Does demographics have an impact on income?

Sai Ajay Vutukuri, Himani Anil Deshpande, Murali Krishna Parvataneni, Tushar Kant Samantaray

November 30, 2021

Introduction

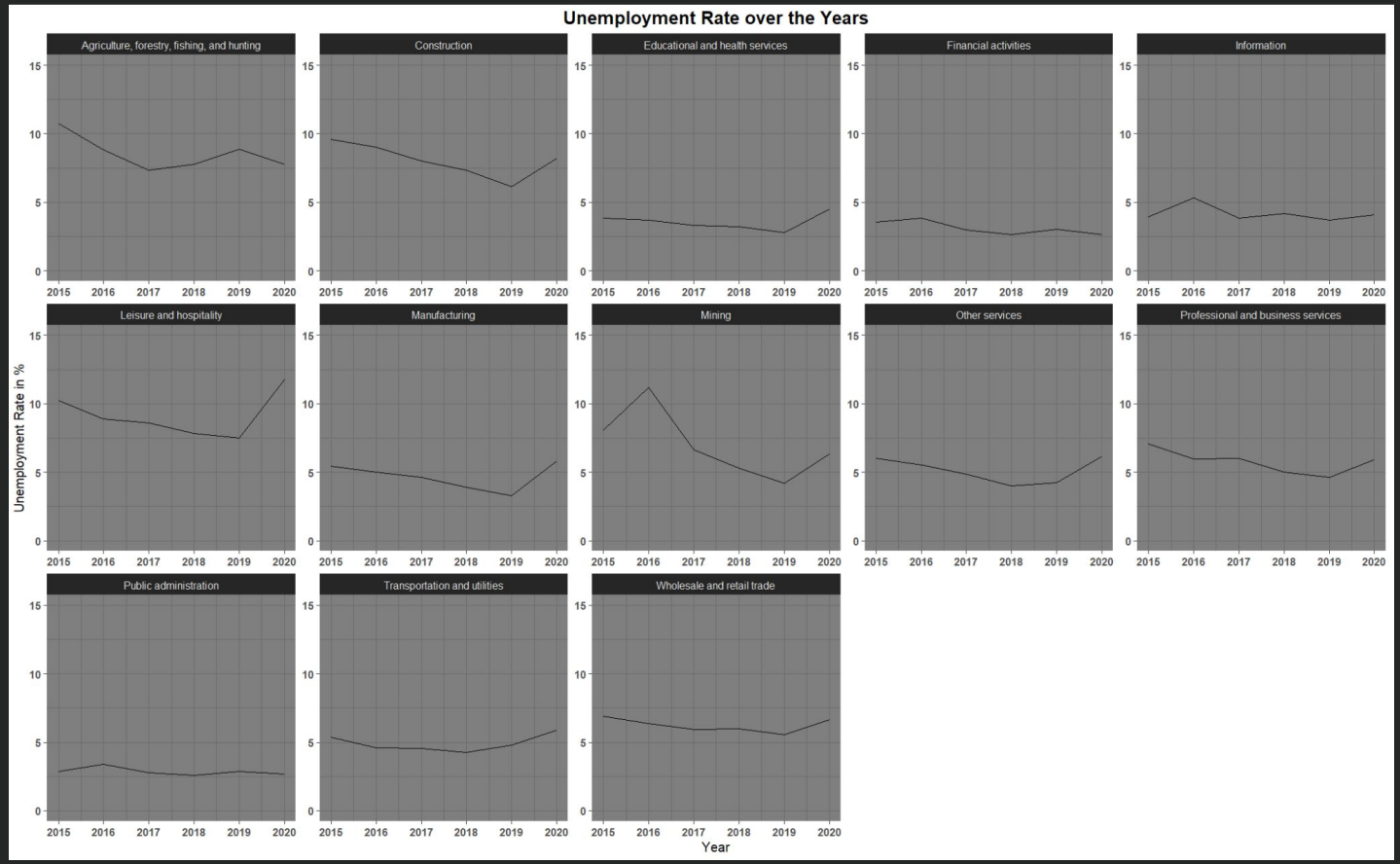
- The goal of the research is to learn about the numerous factors that influence median income in the United States from 2015 to 2020.
- The data used comes from the United States Bureau of Labor Statistics, which keeps track of employment and pay.
- We want to see if demographic parameters like gender, age, race, the industry of employment, and educational qualification have an impact on an employee's pay.
- Fit a linear model to accurately estimate an employee's income based on their demographic data, and determine whether the demographic details have interaction.

Data and Pre Processing

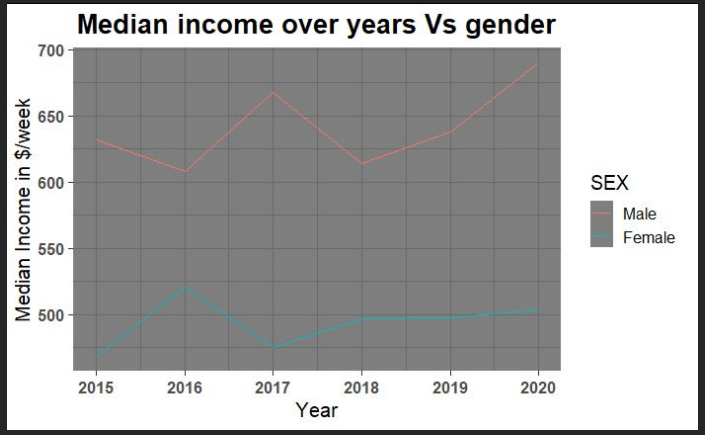
- We used IPUMS data from the US Census Bureau and the Bureau of Labor Statistics, which was gathered and made available by CPS
- The Current Population Survey is a valuable source of information for studying social and economic developments in the United States over the last 50 years
- We look at data from 2015 to 2020 in terms of employment
- For our analysis, we divided the following demographic variables into groups:
  - Create distinct groups depending on 'Age,' which ranges from 0 to 60
  - White, Black, Asian, and Others are four separate 'races'
  - There are fourteen major industries
  - There are several educational groups

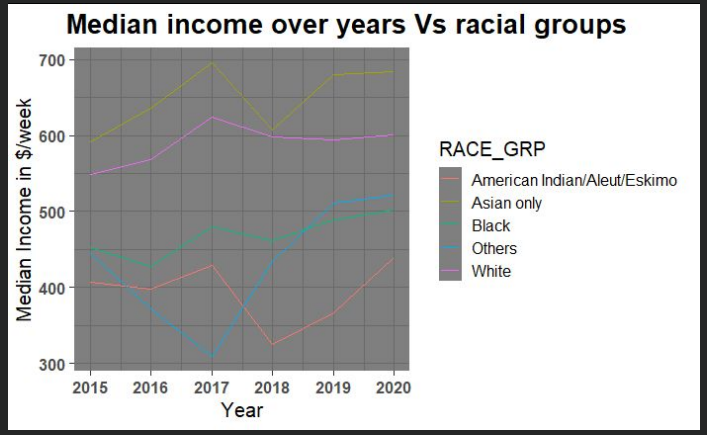
Source: https://cps.ipums.org/cps/

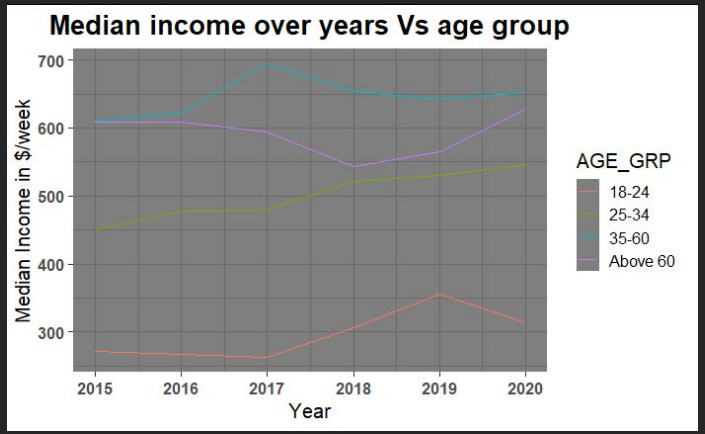
How does unemployment rate changes across different industries over the years?

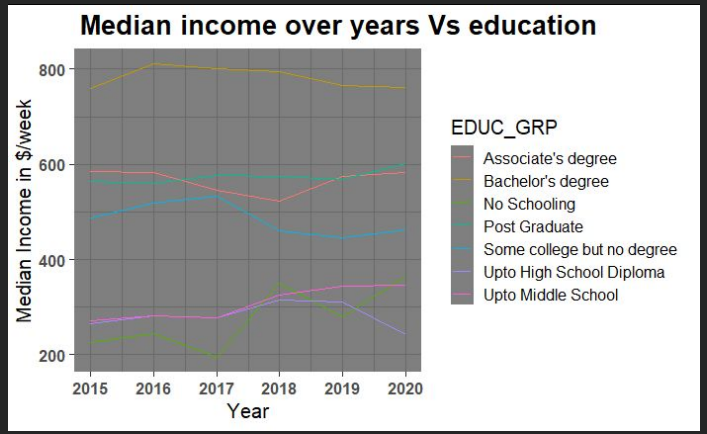


How various demographics are affecting the median income?

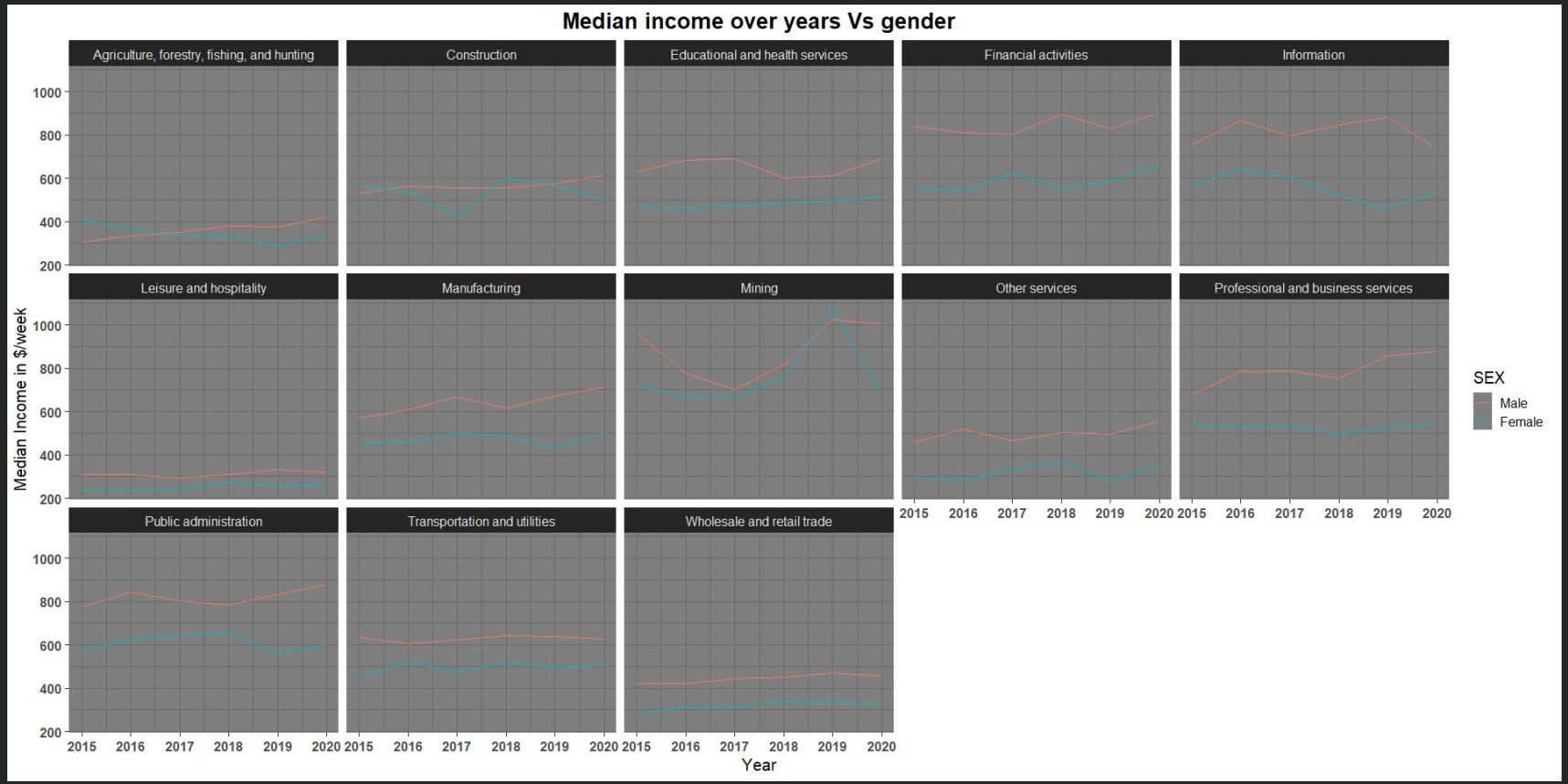




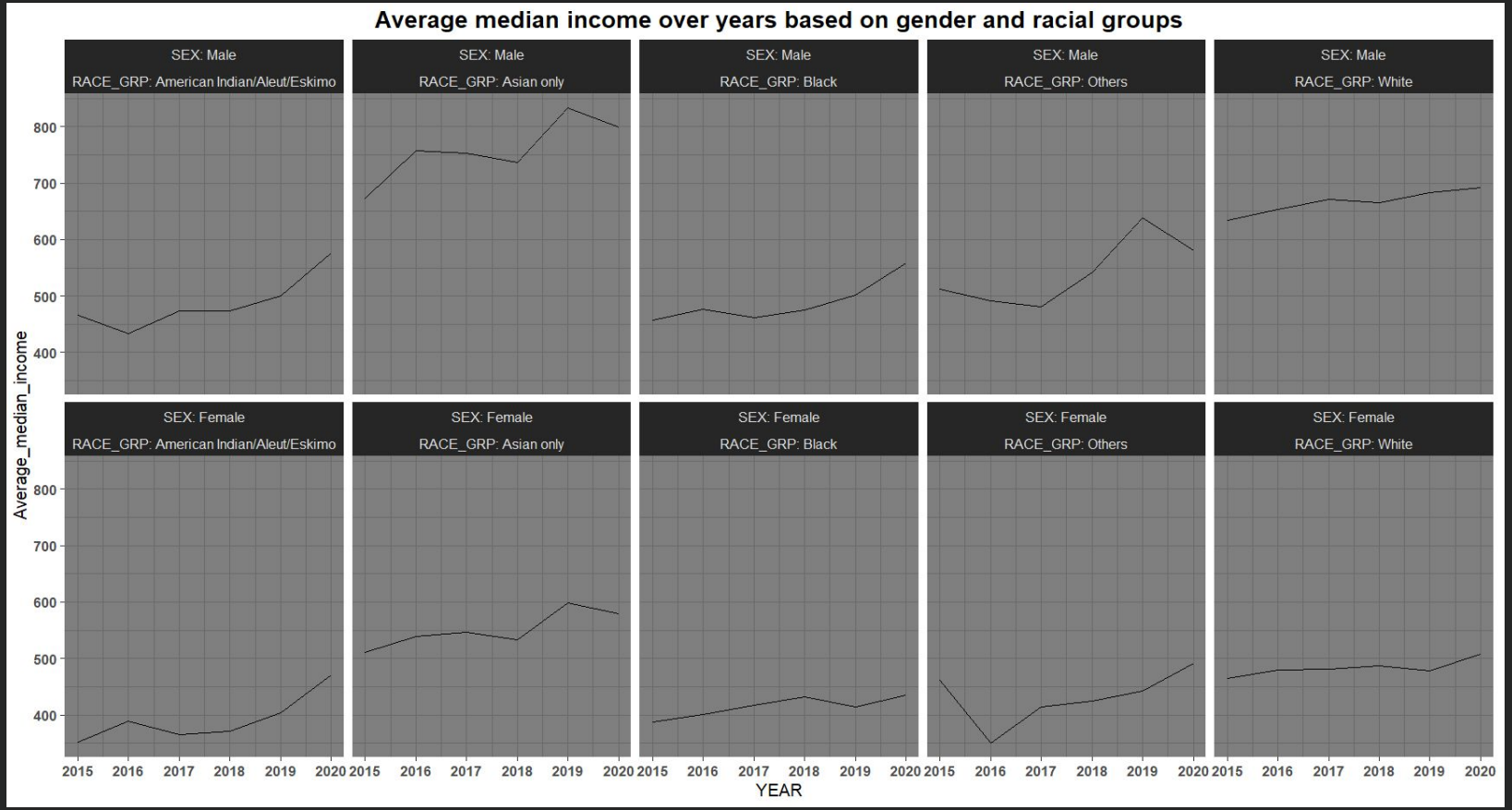




Does median income varies significantly for both genders across various industries?



Does gender and racial groups affects employees average median income over the years?



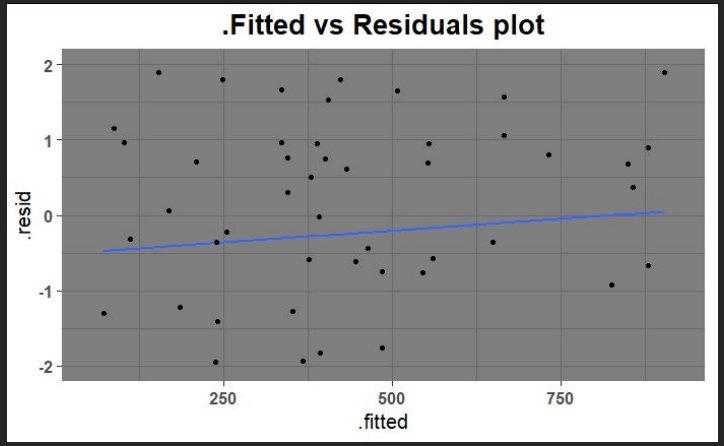
Findings so far

- Most industries saw an increase in unemployment except the Agriculture, forestry, fishing, and hunting, which instead saw an decrease in unemployment rate
- Racial group, gender, education and age effects the median income
- We notice significant wage gap between male and females, and it varies across various industries
- The trend remains same when we consider median income based on both gender and racial group

Creating a model to predict median income based on demographic variables as predictors

- Model Equation:

  lm(median_income ~ SEX + EDUC_GRP + RACE_GRP + YEAR, data = filtered_employee_data)
- AIC - 1016245



Further work

- Explore the interactions between the demographic variables
- Try to fit a model with all meaningful interactions and see if the model is a better fit

Questions/Suggestions ?