For the Wage dataset, what years are recorded here? (b) How many rows of data are recorded for each year on record? (c) Find the mean, min and max wage for each of these years on record. (d) Does the result make sense?

Simple Linear Regression on ISLR2 (Hitters): (a) Build a linear regression model for each of the following variables to predict salary: AtBat, Hits, HmRun, Runs, RBI, Walks, Years (b) Which of these performs the best in terms of evidence of being linearly associated with salary?

Multiple Linear Regression ISLR2 (Credit): (a) Build a multiple linear regression model to predict the Rating. (b) Next build the single variable models and compare the difference. (c) How much better or worse in the multiple linear regression model at predicting the Rating versus the best single variable model?

Multiple Linear Regression ISLR2 (College): (a) Build a model that can predict the sum of the Outstate, Room.Board ,Books columns (excluding the Expense variable). Here we are aggregating the separate costs for each school into one value and they trying to predict it with the other variables. (b) check the variables to see if there are any potential interactions and then revise your model to have the interaction terms. (c) keep the best model of your results and explain if it is a good performing model or not.

Select he following column vectors of interest: Age, Race1, Education, Poverty, BMI, Pulse, BPSysAve, PhysActive, Diabetes. (b) Clear out any NA values using na.omit and detail the number of valid rows of data for each gender ‘male’ and ‘female’ by plotting the histogram of the gender column vector. (c) Is this distribution even and if not, why do you think this is the case?

Build a multiple-logistic regression classifier (using glm function) on the Diabetes vector with a 10% test split (b) Build a random forest classifier (using randomForest library) with a 10% split. Which model has better accuracy?

Now repeat part [2] but do it for each gender separately by partitioning on the gender column vector. (b) Which is the better classifier now and are the separate gender accuracy measures the same or is one better (make sure to use the baseline distribution of each gender to support your finding)? (c) Why do you think that is the case?

now repeat part [3] but do it using the SVM methods (with the e1071 library). (b) Were you able to improve upon any of the gender-specific models with SVM? (c) Why do you think that is the case?