# Prostate Data Analysis

The errors from the Exhaustive subset selection are shown below:

```
> training_error_store
[1] 0.6646057 0.5536096 0.5210112 0.4897760 0.4786485 0.4558176 0.4393627 0.4391998
> test_error_store
[1] 0.4797387 0.4924823 0.4005308 0.4563321 0.4859242 0.5485933 0.5165135 0.5212740
> aic_store
[1] 168.7642 158.5210 156.4548 154.3127 154.7729 153.4984 153.0350 155.0101
> bic_store
[1] 175.3782 167.3397 167.4783 167.5408 170.2058 171.1359 172.8772 177.0570
> |
```
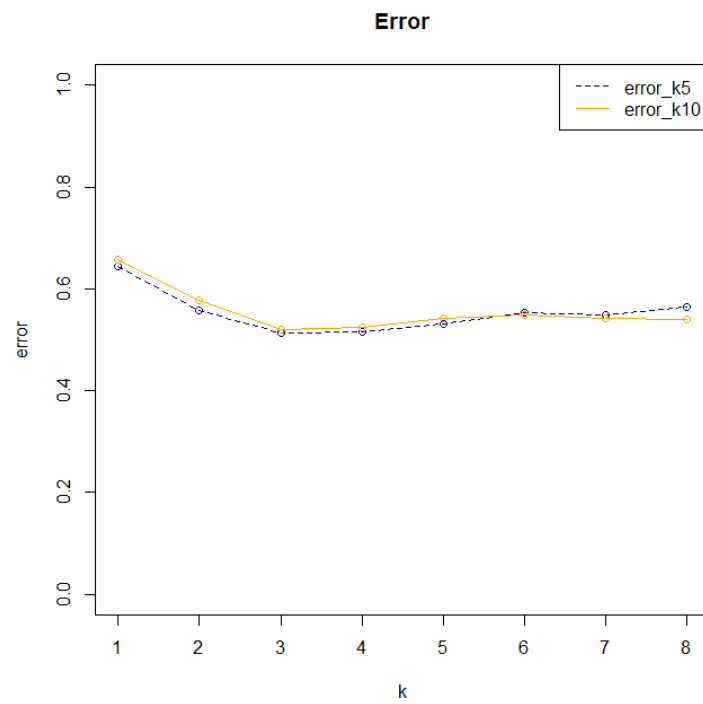

AIC & BIC
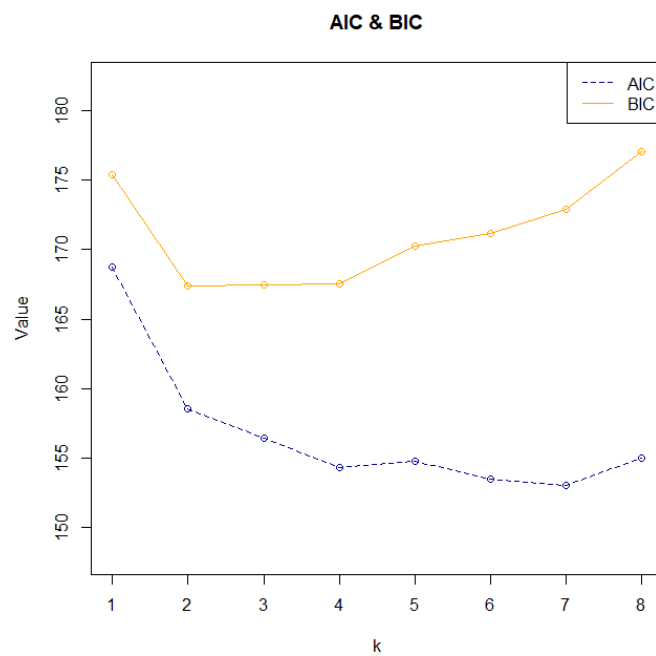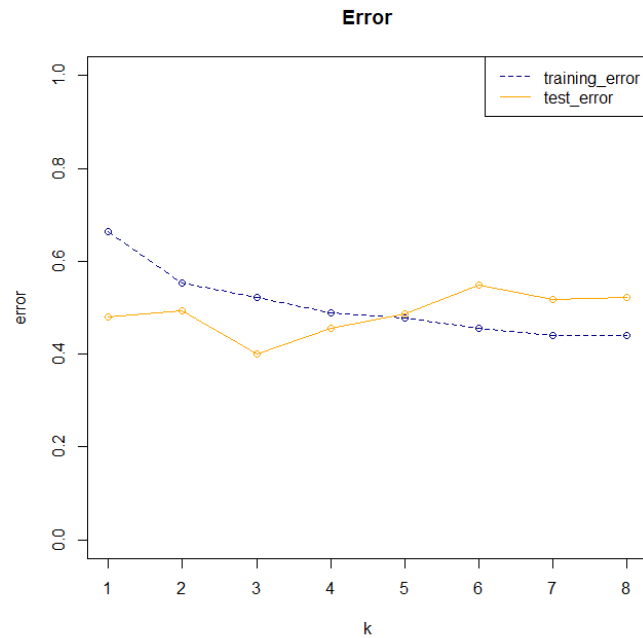
Error


Cross-validation Error for k=5,10 is shown below:

```
' -
> cv_error_k5_store
[1] 0.6441184 0.5567716 0.5134967 0.5162473 0.5304723 0.5534809 0.5476093 0.5629419
> cv_error_k10_store
[1] 0.6578841 0.5774885 0.5199002 0.5249848 0.5409154 0.5492809 0.5417233 0.5399521
>
```
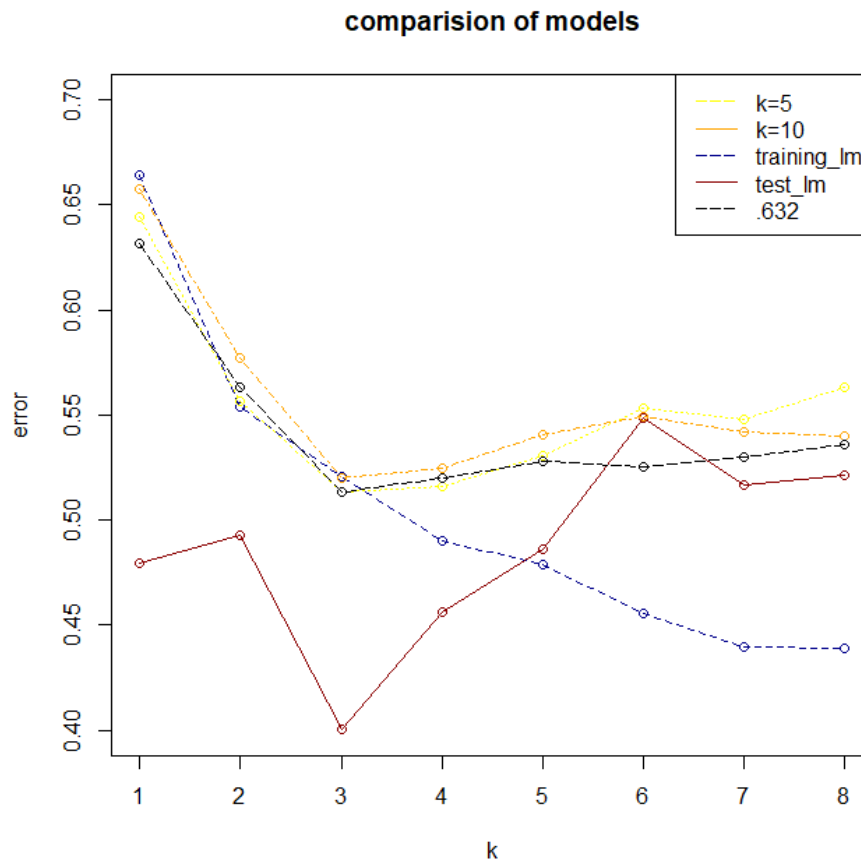
Error

Bootstrap.632 error:

```
> bootstrap.632_error_store
[1] 0.6315205 0.5635025 0.5130604 0.5200012 0.5278691 0.5250263 0.5299995 0.5361999
> |
```
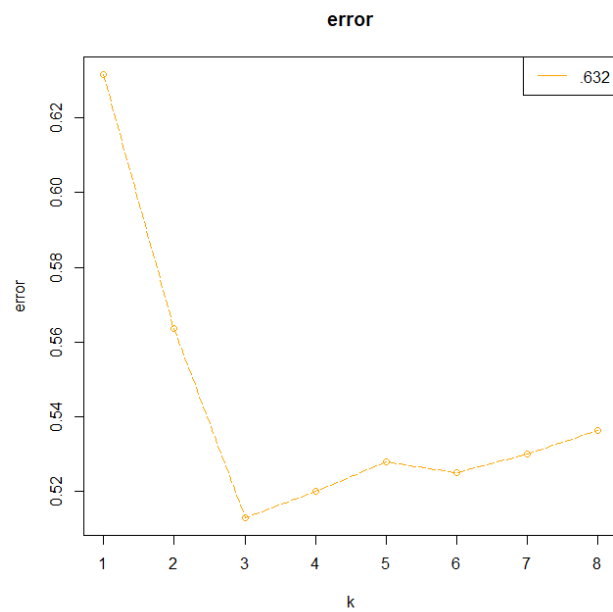
error

Comparing all the models:

comparision of models

As we can see from the above figure the error rate is lowest for the best subset selection method that too the model that has 3 variables.