Dataset: contribution.csv

An NGO operates an online campaign to motivate contribution for child development. They design a new message and want to test its effectiveness.

The NGO sent out fund-raising messages to 297 donors who all contributed various amount of contribution. Each donor received either the old message or the new one, labeled in the dataset as treatment=1. In addition to each individual’s contribution (in k$) and which message this person received (treatment=1 if new message, otherwise 0), the NGO also manages to collect data concerning the donor’s individual characteristics: gender (male=1 if the person is male and 0 if the person is female), income (in k$), and age (in year).

Please write a report answering the following questions.

1. Make a frequency table of the donor’s **income**. Motivate how many categories you want to divide them into and the range of each category.
2. Summarize the distribution of **contribution**, **age**, and **gender** of all the donors in this sample with numerical measures you see proper.
3. Summarize the contributions in each **treatment** using the same set of measures as in question 2.
4. Is the treatment (new message) effective in increasing people’s **contribution**? Is this relationship causal or merely correlational?
5. Among the subgroup of donors with **income** more than 15k, is the **treatment** effective in increasing their **contribution**?
6. Pick one of the two multiple linear regressions to explain contribution. Which one do you think can explain the variance of **contribution** better? Why?
   1. contribution~age+male+treatment
   2. contribution~income+treatment
7. Give your interpretation of the coefficients of **treatment** and **income** in the following regression: contribution~treatment+income+age+male
8. Use the model in question 6 to predict the contribution of
   1. **Mr.** Benz, **age** 62, **income**=45k, received the **old** message
   2. **Miss.** Valaria, **age** 27, **income**=19k, received the **new** message
9. The NGO is planning to send out invitation for contribution to 1000 more people. How much more contribution could be raised if they use the **new message** instead of the old one?