A graph of a bar graph

Description automatically generatedFigure 1: This is a bar plot showing the site location of base pairs found in dogs.

* Main Takeaway: There was a repeat of islands found in the Dog. This caused the cluster of data in the bar plot. This could be due to the dog’s formation, which would cauee multiple islands to be the exact same. It shows the number of sequence that is in the Site Location (base pairs). The Y axis is the island and the X axis is the site location.

A graph with a line

Description automatically generated

Figure 2: This is a box and whisker plot of the GC content within the islands for the dogs

* Main Takeaway: The average of the plot is 70%. The standard deviation goes up towards 75%, which is 5 away from the average. There are also some outliers that go up a little past 85%.

A graph with blue lines

Description automatically generated

Figure 3: This is another box and whisker plot that has the length amount of the islands within the dog sequences.

* Main Takeaway: In this, there are no outliers. The average value is around 1800. The lower quartile range is about 1000 whereas the interquratile range is about 2000.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | min | median | max | mean |
| GC\_CONTENT | 65.9 | 70.2 | 85.8 | 71.11 |
| Length | 206 | 1745 | 2761 | 1775.6 |
|  |  |  |  |  |

Figure 4: This table shows the min, median, max, and mean for both the GC\_content in dogs as well as the length of the islands of the sequences.

* Main Takeaway: The difference between the min and max between the GC conent is only 19.9 compared to the length which has a difference of 2555. The mean and the median for both the GC content and the length are very simmilar, which shows that our data was actually consistent.