

**Peer review assignment.** In order to get credit you must fulfill the following components:

1. Submit your project to gradescope on time
2. Also post on github and email Professor Huang the link to the repository, also on time.
3. For each of your two [peers](#) make a copy of this spreadsheet and fill in using instructions below. You get credit if you follow all instructions. This includes providing **specific** feedback on **every category**, and checking that their **code runs**.
4. You will submit on Gradescope two PDFs one for each peer (see notes below about converting this to PDF)
5. Also upload your PDF to [this link](#) . Name your PDF like "row2\_peer17.pdf" if you reviewed the submission on row2 and your peer number is 17.

**Instructions:** For each question give a score (column C) using the scale below, and for each score, provide feedback on what they did well or could improve (1 sentence per question is fine). Important: for each question in column E, if they got full credit, be **specific** what they did to get full credit. If they get less than full, again be **specific** what they need to change to get full credit.

This includes running the code for code-related questions. (the ones not grayed out). If the code did not work or if you had to change something, just say what problem you saw in column D.

When complete, download as PDF (File --> Download --> PDF). **Before you submit open your PDF to see if that all columns for a row are on the same page (if you drag the columns to be wider, they might not be on the same page, and I will have to ask you to resubmit).**

Scale for column C:

*2 - Satisfactory:* This is approximately equivalent to "B" work, where the thought process is logical and justified for most answers, even if not all answers are correct.

*1 - Progressing:* This work may not be correct and not all answers are logically justified, but the work shows thoughtful engagement with the assignment.

*0 - Incomplete:* This would describe missing or hastily performed work that shows little effort.

|               | Assignment requirements   | Score (0, 1, 2) | Did code work/ what problems arose | Explanation of score and suggestions/feedback  |
|---------------|---|-----------------|------------------------------------|--|
| <b>Part 2</b> | 2.1: Include the contents of Part 2. You must address all comments where you lost points. You do not need to get a perfect score on Part 2 to get a Satisfactory grade on this project but you must work to address all comments. | 2               |                                    | Project part 2 is included, and comments were not addressed, however that is not required for the peer review part |
|               | 2.2: At the end of Part 2, include a summary of what you improved from the previous submission.   | 2               |                                    | No summary included, but this was not required for the peer review portion   |

|                                       |   |   |             |  |
|---------------------------------------|---|---|-------------|--|
| <b>3.1 Load and clean data</b>        | Describe steps someone can take to obtain the data files you are working with. For example, provide the link to a the paper's supplementary Excel file, and explain that you opened it in excel and then saved it as a CSV file called "frogs.csv" in the same directory as the notebook (or whatever, as long as what you instruct them to do matches what your code is so your code works). ONLY IN RARE CASES: Only if you check with Dr. Melamed first, and you have some data that requires special permission to access, or some other big hassle to access, you can do this some alternate way. Your data should be directly downloadable from the source and not your personal shared folder. | 2 | Code worked | The link works, the steps are clear, and the code also works.  |
|                                       | Provide code to read in all relevant data files into data frames. Explain your code and why you did it that way. Show the "head" (first few lines/rows/columns) of each data frame.   | 2 | Code worked | Head is shown of the data frames, explanation makes sense  |
|                                       | If any cleaning steps were needed at this point, explain these cleaning steps. Otherwise, explain how you checked that the data frames were suitable for the further analyses.  | 2 | N/A         | There are some NaNs in the dataframe which might need to be addressed for analysis, but otherwise, the dataframe is appropriately formatted. No index was set, but I am not sure if that is essential for analysis.  |
| <b>3.2: Describe data numerically</b> | Provide code to obtain the shape of the data files. Describe how this shape relates to the number of observations and the number of features. Be precise, such as "This data frame has 6000 rows which is the number 500 mice times the 2 treatments times the 6 time points per treatment".  | 2 | works       | Observations and features make sense. It might help to clarify the number of fish of each type or treatment types.   |
|                                       | <b>Feature 1:</b> Explain what you expect the "describe" function would output, based on your understanding of that features. How many observations have a recorded value of that feature and what is the average across observations?  | 1 |             | What they got was correct, I don't know about the reasoning since I didn't read the paper, but there isn't much justification for why the describe function returns those values for the features. I assume this is because the first bone of the maxilla just has much less bone deposition on average because of some factors relating to blood flow and hormones? |
|                                       | <b>Feature 1:</b> Run the "describe" function and compare the results to what you predicted.  | 2 | Works       | Pretty much spot on  |

|                            |  |   |            |  |
|----------------------------|--|---|------------|--|
| 3.3<br>Visual<br>izations. | <b>Feature 2:</b> Explain what you expect the "describe" function would output, based on your understanding of that features. How many observations have a recorded value of that feature and what is the average across observations? | 1 |            | Same as feature 1, there should probably be a reason pointing to why this feature is predicted to be like this   |
|                            | <b>Feature 2:</b> Run the "describe" function and compare the results to what you predicted.   | 2 | Works      | Good prediction  |
|                            | <b>Visualization 1:</b> Describe what kind of visualization you want to make, why this is appropriate for this feature and data set, and how the visualization will provide insight into the data.                                     | 2 |            | A scatterplot is ideal for plotting two types of numerical data and observing trends between them  |
|                            | <b>Visualization 1:</b> Provide code and explain your code to make the visualization.  | 2 | Code works | Perhaps it might be good to include a trendline, since we recently learned the code to plot the linear regression line (I think)   |
|                            | <b>Visualization 1:</b> Interpret the visualization: compare it to the "describe" function output from 3.2, and explain what insight into the data you can make with the visualization   | 1 |            | This was very briefly touched on, and no specific numbers were mentioned. Ideally this will point back to the describe and compare the means and what the visualization tells you that might not be clear from the describe function, like if the data clumps somewhere or if theres a lot of variation. |
|                            | <b>Visualization 1:</b> Describe how your visualization relates to one of the hypotheses or figures from the paper.  | 1 |            | This works, but its really short and it should probably be more specific on how it relates to the figure in the paper  |
|                            | <b>Visualization 2:</b> Describe what kind of visualization you want to make, why this is appropriate for this feature and data set, and how the visualization will provide insight into the data.                                     | 1 |            | I'm not sure if I'm misreading the explanation, but a histogram ideally visualizes the distribution of the data. Since each observation is a different fish, then wouldn't the histogram show the distribution for the same bone, in all fish, rather than different bones in the same fish?             |
|                            | <b>Visualization 2:</b> Provide code and explain your code to make the visualization.  | 2 | Code works | Good visualization, nice use of kde = True   |
|                            | <b>Visualization 2:</b> Interpret the visualization: compare it to the "describe" function output from 3.2, and explain what insight into the data you can make with the visualization   | 1 |            | Very brief, as for visualization 1   |

|   |   |  |               |
|---|---|--|---------------|
| <b>Visualization 2:</b> Describe how your visualization relates to one of the hypotheses or figures from the paper. | 1 |  | Same as above |
|---|---|--|---------------|