Part A

1.

* I split the states into blue and red based off political party.
* I split the states into 4 regions: North, South, Northwest, and West.
* I split the states into half high income and half low income using the median household income of each state.
* I split the states based off median household income again, but more segmented. 1 to 10, 11 to 20, 21 to 30, 31 to 40, 41 to 52.
* I split the states based off population density. 1 to 10, 11 to 20, 21 to 30, 31 to 40, 41 to 52.

I chose all these ways of splitting up the states to see if I can find which independent variables have the largest effect on the death total by state. I also chose all these ways to see if the data matched my intuition and to see if there were any unexpected trends.

2.

* Population had by far the largest effect on the death rate. This makes sense as the more people there are, the more people that can get the virus and die.
* Area had little effect on the death rate. This makes sense as there are plenty of states in the northeast with small areas, but a high, condensed population. This is while a farm state like Wyoming is very large but does not have many people to get or spread the virus.
* Being a red state had a slightly higher death rate, but negligible. This makes sense as republicans are on average less likely to take as extreme of precautions as democrats when it comes to the virus.
* The Northeast region had the largest effect on the death rate of all the regions. This makes sense because it contains a lot of states with high population density, making the spread easier compared to the other regions. These states also make up a good portion of the U.S. population.
* The bottom half of the states in median household income had slightly higher death rate, but again negligible. This makes sense because more money means more resources to fight the virus.
* The top 10 states in median household income performed noticeably better than the rest, with fewer deaths. After that, the rest of the states were about even.
* The top 10 states in population density performed well worse than the rest, with more deaths. This makes sense as the more compact people are, the easier the spread. The remaining states were all about even.

I added several independent variables, but their effect on the death rate was negligible compared to population size. The addition of these independent variables demonstrated some trends that matched my intuition, but had little effect on the model in the big picture. Population size had an F value of 88, the next closest was region at 5.