

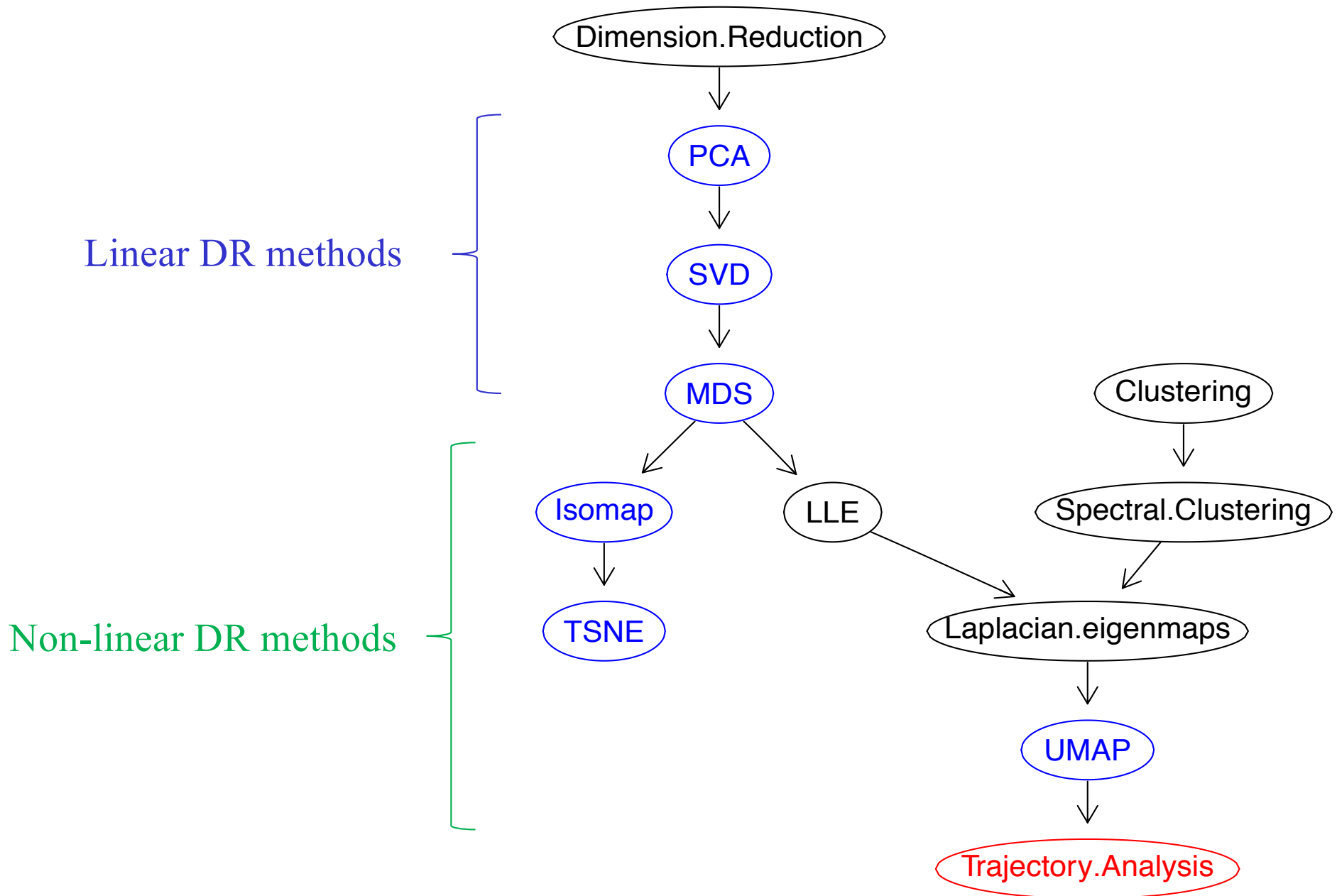
Dimension Reduction Methods: From PCA to TSNE and UMAP

Maxwell Lee

High-dimension Data Analysis Group
Laboratory of Cancer Biology and Genetics
Center for Cancer Research
National Cancer Institute

June 18, 2020

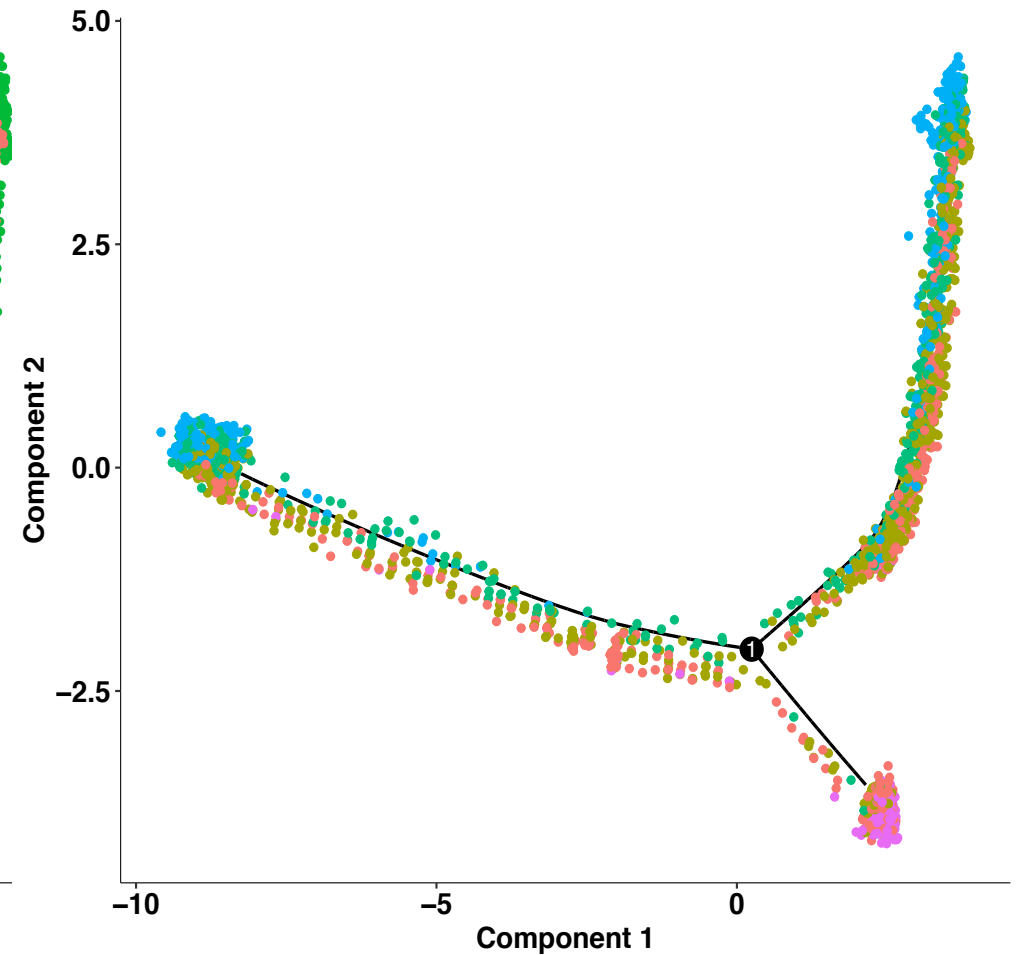
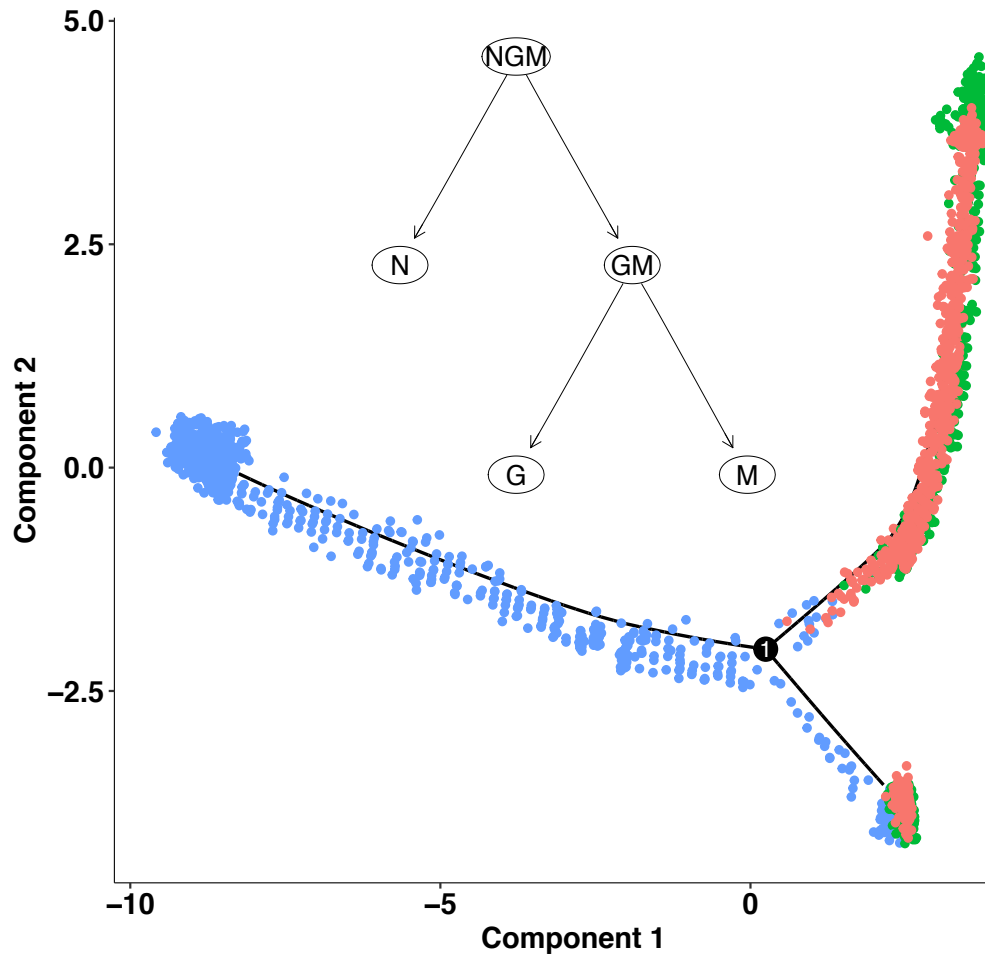
Road Map for Dimension Reduction Methods



Neural Crest Lineage Analysis with Monocle 2

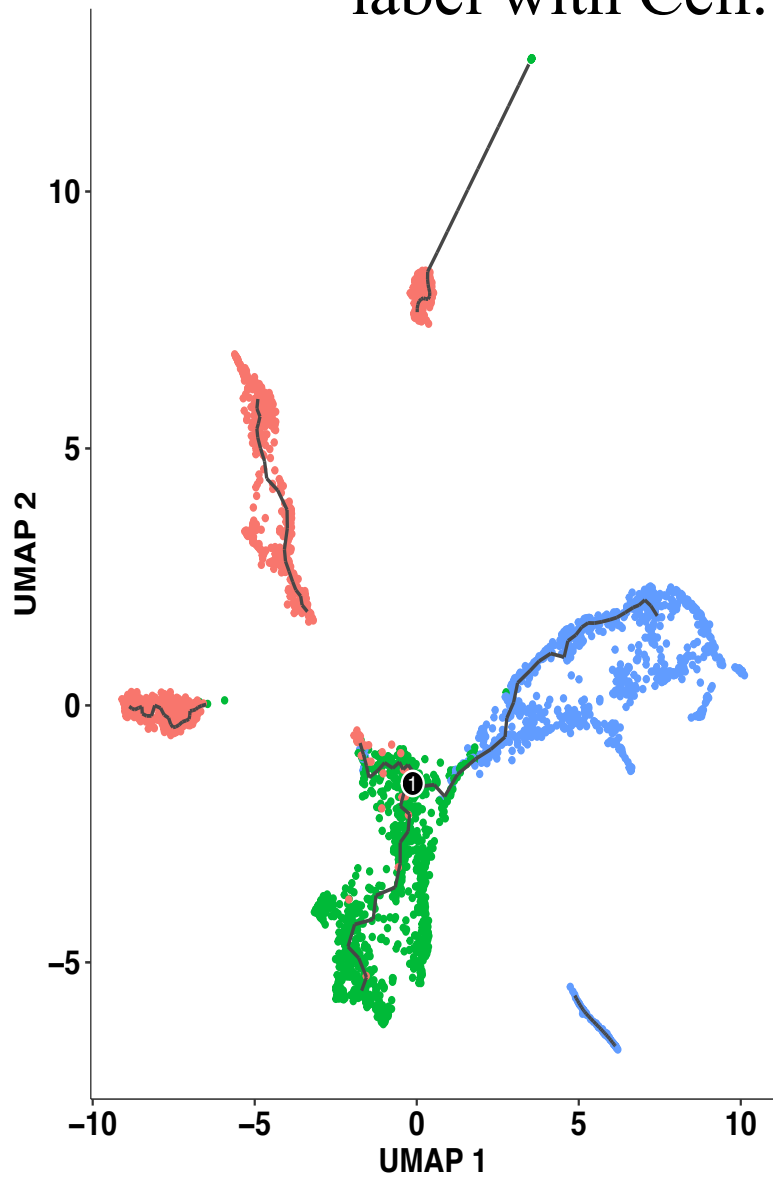
Cell.type • Melanocytes • Schwann • Sensory

day • 10.5 • 11.5 • 12.5 • 13.5 • 9.5

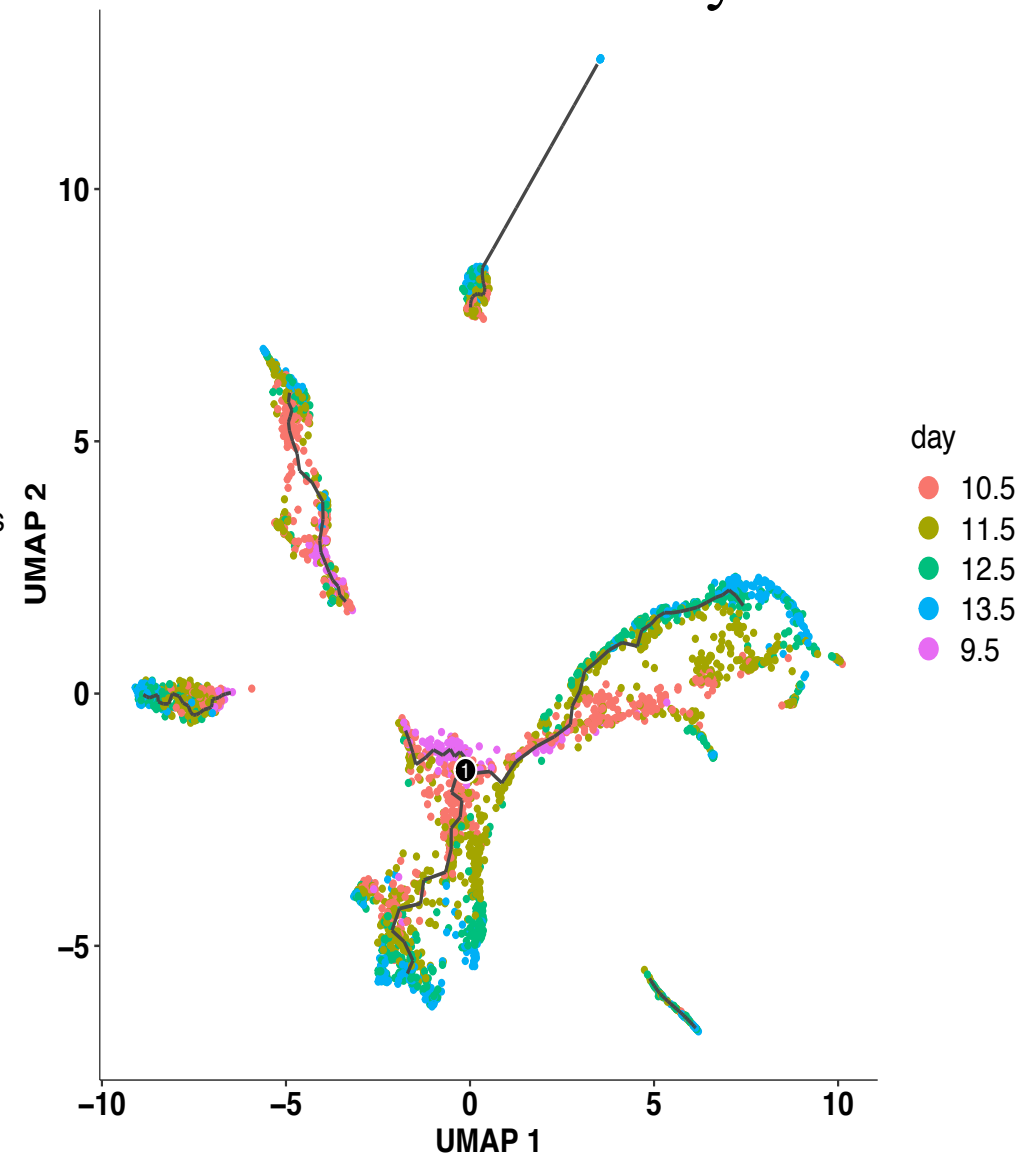


Neural Crest Lineage Analysis with Monocle 3

label with Cell.type

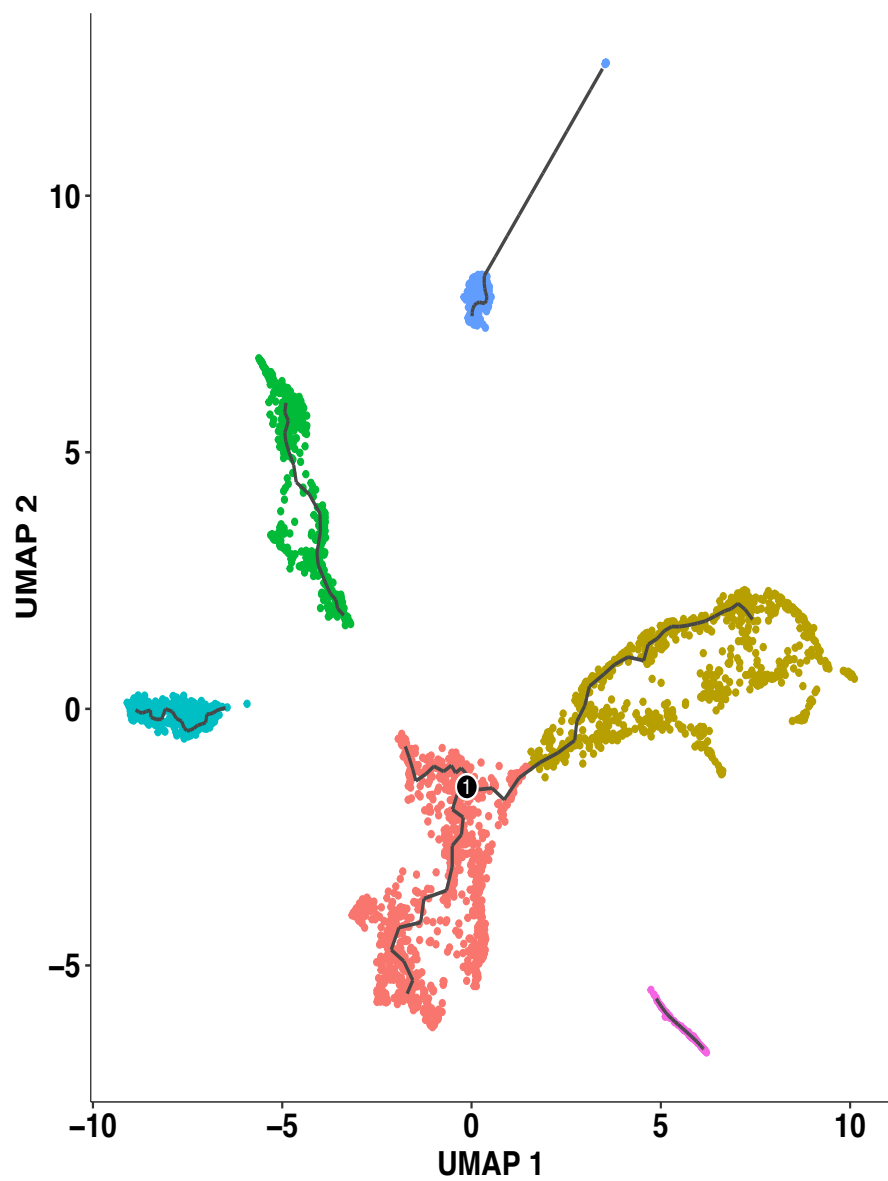


label with day

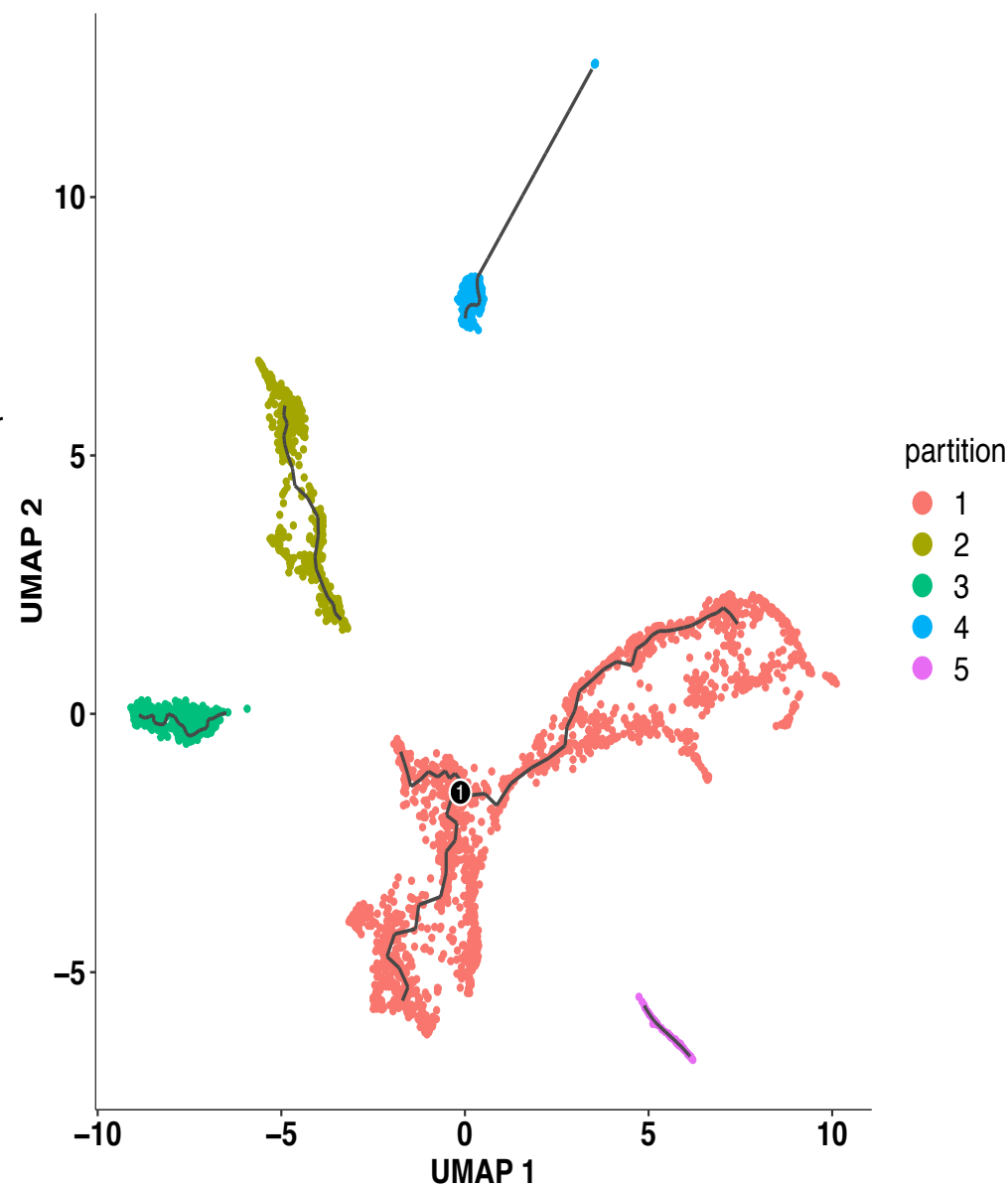


Neural Crest Lineage Analysis with Monocle 3

label with cluster

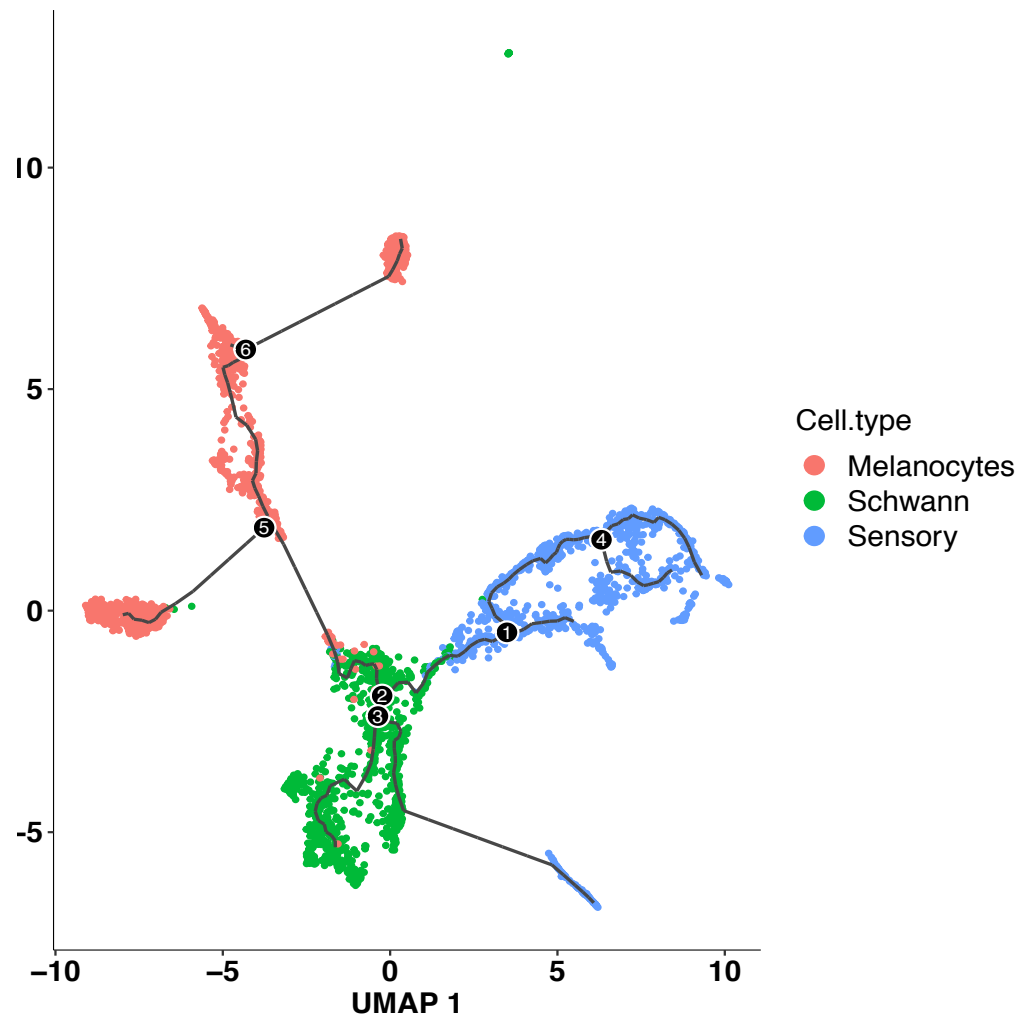
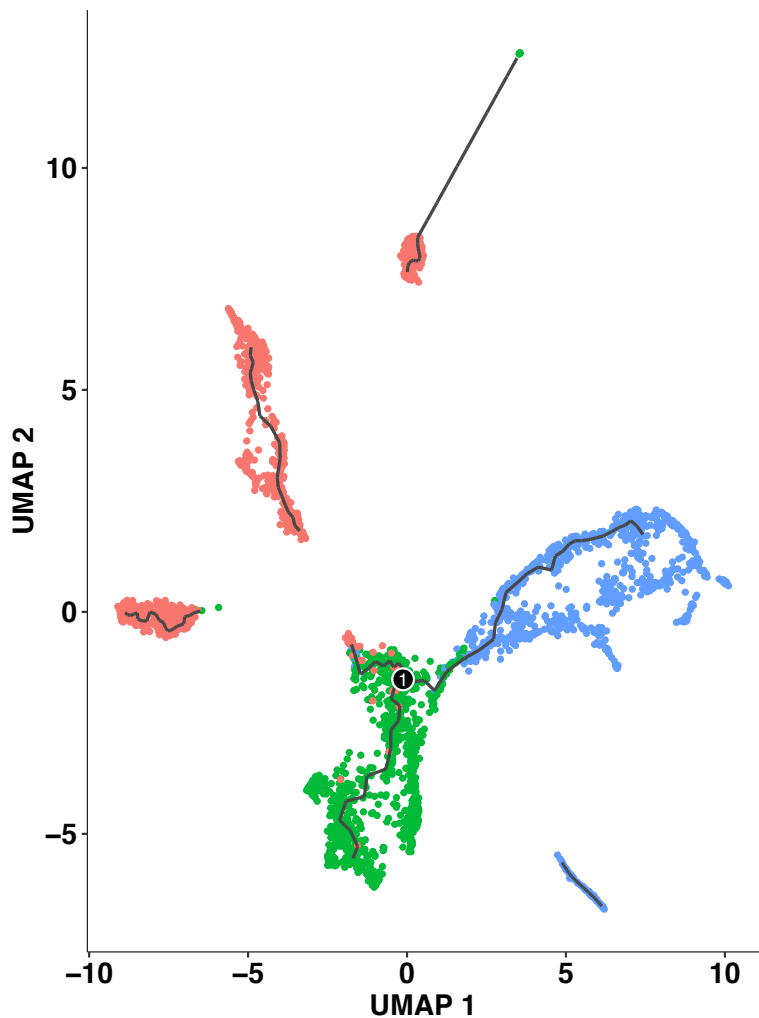


label with partition



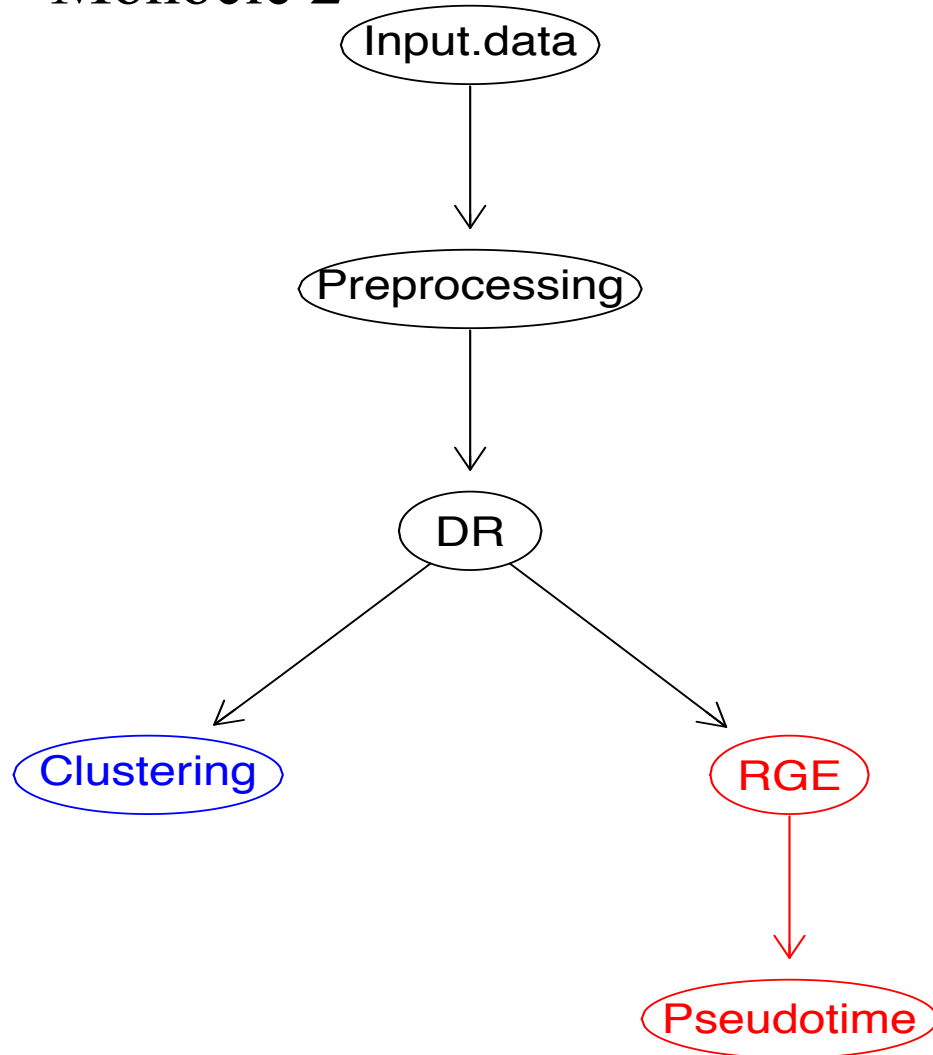
Neural Crest Lineage Analysis with Monocle 3

	default	current analysis
close_loop	true	true
use_partition	true	false

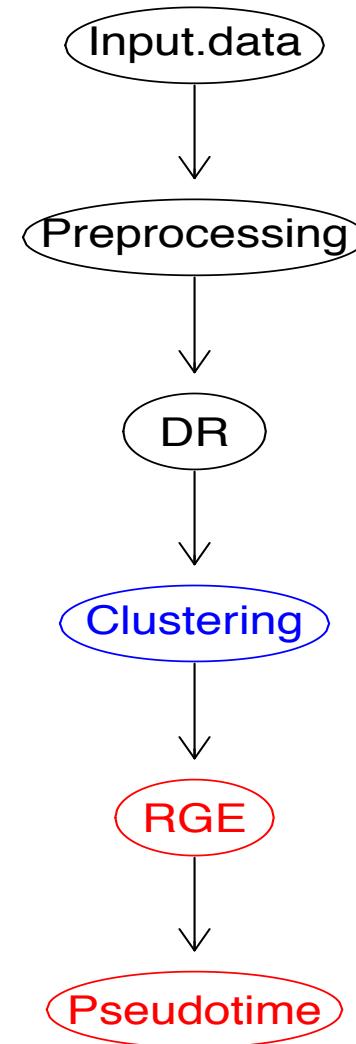


Outline of Trajectory Analysis with Monocle 2 and 3

Monocle 2



Monocle 3



DR: dimension reduction

RGE: reversed graph embedding

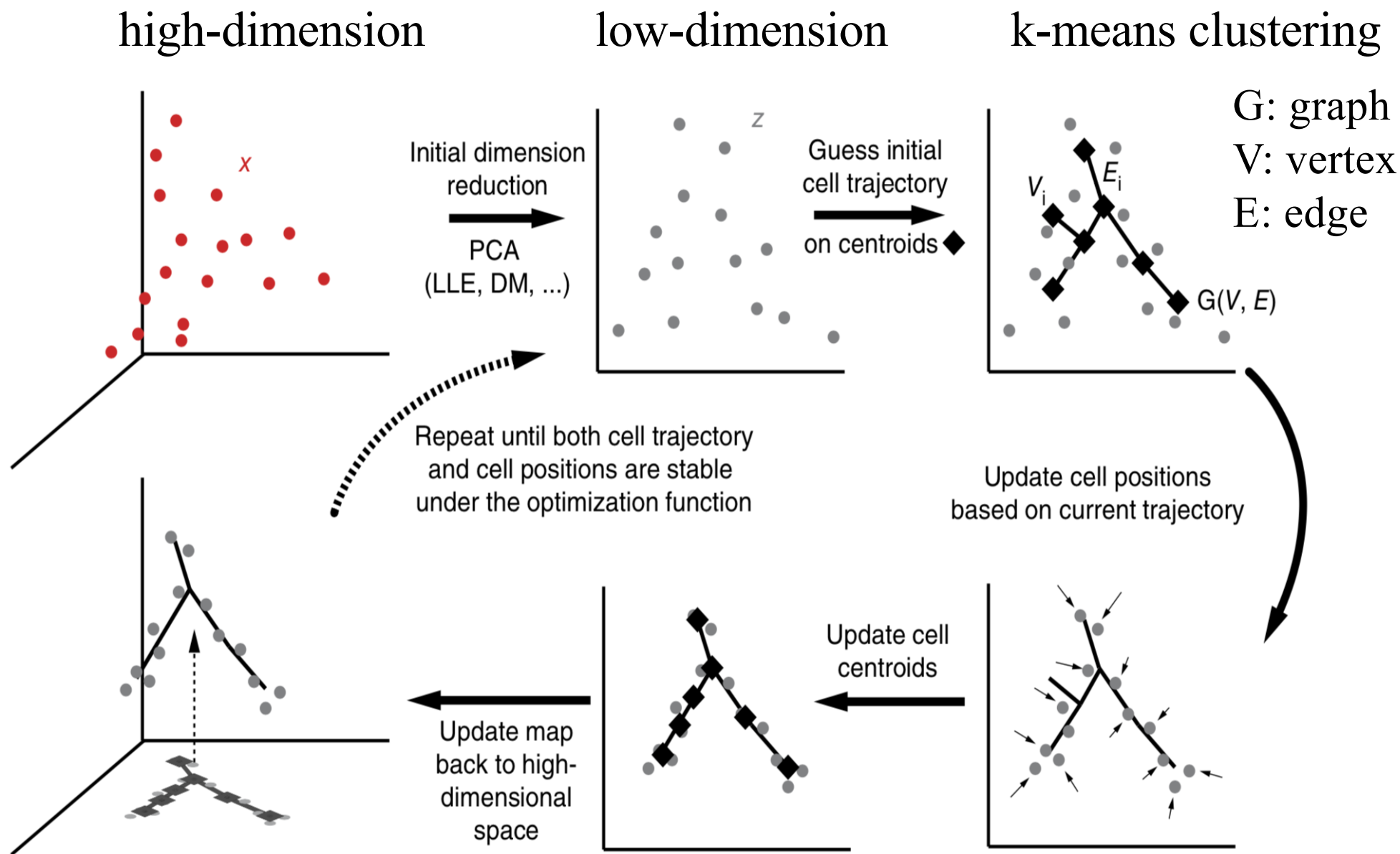
Comparison between Monocle 2 and 3

	Monocle 2	Monocle 3
Feature selection	suggested	possible
Clustering and RGE	parallel	sequential
Initial DR	PCA	UMAP
Graph	Tree	Graph
Partition	Single tree	Multiple graphs
Number of node	133	400
Plot	Projection from RGE	UMAP + trajectory

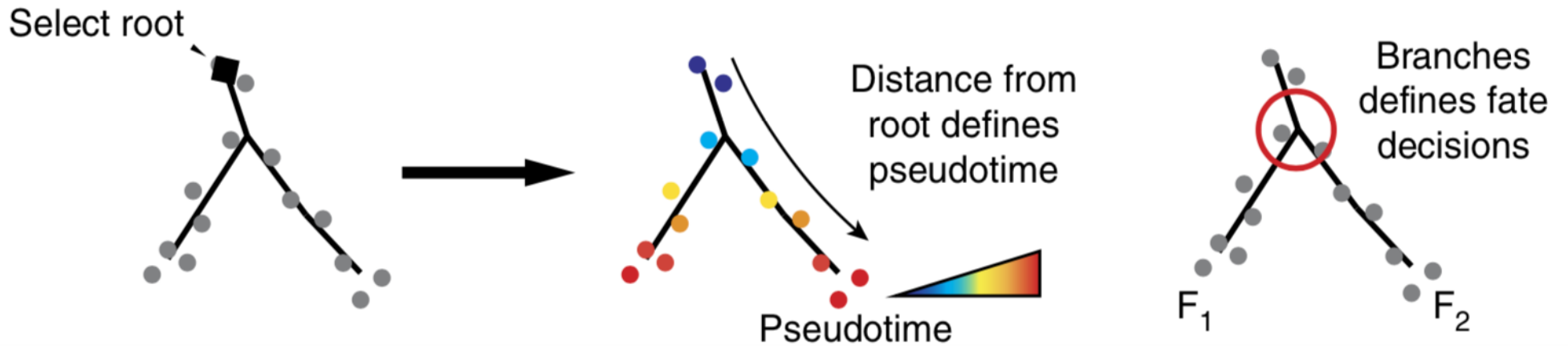
DR: dimension reduction

RGE: reversed graph embedding

Trajectory Analysis (Monocle 2)



Pseudotime (Monocle 2)



Reversed Graph Embedding

$$\text{Obj} = \text{obj}_1 + \lambda(\text{obj}_2) + \gamma(\text{obj}_3)$$

Minimizing objective function

λ : tuning parameter

γ : tuning parameter

Reversed Graph Embedding

$$\text{Obj}_1 = \|X - WZ\|^2$$

$\|A\|$: Frobenius norm = $(\sum_{ij} (a_{ij})^2)^{1/2}$

X : data matrix in high-dimension

Z : data matrix in low-dimension

W : rotation matrix, subject to $W^tW = I$

W^t : transposition of W

I : identity matrix

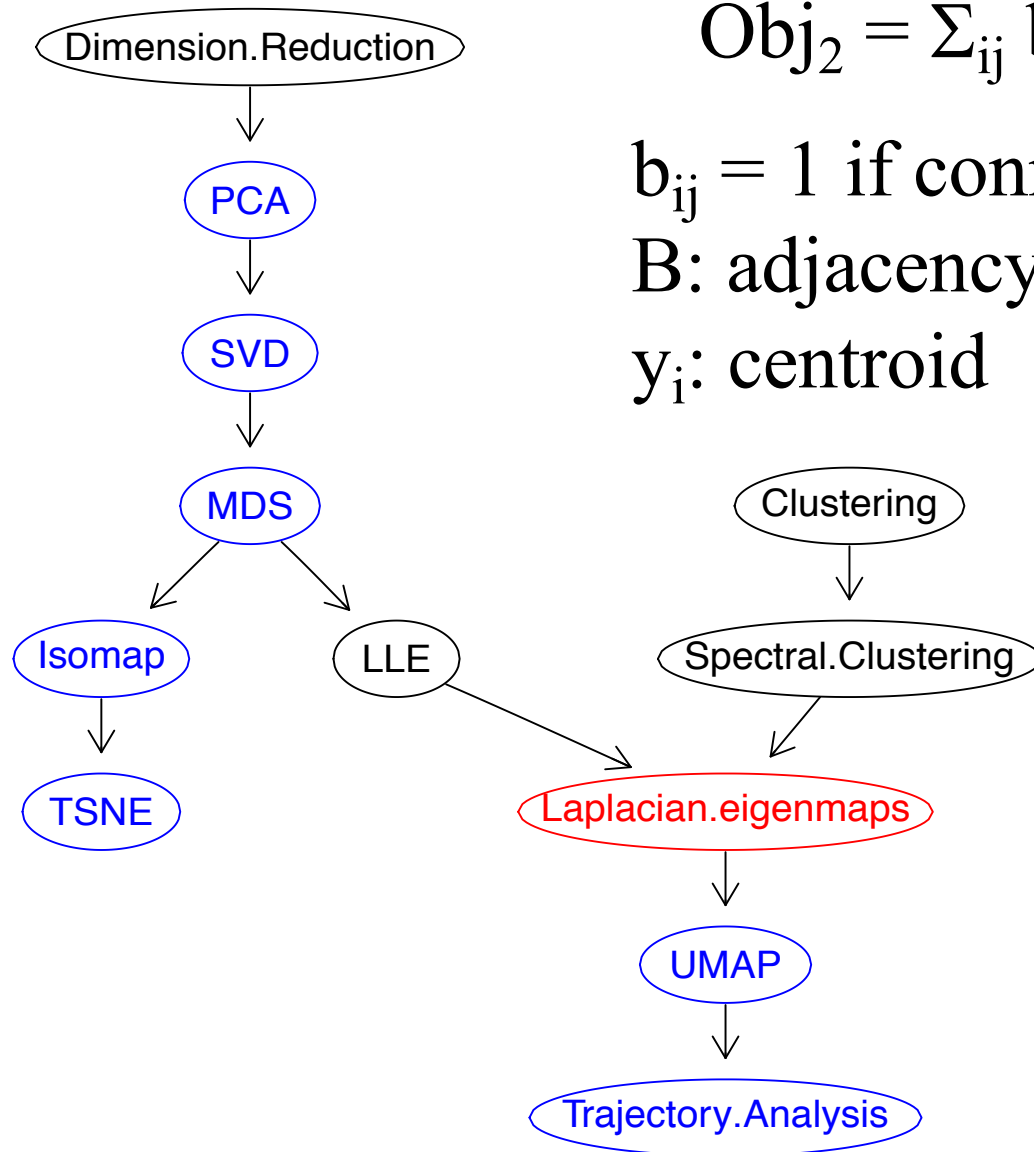
Reversed Graph Embedding

$$\text{Obj}_2 = \sum_{ij} b_{ij} \|y_i - y_j\|^2$$

$b_{ij} = 1$ if connected, 0 otherwise

B: adjacency matrix

y_i : centroid



Reversed Graph Embedding

$$\text{Obj}_3 = \sum_{i,k} r_{i,k} (\|z_i - y_k\|^2 + \sigma \log(r_{i,k}))$$

y_k : centroid

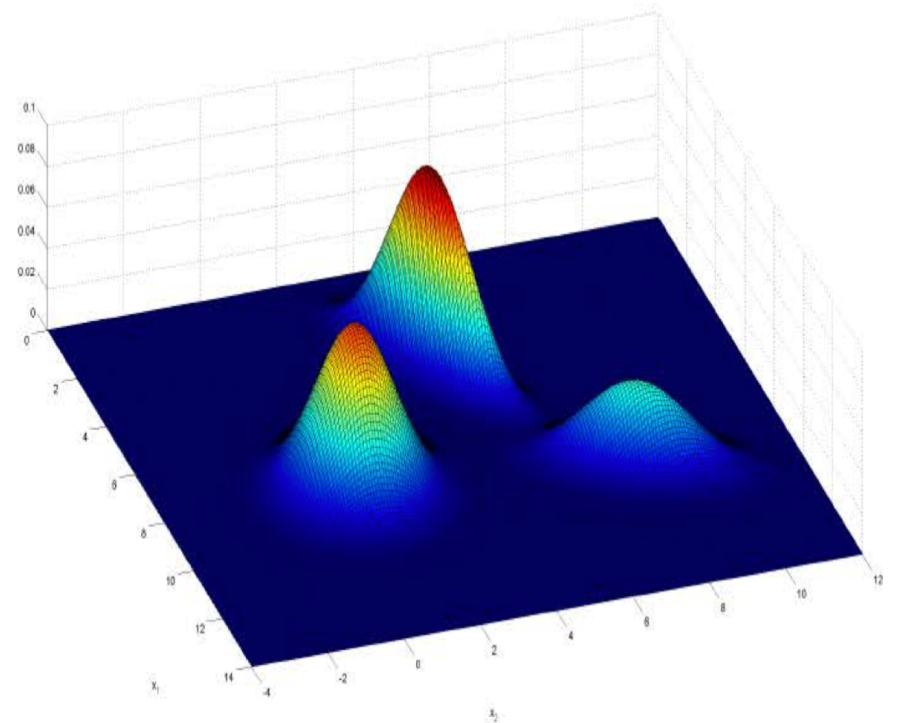
z_i : point in low-dimension

$r_{i,k}$: probability of i in cluster k

$$\sum_k r_{i,k} = 1$$

σ : regularization parameter

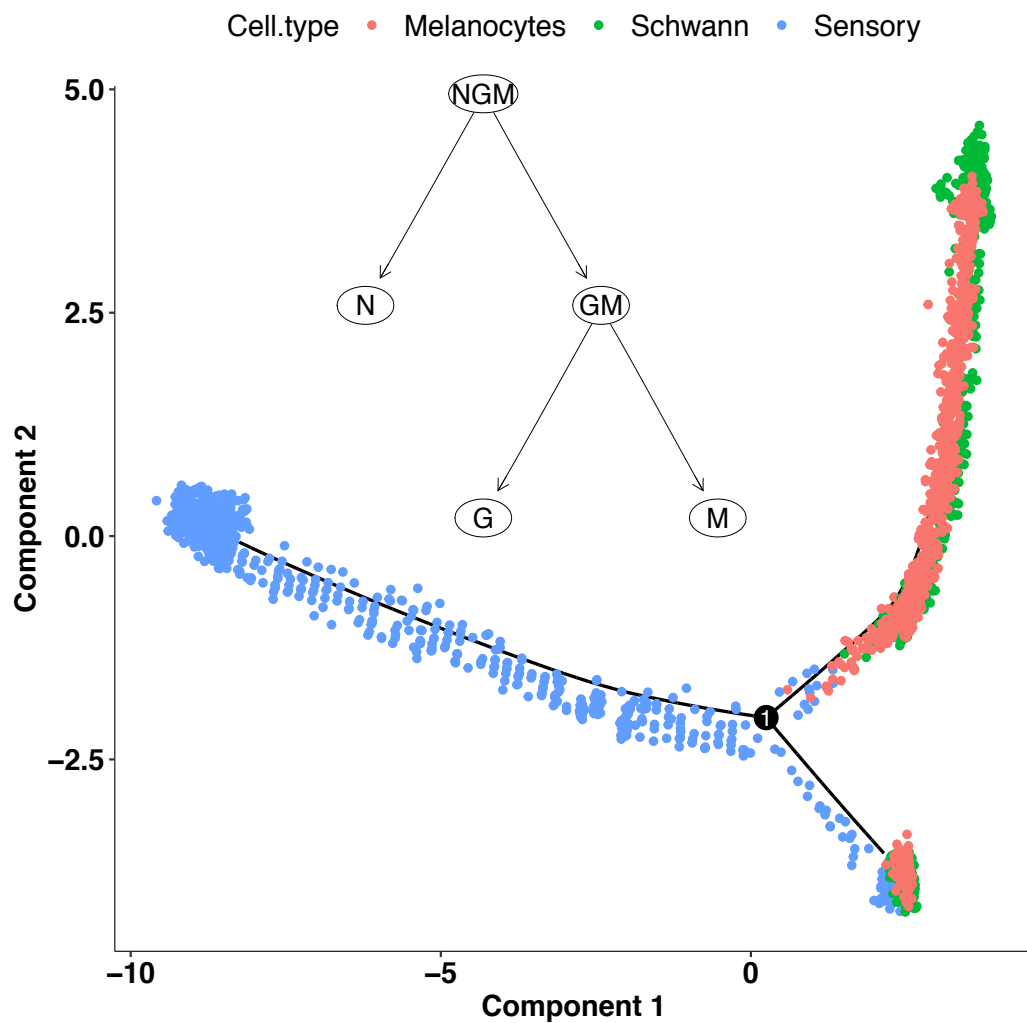
modified k-means clustering



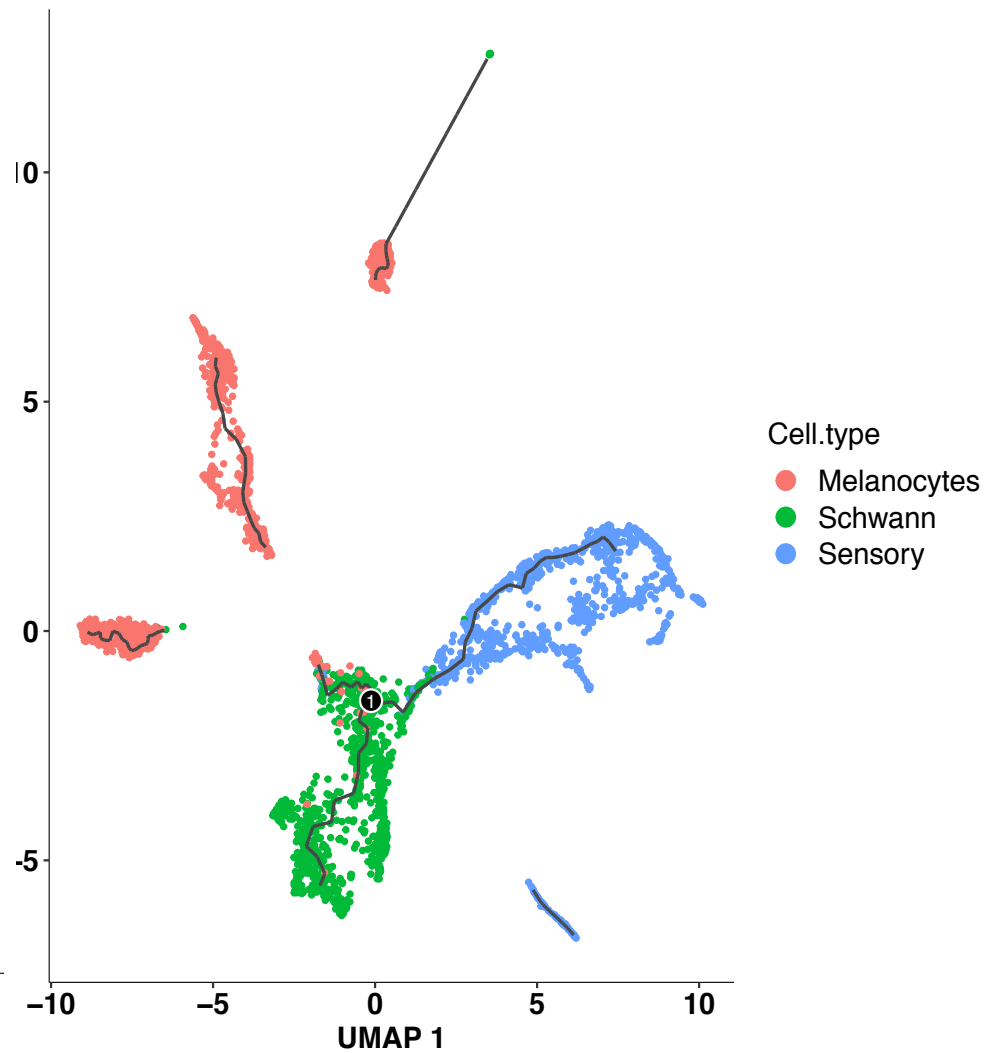
Gaussian Mixture Model (GMM)

Monocle 2 or Monocle 3?

Monocle 2

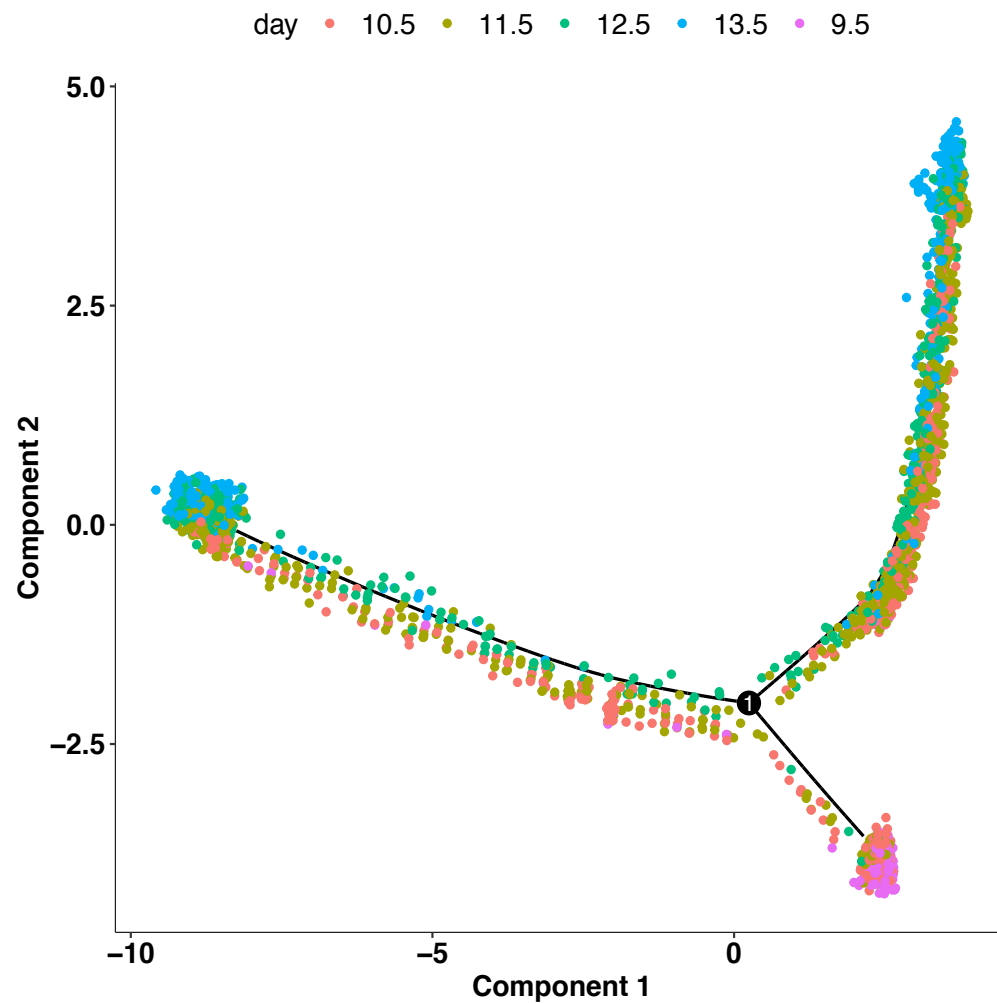


Monocle 3

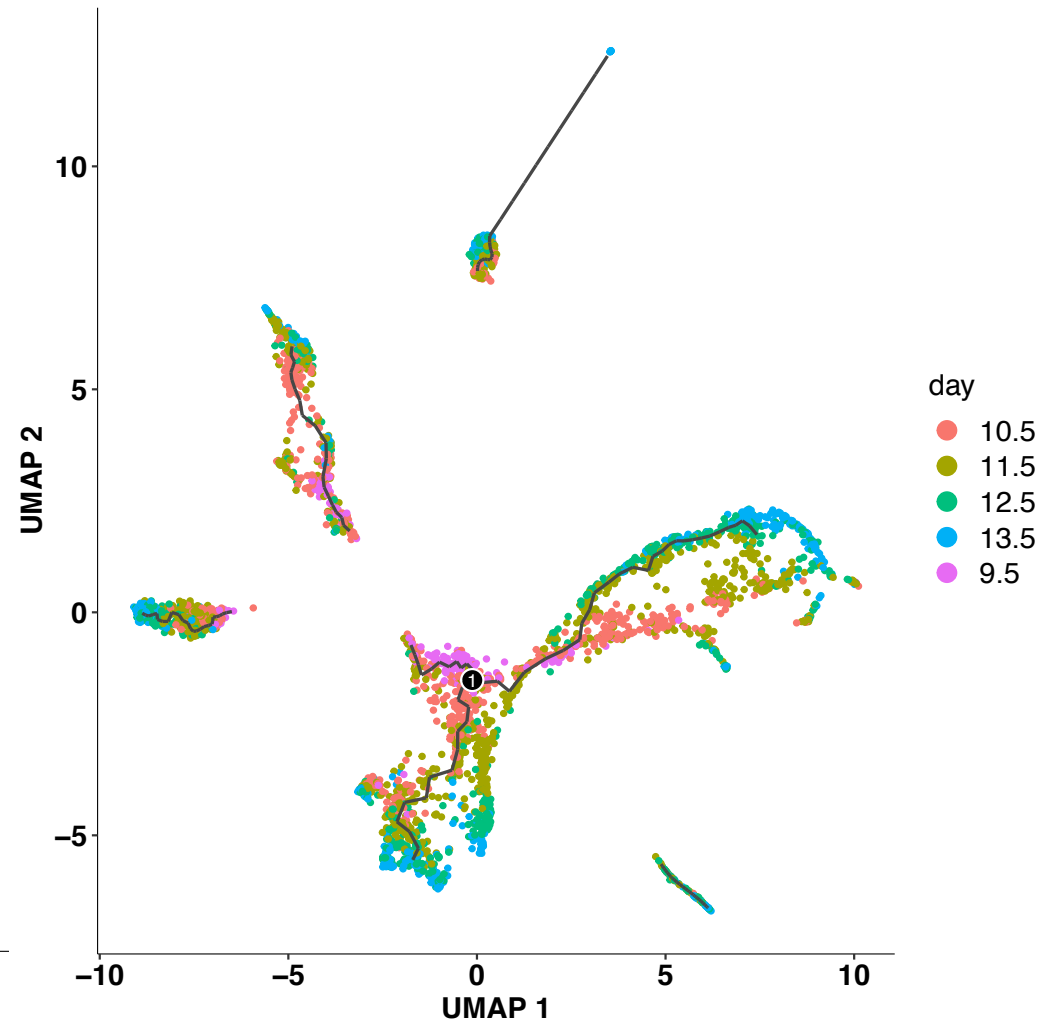


Monocle 2 or Monocle 3?

Monocle 2



Monocle 3



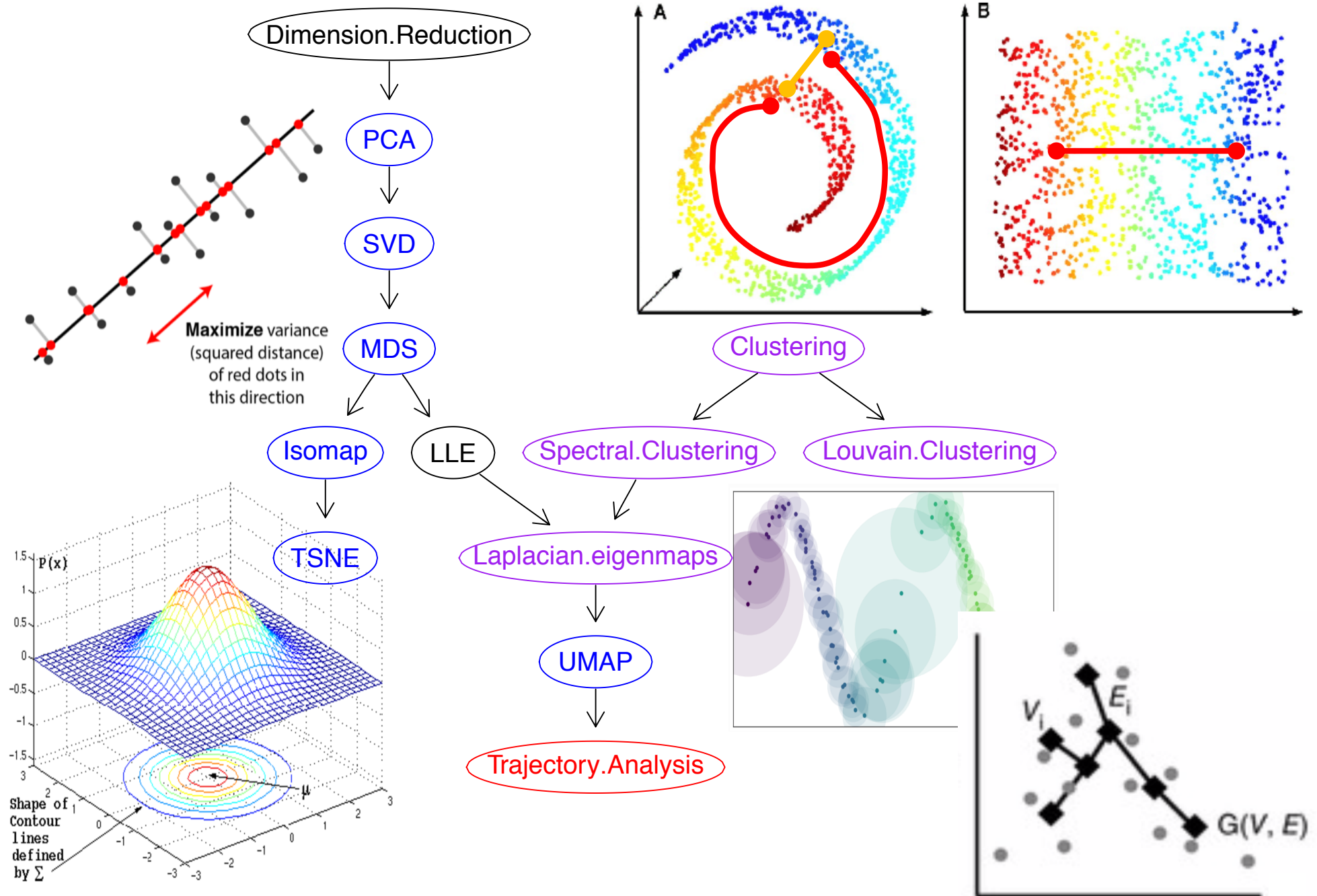
Law of Parsimony (Occam's Razor)

Everything should be made as simple as possible,
but not simpler.

Albert Einstein



Summary of Dimension Reduction Methods



Acknowledgement

- Christina Stuelten
- Emily Tai
- Yoshimi Greer
- Eva Perez Guijarro
- Chi-Ping Day