| **Requirement** | **Met Requirement** | **Element Fulfilling the Requirement** |
| --- | --- | --- |
| Proposal |  |  |
| Motivation for performing this analysis. | Yes | Proposal described motivations |
| Likely data sources. | Yes | Proposal included likely data sources |
|  |  |  |
| Final Project |  |  |
| Has recognizable “data science workflow” | Yes | 1. Plan 2. Acquire, clean, transform individual data sets 3. Combine datasets 4. Analysis 5. Conclusions 6. Statements for possible future expansions 7. Presentation work |
| Includes data from at least two difference types of data sources   * + relational or CSV   + Neo4J   + webpage [scraped or API]   + MongoDBetc   + etc. | Yes | Relational or CSV   * 2015-16 Proficiency….csv from data request * 2015-16 Entity Demographics.csv downloaded from [www.mischooldata.org](http://www.mischooldata.org) * 2015-16 Entity Data from data request * 2016 US Poverty Thresholds.xls from [www.census.gov](http://www.census.gov)   Webpage [API]   * <https://api.census.gov/data/2016/acs/acs5/profile> |
| Includes at least one (for each of the following): |  |  |
| Data transformation | Yes | Most all of the data for school proficiency (section 2.1) |
| Statistical analysis that describes or validates the data | Yes | Summary function was run on each separate data set to validate categories, ranges, and spread. |
| Graphic that describes or validates the data | Yes |  |
| Statistical analysis that supports your conclusions | Yes |  |
| Graphic that supports your conclusions | Yes |  |
| A feature not covered in class | Yes | * Linear regression (section4); * embedded values in Rmarkdown narrative (section 3.1) * point transparency (section 4) |
|  |  |  |
| Presentation |  |  |
| Presentation was delivered in the allotted time (3 to 5 minutes)   * Presentations exceeding 7 minutes will receive no higher than 50% |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Code & Data |  |  |
| Code and data were delivered where they are self-contained   * Preferably via RPubs and GitHub | Yes |  |
| Results are fully reproducible with what has been delivered   * Full credit will not be given if code references data on a local machine! | Yes |  |
| All delivered code runs without errors | Yes |  |
| Code and conclusions were delivered using a “reproducible research” tool such as RMarkdown | Yes |  |
|  |  |  |
| Deadline Management |  |  |
| Project Proposal, Project, and Presentation were delivered on time   * *Any part of the project that is turned in late will receive a maximum grade of 80%*. * Please turn in your work on time! You are of course welcome to deliver ahead of schedule! |  |  |