HW2

109550121 温柏萱

Task1

Part 1

(a)

Coal

```
> pizza[pizza$heat == "Coal", "rating"]
 [1] 4.89 5.00 4.31 4.99 4.92 4.96 4.05 4.96 4.78 4.90 5.00 4.85 4.99 4.93 4.99 3.63 3.56
> coal_rating <- pizza[pizza$heat == "Coal", "rating"]
> mean(coal_rating)
[1] 4.688824
> sd(coal_rating)
[1] 0.4867479
```

Wood

```
> pizza[pizza$heat == "Wood", "rating"]
 [1] 4.73 2.45 4.87 4.98 4.99 4.99 5.00 2.79 4.89 2.17 3.30 4.96 4.99 2.09 2.10 4.96 4.90 4.99
1.75 4.86 0.37
[22] 4.99 4.96 0.88 4.95
> wood_rating <- pizza[pizza$heat == "Wood", "rating"]
> mean(wood_rating)
[1] 3.8764
> sd(wood_rating)
[1] 1.537248
```

Gas

```
> pizza[pizza$heat == "Gas", "rating"]
  [1] 0.03 0.13 2.62 4.38 0.12 4.09 0.30 4.71 0.15 4.74 0.53 4.85 0.21 4.95 3.81 0.31 4.95 4.76
4.75 4.74 4.79
 [22] 0.07 4.71 4.78 3.06 4.99 0.93 4.85 5.00 0.53 0.18 4.61 0.09 4.34 4.94 4.92 0.90 0.12 1.08
4.92 1.99 3.28
 [43] 0.08 4.69 4.31 4.39 4.93 1.22 3.03 3.33 2.51 3.74 3.75 4.93 4.78 0.03 4.89 3.84 1.46 1.45
1.99 1.41 4.52
 [64] 4.78 4.16 4.60 2.43 0.06 4.53 4.53 0.54 2.05 2.63 0.80 0.43 1.32 4.87 0.40 4.08 0.76 4.80
4.40 4.78 4.26
 [85] 4.03 3.34 0.56 4.20 2.72 3.71 0.31 4.81 1.87 0.86 0.78 4.83 4.42 1.08 4.54 2.35 3.66 0.48
1.71 4.84 0.60
```

```
[106] 4.01 4.64 2.92 0.85 0.16 1.19 0.23 2.75 4.98 4.92 3.27 4.80 4.77 4.95 4.96 0.30 0.81 4.96
4.92 0.75 3.66
[127] 3.42 4.75 4.78 3.00 1.43 0.35 3.98 1.63 3.34 3.83 4.99 3.35 4.09 4.88 4.82 0.17 4.91 0.12
3.88 1.14 1.92
[148] 1.75 4.17 0.65 4.37 4.09 4.94 4.12 3.69 3.14 4.27 0.67
> gas_rating <- pizza[pizza$heat == "Gas", "rating"]
> mean(gas_rating)
[1] 2.961013
> sd(gas_rating)
[1] 1.817251
```

	Coal	Wood	Gas
Mean	4.688824	3.8764	2.961013
Standard Deviation	0.4867479	1.537248	1.817251

Pizza heated by coal are the most popular observed by having the highest rating and the smallest standard deviation, indicating that most customers appreciate pizzea heated by coal and agree with each other. Followed by pizza heated by wood, which has the mean rating as the runner-up and the runner-up standard deviation. The most unrecognized heating method is gas, which has a mean rating of 2.961013, but the mean rating is not well-representing since the standard deviation is large, indicating that not all customers hate pizza heated with gas.

(b)

```
> anova_result <- aov(rating ~ heat, data=pizza)
> summary(anova_result)
             Df Sum Sq Mean Sq F value   Pr(>F)
heat          2     58  29.022   9.875 8.18e-05 ***
Residuals   197    579   2.939
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value associated with the F statistic is $8.18 \times 10^{-5} << 0.05$, indicating there is a statistically significant difference between the average ratings of pizzas based on the heat source used.

The ANOVA test's p-value is highly significant (***), meaning the null hypothesis can be rejected. This suggests that the type of heat source used to bake the pizzas (`Coal`, `Wood`, `Gas`) has significant effect on their ratings. Given the significant difference detected by the ANOVA, it's clear that not all heat sources lead to the same average pizza rating.

(c)

```
> model <- lm(rating ~ heat, data=pizza)
> summary(model)

Call:
lm(formula = rating ~ heat, data = pizza)

Residuals:
   Min     1Q Median     3Q    Max
-3.506 -1.715  0.379  1.562  2.039

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6888     0.4158  11.277  < 2e-16 ***
heatGas      -1.7278     0.4376  -3.948 0.000109 ***
heatWood     -0.8124     0.5389  -1.507 0.133289
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.714 on 197 degrees of freedom
Multiple R-squared:  0.09112, Adjusted R-squared:  0.08189
F-statistic: 9.875 on 2 and 197 DF,  p-value: 8.184e-05
```

Since `R` will create dummy variables for `Wood` and `Gas` automatically, I just applied the `lm` instruction to fit the linear regression model.

$$y = 4.6888 - 0.8124 \times \text{headWood} - 1.7278 \times \text{headGas} \tag{1}$$

- Multiple R^2: Multiple R^2 is 0.09112 is rather small, indicating that there may be some important variables missing.
- Adjusted R^2: The small decrease from the R^2 value suggests the model isn't overfitting.
- F-statistic: The F-statistic of 9.875 with a p-value of 8.184×10^{-5} indicates that the model is statistically significant. This means that there's a relationship between heat source and rating.

The model indicates that the type of heat used for cooking pizza significantly affects its rating, with `Coal` being the baseline for comparison. `Gas` significantly lowers the rating compared to `Coal`, while `Wood` also lowers it but not significantly so. These results suggest that if aiming for higher ratings, `Coal` is the preferable heat source according to the data, with `Wood` being a better choice than `Gas`, albeit not significantly better compared to `Coal`.

(d)

(a) only computed the average rating and standard deviation of each heat source and did not fit a model to it.

(b) and (c) both presented the p-value and showed the same value (8.18×10^{-5}), indicating the existance of the relationship between heat soruce and rating.

(b) tests whether there are any overall differences among the groups without specifying what those differences are. It tells us that not all heat sources lead to the same rating, but it doesn't detail how each one compares to the others directly.

(c) Linear regression quantifies the relationship between the heat source and pizza ratings. It provides specific coefficients that indicate how much the choice of heat source (Wood or Gas) decreases the rating compared to the baseline (Coal). It also privides p-values for each coefficient, directly indicating whether the difference in ratings between Coal and each of the other heat sources (Gas and Wood) is statistically significant.

To sum up, ANOVA detects whether there are any differences in pizza ratings based on the heat source, linear regression provides a more detailed result by quantifying these differences and assessing the significance of each heat source.

Part 2

(e)

```
> model_e <- lm(rating ~ heat + area + cost, data=pizza)
> summary(model_e)

Call:
```

```
lm(formula = rating ~ heat + area + cost, data = pizza)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.98864 -0.52516  0.00599  0.51428  1.92332 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.72260    0.34461   2.097  0.03731 *  
heatGas         -1.59555    0.20526  -7.773 4.52e-13 ***
heatWood        -0.45753    0.26056  -1.756  0.08069 .  
areaEVillage     4.17970    0.24628  16.971  < 2e-16 ***
areaLES          2.37294    0.26106   9.089  < 2e-16 ***
areaLittleItaly  0.78700    0.25268   3.115  0.00212 ** 
areaSoHo         3.65362    0.24498  14.914  < 2e-16 ***
cost             0.43865    0.06613   6.633 3.26e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7957 on 192 degrees of freedom
Multiple R-squared:  0.8092,	Adjusted R-squared:  0.8022 
F-statistic: 116.3 on 7 and 192 DF,  p-value: < 2.2e-16
```

The coefficients for `heatGas` and `heatWood` directly compare the effect of Gas and Wood heat sources on pizza ratings against the baseline category, `heatCoal`. The coefficients of `heatGas` and `heatWood` being negative aligns with the assumption whre pizza heated using coal has the highest rating.

(f)

```
> model_f <- lm(rating ~ heat_re + area + cost, data=pizza)
> summary(model_f)

Call:
lm(formula = rating ~ heat_re + area + cost, data = pizza)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.97759 -0.51011 -0.02969  0.52497  2.15583 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.96212    0.31668   3.038  0.00271 ** 
heat_re         -0.87601    0.09242  -9.479  < 2e-16 ***
areaEVillage     4.10646    0.24378  16.845  < 2e-16 ***
areaLES          2.26091    0.25405   8.900 4.08e-16 ***
areaLittleItaly  0.69163    0.24774   2.792  0.00577 ** 
areaSoHo         3.54383    0.23768  14.910  < 2e-16 ***
cost             0.44911    0.06618   6.786 1.38e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.7997 on 193 degrees of freedom
Multiple R-squared:  0.8062,	Adjusted R-squared:  0.8002
F-statistic: 133.8 on 6 and 193 DF,  p-value: < 2.2e-16
```

The coefficient for `heat_re` assumes a linear relationship between the heating method ranking and the pizza rating. This oversimplifies the distinct effects that each heating method might have on the rating.

```
> new_data <- data.frame(heat="Coal", area="LittleItaly", cost=2.50, heat_re=0) # Assuming heat_re=0 for Coal
>
> # Predictions from both models
> prediction_e <- predict(model_e, new_data, interval="prediction")
> prediction_f <- predict(model_f, new_data, interval="prediction")
>
> # Print the predictions and intervals
> print(prediction_e)
       fit       lwr      upr
1 2.606232 0.9747882 4.237676
> print(prediction_f)
       fit     lwr      upr
1 2.776521 1.14876 4.404281
```

The `heat` variable with dummy coding allows for each category to have its unique impact on the ratings, which is more aligned with real-world scenarios where each heat source might affect the outcome differently, which cannot be done through `heat_re`.

The coefficients in the dummy variables directly convey the difference in rating when using different heat sources, for instance, the difference in rating when using wood to heat the pizza is simply the coefficient of the dummy variable `heatWood`, however, the `heat_re` model's coefficient gives a single slope for moving through the heat source ranking, which is less intuitive to explain.

Also, using one variable assumes a uniform effect across ranking, which is not the case in real-world scenarios.

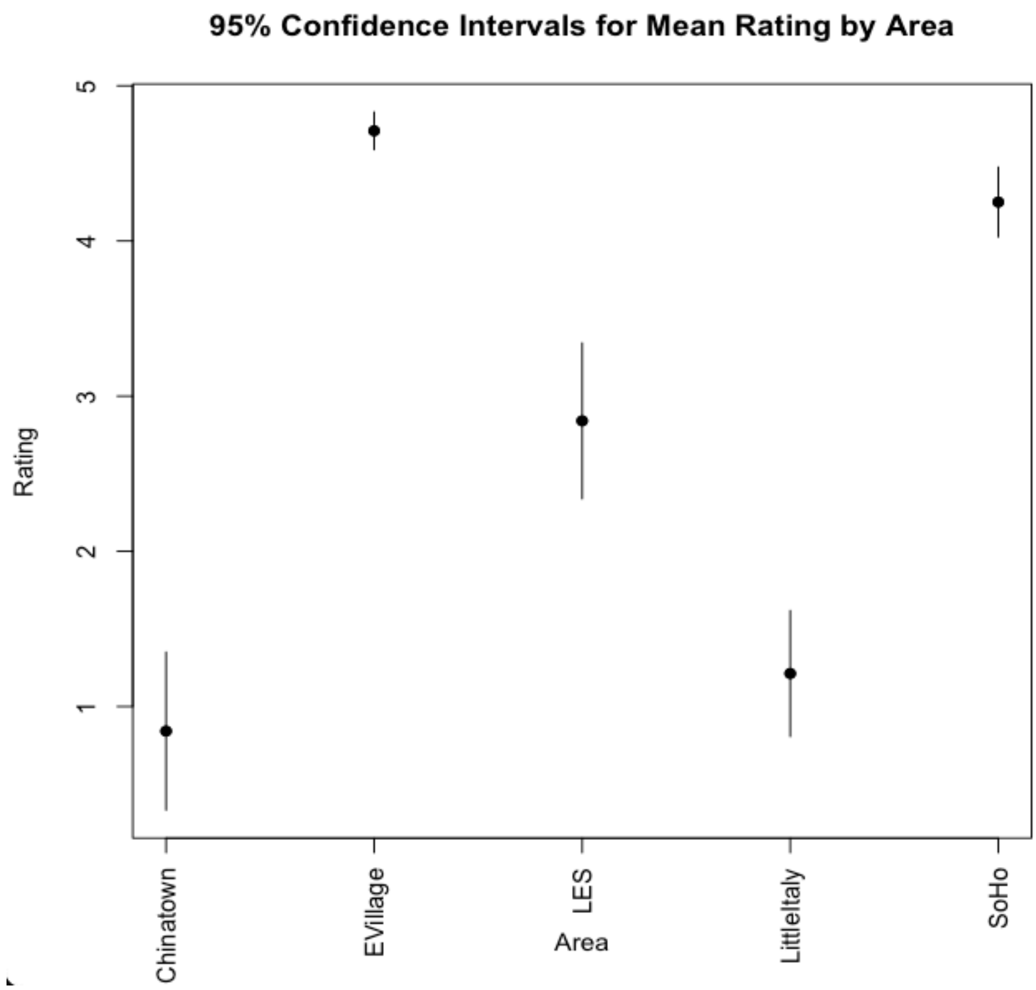
The predicted ratings for a coal-baked pizza that costs $2.50 per slice in LittleItaly are quite similar between the two models with the model in (e) with categorical dummy variables being closer to the ground truth $2.5, but the intervals suggest there's considerable uncertainty about these predictions.

- `heat`: $[0.9747882, 4.237676]$
- `heat_re`: $[1.14876, 4.404281]$

Part3

```
> install.packages("dplyr")
> library(dplyr)
> conf_intervals <- pizza %>%
+   group_by(area) %>%
+   summarise(
+     mean_rating = mean(rating),
+     n = n(),
+     se = sd(rating) / sqrt(n)
+   ) %>%
```

```
+   mutate(
+     lower = mean_rating - qt(0.975, df = n - 1) * se,
+     upper = mean_rating + qt(0.975, df = n - 1) * se
+   )
>
> plot(1:nrow(conf_intervals), conf_intervals$mean_rating, ylim = range(c(conf_intervals$lower,
conf_intervals$upper)),
+      pch = 16, xaxt = 'n', xlab = "Area", ylab = "Rating", main = "95% Confidence Intervals
for Mean Rating by Area")
> axis(1, at = 1:nrow(conf_intervals), labels = conf_intervals$area, las = 2)
> segments(1:nrow(conf_intervals), conf_intervals$lower, 1:nrow(conf_intervals),
conf_intervals$upper)
> points(1:nrow(conf_intervals), conf_intervals$mean_rating, pch = 16)
```



	Chinatown	EVillage	LES	LittleItaly	SoHo
Mean	0.8414286	4.709792	2.840857	1.212558	4.250667
Standard deviation	0.8822598	0.4190338	1.462844	1.318307	0.8744526

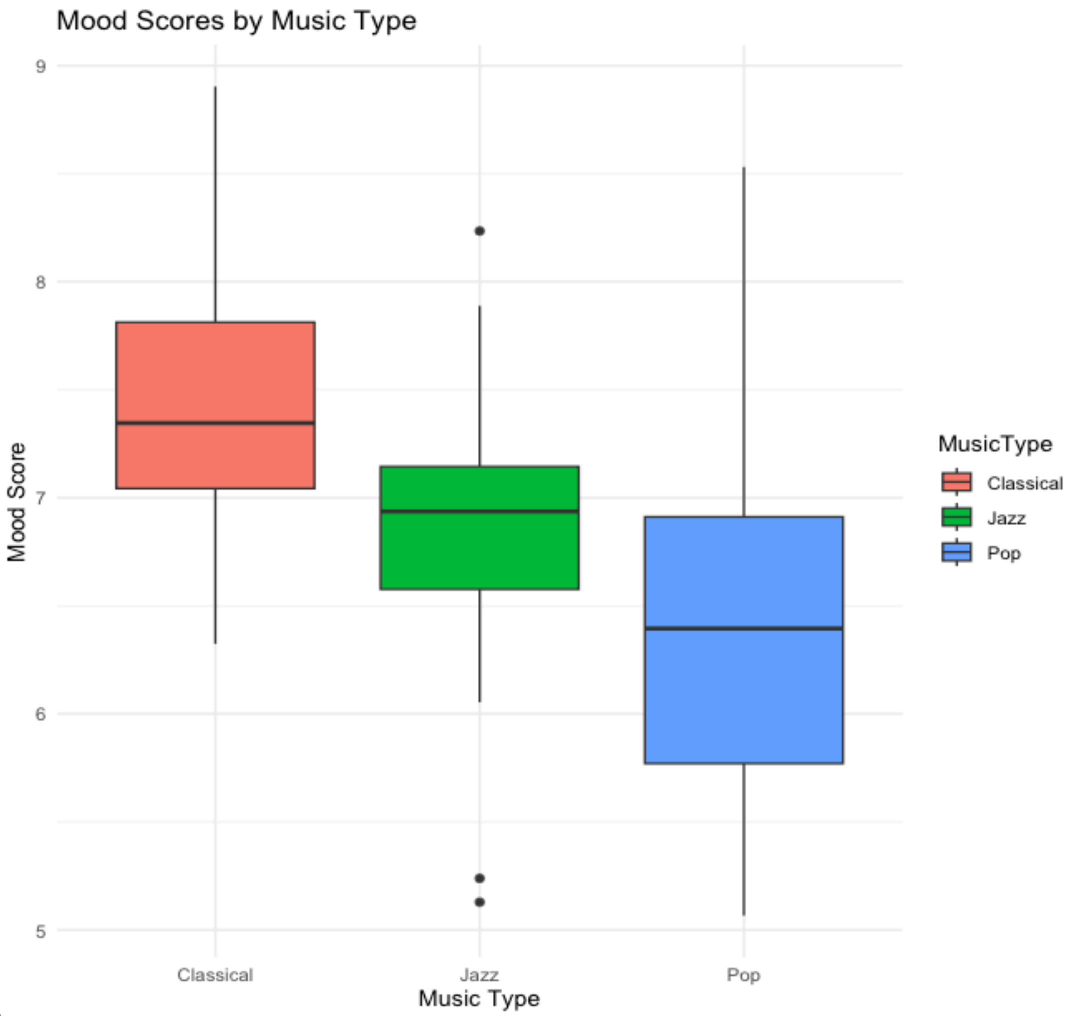
- Chinatown: The confidence interval of China town is wide due to the large standard deviation, indicating the high level of uncertainty and variability in ratings. There may be two reasons, either pizzerias in China town are all horrible with low ratings, or there is considerable inconsistency among individual ratings.

- EVillage: The confidence interval in EVillage is small and with a high rating, indicating the pizzerias in EVillage are rated consistently high.
- LES: The confidence interval for LES is also wide, with a higher mean than China town. Implying varied rating results.
- Little Italy: The confidence interval in little is also wide, could result from a few pizzerias pulling the average down or a genuine spread in quality.
- SoHo: The confidence interval of SoHo is narrow with a high mean, suggesting that pizzerias in SoHo are generally well-rated with a high degree of certainty.

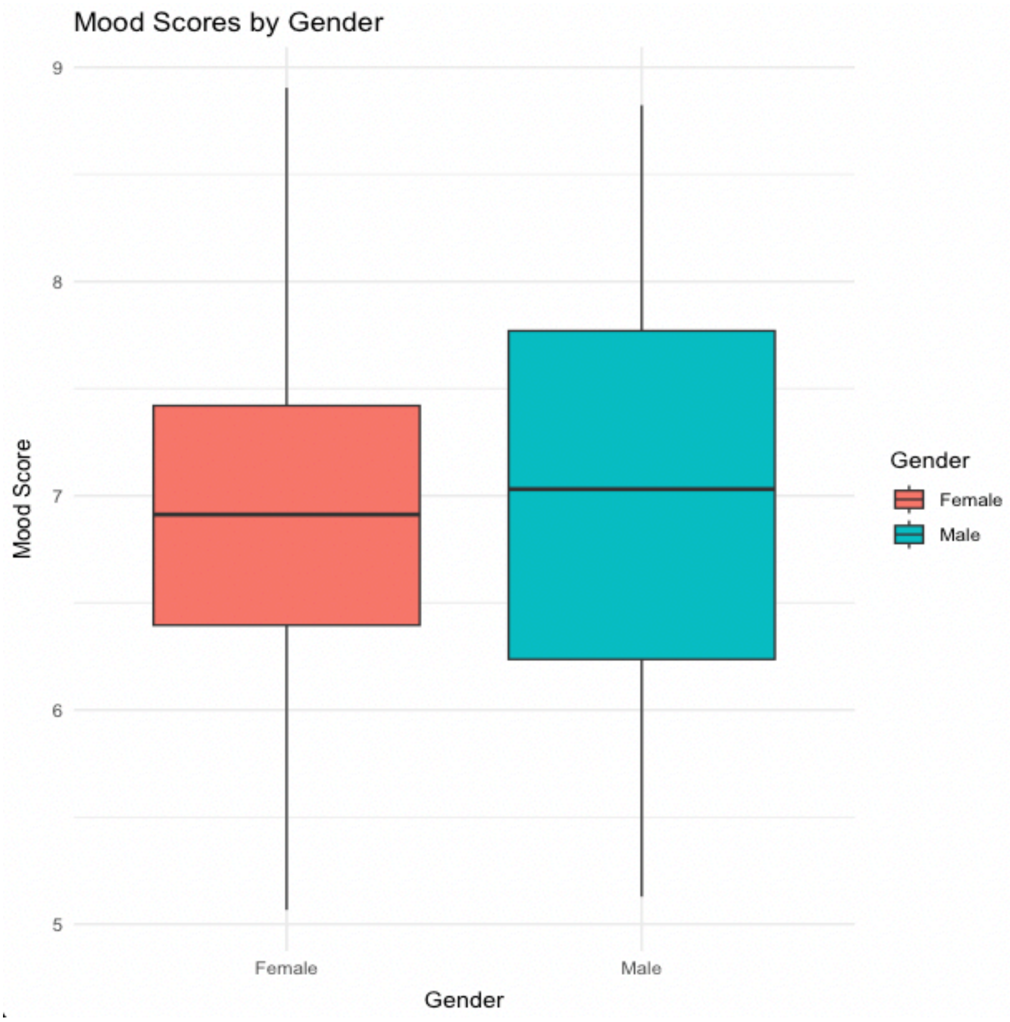
Task 2

(a)

```
> install.packages("ggplot2")
> library(ggplot2)
> ggplot(mood, aes(x=MusicType, y=MoodScore, fill=MusicType)) +
+   geom_boxplot() +
+   labs(title="Mood Scores by Music Type", x="Music Type", y="Mood Score") +
+   theme_minimal()
> ggplot(mood, aes(x=Gender, y=MoodScore, fill=Gender)) +
+   geom_boxplot() +
+   labs(title="Mood Scores by Gender", x="Gender", y="Mood Score") +
+   theme_minimal()
```



The median mood score appears to be highest for Classical music, followed by Jazz, with Pop music having the lowest median mood score. The interquartile range for Pop music is the largest among the three music types, suggesting that those listening to pop music has more varied mood scores.



Both genders show similar variability, as indicated by the similar length of the whiskers on both boxplots. The interquartile range for female is shorter than that of male, indicating that the mood scores of women are more condensed.

(b)

```
> reduced_model <- lm(MoodScore ~ 1, data=mood)
> summary(reduced_model)

Call:
lm(formula = MoodScore ~ 1, data = mood)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.88933 -0.57837  0.01026  0.68716  1.94760 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6.9560     0.1176   59.17   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9106 on 59 degrees of freedom

> full_model <- lm(MoodScore ~ MusicType, data=mood)
```

```
> summary(full_model)

Call:
lm(formula = MoodScore ~ MusicType, data = mood)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6666 -0.4780 -0.1157  0.3584  2.0050 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)     7.4903     0.1748  42.848  < 2e-16 ***
MusicTypeJazz  -0.6948     0.2648  -2.624 0.011131 *  
MusicTypePop   -0.9642     0.2501  -3.854 0.000297 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8199 on 57 degrees of freedom
Multiple R-squared:  0.2167,	Adjusted R-squared:  0.1892 
F-statistic: 7.883 on 2 and 57 DF,  p-value: 0.0009496


> anova_result <- anova(reduced_model, full_model)
> anova_result
Analysis of Variance Table

Model 1: MoodScore ~ 1
Model 2: MoodScore ~ MusicType
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1     59 48.920                                  
2     57 38.321  2    10.599 7.8828 0.0009496 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The significant effect of MusicType has on MoodScore can be tested through having a reduced model with only intercepts and a full model including MusicType as a predictor. The partial F-test will test whether the additional variables (MusicType) in the full model significantly improve the fit of the model compared to the reduced model.

The reduced model is merely the mean of MoodScore, expressing the idea that the mean MoodScore is identical across all music types.

The full model included the MusicType as the predictor.

$$\hat{\text{MoodScore}} = 7.4903 - 0.6948 \times \text{MusicTypeJazz} - 0.9642 \times \text{MusicTypePop} \tag{2}$$

Where we use the MusicType being Classical as the baseline. The R^2 and adjusted R^2 lies within an acceptable range.

The F-statistic of 7.8828 and its associated low p-value ($0.00095 < 0.001$) indicate that the full model with the music type predictor provides a statistically significant better fit for the MoodScore than the reduced model that does not take MusicType into account. This means that the type of music listened to is a significant predictor of the mood score.

(c)

```
> reduced_model_gender <- lm(MoodScore ~ Gender, data=mood)
> summary(reduced_model_gender)

Call:
lm(formula = MoodScore ~ Gender, data = mood)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.85738 -0.58530  0.01986  0.65727  1.97955 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.92401    0.17044  40.624   <2e-16 ***
GenderMale   0.06184    0.23712   0.261    0.795    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9179 on 58 degrees of freedom
Multiple R-squared:  0.001171,	Adjusted R-squared:  -0.01605 
F-statistic: 0.06801 on 1 and 58 DF,  p-value: 0.7952

> full_model_music_gender <- lm(MoodScore ~ MusicType + Gender, data=mood)
> summary(full_model_music_gender)

Call:
lm(formula = MoodScore ~ MusicType + Gender, data = mood)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6510 -0.4725 -0.1094  0.3666  2.0205 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    7.50775    0.21735  34.542  < 2e-16 ***
MusicTypeJazz -0.69834    0.26833  -2.603  0.01182 *  
MusicTypePop  -0.96757    0.25353  -3.816  0.00034 ***
GenderMale    -0.02956    0.21505  -0.137  0.89115    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8271 on 56 degrees of freedom
Multiple R-squared:  0.2169,	Adjusted R-squared:  0.175 
F-statistic: 5.171 on 3 and 56 DF,  p-value: 0.003172

> anova_result_music_gender <- anova(reduced_model_gender, full_model_music_gender)
> anova_result_music_gender
Analysis of Variance Table

Model 1: MoodScore ~ Gender
Model 2: MoodScore ~ MusicType + Gender
```

```
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     58 48.862
2     56 38.308  2    10.555 7.7147 0.001098 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Model 1: partial F is 7.4147 < F_{2,6}=9.32553

  We cannot reject H0, the full model is better

$$\widehat{\text{MoodScore}} = 6.92401 + 0.06184 \times \text{GenderMale} \tag{3}$$

  Given the p-value for GenderMale (0.795), Gender is not a statistically significant predictor of MoodScore in this model. The small coefficient for GenderMale also indicates that the difference in mood scores between genders is minimal according to this model's fit.

- Model 2:

$$\widehat{\text{MoodScore}} = 7.50775 - 0.69834 \times \text{MusicTypeJazz} - 0.96757 \times \text{MusicTypePop} - 0.02956 \times \text{GenderMale} \tag{4}$$

  The fitted regression line indicates that the music type has a larger influence on MoodScore than Gender.

When only including Gender as the predictor for MoodScore, the amount of variation unexplained by the model is 48.862, including the MusicType lowers the unexplained variation to 38.308.

By the partial F-test, MoodScore is stronly related to MusicType where Gender has a minor influence on the MoodScore.

(d)

```
> interaction_model <- lm(MoodScore ~ Gender * MusicType, data=mood)
> summary(interaction_model)

Call:
lm(formula = MoodScore ~ Gender * MusicType, data = mood)

Residuals:
     Min       1Q   Median       3Q      Max
-1.67484 -0.49535 -0.07491  0.39122  1.97658

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                7.5778     0.2803  27.033  < 2e-16 ***
GenderMale                -0.1481     0.3647  -0.406  0.68629
MusicTypeJazz             -0.7896     0.3964  -1.992  0.05146 .
MusicTypePop              -1.0775     0.3780  -2.851  0.00616 **
GenderMale:MusicTypeJazz   0.1636     0.5477   0.299  0.76630
GenderMale:MusicTypePop    0.2024     0.5177   0.391  0.69736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8409 on 54 degrees of freedom
Multiple R-squared:  0.2194,  Adjusted R-squared:  0.1471
F-statistic: 3.035 on 5 and 54 DF,  p-value: 0.01739
```

$$
\begin{aligned}
\widehat{\text{MoodScore}} = {} & 7.5778 - 0.1481 \times \text{GenderMale} - 0.7896 \times \text{MusicTypeJazz} \\
& - 1.0775 \times \text{MusicTypePop} + 0.1636 \times \text{GenderMale} \times \text{MusicTypeJazz} \\
& + 0.2024 \times \text{GenderMale} \times \text{MusicTypePop}
\end{aligned} \tag{5}
$$

The Multiple R^2 value is 0.2194, indicating that approximately 21.94% of the variability in mood scores is explained by the model. The Adjusted R^2 is lower at 0.1471, accounting for the number of predictors in the model. The F-statistic is 3.035 on 5 and 54 degrees of freedom, with a p-value of 0.01739, suggesting that the model is significant.

MusicType has a larger impact on the model while Gender showcased minor effect on the MoodScore, including the interaction terms between Gender and MusicType, indicating the choice of MusicType does not differ by Gender.