

Document Map

Displays geographic locations mentioned in the text.

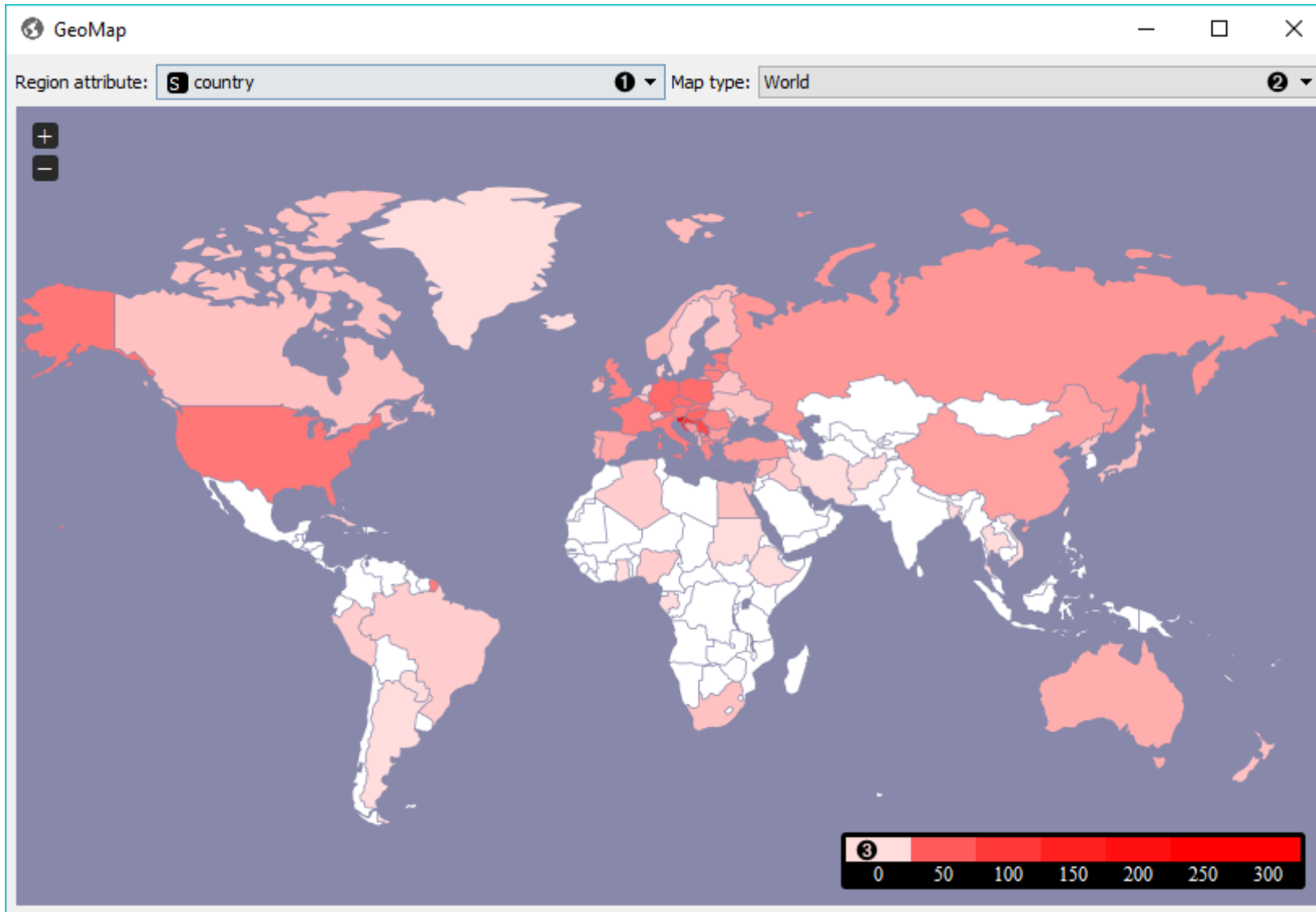
Inputs

- Data: Data set.

Outputs

- Corpus: Documents containing mentions of selected geographical regions.

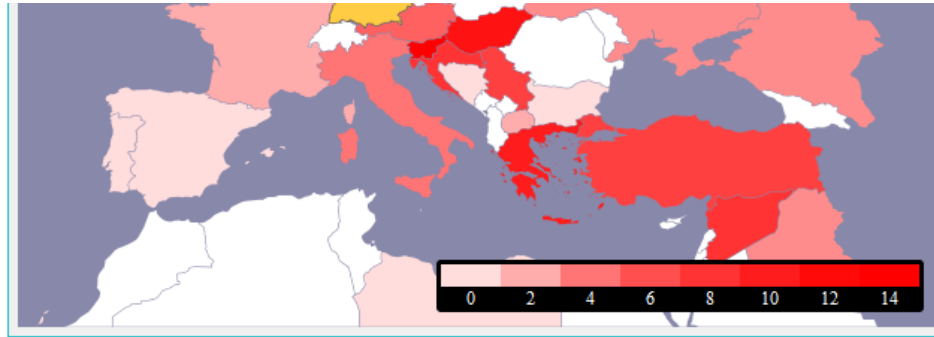
Document Map widget shows geolocations from textual (string) data. It finds mentions of geographic names (countries and capitals) and displays distributions (frequency of mentions) of these names on a map. It works with any Orange widget that outputs a data table and that contains at least one string attribute. The widget outputs selected data instances, that is all documents containing mentions of a selected country (or countries).



1. Select the meta attribute you want to search geolocations by. The widget will find all mentions of geolocations in a text and display distributions on a map.
2. Select the type of map you wish to display. The options are *World*, *Europe* and *USA*. You can zoom in and out of the map by pressing + and - buttons on a map or by mouse scroll.
3. The legend for the geographic distribution of data. Countries with the boldest color are most often mentioned in the selected region attribute (highest frequency).

Example

The screenshot displays the Orange3 data mining environment. A workflow is constructed with the following widgets: 'NY Times' (data source), 'GeoMap' (map widget), 'Corpus Viewer' (text viewer), 'Preprocess Text' (text preprocessing), 'Corpus Viewer (1)' (text viewer), and 'Word Cloud' (word cloud visualization). The 'Word Cloud' widget is open, showing settings for 'Color words' (checked) and 'Words tilt' (set to 'no'). A 'Regenerate word cloud' button is visible. The 'GeoMap' widget is also open, showing a map of Europe with a legend for 'Region attribute: country'. The 'Word Cloud' widget displays a word cloud with terms like 'germany', 'migrants', 'speech', 'immigrant', 'violence', 'debate', 'austria', 'anti', 'breaking point', 'reunification', 'road', 'tourists', 'rules', 'center', 'bound', 'threats', 'trade', 'pour', 'reached', 'irish', 'perspective', 'hans', 'hate', 'distribution', 'role', 'grow', 'dies', 'takes', 'architect', 'dietrich', 'chinese', 'tensions', 'slovenia', 'genscher', 'brexit', '89', 'new', 'german', 'europe'.



Then we sent the data to **Document Map** to see distributions of geolocations by *country* attribute. The attribute already contains country tags for each article, which is why **NY Times** is great in combinations with **Document Map**. We selected Germany, which sends all the documents tagged with Germany to the output. Remember, we queried **NY Times** for articles on Slovenia.

We can again inspect the output with **Corpus Viewer**. But there's a more interesting way of visualizing the data. We've sent selected documents to **Preprocess Text**, where we've tokenized text to words and removed stopwords.

Finally, we can inspect the top words appearing in last year's documents on Slovenia and mentioning also Germany with **Word Cloud**.