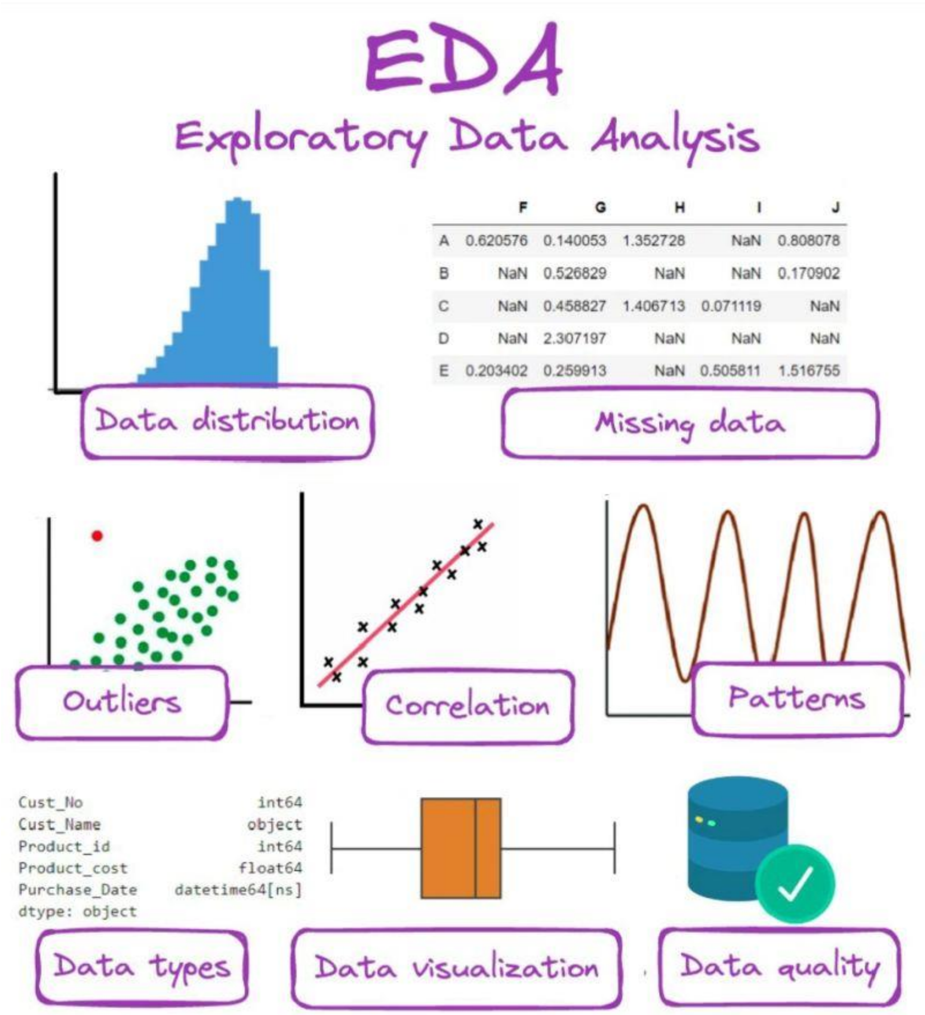
EXPERIMENT – 6


EDA
Exploratory Data Analysis
Data distribution
Missing data
Outliers
Correlation
Patterns
Data types
Data visualization
Data quality
F G H I J
A 0.620576 0.140053 1.352728 NaN 0.808078
B NaN 0.526829 NaN NaN 0.170902
C NaN 0.458827 1.406713 0.071119 NaN
D NaN 2.307197 NaN NaN NaN
E 0.203402 0.259913 NaN 0.505811 1.516755
Cust_No int64
Cust_Name object
Product_id int64
Product_cost float64
Purchase_Date datetime64[ns]
dtype: object

NAME : JVN GANESH

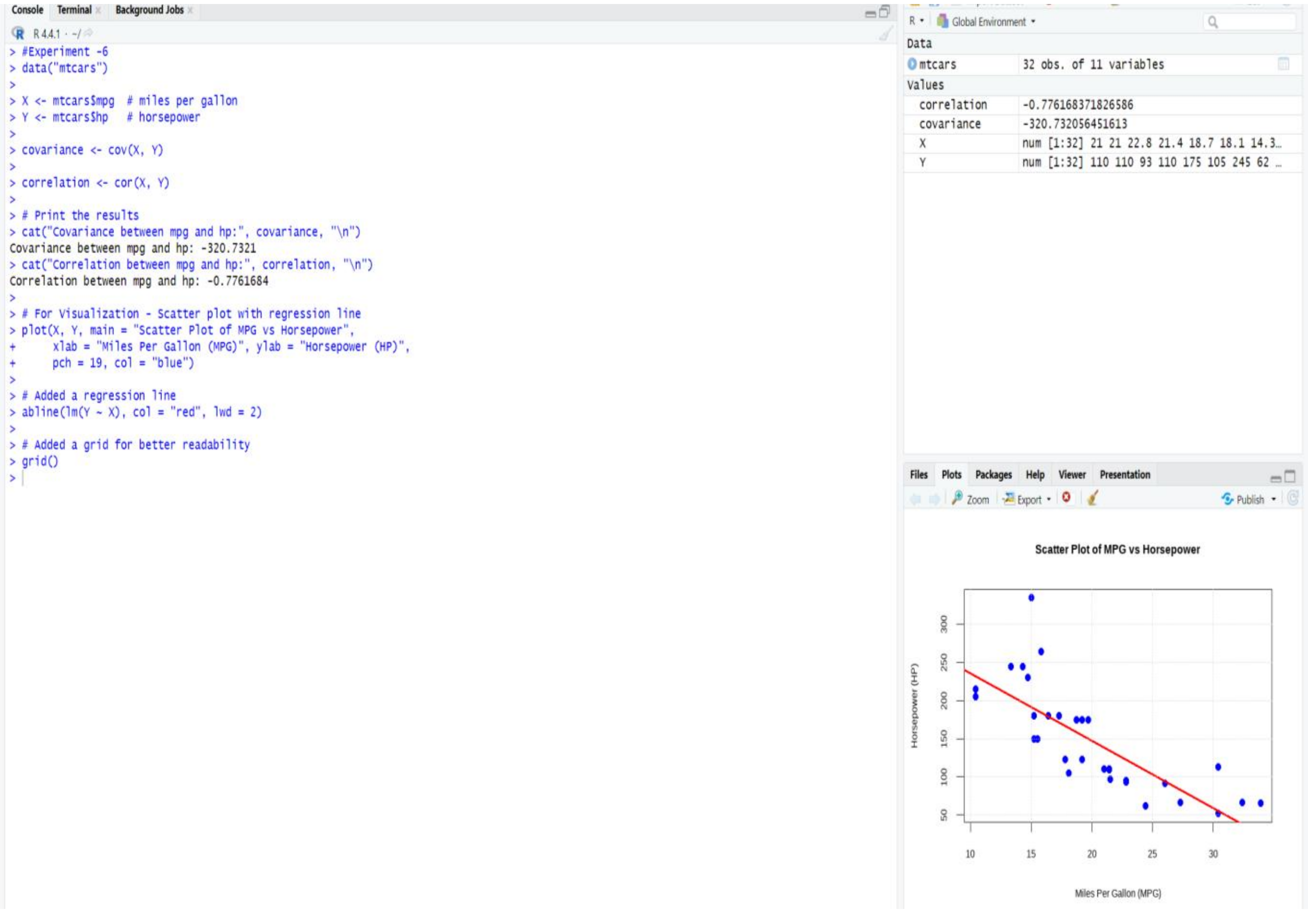
Roll No : 21BDS0085

Code :

```
library(corrplot)  # for the correlation heatmap

library(ggplot2)   # for the scatter plot

data("mtcars")


# Extract the X (mpg) and Y (hp) values

X <- mtcars$mpg  # miles per gallon

Y <- mtcars$hp   # horsepower


# Calculate covariance

covariance <- cov(X, Y)


# Calculate correlation

correlation <- cor(X, Y)


# Print the results

cat("Covariance between mpg and hp:", covariance, "\n")

cat("Correlation between mpg and hp:", correlation, "\n")


# Visualization - Scatter plot with regression line

plot(X, Y, main = "Scatter Plot of MPG vs Horsepower",

    xlab = "Miles Per Gallon (MPG)", ylab = "Horsepower (HP)",

    pch = 19, col = "blue")


# Add a regression line

abline(lm(Y ~ X), col = "red", lwd = 2)
```

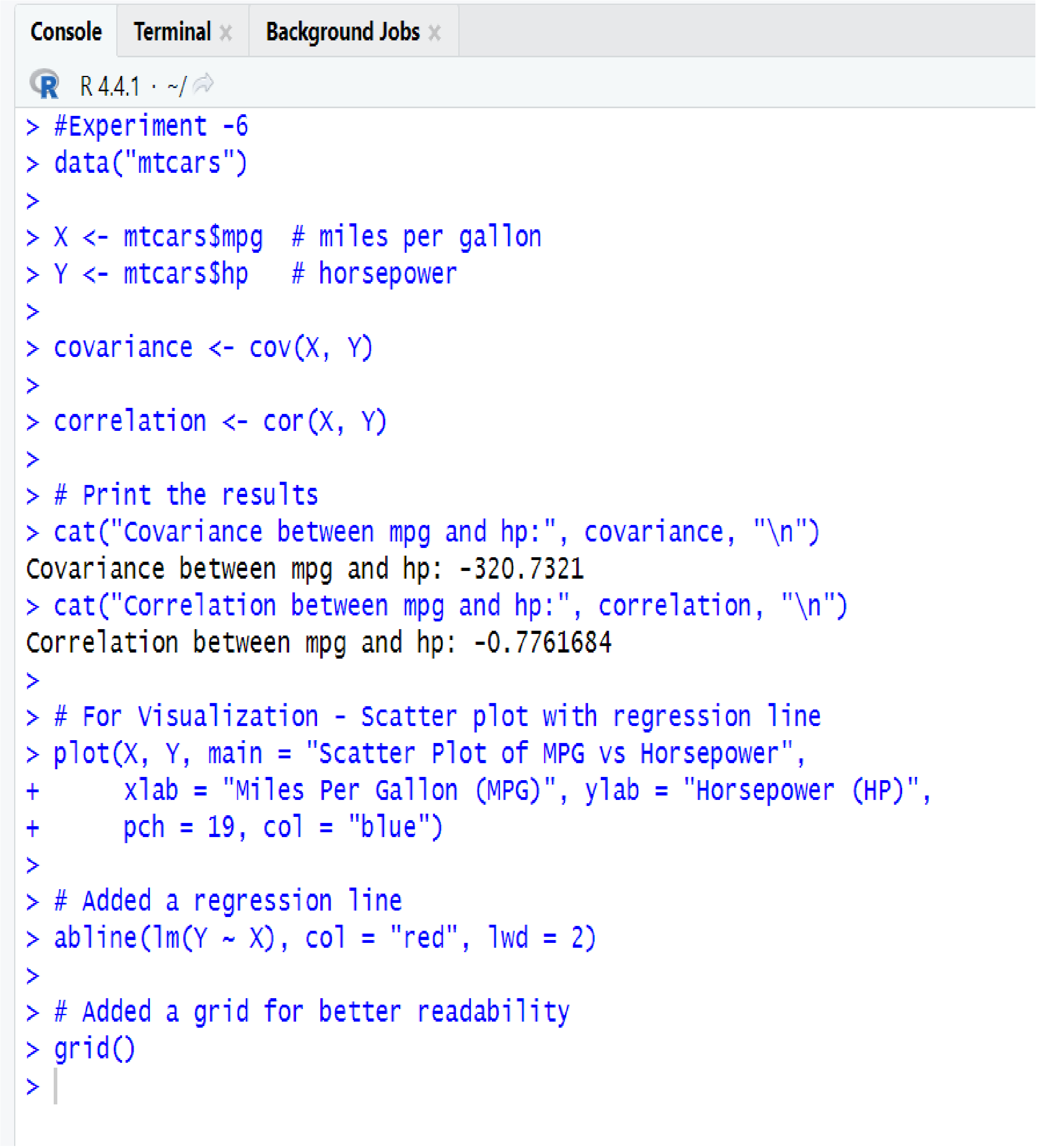
```
# Add a grid for better readability
grid()

# Calculate correlation matrix for all numeric variables in mtcars
cor_matrix <- cor(mtcars)

# Heatmap visualization of the correlation matrix
corrplot(cor_matrix, method = "color", type = "lower",
         tl.col = "black", tl.srt = 45,
         addCoef.col = "black", number.cex = 0.8,
         col = colorRampPalette(c("red", "white", "blue"))(200))

# Alternatively, for a heatmap with ggplot2, you can do:
library(reshape2)

# Melt the correlation matrix for ggplot2
cor_melted <- melt(cor_matrix)

# Plot the heatmap
ggplot(cor_melted, aes(Var1, Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "blue", mid = "white",
                       midpoint = 0, limit = c(-1, 1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                   size = 12, hjust = 1)) +
  coord_fixed()
```

Outputs Screenshots

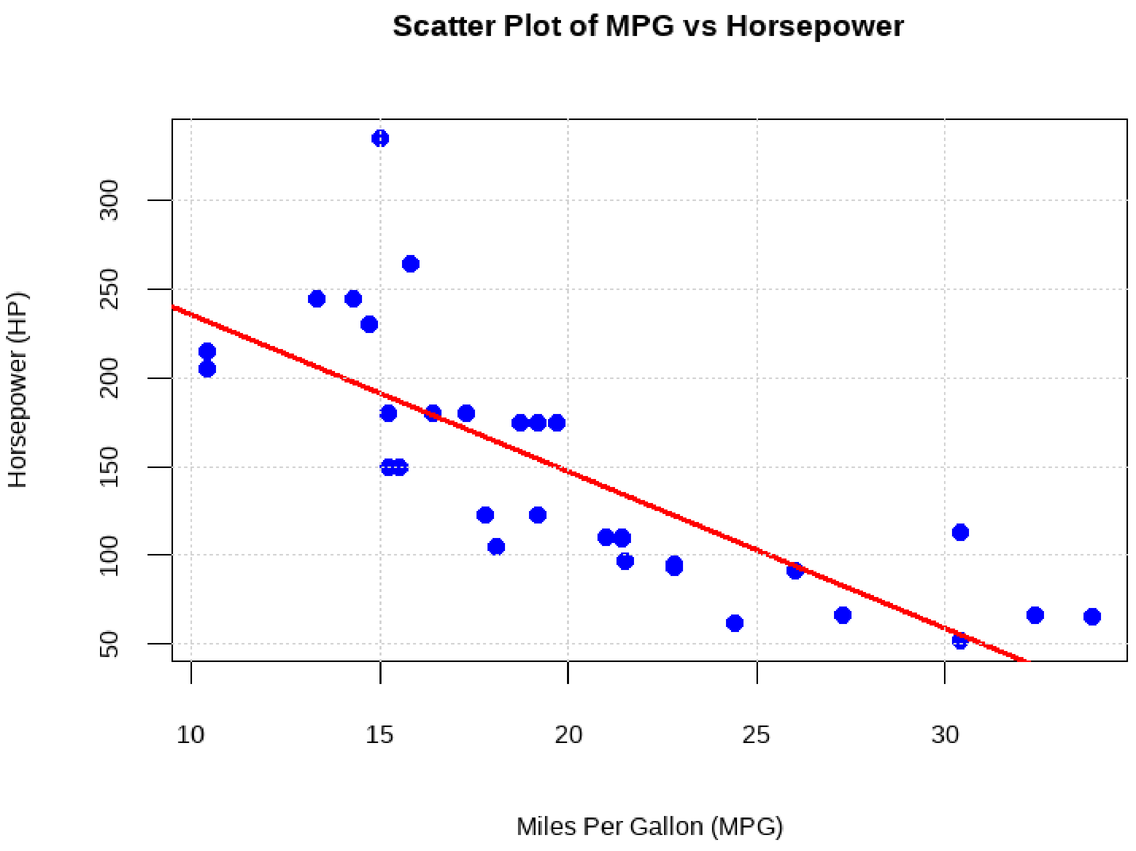


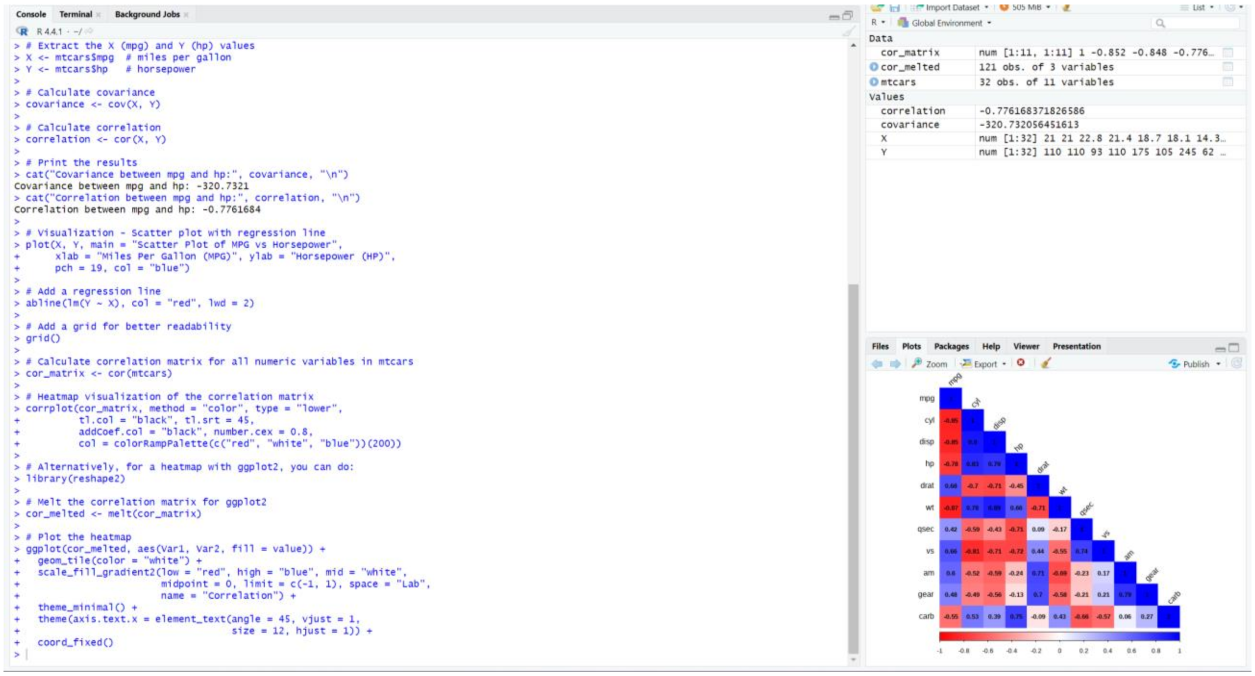
```
> #Experiment -6
> data("mtcars")
>
> X <- mtcars$mpg  # miles per gallon
> Y <- mtcars$hp   # horsepower
>
> covariance <- cov(X, Y)
>
> correlation <- cor(X, Y)
>
> # Print the results
> cat("Covariance between mpg and hp:", covariance, "\n")
Covariance between mpg and hp: -320.7321
> cat("Correlation between mpg and hp:", correlation, "\n")
Correlation between mpg and hp: -0.7761684
>
> # For Visualization - Scatter plot with regression line
> plot(X, Y, main = "Scatter Plot of MPG vs Horsepower",
+      xlab = "Miles Per Gallon (MPG)", ylab = "Horsepower (HP)",
+      pch = 19, col = "blue")
>
> # Added a regression line
> abline(lm(Y ~ X), col = "red", lwd = 2)
>
> # Added a grid for better readability
> grid()
>
```

```
Data
mtcars        32 obs. of 11 variables
Values
  correlation -0.776168371826586
  covariance  -320.732056451613
  X           num [1:32] 21 21 22.8 21.4 18.7 18.1 14.3…
  Y           num [1:32] 110 110 93 110 175 105 245 62 …
```



```
> #Experiment -6
> data("mtcars")
>
> X <- mtcars$mpg  # miles per gallon
> Y <- mtcars$hp   # horsepower
>
> covariance <- cov(X, Y)
>
> correlation <- cor(X, Y)
>
> # Print the results
> cat("Covariance between mpg and hp:", covariance, "\n")
Covariance between mpg and hp: -320.7321
> cat("Correlation between mpg and hp:", correlation, "\n")
Correlation between mpg and hp: -0.7761684
>
> # For Visualization - Scatter plot with regression line
> plot(X, Y, main = "Scatter Plot of MPG vs Horsepower",
+      xlab = "Miles Per Gallon (MPG)", ylab = "Horsepower (HP)",
+      pch = 19, col = "blue")
>
> # Added a regression line
> abline(lm(Y ~ X), col = "red", lwd = 2)
>
> # Added a grid for better readability
> grid()
>
```

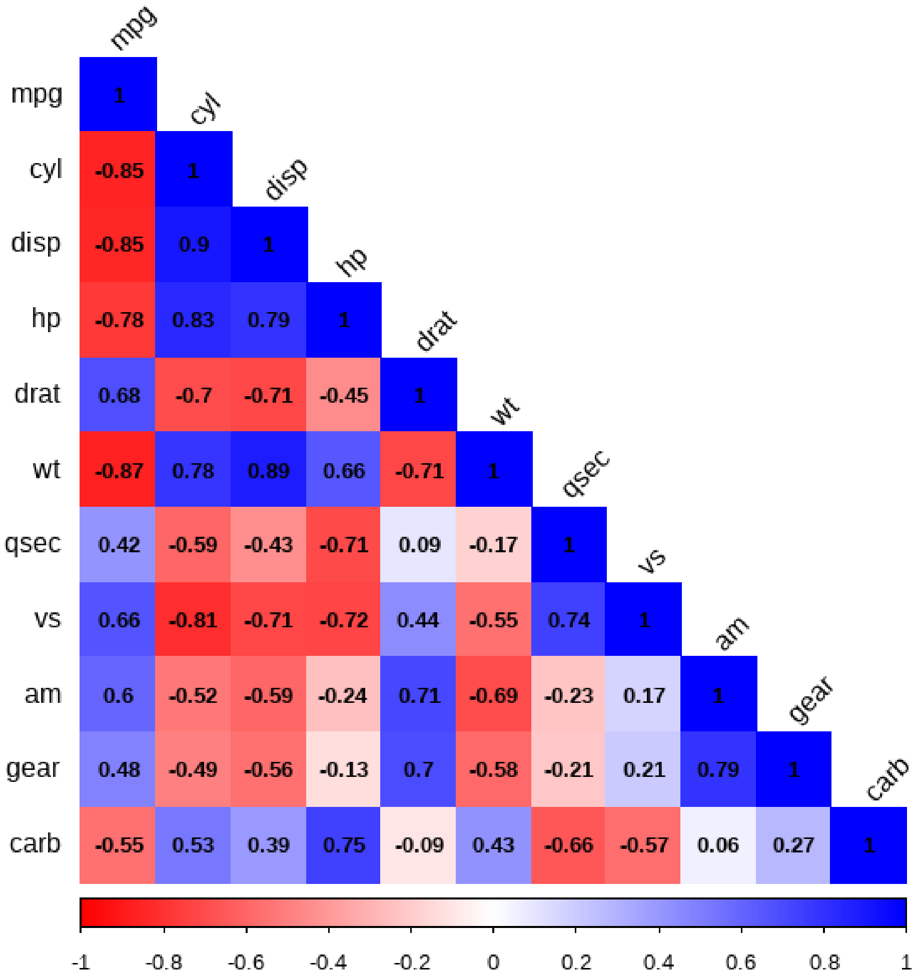


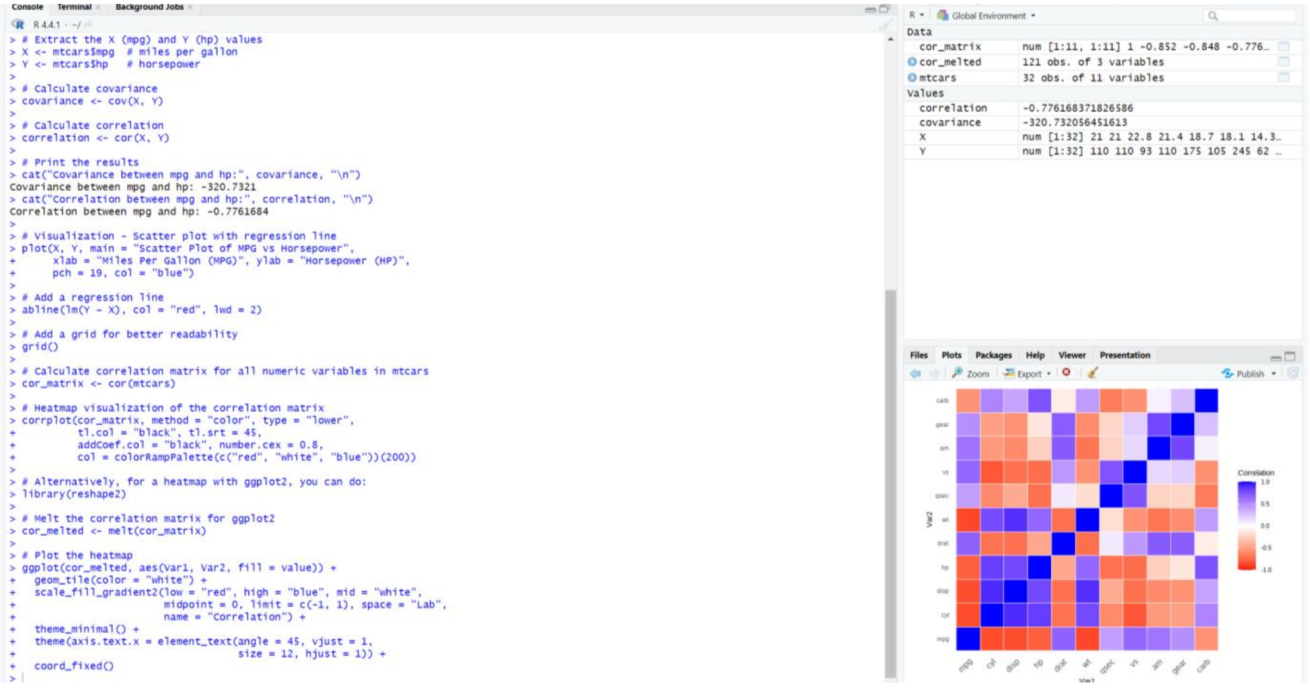


```
> # Extract the X (mpg) and Y (hp) values
> X <- mtcars$mpg  # miles per gallon
> Y <- mtcars$hp   # horsepower
>
> # Calculate covariance
> covariance <- cov(X, Y)
>
> # Calculate correlation
> correlation <- cor(X, Y)
>
> # Print the results
> cat("Covariance between mpg and hp:", covariance, "\n")
Covariance between mpg and hp: -320.7321
> cat("Correlation between mpg and hp:", correlation, "\n")
Correlation between mpg and hp: -0.7761684
>
> # Visualization - Scatter plot with regression line
> plot(X, Y, main = "Scatter Plot of MPG vs Horsepower",
+      xlab = "Miles Per Gallon (MPG)", ylab = "Horsepower (HP)",
+      pch = 19, col = "blue")
>
> # Add a regression line
> abline(lm(Y ~ X), col = "red", lwd = 2)
>
> # Add a grid for better readability
> grid()
>
> # Calculate correlation matrix for all numeric variables in mtcars
> cor_matrix <- cor(mtcars)
>
> # Heatmap visualization of the correlation matrix
> corrplot(cor_matrix, method = "color", type = "lower",
+          tl.col = "black", tl.srt = 45,
+          addCoef.col = "black", number.cex = 0.8,
+          col = colorRampPalette(c("red", "white", "blue"))(200))
>
> # Alternatively, for a heatmap with ggplot2, you can do:
> library(reshape2)
>
> # Melt the correlation matrix for ggplot2
> cor_melted <- melt(cor_matrix)
>
> # Plot the heatmap
> ggplot(cor_melted, aes(Var1, Var2, fill = value)) +
+   geom_tile(color = "white") +
+   scale_fill_gradient2(low = "red", high = "blue", mid = "white",
+                        midpoint = 0, limit = c(-1, 1), space = "Lab",
+                        name = "Correlation") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, vjust = 1,
+                                    size = 12, hjust = 1)) +
+   coord_fixed()
> |
```

Data	
cor_matrix	num [1:11, 1:11] 1 -0.852 -0.848 -0.776…
cor_melted	121 obs. of 3 variables
mtcars	32 obs. of 11 variables
Values	
correlation	-0.776168371826586
covariance	-320.732056451613
X	num [1:32] 21 21 22.8 21.4 18.7 18.1 14.3…
Y	num [1:32] 110 110 93 110 175 105 245 62 …

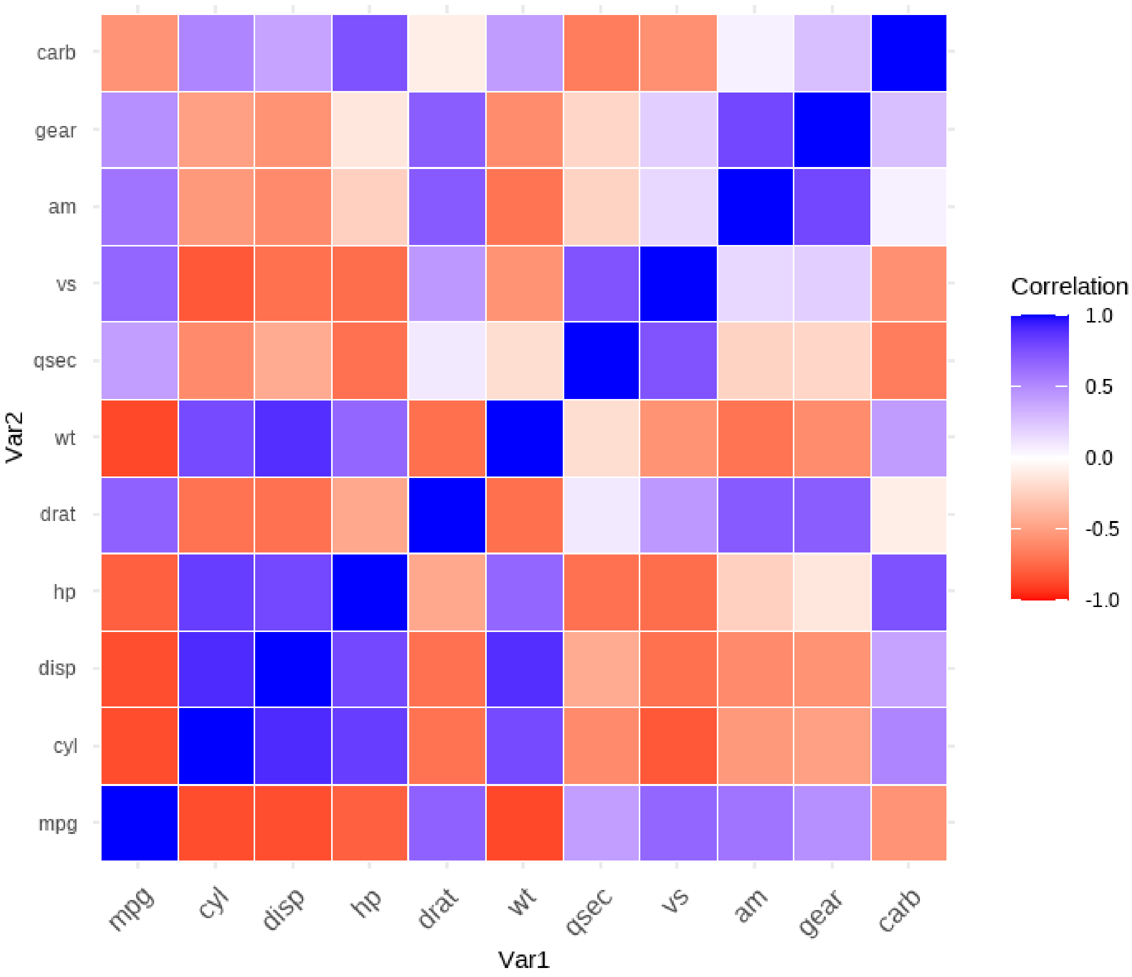
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1										
cyl	-0.85	1									
disp	-0.85	0.9	1								
hp	-0.78	0.83	0.79	1							
drat	0.68	-0.7	-0.71	-0.45	1						
wt	-0.87	0.78	0.89	0.66	-0.71	1					
qsec	0.42	-0.59	-0.43	-0.71	0.09	-0.17	1				
vs	0.66	-0.81	-0.71	-0.72	0.44	-0.55	0.74	1			
am	0.6	-0.52	-0.59	-0.24	0.71	-0.69	-0.23	0.17	1		
gear	0.48	-0.49	-0.56	-0.13	0.7	-0.58	-0.21	0.21	0.79	1	
carb	-0.55	0.53	0.39	0.75	-0.09	0.43	-0.66	-0.57	0.06	0.27	1





```
> # Extract the X (mpg) and Y (hp) values
> X <- mtcars$mpg  # miles per gallon
> Y <- mtcars$hp   # horsepower
>
> # Calculate covariance
> covariance <- cov(X, Y)
>
> # Calculate correlation
> correlation <- cor(X, Y)
>
> # Print the results
> cat("Covariance between mpg and hp:", covariance, "\n")
Covariance between mpg and hp: -320.7321
> cat("Correlation between mpg and hp:", correlation, "\n")
Correlation between mpg and hp: -0.7761684
>
> # Visualization - Scatter plot with regression line
> plot(X, Y, main = "Scatter Plot of MPG vs Horsepower",
+      xlab = "Miles Per Gallon (MPG)", ylab = "Horsepower (HP)",
+      pch = 19, col = "blue")
>
> # Add a regression line
> abline(lm(Y ~ X), col = "red", lwd = 2)
>
> # Add a grid for better readability
> grid()
>
> # Calculate correlation matrix for all numeric variables in mtcars
> cor_matrix <- cor(mtcars)
>
> # Heatmap visualization of the correlation matrix
> corrplot(cor_matrix, method = "color", type = "lower",
+          tl.col = "black", tl.srt = 45,
+          addCoef.col = "black", number.cex = 0.8,
+          col = colorRampPalette(c("red", "white", "blue"))(200))
>
> # Alternatively, for a heatmap with ggplot2, you can do:
> library(reshape2)
>
> # Melt the correlation matrix for ggplot2
> cor_melted <- melt(cor_matrix)
>
> # Plot the heatmap
> ggplot(cor_melted, aes(Var1, Var2, fill = value)) +
+   geom_tile(color = "white") +
+   scale_fill_gradient2(low = "red", high = "blue", mid = "white",
+                        midpoint = 0, limit = c(-1, 1), space = "Lab",
+                        name = "Correlation") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, vjust = 1,
+                                    size = 12, hjust = 1)) +
+   coord_fixed()
> 
```

```
Data
  cor_matrix   num [1:11, 1:11] 1 -0.852 -0.848 -0.776…
  cor_melted   121 obs. of 3 variables
  mtcars       32 obs. of 11 variables
Values
  correlation  -0.776168371826586
  covariance   -320.732056451613
  X            num [1:32] 21 21 22.8 21.4 18.7 18.1 14.3…
  Y            num [1:32] 110 110 93 110 175 105 245 62 …
```


carb
gear
am
vs
qsec
wt
drat
hp
disp
cyl
mpg
Var2
mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
Var1
Correlation
1.0
0.5
0.0
-0.5
-1.0