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 codigo (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 repesentación es que toma en cuenta todo los n, asi que hay que separar

# en unos sin y con azucar.

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))

######

# organisacion 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 tés (negro, earl gray, etc.)?

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

# que las bolsitas mientas 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))

A graph with green and red dots

Description automatically generated

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