

## Write up ASGN7

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -f 20 < newspeak.txt
Average BST size: 1.113072
Average BST height: 1.110016
Average branches traversed: 0.244328
```

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -f 40 < newspeak.txt
Average BST size: 1.113072
Average BST height: 1.110016
Average branches traversed: 0.244328
```

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -f 60 < newspeak.txt
Average BST size: 1.113072
Average BST height: 1.110016
Average branches traversed: 0.244328
```

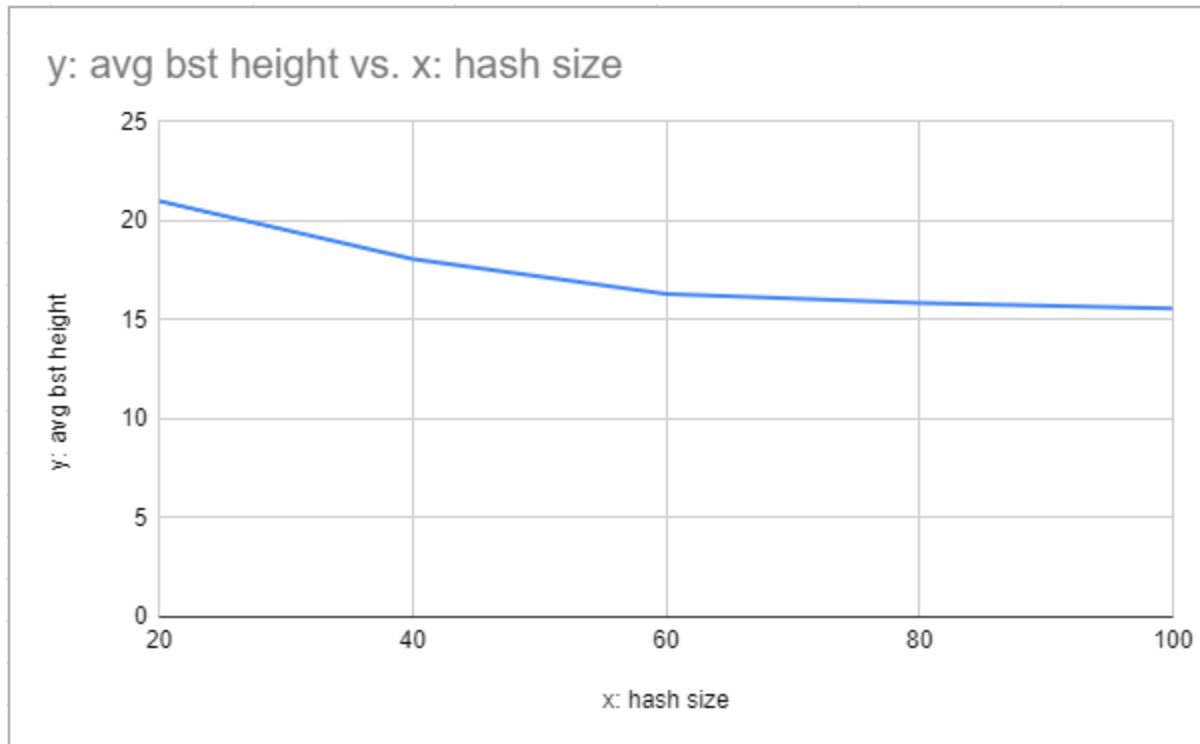
As you can see from the evidence of the output, changing the bloom filter size won't affect anything. The average BST height stayed the same given the different bloom filter size I have inputted. If you take at the evidence below,

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -t 20 < newspeak.txt
Average BST size: 728.450000
Average BST height: 21.000000
Average branches traversed: 10.552918
```

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -t 40 < newspeak.txt
Average BST size: 364.225000
Average BST height: 18.075000
Average branches traversed: 9.209246
```

```
wichapas@wichapas-VirtualBox:~/wpichetp/asgn7$ ./banhammer -s -t 60 < newspeak.txt
Average BST size: 242.816667
Average BST height: 16.300000
Average branches traversed: 8.413639
```

With the increment of 20 for each hash table size inputted, there's a slight change in height.



So looking at this graph, we can see that as the size inputted increases, the average binary search tree decreases. The x-axis goes from 20, 40, 60, 80, 100 and the y-axis goes from 21, 28.075, 16.3, 15.8625, 15.57. My hypothesis for this is that as we increment the hash table size entries, the average BST height decreases due to the formula being total divided by `ht_count` and thus giving us a negative trend to the graph. The larger the input, the more work that is done in the hash table will decrease the amount of work being done in BST. Thus, explains the situation where the average height and size that decreases overtime as hash table size entries increases. It's also safe to say that this is one of the factor that changes the BST height. It's even more evident when I give it some large number like here:

```
wlchapas@wlchapas-VirtualBox:~/wp1ch2p/asgn7$ ./banhammer -s -t 5000 < newspeak.txt
Average BST size: 3.085999
Average BST height: 2.605168
Average branches traversed: 1.274820
```

However, one interesting fact to keep in mind is that the numbers of the AVG branches traversed output has been decreasing over the span of increasing hash table size. This is where another interesting observation about the effect of changing the bloom filter size comes in. Since changing the number of hash table size, logically it should mean that the AVG branches traversed should decrease since the number of lookups increases. The formula for AVG branches traversed is  $\text{branches} / \text{lookups}$ . But since branches may also increase thus causing the changes to the AVG branches traversed to be slight.

All in All, I think I have observed how such changes in hash table size can greatly affect the statistics. My hypothesis was that as we increase the hash table size, the BST height decreases. It stays true as the graph is shown above but the trend is that as the size gets bigger, the decrease in height diminishes.