Peer Grader Guidance

Please review the student expectations for peer review grading and peer review comments. Overall, we ask that you score with accuracy. When grading your peers, you will not only learn how to improve your future homework submissions but you will also gain deeper understanding of the concepts in the assignments. When assigning scores, consider the responses to the questions given your understanding of the problem and using the solutions as a guide. Moreover, please give partial credit for a concerted effort, but also be thorough. **Add comments to your review, particularly when deducting points, to explain why the student missed the points.** Ensure your comments are specific to questions and the student responses in the assignment.

Background

You have been contracted as an automobile consulting company to understand the factors on which the pricing of cars depends.

Data Description

The data consists of a data frame with 205 observations on the following 8 variables:

1. price: Response variable ($)
2. fueltype: Qualitative variable
3. carbody: Qualitative variable
4. carlength: Quantitative variable
5. enginesize: Quantitative variable
6. horsepower: Quantitative variable
7. peakrpm: Quantitative variable
8. highwaympg: Quantitative variable

Instructions on reading the data

To read the data in `R`, save the file in your working directory (make sure you have changed the directory if different from the R working directory) and read the data using the `R` function `read.csv()`

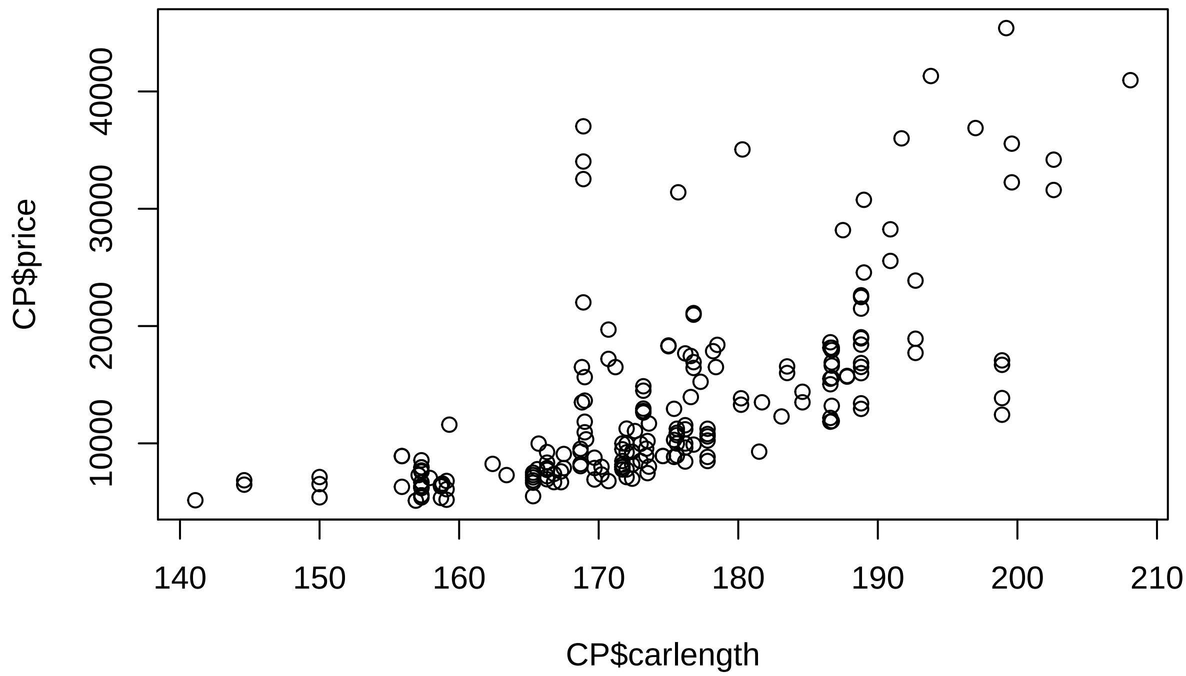
```
CP = read.csv("C:/Users/mjpearl/Desktop/omsa/ISYE-6414-OAN/hw2/carprice.csv", head = TRUE)
head(CP)
```

```
##   fueltype     carbody carlength enginesize horsepower peakrpm highwaympg price
## 1      gas convertible     168.8        130        111    5000         27 13495
## 2      gas convertible     168.8        130        111    5000         27 16500
## 3      gas   hatchback     171.2        152        154    5000         26 16500
## 4      gas       sedan     176.6        109        102    5500         30 13950
## 5      gas       sedan     176.6        136        115    5500         22 17450
## 6      gas       sedan     177.3        136        110    5500         25 15250
```

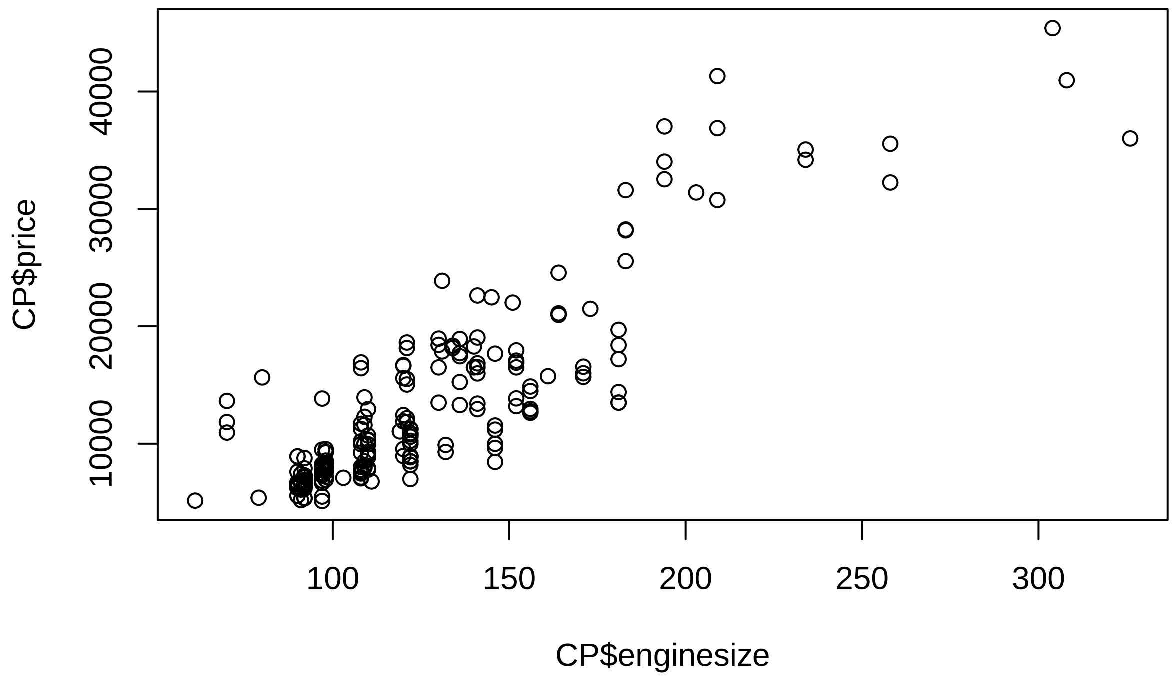
Question 1: Exploratory Data Analysis [12 points]

a. **3 pts** Create plots of the response, *price*, against three quantitative predictors (for simplicity) *carlength*, *enginesize*, and *horsepower*. Describe the general trend (direction and form) of each plot.

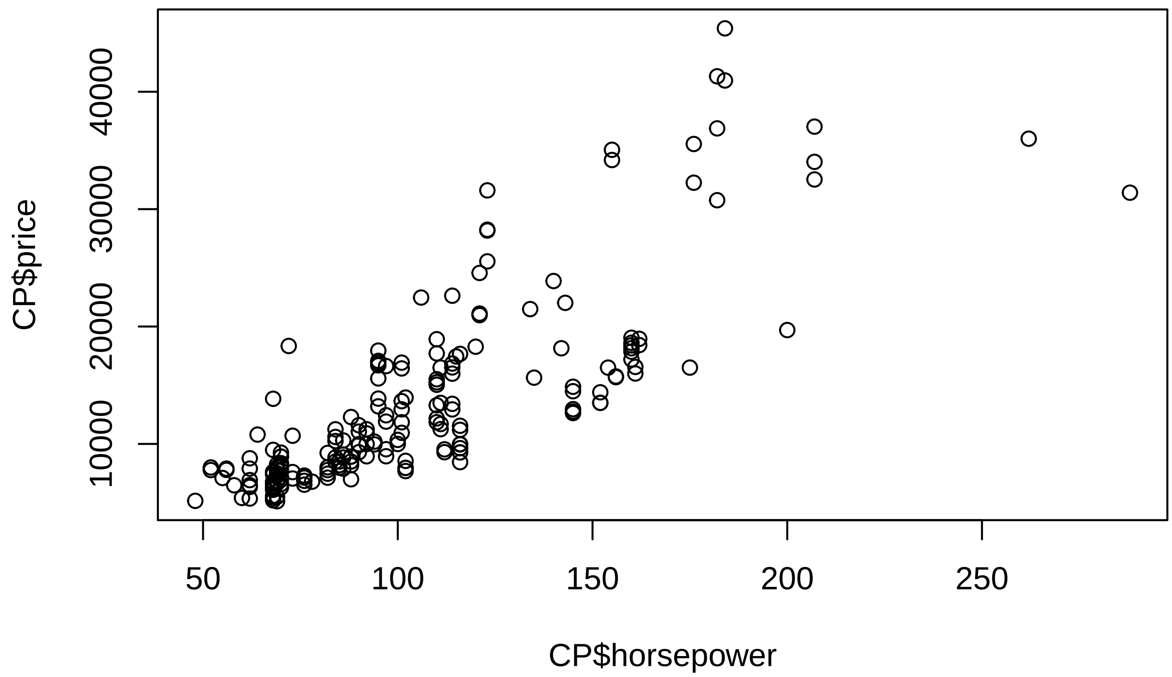
```r
plot(CP$carlength,CP$price)
```




```r
plot(CP$enginesize,CP$price)
```




```r
plot(CP$horsepower,CP$price)
```



b. **3 pts** What is the value of the correlation coefficient for each of the above pair of response and predictor variables? What does it tell you about your comments in part (a).

```
lm1 <- lm(CP$price~CP$carlength)
summary(lm1)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$carlength)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11826  -3579  -1384   1968  26028
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -63690.7     5792.8  -10.99   <2e-16 ***
## CP$carlength    442.2       33.2   13.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5850 on 203 degrees of freedom
## Multiple R-squared:  0.4664, Adjusted R-squared:  0.4638
## F-statistic: 177.4 on 1 and 203 DF,  p-value: < 2.2e-16
```

```r
lm2 <- lm(CP$price~CP$horsepower)
summary(lm2)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$horsepower)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11897.5  -2350.4   -711.1   1644.6  19081.4
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3721.761    929.849  -4.003 8.78e-05 ***
## CP$horsepower   163.263      8.351  19.549  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4717 on 203 degrees of freedom
## Multiple R-squared:  0.6531, Adjusted R-squared:  0.6514
## F-statistic: 382.2 on 1 and 203 DF,  p-value: < 2.2e-16
```

```r
lm3 <- lm(CP$price~CP$enginesize)
summary(lm3)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$enginesize)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10664.2  -2225.0   -482.4   1588.0  14271.5
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8005.446    873.221  -9.168   <2e-16 ***
## CP$enginesize   167.698      6.539  25.645   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3889 on 203 degrees of freedom
## Multiple R-squared:  0.7641, Adjusted R-squared:  0.763
## F-statistic: 657.6 on 1 and 203 DF,  p-value: < 2.2e-16
```

When taking the individual linear regression results for each variable individually, we can see the correlation against the target variable is the following:

Car Length = 0.4638 Horse Power = 0.6514 Engine Size = 0.763

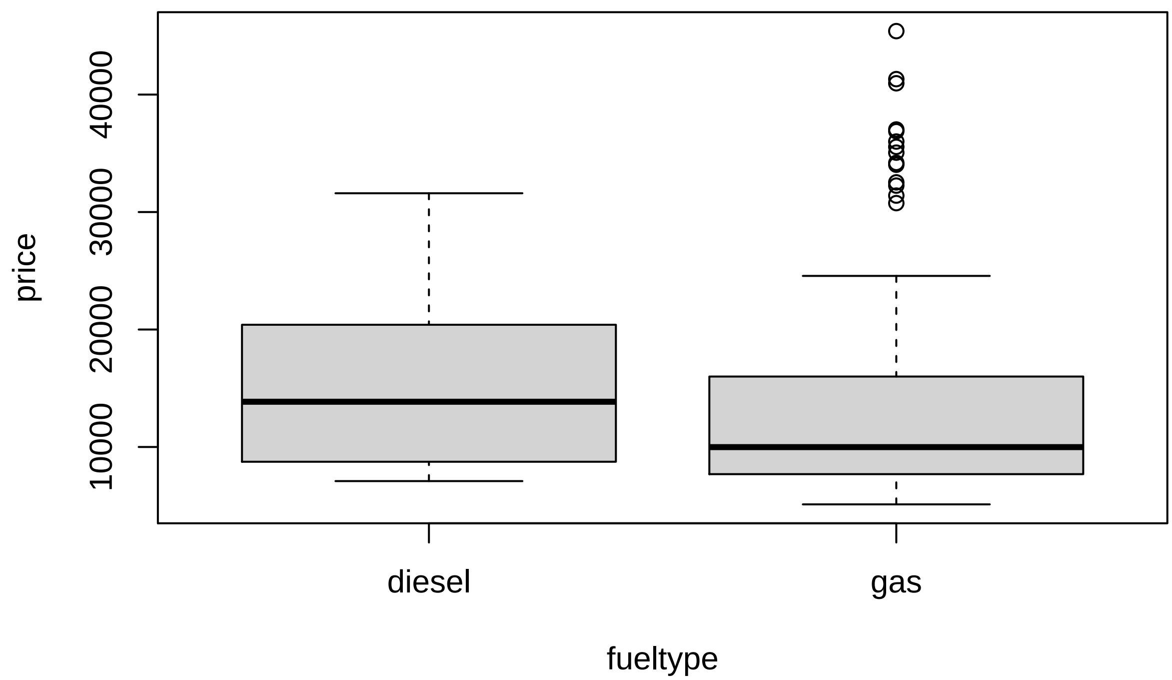
We can see that with each variable that the level of correlation increased as we start to get to the Engine Size variable.

Car Length does not seem to exhibit that much correlation with price, which makes sense since there's a much higher population of cars with modest to small lengths that are much more expensive than the longer sedans / trucks currently on the market.

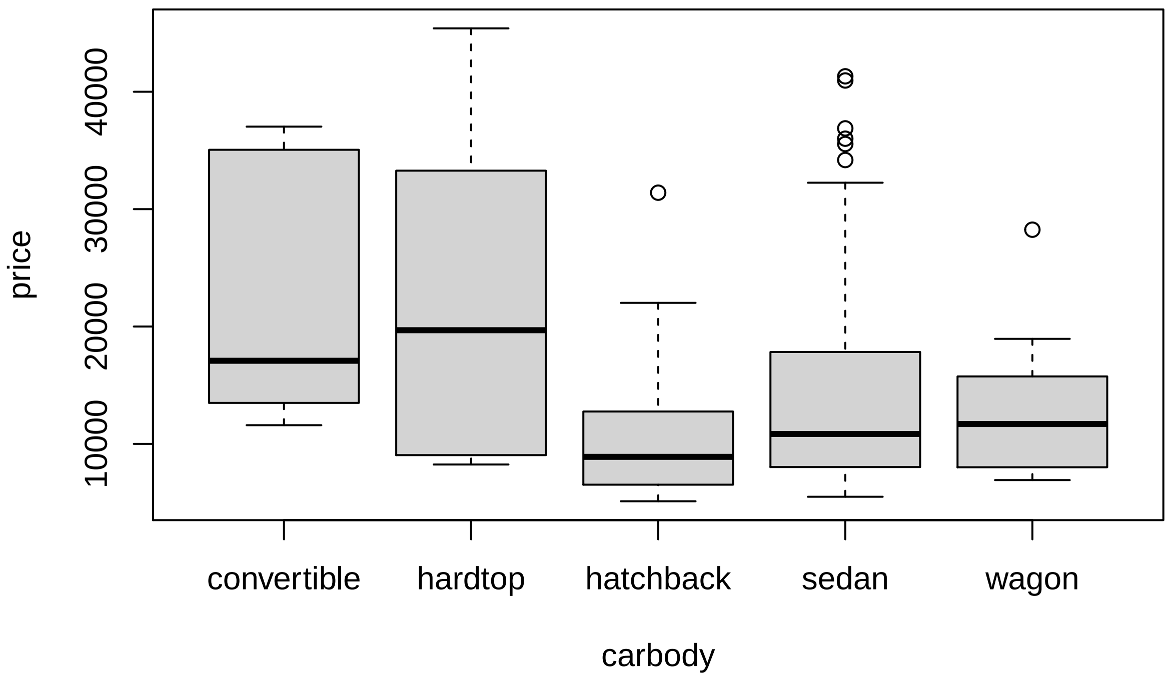
The other variables Horse Power and Engine Size seem to exhibit a strong positive correlation, which makes sense due to our capitalists culture's preference to powerful / fast cars that come at a higher price point.

c. **3 pts** Create box plots of the response, *price*, and the two qualitative predictors *fueltype*, and *carbody*. Based on these box plots, does there appear to be a relationship between these qualitative predictors and the response?

```
boxplot(price~fueltype,data=CP)
```




```
boxplot(price~carbody,data=CP)
```



d. **3 pts** Based on the analysis above, does it make sense to run a multiple linear regression with all of the predictors?

Based on our results, it's evident that we're going to need to use more than just the variables above in order to account for all of the noise when predicting a car price. This makes sense, as there's quite a few features that consumers will look for when trying to buy a new vehicle, not just 1-3 features. We see strong correlation with some features, but based on the size of our dataset, we don't need to worry about filtering the number of predictors used in order to reduce the amount of training time.

Utilizing more features in this case will increased our overall R^2 of the model and lead to better predictions so long as we try and avoid overfitting as best as we can.

From our boxplot, we can see that gas powered vehicles, contains several outliers past the 3rd quartile, but this makes sense as luxury cars at a higher pricepoint represent a smaller % of the population of cars since not all consumers can afford them. Those cars also tend to be classified as sedans a lot of the time, so this makes sense that we're seeing outliers for the carbody type of Sedan.

From our results, it's justified for us to use a multiple linear regression model to predict price.

Note: Please work on non-transformed data for all of the following questions.

Question 2: Fitting the Multiple Linear Regression Model [10 points]

Build a multiple linear regression model, named *model1*, using the response, *price*, and all 7 predictors, and then answer the questions that follow:

a. **5 pts** Report the coefficient of determination for the model and give a concise interpretation of this value.

```
model1 <- lm(CP$price ~ CP$fueltype + CP$carbody + CP$carlength + CP$enginesize + CP$horsepower + CP$pea
```

b. **5 pts** Is the model of any use in predicting price? Conduct a test of overall adequacy of the model, using $\alpha = 0.05$. Provide the following elements of the test: null hypothesis H_0, alternative hypothesis H_a, F- statistic or p-value, and conclusion.

```
summary(model1)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$fueltype + CP$carbody + CP$carlength +
##     CP$enginesize + CP$horsepower + CP$peakrpm + CP$highwaympg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9012.6 -1848.1   -48.1  1658.0 13011.4
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.235e+04  9.325e+03  -2.397 0.017502 *
## CP$fueltypegas      -3.810e+03  9.596e+02  -3.970 0.000101 ***
## CP$carbodyhardtop   -2.904e+03  1.790e+03  -1.622 0.106401
## CP$carbodyhatchback -5.128e+03  1.436e+03  -3.571 0.000449 ***
## CP$carbodysedan     -4.305e+03  1.477e+03  -2.914 0.003985 **
## CP$carbodywagon     -5.504e+03  1.618e+03  -3.402 0.000811 ***
## CP$carlength         9.564e+01  3.915e+01   2.443 0.015471 *
## CP$enginesize        1.032e+02  1.277e+01   8.082 6.69e-14 ***
## CP$horsepower        4.703e+01  1.383e+01   3.400 0.000818 ***
## CP$peakrpm           2.126e+00  6.605e-01   3.218 0.001512 **
## CP$highwaympg       -6.235e+01  6.868e+01  -0.908 0.365114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3272 on 194 degrees of freedom
## Multiple R-squared:  0.8405, Adjusted R-squared:  0.8323
## F-statistic: 102.2 on 10 and 194 DF,  p-value: < 2.2e-16
```

Question 3: Model Comparison [12 points]

a. **4 pts** Assuming a marginal relationship between the car's body type and its price, perform an ANOVA F-test on the means of the car's body types. Using an $\alpha - level$ of 0.05, can we reject the null hypothesis that the means of the car body types are equal? Please interpret.

```
summary(aov(price ~ carbody, data=CP))
```

```
##              Df    Sum Sq   Mean Sq F value   Pr(>F)
## carbody       4 1.802e+09 450499206   8.032 5.03e-06 ***
```

```
## Residuals   200 1.122e+10  56088213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H0: There are not equal means for each group representing the carbody types and price (u1 != u2) HA: There are equal means for the two groups representing carbody types and price (u1 = u2)

Based on the p-value of 0.00000503, we reject the null hypothesis that the mean are equal since the p-value is much smaller than the alpha level of 0.05. This means we can conclude that the carbody type does have an impact on the price.

b. **4 pts** Now, build a second multiple linear regression model, called *model2*, using *price* as the response variable, and all variables except *carbody* as the predictors. Conduct a partial F-test comparing *model2* with *model1*. What is the partial-F test p-value? Can we reject the null hypothesis that the regression coefficients for *carbody* are zero at $\alpha - level$ of 0.05?

```r
model2 <- lm(CP$price ~ CP$fueltype + CP$carlength + CP$enginesize + CP$horsepower + CP$peakrpm + CP$hi
summary(model2)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$fueltype + CP$carlength + CP$enginesize +
##     CP$horsepower + CP$peakrpm + CP$highwaympg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9094.0 -1887.5   -55.5  1581.8 12865.8
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.390e+04  8.808e+03  -2.713  0.00725 **
## CP$fueltypegas -4.099e+03  9.862e+02  -4.156 4.81e-05 ***
## CP$carlength    7.304e+01  3.388e+01   2.156  0.03230 *
## CP$enginesize   1.145e+02  1.276e+01   8.980  < 2e-16 ***
## CP$horsepower   4.390e+01  1.400e+01   3.135  0.00198 **
## CP$peakrpm      2.241e+00  6.732e-01   3.329  0.00104 **
## CP$highwaympg  -7.903e+01  6.850e+01  -1.154  0.25003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3375 on 198 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8216
## F-statistic: 157.5 on 6 and 198 DF,  p-value: < 2.2e-16
```

```r
model2 <- lm(CP$price ~ CP$fueltype + CP$carlength + CP$enginesize + CP$horsepower + CP$peakrpm + CP$hi
anova(model2,model1)
```

```
## Analysis of Variance Table
##
## Model 1: CP$price ~ CP$fueltype + CP$carlength + CP$enginesize + CP$horsepower +
##     CP$peakrpm + CP$highwaympg
## Model 2: CP$price ~ CP$fueltype + CP$carbody + CP$carlength + CP$enginesize +
##     CP$horsepower + CP$peakrpm + CP$highwaympg
```

```
##   Res.Df        RSS Df Sum of Sq      F   Pr(>F)
## 1    198 2254940295
## 2    194 2076673688  4 178266607 4.1634 0.002941 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the result of our p-value of 0.002, we will reject the null hypothesis which means that the model with the variable included is better.

c. **4 pts** What can you conclude from a and b? Do they provide the exact same results?
   Yes, based on our results, we've rejected the null hypothesis in both scenarios since we've concluded that the carbody has an impact on the carpice.

Question 4: Coefficient Interpretation [6 points]

a. **3 pts** Interpret the estimated coefficient of *fueltypegas* in the context of the problem. *Mention any assumption you make about other predictors clearly when stating the interpretation.*

CP$fueltypegas -4.099e+03 9.862e+02 -4.156 4.81e-05 ***

Based on the result of our model, we can see based on a p-value of 0.0000481 that is < 0.05, and we can conclude that is statistically significant for predicting the target variable price.

When converting the scientific notation, the results in a value of 99. Meaning that an increase in one unit of price accounts for a 99 unit increase in the fueltypegas variable.

b. **3 pts** If the value of the *enginesize* in the above model is increased by 0.01 keeping other predictors constant, what change in the response would be expected?

Our value of the coefficient 1.145e+02 translates to a real number of 114.5. Taking 114.5 * 0.01 = 1.45 unit increase in the price variable.

Question 5: Confidence and Prediction Intervals [10 points]

a. **5 pts** Compute 90% and 95% confidence intervals (CIs) for the parameter associated with *carlength* for the model in Question 2. What observations can you make about the width of these intervals?

```r
confint(model1,level=0.90)
```

```
##                              5 %         95 %
## (Intercept)         -37759.665158 -6935.670777
## CP$fueltypegas       -5396.056360 -2223.979529
## CP$carbodyhardtop    -5862.385996    54.815498
## CP$carbodyhatchback  -7501.621525 -2754.430895
## CP$carbodysedan      -6746.385935 -1863.429254
## CP$carbodywagon      -8177.827674 -2830.381319
## CP$carlength            30.932219   160.355997
## CP$enginesize           82.079645   124.280193
## CP$horsepower           24.167048    69.888986
## CP$peakrpm               1.033955     3.217264
## CP$highwaympg         -175.859720    51.163988
```

```r
confint(model1,level=0.99)
```

```
##                            0.5 %       99.5 %
## (Intercept)          -4.660603e+04  1910.690453
## CP$fueltypegas       -6.306430e+03 -1313.606359
## CP$carbodyhardtop    -7.560599e+03  1753.028301
## CP$carbodyhatchback -8.864046e+03 -1392.006460
## CP$carbodysedan      -8.147775e+03  -462.040513
## CP$carbodywagon      -9.712523e+03 -1295.685928
## CP$carlength         -6.211880e+00   197.500095
## CP$enginesize         6.996826e+01   136.391579
## CP$horsepower         1.104504e+01    83.010996
## CP$peakrpm            4.073544e-01     3.843865
## CP$highwaympg        -2.410146e+02   116.318870
```

Confint(90%): CP$carlength 30.932219 160.355997 Confint(99.5$carlength -6.211880e+00 197.500095

We can see that the width for the 99.5 confidence interval is obviously much wider. But more importantly, the 90 confidence interval does not include 0 within the range. If 0 is within the range then this means that the variable has the posibility of not being statistically significant. In our case it doesn't include 0 for a critical value of 0.9, thus there's a high chance it will be statistically significant.

b. **2.5 pts** Using *model1*, estimate the average price for all cars with the same characteristics as the first data point in the sample. What is the 95% confidence interval for this estimation? Provide an interpretation of your results.

```r
newdata = CP[1,]
predict(model1,newdata,interval="confidence")
```

```
## Warning: 'newdata' had 1 row but variables found have 205 rows
```

```
##           fit        lwr        upr
## 1   17565.194 14885.3674 20245.0205
## 2   17565.194 14885.3674 20245.0205
## 3   17021.224 15851.0268 18191.4221
## 4   12292.041 11455.2863 13128.7960
## 5   16188.046 14917.0962 17458.9960
## 6   15832.813 14811.9611 16853.6655
## 7   17305.733 16007.1228 18604.3423
## 8   16106.536 14508.9397 17704.1316
## 9   18512.413 17185.1322 19839.6934
## 10  17118.319 15803.0795 18433.5589
## 11  12860.993 11838.1809 13883.8046
## 12  12860.993 11838.1809 13883.8046
## 13  16347.282 14941.3649 17753.1984
## 14  16347.282 14941.3649 17753.1984
## 15  17701.183 16398.8504 19003.5163
## 16  27844.483 26582.6299 29106.3367
## 17  28303.575 27012.8520 29594.2980
## 18  28734.332 27446.4087 30022.2550
## 19  -1702.837 -3522.9827   117.3087
## 20   5000.691  3817.7401  6183.6425
```

```
## 21   6101.178  4884.0619  7318.2937
## 22   5377.794  4276.3532  6479.2342
## 23   5564.837  4579.1266  6550.5481
## 24   8488.012  7340.2216  9635.8027
## 25   5564.837  4579.1266  6550.5481
## 26   6387.956  5268.1359  7507.7760
## 27   6387.956  5268.1359  7507.7760
## 28   9311.131  7766.9215 10855.3400
## 29  10521.698  9149.5148 11893.8811
## 30  17326.676 16239.9964 18413.3554
## 31   1600.744  -506.6738  3708.1623
## 32   5995.546  4383.8921  7607.1996
## 33   3106.041  1982.4189  4229.6623
## 34   6761.416  5211.1419  8311.6891
## 35   6761.416  5211.1419  8311.6891
## 36   8866.165  7593.3607 10138.9696
## 37   7064.410  5091.8589  9036.9618
## 38  10399.932  9182.8881 11616.9761
## 39  10399.932  9182.8881 11616.9761
## 40  11978.639 10846.9794 13110.2989
## 41  11978.639 10846.9794 13110.2989
## 42  12995.799 11956.9663 14034.6311
## 43  11521.486 10663.0182 12379.9545
## 44   9379.850  8221.7360 10537.9631
## 45   5823.810  4569.4636  7078.1563
## 46   5823.810  4569.4636  7078.1563
## 47  10553.352  9584.5794 11522.1250
## 48  32437.356 30651.3161 34223.3962
## 49  32437.356 30651.3161 34223.3962
## 50  43398.510 40584.7746 46212.2448
## 51   5213.807  3985.8827  6441.7311
## 52   4777.372  3857.8795  5696.8642
## 53   4777.372  3857.8795  5696.8642
## 54   6336.950  5392.4041  7281.4961
## 55   6336.950  5392.4041  7281.4961
## 56   8170.222  6447.7009  9892.7436
## 57   8170.222  6447.7009  9892.7436
## 58   8170.222  6447.7009  9892.7436
## 59  10800.974  9068.9444 12533.0037
## 60  10465.908  9266.2485 11665.5669
## 61  11289.026 10410.3589 12167.6938
## 62  10465.908  9266.2485 11665.5669
## 63  11289.026 10410.3589 12167.6938
## 64  13216.164 11584.2770 14848.0507
## 65  10465.908  9266.2485 11665.5669
## 66  15308.331 14388.6313 16228.0311
## 67  13793.263 12236.6877 15349.8379
## 68  23959.961 22268.0825 25651.8386
## 69  22760.764 20683.4827 24838.0446
## 70  25035.893 22281.6171 27790.1687
## 71  25078.997 23361.5444 26796.4489
## 72  29322.730 27665.3800 30980.0798
## 73  31494.774 28544.6303 34444.9175
## 74  38441.192 35840.7081 41041.6766
```

```
## 75   38578.362 35504.2126 41652.5123
## 76   17583.987 15931.7994 19236.1747
## 77    5584.154  4467.6939  6700.6131
## 78    5771.197  4767.7532  6774.6411
## 79    5771.197  4767.7532  6774.6411
## 80    8488.012  7340.2216  9635.8027
## 81   11886.176 10911.7532 12860.5986
## 82   10620.050  9650.3155 11589.7844
## 83   17326.676 16239.9964 18413.3554
## 84   17326.676 16239.9964 18413.3554
## 85   17326.676 16239.9964 18413.3554
## 86   11385.782 10594.6952 12176.8690
## 87   11385.782 10594.6952 12176.8690
## 88   12651.908 11727.7101 13576.1060
## 89   12651.908 11727.7101 13576.1060
## 90    7347.061  6472.4163  8221.7062
## 91    9457.000  7493.5755 11420.4252
## 92    7347.061  6472.4163  8221.7062
## 93    7347.061  6472.4163  8221.7062
## 94    6616.520  5155.6883  8077.3526
## 95    7347.061  6472.4163  8221.7062
## 96    6552.636  5568.8281  7536.4436
## 97    7347.061  6472.4163  8221.7062
## 98    6616.520  5155.6883  8077.3526
## 99    8470.816  5984.3618 10957.2696
## 100 11175.626 10140.5662 12210.6860
## 101 11998.745 11222.8346 12774.6548
## 102 22421.281 21333.0046 23509.5577
## 103 21499.452 19930.3286 23068.5757
## 104 22511.605 21569.7574 23453.4536
## 105 20735.258 19465.7245 22004.7912
## 106 22741.074 21014.7548 24467.3938
## 107 21481.282 20375.5756 22586.9883
## 108 13469.168 12322.8137 14615.5225
## 109 18118.989 16603.8074 19634.1696
## 110 13436.829 11669.3920 15204.2666
## 111 18585.433 16573.8782 20596.9872
## 112 13375.112 12215.3225 14534.9017
## 113 18118.989 16603.8074 19634.1696
## 114 13342.773 11570.3856 15115.1609
## 115 18585.433 16573.8782 20596.9872
## 116 13469.168 12322.8137 14615.5225
## 117 18118.989 16603.8074 19634.1696
## 118 18305.313 17213.4639 19397.1629
## 119  5377.794  4276.3532  6479.2342
## 120  8488.012  7340.2216  9635.8027
## 121  5564.837  4579.1266  6550.5481
## 122  7344.397  6369.6657  8319.1283
## 123  8169.836  7165.1559  9174.5168
## 124 10521.698  9149.5148 11893.8811
## 125 17326.676 16239.9964 18413.3554
## 126 17181.212 16150.5668 18211.8568
## 127 27826.922 25139.6013 30514.2424
## 128 27826.922 25139.6013 30514.2424
```

```
## 129 30730.707 27705.2411 33756.1731
## 130 30485.064 27000.4503 33969.6785
## 131 12457.711 11045.2068 13870.2145
## 132 12384.262 11239.8659 13528.6573
## 133 13633.040 12170.8211 15095.2591
## 134 14456.159 13428.2535 15484.0640
## 135 13633.040 12170.8211 15095.2591
## 136 14456.159 13428.2535 15484.0640
## 137 16640.539 14835.1199 18445.9582
## 138 17463.658 15820.6206 19106.6948
## 139  5145.197  4120.2301  6170.1641
## 140  5812.867  4206.3625  7419.3715
## 141  5755.481  4121.7009  7389.2601
## 142  8883.976  7862.4955  9905.4570
## 143  8283.124  7004.0691  9562.1788
## 144 10610.296  9841.6918 11378.8995
## 145  9632.151  8153.1404 11111.1610
## 146 10746.572  9458.9785 12034.1648
## 147  8139.985  6735.8974  9544.0723
## 148  9616.913  8264.4906 10969.3349
## 149  8336.593  6878.5264  9794.6594
## 150 10074.493  8192.8746 11956.1105
## 151  4072.656  3036.6333  5108.6793
## 152  4135.004  3111.0755  5158.9327
## 153  4135.004  3111.0755  5158.9327
## 154  4873.359  3385.9140  6360.8038
## 155  5185.098  3690.8770  6679.3195
## 156  5185.098  3690.8770  6679.3195
## 157  6742.669  5740.4373  7744.9016
## 158  5919.551  4881.3236  6957.7781
## 159 10557.119  8724.6094 12389.6291
## 160  9048.174  7082.8170 11013.5311
## 161  6119.191  4361.2280  7877.1536
## 162  6106.594  5109.4662  7103.7227
## 163  6929.713  5883.6377  7975.7885
## 164  7159.259  6171.9038  8146.6140
## 165  6336.140  5302.2397  7370.0409
## 166 13272.272 11575.3481 14969.1961
## 167 12449.153 10692.7959 14205.5111
## 168 16643.028 14256.4859 19029.5710
## 169 16643.028 14256.4859 19029.5710
## 170 14418.787 13429.1815 15408.3934
## 171 16643.028 14256.4859 19029.5710
## 172 14418.787 13429.1815 15408.3934
## 173 19546.814 16762.8518 22330.7756
## 174 10054.772  8618.0745 11491.4694
## 175 12433.129 10777.4631 14088.7955
## 176  9356.349  7896.4464 10816.2518
## 177 10179.468  8763.3486 11595.5868
## 178  9356.349  7896.4464 10816.2518
## 179 21037.079 19924.7128 22149.4455
## 180 21037.079 19924.7128 22149.4455
## 181 22036.327 21102.3496 22970.3052
## 182 19805.331 18286.2084 21324.4541
```

```
## 183  9558.351  7806.3916 11310.3096
## 184 10243.115  9481.9423 11004.2874
## 185  9558.351  7806.3916 11310.3096
## 186 10243.115  9481.9423 11004.2874
## 187 10243.115  9481.9423 11004.2874
## 188  9922.507  8277.0987 11567.9162
## 189 11604.633 10804.3144 12404.9521
## 190 14440.317 11694.3435 17186.2912
## 191  9924.413  8879.2234 10969.6034
## 192 16172.529 15126.7795 17218.2786
## 193 10984.874  9366.2614 12603.4863
## 194 10993.791  9476.8927 12510.6890
## 195 17237.128 16150.5634 18323.6919
## 196 16037.931 14517.6877 17558.1738
## 197 17237.128 16150.5634 18323.6919
## 198 16037.931 14517.6877 17558.1738
## 199 18095.898 16466.2571 19725.5383
## 200 16896.701 14992.4775 18800.9240
## 201 17237.128 16150.5634 18323.6919
## 202 19374.899 18018.0584 20731.7397
## 203 22003.746 20876.9786 23130.5128
## 204 19870.623 18215.5690 21525.6776
## 205 17424.171 16392.6513 18455.6912
```

1 17565.194 14885.3674 20245.0205

This is the result of the prediction interval meaning the following:

fitted value = 17565.19 lower bound = 14885.37 upper bound = 20245.02

c. **2.5 pts** Suppose that the *carlenght* value for the first data point is increased to 200, while all other values are kept fixed. Using *model1*, predict the price of a car with these characteristics. What is the 95% prediction interval for this prediction? Provide an interpretation of your results.

```r
summary(model1)
```

```
##
## Call:
## lm(formula = CP$price ~ CP$fueltype + CP$carbody + CP$carlength +
##     CP$enginesize + CP$horsepower + CP$peakrpm + CP$highwaympg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9012.6 -1848.1   -48.1  1658.0 13011.4
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.235e+04  9.325e+03  -2.397 0.017502 *
## CP$fueltypegas      -3.810e+03  9.596e+02  -3.970 0.000101 ***
## CP$carbodyhardtop   -2.904e+03  1.790e+03  -1.622 0.106401
## CP$carbodyhatchback -5.128e+03  1.436e+03  -3.571 0.000449 ***
## CP$carbodysedan     -4.305e+03  1.477e+03  -2.914 0.003985 **
## CP$carbodywagon     -5.504e+03  1.618e+03  -3.402 0.000811 ***
## CP$carlength         9.564e+01  3.915e+01   2.443 0.015471 *
```

```
## CP$enginesize        1.032e+02  1.277e+01   8.082 6.69e-14 ***
## CP$horsepower        4.703e+01  1.383e+01   3.400 0.000818 ***
## CP$peakrpm           2.126e+00  6.605e-01   3.218 0.001512 **
## CP$highwaympg       -6.235e+01  6.868e+01  -0.908 0.365114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3272 on 194 degrees of freedom
## Multiple R-squared:  0.8405, Adjusted R-squared:  0.8323
## F-statistic: 102.2 on 10 and 194 DF,  p-value: < 2.2e-16
```

Since the fitted value for the regression line of the first observation is 17565.19, and the new coefficient value for the carlength variable is 295.64. We can predict the impact of this variable. With 95.64 as the original coefficient value, 168.8 * 95.64 = 16144.03 is the original value for the coefficient B1 * X11. Now with the new coefficient value, the total is 200 * 95.64 = 19128. Therefore the impact on the target variable is 2983.97 $ in the target variable price.

```
CP[1,]
```

```
##   fueltype     carbody carlength enginesize horsepower peakrpm highwaympg price
## 1      gas convertible     168.8        130        111    5000         27 13495
```