

# Project 1

## *Rubric*

**Name:**

Category	# Points Earned
Proposal	25
Design Elements	25
Execution	25
Code Quality	10
Website Presentation & Discussion	19
TOTAL	104

**Proposal (25 points)**

- \_\_\_\_\_ (4 points) 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. “I will rank all senators” is not a testable hypothesis – it is an activity that might result in some evidence that a hypothesis is true or false. Instead, a hypothesis is a statement one might suspect is true and can evaluate methodically. For instance, “Senators with democratic voting records are more likely to be called ‘liberal’ in the media than senators with more conservative voting records.”
- \_\_\_\_\_ (2 points) The hypothesis is placed in the **context** of a problem. Why is the hypothesis interesting to explore?
- \_\_\_\_\_ (2 point) There is a brief **description of the data** to be used in the project, and the **data source** is specified, including a URL if the data is coming from the internet.
- \_\_\_\_\_ (2 points) There is a brief description of the **format of the data** (e.g., file type and organization of file) and how it will be imported into a spreadsheet.
- \_\_\_\_\_ (10 points) The **steps** of the analysis are numbered, specific, and manageable. “I will import the data and cluster according to votes” is not clear. “I will import all the data for all 893 congressional meetings” is not manageable. Be specific. Break your tasks into reasonable chunks.
- \_\_\_\_\_ (2 points) There is a description of how the hypothesis will be evaluated the **using the results**. What possible results would be confirming evidence for the hypothesis, and what would be disconfirming evidence for the hypothesis?
- \_\_\_\_\_ (1 point) There is a description of a chart or visualization that will help **present the future results**. For example, “I will create a bar chart comparing these three averages.” If no chart or visualization seems appropriate for presenting the results, there is an explanation why.
- \_\_\_\_\_ (2 points) There are some **roadblocks** listed – what could go wrong with the steps you listed? For full credit, no obvious roadblocks are missed. Obvious roadblocks are things like “I want to perform a particular statistical test, but we haven’t covered that formula in class and I don’t know how to do it.”
- \_\_\_\_\_ Total
- \_\_\_\_\_ (0 -100%) Degree of Difficulty: If the problem is trivial (I’ll test whether  $\pi$  is less than 3), the multiplier is 0%. If it’s similar to the senate comparison (but not just a repetition), it’s 100%. If it’s harder than that, we’ll steer you to something more reasonable.

**Design Elements (25 points)**

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

- \_\_\_\_\_ (10 points) Data is imported and formatted reasonably. Errors or inconsistencies in the data are addressed.
- \_\_\_\_\_ (10 points) Intermediate steps (including new formulas) are used to produce the result.
- \_\_\_\_\_ (5 points) The spreadsheet is interactive in some way - users can modify a cell and view a changed result.
- \_\_\_\_\_ Total

**Execution (25 points)**

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

- \_\_\_\_\_ 25 points to start.
- \_\_\_\_\_ ( $\times$  -1 point each) There is a minor bug that doesn't drastically affect the solution.
- \_\_\_\_\_ ( $\times$  -5 points each) There is a major bug that drastically affects the solution.
- \_\_\_\_\_ Total ( $\geq 0$  points)

**Code Quality (10 points)**

- \_\_\_\_\_ (2 points) The initially-opened sheet contains some explanation of what the other sheets contain. Sheets have informative names.
- \_\_\_\_\_ (2 points) The general layout of the spreadsheets is understandable (that is, there isn't one huge table with a complicated function).
- \_\_\_\_\_ (2 points) All cells except raw data use formulas or are labels (or have legitimate explanations for exception).
- \_\_\_\_\_ (2 points) Each cell is clearly labeled by a row and/or column name or with a comment/text box.
- \_\_\_\_\_ (2 points) Cells of interest are formatted nicely (they are colored, they don't allow bad inputs, etc.). These include both input cells and the final "result" cells.
- \_\_\_\_\_ Total

**Website Presentation & Discussion (19 points)**

- \_\_\_\_\_ (3 points) The website is nicely organized, and the “front page” has a concise statement of the problem context and the hypothesis. There are multiple pages with different content. The data source is referenced appropriately.
- \_\_\_\_\_ (2 points) There is a concise description of the methods.
- \_\_\_\_\_ (3 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. Both expected and unexpected results are addressed. Limitations of the methods are addressed.
- \_\_\_\_\_ (2 points) There is a reflection on the project, including what went well and what was problematic.
- \_\_\_\_\_ (4 points) The writing is good and persuasive. Paragraphs are well formed and have topic sentences. Logical arguments are clear and concise.
- \_\_\_\_\_ Total