Max Tjen
CSDS 313
Homework 4

1) Some plausible reasons for the observed data can be:
    a) SAT scores are scaled/relative to other SAT takers, so for example, getting 2 questions wrong in one year won't necessarily get you the same score as if you get 2 questions wrong the next. Therefore, the latter group of students may be performing the same as before, but other SAT takers are doing better relative to before
    b) The distributions of grades are represented as percentages rather than proportions. With that, the total number of students in the high school changes percentages even if the raw numbers are the same. For example, if in the first year has 5/100 students have an A+ GPA, the percentage would be 5%, but if the next year has 5/90 students, then the percentage would be 5.5%, so the percentage increased even though the raw amount of students stayed the same.
    c) The percentages of students' GPA breakdown is based on the entire high school, while SAT test takers are generally junior year students with some sophomore and senior year students. Because of this, the majority of students' GPAs included in the percentages are not apart of the group of students who took the SAT's

2) 80%
    a) Because everyone in Group II (4,500 or 5,000 per month) earns more than 80% of Group I (4,000 per month) while they also all earn less than 20% of Group II (9,500 per month)

3) The bar chart isn't an accurate representation of the student's algorithm being significantly better because of the y-axis scale. In the chart, it looks like there is a large gap/difference between the performance of the algorithms while in reality, the raw difference is around 5%, which isn't that large. To improve this, it would make more sense to increase the y-axis scale so that the bars' heights look closer together so that the chart is more representative of the performance difference.