Lab 3 GSI

Name of student *

Mengling Liu

Readability of report (5 points) *

	1	2	3	4	5	
Narrative unclear and/or difficult to read and/or there is not much detail in the explanations	○	◉	○	○	○	Narrative very clear and/or easy to read and lots of detail is given

Grammar of report (5 points) *

	1	2	3	4	5	
Incorrect of written grammar pervasive	○	○	○	◉	○	Excellent written grammar

Level of written detail on comparison of R and C++ implementation and runtime *

	0	1	2	3	
Did not write about a comparison of the R and C++ implementation	○	○	○	◉	Wrote a detailed comparison between the R and C++ implementations

Correctly coded the parallelization of k-means and pairwise similarity in R/C++ (3 points) *

	0	1	2	3	
Incorrect implementations	○	○	◉	○	Seems correct to me

Comments on implementation of parallelization or the similarity measure

I am not sure exactly what the issue is, but your similarity scores are not correct in the Ben-Hur figures, so there is something wrong with the code

Efficiency and practicality of R and C++ (3 points) *

	0	1	2	3	
Inefficient (e.g. repeated computations, unnecessarily saved objects, etc.)	○	○	◉	○	Very efficient and practical

Suggestions for improving efficiency of R and/or C++ code *

The computation of the ECDF can be done much more efficiently without for loops, but looks good other than that.

In general in R if you are using lots of for loops think about whether there is another way to do it

Does the author satisfy the following code readability requirements? (3 points)

- [ ] Always spacing before and after variable assignment and addition symbols (" = ", " + ") and after commas
- [ ] No line of code exceeds 80 characters
- [x] Consistent variable naming for all variables (words always separated by one of "_" or ".")

Clarity of variable names (2 points) *

	0	1	2	
Variable names are unclear and meaningless (e,g, 'df', 'x', 'data2)'	○	○	◉	Variable names are helpful and unambiguous

Quality of code comments (2 points) *

	0	1	2	
There are almost no comments	○	○	◉	The comments explain clearly what is being done and why

Suggestion for improving readability of R code *

Keep your code lines less than 80 characters and add spacing around +, =, after commas etc.

Did the student provide all code necessary for recompiling their results AND report (note: report not actually reproduced) (2 points) *

	0	1	2	
Incomplete code or no .Rnw/.Rmd file provided	○	○	◉	Everything was provided

Clarity of folder structure (2 points) *

	0	1	2	
Many excess files not relevant to report	○	◉	○	The purpose of each file is clear and there are no excess files in the lab2 folder

Comments on folder structure

Remember, not to upload the data. Also, keep things in different subfolders rather than just having it all in the lab3 folder without any additional structure.

Correctly produce Ben-Hur type figures (3 points) *

	0	1	2	3	
Did not provide a figure like Ben-Hur	○	◉	○	○	Figures look correct

Discuss one (or more things) that you liked about the author's figure *

Good job using a sequential color scheme for the ECDF

Discuss one (or more things) that could be improved the author's figure *

For comparing histograms make sure to have the x-scale be the same, otherwise it can be misleading

Also, your similarity scores do not appear to be correct. For example, with K = 3 you should get values between 0.9 and 1

Justification of conclusions drawn from Ben-Hur figures (3 points) *

	0	1	2	3	
Did not write about any conclusions drawn from the figures	○	○	◉	○	Clearly outlines interpretation of the figures and drew reasonable conclusions (e.g. found k =3, or some other value, is the best and provides reasons why)

Comments on the conclusions and interpretations of the Ben-Hur type figures *

It would be good to add more detail about why you make these decisions. What does it mean for the similarity score to be high? To have low variance?

Things the author did well *

Good job discussing the R and C++ code!

Things the author could improve on *

Make sure to have an introduction and transition sentences that tie the pieces of the report together, rather than just separate bullet points.