FuelEconomyData.R

lokesko

2024-08-15

```r
# Uncomment the following line if you need to install the packagea
#install.packages("readr")
#install.packages("ggplot2")
#install.packages("dplyr")

# 1. Load necessary libraries
library(readr)  # For reading and writing CSV files
library(ggplot2)
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
```

```r
# 2. Read the CSV file
# Specify the path to your CSV file
file_path <- "vehicles.csv"
vehicle <- read_csv(file_path)
```

```
## Rows: 47523 Columns: 84

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (23): drive, eng_dscr, fuelType, fuelType1, make, model, mpgData, trany,...
## dbl (59): barrels08, barrelsA08, charge120, charge240, city08, city08U, city...
## lgl  (2): phevBlended, tCharger
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
df <- vehicle[, c("make", "model","year", "cylinders","trany", "fuelType",
                  "fuelType1", "range", "rangeCity", "rangeHwy","UCity", "UHighway")]

# Summary statistics for numerical columns
summary(df)
```

```
##      make               model                year        cylinders
##  Length:47523       Length:47523       Min.   :1984   Min.   : 2.000
##  Class :character   Class :character   1st Qu.:1993   1st Qu.: 4.000
##  Mode  :character   Mode  :character   Median :2005   Median : 6.000
##                                        Mean   :2004   Mean   : 5.702
##                                        3rd Qu.:2015   3rd Qu.: 6.000
##                                        Max.   :2025   Max.   :16.000
##                                                       NA's   :801
##     trany              fuelType          fuelType1             range
##  Length:47523       Length:47523       Length:47523       Min.   :  0.000
##  Class :character   Class :character   Class :character   1st Qu.:  0.000
##  Mode  :character   Mode  :character   Mode  :character   Median :  0.000
##                                                           Mean   :  7.358
##                                                           3rd Qu.:  0.000
##                                                           Max.   :640.000
##
##    rangeCity          rangeHwy           UCity           UHighway
##  Min.   :  0.000   Min.   :  0.000   Min.   :  0.00   Min.   :  0.0
##  1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.: 18.70   1st Qu.: 28.2
##  Median :  0.000   Median :  0.000   Median : 22.00   Median : 33.9
##  Mean   :  1.632   Mean   :  1.518   Mean   : 25.07   Mean   : 36.0
##  3rd Qu.:  0.000   3rd Qu.:  0.000   3rd Qu.: 26.67   3rd Qu.: 40.0
##  Max.   :520.800   Max.   :520.500   Max.   :224.80   Max.   :187.1
##
```

```r
# Summary of categorical columns
table(df$make)
```

```
##
##                          Acura                     Alfa Romeo
##                            415                             94
##                     AM General  American Motors Corporation
##                              6                             27
##               ASC Incorporated                   Aston Martin
##                              1                            183
##                           Audi                 Aurora Cars Ltd
##                           1271                              1
##              Autokraft Limited        Avanti Motor Corporation
##                              4                              2
##                 Azure Dynamics                        Bentley
##                              1                            174
##                        Bertone    Bill Dovell Motor Car Company
##                              7                              4
##          Bitter Gmbh and Co. Kg                            BMW
##                              5                           2469
##                     BMW Alpina                        Bugatti
##                              3                             21
##                          Buick                            BYD
##                            716                              7
##                       Cadillac                CCC Engineering
##                            754                              2
##                      Chevrolet                       Chrysler
##                           4420                            759
##                CODA Automotive        Consulier Industries Inc
```

```
##                                2                                3
##                     CX Automotive          Dabryan Coach Builders Inc
##                               17                                9
##                            Dacia                           Daewoo
##                                3                               67
##                          Daihatsu                           Dodge
##                               17                             2690
##                  E. P. Dutton, Inc.                           Eagle
##                                1                              161
##   Environmental Rsch and Devp Corp                 Evans Automobiles
##                                1                                3
##                     Excalibur Autos                  Federal Coach
##                                1                               14
##                            Ferrari                           Fiat
##                              275                               80
##                           Fisker                           Ford
##                                2                             3821
##                     General Motors                         Genesis
##                                1                              123
##                              Geo                            GMC
##                              147                             2777
##                         Goldacre           Grumman Allied Industries
##                                1                                1
##                     Grumman Olson                          Honda
##                                4                             1183
##                           Hummer                         Hyundai
##                               19                              991
##      Import Foreign Auto Sales Inc              Import Trade Services
##                                1                               13
##                   INEOS Automotive                       Infiniti
##                                4                              477
##                   Isis Imports Ltd                         Isuzu
##                                1                              434
##                      J.K. Motors                         Jaguar
##                               36                              540
##                 JBA Motorcars, Inc.                          Jeep
##                                1                             1109
##                            Kandi                          Karma
##                                1                                5
##      Kenyon Corporation Of America                           Kia
##                                4                              801
##                         Koenigsegg              Laforza Automobile Inc
##                                3                                2
##              Lambda Control Systems                    Lamborghini
##                                1                              161
##                         Land Rover                          Lexus
##                              340                              694
##                           Lincoln              London Coach Co Inc
##                              420                                1
##                       London Taxi                       Lordstown
##                                1                                1
##                             Lotus                           Lucid
##                               66                               24
##                           Mahindra                        Maserati
```

```
##                                1                                209
##                          Maybach                              Mazda
##                               31                               1111
##                     Mcevoy Motors                McLaren Automotive
##                                6                                 48
##                     Mercedes-Benz                            Mercury
##                             1914                                609
##                            Merkur                               MINI
##                               14                                530
##                         Mitsubishi              Mobility Ventures LLC
##                             1131                                  4
##                           Morgan                             Nissan
##                                3                               1663
##                        Oldsmobile                             Pagani
##                              462                                  5
##                            Panos              Panoz Auto-Development
##                                1                                  1
##        Panther Car Company Limited                     PAS Inc - GMC
##                                4                                  2
##                          PAS, Inc                           Peugeot
##                                2                                 98
##                       Pininfarina                          Plymouth
##                                6                                526
##                         Polestar                           Pontiac
##                               14                                893
##                          Porsche                Quantum Technologies
##                             1469                                  2
##                            Qvale                               Ram
##                                1                                157
##                   Red Shift Ltd.                           Renault
##                                2                                 56
##                           Rivian                        Rolls-Royce
##                               67                                232
##                  Roush Performance                     RUF Automobile
##                               69                                  4
##               Ruf Automobile Gmbh S and S Coach Company  E.p. Dutton
##                                3                                  1
##                              Saab                             Saleen
##                              432                                  5
##                Saleen Performance                            Saturn
##                                5                                278
##                             Scion                            Shelby
##                               84                                  1
##                            smart                            Spyker
##                               38                                 13
##                              SRT                           Sterling
##                                2                                 12
##                              STI                            Subaru
##                                1                               1004
##   Superior Coaches Div E.p. Dutton                            Suzuki
##                                1                                515
##                       Tecstar, LP                             Tesla
##                                6                                144
##               Texas Coach Company                           Toyota
```

```
##                                4                                2440
##               TVR Engineering Ltd                              Vector
##                                4                                   4
##                           Vinfast                  Vixen Motor Company
##                                4                                   1
##       Volga Associated Automobile                          Volkswagen
##                                1                                1313
##                              Volvo                               VPG
##                              914                                   5
##             Wallace Environmental                               Yugo
##                               32                                   8
```

```
table(df$trany)
```

```
##
##                  Automatic (A1)                  Automatic (A2)
##                             730                              64
##               Automatic (AM-S10)               Automatic (AM-S6)
##                               2                             191
##                Automatic (AM-S7)               Automatic (AM-S8)
##                             763                             285
##                Automatic (AM-S9)                 Automatic (AM5)
##                               6                              14
##                  Automatic (AM6)                 Automatic (AM7)
##                             160                             318
##                  Automatic (AM8)               Automatic (AV-S1)
##                              69                              29
##               Automatic (AV-S10)               Automatic (AV-S6)
##                              53                             324
##                Automatic (AV-S7)               Automatic (AV-S8)
##                             214                             183
##                   Automatic (L3)                  Automatic (L4)
##                               2                               2
##                  Automatic (S10)                  Automatic (S4)
##                             588                             233
##                   Automatic (S5)                  Automatic (S6)
##                             848                            3337
##                   Automatic (S7)                  Automatic (S8)
##                             386                            3157
##                   Automatic (S9) Automatic (variable gear ratios)
##                             250                            1104
##                Automatic 10-spd                 Automatic 3-spd
##                             380                            3151
##                 Automatic 4-spd                 Automatic 5-spd
##                           11048                            2203
##                 Automatic 6-spd                 Automatic 7-spd
##                            1752                             720
##                 Automatic 8-spd                 Automatic 9-spd
##                             978                             745
##                    Manual 3-spd                    Manual 4-spd
##                              77                            1483
##            Manual 4-spd Doubled                    Manual 5-spd
##                              17                            8391
##                    Manual 6-spd                    Manual 7-spd
```

```
##                             3079                              176
```

```
table(df$fuelType)
```

```
##
##                         CNG                      Diesel
##                          60                        1274
##                 Electricity             Gasoline or E85
##                         766                        1398
##     Gasoline or natural gas         Gasoline or propane
##                          20                           8
##                    Hydrogen                    Midgrade
##                          32                         164
##                     Premium     Premium and Electricity
##                       14441                         177
##  Premium Gas or Electricity              Premium or E85
##                          55                         128
##                     Regular Regular Gas and Electricity
##                       28890                         106
##  Regular Gas or Electricity
##                           4
```

```
# Average city fuel efficiency
avg_UCity <- mean(df$UCity)
cat("Average City Fuel Efficiency:", avg_UCity, "mpg\n")
```

```
## Average City Fuel Efficiency: 25.07361 mpg
```

```
# Average highway fuel efficiency
avg_UHighway <- mean(df$UHighway)
cat("Average Highway Fuel Efficiency:", avg_UHighway, "mpg\n")
```

```
## Average Highway Fuel Efficiency: 35.99745 mpg
```

```
unique(df$cylinders)
```

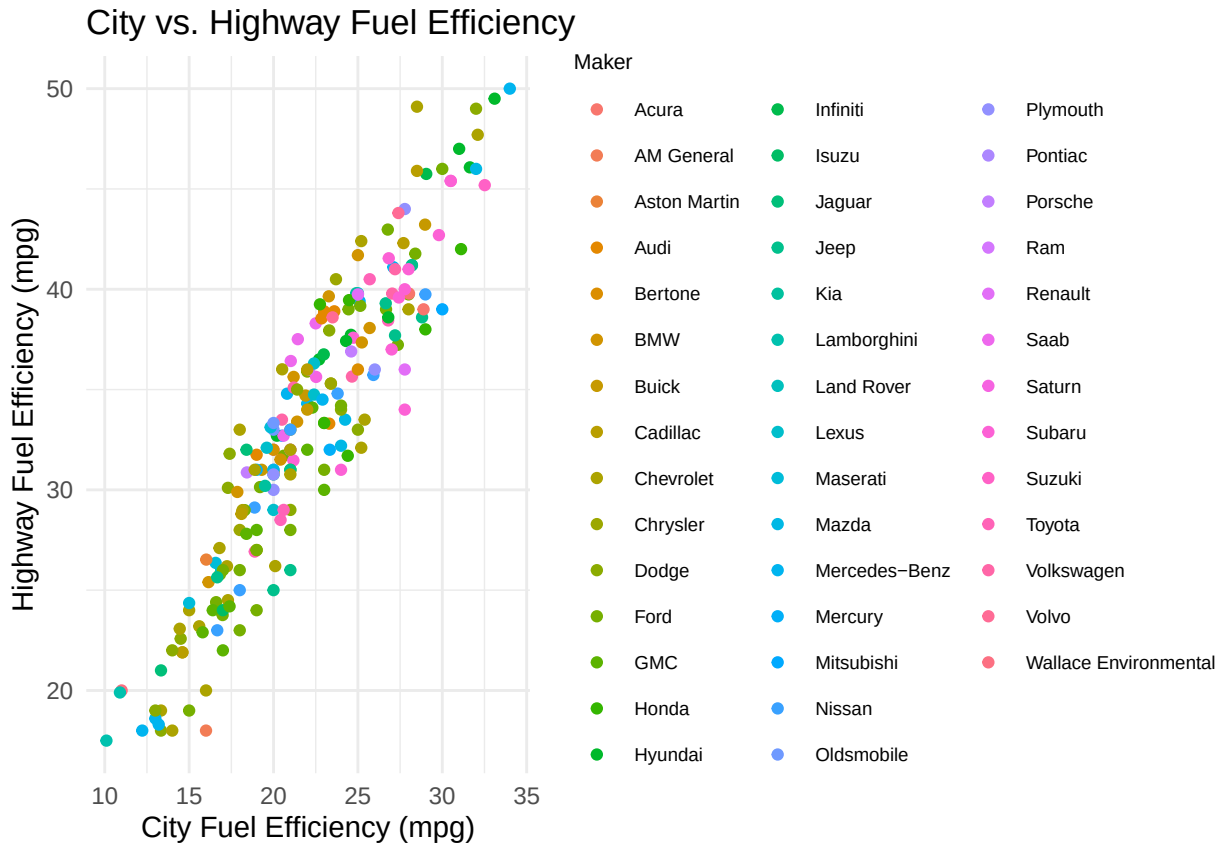
```
##  [1]  4 12  8  6  5 10  2  3 NA 16
```

```
sort(unique(df$year))
```

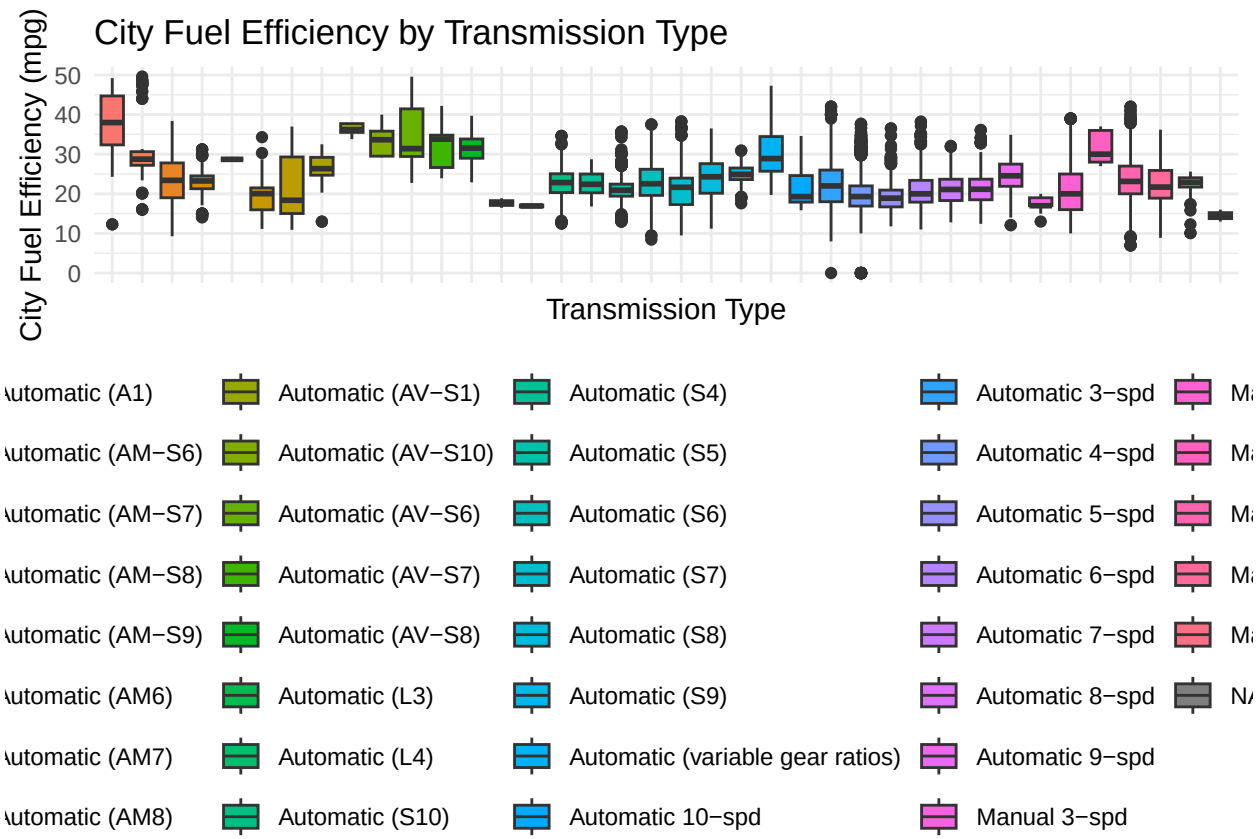
```
##  [1] 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998
## [16] 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
## [31] 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
```

```
# Filtering data , excluding Electric cars
df_filtered <- df %>% filter(fuelType != "Electricity",
                             fuelType != "Premium and Electricity",
                             year != 2025,
                             UCity <= 50,
                             UHighway <= 50)
```

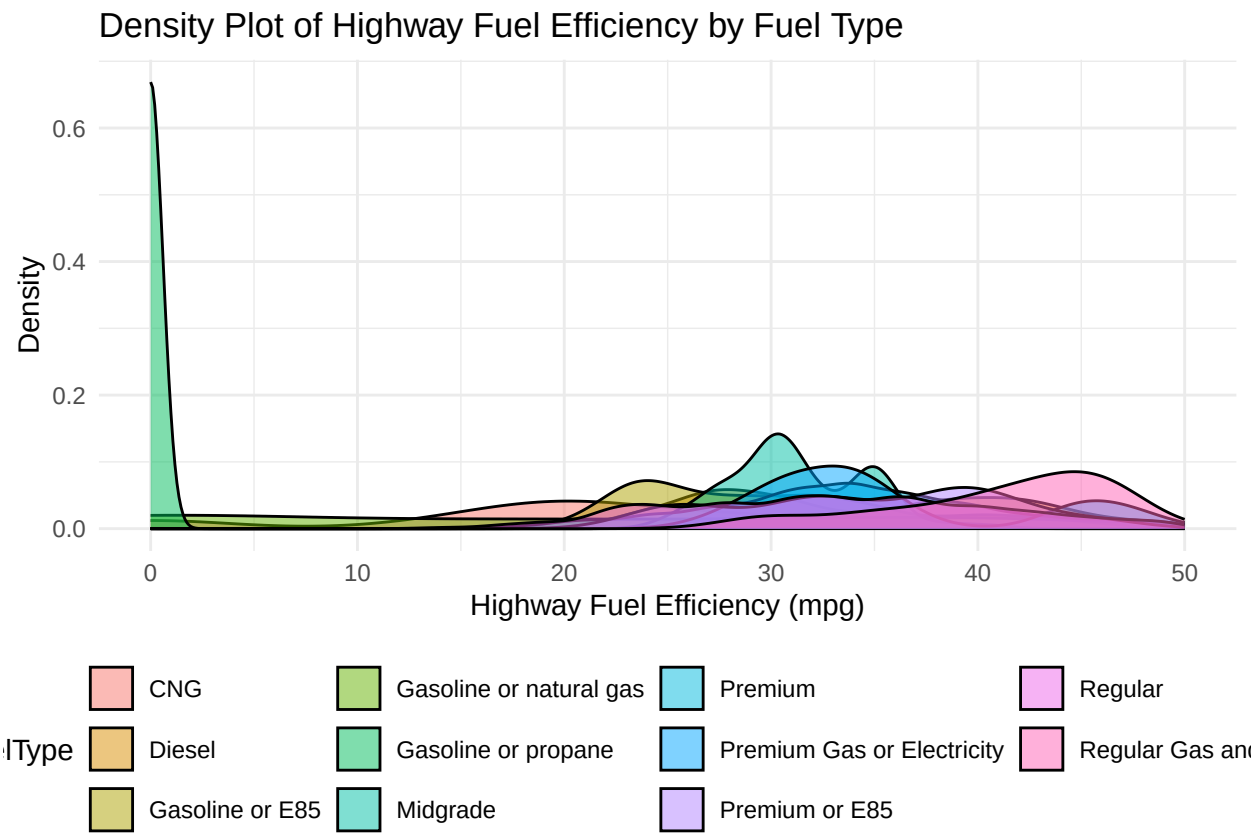
```r
df_sample <- df_filtered %>% sample_frac(0.005)



# Scatter plot of city vs. highway fuel efficiency
ggplot(df_sample, aes(x = UCity, y = UHighway, color = make)) +
  geom_point() +
  labs(title = "City vs. Highway Fuel Efficiency",
       x = "City Fuel Efficiency (mpg)",
       y = "Highway Fuel Efficiency (mpg)",
       color = "Maker") +
    theme_minimal()+
    theme(legend.position = "right",
          legend.text = element_text(size = 7),
          legend.title = element_text(size = 8))
```



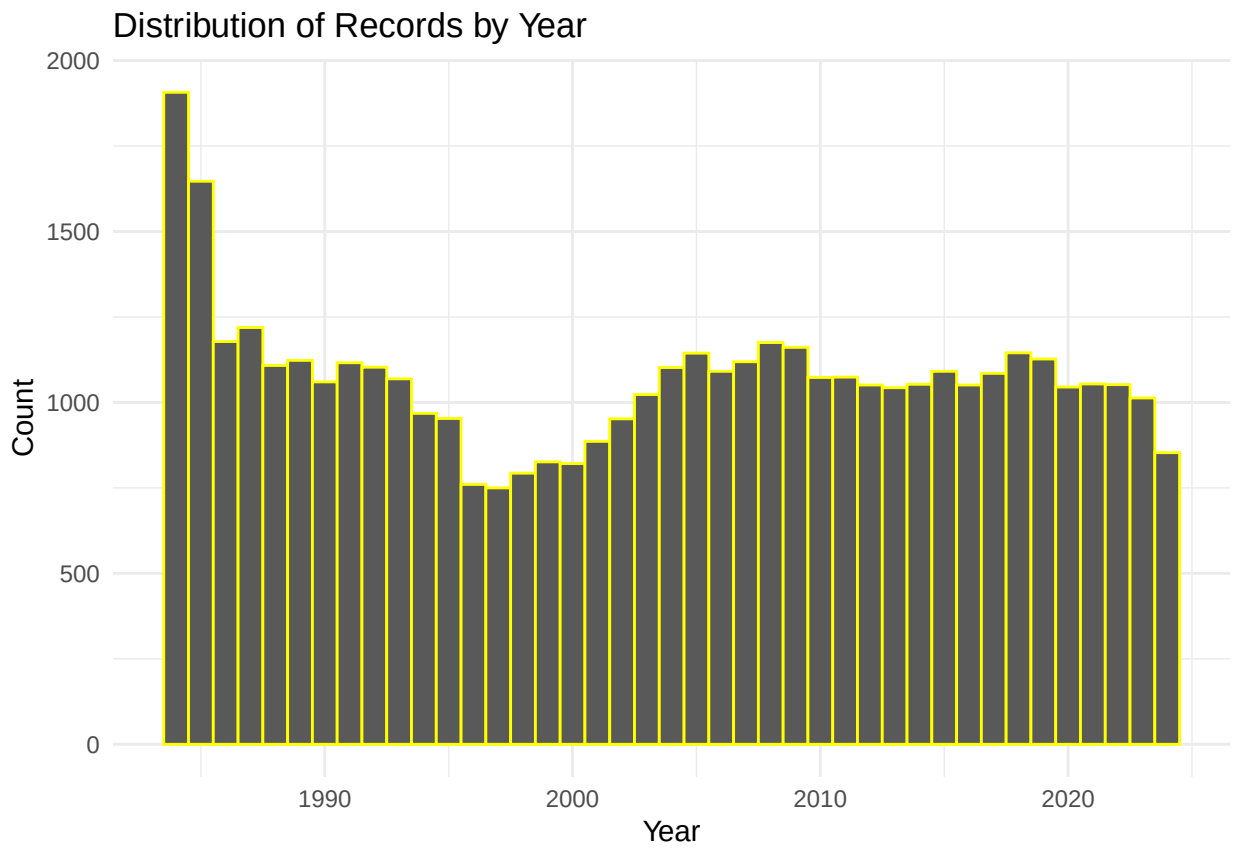

```r
# Box plot of city fuel efficiency by transmission type
ggplot(df_filtered, aes(x = trany, y = UCity, fill = trany)) +
  geom_boxplot() +
  labs(title = "City Fuel Efficiency by Transmission Type",
       x = "Transmission Type",
       y = "City Fuel Efficiency (mpg)") +
  theme_minimal()+
  theme(legend.position = "bottom",
        axis.text.x = element_blank())
```



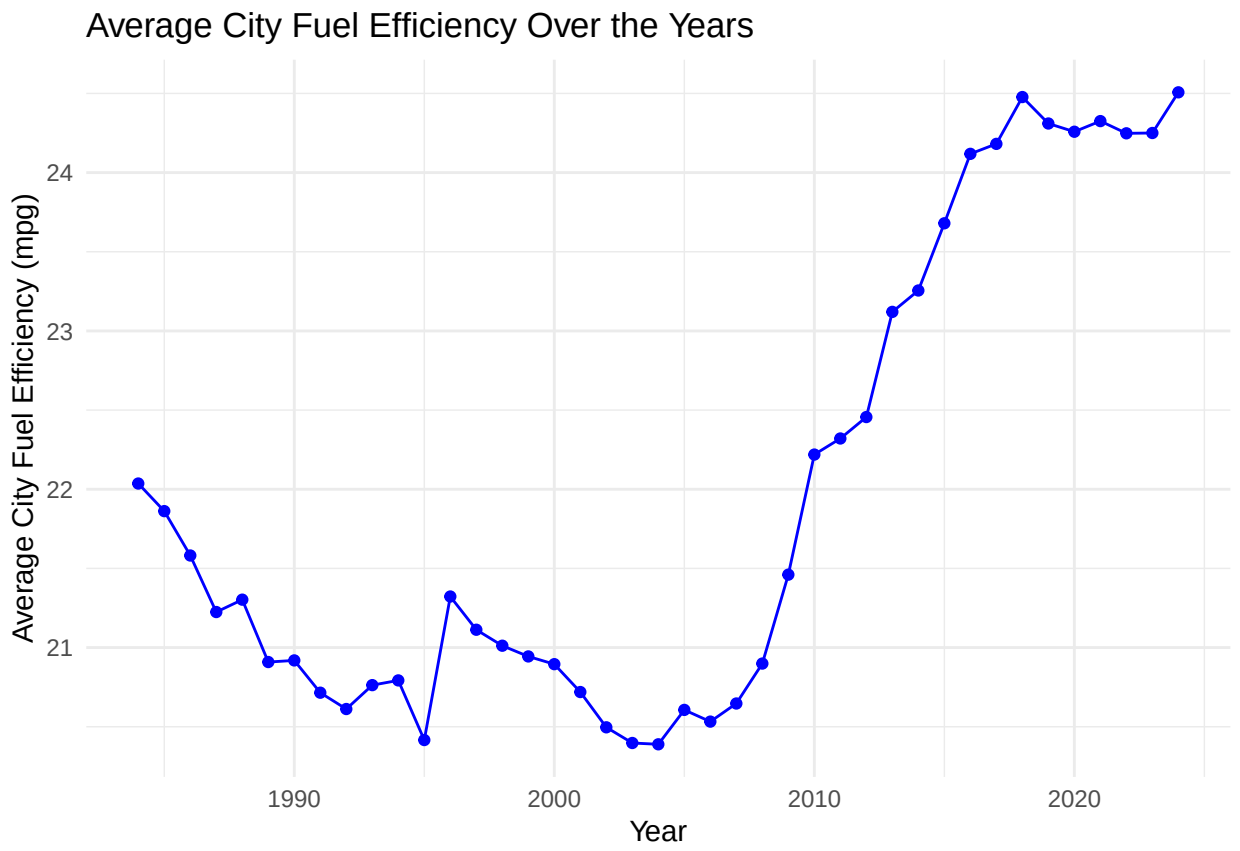

```
# Density plot of highway fuel efficiency by fuel type using the sampled data
ggplot(df_filtered, aes(x = UHighway, fill = fuelType)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Highway Fuel Efficiency by Fuel Type",
       x = "Highway Fuel Efficiency (mpg)",
       y = "Density") +
  theme_minimal() +
  theme(legend.position = "bottom")
```



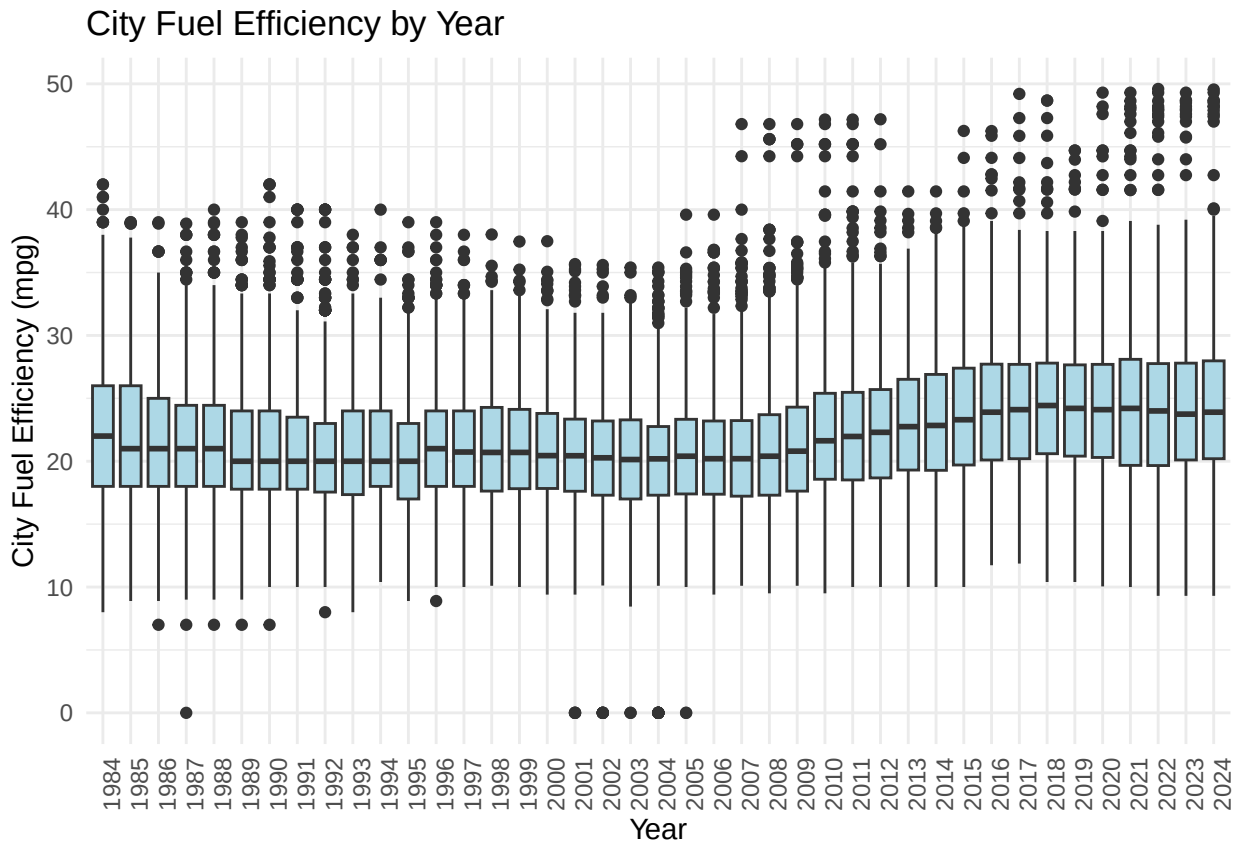

```r
# Histogram of the year column
ggplot(df_filtered, aes(x = year)) +
  geom_histogram(binwidth = 1, color="yellow") +
  labs(title = "Distribution of Records by Year",
       x = "Year",
       y = "Count") +
  theme_minimal()
```



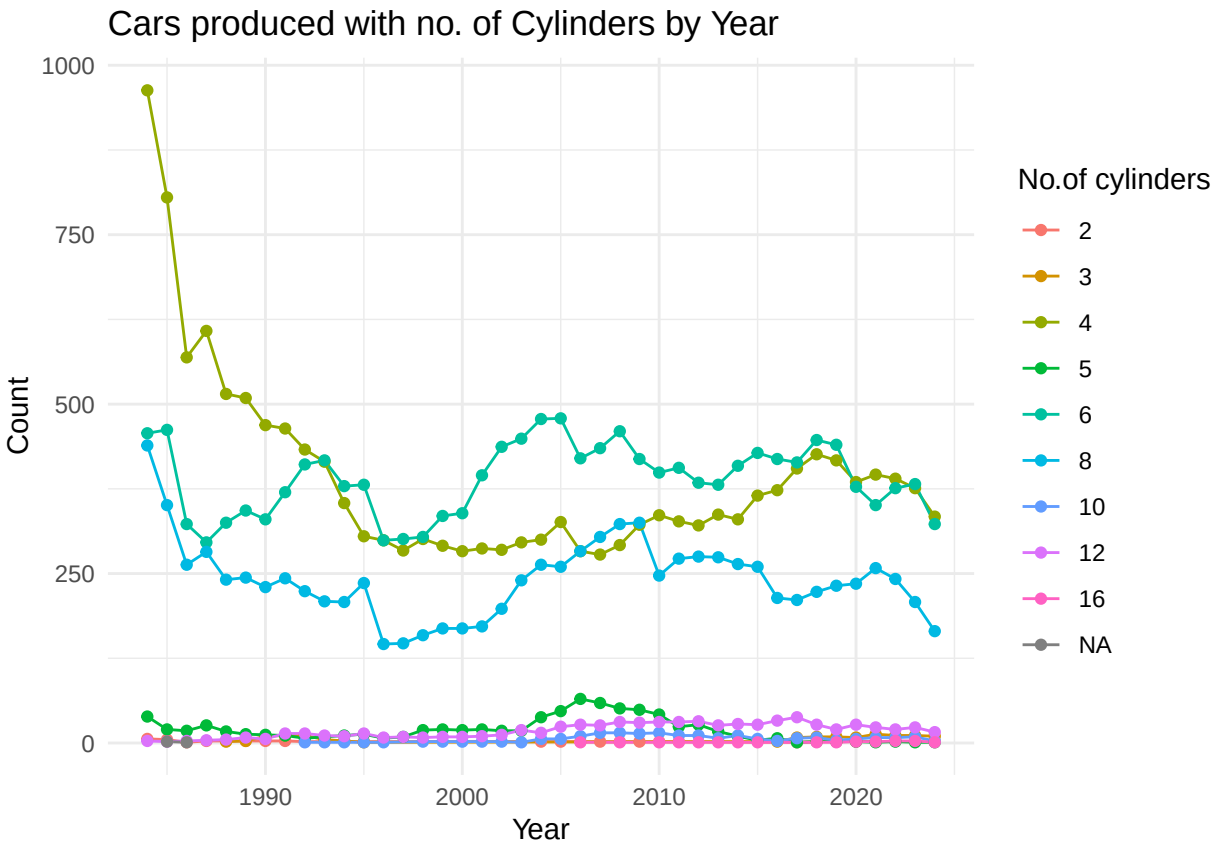

```r
# Aggregate data by year to get the average city fuel efficiency
agg_yearly <- df_filtered %>% group_by(year) %>% summarise(avg_UCity = mean(UCity, na.rm = TRUE))

# Line plot of average city fuel efficiency over the years
ggplot(agg_yearly, aes(x = year, y = avg_UCity)) +
  geom_line(color = "blue") +
  geom_point(color = "blue") +
  labs(title = "Average City Fuel Efficiency Over the Years",
       x = "Year",
       y = "Average City Fuel Efficiency (mpg)") +
  theme_minimal()
```




```r
# Box plot of city fuel efficiency by year
ggplot(df_filtered, aes(x = as.factor(year), y = UCity)) +
  geom_boxplot(fill = "lightblue") +
  labs(title = "City Fuel Efficiency by Year",
       x = "Year",
       y = "City Fuel Efficiency (mpg)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))  # Rotate x-axis text for better readability
```




```r
#  Group by cylinders and then year, and count the number of records
cylinder_year_counts <- df_filtered %>% group_by(cylinders,year) %>% reframe(count = n())

# Line plot of counts by year for each number of cylinders
ggplot(cylinder_year_counts, aes(x = year, y = count, color = as.factor(cylinders),
                                 group = as.factor(cylinders))) +
  geom_line() +
  geom_point() +
  labs(title = "Cars produced with no. of Cylinders by Year",
       x = "Year",
       y = "Count",
       color = "No.of cylinders") +
  theme_minimal()
```


Cars produced with no. of Cylinders by Year
Count
1000
750
500
250
0
1990
2000
2010
2020
Year
No.of cylinders
2
3
4
5
6
8
10
12
16
NA