Assignment Parinya_Sodsai 47817283

Question 1

```
traffic <- read.csv('data/traffic.csv')
head(traffic)
```

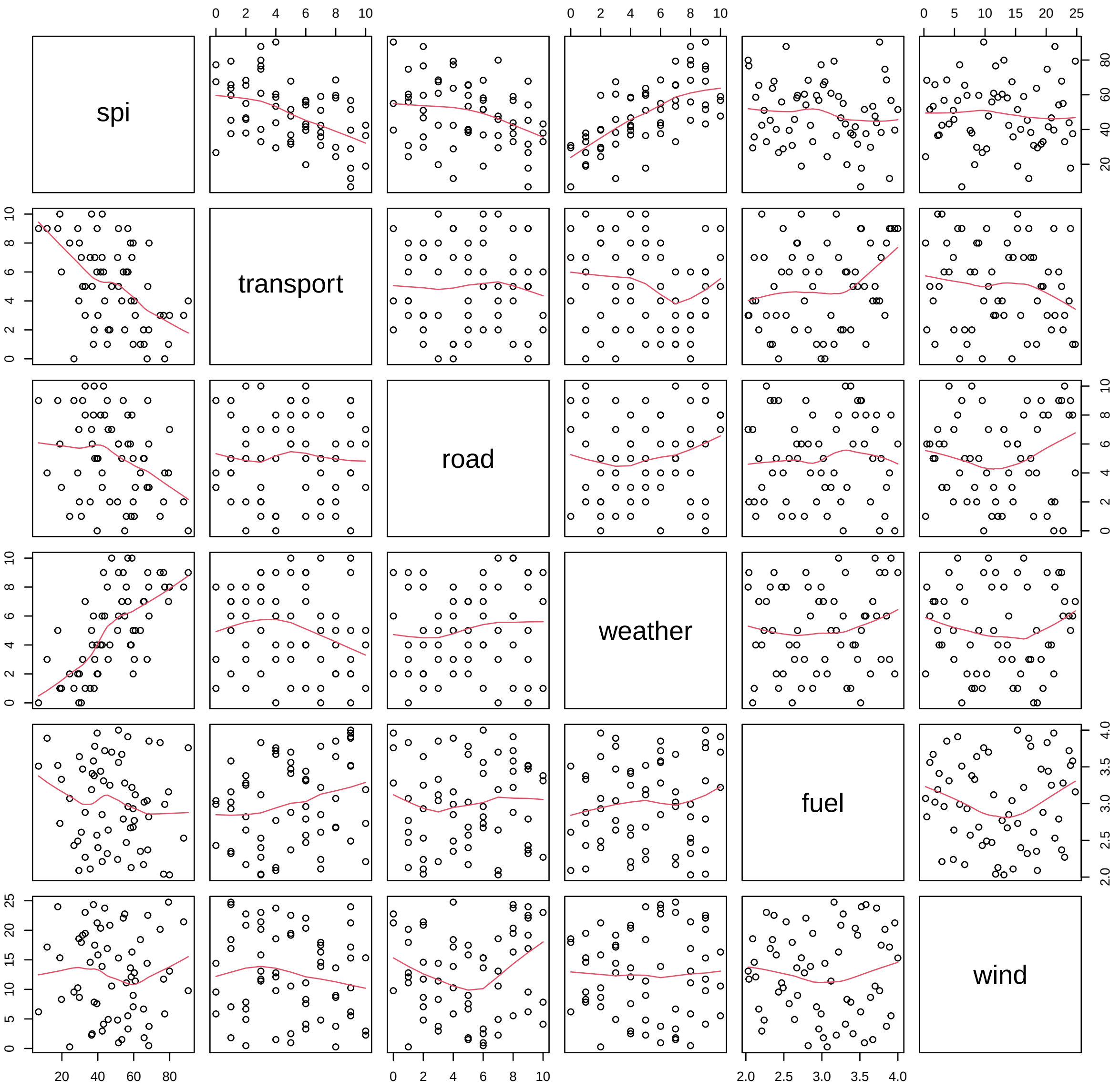
a.Visualization and Correlation Analysis of Data

```
##     spi transport road weather fuel  wind
## 1 39.48         6    5       4 2.57  7.58
## 2 36.87         5    6       4 3.41  2.49
## 3 47.72         5    7      10 3.70 10.56
## 4 58.17         8    6       4 2.67 13.64
## 5 60.33         4    1       3 2.77 12.80
## 6 76.61         3    2       9 2.04 11.73
```

Scatter Plots: To visualize relationships between variables, scatter plots with smoothing lines can provide insights:

```
library(ggplot2)
pairs(traffic, panel = panel.smooth, main="Scatter Plot Matrix")
```

Scatter Plot Matrix



```
#To show correlation matrix
cor(traffic)
```

Correlation Matrix:

```
##                   spi    transport         road     weather         fuel
## spi        1.00000000 -0.472909967 -0.303836850  0.66672345 -0.138153417
## transport -0.47290997  1.000000000 -0.005714728 -0.16971072  0.240947972
## road      -0.30383685 -0.005714728  1.000000000  0.12495993  0.043675635
## weather    0.66672345 -0.169710717  0.124959926  1.00000000  0.110531767
## fuel      -0.13815342  0.240947972  0.043675635  0.11053177  1.000000000
## wind      -0.03466263 -0.131014749  0.080481857  0.00751783  0.006532832
##                  wind
## spi       -0.034662632
## transport -0.131014749
## road       0.080481857
## weather    0.007517830
## fuel       0.006532832
## wind       1.000000000
```

Observations:

Relationships between the response (spi) and predictors:

1. **Weather**: A strong positive correlation with `spi` ($r = 0.667$). This suggests that as weather conditions worsen, traffic congestion (`spi`) tends to increase.
2. **Transport**: A moderate negative correlation with `spi` ($r = -0.473$), indicating that as transport facilities improve or increase, `spi` might decrease.
3. **Road**: A weak negative correlation with `spi` ($r = -0.304$). This might suggest that better road conditions or more roads lead to a decrease in traffic congestion.
4. **Fuel**: A slight positive correlation with `spi` ($r = 0.138$), suggesting that as fuel prices or availability change, it might have a minor effect on `spi`.
5. **Wind**: An almost negligible correlation with `spi` ($r = -0.035$), suggesting that wind conditions don't have a significant impact on traffic congestion.

Relationships between predictors:

1. **Transport vs. Weather**: A mild negative correlation ($r = -0.1697$), suggesting that as transport facilities improve, there might be a decrease in adverse weather events or vice versa.
2. **Transport vs. Fuel**: A mild positive relationship ($r = 0.2409$), which might indicate that as transport facilities increase, so does fuel consumption or availability.
3. **Transport vs. Wind**: A weak negative correlation ($r = -0.1310$), suggesting a minimal relationship between transport and wind conditions.

Question 1b: Linear Regression Model

```r
#Run linear model by target variable = spi with all predictors
M1 <- lm(spi ~ ., data=traffic)
summary(M1)
```

b. Fit a model using all the predictors to explain the spi response.

```
##
## Call:
## lm(formula = spi ~ ., data = traffic)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -18.1596  -4.9415   0.1278   5.1686  21.7415
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  62.8071     7.4080   8.478 1.27e-11 ***
## transport    -2.1750     0.4611  -4.717 1.63e-05 ***
## road         -2.4097     0.4365  -5.520 9.04e-07 ***
## weather       4.2456     0.4473   9.492 2.92e-13 ***
## fuel         -3.6145     2.2759  -1.588    0.118
## wind         -0.1358     0.1764  -0.769    0.445
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 9.913 on 56 degrees of freedom
## Multiple R-squared:  0.7405, Adjusted R-squared:  0.7174
## F-statistic: 31.96 on 5 and 56 DF,  p-value: 3.039e-15
```

b. Using the full model, estimate the impact of weather on spi. Do this by producing a 95% confidence interval that quantifies the change in spi for every one index value increase of weather and comment.

- To estimate the impact of weather on `spi`, we compute a 95% confidence interval for the `weather` coefficient.

The formula used for this is:

$$\hat{\beta}_{\text{weather}} \pm t_{n-p,1-\alpha/2} \times s.e.(\hat{\beta}_{\text{weather}})$$

```r
# 95% confidence interval for the weather coefficient manually
summary.M1 <- summary(M1)
se <- sqrt(diag(summary.M1$cov.unscaled * summary.M1$sigma^2))[4]
t_crit <- qt(0.975, df=56)

upper_bound_CI <- 4.2456  + t_crit * se
lower_bound_CI <- 4.2456 - t_crit * se

cat("Manual Calculation of 95% CI for weather coefficient:\n")
```

Manual Calculation of 95% CI for weather coefficient:

```r
cat("Lower Bound:", round(lower_bound_CI, 2), "\n")
```

Lower Bound: 3.35

```r
cat("Upper Bound:", round(upper_bound_CI, 2), "\n\n")
```

Upper Bound: 5.14

We can also compute this confidence interval using R’s built-in `confint` function:

```r
cat("Using R's confint function:\n")
```

Using R’s confint function:

```r
print(confint(M1, "weather", level=0.95))
```

```
       2.5 %   97.5 %
```

weather 3.349648 5.141639

Conclusion We are 95% confident that for every one index value increase in the `weather` predictor, the `spi` (Speed Performance Index) will increase between 3.35 and 5.14 units on average, keeping all other predictors constant.

Question 1c: Conduct an F-test for the overall regression i.e. is there any relationship between response and the predictors. In your answer:

c1. Mathematical Multiple Regression Model

- **The formula of our multiple regression model is:**

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \epsilon_i, \quad i = 1, 2, ...n$$

Where: Y represents the `spi` (response variable).

- X_{ij} are the predictor variables:
  - X_{i1}: Public Transportation Accessibility
  - X_{i2}: Road Capacity Index
  - X_{i3}: Weather Severity Index
  - X_{i4}: Fuel Price (in USD per gallon)
  - X_{i5}: Average Wind speed (in mph)
- $\epsilon \sim N(0, \sigma^2)$ denotes the random error with constant variance.

c2. Hypotheses for the Overall ANOVA Test

- **Null Hypothesis (H_0):**

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

- **Alternative Hypothesis (H_a):**

$$\text{At least one } \beta_i \neq 0 \quad \text{for } i = 1, 2, \ldots, 5$$

c3. ANOVA Table

```
anova(M1)
```

```
## Analysis of Variance Table
##
## Response: spi
##           Df Sum Sq Mean Sq F value    Pr(>F)
## transport  1 4742.6  4742.6 48.2656 4.228e-09 ***
## road       1 1992.7  1992.7 20.2800 3.441e-05 ***
## weather    1 8651.9  8651.9 88.0507 4.355e-13 ***
## fuel       1  258.1   258.1  2.6264    0.1107
## wind       1   58.2    58.2  0.5921    0.4449
## Residuals 56 5502.6    98.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c4. Computation of the F-Statistic

- We compute the F-statistic to determine the significance of the model:

```
df1 <- 5  #Reg df
df2 <- 56 #Residual df

sum_sq_values <- c(4742.6, 1992.7, 8651.9, 258.1, 58.2) # to find Regression SS
mean_sq_regression <- sum(sum_sq_values) / df1
mean_sq_residuals <- 98.3 #from anova table
f_obs <- mean_sq_regression / mean_sq_residuals

alpha <- 0.05
f_crit <- qf(1 - alpha, df1, df2)

cat("F-statistic (Observed):", round(f_obs, 4), "\n")
```

F-statistic (Observed): 31.9502

```
cat("Critical F-value:", round(f_crit, 4), "\n\n")
```

Critical F-value: 2.3797

c5. Null Distribution for the Test Statistic

- The null distribution for our test statistic is $F(5, 56)$.

c6. P-Value Computation

The p-value associated with our F-statistic is:

```
p_value <- 1 - pf(f_obs, df1, df2)
cat("P-value:", format(p_value, scientific=FALSE), "\n")
```

P-value: 0.000000000000003108624

c7. Conclusions (P-value < 0.05)

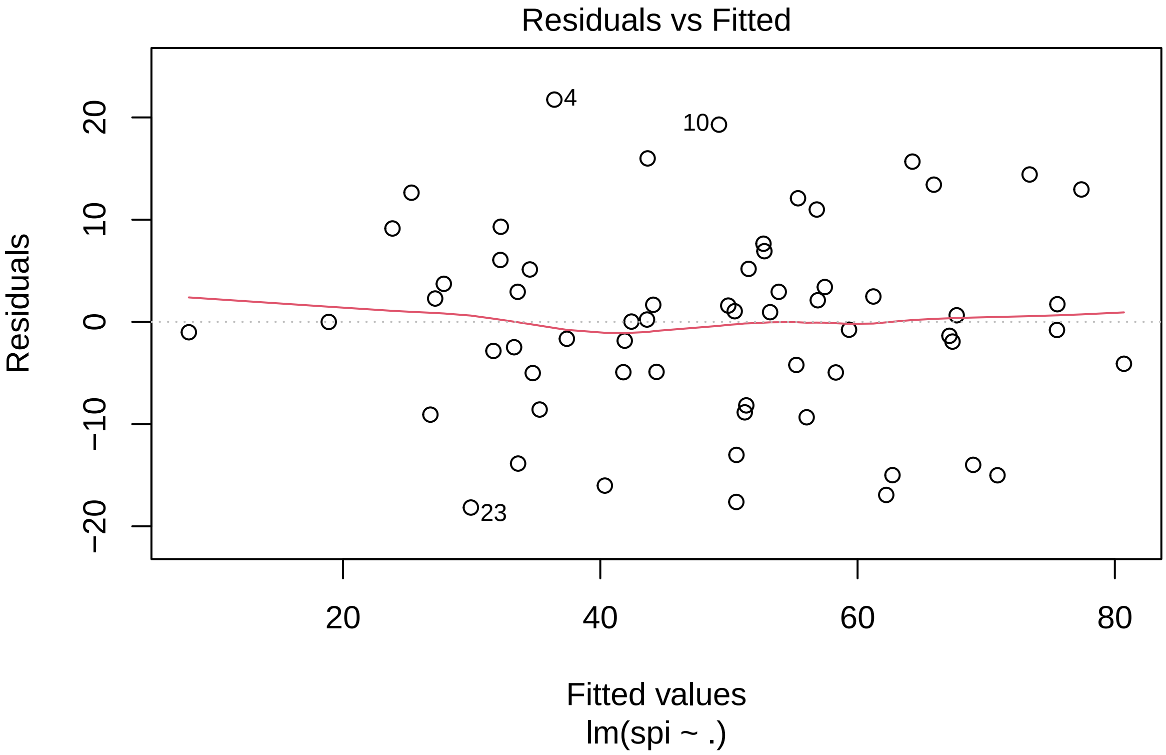
- **Statistical Conclusion:** There is enough evidence to reject H0 at 5% level of significance.
- **Contextual Conclusion:** There's a significant linear relationship between `spi` and at least one of the 5 predictor variables in our model.

Question 1d: Validate the full model and comment on whether the full regression model is appropriate to explain the `spi`

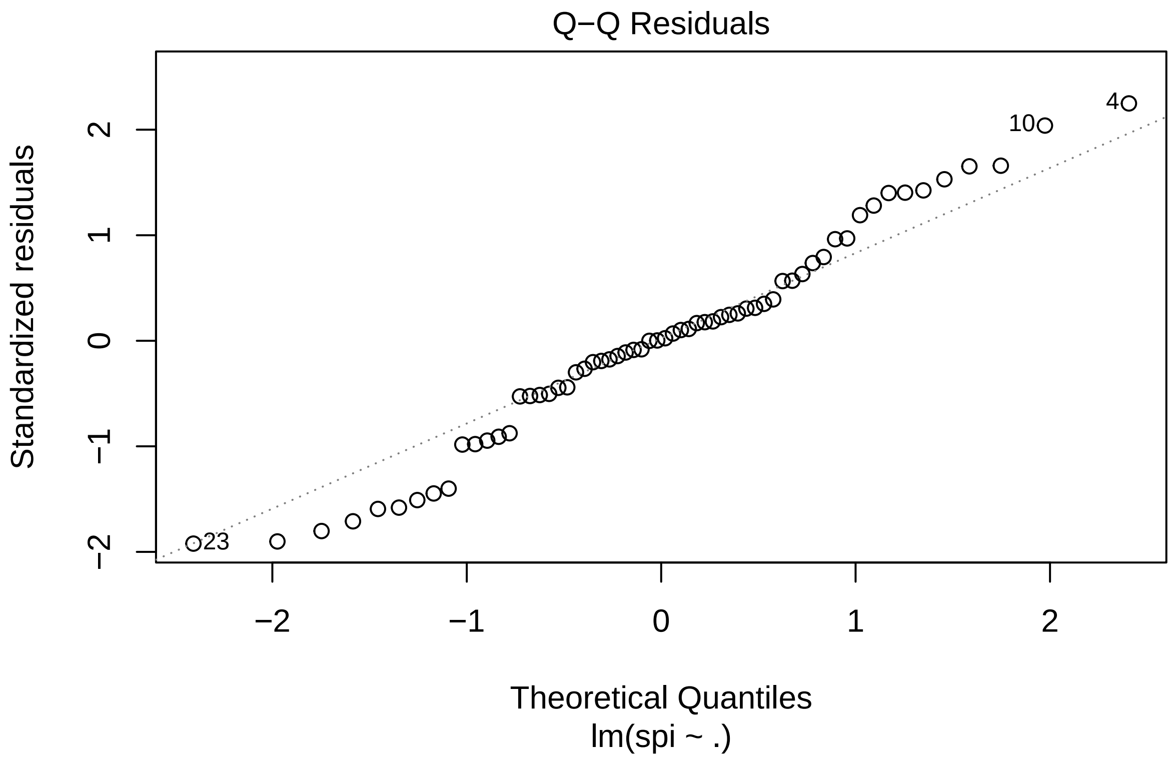
To validate the appropriateness of our full regression model in explaining `spi`, we'll visually inspect various diagnostic plots:

1. **Residuals vs Fitted**: This plot helps us check the assumptions of linearity and homoscedasticity (constant variance of residuals).
2. **Q-Q Plot**: This plot helps us assess the normality of residuals.
3. **Residuals vs Predictors**: By plotting residuals against each predictor, we can further diagnose potential non-linearity or other issues specific to each predictor.

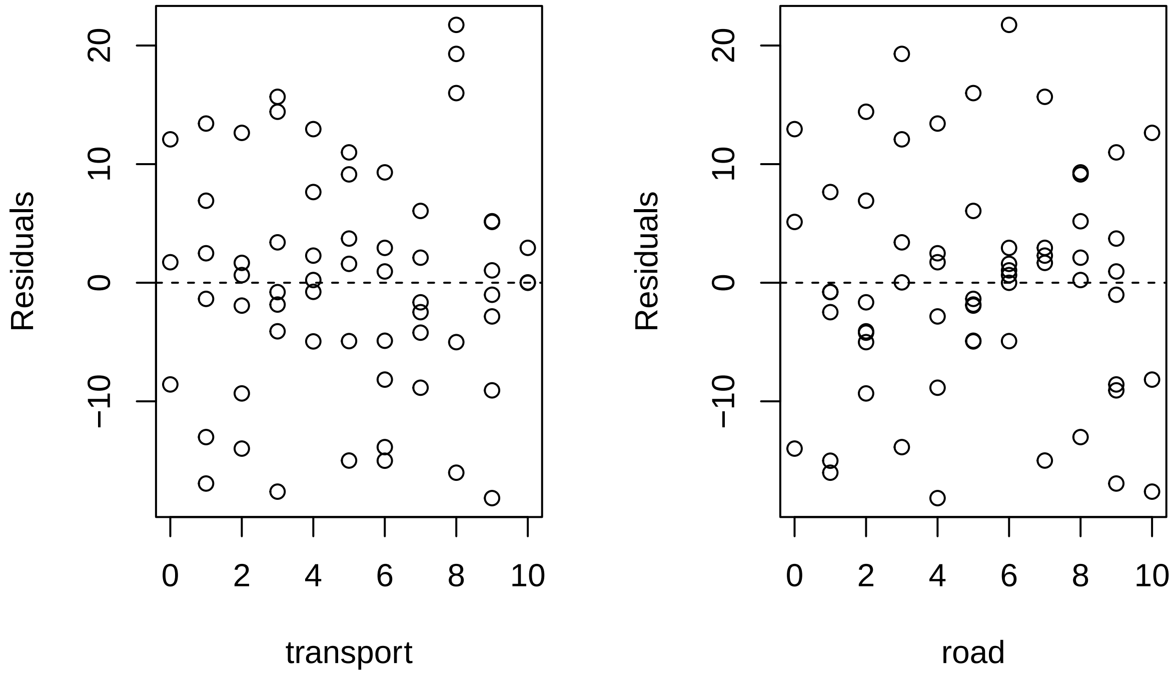
```
#Plot residuals vs fitted values (to check for homoscedasticity and linearity)
plot(M1, which = 1)
```



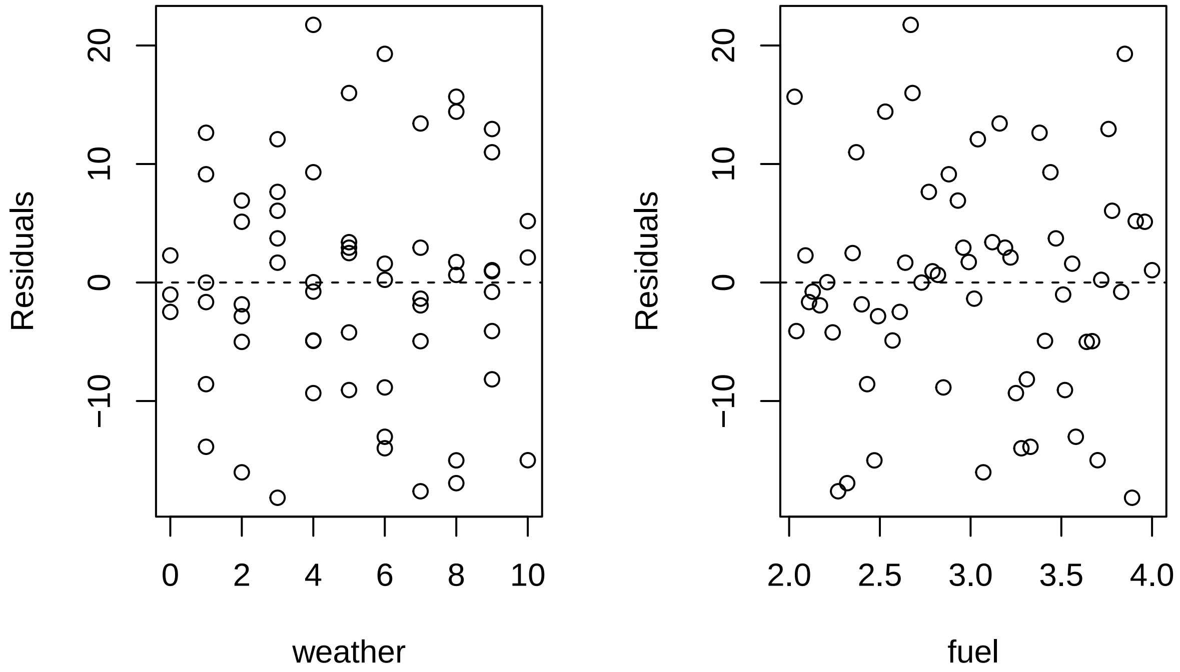
```
#Plot the Q-Q plot (to check for normality of residuals)
plot(M1, which = 2)
```



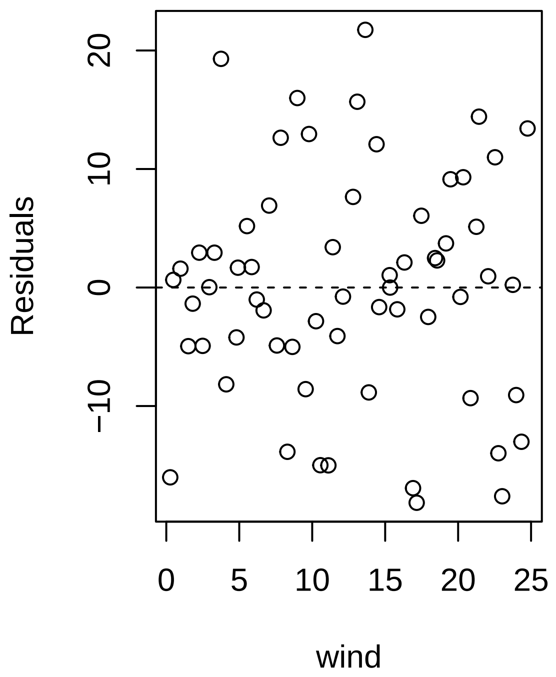
```
#Set up a 1x2 plotting area
par(mfrow = c(1, 2))

#Plot residuals against each predictor
plot(resid(M1) ~ transport, data = traffic, xlab = "transport", ylab = "Residuals")
abline(h = 0, lty = 2)

plot(resid(M1) ~ road, data = traffic, xlab = "road", ylab = "Residuals")
abline(h = 0, lty = 2)
```



```r
plot(resid(M1) ~ weather, data = traffic, xlab = "weather", ylab = "Residuals")
abline(h = 0, lty = 2)

plot(resid(M1) ~ fuel, data = traffic, xlab = "fuel", ylab = "Residuals")
abline(h = 0, lty = 2)
```



```r
plot(resid(M1) ~ wind, data = traffic, xlab = "wind", ylab = "Residuals")
abline(h = 0, lty = 2)
```



Observations:

- **Linearity and Homoscedasticity**: The random scatter in the "Residuals vs Fitted" plot suggests that there is no pattern.This random scatter shows that there is almost a linear relationship between the predictors and the target variable.
- **Normality of Residuals**: The residuals in the Q-Q plot, although largely aligning with the 45-degree reference line in the middle of the distribution, show deviations that suggest they **are not** be normally distributed. These deviations could arise from unnecessary predictors (with insignificant p-values in the full model). Particularly, the tails, especially in the upper tail (top right) and the lower tail (bottom left), deviate from the reference line. This deviation indicates that the residuals might exhibit some skewness or contain potential outliers, both of which are points of concern for the model's predictions.
- **Residuals vs Predictors**: The scatter plots of residuals against each predictor show no patterns, showing that our confidence in the model's linearity assumption for each predictor.

Question 1e: Find the R^2 and comment on what it means in the context of this dataset.

```
model_summary <- summary(M1)
r_squared <- model_summary$r.squared
print(paste("R^2 =", round(r_squared, 4)))
```

```
## [1] "R^2 = 0.7405"
```

Interpretation: An R^2 is equal to 74% which mean 74% of the changes in traffic congestion (`spi`) can be explained by our chosen predictors. On the other hand, the remaining 26% of variability in `spi` remains unexplained by our model.

Question 1f: Using model selection procedures discussed in the unit, find the best multiple regression

To determine the best multiple regression model for explaining our data, we will use a **stepwise backward** elimination approach based on the p-values of predictors. Starting with the full model, we will remove predictors with the highest p-values that are not significant at our chosen level ($\alpha = 0.05$) until all remaining predictors are significant.

```r
####1. Initial Full Model
summary(M1)
```

```
##
## Call:
## lm(formula = spi ~ ., data = traffic)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.1596  -4.9415   0.1278   5.1686  21.7415
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  62.8071     7.4080   8.478 1.27e-11 ***
## transport    -2.1750     0.4611  -4.717 1.63e-05 ***
## road         -2.4097     0.4365  -5.520 9.04e-07 ***
## weather       4.2456     0.4473   9.492 2.92e-13 ***
## fuel         -3.6145     2.2759  -1.588    0.118
## wind         -0.1358     0.1764  -0.769    0.445
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 9.913 on 56 degrees of freedom
## Multiple R-squared:  0.7405, Adjusted R-squared:  0.7174
## F-statistic: 31.96 on 5 and 56 DF,  p-value: 3.039e-15
```

From the above results, the predictor `wind` has the highest p-value, which exceeds our significance threshold of 0.05. Thus, we will eliminate it first and refit the model.

```r
#2. Update the model without wind predictor
M2 <- update(M1, . ~ . - wind)
summary(M2)
```

```
##
## Call:
## lm(formula = spi ~ transport + road + weather + fuel, data = traffic)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.9347  -4.2440   0.0528   5.0544  21.4515
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  61.1610     7.0669   8.655 5.69e-12 ***
## transport    -2.1257     0.4550  -4.672 1.86e-05 ***
## road         -2.4372     0.4335  -5.622 5.92e-07 ***
## weather       4.2565     0.4454   9.555 1.94e-13 ***
## fuel         -3.6853     2.2659  -1.626    0.109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.877 on 57 degrees of freedom
## Multiple R-squared:  0.7378, Adjusted R-squared:  0.7194
## F-statistic: 40.09 on 4 and 57 DF,  p-value: 5.959e-16
```

In this updated model, the predictor `fuel` now has the highest p-value that is greater than 0.05. We will eliminate this predictor next.

```r
#3. Model without `wind` and `fuel`
M3 <- update(M2, . ~ . - fuel)
summary(M3)
```

```
##
## Call:
## lm(formula = spi ~ transport + road + weather, data = traffic)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.672  -5.643   1.067   4.656  23.164
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  51.7370     4.1027  12.611  < 2e-16 ***
## transport    -2.3216     0.4449  -5.218 2.54e-06 ***
## road         -2.4563     0.4394  -5.590 6.40e-07 ***
## weather       4.1450     0.4463   9.286 4.48e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.02 on 58 degrees of freedom
## Multiple R-squared:  0.7256, Adjusted R-squared:  0.7114
## F-statistic: 51.12 on 3 and 58 DF,  p-value: 2.724e-16
```

All predictors are significant and we should kept M3 for the final model. The final (fitted) model equation is

$$\hat{spi} = 51.737 - 2.322 \times \textbf{transport} - 2.456 \times \textbf{road} + 4.145 \times \textbf{weather}$$

Question 1g: Comment on the R^2 and adjusted R^2 in the full and final model you chose in part f.In particular explain why those goodness of fitness measures change.

When analyzing a model, we look at R^2 value to check how well our predictors explain the variations in response variable.However, adding more predictors can increase R^2, even if new predictors aren't significant. Adjust R^2 will provide more balanced view to ensure that we are not just adding unnecessary predictors.

- **Full Model (M1)**:
  - $R^2 = 74\%$
  - Adjusted $R^2 = 71.74\%$
- **Final Model (M3)**: After the removal of the `wind` and `fuel` predictors:
  - R^2 dipped to around 72%
  - Adjusted R^2 settled at approximately 71.14%

The decline in R^2 is expected since using fewer predictors to explain variability in `spi`.However, the comparable levels of adjusted R^2 between M1 and M3 indicate that even though we removed some predictors, the final model (M3) retains nearly the same variance as the full model(M1), but with fewer predictors.This indicates that the model's performance is maintained, showing the robustness of the significant predictors.

Question 2a.For this study, is the design balanced or unbalanced? Explain why.

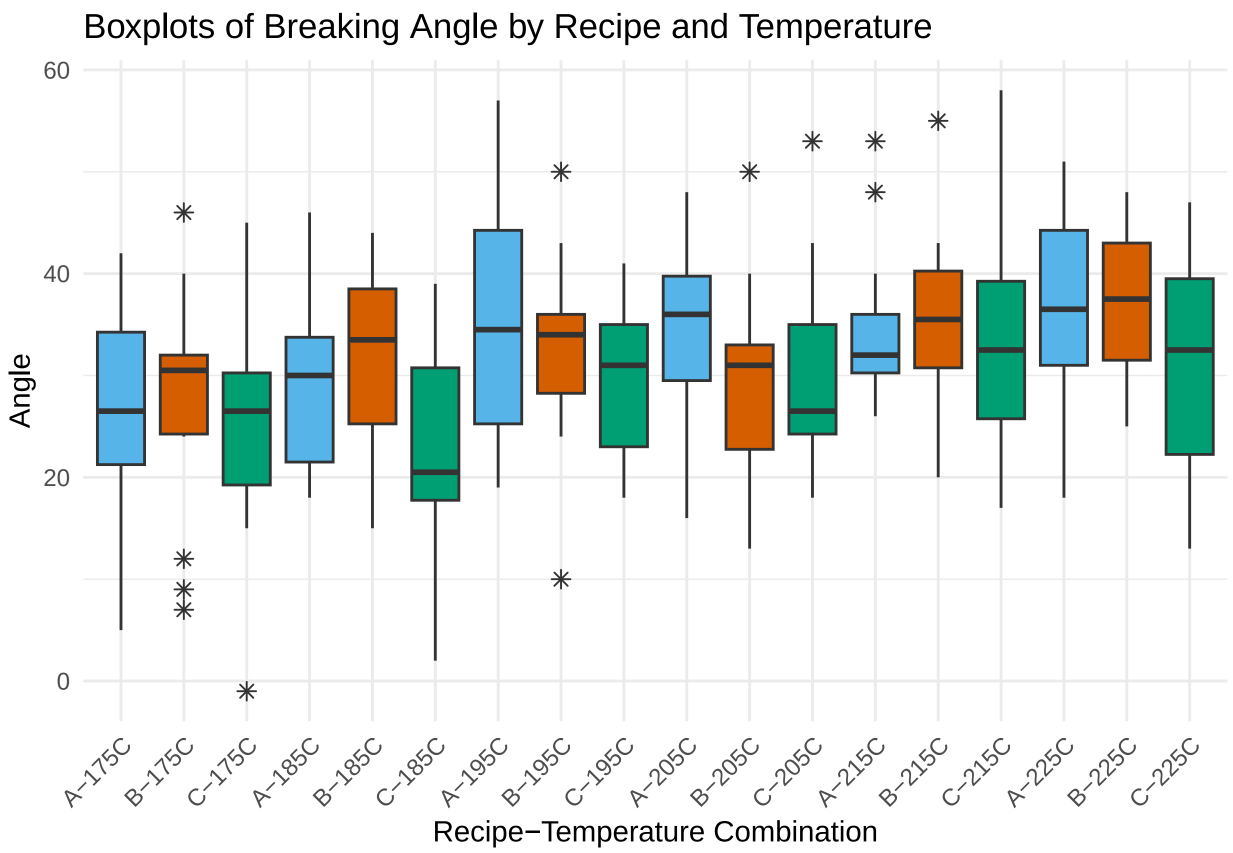
```r
cake <- read.csv("data/cake.csv")
# Create a count table of Recipe and Temp.
count_table <- table(cake$Recipe, cake$Temp)
print(count_table)
```

```
##
##     175C 185C 195C 205C 215C 225C
##   A   14   14   14   14   14   14
##   B   14   14   14   14   14   14
##   C   14   14   14   14   14   14
```

Conclusion: This design is balanced because every level of combination of `'Recipe'` and `'Temp'` has an equal number of observations.

Question 2b.Construct two different preliminary graphs that investigate different features of the data and comment.

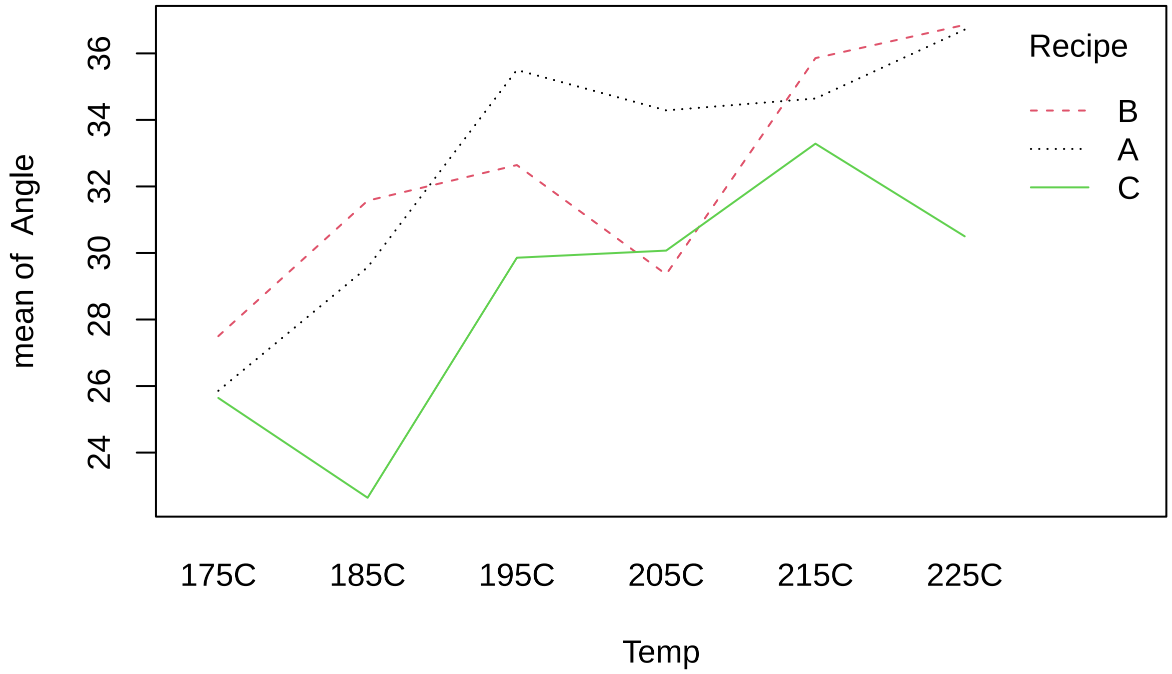
```r
library(ggplot2)
# Define custom colors
colors <- c("#56B4E9", "#D55E00", "#009E73", "#F0E442")

# Enhanced boxplot
ggplot(cake, aes(x=interaction(Recipe, Temp, sep = "-"), y=Angle, fill=Recipe)) +
  geom_boxplot(outlier.shape = 8, outlier.size = 2) +  # Customize outliers
  labs(title="Boxplots of Breaking Angle by Recipe and Temperature",
       x="Recipe-Temperature Combination", y="Angle") +
  scale_fill_manual(values=colors) +
  theme_minimal() +
  theme(legend.position="none", axis.text.x=element_text(angle=45, hjust=1))
```



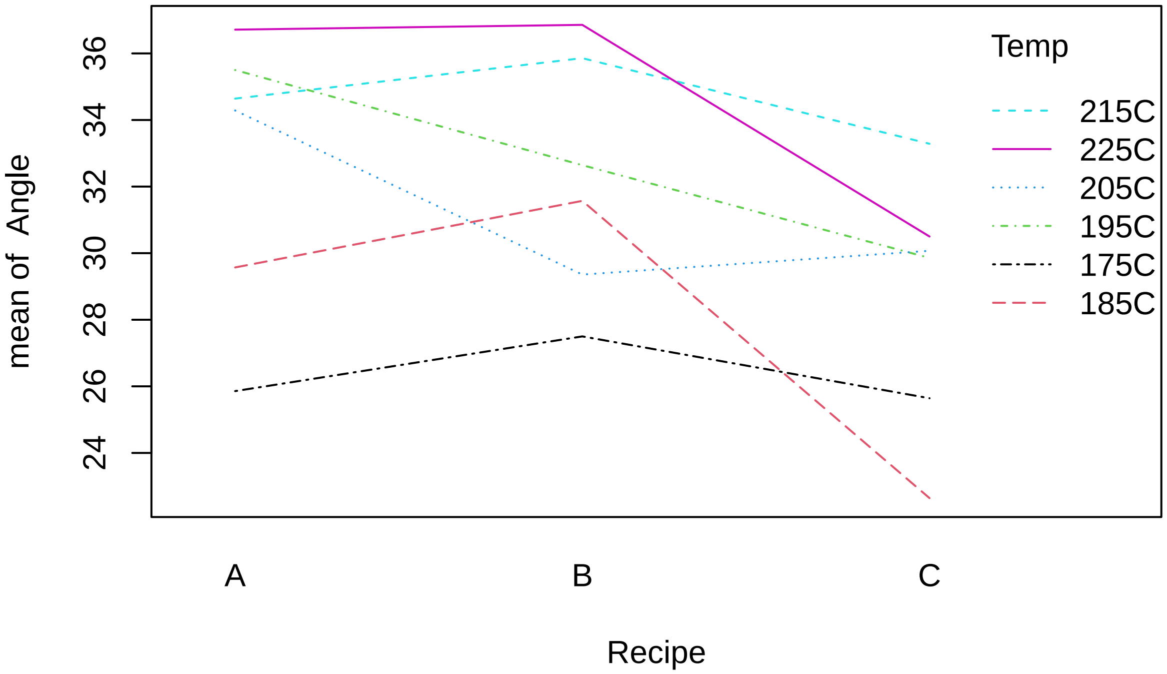
From box plot,

1. We can observe that the median Angle tends to increase with increasing temperature for all recipes.
2. There are some potential outliers in some recipes

```
with(cake, interaction.plot(Temp, Recipe, Angle, col = 1:8))
```



```
with(cake, interaction.plot(Recipe, Temp, Angle, col = 1:8))
```



From both interaction plots, the lines for each line charts seem like **non-parallel** lines.This suggests a potential interaction effect, **but** we need to run anova to recheck the potential of interaction.

Question 2C. Write down the full interaction model for this situation, defining all appropriate parameters.

c. The full two-way ANOVA model with interaction is:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

Where:

- y_{ijk}: the breaking angle for the cake.
- α_i:the `temperature` effect, there are 5 levels - 175C, 185C, 195C, 205C, 215C and 225C.
- β_j: the `recipe` effect, there are 3 levels - A, B and C.
- γ_{ij}: interaction effect, between `temperature` and `recipe`.
- ϵ_{ijk}: is the unexplained variation.

Question 2d.Analyse the data to study the effect of Temp and Recipe on breaking Angle of cake at 5% significance level. Remember to

d1.state the null and alternative hypothesis for each test

Hypotheses:

- H_0: $\gamma_{ij} = 0$ for all i, j. (There's no interaction between Temperature and Recipe.)
- H_a: At least one $\gamma_{ij} \neq 0$. (There's an interaction between Temp and Recipe)

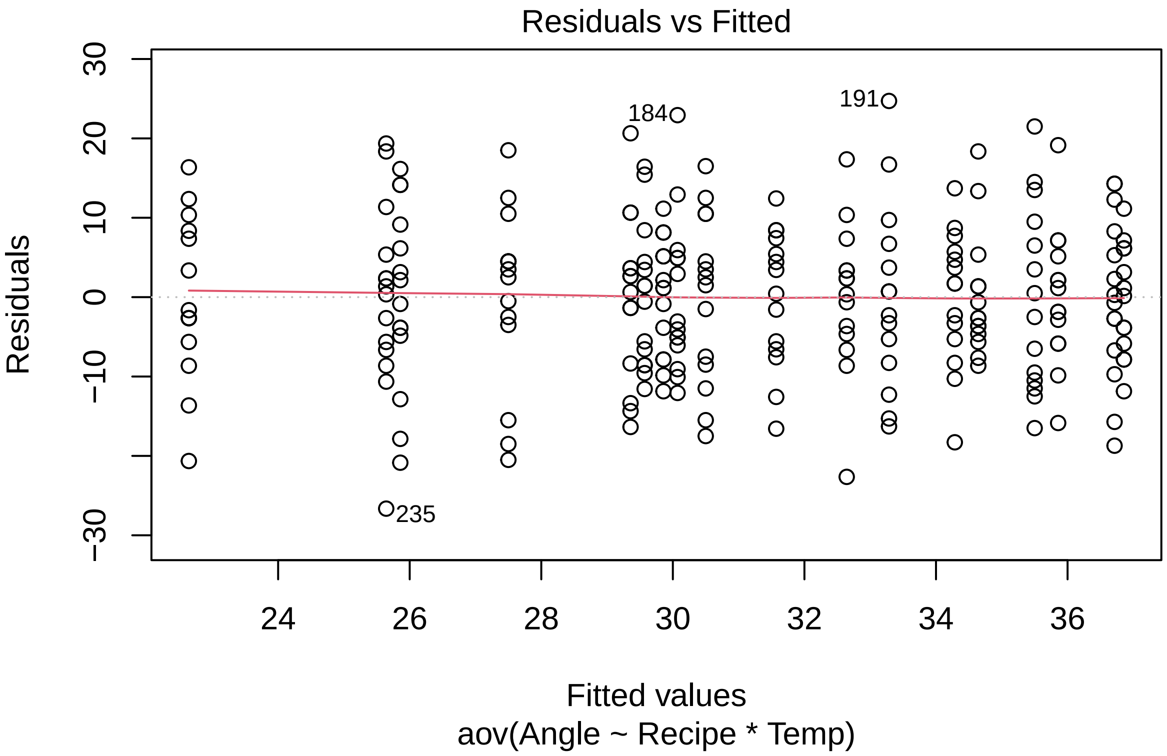
```
# Conduct the Two-Way ANOVA
cake.int <- aov(Angle ~ Recipe * Temp, data=cake)
summary(cake.int)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)    
## Recipe         2    845   422.4   4.276 0.014998 *  
## Temp           5   2530   506.0   5.123 0.000177 ***
## Recipe:Temp   10    636    63.6   0.643 0.775632    
## Residuals    234  23114    98.8                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

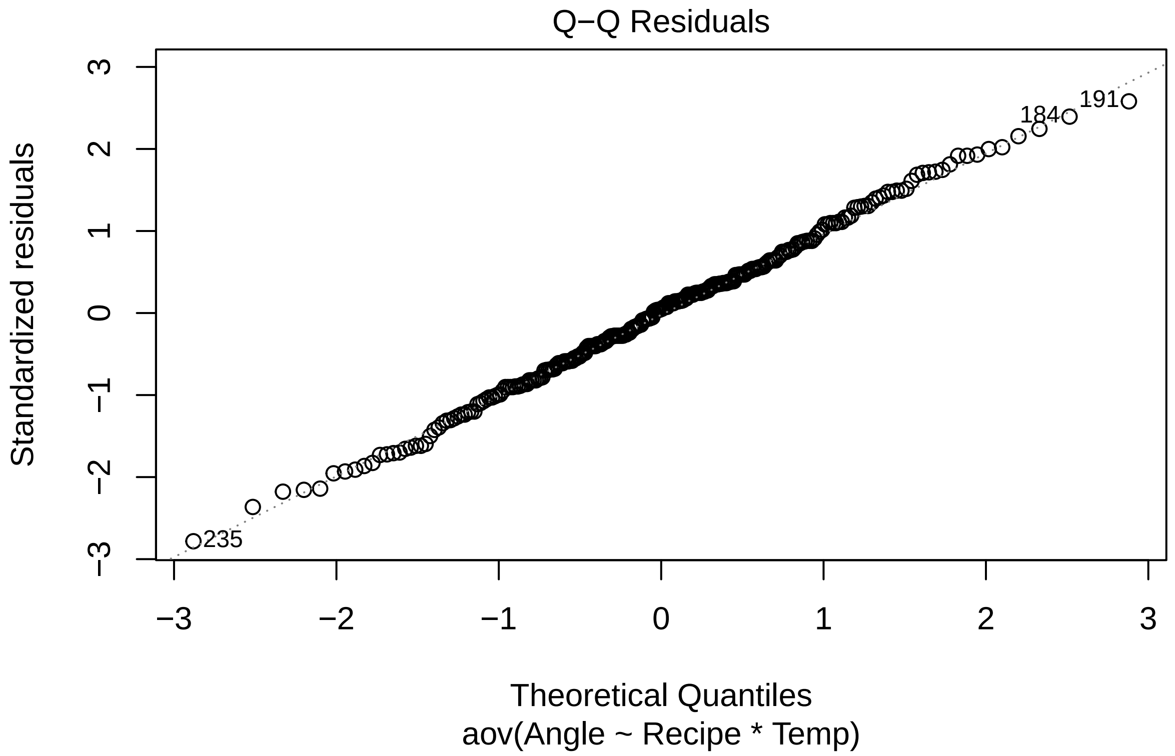
The p-value for the interaction term is much greater than 0.05. Therefore, we **fail** to reject H0 for the interaction effect. This suggests that there's no significant interaction between the Recipe and Temperature in terms of their combined effect on the breaking angle.

d2.check assumptions (model diagnostics)

```
#To diagnostic plots.

plot(cake.int, which = 1)
```



```r
#To diagnostic plots.
plot(cake.int, which = 2)
```



d3.and interpret the results.

- **The residuals plot** are seem to show equal spread around the fitted value and so the constant variance assumption is also appropriate.
- **The normal Q-Q plot** shows that the residuals lie reasonably close to the line.so, the normal assumption should be valid.

Question 2e. Repeat the above test analysis for the main effects.

2e. The full two-way ANOVA model without interaction is:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

Where:

- y_{ijk}: the breaking angle for the cake.
- α_i:the `temperature` effect, there are 5 levels - 175C, 185C, 195C, 205C, 215C and 225C.
- β_j: the `recipe` effect, there are 3 levels - A, B and C.
- ϵ_{ijk}: is the unexplained variation.

Main effects: Recipe

Hypotheses:

- H_0: $\beta_j = 0$ against
- H_a: At least one $\beta_j \neq 0$.

Main effects: Temperature

Hypotheses:

- H_0: $\alpha_i = 0$ against
- H_a: At least one $\alpha_i \neq 0$.

```
# Conduct the Two-Way ANOVA without interaction effect
cake.aov = aov(Angle ~ Temp * Recipe, data=cake)
cake.aov.2 = update(cake.aov, . ~ . - Temp:Recipe)
anova(cake.aov.2)
```

```
## Analysis of Variance Table
##
## Response: Angle
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Temp        5  2530.1  506.01  5.1988 0.0001489 ***
## Recipe      2   844.8  422.38  4.3396 0.0140636 *
## Residuals 244 23749.4   97.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect of Recipe:

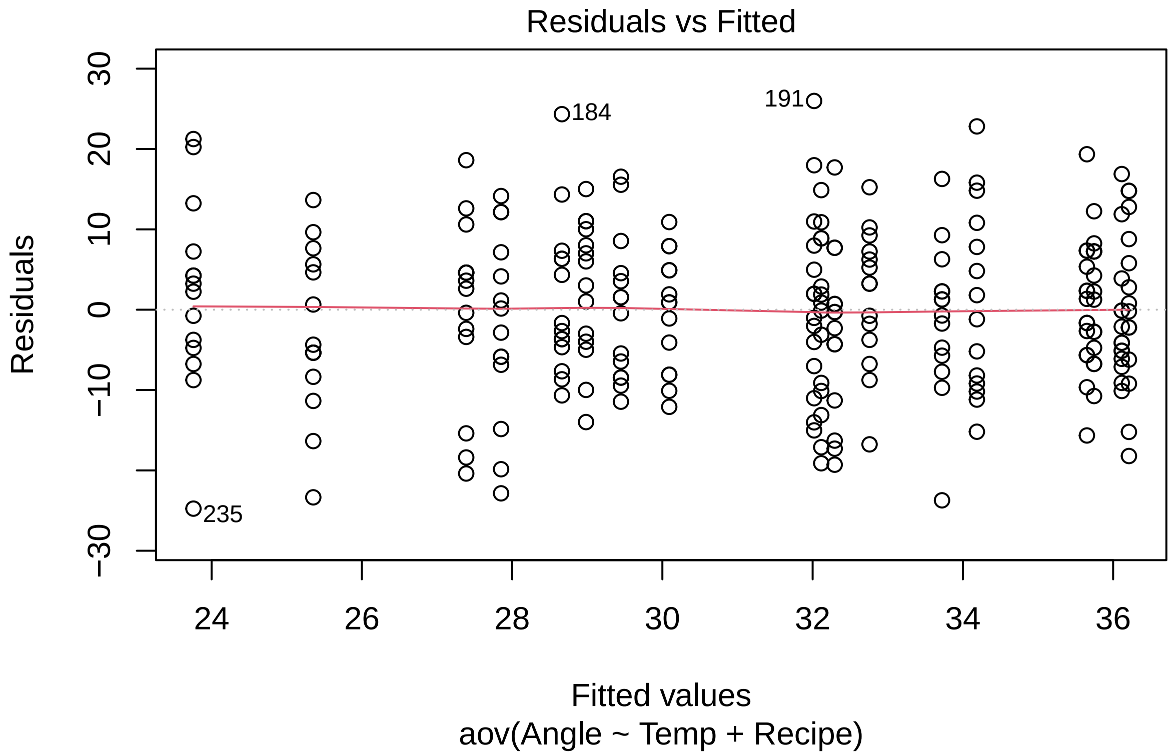
p-value: 0.0140636 .As the p-value is less than 0.05, we reject the null hypothesis for the Recipe effect. This indicates that the `recipe` has a statistically significant effect on the breaking angle.

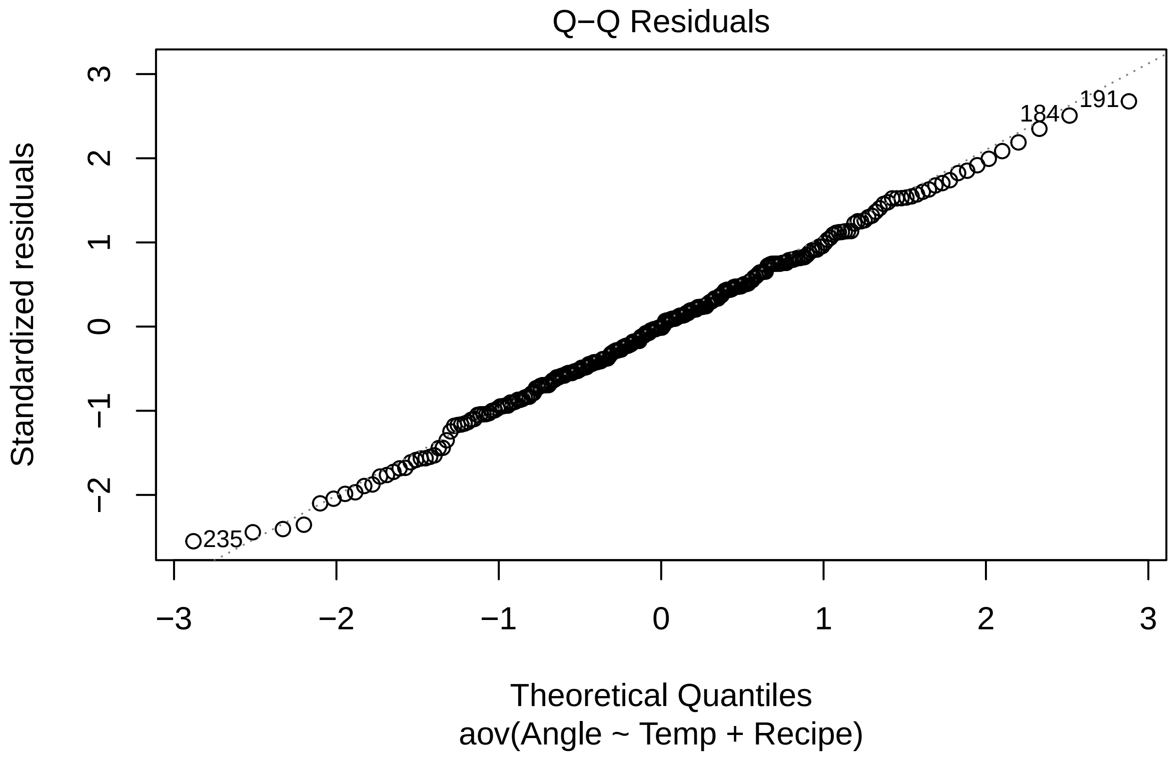
Effect of Temperature:

p-value: 0.000196 .Since the p-value is extremely small, we reject the null hypothesis for the Temperature effect. This indicates that the baking temperature has a statistically significant effect on the breaking angle.

In conclusion, both the recipe and the baking temperature independently influence the breaking angle of the cake.

```
#To diagnostic plots.
plot(cake.aov.2, which =  1:2)
```


Residuals vs Fitted
Residuals
30
20
10
0
−10
−30
24
26
28
30
32
34
36
Fitted values
aov(Angle ~ Temp + Recipe)
184
191
235



d3.and interpret the results.

- **The residuals plot**: are seem to show equal spread around the fitted value and so the constant variance assumption is also appropriate.

- **The normal Q-Q plot**: shows that the residuals lie reasonably close to the line.so, the normal assumption should be valid.

Question 2f. State your conclusions about the effect of Temp and Recipe on the Angle response. These conclusions are only required to be at the qualitative level and can be based off the outcomes of the hypothesis tests in parts d. and e. and the preliminary plots in b.. You do not need to statistically examine the multiple comparisons between contrasts and interactions.

1.Effect of Temperature (Temp) on Angle:

- Based on the preliminary box plot (b1), as the temperature increases, the median angle seems to show an upward trend for all recipes.
- The ANOVA results (from e) confirm this observation by indicating that the effect of temperature on the angle is statistically significant with a very small p-value(by individual).

2.Effect of Recipe on Angle:

- From the ANOVA results (from e), the effect of the recipe on the angle is statistically significant. This means that different recipes produce cakes with different average angles(by individual).

3.Interaction Effect:

- The interaction plots (b2) suggest a **potential** interaction between temperature and recipe, as indicated by the non-parallel lines.
- **However,** the ANOVA results (from d) indicate that this interaction is not statistically significant. This means that while there might appear to be an interaction visually, it's not strong enough to be detected as significant in this dataset.

4.Residual Analysis:

- The residuals plot shows equal spread around the fitted value, suggesting that the constant variance assumption is met.
- The normal Q-Q plot indicates that the residuals are roughly normally distributed, which supports the normality assumption for ANOVA.

5.Potential Outliers:

- The box plot (b1) suggests that there might be some outliers in the dataset. While these outliers could influence the results, the ANOVA is generally robust to a few outliers if the sample size is large.