ONTARIO
NEURODEGENERATIVE
DISEASE RESEARCH
INITIATIVE

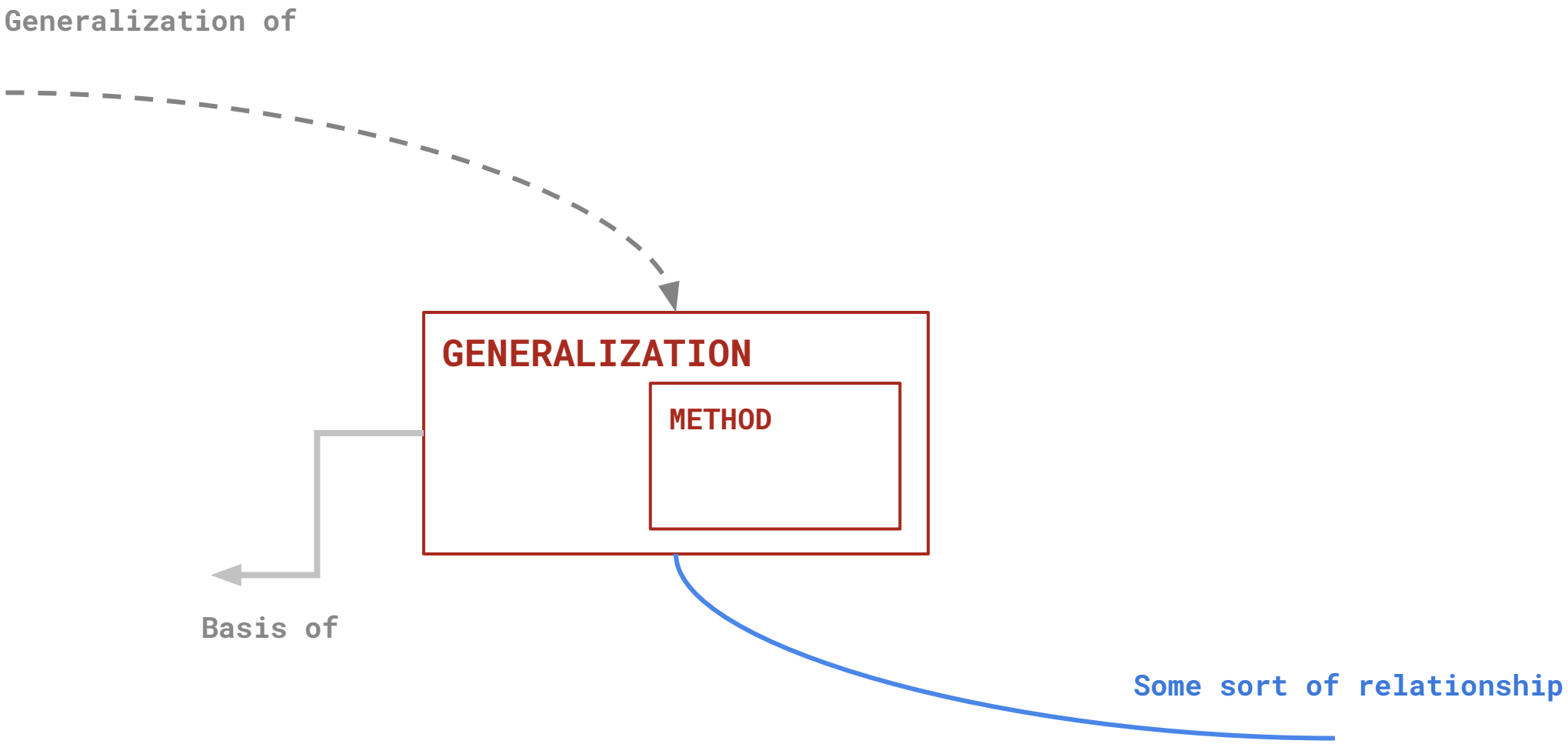
Survey of Data Methods Useful in the ONDRI Environment: Part III

Kelly Sunderland, Malcolm Binns, Derek Beaton, Pradeep Raamana

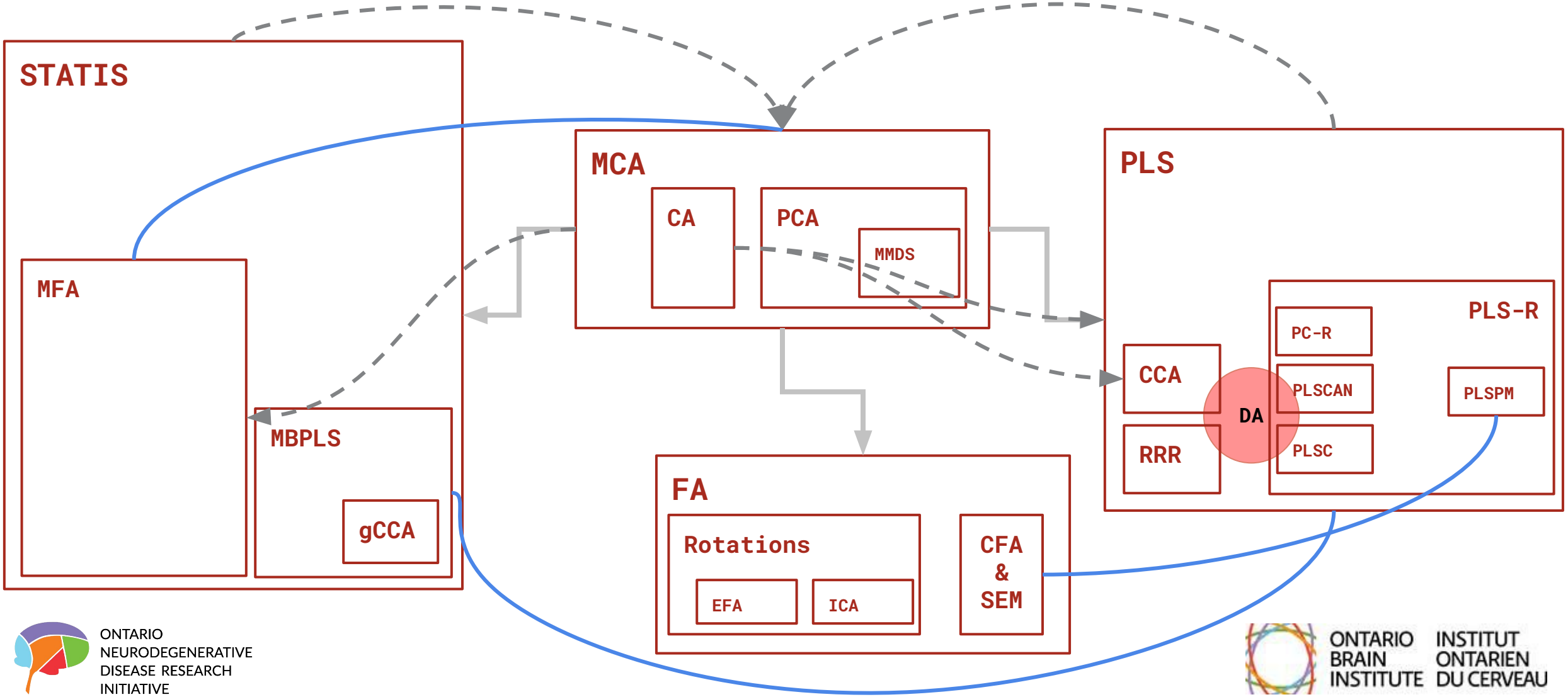
2019-Sep-10

Ordination

- Dimensionality reduction/projection
- Subspace/manifold method (learning)
- Orthogonal transformation
- Diagonalization or matrix factorization
- Matrix approximations
- Matrix decomposition
- Linear autoencoder (single layer a.k.a. stupid neural network)
- Sometimes correctly, sometimes incorrectly: factor analyses
- Spectral decomposition
- (Specific types of) “MVPA” and “RSA” in neuroimaging
- Multivariate statistics
- SURPRISE: It’s all just sort of Principal Components Analysis (PCA)


Generalization of
GENERALIZATION
METHOD
Basis of
Some sort of relationship

Chaos!



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

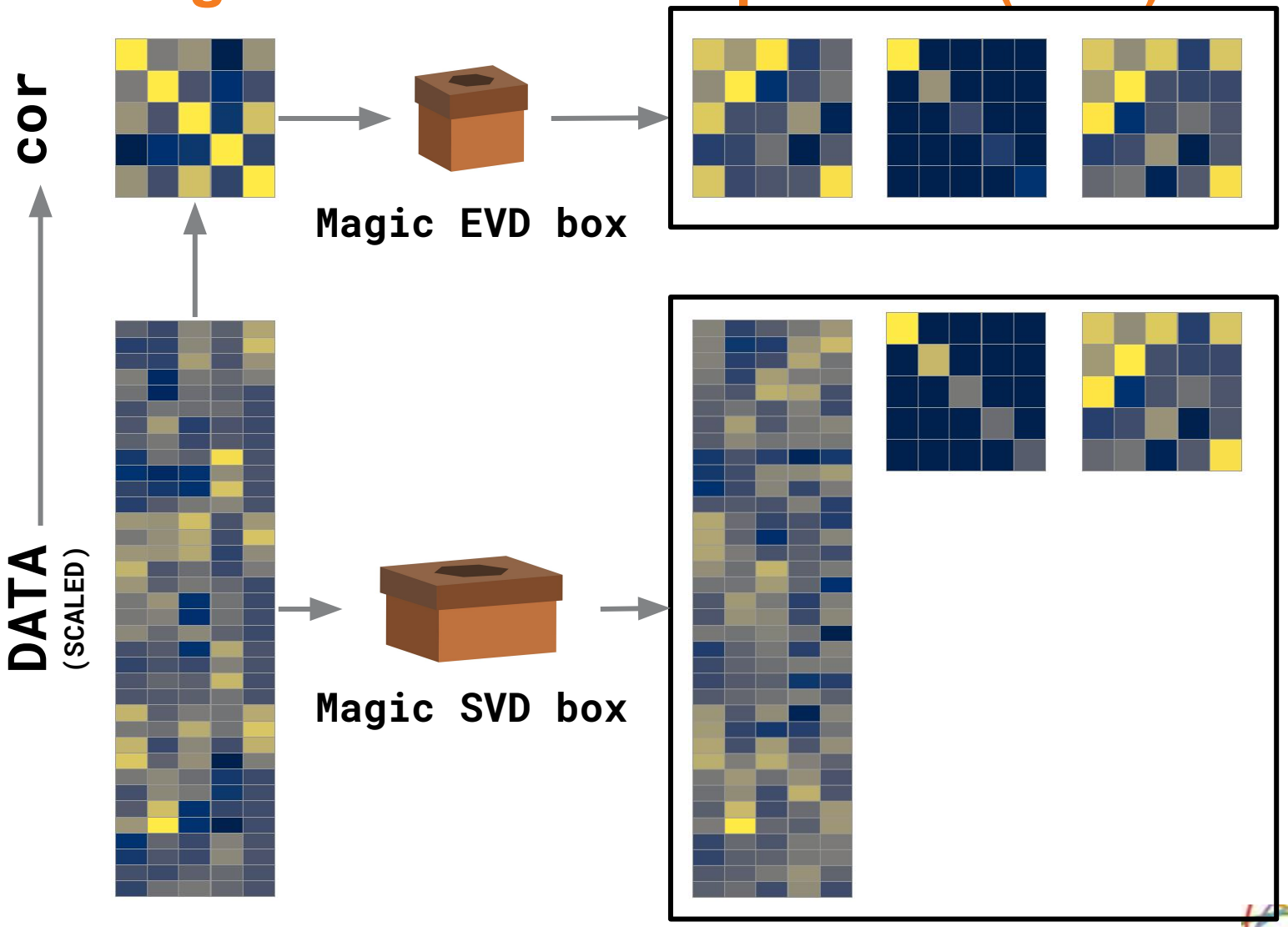
Overview

- PCA
- Something like
  - a PCA but with multiple tables, or structure for the columns?
  - a correlation or regression between tables?
  - a PCA but for all those weird types of data?

A component (sometimes a.k.a. factor)

- A new variable
  - Bits & pieces (weights, “loadings”) of ***all*** original variables
- Each explains a proportion of total variance
- All observations exist along them
- Is orthogonal subsequent to the previous components

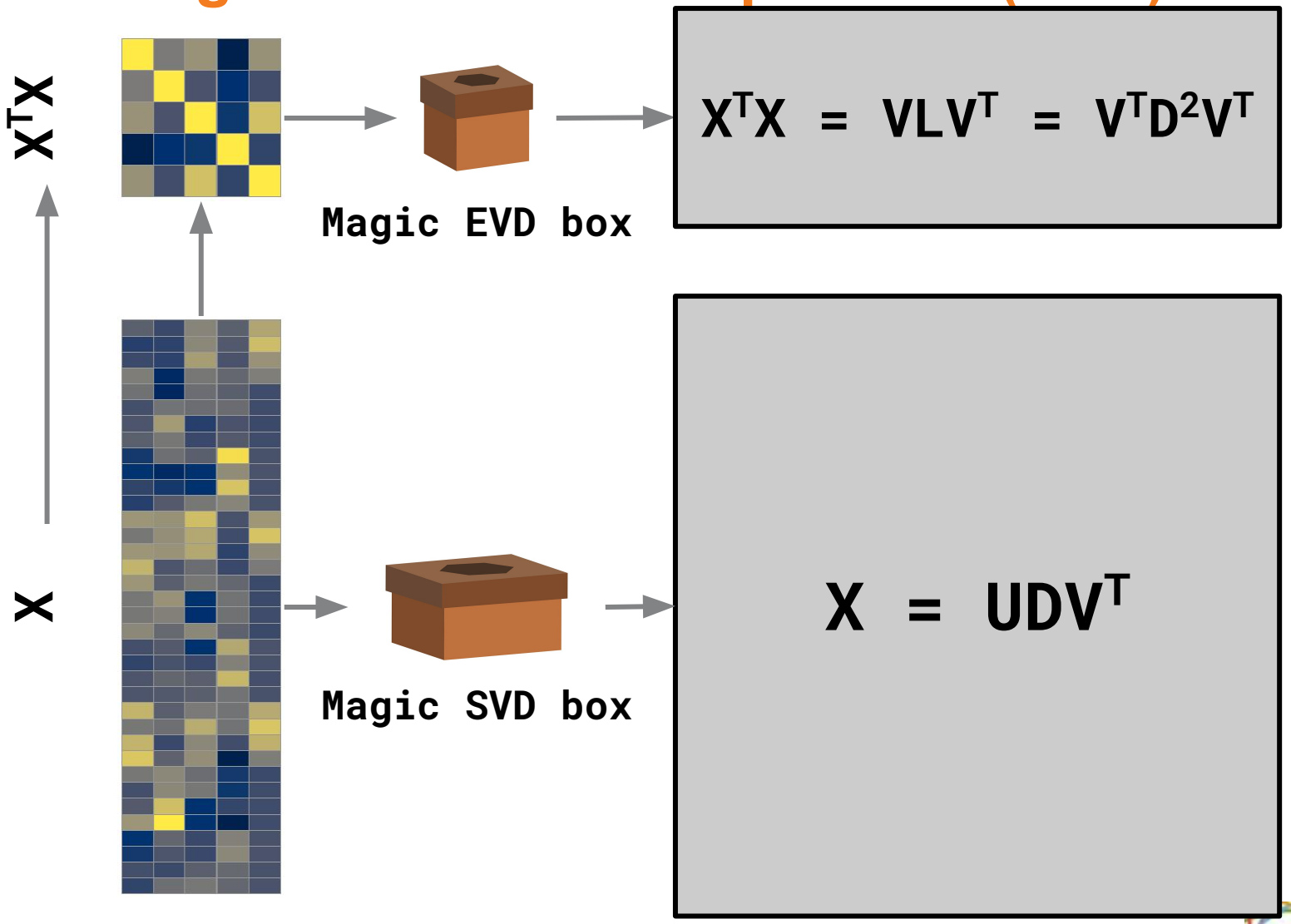
Eigenvalue Decomposition (EVD)
Singular value decomposition (SVD)



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

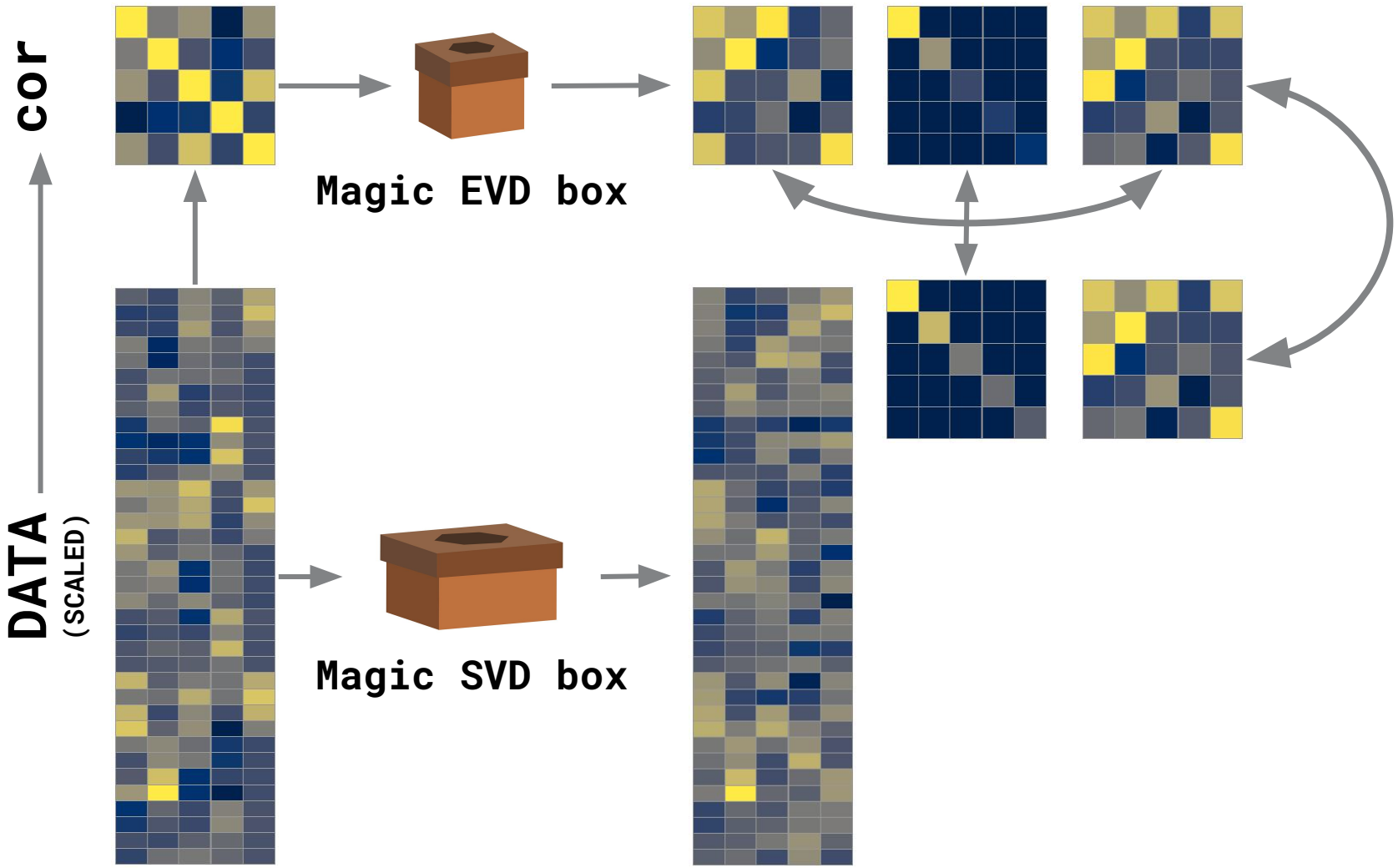
Eigenvalue Decomposition (EVD)
Singular value decomposition (SVD)



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

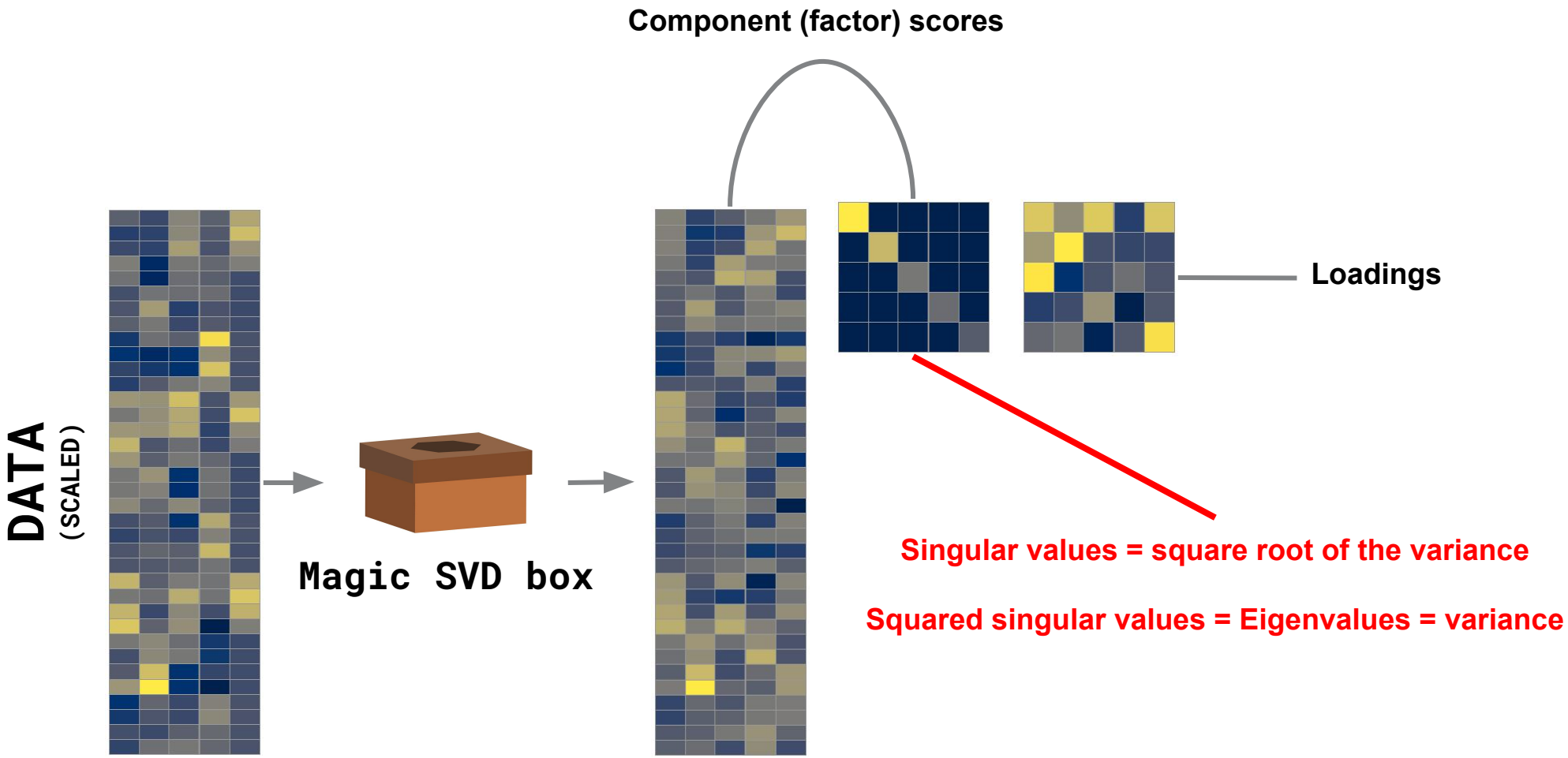
Eigenvalue Decomposition (EVD)
Singular value decomposition (SVD)



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

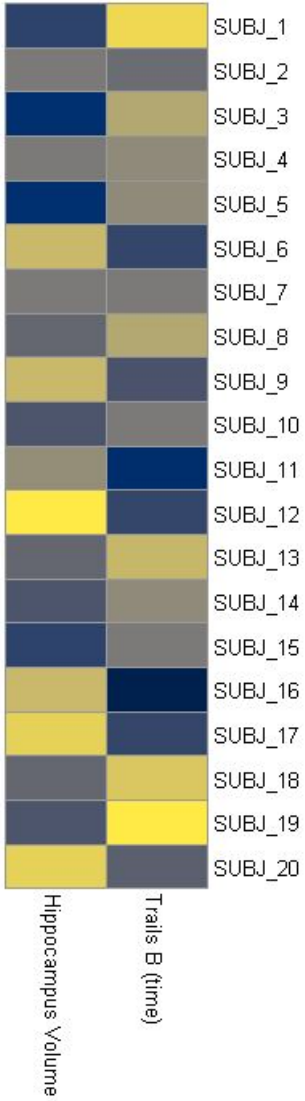
Eigenvalue Decomposition (EVD)
Singular value decomposition (SVD)

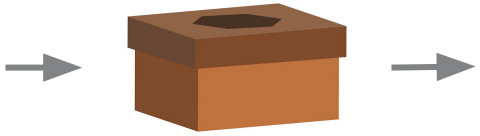


ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

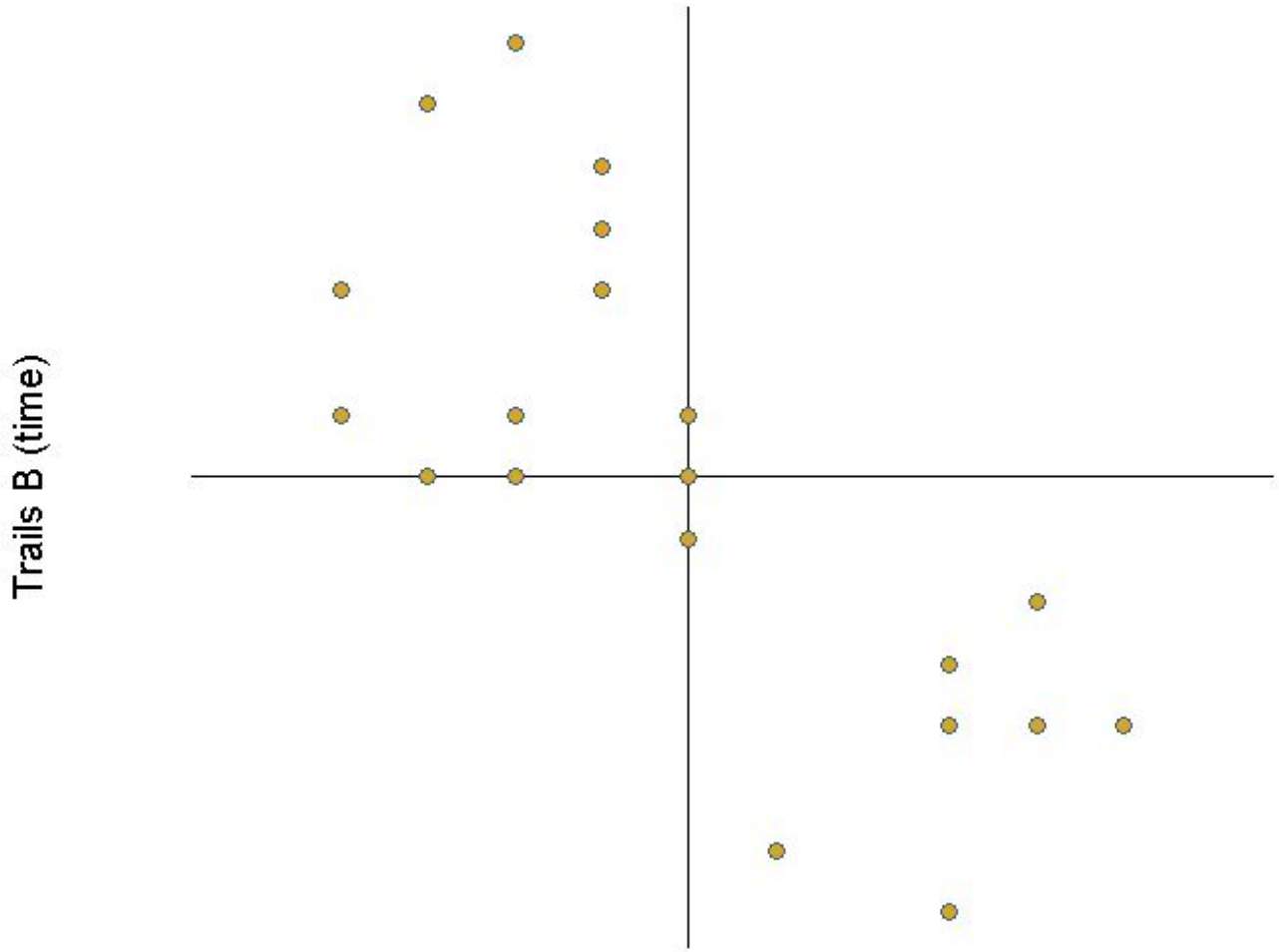
ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Tiny example





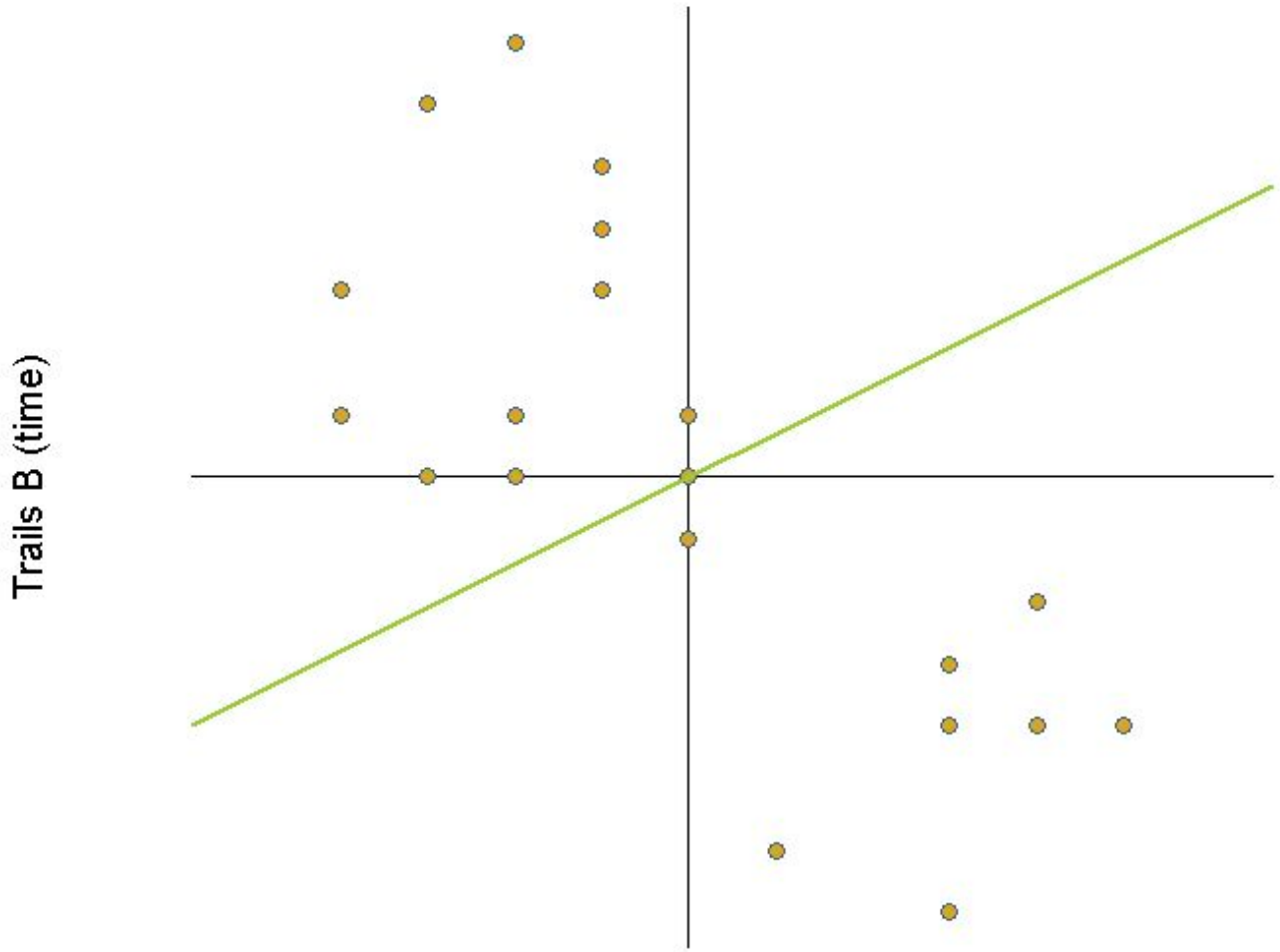
Magic SVD box



ONTARIO
NEURODEGENERATIVE
DISEASE RESEARCH
INITIATIVE

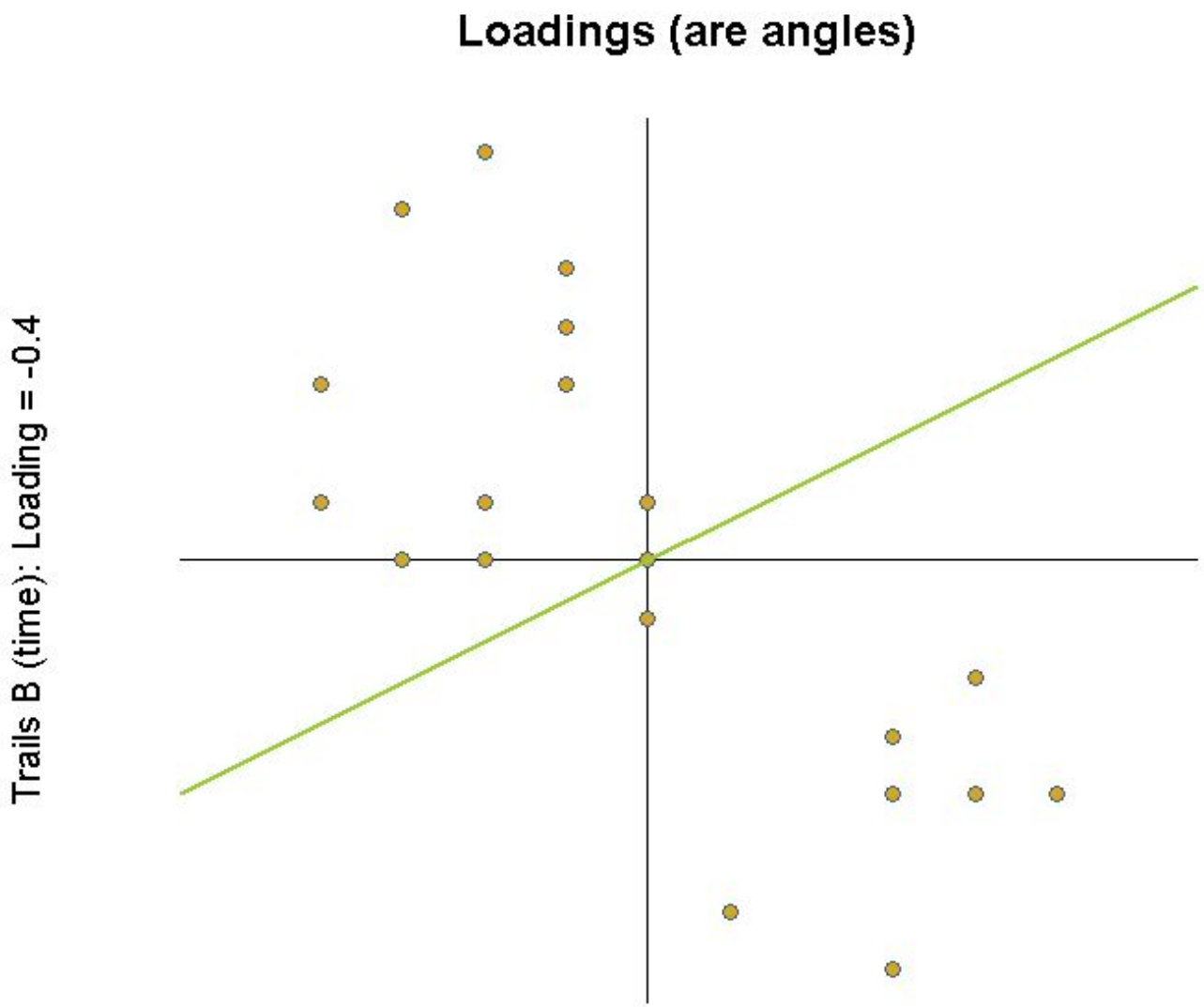
ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

A (terrible) component?



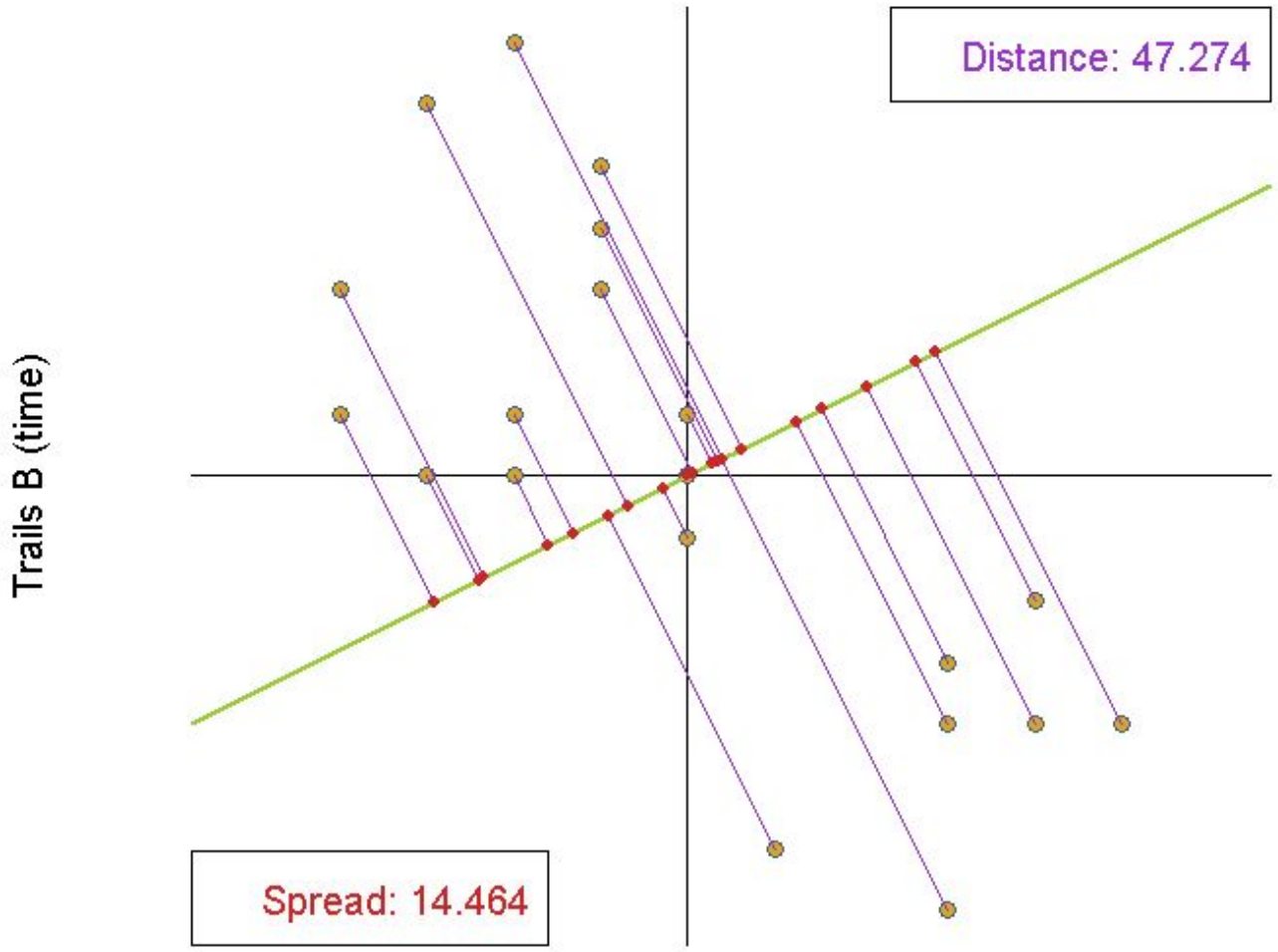
ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU



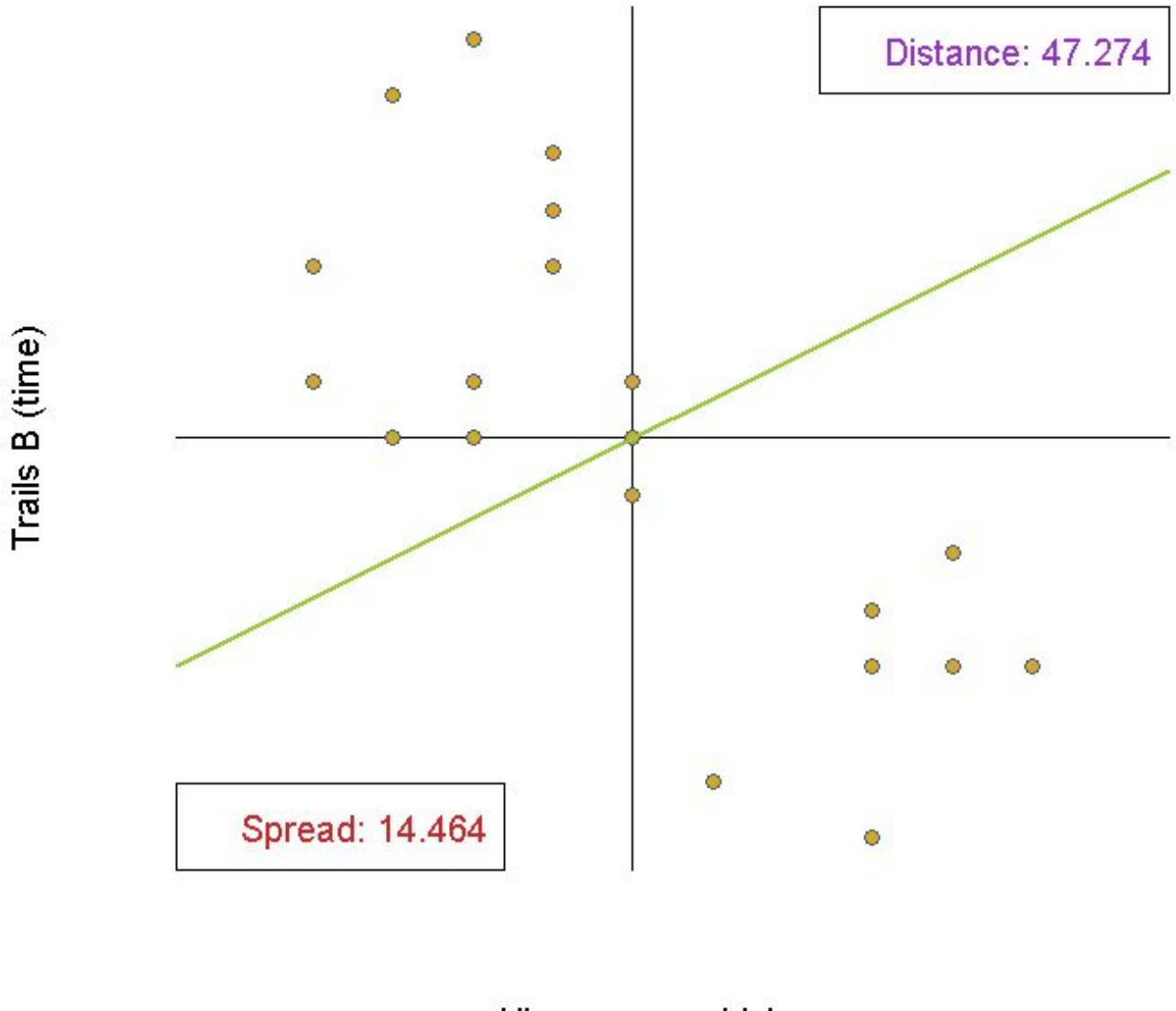
ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

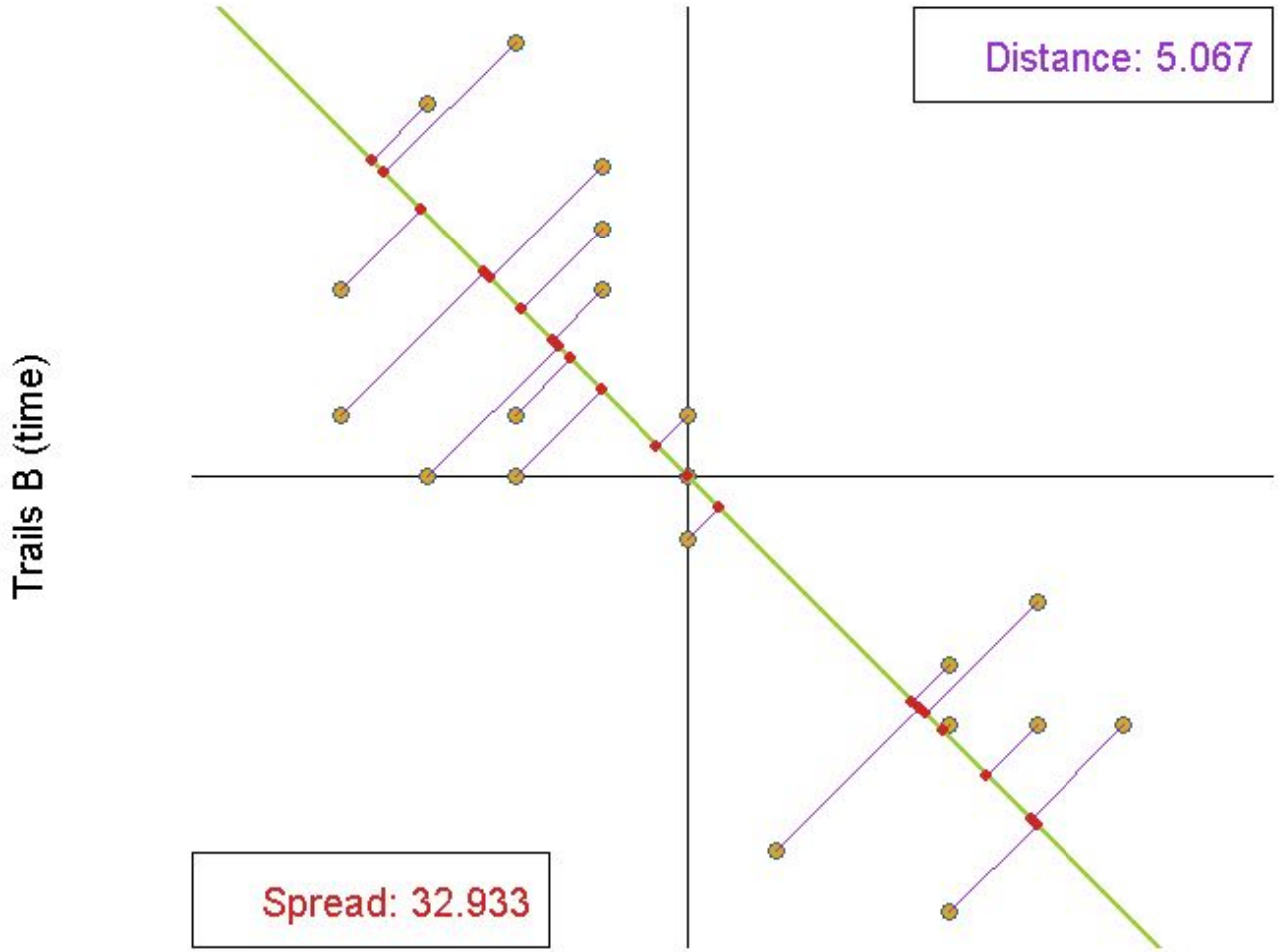
ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

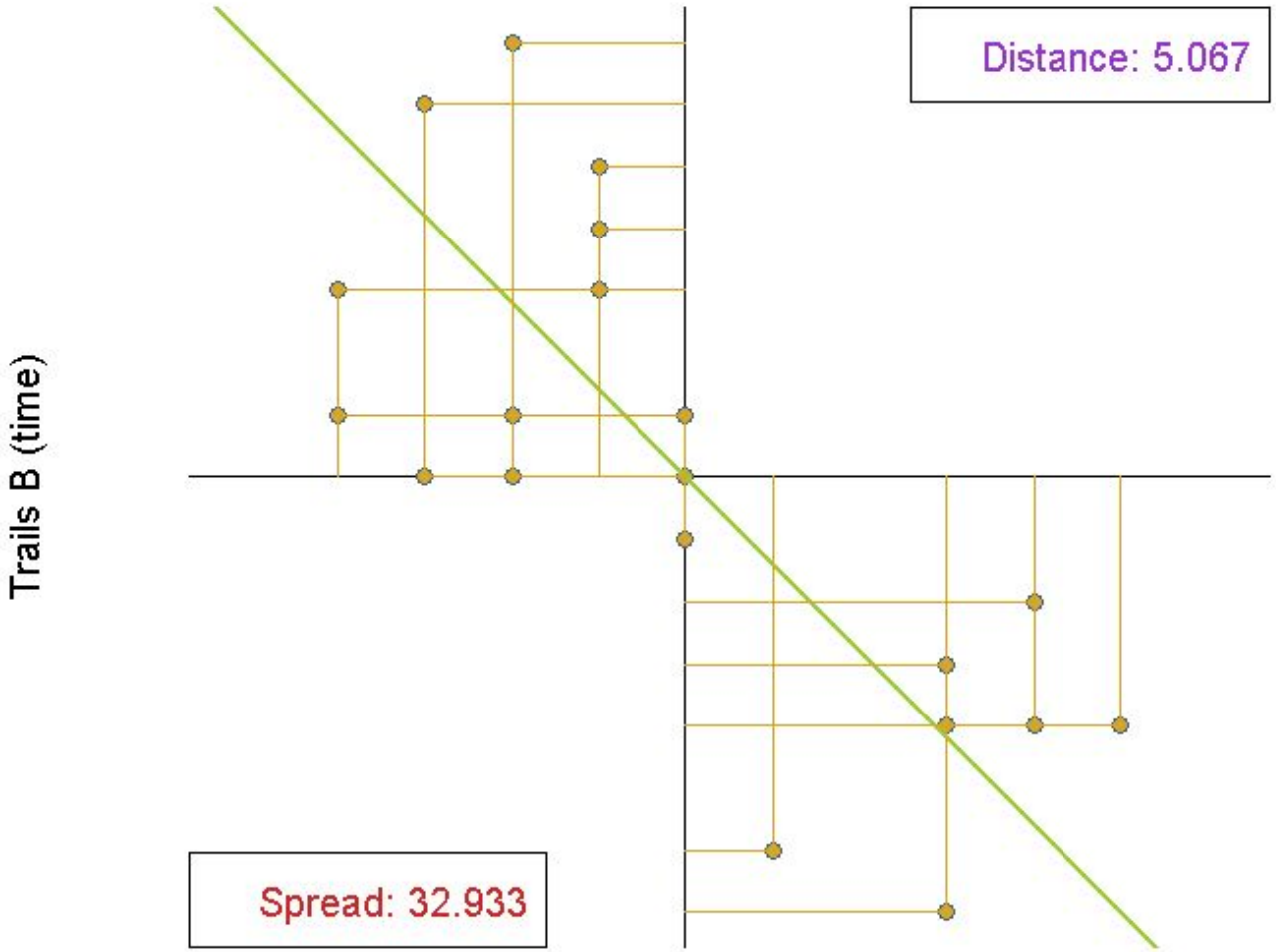
Et Voila!



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

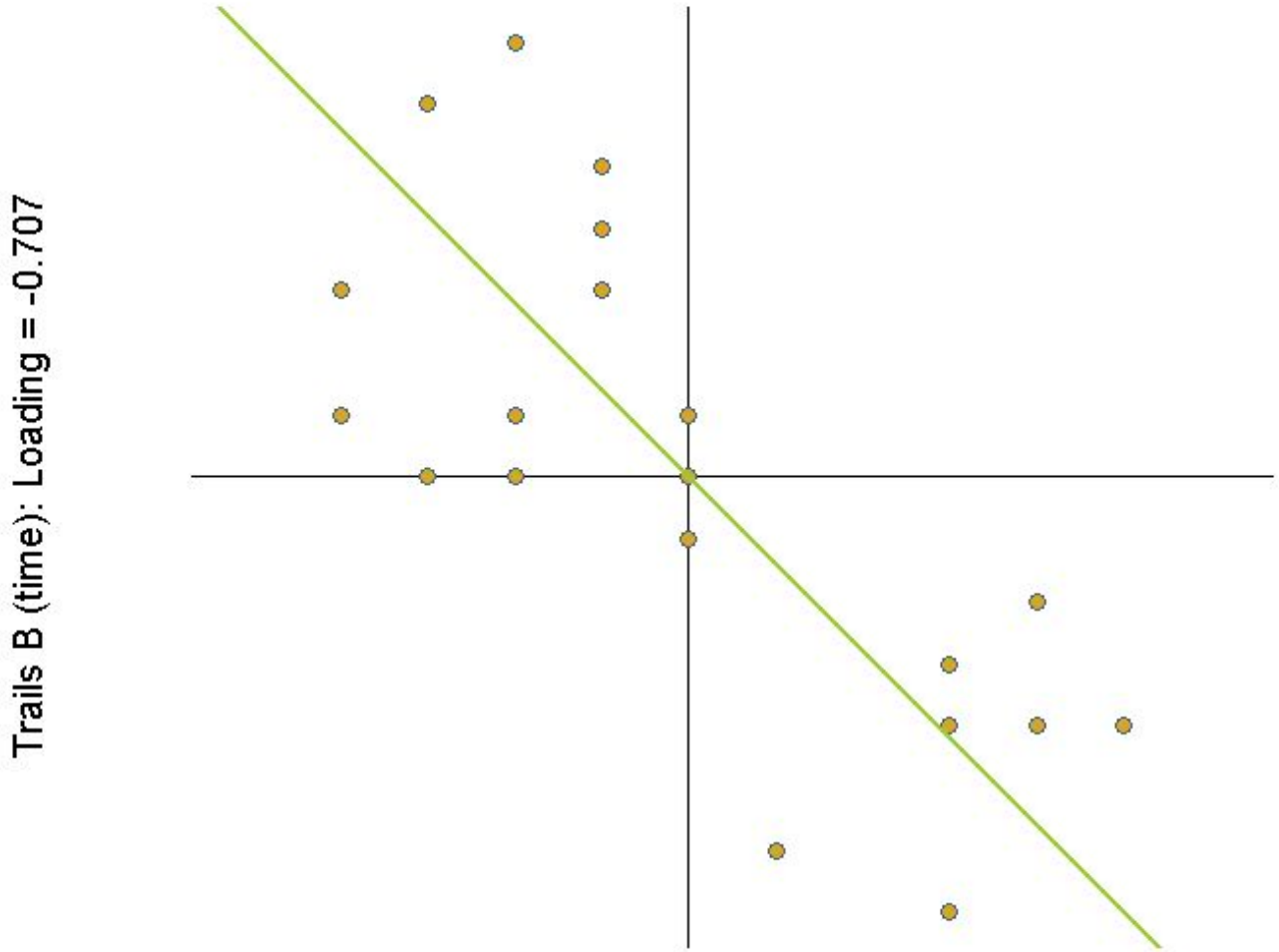
Best fit of rectangles



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

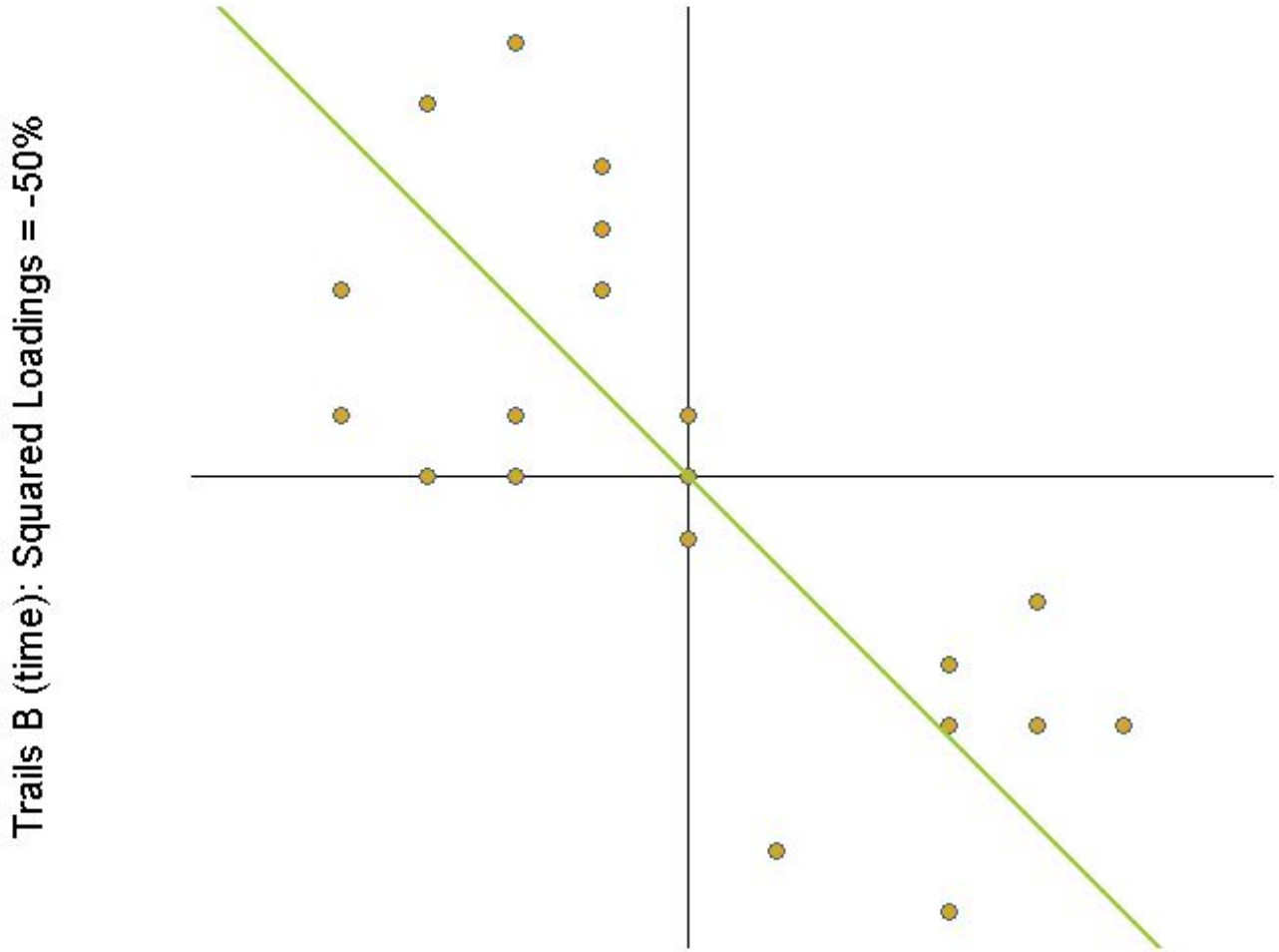
Loadings & Contributions



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

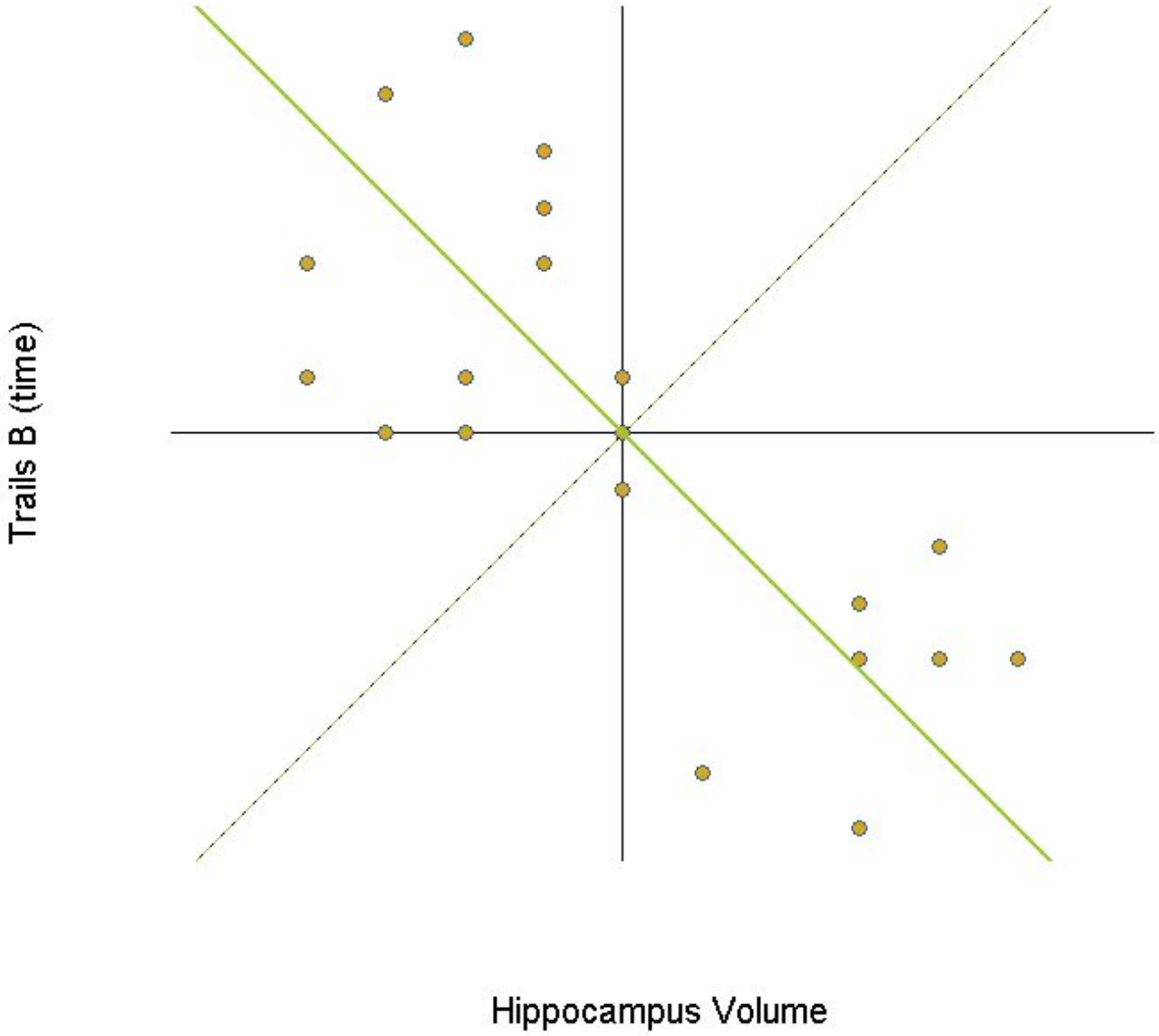
Contributions
are signed squared loadings * 100



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

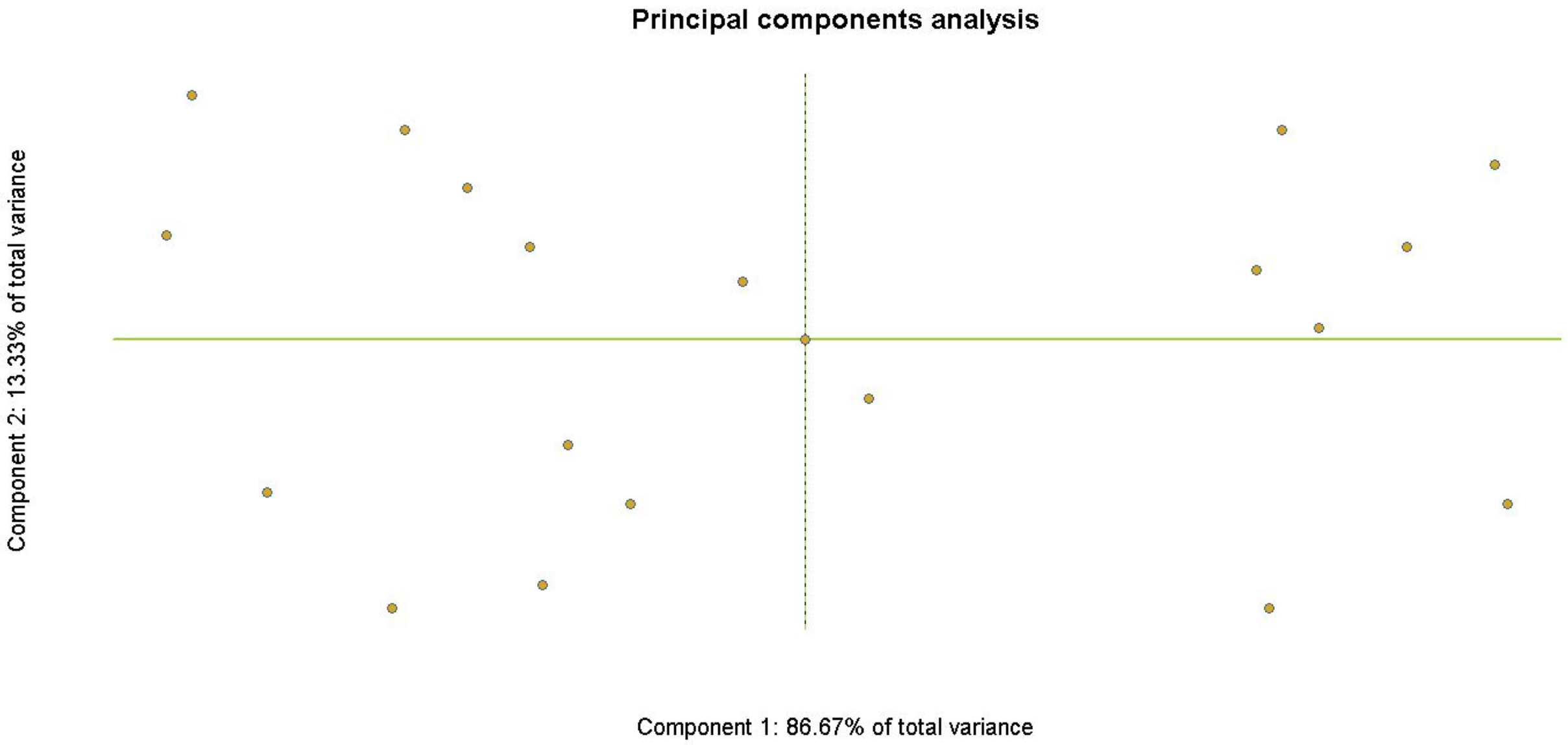
ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

The second component



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

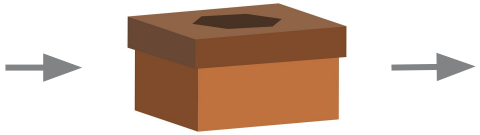
ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Scaling up

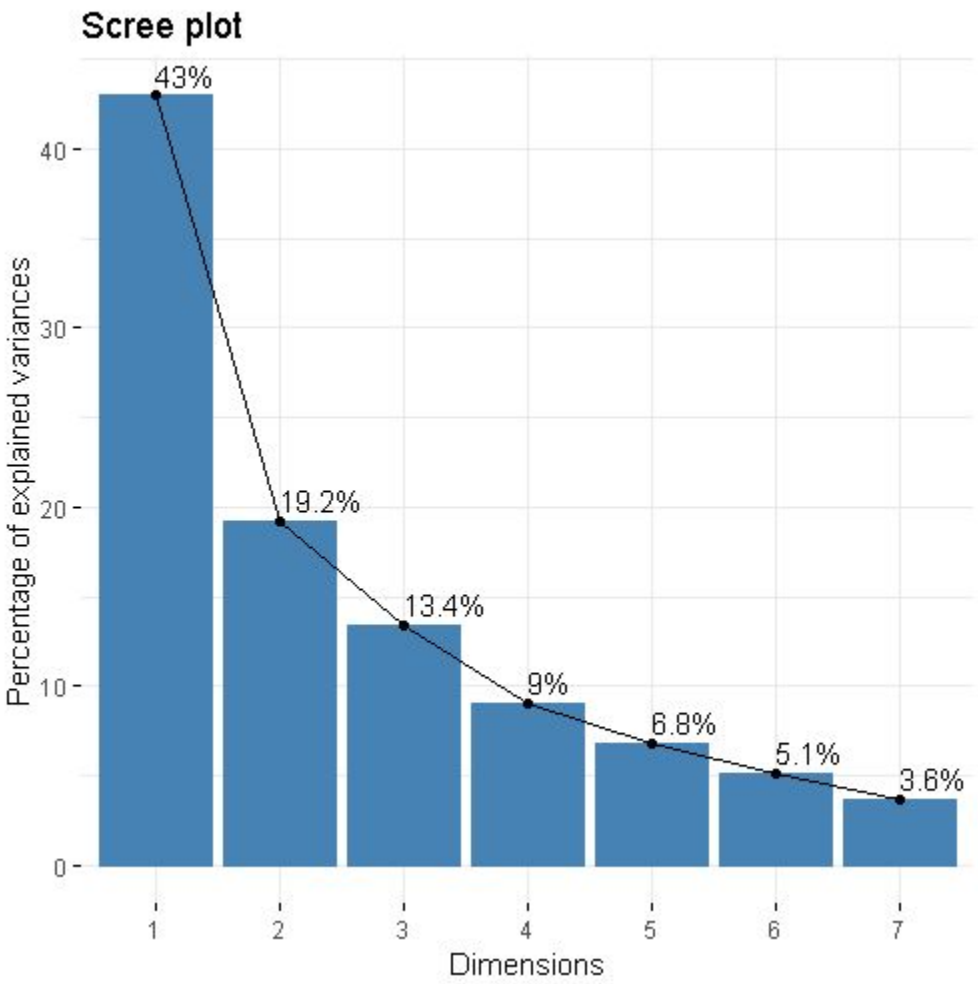


Magic SVD box

ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

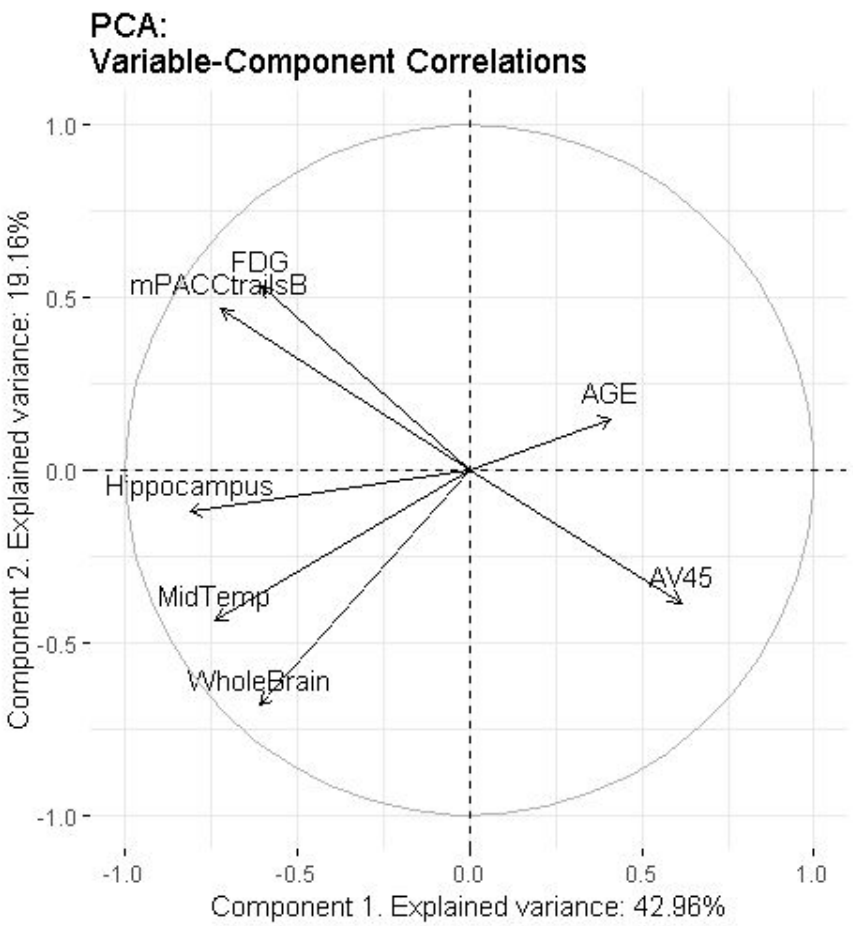
Scaling up

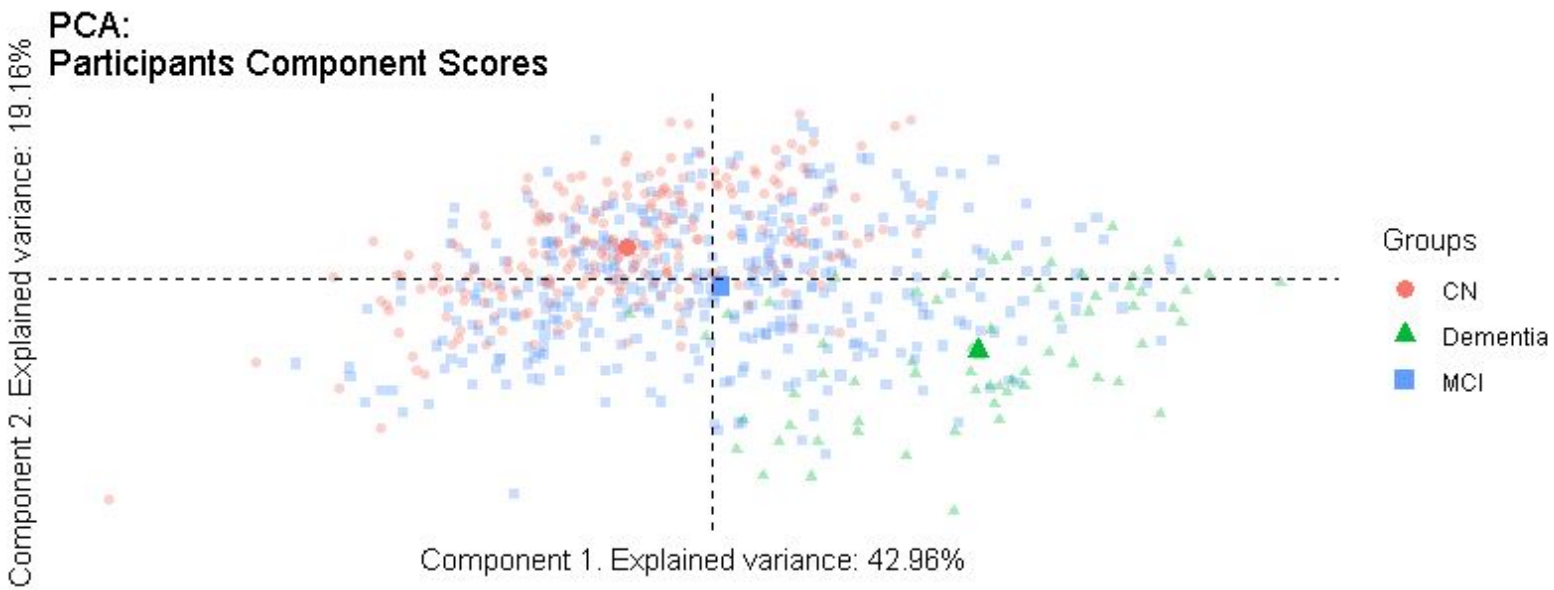


ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Scaling up

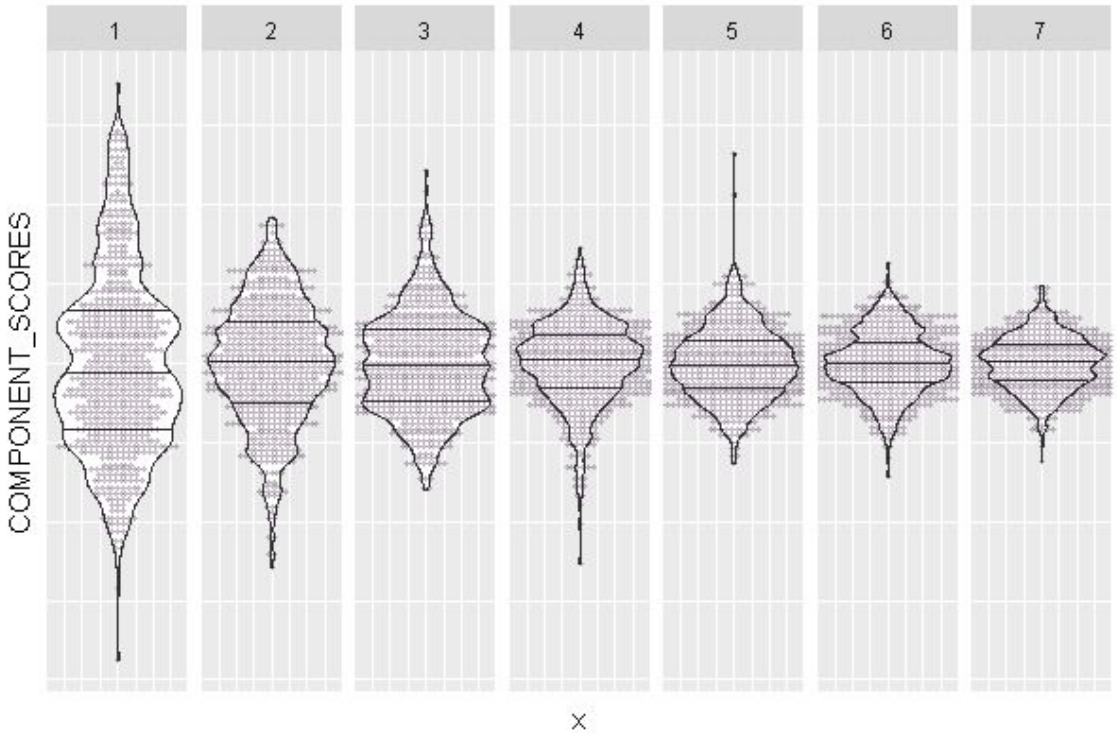


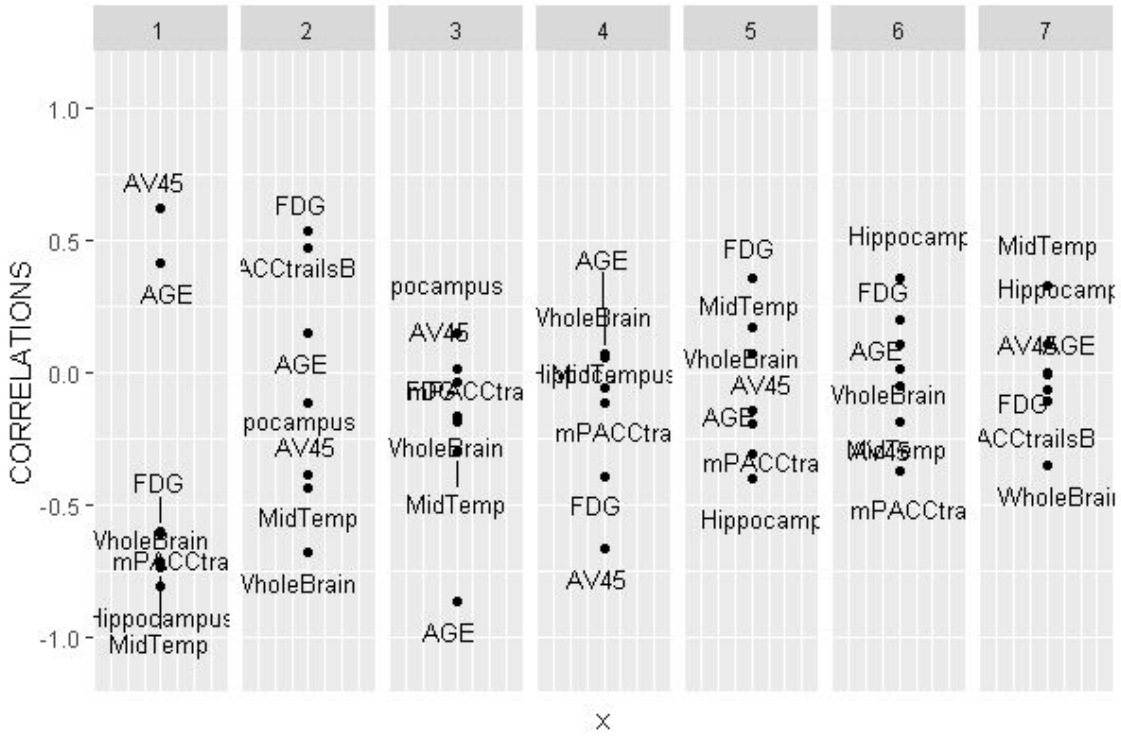


ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Scaling up





ONTARIO
NEURODEGENERATIVE
DISEASE RESEARCH
INITIATIVE

ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

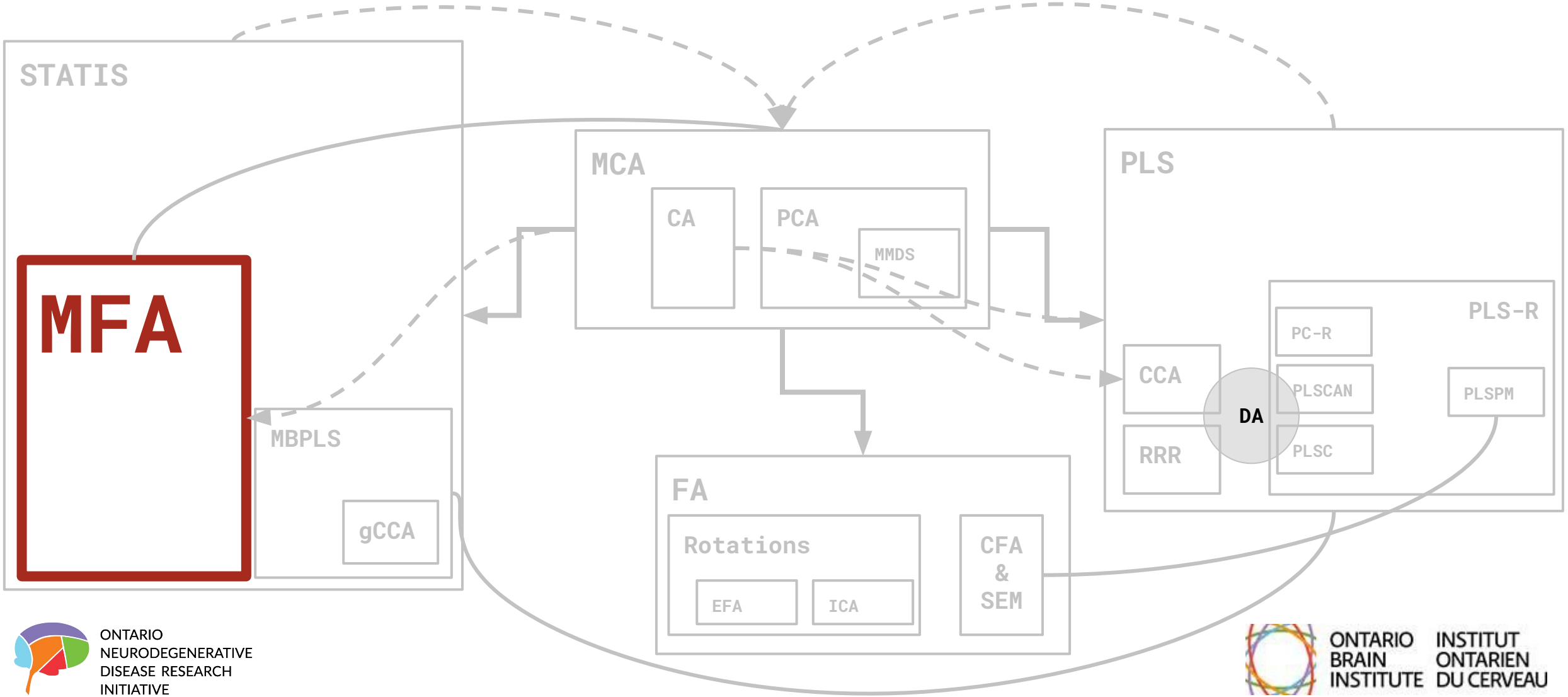
What if things are more complex?

- PCA
- Something like
  - a PCA but with multiple tables, or structure for the columns?
  - a correlation or regression between tables?
  - a PCA but for all those weird types of data?

What if things are more complex?

- PCA
- Something like
  - a PCA but with multiple tables, or structure for the columns?
  - a correlation or regression between tables?
  - a PCA but for all those weird types of data?

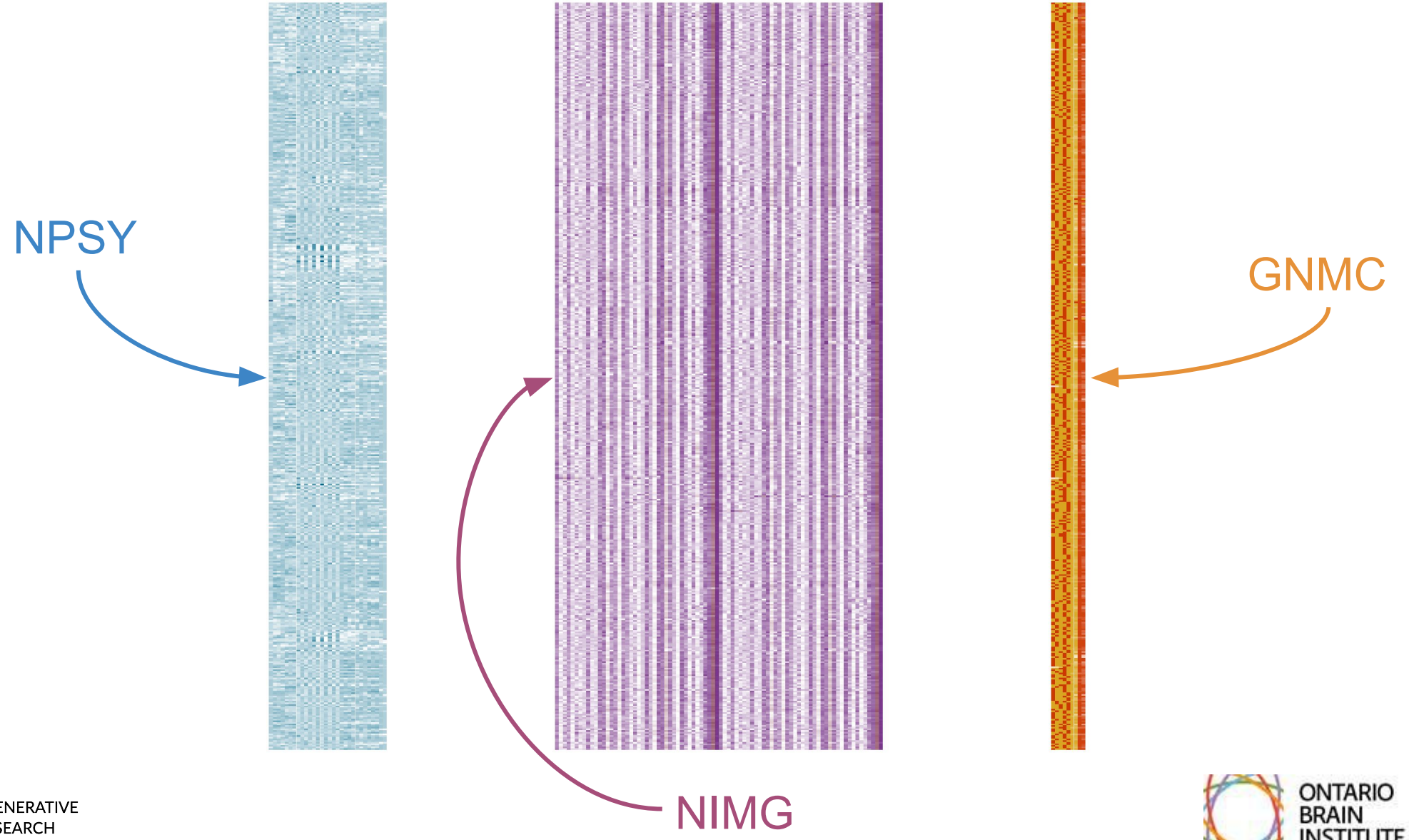
Chaos!



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

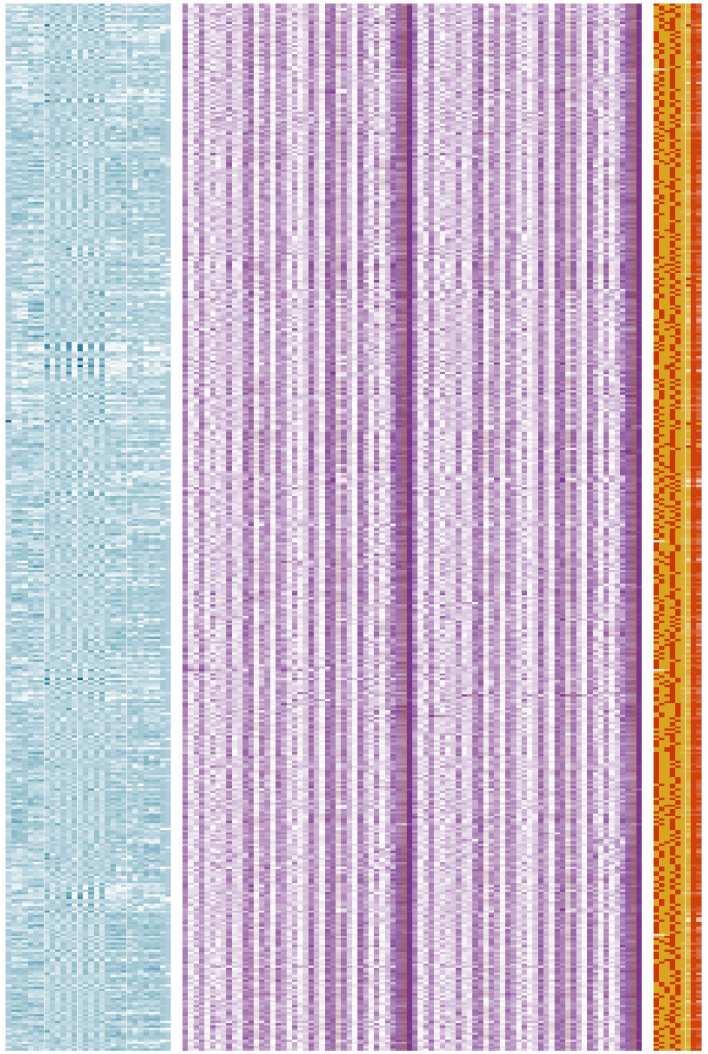
Multiple Factor Analysis

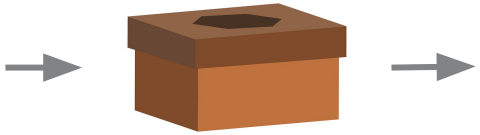


ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Multiple Factor Analysis





Magic SVD box

BAD IDEA

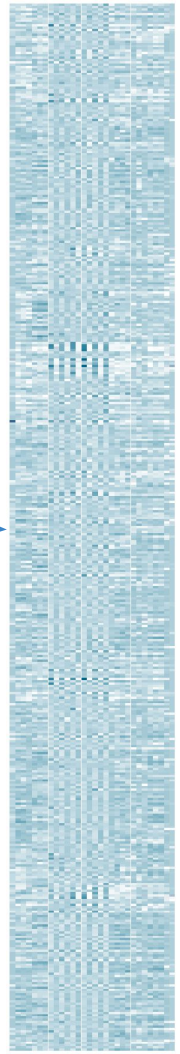
123 total variables
Each variable is normed

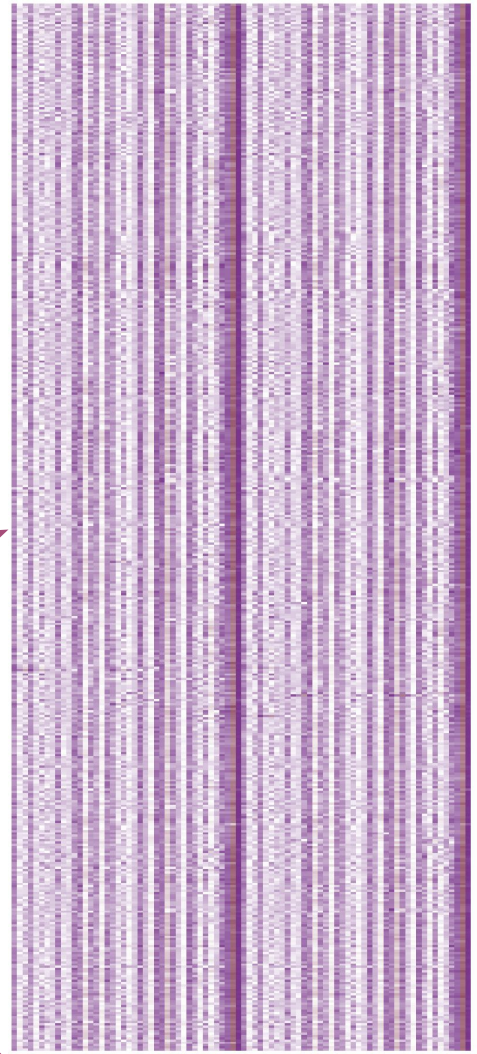
ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Multiple Factor Analysis

30 / 123 = ~24 %



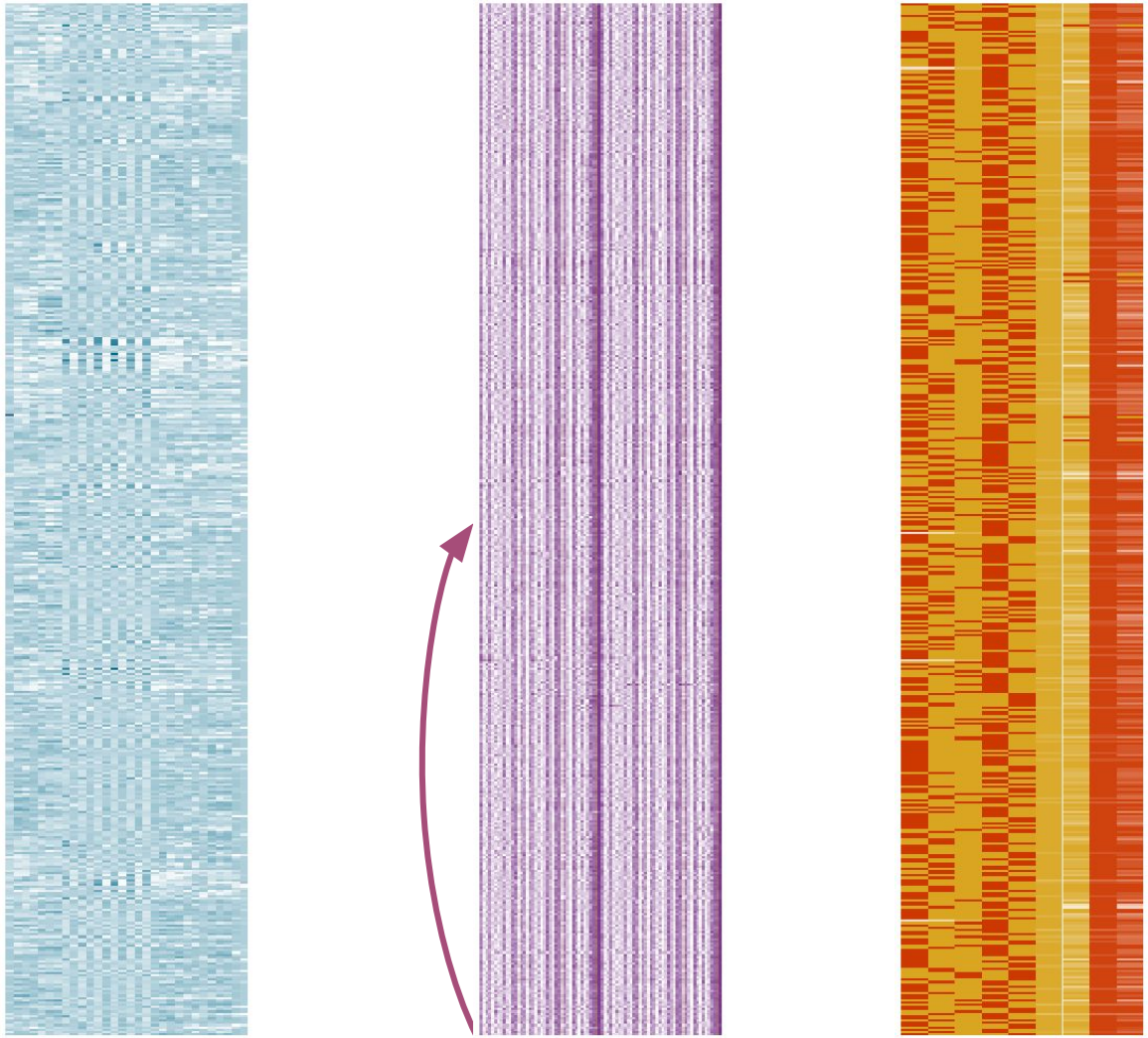


84 / 123 = ~68%

9 / 123 ~7%

Multiple Factor Analysis

30 variables
1 / 3 = ~33 %



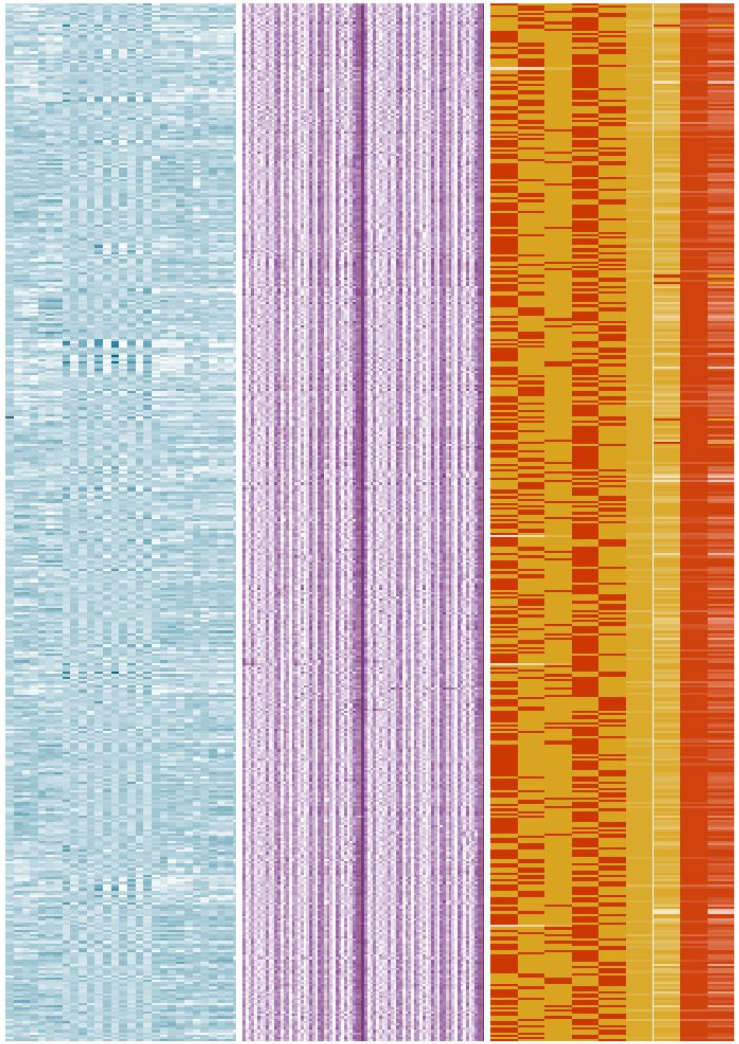
9 variables
1 / 3 = ~33%

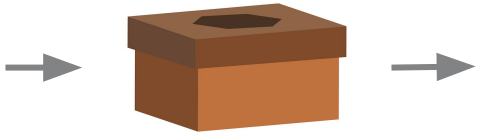
84 variables
1 / 3 = ~33%

ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Making it Fair Analysis





Magic SVD box

123 total variables
Each variable <u>and table</u> is normed

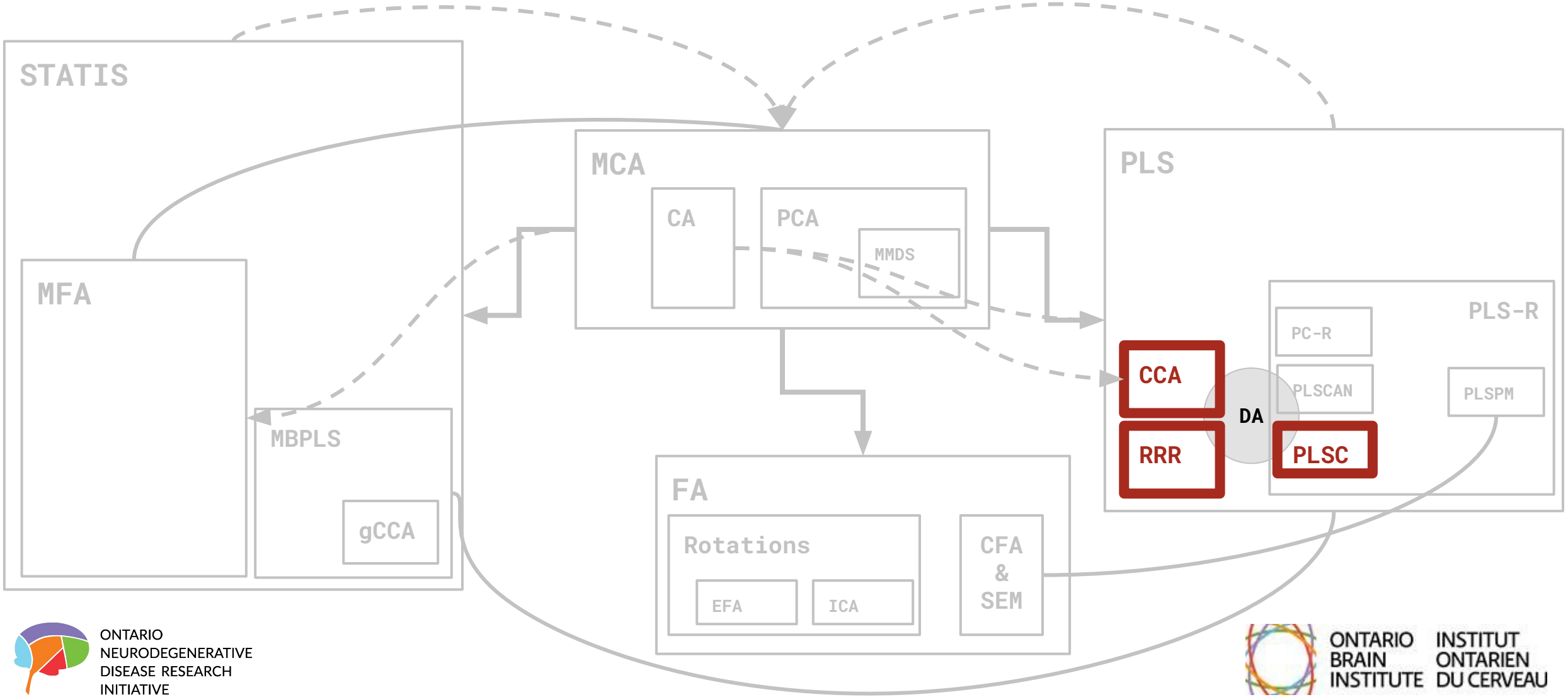
ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

What if things are more complex?

- PCA
- Something like
  - a PCA but with multiple tables, or structure for the columns?
  - a correlation or regression between tables?
  - a PCA but for all those weird types of data?

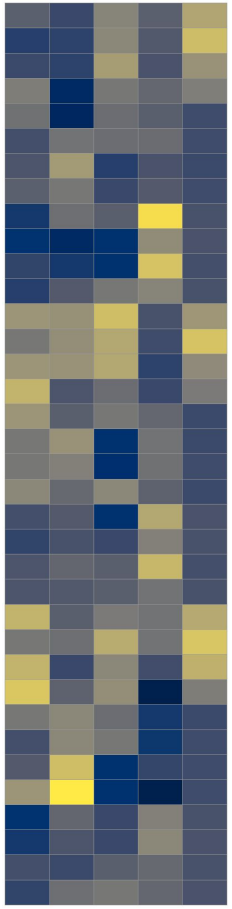
Chaos!



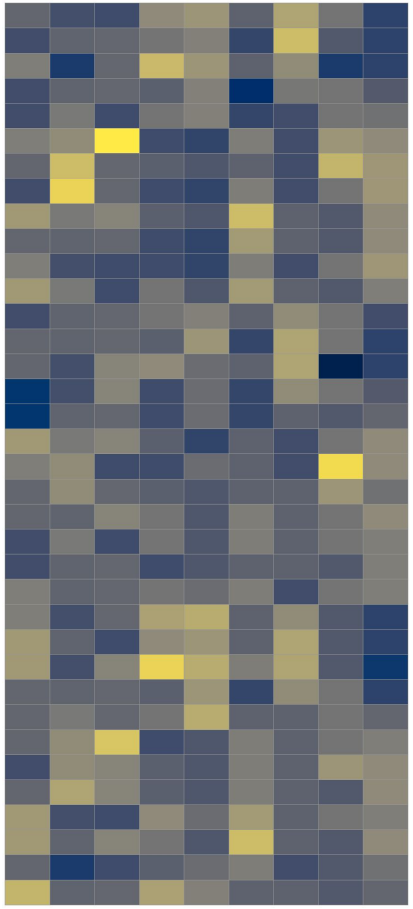
ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Between Two Tables



X

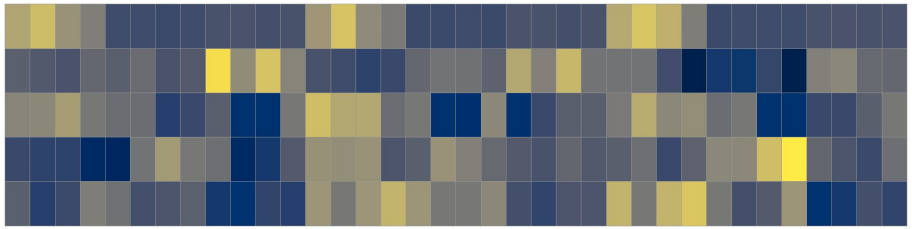


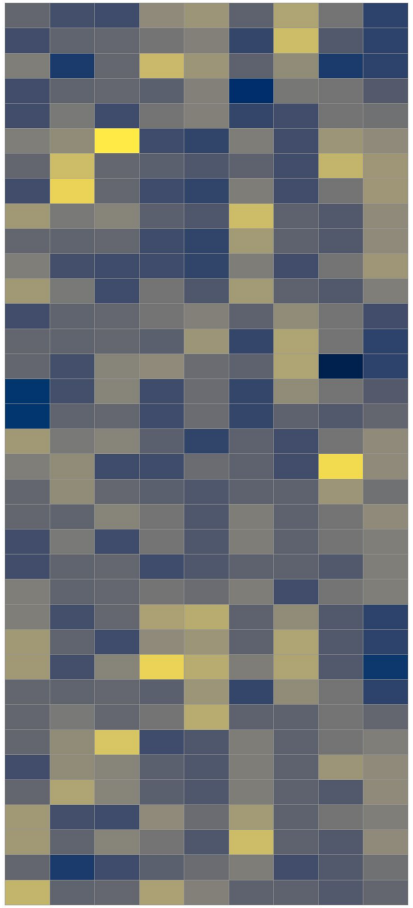
Y

ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

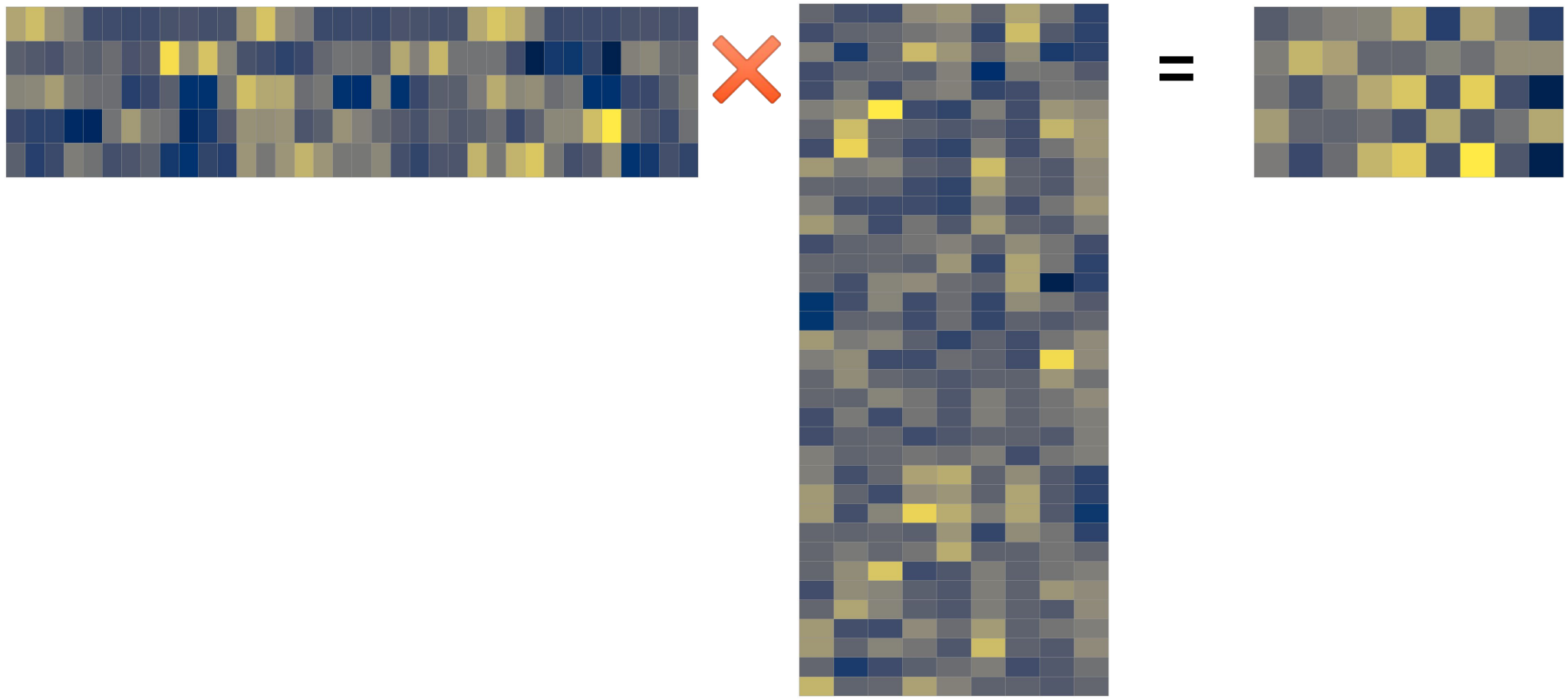
Between Two Tables





X **Y**

Between Two Tables



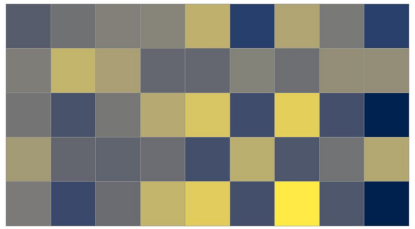
X Y R

ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

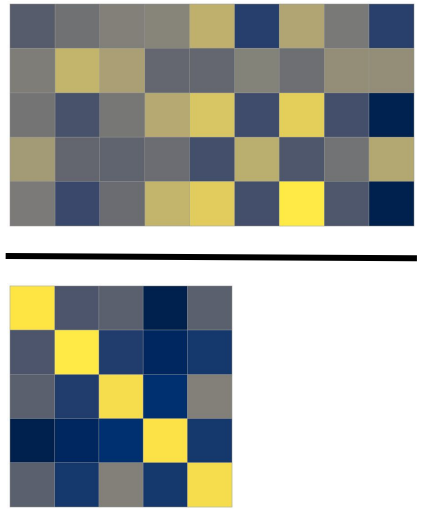
Between Two Tables

Partial least squares ("correlation")



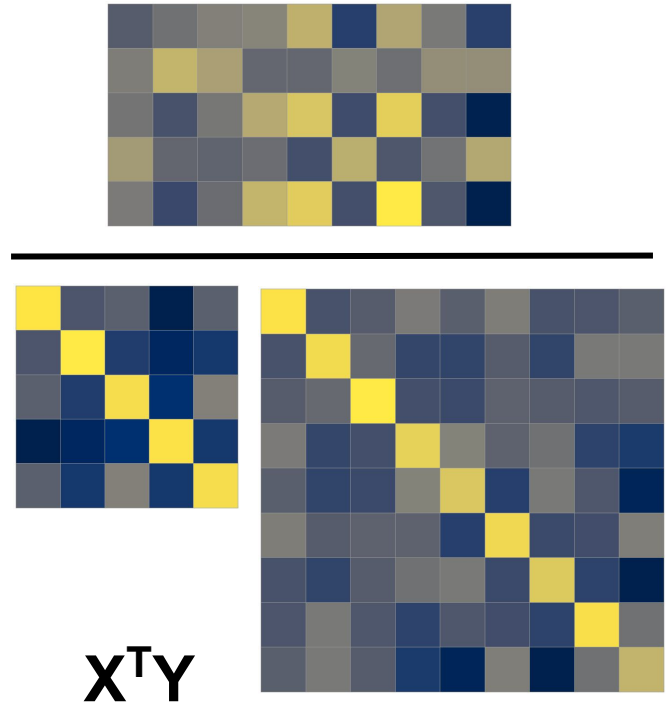
$$X^TY$$

Reduced rank regression



$$\frac{X^TY}{X^TX}$$

Canonical correlation analysis



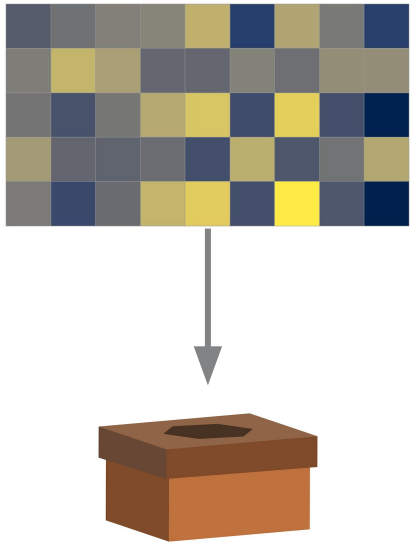
$$\frac{X^TY}{(X^TX)(Y^TY)}$$

ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

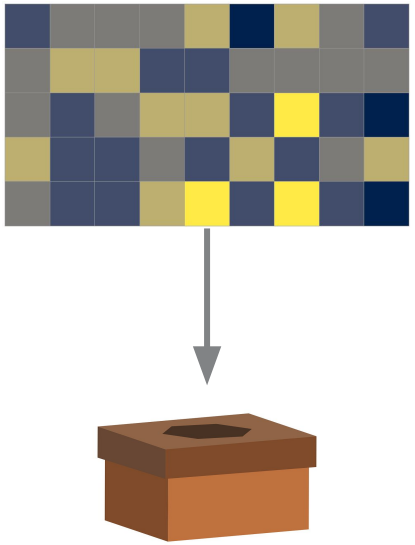
Between Two Tables

Partial least squares
(“correlation”)



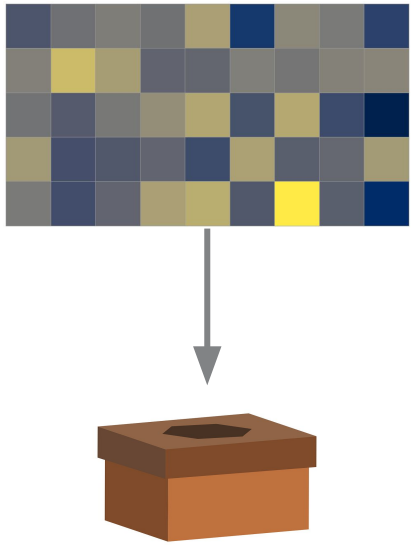
Magic SVD box

Reduced rank regression



Magic SVD box

Canonical correlation
analysis



Magic SVD box

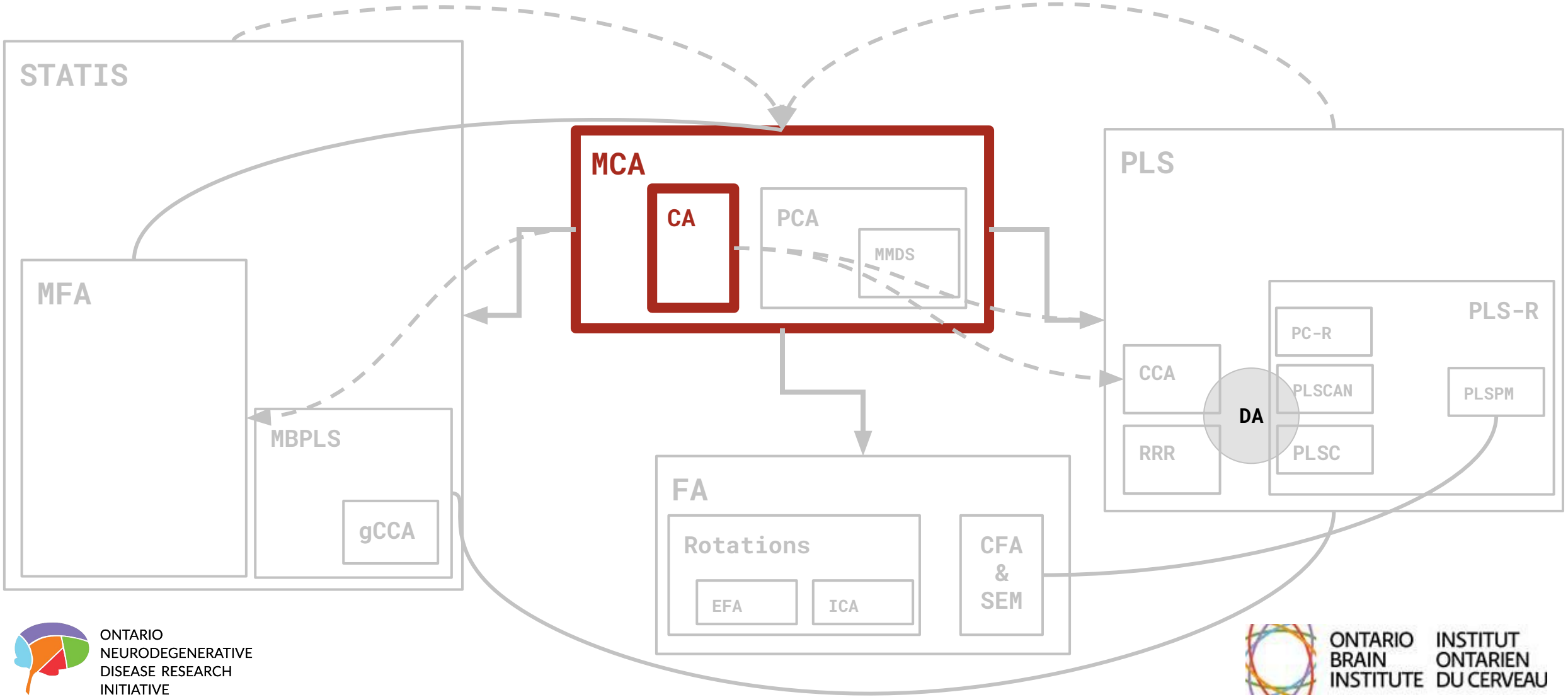
What if things are more complex?

- PCA
- Something like
  - a PCA but with multiple tables, or structure for the columns?
  - a correlation or regression between tables?
  - a PCA but for all those weird types of data?

Everything up until now

- Generally normal(-ish) variables
- Assumed strictly continuous
- What about
  - Non-normal?
  - Counts?
  - Ordinal or Likert?
  - Lots of zeros?
  - Categorical?
- That you can compute a ***meaningful*** correlation matrix

Chaos!



ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

Correspondence Analysis

	DX	PTRACCAT
5023	CN	Asian
5026	MCI	White
5027	Dementia	White
5028	Dementia	White
5031	MCI	White
5037	Dementia	Black
5040	CN	Black
5047	MCI	Black
5054	Dementia	White
5058	Dementia	Asian
5063	Dementia	White

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

Correspondence Analysis

- "coding categorical variables with the indicator matrix of dummy variables and considering them as Gaussian, for instance, is almost a crime."
  - "Jan de Leeuw and the French School of Data Analysis" (Husson, Josse, Saporta)

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
DX.MCI	1	-0.815	-0.363	0.045	0.032	-0.043	-0.072
DX.CN	-0.815	1	-0.243	-0.047	0	0.067	0.003
DX.Dementia	-0.363	-0.243	1	0	-0.053	-0.035	0.116
PTRACCAT.White	0.045	-0.047	0	1	-0.562	-0.657	-0.45
PTRACCAT.Other	0.032	0	-0.053	-0.562	1	-0.031	-0.021
PTRACCAT.Black	-0.043	0.067	-0.035	-0.657	-0.031	1	-0.025
PTRACCAT.Asian	-0.072	0.003	0.116	-0.45	-0.021	-0.025	1

Correspondence Analysis

- **Just like PCA but designed for**
  - Non-normal
  - Counts
  - Ordinal & Likert
  - Lots of zeros
  - Categorical
- Generalizes PCA
  - Through the magic of Chi-squared preprocessing
- It's all you'll ever need if you know
  - But you need to know that it exists

What about everything else?

- Maybe another time?
- 100s, if not 1000s, of PCA-based or PCA-like methods
- Did not cover
  - {Distances & MDS & Clustering} and Networks
  - Discriminant/groups
  - t-SNE/UMA, some types of neural networks & some other types of other neural networks
  - Anything regarding all of the particulars of how/what to interpret
- Significance, stability, selection, and inference

Questions and Comments