

Homework 3 (Quiz)

● Graded

Student

Anthony Wen

Total Points

10 / 10 pts

Question 1

Question 1

6 / 6 pts

✓ - 0 pts Correct

- 0.5 pts incorrect or missing relationship between rental status and time
- 0.5 pts incorrect or missing relationship between rental status and age
- 0.5 pts incorrect or missing cause of the relationship between rental status and time
- 0.5 pts incorrect or missing cause of the relationship between rental status and age
- 1 pt b) incorrect of missing answer to the first question
- 1 pt b) incorrect of missing answer to the second question
- 1 pt c) incorrect of missing answer to the first question
- 1 pt c) incorrect of missing answer to the second question

Question 2

Question 2

4 / 4 pts

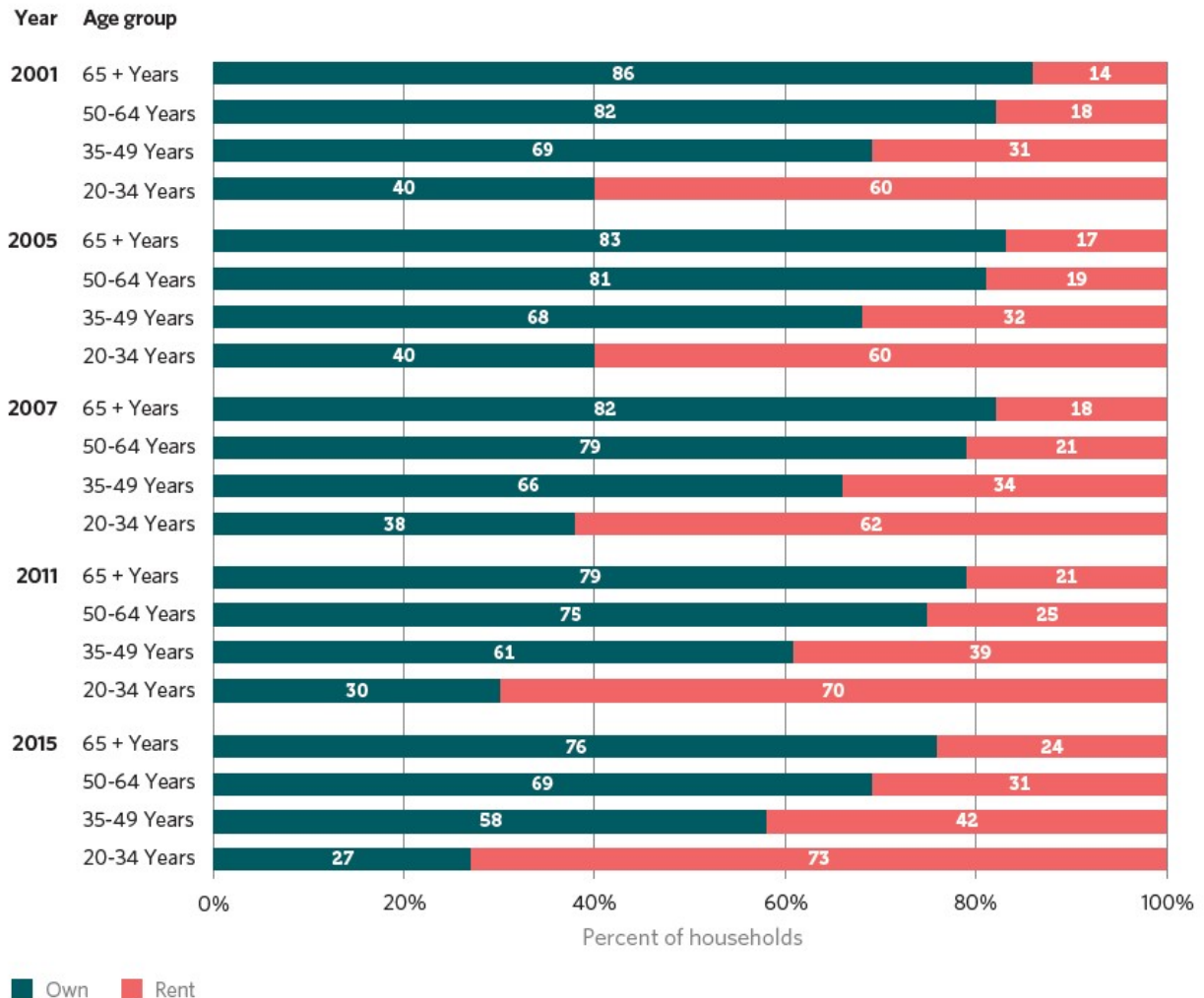
✓ - 0 pts Correct

- 2 pts (a): missing answer
- 2 pts (b): missing answer
- 1 pt (a): the answer is not relevant to the question
- 1 pt (b): the answer is not relevant to the question
- 1 pt (b): did not answer to "How does the system perpetuate or prevent bias in decision making?"

Q1 Question 1

6 Points

Percentage of households that own or rent by age, selected years, 2001-15



Note: Bars may not total 100 percent because of rounding.

Source: Pew analysis of Panel Study of Income Dynamics data

Consider the image above that looks at the relationship between home ownership vs renting of American households at different ages over various years. Address each of the following questions.

a) Describe the relationship between rental status, time, and age. Try and explain the causes of these trends.

b) How would this graphical display change (IE: how would it look), if there were no relationship (IE: independence) between time and rental status? What trends

would still remain?

c) How would this graphical display change (IE: how would it look), if there were no relationship (IE: independence) between age and rental status? What trends would still remain?

a) I noticed that no matter which year the data was taken in, there seems to be a positive correlation between the age and the percentage of owned households. This means that, as the age group gets higher, the percentage that does own a house also increases. In contrast there, is a negative correlation between the age and percentage that rent households. The reason that this might be the case is that in most cases, the older someone gets, the better their financial ability, thus explaining their ability to buy/own a house or rent/own a house.

b) The graph would look pretty similar to a graph that was made if we picked/looked at 1 individual year. This is because of the obviously nearly consistent rates across all years between age and rental/ownership of houses. At 65+ it's always around 80% that owns a house, at 50-64, it's always around 75% that owns a house, and so on. The thing that would change it probably the shape of the graph. The trend between age and ownership would still stay the same though, the older the more owned houses.

c) There would probably be a continuous distribution of people who own houses against ones who rent houses across all age groups. We would basically only think of outside factors other than age like the housing market to check to see how many people buy and own a house in the years 2001 to 2015. The graph would likely be like a bar graph or scatter plot but all kinds of visuals work. The thing that would stay the same is the trend between is the change in home ownership and house rentage rates but just influenced by factors other than age.

Q2 Question 2

4 Points

A bank manager working for a large American bank is interested in reducing their workload, and turn to an AI consulting group to help them set up an automated system for managing loan applications. The consultants ask for a dataset containing a large selection of approved and disapproved loan applications, for which the bank manager provides them all the historical loan data from the bank, going back almost 100 years. The data includes information about the loan applicants, the type of loan, and whether or not the loan was approved or denied.

The consulting group trains a complicated machine learning algorithm to predict which loans will be approved and which will be denied. Upon receiving the algorithm, the bank manager decides that all loan applications will be determined using the new system.

An employee at the bank is concerned that the new system is approving and denying loans in a biased way. The bank manager insists that the system is objective and data driven, and therefor free of any bias.

Address each of the following:

a) At a very high level, how does the machine learning algorithm use the data to make a loan determination?

b) Who do you believe is right in this situation, the employee or the manager? How does the system perpetuate or prevent bias in decision making?

a) The machine would probably analyze patterns and relationships between cases within the historical data that is provided. For example, it would probably compare case characteristics like the applicant's income, credit score, and other information and classify them compared to other applications in the past datasets. When the new data is passed to the machine, they would then classify the new case by comparing it to the old case's characteristics. They would look at which one is the closest match to the new data and then give a prediction based on that. The machine would learn to capture complex trends in the data and use the probability function to see the probability of the new data matching the old data.

b) The employee: this is because if the dataset that was used to train the AI was already biased then the new data would also be biased. Also, if the complex

algorithm had any mistakes like not taking into consideration a few variables like age, race, and other characteristics of the applicant, it might miss out on important variables that should be compared when deciding when an applicant is approved or not. All of this is basically algorithmic bias.