Data Wizards (Group 4) Project 1 R Notebook

Di Chen   Mai Castellano   Tyler Kussee   Spencer Hutchison

Introduction

The American Community Survey is an ongoing survey that provides vital information on a yearly basis about our nation and it's people. The information from the survey generates data that helps inform how trillions of dollars in federal funds are distributed.

With the data from the American Community Survey, the Data Wizards have come up with the following 3 questions they would like to answer:

- Does higher education mean higher income accounting for sex and age?
- Can we predict household income based on education, occupation, and race/ethnicity?
- Is there a relationship between race/ethnicity and home ownership?

In particular, we'll be focusing on the first question to get started.

Obtain the Data

In the case of the ACS dataset, we'll pull it via an API. We've already declared our key earlier, so now we just need to pull our dataset with the variables and filtering we'd like. In this case we'll want to pull our Age (AGEP), Education Level (SCHL), sex (SEX), and Total Person's Income (PINCP) while filtering to the year 2022 and the State of OR.

Scrub the data

Explore the data

Reference on Variables for now

Model the data

Interpret the data