This case study introduces a real dataset which is about diagnosing cancer based on the FNA test. The dependent variable is binary, with outcomes Malignant and Benign. In total we have 32 variables, since the purpose of this exercise is limited to model building and validation, we are not explaining each and every variable and they have also not been labeled in the dataset, however, some(not all) variables are listed below for reference:

1) ID number
2) Diagnosis (M = malignant, B = benign)
3-32)
Ten real-valued features are computed for each cell nucleus:
a) radius (mean of distances from center to points on the perimeter)
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness (local variation in radius lengths)
f) compactness (perimeter^2 / area - 1.0)
g) concavity (severity of concave portions of the contour)
h) concave points (number of concave portions of the contour)
i) symmetry
j) fractal dimension ("coastline approximation" - 1)