

MTCars Dataset Analysis

Alexa Benson

2024-06-27

Data Summary

From the data summary, we can see that the average miles per gallon (mpg) is about 20.1, with a range from 10.4 to 33.9. Similarly, other variables like horsepower (hp), weight (wt), and the number of cylinders (cyl) show considerable variation.

```
data(mtcars)
summary(mtcars)
```

```
##      mpg          cyl          disp          hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##      drat          wt          qsec          vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##      am          gear          carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Correlation Analysis

Investigate the relationships between different variables. The correlation matrix reveals the strength and direction of linear relationships between variables. There is a strong negative correlation between horsepower (hp) and miles per gallon (mpg) (-0.78), suggesting that cars with higher horsepower tend to have lower fuel efficiency. Additionally, weight (wt) and mpg also show a strong negative correlation (-0.87), reinforcing the idea that heavier cars are less fuel-efficient.

```
cor(mtcars)
```

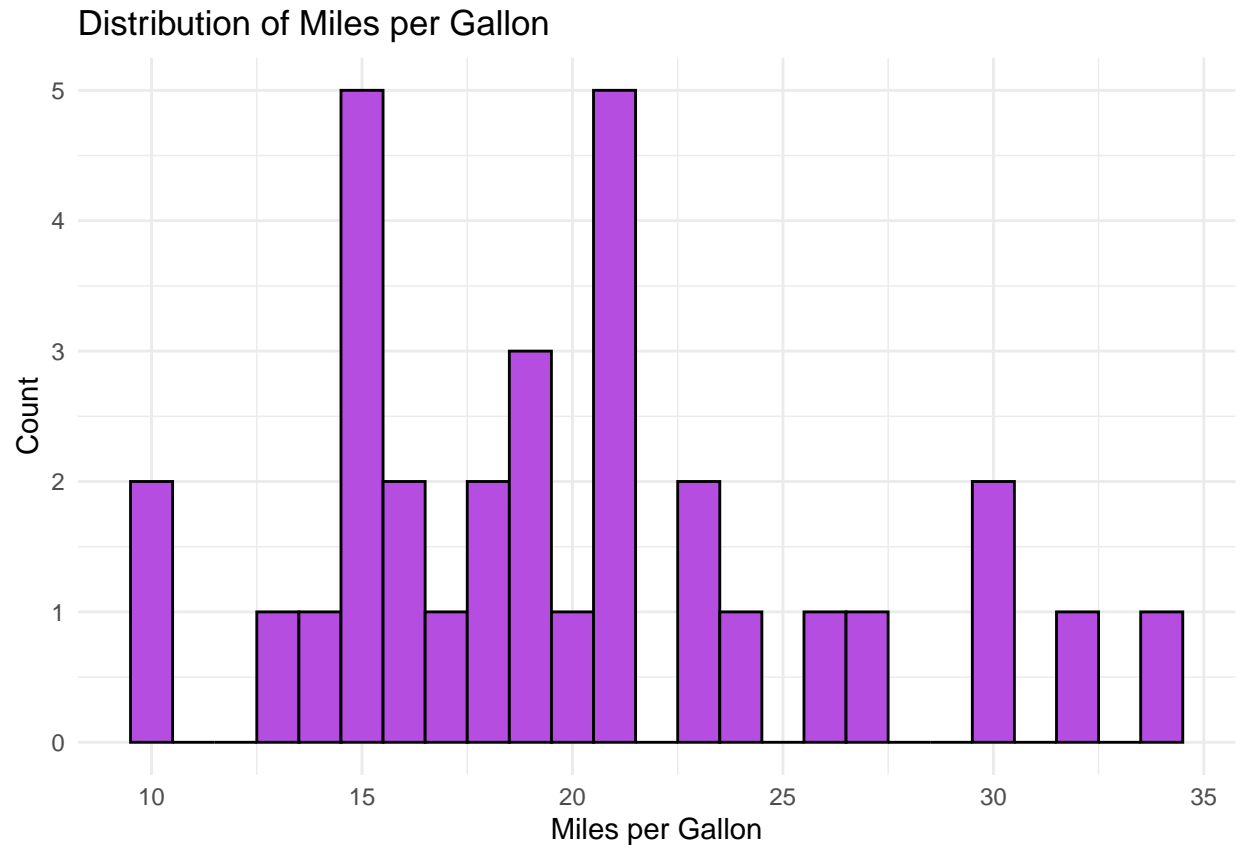
```
##           mpg           cyl           disp           hp           drat           wt
## mpg      1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl     -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp    -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp      -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat     0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt      -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec     0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs       0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am       0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear     0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb    -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##           qsec           vs           am           gear           carb
## mpg      0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl     -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp    -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp      -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat     0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt      -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec     1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs       0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am      -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear    -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb    -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

Distribution of a Single Variable

The histogram of miles per gallon (mpg) shows that most cars have mpg values between 15 and 25. The distribution appears slightly right-skewed, indicating that there are a few cars with exceptionally high mpg.

```
p <- ggplot(mtcars, aes(x=mpg)) +
  geom_histogram(binwidth=1, fill="#b44de0", color="black") +
  theme_minimal() +
  labs(title="Distribution of Miles per Gallon",
       x="Miles per Gallon",
       y="Count")

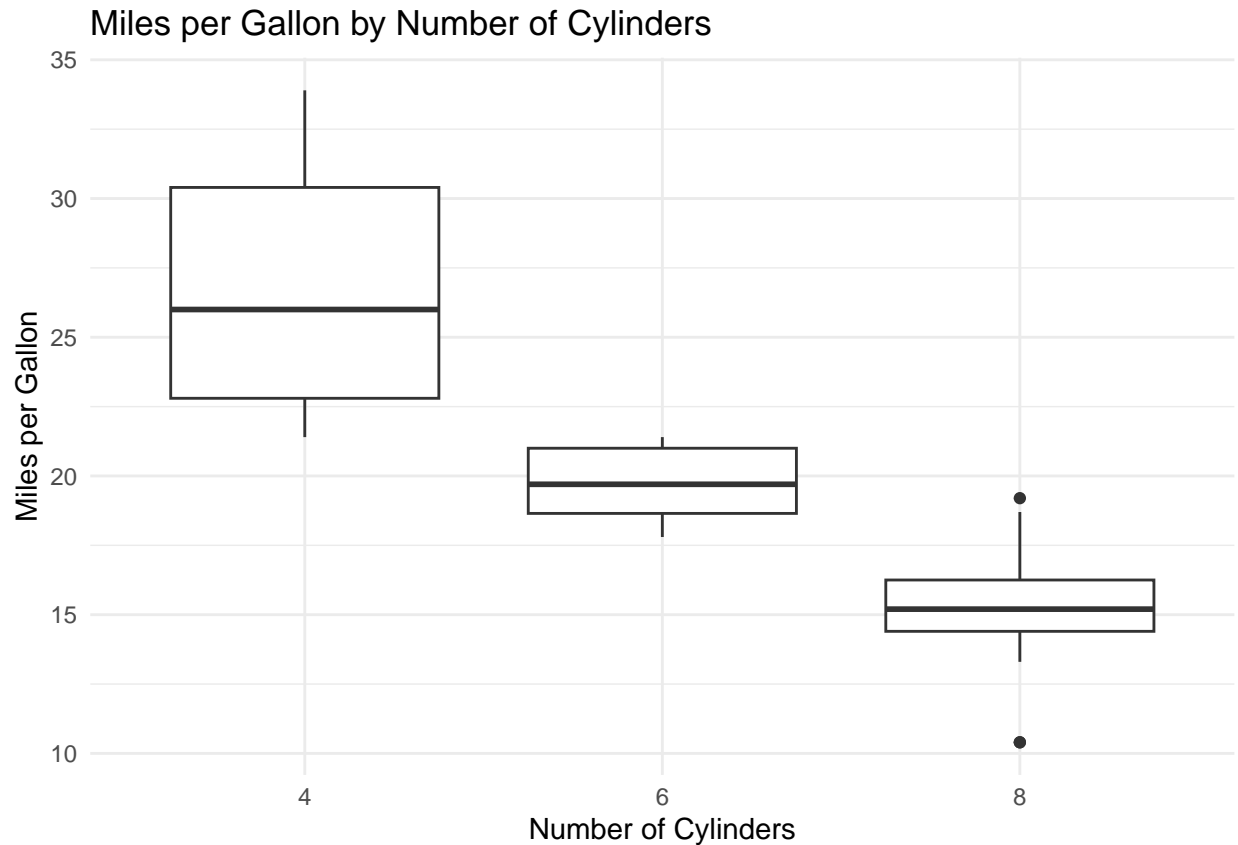
if (knitr::is_html_output()) {
  ggplotly(p)
} else {
  print(p)
}
```



Boxplots for categorical variables

The boxplot comparing mpg across different numbers of cylinders (cyl) shows that cars with fewer cylinders generally have higher mpg. Specifically, 4-cylinder cars have the highest median mpg, followed by 6-cylinder and then 8-cylinder cars. This indicates that cars with more cylinders tend to be less fuel-efficient.

```
p <- ggplot(mtcars, aes(x=factor(cyl), y=mpg)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title="Miles per Gallon by Number of Cylinders",
       x="Number of Cylinders",
       y="Miles per Gallon")
if (knitr::is_html_output()) {
  ggplotly(p)
} else {
  print(p)
}
```

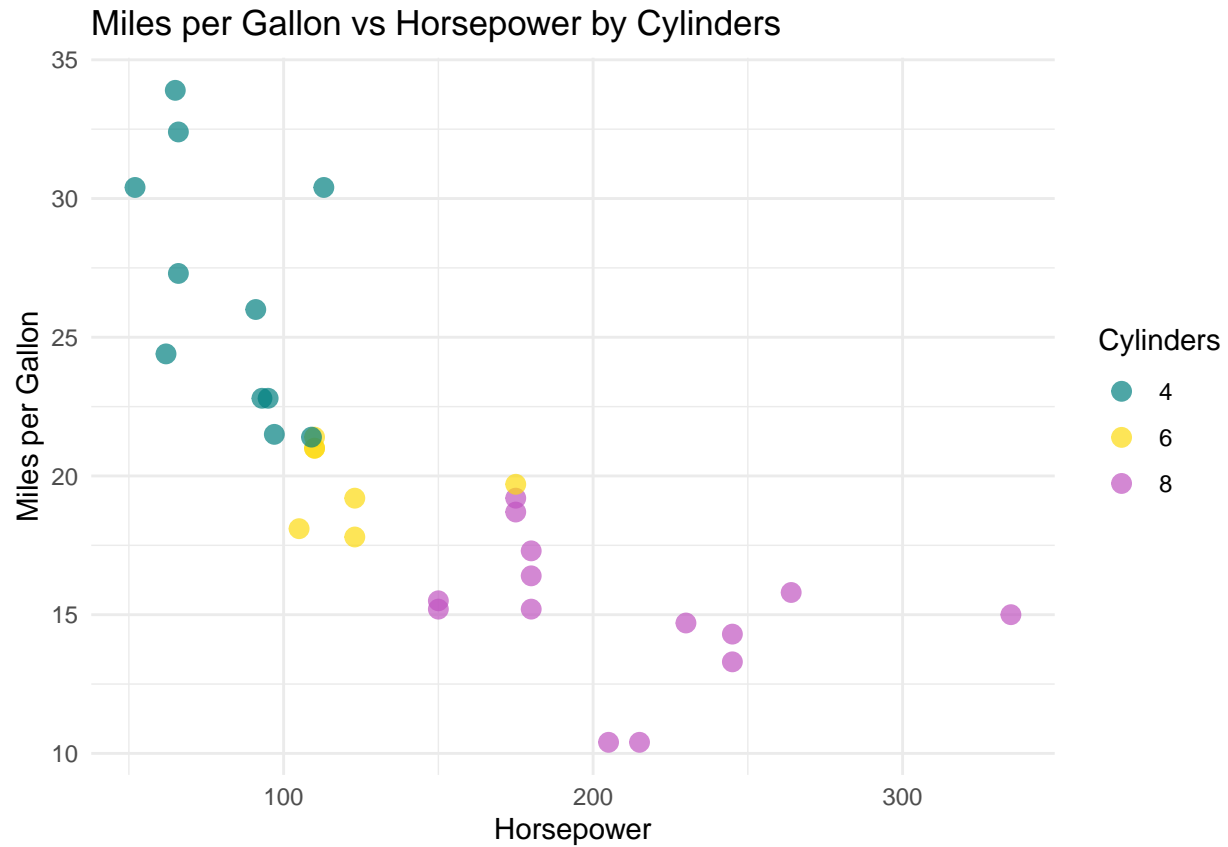


Facited plots

The faceted scatter plots show the relationship between horsepower (hp) and miles per gallon (mpg) across different numbers of cylinders (cyl). Each facet represents a subset of the data for a specific cylinder count, revealing that the negative relationship between hp and mpg is consistent across all cylinder groups, but cars with more cylinders generally have lower mpg.

```
p <- ggplot(mtcars, aes(x=hp, y=mpg, color=factor(cyl))) +
  geom_point(size = 3, alpha = 0.7) +
  theme_minimal() +
  labs(title="Miles per Gallon vs Horsepower by Cylinders",
       x="Horsepower",
       y="Miles per Gallon",
       color = "Cylinders") +
  scale_color_manual(values = c("4" = "#008080", "6" = "#FDDA0D", "8" = "#C154C1"))

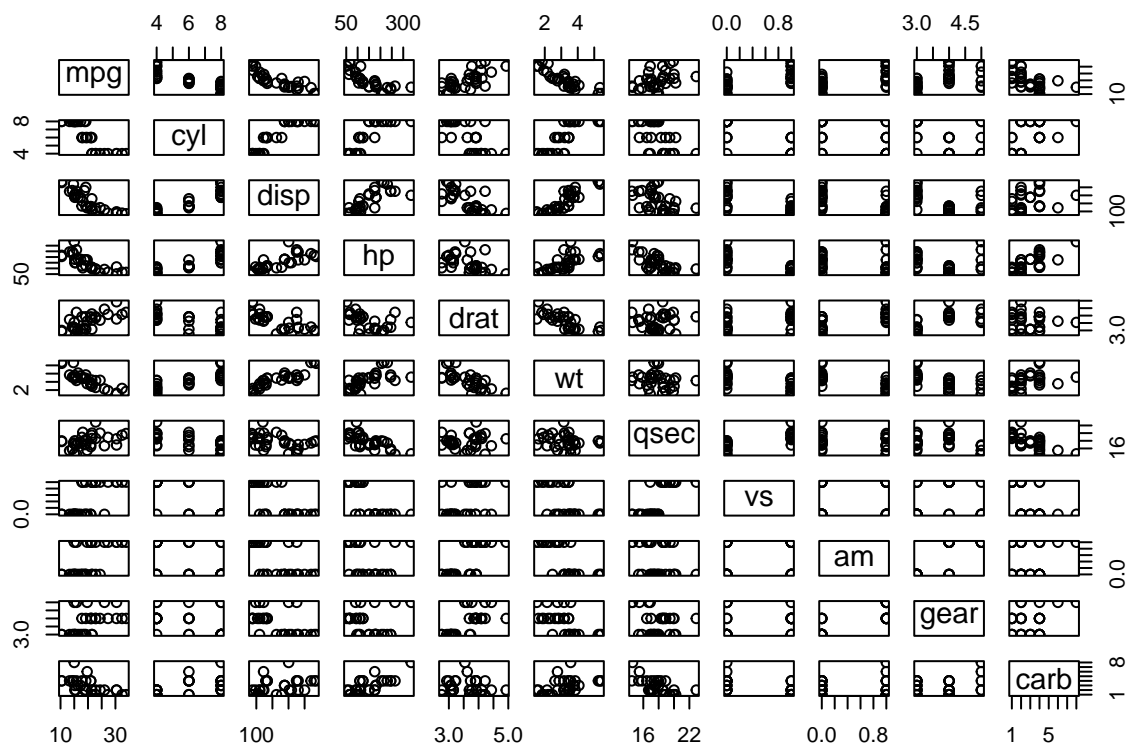
if (knitr::is_html_output()) {
  ggplotly(p)
} else {
  print(p)
}
```



Pairwise plots

The pairwise scatter plots provide a comprehensive view of relationships between all pairs of variables. We can observe that both weight (wt) and displacement (disp) have strong negative relationships with mpg, while positively correlating with each other. This helps identify multicollinearity and understand how variables interact with one another.

```
pairs(mtcars)
```



Linear Regression Analysis

The linear regression analysis models mpg as a function of horsepower (hp), weight (wt), and number of cylinders (cyl). The results show that all three variables significantly impact mpg, with weight and horsepower having the largest negative coefficients. This quantifies the earlier observations that heavier and more powerful cars are less fuel-efficient.

```
model <- lm(mpg ~ hp + wt + cyl, data=mtcars)
summary(model)
```

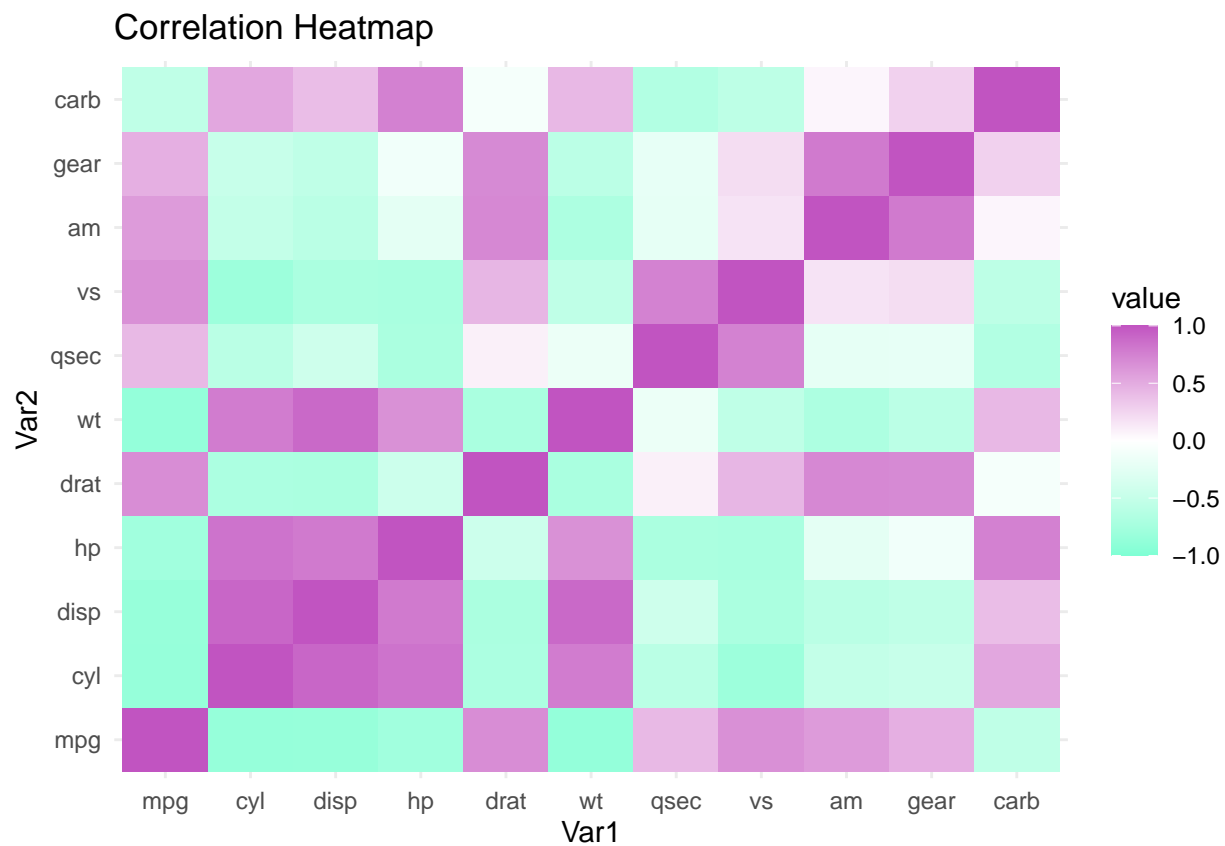
```
##
## Call:
## lm(formula = mpg ~ hp + wt + cyl, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9290 -1.5598 -0.5311  1.1850  5.8986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.75179    1.78686   21.687 < 2e-16 ***
## hp          -0.01804    0.01188   -1.519  0.140015
## wt          -3.16697    0.74058   -4.276  0.000199 ***
## cyl         -0.94162    0.55092   -1.709  0.098480 .
##
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.512 on 28 degrees of freedom
## Multiple R-squared:  0.8431, Adjusted R-squared:  0.8263
## F-statistic: 50.17 on 3 and 28 DF,  p-value: 2.184e-11
```

Heatmap of correlations

The heatmap visualizes the correlation matrix, showing strong negative correlations between mpg and both horsepower (hp) and weight (wt), and strong positive correlations between horsepower (hp) and weight (wt).

```
cor_matrix <- cor(mtcars)
melted_cor_matrix <- melt(cor_matrix)
p <- ggplot(melted_cor_matrix, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  scale_fill_gradient2(low="aquamarine", high="#C154C1", mid="white", midpoint=0, limit=c(-1,1)) +
  theme_minimal() +
  labs(title="Correlation Heatmap")
print(p)
```



Density plots

The density plot for miles per gallon (mpg) shows a smooth distribution curve, indicating the probability density of different mpg values. The plot confirms that most cars have mpg values around 20, with a long tail on the right side, indicating a few highly fuel-efficient cars.

```
p <- ggplot(mtcars, aes(x=mpg)) +  
  geom_density(fill="blue", alpha=0.5) +  
  theme_minimal() +  
  labs(title="Density Plot of Miles per Gallon",  
        x="Miles per Gallon",  
        y="Density")  
  
if (knitr::is_html_output()) {  
  ggplotly(p)  
} else {  
  print(p)  
}
```

