Benjamin Ng
kbng2
CS 440
MP1 Part 2

1.a

The agent gets stuck. It refuses to take the slippery route because it is greedy.

1.b

The agent now successfully finds the packages because it randomly decided to go on the slippery tiles despite its better judgement.

1.c

The route taken gets more and more indirect, which is the result of poor decisions being selected over good decisions too often.

2.a

The agent gets stuck because it is unaware of the location of the thief, and thus leaned to associate the slippery tiles as well as the entire path the thief takes with a penalty.

2.b

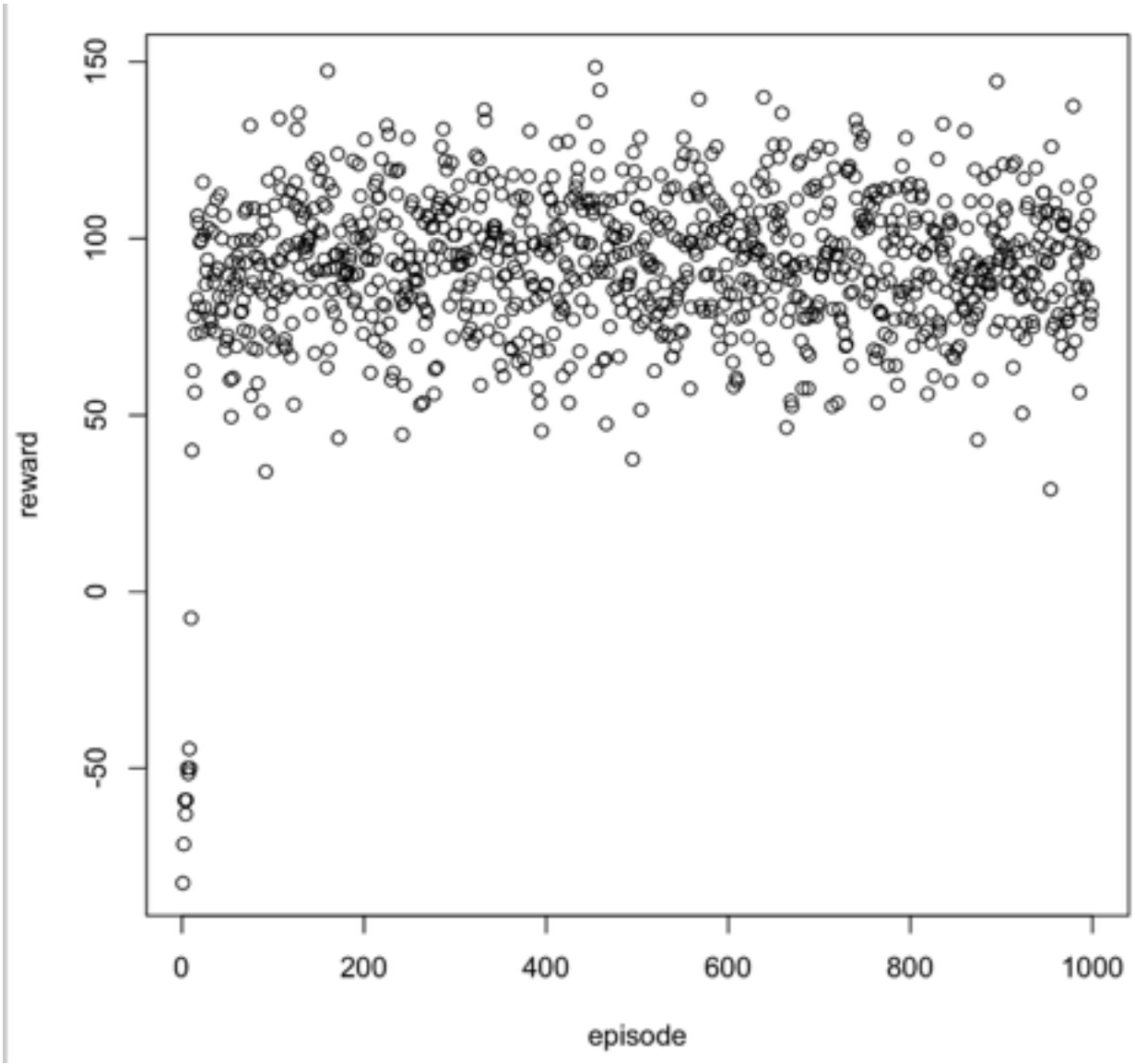
The agent retrieves the packages because it has learned to avoid the thief by using its position.

2.c

I found learning rate = 0.1 and epsilon = 0.2 to give the best results. If rate was too high, the agent ran into the thief too frequently. If epsilon was too high, the agent started behaving erratically.

3.

I did the following by importing episodes.txt into R and averaging the results of ten runs. The reward plateaus quickly at around 100 episodes, which makes sense because there is a limit to how rewarding an episode can be. As the learning algorithm nears this limit, it cannot do much better. The plot is attached.


reward
episode
150
100
50
0
-50
0
200
400
600
800
1000