Project - MPG Data Visualization

Phatchara Soroschokchai

2023-12-11

Data Visualization

Import Library

```
library("tidyverse") # for manipulate data (ggplot in this library)
library("ggthemes") # for setting theme
library("patchwork") # multiple plot in one chart
```

Explore the mpg data

```
## # A tibble: 6 x 12
##   manufacturer model displ  year cyl   trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <chr> <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~
## # i 1 more variable: trans_label <chr>
```

Data Dictionary

Variable	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

Variable	Description
trans_label	group of trans

Top 10 manufacturer with the most average city and highway miles per galon.

```
manu_mean_cty <- mpg %>%
  group_by(manufacturer) %>%
  summarise(mean_cty = mean(cty)) %>%
  arrange(desc(mean_cty)) %>%
  head(10)

manu_mean_hwy <- mpg %>%
  group_by(manufacturer) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(desc(mean_hwy)) %>%
  head(10)

bar1 <- ggplot(manu_mean_cty,
               mapping=aes(x=mean_cty, y=reorder(manufacturer, -mean_cty))) +
  geom_col(fill = "cadetblue") +
  scale_x_continuous(limits = c(0, 35),
                     oob = scales::squish) +
  labs(title = "Top 10 Manufacturer with the most average city miles per galon",
       x = "Manufacturer",
       y = "Avereage of City Miles per Galon",
       caption = "source: mpg dataset in R") +
  theme_minimal()

bar2 <- ggplot(manu_mean_hwy,
               mapping=aes(x=mean_hwy, y=reorder(manufacturer, -mean_hwy))) +
  geom_col(fill = "seagreen4") +
  scale_x_continuous(limits = c(0, 35),
                     oob = scales::squish) +
  labs(title = "Top 10 Manufacturer with the most average highway miles per galon",
       x = "Manufacturer",
       y = "Avereage of Highway Miles per Galon",
       caption = "source: mpg dataset in R") +
  theme_minimal()

bar1 / bar2
```

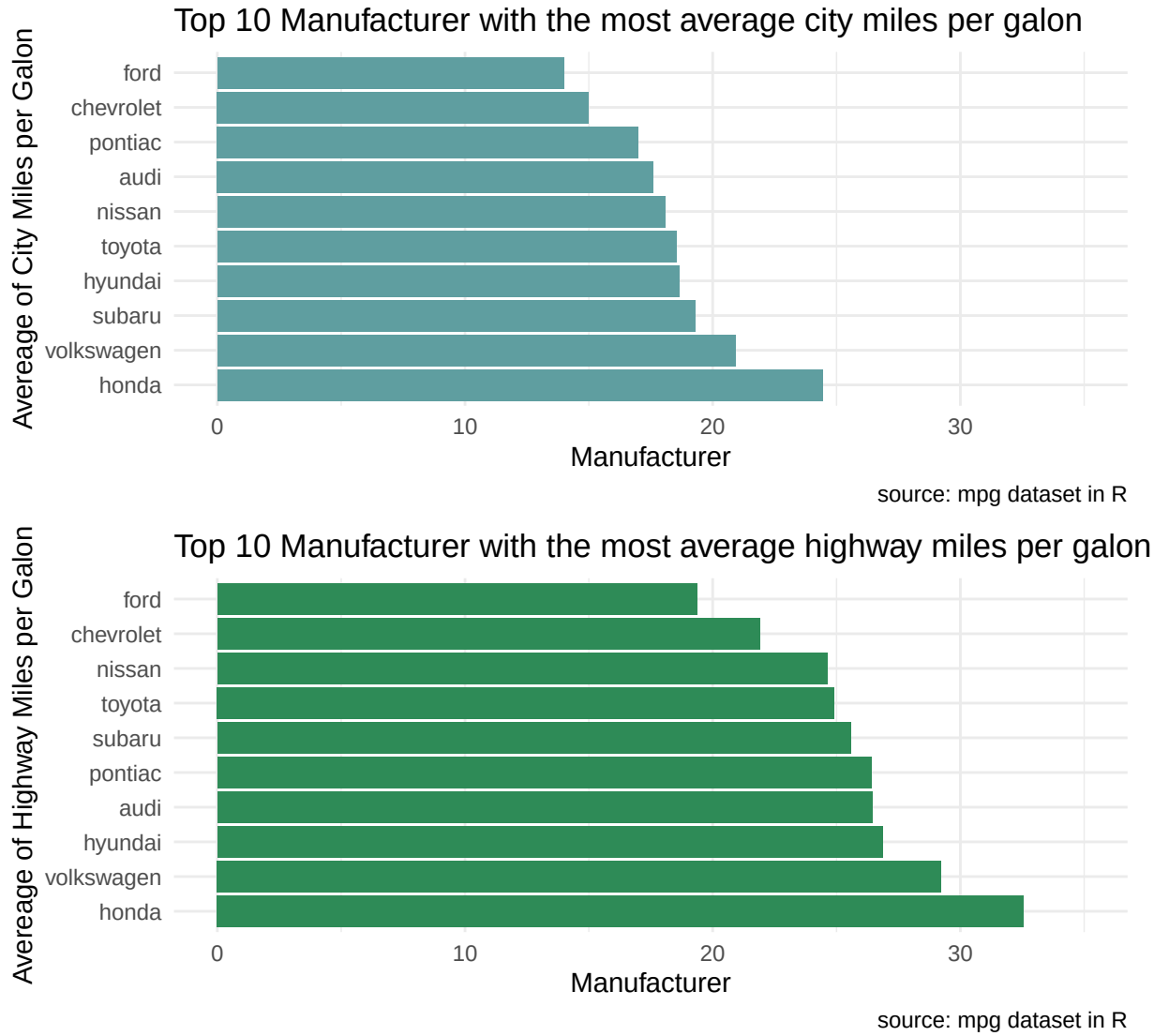

Fig. 1. You can see Honda and Volkswagen manufacturer produce the most fuel-efficient cars when measured in miles per gallon on average. If the car of Honda and Volkswagen manufacturer is driven on the highway and in the city, then Honda and Volkswagen are the best options.

Distribution of city & highway miles per galon by type of drive train

```
box1 <- ggplot(mpg,
               mapping=aes(x=cty, group=drv, col=drv)) +
  geom_boxplot() +
  labs(title = "Box plot of city miles per galon by type of drive train",
       x = "City Miles per Galon",
       y = "Ntype of drive train",
       caption = "source: mpg dataset in R") +
  theme_minimal()

box2 <- ggplot(mpg,
```

```
              mapping=aes(x=hwy, group=drv, col=drv)) +
  geom_boxplot() +
  labs(title = "Box plot of city miles per galon by type of drive train",
       x = "Highway Miles per Galon",
       y = "type of drive train",
       caption = "source: mpg dataset in R") +
  theme_minimal()

hist1 <- ggplot(mpg,
                mapping=aes(x=cty, fill=drv, col=drv)) +
  geom_histogram(bins=20, alpha=0.4, position = "identity") +
  labs(title = "Histogram of highway miles per galon by type of drive train",
       x = "City Miles per Galon",
       y = "Frequency",
       caption = "source: mpg dataset in R") +
       theme_minimal()

hist2 <- ggplot(mpg,
                mapping=aes(x=hwy, fill=drv, col=drv)) +
  geom_histogram(bins=20, alpha=0.4, position = "identity") +
  labs(title = "Histogram of highway miles per galon by type of drive train",
       x = "Highway Miles per Galon",
       y = "Frequency",
       caption = "source: mpg dataset in R") +
       theme_minimal()

(box1 + box2) / (hist1 + hist2)
```

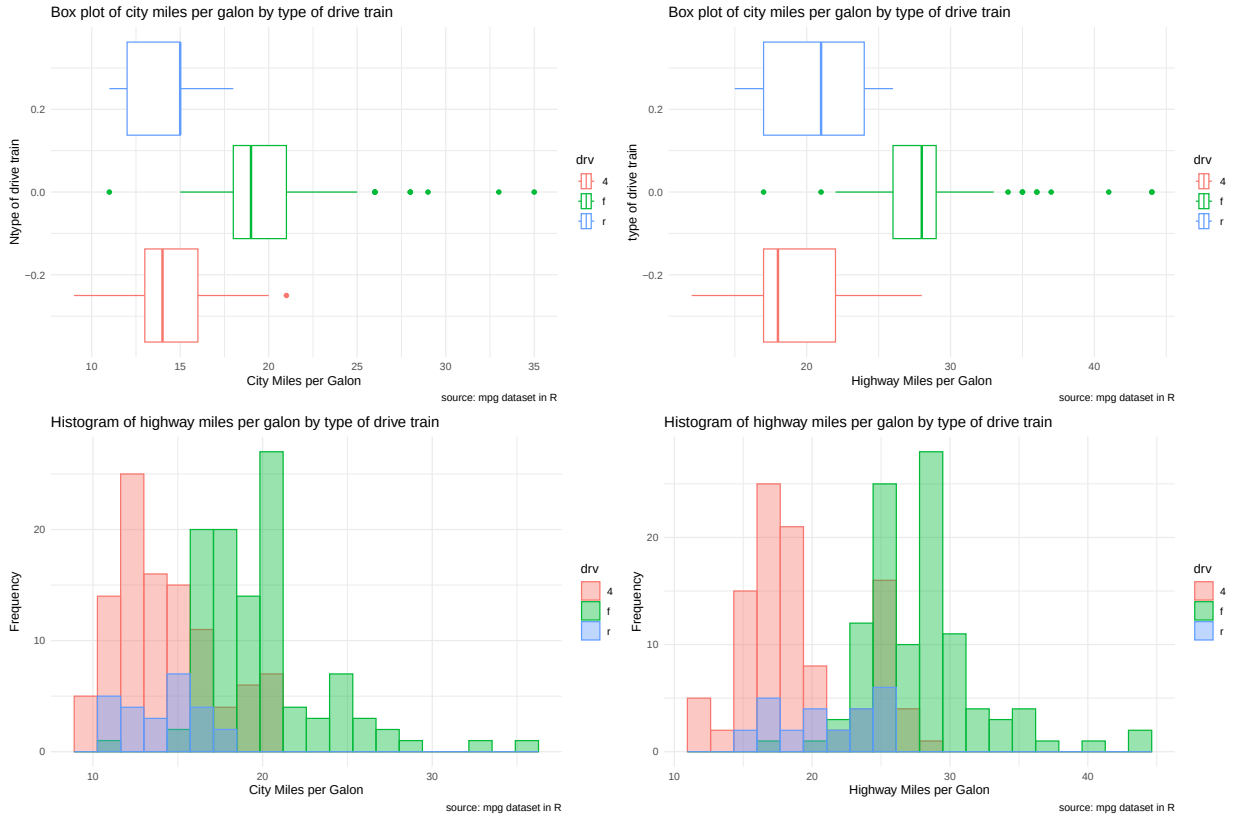

Fig. 2. This chart shows the distribution of miles per galon if a car is driven on a highway or in a city. In conclusion, front-wheel drive cars can get better miles per galon, showing that front-wheel drive cars can save on fuel more than 4-wheel and rear-wheel drive cars.

Comparison between City and Highway Miles per Galon by number of cylinders.

```
ggplot(mpg,
       mapping = aes(x=cty, y=hwy, col=cyl)) +
  geom_point(alpha=0.4, size=2) +
  geom_smooth(method="lm", se=FALSE) +
  labs(title = "Comparison between City and Highway Miles per Galon by number of cylinders",
       x = "City Miles per Galon",
       y = "Highway Miles per Galon",
       caption = "source: mpg dataset in R") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

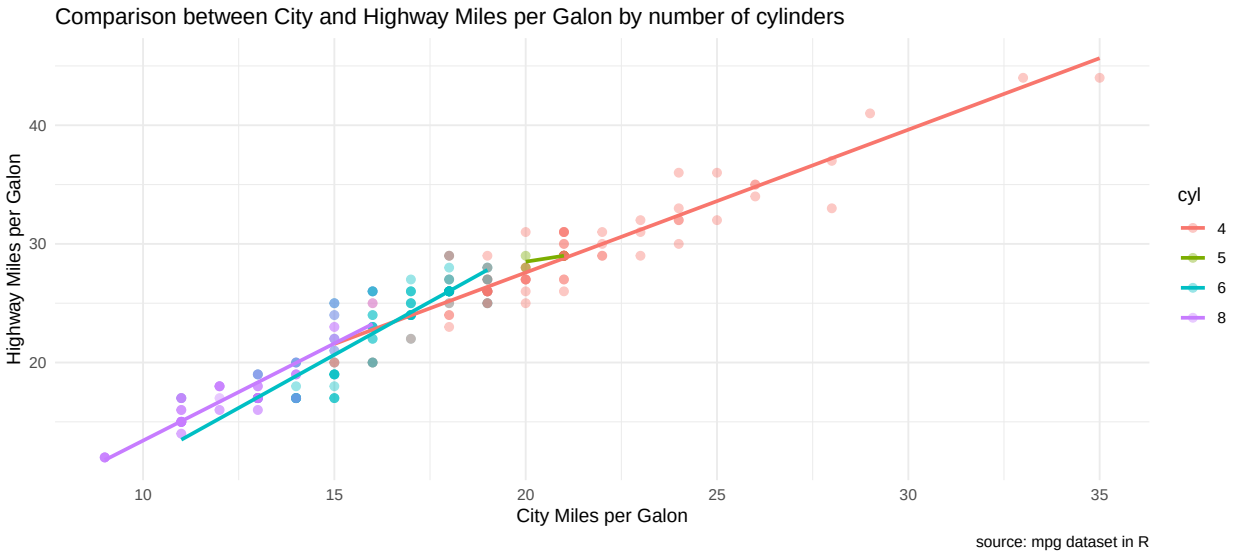

Fig. 3. This chart shows a comparison between City MPG and Highway MPG segments based on number of cylinders. It can be seen that the number of cylinders = 4 can make City MPG and Highway MPG mostly very valuable. And if the number of cylinders is = 8, City MPG and Highway MPG can mostly be low.

Fraction of Car types by Manufaturer

```
ggplot(mpg,
       aes(manufacturer, fill=class)) +
  geom_bar(position="fill") +
  scale_fill_brewer(palette = "PuBuGn") +
  labs(title = "Fraction of Car types by Manufaturer",
       x = "Manufacturer",
       y = "Fraction",
       caption = "source: mpg dataset in R") +
  theme_minimal()
```

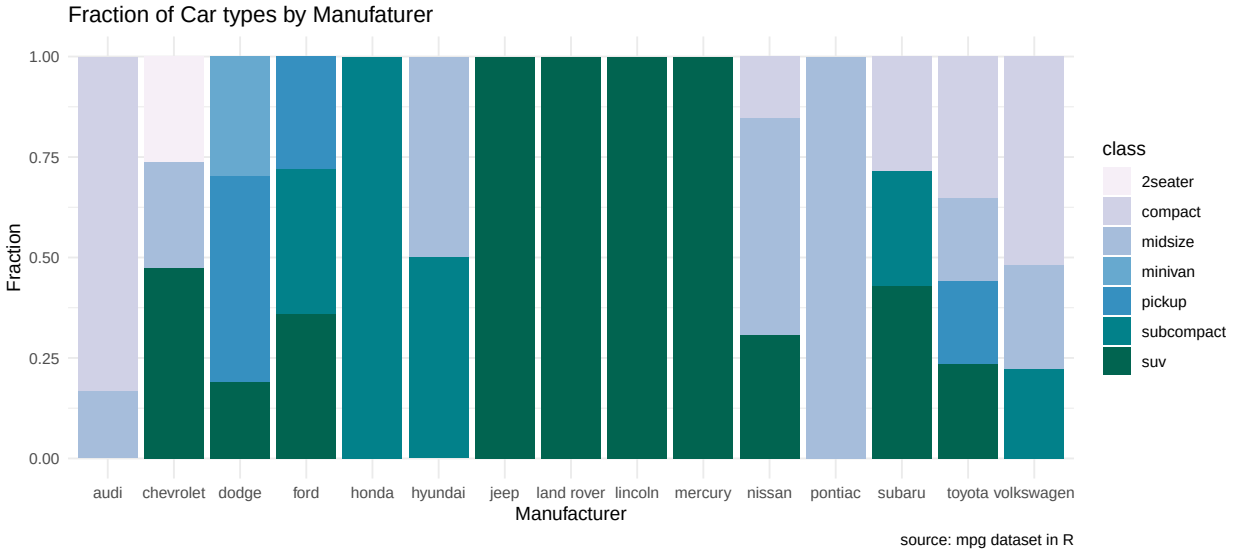

Fig. 4. From this graph, you can see that Jeep, Land Rover, Lincoln, and Mercury all produce SUVs cars. Honda only produces mini vans cars. Pontiac only produces mid size cars.

Box plot of Engine displacement by Transmission and fuel types

```
ggplot(mpg, aes(x = displ, group=trans_label, fill=trans_label)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Box plot of Engine displacement by Transmission and fuel types",
       x = "Engine displacement (litres)",
       y = "Transmission (Auto, Manual)") +
  facet_wrap(~ fl, nrow=5)
```

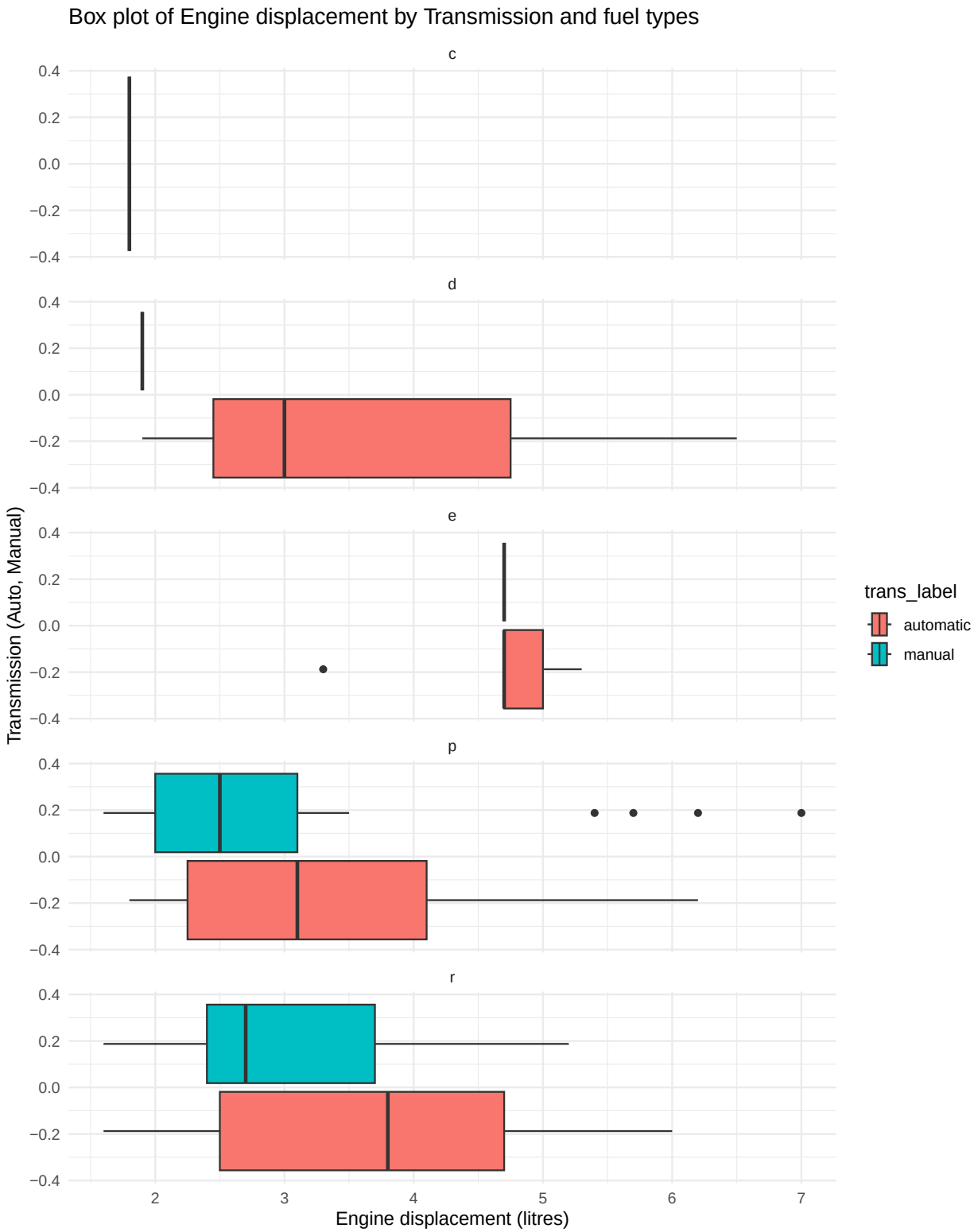

Fig. 5. Automatic transmission cars Always have a larger average cylinder volume (liters) than a car with a manual transmission.