**Exercise 1 - Getting Started with R and Exploring Iris Data**

Top of Form

|  |  |
| --- | --- |
| **Exercise Overview**  In this exercise, you will get familiar with RStudio interface. You will write a script to explore the pre-installed Iris data. The data exploration includes previewing the data, descriptive statistics, grouping, frequency tables, and data visualization.  **Before working on this exercise, install the latest version of R and R studio.  See instructions under Module 1 content.**  **Exercise Instructions** | Iris Flower MeasuermentRStudio Shortcut |
| **Part 0** - Complete the tutorial in the ExploringIrisData.docx Word document. Depending on the R and R Studio version and on the operating system, your output might be slightly different from the output in a Word document. | |

**Part 1** – Write an R script with the following commands in the same order as listed below.

The script contents

* Line 1 - a comment #This is my first R script
* Line 2 - Sys.time() command
* Line 3 - Sys.info() command
* Line 4 - R.version command
* Line 5 – leave blank
* Line 6 – the command to display the first 5 rows in the iris data
* Line 7 – the command to display the maximum value of the Petal.width variable
* Line 8 – the summary command to display the statistics only for Petal.width variable
* Line 9 – the command to display the help page for plot command
* Line 10 – the hist command to build the histogram for the Petal.width variable

Save the file as your first initial Last name\_exercise1 (for example, Yelena Bytenskaya saves the file as YBytenskaya\_exercise1)

Keep in mind that lines 6-10 may have more than one correct answer.

**Part 2** –Part 1 script output and iris data exploration

|  |  |  |
| --- | --- | --- |
| 1. Run all commands in Part 1. Paste the output into the Word document. For each command, the output must show the command and its output. 2. Explain what information could the R commands reveal about the iris data? To answer this question, you may need to review the tutorial in Part 0 | |  |
| **Part 3** - Data visualization   1. Explain what the hist command on line 10 in Part 1 reveals about Petal.width variable in the iris data 2. Why is it important for the data-mining tool to have visualization capabilities? | Iris Histogram | |
|  |  | |
| **Part 4** -  Is an open source data mining tool better than a commercial one?  Why or why not? To answer this question, consider the key differences between open source and close source tools. | | |
|  | | |

**Exercise Deliverables**

* A R script for Part 1
* A single Word file with screenshots for Part 1 and answers to Part 2, 3, and 4.

Submit both files into Exercise 1 Assignment Folder.  **You need to submit both files in a single submission** because only one submission is allowed.

**Exercise Grading**

* **This exercise is worth 2% of the course grade.**
* To get full credit for part 1 of the assignment, the R and RStudio version must be the latest
* An R script for part 1 must return the expected output without error messages
* The commands in the R script must be in the same order as instructed. The corresponding output screenshots must match the order of commands.
* All answers for Part 2-4 must be in order and in your own words to receive full credit for Part 2-4.
* Cite references in text and on a reference list as needed.
* In addition, grammatical and spelling errors may affect the grade.

Keep in mind that some questions in Part 1 may have more than one correct answer.

**Start working on assignment early in a week.**An estimated time to complete this assignment is 3-5 hours.

Post your questions about this exercise in the course discussion.

Good Luck!

Bottom of Form