Up to this point

* Imported the data in MATLAB
* Switch to R to clean the data better and format appropriately
  + In this step I wrote out 5 .csv files containing 100 observations each, 50 of the best cities and 50 of the worst cities
* Switch back to MATLAB to find the best purity index on the data set
* The best purity score was 0.89 on K=5,10,15,20, and 25 was at K=10
* s