

Tarea 3

Rómulo Enrique Troncoso Pacheco

Septiembre 2020

1. Introducción

El ploteo es una práctica comúnmente utilizada en revistas científicas, libros y materiales impresos, que sirven al lector para poder identificar de forma visual en forma de gráficas datos con mayor sencillez y comprensión en la exposición de elementos evaluados. Esto, genera la posibilidad de establecer estadísticas con fines comparativos relacionados a la identificación de cantidades de letras y palabras que se puedan presentar en índices numéricos.

De esta forma, en esta práctica académica se establecen letras que mantienen una mayor frecuencia en el libro de *The King in Yellow* by Robert W. Chambers[1], con la finalidad de generar un ejercicio de graficación de frecuencias en términos numéricos. Con la finalidad de determinar una posible correlación entre las letras más frecuentes en el libro, se tomó como punto de referencia un diccionario con las palabras más frecuentes en inglés [7].

La importancia del ploteo es el establecer fundamentos para realizar un análisis de datos con la finalidad de emplearlo en diversas ramas de las ciencias. De esta forma, esta práctica se orienta a una serie de pasos en los cuales se ejecutaron de forma sistemática mediante instrucciones específicas incluyendo siete apartados donde se describe la secuencia a seguir para la utilización del lenguaje de programación utilizado R[2].

2. Herramientas de desarrollo

Dentro de la práctica se utilizaron lenguajes de programación, librerías y editores de textos que consistieron en:

Lenguaje de programación R[2]: utilizado para el enfoque de análisis estadístico originado desde los cimientos del lenguaje S; cuya finalidad, era convertir ideas en software de forma rápida y fiel, por ello, se generaron las instrucciones que llevan una secuencia y lógica de pasos para cumplir algún objetivo en específico, como lo es el ploteo como instrucción.

Librerías: utilizado mediante un método de “librería” que consiste en fragmento de códigos, con el fin de facilitar las tareas al usuario para realizar una tarea en específico, para esta práctica se utilizaron las siguientes como `gutenbergr`[8], `tidytext`[3], `dplyr`[5] y `readr`[4].

Editores de texto: utilizado el Visual studio code[6], que ayuda a plantear el manejo del texto del scrip.

3. Metodología

En la práctica se describen las letras con mayor impacto en el libro mencionado con antelación, con la finalidad de realizar una descripción exacta de los elementos, relacionado de forma sistemática de acuerdo a la técnica de filtrado que auxilió a concretar elementos exactos descriptivos en función a las letras del libro analizado. Aunado a ello, se analizó un total de 1000 palabras que son mayormente utilizadas en el idioma inglés, con el fin de comparar los resultados de frecuencia del libro con los resultados de frecuencia de las palabras del diccionario.

Para ello, se utilizó las gráficas de barras como herramienta de ploteo que permitió establecer la altura de cada una, en función a la proporción de la frecuencia obtenida del elemento estudiado.

4. Objetivo

Representar gráficamente la frecuencia de letras en el libro The King in Yellow by Robert W. Chambers[1]. Representar gráficamente la frecuencia de las letras de las palabras del diccionario.

4.1. Objetivo específico

Graficar letras filtradas relacionadas a la frecuencia del libro The King in Yellow by Robert W. Chambers[1]. Graficar letras filtradas relacionadas a la frecuencia de las palabras del diccionario. Comparar los resultados de los objetivos específicos previamente mencionados.

5. Procedimiento

Para poder llevar a cabo el objetivo planteado de la práctica fue necesario realizar lo siguiente de acuerdo a las herramientas mencionadas anteriormente:

- Cargar el libro
- Separar letras
- Graficar letras filtradas
- Cargar diccionario
- Separar letras de las palabras
- Crear tabla de frecuencia de letras del diccionario

- Ordenar de forma decreciente las frecuencias
- Graficar las palabaras

6. Conclusión

1. Se pudo graficar la frecuencia de las letras del libro analizado, referido en la sección de salida (Figura 1).
2. Se pudo representar gráficamente la frecuencia de las letras de las palabras del diccionario (Figura 2).
3. Se logró comparar los resultados de la gráfica de frecuencia del libro con la gráfica de frecuencia del diccionario.
4. Se llegó a la conclusión de que las letras a, e, tuvieron un comportamiento similiar, puesto que en ambas gráficas aparecen como las letras con mayor frecuencia; por otra parte, las letras k, v, en ambos casos, fueron las que menor frecuencia tuvieron.

7. Salida

En esta sección se muestran las salidas del sistema donde se incluyen las figuras 1 y 2:

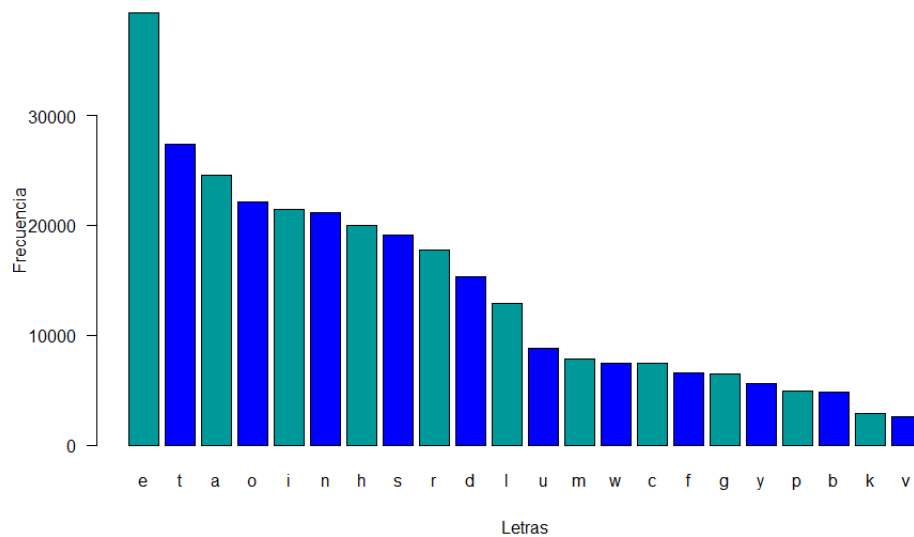


Figura 1: En la figura se puede apreciar las letras con mayor frecuencia en libro analizado con una restricción de presentación a letras mayores a 1000 repeticiones.

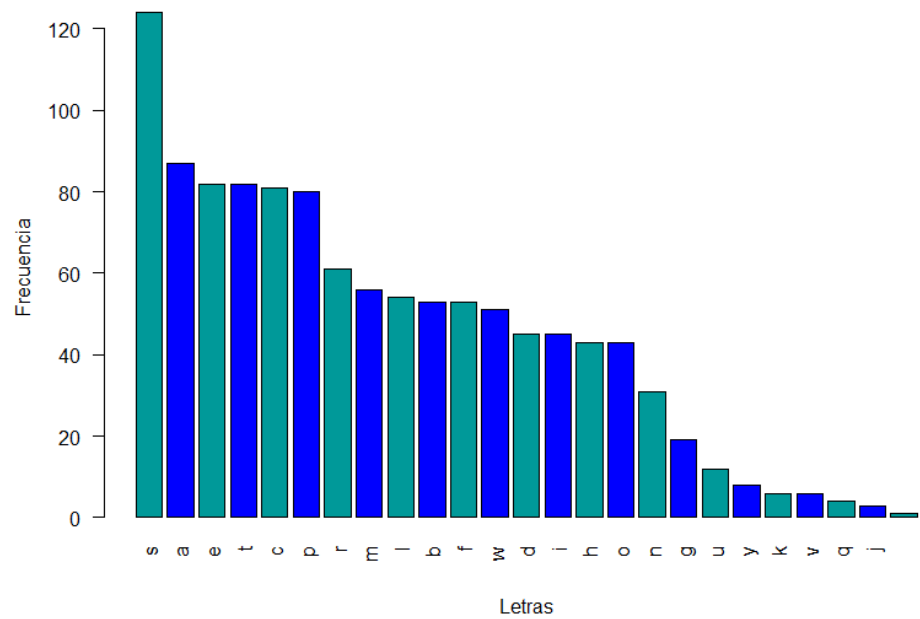


Figura 2: En la figura se aprecia las letras con más repeticiones en el diccionario analizado con una restricción de búsqueda a las 1000 palabras con mayor frecuencia en el idioma inglés.

Referencias

- [1] Robert W. Chambers. *The King in Yellow*. 2005.
- [2] The R Foundation. The r project for statistical computing. <https://www.r-project.org/>, 2020.
- [3] Emil Hvitfeldt Os Keyes ORCID Kanishka Misra Tim Mastny Jeff Erickson David Robinson Julia Silge ORCID Gabriela De Queiroz, Colin Fay ORCID. Tidytext v0.2.5. <https://cran.r-project.org/web/packages/tidytext/index.html>, 2020.
- [4] David W Gerbing. Read v3.7.6. <https://www.rdocumentation.org/packages/lessR/versions/3.7.6/topics/Re> 2020.
- [5] Lionel Henry Kirill Müller ORCID RStudio Hadley Wickham ORCID, Romain François ORCID. Dplyr v1.0.2. <https://cran.r-project.org/web/packages/dplyr/index.html>, 2020.
- [6] Microsoft. Visual studio code - code editing. <https://code.visualstudio.com/>, 2020.
- [7] Corpus of Contemporary American English (COCA). Word frequency data. <https://www.wordfrequency.info/>, 2019.
- [8] David Robinson. Gutenbergr v0.1.5. <https://www.rdocumentation.org/packages/gutenbergr>, 2020.