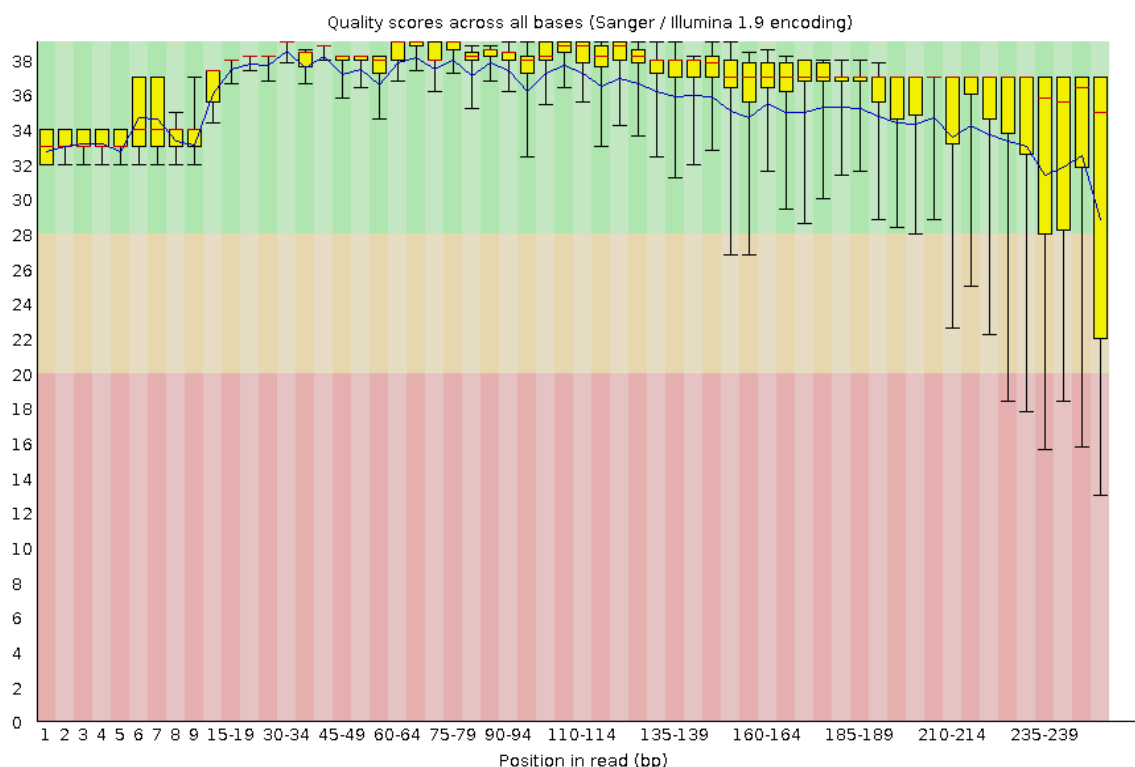


Para esta questão utilizei o arquivo “F3D0_S188_L001_R1_001.fastq” disponibilizado no pelo próprio desafio. Os softwares utilizados para realizar a análise da qualidade e a limpeza dos dados foram FastQC e Trimmomatic, respectivamente.

Primeiramente, o arquivo “F3D0_S188_L001_R1_001.fastq” foi aberto no software FastQC, onde foi possível observar o número total de sequências, bem como o número de sequências marcadas como sendo de baixa qualidade. Na aba referente à qualidade das sequências por bases (*Per base sequence quality*) foi possível observar que uma parte considerável das sequências estava abaixo de 30 na escala phred/phrap (Figura 1).

Figura 1. Score de qualidade das bases antes da trimagem.



Em seguida, o arquivo utilizado foi modificado pelo software Trimmomatic, disponível na plataforma Galaxy, a fim de eliminar porções de baixa qualidade das sequências. Após a trimagem, um novo arquivo fastq foi gerado (“Galaxy13-[Trimmomatic_on_F3D0_S188_L001_R1_001.fastq].fastqsanger”), o qual também foi submetido ao FastQC. O resultado da qualidade das bases observado foi melhor do que o observado antes da trimagem, tendo uma porção muito menos de bases estando abaixo de 30 na escala phred/phrap e nenhuma abaixo de 20 (Figura 2.).

Figura 2. Score de qualidade das bases após a trimagem.

