



**UPH**  
UNIVERSITAS PELITA HARAPAN

**Data Science**

# **Introduction to Orange3**



## **Training Series**

### **Week 2**

**26 Januari 2024**

Orange3

Prepared by I Made Murwantara, Universitas Pelita Harapan

©IMM 2024



# Overview

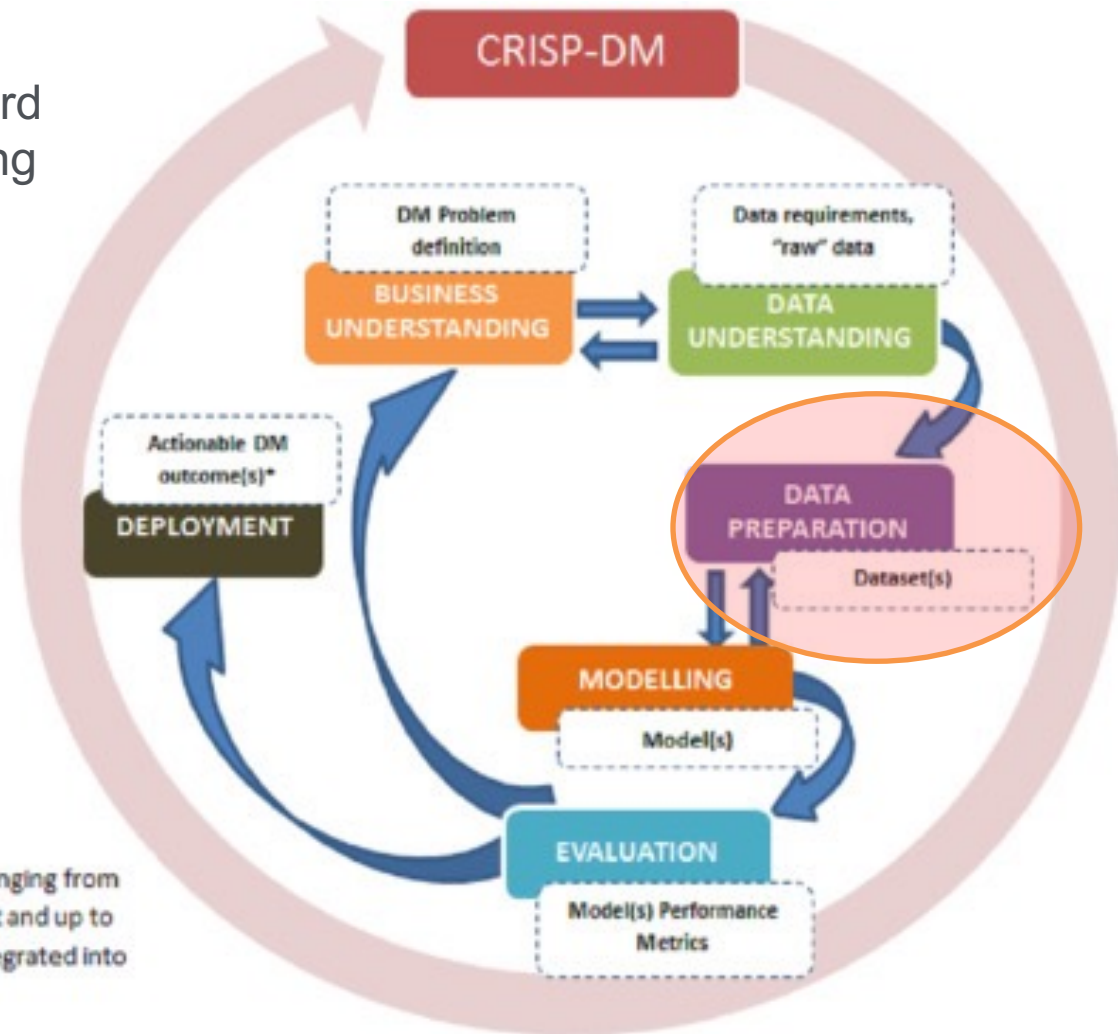
1. Data Preparation
2. Exploratory Data Analysis (EDA)
3. Visualization
4. Case Study

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

Source

☒ File:  ... Reload

☐ URL:



		1		2		3		4		5		6		7		8		9		10		11	
1		Attrition		Age		BusinessTravel		DailyRate		Department		DistanceFro...		Education		EducationField		Environment...		Gender		HourlyRate	Jc
2	Yes			41		Travel_Rarely		1102		Sales		1	2		Life Sciences		2		Female		94	3	
3	No			49		Travel_Frequ...		279		Research & ...		8	1		Life Sciences		3		Male		61	2	
4	Yes			37		Travel_Rarely		1373		Research & ...		2	2		Other		4		Male		92	2	
5	No			33		Travel_Frequ...		1392		Research & ...		3	4		Life Sciences		4		Female		56	3	
6	No			27		Travel_Rarely		591		Research & ...		2	1		Medical		1		Male		40	3	

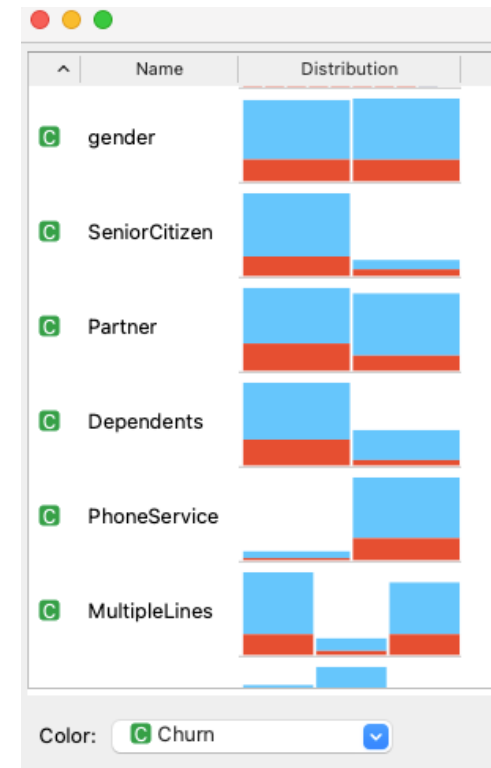
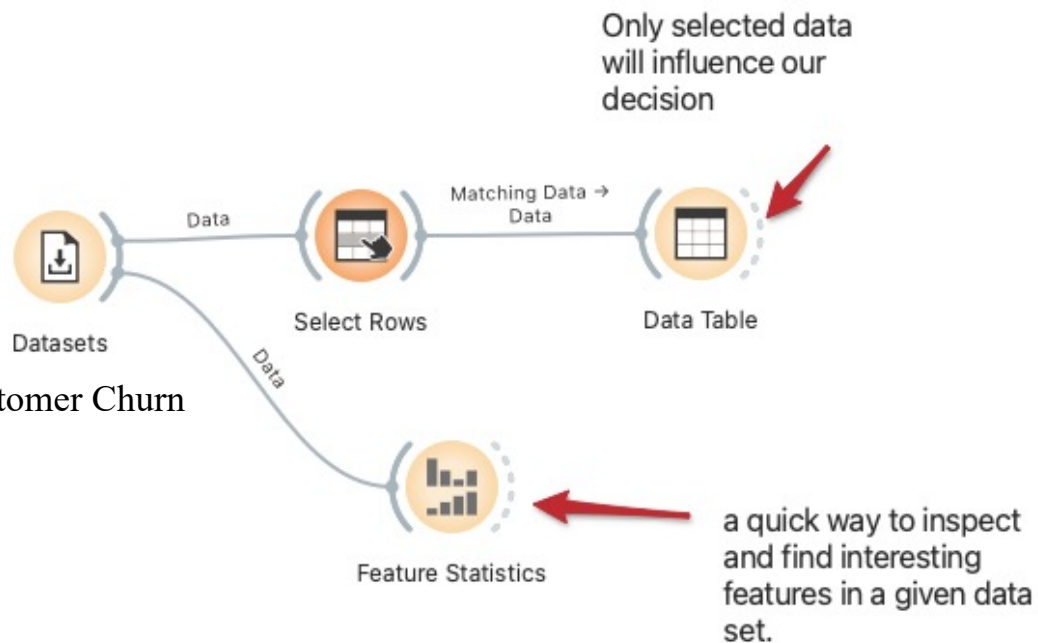
	Title	Size	Instances	Variables	Target	Tags
	Slovenian National Assembly	17.8 KB	84	18		image analytics, ...
	Smoking effect on B lymphocytes	1.8 MB	79	3000	categorical	genomics
	TKI resistance	1.2 MB	280	467	categorical	spectral
	Telecom customer churn	705.1 KB	7043	20	categorical	churn, telecom
	Titanic	44.1 KB	2201	4	categorical	

# Data Discovery

Activity that include cataloging, classifying & evaluation

Conditions

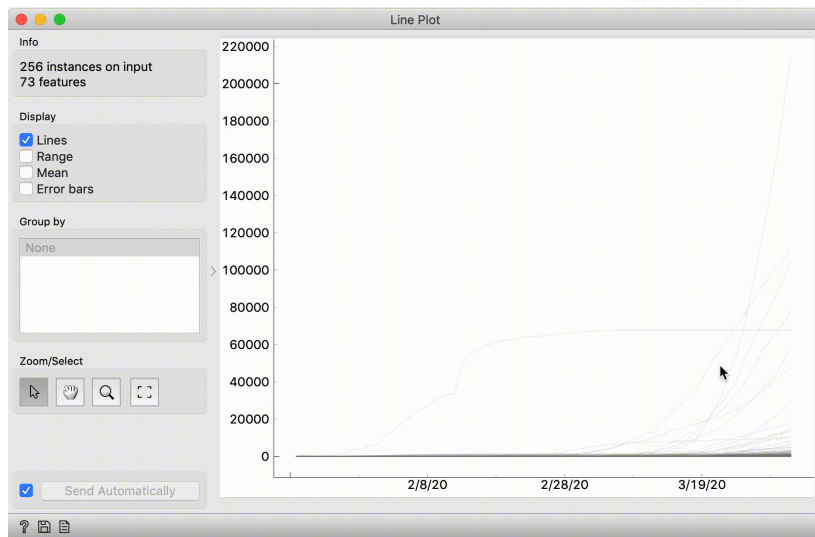
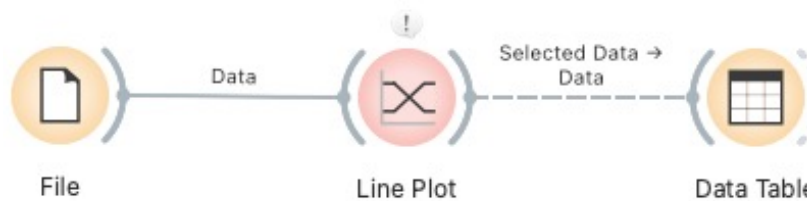
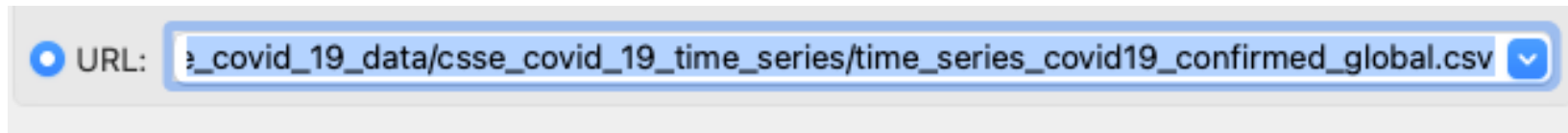
PaymentMethod is Bank transfer (automatic) x



# Data Discovery

<https://orangedatamining.com/blog/data-mining-covid-19-epidemics-part-1/>

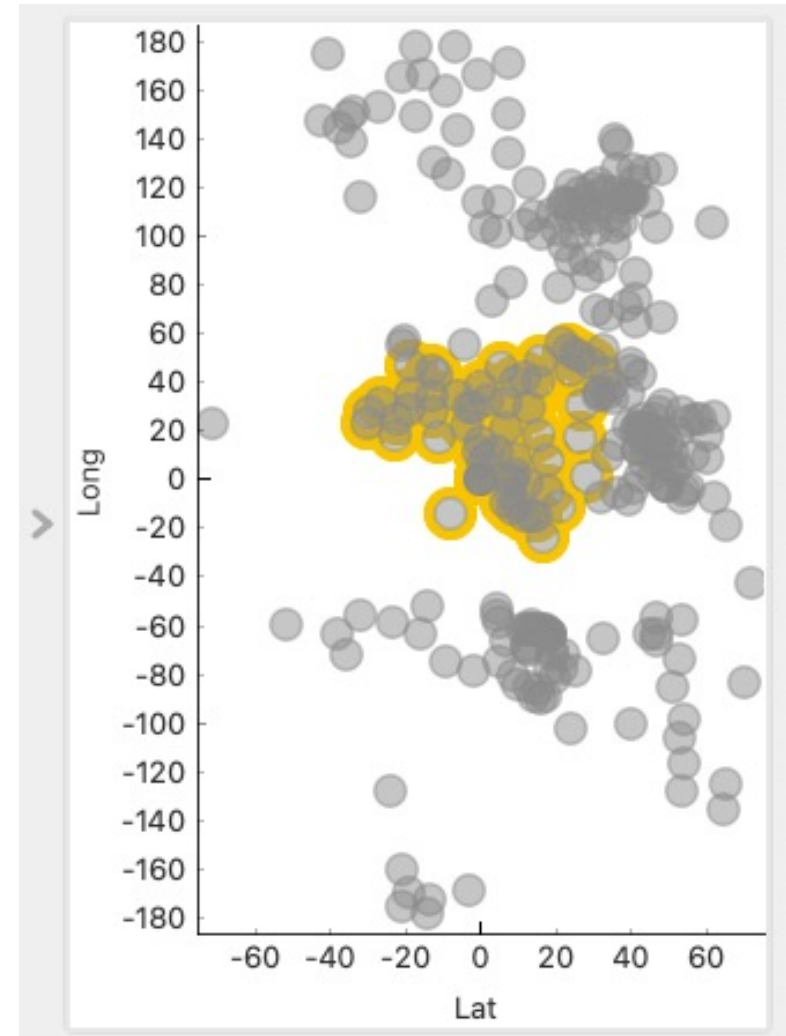
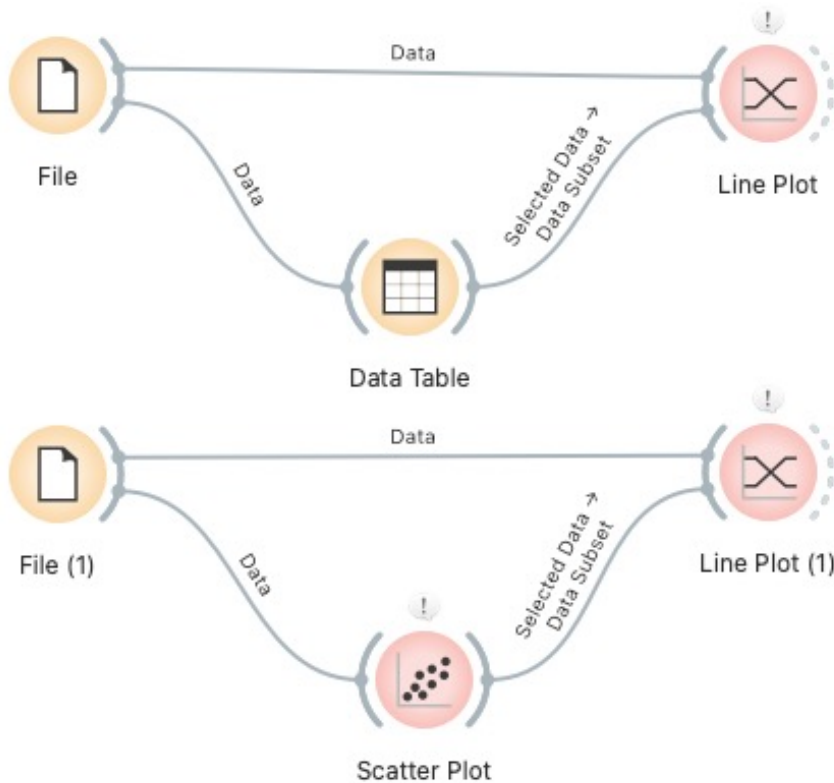
Case on Covid-19 [Github Dataset](#)



Interactive Result

# Exploratory Data Analysis

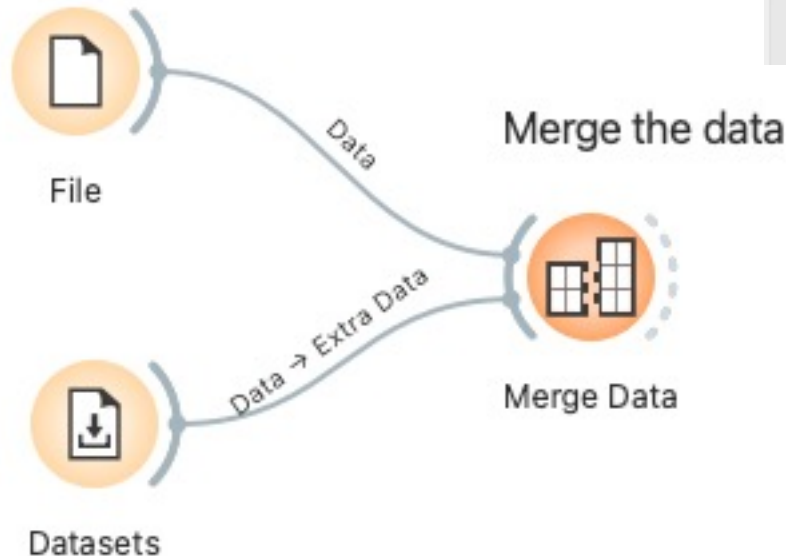
Data Selection based on table & scatter plot





# Exploratory Data Analysis

## Merge Table



Merging

- ☒ Append columns from Extra data
- ☐ Find matching pairs of rows
- ☐ Concatenate tables

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

Row matching

☒ Country/Region matches ☒ Country

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

Conditions

☒ Population Median age (years) 2015 is greater than 0

File Merge the data Merge Data Select Rows Formula Select Columns Data Table

Datasets

Variable Definitions

New ☒ Total Case \_11\_29\_21/Population\_Urban\_2015\_\_

Remove ☐ Meta attribute Select Feature Select Function

☒ Total Case := \_11\_29\_21/Population\_Urban\_2015\_\_

Ignored (1197)

Filter

- ☒ Life expectancy at age 65 (years) 20...
- ☒ Physicians (per 10,000 people) 20...
- ☒ Public health expenditure (% of GD...
- ☒ Employment to population ratio (% ...
- ☒ Labour force participation rate (% a...
- ☒ Employment in agriculture (% of tot...
- ☒ Employment in services (% of total ...
- ☒ Total Unemployment (% of labour f...
- ☒ Unemployment Youth (% ages 15-2...
- ☒ Unemployment Youth not in school ...
- ☒ Vulnerable employment (% of total ...
- ☒ Working poor at PPP\$3.10 a day (%...
- ☒ Old-age pension recipients (% of s...
- ☒ Internet users (% 2010 -2015)
- ☒ Coefficient of human inequality
- ☒ Inequality in life expectancy (% 20...
- ☒ Inequality-adjusted life expectancy ...
- ☒ Child labour (% ages 5-14) 2009-2...
- ☒ Inequality in education (%)
- ☒ Inequality in income (%)
- ☒ Mandatory paid maternity leave (da...

Features (1)

Filter

☒ Total Case

Target

Metas (2)

☒ Country/Region

☒ Province/State

Data Table

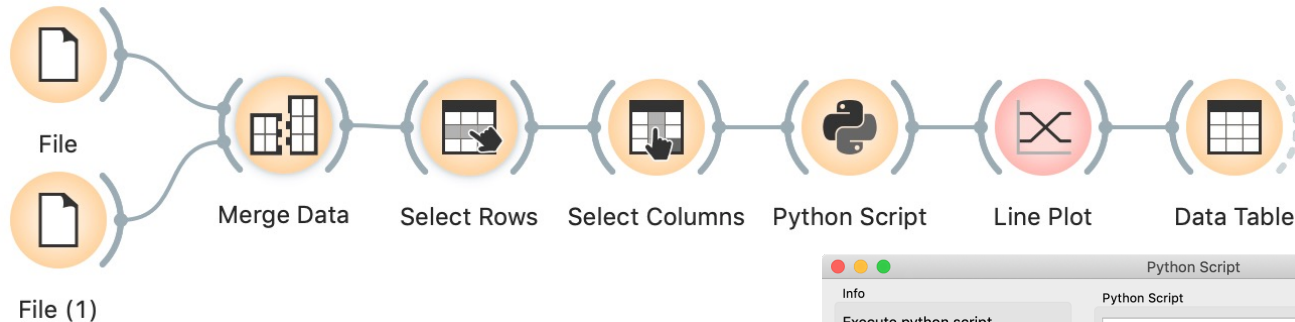
Country/Region	Province/State	Total Case
Canada	Quebec	5469.27
Canada	Repatinated Travellers	0.159924
Canada	Saskatchewan	995.609
Canada	Yukon	16.423
Central Africa	?	92.7
Chad	?	226.976
Chile	?	19680.1
China	Arhui	16.2595
China	Beijing	21.4209
China	Chongqing	10.9712
China	Fujian	23.755
China	Gansu	6.16705
China	Guangdong	59.5324

UPH 2

10

# Exploratory Data Analysis

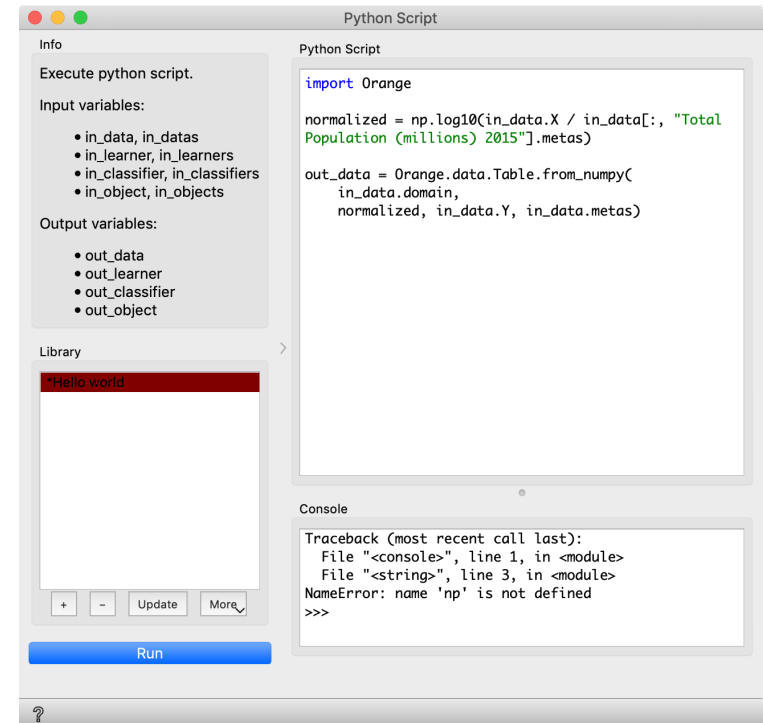
## Example using python code



```
import numpy as np
import Orange

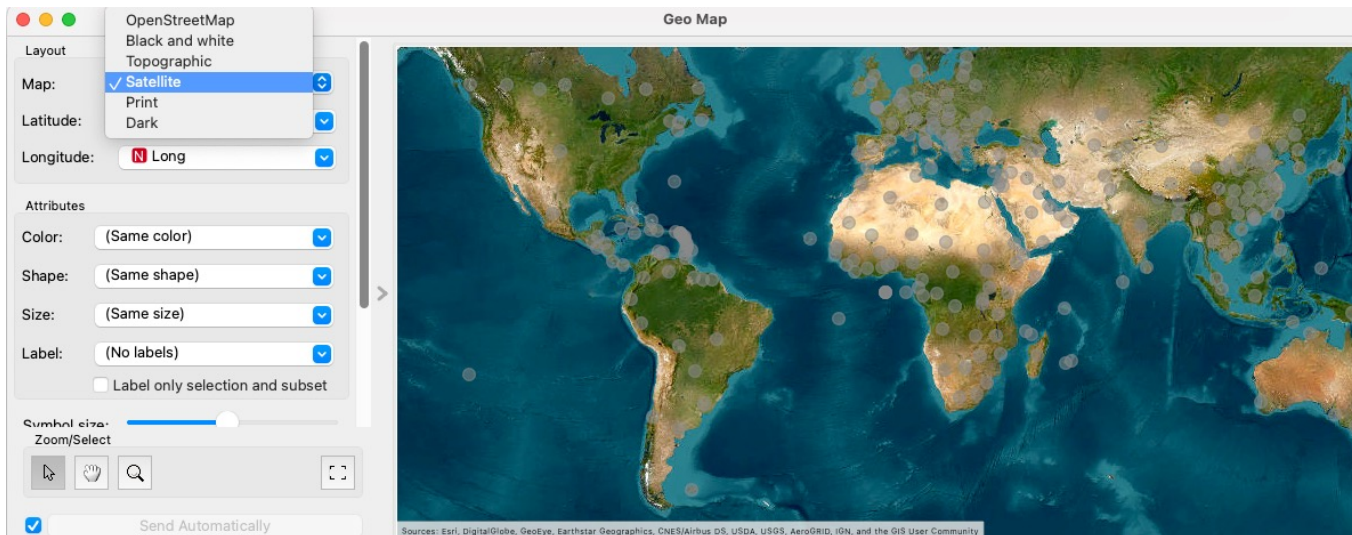
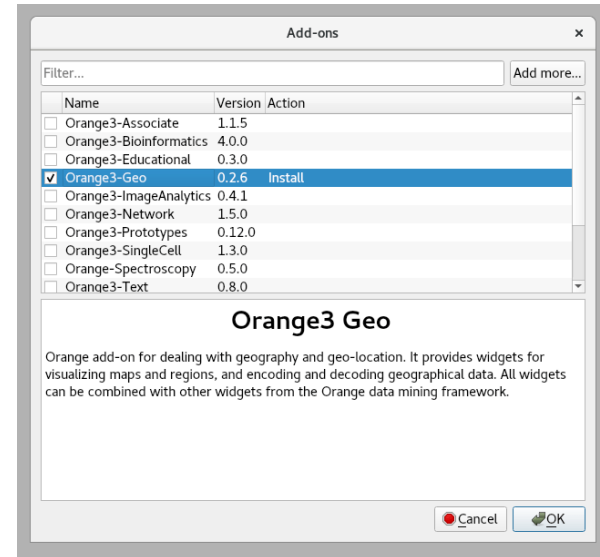
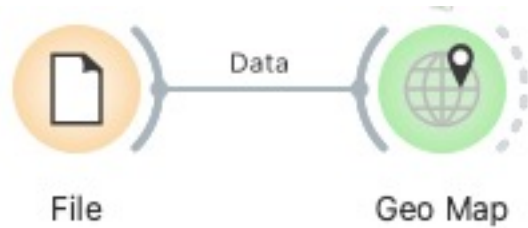
normalized = in_data.X / in_data[:, "Total Population (millions) 2015"].metas

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



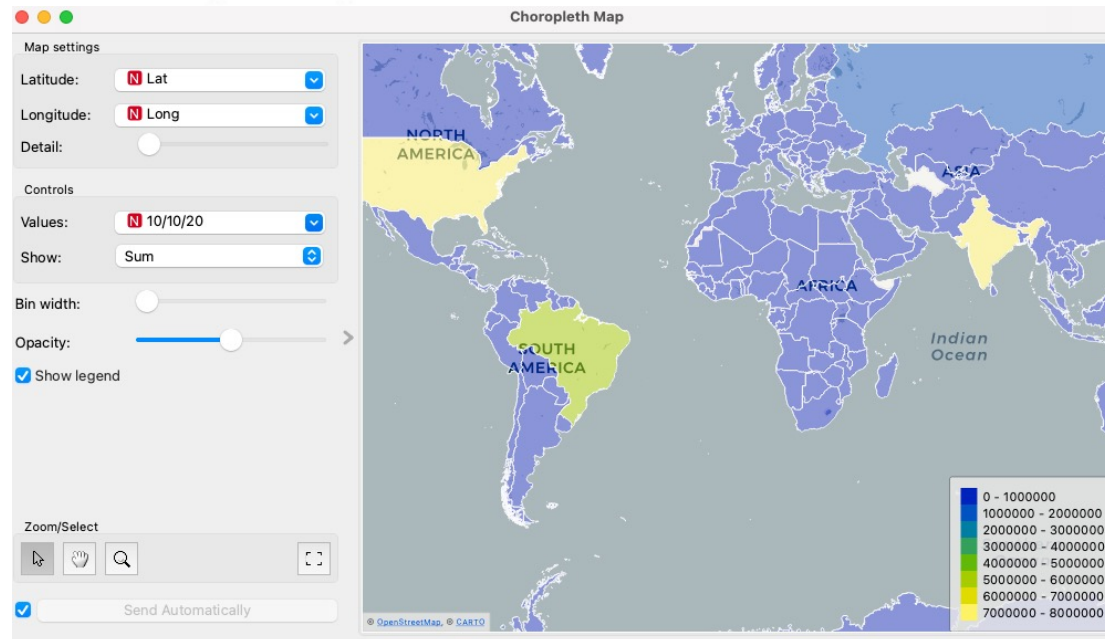
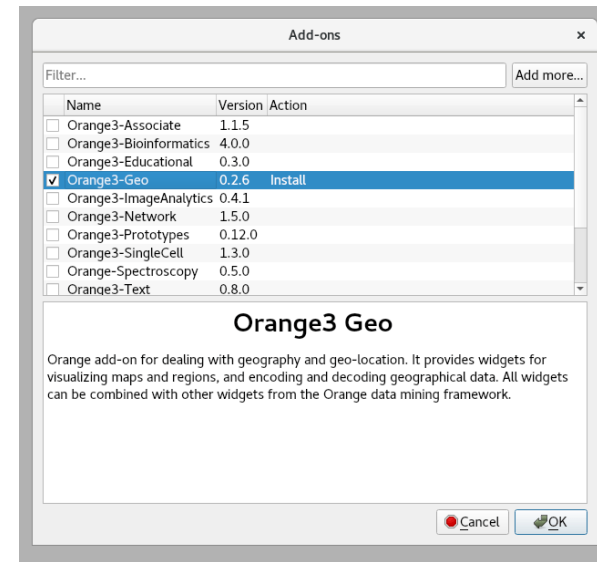
# Visualization

## Geo map



# Visualization

## Choropleth map



# Visualization

## Geo Animation – Step 1



	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20
12	Northern Ter...	Australia	-12.4634	130.846	0	0	0	0
13	Queensland	Australia	-27.4698	153.025	0	0	0	0
14	South Austra...	Australia	-34.9285	138.601	0	0	0	0
15	Tasmania	Australia	-42.8821	147.327	0	0	0	0
16	Victoria	Australia	-37.8136	144.963	0	0	0	0
17	Western Aus...	Australia	-31.9505	115.861	0	0	0	0
18	?	Austria	47.5162	14.5501	0	0	0	0
19	?	Azerbaijan	40.1431	47.5769	0	0	0	0

Variables

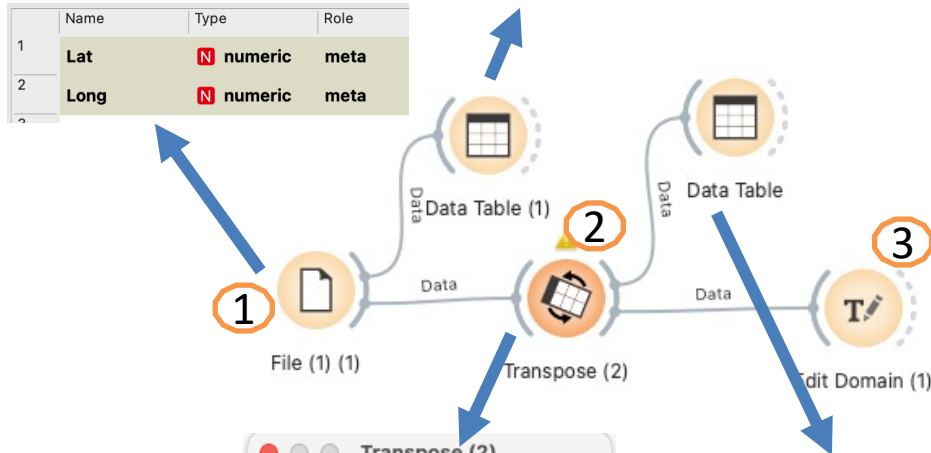
☒ Afghanistan  
☒ Albania  
☒ Algeria  
☒ Andorra  
☒ Angola  
☒ Antigua and Barbuda  
☒ Argentina  
☒ Armenia  
☒ Australia (1)  
☒ Australia (2)  
☒ Australia (3)  
☒ Australia (4)  
☒ Australia (5)  
☒ Australia (6)  
☒ Australia (7)  
☒ Australia (8)  
☒ Austria  
☒ Azerbaijan  
☒ Bahamas  
☒ Bahrain  
☒ Bangladesh

Edit

Name: Afghanistan  
 Type: ☒ Numeric  
 Labels:
 

Key	Value
Country/Reg...	Afghanistan
Lat	33
Long	65

Reset Selected   Reset All   Apply



Transpose (2)

Feature names

☐ Generic  
 Type a prefix ...

☒ From variable:  
 Country/Region

☐ Remove redundant instance

☒ Apply Automatically

Province/State	Australia (1) Northern Territory	Australia (2) New South Wales	Australia (3) Northern Territory	Australia (4) Queensland	Australia (5) South Australia	Australia (6) Tasmania	Australia (7) Victoria
1	-12.4735	-33.8688	-12.4634	-27.4698	-34.9285	-42.8821	-37.8136
2	153.025	151.209	130.846	153.025	138.601	147.327	144.963
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0



# Visualization

## Geo Animation – Step 2





# Visualization – Case Study

## Geo Animation

### Datasets

name	size	rows	columns	type
● Traffic accidents - events	4.3 MB	17931	18	geo, timeseries

#### Your Task:

1. Animate Traffic Accidents by dates
2. Remove Row with empty Latitude & Longitude



# Visualization – Case Study - Solution



## Geo Animation

### Datasets

name	44.1 KB	2201	4	✓ Categorical
● Traffic accidents - events	4.3 MB	17931	18	geo, timeseries





# Any Question ?