

# Tarea - Transformaciones Estadísticas con ggplot

Jesus Mudarra Luján

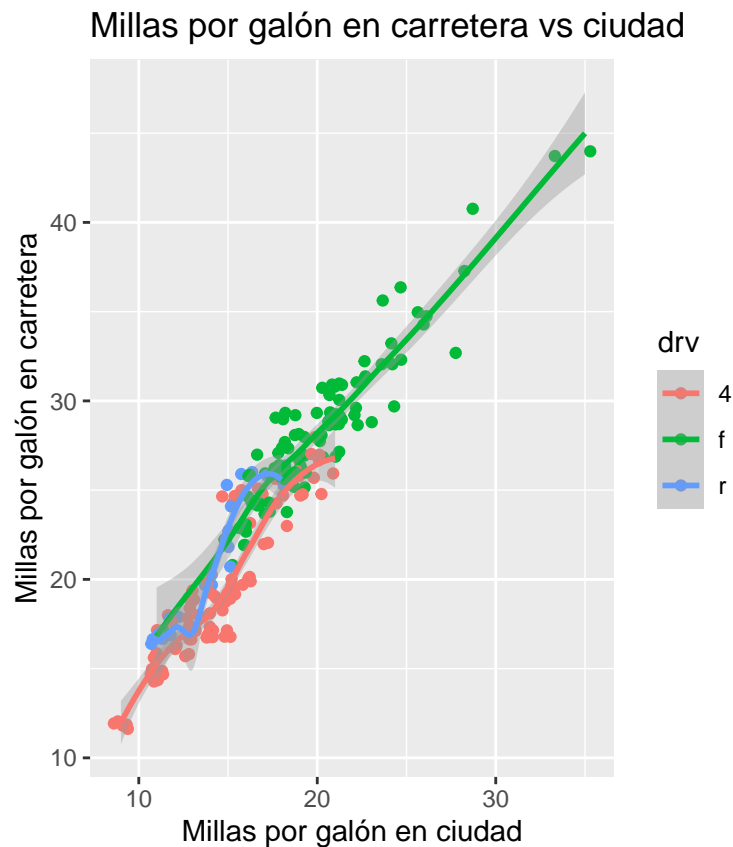
2022-09-07

## Ejercicio 1

El siguiente gráfico que genera el código de R es correcto, pero puede mejorarse. ¿Qué cosas añadirías para mejorarlo?

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy, color = drv)) +  
  geom_jitter() +  
  geom_smooth() +  
  coord_fixed() +  
  labs(x = "Millas por galón en ciudad", y = "Millas por galón en carretera",  
       title = "Millas por galón en carretera vs ciudad")
```

## 'geom\_smooth()' using method = 'loess' and formula 'y ~ x'



## Ejercicio 2

Investiga la documentación de `geom_jitter()`. ¿Qué parámetros controlan la cantidad de ruido aleatorio (jitter)?

Los parámetros que controlan la cantidad de ruido son `width` y `height`.

## Ejercicio 3

Compara las funciones `geom_jitter` contra `geom_count` y busca semejanzas y diferencias entre ambas.

`geom_jitter` añade una pequeña cantidad de ruido aleatorio a cada punto, y es una forma útil de manejar el overlapping causado por la discreción en datasets pequeños.

`geom_count` cuenta el número de observaciones en cada ubicación, y luego mapea el recuento en el área de puntos. Es útil cuando se tienen datos discretos u overlapping

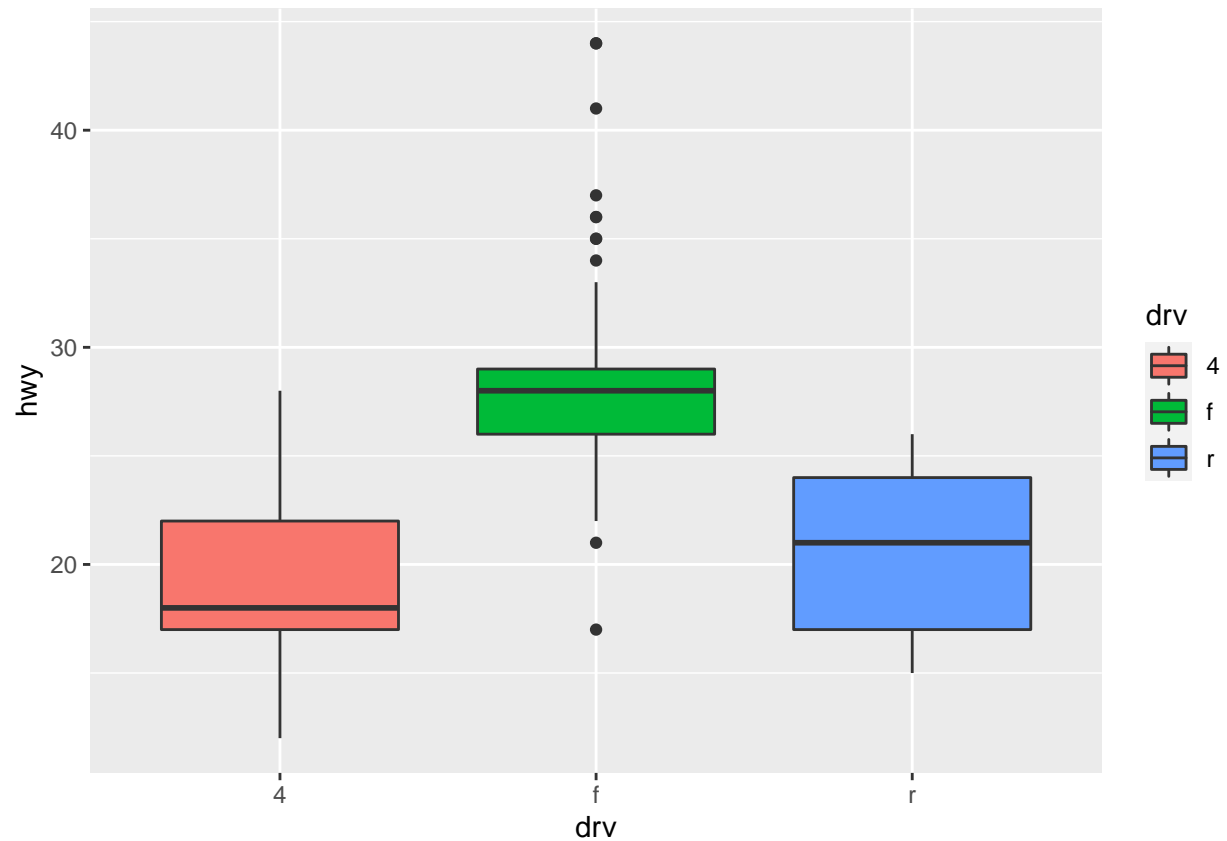
Tienen en común que las dos son variantes de `geom_point` y son extremadamente útiles cuando se solapan los puntos.

## Ejercicio 4

¿Cuál es el valor por defecto del parámetro `position` de un `geom_boxplot`? Usa el dataset de `diamonds` o de `mpg` para hacer una visualización que lo demuestre.

El valor por defecto del parámetro `position` es `position = "dodge2"`.

```
ggplot(data = mpg, mapping = aes(x = drv,
                                  y = hwy,
                                  fill = drv)) +
  geom_boxplot()
```



## Ejercicio 5

Convierte un diagrama de barras apilado en un diagrama de sectores o de tarta usando la función `coord_polar()`.

Diagrama de barras apilado:

```
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, fill = color))
```

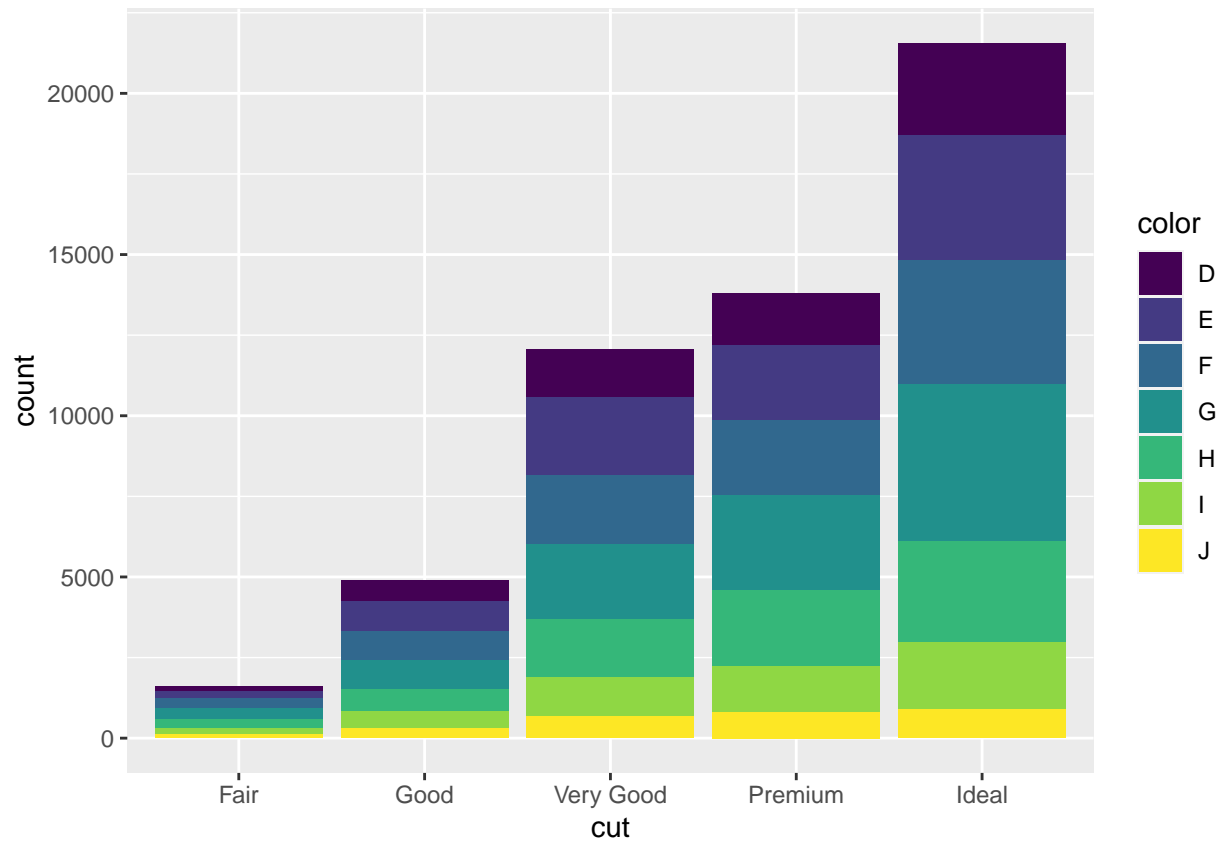
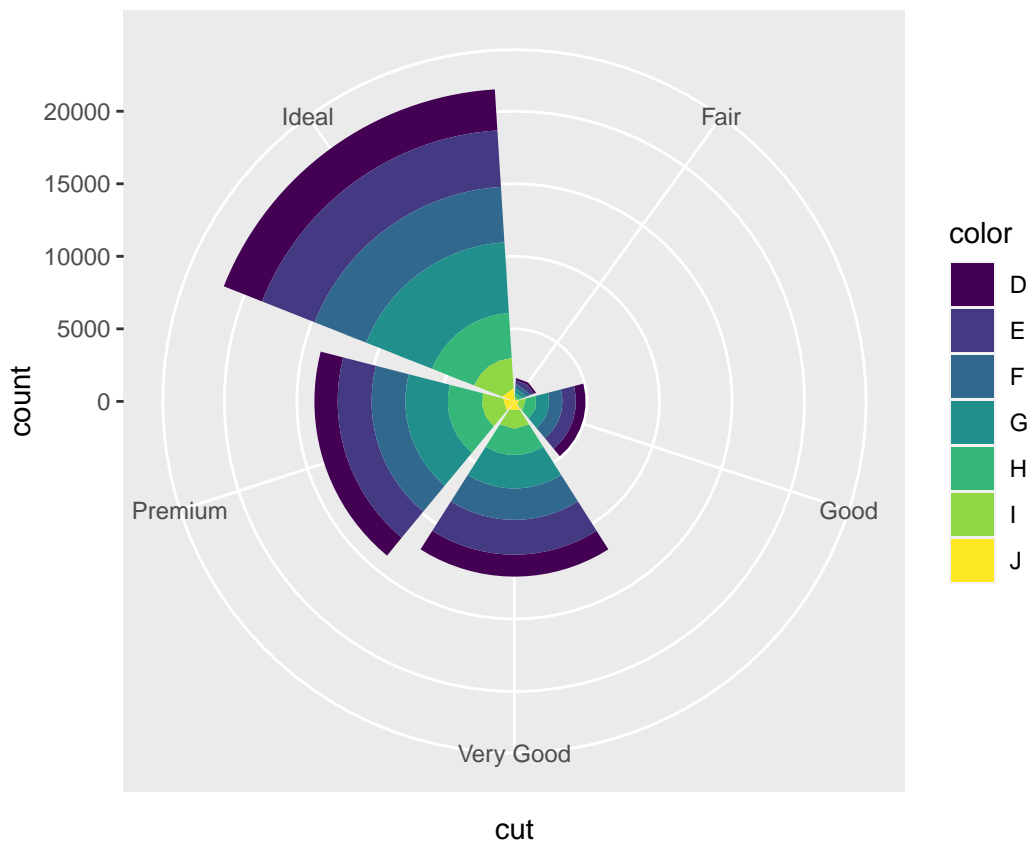


Diagrama de barras apilado convertido a diagrama de sectores:

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = color)) +  
  coord_polar()
```



## Ejercicio 6

¿Qué hace la función `labs()`? Lee la documentación y explícalo correctamente.

Se trata de una función que modifica los ejes x e y, la leyenda y las etiquetas del gráfico.

## Ejercicio 7

¿En qué se diferencian las funciones `coord_quickmap()` y `coord_map()`?

`coord_map()` proyecta una porción de la tierra (aproximadamente esférica), sobre un plano 2D utilizando cualquier proyección definida por el paquete `mapproj`. En general, las proyecciones de mapas no conservan las líneas rectas, por lo que requiere un cálculo importante. `coord_quickmap()`, es una aproximación rápida que sí conserva las líneas rectas. Funciona mejor para áreas pequeñas más cercanas al ecuador.

## Ejercicio 8

Investiga las coordenadas `coord_fixed()` e indica su función.

`coord_fixed()` obliga a establecer una relación específica entre la representación física de las unidades de datos en los ejes. La relación representa el número de unidades en el eje y equivalente a una unidad en el eje x.

Argumentos:

- **ratio**: Relación de aspecto, expresada como  $y/x$ .
- **xlim**: Límites para el eje x.
- **ylim**: Límites para el eje y.
- **expand**: Si es **TRUE**, el valor por defecto, añade un pequeño factor de expansión a los límites para asegurar que los datos y los ejes no se solapen. Si es **FALSE**, los límites se toman exactamente de los datos o de **xlim/ylim**.
- **clip**: ¿Debe recortarse el dibujo a la medida del panel de la parcela?

## Ejercicio 9

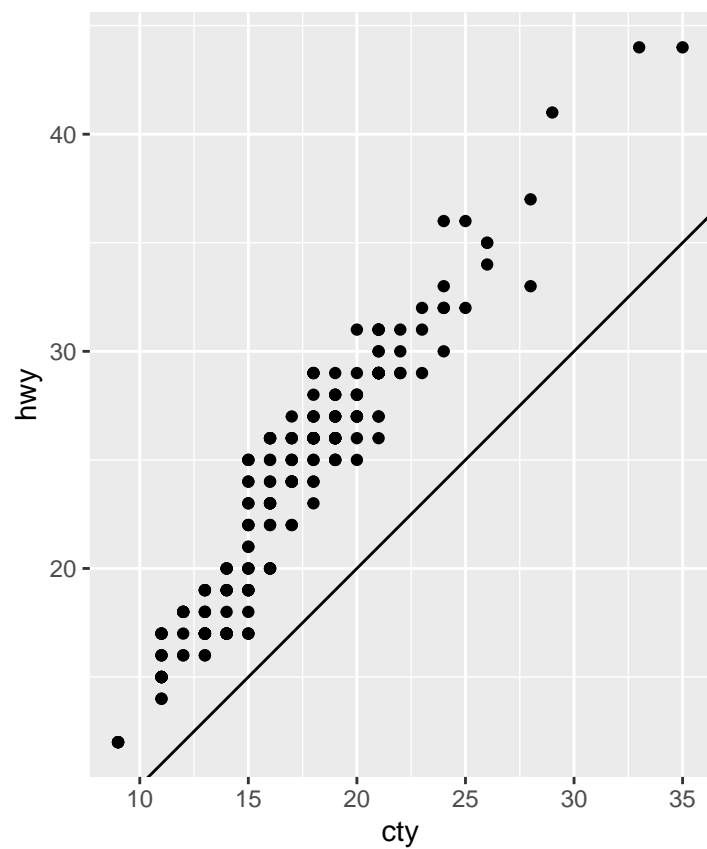
Investiga la geometría de la función `geom_abline()`, `geom_vline()` y `geom_hline()` e indica su función respectivamente.

- `geom_abline()`: Añade líneas de referencia a un gráfico diagonales.
- `geom_vline()`: Añade líneas de referencia a un gráfico verticales.
- `geom_hline()`: Añade líneas de referencia a un gráfico horizontales.

## Ejercicio 10

¿Qué nos indica el gráfico siguiente acerca de la relación entre el consumo en ciudad y en autopista del dataset de mpg?

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +
  geom_point() +
  geom_abline() +
  coord_fixed()
```



El gráfico nos indica que tienen una relación lineal. Cuando aumenta la consumo en ciudad aumenta el consumo en autopista y viceversa.