**Question 1**

How many particles labeled "synecho" are in the file provided?

Answer for Question 1:



**Question 2**

What is the 3rd Quantile of the field fsc\_small? (the summary function computes this on your behalf)

Answer for Question 2



**Question 3**

What is the mean of the variable "time" for your training set?

Answer for Question 3



**Question 4**

In the plot of pe vs. chl\_small, the particles labeled ultra should appear to be somewhat "mixed" with two other populations of particles. Which two populations?

pico

crypto

nano

synecho

**Question 5**

Use print(model) to inspect your tree. Which populations, if any, is your tree incapable of recognizing? (Which populations do not appear on any branch?) (It's possible, but very unlikely, that an incorrect answer to this question is the result of improbable sampling.) Hint: Look

crypto

synecho

nano

ultra

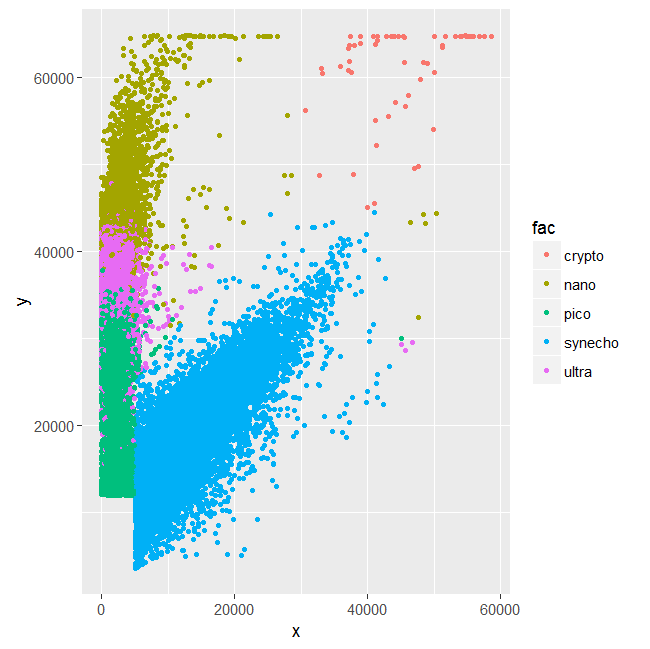
pico

**Question 6**

Most trees will include a node near the root that applies a rule to the pe field, where particles with a value less than some threshold will descend down one branch, and particles with a value greater than some threshold will descend down a different branch. If you look at the plot you created previously, you can verify that the threshold used in the tree is evident visually. What is the value of the threshold on the pe field learned in your model?

Answer for Question 6





**Question 7**

Based on your decision tree, which variables appear to be most important in predicting the class population?

fsc\_perp

pe

chl\_small

fsc\_big

fsc\_small

chl\_big

**Question 8**

How accurate was your decision tree on the test data? Enter a number between 0 and 1.

Answer for Question 8



**Question 9**

What was the accuracy of your random forest model on the test data? Enter a number between 0 and 1.

Answer for Question 9



**Question 10**

After calling importance(model), you should be able to determine which variables appear to be most important in terms of the gini impurity measure. Which ones are they?

fsc\_perp

chl\_big

fsc\_small

fsc\_big

pe

chl\_small

**Question 11**

What was the accuracy of your support vector machine model on the test data? Enter a number between 0 and 1.

Answer for Question 11



**Question 12**

Construct a confusion matrix for each of the three methods using the table function. What appears to be the most common error the models make?

synecho is mistaken for pico

ultra is mistaken for nano

ultra is mistaken for pico

nano is mistaken for ultra

crypto is mistaken for ultra

pico is mistaken for ultra

**Question 13**

The variables in the dataset were assumed to be continuous, but one of them takes on only a few discrete values, suggesting a problem. Which variable exhibits this problem?

fsc\_big

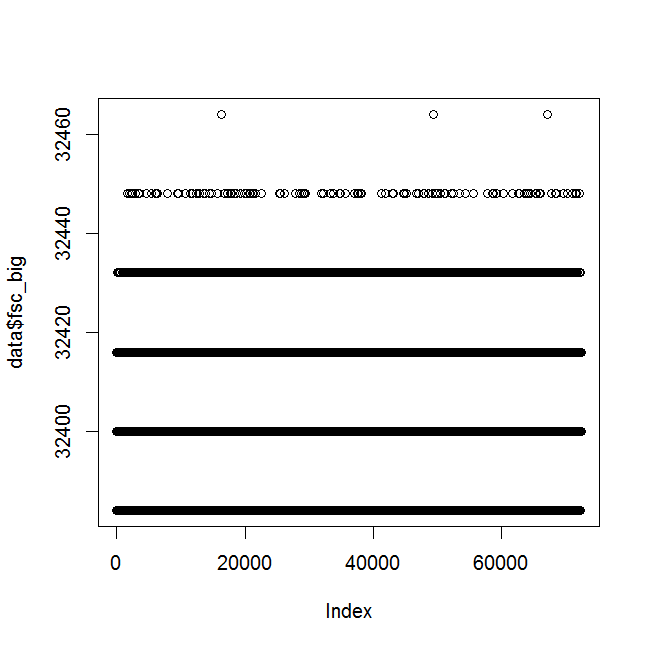
pe

chl\_big

chl\_small

fsc\_small

fsc\_perp



**Question 14**

After removing data associated with file\_id 208, what was the effect on the accuracy of your svm model? Enter a positive or negative number representing the net change in accuracy, where a positive number represents an improvement in accuracy and a negative number represents a decrease in accuracy.

SVM: accuracy increased by 0.05 to 0.97 (from 0.92).