

Ejercicio # 2

Ronald Bailey

2023-05-28

Ejercicio #2:

Para este ejercicio se le solicita que desarrolle las siguientes actividades utilizando RStudio Con el dataset Admissions adjunto a este laboratorio realice lo siguiente:

```
dataset = read.csv("Admissions.csv")
```

1. Realice un análisis estadístico sobre todas las variables del dataset, recuerde que puede usar la función `summary()`.

```
summary(dataset)
```

```
##      Serial.No.      GRE.Score      TOEFL.Score      University.Rating
## Min.       : 1.0      Min.       :290.0      Min.       : 92.0      Min.       :1.000
## 1st Qu.:125.8      1st Qu.:308.0      1st Qu.:103.0      1st Qu.:2.000
## Median :250.5      Median :317.0      Median :107.0      Median :3.000
## Mean     :250.5      Mean     :316.5      Mean     :107.2      Mean     :3.114
## 3rd Qu.:375.2      3rd Qu.:325.0      3rd Qu.:112.0      3rd Qu.:4.000
## Max.     :500.0      Max.     :340.0      Max.     :120.0      Max.     :5.000
##      SOP           LOR           CGPA           Research
## Min.       :1.000      Min.       :1.000      Min.       :6.800      Min.       :0.00
## 1st Qu.:2.500      1st Qu.:3.000      1st Qu.:8.127      1st Qu.:0.00
## Median :3.500      Median :3.500      Median :8.560      Median :1.00
## Mean     :3.374      Mean     :3.484      Mean     :8.576      Mean     :0.56
## 3rd Qu.:4.000      3rd Qu.:4.000      3rd Qu.:9.040      3rd Qu.:1.00
## Max.     :5.000      Max.     :5.000      Max.     :9.920      Max.     :1.00
## Chance.of.Admit
## Min.       :0.3400
## 1st Qu.:0.6300
## Median :0.7200
## Mean     :0.7217
## 3rd Qu.:0.8200
## Max.     :0.9700
```

Según la descripción proporcionada por la función `summary()`, el conjunto de datos tiene varias variables:

- **Serial.No.** : Esta parece ser una identificación única para cada estudiante. Rango de 1 a 500, y parece estar distribuido uniformemente dado que la mediana y la media son aproximadamente la misma (250.5).

- **GRE.Score** : Este es el puntaje GRE de los estudiantes. La puntuación mínima es de 290, la máxima es de 340, y la media es de 316.5, lo que indica que los puntajes son relativamente altos en promedio.
- **TOEFL.Score** : Este es el puntaje TOEFL de los estudiantes. La puntuación mínima es de 92, la máxima es de 120 y la media es de 107.2, lo que también indica un nivel de puntajes relativamente alto.
- **University.Rating** : Esta es la calificación de la universidad. Las calificaciones van de 1 a 5, y la media está en 3.114, lo que sugiere que las calificaciones están distribuidas bastante uniformemente.
- **SOP** : Esta parece ser una medida de la calidad de la declaración de propósito de los estudiantes. Va de 1 a 5, y la media es 3.374, lo que indica una calidad promedio bastante buena.
- **LOR** : Esta es una medida de la calidad de las cartas de recomendación. También va de 1 a 5, y la media es 3.484, lo que también indica una buena calidad promedio.
- **CGPA** : Este es el promedio de calificaciones acumulado de los estudiantes. Va de 6.8 a 9.92, y la media es 8.576, lo que indica un alto nivel de rendimiento académico.
- **Research** : Este parece ser un indicador binario de si el estudiante ha hecho investigación o no. La media es 0.56, lo que indica que más de la mitad de los estudiantes ha hecho investigación.
- **Chance.of.Admit** : Esta es la probabilidad de admisión. Va de 0.34 a 0.97, y la media es 0.7217, lo que indica que la probabilidad promedio de admisión es bastante alta.

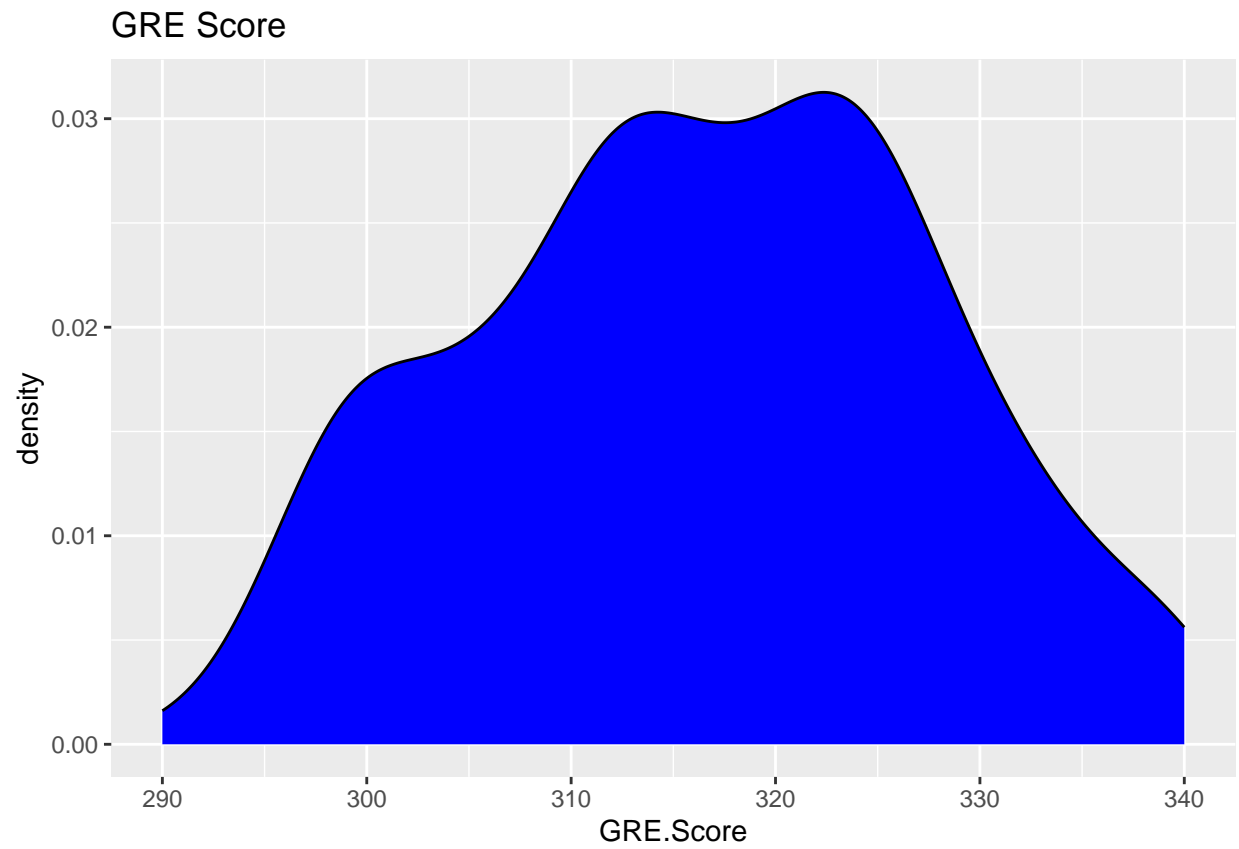
Cabe señalar que aunque estos análisis proporcionan información valiosa, es importante complementarlos con visualizaciones de datos y análisis adicionales, como verificar la correlación entre diferentes variables, realizar pruebas de hipótesis, etc.

2. Realice una gráfica de densidad para cada una de las variables numéricas en el dataset: GRE.Score, TOEFL.Score, CGPA y Chance of Admit.

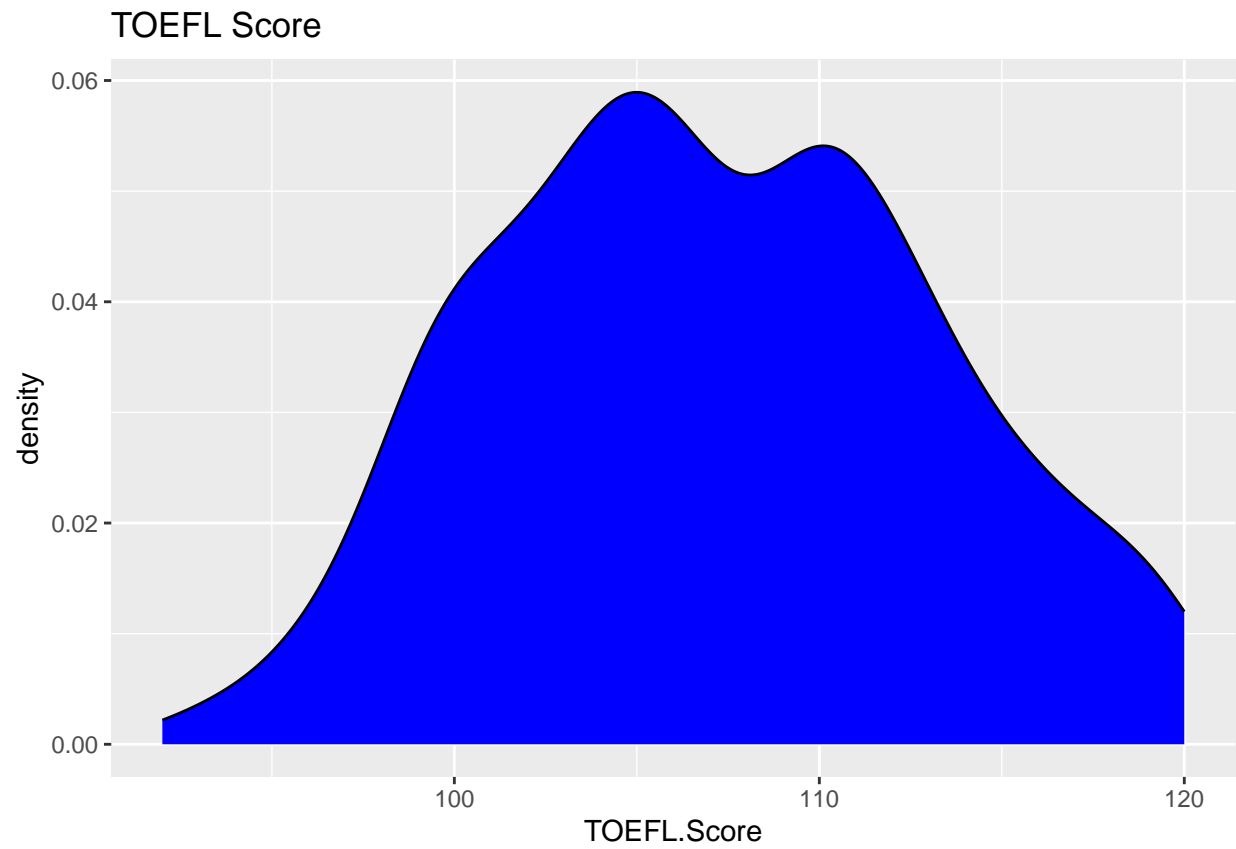
```
# Cargar la librería ggplot2
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

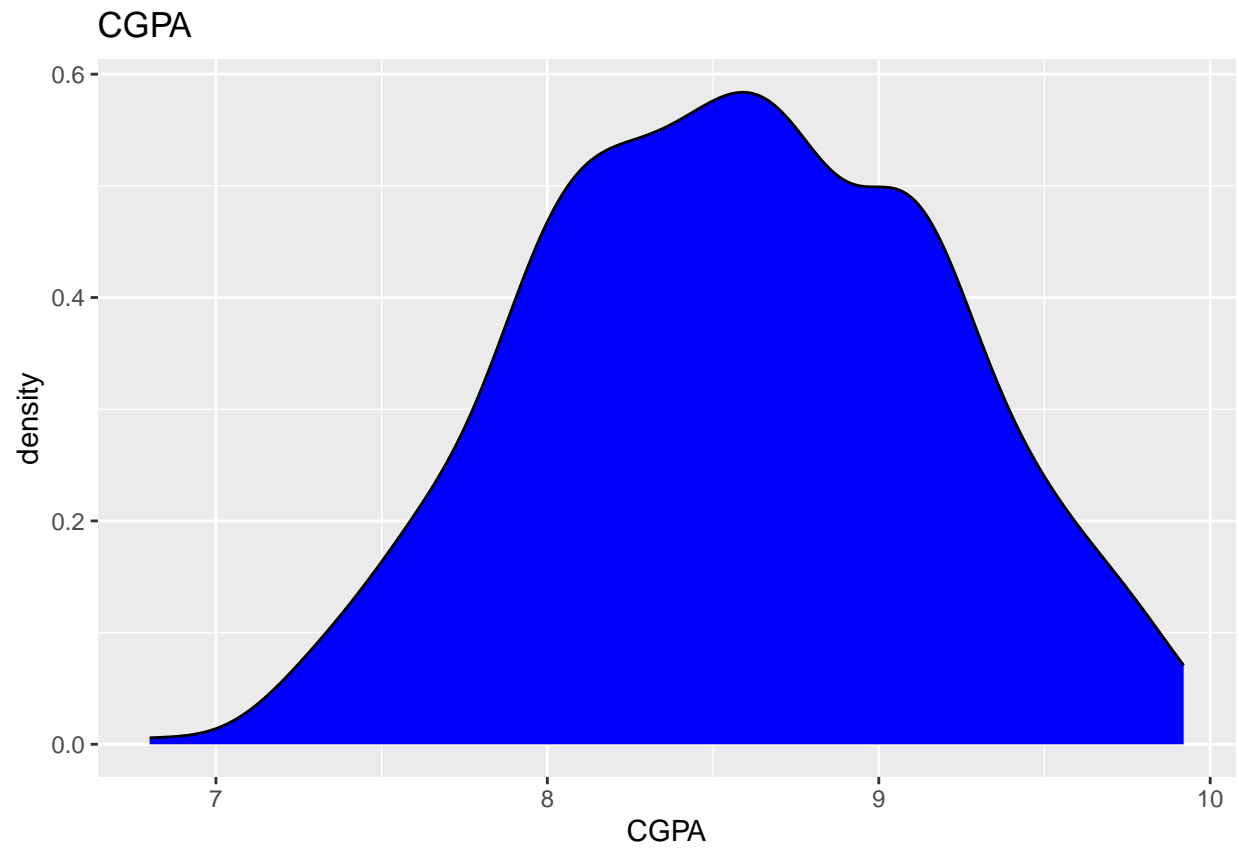
```
# Crear gráficos de densidad
ggplot(dataset, aes(x = GRE.Score)) + geom_density(fill = "blue") + labs(title = "GRE Score")
```



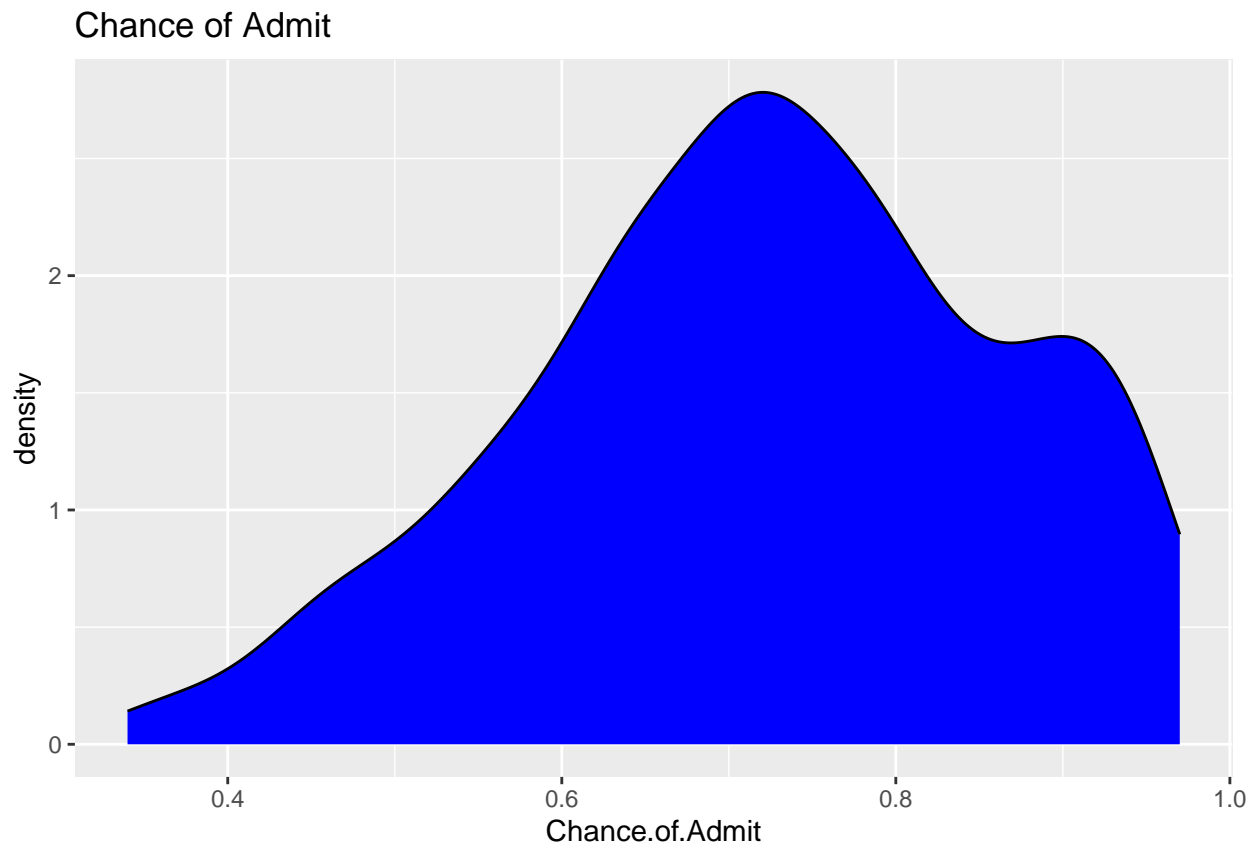
```
ggplot(dataset, aes(x = TOEFL.Score)) + geom_density(fill = "blue") + labs(title = "TOEFL Score")
```



```
ggplot(dataset, aes(x = CGPA)) + geom_density(fill = "blue") + labs(title = "CGPA")
```



```
ggplot(dataset, aes(x = Chance.of.Admit)) + geom_density(fill = "blue") + labs(title = "Chance of Admit")
```



```
# Cargar la biblioteca corrplot
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.2.3
```

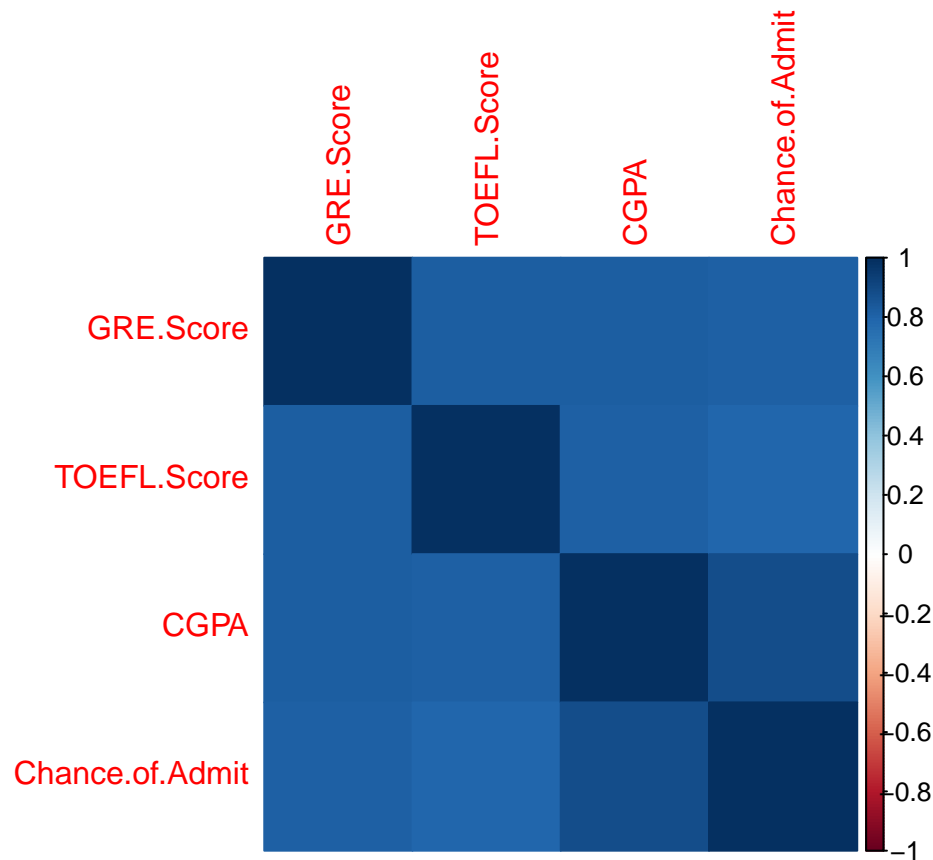
```
## corrplot 0.92 loaded
```

```
# Subconjunto de datos con las variables de interés
data_subset <- dataset[, c("GRE.Score", "TOEFL.Score", "CGPA", "Chance.of.Admit")]
```

```
# Calcular la matriz de correlación
cor_matrix <- cor(data_subset)
print(cor_matrix)
```

```
##           GRE.Score TOEFL.Score      CGPA Chance.of.Admit
## GRE.Score      1.0000000  0.8272004 0.8258780      0.8103506
## TOEFL.Score    0.8272004  1.0000000 0.8105735      0.7922276
## CGPA           0.8258780  0.8105735 1.0000000      0.8824126
## Chance.of.Admit 0.8103506  0.7922276 0.8824126      1.0000000
```

```
# Crear el mapa de calor de correlación
corrplot(cor_matrix, method = "color")
```



4. Realice comentarios sobre el análisis estadístico de las variables numéricas y la gráfica de correlación.

- GRE.Score y TOEFL.Score tienen una correlación de 0.827, lo que indica una relación lineal positiva fuerte. Esto significa que a medida que las puntuaciones GRE aumentan, también lo hacen las puntuaciones TOEFL, y viceversa.
- GRE.Score y CGPA tienen una correlación de 0.826, lo que también indica una fuerte relación lineal positiva.
- GRE.Score y Chance.of.Admit tienen una correlación de 0.810, lo que indica una fuerte relación lineal positiva.
- TOEFL.Score y CGPA tienen una correlación de 0.811, lo que indica una fuerte relación lineal positiva.
- TOEFL.Score y Chance.of.Admit tienen una correlación de 0.792, lo que indica una fuerte relación lineal positiva.
- CGPA y Chance.of.Admit tienen la correlación más fuerte en tu matriz de 0.882, lo que indica una muy fuerte relación lineal positiva.

```
# Utilizar un bucle para crear gráficos para todas las variables
for (var in as.vector(colnames(dataset))){
  p <- ggplot(dataset, aes_string(x = var, y = "Chance.of.Admit")) +
    geom_point() +
    labs(title = paste("Grafica de puntos", var, "vs Chance of Admit"),
         x = var,
```

```

    y = "Chance of Admit")
  print(p)
}

```

```

## Warning: 'aes_string()' was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with 'aes()'.
## i See also 'vignette("ggplot2-in-packages")' for more information.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

```

