

# Spatio-temporal Data Analytics with KNIME

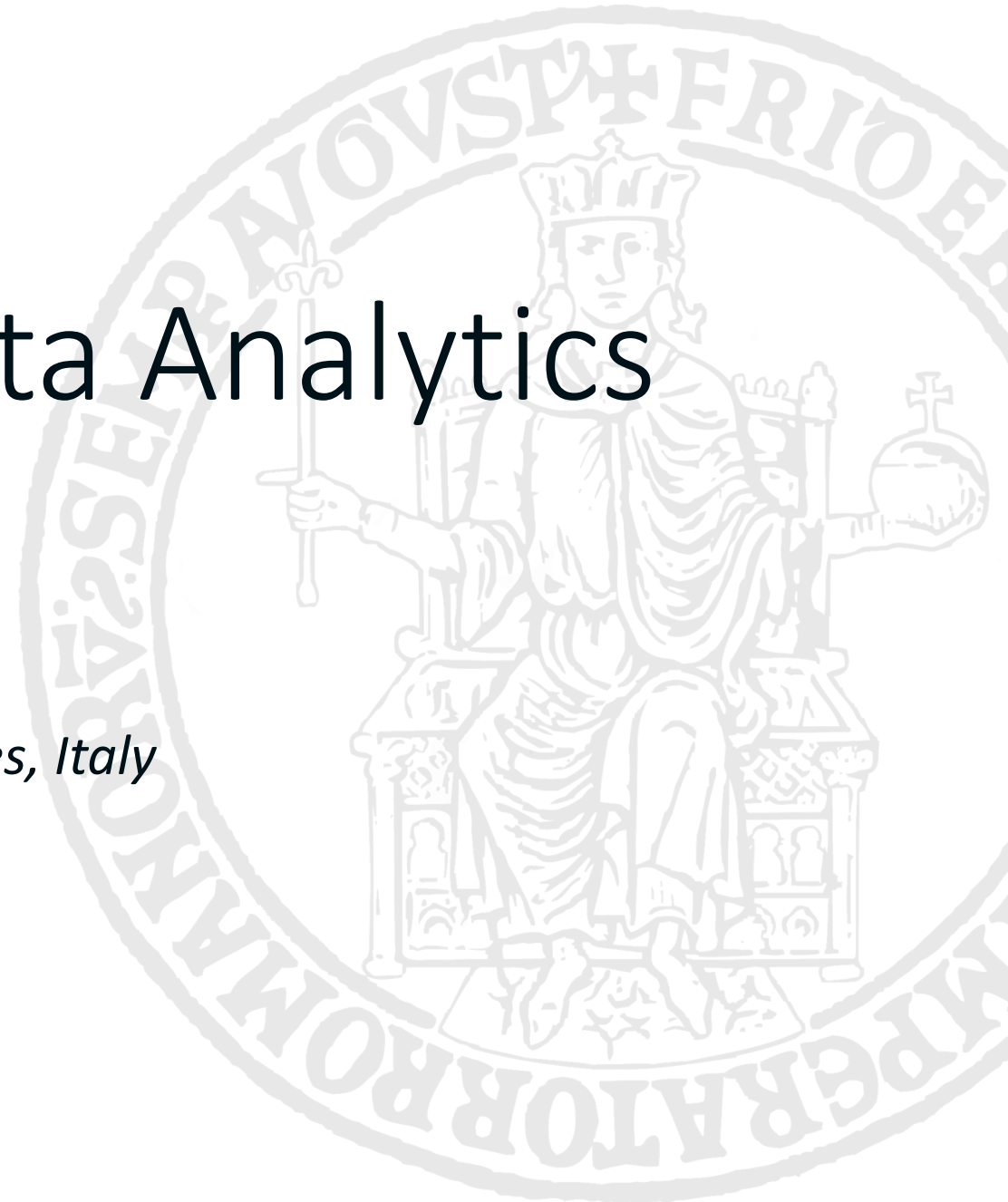
Luigi Libero Lucio Starace, Ph.D.

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

[luigiliberolucio.starace@unina.it](mailto:luigiliberolucio.starace@unina.it)

<https://luistar.github.io>

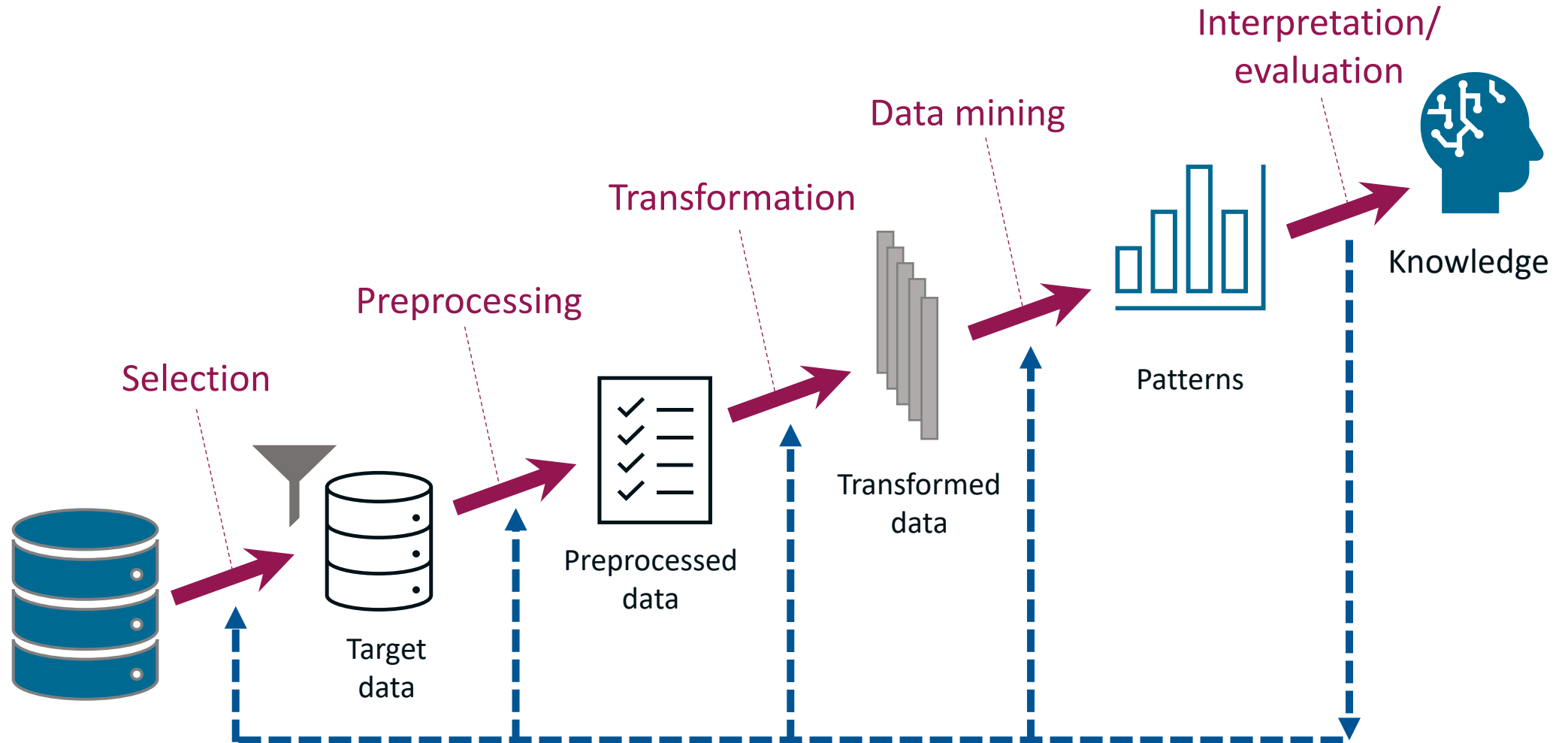
April 18, 2023 – University College Dublin



# 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

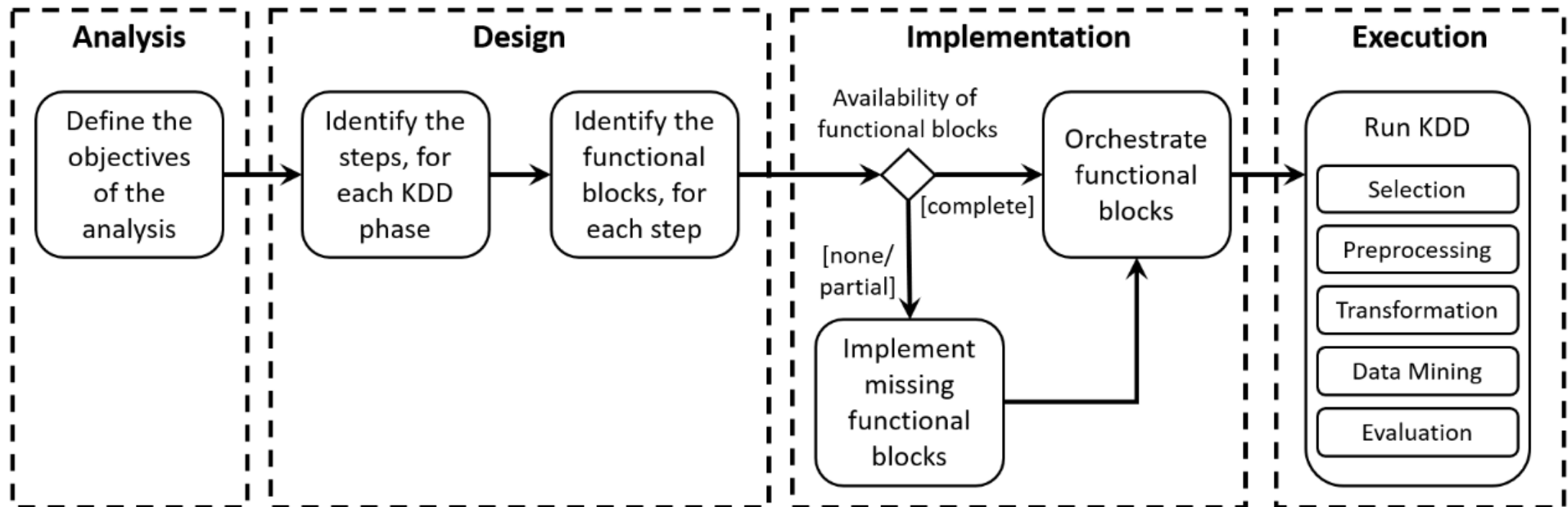
# Knowledge Discovery from Data



# 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)

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow with the following nodes: **Table Reader**, **Extract Column Header** (with sub-nodes **Get column names** and **Create column with column names**), **Transpose**, **Color Manager** (with sub-nodes **Create column with column names** and **Configure colors for each column**), and two **Bar Chart** nodes. The first bar chart is a horizontal sum chart, and the second is a grouped chart with custom colors. A yellow label **WORKFLOW** is placed over the workflow diagram.

On the left, the **Node Repository** panel is visible, with a yellow label **AVAILABLE NODES** highlighting it. The repository lists various node categories such as IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, and KNIME Labs.

On the right, the **Description** panel for the workflow is shown, with a yellow label **INTERACTIVE VISUALIZATION** highlighting it. The description includes the title **04\_Example...** and a description of bar charts. Below the description is an interactive visualization window titled **Grouped Bar Chart**. This window shows a legend with four categories: **sepal length** (blue), **sepal width** (orange), **petal length** (green), and **petal width** (red). The chart displays grouped bars for three species: **Iris-setosa**, **Iris-versicolor**, and **Iris-virginica**. The x-axis ranges from 0 to 329.4. At the bottom of the window are **Reset**, **Apply**, and **Close** buttons.

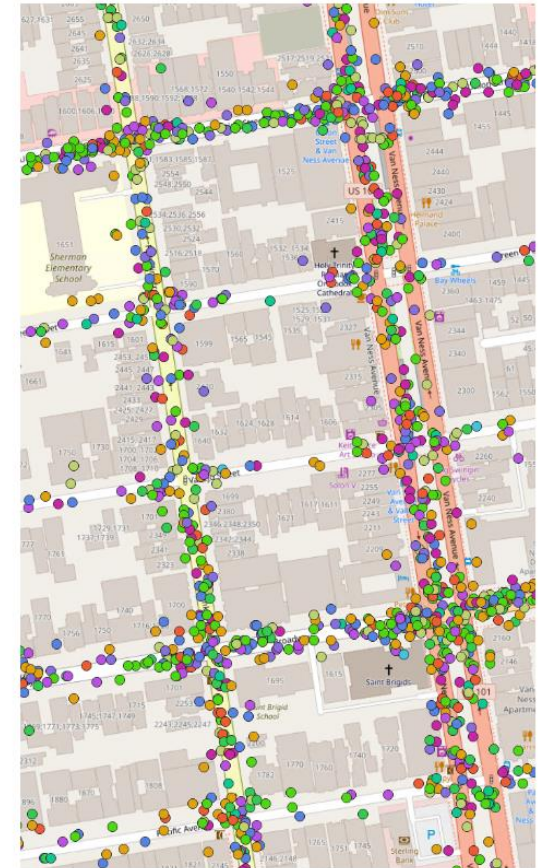
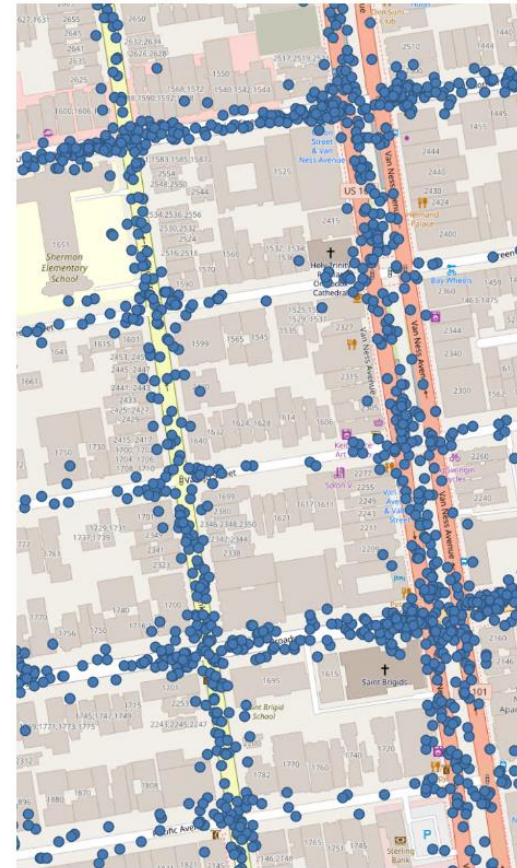
The **KNIME Console** at the bottom shows log messages, including a warning about the log file location and several **WARN** messages for **Radial Plot Appender**, **GroupBy**, and **Color Manager**.

# 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

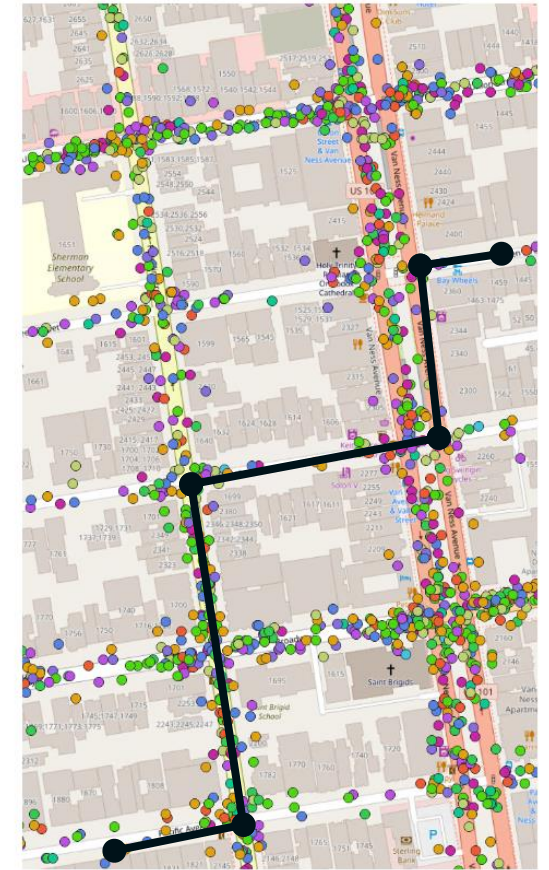
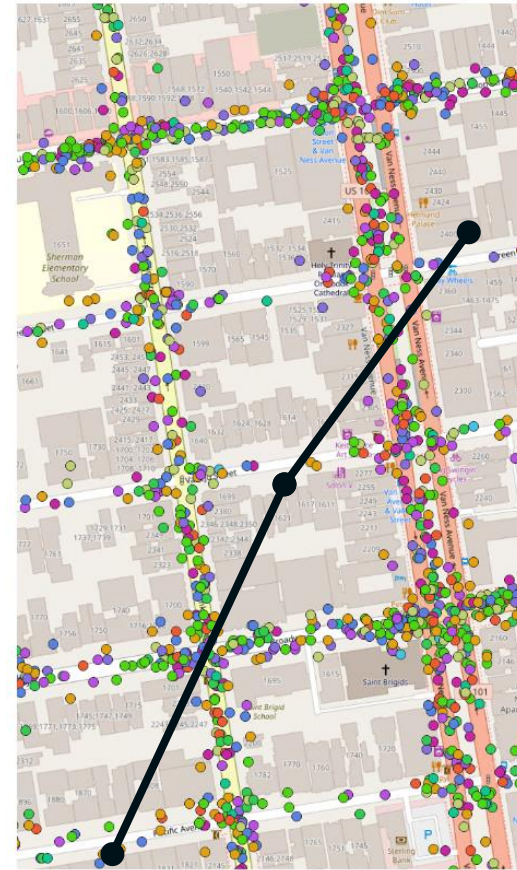
- Many datasets consist of a single stream of spatio-temporal data
- A first step in most analyses is to split such data stream according to some criteria (e.g.: by vehicle, by trip, etc...)
- In most cases, the goal is to get a set of independent routes from an origin to a destination





# 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: 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.

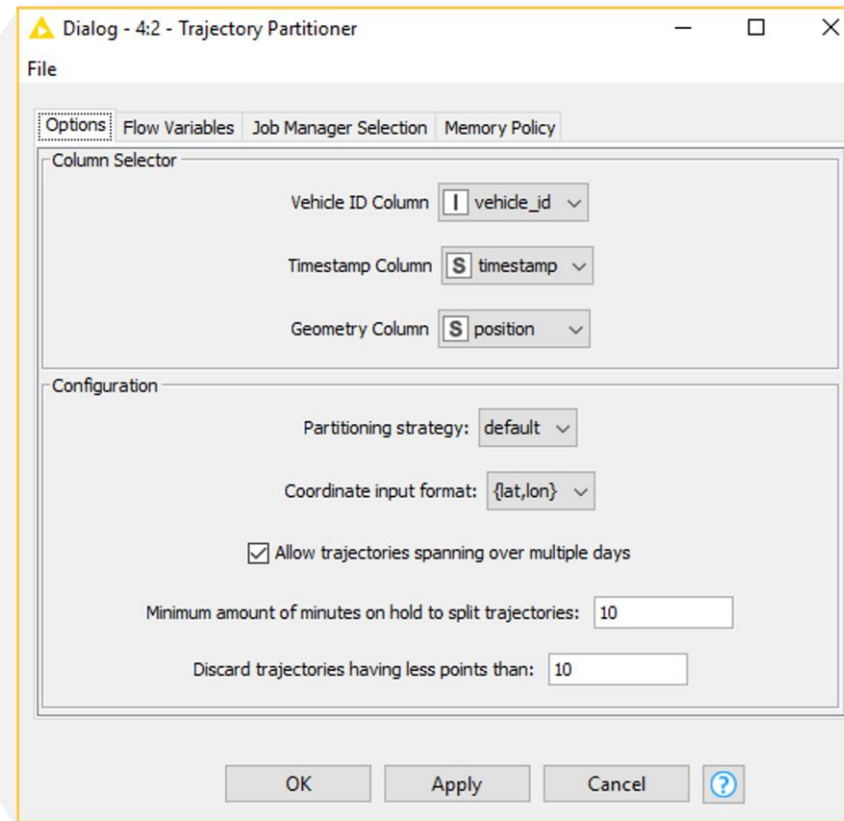
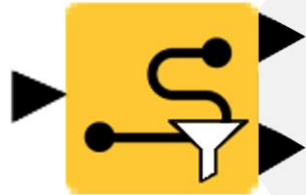
## Leverage the power of KNIME

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.

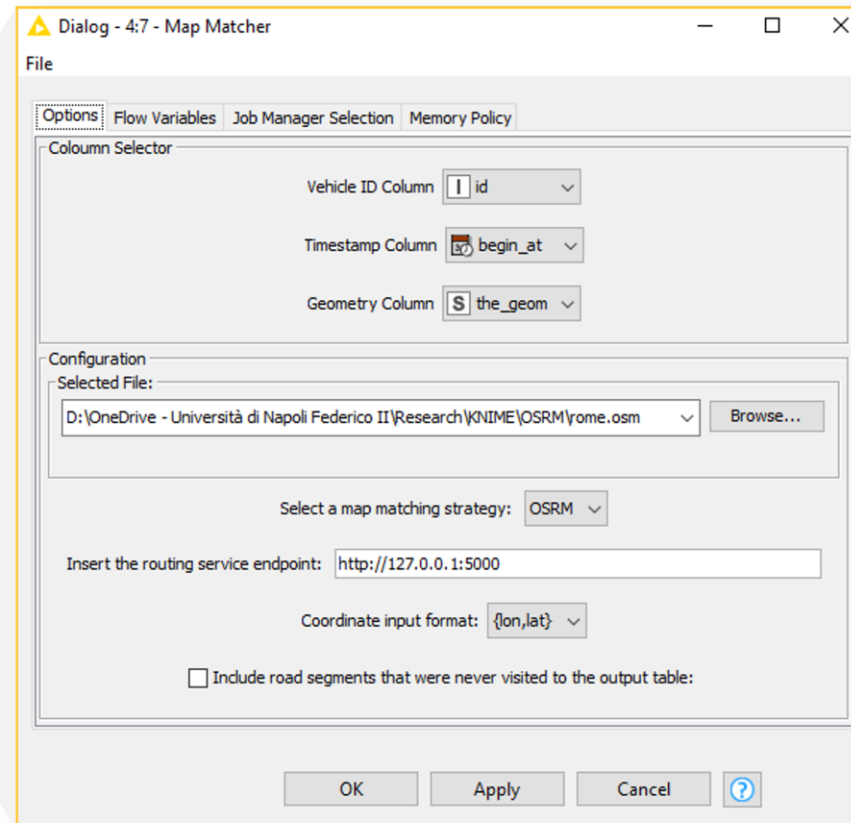
# The KNIME Mobility Toolkit

- 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

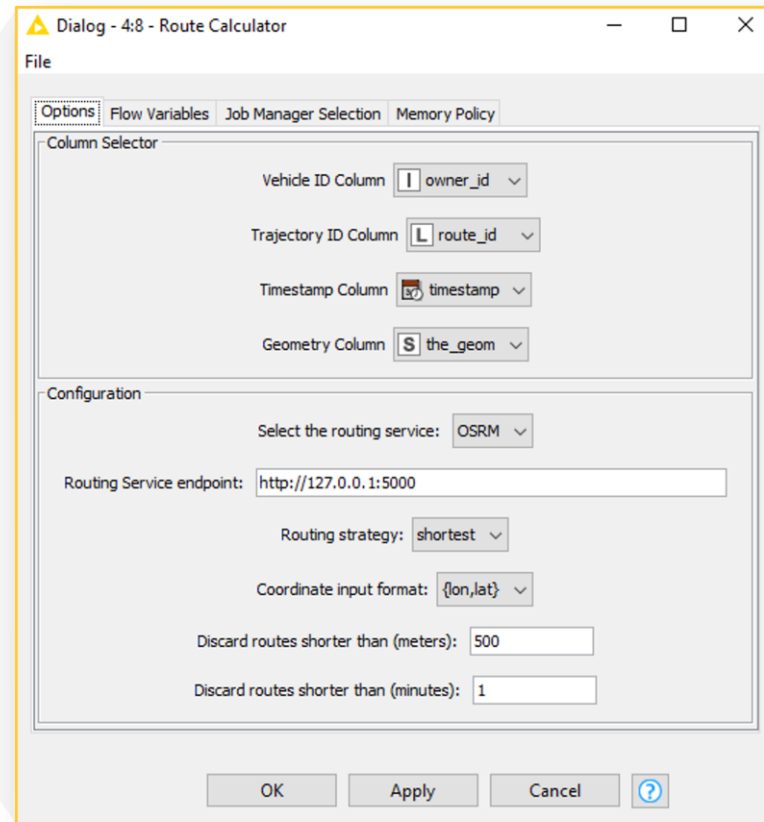
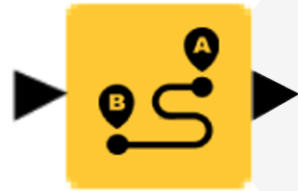


# Map Matcher

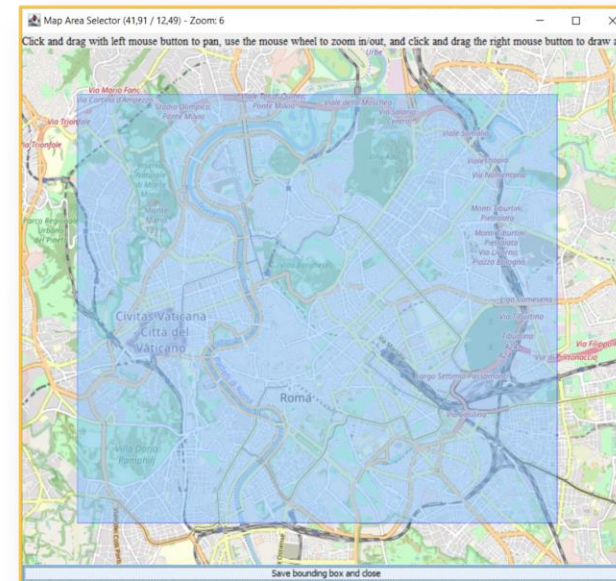
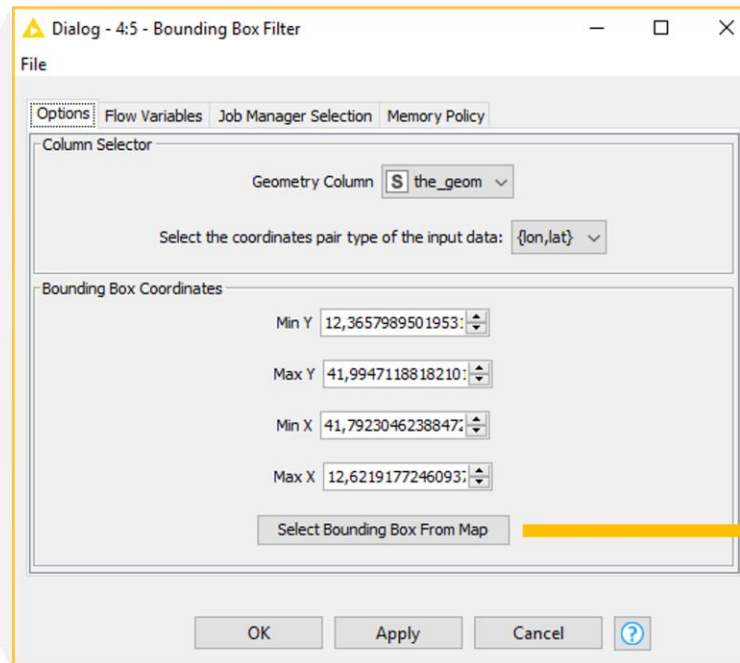




# Route Calculator



# Bounding Box Filter



# Segment Coverage Analyzer



Dialog - 4:11 - Segment Coverage Analyzer

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Column Selector

Vehicle ID Column

Start Visit Timestamp Column

End Visit Timestamp Column

Origin OSM Node ID Column

Destination OSM Node ID Column

Tags Column

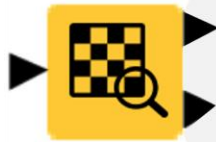
Geometry Column

Configuration

Minimum number of seconds between distinct visits for the same vehicle:

OK Apply Cancel ?

# Grid Coverage Analyzer



Dialog - 4:10 - Grid Coverage Analyzer

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Column Selector

Vehicle ID Column:

Start Visit Timestamp Column:

End Visit Timestamp Column:

Geometry Column:

Configuration

Minimum Y:

Maximum Y:

Minimum X:

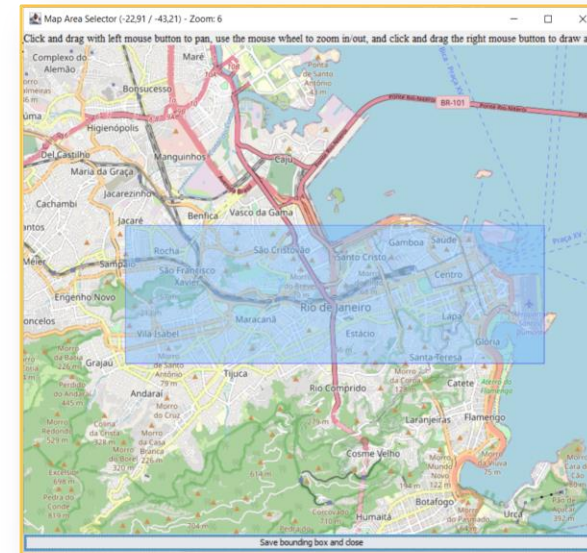
Maximum X:

Select the coordinates pair type of the input data:

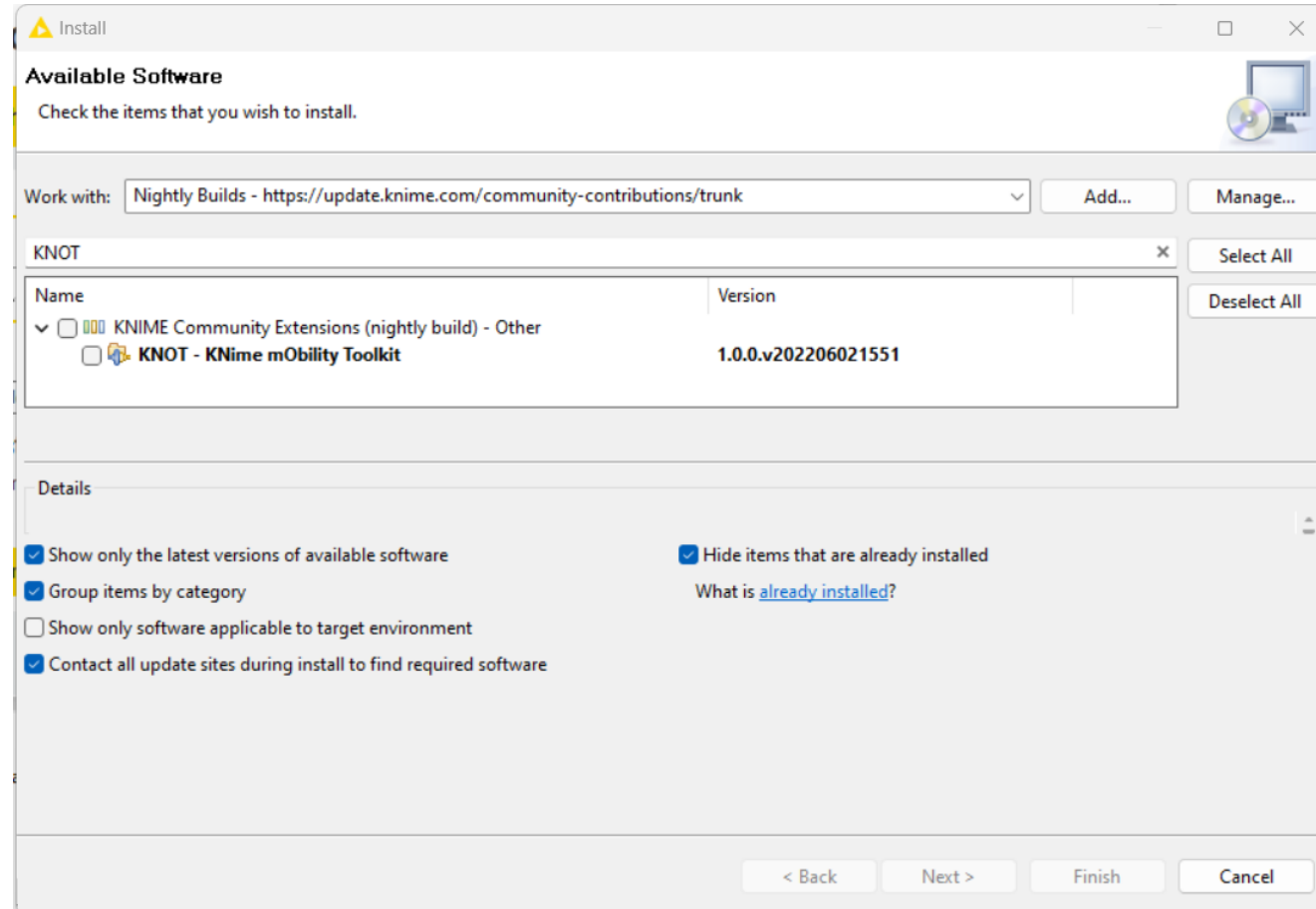
Number of grid rows:

Number of grids columns:

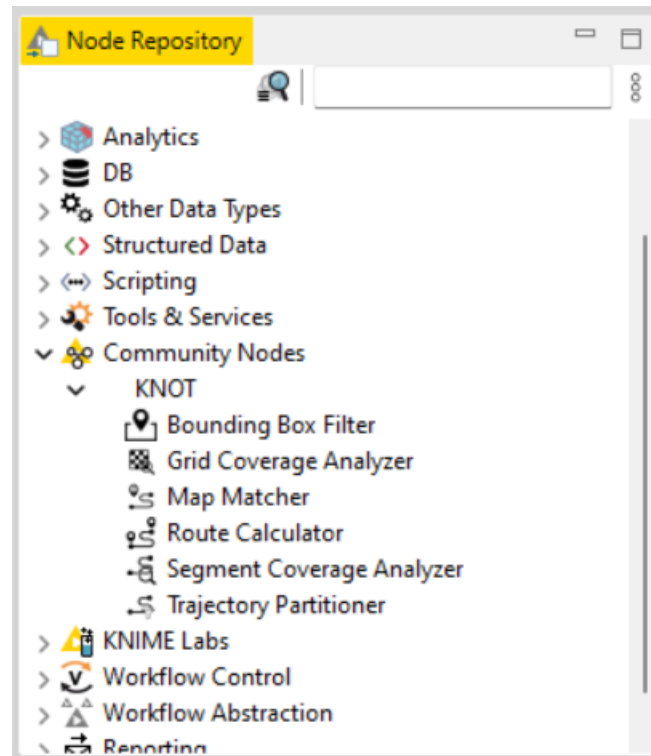
Minimum number of minutes between distinct visits from the same vehicle:



# Installing KNOT



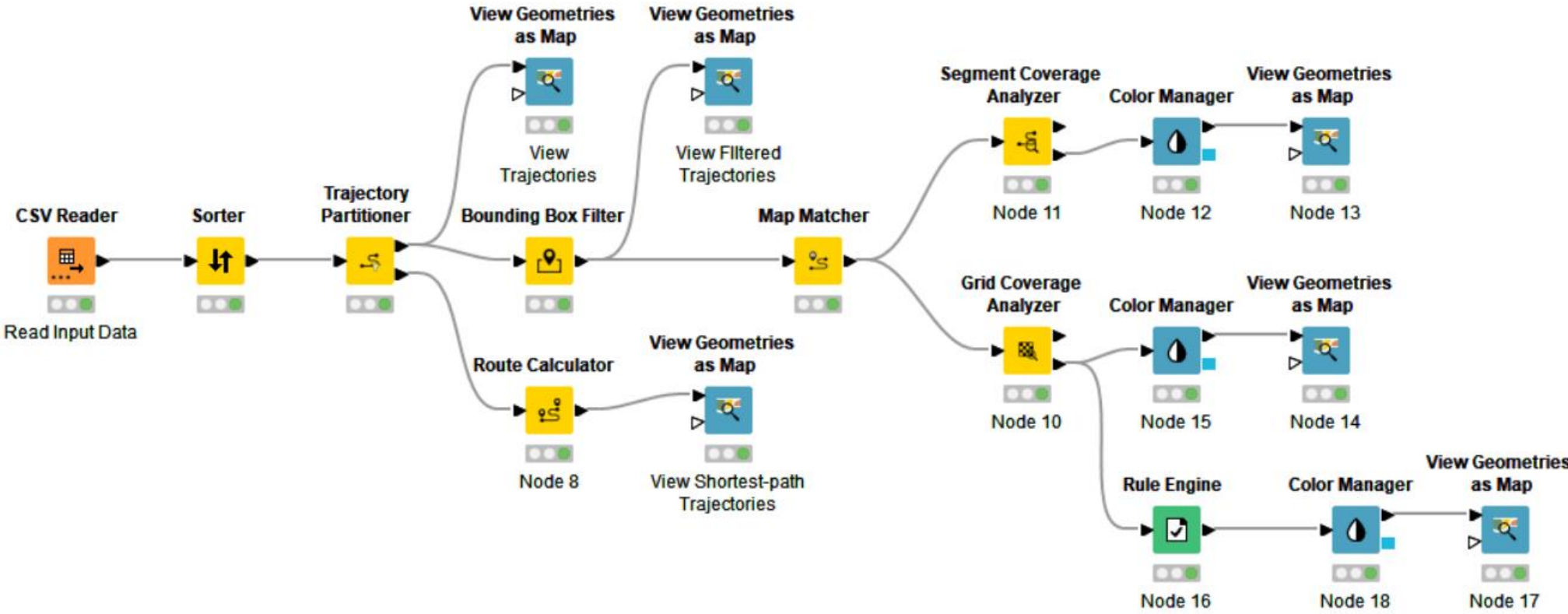
# Installing KNOT



# Spatio-temporal Data Analytics: Real-world scenarios

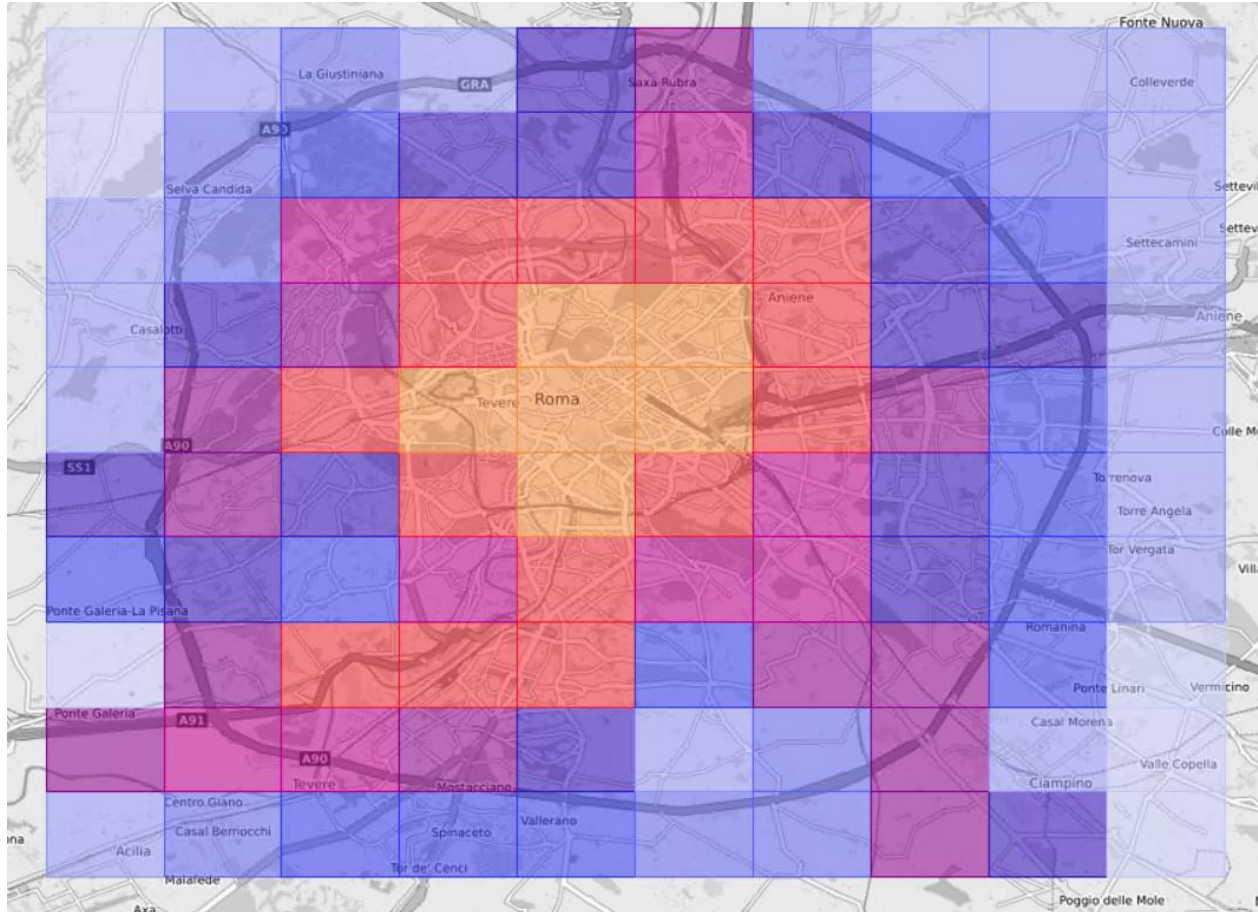
- Can taxis be used for Vehicular Crowd-Sensing in Rome?
- How pervasive is public transportation in Rio de Janeiro?

# Road Network Coverage in Rome





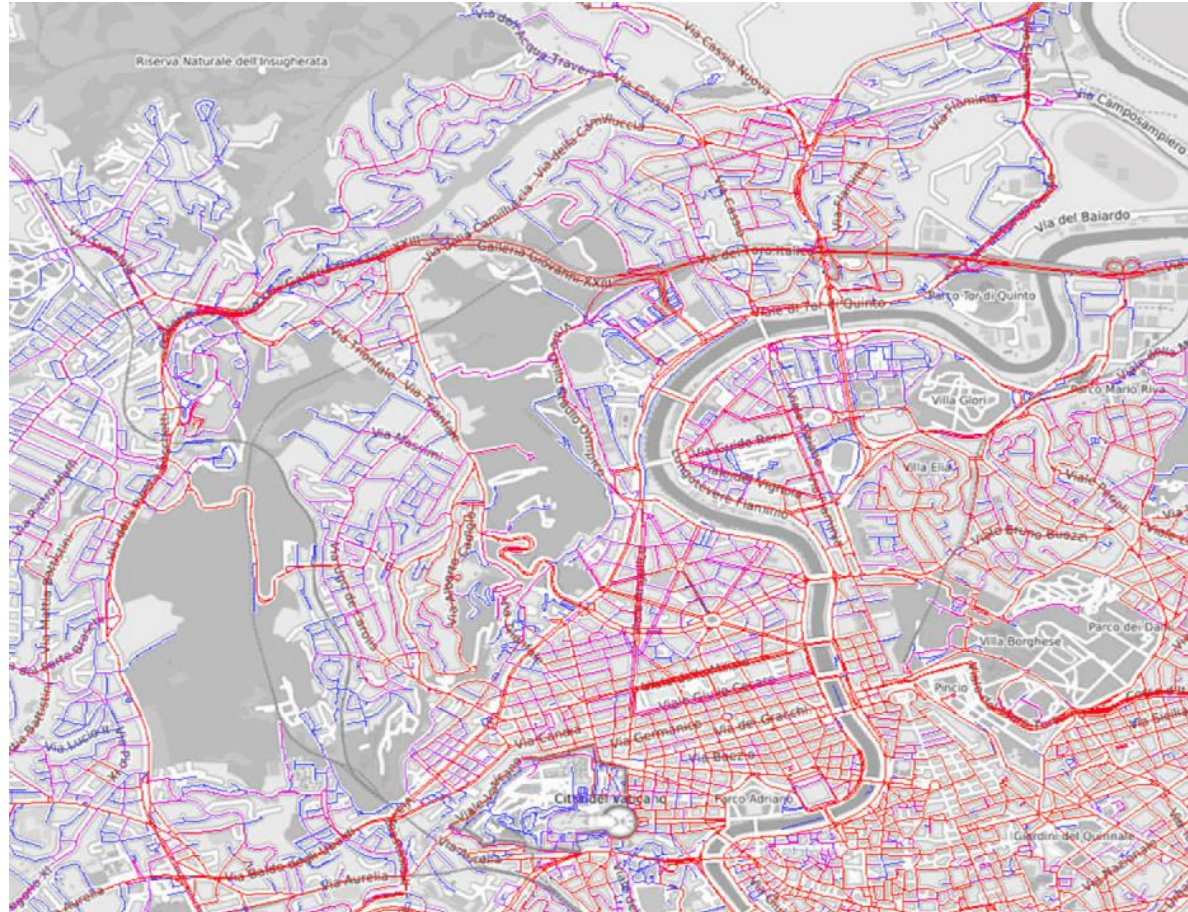
# Area Coverage



Number of sensings in the considered period



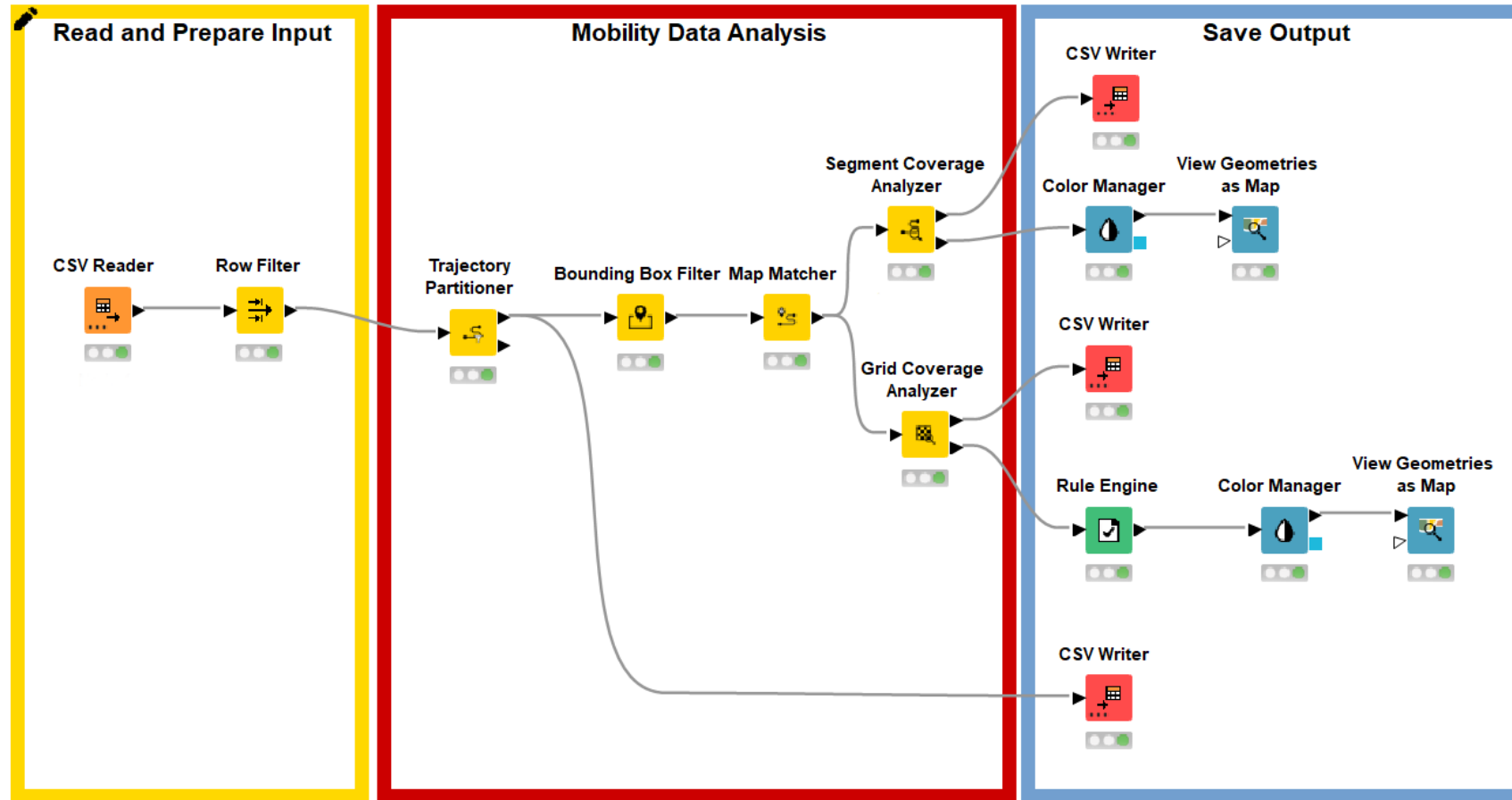
# Segment Coverage



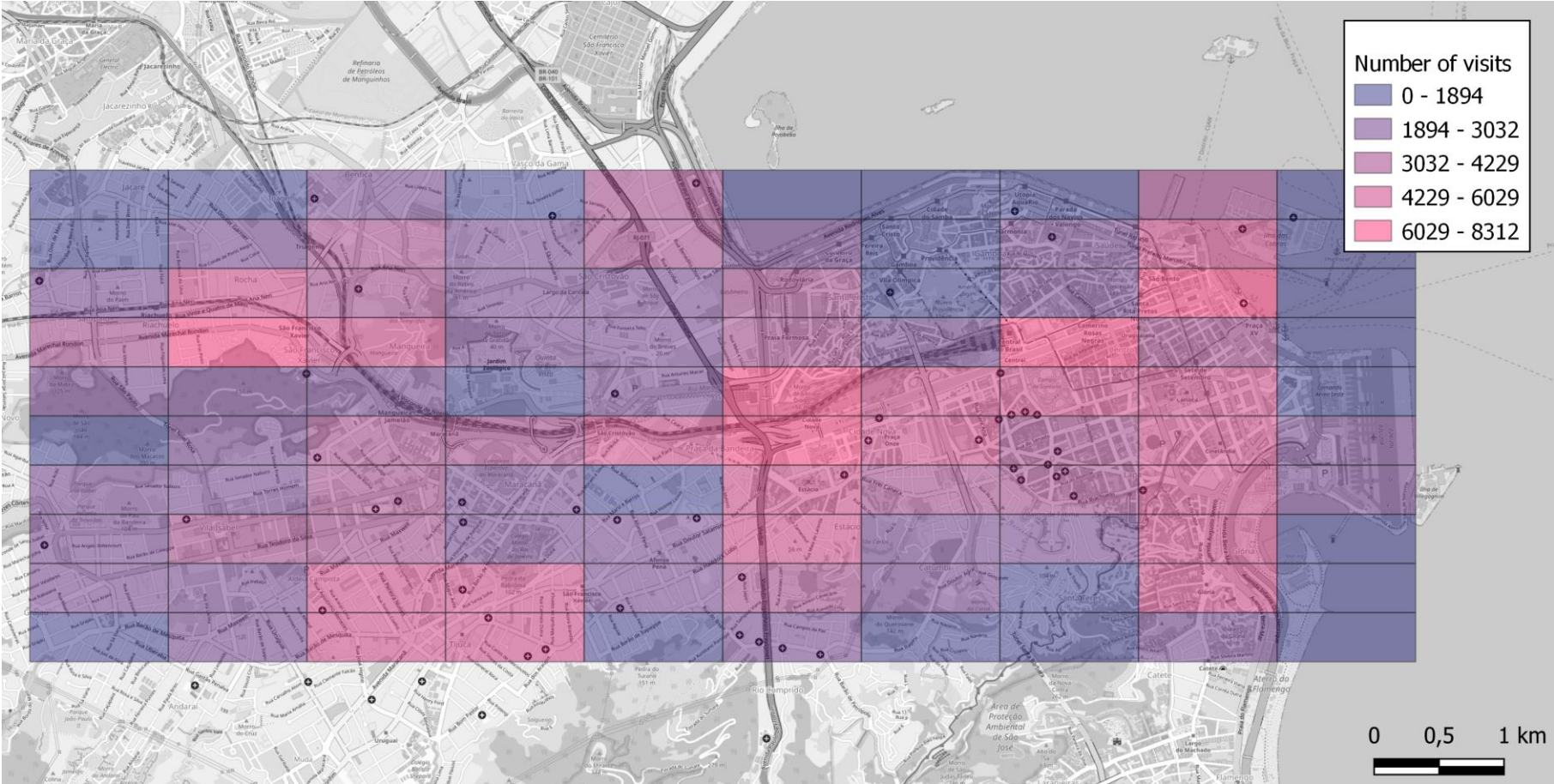
Number of visits in the considered period:

- Less than 5
- Between 6 and 10
- Between 11 and 25
- Between 26 and 50
- Between 51 and 100
- More than 100

# 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