

PROYECTO DE VISUALIZACIÓN 2021

25 NOVIEMBRE

UNIVERSIDAD DE VALLADOLID

Creado por: Héctor Toribio González



Informe inicial

Introducción

Para empezar con este trabajo estuve buscando en numerosas páginas de internet hasta encontrar un dataset que me interesase. Finalmente me quedé con el dataset sobre la predicción de ataques al corazón, el colesterol en mujeres y hombres.

He escogido este dataset porque me parece algo muy importante el saber cuando hay que preocuparse por el colesterol, cuando es necesario actuar y tu corazón puede estar en peligro. Además de que me interesa informarme sobre el tema ya que corre mucha desinformación e información falsa por las redes.

Un colesterol alto se considera a partir de 240 mg/dL, el deseable es menor de 200 mg/dL y tenerlo algo elevado se encuentra entre 200 y 239 mg/dL. En estos datos podemos observar varias personas que llegaron a tener el colesterol incluso por encima de 361 mg/dL. Esto es muy peligroso y por eso voy a utilizar estos datos, cuidarse es muy importante.

Para expresar estos datos voy a utilizar un diagrama de barras con animaciones y distintos botones que te muestren los datos.

Estructura de la memoria:

1. [Información sobre los datos](#)
2. [Planificación de la visualización](#)
3. [Diseño de la visualización y boceto](#)
4. [Referencias y bibliografía](#)

1. Información y análisis de los datos

Voy a utilizar un csv obtenido de una página web²:

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>

El csv es sobre predicciones de ataques al corazón, con los datos que se pueden observar en la imagen 1. Modificaré este csv ya que trae demasiada información (Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, ST_Slope, HeartDisease) para analizar solo tres tipos de datos (Age, Sex, Cholesterol) que nos interesan para la visualización. En este csv aparecen estos datos recogidos de 918 personas.

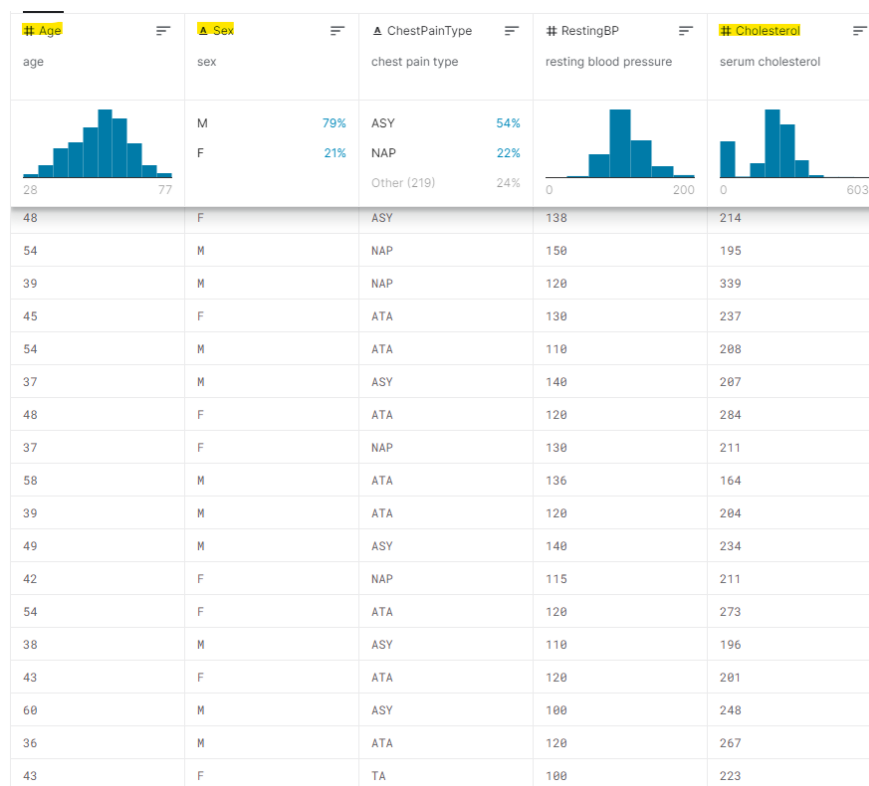


Imagen 1

Hay tres atributos:

- Edad: Oscila entre 28 y 77. Las edades más comunes en las que ocurren estos ataques al corazón son entre los 52 y los 57. Es un tipo de datos cuantitativo natural.

- Sexo: Un 79% son hombres y un 21% son mujeres, son datos categoricos sin orden.
- Colesterol: Oscila entra 0 y 603 mg/dL. La mayoría oscila entre 180 y 241 mg/dL. Son datos cuantitativos naturales.

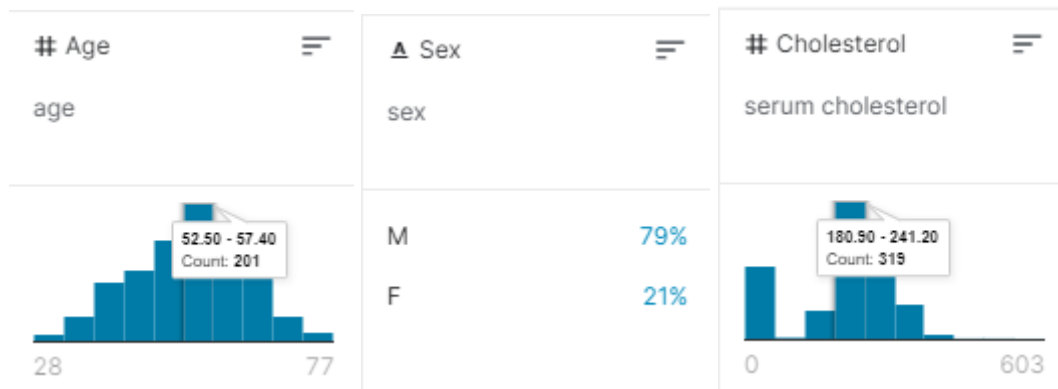


Imagen 2

Los datos van a necesitar transformaciones. Estas transformaciones consisten en hacer la media de todos los colesteroles de personas de la misma edad y el mismo sexo. Por ejemplo, habrá varios hombres de 40 años de los que se han almacenado datos, vamos a necesitar hacer la media de los colesteroles de esos hombres de 40 años para hacer la barra correspondiente en nuestra visualización.

La edad se representa en años, el sexo está representado en mujeres y hombres y el colesterol está expresado en mg/dL.

2. Planificación de la visualización

Mi **objetivo** en esta visualización es informar a las personas de la posibilidad de sufrir un ataque al corazón en función del colesterol (sin contar otras causas como la presión sanguínea o la glucemia en ayunas), de la edad y del sexo.

Podemos definir la **función** como explicativa, ya que vamos a intentar transmitir los niveles de colesterol mas peligrosos dependiendo de la edad y el sexo.

En cuanto al **tono** voy a enfocarlo más a un tono emotivo.

El **efecto** que quiero causar es encontrar relaciones entre el nivel de colesterol, la edad, el sexo y la posibilidad de sufrir un ataque al corazón, que el usuario se interese sobre el tema y se pregunte si se está cuidando como debe. Mi objetivo con esta visualización es concienciar de la importancia de cuidar tu nivel de colesterol.

En cuanto al **enfoque**, si queremos que el lector razone y reflexione usaremos animaciones y fotos. Tengo en mente animaciones (que comentaré en la parte de diseño) que apoyan este tono emotivo del que hablo. La interacción se hará con botones que te mostrarán diferentes gráficas dependiendo de los datos que quieras visualizar. El rango de edades irá desde los 28 años a los 77 y trataremos las edades como dos gráficas distintas o, en caso de querer compararlas, una misma gráfica con unas barras de un color para los hombres y otras de otro color para las mujeres.

3.1. Diseño de la visualización

Mi visualización se basa en dos diagramas de barras. Los ejes X van a ser la edad de las personas y los ejes Y van a ser el colesterol. La grafica podrá cambiarse para hombres o para mujeres apareciendo unas barras con unos datos o otras con los otros pero, en caso de necesitar comprar los datos, habrá un botón para mostrar las barras de las dos gráficas al mismo tiempo diferenciadas por colores. Se podrá cambiar de gráfica con tres botones situados en la parte derecha de la visualización e intentaré hacer un gráfico circular que muestre el porcentaje de hombres y de mujeres a los que se les tomaron los datos.

En cuanto a las animaciones. Mi idea es hacer que cuando aparezcan las barras hagan dos movimientos en los que estas barras se amplien o bajen y suban al mismo ritmo, para asemejarlo a dos latidos del corazón y que, tras esto, las barras bajen y se coloquen en sus correspondientes posiciones. Voy a hacer las barras rojas, ya que es el color de la sangre y estamos hablando de posibles problemas cardiovasculares y me gustaría que los botones de mujeres y hombres fuesen glóbulos rojos, pero esto depende de como quede lo dejaré o pondré botones normales.

3.2. Boceto

La **imagen 3** corresponde a pulsar el botón de “HOMBRES”, la **imagen 4** corresponde a pulsar el botón “MUJERES” y la **imagen 5** corresponde a pulsar el botón “HOMBRES + MUJERES”.

El colesterol va desde 0 a 600 porque son los datos correspondientes a los datos y las edades van de 28 a 54 porque no puedo hacer un boceto tan grande que abarque hasta los 77 años, pero en la visualización habrá hasta 77 años.

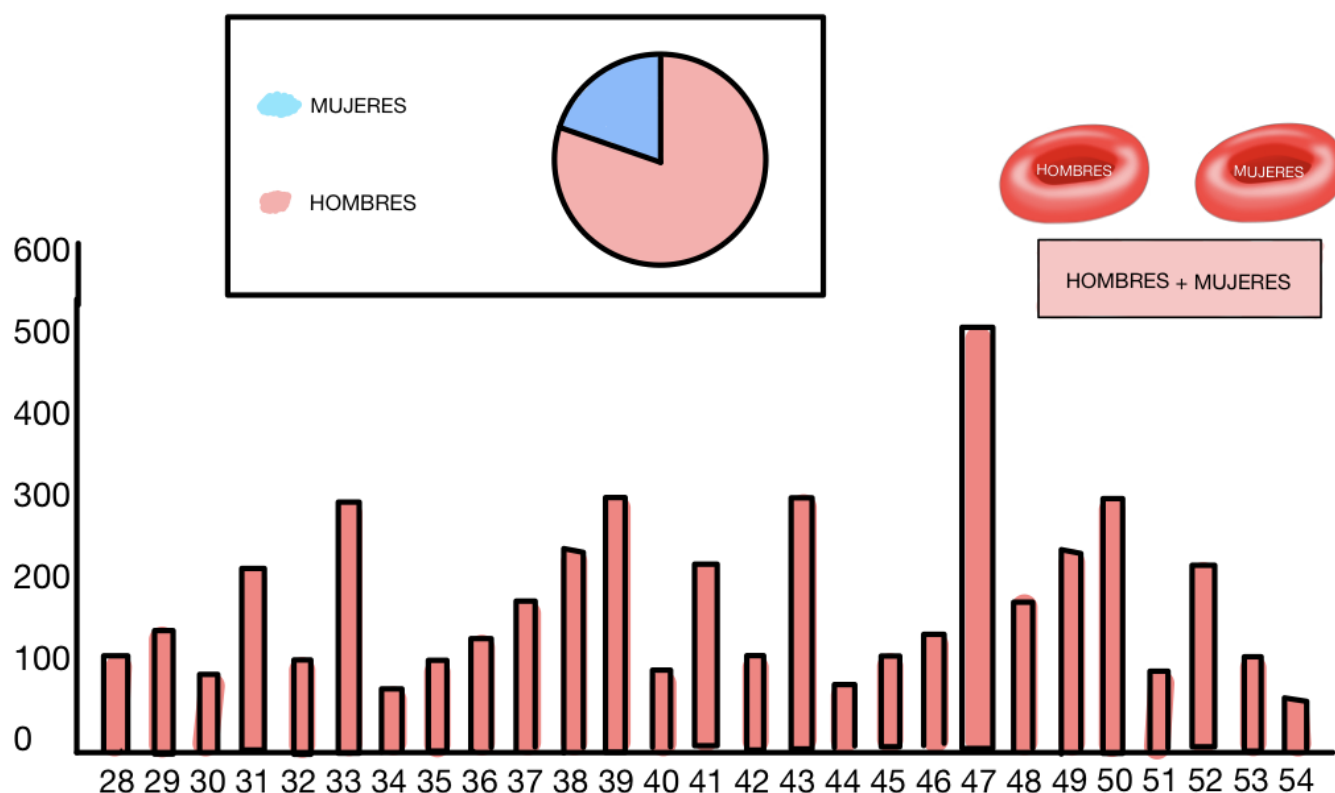


Imagen 3

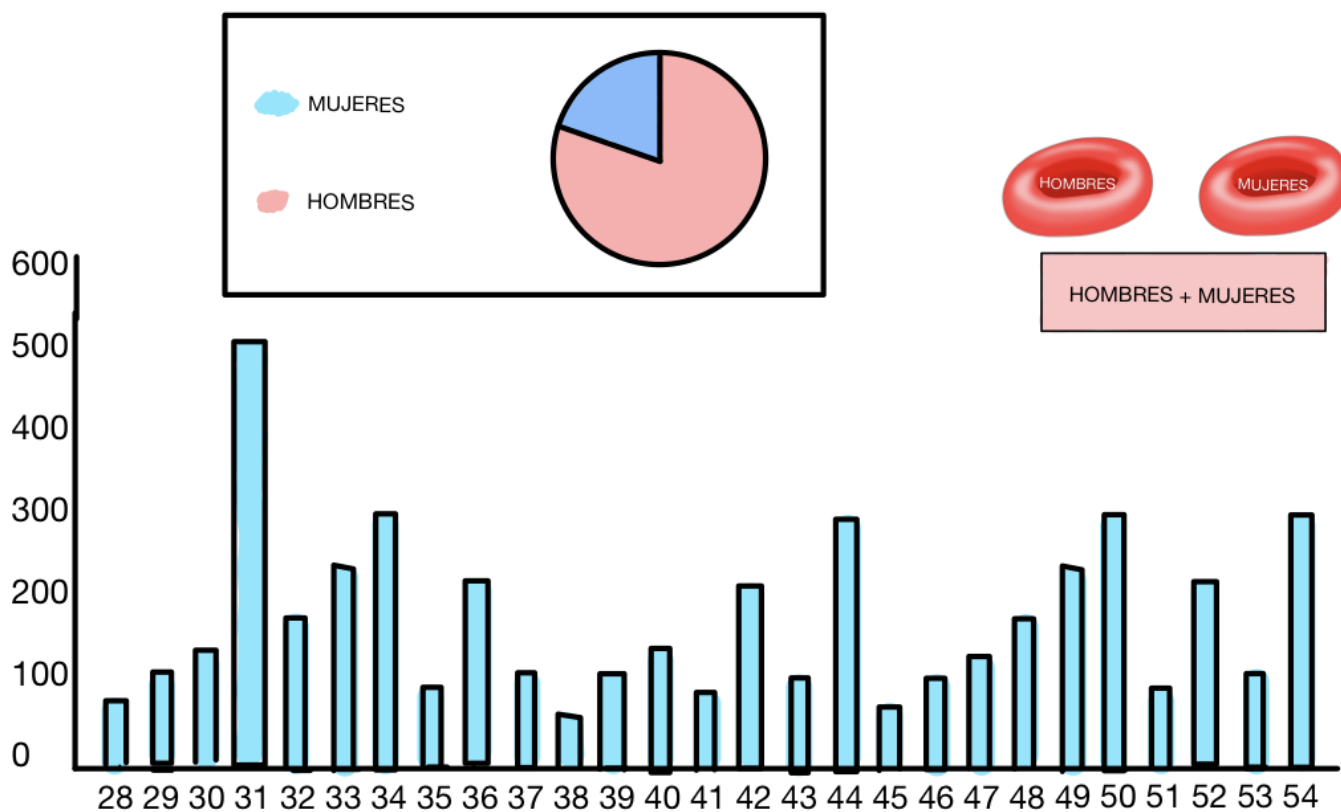


Imagen 4

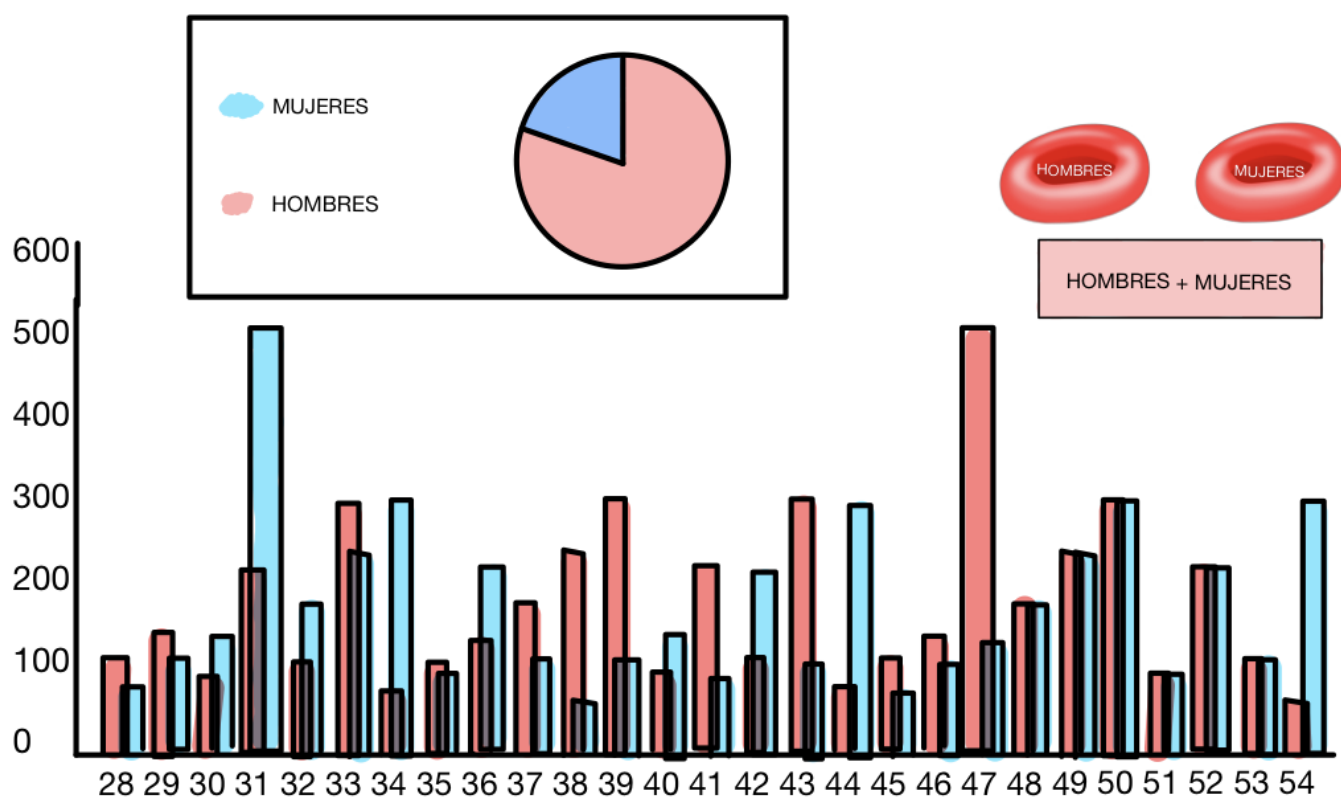


Imagen 5

4. Referencias y bibliografía

1. Para hacer esta visualización me voy a fijar en otra visualización de otro año, concretamente la visualización de la relación entre id + h, horas de trabajo y felicidad:

<https://docs.google.com/document/d/1Wf7hEVivmA41ojXoBEaP-ExoviW6Fml8cxzEX2mUugk/edit>

2. He cogido los datos del siguiente enlace:

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>