Exploring Adult Income

Identifying strong predictors of incomes over $50,000

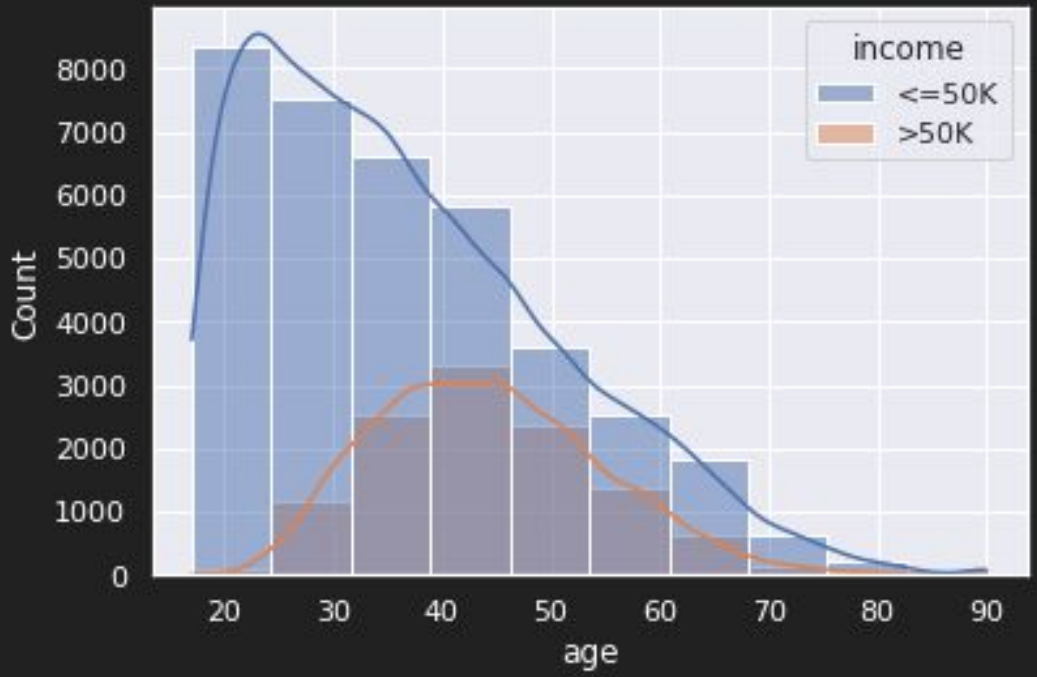
The Data

- Contains individuals described through various **demographic** and **socioeconomic** factors
- Our goal is to predict whether each individual makes over $50,000 in annual income

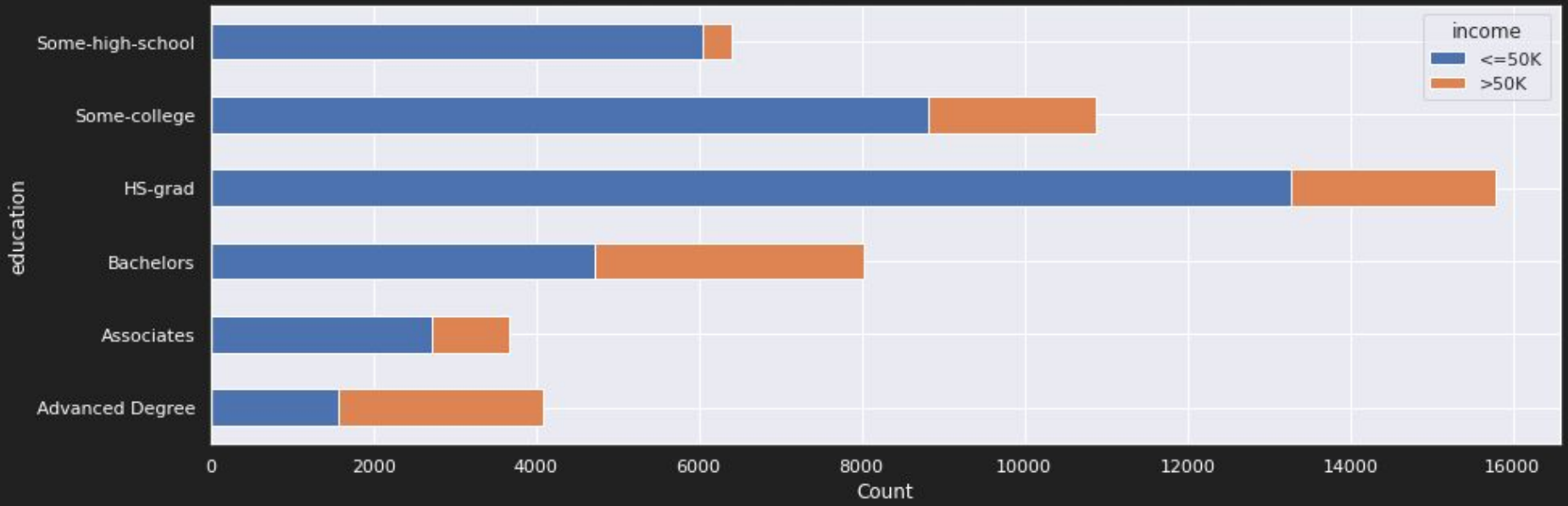
What are the given features?

- Age
- Work Class
- Education
- Marital Status
- Occupation
- Race and Gender
- Hours Worked Per Week
- Capital Gains and Losses
- Native Country

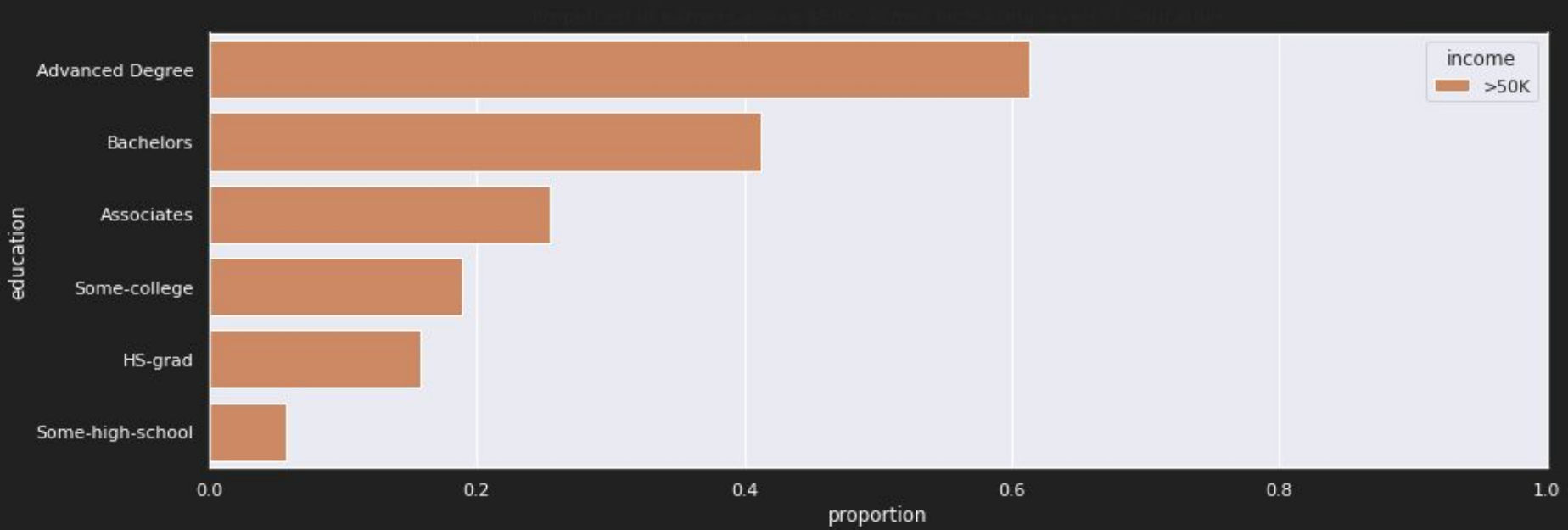
Do older people tend to make more money?



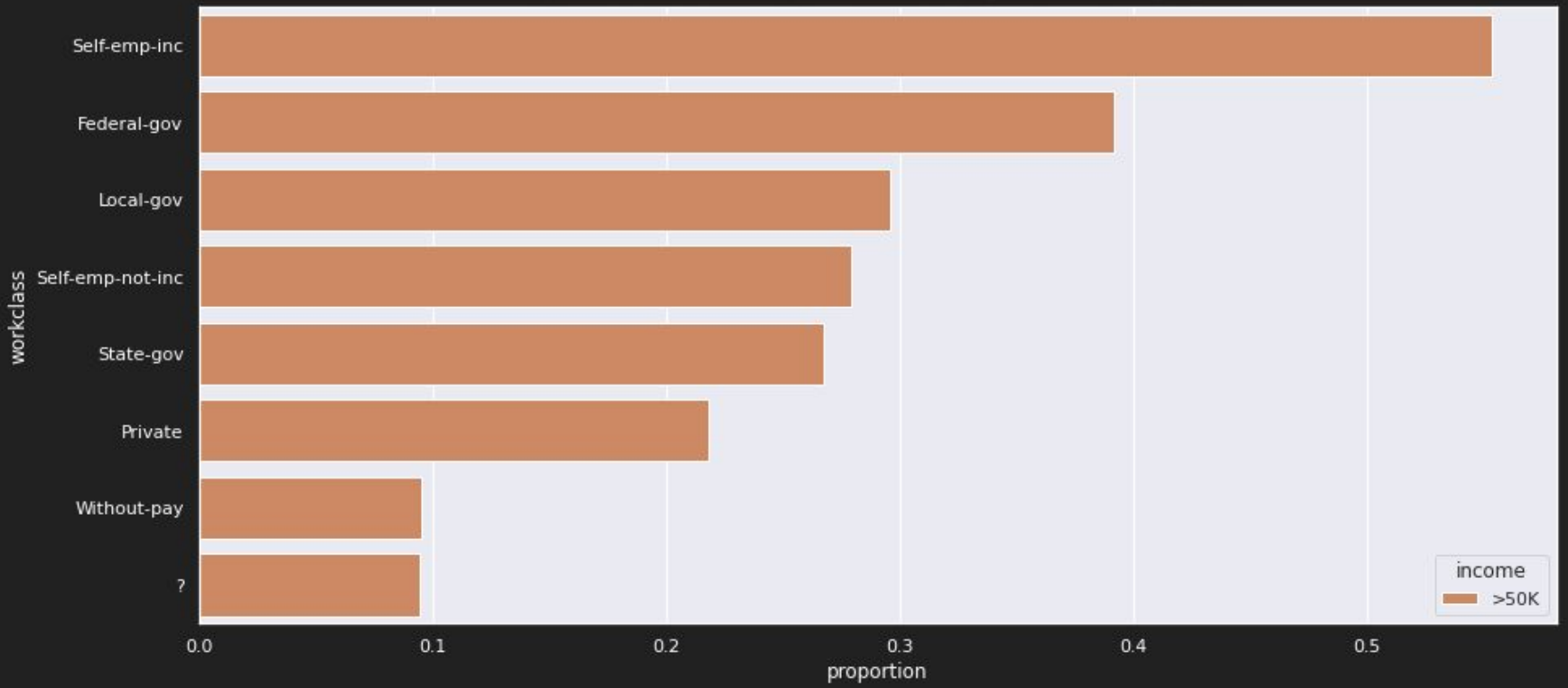
Does higher education yield higher earnings?



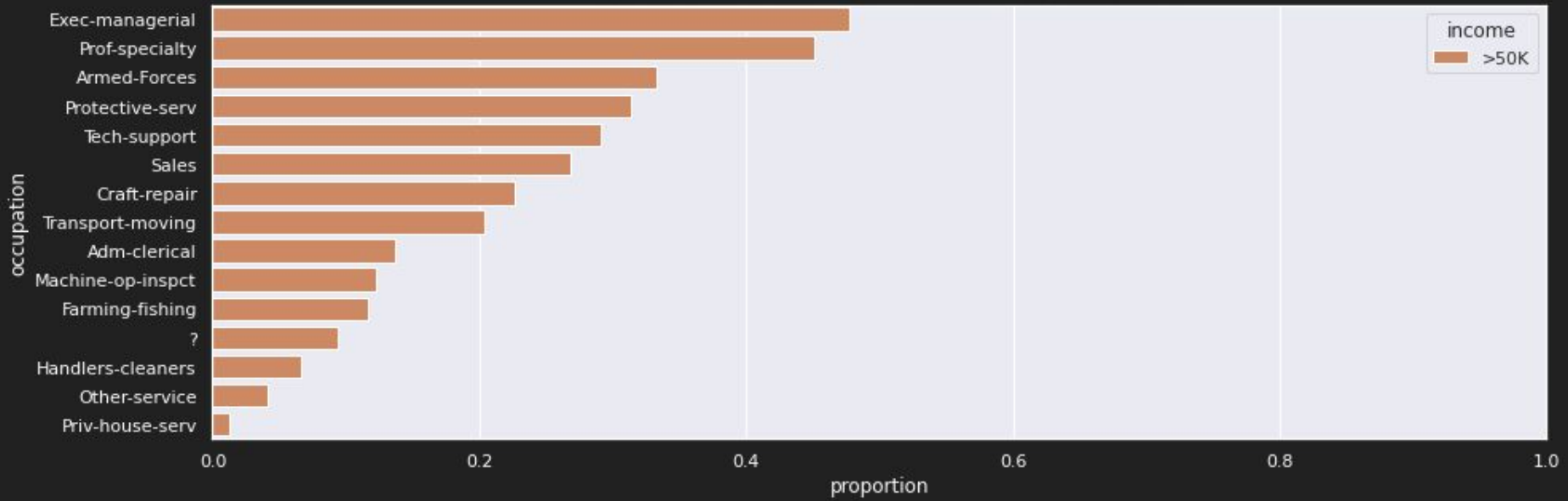
Higher education pays off



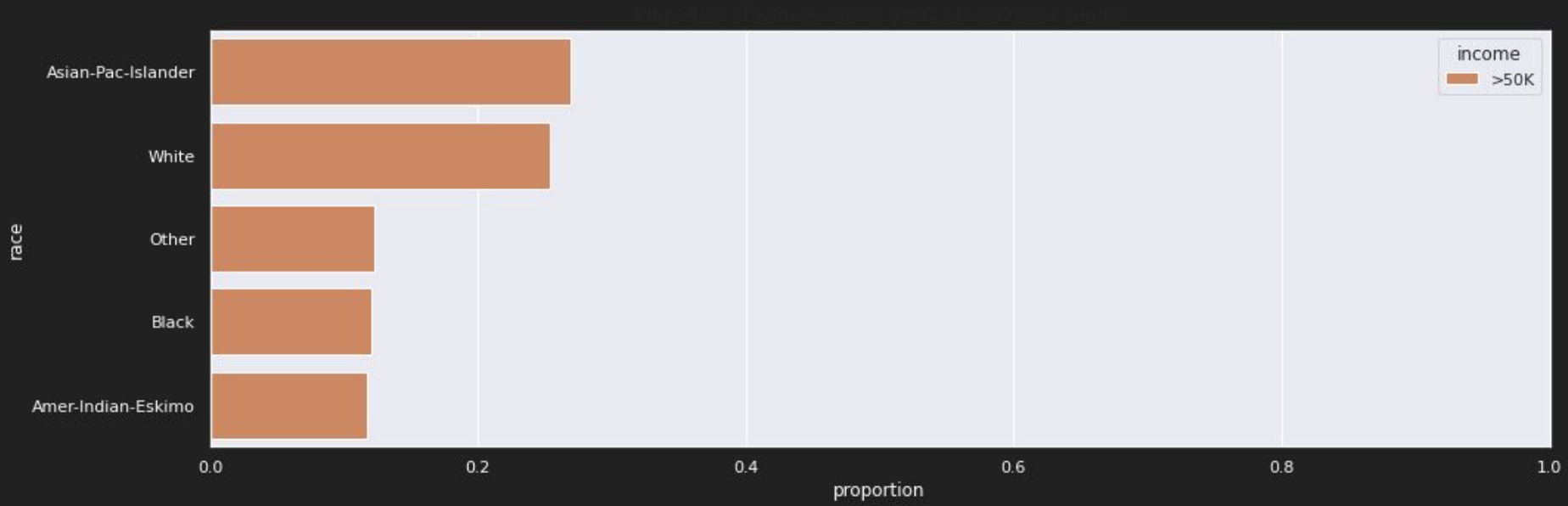
Which type of working class makes the most money?



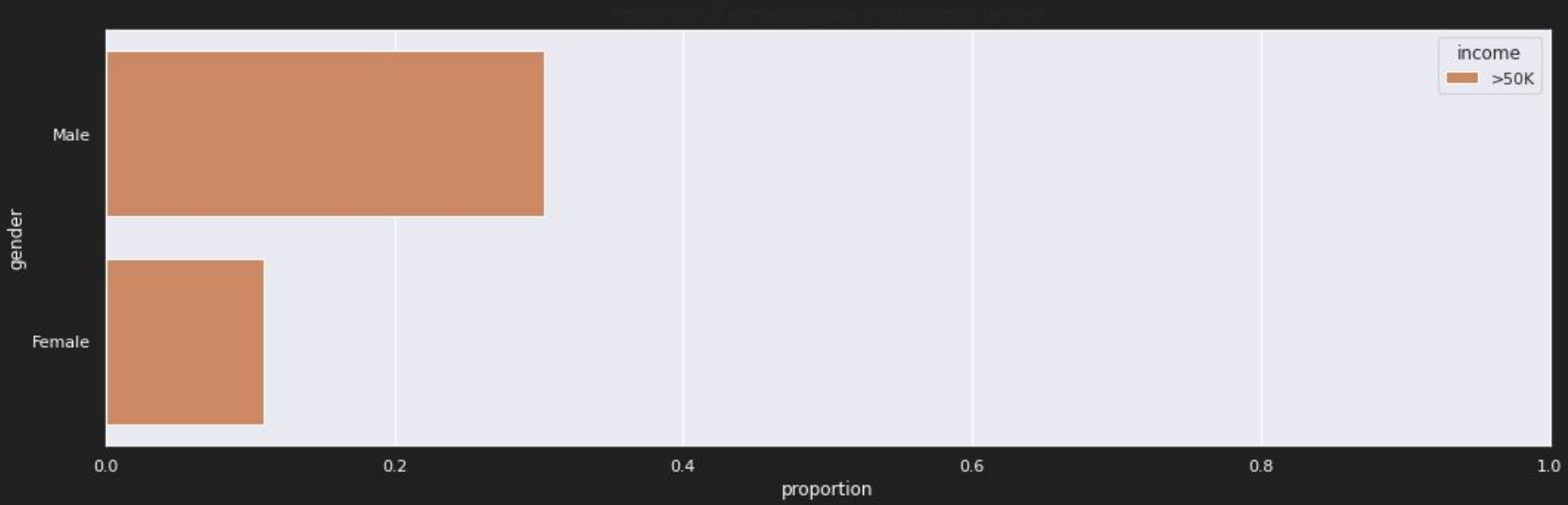
Which type of occupation makes the most money?



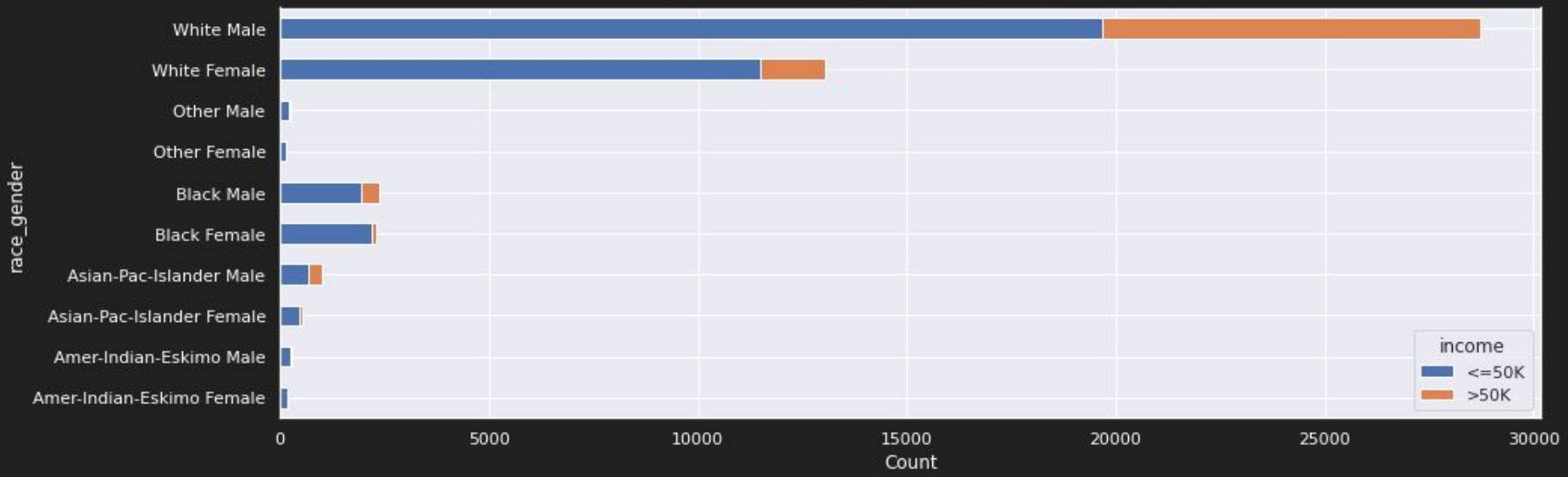
How does race affect income?



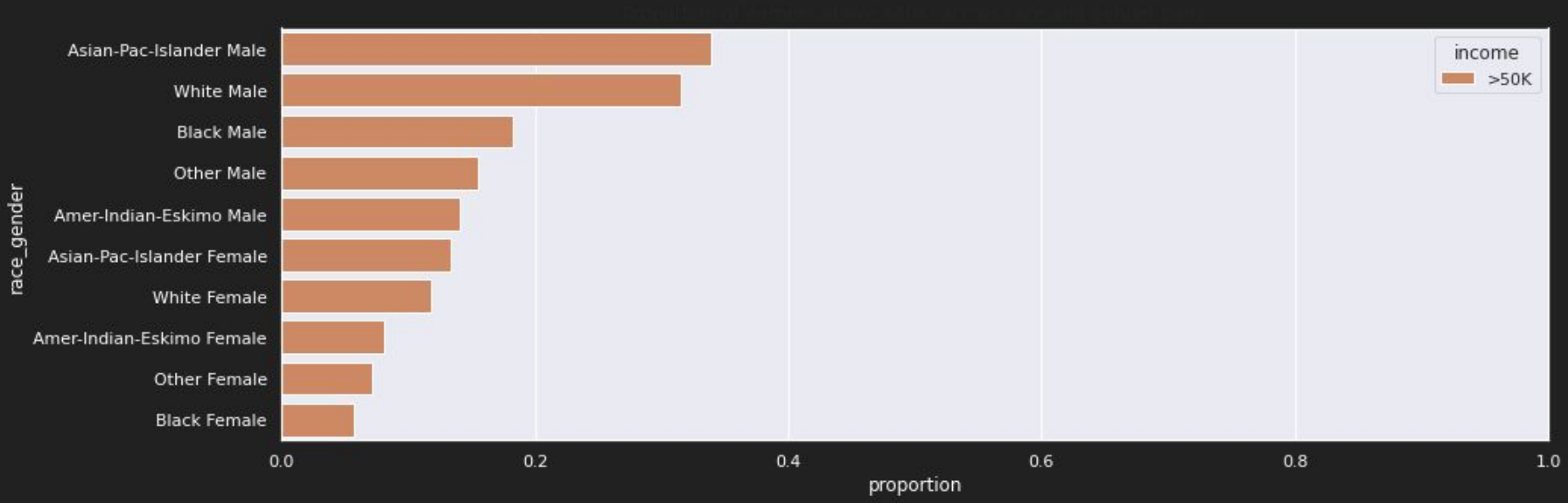
How does gender affect income?



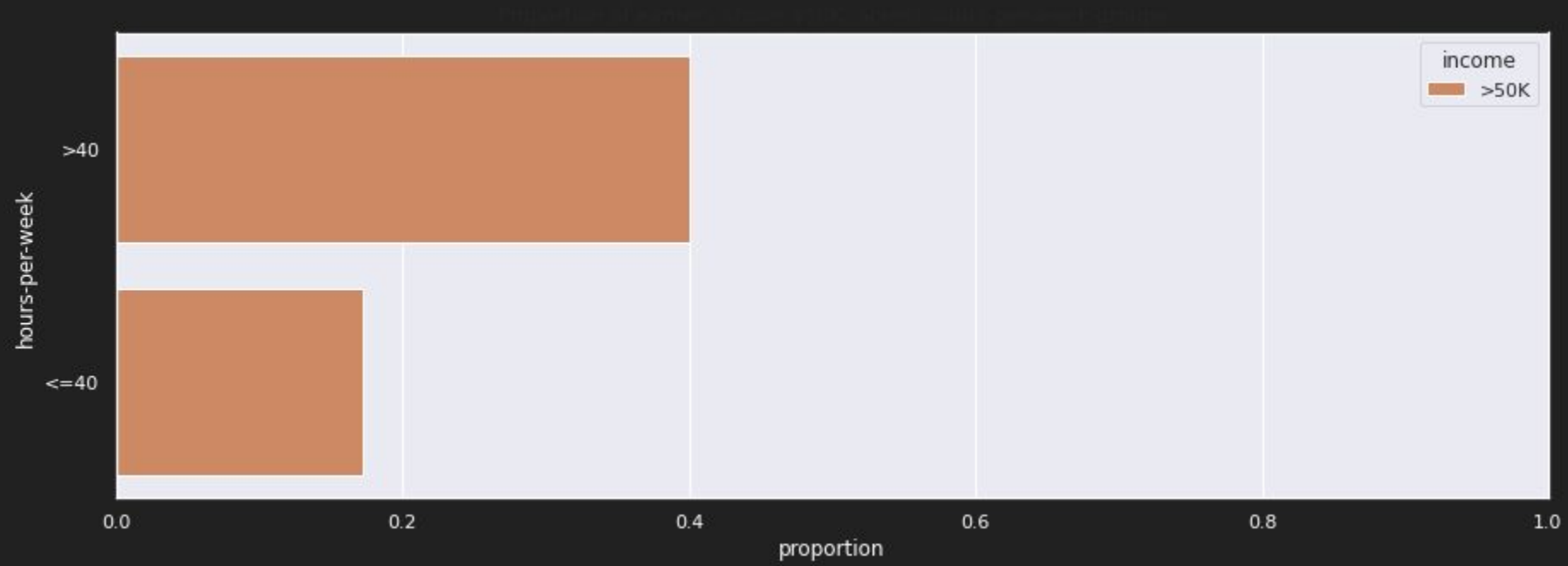
Race and gender



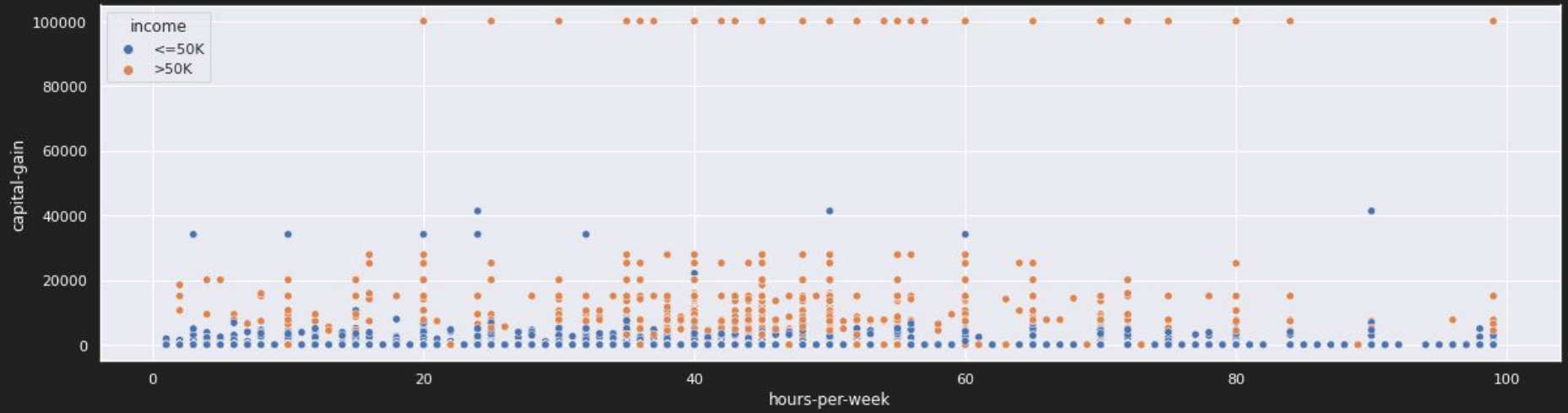
Race and gender



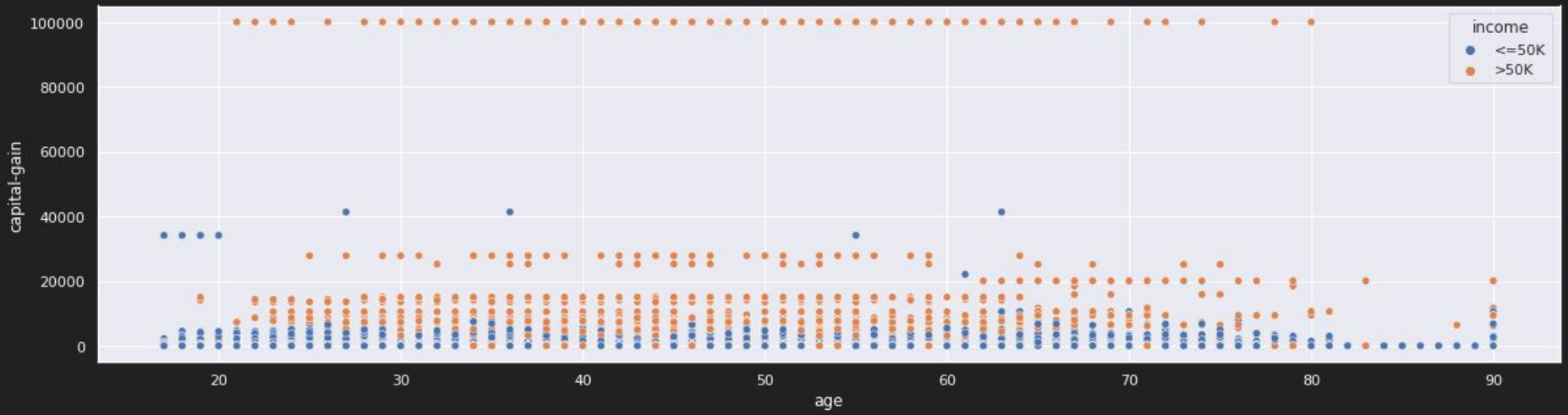
How do hours worked per week affect income?



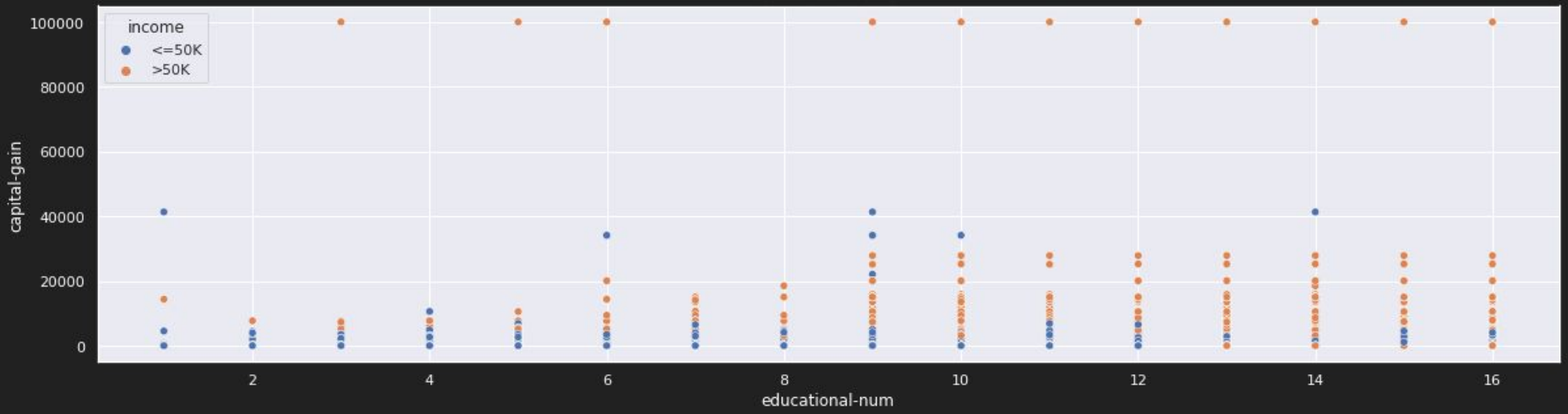
Separability of Income Groups: Hours Per Week and Capital Gains



Separability of Income Groups: Age and Capital Gains



Separability of Income Groups: Education and Capital Gains



What appear to be strong predictors of incomes greater than $50K?

- Age
- Work Class
- Education
- Marital Status
- Relationship
- Occupation
- Race and Gender
- Hours Worked Per Week
- Capital Gains and Losses
- Native Country

Afterword

- Specialized work often necessitates higher compensation. As a result, features that predict higher specialization like education or working class are useful in predicting higher income.
- In many industries, compensation is a function of work experience. Older employees that have invested more time in their craft are often rewarded for their contributions.
- America has a longwithstanding history of racial and gender discrimination in the workplace, which may lend predictive power to demographic information.

Afterword

- Lastly, after performing EDA, it’s imperative to conduct experiments to quantify the power and significance of the patterns encountered in visualization.