Napari-clusters-plotter plugin

Marcelo Leomil Zoccoler

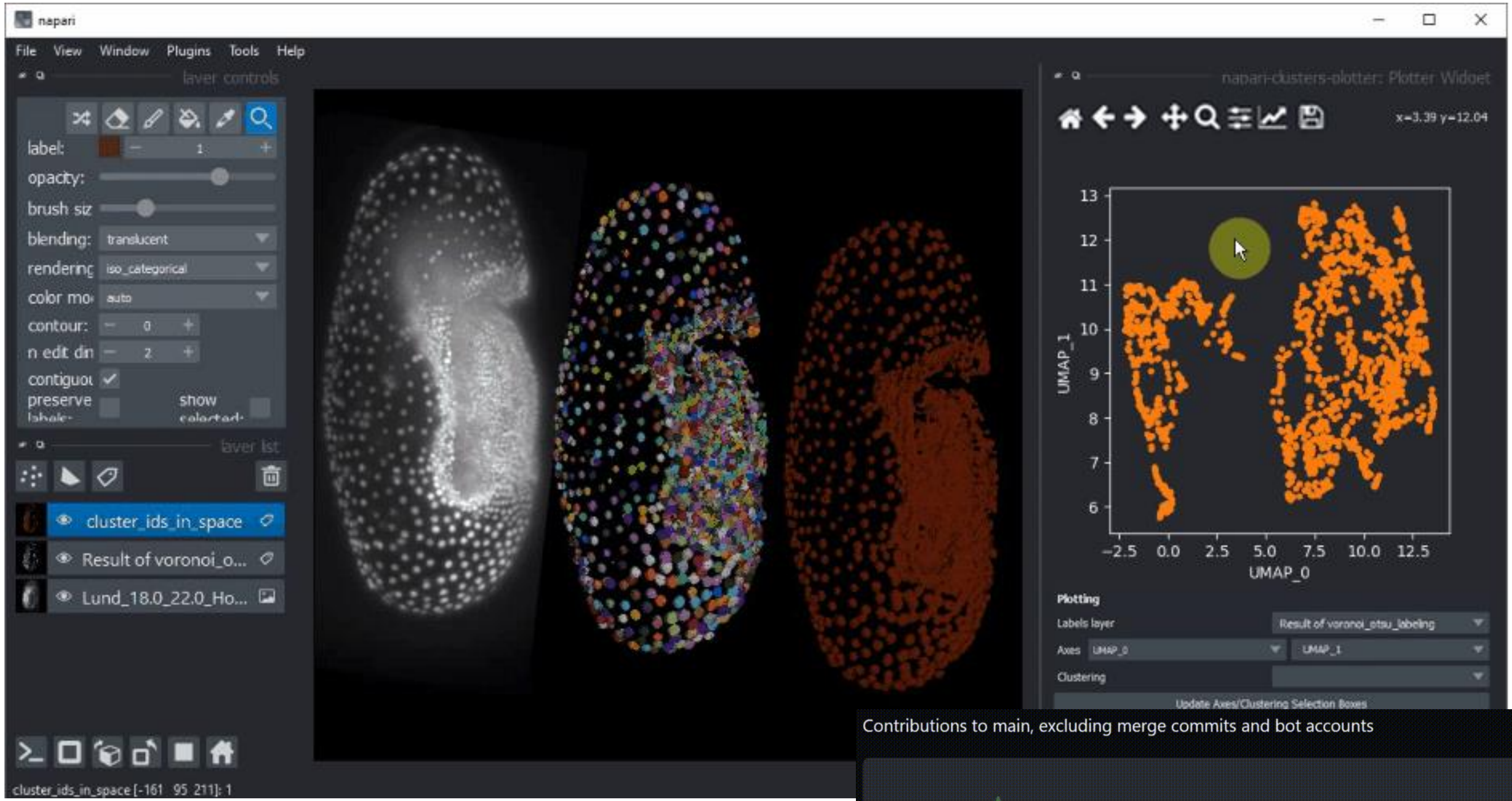
Reusing material from

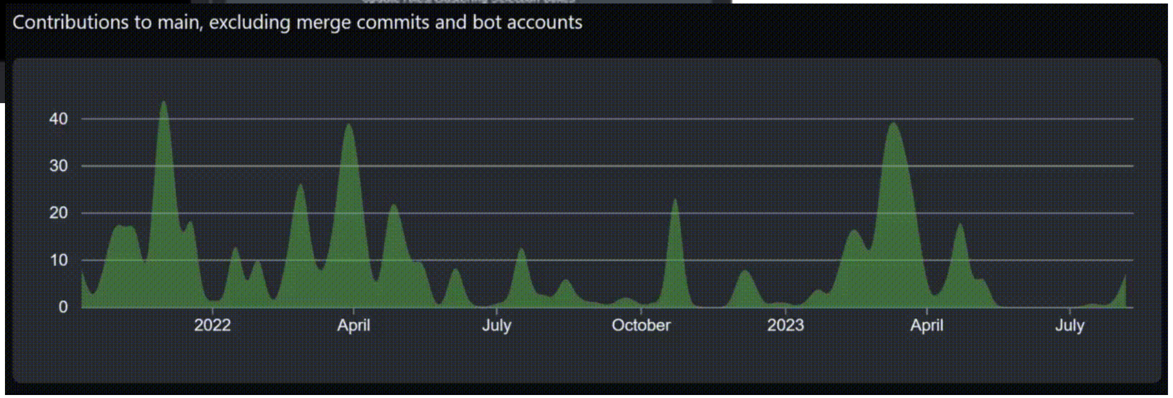
Robert Haase, Johannes Soltwedel, Till Korten,
Ryan Savill, Laura Zigutyte and Mara Lampert, PoL TU Dresden

@zoccolermarcelo

August 2023

Napari-clusters-plotter plugin





For today’s demo, please use this environment:

mamba create -n napari-clusters-plotter-env python=3.10 napari=0.4.17 devbio-napari

mamba activate napari-clusters-plotter-env

Quantitative Bio-Image Analysis with Visualization

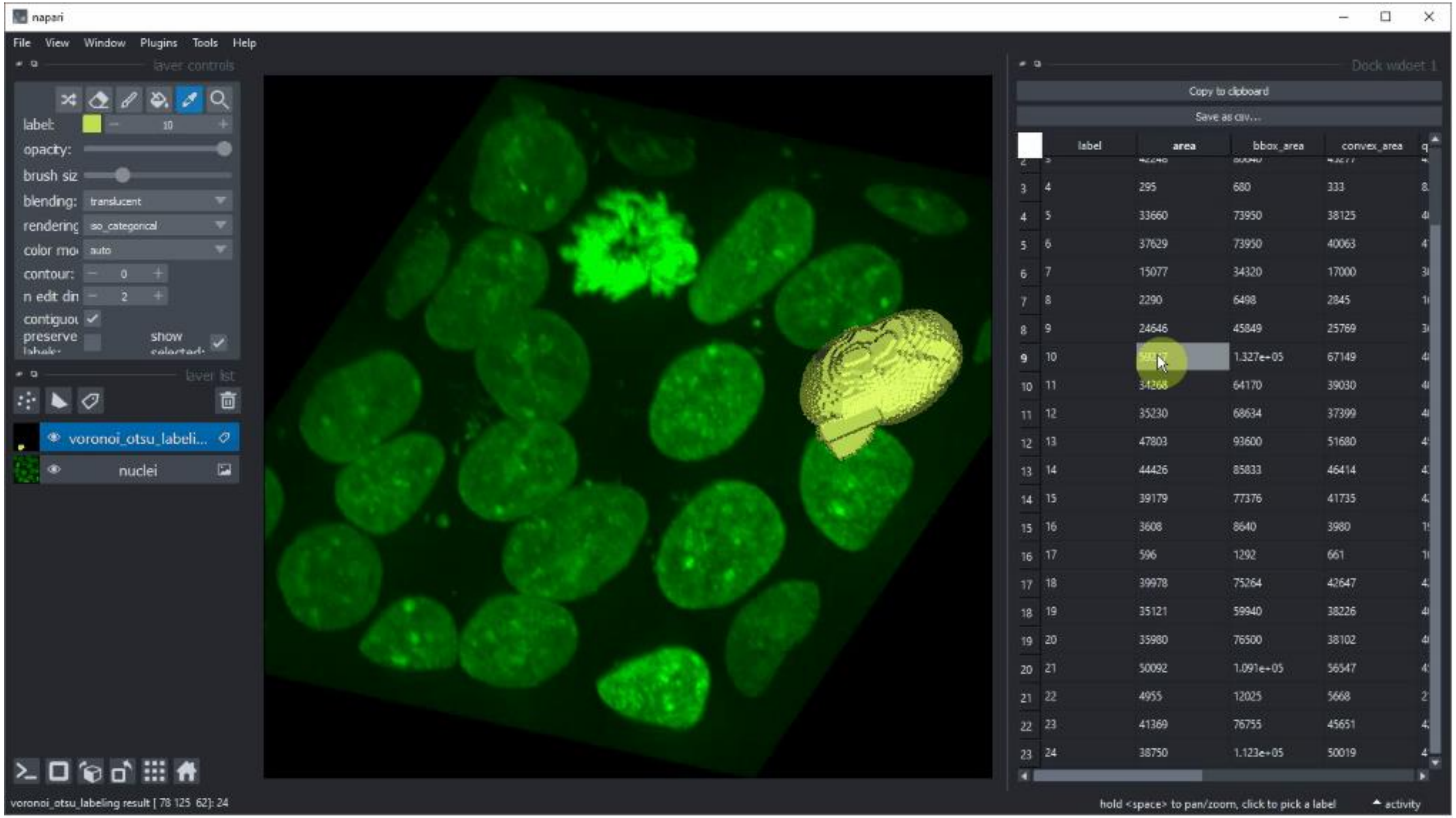
@zoccolermarcelo

August 2023

Quantitative bio-image analysis

PoL
Physics of Life
TU Dresden

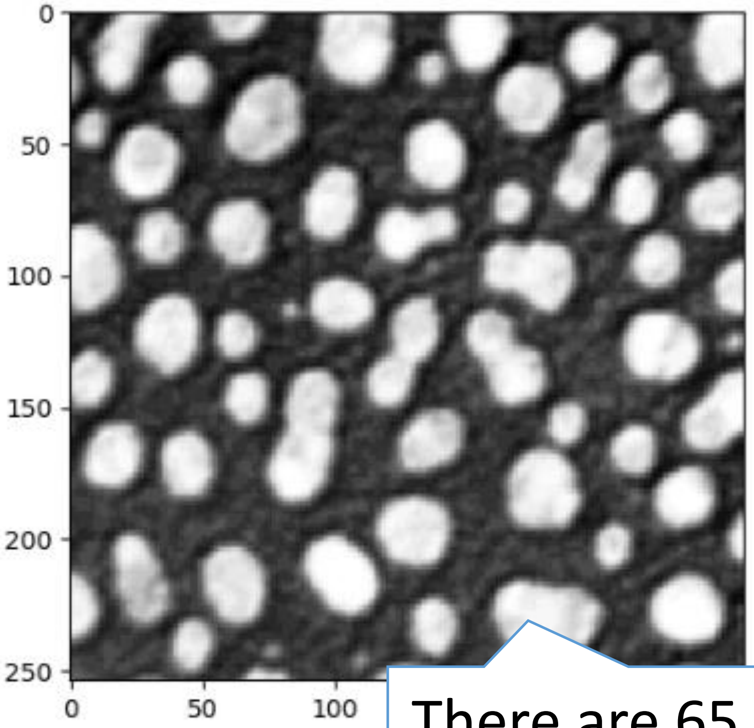
- Deriving quantitative information from images of biological samples taken with microscopes + visualization



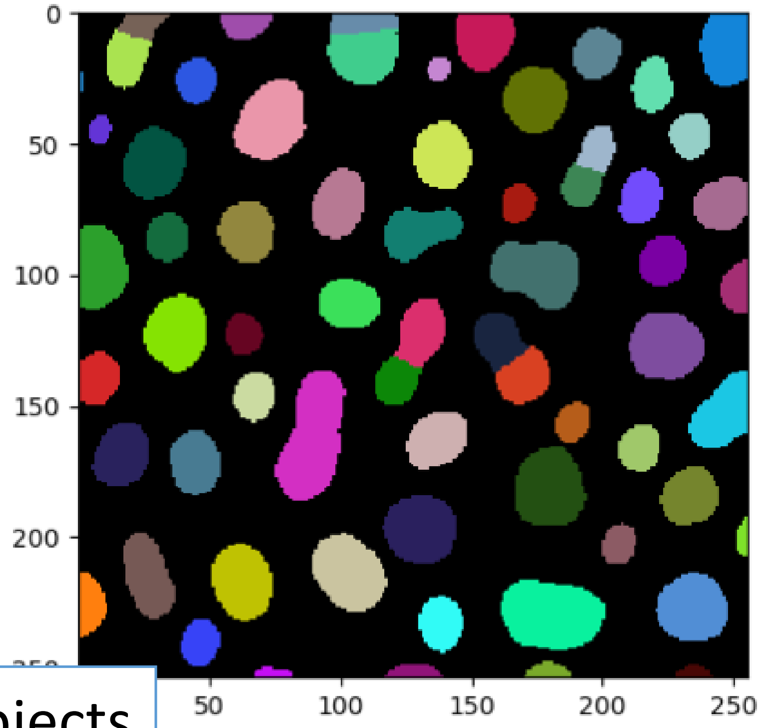
Reliable bio-image analysis

- Algorithms must be reliable (trustworthy).
- Visualization helps gaining trust in automated methods.

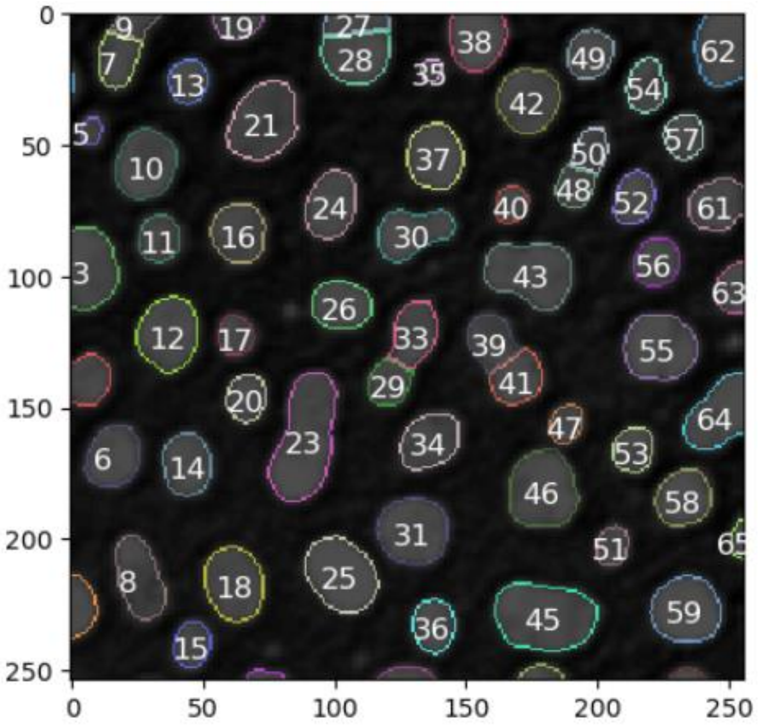
Original image



Label image



Overlay

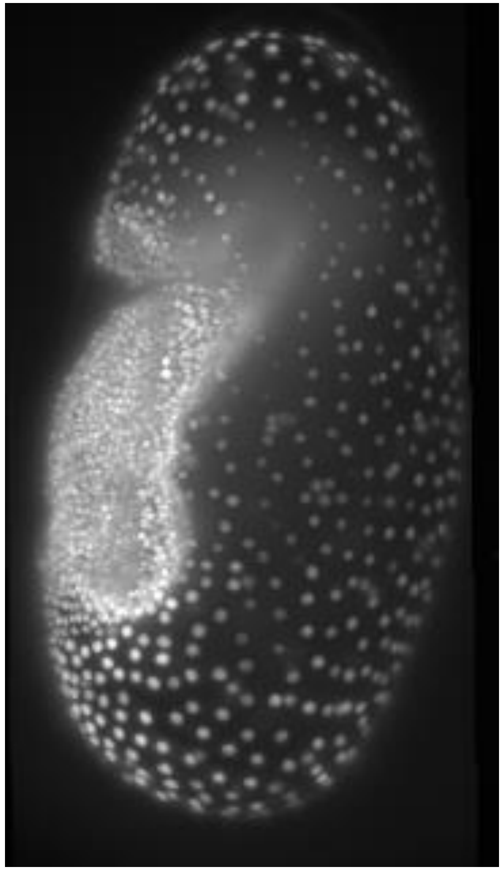


There are 65 objects in this image.

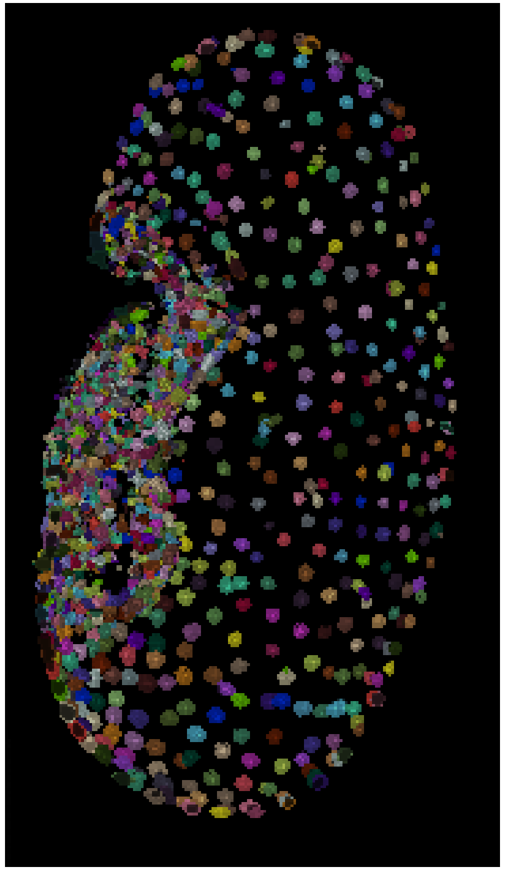
Source: M. Zoccoler & R. Haase licensed CC-BY
https://haesleinhuepf.github.io/BioImageAnalysisNotebooks/60_data_visualization/overlay_text_on_image.html

Object classification

- Goal: Separate objects
  - According to an expert's annotation (supervised)
  - According to separability of object properties (unsupervised)

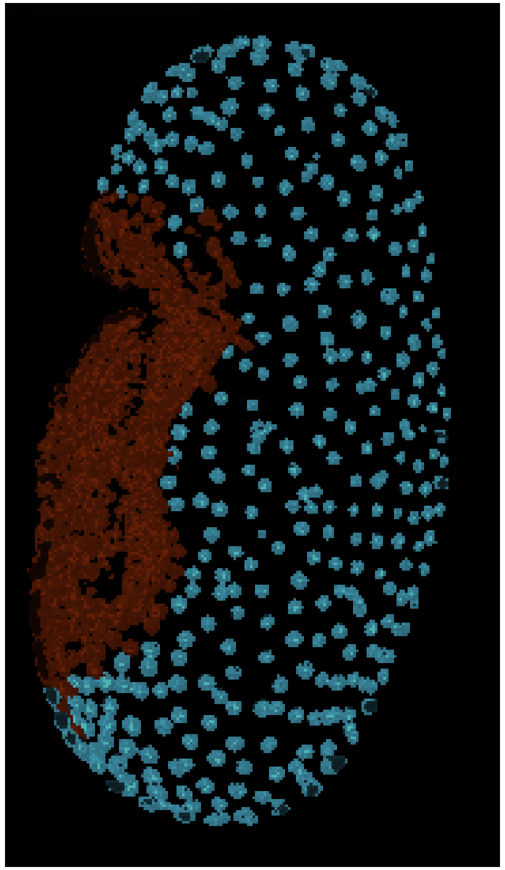


Raw data



Instance segmentation

Object classification



Semantic segmentation

Image Data Source: Daniela Vorkel, Myers lab, MPI-CBG/CSBD Dresden

Unsupervised Machine Learning for Object Characterization

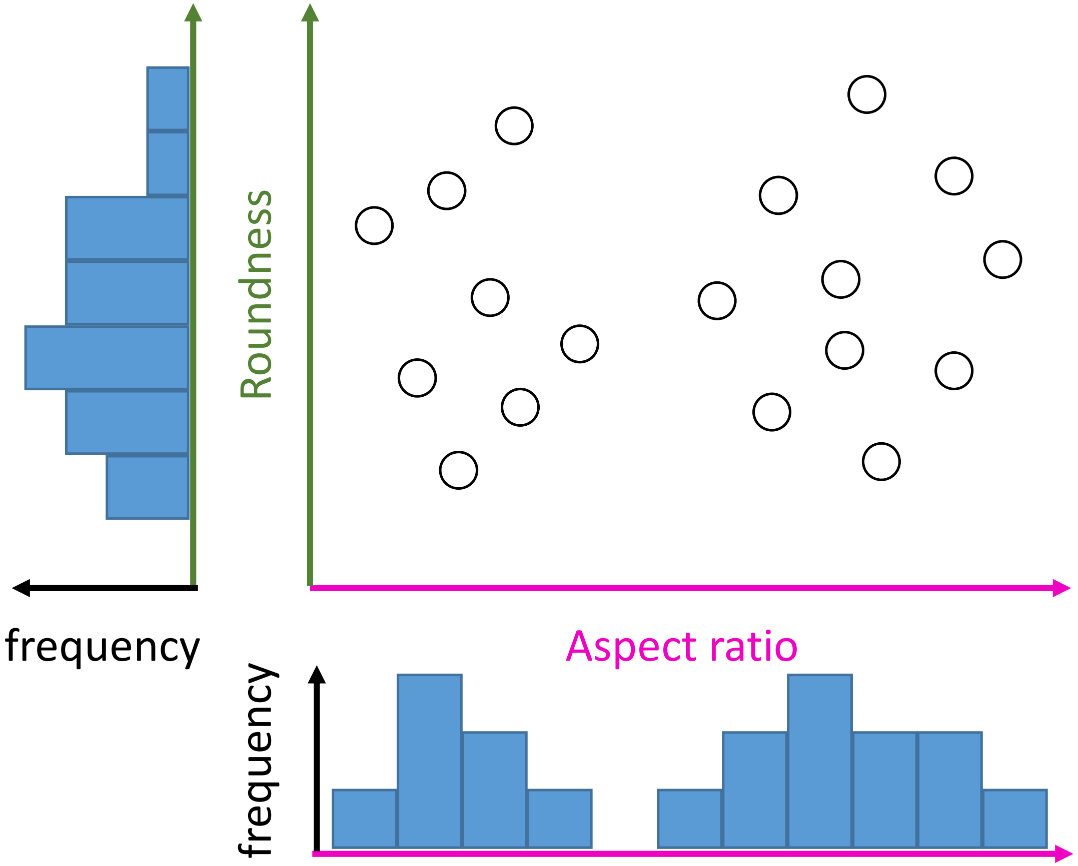
Reusing material from
Robert Haase, Johannes Soltwedel and Till Korten, PoL, TU Dresden and
OpenStreetMap foundation

@zoccolermarcelo

August 2023

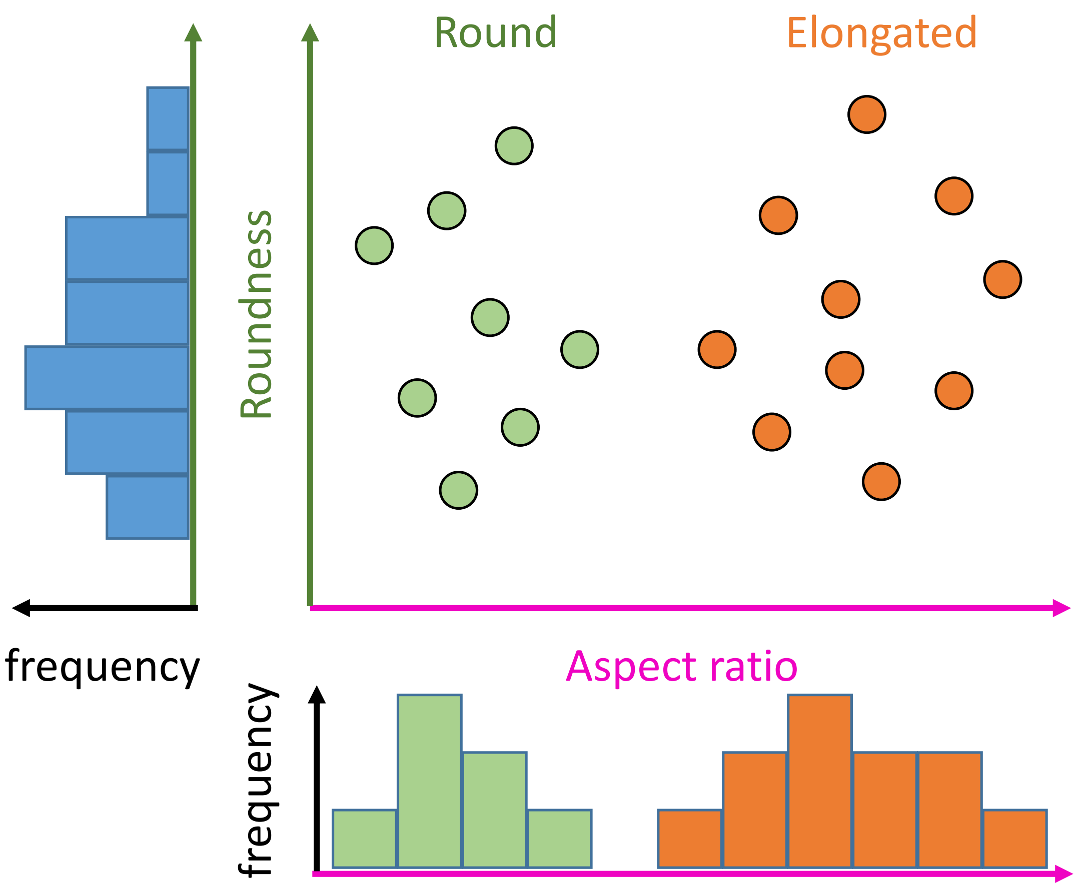
Unsupervised machine learning

- If you don't provide ground truth, the algorithm is *unsupervised*.



Unsupervised machine learning

- If you don't provide ground truth, the algorithm is unsupervised.
- Nevertheless, algorithms can tell us something about the data



Hypothesis-driven quantitative biology

- Hypothesis: Cell shape can be influenced by modifying X.
- Null-Hypothesis: Circularity of modified cells is similar to cells in the control group.

- Sample preparation
- Imaging
- Cell segmentation
- Circularity measurement
- Statistics

Shall we use a different microscope?

Should we use a different segmentation algorithm?

Is circularity the right parameter to measure?

Hypothesis *generating* quantitative biology

PoL
Physics of Life
TU Dresden

- ~~Hypothesis: Cell shape can be influenced by modifying X.~~
- Question: Which image-derived parameter is influenced when modifying X?
  - Sample preparation
  - Imaging
  - Cell segmentation algorithm A, algorithm B, algorithm C
  - Measurement of circularity, solidity, elongation, extend, texture, intensity, topology …
  - Statistics

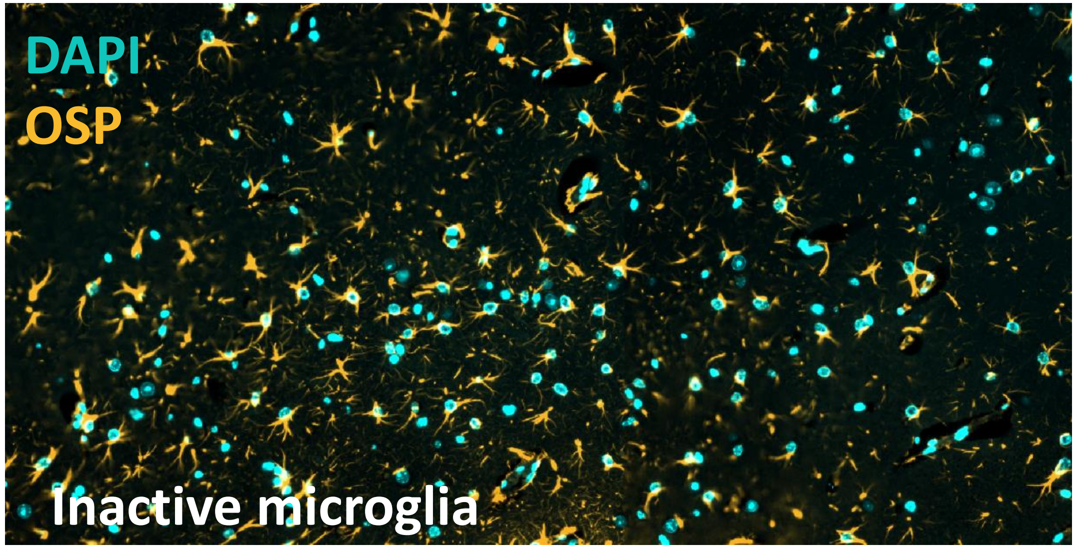
Which segmentation algorithms allow measurements that show a relationship with X?

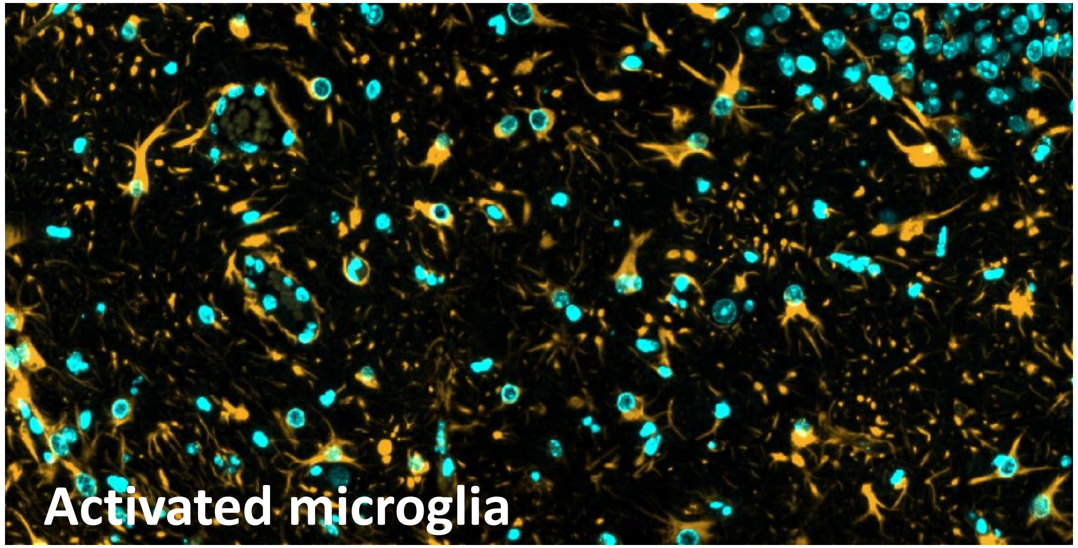
Why?

Which parameter shows any relationship with X?

Identifying features to measure

Example: Inactive vs. activated microglia in mouse brain





Challenge: Which of these features reflect biological propeties?

	label	area	bbox_area	convex_area	quivalent_diamete	max_intensity	mean_intensity	min_intensity	solidity	extent	eret_diameter_ma:	local_centroid-0	l
1	1	3379	13949	5120	18.61786412639...	613.0	345.6717963894...	259.0	0.6599609375	0....	37.3496987939662	15.77952056821...	18
2	2	2319	7448	3491	16.42230229224...	421.0	297.8434670116...	240.0	0....	0....	38.65229618017...	4....	17
3	3	2304	14415	4281	16.38681751812...	456.0	300.8298611111...	245.0	0....	0....	34.19064199455...	17.73828125	13
4	4	3278	13804	5139	18.43048549951...	467.0	316.1446003660...	249.0	0....	0....	34.84250278036...	15.52287980475...	10
5	5	1501	3315	1681	14.20563625190...	458.0	302.147235176549	236.0	0....	0....	17.97220075561...	6....	6.
6	6	2341	6061	2714	16.47407088948...	594.0	355.4446817599...	261.0	0....	0....	30.67572330035...	16.54250320375...	6.
7	7	1725	3584	1940	14.87979081163...	568.0	343.7866666666...	257.0	0....	0....	17.72004514666...	7.80463768115942	7.
8	8	1502	3840	1753	14.20879025650...	431.0	290.0659121171...	235.0	0....	0....	18.57417562100...	8....	6.

Source: Slice2Volume, https://rodare.hzdr.de/record/1849

Dimensionality Reduction

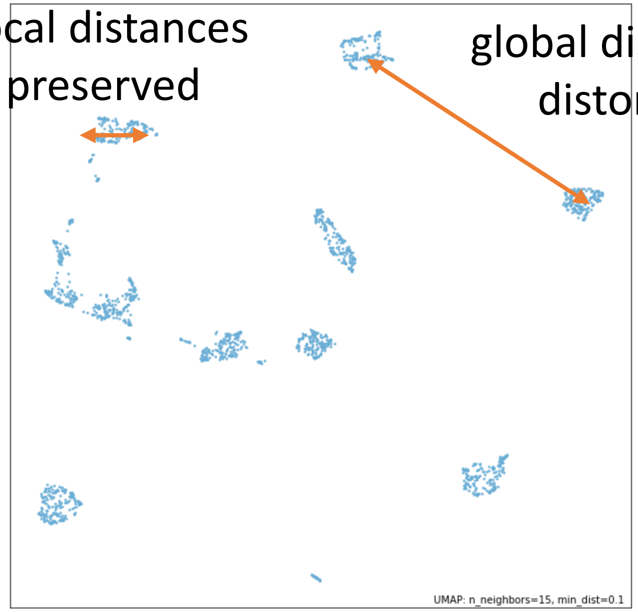
August 2023

Dimensionality reduction

- Challenge: Find a representation (embedding) of your data that represents the data in fewer dimensions
- Preserve local distances at the expense of global distortions

	label	area	bbox_area	convex_area	quivalent_diamete	max_intensity	mean_intensity	min_intensity	solidity	extent	eret_diameter_ma:	local_centroid-0	l
1	1	3379	13949	5120	18.61786412639...	613.0	345.6717963894...	259.0	0.6599609375	0....	37.3496987939662	15.77952056821...	18
2	2	2319	7448	3491	16.42230229224...	421.0	297.8434670116...	240.0	0....	0....	38.65229618017...	4....	17
3	3	2304	14415	4281	16.38681751812...	456.0	300.8298611111...	245.0	0....	0....	34.19064199455...	17.73828125	13
4	4	3278	13804	5139	18.43048549951...	467.0	316.1446003660...	249.0	0....	0....	34.84250278036...	15.52287980475...	10
5	5	1501	3315	1681	14.20563625190...	458.0	302.147235176549	236.0	0....	0....	17.97220075561...	6....	6.
6	6	2341	6061	2714	16.47407088948...	594.0	355.4446817599...	261.0	0....	0....	30.67572330035...	16.54250320375...	6.
7	7	1725	3584	1940	14.87979081163...	568.0	343.7866666666...	257.0	0....	0....	17.72004514666...	7.80463768115942	7.
8	8	1502	3840	1753	14.20879025650...	431.0	290.0659121171...	235.0	0....	0....	18.57417562100...	8....	6.
9	9	1602	4080	1894	14.51737058294...	475.0	297.8008739076...	241.0	0....	0....	18.70828693386...	8....	8.
10	10	1395	3600	1624	13.86304166283...	424.0	304.8494623655...	247.0	0....	0.3875	17.60681686165...	7....	7.
11	11	609	1100	697	10.51654029260...	323.0	274.2528735632...	241.0	0....	0....	13.45362404707...	3....	4.
12	12	1686	3757	1894	14.76679738567...	460.0	303.8303677342...	240.0	0....	0....	17.97220075561...	9....	7.
13	13	2157	5184	2531	16.03062694504...	576.0	339.990264255911	270.0	0....	0....	19.54482028569...	8....	8.
14	14	863	2340	1032	11.81237949737...	327.0	272.4449594438...	237.0	0....	0....	16.0312195418814	6....	5.

Many dimensions

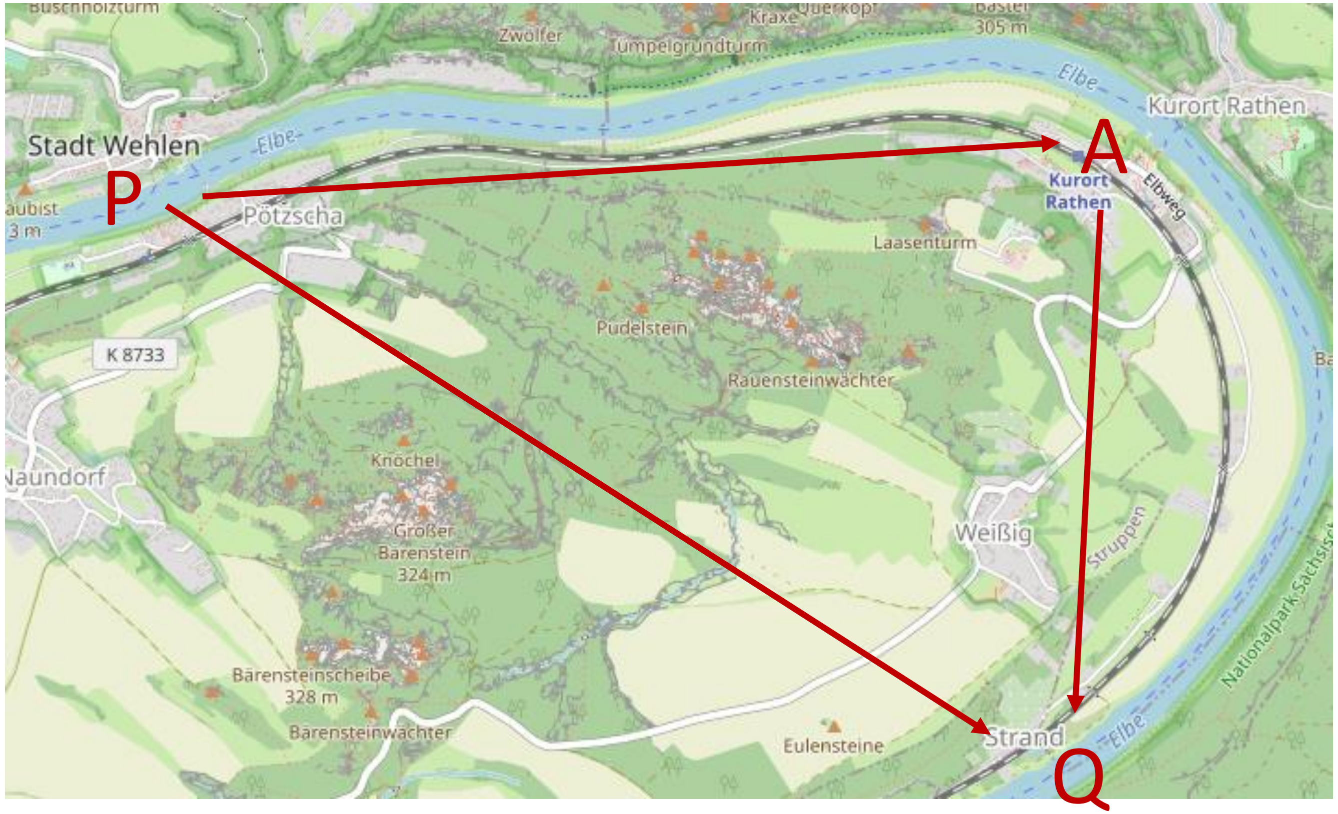


Few dimensions

https://umap-learn.readthedocs.io/en/latest/index.html

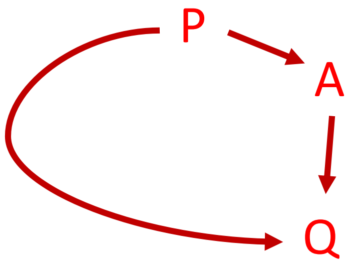
Dimensionality reduction

Identifying the right features might require deforming feature space.



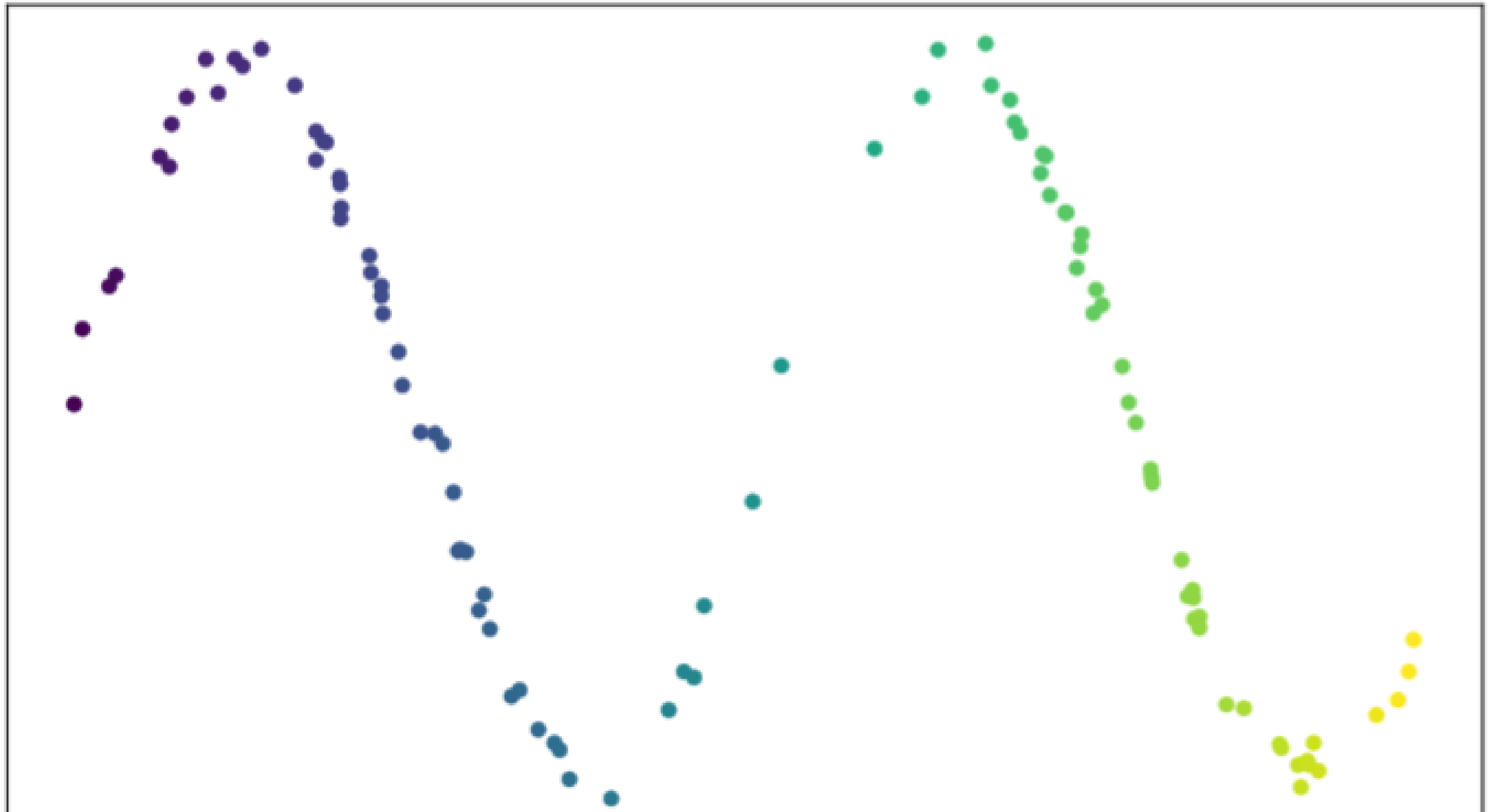
Example: Euclidean distance and travel time while hiking. Travelling from Wehlen to Strand might be faster if you take a detour through Rathen.

→ The data might be better represented like this:



Dimensionality reduction example

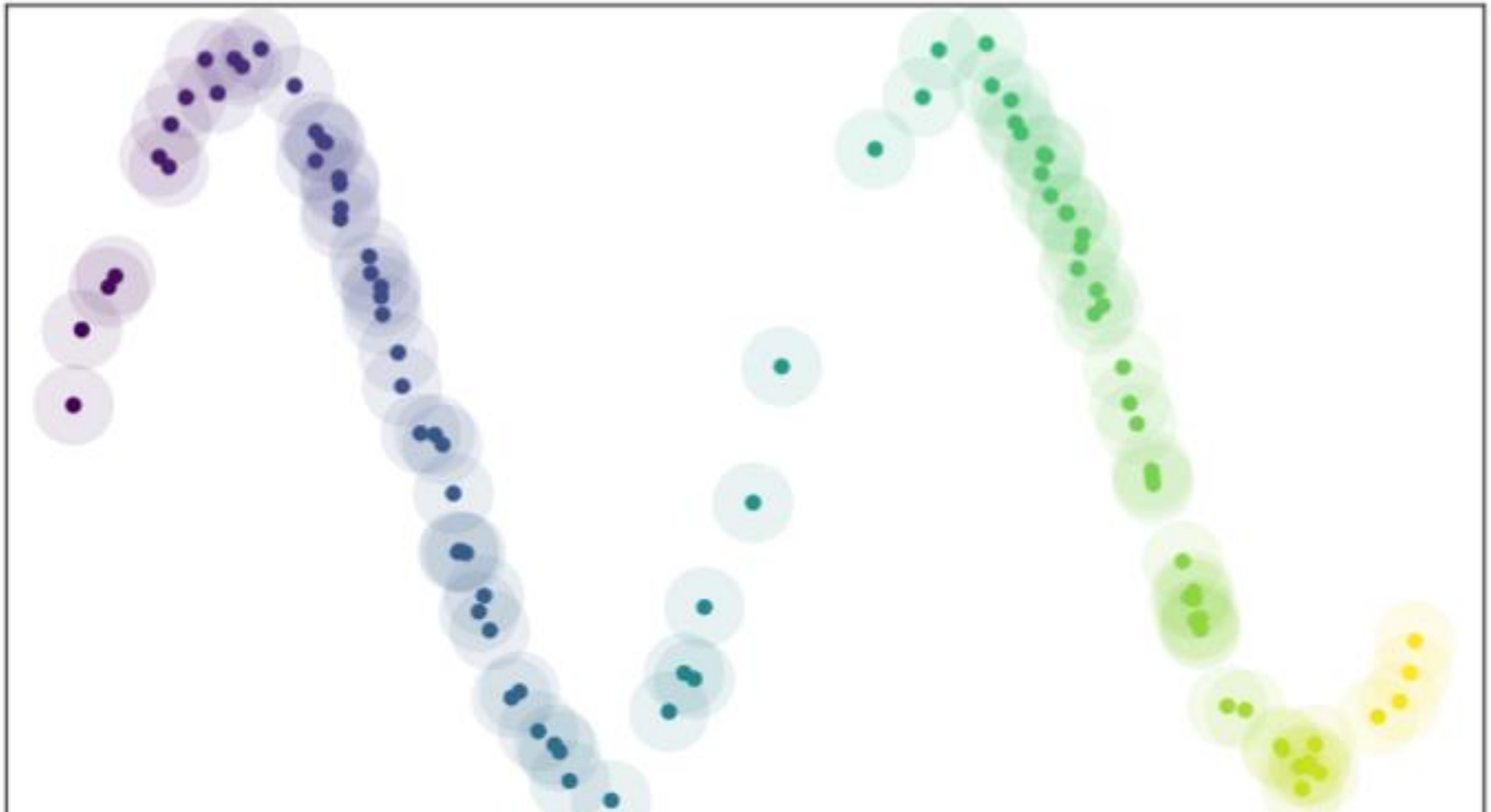
- Dimensionality reduction from 2D to 1D



Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

Dimensionality reduction example

- Dimensionality reduction from 2D to 1D

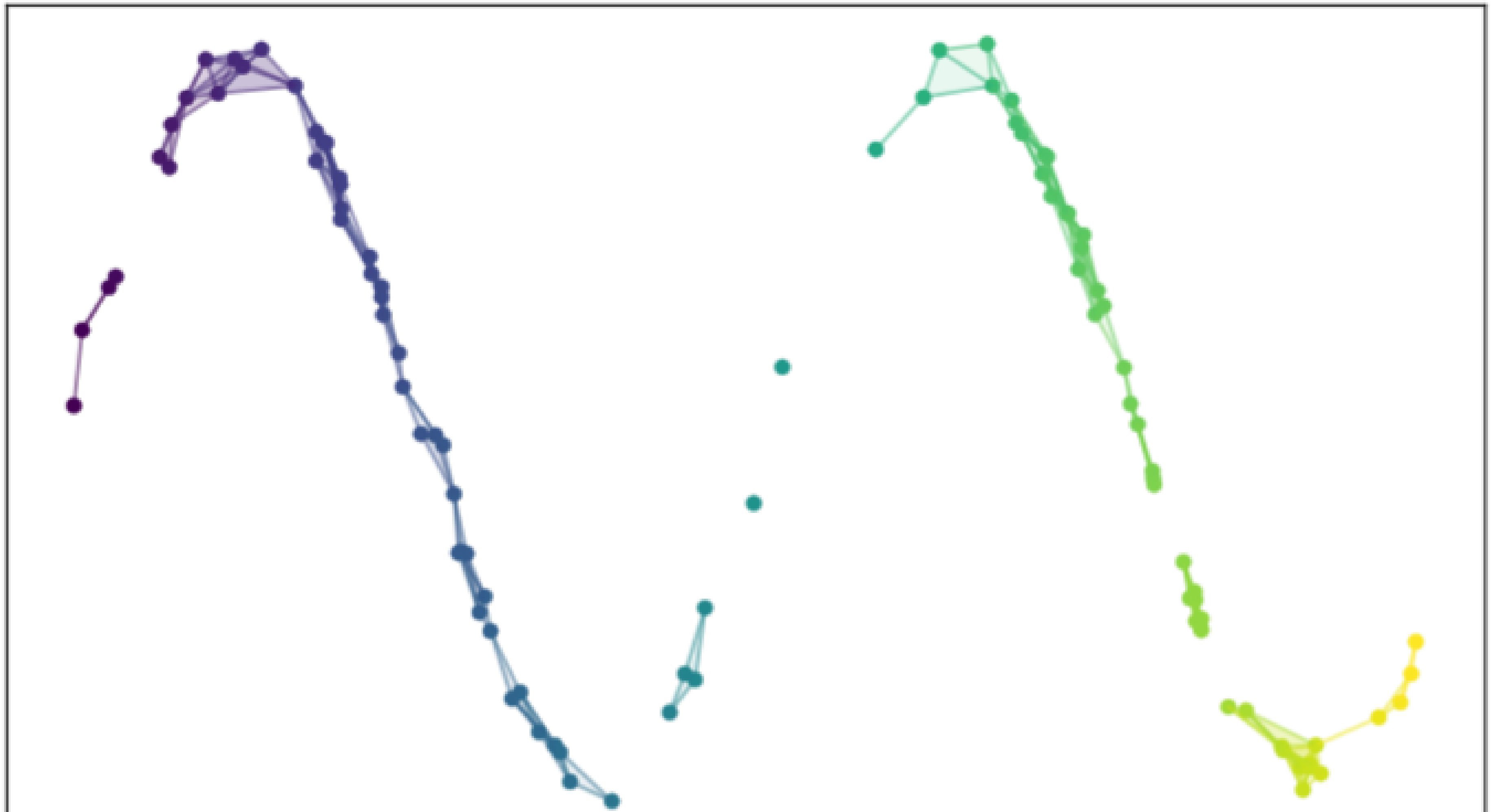


Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

Approach:
Points within a defined radius are considered neighbors

Dimensionality reduction example

- Dimensionality reduction from 2D to 1D



Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

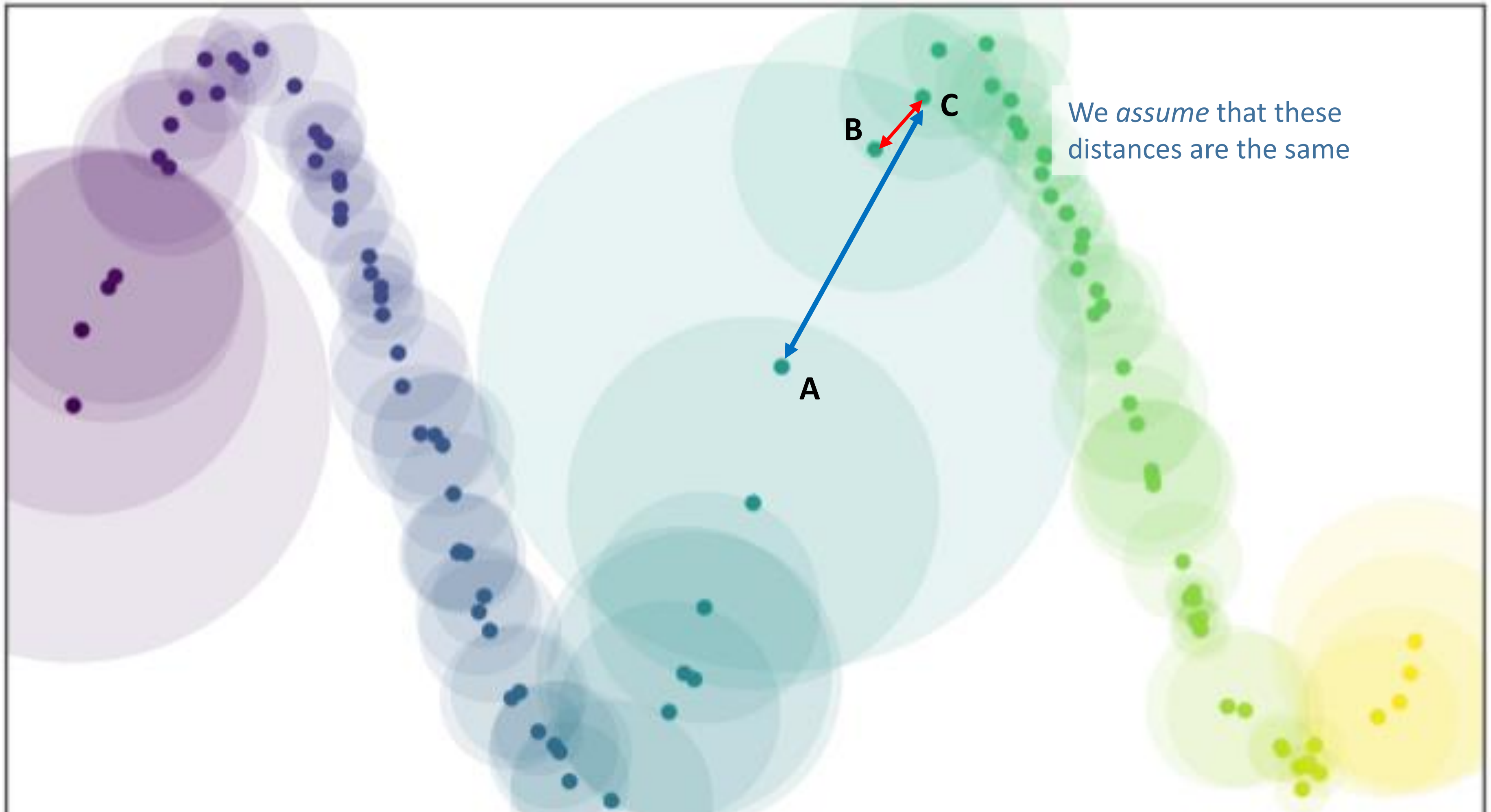
Result:
Neighborhood graph of close points

Problem:
The scarcity of the points leads to a disconnected neighborhood

→Searching neighbors by radius may not be viable

Dimensionality reduction example

- Dimensionality reduction from 2D to 1D



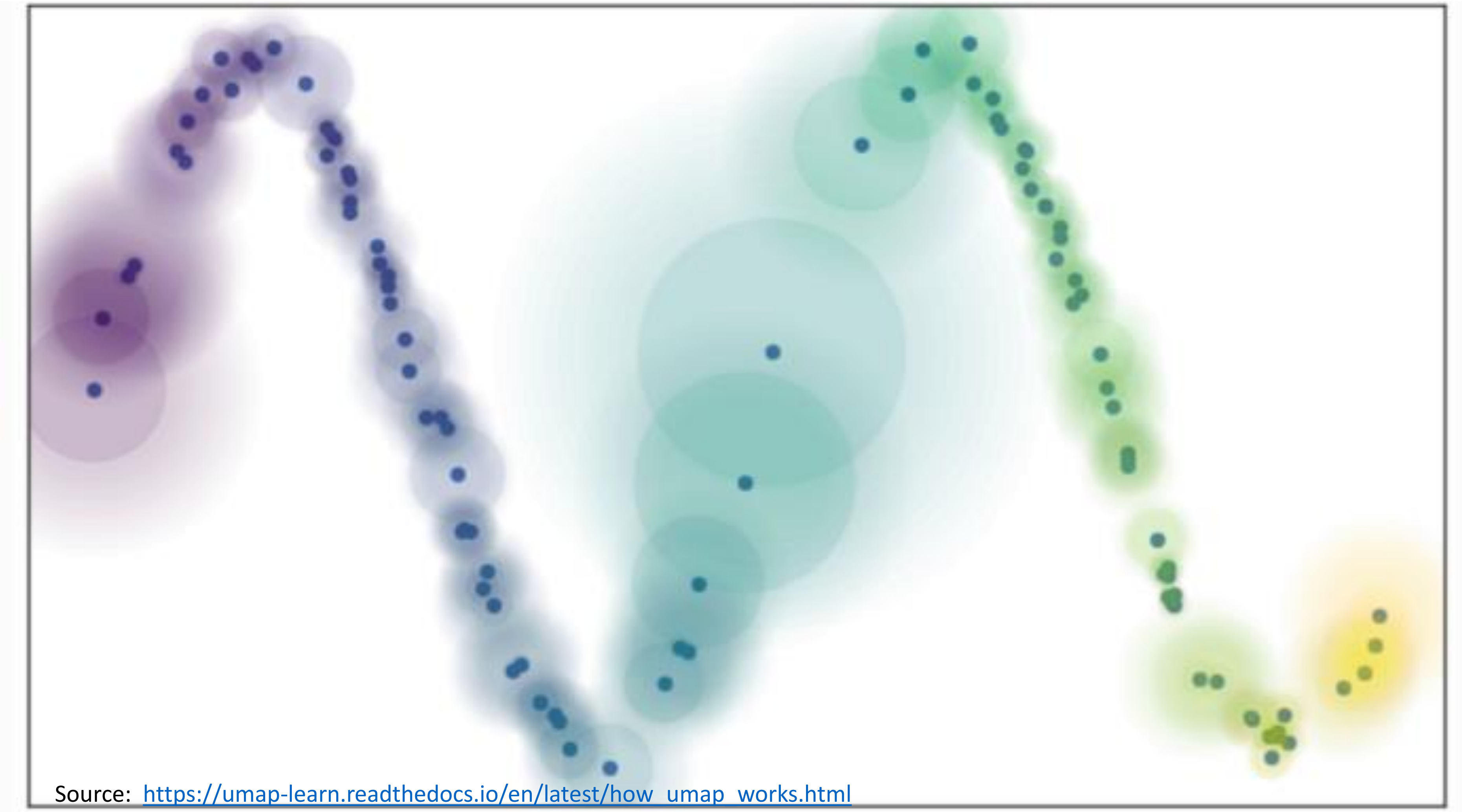
Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

Each point now has its own distance metric assigned to it
→ d(A, C) from A's point of view: 1
→ d(B,C) from C's point of view: 1

Result:
We choose a number of neighbors that each radius should cover rather than a fixed distance threshold

Dimensionality reduction example

- Dimensionality reduction from 2D to 1D



Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

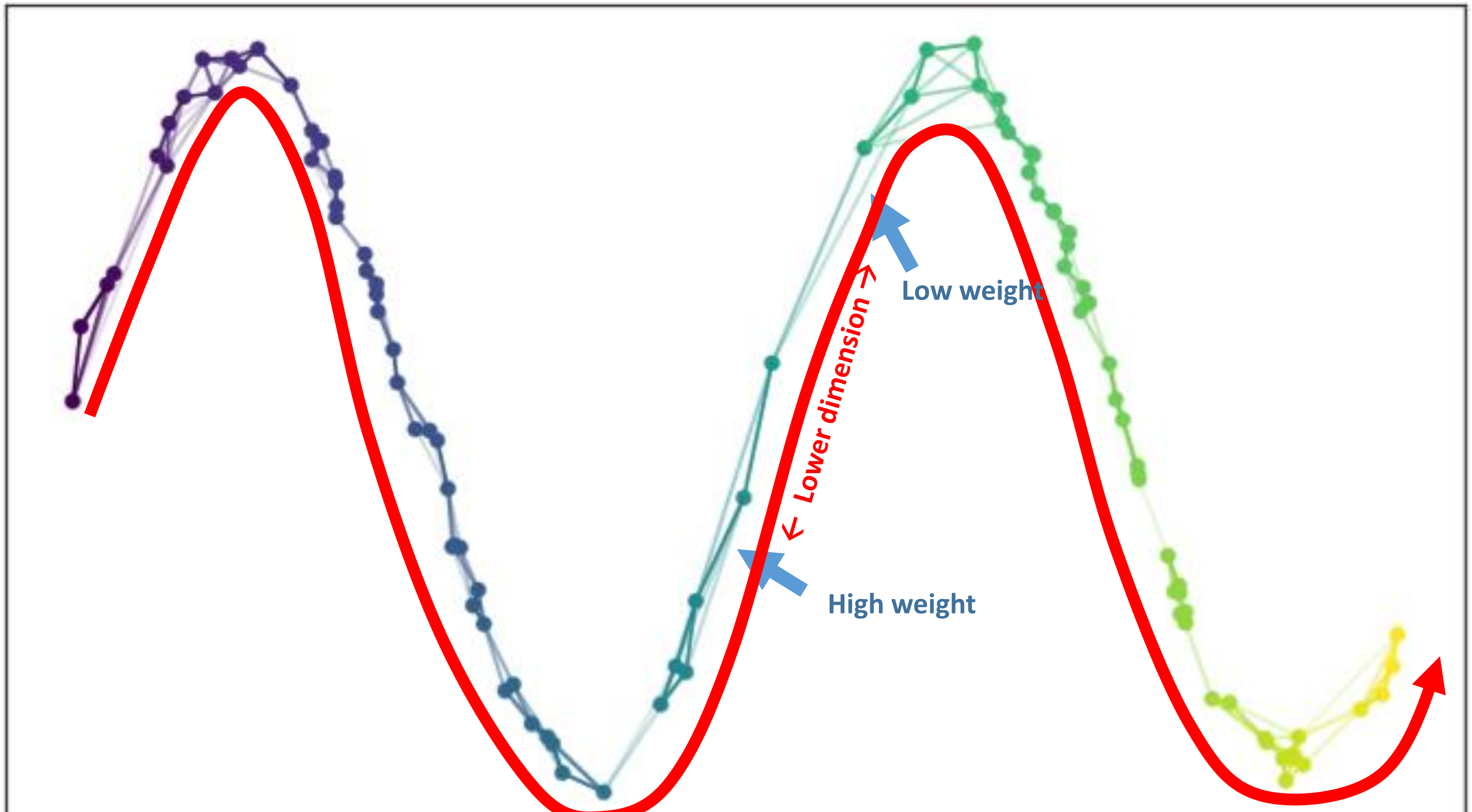
→ Demand that each point is connected to its closest neighbor

→ Weigh connection to further neighbors with distance beyond nearest neighbor

→ Results in a fuzzy topology

Dimensionality reduction example

- Dimensionality reduction from 2D to 1D



Source: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

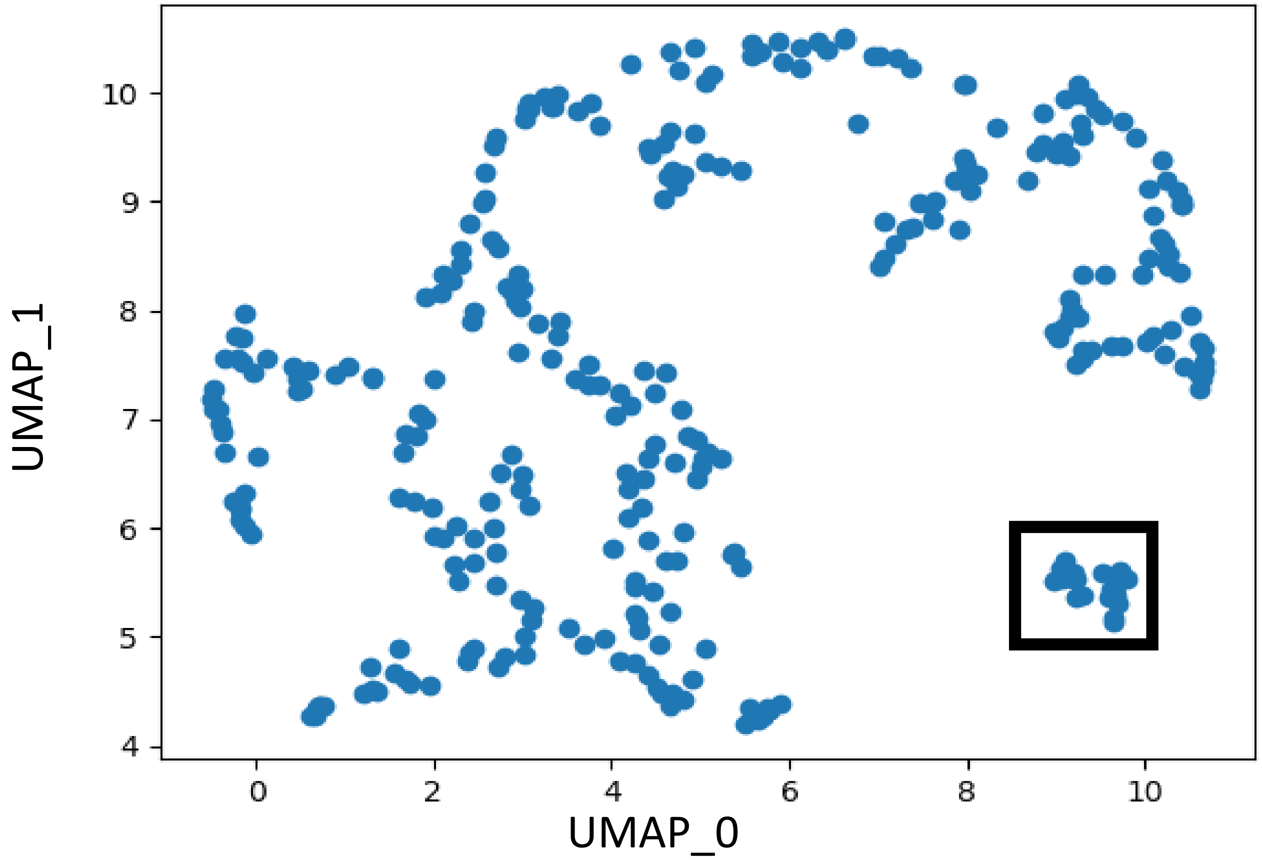
Result:
→ Global neighborhood graph
→ Local scarcity is reflected through edge weights

Last step:
Project this structure into a lower dimension so that overall topology is reflected

Clustering – k-means

Introduction

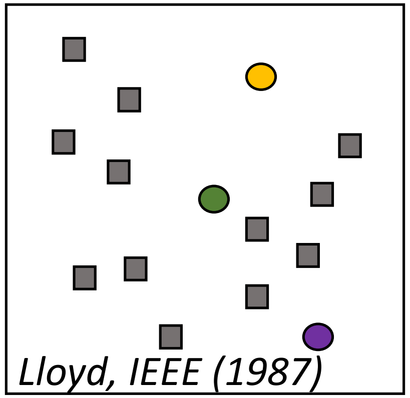
- **Starting point:** Feature space or dimensionality reduction reveals “groups” in our data
- **Can we automatically identify these groups?**



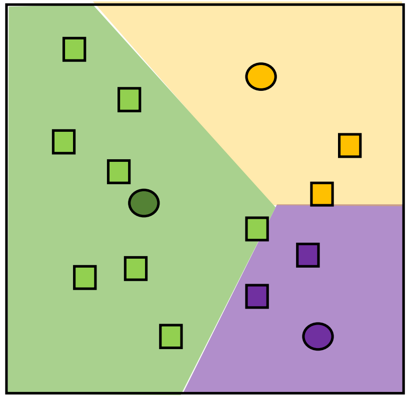
→ **Clustering** allows to stratify data into groups *without previous annotations*

K-means clustering

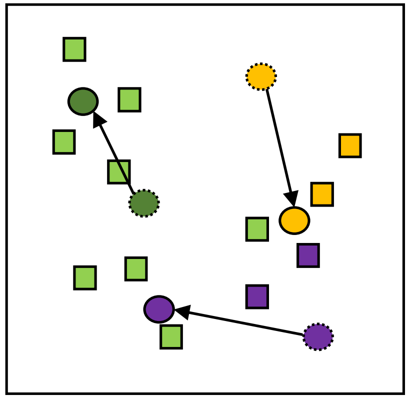
Strategy: Group data points into n groups so that variance within group is minimal



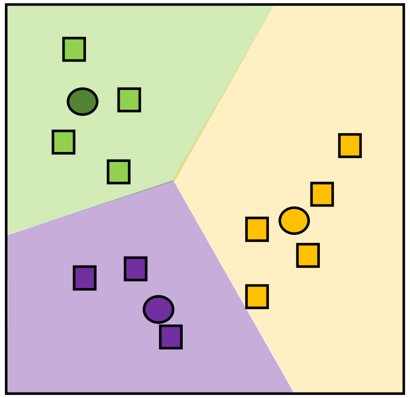
Step1: Random initialization of cluster centers



Step2: Tessellation of space into cluster regions



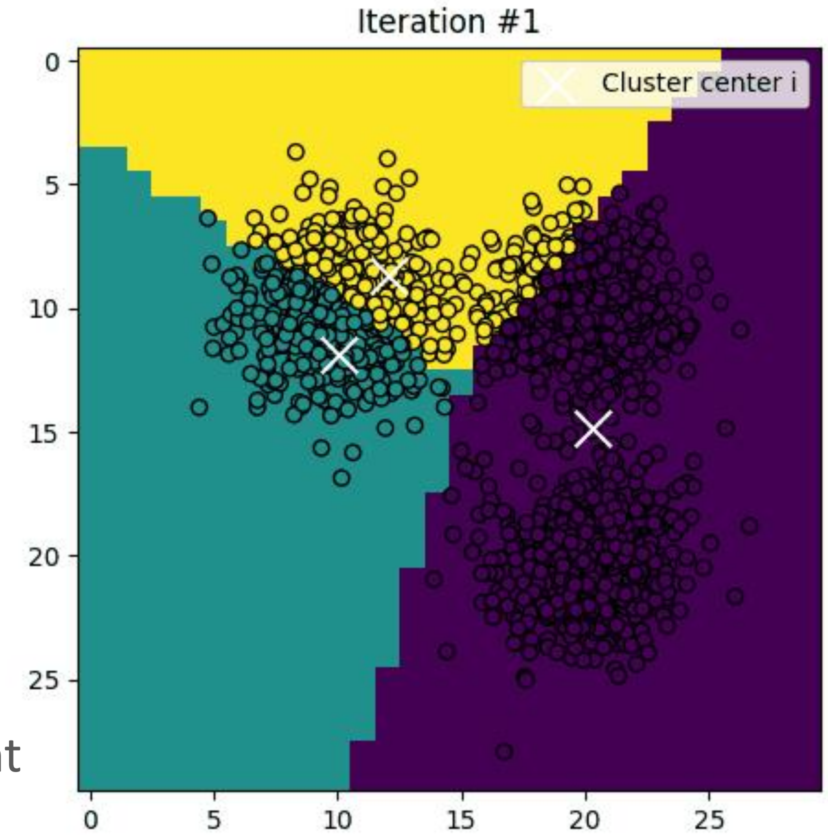
Step3: Replace cluster center with centroids



Step4: Repeat 2&3 until convergence

Strength and weaknesses

- Number of clusters needs to be known
- Clusters can not capture more complex topologies
- Based on Euclidian metrics → every new point can be assigned to a cluster

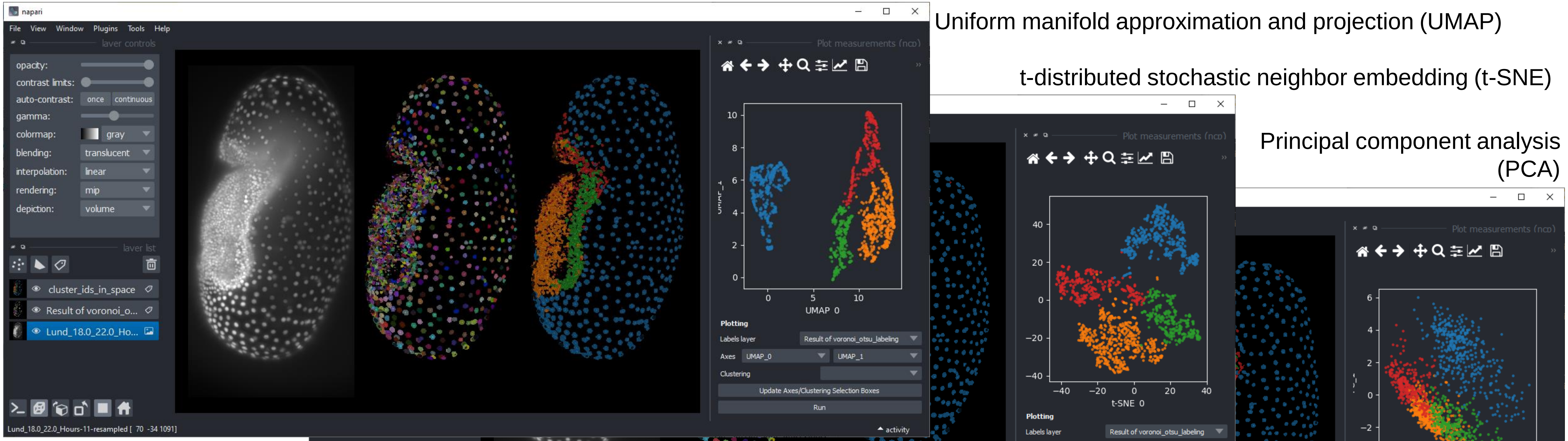


Dimensionality reduction

Uniform manifold approximation and projection (UMAP)

t-distributed stochastic neighbor embedding (t-SNE)

Principal component analysis (PCA)



Laura Žigutytė
@zigutyte

Ryan Savill
@RyanSavill4

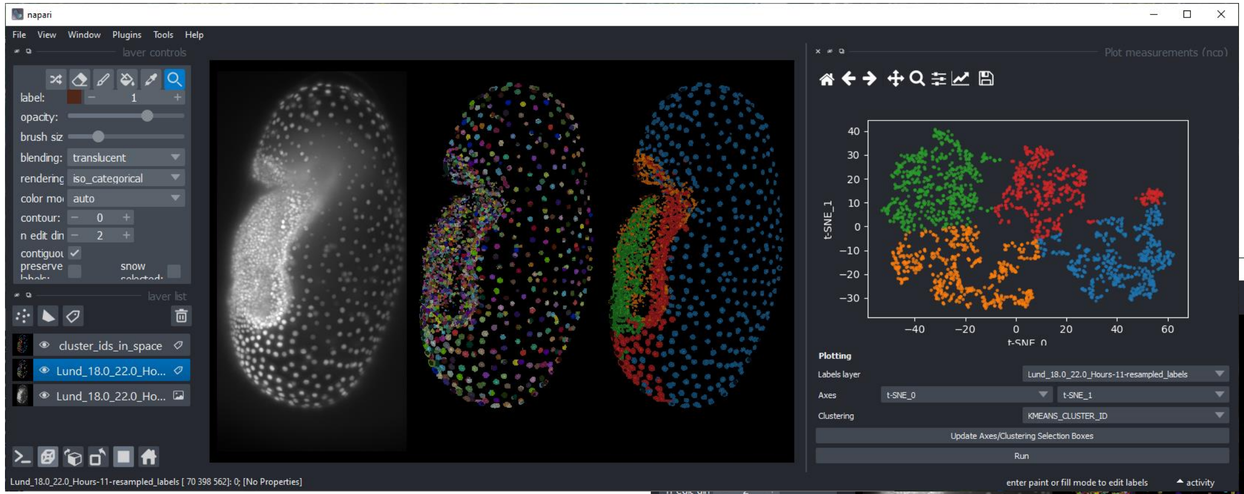
Marcelo L. Zoccoler
Robert Haase,
Thorsten Wagner,
Johannes Soltwedel

@zoccolermarcelo

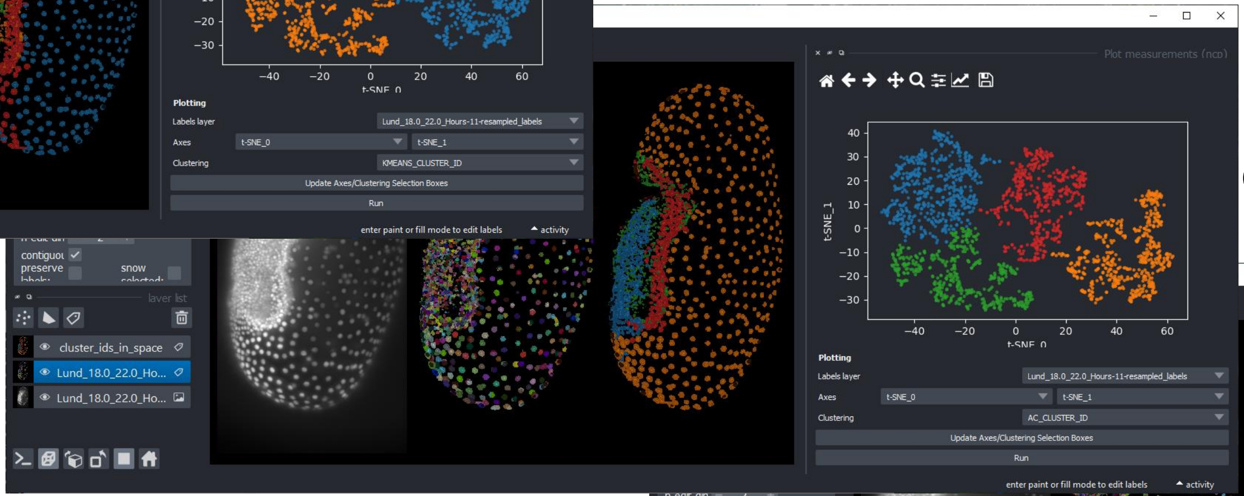
https://github.com/BiAPoL/napari-clusters-plotter

Image Data Source: Daniela Vorkel, Myers lab, MPI-CBG/CSBD Dresden

Clustering

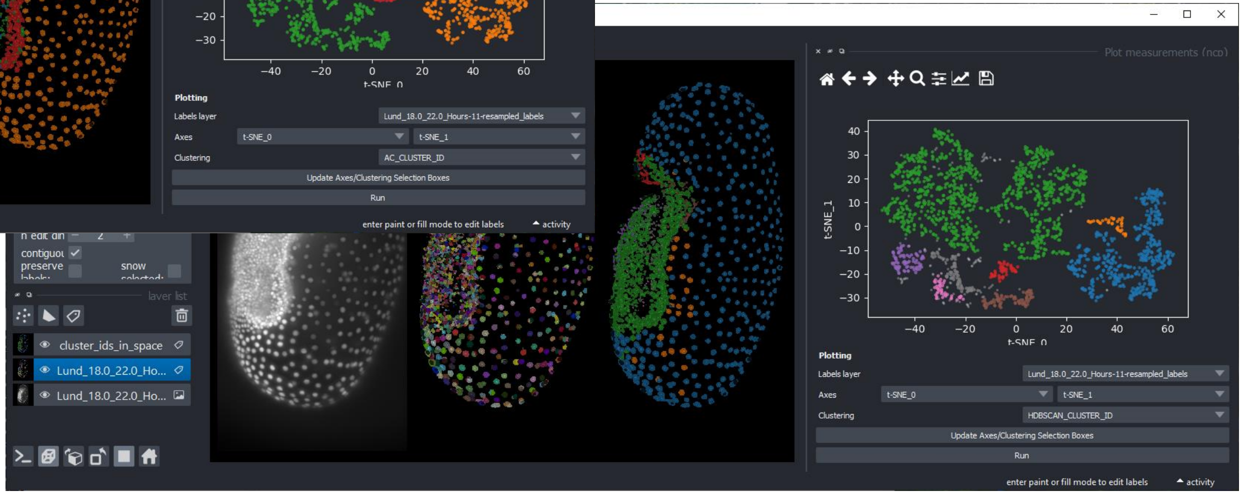


K-means clustering



Agglomerative clustering

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)



Laura Žigutytė
@zigutyte

Ryan Savill
@RyanSavill4

Marcelo L. Zoccoler
Robert Haase,
Thorsten Wagner,
Johannes Soltwedel

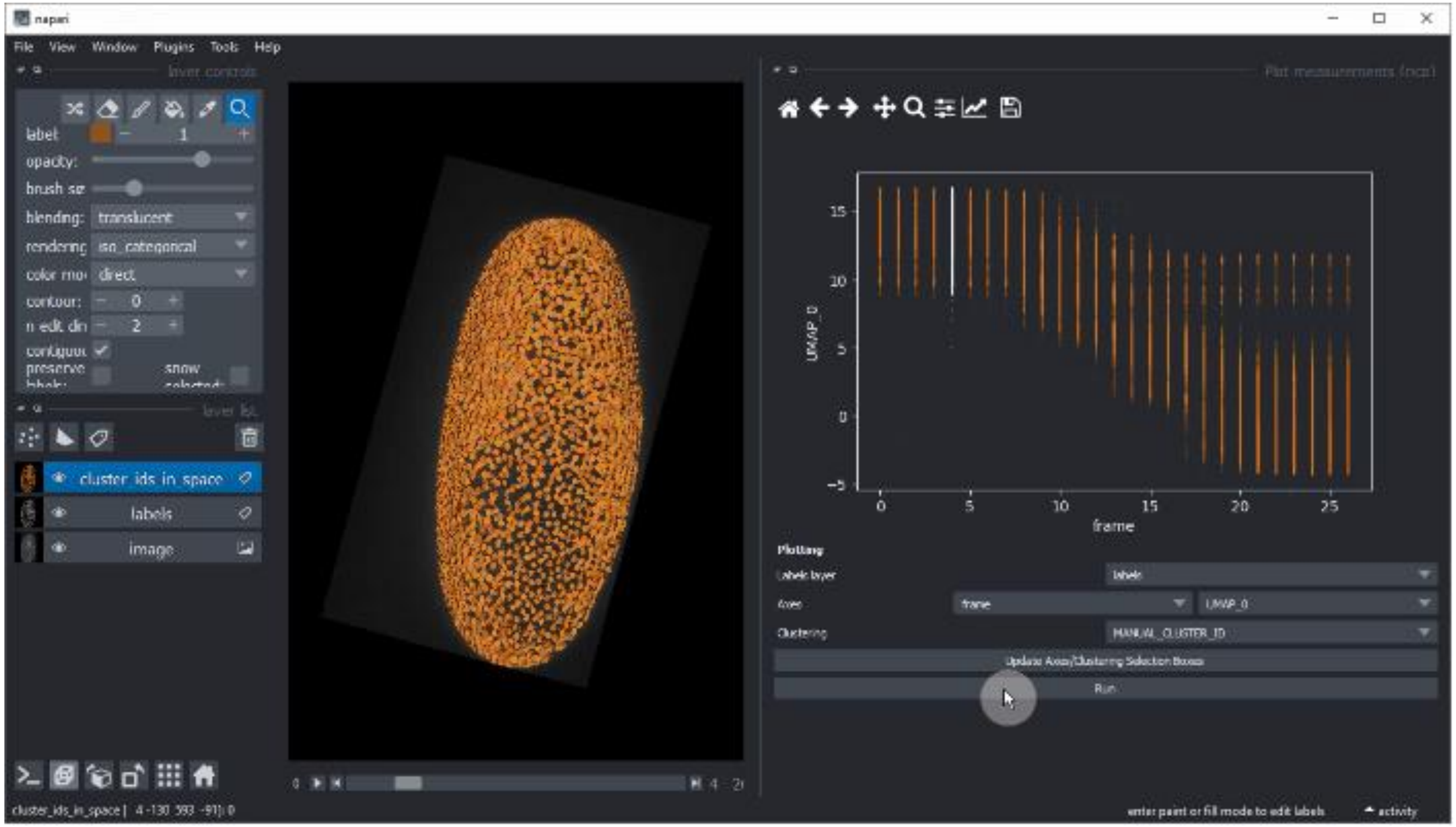
@zoccolermarcelo

https://github.com/BiAPoL/napari-clusters-plotter

Image Data Source: Daniela Vorkel, Myers lab, MPI-CBG/CSBD Dresden

Data Exploration

- … using interactive clustering



Laura Žigutytė
@zigutyte

Ryan Savill
@RyanSavill4

Marcelo L. Zoccoler
Robert Haase,
Thorsten Wagner,
Johannes Soltwedel

Image data source: Daniela Vorkel, Myers lab, MPI-CBG/CSBD

Exercise

@zoccolermarcelo

August 2023

Separate different cell populations

1. Open the human_mitosis image:
   - File > Open Sample > napari > Human Mitosis

2. Apply the ‘voronoi_otsu_labeling’ function
   - Tools > Segmentation / labeling > Voronoi-Otsu-labeling
   - Choose parameters and click on ‘Run’
   - *Close widgets (optional, but recommended)*

3. Use napari-skimage-regionprops to extract object features
   - Tools > Measurement tables > Object Features / Properties
   - Select ‘intensity’, ‘size’ and ‘shape’ and click ‘Run’
   - *Close widgets, including table (optional, but recommended)*

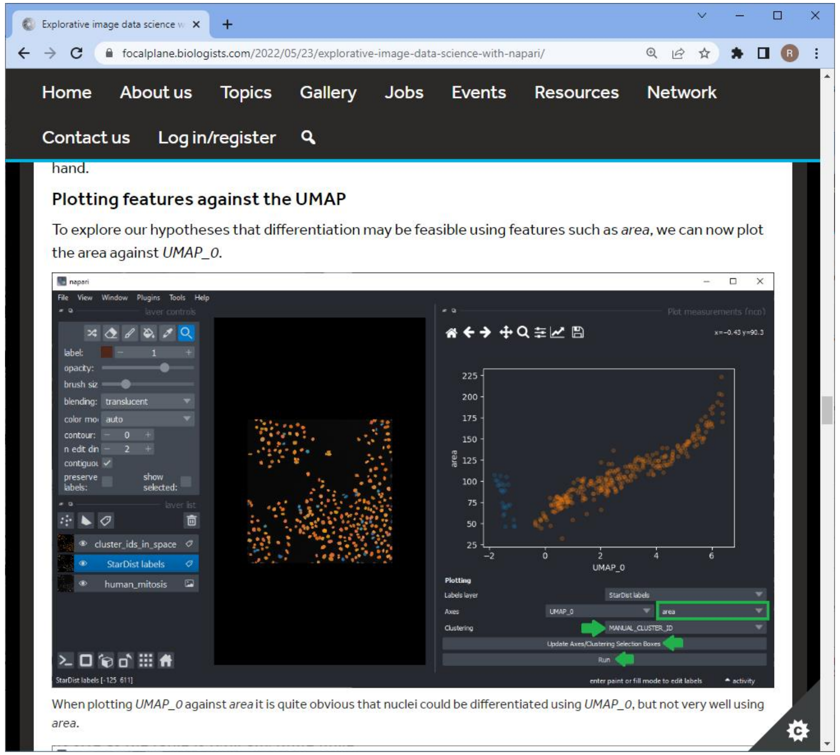
4. Use napari-clusters-plotter to apply dimensionality reduction
   - Tools > Measurements post-processing > Dimensionality Reduction
   - Select UMAP in the ‘Dimensionality Reduction Method’ dropdown and click ‘Run’
   - *Close widgets, including table (optional, but recommended)*

5. Use napari-clusters-plotter to apply k-means clustering
   - Tools > Measurements post-processing > Clustering
   - In the Measurements field, hold CTRL (CMD on MAC) and click on ‘UMAP-0’ and ‘UMAP-1’
   - In Clustering Method, select ‘K-MEANS’ and click ‘Run’

6. Plot UMAP dimensions and display K-MEANS clusters with different colors
   - Tools > Visualization > Plot Measurements
   - In Axes, select ‘UMAP-0’ and ‘UMAP-1’
   - In Clustering, select ‘K-MEANS-CLUSTER-ID’
   - Click ‘Plot’

Further reading

- Blog posts on the Focalplane about Unsupervised Machine Learning



https://focalplane.biologists.com/2022/05/23/explorative-image-data-science-with-napari/

Acknowledgements

PoL Physics of Life TU Dresden

BiAPoL team

- Mara Lampert
- Marcelo Zoccoler
- Johannes Soltwedel
- Maleeha Hassan
- Allyson Ryan
- Till Korten
- Stefan Hahmann
- Somashekhar Kulkarni

Former lab members:

- Ryan George Savill
- Laura Zigutyte

Networks

BioDIP Biopolis Dresden Imaging Platform

CENTER FOR SYSTEMS BIOLOGY DRESDEN

NFDI4BIOIMAGE

neubias network of european bioimage analysts

image.sc

German BioImaging Gesellschaft für Mikroskopie und Bildanalyse

Funding

DFG public

Federal Ministry of Education and Research

Chan Zuckerberg Initiative