**2016 US Presidential Election Results Analysis**



**Name of author**: Hans Peter Ndeffo

**Department**: Mathematics

**Date**: 9/30/2018

**2016 US Presidential Election Results Analysis**

**Name of author**: Hans Peter Ndeffo

**Department**: Mathematics

**Date**: 9/30/2018

**Summary**

The results of US 2016 presidential election surprised many people throughout the world. Hillary Clinton was the favorite candidate with potentially the highest number of voters based on preliminary analysis. Therefore, how can Donald Trump can outclass all the predictions against him and win? The purpose of this report is, consequently, to establish if the outcome of the election has been influenced by Russia hacking or it is natural. Our null hypothesis is the US presidential election has been hacked by Russia.

Our analysis will be split into 5 steps:

* Getting the US 2016 presidential election tally.
* Pre-process the data using numpy, sklearn to check for empty rows or columns.
* Sorting the data using R studio to get only the last digit from 0 to 9 of the total number of voters of each states and count how many times they appear.
* Use the Chi-Squared to see if the vote tally is uniformly distributed based on the previous elections distribution. That will determine if there is a fraud or not with confidence level of 95%
* Conclude

**Contents**

Summary 2

Getting the data 4

Sorting and counting 5

Accepting or rejecting the result using Chi-Squared test 6

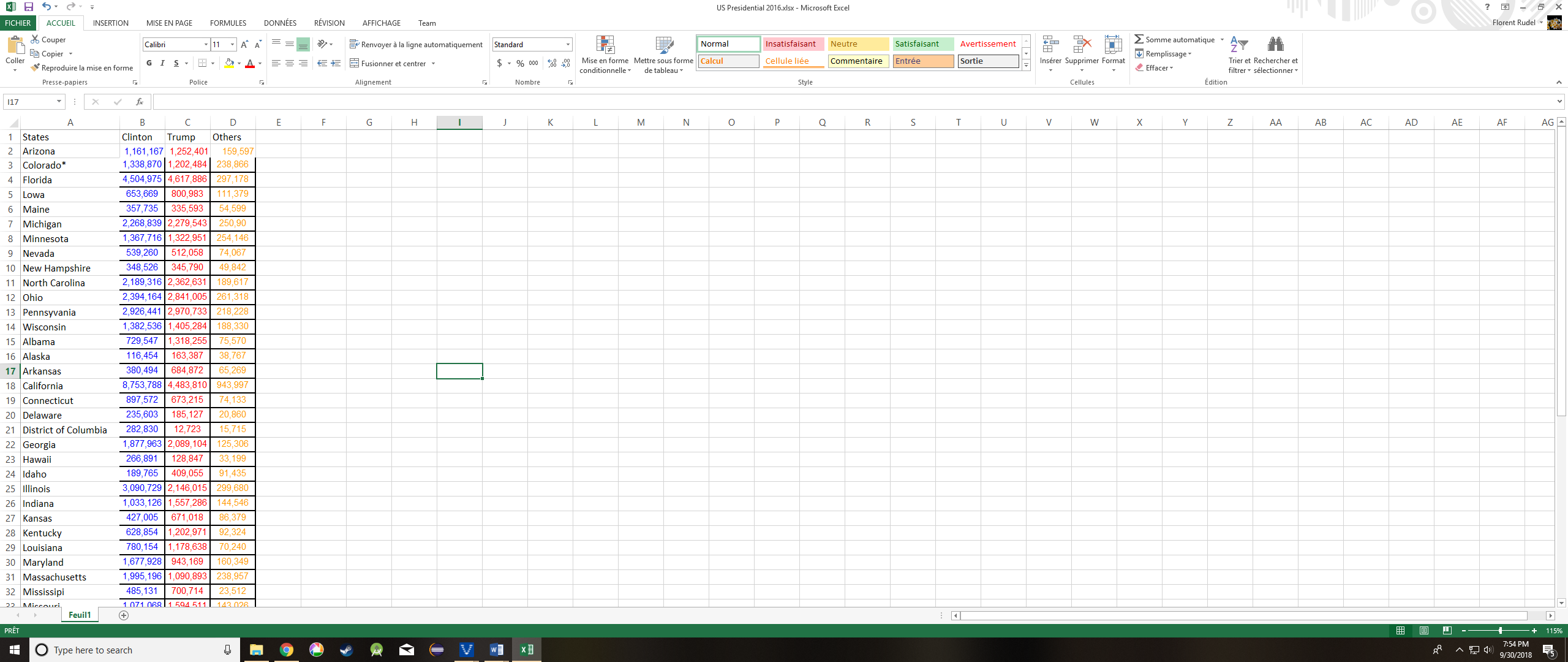
Conclusion 7

Appendix A: 8

References 9

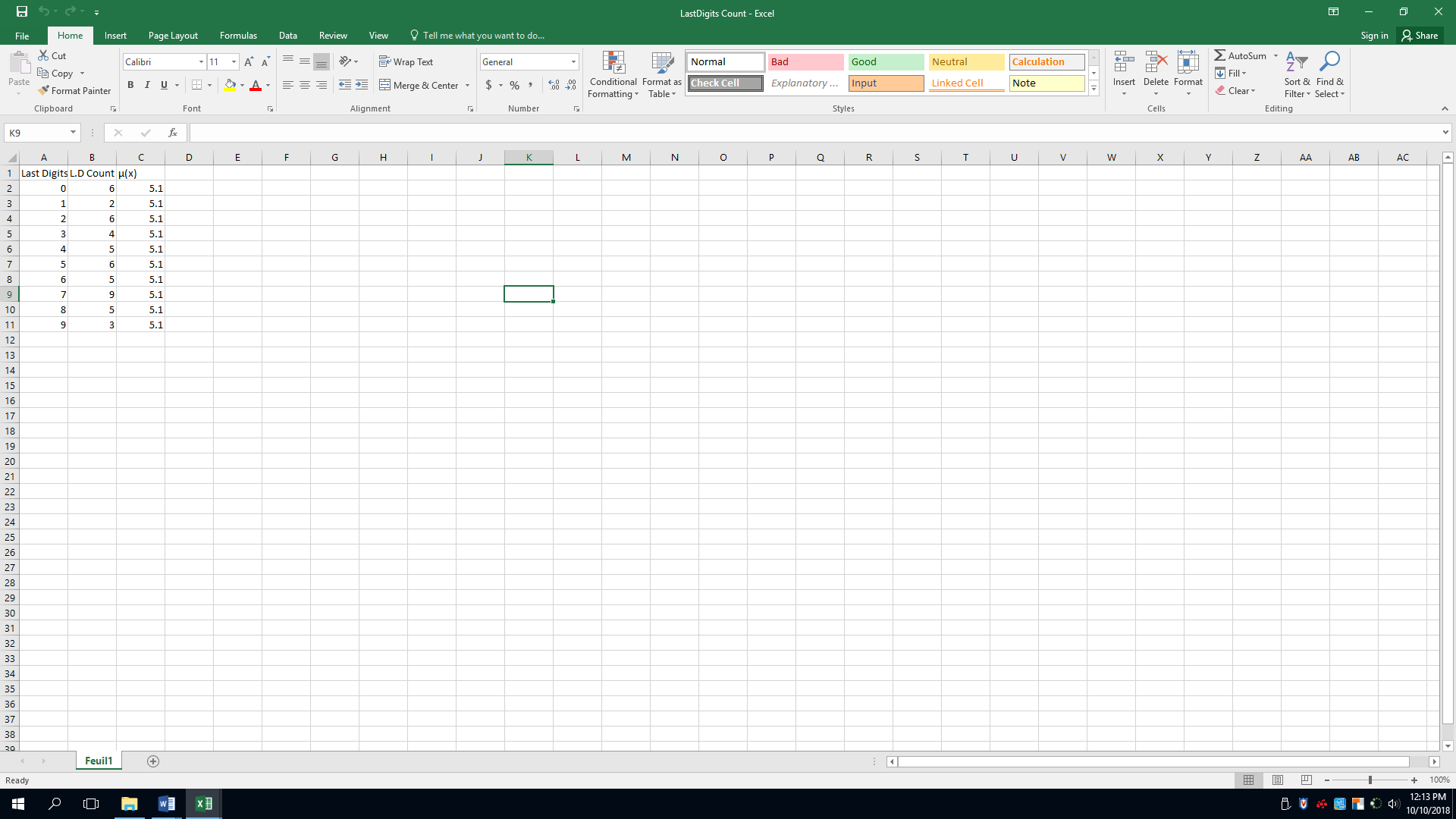
**Getting the data**

The data were downloaded from <https://catalog.data.gov/> in csv file then loaded using excel. The actual numbers look like.



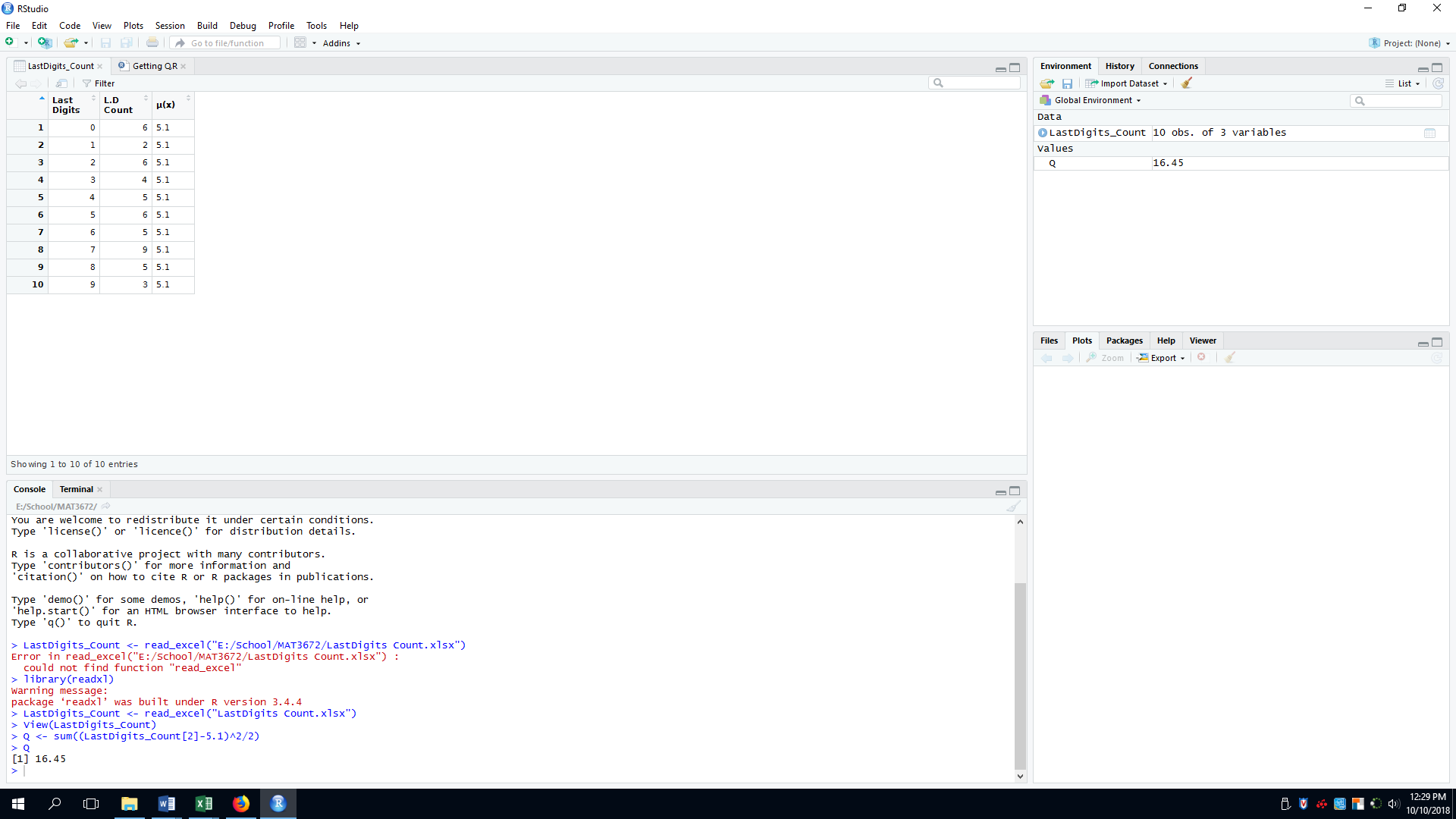
**Sorting and counting**

The data are sorted to get the last digit using the command in excel RIGHT(Cell,1) and counted the number of last digit from 0 to 9 using COUNTIF() command. μ represents the average last digit for all 10 numbers( 0 to 9m).

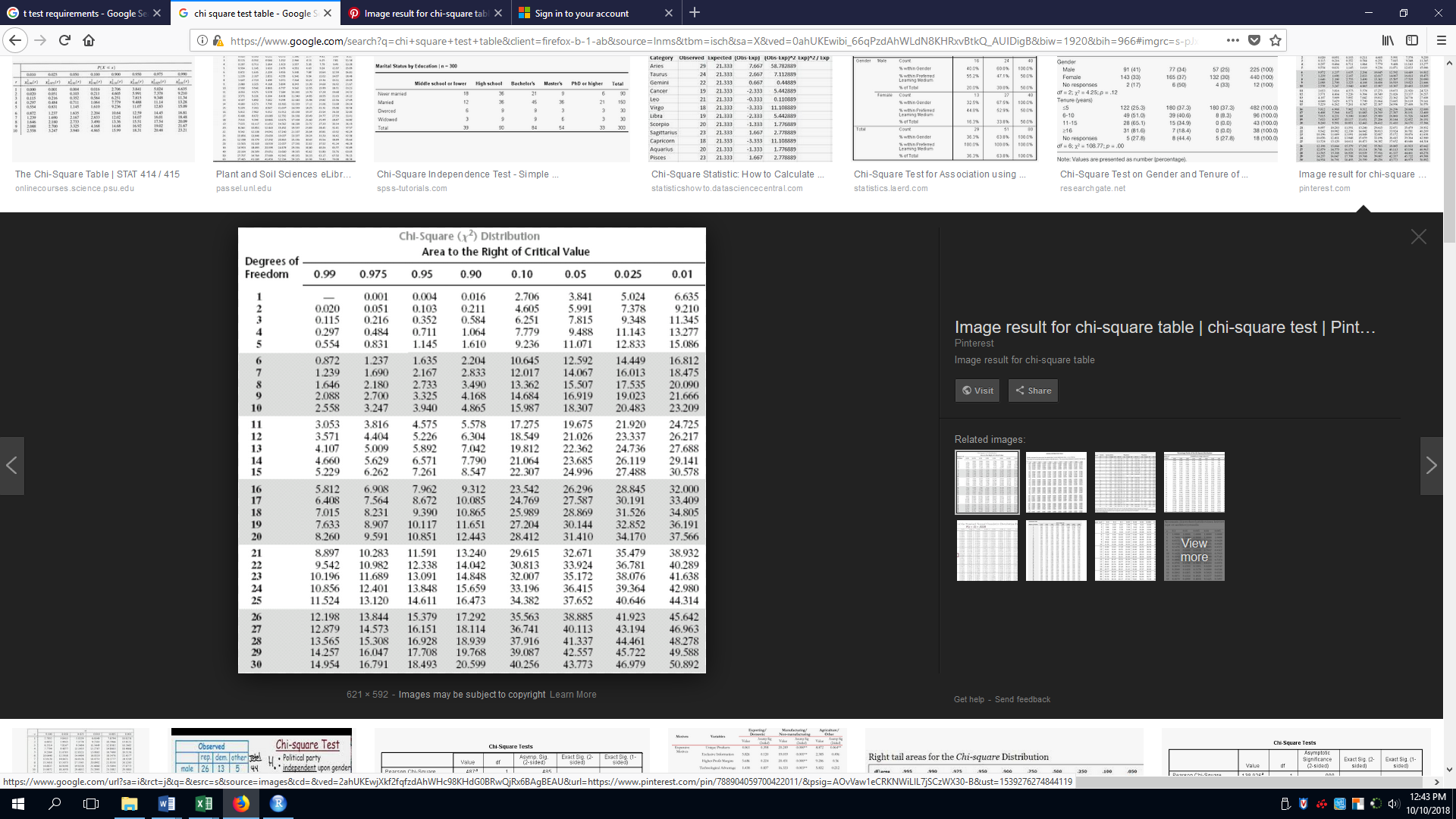


**Chi-Square Test**

Data are imported into then Q is computed. As it can be seen Q = 16.45 and we made 1 assumption. The first is that the data are normally distributed based on previous elections and 1 other degree of freedom is lost naturally. The degree of freedom left are 10 – 2= 8



16.45>15.05 as it is seen the Chi Square table below for 95% confidence



**Conclusion**

Based on the result of the Chi-Square Test, The null hypothesis which was the election was not fair is likely to be true because the votes seem not to be uniformly distributed.