

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**

	Assignment requirements	Score (0, 1)	Did code work/ what problems	Explanation of score and suggestions/feedback
Part 2	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	N/A		N/A

	2.2: At the end of Part 2, include a summary of what you improved from the previous submission.	N/A		N/A
	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	2	Code worked when copying the path from where the file is located in my computer.	A score of 2 was given because the steps described on how to obtain the files was thoroughly and clearly explained. The code worked when updating the file path location as given by the instructions. The only suggestion I would make I do not have a windows computer to test out on retrieving the file path but with the MAC, it may be different with other MAC users, but I had to also click 'Get info' to show the information for 'where'.

3.1 Load and clean data	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	The code provided worked and the code provided was explained.	A score of 2 was given because the code provided was explained and the code provided worked. I did not understand why the data types column with df.dtypes was included no explanation for that code was given unless that was inspecting the dataset by each aspect, i.e. rows/columns and in that case you should also print(Columns include Data Types:). Or another suggestion is to move the cell above the inspection of the datasets to right below the rows and before the data type cell. Also you may be able to break up the cell for the rows to print first few rows then the next line have the df.head() to have the dataframe in the table format output. Another suggestion is that I do not know if I would consider the second data set a relevant data file only because it was not used for the visualizations for the features chosen provided in this notebook.
	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	The code provided worked with no corrections needed.	A score of 2 was given because the code provided to clean the data frame removed the Na's found within the data and they explained why and how they wanted to remove the missing values from both datasets. I do not have any suggestions except I did not understand why 20/50 rows were checked after cleaning instead of just the first five like previously.

3.2:	<p>Provide code to obtain the shape of the data files.</p> <p>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".</p>	2	<p>The code provided worked and the code provided was explained.</p>	<p>A score of 2 was given because the code provided worked and showed the shape of the dataset in regards to the observations and features. They also explained how the shape of each dataframe related to the observations/features between the fish species and location sites. I would change the code for the second part in <code>df.shape[1]</code> for the first dataset to say features, unless both are observations and they are not 20 features. I was confused by why the unique features were summarized showing the summary. I could understand if you were cleaning the data further to have the unique data only, but you could potentially show the dataframe shape changed after cleaning instead.</p>
	<p>Feature 1: Explain what you expect the "describe" function would output, based on your understanding of that features.</p> <p>How many observations have a recorded value of that feature and what is the average across observations?</p>	1		<p>A score of 1 was given because the predicted describe function was predicted and explained for the feature chosen for the Total Body length, but I feel that both features were combined and not fully explained. A range of values could have been given for the prediction of TL and why you would see that many observations.</p>

Describe data numerically	<p>Feature 1: Run the "describe" function and compare the results to what you predicted.</p>	1	<p>The code provided worked and the code provided was explained.</p>	<p>A score of 1 was given because the describe function provided an output that described the total body length, but the results were not compared to the predicted values. I would show the comparison of the describe function and the predicted after you run the describe</p>
	<p>Feature 2: 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?</p>	1		<p>A score of 1 was given because the predicted describe function was predicted and explained for the feature chosen for the Pectoral Fin base to Body depth ratio, but I feel that both features were combined and not fully explained. A range of values could have been given for the prediction of PBd and why you would see that many observations.</p>
	<p>Feature 2: Run the "describe" function and compare the results to what you predicted.</p>	1	<p>The code provided worked and the code provided was explained.</p>	<p>A score of 1 was given because the describe function provided an output that described the PBd, but the results were not compared to the predicted values. I would show the comparison of the describe function and the predicted after you run the describe</p>

	<p>Visualization 1: 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.</p>	1		<p>A score of 1 was given because an explanation was given for the type of visualization to make for the TL feature as a box plot to look at the variance of total body length of the fish species. I would have specified individually for each feature the type of visualization, it was only explained that a box plot would be a good visualization.</p>
	<p>Visualization 1: Provide code and explain your code to make the visualization.</p>	2	<p>The code provided worked and the code provided was explained.</p>	<p>A score of 2 was given because the code provided to visualize the feature as a box plot worked with no corrections needed and the code was explained. Previously the instructions advised to install the packages used this was not done with the visualization code.</p>
	<p>Visualization 1: 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</p>	1		<p>A score of 1 was given because the interpretation of the visualization was thorough and insightful for measuring diversity among the species observed regarding total body length. I would relate the median range described as seen in the box plots to the actual output from the describe function for total body length. I would also specify that you are referencing TL for the outliers in comparison to PBd.</p>

**3.3
Visual
izations.**

<p>Visualization 1: Describe how your visualization relates to one of the hypotheses or figures from the paper.</p>	2		<p>A score of 2 was given because the total body length description from the box plot visualization related to the hypothesis with variability within the traits of the fish species. I would specify which of the two hypotheses or if the total body length visualization is related to both hypotheses.</p>
<p>Visualization 2: 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.</p>	1		<p>A score of 1 was given because an explanation was given for the type of visualization to make for the PBd feature as a box plot to look at the variance of pectoral fin base for the fish species. I would have specified individually for each feature the type of visualization, it was only explained that a box plot would be a good visualization of what to expect from the feature.</p>
<p>Visualization 2: Provide code and explain your code to make the visualization.</p>	2	<p>The code provided worked and the code provided was explained.</p>	<p>A score of 2 was given because the code provided to visualize the feature as a box plot worked with no corrections needed and the code was explained. Previously the instructions advised to install the packages used this was not done with the visualization code. Both features were visualized individually but were compared to each other in terms of variability.</p>

<p>Visualization 2: 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</p>	1		<p>A score of 1 was given because the interpretation of the visualization was thorough and insightful for measuring diversity among the species observed pectoral fin base. It was stated that the feature was conserved and should not be used to measure diversity. I would relate the ranges described as seen in the box plots to the actual output from the describe function for PBd.</p>
<p>Visualization 2: Describe how your visualization relates to one of the hypotheses or figures from the paper.</p>	2		<p>A score of 2 was given because the pectoral fin base description from the box plot visualization related to the hypothesis with variability within the traits of the fish species. I would specify that the feature PBd relates to which of the two hypotheses, it seems that it was more focused on the TL feature when stating this finding, instead to include both features to say these findings.</p>