COVID MORTALITY

1. I split by the regions because I wanted to use the built in R library/functions that would automatically do the splitting for me.
2. By creating my own categorical variable, I believe it did make the model better, because the variable had a discrete amount of values (only 4 regions) so it made splitting easier. Creating your own independent variable when building a multiple linear regression model does not necessarily make the model better. In fact, creating an independent variable that is not relevant or meaningful to the response variable can even harm the model's predictive accuracy. The goal of building a multiple linear regression model is to find the combination of independent variables that best predicts the response variable. The independent variables should be chosen based on their relevance to the response variable and their ability to improve the predictive power of the model. Creating your own independent variable may be useful if it captures some relevant information that was not present in the original set of independent variables. However, it is important to ensure that the variable is statistically significant and improves the model's overall fit before including it in the model.

Lending Club

1. The summary of this dataset describes the statistical properties of a dataset containing information on loan defaults and various borrower characteristics. The dataset includes 88,451 observations with 12 variables. The first variable, "loan\_default", indicates whether the borrower defaulted on their loan, and has a mean of 0.13, suggesting that about 13% of borrowers in the dataset defaulted on their loans. The "loan\_amnt" variable represents the loan amount, with an average loan size of $12,435.17 and a standard deviation of $7,186.94. The "adjusted\_annual\_inc" variable represents the borrower's annual income, with an average of $57,013.98 and a very high standard deviation of $55,124.23, indicating a large variation in income levels in the dataset. The "bc\_util" variable represents the borrower's credit utilization ratio, with an average of 66.71%, indicating that borrowers are using roughly 67% of their available credit. The "num\_accts\_ever\_120\_pd" variable represents the number of accounts that have ever been 120 or more days past due, with an average of 0.32. The "pub\_rec\_bankruptcies" variable represents the number of public record bankruptcies the borrower has, with an average of 0.09. Overall, this dataset provides valuable information for lenders to assess the creditworthiness of potential borrowers and make informed decisions about loan approvals and interest rates.

Chart, histogram

Description automatically generated

Graphical user interface, text

Description automatically generated

Text

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence

3.

Chart, line chart

Description automatically generated