

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

24092167

Completeness of report *

- ☒ Provided kernel density plots for temperature
- ☒ Provided Loess plots for temperature against humidity
- ☒ Discussed data cleaning for the linguistic data
- ☒ Investigated two survey questions in terms of geography and one another
- ☒ Discussed dimension reduction (e.g. PCA)
- ☒ Discussed clustering the survey respondents
- ☒ Assessed robustness of a finding
- ☒ Provided code necessary to recompile the report (even if you didn't manage to recompile the report)

Readability of report (5 points) *

| | 1 | 2 | 3 | 4 | 5 | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------------|--|
| Narrative unclear and/or difficult to read | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | Narrative very clear and/or easy to read |

Grammar of report (5 points) *

| | 1 | 2 | 3 | 4 | 5 | |
|-------------------------------------|-----------------------|-----------------------|-----------------------|----------------------------------|-----------------------|---------------------------|
| Incorrect written grammar pervasive | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | 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 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | 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 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | 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 that the author adds shadow under the density curve, and that different bandwidths are included in one figure.

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

The figure is very visually appealing overall. I would suggest maybe just add 2 or 3 bandwidths in one figure though. Now it seems hard to compare $bw=.5$ with $bw=1$.

Detail of loess smoothing analysis (3 points) *

| | 0 | 1 | 2 | 3 | |
|--|-----------------------|-----------------------|-----------------------|----------------------------------|--|
| Did not conduct an analysis using a loess smoother | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | 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 | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | Provided clear, relevant and visually appealing figures |

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

confidence intervals are plotted!

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

could set $\alpha < 1$

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) | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | 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

I think the correlation between Q54 and Q55 is probably less interesting than the other example the author gives.

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

*

| | 0 | 1 | 2 | 3 | |
|----------------------------|-----------------------|----------------------------------|-----------------------|-----------------------|--|
| Did not provide figures | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | Provided clear, informative, and visually appealing figures |

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

The messages the figures convey are pretty clear. I like that the author encoded choices to english words.

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

There's a typo in the Q54&Q55 plot (anyone-anyone). When only plotting the geographical distribution of a single answer, it might be better to use filter and set color in geom_point instead of in aes. The legend seems unnecessary. Also there are no captions for the figure, which makes it hard to follow.

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 | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | Found that PCA was ineffective 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 | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | Discussed in detail the clusters found in the data and how they related to geography |

Optional comments on cluster analysis

It might be useful to leave out part of the data for validation. Could also provide a plot for the "continuum".

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

| | 0 | 1 | 2 | 3 | |
|---------------------|-----------------------|----------------------------------|-----------------------|-----------------------|---|
| No figures provided | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | Provided clear, informative, and visually appealing figures |

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

I really like the PCA plots, especially the size of the figures.

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

I think it suffices to show the result $K=3$. The plots for $K=4$ seem a bit irrelevant here.

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 | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | 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)

☐ 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 | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | 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

took some time, but managed to recompile in the end.

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 | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | Code easy to read with clear documentation |

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

Code is easy to read in general, although comments are missing for some chunks.

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

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 *

The author discussed historical and political issues related to some of the questions, and used them to explain the findings.

One or more things that could be improved upon *

I would suggest more time be spent on analyzing clustering results and robustness.

Any other comments that you would like to add?

This form was created inside of UC Berkeley.

Google Forms