TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Dimensionality reduction and UMAP

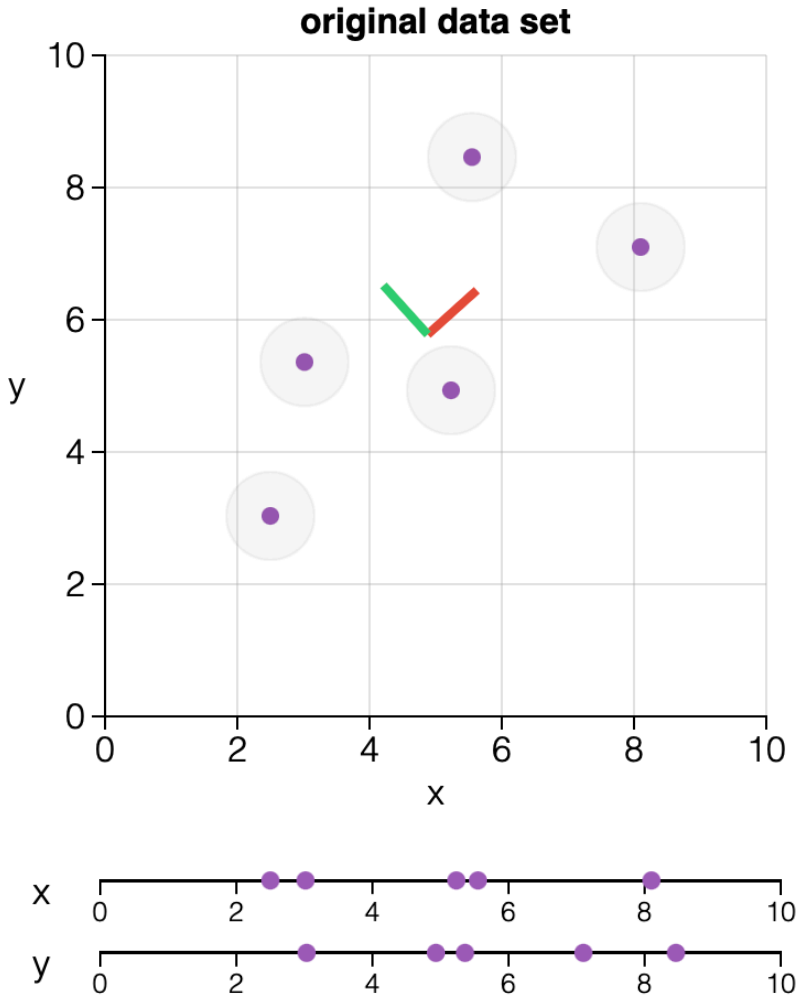
Anna Poetsch

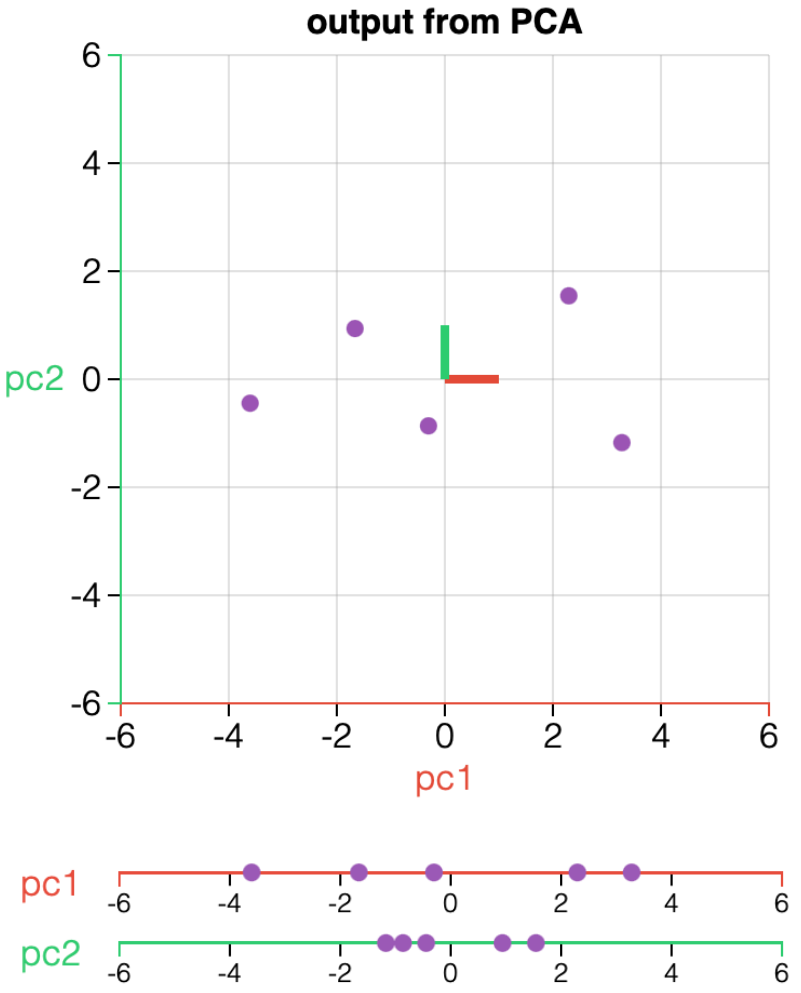
Research Group „Biomedical Genomics“, Biotechnology Center TU Dresden, NCT Dresden, and MSNZ

27.06.22

@apoetsch

Principle Component Analysis (PCA)

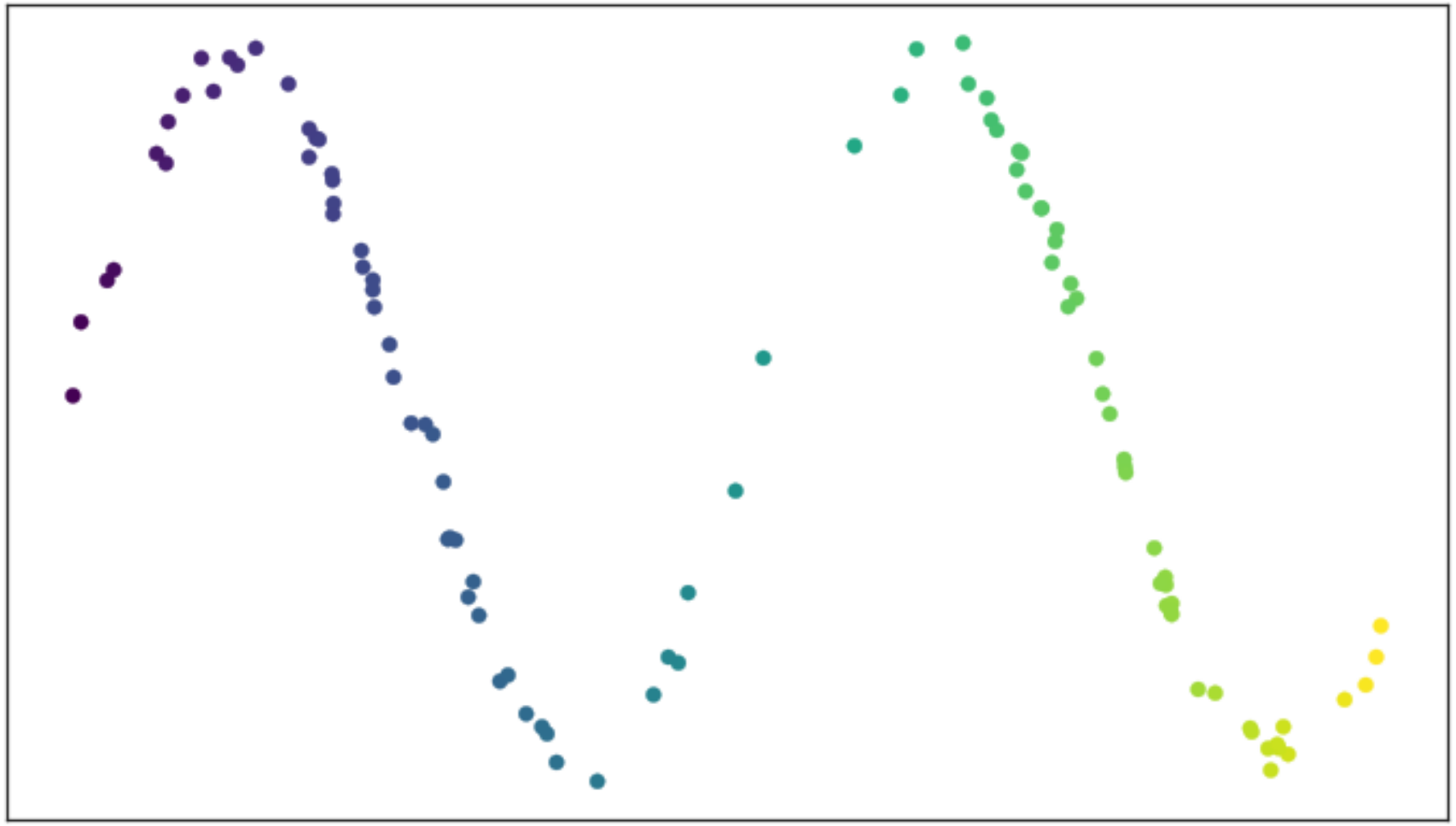




Principle Component Analysis (PCA)

- PCA builds linear projections of data into a new coordinate system
- The coordinate system is chosen and ranked by the variance it explains in the data
- Usually the first principle components, the ones with the highest variance explained are shown
- How much variance they explain is indicative of how well the dimensionality reduction has worked

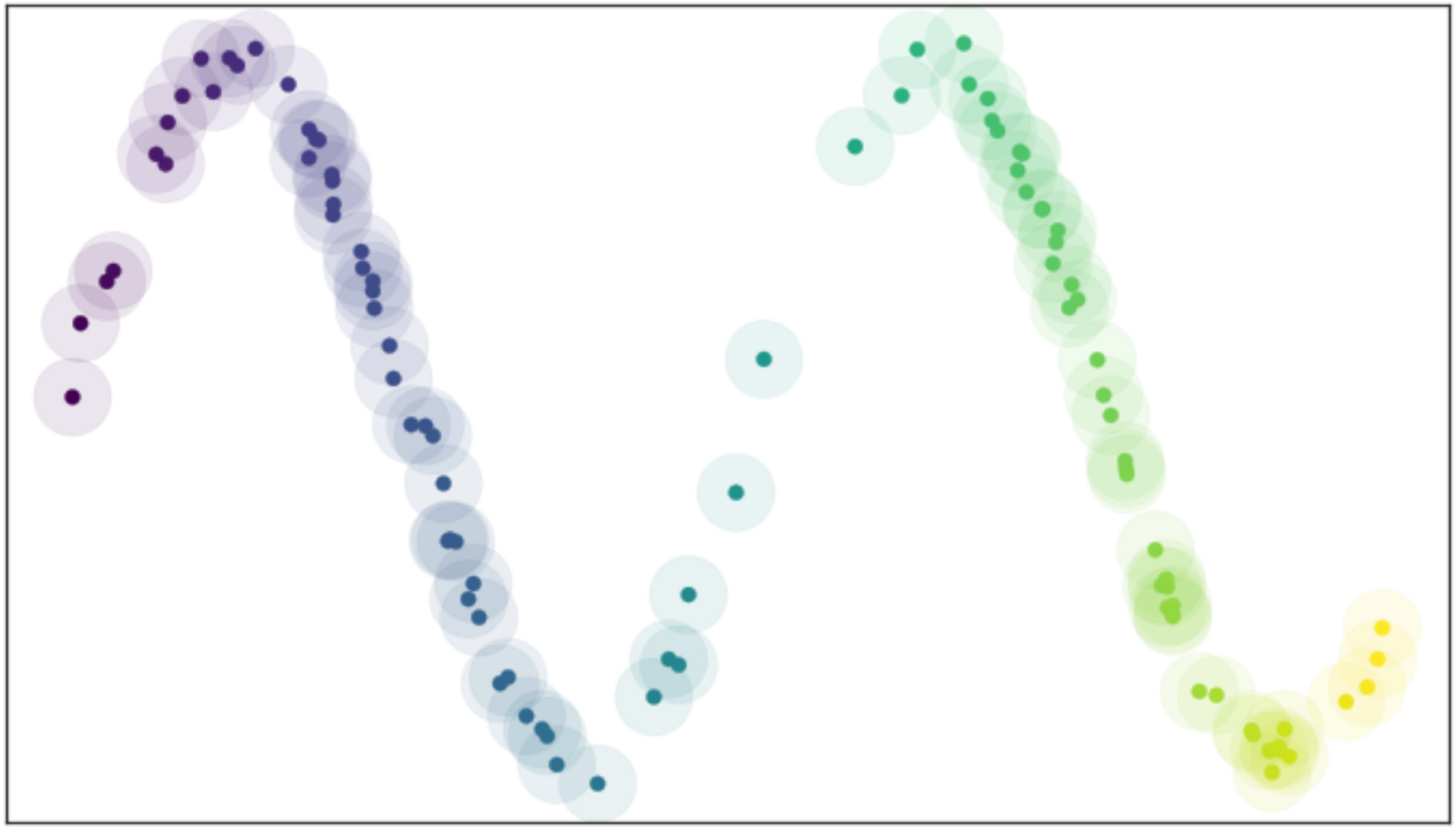
UMAP with two dimensions



As example a sinus curve with some noise

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

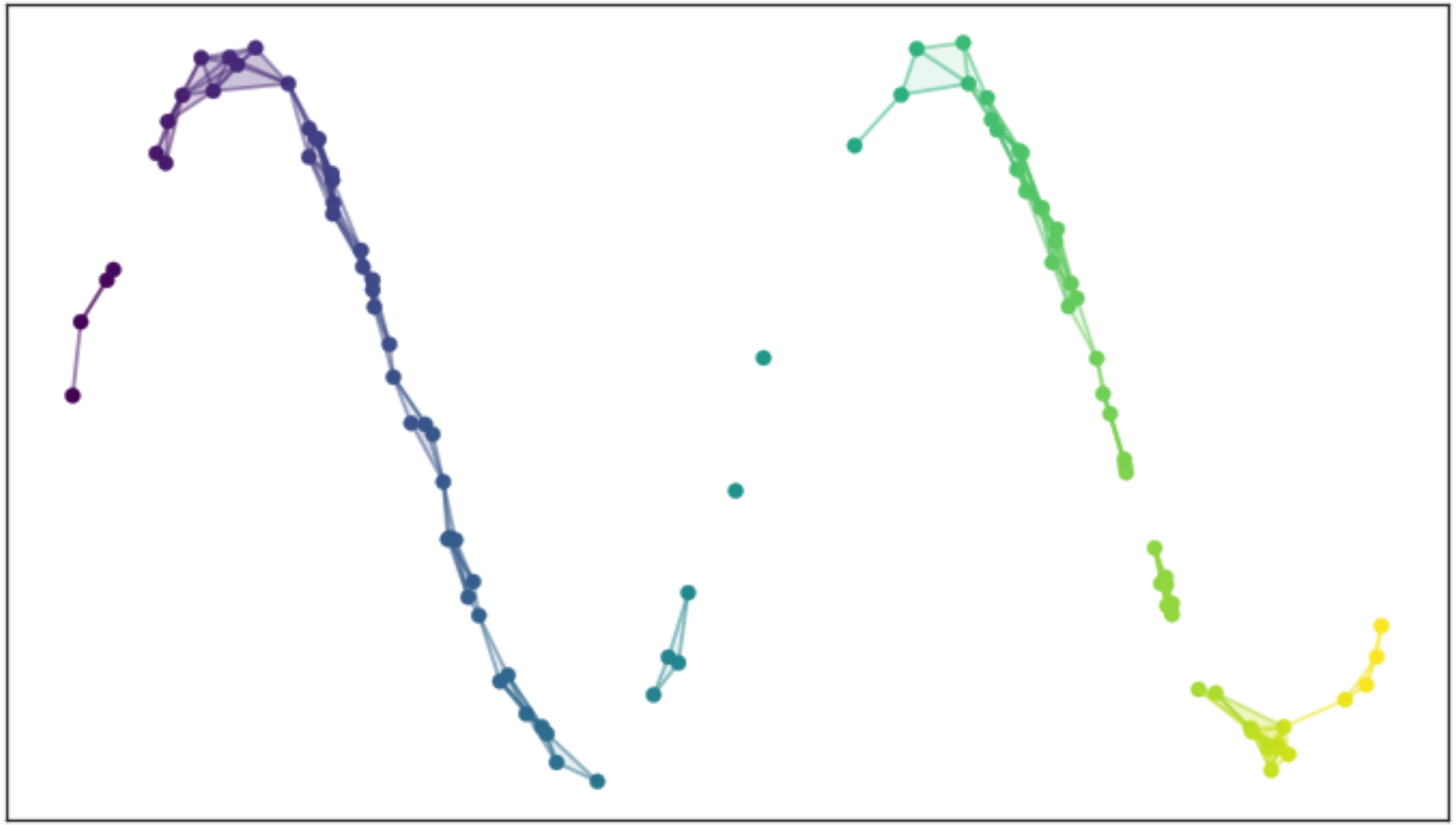
UMAP with two dimensions



A radius shows us whether there are neighbours

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

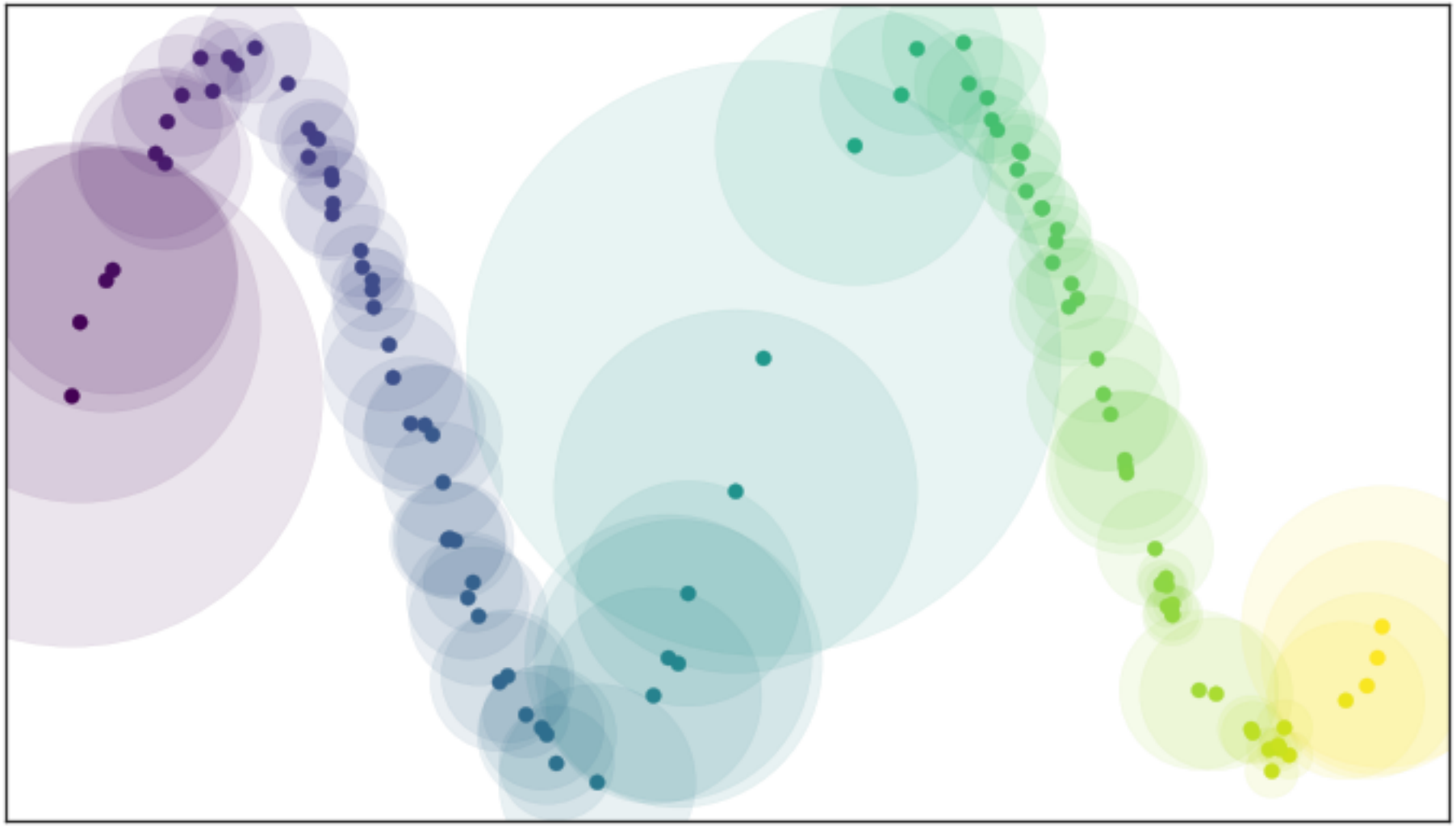
UMAP with two dimensions



The resulting yes-no answer is a bit unsatisfactory

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

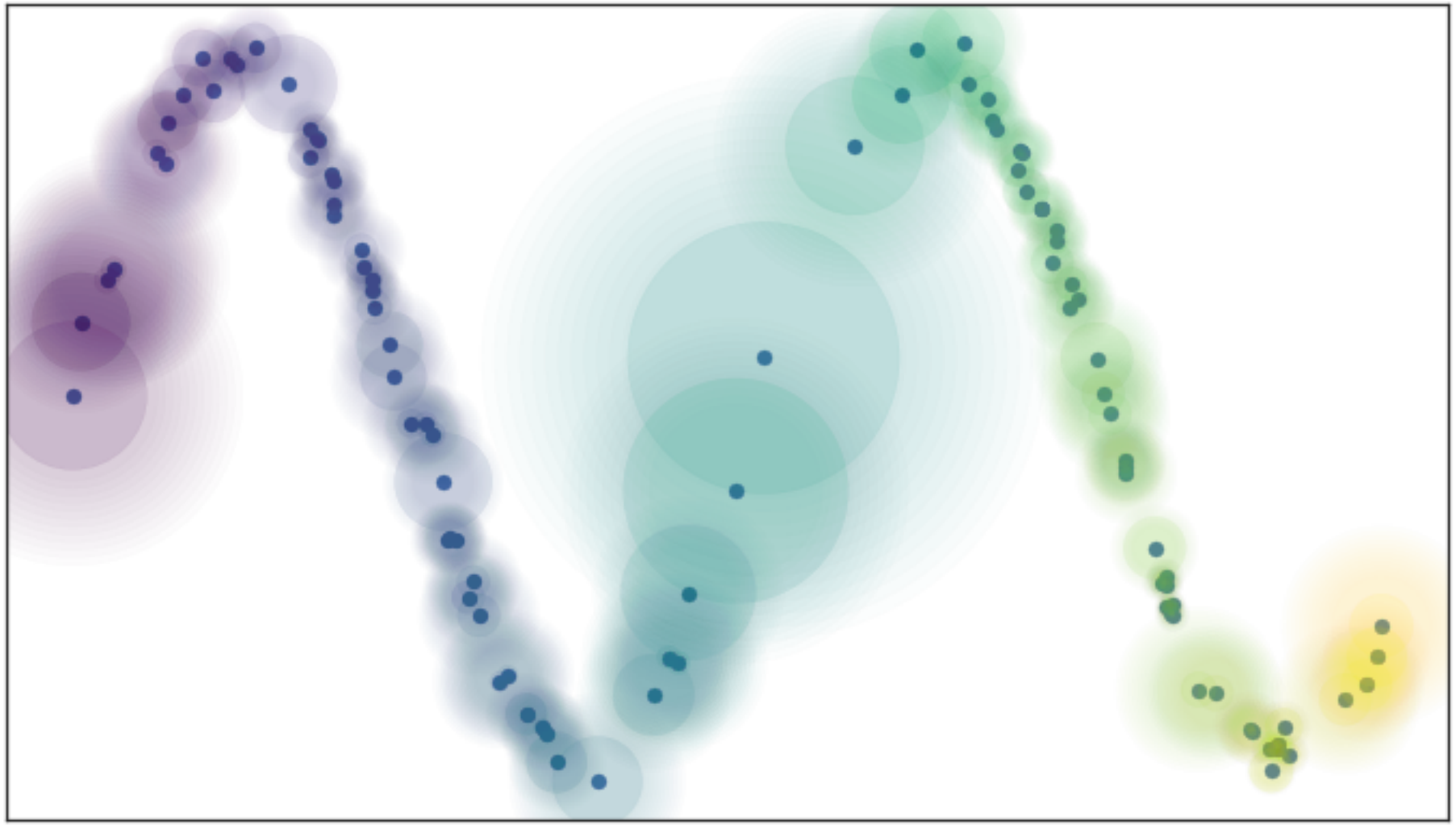
UMAP with two dimensions



Flexible radia also allow finding of isolated points

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

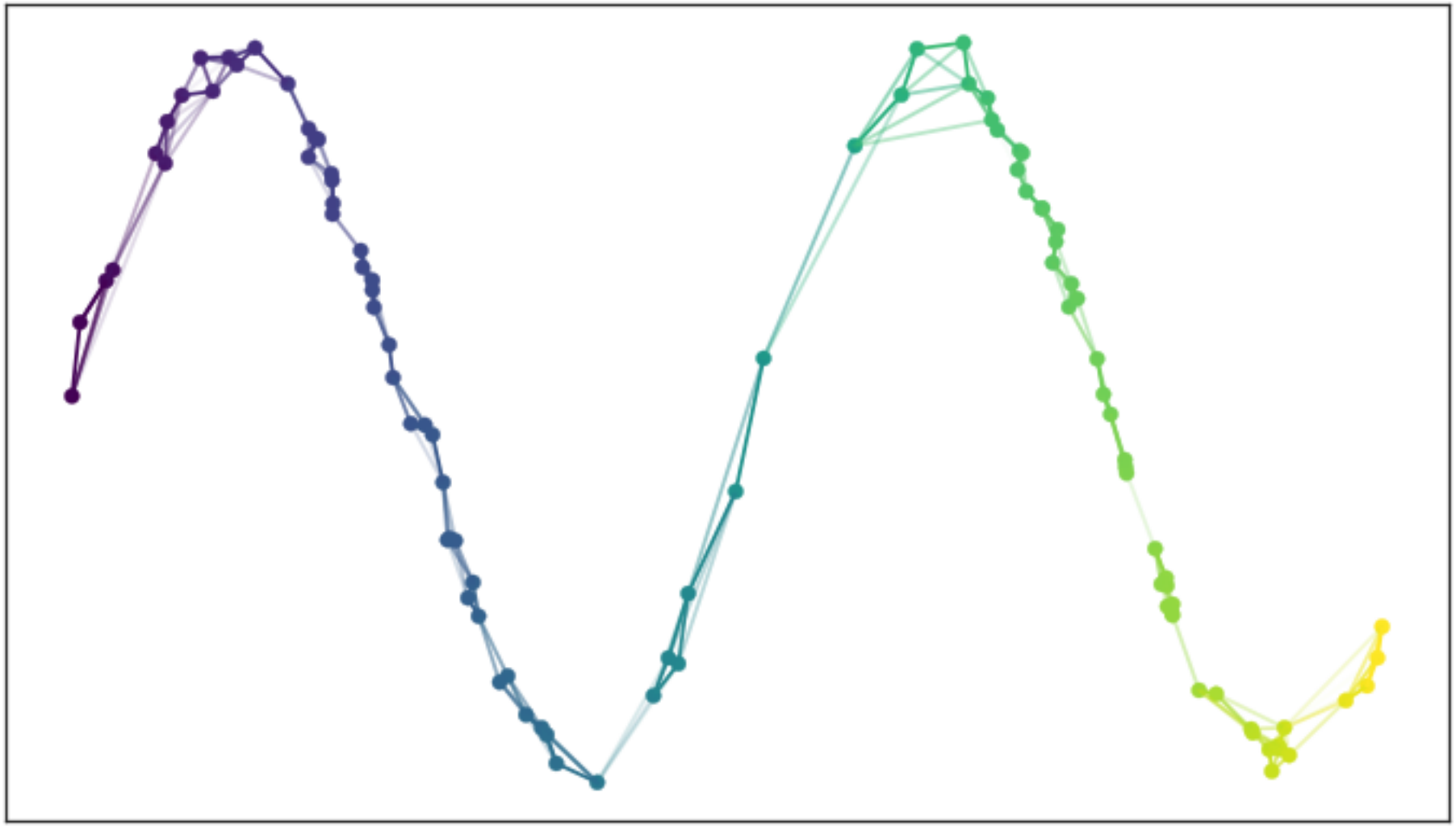
UMAP with two dimensions



A combination of a hard radius to the next neighbour and a flexible one beyond is more practical, because…

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

UMAP with two dimensions



...it allows us to calculate probabilities, whether there are connections between the points.

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

Dimensionality reduction with UMAP

- Optimisation algorithm for a “flexible” distance measurement to find a place in a low-dimensional space

- The result of the optimisation is dependent on the data and a random component

- UMAP can be tweaked with…
  …how distance is measured (metric)
  …how many neighbours are considered (n_neighbors)
  …how much points are allowed to overlay (min-dist)

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

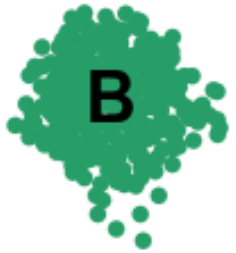
Things to consider

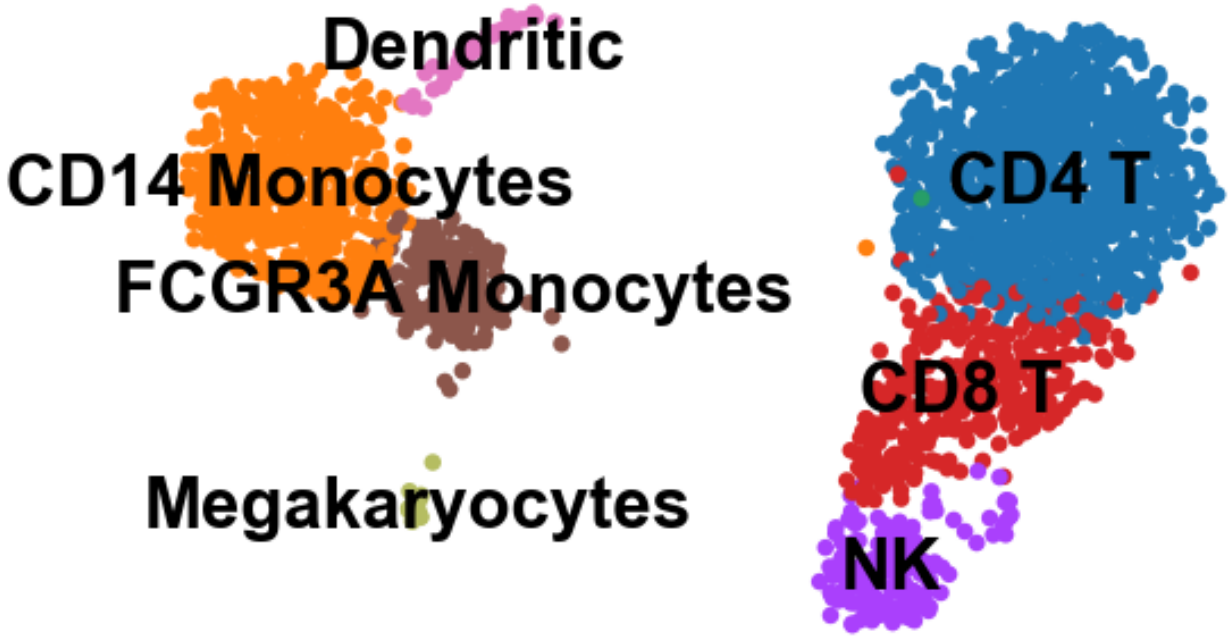
- Many parameters invite to “adjust” the data analysis
- Danger to over-interpret the visual “distance”
- How much data structure is preserved is still a matter of debate

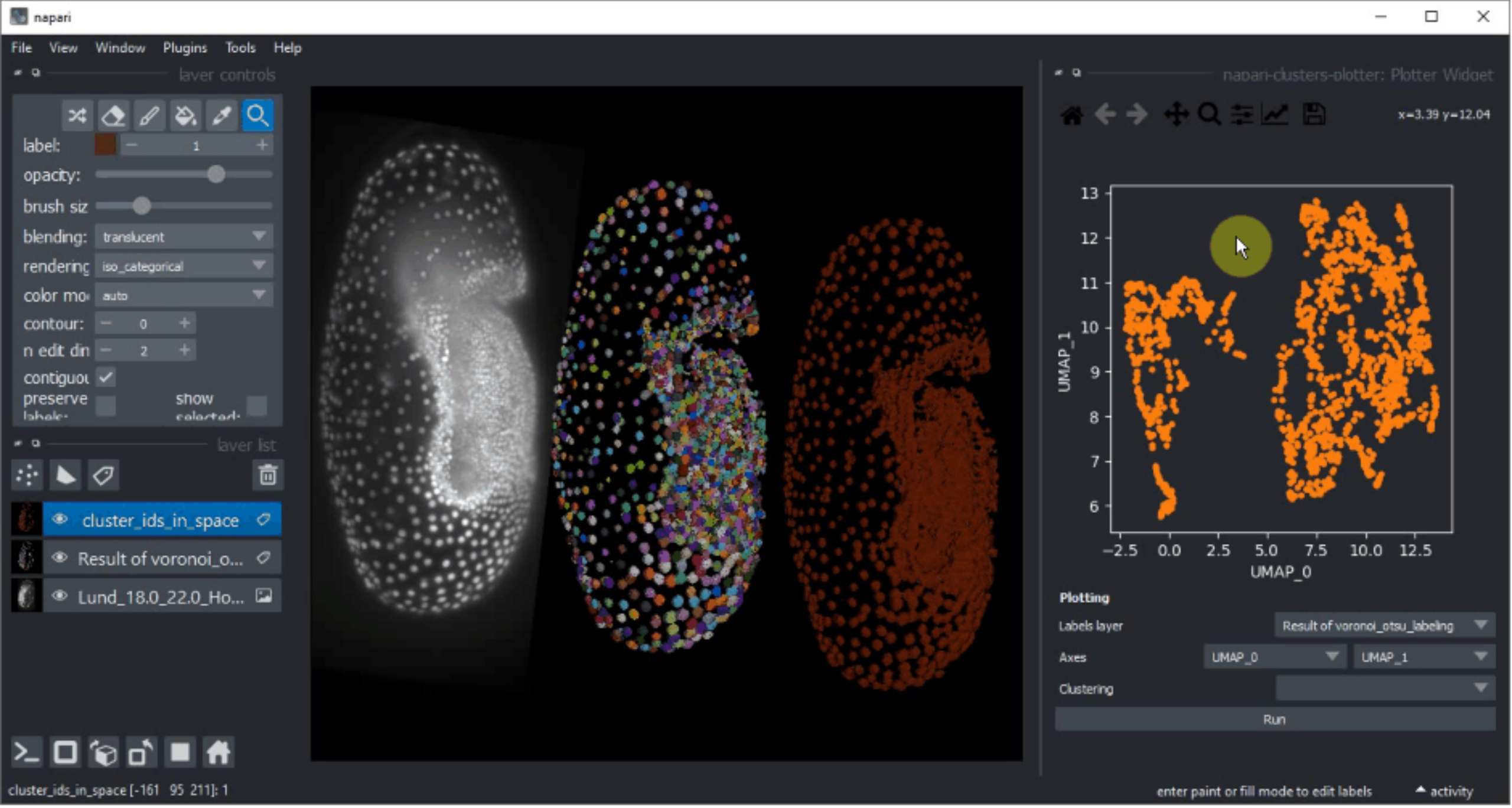
https://www.nature.com/articles/s41587-020-00809-z

Applications of UMAP

Ready to go Tutorials:
scRNA-Seq tutorial in Python: https://github.com/theislab/single-cell-tutorial
scRNA-Seq blood analysis in Python: https://scanpy-tutorials.readthedocs.io/en/latest/pbmc3k.html
scRNA-Seq blood analysis in R: https://satijalab.org/seurat/articles/pbmc3k_tutorial.html






napari
File View Window Plugins Tools Help
layer controls
label:
opacity:
brush siz
blending: translucent
rendering iso_categorical
color mo auto
contour:
n edit din
contiguou
preserve
show
layer list
cluster_ids_in_space
Result of voronoi_o...
Lund_18.0_22.0_Ho...
napari-clusters-plotter: Plotter Widget
x=3.39 y=12.04
UMAP_1
UMAP_0
Plotting
Labels layer
Result of voronoi_otsu_labeling
Axes
UMAP_0
UMAP_1
Clustering
Run
cluster_ids_in_space [-161 95 211]: 1
enter paint or fill mode to edit labels
activity

Sources and Material:

Tutorial: https://umap-learn.readthedocs.io/en/latest/
Paper: https://arxiv.org/abs/1802.03426
scRNA-Seq tutorial in Python: https://github.com/theislab/single-cell-tutorial
blood analysis in Python: https://scanpy-tutorials.readthedocs.io/en/latest/pbmc3k.html
blood analysis in R: https://satijalab.org/seurat/articles/pbmc3k_tutorial.html