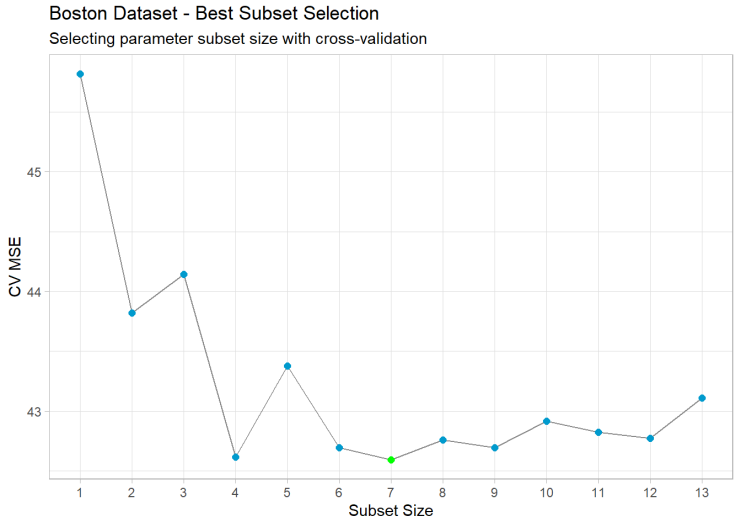
Best Subset Selection:

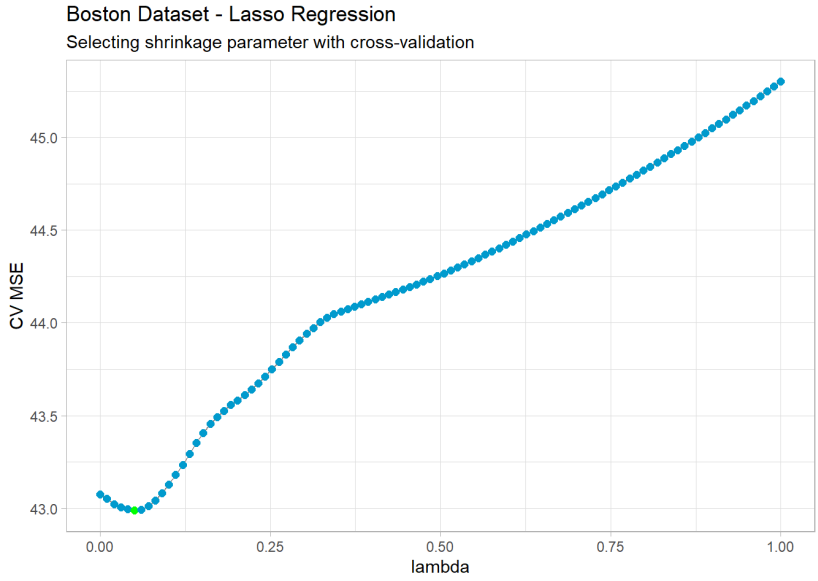
I used 10 fold cross validation for dataset, as we can see the MSE for Best subset selection is the minimum among other techniques.



Minimum CV MSE:

```
## [1] 42.58941
```

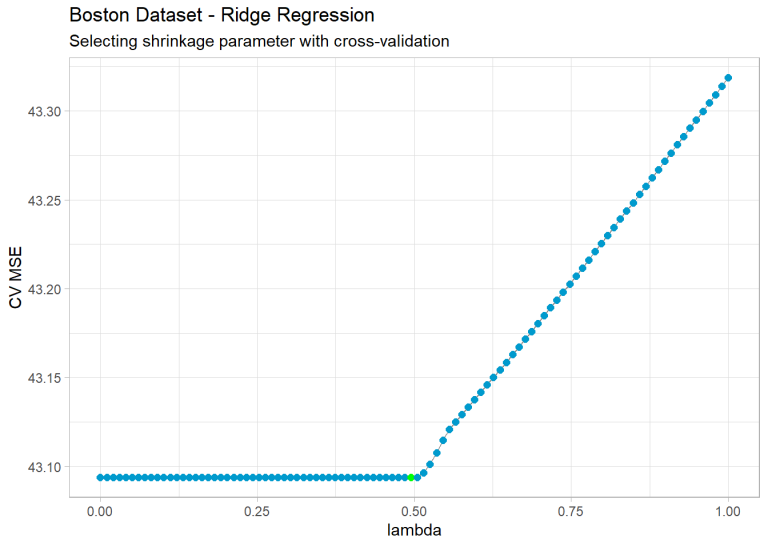
Lasso Regression:



Minimum CV MSE:

```
## [1] 42.98778
```

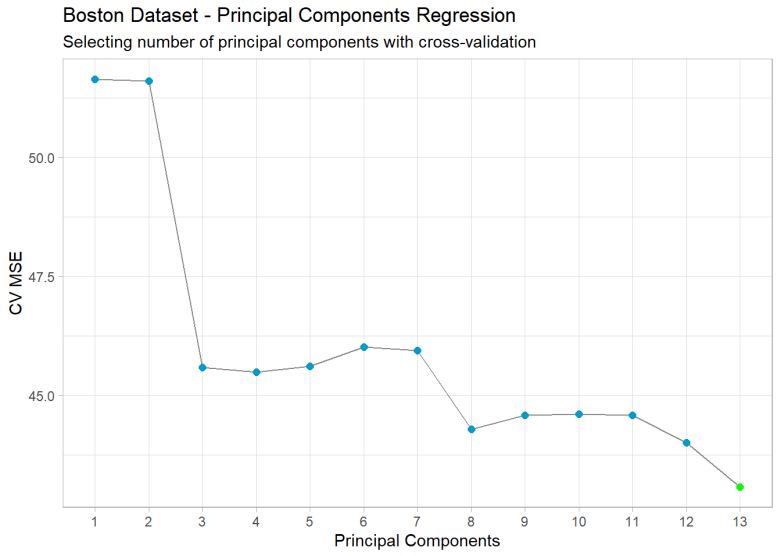
Ridge Regression:



Minimum CV MSE:

```
## [1] 43.09366
```

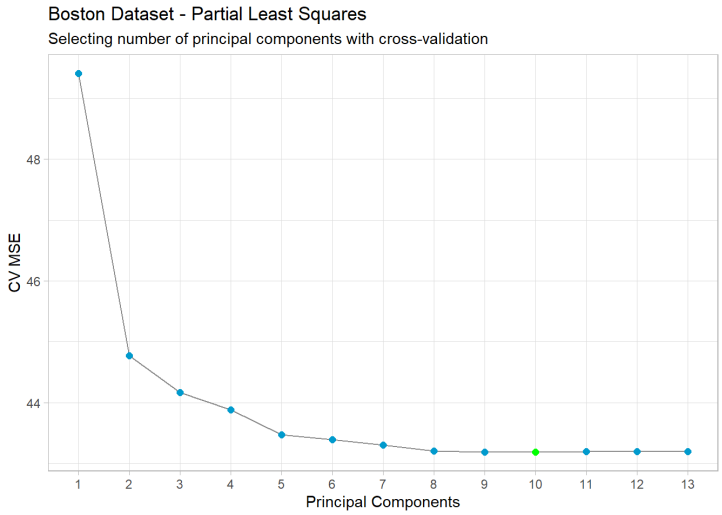
Principal Components Regression:



Minimum CV MSE:

```
## [1] 43.06695
```

Partial Least Squares:



Minimum CV MSE:

```
## [1] 43.18333
```

(b) Proposed Model

A: The model that achieved the lowest cross-validation Mean Squared Error (MSE) was selected using best subset selection, with an MSE of 42.59. However, utilizing non-linear techniques could potentially result in even better performance in practical applications.

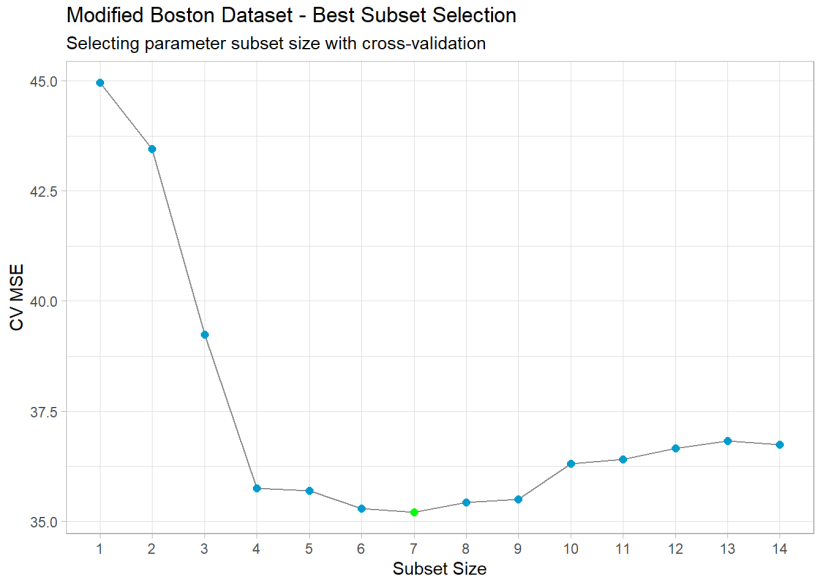
To support this claim, I made some slight modifications to the dataset, which I believe would yield the most promising outcomes. By making some changes to dataset we proposed new model:

- adding quadratic & cubic terms for `medv` (the most obvious case of a non-linear predictor)
- creating `rad_cat` - a binary variable indicating whether `rad` is > 20
- removing `rad` & `tax` (their useful information is captured by `rad_cat`)

```
## Rows: 506

## Columns: 15

## $ crim     <dbl> 0.00632, 0.02731, 0.02729, 0.03237, 0.06905, 0.02985, 0.0
8...
```

```
## $ zn       <dbl> 18.0, 0.0, 0.0, 0.0, 0.0, 0.0, 12.5, 12.5, 12.5, 12.5, 12
....

## $ indus    <dbl> 2.31, 7.07, 7.07, 2.18, 2.18, 2.18, 7.87, 7.87, 7.87, 7.8
7...

## $ chas     <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0...

## $ nox      <dbl> 0.538, 0.469, 0.469, 0.458, 0.458, 0.458, 0.524, 0.524, 0
....

## $ rm       <dbl> 6.575, 6.421, 7.185, 6.998, 7.147, 6.430, 6.012, 6.172, 5
....

## $ age      <dbl> 65.2, 78.9, 61.1, 45.8, 54.2, 58.7, 66.6, 96.1, 100.0, 85
....

## $ dis      <dbl> 4.0900, 4.9671, 4.9671, 6.0622, 6.0622, 6.0622, 5.5605, 5
....

## $ ptratio  <dbl> 15.3, 17.8, 17.8, 18.7, 18.7, 18.7, 15.2, 15.2, 15.2, 15.
2...

## $ black    <dbl> 396.90, 396.90, 392.83, 394.63, 396.90, 394.12, 395.60, 3
9...

## $ lstat    <dbl> 4.98, 9.14, 4.03, 2.94, 5.33, 5.21, 12.43, 19.15, 29.93,
1...

## $ medv     <dbl> 24.0, 21.6, 34.7, 33.4, 36.2, 28.7, 22.9, 27.1, 16.5, 18.
9...

## $ medv_sq  <dbl> 576.00, 466.56, 1204.09, 1115.56, 1310.44, 823.69, 524.41
,...

## $ medv_cub <dbl> 13824.000, 10077.696, 41781.923, 37259.704, 47437.928, 23
6...

## $ rad_cat  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0...
```

In a similar manner as previously, I carry out best subset selection, which involves obtaining the best-fitting model for each subset size, and then utilizing repeated cross-validation to select the optimal subset.



In terms of predictive accuracy, the linear model using the raw data showed a noteworthy enhancement compared to other linear techniques, with an MSE of 35.2.
Although one may argue that this model is prone to overfitting, as the transformations were derived from the entire Boston dataset instead of just the training sample, I believe that these transformations would have been discovered even with a smaller training sample.
Overall, this serves as an indication that non-linear methods that can effectively capture these relationships with less bias hold greater potential.

(c) Selected Model

A: The selected model uses just 7 variables:

```
##   (Intercept)           nox           dis       ptratio          medv
##  52.291186700 -11.960713097  -0.468287570  -0.328043396  -3.885786605
##       medv_sq      medv_cub       rad_cat
##   0.123927558  -0.001246049   9.187000141
```

Incorporating additional features may decrease the training Residual Sum of Squares (RSS), but it may also increase the cross-validation Mean Squared Error (MSE). thus there is no proof that they will enhance the predictive performance outside the sample.
To opt for a more efficient model, we could choose the four-variable model that solely includes the four variables that I constructed.

```
##  (Intercept)          medv       medv_sq      medv_cub       rad_cat
## 38.178959648 -4.035693216  0.132937552 -0.001358685  8.101325067
```