# Lab 3 GSI grading

001	feed	ا م ما
GOL	reea	Dack

The respondent's email address (rebeccabarter@berkeley.edu) was recorded on submission of this form.

The name of the student whose paper you are grading \*

Yizhou Zhao

The student ID of the student whose paper you are grading \*

3032130362



Readability and grammar of written report (5 points) \*

	1	2	3	4	5	
Difficult to read and/or poor	$\bigcirc$	$\bigcirc$		$\bigcirc$	$\bigcirc$	Clearly written and excellent
grammar						grammar

Level of written detail on comparison of R and C++ implementation and runtime (3 points) \*

	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



Review the code written by the author. If you aren't sure of the correctness of the implementation, that's fine, just give a grade and say so in the comments.

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



#### Comments on implementation of parallelization or the similarity measure?

You said you are running the code on the cluster, but you are not correctly setting up the parallelization. For instance, you should be explicit about the number of cores to be used (rather than using detectCores) and then you need to specify now many cores per task in the shell script.

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



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

It looks ok. It is at times unclear what is going on since there aren't quite enough comments (e.g. in the getCoeff function)

Does the author s points)	satisfy the fo	llowing code rea	dability requi	rements? (3
Consistent spaci	_	fter variable assignn	nent and addition	n symbols (" = ", " +
No line of code e	xceeds 80 chara	acters		
Consistent variab	ole naming (word	ds always separated	by one of "_" or	".")
Clarity of variable	e names (2 po	oints) *		
	0	1	2	
variable names are unclear and meaningless (eg `df`, `x`, `data2`, etc)				variable names are helpful and unambiguous
Quality of code co	omments (2	points) *		
	0	1	2	
there are almost no comments				the comments explain clearly what is being done and why

### Suggestions for improving \*readability\* of R code \*

"stabalize" isn't a word... the word is "stabilize" but we aren't trying to stabilize the results anyway, we are trying to assess the stability of the results with various k.

You could vastly improve your spacing (e.g. for " = " and after commas) it is very inconsistent.

Why didn't you provide the raw .Rnw/.Rmd files for generating your report and figures?

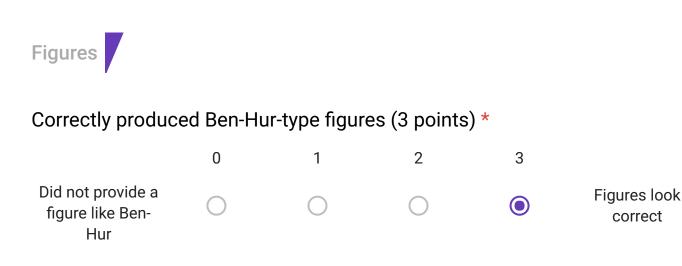
your commenting is pretty good overall, but there are a few places where more comments would have been helpful.

Did the student provide all code necessary for recompiling their results AND report (note: you do not have to actually reproduce their report) (2 points) \*

	0	1	2	
Incomplete code or no .Rnw/.Rmd file provided			0	Everything was provided
Clarity of folder	structure (2 po	oints) *		
	0	1	2	
The folder structure was very confusing	0			It was clear what each file corresponded to and there were no surplus files floating around

Optional comments on folder structure and files provided (please provide comments if you docked points for any reason)

Providing a detailed readme would have been helpful for making it clear what each script was for.



If the Ben-Hur figures do not look correct, what is wrong?

Quality of Ben-H	ur Figure 3	3 replication	ı figures (3 p	oints) *	
	0	1	2	3	
Did not provide a figure like Ben- Hur	$\circ$			$\circ$	Provided clear and visually appealing figures
Discuss one (or figures *	more) thin	ngs that you	liked about	the autho	r's Ben-Hur
Nice use of ggplot ov	verall.				
Discuss one (or Hur figures *	ŕ		·		
its own range for the		-		molecaumy.	Edon plot also has
The legend for the cu	_	_	-	(a title of "nu	ımber of clusters"
Justification of c	conclusion	ns drawn fro	m the Ben-H	lur-type fig	gures (3 points)
	0	1	2	3	
Did not write about any conclusions drawn from the figures					Clearly outlined interpretations 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 \*

What about the cumulative plot? Doesn't that make it look like k=2 is the best?

What do you mean "transitions between 2 and 3 clusters and 3 to 4 cluster..."?

Your conclusion doesn't seem to follow from the figures...



Provide concluding comments

One or more things that you thought was well done overall \*

I really like that you tried multiple similarity measures

### One or more things that could be improved upon overall \*

Take care to really understand the results of the figures you make.

Also, try to write with a bit more clarity. It was unclear what you were trying to say in many places (particularly on the final page).

Any other comments that you would like to add?

This form was created inside of UC Berkeley.

Google Forms