Lab 1 Peer Review

Student ID (name of folder) of report you are reviewing *

3033122406

Completeness *

- [ ] Discussed the measurement of interesting variables
- [x] Discussed data cleaning
- [x] Provided a graphical critique
- [x] Discussed finding 1
- [x] Discussed finding 2
- [x] Discussed finding 3
- [x] Provided code necessary for compiling report

Reproducibility of report (were you able to reproduce the report? If not, what was the error?) *

I was able to run everything except for the code defining "df1" starting on line 344. However, I realized this was just test code so it didn't have any influence on the rest of the code (i.e. "df1" was not used anywhere else in the code).

Readability of code (Did the code follow google style guide? Was it well documented?) *

A couple of lines were longer than 80 characters (specifically, lines 396, 398, 401, 408, 411, 447, 450, 477, 478, 480, 497, and 500). The line breaks inside the parentheses on lines 401, 411, 450, 480, 500 were not aligned. Identifiers, spacing, and assignments followed the Google Style Guide. Descriptive comments accompanied the code; however there were some commented-out code that should have been deleted.

Readability of report (Was the narrative clear and easy to read? Or did you find it hard to follow?) *

I understood the author's train of thought, but there were some minor issues with grammar, particularly with tense usage and sentence structure.

Discuss the data cleaning (Was the cleaning described in detail? Were there any inconsistencies in the data that were missed?) *

Yes, the data cleaning process was described well. The author cleaned the data by considering missing values, values of voltage, humidity, incident PAR, reflected PAR, and repeated values in the two data sets. There might have been additional variables that could have been cleaned (such as temperature, especially since it was described in two figures), but overall the author considered the major variables that definitively contained incorrect measurements and/or recordings.

Relevance of figures - excluding findings (Were the figures relevant and discussed in the report?) *

The figures corresponded to part of the data exploration process - when the author explored the pairwise relationships between temperature, humidity, and voltage using histograms. However, the histograms displayed the frequency of values for each of the three variables, which is not the same as exploring pairwise relationships (from my understanding, pairwise relationships would be looking at the relationship between each pair of the three variables). To effectively visualize pairwise relationships, I would prefer graphs such as pairwise scatterplots. Furthermore, since there were no comments on the relevance of the distribution of these three variables, some explanation for including the plots would have been helpful. Nevertheless, it was helpful to visualize the voltage distributions because it highlighted the potential outliers in the data sets.

The incident versus reflected PAR by time plot was a beneficial visualizing tool, even though it didn't exactly correspond to what was discussed in the paragraph preceding the figure.

In general, I think referring to the figures in the body of the report would have been more helpful.

Quality of figures - excluding findings (Were the figures easy to understand? Was there a caption? Were the axes labeled? Were they visually appealing? If not, what would you have changed?) *

Histograms: All figures were easily understood with clear labels but there were no captions or units listed for any of the figures. The figures, albeit generic in style, did the job, and the clear division between each bar in the histogram was useful. While I would have tried to make the histograms look less generic, I'm not sure how necessary that is for the initial exploration stage.

Scatterplot: Instead of using "hamatop" and "hamabot," "Incident PAR" and "Reflected PAR" would have been better because then people who haven't seen the dataset or the readme text file would still be able to understand what variables the labels correspond to. The colors were distinct enough to make out clusters.

Finding 1 (Discuss whether you found the finding interesting. Why or why not?) *

The conclusions that sunlight hours of a day doesn't have much impact on directed or reflected sunlight in addition to a clear segmentation in humidity levels were interesting, but it was hard to deduce these findings from the figures provided.

Finding 1, figure quality (Discuss the quality of the figure) *

The figures were informative, but there colors were too similar. Having the color scale go from one color to another would likely be more helpful. No caption was present.

Finding 2 (Discuss whether you found the finding interesting. Why or why not?) *

There was no explanation as to why the author specifically chose to analyze trends of three variables for node 105. Was something particularly interesting about node 105? It was nice to get a sense of the behavior of these variables for one node, but the finding seems pretty basic and is one step above something found in the initial exploration stage. Also, since no caption or description was present for either figure, it is unclear what the second figure shows. It appears from the title that it simply plots the observations for one specific day (April 29 2004) at node 105. Unfortunately, no explanation was provided for why this specific day was chosen or the relevance of this figure to the finding.

Finding 2, figure quality (Discuss the quality of the figure) *

It would have been more helpful if each of the four plots constituting one figure were stacked on top of each other instead of presented side by side. The current configuration compresses the plots horizontally, making it difficult to see specific points. However, the distribution was still clear, especially in the second figure.

Finding 3 (Discuss whether you found the finding interesting. Why or why not?) *

The finding (that there is a negative linear correlation between temperature and humidity) was discussed in Tolle et al's paper and could have been easily observed by observing pairwise relationships between the variables, which was explored in the Data Exploration section. The figure was helpful for providing a visualization for this fact, but I was a little confused as to whether the points represented every single observation in the data set given that the colors represent hour.

Finding 3, figure quality (Discuss the quality of the figure) *

I think the figure contains way too many overlapping points. Finding a way of plotting a subset of all the points in a equally helpful way would improve the quality of the figure. While it is less of an issue in this figure than in the figure corresponding to the first finding, a different color scale with more distinct colors would have been more helpful. The gridded white background is appropriate.

Any additional comments