Generalized Matrix Decomposition:

From Exploratory Analysis to High-Dimensional Inference

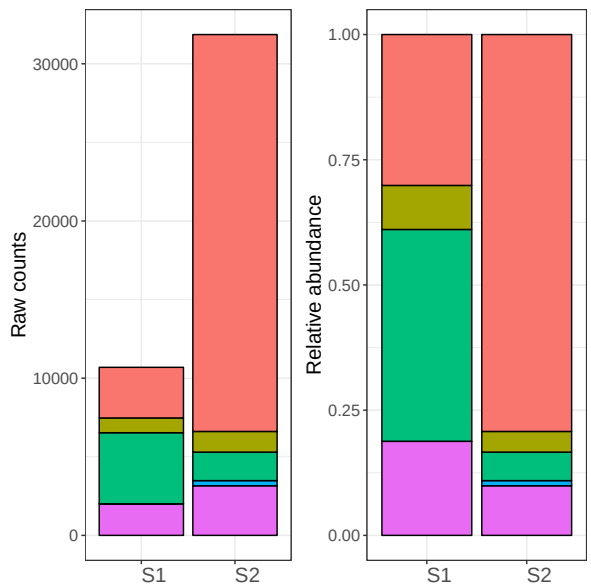
Jing Ma

Department of Statistics, Texas A&M

15 Dec 2019

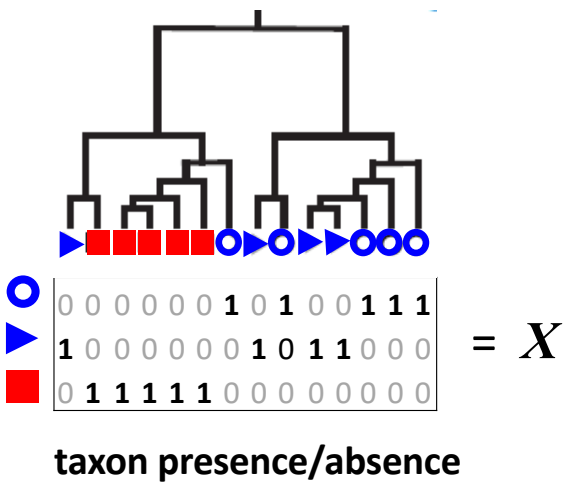
Microbiome Data

- $\boldsymbol{X} = (x_{ij})_{n \times p}$ matrix of microbiome data for n samples and p taxa
- Due to differences in effort, often work with relative abundances



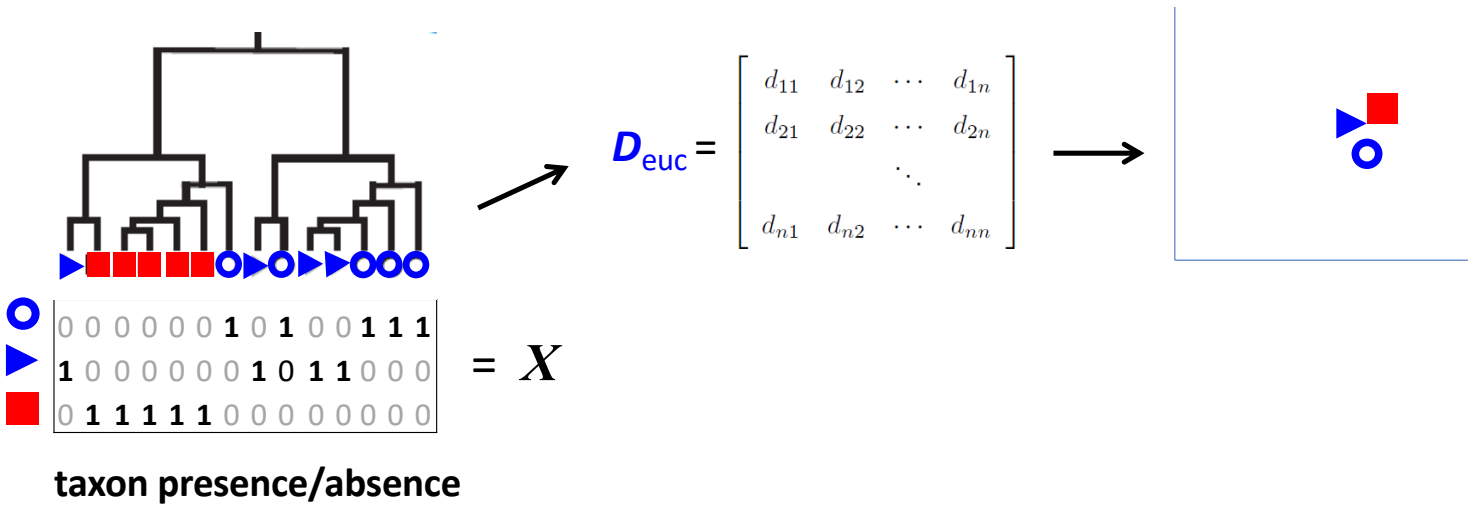
Microbiome Data

- Similarities among samples better captured by phylogenetic tree
- Many methods for capturing phylogenetic distances, e.g. UniFrac dist.



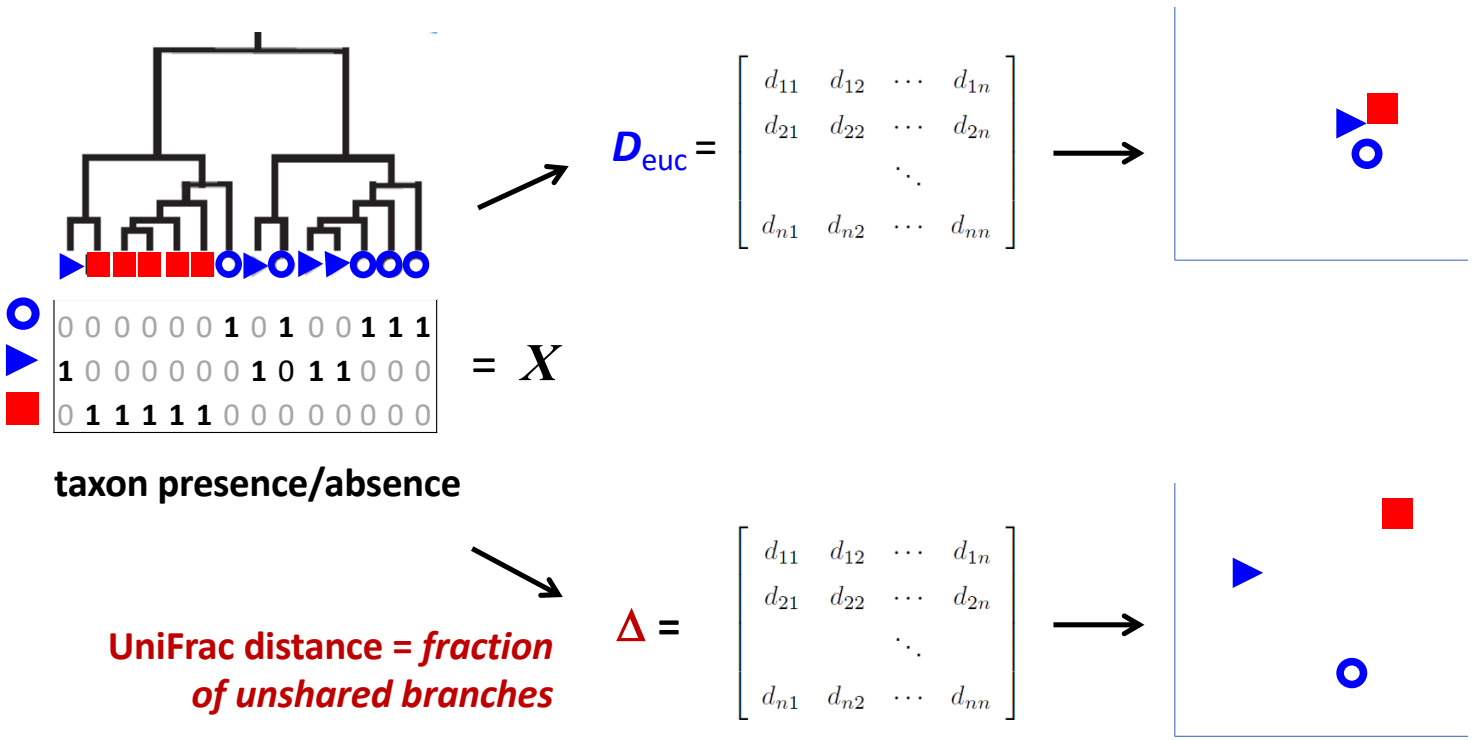
Microbiome Data

- Similarities among samples better captured by phylogenetic tree
- Many methods for capturing phylogenetic distances, e.g. UniFrac dist.



Microbiome Data

- Similarities among samples better captured by phylogenetic tree
- Many methods for capturing phylogenetic distances, e.g. UniFrac dist.

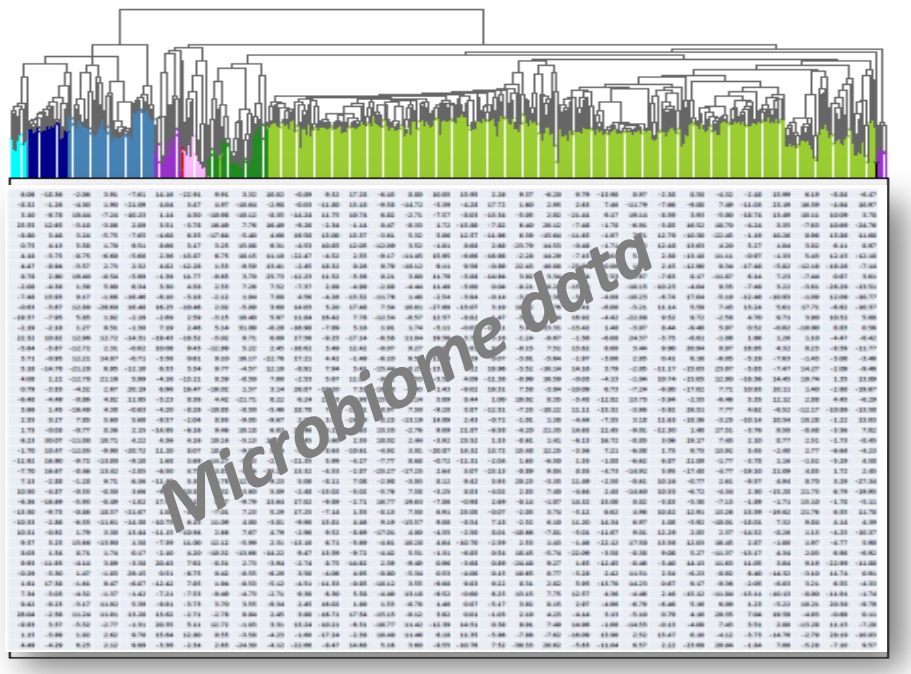


Microbiome Data

- Similarities among samples better captured by phylogenetic tree
- The phylogenetic tree also captures similarities among taxa.

Q =

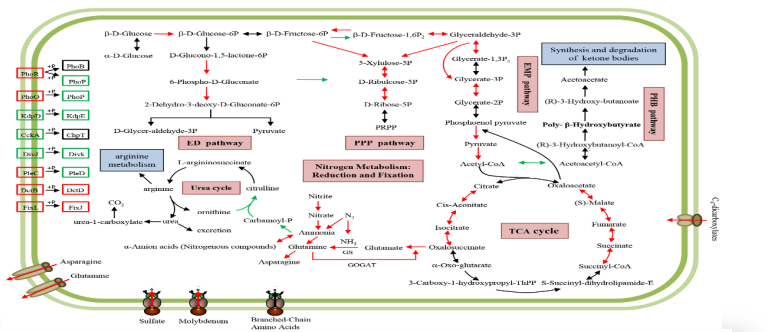
X =



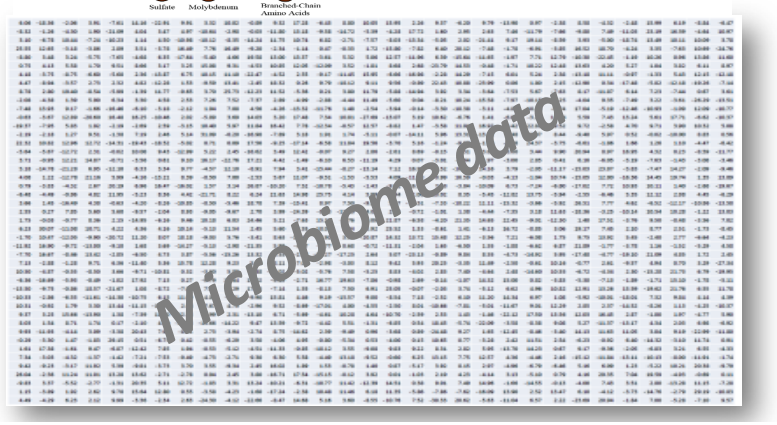
Microbiome Data

- Similarities among samples better captured by phylogenetic tree
- Alternatively, can consider information from metabolic pathways.

Q =



X =



Microbiome Data Analysis

- Need to capture
  - Similarities among samples – non-Euclidean (e.g. UniFrac distance)
  - Similarities among taxa – phylogenetic tree, pathway information, etc.

Microbiome Data Analysis

- Need to capture
  - Similarities among samples – non-Euclidean (e.g. UniFrac distance)
  - Similarities among taxa – phylogenetic tree, pathway information, etc.
- Often use exploratory data analysis tools
  - PCoA (aka MDS)
  - DPCoA (Double PCoA)

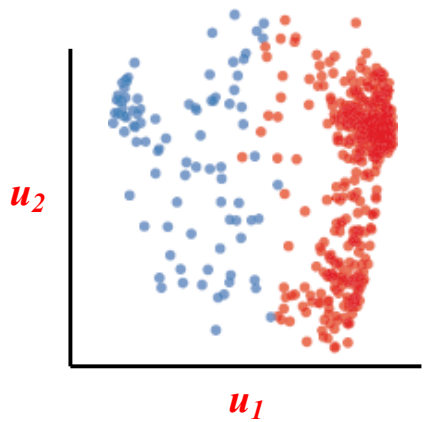
Exploratory Analysis using PCoA

First recall PCA

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ x_3^1 & x_3^2 & \cdots & x_3^p \\ & & \ddots & \\ x_n^1 & x_n^2 & \cdots & x_n^p \end{bmatrix}$$

n samples (rows)
p variables (cols)

SVD $\longrightarrow$ $X = USV^T$



Exploratory Analysis using PCoA

First recall PCA

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ x_3^1 & x_3^2 & \cdots & x_3^p \\ & & \ddots & \\ x_n^1 & x_n^2 & \cdots & x_n^p \end{bmatrix}$$

SVD $\longrightarrow$ $X = USV^T$

$\downarrow$

$\downarrow$

Euclidian distances^2

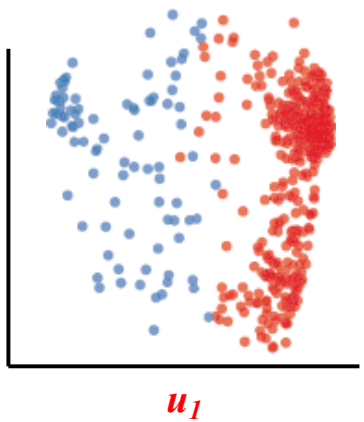
$$D_{euc} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ & & \ddots & \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

PCA $\longrightarrow$

$\mathbf{H} = -\frac{1}{2} J D J$

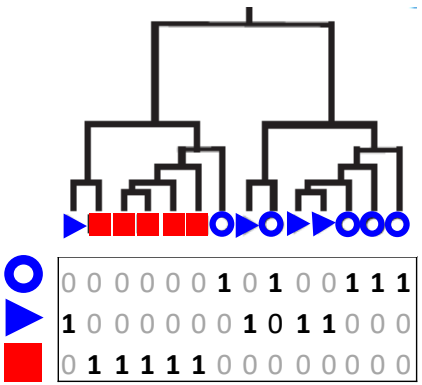
$= USU^T$

(J=centering matrix)



Exploratory Analysis using PCoA

UniFrac PCoA



$= X$

~~$X = USV^T$~~

taxon presence/absence

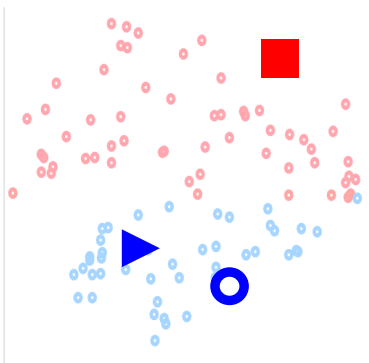
↓

distance = *fraction of unshared branches*

$$\Delta = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ & & \ddots & \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

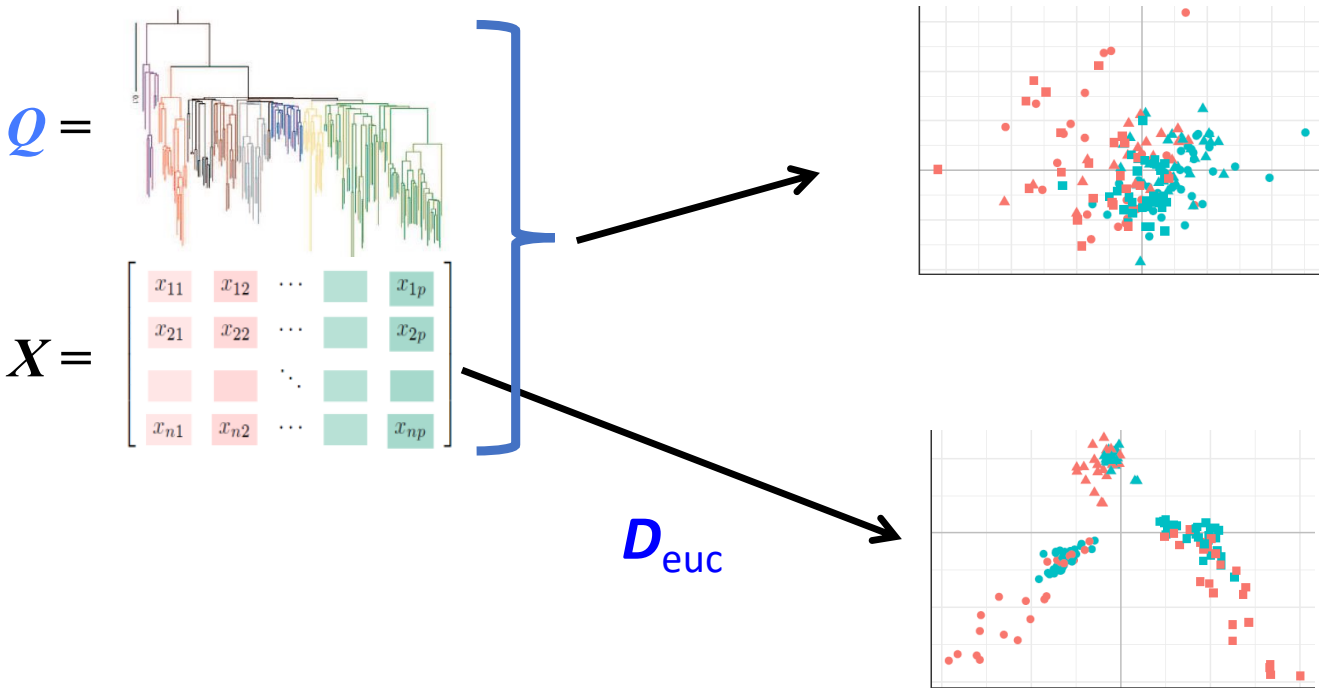
PCoA →

$$\mathbf{H} = -\tfrac{1}{2}\, J\, \Delta\, J$$



Exploratory Analysis using DPCoA

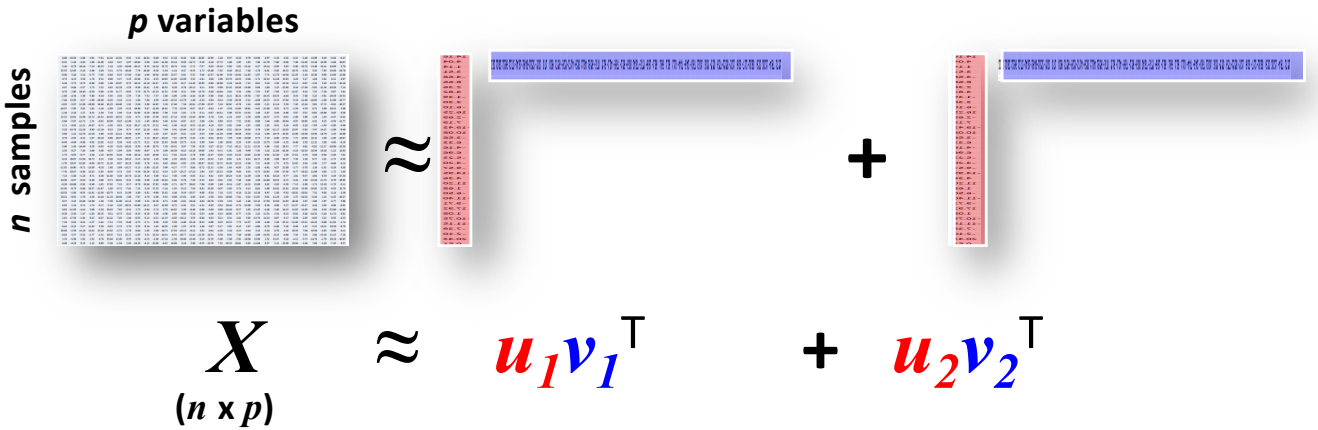
DPCoA[1] based on phylogenetic tree

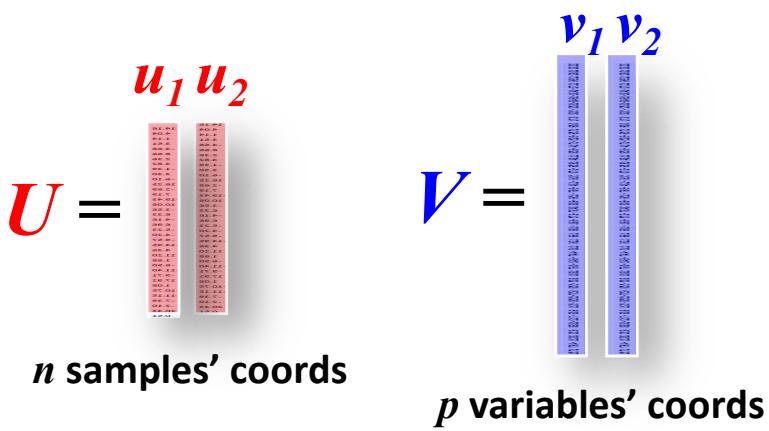


[1] Pavoine et al. 2004; Purdom. AOAS, 2011

Important Variables

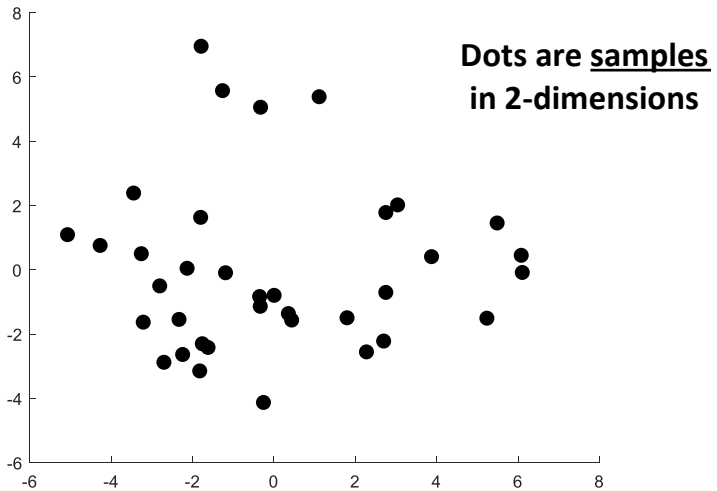
Biplots provide simultaneous visualization of samples and variables





Important Variables

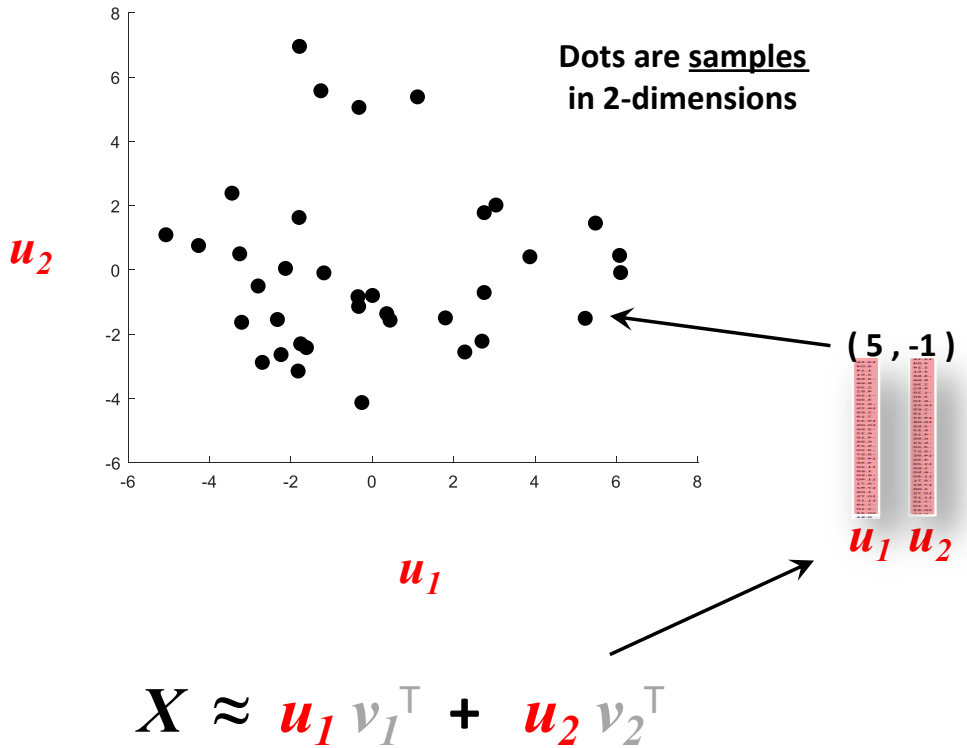
Biplots provide simultaneous visualization of samples and variables



$$X \approx u_1 v_1^{\mathsf{T}} + u_2 v_2^{\mathsf{T}}$$

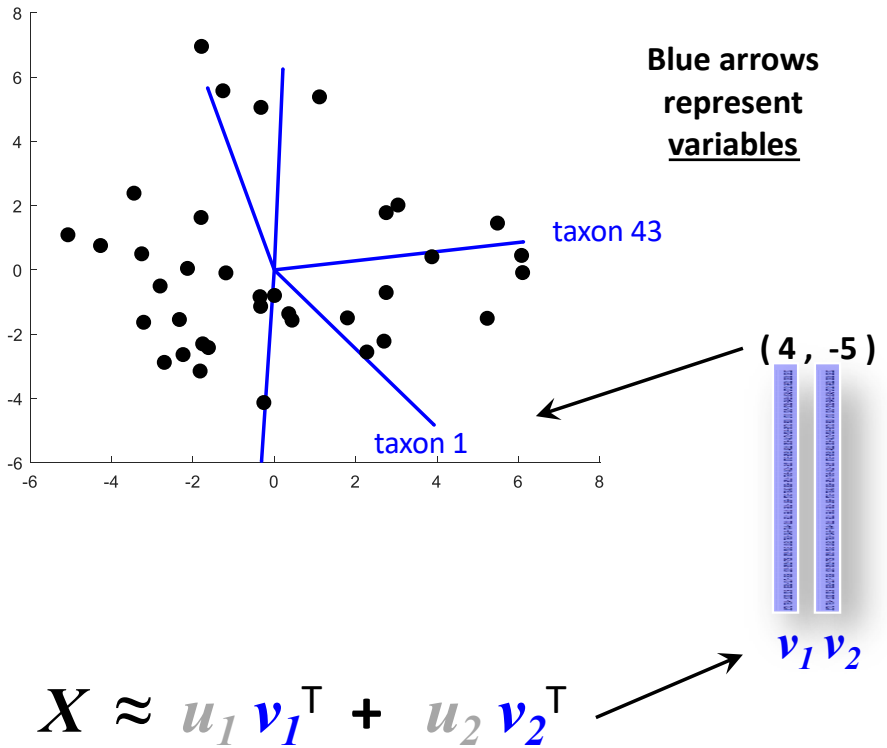
Important Variables

Biplots provide simultaneous visualization of samples and variables



Important Variables

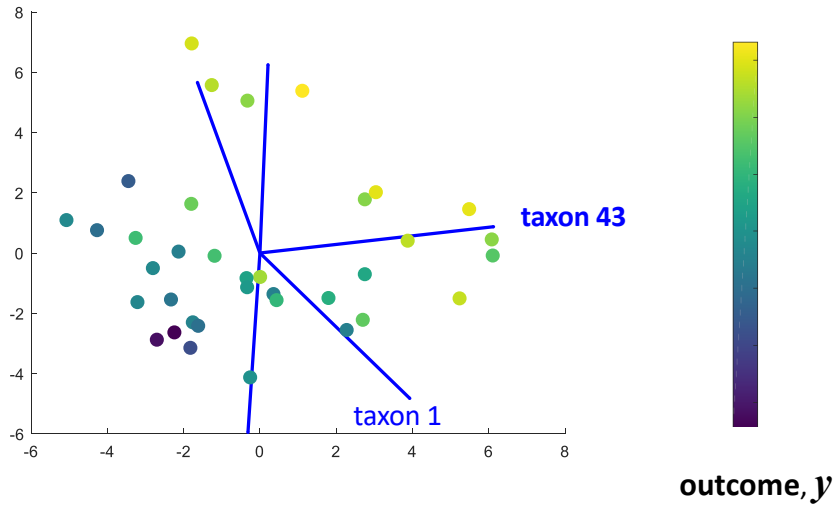
Biplots provide simultaneous visualization of samples and variables



Important Variables

Biplots provide simultaneous visualization of samples and variables

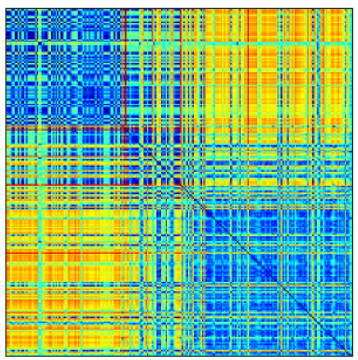
- Can also overlay outcome



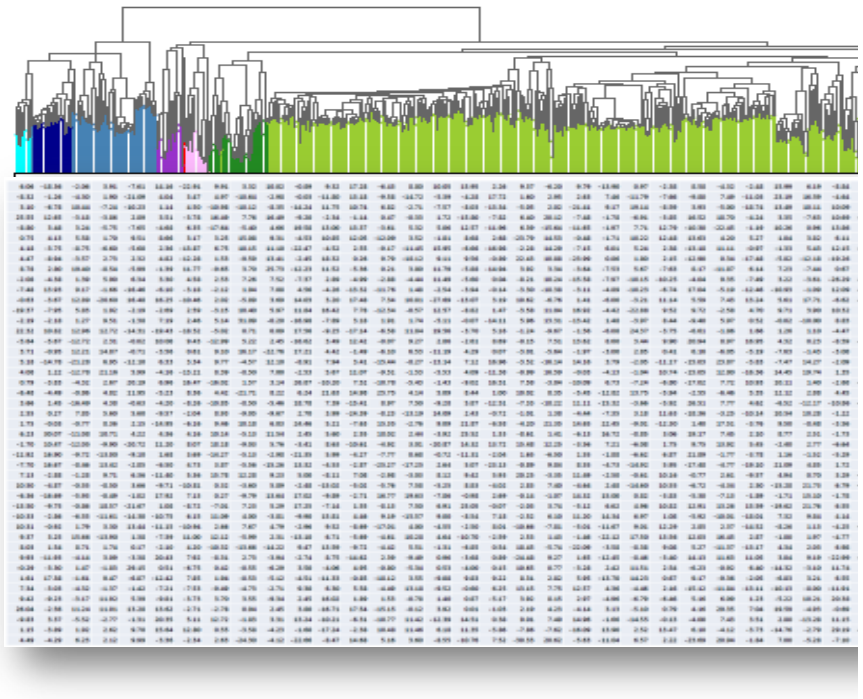
$$X \approx u_1 v_1^{\mathsf{T}} + u_2 v_2^{\mathsf{T}}$$

Microbiome Data are Doubly Structured

$\mathbf{H} = -\frac{1}{2} \boldsymbol{J} \boldsymbol{\Delta} \boldsymbol{J}$



$\boldsymbol{Q} =$



Important Variables

Can we draw a biplot that accounts for the external structures?

[2] Satten et al. PLOS One, 2017

Important Variables

Can we draw a biplot that accounts for the external structures?

- Unfortunately, there is no biplot with PCoA because
  - PCA uses $\boldsymbol{X}$ and gives: $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\intercal$ (SVD)
  - PCoA uses $\boldsymbol{\Delta}$ and only gives: $\boldsymbol{U}\boldsymbol{S}^2\boldsymbol{U}^\intercal$ (no $\boldsymbol{V}$)

[2] Satten et al. PLOS One, 2017

Important Variables

Can we draw a biplot that accounts for the external structures?

- Unfortunately, there is no biplot with PCoA because
  - PCA uses $\boldsymbol{X}$ and gives: $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\intercal$ (SVD)
  - PCoA uses $\boldsymbol{\Delta}$ and only gives: $\boldsymbol{U}\boldsymbol{S}^2\boldsymbol{U}^\intercal$ (no $\boldsymbol{V}$)
- Existing approaches[2] in the field are approximate/add hoc.

[2] Satten et al. PLOS One, 2017

Important Variables

Can we draw a biplot that accounts for the external structures?

- SVD gives $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}$ by solving

$$\underset{\boldsymbol{U},\boldsymbol{S},\boldsymbol{V}}{\arg\min} \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\|_F$$

where $\|A\|_F = \text{trace}(A^\mathsf{T}A)$.

[3] Allen et al. JASA, 2014

Important Variables

Can we draw a biplot that accounts for the external structures?

- SVD gives $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}$ by solving

$$\underset{\boldsymbol{U},\boldsymbol{S},\boldsymbol{V}}{\arg\min} \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\|_F$$

  where $\|A\|_F = \text{trace}(A^\mathsf{T}A)$.

- Consider instead a general norm to incorporate $\boldsymbol{H}$ and $\boldsymbol{Q}$:

$$\|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\|_{\boldsymbol{H},\boldsymbol{Q}}$$

  where $\|A\|_{\boldsymbol{H},\boldsymbol{Q}} = \text{trace}(A^\mathsf{T}\boldsymbol{H}A\boldsymbol{Q})$.

[3] Allen et al. JASA, 2014

Important Variables

Can we draw a biplot that accounts for the external structures?

- SVD gives $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ by solving

$$\underset{\boldsymbol{U},\boldsymbol{S},\boldsymbol{V}}{\arg\min} \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top\|_F$$

  where $\|A\|_F = \text{trace}(A^\top A)$.

- Consider instead a general norm to incorporate $\boldsymbol{H}$ and $\boldsymbol{Q}$:

$$\|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top\|_{\boldsymbol{H},\boldsymbol{Q}}$$

  where $\|A\|_{\boldsymbol{H},\boldsymbol{Q}} = \text{trace}(A^\top \boldsymbol{H} A \boldsymbol{Q})$.

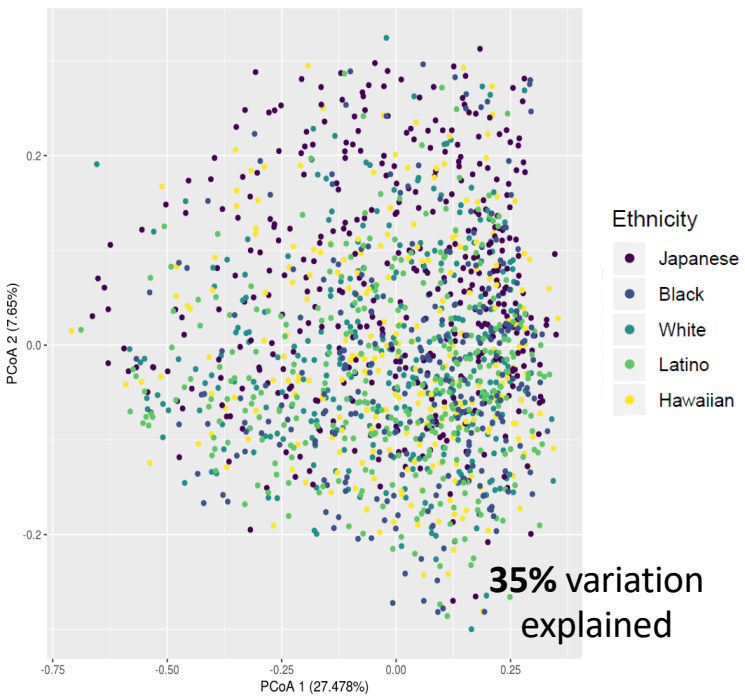
- The GMD (Gen'zd Matrix Decomp[3]) gives $\boldsymbol{X} = \mathcal{U}\mathcal{S}\mathcal{V}^\top$ such that $\mathcal{U}^\top \boldsymbol{H}\mathcal{U} = \mathcal{V}^\top \boldsymbol{Q}\mathcal{V} = I_K$, and $\mathcal{S}$ is the diagonal matrix of GMD values.

[3] Allen et al. JASA, 2014

GMD Biplot

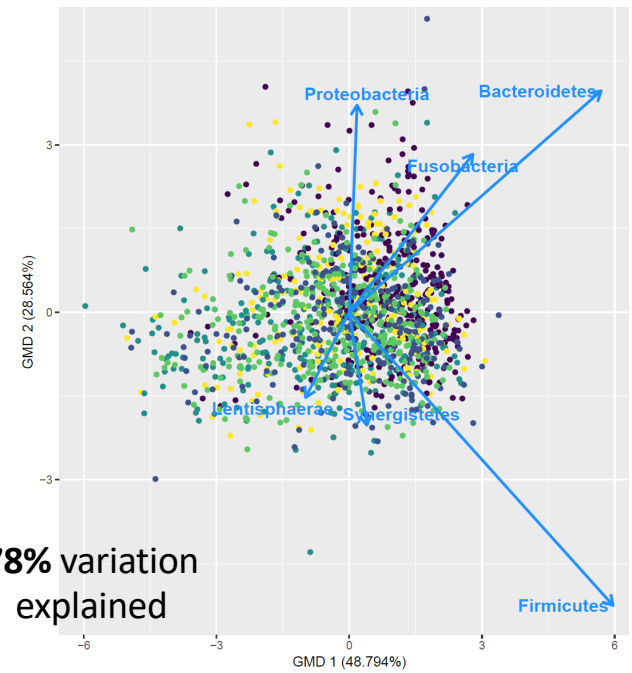
The GMD-biplot displays samples and variables using columns of $\mathcal{U}$ and $\mathcal{V}$

PCoA using UniFrac $\mathbf{\Delta}$



35% variation explained

PCoA using $\boldsymbol{X}$ and UniFrac $\mathbf{\Delta}$; phyla (arrows) from **GMD biplot**



78% variation explained

Supervised Learning with GMD

- GMD generalizes SVD for doubly structured data
- Can thus use GMD for supervised learning, similar to PCR

Supervised Learning with GMD

- GMD generalizes SVD for doubly structured data
- Can thus use GMD for supervised learning, similar to PCR
  - Suppose $y = \boldsymbol{X}\beta^* + \varepsilon$

Supervised Learning with GMD

- GMD generalizes SVD for doubly structured data
- Can thus use GMD for supervised learning, similar to PCR
  - Suppose $y = \boldsymbol{X}\beta^* + \varepsilon$
  - Let $\boldsymbol{X} = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal$ be the GMD of $\boldsymbol{X}$ w.r.t $\boldsymbol{H}$ and $\boldsymbol{Q}$

Supervised Learning with GMD

- GMD generalizes SVD for doubly structured data
- Can thus use GMD for supervised learning, similar to PCR
  - Suppose $y = \boldsymbol{X}\beta^* + \varepsilon$
  - Let $\boldsymbol{X} = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal$ be the GMD of $\boldsymbol{X}$ w.r.t $\boldsymbol{H}$ and $\boldsymbol{Q}$
  - Sample scores are given by columns of $\mathcal{U}\mathcal{S} = \boldsymbol{X}\mathcal{V}$
  - For any $\mathcal{J} \subset \{1, \ldots, \text{rank}(\boldsymbol{X})\}$, let $\Upsilon = (\mathcal{U}\mathcal{S})_{\mathcal{J}}$

Supervised Learning with GMD

- GMD generalizes SVD for doubly structured data
- Can thus use GMD for supervised learning, similar to PCR
  - Suppose $y = \boldsymbol{X}\beta^* + \varepsilon$
  - Let $\boldsymbol{X} = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal$ be the GMD of $\boldsymbol{X}$ w.r.t $\boldsymbol{H}$ and $\boldsymbol{Q}$
  - Sample scores are given by columns of $\mathcal{U}\mathcal{S} = \boldsymbol{X}\mathcal{V}$
  - For any $\mathcal{J} \subset \{1, \ldots, \text{rank}(\boldsymbol{X})\}$, let $\Upsilon = (\mathcal{U}\mathcal{S})_\mathcal{J}$
  - The GMDR estimator is

$$\hat{\gamma}(\mathcal{J}) = \arg\min \|y - \Upsilon\gamma\|^2_{\boldsymbol{H}},$$
$$\hat{\beta}_{GMDR}(\mathcal{J}) = (\boldsymbol{Q}\mathcal{V})_\mathcal{J}\hat{\gamma}(\mathcal{J})$$

GMD Regression

- Linear model $y = \boldsymbol{X}\beta^* + \varepsilon$
- Incorporating $\boldsymbol{H}$ and $\boldsymbol{Q}$

$$y = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal\beta^* + \varepsilon$$

- Coefficient $\hat{\beta}_{GMDR}(\mathcal{J}) \in \mathcal{B}_{GMD}$, where

$$\mathcal{B}_{GMD} = \{\hat{\beta}^{\mathcal{W}} = \boldsymbol{Q}\mathcal{V}\mathcal{W}\mathcal{S}^{-1}\mathcal{U}^\intercal\boldsymbol{H}y : \mathcal{W} = \text{diag}(w_1, \ldots, w_K), w_j \geq 0\}.$$

[4] Randolph et al. AOAS, 2018

GMD Regression

- Linear model $y = \boldsymbol{X}\beta^* + \varepsilon$
- Incorporating $\boldsymbol{H}$ and $\boldsymbol{Q}$

$$y = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal\beta^* + \varepsilon$$

- Coefficient $\hat{\beta}_{GMDR}(\mathcal{J}) \in \mathcal{B}_{GMD}$, where

$$\mathcal{B}_{GMD} = \{\hat{\beta}^{\mathcal{W}} = \boldsymbol{Q}\mathcal{V}\mathcal{W}\mathcal{S}^{-1}\mathcal{U}^\intercal\boldsymbol{H}y : \mathcal{W} = \text{diag}(w_1, \ldots, w_K), w_j \geq 0\}.$$

Two examples of the weight matrix $\mathcal{W}$ are:

- $\mathcal{W}_j = \mathbf{1}_{j\in\mathcal{J}} \rightarrow \hat{\beta}_{GMDR}(\mathcal{J})$,

[4] Randolph et al. AOAS, 2018

GMD Regression

- Linear model $y = \boldsymbol{X}\beta^* + \varepsilon$
- Incorporating $\boldsymbol{H}$ and $\boldsymbol{Q}$

$$y = \mathcal{U}\mathcal{S}\mathcal{V}^\intercal\beta^* + \varepsilon$$

- Coefficient $\hat{\beta}_{GMDR}(\mathcal{J}) \in \mathcal{B}_{GMD}$, where

$$\mathcal{B}_{GMD} = \{\hat{\beta}^{\mathcal{W}} = \boldsymbol{Q}\mathcal{V}\mathcal{W}\mathcal{S}^{-1}\mathcal{U}^\intercal\boldsymbol{H}y : \mathcal{W} = \text{diag}(w_1, \ldots, w_K), w_j \geq 0\}.$$

Two examples of the weight matrix $\mathcal{W}$ are:

- $\mathcal{W}_j = \mathbf{1}_{j\in\mathcal{J}} \rightarrow \hat{\beta}_{GMDR}(\mathcal{J})$,
- $\mathcal{W} = \mathcal{S}^2(\mathcal{S}^2 + \lambda I_n)^{-1} \rightarrow \hat{\beta}_{KPR} = \arg\min_\beta\{\|y - \boldsymbol{X}\beta\|^2_{\boldsymbol{H}} + \lambda\|\beta\|^2_{\boldsymbol{Q}^{-1}}\}$[4]

[4] Randolph et al. AOAS, 2018

GMD Inference

- GMDR is good for prediction, but which covariates are associated with the outcome?

[5] Bühlmann. 2013; Zhao & Shojaie, 2016

GMD Inference

- GMDR is good for prediction, but which covariates are associated with the outcome?
- For $\hat{\beta}^{\mathcal{W}} \in \mathcal{B}_{GMD}$,

$$\mathbb{E}[\hat{\beta}^{\mathcal{W}} - \beta^*] = \underbrace{\boldsymbol{Q}\mathcal{V}(\mathcal{W} - \boldsymbol{I})\mathcal{V}^\intercal\beta^*}_{\text{estimation bias}} + \underbrace{\theta^* - \beta^*}_{\text{projection bias}^5}$$

[5] Bühlmann. 2013; Zhao & Shojaie, 2016

GMD Inference

- GMDR is good for prediction, but which covariates are associated with the outcome?
- For $\hat{\beta}^{\mathcal{W}} \in \mathcal{B}_{GMD}$,

$$\mathbb{E}[\hat{\beta}^{\mathcal{W}} - \beta^*] = \underbrace{\boldsymbol{Q}\mathcal{V}(\mathcal{W} - \boldsymbol{I})\mathcal{V}^\intercal\beta^*}_{\text{estimation bias}} + \underbrace{\theta^* - \beta^*}_{\text{projection bias}^5}$$

- Correct the estimation and projection bias via an initial estimator ($\boldsymbol{Q} = D\Delta D^\intercal$)

$$\tilde{\beta}(\lambda) = \arg\min_{\beta}\{\|y - \boldsymbol{X}D\beta\|_{\boldsymbol{H}}^2 + \lambda\|\Delta^{-1/2}\beta\|_1\}, \quad \hat{\beta}^{init} = D\tilde{\beta}(\lambda)$$

[5] Bühlmann. 2013; Zhao & Shojaie, 2016

GMD Inference

- GMDR is good for prediction, but which covariates are associated with the outcome?
- For $\hat{\beta}^{\mathcal{W}} \in \mathcal{B}_{GMD}$,

$$\mathbb{E}[\hat{\beta}^{\mathcal{W}} - \beta^*] = \underbrace{\boldsymbol{Q}\mathcal{V}(\mathcal{W} - \boldsymbol{I})\mathcal{V}^\intercal\beta^*}_{\text{estimation bias}} + \underbrace{\theta^* - \beta^*}_{\text{projection bias}^5}$$

- Correct the estimation and projection bias via an initial estimator ($\boldsymbol{Q} = D\Delta D^\intercal$)

$$\tilde{\beta}(\lambda) = \arg\min_{\beta}\{\|y - \boldsymbol{X}D\beta\|_{\boldsymbol{H}}^2 + \lambda\|\Delta^{-1/2}\beta\|_1\}, \quad \hat{\beta}^{init} = D\tilde{\beta}(\lambda)$$

- Obtain inference for $H_{0,j} : \beta_j^* = 0$

[5] Bühlmann. 2013; Zhao & Shojaie, 2016

GMD Inference

- We need to make the following assumptions:
  - A compatibility assumption w.r.t $\boldsymbol{Q}$ and $\boldsymbol{H}$
  - β^* is $\boldsymbol{Q}$-smooth:

$$\|\boldsymbol{Q}^{-1/2}\beta^*\|_0 = o\{(n/\log p)^r\} \text{ for } r \in (0, 1/2).$$

GMD Inference

- We need to make the following assumptions:
  - A compatibility assumption w.r.t $\boldsymbol{Q}$ and $\boldsymbol{H}$
  - β^* is $\boldsymbol{Q}$-smooth:

$$\|\boldsymbol{Q}^{-1/2}\beta^*\|_0 = o\{(n/\log p)^r\} \text{ for } r \in (0, 1/2).$$

- However, no sparsity assumption on β^*

GMD Inference

- We need to make the following assumptions:
  - A compatibility assumption w.r.t $\boldsymbol{Q}$ and $\boldsymbol{H}$
  - β^* is $\boldsymbol{Q}$-smooth:

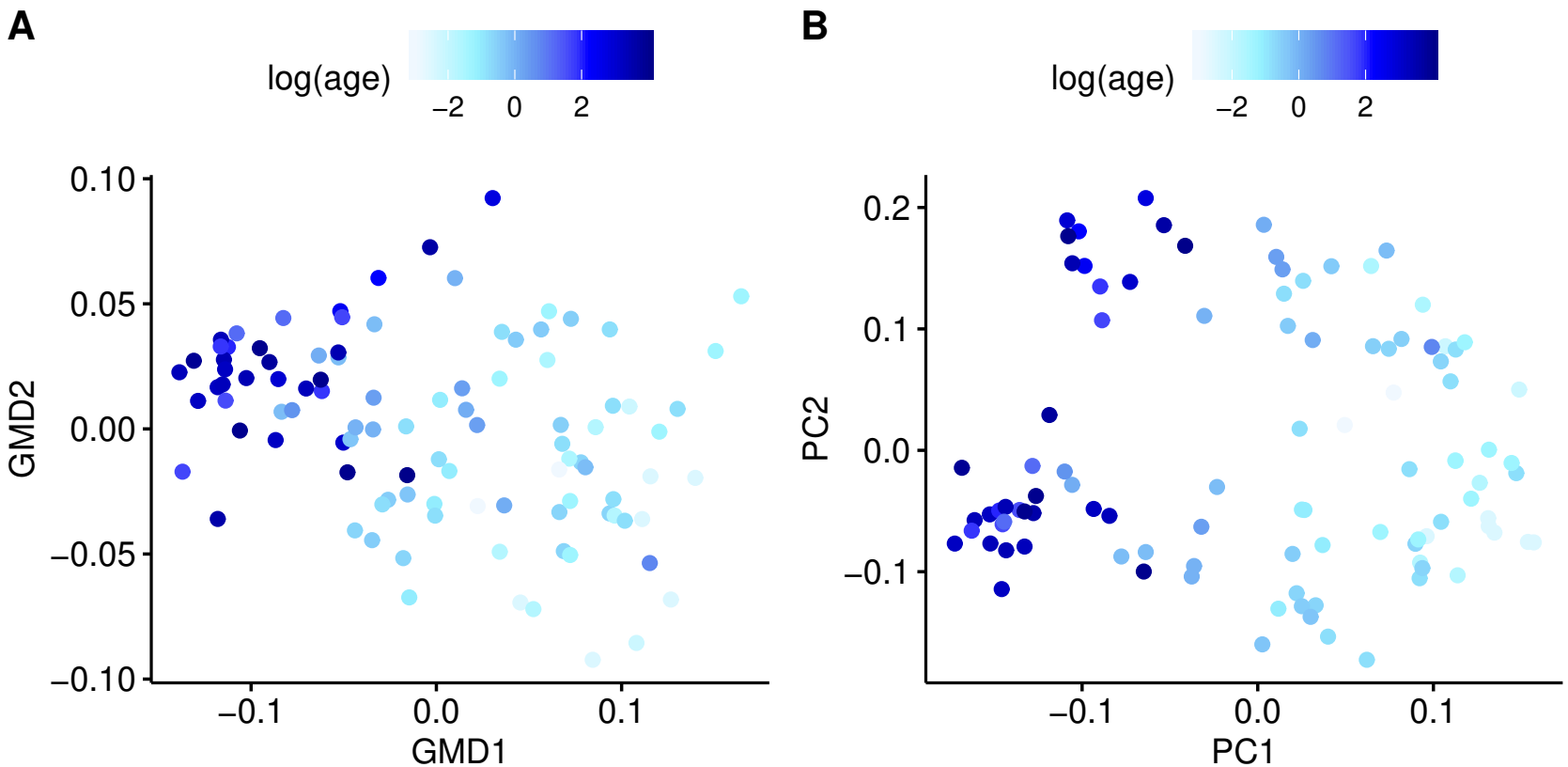
$$\|\boldsymbol{Q}^{-1/2}\beta^*\|_0 = o\{(n/\log p)^r\} \text{ for } r \in (0, 1/2).$$

- However, no sparsity assumption on β^*
- Can characterize the detection level of the test under the alternative

Application to Microbiome Data

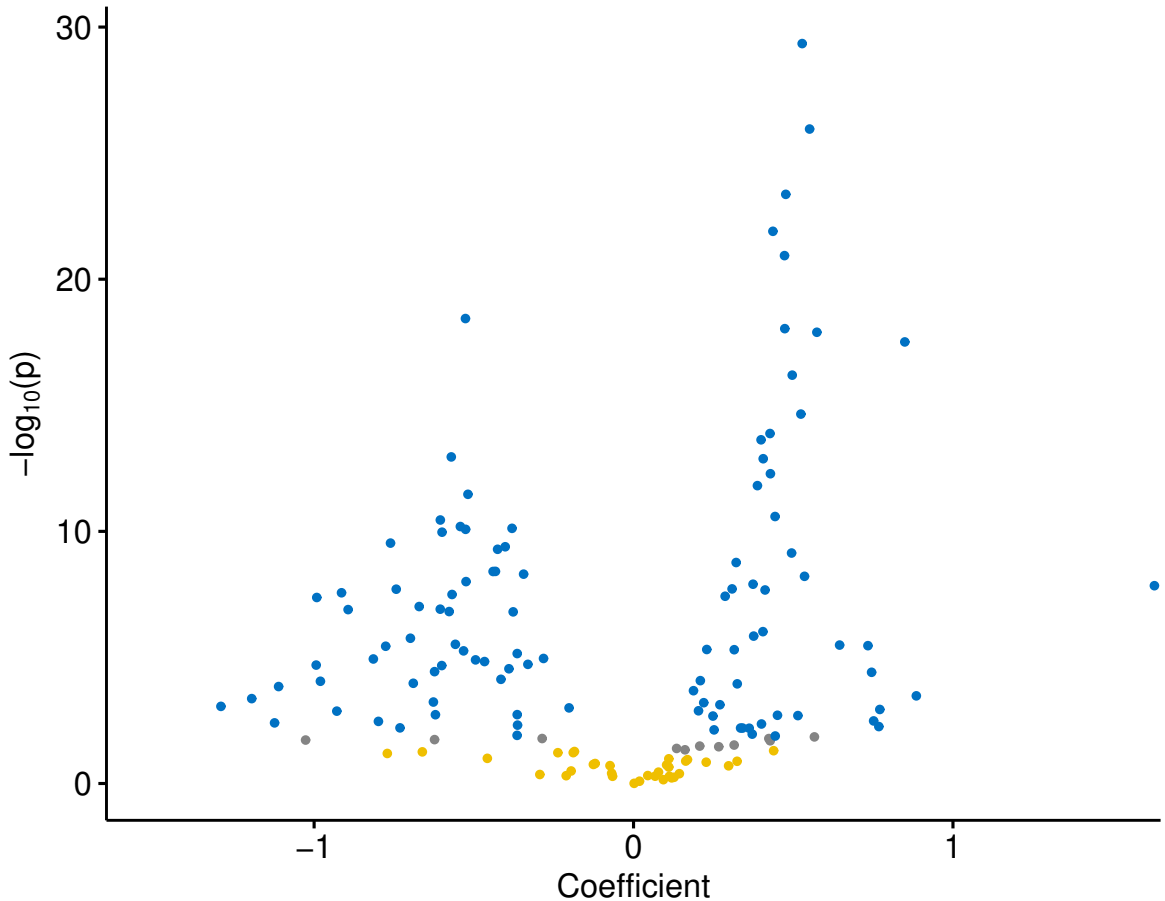
- Data from Yatsunenko et al. Nature. (2012)
- $p = 149$ and $n = 100$
- $\boldsymbol{Q}$ is derived from the patristic distance between each pair of the tips of the phylogenetic tree
- $\boldsymbol{H}$ is derived from Enzyme Commision (EC) numbers.

GMD Improves Visualization of Samples



Which Bacteria are Associated with Age?

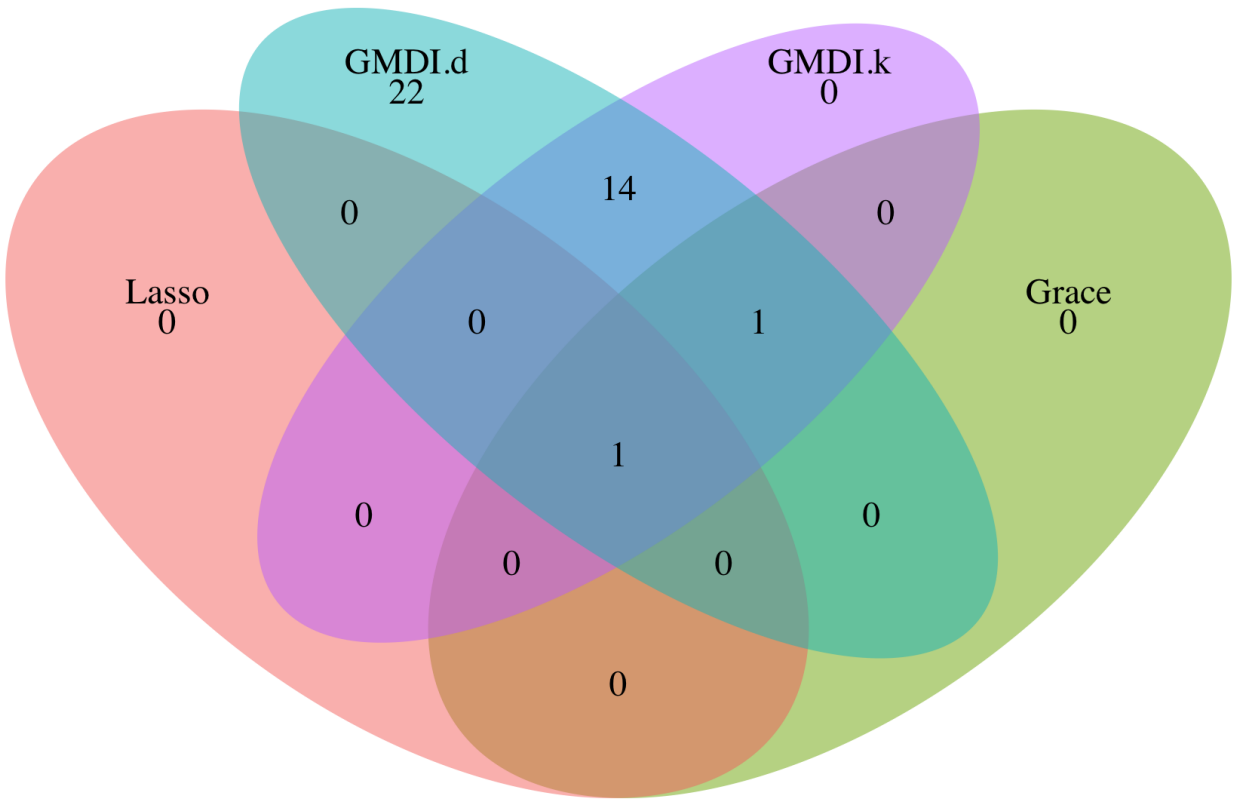
- Many significant marginal associations[6] (FDR=0.1)



[6] Results from Yatsunenko et al. Nature. (2012)

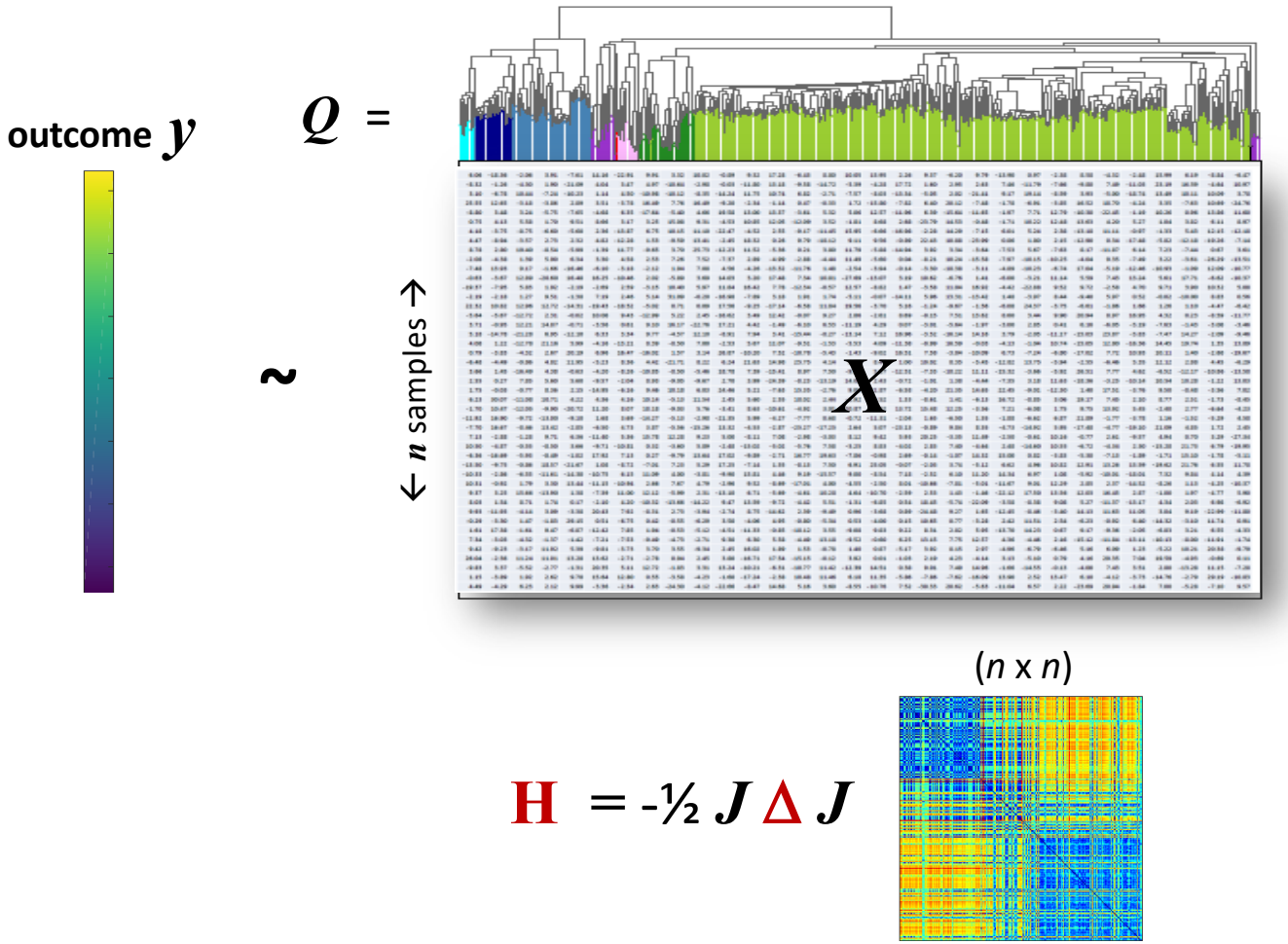
Which Bacteria are Associated with Age?

- Significant associations from multivariate methods[7] (FDR=0.1)



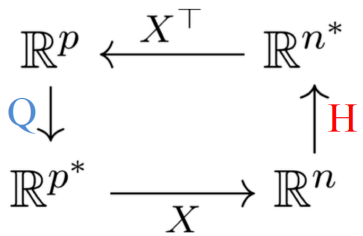
[7] Ridge test by Bühlmann (2013) returns 0 sig association.

Summary

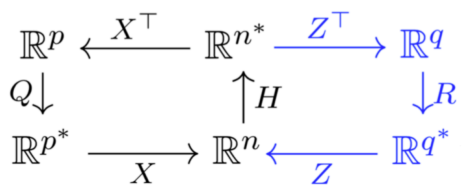


Summary

- Discussed visualization and regression based on GMD
- GMD is closely related to the duality[8] between viewing the data from the perspective of samples and variables



- Our framework
  - encompasses classical methods, both unsupervised (PCA, PCoA/MDS, biplots) and supervised (ridge, GLS) methods
  - extends them to non-standard settings (multi-view data)



[8] Escoufier. 1977; de la Cruz & Holmes. AOAS, 2011

Aknowledgement

Yue Wang

Ali Shojaie

Tim Randolph

- The GMD-biplot and its application to microbiome data. *mSystems*. 2019
- Generalized matrix decomposition: estimation and inference for two-way structured data. 2019+

Thank You!

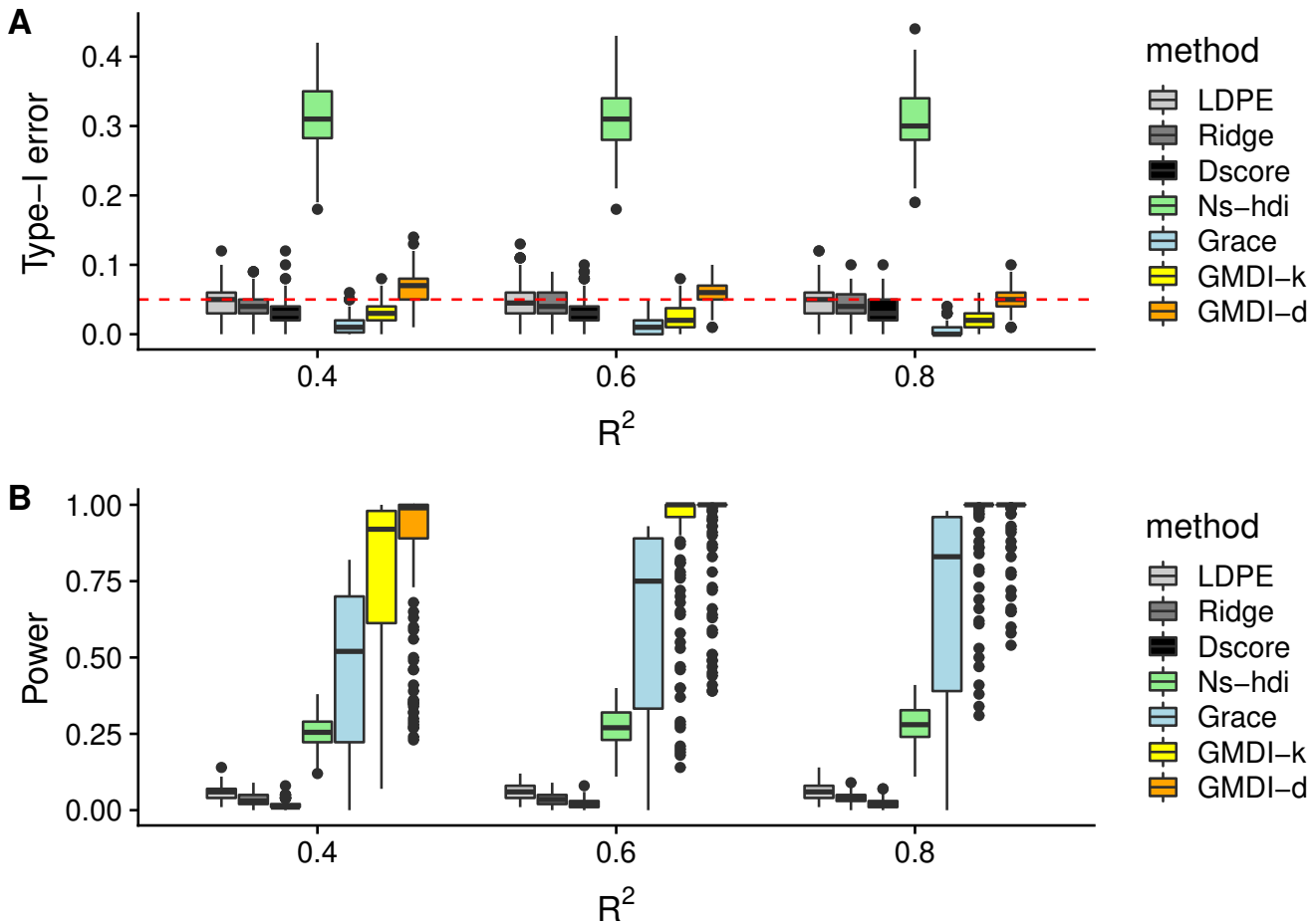
Simulation Examples

- Data generated from linear model with $p = 300$ & $n = 200$
- Compare GMDI-k and GMDI-d with
  - Low-dimension Projection Estimator (LDPE)[9]
  - Ridge-based inference[10]
  - Decorrelated score test (Dscore)[11]
  - Non-sparse high-dimensional inference (Ns-hdi)[12]
  - Grace test[13]

[9] van de Geer et al. 2014; Zhang & Zhang. 2014
[10] Bühlmann. 2013
[11] Ning & Liu. 2017
[12] Zhu & Bradic. 2018
[13] Zhao & Shojaie. 2015

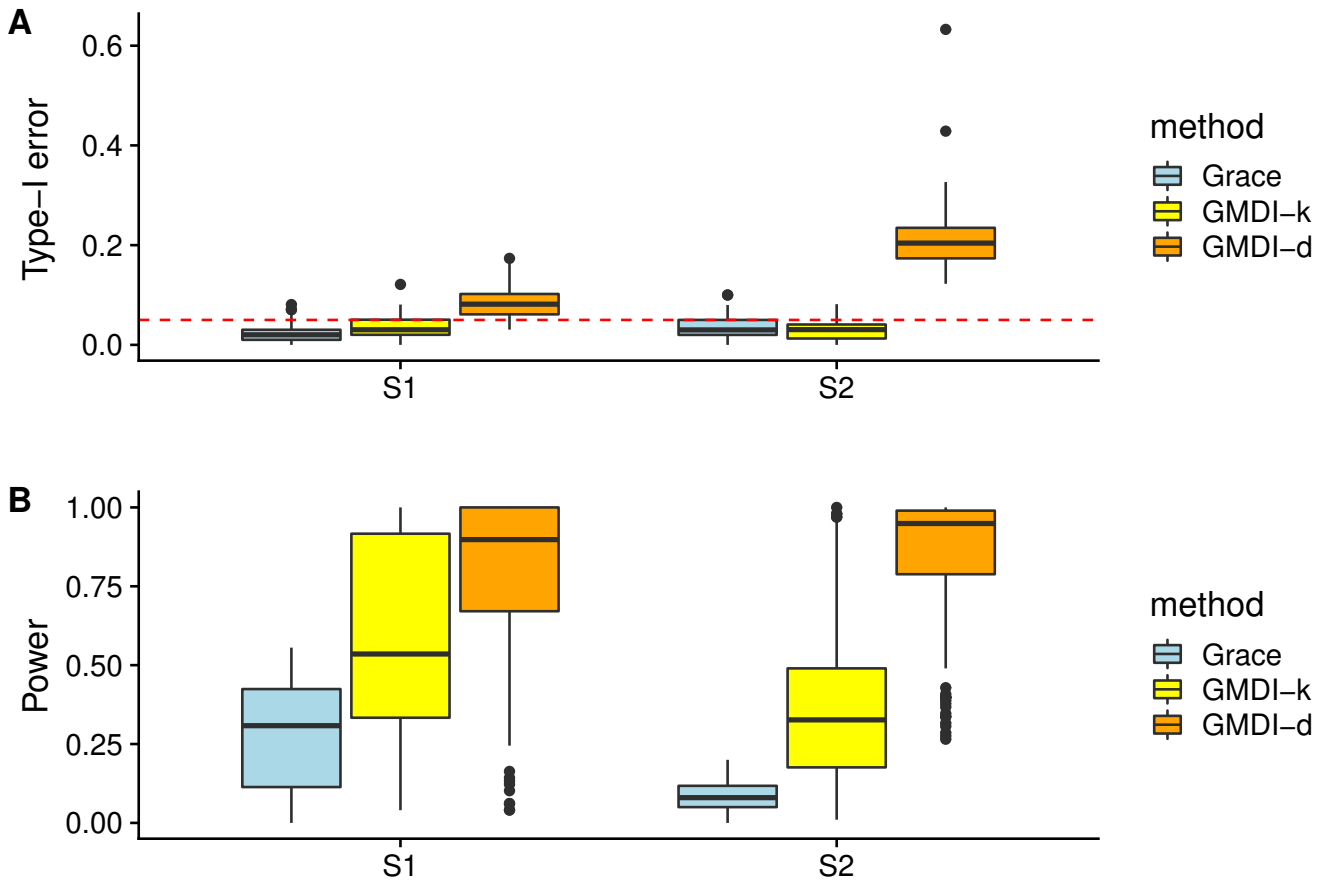
Simulation Examples

- Setting 1: $\boldsymbol{H} = \mathbf{I}$, $\|\boldsymbol{Q}^{-1/2}\beta^*\|_0 = 10$ but β^* is not sparse.



Simulation Examples

- Setting 2: similar to Setting 1, but with Q perturbed.



Simulation Examples

- Setting 3: $\boldsymbol{H}$ is block diagonal, $\boldsymbol{Q}$ is the same as in setting 1.