

## Project Deliverable 2

More data should have been collected to perform a more thorough analysis of the data and attempt to answer one additional question relevant to your project proposal which you will submit as a pull request.

## Checklist

1. Collect and pre-process a secondary batch of data
  - Collected insurance data of each county in the US and combined the data with the covid case and death data for each county.
2. Refine the preliminary analysis of the data performed in PD1
  - Removed duplicate or unassigned counties i.e. Utah's four divisional sections
3. Answer another key question
  - What is the distribution of health insurance among US states? US counties?
    - The average percentages of the insured population among the US states are all at least 80%. The states with the most insured population are the states in the East Coast and states with the highest population of people. For example, Massachusetts has the highest percentage of insured population (96.7%) and high ranking states include New York, Connecticut, and California.
  - What is the relationship between confirmed COVID rates and COVID death rates?
    - They are directly proportional, as cases increase in a county/state deaths also increase.
  - What is the relationship between health insurance and confirmed COVID rates?  
Between health insurance and COVID death rates?
    - No distinct relation, may need to break down data or get a linear/logistic regression of the points.
4. Refine project scope and list of limitations with data and potential risks of achieving project goal
  - Not seeing a clear correlation between health insurance and covid rates
  - Split out data into different subsets of states based on population, climate, wealth, education, lockdown period, etc.
5. Submit a PR with the above report and modifications to original proposal