**G440/540 Final Project Rubric**

NOTE: We have learned three uses of coding: 1) cleaning/tidying your data and preparing it for analysis; 2) Exploratory data analysis; 3) analyzing the data (often using statistical methods) to answer a specific question about the patterns contained in the data and what we can learn from them.

To achieve full marks for this project it must contain all three elements. You may strongly focus on one aspect, with slightly reduced focus on the second aspect, and a smaller focus on the third aspect.; or, you may present an equal focus on all three. If you strongly focus on one aspect there must be substantial work presented. If you are focusing on cleaning and tidying your data, you must have a good and clearly explained reason for doing so. For example, if you have a lot of data to process (many, large files with multiple issues of missing data, incorrect units etc). If you only have one dataset that does not require a lot of pre-processing, you must instead focus on extensive EDA (that leads you a long way to answering your question or posing new analyses) or an extensive statistical data analysis.

|  |  |  |
| --- | --- | --- |
| **Section** | **Points** | **Marks for:** |
| **Introduction** | **10** | * Clear out line of the purpose of your proposed project with aims you wish to address and questions you wish to answer. For example, an aim might be to tidy and clean your data in a certain way. The questions will be related to something you want to learn from the data by analyzing the patterns in the data. * Supporting literature or references (including websites) if any |
| **Methods** | **25** | * Appropriate analysis for your project design 🡪 adequate to meet the aims/questions you defined in the introduction * Clear explanation of all the methods and functions you will use, and why * The number of different methods, functions and libraries that you use (appropriate for your project) * Description of the data you are using, including where you got it from, its attributes/characteristics, which sites, why these data are appropriate. |
| **Results** | **60** | * Clear, efficient and commented code (for example if you are reading in multiple files you should do so in a loop). * Analysis that corresponds to the methods you have proposed. * Appropriate and accurate implementation of your analysis methods * Well formatted plots and analysis outputs that clearly show all the results of your analysis. |
| **Summary and discussion** | **5** | * A discussion of what you have learned from your analysis * Explanation of any pitfalls you encountered and how you addressed them. |
| **TOTAL** | **100** |  |