

# Tarea 1

---

Victor Manuel Gómez Espinosa

12 de febrero de 2020

## 1. PROBLEMA 1

El procedimiento para resolver el primer problema para identificar regiones y áreas dentro de las regiones en los datos **olive**, consistió en primero explorar los datos con la herramienta ggobi ya que permite explorar los datos rápidamente, después cuando se encontraron los gráficos más representativos se recrearon utilizando Python.

Se utilizaron principalmente dos tipos de gráficos, los de ejes paralelos y los de matriz de dispersión, debido a que en el primero nos permite comparar al mismo tiempo todas las variables e identificar las bandas que estén más cargadas en determinadas variables para después, quedarnos con las variables más representativas y verificar si en la matriz de dispersión encontramos nubes de puntos que se separen.

### 1.1 REGIONES

Observe la Figura 1.1, note que las regiones 2 y 3 en el eje de eicosenoic se separan, y también están muy agrupados en los ejes oleic y linoleic. Compare esos componentes con Figura 1.2 y note como efectivamente las regiones 2 y 3 se separan de la 1.

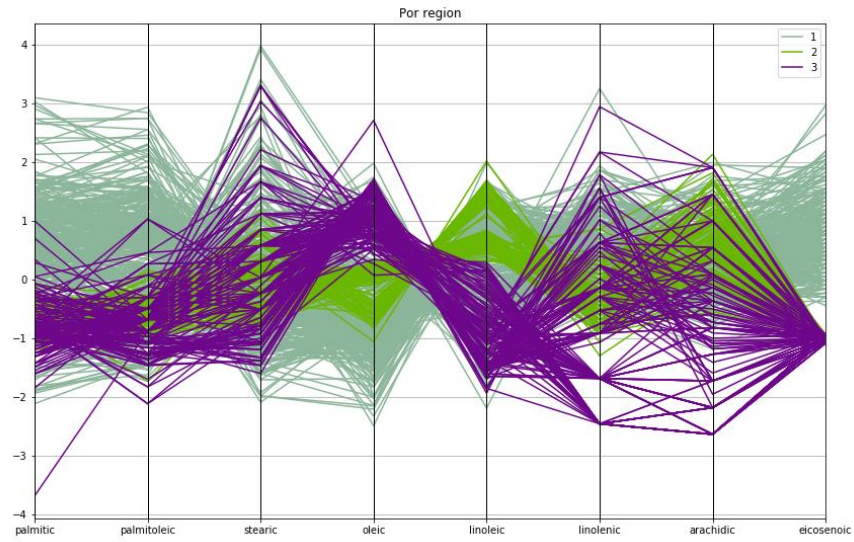


Figura 1.1: Gráfico de ejes paralelos para las regiones.

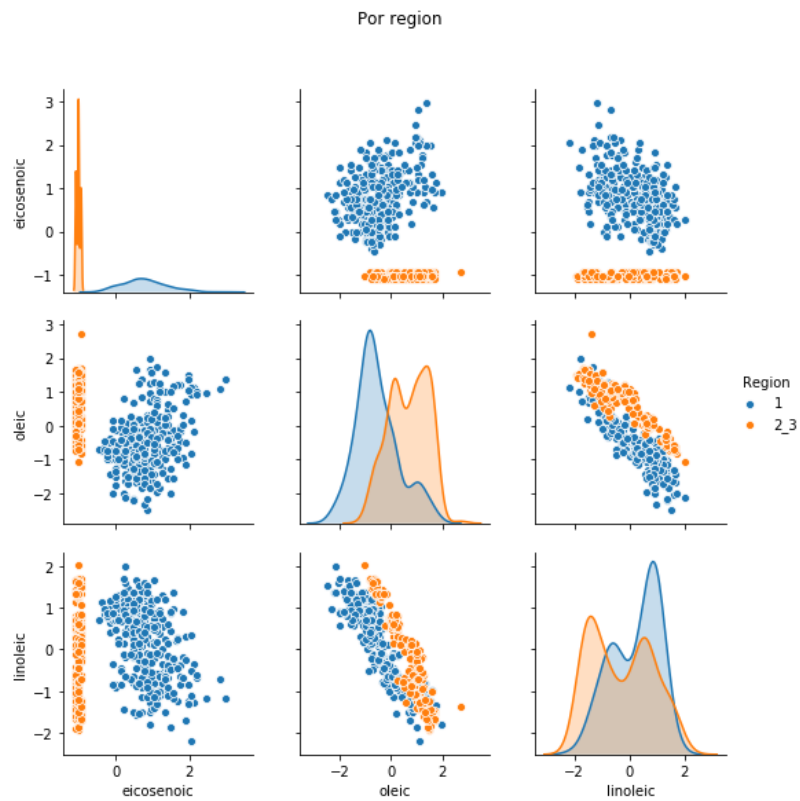


Figura 1.2: Matriz de dispersión para las regiones.

## 1.2 ÁREAS POR REGIÓN

Se siguió el procedimiento anterior para identificar las áreas en cada región.

### 1.2.1 ÁREAS REGIÓN 1

En la región 1 observe las Figuras 1.3 y 1.4 y note que se identificaron 3 áreas, North Apulia, South Apulia y se juntó las áreas de Sicily y Calabria en una misma categoría, ya que no hay diferencia notable.

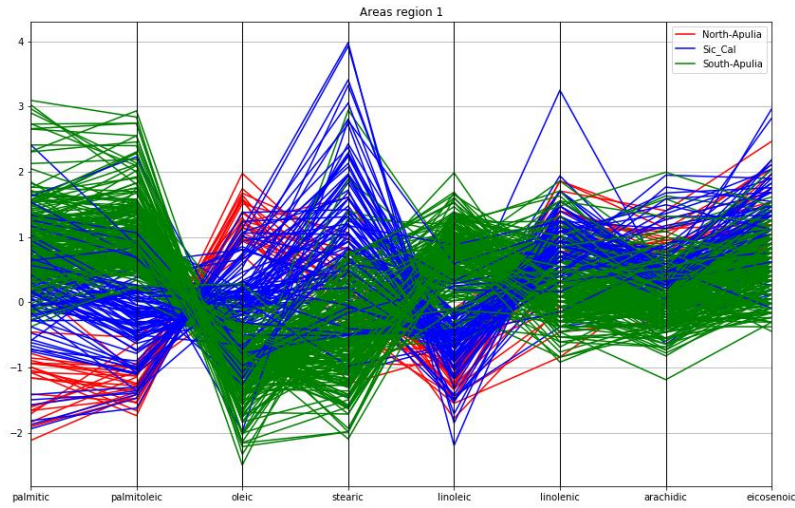


Figura 1.3: Gráfico de ejes paralelos para las áreas de la región 1.

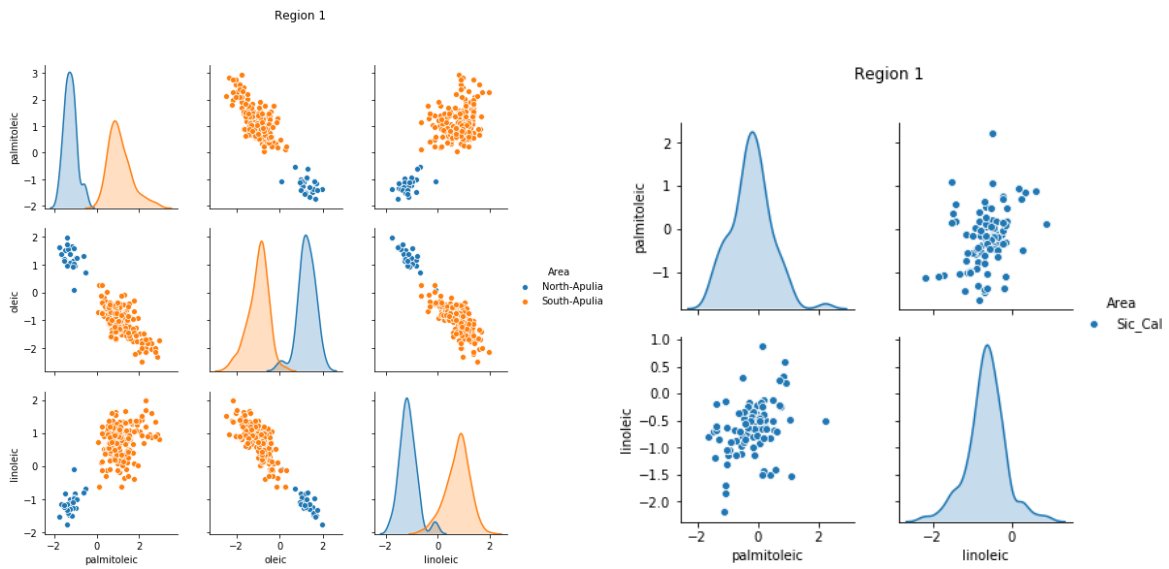


Figura 1.4: Matriz de dispersión para las áreas de la región 1.

### 1.2.2 ÁREAS REGIÓN 2

En la región 2 observe las Figuras 1.5 y 1.6 y note que se identificaron 2 áreas, Inland-Sardinia y Coast-Sardinia.

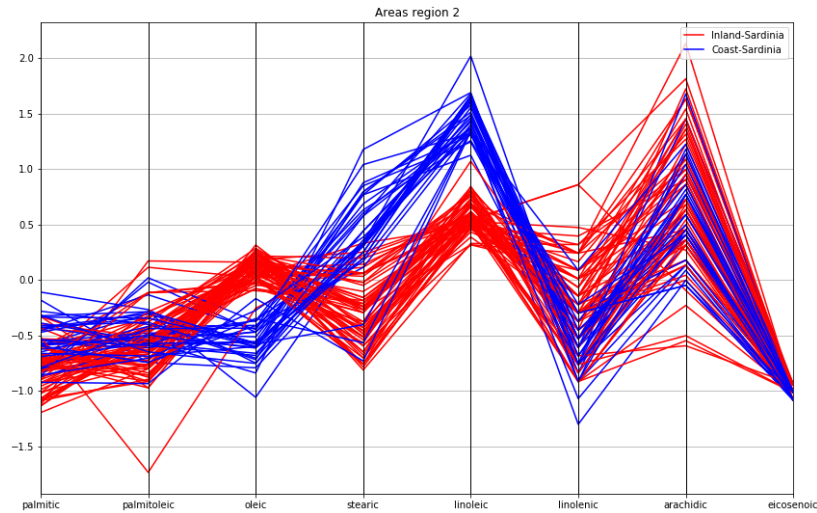


Figura 1.5: Gráfico de ejes paralelos para las áreas de la región 2.

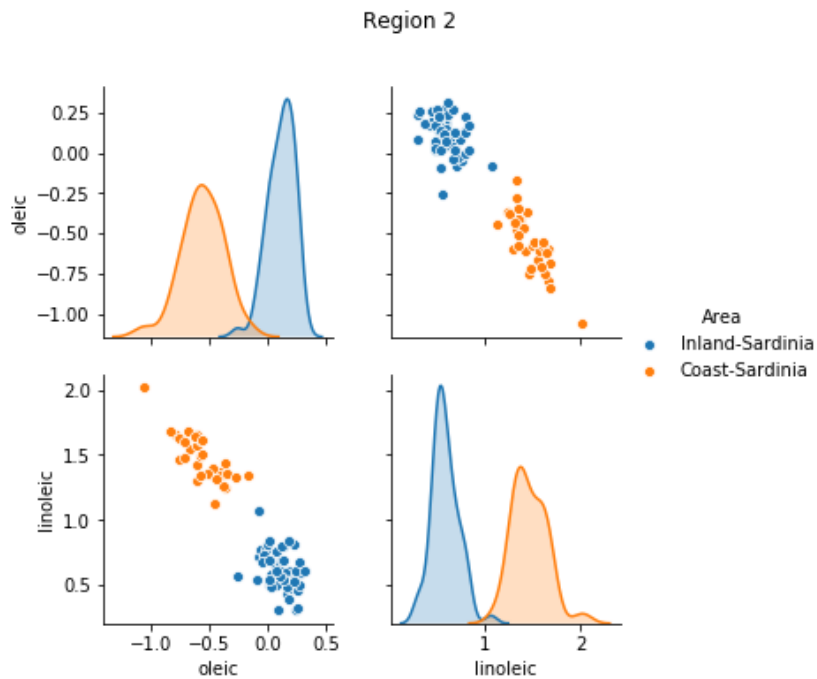


Figura 1.6: Matriz de dispersión para las áreas de la región 2.

### 1.2.3 ÁREAS REGIÓN 3

En la región 3 observe las Figuras 1.7 y 1.8 y note que se identificaron 2 áreas, Umbria y se juntó las áreas de East y West Liguria en una misma categoría, ya que se puede notar que se separa de Umbria, pero no por mucho porque está muy dispersa.

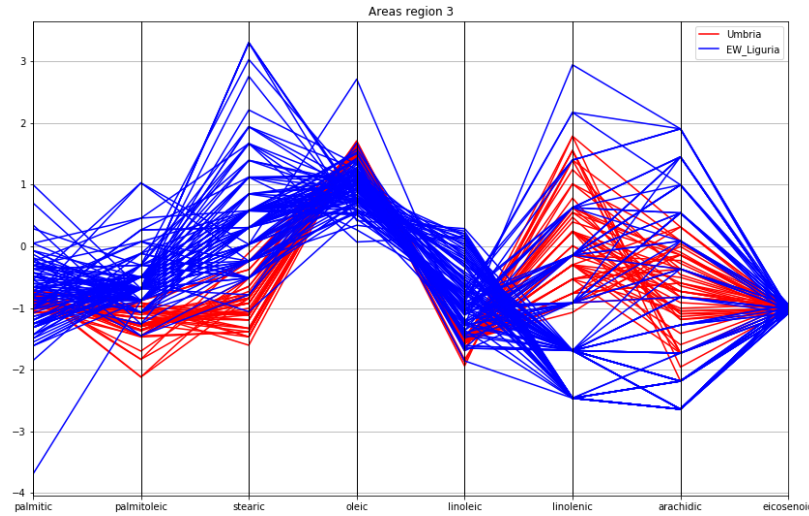


Figura 1.7: Gráfico de ejes paralelos para las áreas de la región 3.

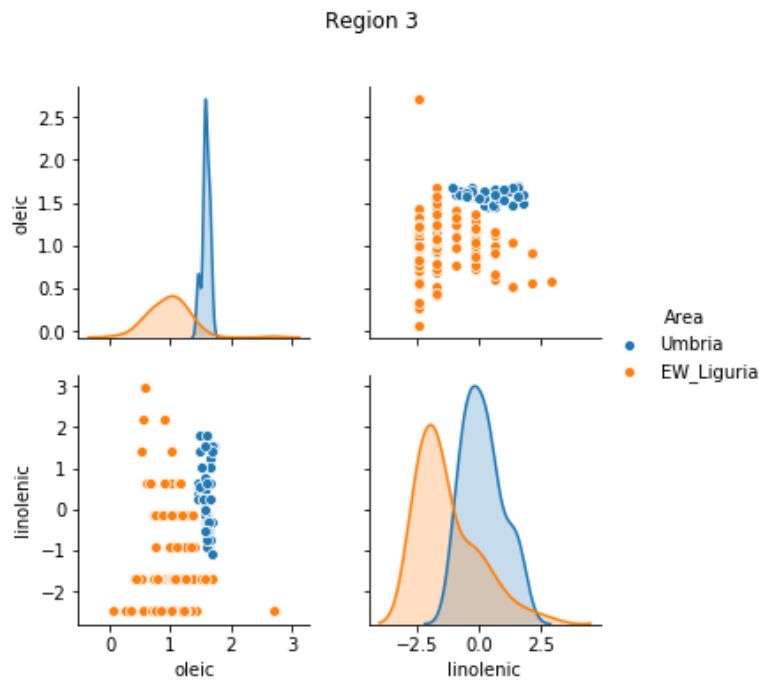


Figura 1.8: Matriz de dispersión para las áreas de la región 3.

## 2. PROBLEMA 2

### 2.1 PREGUNTAS DE INTERÉS

- ¿Alguna generación es más propensa a cometer suicidio?
- ¿La edad o sexo están relacionadas con los casos de suicidio?
- ¿Los habitantes de países con mayor HDI son menos propensos a cometer suicidio?

### 2.2 ESTRUCTURAS O CORRELACIONES

El procedimiento para resolver el segundo problema para analizar el fenómeno del suicidio en 101 países y particularmente en México y Estados Unidos, consistió en primero explorar los datos, todas las variables con las que contamos, que tipo de variable son (categóricas, numéricas), el rango de las variables, si había datos faltantes.

Con respecto a los datos faltantes, correspondientes a la variable HDI, se creó una copia de los datos para todas las variables únicamente donde tuviéramos datos disponibles de HDI para ver si estaba correlacionada con alguna otra variable que pudiera utilizar en su lugar, encontrando el Producto Interno Bruto per cápita, adicionalmente esto también sirvió para inspeccionar si el número de suicidios estaba correlacionado con alguna variable, resultando en una ligera correlación con la población (Figura 2.1 y 2.2).

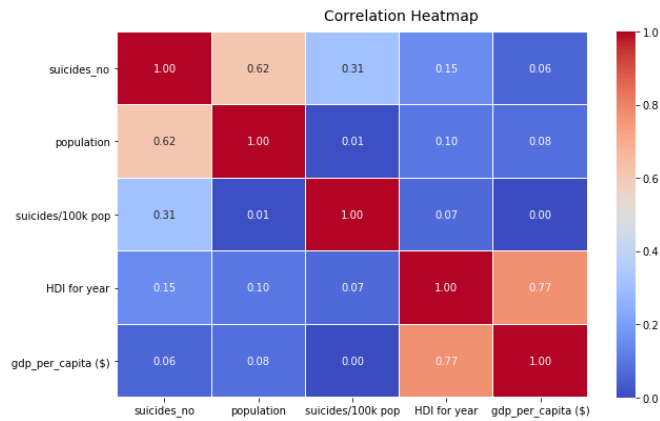


Figura 2.1: Mapa de correlación de las variables para todos los países.

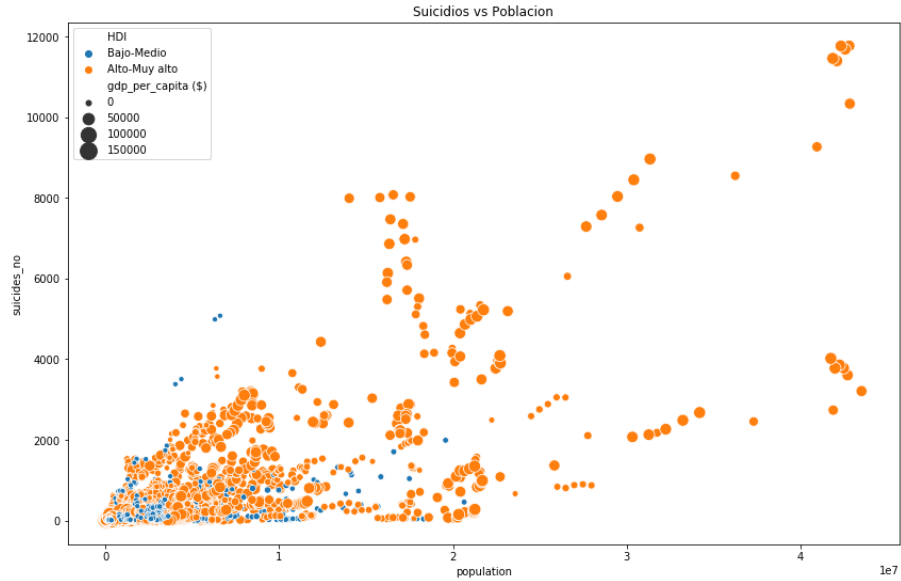


Figura 2.2: Gráfico de dispersión de suicidios contra población para los 101 países.

### 2.3 RESPUESTAS A LAS PREGUNTAS DE INTERÉS

Se observó que la generación Boomers es la que tiene mayor numero de suicidios, la cual podría ser la mas propensa a este fenomeno (Figura 2.3), predominantemente para el rango de edades de 35 a 54 años, predominantemente en hombres (Figura 2.4). Tambien se pudo observar que los paises con un lato HDI son mas propensos a cometer suicidio (Figura 2.2 y 2.4), lo cual parece interesante ya que lo que parece logico es que con un HDI alto las condiciones de vida de los individuos serían mejores y por lo tanto tendrían menores cifras de suicidio.

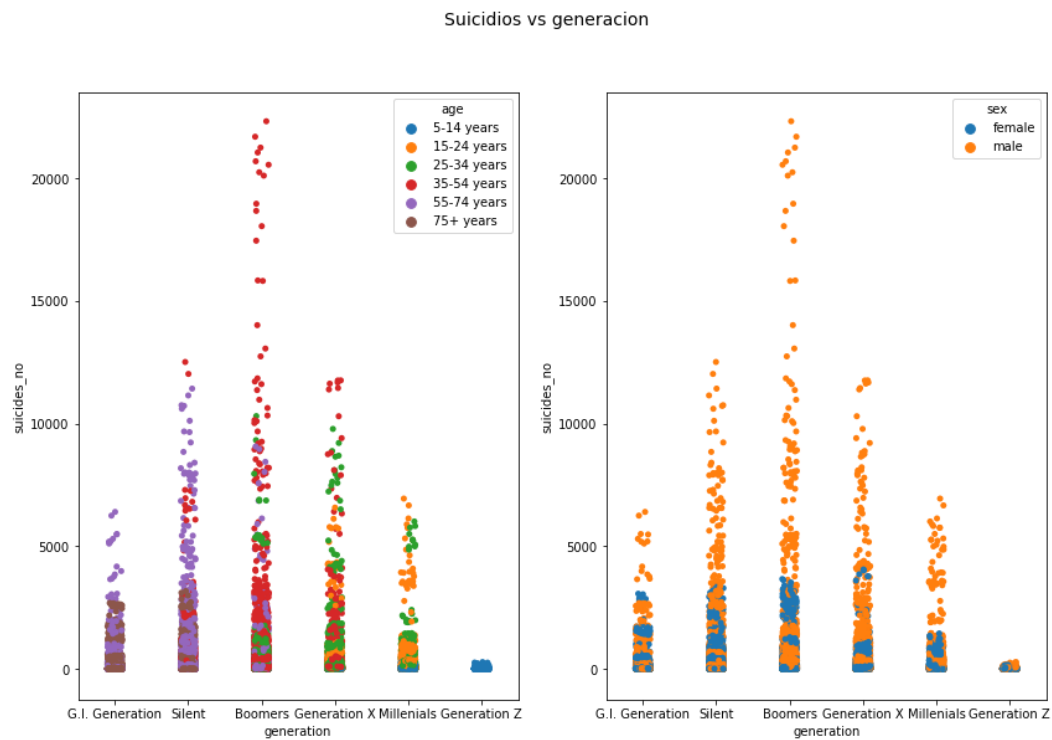


Figura 2.3: Suicidios por generación.

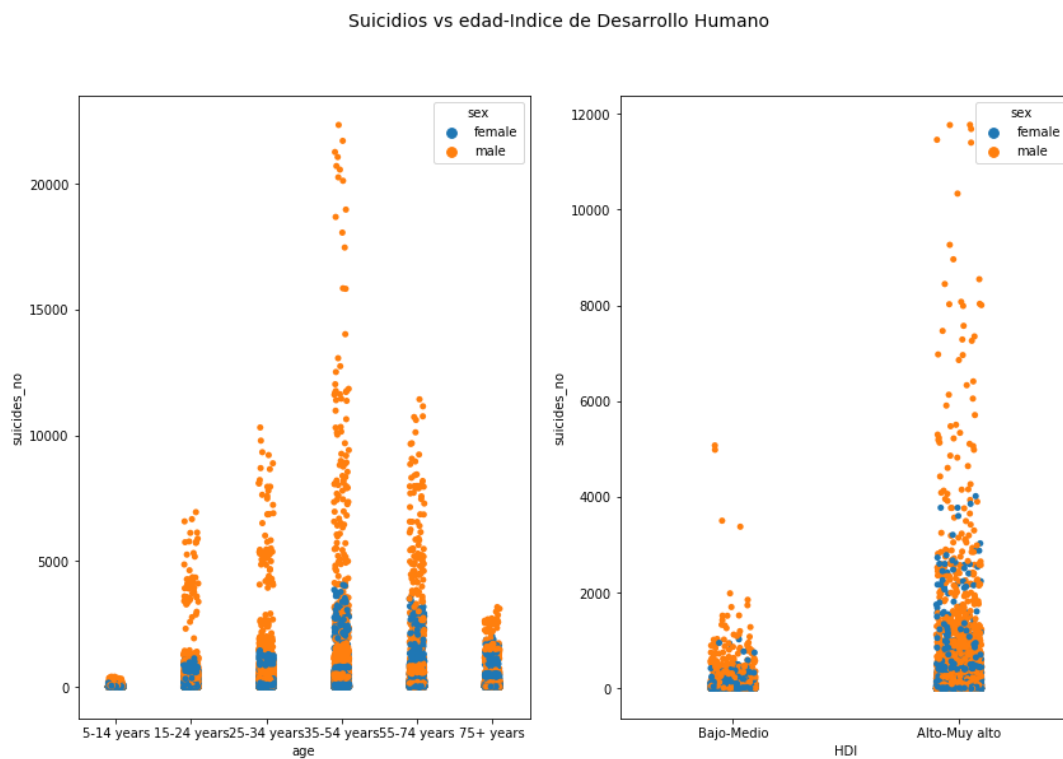


Figura 2.4: Suicidios por edades e HDI.



## 2.4 MÉXICO Y ESTADOS UNIDOS

Realizando una observacion de los datos, México y Estados Unidos se encuentran dentro de los 10 países con mayor numero de suicidios (Figura 2.5). Adicionalmente se encontró que para ambos países, el numero de suicidios muestra una correlacion ligera positiva con la poblacion (Figura 2.6).

Se encontró que los suicidios por sexo siguen siendo predominantemente en hombres tanto para México como Estados Unidos, de 35 a 54 años, teniendo en Estados Unidos una diferencia muy marcada en cuanto a sexo (Figura 2.7).

En cuanto a la generacion con mas tendencia al suicidio, se observo que en Estados Unidos prevalecen los Boomers, pero en México es la Generación X (Figura 2.8).

Respecto a la relacion del numero de suicidios con la población nuevamente se notó que para ambos países se tiene una correlación positiva, notando que existen 2 clases bien definidas por el sexo (Figura 2.9).

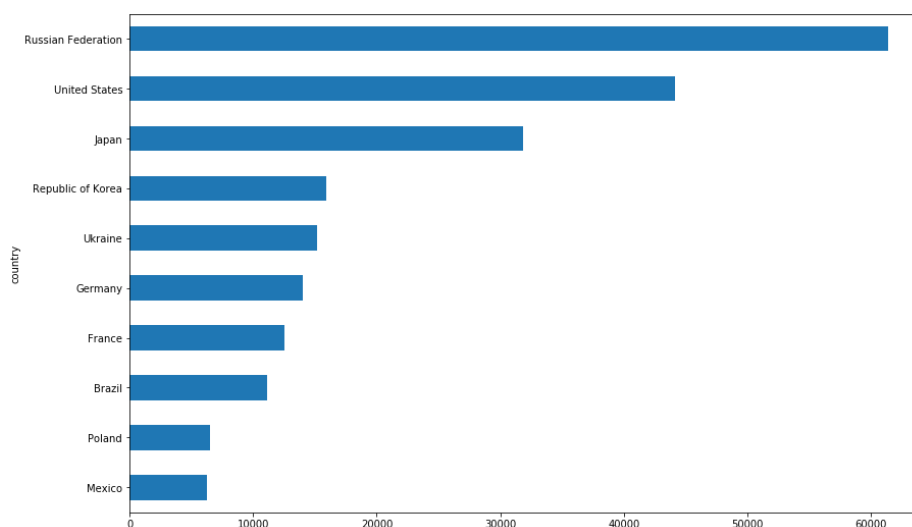


Figura 2.5: Top 10 países con mayor número de suicidios.

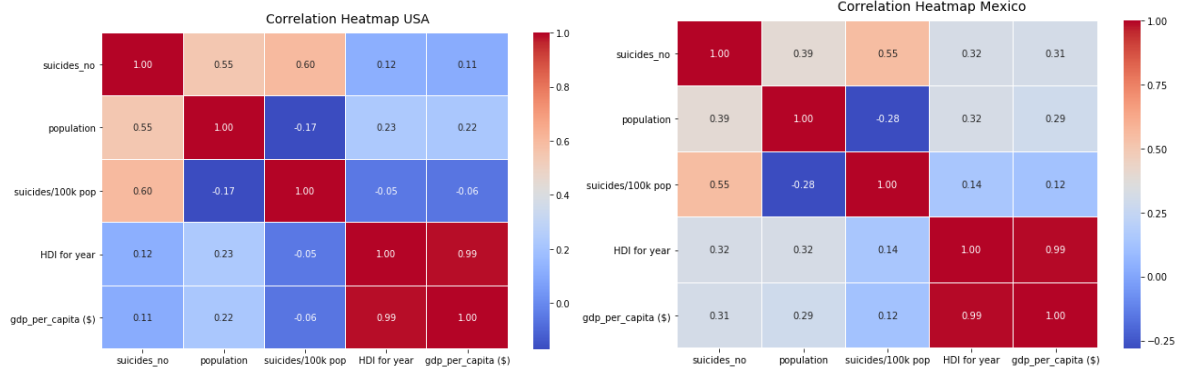


Figura 2.6: Mapa de correlación de las variables para Estados Unidos y México.

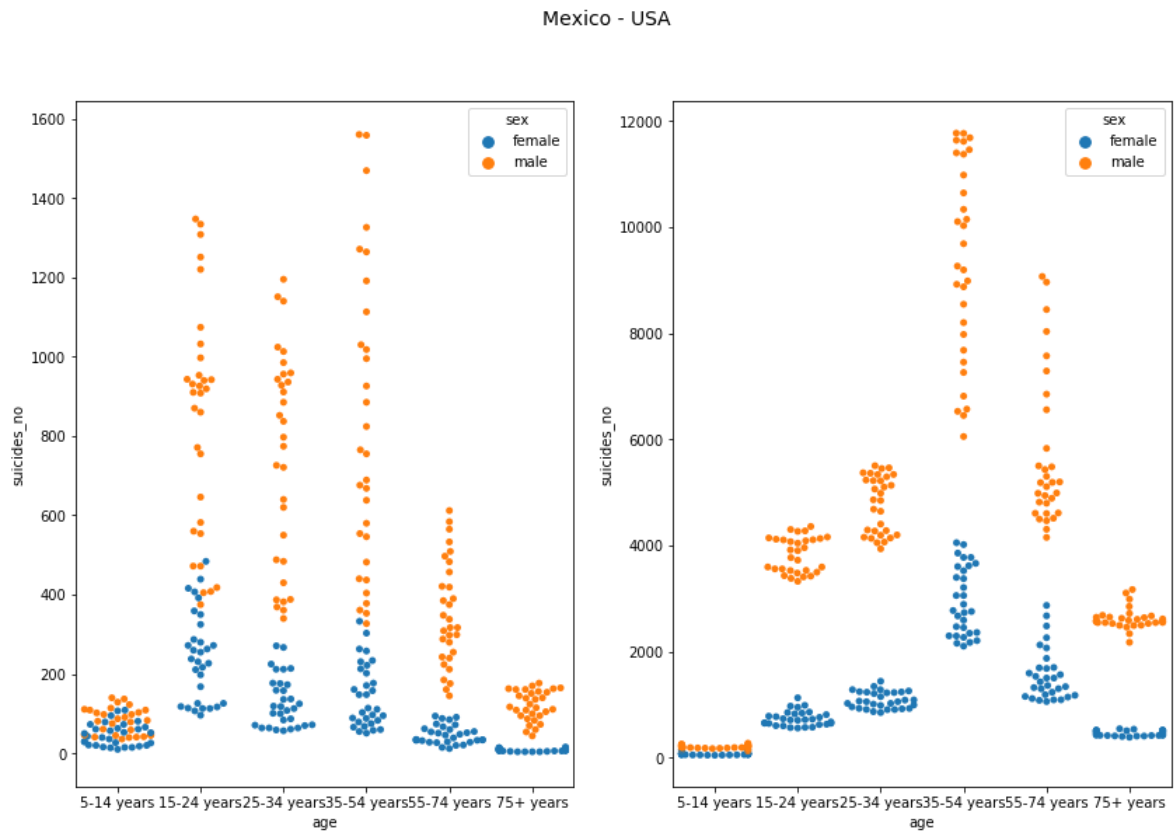
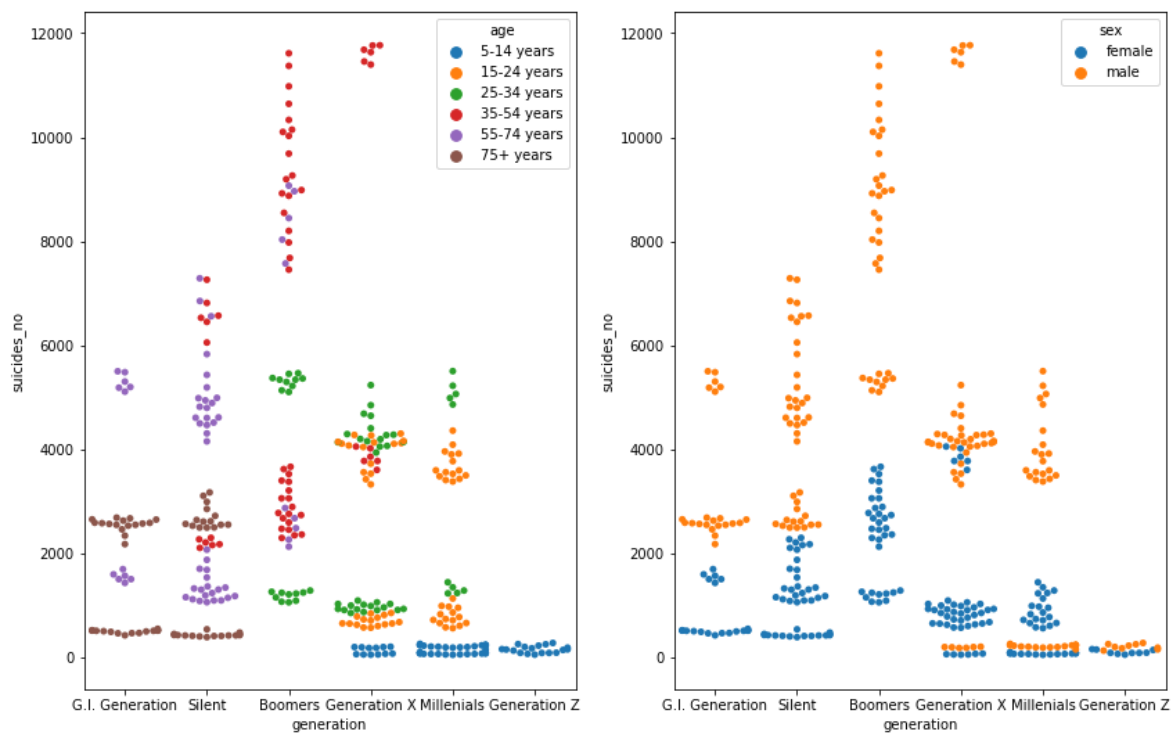


Figura 2.7: Suicidios por edad para México y Estados Unidos.

# USA



# Mexico

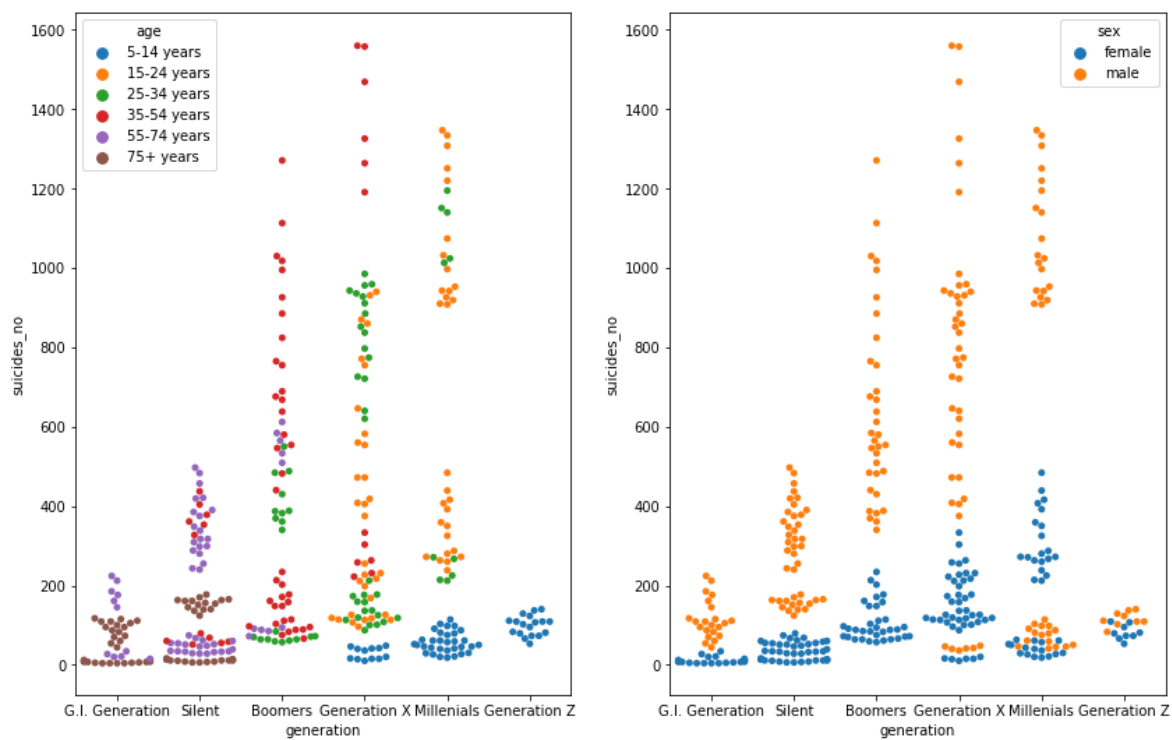


Figura 2.8: Suicidios por generación para Estados Unidos y México.

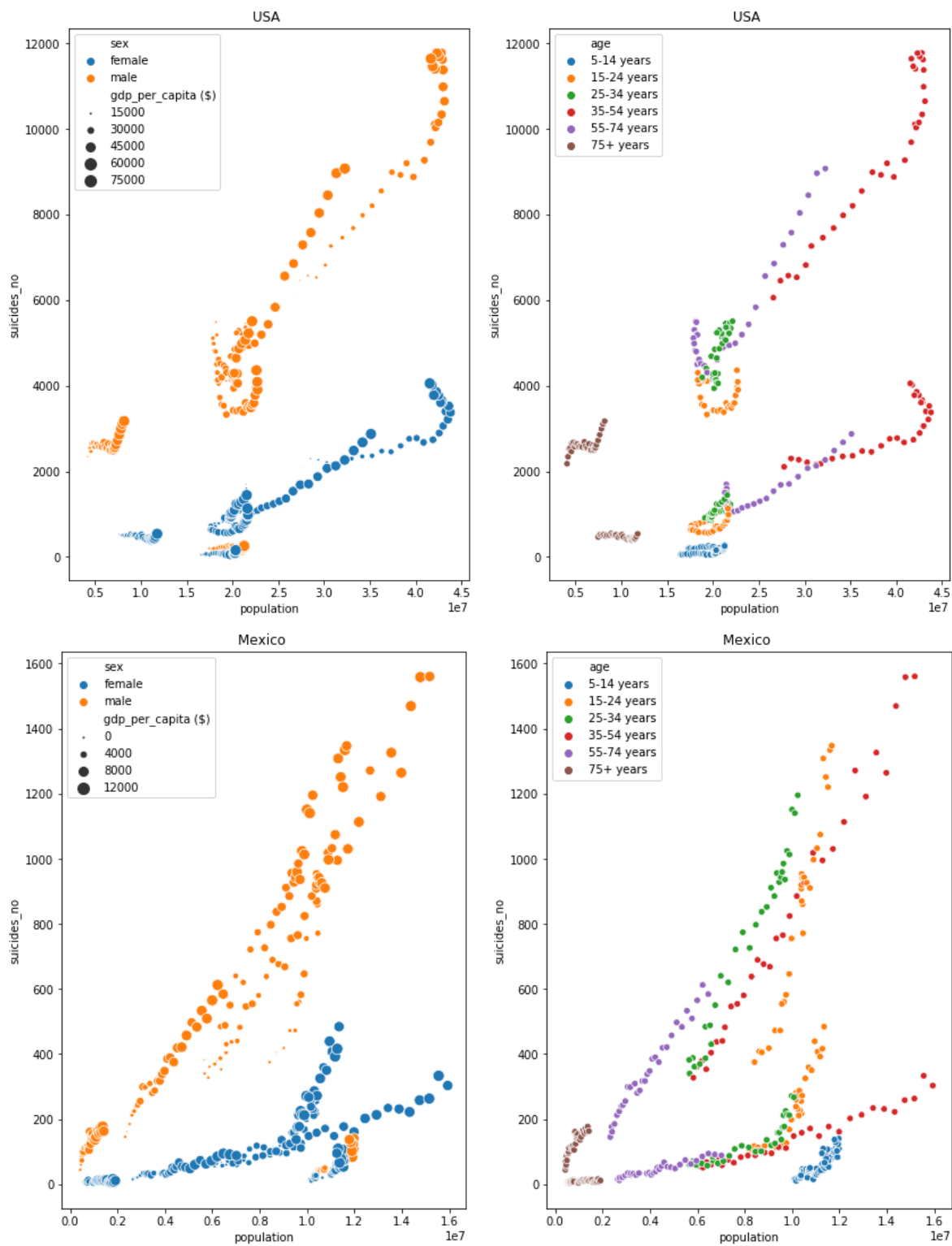


Figura 2.9: Suicidios por población para Estados Unidos y México.

## 2.5 OTRA INFORMACIÓN DISPONIBLE

Para continuar estudiando más a fondo este fenómeno sería más interesante tener información del número de personas que se han suicidado dado que un familiar o conocido se suicidó, o el número de personas que se han suicidado dado que han perdido sus trabajos o debido a la ruptura de una relación amorosa, o enfermedad. Considero que esta información refleja más el estado de salud de una persona y podría ayudar a comprender más el fenómeno.