Question 1

1. What is correlation between variables or features of a dataset? Why we need to find correlation to extract the best predictor?

2. In the following table each row represents one observation, or the data about one employee (either Ann, Rob, Tom, or Ivy). Each column shows one property or feature (name, experience, or salary) for all the employees.

Name	Years of Experience	Annual Salary
Ann	30	120,000
Rob	21	105,000
Tom	19	90,000
Ivy	10	82,000

If you analyze any two features of a dataset, then you'll find some type of correlation between those two features. Explain what type of correlation is expressed in the following three graphs?

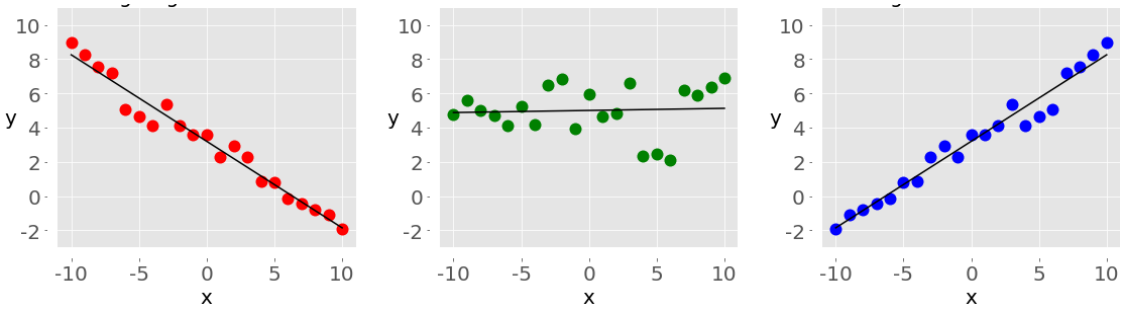

Figure 1: Different forms of Correlations

3. What is the difference between correlation and linear regression?

4. Which one will be more suitable, correlation or regression, for each of the following scenarios:

   (a) You want to predict blood pressure for different doses of a drug.

   (b) A clinical trial has multiple endpoints and you want to know which pair of endpoints has the strongest linear relationship.

   (c) You want to know how much the response (Y) changes for every one-unit increase in (X).