Structure of the SPL Paper

Structure:

1. Introduction
2. Theory of Cluster Analysis in the Context of Mixed Type Data
    a. Present standard. procedure of cluster analysis (for mixed data in contrast to numerical data)
        i. Dissimilarity matrices
        ii. Algorithms
        iii. Determining the number of clusters
        iv. Visualization
    b. Literature review of cluster analysis for mixed data
3. Data and research set up
4. Simulation
    a. Data Preprocessing
    b. Display results of the different techniques
    c. Visualization
    d. Interpretation
5. Evaluation of Different Techniques
6. Conclusion

Notes:

- Include Code in appendix
- Make Quantlets to graphics
- The above structure also functions as a description of the content for you to understand with to include in each paragraph

Literature

RFM Concept

- https://en.wikipedia.org/wiki/RFM_(customer_value)

Missing imputation by rpart

- https://www.r-bloggers.com/missing-value-treatment/

Overview of diverse clustering techniques

- http://girke.bioinformatics.ucr.edu/GEN242/pages/mydoc/Rclustering.html
- copies from Härdle textbook in our github

Kmodes algorithm

- https://dabblingwithdata.wordpress.com/2016/10/10/clustering-categorical-data-with-r/

Structure of the SPL Paper

Determining number of cluster

- https://stackoverflow.com/questions/15376075/cluster-analysis-in-r-determine-the-optimal-number-of-clusters

Hierarchical Clustering

- https://medium.com/@anastasia.reusova/hierarchical-clustering-on-categorical-data-in-r-a27e578f2995

Procedure for PAM Algorithm

- https://www.r-bloggers.com/clustering-mixed-data-types-in-r/
- L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9(Nov):2579-2605, 2008.