Exploring the mtcars Dataset

Ken Wood

2024-10-05

Overview: In this assignment, you will dive into the world of automotive data using the `mtcars` dataset available in the ggplot2 package. Similar to our previous lesson where we explored the preloaded diamonds dataset, you will apply various data manipulation and visualization techniques to uncover insights and analyze automotive trends.

Data Overview:

The mtcars dataset contains data on various automobile models, including attributes such as miles per gallon (mpg), number of cylinders (cyl), engine displacement (disp), horsepower (hp), and more. To learn more about the dataset, including a data dictionary, you can use the following R command: `?mtcars`.

```
# Load the datasets package (usually not necessary as it's loaded by default)
library(datasets)
# Import the mtcars dataset
data(mtcars)
# Display the first few rows of the dataset
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
```

1. Selecting Columns: Use the select() function to choose only the columns mpg, cyl, disp, hp, and drat. Save the resulting dataframe as selected_cars.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
selected_cars <- mtcars %>% select(mpg, cyl, disp, hp, drat)
selected_cars
```

```
##                    mpg cyl  disp  hp drat
## Mazda RX4         21.0   6 160.0 110 3.90
```

```
## Mazda RX4 Wag       21.0   6 160.0 110 3.90
## Datsun 710          22.8   4 108.0  93 3.85
## Hornet 4 Drive      21.4   6 258.0 110 3.08
## Hornet Sportabout   18.7   8 360.0 175 3.15
## Valiant             18.1   6 225.0 105 2.76
## Duster 360          14.3   8 360.0 245 3.21
## Merc 240D           24.4   4 146.7  62 3.69
## Merc 230            22.8   4 140.8  95 3.92
## Merc 280            19.2   6 167.6 123 3.92
## Merc 280C           17.8   6 167.6 123 3.92
## Merc 450SE          16.4   8 275.8 180 3.07
## Merc 450SL          17.3   8 275.8 180 3.07
## Merc 450SLC         15.2   8 275.8 180 3.07
## Cadillac Fleetwood  10.4   8 472.0 205 2.93
## Lincoln Continental 10.4   8 460.0 215 3.00
## Chrysler Imperial   14.7   8 440.0 230 3.23
## Fiat 128            32.4   4  78.7  66 4.08
## Honda Civic         30.4   4  75.7  52 4.93
## Toyota Corolla      33.9   4  71.1  65 4.22
## Toyota Corona       21.5   4 120.1  97 3.70
## Dodge Challenger    15.5   8 318.0 150 2.76
## AMC Javelin         15.2   8 304.0 150 3.15
## Camaro Z28          13.3   8 350.0 245 3.73
## Pontiac Firebird    19.2   8 400.0 175 3.08
## Fiat X1-9           27.3   4  79.0  66 4.08
## Porsche 914-2       26.0   4 120.3  91 4.43
## Lotus Europa        30.4   4  95.1 113 3.77
## Ford Pantera L      15.8   8 351.0 264 4.22
## Ferrari Dino        19.7   6 145.0 175 3.62
## Maserati Bora       15.0   8 301.0 335 3.54
## Volvo 142E          21.4   4 121.0 109 4.11
```

2. Filtering Data: Filter the mtcars dataset to include only cars with more than 150 horsepower (hp). Save the filtered dataframe as high_hp_cars.

```
high_hp_cars <- mtcars %>% filter(mtcars$hp > 150)
high_hp_cars
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
```

3. Arranging Data: Arrange the high_hp_cars dataframe in descending order of miles per gallon (mpg). Save the arranged dataframe as sorted_cars.

```
sorted_cars <- high_hp_cars %>% sort_by(high_hp_cars$mpg,decreasing=TRUE)
sorted_cars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
```

4. Using the Pipe Operator (%>%):

Rewrite the previous tasks using the pipe operator (%>%) to create a streamlined workflow.

5. Adding a New Column:

Use the mutate() function to create a new column named mileage_category in the mtcars dataset. Classify cars with mpg greater than or equal to 20 as "High Mileage" and others as "Low Mileage".

```
# Use mutate to create a new column mileage_category
mtcars <- mtcars %>%
  mutate(mileage_category = ifelse(mpg >= 20, "High Mileage", "Low Mileage"))

mtcars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
```

```
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
##                     mileage_category
## Mazda RX4                 High Mileage
## Mazda RX4 Wag             High Mileage
## Datsun 710                High Mileage
## Hornet 4 Drive            High Mileage
## Hornet Sportabout          Low Mileage
## Valiant                    Low Mileage
## Duster 360                 Low Mileage
## Merc 240D                 High Mileage
## Merc 230                  High Mileage
## Merc 280                   Low Mileage
## Merc 280C                  Low Mileage
## Merc 450SE                 Low Mileage
## Merc 450SL                 Low Mileage
## Merc 450SLC                Low Mileage
## Cadillac Fleetwood         Low Mileage
## Lincoln Continental        Low Mileage
## Chrysler Imperial          Low Mileage
## Fiat 128                  High Mileage
## Honda Civic               High Mileage
## Toyota Corolla            High Mileage
## Toyota Corona             High Mileage
## Dodge Challenger           Low Mileage
## AMC Javelin                Low Mileage
## Camaro Z28                 Low Mileage
## Pontiac Firebird           Low Mileage
## Fiat X1-9                 High Mileage
## Porsche 914-2             High Mileage
## Lotus Europa              High Mileage
## Ford Pantera L             Low Mileage
## Ferrari Dino               Low Mileage
## Maserati Bora              Low Mileage
## Volvo 142E                High Mileage
```

6. Grouping and Summarizing:

Group the mtcars dataset by the number of cylinders (cyl) and summarize the average horsepower (hp) for each cylinder category.

```
grouped_data <- mtcars %>% group_by(cyl) %>% summarize(avg_hp = mean(hp))
grouped_data
```

```
## # A tibble: 3 x 2
##     cyl avg_hp
##   <dbl>  <dbl>
## 1     4   82.6
## 2     6  122.
```

```
## 3     8  209.
```