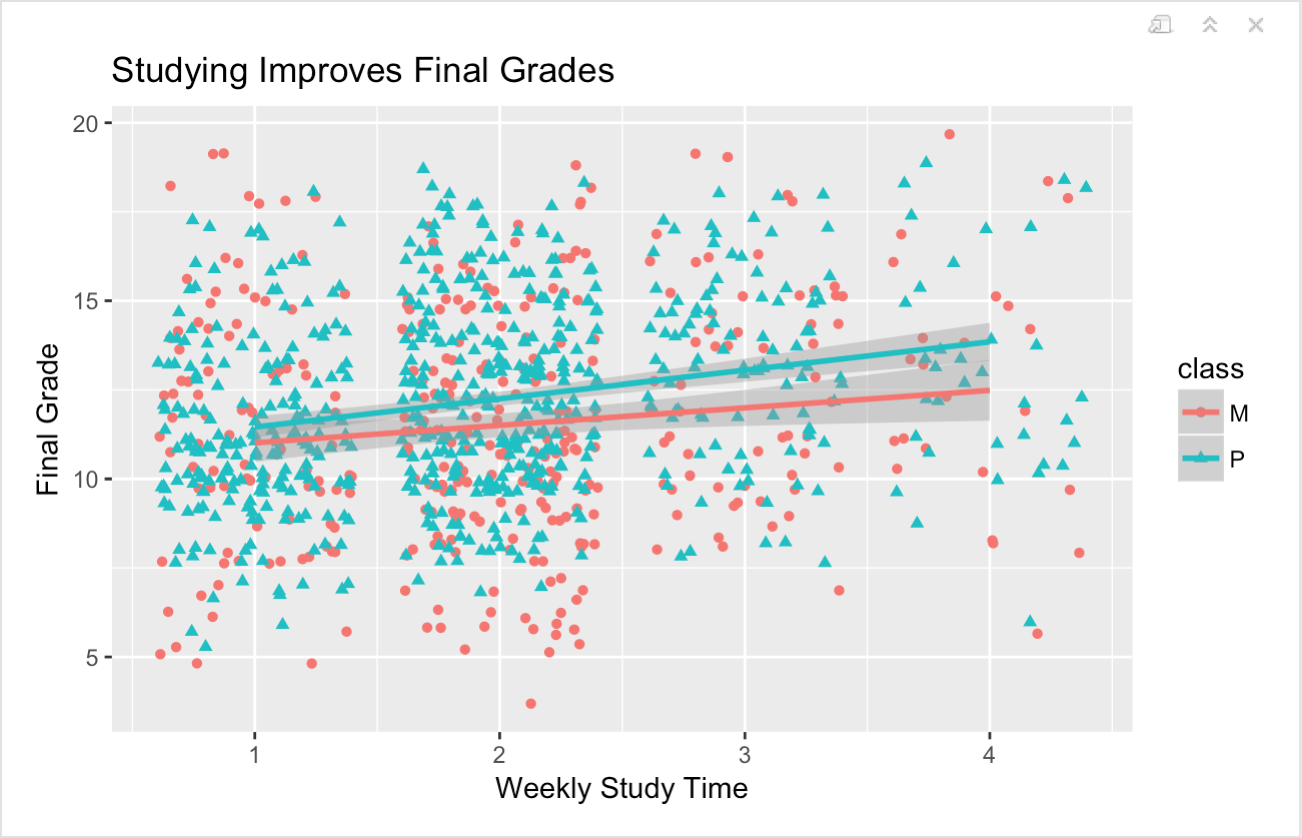
Matthew Conway, Trent Hollandsworth, Valerie Davis

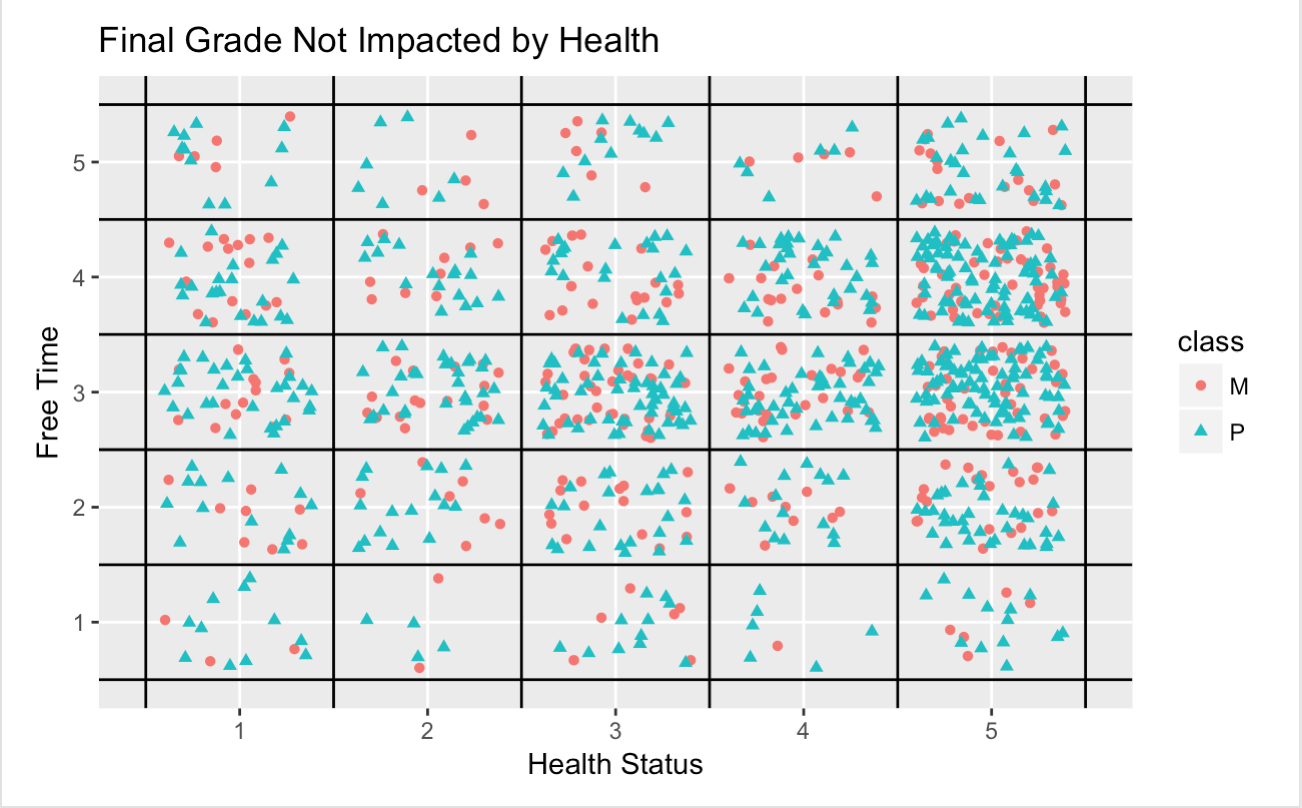
STOR 390 Final Project

Summary

The two data sets, one from a Math class and the other from a Portuguese Class, are imported and merged. Observations with a 0 as the final grade were dropped, as the other variables did not support this outcome, meaning there were other factors beyond this data that influenced the scores.

Does studying improve grades? For the most part, yes. There is a general increasing trend between the 4 study time groups. This variable was also considered significant by the linear model, meaning the trend is statistically significant as well.

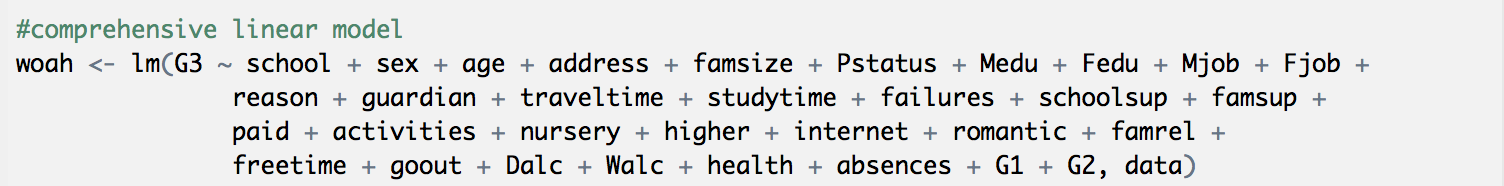


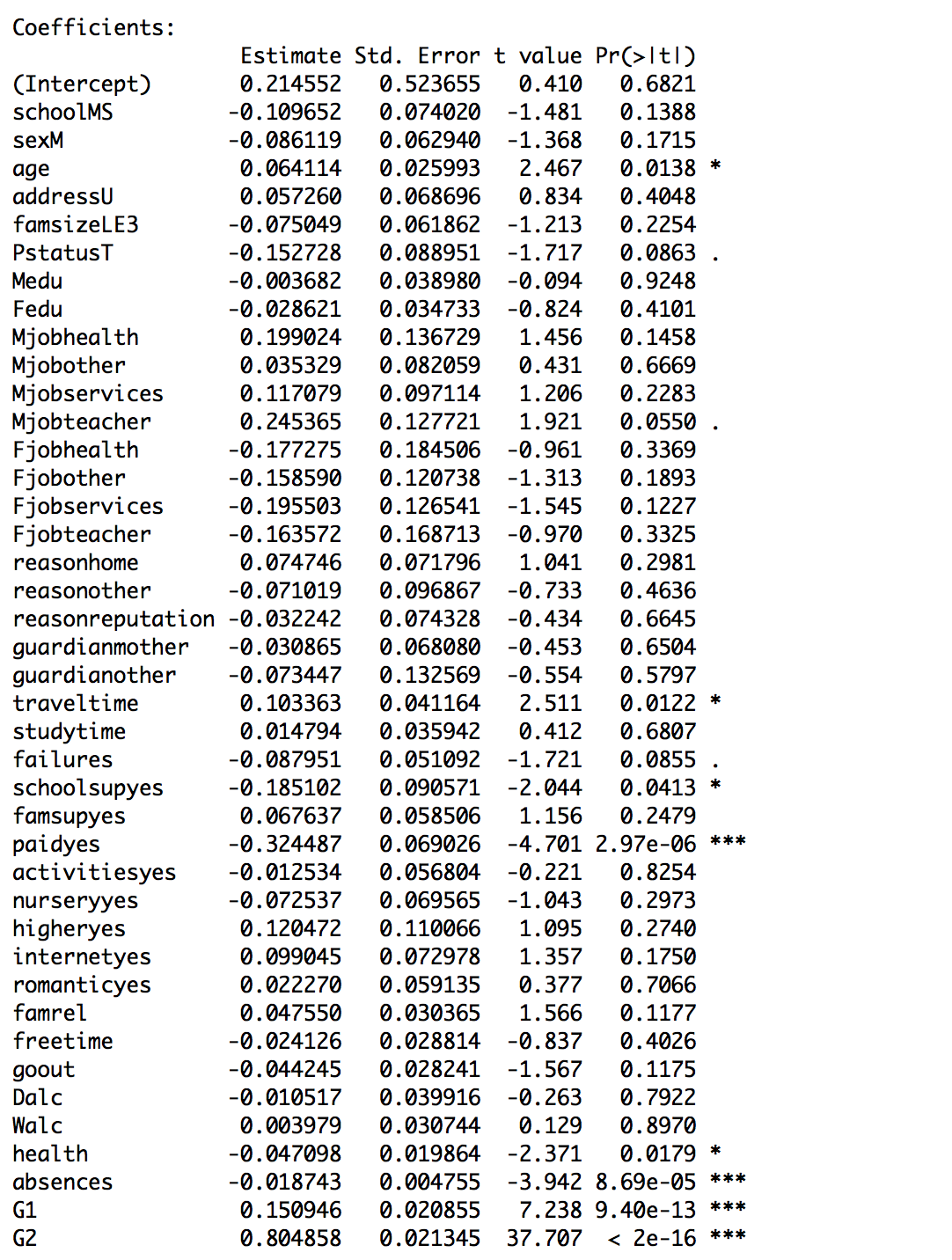
Examining heath vs. performance in class has the most interesting result, for the healthiest individuals (category 5) have a lower average grade than every other health level. This could be for many reasons, such as a greater sample size or these students focus on their health more than school. This theory can be tested with another plot of health vs. free time, but that graph is inconclusive just by looking at it. 

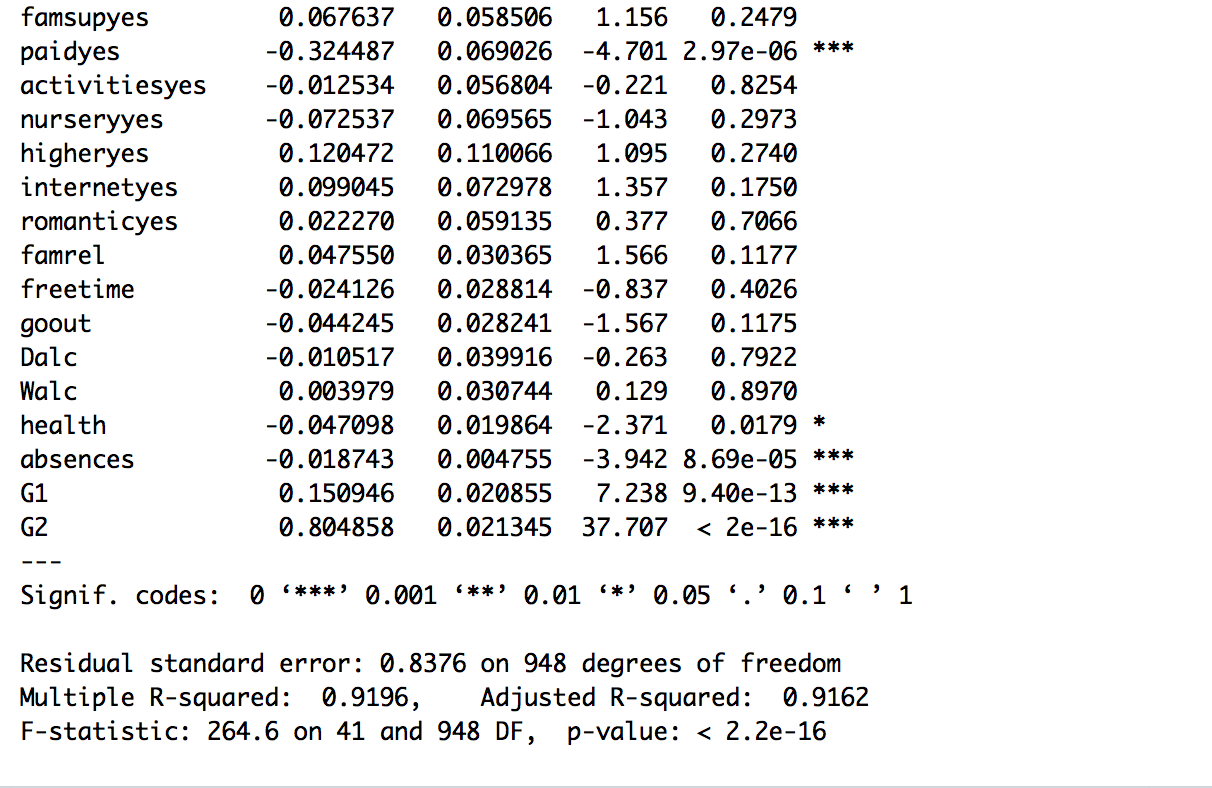
As for the initial model, a simple linear regression has a high R^2 value and is a good starting place for modeling. The predicted values come close to the actual ones.

Initial comprehensive linear model regresses G3 on all variables. From this model it is seen that the best predictors of G3 are scores received in G1 and G2. However this knowledge does not serve us well, there are too many predictors and the best ones are coming from data points received during the school year. The model shows significance at the 95% confidence level at least in:

* Age
* Travel time
* School supplies
* Paid yes
* Health
* Absences
* G1
* G2







Next steps include exploring additional models and comparing training error. Additionally completing more comprehensive figures to help answer the questions.