

Lab 4: Creating models with Orange

When COVID-19 started propagating throughout the world, governments didn’t know which measures were the best to take. If governments raised very heavy actions on the population such as closing schools, businesses, public transport and so on, that would cause severe damage to the economy and also the mental health of the population. On the other hand, not applying any action will probably derive in an enormous amount of deaths and overwhelmed hospitals. Therefore to reach a balance, governments needed to take decisions based on different models that simulated the different scenarios.

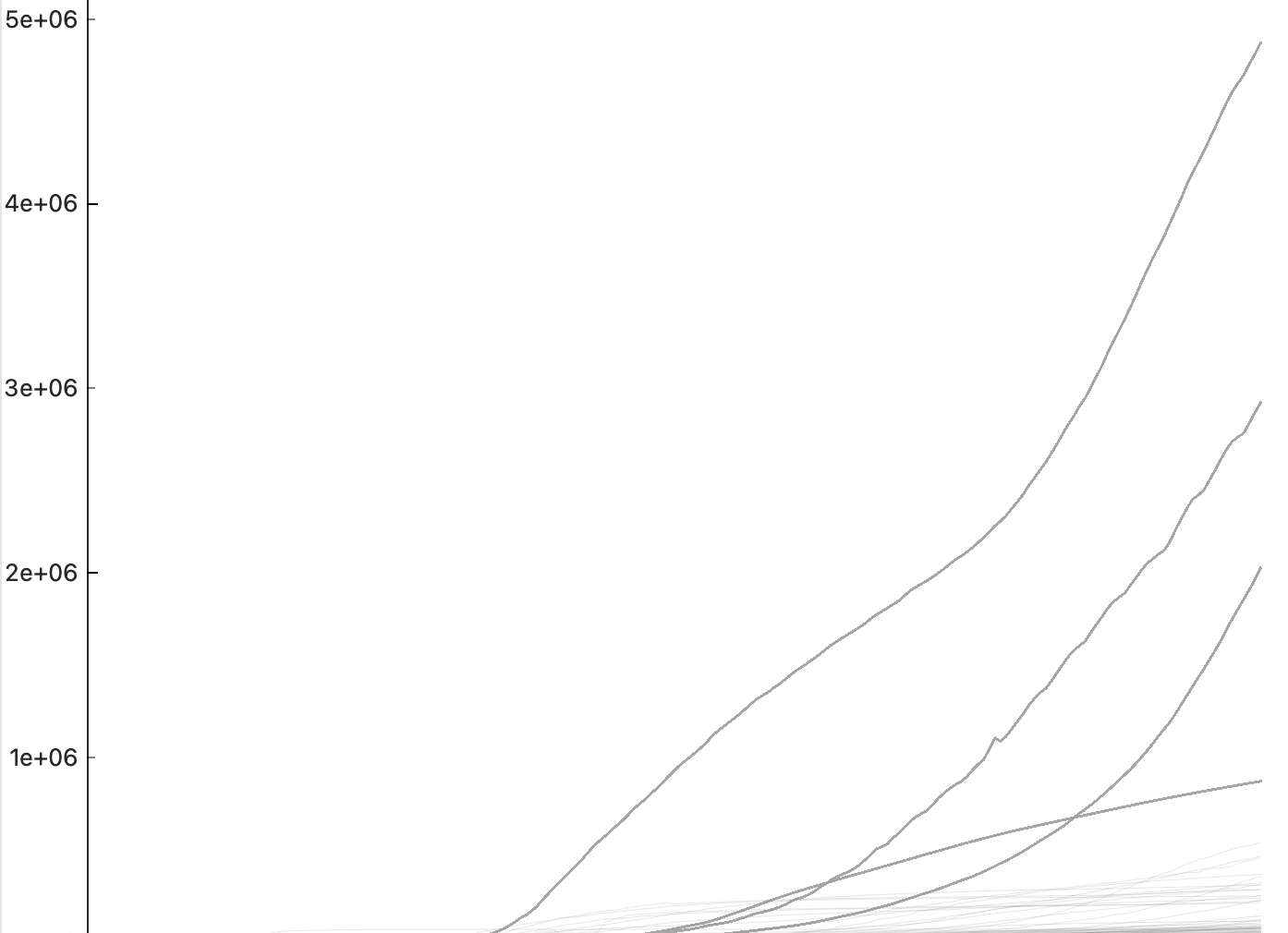
In this assignment, we are going to explore the Covid-19 time series:

This tutorial can be seen in more detail in the following video: [Explore the Covid-19 Data with Orange](https://youtu.be/EfsqjFJC434)

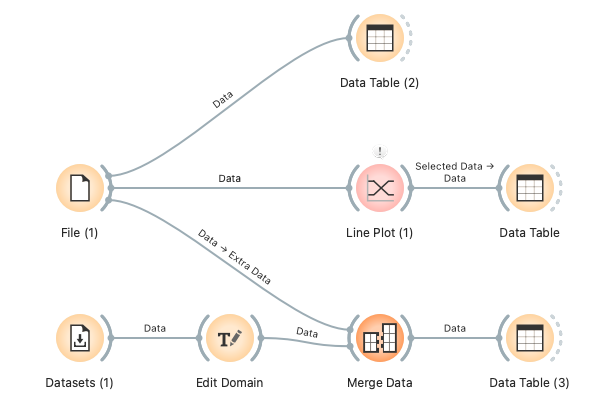
* Step 1: Start a new Orange data mining project, create a widget called “File” and in the URL paste the following URL address:

<https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv>

* Step 2: To visualize the data, create use a “Line plot” widget which will show which countries have the most number of cases. You can use a line to select some of the countries and show them in a data table as follows:



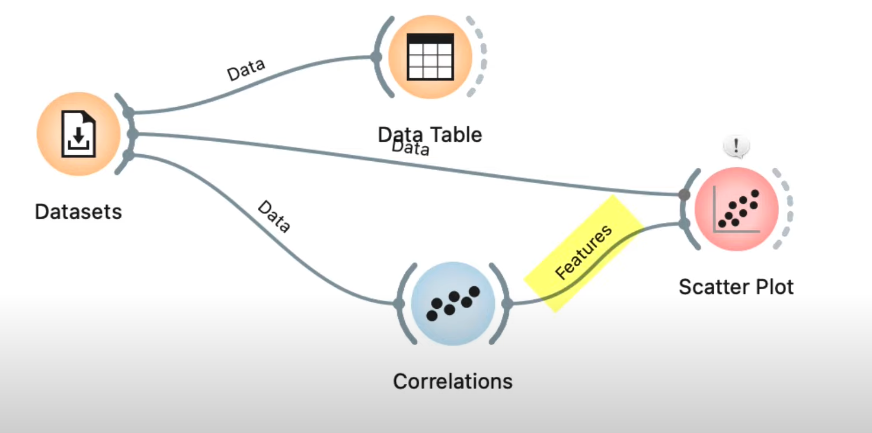
* Step 3: To get better insights from our data, we are going to merge two datasets by the field “country” to create a new one. To do so, we will create a pipeline with two new widgets “Merge Data” and “Edit domain”. You will need to merge the population per country which comes in the following dataset: [Population by Country - 2020 | Kaggle](https://www.kaggle.com/datasets/tanuprabhu/population-by-country-2020)
* Step 4: Merge the table of covid cases with the table of step 3. Note: You can remove the rows that do not have a specific value. When selecting the rows use the option Country “is defined”. You can find a file with a pipeline about COVID in bright space.



* Step 5: To learn how to use correlations, take a look at the following video

[Correlations - Introduction to Data Science](https://www.youtube.com/watch?v=Bd7n9FdY3g0&ab_channel=OrangeDataMining)

The video uses the dataset HDI and uses the widget correlations to find interesting relationships between the variables.



* Step 6: The last part of the assignment consists of representing the different covid points on a map. To do this, follow the steps on the video <https://youtu.be/4SIv0fRScbs>

