

Project 2

Rubric

Name:

Category	# Points Earned
Proposal	25
Design Elements	20
Execution	25
Code Quality	15
Website Presentation and Discussion	15
TOTAL	100

Proposal (25 points)

_____ (3 points) One of the two following is present:

1. A **hypothesis** is stated that can be tested using data and computation. It is specific enough that you can reasonably evaluate it within the time frame for the project. A hypothesis is an assertion you suspect is true and plan to evaluate methodically. An example is, “Retweeted tweets tend to be longer, on average, than those that do not get retweeted.”
2. A **computation** is described that will take some data and produce some useful output. For instance, “I will write a book that produces a concordance for any text in Project Gutenberg,” would be a good description of a computation, as would “I will search a text (given as a string) for all occurrences of words that appear twice in a row, because these are often mistakes. I’ll apply this to several Project Gutenberg texts to demonstrate that it works.”

_____ (2 points) The hypothesis is placed in the **context** of a problem. Why is the hypothesis interesting to explore? Alternatively, the computation is placed in context: Why is this thing worth doing, or interesting to you, or challenging?

_____ (2 point) There is a brief **description of the data** to be used in the project, and the **data source** is specified.

_____ (2 points) There is a brief description of the **format of the data** (e.g., file type and organization of file, or the URL and the format of the data at that URL) and how it will be loaded by your Python program. It is fine if your program loads data from files or from a URL within your program.

_____ (5 points) The **steps** of the program are numbered, specific, and manageable.

_____ (5 points) **Skeleton code** for functions used in the program, with comments.

- You **must** indicate any inputs and outputs of each of your functions inside the comment under each function definition. We ask that you use the “arrow” notation (as in “`int -> string`”) which will be described in class.

- You **must** indicate which function in your code will be the “main” function that calls any other functions you’ve defined.
- You **must** indicate whether the user will run your analysis (or computation) by evaluating your .py file (e.g., using F5 in IDLE) then calling your main function with the interactive shell, or whether your main function will be called in a statement at the end of your .py file. In other words, tell us how we will be able to run your program. Either of these options is OK.

_____ (1 point) The **skeleton code compiles**.

_____ (1 points) There is a description of how the hypothesis will be evaluated **using the results**. What possible results would be confirming evidence for the hypothesis, and what would be disconfirming evidence for the hypothesis? Alternatively, you must describe how the results of the computation will be presented in a useful format.

_____ (1 point) There is a description of a chart or visualization that will help **present the results**. For example, “I will create a bar chart comparing these three averages.” Consider using Excel to make a visualization of your results. If no chart or visualization seems appropriate for presenting the results, explain why.

_____ (1 point) There is a description of how user input (i.e., using the `raw_input()` function) could be used to make some aspect of the program interactive. If no interaction seems appropriate, explain why.

_____ (2 points) There are **roadblocks** listed. What could go wrong with the steps you outlined? For full credit, no obvious roadblocks are missed.

_____ Total

Design Elements (20 points)

A *bug* is a problem with the code, whereas an *error* might be an problem with the data or the user inputs.

_____ (5 points) Data is read in from one or more files or URLs correctly. Errors or inconsistencies in the data are addressed.

- _____ (5 points) Results of the analysis or computation are either printed to the screen (using `print` statements) or written into an output file in a reasonable format.
- _____ (10 points) Functions and variables are used to break the problem down into meaningful pieces. One could reuse these functions in a different context to compute something different.
- _____ Total

Execution (25 points)

A *bug* is a problem with the code, whereas an *error* might be an issue with the data or the user inputs.

- _____ 25 points to start.
- _____ (\times -1 point) Bugs in the code.
- _____ (\times -1 point) Errors are not handled (either the program returns a wrong value or dies with an error).
- _____ Total (≥ 0 points)

Code Quality (15 points)

- _____ (3 points) Variable and function names are readable and descriptive.
- _____ (5 points) Every function has a header-string comment that describes the type of the function (e.g., `wordlist -> int` or `int * string -> string`) and what the function does (i.e., “return the length of the longest word in the list, or zero if the list is empty”). Variables are commented if needed.
- _____ (5 points) Blocks of code that are run multiple times are not cut-and-pasted, but instead are written as loops and functions.
- _____ (2 points) For functions that could produce incorrect results when given bad input: functions check to make sure that input values are valid.
- _____ Total

Website Presentation & Discussion (15 points)

- _____ (3 points) The website is nicely-organized, and the “front page” has a concise statement of the problem context and the hypothesis, or of the computational goal and its context. There are multiple pages with different content. The data source is referenced appropriately.
- _____ (4 points) There is a concise but complete description of the methods.
- _____ (2 points) The results are presented in a useful way (as a visualization or chart from Excel, as a table, as a screenshot, etc.) as discussed in the proposal.
- _____ (5 points) There is a discussion of the findings or results. Both expected and unexpected results are addressed. Limitations of the methods are addressed.
- _____ (1 points) There is a reflection on the project, including what went well and what was problematic.
- _____ Total