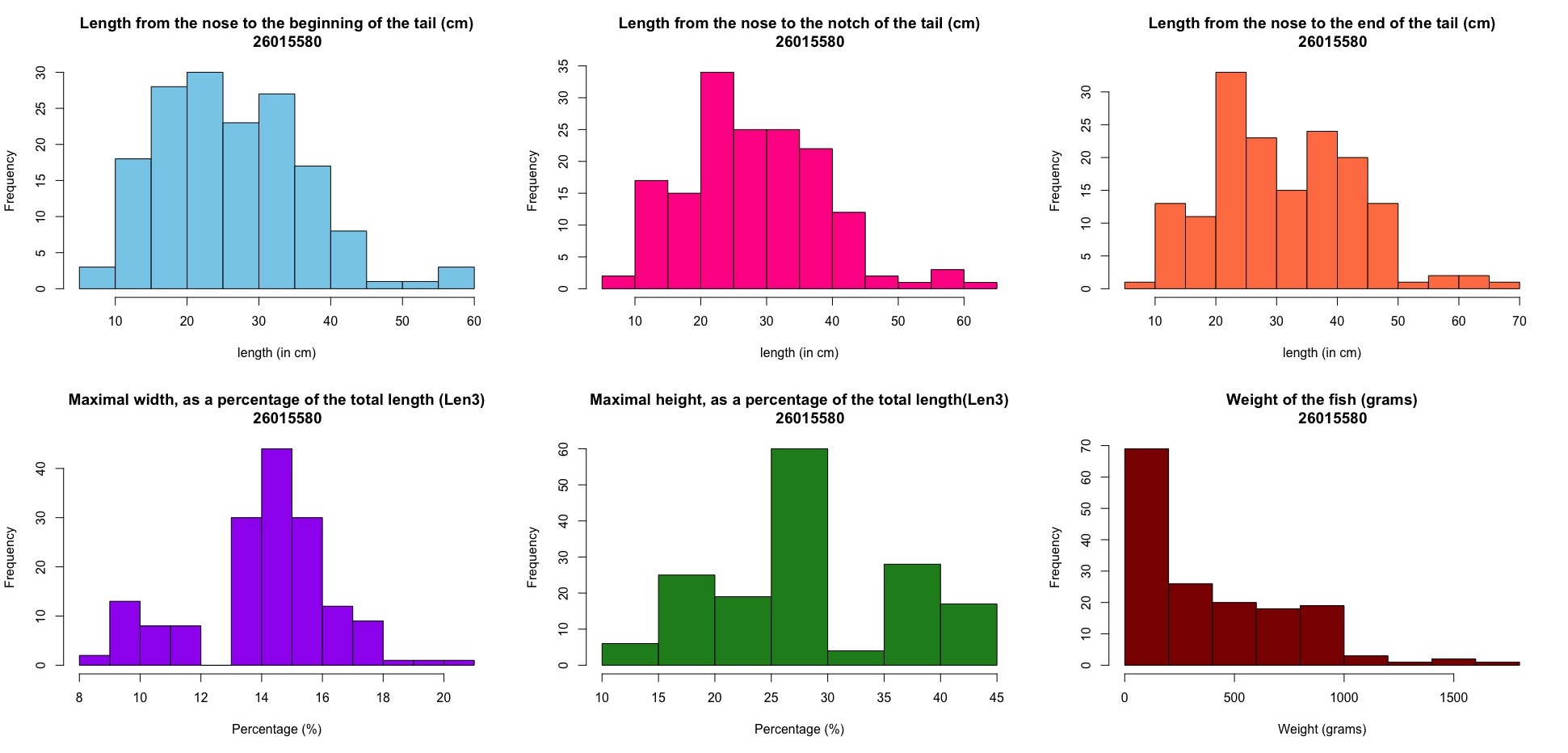
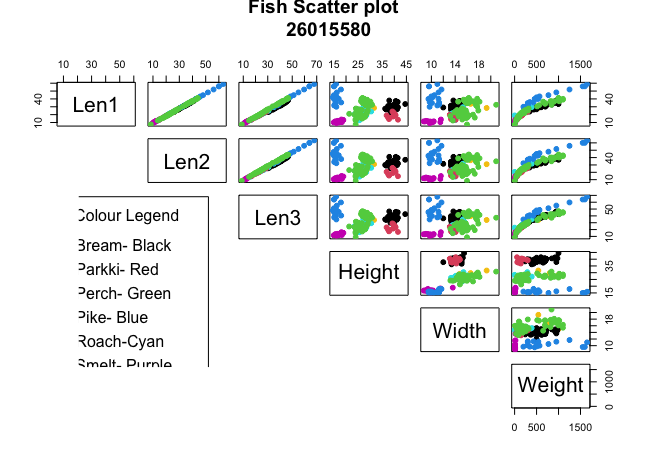
**Multivariate Analysis of the Measurements of fish**

Exploratory analysis - histograms



The histograms for the maximal width and height have very clear bimodal distributions in the width histogram is show two groups a group of fish with ~9.5 units width and another with more than ~14.5 units in width, the second being the maximal height shows a group with ~27 units tall and another with ~37 units tall. There are also slight hints of a bimodal distribution in Len 3 and Len 1. I used a histogram plot for each variable because it provides a starting point and allows me to start seeing signs of groupings among the fishes. I also left out the sex of the fishes as they serve no purpose to the analysis also, they are a Boolean variable type which will be incompatible with the plots I will be using.

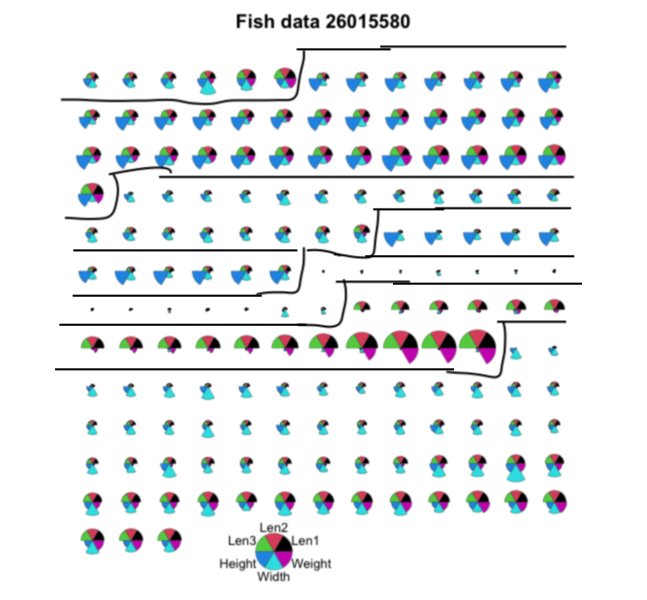
Scatterplot matrix



The results from the scatterplots show that there are positive correlations between variables which show that if a fish has a large measurement in one area it will most likely have a large measurement in another area, this is seen in the length from the nose to the beginning of the tail, Length from the nose to the notch of the tail, Length from the nose to the end of the tail and even in the weight, the only exceptions being in the maximal height and the maximal width. There is also a positive linear relationship between the Length from the nose to the beginning of the tail, Length from the nose to the notch of the tail, Length from the nose to the end of the tail. In the plot of the Length from the nose to the beginning of the tail against Maximal height variable there are three separate groups of fish. And we can see there are three significant groupings in the height variable this shows this can be very useful to tell species apart.

This scatter plot graph was used because it allows us to see the relationships between all variables clearly in one plot and it also shows grouping which is useful for finding similarities between the fishes.

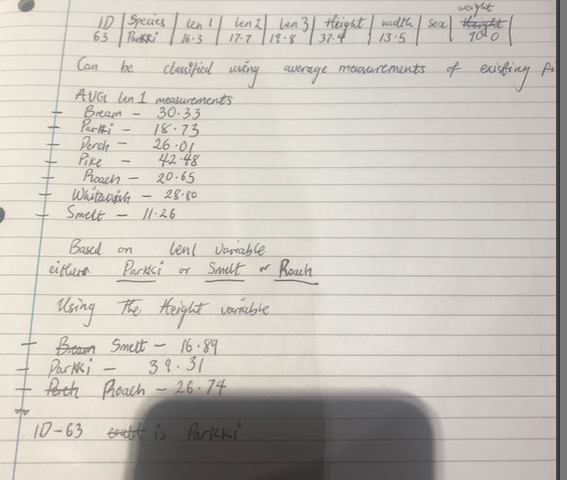
Segment plot



I have separated the fishes into true species. A total of 7 groups have been made - which I would be referring to them as letters A-G, from these groupings we can see that group B (Bream) fish then to be very long in terms of height irrespective of their other measurements, while group C(Roach) fish have small measurements from the nose to the tail and weigh less than other groups on average, group D(Parkki) fish tend to have very large heights just like the group B(Bream) fish but very low length measurement comparatively , group e(Smelt) fish all tend to have very low measurements overall ,group F(Pike) fish have very low heights and widths but unlike the group e(Smelt) fish have very large measurements from the nose to the tail. The remaining groups being group A(Whitewish) and group G(Perch) in terms of the measurements given are outliers and do not have much uniform differences from each other as the measurements of their variables are random and do not form and groups that fit in with the others of themselves even, because of this they will be removed from the analysis.

I used the segment plot because it helps easily identify the attributes of the different groups and it is easier to read than the Chernoff's faces and star plot.

Analysis Demonstration



For the demonstration I used the average measurement of the existing fish and a method of elimination to narrow down the possible groups the fish could belong to. I used Len 1 measurement as it served as a good starting point and it helped me to eliminate 4 of the other fishes, I then used the height measurement because it was the variable that split the fishes into groups the most on the scatter graph.