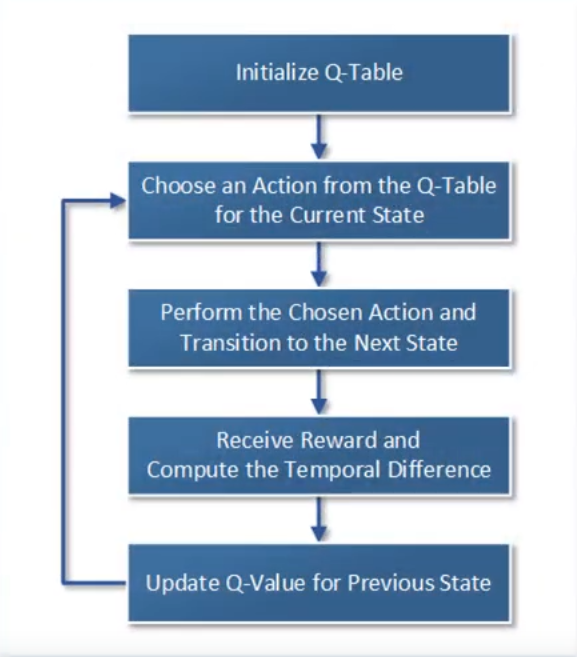
Q learning:

* Q-value indicates the quality of a particular action a in given state s
  + Current estimate of the sum of future rewards
    - Will increase as agent gets closer to highest reward.
* Q-values stored in Q-table
  + One row for each possible state, and one column for reach action
  + Represents the policy for acting in the current environment.
* Temporal differences provide method of calculating how much the Q-value for the action taken in the previous state should be changed, based on what AI agent has learned.
* Bellman Backup

Q Learning Process:



Notes:

* In order to prevent the highest Q-value from being selected every time, implement an “Epsilon greedy algorithm” which includes an element of chance in which action it selects.