Lab 2 GSI

Name of student *

Todd Faulkenberry

Student ID *

3033134639

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

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 *

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

Comments on what you liked about the author's kernel density estimation figures *

I like your color choices, and how you included a separate plot for the extreme cases!

Comments on what could be improved in the author's kernel density estimation figures *

Make sure to include a legend showing the values of the different bandwidths you choose. Although I know from previous knowledge that the bumbier the estimate the smaller the bandwdith if I did not it wouldn't be clear which was which.

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

Comments on what you liked about the author's loess figures *

The figure looks really good! The choice of using gray points is a good one

Comments on what could be improved in the author's loess figures *

I don't love the yellow background, and transparency may have good, although there are not too many points in your figure so it is not 100% necessary

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)

Comments on authors analysis of the 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

Comments on what you liked about the figure(s) *

I think here the yellow background is useful in grouping all the question and title into one figure.

Comments on what could be improved in the author's figure(s) *

The color choices here are somewhat difficult to pick out from one another especially in the first plot where the dark red answer drowns out the other points.

Discovered that the binary encoding should be aggregated (e.g. lat-lon bins) in order to perform meaningful PCA (or other dimension 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 ineffective for binary encoding and used aggregated data instead

Discussed clustering and related these clustering result to geography (3 points) NOTE: point subtracted if lat-long values included in cluster algorithm *

	0	1	2	3	
Did not discuss clustering	○	○	◉	○	Discussed in the detail the clusters found in the data and how they relate to geography and thought carefully about number of clusters

Comments on clustering analysis

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

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

Comments on what you liked about the figure(s) related to clustering and geography *

I like that you don't include any axes in your maps, that would just add clutter to the figure!

Comments on what could be improved in the author's figure(s) related to clustering and geography *

Here it is really hard to tell the difference between cluster 6 and 8 for example

Analyzed the robustness/stability of a finding (3 points) NOTE: partial point 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 findings (e.e. 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 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 point)

- [ ] The author performed a creative analysis

Reproducibility of report (4 points) *

	1	2	3	4	
Could not recompile	○	○	○	◉	Could recompile the report and figures without manual effort and got the same output

If you could not compile, or got different output, explain what went wrong

I couldn't compile because one of your packages would not install in my version of R, but you likely just have a older version and you would've had no way on knowing.

Readability of code (4 points) *

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

Comments on code

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 what the author did well

Your discussion of your findings is really nice. I like how you related it to your everyday life!

Comments on what the author could improve

Obviously, time is the biggest issue here. If you had been able to complete the report I am confident you would have done quite well.
