Yoonseok Choi // 677261654 // choi156
MP1 Part 2 : Experimentation

1.
(a) It produced negative reward at first and zero for the rest.

(b) It produced small sized negative rewards for the first few times and produced positive rewards for the rest in average about 100~150.
ex)first ten and last ten

-19.5 -13.0 -15.5 -12.5 -11.5 -6.5 -13.0 -4.5 -4.0 -9.0 ...->	208.5 93.5 180.0 72.0 104.0 173.5 164.5 70.5 154.5 123.5

(c) It produced negative rewards for most of the time.
ex) first ten and last ten

-36.5 -36.5 -30.0 -36.5 -22.0 -25.0 -26.5 -16.0 -31.0 -22.5 ...->	-49.5 -58.0 -13.0 -58.0 -23.5 -39.5 -42.5 -25.5 -36.0 -28.5

2
(a) It produced negative rewards for all the time however for average about -10
ex) first ten and last ten

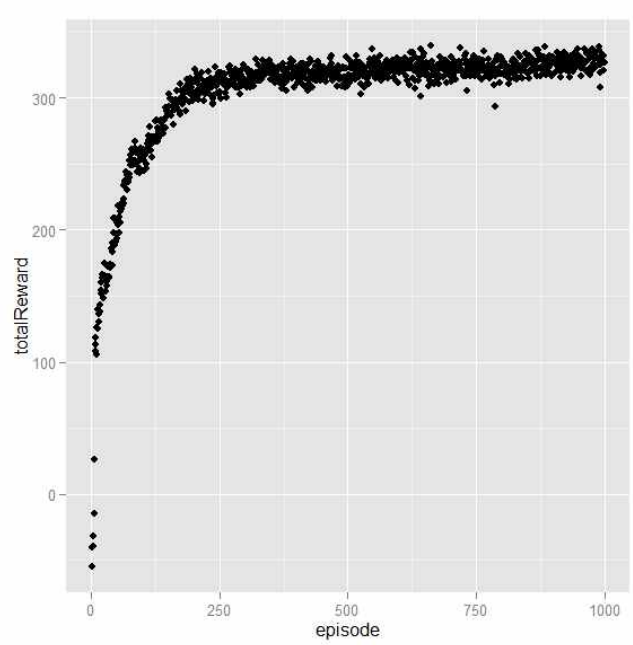
-40.5 -24.5 -17.0 -12.0 -19.0 -11.5 -9.5 -14.5 -12.5 -13.0 ...->	-16.0 -9.0 -16.0 -10.5 -12.5 -10.0 -13.0 -10.0 -9.0 -13.0

(b) It produced negative rewards for the first few and positive for the rest of simulation. The rewards increased in the beginning of the simulation until later 200's. Average about 200.

-71.0 -52.0 -50.0 -35.5 -23.5 11.0 61.0 83.0 82.0 102.5	275.5 268.5 278.5 283.5 275.5 279.5 279.5 297.5 295.0 299.0

(c) I have found best learning rate to be 0.1 and epsilon to be 0.01. This produces rewards about average 300.

3. The total reward increases each episode. Data is yet scattered a little.



For the averaged rewards, the graph looks similar to the first one, however, it looks neat since data are less scattered due to averaged process.