

## Actividad 6 - Martín Olivera - CI 4845488-3

STAT\_NT

8/5/2021

### Ejercicio 1

```
library(ggplot2)
ggplot(mpg, aes(hwy, cty)) + geom_point() + theme(aspect.ratio = 1) +
  labs(x = "Millas de ruta por galón" , y = "Número de cilindros")
```

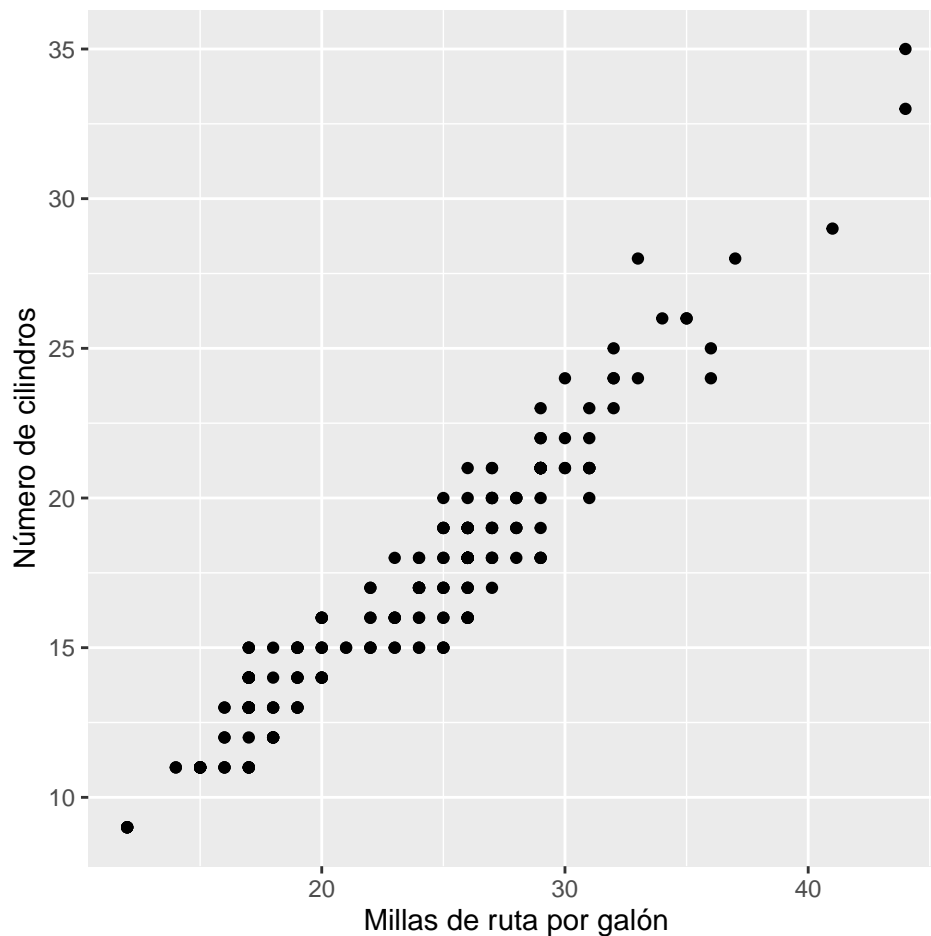


Figure 1: Nube de puntos (scatterplot) entre hwy y cty

Los datos surgen de la base mpg que contiene datos economicos del combustible entre 1999 y 2008 para 38 modelos de autos. Parece existir una relación de asociación positiva entre el número de cilindros y las millas recorrida por galon de combustible. A su vez esta relación parece estar fuertemente guiada por una relación

lineal entre ambas variables. Sin embargo en altos valores de ambas variables, la cantidad de observaciones se reduce y por tanto, el grado de asociación no parece del todo claro. (?mpg).

## Ejercicio 2

```
ggplot(mpg, aes(hwy, cty)) + geom_point(alpha = 1/3) + theme(aspect.ratio = 1) +  
  labs(x = "Millas de ruta por galón" , y = "Número de cilindros")
```

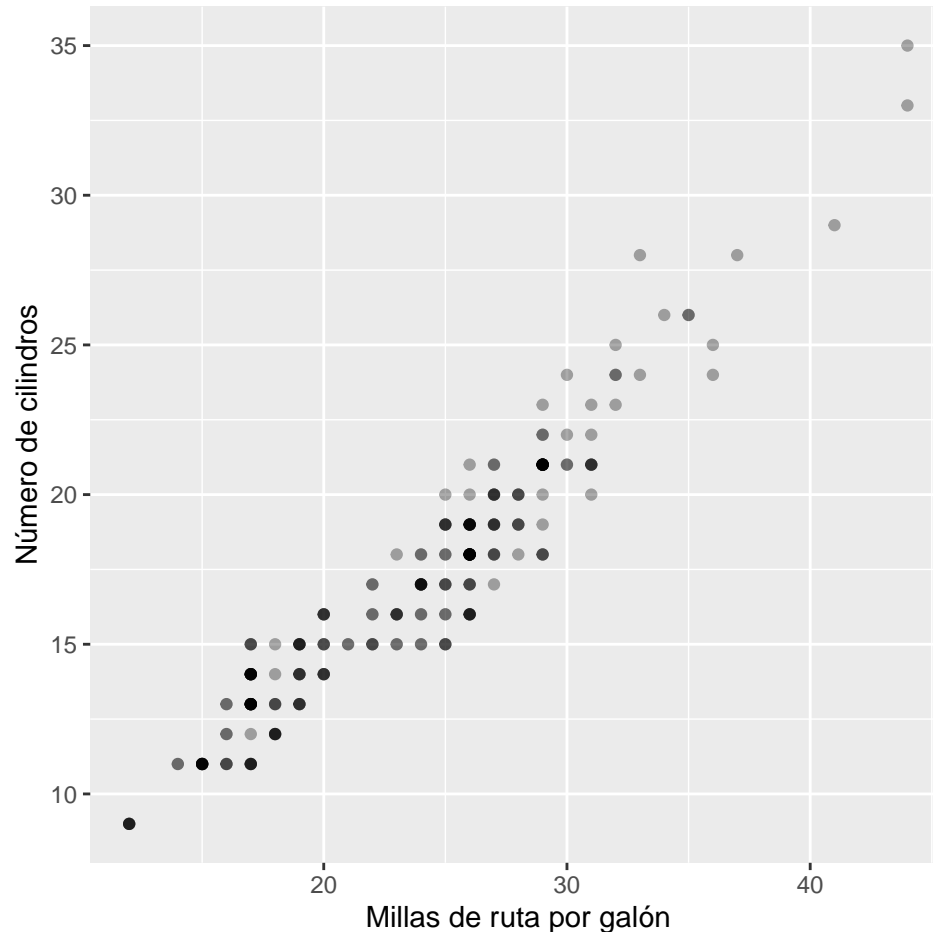


Figure 2: Nube de puntos (scatterplot) entre hwy y cty

Dada la existencia del problema del sobreploteo, se propone diferenciar la cantidad de observaciones usando las transparencias con  $\alpha=1/3$ , es decir, donde cada observación sea  $1/3$  del color del punto. Así, los puntos más oscuros resultan de mayor cantidad de observaciones en esa posición. Aún así se puede interpretar que se mantiene la relación lineal positiva (y sin demasiada variabilidad) entre el número de cilindros de los autos y la cantidad de millas por galon de fuel consumido. Observando este gráfico los valores estan concentrados en valores menores a 30 millas por galon y 20 cilindros. Aunque hay algunos valores aislados, no parecen tener carácter de outlier ni de observación influyente.

Comentario: "no parecen tener carácter de outlier" según que? A veces mejor decir menos que más y errarle.

```
ggplot(mpg, aes(hwy, cty)) + geom_point() + theme(aspect.ratio = 1) + geom_jitter() +  
  labs(x = "Millas de ruta por galón" , y = "Número de cilindros")
```

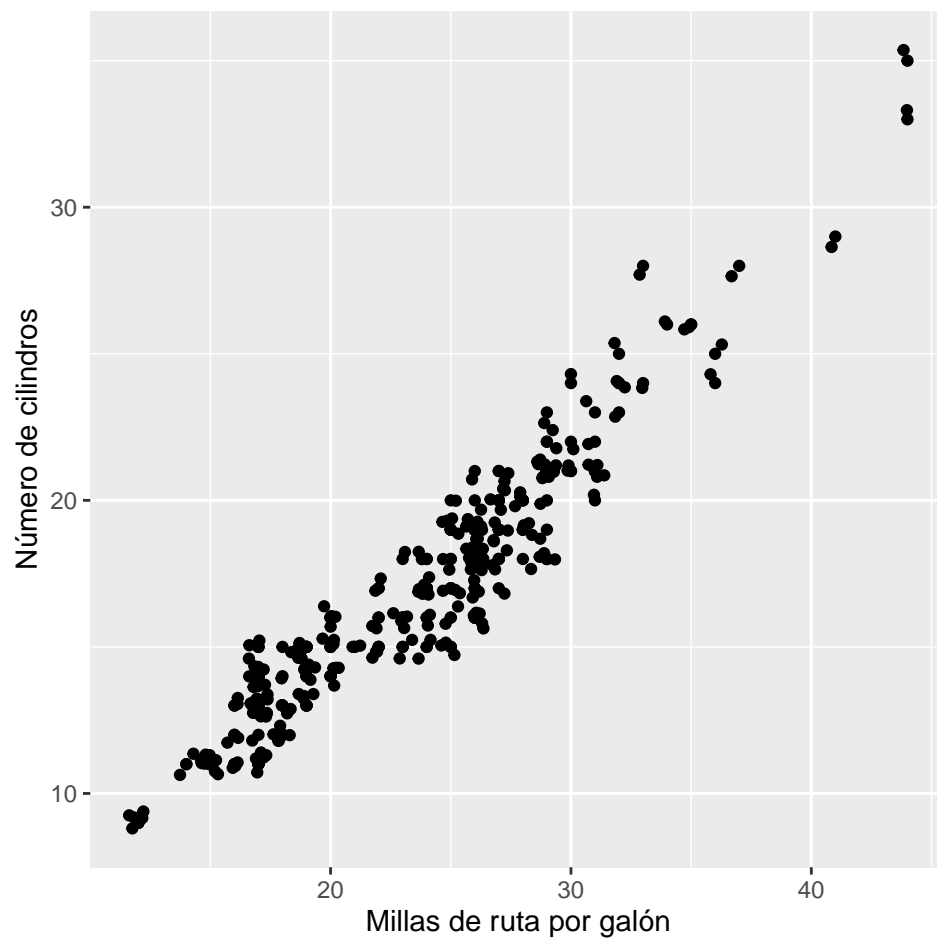


Figure 3: Nube de puntos (scatterplot) entre hwy y cty

El gráfico anterior usando la opción `geom_jitter()` colabora a la solución del problema del sobreploteo incluyendo variaciones aleatorias de los puntos, a fin de distinguir unos de otros. Las conclusiones del gráfico son análogas a los items anteriores. Sin embargo, en comparación con el gráfico con transparencias, este no parece ser tan claro en comparación al anterior, debido a la cantidad de puntos y la similitud entre las posiciones de los puntos. Visualmente las transparencias colaboran de mejor manera a entender sobre concentración de puntos a partir de los colores y discutir sobre la relación entre los indicadores de forma más certera.

### Ejercicio 3

```
ggplot(mpg, aes(x=hwy, y=cty, colour = class)) + geom_point() + theme(aspect.ratio = 1) + geom_jitter()
labs(x = "Millas de ruta por galón" , y = "Número de cilindros")
```

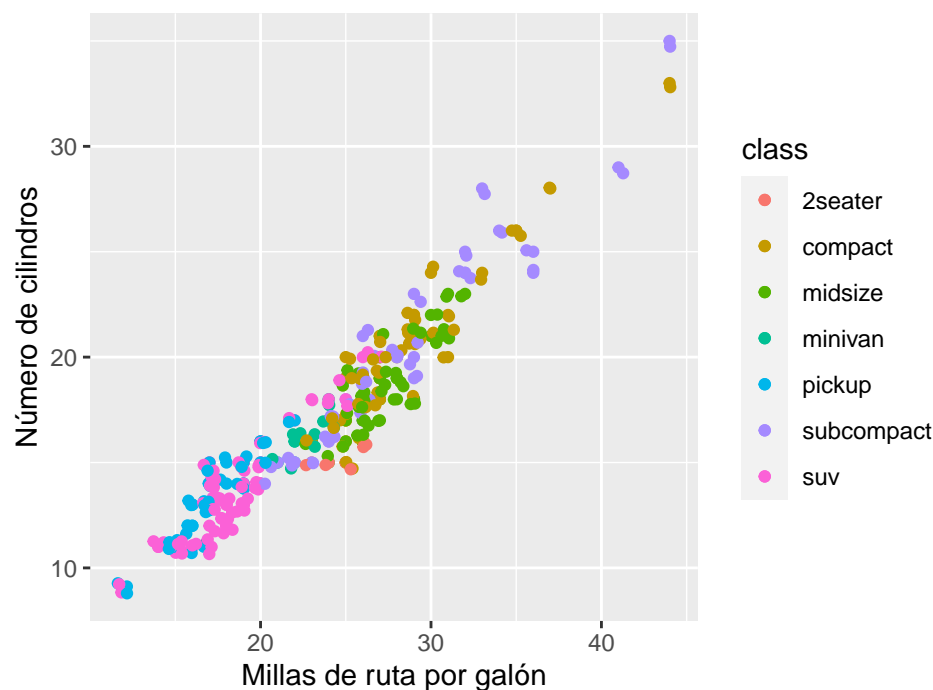


Figure 4: Nube de puntos (scatterplot) entre hwy y cty

A priori parecería que pickup y suv son los modelos con menor número de cilindros y menor rendimiento del combustible, mientras que compact y midsize toman valores intermedio y subcompact es el modelo que tiene mayor cantidad de cilindros y además mayor cantidad de millas por galón.

```
ggplot(mpg, aes(x=hwy, y=cty, colour = class)) + geom_point() + theme(aspect.ratio = 1) + scale_color_b
labs(x = "Millas de ruta por galón" , y = "Número de cilindros")
```

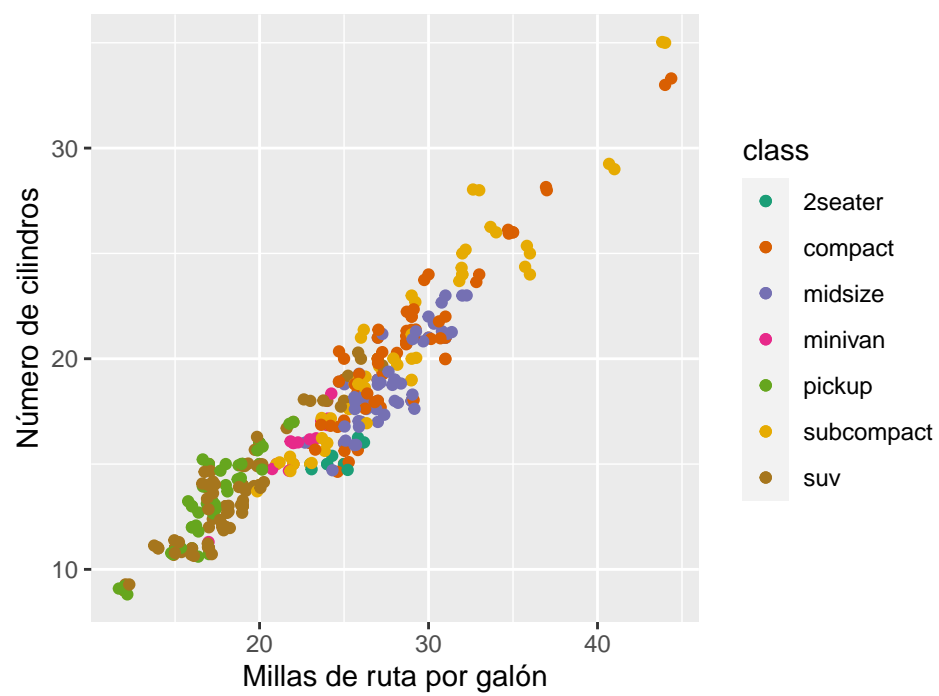


Figure 5: Nube de puntos (scatterplot) entre hwy y cty

```
ggplot(mpg, aes(x=hwy, y=cty, colour = class)) + geom_point() + theme(aspect.ratio = 1) + scale_color_b
```

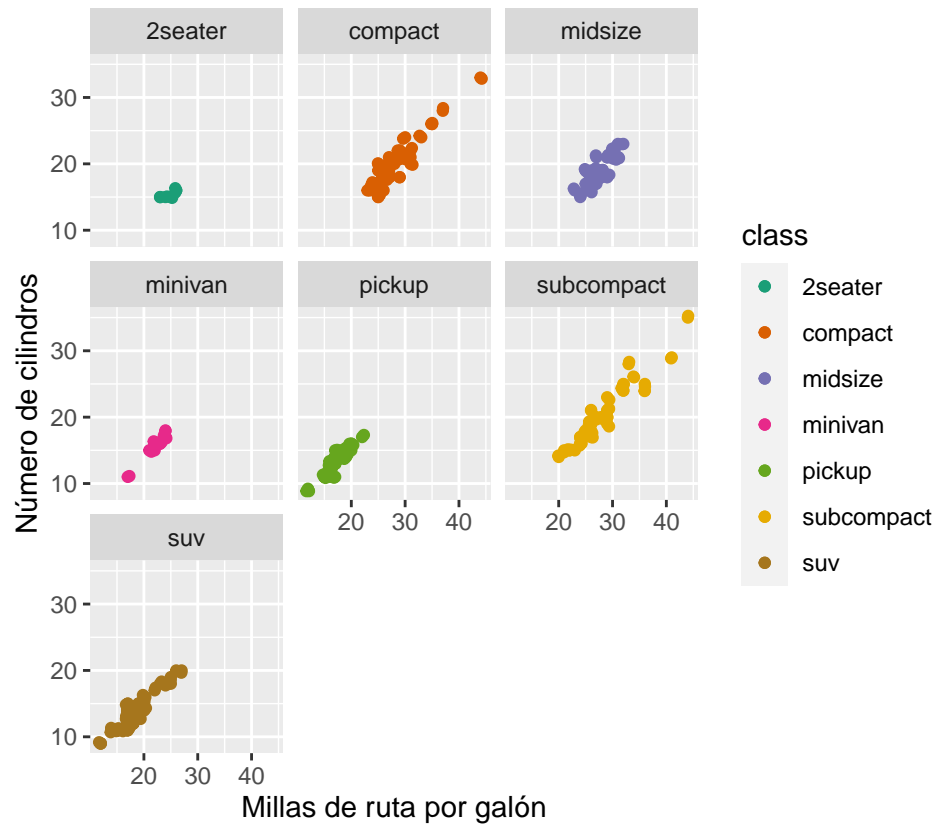


Figure 6: Nube de puntos (scatterplot) entre hwy y cty

De lo anterior se puede observar que nuevamente midsize y compact contienen sus datos en niveles intermedios de las variables, mientras que subcompact toma valores medios y altos de las variables. Por último las restantes clases toman bajos valores de la cantidad de millas por galón y el número de cilindros que tiene ese modelo.

**Comentario:** Muy buen trabajo. A mejorar: revisar la ortografía (por favor) y cuidado en las interpretaciones (pero excelente por realizarlas!).