Alex Te (861215867)
CS 171 Spring 2018
Professor Papalexakis
Late Days used: 1
Total Late Days used so far: 1

Question 0: Getting real data

       I imported the data from the iris and wine dataset by creating a script in python that would read in the data, store it into an array, and parse it.

Question 1: Feature distribution

1. To calculate and plot equal width histograms, I first prompted the user for the number of desired bins that they want to see, either 5, 10, 50, or 100. Then I asked them what dataset / attribute they want. To calculate equal width bin sizes, I took the range of their chosen attribute and I divided it by the chosen bin size. I calculated the range of each bin size by taking the min of the dataset and adding (range - binsize) to get the upper bound. Continue this until you reach the true upper bound. I then stored it into a frequency array so that I can pass it into python's box function. This will yield equal width bin sizes for the histograms.

2. Plotting the boxplot was straightforward since I stored the frequency values into a frequency array. All I needed to do was to call python's boxplot function and passed the frequency array as the parameter to get the image.

3. All The figures are below, categorized by bin size and then the type of Iris / Wine. Also determined skewed/symmetric and unimodal / bimodal / multimodal in the description
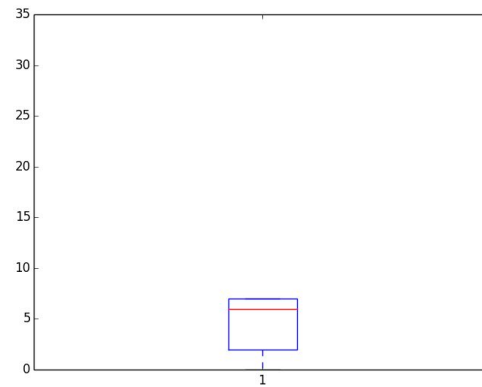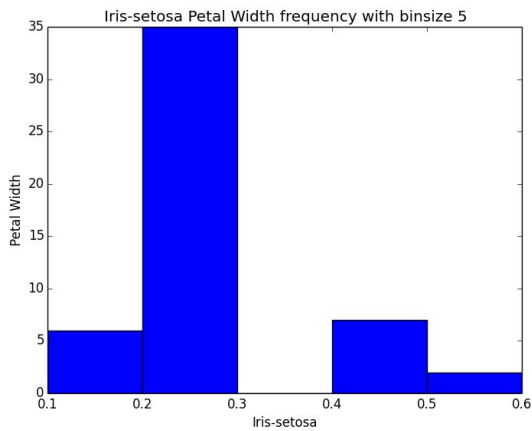
Assumption: Only doing first 3 attributes for wine dataset (Alcohol, Malic Acid, Ash)
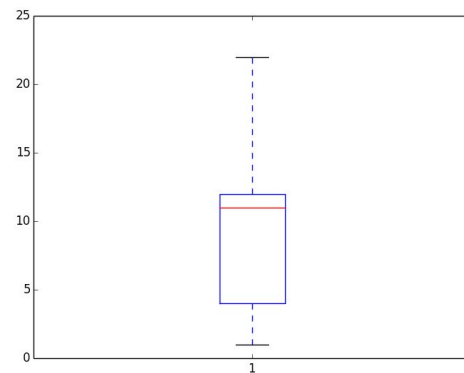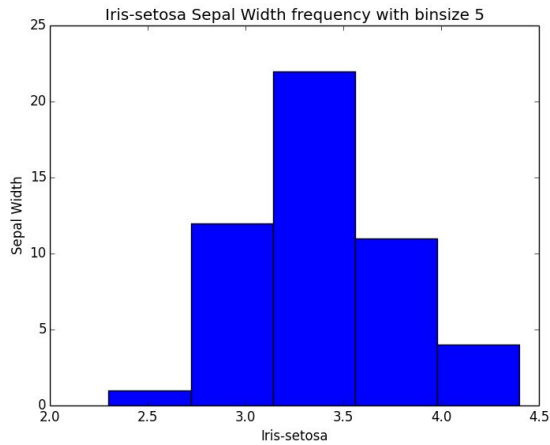
Bin Size: 5

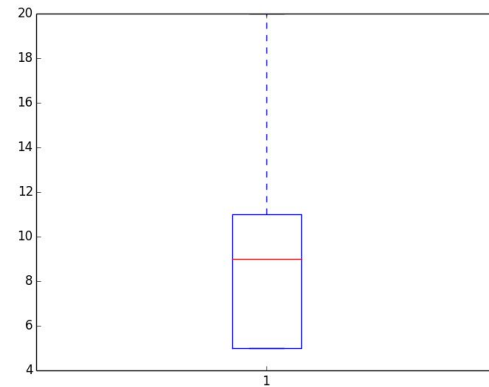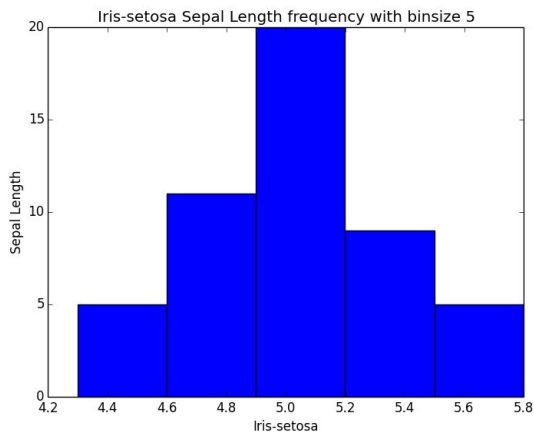Iris-Setosa Petal Length Histogram and Boxplot (symmetric, mostly unimodal):
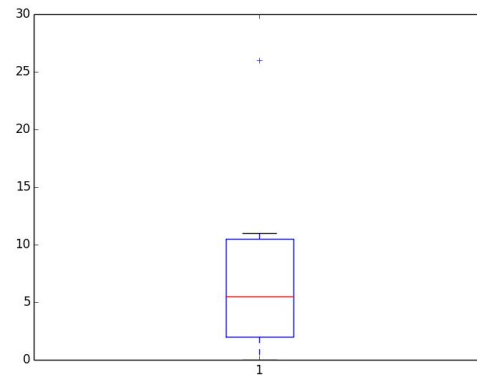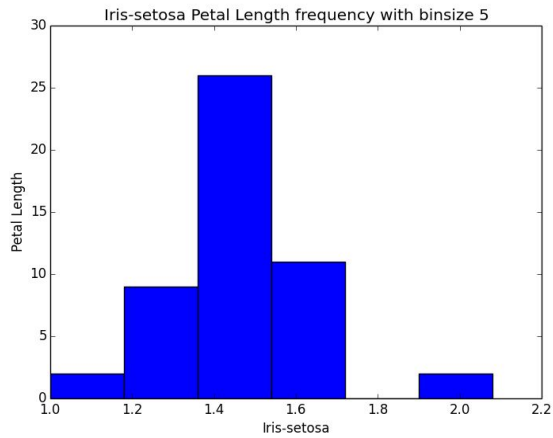
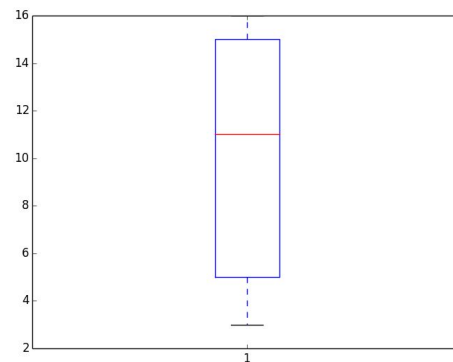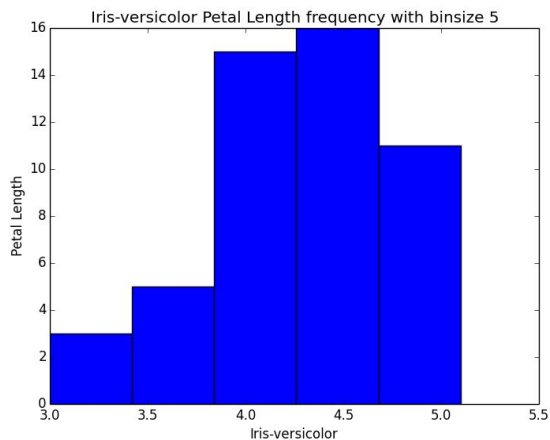Iris-Setosa Petal Width Histogram and Boxplot (Skewed, unimodal):

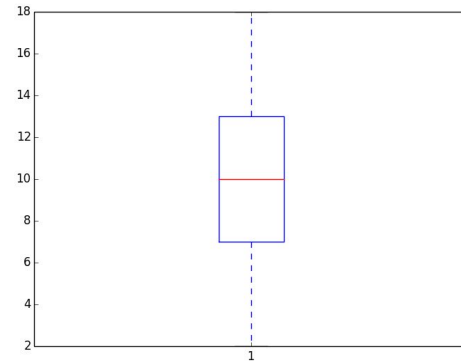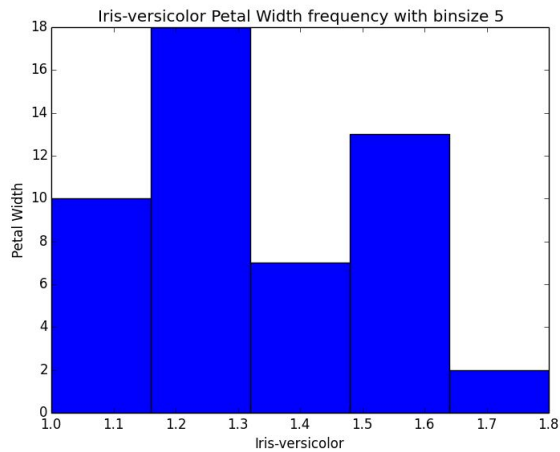## Iris-Setosa Sepal Width Histogram and Boxplot (symmetric, unimodal):

Iris-Setosa Sepal Length Histogram and Boxplot (Symmetric, Unimodal):
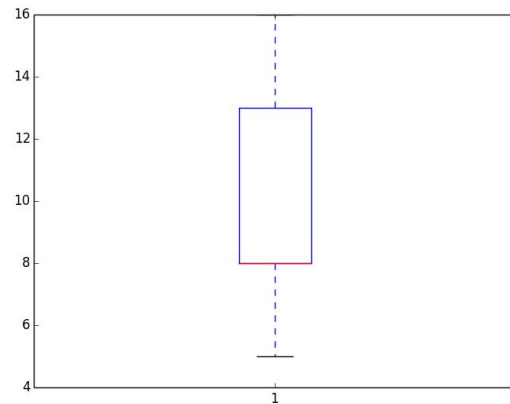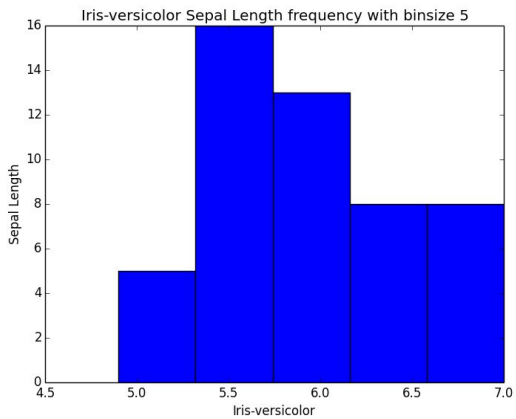

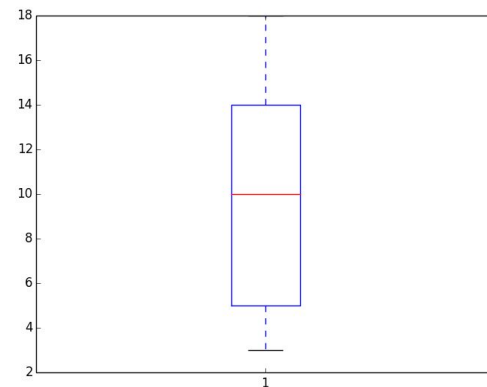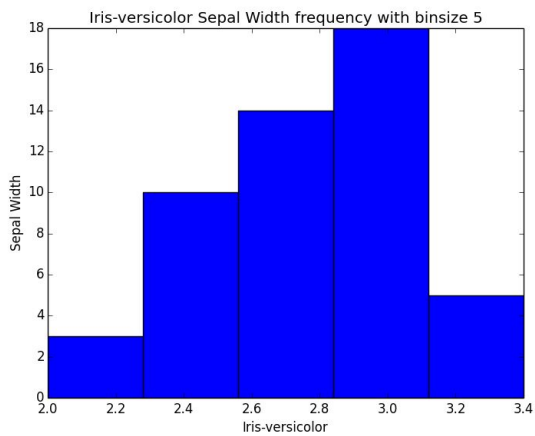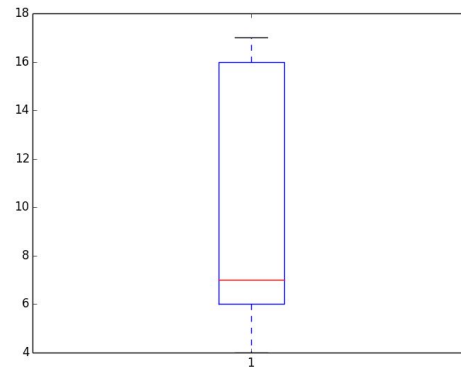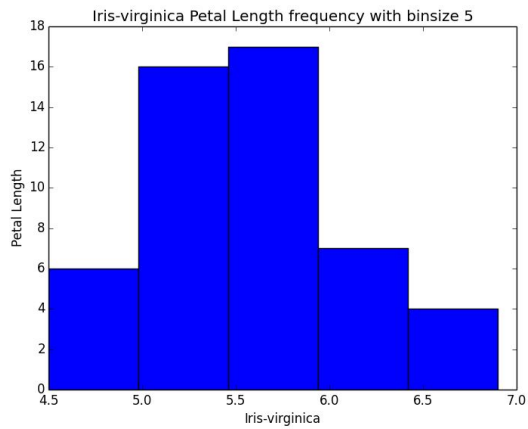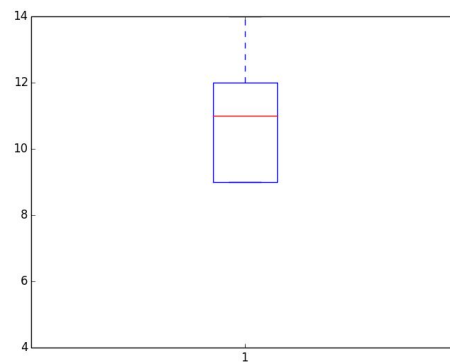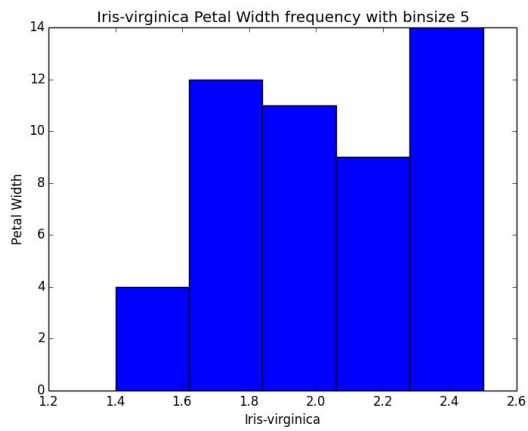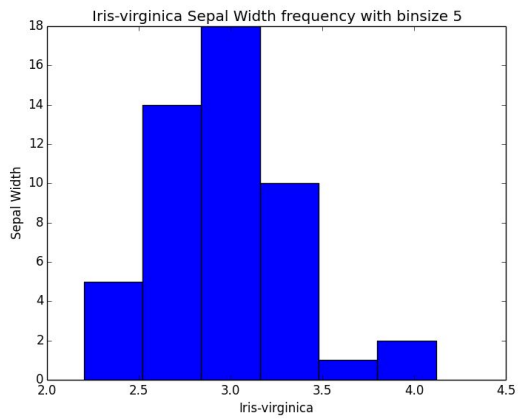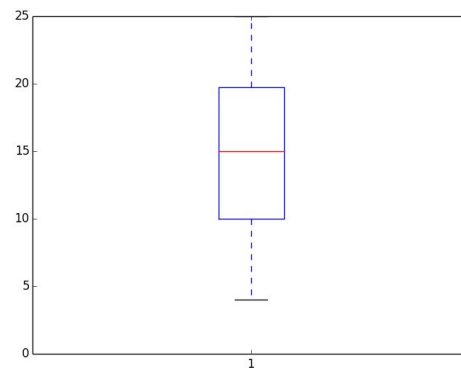## Iris-Versicolor Petal Length Histogram and BoxPlot (Skewed, Unimodal):

Iris-Versicolor Petal Width Histogram and BoxPlot (Skewed, Bimodal):

Iris-versicolor Petal Width frequency with binsize 5

Iris-Versicolor Sepal Length Histogram and BoxPlot (Skewed, Unimodal):

Iris-versicolor Sepal Length frequency with binsize 5

Iris-Versicolor Sepal Width Histogram and BoxPlot (Skewed, Unimodal):

Iris-versicolor Sepal Width frequency with binsize 5

Iris-Virginica Petal Length Histogram and BoxPlot (Symmetric, Unimodal):

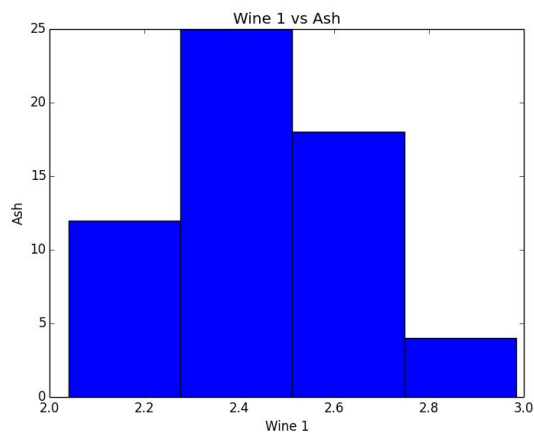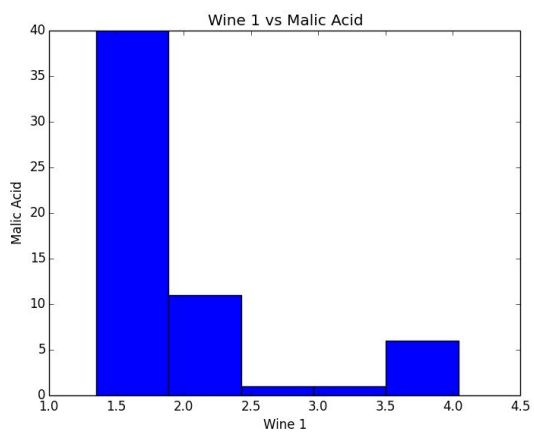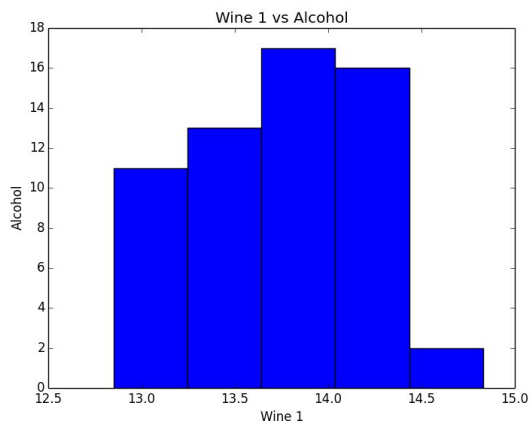Iris-Virginica Petal Width Histogram and BoxPlot (Skewed, mostly uniform):

Iris-Virginica Sepal Length Histogram and BoxPlot (Symmetric, Unimodal):
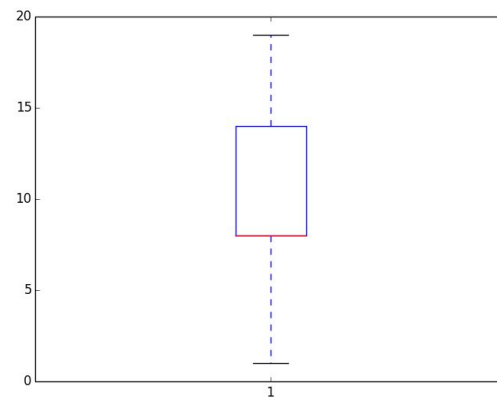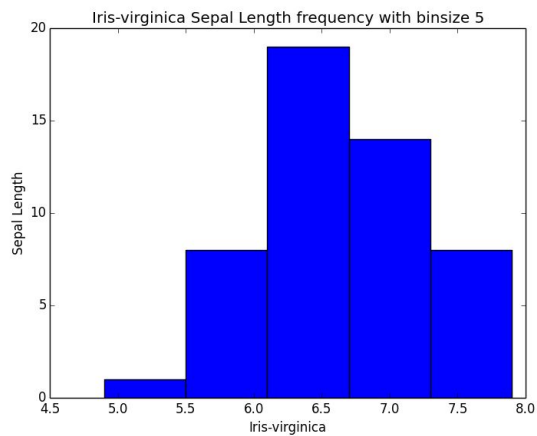
Iris-Virginica Sepal Width Histogram and BoxPlot (Symmetric, Unimodal):

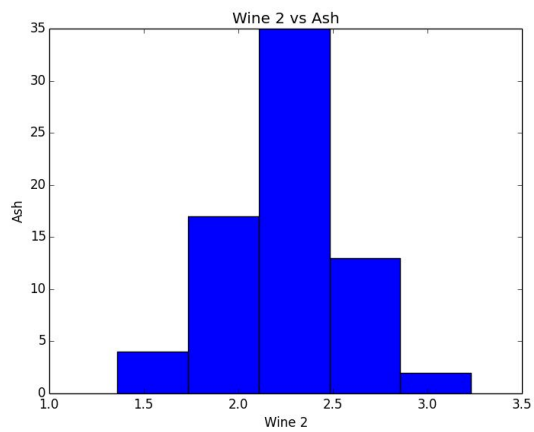Wine1 Ash Histogram and Boxplot (mostly symmetric, unimodal)

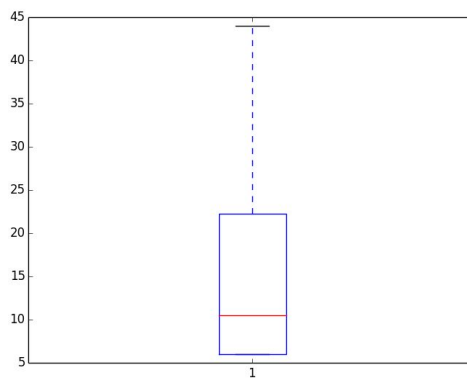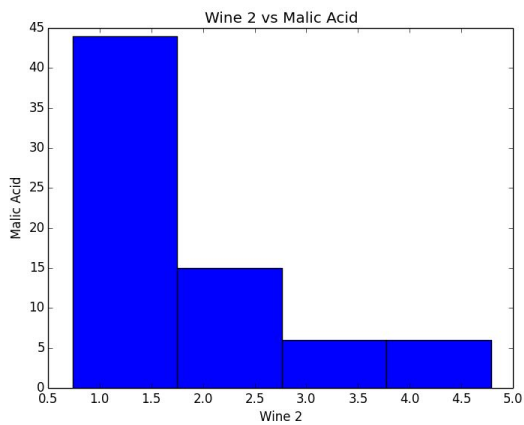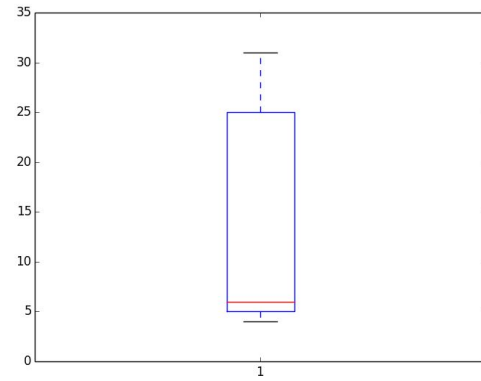Wine1 Malic Acid Histogram and Boxplot (mostly skewed, unimodal)

Wine1 Alcohol Histogram and Boxplot (mostly symmetric, mostly uniform)
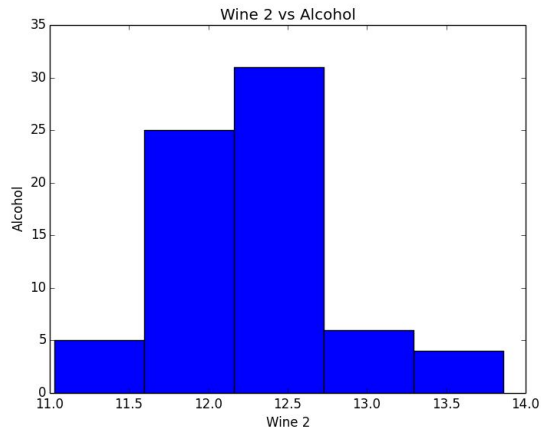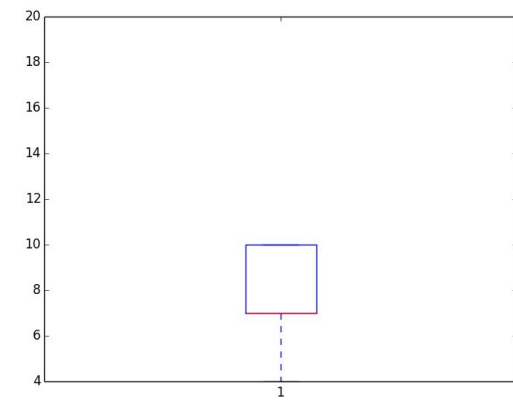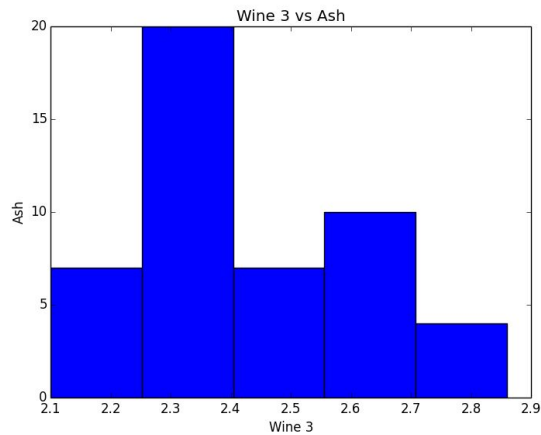
Wine2 Ash Histogram and Boxplot (mostly symmetric, unimodal)

Wine2 Malic Acid Histogram and Boxplot (mostly skewed, unimodal)
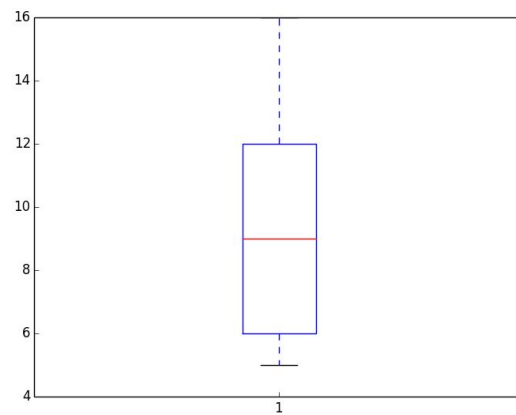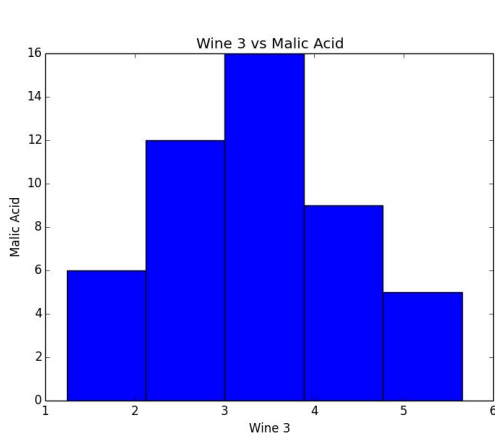
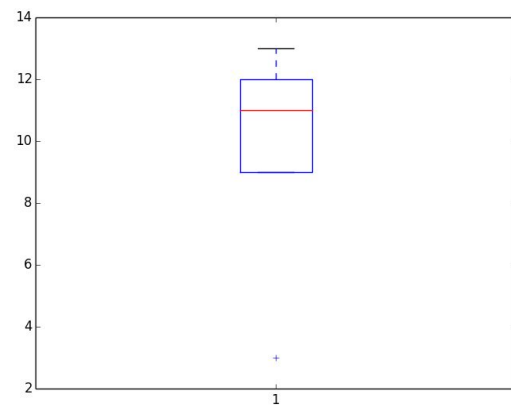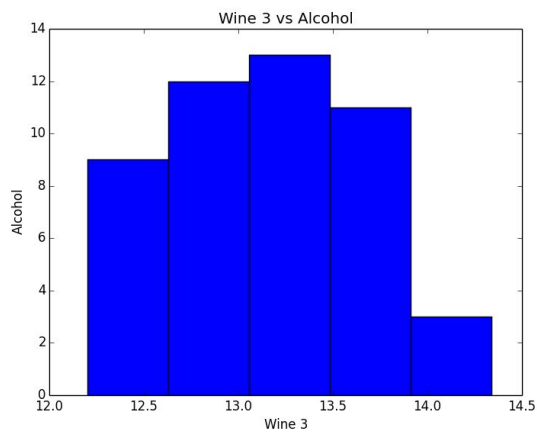# Wine2 Alcohol Histogram and Boxplot (mostly symmetric, unimodal)

Wine3 Ash Histogram and Boxplot (mostly skewed, bimodal)

Wine3 Malic Acid Histogram and Boxplot (mostly symmetric, unimodal)
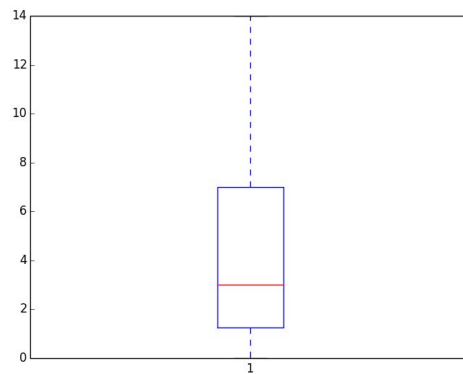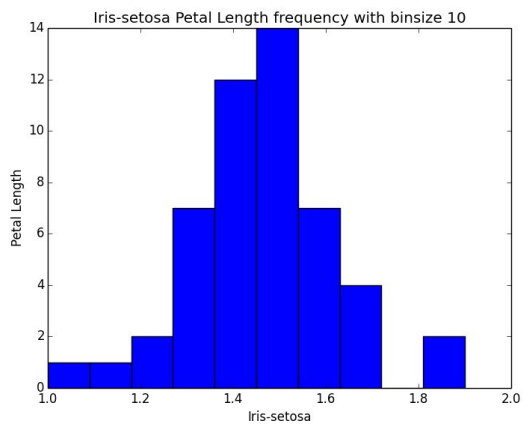
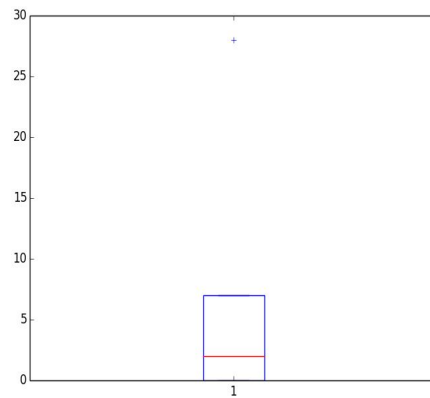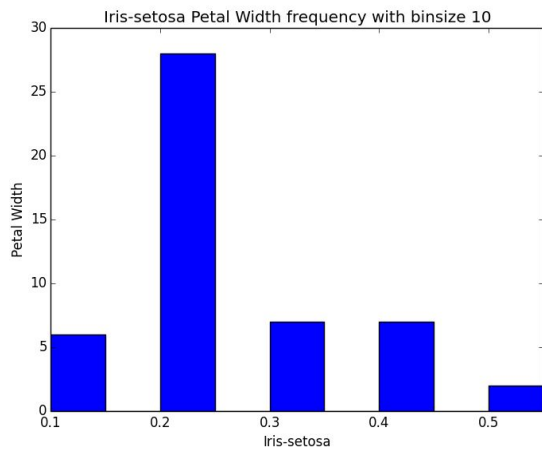Wine3 Alcohol Histogram and Boxplot (mostly symmetric, mostly uniform)
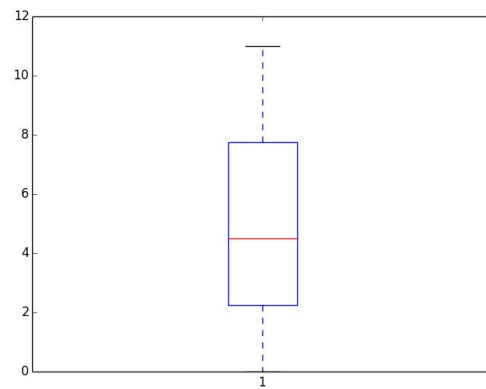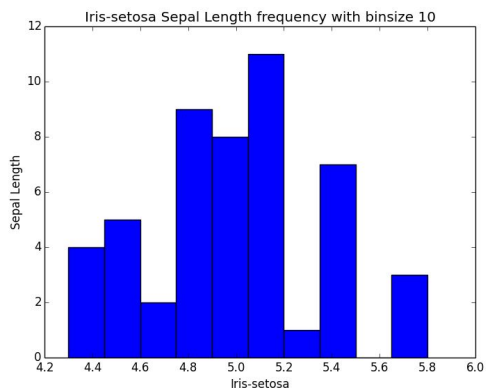

Bin Size: 10

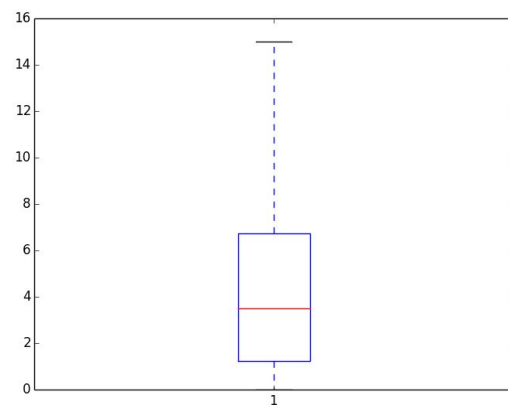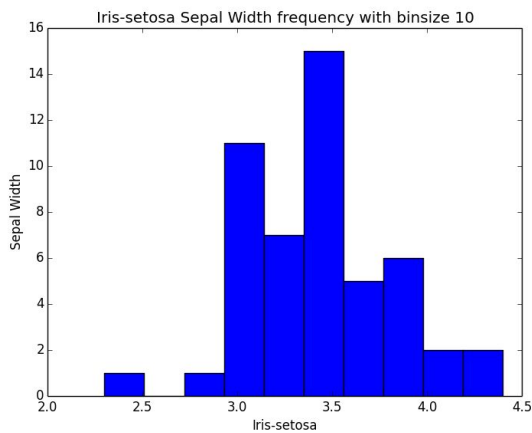Iris-Setosa Petal Length Histogram and Boxplot (symmetric, mostly unimodal):

Iris-Setosa Petal Width Histogram and Boxplot (Skewed, mostly unimodal):
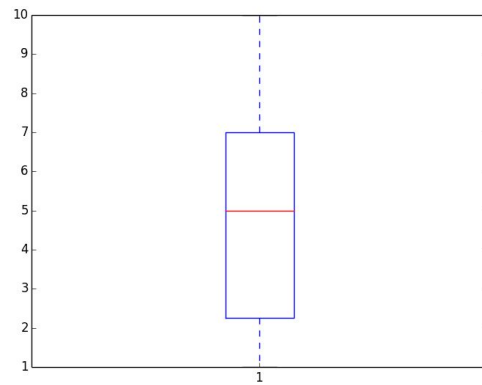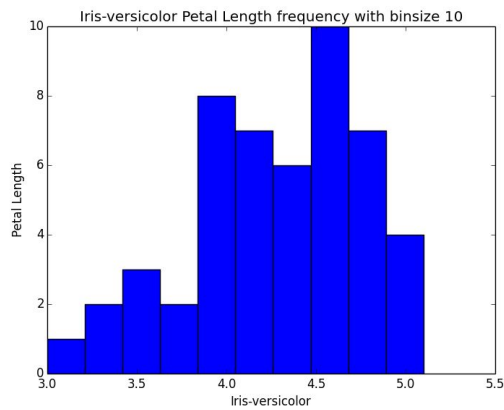
Iris-Setosa Sepal Length Histogram and Boxplot (symmetric, mostly unimodal):
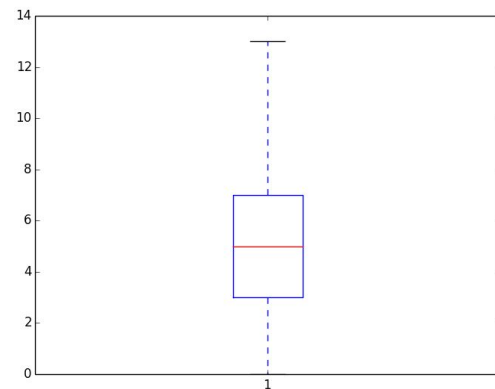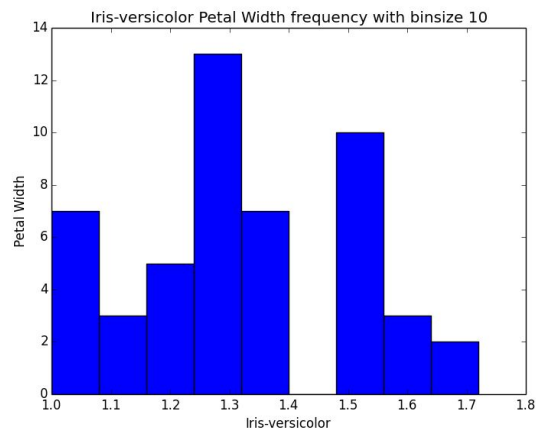

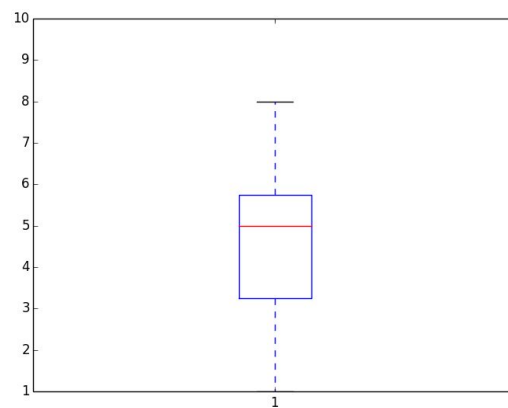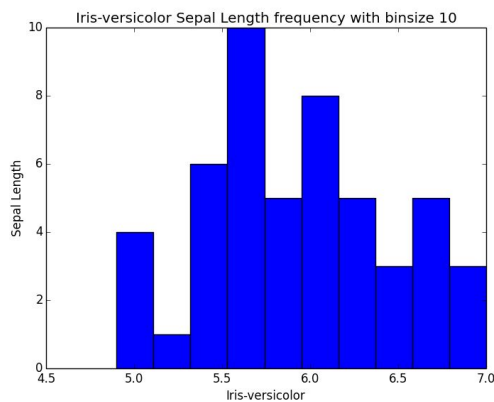Iris-Setosa Sepal Width Histogram and Boxplot (symmetric, multimodal):

Iris-Versicolor Petal Length Histogram and Boxplot (Skewed, multimodal):
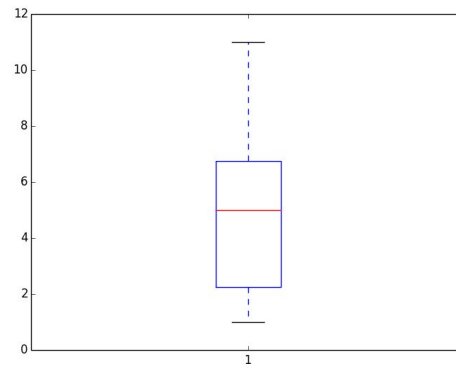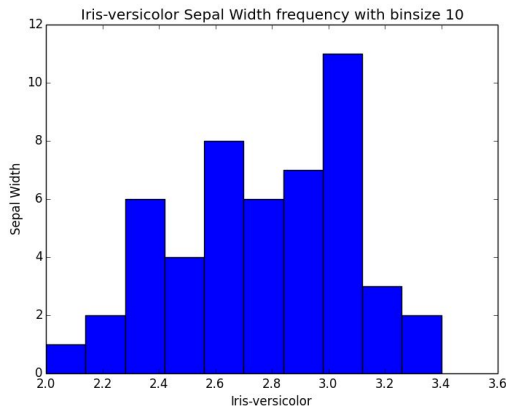
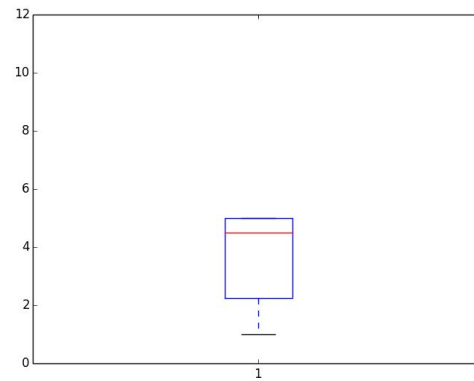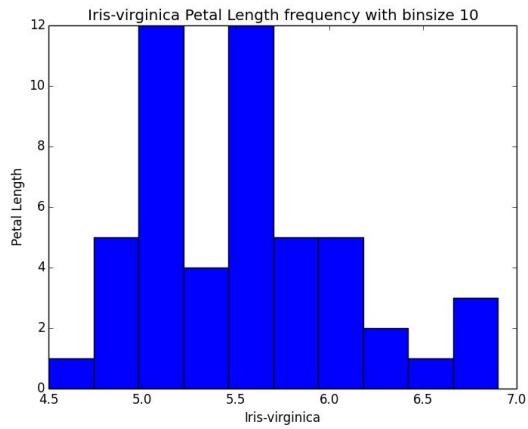Iris-Versicolor Petal Width Histogram and Boxplot (Skewed, bimodal):

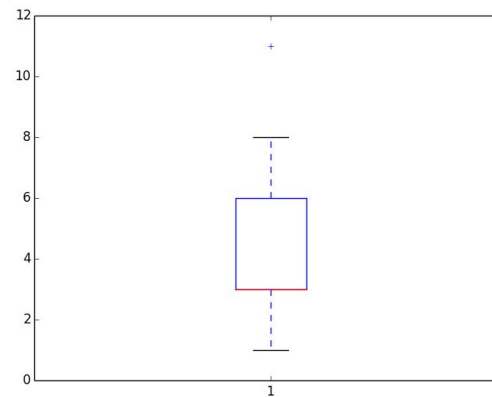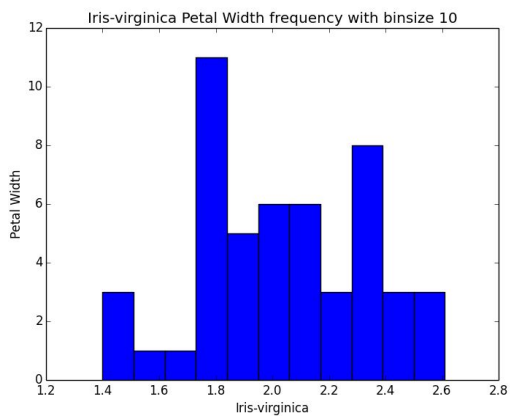Iris-versicolor Sepal Length Histogram and Boxplot (skewed, multimodal):

Iris-Versicolor Sepal Width Histogram and Boxplot (Skewed, multimodal):

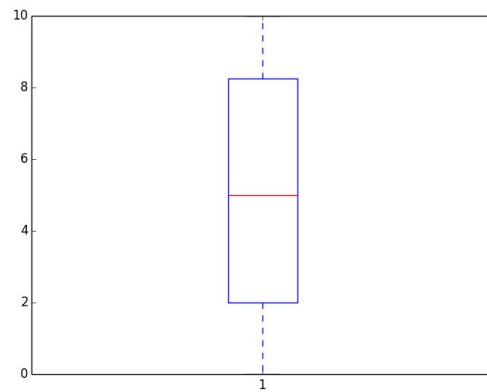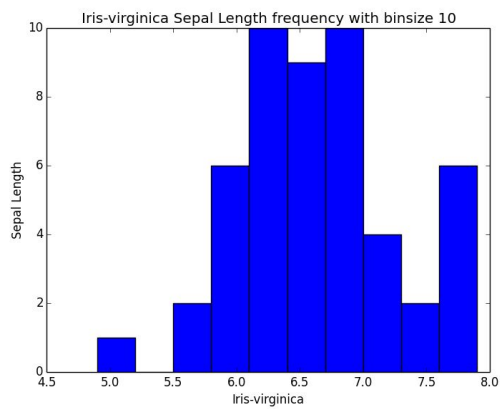Iris-versicolor Sepal Width frequency with binsize 10

Iris-virginica Petal Length Histogram and Boxplot (Skewed, bimodal):

Iris-virginica Petal Length frequency with binsize 10
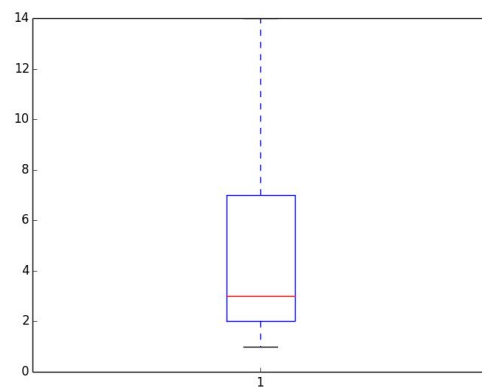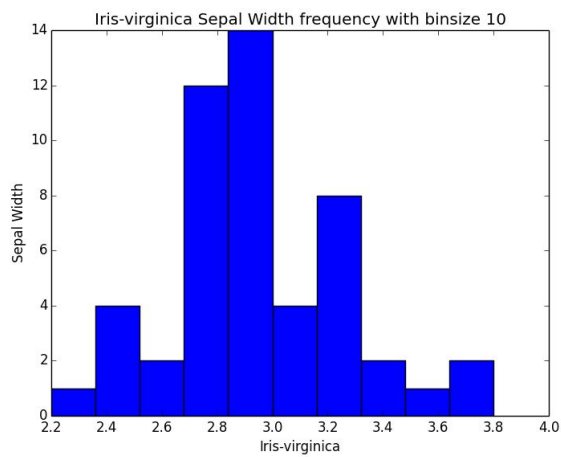
Iris-virginica Petal Width Histogram and Boxplot (Skewed, multimodal):

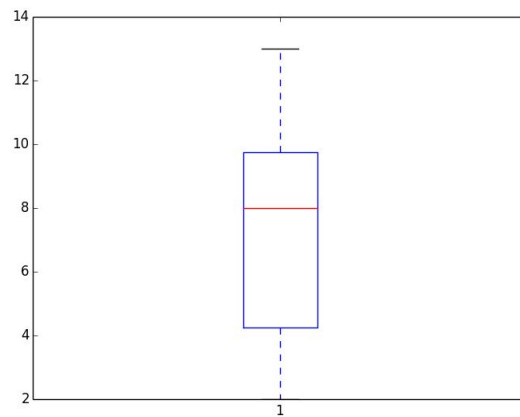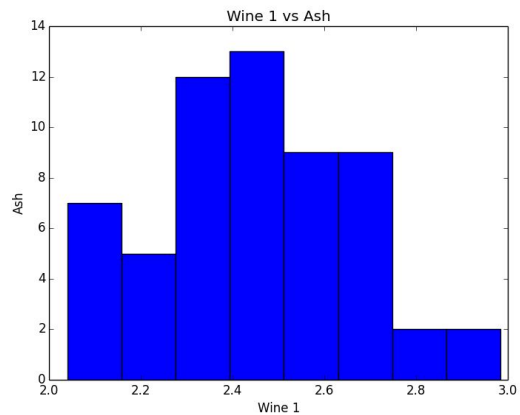
Iris-virginica Petal Width frequency with binsize 10

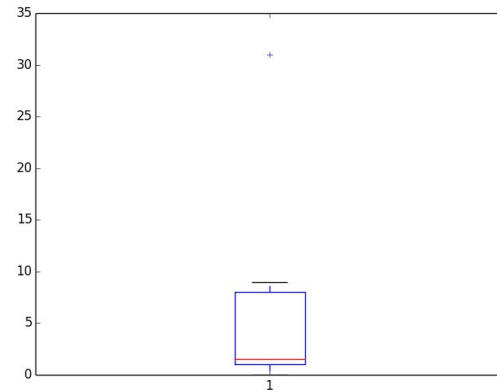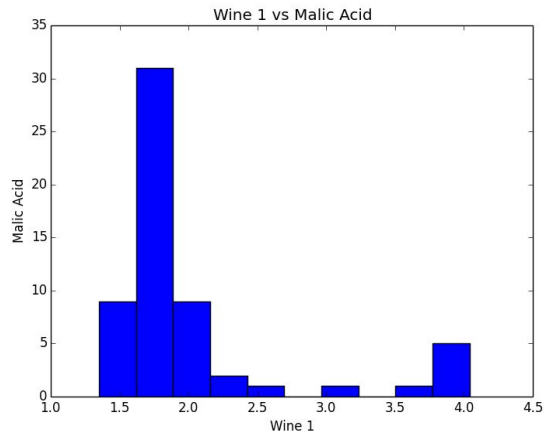## Iris-virginica Sepal Length Histogram and Boxplot (Skewed, mostly uniform):

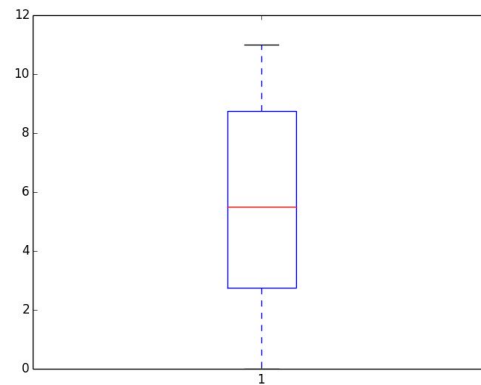## Iris-virginica Sepal Width Histogram and Boxplot (Symmetric, multimodal):
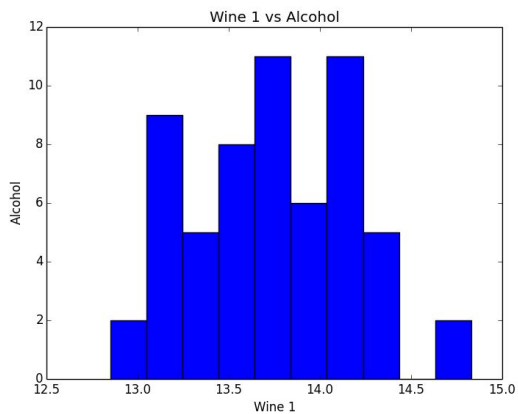

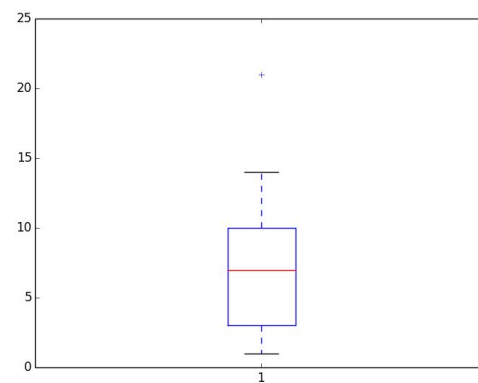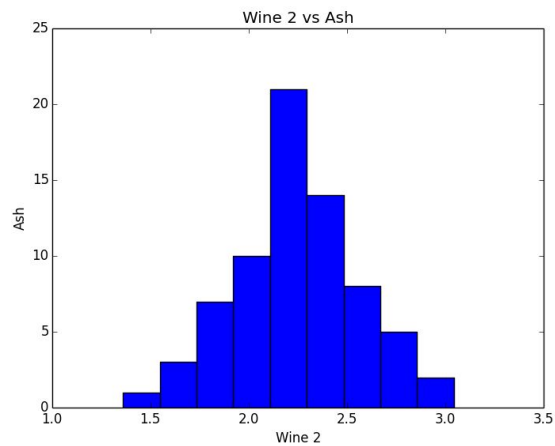## Wine1 Ash Histogram and Boxplot (mostly symmetric, unimodal)

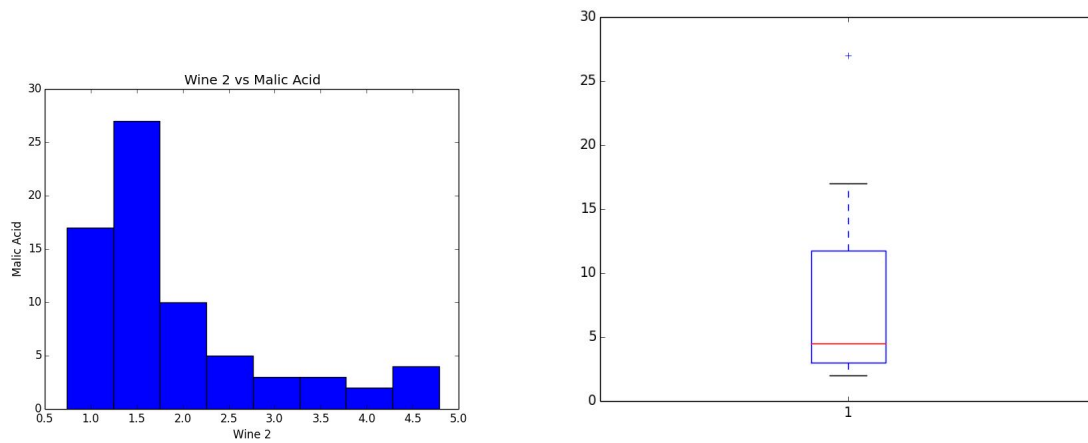## Wine1 Malic Acid Histogram and Boxplot (mostly skewed, unimodal)

Wine1 Alcohol Histogram and Boxplot (mostly symmetric, mostly uniform)
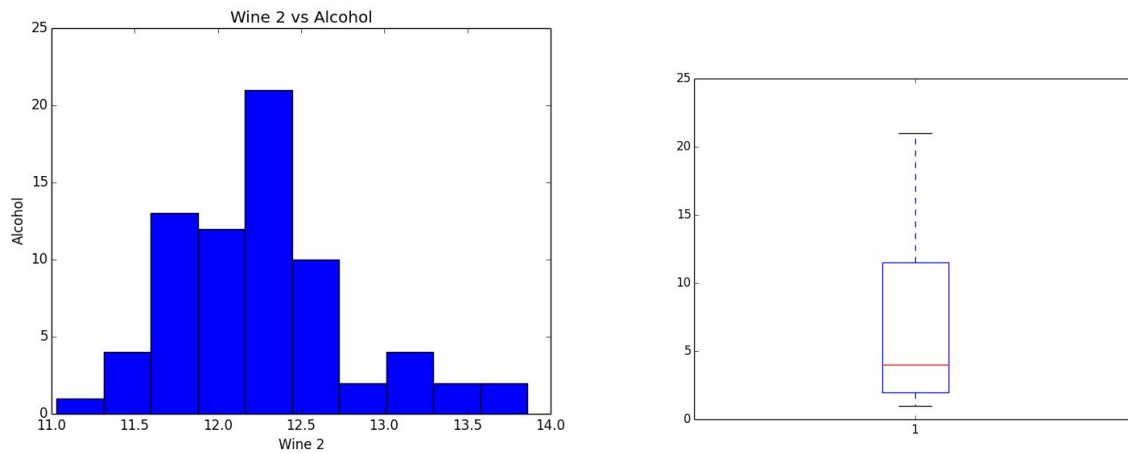

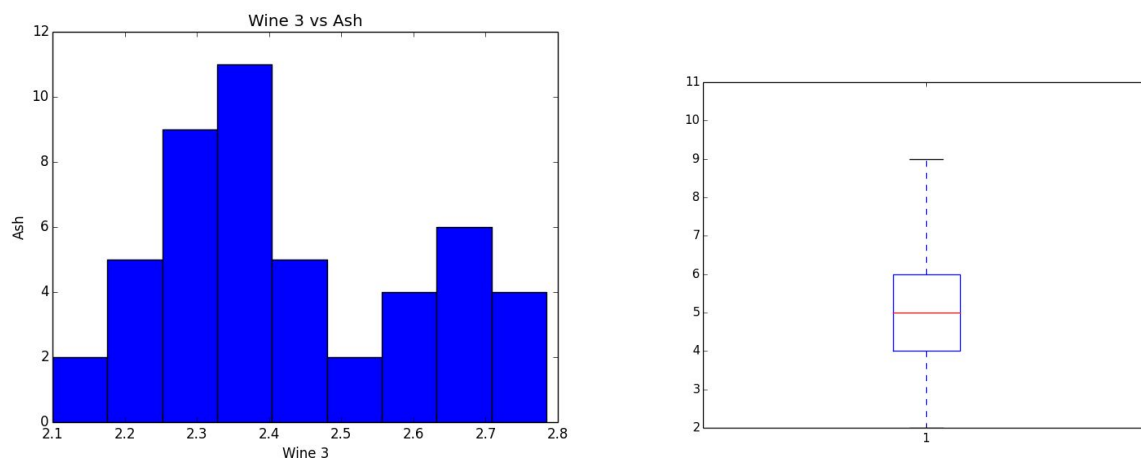## Wine2 Ash Histogram and Boxplot (mostly symmetric, unimodal)

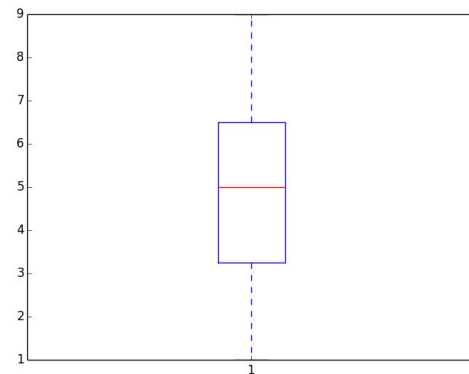# Wine2 Malic Acid Histogram and Boxplot (mostly skewed, unimodal)

Wine 2 vs Malic Acid
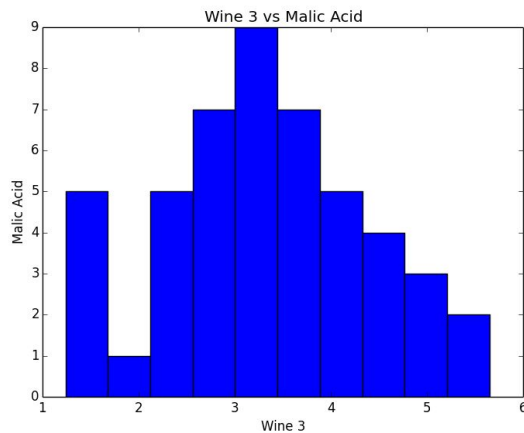
Wine2 Alcohol Histogram and Boxplot (mostly symmetric, bimodal)
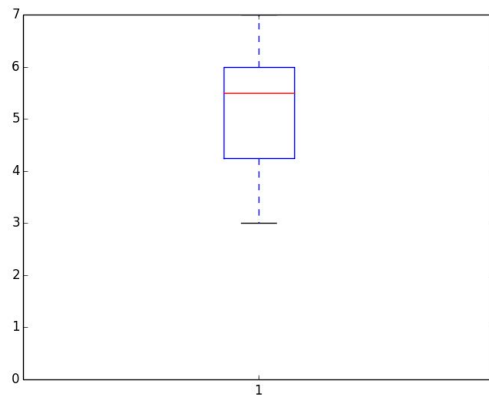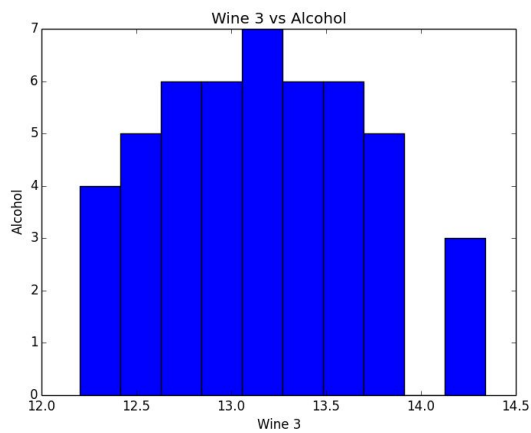

Wine 2 vs Alcohol

Wine3 Ash Histogram and Boxplot (mostly symmetric, bimodal)

Wine 3 vs Ash

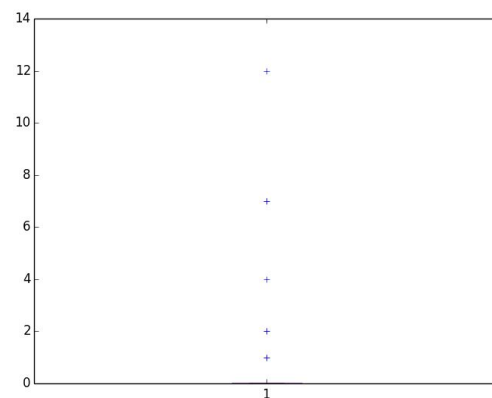Wine3 Malic Acid Histogram and Boxplot (mostly symmetric, unimodal)
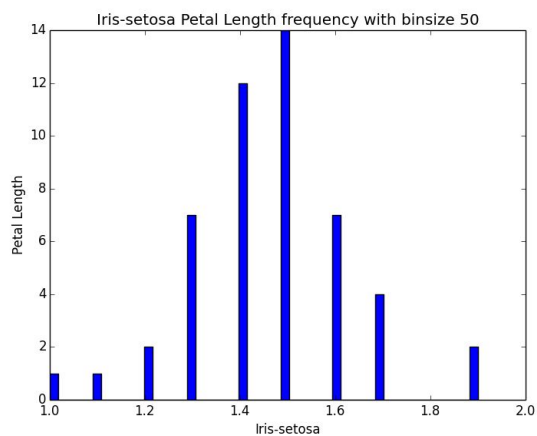

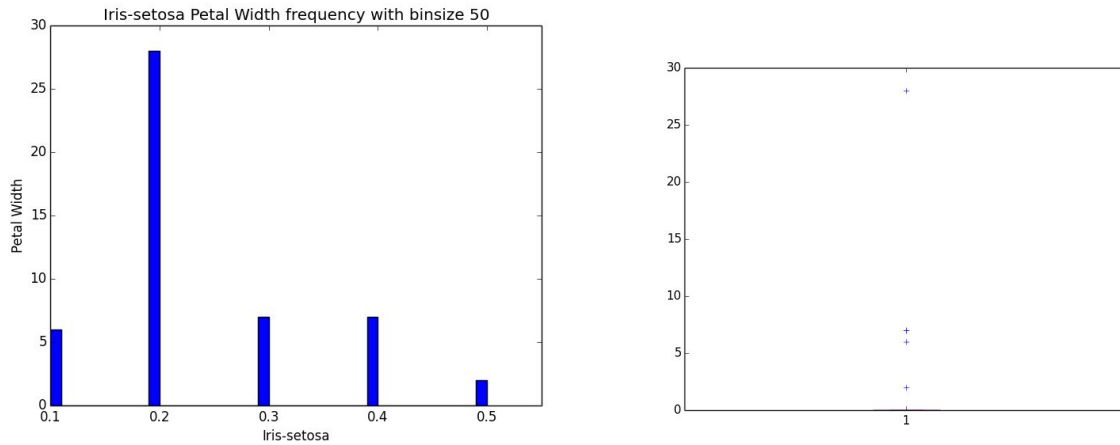Wine3 Alcohol Histogram and Boxplot (mostly symmertic, mostly uniform)
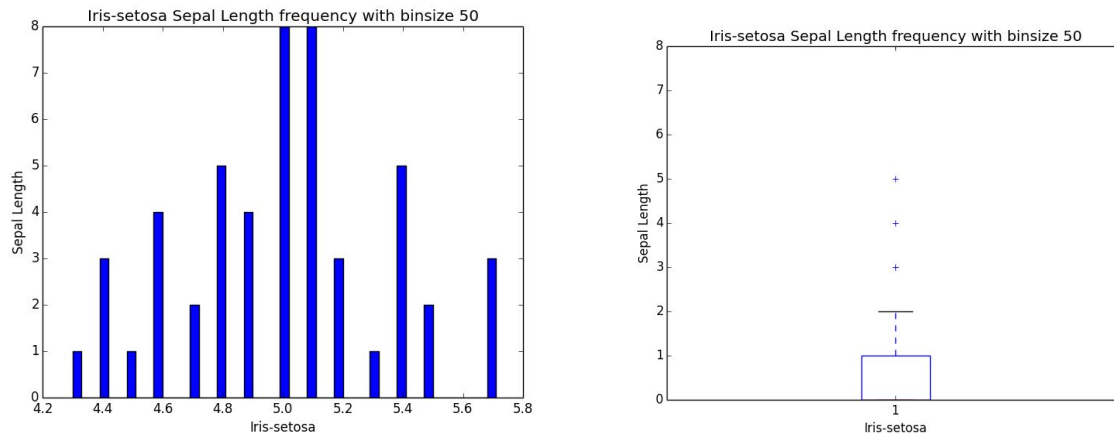

Bin Size: 50

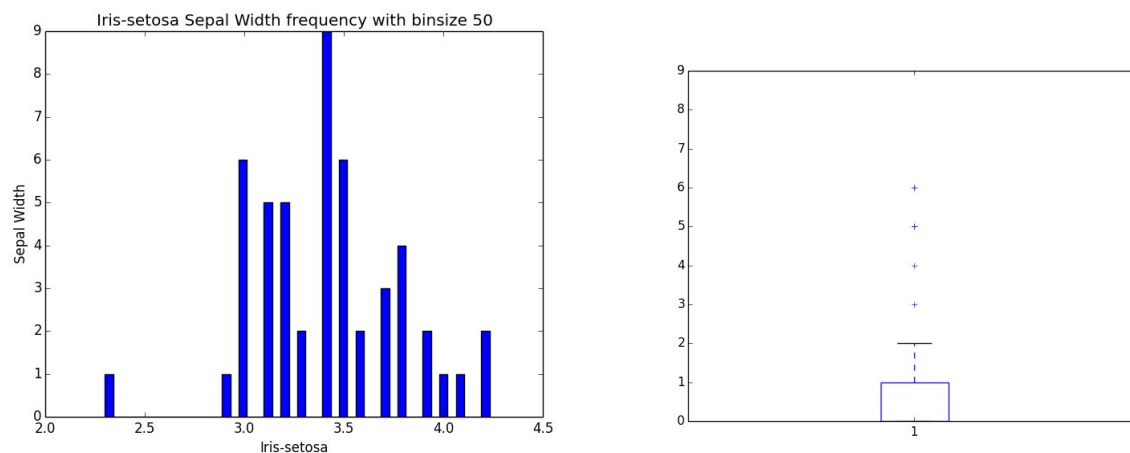Iris-Setosa Petal Length Histogram and Boxplot (mosly symmetric, mostly unimodal):

Iris-Setosa Petal Width Histogram and Boxplot (mostly skewed, mostly unimodal):
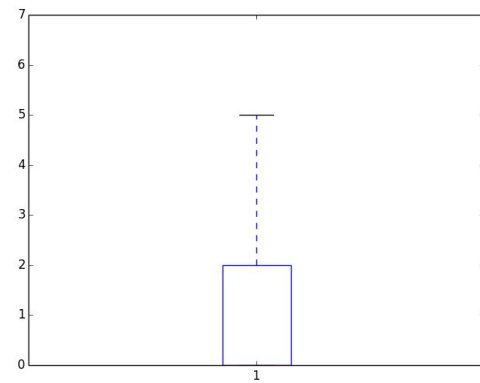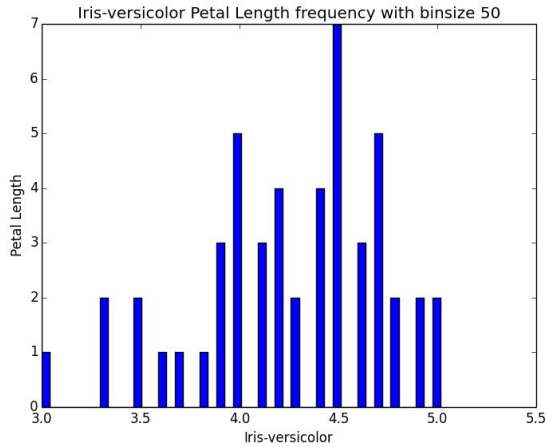

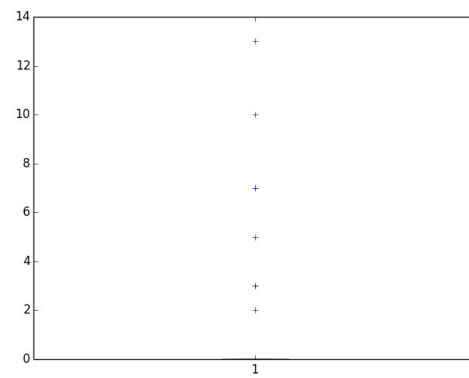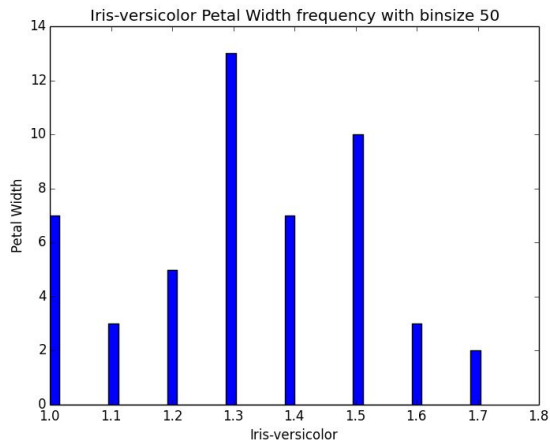# Iris-Setosa Sepal Length Histogram and Boxplot (symmetric, mostly unimodal):

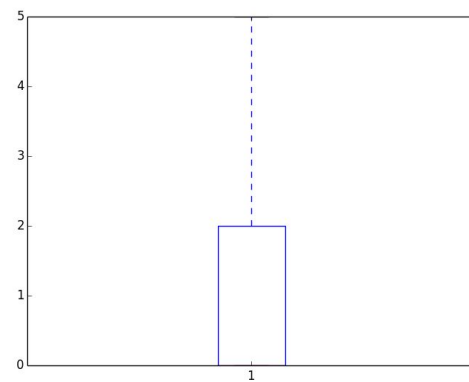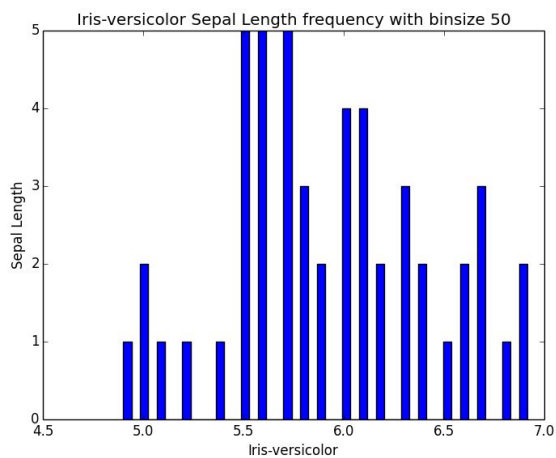# Iris-Setosa Sepal Width Histogram and Boxplot (mosly symmetric, mostly multimodal):

Iris-versicolor Petal Length Histogram and Boxplot (mosly skewed, mostly multimodal):

Iris-versicolor Petal Width Histogram and Boxplot (mosly symmetric, mostly bimodal):

Iris-versicolor Sepal Length Histogram and Boxplot (mosly skewed, mostly multimodal):

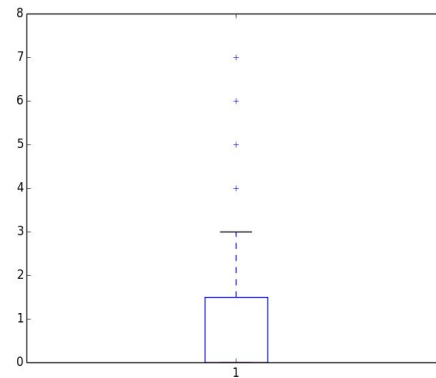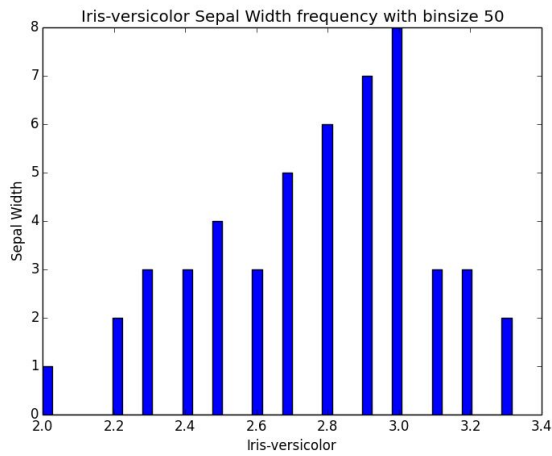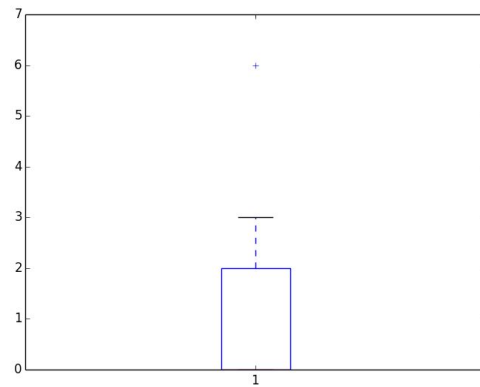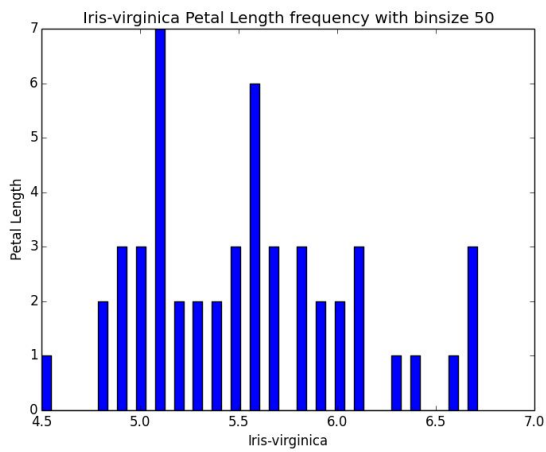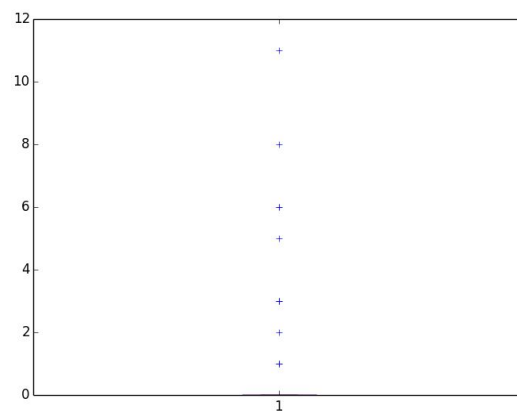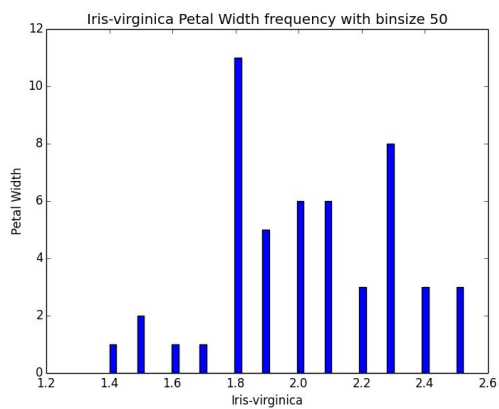

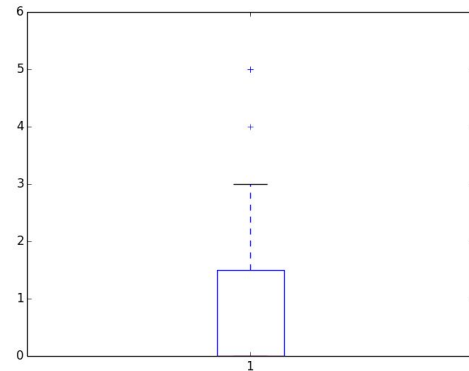Iris-versicolor Sepal Width Histogram and Boxplot (mosly skewed, mostly bimodal):
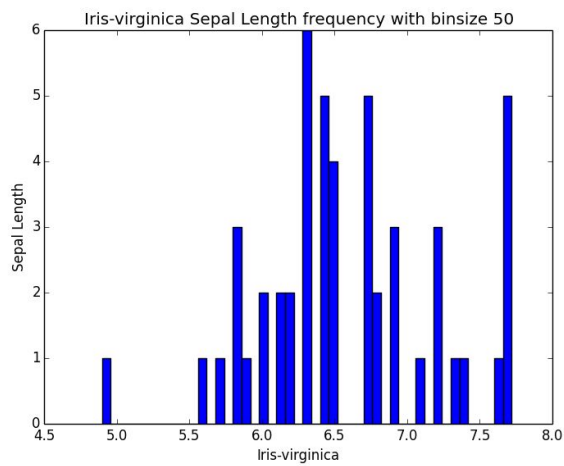
Iris-virginica Petal Length Histogram and Boxplot (mosly skewed, mostly multimodal):
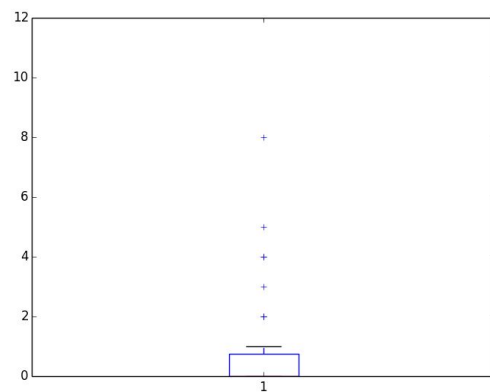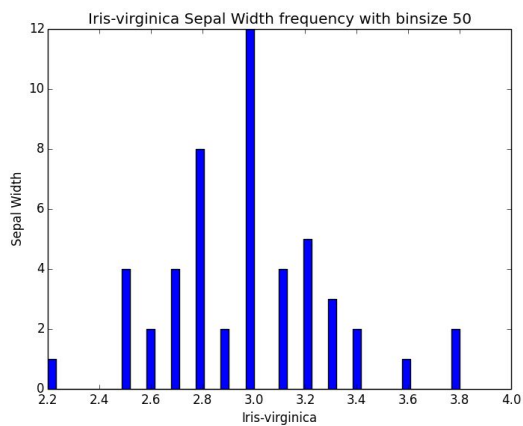

Iris-virginica Petal Width Histogram and Boxplot (mosly skewed, mostly bimodal):
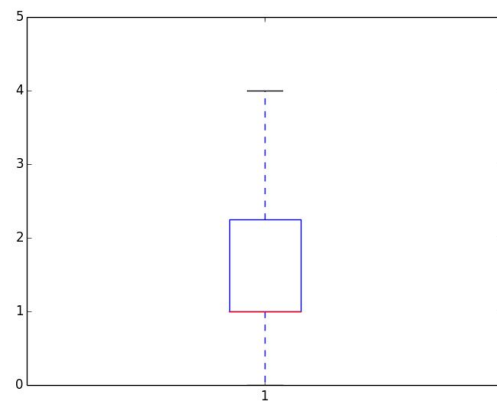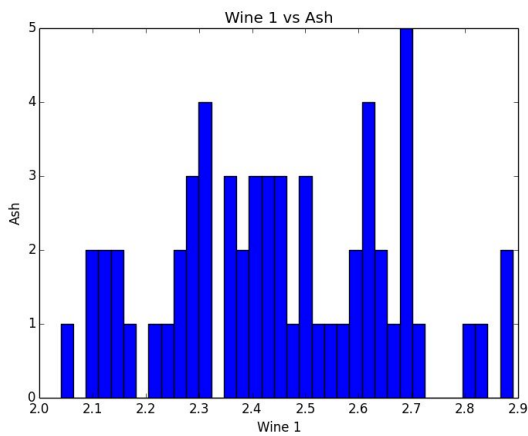
Iris-virginica Sepal Length Histogram and Boxplot (mosly symmetric, mostly multimodal):
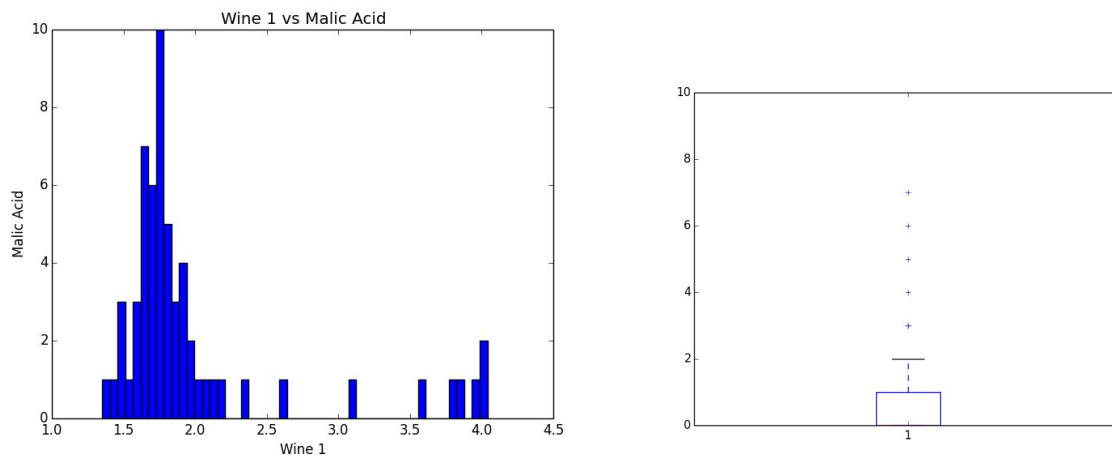

Iris-virginica Sepal Width Histogram and Boxplot (mosly symmetric, mostly multimodal):
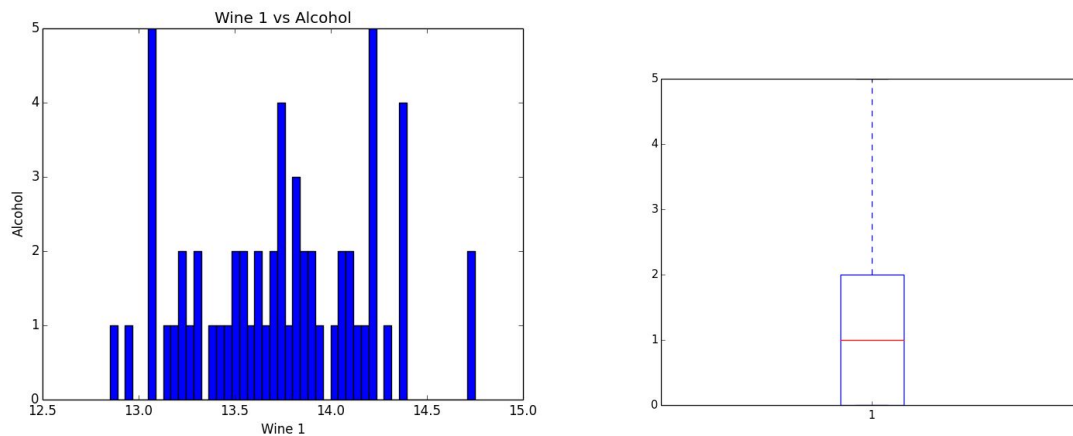

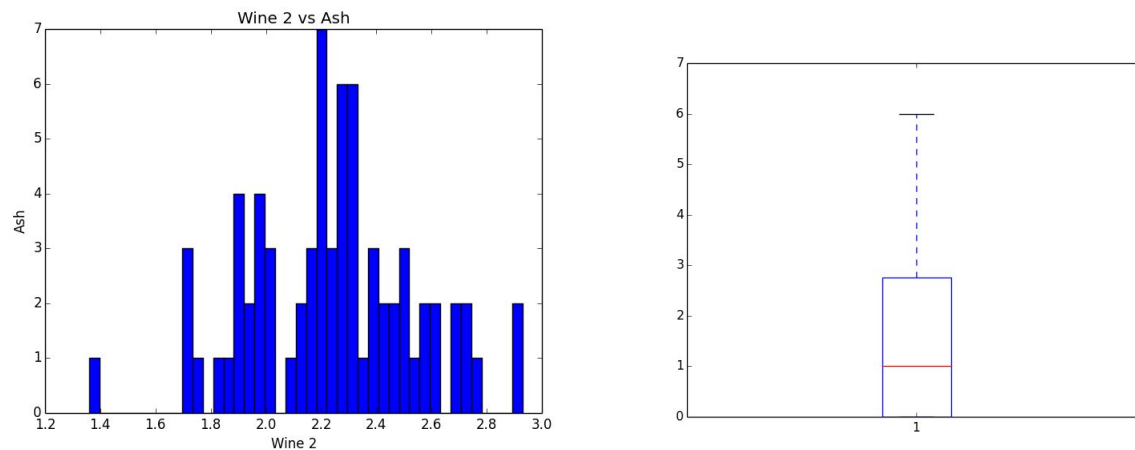## Wine1 Ash Histogram and Boxplot (mostly skewed, multimodal)

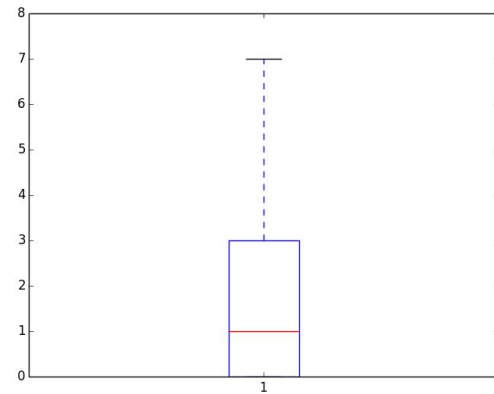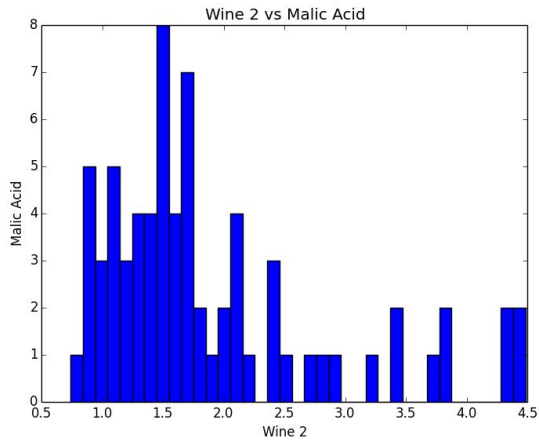# Wine1 Malic Acid Histogram and Boxplot (mostly skewed, multiimodal)

Wine1 Alcohol Histogram and Boxplot (mostly symmetric, bimodal)
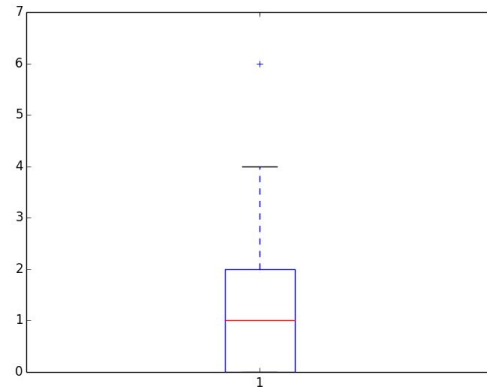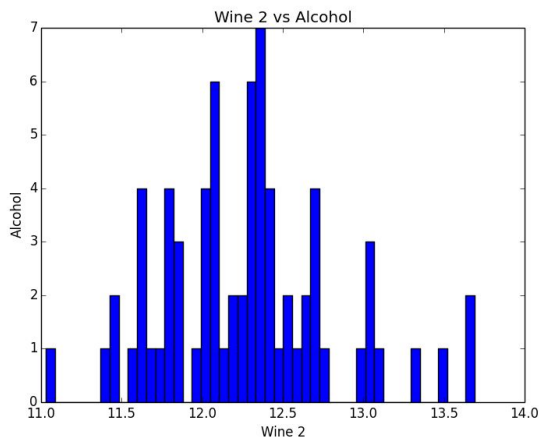

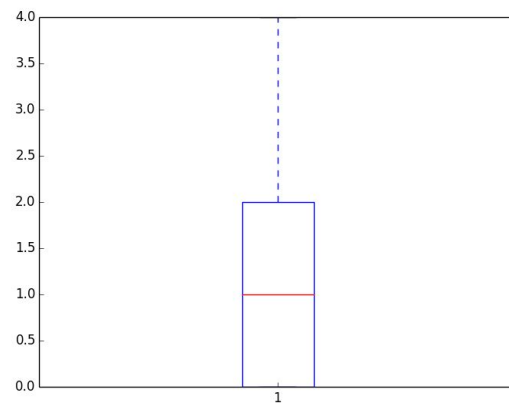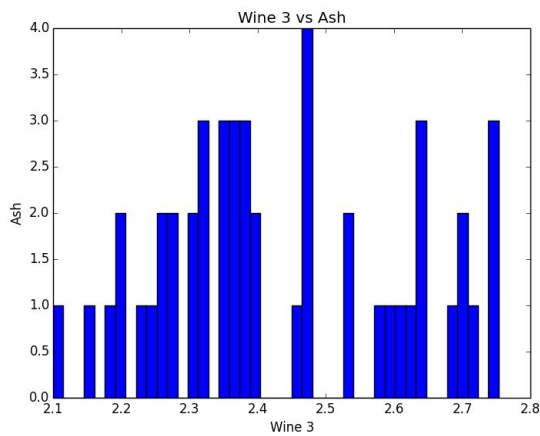# Wine2 Ash Histogram and Boxplot (mostly symmetric, multimodal)

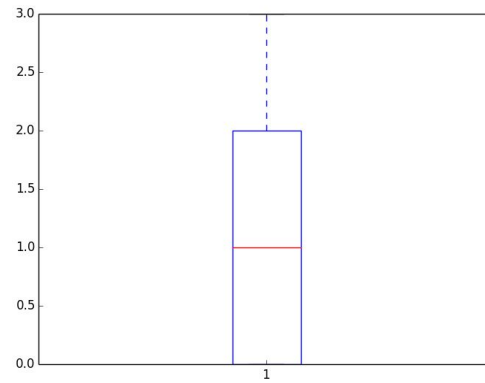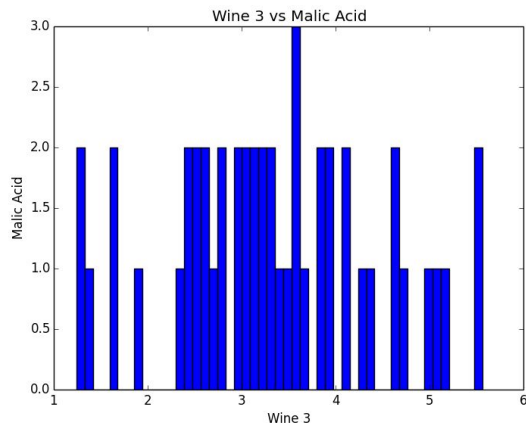## Wine2 Malic Acid Histogram and Boxplot (mostly skewed, multimodal)

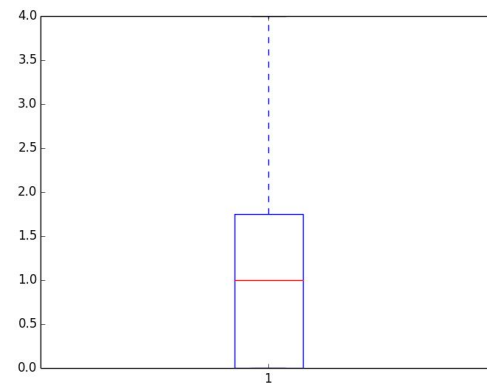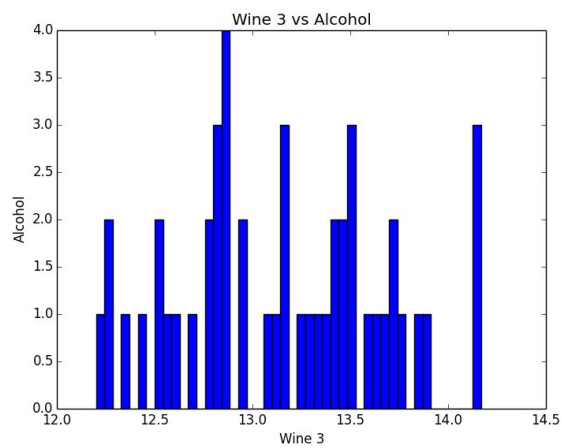## Wine2 Alcohol Histogram and Boxplot (mostly symmetric, multimodal)

Wine3 Ash Histogram and Boxplot (mostly symmetric, mostly uniform)

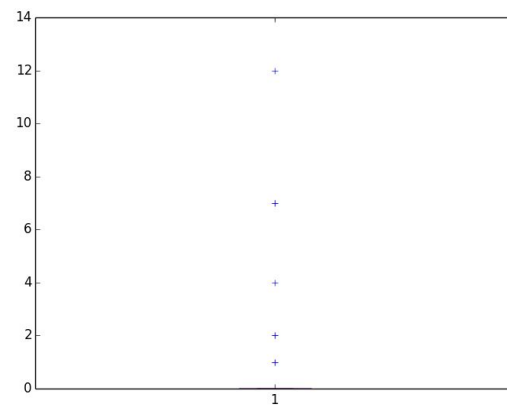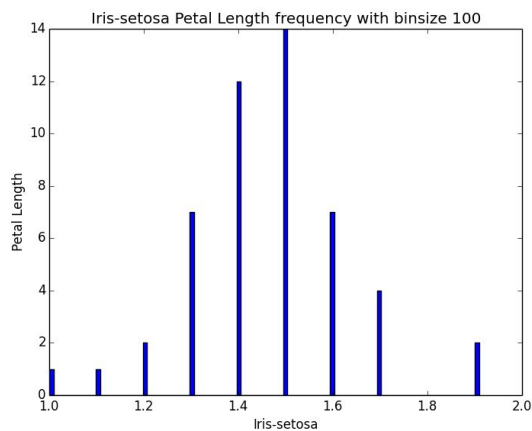## Wine3 Malic Acid Histogram and Boxplot (mostly symmetric, unimodal)

Wine3 Alcohol Histogram and Boxplot (mostly skewed, multimodal)
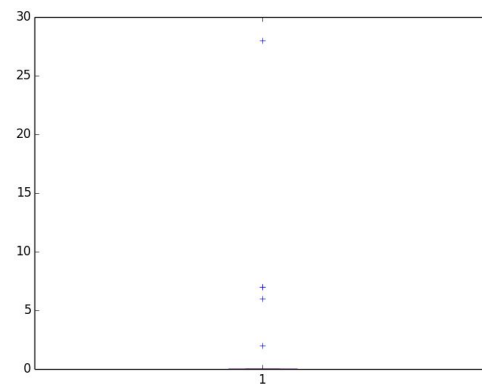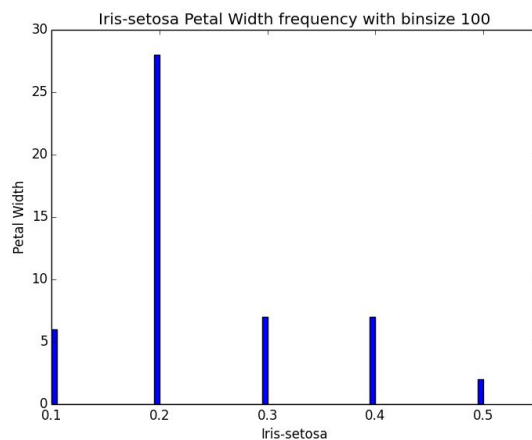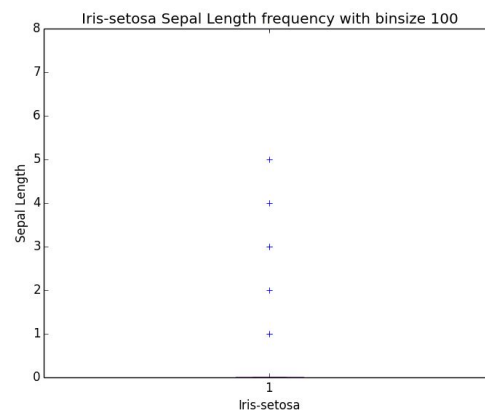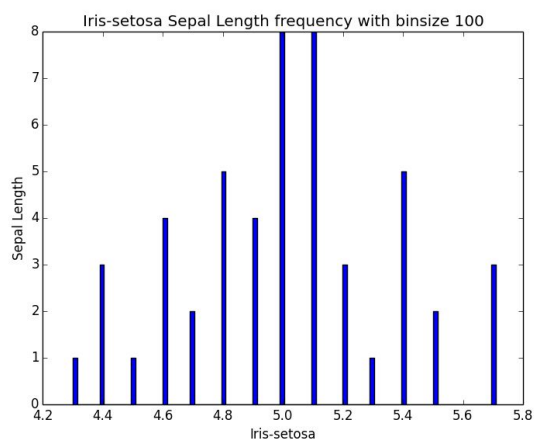

Bin Size: 100

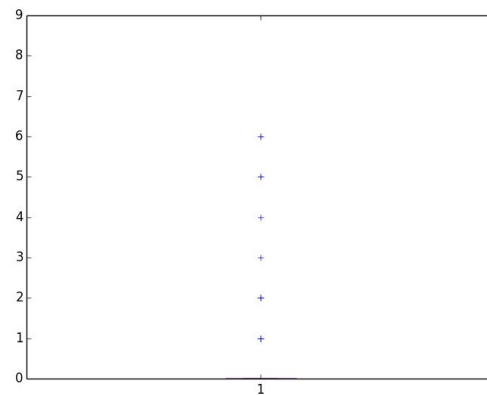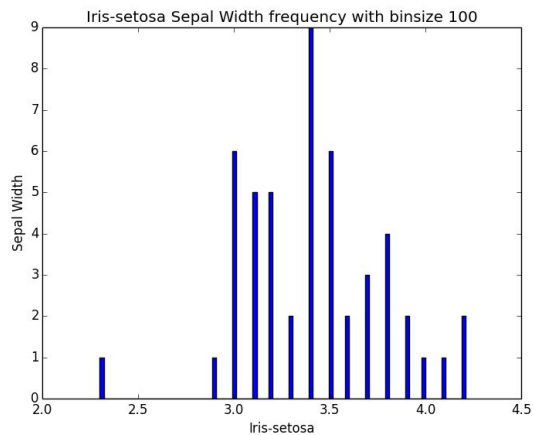Iris-Setosa Petal Length Histogram and Boxplot (mostly symmetric, mostly unimodal):

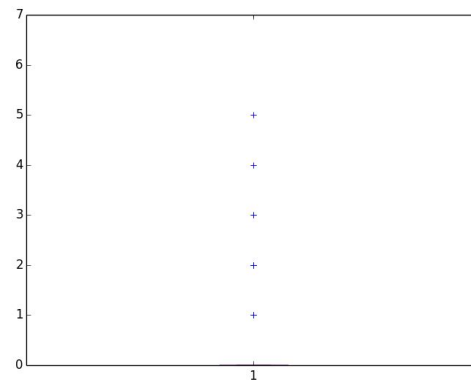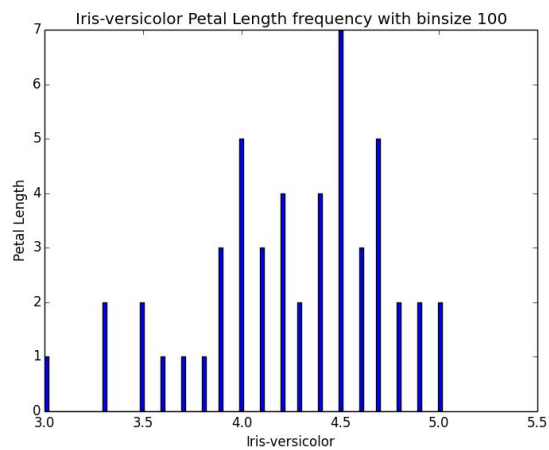Iris-Setosa Petal Width Histogram and Boxplot (mostly skewed, mostly unimodal):

Iris-Setosa Sepal Length Histogram and Boxplot (mostly symmetric, mostly multimodal):
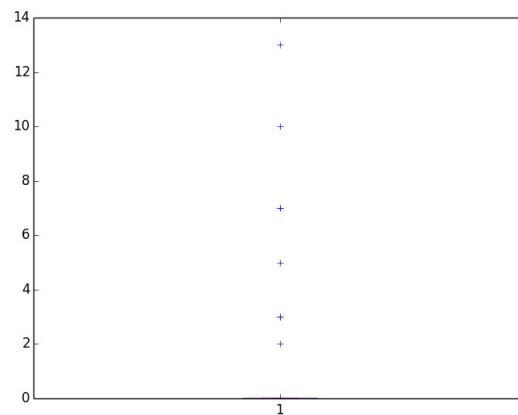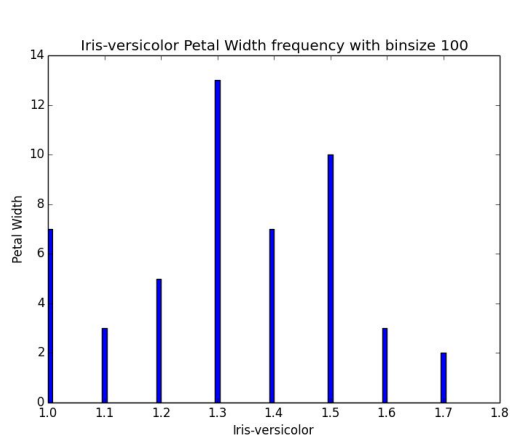

Iris-Setosa Sepal Width Histogram and Boxplot (mostly symmetric, mostly multimodal):

Iris-Versicolor Petal Length Histogram and Boxplot (mostly skewed, mostly multimodal):

Iris-Versicolor Petal Width Histogram and Boxplot (mostly symmetric, bimodal):

Iris-Versicolor Sepal Length Histogram and Boxplot (mostly skewed, mostly multimodal):

Iris-Versicolor Sepal Width Histogram and Boxplot (mostly skewed, mostly bimodal):

Iris-Virginica Petal Length Histogram and Boxplot (mostly skewed, mostly multimodal):

Iris-Virginica Petal Width Histogram and Boxplot (mostly skewed, mostly bimodal):

Iris-Virginica Sepal Length Histogram and Boxplot (mostly symmetric, mostly multimodal):

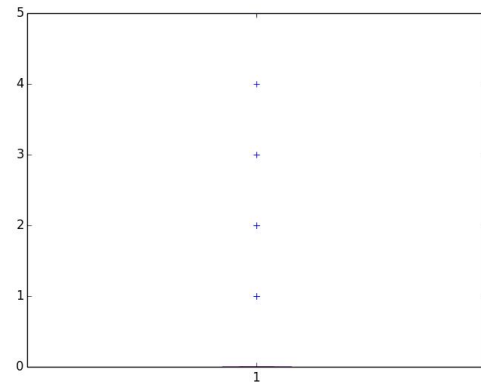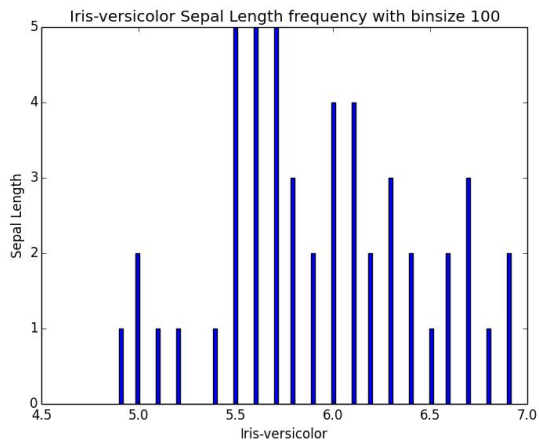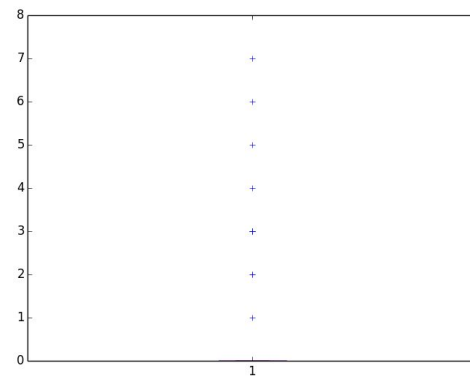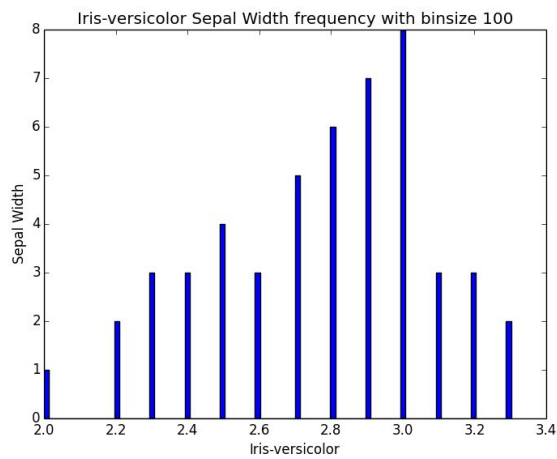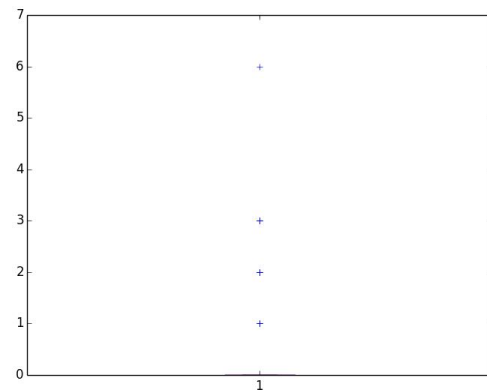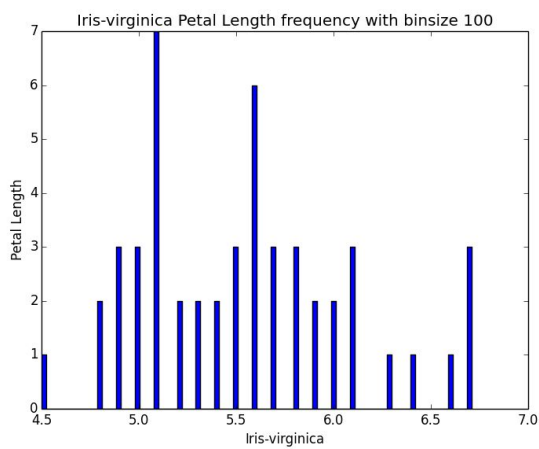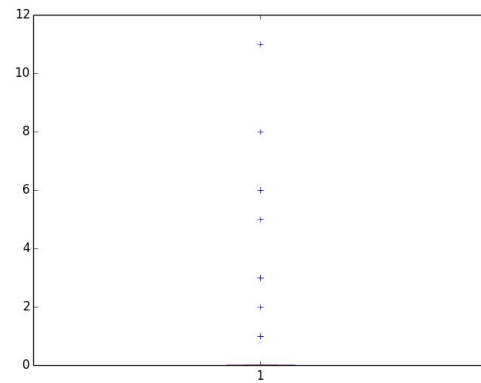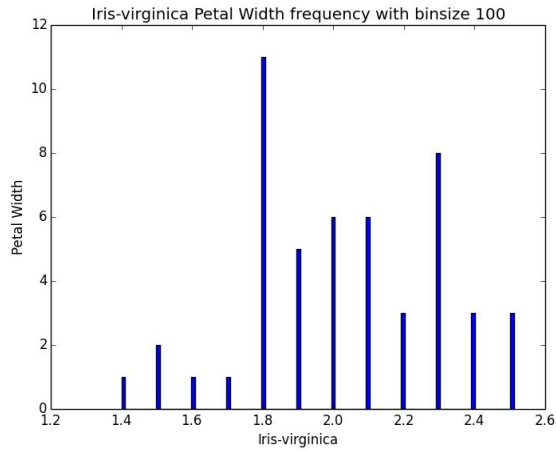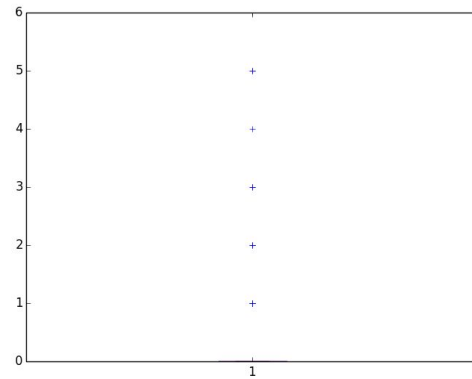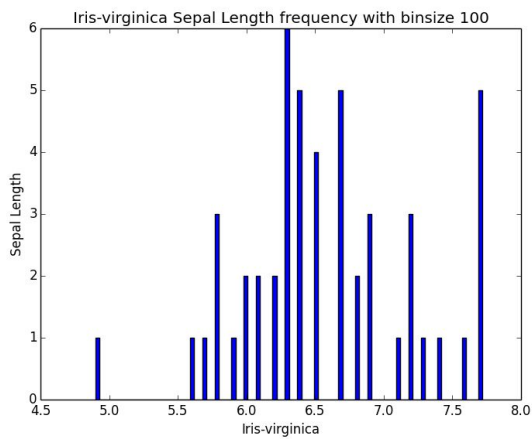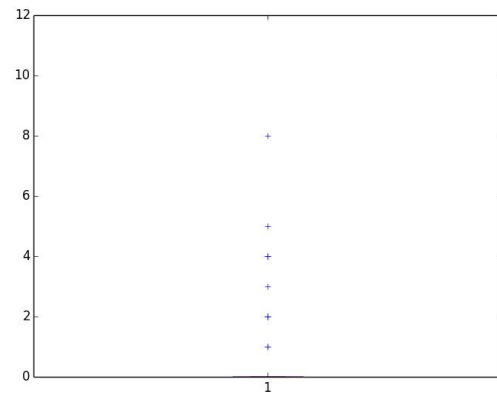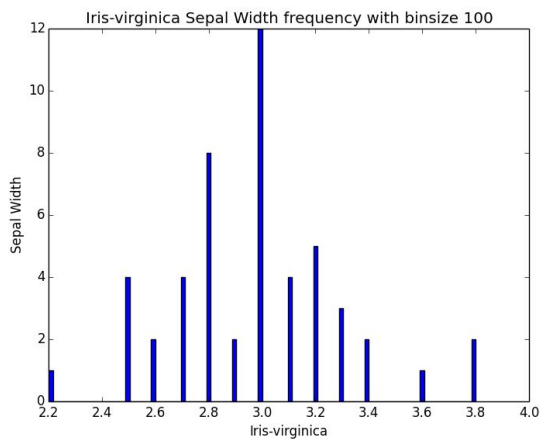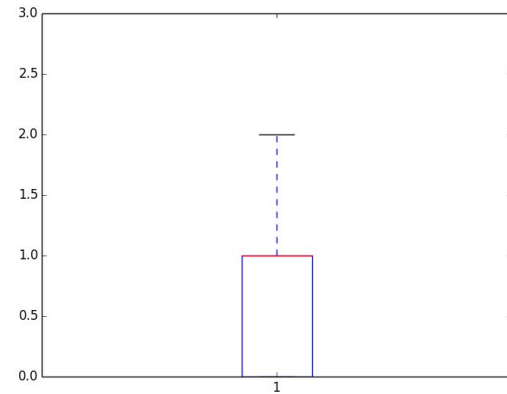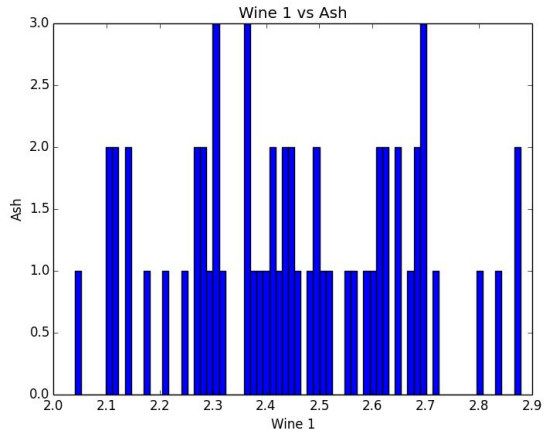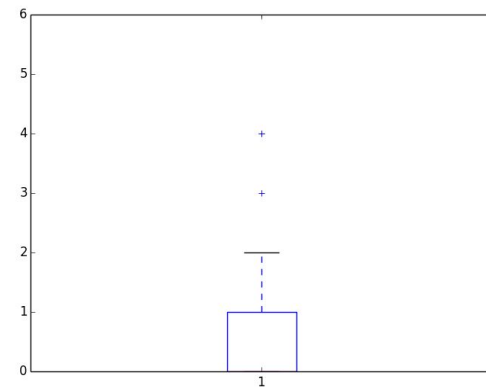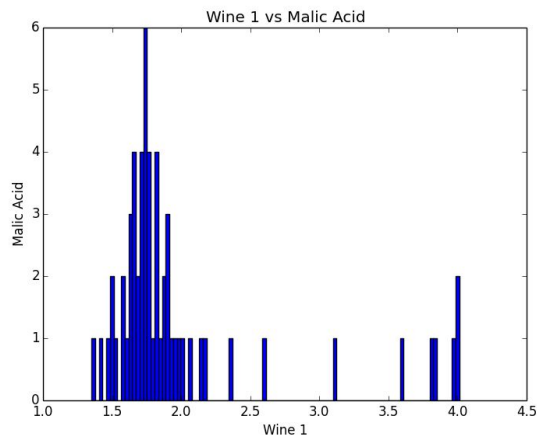Iris-Virginica Sepal Width Histogram and Boxplot (mostly symmetric,mostly multimodal):

Wine1 Ash Histogram and Boxplot (mostly symmetric, mostly uniform)

Wine1 Malic Acid Histogram and Boxplot (mostly skewed, multimodal)

Wine1 Alcohol Histogram and Boxplot (mostly skewed, multimodal)

Wine2 Ash Histogram and Boxplot (mostly symmetric, multimodal)

Wine2 Malic Acid Histogram and Boxplot (mostly skewed, multimodal)

Wine2 Alcohol Histogram and Boxplot (mostly symmetric, multimodal)

Wine3 Ash Histogram and Boxplot (mostly symmetric, multimodal)
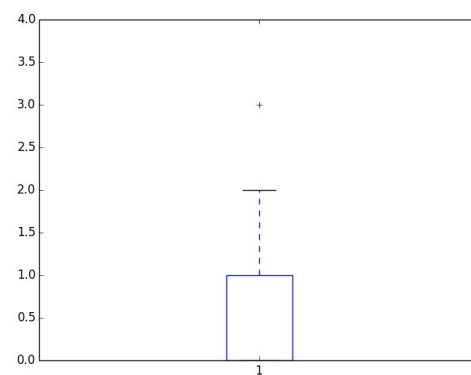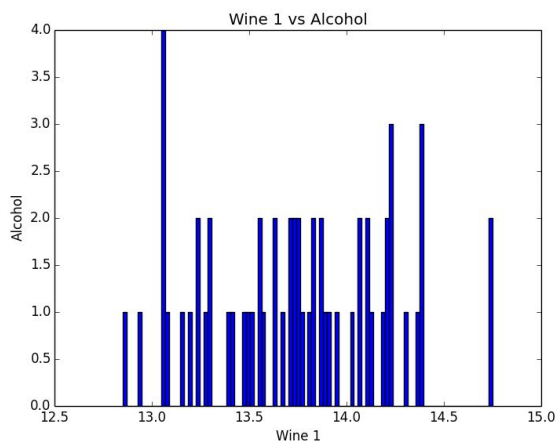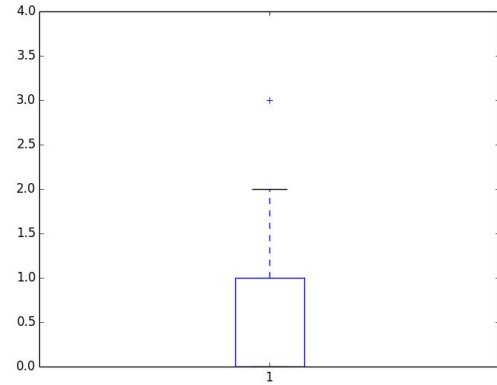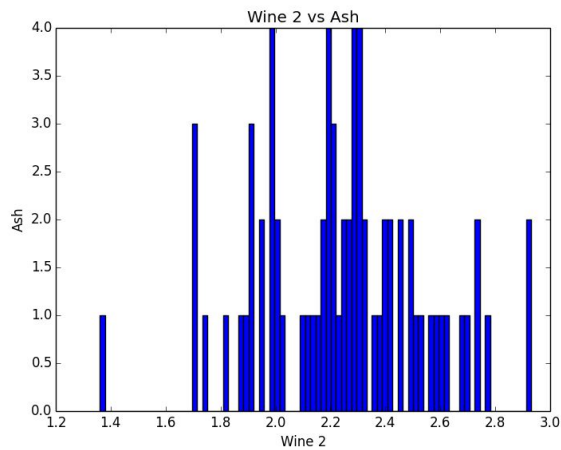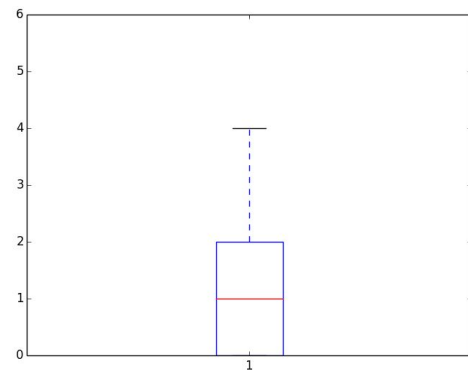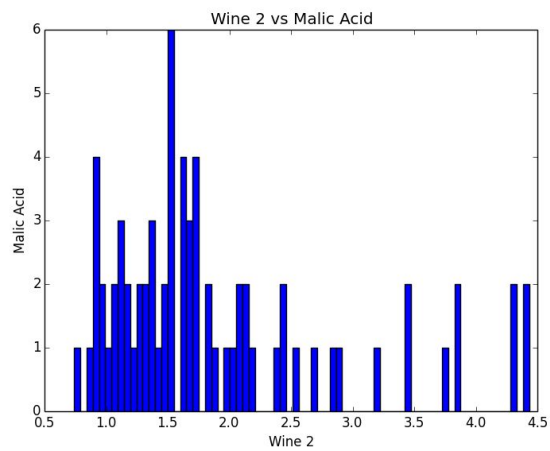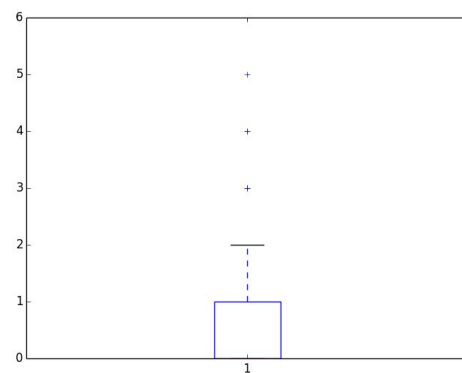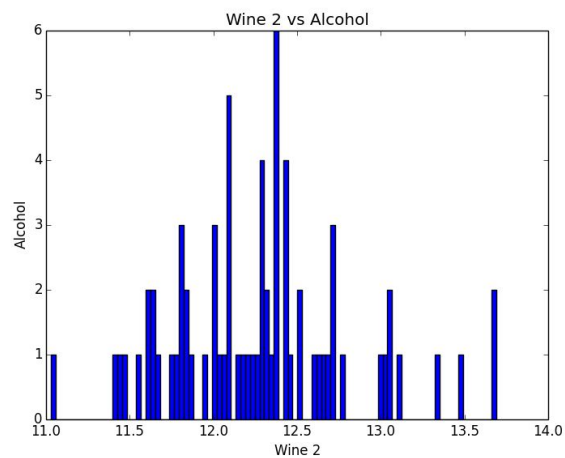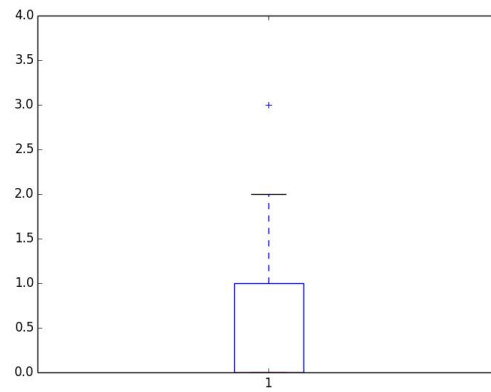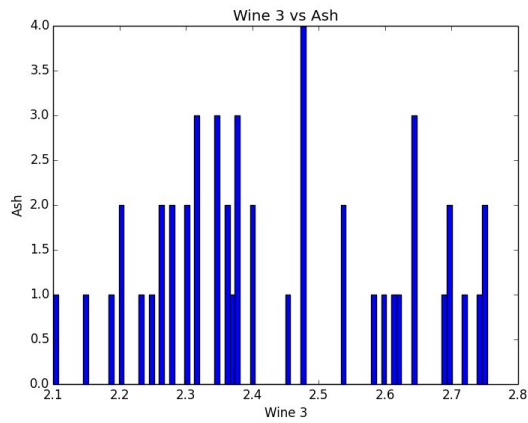
Wine 3 vs Ash

Wine3 Malic Acid Histogram and Boxplot (mostly symmetric, mostly uniform)

Wine 3 vs Malic Acid

Wine3 Alcohol Histogram and Boxplot (mostly symmetric, mostly uniform)

Wine 3 vs Alcohol
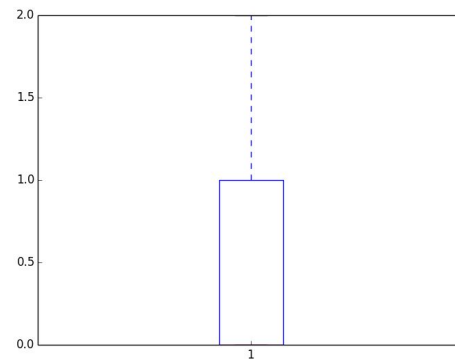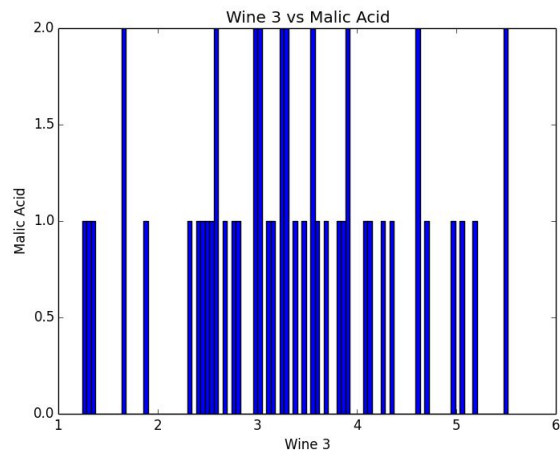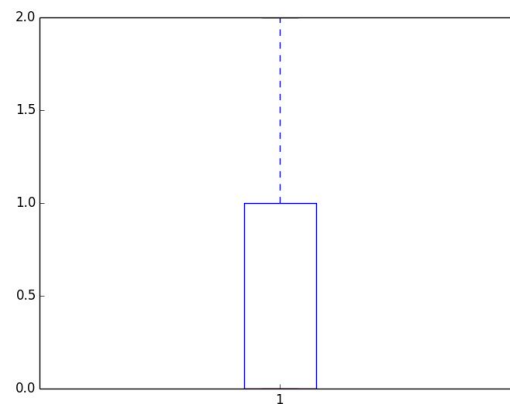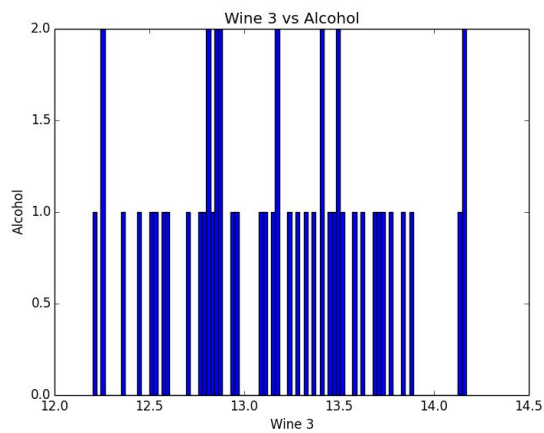
Question 2: Relations between features and datapoints
        #ASSUMPTION COMBINING QUESTION 2 PART 1 B AND C. C is answered in line
1. Correlation Plots:
    a. Pearson Correlation Coefficient - done
    b. Feature-by-Feature Correlation Matrix
        i.   Iris Heat Map
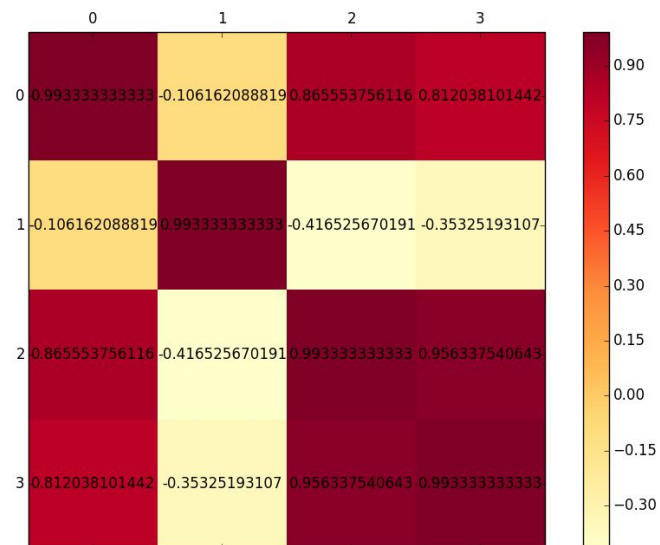
Minimum number of calls to correlation to fill matrix: 6

ii.   Wine Heat Map

Minimum number of calls to correlation to fill matrix: 78

c. Correlated Features:
   i.   The correlated features are the ones that have a darker red color.
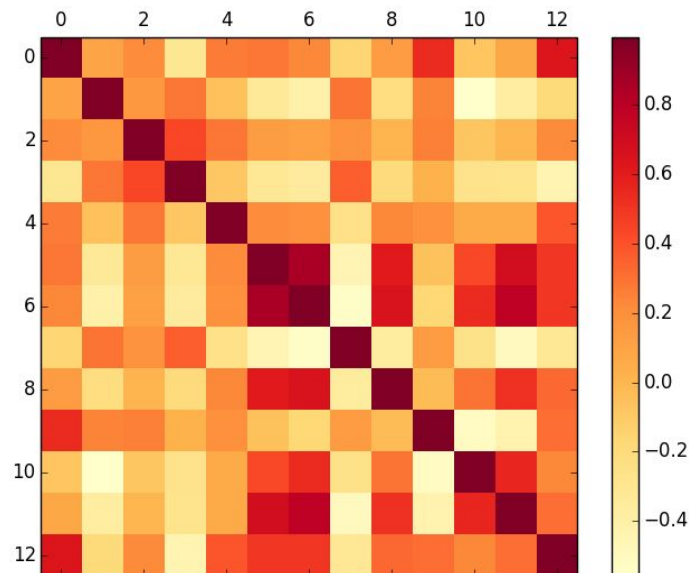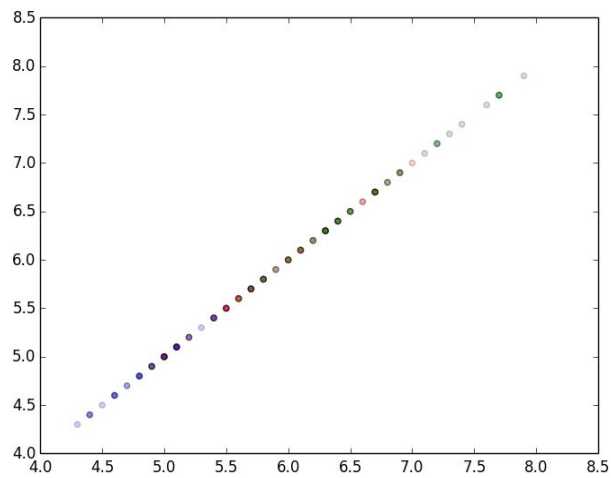   This makes sense because it is being compared to itself, or to a
         number that is next to its value
   ii.   This information can be useful because it will tell you if the data is
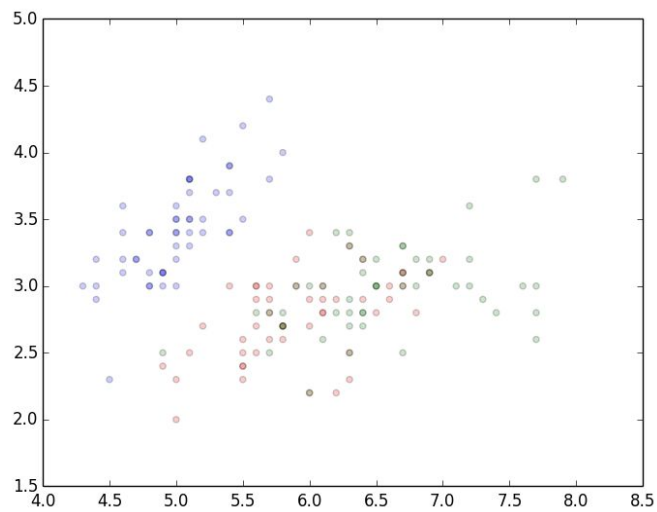   discriminative or no.

2. ScatterPlots:
   JUDGING WHETHER IT IS DISCRIMINATIVE OR NOT IN DESCRIPTION
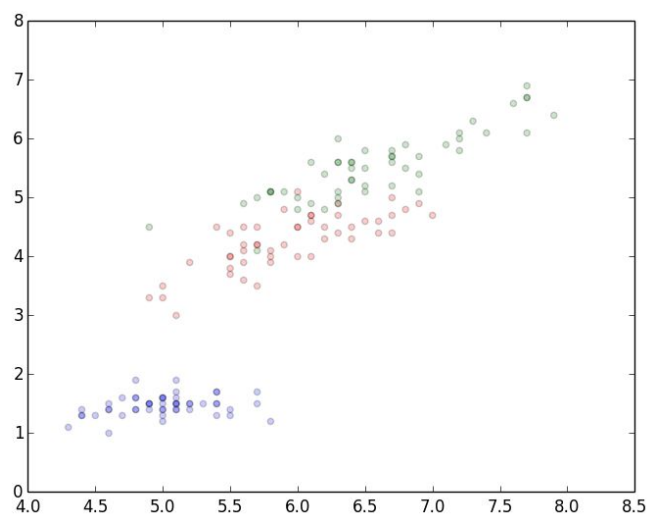   **Part c is answered at the end**

Sepal Length vs Sepal Length (not discriminative)
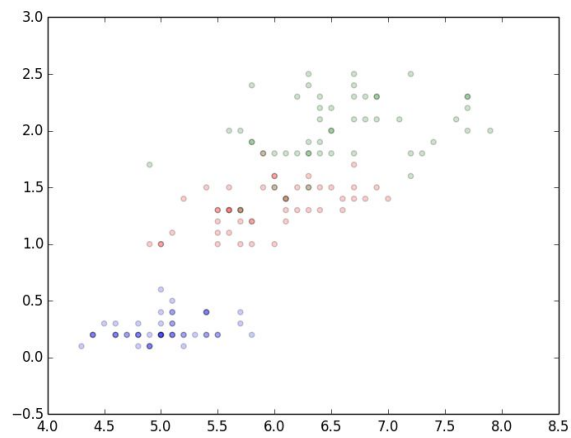

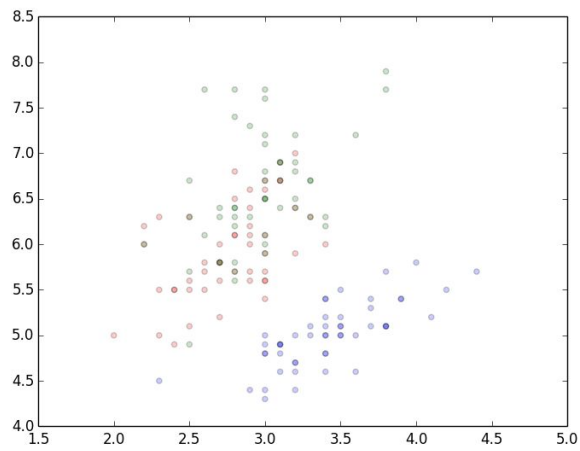## Sepal Length vs Sepal Width (not discriminative)

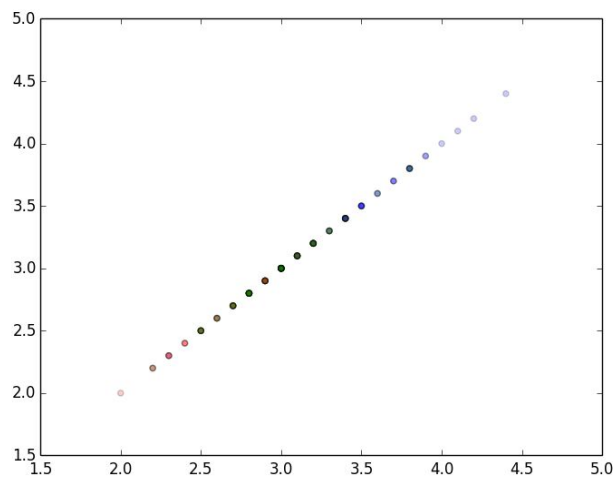## Sepal Length vs Petal Length (discriminative)
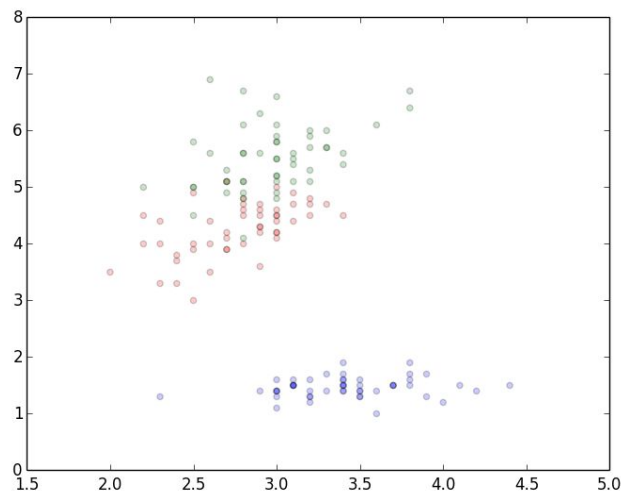
Sepal Length vs Petal Width (discriminative)

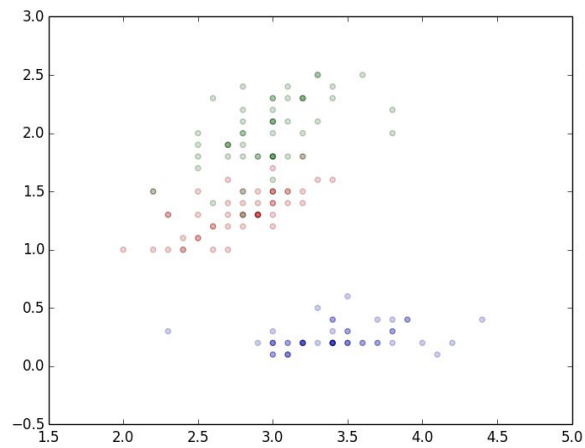## Sepal Width vs Sepal Length (not discriminative)

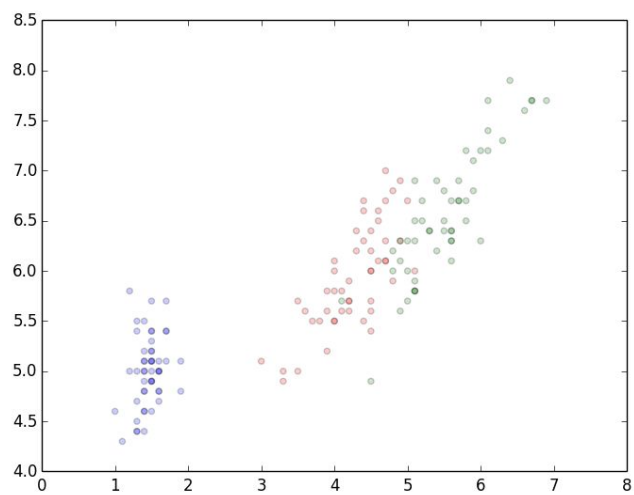## Sepal Width vs Sepal Width (not discriminative)

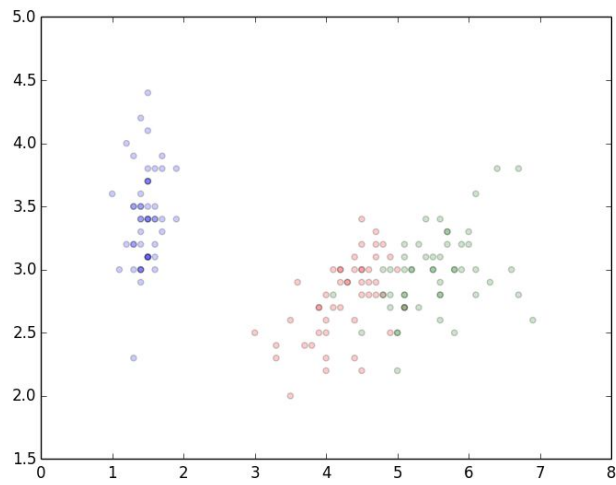## Sepal Width vs Petal Length (discriminative)

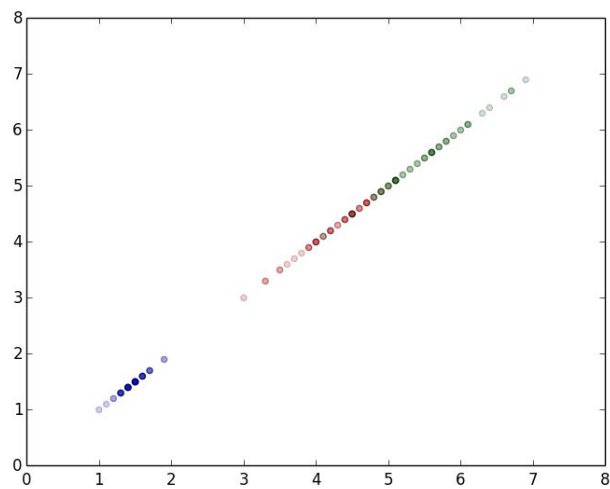## Sepal Width vs Petal Width (discriminative)

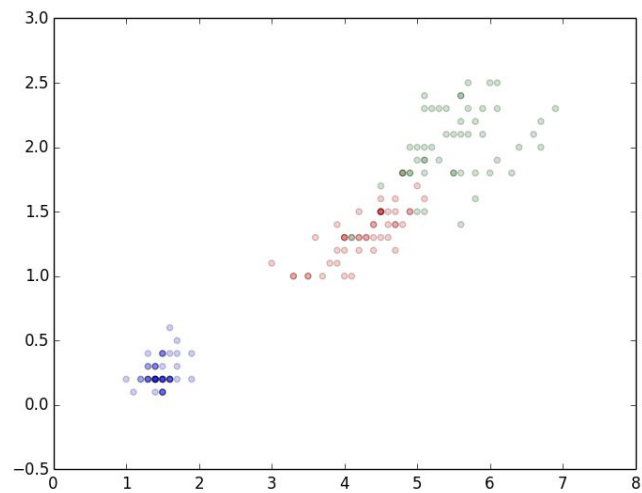## Petal Length vs Sepal Length (discriminative)

Petal Length vs Sepal Width (discriminative)
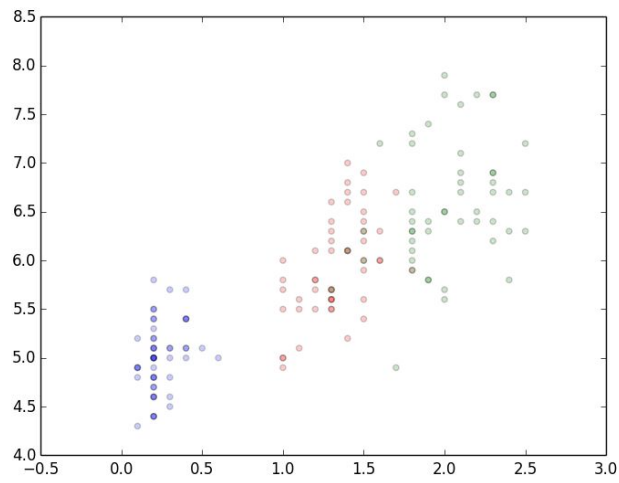

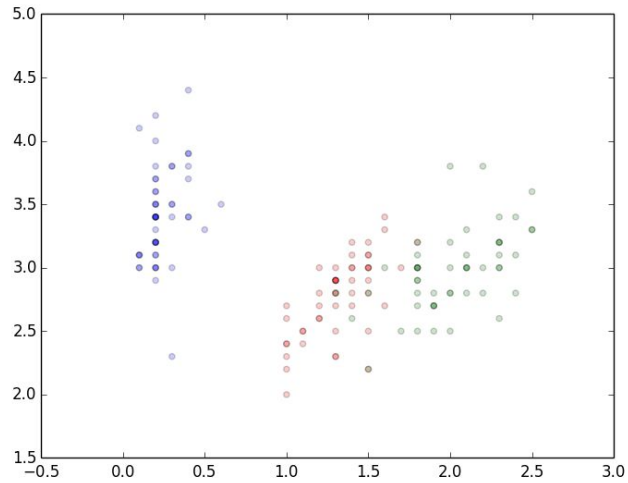## Petal Length vs Petal Length (not discriminative)

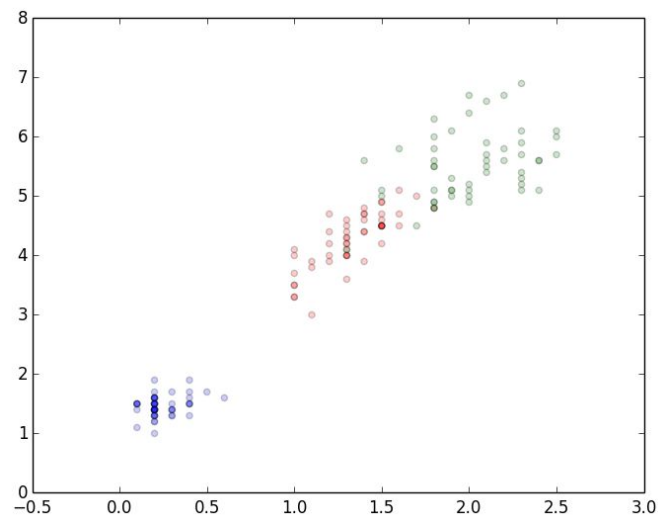
Petal Length vs Sepal Width (discriminative)

Petal Width vs Sepal Length (not discriminative)
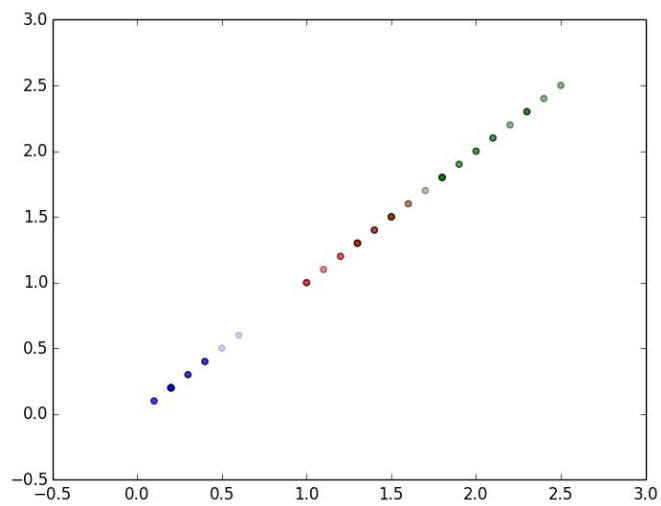

Petal Width vs Sepal Width (not discriminative)

Petal Width vs Petal Length (discriminative)

Petal Width vs Petal Width (not discriminative)

The closer the number is to 0 is not discriminative and the closer the number is to 1 it is discriminative when you look at the matrix.