Context

The insurance.csv dataset contains 1338 observations (rows) and 7 features (columns). The dataset contains 4 numerical features (age, bmi, children and expenses) and 3 nominal features (sex, smoker and region) that were converted into factors with numerical value designated for each level.

Acknowledgements

Insurance.csv file is obtained from the Machine Learning course website (Spring 2017) from Professor Eric Suess at <http://www.sci.csueastbay.edu/~esuess/stat6620/#week-6>.

Inspiration

The purposes of this exercise to look into different features to observe their relationship, and plot a multiple linear regression based on several features of individual such as age, physical/family condition and location against their existing medical expense to be used for predicting future medical expenses of individuals that help medical insurance to make decision on charging the premium.