# Week 6 Homework - mtcars dataset analysis

## Step 1: Load the mtcars dataset

data("mtcars")

## Step 2: Double the dataset by appending it to itself

extended\_mtcars <- rbind(mtcars, mtcars)  
  
# Print the dimensions to confirm that the data has been doubled  
cat("Original dataset dimensions: ", dim(mtcars), "\n")

## Original dataset dimensions: 32 11

cat("Extended dataset dimensions: ", dim(extended\_mtcars), "\n")

## Extended dataset dimensions: 64 11

## Step 3: Perform summary statistics for the original and extended datasets

# Original dataset summary  
cat("\nSummary statistics for the original dataset:\n")

##   
## Summary statistics for the original dataset:

print(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

# Extended dataset summary  
cat("\nSummary statistics for the extended dataset:\n")

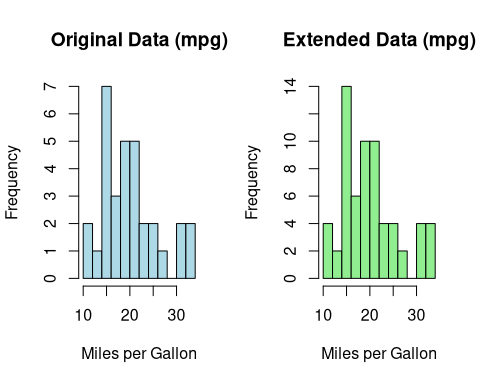
##   
## Summary statistics for the extended dataset:

print(summary(extended\_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

## Step 4: Create visualizations to compare the original and extended datasets

# Set plotting area to show two plots side by side  
par(mfrow = c(1, 2))  
  
# Plot histogram for 'mpg' (Miles Per Gallon) in the original dataset  
hist(mtcars$mpg, main = "Original Data (mpg)",   
 xlab = "Miles per Gallon", col = "lightblue",   
 border = "black", breaks = 10)  
  
# Plot histogram for 'mpg' (Miles Per Gallon) in the extended dataset  
hist(extended\_mtcars$mpg, main = "Extended Data (mpg)",   
 xlab = "Miles per Gallon", col = "lightgreen",   
 border = "black", breaks = 10)



## Step 5: Explanation of differences

cat("\nExplanation of Results:\n")

##   
## Explanation of Results:

cat("Since we duplicated the dataset, the distribution of 'mpg' values remains the same, but the number of observations has doubled. This means the histograms look similar, and summary statistics should reflect identical measures of central tendency and spread, but with doubled row counts.\n")

## Since we duplicated the dataset, the distribution of 'mpg' values remains the same, but the number of observations has doubled. This means the histograms look similar, and summary statistics should reflect identical measures of central tendency and spread, but with doubled row counts.