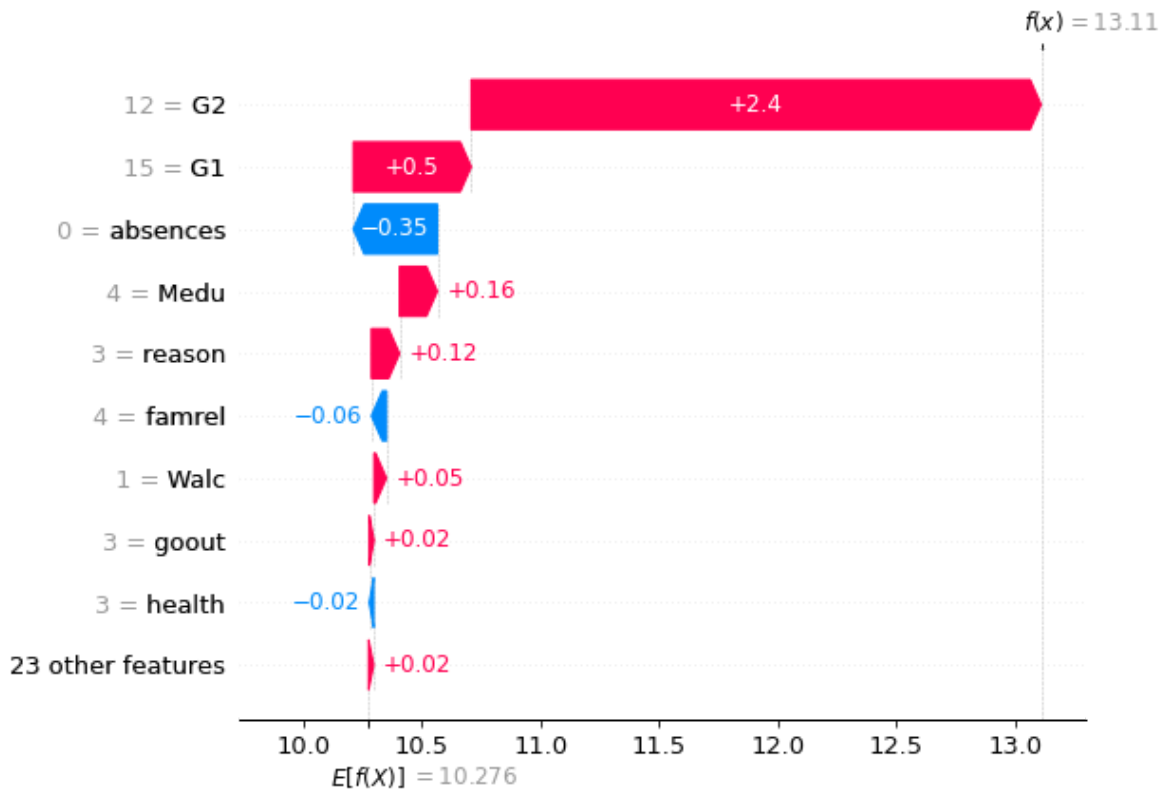
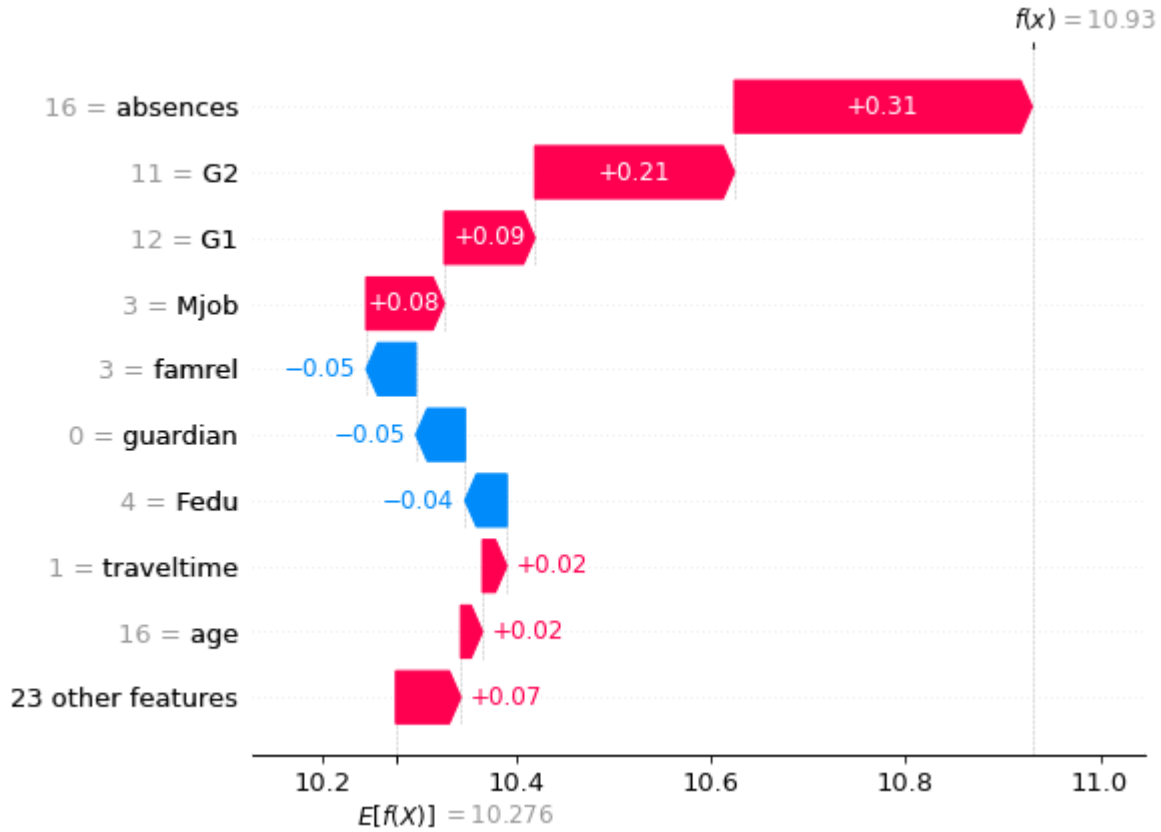
4. Find two observations in the data set, such that they have different variables of the highest importance



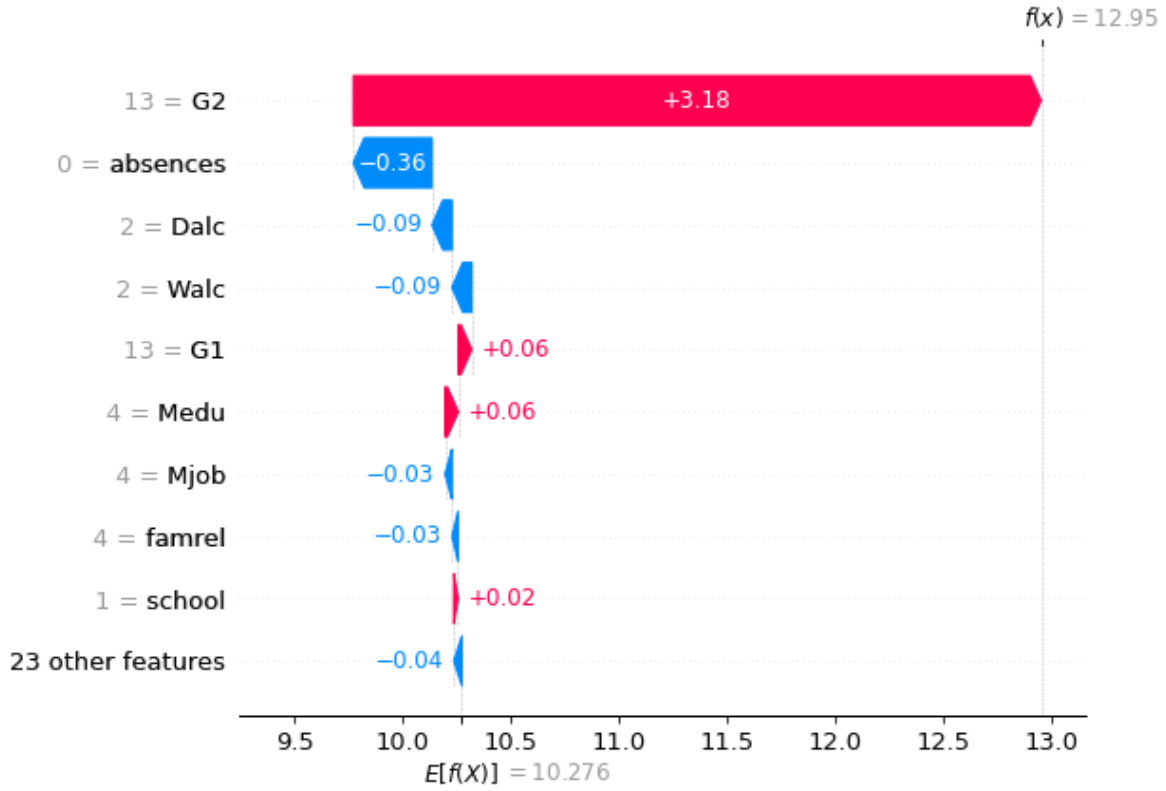
Random forest observation 1



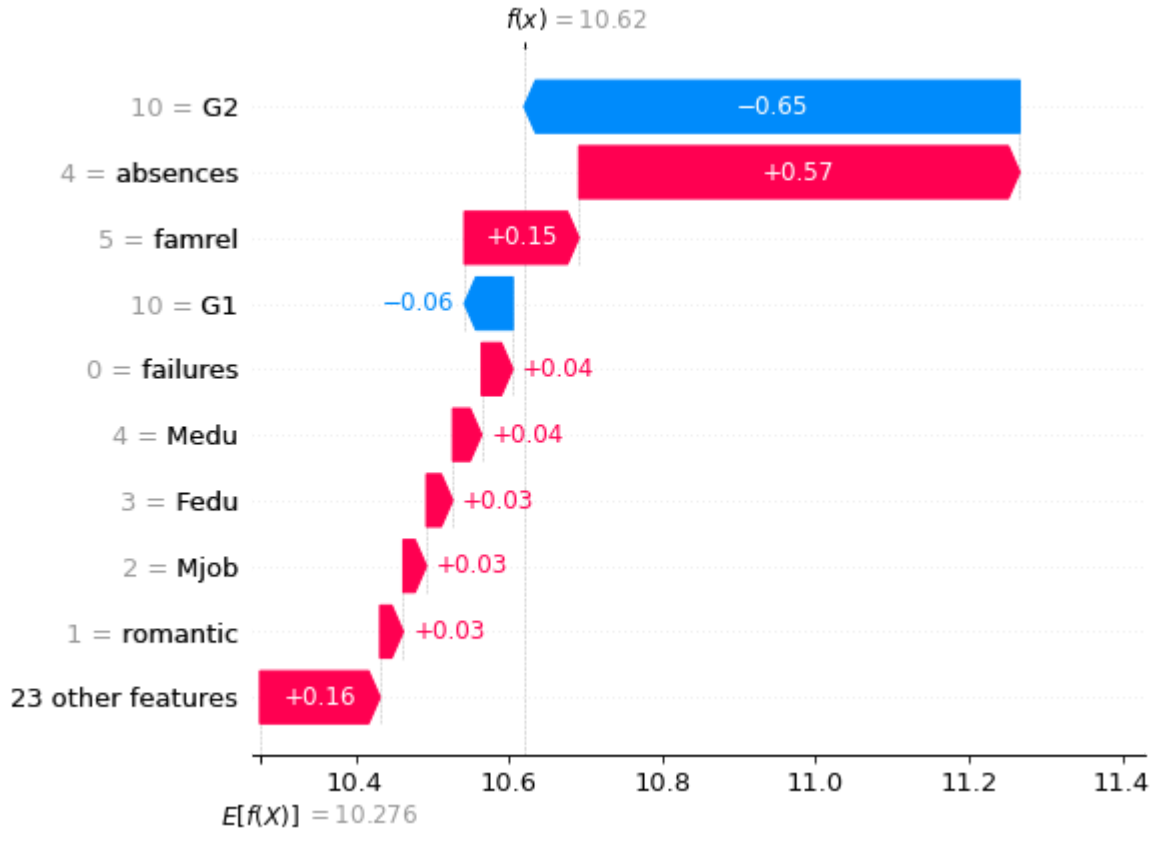
Random forest observation 14

Between my observations, the first's most important features are G2 and G1, which are expected according to the data description. In case of the second observation, the first most important feature are absences, which suggests that the student had so many of them (16) that even previous grades could not save him.

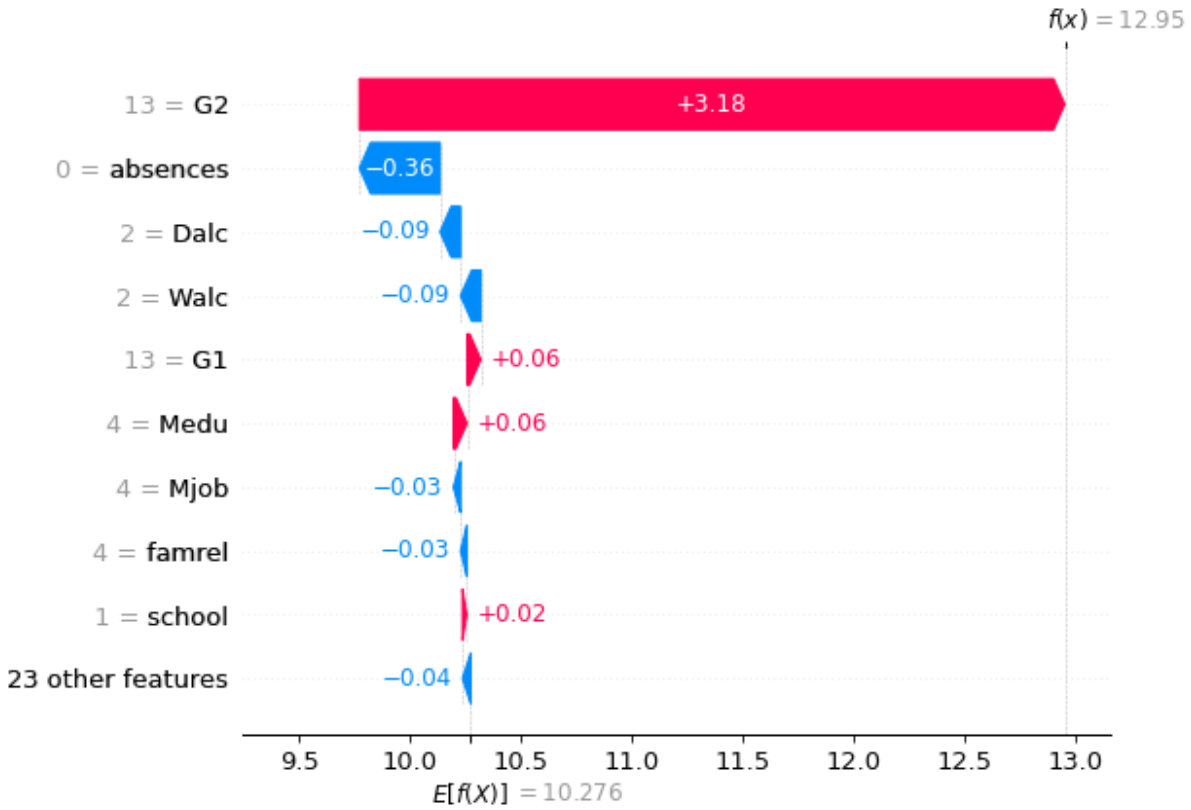
5. Select one variable and find two observations in the data set such that for one observation this variable has a positive effect and for the other a negative effect.

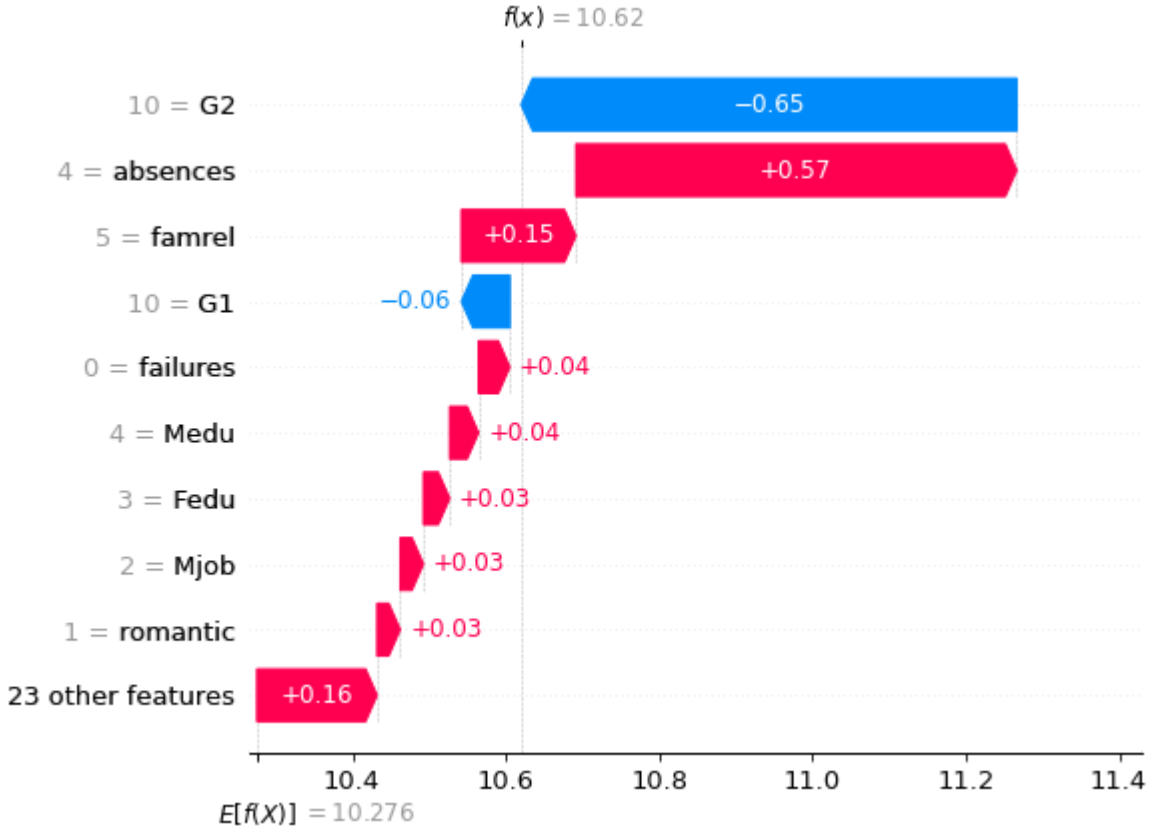


Random forest observation 3



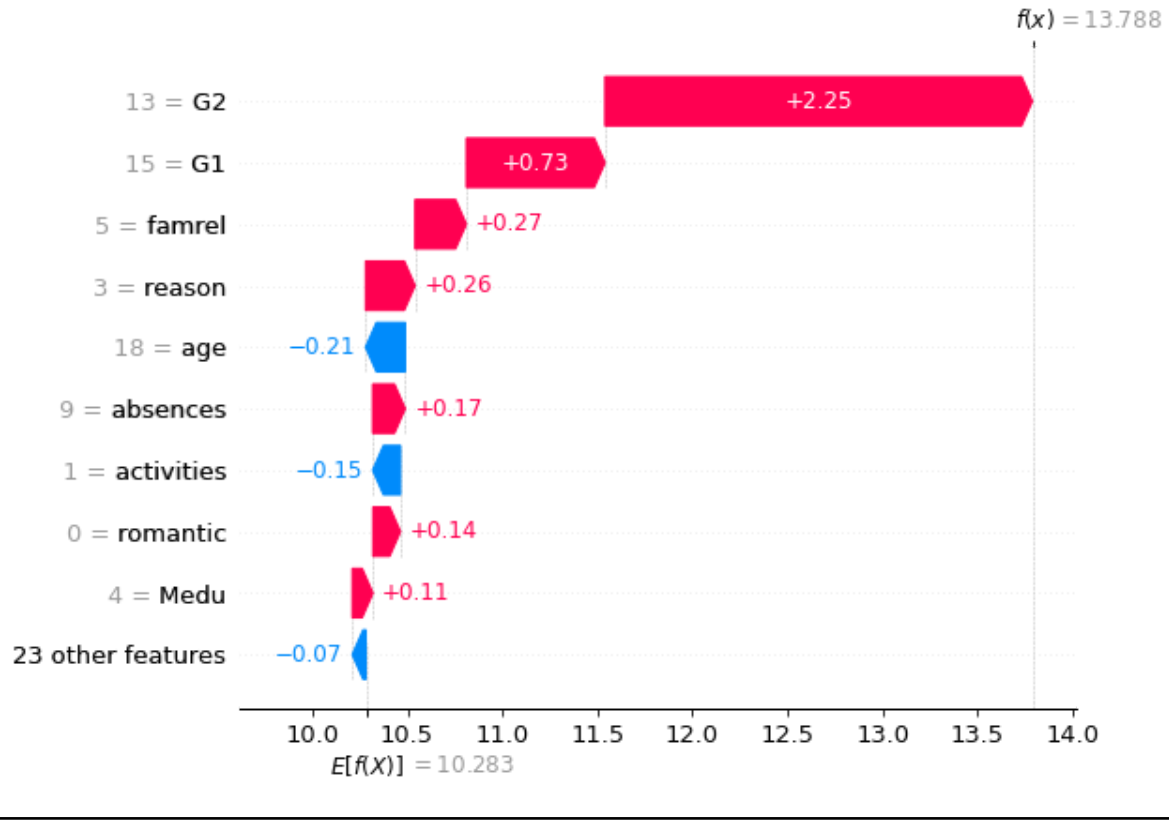
Random forest observation 33



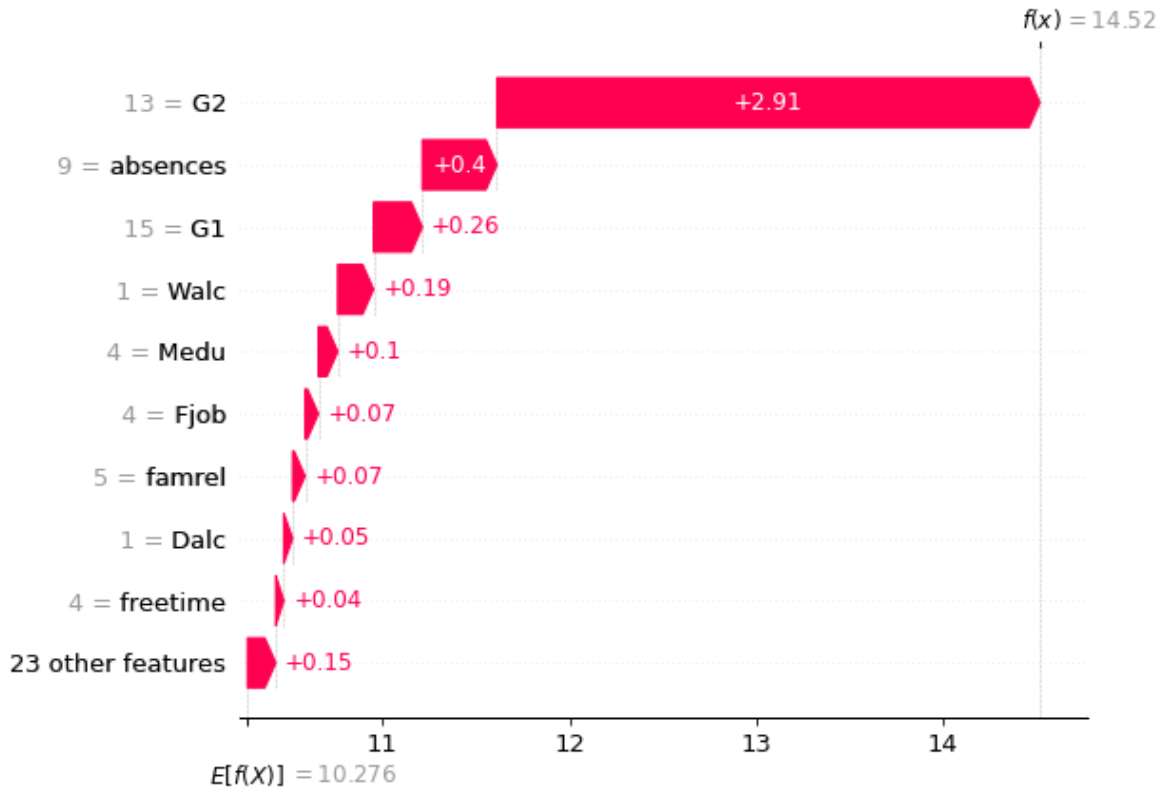


For observation 3 G2 was the most important predictor since it was consistent with the final grade. For observation 3 however, the case was not that simple, because the student had more absences, and the final grade was not the same as G2.

6. Train a second model and find an observation for which BD/shap attributions are different between the models.



Linear regression observation 11



Random forest observation 11

In observation 11, the explanation method showed that both models found variable G2 as important, however, for linear regression the absences are a lot less important than for the random forest, which placed absences as the second most important feature, and valued G1 more.