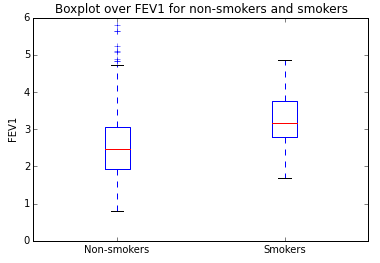
Assignment 1

## Exercise 1

b) The average lung function for smokers is 3.28 FEV1, whereas the average lung function for non-smokers is 2.57 FEV1. The result is surprising because the average non-smoker has a lower FEV1 than the average smoker, which suggests that the average smoker has better lung function than the average non-smoker.

## **Exercise 2**



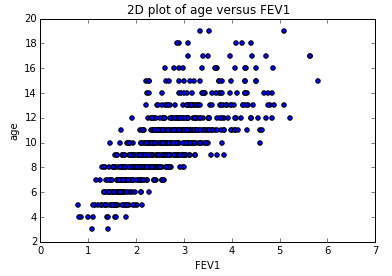
The median FEV1 for smokers is surprisingly higher than the median FEV1 for non-smokers, but the FEV1 range for non-smokers is wider than for smokers and FEV1 for non-smokers has several outliners past the maximum, which were probably not accounted for in the median.

## Exercise 3

b) The value of t-statistic, the degrees of freedom and the returned p-value was 7.15, 83, and 3.07×10-10 respectively. The hypothesis was rejected, but the two samples do not have the same mean as assumed in the null hypothesis, so the result of the hypothesis test does provide a clear conclusion about the difference between FEV1 in the smokers and non-smokers group.

## Exercise 4

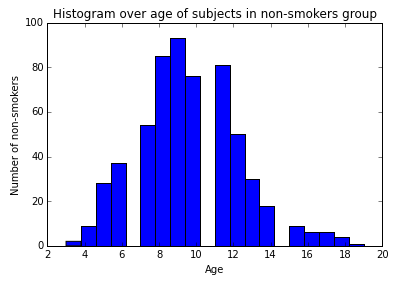
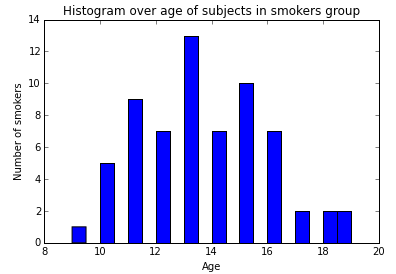
a)



The 2D plot of age versus FEV1 shows that there is a positive linear relationship between them.

c) The correlation between age and FEV1 is 0.756, suggesting that there is a strong positive correlation between them.

## Exercise 5



The two histograms show that there are quite significant differences in the size and the age of the subjects in the smokers and non-smokers groups. The distribution of the age of subjects in the non-smokers group is right-skewed compared to that of the smokers group. The mode age of subjects in the smokers group is between 13 and 14 years old, whereas the mode age in the non-smokers group is between 9 and 10 years old. Since the correlation between age and FEV1 is strongly positive, this could explain the lower mean FEV1 for non-smokers compared to the mean FEV1 for smokers.