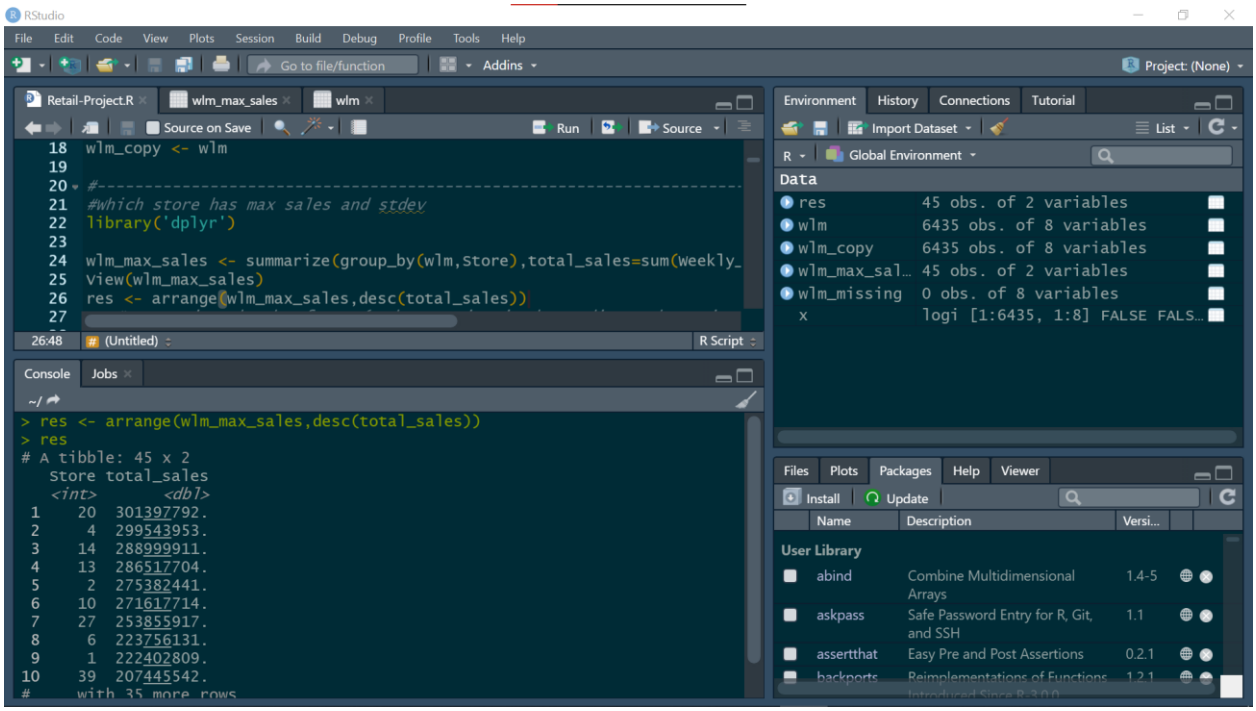
RETAIL ANALYSIS on WALMART by **Piyush Kumar**

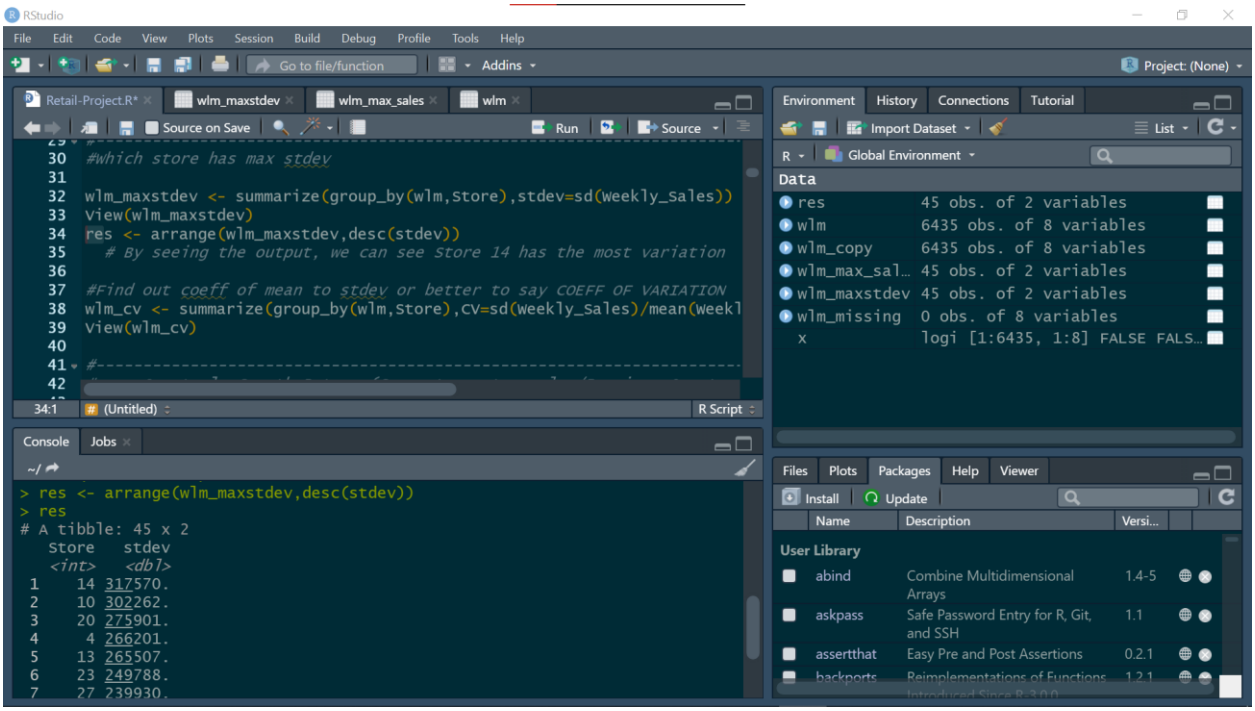
//This file contains all the necessary insights and outputs for the dataset “Walmart_Store_sales.csv”

Statistical Analysis

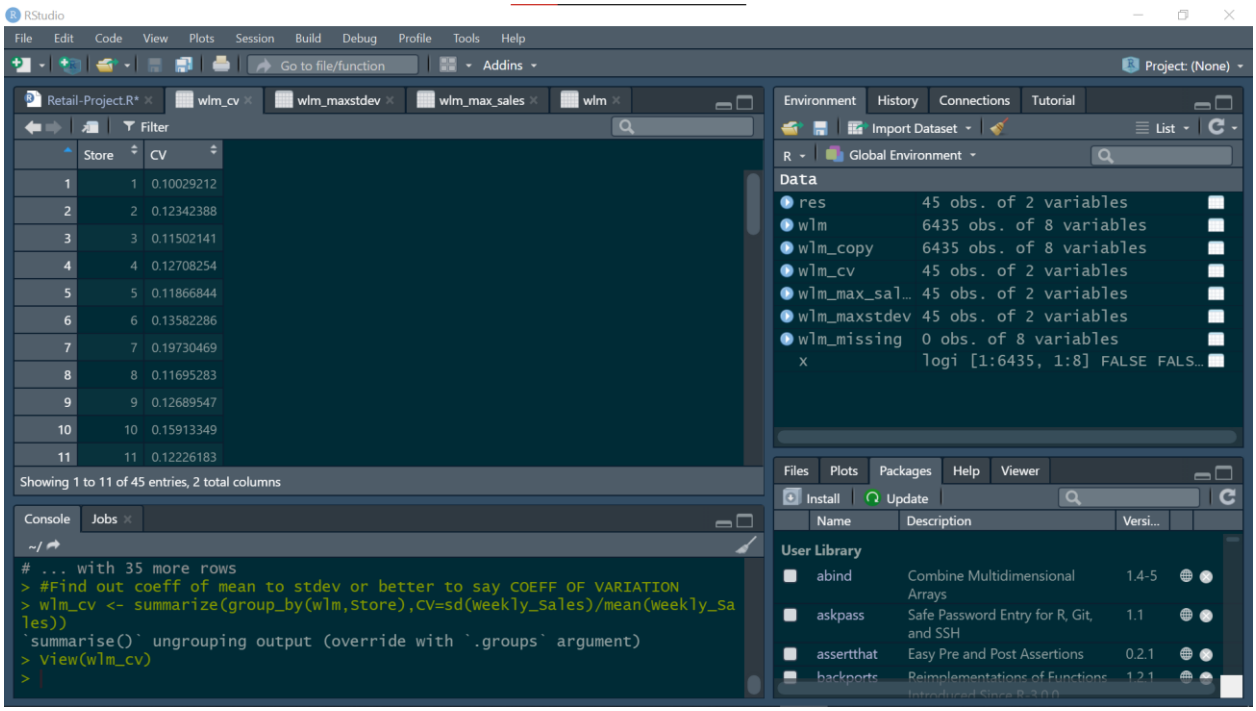
- Maximum sales : Store 20



- Max Standard Deviation : Store 14

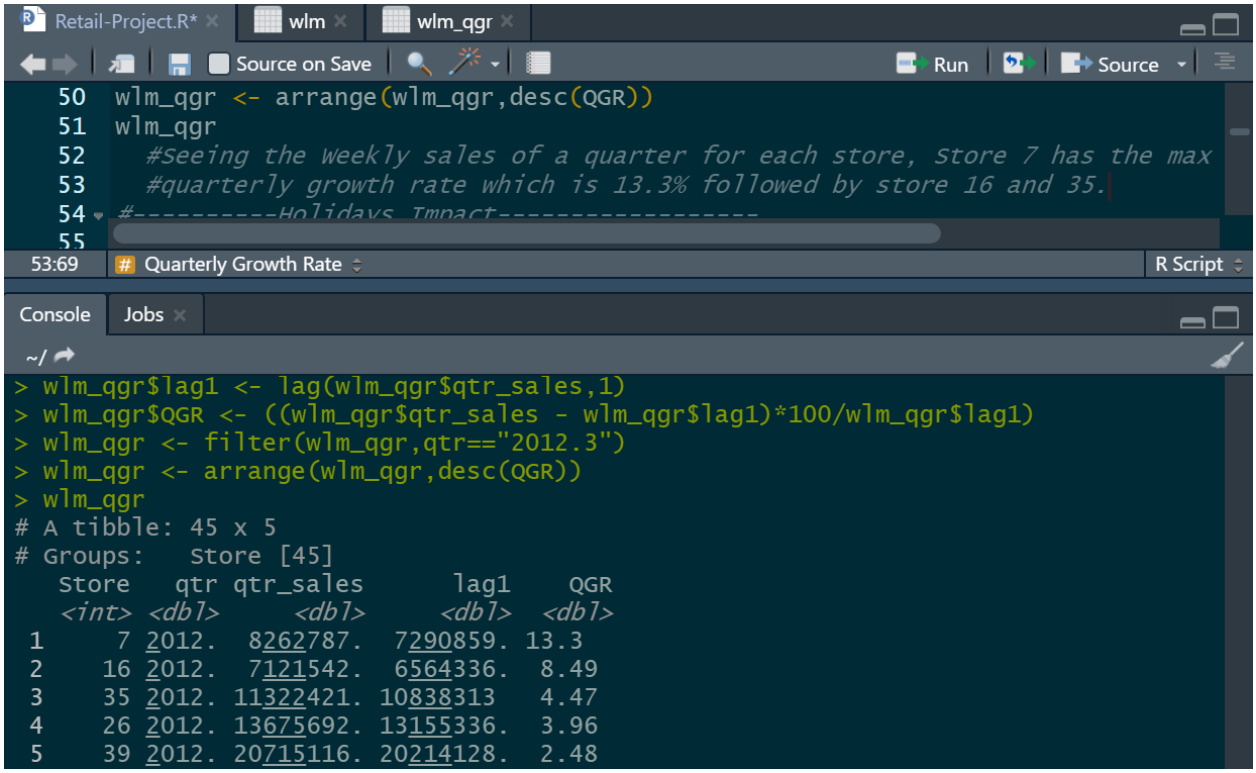


➢ Coefficient of Variance :

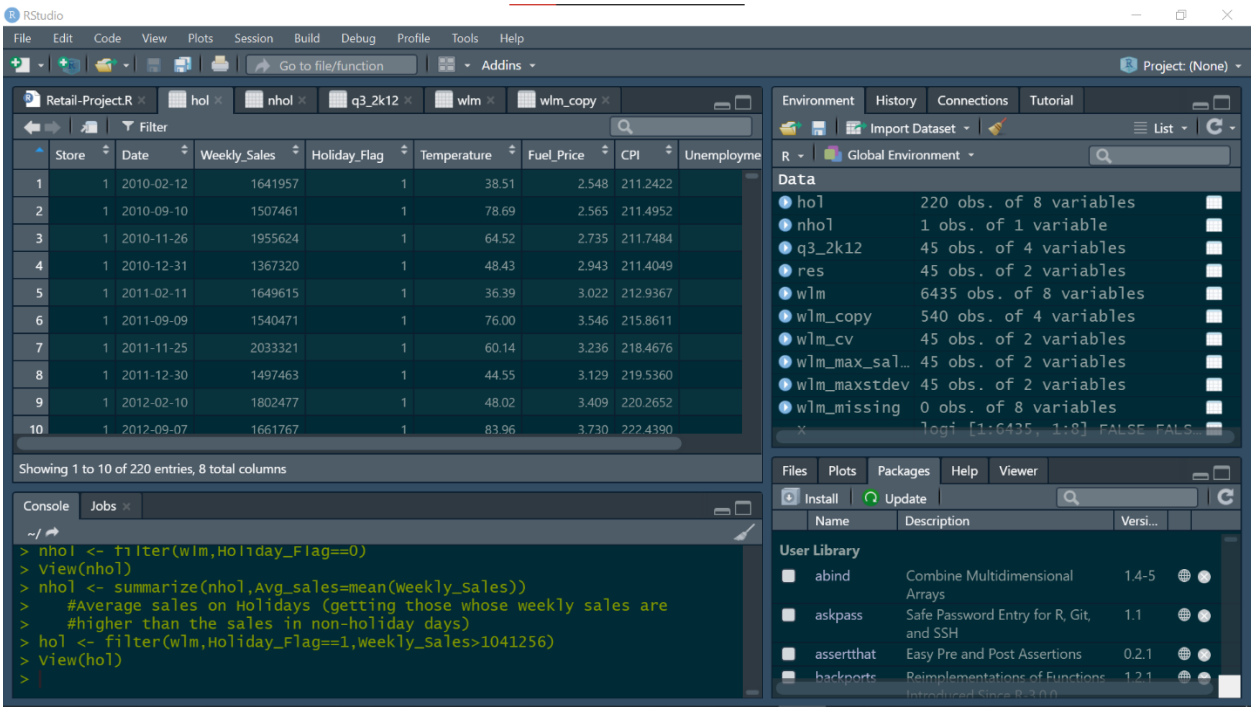


➢ Good quarterly growth rate in Quarter 3 of 2012:

Store 7 has max quarterly growth rate (refer to source code for better explanation)

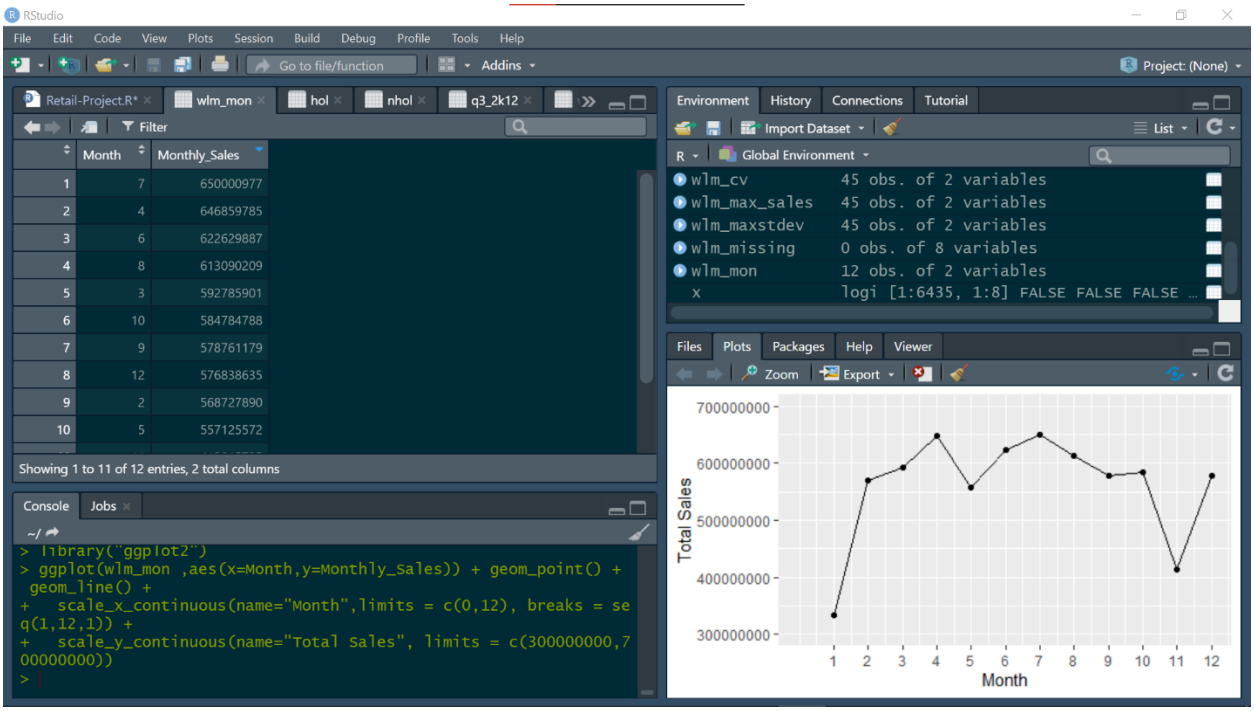


- Holidays with higher sales than the average sales in non-holiday season:
  "hol" dataset contains all the desired entries

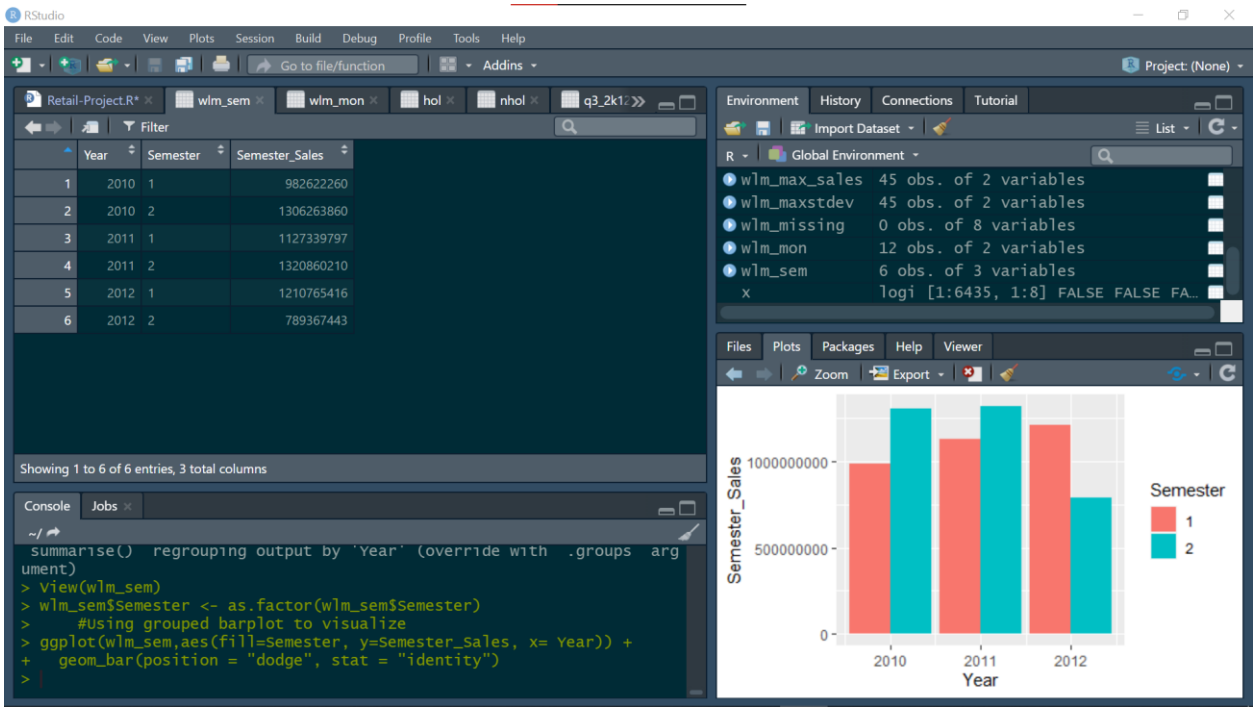


1. Monthly and semester view of sales in units and insights

>> Monthly Sales Insights: Sales is least in JAN but shoots up in FEB by a lot.
Main insight is Sales decreases a lot in NOV but increases a lot in DEC.



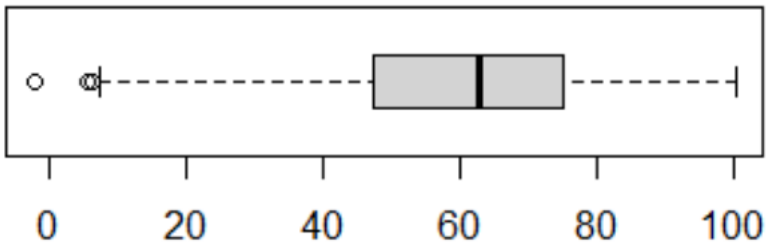
>> Semester sales insights: Sales increased in every second semester except for year 2012



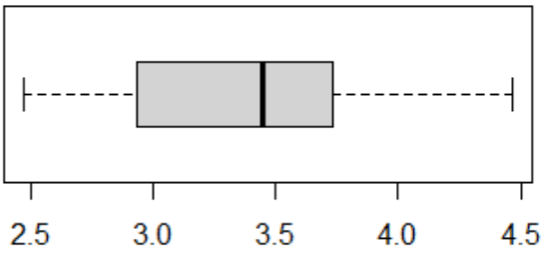
LINEAR REGRESSION MODEL

Boxplots for OUTLIER DETECTION

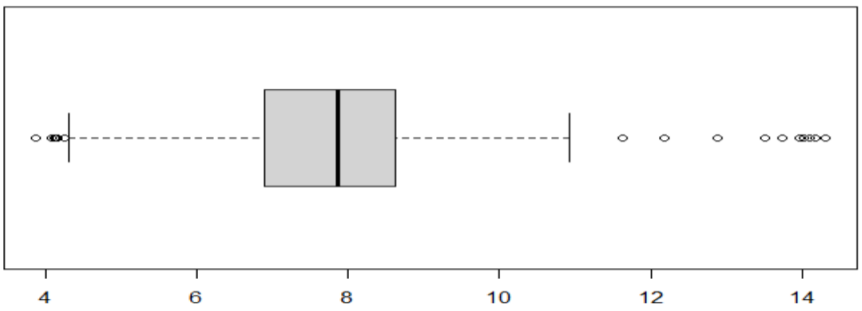
- Temperature (OUTLIERS <10)



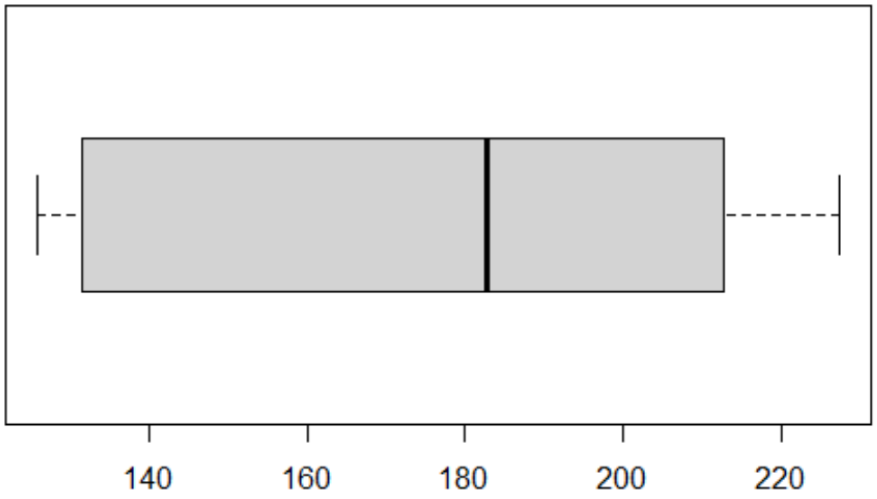
- Fuel Price (No outliers)



- Unemployment (Outliers <4.5 and >10)



- CPI (No outliers)



REGRESSION MODEL OUTPUTS :

Model before hypothesizing / dropping variables on the basis of p-values

```
Console  Jobs
~/
> wlm_test <- subset(wlm_new,sel==FALSE)
> wlm_model <- lm(Weekly_Sales ~ Fuel_Price+Temperature+Holiday_Flag+CPI+Unemployment, dat
a = wlm_new)
> summary(wlm_model)

Call:
lm(formula = Weekly_Sales ~ Fuel_Price + Temperature + Holiday_Flag +
    CPI + Unemployment, data = wlm_new)

Residuals:
    Min      1Q  Median      3Q     Max
-305166  -78247  -18260   53643  854412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2427856    1752958  -1.385   0.1683
Fuel_Price      -24337      47335  -0.514   0.6080
Temperature       -2160        922  -2.343   0.0206 *
Holiday_Flag1     89376      49338   1.811   0.0723 .
CPI               16632       6786   2.451   0.0155 *
Unemployment      80209      58727   1.366   0.1742
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 146500 on 137 degrees of freedom
Multiple R-squared:  0.1495,    Adjusted R-squared:  0.1184
F-statistic: 4.815 on 5 and 137 DF,  p-value: 0.0004359
```

Output after dropping variables on the basis of p-values :

Console | Jobs

~/

```
Residual standard error: 146500 on 137 degrees of freedom
Multiple R-squared:  0.1495,    Adjusted R-squared:  0.1184
F-statistic: 4.815 on 5 and 137 DF,  p-value: 0.0004359

> #Dropping non-significant variables seeing the p-values : Holiday,Day,Year
> wlm_model <- lm(Weekly_Sales ~ Temperature+CPI, data = wlm_new)
> summary(wlm_model)

Call:
lm(formula = Weekly_Sales ~ Temperature + CPI, data = wlm_new)

Residuals:
    Min      1Q  Median      3Q     Max
-312205  -85704   -9198   57222  830489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -233190     616327  -0.378  0.70574
Temperature    -2769        877  -3.157  0.00195 **
CPI             9156       2872   3.187  0.00177 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 147900 on 140 degrees of freedom
Multiple R-squared:  0.1139,    Adjusted R-squared:  0.1012
F-statistic: 8.998 on 2 and 140 DF,  p-value: 0.0002107

>
```