

1. ¿Cuántas variables tiene el dataset? ¿Cuántos datos?

R: El dataset tiene 1338 datos, cada uno con siete variables

2. ¿Qué variables se podrían usar para hacer un scatter plot? ¿Cómo se imaginan que va a resultar el gráfico en cada caso? Hagan el esquema en papel y lápiz.

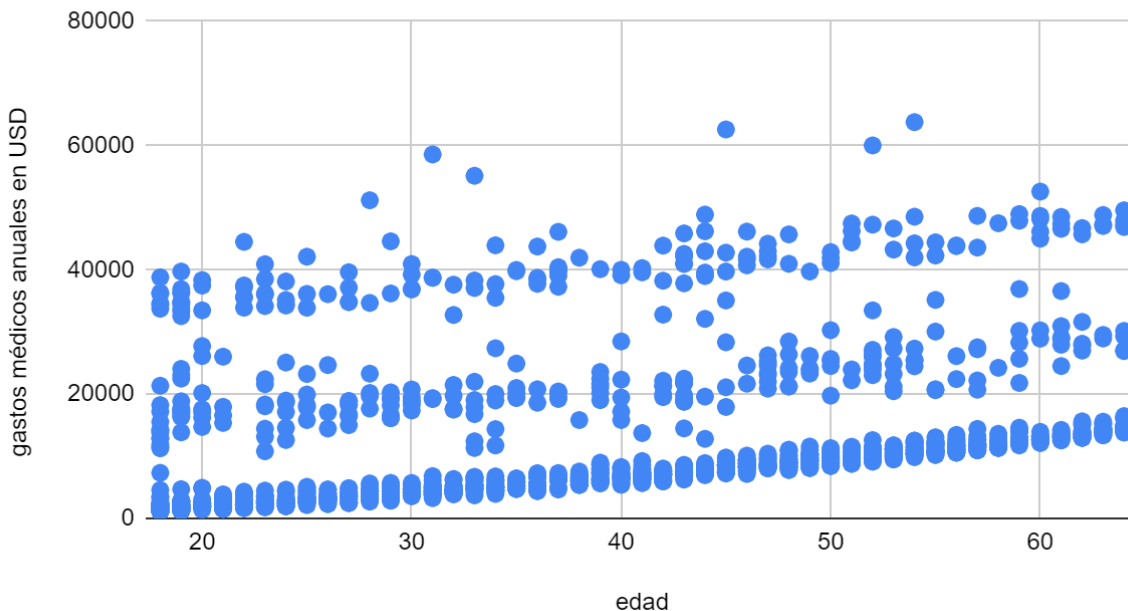
R: Podríamos usar como variable independiente la edad y como variable dependiente el costo de seguro. Tendrá una forma lineal creciente

3. Realizar alguno de los gráficos que pensaron en el punto 3. ¿Qué resultados obtuvieron? ¿Resultó como se habían imaginado? ¿Por qué?

- a. ¿Cómo mejorarían el gráfico?
- b. ¿Es necesario ajustar los límites de los ejes para poder apreciar mejor la relación?
- c. ¿Hay algún/algunos puntos que sobresale o que les llame la atención?

R: El gráfico resultó como esperaba. Podría mejorarse ajustando las unidades del eje vertical y buscando la o las variables que generan los tres grupos visibles (que es lo que más llama la atención) para colorearlos distinto.

Gastos Médicos en función de la edad



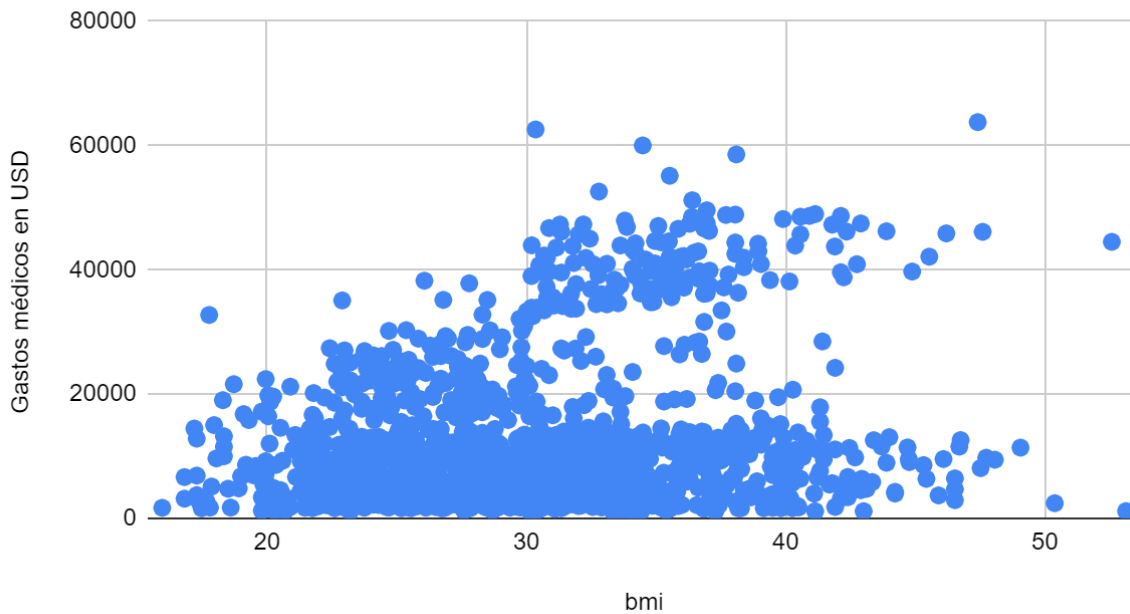
4. Volver a realizar el punto 4 para otros pares de variables.

a. De todos los gráficos que fueron haciendo ¿Cuál transmite más información?
¿Identifican algún comportamiento llamativo? ¿Cuál es más claro?

R: Para mí el más informativo es el primer scatter. Transmite un mensaje claro: a mayor edad, más se paga de seguro médico y tiene esos tres grupos llamativos que sería bueno identificar.

Este es el otro scatter que propuse:

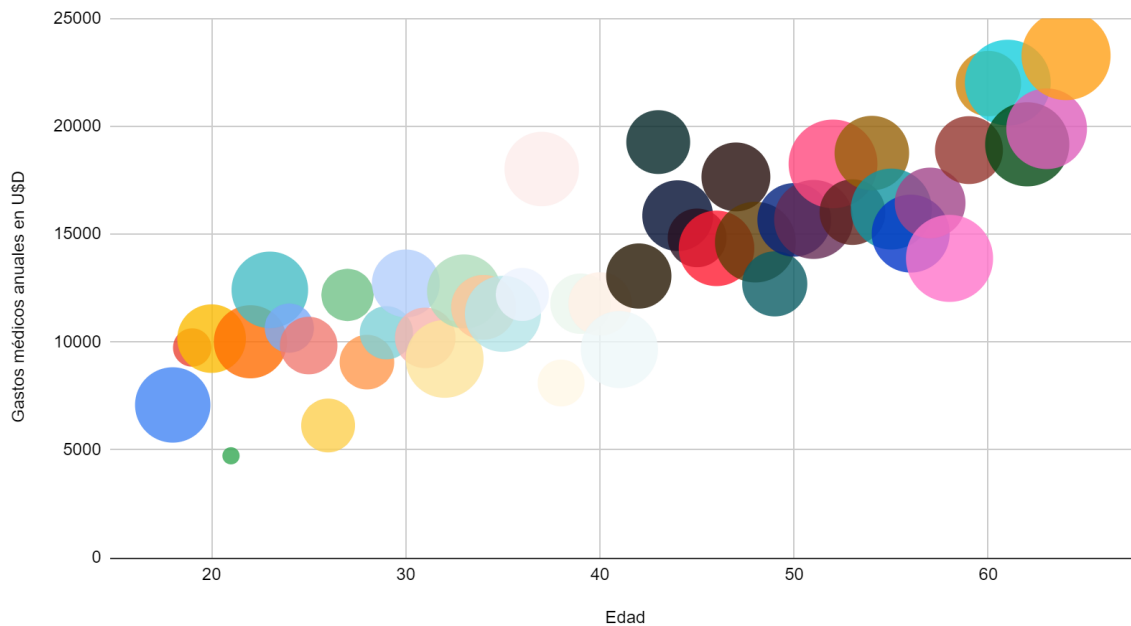
Gastos médicos según BMI



5. Elegir uno de los scatter plots realizados en el punto anterior al que le agregaremos una “tercera dimensión” utilizando un bubble chart (ver Cápsula 1). Elegir esa variable de tal manera que se pueda aprovechar la nueva información que están sumando al gráfico.

R: acá el tamaño de la burbuja es el BMI promedio para esa edad. Los gastos médicos también son promedios. Acá no supe por qué ni cómo resolver el hecho de que para cada BMI me hizo un color distinto, pero respondiendo al punto 6 yo quisiera que el color esté dado por fumador o no fumador.

Gastos médicos según Edad y BMI



6. Sumarle al gráfico anterior la información de alguna variable categórica en forma de color. Pensar cuál y por qué.

R: Como no pude cumplir con lo anterior me mandé a hacer histogramas separando en cuatro grupos según si tienen o no sobrepeso (que es lo que querría poner como variable continua que me de el tamaño de la burbuja en vez de dicotómica) y si fuman o no (la categórica que quisiera que me de el color de las burbujas). No logra mostrar completamente lo que quisiera porque no se ve la relación con la edad pero sí se deja ver que claramente los que más gastan son los fumadores con sobrepeso, luego los fumadores sin sobrepeso y por último los no fumadores

