# Data Science
# Introduction to Orange3

# Training Series

# Week 2

**26 Januari 2024**

Orange3
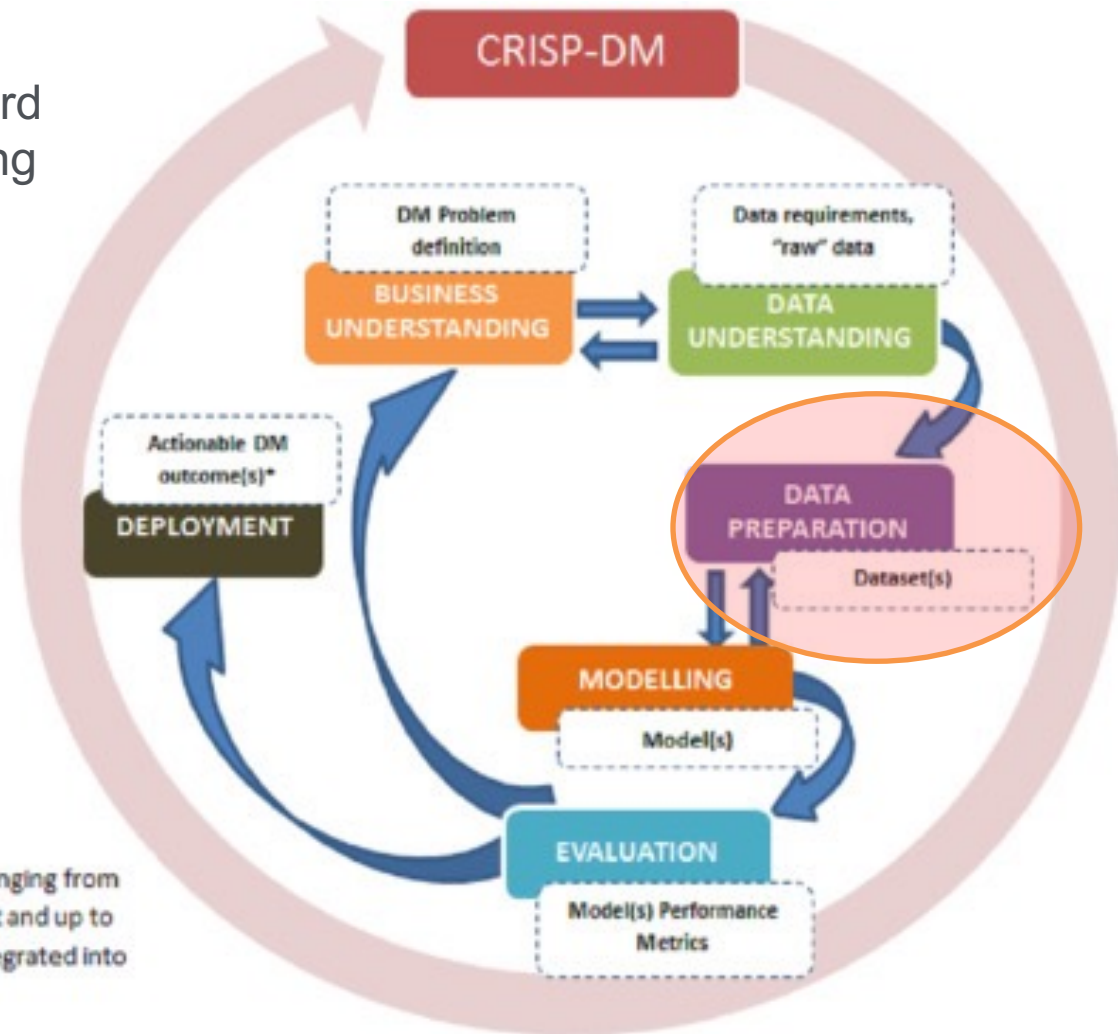Prepared by I Made Murwantara, Universitas Pelita Harapan

# Overview

1. Data Preparation
2. Exploratory Data Analysis (EDA)
3. Visualization
4. Case Study

Add-on:
- Geo
- Time Series

# Data Preparation

Cross Industry Standard
Process for Data Mining

# Data Preparation

- Data collection
- Data Discovery & Profiling
- Data Formatting
- Data Combining
- Validate & Visualize

# Data Collection

# Data Discovery

Activity that include cataloging, classifying & evaluation



Data Customer Churn

# Data Discovery

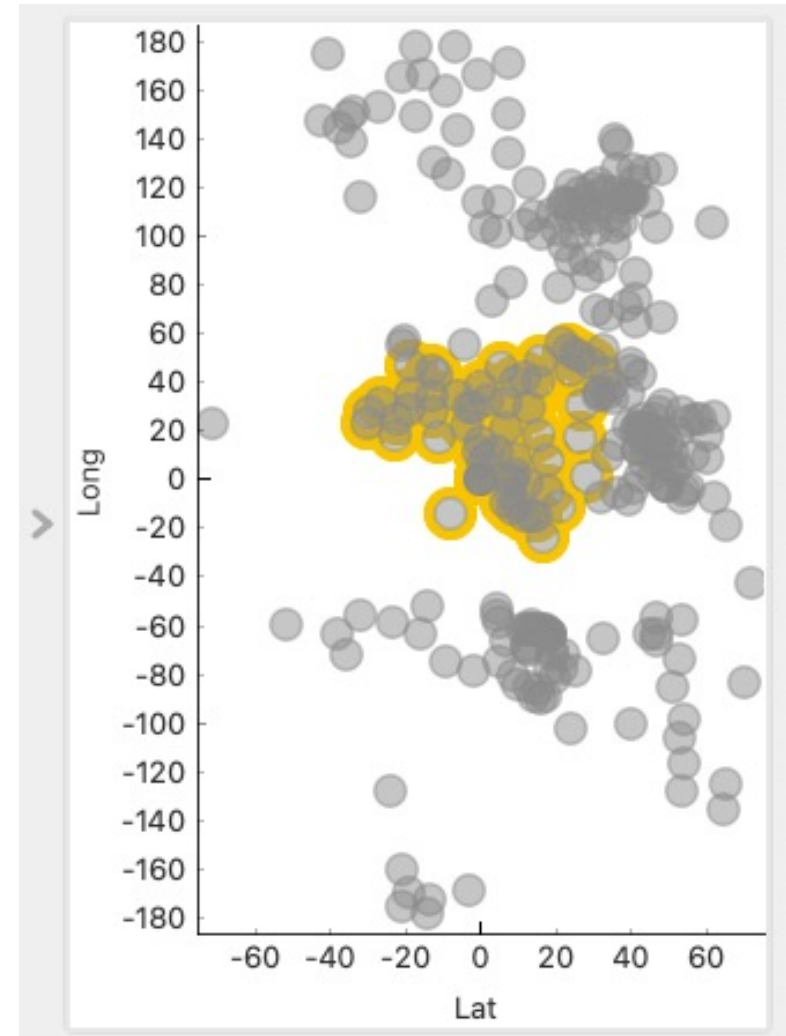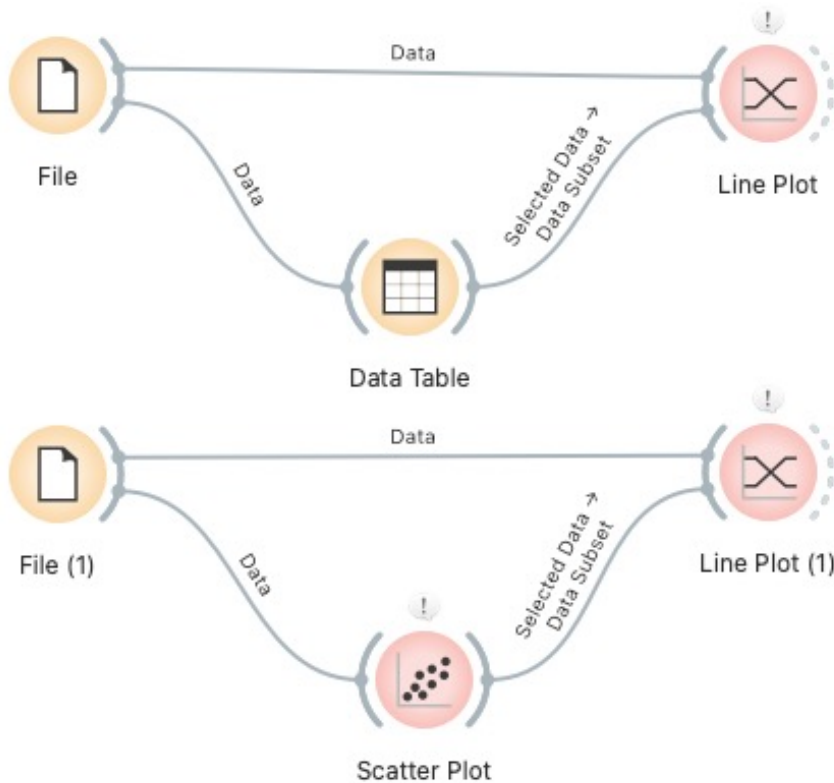https://orangedatamining.com/blog/data-mining-covid-19-epidemics-part-1/
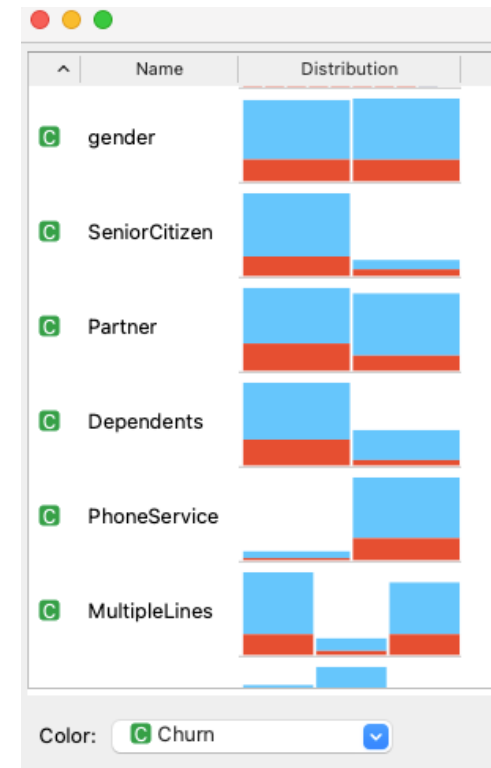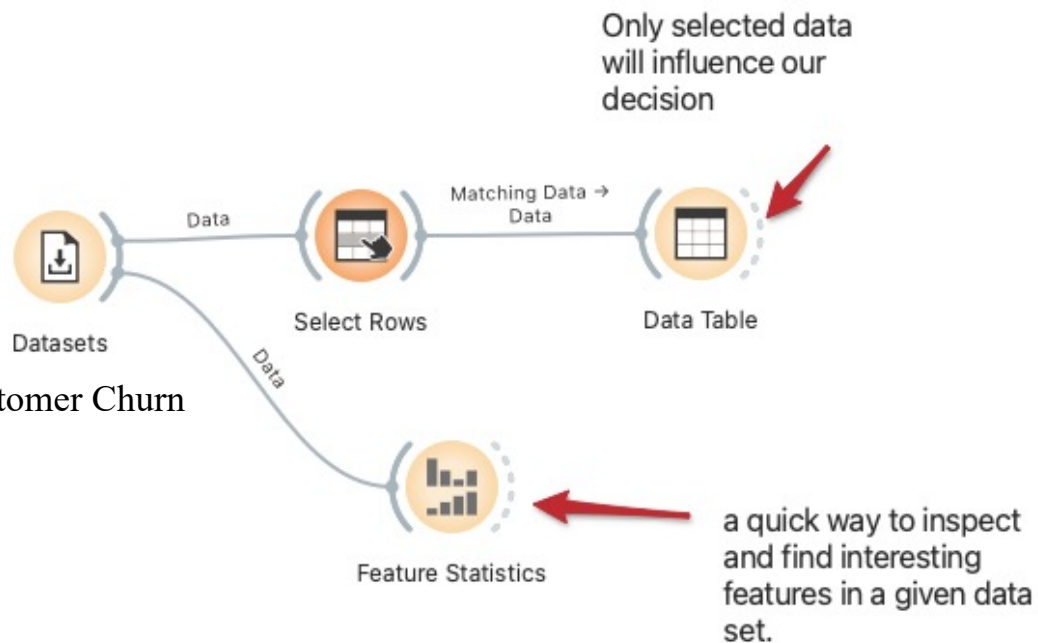
Case on Covid-19     Github Dataset
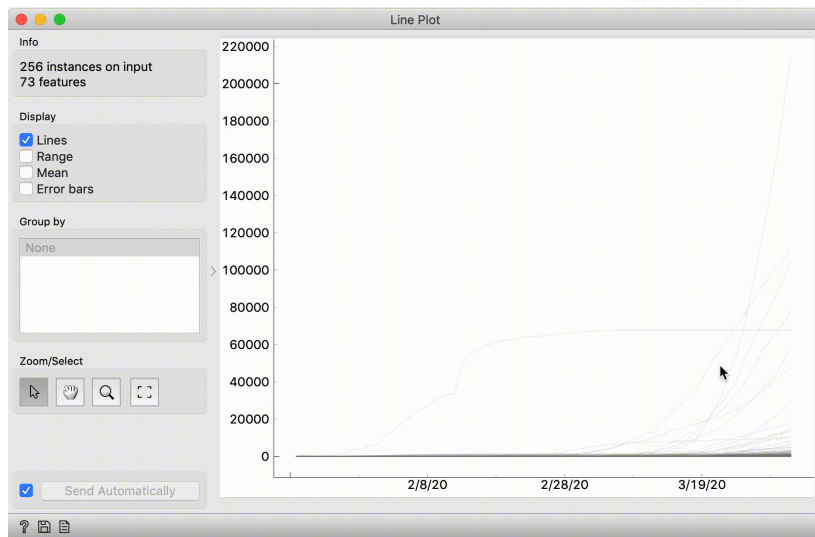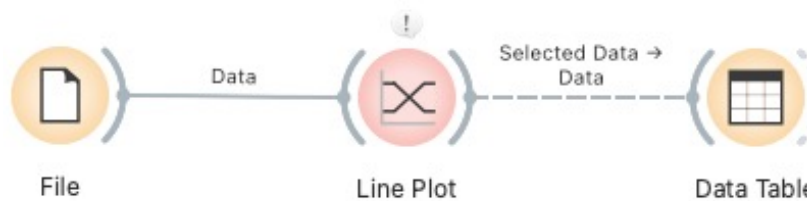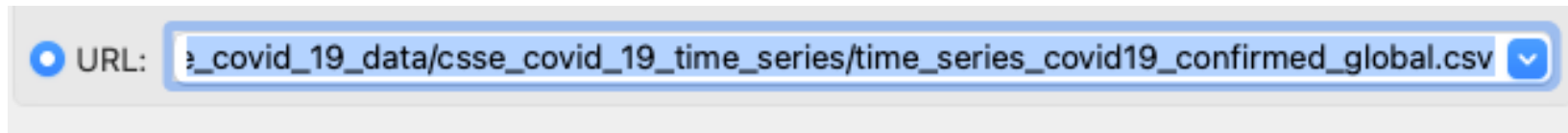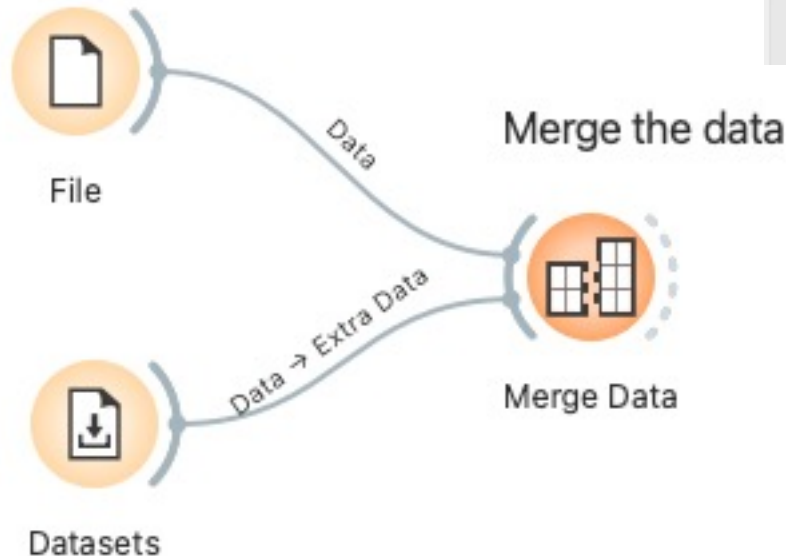


Interactive Result

# Exploratory Data Analysis

Data Selection based on table & scatter plot

# Exploratory Data Analysis

## Merge Table

Merging
- ⦿ Append columns from Extra data
- ○ Find matching pairs of rows
- ○ Concatenate tables

The first table may contain, for instanc
and the second would be a list of cities
Columns with coordinates would then

Row matching

| S Country/Region | ⌄ | matches | S Country | ⌄ |

Merge the data

File

Data

Data → Extra Data

Datasets

Merge Data

| hdi-edu | | Show data sets in | English ⌄ |

| | Title | ^ | Size | Instances | Variables | Ta |
|---|---|---|---|---|---|---|
| ● | HDI-edu | | 54.2 KB | 188 | 53 | |

# Exploratory Data Analysis

## Compute specific data

# Exploratory Data Analysis

## Example using python code
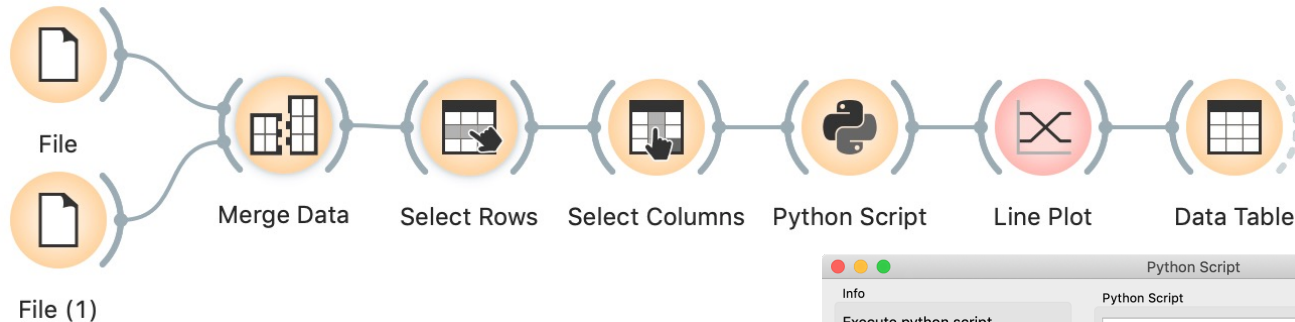


```python
import numpy as np
import Orange

normalized = in_data.X / in_data[:, "Total Population (millions) 2015"].metas

out_data = Orange.data.Table.from_numpy(
    in_data.domain,
    normalized, in_data.Y, in_data.metas)
```

File · Merge Data · Select Rows · Select Columns · Python Script · Line Plot · Data Table

File (1)

Python Script
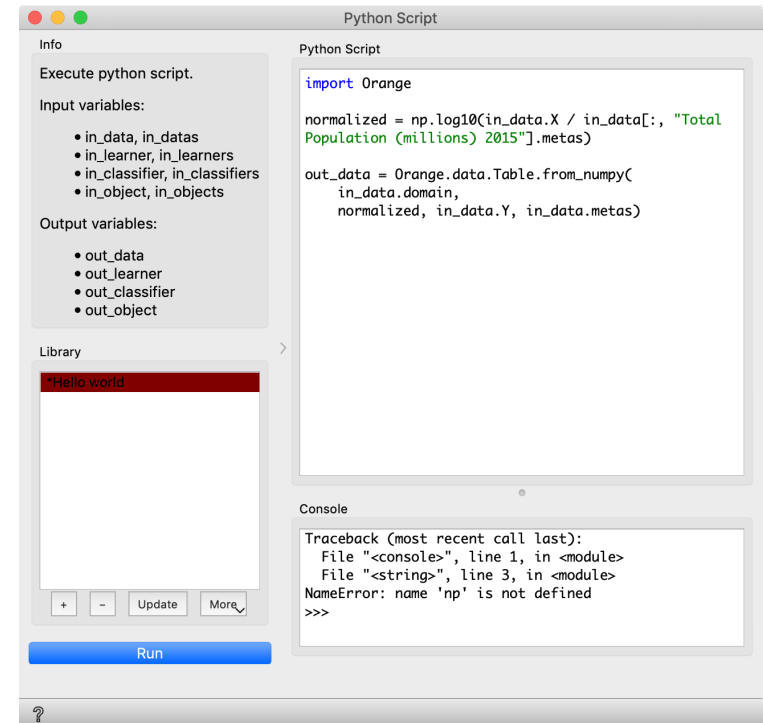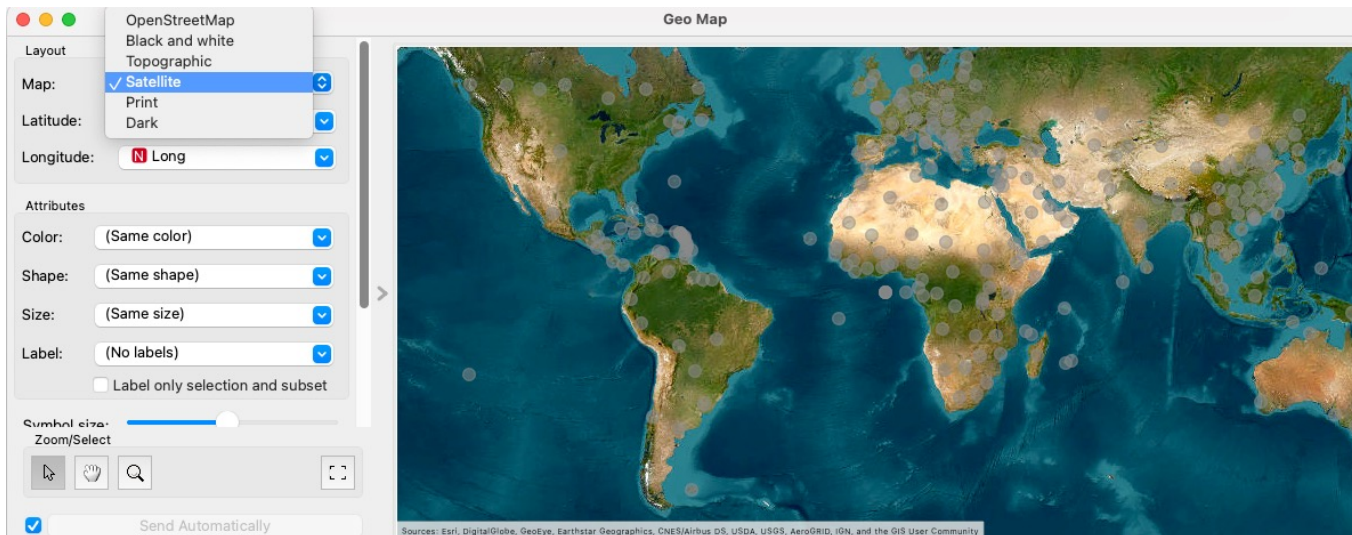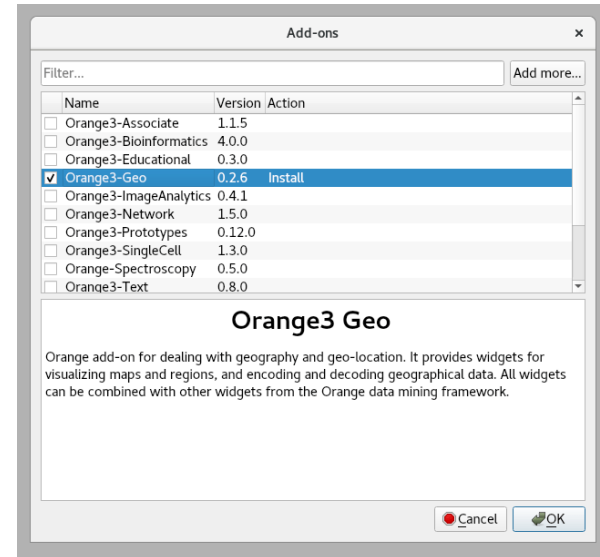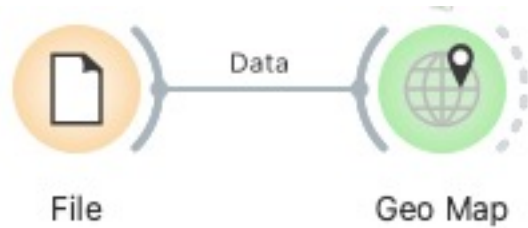
Info

Execute python script.

Input variables:
- in_data, in_datas
- in_learner, in_learners
- in_classifier, in_classifiers
- in_object, in_objects

Output variables:
- out_data
- out_learner
- out_classifier
- out_object

Library

*Hello world

Python Script

```python
import Orange

normalized = np.log10(in_data.X / in_data[:, "Total Population (millions) 2015"].metas)

out_data = Orange.data.Table.from_numpy(
    in_data.domain,
    normalized, in_data.Y, in_data.metas)
```
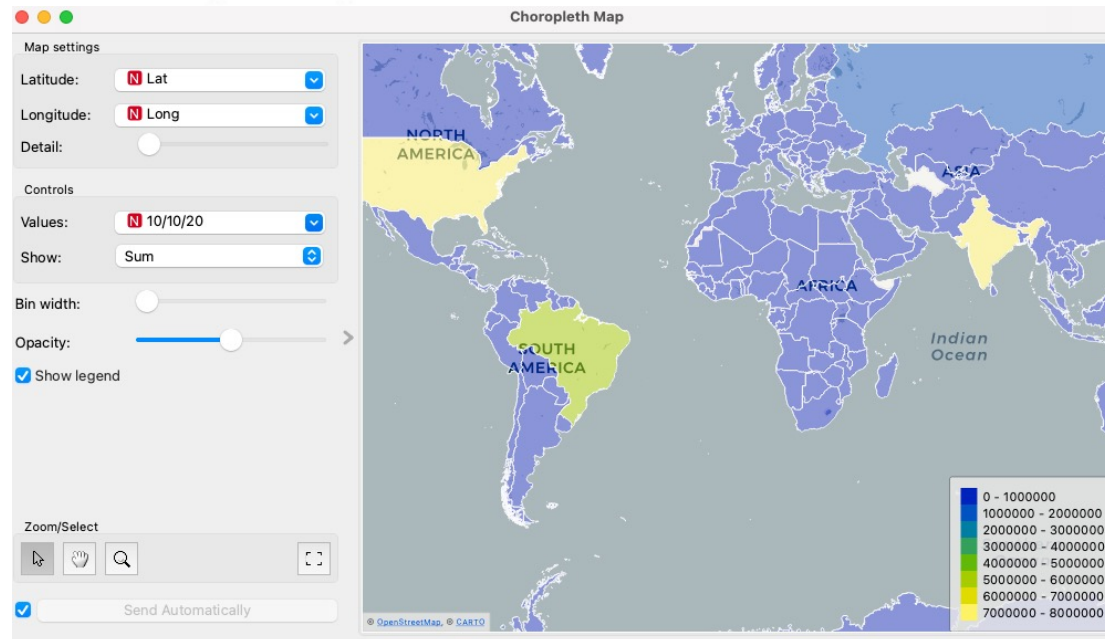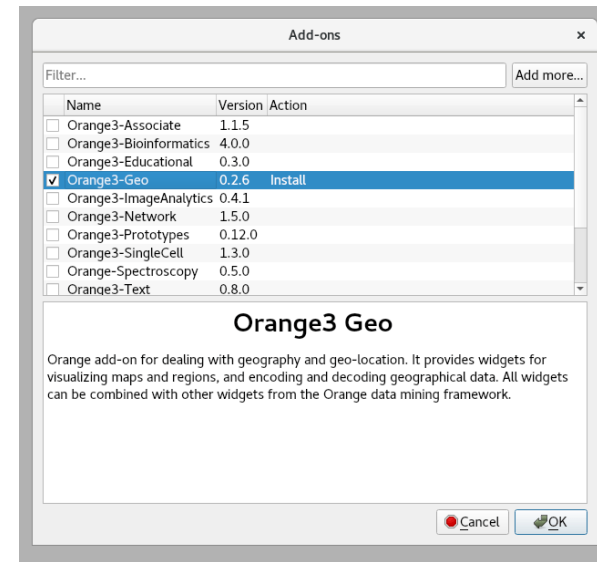
Console

```
Traceback (most recent call last):
  File "<console>", line 1, in <module>
  File "<string>", line 3, in <module>
NameError: name 'np' is not defined
>>>
```

+  −  Update  More

Run

# Visualization

## Geo map

https://orangedatamining.com/blog/data-mining-covid-19-epidemics-part-2/

# Visualization

## Choropleth map

https://orangedatamining.com/blog/data-mining-covid-19-epidemics-part-2/

# Visualization

https://orangedatamining.com/blog/data-mining-covid-19-epidemics-part-2/

# Visualization

## Geo Animation – Step 2

# Visualization – Case Study

## Geo Animation

### Datasets

| | | | | | |
|---|---|---|---|---|---|
| Titanic | | 44.1 KB | 2201 | 4 categorical | |
| ● Traffic accidents - events | | 4.3 MB | 17931 | 18 | geo, timeseries |

Your Task:
1. Animate Traffic Accidents by dates
2. Remove Raw with empty Latitude & Longitude

# Visualization – Case Study - Solution

## Geo Animation

### Datasets

# Any Question ?