

## Lecture 2 worksheet:

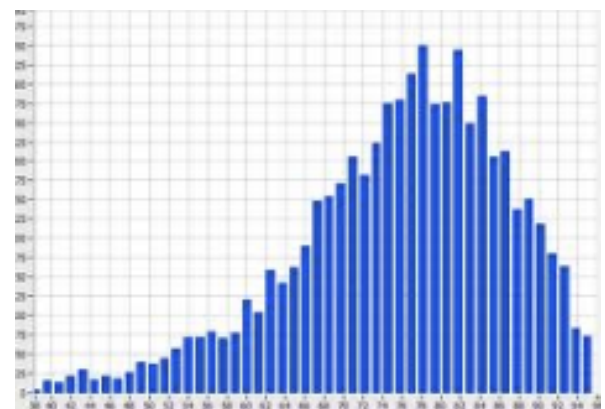
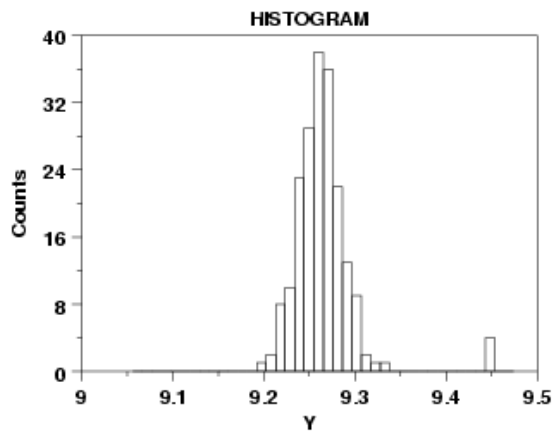
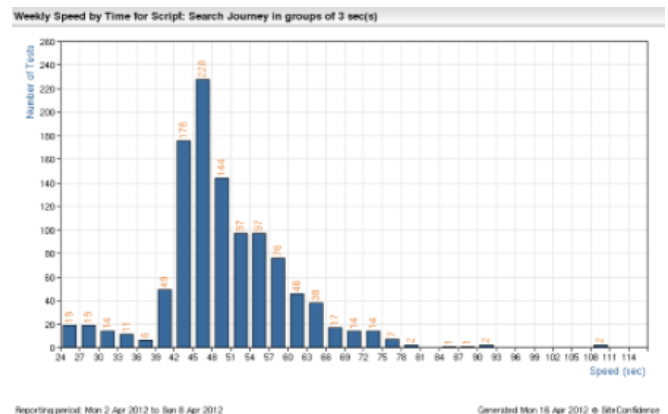
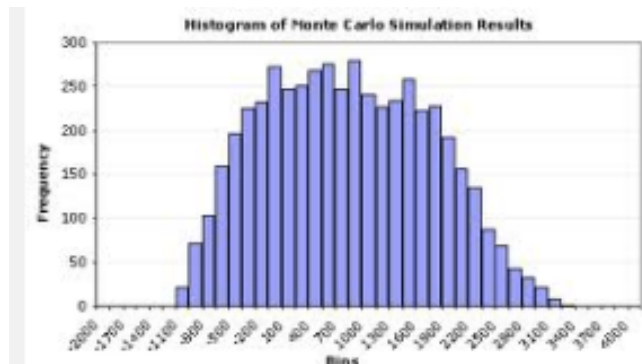
### Exercise 1:

Look at the following table describing attributes in a dataset. Identify three quality issues:

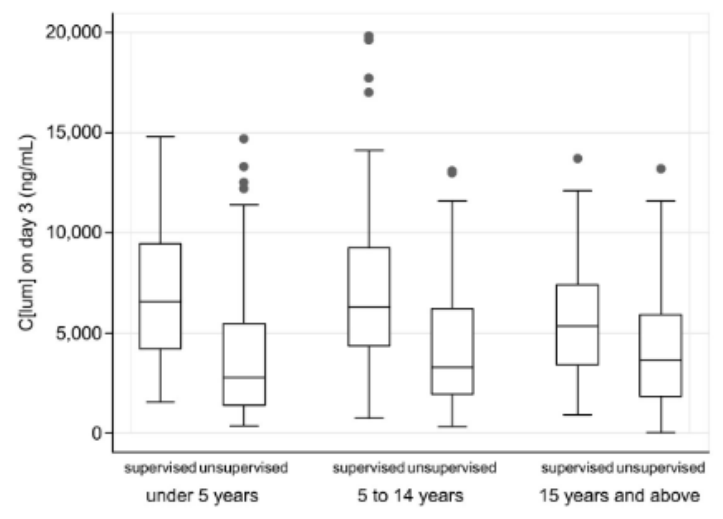
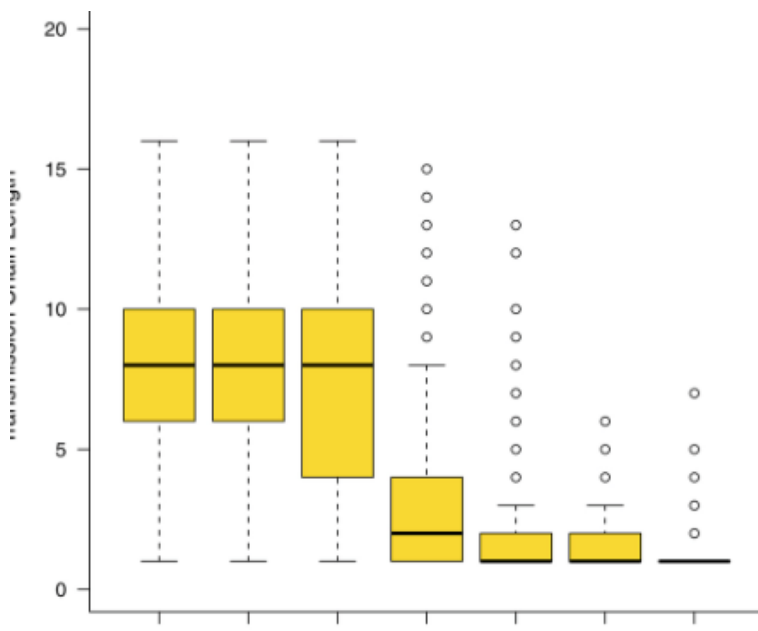
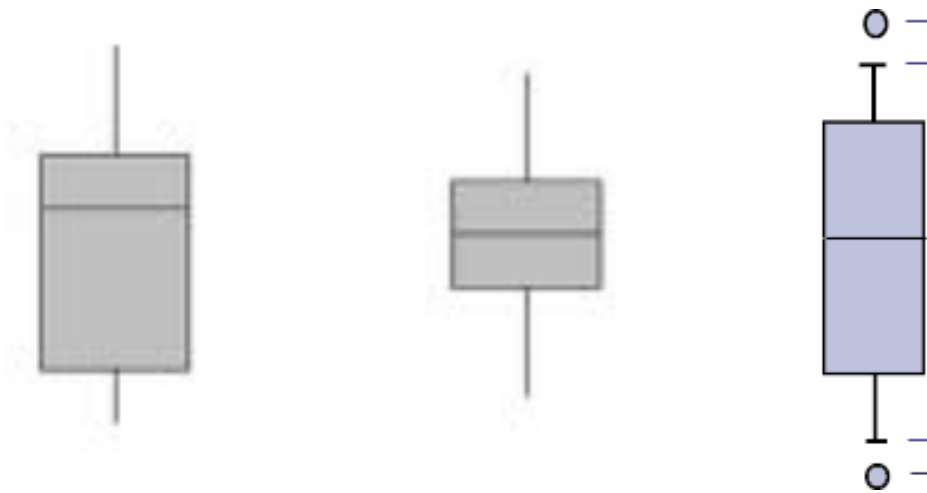
Attribute name	Data type	Mean	Standard deviation	Range	Missing values
Age	Integer	23	5	[18,60]	0
Gender	Integer	0.3	0.1	[0,1]	3
Salary	Numeric	45000	20000	[-100,500000]	0

If 0=female, and 1=male, are there more males or females in the dataset?

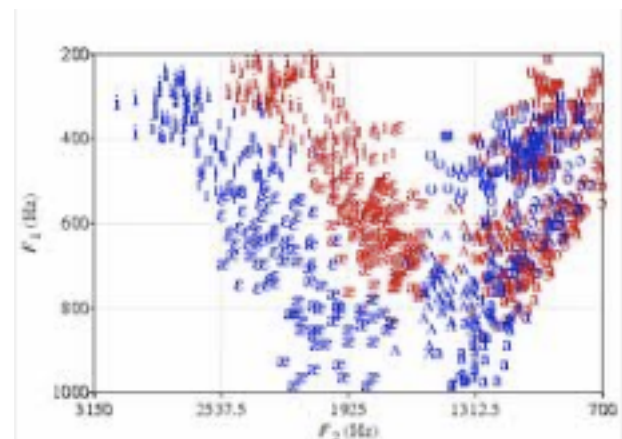
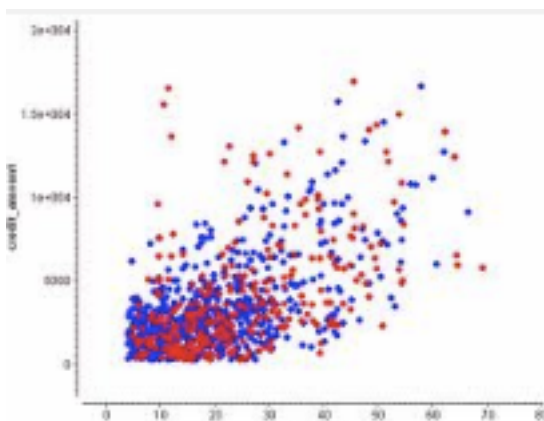
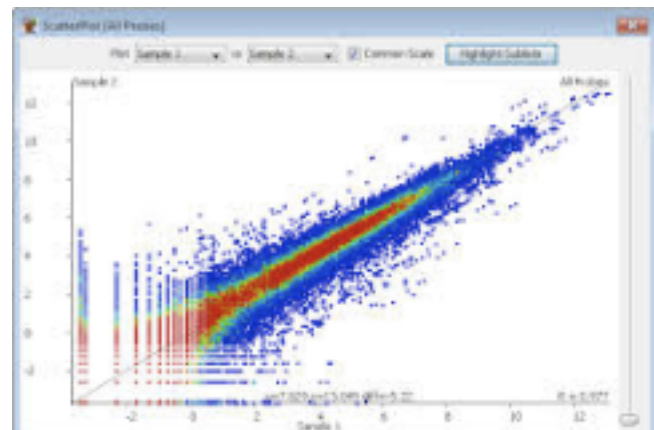
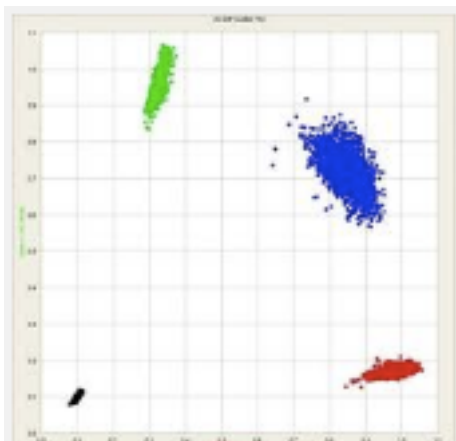
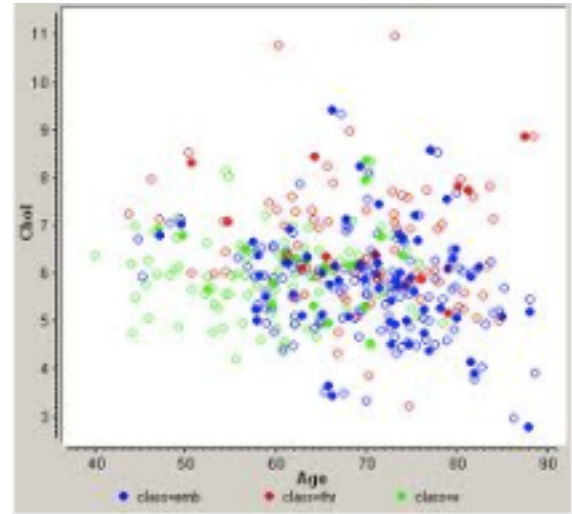
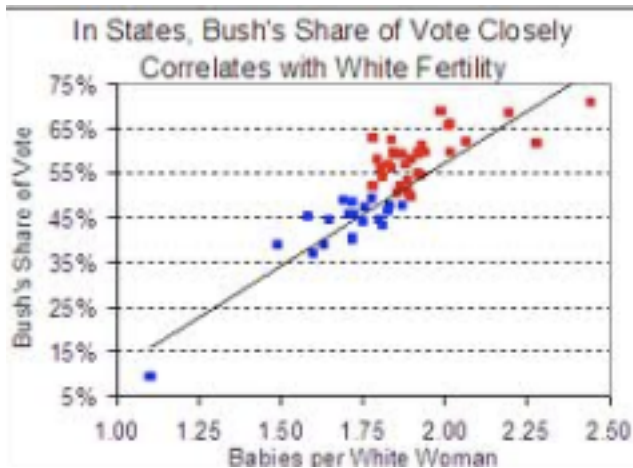
### Exercise 2: Which of the following four histograms identify outliers?



**Exercise 3:** For each of the following box plots, say if the distribution is normal or skewed. Which box plots identify outliers?



**Exercise 4:** are the attributes depicted in the scatter plots below useful in predicted the class label?



## Exercise 5:

Based on the parallel plots below, will it be easy to distinguish between the classes?

