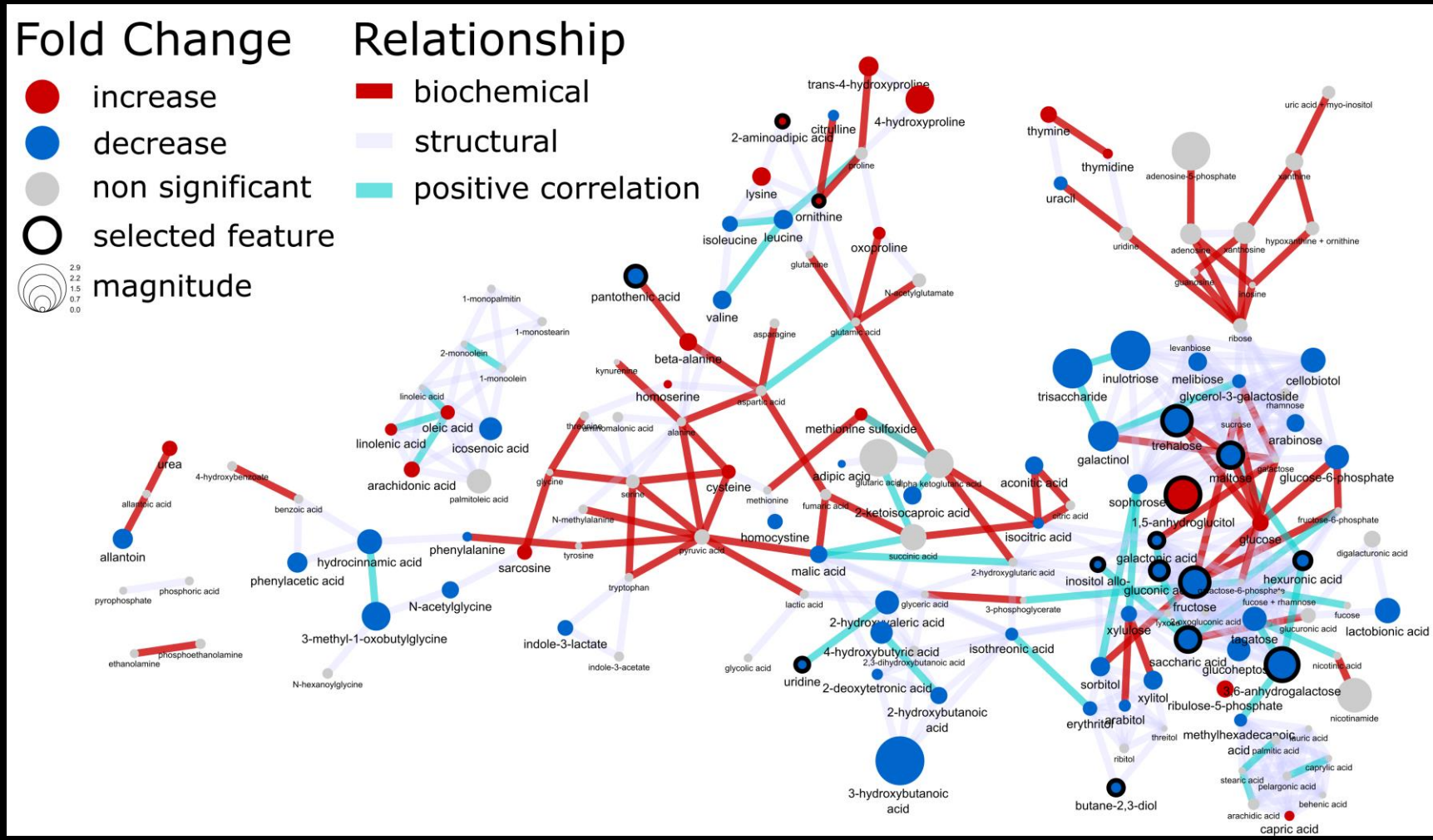
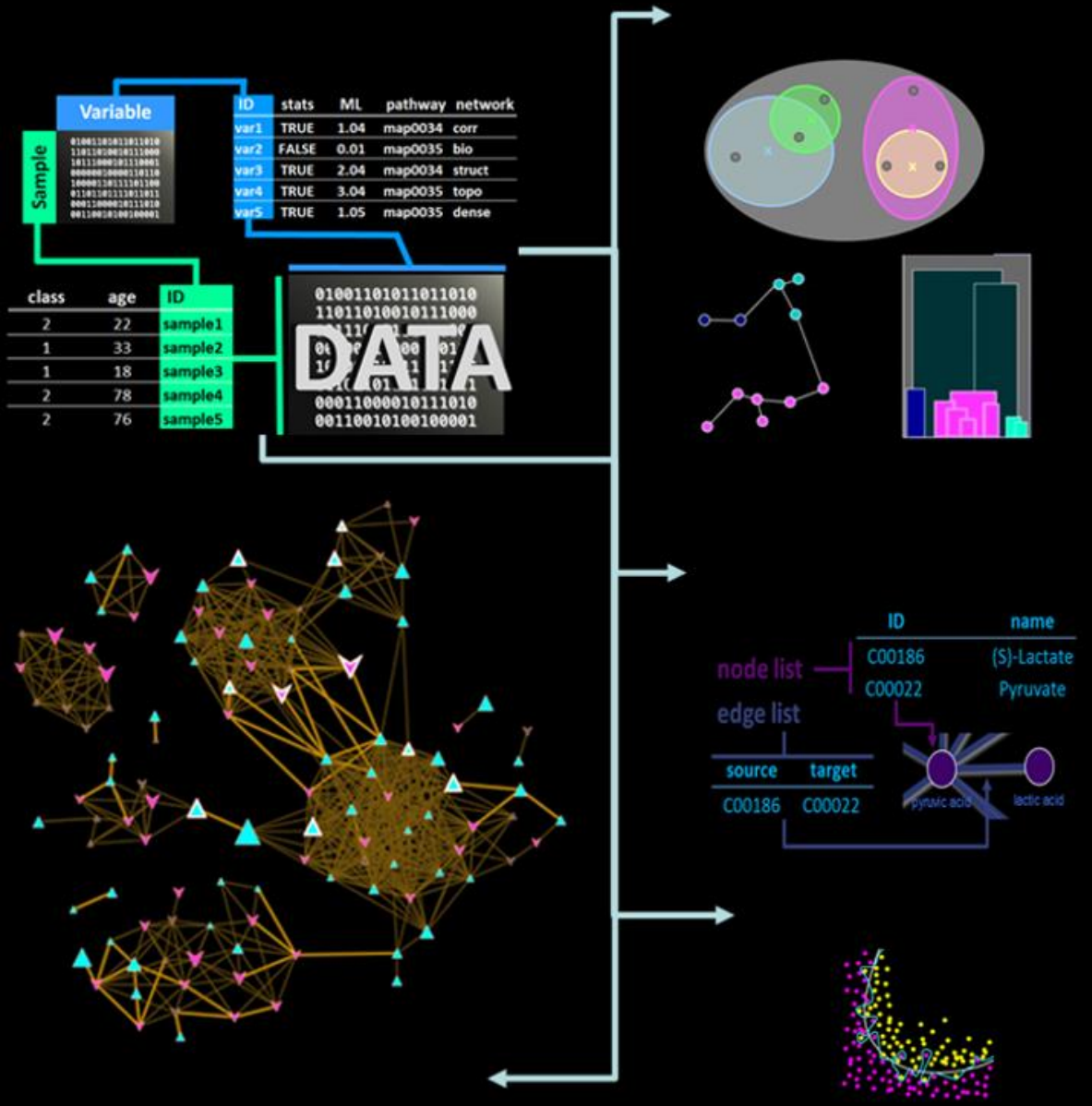
Welcome to network mapping 101

In the following course you will learn how to integrate statistical, multivariate and machine learning results within a publication quality biochemical network.

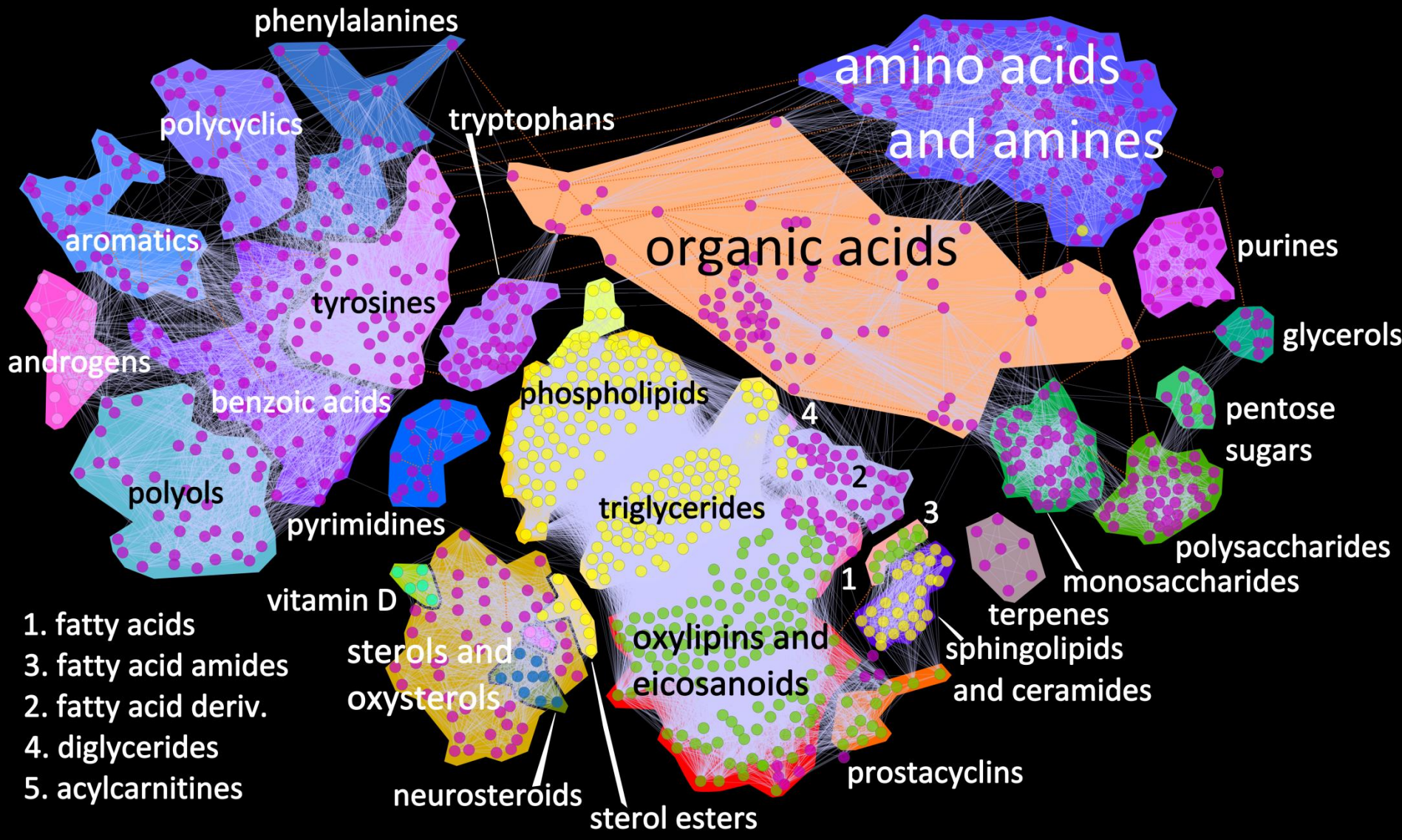


Tutorials

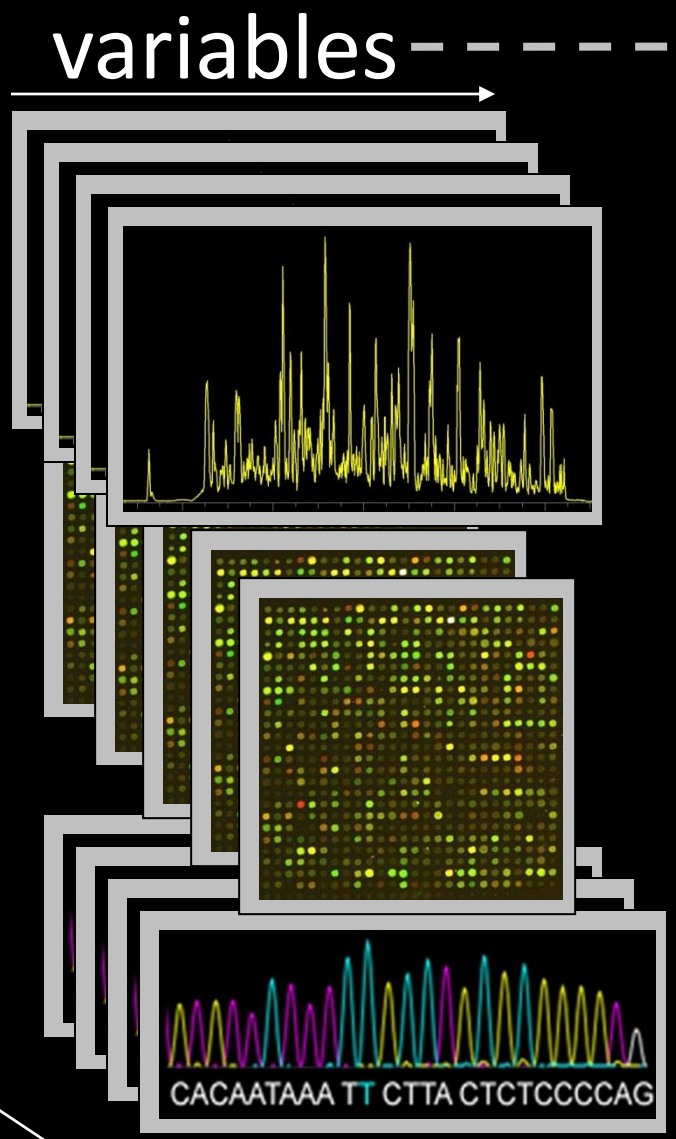
- Preparing raw data for analysis
- Statistical analysis
- Multivariate data exploration
- Supervised clustering
- Machine learning
  - classification
  - model validation
  - feature selection
- Network analysis
  - biochemical
  - structural similarity
  - correlation
- Network mapping – putting it all together

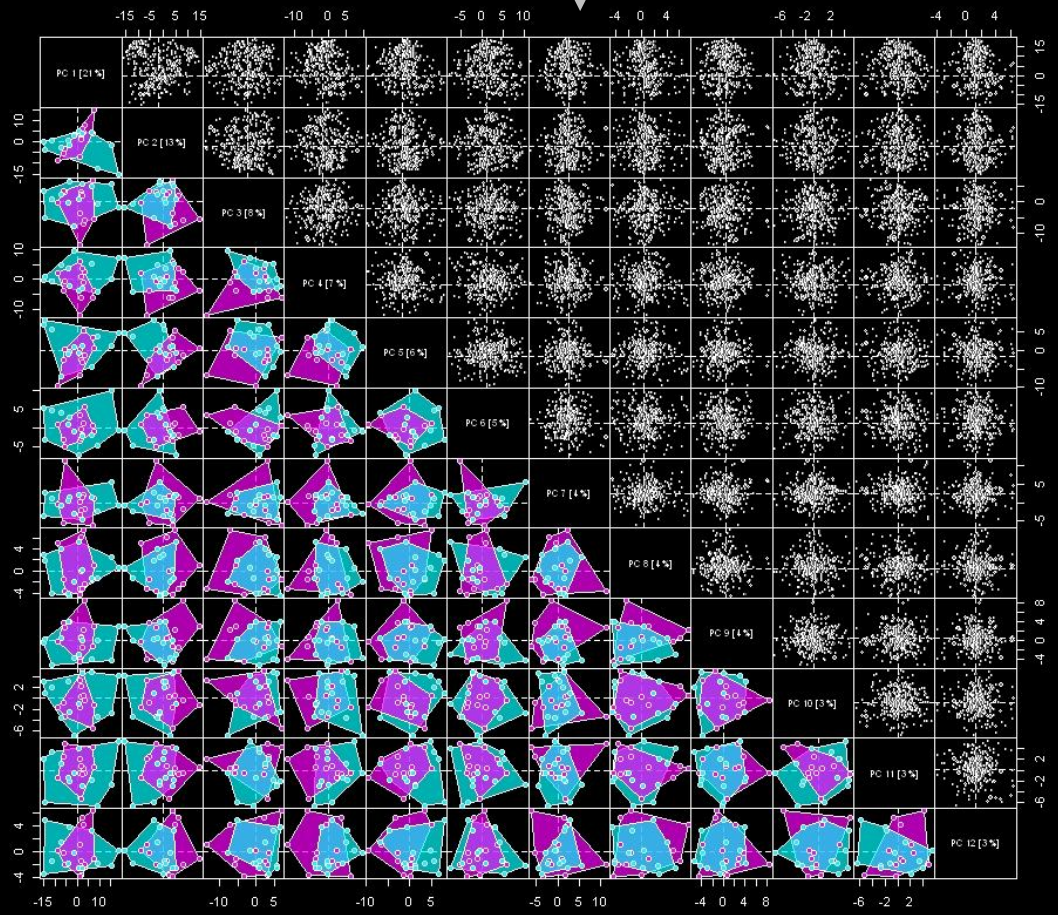


Analysis at the metabolomic scale



Integrate high-dimensional data





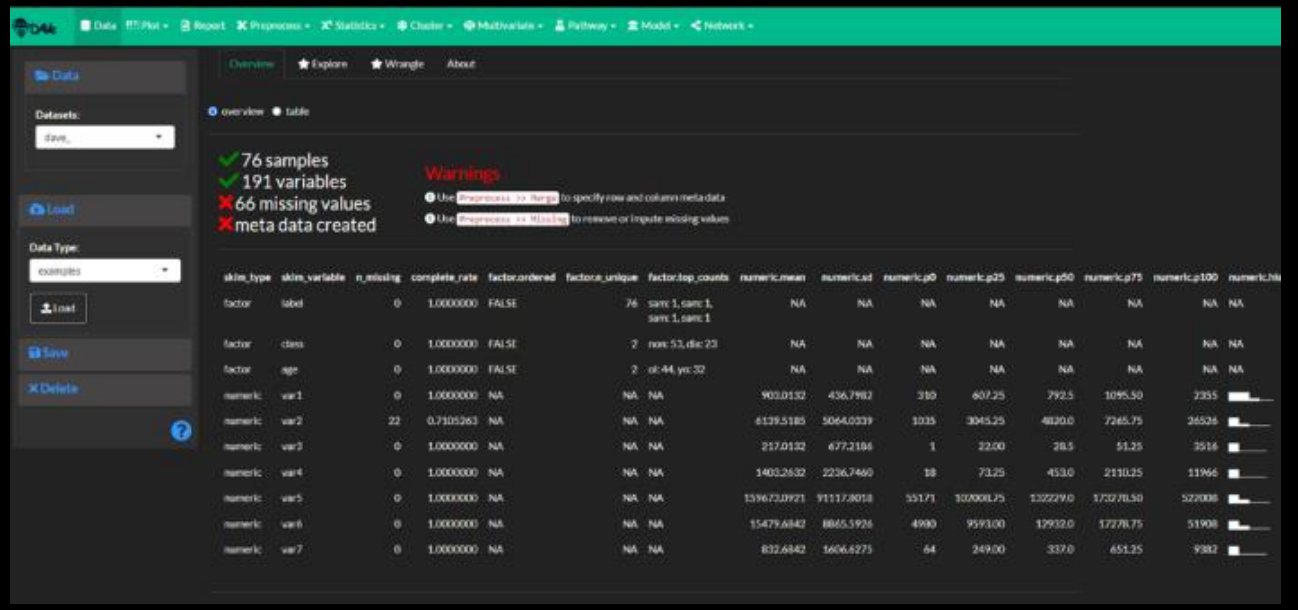
Identify what matters

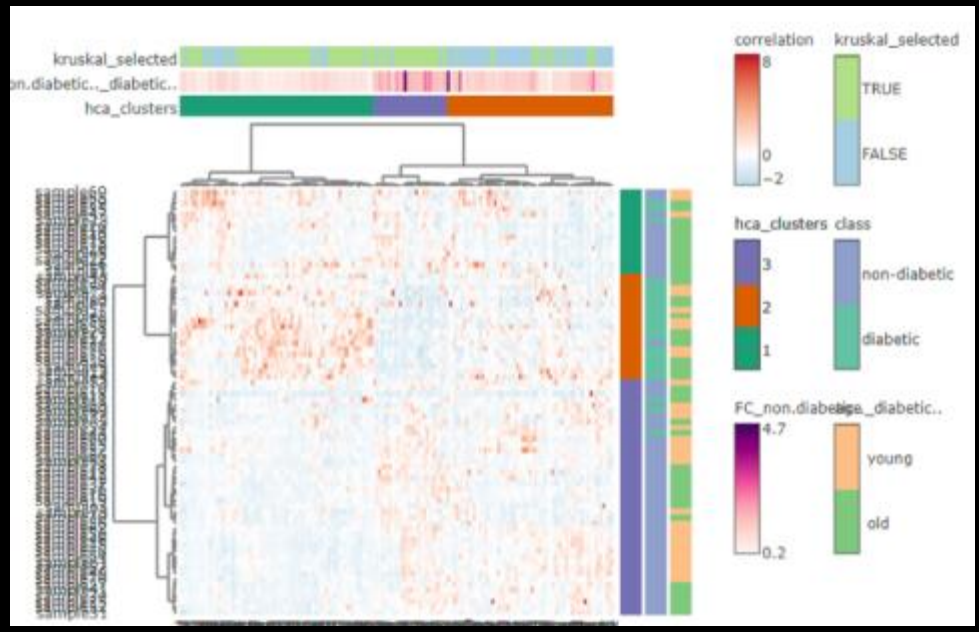
CDS Creative Data Solutions

	Univariate	Multivariate	Predictive Modeling
Group 1			
Group 2			
	ANOVA	PCA	PLS

Topics

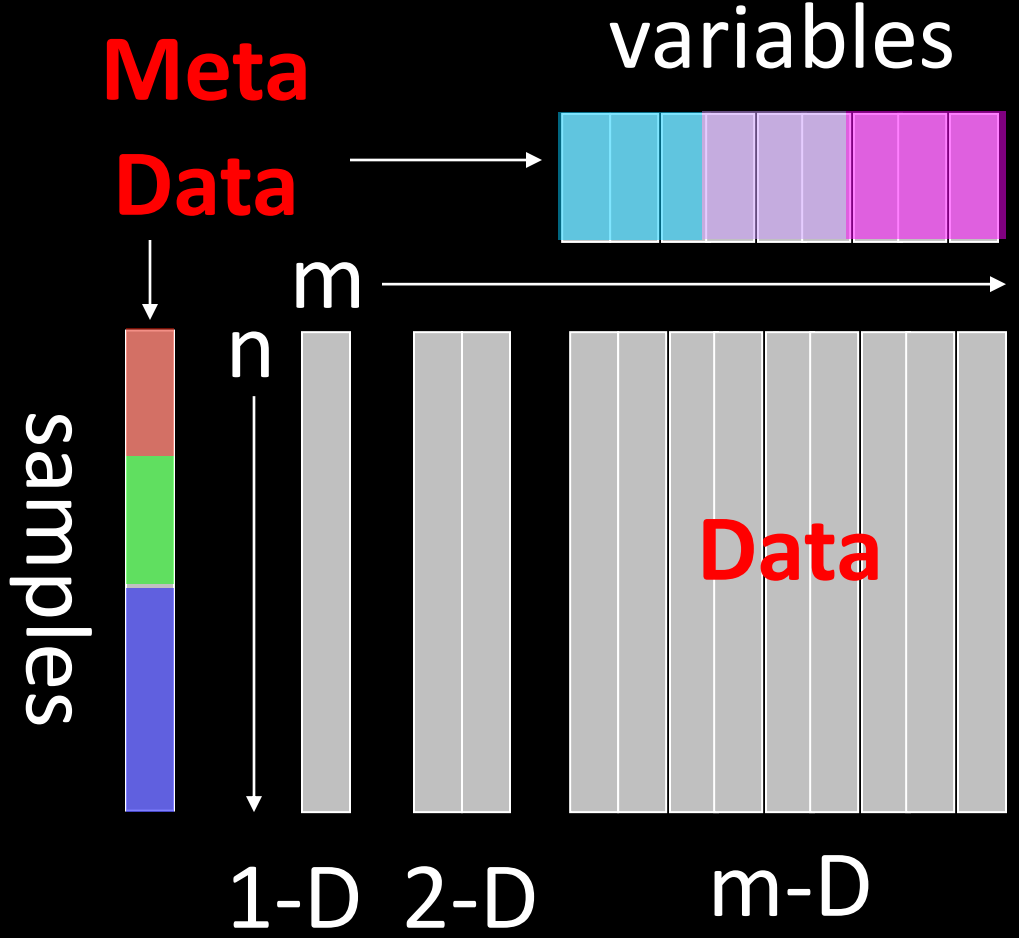
- Data preparation
- Differential expression
- Hierarchical Clustering
- Principal Components Analysis (PCA)
- Statistical analysis
- Machine learning
- Network analysis
- Network mapping



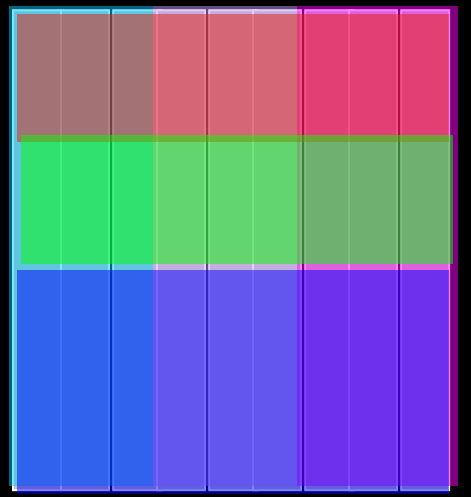


https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/#topics

How to think about data complexity



Experimental Design = complexity

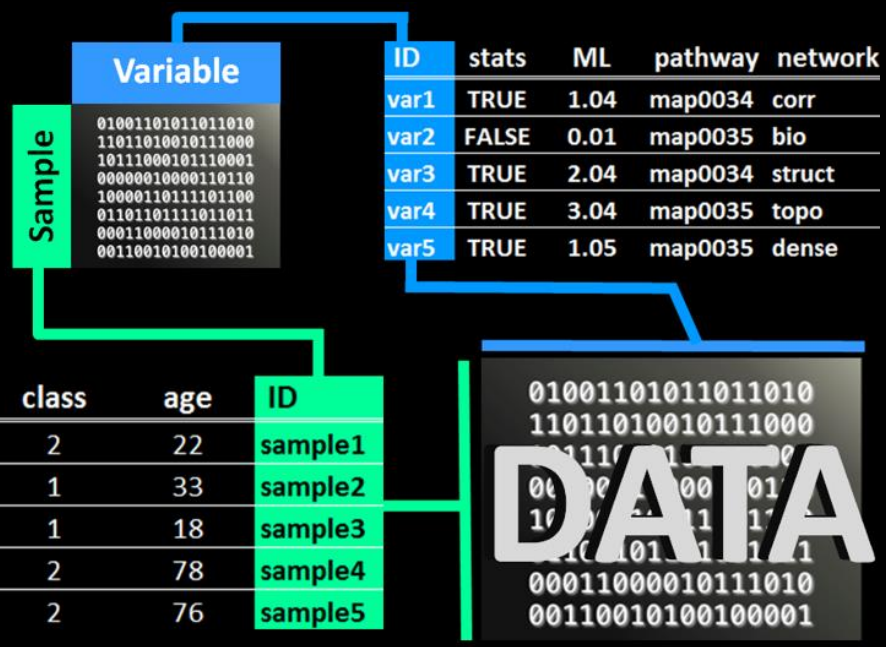


Variable # = dimensionality

Data preprocessing

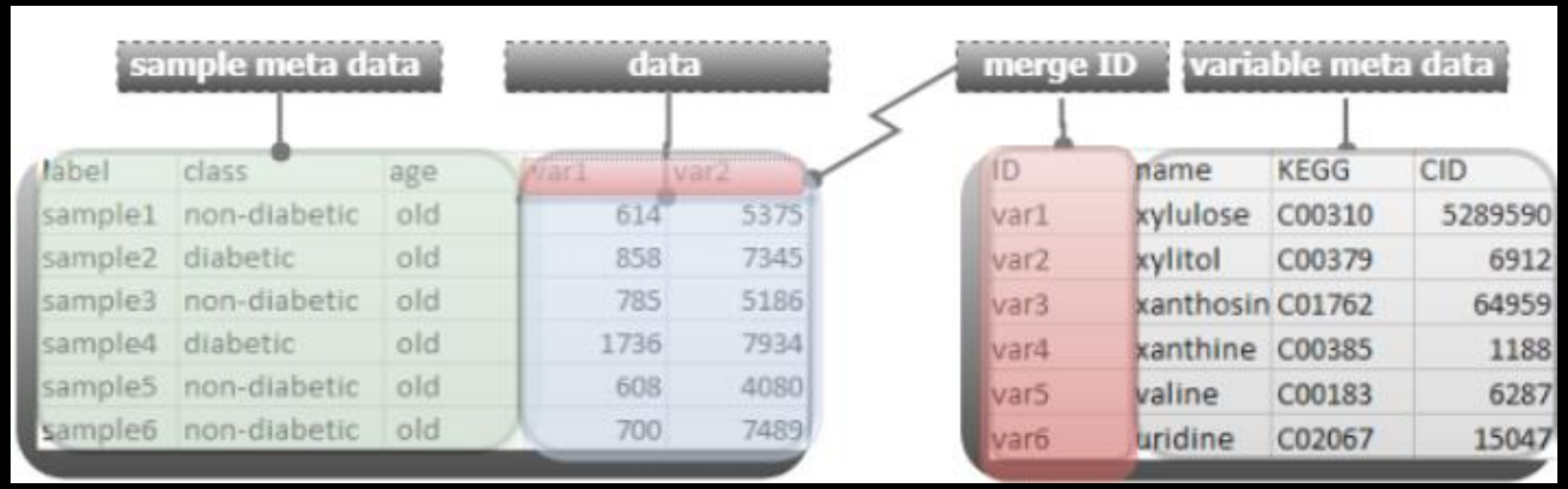
define

- data
- row meta data
- column meta data



remove and/or impute

- missing values



Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/preprocess/

Differential expression

compare

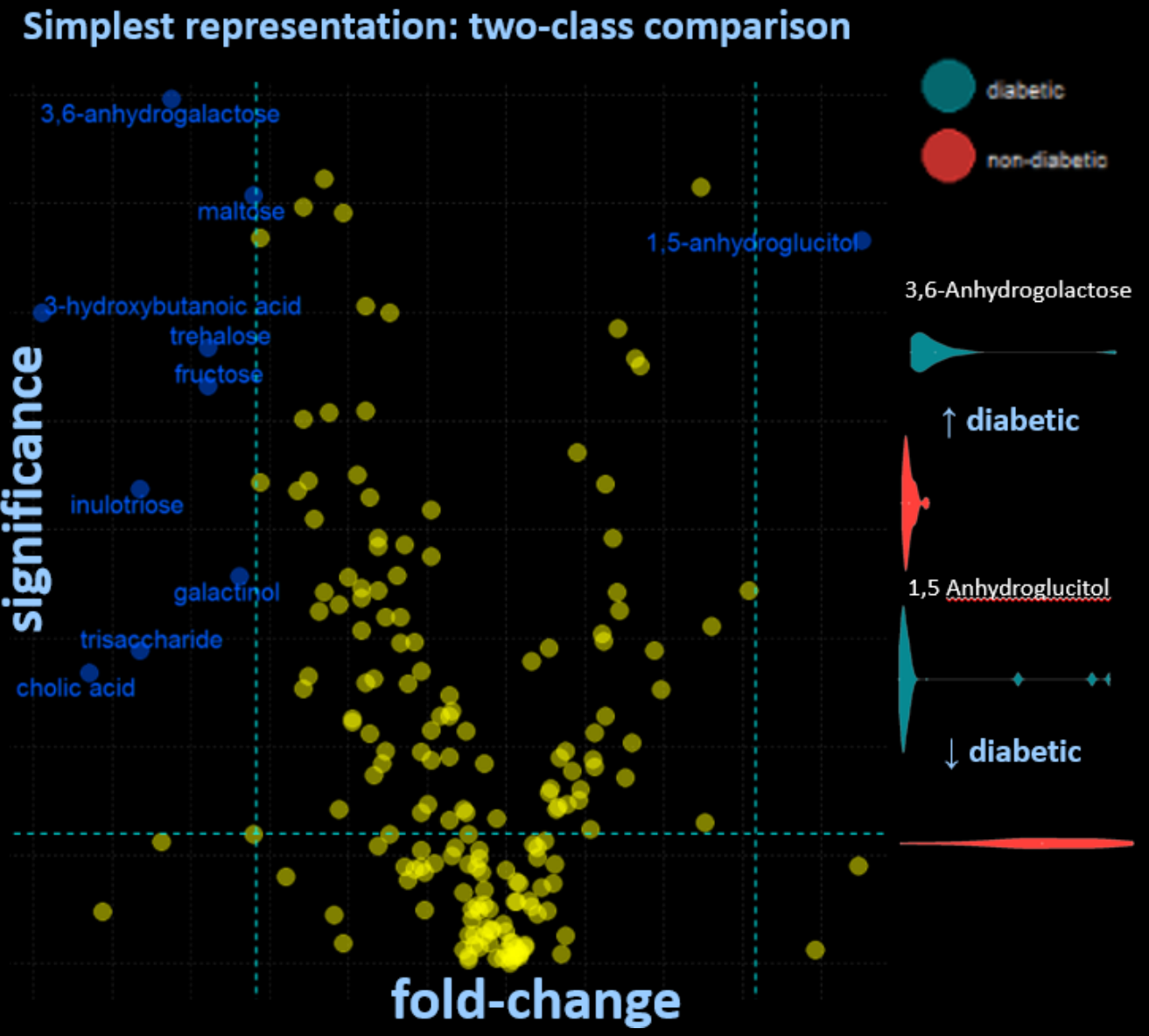
- class means

identify

- significant differences

visualize

- volcano plots
- violin plots



https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/statistics/

Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/statistics/

Hierarchical clustering

group

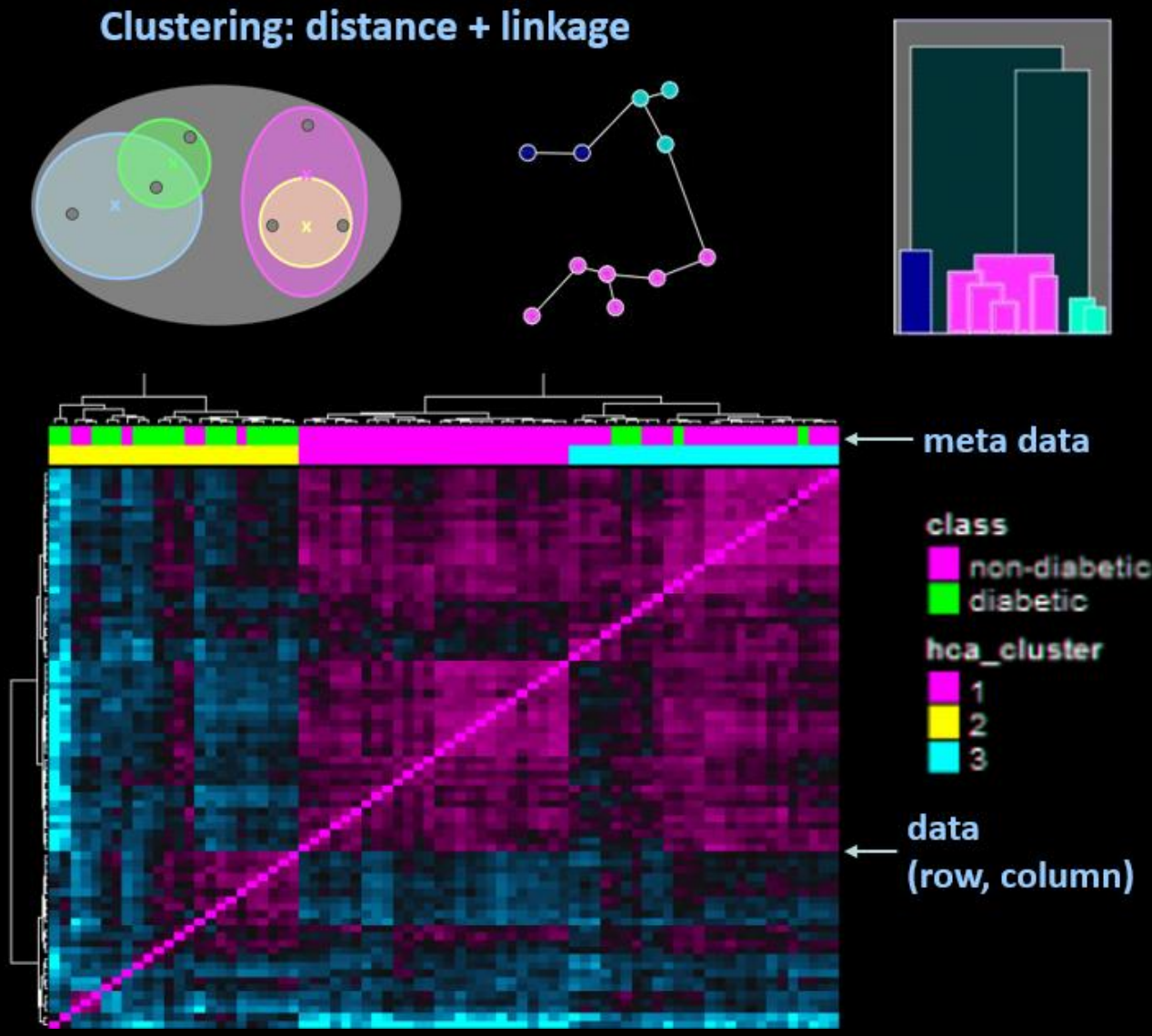
- samples and/or variables

define similarity

- correlation
- distance
- linkage

visualize

- heatmaps
- dendrograms



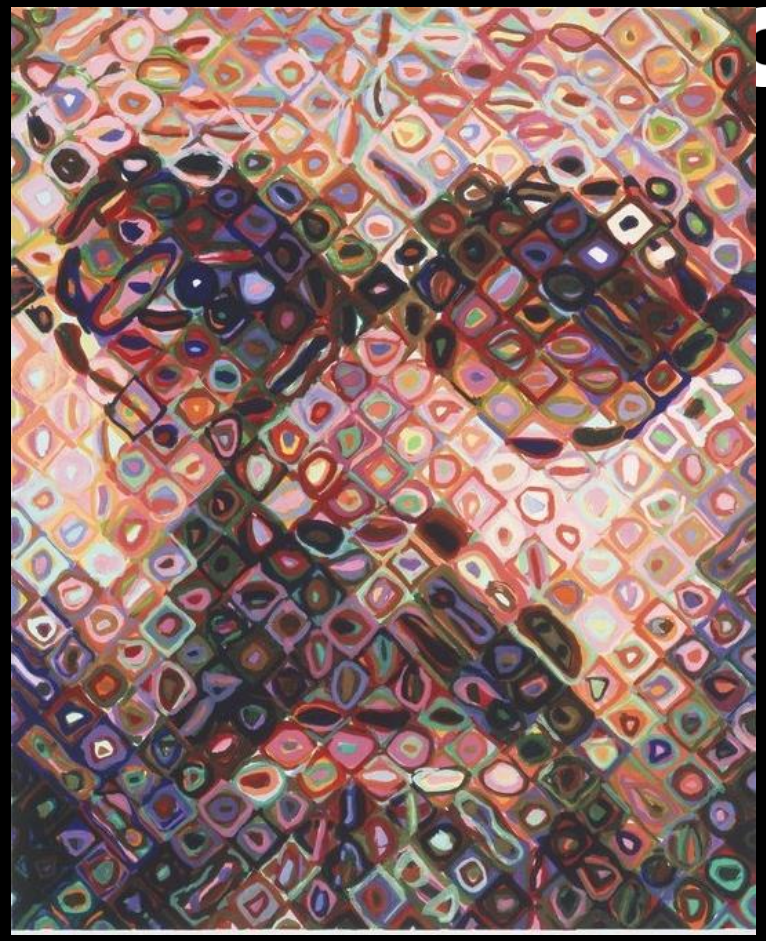
https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/clustering/#heirarchical-clustering

Clustering

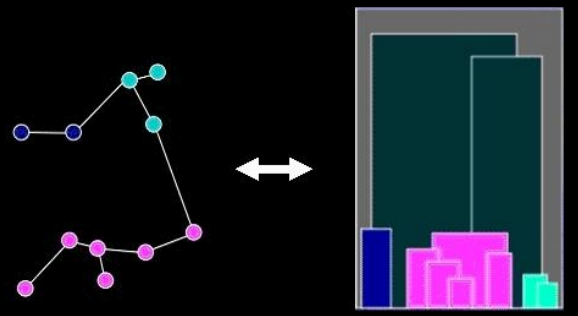
Identify

- patterns
- group structure
- relationships

- Evaluate/refine hypothesis
- Reduce complexity



Artist: Chuck Close



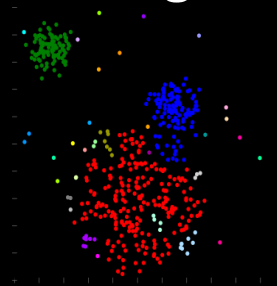
Cluster Analysis

CDS Creative Data Solutions

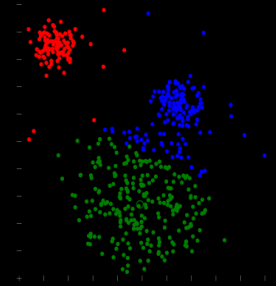
Use the concept similarity/dissimilarity to group a collection of samples or variables

Approaches
- **hierarchical (HCA)**
- **non-hierarchical (k-NN, k-means)**
- **distribution (mixtures models)**
- **density (DBSCAN)**
- **self organizing maps (SOM)**

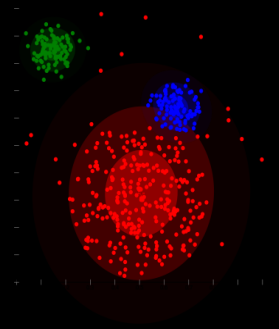
Linkage



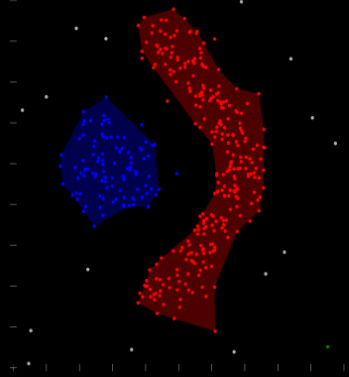
k-means



Distribution

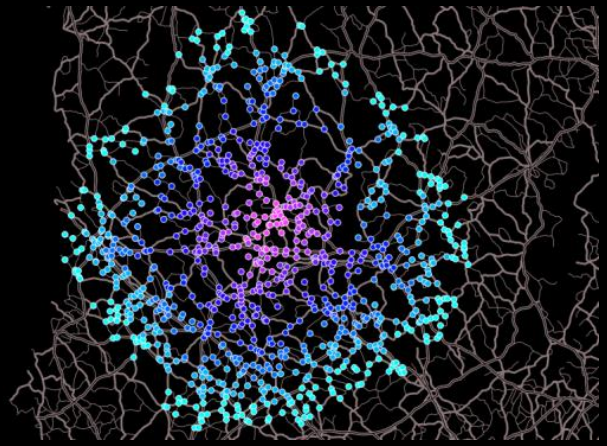


Density

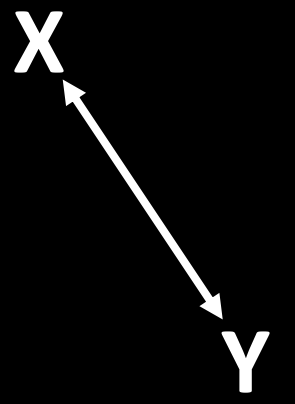


Hierarchical Cluster Analysis

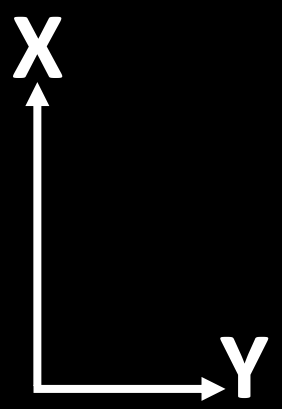
- similarity/dissimilarity defines “nearness” or distance



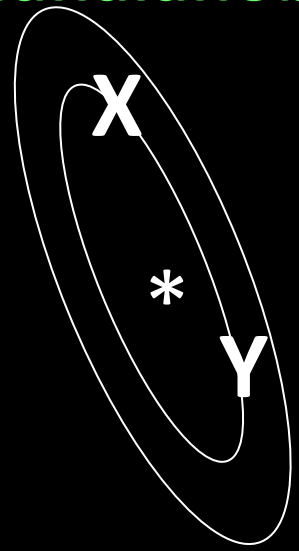
euclidean



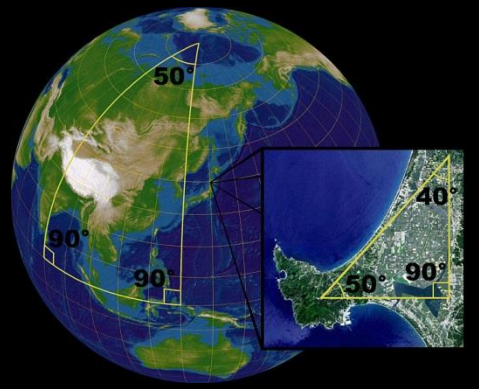
manhattan



Mahalanobis

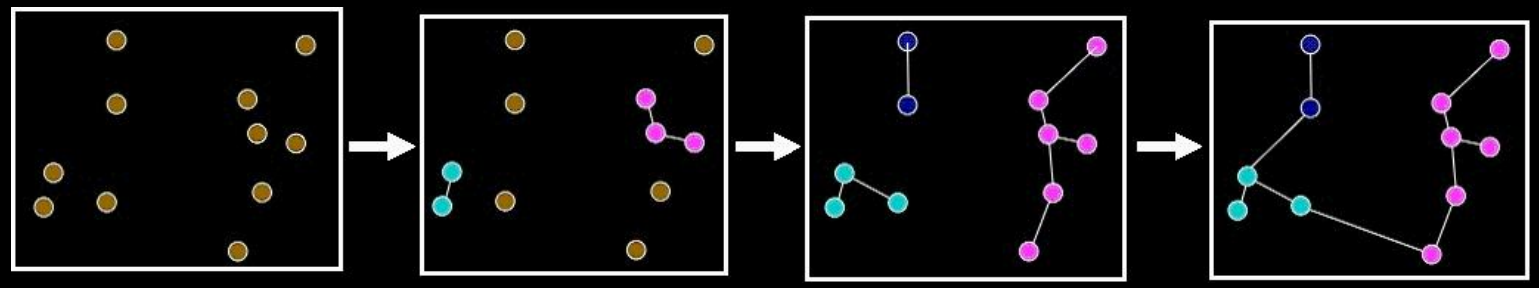


non-euclidean

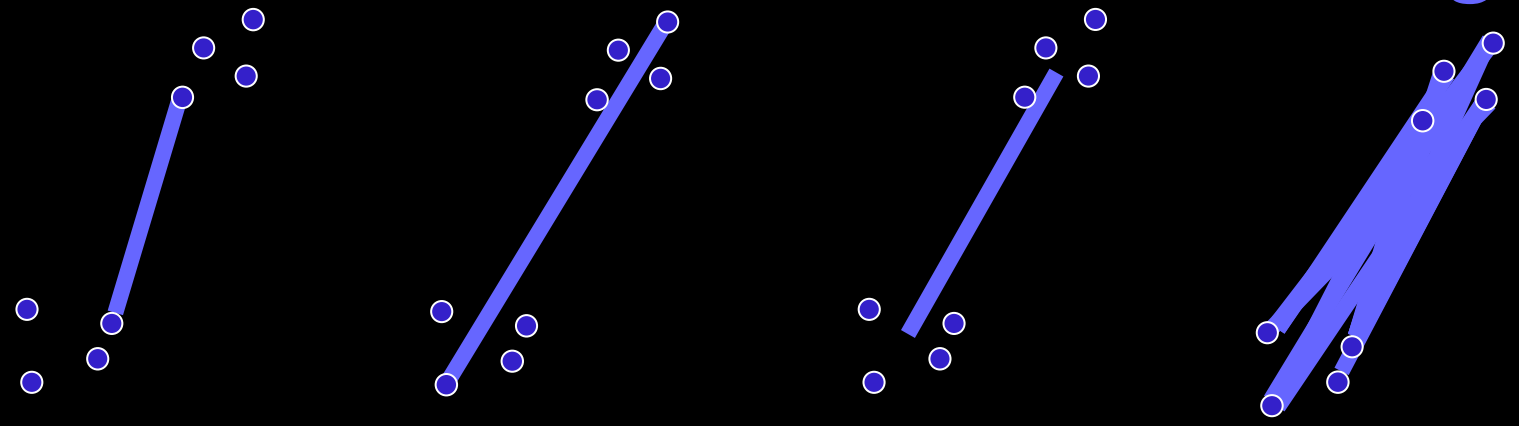


Hierarchical Cluster Analysis

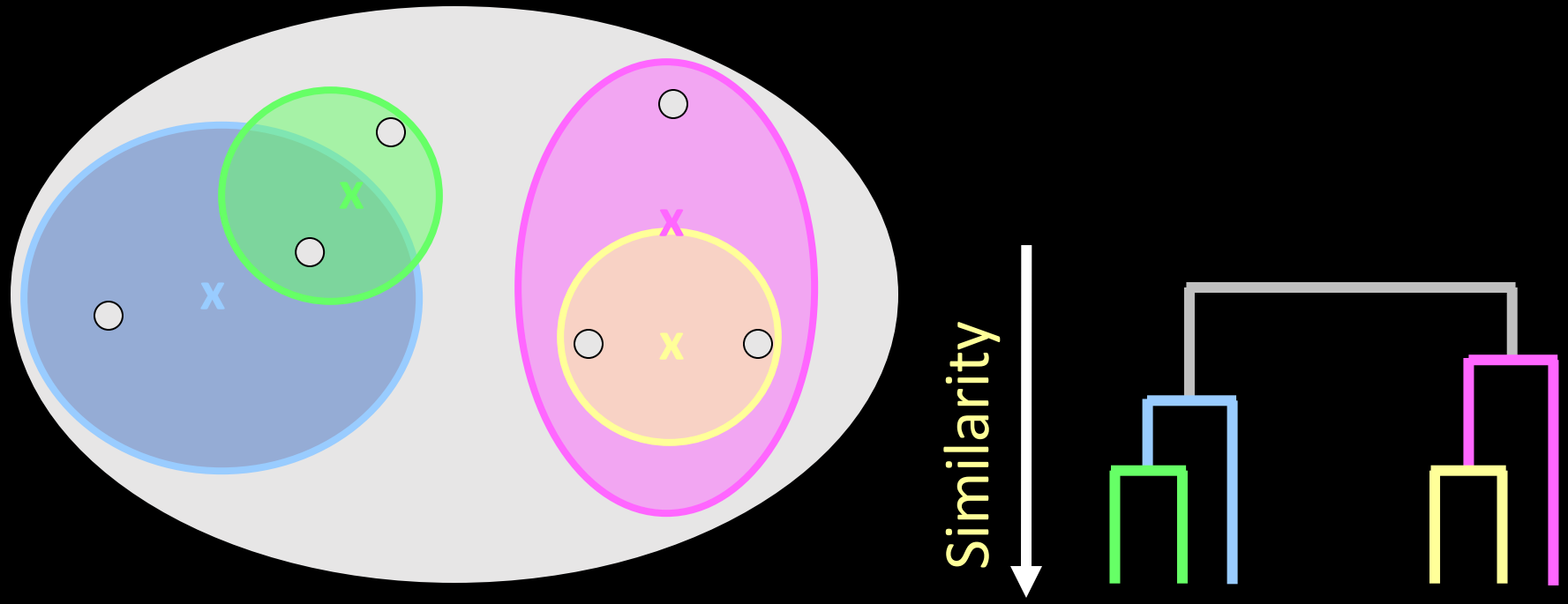
Agglomerative/linkage algorithm defines how points are grouped



single complete centroid average



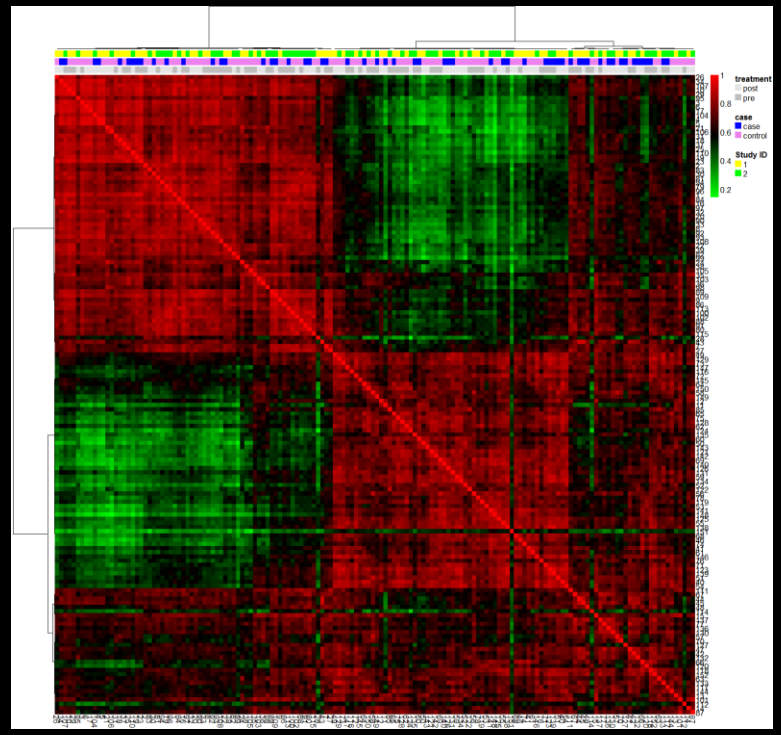
Hierarchical Cluster Analysis (cont.)



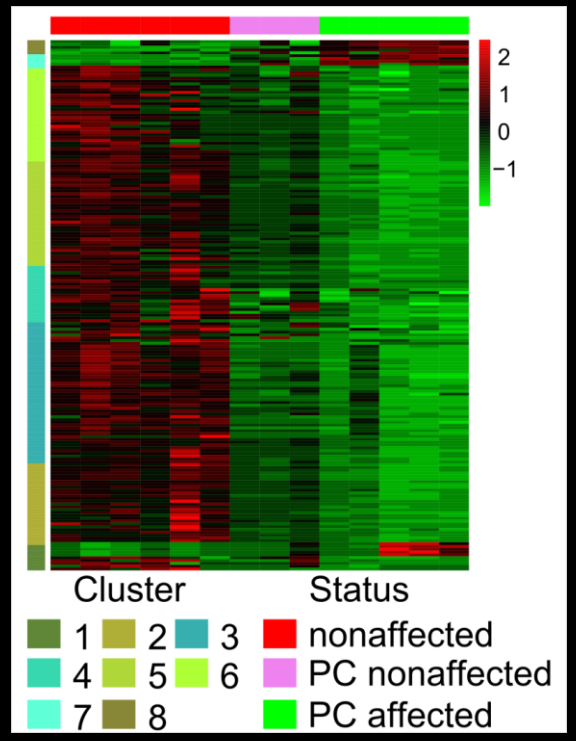
Hierarchical Cluster Analysis

How does my metadata match my data structure?

Overview



Confirmation



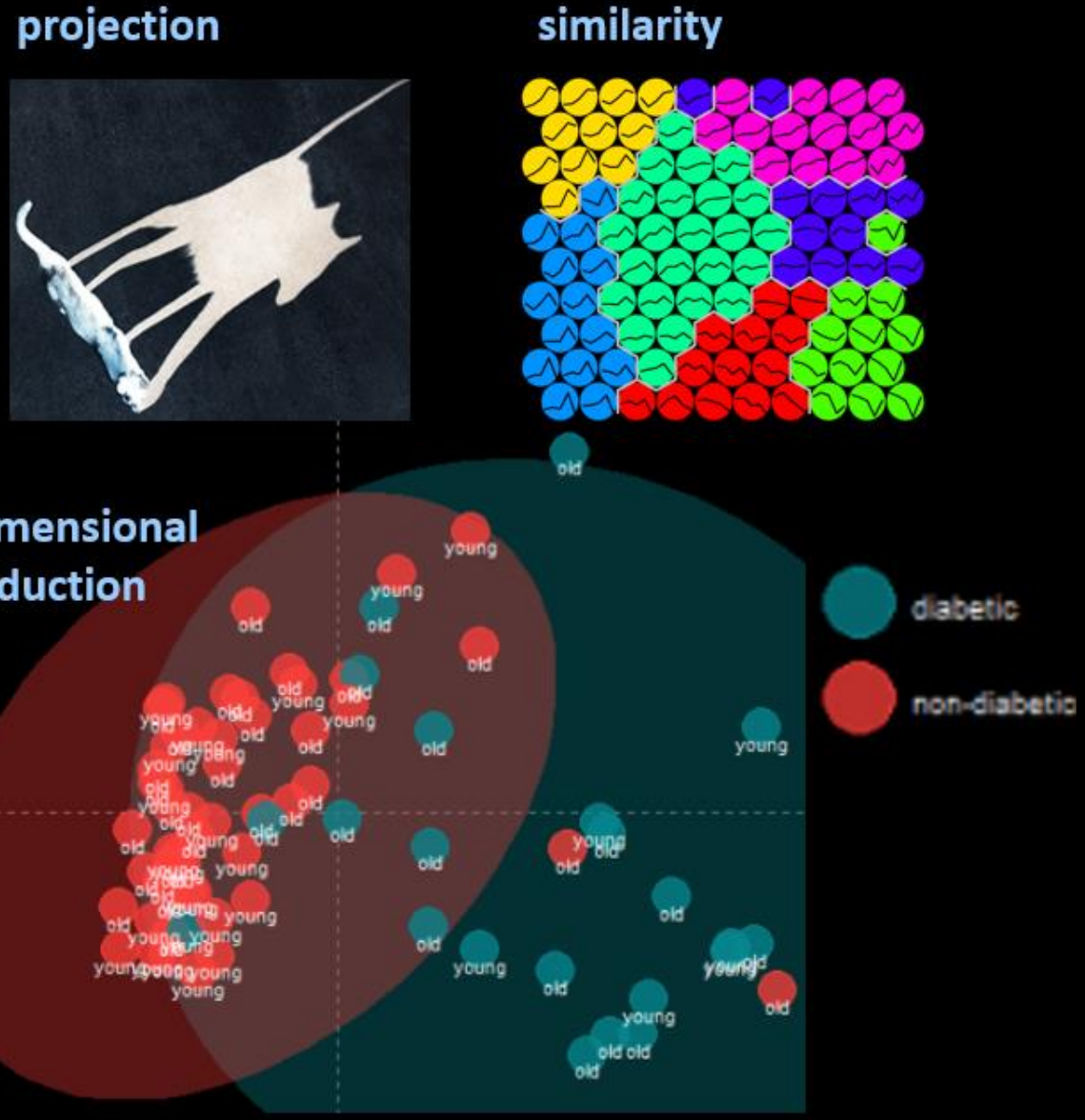
Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/clustering/#heirarchical-clustering

Principal Components Analysis (PCA)



reduce

- dimensionality

maximize

- variance explained

visualize

- variance explained
- outliers
- sample scores
- variable loadings

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/multivariate/#pca

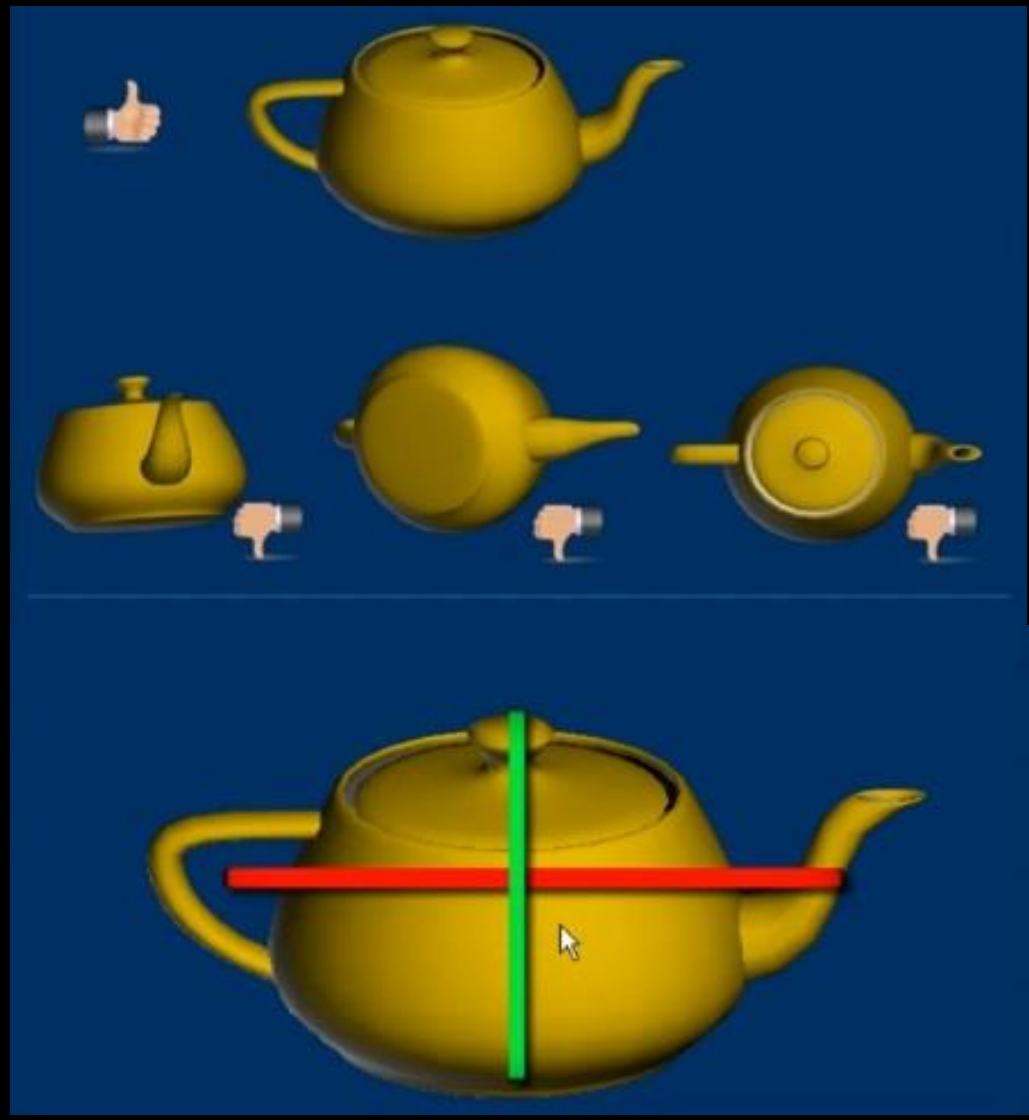
PCA goals

Principal Components (PCs)

- non-supervised
- projection of the data which maximize variance explained

results

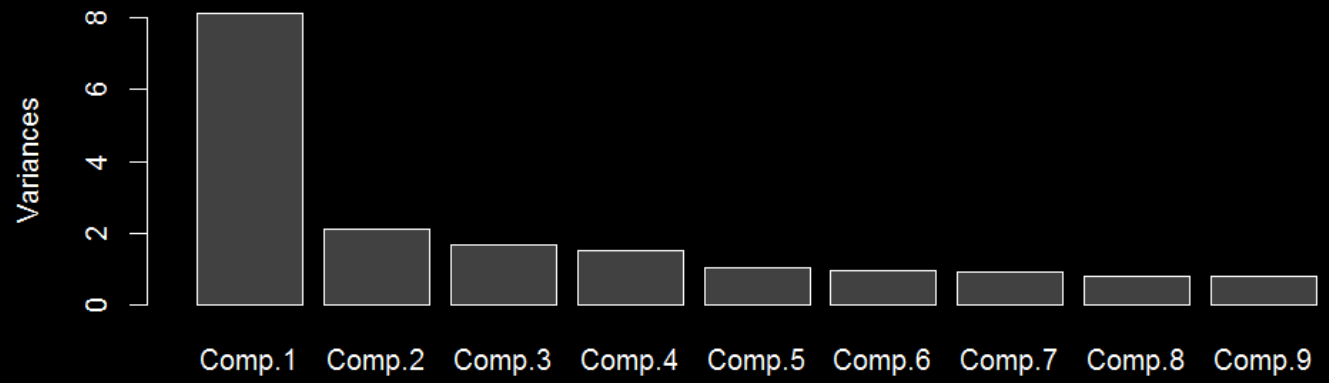
1. eigenvalues = variance explained
2. scores = new coordinates for samples (rows)
3. loadings = linear combination of original variables



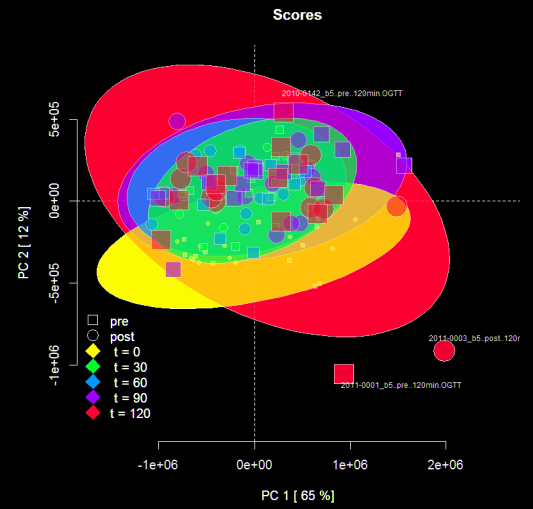
James X. Li, 2009, VisuMap Tech.

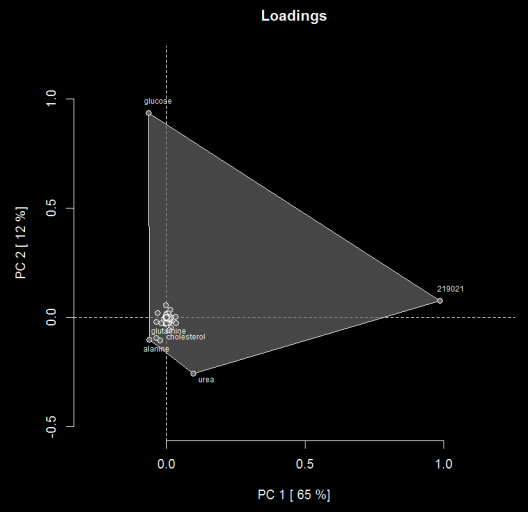
PCA interpretation

Variance explained (eigenvalues)

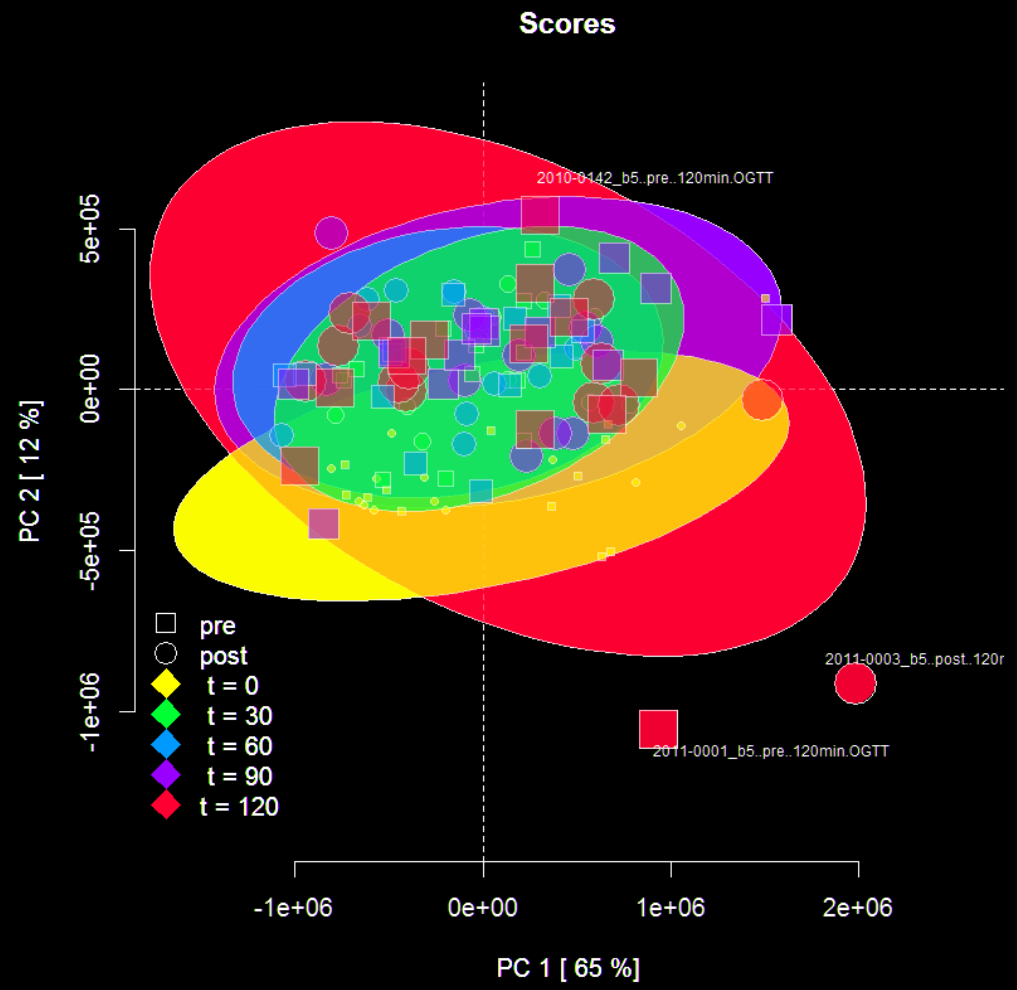


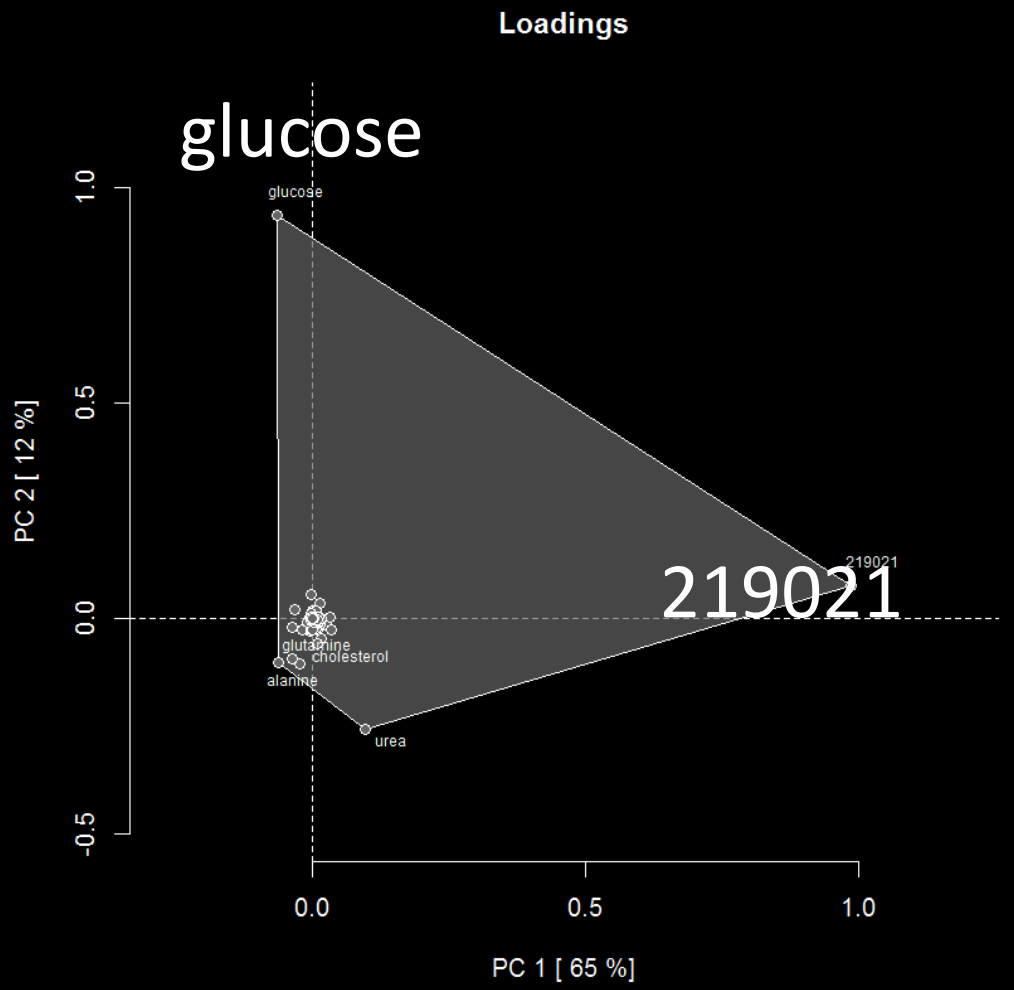
Row (sample) scores and column (variable) loadings





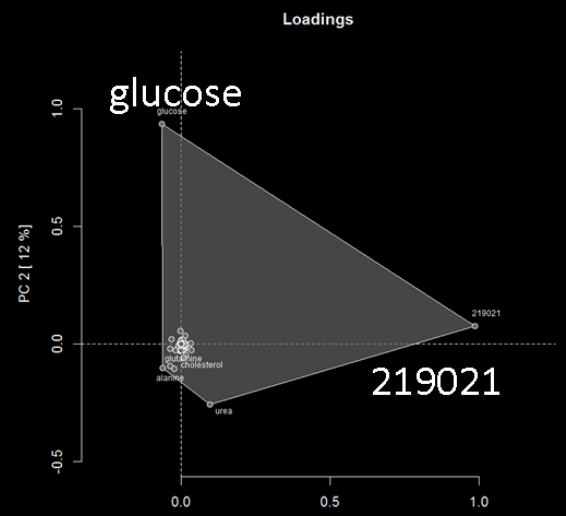
PCA example





*no scaling or centering

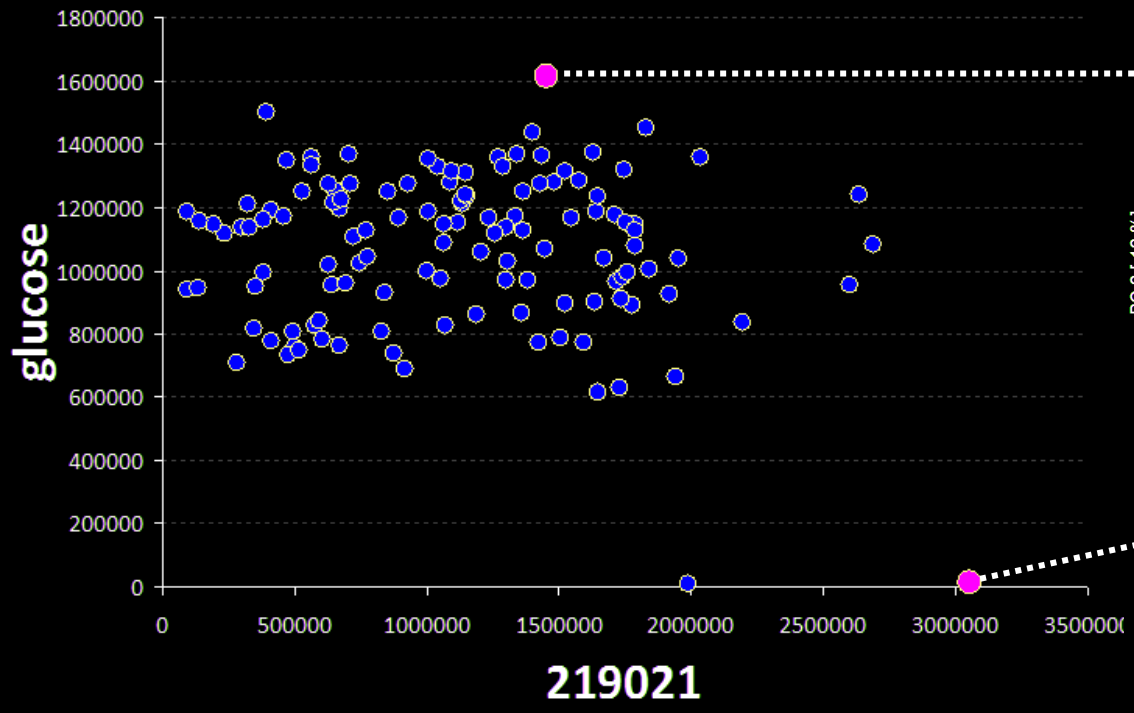
Relationship between scores and loadings

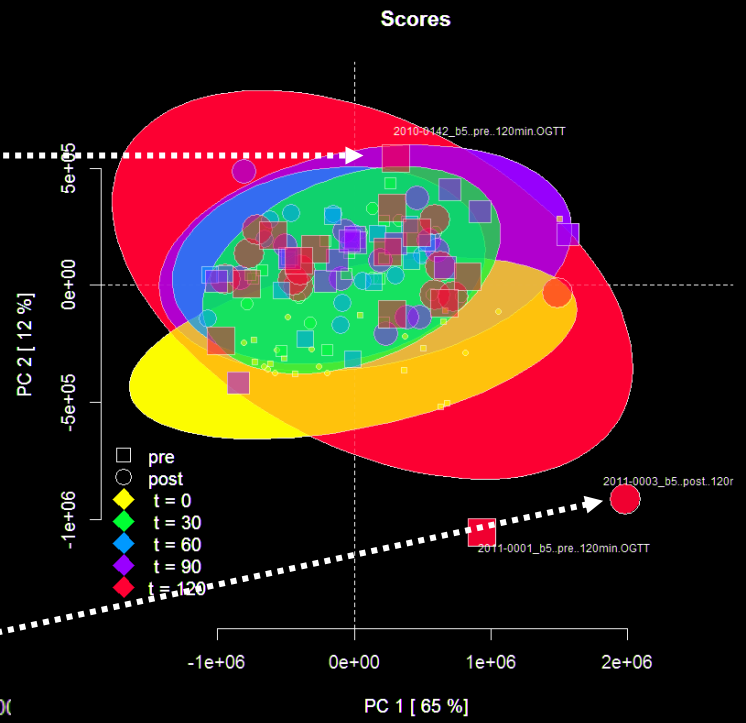


loadings

219021

top loading variable's scatterplot





scores

Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/multivariate/#pca

Machine learning

predict

- sample classification

optimize

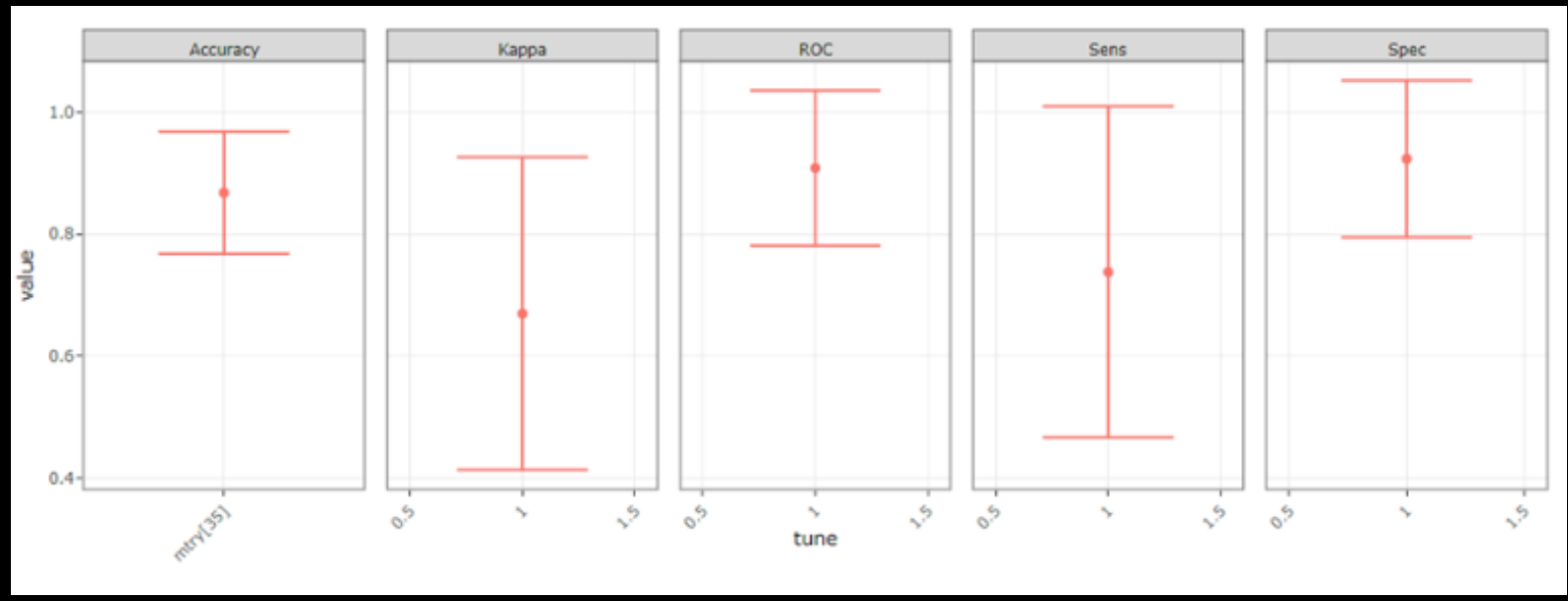
- model performance

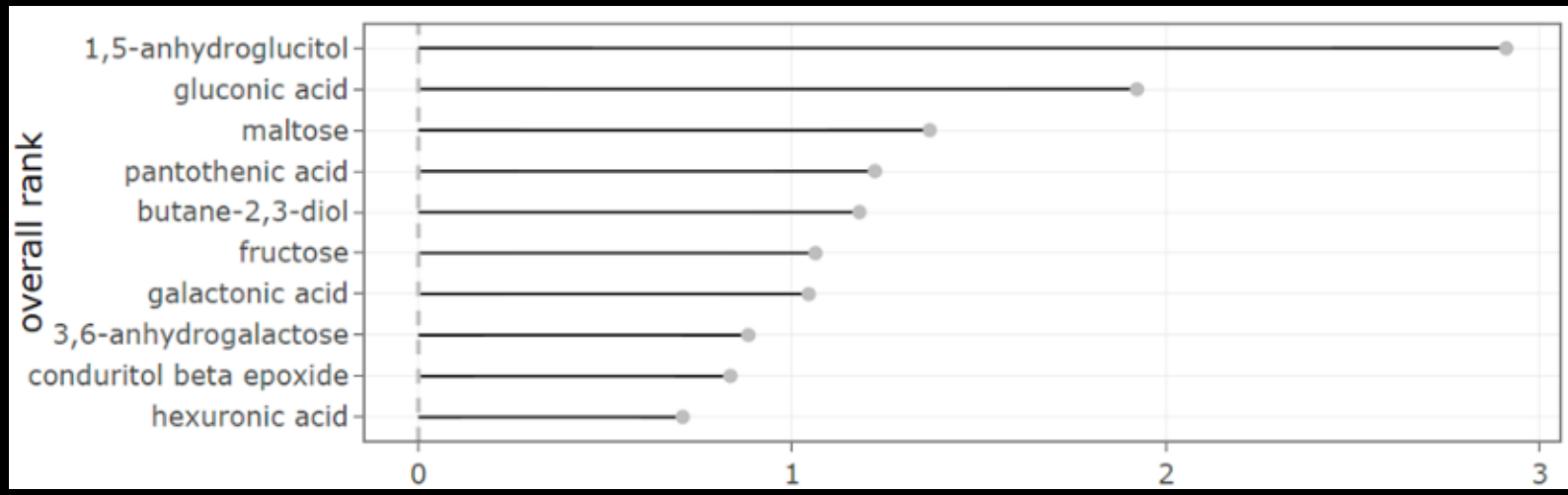
select

- important features

visualize

- model performance
- feature importance





https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/model/

Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/model/

Network analysis

network mapping

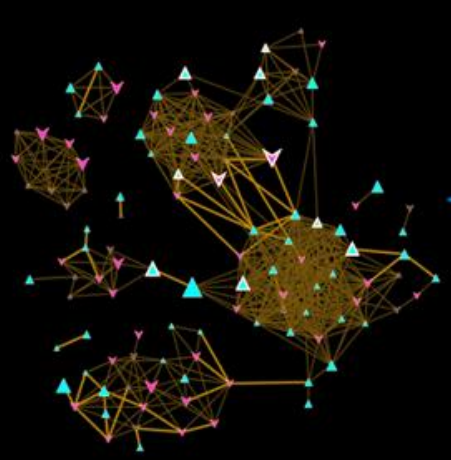
- transform variables

network calculation

- regularized correlation
- biochemical
- structural similarity
- model performance

visualize

- interactive networks

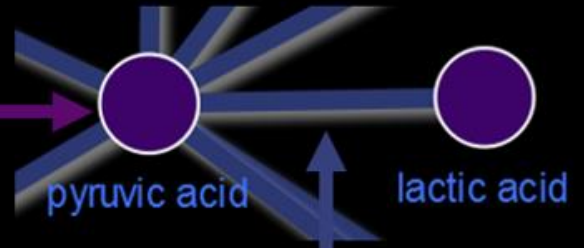


node list

ID	name	stats	diff	rank
C00186	(S)-Lactate	0.005	up	1
C00022	Pyruvate	0.100	down	2

edge list

source	target	type	weight	dir
C00186	C00022	KEGG	1	NA
C00186	C00022	corr	1	+



https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/network/

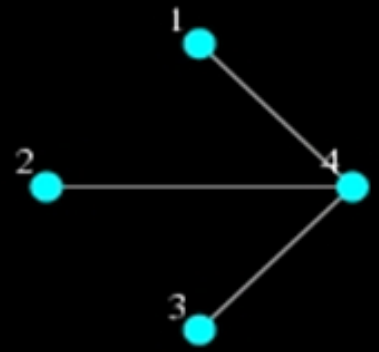
Components for network mapping

connections (edges)

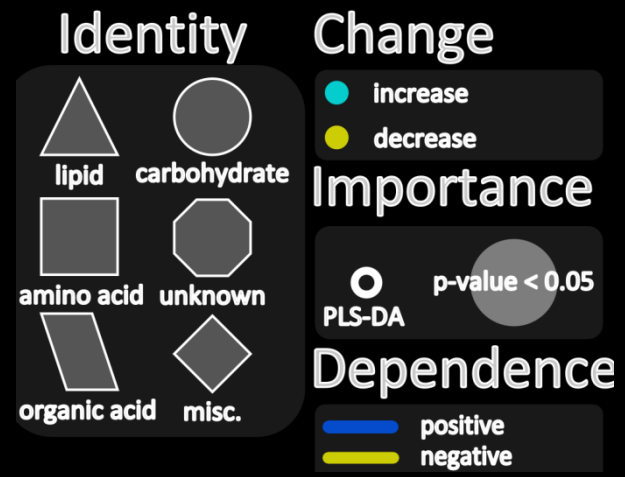
- empirical dependency (correlation)
- biochemical (substrate/product)
- chemical similarity
- …

nodes (vertices)

- magnitude
- importance
- direction
- relationships
- …



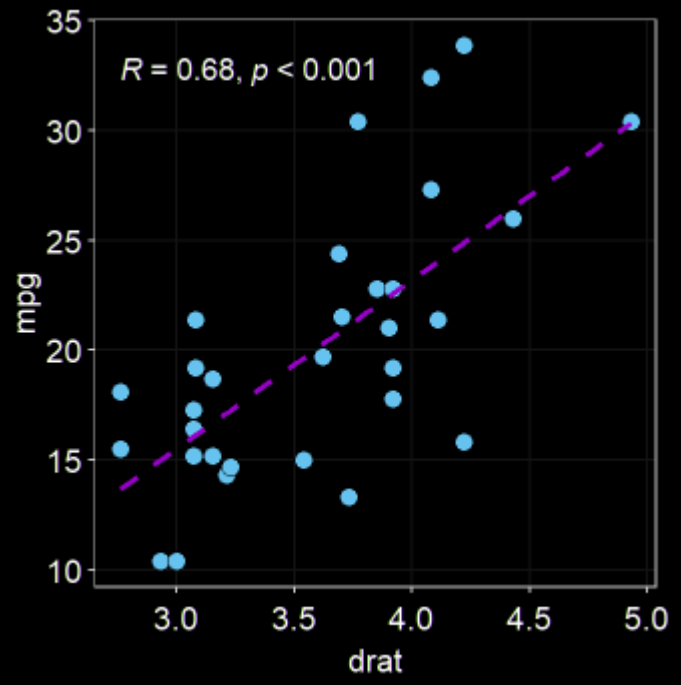
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$



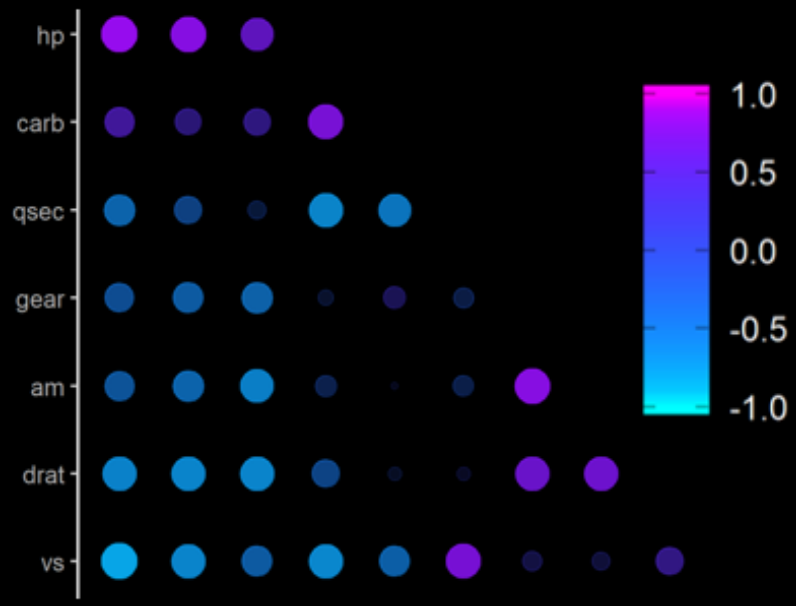
Correlation networks

Connect molecules based on strength of their correlation or partial-correlation

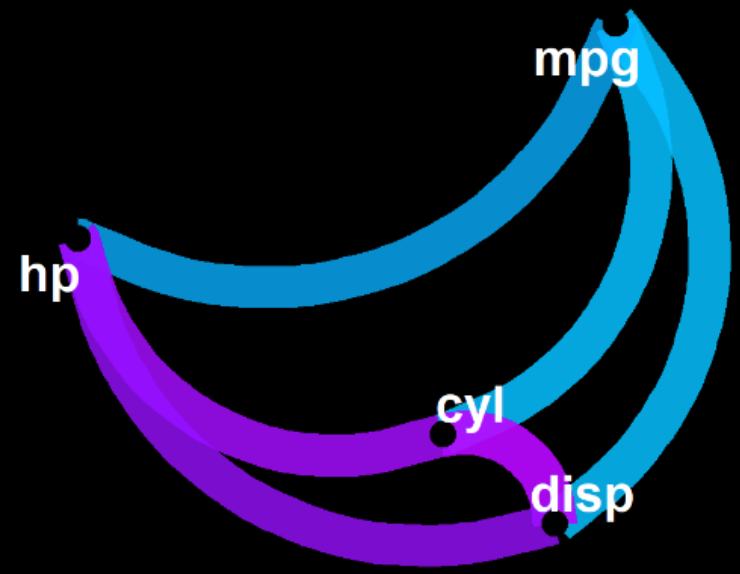
bivariate



multivariate

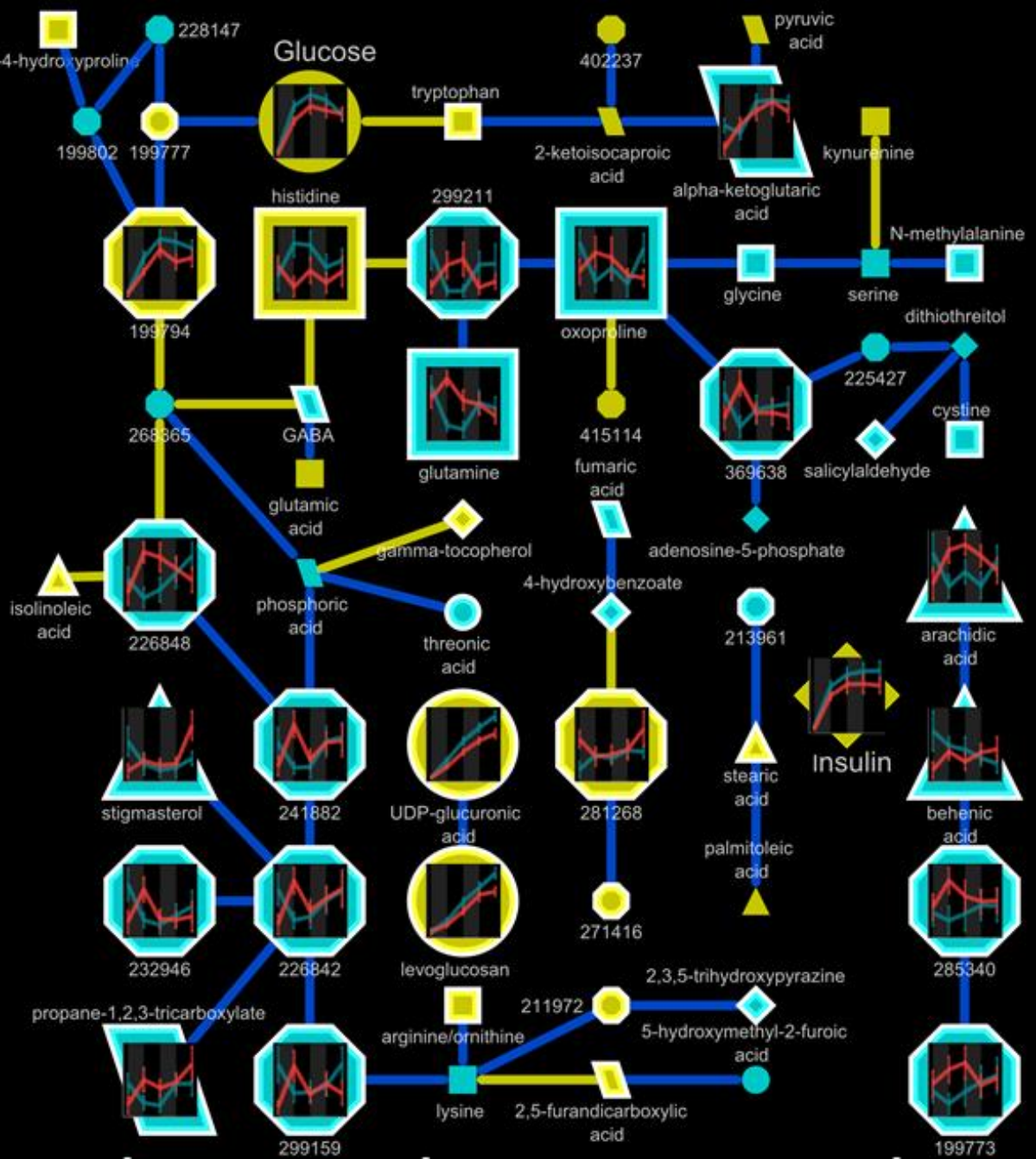


network

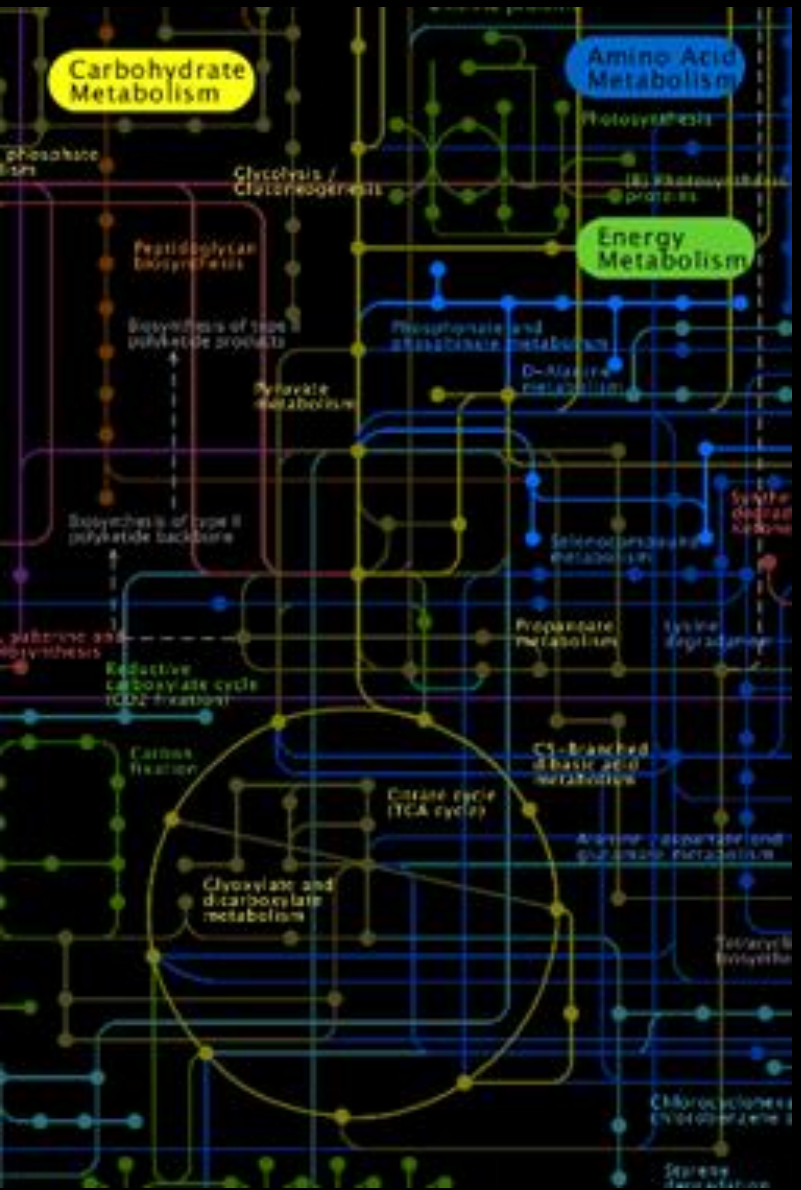


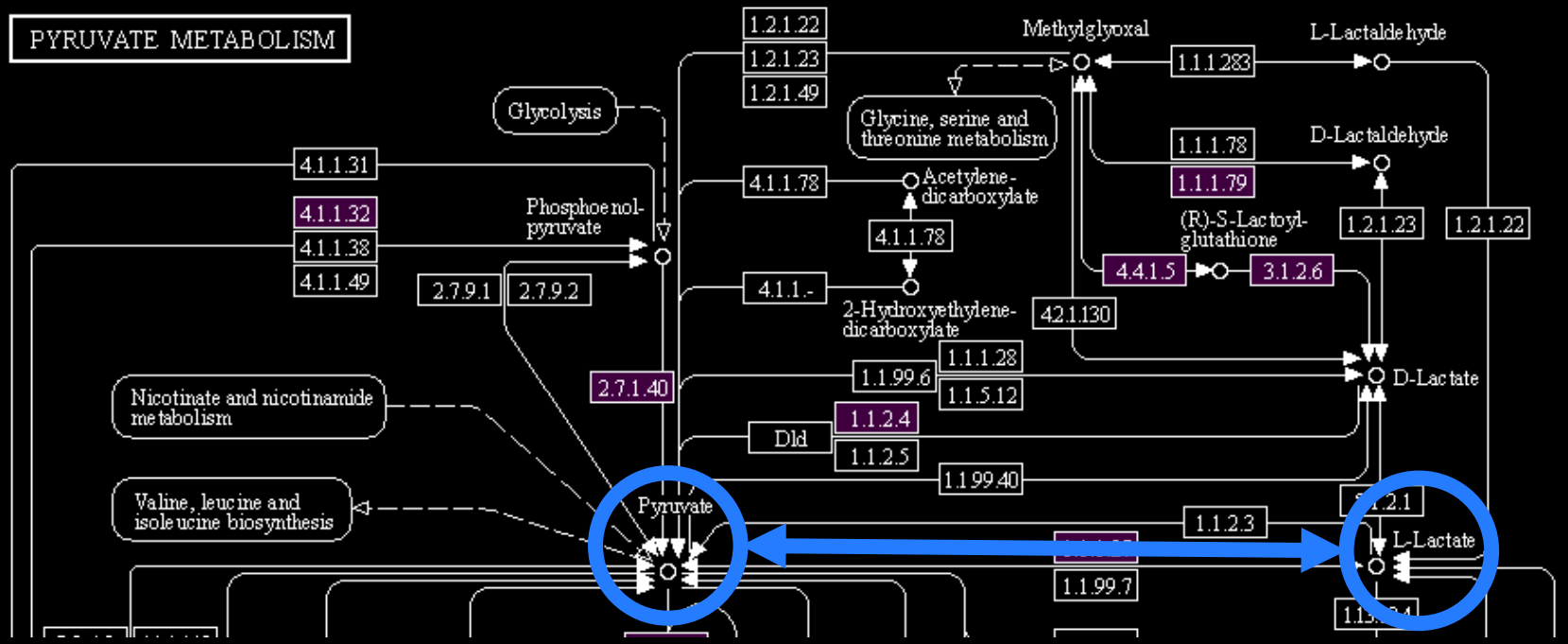
Correlation example

regularized correlation network showing relationships in metabolic timeseries measurements for two classes of samples



Biochemical networks





nodes

ID	name	stats	diff	rank
C00186	(S)-Lactate	0.005	up	1
C00022	Pyruvate	0.100	down	2

edges

source	target	type	weight	dir
C00186	C00022	KEGG	1	NA

Structural similarity networks

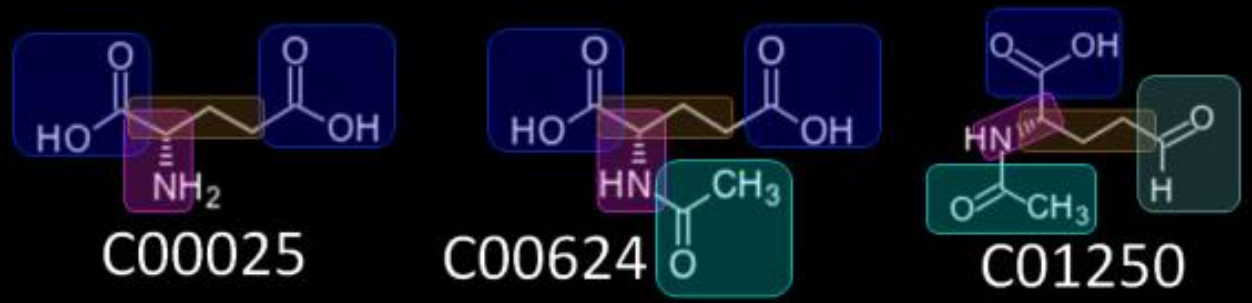
- Use structure to generate molecular fingerprint
- Calculate similarities between metabolites based on fingerprint
- PubChem service for similarity calculations

http://pubchem.ncbi.nlm.nih.gov//score_matrix/score_matrix.cgi

- online tools

http://uranus.fiehnlab.ucdavis.edu:8080/MetaMapp/homePage

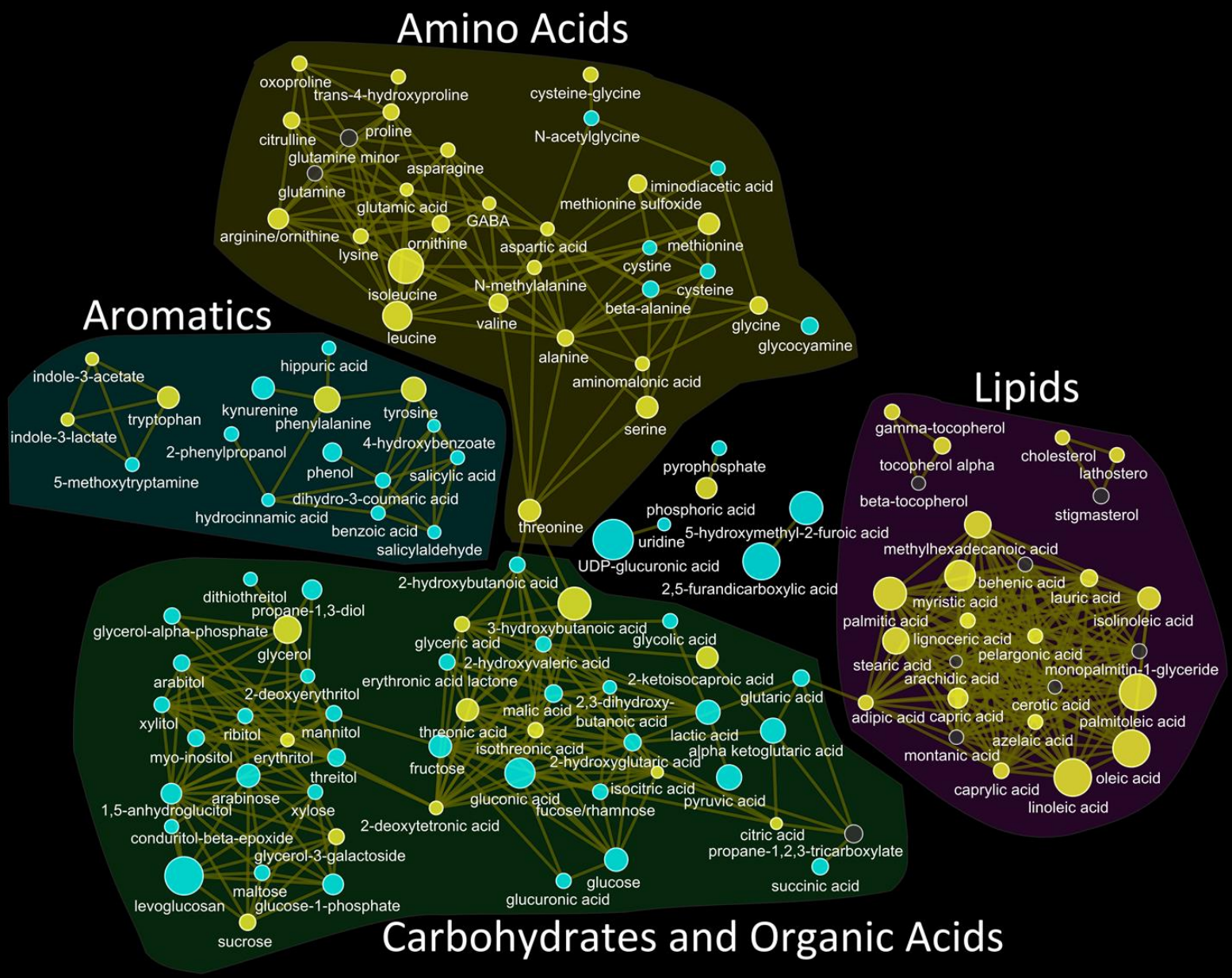
Chemical mapping
of substructure comparison
using PubChem



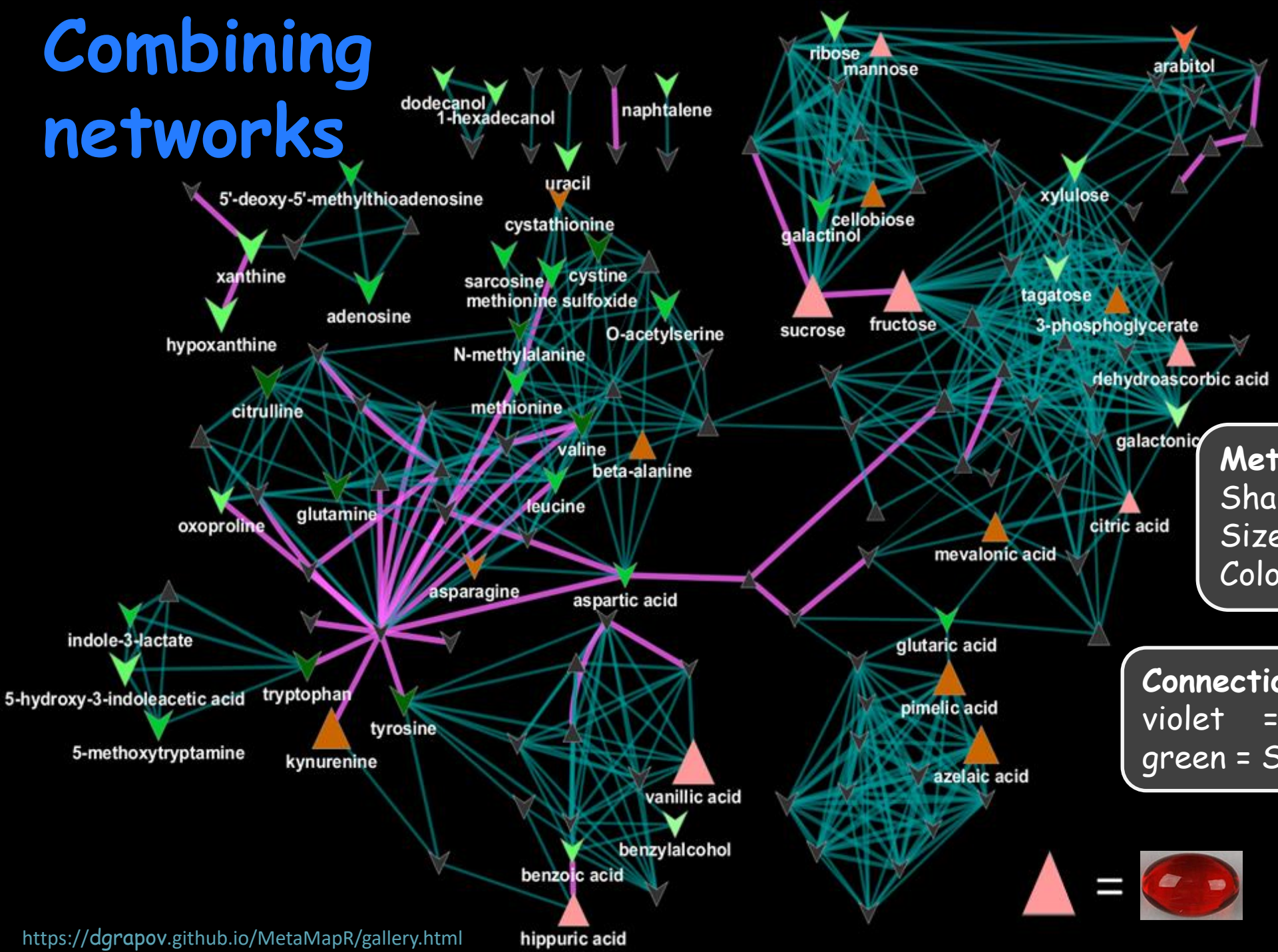
substructure matrix decomposition and
Tanimoto chemical similarity calculations

BMC Bioinformatics 2012, **13**:99 doi:10.1186/1471-2105-13-99

Structural similarity example



Combining networks



Metabolites
Shape = increase/decrease
Size = importance (loading)
Color = correlation

Connections
violet = Biochemical relationships
green = Structural similarity

https://dgrapov.github.io/MetaMapR/gallery.html

Your turn

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/network/

Network refinement and visualization

learn

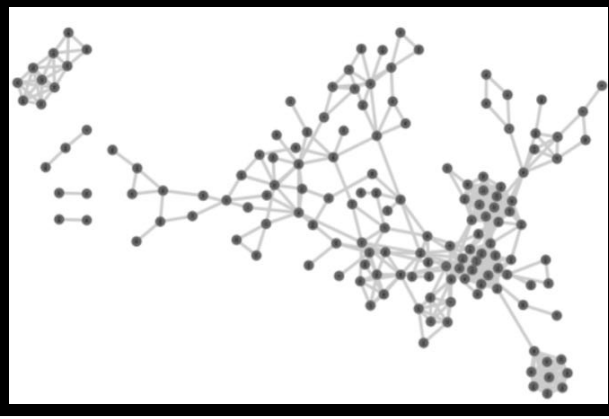
- Cytoscape basics

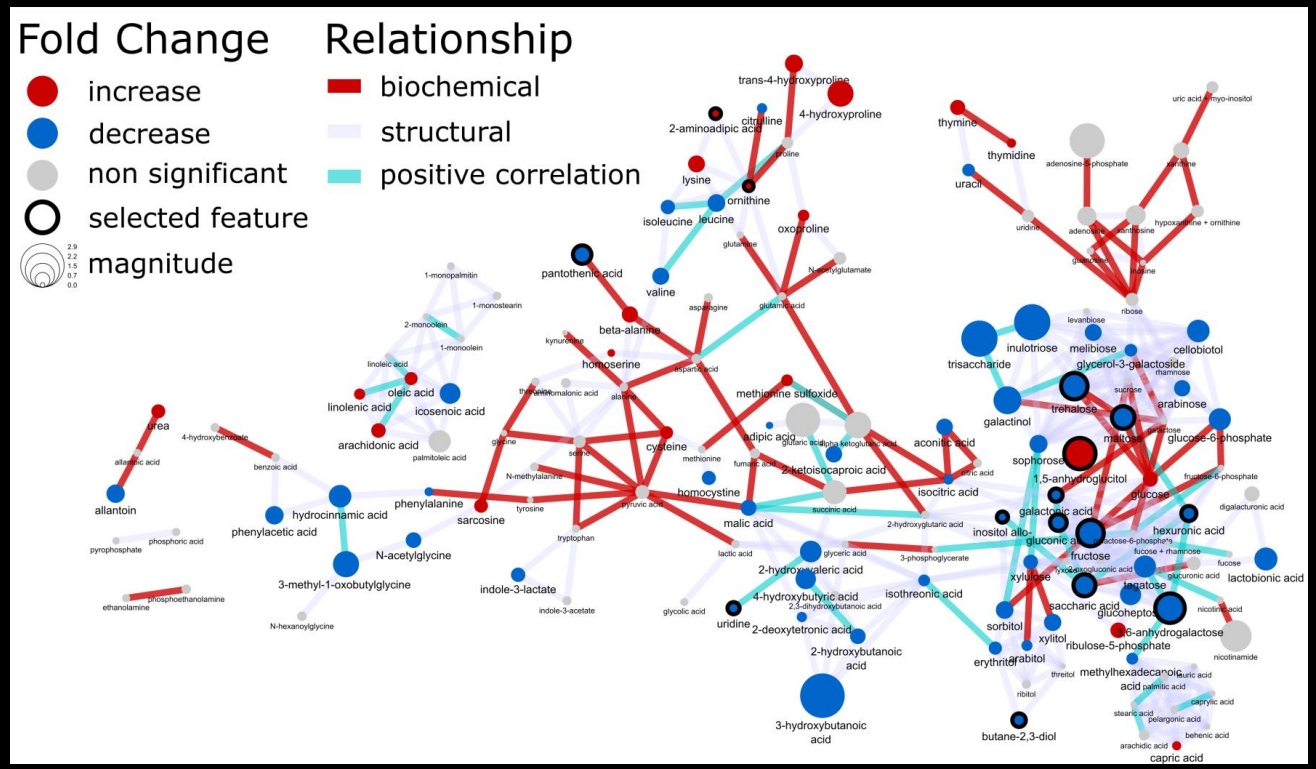
map variables to

- node attributes
- edge attributes

optimize

- layout
- legend
- publication quality figure





https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/cytoscape/

Your turn

Creative Data Solutions

Follow along with the following tutorial:

https://creativedatasolutions.github.io/CDS.courses/courses/network_mapping_101/docs/partial/cytoscape/