Project_(631)

Our research question is - **"How do Storage Capacity and RAM Size affect laptop price, and is there an interaction between them?"**

Loading the dataset and necessary libraries

```
laptop <- read.table("Laptop_price.csv", header = TRUE, sep = ",")

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
```

```
library(ggplot2)
library(car)
```

```
Warning: package 'car' was built under R version 4.4.2

Loading required package: carData

Warning: package 'carData' was built under R version 4.4.2
```

```
Attaching package: 'car'

The following object is masked from 'package:dplyr':

    recode
```

Displaying summary statistics and first 6 rows of the dataset

```
summary(laptop)
```

```
    Brand           Processor_Speed    RAM_Size    Storage_Capacity
 Length:1000        Min.   :1.512   Min.   : 4.0   Min.   : 256.0
 Class :character   1st Qu.:2.089   1st Qu.: 8.0   1st Qu.: 256.0
 Mode  :character   Median :2.761   Median :16.0   Median : 512.0
                    Mean   :2.751   Mean   :15.5   Mean   : 584.6
                    3rd Qu.:3.363   3rd Qu.:32.0   3rd Qu.:1000.0
                    Max.   :3.999   Max.   :32.0   Max.   :1000.0
  Screen_Size        Weight          Price
 Min.   :11.01   Min.   :2.001   Min.   : 8570
 1st Qu.:12.64   1st Qu.:2.717   1st Qu.:10114
 Median :14.10   Median :3.465   Median :17287
 Mean   :14.06   Mean   :3.467   Mean   :19604
 3rd Qu.:15.53   3rd Qu.:4.213   3rd Qu.:31566
 Max.   :16.99   Max.   :4.991   Max.   :33504
```

```
head(laptop)
```

```
   Brand Processor_Speed RAM_Size Storage_Capacity Screen_Size   Weight
1   Asus        3.830296       16              512    11.18515 2.641094
2   Acer        2.912833        4             1000    11.31137 3.260012
3 Lenovo        3.241627        4              256    11.85302 2.029061
4   Acer        3.806248       16              512    12.28036 4.573865
5   Acer        3.268097       32             1000    14.99088 4.193472
6     HP        1.881348       16              256    11.94396 4.840268
      Price
1 17395.093
2 31607.606
3  9291.024
4 17436.728
5 32917.991
6  9543.720
```

Converting columns to factors for categorical analysis

```
laptop <- laptop %>%
  mutate(across(c(Storage_Capacity, RAM_Size, Brand), as.factor))
```

Checking whether our dataset is balanced or not

```
tapply(laptop[-1,]$Price,
laptop[-1,]$Storage_Capacity:laptop[-1,]$RAM_Size, length)
```

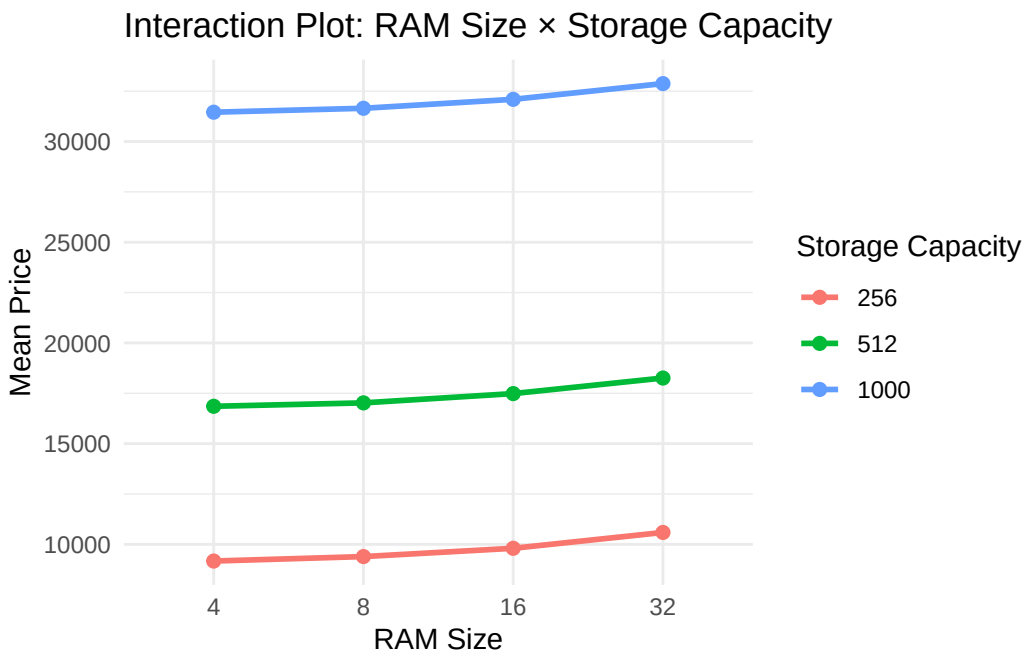
```
  256:4   256:8  256:16  256:32   512:4   512:8  512:16  512:32  1000:4  1000:8
     88      87      84      98      80      64      80      82      75      95
1000:16 1000:32
     72      94
```

Checking structure and data types of the dataset

```
str(laptop)
```

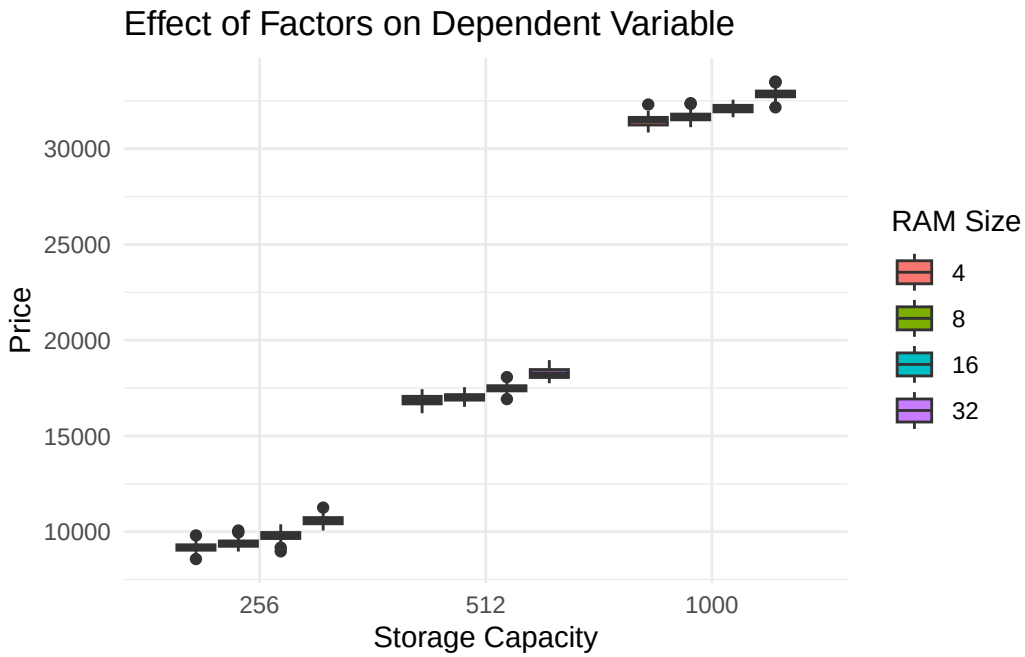
```
'data.frame':   1000 obs. of  7 variables:
 $ Brand           : Factor w/ 5 levels "Acer","Asus",..: 2 1 5 1 1 4 5 5 5 1 ...
 $ Processor_Speed : num  3.83 2.91 3.24 3.81 3.27 ...
 $ RAM_Size        : Factor w/ 4 levels "4","8","16","32": 3 1 1 3 4 3 4 1 1 4 ...
 $ Storage_Capacity: Factor w/ 3 levels "256","512","1000": 2 3 1 2 3 1 1 1 2 2 ...
 $ Screen_Size     : num  11.2 11.3 11.9 12.3 15 ...
 $ Weight          : num  2.64 3.26 2.03 4.57 4.19 ...
 $ Price           : num  17395 31608 9291 17437 32918 ...
```

```
# Summarize mean price by group
summary_data <- laptop %>%
  group_by(RAM_Size, Storage_Capacity) %>%
  summarise(mean_price = mean(Price, na.rm = TRUE), .groups = "drop")

# Interaction plot
ggplot(summary_data, aes(x = RAM_Size, y = mean_price, color = Storage_Capacity, group = Sto
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "Interaction Plot: RAM Size × Storage Capacity",
       x = "RAM Size",
       y = "Mean Price",
       color = "Storage Capacity")
```



Boxplot to visualize distribution of Price by Storage_Capacity and RAM_Size

```
ggplot(laptop, aes(x = Storage_Capacity, y = Price, fill = RAM_Size)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Effect of Factors on Dependent Variable",
       x = "Storage Capacity",
       y = "Price",
       fill = "RAM Size")
```



Fitting Two-Way ANOVA: Storage_Capacity and RAM_Size

```
# Fitting Two-Way ANOVA: Storage_Capacity and RAM_Size
model <- aov(Price ~ Storage_Capacity * RAM_Size, data = laptop)
Anova(model,type=2)
```

```
Anova Table (Type II tests)

Response: Price
                            Sum Sq  Df    F value Pr(>F)
Storage_Capacity          8.7763e+10   2 720444.501 <2e-16 ***
RAM_Size                  3.0919e+08   3   1692.108 <2e-16 ***
Storage_Capacity:RAM_Size 7.4541e+04   6      0.204 0.9756
Residuals                 6.0178e+07 988
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A **p-value $<$ 0.05** indicates a **statistically significant** effect.

For both Storage_Capacity and RAM_Size have p value less than 0.05.

- **Storage_Capacity** has a strong and significant impact on laptop price.

- **RAM_Size** also has a strong and significant impact on laptop price.

Fitting updated Two-Way ANOVA: Storage_Capacity and RAM_Size

```
# Fitting updated Two-Way ANOVA: Storage_Capacity and RAM_Size
redmodel <- aov(Price ~ Storage_Capacity + RAM_Size, data = laptop)
Anova(redmodel,type=2)
```

```
Anova Table (Type II tests)

Response: Price
                     Sum Sq  Df  F value    Pr(>F)
Storage_Capacity 8.7763e+10   2 723923.0 < 2.2e-16 ***
RAM_Size         3.0919e+08   3   1700.3 < 2.2e-16 ***
Residuals        6.0252e+07 994
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

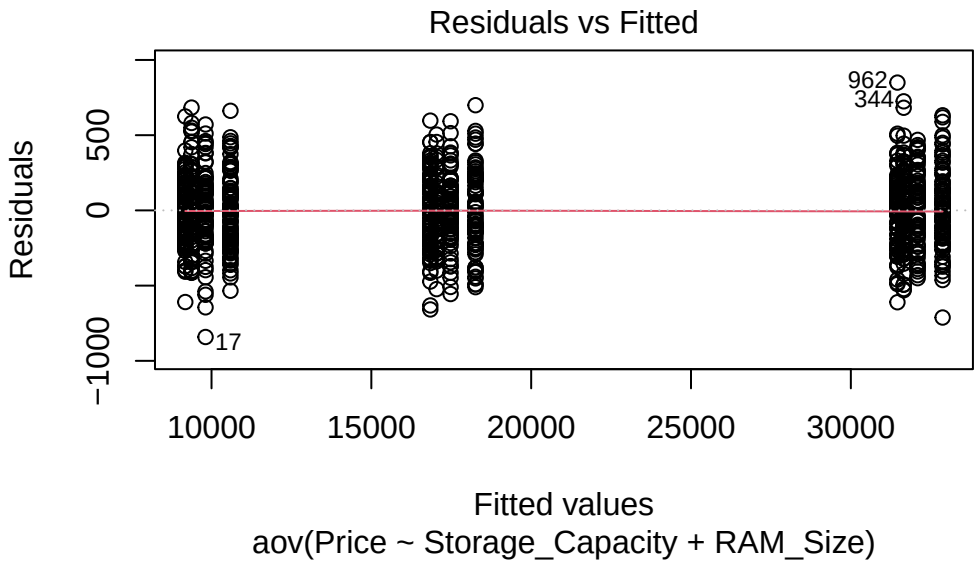
Shapiro-Wilk test for normality of residuals

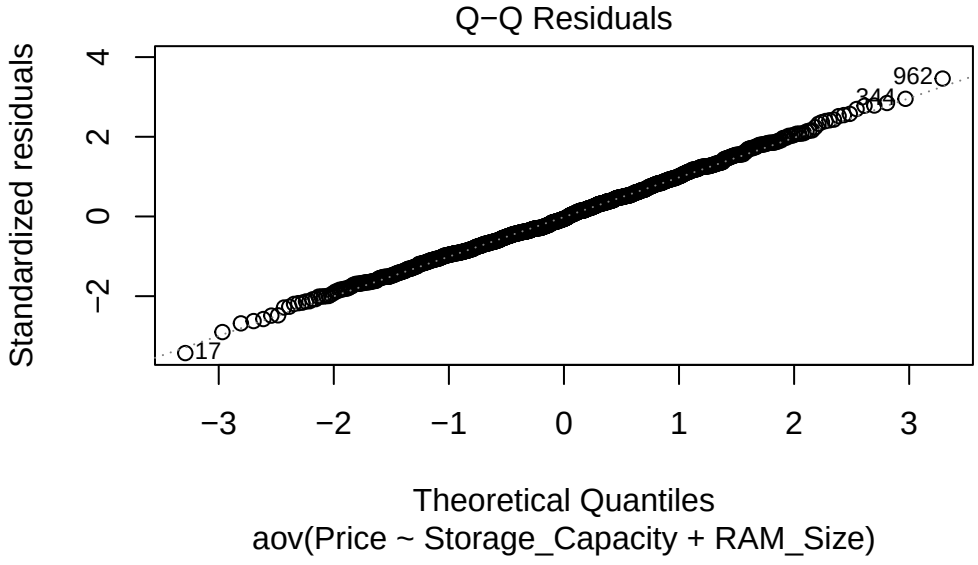
```
shapiro.test(residuals(redmodel))
```

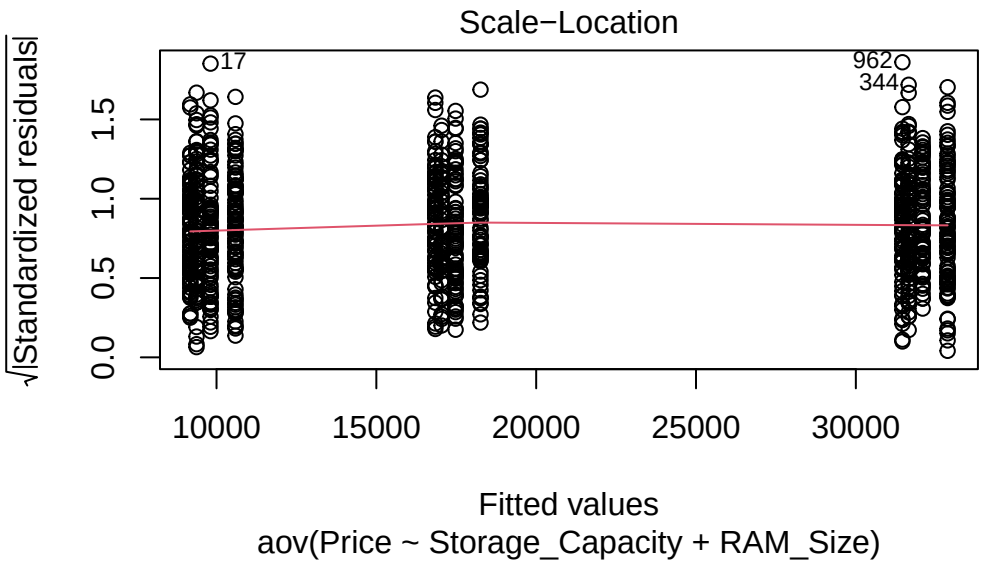
```
    Shapiro-Wilk normality test

data:  residuals(redmodel)
W = 0.99825, p-value = 0.4054
```

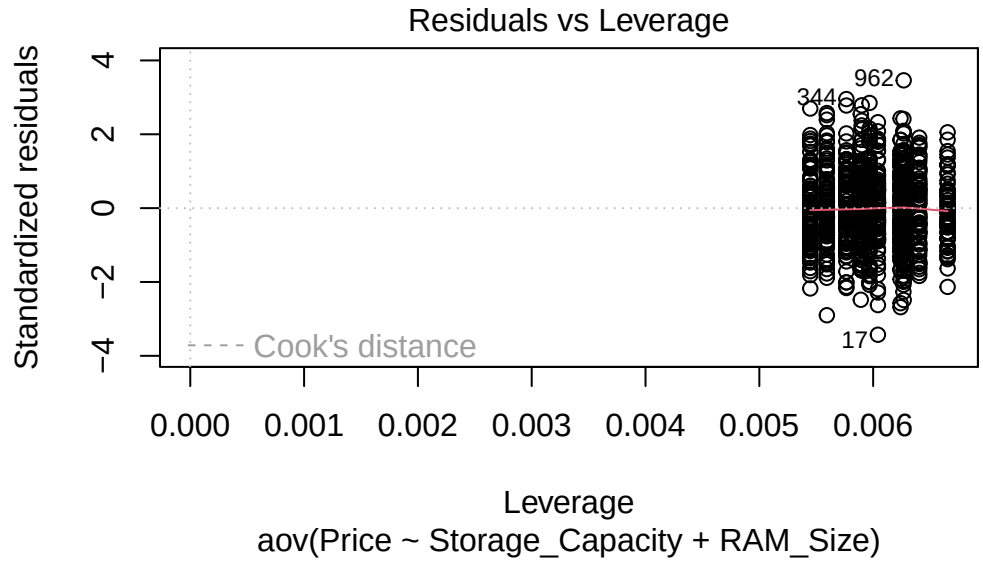
Plotting to check for assumptions

```
plot(redmodel)
```


Residuals vs Fitted
Residuals
−1000
0
500
962
344
17
10000
15000
20000
25000
30000
Fitted values
aov(Price ~ Storage_Capacity + RAM_Size)


Q−Q Residuals
Standardized residuals
−2
0
2
4
962
344
17
−3
−2
−1
0
1
2
3
Theoretical Quantiles
aov(Price ~ Storage_Capacity + RAM_Size)





- **Residuals vs Fitted** - linearity and equal variance (homoscedasticity) assumptions are met
- **Q-Q plot** - Normality assumption is satisfied.

- **Scale-Location** - Homoscedasticity assumption satisfied.
- **Residuals vs Leverage** - No concerning outliers.
- **All assumptions are satisfied**.

Levene's test to check equality of variances

```
leveneTest(Price ~ Storage_Capacity * RAM_Size, data = laptop)
```

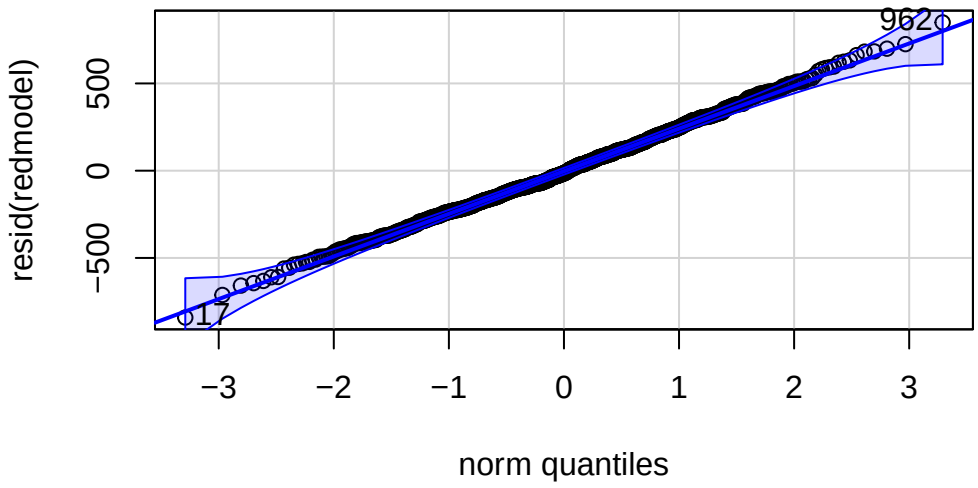
```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group  11  0.9494 0.4917
      988
```

```
#Equal variances are satisfied
```

Residuals are normal

Q-Q plot for reduced model residuals

```
qqPlot(resid(redmodel))
```



```
[1] 962  17
```

Tukey HSD test for pairwise comparisons among groups

```
tukey <- TukeyHSD(redmodel)
tukey
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Price ~ Storage_Capacity + RAM_Size, data = laptop)

$Storage_Capacity
             diff       lwr       upr p adj
512-256   7666.62  7621.639  7711.601     0
1000-256 22280.25 22236.320 22324.170     0
1000-512 14613.63 14567.999 14659.252     0

$RAM_Size
           diff       lwr       upr p adj
8-4    196.2716  138.9680  253.5753     0
16-4   633.9371  576.0953  691.7789     0
32-4  1413.9737 1358.1440 1469.8034     0
16-8   437.6655  379.9981  495.3329     0
32-8  1217.7021 1162.0531 1273.3511     0
32-16  780.0366  723.8336  836.2396     0
```