**Problem 3**

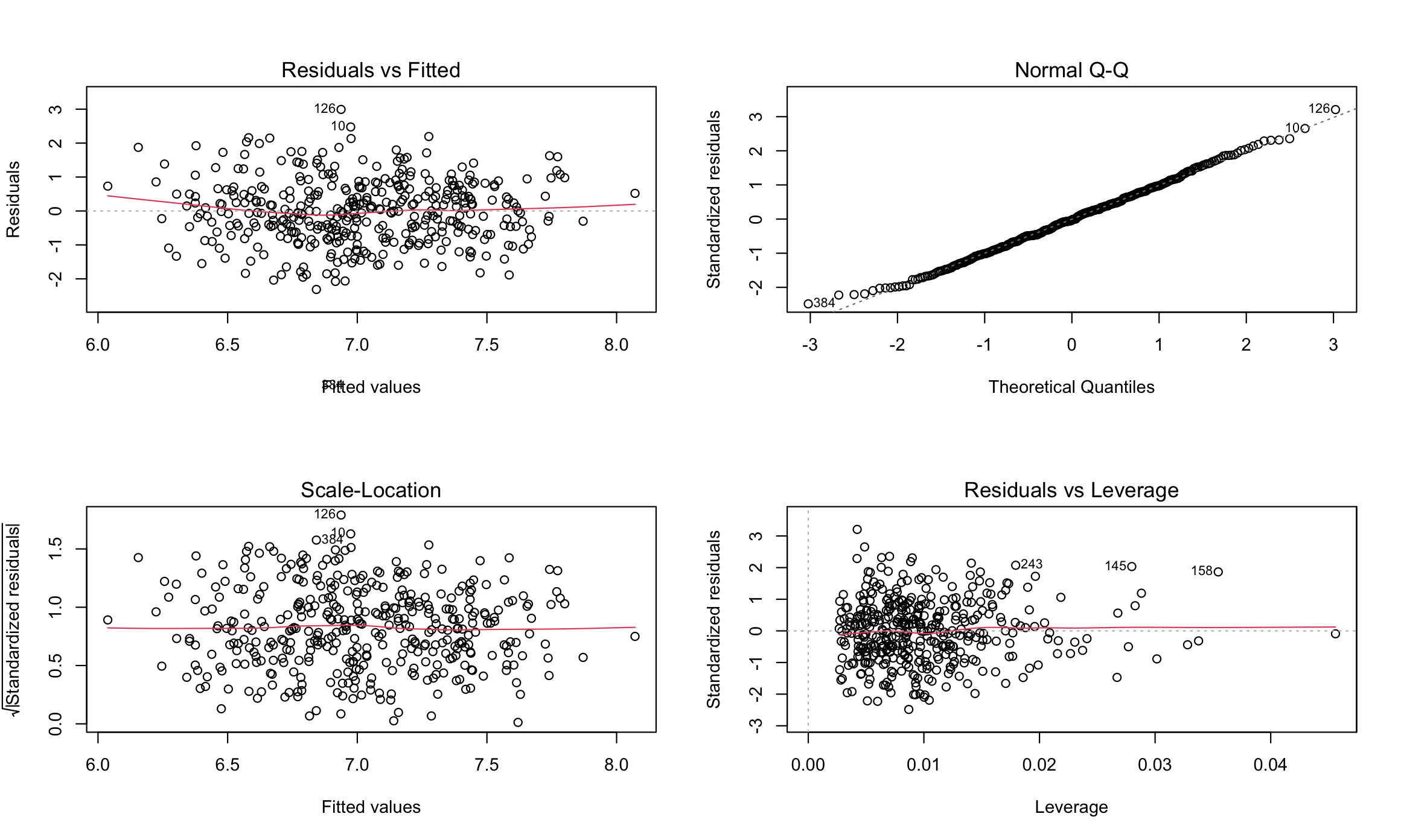
We are going to study the danceability of songs using the designed linear model.

Fitting the model on our training set we obtain the following parameters:

Intercept loudness energy tempo

9.18676786 0.09930218 0.07258300 -0.00888967

And sigma is 0.9314841



We can trust the model since by looking at the residual we see no patterns, they seem to be gaussian by the qqplot and there are no leverage points. The Shapiro test on the residuals gives a p-values of 2.112e-06 so we confirm the gaussian intuition on the residuals.

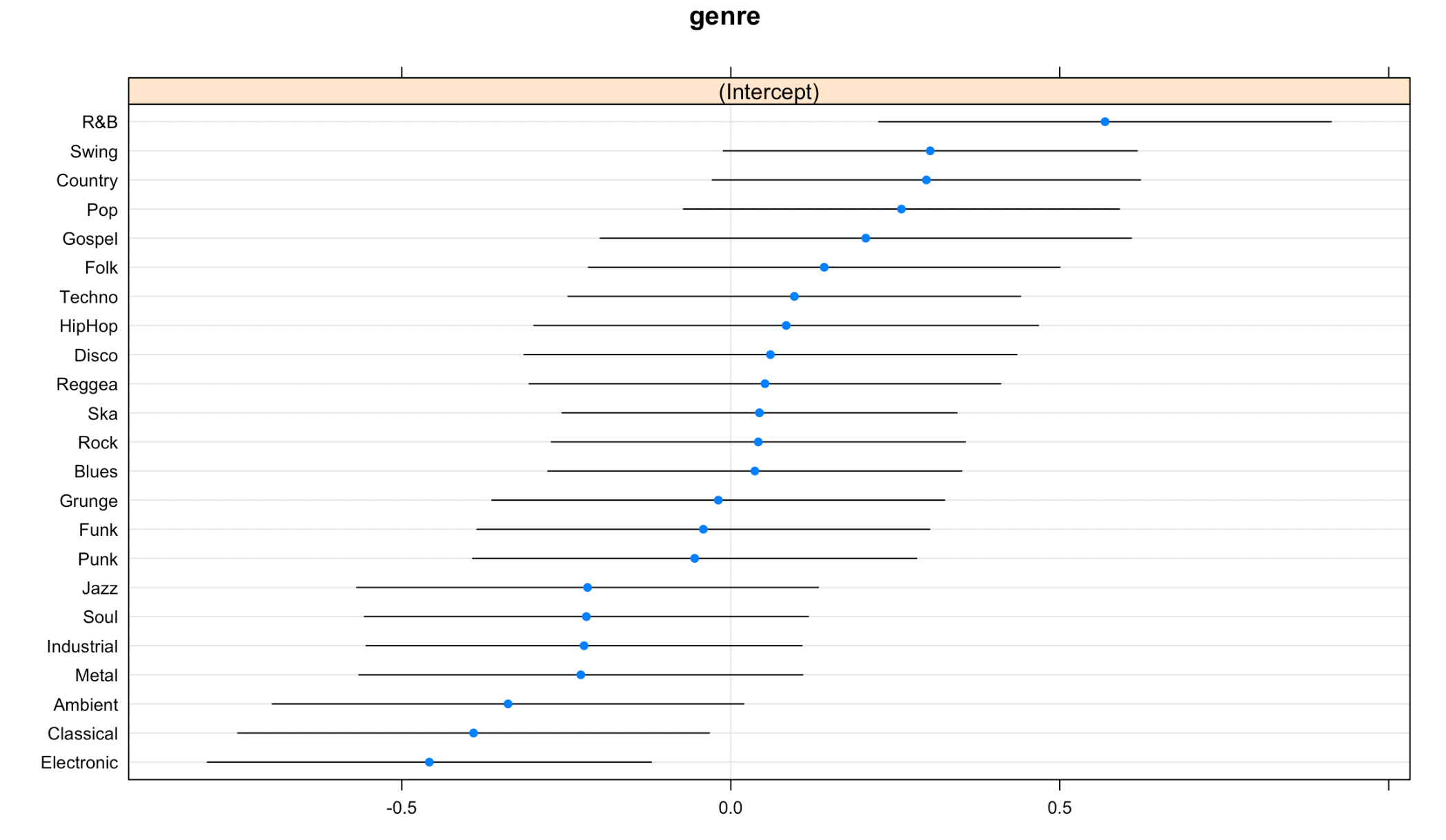
The VIF values are: loudness 2.282505, energy 2.279888 and tempo 1.001943 so we can say that there is not much collinearity between variables.

We perform statistical tests using Linear Hypothesis to see if energy and loudeness are significant. The p-value of the test is 2.112e-06 so we reject the null hypothesis (at 5%) of null coefficient and we keep them in the model.

We perform the same test for the single elements but we p-values of 0.0603 for loudness and 0.07311 for energy so we keep both at 5% in our model. The other variables got even better significance so we stay with the complete model.

We now create a model with mixed effects in order to consider the genre of the song. It will influence on the intercept of the model. Fitting the model we find that the PVRE is 0.1046363.

The resulting dot plot is the following:



And the most danceable genre is R&B