

BIOT: An R Package for Explaining the Axes of MDS Visualizations

Rebecca Marion*

NADI, TRAIL

Université de Namur

Adrien Bibal

CENTAL

Université catholique de Louvain

Rainer von Sachs

ISBA, LIDAM

Université catholique de Louvain

Benoît Frénay

NADI, TRAIL

Université de Namur

5 April, 2022



*Research completed as a part of the ARIAC project funded by the Walloon Region.

Table of Contents

Context

Example Research Problem

Motivation for BIOT

Proposed Solution

BIOT Method

BIOT Package

Example Problem from Psychology

Example Problem from Psychology

Research Question: What are the major **attributes** people use to **judge the similarity/difference** between social groups?

Example Problem from Psychology

Research Question: What are the major **attributes** people use to **judge the similarity/difference** between social groups?

Example: why are **artists** and **conservatives** seen as different?

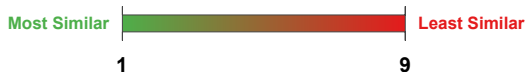
Example Problem from Psychology

Research Question: What are the major **attributes** people use to **judge the similarity/difference** between social groups?

Example: why are **artists** and **conservatives** seen as different?

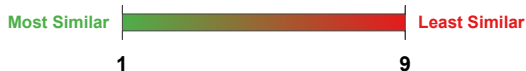
Example Problem from Psychology

Data set up: ask participants to rate the similarity of pairs of social groups



Example Problem from Psychology

Data set up: ask participants to rate the similarity of pairs of social groups



	artists	conservatives	educated	sportsfans	students	...
artists	1	6	4	7	4	...
conservatives	6	1	4	5	6	...
educated	4	4	1	5	2	...
sportsfans	7	5	5	1	3	...
students	4	6	2	3	1	...
...

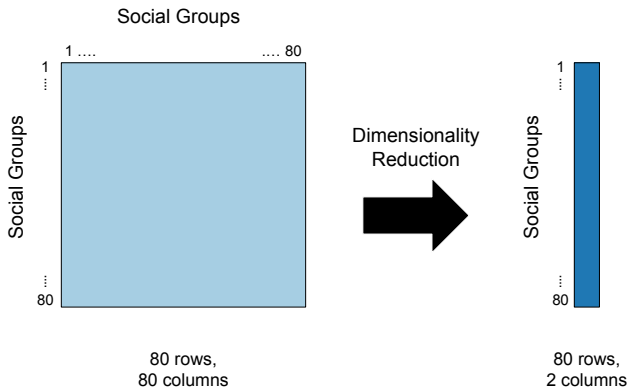
Example Problem from Psychology

Problem: Difficult to analyze (too many columns for visualization)

Example Problem from Psychology

Problem: Difficult to analyze (too many columns for visualization)

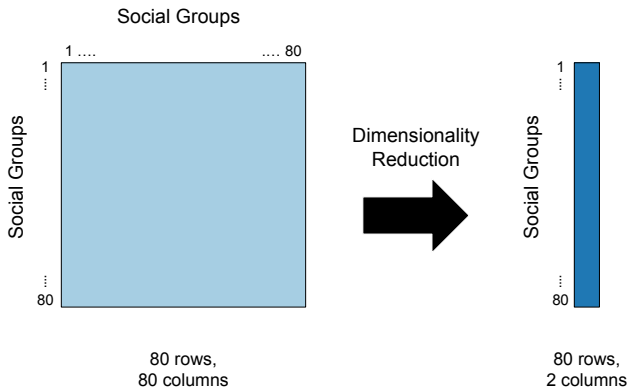
Solution: Dimensionality reduction



Example Problem from Psychology

Problem: Difficult to analyze (too many columns for visualization)

Solution: Dimensionality reduction



★ Now the data can be visualized in a 2D plot!

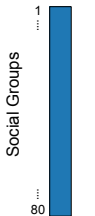
Example Problem from Psychology

In psychology, **Multidimensional Scaling (MDS)** is often used to perform dimensionality reduction

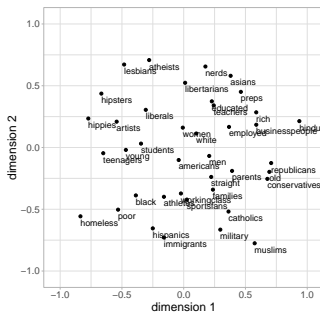
Example Problem from Psychology

In psychology, **Multidimensional Scaling (MDS)** is often used to perform **dimensionality reduction**

Embedding



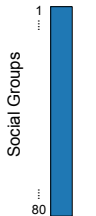
2D Scatter Plot



Example Problem from Psychology

In psychology, **Multidimensional Scaling (MDS)** is often used to perform dimensionality reduction

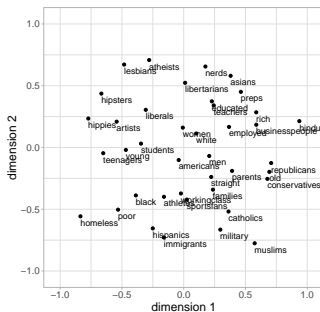
Embedding



80 rows,
2 columns



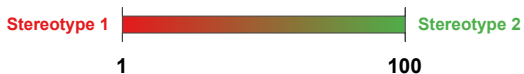
2D Scatter Plot



Problem: The axes have no pre-defined meaning

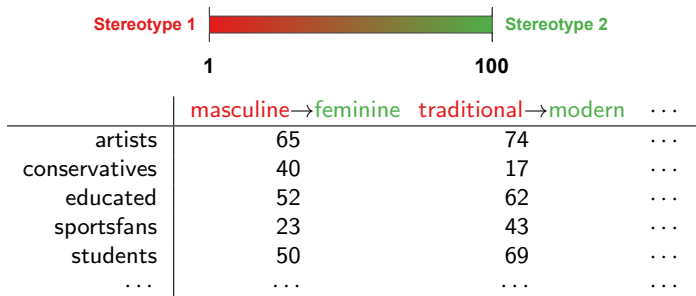
Example Problem from Psychology

Second dataset: social groups are scored according to different attributes (e.g. stereotypes)



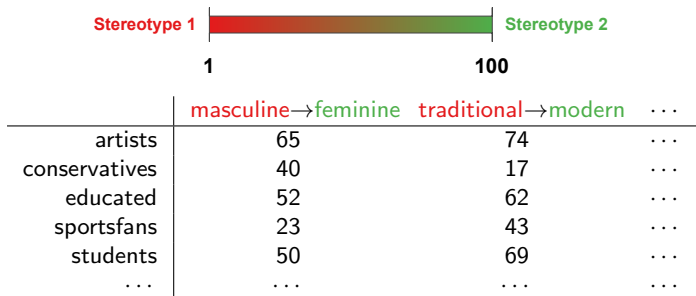
Example Problem from Psychology

Second dataset: social groups are scored according to different attributes (e.g. stereotypes)



Example Problem from Psychology

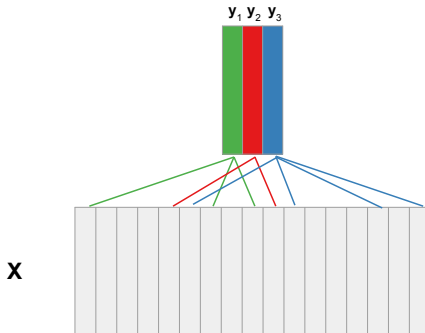
Second dataset: social groups are scored according to different attributes (e.g. stereotypes)



Objective: Identify attributes (e.g. stereotypes) that explain each embedding dimension

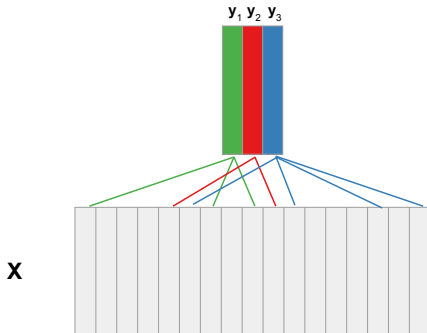
Example Problem from Psychology

Potential Strategy: Use sparse regression models to identify important attributes for predicting each embedding dimension



Example Problem from Psychology

Potential Strategy: Use sparse regression models to identify important attributes for predicting each embedding dimension



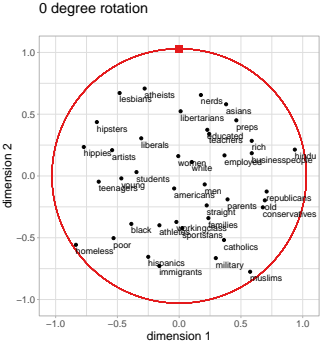
★ Attributes selected for the same dimension **give meaning** to the dimension

Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary

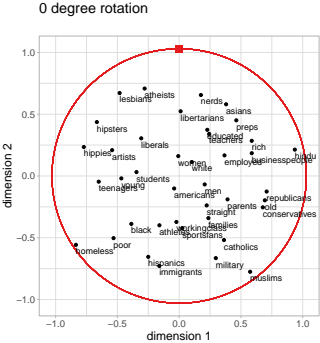
Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary



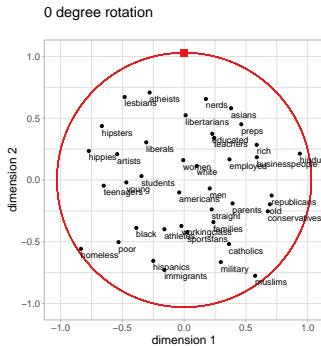
Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary



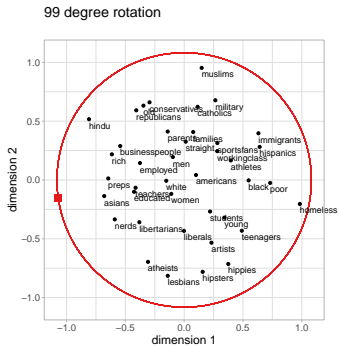
Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary



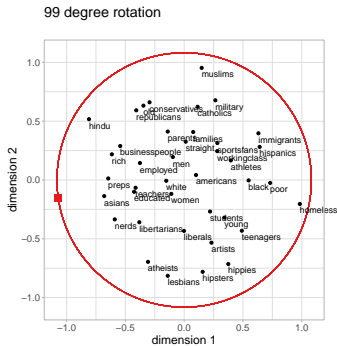
Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary



Rotation Problem

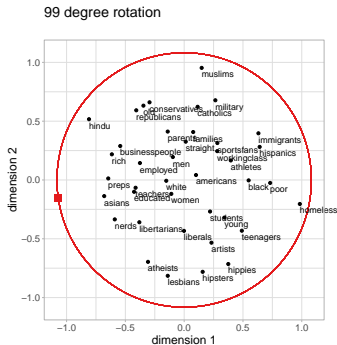
Problem: The rotation of MDS embeddings is arbitrary



- ▶ Both embeddings are **equally faithful representations** of the original dataset

Rotation Problem

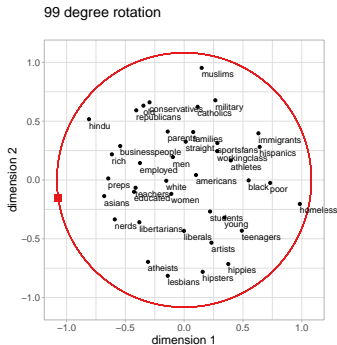
Problem: The rotation of MDS embeddings is arbitrary



- ▶ Both embeddings are **equally faithful representations** of the original dataset
- ▶ The **error and interpretability** of sparse regression models are **not equal**

Rotation Problem

Problem: The rotation of MDS embeddings is arbitrary



- ▶ Both embeddings are **equally faithful representations** of the original dataset
- ▶ The **error and interpretability** of sparse regression models are **not equal**
 - ▶ **Need to find the best rotation of the embedding**

Table of Contents

Context

Example Research Problem

Motivation for BIOT

Proposed Solution

BIOT Method

BIOT Package

Proposed Method: Best Interpretable Orthogonal Transformation (BIOT)



Bibal, Adrien, Rebecca Marion, Rainer von Sachs, and Benoît Frénay (2021). "BIOT: Explaining Multidimensional Nonlinear MDS Embeddings using the Best Interpretable Orthogonal Transformation". In: *Neurocomputing*.

Proposed Method: Best Interpretable Orthogonal Transformation (BIOT)



Bibal, Adrien, Rebecca Marion, Rainer von Sachs, and Benoît Frénay (2021). "BIOT: Explaining Multidimensional Nonlinear MDS Embeddings using the Best Interpretable Orthogonal Transformation". In: *Neurocomputing*.

Main Idea:

Proposed Method: Best Interpretable Orthogonal Transformation (BIOT)



Bibal, Adrien, Rebecca Marion, Rainer von Sachs, and Benoît Frénay (2021). "BIOT: Explaining Multidimensional Nonlinear MDS Embeddings using the Best Interpretable Orthogonal Transformation". In: *Neurocomputing*.

Main Idea:

Given an MDS embedding and a dataset of attributes,

Proposed Method: Best Interpretable Orthogonal Transformation (BIOT)



Bibal, Adrien, Rebecca Marion, Rainer von Sachs, and Benoît Frénay (2021). "BIOT: Explaining Multidimensional Nonlinear MDS Embeddings using the Best Interpretable Orthogonal Transformation". In: *Neurocomputing*.

Main Idea:

Given an MDS embedding and a dataset of attributes,

- ▶ find the best **orthogonal transformation** of the embedding such that the **sparse linear model** used to explain the embedding has low error and low model complexity (i.e. number of attributes used)

Proposed Method: Best Interpretable Orthogonal Transformation (BIOT)



Bibal, Adrien, Rebecca Marion, Rainer von Sachs, and Benoît Frénay (2021). "BIOT: Explaining Multidimensional Nonlinear MDS Embeddings using the Best Interpretable Orthogonal Transformation". In: *Neurocomputing*.

Main Idea:

Given an MDS embedding and a dataset of attributes,

- ▶ find the best **orthogonal transformation** of the embedding such that the **sparse linear model** used to explain the embedding has low error and low model complexity (i.e. number of attributes used)
 - ▶ Relative importance of error vs. complexity = defined by the user (sparsity hyperparameter)

What can you do with the BIOT package?

What can you do with the BIOT package?

- ▶ Train an explanation model

What can you do with the BIOT package?

- ▶ Train an explanation model
- ▶ Evaluate performance for different sparsity hyperparameters

What can you do with the BIOT package?

- ▶ Train an explanation model
- ▶ Evaluate performance for different sparsity hyperparameters
- ▶ Tune the sparsity hyperparameter via k -fold cross-validation

What can you do with the BIOT package?

- ▶ Train an explanation model
- ▶ Evaluate performance for different sparsity hyperparameters
- ▶ Tune the sparsity hyperparameter via k -fold cross-validation
- ▶ Compare performance for different embedding sizes

What can you do with the BIOT package?

- ▶ Train an explanation model
- ▶ Evaluate performance for different sparsity hyperparameters
- ▶ Tune the sparsity hyperparameter via k -fold cross-validation
- ▶ Compare performance for different embedding sizes
- ▶ Display the explanation model in the MDS scatterplot

What can you do with the BIOT package?

- ▶ Train an explanation model
- ▶ Evaluate performance for different sparsity hyperparameters
- ▶ Tune the sparsity hyperparameter via k -fold cross-validation
- ▶ **Compare performance for different embedding sizes**
- ▶ **Display the explanation model in the MDS scatterplot**

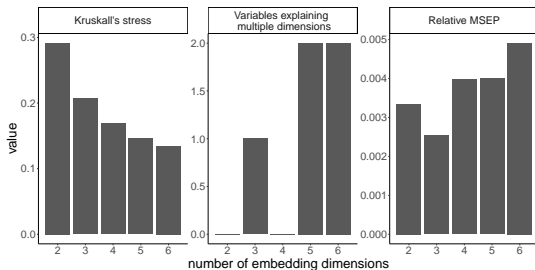
BIOT Package: Compare performance for different embedding sizes

BIOT Package: Compare performance for different embedding sizes

- ▶ `PlotExplanationStats()` → choose the number of embedding dimensions

BIOT Package: Compare performance for different embedding sizes

- ▶ `PlotExplanationStats()` → choose the number of embedding dimensions



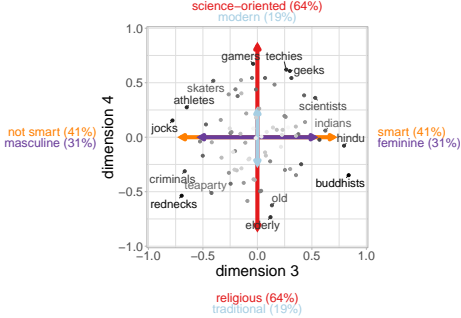
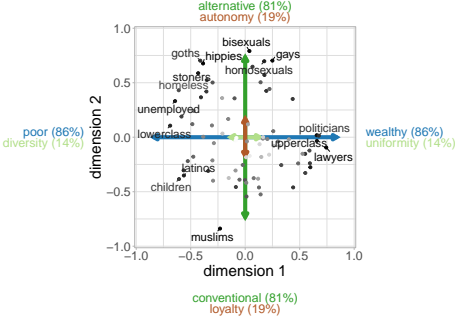
BIOT Package: Display the explanation model

BIOT Package: Display the explanation model

- ▶ `PlotExplanations()` → plot the explanation model

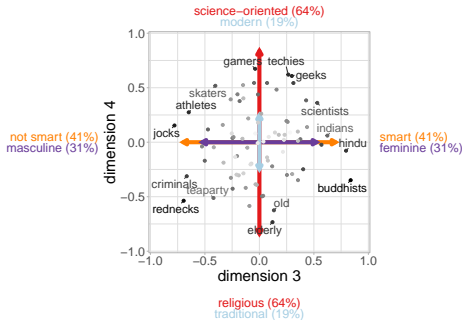
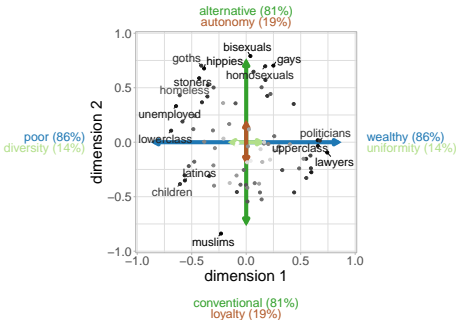
BIOT Package: Display the explanation model

► PlotExplanations() → plot the explanation model



BIOT Package: Display the explanation model

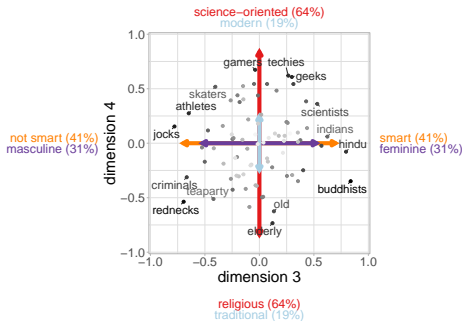
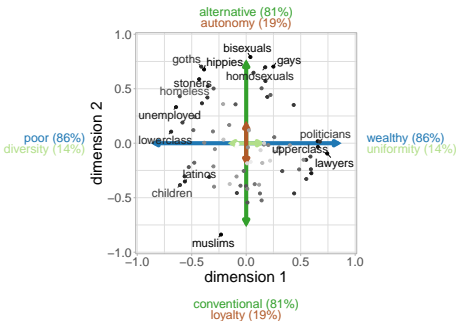
- ▶ PlotExplanations() → plot the explanation model



Important user options:

BIOT Package: Display the explanation model

- ▶ `PlotExplanations()` → plot the explanation model

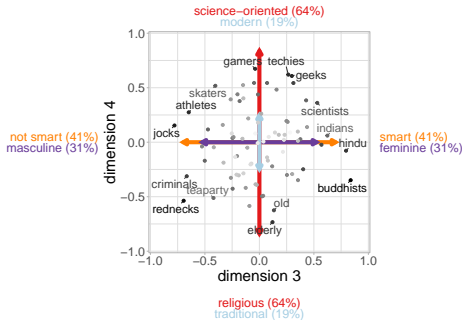
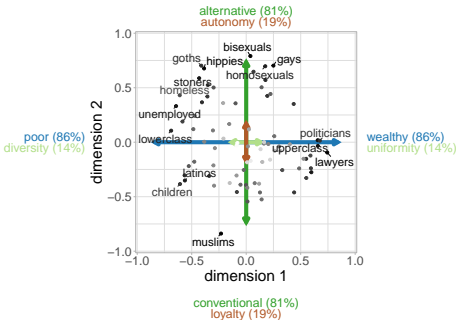


Important user options:

- ▶ number of points to label (example: 15)

BIOT Package: Display the explanation model

- ▶ `PlotExplanations()` → plot the explanation model

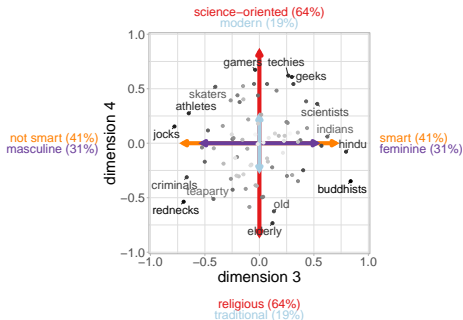
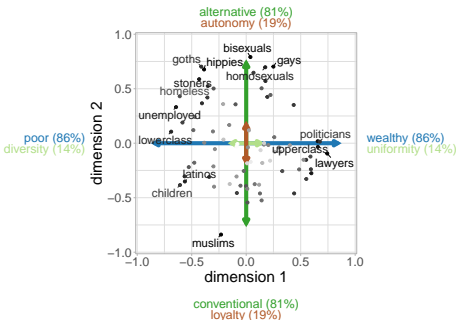


Important user options:

- ▶ number of points to label (example: 15)
- ▶ max number of attributes to show per axe (example: 2)

BIOT Package: Display the explanation model

- ▶ `PlotExplanations()` → plot the explanation model



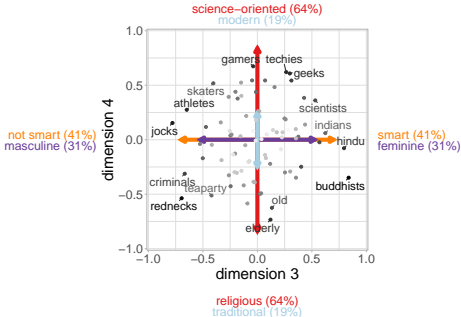
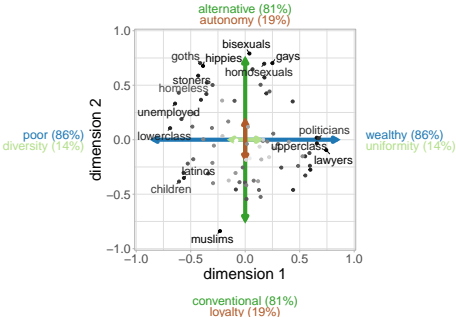
Important user options:

- ▶ number of points to label (example: 15)
- ▶ max number of attributes to show per axe (example: 2)
- ▶ shading of points proportional to reconstruction error (example: TRUE)

BIOT Package

Research Question: What are the major **attributes** people use to **judge the similarity/difference** between social groups?

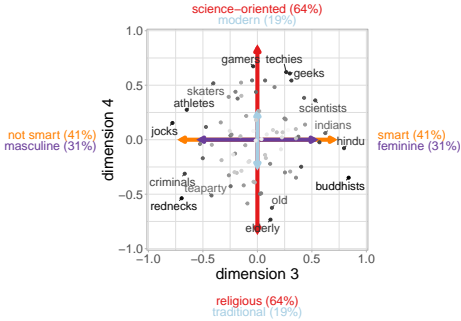
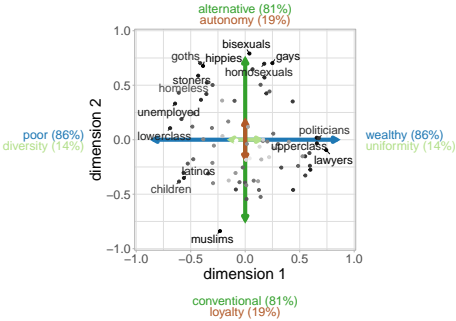
BIOT → Visualize an interpretable explanation model



BIOT Package

Research Question: What are the major **attributes** people use to **judge the similarity/difference** between social groups?

BIOT → Visualize an interpretable explanation model



Thank you for listening!