Base Statistics:

File Name - Name of the file
File Type - The file contains actual base calls or colorspace data that gets converted to base calls
Encoding - The ASCII encoding of quality value that is in this file
Total Sequences - The number of sequences processed
Sequences Flagged as Poor Quality - The number of poor quality sequences
Sequence Length - The length of the shortest and longest sequence in the set
%GC - The percent of guanine/cytosine bases in all of the sequences

Per Base Sequence Quality:

- The central red line is the median value
- The yellow box represents the inter-quartile range (25-75%)
- The upper and lower whiskers represent the 10% and 90% points
- The blue line represents the mean quality
- The quality scores are on the y-axis. The higher the score, the higher the quality. Green is good quality, orange is moderate quality, and red is poor quality.

Per Tile Sequence Quality:

- The colors on the plot are on a cold to warm scale, with cold colors at positions where the quality was at or above the average for that base in the run, and warm colors indicate that a tile had worse qualities than other tiles for that base. The best quality data is completely blue.

Per Sequence Quality Scores:

- Per sequence quality shows the number of reads for each quality scores. If you got the highest peak for "30" it means Q30 has more reads than other quality scores. x-axis = the quality scores and the y-axis = the number of reads.

Per Base Sequence Content:

- Per Base Sequence Content plots out the proportion of each base position in a file for which each of the four normal DNA bases has been called. It is the percent of each base (ATGC) in the data.

Per Sequence GC Content:

- The bell curve measures the GC content across the whole length of each sequence in a file and compares it to a modeled normal distribution of GC content.

Per Base N Content:

- This graph raises a warning if any position shows an N content of >5%. Data is poor if N content is too high (>5%). If a sequencer is unable to make a base call with sufficient confidence then it will normally substitute an N rather than a conventional base call (ATGC). This module plots out the percentage of base calls at each position for which an N was called.

Sequence Length Distribution:

- The graph shows the distribution of fragment sizes in the file which was analyzed.

Sequence Duplication Levels:

- This graph shows the degree of duplication for every sequence in a library and creates a plot showing the relative number of sequences with different degrees of duplication.

Overrepresented Sequences:

- A normal high-throughput library will contain a diverse set of sequences, with no individual sequence making up a tiny fraction of the whole. Finding that a single sequence is very overrepresented in the set either means that it is highly biologically significant, or indicates that the library is contaminated, or not as diverse as you expected.

- This module lists all of the sequences which make up more than 0.1% of the total.

<u>Adapter Content:</u>

- The plot shows a cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position. Once a sequence has been seen in a read it is counted as being present right through to the end of the read so the percentages you see will only increase as the read length goes on.