

CSC 230 Requirements Document

Current Understand of the requirements of the project:

- Create software that makes it easy for a political science student to generate code that runs a statistical regression on some given data
- Requirements:
 - Ease of Use
 - Students may not know R so the interface must be simple and should guide the students
 - Education
 - Want the students to understand the “why” of the outputs, not the outputs themselves. NEED tooltips and explanations
 - Code Visibility
 - Must display the exact R code used so students can learn and reuse code
 - Statistical Clarity
 - Want the students to understand the context of their data via p-values, coefficients, correlations, etc.
 - Data Types
 - Should automatically detect the type of data: nominal/ordinal/ratio/interval, along with explanations (should also have the ability to manually override selection)
 - Outputs
 - Tables, plots (types of plots), and the downloadable R script, marginal effects plots
 - Make sure to have libraries imported features
 - Scope
 - Want to focus on common regression models (OLS, logistic, multinomial logistic, multinomial, ordered, ordered logistic, Poisson, negative binomial for starters)
 - How to interpret these models
 - Type of variable (nominal, ordinal, continuous, etc.), may not get the regression required for the variable
- Requirements simplified:
 - Upload csv/xlsx
 - Detect variable types
 - Suggest regression model (OLS, Logistic, etc.)
 - Run curated R templates with plumber
 - Show tidy results with glossary explanations
 - Run locally via Docker
- Constraints/Considerations:
 - Students will have limited technical background
 - Data sets will come from excel or csv
 - Tool needs to be able to run in browser, so no complex installations