

# Resolución tarea dplyr

## Enunciado

Usando la base de personas y gastos de la VIII EPF:

1. Calculen el gasto promedio de los hogares por división y zona (deben quedar dos filas por cada división)
2. Calculen la suma del gasto por hogar en la división 1 (debe quedar una observación por hogar)
3. Unir el gasto calculado de la división 1 con la base de personas

**Nota:** en primera instancia, no se preocupen de los factores de expansión, pero si se animan, es un muy buen ejercicio hacer los cálculos considerando la expansión.

Para cumplir con la tarea, deberán usar casi todas las funciones que vimos hoy, en especial: `group_by`, `summarise` y `left_join`

Es posible que algunas partes no sean tan fáciles, pero no se rindan. Sufrir un rato tiene su recompensa :)

## Ejercicio 1

```
names(gastos) <- tolower(names(gastos))
promedio_gasto <- gastos %>%
  filter(folio != "" & d != "") %>% #eliminar challa
  group_by(folio) %>%
  mutate(n = row_number() == 1) %>% #marca con 1 la primera observación
  group_by(zona) %>%
  mutate(hogares_zona = sum(n)) %>% #sumar hogares por zona
  group_by(zona, d) %>%
  mutate(gasto_medio = sum(gasto) / hogares_zona) %>% #promedio división-zona
  slice(1) %>% #una observación por division-zona
  select(zona, d, hogares_zona, gasto_medio)
```

## Ejercicio 2

Notar que dado que no todos los hogares tienen gasto en la división 1, algunos folios se pierden. Esto se soluciona en el ejercicio 3.

```
gasto_div1 <- gastos %>%
  filter(folio != "" & d != "") %>% #eliminar challa
  filter(d == "01") %>% #OJO: se pierden algunos folios
  group_by(folio) %>%
  mutate(suma_d1 = sum(gasto)) %>%
  slice(1) %>%
  select(folio, suma_d1)
```

## Ejercicio 3

Es importante la última línea, ya que allí se soluciona el problema de los hogares desaparecidos en el ejercicio 2.

```
final <- personas %>%  
  left_join(gasto_div1, by = "folio") %>%  
  mutate(suma_d1 = if_else(is.na(suma_d1), 0, suma_d1)) #hogares que no pegaron
```