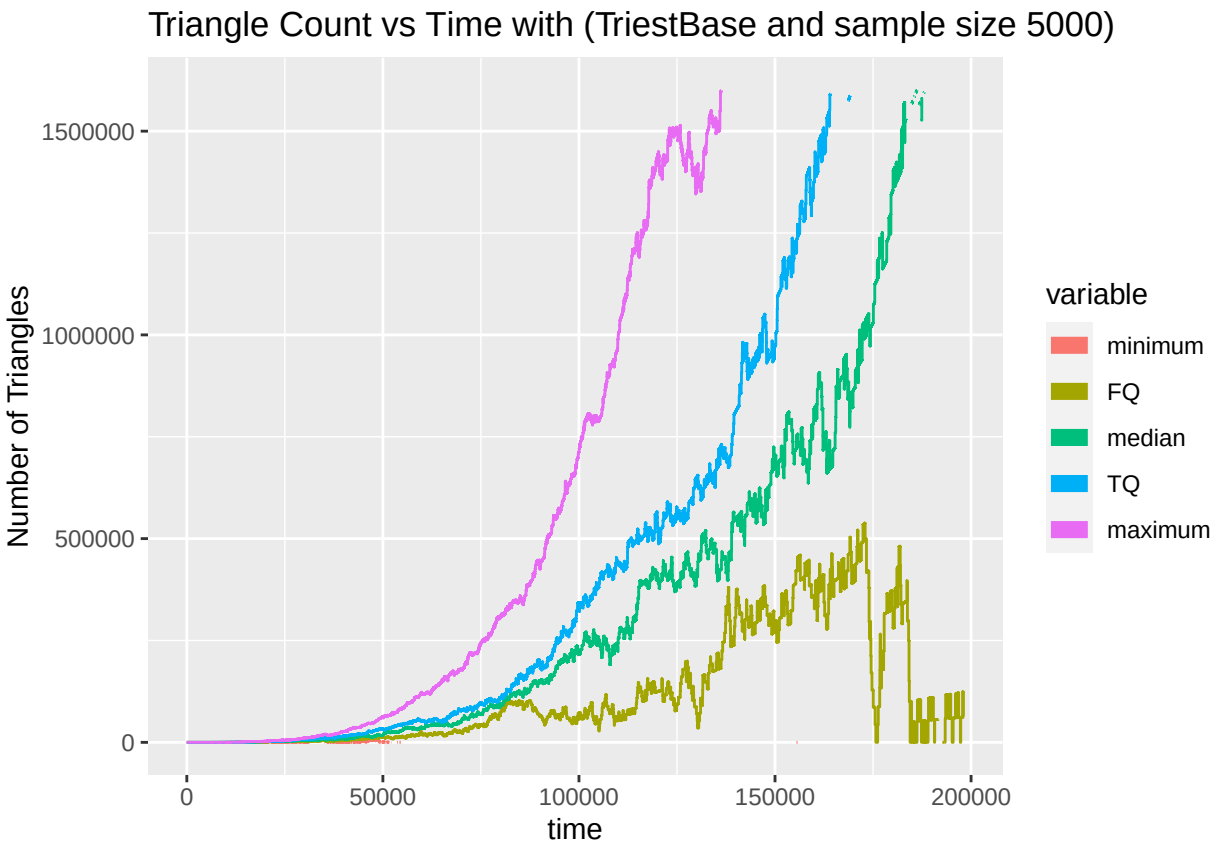
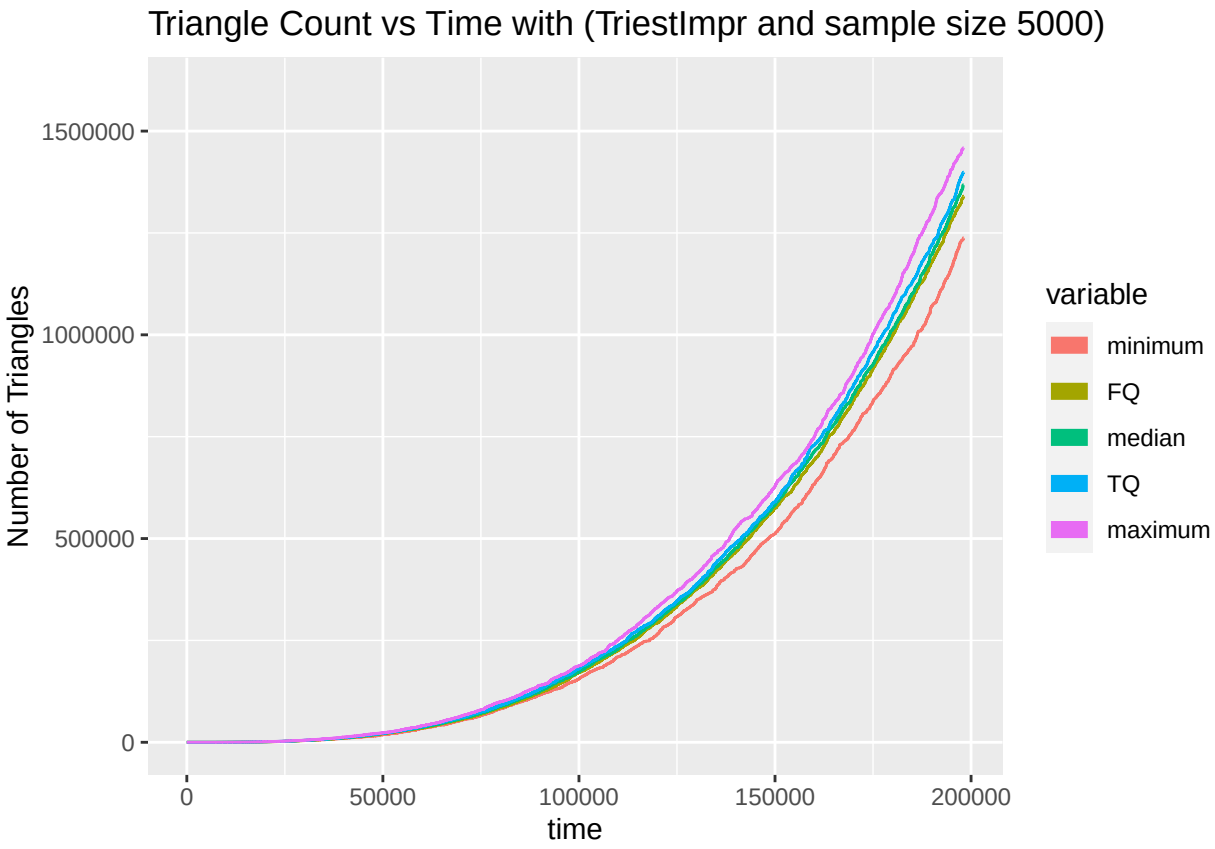
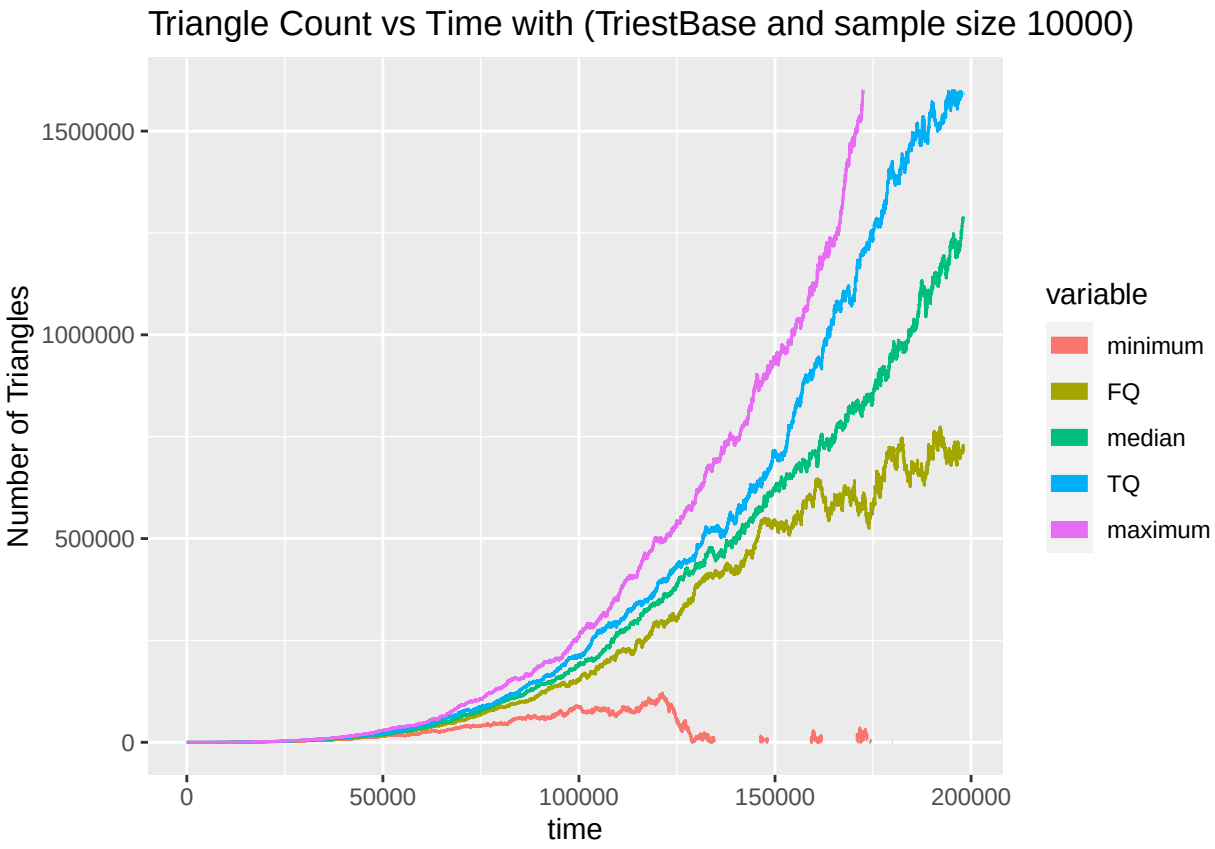
HW03: Approximate Triangle Counting from Data Stream

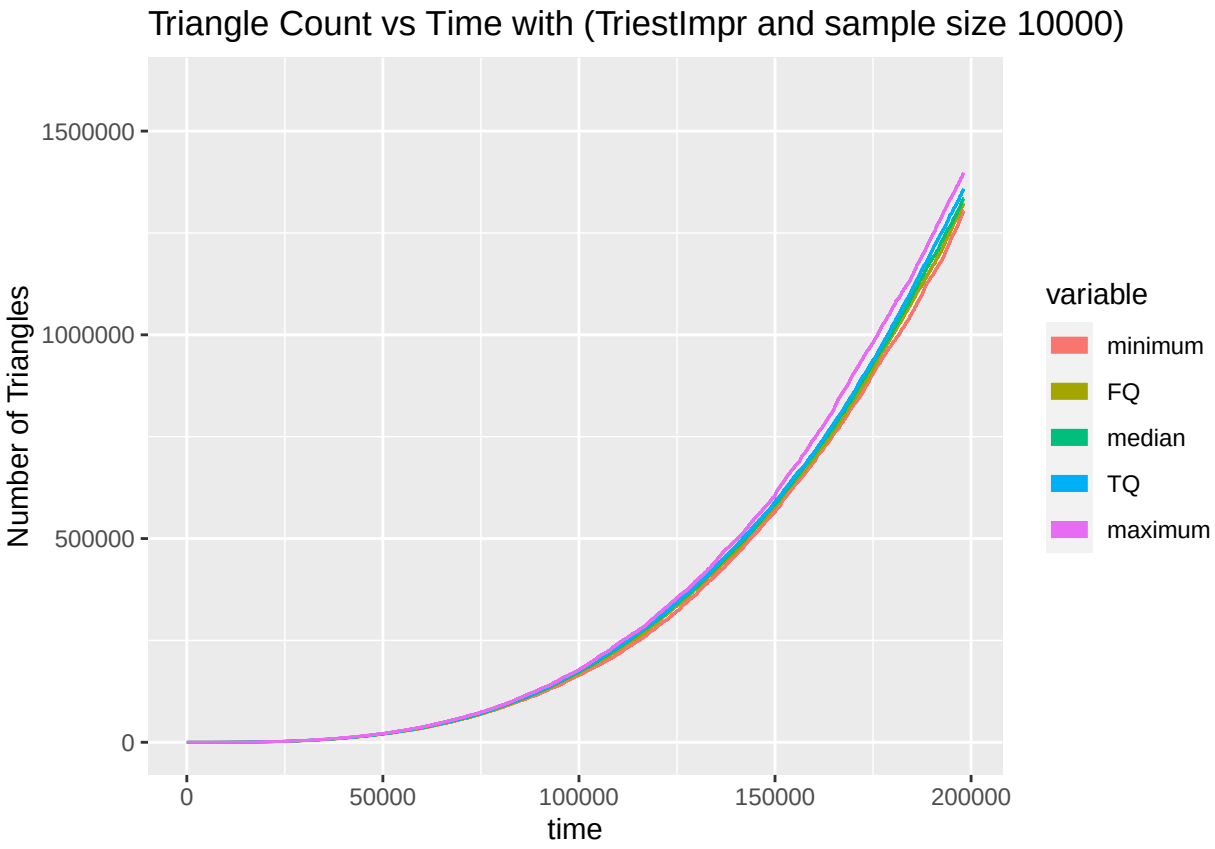
Drake Kufwafwa

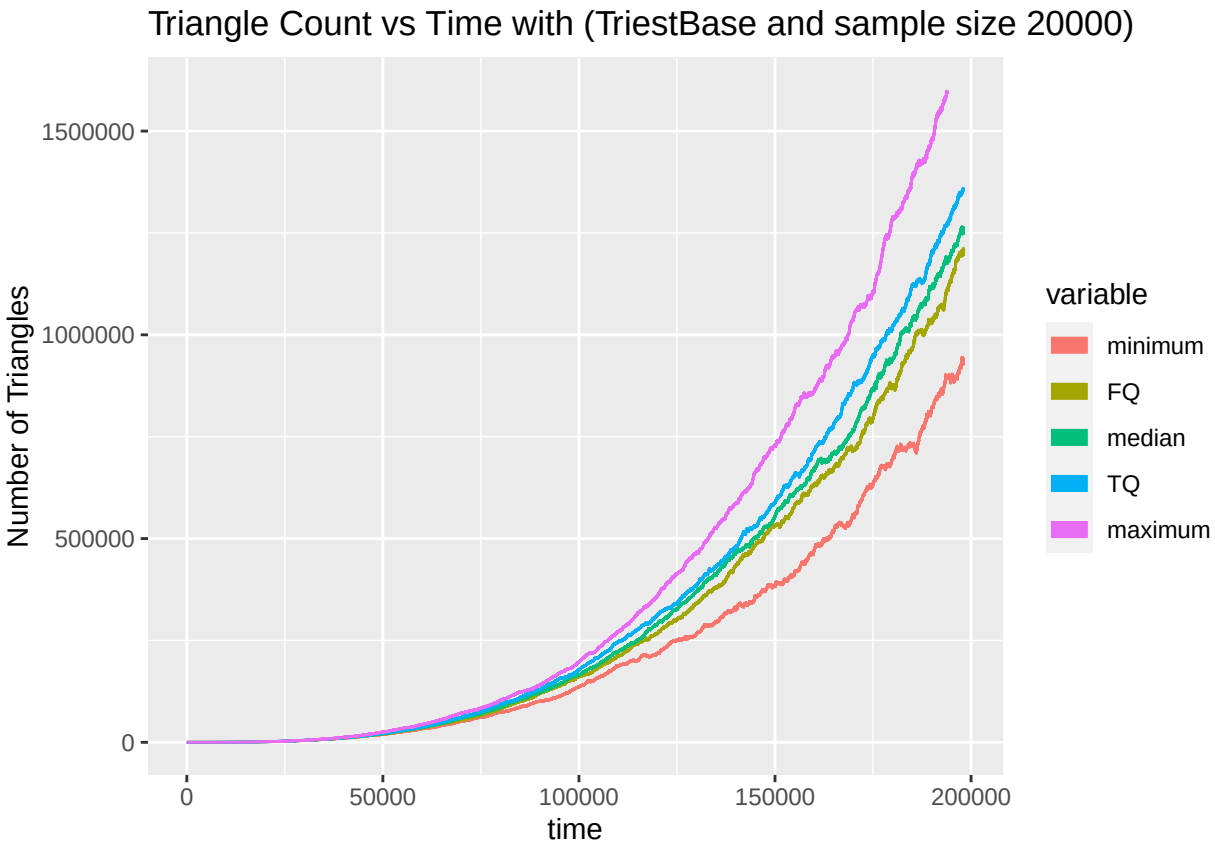
5/7/2021

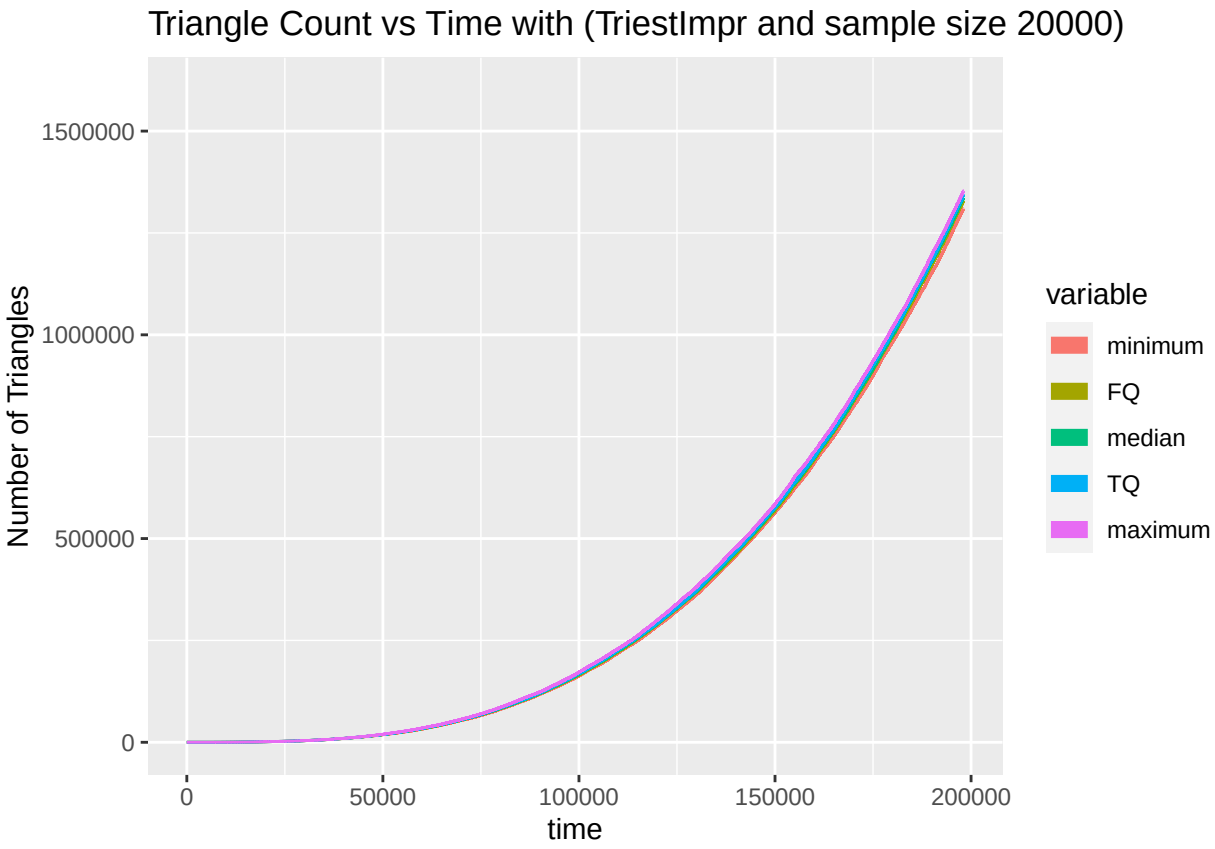


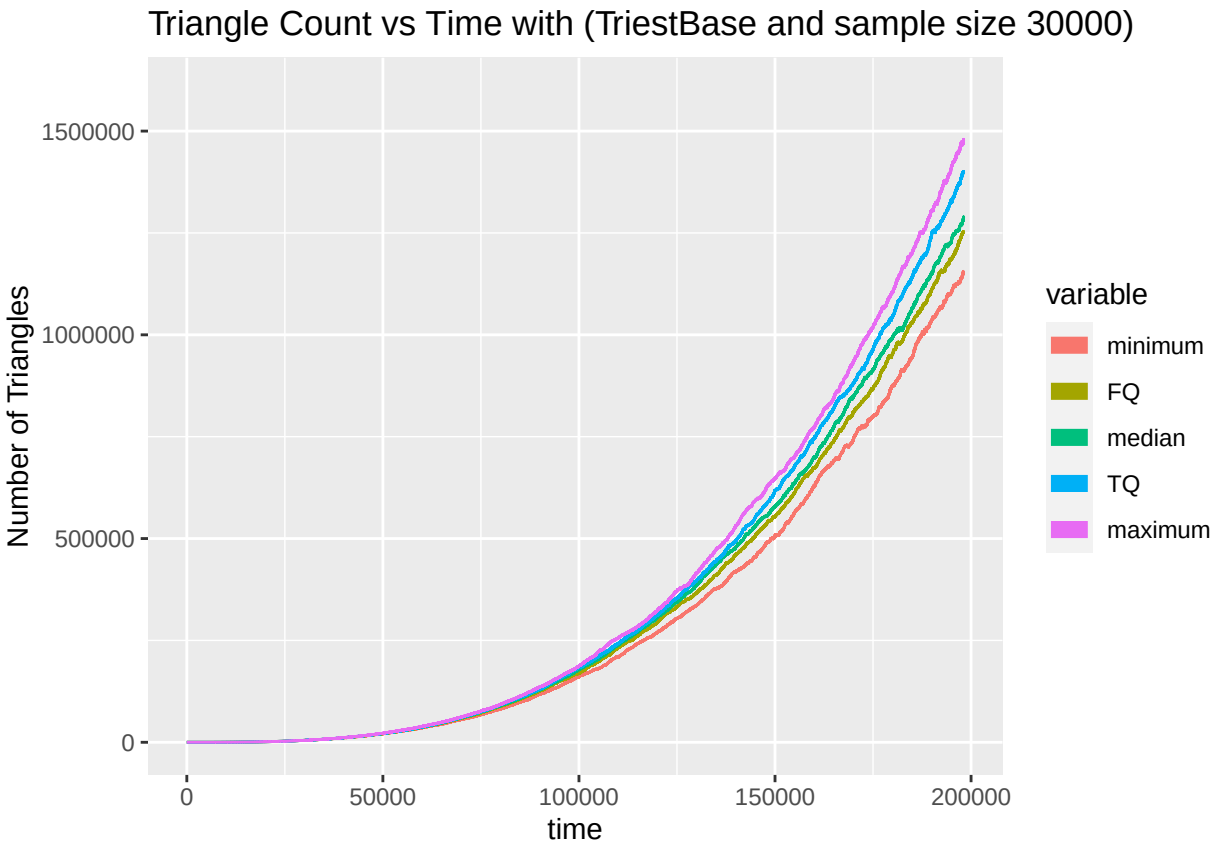

Triangle Count vs Time with (TriestImpr and sample size 5000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum

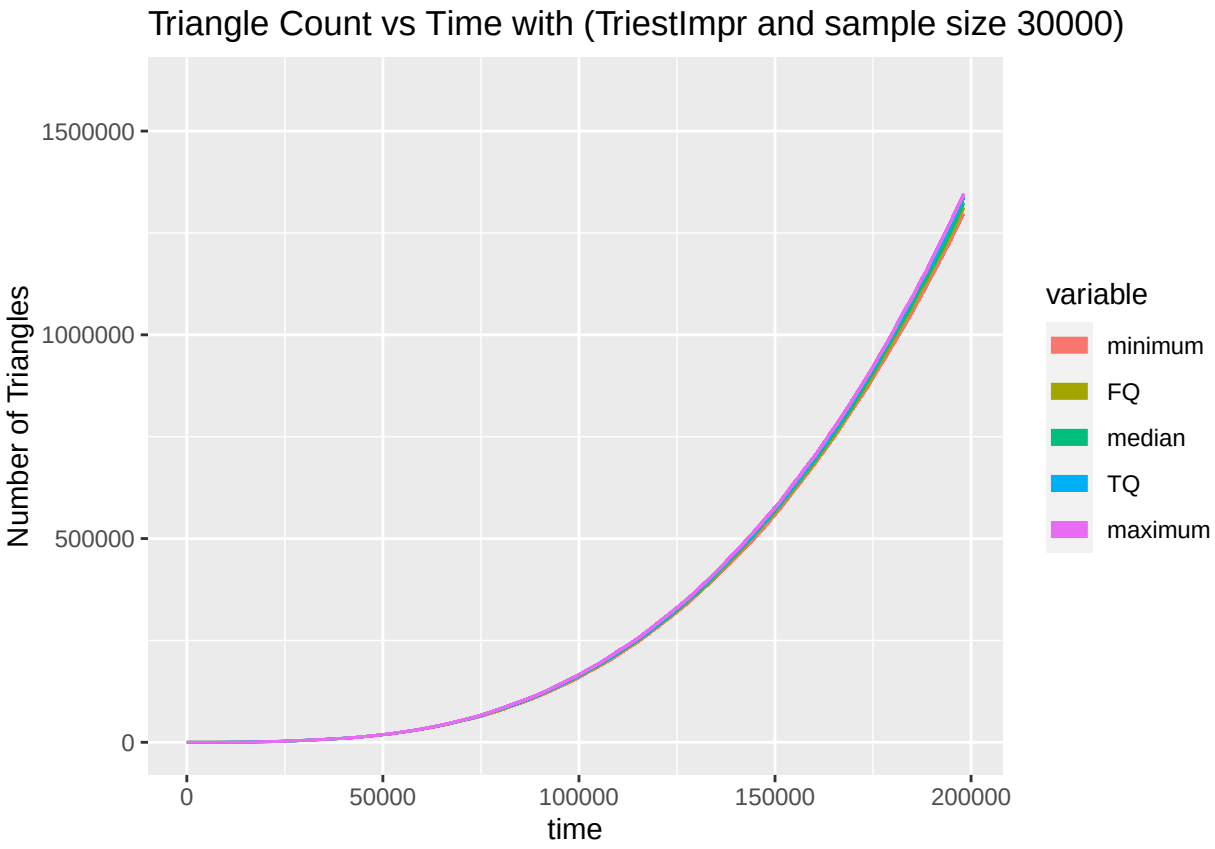

Triangle Count vs Time with (TriestBase and sample size 10000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum

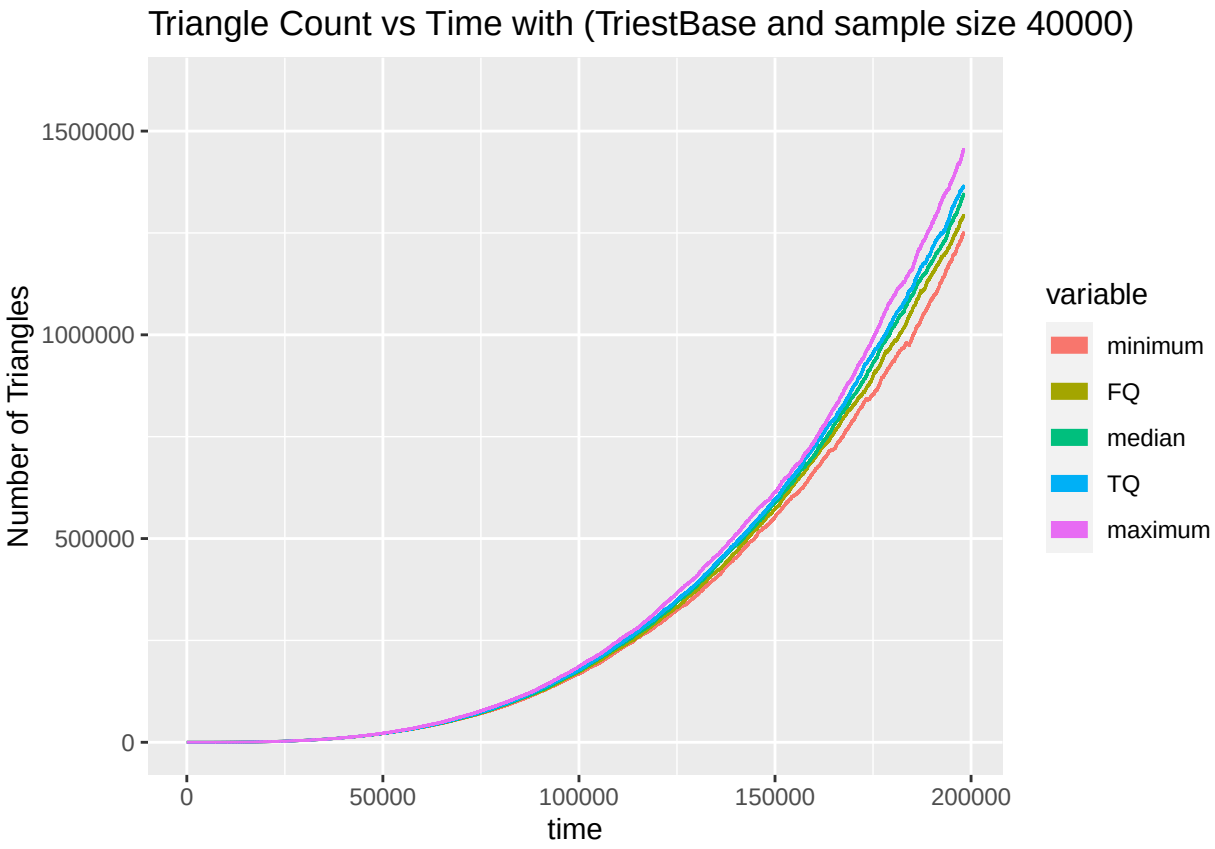

Triangle Count vs Time with (TriestImpr and sample size 10000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum

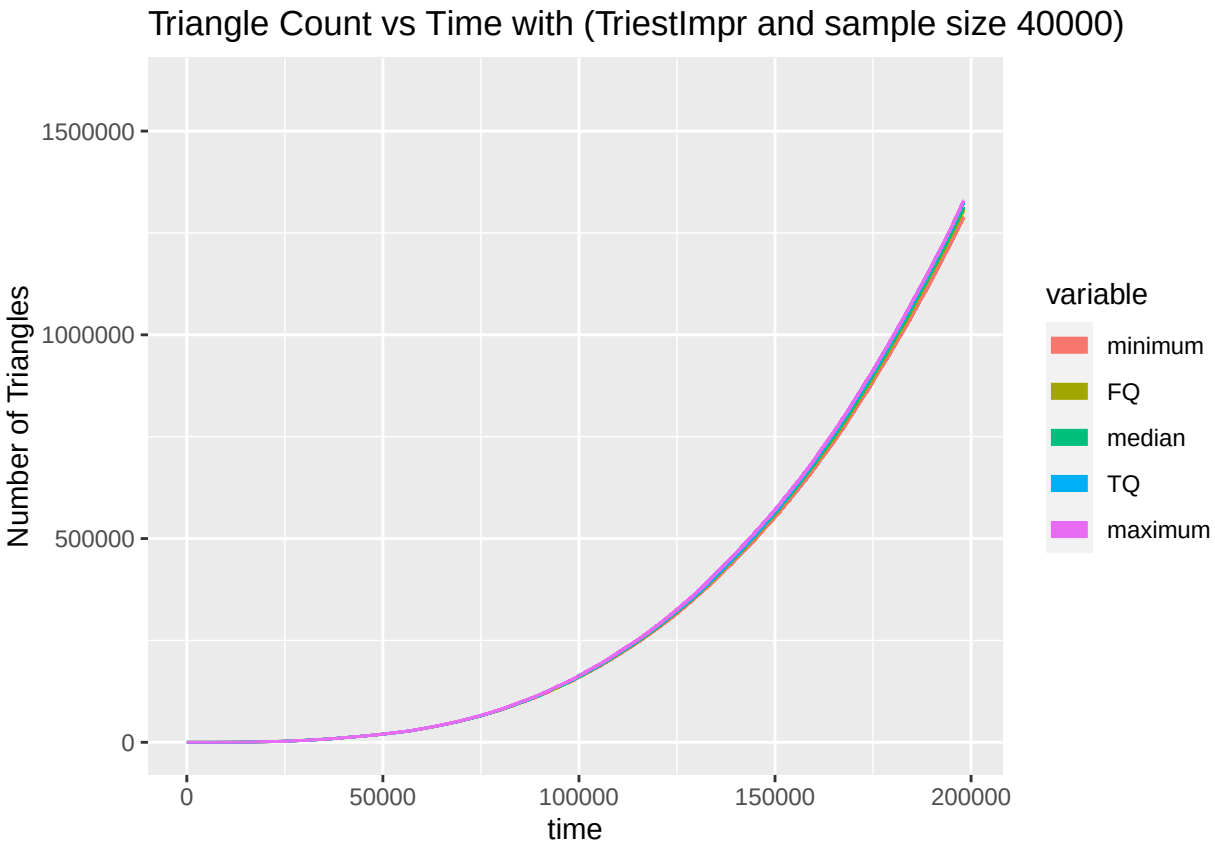

Triangle Count vs Time with (TriestBase and sample size 20000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum


Triangle Count vs Time with (TriestImpr and sample size 20000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum


Triangle Count vs Time with (TriestBase and sample size 30000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum


Triangle Count vs Time with (TriestImpr and sample size 30000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum


Triangle Count vs Time with (TriestBase and sample size 40000)
Number of Triangles
1500000
1000000
500000
0
0
50000
100000
150000
200000
time
variable
minimum
FQ
median
TQ
maximum



I used `ca-AstroPh.txt` as my input graph. From the plots, we notice that, as sample size increases, the algorithm is more accurate and smooth. This is supported by the Central Limit Theorem [big enough sample sizes]. As `m` increases the curves for each of the minimum, maximum, median, and first and third quartile over the 20 runs for each algorithm diverge less and less from each other. Furthermore, the gaps between the curves for the measures are much narrower for the graphs by `TriestImpr` than `TriestBase`.