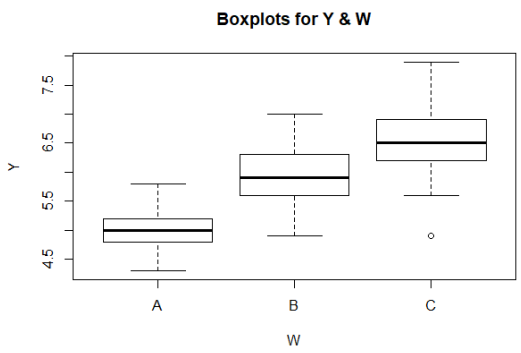
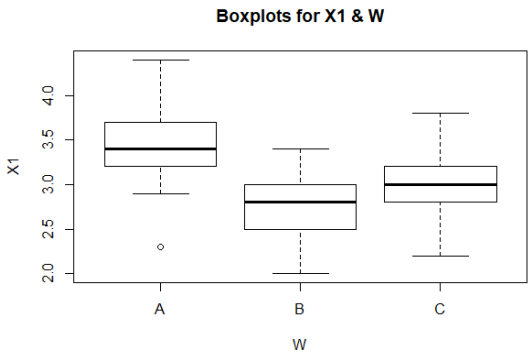
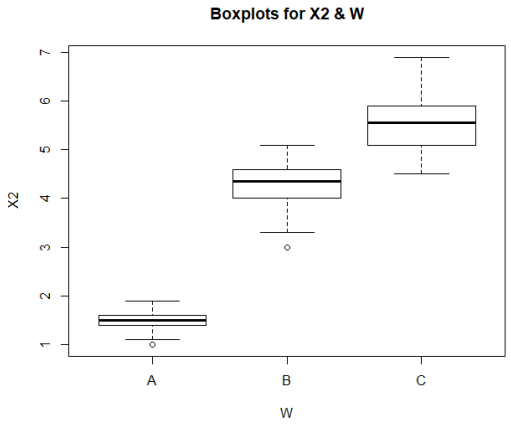
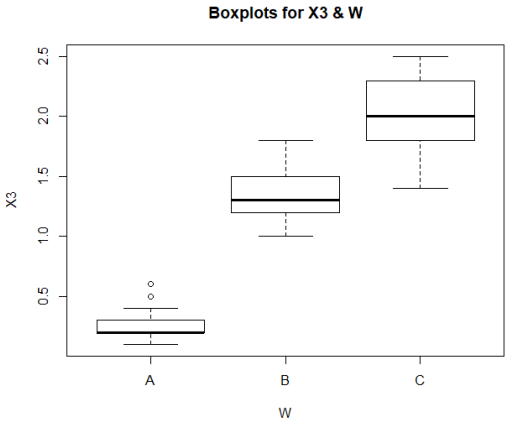
A. i. Below there are four boxplots of Y, X1, X2, X3 versus W. The mean of each boxplot seems to be different. The only possible difference could be in variable X1 versus W where group B and C may not be significant different.









ii.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
W	2	63.21	31.606	119.3	<2e-16 ***
Residuals	147	38.96	0.265		

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

The ANOVA test between Y and W shows that we reject H0 which states that all the means of each category are equal (p-value close to zero) and there is at least one mean that is different from the rest.

```
             Df Sum Sq Mean Sq F value Pr(>F)
W             2  11.35   5.672   49.16 <2e-16 ***
Residuals   147  16.96   0.115
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The ANOVA test between X1 and W shows that we reject H0 which states that all the means of each category are equal (p-value close to zero) and there is at least one mean that is different from the rest.

```
             Df Sum Sq Mean Sq F value Pr(>F)
W             2  437.1  218.55    1180 <2e-16 ***
Residuals   147   27.2    0.19
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The ANOVA test between X2 and W shows that we reject H0 which states that all the means of each category are equal (p-value close to zero) and there is at least one mean that is different from the rest.

```
             Df Sum Sq Mean Sq F value Pr(>F)
W             2  80.41   40.21     960 <2e-16 ***
Residuals   147   6.16    0.04
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The ANOVA test between X3 and W shows that we reject H0 which states that all the means of each category are equal (p-value close to zero) and there is at least one mean that is different from the rest.

iii.

```
Bartlett test of homogeneity of variances

data:  Y by W
Bartlett's K-squared = 16.006, df = 2, p-value = 0.0003345
```

```
        Shapiro-Wilk normality test

data:  fitY$residuals
W = 0.9879, p-value = 0.2189
```

By checking the assumptions for homogeneity and normality for Y on W, we observe in the Bartlett test for homogeneity of variances that we reject H0 which states that the variances of

residuals in each category are equal (p-value=0.0003<a). Also, in Shapiro-Wilk normality test, we cannot reject H0 which states that the residuals follow normal distribution (p-value=0.21>a).

By transforming Y to log(Y) and running the same tests we obtain a p-value=0.08>a (at a=5%) for Bartlett test and a p-value=0.24>a for Shapiro-Wilk test. Hence, both assumptions hold with this transformation.

```
        Bartlett test of homogeneity of variances

data:  X1 by W
Bartlett's K-squared = 2.0911, df = 2, p-value = 0.3515
```

```
        Shapiro-Wilk normality test

data:  fitX1$residuals
W = 0.98948, p-value = 0.323
```

By checking the assumptions for homogeneity and normality for X1 on W, we observe in the Bartlett test for homogeneity of variances that we cannot reject H0 which states that the variances of residuals in each category are equal (p-value=0.35>a). Also, in Shapiro-Wilk normality test, we cannot reject H0 which states that the residuals follow normal distribution (p-value=0.21>a). Both assumptions hold for X1 on W.

```
        Bartlett test of homogeneity of variances

data:  X2 by W
Bartlett's K-squared = 55.423, df = 2, p-value = 9.229e-13
```

```
        Shapiro-Wilk normality test

data:  fitX2$residuals
W = 0.98108, p-value = 0.03676
```

By checking the assumptions for homogeneity and normality for X2 on W, we observe in the Bartlett test for homogeneity of variances that we reject H0 which states that the variances of residuals in each category are equal (p-value=0<a). Also, in Shapiro-Wilk normality test, we reject H0 (in a=5%) which states that the residuals follow normal distribution (p-value=0.03<a). Both assumptions do not hold for X2 on W.

By transforming X2 to log(X2) and running the same tests we obtain a p-value=0.31>a for Bartlett test and a p-value=0.052>a (for a=5%) for Shapiro-Wilk test. Hence, both assumptions hold with this transformation.

Bartlett test of homogeneity of variances
data: X3 by W
Bartlett's K-squared = 39.213, df = 2, p-value = 3.055e-09

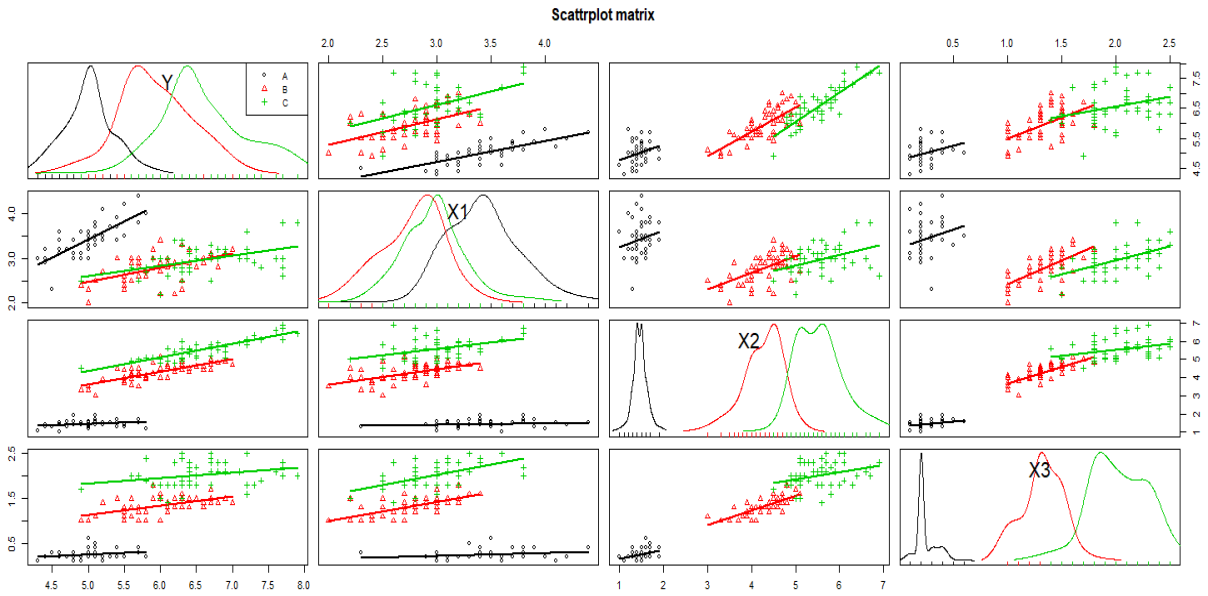
Shapiro-Wilk normality test
data: fitX3$residuals
W = 0.97217, p-value = 0.003866

By checking the assumptions for homogeneity and normality for X3 on W, we observe in the Bartlett test for homogeneity of variances that we reject H0 which states that the variances of residuals in each category are equal (p-value=0<a). Also, in Shapiro-Wilk normality test, we reject H0 which states that the residuals follow normal distribution (p-value=0.003<a). Both assumptions do not hold for X3 on W.

By transforming X3 to 1/X3 and running the same tests we obtain a p-value=0.55>a for Bartlett test and a p-value=0.02>a (for a=1%) for Shapiro-Wilk test. Hence, both assumptions hold with this transformation, only if a=1%.

There are many transformations that can be applied if the assumptions of homogeneity and normality of the residuals do not hold (for example Garch model for conditional heteroskedasticity).

B. Scatter plot matrix of Y, X1, X2, X3 with different levels of W in each plot with different colors (A with black, B with Red, C with green)



C.

```
Call:
lm(formula = Y ~ X1, data = datas)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5561 -0.6333 -0.1120  0.5579  2.2226

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5262     0.4789   13.63   <2e-16 ***
X1           -0.2234     0.1551   -1.44    0.152
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8251 on 148 degrees of freedom
Multiple R-squared:  0.01382,  Adjusted R-squared:  0.007159
F-statistic: 2.074 on 1 and 148 DF,  p-value: 0.1519
```

The regression model of Y on X1 does not provide any significant information, as the coefficient of X1 is not significant (p-value=0.12>a) and the r-squared=1% is low as expected. Hence, the variable X1 cannot explain with significance the dependent variable Y.

D.

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + W + W * X1 + W * X2 + W * X3,
    data = datas)

Residuals:
     Min       1Q   Median       3Q      Max
-0.73883 -0.21607  0.00051  0.21813  0.74427

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.3519     0.4965   4.737 5.33e-06 ***
X1            0.6548     0.1168   5.605 1.09e-07 ***
X2            0.2376     0.2629   0.904   0.3678
X3            0.2521     0.4384   0.575   0.5661
WB           -0.4564     0.6727  -0.678   0.4987
WC           -1.6520     0.7067  -2.338   0.0208 *
X1:WB        -0.2680     0.2172  -1.234   0.2194
X1:WC        -0.3245     0.2016  -1.610   0.1098
X2:WB         0.6708     0.3017   2.223   0.0278 *
X2:WC         0.7080     0.2765   2.561   0.0115 *
X3:WB        -0.9313     0.5866  -1.588   0.1146
X3:WC        -0.4219     0.4765  -0.885   0.3775
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2997 on 138 degrees of freedom
Multiple R-squared:  0.8787,   Adjusted R-squared:  0.869
F-statistic: 90.87 on 11 and 138 DF,  p-value: < 2.2e-16
```

The regression model of Y on X1, X2, X3, W and the interactions between them W*X1, W*X2, W*X3. The independent variable X1 becomes significant, while X2 and X3 are not significant. The intercept represents WA and is significant along with WC. The interactions of X2 with W are also significant.

E.

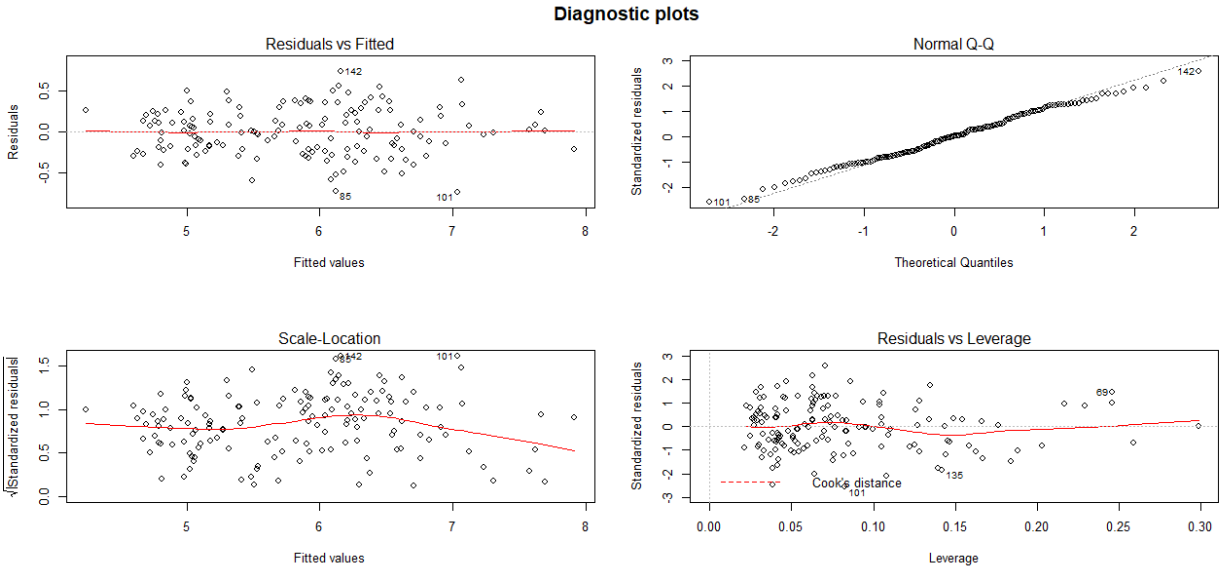
```
Analysis of Variance Table

Response: Y
           Df Sum Sq Mean Sq  F value    Pr(>F)
X1          1  1.412   1.412  15.7236 0.0001172 ***
X2          1 84.427  84.427 939.9959 < 2.2e-16 ***
X3          1  1.883   1.883  20.9689 1.032e-05 ***
W           2  0.889   0.444   4.9485 0.0084033 **
X1:W        2  0.385   0.192   2.1421 0.1212968
X2:W        2  0.543   0.272   3.0245 0.0518166 .
X3:W        2  0.234   0.117   1.3010 0.2755893
Residuals 138 12.395   0.090
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
        Shapiro-Wilk normality test

data:  fit$residuals

W = 0.99341, p-value = 0.7269
```



The results of ANOVA table provide information of different means in groups between X1, X2, X3 and W. Also, the Shapiro Wilk normality test provides information of normality in residuals of the model (p-value=0.72>a). Lastly, the diagnostic plots provide information of linear relationship from the first plot, normal distributed residuals from the second plot, homoskedasticity from the third plot and that there are not outliers and influential points from the last plot. Hence, there is no need for alternatives.

F.

```
Step:  AIC=-351.51
Y ~ X1 + X2 + W + X1:W + X2:W

       Df Sum of Sq    RSS     AIC
<none>              12.771 -351.51
+ X3    1   0.14311 12.628 -351.20
- X1:W  2   0.61227 13.384 -348.49
- X2:W  2   0.62995 13.402 -348.29

Call:
lm(formula = Y ~ X1 + X2 + W + X1:W + X2:W, data = datas)

Coefficients:
(Intercept)           X1           X2           WB           WC        X1:WB        X1:WC        X2:WB
     2.3037       0.6674       0.2834      -0.1873      -1.6790      -0.4198      -0.4075       0.4522
      X2:WC
     0.6514
```

Using the stepwise regression method, we obtain the optimal dimensions of the model based on AIC. The final model includes X1, X2, W and the interactions X1*W and X2*W. The intercept is equal to 2.3 which means if X1 and X2 are zero and W is A then y in average will be 2.3. X1 has a coefficient of 0.66 which means that if X1 increases by one unit and all the other variables remain constant, then Y will increase by 0.66. Similar results we obtain from the coefficient of X2 which is equal to 0.28. The coefficient of WB is equal to -0.18. This means that if W equals the B category, then Y will decrease by 0.18. Similar results we obtain from WC which have a coefficient of -0.1.67. Lastly, the interaction variables show the slope between these variables. For example, the X1:WB shows that if X1 increases by one unit and W is the B category and all the other variables remain constant then Y will decrease by 0.41. Similar results we obtain for the other interaction variables.

G. The first table below is the contingency table of X3 on W and the second table of X3 on Z. They depict the number of times each value of X3 belongs to one of the categories of W or Z.

X3	A	B	C
0.1	5	0	0
0.2	29	0	0
0.3	7	0	0
0.4	7	0	0
0.5	1	0	0
0.6	1	0	0
1	0	7	0
1.1	0	3	0
1.2	0	5	0
1.3	0	13	0
1.4	0	7	1
1.5	0	10	2
1.6	0	3	1
1.7	0	1	1
1.8	0	1	11
1.9	0	0	5
2	0	0	6
2.1	0	0	6
2.2	0	0	3
2.3	0	0	8
2.4	0	0	3
2.5	0	0	3

X3	(0.0976,0.7]	(0.7,1.3]	(1.3,1.9]	(1.9,2.5]
0.1	5	0	0	0
0.2	29	0	0	0
0.3	7	0	0	0
0.4	7	0	0	0
0.5	1	0	0	0
0.6	1	0	0	0
1	0	7	0	0
1.1	0	3	0	0
1.2	0	5	0	0
1.3	0	13	0	0
1.4	0	0	8	0

1.5	0	0	12	0
1.6	0	0	4	0
1.7	0	0	2	0
1.8	0	0	12	0
1.9	0	0	5	0
2	0	0	0	6
2.1	0	0	0	6
2.2	0	0	0	3
2.3	0	0	0	8
2.4	0	0	0	3
2.5	0	0	0	3

H.

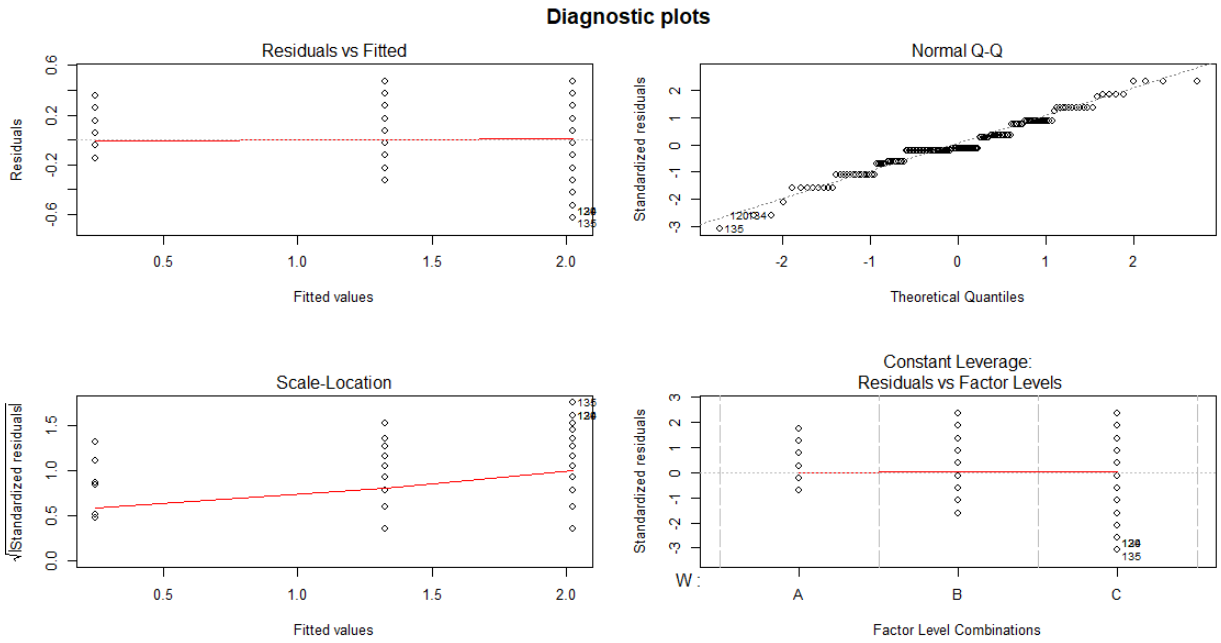
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
W	2	63.21	31.606	138.70	< 2e-16	***
Z	2	5.91	2.957	12.97	6.55e-06	***
Residuals	145	33.04	0.228			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

By running the two-way Anova model, we observe from the table above that both factor W and Z have a significant and really small p-value meaning that we reject H0 that all the means of each category are equal.

Shapiro-Wilk normality test
data: fit$residuals
W = 0.98249, p-value = 0.05336

The Shapiro Wilk normality test provides information of normality in residuals of the model (p-value=0.053>a if a=5%)



Lastly, the diagnostic plots provide information of linear relationship from the first plot, normal distributed residuals from the second plot, homoskedasticity from the third plot and that there are no outliers and influential points from the last plot. Hence, there is no need for alternatives.