Assignment 2

Q1. What does each row of the data matrix represent?
-Each row of the data matrix represents every person in the UK who has taken a survey regarding his/her smoking habits.

Q2. How many participants were included in the survey?
-1691 participants were included in the survey.

Q3. Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.
-

Gender	→	Categorical and not ordinal
Age	→	Numerical and continuous
MartialStatus	→	Categorical and not ordinal
HighestQualification	→	Categorical and not ordinal
Nationality	→	Categorical and not ordinal
Ethnicity	→	Categorical and not ordinal
GrossIncome	→	Categorical and ordinal
Region	→	Categorical and not ordinal
Smoke	→	Categorical and not ordinal
AmtWeekends	→	Numerical and discrete
AmtWeekdats	→	Numerical and discrete
Type	→	Categorical and not ordinal

Q4. According to the CAY classification discussed in the Jacoby reading, how many modes does this data have? How many ways?, How many levels are there for each way?
-
The data is having one mode (which is answers of the survey), there are two ways as it's a two-dimensional matrix, where one way has 12 levels and the other has 1691 levels.

Q5. Consider this proposal: if the survey had asked for "years of formal education" instead of "highest qualification," then we could interpret the responses as values of a ratio level variable. Take a position on whether that interpretation would be reasonable and useful. Explain and justify your argument in no more than two paragraphs.
-
A ratio variable has a clear definition of 0.0. As we are changing the variable from "highest qualification" to "years of formal education", we are changing the data from categorical to continuous numerical variable. A ratio level variable is also an interval variable. Which means when the variable equals 0.0, there is none of that variable. In our case, if a person has taken 0 years of formal education, it means that the person has no education.
For example, if we have to compare two variables to calculate the ratio of the two variables, suppose if person A has 4 years of education, and the person B has 2 years of education then the person A has twice the the years of education as person B, because number of years is a ratio variable.

Q6. Propose two different interpretations of the gross income variable, following the four way Coombs classification discussed in the Jacoby reading. In which two categories would observations on income be classified. Explain your reasoning.
-

Assignment 2

Gross income variable can be classified into two types: Stimulus comparison and Proximity Similarities. For example; if person A has more gross annual income than person B, then we can say that person A has a dominance relation over person B. And since both the observations are drawn from a single data set, the observations can be classified as stimulus comparison data.
Second condition is; if person A has similar gross income as person B, then the observations are set to be in proximity similarity.