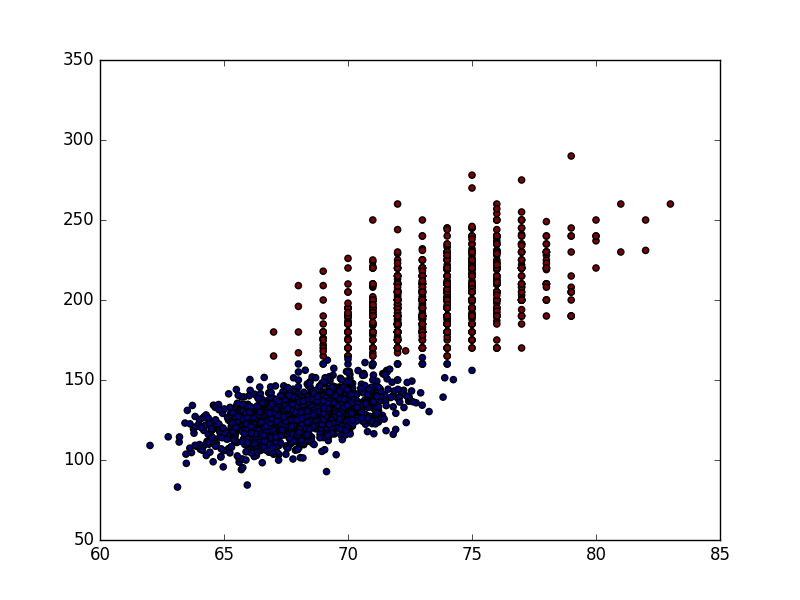
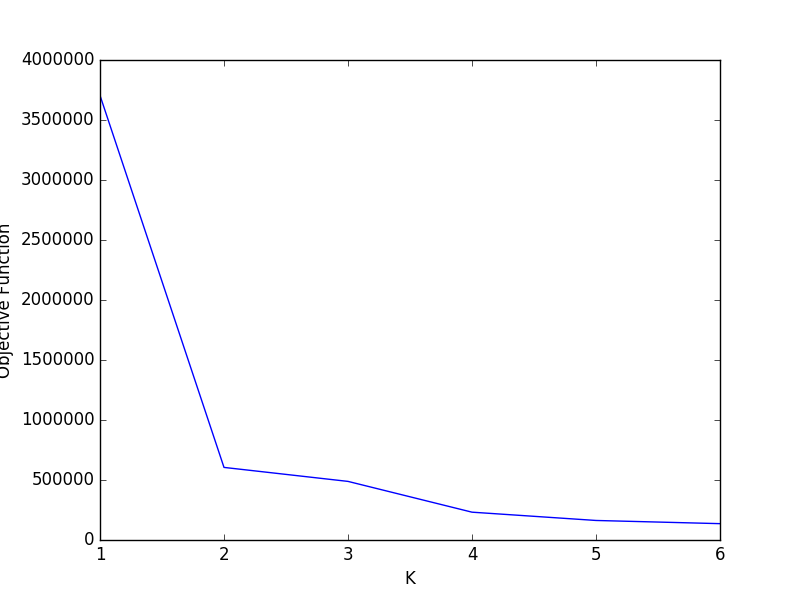
Q3

Q3:



Q4:

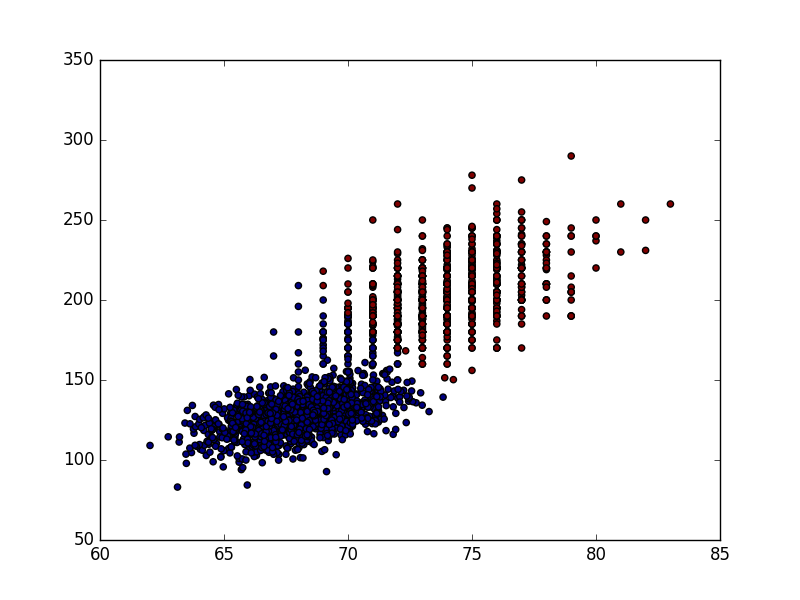


The graph clearly suggests K=2 is the best pick of K for this data.

Q5: using the K=2 value found in our knee-finding process, these are the purity values of those 2 clusters after 50 iterations:

Purity values for each cluster: [0.97957516339869277, 0.99900990099009901]

Q7:



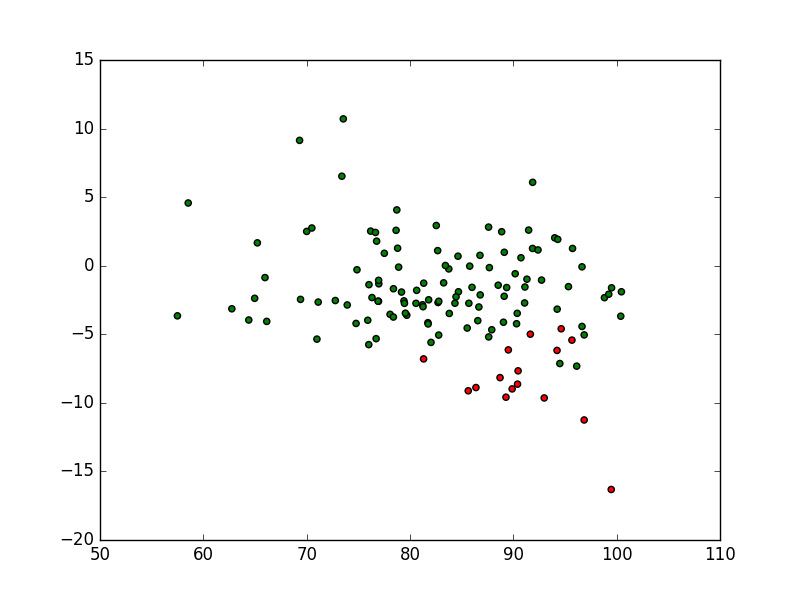
Diag:

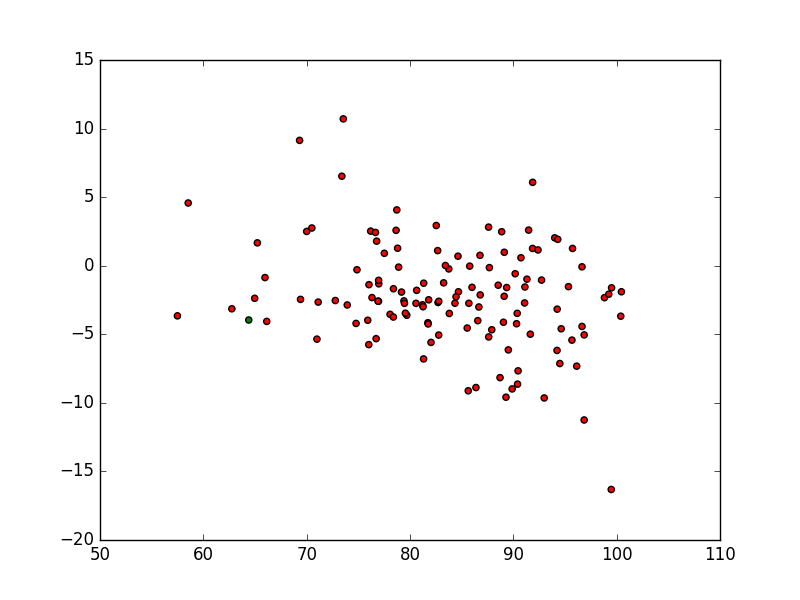
Full:

For some reason my dataset1 stopped working when I got dataset2 to work, I don’t know why but the above graph is from that.

Q8:

Diag:





Full:

Q9:

Purities:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Covariance type | Iterations | Purity Group 1 | Purity Group 2 |
| Dataset 1 | Diag | 20 | .93 | .9779 |
| Dataset 1 | Full | See above for explanation |  |  |
| Dataset 2 | Diag | 200 | 0.6875 | 0.5267857142857143 |
| Dataset 2 | Full | 200 | 1.0 | 0.50393700787401574 |