3.2 Dataset

Attribute Information and Measurement

#	Attributes	Type	Domain
1	Sample code number	Nominal	ID number
2	Clump Thickness	Ordinal	1-10
3	Uniformity of Cell Size	Ordinal	1-10
4	Uniformity of Cell Shape	Ordinal	1-10
5	Marginal Adhesion	Ordinal	1-10
6	Single Epithelial Cell Size	Ordinal	1-10
7	Bare Nuclei	Ordinal	1-10
8	Bland Chromatin	Ordinal	1-10
9	Normal Nucleoli	Ordinal	1-10
10	Mitoses	Ordinal	1-10
11	Class	Nominal	2 = benign,4 = malignant

3.3 Data Analytics

Descriptive Analytics

1. Five number summary by constructing box plot
2. Check correlations by chi-square and correlation-coefficient

Predictive Analytics

1. Build predictive model with all the independent variables by conducting logistic regression
2. Tune the model to a minimal and more accurate model

Note: 16 missing values were dropped from the dataset before conducting the analytics