

# Tarea 1

Enrique Santibáñez Cortés

9 de febrero de 2021

## 1. PROBLEMA 1

Concluye el ejercicio de análisis en los datos *olive* según los objetivos que discutimos en clase, es decir, tratar de identificar las regiones de los aceites y también las áreas dentro de cada región.

### 1.1. CLASIFICACIÓN DE REGIONES

Por lo visto en clase, tenemos que para clasificar las regiones de los aceites podemos utilizar las variables: palmitoleic, oleic, linoleic, eicosenoic. (ver Figura 1.1 y 1.2) Utilizando la variable eicosenoic se logra observar una separación entre la regiones 1 y el resto de las regiones, esta variable es la más relevante para identificar esta región. Ahora, la variable oleic nos ayuda a identificar la región 3 de las demás. Para tener una mejor clasificación podemos utilizar las variables palmitoleic y linoleic, por lo que podríamos clasificar los aceites de cada región con estas simples 4 variables.

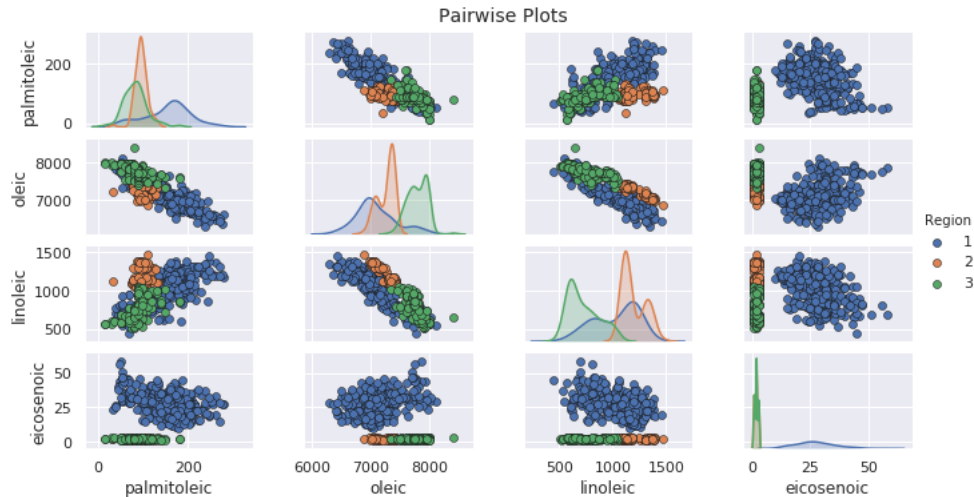


Figura 1.1: Scatteplot y FDE.

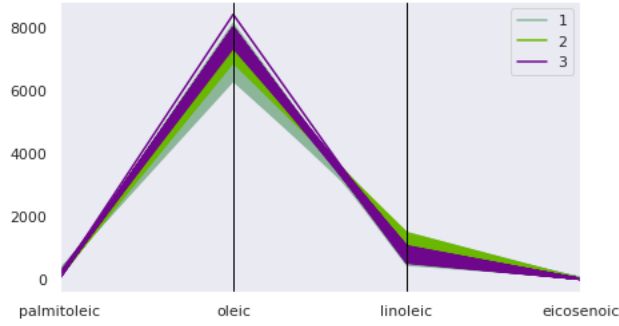


Figura 1.2: Parallel coordinate plots

## 1.2. CLASIFICACIÓN DE ÁREAS EN CADA REGIÓN.

Para cada región se encontraron de forma visual las variables que separaban las áreas sobre las regiones. **Para la región 1, no fue posible obtener una o un conjunto de variables que separaran las 4 áreas** (ver Figura 1.3 y 1.4). Si no existiera el Área Scity, las variables palmitic, palmitoleic y deic servirían para clasificar las otras 3 áreas de la región.

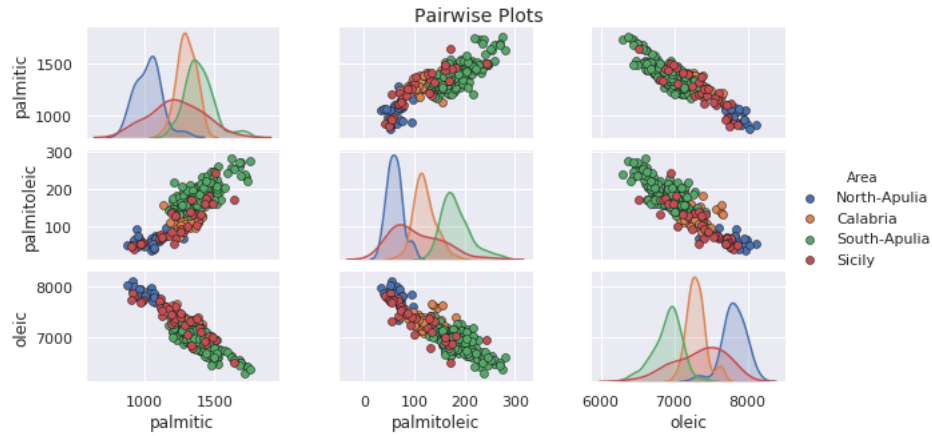


Figura 1.3: Scatteplot y FDE Región 1

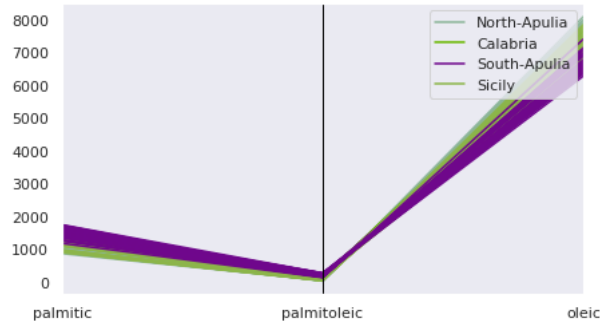


Figura 1.4: Parallel coordinate plots Región 1

Para la región 2, las variables que ayudan a clasificar las áreas de esta región son oleic y linoleic

(ver Figura 1.5 y 1.6). En este caso es muy notoria la separación de estas dos áreas dentro de la región 2.

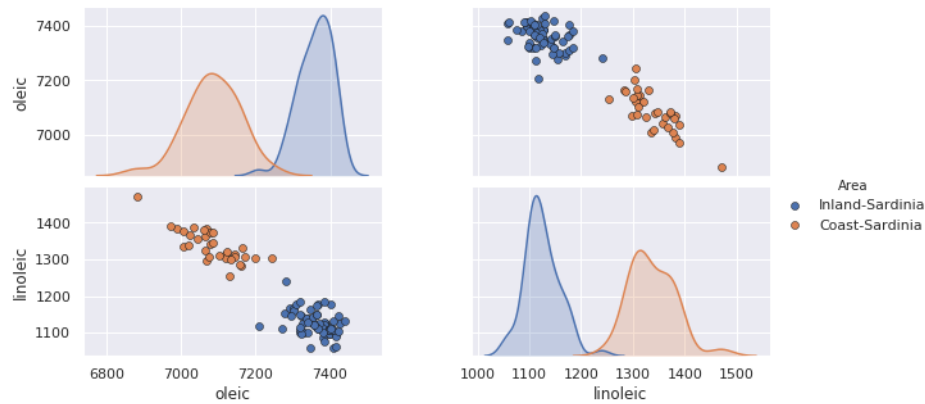


Figura 1.5: Scatteplot y FDE Región 2

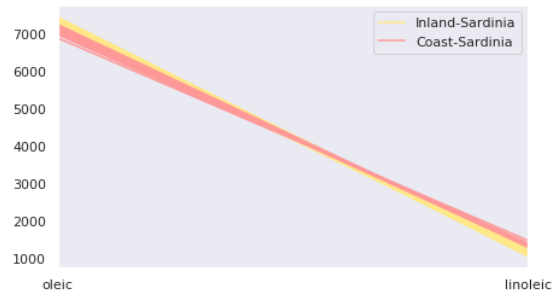


Figura 1.6: Parallel coordinate plots Región 2

Y por último, para la región 3 considero que las mejores variables para clasificar las áreas dentro de la región son palmitic y linoleic (ver figura 1.7 y 1.8). Observemos que la variable linoleic separa a la área West-Liguria de las otras dos, entonces haciendo esta separación podemos separar las otras dos áreas con la ayuda de la variable palmitic.

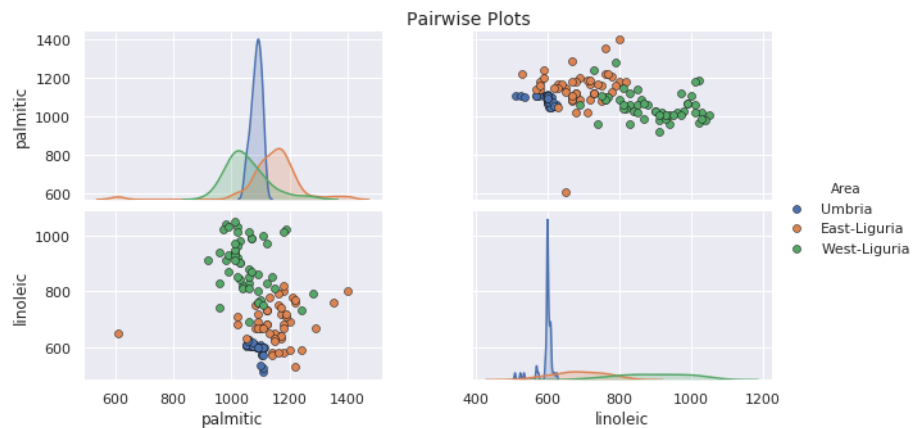


Figura 1.7: Scatteplot y FDE Región 3

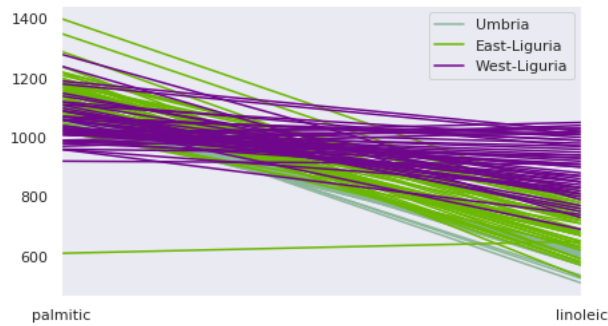


Figura 1.8: Parallel coordinate plots Región 3

## 2. PROBLEMA 2

Considera los datos que se encuentran en el archivo *suicide\_data.csv*, que compila información relacionada con suicidios en 101 países en diferentes años e incluye también algunos indicadores de desarrollo, como el Producto Interno Bruto (GPD) y percapita, Índice de Desarrollo Humano (HDI), entre otra información.

### 2.1. A) ¿QUÉ PREGUNTAS DE INTERÉS PUEDES FORMULAR PARA ANALIZAR ÉSTE FENÓMENO EN BASE A ESTOS DATOS?

Considero que las primeras preguntas a contestar serían: **¿Actualmente los suicidios por año son mayores que hace 20 años? ¿Cómo es la tendencia de los suicidios (o tasa de suicidios) a nivel mundial?** Esta pregunta es muy general, pero creo que nos daría un mucha información acerca de la importancia que tienen los suicidios actualmente.

Ahora, considerando las variables económicas podemos contestar las siguientes preguntas: **¿Existe una relación entre el desarrollo económico de un país y el número de suicidios(o la tasa de suicidios)?** Claramente se esperaría que si existía una relación, debido a que si consideramos que el desarrollo económico de un país esta asociado en sus recursos que se les asignan a cada sector del país (educación, salud, infraestructura, etc), los países bajo desarrollo económico en donde los recursos para el sector salud es menor esto implicaría que la identificación y el tratamiento de los suicidios serían más escasos que los países con mayor desarrollo económico.

Y por último, me gustaría contestar lo siguiente: **¿Existen algún rango de edad en la cuál las personas estén más propensas a suicidarse? ¿Qué genero esta más propenso a suicidarse? ¿Qué generación es más propensa a suicidarse? ¿esto es diferente en cada país?** Considero que las personas jóvenes-adultas estarían más propensos a suicidarse, esta hipótesis se basa en que a esa edad no se esta consiente de muchas cosas y es cuando las personas se desarrollan personal y profesionalmente lo que provoca que estén más expuesto a estrés, problemas, etc. Por otro lado, creo que la puede existe una gran diferencia entre los suicidios por genero en cada país, ya que podemos pensar que los países en los que los derechos de las mujeres no sean los adecuados o igualitarios que los hombres esto provoque una diferencia en la tasa de suicidios.

Dependiendo de las resultados podemos ampliar las preguntas considerando las variables económicas y las variables de edad y género, pero para no hacer el análisis tan complejo empezaremos con estas

preguntas.

## 2.2. B) REALIZA UN ANÁLISIS EXPLORATORIO Y DESCRIPTIVO DE LOS DATOS CON LAS HERRAMIENTAS QUE CONSIDERES APROPIADAS. ¿QUÉ ESTRUCTURAS O CORRELACIONES ENCUENTRAS?

El lapso de tiempo estudiado es de 1985 a 2016 lo equivale un total de 31 años. Ahora, el índice de desarrollo humano no esta reportado en 22 años específicos (Ver Figura 2.1), lo que se espera que esta variable no sea de gran importancia para este análisis.



Figura 2.1: Total de registros nulos HDI por año

Por otro lado, 24 países tienen menos de 20 años con registros (Ver Figura 2.2), no se filtraron es debido a que la gran mayoría de los países pertenecen a países con un bajo desarrollo por lo que desequilibraría el conjunto de datos y podríamos crear un sesgo en la información.

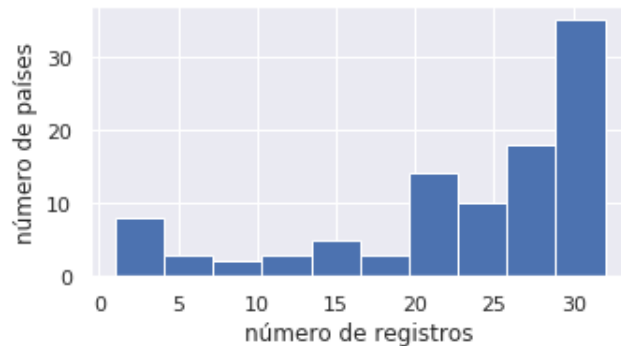


Figura 2.2: Total de registros nulos HDI por año

Un enfoque muy general de como se relacionan los datos a más bajo nivel es utilizando una matriz de correlaciones (ver Figura 2.3). Controlando al nivel más bajo de los datos, es decir, considerando el sexo y los rangos de edades por año de cada país podemos observar que no relación interesante de los datos a este nivel. Por lo que, se hará una análisis más extenso en la siguiente sección de este documento.

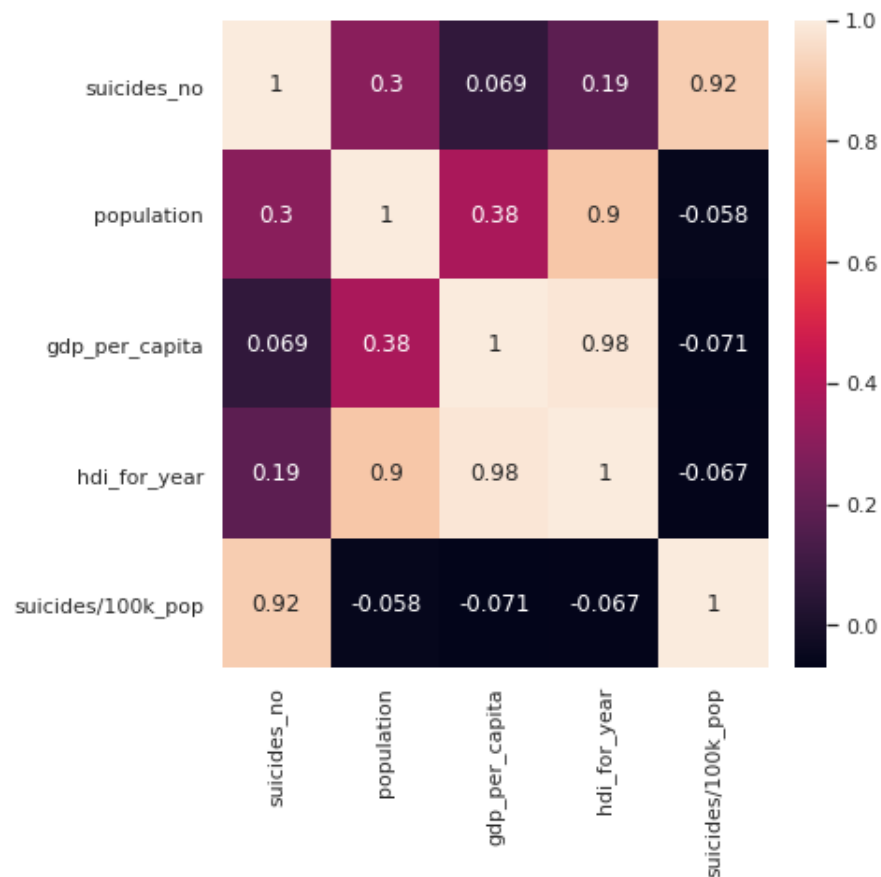


Figura 2.3: Matriz de correlación

### 2.3. C) TRATA DE RESPONDER LAS PREGUNTAS QUE FORMULASTE EN EL PRIMER INCISO. ¿PUEDES OBTENER CONCLUSIONES INTERESANTES?

El primer acercamiento fue considerar los totales y las tasas a nivel mundial de los suicidios (ver Figura 2.4, 2.5). Si analizamos el número total de suicidios no se observa claramente algún tendencia, pero esto se puede explicar debido a que los datos que tenemos no considera en cada año los mismos países. Observamos que en el año 2016 existe una gran caída en el número de suicidios, pero la explicación es simple, en este año solo se han reportado de 16 países lo que apenas representa el 10% del total de países analizados.

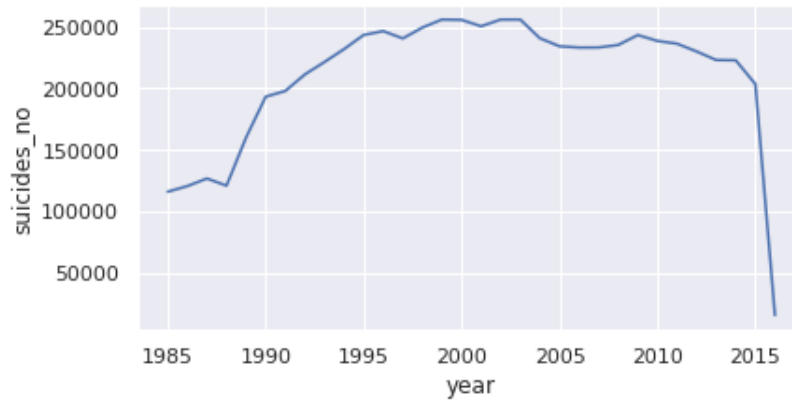


Figura 2.4: Total de suicidios por sexo

Por otro, considerando **la tasa de suicidios por cada 100 mil habitantes podemos observar que esta tiene una tendencia a la baja**. Para consolidar más este hecho se hará un análisis más profundo posteriormente.

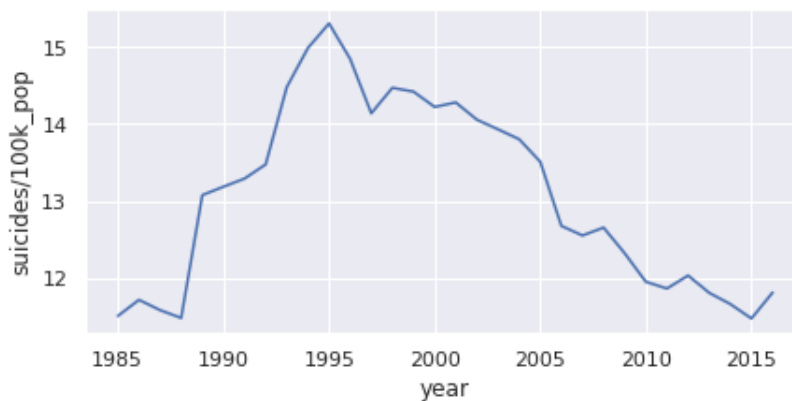


Figura 2.5: Tasa de suicidios por sexo

Ahora analicemos la tasa de los suicidios pero considerando el sexo (ver Figura 2.6), se observa claramente que **la tasa de suicidio de las mujeres es 6 o 7 veces más grande que la tasa de los hombres en todos los años**. Esta observación creo que es muy importante para tomar medidas de prevención del suicidio.

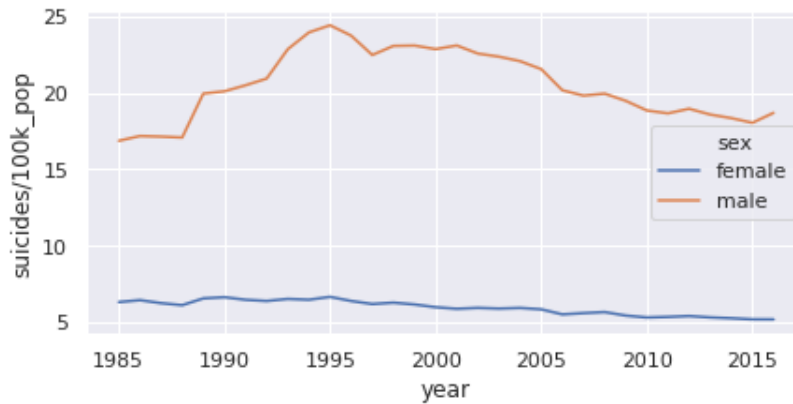


Figura 2.6: Tasa de suicidios por sexo, anual

Si consideramos las tasas de suicidios de los diferentes rangos de edades(ver Figura 2.7), podemos observar que **la tasa de suicidio es mayor para los rangos con las edades más grande**. Este resultado es muy interesante, ya que contradice un poco a la hipótesis planteada sobre que el rango de edad más propensos a suicidarse eran los universitarios.

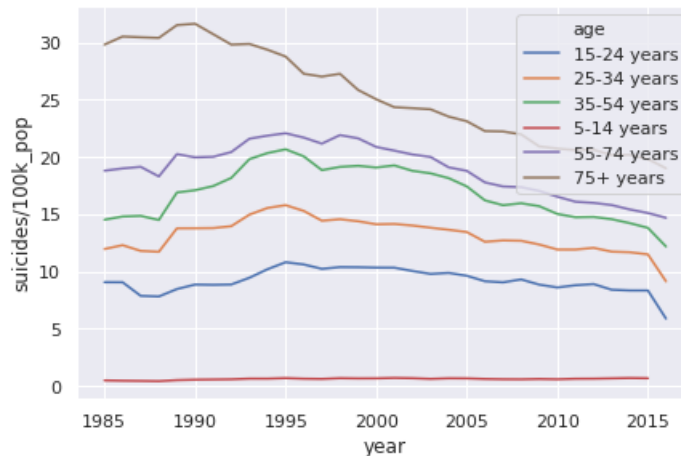


Figura 2.7: Tasa de suicidios por rango de edad

Por último, si consideramos la tasa de suicidios a nivel controlando por el sexo y los rangos de edad(ver Figura 2.8) podemos obtener conclusiones muy parecidas a las que cuando se analizan las variables por separado. Es decir, **las tasas de las mujeres es mayor en los rangos con edades más grandes al igual que los hombres, y además estas tasas son mayores para las mujeres en todos los casos**. De lo anterior podemos decir que el grupo más vulnerable son las mujeres de 75 o más. Cabe resaltar que solo el caso en que las mujeres están en un rango infantil, la tasa es menor que cualquier rango de hombre a excepción del rango infantil.



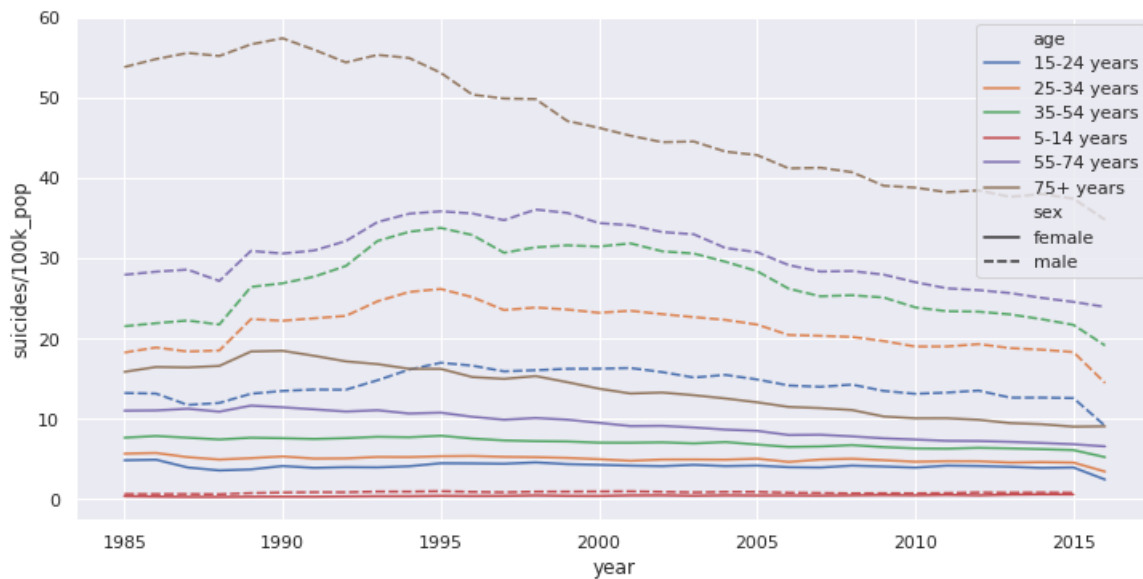


Figura 2.8: Tasa de suicidios por rango de edad y sexo

Realizando este mismo análisis pero considerando cada país por separado, cuando controlamos la tasa de suicidio por el sexo (ver Figura 2.10) podemos observar que el comportamiento que vimos anteriormente se sigue conservando. Es decir, **las mujeres tiene tasas de suicidios más altos que los hombres**. De lo anterior podemos decir que no importa el país, el sexo femenino esta más propenso a suicidarse.

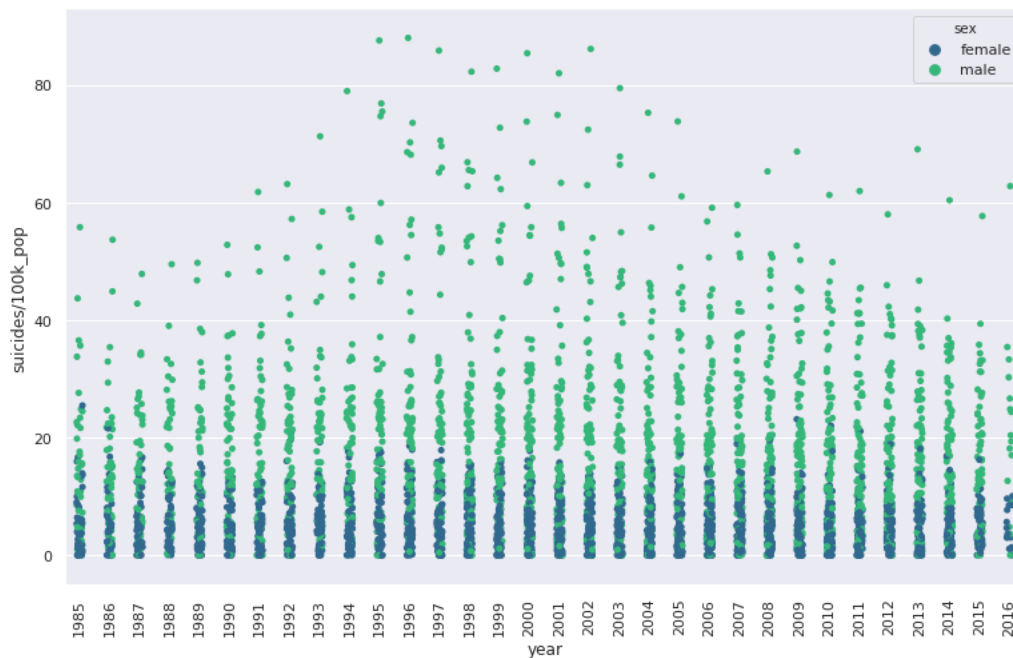


Figura 2.9: Tasa de suicidios por rango de edades, de cada país

Si controlamos a los países por los rangos de edades (ver Figura 2.10), podemos observar que el

rango infantil (5 a 14 años) son los que tiene una tasa menor. Además de lo anterior, no existe un patrón tan claro para las demás categorías como cuando se considera a nivel global.

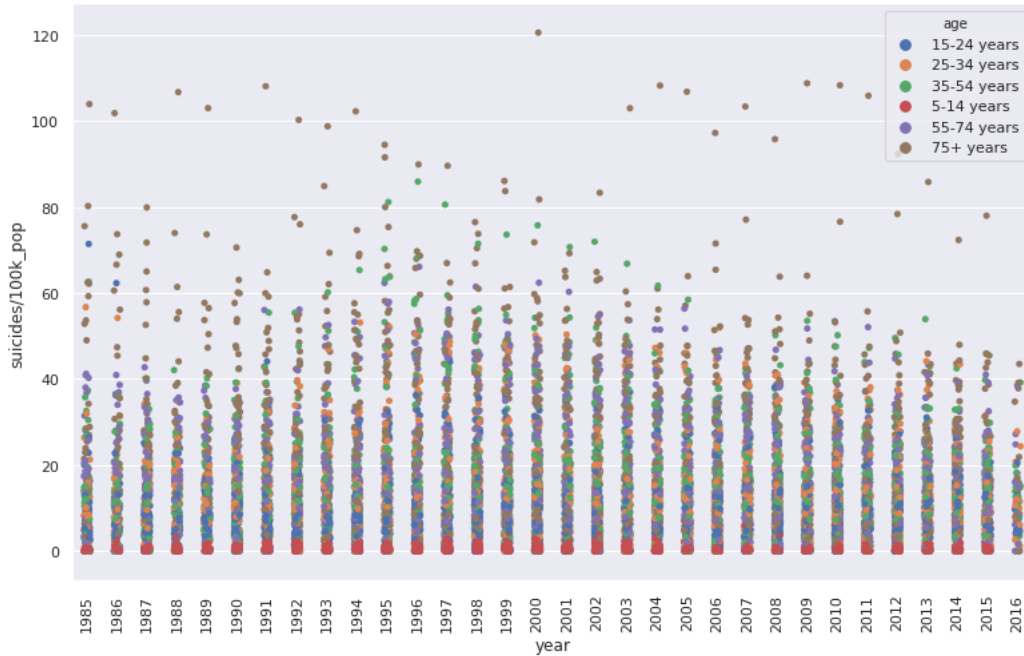


Figura 2.10: Tasa de suicidios por sexo, de cada país

El banco mundial sugiere que los países se pueden clasificar considerando el GDP per capital considerando los límites que se muestran en el Cuadro 2.1.

Gropo	GDP per capital
Bajos ingresos	< 1026
Ingresos medianos bajos	1026 – 3995
Ingresos medianos altos	3996 – 12375
Ingresos altos	12535

Cuadro 2.1: Clasificación según el GPD per capital

Entonces si consideramos la clasificación anterior y la tasa anual de cada país (ver Figura 2.11), no se observa un un patrón en los datos. Se esperaría que existiera este patrón, **pero si esto valida (al igual que las correlaciones de la sección anterior) que no existe una relación del GDP y la tasa de suicidios.**

La otra variable económica que tenemos es el HDI, como ya se comento anteriormente este indicador tiene años sin registros en ningún solo país por lo que graficamos los años en los que tenemos registros (ver Figura 2.12). De igual forma, **no podemos concluir algo relevante para esta variable.**

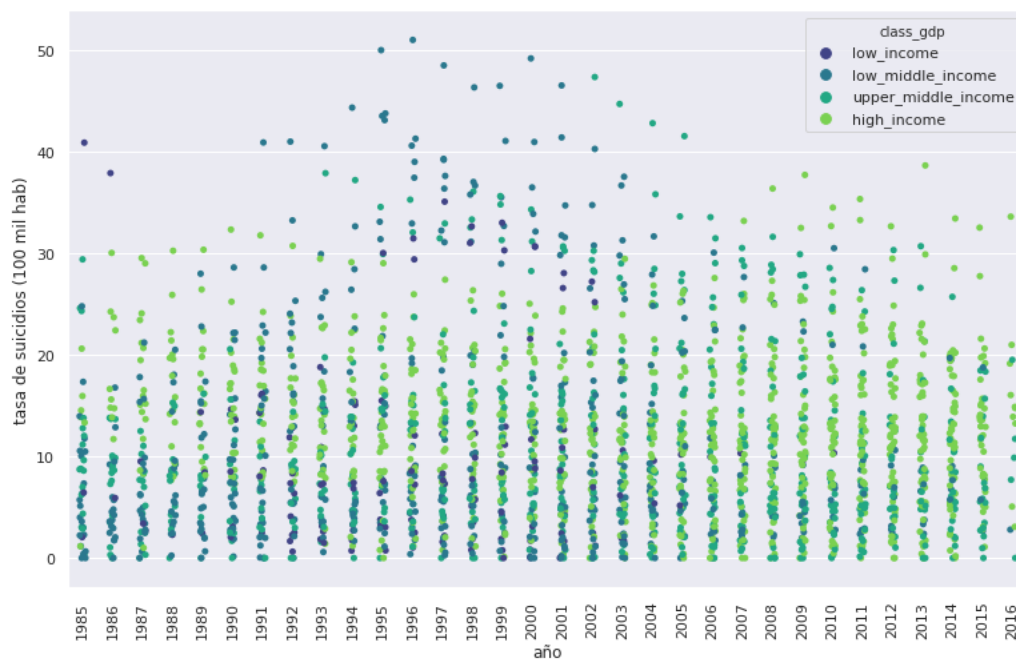


Figura 2.11: GDP per capital de los países clasificados

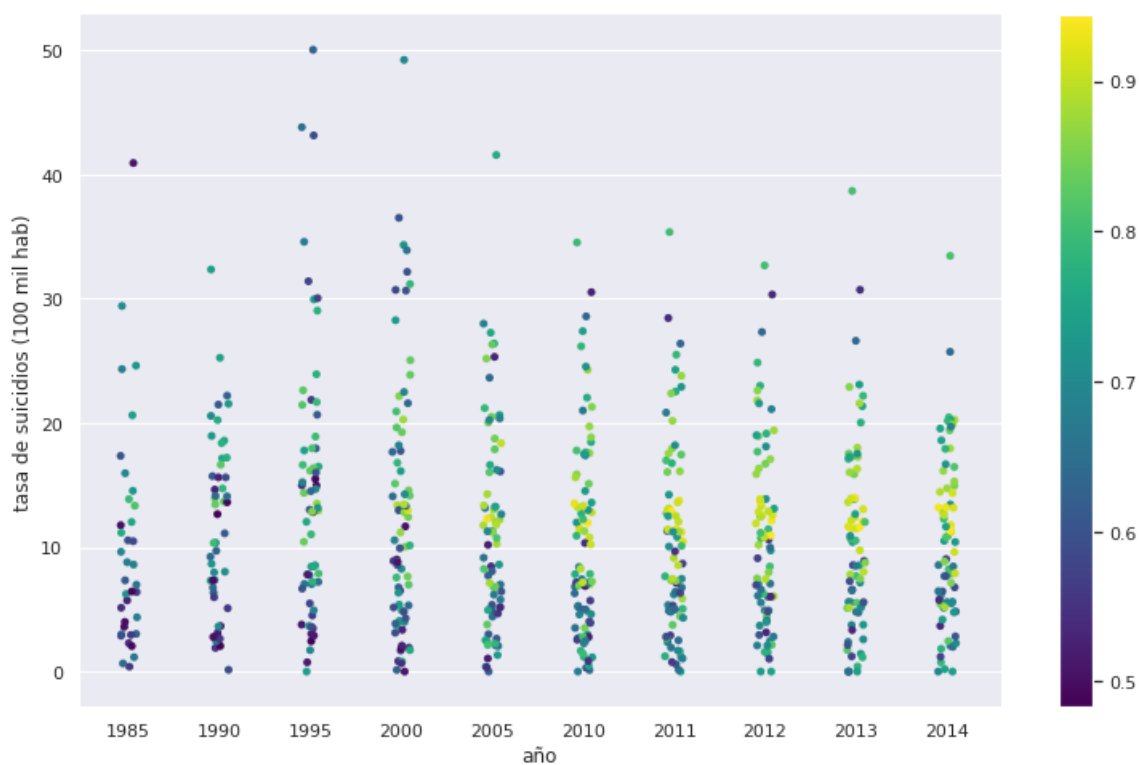


Figura 2.12: HDI de los países para algunos años

En conclusión, **claramente se observa la tasa de suicidios de las mujeres es mayor que la**

**de los hombres** este comportamiento se observa en la mayoría de los países. Y observamos que existe una relación del rango de edad y la tasa de los suicidios, pero esta relación no es la misma para todos los países.

- 2.4. D) CONSIDERA SOLAMENTE LOS DATOS DE MÉXICO Y ESTADOS UNIDOS.  
REALIZA LOS INCISOS 2b Y 2c POR SEPARADO Y COMPARA LOS RESULTADOS.  
¿QUÉ PUEDES CONCLUIR?

Veamos la tendencia del número total de suicidios para cada país (ver Figura 2.13), ambos países presentan una pequeña tendencia a la alza del número de suicidios.

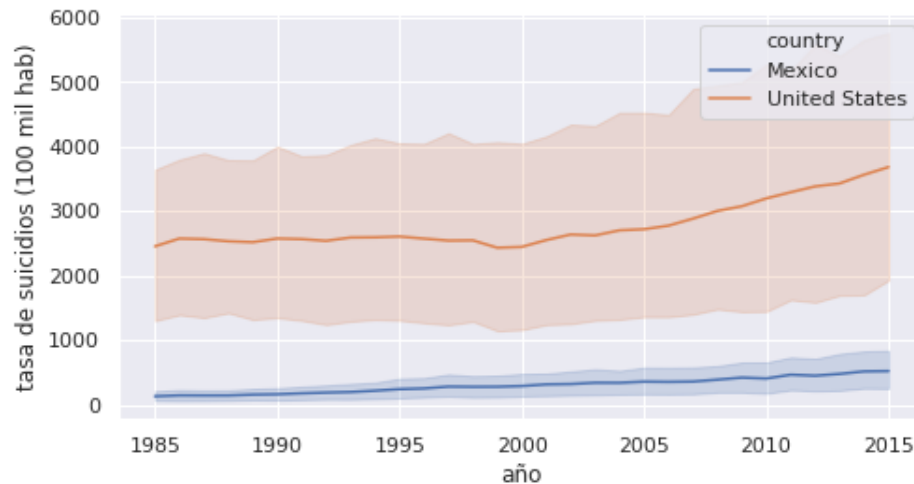


Figura 2.13: Suicidios de México y USA

Ahora consideremos la tasa de suicidios pero controlando por el sexo (ver Figuras 2.14 y 2.15). **En ambos países es claro la diferencia entre las tasa de las mujeres y los hombres, además de observarse que ambos tiene una alza creciente.** De igual, manera podemos señalar que las tasas en México son menores que las tasas de USA para ambos sexos.

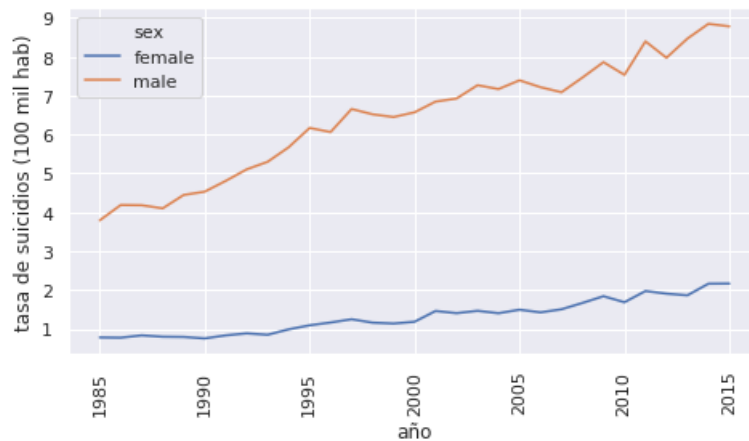


Figura 2.14: Tasa de suicidios para México

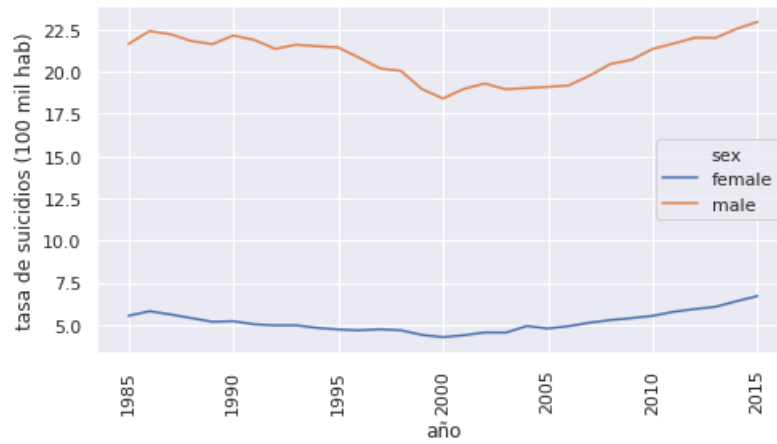


Figura 2.15: Tasa de suicidios para USA

Veamos como se relaciona el rango de edades en cada país (ver Figuras 2.16 y 2.17). En México no se observa el patrón que se veía a nivel mundial, aquí podemos concluir que **la tasa de suicidios en los jóvenes es la mas alta.**

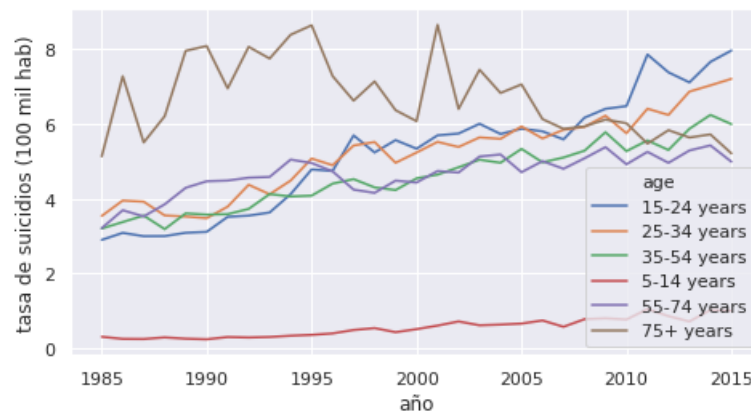


Figura 2.16: Tasa de suicidios para México rangos de edades

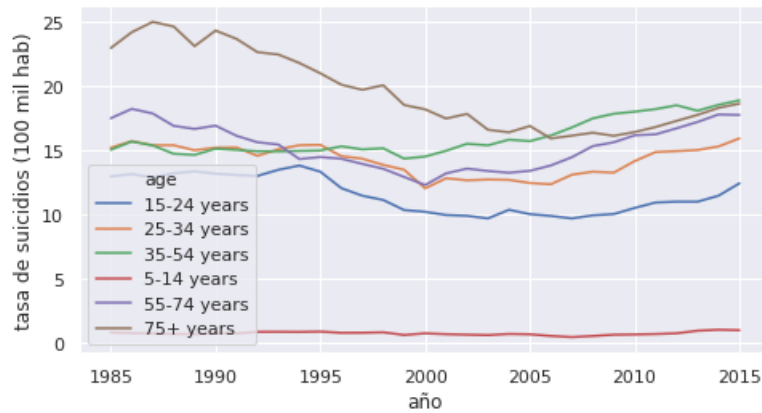


Figura 2.17: Tasa de suicidios para USA rangos de edades

En USA podemos observar un patrón parecido al mundial, ya que **los rangos de las edades más altas presentan las tasas más grandes**. Esta diferencia podría deberse a las costumbres que tenemos en México, es común que los hijos cuiden de sus padres en la vejez que a diferencia de USA utilizan con más frecuencia los asilos.

Por último, controlemos el sexo y el rango de las edades de estos países (ver Figuras 2.18 y 2.19). En México se observa la separación de las tasas por sexo, pero además para la tasa de las mujeres tiene un patrón por rango de edades, es decir, **la tasa de las mujeres va creciendo conforme al rango de edad**. Que a diferencia de la tasa de los hombres, esta presenta la mayor tasa en los jóvenes.

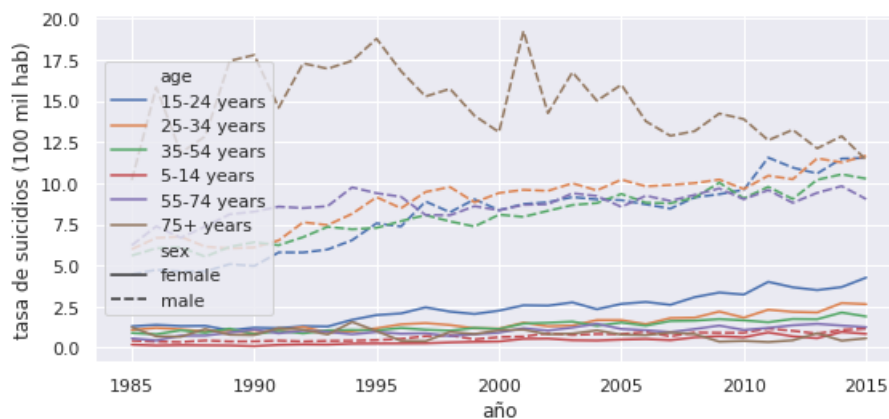


Figura 2.18: Tasa de suicidios para México rangos de edades y sexo

De igual manera que paso en México, en USA se observa la diferencia entre las tasas por sexo y rango de edad.

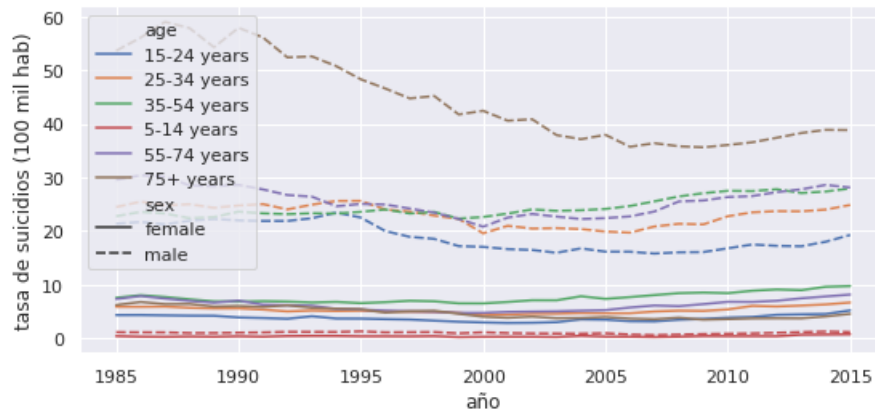


Figura 2.19: Tasa de suicidios para USA rangos de edades y sexo

En esta sección no se realizó las correlaciones sobre las variables económicas debido a que el objetivo era detectar si el número de suicidios tenía relación con el HDI o GDP, pero no estábamos a entender si relación del cambio de estos indicadores durante el tiempo y el suicidios. Es por esto que el análisis solo cubren las variables de sexo y edad.

## 2.5. E) ¿QUÉ OTRA INFORMACIÓN TE GUSTARÍA TENER DISPONIBLE PARA ANALIZAR ÉSTE FENÓMENO Y CÓMO LA UTILIZARÍA?

Claramente se observó la relevancia que tiene de las variables: sexo y la edad de las personas. Pero no se identificó algo tan claro cuando se analizó el DHI y el GDP, entonces para intentar ampliar este análisis por este enfoque creo que considerar la desigualdad social (coeficiente de gini) que tiene cada país ayudaría al análisis.

Teniendo como objetivo prevenir los suicidios, tener más información sobre el suicidio y las personas que se suicidaron podríamos segmentar la población para poder que personar estaría más propensas a realizarlo. Por ejemplo: ¿cómo se realizó el suicidio?, ¿hay intentos anteriores de suicidios?, ¿antecedentes de consumo de drogas?, ¿problemas familiares?, ¿nivel escolar de la persona?, ¿familiares cercanos con signos de depresión?, etc. Posiblemente tener las respuestas a estas preguntas nos ayudarían a entender más a fondo el problema.