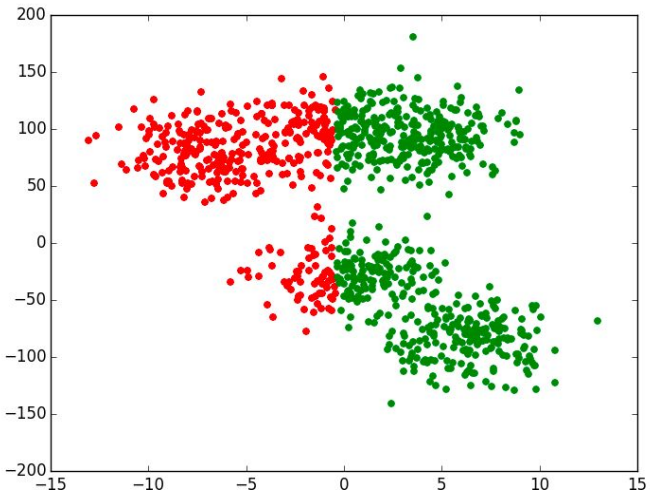
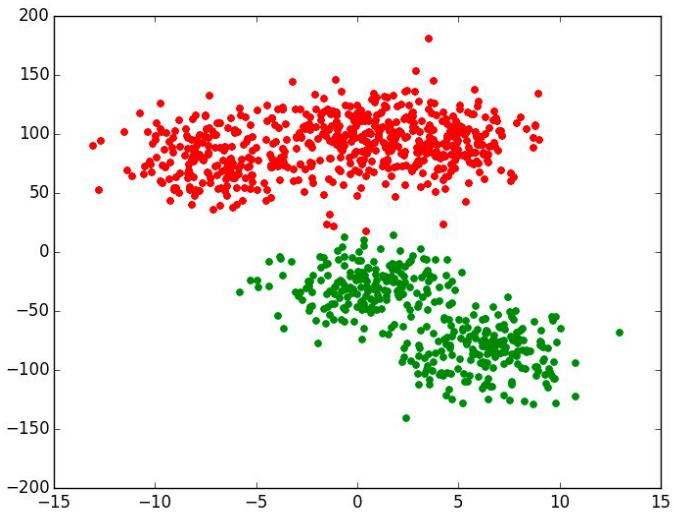
We implemented the function `kmeans` in the included `coli_ex_8.py` to calculate the k-means clustering of a data set with any number of dimensions.

Task 1

We found the following clustering just using the x-values:



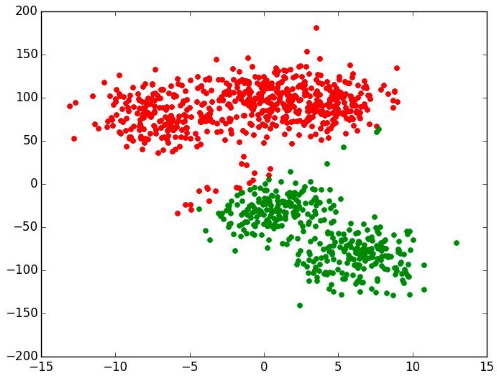
And just using the y-values:

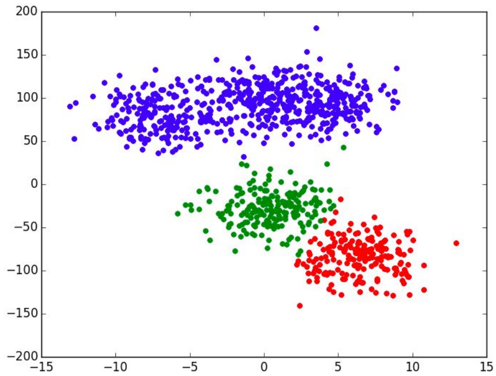


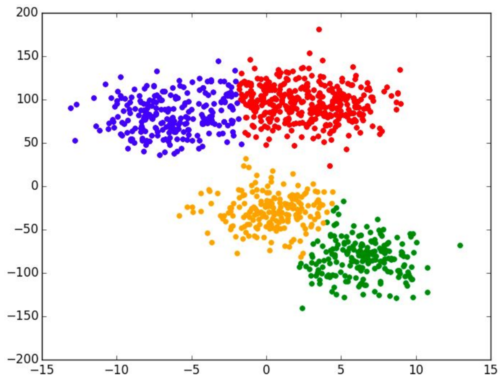
The y-value clustering is clearly better for this data set.

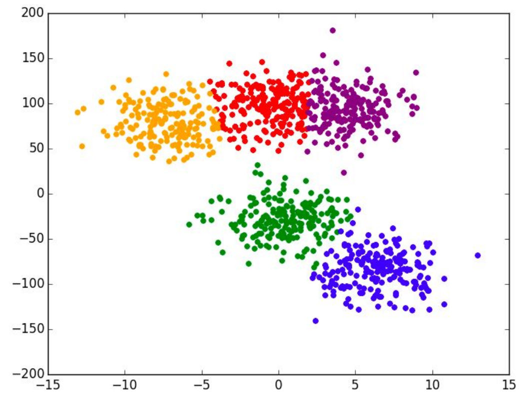
Task 2

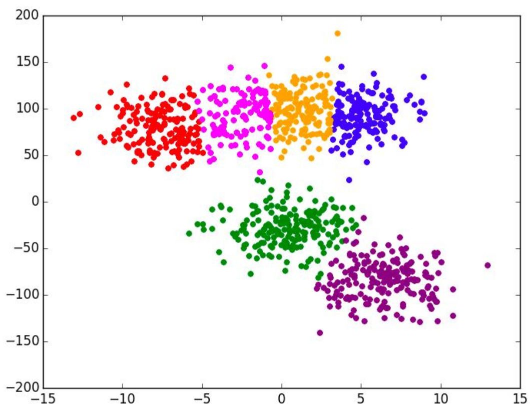
We normalized by calculating the *z*-score of each data point, in the function `normalize`.
We found the following clusters for *k* between 2 and 6:











Task 3

We implemented the Variance Ratio Criterion in the function `vrc`, and the function to calculate the minimum in `min_vrc`.
Using this, we found that the optimal value for *k* was 5 (or occasionally 4, depending on the particular clusterings from the `kmeans` function).
We were unable to include the results for *k* = 2, however, since we were unsure how to calculate VRC(1) without dividing by zero.