Badge-3 Lab-5 [SOM]

Out date: Aug 10, 2022, 2020
Due date: Aug 14, 2020 at 11:59PM

Submission

1. Prepare your solution in Orange and save the workspace (e.g., Lab-5_SOM_lastname.ows).
2. Complete the table given below and save the file (e.g., Lab-5_SOM_lastname.docx)
3. Upload the files to the Canvas.

Background information: Oil and gas reservoirs lie deep beneath the Earth's surface. Geologists and engineers cannot examine the rock formations in situ, so tools called sondes go there for them. Specialists lower these tools into a wellbore and obtain measurements of subsurface properties. The data are displayed as a series of measurements covering a depth range in a display called a well log. Often, several tools are run simultaneously as a logging string, and the combination of results is more informative than each individual measurement (https://www.slb.com/resource-library/oilfield-review/defining-series/defining-logging).

Link below gives an overview of interpreting lithology using Gamma Ray, Density porosity and Neutron Porosity logs.
http://www.kgs.ku.edu/Publications/Bulletins/LA/05_overlay.html

LAS file (1033440835.las) containing Gamma Ray, Caliper, Density Porosity and Neutron Porosity for well Beck 'A' #1 that is used in the overview link above was downloaded from the link below.
https://chasm.kgs.ku.edu/ords/las.lasd5.SelectWells

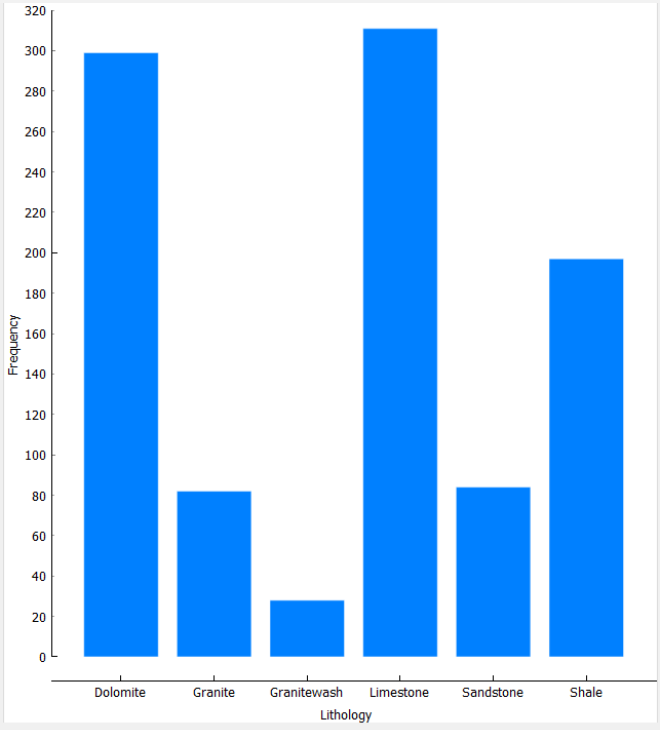
Please refer to http://www.kgs.ku.edu/General/copyright.html regarding use of data / information from Kansas Geological Survey.

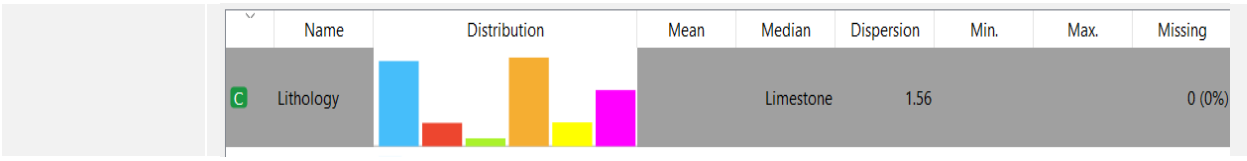
Objective: To use SOM in Orange to cluster the input log dataset.

Data: Relevant log data was extracted to an excel file (***Log Lithology classification example.xlsx***) and lithology labels were created to be used for the hands-on lab.

Lab Instructions

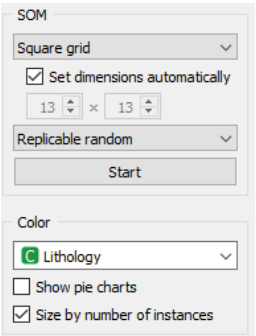
1. Download from Canvas the *Log Lithology classification example.xlsx* file and *Lab5-SOM-start.ows* to your working folder
2. Launch Orange and open the *Lab5-SOM-start.ows* file. Inspect the pipeline and complete the table below: (15 points)

How many features?	1001
How many samples?	6, no meta
What is the distribution of the different Lithology labels?	Dolomite, Granite, Granitewash, Limestone, Sandstone, Shale All looks to be histograms and skewed 



3. Add **Self-Organizing Map** widget to the **Preprocess** widget.

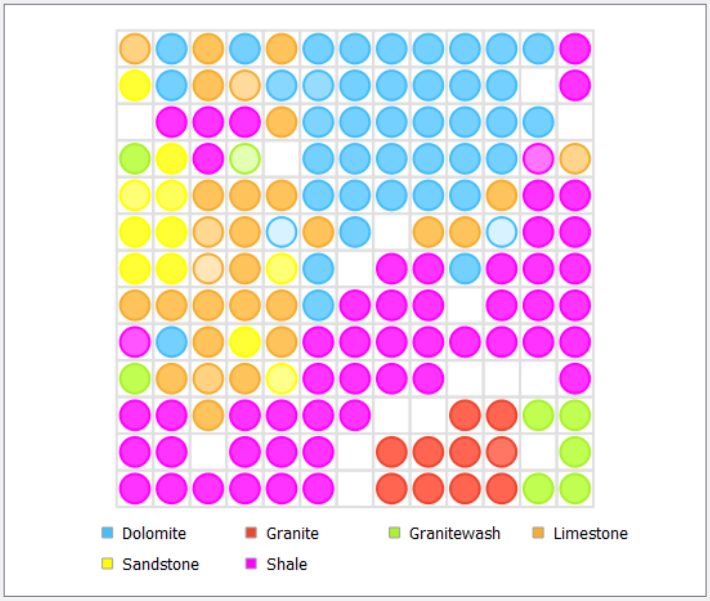
4. Open **Self-Organizing Map** widget and set parameters as below:



5. Change parameters in SOM and complete the table below: (85 points)

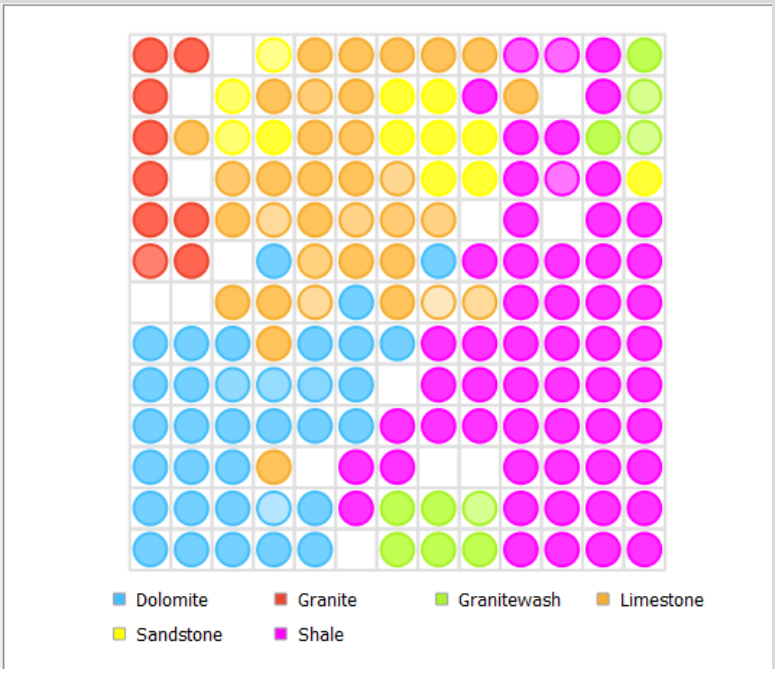
1. With parameters as shown in the above image, how many clusters do you see?

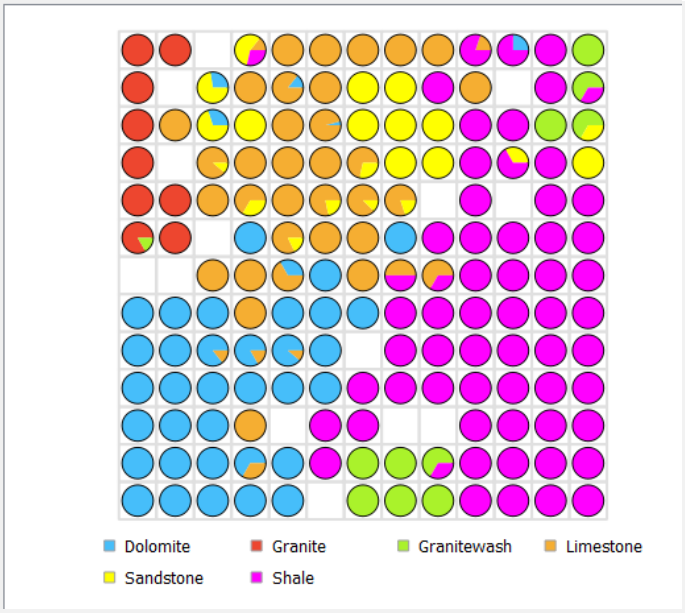
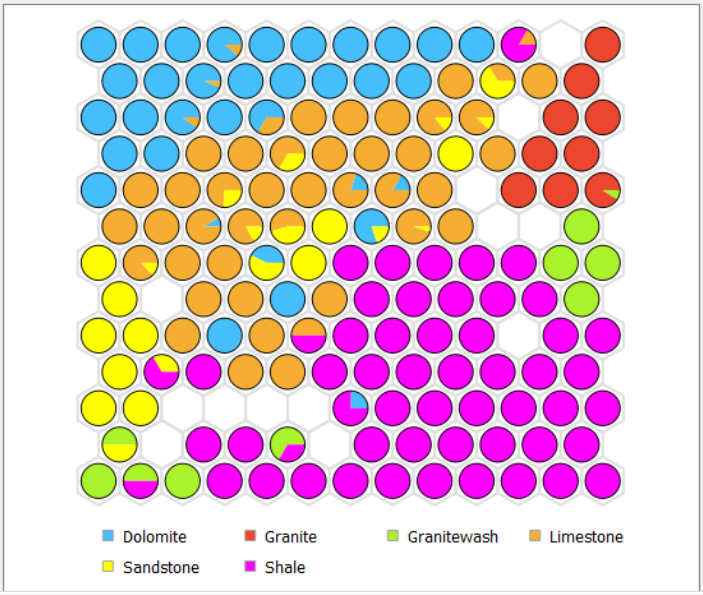
6



2. Remove Caliper variable from Select Columns. Run SOM. How many clusters do you see? Any improvement from the earlier SOM viz?

6, yes it did visually

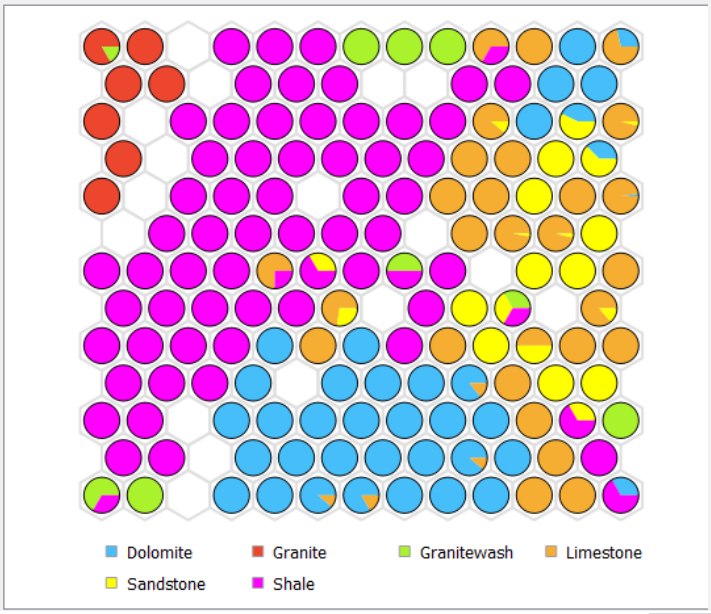


3. Uncheck ‘Size by number of instances’ and select ‘Show pie charts’. What can you comment about the clusters now?	 this will tell you how each neuron contains the different samples, we will some really solid and others are split/overlap (indicates not clear distinction)
4. Change grid to Hexagonal. What about the clusters now?	Looks visually better to distinguish different clusters 
5. Change weight initialization to PCA. What is your	Changes the formation of the clusters, meaning their positions are not the same as before.

observation about the clusters?

PCA, to reduce the dimension

But not the best in grouping the cluster based what we see.

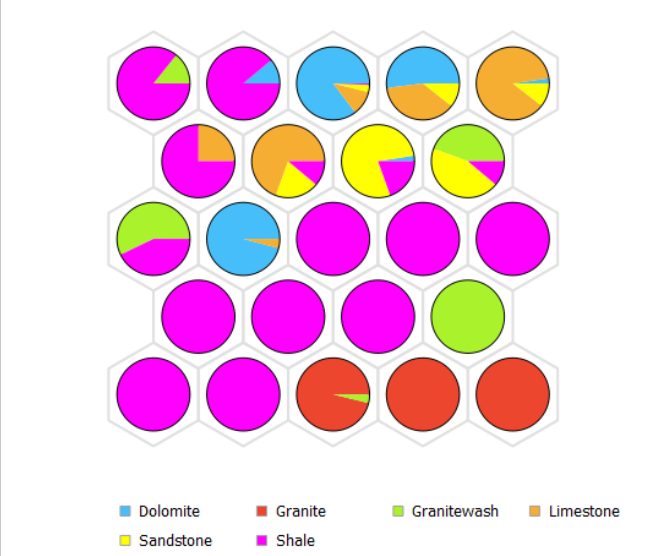


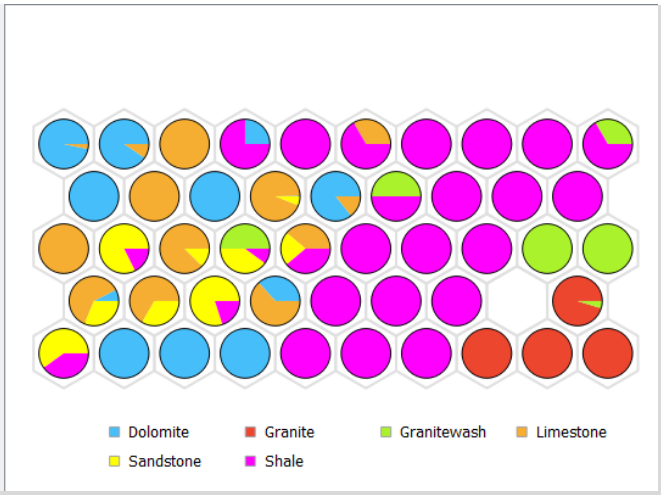
6. Set dimensions manually to the following and observe the SOM clusters.

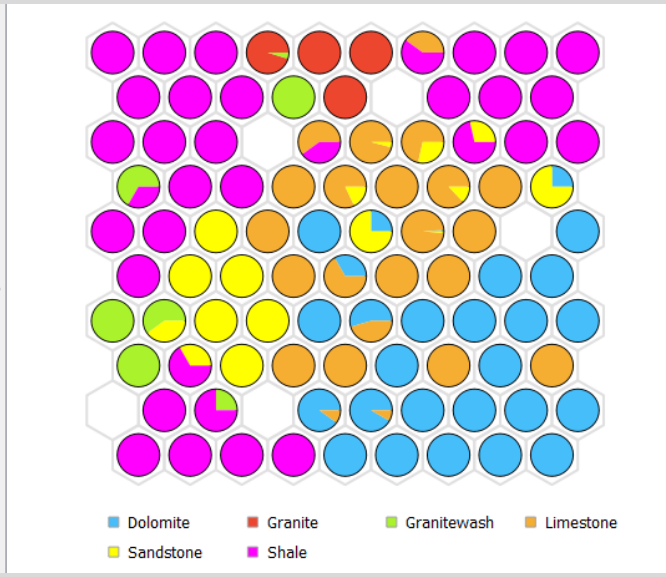
   Use Replicable Random for weights.

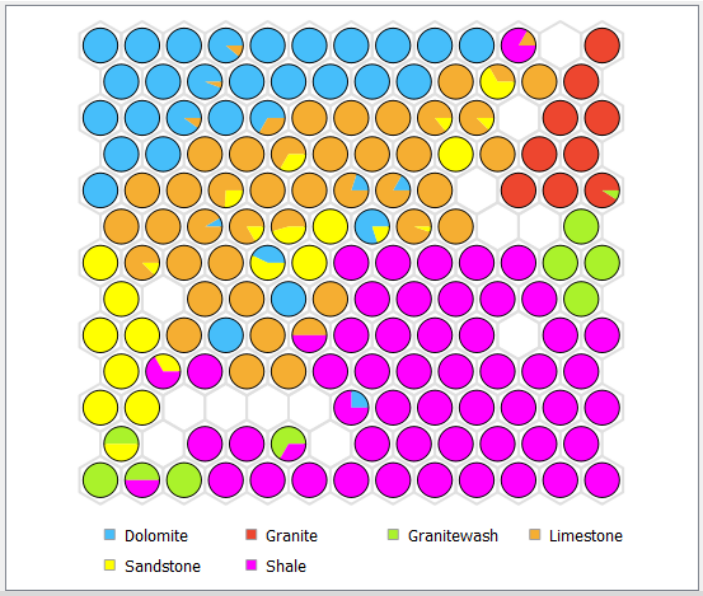
   Which dimension would you prefer?

5x5
10x5
10x10
13x13 (automatic dimensions)




Dolomite
Granite
Granitewash
Limestone
Sandstone
Shale


Dolomite
Granite
Granitewash
Limestone
Sandstone
Shale



I would still prefer the automatic because much better visualization of all the nodes and the clusters of the nodes compared to the others. It tells more story and groups than limit to smaller size.