

# Project1

Ruth Walters, Micheal Purtle, William Buckley, and Sidney Gehring

2025-02-05

```
# Import dependencies
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.1
## corrplot 0.95 loaded

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.4.1

theme_set(theme_bw())

# Import data
mobility <- read.csv("mobility-all.csv", header = TRUE)
```

## Introduction

*Write four to five sentences introducing the research problem and describing specific research hypotheses. Cite any information sources in parentheses or foot- or end- notes.*

### Research questions:

1. Which variables are the most important variables for predicting economic mobility?
2. To what extent do measures of better education predict higher levels of economic mobility?
3. To what extent do measures of integration across social groups predict economic mobility?
4. To what extent do variables which can be directly affected by government policy predict economic mobility?

## Exploratory data analysis

*Visually and numerically investigate which variables seem associated with economic mobility? - Examine the (predictor and response) variables univariately and multivariately. You will likely not be able to include all of*

*the plots, think carefully about which ones would be good to include. - Are there any variables that you would consider transforming based on the plots?*

## **Model selection**

### **Initial modeling**

*Start by building a multivariate linear regression using the covariates to predict mobility variable. Address the specific questions of above when building the model. Be sure to justify the choices you made in building this initial model*

### **Diagnostics**

- Are the basic assumptions met for your multivariate linear regression model? Why or why not?
- What transformations do you choose (if any)? Why?
- Are there any outliers in your sample overly influencing your model? Identify any outlier candidates and decide whether or not to remove them. Give details.
- Do you exclude any variables? Why? All exclusions/inclusions must be justified

### **Final model selection**

### **Model results**

Create a table that summarizes your final model (coefficients, standard errors, confidence intervals, p-values). Provide interpretations of all your coefficients in the context of the problem. Be sure to address the specific questions of the client (above).

### **Discussion**

What are your conclusions? Identify a few key findings, and discuss, with reference to the supporting evidence. Can you come up with explanations for the patterns you have found? Suggestions or recommendations for the client? How could your analysis be improved? (6-8 sentences)