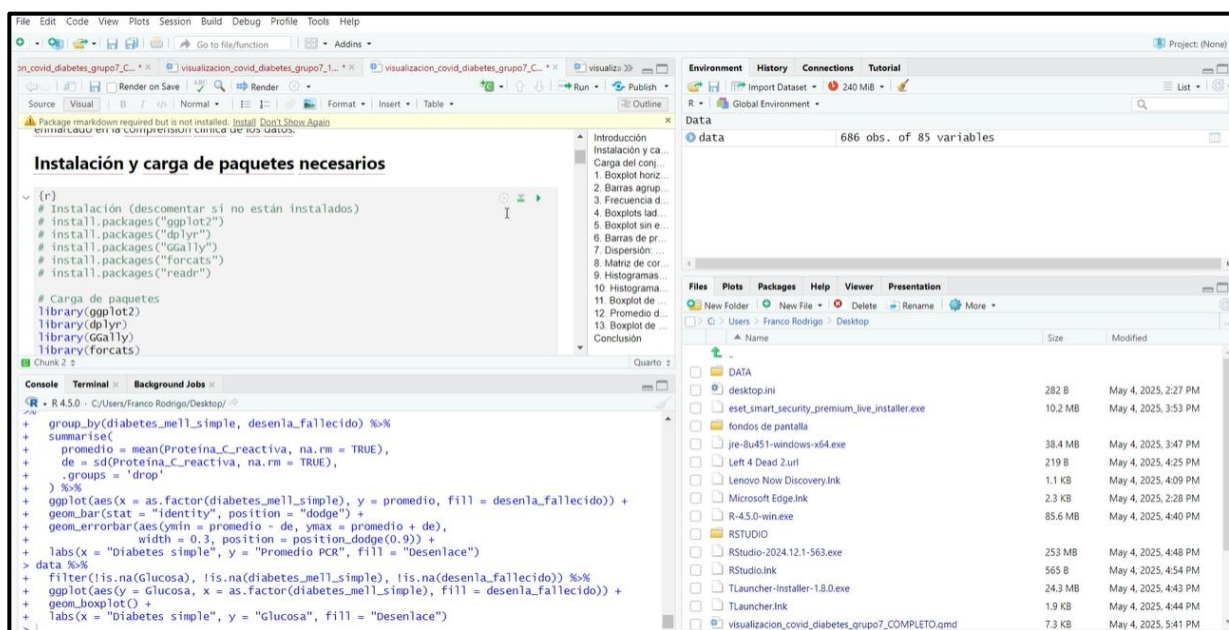


# ACTIVIDAD 5: (DIABETES\_COVID\_19)

## Integrantes del equipo:

- Meyli Llancari Nivin
- Christian Raul Flores Diaz
- Juana July Vera Fonseca
- Lucia Rosa Ccuro Minaya
- Melissa Alejo Huaman

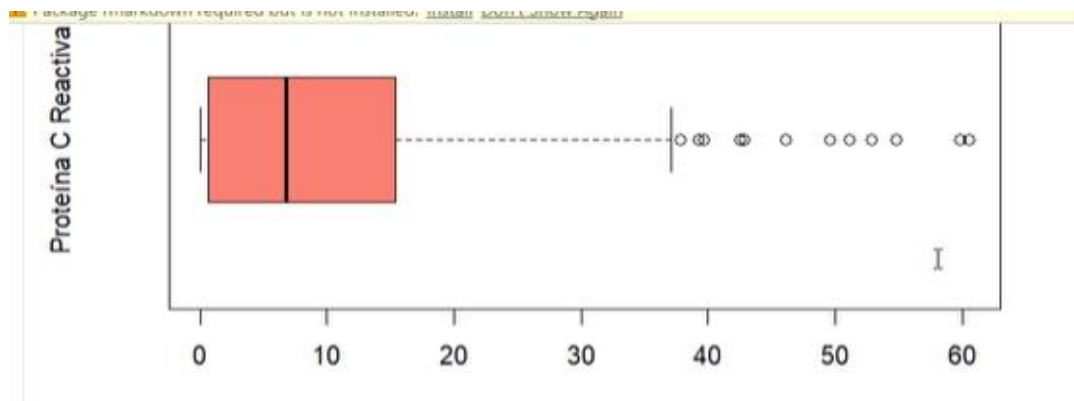
## DESARROLLO DE LA ACTIVIDAD:



## 1. Boxplot horizontal de proteína C reactiva

**Descripción:** El gráfico de caja (boxplot) permite visualizar la distribución de una variable numérica, mostrando la mediana, los cuartiles y los valores atípicos. Aquí analizamos la variable Proteína\_C\_reactiva.

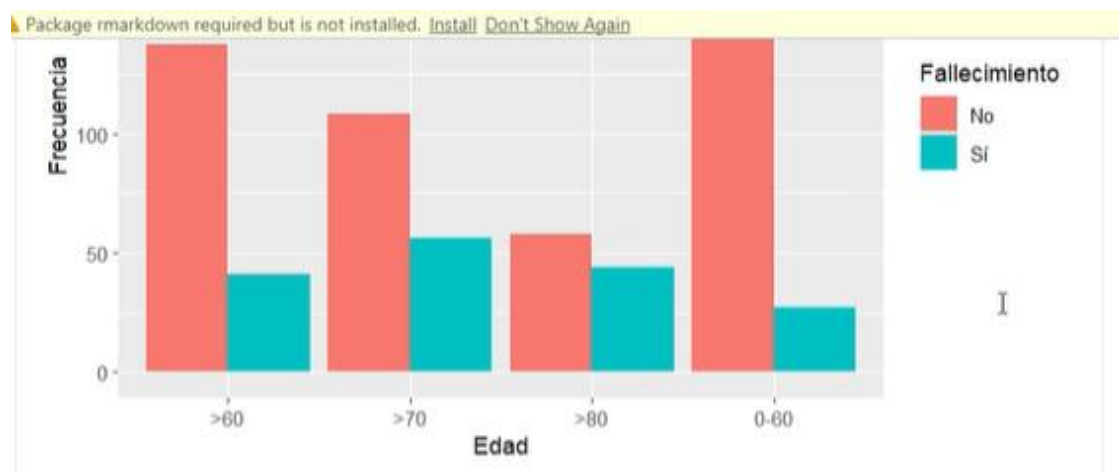
```
{r}  
boxplot(data$`Proteína_C_reactiva`, horizontal = TRUE, col = "salmon", ylab =  
"Proteína C Reactiva")
```



## 2. Barras agrupadas: Edad vs Fallecimiento

**Descripción:** Se analiza cómo se distribuyen los desenlaces (fallecido o no) entre los diferentes rangos de edad. Las barras muestran la frecuencia de pacientes por grupo.

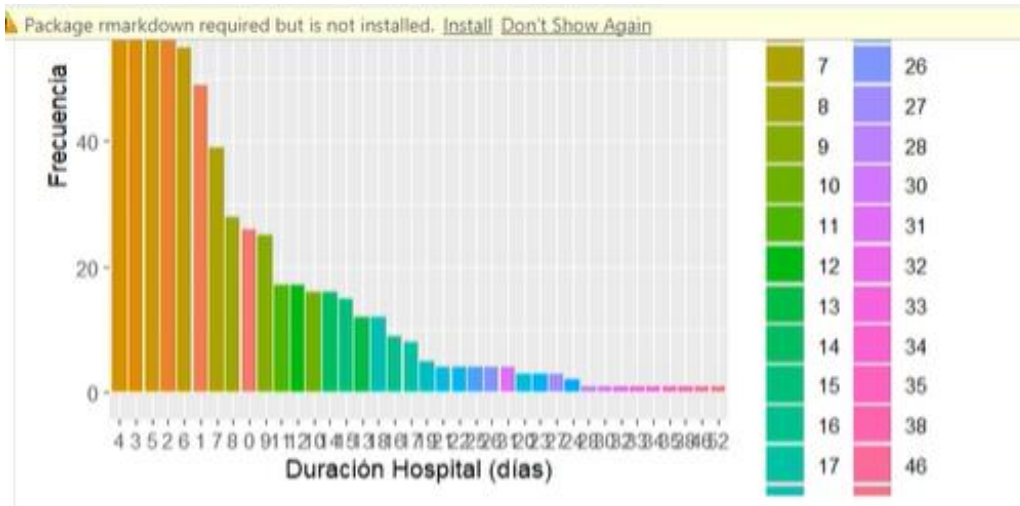
```
{r}  
ggplot(data, aes(x = edad, fill = desenla_fallecido)) +  
  geom_bar(position = "dodge") +  
  labs(y = "Frecuencia", x = "Edad", fill = "Fallecimiento")
```



### 3. Frecuencia de duración hospitalaria

**Descripción:** Se analiza cuántos pacientes permanecieron hospitalizados una cantidad determinada de días. Se usa `fct_infreq` para ordenar por frecuencia.

```
{r}  
ggplot(data, aes(x = fct_infreq(as.factor(duraci_hospita_diaz)), fill = as  
.factor(duraci_hospita_diaz))) +  
  geom_bar() +  
  labs(y = "Frecuencia", x = "Duración Hospital (días)")
```



### 4. Boxplots lado a lado: Proteína C reactiva por desenlace

**Descripción:** Se compara la distribución de los niveles de proteína C reactiva entre pacientes fallecidos y no fallecidos mediante boxplots.

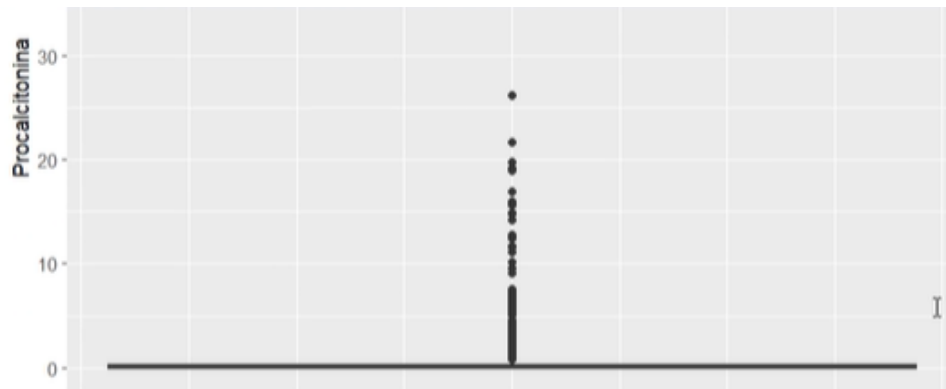
```
{r}  
data %>%  
  filter(!is.na('Proteína_C_reactiva') & !is.na(desenla_fallecido)) %>%  
  ggplot(aes(y = 'Proteína_C_reactiva', x = desenla_fallecido)) +  
  geom_boxplot() +  
  labs(y = "Proteína C Reactiva", x = "Fallecimiento")
```



## 5. Boxplot sin eje X: Procalcitonina

**Descripción:** Representa la distribución de la variable Procalcitonina eliminando los textos del eje X para centrarse solo en la variable dependiente.

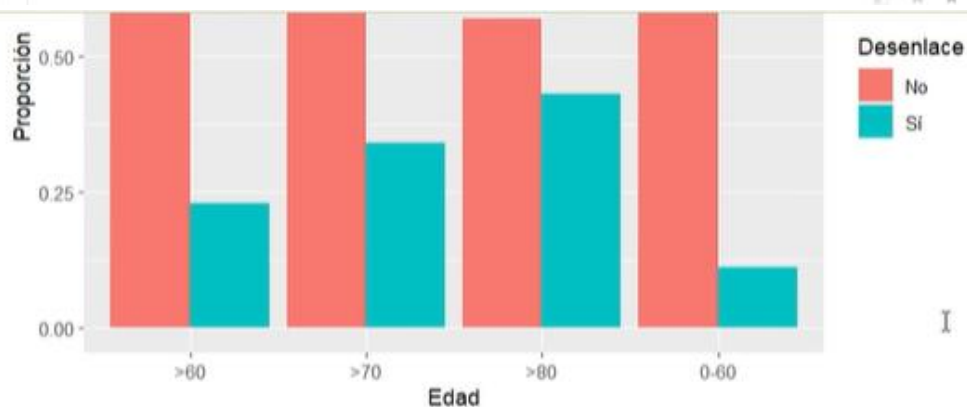
```
{r}
ggplot(data, aes(y = Procalcitonina)) +
  geom_boxplot() +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  labs(y = "Procalcitonina")
```



## 6. Barras de proporciones por edad y desenlace

**Descripción:** Se calcula la proporción de desenlaces (fallecimiento) dentro de cada grupo etario, permitiendo ver qué grupo tiene mayor tasa de mortalidad.

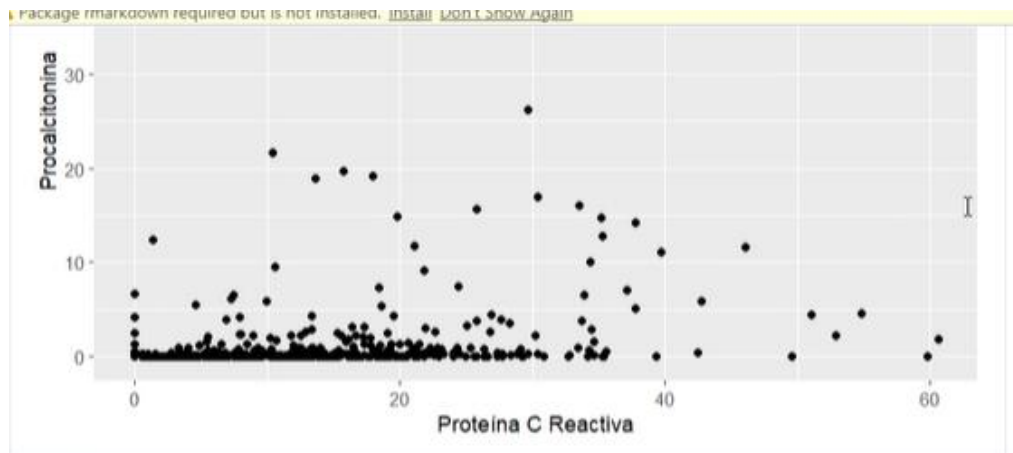
```
{r}
data %>%
  group_by(edad, desenla_fallecido) %>%
  count() %>%
  group_by(edad) %>%
  mutate(Proportion = n / sum(n)) %>%
  ggplot(aes(x = edad, y = Proportion, fill = desenla_fallecido)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(y = "Proporción", x = "Edad", fill = "Desenlace")
```



## 7. Dispersión: Proteína C reactiva vs Procalcitonina

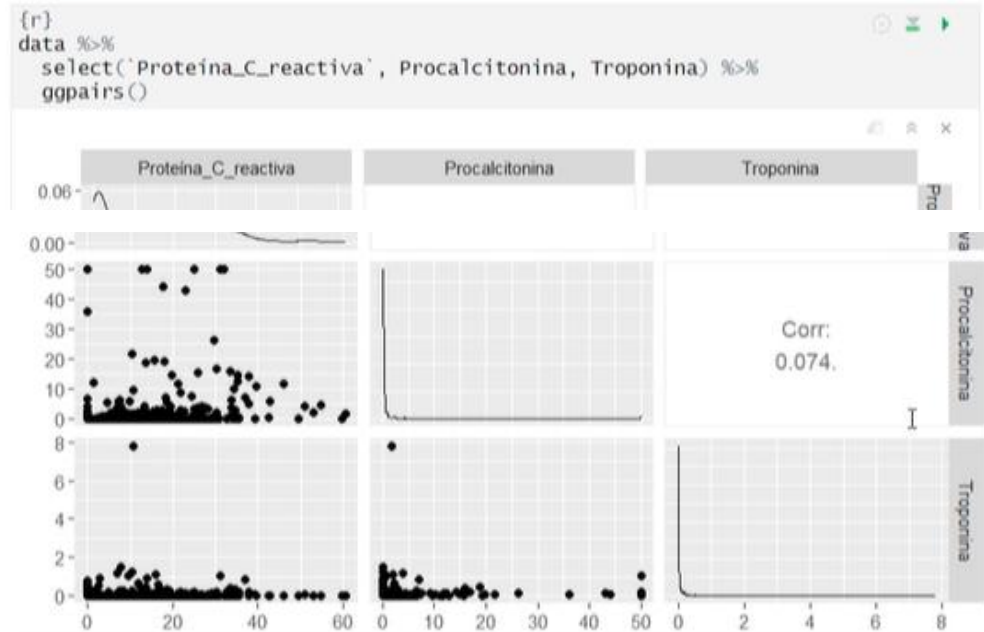
**Descripción:** Se usa un gráfico de dispersión para observar la posible relación entre dos biomarcadores: proteína C reactiva y procalcitonina.

```
{r}
ggplot(data, aes(x = 'Proteína_C_reactiva', y = Procalcitonina)) +
  geom_point() +
  labs(x = "Proteína C Reactiva", y = "Procalcitonina")
```



## 8. Matriz de correlación

**Descripción:** Se genera una matriz de correlaciones entre tres variables continuas clave, mostrando gráficos de dispersión y coeficientes de correlación.

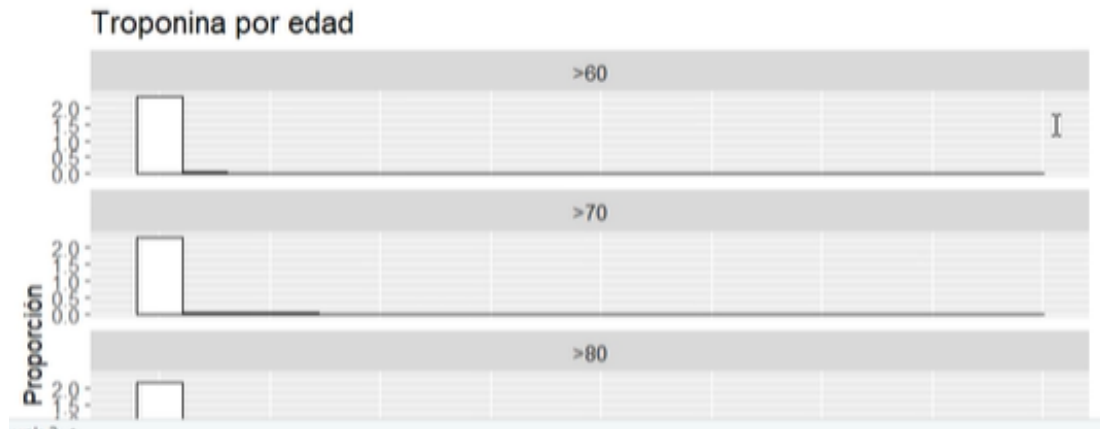




## 9. Histogramas por grupo etario

**Descripción:** Se muestran histogramas de troponina distribuidos en facetas por grupo de edad para observar patrones de distribución por subgrupo.

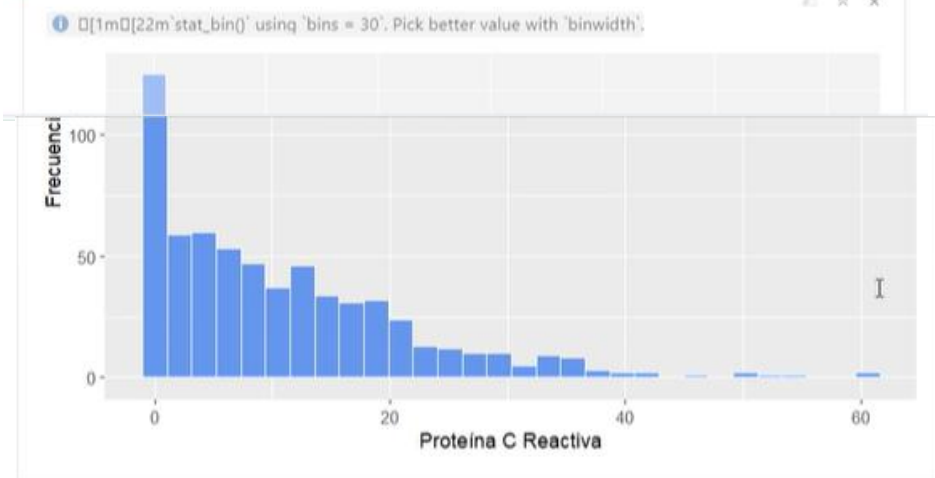
```
{r}  
data %>%  
  filter(!is.na(edad) & !is.na(Troponina)) %>%  
  ggplot(aes(x = as.numeric(Troponina))) +  
  geom_histogram(aes(y = ..density..), bins = 20, color = "black", fill =  
  "white") +  
  facet_wrap(~edad, nrow = 4) +  
  labs(x = "Troponina", y = "Proporción") +  
  ggtitle("Troponina por edad")
```



## 10. Histograma coloreado de proteína C reactiva

**Descripción:** Se representa la frecuencia de valores de proteína C reactiva usando un histograma con barras coloreadas.

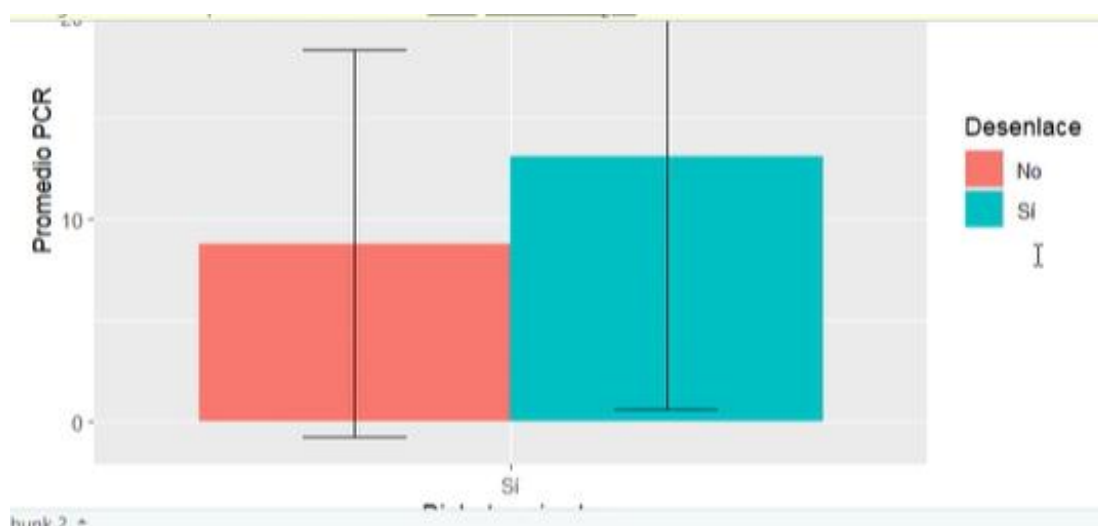
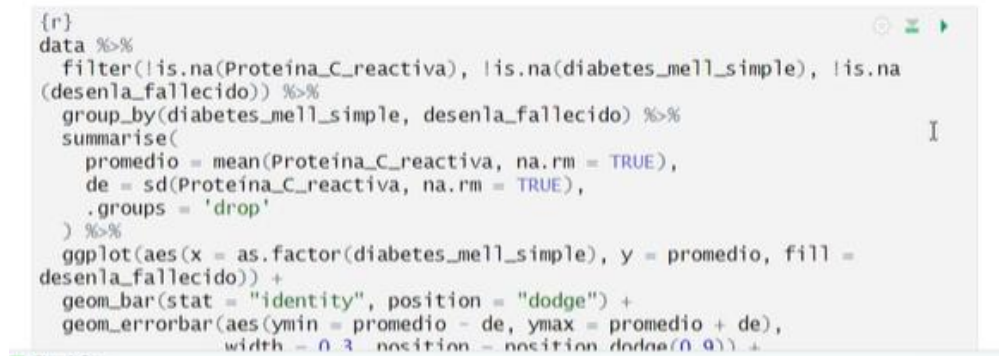
```
{r}  
ggplot(data, aes(x = 'Proteina_C_reactiva')) +  
  geom_histogram(color = "white", fill = "cornflowerblue") +  
  labs(y = "Frecuencia", x = "Proteina C Reactiva")
```



## 11. Boxplot de Troponina por desenlace, coloreado por glucosa crítica

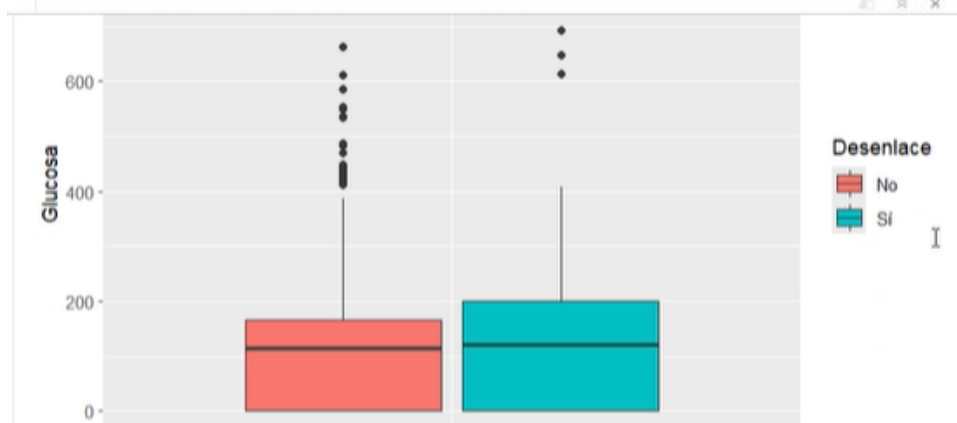


## 12. Promedio de Proteína C Reactiva por Diabetes y Desenlace



### 13. Boxplot de Glucosa según Diabetes, coloreado por desenlace

```
{r}  
data %>%  
  filter(!is.na(Glucosa), !is.na(diabetes_mell_simple), !is.na(desenla_fallecido)) %>%  
  ggplot(aes(y = Glucosa, x = as.factor(diabetes_mell_simple), fill = desenla_fallecido)) +  
  geom_boxplot() +  
  labs(x = "Diabetes simple", y = "Glucosa", fill = "Desenlace")
```



### Conclusión

Gracias al uso de múltiples técnicas de visualización, este informe permite comprender la distribución y relaciones entre variables clave del conjunto de datos de pacientes con COVID-19 y diabetes. El uso de colores, agrupaciones y descripciones mejora la interpretación y resalta diferencias importantes entre grupos clínicos.