Note: Please email your HW1 report (with name and ID, in pdf format) to the TA (109354004@nccu.edu.tw) before the deadline (13:00, 3/24/2022).

Download the dataset "2020-QS-World-University-Rankings-100.csv" from the WM5 website. The data contains variables based on which the 2020 world college ranking was given by the QS. The variables are:

Rank: college ranking based on the "Overall_Score"
College Name: name of colleges from all places
Academic_Reputation: score given for the college's overall academic reputation
Employer_Reputation: score given by employers based on the quality of graduates
Faculty/Student: score based on the college's faculty/student ratio
Faculty_Citation: score based on citations per faculty
International_Faculty: score based on the college's international faculty ratio
International_Students: score based on the college's international student ratio
Overall_Score: score calculated based on the above category scores with weights

Let the response variable be y = "Academic_Reputation" and consider 5 independent variables x = "Employer_Reputation", "Faculty_Student", "Faculty_Citation", "International_Faculty", "International_Students".

Fit the following 3 regression models by using R and evaluate their performance in terms of the **accuracy of predicting the response y** = "Academic_Reputation" (using a **10-fold Cross Validation**):

Model 1: The Least Squares (LS) regression model without the intercept term.
Model 2: The Principal Component Regression (PCR) without the intercept term. For this method, please choose the best number of components based on the model predictability.
Model 3: The Partial Least Squares (PLS) regression without the intercept term. Analogously, please choose the best number of components based on the model predictability.

Questions: (1) Are the above 3 prediction models similar, or different? (2) Which model is best for predicting the college's "Academic Reputation"? Explain why.

[Hint]: For Q1, running the regression model by using function lm(); then the CV error (prediction error) can be computed by using function cv.lm(), which requires installation of package "lmvar". For Q2, you need to install package "pls". Please refer to the R codes on page 256-258 of the book I gave you: "An Introduction to Statistical Learning: With Applications in R" by James et al. (2017).