

RTutor report

RTutor, Powered by ChatGPT

Mon Aug 11 21:19:27 2025

Developed by Steven Ge using API access via the OpenAI package and custom chat completion package to OpenAI's "GPT-4o" model.

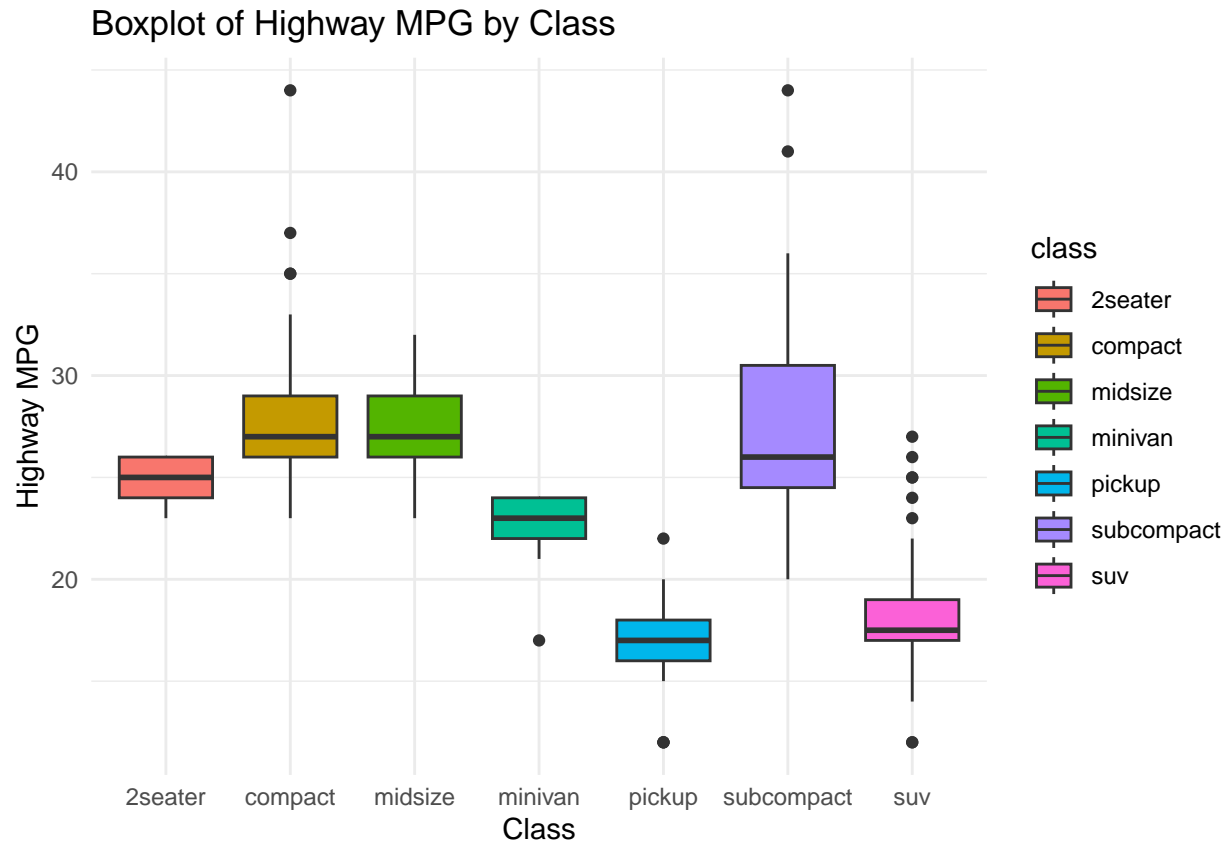
RTutor Website: <https://RTutor.ai> Source code: GitHub.

1. Create a boxplot of highway vs. class. Color by class.

Azure GPT-4o (Temperature = 0.2)

```
library(ggplot2)
data(mpg)
df <- mpg

# Create boxplot of highway vs. class, colored by class
ggplot(df, aes(x = class, y = hwy, fill = class)) +
  geom_boxplot() +
  labs(title = "Boxplot of Highway MPG by Class", x = "Class", y = "Highway MPG") +
  theme_minimal()
```

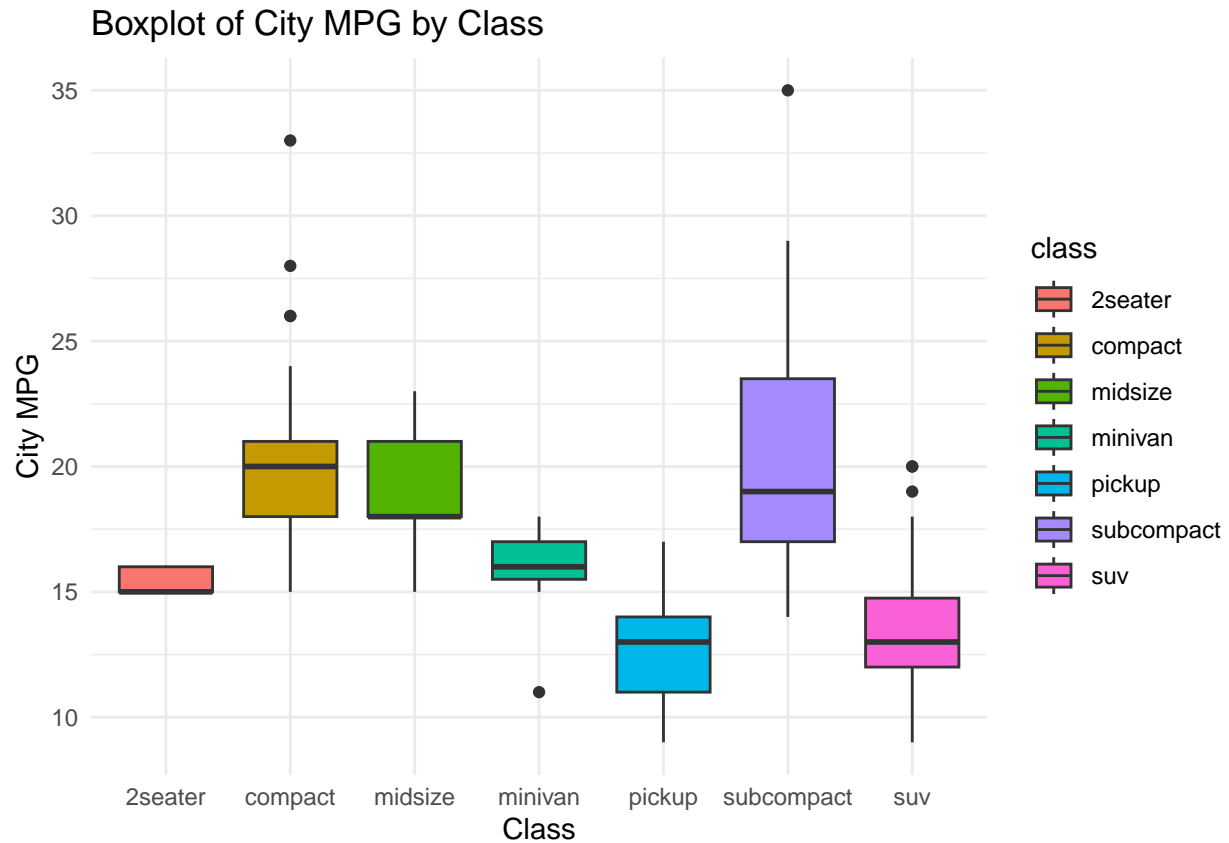


2. Create a boxplot of city vs. class.

Azure GPT-4o (Temperature = 0.2)

```
library(ggplot2)

# Create boxplot of city MPG vs. class
ggplot(df, aes(x = class, y = cty, fill = class)) +
  geom_boxplot() +
  labs(title = "Boxplot of City MPG by Class", x = "Class", y = "City MPG") +
  theme_minimal()
```

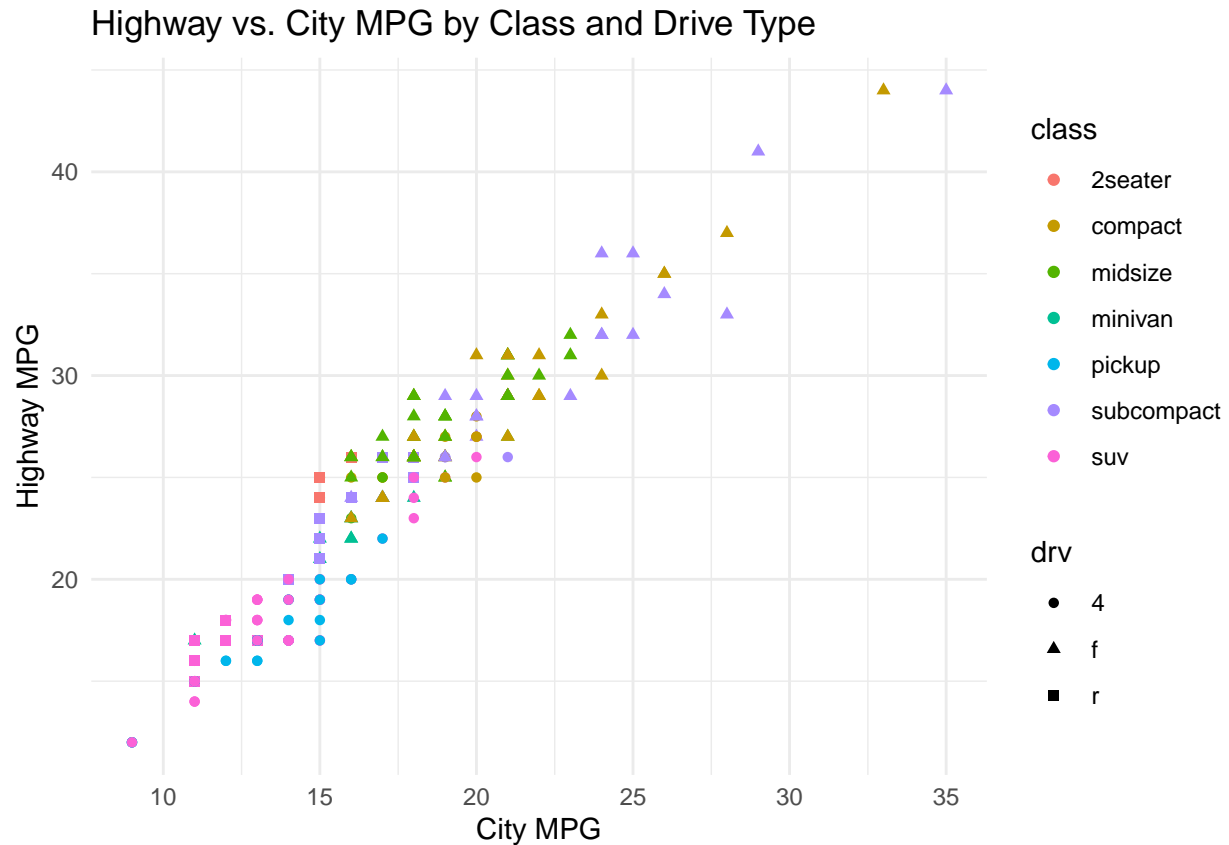


3. Plot hwy vs. cty, colored by class. Change shape by drv.

Azure GPT-4o (Temperature = 0.2)

```
library(ggplot2)

# Plot highway vs. city MPG, colored by class and shaped by drive type
ggplot(df, aes(x = cty, y = hwy, color = class, shape = drv)) +
  geom_point() +
  labs(title = "Highway vs. City MPG by Class and Drive Type", x = "City MPG", y = "Highway MPG") +
  theme_minimal()
```



4. Conduct ANOVA of hwy by class.

Azure GPT-4o (Temperature = 0.2)

```
# Conduct ANOVA of highway MPG by class
anova_result <- aov(hwy ~ class, data = df)

# Display ANOVA summary
summary(anova_result)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## class         6   5683   947.2    83.39 <2e-16 ***
## Residuals    227   2578    11.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Conduct ANOVA of log-transformed hwy by class and drv.

Azure GPT-4o (Temperature = 0.2)

```
# Conduct ANOVA of log-transformed highway MPG by class and drive type
anova_result <- aov(log(hwy) ~ class + drv, data = df)
```

```
# Display ANOVA summary
summary(anova_result)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## class         6 11.477   1.9128  115.82 < 2e-16 ***
## drv           2  0.606   0.3028   18.34 4.2e-08 ***
## Residuals    225  3.716   0.0165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6. Create a correlation map of all the columns that contain numbers.

Azure GPT-4o (Temperature = 0.2)

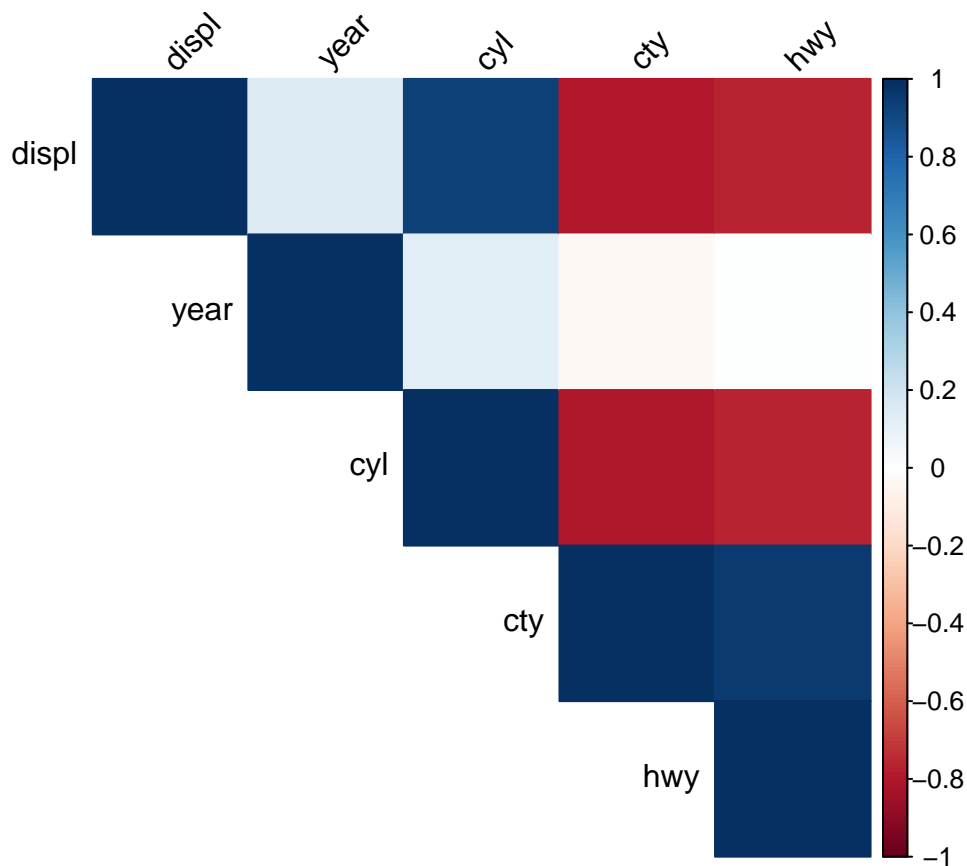
```
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
# Select numeric columns
numeric_df <- df[, sapply(df, is.numeric)]
```

```
# Calculate correlation matrix
correlation_matrix <- cor(numeric_df)
```

```
# Plot correlation map
corrplot(correlation_matrix, method = "color", type = "upper", tl.col = "black", tl.srt = 45)
```



7. Use `corrplot` to create a correlation plot of all the columns that contain numbers.

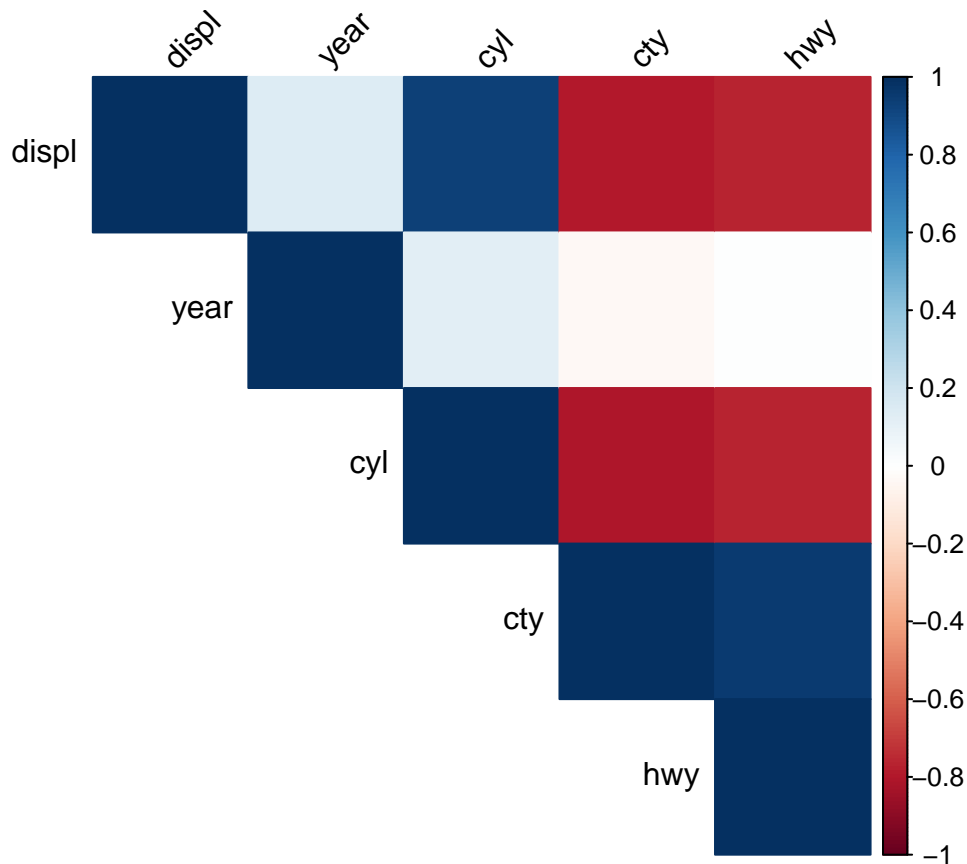
Azure GPT-4o (Temperature = 0.2)

```
library(corrplot)

# Select numeric columns
numeric_df <- df[, sapply(df, is.numeric)]

# Calculate correlation matrix
correlation_matrix <- cor(numeric_df)

# Plot correlation map using corrplot
corrplot(correlation_matrix, method = "color", type = "upper", tl.col = "black", tl.srt = 45)
```



8. Calculate average cty by year and class.

Azure GPT-4o (Temperature = 0.2)

```
# Calculate average city MPG by year and class
average_cty <- aggregate(cty ~ year + class, data = df, FUN = mean)

# Display the result
average_cty
```

```
##   year   class   cty
## 1 1999 2seater 15.50000
## 2 2008 2seater 15.33333
## 3 1999 compact 19.76000
## 4 2008 compact 20.54545
## 5 1999 midsize 18.15000
## 6 2008 midsize 19.33333
## 7 1999 minivan 16.16667
## 8 2008 minivan 15.40000
## 9 1999 pickup 13.00000
## 10 2008 pickup 13.00000
## 11 1999 subcompact 21.57895
## 12 2008 subcompact 18.93750
## 13 1999   suv 13.37931
## 14 2008   suv 13.60606
```

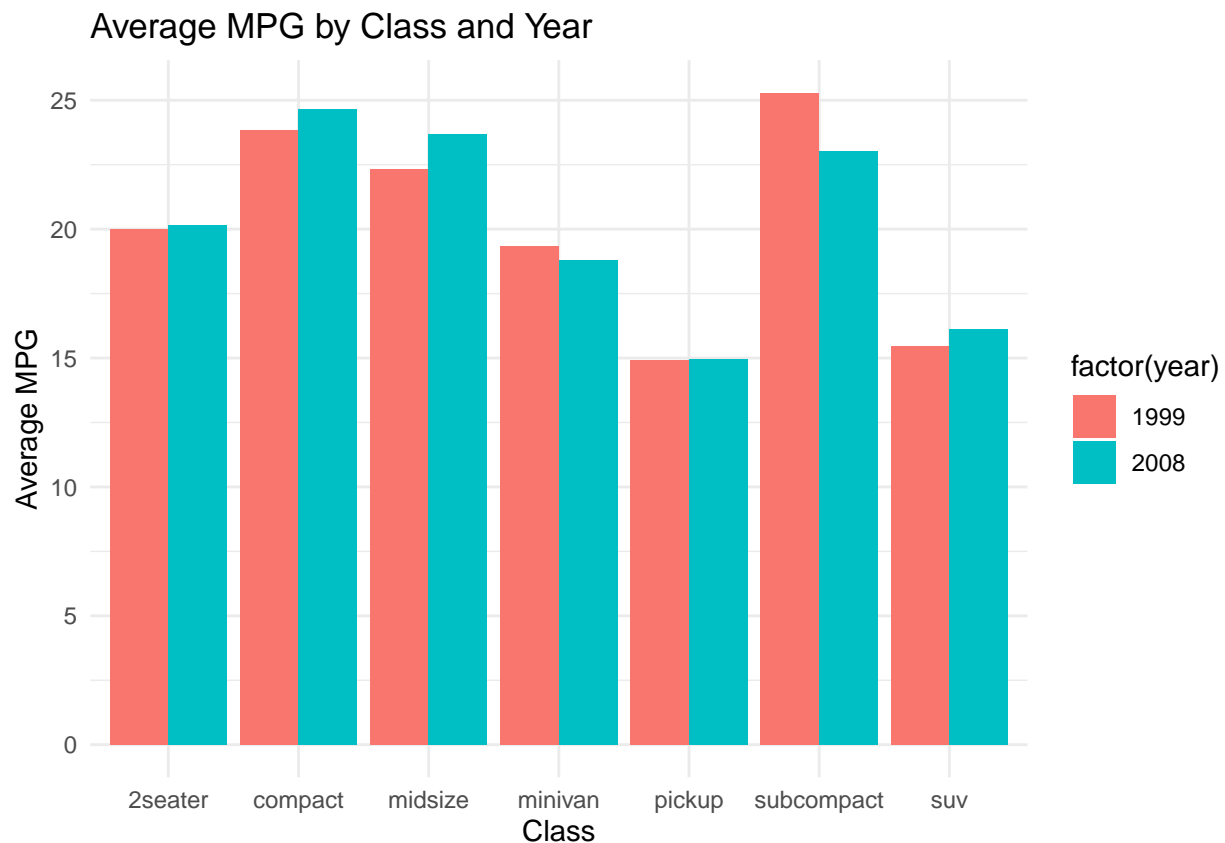
9. Create a barplot of average mpg by class, colored by year. The years should be side by side.

Azure GPT-4o (Temperature = 0.2)

```
library(ggplot2)

# Calculate average MPG by class and year
average_mpg <- aggregate(cbind(cty, hwy) ~ class + year, data = df, FUN = mean)

# Create a barplot of average MPG by class, colored by year
ggplot(average_mpg, aes(x = class, y = (cty + hwy) / 2, fill = factor(year))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Average MPG by Class and Year", x = "Class", y = "Average MPG") +
  theme_minimal()
```



10. Calculate the correlation of cty vs hwy. Repeat that after log transformation. Collect these results and show them.

Azure GPT-4o (Temperature = 0.2)

```
# Calculate correlation of city vs highway MPG
cor_cty_hwy <- cor(df$cty, df$hwy)

# Calculate correlation of log-transformed city vs highway MPG
```



```
cor_log_cty_hwy <- cor(log(df$cty), log(df$hwy))

# Collect and show results
results <- data.frame(
  Metric = c("Correlation cty vs hwy", "Correlation log(cty) vs log(hwy)"),
  Value = c(cor_cty_hwy, cor_log_cty_hwy)
)

results
```

```
##               Metric      Value
## 1      Correlation cty vs hwy 0.9559159
## 2 Correlation log(cty) vs log(hwy) 0.9556827
```

11. Hwy and cty represent miles per gallon (MPG) on the highway and in the city, respectively. Only keep cars more efficient than 15 MPG, but less than 40, on the highway. Add 0.5 to city MPG for correction. Perform log transformation on city MPG. Raise highway MPG to the second power. Calculate correlation coefficient of the two transformed variables.

Azure GPT-4o (Temperature = 0.2)

```
# Filter cars with highway MPG between 15 and 40
filtered_df <- df[df$hwy > 15 & df$hwy < 40, ]

# Add 0.5 to city MPG and perform log transformation
filtered_df$log_cty <- log(filtered_df$cty + 0.5)

# Raise highway MPG to the second power
filtered_df$squared_hwy <- filtered_df$hwy^2

# Calculate correlation coefficient of the two transformed variables
correlation_coefficient <- cor(filtered_df$log_cty, filtered_df$squared_hwy)

# Show the result
correlation_coefficient
```

```
## [1] 0.9334202
```