Project Milestone #3: Nov 11	
Criteria	
Subset rows or columns as needed	2
Create new variables needed for analysis (minimum 2)	3
Clean variables needed for analysis (minimum 2)	3
Restructure datasets, as needed	2
Data dictionary based on clean dataset (minimum 4 data elements)	4
One or more tables with descriptive statistics for 4 data elements	4
HTML document that is professionally prepared for presentation	2
Total Points:	**20**

The sim_novelid_CA.csv file and sim_novelid_LACounty.csv file together represent simulated morbidity for the entire state of California. While it's certainly possible for data from different sources to adhere to a format standard, one of the challenges of working with secondary data is that it often does not work out that way. Therefore to prepare each dataset (milestone #3) so that they can be combined into one, whole state, dataset you will need to do the following:

Morbidity datasets (from CA and LA County)

- Recode column names, values, or formats that are in discordant (dates, etc)
- Combine morbidity datasets into a single dataset
- Select demographic and geographic strata (s) of interest
- Aggregate the data into a new dataframe to include only one row per strata of interest

Population dataset

- Recode values to be consistent with morbidity datasets so they can be joined
- Select demographic or geogrphic strata(s) of interest
- Create a rate metric and summarize to include only one row per strata of interest