# Project Description

Econometric analysis plays a vital role in decision-making across various fields, including economics, finance, social sciences, and public policy. Conducting econometric analyses often involves complex data manipulation, statistical modeling, and graph visualization. Many economists and policy analysts lack extensive programming or software engineering skills, which can be a significant barrier to efficiently conducting econometric research. While econometricians may program in statistical software, often, the code is written in a 'quick and dirty' manner, without long term maintainability in mind, leading to code that is impossible to change or debug.

Despite this, most popular econometrics programs are little more than a menu wrapped around a command line. While some point-and-click style programs with Graphical User Interfaces *GUIs* exist – for example, Gauss, SPSS and SAS - they are unpopular because econometricians need to document every data transformation for replicability in empirical research. Point-and-click GUIs make this documentation challenging since they lack a proper way to track every action taken. The naive route most GUIs take is to log every action, however, this leads to messy documentation. In contrast, command-line systems usually rely on a single script to record all steps, making documentation straightforward.

The proposed project aims to develop a novel GUI tailored to the unique needs of econometricians. The GUI will promote:

1. **Modularity**: Since many econometric workflows share common transformation and analysis steps, the GUI should promote reuse, making it easy to save and reload specific analysis units. This would also make debugging easier - one could test and debug each analysis unit separately.

2. **Reproducibility**:Additionally, the GUI should allow users to document and share their pipeline easily, so others can easily reproduce their work.

# User Personas/Stories

## Bob

### Overview

Bob is an econometrician working for the public service. He has a Masters in applied economics. Bob is relatively comfortable with technology - he is able to install and setup software on his computer, and has some basic programming knowledge. However, he has not formally studied computer science, and has obtained most of his programming knowledge through online videos and copying code from examples. Bob often uses a mixture of R and Excel to conduct regression analyses for his work.

### Journey and Pain Points

Bob wishes to analyse the effect of green ratings on house prices in a particular district. His dataset spans several years and is very large. Bob finds it easier to manipulate such large datasets in R, so he opts to use R for his analyses. Bob decides to start with a simple linear regression and ordinary least squares estimation for his analysis, however, he cannot remember how to run the regression in R. He pulls up several older projects until he finds one that uses linear regression, and copies the code in. However, all the variable names are wrong, and he has to spend sometime fixing these up. Finally, the code runs, however, Bob wants to try several other regressions and do some further analysis. Several hours later, Bob runs into an error he has never seen before. His code is a tangle of bits and pieces copied from different projects, and he does not know where to even start debugging. Eventually he gives up and starts again from scratch. Eventually, Bob gets the results he needs. As a public servant, he needs to ensure his results are quality assured, and part of that means having well documented and commented code. Having struggled to get the code to work in the first place, Bob struggles even more to document it, not remembering what any of the functions from older projects do. Eventually, he manages to figure out what most things do, but he wishes there was an easier way.

## Sally

### Overview

Sally is a 20 year old undergraduate studying psychology at the University of Cambridge. She is comfortable with technology, but has no prior coding experience at all.

### Journey and Pain Points

As part of one of her courses, Sally needs to study the effect of genome sequencing and family history of disease on anxiety and depression in patients. Being an overworked student, Sally has left this assignment to the last minute, and only now realizes that the work is due in a couple of hours. She has never done such regression analysis before only has a vague understanding of R based on resources from her lecturer. She gets started, initially making good progress based on examples from her lecturer, however, as soon as the analysis becomes more complex, she looses track of what she is doing, and ends up with spaghetti R code. Desperate to submit her work on time, she asks a computer scientist for help, who manages to make her regression analysis work. Sally submits her code without actually understanding most of what she has done, and wishes there was a more beginner friendly platform to learn the ropes. She brings this up to her lecturer, who suggests some higher level point-and-click platforms, but says there's no way for him to use these platforms for assignments, because he cannot grade work done on them using an autograder. Theres no way for her lecturer to easily replicate students work done on such programs, because the program doesn't provide an easy documentation of exactly what was done.

# Requirements

## Functional

- The program will allow users to import data in .xlsx and .csv formats
- The program will allow users to perform simple linear regression and multivariate regression
- The program will allow users to use ordinary least squares and maximum likelihood estimation methods
- The program will allow users to change axis names, graph colours and point marker shape and size.
- The program will allow users to save parts of or all their work in a single, independent .template file, that can be run on any other instance of the program.
- The program will allow users to export their results as a pdf.

## Non Functional

- The program will use a statistical package to do regression and estimations
- The program will use a React Native package to implement drag and drop functionality
- Users should be comfortable with technology and using / installing software on their own, but don't necessarily need to have any programming experience.
- The program should prioritise utility and learnability. In the context of econometric analysis, efficiency is less of a priority.