ECON 390 SID 2

Question 1: Reading in Data

Read in Sales Data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
Sales = read.csv('sales_data.csv')
```

Summary statistics for price variable

```
summary(Sales$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3959  1.1441  1.4580  1.5623  1.8532 11.9022
```

Mean of price per store

```
pricePerStore = aggregate(Sales$price, list(Sales$store), FUN = mean)
```

Question 2: Creating Product Cross-Section

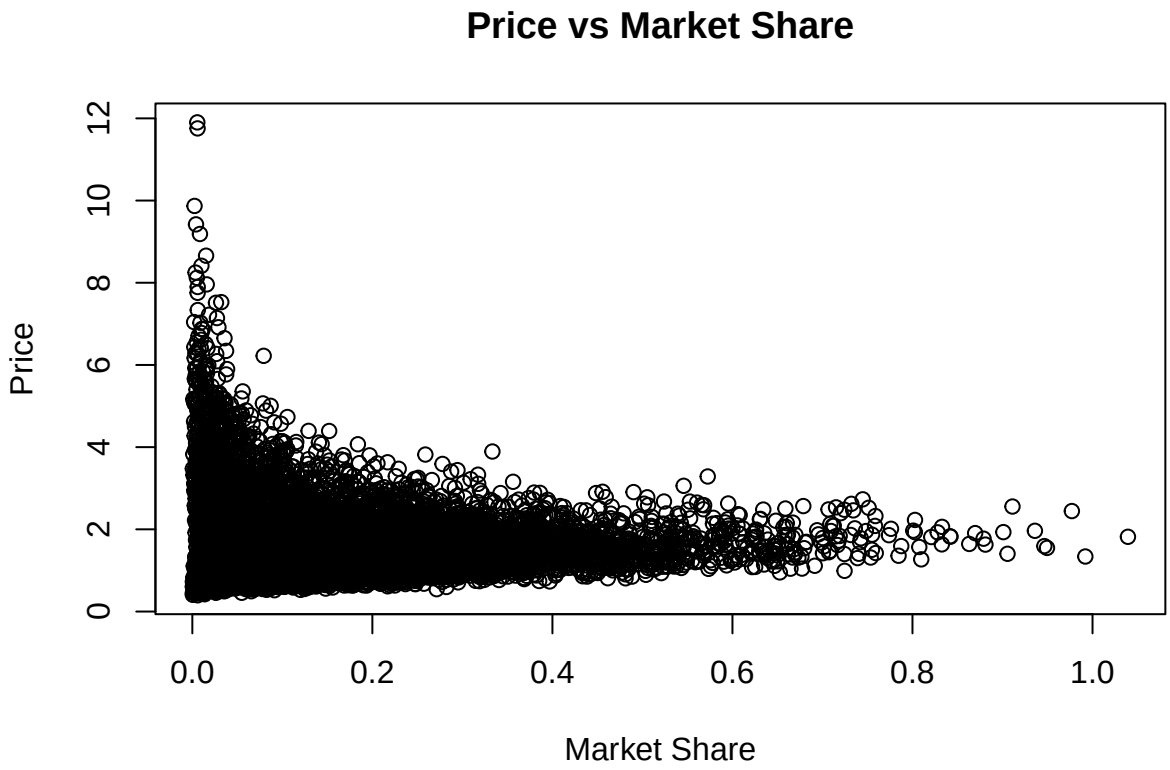
Cross Section df

```r
# Sjm = Qjm / Nm
# Qjm = total quantity sold of product j in market M (store)
# Nm = total size of market (store), surrounding area of each store
Population = read.csv('population_data.csv')
Sales =  merge(Population, Sales, by=c("store"), all = TRUE)
PricePerProduct = aggregate(Sales$price, list(Sales$product, Sales$store), FUN = mean)
SalesPerProduct = aggregate(Sales$quantity, list(Sales$product, Sales$store), FUN = sum)
PopulationPerStore = aggregate(Sales$pop_size, list(Sales$product, Sales$store), FUN = mean)
MarketSharePerStoreProduct = data.frame(PricePerProduct[2], PricePerProduct[1], SalesPerProduct[3], Popu
MarketSharePerStoreProduct = data.frame(MarketSharePerStoreProduct[1], MarketSharePerStoreProduct[2], Ma
colnames(MarketSharePerStoreProduct) = c('store','product','sales','pop_size', 'share', 'price')
```

Question 3: Identify Supply and Demand

Scatter Plot of Price vs Quantity

```r
plot(MarketSharePerStoreProduct$share, MarketSharePerStoreProduct$price, ylab = "Price", xlab = "Market
```



There seems to be an positive weak correlation between Price and Market Share. The law of supply and demand may hold

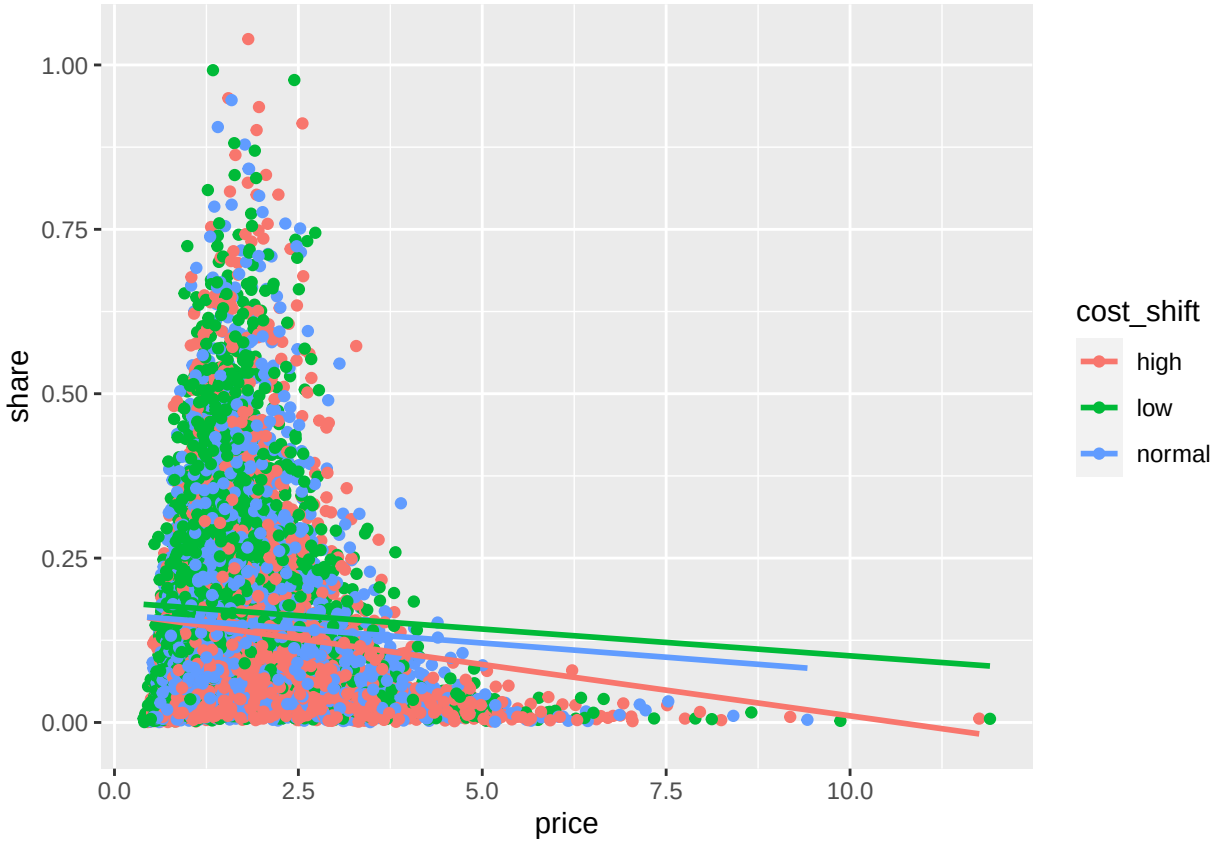
Merged Cost Data

```
Costs = read.csv("cost_data.csv")
MarketSharePerStoreProduct = data.frame(MarketSharePerStoreProduct[1], MarketSharePerStoreProduct[2], Ma
colnames(MarketSharePerStoreProduct) = c('store','product','sales','pop_size', 'share', 'price','cost_sh
```

ggplot

```
library(ggplot2)
ggplot(MarketSharePerStoreProduct, aes(y = share, x = price, color = cost_shift) ) + geom_point() + geom
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Question 4: Price Aggregation

```
eps = rnorm(NROW(Sales),0, 0.05^2)

#Repeated from 2 and 3
Population = read.csv('population_data.csv')
```

```
Sales = read.csv('sales_data.csv')

# Acounting for EPS
SalesEps =  merge(Population, Sales, by=c("store"), all = TRUE)
SalesEps$price =  Sales$price + eps
PricePerProduct = aggregate(SalesEps$price, list(SalesEps$product, SalesEps$store), FUN = mean)

#Repeated from 2 and 3
SalesPerProduct = aggregate(SalesEps$quantity, list(SalesEps$product, SalesEps$store), FUN = sum)
PopulationPerStore = aggregate(SalesEps$pop_size, list(SalesEps$product, SalesEps$store), FUN = mean)
MarketSharePerStoreProduct = data.frame(PricePerProduct[2], PricePerProduct[1], SalesPerProduct[3], Popu
MarketSharePerStoreProduct = data.frame(MarketSharePerStoreProduct[1], MarketSharePerStoreProduct[2], Ma
colnames(MarketSharePerStoreProduct) = c('store','product','sales','pop_size', 'share', 'price')
MarketSharePerStoreProduct = data.frame(MarketSharePerStoreProduct[1], MarketSharePerStoreProduct[2], Ma
colnames(MarketSharePerStoreProduct) = c('store','product','sales','pop_size', 'share', 'price','cost_sh
```

Perhaps an alternative appraoch would be to take the median or mode to get rid of outliers

Question 5: Basic Descriptive Analysis

Price Breakdown

```
MarketSharePerStoreProduct %>% group_by(cost_shift) %>% summarise(price = mean(price))
```

```
## # A tibble: 3 x 2
##   cost_shift price
##   <chr>      <dbl>
## 1 high        1.72
## 2 low         1.54
## 3 normal      1.62
```

As expected as prices are higher the associated average price seems to be higher as well

Regression

```
model = glm(price ~ factor(cost_shift), data = MarketSharePerStoreProduct)
summary(model)
```

```
##
## Call:
## glm(formula = price ~ factor(cost_shift), data = MarketSharePerStoreProduct)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2662  -0.5766  -0.2097   0.3139  10.3598
##
## Coefficients:
```

```
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.72067    0.01409 122.088  < 2e-16 ***
## factor(cost_shift)low      -0.17806    0.01993  -8.935  < 2e-16 ***
## factor(cost_shift)normal -0.10343    0.02005  -5.159 2.53e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.7748658)
##
##     Null deviance: 9060.1  on 11614  degrees of freedom
## Residual deviance: 8997.7  on 11612  degrees of freedom
## AIC: 30004
##
## Number of Fisher Scoring iterations: 2
```

The intercept seems to be about the same as the high cost average. Again this is similar to our results from before, as subtracting coefficients will give low and normal average costs. As production cost is more so is the general price