Lab 2 GSI grading

This form contains the final scores and extra comments from the GSI.

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

Name of student

Yizhou Zhao

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

3032130362

Readability of report (5 points) *

	1	2	3	4	5	
Narrative unclear and/or difficult to read	○	○	◉	○	○	Narrative very clear and/or easy to read

Grammar of report (5 points) *

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

Analysis: redwood trees

In this section you will assess the actual analysis using kernel density estimation and loess on the redwood trees data.

Detail of kernel density estimation analysis (3 points) *

	0	1	2	3	
Did not explore different bandwidths or kernels	◉	○	○	○	Explored a variety of bandwidths and kernels and clearly related these to the bias-variance-tradeoff

Relevance and quality of figures related to kernel density estimation (3 points) *

	0	1	2	3	
Did not provide any figures	○	○	◉	○	Provided clear, relevant and visually appealing figures

Discuss one (or more) things that you liked about the author's kernel density estimation figures

I like the simplicity of your plots

Discuss one (or more) things that could be improved for the author's kernel density estimation figures

You unfortunately didn't write anything about this section!

Try to use more human-readable labels ("temperature" instead of "humid_temp").

It would be much clearer if you directly labeled the plots with their bandwidth and kernel rather than stated this in the caption.

Choosing a wider range of bandwidths would help display the bias-variance tradeoff.

Detail of loess smoothing analysis (3 points) *

	0	1	2	3	
Did not conduct an analysis using a loess smoother	○	◉	○	○	Explored a variety of bandwidths and polynomials and clearly related these to the bias-variance-tradeoff

Relevance and quality of figures related to loess smoothing (3 points) *

	0	1	2	3	
Did not provide any figures	○	○	◉	○	Provided clear, relevant and visually appealing figures

Discuss one (or more) things that you liked about the author's loess figures

simple and clean.

Discuss one (or more) things that could be improved for the author's loess figures

What's going on with the last plot?

Making the points a light, transparent grey would help the viewer focus on the loess curves.

Analysis: linguistic survey

Level of detail in the written comparison between two questions (3 points) *

	1	2	3	
Little detail (barely described the relationships between the two questions)	◉	○	○	Very detailed (described clearly the geographical groups formed by each question and discussed how the questions were related to one another)

Optional comments about author's analysis of the two questions

You did not compare two questions

Quality and relevance of figures (e.g. maps) for the two questions (3 points) *

	0	1	2	3	
Did not provide figures	○	○	◉	○	Provided clear, informative, and visually appealing figures

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

The maps are very clear.

Discuss one (or more) things that could have been improved for the author's figure(s)

It is a bit redundant to both show the answers separately as well as together. I think showing them together is sufficient.

Removing the gridlines in the background and the axes would help reduce redundancy in the figure.

Providing the actual question in the title of the figure would be an excellent idea.

Discovered that the binary encoding should be aggregated (e.g. in lat-long bins) in order to perform meaningful PCA (or other dimensionality reduction technique) (2 points) *

	0	1	2	
Did not mention that dimensionality reduction did not work well on the binary encoded data	○	◉	○	Found that PCA was inneffective for binary encoding and used aggregated data instead (e.g. grouped by ZIP or lat/long bins)

Discussed clustering and related these clustering results to geography (3 points) (note: deduct a point if the author used lat/long as a variable in their cluster algorithm) *

	0	1	2	3	
Did not discuss clustering	○	○	◉	○	Discussed in detail the clusters found in the data and how they related to geography

Optional comments on cluster analysis

You could have discussed a lot more about the PCA you performed.

How

Quality and relevance of figures related to clustering and geography (3 points) *

	0	1	2	3	
No figures provided	○	○	◉	○	Provided clear, informative, and visually appealing figures

Discuss one (or more) things you liked about the author's clustering figures

Discuss one (or more) things that could be improved for the author's clustering figures

There is a lot of overplotting all figures in this section. Making the points smaller and transparent would help a lot.

Analyzed the robustness/stability of a finding (3 points) (give partial points if the author showed stability only by re-running K-means without perturbing the data) *

	0	1	2	3	
Did not study robustness	○	○	◉	○	Tested in detail the robustness of their finding (e.g. using repeated data perturbations, subsamples, or bootstrapped samples)

Bonus point for a particularly cool visualization (i.e. not just scatter points on a map) (1 bonus point)

- [ ] The author made a really creative map!

Bonus point for a particularly cool analysis (i.e. answering a question of the data not required by the lab) (1 bonus point)

- [x] The author performed a really creative analysis!

Reproducibility

In this section you will assess the reproducibility of the your peer's report. Be sure to take note of any extra README files that explain any extra steps you might need to take to recompile the report. If they have saved their figures in a separate folder, check to see whether there is a script that will automatically produce AND SAVE their figures. If not, take a point off for reproducibility.

Several people will have saved a large file (probably geocoded locations) and used this file in analysis. This is fine if they also provided clear instructions concerning how the reviewer could reproduce this file in an automated way (e.g. by running an R script or calling a function). If they rely on such a file but do not provide

instructions about how one could reproduce this file, then take a point off for reproducibility. You do not need to actually regenerate this file.

Reproducibility of report (4 points) *

	1	2	3	4	
Could not recompile the report	◉	○	○	○	Could recompile the report and figures without manual effort and got the same output as provided in the original pdf

If you could not recompile the report, or got different output, explain what went wrong

You did not include the r or python files necessary to recompile your report. For that matter, I'm not particularly pleased that you used python without asking if it is ok. I have nothing against python, but there are many other people who would also like to do this and it's not fair if some people can and others can't.

I would also suggest thinking about who is reading your report: me, an R enthusiast. Stating that "Python, which is more effective..." is not going to put me in a good mood when reading your report... I would suggest that you see that this class is an opportunity to learn a new language, afterall, you are a statistics masters student, and R is the language used by most statisticians.

Readability of code (4 points) - be sure to look at any files in the R/ folder *

	1	2	3	4	
Code very difficult to read with little documentation	○	◉	○	○	Code easy to read with clear documentation

Suggestions to improve code (either provide specific examples or general comments)

You did not provide all of your code so I can only comment on what I can see.

Overall, you need to provide more comments and take care to have consistent spacing and line length. Specifically, you should always have a space before and after arguments such as " = ", " + ", and after commas ", ".
No line of code should exceed 80 characters.

I do however like that you are using piping!!

Clarity of folder structure (2 points) *

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

Optional suggestions for improving folder structure

Please don't include the data, per the instructions.

Concluding remarks

In this section you will provide some general feedback to the author.

One or more things that you liked about the report overall

Good use of color overall for your figures.

One or more things that could be improved upon

Your report could be substantially more professional. You have phrases such as ""Ummm....nothing special", and have not deleted some of the instructions from the original template. Taking the time to read through and polish your report would make a big difference.

Your file size is also far too large -- please try to avoid this for next time by embedding your figures as png (include dev = "png" in the chunk options).

Any other comments that you would like to add?