

CONSTRUCCIÓN DE UNA FUNCIÓN EN R

Carla Rojas – Fernando López

Table of contents

1	Ojetivo	1
1.1	Motivación	1
1.2	Desarrollo de la función:	3
2	Uso de los resultados de la función	8
3	Conclusión Final	31

1 Ojetivo

Crear una función en R Studio de mediana a alta complejidad

1.1 Motivación

El manejo de bases de datos con gran cantidad de observaciones, requiere de gran tiempo de procesamiento de datos y el uso de distintos software para optimizar el tiempo de analisis y de respuesta a un trabajo requerido. Por esto el uso del software R y y su plataforma R Studio, al poder programar una función que resuelva la necesidad de optimizar el tiempo de respuesta y análisis en la actualidad, es de gran ayuda para alcanzar este objetivo.

En este caso nos encontramos con una base de datos que posee registros de hospitalizaciones en el servidio de psiquiatría del hospital regional de concepcion durante los años 2018 al 2023, con más de 3000 observaciones durante este periodo de tiempo, con información valiosa como la edad, el género, el tiempo de hospitalización y los diagnóstico entre otros. La necesidad de poder clasificar estos diagnósticos y poder realizar un análisis exploratorio de datos es que nos lleva a programar la siguiente funcion:

1.1.1 Filtrado y generación de tablas de datos

Lo que se pretende es que la función sea capaz de filtrar la base de datos creando nuevos datasets más pequeños y de fácil manejo. Para filtrar se toman en referencia las patologías de salud mental contenidas en el GES en sus respectivas guías clínicas, lográndose identificar 5 grupos:

- Depresión
- Trastorno Afectivo Bipolar
- Esquizofrenia
- Consumo de Alcohol y Droga
- Alzheimer

Luego de una inspección manual de la base de datos, se identificaron los siguientes patrones para la creación de estos grupos, además consiguiendo identificar un sexto grupo de patologías no clasificadas dentro de los grupos anteriores, los parámetros son:

- **Depresión:**
 - Edad mayor o igual a 15 años,
 - Que dentro de la variable “Diagnostico” se encuentren las palabras “depresion”, “depresivo”, “depresivos”
 - Que se excluyan las palabras “antidepresivo”, “antidepresivos”
- **Trastorno Afectivo Bipolar:**
 - Edad mayor o igual a 15 años,
 - Que dentro de la variable “Diagnostico” se encuentren las palabras “mania”, “bipolar”
- **Esquizofrenia:**
 - Que dentro de la variable “Diagnostico” se encuentren las palabras “esquizofrenia”, “Esquizofrenias”, “esquizaafectivo”, “esquizaafectivos”
- **Consumo de Alcohol y Drogas:**
 - Que dentro de la variable “Diagnostico” se encuentren las palabras “alcohol”, “drogas”, “dependencia”, “nocivo”, “cannabinoides”, “sedantes”, “estimulantes”
- **Otras Patologías:**
 - Que se agrupen todas las patologías que hayan quedado fuera de la clasificación anterior.

1.2 Desarrollo de la función:

1.2.0.1 1. Carga de paquetes e instalación automática

Antes de todo, vamos a cargar las librerías que nos ayudarán a ejecutar nuestra función, si estas no estan instaladas, las instalaremos.

```
# Instalación automática de paquetes
paquetes <- c("readxl", "dplyr", "stringr", "purrr", "ggplot2", "writexl", "assertthat")

instalar_si_falta <- function(pkg) {
  if (!requireNamespace(pkg, quietly = TRUE)) {
    install.packages(pkg)
  }
}
invisible(lapply(paquetes, instalar_si_falta))

# Carga de paquetes
library(readxl) # para leer los archivos excel
library(dplyr) # para manipular y transformar los datos
```

Adjuntando el paquete: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(stringr) # para la búsqueda de palabras claves
library(purrr) # para trabajar con listas o aplicar funciones a múltiples objetos
library(ggplot2) # para crear los gráficos
library(writexl) # para guardar los nuevos archivos creados
```

Warning: package 'writexl' was built under R version 4.5.1

```
library(assertthat) # para evitar errores por datos o columnas incompletas
```

Warning: package 'assertthat' was built under R version 4.5.1

1.2.0.2 2. Validación de columnas esperadas

Se revisan las listas, columnas y datos antes de iniciar el filtrado de la base de datos.

```
validar_datos <- function(data) {  
  required_cols <- c("EDAD", "DIAGNOSTICO_EGRESO", "AÑO_EGRESO", "SEXO", "DIAS_HOSPITALI  
  for (col in required_cols) {  
    assert_that(col %in% names(data), msg = paste("Falta la columna:", col))  
  }  
}
```

1.2.0.3 3. Función auxiliar para filtrado

Esta función previa nos va a permitir reutilizar la función cada vez que se analice la base de datos sin tener que reescribir todo por cada grupo a filtrar, así mantenemos el orden y claridad del código evitando errores.

```
# Crear función que automatice la función principal  
filtrar_diagnostico <- function(data, palabras_clave, edad_minima = 0) {  
  data %>%  
    filter(  
      EDAD >= edad_minima,  
      str_detect(tolower(DIAGNOSTICO_EGRESO), str_c(palabras_clave, collapse = "|"))  
    )  
}
```

1.2.0.4 4. Función principal.

Desarrollo de la función principal, aquí vamos a incluir todas las observaciones que se realizaron al inicio para cada grupo a crear, vamos a establecer los límites de edad y la búsqueda de las palabras claves definidas para cada grupo, también se generará un grupo que reúne a todas las patologías diagnosticadas que no entran en las clasificaciones anteriores.

```
analizar_psiquiatria <- function(data) {  
  validar_datos(data)  
  
  data <- data %>%  
    mutate(  
      PERIODO = case_when(  
        AÑO_EGRESO %in% 2018:2019 ~ "prepandemia",  
        AÑO_EGRESO %in% 2020:2022 ~ "pandemia",  
        AÑO_EGRESO == 2023 ~ "postpandemia",  
        TRUE ~ "otro"  
      )  
    )  
}
```

```

)

grupos <- list(

  # configuracion filtro para depresion
  depresion = filtrar_diagnostico(data, c("depresion", "depresivo", "depresivos"), 15),

  # configuracion filtro para trastorno bipolar
  trastorno_bipolar = filtrar_diagnostico(data, c("mania", "bipolar"), 15),

  # configuracion filtro para esquizofrenia
  esquizofrenia = filtrar_diagnostico(data, c("esquizofrenia", "esquizofrenias", "esqu

  # configuracion filtro para drogas y alcohol
  consumo = filtrar_diagnostico(data, c("alcohol", "drogas", "dependencia", "nocivo",

  # configuracion filtro para alzheimer
  alzheimer = filtrar_diagnostico(data, c("alzheimer"))
)

# Generar grupo "No GES"
otros_dg_psq <- data %>%
  filter(!row_number() %in% unlist(map(grupos, rownames)))

grupos$otros_dg_psq <- otros_dg_psq

# Guardar datasets
walk2(grupos, names(grupos), ~ write_xlsx(.x, paste0("grupo_", .y, ".xlsx")))

# Crear resumen estadístico
resumen_grupo <- function(df, nombre) {
  if (nrow(df) == 0) return(tibble(Grupo = nombre, Casos = 0))
  df %>%
    summarise(
      Grupo = nombre,
      Casos = n(),
      Edad_Prom = round(mean(EDAD, na.rm = TRUE), 1),
      Dias_Prom = round(mean(DIAS_HOSPITALIZADO, na.rm = TRUE), 1),
      Femenino = sum(SEXO == "FEMENINO", na.rm = TRUE),
      Masculino = sum(SEXO == "MASCULINO", na.rm = TRUE)
    )
}

resultados <- map2_dfr(grupos, names(grupos), resumen_grupo)

```

```

write_xlsx(resultados, "resumen_diagnosticos.xlsx")

return(list(resumen = resultados, grupos = grupos))
}

```

1.2.0.5 5. Función para generar gráficos automáticamente

Vamos a crear una función que nos permita generar graficos necesarios para realizar un análisis exploratorio de datos.

```

graficar_edad_por_grupo <- function(df, nombre_grupo) {
  if (nrow(df) == 0) return(NULL)

  ggplot(df, aes(x = EDAD)) +
    geom_histogram(binwidth = 5, fill = "steelblue", color = "black") +
    labs(title = paste("Distribución de la edad -", nombre_grupo),
         x = "Edad", y = "Frecuencia") +
    theme_minimal()
}

```

1.2.0.6 6. Carga de base de datos y ejecución de la función

Cargamos la base de datos del archivo “PSQ.xlsx” y comenzamos a utilizar nuestra función.

```

# Cargar la base de datos y ejecutar la función principal

# 1. Leer la base desde Excel
datos <- read_excel("PSQ.xlsx")

# 2. Renombrar columnas para compatibilidad con la función, debido a errores al correr la función
datos <- datos %>%
  rename(
    FECHA_INGRESO = INGRESO,
    FECHA_EGRESO = EGRESO,
    AÑO_EGRESO = A_EGRESO,
    DIAS_HOSPITALIZADO = ESTANCIA,
    DIAGNOSTICO_EGRESO = DIAGNOSTICO,
    RANGO_ETAREO = RANGO,
    UNIDAD_EGRESO = UNIDAD
  )

# 3. Ejecutar la función con los datos corregidos
resultado_final <- analizar_psiqatria(datos)

```

```
# 4. Visualizar resumen estadístico
resultado_final$resumen
```

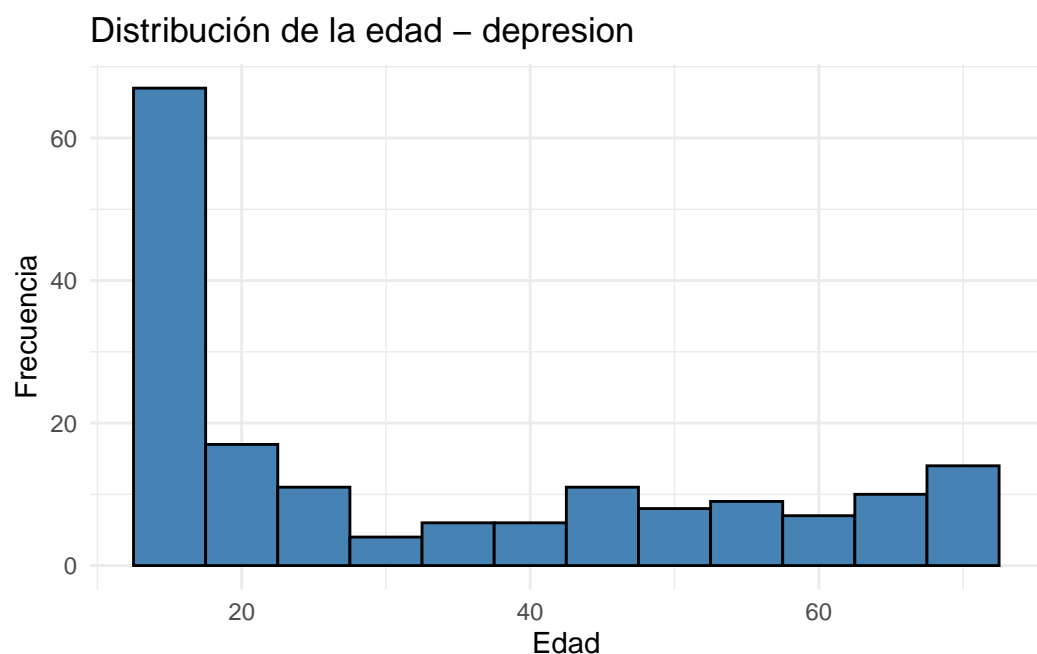
```
# A tibble: 6 x 6
```

	Grupo	Casos	Edad_Prom	Dias_Prom	Femenino	Masculino
	<chr>	<int>	<dbl>	<dbl>	<int>	<int>
1	depression	170	33.3	27.4	107	63
2	trastorno_bipolar	393	38	38	207	186
3	esquizofrenia	372	39.4	105	122	250
4	consumo	567	24.3	33	179	388
5	alzheimer	1	51	538	0	1
6	otros_dg_psq	2669	31.8	65	1262	1407

1.2.0.7 7. Mostrar gráficos por grupo

Se pueba la función con el grupo de depresión como ejemplo.

```
graficos <- map2(resultado_final$grupos, names(resultado_final$grupos), graficar_edad_po
graficos[[1]] # Muestra el gráfico del grupo de depresión como ejemplo
```



1.2.0.8 8. Conclusiones automáticas simples

Se configura una conclusión automática sencilla con el fin de tener una visualización rápida de los resultados, de forma general pensando en investigadores con conocimientos básicos de

estadística y puedan ser usadas como ideas iniciales que puedan decantar en conclusiones más fuertes y desarrolladas.

```
cat("CONCLUSIONES AUTOMÁTICAS\n")
```

CONCLUSIONES AUTOMÁTICAS

```
res <- resultado_final$resumen  
  
# Diagnóstico más frecuente  
mas_casos <- res$Grupo[which.max(res$Casos)]  
cat("- El grupo con más casos es:", mas_casos, "\n")
```

- El grupo con más casos es: otros_dg_psq

```
# Edad promedio mínima y máxima  
cat("- La edad promedio varía entre", min(res$Edad_Prom, na.rm=TRUE),  
      "y", max(res$Edad_Prom, na.rm=TRUE), "años.\n")
```

- La edad promedio varía entre 24.3 y 51 años.

```
# Diferencia por sexo en depresión  
depre <- res %>% filter(Grupo == "depresion")  
if (nrow(depre) > 0 && depre$Femenino > depre$Masculino) {  
  cat("- En depresión hay mayor proporción de mujeres hospitalizadas.\n")  
}
```

- En depresión hay mayor proporción de mujeres hospitalizadas.

2 Uso de los resultados de la función

Una vez obtenidos los nuevos dataset, en total se generaron 6 nuevas bases de datos, se le aplicará un análisis exploratorio de datos a cada tabla, simulando el uso de estas bases de datos independientes, con la posibilidad de exportar cada una de ellas y poder trabajarla por separado por los investigadores.

2.0.1 Análisis por Grupo Diagnóstico

2.0.2 1. Depresión


```
# Cargar tabla de depresión desde el archivo generado
depression <- read_excel("grupo_depresion.xlsx")

# Ver resumen de variables
summary(depression)
```

FECHA_INGRESO		FECHA_EGRESO		AÑO_EGRESO	
Min.	:2017-10-25 00:00:00	Min.	:2018-01-02 00:00:00	Min.	:2018
1st Qu.	:2019-06-15 00:00:00	1st Qu.	:2019-06-20 18:00:00	1st Qu.	:2019
Median	:2021-08-18 00:00:00	Median	:2021-09-04 00:00:00	Median	:2021
Mean	:2021-03-17 13:07:45	Mean	:2021-04-13 23:26:07	Mean	:2021
3rd Qu.	:2022-12-08 18:00:00	3rd Qu.	:2023-01-15 06:00:00	3rd Qu.	:2023
Max.	:2023-12-12 00:00:00	Max.	:2023-12-29 00:00:00	Max.	:2023
DIAS_HOSPITALIZADO	SEXO	EDAD	RANGO_ETAREO		
Min.	: 0.00	Length:170	Min.	:15.00	Length:170
1st Qu.	: 10.00	Class :character	1st Qu.	:16.25	Class :character
Median	: 19.00	Mode :character	Median	:23.00	Mode :character
Mean	: 27.43		Mean	:33.29	
3rd Qu.	: 33.75		3rd Qu.	:49.75	
Max.	:149.00		Max.	:72.00	
UNIDAD_EGRESO	DIAGNOSTICO_EGRESO	PERIODO			
Length:170	Length:170	Length:170			
Class :character	Class :character	Class :character			
Mode :character	Mode :character	Mode :character			

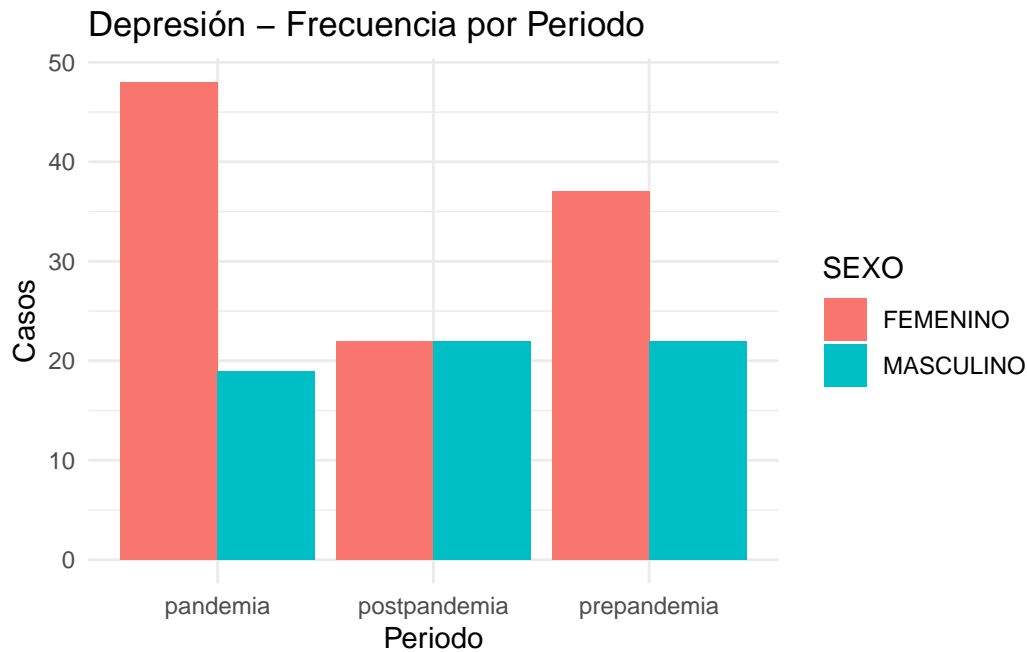
```
# Tabla de frecuencia por sexo
table(depression$SEXO)
```

FEMENINO	MASCULINO
107	63

```
# Tabla por periodo
table(depression$PERIODO)
```

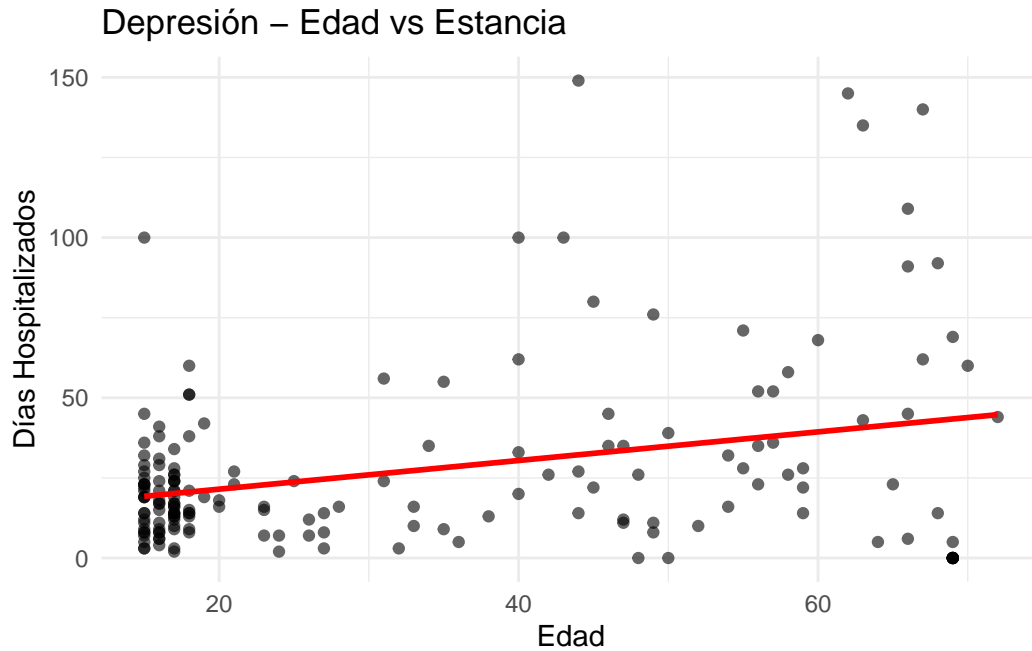
pandemia	postpandemia	prepandemia
67	44	59

```
# Gráfico de barras por periodo
ggplot(depresion, aes(x = PERIODO, fill = SEXO)) +
  geom_bar(position = "dodge") +
  labs(title = "Depresión - Frecuencia por Periodo", x = "Periodo", y = "Casos") +
  theme_minimal()
```



```
# Scatterplot: edad vs días hospitalizado
ggplot(depresion, aes(x = EDAD, y = DIAS_HOSPITALIZADO)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Depresión - Edad vs Estancia", x = "Edad", y = "Días Hospitalizados") +
  theme_minimal()
```

`geom_smooth()` using formula = 'y ~ x'



```
# Regresión lineal
modelo_dep <- lm(DIAS_HOSPITALIZADO ~ EDAD, data = depression)
summary(modelo_dep)
```

Call:

```
lm(formula = DIAS_HOSPITALIZADO ~ EDAD, data = depression)
```

Residuals:

Min	1Q	Median	3Q	Max
-43.389	-13.700	-4.957	5.827	116.785

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	12.5482	4.0759	3.079	0.00243	**
EDAD	0.4470	0.1053	4.243	3.64e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.08 on 168 degrees of freedom

Multiple R-squared: 0.09679, Adjusted R-squared: 0.09142

F-statistic: 18 on 1 and 168 DF, p-value: 3.636e-05

2.0.2.1 Análisis de resultados:

- *Fechas y duración de hospitalización*

- **Período observado:** 2018 a 2023 (aunque hay un caso con ingreso en 2017).
- **Duración promedio:** 27.4 días (máx. 149 días).
- **Mediana de hospitalización:** 19 días → la mayoría permanece de 2 a 4 semanas.

- ***Características demográficas***

- **Edad mínima:** 15 años (criterio aplicado).
- **Edad promedio:** 33.3 años (rango: 15 a 72).
- **Distribución por sexo:** 107 mujeres (63%), 63 hombres (37%)

2.0.2.2 Conclusion:

Mayor proporción de mujeres hospitalizadas por depresión, consistente con la literatura sobre prevalencia de trastornos depresivos.

- ***Distribucion por periodo epidemiológico***

- **Prepandemia (2018–2019):** 59 casos
- **Pandemia (2020–2022):** 67 casos
- **Postpandemia (2023):** 44 casos

2.0.2.3 conclusiones

- Se mantuvo una alta frecuencia durante la pandemia, lo que puede reflejar el impacto psicosocial del confinamiento.
- En 2023 se observa una disminución, pero aún sin volver a los niveles prepandémicos. Faltan datos actuales para evaluar dicha tendencia.
- ***MODELO DE REGRESIÓN: Días de hospitalización según edad***
 - **Ecuación:** días $12.5 + 0.45 \times \text{edad}$
 - **Interpretación del coeficiente:** Por cada año adicional de edad, se incrementan en promedio 0.45 días de hospitalización.
 - **p-value:** $3.6e-05$ → El modelo es estadísticamente significativo.
 - **$R^2 = 0.097$** → La edad explica 9.7% de la variabilidad en los días hospitalizados.

2.0.2.4 conclusiones:

- Aunque existe una relación significativa entre edad y duración de hospitalización, la edad no es un predictor fuerte. Existen otros factores clínicos no incluidos que explican mejor la duración de la hospitalización en el caso de la depresión.

2.0.2.5 Conclusion general del análisis:

En pacientes hospitalizados por depresión entre 2018 y 2023, se observa una mayor proporción de mujeres, con una edad promedio de 33 años. La duración media de hospitalización es de 27 días. El periodo de pandemia concentra la mayor cantidad de casos, lo que refuerza el impacto de la crisis sanitaria en la salud mental. Si bien hay una relación estadísticamente significativa entre la edad y los días de hospitalización, esta relación es débil desde el punto de vista explicativo ($R^2 = 10\%$).

2.0.3 2. Trastorno afectivo bipolar

```
# Cargar tabla de depresión desde el archivo generado
bipolar <- read_excel("grupo_trastorno_bipolar.xlsx")

# Ver resumen de variables
summary(bipolar)
```

FECHA_INGRESO		FECHA_EGRESO		AÑO_EGRESO			
Min.	:2017-08-18 00:00:00	Min.	:2018-01-05 00:00:00	Min.	:2018		
1st Qu.	:2019-05-29 00:00:00	1st Qu.	:2019-06-19 00:00:00	1st Qu.	:2019		
Median	:2021-01-13 00:00:00	Median	:2021-01-29 00:00:00	Median	:2021		
Mean	:2020-11-27 19:06:52	Mean	:2021-01-04 18:59:32	Mean	:2021		
3rd Qu.	:2022-04-29 00:00:00	3rd Qu.	:2022-06-13 00:00:00	3rd Qu.	:2022		
Max.	:2023-11-17 00:00:00	Max.	:2023-12-22 00:00:00	Max.	:2023		
DIAS_HOSPITALIZADO		SEXO		EDAD		RANGO_ETAREO	
Min.	: 0.00	Length:	393	Min.	:15.00	Length:	393
1st Qu.	: 19.00	Class	:character	1st Qu.	:26.00	Class	:character
Median	: 30.00	Mode	:character	Median	:35.00	Mode	:character
Mean	: 37.99			Mean	:38.04		
3rd Qu.	: 44.00			3rd Qu.	:50.00		
Max.	:575.00			Max.	:78.00		
UNIDAD_EGRESO		DIAGNOSTICO_EGRESO		PERIODO			
Length:	393	Length:	393	Length:	393		
Class	:character	Class	:character	Class	:character		
Mode	:character	Mode	:character	Mode	:character		

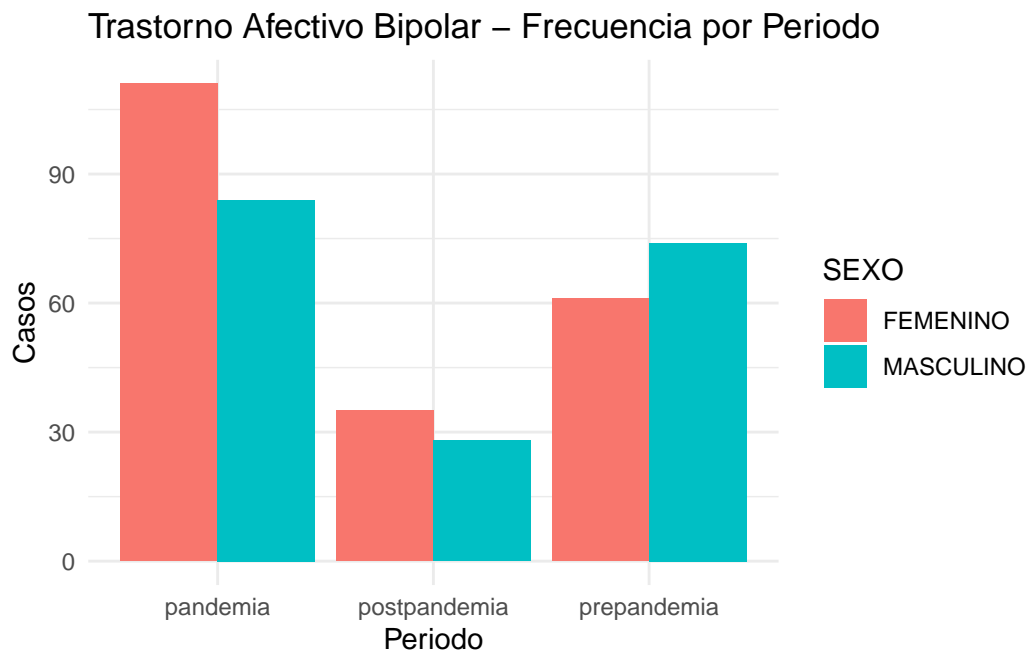
```
# Tabla de frecuencia por sexo
table(bipolar$SEXO)
```

FEMENINO	MASCULINO
207	186

```
# Tabla por periodo
table(bipolar$PERIODO)
```

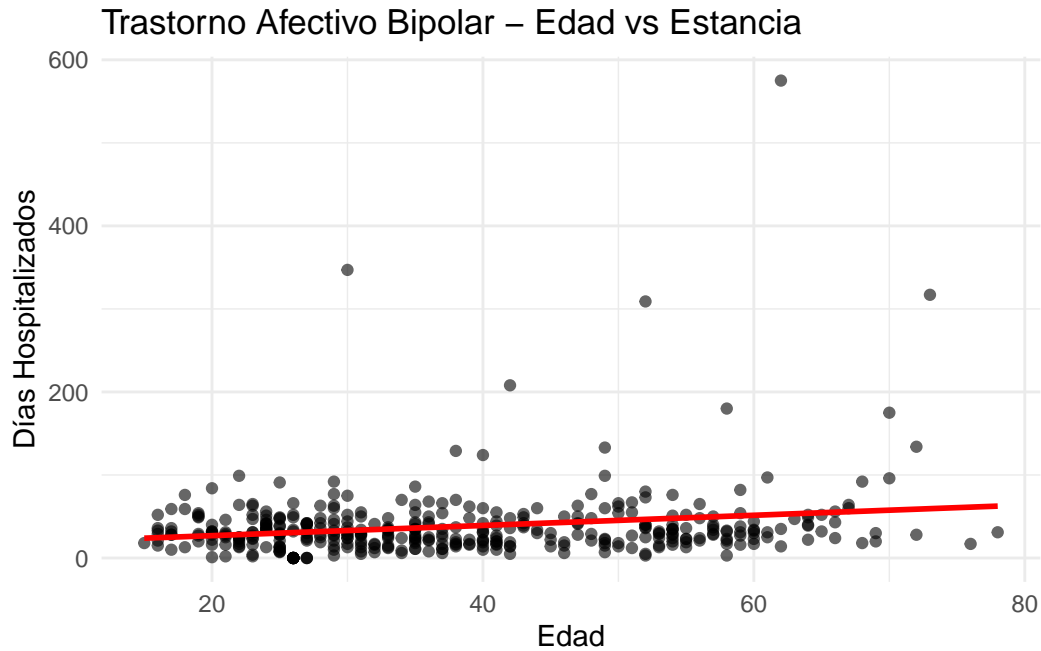
pandemia	postpandemia	prepandemia
195	63	135

```
# Gráfico de barras por periodo
ggplot(bipolar, aes(x = PERIODO, fill = SEXO)) +
  geom_bar(position = "dodge") +
  labs(title = "Trastorno Afectivo Bipolar - Frecuencia por Periodo", x = "Periodo", y = "Casos") +
  theme_minimal()
```



```
# Scatterplot: edad vs días hospitalizado
ggplot(bipolar, aes(x = EDAD, y = DIAS_HOSPITALIZADO)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Trastorno Afectivo Bipolar - Edad vs Estancia", x = "Edad", y = "Días Hospitalizados") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



```
# Regresión lineal
modelo_dep <- lm(DIAS_HOSPITALIZADO ~ EDAD, data = bipolar)
summary(modelo_dep)
```

Call:

```
lm(formula = DIAS_HOSPITALIZADO ~ EDAD, data = bipolar)
```

Residuals:

Min	1Q	Median	3Q	Max
-47.23	-21.46	-6.68	8.25	522.31

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.6787	6.2884	2.334	0.0201 *
EDAD	0.6130	0.1545	3.968	8.64e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.36 on 391 degrees of freedom

Multiple R-squared: 0.0387, Adjusted R-squared: 0.03624

F-statistic: 15.74 on 1 and 391 DF, p-value: 8.638e-05

2.0.3.1 Análisis de resultados:

- *Fechas y duración de hospitalización*

- **Período observado:** 2018 a 2023 (aunque hay un caso con ingreso en 2017).
- **Duración promedio:** 27.4 días (máx. 149 días).
- **Mediana de hospitalización:** 19 días → la mayoría permanece de 2 a 4 semanas.

- ***Características demográficas***

- Total de pacientes hospitalizados: **393**
- Edad Mínima: 15 años
 - * Promedio: **38.0 años**
 - * Mediana: 35 años
 - * Máxima: 78 años
- Sexo:
 - * **207 mujeres (52.7%)**
 - * **186 hombres (47.3%)**

2.0.3.2 Conclusion:

La hospitalización por TAB afecta a ambos sexos casi por igual, con leve predominio femenino. La edad promedio indica que este grupo abarca principalmente adultos jóvenes y de mediana edad.

- ***Distribucion por periodo epidemiológico***

- **Prepandemia (2018–2019):** 135 casos
- **Pandemia (2020–2022):** 195 casos (49.6% del total)
- **Postpandemia (2023):** 63 casos

2.0.3.3 conclusiones

- La pandemia COVID-19 representa el periodo con mayor número de hospitalizaciones, lo que podría asociarse a:
 - Mayor descompensación afectiva en pacientes con Trastorno Afectivo Bipolar
 - Disminución del acceso ambulatorio
 - Aumento del estrés sostenido y rupturas terapéuticas
- **Duración de hospitalización**
 - Promedio: 38 días

- Mediana: 30 días
- Máximo: 575 días

2.0.3.4 conclusiones

Aunque la mayoría se hospitaliza alrededor de 1 mes, algunos pacientes presentan estancias extraordinariamente largas, posiblemente por episodios severos, falta de red de apoyo o escasa respuesta al tratamiento.

• *MODELO DE REGRESIÓN: Días de hospitalización según edad*

– Ecuación del modelo:

$$\text{Estancia (días)} = 14.68 + 0.61 \times \text{Edad}$$

– Significancia:

- Coef. de edad: $p = 8.64e-05$ (muy significativo)
- $R^2 = 0.0387 \rightarrow$ Edad explica solo 3.9% de la variabilidad en los días hospitalizados.

2.0.3.5 conclusiones:

Aunque la relación entre edad y duración de hospitalización es significativa, su capacidad explicativa es baja. La duración probablemente está determinada por variables clínicas como:

- Presencia de síntomas psicóticos
- Número de episodios previos
- Adherencia a tratamiento
- Apoyo familiar o institucional

2.0.3.6 Conclusion general del análisis:

El análisis de los 393 pacientes hospitalizados por Trastorno Afectivo Bipolar muestra una distribución por sexo relativamente equilibrada, con un predominio de adultos entre los 30 y 50 años. La duración media de hospitalización es de 38 días, aunque existen casos extremos. Durante el periodo de pandemia se observó la mayor concentración de hospitalizaciones, lo que sugiere un fuerte impacto de la crisis sanitaria en esta población. Si bien existe una relación estadísticamente significativa entre edad y estancia, esta es débil desde el punto de vista clínico, lo que refuerza la necesidad de considerar múltiples factores para comprender y planificar la atención en salud mental de personas con TAB.

2.0.4 3. Esquizofrenia

```
# Cargar tabla de depresión desde el archivo generado
esquizofrenia <- read_excel("grupo_esquizofrenia.xlsx")

# Ver resumen de variables
summary(esquizofrenia)
```

FECHA_INGRESO		FECHA_EGRESO		AÑO_EGRESO	
Min.	:2013-01-08 00:00:00	Min.	:2018-01-03 00:00:00	Min.	:2018
1st Qu.	:2018-10-11 18:00:00	1st Qu.	:2019-01-25 18:00:00	1st Qu.	:2019
Median	:2020-01-23 12:00:00	Median	:2020-05-29 00:00:00	Median	:2020
Mean	:2020-05-29 00:07:44	Mean	:2020-09-11 00:50:19	Mean	:2020
3rd Qu.	:2022-03-06 18:00:00	3rd Qu.	:2022-06-29 12:00:00	3rd Qu.	:2022
Max.	:2023-11-23 00:00:00	Max.	:2023-12-28 00:00:00	Max.	:2023
DIAS_HOSPITALIZADO		SEXO	EDAD	RANGO_ETAREO	
Min.	: 0.0	Length:372	Min.	:12.0	Length:372
1st Qu.	: 31.0	Class :character	1st Qu.	:27.0	Class :character
Median	: 51.5	Mode :character	Median	:37.0	Mode :character
Mean	: 105.0		Mean	:39.4	
3rd Qu.	: 98.5		3rd Qu.	:52.0	
Max.	:2784.0		Max.	:83.0	
UNIDAD_EGRESO		DIAGNOSTICO_EGRESO	PERIODO		
Length:372		Length:372	Length:372		
Class :character		Class :character	Class :character		
Mode :character		Mode :character	Mode :character		

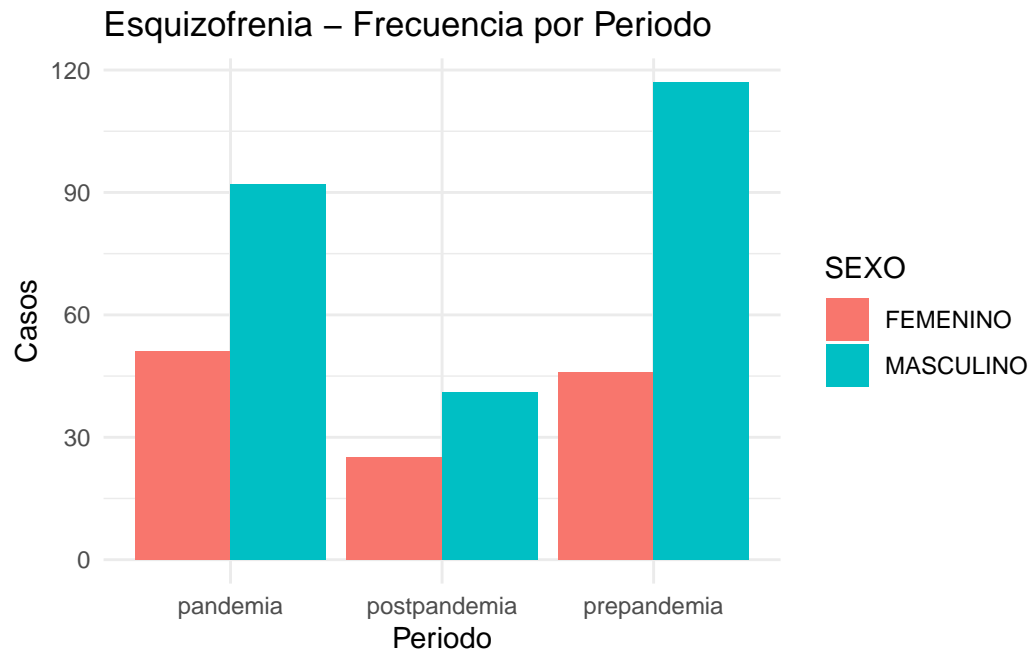
```
# Tabla de frecuencia por sexo
table(esquizofrenia$SEXO)
```

FEMENINO	MASCULINO
122	250

```
# Tabla por periodo
table(esquizofrenia$PERIODO)
```

pandemia	postpandemia	prepandemia
143	66	163

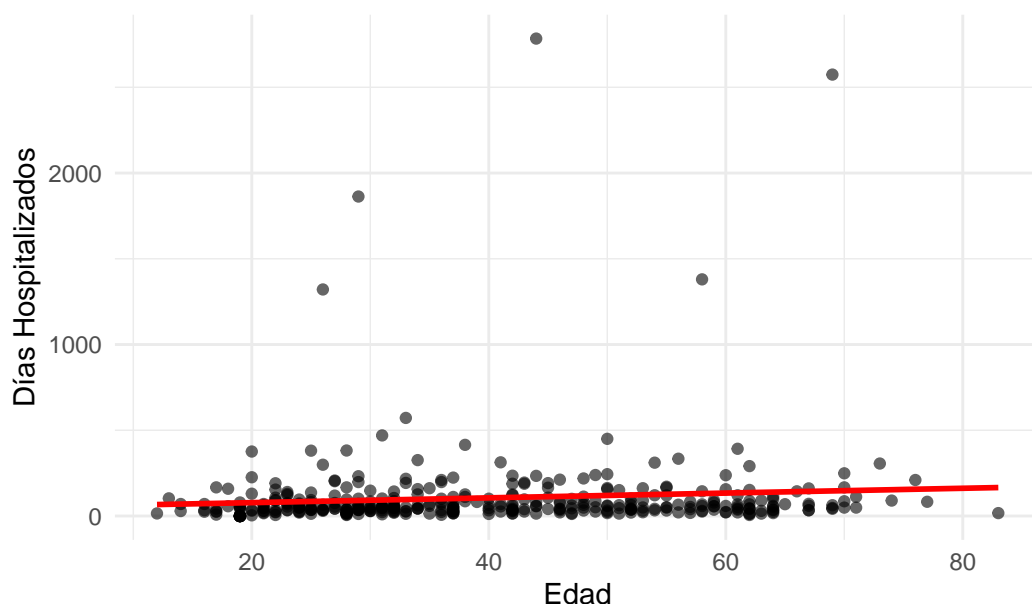
```
# Gráfico de barras por periodo
ggplot(esquizofrenia, aes(x = PERIODO, fill = SEXO)) +
  geom_bar(position = "dodge") +
  labs(title = "Esquizofrenia - Frecuencia por Periodo", x = "Periodo", y = "Casos") +
  theme_minimal()
```



```
# Scatterplot: edad vs días hospitalizado
ggplot(esquizofrenia, aes(x = EDAD, y = DIAS_HOSPITALIZADO)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Esquizofrenia - Edad vs Estancia", x = "Edad", y = "Días Hospitalizados") +
  theme_minimal()
```

`geom_smooth()` using formula = 'y ~ x'

Esquizofrenia – Edad vs Estancia



```
# Regresión lineal
modelo_dep <- lm(DIAS_HOSPITALIZADO ~ EDAD, data = esquizofrenia)
summary(modelo_dep)
```

Call:

```
lm(formula = DIAS_HOSPITALIZADO ~ EDAD, data = esquizofrenia)
```

Residuals:

Min	1Q	Median	3Q	Max
-148.91	-75.47	-48.66	-4.32	2672.55

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.002	34.264	1.459	0.1453
EDAD	1.397	0.808	1.728	0.0848 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 244.3 on 370 degrees of freedom

Multiple R-squared: 0.008009, Adjusted R-squared: 0.005328

F-statistic: 2.987 on 1 and 370 DF, p-value: 0.08475

2.0.4.1 Análisis de resultados:

- *Fechas y duración de hospitalización*

- **Período observado:** 2018 a 2023 (aunque hay un caso con ingreso en 2017).
- **Duración promedio:** 27.4 días (máx. 149 días).
- **Mediana de hospitalización:** 19 días → la mayoría permanece de 2 a 4 semanas.

- ***Características demográficas***

- Total de pacientes hospitalizados: 372
- Edad Máxima: 83 años
- Edad promedio no indicada directamente, pero el rango sugiere amplia dispersión
- Sexo:
 - * 122 mujeres (32.8%)
 - * 257 hombres (67.2%)

2.0.4.2 Conclusion:

Hay un predominio masculino claro en la hospitalización por esquizofrenia, lo que coincide con estudios clínicos que indican mayor prevalencia, inicio más temprano y evolución más crónica en hombres.

- ***Distribución por periodo epidemiológico***

- **Prepandemia (2018–2019):** 163 casos
- **Pandemia (2020–2022):** 143 casos (49.6% del total)
- **Postpandemia (2023):** 66 casos

2.0.4.3 conclusiones

- La mayor frecuencia se observa en el periodo prepandemia, con una disminución progresiva. Esto podría estar asociado a:
 - Reducción del acceso durante la pandemia
 - Cambios en las políticas de hospitalización
 - Derivaciones a servicios ambulatorios u otras instituciones
- **Duración de hospitalización**
 - No se entrega el promedio directamente, pero el máximo alcanza 2672 días (más de 7 años), lo que indica presencia de casos crónicos, judicializados o casos sociales.

2.0.4.4 conclusiones

Esquizofrenia es una patología severa con potencial de hospitalización prolongada. La mediana de los residuos negativos sugiere que muchos pacientes tienen estancias prolongadas, posiblemente institucionalizadas.

• *MODELO DE REGRESIÓN: Días de hospitalización según edad*

– Ecuación del modelo:

$$\text{Estancia (días)} = 50.0 + 1.40 \times \text{Edad}$$

– Significancia:

– Coef. de edad: $p = 0.0848 \rightarrow$ No significativa al nivel 5%, pero marginal ($p < 0.10$)

– $R^2 = 0.008 \rightarrow$ Edad explica solo el 0.8% de la variabilidad.

2.0.4.5 conclusiones:

La edad no predice de forma significativa ni útil la duración de la hospitalización en este grupo. Esto es esperable, ya que las estancias suelen depender de la severidad clínica, funcionalidad social, abandono familiar o medidas legales

2.0.4.6 Conclusion general del análisis:

En el grupo de pacientes hospitalizados por esquizofrenia, se observa un marcado predominio masculino (más de 2:1), con un amplio rango etario. La mayor frecuencia de ingresos ocurrió en el periodo prepandemia, con una disminución posterior. Las hospitalizaciones pueden alcanzar una duración extrema, lo que sugiere la existencia de casos crónicos o con internación prolongada por medidas judiciales. No se encontró una relación estadísticamente significativa entre la edad y la duración de hospitalización, indicando que el tiempo de estancia responde principalmente a factores clínicos y sociales distintos a la edad. Este patrón refuerza la necesidad de enfoques integrales para el manejo de la esquizofrenia, considerando aspectos como adherencia, soporte social, reinserción y comorbilidades psiquiátricas o neurológicas.

2.0.5 4. Consumo alcohol y drogas

```
# Cargar tabla de depresión desde el archivo generado
droga <- read_excel("grupo_consumo.xlsx")

# Ver resumen de variables
summary(droga)
```

FECHA_INGRESO		FECHA_EGRESO		AÑO_EGRESO	
Min.	:2016-06-14 00:00:00	Min.	:2018-01-03 00:00:00	Min.	:2018
1st Qu.	:2019-01-29 00:00:00	1st Qu.	:2019-03-04 12:00:00	1st Qu.	:2019
Median	:2020-05-18 00:00:00	Median	:2020-08-03 00:00:00	Median	:2020
Mean	:2020-09-27 14:46:20	Mean	:2020-10-30 14:33:39	Mean	:2020
3rd Qu.	:2022-08-09 12:00:00	3rd Qu.	:2022-09-06 00:00:00	3rd Qu.	:2022
Max.	:2023-12-12 00:00:00	Max.	:2023-12-26 00:00:00	Max.	:2023

DIAS_HOSPITALIZADO		SEXO	EDAD	RANGO_ETAREO
Min.	: 0.00	Length:567	Min.	:10.00
1st Qu.	: 12.00	Class :character	1st Qu.	:16.00
Median	: 20.00	Mode :character	Median	:18.00
Mean	: 32.99		Mean	:24.34
3rd Qu.	: 30.00		3rd Qu.	:30.00
Max.	:1658.00		Max.	:67.00

UNIDAD_EGRESO	DIAGNOSTICO_EGRESO	PERIODO
Length:567	Length:567	Length:567
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

```
# Tabla de frecuencia por sexo
table(droga$SEXO)
```

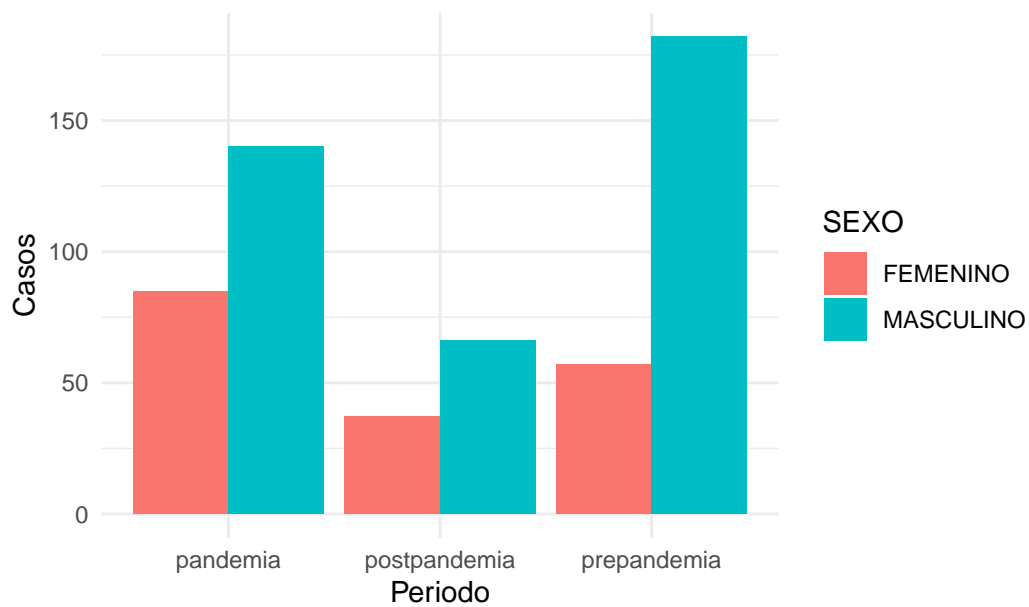
```
FEMENINO MASCULINO
      179      388
```

```
# Tabla por periodo
table(droga$PERIODO)
```

```
      pandemia postpandemia prepandemia
      225          103          239
```

```
# Gráfico de barras por periodo
ggplot(droga, aes(x = PERIODO, fill = SEXO)) +
  geom_bar(position = "dodge") +
  labs(title = "Consumo de Alcohol y Drogas - Frecuencia por Periodo", x = "Periodo", y = "Frecuencia")
theme_minimal()
```

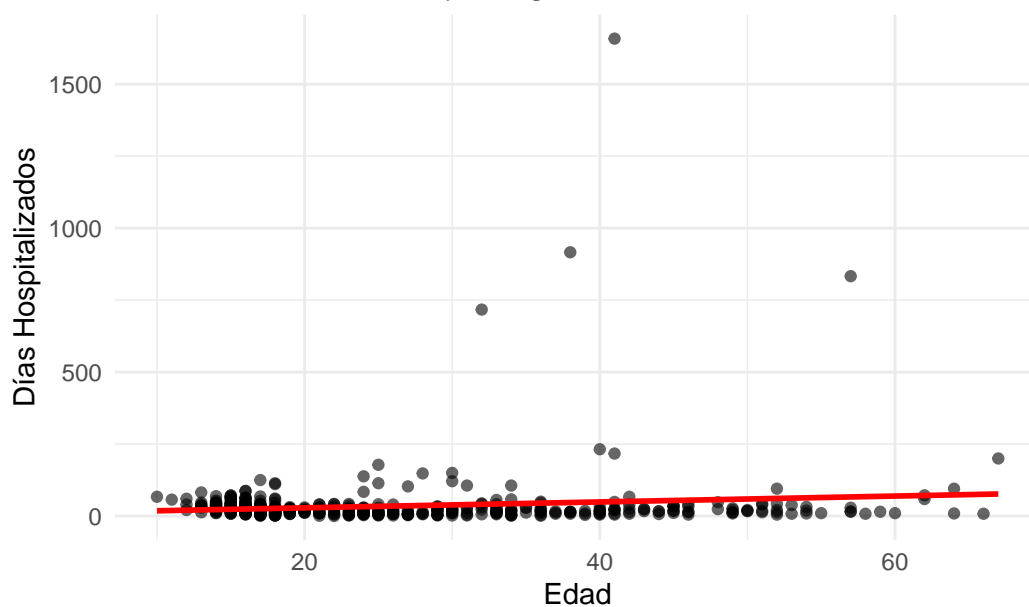
Consumo de Alcohol y Drogas – Frecuencia por Periodo



```
# Scatterplot: edad vs días hospitalizado
ggplot(droga, aes(x = EDAD, y = DIAS_HOSPITALIZADO)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Consumo de Alcohol y Droga - Edad vs Estancia", x = "Edad", y = "Días Ho
  theme_minimal()
```

`geom_smooth()` using formula = 'y ~ x'

Consumo de Alcohol y Droga – Edad vs Estancia




```
# Regresión lineal
modelo_dep <- lm(DIAS_HOSPITALIZADO ~ EDAD, data = droga)
summary(modelo_dep)
```

Call:

```
lm(formula = DIAS_HOSPITALIZADO ~ EDAD, data = droga)
```

Residuals:

Min	1Q	Median	3Q	Max
-67.55	-22.71	-9.51	2.52	1607.99

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.1241	9.0299	0.900	0.36867
EDAD	1.0216	0.3347	3.052	0.00238 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 92.7 on 565 degrees of freedom

Multiple R-squared: 0.01622, Adjusted R-squared: 0.01448

F-statistic: 9.315 on 1 and 565 DF, p-value: 0.00238

2.0.5.1 Análisis de resultados:

- *Características demográficas*

- Total de pacientes hospitalizados: 567

- Edad:

- * Mínima: 10 años

- * Mediana: 18 años

- * Promedio: 24.3 años

- * Máxima: 67 años

- Sexo:

- * 179 mujeres (31.6%)

- * 388 hombres (68.4%)

2.0.5.2 Conclusion:

Este grupo presenta un claro predominio masculino, y está compuesto mayoritariamente por adolescentes y adultos jóvenes. La edad mediana de 18 años sugiere un inicio temprano del consumo problemático, lo cual es coherente con los patrones de adicción en población joven.

- ***Distribucion por periodo epidemiológico***

- **Prepandemia (2018–2019):** 239 casos
- **Pandemia (2020–2022):** 225 casos
- **Postpandemia (2023):** 103 casos

2.0.5.3 conclusiones

- La distribución temporal es bastante equilibrada. Se observa una ligera baja durante la pandemia, pero se mantiene alta, probablemente debido a:
 - Deterioro de las redes de apoyo
 - Aumento del consumo durante confinamientos
 - Mayor visibilidad de síntomas conductuales que requieren hospitalización
- **Duración de hospitalización**
 - **Promedio:** 33 días
 - **Mediana:** 20 días
 - **Máximo:** 1658 días (más de 4 años)

2.0.5.4 conclusiones

Si bien la mayoría de los pacientes tienen estancias cortas a medianas, existen casos de hospitalización extremadamente prolongada, lo que podría indicar:

- Medidas judiciales (internación forzosa)
- Comorbilidades graves
- Ausencia de redes sociales o familiares de apoyo
- ***MODELO DE REGRESIÓN: Días de hospitalización según edad***
 - **Ecuación del modelo:**
$$\text{Estancia (días)} = 8.12 + 1.02 \times \text{Edad}$$
 - **Significancia:**

- Coef. de edad: $p = 0.00238$ (estadísticamente significativa)
- $R^2 = 0.016 \rightarrow$ Edad explica solo el 1.6% de la variabilidad.

2.0.5.5 conclusiones:

A mayor edad, se observa una leve tendencia a estancias más largas. Sin embargo, la edad no explica de forma relevante la duración de hospitalización, lo que sugiere que la permanencia está influida por factores clínicos, legales o sociales más que por la edad misma.

2.0.5.6 Conclusion general del análisis:

El grupo de pacientes hospitalizados por consumo de alcohol y drogas presenta un predominio masculino y juvenil, con una edad media de 24 años y una mediana de solo 18 años, lo que destaca la precocidad del ingreso a hospitalizaciones psiquiátricas por consumo. La duración de hospitalización promedio es de 33 días, aunque se registran casos excepcionales de estancia prolongada, lo que indica posibles intervenciones judiciales o severas condiciones sociales y clínicas. La relación entre edad y días hospitalizados es significativa pero clínicamente débil, reflejando que la hospitalización en estos casos depende principalmente de factores como la gravedad del consumo, la presencia de psicosis inducida, recaídas frecuentes o la falta de recursos externos. Este grupo representa una prioridad para políticas de intervención temprana y reinserción social.

2.0.6 5. Alzheimer

```
# Cargar tabla de depresión desde el archivo generado
alz <- read_excel("grupo_alzheimer.xlsx")

# Ver resumen de variables
summary(alz)
```

FECHA_INGRESO	FECHA_EGRESO	AÑO_EGRESO	DIAS_HOSPITALIZADO
Min. :2021-05-19	Min. :2022-11-08	Min. :2022	Min. :538
1st Qu.:2021-05-19	1st Qu.:2022-11-08	1st Qu.:2022	1st Qu.:538
Median :2021-05-19	Median :2022-11-08	Median :2022	Median :538
Mean :2021-05-19	Mean :2022-11-08	Mean :2022	Mean :538
3rd Qu.:2021-05-19	3rd Qu.:2022-11-08	3rd Qu.:2022	3rd Qu.:538
Max. :2021-05-19	Max. :2022-11-08	Max. :2022	Max. :538

SEXO	EDAD	RANGO_ETAREO	UNIDAD_EGRESO
Length:1	Min. :51	Length:1	Length:1
Class :character	1st Qu.:51	Class :character	Class :character
Mode :character	Median :51	Mode :character	Mode :character
	Mean :51		

```
3rd Qu.:51
Max.    :51
DIAGNOSTICO_EGRESO PERIODO
Length:1          Length:1
Class :character  Class :character
Mode  :character  Mode  :character
```

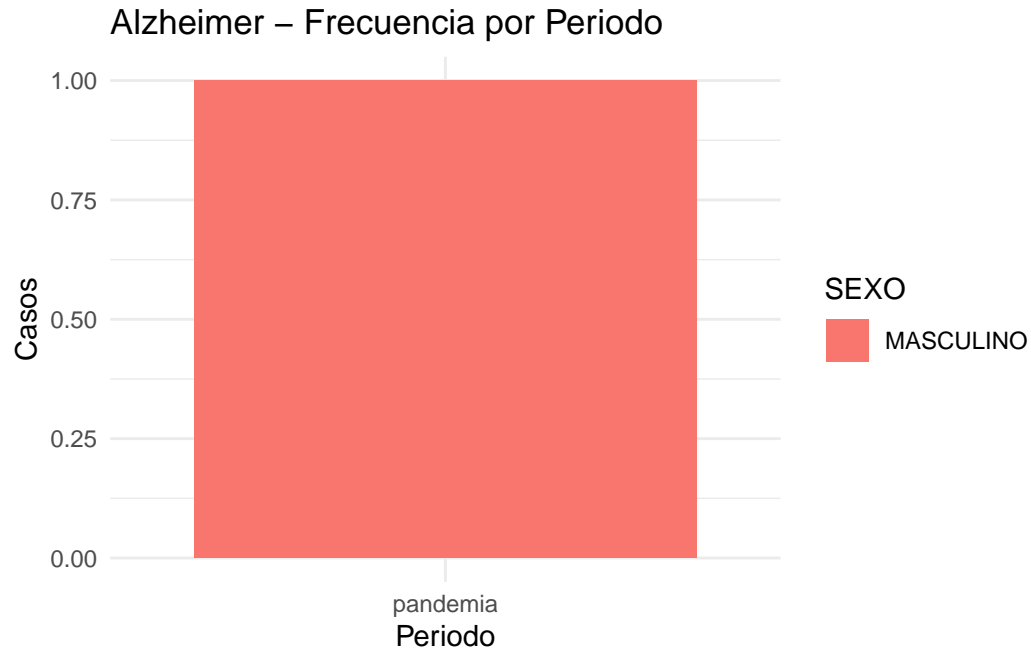
```
# Tabla de frecuencia por sexo
table(alz$SEXO)
```

```
MASCULINO
1
```

```
# Tabla por periodo
table(alz$PERIODO)
```

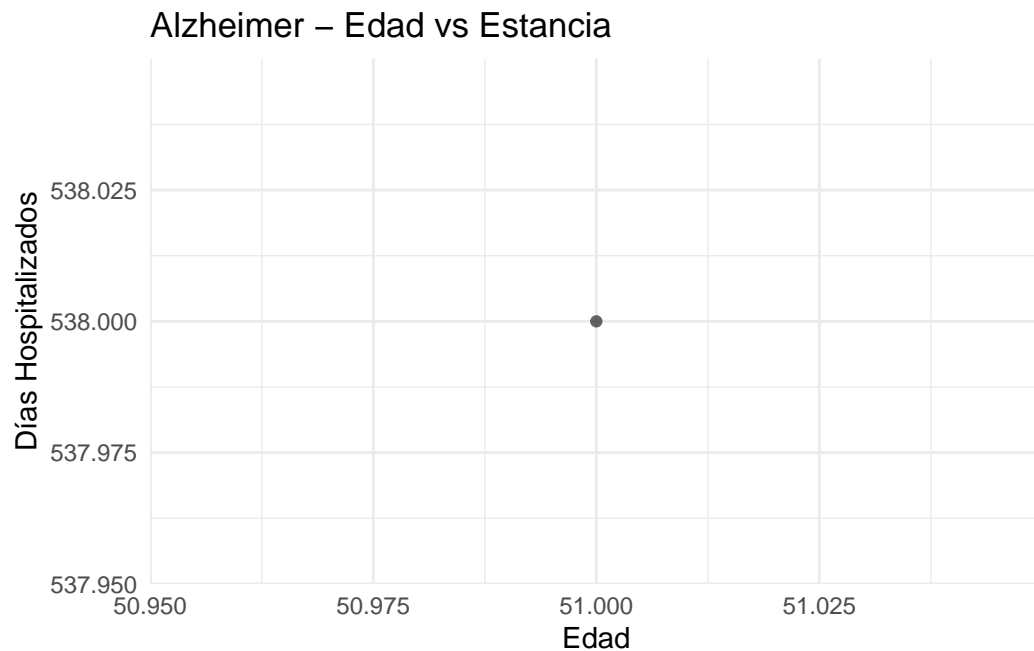
```
pandemia
1
```

```
# Gráfico de barras por periodo
ggplot(alz, aes(x = PERIODO, fill = SEXO)) +
  geom_bar(position = "dodge") +
  labs(title = "Alzheimer - Frecuencia por Periodo", x = "Periodo", y = "Casos") +
  theme_minimal()
```



```
# Scatterplot: edad vs días hospitalizado
ggplot(alz, aes(x = EDAD, y = DIAS_HOSPITALIZADO)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Alzheimer - Edad vs Estancia", x = "Edad", y = "Días Hospitalizados") +
  theme_minimal()
```

`geom_smooth()` using formula = 'y ~ x'



```
# Regresión lineal
modelo_dep <- lm(DIAS_HOSPITALIZADO ~ EDAD, data = alz)
summary(modelo_dep)
```

Call:

```
lm(formula = DIAS_HOSPITALIZADO ~ EDAD, data = alz)
```

Residuals:

ALL 1 residuals are 0: no residual degrees of freedom!

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	538	NaN	NaN	NaN
EDAD	NA	NA	NA	NA

Residual standard error: NaN on 0 degrees of freedom

En este caso, solo se tiene información de una persona hospitalizada con esta patología, lo que no permite realizar un análisis completo de datos. Esto no quita relevancia a la información que se tiene de él debido a lo prolongado de su hospitalización (538 días equivalente a 1 año y 5 meses aproximadamente), por ejemplo:

- Se trata de un paciente varón relativamente joven para un diagnóstico de Alzheimer (que suele presentarse en mayores de 65 años), lo que sugiere un caso de Alzheimer de inicio temprano, o bien una clasificación diagnóstica inusual o errónea.
- Su caso es relevante debido a la extrema duración de hospitalización (538 días), lo que sugiere:
 - Posible abandono familiar
 - Comorbilidades psiquiátricas o neurológicas
 - Internación judicial o social por falta de red de cuidados
- No se puede inferir una tendencia poblacional a partir de un solo caso, pero este tipo de casos ayuda a ilustrar problemáticas graves de institucionalización prolongada, particularmente en enfermedades neurodegenerativas.

Este caso, único en la base de datos, no permite extraer conclusiones estadísticas, pero sí plantea interrogantes clínicas importantes. La edad sugiere un posible Alzheimer de inicio temprano o un error de codificación diagnóstica. La duración extrema de la hospitalización podría estar asociada a la ausencia de una red de apoyo o medidas judiciales. Este tipo de casos refleja una situación crítica de institucionalización prolongada que merece atención desde una perspectiva de salud pública y ética del cuidado.

2.0.6.1 6. Otros diagnóstico psiquiátricos

Este grupo reúne todos los casos que no cumplen criterios específicos para ser clasificados dentro de los cinco principales grupos analizados (depresión, trastorno afectivo bipolar, esquizofrenia, consumo de sustancias y Alzheimer). Debido a su alta heterogeneidad diagnóstica, que incluye trastornos de personalidad, cuadros adaptativos, neurodesarrollo, síntomas inespecíficos o diagnósticos mal codificados, este grupo no fue sometido a un análisis estadístico detallado en el presente informe.

Consideramos que su evaluación adecuada requiere una exploración específica y más profunda, idealmente con reclasificación diagnóstica interna y segmentación temática. Este análisis excede los objetivos y extensión definidos para esta evaluación. No obstante, se reconoce que este grupo representa una proporción importante de los egresos psiquiátricos y plantea desafíos significativos en términos de clasificación clínica, gestión institucional y asignación de recursos.

3 Conclusión Final

El desarrollo de este trabajo tuvo como objetivo principal la creación de una función en R capaz de automatizar el análisis de hospitalizaciones psiquiátricas, utilizando una base de datos real proveniente del Hospital Guillermo Grant Benavente. Desde su planificación, se planteó la necesidad de contar con una herramienta flexible, escalable y reutilizable, que permitiera clasificar y analizar grandes volúmenes de información de manera eficiente y ordenada.

En lo personal, este proyecto nos permitió comprobar el potencial de la programación funcional en R para resolver desafíos reales del ámbito clínico. La posibilidad de generar bases de datos filtradas por diagnóstico, exportarlas, analizarlas estadísticamente y visualizarlas, no solo facilita el trabajo individual, sino que abre la puerta para que otros profesionales o investigadores puedan utilizar estos datos de manera independiente, fomentando la colaboración interdisciplinaria.

Nos sorprendió especialmente encontrar una proporción importante de casos agrupados como “otros diagnósticos psiquiátricos”, que si bien no forman parte de las categorías tradicionales ni de las políticas GES en salud mental, contienen información valiosa que amerita un análisis más profundo. Esta observación refuerza la importancia de una codificación diagnóstica precisa y de políticas públicas que se ajusten a la realidad clínica observada en los servicios de salud.

En resumen, este fue un trabajo largo y desafiante, pero que logró cumplir sus objetivos. Más allá de lo técnico, me permitió reflexionar sobre la relevancia del análisis de datos en contextos clínicos y su potencial para apoyar la toma de decisiones en salud pública y mejorar la atención de los pacientes.