

DAT 640 Practical R Activity Five Guidelines and Rubric: Cluster Analysis

Cluster analysis is used to group the observations in your data into *clusters* in which the data in a cluster is more similar than it is to data in other clusters.

For this assignment, you will have two deliverables:

Part I

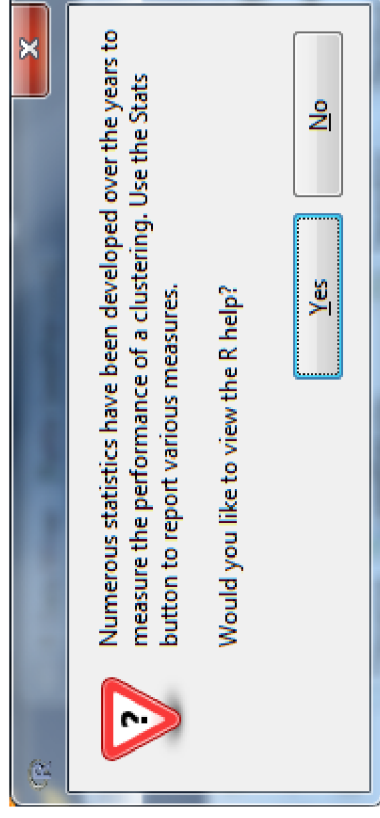
Submit your commands and results of three k-means models:

- Load the data set weather into Rattle (select “Execute” in the first tab and accept the default weather data set).
- From Rattle, select the Cluster tab to be presented with various clustering algorithms.
- The k-means algorithm is the default option, and by default, 10 clusters will be built as the model. A random seed is already provided, and changing the seed will result in a randomly different collection of starting points for our means. With the weather data loaded from the Data tab, click the Execute button while on the Cluster tab. Finally, click the “Stats” button to retrieve additional statistics on the model results.
- Copy and paste the results into your document and, in 2–3 paragraphs, compare your output to the one in Figure 9.3 of the textbook and describe the information contained in the main results (those above the “General cluster statistics”).
- Execute the algorithm two additional times using a different number of clusters other than the default 10, and paste the results into your document.
- Compare the output from the two procedures.

Part II

Validate the model:

- Use the Help ->Cluster->Stats menu to open the help documents in RStudio.



- Research and choose one validation index and provide:
 - A basic overview of how the chosen index measures the quality of the model
 - A description of which of the models should be considered more accurate based on the chosen index

Rubric

Guidelines for Submission: Your submission should follow these formatting guidelines: double spacing, 12-point Times New Roman font, one-inch margins, and citations, if any, in APA format when appropriate.

Critical Elements	Exemplary (100%)	Proficient (90%)	Needs Improvement (70%)	Not Evident (0%)	Value
Demonstration of R Command Execution	Includes demonstration of all R command executions in the form of screenshots or command output listings	Includes demonstration of most of R command executions in the form of screenshots or command output listings	Includes limited demonstration of R command executions in the form of screenshots or command output listings	Does not include demonstration of R command executions in the form of screenshots or command output listings	15
Required Elements	Meets “Proficient” criteria and provides additional analyses or relevant supporting scholarly material	Submission includes all the required elements of the analysis	Submission includes more than 50% of the required elements of the analysis	Submission includes less than 50% of the required elements of the analysis	15
Model Computation	Meets “Proficient” criteria and includes assessment of alternative settings/tuning procedures	Demonstrates accurate utilization of the R/Rattle packages per the model specifications	Demonstrates utilization of the R/Rattle packages with inaccuracies or does not adhere to the model specifications	Does not demonstrate utilization of the R/Rattle packages	30
Model Analysis	Meets “Proficient” criteria and draws insightful conclusions that are thoroughly defended with evidence and examples	Provides in-depth analysis that demonstrates complete understanding of concepts and draws informed conclusions that are justified with evidence	Provides an analysis that demonstrates a general understanding of concepts or draws logical conclusions, but does not defend with evidence	Does not provide in-depth analysis	30
Articulation of Response	Submission is free of errors related to citations, grammar, spelling, syntax, and organization and is presented in a professional and easy-to-read format	Submission has no major errors related to citations, grammar, spelling, syntax, or organization	Submission has major errors related to citations, grammar, spelling, syntax, or organization that negatively impact readability and articulation of main ideas	Submission has critical errors related to citations, grammar, spelling, syntax, or organization that prevent understanding of ideas	10
Earned Total					100%