Shiny Project Proposal

Macroeconomics, Education, & Marriage

Project summary:

In this project we will explore macro level data of economic trends and education with an intent of determining how much, if at all, these factors affect marriage, marriage tenure, and divorce rates at scale. We want to know if there is an income and educational level that increases the probability of a person between the ages of 21 and 45 getting married and staying married for 20 years or more; as well as, how much do macroeconomic trends influence marriage rates. The goal is to create a model that can answer these inquiries. We will look at **the census, bureau of labor and statistics** and **CDC** data.

We will also pull in some micro level data from surveys. The challenge here will be cross referencing this data appropriately and avoiding ecological fallacies. The **ecological fallacy is a logical error where characteristics of a population are incorrectly attributed to an individual member of that population.** Other factors such as gender, age, and race will be considered to determine if things shake out differently for certain demographics of people. If possible, I will remove outliers such as celebrities from the datasets because these individuals are extreme outliers in terms of income, lifestyle, social interactions and public persona. Some other issues involve incongruent datasets and timelines.

Motivation:

I have chosen this project because I am curious about how income, macroeconomic shifts and education level affect individuals and communities when it comes to getting and staying married. As someone who has been married I would like to know about the larger factors beyond our own individual characteristics (personality, emotional state, physical appearance) that determine the likelihood of getting married, staying married and getting divorced.

This data could also be useful to industries by helping them determine how they should market their goods or services to potential customers. They could potentially use this data to determine how they should prepare to position their offering to win, serve and retain customers now and in the future based on the demographic they want to target and the current macroeconomic environment and shifts that are on the horizon.

MVP:

Initially I would like to build an app where you can input things like your income, education level, race, age, and gender and have it provide a percentage on your likelihood of getting married, staying married, and getting divorced while accounting for the current and possibly future macroeconomic environments (we may also need to consider if the person has been married before and/or has children). Hopefully this doesn't get too hairy. If I can accurately do this I would love it and the work would be worth it.

Schedule:

1. Get the Data (02/06/2026)
2. Clean & Explore the Data (02/10/2026)
3. Create Presentation and Shiny App (02/14/2026)
4. Internal Demos (02/17/2026)
5. Midcourse Project Presentations (02/21/2026)

Data Sources:

- Bureau of labor and statistics - https://www.bls.gov/
- Census - https://www.census.gov/
- CDC - https://www.cdc.gov/index.html
- Bureau of Economic Analysis - https://www.bea.gov/
- ADP Research - https://www.adpresearch.com/

Known Issues and Challenges:

- This analysis will only examine divorces among people who got married. This may bias results because individuals who chose to marry may differ from those who did not.
    - **Managing** - include a disclaimer in my project
- Causation vs. Correlation. I cannot say that these factors “caused” a person to get married or divorced since this is observational data. I can say that factors are associated with higher marriage rates or lower divorce rates.
    - **Managing** - be careful with the language I use. No “matter of fact” claims.
- Ecological fallacy - the idea that individuals with higher incomes or education levels have lower divorce rates so people who are highly educated with high incomes are less likely to divorce.
    - **Managing** - not sure here. The world has taught us to live on ecological fallacies.
- Data compatibility - The data from different data sets will not be congruent. Data cleaning and data manipulation will be necessary.
    - Data may stop at different times in different datasets
    - **Managing** - pick specific timeframes such as from 2013-2020.
- Use probability instead of words such as likelihood.
    - **Managing** - pick one definition and stick to it.