As mentioned in the class, the conversion from ICD10 to ICD9 is the way to go.

Looking at file “2018\_I10gem.txt” we can see that what ICD10 code corresponds to what ICD9 code.

Also, from the California ICD9 data, we can find out that what ICD9 codes and with what frequency had been used in the first three quarters of 2015.

My strategy to map ICD10 codes to ICD9 was to find out the repeated ICD10 codes from “2018\_I10gem.txt” and then find the corresponding ICD9 codes and the frequency they had been used and then pick the ICD9 code with highest frequency and assign it to the ICD10 code.

My Python code, ontology2.py, reads “2018\_I10gem.txt” and “ICD9.csv” as inputs and print out a matrix containing of two columns, ICD10 and ICD9, where each row shows which ICD10 code should map to which ICD9 code.

Using the above strategy, we are able to map each ICD10 code to either a unique ICD9 code or if there was many corresponding IC9 codes, pick the one that had been used more than the others.

Before processing the ICD10 codes in “2018\_I10gem.txt”, there are 70940 repeated codes. After processing, we will have 7474 unique codes out of 70940 that is mapped to the most frequent corresponding ICD9 code.