Beyond vanilla GLLVMs

Bert van der Veen

Department of Mathematical Sciences, NTNU

Questions so far?

Some topics covered in this workshop

▶ Vector GLMs, GLMMs and recap (day 1)
▶ Model-based unconstrained ordination: UO, UQO (day 2)
▶ Ordination with covariates (day 3)
▶ JSDMs: 4th corner model, Phylogenetic random effects (day 3)

Topics not covered in this workshop

▶ Joint models with imperfect detection (Tobler et al. 2019, Doser et al. 2023, Hogg et al. 2021)

▶ Mixture models for clustering species (Hui et al. 2015, Hui 2017)

▶ Spatio-temporal analysis

▶ Other methods for analysing community ecological data

▶ Much, much more.

Next steps

More response types for unimodal response model?

Spatio-temporal modeling is still a challenge.

▶ In space and/or time dimensions are usually large
▶ This slows down models considerably
▶ Requires additional approximations (e.g, as in INLA, CBFM, or NNGPs as in HMSC)

Next steps

In `gllvm` we mostly think about modeling ecological processes. However,

▶ Data are often not randomly sampled
▶ Might contain bias
▶ Presence-only from citizen scientists
▶ Mixed response types

This warrants further developments.

Next steps for me

I have an article on Phylogenetic random effects in the works.
However, it is the precursor for a few other developments:

▶ Improvements in usability (get methods to a larger audience)
▶ Incorporate traits and Phylogeny into ordinations
▶ Improvements of robustness (optimisation)

Trait-based analysis

1. CWM + RDA *Doledec et al. (1996)*
2. Double constrained ordination
   Lebreton et al. (1988), ter Braak et al. (2018)
3. Fourth corner (LV) Models *Brown et al. (2014), Ovaskainen at al. (2017), Niku et al. (2021)*
4. Model-based double constrained ordination (VGAM)

Figure 1: Quadriplot
ter Braak et al. (2018)

Reduced-rank modeling

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{E})\right\} = \mathbf{1}\beta^{\top} + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \qquad (1)$$

This is a GLMM

Reduced-rank modeling

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{E})\right\} = \mathbf{1}\beta^\top + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \quad (1)$$

This is an unconstrained ordination

Reduced-rank modeling

$$g\left\{ \mathbb{E}(\mathbf{Y}|\mathbf{E}) \right\} = \mathbf{1}\beta^\top + \mathbf{E}, \qquad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}\mathbf{\Gamma}^\top) \qquad (1)$$

This is an unconstrained ordination

$$g\left\{ \mathbb{E}(\mathbf{Y}|\mathbf{X}) \right\} = \mathbf{1}\beta^\top + \mathbf{XB} \qquad (2)$$

This is a GLM

Reduced-rank modeling

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{E})\right\} = \mathbf{1}\beta^\top + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top) \qquad (1)$$

This is an unconstrained ordination

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{X})\right\} = \mathbf{1}\beta^\top + \mathbf{X}\mathbf{B}_{lv}\boldsymbol{\Gamma}^\top \qquad (2)$$

This is a constrained ordination

Reduced-rank modeling

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{E})\right\} = \mathbf{1}\beta^\top + \mathbf{Z}\boldsymbol{\Gamma}^\top, \qquad \mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad (1)$$

This is an unconstrained ordination

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{X})\right\} = \mathbf{1}\beta^\top + \mathbf{Z}\boldsymbol{\Gamma}^\top \qquad (2)$$

This is a constrained ordination

4th corner LVM

From yesterday:

$$\eta_{ij} = \beta_{0j} + \mathbf{x}_i^\top (\boldsymbol{\beta}_x + \mathbf{b}_j) + \mathbf{tr}_j^\top \mathbf{B}_{xtr} \mathbf{x}_i \tag{3}$$

Usually we have many traits. And many covariates.
So, how can we reduce this down to fewer dimensions?

Hierarchical ordination

$$g\left\{\mathbb{E}(\mathbf{Y}|\mathbf{Z}, \mathbf{\Gamma})\right\} = \mathbf{1}\beta^\top + \mathbf{Z}\mathbf{\Sigma}\mathbf{\Gamma}^\top \tag{4}$$

Intercept ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯ Ordination

All effects go into the ordination
Pros: fewer parameters, easy visualization
Cons: generally more difficult to fit/estimate

Hierarchical ordination
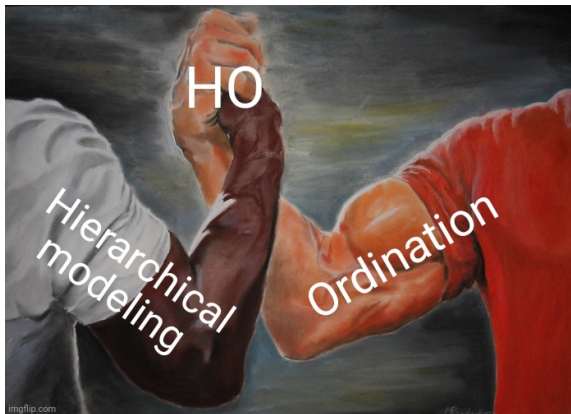
The "Hierarchical" in ordination

Hierarchically modeling the latent variables:

$$\mathbf{Z} = \mathbf{X}_{lv}\mathbf{B}_{lv} + \mathbf{E} \tag{5}$$

1. Fixed effects ————————————————

2. LV-level errors ————————————————

3. (Additional random effects)

van der Veen et al. (2023)

The "Hierarchical" in ordination

Hierarchically modeling the latent variables:

$$\mathbf{Z} = \mathbf{X}_{lv}\mathbf{B}_{lv} + \mathbf{E} \tag{5}$$

1. Fixed effects
2. LV-level errors
3. (Additional random effects)

van der Veen et al. (2023)

Next: hierarchically modeling the loadings

Hierarchical ordination
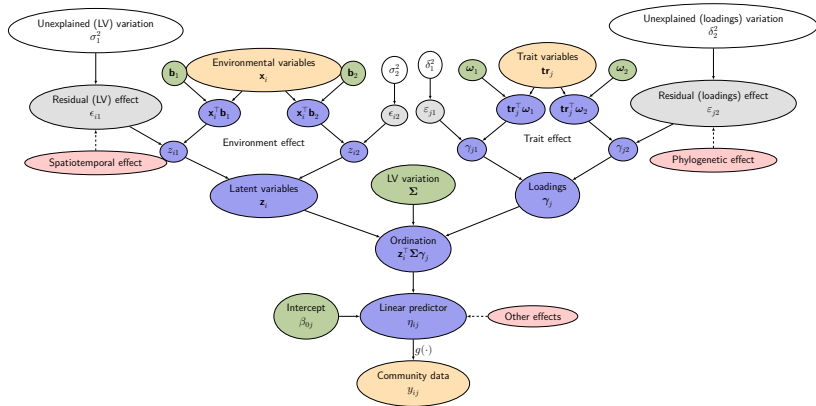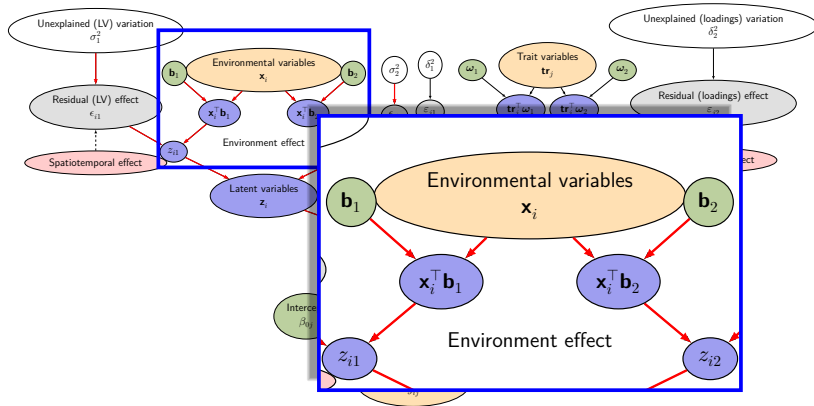
Hierarchical ordination

Hierarchical ordination
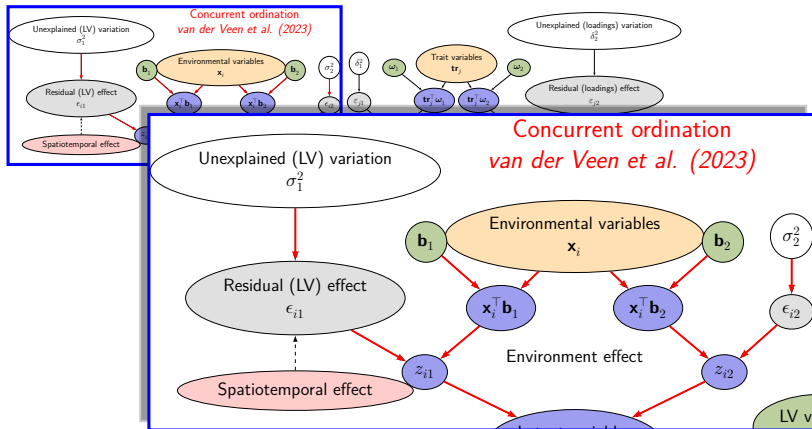
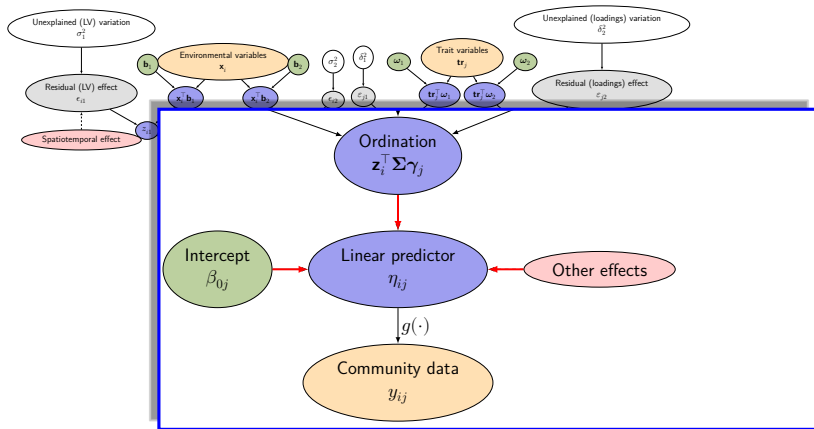# Hierarchical ordination
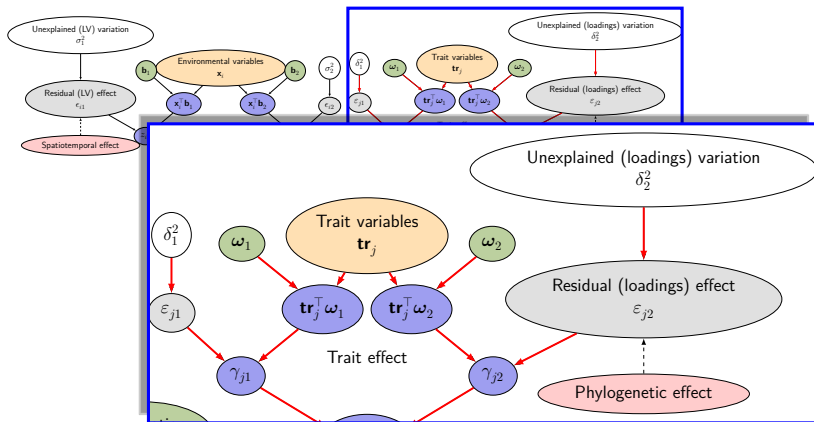
Hierarchical ordination
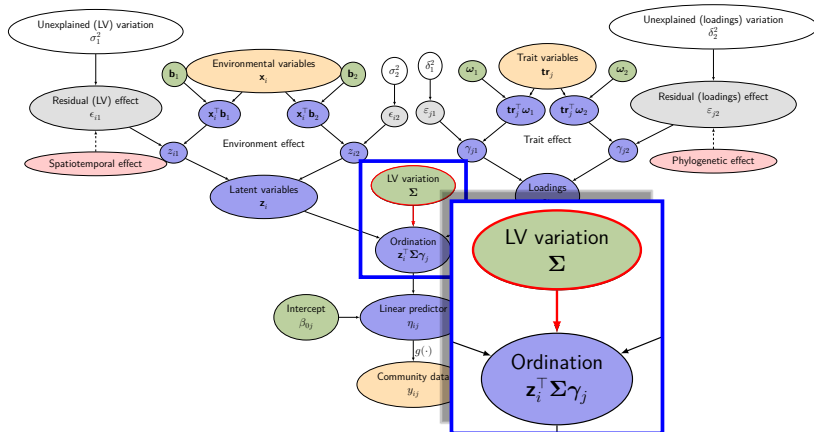
Hierarchical ordination

HO → 4th corner

$$\mathbf{Z}\,\boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^{\top} =$$

$$\underbrace{\mathbf{X}\mathbf{B}\,\boldsymbol{\Sigma}\,\mathcal{E}^{\top}}_{\text{Main effects}} + \overbrace{\mathbf{X}\mathbf{B}\,\boldsymbol{\Sigma}\,\boldsymbol{\Omega}^{\top}\mathbf{T}\mathbf{R}^{\top}}^{\text{4th corner terms}} + \underbrace{\mathbf{E}\,\boldsymbol{\Sigma}\,(\mathbf{T}\mathbf{R}\boldsymbol{\Omega} + \mathcal{E})^{\top}}_{\text{Other stuff}} \qquad (6)$$

1. Induces site correlations ⌐

2. Induces species correlations ──────────

HO → 4th corner

$$\mathbf{Z}\ \boldsymbol{\Sigma}\ \boldsymbol{\Gamma}^{\top} =$$

$$\underbrace{\mathbf{XB}\ \boldsymbol{\Sigma}\ \mathcal{E}^{\top}}_{\text{Main effects}} + \overbrace{\mathbf{XB}\ \boldsymbol{\Sigma}\ \boldsymbol{\Omega}^{\top}\mathbf{TR}^{\top}}^{\text{4th corner terms}} + \underbrace{\mathbf{E}\ \boldsymbol{\Sigma}\ (\mathbf{TR}\boldsymbol{\Omega} + \mathcal{E})^{\top}}_{\text{Other stuff}} \quad (6)$$

1. Induces site correlations
2. Induces species correlations

HO → 4th corner

Then we also have:

$$\beta = \mathbf{B}\ \boldsymbol{\Sigma}\ \mathcal{E}^{\top} + \mathbf{B}\ \boldsymbol{\Sigma}\ \boldsymbol{\Omega}^{\top}\mathbf{T}\mathbf{R}^{\top} \tag{7}$$

I.e., a hierarchical linear regression again of species responses to the environment. The "intercept" term follows from including an intercept column in the trait matrix.

Example: carabid beetles

▶ Data from *Ribera et al. (2001)*
   ▶ Counts
   ▶ 87 sites
   ▶ 68 species
   ▶ Beetles caught with pitfall traps
   ▶ 26 environmental variables and
     27 traits
   ▶ in Scotland
      ▶ Management, elevation, soil
        moisture, pH

Example: carabid beetles

▶ Data from *Ribera et al. (2001)*
  ▶ Counts
  ▶ 87 sites
  ▶ 68 species
  ▶ Beetles caught with pitfall traps
  ▶ 26 environmental variables and
    27 traits
  ▶ in Scotland
    ▶ Management, elevation, soil

4887 → 484 parameters (for two dimensions)

Case study: results[*]

▶ Reduced-rank approximated terms (e.g., 4th corner)

▶ Ordination plots

▶ Residual correlation plots

Sites

Species
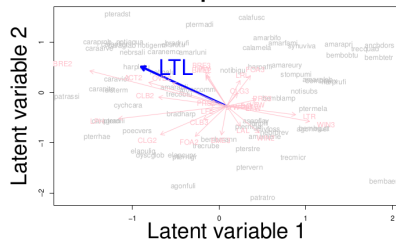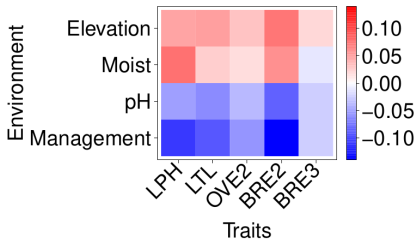
fitted with Nimble (de Valpine et al., 2017)

Hierarchical ordination

Read more here:

O'Hara & van der Veen (2022)

Algorithm-based ordination

This workshop focussed strongly on model-based methods for dimension reduction. However, recall that there are three **schools** of ordination:

1) "Simple-method"

2) "Algorithm-based" (also called "distance-based")

3) "Model-based"

Algorithm-based ordination

▶ Hierarchical Ordination doubles down on the "model" and ordination angles of these methods

▶ JSDMs are in the opposite direction: ordination is an afterthought

▶ Algorithm-based methods go the other direction: they double down on the ordination

JSDMs and Algorithm-based oridnation both focus on **prediction**

Algorithm-based ordination

A set of methods that we have not covered:

▶ Machine learning methods for dimension reduction
▶ These can be thought of as similar to NMDS
▶ I.e., iterative algorithms that have a geometric interpretation
▶ This includes for example: t-SNE, UMAP, Laplacian eigenmaps, isomap, LLE
▶ Many, many more
▶ No focus (as far as I know) on including covariates

Algorithm-based ordination

NMDS is still very popular, but there have been advances. In all directions, new approaches for co-occurrence analysis are not really adopted in ecology (besides JSDMs).

I will go through a few contemporary algorithm-based ordination methods here.

Disclaimer: I am not an expert at these methods

The paradigm

The idea is very different from model-based methods:

1) There is no model
2) We have a manifold (high dimensional topology)
3) Dimension reduction is performed to visualize it in (usually) 2D
4) Sites are located the that manifold, and we want to represent their association (no interest in species)

Common in (for example) processing large amounts of data; image processing. Or a biological example: morphometrics.

Topologies

▶ In ecology, we do not (yet) have hypotheses about high dimensional topologies

▶ Well, kind of. The unimodal model

▶ If there is a shape, it is usually consider an "artifact"

Consequently, not many of these methods have made it to ecology.

Clustering

Clustering and ordination are different things. But:

▶ Often has focus for algorithm-based ordination
▶ A good method supposedly retrieves either, or both, global and local structure
  ▶ Global: overal patterns, clusters, orderings
  ▶ Local: relationship of sites that are close together
▶ Many of these methods artificially introduces clusters when they do not exist

Example with handwritten letters

```
data(usps, package = "Rdimtools")
image(t(matrix(usps$data[4400,],nrow=16)[16:1,]))
image(t(matrix(usps$data[9900,],nrow=16)[16:1,]))
image(t(matrix(usps$data[6600,],nrow=16)[16:1,]))
```
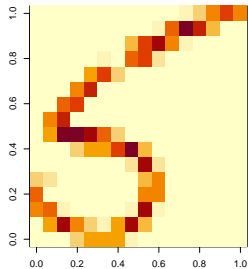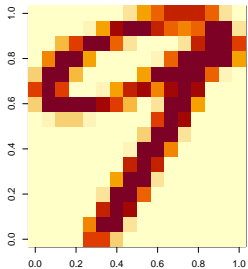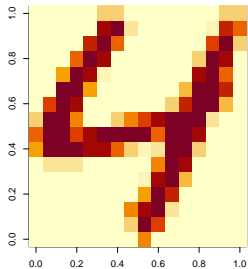
GLLVM

```
HW = t(matrix(usps$data[6600,],nrow=16)[16:1,]) # 6
model<-gllvm::gllvm(HW, num.lv = 2, family = "negative.binomial",
          starting.val="res",quadratic=TRUE, n.init = 1, seed = 2)
image(predict(model,type="response"), main = "Prediction")
image(HW, main = "Data")
```
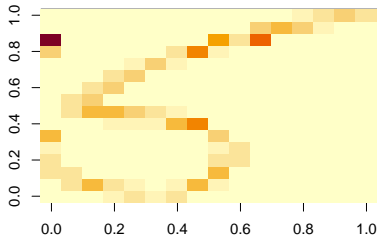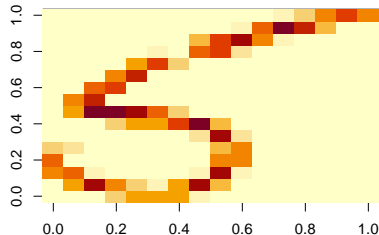
t-nearest Stochastic Neighbour Embedding ("t-SNE")

t-SNE was popular a few years ago. Developed by van der Maarten (2008)

▶ Mostly because it is very fast
▶ Focus is on placing dissimilar sites far away
▶ Distances in the data are represented with a normal distribution, in the ordination with a t-distribution
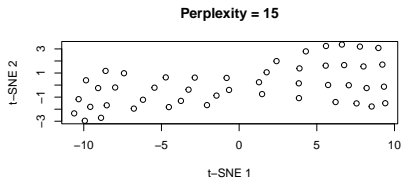▶ Just as in VA it minimizes Kullback-Leibler divergence

Example with Podani data

```
ord1 <- labdsv::besttsne(vegan::vegdist(PM1), perplexity = 13)
plot(ord1)
```
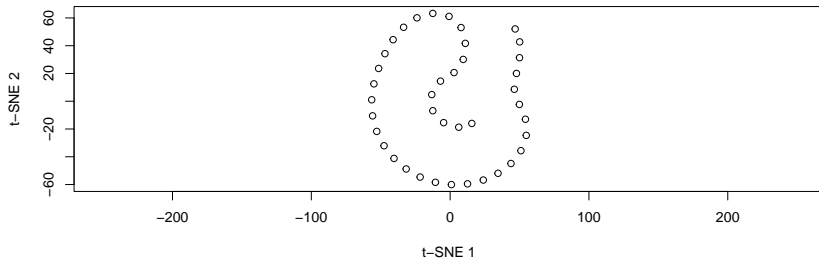

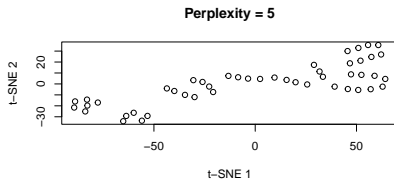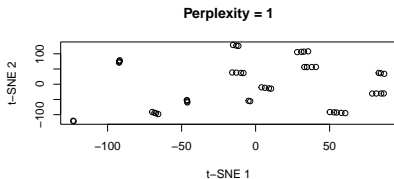Here, "perplexity" is a trade-off between local and global structure.

Example with the Minchin data

```
ord2 <- labdsv::besttsne(vegan::vegdist(MC), perplexity=15)
plot(ord2)
```
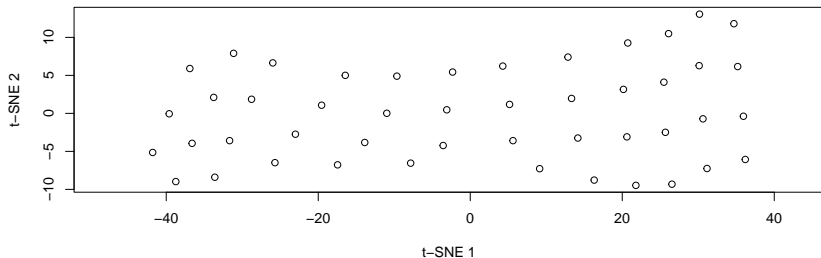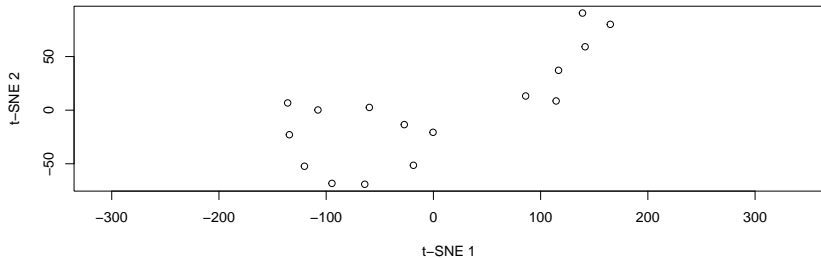

OK, we can see the lattice!

Example with the Minchin data
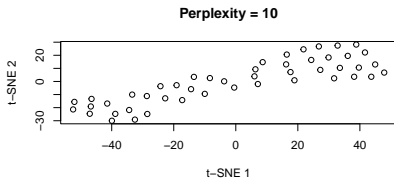
Not better than NMDS or GLLVMs, but definitely not bad.
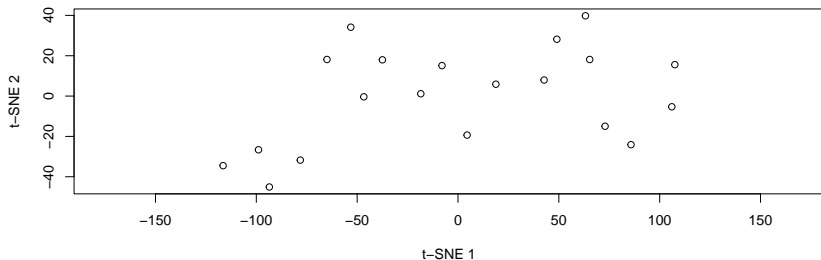
Example with handwritten letters

```
ord3 <- labdsv::besttsne(vegan::vegdist(HW),perplexity=5)
plot(ord3)
```

Example the dune dataset

Finally a real dataset:

```
ord3 <- labdsv::besttsne(vegan::vegdist(dune), perplexity=6)
plot(ord3)
```


Now we see clusters of sites that might not be real.

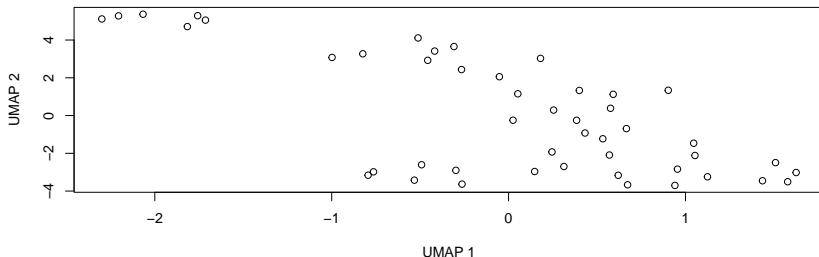# Uniform manifold approximation and projection for dimension reduction ("UMAP")

Is even more recent than t-SNE, developed by McInnes and Healey (2018)

▶ Supposed to be an improvement over t-SNE
▶ Better ability to retain global structure
▶ Faster and scales better

Example with Podani data

```
ord4 <- umap::umap(MC)
plot(ord4$layout, xlab = "UMAP 1", ylab = "UMAP 2")
```
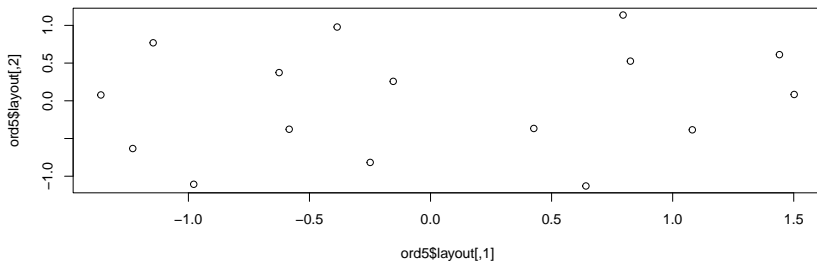

Various R-packages implement UMAP, others have some settings to tune
("epochs", and nearest neighbours in uwot).
This looks terrible (vegdist did not improve it much).

Example with handwritten letters

```
ord5 <- umap::umap(HW)
plot(ord5$layout)
```


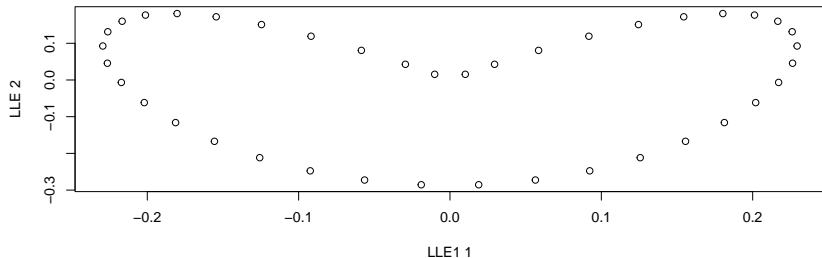Clearly, it does not do better than t-SNE.

Locally-linear embedding

Introduced by Roweis and Saul (2000)

▶ Also based on finding a fixed number of nearest neighbours
▶ Based on euclidean distance
▶ McCune and Grace (2002): it is more similar to PCA than to NMDS

Example with Podani data

```
ord6 <- Rdimtools::do.lle(as.matrix(PM1), type = c("proportion", 0.2))
plot(ord6$Y, xlab = "LLE1 1", ylab = "LLE 2")
```
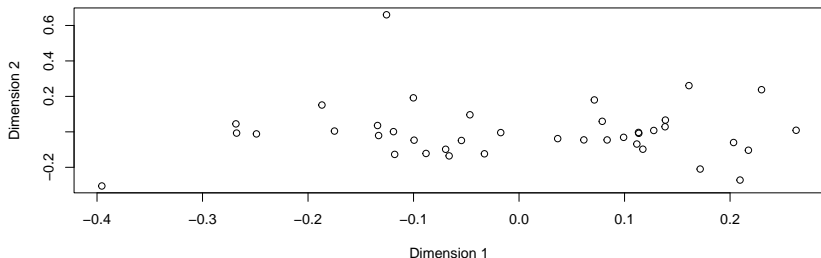

It made us a heart. But clearly not a good method.

Laplacian eigenmaps

Also a method based on eigendecomposition, by Belkin and Niyogi (2003)

```
ord7 <- Rdimtools::do.lapeig(PM1, type=c("proportion", 0.5))
plot(ord7$Y, xlab = "Dimension 1", ylab = "Dimension 2")
```
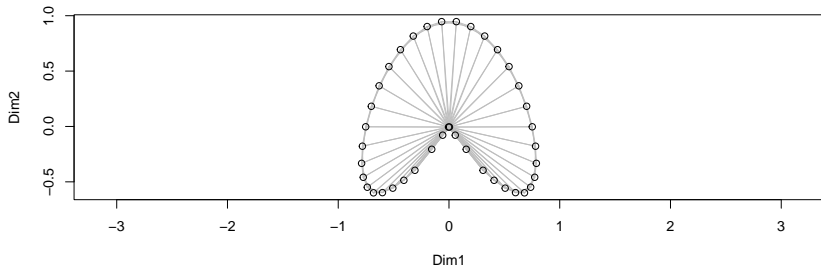
isomap

Isomap, also known as Principal Curves (Hastie 1984)

▶ Has a constrained variant (De'eath 1999)
▶ Uses geodesic distances
▶ Various parameters to tune: `epsilon`, `k`

Example with Podani data

```r
ord8 <- vegan::isomap(vegan::vegdist(PM1), k = 3)
plot(ord8)
```

Constrained algorithm-based ordination

There are some constrained methods

▶ constrained NMDS exists (and not difficult to implement)
▶ db-RDA

But general focus is on unconstrained ordination.

Next generation ordination

I have given a brief overview of a few alternative ordination methods.

▶ Context is important to see: `gllvm` peforms very well
▶ GLLVMs have a wide range of tools for inference, but also prediction
▶ They could be adjusted towards topological spaces, but why would we?
▶ Dimension reduction is a very broad class of methods
▶ Developments are ongoing, and it is important to keep an eye on the competition!

One thing is clear: some very good methods have been developed in the last decades.