Lab 1 Peer Review

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

3033122406

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?) *

Yes, the report was reproducible.

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

The code followed the stylistic guidelines, but was not well documented. There were many commented out lines that didn't make it onto the figures or calculations, and the code had minimal comments to see what was actually happening.

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

It was rather unclear how formal this report should be (a formal scientific journal critique-level report vs. class project), but the writing style was rather inconsistent throughout the report (ex. the author interchangeably uses "I" and "we", formatting of each sections, etc.). Furthermore, some sections could definitely merit a lot more explanation and discussion than merely stating what's in the figure/in the analysis.

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

The cleaning process was well documented, but the steps and justifications were rather incongruent and scant. First, the author did notice the voltage readings were vastly different for net and log, but the author ends up employing the same cleaning strategy without looking into the voltage readings causing a lot of the net data to be thrown out without an appropriate voltage transform. While this can be acceptable, this could have been noted in the report rather than applying same steps on the net and log dataset. Furthermore, the reading ranges set forth by the author didn't have much justification - they physically made sense, but no mention of why these were chosen to be such as the voltage and humidity cutoffs were highly reminiscent of Tolle et al.'s paper. Other steps, like analyzing NA columns and combining with respect to nodeid and epoch, were well laid out.

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

While most of the figures were relevant, they occupied too much space. For instance, the histogram plots could be a lot smaller as the size needs to be sufficiently large to see the trend but not too big so that it wastes space. Furthermore, the author described having plotted the pairwise relationships, but there were no figures to justify the claim. The data exploration section also could benefit from having more relevant figures as there was only one.

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?) *

Title and axes were clearly labeled in each figures, but none of the figures had any captions. The histograms were a bit hard to read as the bars were not separated from each other, which made it harder to read- perhaps separating edges would help. Indeed, all the figures were rather too big and took up unnecessary amount of space. Outliers/outlier bins could have been colored differently to show that they were outliers. The last exploration figure of hamatop vs. hamabot was rather confusing, as the time was actually set as the color variable rather than it being an axis; temporal trends would be a lot more effectively visualized if it were plotted along an axis rather than represented as a color. There were also too many data points to make out which points were the relevant ones.

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

The idea behind the finding was interesting, but not the way the variables were analyzed. In the first figure, the author compared incident PAR and reflected PAR with the "hour difference" as colors. First, it was unclear what "hour difference" variable meant or how the author obtained that quantity. Second, it seemed like the author missed there were a lot of "12 hour difference" points clustered together where incident PAR was 0 and the "0 hour difference" points where incident PAR was very high, which may go against the author's claim that time didn't have too much impact on these two variables. In the second figure, it was a bit unclear what's meant by the "segmentation" nor what the actual trend was, but if the author was referring to the stratification, this might be due to the location of the nodes rather than it being an actual trend. Also, very minimal descriptions/discussions were provided.

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

Figure framing was good (title, legend, axis), but the color scheme was very hard to differentiate. Perhaps a color slider that goes from red to blue might be more effective than blue to darker blue. Furthermore, it's hard to see what the actual message of the plot was from this diagram, but that could be attributed to how the author analyzed the findings rather than the figure itself.

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

The observations were valid, but the finding itself wasn't as compelling since the author just plotted each variables with respect to epoch for a single node. This seemed to fall more under exploratory data analysis; while the observations were valid, the author didn't take extra steps to analyze the trends of each covariates over time. Furthermore, the author didn't mention normalization which was indeed done as we see the figures.

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

The figures were rather disorganized. All of the figures did not have any captions, and the x-axes were rather cluttered. First, the author plotted all the covariates within four plots that were put side by side rather than stacked. For such variability within the given experiment time, it'd be a lot better option to see the seasonality in a wide graph than a narrow one. Furthermore, this added on to the fact that there are so many epochs, which actually obstructed us from seeing any daily trends in the node. The author claimed that the author "can tell from the graph," but as the graphs were shown, it's actually rather had to tell, especially since the x-axis was labeled with epoch not time. The normalized graph was slightly better, but still suffered the same problem. In fact, the node id in the second graph was cut off from the PDF.

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

The finding itself was interesting, and the temporal trend was rather visible here compared to the other figures. However, the discussion that followed up was rather short and not as descriptive.

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

The figure was the best among all findings. Good title, axes, legend placement, and the color scheme made sense here.

Any additional comments