Team 8
Spencer Howlett
Nick Kowalczyk
Trevor Isaacson

CS455 HW3 Results and Analysis

1. The best day of the week for AQI scores was Monday with an average score of 30.88. The worst day of the week for AQI scores was Friday with an average score of 31.38. To find these results, we used a simple map reduce program. The mapper parsed the csv file into AQI score and day of week and the reducer computed the average associated with each day of the week.
2. The best month for AQI scores was October with an average score of 28.63. The worst month for AQI scores was May with an average score of 32.89. Similar to the first question, the mapper parsed the csv file into AQI score and month. The reducer then took those outputs and computed the average associated with each month of the year. Its interesting to see the variation between average AQI scores and the month that is occurred.
3. The 10 counties that had the best average AQI scores for the year 2020 were G0800430, G0801130, G0800970, G5300130, G0801070, G1300550, G1201070, G1300730, G1200230 and G5300570. County G0800430 had an average of 12.64. The mapper parsed the csv file into county code, year and AQI score. It also limited outputs to just the year 2020. The reducer then computed the average scores for each county and outputted the ten lowest and best AQI scores.
4. The 10 counties that had the worst average AQI scores for the year 2020 were G0601010, G0600070, G0600630, G0601070, G0600510, G0600710, G0600390, G0600290, G0600190 and G0600310. County G0600310 had the worst score of 64.37 and each of the ten worst counties were located in California. The mapper and reducer are very similar to Q3 except the ten highest and worst AQI scores are outputted.
5. For this question, we were unable to reach a specific answer. We approached the problem using chained jobs. The first job has the responsibility of mapping county codes, times and AQI scores to the reducer. The first reducer would then compute the average for a given week. The second mapper would map county averages in time order and the second reducer would compute differences and output the largest weekly change.
6. For our own analysis, we decided to analyze the average AQI score by state for the year 2020. We did this by getting the 3 digit state code from the GIS county code, and mapping each state to a reducer which computed the average. Our analysis found that Florida had the best average AQI scores in 2020 with an average of 26.54 and California had the worst average AQI scores with an average of 42.12. California had an average about 10 points higher than the next worst. We were interested to see how air quality differs by state. This analysis could then be applied to other years and a timeline of air quality averages could help determine future air qualities for different states. Additionally, this analysis can be utilized further to compare counties with the state average, as well as seeing whether there is a correlation between the best/ worst AQI score counties and

the best/worst AQI scoring states. We believe this is important because there may be socio economic, as well as geographic differences that can cause it.