PCA WHO SDGs

Assignment 5

AJ Smit

2020-06-28

Table of contents

Submit a Rmarkdown script wherein you provide answers to Questions 1–5, and provide the associated compiled html output. Label the script as follows: `BCB743_<Name>_<Surname>_Assignment_5.R`, e.g. `BCB743_AJ_Smit_Assignment_5.R`.

Note that these questions also cover the Cluster Analysis lecture. Refer to the Principal Component Analysis SDG and the Cluster Analysis lecture material to see the questions in context.

The deadline for this submission is Monday 1 August 2022.

0.1 Assignment 5 Questions

Question 1: Explain the code section-by-section in long-form text. Include also the reasoning/rationale behind each section.

Please see the Cluster Analysis section for additional questions.

Question 2: What happens if we use `pam()` to create four, five, or even six clusters?

Question 3: In your reasoned opinion, what would be the optimal number of clusters to use?

Question 4: Repeat the analysis using either `kmeans()` or `hclust()`, and feel free to use the **factoextra** helper functions and visualisations. Are the results markedly different? Which clustering approach do you wish to proceed with—i.e., `pam()`, `hclust()` or `kmeans()`?

Question 5: Describe the patterns that you observe at the end of your ordination and final cluster selection (i.e. based on the optimal number of clusters and whichever cluster technique you deem most appropriate). How does South Africa fare in terms of attaining SDGs? Contrast with some key countries of your choice to make your points. Label the key countries that you refer to in your text by updating the code accordingly. Continue to explain these patterns in terms of the global socio-political/socio-economic landscape. Provide a discourse about possible explanations for the patterns observed globally and regionally.