Question 3.1

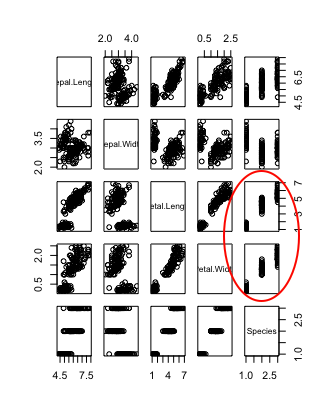
Question 4.1

A clustering model would be appropriate for segmenting voters into categories to determine whether they would be likely to support certain candidates or issues. Predictors could include income level, political party, age, occupation, or education levels.

Question 4.2

Please review the included R file 4.2.R along with this explanation.

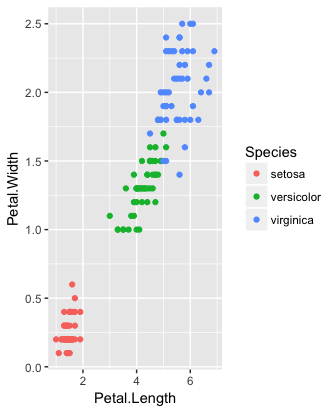
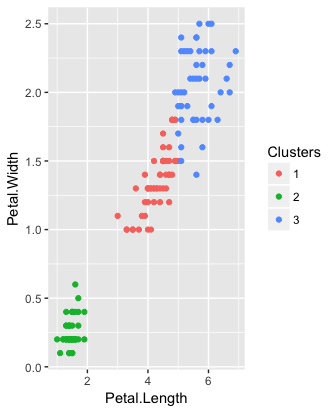
First, I loaded and inspected the data. I then plotted all combos of predictors. From the plot below, it appears that Petal Length and Petal Width are likely to be the best predictors, because they both appear to have a fairly linear relationship between the predictors and the species. There is little overlap between each species in a given range of petal lengths and widths.



I then initialized the model and tried a variety of predictors. The best results were with k=3 and Petal Length and Petal Width as predictors, as expected. This combination provided 94.3% accuracy.

I then compared the predicted clusters from this model with the actual clusters by species:

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted | Actual Species | | |
| Clusters | setosa | versicolor | virginica |
| 1 | 50 | 0 | 0 |
| 2 | 0 | 2 | 46 |
| 3 | 0 | 48 | 4 |



The table and plots above show that predicted clusters are generally very close to the actual species, correctly classifying all 50 setosa data points and 94 out of 100 of the remaining points.