

# Data visualization

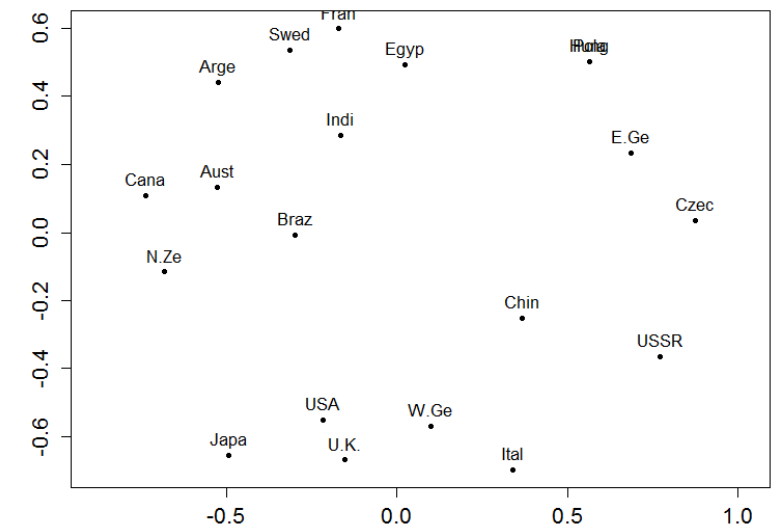
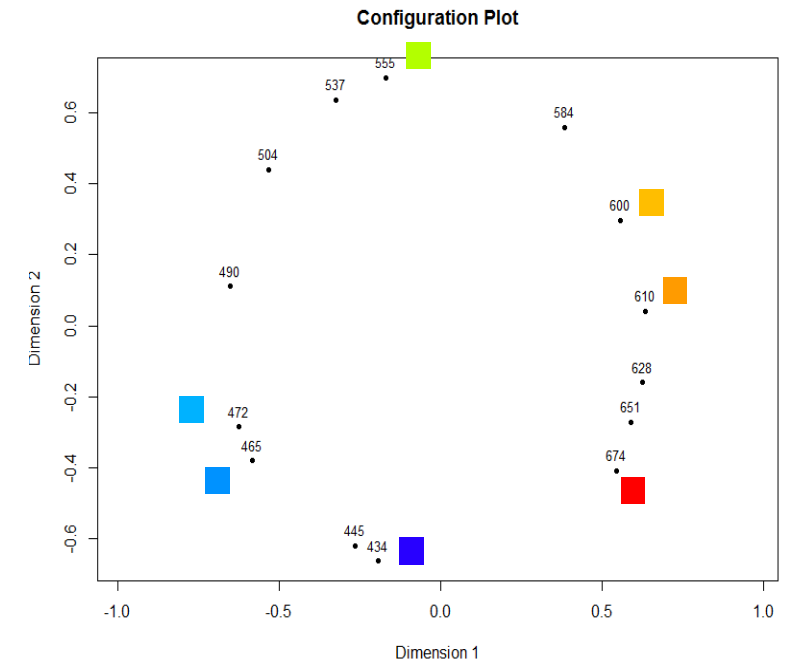
Q&A 8

David Hoksza

<http://siret.ms.mff.cuni.cz/hoksza>

# Příklady

- Psychometrický experiment
  - Data z aggregate proximity matrix? ANO
  - Různé vzdálenosti v poměrně jinak rovnoměrném rozložení barev?
- Obchod mezi zeměmi
  - dim:  $N-1 \rightarrow 2$
  - Záporná hodnota os – interpretace?



# MDS

- Kdy je lepší Kruskalův stress než normalizovaný stress?  
→ typicky se stress hodnotí pomocí  $\sigma_1$
- Lze použít stress i při metrickém MDS? → ANO  
(klasické MDS vs metrické MDS)
- Existují chytřejší metody snažící se vyhnout lokálnímu maximu v iteraci?
  - Nedal by se pro hledání optima použít např. SGD algoritmus?
- Dají se váhy upravovat i dynamicky za běhu? Popřípadě, jak by se toho dalo využít?

$$\sigma_n^2(X) = \frac{\sum_{i<j} w_{ij} (d_{ij}(X) - \delta_{ij})^2}{\sum_{i<j} w_{ij} \delta_{ij}^2}$$
$$\sigma_1(X) = \left( \frac{\sum_{i<j} w_{ij} (d_{ij}(X) - \delta_{ij})^2}{\sum_{i<j} w_{ij} \delta_{ij}^2} \right)^{1/2}$$

# MDS interpretace

- Existuje heuristika pro interpretaci os?
- Jak poznat, že metrický MDS dává špatné výsledky? → stress
- Může MDS (configuration plot) pomoci najít společné vlastnosti nebo málo viditelné odlišnosti?

# Vhodnost MDS

- Kdy je lepší MDS → neočekávám lineární závislosti / mám váhy
- Je pro obrázky lepší MDS nebo PCA? → spíše MDS
- Lze MDS použít pro kreslení grafů? → viz. Kamada-Kawai
- Alternativy MDS? (ML?)
  - Sammon mapping, Isomap, Locally Linear Embedding, **t-SNE**, **UMAP**, ...

# Ostatní

- Proximita – velká vs malá? Jak převést proximitu na vzdálenost?
- Odvození vzorce na zkoušku?