

Tarea 2 Graficación de palabras y letras

Rómulo Enrique Troncoso Pacheco

September 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 y palabras que mantienen una mayor frecuencia en el libro de *The King in Yellow* by Robert W. Chambers[5], con la finalidad de generar un ejercicio de graficación de frecuencias en términos numéricos. 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[6].

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[6]: 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 generó 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 consistente 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`[1], `tidytext`[2] y `dplyr`[3].

Editores de texto: utilizado el Visual studio code[4], que ayuda a plantear el manejo del texto del scrip.

3 Metodología

En la práctica se describen las palabras con mayor impacto en el libro mencionado con antelación, con la finalidad de realizar una descripción exacta de elementos, relacionado de forma sistemática de acuerdo a la técnica de filtrado que auxilió a concretar elementos exactos descriptivos en función a las letras y palabras del libro analizado.

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 las frecuencias de palabras y letras en el libro The King in Yellow by Robert W. Chambers[5].

4.1 Objetivo específico

Graficar palabras y letras filtradas relacionadas a la frecuencia del libro The King in Yellow by Robert W. Chambers[5].

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 palabras y letras.
- Graficar palabras.
- Graficar palabras filtradas.
- Graficar letras.
- Graficar letras filtradas.

6 Conclusión

1.- Se pudo graficar las frecuencias relacionadas a las palabras y letras del libro analizado, referido en la sección de salida.

2.- Se pudo especificar de acuerdo a las herramientas y técnicas utilizadas la función de graficación para identificar las apartados correspondientes de palabras y letras repetidas con mayor frecuencia en el libro analizado.

3.- Relacionado al ploteo de las frecuencias de letras la "e" es la que tiene mayor frecuencia y la menor corresponde a la letra "z", señalando que el número con menor frecuencia corresponde al "6".

4.- Se pudo identificar mediante la base de datos que la palabra con mayor frecuencia corresponde a "the", como se observa en la figura 3 y con la menor frecuencia la palabra "yell".

7 Salida

En esta sección se muestran las salidas del sistema donde se incluyen las siguientes gráficas:

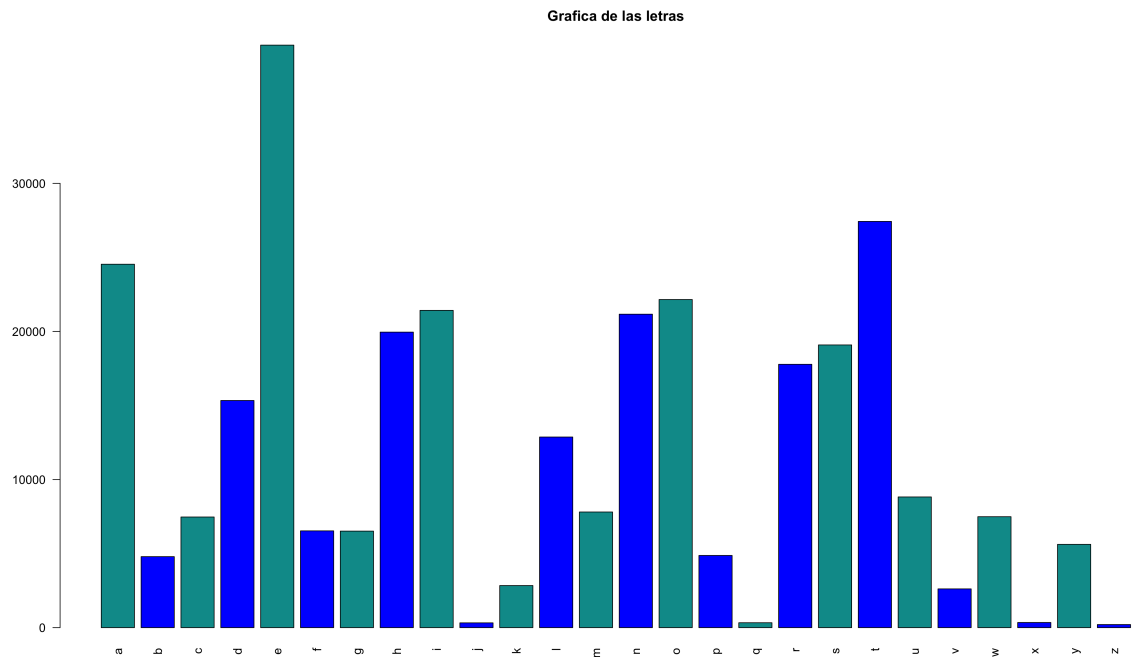


Figure 1: En esta gráfica se expone todas las letras utilizadas en el libro, con establecimiento de rango superior a 100 sin ordenamiento.

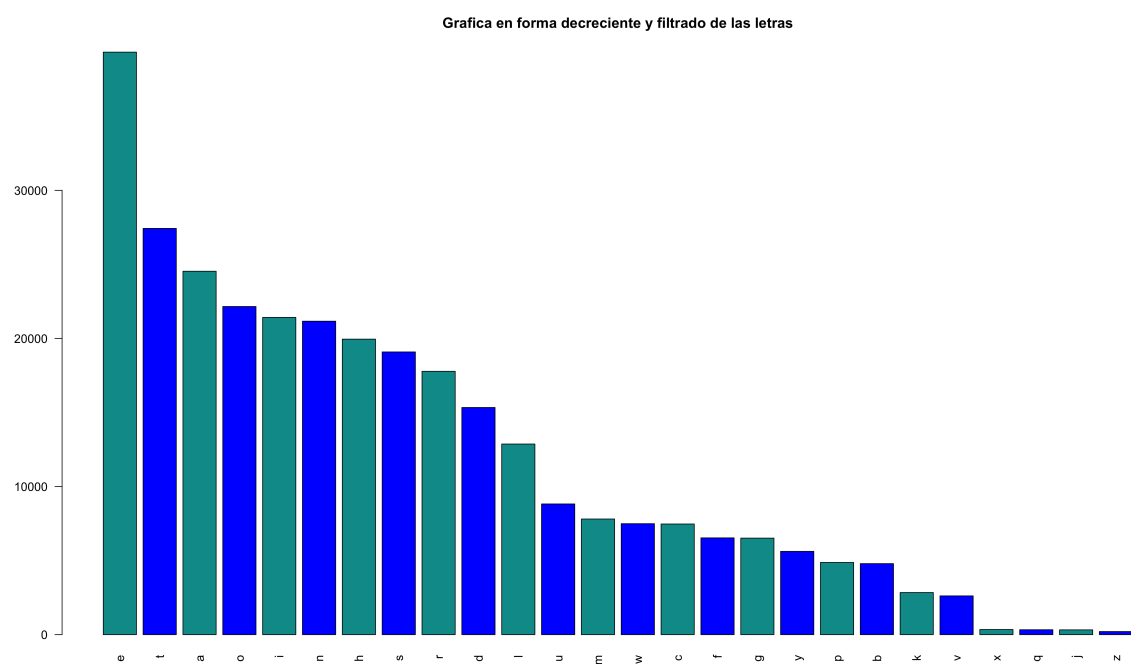


Figure 2: En esta gráfica se expone las letras de forma decreciente en rangos superior a 100

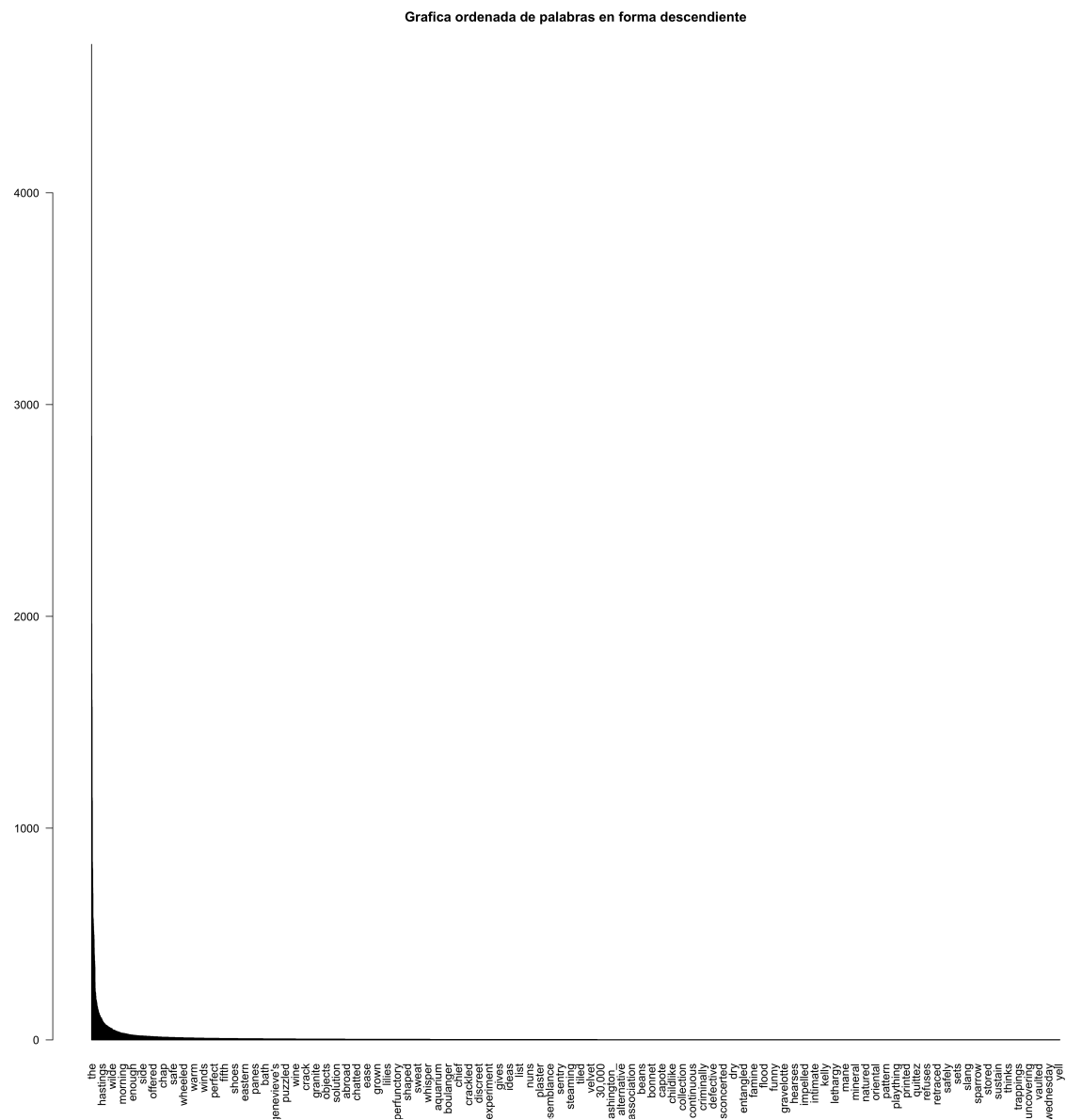


Figure 3: En esta gráfica se expone todas las palabras en orden descendiente.

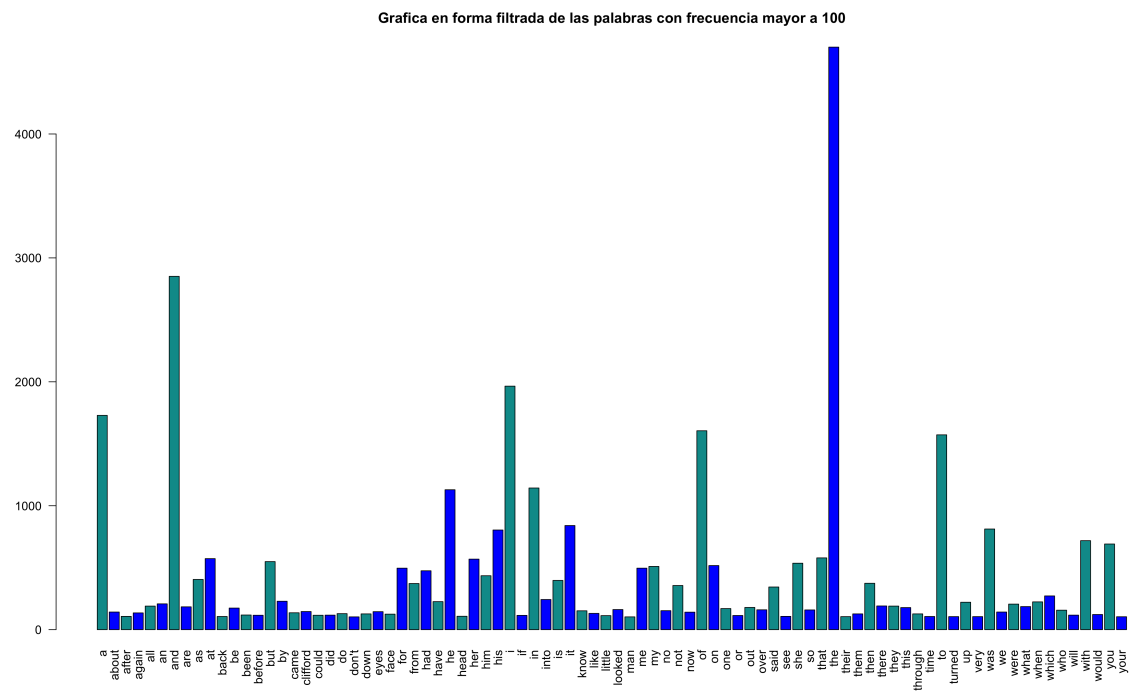


Figure 4: En esta gráfica se expone el filtrado de palabras mayor a 100.

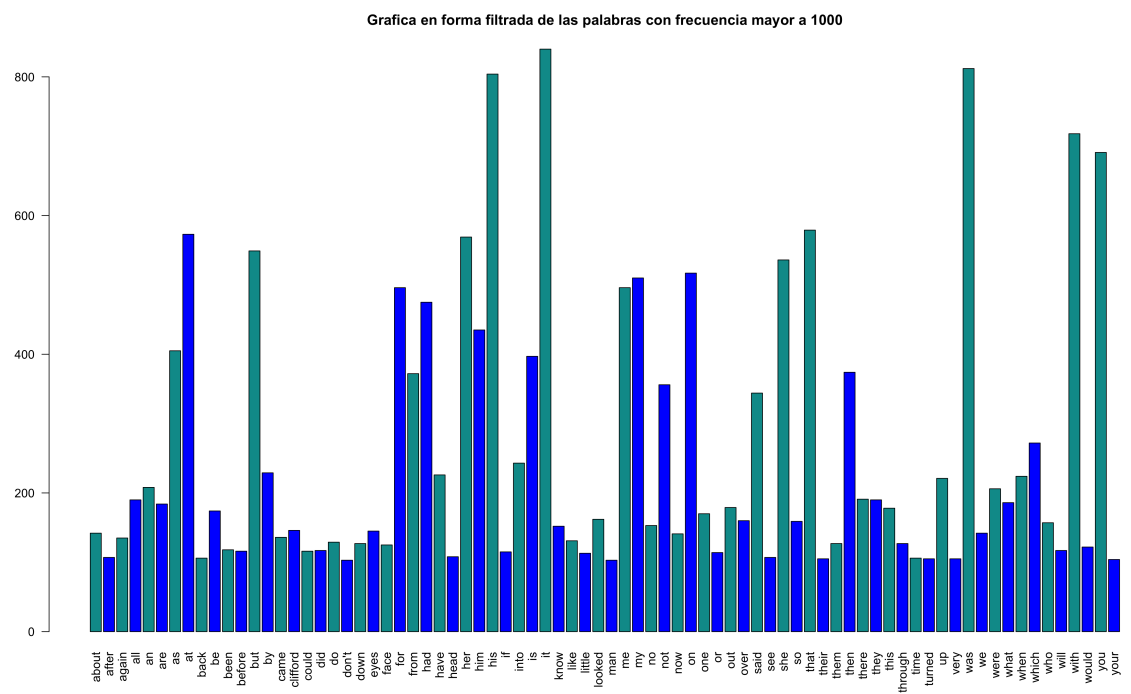


Figure 5: En esta gráfica se expone el filtrado de palabras mayor a 1000 sin ordenamiento.

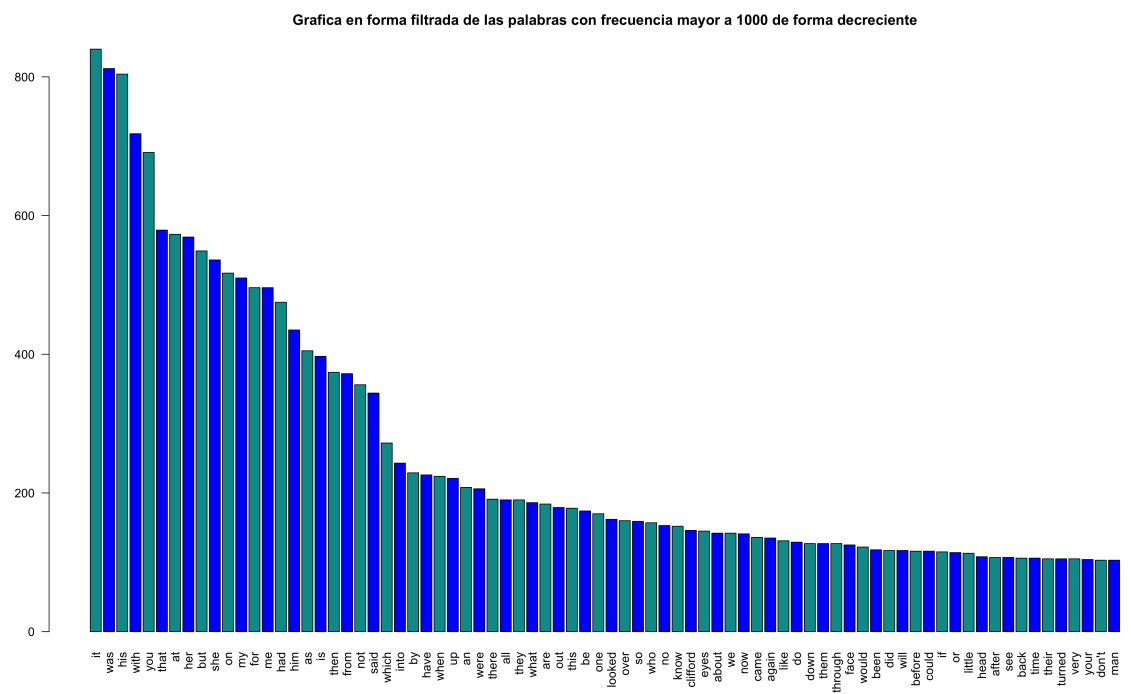


Figure 6: En esta gráfica se expone el filtrado de palabras mayor a 1000 de forma ordenada.

References

- [1] David Robinson. Gutenbergr V0.1.5. <https://www.rdocumentation.org/packages/gutenbergr>, 2020.
- [2] Gabriela De Queiroz, Colin Fay ORCID, Emil Hvitfeldt, Os Keyes ORCID , Kanishka Misra, Tim Mastny , Jeff Erickson, David Robinson, Julia Silge ORCID . Tidytext V0.2.5. <https://cran.r-project.org/web/packages/tidytext/index.html>, 2020.
- [3] Hadley Wickham ORCID, Romain François ORCID, Lionel Henry, Kirill Müller ORCID, RStudio. Dplyr V1.0.2. <https://cran.r-project.org/web/packages/dplyr/index.html>, 2020.
- [4] Microsoft. Visual Studio Code - Code Editing. <https://code.visualstudio.com/>, 2020.
- [5] Robert W. Chambers. The King in Yellow. <http://www.gutenberg.org/ebooks/8492>, 2005.
- [6] The R Foundation. The R Project for Statistical Computing. <https://www.r-project.org/>, 2020.