

```
library(tidyverse)
```

```
casas <- read_csv("casas.csv")
```

```
glimpse(casas)
```

```
# 1. Condición y calidad general vs precio x m2
```

```
# haz una tabla de conteos de los valores
```

```
# de calidad general de construcción y terminados (calidad_gral)
```

```
# tip: usa la función tally o summarise
```

```
casas %>% group_by(calidad_gral) %>%
```

```
  summarise(n = n())
```

```
casas %>% group_by(condicion_venta) %>%
```

```
  summarise(n = n())
```

```
# Haz una gráfica de caja y brazos del precio x m2 para cada
```

```
# nivel de calidad gral ¿Qué interpretas?
```

```
# aquí tu código (tip: usa factor(variable) para convertir una variable
```

```
# numérica a factor)
```

```
ggplot(casas,
```

```
  aes(x = calidad_gral, y = precio_m2, group = calidad_gral)) +
```

```
  geom_boxplot()
```

```
ggplot(casas,
```

```
  aes(x = factor(calidad_gral), y = precio_m2)) +
```

```
  geom_boxplot()
```

```
# 2. Repite el anterior con número de coches que caben en el garage.
```

```
# ¿Cuál es la relación? ¿Qué puedes malinterpretar de esta gráfica?
```

```
# num_coches

ggplot(casas ,
       aes(x = factor(num_coches), y = precio_m2)) +
  geom_boxplot() + geom_jitter(width = 0.2)

casas %>% group_by(num_coches) %>% summarise(n = n())
```

```
# aquí tu respuesta
```

```
# 3. Tablas cruzadas: azúcar y tipo de té
```

```
#Lee los datos de tomadores de té
```

```
tea <- read_csv("tea.csv")
```

```
## Haz una tabla cruzada de uso de azúcar (sugar) con tipo de té (Tea),
```

```
## mostrando el número de casos en cada cruce
```

```
# Pon el uso de azúcar en las columnas
```

```
tea %>% group_by(Tea)# aquí tu código

tea_analysed_sugar_n <- tea |> select(Tea, sugar) |>
  count(sugar, Tea) |>
  group_by(sugar)
```

```
# results:
```

```
# sugar  Tea      n
```

```
# <chr>  <chr>   <int>
```

```
# 1 No.sugar Earl Grey  84
```

```
# 2 No.sugar black     51
```

```
# 3 No.sugar green     20
```

```
# 4 sugar Earl Grey 109
```

```
# 5 sugar black 23
```

```
# 6 sugar green 13
```

```
## ¿Cómo se relaciona el uso de azúcar con el té que toman las personas?
```

```
## Haz una tabla de porcentajes por renglón (para cada tipo de té) de tu tabla anterior
```

```
#tea %>% # aquí tu código
```

```
tea_analysed_sugar_pct <- tea |> select(Tea, sugar) |>
```

```
count(sugar, Tea) |>
```

```
group_by(sugar) |>
```

```
mutate(pct = 100*n/sum(n)) |>
```

```
select(-n)
```

```
# results:
```

```
# sugar Tea pct
```

```
# <chr> <chr> <dbl>
```

```
# 1 No.sugar Earl Grey 54.2
```

```
# 2 No.sugar black 32.9
```

```
# 3 No.sugar green 12.9
```

```
# 4 sugar Earl Grey 75.2
```

```
# 5 sugar black 15.9
```

```
# 6 sugar green 8.97
```

```
# El problema con esta representación es que toma en cuenta todo los n, así que hay que separar
```

```
# en unos sin y con azúcar.
```

```
tea_EG <- filter(tea, Tea=="Earl Grey")
```

```
tea_green <- filter(tea, Tea=="green")
```

```
tea_black <- filter(tea, Tea=="black")
```

```

tea_analysed_sugar_pct_EG <- tea_EG |> select( sugar, Tea) |>
  group_by(Tea) |>
  count(sugar) |>
  mutate(pct = 100*n/sum(n)) |>
  select(-n)

```

```

tea_analysed_sugar_pct_black <- tea_black |> select( sugar, Tea) |>
  group_by(Tea) |>
  count(sugar) |>
  mutate(pct = 100*n/sum(n)) |>
  select(-n)

```

```

tea_analysed_sugar_pct_green <- tea_green |> select( sugar, Tea) |>
  group_by(Tea) |>
  count(sugar) |>
  mutate(pct = 100*n/sum(n)) |>
  select(-n)

```

```
# results
```

```

# Tea    sugar    pct
# <chr>  <chr>    <dbl>
# 1 Earl Grey No.sugar 43.5
# 2 Earl Grey sugar    56.5

```

```

# Tea  sugar    pct
# <chr> <chr>    <dbl>
# 1 green No.sugar 60.6

```

```
# 2 green sugar    39.4
```

```
# Tea  sugar    pct
```

```
# <chr> <chr>    <dbl>
```

```
# 1 black No.sugar 68.9
```

```
# 2 black sugar    31.1
```

```
## 4. Haz una tabla cruzada para la variable Tea y la presentación (how)
```

```
## donde las columnas son el tipo de Té
```

```
#tea %>% # tu código
```

```
tea_analysed_Tea_present <- tea |> select(Tea, how) |>
```

```
  group_by(how) |>
```

```
  count(Tea)
```

```
# result
```

```
# Tea      how      n
```

```
# <chr>    <chr>    <int>
```

```
# 1 Earl Grey tea bag      117
```

```
# 2 Earl Grey tea bag+unpackaged  60
```

```
# 3 Earl Grey unpackaged     16
```

```
# 4 black   tea bag      36
```

```
# 5 black   tea bag+unpackaged  23
```

```
# 6 black   unpackaged     15
```

```
# 7 green   tea bag      17
```

```
# 8 green   tea bag+unpackaged  11
```

```
# 9 green    unpackaged      5
```

```
##### not necessary
```

```
num_Teas <- n_distinct(tea |> select(Tea))
```

```
num_how <- n_distinct(tea |> select(how))
```

```
#####
```

```
# organizacion de la table con las columnas tipo de Tea, las celdas contienen en numero
```

```
tabla_cruzada <- tea |>
```

```
  count(how, Tea) |>
```

```
  group_by(how)
```

```
# reorganizar para que las columnas sea de Tea, las celdas los numeros
```

```
tabla_cruzada |>
```

```
  pivot_wider(names_from = Tea, values_from = n)
```

```
# result N
```

```
# how      `Earl Grey` black green
```

```
# <chr>      <int> <int> <int>
```

```
# 1 tea bag      117  36  17
```

```
# 2 tea bag+unpackaged      60  23  11
```

```
# 3 unpackaged      16  15  5
```

```
## Ahora calcula porcentajes por columna de la tabla anterior
```

```
#tea %>% # tu código aquí
```

```
## overwriting the former result to replace number with percentage
```

```
tabla_cruzada <- tea |>
```

```
count(how, Tea) |>
```

```
group_by(how) |>
```

```
mutate(pct = round(100 * n / sum(n))) |>
```

```
select(-n)
```

```
tabla_cruzada |>
```

```
pivot_wider(names_from = Tea, values_from = pct)
```

```
# result %
```

```
# how          `Earl Grey` black green
```

```
# <chr>         <dbl> <dbl> <dbl>
```

```
# 1 tea bag      69   21   10
```

```
# 2 tea bag+unpackaged  64   24   12
```

```
# 3 unpackaged    44   42   14
```

```
# ¿Cómo son diferentes en cuanto a la presentación
```

```
# los tomadores de distintos té (negro, earl gray, etc.)?
```

```
# Parece que los tomadores de té negro prefieren la presentación "unpackaged" mas
```

```
# que las bolsitas mientras el Earl Gray se toma mas de bolsas. Parece que el
```

```
# tiene poco consumidores o por lo menos no depende de la presentación.
```

```
g1 <- ggplot(tabla_cruzada) +
```

```
geom_point(aes(x = how, y = pct, color = tabla_cruzada$Tea))
```

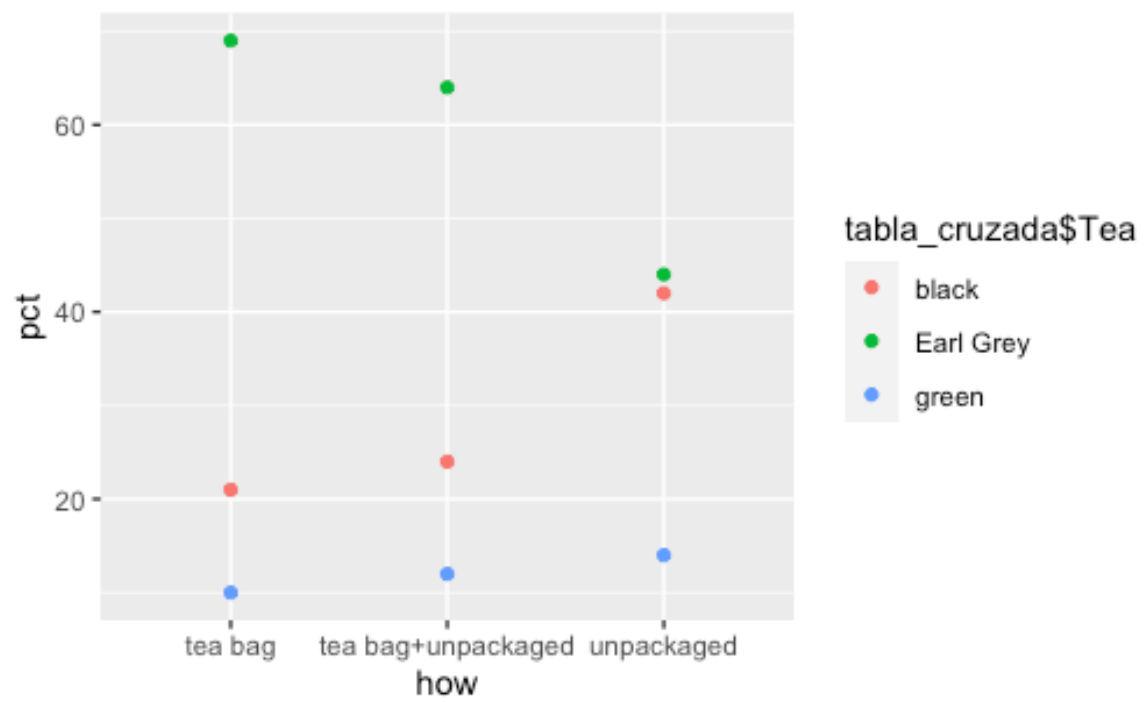


Ilustración 1: resultado de la tabla cruzada Tea vs how y porcentaje.