Project Continuation:

Before running the two-way ANOVA tests we check for interaction between the variables for the two cases, one including the age of the books and the other including the Price classification (i.e. expensive)

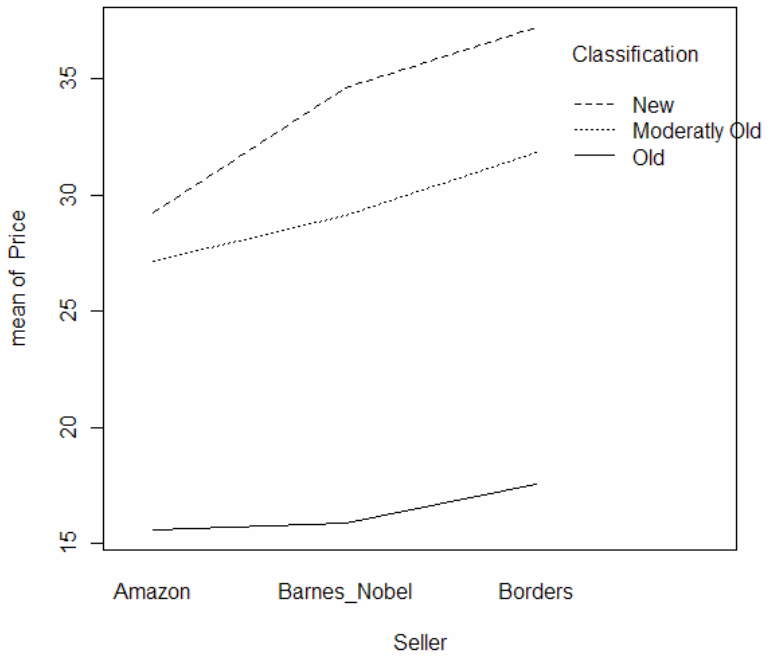
- We first check to make sure that both variables are factors or in other words categorical.

```
> getwd()
[1] "D:/Documents and Settings/Class2009/My Documents"
> setwd("D:/Documents and Settings/Class2009/My Documents/Fall08/MA331")
> Data <- read.csv("Project2.csv", header = TRUE)
> attach(Data)
> is.factor(Classification)
[1] TRUE
> is.factor(Seller)
[1] TRUE
```

- Next we take a look at the **Interaction Plot**. For **Seller/Vendor and Age Classification**

  Note: in the below plot formula, the first two arguments are the cathegorical variables and the last is the quantitative response

```
> interaction.plot(Seller, Classification, Price)
```

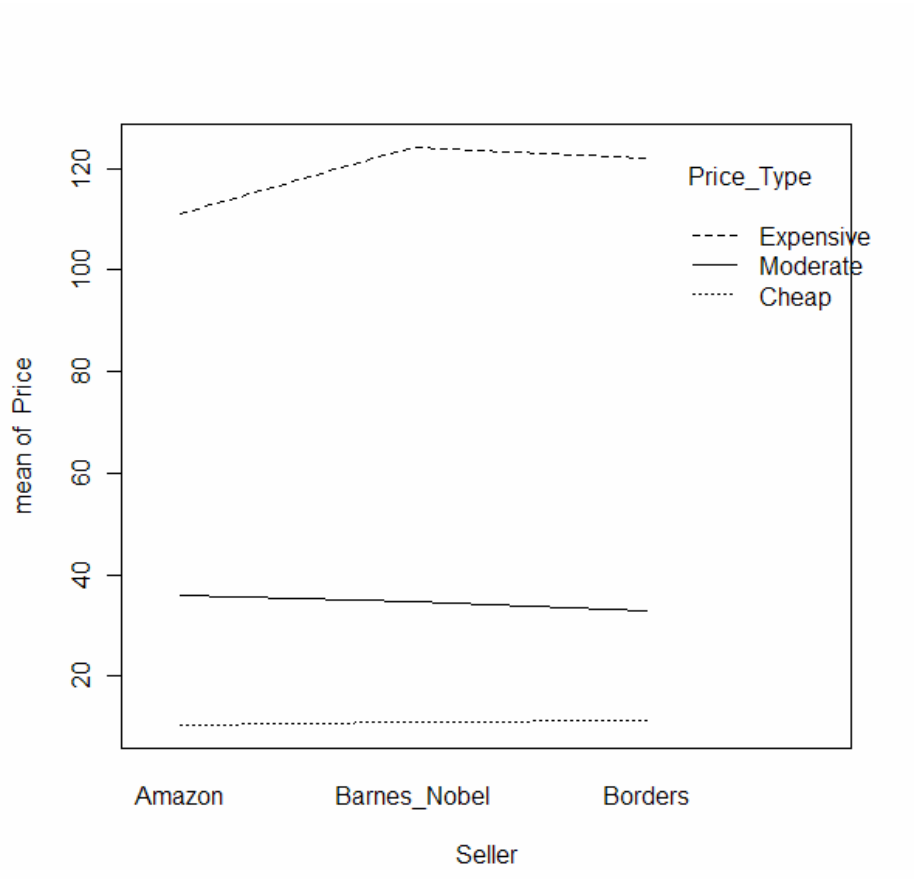


From the graph, we can gather that,

- There is little to no interaction between the Seller and the Age Classification of the books.
- We see that the Age appears to have some significant influence on the mean of Price
- As was found in the earlier studies, the average price of Borders books is a bit higher than that of Barnes and Nobel's, which is in tern a bit higher than Amazon.

- Next we take a look at the **Interaction Plot**. For **Seller/Vendor and Price Type** (i.e. expensive)

```
> interaction.plot(Seller, Price_Type, Price)
```

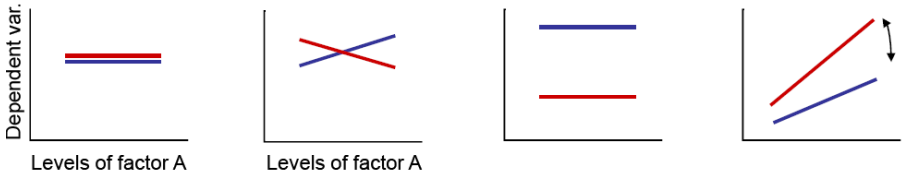


From the graph, we can gather that,

- There is little to no interaction between the Seller and the Price type of the books.
- We see that the mean for the price of Expensive Amazon books seems significantly lower than that of both Barnes and Nobel and Borders.

The analysis of the graphs was based off of the following observations from the lecture.

Neither factor is significant	Neither factor is significant	Only 1 factor is significant	Both factors are significant
No interaction	Interaction effect is significant	No interaction	With or without significant interaction



- Verification of the previous observations

The below summary clearly indicates that there is not interaction between seller and age of books, with effect on the price. **This means that for that with regards to these two variables a one-way analysis is sufficient.**

Analysis of Variance Table:

```
> anova(lm(Price~Seller+Classification+Seller:Classification))
Analysis of Variance Table

Response: Price
                     Df Sum Sq Mean Sq F value Pr(>F)
Seller                2    664     332  0.2164 0.8057
Classification        2   6635    3317  2.1625 0.1188
Seller:Classification 4    161      40  0.0262 0.9987
Residuals           141 216302    1534
```

```
> summary(lm(Price~Seller+Classification+Seller:Classification))

Call:
lm(formula = Price ~ Seller + Classification + Seller:Classification)

Residuals:
    Min      1Q  Median      3Q     Max
-30.197 -21.201 -11.828  -1.738 172.846

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           27.154      8.547   3.177  0.00183 **
SellerBarnes_Nobel                     1.969     12.087   0.163  0.87086
SellerBorders                          4.669     12.087   0.386  0.69990
ClassificationNew                      2.037     12.778   0.159  0.87358
ClassificationOld                    -11.554     14.174  -0.815  0.41633
SellerBarnes_Nobel:ClassificationNew   3.460     18.071   0.191  0.84843
SellerBorders:ClassificationNew        3.327     18.071   0.184  0.85418
SellerBarnes_Nobel:ClassificationOld  -1.704     20.044  -0.085  0.93236
SellerBorders:ClassificationOld       -2.708     20.044  -0.135  0.89274
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.17 on 141 degrees of freedom
Multiple R-squared: 0.03334,    Adjusted R-squared: -0.02151
F-statistic: 0.6078 on 8 and 141 DF,  p-value: 0.7702
```

According to the below summary there seems to be interaction between seller and Price_Type, with effect on the price. This is indicated by the strong P-value in the "Seller:Price_Type" row of the ANOVA results. However, by observing the summary of the lm we find that the "Price_TypeExpensive" and Price_TypeModerate" rows are the two rows having a strong effect. This is expected as a Price categorical value for price should follow the general trend of the mean prices for the books.

Analysis of Variance Table:

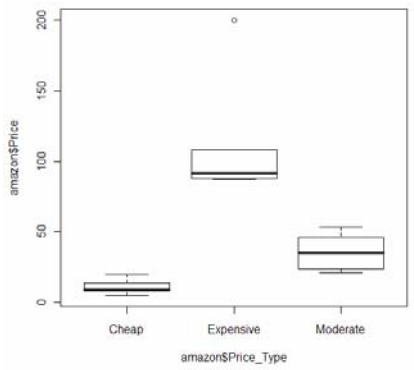
```
> anova(lm(Price~Seller+Price_Type+Seller:Price_Type))
Analysis of Variance Table

Response: Price
                    Df Sum Sq Mean Sq  F value Pr(>F)
Seller               2    664     332   1.4885 0.2292
Price_Type           2 191097   95548 428.4367 <2e-16 ***
Seller:Price_Type    4    556     139   0.6229 0.6469
Residuals          141  31445     223
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(lm(Price~Seller+Price_Type+Seller:Price_Type))

Call:
lm(formula = Price ~ Seller + Price_Type + Seller:Price_Type)

Residuals:
    Min      1Q  Median      3Q     Max
-37.177  -3.891  -2.486   3.121  89.003

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                           10.4762     2.3913   4.381 2.29e-05 ***
SellerBarnes_Nobel                     0.4049     3.4272   0.118 0.906117
SellerBorders                          0.7601     3.4771   0.219 0.827269
Price_TypeExpensive                  100.5205     6.5489  15.349  < 2e-16 ***
Price_TypeModerate                    25.3518     7.0938   3.574 0.000482 ***
SellerBarnes_Nobel:Price_TypeExpensive 12.7301     9.2782   1.372 0.172228
SellerBorders:Price_TypeExpensive      10.4203     9.0066   1.157 0.249242
SellerBarnes_Nobel:Price_TypeModerate  -1.5044     9.3919  -0.160 0.872972
SellerBorders:Price_TypeModerate       -3.7531     9.1962  -0.408 0.683807
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.93 on 141 degrees of freedom
Multiple R-squared: 0.8595,     Adjusted R-squared: 0.8515
F-statistic: 107.8 on 8 and 141 DF,  p-value: < 2.2e-16
```

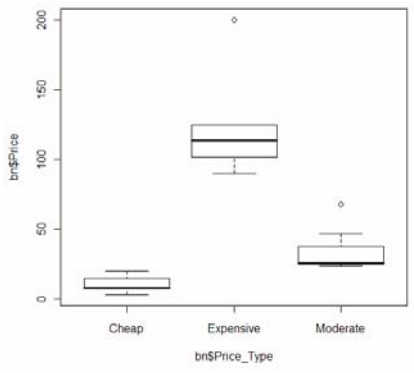
Do to the findings in the summary for he interaction between vendor and price type we will look at the individual graphs of the price~price type for each vendor. We specifically need to pay attention to the behavior for Expensive and moderate books.

```
> plot(bn$Price~bn$Price_Type)
> plot(amazon$Price~amazon$Price Type)
> plot(borders$Price~borders$Price_Type)
```

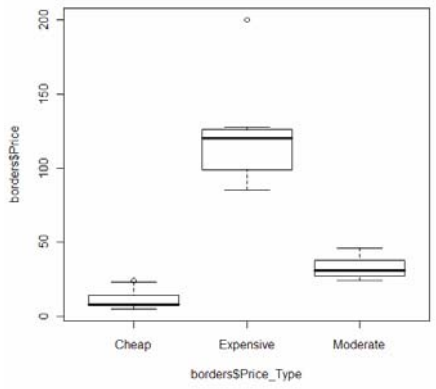
Amazon: Price~Price_Type



Barns & Nobel: Price~Price_Type



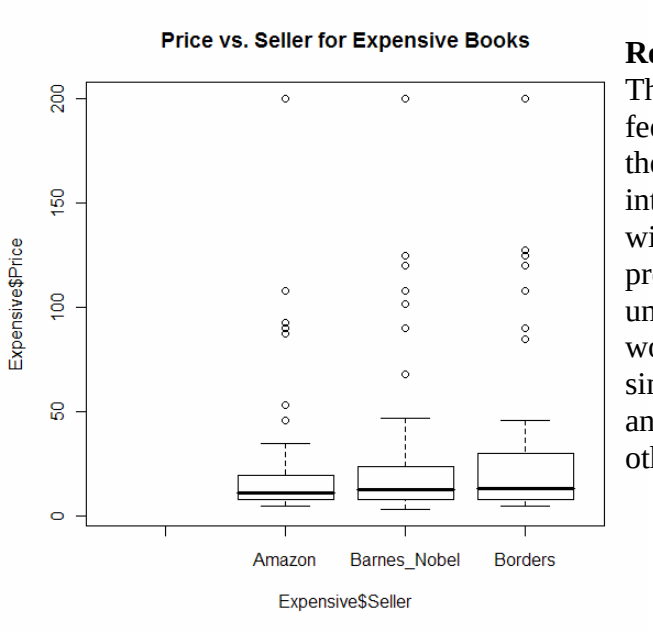
Borders:
Price~Price_Type



From these Graphs & the Graphs that follow:
It is evident that the behavior of the three vendors varies in the case of books which are classified as either Expensive or Mederate. A buyer can expect to be much more likely to pay more for Expensive books if they are buying from Borders. A middle ground lies in Barnes & Nobel and a much lower range with Amazon.

In contrast to my earlier findings without the use of the two-way ANOVA, it seems from this data that my large amount of "Cheap" books was masking the variance that could be found in the smaller data sets of Expensive and Moderate books.

ExpensiveBooks:



Reflection:

This project has been extremely useful; I feel that I have learned a great deal about the overall process of statistics. It was interesting to pick a problem and just go with it. I was happy to have chosen a project in which I could gain an understanding of ANOVA as apposed to working on a problem where I would simply be using the same regression analysis, which was the case for many other groups.

ModerateBooks: