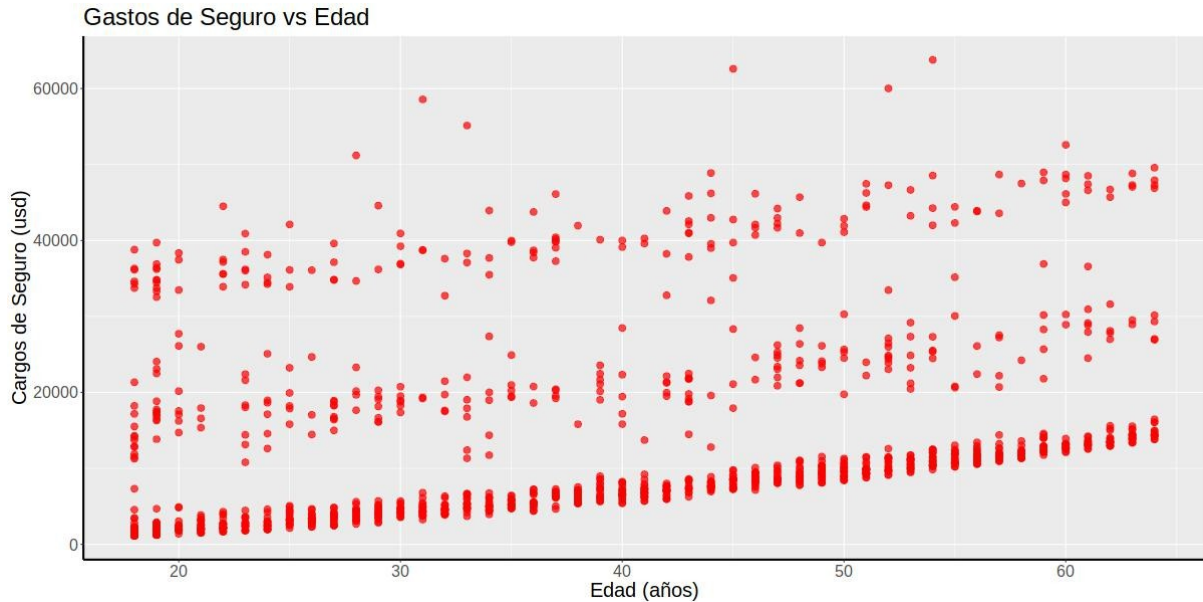
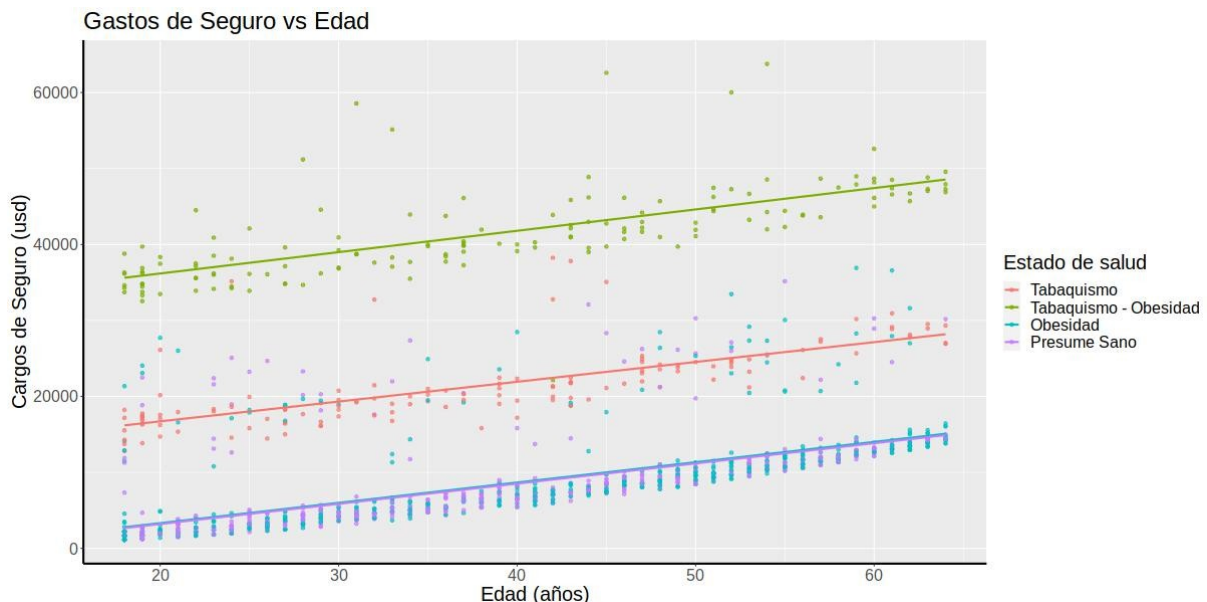


ANÁLISIS DEL DATASET INSURANCE

En esta oportunidad se analiza el dataset de cargos de seguro médico en EEUU, dominio público [1]. El dataset cuenta con un total de 1338 datos y 7 variables, de las cuales 4 son categóricas: sex, smoker, children y region. La variable charges es la dependiente y el resto son las independientes, por lo tanto, se podrían graficar todas las variables en función de charges. Por ejemplo, se podría graficar como varían los costos del seguro en función de la edad. Probablemente, exista una dependencia en el costo del seguro para personas de riesgo, como personas fumadoras o que se encuentren fuera de los parámetros estándar de BMI, por lo cual puede ser interesante incorporarlas al gráfico en distintos colores o formas de los puntos.



En la gráfica superior se observa que el costo del seguro aumenta con la edad. Incluso se logran apreciar tres agrupaciones de puntos distintos, asociadas a otras variables. Debido a esto se decide incorporar en el siguiente gráfico otras variables en otra dimensión, color.



En esta figura se puede apreciar cómo se distribuyen principalmente las personas en grupos de riesgo como tabaquismo y BMI altos hacia seguros de mayor costo. Los fumadores se aprecian principalmente en el grupo del medio y en el superior los fumadores con valores de BMI elevados, que sufren obesidad, así como en el inferior estos últimos junto a los presuntos sanos.

REFERENCIAS

1. US Health Insurance Dataset, <https://www.kaggle.com/datasets/teertha/ushealthinsurancedataset>