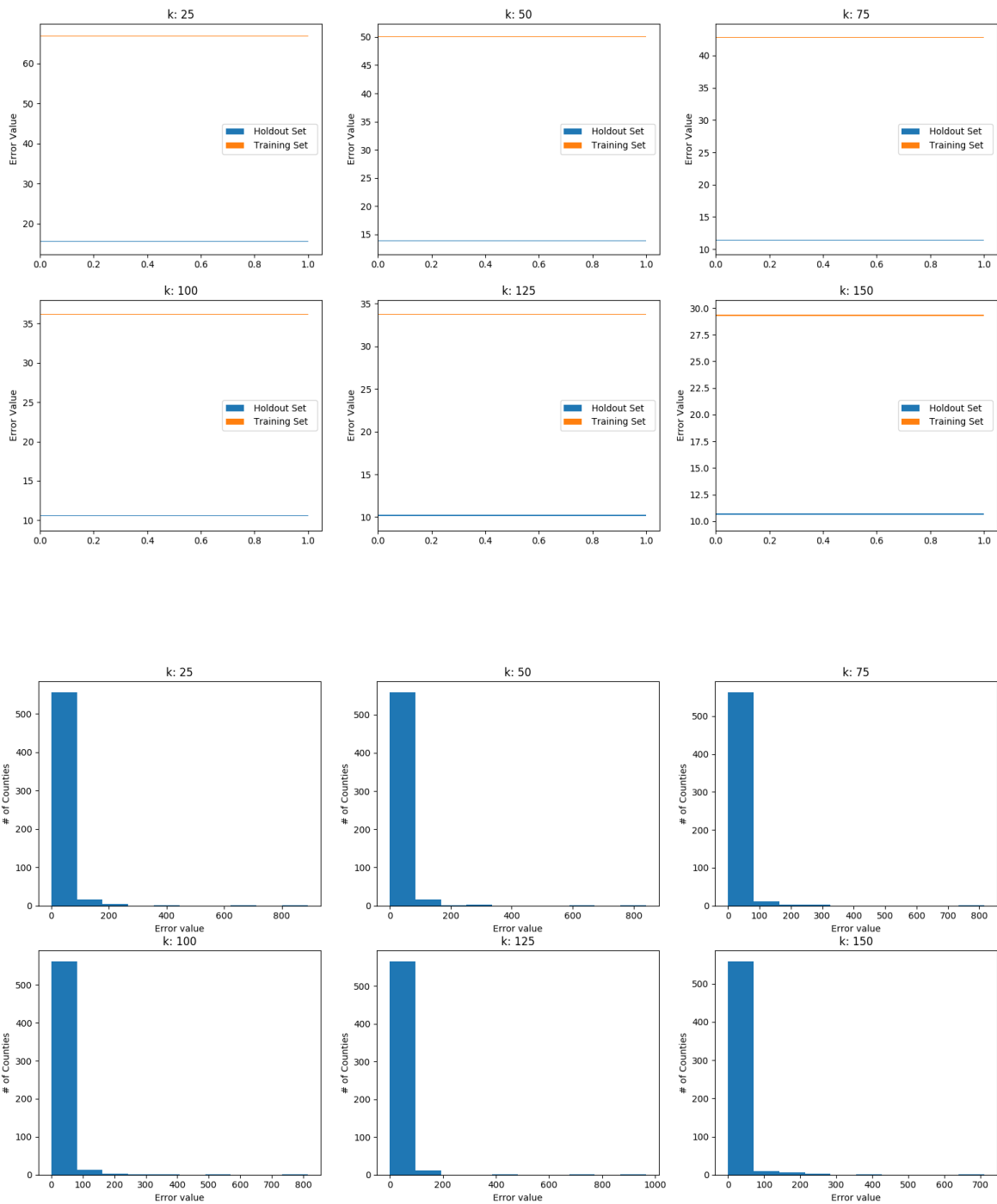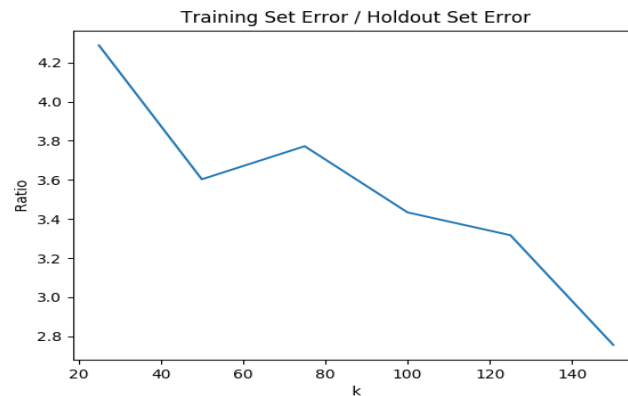# Problem #1: Graphing Removed Error

Looking at the first graph, while k-means increases, the error of the current dataset decreases. Larger dataset has more acceleration trend.

Training Set Error / Holdout Set Error

Even if larger datasets have more acceleration trend regarding k-means value, larger k-means means less gap between two total error values.

First graph indicated relevance between less acceleration trend and smaller datasets. More data means more rare situations arises and standard deviation likewise.

Bigger dataset, bigger total error value.

Problem #2: k-means and you

1. No, because k-means choose random initial centroids for k cluster. Similar, not the same.
2. Yes, in a similar cluster with error value. Because, even if k-means is somewhat random, after initial points it clusters likely aka. nearest points.

Problem #3: Graphing Prediction Error

If k-means value goes higher, prediction value is likely to get higher too. As we are calculating here, we do not include the value in the clustering criteria in k-means algorithm.

While k goes higher, the lost criteria's – which is poverty in this graph, influence on clustering algorithms gets more significant too.

Problem #4: Graphing Prediction Error (AGAIN!)

1. New Vector = [1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1]
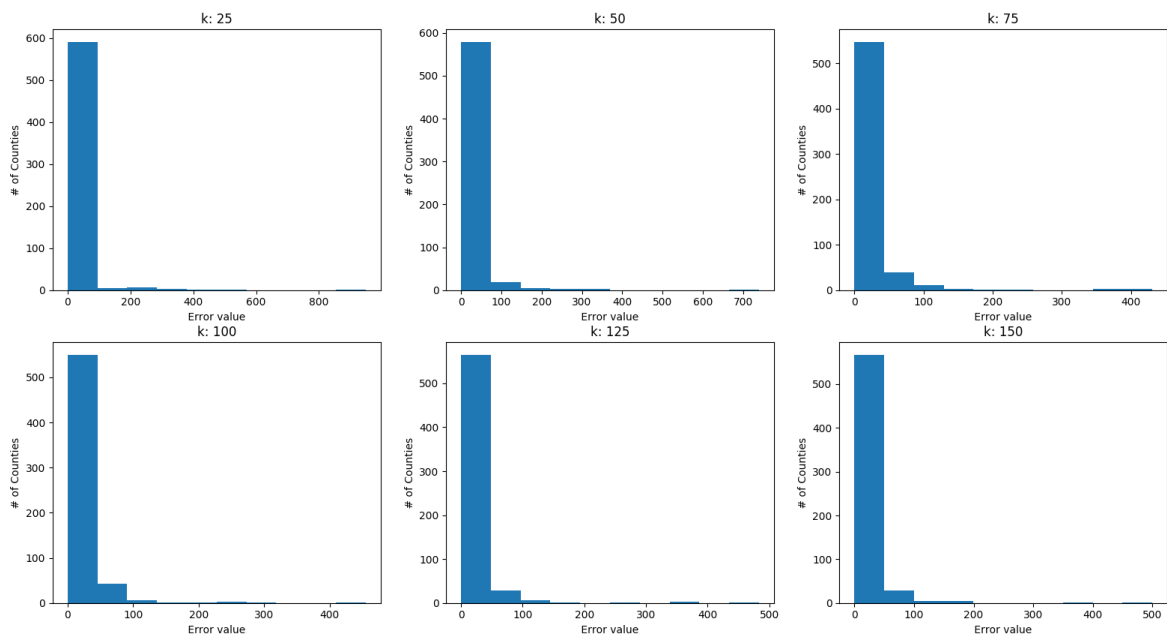
If poverty is high, home value, below 18 which is birth rate and collage graduate will drop; they have reverse effects. If you eliminate those values, result will be more likely close to poverty parameter. Predicting poverty will be much easier with k-means clustering.

2. New Vector = [1, 0, 0, 0, 1, 0.25, 0, 0, 0.5, 0, 1, 1, 1, 1]
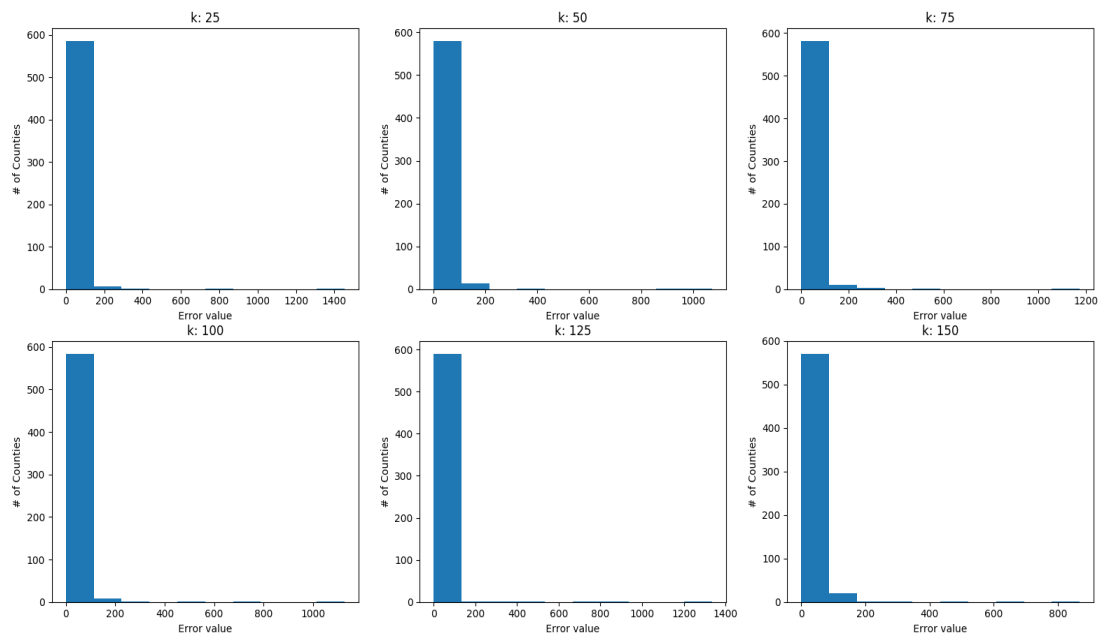
Percent65+ and PercentHSGraduate are less relevant to Poverty than other paramaters. PercentFemale and PopDensity are irrelevant enough to disregard for Poverty while clustering.