Interactive Visual Analysis with Dimensionality Reduction

Thomas Höllt - Computer Graphics & Visualization - TU Delft

Visualization

Present

- "*Everything*" known about the data,

⇨ Visualization used for **Communication** of results

Visualization

Present

- "*Everything*" known about the data,

➯ Visualization used for **Communication** of results

Analyze

- There are hypotheses

➯ Visualization used **Verification** or **Falsification**

Visualization

Present

- "*Everything*" known about the data,

➫ Visualization used for **Communication** of results

Analyze

- There are hypotheses

➫ Visualization used **Verification** or **Falsification**

Explore

- *Nothing* is known

➫ Visualization used for data **Exploration**

Visualization

Present

- “Ever

➩ Visua

Analyze

- There

➩ Visua

Explore

- *Nothing* is known

➩ Visualization used for data **Exploration**

Visualization

Present

- “Ever

➩ Visua

Analyze

- There

➩ Visua

Explore

- *Nothing* is known

➩ Visualization used for data **Exploration**

The case for interactive visual analysis

Numbers do not tell the whole story...

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Anscombe's Quartet

Statistics:

Mean(x):

Variance(x):

Mean(y):

Variance(y):

Correlation(x,y):

Numbers do not tell the whole story...

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

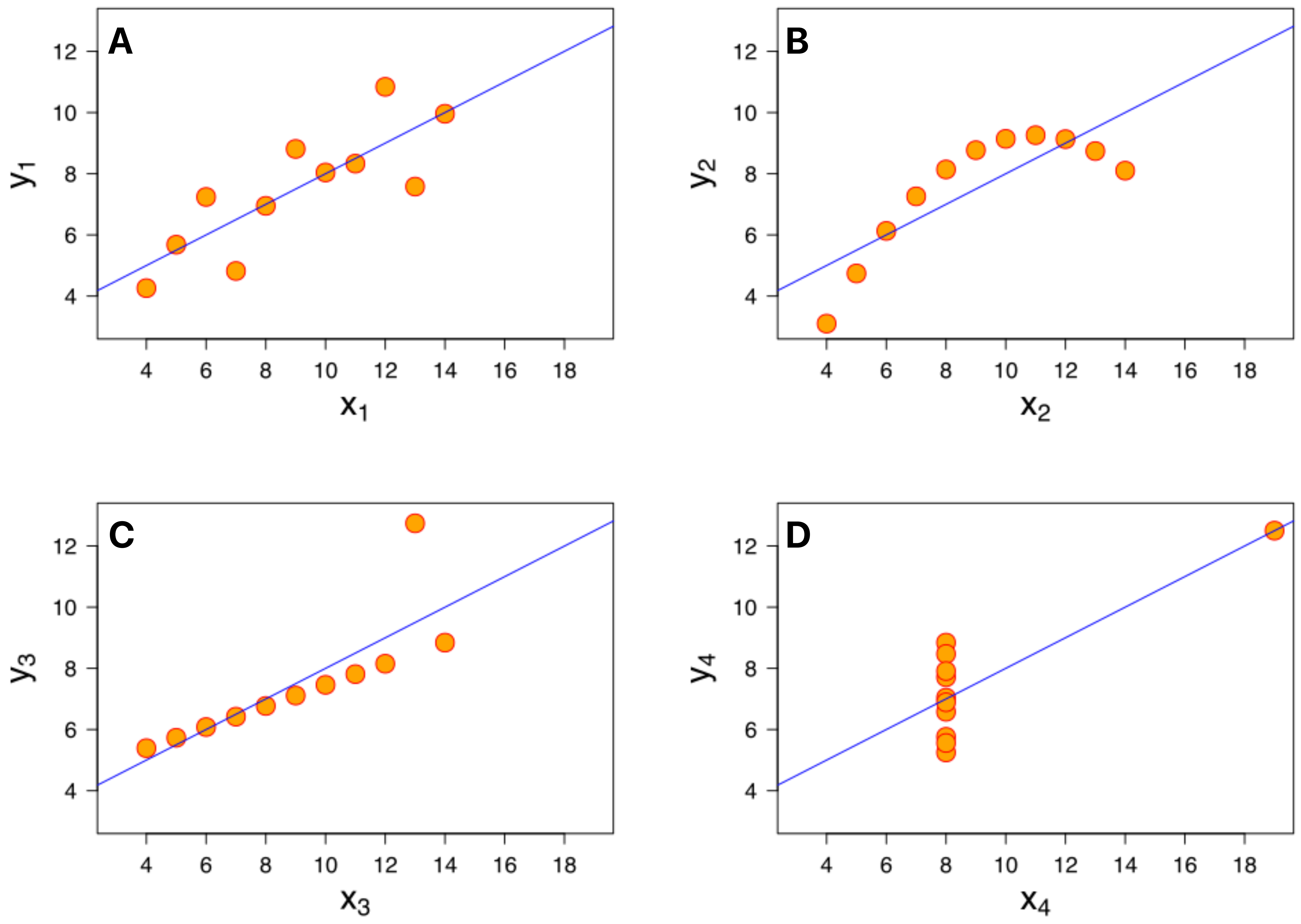
Anscombe's Quartet

Statistics:

Mean(x):	A=B=C=D=9	exact
Variance(x):	A=B=C=D=11	exact
Mean(y):	A=B=C=D=7.50	(2 decimals)
Variance(y):	A=B=C=D=4.125	(+/- 0.003)
Correlation(x,y):	A=B=C=D=816	(3 decimals)

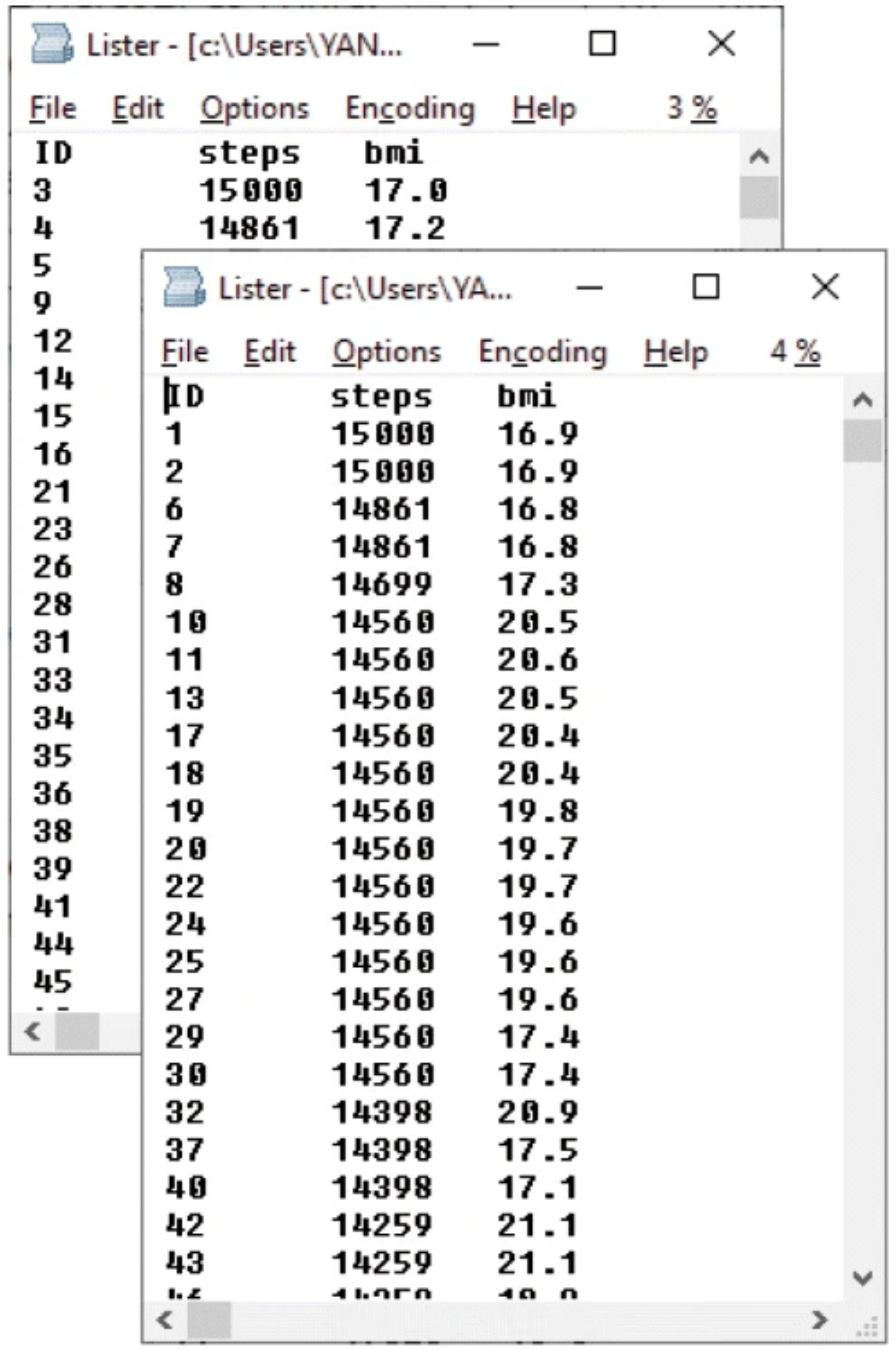
Numbers do not tell the whole story...

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



Anscombe's Quartet

Do we look for the whole story?



Lister - [c:\Users\YAN...

File Edit Options Encoding Help 3 %

ID	steps	bmi
3	15000	17.0
4	14861	17.2
5		
9		
12		
14		
15		
16		
21		
23		
26		
28		
31		
33		
34		
35		
36		
38		
39		
41		
44		
45		

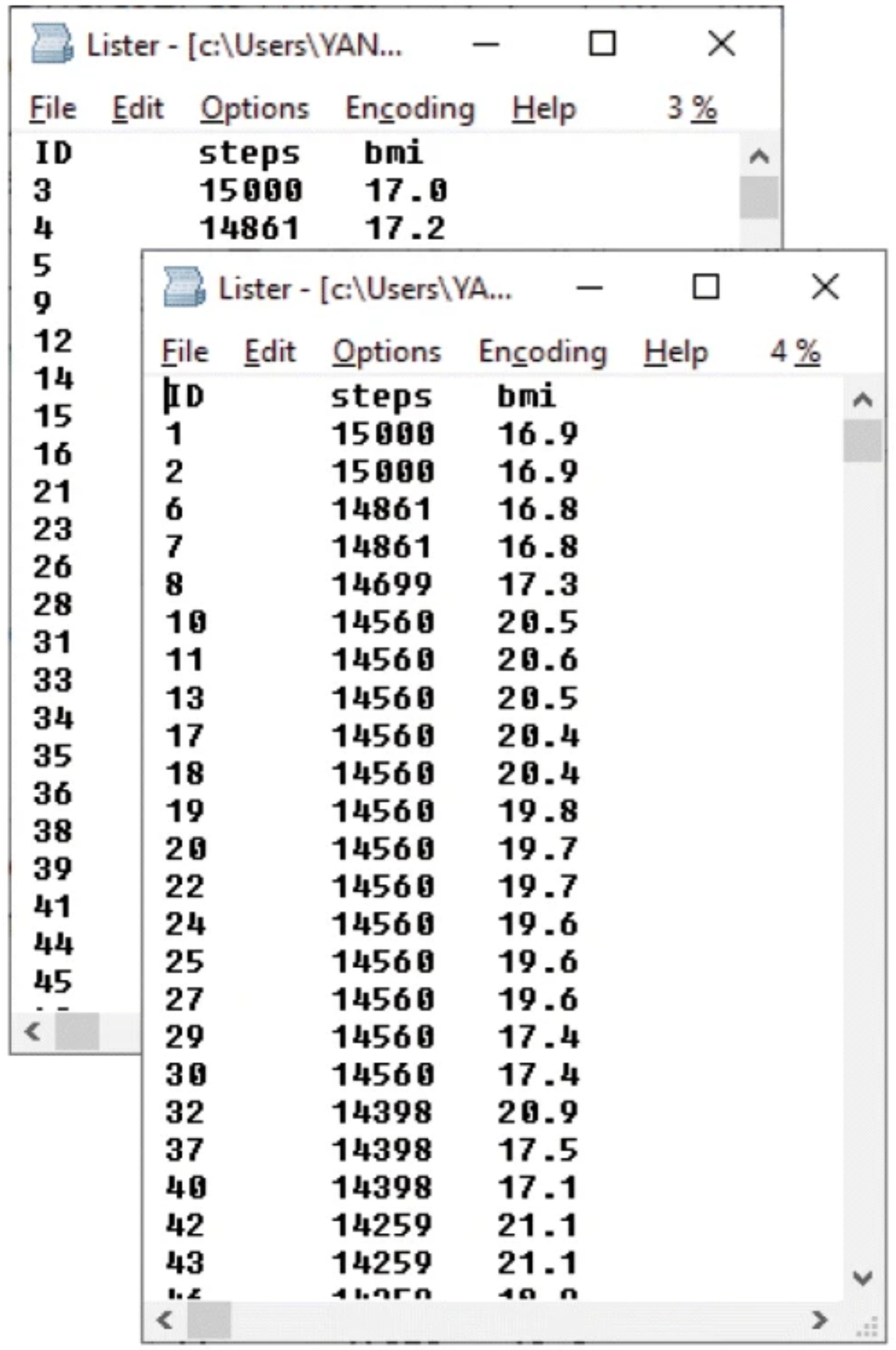
Lister - [c:\Users\YA...

File Edit Options Encoding Help 4 %

ID	steps	bmi
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1

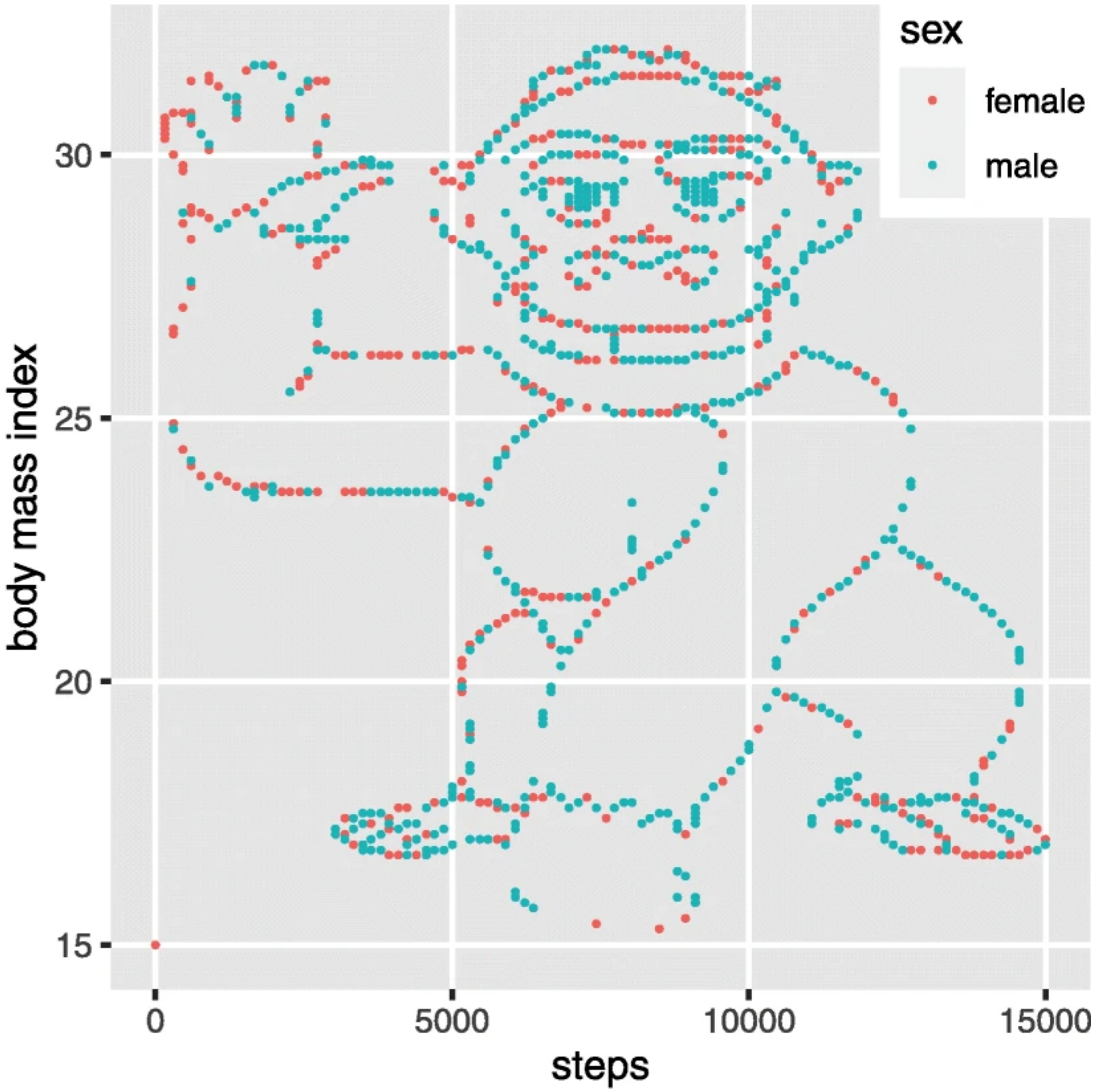
Yanai and Lercher, BMC Genome Biology, 2020

Do we look for the whole story?



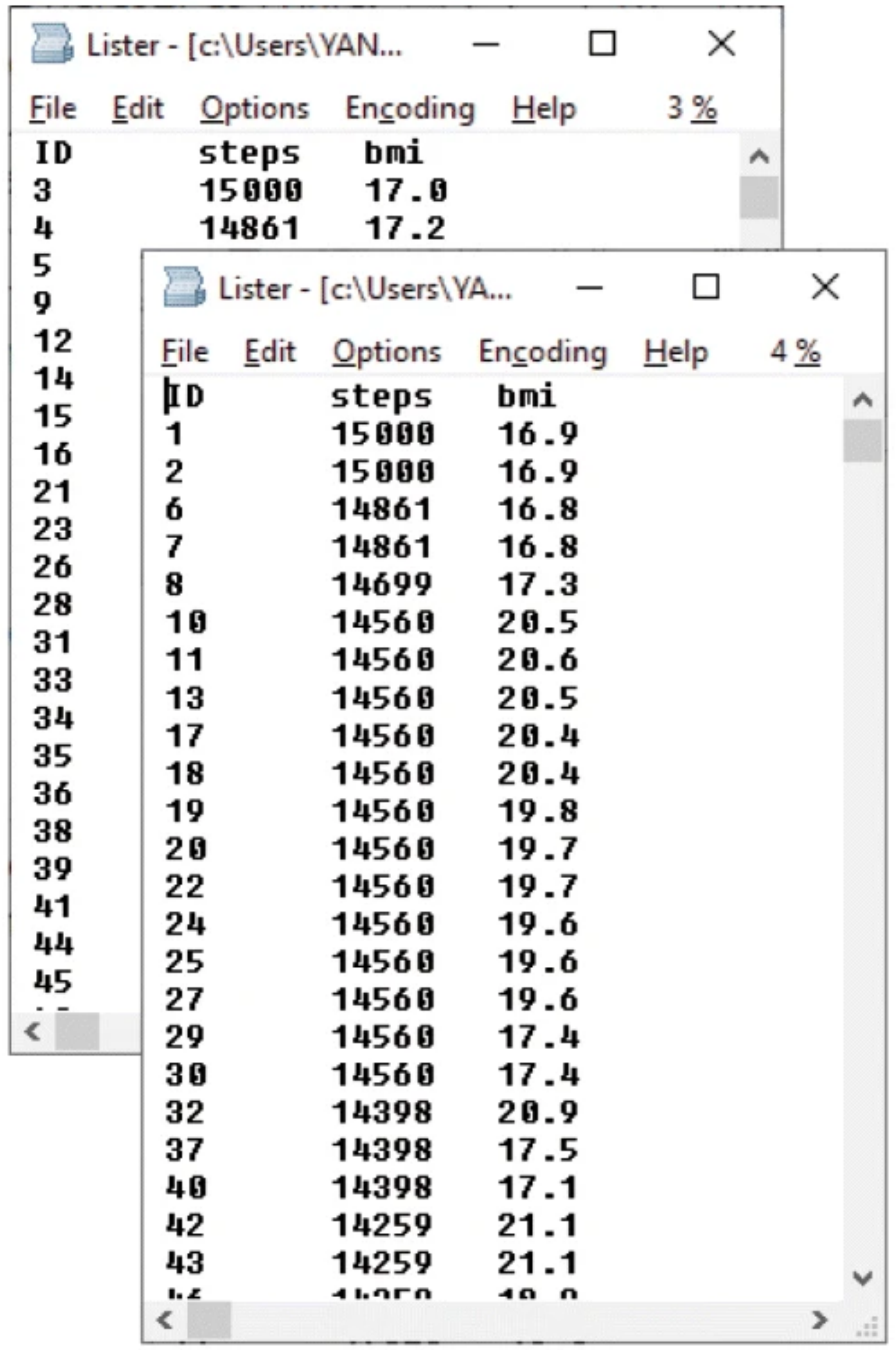
ID	steps	bmi
3	15000	17.0
4	14861	17.2
5		
9		
12		
14		
15		
16		
21		
23		
26		
28		
31		
33		
34		
35		
36		
38		
39		
41		
44		
45		

ID	steps	bmi
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1



Yanai and Lercher, BMC Genome Biology, 2020

Do we look for the whole story?



Lister - [c:\Users\YAN...

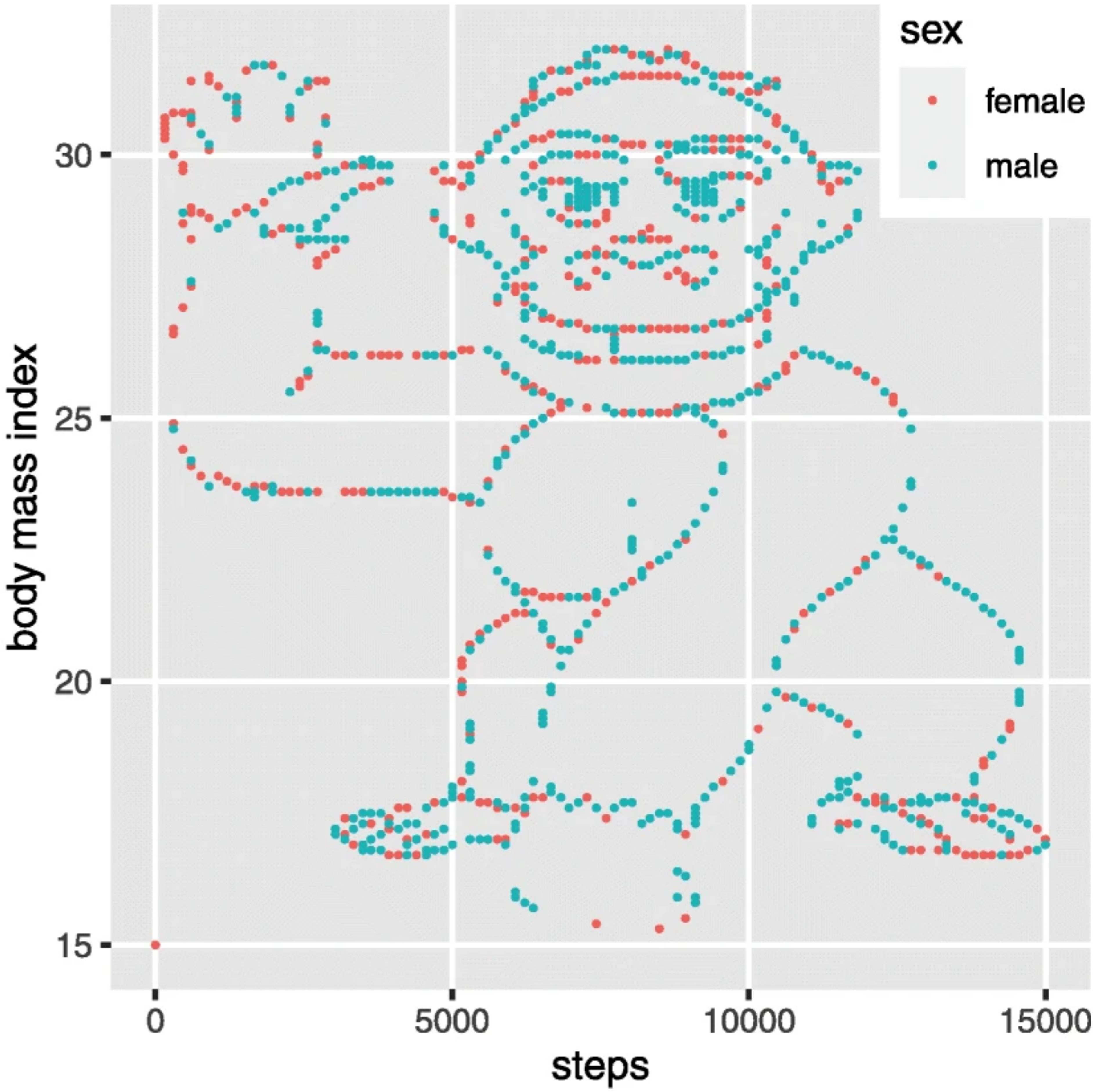
File Edit Options Encoding Help 3 %

ID	steps	bmi
3	15000	17.0
4	14861	17.2
5		
9		
12		
14		
15		
16		
21		
23		
26		
28		
31		
33		
34		
35		
36		
38		
39		
41		
44		
45		

Lister - [c:\Users\YA...

File Edit Options Encoding Help 4 %

ID	steps	bmi
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1



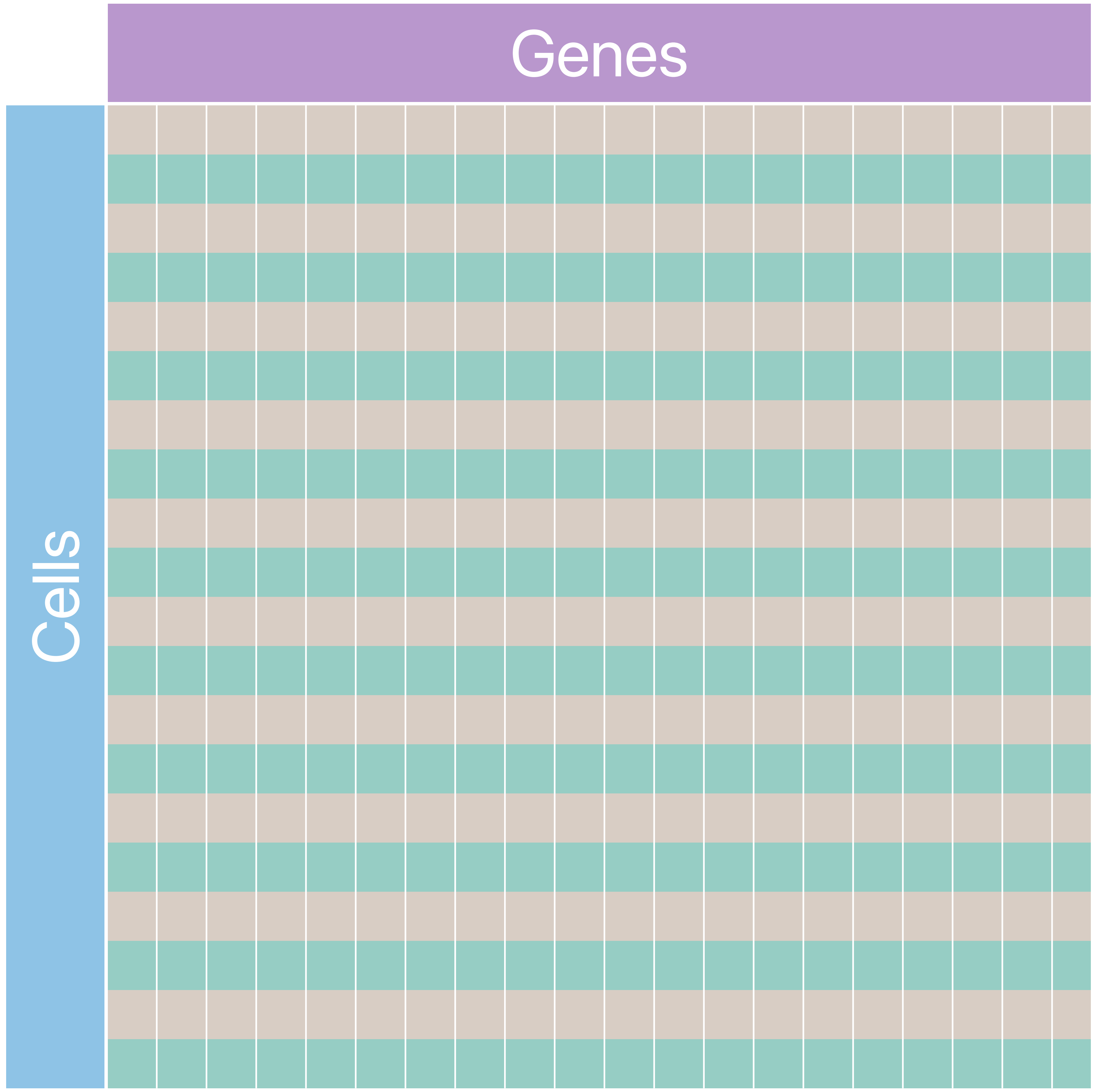
🦍 / Hypothesis	🚫	✅
✅	14	5
🚫	5	9

Yanai and Lercher, BMC Genome Biology, 2020

Single-Cell Visualization

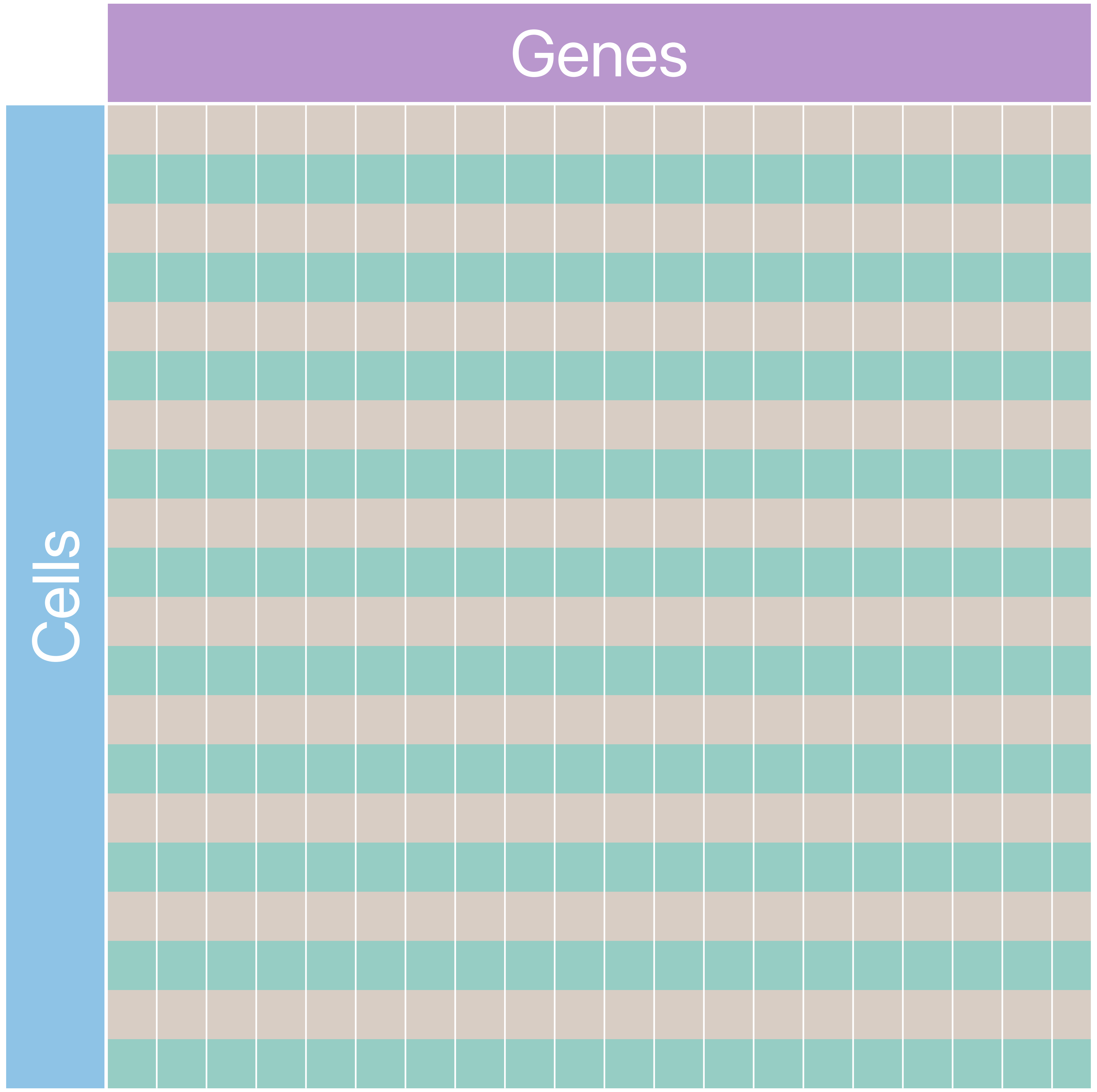
Dimensionality Reduction?

- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - Clustering
  - Dimensionality Reduction



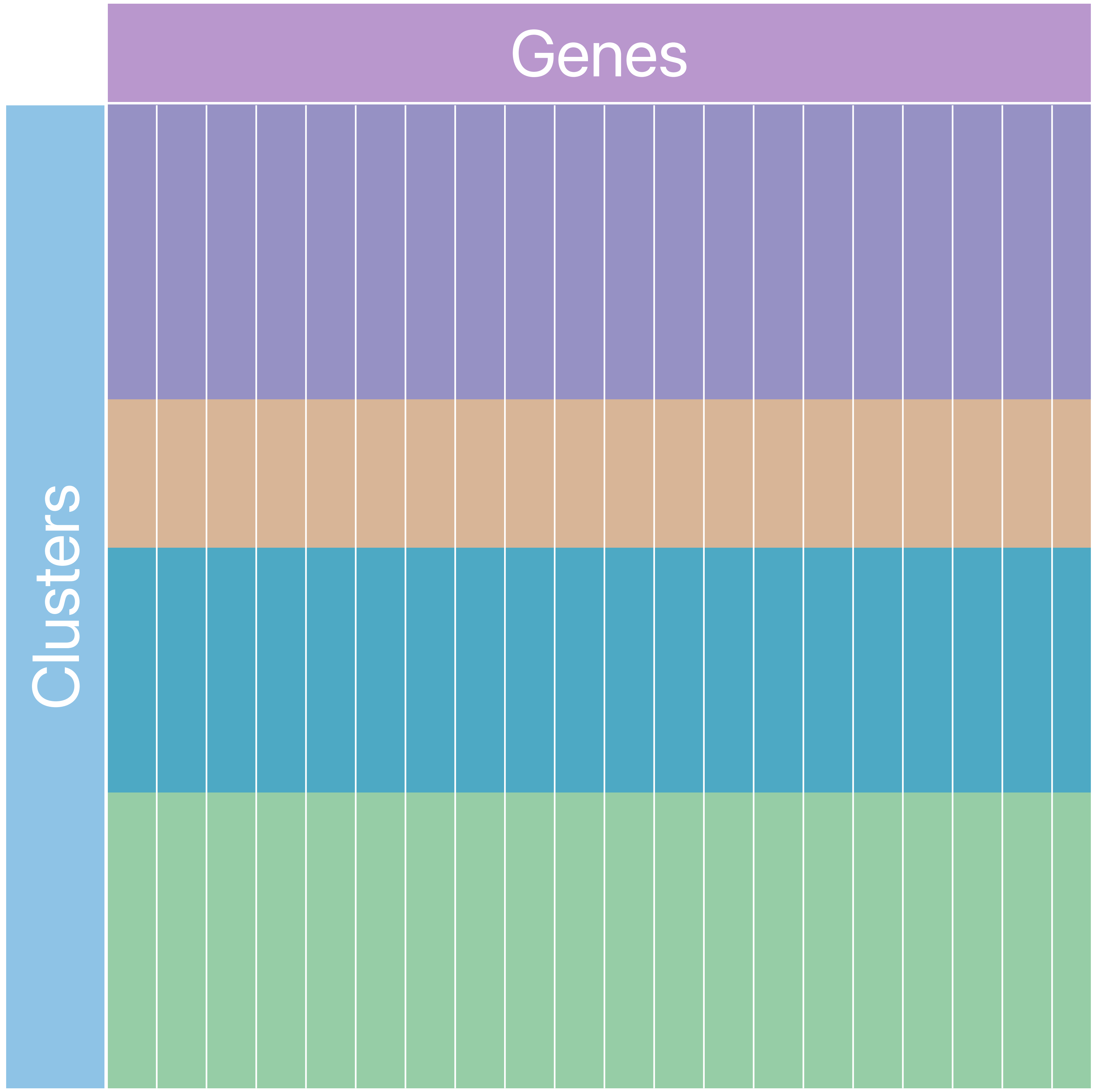
Dimensionality Reduction?

- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - **Clustering**
  - Dimensionality Reduction



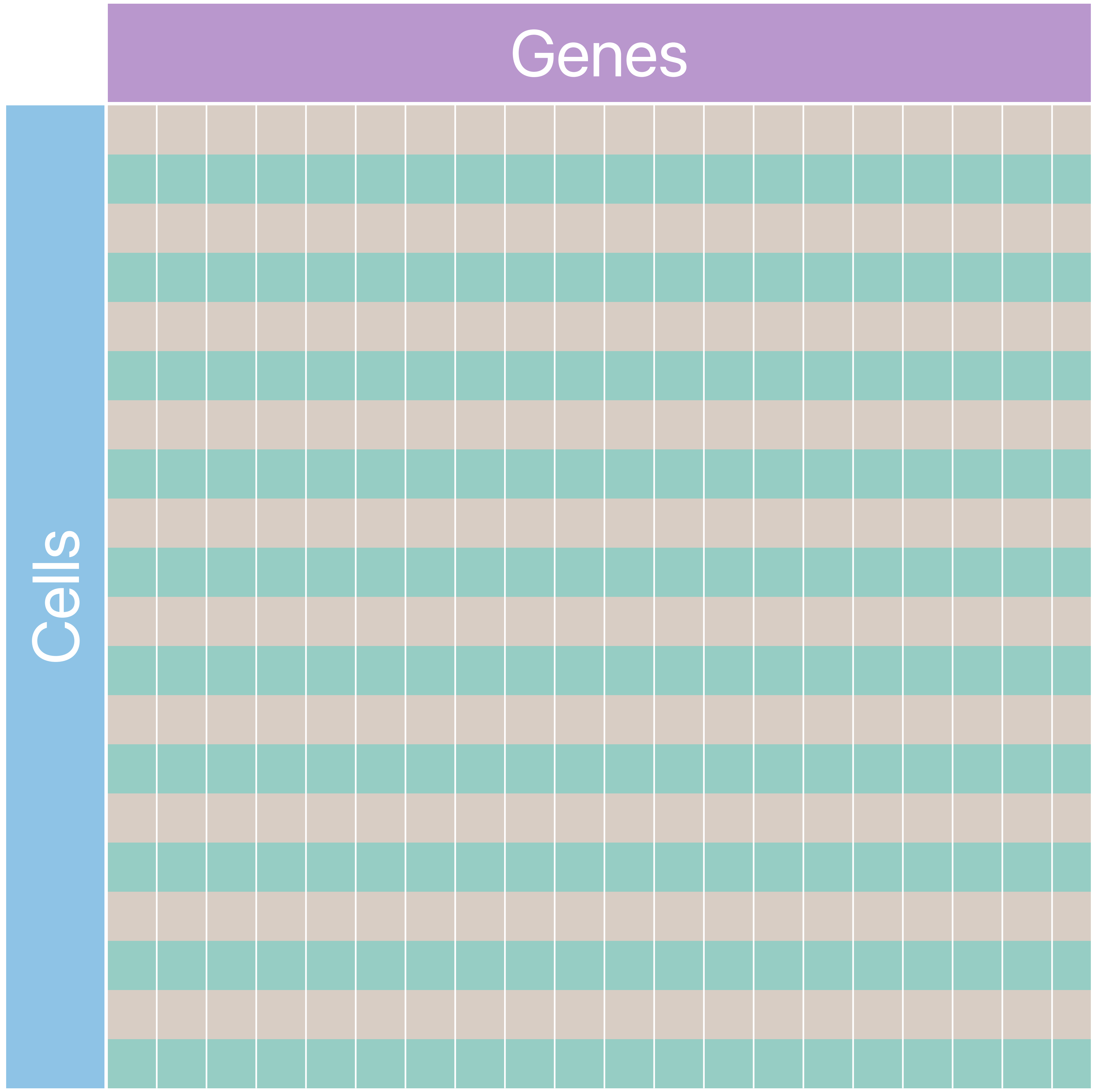
Dimensionality Reduction?

- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - **Clustering**
  - Dimensionality Reduction



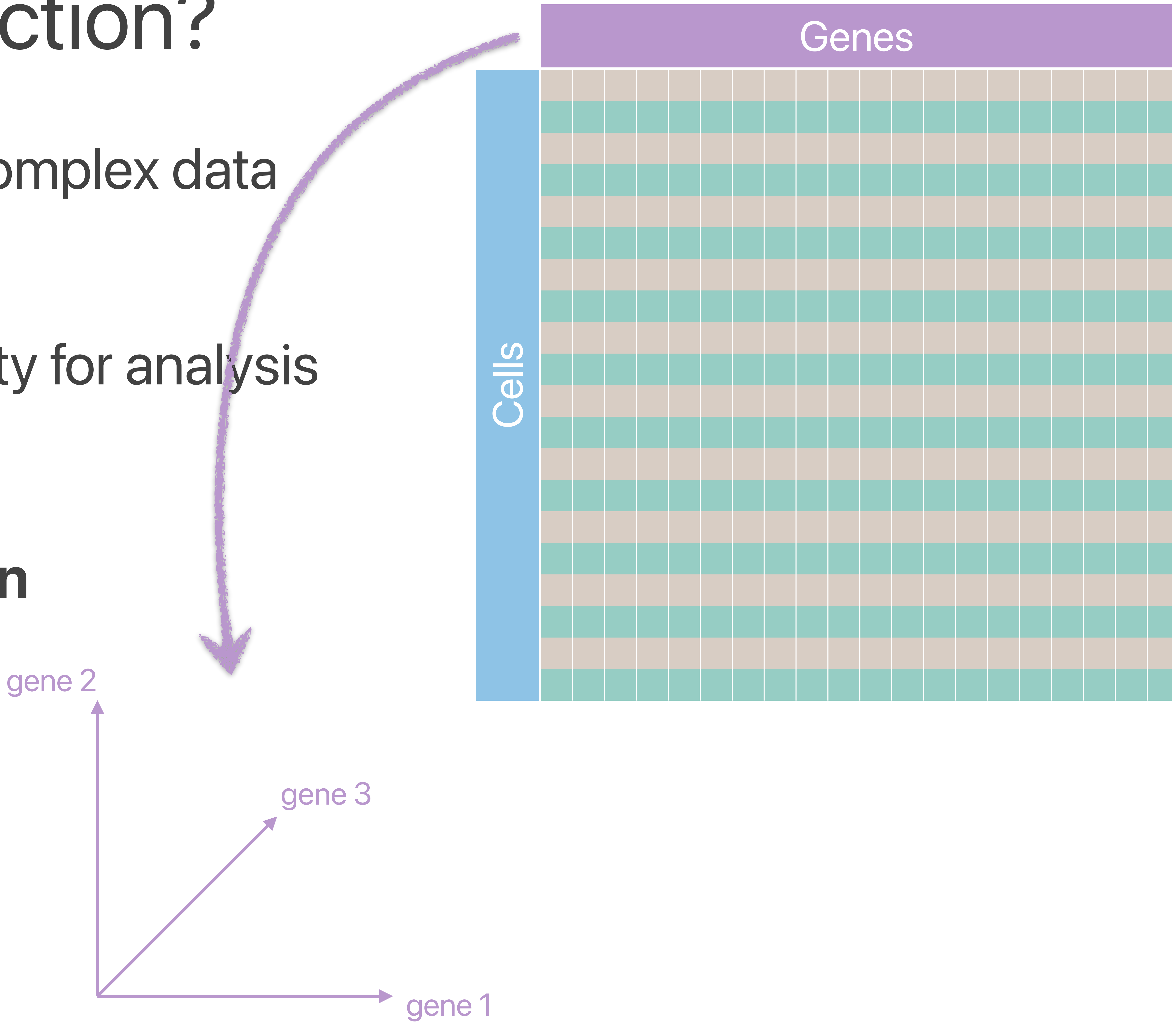
Dimensionality Reduction?

- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - Clustering
  - **Dimensionality Reduction**



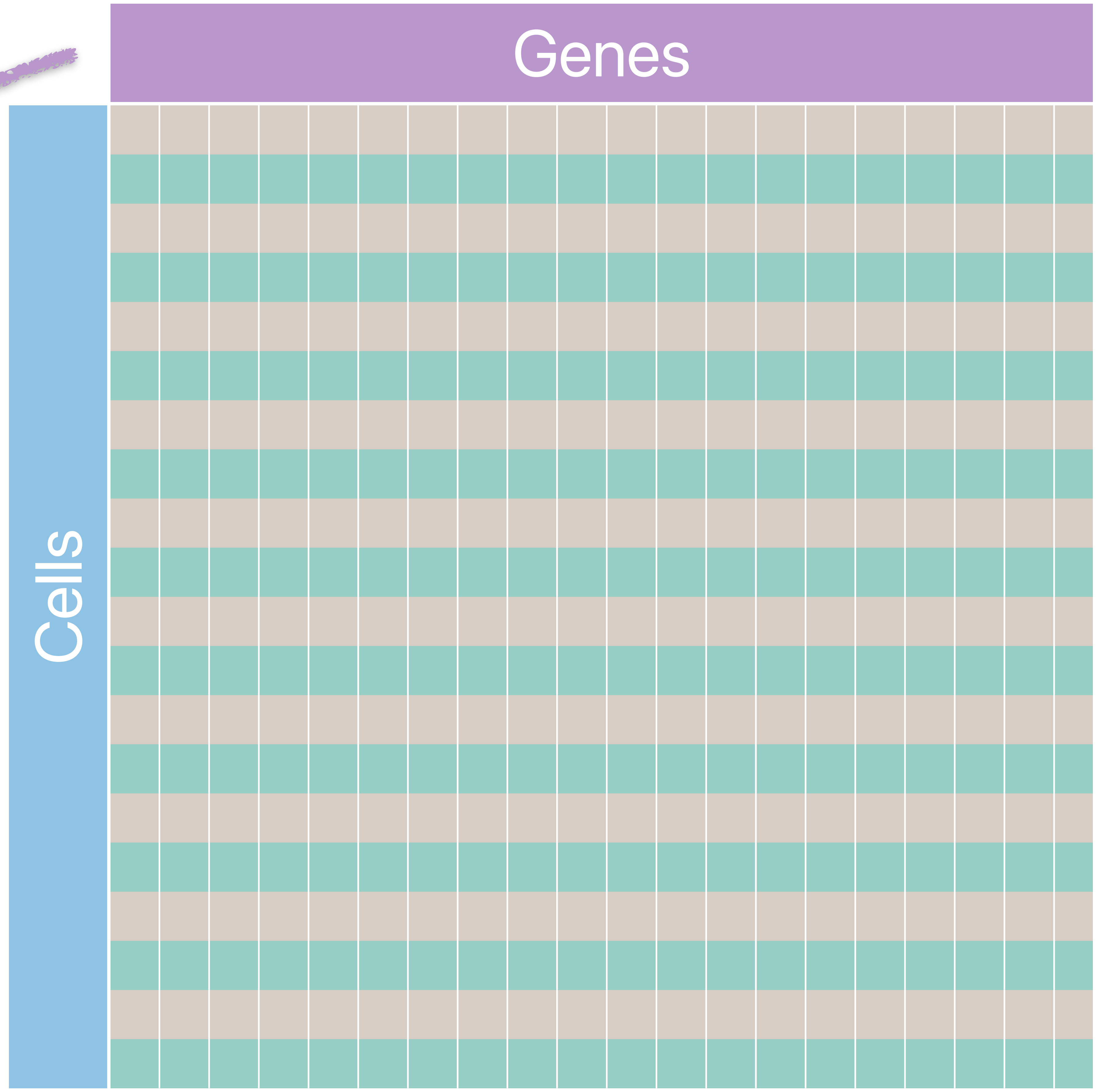
Dimensionality Reduction?

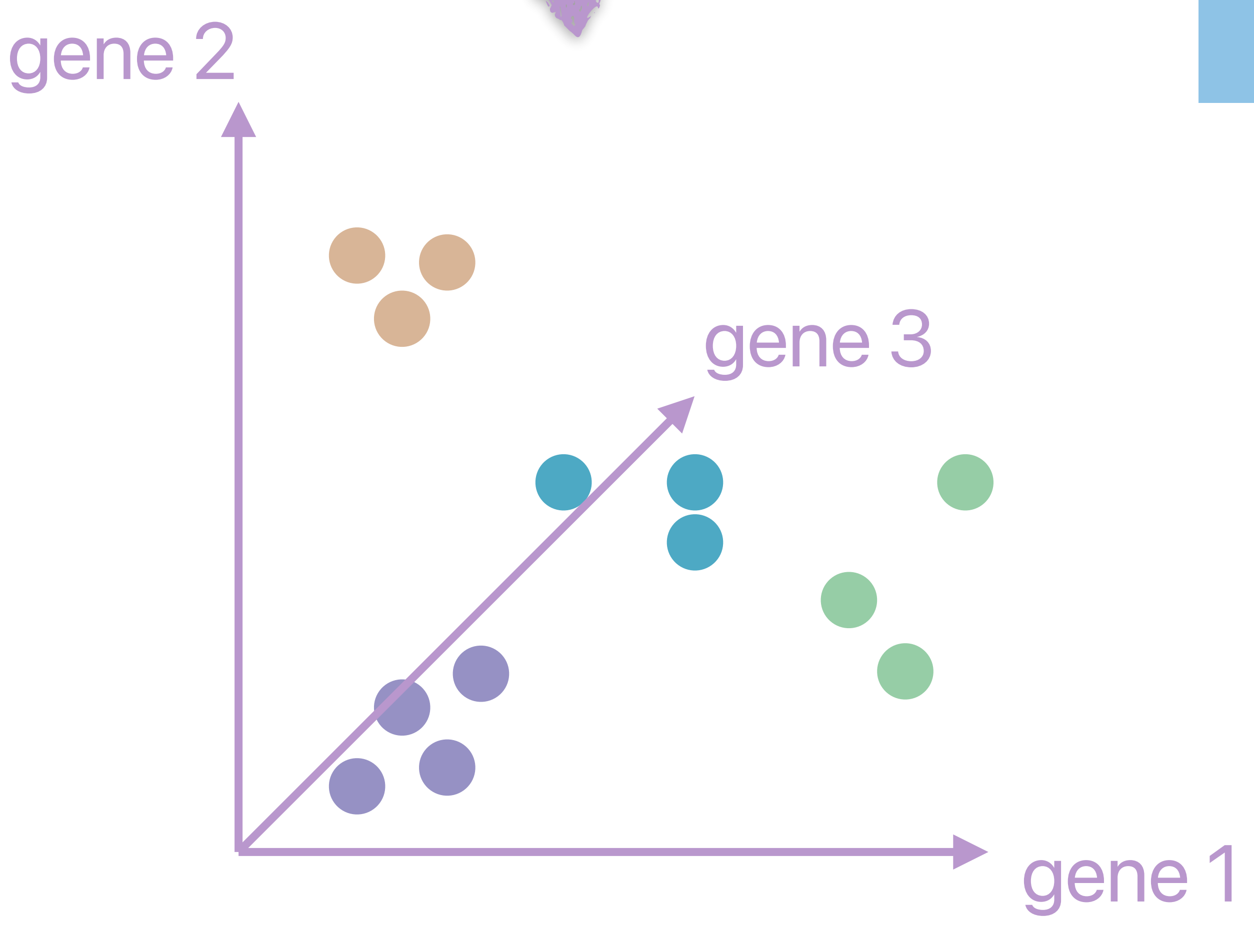
- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - Clustering
  - **Dimensionality Reduction**



Dimensionality Reduction?

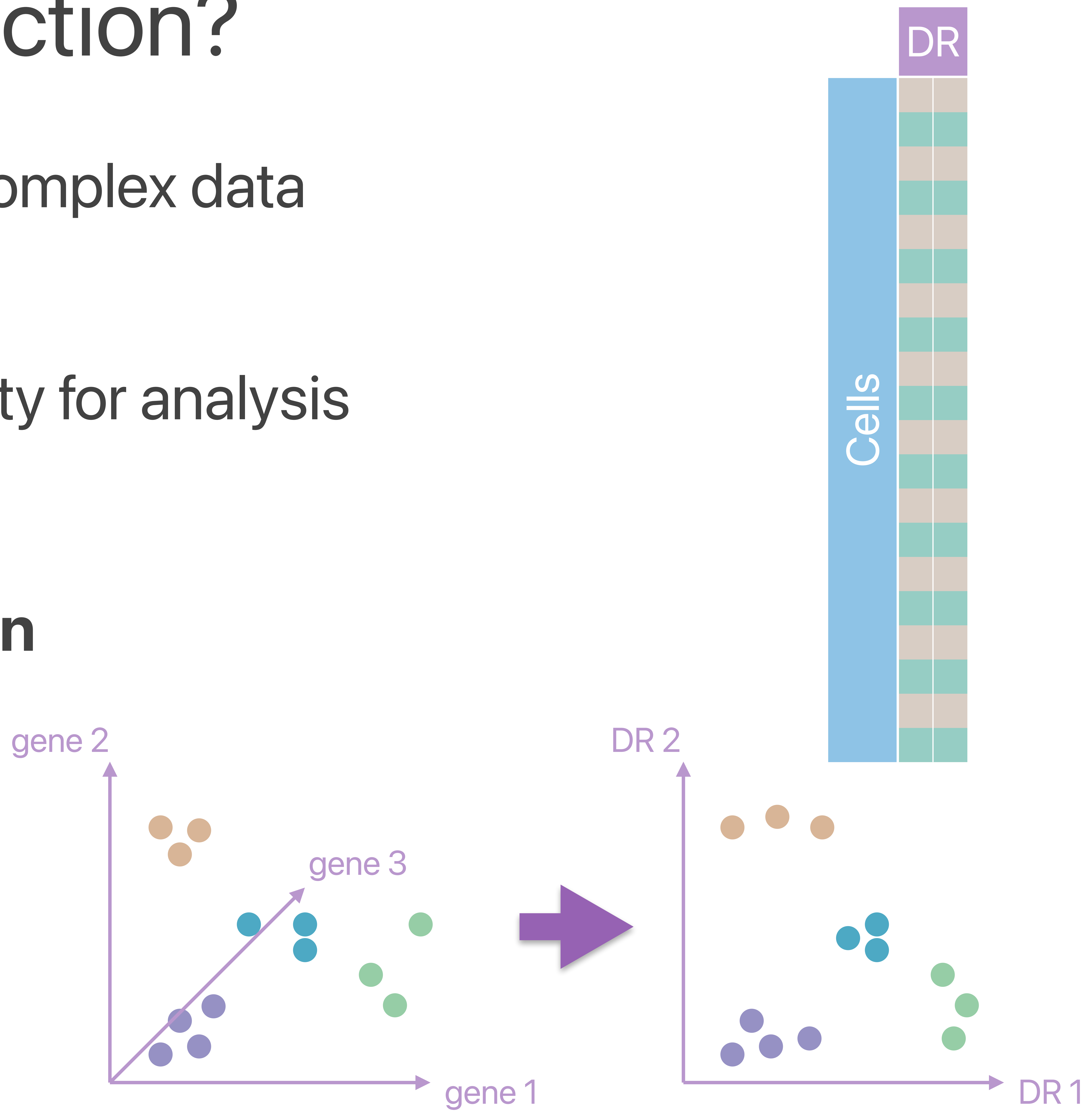
- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - Clustering
  - **Dimensionality Reduction**





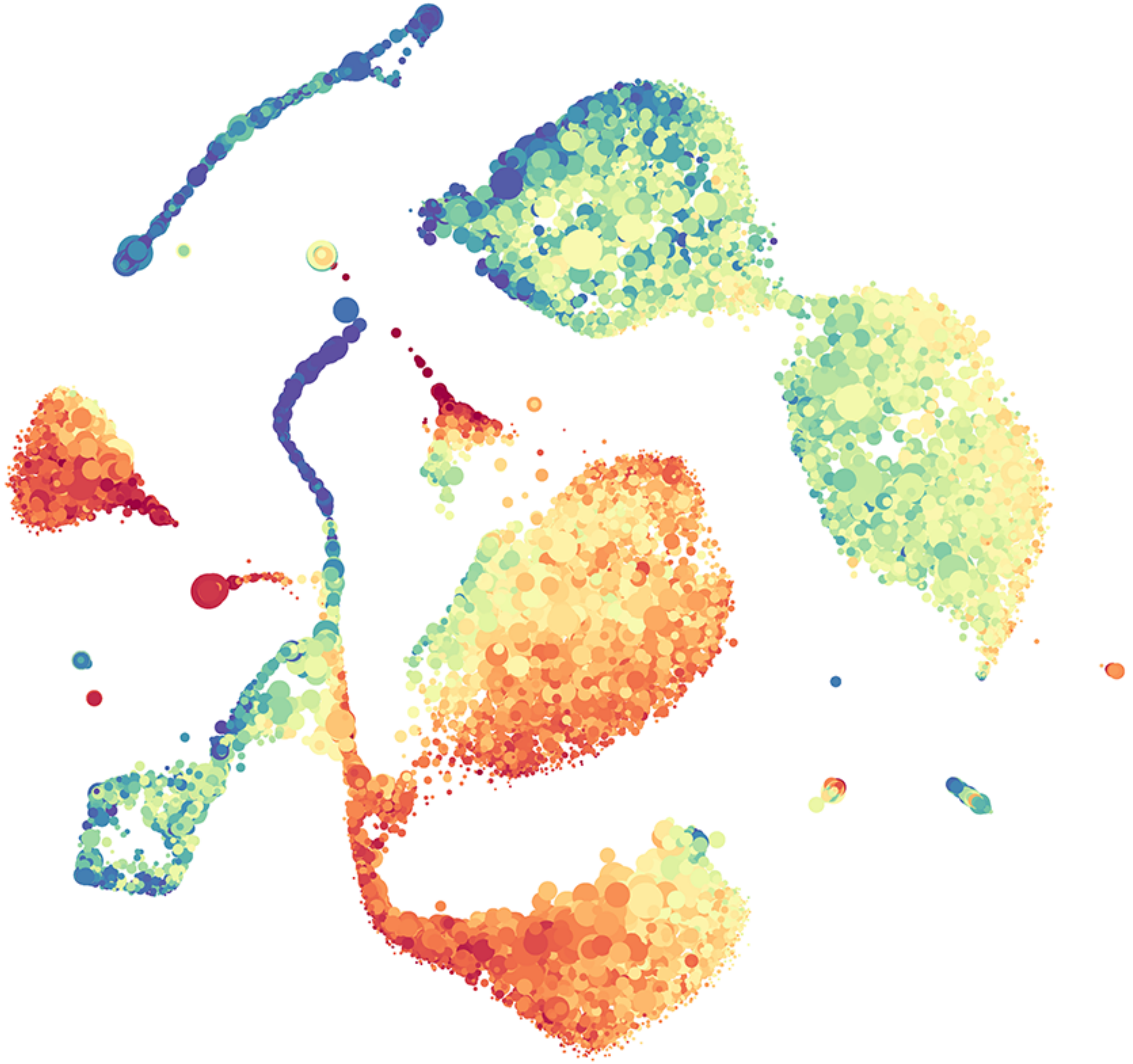
Dimensionality Reduction?

- We have huge amounts of complex data (many cells x many genes)
- We want to reduce complexity for analysis
  - Clustering
  - **Dimensionality Reduction**



Why Visualization for Data Exploration?

- Can't machines do (learn) that?
- Machine learning is great for
  - Well defined problems
  - Verifying Hypothesis
- ML not so great for
  - Finding the unknown
  - Fuzzy problems
  - Hypothesis generation



Algorithms

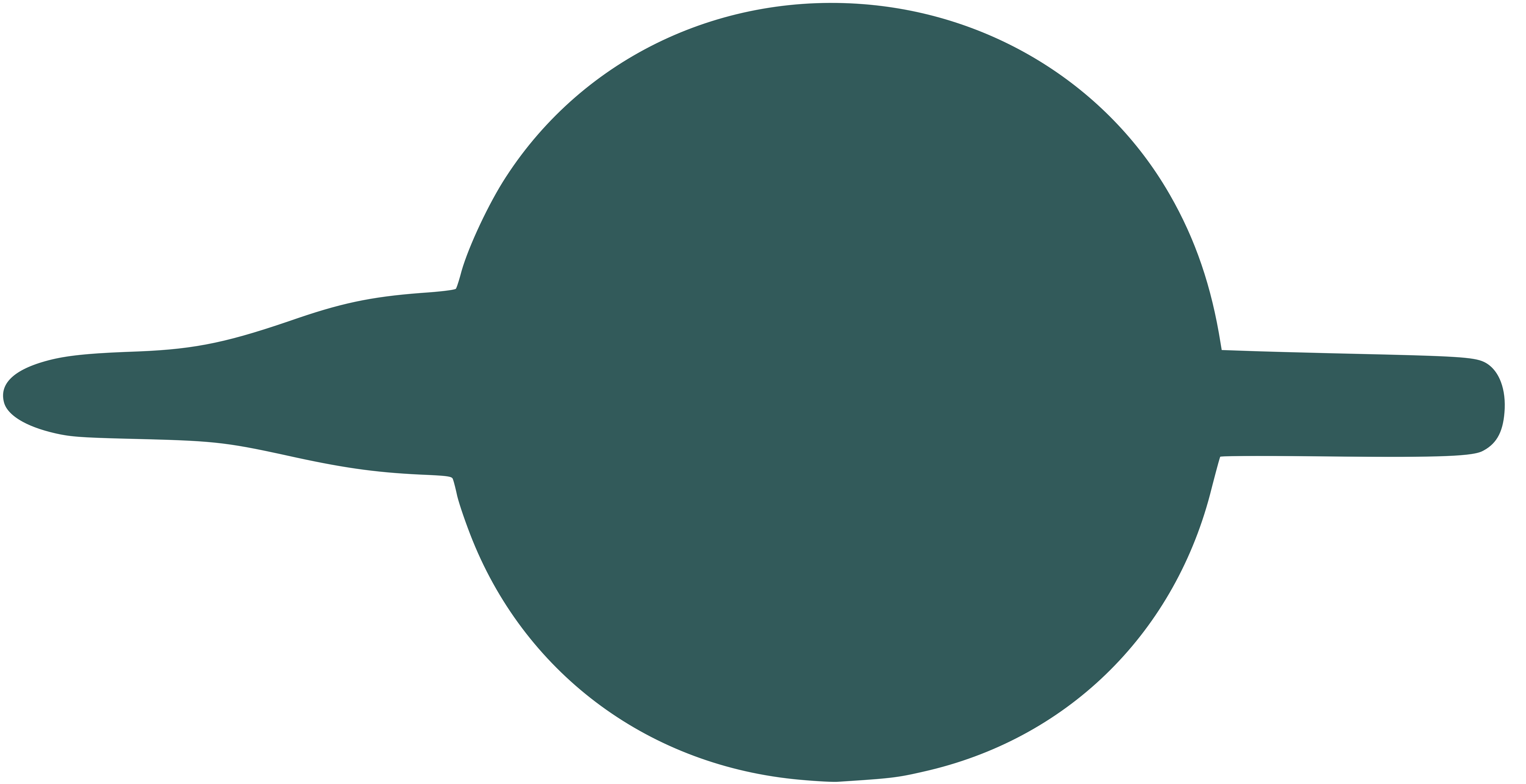
PCA	linear	Matrix Factorization	
ICA	linear	Matrix Factorization	
MDS	non-linear	Matrix Factorization	
cPCA	non-linear	Matrix Factorization	https://doi.org/10.1038/s41467-018-04608-8
ZIFA	non-linear	Matrix Factorization	https://doi.org/10.1186/s13059-015-0805-z
ZINB-WaVE	non-linear	Matrix Factorization	https://doi.org/10.1038/s41467-017-02554-5

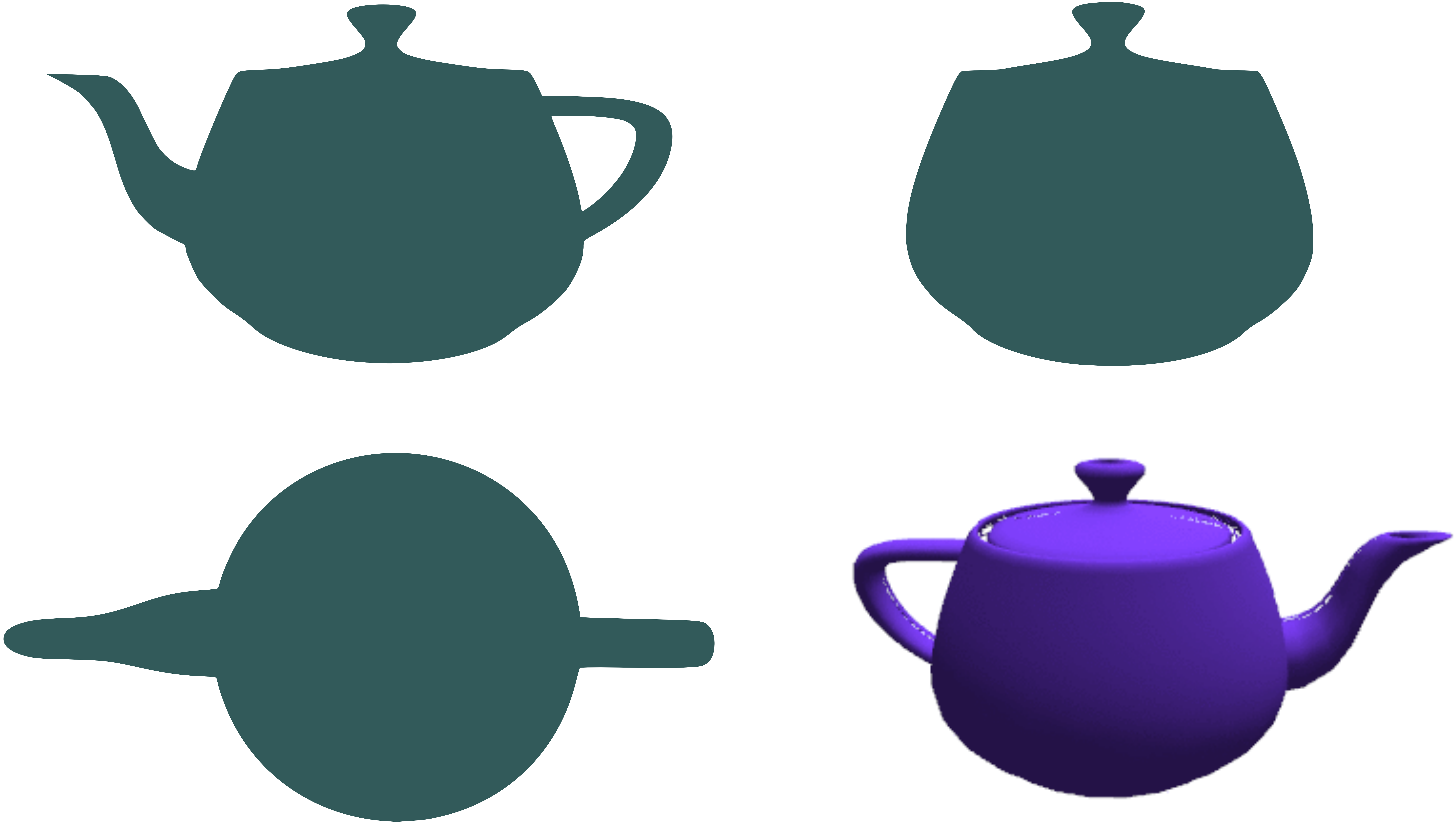
Diffusion maps	non-linear	Graph-based	https://doi.org/10.1073/pnas.0500334102
Isomap	non-linear	Graph-based	https://doi.org/10.1126/science.290.5500.2319
t-SNE	non-linear	Graph-based	https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf
HSNE	non-linear	Graph, hierarchical	https://dx.doi.org/10.1038/s41467-017-01689-9
LargeVis	non-linear	Graph-based	arXiv:1602.00370
UMAP	non-linear	Graph-based	arXiv:1802.03426
PHATE	non-linear	Graph-based	https://doi.org/10.1101/120378

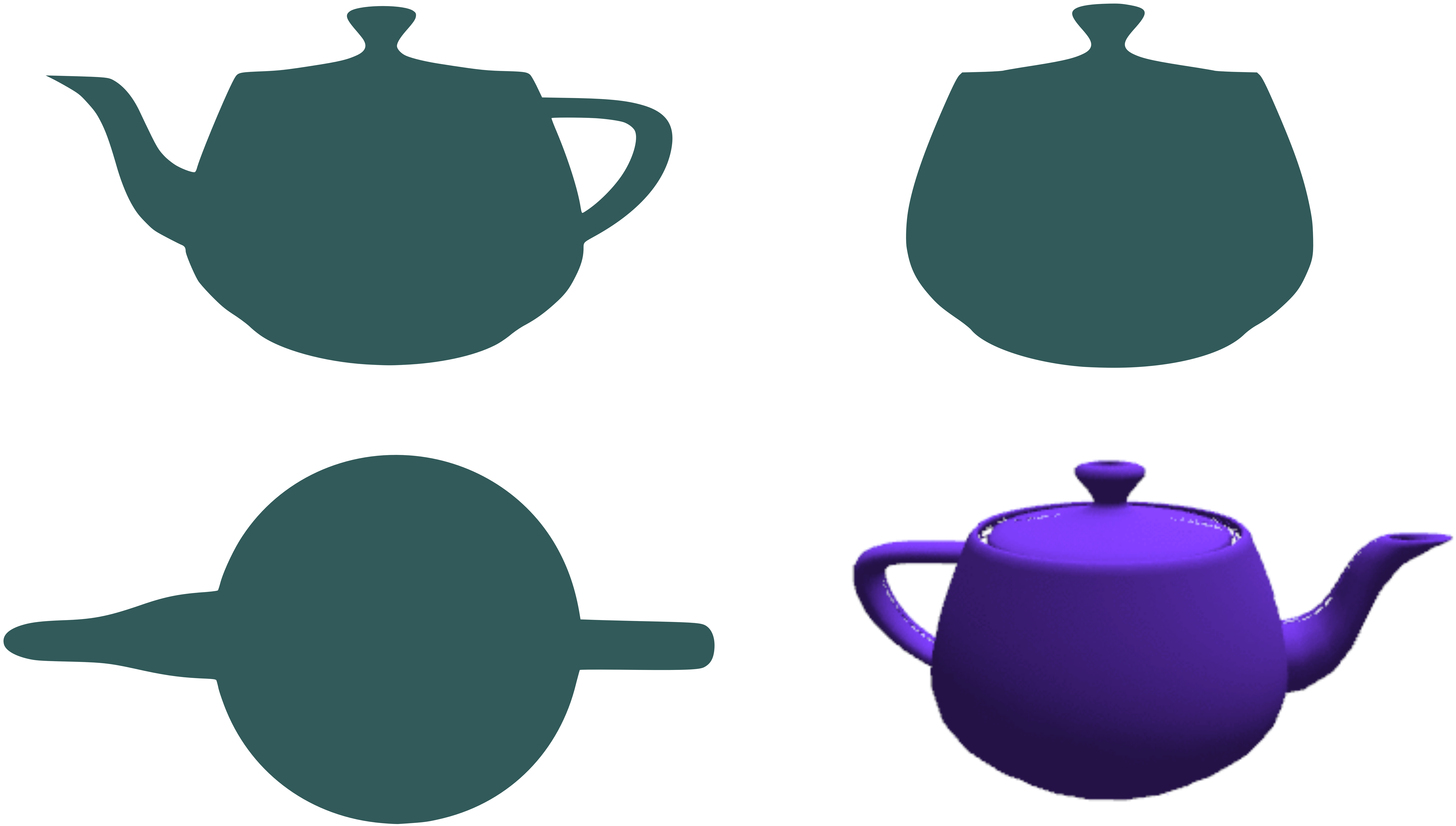
scvis	non-linear	Autoencoder	https://doi.org/10.1038/s41467-018-04368-5
VASC	non-linear	Autoencoder	https://doi.org/10.1016/j.gpb.2018.08.003

PCA

Principal Component Analysis



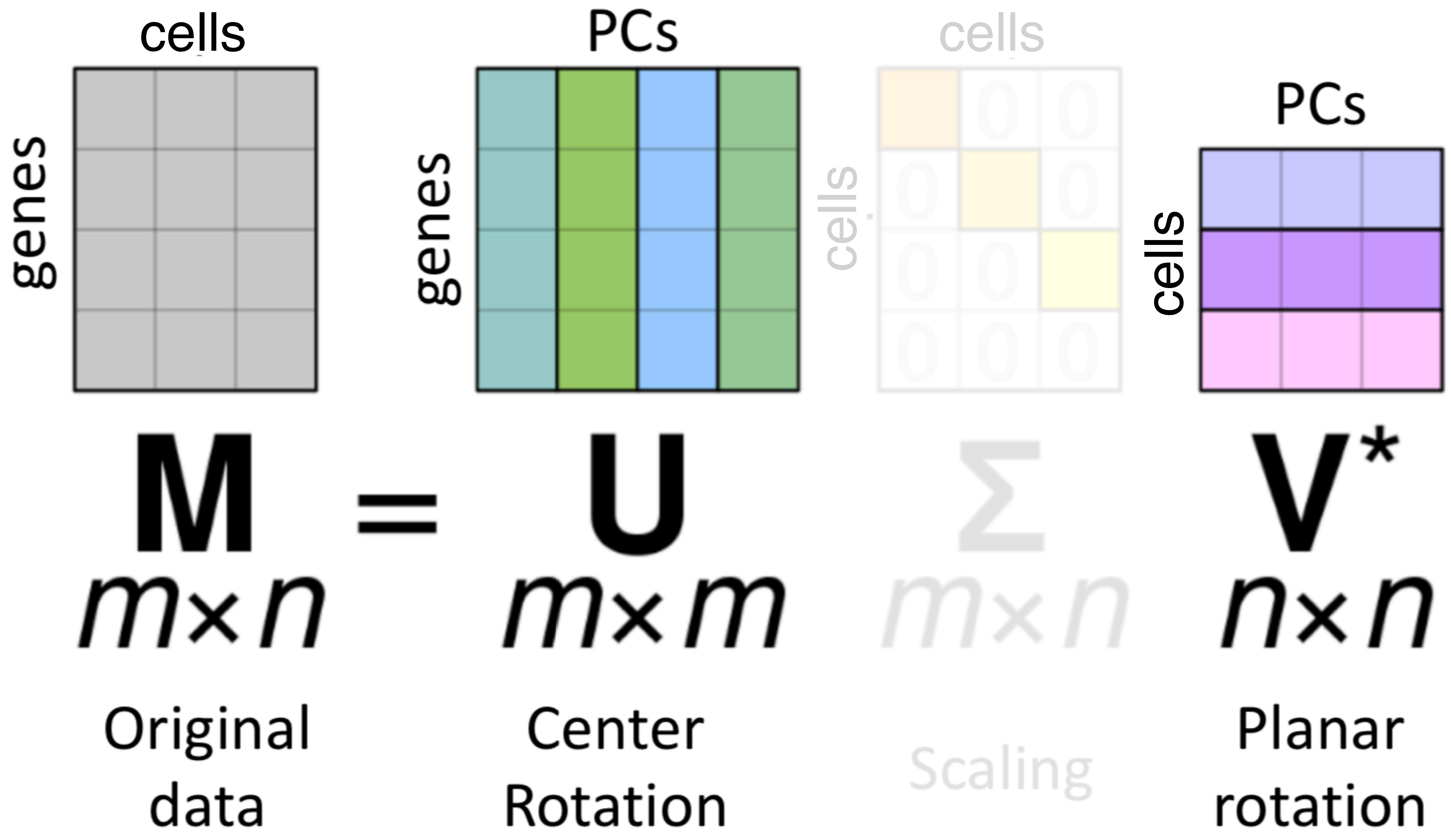




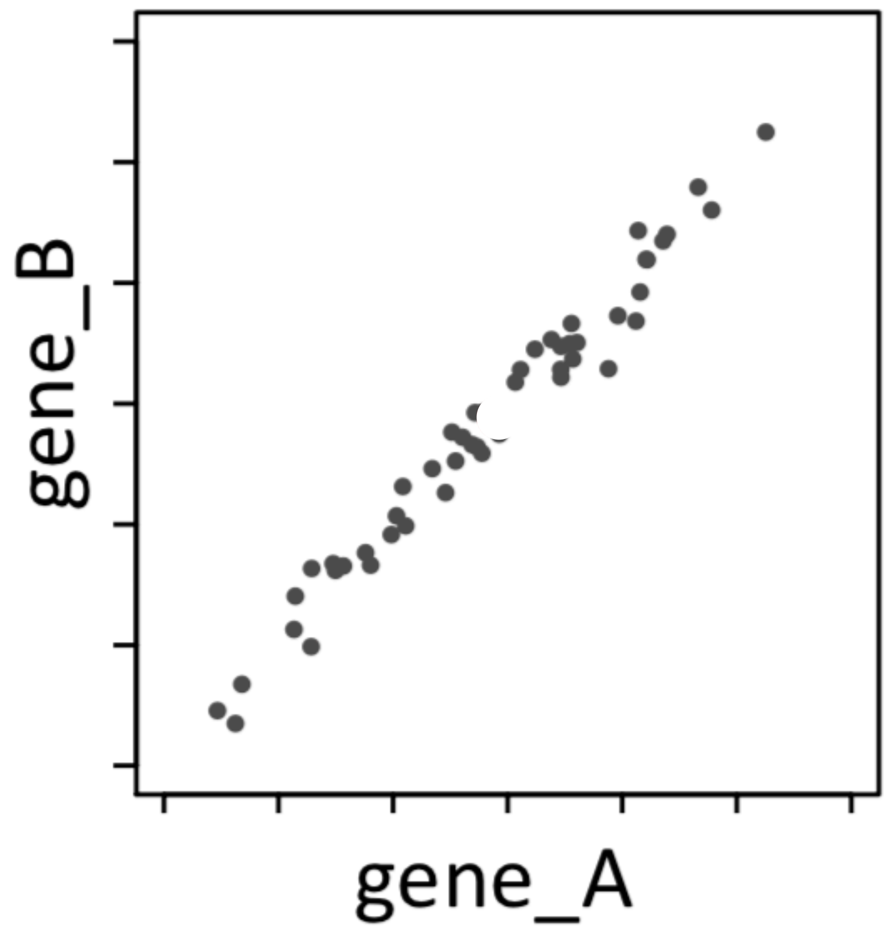
PCA - Intuition

- Given a dataset, compute/measure a number of features
- These features represent an N-dimensional problem
- PCA finds a new coordinate system obtained from the previous one by translation and rotation only
  change the point of view
- Moves the center of the coordinate system with center of the data
- Moves the x-axis into the principal axis of variation
- Orders axes by amount of variation (importance)

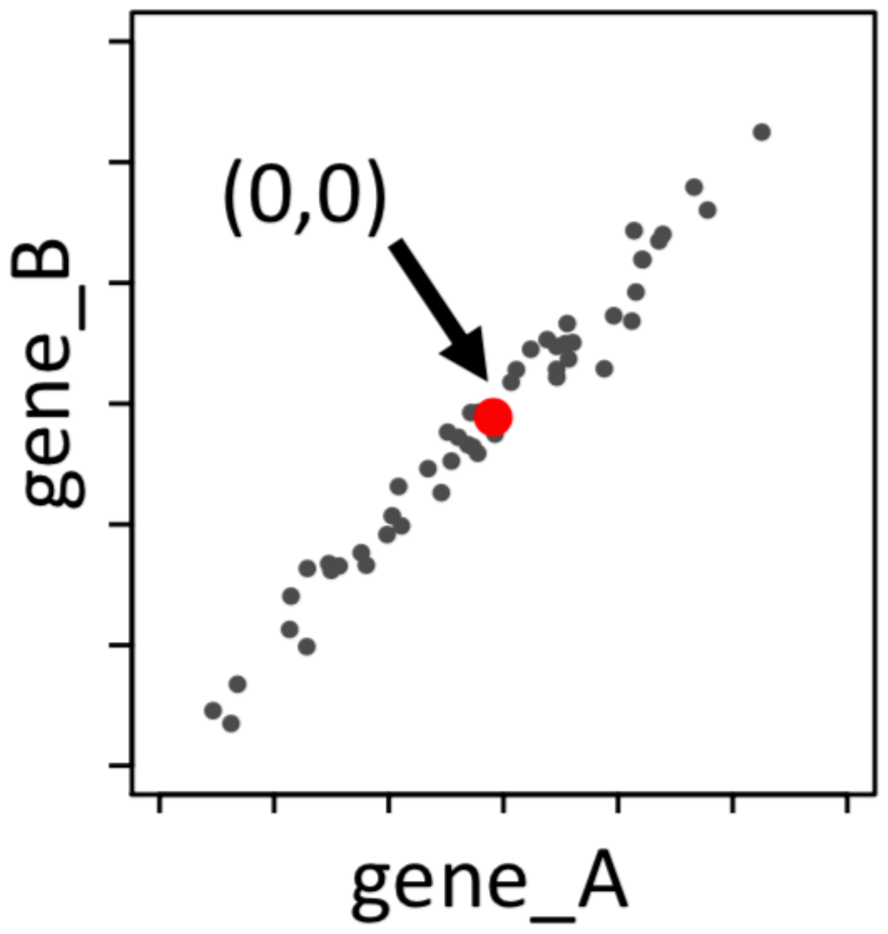
PCA in Brief


cells
genes
PCs
genes
cells
cells
0 0
0 0
0 0
0 0 0
PCs
cells
M = U Σ V*
m×n
m×m
m×n
n×n
Original data
Center Rotation
Scaling
Planar rotation

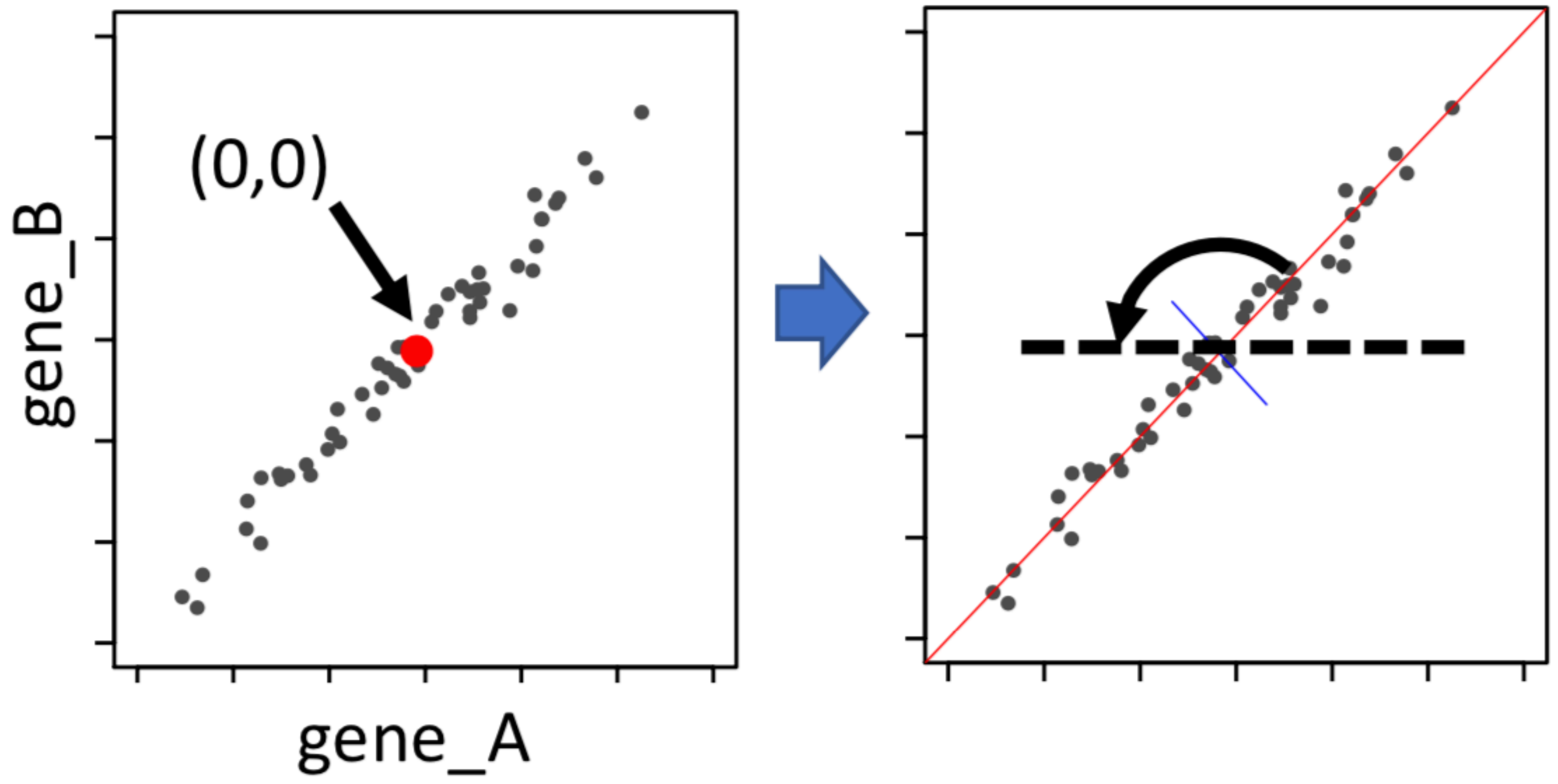
PCA in Brief



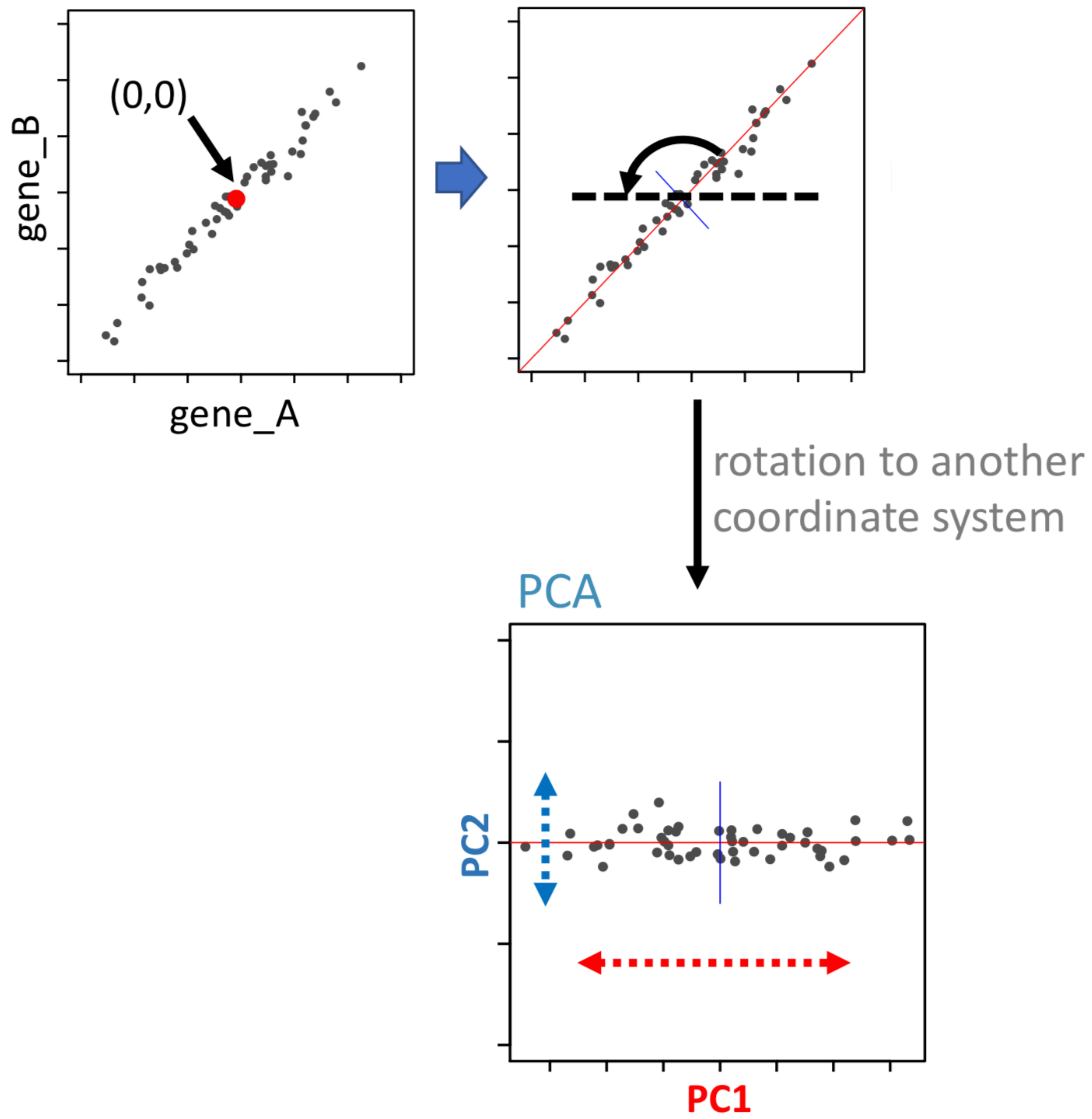
PCA in Brief



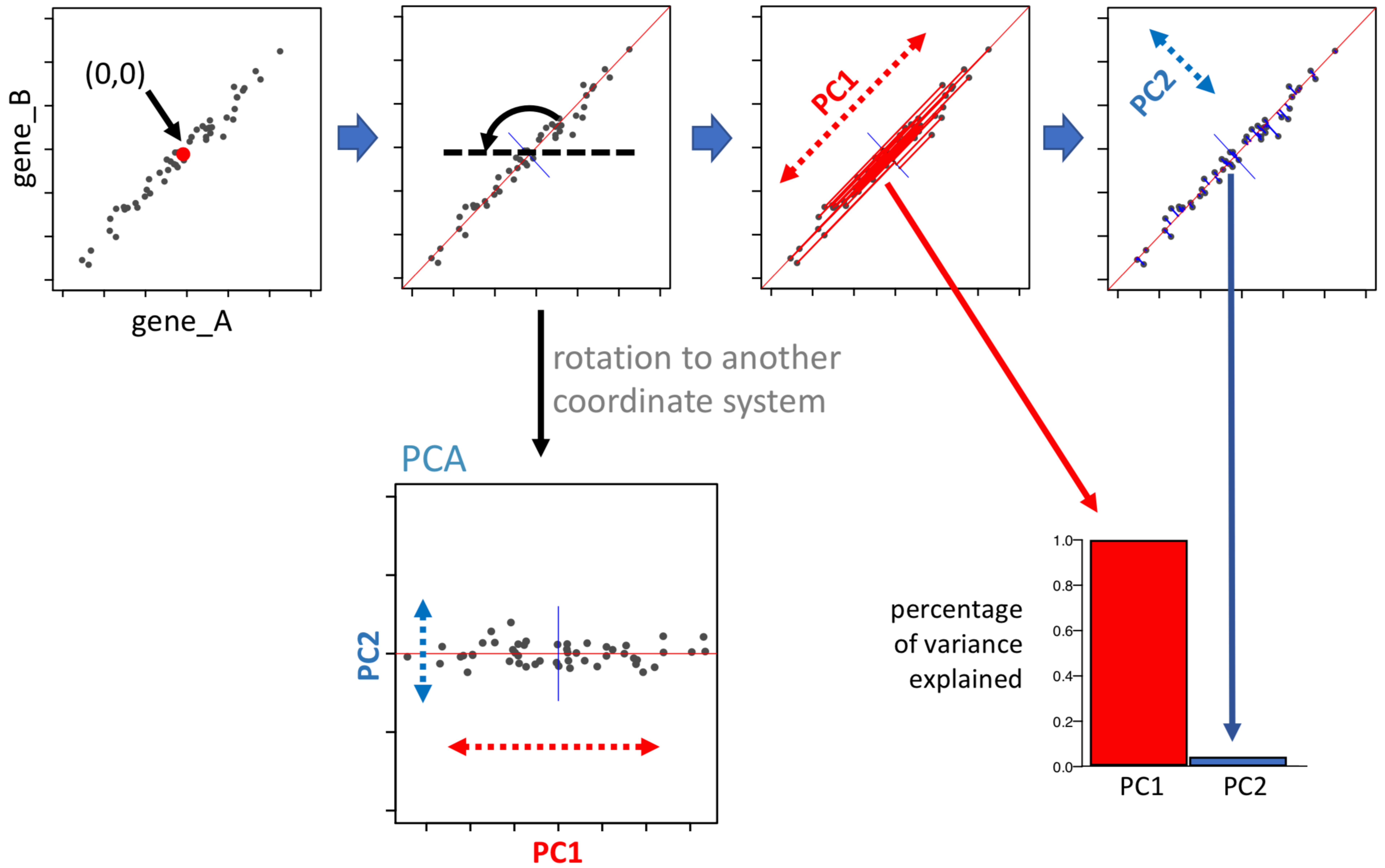
PCA in Brief



PCA in Brief


(0,0)
gene_B
gene_A
rotation to another
coordinate system
PCA
PC2
PC1

PCA in Brief



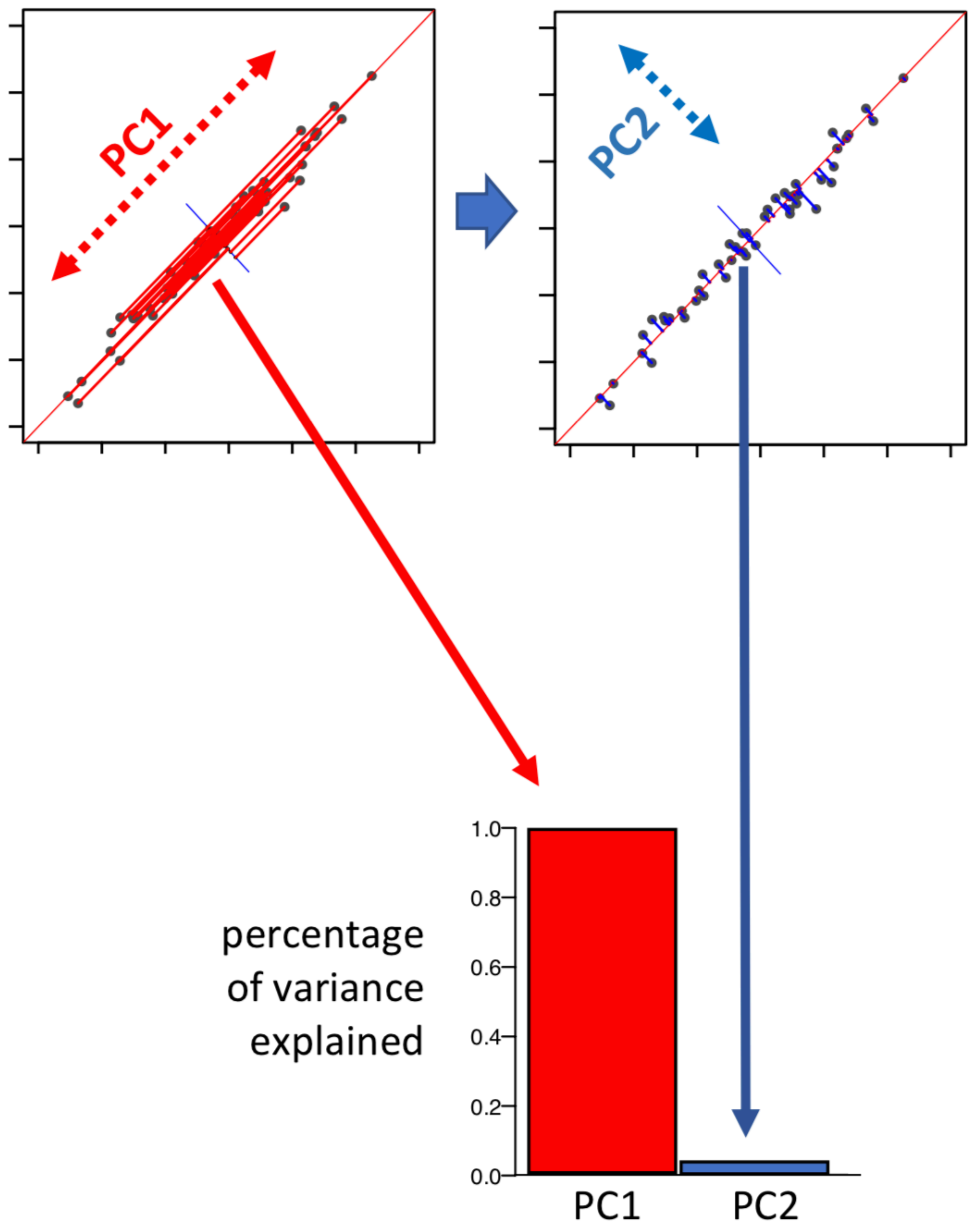
(0,0)
gene_B
gene_A
PC1
PC2
rotation to another coordinate system
PCA
PC2
PC1
percentage of variance explained
1.0
0.8
0.6
0.4
0.2
0.0
PC1
PC2

PCA in Brief

PC1 explains >98% of the variance

1 PC thus represents 2 genes very well
"Removing" redundancy

PC2 is nearly insignificant in this example
Could be disregarded



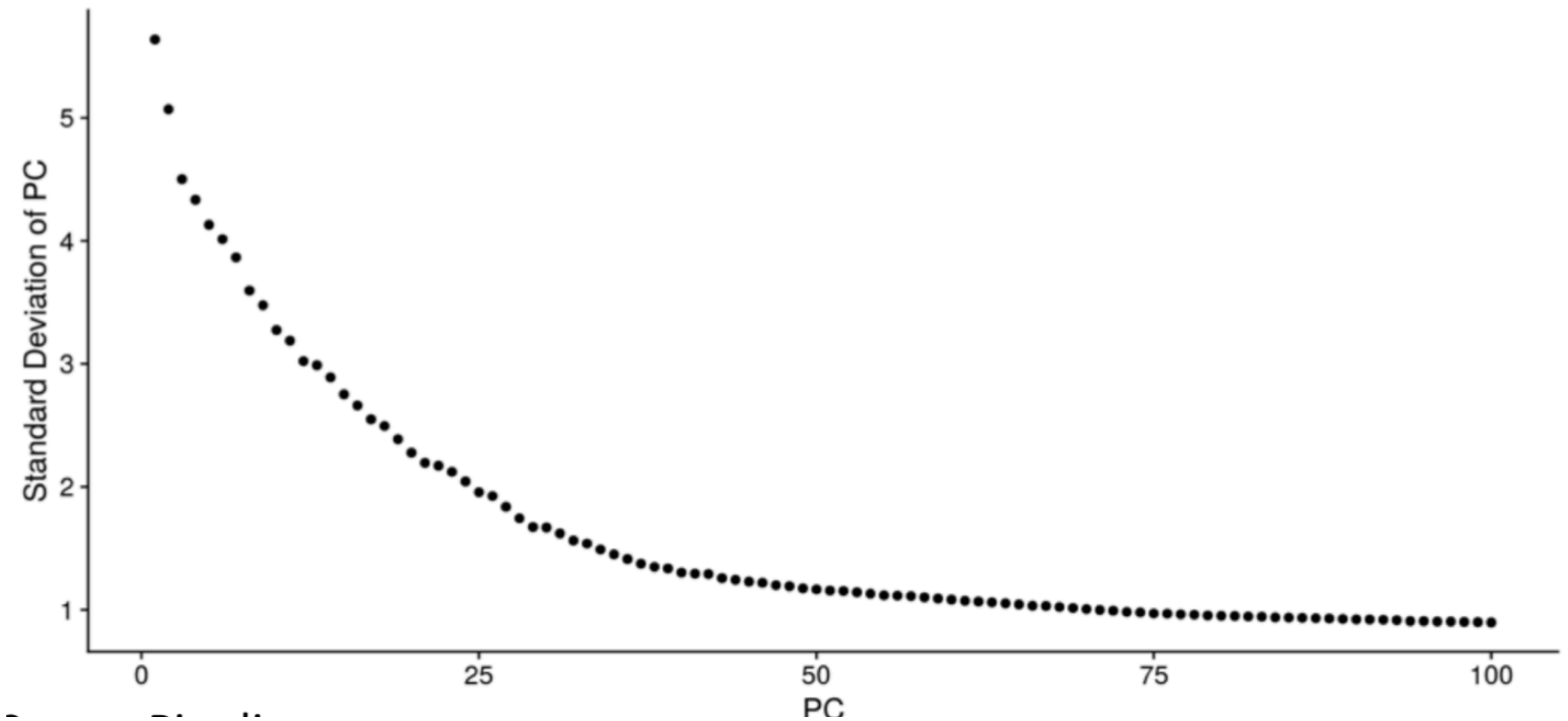
PCA in Brief

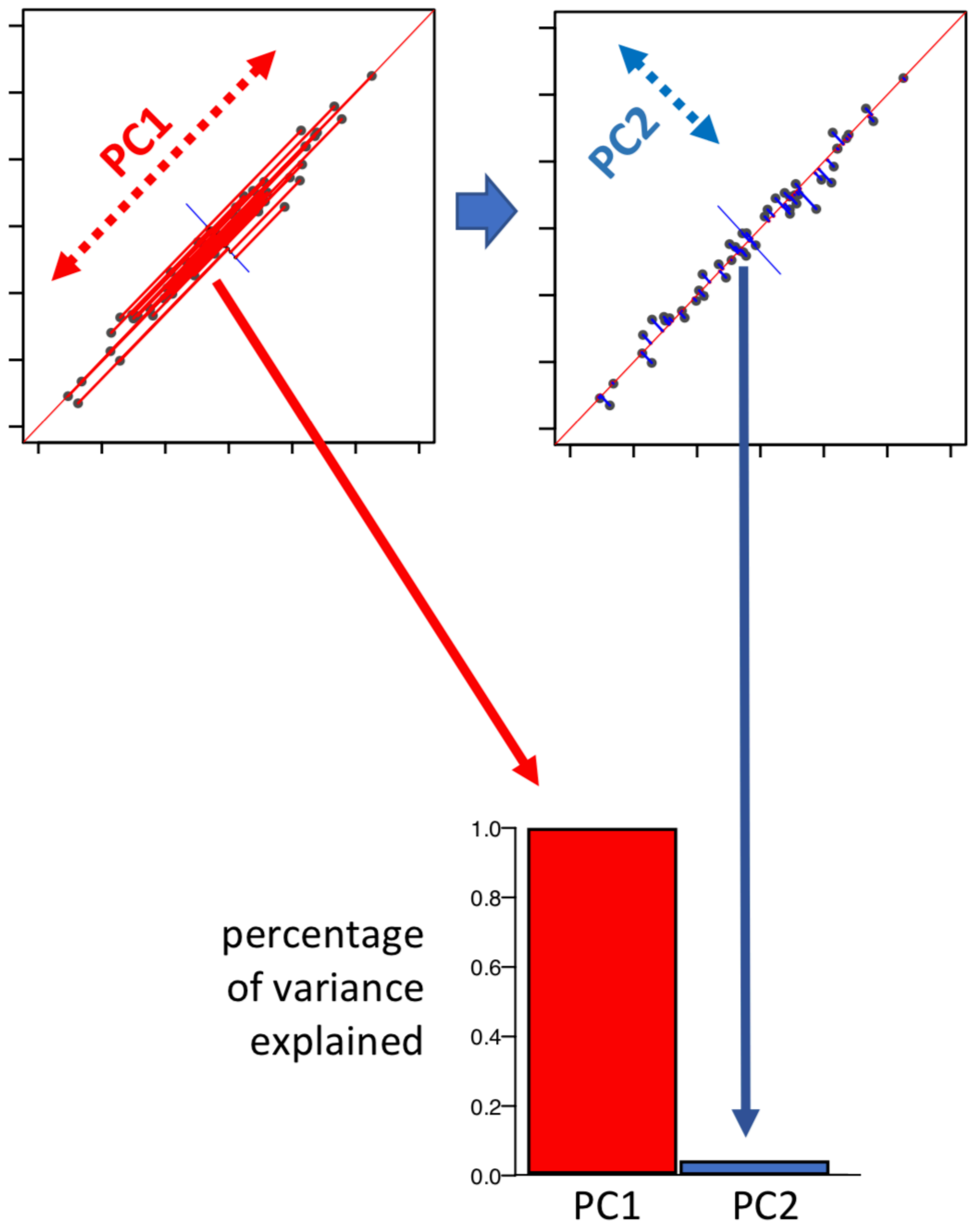
PC1 explains >98% of the variance

1 PC thus represents 2 genes very well
"Removing" redundancy

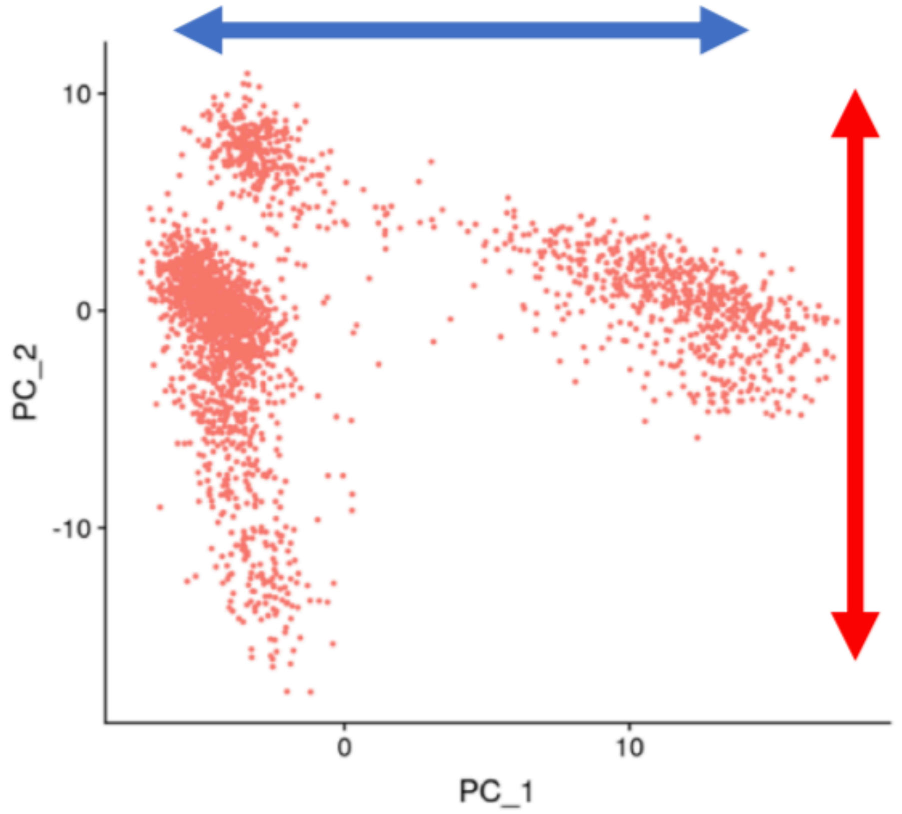
PC2 is nearly insignificant in this example
Could be disregarded

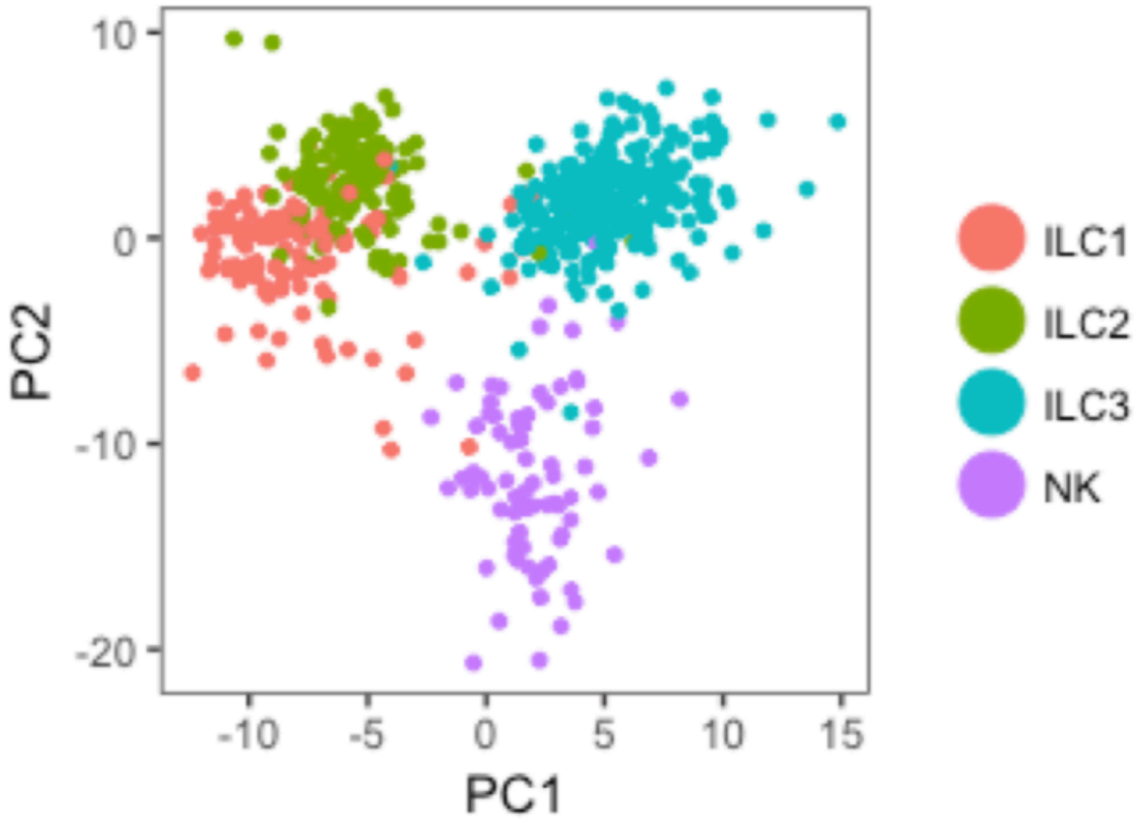
In real life …





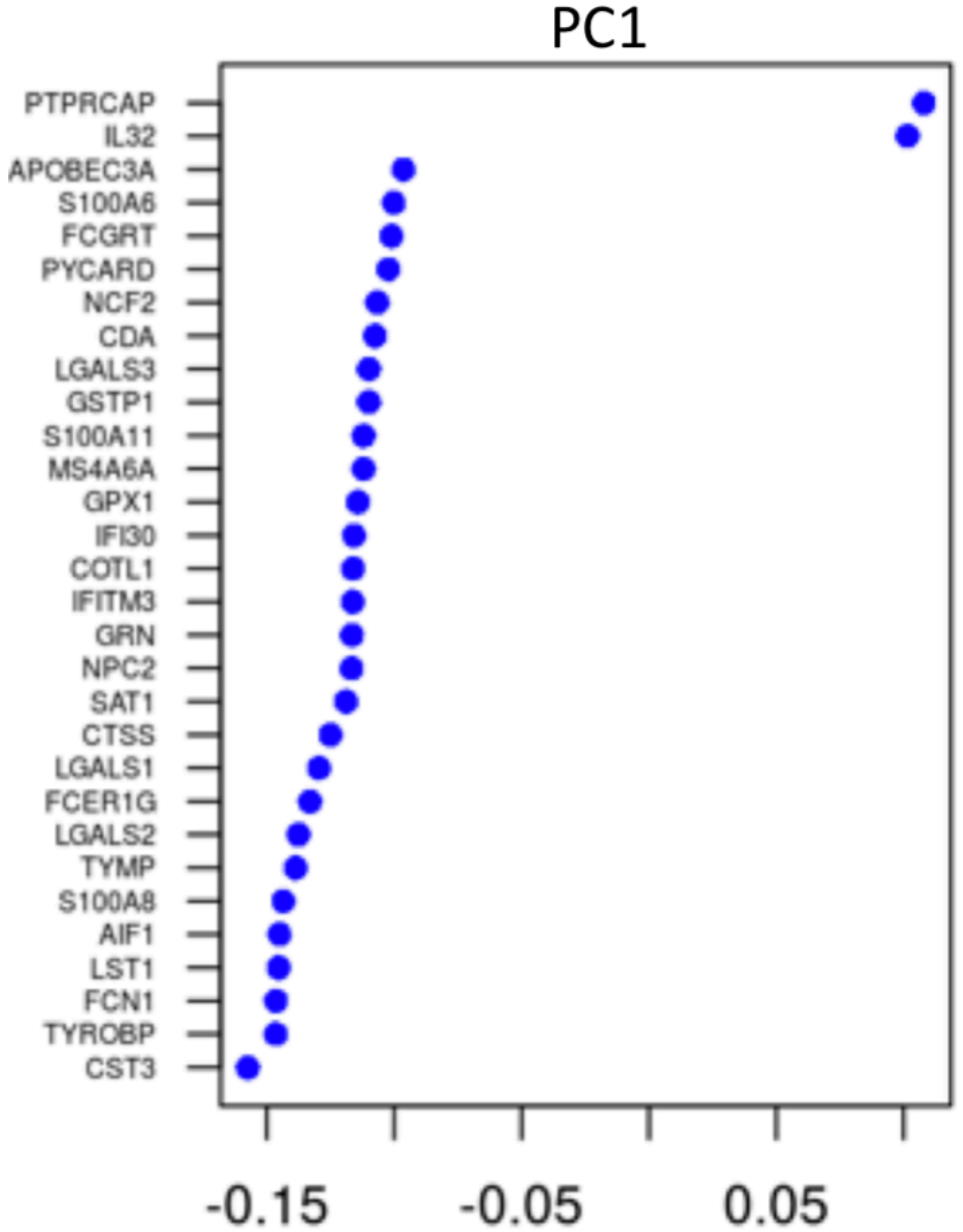
PCA in Brief

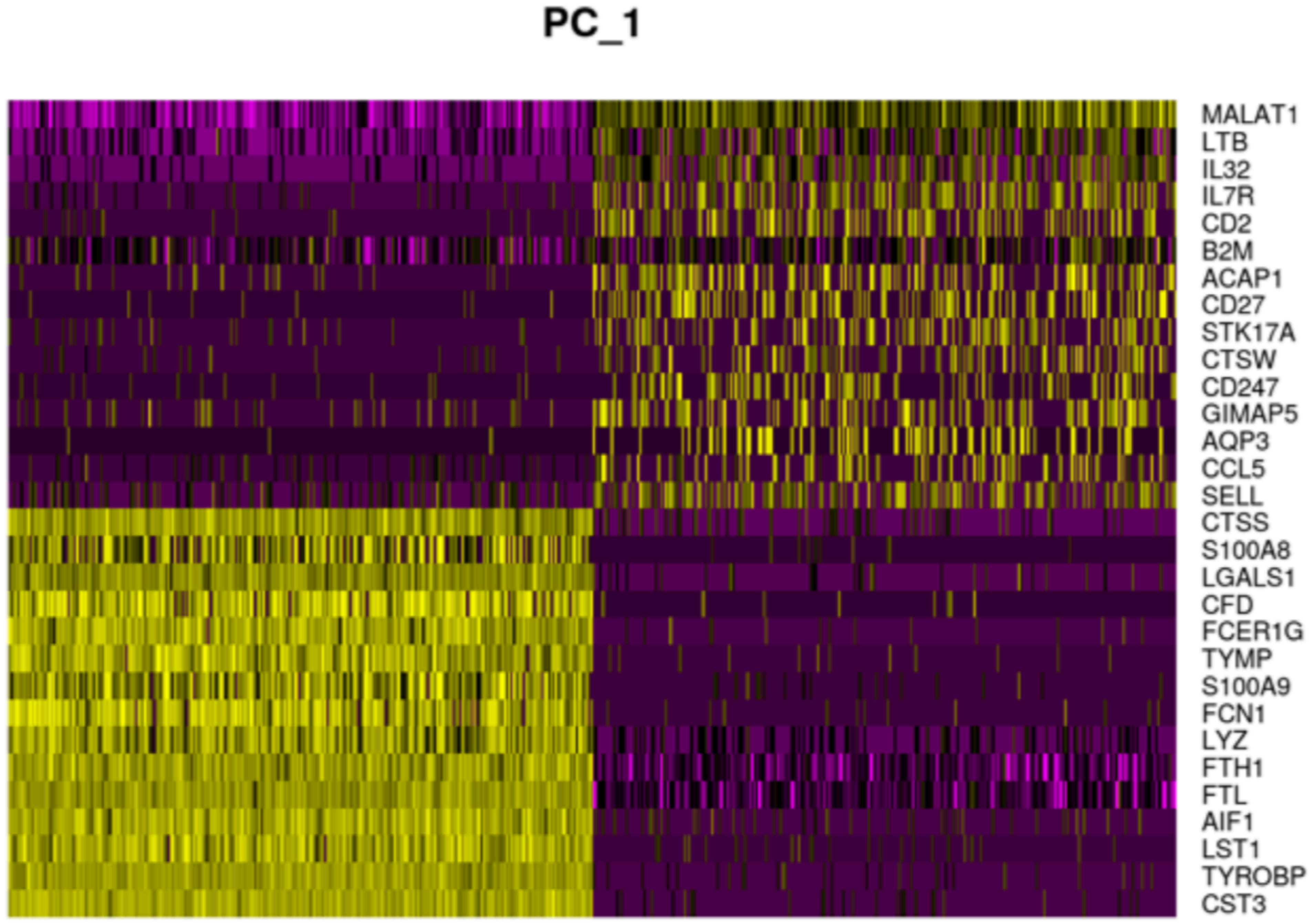




PC1 and PC2 are commonly correlated to sequencing depth and cell heterogeneity/complexity

(but not always …)





Seurat Pipeline / Forkel et all 2015

Summary: PCA

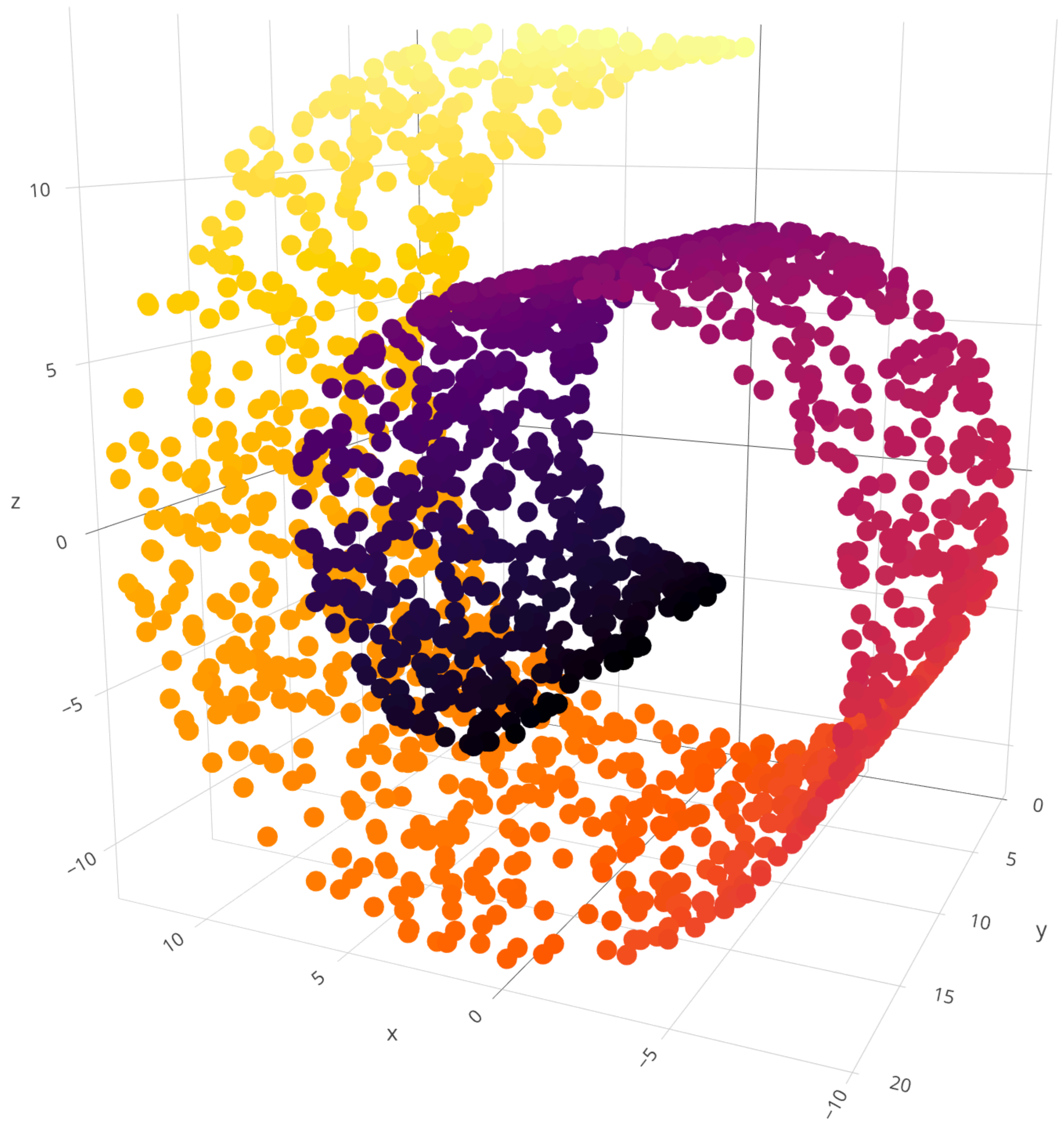
- LINEAR method of dimensionality reduction
- The TOP principal components contain higher variance from the data
- Can be used as FILTERING, by selecting only the top significant PCs
- *It is an interpretable/parametric dimensionality reduction*

- **Problems:**
- It performs poorly to separate cells in 0-inflated data types
- Cell sizes and sequencing depth are usually captured in the top PCs

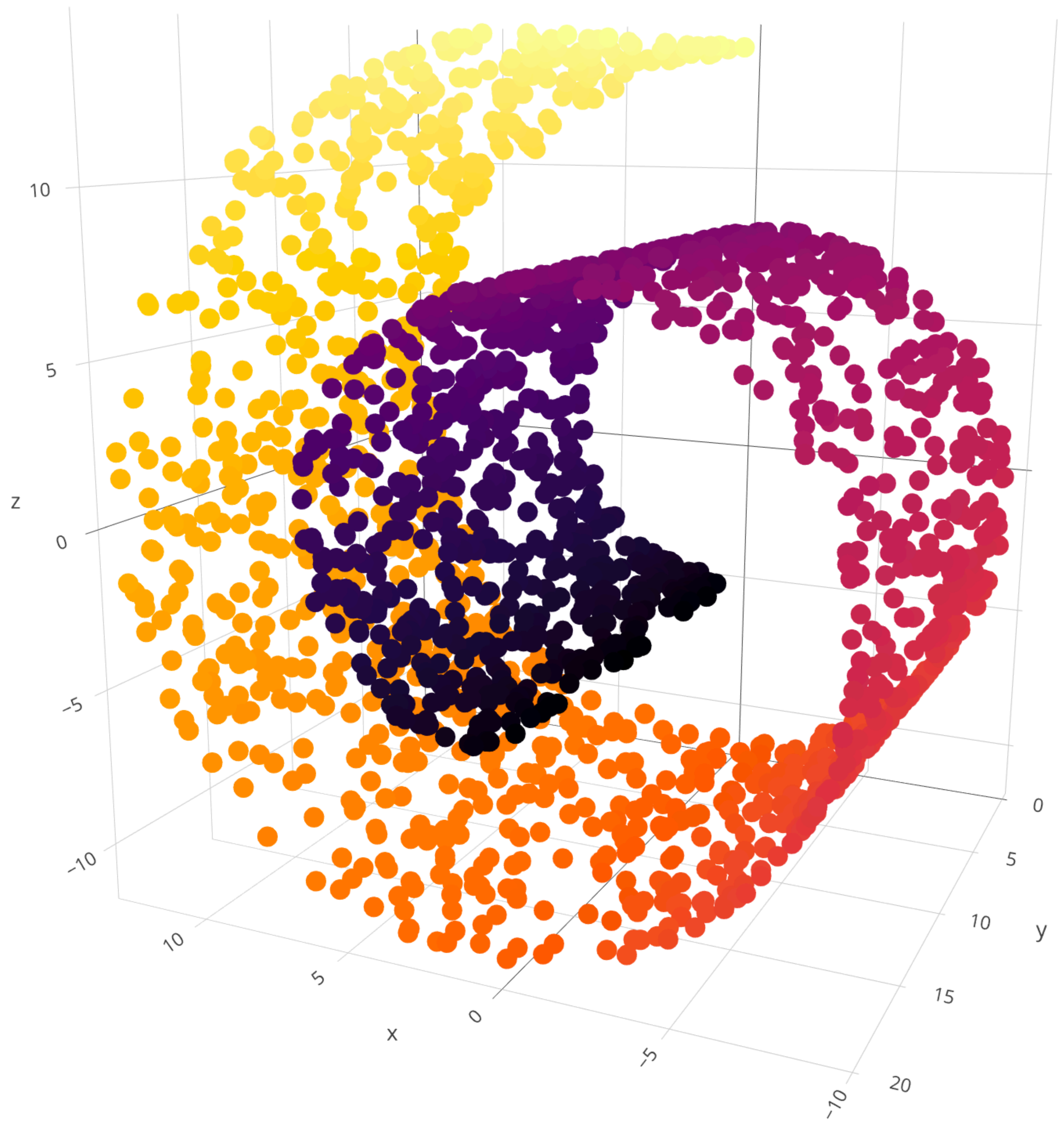
t-SNE

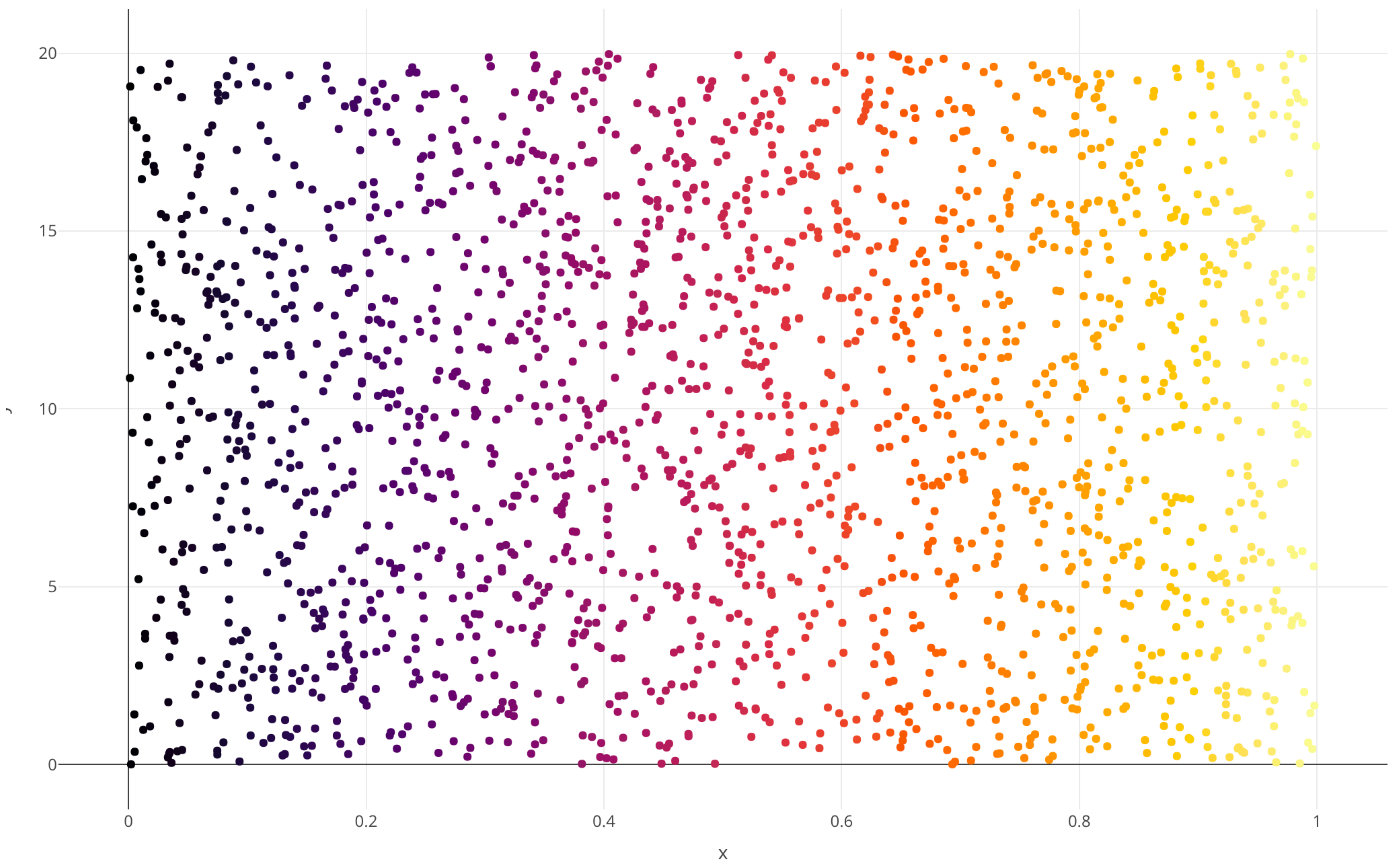
t-distributed Stochastic Neighborhood Embedding

Manifold Learning

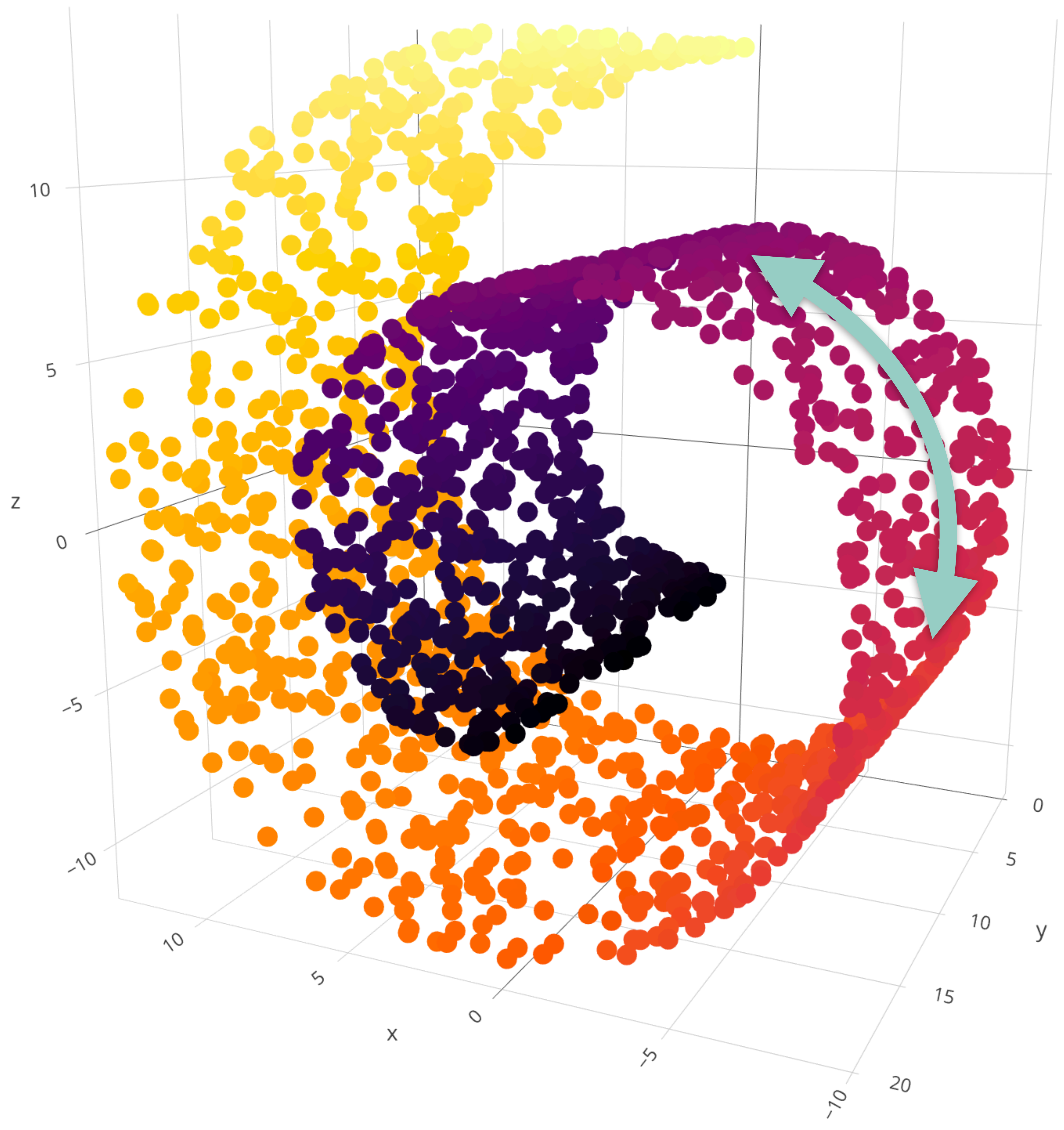


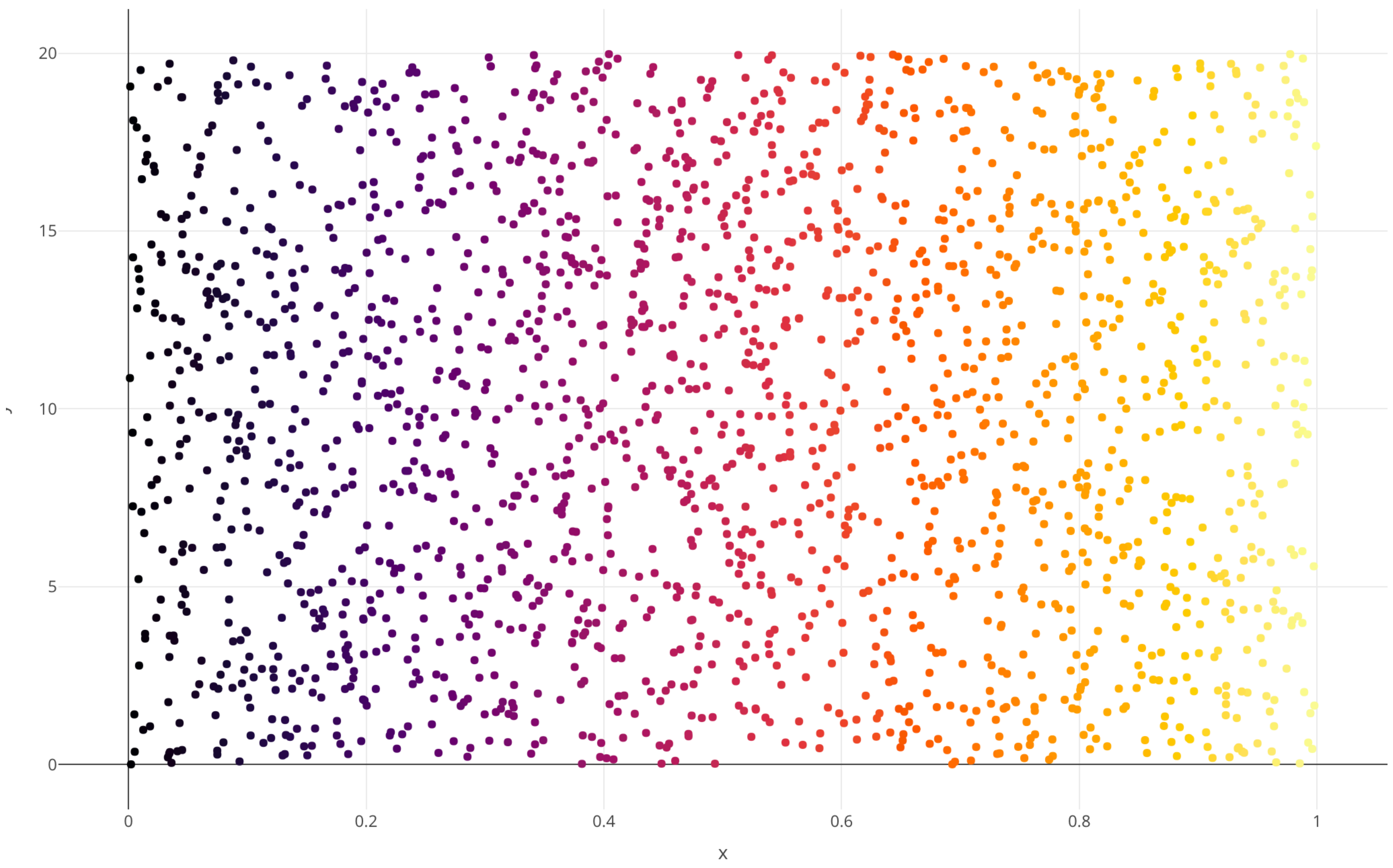
Manifold Learning



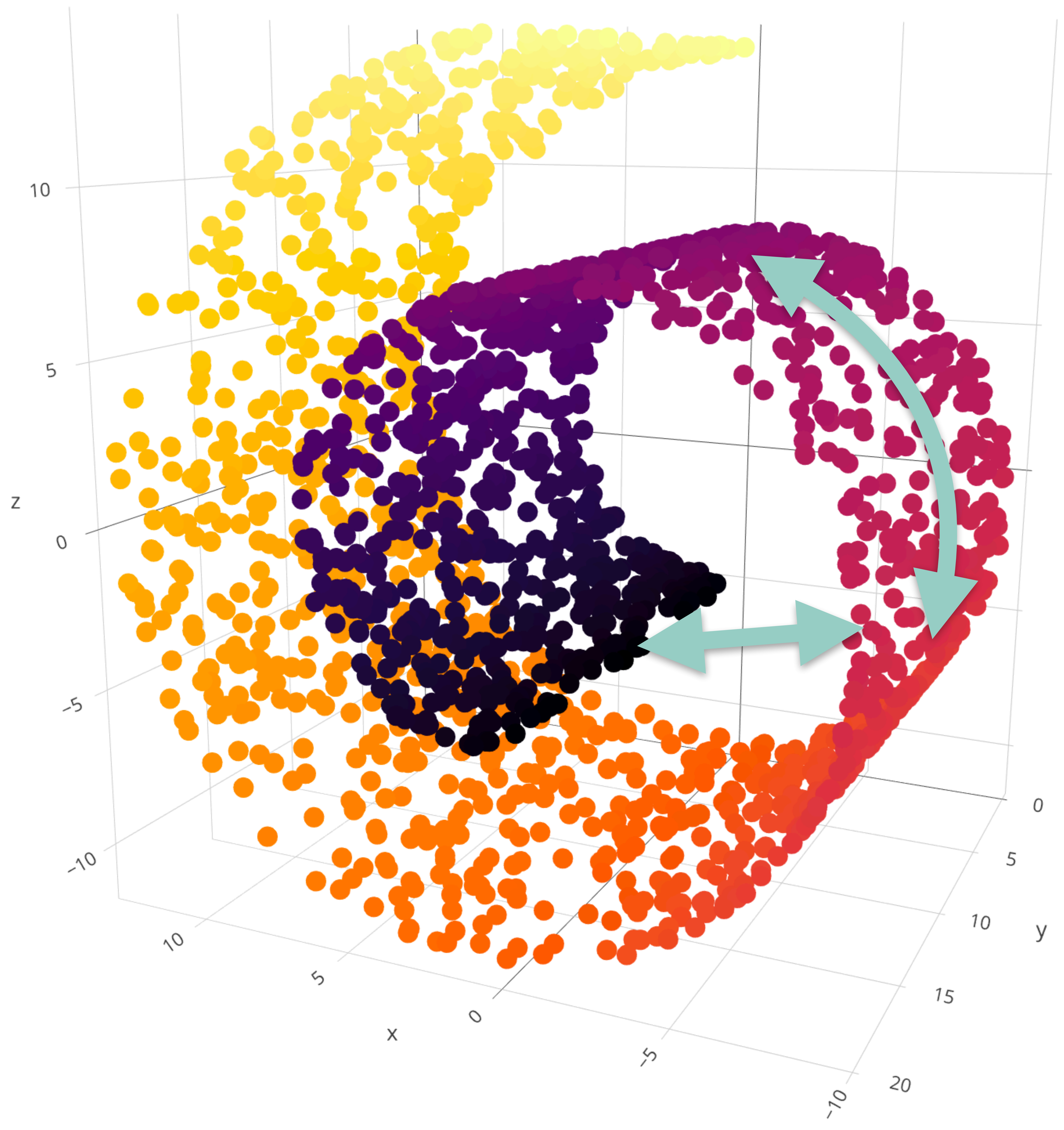


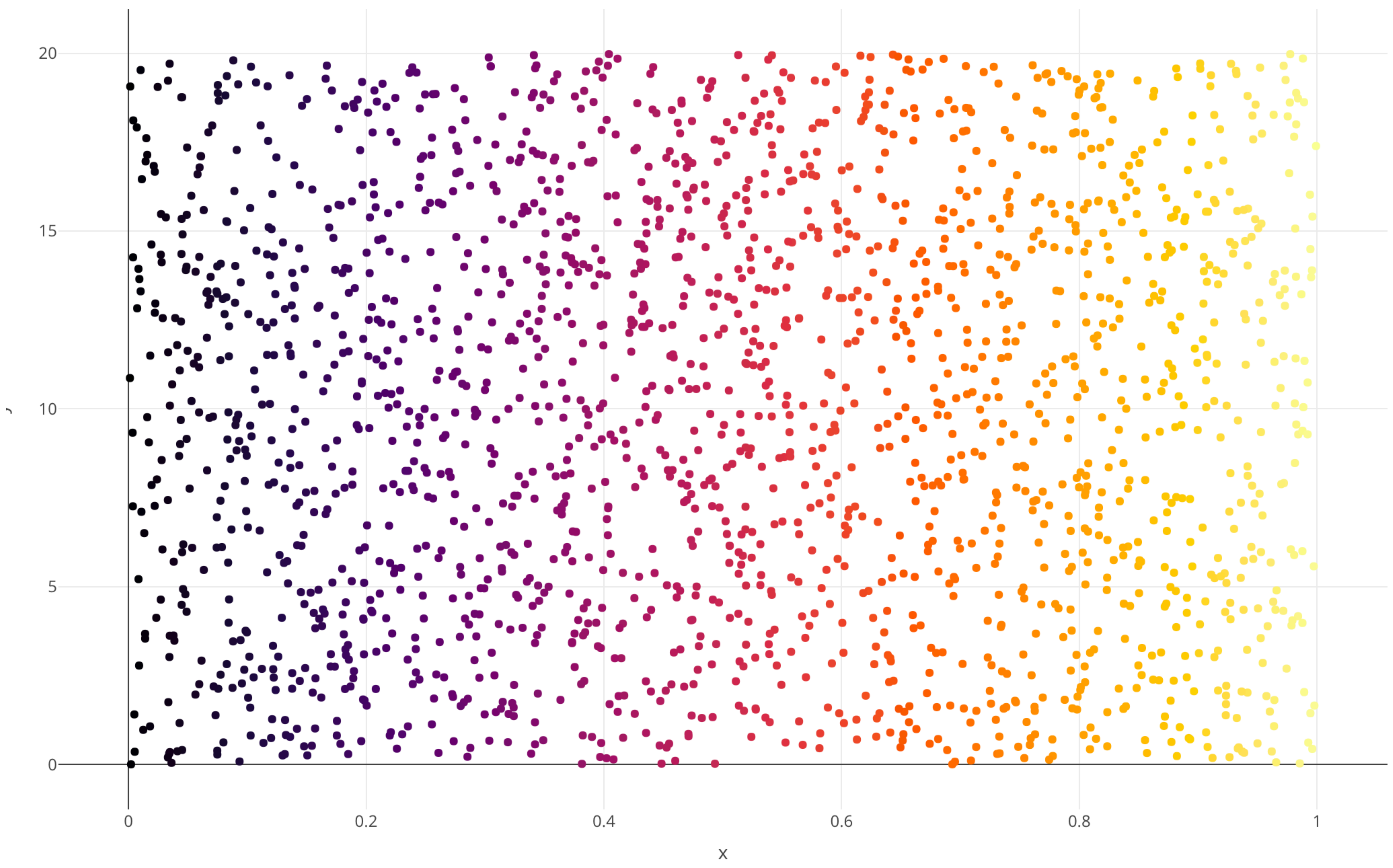
Manifold Learning



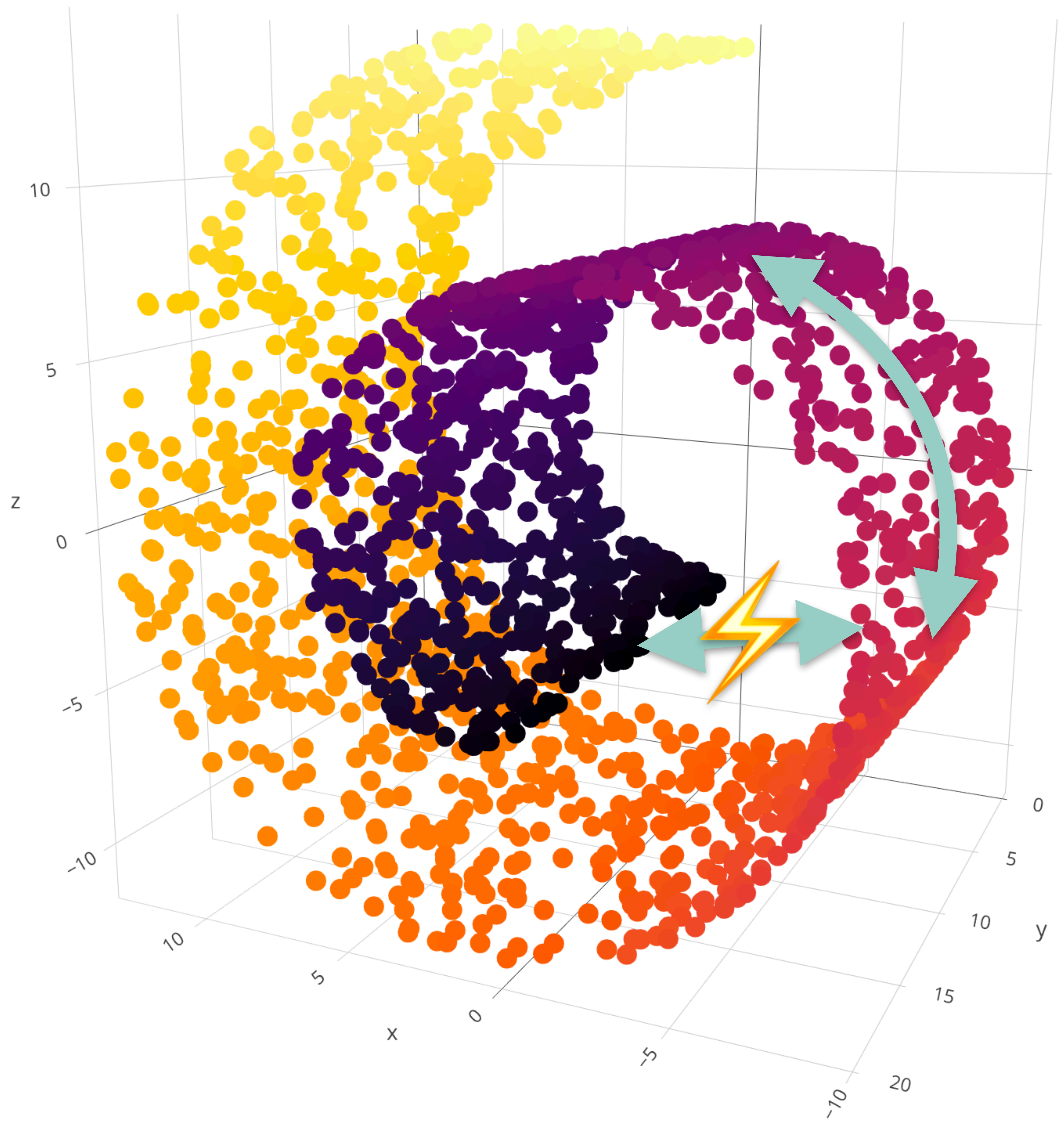


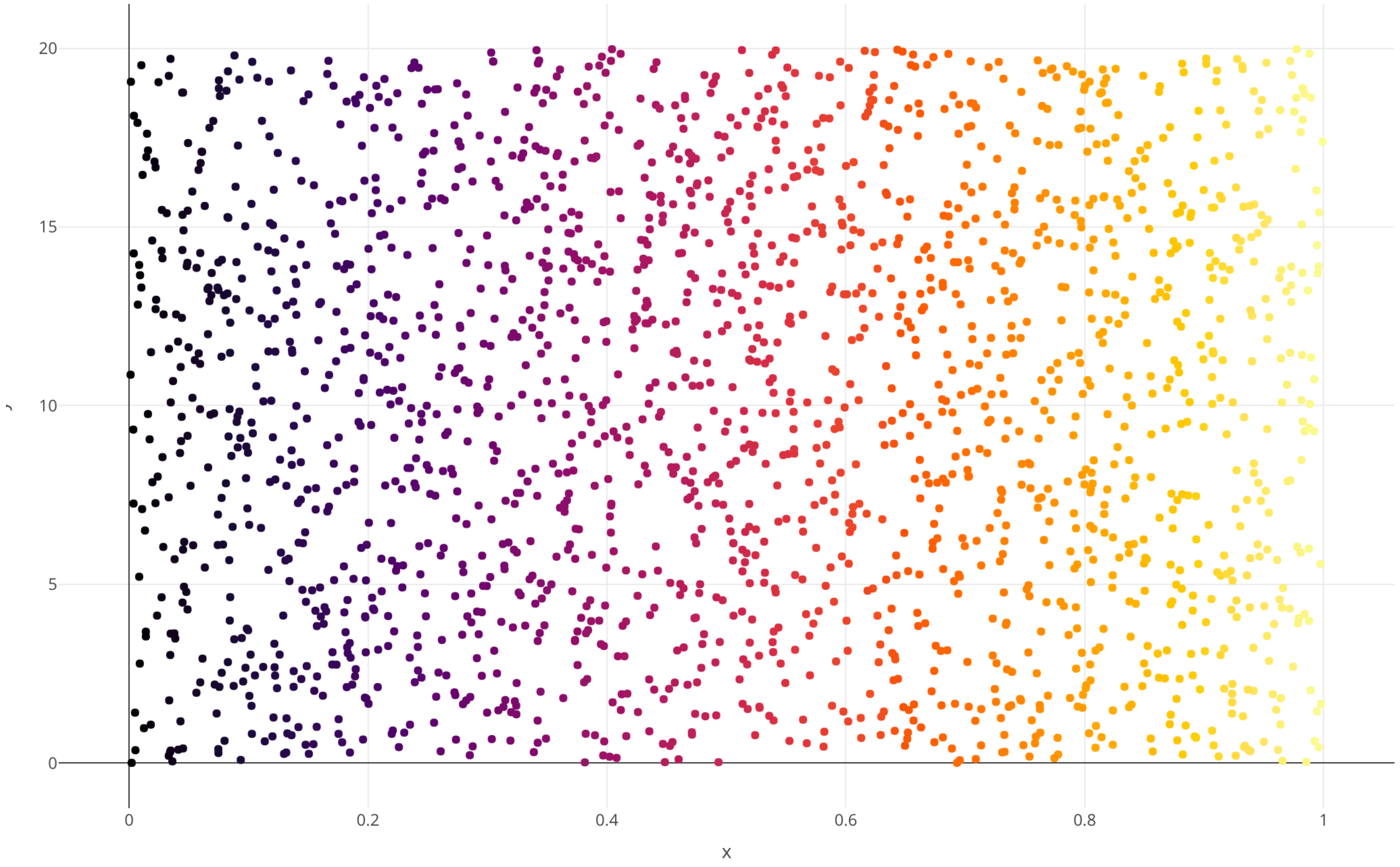
Manifold Learning



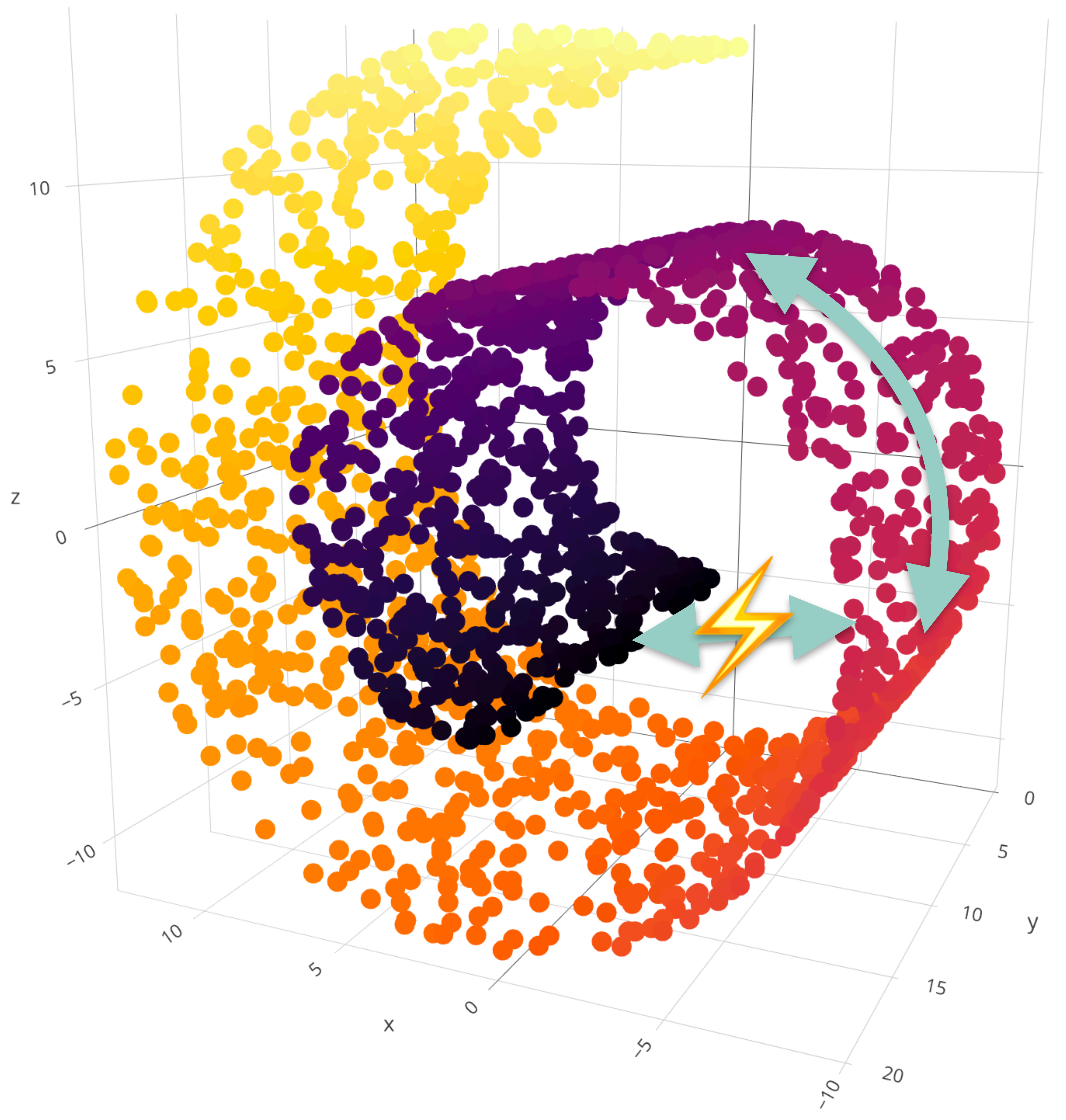


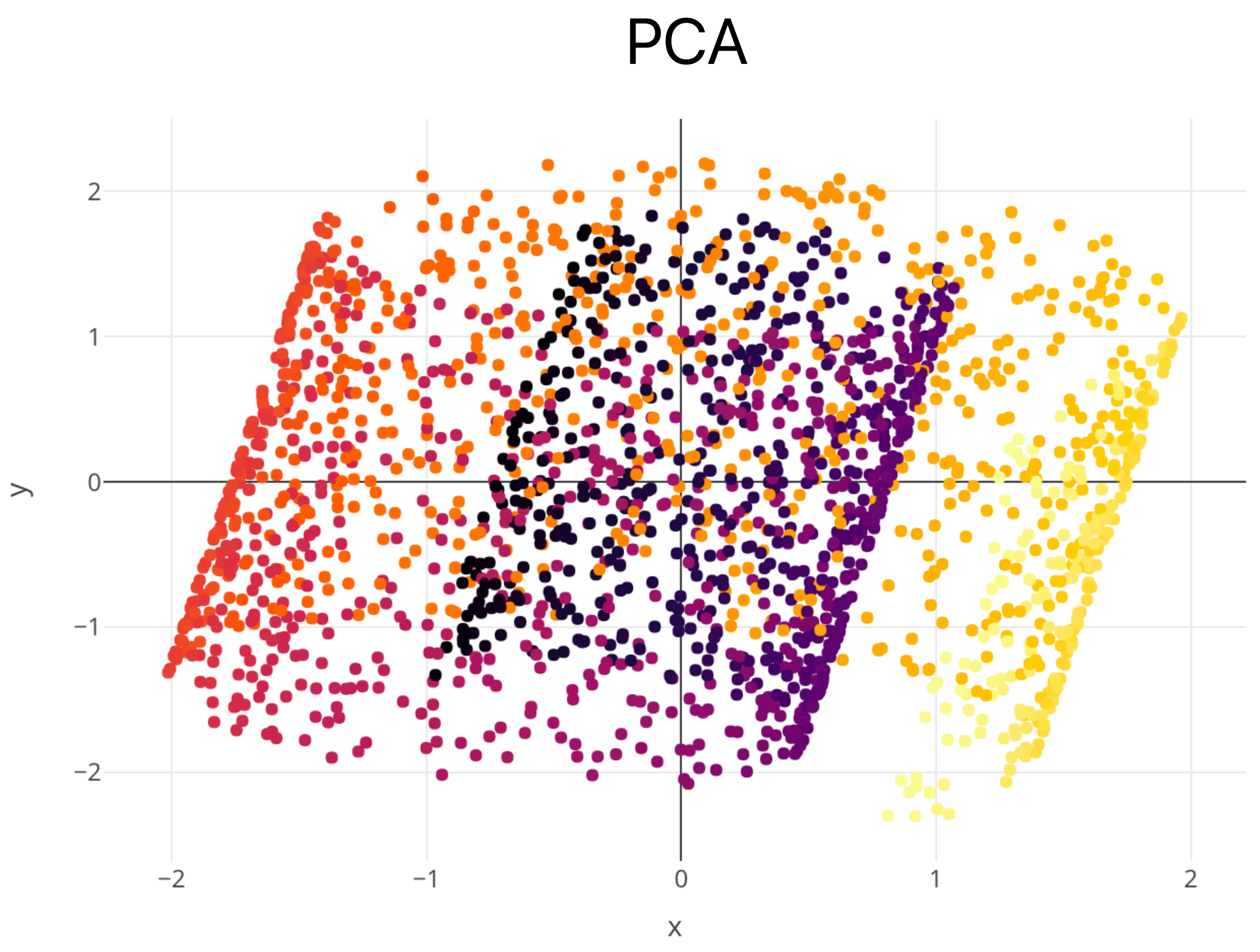
Manifold Learning





Manifold Learning





t-SNE Intuition

- Non-linear dimensionality reduction
  - Compute neighborhoods in hi-D
  - Model low-D to preserve neighborhoods
- Preserves local neighborhoods

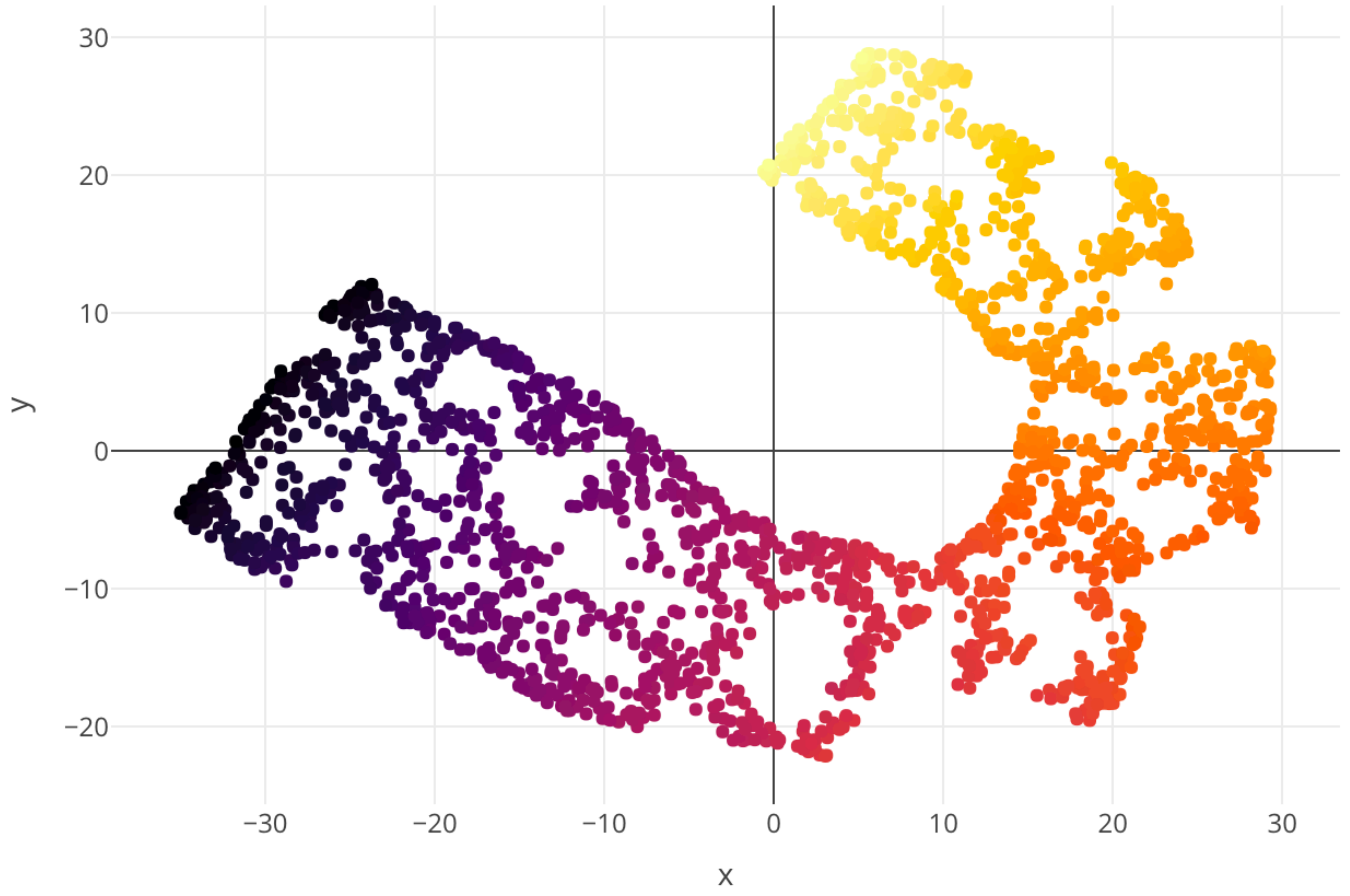
➫ Preserves high-D clusters!

t-SNE Intuition

- Non-linear dimensionality reduction
  - Compute neighborhoods in hi-D
  - Model low-D to preserve neighborhoods
- Preserves local neighborhoods

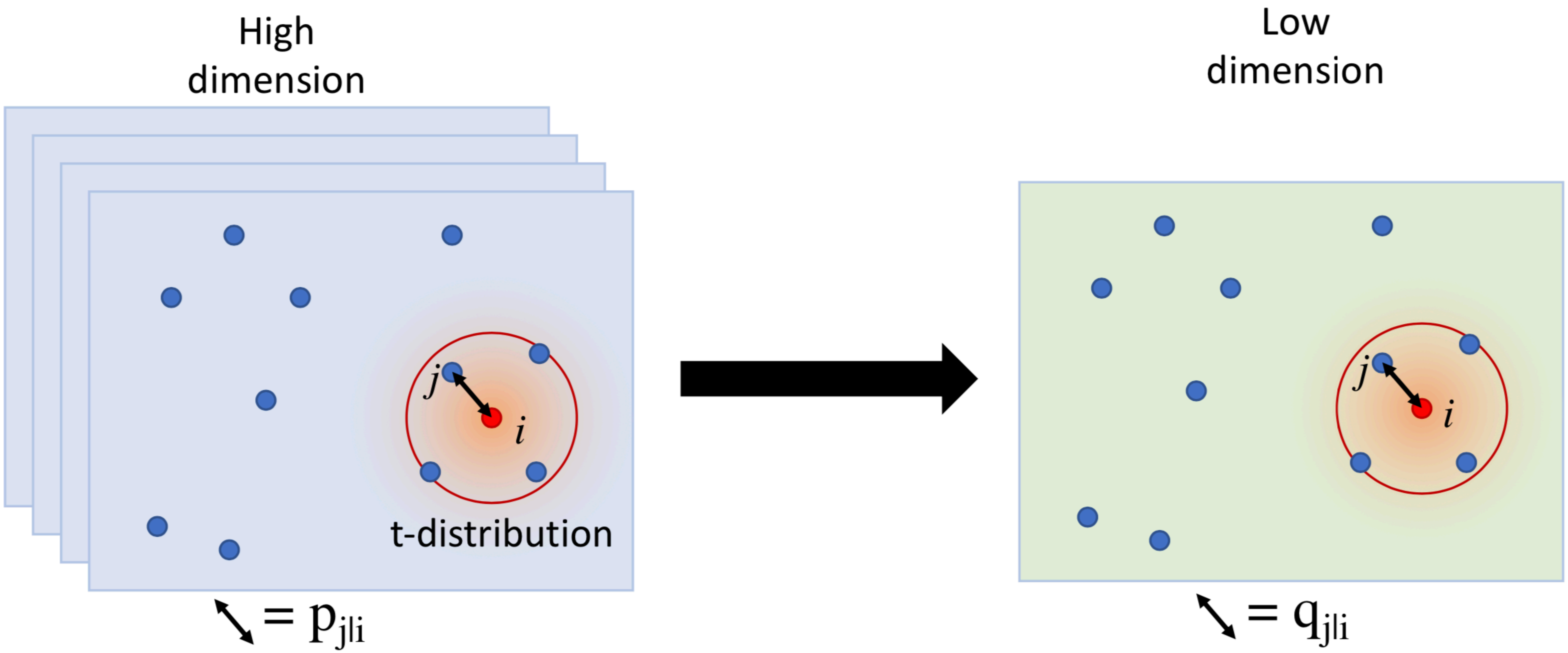
➯ Preserves high-D clusters!

t-SNE

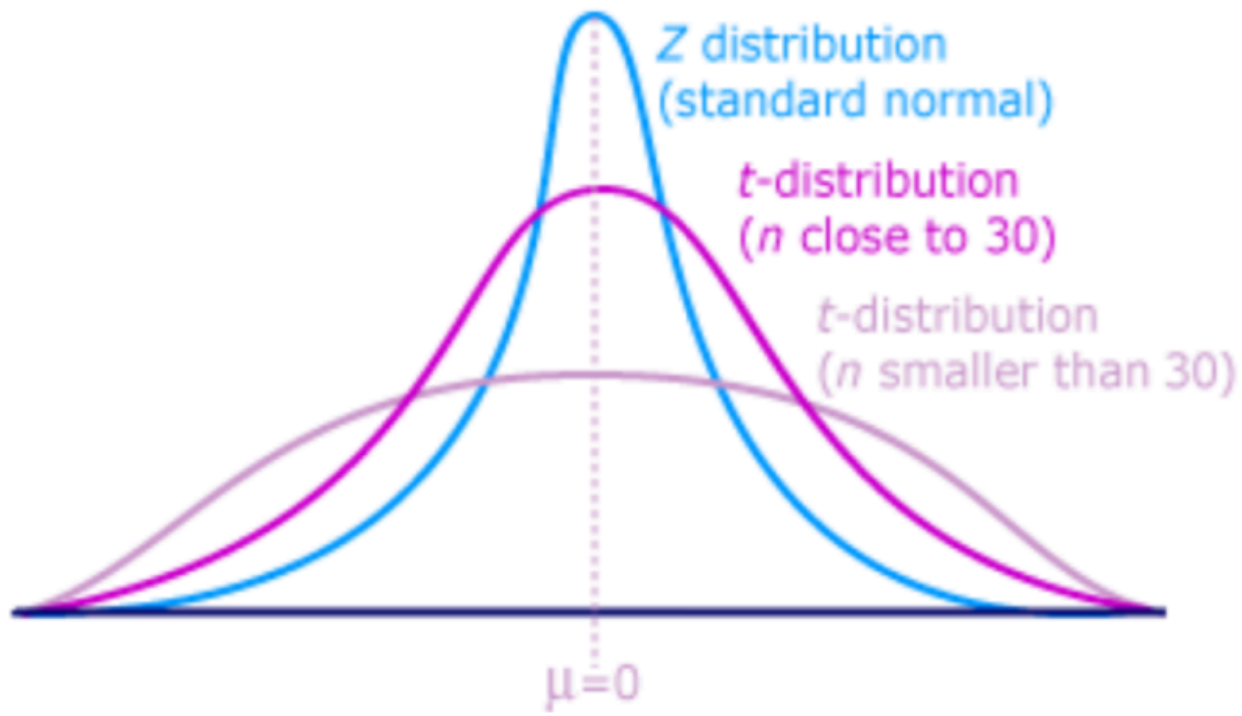


*actually also not great with Swiss Roll

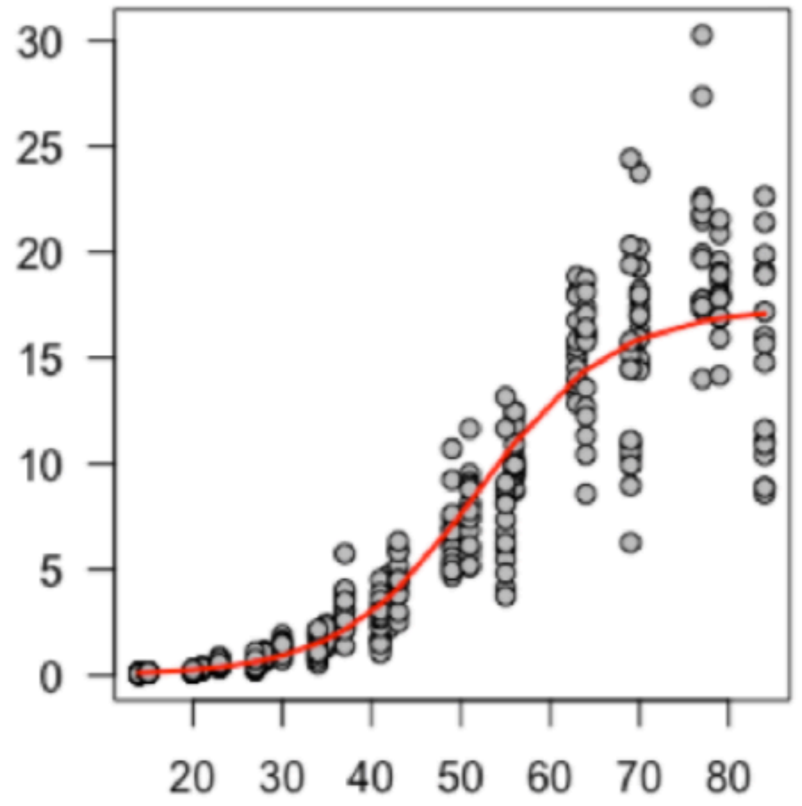
t-SNE in Brief



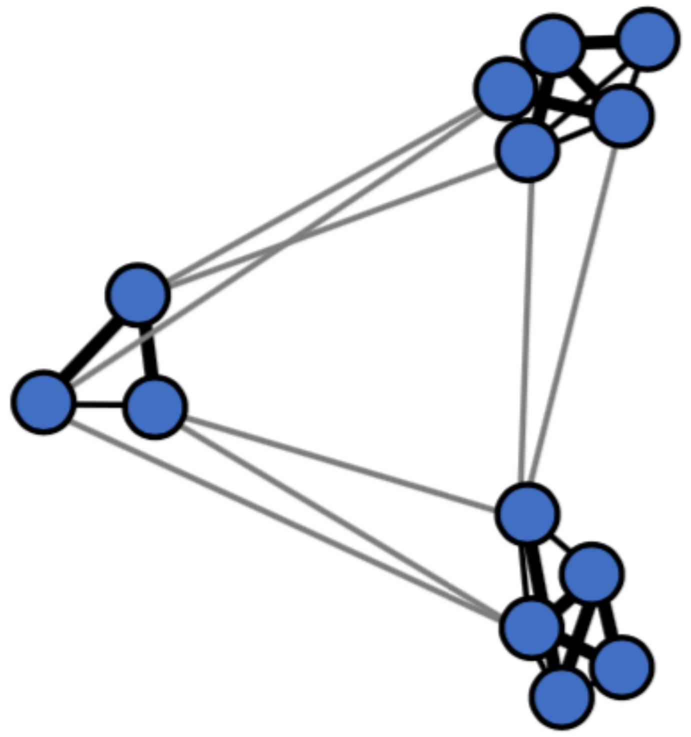
$p_{j|i}$ and $q_{j|i}$ measure the <u>conditional probability</u> that a point i would pick point j as it's nearest neighbor, in high (p) and low (q) dimensional space respectively.



Sidestep: Graphs

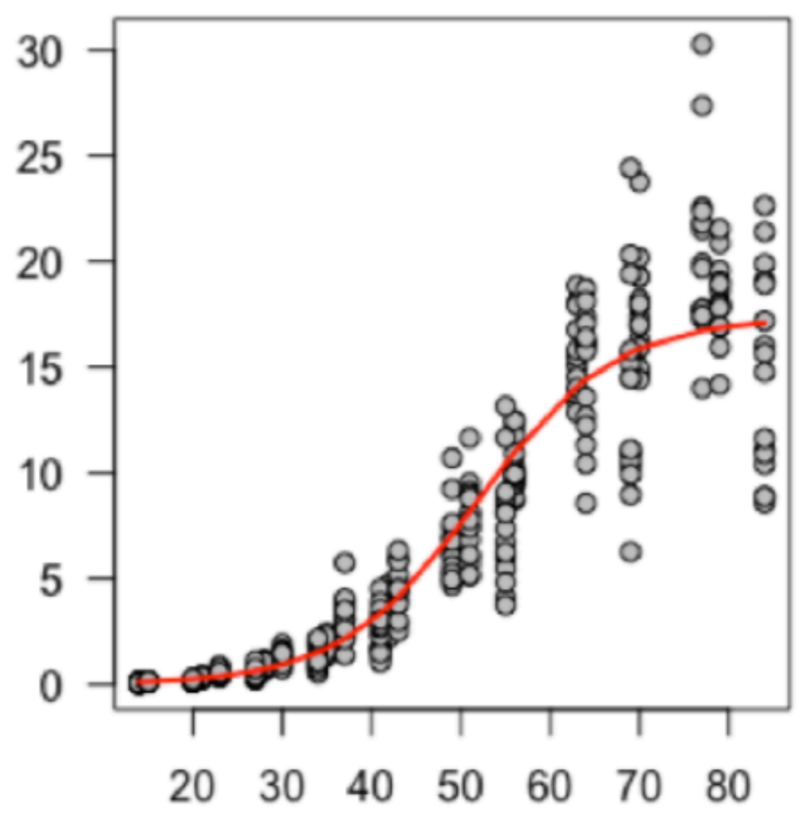


This is a PLOT

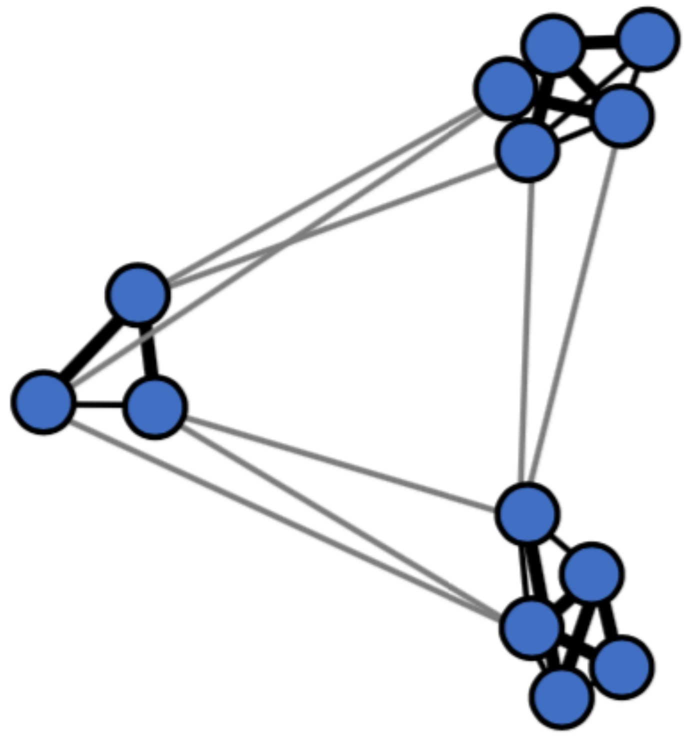


This is GRAPH
(a.k.a. network)

Sidestep: Graphs



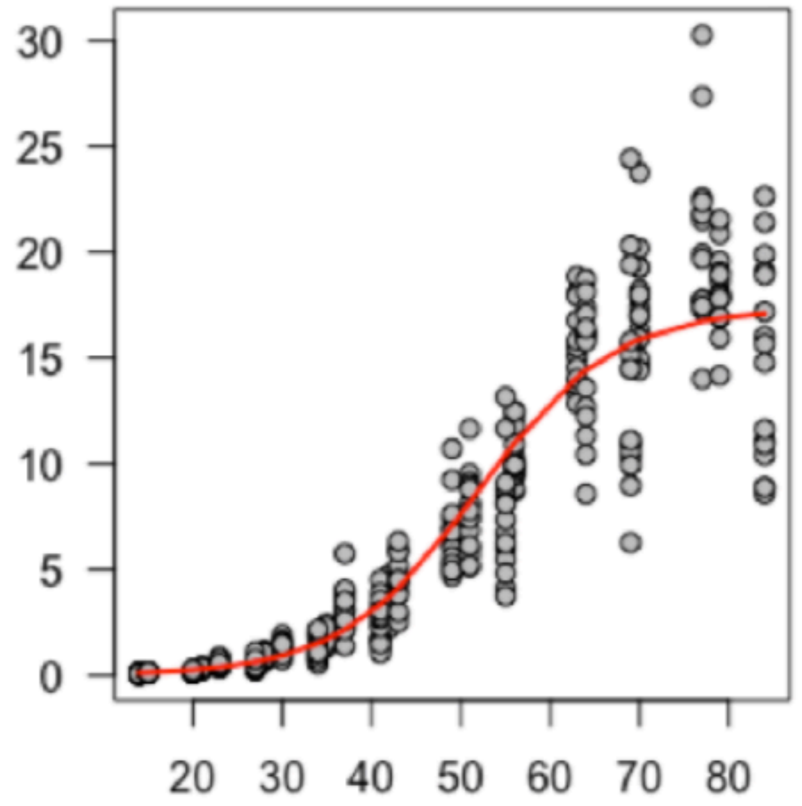
This is a <u>PLOT</u>



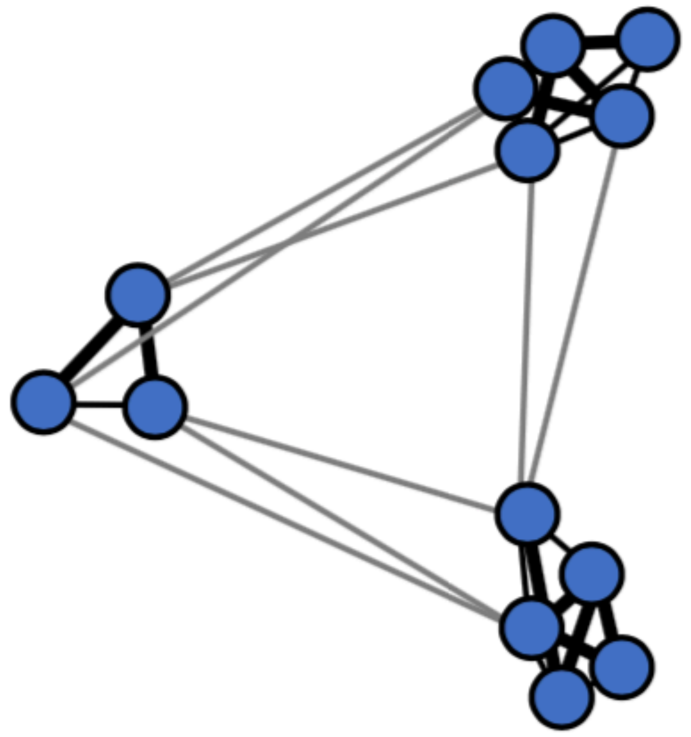
This is <u>GRAPH</u>
(a.k.a. network)

- Each dot is a cell (or a gene)
- Each line represents a connection between 2 cells
- Each connection can be weighted as a proximity between cells
  - Correlation (high and positive)
  - Euclidean distance (low)
  - etc.

Sidestep: Graphs



This is a PLOT



This is GRAPH
(a.k.a. network)

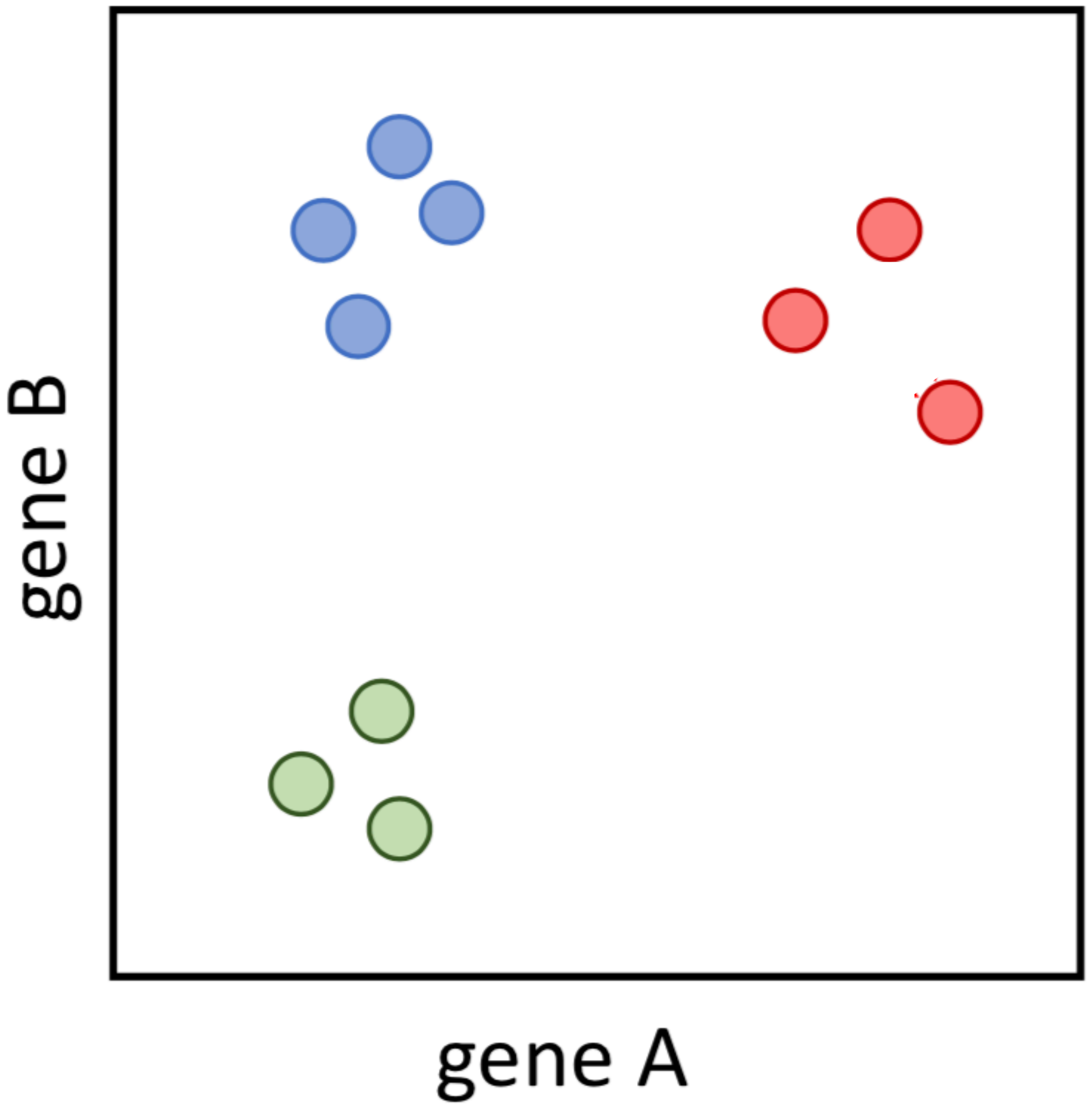
- Each dot is a cell (or a gene)
- Each line represents a connection between 2 cells
- Each connection can be weighted as a proximity between cells
  - Correlation (high and positive)
  - Euclidean distance (low)
  - etc.

Graph-based dimensionality reduction algorithms can be divided into 2 main steps:

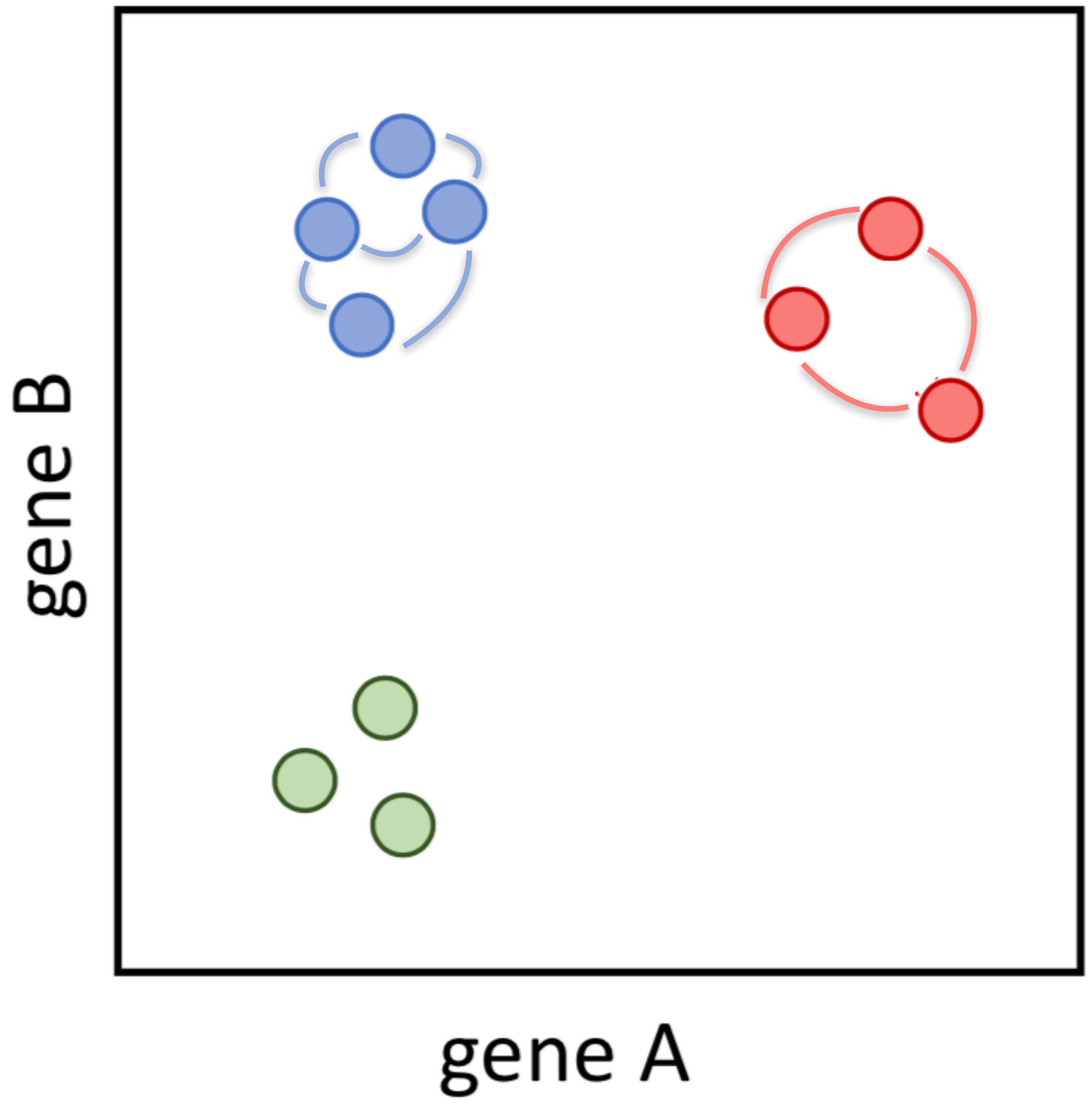
1. Construct a weighted graph based on the top k connections (a.k.a. k-nearest neighbors, KNN)

2. The low dimensional layout of the graph is computed and optimized

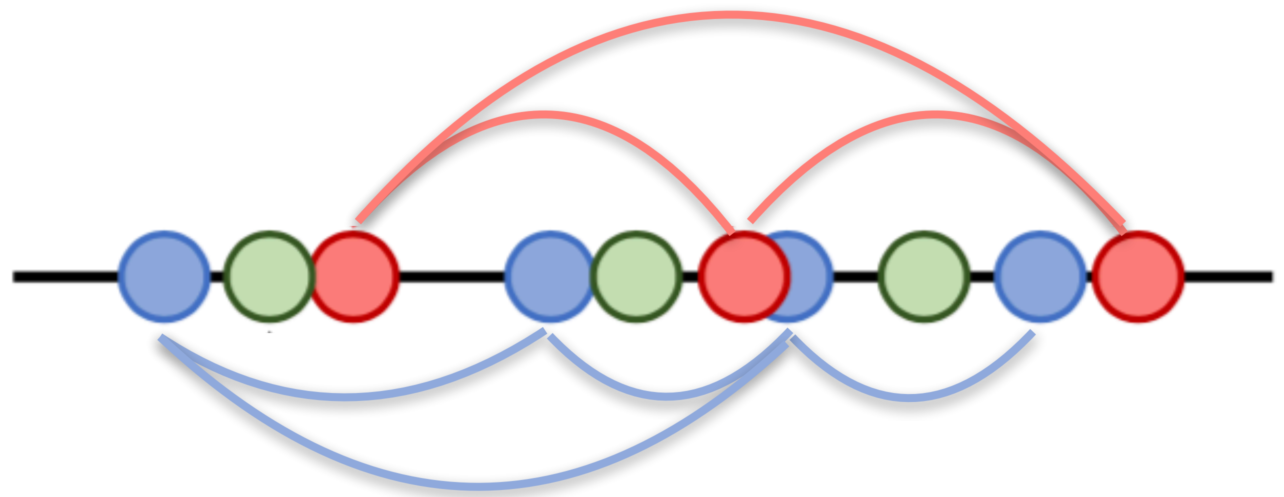
t-SNE in Brief



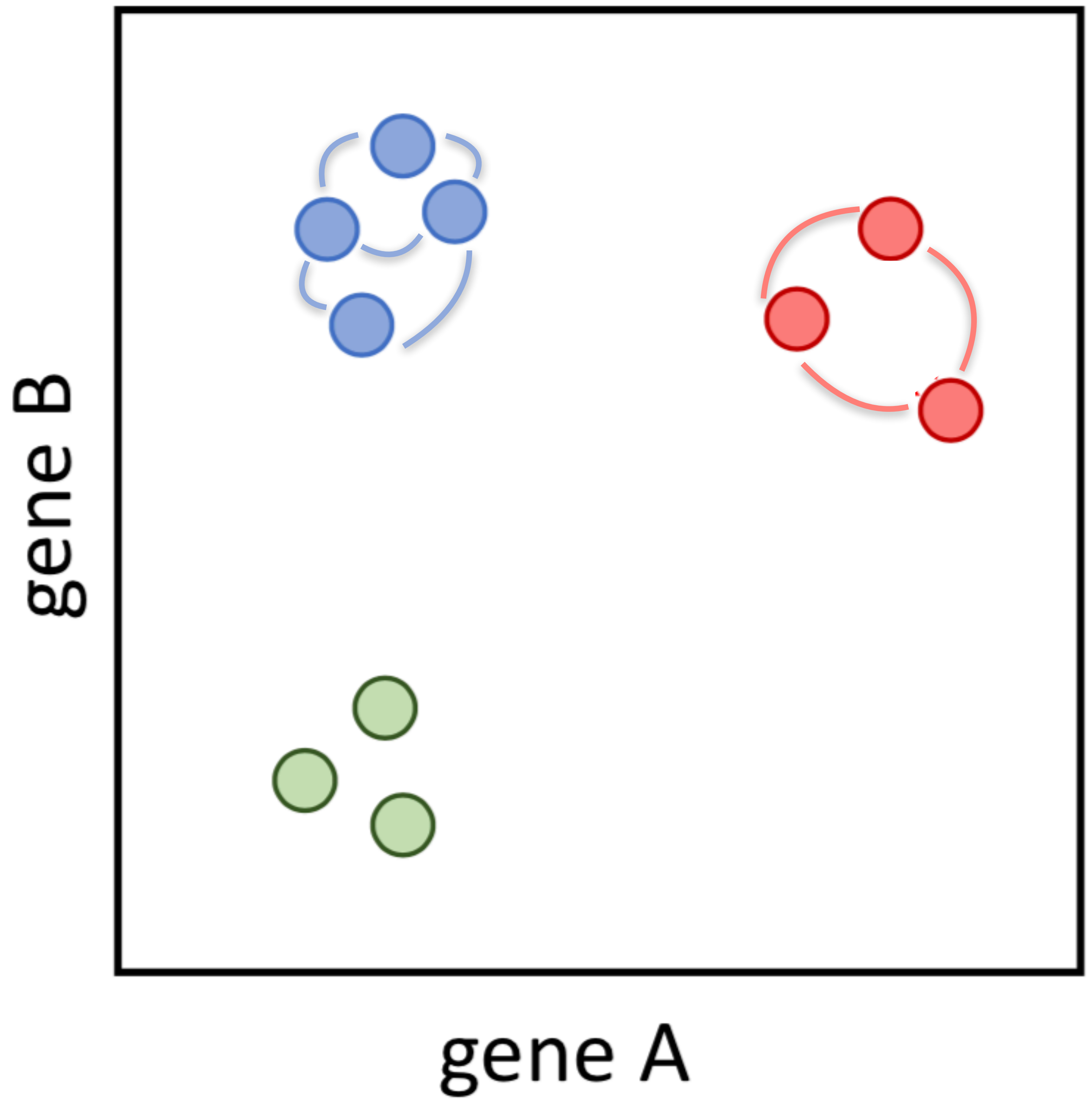
t-SNE in Brief



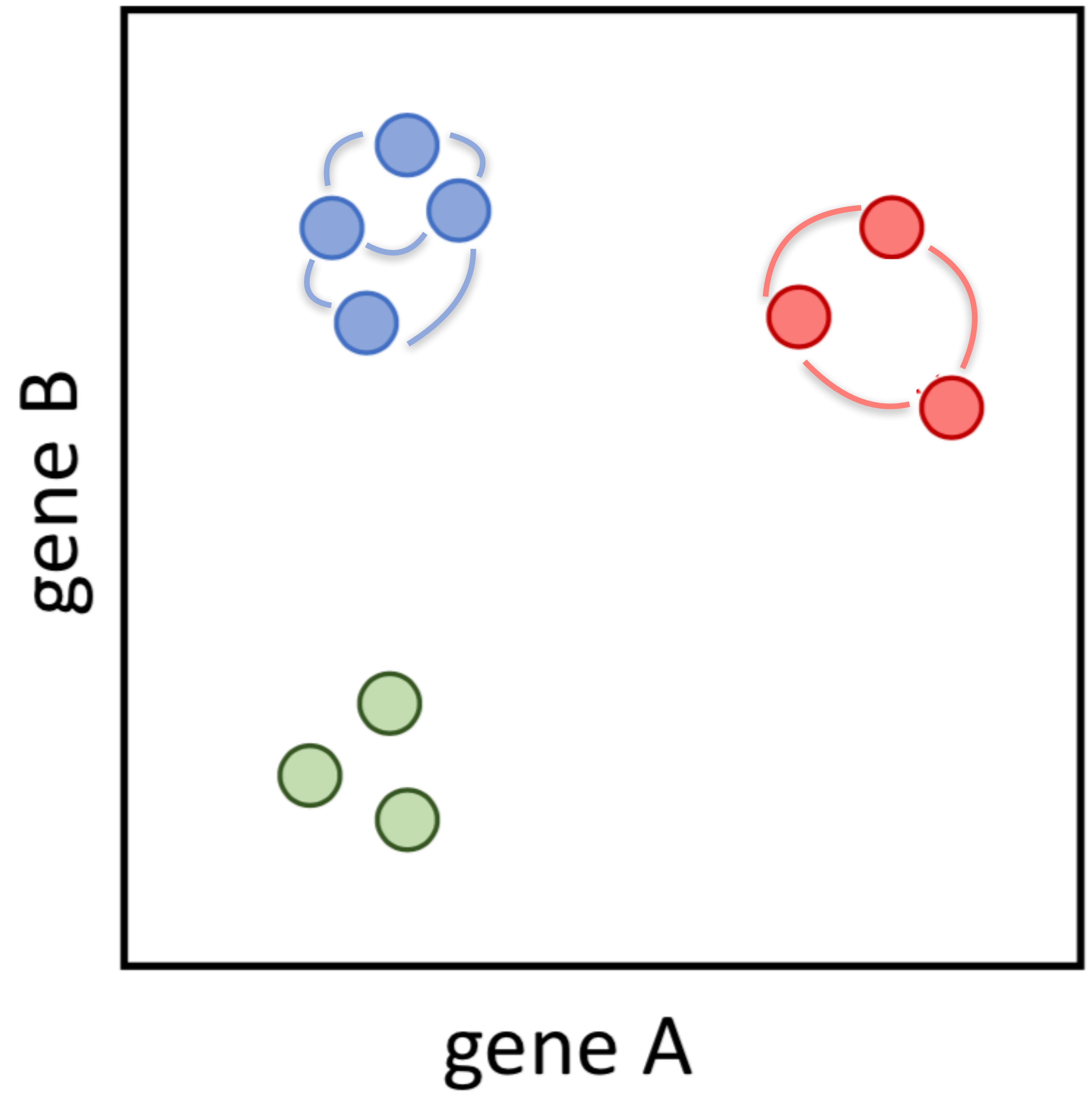
t-SNE in Brief

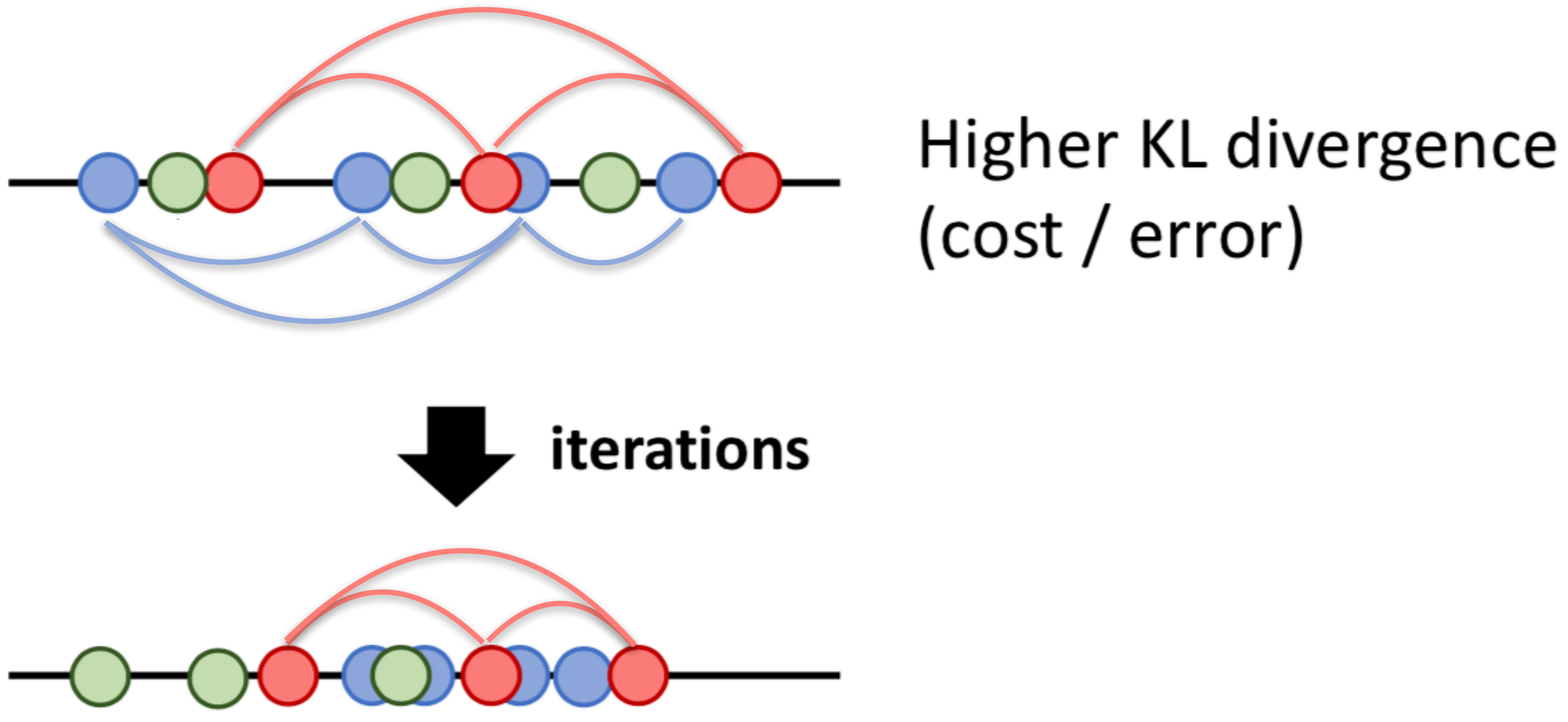


Higher KL divergence
(cost / error)

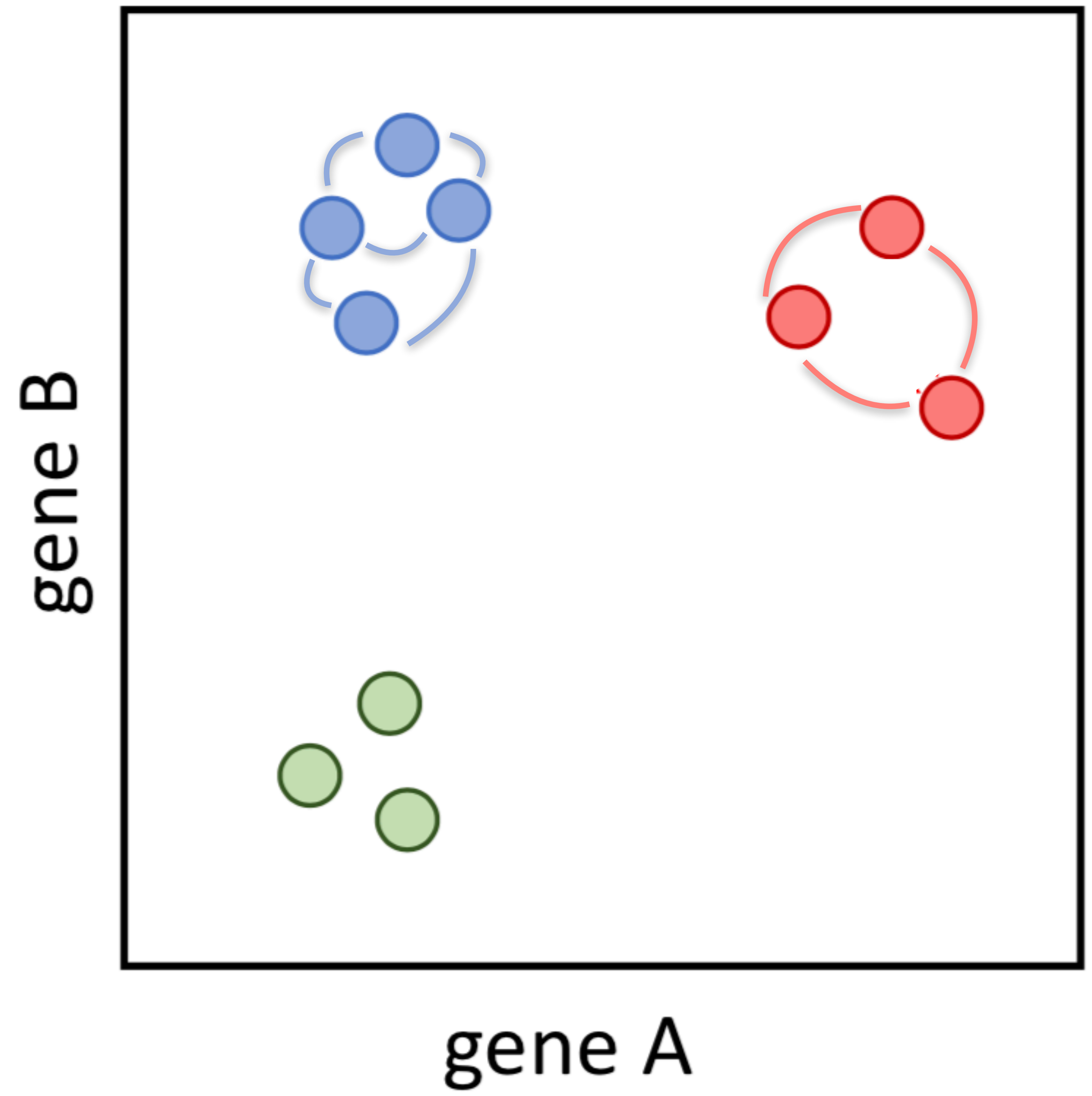


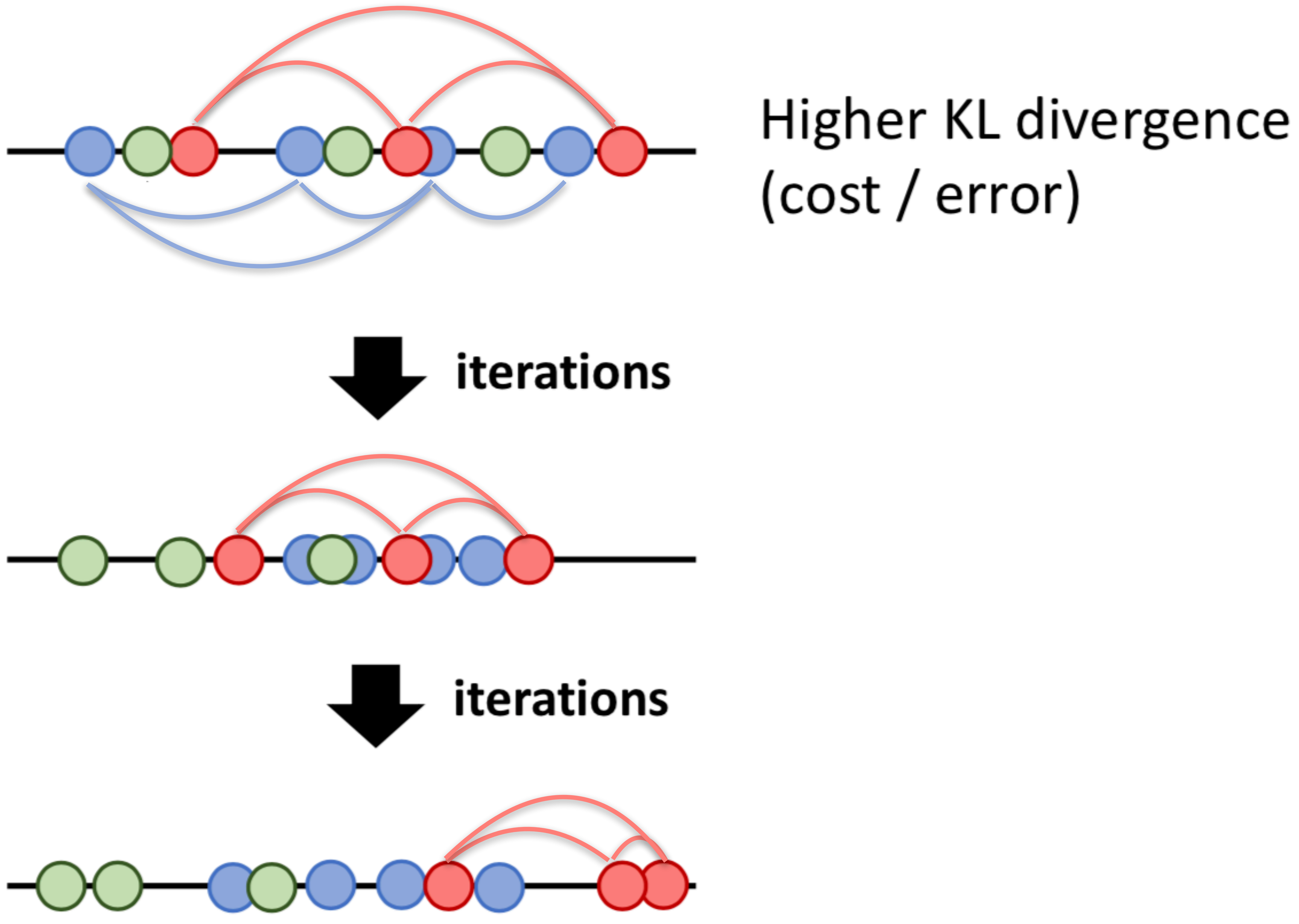
t-SNE in Brief



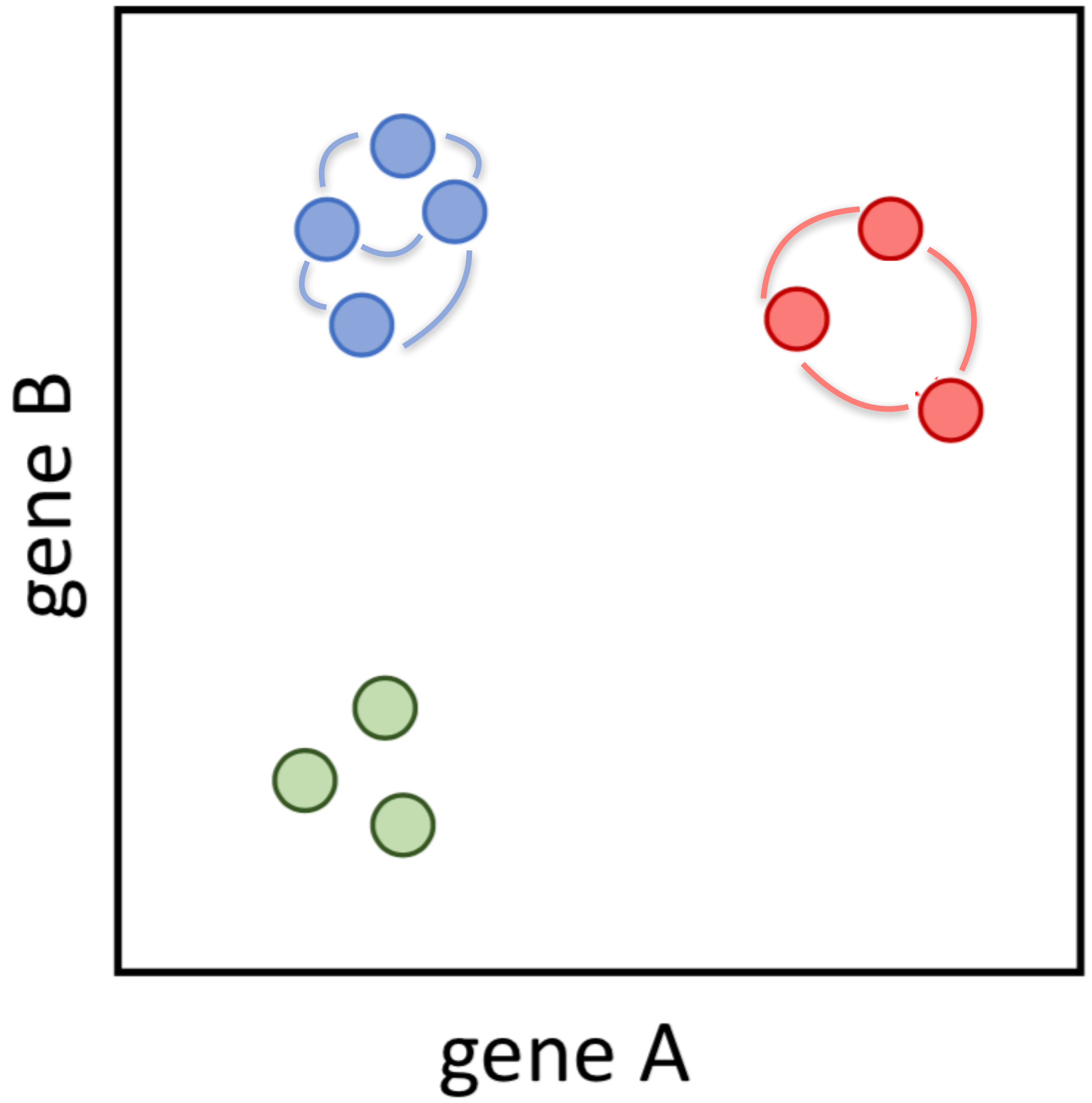


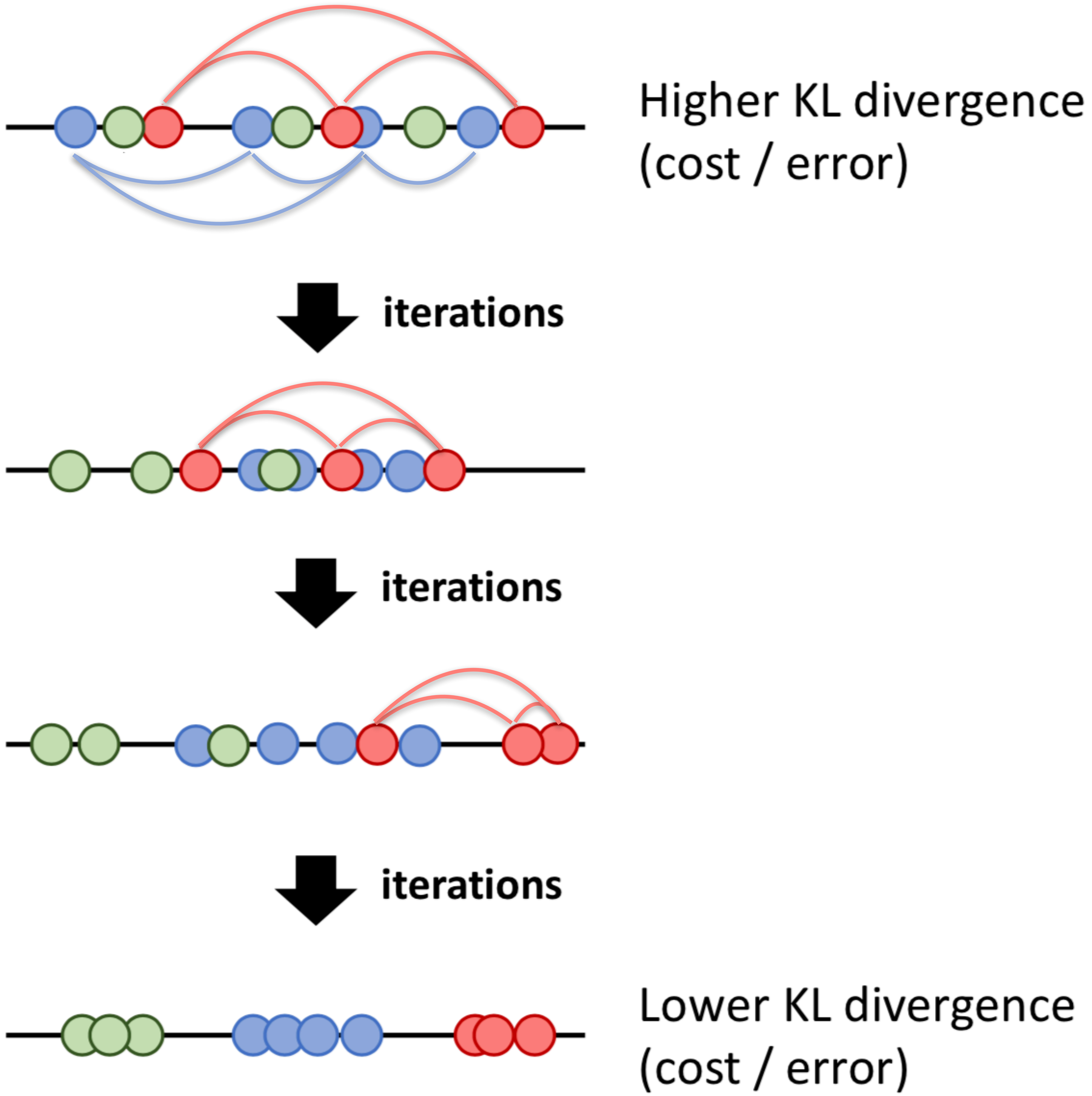
t-SNE in Brief





t-SNE in Brief





t-SNE in Brief

- 2 major computation parts in tSNE
  - compute high dimensional neighborhoods
  - optimize low dimensional neighborhoods
- Computationally intensive
  - Barnes-Hut implementation improves this
- Several parameters
  - Some can severely impact results

https://distill.pub/2016/misread-tsne/

t-SNE in Brief

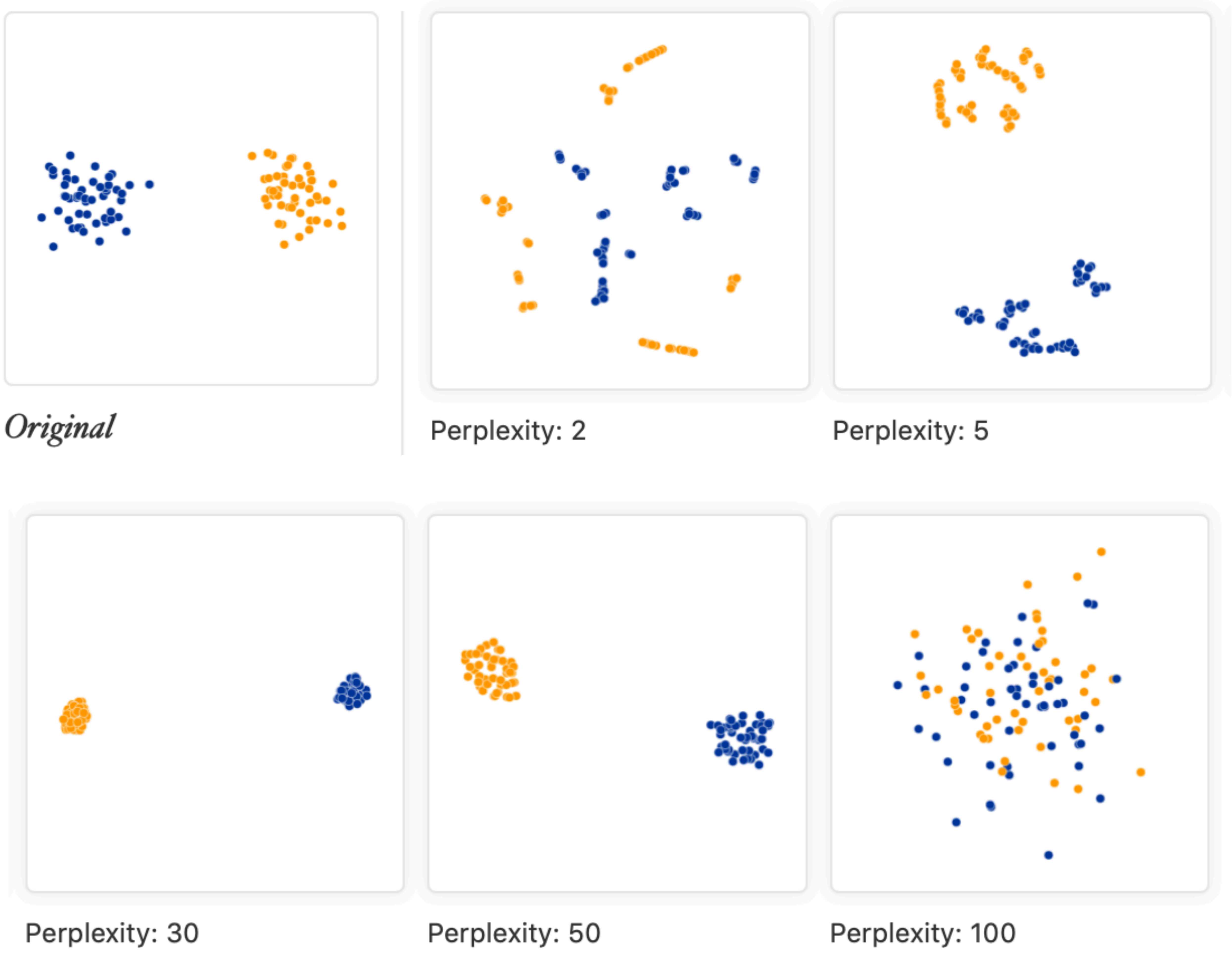
- 2 major computation parts in tSNE
  - compute high dimensional neighborhoods
  - optimize low dimensional neighborhoods
- Computationally intensive
  - Barnes-Hut implementation improves this
- Several parameters
  - Some can severely impact results

https://distill.pub/2016/misread-tsne/

t-SNE Parameters

- **Perplexity**
- Number of iterations
- Learning rate
- Theta (for BH t-SNE)

…

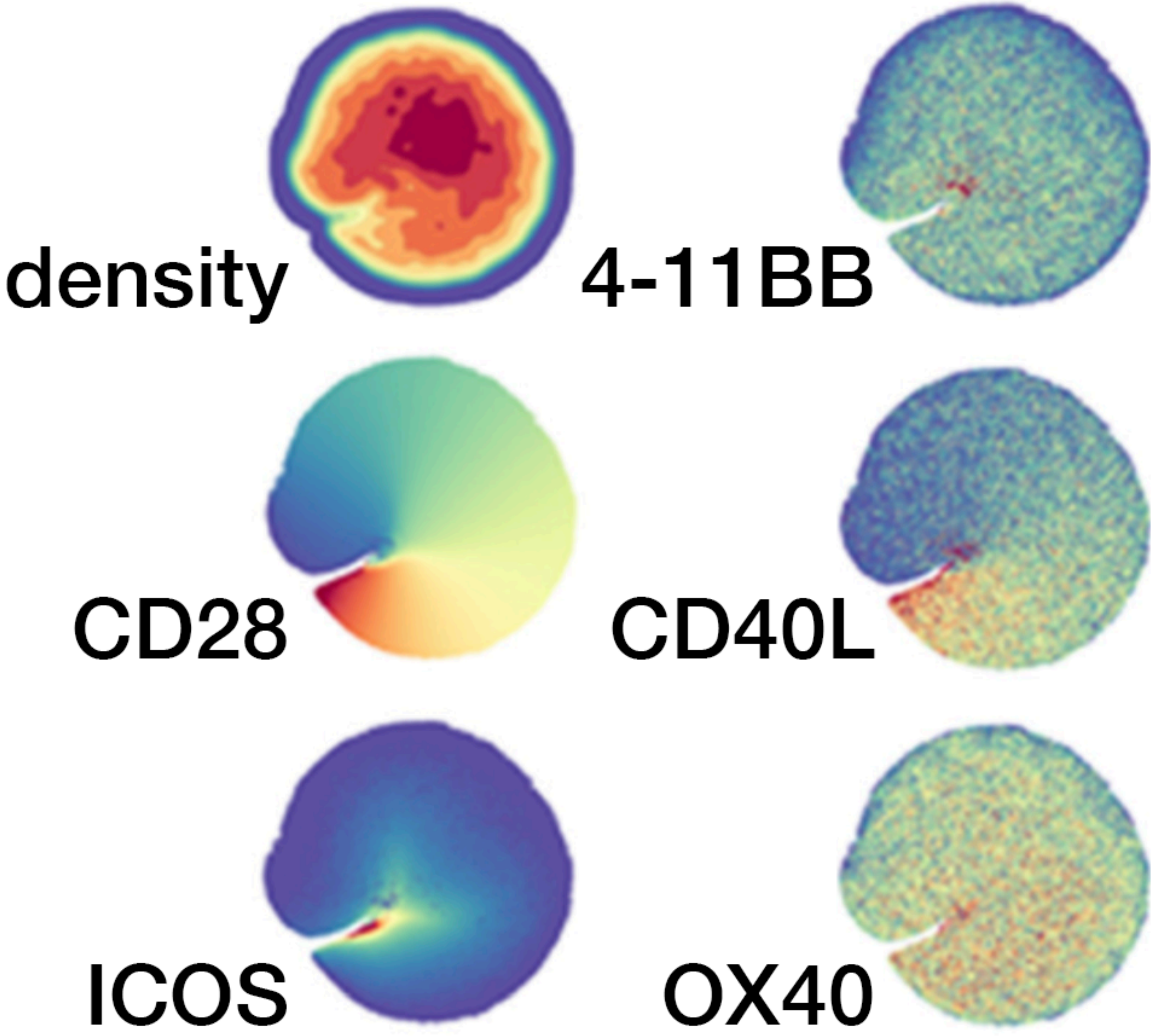


https://distill.pub/2016/misread-tsne/

t-SNE Parameters

- Perplexity
- **Number of iterations**
- Learning rate
- Theta (for BH t-SNE)

…



Important Notes

- Typically, the optimization is initialized randomly
  Multiple runs will produce different results

- The cost function never reaches the minimum

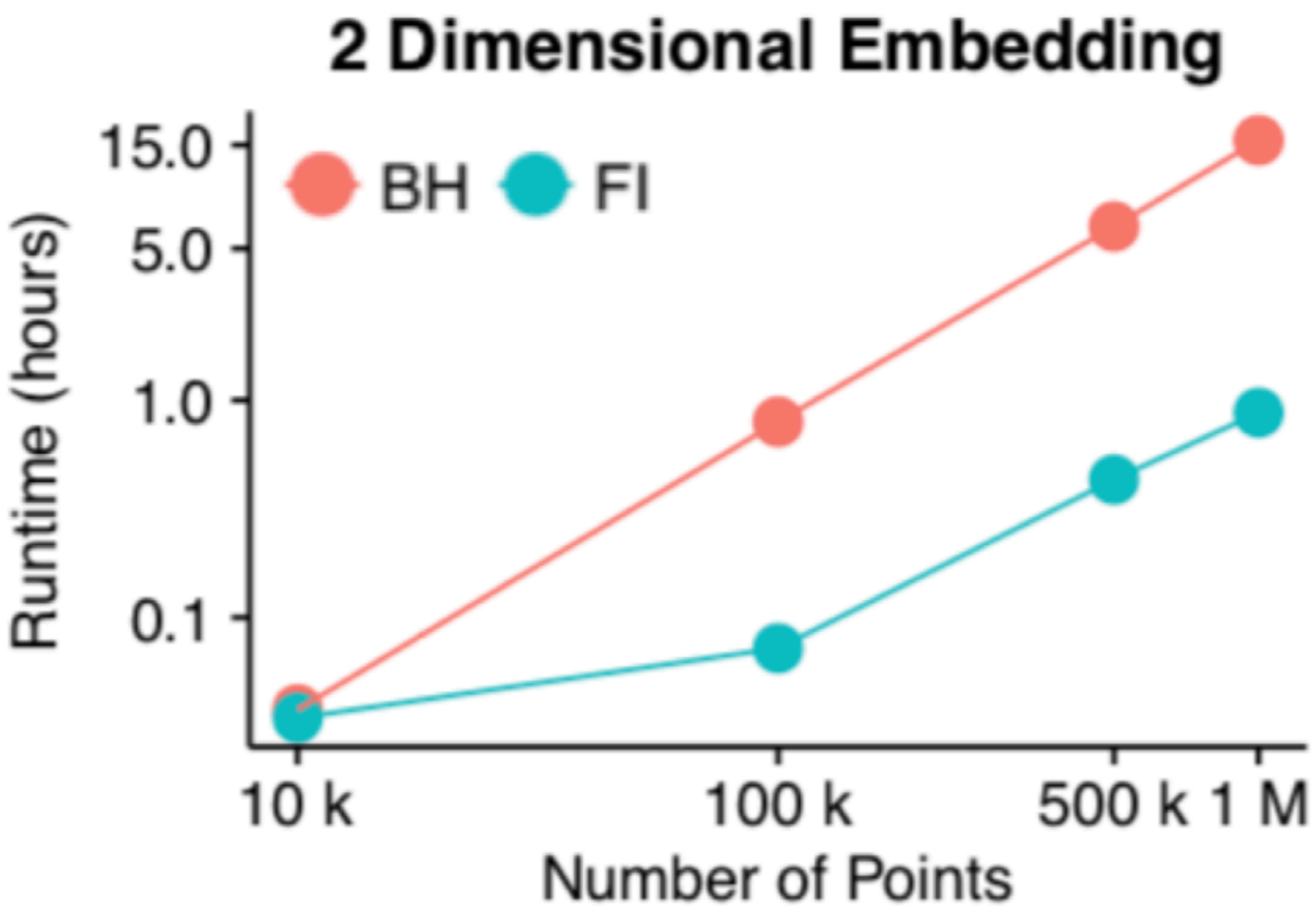
- t-SNE optimizes the distance between close points (local embedding)
  Distances within a group are slightly meaningful, but not between groups!

- To add more samples, you need to re-run the algorithm from start.

t-SNE Implementations

- Many Implementations available
- Fast Fourier Transform-accelerated

https://www.nature.com/articles/s41592-018-0308-4



Implementations

Below, implementations of t-SNE in various languages are available for download. Some of these implementations were developed by me, and some by other contributors. For the standard t-SNE method, implementations in Matlab, C++, CUDA, Python, Torch, R, Julia, and JavaScript are available. In addition, we provide a Matlab implementation of parametric t-SNE (described here). Finally, we provide a Barnes-Hut implementation of t-SNE (described here), which is the fastest t-SNE implementation to date, and which scales much better to big data sets.

You are free to use, modify, or redistribute this software in any way you want, but only for non-commercial purposes. The use of the software is at your own risk; the authors are not responsible for any damage as a result from errors in the software.

NOTE: t-SNE is now built-in functionality in Matlab and in SPSS!

Matlab implementation (user guide)	All platforms
CUDA implementation (by David, Roshan, and Forrest; see paper)	All platforms
Python implementation	All platforms
Go implementation (by Daniel Salvadori)	All platforms
Torch implementation	All platforms
Julia implementation (by Leif Jonsson)	All platforms
Java implementation (by Leif Jonsson)	All platforms
R implementation (by Justin)	All platforms
JavaScript implementation (by Andrej; online demonstration)	All platforms
Parametric t-SNE (outdated; see here)	All platforms
Barnes-Hut t-SNE (C++, Matlab, Python, Torch, and R wrappers; see here)	All platforms / Github
MNIST Dataset	Matlab file

https://lvdmaaten.github.io/tsne/

GPU t-SNE

PyPI

pip install nptsne

https://pypi.org/project/nptsne/
https://www.github.com/biovault/

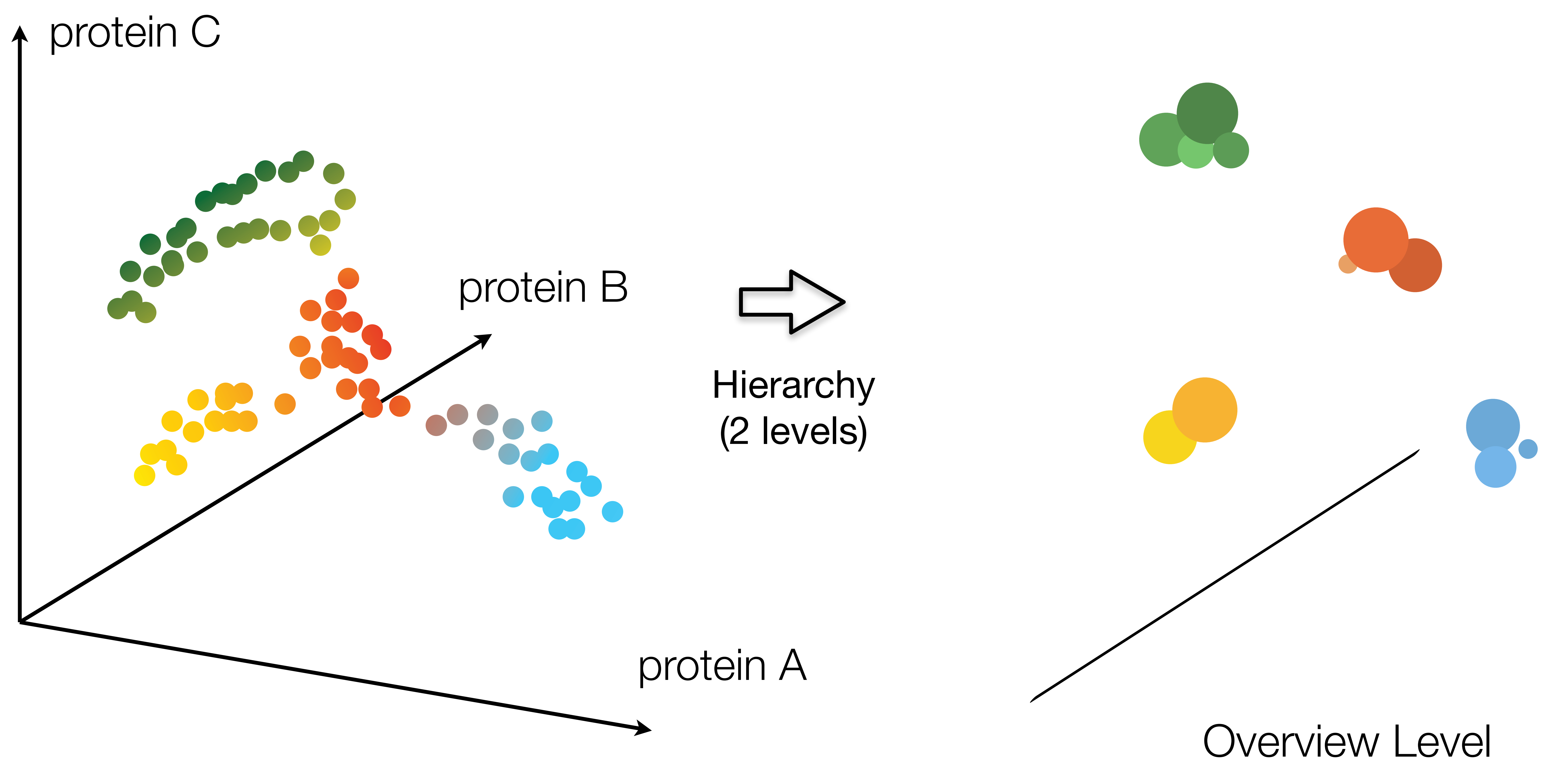
Summary: t-SNE

- NON-LINEAR method of dimensionality reduction
- It is the current GOLD-STANDARD method in single cell data (including scRNA-seq)
- Can be run from the top PCs (e.g.: PC1 to PC10)

Problems:

- It does not learn an explicit function to map new points
- It’s cost function is not convex – This means that the optimal t-SNE cannot be computed
- Many hyper-parameters need to be defined empirically (dataset-specific)
- It does not preserve global structure (in practice)

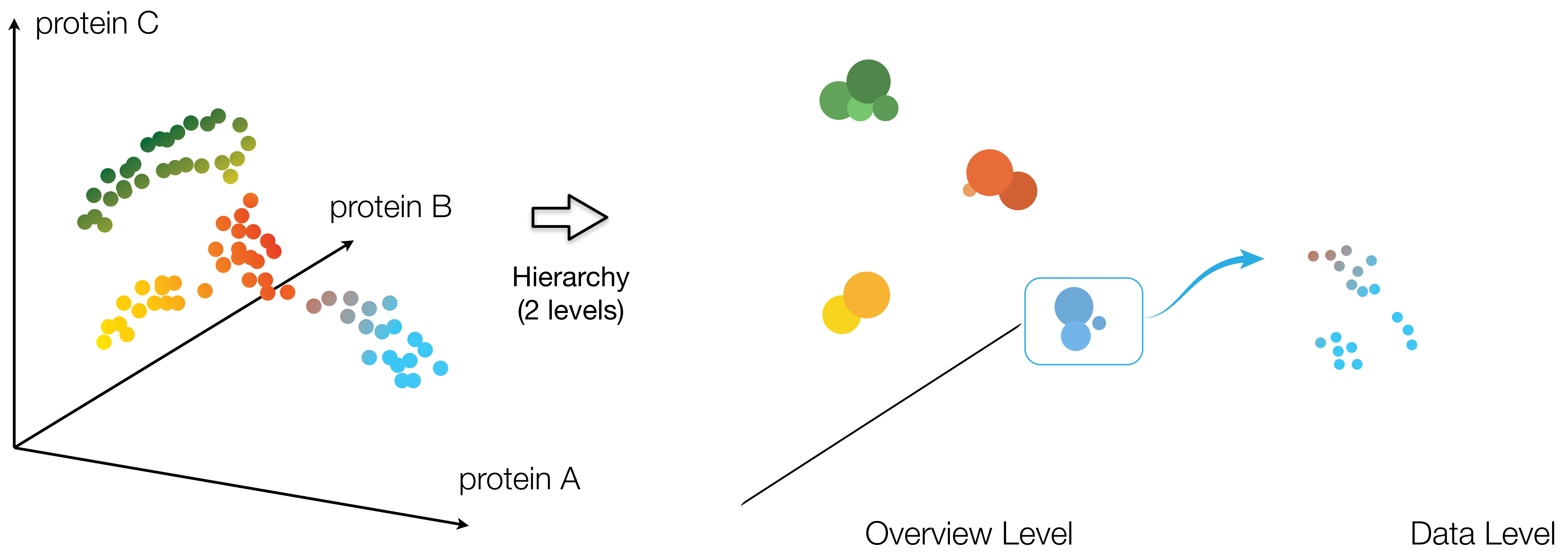
Hierarchical SNE



Pezzotti et al., Computer Graphics Forum, 2016
Van Unen, Höllt, Pezzotti, et al., Nature Communications, 2017

https://www.cytosplore.org

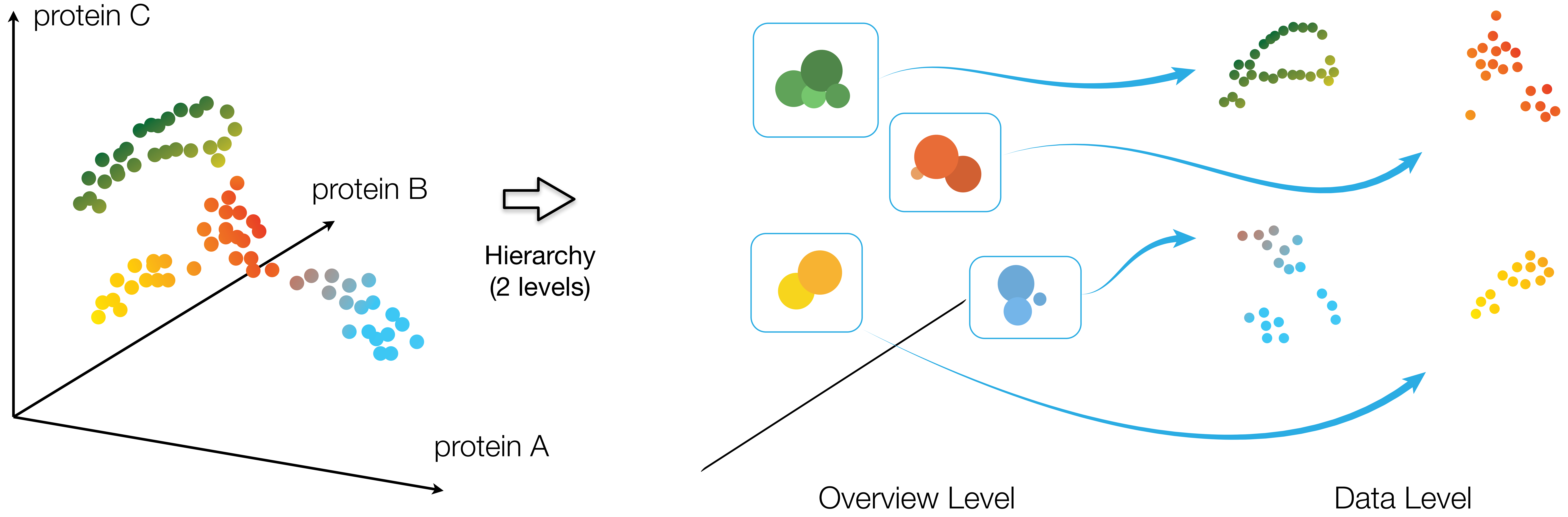
Hierarchical SNE



Pezzotti et al., Computer Graphics Forum, 2016
Van Unen, Höllt, Pezzotti, et al., Nature Communications, 2017

https://www.cytosplore.org

Hierarchical SNE



Pezzotti et al., Computer Graphics Forum, 2016
Van Unen, Höllt, Pezzotti, et al., Nature Communications, 2017

https://www.cytosplore.org

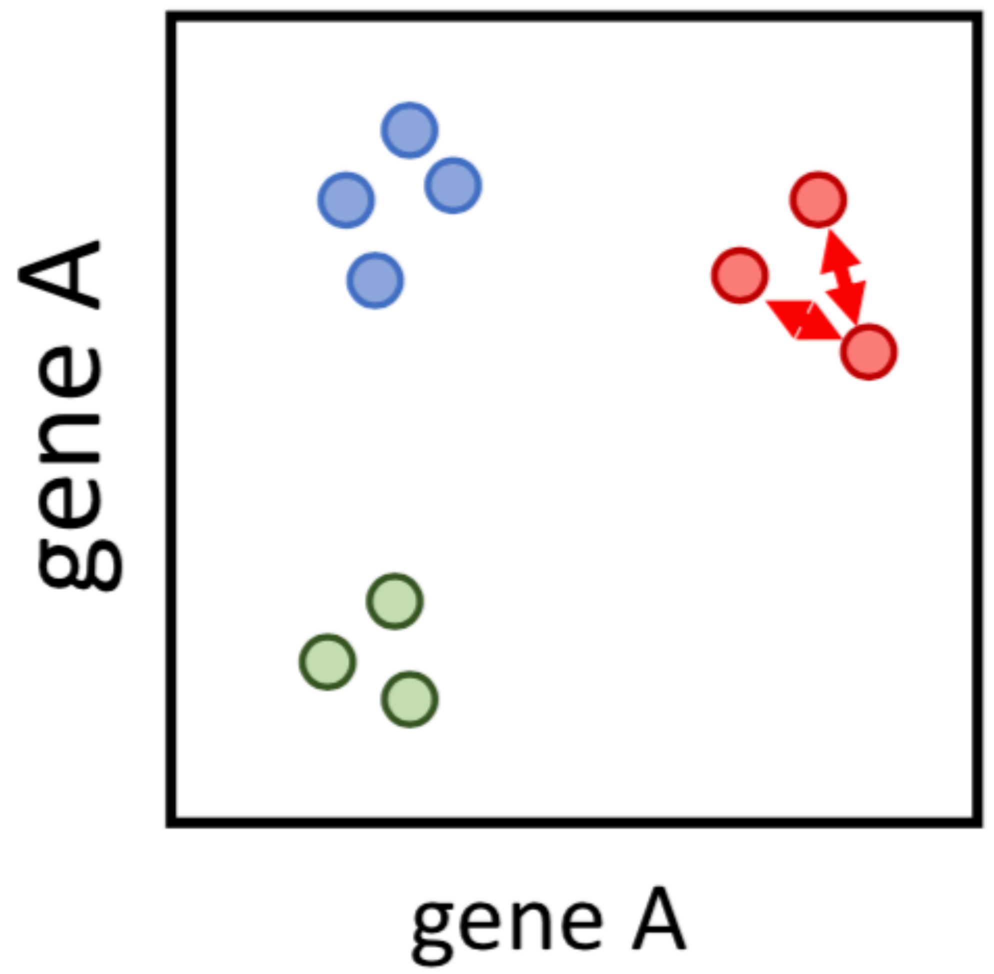

Cytosplore
Settings
HSNE Analysis
Input Data
All Points
Downsampling ...
Active Markers
Time
Event_length
Viability
Ba138Di
CD103
Bead1
CCR6
CD19
C-KIT
CD11b
CD4
CD8a
CD7
CD34
CD25
TCRab
CD123_Bead2
TCRgd
IL-21R_Bead3
CD45
CRTH2
CD122
CCR7
CD14
CD11c
CD161
CD127_Bead4
CD8b
CD27
IL-15Ra
CD45RA
CD3
CD28
CD38
NKp46
PD-1_Bead5
CD56
BCKG190Di
DNA
DNA
Pb206Di
sampleTag
CSPLR_STATE
CSPLR_SUBSET
Inverse Markers
Load Selection
Set Default
Advanced Parameters
Parameters
Number of Scales 5
Compute HSNE
HSNE Analysis - Overview Level
high
low
Single-Cell
CD3
0
6.83
0.00
1.71
3.42
5.12
6.83
Darkmode
Point Size
Point Opacity
Halo Size
HeatMap View
6.39
0.00
CD103
CCR6
CD19
C-KIT
CD11b
CD4
CD8a
CD7
CD34
CD25
TCRab
CD123_Bead2
TCRgd
IL-21R_Bead3
CD45
CRTH2
CD122
CCR7
CD14
CD11c
CD161
CD127_Bead4
CD8b
CD27
IL-15Ra
CD45RA
CD3
CD28
CD38
NKp46
PD-1_Bead5
CD56
Density Clusters
Variaton
Dendrogram
Labels
0.00
1.60
3.19
4.79
6.39


Cytosplore
Settings
HSNE Analysis
Input Data
All Points
Downsampling ...
Active Markers
Time
Event_length
Viability
Ba138Di
CD103
Bead1
CCR6
CD19
C-KIT
CD11b
CD4
CD8a
CD7
CD34
CD25
TCRab
CD123_Bead2
TCRgd
IL-21R_Bead3
CD45
CRTH2
CD122
CCR7
CD14
CD11c
CD161
CD127_Bead4
CD8b
CD27
IL-15Ra
CD45RA
CD3
CD28
CD38
NKp46
PD-1_Bead5
CD56
BCKG190Di
DNA
DNA
Pb206Di
sampleTag
CSPLR_STATE
CSPLR_SUBSET
Inverse Markers
Load Selection
Set Default
Advanced Parameters
Parameters
Number of Scales 5
Compute HSNE
HSNE Analysis - Overview Level
high
low
Single-Cell
CD3
0
6.83
0.00
1.71
3.42
5.12
6.83
Darkmode
Point Size
Point Opacity
Halo Size
HeatMap View
CD103
CCR6
CD19
C-KIT
CD11b
CD4
CD8a
CD7
CD34
CD25
TCRab
CD123_Bead2
TCRgd
IL-21R_Bead3
CD45
CRTH2
CD122
CCR7
CD14
CD11c
CD161
CD127_Bead4
CD8b
CD27
IL-15Ra
CD45RA
CD3
CD28
CD38
NKp46
PD-1_Bead5
CD56
6.39
0.00
Density Clusters
Variaton
Dendrogram
Labels
0.00
1.60
3.19
4.79
6.39

UMAP
Uniform Manifold Approximation and Projection

UMAP Intuition

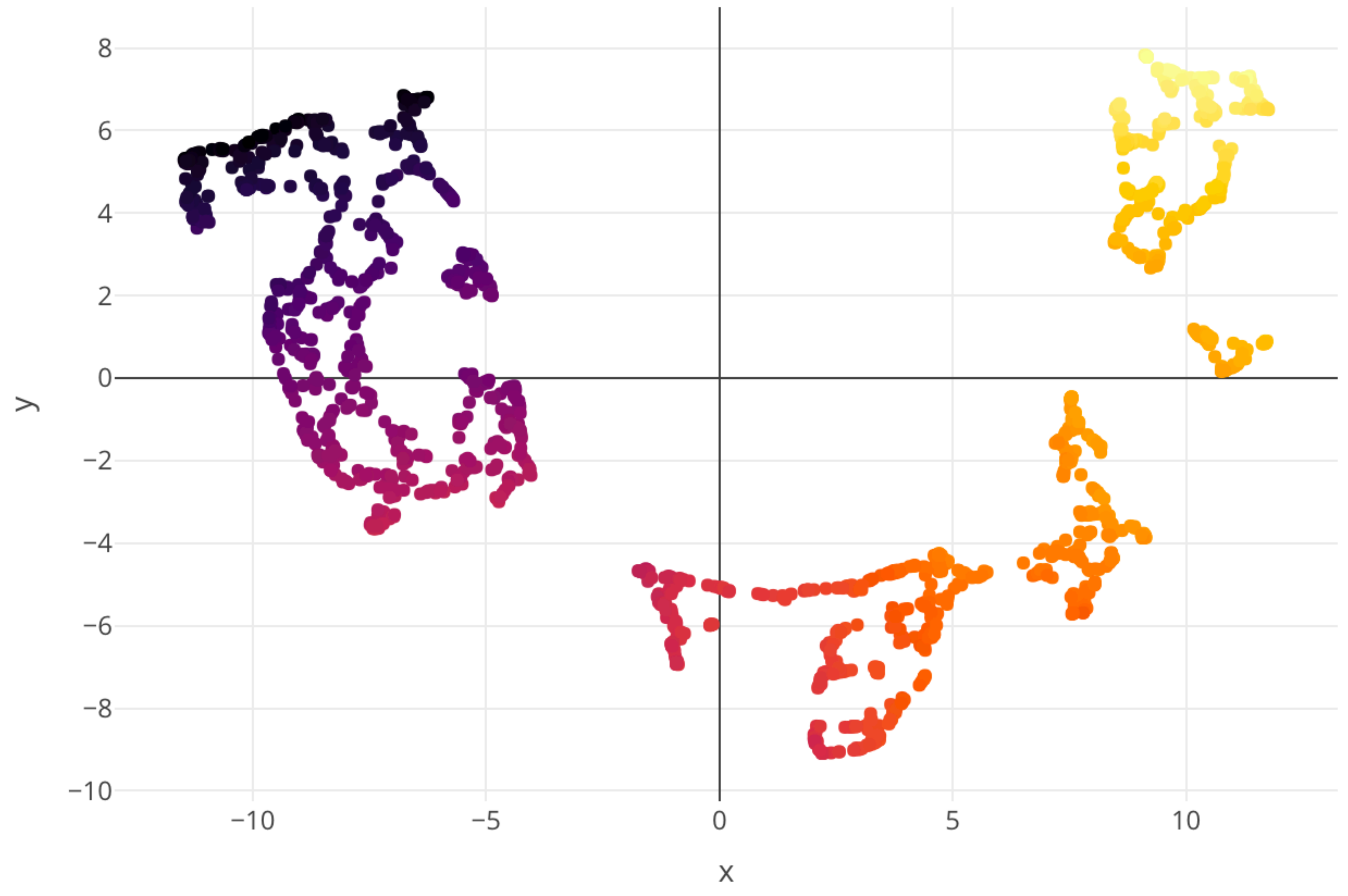
- Similar principle as t-SNE
- Initializes with a non-heuristic “guess"
  - Same result every time
- Resolves global structure somewhat better (due to initialization)



McInnes et al (2018) arXiv https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

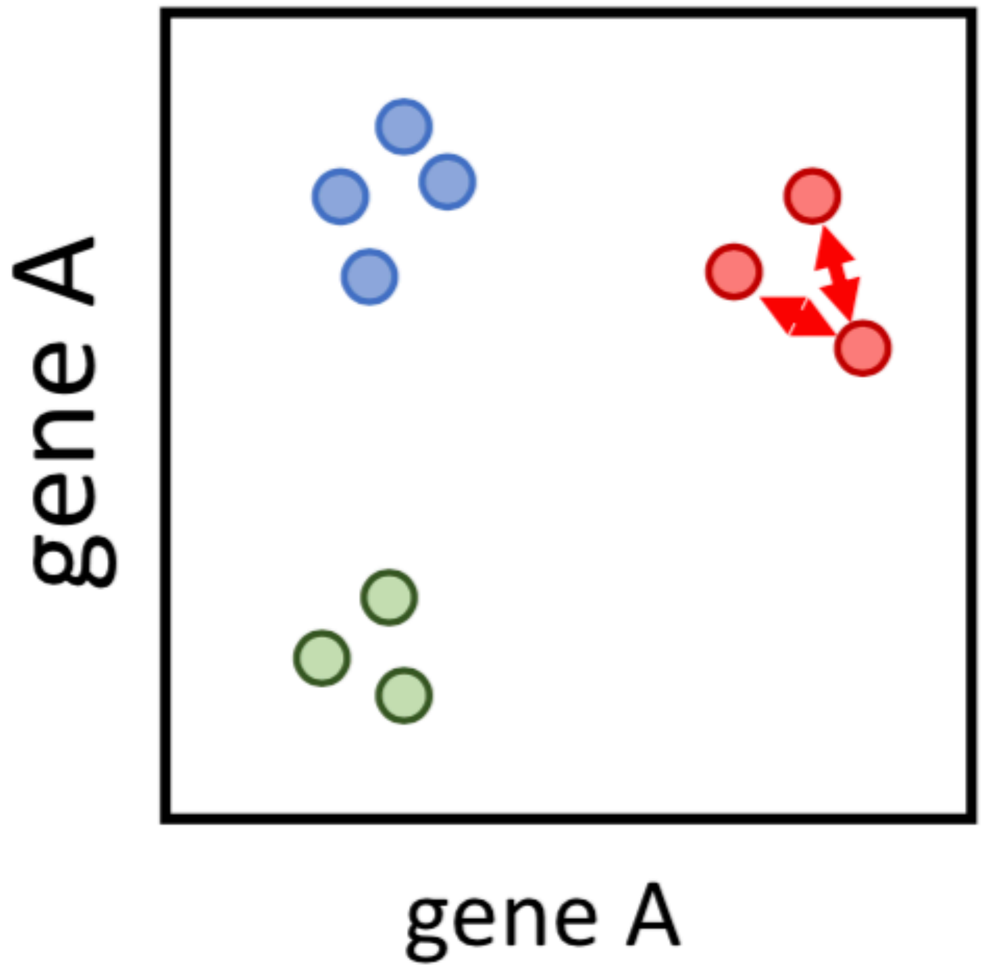
UMAP Intuition

- Similar principle as t-SNE
- Initializes with a non-heuristic “guess"
  - Same result every time
- Resolves global structure somewhat better (due to initialization)



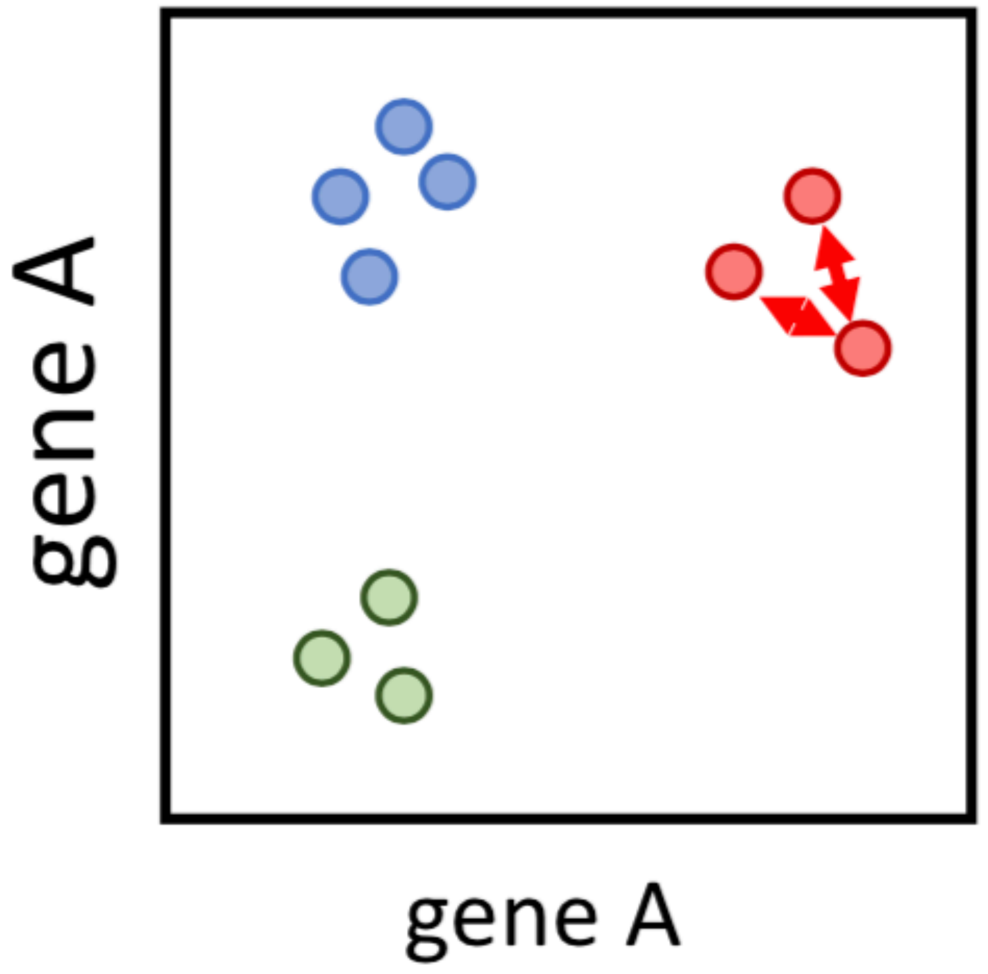
McInnes et al (2018) arXiv https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

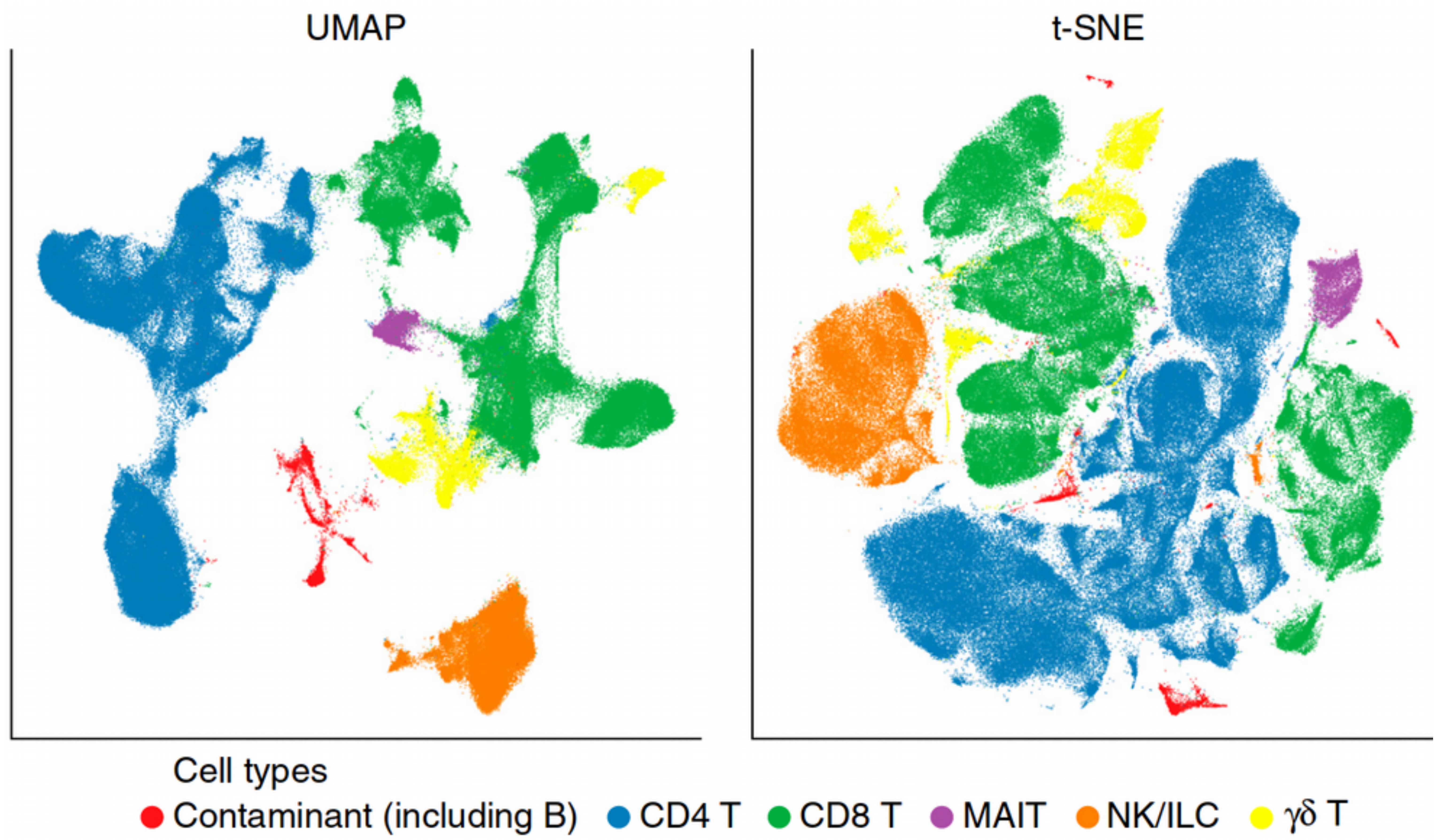
UMAP in Brief



Becht & McInnes et al (2019) *Nat Biot*

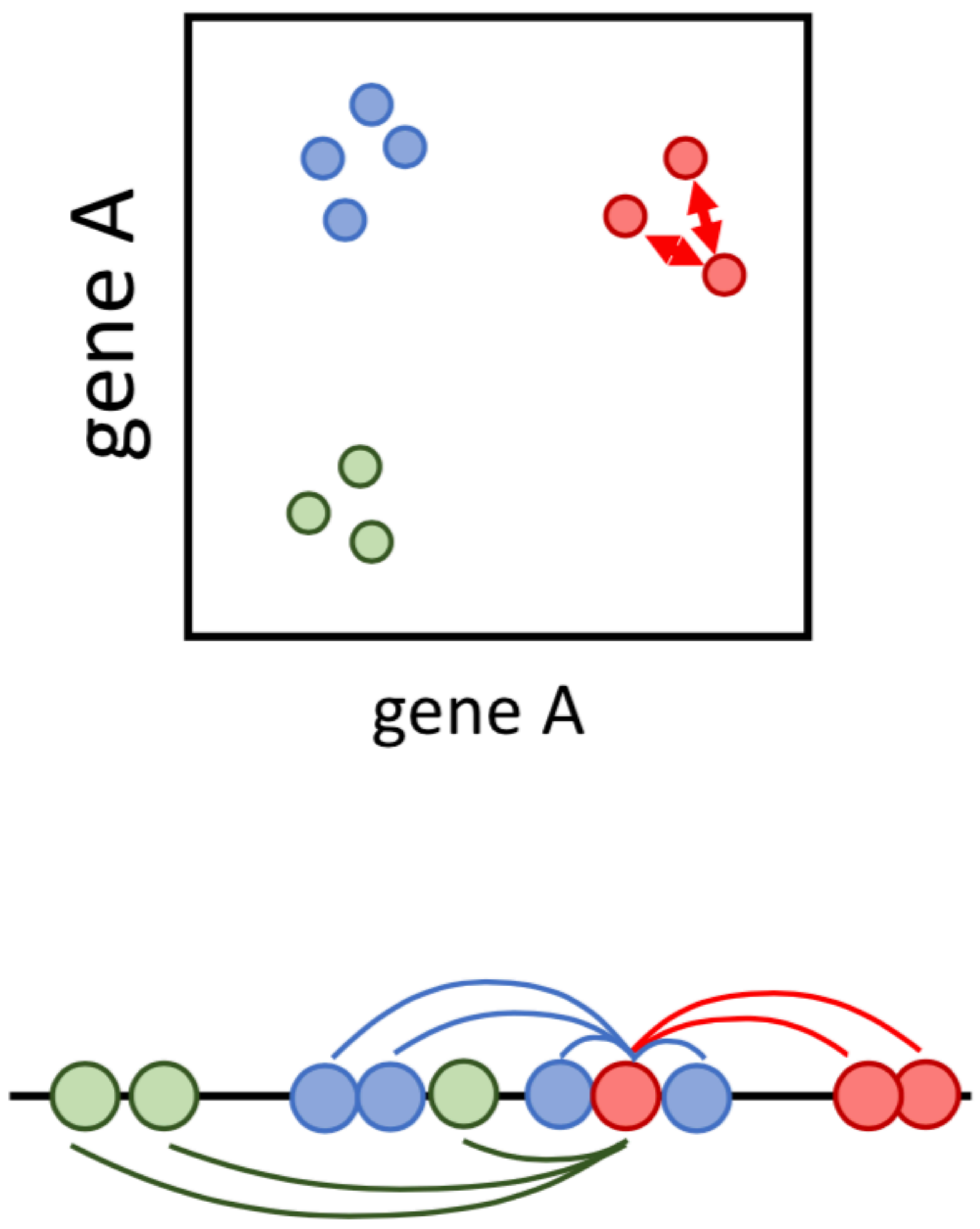
UMAP in Brief

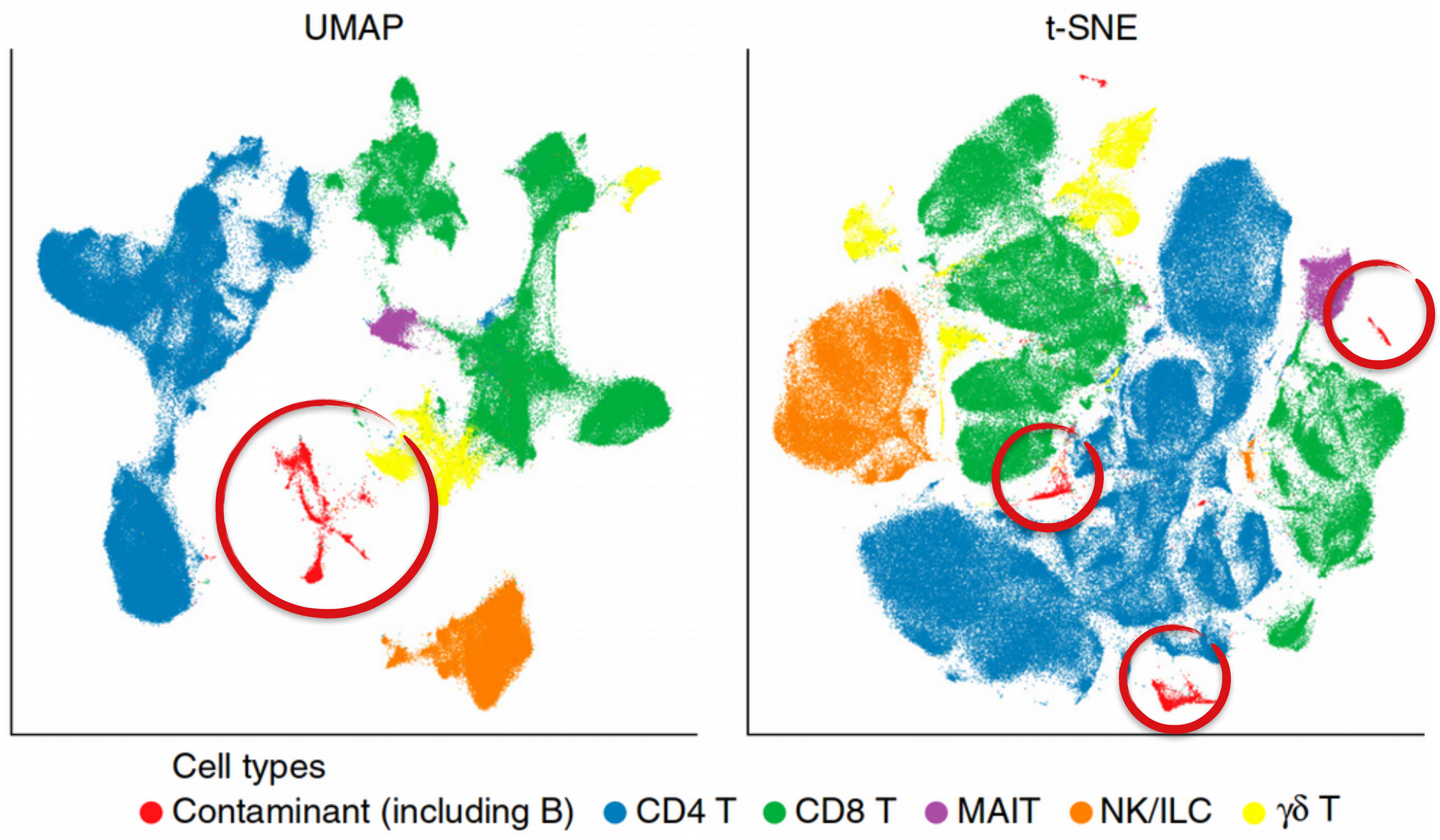




Becht & McInnes et al (2019) *Nat Biot*

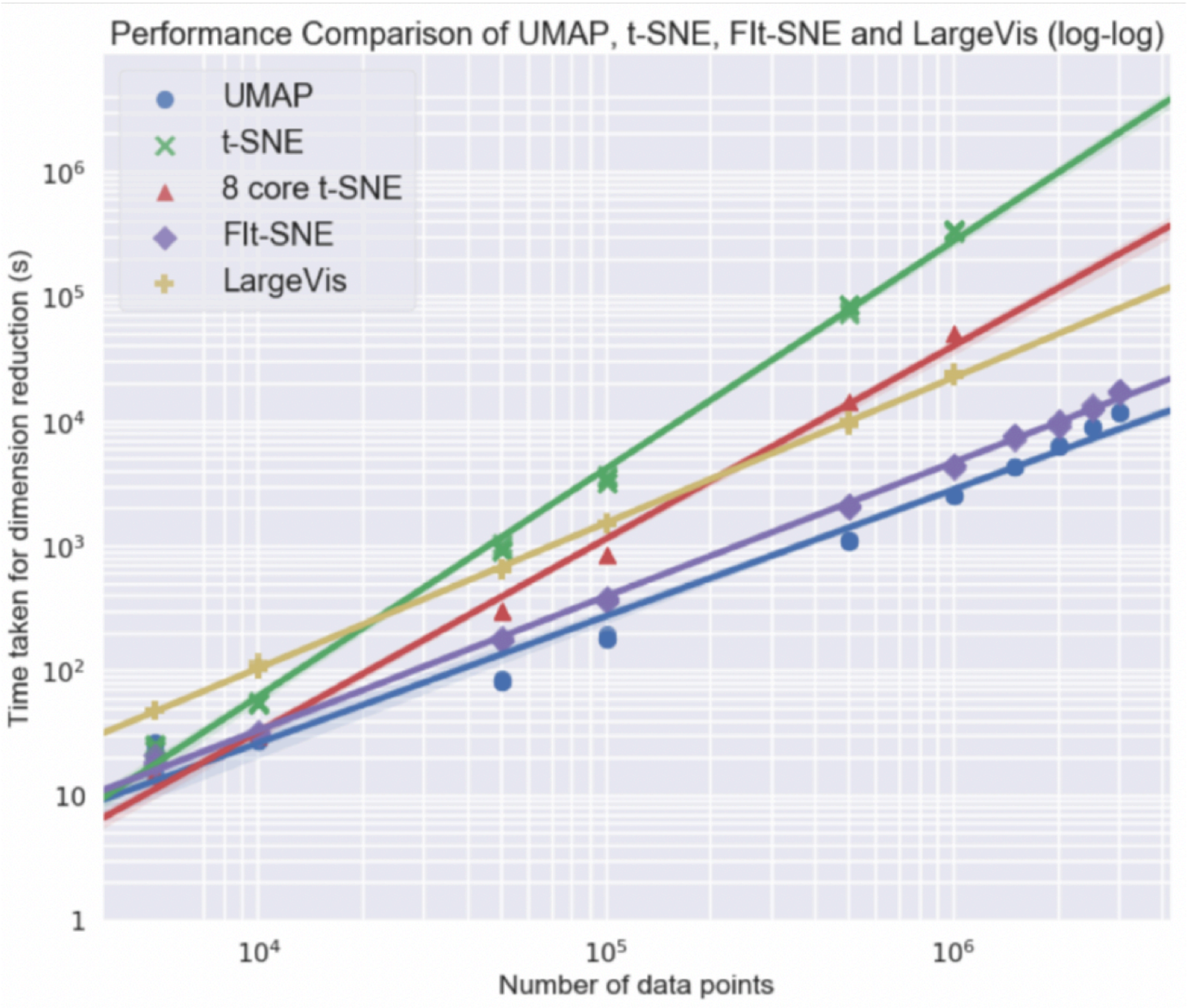
UMAP in Brief





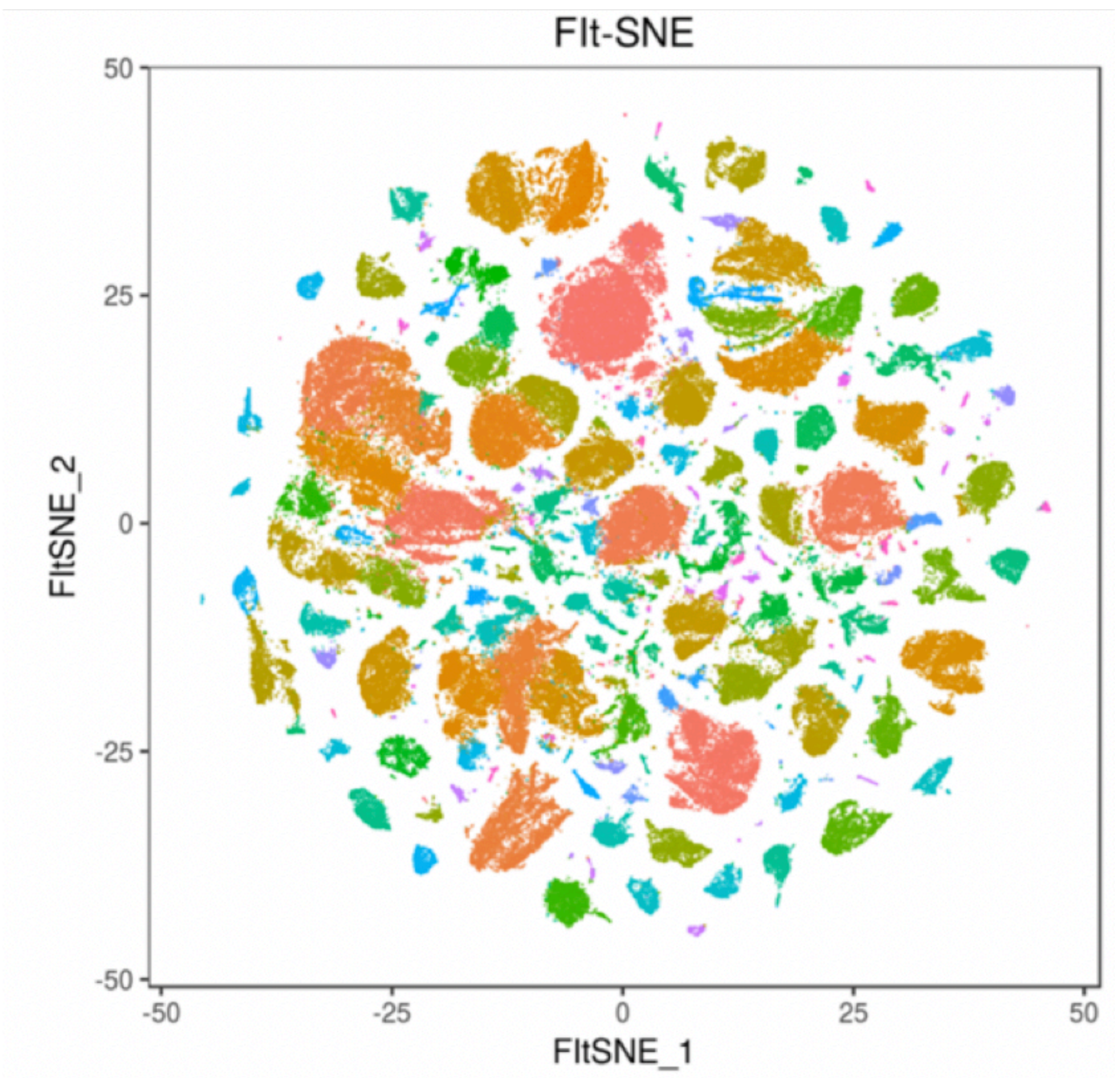
Becht & McInnes et al (2019) *Nat Biot*

UMAP Parameters

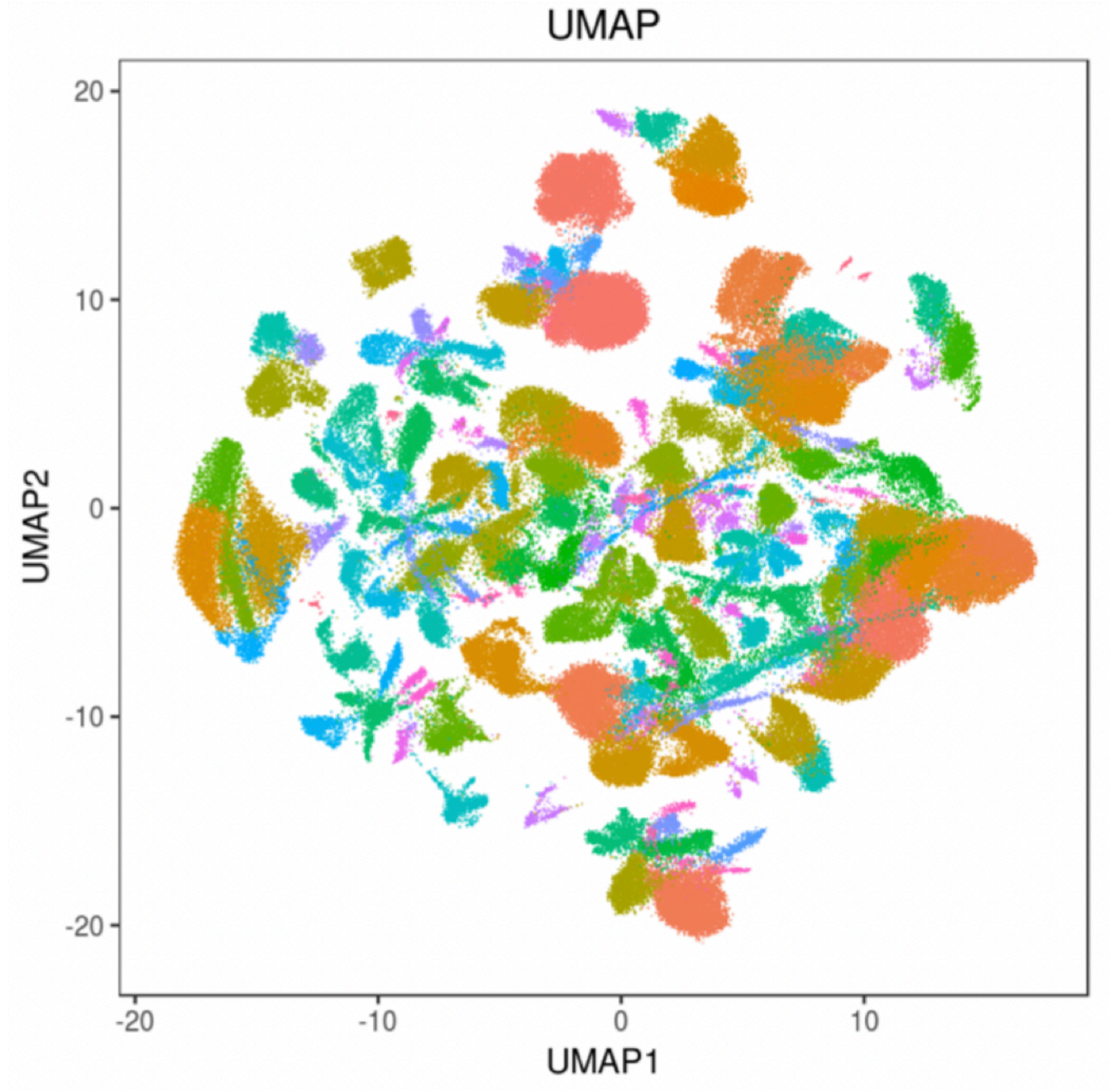


250k cells

FIt-SNE
1 hour



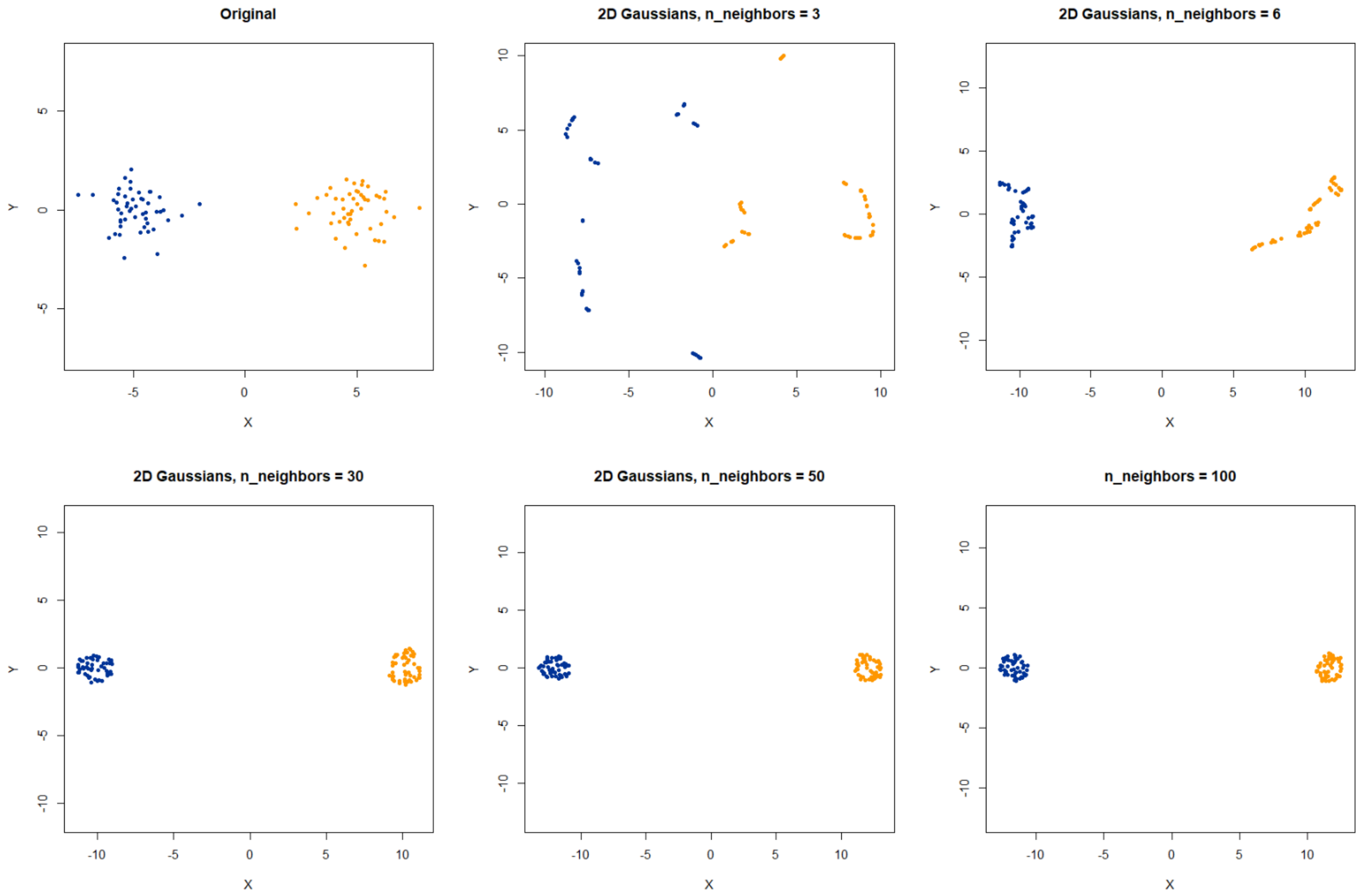
UMAP
7 minutes



UMAP Parameters

- Number of neighbors
- Number of iterations
- Minimun Distance (low-D)
- Metrics

...



https://jlmelville.github.io/uwot/umap-simple.html

Summary: UMAP

- NON-LINEAR method of dimensionality reduction
- Very efficient to compute
- Can be run from the top PCs (e.g.: PC1 to PC10)
- Is not randomly initialized and allows
- It should preserve global structure

Problems:

- It is designed to group cells stronger than t-SNE to show meaningful larger distances
- Similar number of hyper-parameters as t-SNE

www.cytosplore.org
graphics.tudelft.nl
@thomashollt

Acknowledgements:
STW / NWO Grant 12720 VAnPIRe
LKEB, IHB, LCBC @ LUMC
CGV @ TU Delft