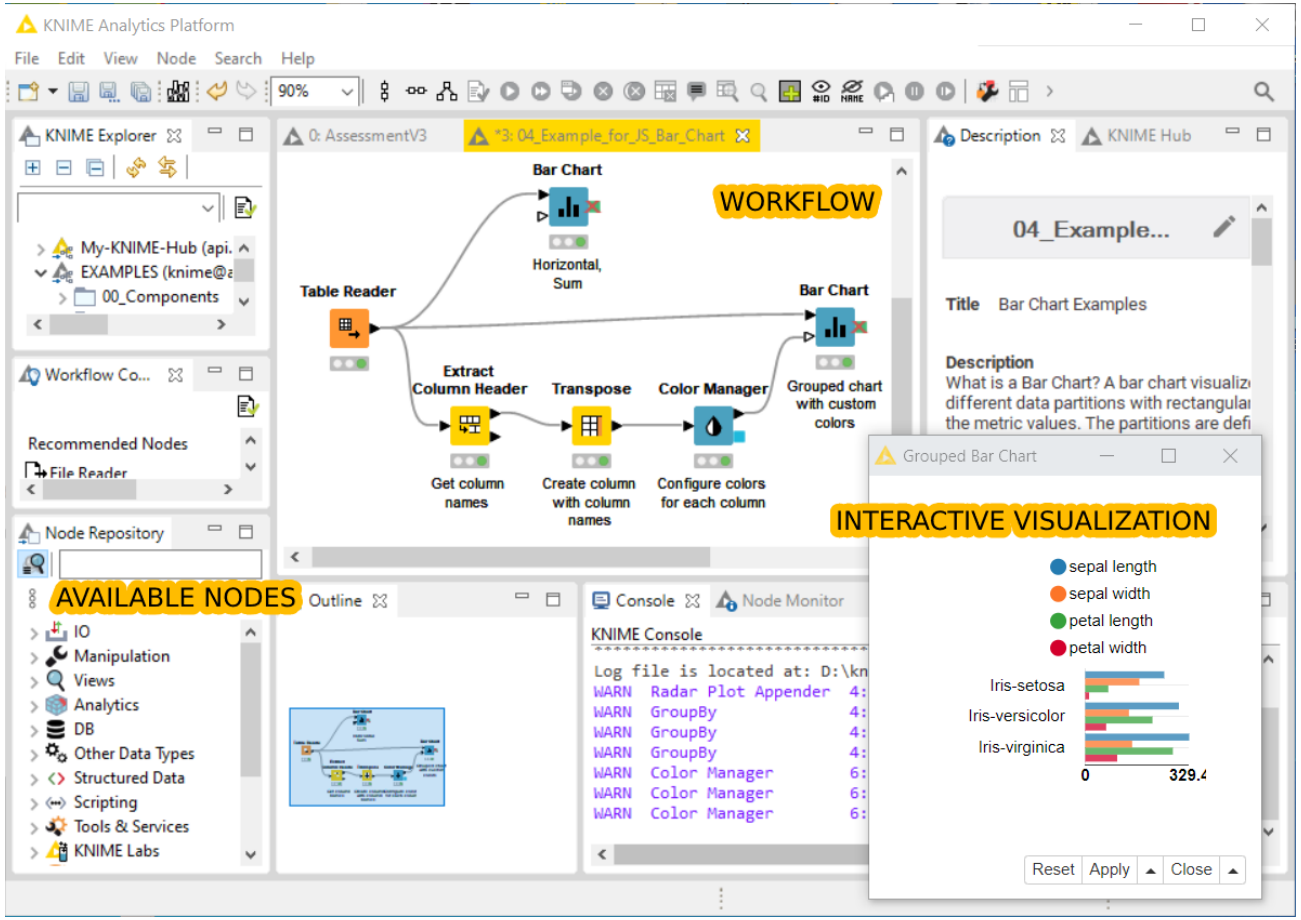
# KDD pipelines

- Can be designed and implemented as **modular software systems**
- Each step is carried out by a specialized module (**functional block**)
- Modular design can help improve
  - **Understandability**: blocks are a first abstraction of the pipeline -> easier to focus on **what** we should do rather than on **how** should we implement it
  - **Reusability**: specialized functional blocks can be reused in other pipelines

# Implementing KDD pipelines



# Visual Data Analytics Platforms (Demo)

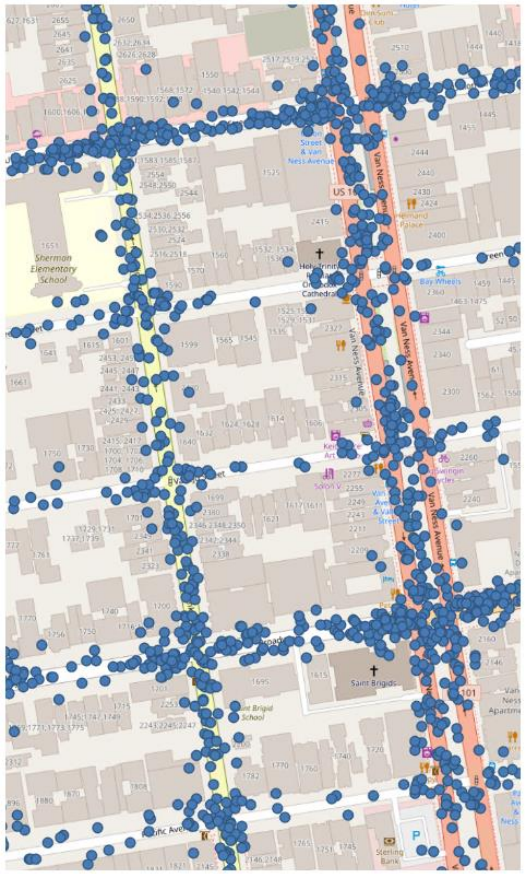


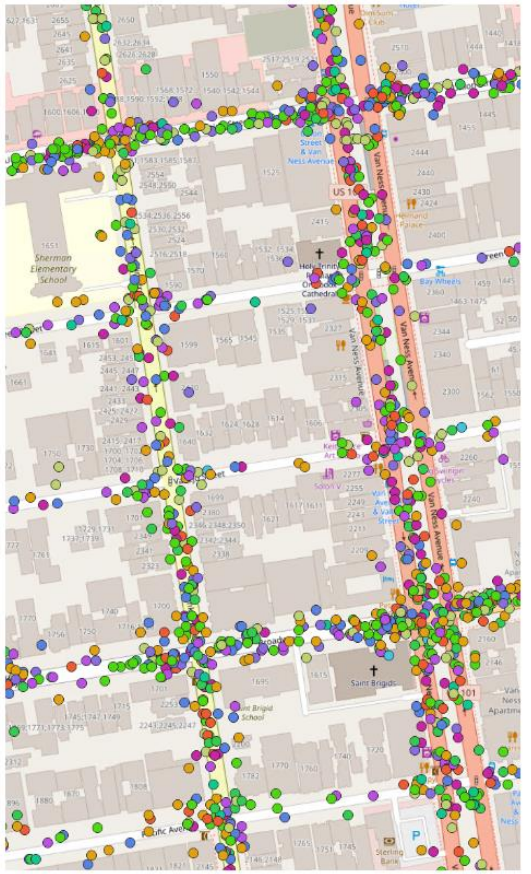
# Spatio-temporal Mobility Data

- More and more sensors on vehicles + telematics = large amounts of spatio-temporal mobility data becoming available
- Typically consists of massive amounts of structured data
  - Timestamp
  - GPS position
  - Additional information (vehicle id, recorded temperature, speed, etc…)
- Some peculiar pre-processing steps are required to apply KDD

# Trajectory Partitioning

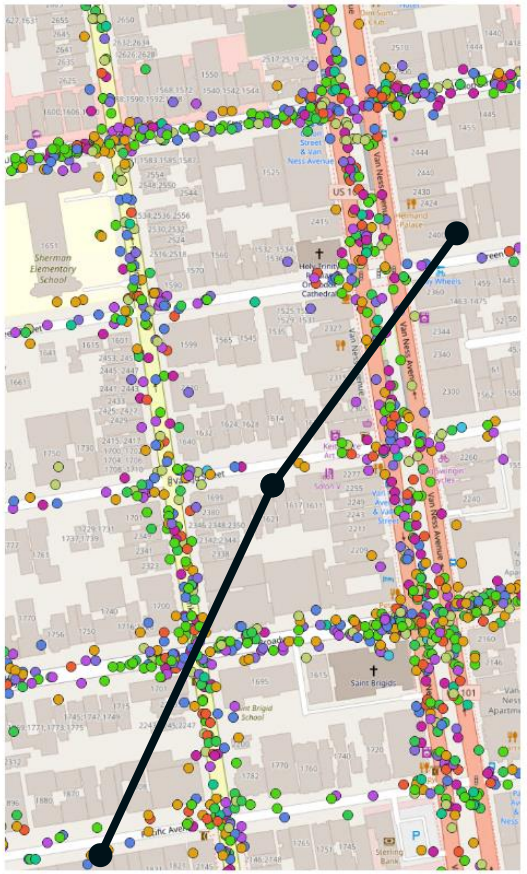
- Many datasets consist of a single stream of spatio-temporal data
- A first step in most analyses is to split such data stream according to some criteria (e.g.: by vehicle, by trip, etc…)
- In most cases, the goal is to get a set of independent routes from an origin to a destination

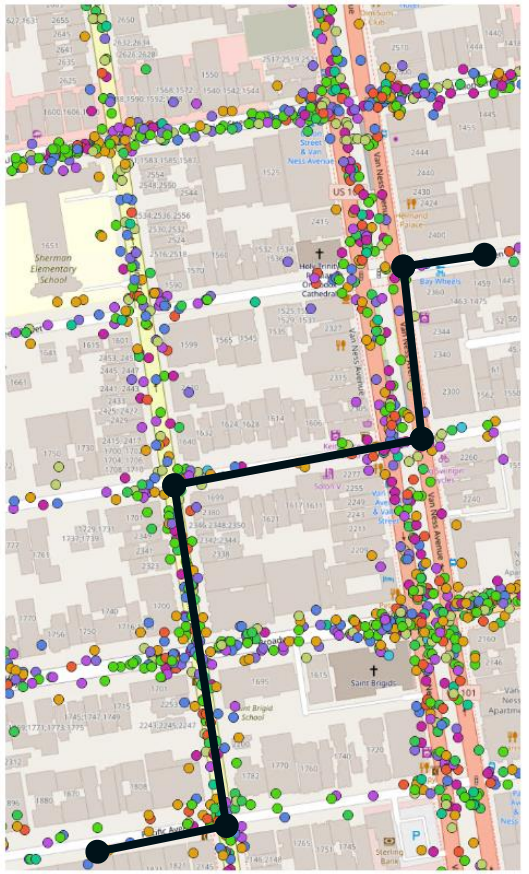




# Map Matching

- Connecting subsequent GPS positions is typically not good enough to reconstruct accurate trajectories
- Map Matching aims at **aligning** raw (possibly inaccurate) positioning data with an underlying **logical representation of the road network**





# Trajectory Restoration/Interpolation

- In presence of significant positioning errors and/or insufficient sampling rates, further processing of the trajectories might be required to reconstruct plausible trajectories
- In some cases, analysts might also be interested in manipulating the original trajectories to investigate what-if scenarios
  - What if all the taxis followed the shortest-path to destination?

# Next Steps

- After trajectory partitioning, map matching and trajectory restoration, the next steps depend on the goals of the analysis
- In many cases, **coverage analysis** might be useful
- Given a set of trajectories:
  - **How many times a certain road segment has been visited?**
  - **How frequently a certain area is visited by one of the monitored entities?**

# Working with spatio-temporal data: Challenges

- Visual-based Tools do not support these peculiar preprocessing steps
- Practitioners typically re-implement the entire pipeline from scratch
  - Limited re-usability
  - Limited replicability
  - Hinders productivity

KNOT Docs Blog Contributors Publications Search docs... /

# KNOT: a KNime mObility Toolkit

KNOT is a KNIME plugin helping you visually compose intuitive, reproducible, and easy to distribute Knowledge Discovery pipelines for massive mobility datasets including — but not limited to — Floating Car Data (FCD) ones.

release v2.0.0 license MIT issues 0 open stars 1

Get started

## Get started in minutes

Getting started with KNOT takes only a few minutes. You install our nodes in KNIME and you're ready to go solve challenging issues!

## Focus on what *really* matters

KNOT offers out-of-the-box support for map matching raw trajectories to OpenStreetMap models, so that you can focus on the analyses that really matter.

## We ❤ Open Data

KNOT supports OpenStreetMap data by default, and all the produced spatial objects are encoded in WKT format to maximize interoperability.

## Customizable

Most of the nodes provided by KNOT can be fine-tuned via the built-in dialogs, so that they can be effortlessly adapted to work with a number of different datasets and scenarios.

## Easily Extendable

KNOT is Open Source, and is designed to be easily extended. For example, if you are working on a novel map matching algorithm, you could easily extend our map matcher node to support it by implementing a simple Java interface.

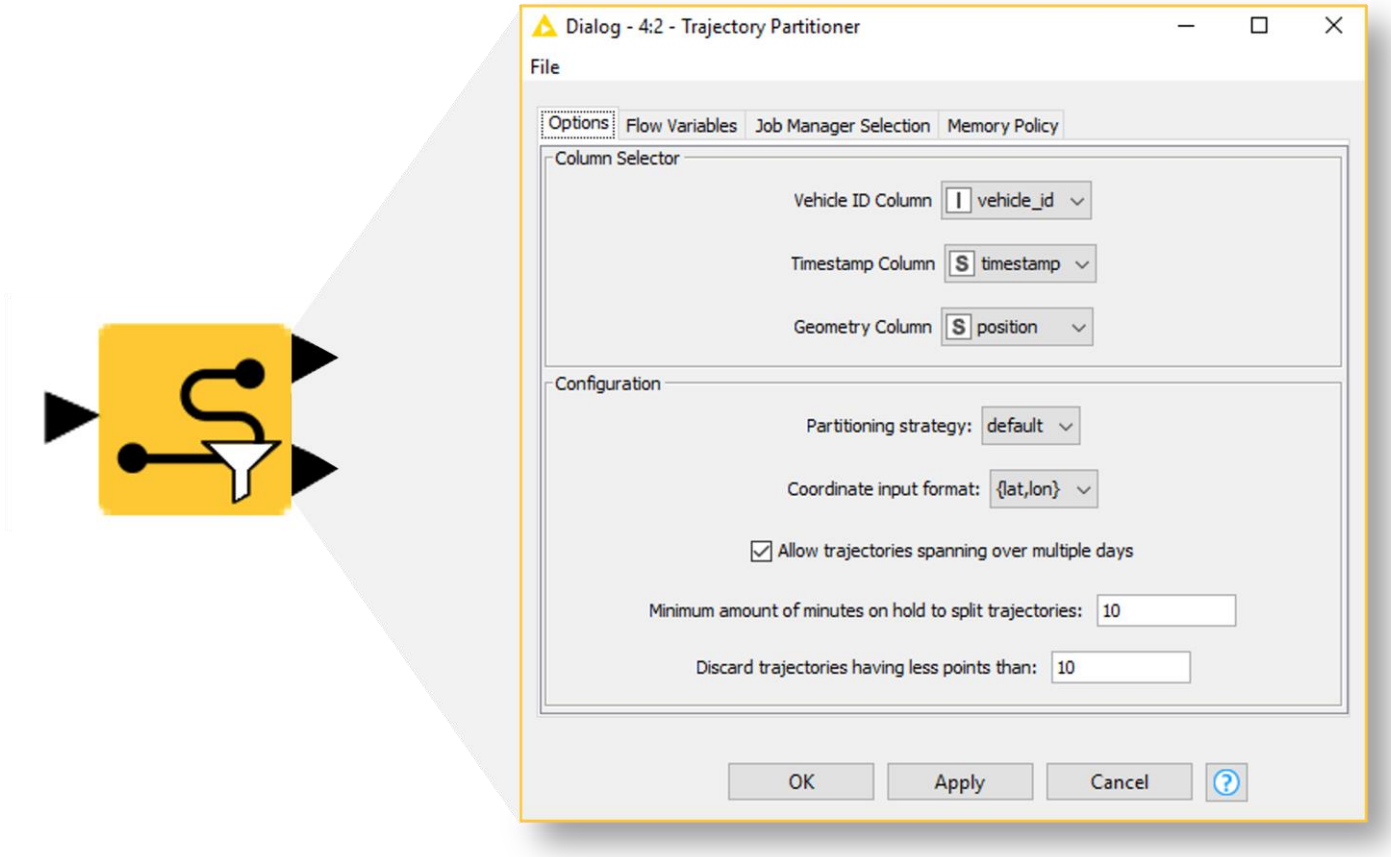
## Leverage the power of KNIME

Along with the custom nodes KNOT provides, you can leverage the full power of the well-known KNIME Analytics Platform and streamline your knowledge discovery process with hundreds of additional nodes provided out of the box.

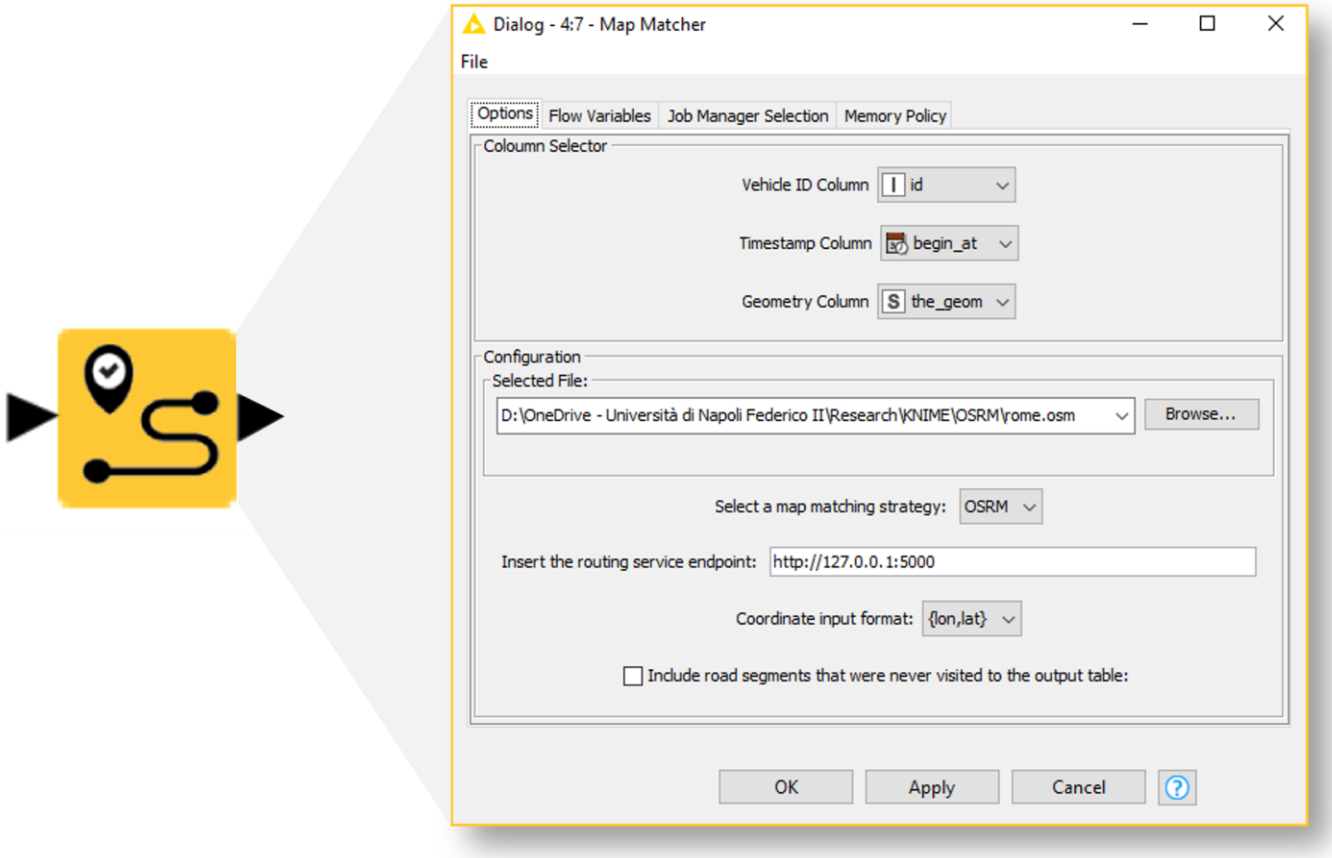
# The KNIME Mobility Toolkit

- Collection of dedicated KNIME components to support typical spatio-temporal data processing steps
- Open-source, freely available (currently in the nightly builds)
- Each component can be customized and extended
- https://github.com/knot/

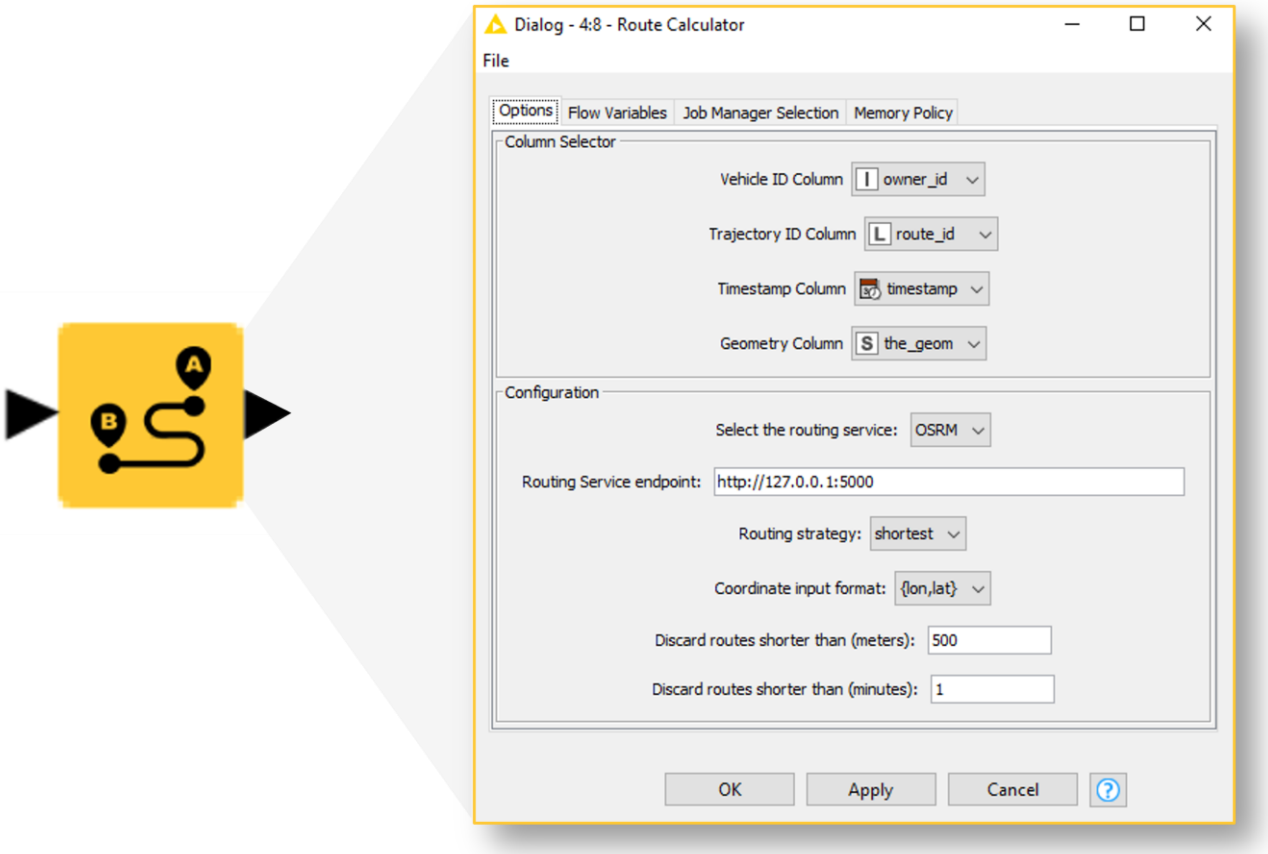
# Trajectory Partitioner



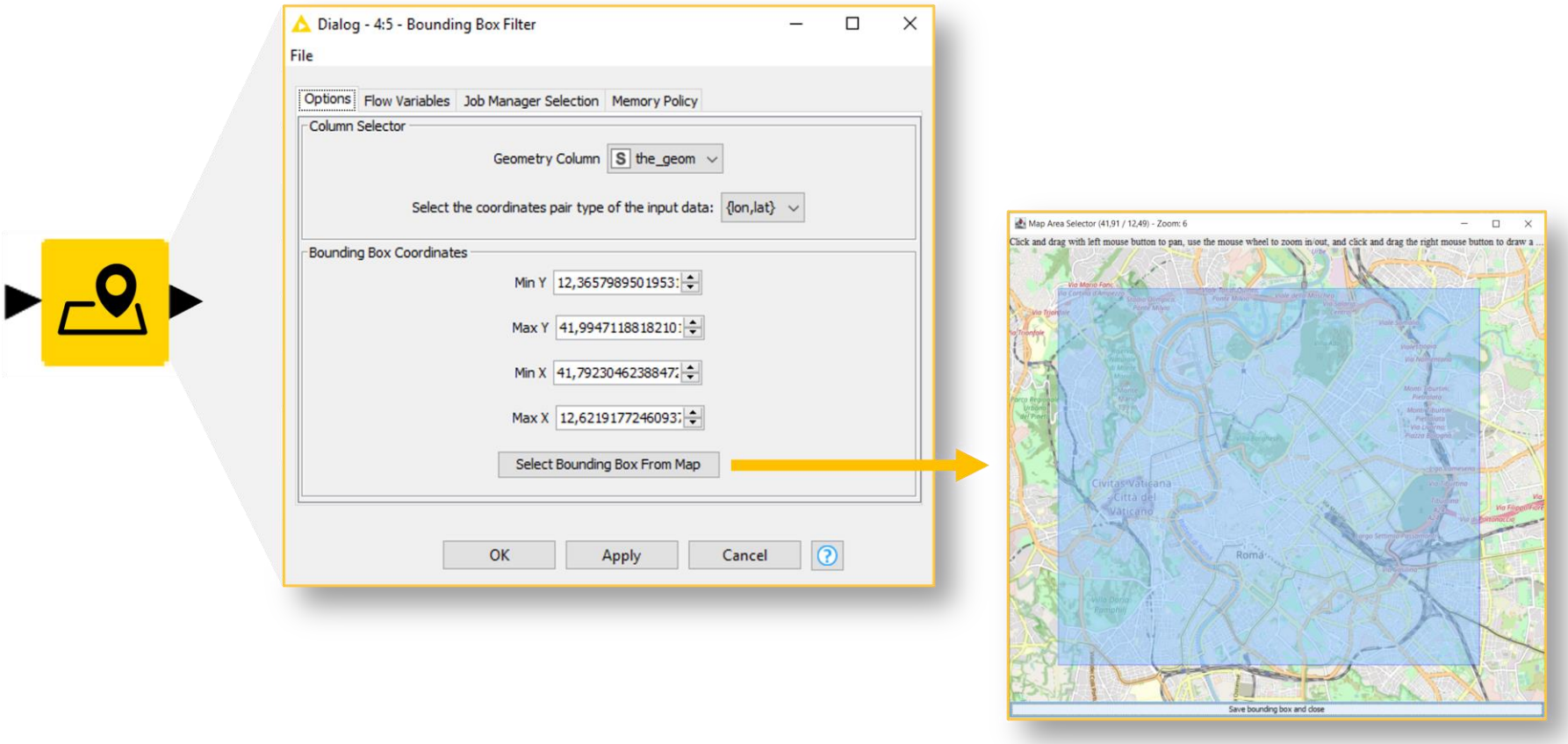
# Map Matcher


Dialog - 4:7 - Map Matcher
File
Options
Flow Variables
Job Manager Selection
Memory Policy
Coloumn Selector
Vehicle ID Column
id
Timestamp Column
begin_at
Geometry Column
the_geom
Configuration
Selected File:
D:\OneDrive - Università di Napoli Federico II\Research\KNIME\OSRM\rome.osm
Browse...
Select a map matching strategy:
OSRM
Insert the routing service endpoint:
http://127.0.0.1:5000
Coordinate input format:
{lon,lat}
Include road segments that were never visited to the output table:
OK
Apply
Cancel

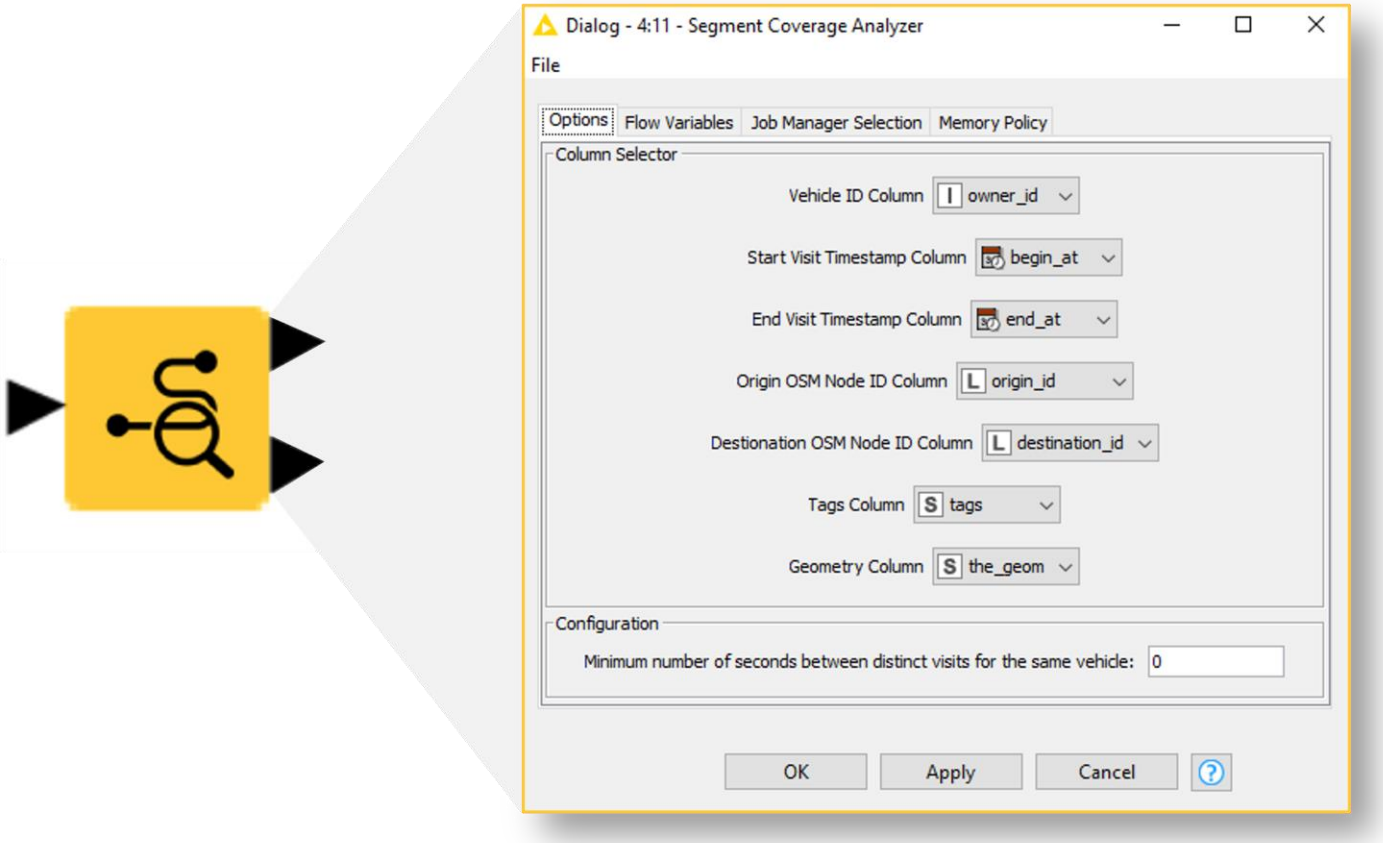
# Route Calculator



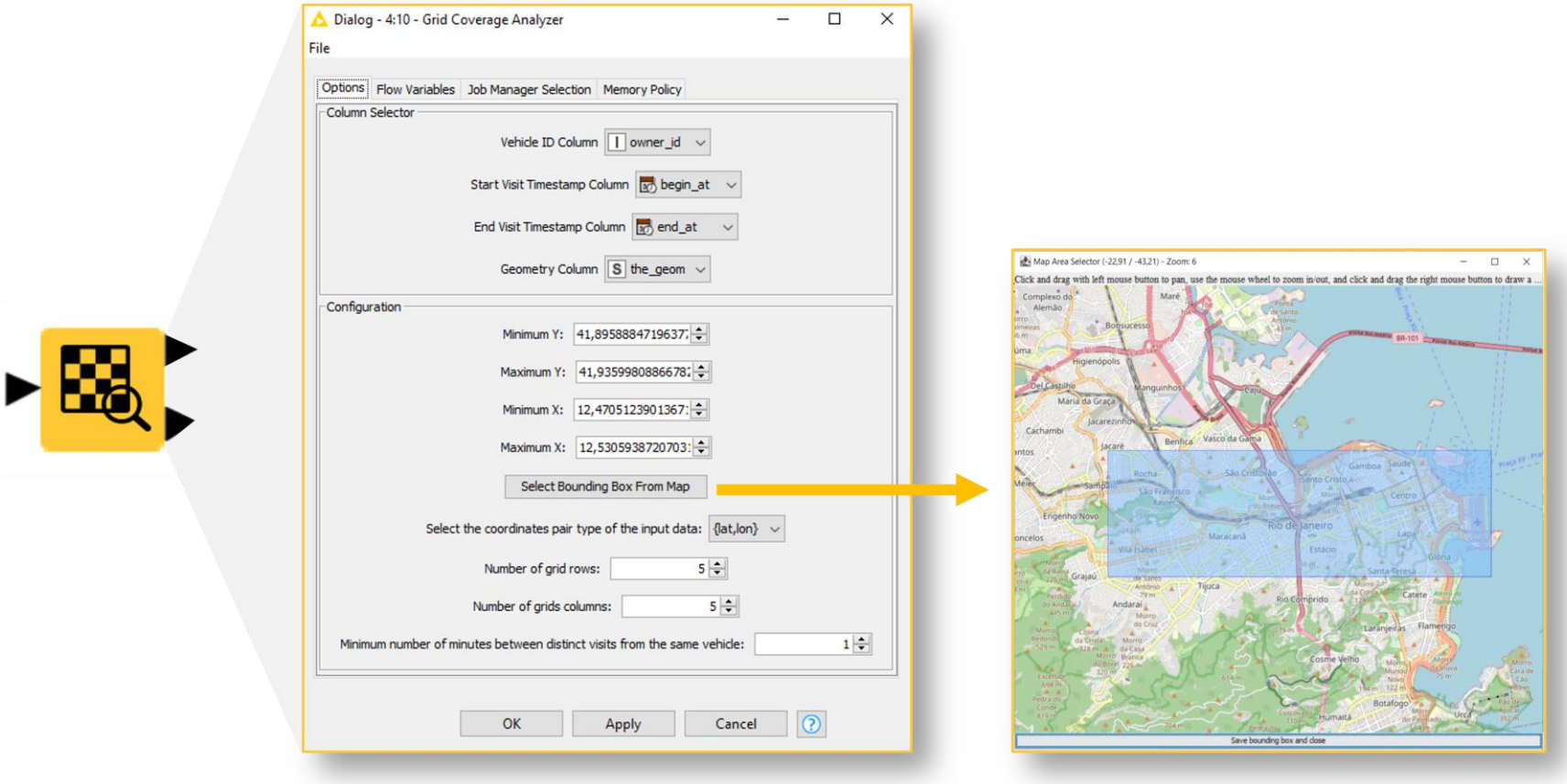
# Bounding Box Filter



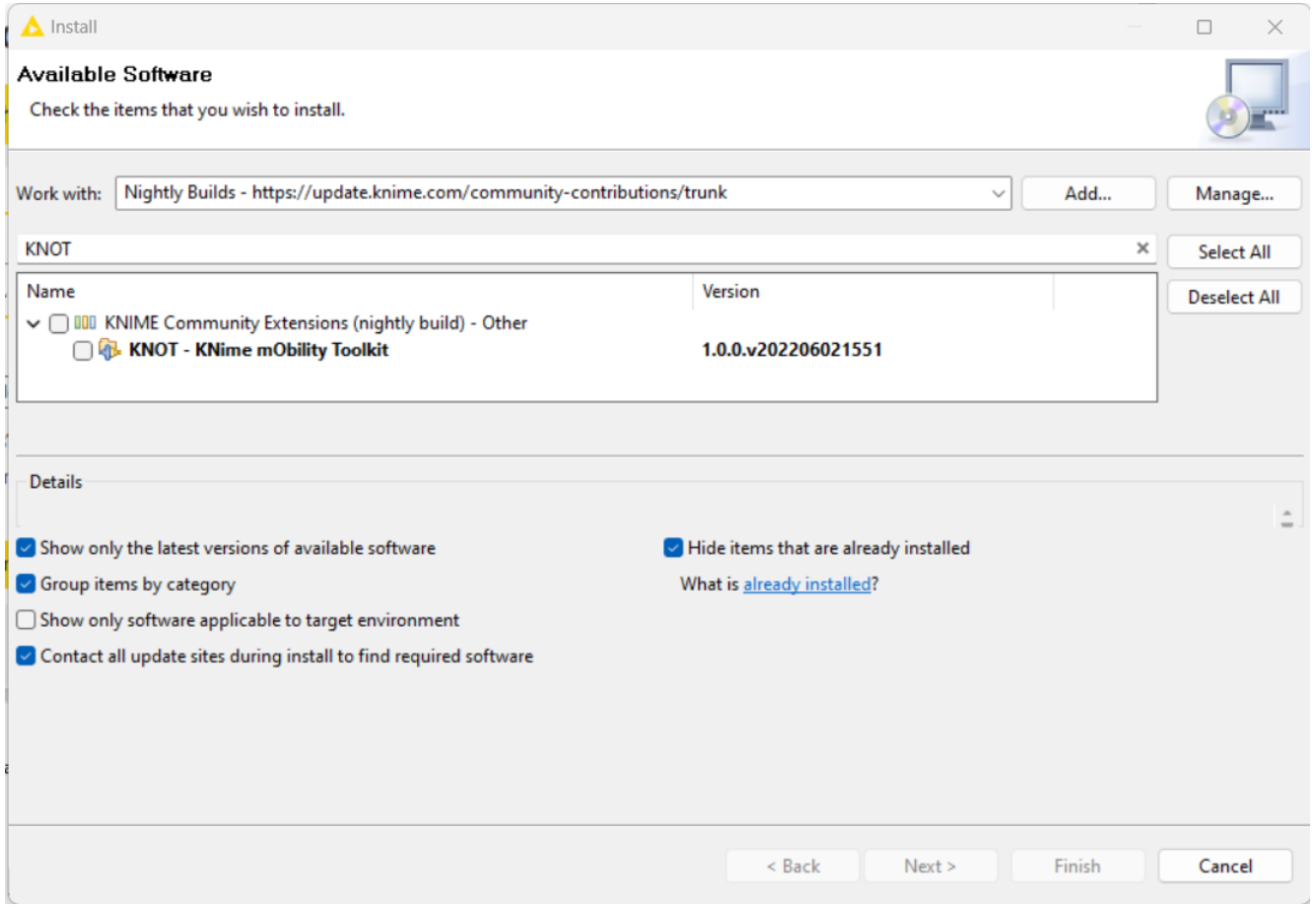
# Segment Coverage Analyzer



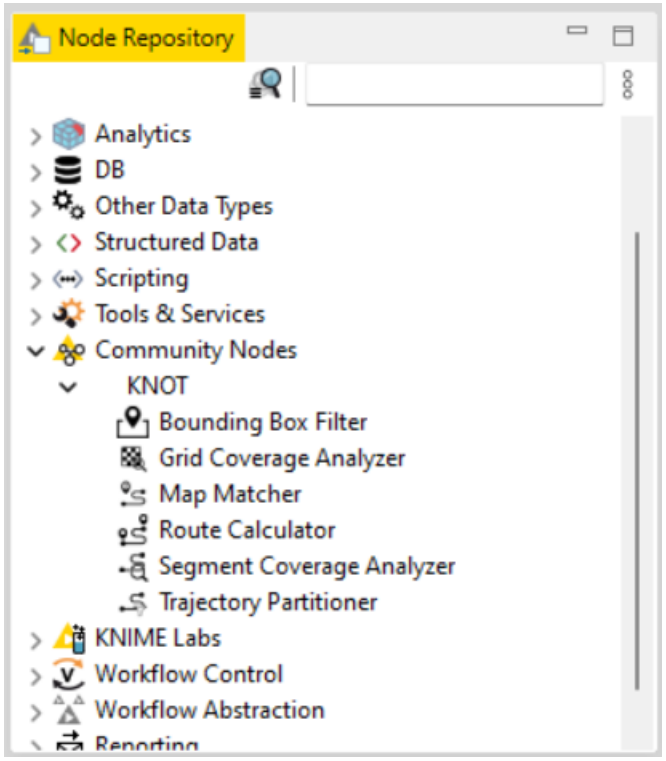
# Grid Coverage Analyzer



# Installing KNOT



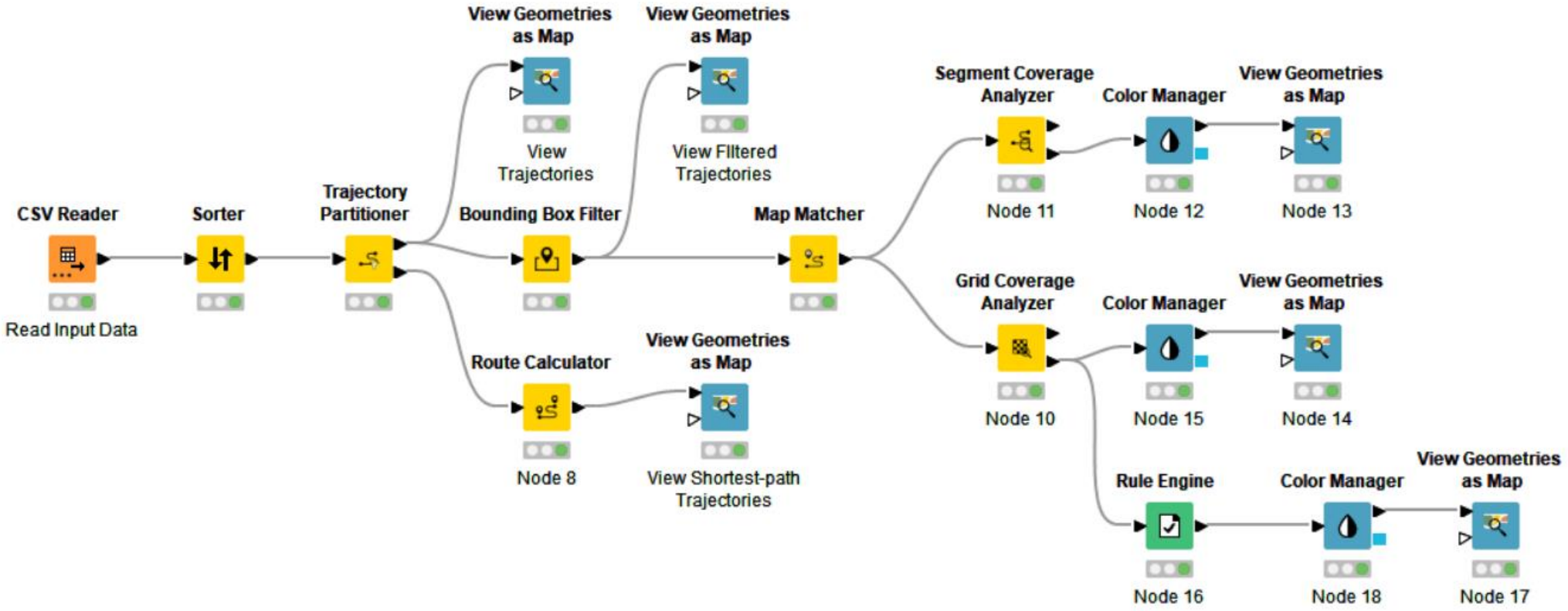
# Installing KNOT



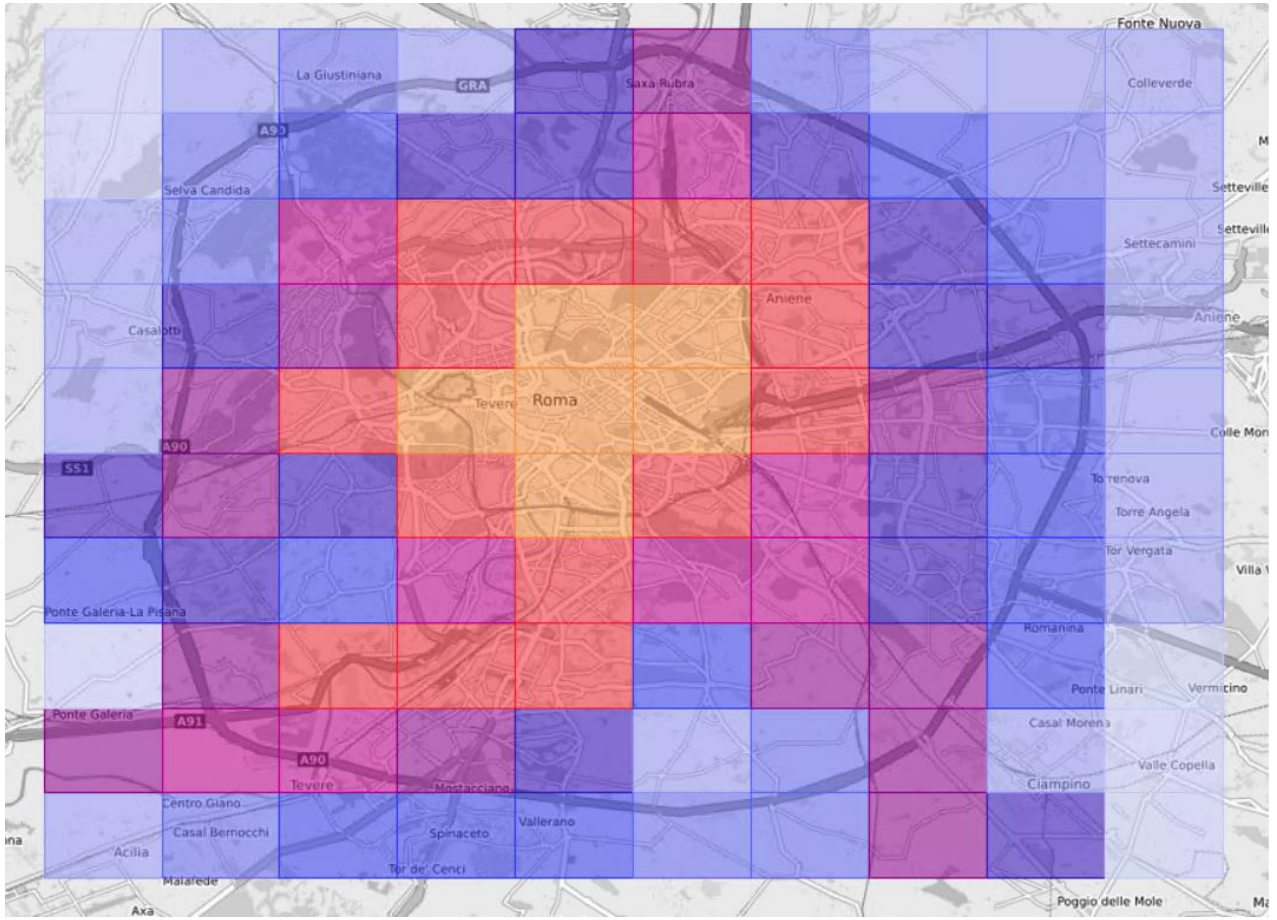
# Spatio-temporal Data Analytics: Real-world scenarios

- Can taxis be used for Vehicular Crowd-Sensing in Rome?
- How pervasive is public transportation in Rio de Janeiro?

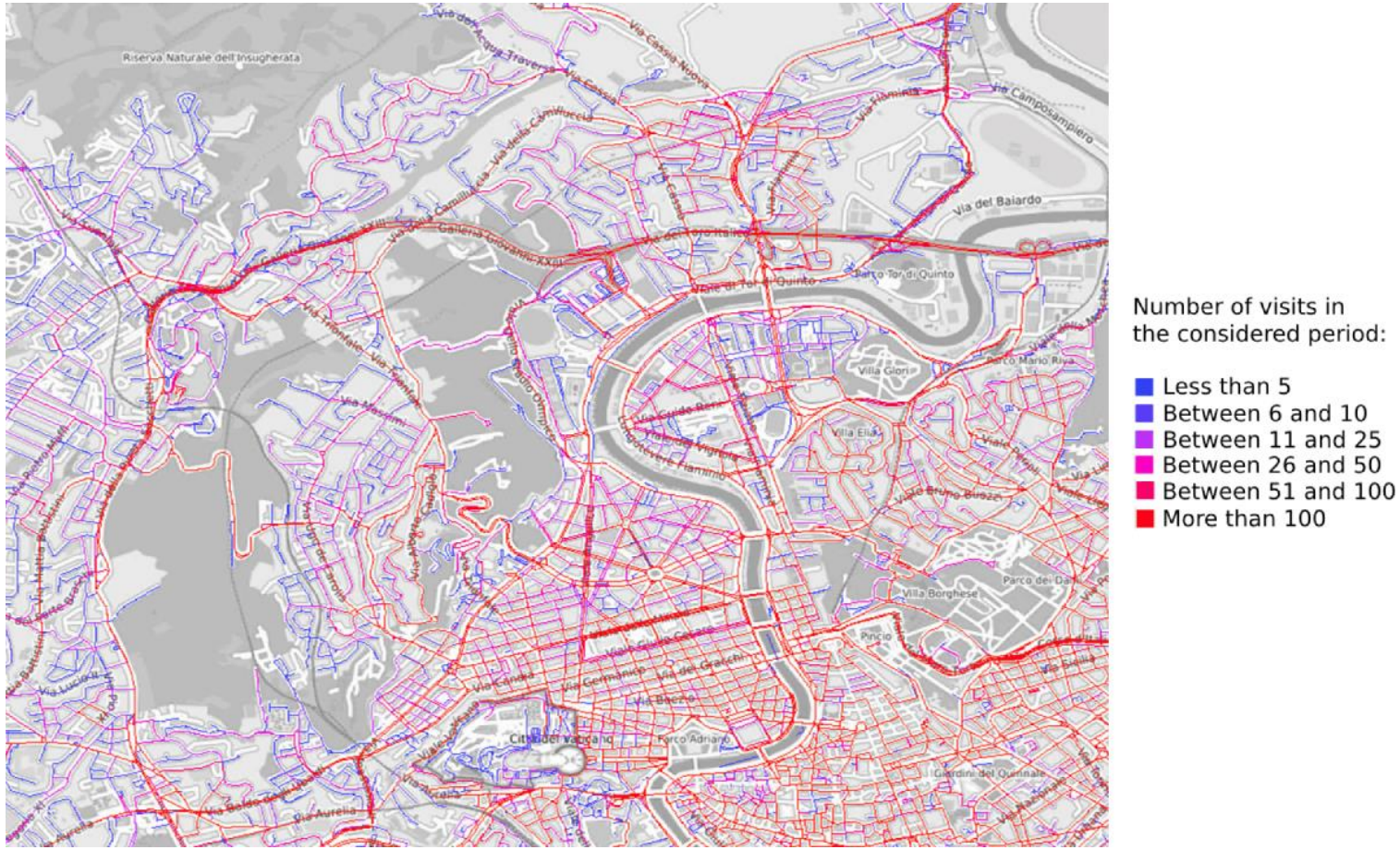
# Road Network Coverage in Rome



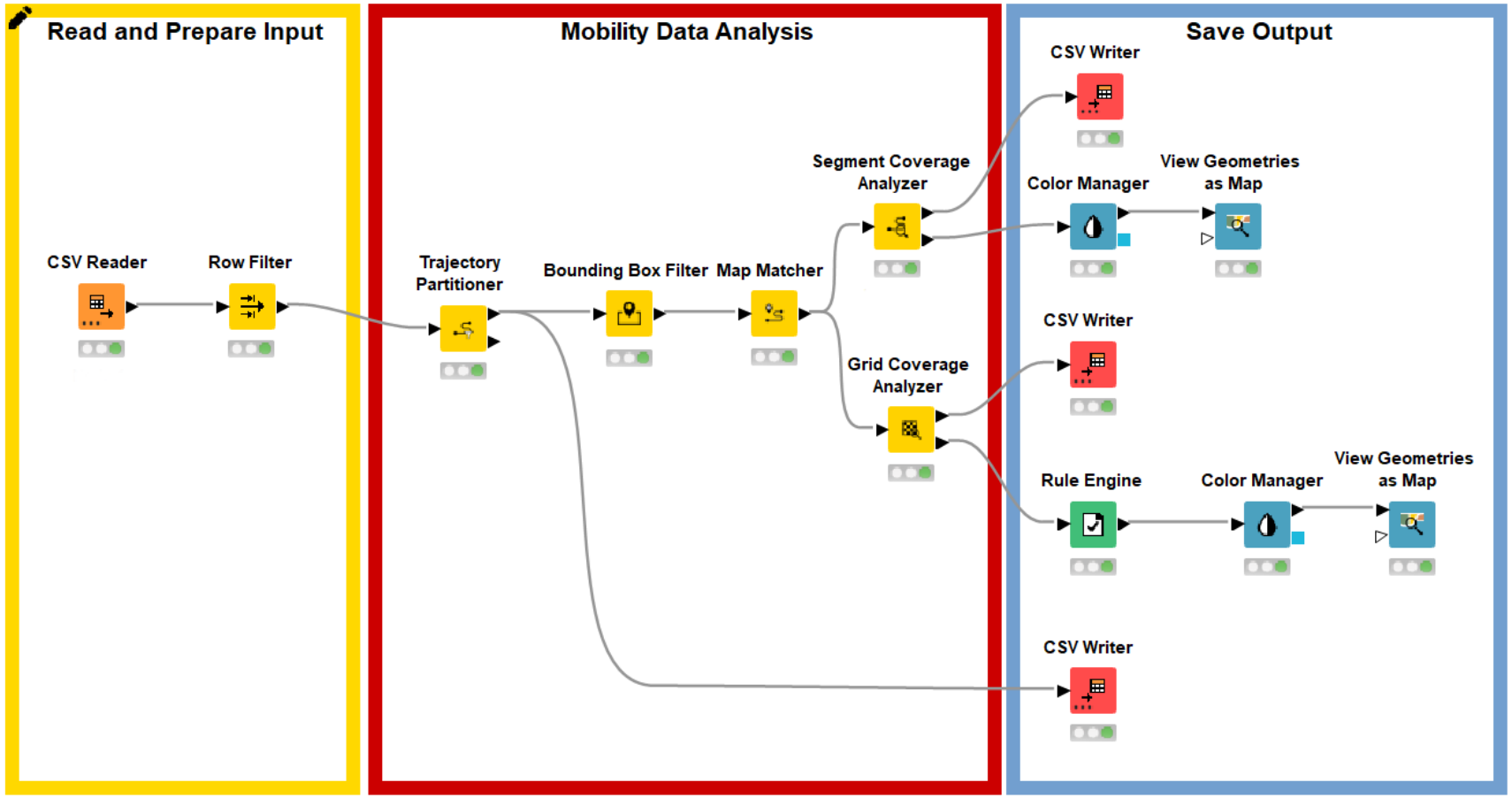
# Area Coverage



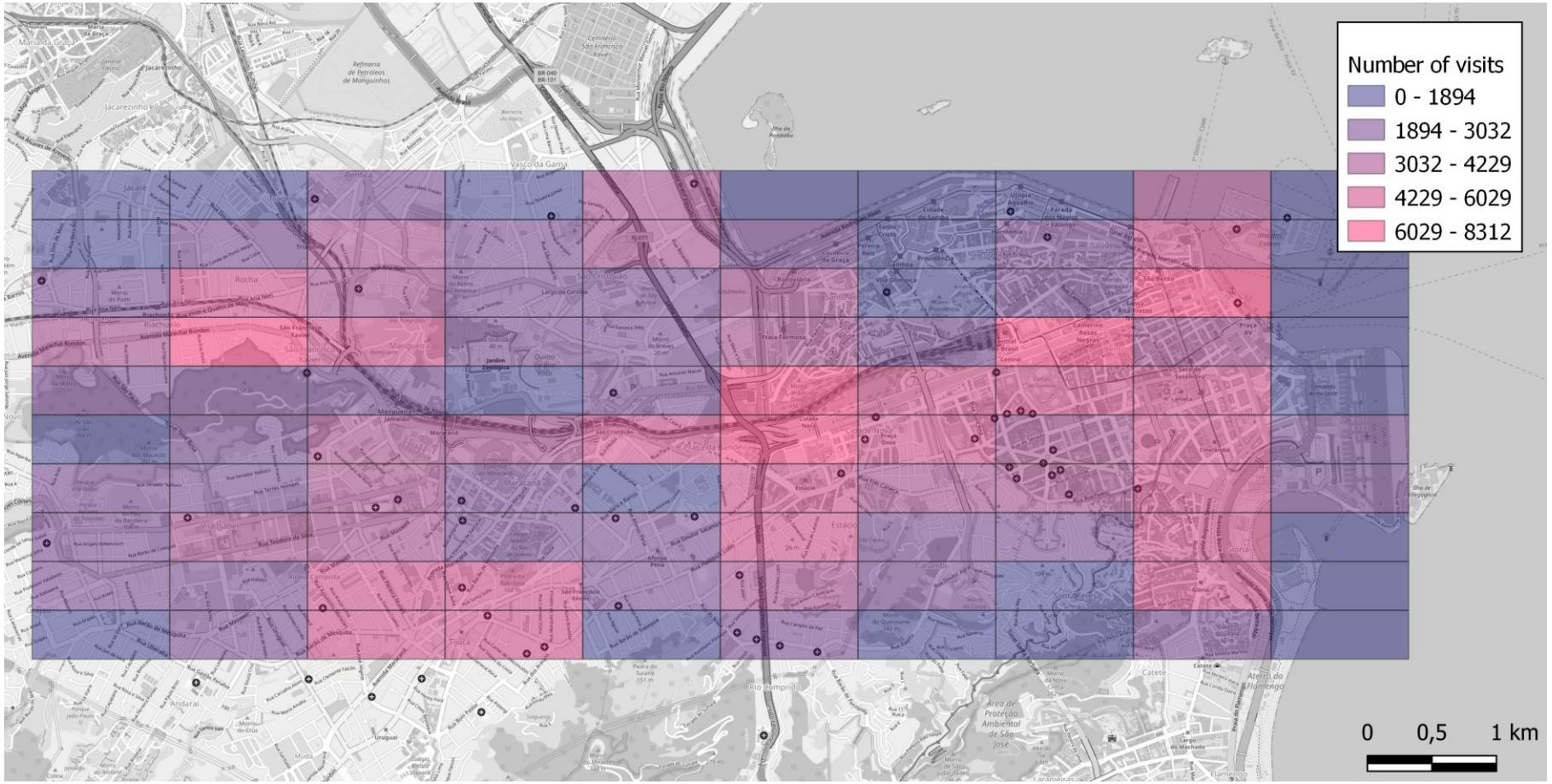
# Segment Coverage



# Public Transit Coverage in Rio De Janeiro



# Area Coverage



# Take Home Messages

- Knowledge Discovery from Data
- Implementing KDD pipelines
- Visual Analytics Platform (KNIME)
- Challenges of KDD on Spatio-Temporal Mobility Data
- The KNIME Mobility Toolkit