Names: Bob Skowron, Jason Walker
Keys: rskowron, jwalker
SVN: jwalker: https://svn.seas.wustl.edu/repositories/jwalker/cse427s_fl17/

1. a. See SVN

   b. 103701 Ferris Bueller's Day Off
   95216 Rain Man
   94398 Seven
   92377 The Godfather
   92029 The Incredibles
   90891 Pretty Woman
   88670 As Good as It Gets
   82862 The Italian Job
   81889 Terminator 2: Extreme Edition
   78936 When Harry Met Sally
   78892 National Lampoon's Vacation
   76587 Beverly Hills Cop
   76473 Office Space
   75145 Air Force One
   70400 Sweet Home Alabama

Names: Bob Skowron, Jason Walker
Keys: rskowron, jwalker
SVN: jwalker: https://svn.seas.wustl.edu/repositories/jwalker/cse427s_fl17/

2. a. job.setCombinerClass(SumReducer.class);
No, the output does not change.

b.

Counter	With Combiner	Without Combiner
FILE: number of bytes read	99,540	156,981,268
FILE: number of bytes written	373,984	236,309,496
HDFS: number of bytes read	58,524,136	58,524,136
HDFS: number of bytes written	47,019	47,019

The number of bytes written and read from HDFS do not change because the data read from HDFS (namely the TrainingRatings dataset) is the same in both cases. They both need to read the entire set and regardless of using the Combiner or not, we get the same answer and thus the same output bytes. The bytes written and read though on the filesystem do change, and are reduced with the Combiner. This is because the Combiner reduces the amount of intermediate data produced by the mappers by collapsing all the Mapper output from a given compute node before writing the intermediate data.
Number of key-value pairs combined: 3,192,295 - 1,789 = 3,190,506

c. We can use the SumReducer as the Combiner because:

- Our reduce step is summation, which is commutative and associative
- The output data types of the Combiner and Mapper are the same (Text, IntWritable)

d. A Reducer that finds the median would not be able to be used as the Combiner because it is not associative