James Cook University

CP2407 - Information Visualization

Assignment One - Frequency Analysis of Text

Name: Kong Wen Yao

Student ID: 13101845

Lecturer: Mr. Randy Zhu

Submission Date: December 23, 2016

**Frequency Analysis of Text Program**

**Introduction**

Frequency analysis of text web page is a simple visualization program done using JavaScript, HTML, and RGraph libraries. This program is designed to allow user to select a text file from their computer and produces a histogram showing the number of occurrences of alphabetical characters from the text context in the file, ignoring cases as well as punctuation within the context. The main goal in coming out with this program is to acquire the experience of implementing information visualization for myself and to design a simple and easy-to-understand frequency analysis tool for the target audience.

**Nature of The Data**

The histogram in this program is composed in total of two variables of data, a numerical variable that measured in ratio scale of measurement and a categorical variable that measured in ordinal scale of measurement. Each observation in the collected data includes the number of occurrences of alphabetical character (frequency) and its corresponding letter (character). Table 1 shown the variables, nature and scale of measurement of data used in displaying the histogram.

Table 1. Variables, Nature and scale of measurement of data.

|  |  |  |
| --- | --- | --- |
| **Variables** | **Nature of Data** | **Scale of Measurement** |
| Alphabetical Character | Categorical | Ordinal |
| Number of Occurrences (frequency) | Numerical | Ratio |

**Target Audience**

The primary purpose or goal of the provision of this program is designed as a tool to allow people, ranging from students to industrial professionals, who are interested in frequency analysis of alphabetical order or needed the program as a tool to perform tasks or studies related to frequency analysis, to simplify or reduce their amount of work needed to perform their task. An example of target audience is those involve in cryptography, who use frequency analysis as a method in their study of letters or cipher text to decrypt or reveal an extent of messages, as stated in “Learn Cryptography” (2016).

**Suitability of This Visualization**

The whole implementation of using histogram in displaying this type of data is one of the approach that can be used. Other approaches like horizontal bar graph or normal bar chart can also be used in my opinion for different preferences of representation of the data in this program. With the limited data used in this program, I will say that I am pleased on the suitability of this visualization using just histogram, despite that certain improvement can still be done in design or colorization of the webpage for visual satisfaction, interactions of the visualization itself, and presentation of the data and context.

With the added CSS styling and enlarged visualization of the graph in this program that I designed in precise and concise manner and with user experience in mind, I believe that the target audience will find the program useful and be satisfy with the easy-to-read and understandable presentation of the data.

**Conclusion**

This report has included the introduction, nature of data, target audience, and suitability of visualization, explaining the data itself as well as sharing reasons of the suitability of visualization in my opinions. All in all, the whole visualization of this frequency analysis web page was interesting to me in term of experiences in planning, coding, implementing, and prototyping. Having no prior experience in information visualization, it was challenging at first in the planning phase of the program itself as well as part of the coding for the program, solving loops and problems in JavaScript. However, I am grateful and pleased with the outcome of my result.

**References**

Frequency Cryptography. (2016). Retrieved from <https://learncryptography.com/cryptanalysis/frequency-analysis>