

DAT 640 Practical R Activity Two Guidelines and Rubric: Summarizing Data

Summarizing data provides an entry point to building a model. Summaries may alert you to a particular pattern that may be exploited in your model or identify issues that need to be corrected. In this activity, you will use R and Rattle to produce such a summary.

Your submission should include the following **critical elements**:

- Provide evidence of loading any of the built-in R data sets (**not** the *default weather data set*) and invoking Rattle.
- You can see the sample data sets available by entering the command *library(help = "datasets")*. Select a data set that is not a time series but, instead, has columns of values. (*iris* and *swiss* are good examples of the type of data set you want to use.) "Load" the data set with the command *data(<dataset name>)*, for example, *data(iris)*.
- In R, summarize the contents of that data set with:
 - `names()`
 - `dim()`
- In Rattle, create a "text" summary of the data using the method outlined in Sections 2.4 and 3.7. This will produce a summary table about your data set. **Tip:** Be sure to save your summarizing program for later use.
- In two to three paragraphs, discuss the relationship between the exercise in Section 3.7, Documenting Data, and the concepts in Section 7.1, Data Issues. Explain the necessity of both of these concepts in a data mining report.
- Include three distribution charts from the Explore tab with a brief explanation for each.

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 50% or more, but not all, 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%