Lab2

2023-04-16

R Markdown

```
##  Item_Identifier    Item_Weight     Item_Fat_Content  Item_Visibility
## Length:14204       Min.   : 4.555  Length:14204      Min.   :0.00000
## Class :character   1st Qu.: 8.710  Class :character  1st Qu.:0.02704
## Mode  :character   Median :12.600  Mode  :character  Median :0.05402
##                    Mean   :12.793                    Mean   :0.06595
##                    3rd Qu.:16.750                    3rd Qu.:0.09404
##                    Max.   :21.350                    Max.   :0.32839
##                    NA's   :2439
##   Item_Type          Item_MRP       Outlet_Identifier
## Length:14204       Min.   : 31.29  Length:14204
## Class :character   1st Qu.: 94.01  Class :character
## Mode  :character   Median :142.25  Mode  :character
##                    Mean   :141.00
##                    3rd Qu.:185.86
##                    Max.   :266.89
##
## Outlet_Establishment_Year Outlet_Size      Outlet_Location_Type
## Min.   :1985              Length:14204     Length:14204
## 1st Qu.:1987              Class :character Class :character
## Median :1999              Mode  :character Mode  :character
## Mean   :1998
## 3rd Qu.:2004
## Max.   :2009
##
## Outlet_Type        Item_Outlet_Sales
## Length:14204       Min.   :    33.29
## Class :character   1st Qu.:  878.86
## Mode  :character   Median : 1828.27
##                    Mean   : 2099.33
##                    3rd Qu.: 2949.30
##                    Max.   :13086.97
##


##  Item_Identifier    Item_Weight     Item_Fat_Content  Item_Visibility
## Length:13325       Min.   : 4.555  Length:13325      Min.   :0.003575
## Class :character   1st Qu.: 8.710  Class :character  1st Qu.:0.031396
## Mode  :character   Median :12.600  Mode  :character  Median :0.057851
##                    Mean   :12.807                    Mean   :0.070303
##                    3rd Qu.:16.750                    3rd Qu.:0.097618
##                    Max.   :21.350                    Max.   :0.328391
##   Item_Type          Item_MRP       Outlet_Identifier
## Length:13325       Min.   : 31.29  Length:13325
```

```
## Class :character   1st Qu.: 94.14   Class :character
## Mode  :character   Median :142.28   Mode  :character
##                    Mean   :141.11
##                    3rd Qu.:185.86
##                    Max.   :266.89
## Outlet_Establishment_Year Outlet_Size      Outlet_Location_Type
## Min.   :1985              Length:13325     Length:13325
## 1st Qu.:1987              Class :character  Class :character
## Median :1999              Mode  :character  Mode  :character
## Mean   :1998
## 3rd Qu.:2004
## Max.   :2009
## Outlet_Type       Item_Outlet_Sales
## Length:13325      Min.   :   33.29
## Class :character  1st Qu.:  875.53
## Mode  :character  Median : 1821.63
##                   Mean   : 2096.48
##                   3rd Qu.: 2942.84
##                   Max.   :13086.97


##          Group.1        x
## 1     Grocery Store 154.7059
## 2 Supermarket Type1 147.3417
## 3 Supermarket Type2 146.5251
## 4 Supermarket Type3 148.8826


##          Group.1        x
## 1     Grocery Store 141.2198
## 2 Supermarket Type1 144.1325
## 3 Supermarket Type2 144.4011
## 4 Supermarket Type3 143.4028


##          Group.1        x
## 1     Grocery Store 140.2424
## 2 Supermarket Type1 137.9834
## 3 Supermarket Type2 136.9866
## 4 Supermarket Type3 138.5223


##          Group.1        x
## 1     Grocery Store 148.0403
## 2 Supermarket Type1 146.6913
## 3 Supermarket Type2 145.8619
## 4 Supermarket Type3 145.7730
```
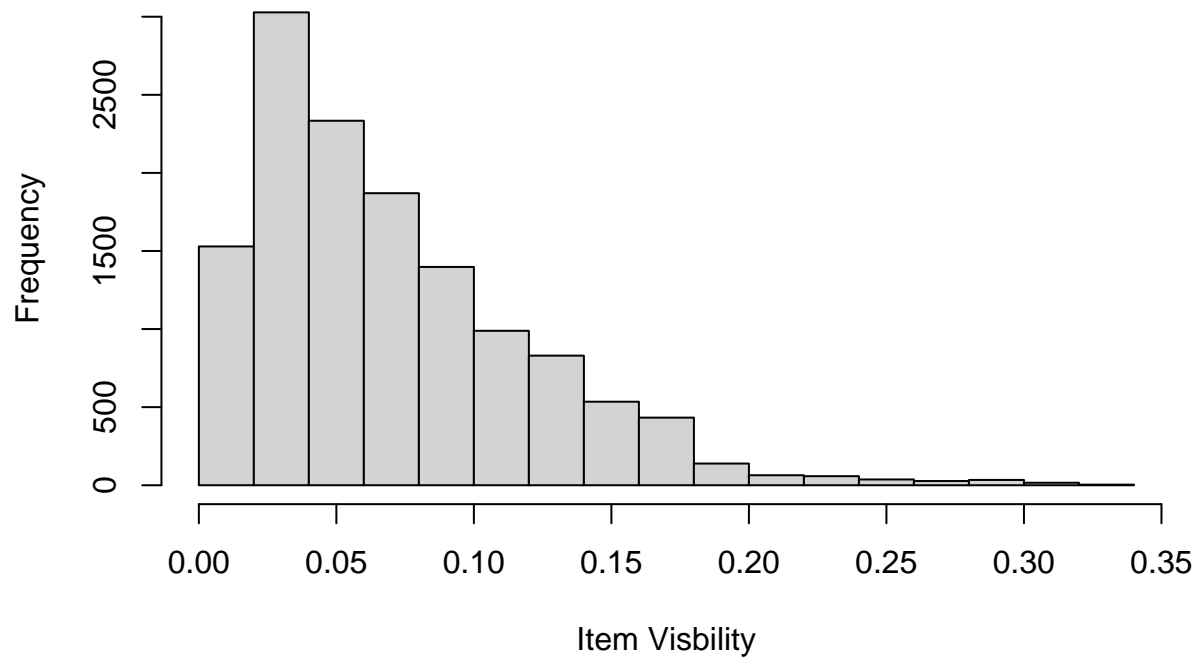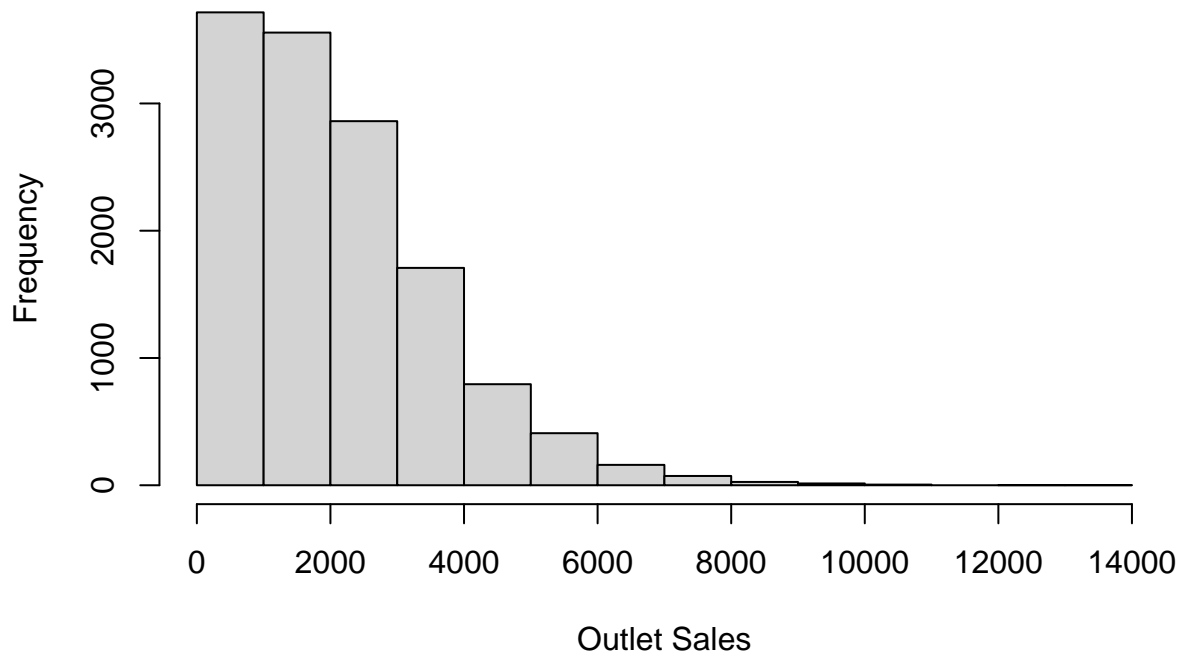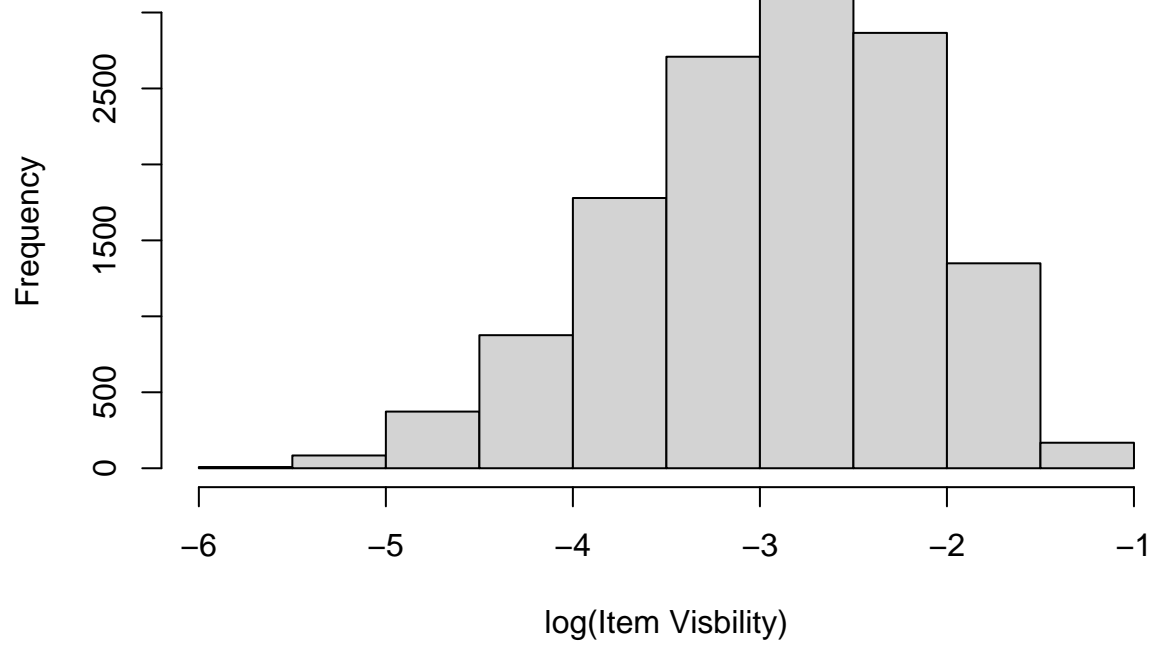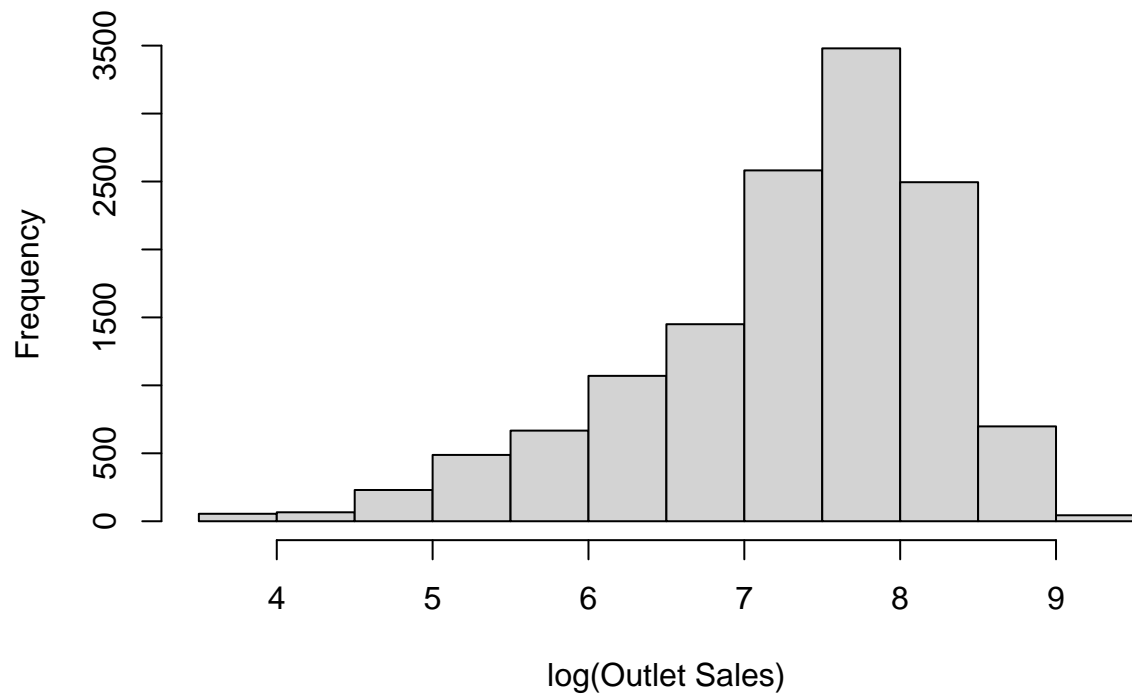
Histogram of Item Visibility

Item Visbility

Histogram of Outlet Sales

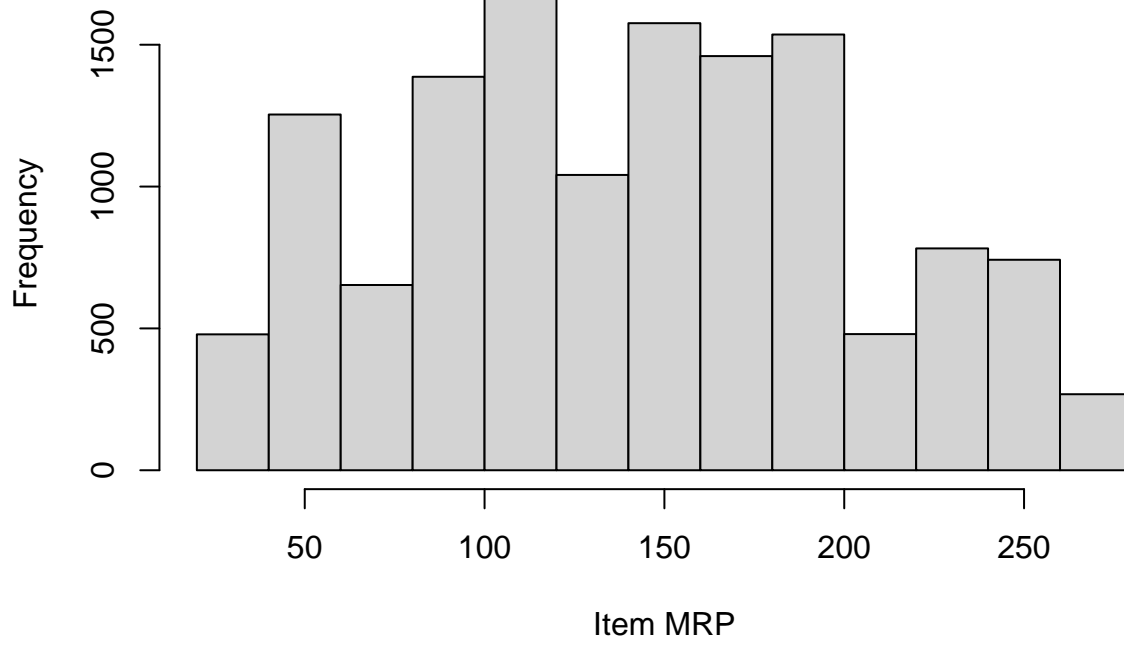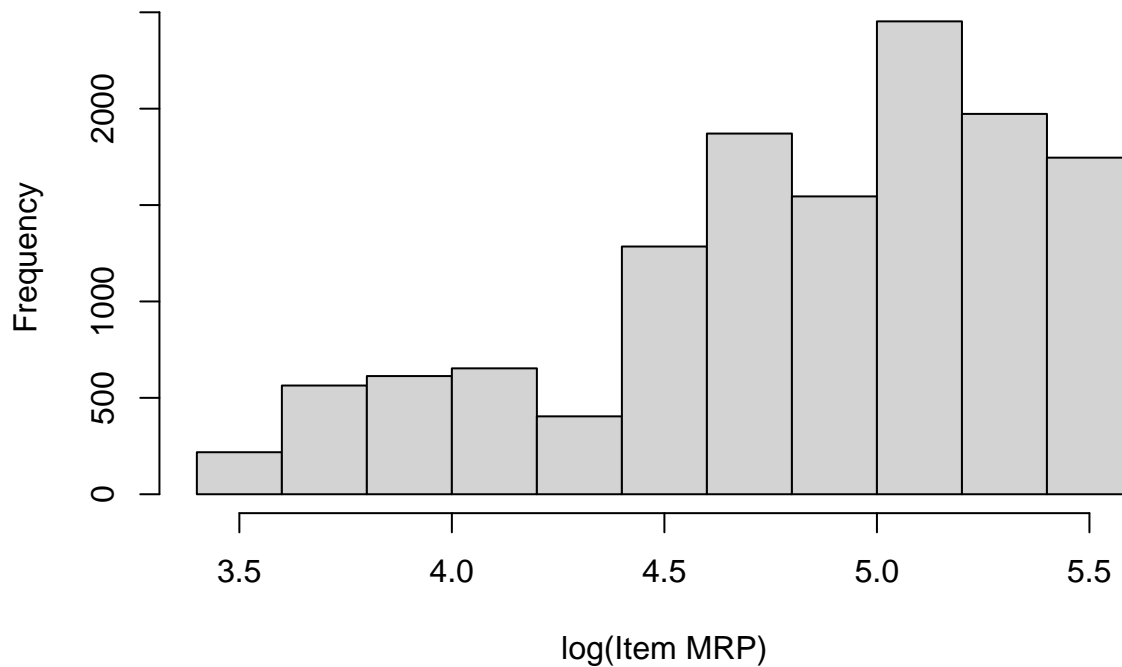# Histogram of Log Item Visibility

Histogram of Log Outlet Sales

Histogram of Item MRP

Histogram of Log Item MRP


```
##   Item_Identifier     Item_Weight      Item_Fat_Content   Item_Visibility
## Length:3998         Min.   : 4.555    Length:3998        Min.   :0.003598
## Class :character    1st Qu.: 8.645    Class :character   1st Qu.:0.031431
## Mode  :character    Median :12.600    Mode  :character   Median :0.057367
##                     Mean   :12.813                       Mean   :0.070460
##                     3rd Qu.:16.750                       3rd Qu.:0.097753
##                     Max.   :21.350                       Max.   :0.325781
##   Item_Type            Item_MRP       Outlet_Identifier
## Length:3998         Min.   : 31.29   Length:3998
## Class :character    1st Qu.: 93.81   Class :character
## Mode  :character    Median :141.91   Mode  :character
##                     Mean   :140.98
##                     3rd Qu.:186.58
##                     Max.   :266.69
## Outlet_Establishment_Year Outlet_Size     Outlet_Location_Type
## Min.   :1985              Length:3998     Length:3998
## 1st Qu.:1987              Class :character Class :character
## Median :1999              Mode  :character Mode  :character
## Mean   :1998
## 3rd Qu.:2004
## Max.   :2009
## Outlet_Type         Item_Outlet_Sales
## Length:3998         Min.   :   33.96
## Class :character    1st Qu.:  867.48
## Mode  :character    Median : 1824.82
##                     Mean   : 2102.60
```

```
##                          3rd Qu.: 2965.59
##                          Max.   :10306.58


##   Item_Identifier      Item_Weight      Item_Fat_Content   Item_Visibility
## Length:9327         Min.   : 4.555   Length:9327         Min.   :0.003575
## Class :character    1st Qu.: 8.770   Class :character    1st Qu.:0.031388
## Mode  :character    Median :12.500   Mode  :character    Median :0.057993
##                     Mean   :12.805                       Mean   :0.070236
##                     3rd Qu.:16.750                       3rd Qu.:0.097605
##                     Max.   :21.350                       Max.   :0.328391
##   Item_Type            Item_MRP        Outlet_Identifier
## Length:9327         Min.   : 31.89   Length:9327
## Class :character    1st Qu.: 94.34   Class :character
## Mode  :character    Median :142.42   Mode  :character
##                     Mean   :141.16
##                     3rd Qu.:185.66
##                     Max.   :266.89
## Outlet_Establishment_Year Outlet_Size       Outlet_Location_Type
## Min.   :1985              Length:9327       Length:9327
## 1st Qu.:1987              Class :character  Class :character
## Median :1999              Mode  :character  Mode  :character
## Mean   :1998
## 3rd Qu.:2004
## Max.   :2009
## Outlet_Type         Item_Outlet_Sales
## Length:9327         Min.   :   33.29
## Class :character    1st Qu.:  884.18
## Mode  :character    Median : 1821.63
##                     Mean   : 2093.85
##                     3rd Qu.: 2931.52
##                     Max.   :13086.97


## Analysis of Variance Table
##
## Model 1: log(Item_Outlet_Sales) ~ log(Item_Visibility)
## Model 2: log(Item_Outlet_Sales) ~ log(Item_Visibility) * Item_MRP
## Model 3: log(Item_Outlet_Sales) ~ log(Item_Visibility) * Item_MRP + Outlet_Type
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   3996 3727.0
## 2   3994 2630.7  2    1096.3 3000.1 < 2.2e-16 ***
## 3   3991 729.2  3    1901.5 3468.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility), data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8772 -0.5390  0.1882  0.6938  1.9750
##
## Coefficients:
```

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.62848    0.05921  111.94   <2e-16 ***
## log(Item_Visibility) -0.22635   0.01956  -11.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9658 on 3996 degrees of freedom
## Multiple R-squared:  0.03241,    Adjusted R-squared:  0.03217
## F-statistic: 133.8 on 1 and 3996 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility) *
##     Item_MRP, data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1425 -0.2266  0.2070  0.4991  1.6907
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 5.3951221  0.1233659  43.733  < 2e-16 ***
## log(Item_Visibility)        -0.2412542 0.0410961  -5.870  4.7e-09 ***
## Item_MRP                    0.0088424  0.0008068  10.960  < 2e-16 ***
## log(Item_Visibility):Item_MRP 0.0001377 0.0002689   0.512    0.609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8116 on 3994 degrees of freedom
## Multiple R-squared:  0.317,  Adjusted R-squared:  0.3165
## F-statistic:   618 on 3 and 3994 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility) *
##     Item_MRP + Outlet_Type, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23221 -0.18391  0.04723  0.22635  1.34806
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.2447043  0.0660345  64.280   <2e-16 ***
## log(Item_Visibility)        -0.0490431 0.0217316  -2.257   0.0241 *
## Item_MRP                    0.0089782  0.0004250  21.124   <2e-16 ***
## Outlet_TypeSupermarket Type1 1.9471448 0.0206992  94.069   <2e-16 ***
## Outlet_TypeSupermarket Type2 1.8098845 0.0280190  64.595   <2e-16 ***
## Outlet_TypeSupermarket Type3 2.4757051 0.0277120  89.337   <2e-16 ***
## log(Item_Visibility):Item_MRP 0.0002159 0.0001417  1.524   0.1275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4274 on 3991 degrees of freedom
```

```
## Multiple R-squared:  0.8107, Adjusted R-squared:  0.8104
## F-statistic:  2848 on 6 and 3991 DF,  p-value: < 2.2e-16


##              Group.1        x
## 1     Grocery Store 157.9325
## 2 Supermarket Type1 148.9183
## 3 Supermarket Type2 146.3910
## 4 Supermarket Type3 154.1648


##              Group.1        x
## 1     Grocery Store 141.9583
## 2 Supermarket Type1 145.1227
## 3 Supermarket Type2 146.1616
## 4 Supermarket Type3 143.5479


##              Group.1        x
## 1     Grocery Store 146.7284
## 2 Supermarket Type1 136.0410
## 3 Supermarket Type2 133.6612
## 4 Supermarket Type3 131.8238


##              Group.1        x
## 1     Grocery Store 150.0489
## 2 Supermarket Type1 145.8780
## 3 Supermarket Type2 146.9438
## 4 Supermarket Type3 142.0538


##
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility) *
##     Item_MRP + Outlet_Type, data = test_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3558 -0.1938  0.0527  0.2346  1.3780
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.470e+00  4.453e-02 100.375   <2e-16 ***
## log(Item_Visibility)          2.815e-02  1.459e-02   1.929   0.0537 .
## Item_MRP                      7.661e-03  2.847e-04  26.907   <2e-16 ***
## Outlet_TypeSupermarket Type1  1.954e+00  1.448e-02 134.890   <2e-16 ***
## Outlet_TypeSupermarket Type2  1.774e+00  1.921e-02  92.325   <2e-16 ***
## Outlet_TypeSupermarket Type3  2.476e+00  1.920e-02 128.976   <2e-16 ***
## log(Item_Visibility):Item_MRP -1.998e-04  9.493e-05  -2.105   0.0353 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.44 on 9320 degrees of freedom
## Multiple R-squared:  0.789,  Adjusted R-squared:  0.7888
## F-statistic:  5808 on 6 and 9320 DF,  p-value: < 2.2e-16


##
```

```
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility) *
##     Item_MRP + Outlet_Type, data = fruits_and_veg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23396 -0.17294  0.04911  0.24525  1.15991
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.4716271  0.1306289  34.232   <2e-16 ***
## log(Item_Visibility)          0.0301549  0.0440031   0.685    0.493
## Item_MRP                      0.0078560  0.0008388   9.365   <2e-16 ***
## Outlet_TypeSupermarket Type1  1.9661645  0.0375470  52.365   <2e-16 ***
## Outlet_TypeSupermarket Type2  1.7966618  0.0500597  35.890   <2e-16 ***
## Outlet_TypeSupermarket Type3  2.5930834  0.0503167  51.535   <2e-16 ***
## log(Item_Visibility):Item_MRP -0.0001098  0.0002825  -0.389    0.698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4419 on 1320 degrees of freedom
## Multiple R-squared:  0.7958, Adjusted R-squared:  0.7949
## F-statistic: 857.6 on 6 and 1320 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = log(Item_Outlet_Sales) ~ log(Item_Visibility) *
##     Item_MRP + Outlet_Type, data = snacks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.89306 -0.20161  0.04725  0.23347  1.26247
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.4848276  0.1313971  34.132   <2e-16 ***
## log(Item_Visibility)          0.0326450  0.0418966   0.779    0.436
## Item_MRP                      0.0075692  0.0008176   9.257   <2e-16 ***
## Outlet_TypeSupermarket Type1  2.0025889  0.0402930  49.701   <2e-16 ***
## Outlet_TypeSupermarket Type2  1.8556303  0.0530473  34.981   <2e-16 ***
## Outlet_TypeSupermarket Type3  2.5264493  0.0533528  47.354   <2e-16 ***
## log(Item_Visibility):Item_MRP -0.0001221  0.0002667  -0.458    0.647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4502 on 1308 degrees of freedom
## Multiple R-squared:  0.7676, Adjusted R-squared:  0.7665
## F-statistic:   720 on 6 and 1308 DF,  p-value: < 2.2e-16


## [1] 15.19786 15.19786


##                      GVIF Df GVIF^(1/(2*Df))      Interacts With
## Item_Visibility 15.20331  1        3.899142             Item_MRP
```

```
## Item_MRP        15.20331  1        3.899142 log(Item_Visibility)
## Outlet_Type      1.05888  3        1.009581                   --
##                                     Other Predictors
## Item_Visibility Item_Visibility, Outlet_Type
## Item_MRP        Item_Visibility, Outlet_Type
## Outlet_Type        Item_Visibility, Item_MRP
```

Table 1: Regression Model Results Comparing Models

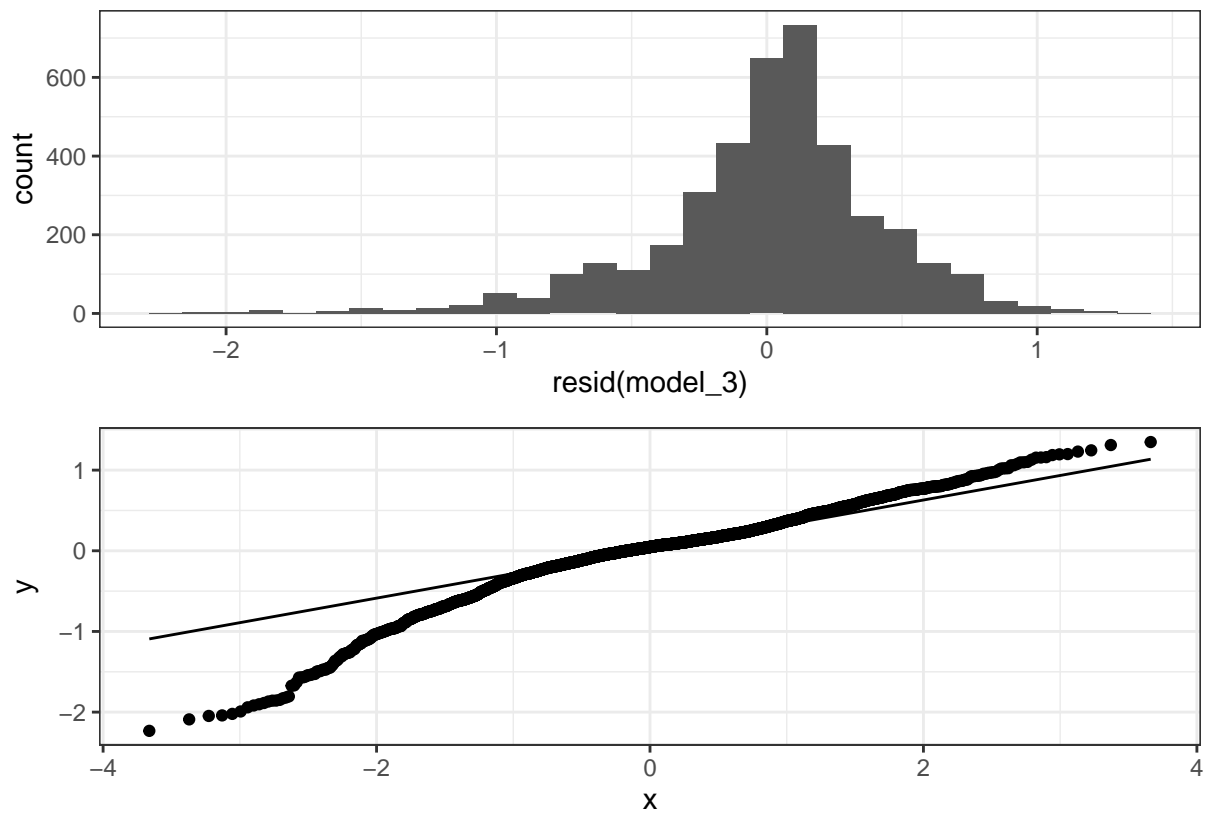|	*Dependent variable:*		
	Log(Outlet Sales)		
	(1)	(2)	(3)
Log(Visibility)	−0.226***	−0.241***	−0.049**
	(0.020)	(0.041)	(0.022)
Maximum Retail Price (MRP)		0.009***	0.009***
		(0.001)	(0.0004)
Type 1 Supermarket			1.947***
			(0.021)
Type 2 Supermarket			1.810***
			(0.028)
Type 3 Supermarket			2.476***
			(0.028)
Log(Visibility) - MRP Interaction		0.0001	0.0002
		(0.0003)	(0.0001)
Constant	6.628***	5.395***	4.245***
	(0.059)	(0.123)	(0.066)
Observations	3,998	3,998	3,998
R^2	0.032	0.317	0.811
Adjusted R^2	0.032	0.317	0.810
Residual Std. Error	0.966 (df = 3996)	0.812 (df = 3994)	0.427 (df = 3991)
F Statistic	133.843*** (df = 1; 3996)	618.009*** (df = 3; 3994)	2,848.379*** (df = 6; 3991)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Model Results Comparing Item Data

	Dependent variable:		
	Log(Outlet Sales)		
	(1)	(2)	(3)
Log(Visibility)	0.028*	0.030	0.033
	(0.015)	(0.044)	(0.042)
Maximum Retail Price (MRP)	0.008***	0.008***	0.008***
	(0.0003)	(0.001)	(0.001)
Type 1 Supermarket	1.954***	1.966***	2.003***
	(0.014)	(0.038)	(0.040)
Type 2 Supermarket	1.774***	1.797***	1.856***
	(0.019)	(0.050)	(0.053)
Type 3 Supermarket	2.476***	2.593***	2.526***
	(0.019)	(0.050)	(0.053)
Log(Visibility) - MRP Interaction	−0.0002**	−0.0001	−0.0001
	(0.0001)	(0.0003)	(0.0003)
Constant	4.470***	4.472***	4.485***
	(0.045)	(0.131)	(0.131)
Observations	9,327	1,327	1,315
R^2	0.789	0.796	0.768
Adjusted R^2	0.789	0.795	0.767
Residual Std. Error	0.440 (df = 9320)	0.442 (df = 1320)	0.450 (df = 1308)
F Statistic	5,807.752*** (df = 6; 9320)	857.568*** (df = 6; 1320)	720.034*** (df = 6; 1308)

Note: *p<0.1; **p<0.05; ***p<0.01

```
##
##   studentized Breusch-Pagan test
##
## data:  model_3
## BP = 14.595, df = 6, p-value = 0.02365
```