

Mini Project : Data Vizualization

Phumpanlop Klinyoo

2024-01-08

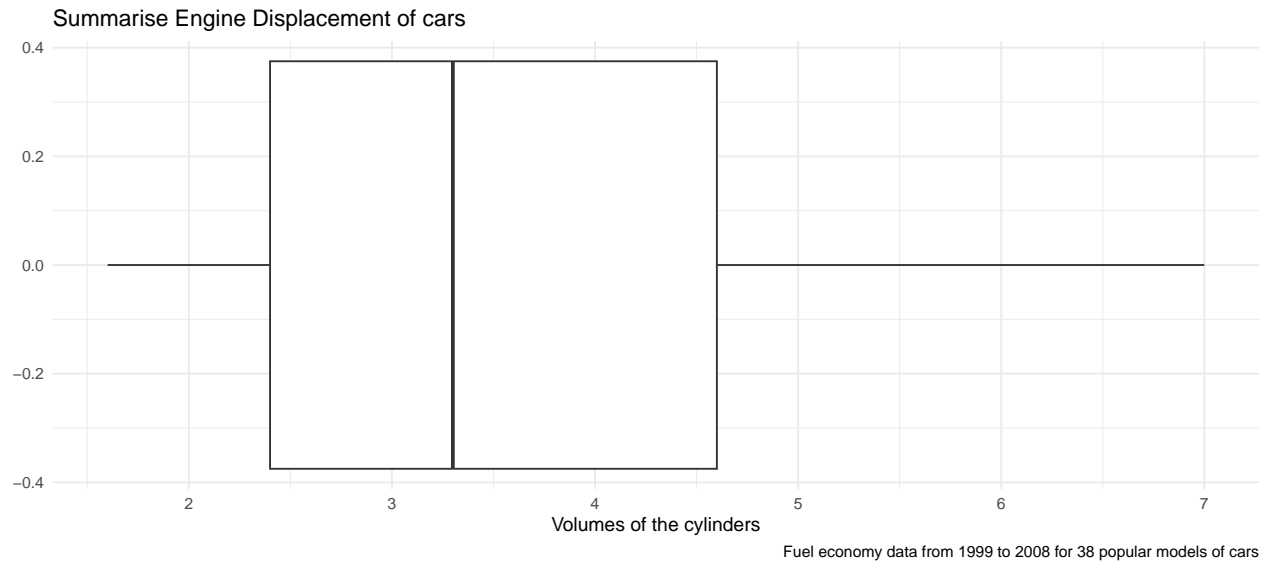
```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv   cty   hwy fl   class
##   <chr>         <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)  f      18    29 p   compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f      21    29 p   compa~
## 3 audi         a4      2    2008     4 manual(m6) f      20    31 p   compa~
## 4 audi         a4      2    2008     4 auto(av)   f      21    30 p   compa~
## 5 audi         a4      2.8  1999     6 auto(l5)  f      16    26 p   compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f      18    26 p   compa~
```

MPG => A data frame with 234 rows and 11 variables:

Variables	Description
manufacturer	manufacturer name
model	model name
displ	engine displacement, in litres
year	year of manufacture
cyl	number of cylinders
trans	type of transmission
drv	the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd
cty	city miles per gallon
hwy	highway miles per gallon
fl	fuel type
class	type of car

1. Engine Displacement(litres)

```
ggplot(mpg , aes(x=displ)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title="Summarise Engine Displacement of cars",
       x = "Volumes of the cylinders",
       caption="Fuel economy data from 1999 to 2008 for 38 popular models of cars")
```



```
mpg %>%
  select(manufacturer, model, displ) %>%
  reframe(min_displ = min(displ),
          q13_displ = quantile(displ, c(0.25, 0.75)), prob = c(0.25, 0.75),
          med_displ = median(displ),
          max_displ = max(displ))
```

```
## # A tibble: 2 x 5
##   min_displ q13_displ  prob med_displ max_displ
##   <dbl>    <dbl> <dbl>    <dbl>    <dbl>
## 1      1.6      2.4  0.25      3.3      7
## 2      1.6      4.6  0.75      3.3      7
```

```
mpg %>%
  select(manufacturer, model, displ) %>%
  filter(displ == c(1.6, 7.0))
```

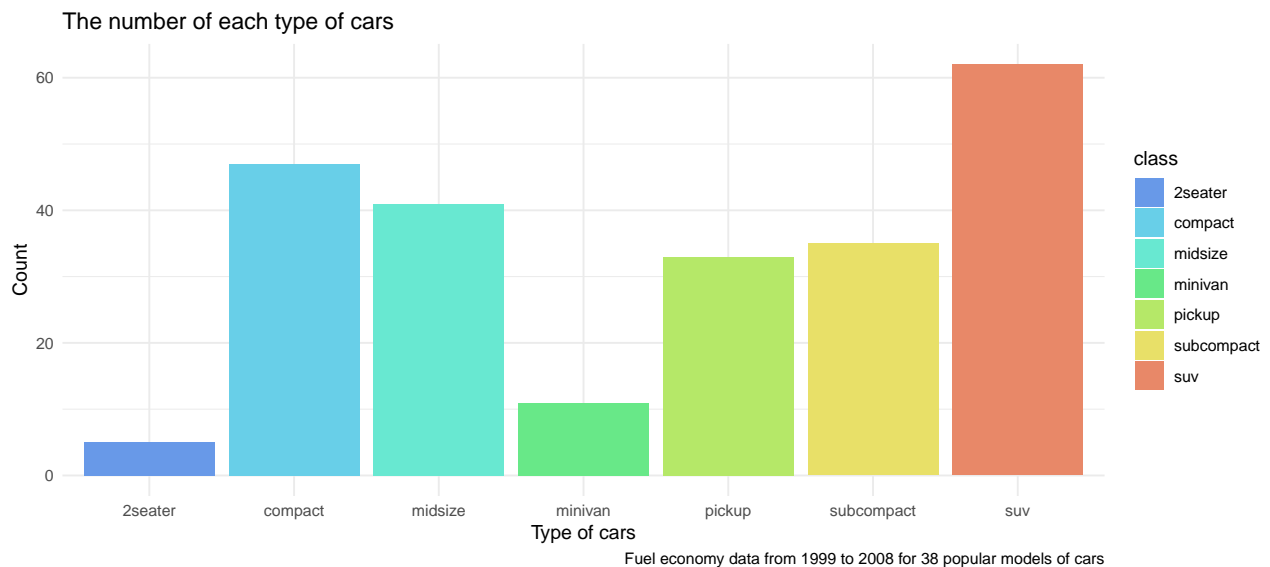
```
## # A tibble: 3 x 3
##   manufacturer model    displ
##   <chr>         <chr>    <dbl>
## 1 chevrolet    corvette    7
## 2 honda        civic      1.6
## 3 honda        civic      1.6
```

From the boxplot, it was found that the minimum, median, and maximum volumes of the cylinders inside the engine were 1.6 liters (Honda Civic model), 3.3 liters and 7.0 liters (Chevrolet Corvette model), respectively.

2. The number of each type of cars

```
ggplot(mpg, aes(class, fill=class)) +
  geom_bar() +
  theme_minimal() +
  scale_fill_manual(values = c(
    "#6899e8", "#68cfe8", "#68e8d1", "#68e888", "#b5e868", "#e8e068", "#e88868")) +
  labs(title="The number of each type of cars",
```

```
x = "Type of cars",
y = "Count",
caption="Fuel economy data from 1999 to 2008 for 38 popular models of cars")
```



```
mpg %>%
  count(class) %>%
  arrange(desc(n))
```

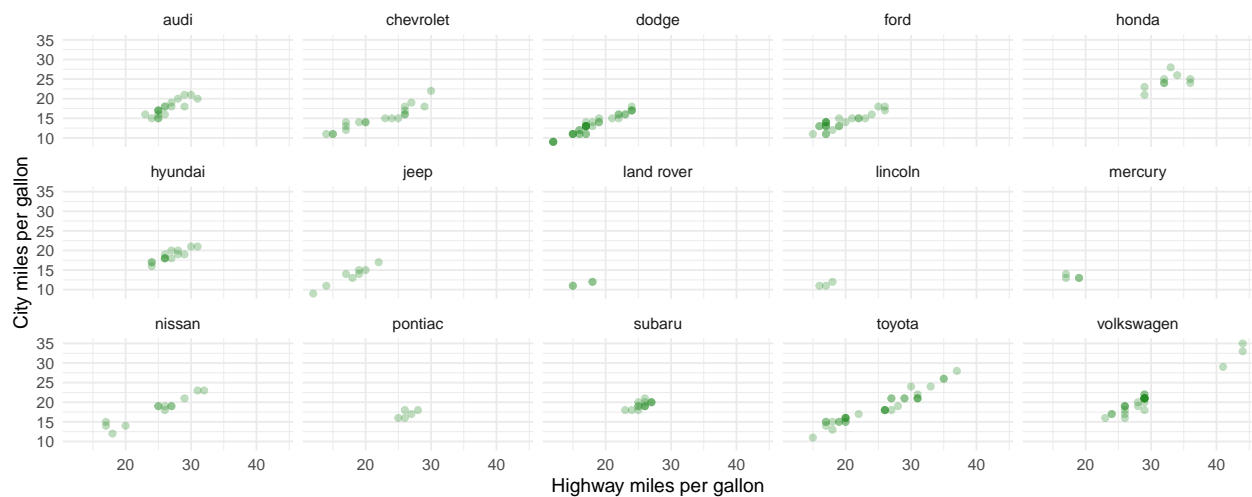
```
## # A tibble: 7 x 2
##   class      n
##   <chr>    <int>
## 1 suv      62
## 2 compact  47
## 3 midsize  41
## 4 subcompact 35
## 5 pickup   33
## 6 minivan  11
## 7 2seater   5
```

from bar chart It was found that between 1999 and 2008, in the top 3 categories, there were suv, compact and midsize, respectively.

3. City miles and highway miles of each cars manufacturer

```
ggplot(mpg, aes(hwy, cty)) +
  geom_point(alpha = 0.3, color="forestgreen") +
  theme_minimal() +
  facet_wrap(~ manufacturer, ncol = 5) +
  labs(title="City miles and highway miles of each cars manufacturer ",
       x = "Highway miles per gallon",
       y = "City miles per gallon",
       caption="Fuel economy data from 1999 to 2008 for 38 popular models of cars")
```

City miles and highway miles of each cars manufacturer



Fuel economy data from 1999 to 2008 for 38 popular models of cars

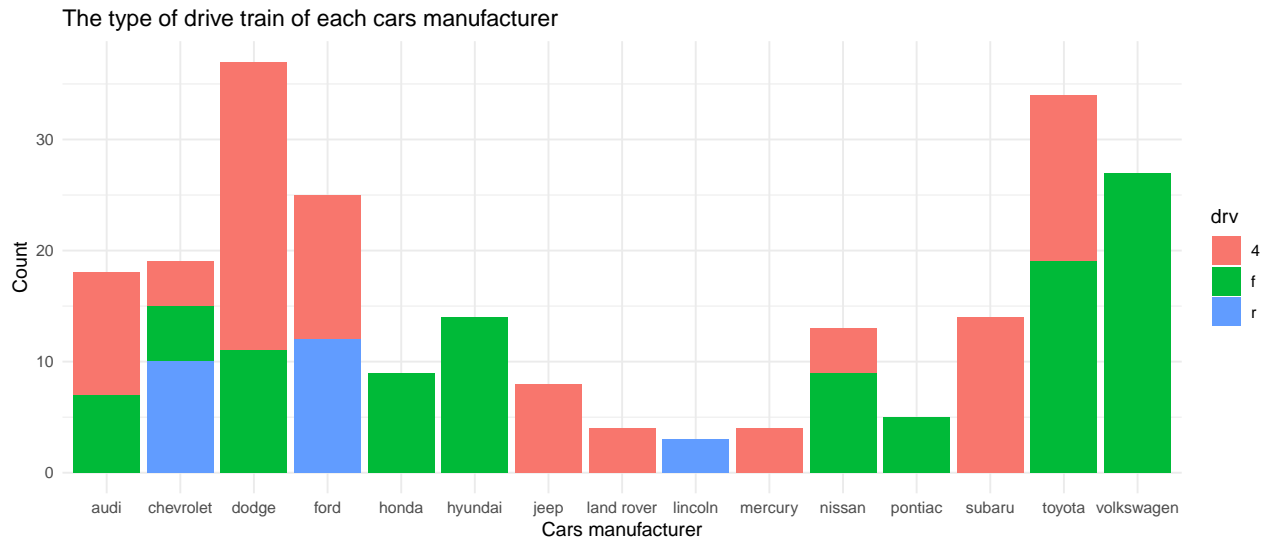
```
mpg %>%
  select(manufacturer, cty, hwy) %>%
  group_by(manufacturer) %>%
  summarise(mean_cty = mean(cty),
            mean_hwy = mean(hwy),
            Count = n()) %>%
  arrange(desc(mean_cty)) %>%
  head(5)
```

```
## # A tibble: 5 x 4
##   manufacturer mean_cty mean_hwy Count
##   <chr>         <dbl>   <dbl> <int>
## 1 honda         24.4     32.6     9
## 2 volkswagen    20.9     29.2    27
## 3 subaru        19.3     25.6    14
## 4 hyundai       18.6     26.9    14
## 5 toyota        18.5     24.9    34
```

From the scatter plot and table above, the top 5 city and highway miles for each cars manufacturer are honda, volkswagen, subaru, hyundai and toyota, respectively.

4. The type of drive train of each cars manufacturer

```
ggplot(mpg, aes(manufacturer, fill=drv)) +
  geom_bar(position="stack") +
  theme_minimal() +
  labs(title="The type of drive train of each cars manufacturer",
       x = "Cars manufacturer",
       y = "Count",
       caption="Fuel economy data from 1999 to 2008 for 38 popular models of cars")
```



Fuel economy data from 1999 to 2008 for 38 popular models of cars

```
mpg %>%
  select(manufacturer, drv) %>%
  filter(drv == "4") %>%
  group_by(manufacturer) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  head(1)
```

```
## # A tibble: 1 x 2
##   manufacturer      n
##   <chr>          <int>
## 1 dodge          26
```

```
mpg %>%
  select(manufacturer, drv) %>%
  filter(drv == "f") %>%
  group_by(manufacturer) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  head(1)
```

```
## # A tibble: 1 x 2
##   manufacturer      n
##   <chr>          <int>
## 1 volkswagen      27
```

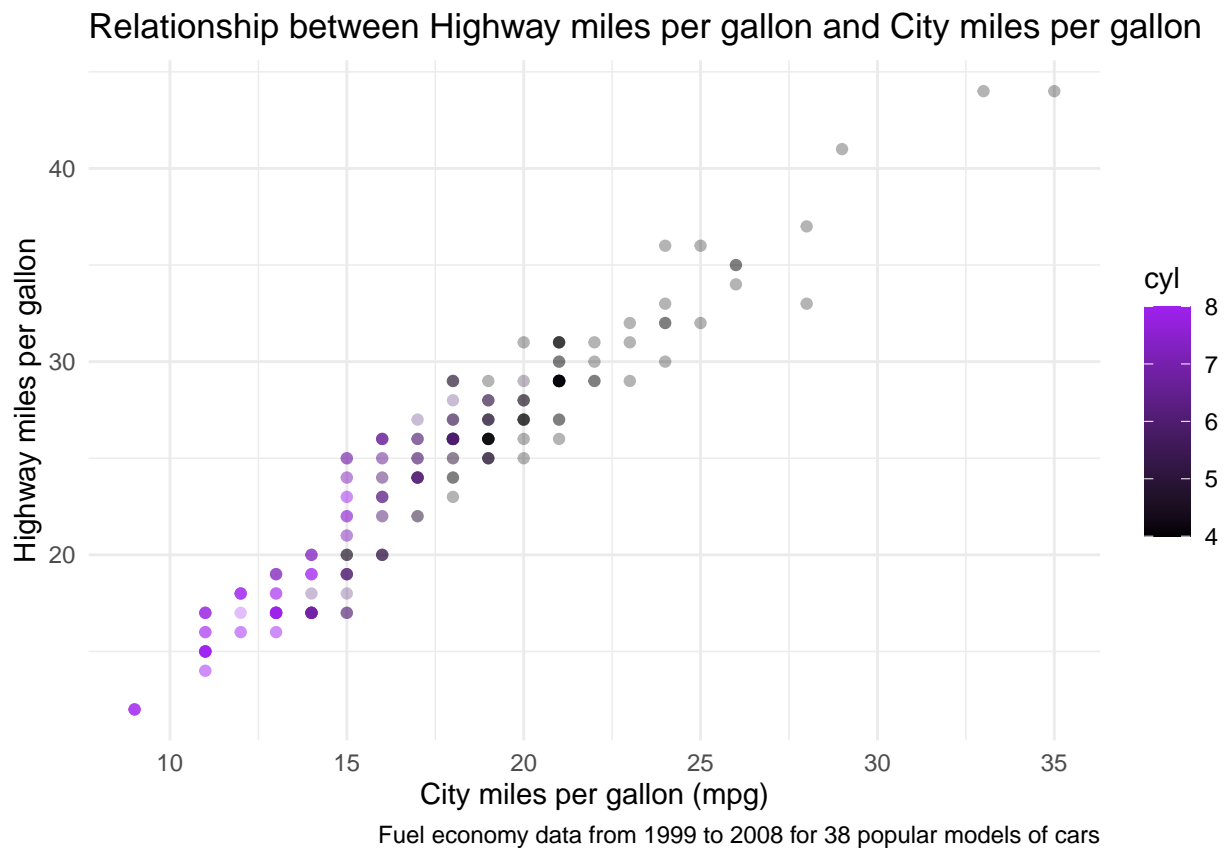
```
mpg %>%
  select(manufacturer, drv) %>%
  filter(drv == "r") %>%
  group_by(manufacturer) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  head(1)
```

```
## # A tibble: 1 x 2
##   manufacturer      n
##   <chr>          <int>
## 1 ford           12
```

From the bar graph, it can be found that the most popular types of drive train of cars manufacturers are Dodge as 4 wheel drive, Volkswagen as front-wheel drive, Ford as rear-wheel drive.

5. Efficiency in saving fuel

```
ggplot(mpg, aes(x=cty, y=hwy, color=cyl)) +
  geom_point(alpha=0.3) +
  theme_minimal() +
  scale_color_gradient(low = "black", high = "purple") +
  labs(title="Relationship between Highway miles per gallon and City miles per gallon",
       x = "City miles per gallon (mpg)",
       y = "Highway miles per gallon",
       caption="Fuel economy data from 1999 to 2008 for 38 popular models of cars")
```



From scatter chart it is found that cars with a lot of cylinders Efficient in using less oil or not saving fuel But cars with fewer cylinders will have better fuel economy.