What important fields and information does the data set have?

What are its limitations i.e. what are some questions that you cannot answer with this data set?

\*What kind of cleaning and wrangling did you need to do?

Any preliminary exploration you’ve performed and your initial findings.

Based on these findings, what approach are you going to take? How has your approach changed from what you initially proposed, if applicable?

Important information that the data contains are age, gender, work class, occupation and education level. These factors help to create a profile which can be further analyzed to increase predictability for a predetermined income level. The use of character defining traits produces more efficient training of data sets further strengthening algorithms when it comes to testing.

However, the data set does provide some limitations. The absence of specified states/cities in the survey makes it impossible to determine which regions have the highest income level. This piece of information could have further aided the algorithms in determining if an individual makes over a certain amount of income per year. Also, the knowledge of state tax levels would help us to determine which areas of the U.S. did individuals retain more of their earnings.