

## Missing Data

In this assignment you will complete a variety of tasks related to working with missing data.

**Libraries:** For this assignment you will need the following libraries: tidyverse, VIM, and mice.

Before beginning the assignment tasks, you should read-in the data for the assignment into a data frame called grades. This data contains grade information from an engineering course. The dataset was originally used to investigate how student performance in the course would be predictive of student grades on the “Final” exam. The “Prefix” column is a surrogate for enrollment year in the engineering program. Smaller values imply older (more mature?) students.

**Task 1:** For which variables is there missing data?

**Task 2:** Use the VIM package to visualize missingness. Are there any students that are missing multiple pieces of data?

**Task 3:** Use row-wise deletion of missing values to create a **new data frame**. How many rows remain in this data frame?

**Task 4:** Use column-wise deletion of missing values to create a **new data frame** (from the original data frame not from the data frame created in Task 3). How many columns remain in this data frame?

**Task 5:** Which approach (Task 3 or Task 4) seems preferable for this dataset? Briefly discuss your answer.

**Task 6** Use the “mice” package to impute the missing data in this dataset. Set “m = 5” and “method = “pmm”” in the imputation. You MUST use a random number seed of 123 for this imputation. What is the mean value for the “Final” variable after imputation?