

# Master thesis proposal

## Development of visualization possibilities of the **RMaCzek** package

Krzysztof Bartoszek

October 6, 2022

The **RMaCzek**<sup>1</sup> R package [1, 4] offers the possibility of plotting both the symmetric and asymmetric version of Czekanowski’s Diagram [2]—a visualization technique for representing similarity between (arbitrarily dimensional) observations. Recently, clustering capabilities have been added to it [3].

An issue with the current **RMaCzek** package is that for even relatively small data (30 observations) the graphic and the axes labels (observations’ names) can become unreadable. To change the situation one would need to implement an algorithm that factors in the number of observations and their names, the identified clusters and based on these proposes an optimal font for the axes, size of the points representing the distances and distance between the points on the diagram. Potentially one might need to pre-cluster observations and only present similarities between grouped objects.

The goals of an ideal thesis are to develop **RMaCzek**’s implementation of Czekanowski’s Diagram for big data.

For the thesis simulated and from the literature data can be used.

## References

- [1] K. Bartoszek and A. Västerlund. “Old Techniques for New Times”: the **RMaCzek** package for producing Czekanowski’s Diagrams. *Biometrical Letters*, 57(2):89–118, 2020.
- [2] J. Czekanowski. Zur Differentialdiagnose der Neandertalgruppe. *Korespondentblatt der Deutschen Gesellschaft für Anthropologie, Ethnologie und Urgeschichte*, XL(6/7):44–47, 1909.
- [3] Y. Luo. Czekanowski’s clustering, 1 2021. Master thesis in Statistics, Division for Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, Sweden.
- [4] A. Västerlund. Czekanowski’s Diagram: Implementing and exploring Czekanowski’s Diagram with different seriation methods, 6 2019. Master thesis in Statistics, Division

---

<sup>1</sup><https://cran.r-project.org/web/packages/RMaCzek/index.html>

for Statistics and Machine Learning, Department of Computer and Information Science,  
Linköping University, Sweden.