Final Project Peer Review

Below is a rubric I will use when grading. Each sub-bullet will be graded on a 6 point scale, with an average of 5 giving a perfect score.

6 - exceptional
5 - meets expectations, few minor errors
4 - adequate, room for improvement, potentially several minor errors
3 - below expectations, major errors
2 - well below expectations, partially complete, multiple major errors
1 - missing / no effort to address

For each section, please provide your peer with 1-3 comments on how their report could be improved. Focus on the most important areas for improvement. Please follow the peer review guidelines from the midterm project: https://jbhender.github.io/Stats506/F20/peer_review.html

- Basic Requirements [20 points]
  - Is the report (if printed) ~2 pages long and less than 3 pages?
  - Is the text between 200-600 words?
  - Are there 1-3 (no more) graphical elements?
  - Are all required sections present?

- Introduction [20 points]: Approximately 2-3 paragraphs explaining what your question is, why it is interesting, and ending with a high level description of the analysis you did (not the results).
  - Quality of motivation
  - Clarity of question
  - Question is about the population, not about the dataset.
  - Quality of high-level description of analysis:
    - High-level = "big picture" e.g. a succinct description of the analysis
    - Does not state results
    - Does not use specific variable names.

- Data / Methods [20 points]: Describe your data source and the methods you used. There should be enough detail here that I could repeat your analysis. Focus on **what** you did, not **how** you did it. Include a sentence with a link to a GitHub repository containing your code.
  - Data source
    - Is the data source (including years, if relevant) clearly defined?
    - If a Monte Carlo study, are the simulation parameters clearly defined?

    - Variables
        - Is it clear what variables were used and what role each played?
        - Are any derived variables defined? For example, if a variable is dichotomized is made clear that "(new concept) was defined as values (less than/greater than) X".
    - Detail
        - Is there enough detail to repeat the analysis?
        - Is the description free of "how" details? For example, ...
            - [Good] "Average energy use per square foot was computed for each Census region and compared visually. All estimates are given with 95% confidence intervals computed using the balance repeated replicate weights as described in the documentation [link]. "
            - [Worse] "I used data.table to multiply building weight by square footage by energy use and then divided by building weight times square footage. This was done by group using Census region in `by` of data.table. I reported this for each replicate weight using long format and then computed confidence intervals by ..."
        - Is there a link to a GitHub repository with code for the analysis?
    - Completeness
        - Are all aspects of analysis from the scripts and results described?

- Results [50 points]: What did you find? This should be the largest section and is where all of your tabular/graphical elements go.
    - Is this section written in a factual manner, with minimal interpretation?
    - Are the results presented pertinent to the question posed?
    - Are the results clearly organized in a logical way?
    - For NHANES and CBECS, is there a "table one" with descriptive statistics about the sample? Is this table organized according to a key exposure or grouping variable?
    - Are all point estimates reported with 95% confidence intervals?
    - Tabular / graphical elements [25 points]
        - Do these help to answer the question posed?
        - Is it clear how the values presented relate to the analysis performed?
        - Do these support the results stated?
        - Are these elements well organized with appropriate mappings, e.g.:
            - For tables, the most important comparisons are across rows
            - For graphs, color / shape / facets are used to emphasize the most important comparisons.
        - Are these elements "polished" with captions, clear axis labels, legends and free of code conventions (e.g. snake_case)?

- Conclusion / Discussion [10 points]: What do your results allow us to conclude about the question you posed? What are the strengths and limitations of your analysis?
  - Is the conclusion supported by the results?
  - Are the strengths and limitations discussed?
    - For NHANES / CBECS, limitations will often be potential sources of confounding not accounted for.
    - For Monte Carlo studies, consider the scope of simulations.

- Code [20 points]
  - Are the files at the GitHub link clearly organized? [5]
  - Does the code follow the style guidelines? [15]
    - https://jbhender.github.io/Stats506/F20/style_guide.html
    - Headers
    - Line length
    - Use of comments
    - Spacing

- Writing Quality [10 points]
  - Is the writing sufficiently clear to allow one to focus on the analysis and results?
  - Is the writing free from errors that could be caught with a word processor such as MS Word?
  - Are any references clearly cited? (Not all reports will have references.)