Assignment 1 Data Visualization

Yazeed Mshayekh 0202090

2024-03-19

Call the Required Frameworks

```
suppressWarnings(library(ggplot2))
```

Read the built-in mtcars data in R

```
data(mtcars)
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
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
```

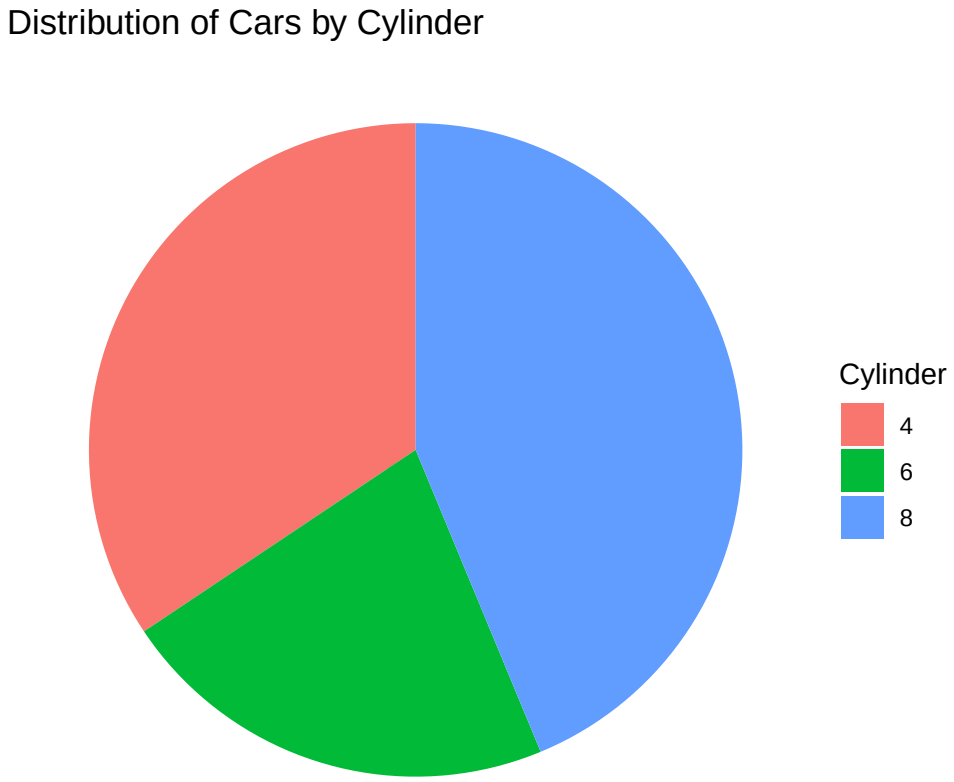
```
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Pie Chart of Car Distribution by Cylinder

Generate a pie chart using ggplot2 to illustrate the distribution of cars based on their cylinder (cyl) values from the mtcars dataset.

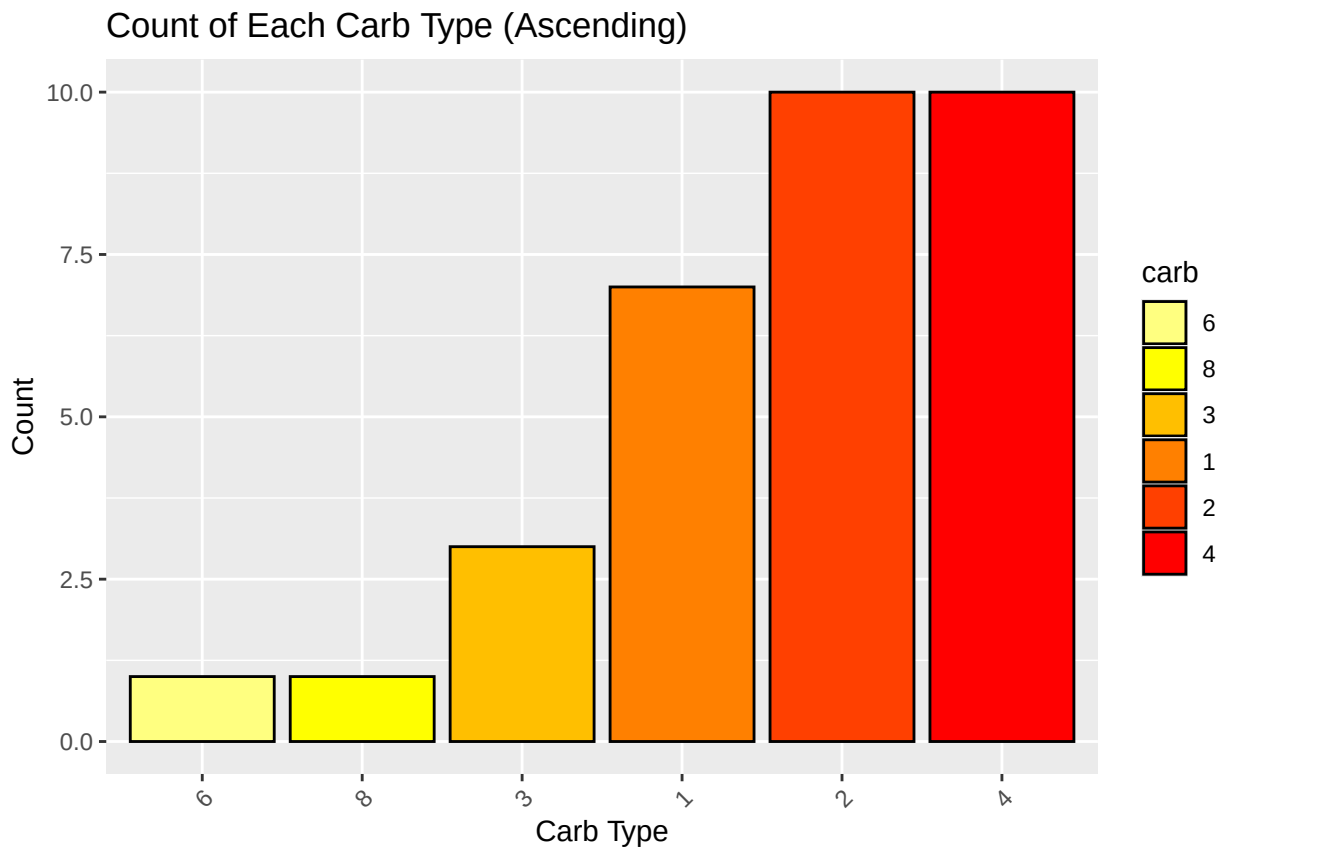
```
pie_data <- table(mtcars$cyl)
pie_labels <- paste(names(pie_data), "cylinders", sep = " ")
pie_chart <- ggplot() +
             geom_bar(aes(x = "", y = pie_data, fill = factor(names(pie_data))), stat = "identity") +
             coord_polar("y", start = 0) +
             theme_void() +
             labs(title = "Distribution of Cars by Cylinder", fill = "Cylinder")

print(pie_chart)
```

```
## Don't know how to automatically pick scale for object of type <table>.
## Defaulting to continuous.
```



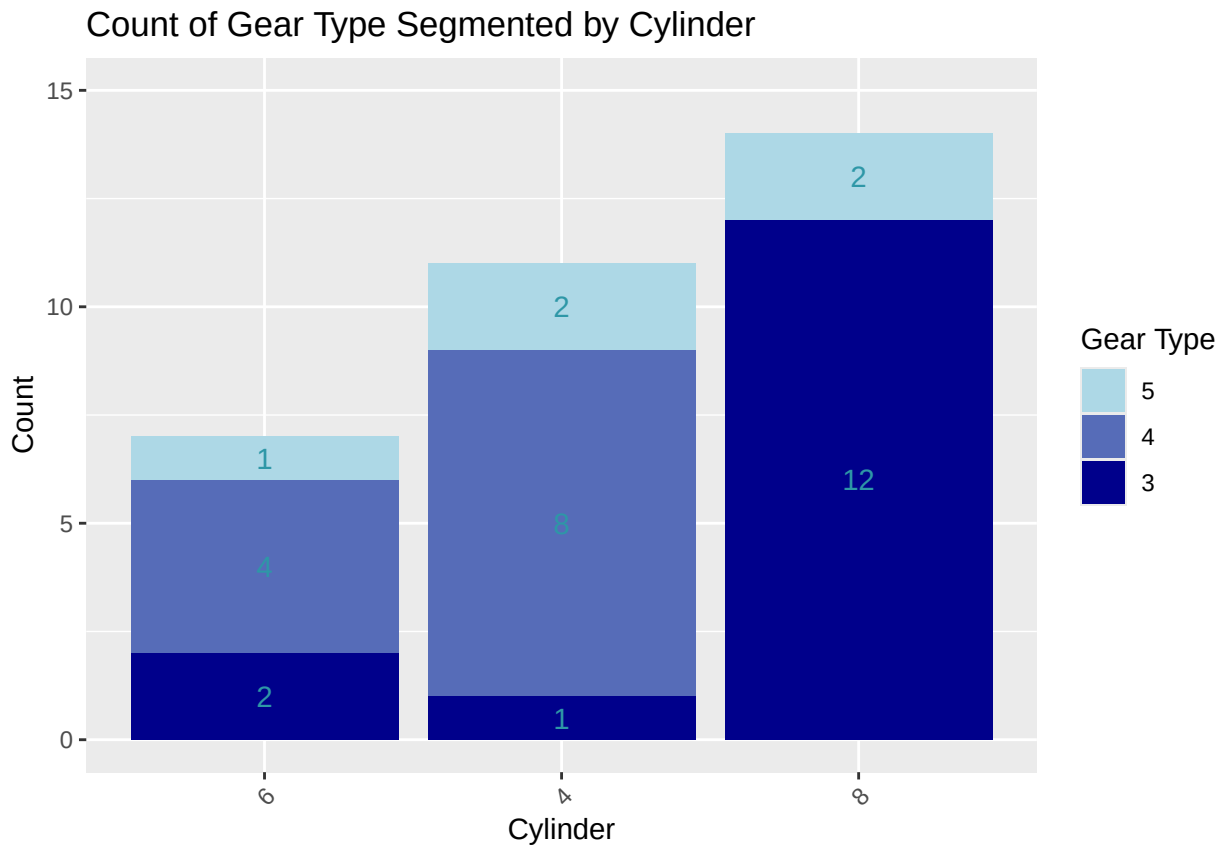
Bar Plot of Carb Type Count

```
suppressWarnings({
carb_counts <- table(mtcars$carb)

carb_levels <- names(sort(carb_counts))

mtcars$carb <- factor(mtcars$carb, levels = carb_levels)

color_range <- rev(heat.colors(length(carb_levels)))

color_palette <- colorRampPalette(color_range)(length(carb_levels))

carb_colors <- setNames(color_palette, carb_levels)

bar_plot <- ggplot(mtcars, aes(x = carb)) +
            geom_bar(aes(fill = carb), color = "black") +
            scale_fill_manual(values = carb_colors) +
            labs(title = "Count of Each Carb Type (Ascending)", x = "Carb Type", y = "Count") +
            theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(bar_plot)
})
```



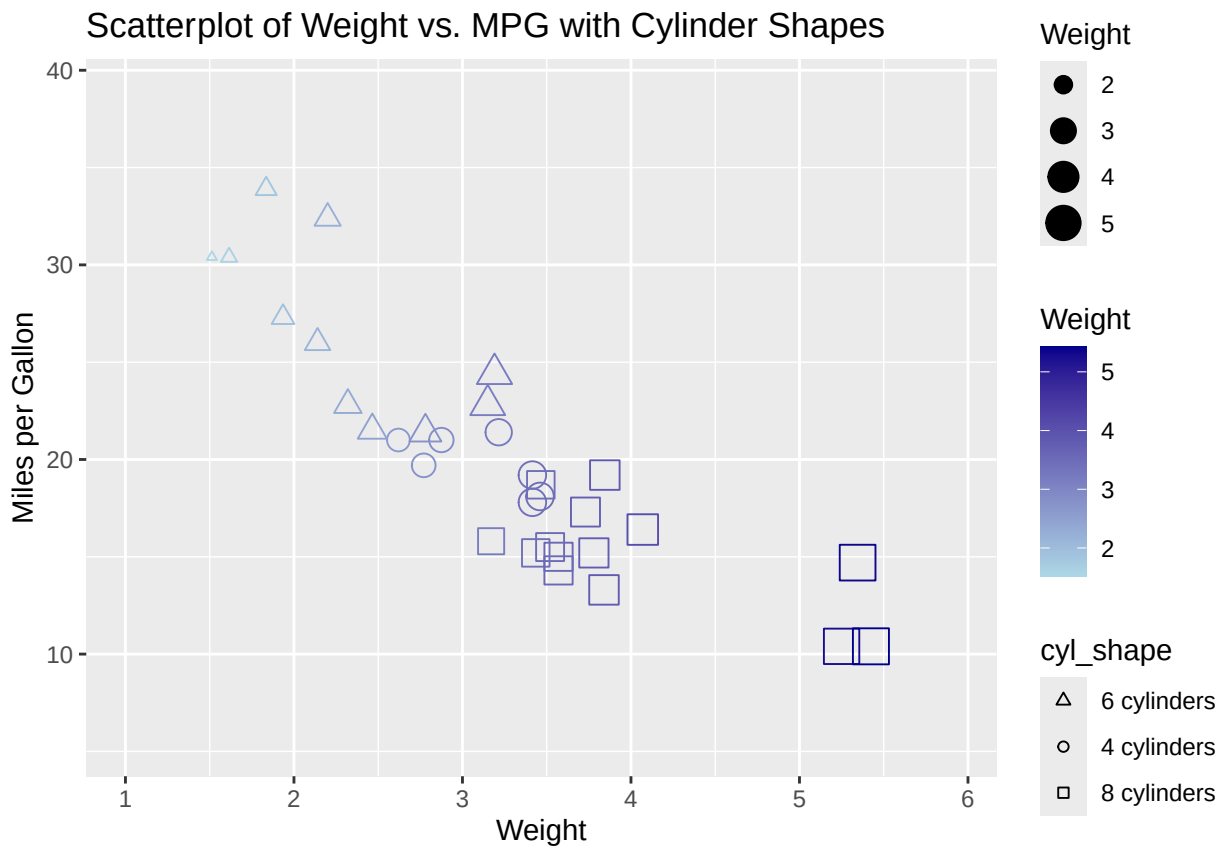
Stacked Bar Plot of Gear Type by Cylinder

```
suppressWarnings({
max_count <- max(table(mtcars$gear))

mtcars$cyl <- factor(mtcars$cyl, levels = unique(mtcars$cyl), ordered = TRUE)

gear_counts <- table(mtcars$gear)
gear_levels <- names(sort(gear_counts))

blue_palette <- colorRampPalette(c("lightblue", "darkblue"))(length(gear_levels))
gear_colors <- setNames(blue_palette, gear_levels)

stacked_bar <- ggplot(mtcars, aes(x = cyl, fill = factor(gear, levels = gear_levels))) +
               geom_bar(position = "stack") +
               scale_fill_manual(values = gear_colors) +
               labs(title = "Count of Gear Type Segmented by Cylinder", x = "Cylinder", y = "Count", fi
               theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
               geom_text(aes(label = stat(count)), stat = "count", position = position_stack(vjust = 0.5
               ylim(0, max_count)

print(stacked_bar)

})
```



##

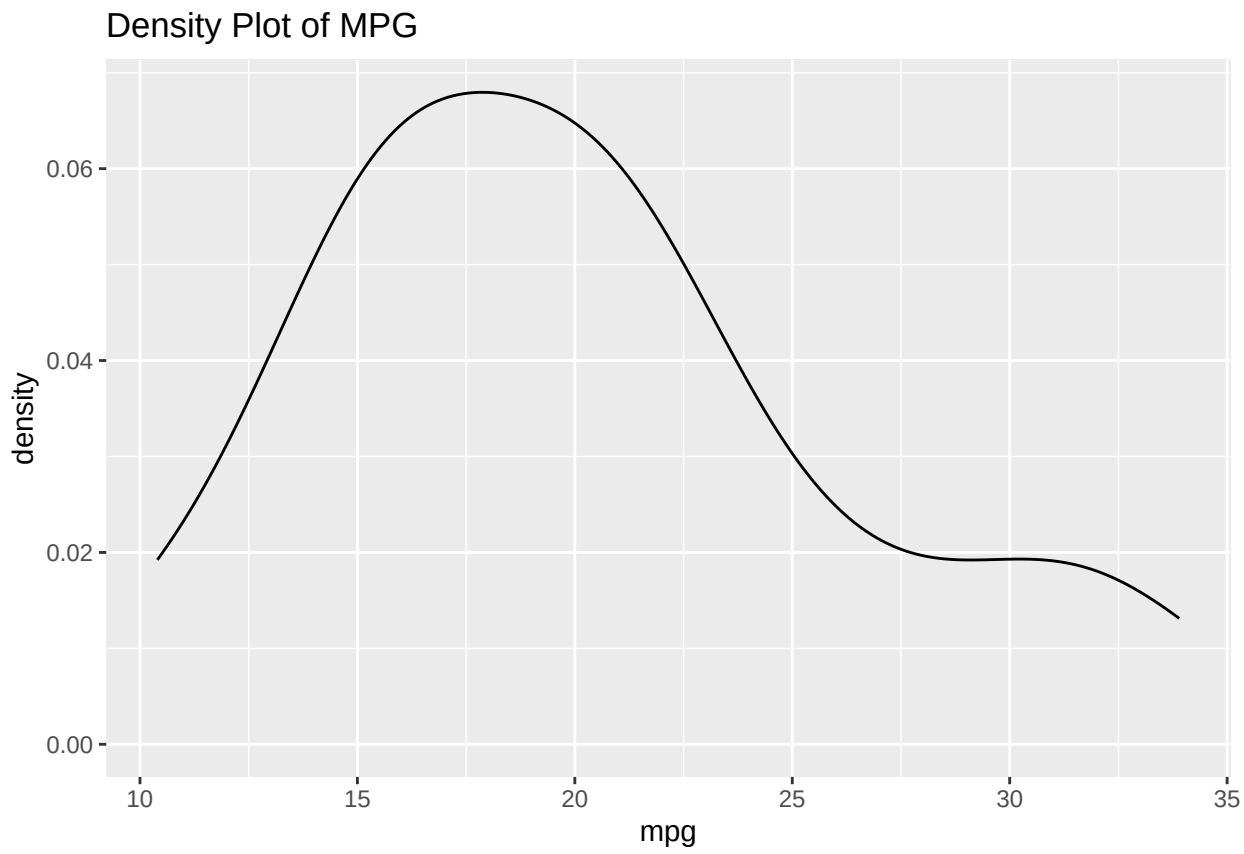
```r
suppressWarnings({

mtcars$cyl_shape <- factor(ifelse(mtcars$cyl == 6, "6 cylinders", ifelse(mtcars$cyl == 4, "4 cylinders"

color_palette <- colorRampPalette(c("lightblue", "darkblue"))(4)

mtcars$size <- mtcars$wt

x_range <- c(min(mtcars$wt) - 0.5, max(mtcars$wt) + 0.5)
y_range <- c(min(mtcars$mpg) - 5, max(mtcars$mpg) + 5)

scatterplot <- ggplot(mtcars, aes(x = wt, y = mpg, shape = cyl_shape, size = size, color = wt)) +
               geom_point(position = position_dodge(width = 0.1)) +
               labs(title = "Scatterplot of Weight vs. MPG with Cylinder Shapes", x = "Weight", y = "Mi
               scale_shape_manual(values = c("6 cylinders" = 1, "4 cylinders" = 2, "8 cylinders" = 0),
                                  labels = c("6 cylinders", "4 cylinders", "8 cylinders")) +
               scale_color_gradient(low = color_palette[1], high = color_palette[4], limits = range(mtc
               guides(size = guide_legend(title = "Weight"), color = guide_colorbar(title = "Weight")) +
               xlim(x_range) +
               ylim(y_range)

print(scatterplot)

})
```



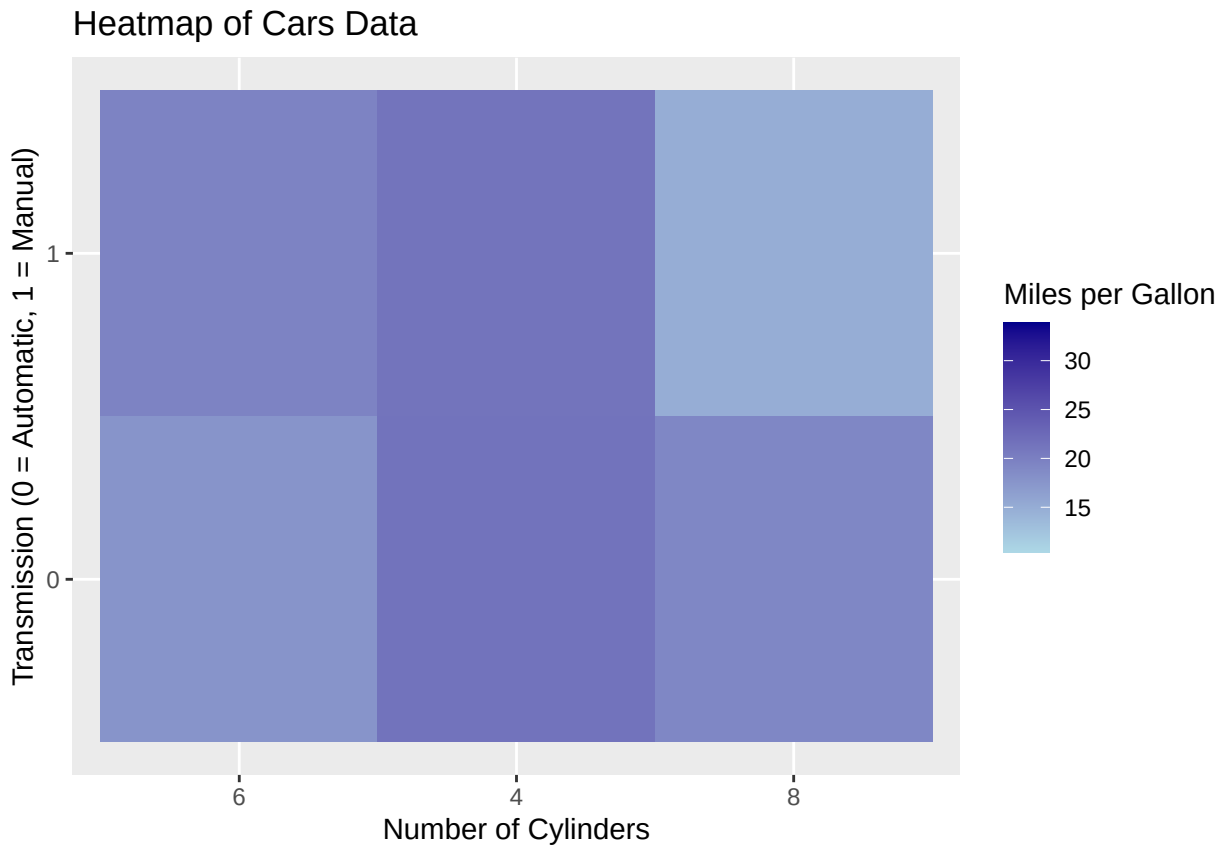
Other Plots

```r
# Density Plot
density_plot <- ggplot(mtcars, aes(x = mpg)) +
                geom_density() +
                labs(title = "Density Plot of MPG")

# Heatmap
heatmap <- heatmap <- ggplot(mtcars, aes(x = factor(cyl), y = factor(am), fill = mpg)) +
           geom_tile() +
           scale_fill_gradient(low = "lightblue", high = "darkblue") +
           labs(title = "Heatmap of Cars Data", x = "Number of Cylinders", y = "Transmission (0 = Automa

# Dot Plot
dot_plot <- ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
            geom_dotplot(binaxis = "y", stackdir = "center", fill = "blue") +
            labs(title = "Dot Plot of MPG by Cylinder", x = "Cylinder", y = "Miles per Gallon")

# ECDF Plot
ecdf_plot <- ggplot(mtcars, aes(x = mpg)) +
             stat_ecdf() +
             labs(title = "ECDF of MPG")

# Q-Q Plot
qq_plot <- ggplot(mtcars, aes(sample = mpg)) +
           stat_qq() +
           labs(title = "Q-Q Plot of MPG")

# Comment on best display method
comment <- "Each type of plot has its own advantages depending on the purpose. For exploring distributi

list(density_plot, heatmap, dot_plot, ecdf_plot, qq_plot, comment)
```

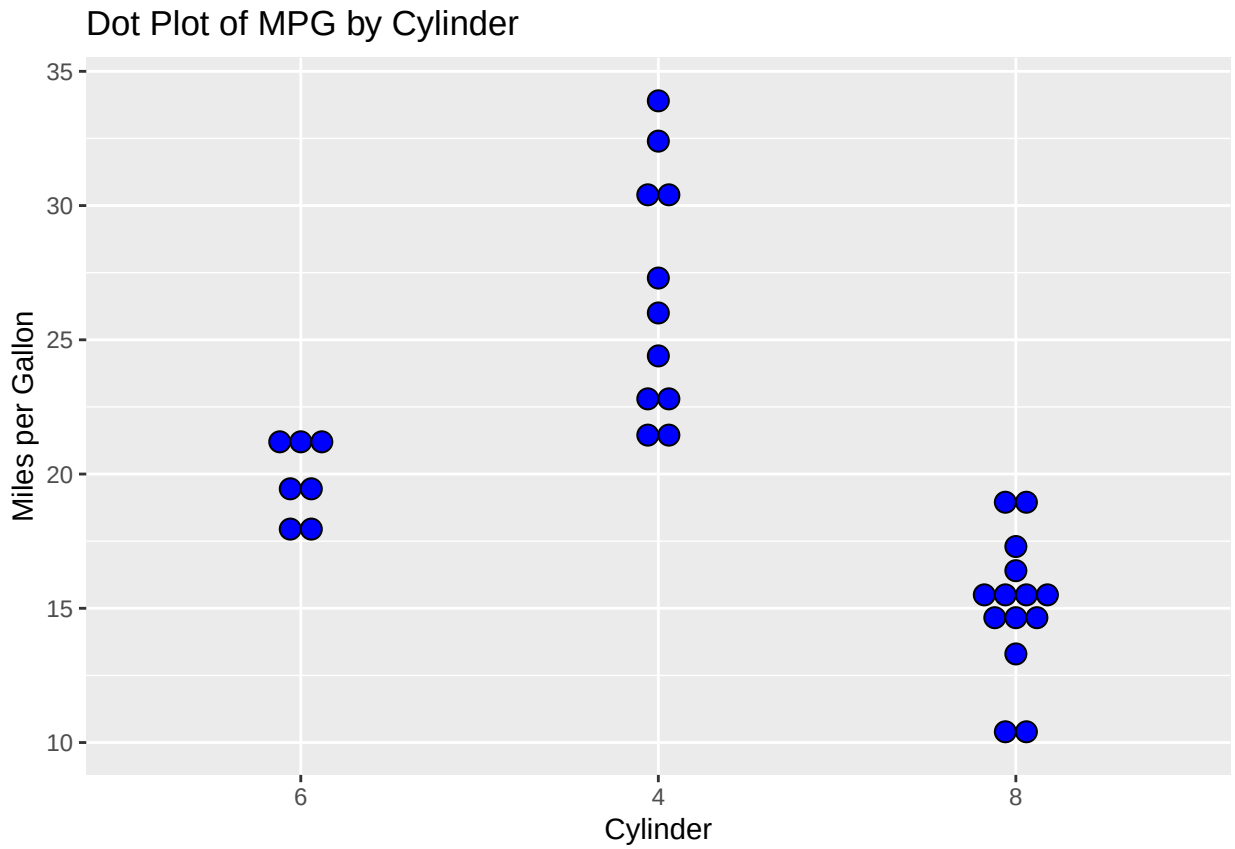
```
## [[1]]
```


Density Plot of MPG
density
0.06
0.04
0.02
0.00
10
15
20
25
30
35
mpg

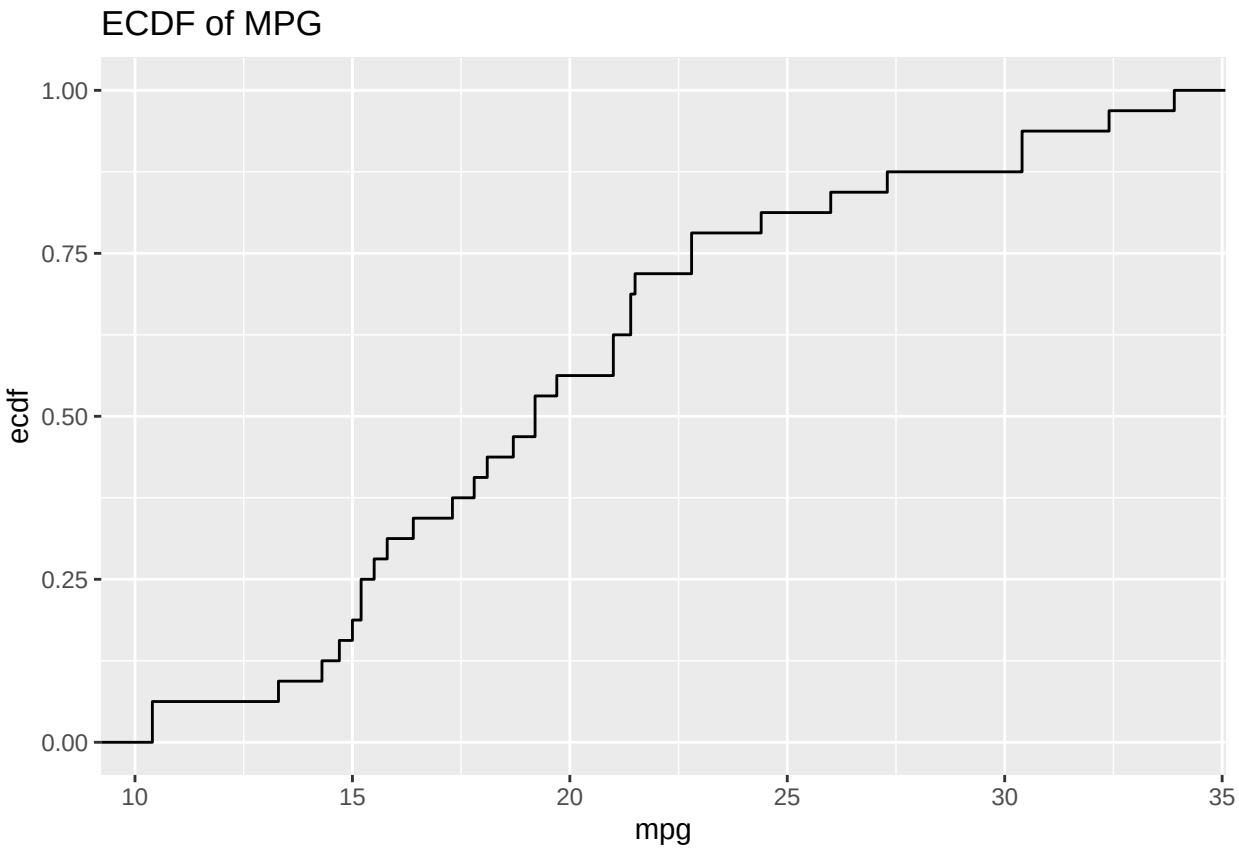
```
##
## [[2]]
```



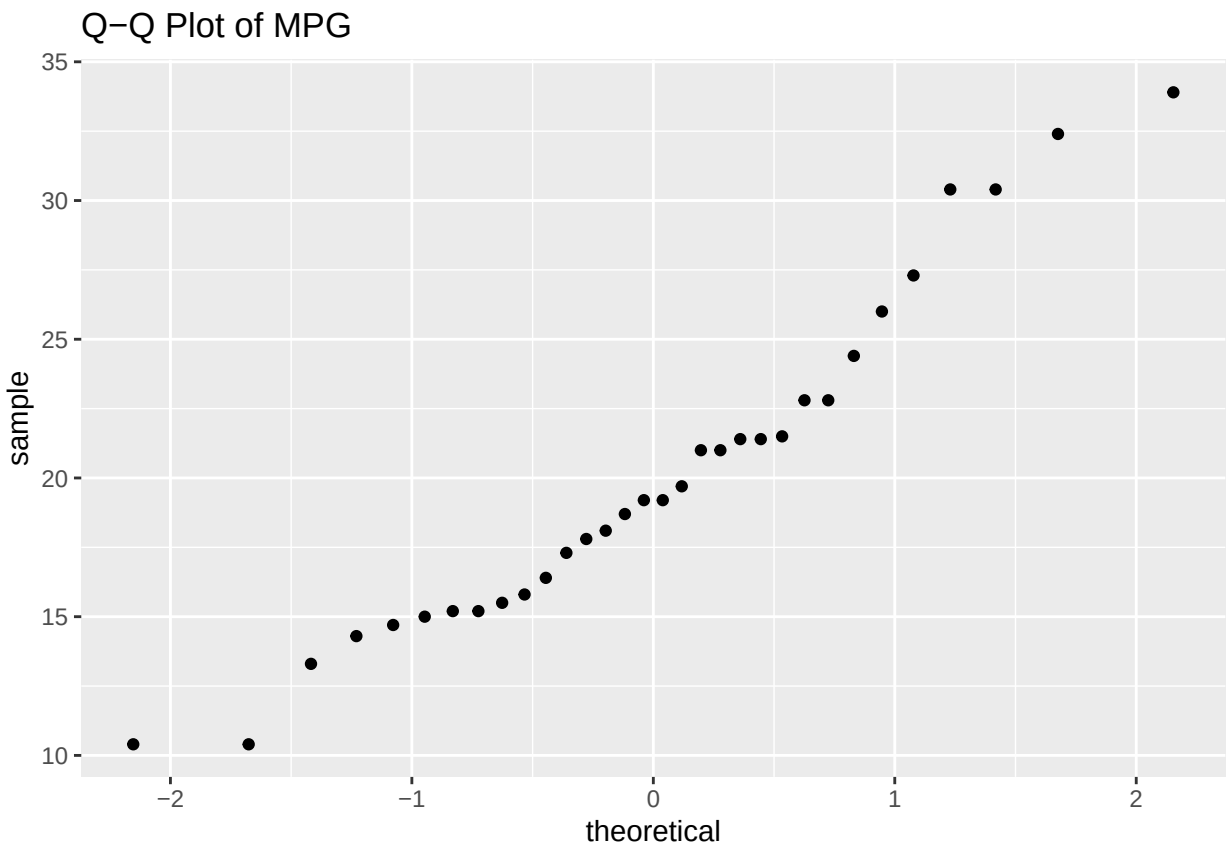

```
##
## [[3]]

## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```


Dot Plot of MPG by Cylinder
Miles per Gallon
35
30
25
20
15
10
6
4
8
Cylinder

```
##
## [[4]]
```




```
##
## [[5]]
```




```
##
## [[6]]
## [1] "Each type of plot has its own advantages depending on the purpose. For exploring distributions,
```

Dataset

```
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

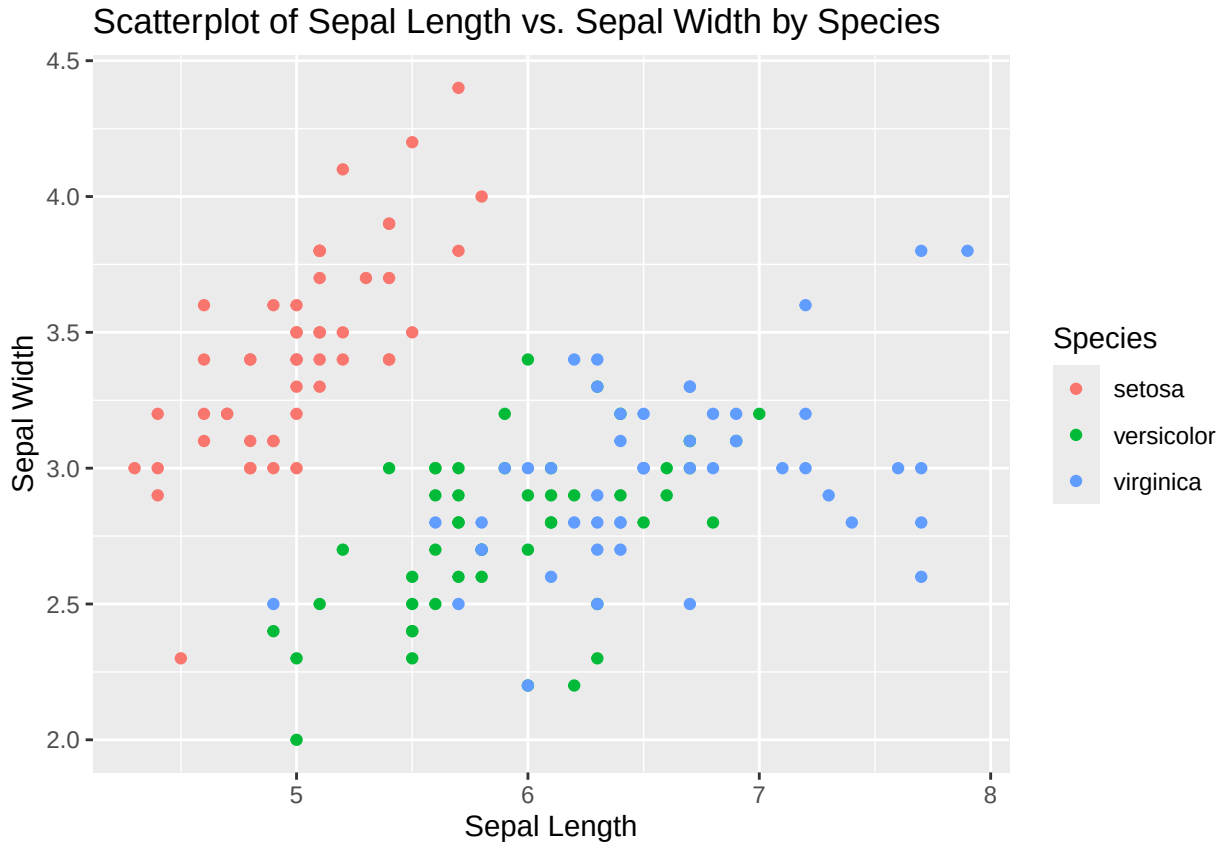
```
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
```

```
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

Scatter of Sepal Length vs. Sepal Width by Species

```
scatterplot <- ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +
               geom_point() +
               labs(title = "Scatterplot of Sepal Length vs. Sepal Width by Species",
                    x = "Sepal Length", y = "Sepal Width", color = "Species")

print(scatterplot)
```



Explanation

- **Dataset Selection**: The Iris dataset was chosen because it contains measurements of iris flowers from three species, making it suitable for showing how to visualize and group data.
- **Visualization Type and Rationale**: A scatterplot was chosen because it effectively shows the relationship between two measurements (Sepal.Length and Sepal.Width). By using colors to represent species, we can easily see differences and patterns among them. This helps in understanding the characteristics of the Iris dataset.