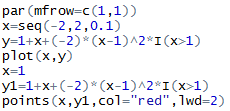
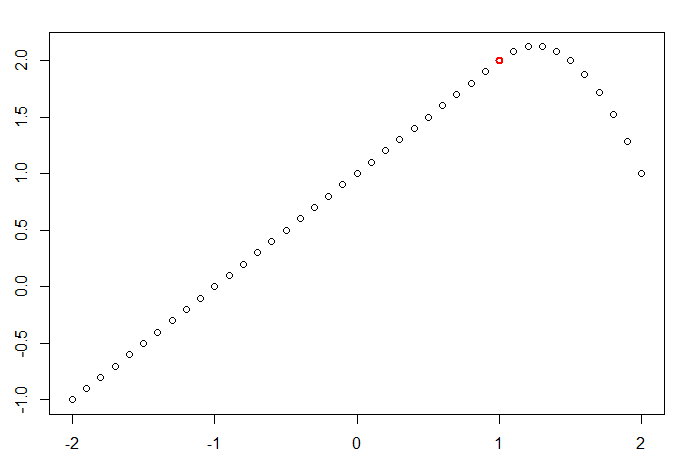
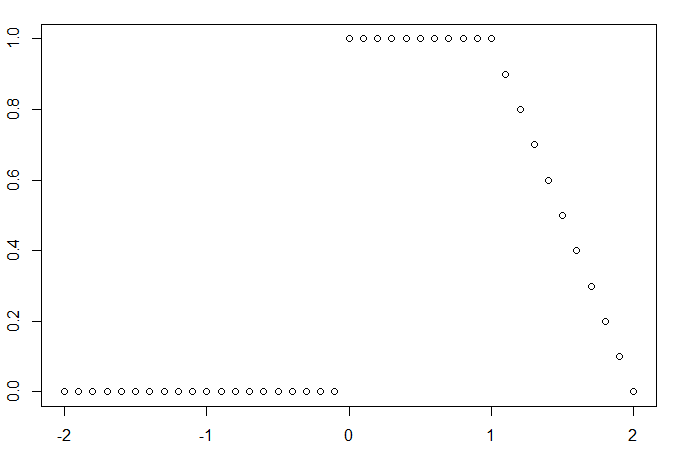
**Q3. Suppose we fit a curve with basis functions *b1(X)=X*,  *b2(X)=(X−1)2I(X≥1)*. We fit the linear regression model**

***Y=β0+β1b1(X)+β2b2(X)+ε,***

**and obtain coefficient estimates . Sketch the estimated curve between *X=−2* and *X=2*. Note the intercepts, slopes, and other relevant information.**

**Ans3:** The figure above shows the Y curve respect to X. We can see the curve is linear between -2 and 1, where Y=X+1. The curve is quadratic between 1 and 2, where Y=1+X-2(X-1)2 The knot is labeled in red in the figure.

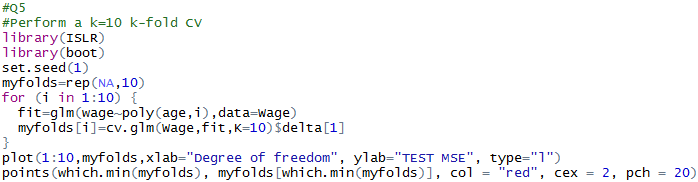
**Q4.** **Suppose we fit a curve with basis functions b1(X) = I(0 ≤ X ≤ 2) - (X -1)I(1 ≤ X ≤ 2), b2(X) = (X -3)I(3≤ X ≤ 4)+ I(4 < X ≤ 5). We fit the linear regression model**

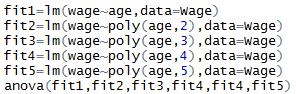
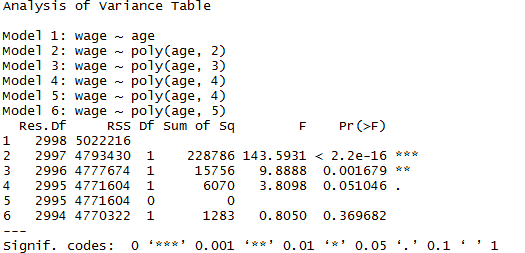
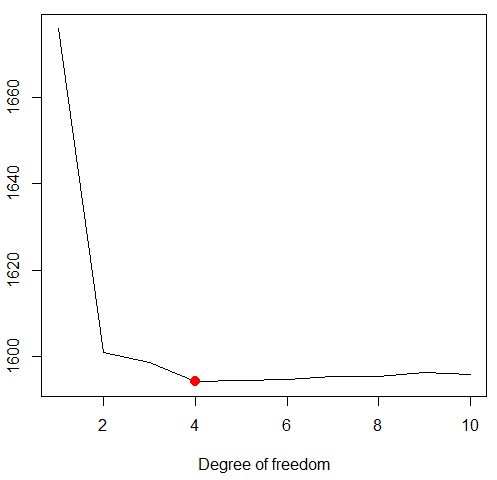
**Y = β0 + β1b1(X) + β2b2(X) +** ε**,and obtain coefficient estimates . Sketch the estimated curve between X = -2 and X = 2. Note the intercepts, slopes, and other relevant information.**

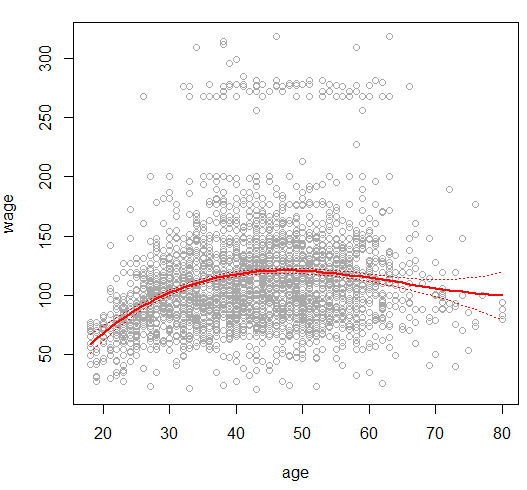
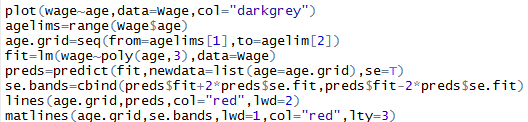
**Ans4:** The curve is constant between −2 and 0 where y=1, constant between 0and 1, where y=2, and linear between 1 and 2 where y=3−xy=3−x.

**Q5, In this exercise, you will further analyze the Wage data set considered throughout this chapter.**

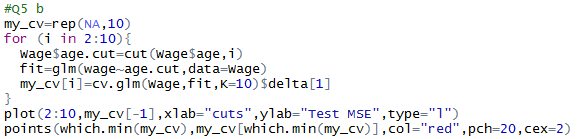
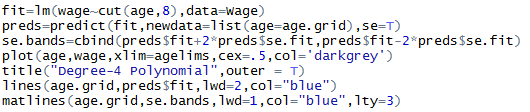
1. **Perform polynomial regression to predict wage using age. Use cross-validation to select the optimal degree d for the polynomial. What degree was chosen, and how does this compare to the results of hypothesis testing using ANOVA? Make a plot of the resulting polynomial fit to the data.**

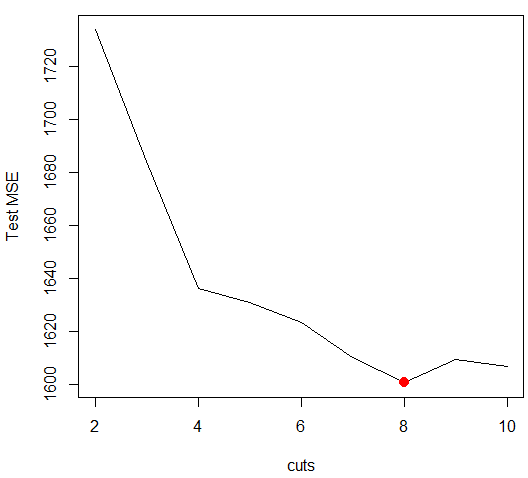
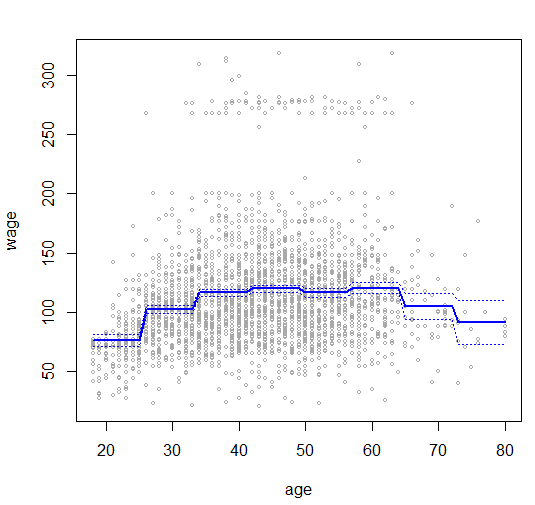
**Ans5a:** The result indicates that degree of polynomial of 4 is the optimized point ().



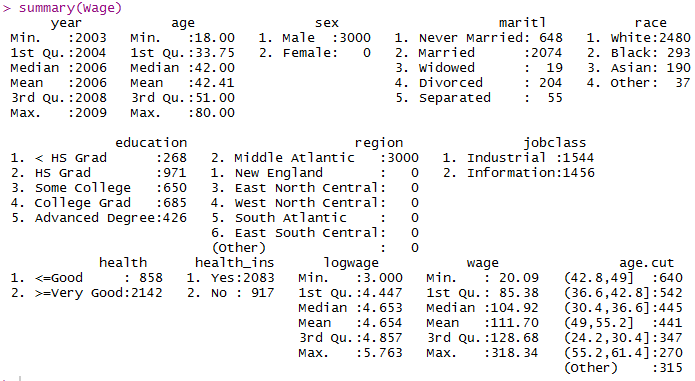
Using the ANOVA to do the hypothesis test, we can find that polynomial model with degree of 4 have the most reasonable P-value compared with other models. Degree of 3 is also reasonable. Higher or lower degrees’ models are not justified. The result is plotted below.

1. **Fit a step function to predict wage using age, and perform cross validation to choose the optimal number of cuts. Make a plot of the fit obtained.**

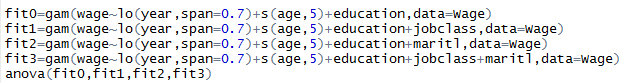
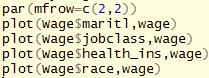
**Ans5b:** This result indicates that step function with 8 cuts is the best choice. The step function curve result is shown as follows.

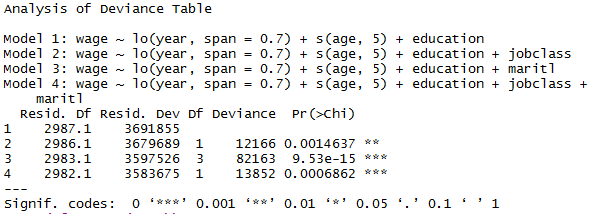
 7. The Wage data set contains a number of other features not explored in this chapter, such as marital status (maritl), job class (jobclass), and others. Explore the relationships between some of these other  
predictors and wage, and use non-linear fitting techniques in order to fit flexible models to the data. Create plots of the results obtained, and write a summary of your findings.

**Ans7**: First we made a summary over the data set, there are totally 12 variables in the original dataset.

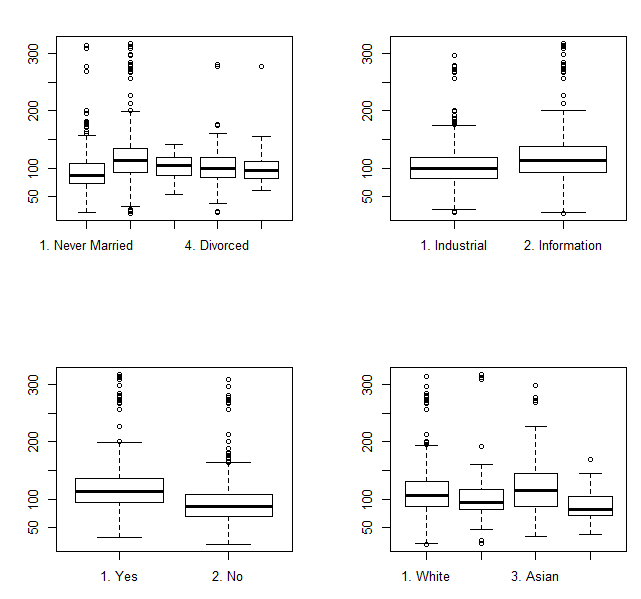


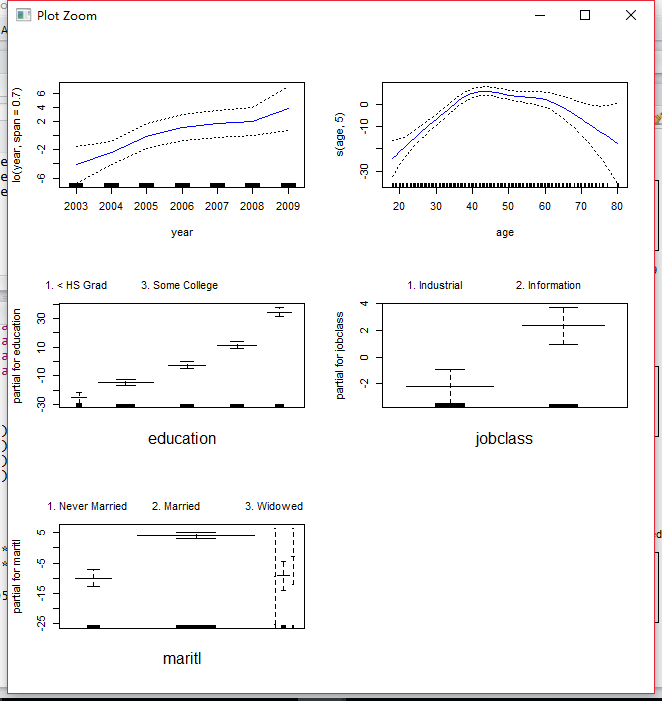
Then, we randomly choose some features from the table and plot their box charts by the codes as follows. We used GAM for the model fitting. The model we chosen for predicting is based on features of*“year”, “age”, “education”, “jobclass”, “maritl”.* The ANOVA is used to evaluate the model.





We can see the models of fit1,fit2,fit3 are all valid for this data set, the smallest P-Value is observed in model fit3. The fitting curve of fit3 is shown as follows.





Coefficients in fit3 model is shown below.

