In question 2 we implemented Q learning on frozen lake.  
The value of the starting state is measured at each episode using first-visit Monte Carlo.  
The action at each step is chosen using an epsilon greedy policy, with epsilon = 0.3, which chooses a random action with probability 0.3, else a greedy action. The Q-table is updated after each step using the formula .  
The graph plotted shows the value of the starting state which is sampled every 10,000 iterations.   
A graph with a line

Description automatically generated