Exercise 1

The plot shows that with more variable(regressors) added to the model, the RSME value are getting smaller, but the reduced

value changes of RMSE for each time are different (Overall the value changes tend to get smaller with more variable added).

I think we should not use full-size model. Since to determine the "best" model that fits the criteria the most, several predictors that are not fitting the model appropriately should be removed to decreases RMSE value. By selecting appropriate variables(regressors), we could cut down considerably the number of possible regression models to consider.

If there is at least one model that contains less than 15 variables results a smaller RMSE value than that with full-size model, the assumption above would be correct.

Exercise 2

The resulting model used 14 variables that contains "YearBuilt*GarageCars+ KitchenAbvGr*TotRmsAbvGrd + GrLivArea*PoolArea + BedroomAbvGr + TotalBsmtSF + BsmtFinSF1 + YearRemodAdd*Fireplaces + ScreenPorch*LotArea + WoodDeckSF" as regressors.

We tried multiple different combinations of variables that in full-size model, calculated the RMSE value for each of them and compared to select the "best" model, which results the minimum RMSE.

So far, the RMSE value of our final model is 34485.711, which is much smaller than the initial value that result by the full-size model (38090.605).