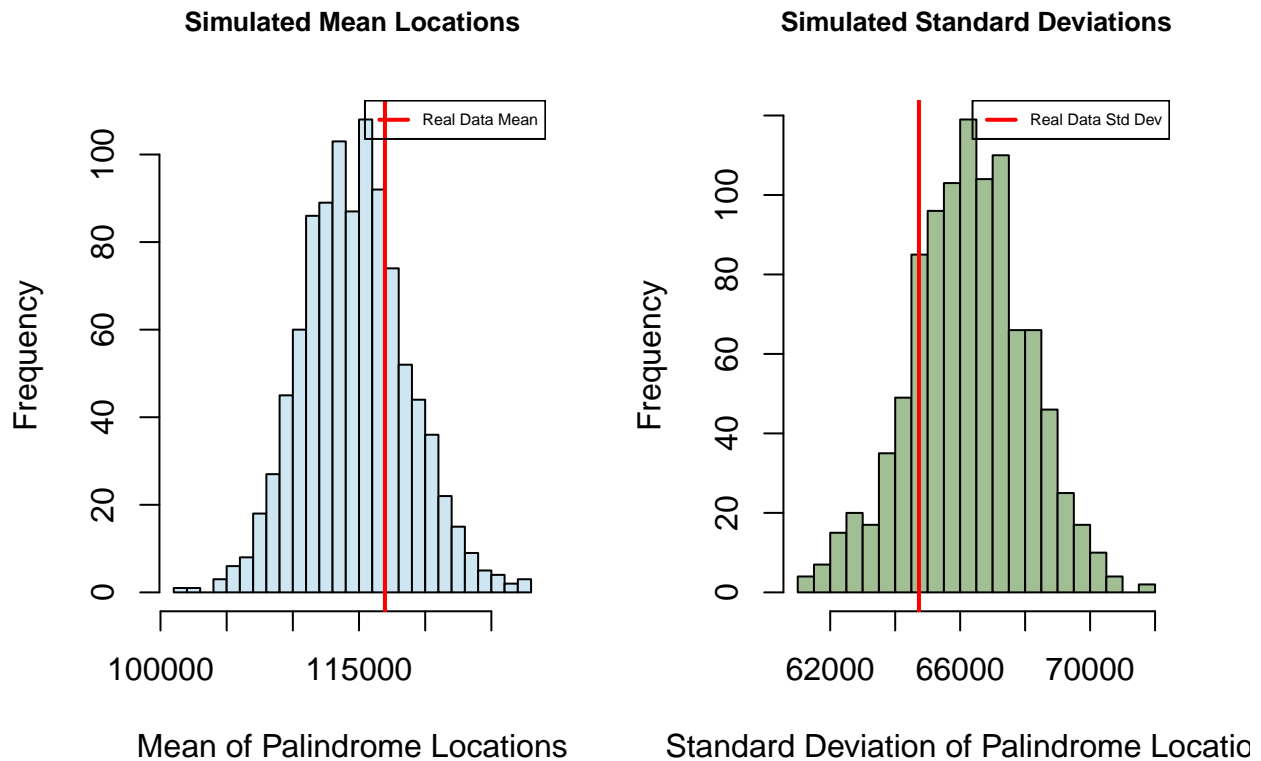


Question 1: Compare Data with Random Scattering of Palindromes

The dataset contains the locations of palindromic sequences in a series of 229,354 base pairs. We will do multiple simulations of random locations of palindromic sites along these 229,354 base pairs and compare it to the real data.

##	Statistic	Real_Locations	Simulated_Locations
## 1	Mean	1.169601e+05	1.147081e+05
## 2	Median	1.178260e+05	1.146895e+05
## 3	Variance	4.190236e+09	4.394407e+09
## 4	Standard Deviation	6.473203e+04	6.626665e+04

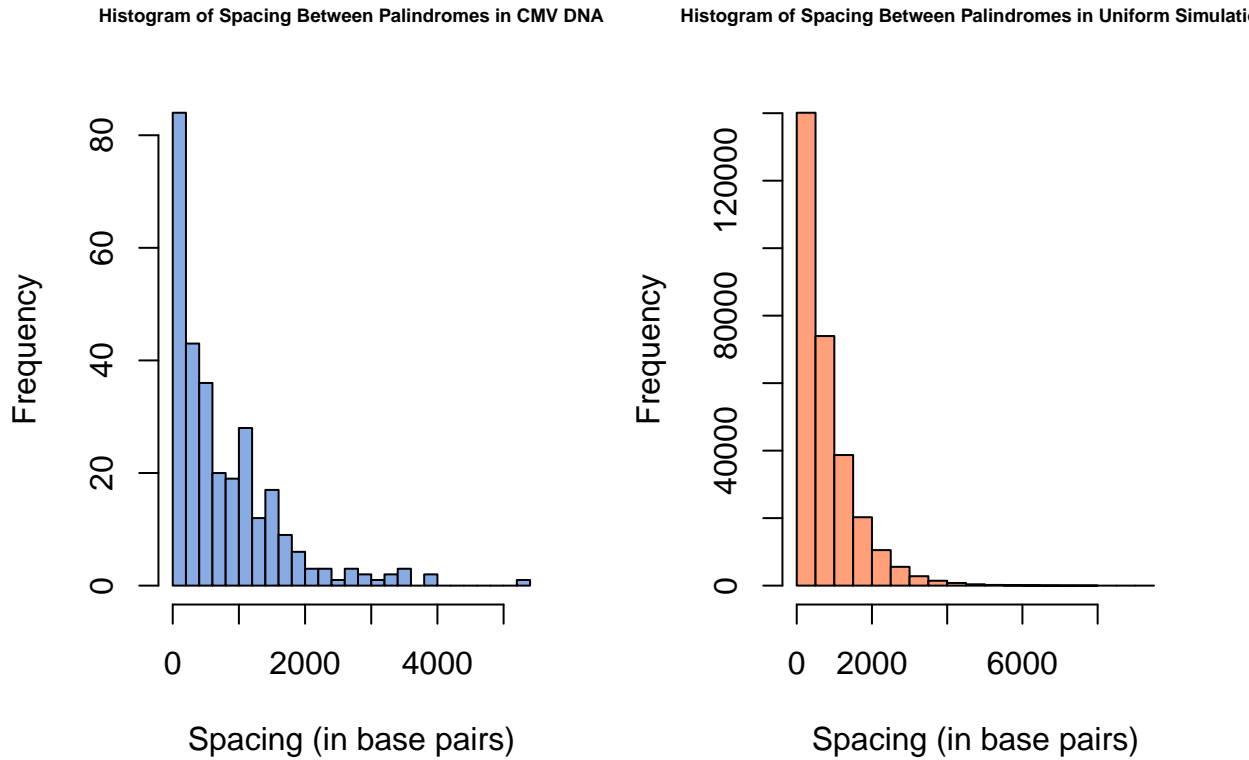
To qualitatively compare the real palindrome locations to the simulated locations, I plot a histogram of the means of each simulated palindrome sequence and analyze where the real mean is on the distribution. The same process is also done for the standard deviations.



Question 2) [Locations and Spacings] Use graphical methods to analyze the patterns in:

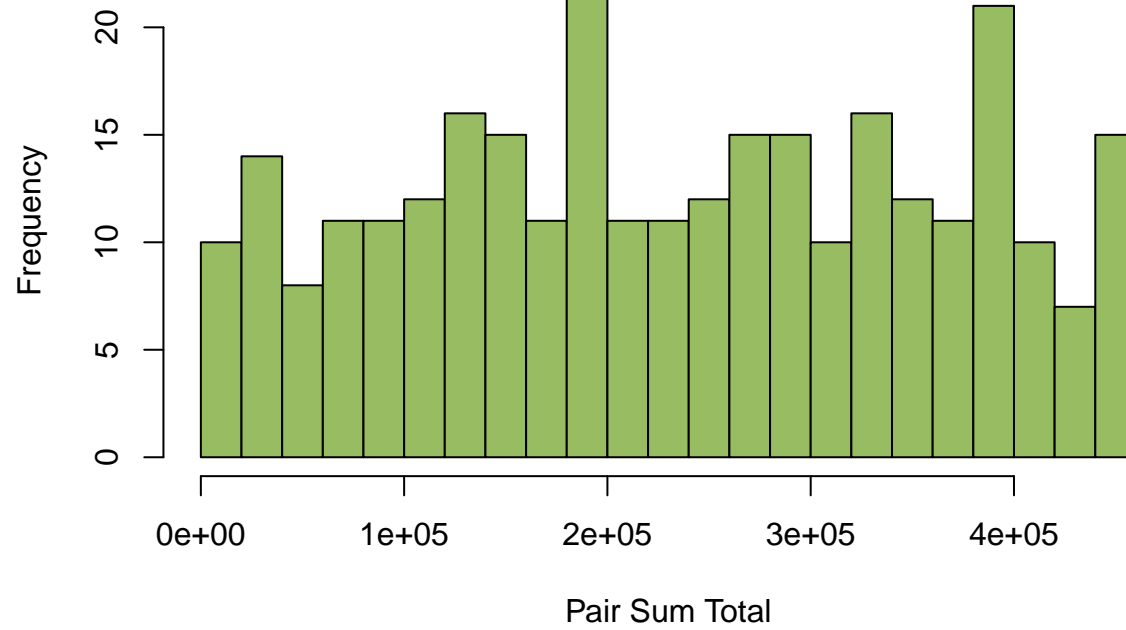
In order to further analyze the distribution of the palindromes, let's see three different variations of the data and how they deviate from a uniform scatter across the DNA sequence. To do this, we will compare the observed patterns below to expected uniform random distributions to identify any significant clusters of unusual spacing in these palindrome locations.

a) Spacing between consecutive palindromes



b) Sums of Palindrome Pairs

Histogram of Pair Sums of Palindromes in DNA



c) Sums of Palindrome Triplets

Histogram of Triplet Sums of Palindromes in DNA

