HSNE Demo

Purpose

This demo illustrates the capabilities of the nptsne HSNE wrapping combined with *nptsne.hsne_analysis* classes for visual analytics support. The demo application is a Qt based GUI for the analysis of multidimensional data. It allows a user to create or load an HSNE hierarchy using the *nptsne.HSne* class and then navigate the HSNE model using the supporting *AnalysisModel* and *Analysis* classes in the *nptsne.hsne_analysis* sub module.

A number of pre-packages demos provide a quick start to the application.

Installing

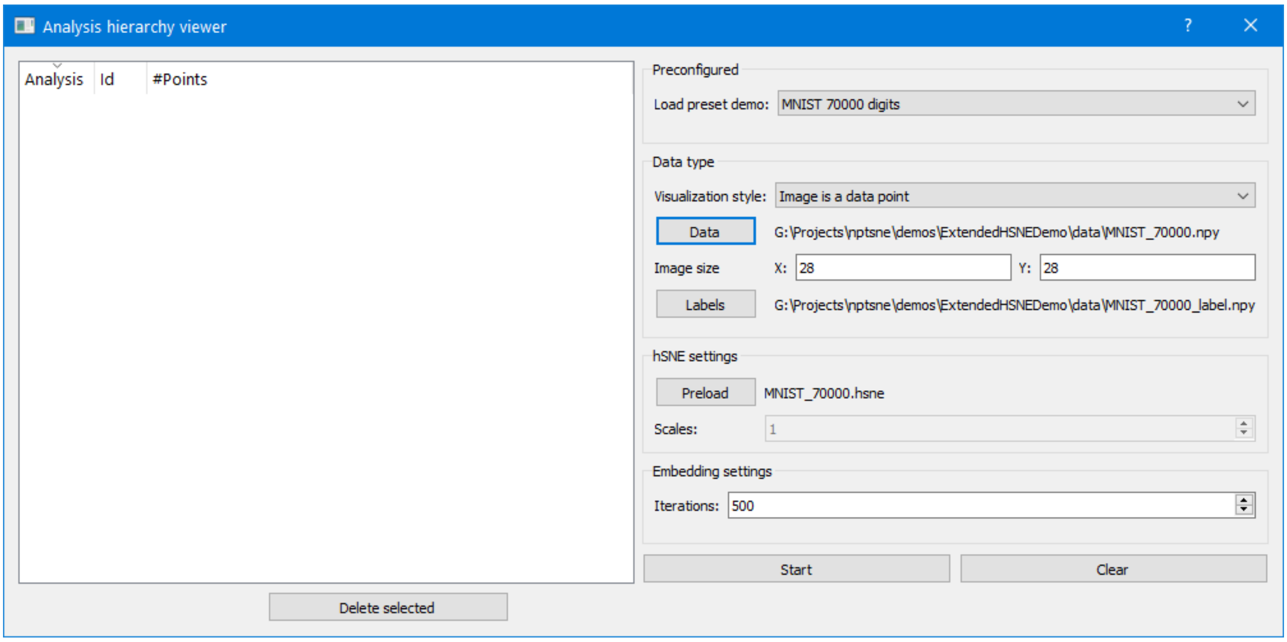
```
> # Unpack data on linux
> python ../unpack_data.py
> # or unpack data on windows
> python ..\unpack_data.py
> pip install -r requirements.txt
```

Running

```
> python hsnedemo.py
```

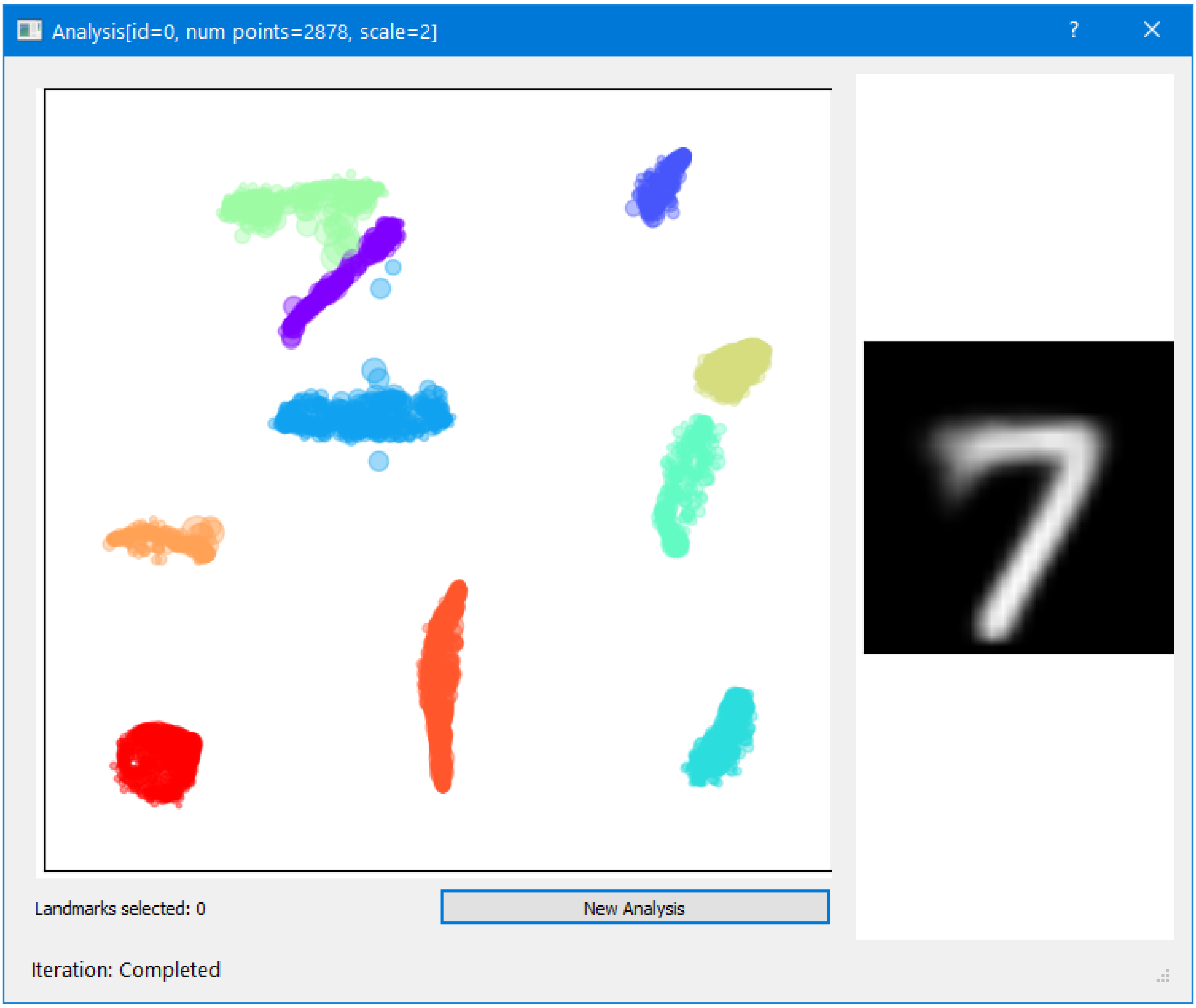
Usage

1. Select a **MNIST 70000 digits** from the **Load preset demo** list. The data type and information concerning the demo will be loaded.



Preconfigured
Load preset demo: MNIST 70000 digits
Data type
Visualization style: Image is a data point
Data G:\Projects\nptsne\demos\ExtendedHSNEDemo\data\MNIST_70000.npy
Image size X: 28 Y: 28
Labels G:\Projects\nptsne\demos\ExtendedHSNEDemo\data\MNIST_70000_label.npy
hSNE settings
Preload MNIST_70000.hsne
Scales: 1
Embedding settings
Iterations: 500

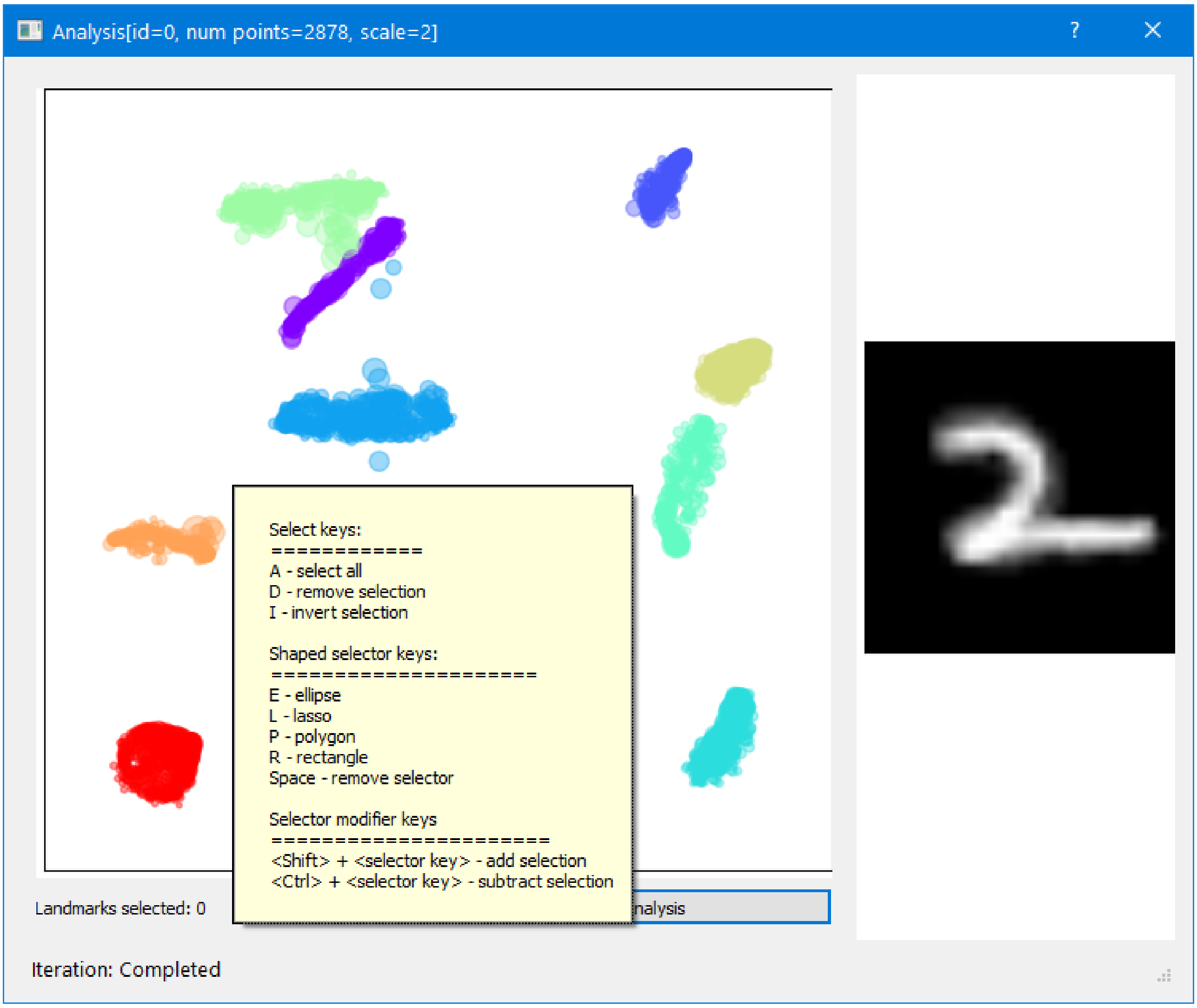
2. Click **Start** and the top-level Analysis embedding will be created.



Landmarks selected: 0

Iteration: Completed

3. Use the select keys to start a selection operation

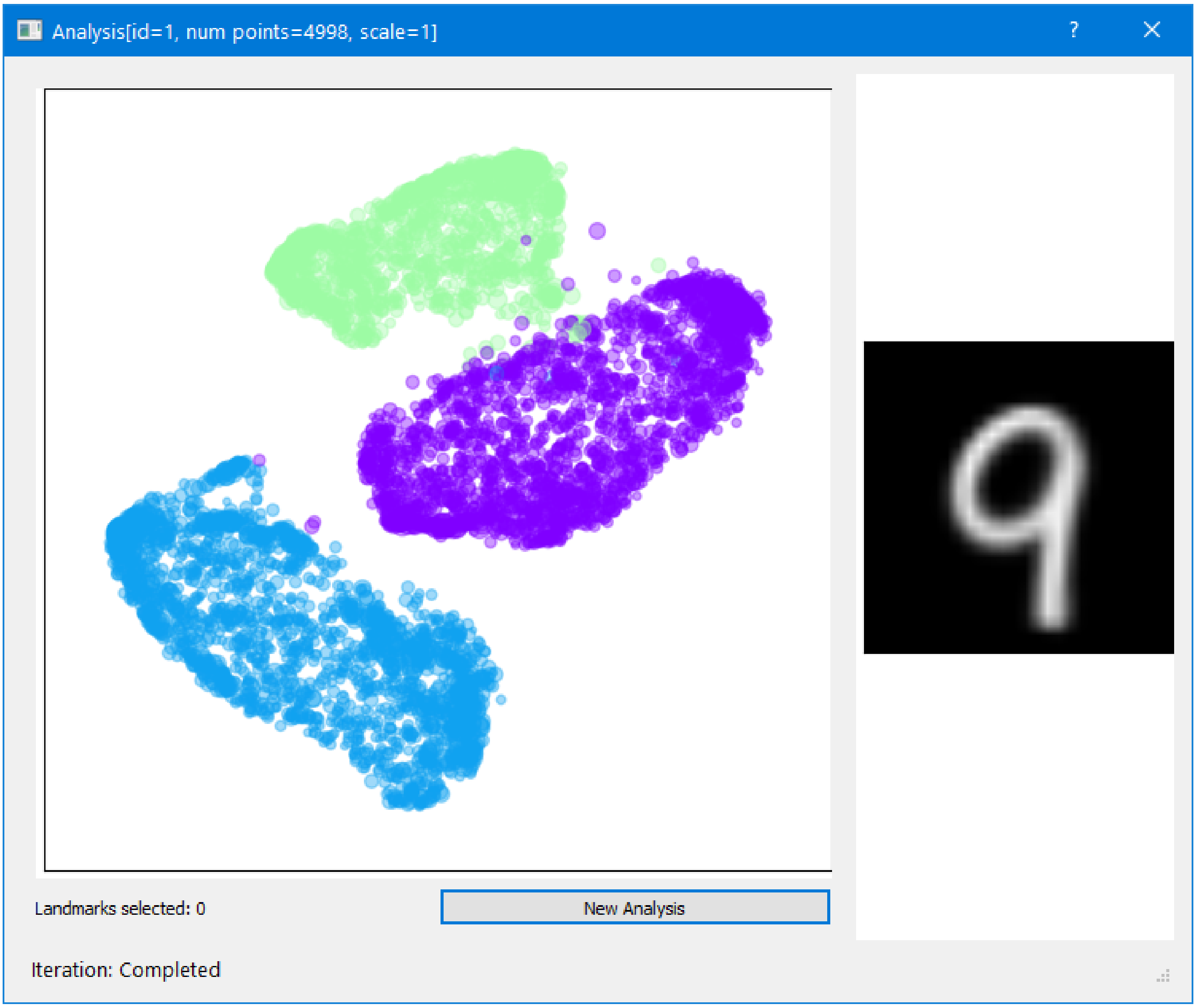


```
Select keys:
============
A - select all
D - remove selection
I - invert selection

Shaped selector keys:
=====================
E - ellipse
L - lasso
P - polygon
R - rectangle
Space - remove selector

Selector modifier keys
======================
<Shift> + <selector key> - add selection
<Ctrl> + <selector key> - subtract selection
```

Then click **New Analysis**. 4. A new embedding is created for the chosen points at a more fine grained HSNE scale.

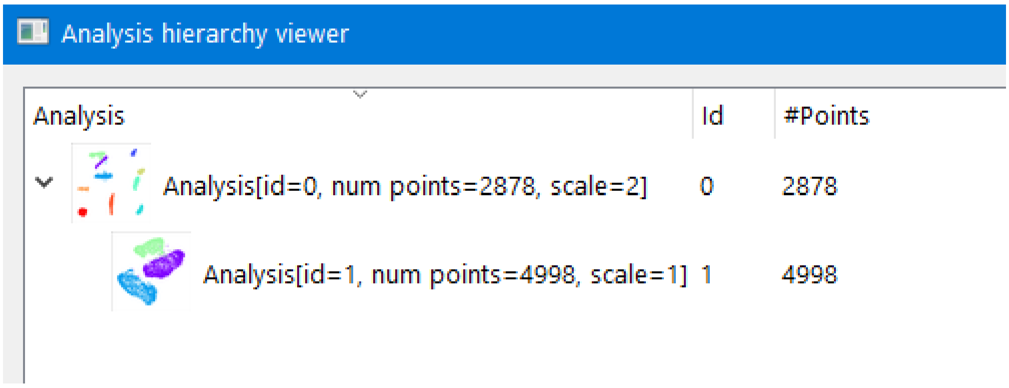


Analysis[id=1, num points=4998, scale=1]

Landmarks selected: 0

Iteration: Completed

5. Meanwhile the ModelGui is updated to reflect the hierarchy.



Analysis hierarchy viewer

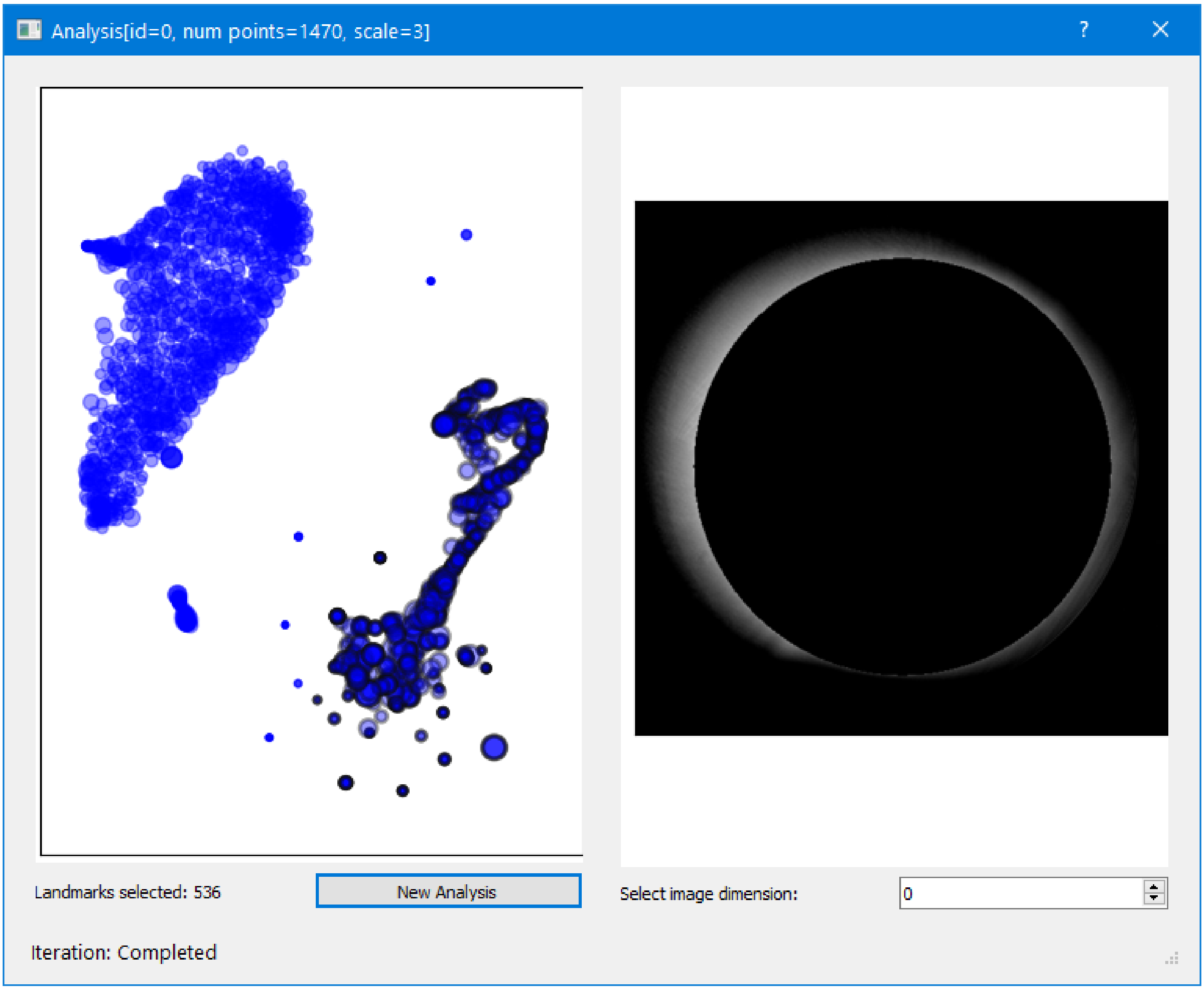
Analysis	Id	#Points
Analysis[id=0, num points=2878, scale=2]	0	2878
Analysis[id=1, num points=4998, scale=1]	1	4998

6. Analyses can be brought to the front by clicking in the ModelGui. Once selected in the ModelGui and Analysis and its sub-analyses can be deleted.

7. Before loading new data restart the program.

8. Load the **Hyperspectral sun (512x512)** data. Two clusters emerge representing

the background pixels (including the corona)



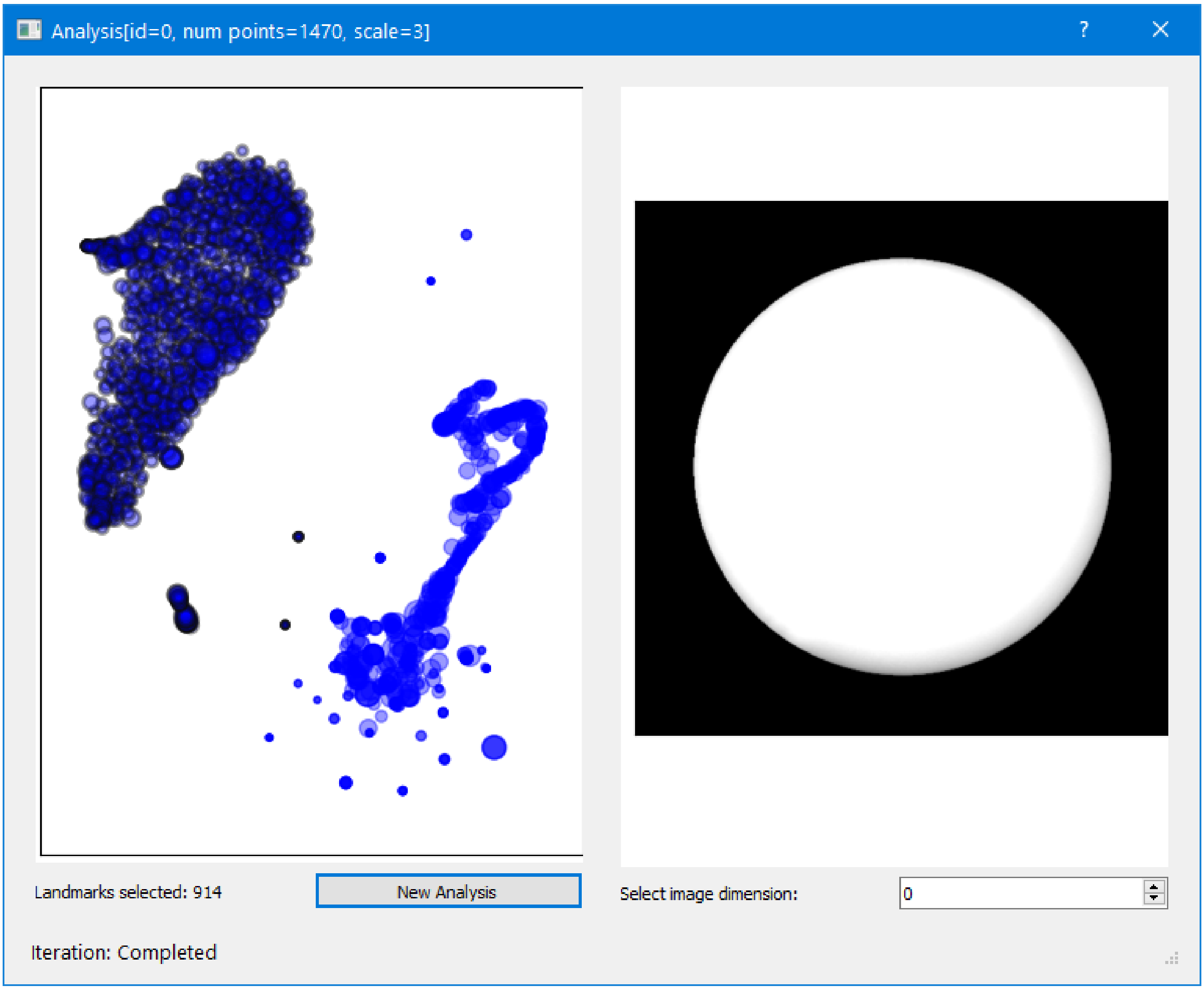
Analysis[id=0, num points=1470, scale=3]

Landmarks selected: 536

Select image dimension: 0

Iteration: Completed

or the photosphere

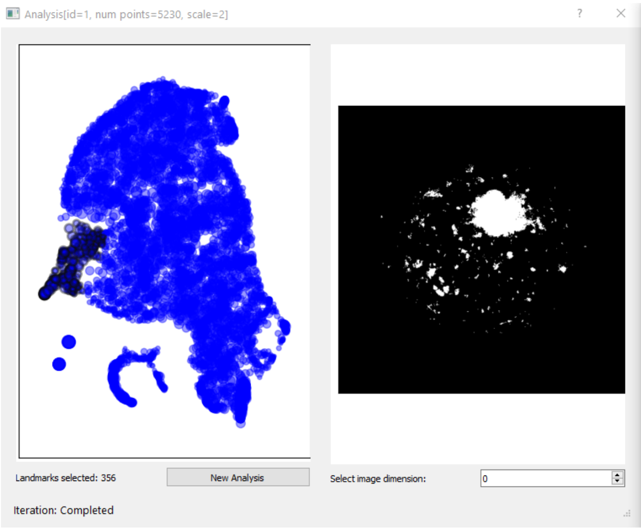

Analysis[id=0, num points=1470, scale=3]

Landmarks selected: 914

Select image dimension: 0

Iteration: Completed

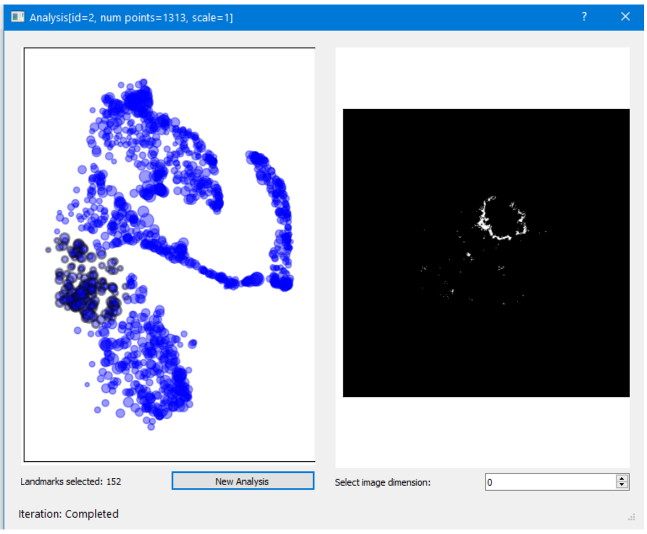
9. Selecting the photosphere for analysis reveals structure in the cluster allowing us to extract (using lasso selector) and examine details at further sub levels


Analysis[id=1, num points=5230, scale=2]

Landmarks selected: 356

Select image dimension: 0

Iteration: Completed


Analysis[id=2, num points=1313, scale=1]

Landmarks selected: 152

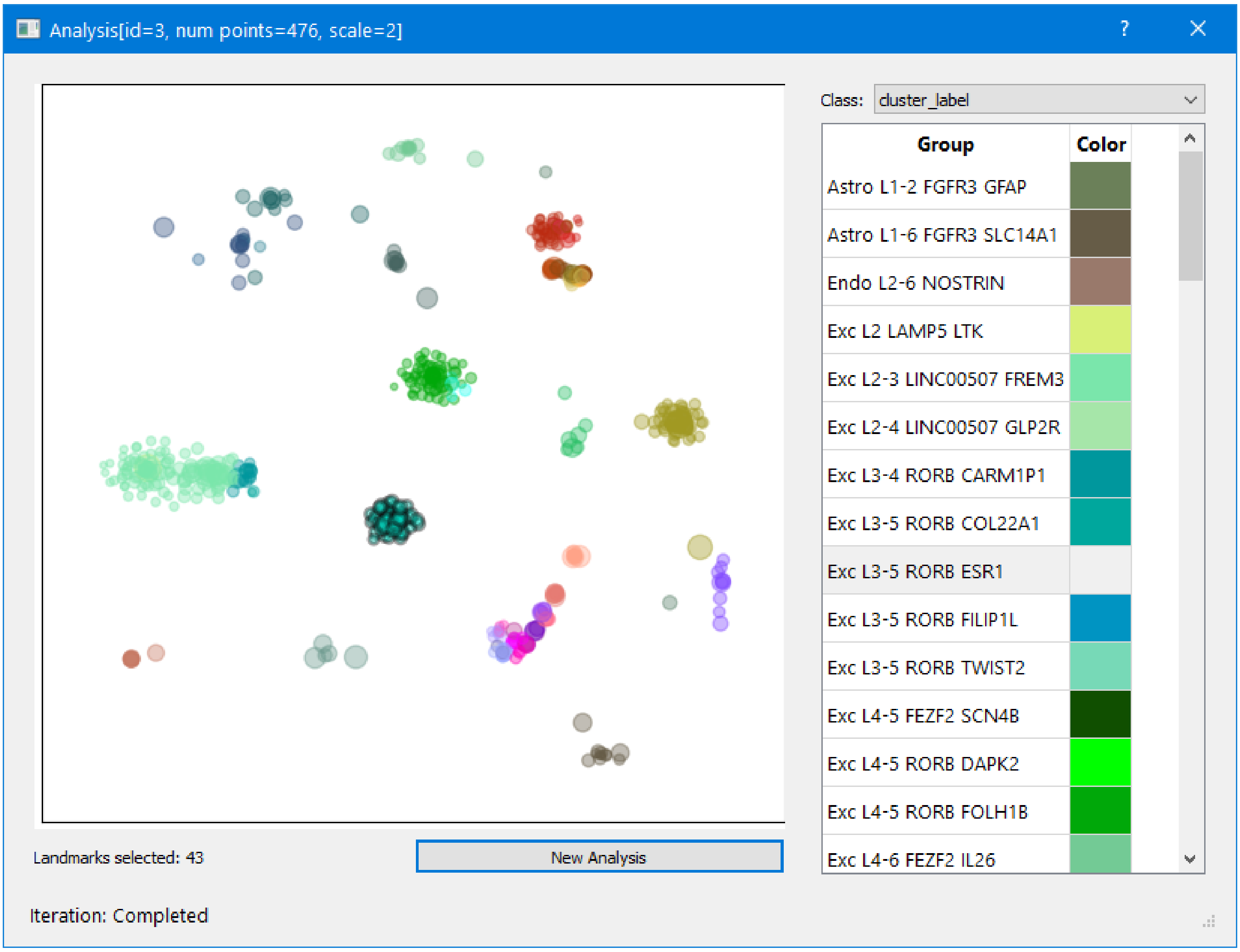
Select image dimension: 0

Iteration: Completed

10. Before loading new data restart the program.

11. Load the **MTG cell** data*

12. The analysis immediately reveals multiple well separate clusters some of which have a clear link with cluster labels (select the cluster label in the table - in this case the Exc L3-5 RORB ESR1 cluster label strongle associated with the group below center)



Analysis[id=3, num points=476, scale=2]

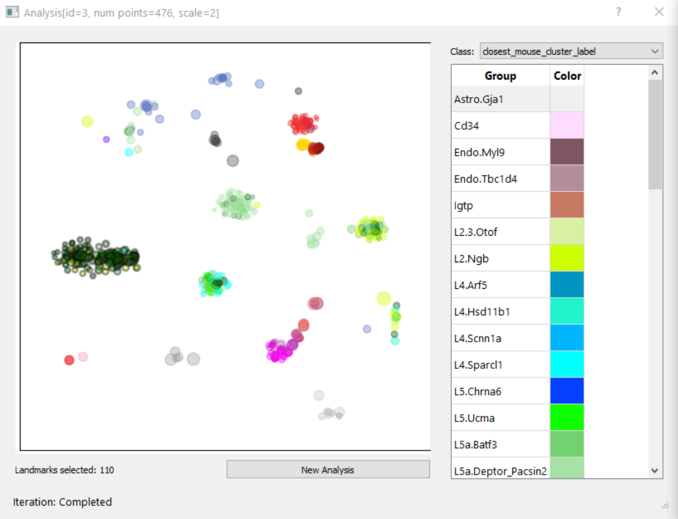
Class: cluster_label

Group	Color
Astro L1-2 FGFR3 GFAP	■
Astro L1-6 FGFR3 SLC14A1	■
Endo L2-6 NOSTRIN	■
Exc L2 LAMP5 LTK	■
Exc L2-3 LINC00507 FREM3	■
Exc L2-4 LINC00507 GLP2R	■
Exc L3-4 RORB CARM1P1	■
Exc L3-5 RORB COL22A1	■
Exc L3-5 RORB ESR1	
Exc L3-5 RORB FILIP1L	■
Exc L3-5 RORB TWIST2	■
Exc L4-5 FEZF2 SCN4B	■
Exc L4-5 RORB DAPK2	■
Exc L4-5 RORB FOLH1B	■
Exc L4-6 FEZF2 IL26	■

Landmarks selected: 43

Iteration: Completed

13. Selection of othe label groubs and sub analyses are possible



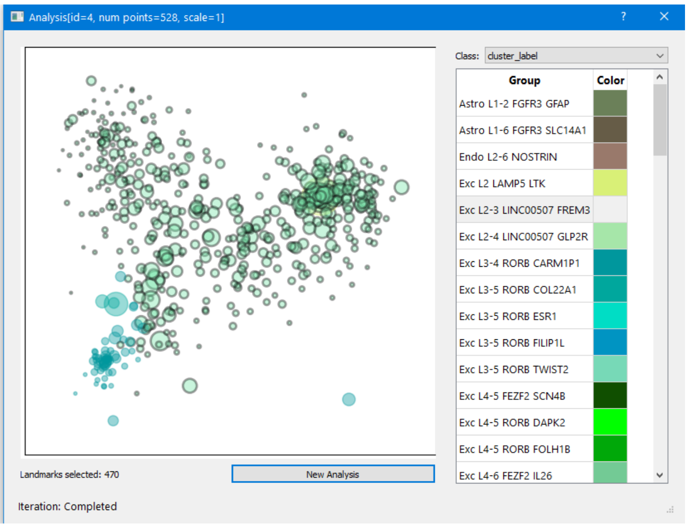
Analysis[id=3, num points=476, scale=2]

Class: closest_mouse_cluster_label

Group	Color
Astro.Gja1	
Cd34	■
Endo.Myl9	■
Endo.Tbc1d4	■
Igtp	■
L2.3.Otof	■
L2.Ngb	■
L4.Arf5	■
L4.Hsd11b1	■
L4.Scnn1a	■
L4.Sparcl1	■
L5.Chrna6	■
L5.Ucma	■
L5a.Batf3	■
L5a.Deptor_Pacsin2	■

Landmarks selected: 110

Iteration: Completed



Analysis[id=4, num points=528, scale=1]

Class: cluster_label

Group	Color
Astro L1-2 FGFR3 GFAP	■
Astro L1-6 FGFR3 SLC14A1	■
Endo L2-6 NOSTRIN	■
Exc L2 LAMP5 LTK	■
Exc L2-3 LINC00507 FREM3	
Exc L2-4 LINC00507 GLP2R	■
Exc L3-4 RORB CARM1P1	■
Exc L3-5 RORB COL22A1	■
Exc L3-5 RORB ESR1	■
Exc L3-5 RORB FILIP1L	■
Exc L3-5 RORB TWIST2	■
Exc L4-5 FEZF2 SCN4B	■
Exc L4-5 RORB DAPK2	■
Exc L4-5 RORB FOLH1B	■
Exc L4-6 FEZF2 IL26	■

Landmarks selected: 470

Iteration: Completed

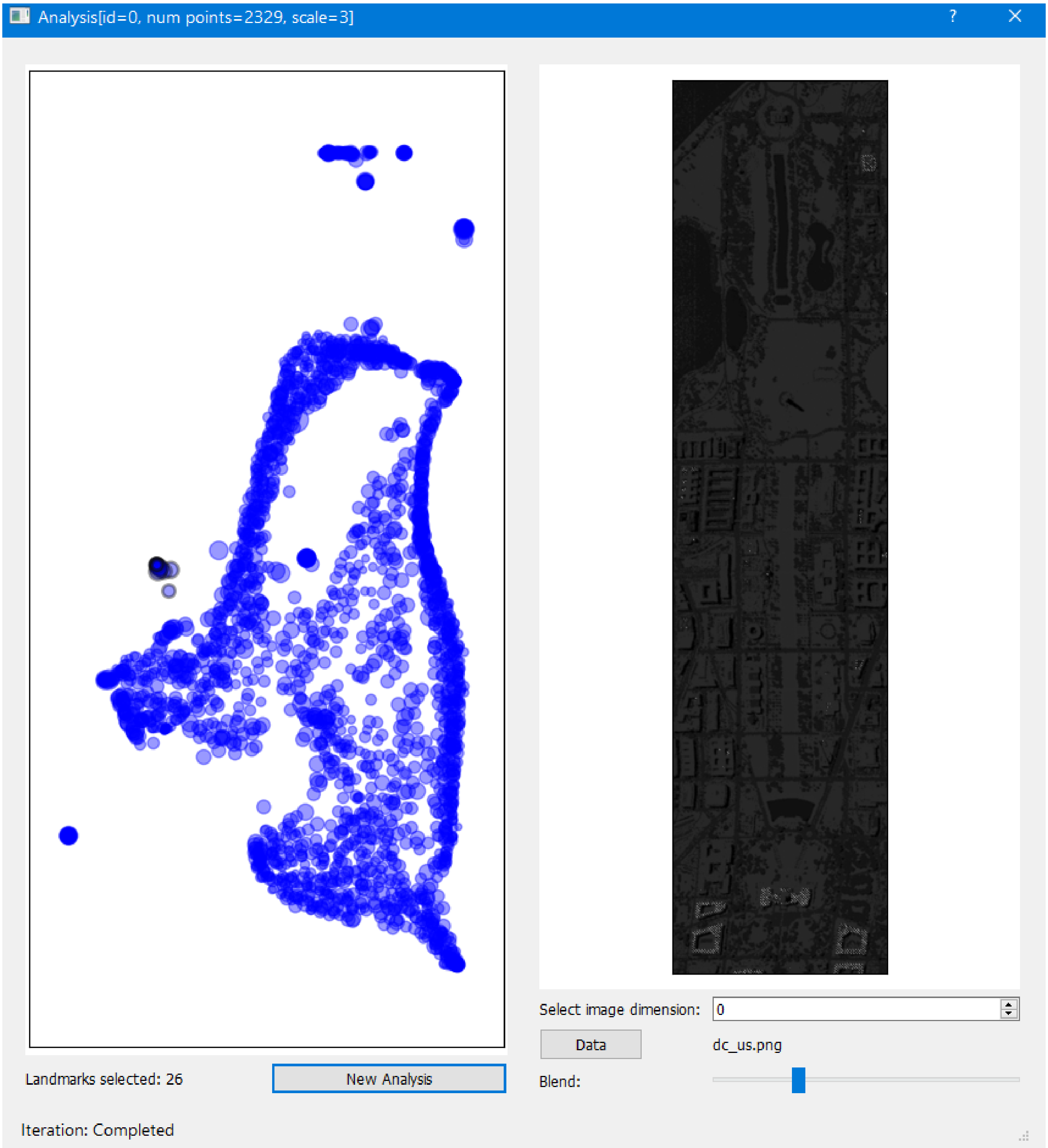
14. Before loading new data restart the program.

15. Load the **DC Mall hyperspectral** data

16. Load the dc_us.png image as background and select a blend level.

17. Moving the cursor over the top level of the HSNE reveals a number of interesting clusters

18. Copper roofs: including the US Capitol (bottom center) and the National Academy of Sciences building (top right)



Analysis[id=0, num points=2329, scale=3]

Select image dimension: 0

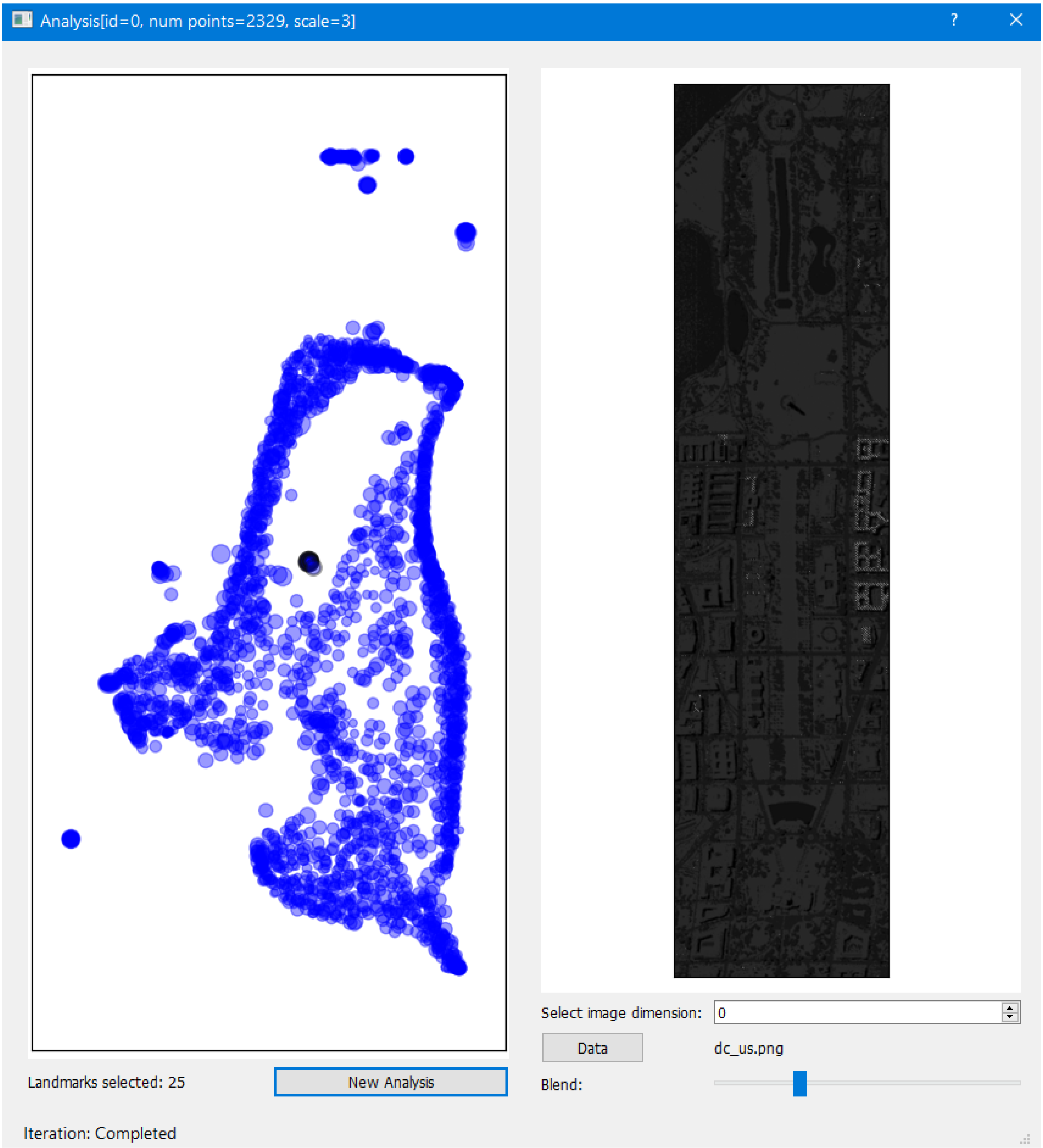
Data dc_us.png

Blend:

Landmarks selected: 26

Iteration: Completed

19. Terracotta tile roofs: including the Herbert C. Hoover building, the US EPA building (right hand side half-way down).


Analysis[id=0, num points=2329, scale=3]

Landmarks selected: 25

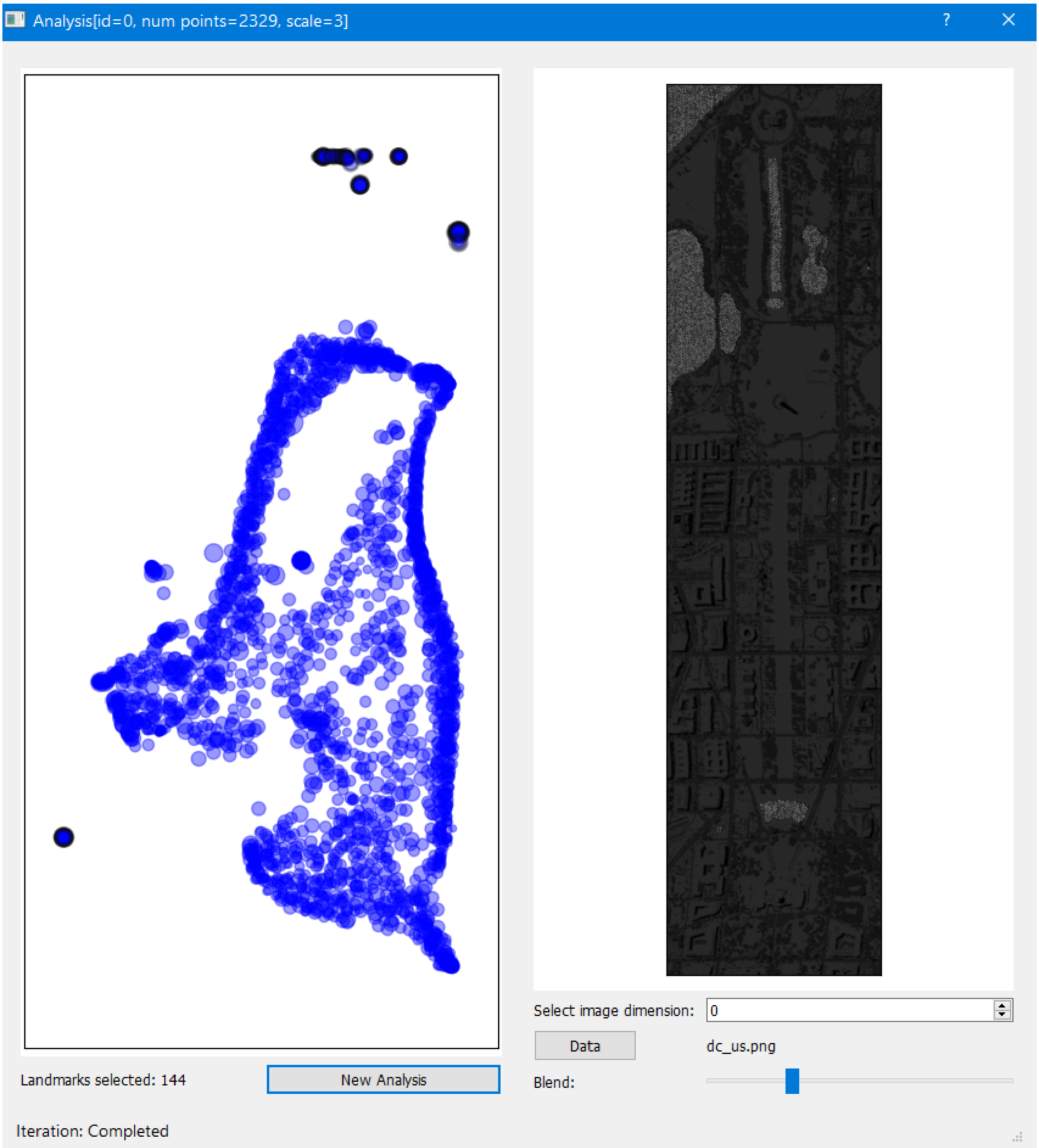
Select image dimension: 0

dc_us.png

Blend:

Iteration: Completed

20. Water features. The Lincoln Memorial Reflecting Pool, Potomac River, Tidal inlets and other water features mostly form distinct clusters. All the clusters have been selected here using the add selection feature.



Analysis[id=0, num points=2329, scale=3]

Landmarks selected: 144

Iteration: Completed

Select image dimension: 0

Data dc_us.png

21. Use the blend slider to increase/decrease background visibility to help locate the highlighted areas.

Summary pre-packaged data

The demo come with 3 types of pre-packaged data

Demo name	Data type	Description
MNIST 70000 digits	Image is data point	MNIST style data
MTG cell data	Point and metadata	Multidimensional points with associated metadata
Hyperspectral sun (512x512)	Hyperspectral image	Each pixel has multiple values
DC Mall (hyperspectral)	Hyperspectral image	Each pixel has multiple values

Using your own data

You can load your own data files if they correspond to one of the supported data types. In detail these types are as follows:

Data types

1. **Image is a data point:**
   Single data file

- npy file - shape = (<number_of_images>, <number_of_pixels>)

2. **Point and metadata**
   Two data files

- npy file for point data - shape = (<number_of_points>, <number_of_dimensions>)
- csv file containing an id followed by pairs of label + color columns

```
id   label0 color0   label1 color1   ---   labelN colorN
.    ...    ...      ...    ...      ---   ...    ...
.    ...    ...      ...    ...      ---   ...    ...
```

3. **Hyper-spectral image**
   Single data file

- npy file - shape = (<number_of_images>, <number_of_pixels>)

Data conversion

For the DC Mall data an offset was applied to the original tiff values to make everything positive and the result was converted to uint32. The 'us' in the filename stands for unsigned.

Loading a precalculated .hsne file

A pre-calculated HSNE model can be saved and read from a file with the **.hsne** extension.

Architecture

For the benefit of developers a short summary of the demo architecture is given here.

There are two main GUI elements:

1. The ModelGui displaying the load controls and the hierarchy of analyses
2. The AnalysisController, one or more dialogs containing the interactive embedding plot widget for an analysis and an associated viewer widget for image or meta-data.

Once data is loaded an initial AnalysisController dialog is displayed for the top level analysis containing all the top scale points.

Software Components

Component	Type	Function
ModelController	Control	Creates the main Qt app and coordinates the creation of an *HSNE* analysis and the navigation of the *nptsne.hsne_analysis.AnalysisModel* and the creation of new *nptsne.hsne_analysis.Analysis* instances. Starts the *ModelGui*
ModelGui	GUI	Displays the hierarchy of *nptsne.hsne_analysis.Analysis* instances in the *AnalysisModel*. Permit deletion of analyses.
AnalysisController	Control & GUI	Controls creation of a new *nptsne.hsne_analysis.Analysis* based on a user selection. Creates an embedding for the *nptsne.hsne_analysis.Analysis*. Starts the *EmbeddingGui* to display the *nptsne.hsne_analysis.Analysis*. Next to the *EmbeddingGui* the appropriate (depending on data type) Viewer GUI is displayed
EmbeddingGui	Widget	Display an interactive scatter plot for the *nptsne.hsne_analysis.Analysis* embedding. Supports multiple selection options: Select brush: *Rectangle:* **R**, *Ellipse:* **E**, *Lasso:* **L**, *Polygon:* **P** *Clear* **<space>** Select function: *Select All:* **A**, *Remove Selection:* **D**, *Invert Selection:* **I** Select modifiers: *Add* **<shift>+<key>**, *Subtract:* **<ctrl>+<key>**
CompositeImageViewer	Widget	View an image based on a user selection in the *EmbeddingGui* where the data represents one image per point (e.g. MNIST)
HyperSpectralImageViewer	Widget	View a hyperspectral image based on selections in the *EmbeddingGui*. For Hyperspectral data.
MetaDataViewer	Widget	Table view of meta data base on user selection in the *EmbeddingGUI*
ConfigClasses	Data	Data containers for application configuration
DemoConfig	Data	The config for this application

Data sources

MNIST Data - retrieved from mnist-orignal.mat (https://github.com/amplab/datascience-sp14/raw/master/lab7/mldata/mnist-original.mat) a Matlab format data file derived from the original Yann LeCun MNIST data base at http://yann.lecun.com/exdb/mnist/

Hyperspectral Solar Images - downloaded from the Solar Dynamics Observatory as in the original Hierarchical Stochastic Neighbor Embedding paper.

MTG cell data - 15603 cell samples from the human middle temporal gyrus with single-nucleus transcriptomes. Data is derived from the complete set at Allen Institute for Brain Science

DC Mall data - downloaded from the Purdue University MultiSpec(c) page. The original 191 band hyperspectral data was created by the Spectral Information Technology Application Center of Virginia.