Lab 1 Peer Review

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

3033134639

Completeness *

- [x] 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 fully reproduce the report

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

Code was readable. Code was well documented

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

The report was clear and the narrative was mostly easy to follow (with the exception of voltage-based outlier rejection). The authors motivations were clear, although I have some disagreements with their approach and interpretations. There were a few typos and grammatical errors, but these did not prevent from the overall understanding of the report.

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

The interpretation of the raw data was different than my interpretation of the raw data. They interpreted “interior” and “edge” as the distance of the sensors from the trunk, as opposed to two distinct trees.
The reasoning behind the removal specific nodes made sense (node-specific failures).
I did not understand their reasoning behind voltage-based outlier rejection.
I disagree with their outlier-elimination based on calibration min/max ranges. They did not take into account the accuracy error of the min and max (i.e. 100.2 +- 3.5 % is 103.7, not 100.2). Furthermore, they did not attempt to identify scaling behavior of the data which may have put it within the calibration min/max ranges (relevant to the PAR data). This caused them to eliminate all the PAR measurements.

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

The tables and voltage histogram was relevant to their explanation (with the caveat that I did not understand the voltage explanation).
The humidity histogram and density plots chose to characterize the “humidity_adj” variable rather than the “humidity” variable. Given that there was not interpretation of the “humidity_adj” variable in the variable-key, this seems like a peculiar choice. This plot also continued with the misinterpretation of interior vs. edge as related to one tree vs. two tree.
The figures are relevant to the explanation of why they chose to explore humidity-based findings.

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?) *

Overall - the quality of the figures was not exceptional. There were no units on the axis labels. Axis labels used the default variable/column names from the dataframe, rather than readable labels and titles. The figures had no captions or titles. While the figures had a visually appealing color scheme, better trend-labeling and general labeling would have been helpful.

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

I found the first finding to be of little interest. If I understood the argument correctly, the author claimed that humidity follows weather patterns. This is somewhat the banal definition of weather, isn't it?

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

The figures had readable axis labels and titles, but still did not include units. I did not understand the additional information the second figure (sorted median humidity) was meant to provide.
I don't believe box plots were the best method of conveying the relationship the author attempted to present.

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

I found the second finding to be of medium interest. If I understood the argument correctly, the author presents the inverse relationship between temperature and humidity for humidities above 70% ("higher humidity"). The inverse relationship between temperature and humidity is quite well known, but it is always good to validate. Also, the specification of "only in high humidity" is of some interest.

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

The figures had readable axis labels and titles, but still did not include units. I don't believe the least-square line provided additional information. It is unclear why height was added to the figure, since the finding does not seem to talk about this variable (this variable might have been useful during data exploration, but is not relevant to the finding). Nevertheless, I like the use of a color gradient to represent height, but a more distinct color palette could be more visually appealing.

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

I found the second finding to be of medium interest. If I understood the argument correctly, the author presents the inverse relationship between temperature and humidity for humidities below 70%(“lower humidity”) as a weaker relationship compared to “high humidity”.
I feel this finding somewhat inappropriate for several reasons: The others observed a very short date-range (4/27-4/29) and extrapolated based on this date range. Furthermore, humidity is on a scale of 0-100% while temperature is on a different scale, and therefore a non-linear model (which would have combined the last two findings) would have been a more appropriate finding for this case in my opinion.

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

The figures had readable axis labels and titles, but still did not include units. I don’t believe the least-square line provided additional information. It is unclear why the “edge” vs “interior” was added to the figure. While the author does attempt to slightly analyze this variable (unlike the previous finding), they do not provide a plausible scientific explanation for the reasoning behind the distinction between “edge” and “interior” (and again, I believe this analysis is derived from a flawed interpretation of “edge” vs “interior”) .

Any additional comments

I believe the conclusions of this report are derived from a flawed understanding of the underlying data. I have a sense that the author did not fully understand the meaning of each variable. The elimination of all the PAR observations restricted the author to analyzing only temperature and humidity. I believe the author attempted to utilize techniques presented in class, without much justification for many of the plots. Nevertheless, given the base-assumptions, the findings are of some interest, and the plots are visually appealing (albeit - insufficiently labeled).