In what follows, use CDC data. Each discussion point should be supported with the respective code. Submit the code along with your answers.

1. investigate how many missing data points each column has.
2. Following our class discussion, comment on whether the presence of missing data could impede the analysis of measured health characteristics. For example, consider how missing data may affect identifying patterns and significant differences at the level of locations/cities and when aggregated by states.
3. Discuss how the spatial locations are represented in the dataset (e.g., list the names of the columns related to location measures and their relationships to each other).
4. Following our class discussion, check and explain any inconsistencies and the one-to-one correspondence between ('StateAbbr' and 'StateDesc') and between/among ('LocationName', 'LocationID', and 'Geolocation').
5. Discuss how health outcomes are represented in the data and how many unique health outcomes are being measured.
6. Is there a time variable in the data? If so, discuss whether the repeated measures over time and the health measures of interest differ across each time period.
7. Create a DataFrame where the rows represent unique locations and the columns represent health measures.
8. List the top 3 highest correlated health measures.
9. Create a DataFrame where the rows represent states and the columns represent health measures, with the values inside the table being the averages of the measures by state.