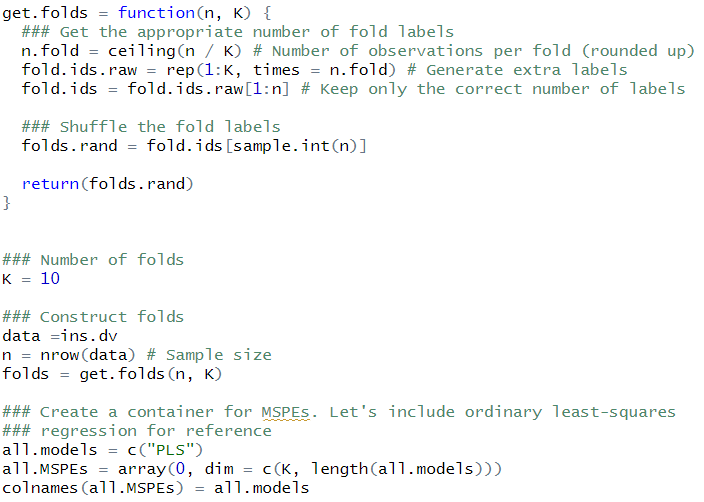
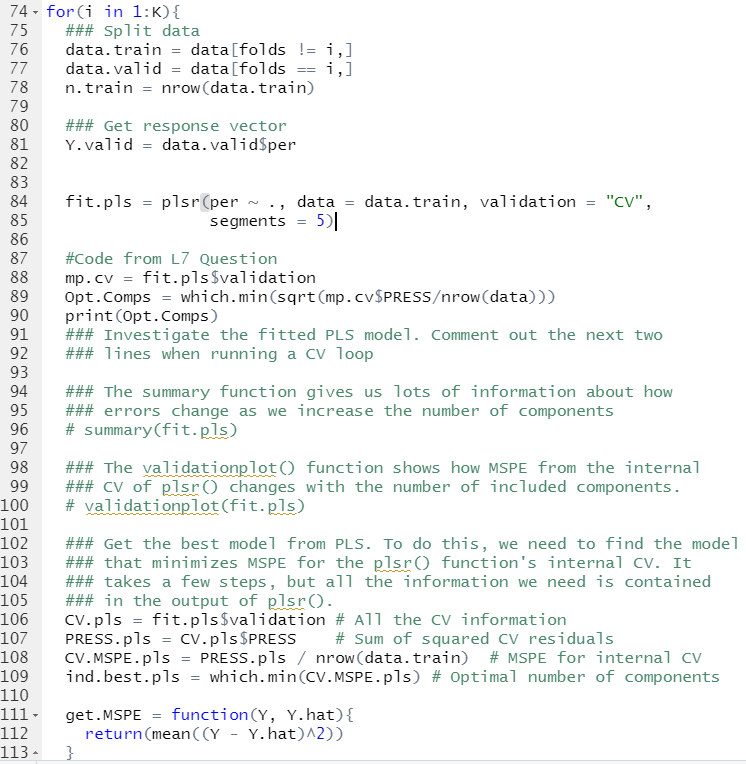
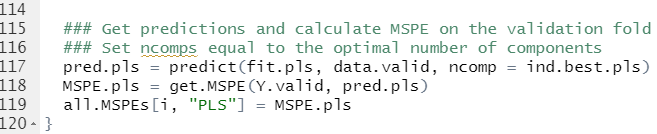
**OZONE DATA**

Refer Lecture 6, Problem 3, where you used 10-fold CV for comparing LASSO models

Using *the same 10 folds* estimate the MSPE for PLS







(a) On each training set, run the PLS function, choosing the optimum number of

components using 10-fold CV. You can figure out how many components is optimal

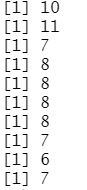
according to CV using code like this, assuming that I have named my model,

mod.pls

mp. cv = mod. p l s $ v a l i d a t i o n

Opt .Comps = which .min( s q r t (mp. cv$PRESS/nrow( i n s . dv ) ) )

**Report the optimal number of components for each of the 10 folds.**

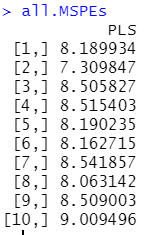


(b) Use the predict() function to predict responses on the test set. You need to

specify ncomps for this function.

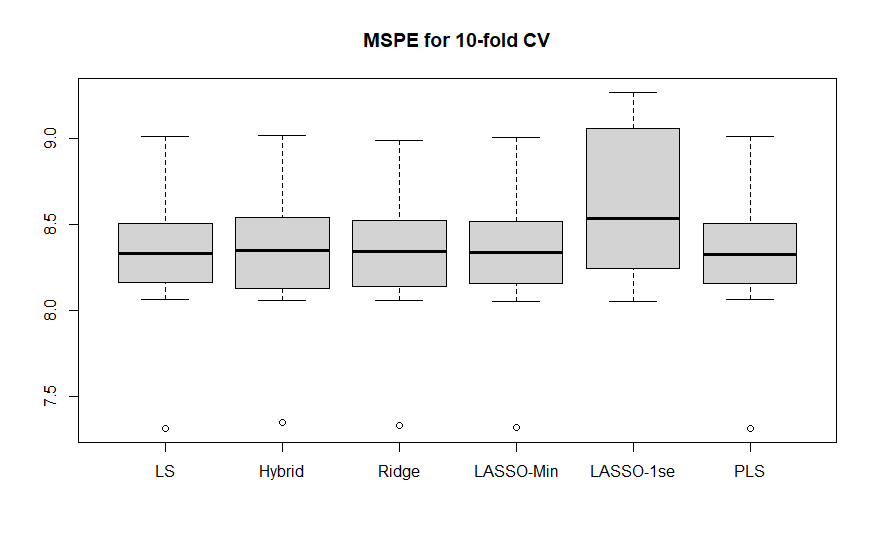
(c) **Report the separate MSPEs from each fold,** *MSPEv, v* = 1*, . . . ,* 10 **and**

**the MSPE for the full data.**



(d) **ADD a boxplot for PLS to the boxplots you made for other models.**

**Comment on how well PLS compares to the other models.**



**->** PLS does pretty good because it seems to have the lowest MSPE value.

(e) **Remake the plot using relative MSPE and comment.**

