**Introduction/Methods:**

The objective of this project was to apply the mapreduce() frame work on a large data set and then process and classify the data.

MapReduce helps with processing and generating large data sets. The Map function processes a key pair that generates a set of key pairs. The reduce function then merges the values associated with the same key. Below is a flow diagram of how the MapReduce Function works.

File Offset

Key

Key

Key

Key

Key sorter

Input Data File

Mapper

Reducer

Output Data File

Value Line content

Value

Value

Value

The query assigned was:

1.      Extract all samples belonging to the group “Respiratory” such that  
“discharge\_disposition\_id” = 1 (“Discharged to home” type).  
2       Extract all samples belonging to the group “Respiratory” such that  
“discharge\_disposition\_id” != 1 (not “Discharged to home” type).

Using Rhadoop and the MapReduce function, the data set was extracted pertaining to the query. The next step was to choose a machine learning tool in order to produce quantifiable associations.

The machine learning tool that was chosen was the Bayes decision rule. This was done because Bayesian inference allows seeing whether there is a relationship between the feature and the class presented. And if there is then how can we use that information in the future in order to predict future needs. The flow chart below explains the process used:

Classification of Features

Naive Bayes

MapReduce

Query

RHadoop

Data

**Result:**

*Admissions:*

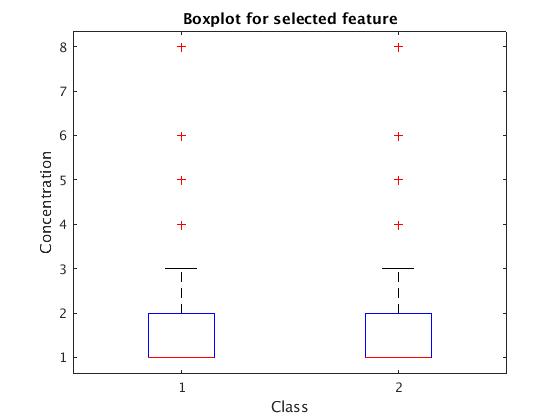
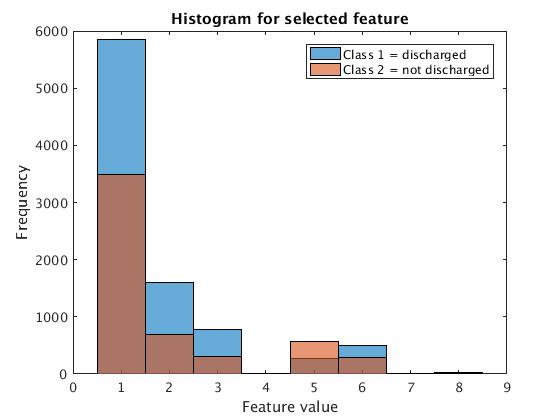


Figure 1,2: Histogram and Box Plot for Admissions Type; 1 = Emergency, 2 = Urgent, 3 = Elective, 4 = Newborn, 5 = Not Available, 6: NULL, 7 = Trauma Center, 8 = Not Mapped

*Age:*

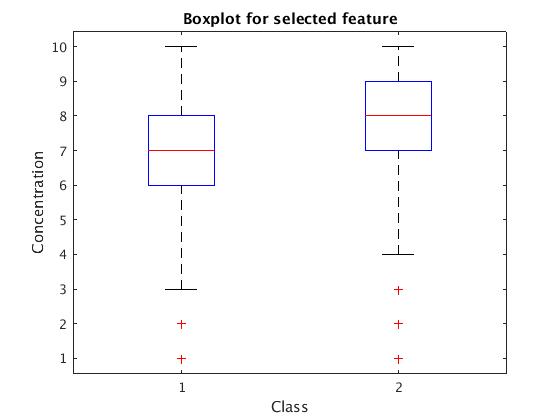
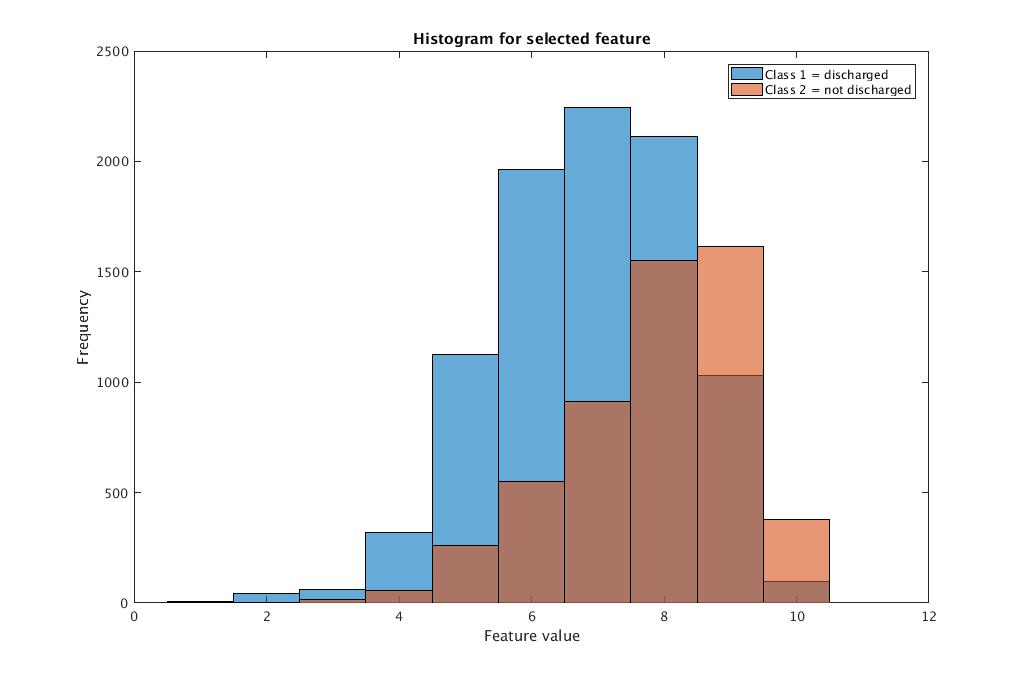


Figure 3,4: Histogram and Box Plot for Age; 1 = (0-10), 2 = (10-20), 3 = (20-30), 4 = (30-40), 5 = (40-50), 6 = (50-60), 7 = (60-70), 8 = (70-80), 9 = (80-90), 10 = (90-100)

*Gender:*

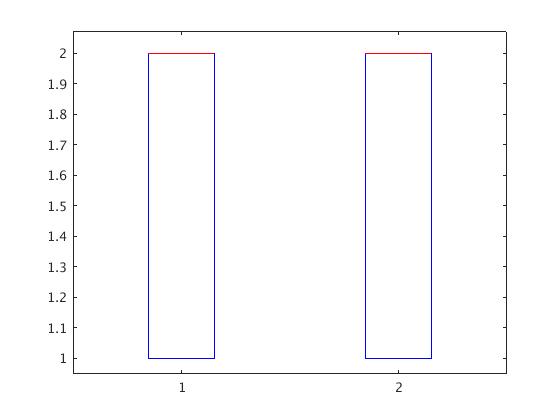
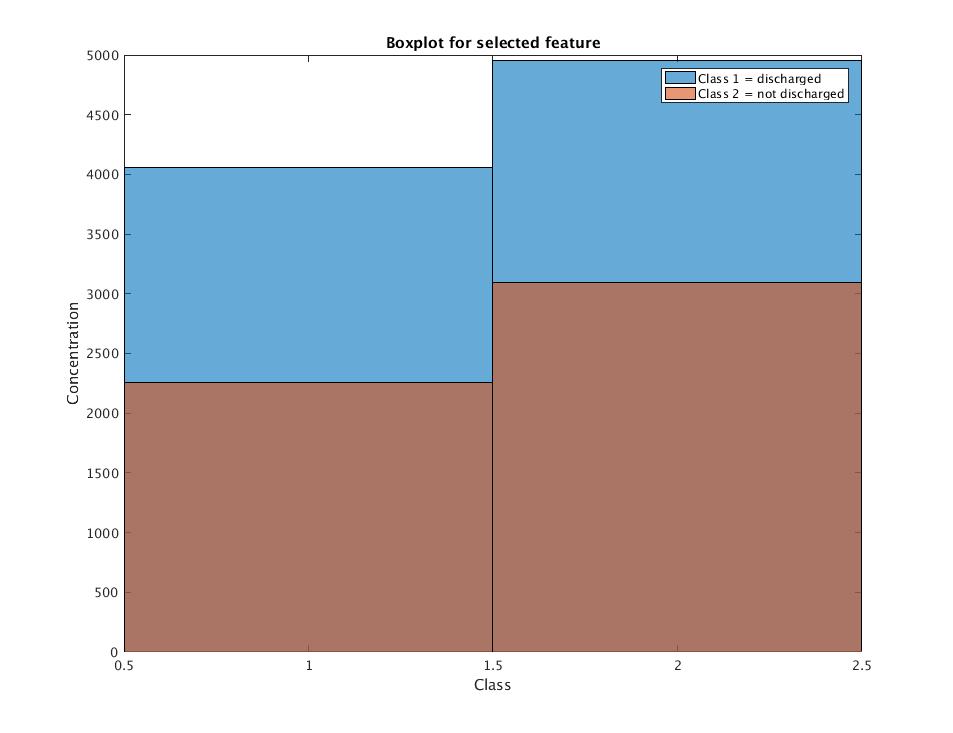


Figure 5,6: Histogram and Box Plot for Gender; Male = 1, Female = 2

*Medication:*

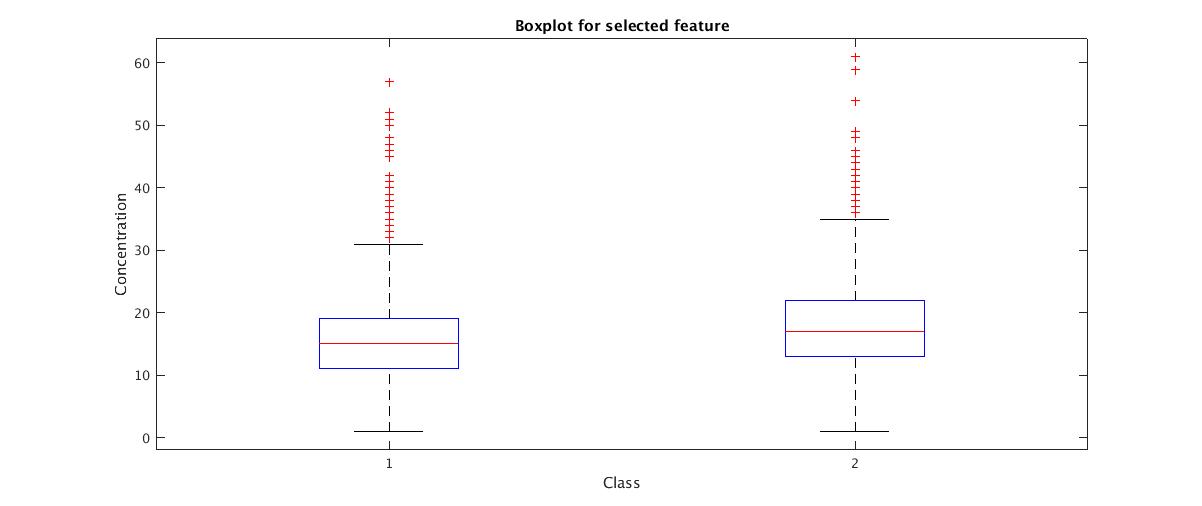
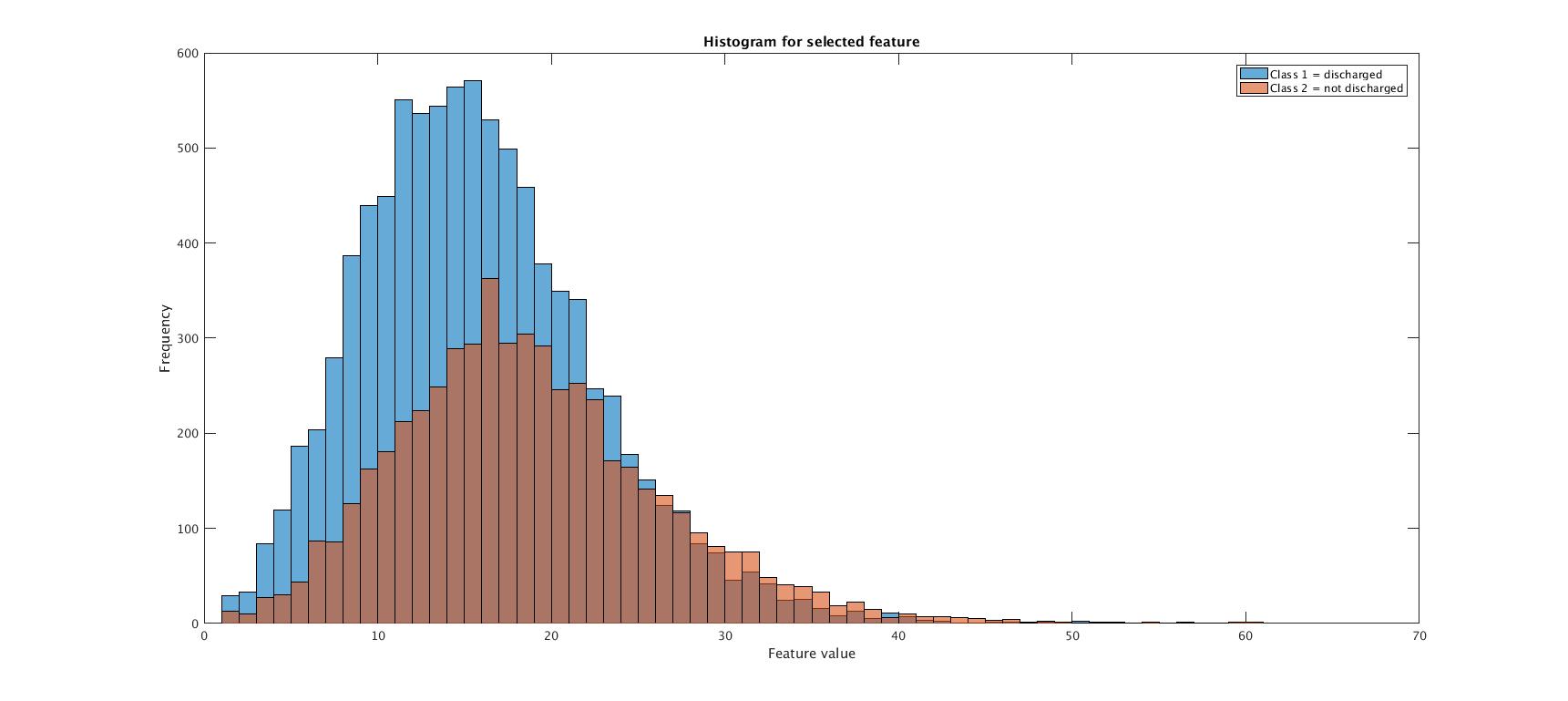


Figure 7,8: Histogram and Box Plot for Medication

*Procedure:*

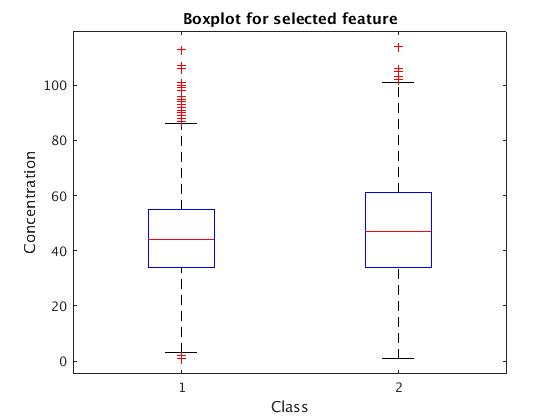
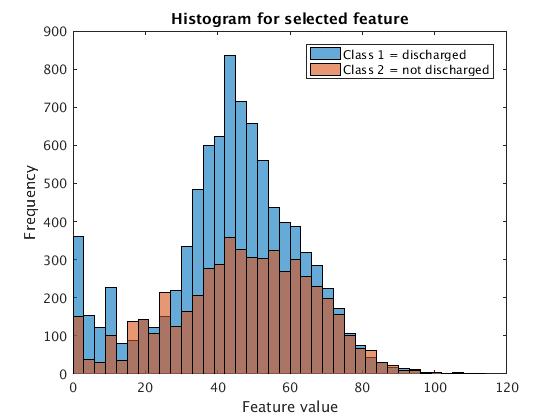


Figure 9,10: Histogram and Box Plot for Procedure

*Readmission:*

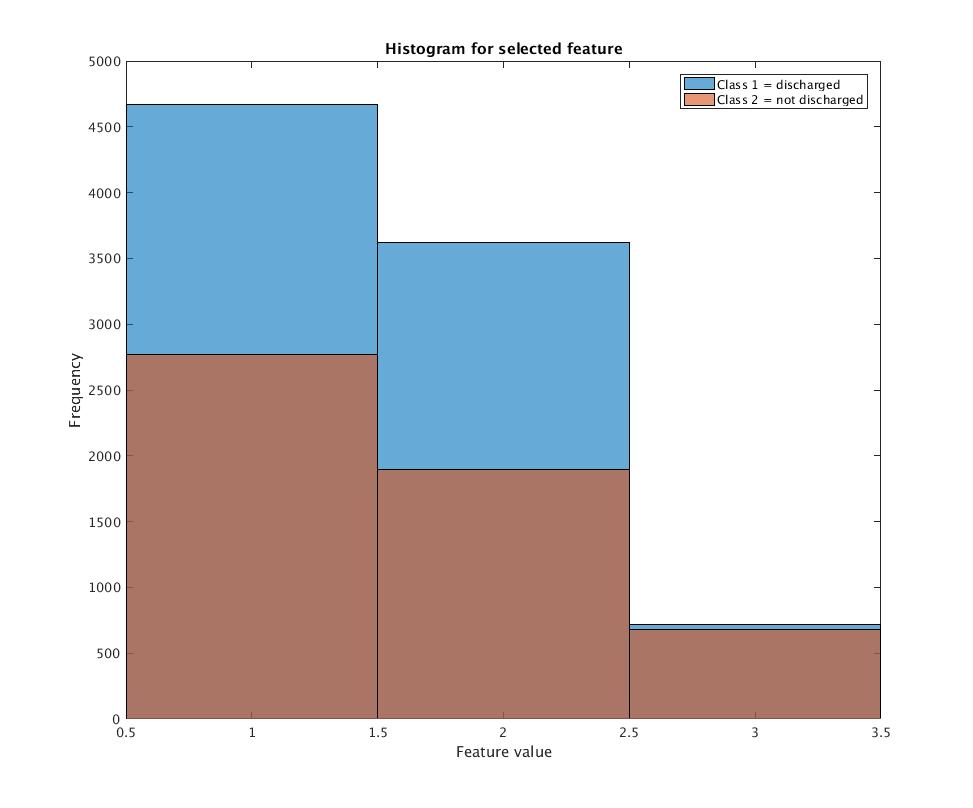


Figure 11,12: Histogram and Box Plot for Readmission; 1 = NO, 2 = >30, 3 = <30

*A1CResult:*

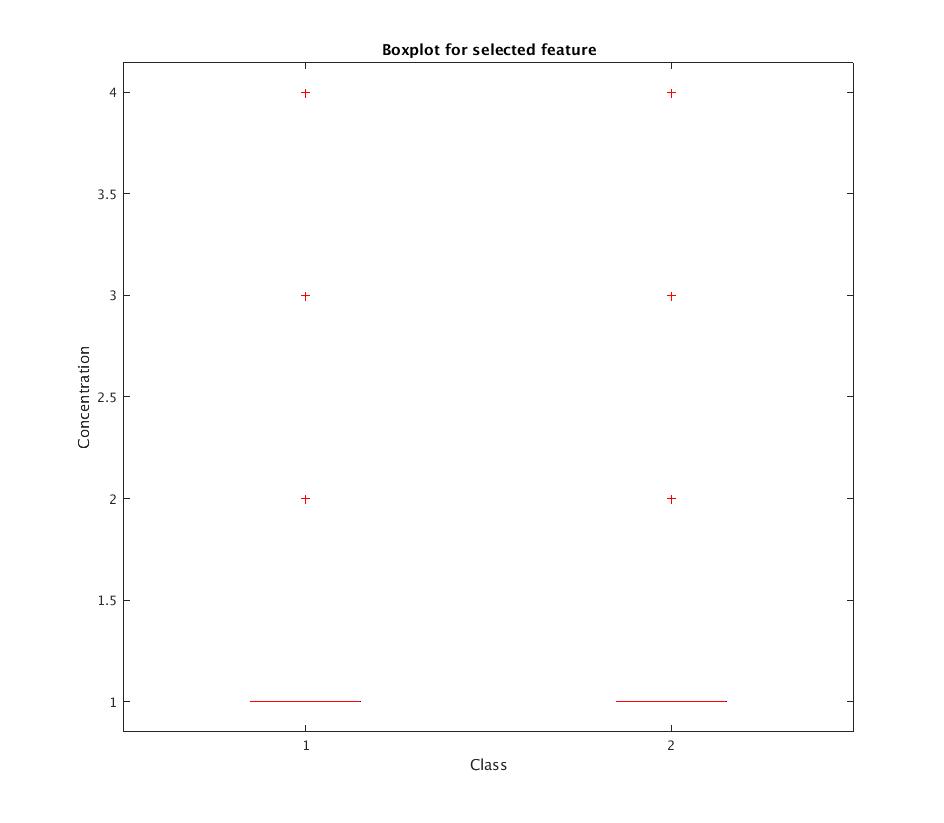
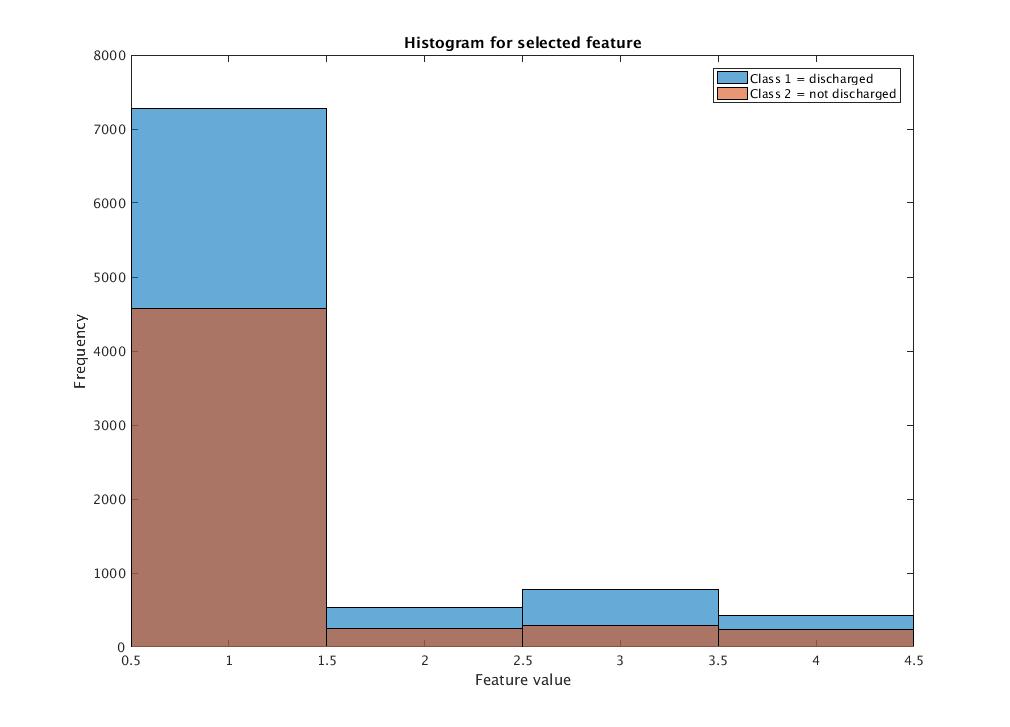


Figure 12,13: Histogram and Box Plot for Result: 1 = None, 2 = Norm, 3 = >8, 4 = >7

*Time:*

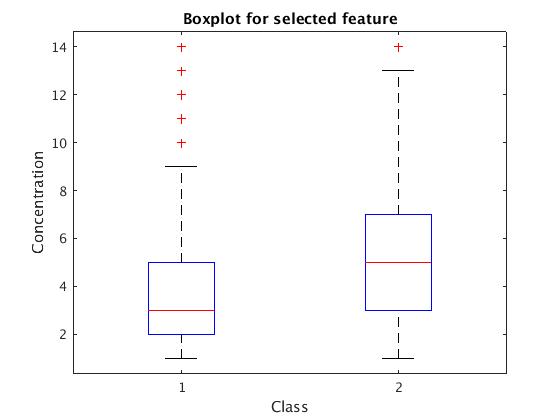
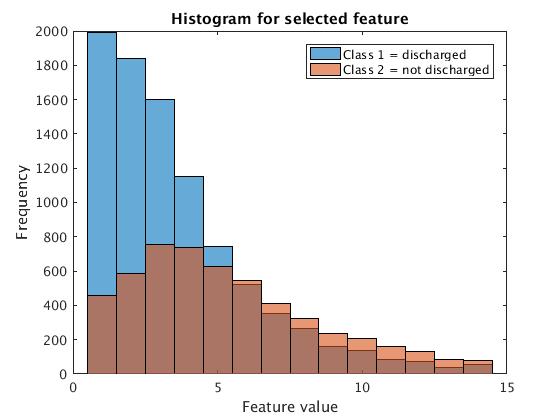


Figure 14,15: Histogram and Box Plot for Time

Discussion: