Análisis de estilo de vida y salud

[Actualizar con nombres del equipo]

16 de September de 2025

Contents

0. Cargar paquetes y datos

```
required_packages <- c(
  "tidyverse", "janitor", "skimr", "gt",
  "GGally", "broom", "gridExtra", "viridis"
)

invisible(lapply(required_packages, function(pkg) {
  if (!requireNamespace(pkg, quietly = TRUE)) {
    stop(paste0("El paquete '", pkg, "' no está instalado. Instálalo antes de compilar el notebook."))
  }
  library(pkg, character.only = TRUE)
}))

theme_set(theme_minimal(base_size = 11))

datos <- readr::read_csv("data/health_lifestyle_classification.csv") |>
  janitor::clean_names()

# Ajustar tipos de variables categóricas
categoricas <- c(
  "gender", "sleep_quality", "alcohol_consumption", "smoking_level",
  "mental_health_support", "education_level", "job_type", "occupation",
  "diet_type", "exercise_type", "device_usage", "healthcare_access",
  "insurance", "sunlight_exposure", "caffeine_intake", "family_history",
  "pet_owner", "target"
)

datos <- datos |> mutate(across(all_of(categoricas), as.factor))

# Variables ordinales (ajustar el orden según interpretación del equipo)
ord_sleep <- c("Poor", "Fair", "Good", "Excellent")
```

```r
ord_alcohol <- c("None", "Occasionally", "Regularly")
ord_smoking <- c("Non-smoker", "Light", "Moderate", "Heavy")
ord_healthcare <- c("Poor", "Moderate", "Good", "Excellent")
ord_device <- c("Low", "Moderate", "High")

ordenes <- list(
  sleep_quality = ord_sleep,
  alcohol_consumption = ord_alcohol,
  smoking_level = ord_smoking,
  healthcare_access = ord_healthcare,
  device_usage = ord_device
)

for (var in names(ordenes)) {
  if (var %in% names(datos)) {
    datos[[var]] <- factor(datos[[var]], levels = ordenes[[var]], ordered = TRUE)
  }
}

# Guardar un tibble con metadatos básicos
metadata <- tibble(
  variable = names(datos),
  tipo = sapply(datos, function(x) class(x)[1]),
  descripcion = NA_character_
)
```

1. Marco metodológico y objetivo

Sugerencia: redactar este apartado en estilo narrativo. Incluya detalles sobre el origen simulado de los datos, descripción general del instrumento de recolección y contexto de aplicación.

```r
tipos <- vapply(datos, function(x) class(x)[1], character(1))
marco_general <- tibble(
  indicador = c(
    "Registros",
    "Variables",
    "Variables cuantitativas",
    "Variables categóricas"
  ),
  valor = c(
    nrow(datos),
    ncol(datos),
    sum(tipos %in% c("numeric", "integer", "double")),
    sum(tipos %in% c("factor", "ordered"))
  )
)

gt::gt(marco_general)
```

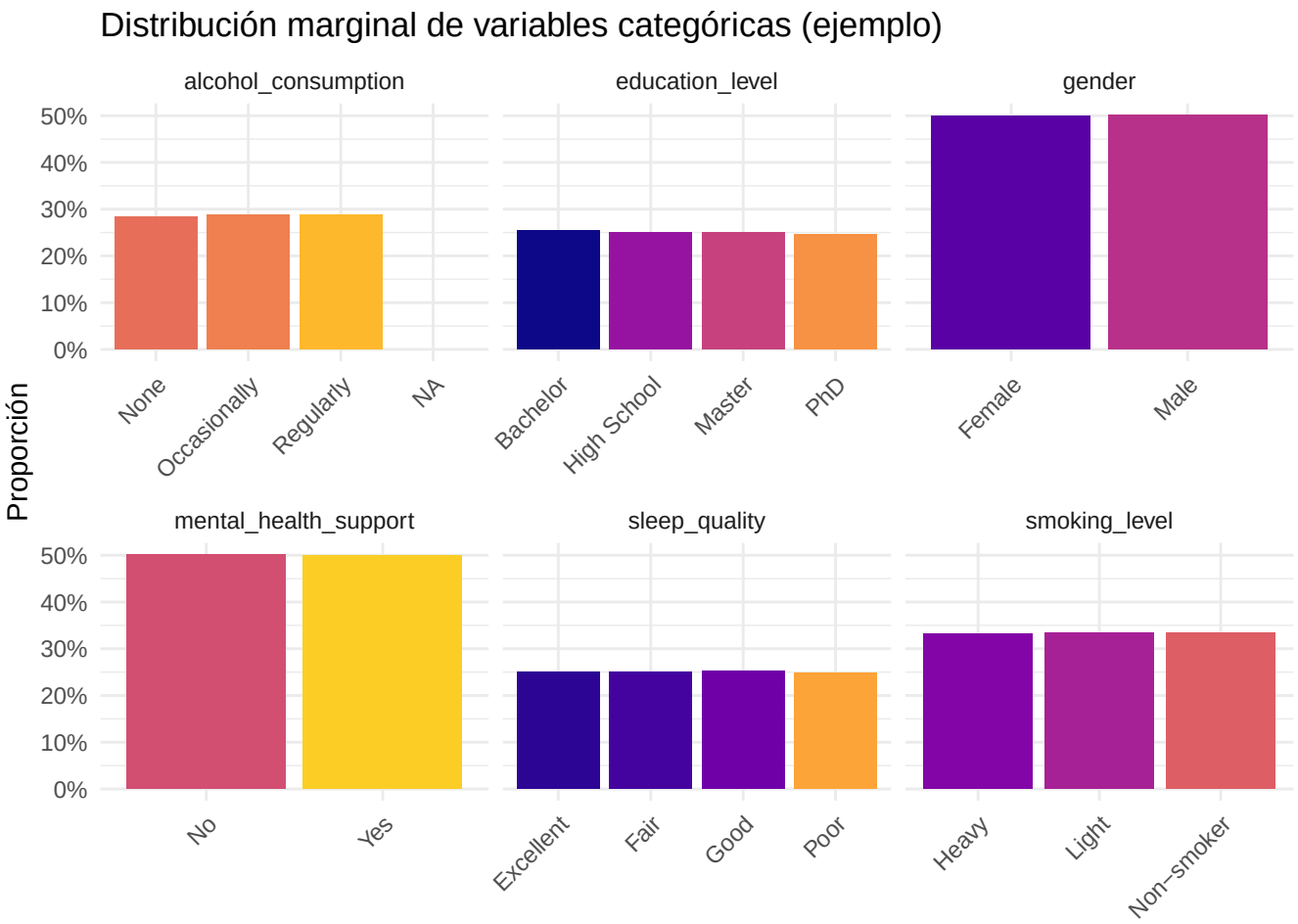
Objetivo del análisis: `<<Redactar un objetivo claro y medible para el estudio>>`

indicador	valor
Registros	100000
Variables	48
Variables cuantitativas	30
Variables categóricas	18

2. Análisis descriptivo de variables categóricas

2.1 Análisis univariado

```
cat_vars <- metadata |>
  filter(tipo %in% c("factor", "ordered")) |>
  pull(variable)

frecuencias_cat <- datos |>
  select(all_of(cat_vars)) |>
  mutate(across(everything(), as.character)) |>
  pivot_longer(everything(), names_to = "variable", values_to = "categoria") |>
  group_by(variable, categoria) |>
  summarise(n = n(), .groups = "drop") |>
  group_by(variable) |>
  mutate(
    porcentaje = n / sum(n)
  )

frecuencias_cat |>
  filter(variable %in% cat_vars[1:6]) |>
  ggplot(aes(x = categoria, y = porcentaje, fill = categoria)) +
  geom_col() +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_viridis_d(option = "C", end = 0.9, guide = "none") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(x = NULL, y = "Proporción", title = "Distribución marginal de variables categóricas (ejemplo)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



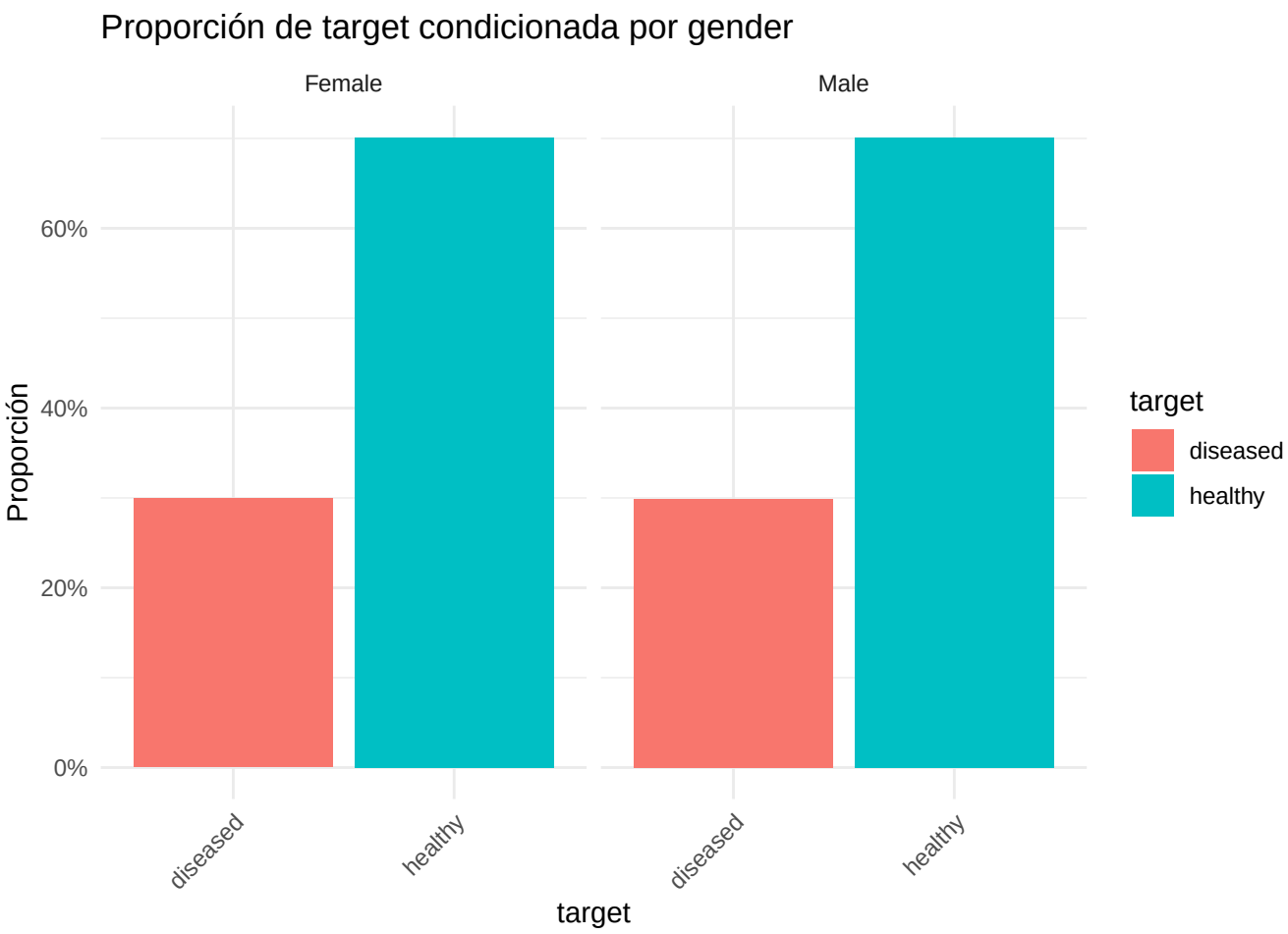

```
# Tabla resumen para documento (ajustar número de variables a mostrar)
frecuencias_cat |>
  filter(variable %in% cat_vars[1:3]) |>
  mutate(porcentaje = scales::percent(porcentaje)) |>
  arrange(variable, desc(n)) |>
  gt::gt(groupname_col = "variable")
```

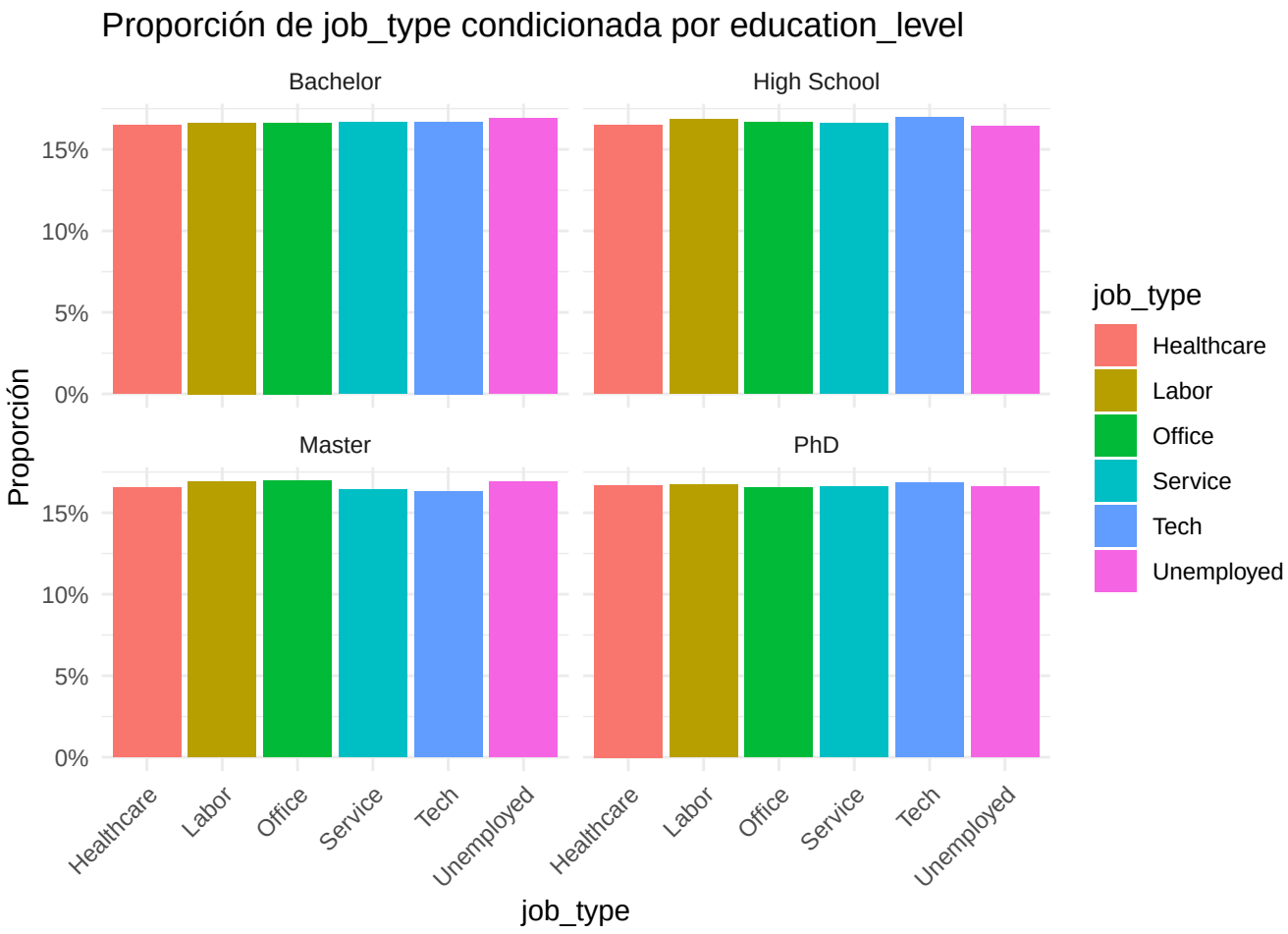
2.2 Análisis bivariado entre variables categóricas

```
# Seleccionar pares de interés (ajustar según narrativa)
pares_cat <- tibble(
  variable_x = c("gender", "education_level", "diet_type"),
  variable_y = c("target", "job_type", "exercise_type")
)

coincidencias <- map2(pares_cat$variable_x, pares_cat$variable_y, ~ {
  x_sym <- rlang::sym(.x)
  y_sym <- rlang::sym(.y)
  tabla <- datos |>
    count(!!x_sym, !!y_sym, name = "n") |>
    group_by(!!x_sym) |>
    mutate(prop = n / sum(n)) |>
    ungroup()
```

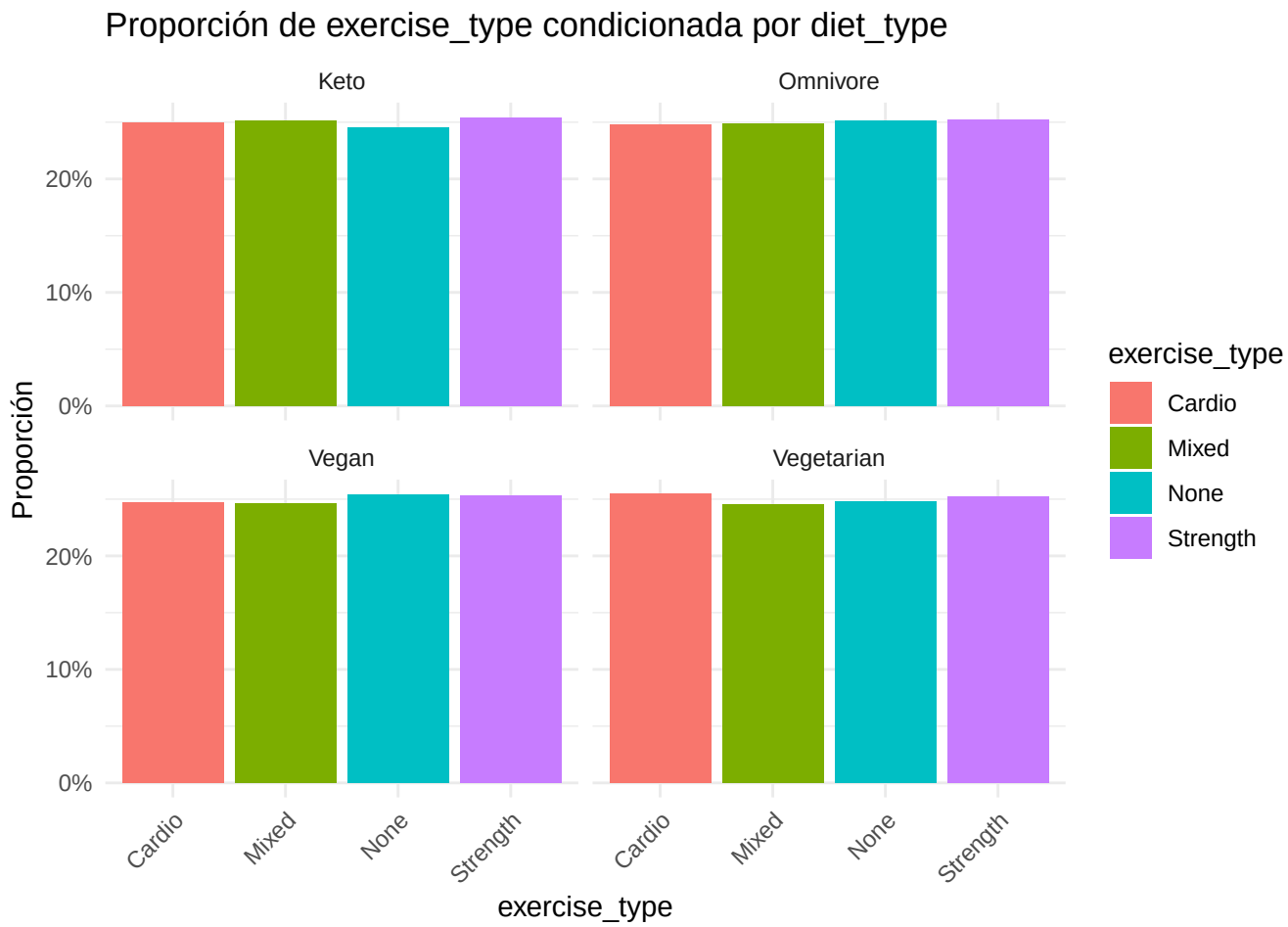
categoria	n	porcentaje
alcohol_consumption		
Occasionally	28831	28.831%
Regularly	28782	28.782%
None	28477	28.477%
NA	13910	13.910%
gender		
Male	50132	50.13%
Female	49868	49.87%
sleep_quality		
Good	25147	25.147%
Excellent	25091	25.091%
Fair	25008	25.008%
Poor	24754	24.754%

```
  grafico <- tabla |>
    ggplot(aes(x = !!y_sym, y = prop, fill = !!y_sym)) +
    geom_col() +
    scale_y_continuous(labels = scales::percent_format()) +
    labs(
      title = paste("Proporción de", .y, "condicionada por", .x),
      x = .y,
      y = "Proporción",
      fill = .y
    ) +
    facet_wrap(vars(!!x_sym)) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
  list(
    descripcion = paste(.x, "vs", .y),
    tabla = tabla,
    grafico = grafico
  )
})

# Mostrar gráficos de ejemplo
walk(coincidencias, ~ print(.x$grafico))
```


Proporción de target condicionada por gender
Female
Male
60%
40%
20%
0%
Proporción
diseased
healthy
diseased
healthy
target
target
diseased
healthy


Proporción de job_type condicionada por education_level
Bachelor
High School
Master
PhD
Proporción
15%
10%
5%
0%
Healthcare
Labor
Office
Service
Tech
Unemployed
job_type
job_type
Healthcare
Labor
Office
Service
Tech
Unemployed

gender	target	n	prop
Female	diseased	14933	0.2994505
Female	healthy	34935	0.7005495
Male	diseased	14970	0.2986117
Male	healthy	35162	0.7013883




```
# Tabla de frecuencias para el primer par
coincidencias[[1]]$tabla |>
  gt::gt()
```

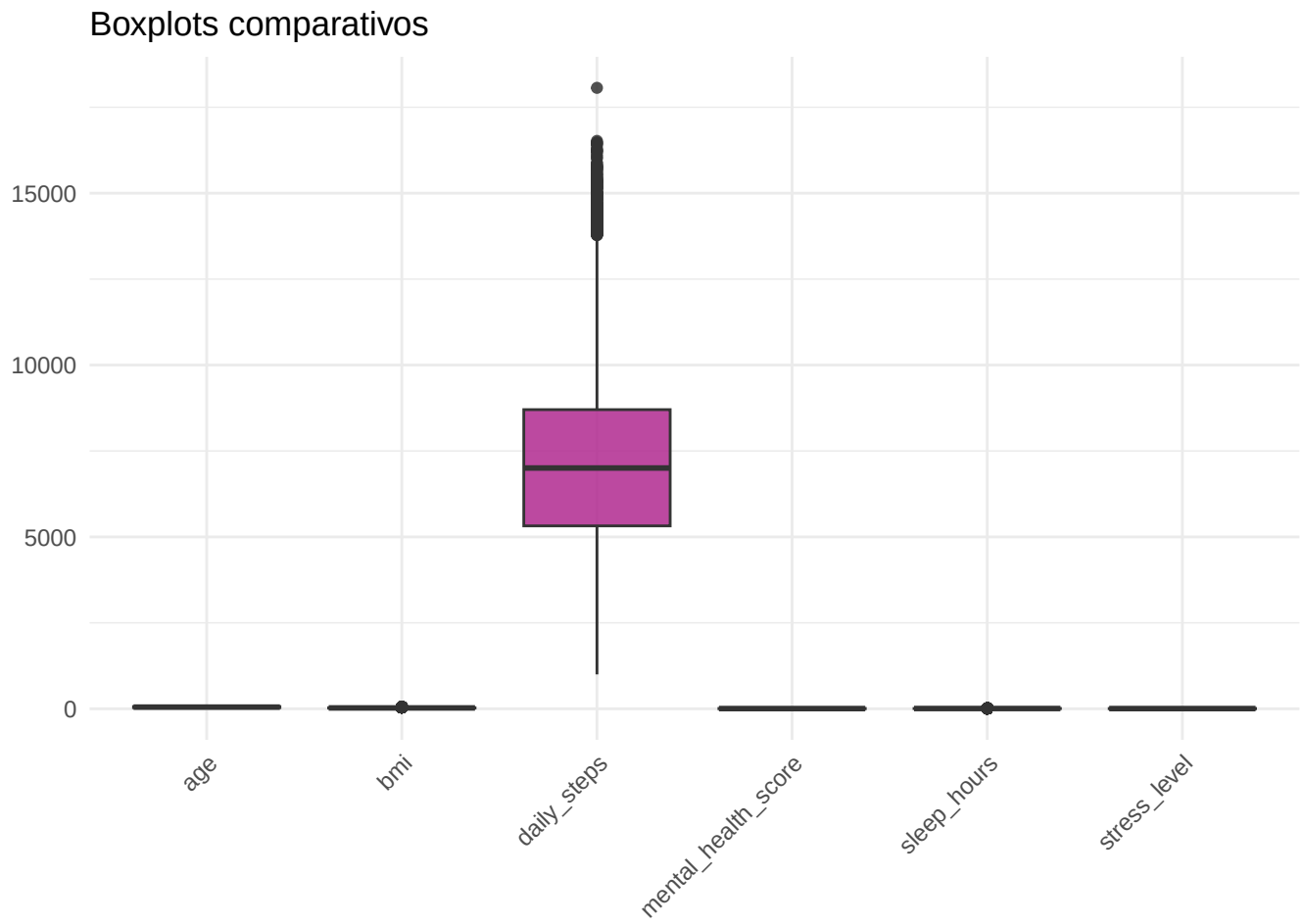
3. Análisis descriptivo de variables cuantitativas

3.1 Análisis univariado

```
cuant_vars <- metadata |>
  filter(tipo %in% c("numeric", "integer")) |>
  pull(variable)

skewness <- function(x) {
  x <- x[!is.na(x)]
```

```r
  if (length(x) < 3 || sd(x) == 0) return(NA_real_)
  mean(((x - mean(x)) / sd(x))^3)
}

kurtosis_excess <- function(x) {
  x <- x[!is.na(x)]
  if (length(x) < 4 || sd(x) == 0) return(NA_real_)
  mean(((x - mean(x)) / sd(x))^4) - 3
}

resumen_cuant <- datos |>
  select(all_of(cuant_vars)) |>
  pivot_longer(everything(), names_to = "variable", values_to = "valor") |>
  group_by(variable) |>
  summarise(
    media = mean(valor, na.rm = TRUE),
    mediana = median(valor, na.rm = TRUE),
    desviacion = sd(valor, na.rm = TRUE),
    q1 = quantile(valor, 0.25, na.rm = TRUE),
    q3 = quantile(valor, 0.75, na.rm = TRUE),
    minimo = min(valor, na.rm = TRUE),
    maximo = max(valor, na.rm = TRUE),
    asimetria = skewness(valor),
    curtosis = kurtosis_excess(valor),
    faltantes = sum(is.na(valor)),
    .groups = "drop"
  )

resumen_cuant |>
  mutate(across(where(is.numeric), round, 3)) |>
  gt::gt()
```

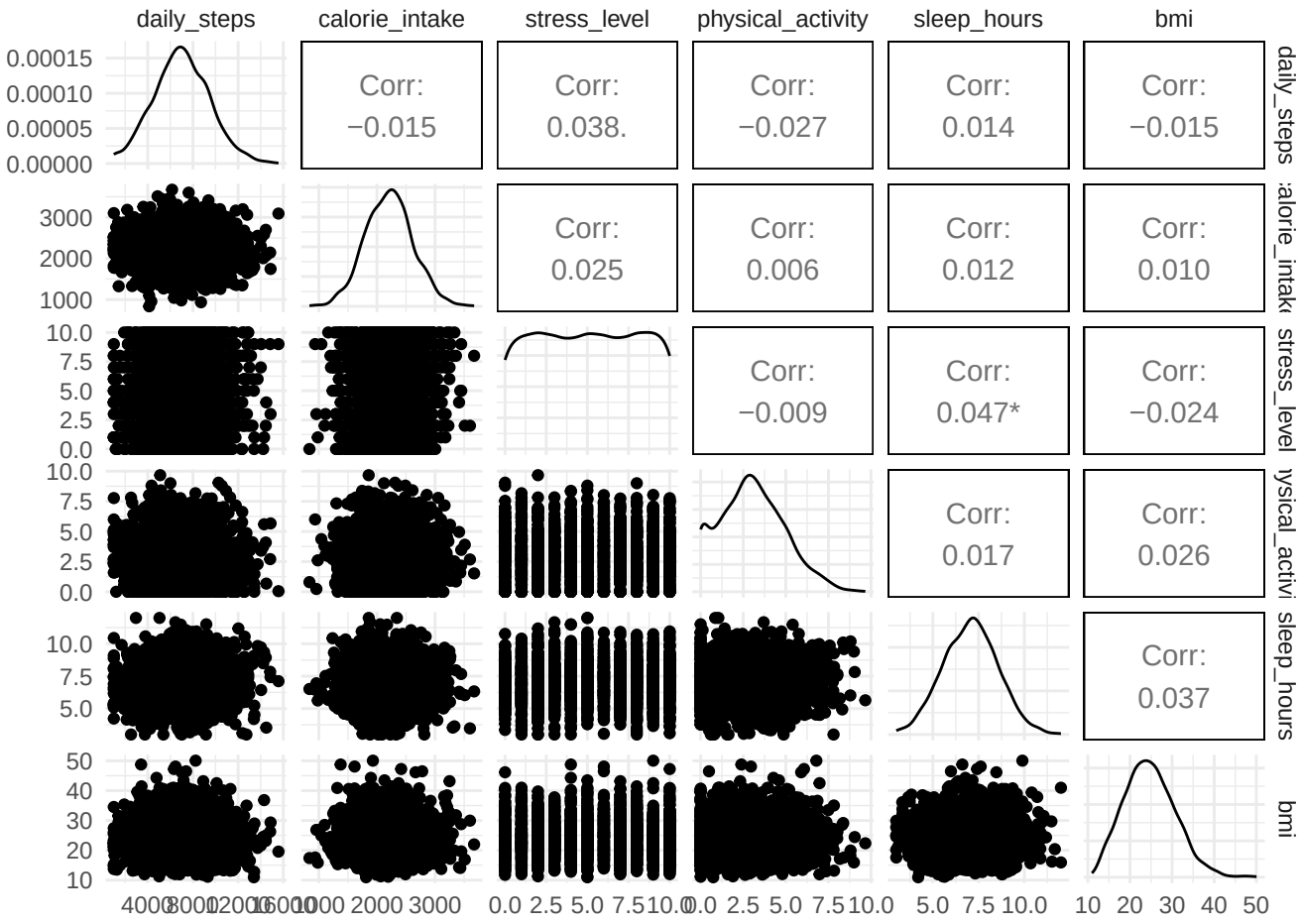
```r
vars_para_hist <- c("age", "bmi", "sleep_hours", "daily_steps", "stress_level", "mental_health_score")

# Histogramas
if (all(vars_para_hist %in% names(datos))) {
  datos |>
    select(all_of(vars_para_hist)) |>
    pivot_longer(everything(), names_to = "variable", values_to = "valor") |>
    ggplot(aes(x = valor)) +
    geom_histogram(fill = "#2a9d8f", color = "white", bins = 30) +
    facet_wrap(~ variable, scales = "free") +
    labs(title = "Distribuciones univariadas de variables cuantitativas", x = NULL, y = "Frecuencia")

  datos |>
    select(all_of(vars_para_hist)) |>
    pivot_longer(everything(), names_to = "variable", values_to = "valor") |>
    ggplot(aes(x = variable, y = valor, fill = variable)) +
    geom_boxplot(alpha = 0.85) +
    scale_fill_viridis_d(option = "C", guide = "none") +
    labs(x = NULL, y = NULL, title = "Boxplots comparativos") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}
```



3.2 Análisis bivariado cuantitativo

```
pares_cuant <- datos |>
  select(daily_steps, calorie_intake, stress_level, physical_activity, sleep_hours, bmi)

cor_matrix <- round(cor(pares_cuant, use = "pairwise.complete.obs"), 2)

cor_matrix |>
  as.data.frame() |>
  rownames_to_column("variable") |>
  gt::gt()
```

```
set.seed(123)
muestra_cuant <- datos |>
  select(daily_steps, calorie_intake, stress_level, physical_activity, sleep_hours, bmi) |>
  drop_na() |>
  slice_sample(n = 2000)

GGally::ggpairs(muestra_cuant)
```



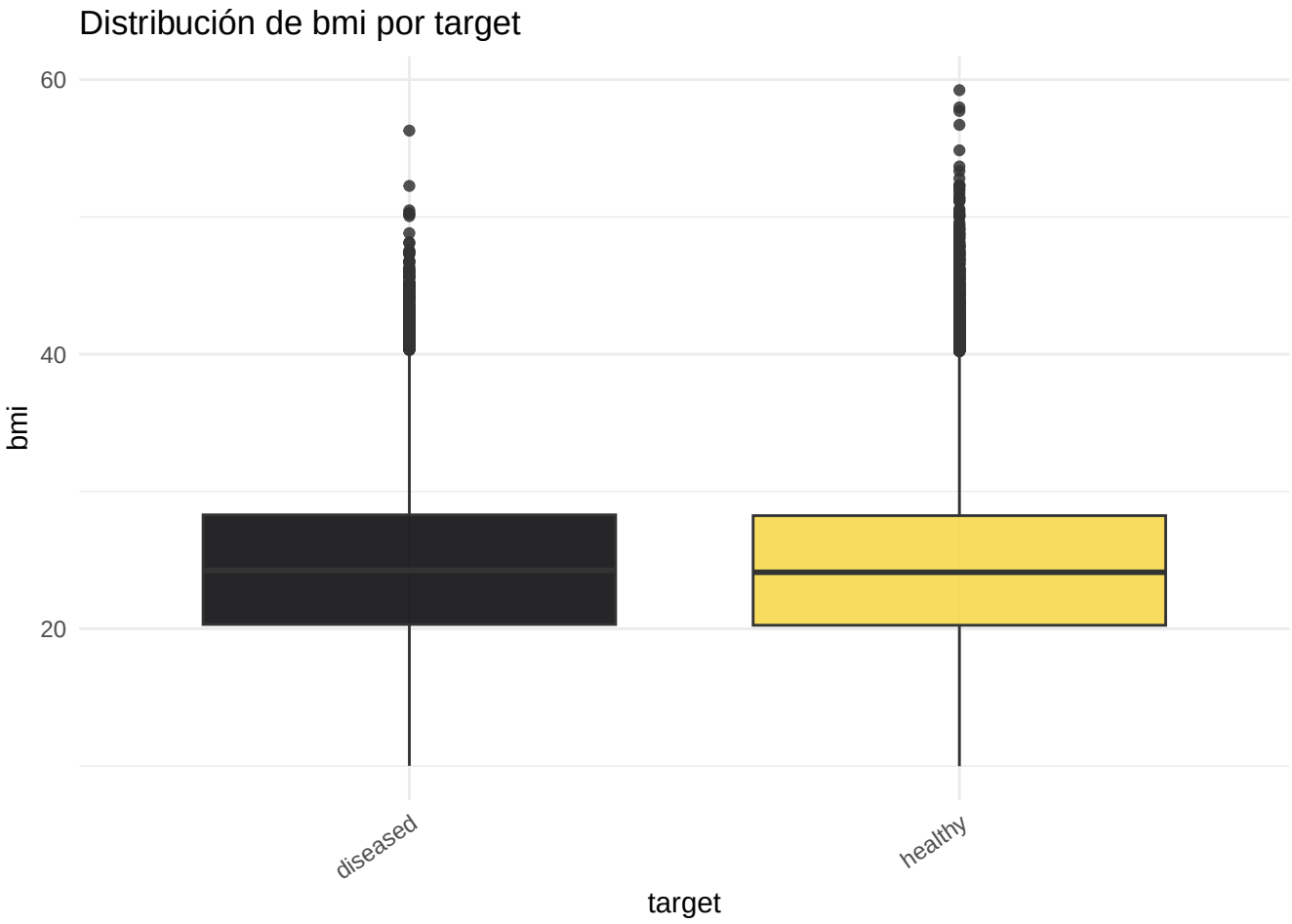
Sugerencia analítica: describa si las variables seleccionadas se ajustan a distribuciones conocidas (normal, log-normal, Poisson, etc.) en función de los gráficos, medidas de asimetría y curtosis.

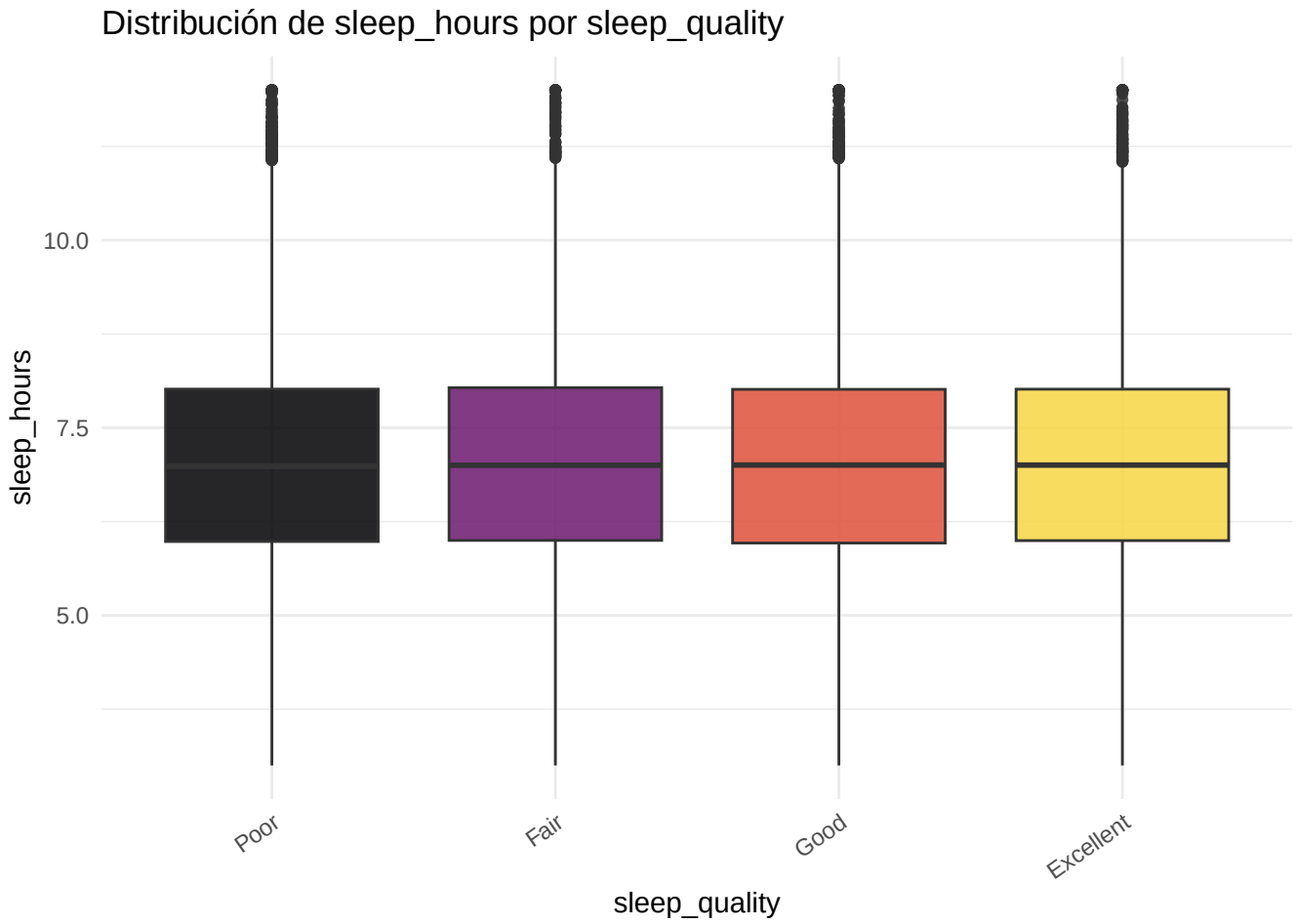
4. Relación entre variables categóricas y cuantitativas

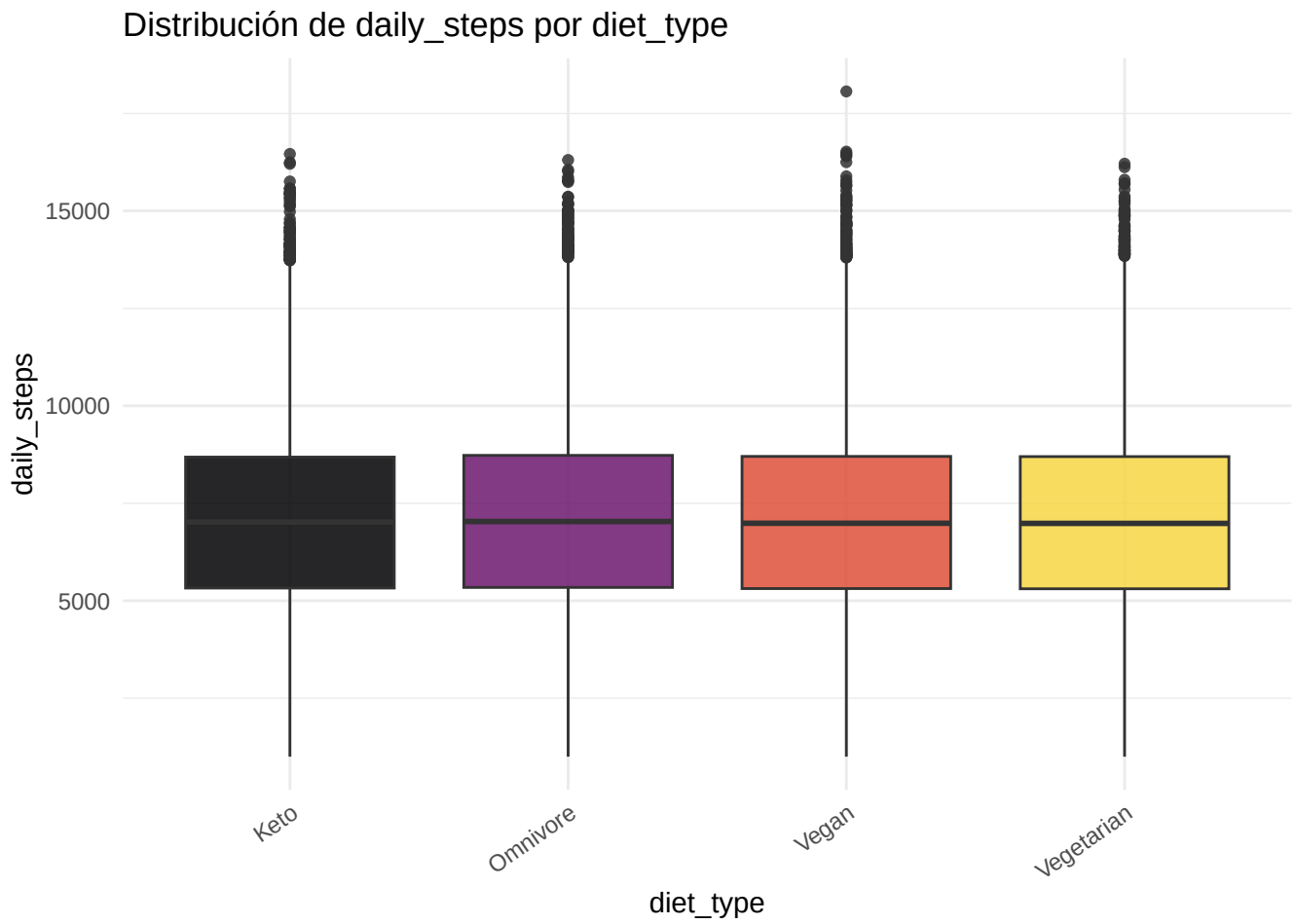
```
ejemplos_cat <- c("target", "sleep_quality", "diet_type", "smoking_level")
cuant_vars_interes <- c("bmi", "sleep_hours", "daily_steps", "stress_level")

resumen_cat_cuant <- map_dfr(ejemplos_cat, function(cat_var) {
  cat_sym <- rlang::sym(cat_var)
  datos |>
    select(!!cat_sym, all_of(cuant_vars_interes)) |>
    pivot_longer(cols = all_of(cuant_vars_interes), names_to = "variable_cuant", values_to = "valor") |
    group_by(grupo = !!cat_sym, variable_cuant) |>
    summarise(
      n = n(),
      media = mean(valor, na.rm = TRUE),
      mediana = median(valor, na.rm = TRUE),
      sd = sd(valor, na.rm = TRUE),
      q1 = quantile(valor, 0.25, na.rm = TRUE),
      q3 = quantile(valor, 0.75, na.rm = TRUE),
      .groups = "drop"
    ) |>
```

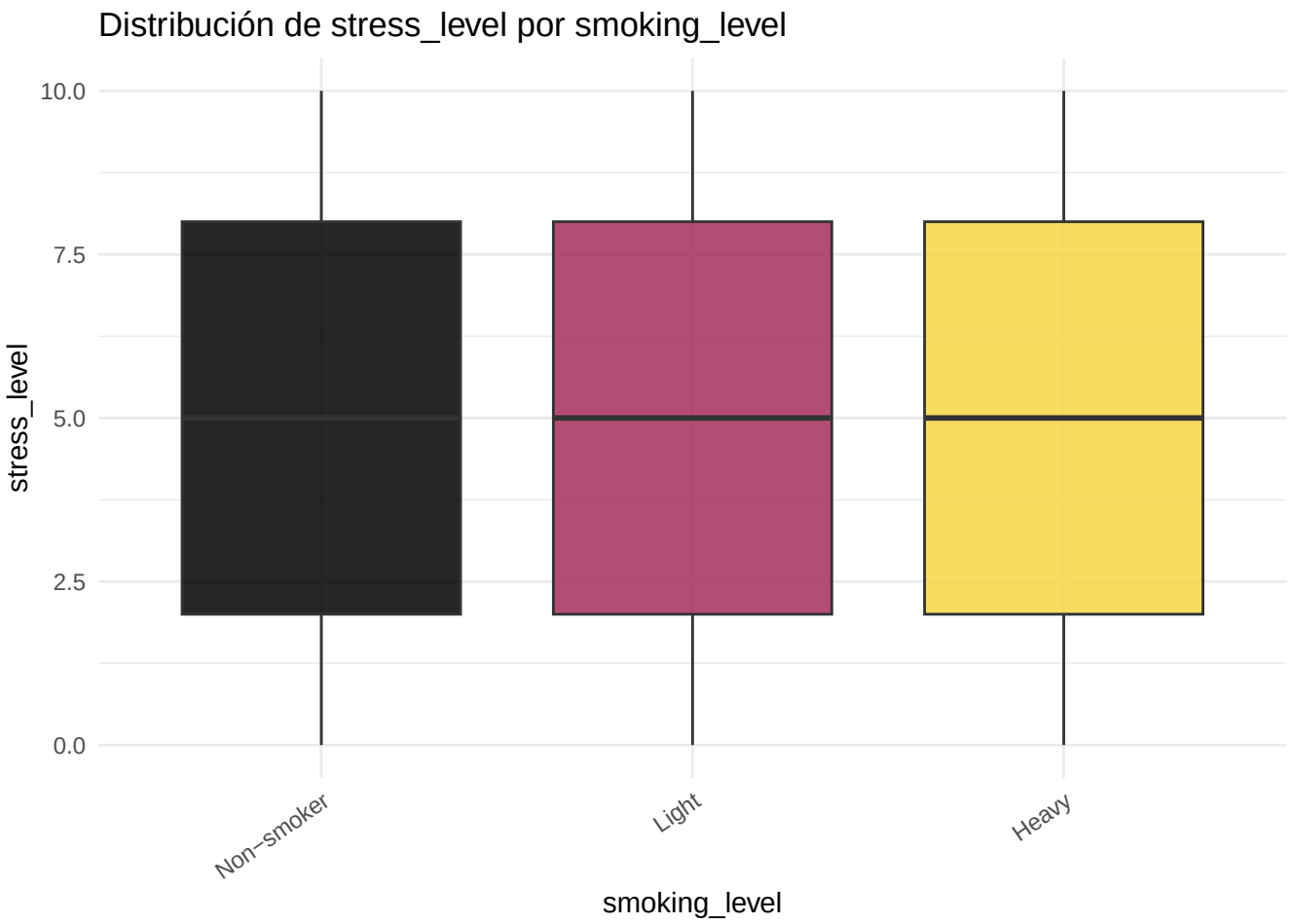
```
    mutate(grupo = as.character(grupo)) |>
    mutate(variable_categorica = cat_var)
})

resumen_cat_cuant |>
  mutate(across(where(is.numeric), round, 3)) |>
  gt::gt(groupname_col = "variable_categorica")
```

```
pares_cat_cuant <- tribble(
  ~cat, ~cuant,
  "target", "bmi",
  "sleep_quality", "sleep_hours",
  "diet_type", "daily_steps",
  "smoking_level", "stress_level"
)

graficos_cat_cuant <- map2(pares_cat_cuant$cat, pares_cat_cuant$cuant, function(cat_var, cuant_var) {
  cat_sym <- rlang::sym(cat_var)
  cuant_sym <- rlang::sym(cuant_var)
  datos |>
    ggplot(aes(x = !!cat_sym, y = !!cuant_sym, fill = !!cat_sym)) +
    geom_boxplot(alpha = 0.85, outlier_alpha = 0.25) +
    scale_fill_viridis_d(option = "B", end = 0.9, guide = "none") +
    labs(
      title = paste("Distribución de", cuant_var, "por", cat_var),
      x = cat_var,
      y = cuant_var
    ) +
    theme(axis.text.x = element_text(angle = 35, hjust = 1))
})

walk(graficos_cat_cuant, print)
```


Distribución de bmi por target
60
40
20
bmi
diseased
healthy
target


Distribución de sleep_hours por sleep_quality
sleep_hours
10.0
7.5
5.0
Poor
Fair
Good
Excellent
sleep_quality


Distribución de daily_steps por diet_type
daily_steps
15000
10000
5000
Keto
Omnivore
Vegan
Vegetarian
diet_type



5. Pruebas de hipótesis

```
# 5.1 Prueba para la media: horas de sueño promedio vs. referencia de 7 horas
prueba_media <- t.test(datos$sleep_hours, mu = 7)
resultado_media <- broom::tidy(prueba_media)

resultado_media |>
  mutate(across(where(is.numeric), round, 4)) |>
  gt::gt()
```

```
# 5.2 Prueba para una proporción: proporción de participantes "healthy" vs. 70%
conteo_target <- datos |>
  summarise(
    exito = sum(target == "healthy", na.rm = TRUE),
    total = n()
  )

prueba_prop <- prop.test(x = conteo_target$exito, n = conteo_target$total, p = 0.70, alternative = "two
resultado_prop <- broom::tidy(prueba_prop)

resultado_prop |>
  mutate(across(where(is.numeric), round, 4)) |>
  gt::gt()
```

Nota: Interprete cada prueba definiendo claramente las hipótesis nula y alternativa, revisando supuestos (normalidad, tamaño muestral, etc.) y contextualizando los hallazgos con la narrativa del informe.

6. Anexo: Diccionario de variables

```
diccionario <- metadata |>
  mutate(
    tipo_variable = case_when(
      tipo %in% c("numeric", "integer") ~ "Cuantitativa",
      tipo %in% c("factor", "ordered") ~ "Categórica",
      TRUE ~ tipo
    ),
    valores_unicos = map_int(datos[variable], ~ dplyr::n_distinct(.x, na.rm = TRUE))
  ) |>
  select(variable, tipo_variable, valores_unicos, descripcion)

diccionario |>
  gt::gt() |>
  gt::tab_spanner(label = "Completar las descripciones antes de la entrega", columns = everything())
```

```
# Resumen de valores faltantes para complementar la discusión metodológica
faltantes <- datos |>
  summarise(across(everything(), ~ sum(is.na(.)))) |>
  pivot_longer(everything(), names_to = "variable", values_to = "faltantes") |>
  mutate(pct = faltantes / nrow(datos)) |>
  arrange(desc(faltantes))

faltantes |>
  filter(faltantes > 0) |>
  mutate(pct = scales::percent(pct)) |>
  slice_head(n = 15) |>
  gt::gt()
```

> **Recordatorio para el informe:** redacte conclusiones y recomendaciones alineadas con los hallazgos clave. Asegúrese de que las tablas y figuras seleccionadas estén referenciadas en el documento PDF y que incluyan títulos, fuentes y notas interpretativas cuando sea necesario.

variable	media	mediana	desviacion	q1	q3	minimo	m
age	48.526	48.000	17.887	33.000	64.000	18.000	
blood_pressure	119.980	119.952	15.016	109.812	130.121	59.128	18
bmi	24.494	24.157	5.951	20.271	28.259	9.988	
bmi_corrected	24.494	24.152	5.954	20.271	28.248	9.894	
bmi_estimated	24.494	24.157	5.951	20.271	28.259	9.988	
bmi_scaled	73.482	72.470	17.853	60.814	84.776	29.965	1
calorie_intake	2201.429	2200.993	400.516	1932.278	2471.218	527.172	39
cholesterol	189.966	190.045	29.982	169.668	210.222	58.411	3
daily_steps	7012.926	7004.285	2488.989	5320.858	8702.281	1000.000	180
daily_supplement_dosage	0.016	0.016	5.764	-4.981	5.008	-10.000	
electrolyte_level	0.000	0.000	0.000	0.000	0.000	0.000	
environmental_risk_score	5.500	5.500	0.000	5.500	5.500	5.500	
gene_marker_flag	1.000	1.000	0.000	1.000	1.000	1.000	
glucose	99.995	99.987	19.983	86.461	113.509	12.435	18
heart_rate	74.969	75.046	9.942	68.275	81.686	34.745	1
height	170.024	170.017	9.983	163.307	176.729	140.000	2
income	4038.127	4004.601	1930.026	2665.403	5360.013	500.000	120
insulin	14.988	14.983	5.003	11.627	18.362	-6.794	
meals_per_day	2.999	3.000	1.415	2.000	4.000	1.000	
mental_health_score	5.005	5.000	3.164	2.000	8.000	0.000	
physical_activity	3.038	2.971	1.884	1.634	4.327	0.000	
screen_time	6.022	5.991	2.934	3.971	8.024	0.000	
sleep_hours	7.002	6.998	1.497	5.987	8.019	3.000	
stress_level	4.992	5.000	3.155	2.000	8.000	0.000	
sugar_intake	60.047	60.048	19.967	46.504	73.476	-27.882	1
survey_code	50000.500	50000.500	28867.658	25000.750	75000.250	1.000	1000
waist_size	84.933	84.957	12.040	76.795	93.019	34.093	1
water_intake	2.006	2.001	0.689	1.532	2.473	0.500	
weight	70.065	69.924	14.694	59.857	80.027	40.000	1
work_hours	8.001	8.005	1.995	6.651	9.354	0.000	

variable	daily_steps	calorie_intake	stress_level	physical_activity	sleep_hours	bmi
daily_steps	1.00	0.00	0	0	0.00	-0.01
calorie_intake	0.00	1.00	0	0	0.00	0.01
stress_level	0.00	0.00	1	0	0.00	0.00
physical_activity	0.00	0.00	0	1	0.00	0.00
sleep_hours	0.00	0.00	0	0	1.00	0.01
bmi	-0.01	0.01	0	0	0.01	1.00

grupo	variable_cuant	n	media	mediana	sd	q1	q3
target							
diseased	bmi	29903	24.517	24.257	5.930	20.311	28.296
diseased	daily_steps	29903	6993.622	6985.091	2486.166	5313.425	8671.585
diseased	sleep_hours	29903	7.006	7.005	1.494	5.996	8.015
diseased	stress_level	29903	4.988	5.000	3.153	2.000	8.000
healthy	bmi	70097	24.484	24.112	5.960	20.252	28.238
healthy	daily_steps	70097	7021.124	7011.261	2490.161	5322.765	8713.419
healthy	sleep_hours	70097	7.000	6.996	1.498	5.983	8.021
healthy	stress_level	70097	4.993	5.000	3.156	2.000	8.000
sleep_quality							
Poor	bmi	24754	24.528	24.166	5.990	20.297	28.335
Poor	daily_steps	24754	7015.323	7004.389	2473.506	5349.309	8676.502
Poor	sleep_hours	24754	7.000	6.986	1.491	5.984	8.015
Poor	stress_level	24754	4.999	5.000	3.150	2.000	8.000
Fair	bmi	25008	24.484	24.149	5.903	20.303	28.244
Fair	daily_steps	25008	7016.942	6996.000	2487.034	5333.208	8701.300
Fair	sleep_hours	25008	7.007	7.001	1.502	5.999	8.034
Fair	stress_level	25008	4.994	5.000	3.154	2.000	8.000
Good	bmi	25147	24.458	24.157	5.977	20.190	28.213
Good	daily_steps	25147	7021.017	7016.649	2499.082	5314.977	8728.284
Good	sleep_hours	25147	6.993	7.003	1.502	5.965	8.012
Good	stress_level	25147	5.008	5.000	3.157	2.000	8.000
Excellent	bmi	25091	24.505	24.155	5.933	20.293	28.236
Excellent	daily_steps	25091	6998.442	7001.103	2496.111	5282.804	8701.151
Excellent	sleep_hours	25091	7.007	7.003	1.492	5.996	8.014
Excellent	stress_level	25091	4.965	5.000	3.159	2.000	8.000
diet_type							
Keto	bmi	24764	24.485	24.149	5.921	20.296	28.237
Keto	daily_steps	24764	7007.375	7017.225	2477.402	5325.319	8684.420
Keto	sleep_hours	24764	7.025	7.020	1.497	6.005	8.035
Keto	stress_level	24764	5.018	5.000	3.150	2.000	8.000
Omnivore	bmi	25089	24.456	24.130	5.973	20.173	28.214
Omnivore	daily_steps	25089	7026.565	7032.163	2494.846	5341.007	8728.405
Omnivore	sleep_hours	25089	6.991	6.984	1.493	5.975	8.004
Omnivore	stress_level	25089	4.978	5.000	3.150	2.000	8.000
Vegan	bmi	25122	24.564	24.222	5.986	20.304	28.387
Vegan	daily_steps	25122	7014.132	6987.743	2497.552	5312.066	8701.750
Vegan	sleep_hours	25122	6.983	6.979	1.499	5.969	8.014
Vegan	stress_level	25122	4.968	5.000	3.152	2.000	8.000
Vegetarian	bmi	25025	24.470	24.122	5.923	20.321	28.192
Vegetarian	daily_steps	25025	7003.510	6984.109	2486.051	5306.306	8697.439
Vegetarian	sleep_hours	25025	7.010	7.006	1.499	6.000	8.027
Vegetarian	stress_level	25025	5.002	5.000	3.168	2.000	8.000
smoking_level							
Non-smoker	bmi	33355	24.507	24.156	5.973	20.265	28.298
Non-smoker	daily_steps	33355	7010.223	6990.753	2499.472	5310.683	8703.114

estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
7.002	0.4243	0.6714	99999	6.9927	7.0113	One Sample t-test	two.sided

estimate	statistic	p.value	parameter	conf.low	conf.high	method
0.701	0.4434	0.5055	1	0.6981	0.7038	1-sample proportions test with continuity

Completar las descripciones antes de la entrega			
variable	tipo_variable	valores_unicos	descripcion
survey_code	Cuantitativa	100000	NA
age	Cuantitativa	62	NA
gender	Categórica	2	NA
height	Cuantitativa	99843	NA
weight	Cuantitativa	97703	NA
bmi	Cuantitativa	99996	NA
bmi_estimated	Cuantitativa	99996	NA
bmi_scaled	Cuantitativa	99996	NA
bmi_corrected	Cuantitativa	100000	NA
waist_size	Cuantitativa	100000	NA
blood_pressure	Cuantitativa	92331	NA
heart_rate	Cuantitativa	85997	NA
cholesterol	Cuantitativa	100000	NA
glucose	Cuantitativa	100000	NA
insulin	Cuantitativa	84164	NA
sleep_hours	Cuantitativa	99599	NA
sleep_quality	Categórica	4	NA
work_hours	Cuantitativa	99997	NA
physical_activity	Cuantitativa	93167	NA
daily_steps	Cuantitativa	90934	NA
calorie_intake	Cuantitativa	100000	NA
sugar_intake	Cuantitativa	100000	NA
alcohol_consumption	Categórica	3	NA
smoking_level	Categórica	3	NA
water_intake	Cuantitativa	98489	NA
screen_time	Cuantitativa	97730	NA
stress_level	Cuantitativa	11	NA
mental_health_score	Cuantitativa	11	NA
mental_health_support	Categórica	2	NA
education_level	Categórica	4	NA
job_type	Categórica	6	NA
occupation	Categórica	6	NA
income	Cuantitativa	87872	NA
diet_type	Categórica	4	NA
exercise_type	Categórica	4	NA
device_usage	Categórica	3	NA
healthcare_access	Categórica	3	NA
insurance	Categórica	2	NA
sunlight_exposure	Categórica	3	NA
meals_per_day	Cuantitativa	5	NA
caffeine_intake	Categórica	3	NA
family_history	Categórica	2	NA
pet_owner	Categórica	2	NA
electrolyte_level	Cuantitativa	1	NA
gene_marker_flag	Cuantitativa	1	NA
environmental_risk_score	Cuantitativa	1	NA
daily_supplement_dosage	Cuantitativa	100000	NA

variable	faltantes	pct
insulin	15836	15.836%
heart_rate	14003	14.003%
alcohol_consumption	13910	13.910%
gene_marker_flag	10474	10.474%
income	8470	8.470%
daily_steps	8329	8.329%
blood_pressure	7669	7.669%