To be submitted (Friday, March 18)

1. Names of team members.
2. Brief summary of the data: Clearly describe the response variable, feature variables (predictors), and data source.
3. Research objective: What is the goal of the analysis? Is this a regression or classification problem?

1. Jeffrey Kerley and Drew Dahlquist.
2. Link to dataset: https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset.

Abstract: Images of 13,611 grains of 7 different registered dry beans were taken with a high-resolution camera. A total of 1

Data Set Characteristics:	Multivariate	**Number of Instances:**	13611	**Area:**	Computer
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	17	**Date Donated**	2020-09-14
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	1965816

The data source is Selcuk University, Turkey. Each bean image sample went through segmentation and feature extraction. 16 predictors, 13,000+ uniform samples consisting of the 7 different types of beans are available.

List of the feature variables are as follows:

Attribute Information:

1.) Area (A): The area of a bean zone and the number of pixels within its boundaries.
2.) Perimeter (P): Bean circumference is defined as the length of its border.
3.) Major axis length (L): The distance between the ends of the longest line that can be drawn from a bean.
4.) Minor axis length (l): The longest line that can be drawn from the bean while standing perpendicular to the main axis.
5.) Aspect ratio (K): Defines the relationship between L and l.
6.) Eccentricity (Ec): Eccentricity of the ellipse having the same moments as the region.
7.) Convex area (C): Number of pixels in the smallest convex polygon that can contain the area of a bean seed.
8.) Equivalent diameter (Ed): The diameter of a circle having the same area as a bean seed area.
9.) Extent (Ex): The ratio of the pixels in the bounding box to the bean area.
10.)Solidity (S): Also known as convexity. The ratio of the pixels in the convex shell to those found in beans.
11.)Roundness (R): Calculated with the following formula: (4piA)/(P^2)
12.)Compactness (CO): Measures the roundness of an object: Ed/L
13.)ShapeFactor1 (SF1)
14.)ShapeFactor2 (SF2)
15.)ShapeFactor3 (SF3)
16.)ShapeFactor4 (SF4)
17.)Class (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira)

3. The goal of this analysis is to accurately classify at least 3 or 4 (ideally all 7) types of dry beans based on some important subset of the 16 given predictors. We plan on using methods such as LDA, QDA, Decision trees, or SVMs in addition to cross-validation for model tuning and assessment. This is a classification problem with 7 classes, 1 for each type of bean.