Assignment 4

Anthony Cunningham

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE)

# Change working dir in RMarkdown cell
knitr::opts_knit$set(root.dir =
'C:/Users/AC069015/kumc_applied_stats/data_824_data_viz_and_acquisition/3_univariate_eda'
)

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
```
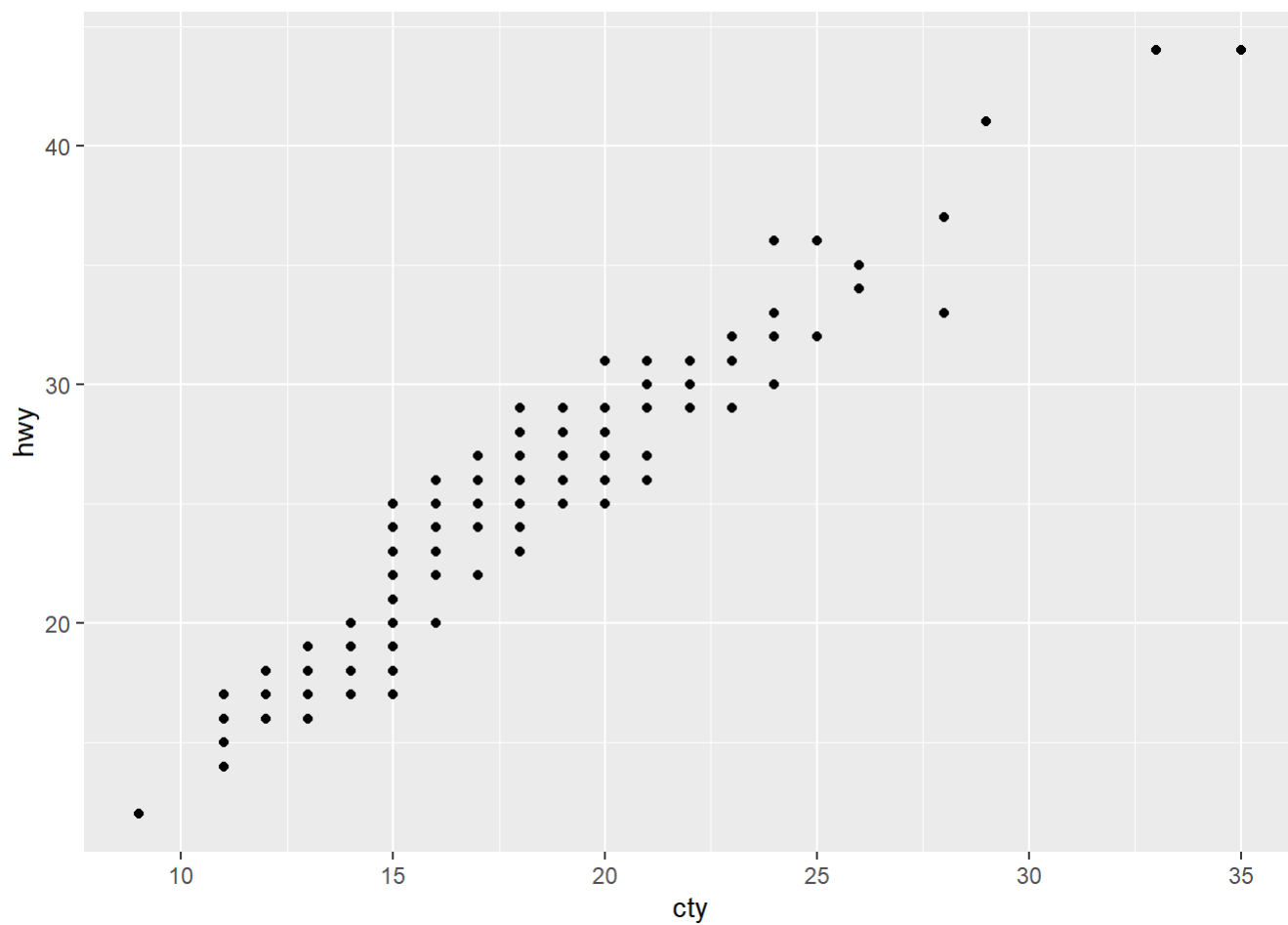
```
library(tidyr)
library(ggplot2)
library(patchwork)
library(hexbin)

df <- mpg
```

Exercise 1
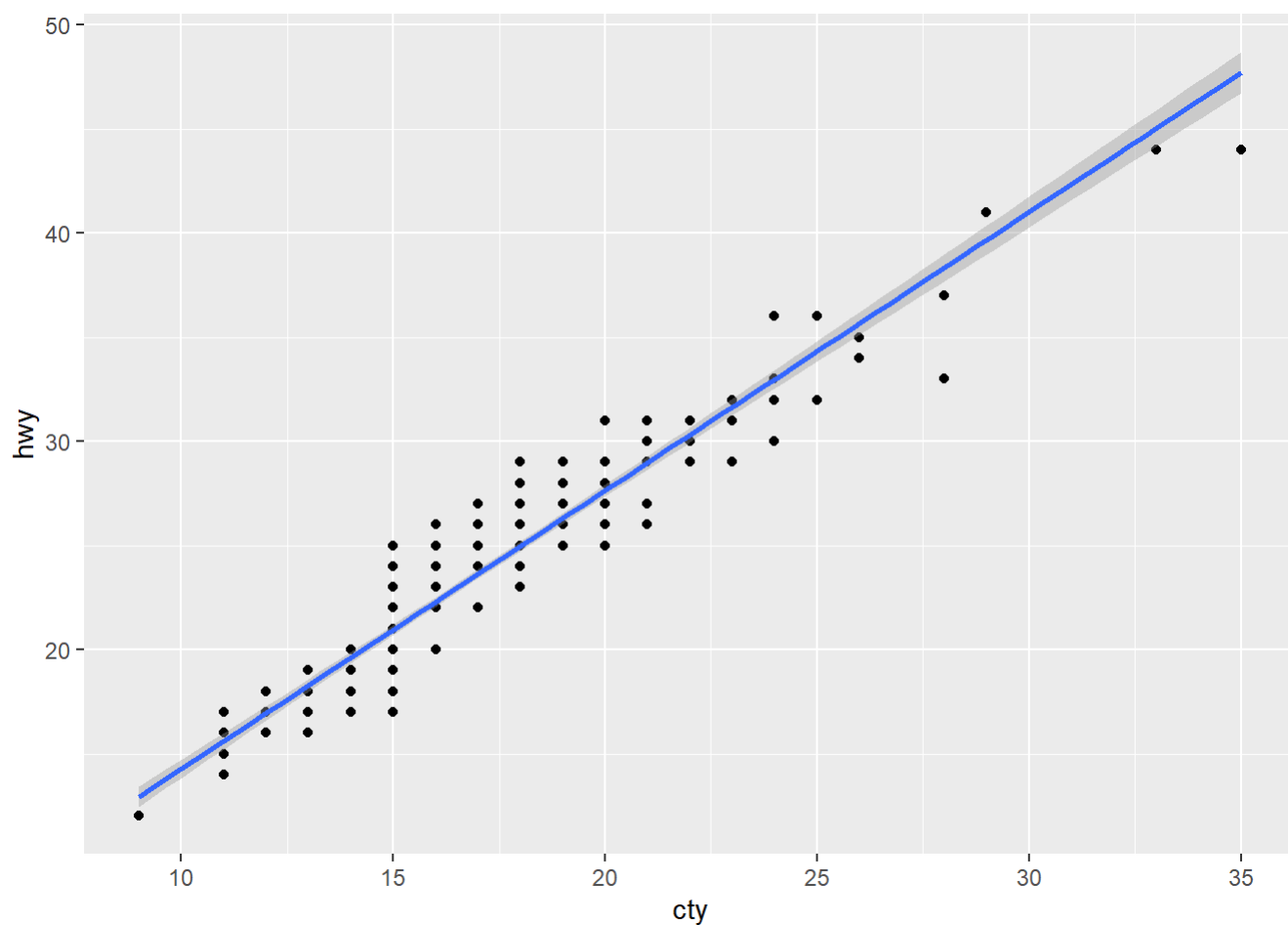
```
df[sapply(df, is.character)] <- lapply(df[sapply(df, is.character)], as.factor)
```

```
df %>%
  ggplot(aes(x = cty, y = hwy)) +
  geom_point()
```

```
df %>%
  ggplot(aes(x = cty, y = hwy)) +
  geom_point() +
  stat_smooth(method = "lm", formula = y ~ x, se = T, level = 0.95)
```

```
df %>%
  ggplot(aes(x = cty, y = hwy)) +
  geom_point(
      color = "black",
      fill = "white",
      alpha = 0.3,
      size = 1.5,
      stroke = 1.5,
      ) +
  stat_smooth(method = "lm", formula = y ~ x, se = T, level = 0.95) +
  xlab("City MPG") +
  ylab("Highway MPG") +
  ggtitle("Vehicle Miles per Gallon (MPG)") +
  theme(
      axis.text = element_text(size = 16),
      axis.title = element_text(size = 20),
      plot.title = element_text(size = 25, face = "bold"),
      panel.border = element_rect(color = "black", fill = NA, size = 1.5)
      )
```

Vehicle Miles per Gallon (MPG)
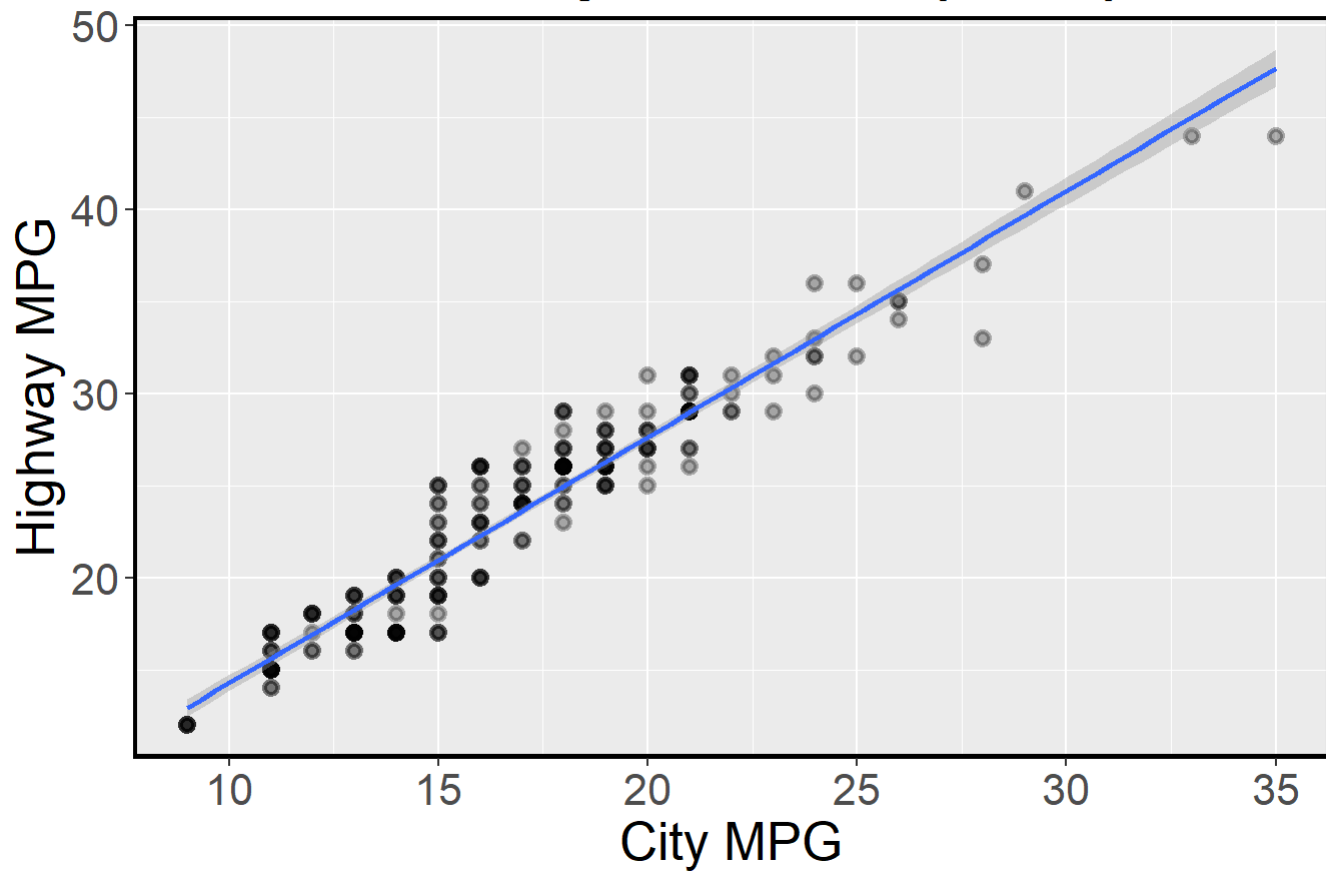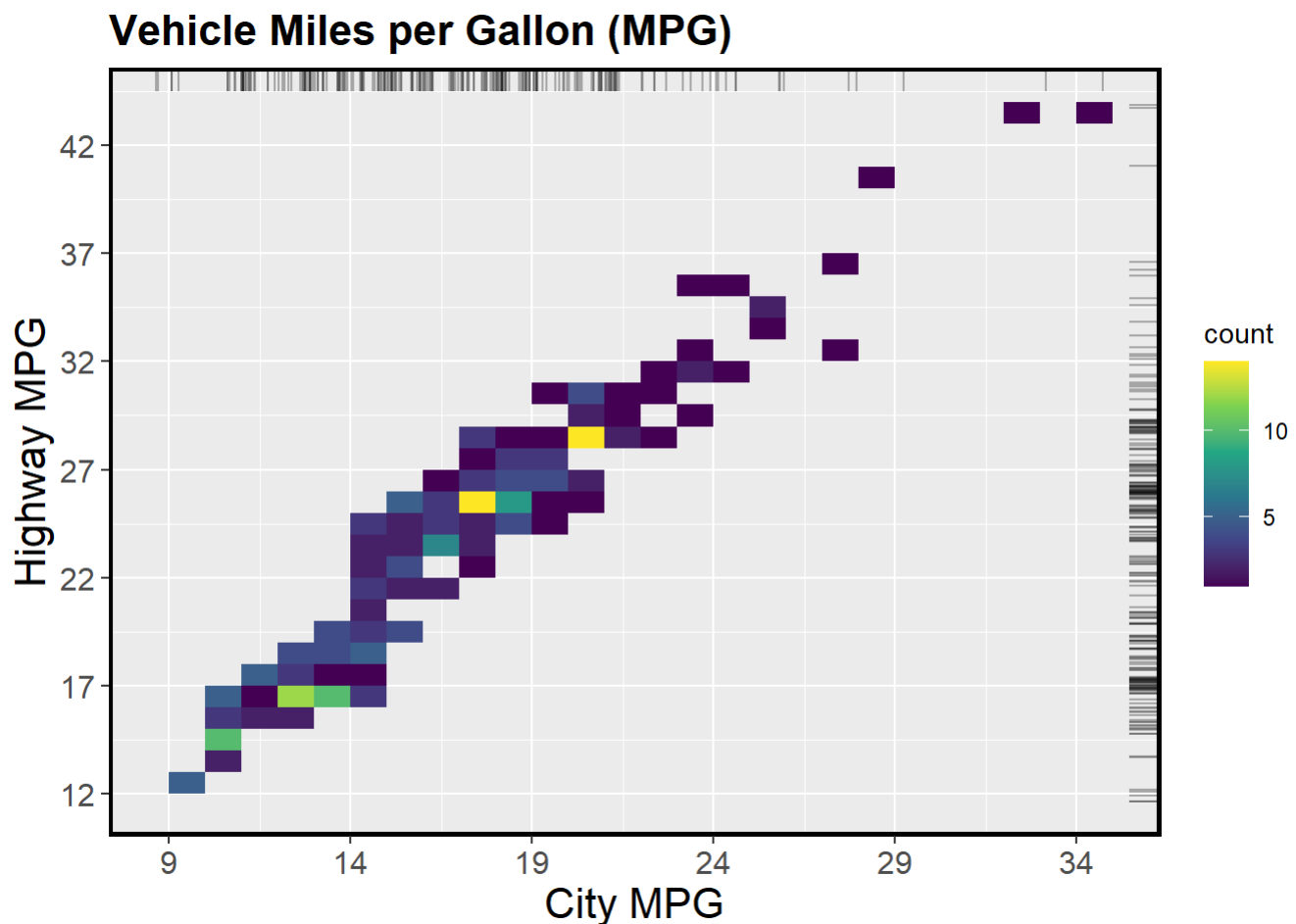


```
ggsave(
    filename = "images/04_assignment_fig1.png",
    units = "cm",
    width = 29.7,
    height = 21,
    dpi = 600
    )
```

```
min_cty <- min(df$cty)
max_cty <- max(df$cty)
min_hwy <- min(df$hwy)
max_hwy <- max(df$hwy)

df %>%
  ggplot(aes(x = cty, y = hwy)) +
  geom_bin2d(binwidth = c(1,1)) +
  geom_rug(position = "jitter", sides = "tr", alpha = 0.3) +
  scale_fill_viridis_c() +
  scale_x_continuous(breaks = seq(min_cty, max_cty, 5)) +
  scale_y_continuous(breaks = seq(min_hwy, max_hwy, 5)) +
  xlab("City MPG") +
  ylab("Highway MPG") +
  ggtitle("Vehicle Miles per Gallon (MPG)") +
  theme(
      axis.text = element_text(size = 12),
      axis.title = element_text(size = 16),
      plot.title = element_text(size = 16, face = "bold"),
      panel.border = element_rect(color = "black", fill = NA, size = 1.5)
      )
```


Vehicle Miles per Gallon (MPG)

```
ggsave(
    filename = "images/04_assignment_fig2.png",
    units = "cm",
    width = 29.7,
    height = 21,
    dpi = 600
    )
```

There appears to be a positive linear relationship between city MPG and highway MPG for vehicles. Both attributes have similar vehicles that are outliers (i.e. if a vehicle has very high city MPG, it also likely has very high highway MPG as well).

Exercise 2

```
box <- df %>%
  ggplot(aes(x = manufacturer, y = hwy)) +
  geom_boxplot(outlier.colour = "red") +
  geom_jitter(alpha = 1/5) +
  scale_y_continuous(breaks = seq(min_hwy, max_hwy, 5)) +
  coord_flip() +
  xlab("Manufacturer") +
  ylab("Highway Miles per Gallon") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 12),
        panel.border = element_rect(color = "black", fill = NA, size = 1.2))
```

```
hex <- df %>%
  ggplot(aes(x = displ, y = hwy)) +
  geom_hex(binwidth = c(0.25,1)) +
  scale_x_continuous(breaks = seq(floor(min(df$displ)), ceiling(max(df$displ)), 1)) +
  scale_y_continuous(breaks = seq(min_hwy, max_hwy, 5)) +
  scale_fill_viridis_c(option = "inferno") +
  xlab("Engine Displacement") +
  ylab("Highway Miles per Gallon") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 12),
        panel.border = element_rect(color = "black", fill = NA, size = 1.2)
        )
```

```
viol <- df %>%
  ggplot(aes(y = hwy, x = class)) +
  geom_violin(
      scale = "area",
      fill = "darkgoldenrod1",
      color = "black",
      draw_quantiles = c(.25, .5, .75)
      ) +
  geom_jitter(alpha = 1/5) +
  scale_y_continuous(breaks = seq(min_hwy, max_hwy, 5)) +
  xlab("Vehicle Type") +
  ylab("Highway Miles per Gallon") +
  coord_flip() +
  theme(axis.text = element_text(10),
        axis.title = element_text(size = 12),
        panel.border = element_rect(color = "black", fill = NA, size = 1.2),
        )
```

```
box + hex + viol + plot_layout(ncol = 2)
```
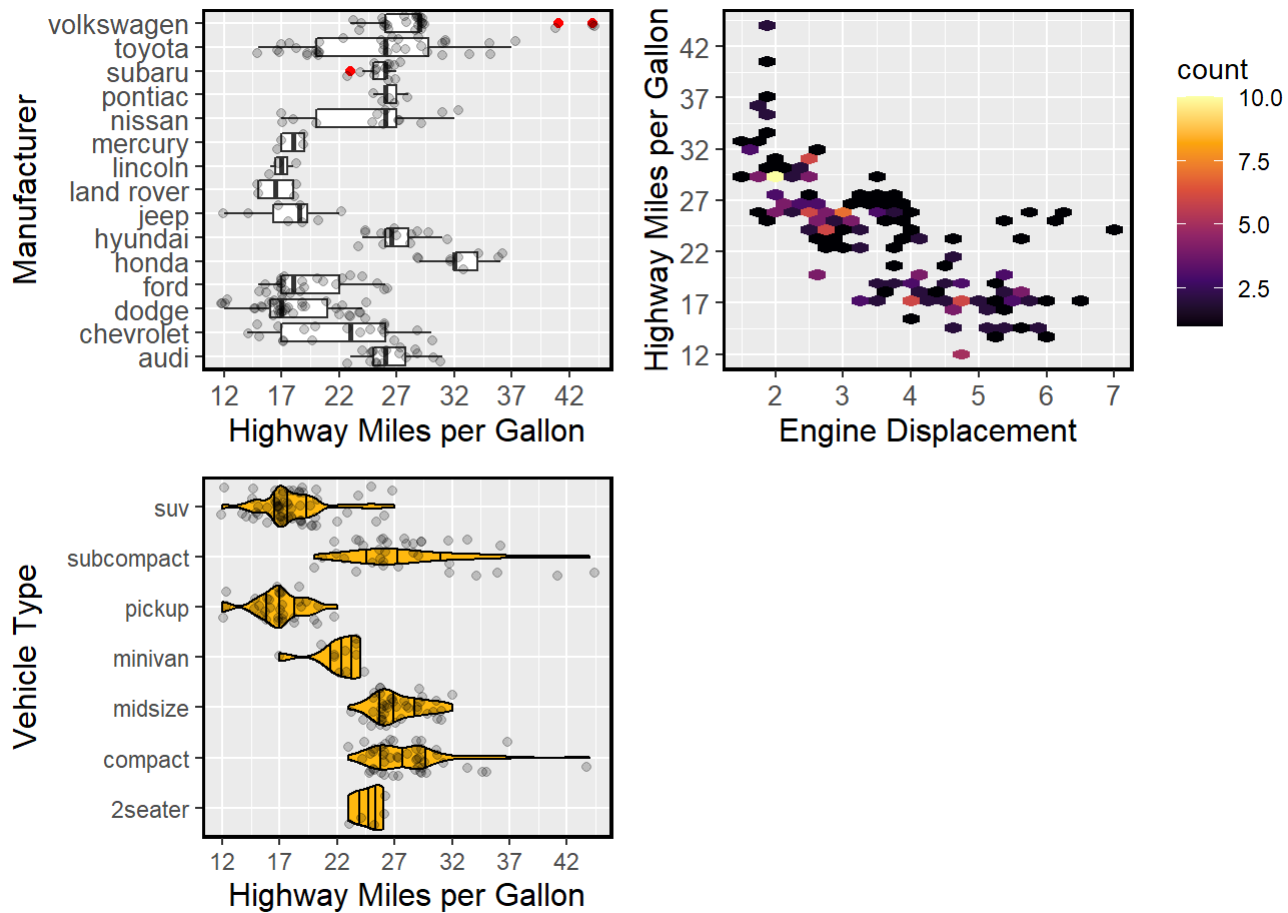
```
ggsave(
    filename = "images/04_assignment_fig3.png",
    units = "cm",
    width = 29.7,
    height = 21,
    dpi = 600
    )
```

From the above plots, it is clear that a vehicle's highway MPG is different for different subsets of vehicles. For instance, if one were to choose a Honda, subcompact, compact or midsize vehicle, that vehicle will likely have a higher highway MPG relative to other vehicles. If one chose to drive a vehicle with low engine displacement (particularly 2 or below), that vehicle will very likely have great fuel efficiency relative to vehicles with moderate or high engine displacement.

On the other hand, pickups and SUVs tend to have much worse fuel efficiency compared to other classes, with median highway MPG much lower than in other vehicle types. Some manufacturers with lower median highway MPG include Lincoln, Mercury, Land Rover, Jeep, Ford and Dodge (with the latter 3 showing more variability). Lastly, vehicles with higher (> 4) engine displacement, tend to have lower highway MPG.