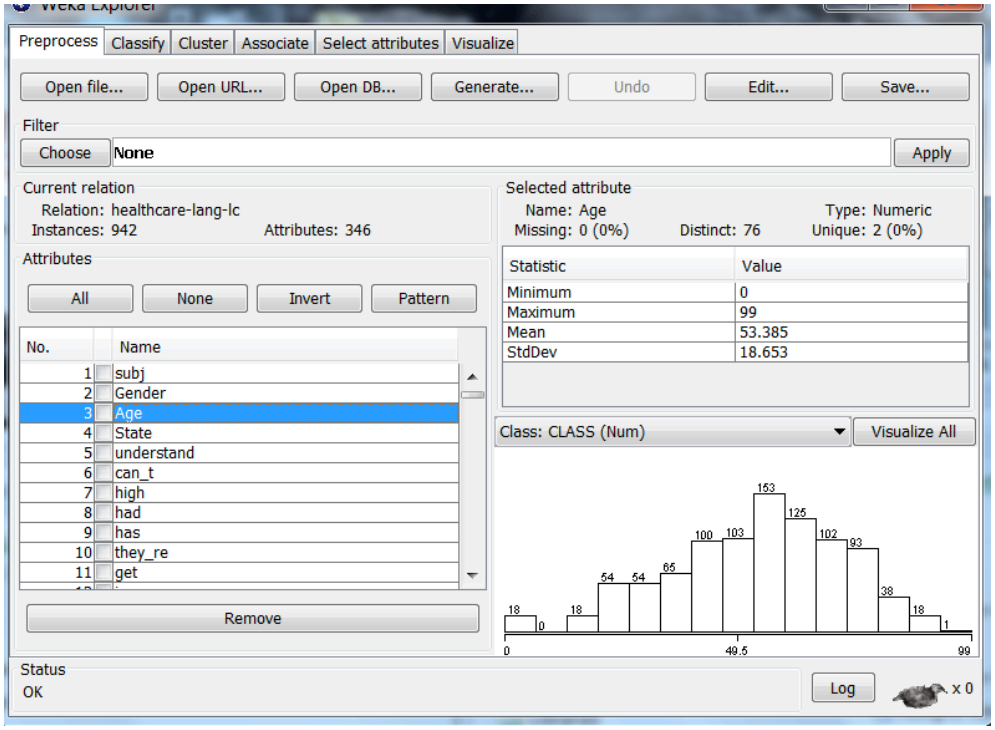
Assignment I

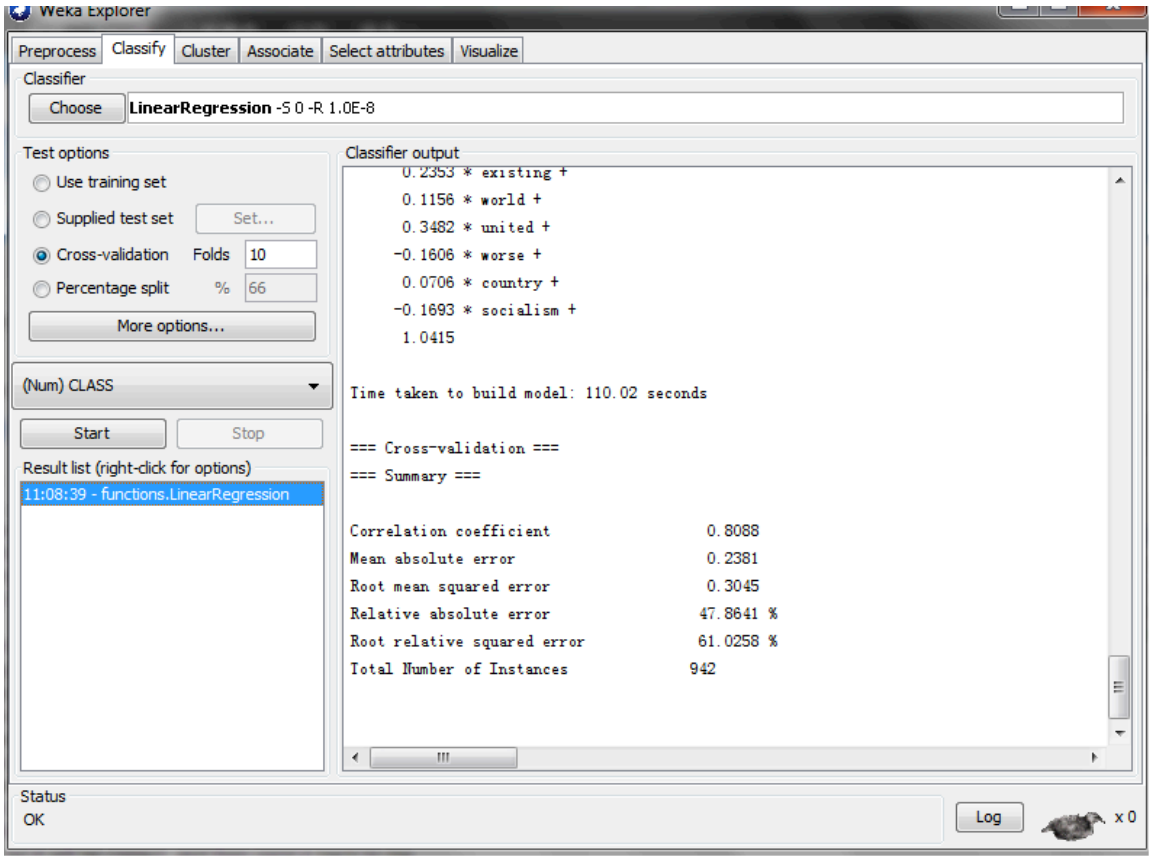
- Save the file as a ".arff" file which is standard input format of Weka.
    a. See example in assign1.arff

- Find out from the GUI interface what the average age of people participating in the poll was.



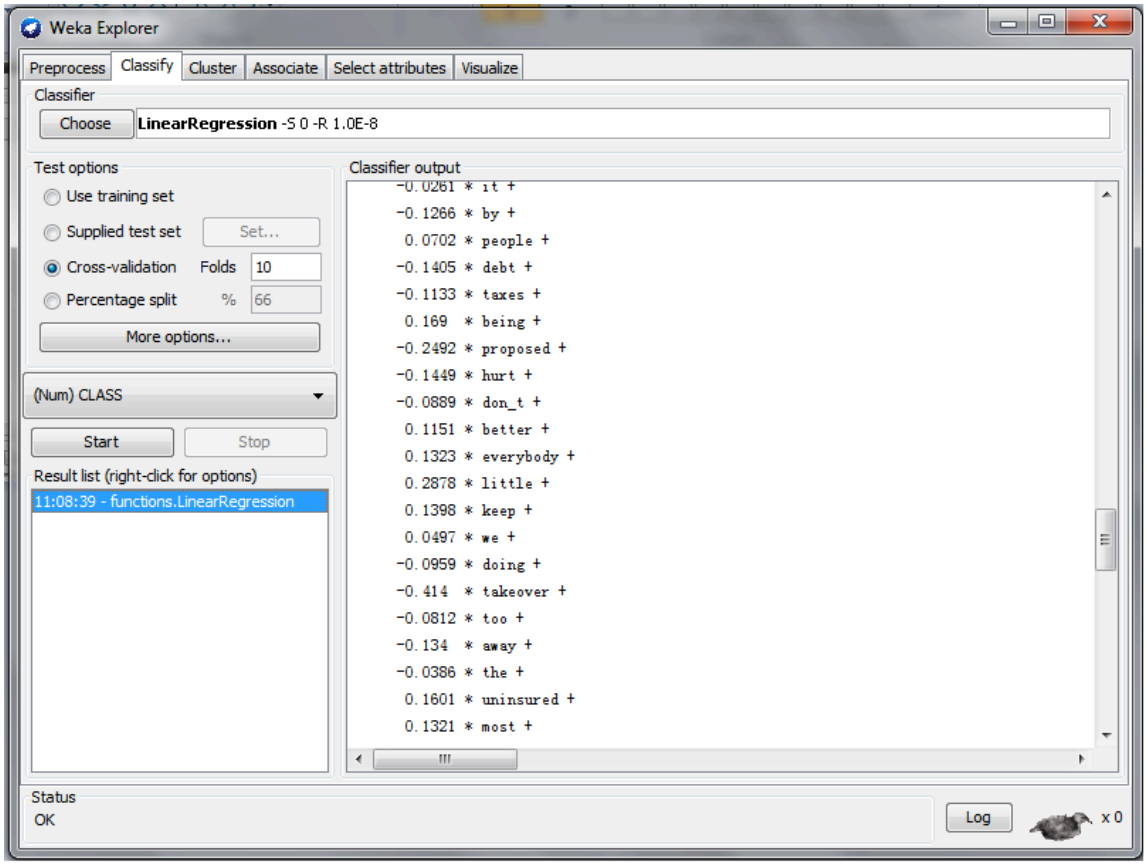
The answer is 53.

- In class you saw how to evaluate the accuracy of a classification model. Here we are evaluating the effectiveness of a regression model. We do this by looking at the Correlation Coefficient, where numbers close to 1 are good, and numbers close to 0 ore poor. What do you get for the Correlation Coefficient when you run the experiment?



Correlation coefficient: 0.8088, which is good.

- And which features were most predictive of voting positively? Which features were most predictive of voting negatively? Did any of this surprise you?



To find the most predictive features, use your favorite post-processing tool (like Excel) to sort the feature weight column from low to high. The most positive and negative weights are associated with the features most predictive of voting positively and negatively, respectively.
Features predictive of voting positively: positive: UNITED, SOCIAL, PRE
Features predictive of voting negatively: negative: STATES, FREE, TAKES, NATIONAL, PROPOSED
Surprises? United and States had opposite weights, which is strange considering that they usually appear together. I would have also expected “social” to be a big concern and that the weight for social could go either way.

- Save your result buffer.

See Result.txt