1. Project format: this data analysis project takes the form of **individual coding and group writing**. At the end of the semester, each student submits an R code file and each group submits an analysis report. The grouping follows the random principle. The size of each group is as equal as possible, with three to four people (the number of people in the course cannot be divided completely equally). The list has been uploaded to the eLearning.

2. Programming part: you are supposed to perform regression analysis on the Boston Housing data set, and implement the learned content using the R language after each chapter. For example, after studying Chapter 5, you may use OLSE to fit a linear model to the data, make point and interval estimates of the regression parameters, and then test the significance of the overall model and each covariate. After that, you may remove the insignificant covariates and fit a reduced model. Dividing the data into a training set and a test set, you can compare the full and reduced models according to their prediction errors. After studying the subsequent chapters, you can test whether the data satisfy the fundamental model assumptions and make corresponding improvements to the basic model. Code to load the data and to present their basic information can be found in the file "Real_Data_Analysis.R". Please add sufficient **comments** to your code to explain the purpose and function of each block or line (readability of comments is one of the criteria for the code scoring). For specific programming methods and comment formats, please refer to the example code flies on eLearning.

3. Report part: the analysis report should mainly focuse on interpreting the programming results, such as the practical implications of the regression parameter estimates, whether the positive or negative signs are consistent with the realistic background, how to intuitively explain the insignificance of some covariates in the test, and how the fitted model reveals the effects of the covariates on the response. Based on these results, you are supposed to provide guidance for all relevant parties in the real estate industry. When writing the report, it is recommended that you assume a relevant identity, such as a government decision-maker or an industry practitioner, so as to make practical interpretations and suggestions. All interpretations must be based on the data analysis results rather than being discussed in general terms. Please keep the text of the report within **ten pages**. The remaining content can be attached as supplementary materials after the text.

4. Analysis content: the main goal is to practice the course content. First, describe the data, and then use the methods learned in the course to model and analyze. On this basis, other modeling methods (such as more complicated regression or machine learning algorithms) can also be used to analyze the data. Results from different methods should be compared while their similarities and differences should be well explained.

5. Scoring criteria: it consists of two parts: individual code and group report, each accounting for 10% of the total course score. Code scores are different for each person while report scores are the same within each group. This project accounts for 20% of the total course score, as stated in the syllabus.

6. Report presentation: in the last two class meetings of the semester, each group will have **at most ten minutes** to present their analysis reports on stage.

7. Deadline: submit the final personal code and group report before **5pm on Monday, December 23th**. The score for this project is based on the final individual code and group report submitted on December 23th.