



**Categories:** Covid-19 (/Categories/Covid-19) | Visualization (/Categories/Visualization) | Addons (/Categories/Addons) | Geo (/Categories/Geo) | Time (/Categories/Time)

# Data Mining COVID-19 Epidemics: Part 2

By: Robert Cvitkovič, Apr 13, 2020

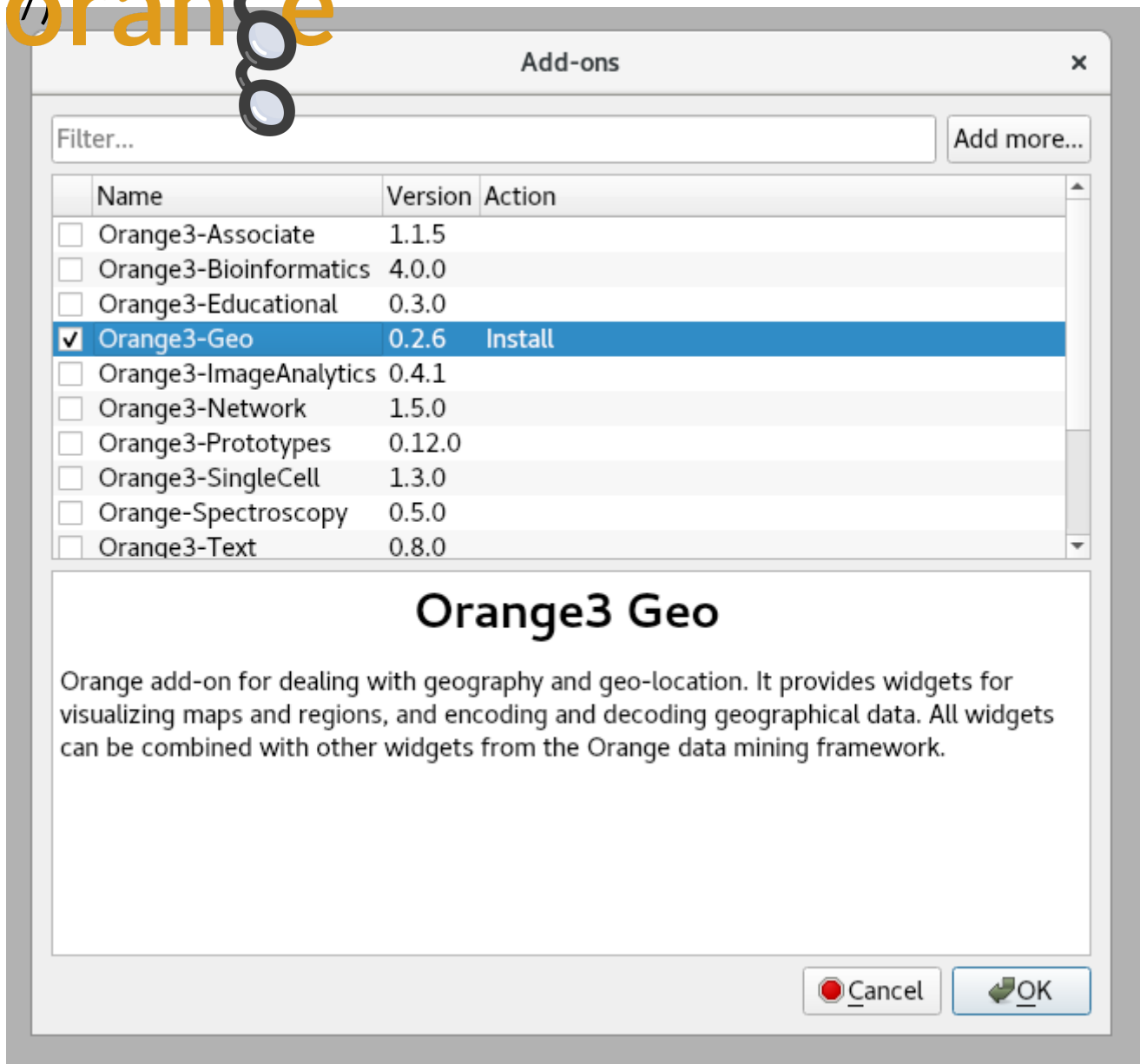
Previously on **Data Mining COVID-19 Epidemics: Part 1** (<https://orange.biolab.si/blog/2020/2020-04-02-covid-19-basic/>) we fired up a COVID-19 epidemics data set and looked at some basic visualizations. If you haven't got your hands dirty with it yet, check that out first.

In this post, we'll try putting the data on a map. We'll also expand on the "data mining is interactive" mantra by creating some animations showing the epidemics spread throughout the world.

## Prerequisite: installing add-ons

Orange already comes with a great selection of widgets that work with all sorts of data. But sometimes we're faced with a particular data set, such as textual data, time-series, or geographical data, which requires widgets with specific functionality – add-ons to the rescue! Add-ons are packages of widgets specially designed to handle specific data types.

We'll be working with the Geo and Timeseries add-ons. To install them, open *Options* in the menu bar and select *Add-ons*. A new window will open, showing a list of officially supported add-ons. Select Orange3-Geo and Orange3-TimeSeries and click *OK*. After installation, restart Orange. Check out the new sections in the widget toolbox; you should see some cool new widgets.

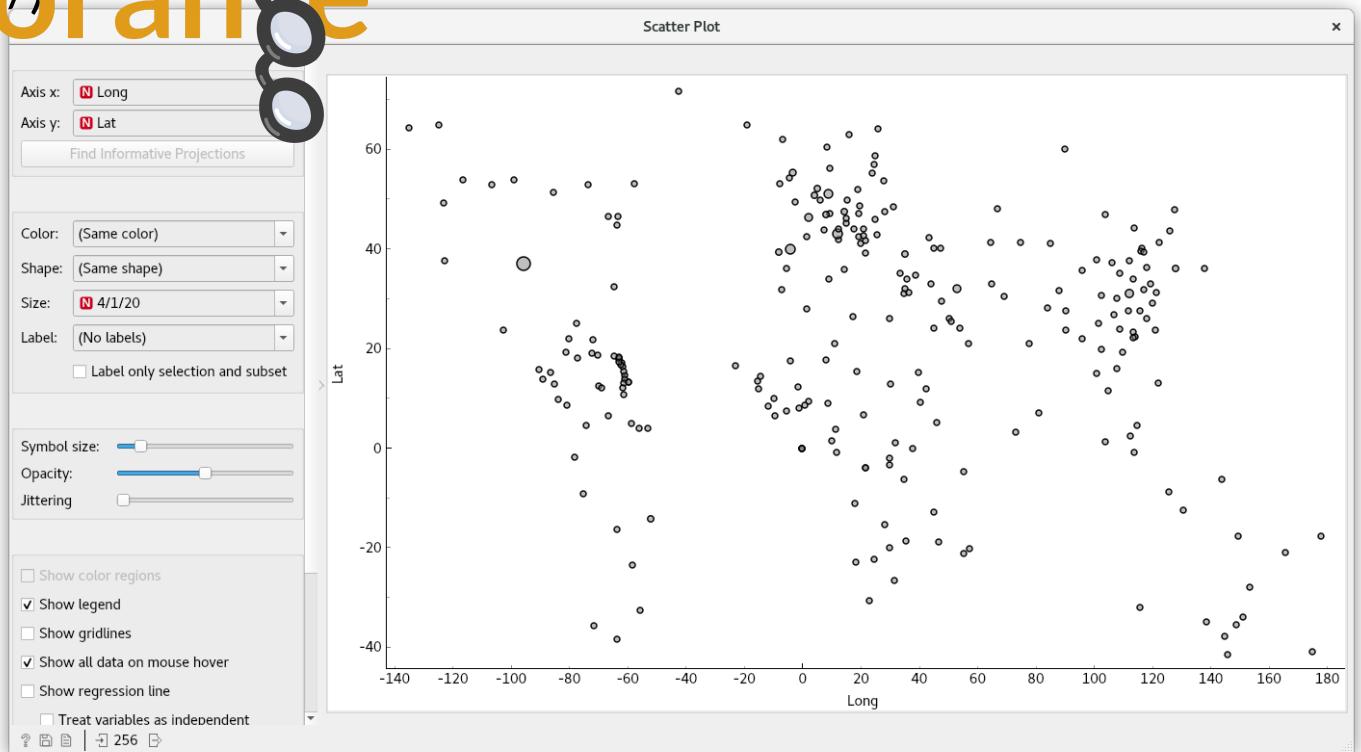


(/blog\_img/2020/2020-04-09-install-addon.png)

**Note:** the two add-ons have been revamped very recently. If you already have them, use the Add-ons dialog to upgrade them to the latest version.

## From Scatter Plot to Geo Map

You should already be familiar with the Scatter Plot widget. In the previous blog post, we used it as an impromptu map by putting latitude and longitude on the axes.



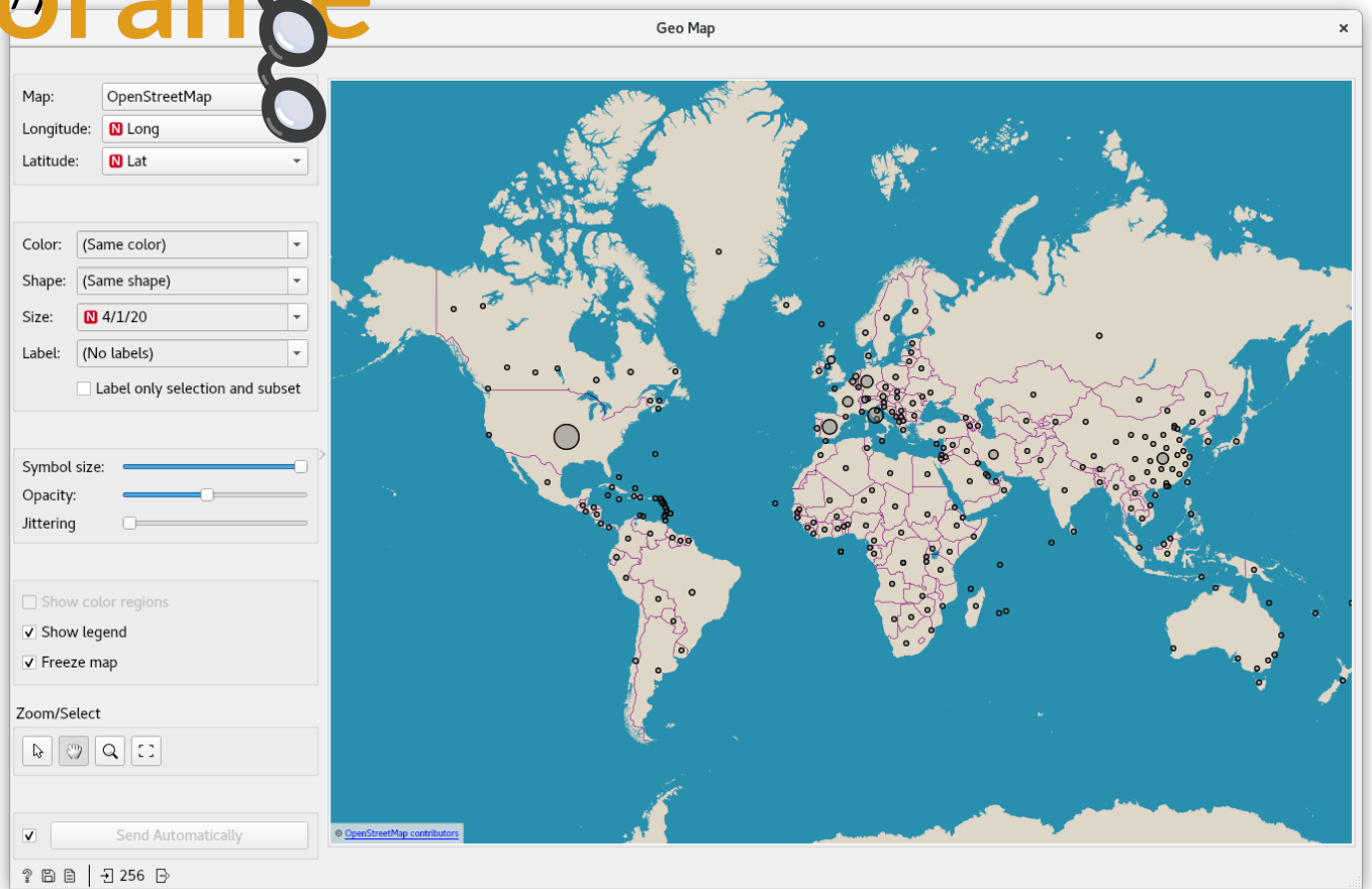
(/blog\_img/2020/2020-04-09-scatter-plot.png)

A better solution is to use Geo Map from the Orange3-Geo add-on. It provides a variety of maps as backgrounds. With them, it's much easier to associate points with locations. To start, first load the **COVID-19 data set** ([https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/time\\_series\\_covid19\\_confirmed\\_global.csv](https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv)) using the File widget, as shown in the previous post (<https://orange.biolab.si/blog/2020/2020-04-02-covid-19-basic/>). Then, add and connect a Geo Map widget.



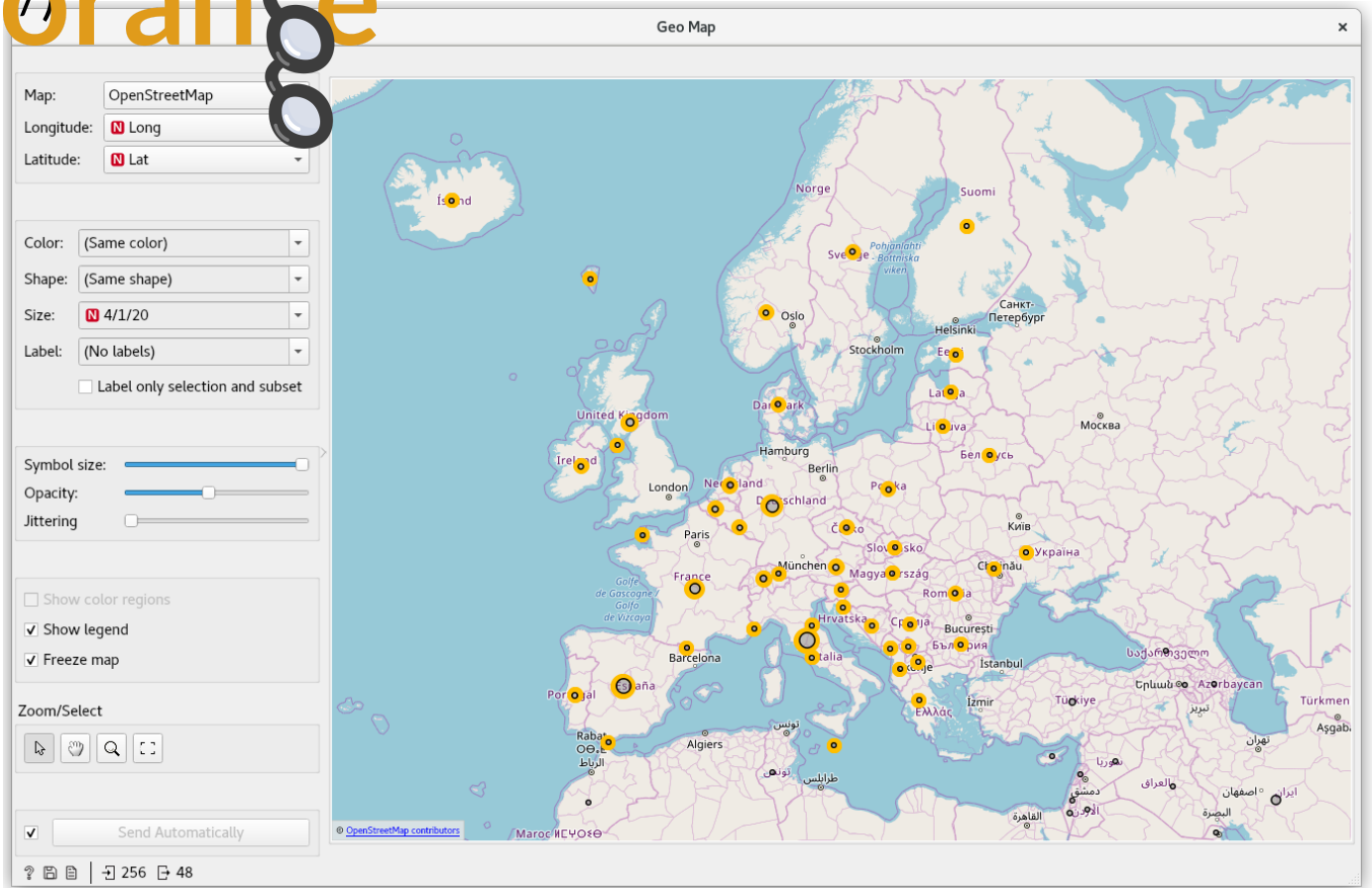
(/blog\_img/2020/2020-04-09-workflow-01.png)

Latitude and longitude are automatically selected. Below, you can define similar point properties, as in the scatter plot. The size of a point can intuitively represent the number of cases, so let's select the latest date in the *Size* dropdown. If you're feeling adventurous, try changing the map type to satellite or topological.



(/blog\_img/2020/2020-04-09-full-geomap.png)

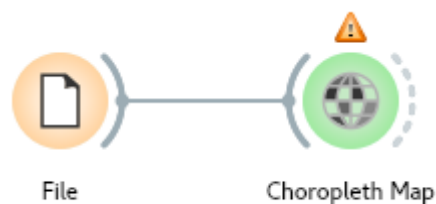
Just like in Scatter Plot, you can scroll to zoom in and out, and click/drag to select data points. As opposed to using Scatter Plot, showing the points over a map of Europe makes it easier to select the countries we want to inspect.



(/blog\_img/2020/2020-04-09-eu-geomap.png)

The chloro... choropl... choropleth... map with areas colored in

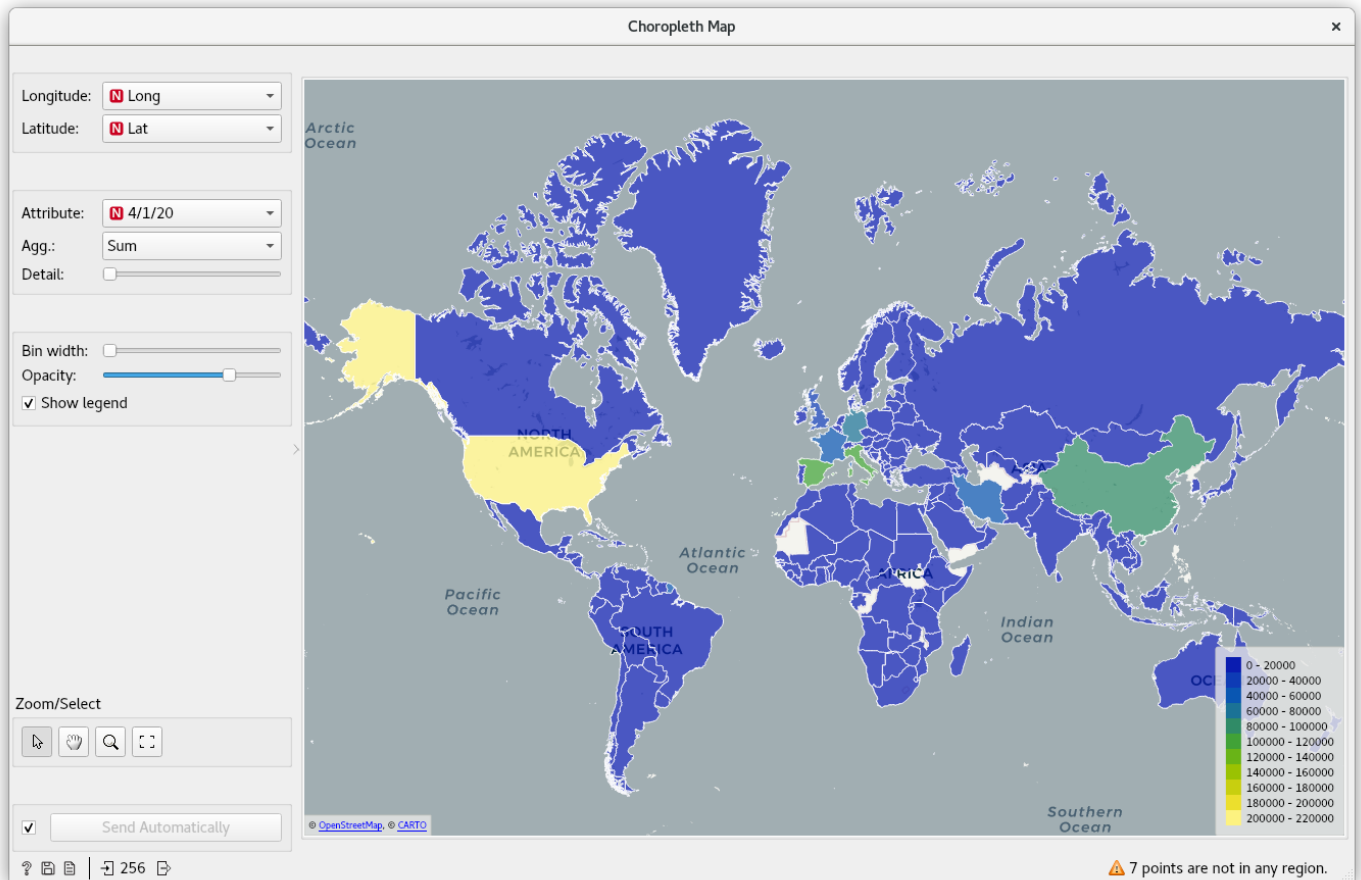
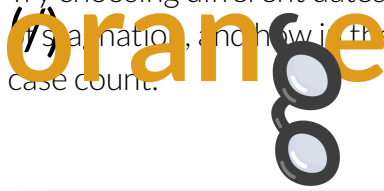
The map is nice and informative, and it's easy to distinguish larger epidemics from smaller ones. Still, those similarly-sized points on the map represent entire countries from big Russia to small Slovenia. We could, alternatively, represent countries by coloring their surface area. This is a job for Choropleth Map. Let's add it and connect it.



(/blog\_img/2020/2020-04-09-workflow-02.png)

This widget is similar to Geo Map, but instead of displaying data as points, it aggregates them into regions (e.g., Country, State, Province), and displays those instead. Under *Attribute*, select the latest date and under *Agg.*, select *sum*. The colored regions now show the sum of confirmed COVID-19 cases for each country.

Try choosing different dates. Notice how China initially experiences rapid growth, followed by a period of stagnation, and how in the past week, the USA overtook all other countries' confirmed COVID-19 case count.



(/blog\_img/2020/2020-04-09-choropleth.png)

A warning that *7 points are not in any region* refers to some small islands and cruise ships included in the data.

## If a picture paints a thousand words, an animation paints a million

When working with time-series data, the coolest insights arise from observing trends – how data changes depending on time. Let's animate our maps!

We'll be using a special widget from Orange3-Timeseries: Time Slice, which lets you select data in a time window. It also features a playback function that slowly moves the time window, outputting corresponding data as it goes.

### Transpose: part I

Time Slice requires time instances in the form of rows, while our data specifies them in columns. Let's start again with the File widget that loads the data. Move *Lat* and *Long* to meta attributes by double-clicking *feature* in the *Role* column and selecting *meta*. This will tell Orange to exclude them from data manipulations.



**Iris flower dataset**  
Classical dataset with 150 instances of Iris setosa, Iris virginica and Iris versicolor.

150 instance(s)  
4 feature(s) (no missing values)  
Classification; categorical class with 3 values (no missing values)  
0 meta attribute(s)

Columns (Double click to edit)

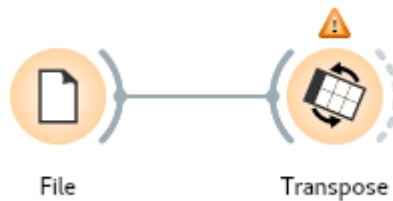
	Name	Type	Role	Values
1	sepal length	N numeric	feature	
2	sepal width	N numeric	feature	
3	petal length	N numeric	feature	
4	petal width	N numeric	feature	
5	iris	C categorical target	target	Iris-setosa, Iris-versicolor, Iris-...

Browse documentation datasets      Reset      Apply

? | 150

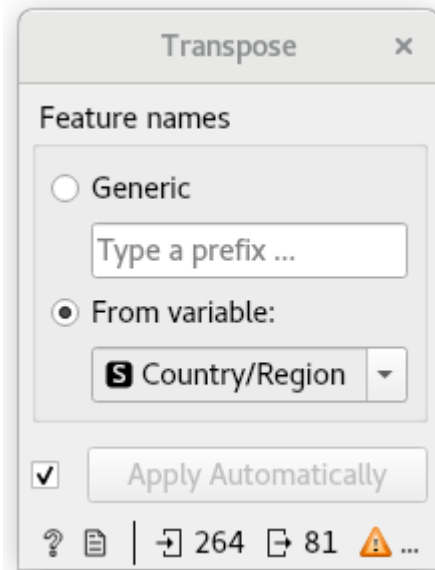
(/blog\_img/2020/2020-04-09-file.gif)

To flip rows and columns, we use the Transpose widget. Connect it to the File widget.



(/blog\_img/2020/2020-04-09-workflow-03.png)

Set *From variable* to *Country/Region*. Ignore the warning that indices were added to variables with the same name; it won't hurt us.



(/blog\_img/2020/2020-04-09-transpose-1.png)

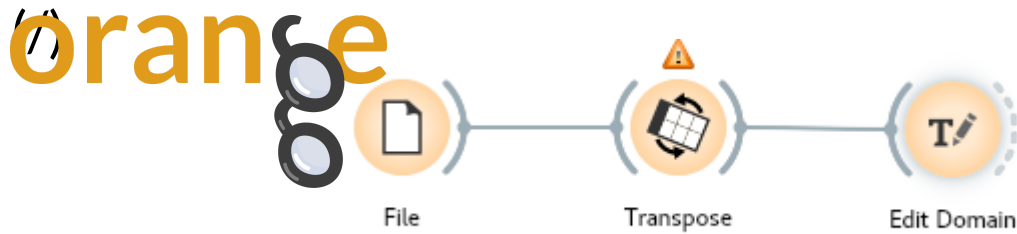
To better understand what just happened, connect a Data Table to Transpose and see the result: rows and columns were swapped. Columns are named after countries (because we chose *Country/Region* in the Transpose widget), while a new column, *Feature name*, contains the original names of the columns.

Country/Region	Feature name	Afghanistan Afghanistan	Albania Albania	Algeria Algeria	Andorra Andorra	Angola Angola	Burkina Faso Burkina Faso	Argentina Argentina
Lat		33	41.1533	28.0339	42.5063	-11.2027	17.0608	-38.416
Long		65	20.1683	1.6596	1.5218	17.8739	-61.7964	-63.616
Province/State								
1	1/22/20	0	0	0	0	0	0	0
2	1/23/20	0	0	0	0	0	0	0
3	1/24/20	0	0	0	0	0	0	0
4	1/25/20	0	0	0	0	0	0	0
5	1/26/20	0	0	0	0	0	0	0
6	1/27/20	0	0	0	0	0	0	0
7	1/28/20	0	0	0	0	0	0	0
8	1/29/20	0	0	0	0	0	0	0
9	1/30/20	0	0	0	0	0	0	0
10	1/31/20	0	0	0	0	0	0	0
11	2/1/20	0	0	0	0	0	0	0
12	2/2/20	0	0	0	0	0	0	0
13	2/3/20	0	0	0	0	0	0	0
14	2/4/20	0	0	0	0	0	0	0
15	2/5/20	0	0	0	0	0	0	0
16	2/6/20	0	0	0	0	0	0	0
17	2/7/20	0	0	0	0	0	0	0
18	2/8/20	0	0	0	0	0	0	0
19	2/9/20	0	0	0	0	0	0	0
20	2/10/20	0	0	0	0	0	0	0
21	2/11/20	0	0	0	0	0	0	0

(/blog\_img/2020/2020-04-09-table-1.png)

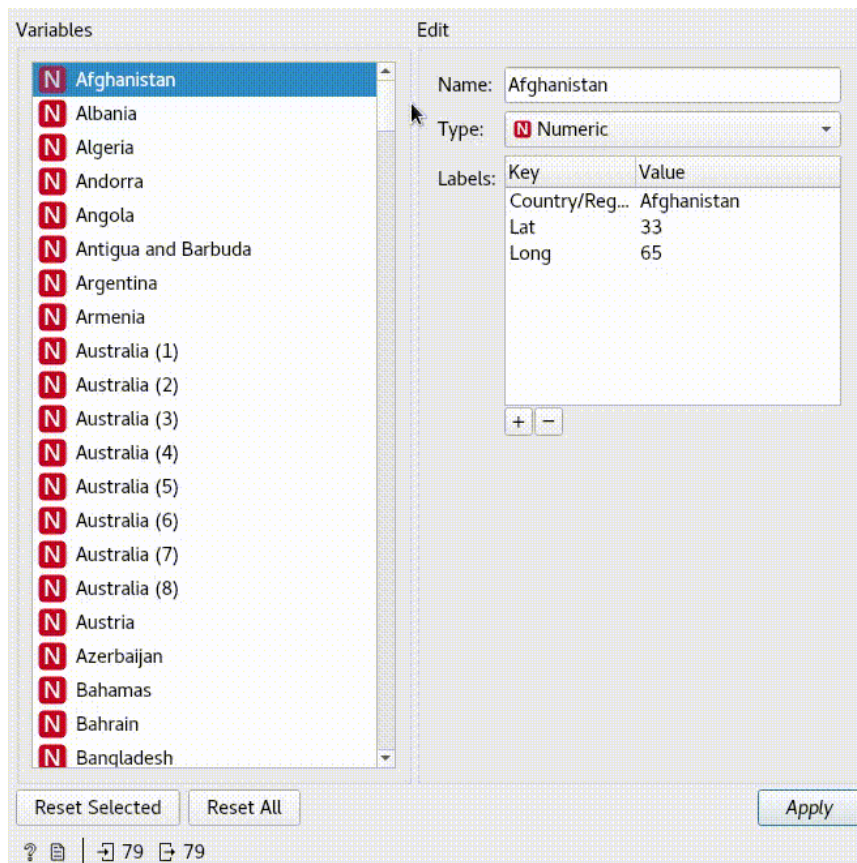
Right now, Orange is treating values in *Feature name* as strings. We need to tell Orange that the feature contains timestamps. Let's connect an Edit Domain widget.





(/blog\_img/2020/2020-04-09-workflow-04.png)

Scroll to the bottom of *Variables* and select *Feature name*. In *Edit*, change its type to *Time*. While you're at it, change its name to something meaningful, like *date*. Finally, click *Apply*

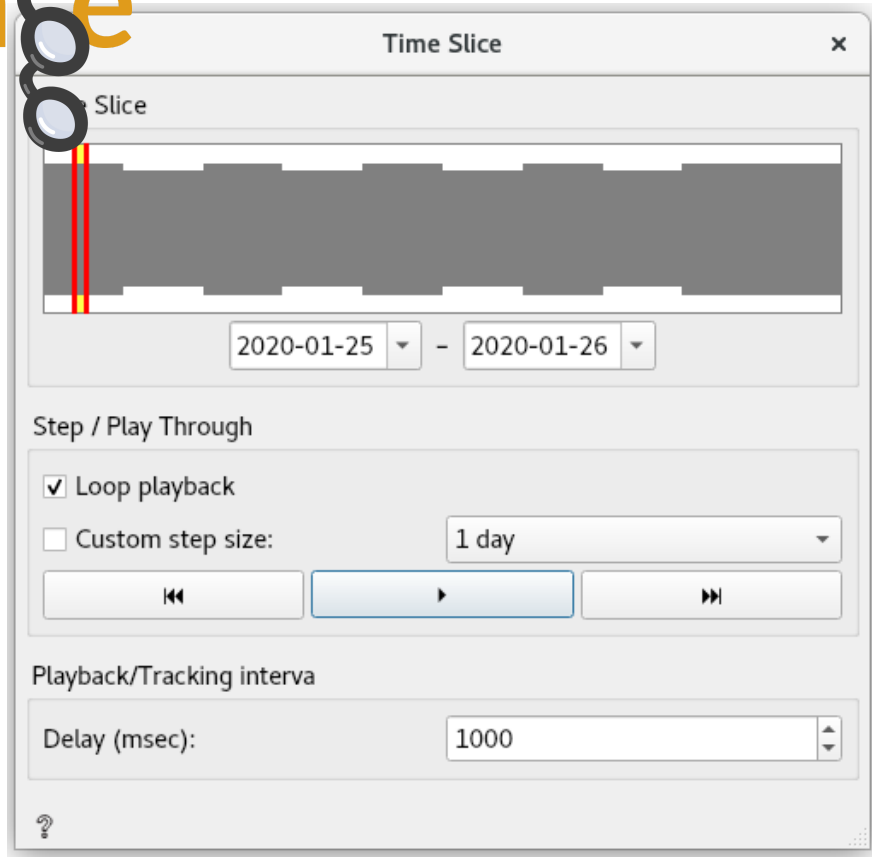


(/blog\_img/2020/2020-04-09-edit-domain.gif)

The data is now ready to be processed by Time Slice.



(/blog\_img/2020/2020-04-09-workflow-05.png)

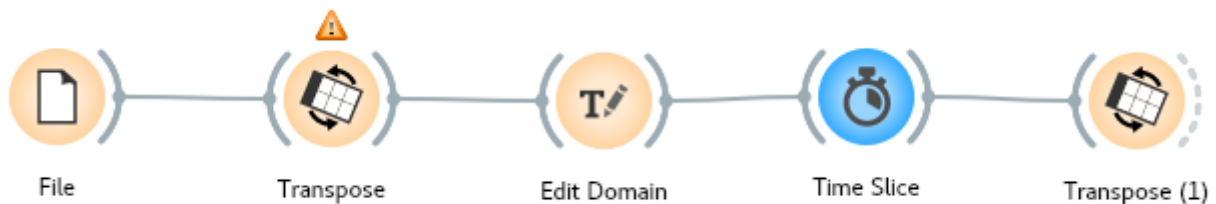
(/blog\_img/2020/2020-04-09-time-slice.png)

The top indicates the currently selected slice of data. You can change the interval by interacting with the histogram, or by adjusting the start and end dates directly below. Let's select a day's slice between the 25th and 26th January 2020.

With a click of the play button, the slice moves, and the data output updates accordingly. To verify that it works, connect a Data Table, and see it in action.

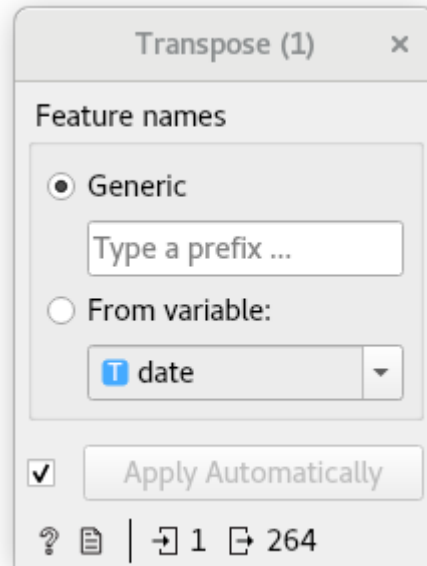
Transpose: part II

Make sure to stop the Time Slice playback before proceeding. We're almost at the finish line; we just need to transform the data back into its original form. To do so, connect another Transpose widget.



(/blog\_img/2020/2020-04-09-workflow-06.png)

Under *Future names* select *Generic*, which will name the new column *Feature 1*.



(/blog\_img/2020/2020-04-09-transpose-2.png)

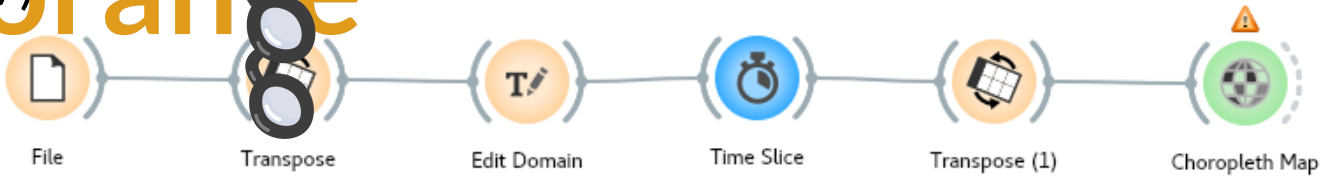
Let's connect a Data Table for a final check of our data before putting it on a map.

	Feature name	Lat	Long	Province/State	Country/Region	Feature 1
1	Afghanistan	33	65	?	Afghanistan	0
2	Albania	41.1533	20.1683	?	Albania	0
3	Algeria	28.0339	1.6596	?	Algeria	0
4	Andorra	42.5063	1.5218	?	Andorra	0
5	Angola	-11.2027	17.8739	?	Angola	0
6	Antigua and ...	17.0608	-61.7964	?	Antigua and ...	0
7	Argentina	-38.4161	-63.6167	?	Argentina	0
8	Armenia	40.0691	45.0382	?	Armenia	0
9	Australia (1)	-35.4735	149.012	Australian ...	Australia	0
10	Australia (2)	-33.8688	151.209	New South ...	Australia	4
11	Australia (3)	-12.4634	130.846	Northern ...	Australia	0
12	Australia (4)	-28.0167	153.4	Queensland	Australia	3
13	Australia (5)	-34.9285	138.601	South ...	Australia	0
14	Australia (6)	-41.4545	145.971	Tasmania	Australia	0
15	Australia (7)	-37.8136	144.963	Victoria	Australia	2
16	Australia (8)	-31.9505	115.861	Western ...	Australia	0
17	Austria	47.5162	14.5501	?	Austria	0
18	Azerbaijan	40.1431	47.5769	?	Azerbaijan	0
19	Bahamas	25.0343	-77.3963	?	Bahamas	0
20	Bahrain	26.0275	50.55	?	Bahrain	0
21	Bangladesh	23.685	90.3563	?	Bangladesh	0
22	Barbados	13.1939	-59.5432	?	Barbados	0

(/blog\_img/2020/2020-04-09-table-2.png)

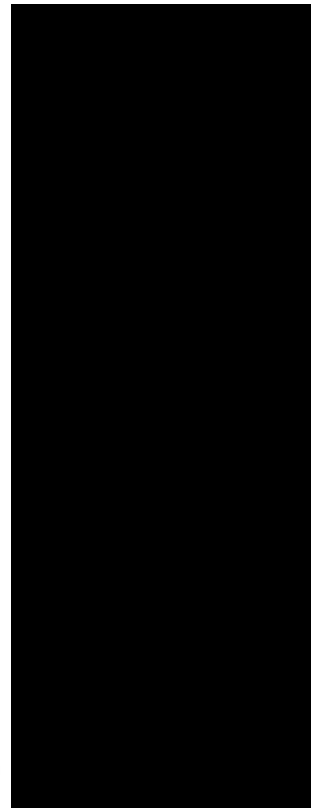
Looking good!

The final step entails connecting the Choropleth Map, selecting *Feature 1* as *Attribute*, and *Sum* as the *Agg*.



(/blog\_img/2020/2020-04-09-workflow-07.png)

To observe the animation, open up both Choropleth and Time Slice at the same time.  
Protip: pressing Ctrl/CMD+Up reveals the workflow, Ctrl/CMD+Down reveals widget windows.  
All that's left is to hit play.



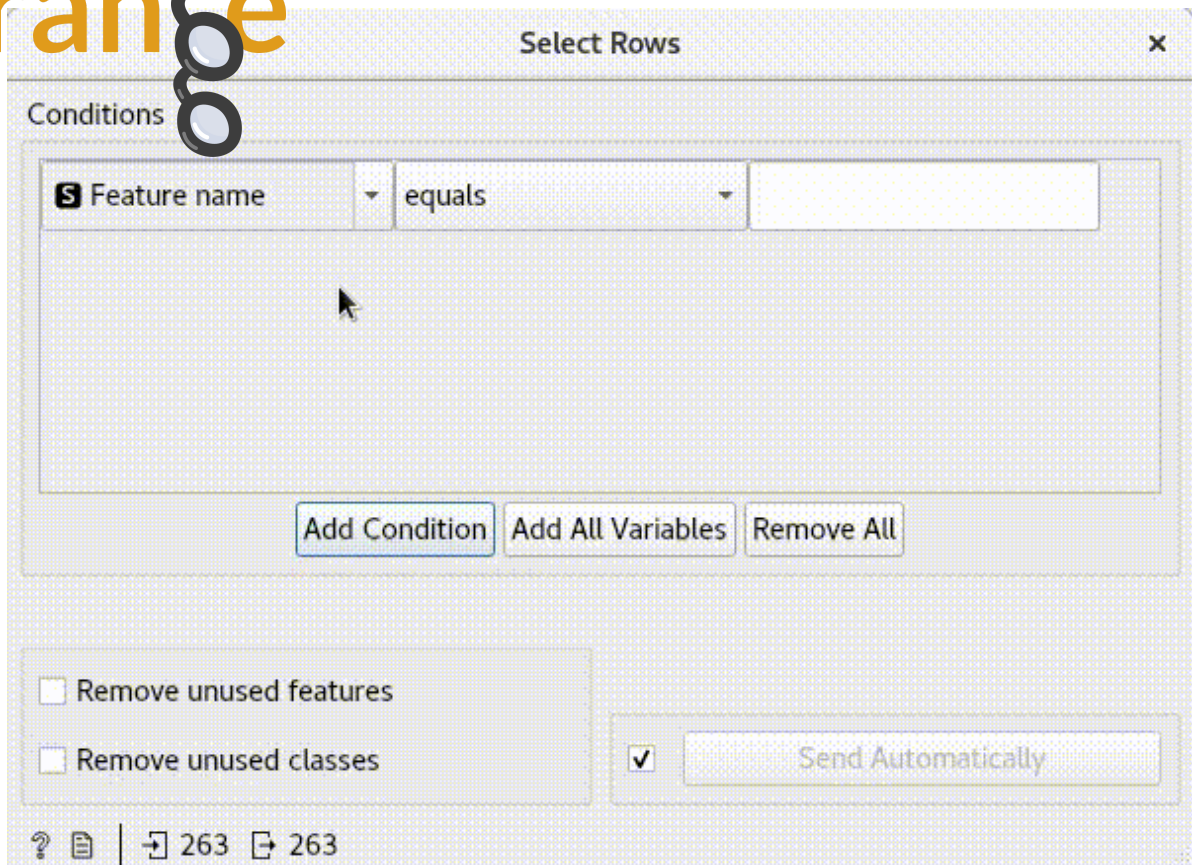
0:00 / 0:17

Now that we're familiar with Time Slice, let's use it with Geo Map too. For bonus points, let's also remove countries with zero confirmed cases. That way, the number of points on the map will increase as the virus spreads throughout the world. For this, we'll use a Select Rows widget.



(/blog\_img/2020/2020-04-09-workflow-08.png)

Under Conditions, select *Feature 1*, select *is greater than*, and type in 0.



(/blog\_img/2020/2020-04-09-select-rows.gif)

Add and connect a Geo Map widget. Zoom out, so you see the whole world, and on the left side, check *Freeze map*. This will prevent the map from jumping around as new points are added.



(/blog\_img/2020/2020-04-09-workflow-09.png)

Press play once again – let's see what we've built!



0:00 / 0:23

## Where from here?

In the blog, we explained how to turn Orange into an interactive data exploration tool for COVID-19 data! With a few widgets we were able to determine the change in time for the confirmed cases and plot the results on a map.

Time Slice isn't the only widget in the Orange3-Timeseries add-on. Stay tuned for next week's blog, where we'll delve deeper into the patterns emerging from time-series data.

---

Orange is a multi-platform open-source machine learning and data visualization tool for beginners and experts alike. **Download Orange** (<https://orange.biolab.si/download>), and load and explore your own data sets!

In addition to a variety of learning materials posted online in the form of **blog posts** (<https://orange.biolab.si/blog/>), **tutorial videos** (<https://www.youtube.com/playlist?list=PLmNPvQr9Tf-ZSDLwOzxpY-HrE0yv-8Fy>), we've created a Discord server. **Join the community, tell us what you think!** (<https://discord.gg/FWrfeXV>)



Comment Community

Privacy Policy



Recommend

Tweet

Share

Sort by Best

Join the discussion...



HughSaysHelloThere • 8 days ago

I installed the geo widget. No choropleth map. Only geo map.

1 ^ | v • Reply • Share ›



Игорь Клейнер → HughSaysHelloThere • 5 days ago

instal update to this model

^ | v • Reply • Share ›



HughSaysHelloThere → Игорь Клейнер • 4 days ago

i uninstalled and reinstalled both orange and the geo add-on. same thing.

Am on Orange 3.25 and orange-geo 0.2.7

No choropleth map. Only geo map



1 ^ | v • Reply • Share ›



Alberto Garcia • 13 days ago

Very nice, I am waiting for next time series article, thanks very much

^ | v • Reply • Share ›



## Download

[Windows \(/download/#windows\)](/download/#windows)

[Mac OS \(/download/#macos\)](/download/#macos)

## Community

[Facebook \(https://www.facebook.com/orangedm\)](https://www.facebook.com/orangedm)

[YouTube \(https://www.youtube.com/channel/UCIKKWBe2SCAEyv7ZNGhle4g\)](https://www.youtube.com/channel/UCIKKWBe2SCAEyv7ZNGhle4g)

[Twitter \(https://twitter.com/orangedataminer\)](https://twitter.com/orangedataminer)

[Stack Exchange \(https://datascience.stackexchange.com/questions/tagged/orange\)](https://datascience.stackexchange.com/questions/tagged/orange)

[Discord \(https://discord.gg/FWrfeXV\)](https://discord.gg/FWrfeXV)

## Documentation

[Get started \(/getting-started/\)](/getting-started/)

[Widgets \(/widget-catalog/\)](/widget-catalog/)

[Scripting \(http://docs.orange.biolab.si/3/data-mining-library/\)](http://docs.orange.biolab.si/3/data-mining-library/)

## Developers

[GitHub \(https://github.com/biolab/orange3\)](https://github.com/biolab/orange3)

## Latest blog posts (/blog/)

20 Apr

[Data Mining COVID-19 Epidemics: Part 3 \(/blog/2020/2020-4-015-covid-19-part-3/\)](/blog/2020/2020-4-015-covid-19-part-3/)

13 Apr

[Data Mining COVID-19 Epidemics: Part 2 \(/blog/2020/2020-04-09-covid-19-part-2/\)](/blog/2020/2020-04-09-covid-19-part-2/)

02 Apr

[Data Mining COVID-19 Epidemics: Part 1 \(/blog/2020/2020-04-02-covid-19-basic/\)](/blog/2020/2020-04-02-covid-19-basic/)

Copyright © University of Ljubljana