

# Task 5\_ggplot\_Moreno\_Ochoa\_Casarrubia

Moreno - Ochoa - Casarrubia

25/10/2021

## **Tema:** Visualización usando ggplot2

*NOMBRES:* Miguel Moreno - Esteban Ochoa - David Casarrubia

```
library(tidyverse); library(ggplot2);library(corrplot);library(knitr)
library(readr); library(dplyr); library(reshape2); library(tibble)
```

1. Carga el dataset de diabetes.csv file

```
getwd()
```

```
## [1] "C:/Users/Miguel Angel/Documents/materias/Ciencias de datos/DS - 2021/tareas en clase/tarea 5"
```

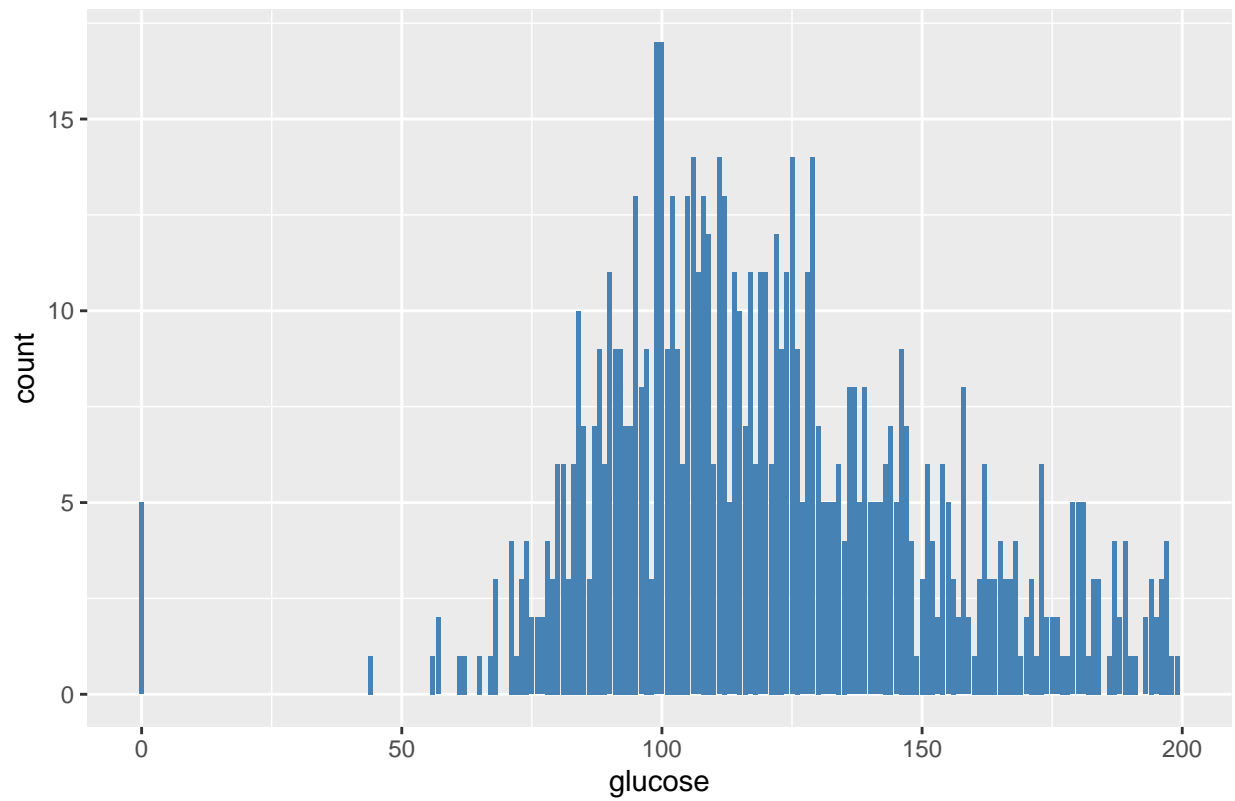
```
setwd("C:/Users/Miguel Angel/Documents/materias/Ciencias de datos/DS - 2021/Scripts_DS/1- Data")
# Asigna la ruta a una variable que crearemos "dataset"
diabetes <- as.data.frame(read_delim("diabetes.csv", ",")) # asigna l
##os datos a una variable llamada dataset y lee los datos
names(diabetes)
```

```
## [1] "...1"      "pregnant" "glucose"   "pressure" "triceps"   "insulin"
## [7] "mass"      "pedigree" "age"       "diabetes"
```

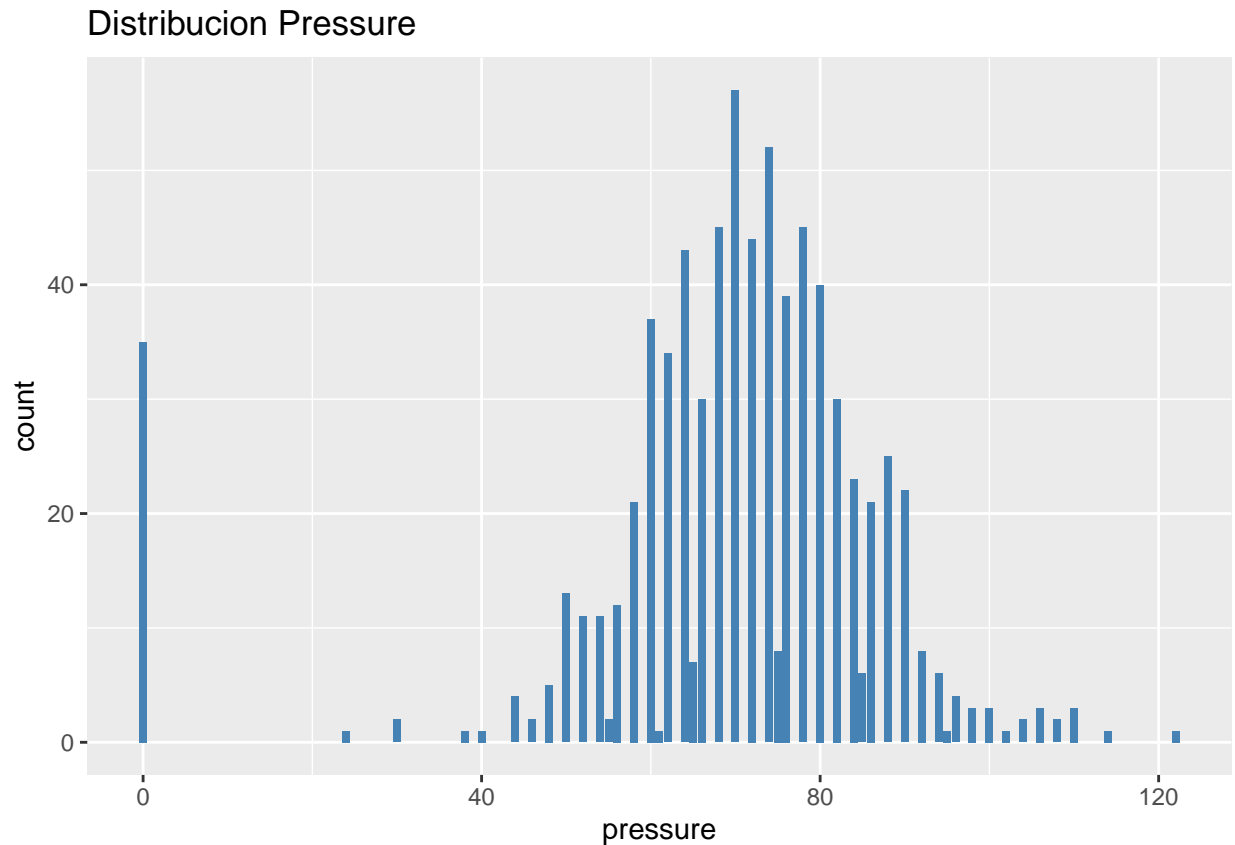
2. Crea una gráfica que muestre la distribución de la variable Glucose y otra para la pressure. ¿Cuál es valor de glucosa con una mayor frecuencia en este dataset?, ¿Existen pacientes con presión arterial sistólica alta en el dataset? Justifique.

```
ggplot(diabetes, aes(x = glucose)) + geom_bar(fill="steelblue") +
  labs(title="Distribucion Glucose")
```

Distribucion Glucose



```
ggplot(diabetes, aes(x = pressure)) + geom_bar(fill = "steelblue") +  
  labs(title="Distribucion Pressure")
```



R// A) El valor con una mayor frecuencia con respecto a la glucosa es 100, ya que se repite con una frecuencia de 80 veces según la información en el dataset.

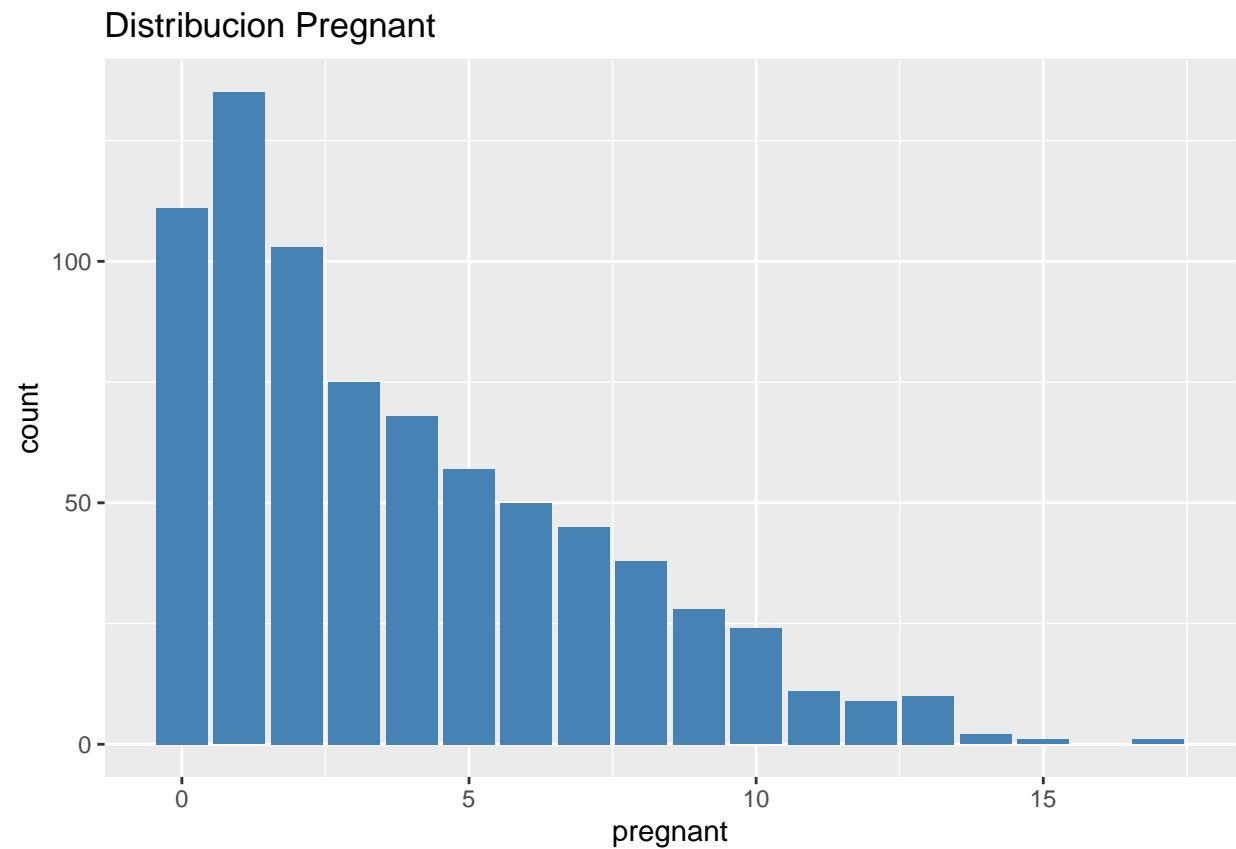
```
summary(diabetes$pressure)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   62.00   72.00   69.11   80.00   122.00
```

B) En el dataset la presión arterial sistólica más alta posee un valor de 122, para lo cual en la medicina no es considerado una presión alta pero dado en el dataset es considerado el valor de la presión más alta.

3. Crea una gráfica que muestre la distribución de la variable pregnant, ¿Existe una mayor cantidad de mujeres que no están embarazadas Vs aquellas que han tenido tienen 3 embarazos?

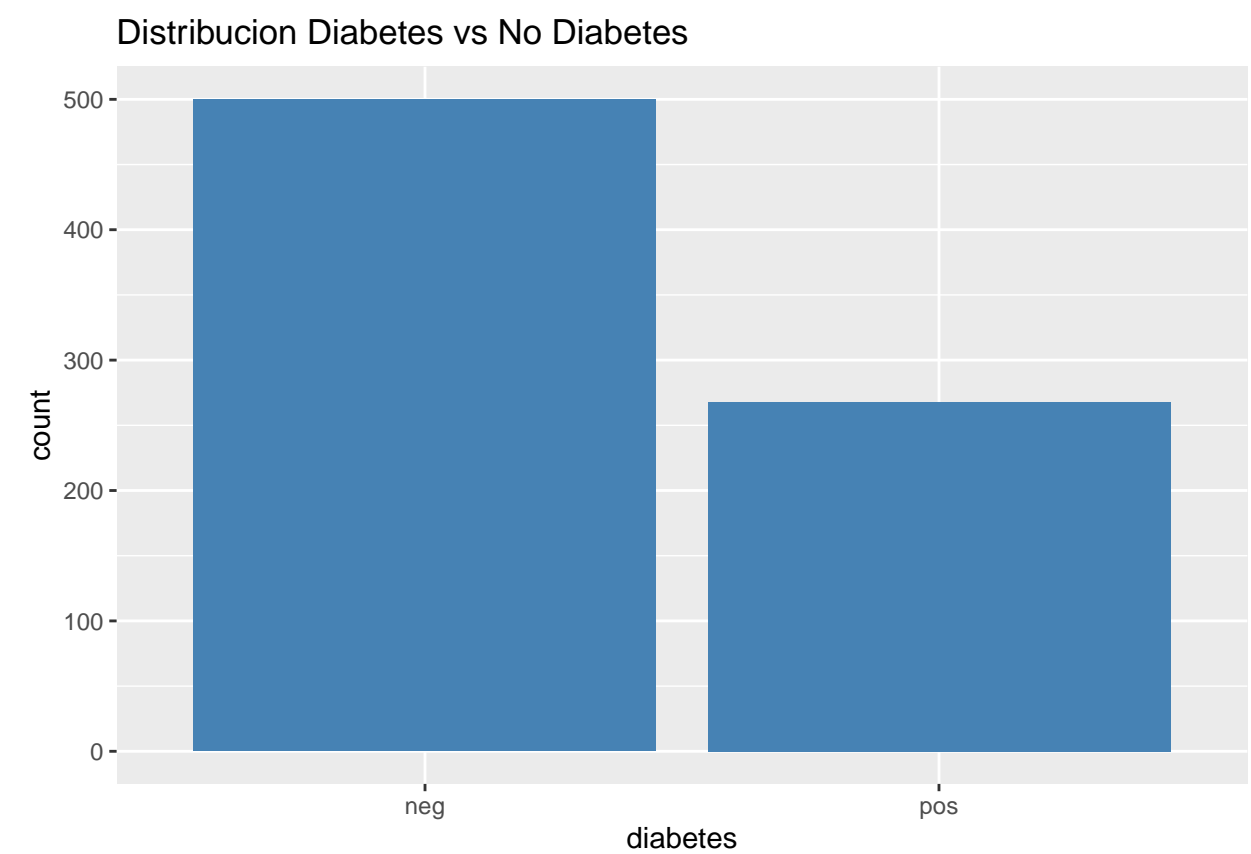
```
ggplot(diabetes, aes(x = pregnant)) + geom_bar(fill = "steelblue") +
  labs(title="Distribucion Pregnant")
```



R// A) Si, según la gráfica existen más de 100 mujeres que no han tenido embarazos y menos de 85 que han tenido 3 embarazos

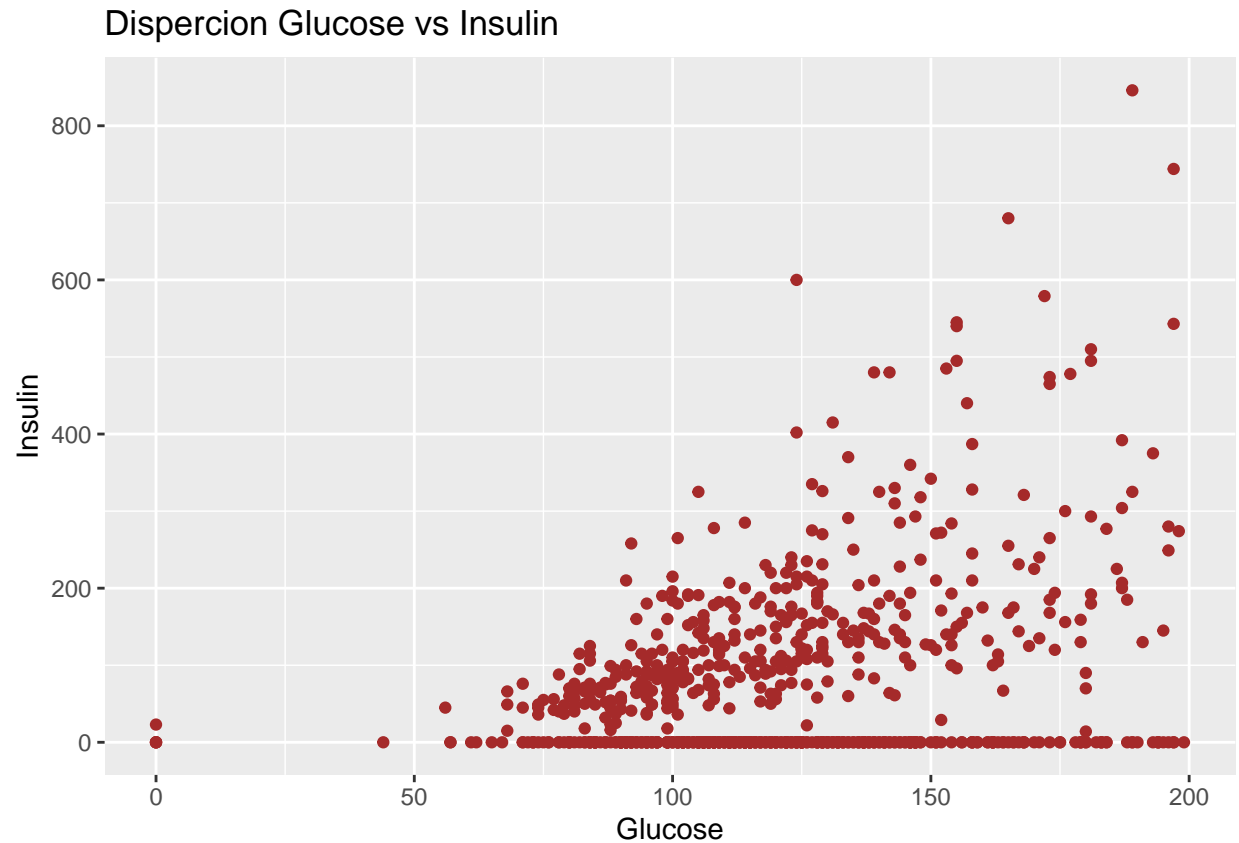
4. Crea un gráfico que muestre la cantidad de personas con diabetes vs no diabetes.

```
ggplot(diabetes, aes(x = diabetes)) + geom_bar(fill = "steelblue") +
  labs(title="Distribucion Diabetes vs No Diabetes")
```



5. Crea una gráfica de dispersión donde se muestre las variables de glucose e insulin. ¿Qué podemos inferir de esta gráfica?

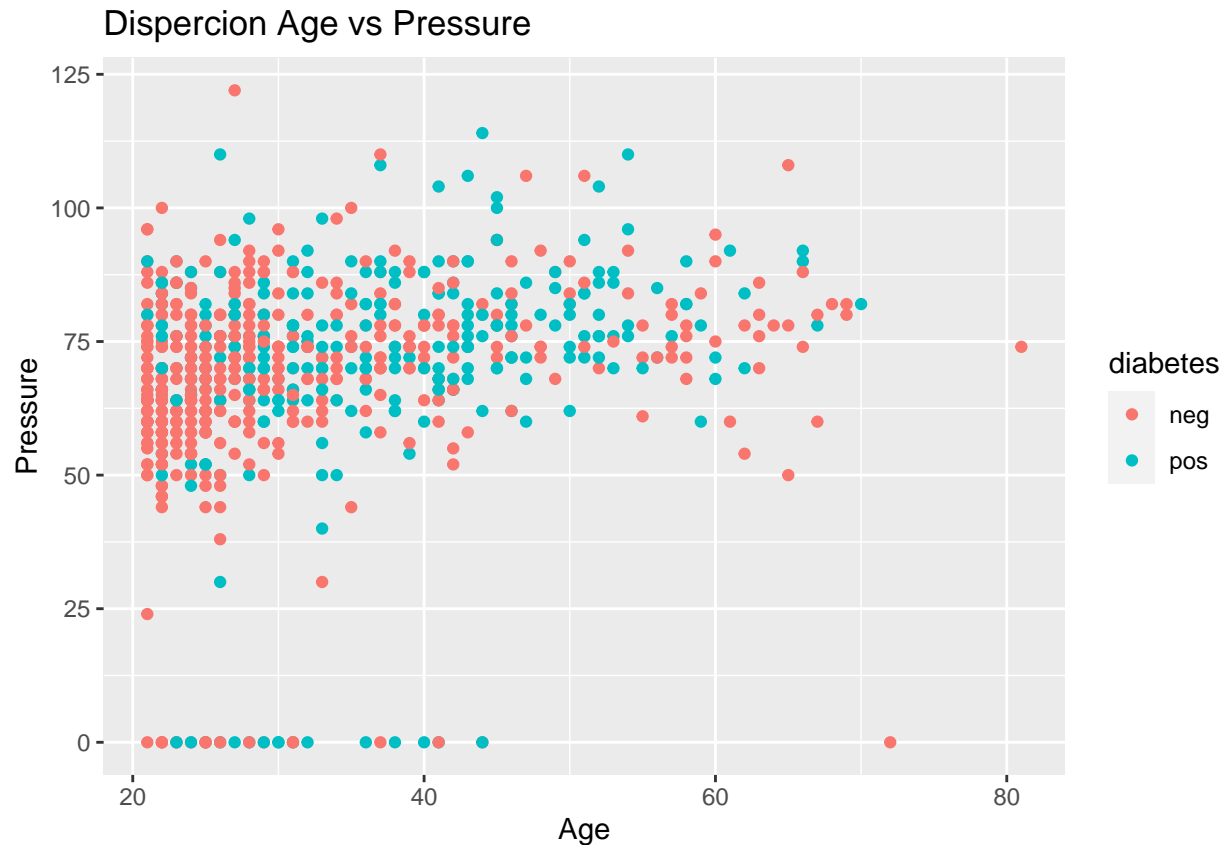
```
ggplot(diabetes, aes(x = glucose, y = insulin)) + geom_point(color = "brown") +  
  labs(title="Dispercion Glucose vs Insulin", x="Glucose", y = "Insulin")
```



*R*// La glucosa y la insulina son proporcionales entre sí, además de que la mayoría de personas poseen una glucosa y una insulina igual o cercana a 100.

6. Crea una gráfica de dispersión donde se muestre las variables age y pressure y grafica los puntos teniendo en cuenta si tiene o no diabetes. ¿Qué podemos inferir de esta gráfica?

```
ggplot(diabetes, aes(x = age, y = pressure, color = diabetes)) + geom_point() +
labs(title="Dispercion Age vs Pressure", x="Age", y = "Pressure")
```



R// Las personas que están entre los 30 y 55 años sufren de diabetes y de 20 a 30 años está la mayor población sin diabetes, pero se presentan pocos casos con diabetes en ese rango de edad de 20 a 30. Adicional las personas de la tercera edad no presentan muchos casos de diabetes.

7. Crea una gráfica de dispersión donde se muestre las variables age y pressure teniendo en cuenta la variable diabetes

```
ggplot(data = diabetes, aes(x = age, y = pressure, color = diabetes)) +
  geom_point() + labs(title="Dispercion Age vs Pressure", x="Age", y = "Pressure")
```

