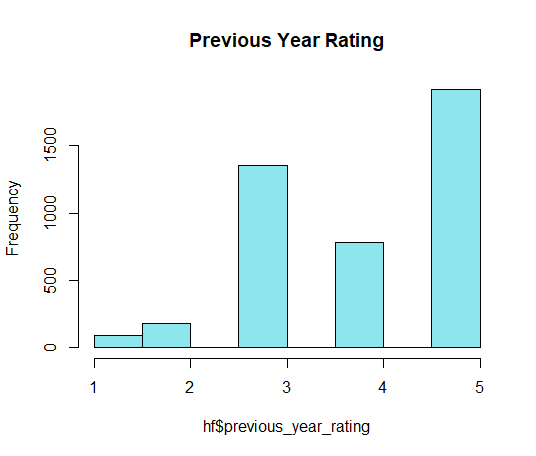
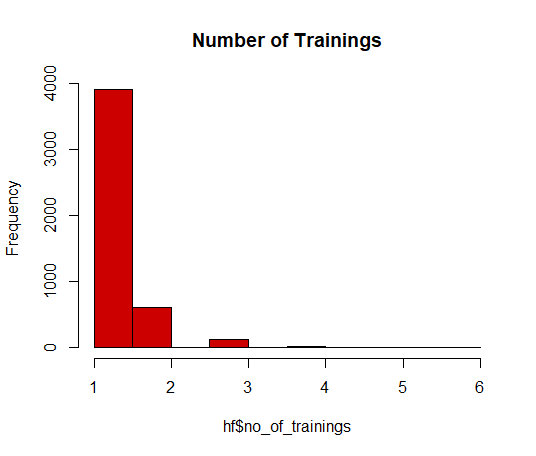
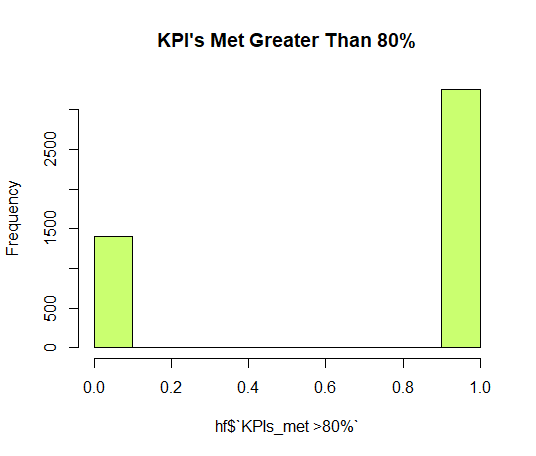
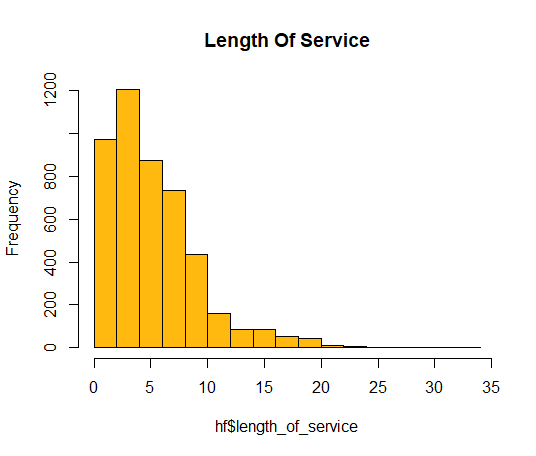
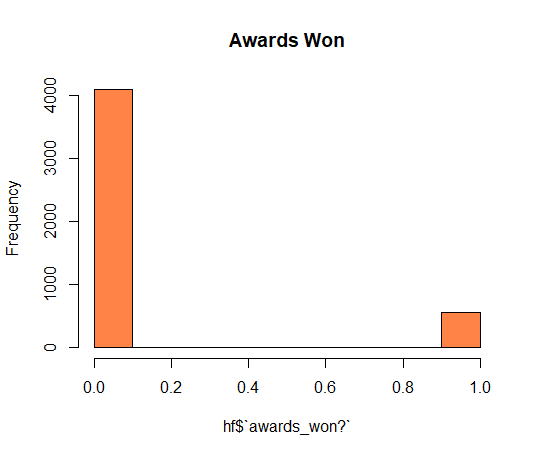
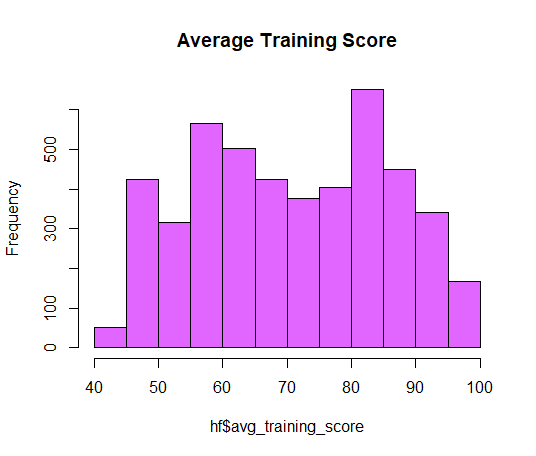
In order to first see how each variable might be impacting the candidate selection for promotion, the first thing I wanted to do was compare the frequencies of all the variables that are most common. Using R, I was able to build these graphs. Only employees in the data set that were selected for promotion were used in order to see what variables likely had the greatest impact on who should get promoted.



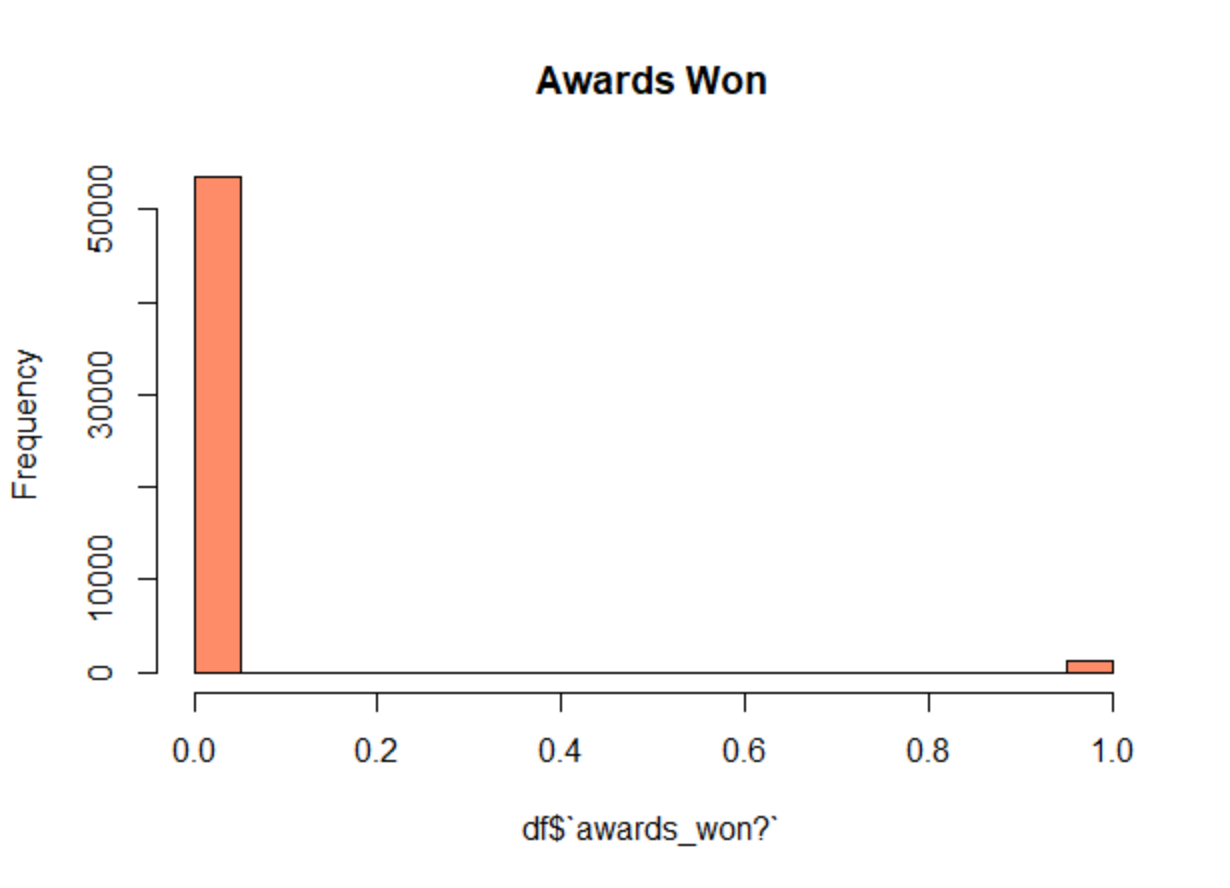
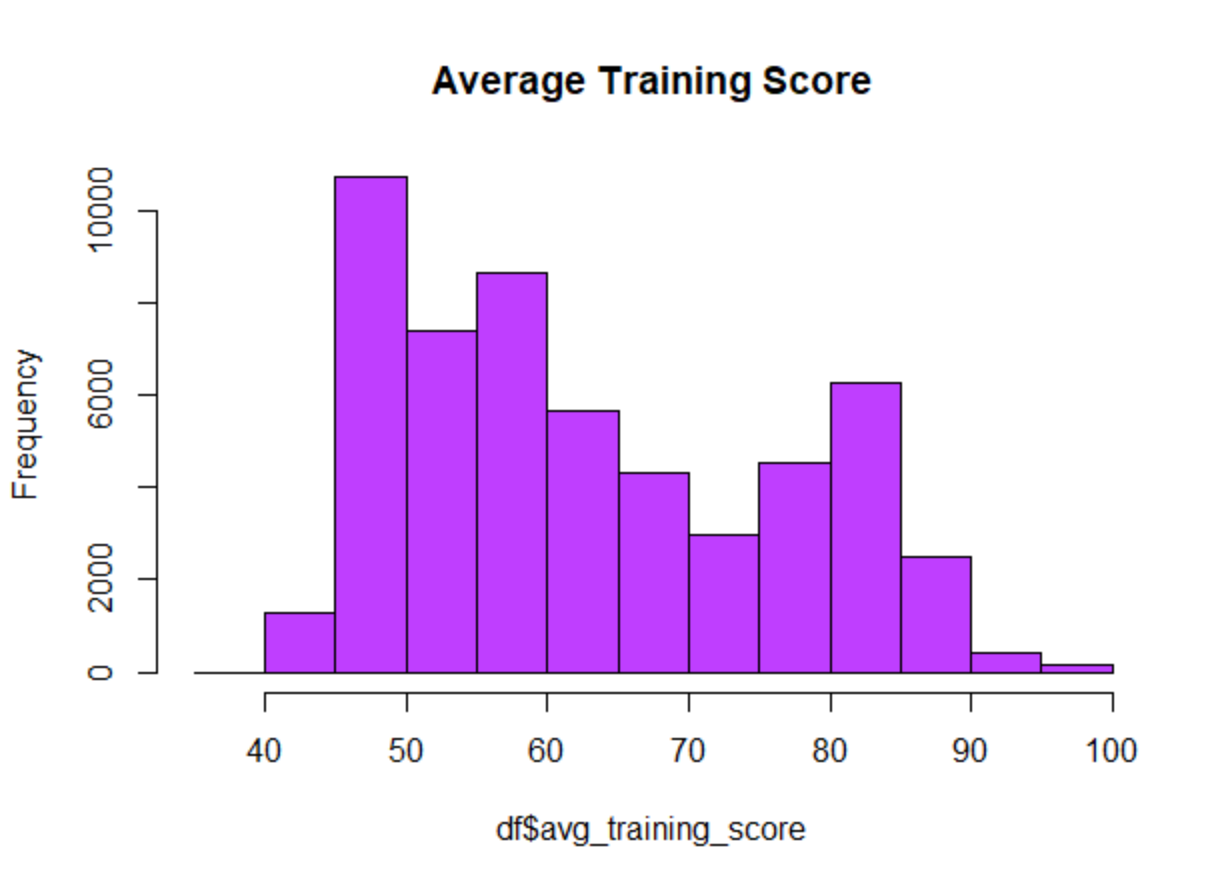
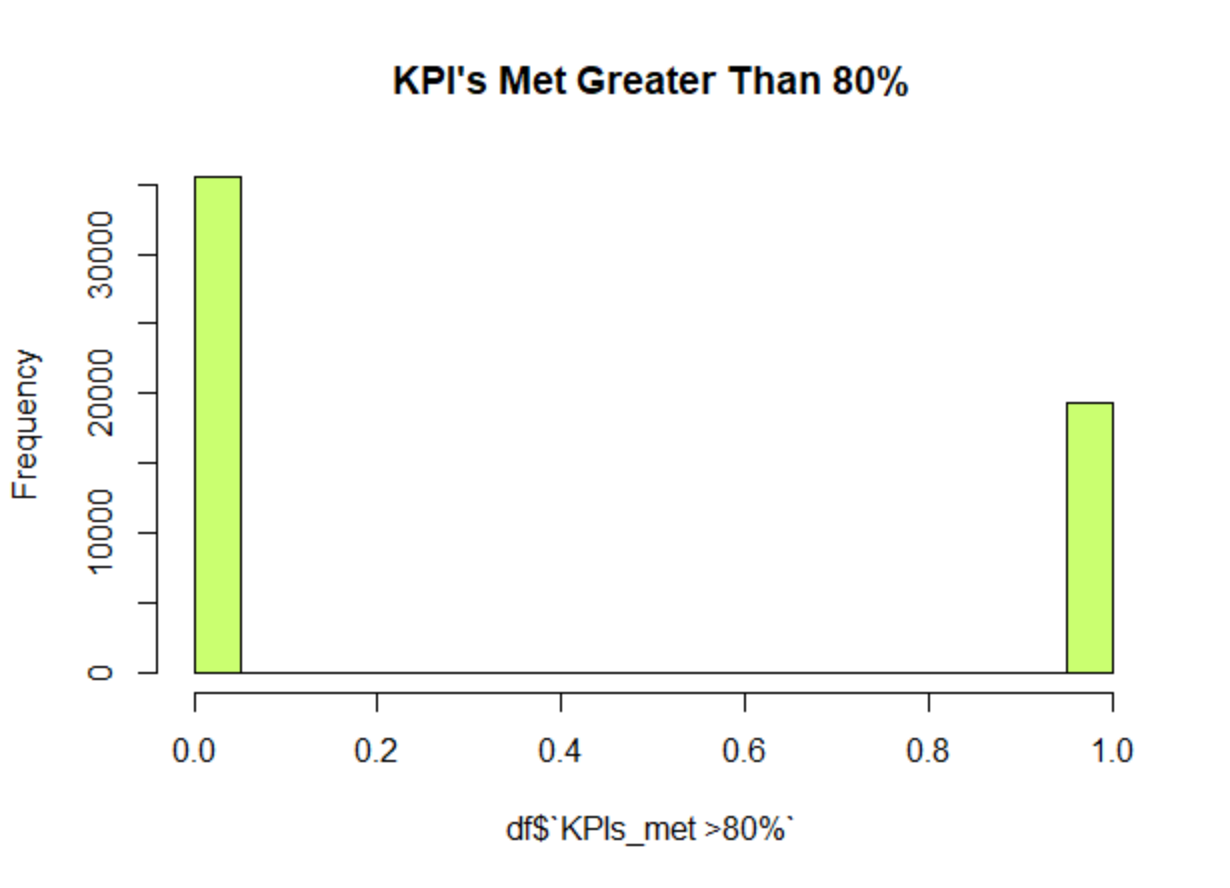
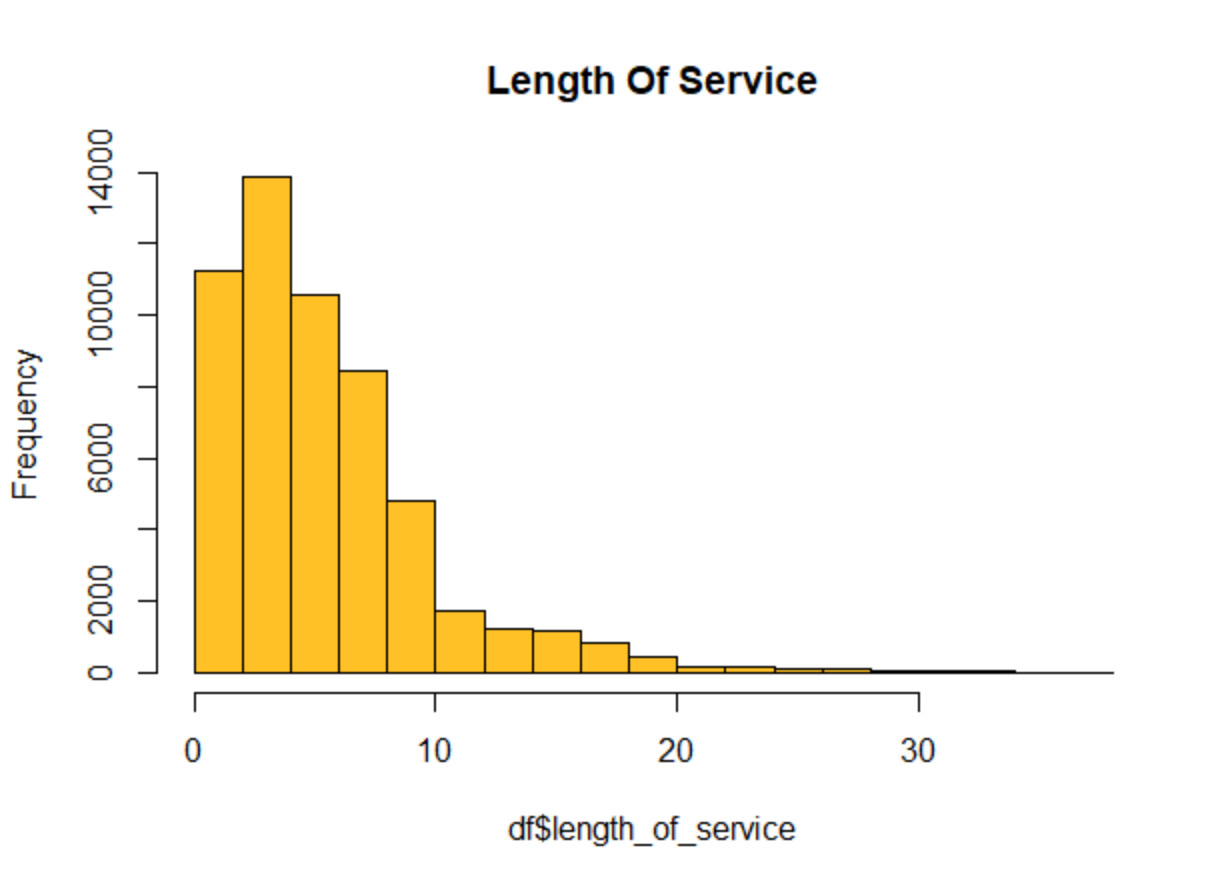
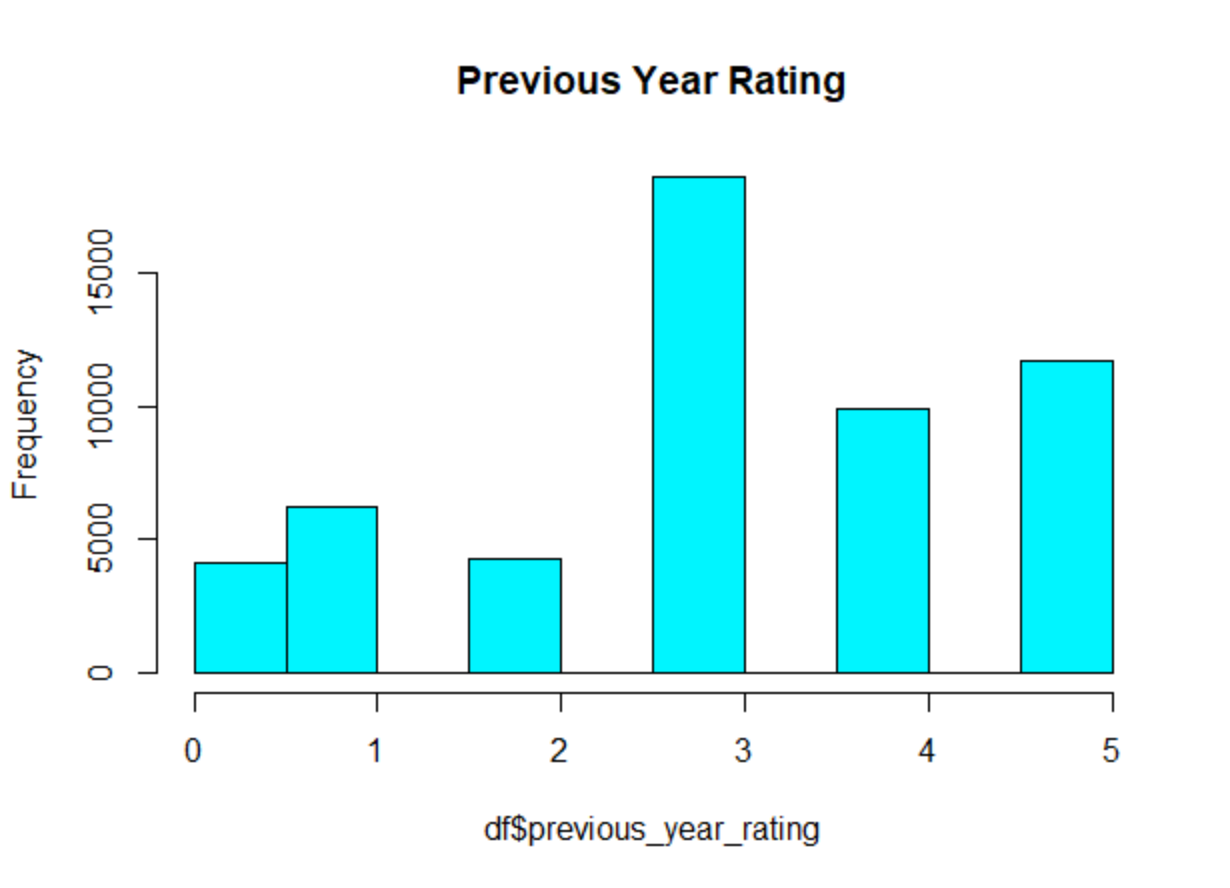
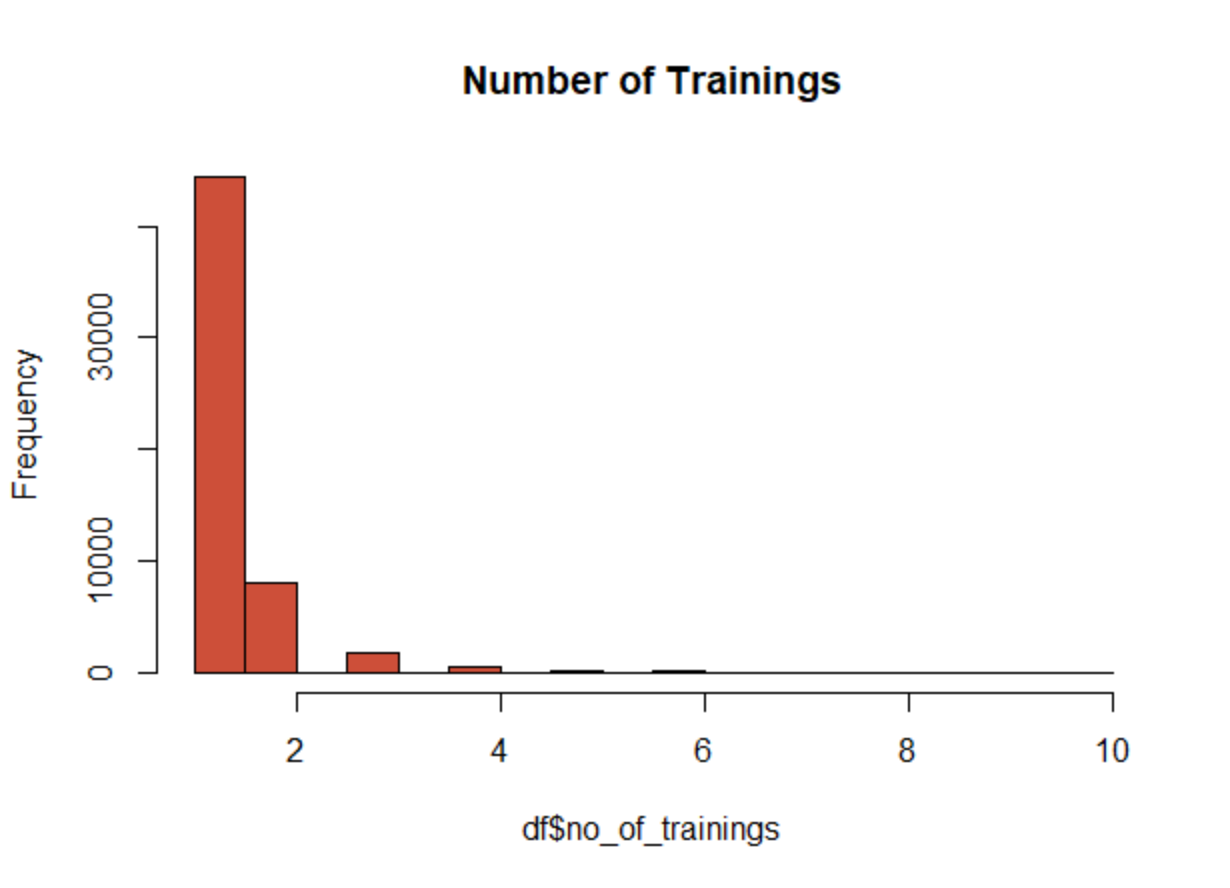




From the graphical results we can see some trends from the promotion group:

1. The majority of employees that were selected for promotion were rated 3 and up last year.
2. Approximately 2/3 of the employees had KPI’s > 80.
3. The average training score is a little bit more evenly distributed.

It is hard to tell though with just the graphs how each variable should be rated. If we take a look at the overall numbers and spot overall trends and compare it might tell us more which variable has the stronger impact.



Comparing the two sets of graphs a couple of items are apparent:

1. The average training score of the whole group is skewed to the left suggesting that the majority of the employees scored worse then the ones selected for promotion.
2. The employees selected for promotion, the KPI’s met seems like a strong indicator.
3. Previous year rating, also appears to be a moderate indicator.

The histograms give a good understanding of the data set overall, and how the variables affect selection for promotion. This isn’t enough to go on to determine what weights each variable should have for the best predictions so the next step is to check statistically the importance of each variable.