

HW#6 Part 2: Due by 11.59pm on 7-17-2018

Use data.csv to

1. Run K-mean clustering with 3 clusters and visualize results using a pairplot with each cluster highlighted different color.
2. Identify number of clusters using the elbow method. Plot Num Clusters vs SSE for Num Clusters $\in \{2,3,4,5,6,7,8,9,10\}$.

Using code on simulating linear regression discussed in class

3. Investigate what happens to the run time when we vary number of observations? (m). Plot number of observations on the x-axis and time it took to run it on the y-axis. Provide a linear, $\log n$, and $n \log n$ benchmarks for comparison.