

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.

		Score (0, 1, 2)	Did code work/ what problems arose	Explanation of score and suggestions/feedback
	Assignment requirements			
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 address all comments.	2		This was not required for this assignment, project part 2 is in a separate file, loading it into this file will be easier to relate the figures back to your analysis completed in project part 2
	2.2: At the end of Part 2, include a summary of what you improved from the previous submission.	n/a		This was not required for this assignment

3.1 Load and clean data	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.		1	Code was a bit confusing to follow	The explanation of retrieving the data could be more thorough. It is not clearly stated that the "gene data" is the experimental data and not the meta data and it is not called experimental data in the research paper. You also could make it more clear how you downloaded it or that you saved it into the same folder in jupyter. I went into the link of the paper and looked at the GEO data sets and it is also not clear how you downloaded the data for this
	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.		1	There is no data frame printed so I cannot see the top of the data frame and I was unable to download the data so I am unable to run the code here.	Since I could not read in the file, I trust that a data frame was made because you were able to complete later parts, but printing the data frame head could be helpful too in case someone else runs into this same issue
	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.		1	Code was provided	It is clear what cleaning steps you took but the explanation of why you did them was a bit vague. "index was set" does not tell the viewer why you set the index col = 0 or why you transposed it.
3.2: Describe data numerically	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	Code was provided	The shape and explanation of what the dataframe is telling the viewer based on the observations and the features is logical and thorough.

	Feature 1: 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		The expectation for the describe function could have been a bit more clear on why you expect to these values, and why for both features you expect to see the exact same thing from the describe function
	Feature 1: Run the "describe" function and compare the results to what you predicted.	1	code was provided	You provided the proper code but did not have an explanation on how this result relates to what you presented as an expectation
	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?	1		Similar to feature 1, why would you expect the mean and standard deviation to be at this value? Why are they the same?
	Feature 2: Run the "describe" function and compare the results to what you predicted.	1	Code was provided	Code was provided but there was missing explanation on how this relates to your prediction
3.3 Visualizations.	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.	2		The explanation for why you chose this kind of plot makes sense, and you explained why you used the concat function
	Visualization 1: Provide code and explain your code to make the visualization.	2	Code was provided	The proper plot was the output of this code
	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	1		Your interpretation of the plot was good. It is unclear though, how this relates to the describe function as you included both features from the previous part in this figure.
	Visualization 1: Describe how your visualization relates to one of the hypotheses or figures from the paper.	1		You mention what kind of insight this might provide the researcher with, but do not specifically state how it relates to the hypothesis or a specific research question

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.	2		You explain why you use a catplot, although this could be more thorough, why would you use a categorical plot vs a continuous plot? What about the data suggests that you should do this? A violin plot would make sense for this type of data though.
Visualization 2: Provide code and explain your code to make the visualization.	2	Code was provided	The code was explained well and the axes were properly labeled
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	1		You gave good insight into a potential interpretation of this plot, but did not relate it to what you saw in the describe function because you did not use these features in the previous part where you should run the describe function.
Visualization 2: Describe how your visualization relates to one of the hypotheses or figures from the paper.	2		You related the information observed in the visual violin plot to the hypothesis of the research well and gave a good explanation of what this could tell the researcher about their data