Project 1

*Rubric*

Name:

<table>
<thead>
<tr>
<th>Category</th>
<th># Points Earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal</td>
<td>30</td>
</tr>
<tr>
<td>Design Elements</td>
<td>20</td>
</tr>
<tr>
<td>Execution</td>
<td>25</td>
</tr>
<tr>
<td>Code Quality</td>
<td>10</td>
</tr>
<tr>
<td>Write-up</td>
<td>20</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>105</strong></td>
</tr>
</tbody>
</table>
Proposal (30 points)

_____ (5 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 points) 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.

_____ (4 points) What potential biases exist within your data? How will your results be affected by these potential weaknesses? Biases are inevitable, but minimizing them and being aware of them are key.

_____ (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.

_____ (10 points) The steps of the analysis are specific and outlined roughly like what you did in Homework #2. Completely polished pseudocode is unnecessary; we just want to see that you’ve given thought to the exact, concrete steps and have a plan, and the more specific you can detail to us, the better we can help you and gauge the difficulty of your proposed project. “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 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

_____ Category of difficulty: We will qualify your problem as being either (1) too easy (you need to revise it, based on our comments and talking with us); (2) on the easier side but sufficient; (3) great and appropriate; (4) potentially on the difficult side for a 3-week project; (5) too difficult (you need to revise it, based on our comments and talking with us).
Design Elements (20 points)

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

_____ (8 points) Data is imported and formatted reasonably. Errors or inconsistencies in the data are addressed.

_____ (5 points) Intermediate steps (using variables and appropriate data structures like lists and dictionaries) are used to produce the result.

_____ (7 points) Your approach for quantifying your findings (i.e., your metric) makes sense and is appropriate for addressing your hypothesis.

_____ 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.

_____ (× -1 point each) There is a minor bug that doesn’t drastically affect the solution.

_____ (× -5 points each) There is a major bug that drastically affects the solution.

_____ Total (≥ 0 points)

Code Quality (10 points)

_____ (2 points) You have a proper entry point into your program, via `if __name__ == '__main__':`

_____ (2 points) You use comments appropriately (i.e., not an overwhelming amount and not just 2 comments in the entire program)

_____ (2 points) Overall style: You use good variable names and whitespace in between paragraphs of code

_____ (4 points) You use functions and data structures appropriately

_____ Total
Write-up (20 points)

_____ (3 points) Document is organized well and it is easy to discern what your project is about

_____ (2 points) There is a concise description of the hypothesis and what it would take for results to support or refute your claim.

_____ (3 points) The results are written in an easy-to-follow, digestible, descriptive build-up and your results are clearly indicated. If you had a graph or chart, its presence makes sense and doesn’t appear distracting.

_____ (5 points) There is a discussion of the findings. Both expected and unexpected results are addressed. Limitations of the methods are addressed.

_____ (3 points) There is a reflection on the project, including what went well and what was problematic.

_____ (4 points) The writing is good and persuasive (shouldn’t be fragmented in thoughts or have egregious, distracting errors). Paragraphs are well formed and have topic sentences. Logical arguments are clear and concise.

_____ Total

Extra Credit

_____ (???) You should first focus on completing each of the above criteria, but if time permits, feel free to expand your project in a way you deem meaningful. For example, maybe another small analysis that is a slight variant from what you already computed would be useful. Maybe you’re great at making graphs, and you can create a compelling, relevant plot which deserves more points. Or, you gave serious attention to the pitfalls of your results and potential weaknesses, attributing it to having bad data, and you spend an hour cleaning up the data and removing certain fields of data that affected your program – evident by re-running your program with your improved data.

_____ Total