

Ethan Ducharme

Report 2.3

MTH 332 Gary Davis

Analysis of DNA Palindrome Data

from Stat Labs Data Site A Report

As Evaluation

Conclusion

After reading the metadata and generating a histogram we found that one area of the strand of Deoxyribonucleic acid (DNA) has a relatively high palindrome count while throughout the remaining DNA the count remains Uniform.

Intro

This data analyzed was cloned from a stat labs data site. It was then Imported using R CRAN Libraries along with RStudio as a programming environment. In said environment The data was accessed and assigned to the variable name x.

Analysis

The next step was to run a built in R function for our Chi-squared test of the data on x.

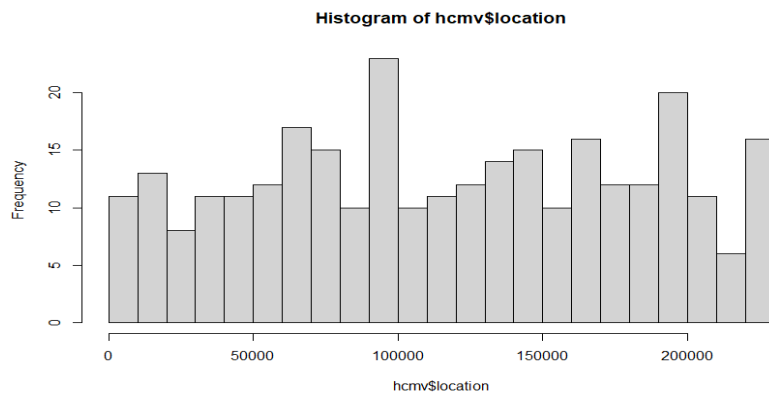
This is what it returned

```
Chi-squared test for given probabilities
```

```
data: x  
X-squared = 10568729, df = 295, p-value < 2.2e-16
```

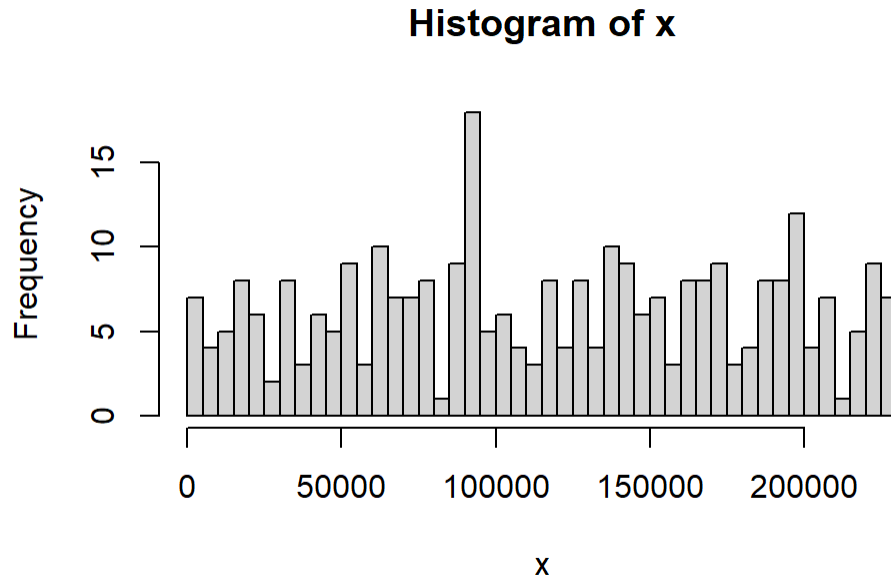
```
< |
```

As you can see the p value is unusually low leading us to believe that the data is significant to the observed difference. As for a visual of the acquired data I choose to plot a Histogram. I did so using the built in R function For histograms labeled hist and I passed it the data and a bin size of 23 and the following was produced by the compiler.



Then I proceeded to investigate further by increasing my breaks/bin-size to around 57.

Again running the hist function the following was produced by the compiler



Now looking at this we get a good picture of a very uniform Data set with a very clear palindrome pattern break below 100000.