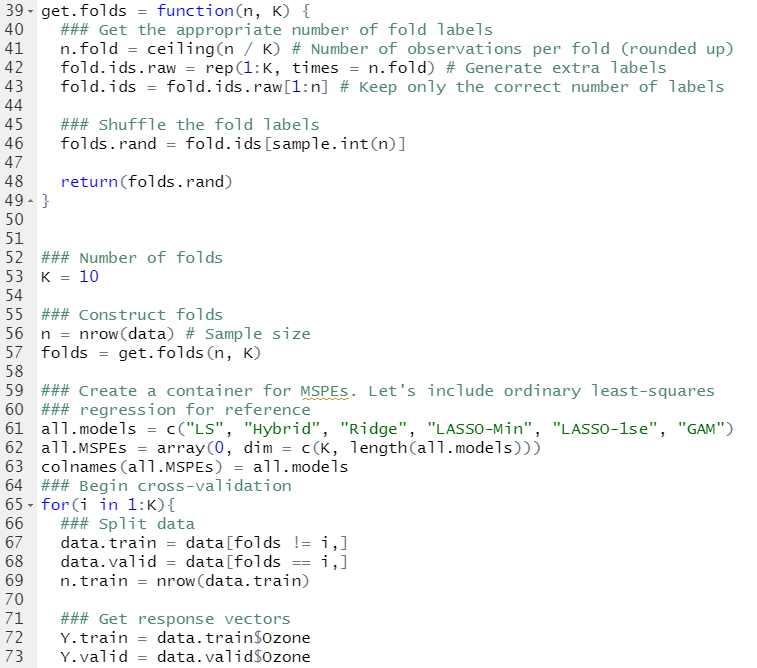
**Application**

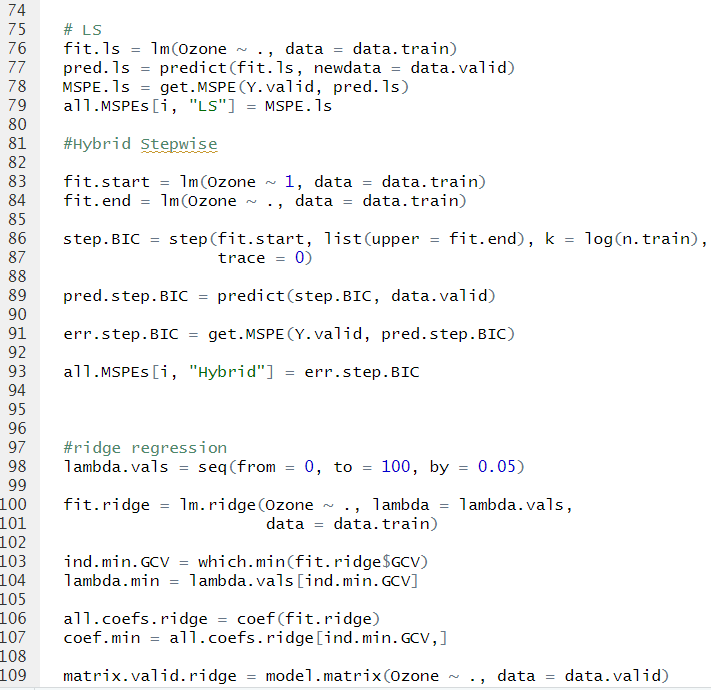
Refer to the Air Quality data described previously, and the analyses we have done with Ozone as the response variable, and the five explanatory variables (including the two engineered features).

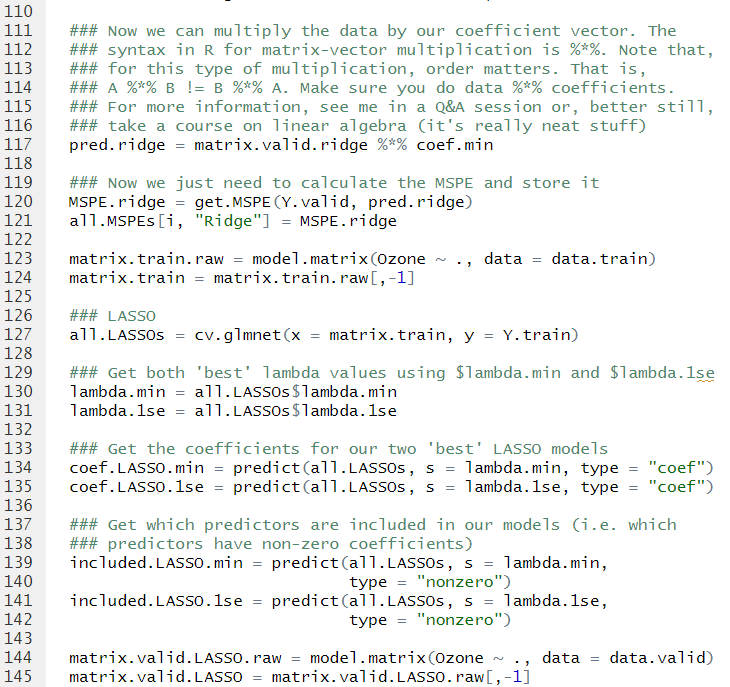
2. Add GAM on all variables to the 10-fold CV comparison that has been used for LASSO,

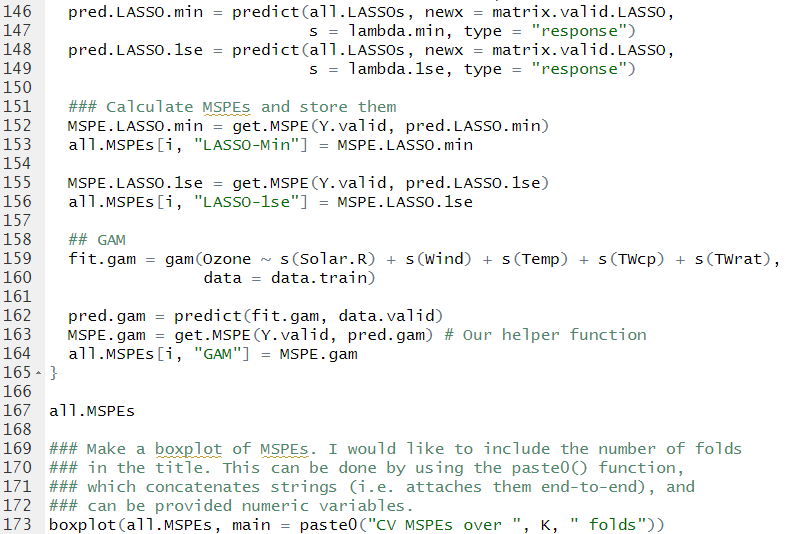
Ridge, and other methods. Use the same folds for GAM that were used for the other

methods.



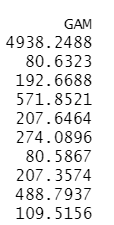






(a) **Report the separate MSPEs from each fold,** *MSPEv, v* = 1*, . . . ,* 10 **and**

**the MSPE for the full data.**

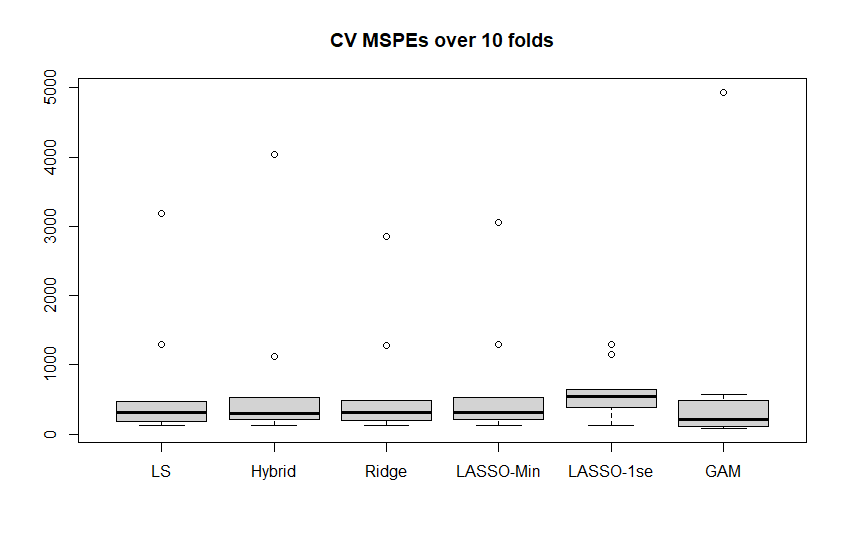


(b) **Starting with boxplots the plots made earlier for least squares, hybrid**

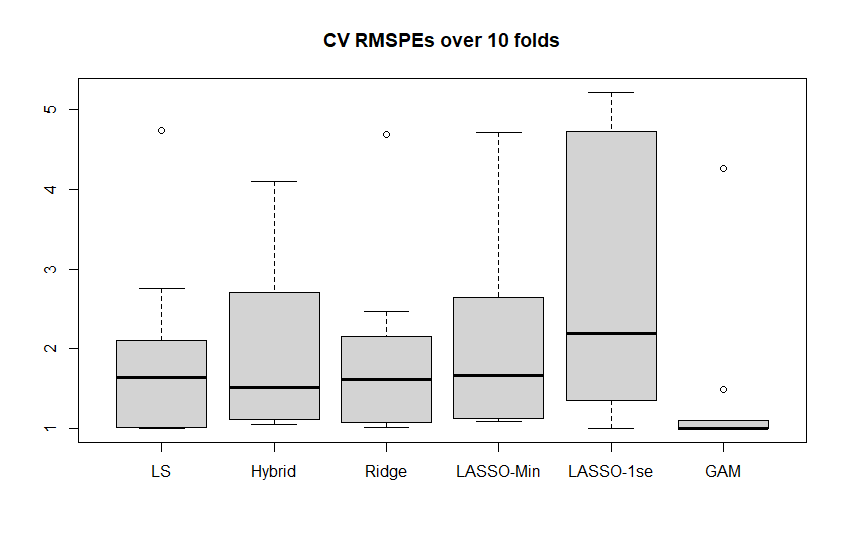
**stepwise, ridge, and LASSO, ADD a boxplot of the 10 CV error estimates**

**for GAM as the last box on the right. Comment on how GAM**

**compares to other methods**



(c) **Repeat this using relative MSPE.**



(d) Using the knowledge gained from the analysis you did in Question 1, give a 1-

sentence explanation for why GAM performs as it does. (If it is better than other

methods, why? If it is no better than other methods, why?)

* Gam may not be the best model if the variables are interactive, however we include two interactions so it covers it.